Extreme Pathways in the Post-Genome Era

Bernhard Palsson

Lecture #8

September 16, 2003

University of California, San Diego
Department of Bioengineering

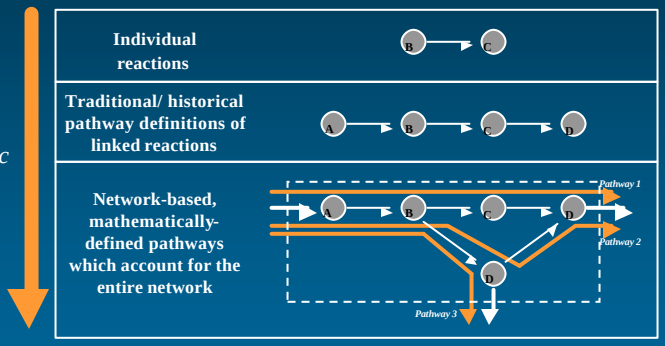
Systems Biology Research Group
http://systemsbiology.ucsd.edu

Outline

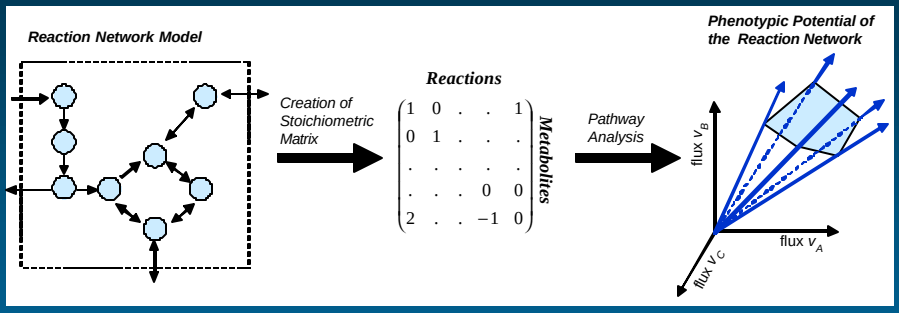
1) Overview of extreme pathways
2) Extreme pathways of human red blood cell metabolism
3) Extreme pathway analysis of *H. influenzae*
4) Extreme pathway analysis of *H. pylori*
5) Methods for analyzing large sets of extreme pathways
6) Interpretation of transcriptional regulation with extreme pathway analysis

Development of the network-based pathway paradigm

Increase in complexity and systemic relevance



Extreme pathways capture the phenotypic potential of metabolic reaction networks



Extreme pathways have the following characteristics:

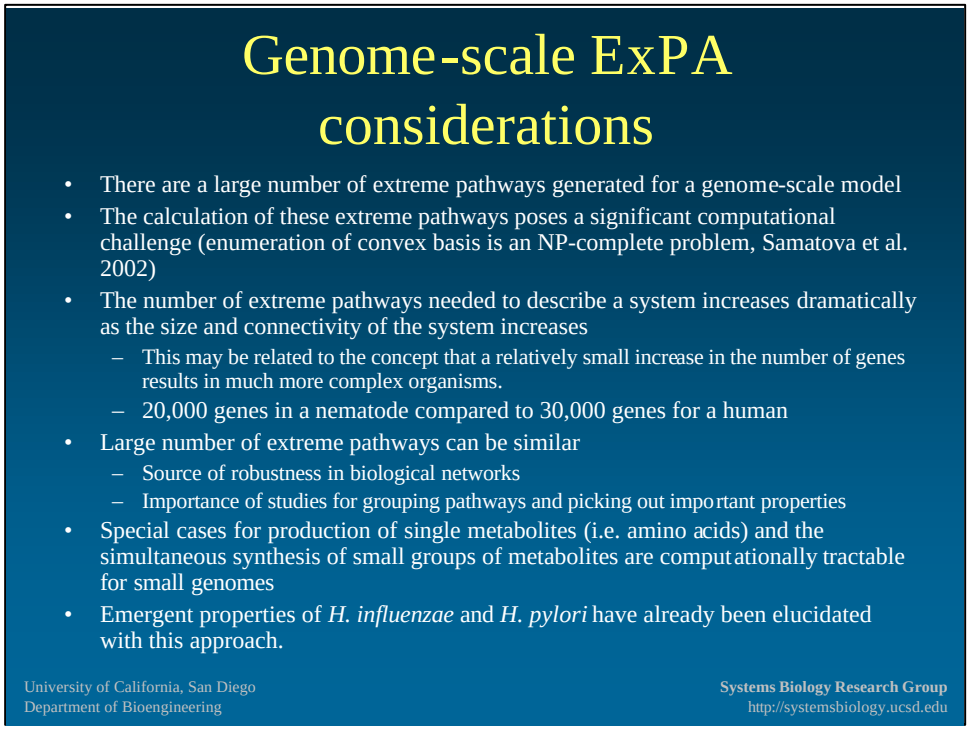
1) They are a unique and minimal set of basis vectors

2) All possible phenotypes can be represented by non-negative linear combinations of the EPs

3) They represent time-invariant properties of the metabolic network

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu



First, some important considerations for extreme pathway analysis.

It is first important to remember the ExPA of genome-scale networks results in a tremendously large number of data. The calculation of these vectors pose a significant computational challenge. In fact the enumeration of a convex basis has been classified as an NP-complete problem…the size of the metabolic networks exponentially increases the computational time to perform the extreme pathway calculation. This idea of increasing complexity is an interesting point when looking at the complexity of organisms with small increases in the number of genes. For example, we see the difference in the number of genes from a nematode to a human, 20,000 genes to 30,000 genes.

An additional consideration for extreme pathway analysis at a genome-scale lies in the properties of the pathways themselves. There are many pathways with only subtle differences in certain reactions that are used. This property leads to the idea of robustness in metabolic networks. For this reason there is a significant need to look at methods for parsing out important properties.

To date, the extreme pathways for amino acid production and ribonucleotide synthesis in H. influenzae and H. pylori have been computed, and emergent properties have been characterized.

Approaches to genome-scale ExPA

- Due to computational intractability, models have previously been segmented into biologically meaningful groups (i.e. amino acid synthesis, nucleotide synthesis; Schilling and Palsson 2000).
- Recent efforts have been made to algorithmically define these groups (Schuster et al. 2002).
- However, the subdivision of the networks excludes the combinatorial possibilities that occur with the interaction between the subdivisions.
- Recent work has involved the analysis of integrated cellular models (Papin et al. 2002, Price et al. 2002).

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

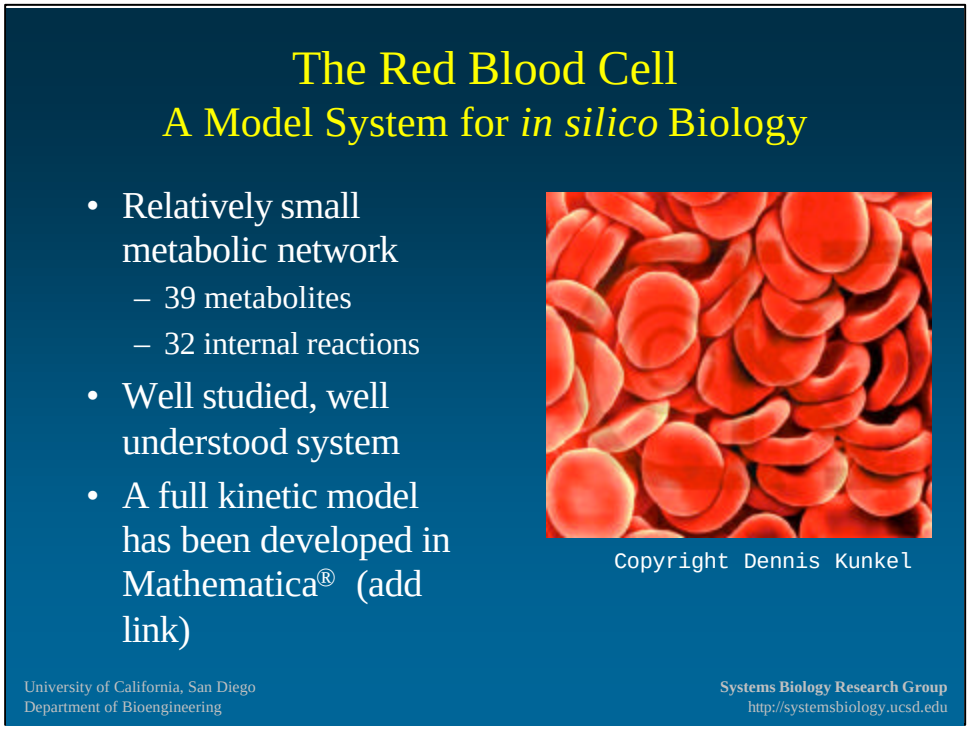
One approach to overcome computational intractability has involved segmenting the networks into subsystems that have biological meaning. For example, a full metabolic network can be broken up into reactions associated with amino acid synthesis, central metabolism, nucleotide synthesis, lipid metabolism, etc. The extreme pathways of each of these subsystems can be computed and the interacting metabolites between these subsystems can be accounted for. However, this approach involves arbitrary grouping and neglects the combinatorial possibilities that arise from the interactions of components of the subsystems.

In order to more precisely define these subsystems, recent work has involved mathematically defining how these subsystems are defined. Recent work has also recently been published in which the entire genome is analyzed under minimal medium conditions.

Extreme pathways of human red blood cell metabolism

University of California, San Diego
Department of Bioengineering

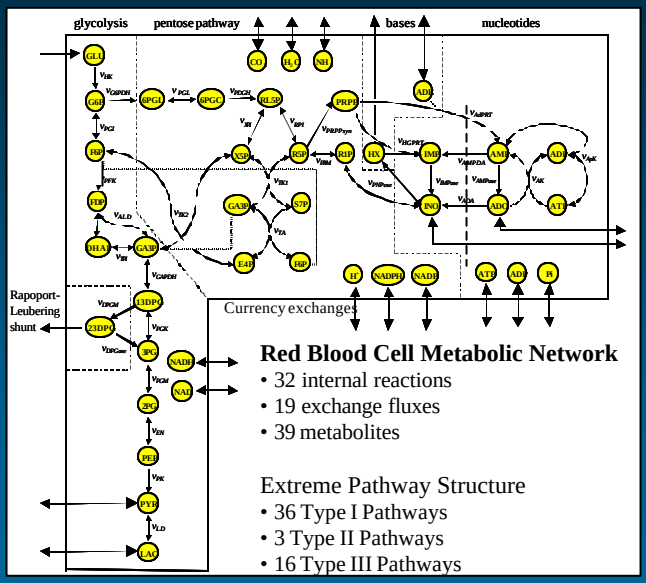
Systems Biology Research Group
http://systemsbiology.ucsd.edu



The Human Red Blood Cell

Previously, extreme pathway analysis was applied to sample systems without real biological meaning (e.g. Schilling et al., 2000). Such systems helped in establishing the algorithm and interpreting the results but provided no real biological insight. At the other extreme, the analysis has been applied to genome-scale metabolic network resulting in an immense number of extreme pathways for which a detailed interpretation is not possible. Only statistical properties of this large set of data could be obtained yielding limited insight into cellular physiology (Papin et al., 2002). The human red blood cell provides an attractive case to study the extreme pathways. Many people argue that you will never know all the kinetics of most cells so a full cell simulation is not possible. The RBC is the exception to this rule. Hence we can test theories of model reduction by working on simplified version of the RBC in which we focus on only the most important parameters of the system. This reduced model can then be compared to the full model to see if the simplified representation is a good approximation of the full cell simulator.

Extreme pathways in the red blood cell



Red Blood Cell Metabolic Network

- 32 internal reactions
- 19 exchange fluxes
- 39 metabolites

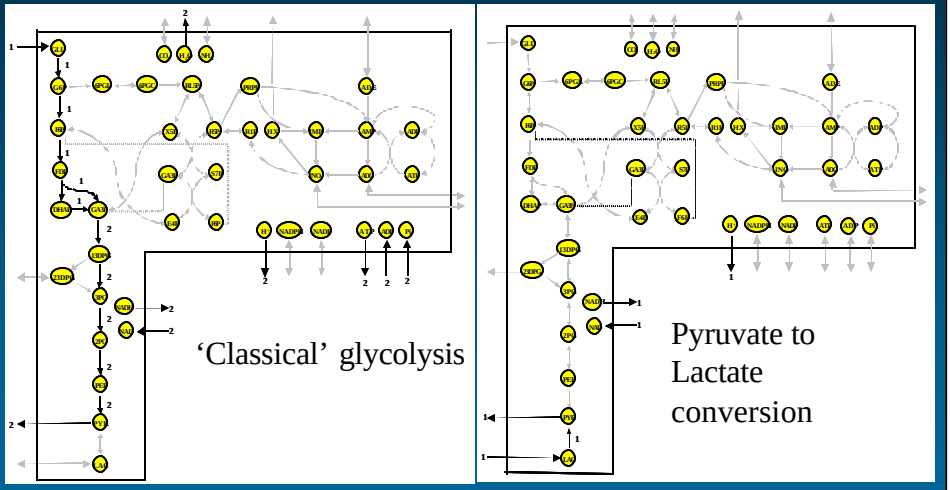
Extreme Pathway Structure

- 36 Type I Pathways
- 3 Type II Pathways
- 16 Type III Pathways

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Extreme pathways in the red blood cell



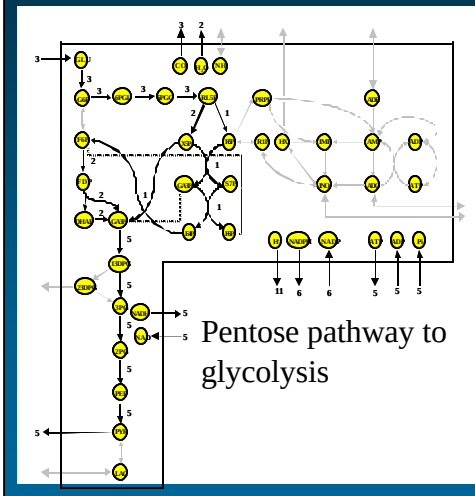
‘Classical’ glycolysis

Pyruvate to Lactate conversion

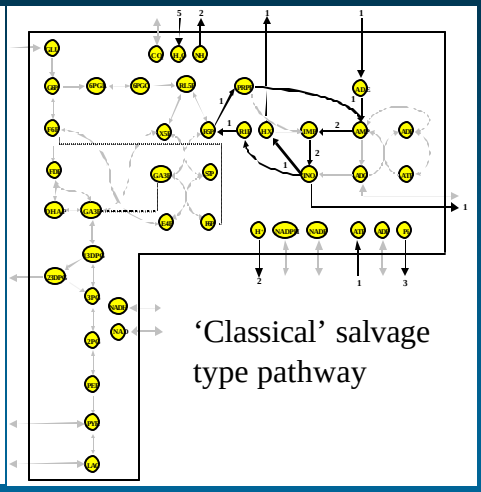
University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Extreme pathways in the red blood cell

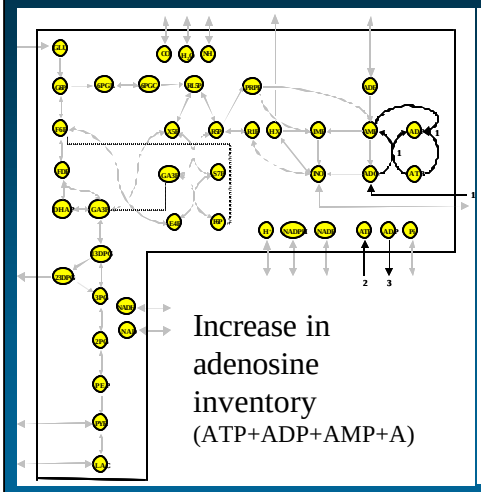


Pentose pathway to glycolysis

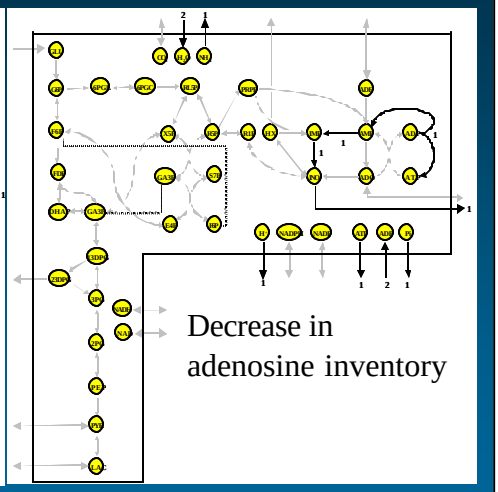


'Classical' salvage type pathway

Extreme pathways in the red blood cell

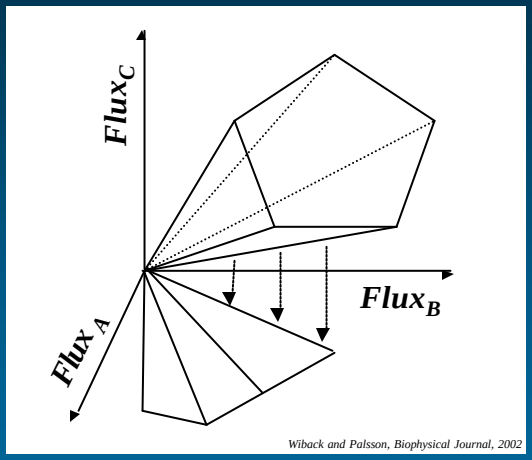


Increase in adenosine inventory
(ATP+ADP+AMP+A)



Decrease in adenosine inventory

Phase Planes as projections of high dimensional cones

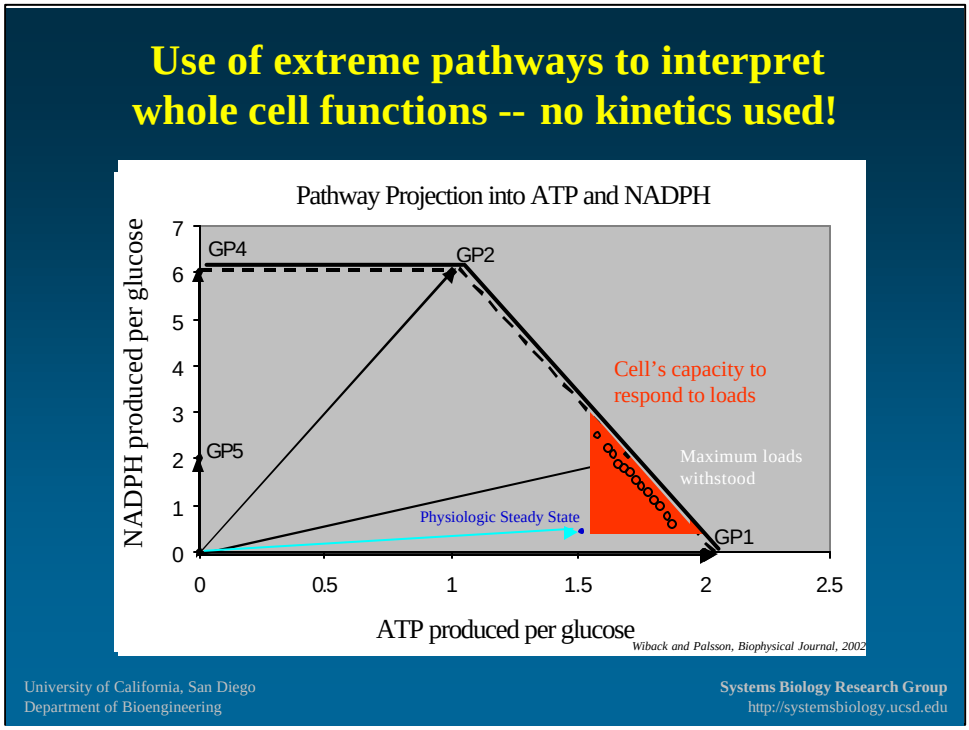


University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

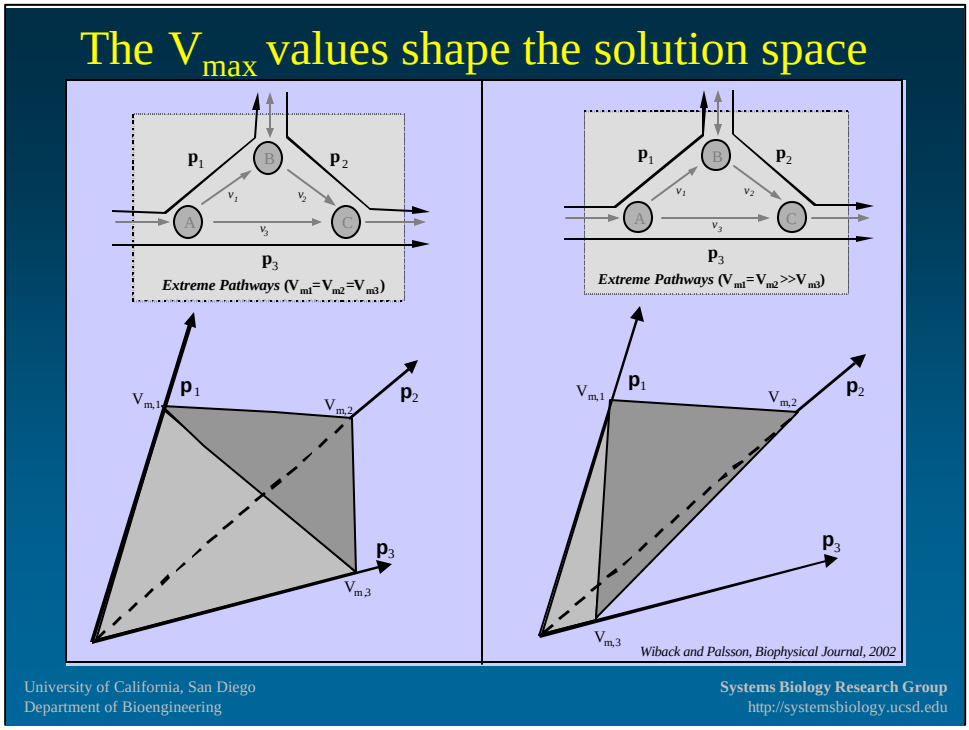
Projection into lower dimensions

As said before, the extreme pathways form the edges of a high-dimensional flux cone which encompasses all steady state flux distributions attainable by the network. Since it is hard for most people to think and see above 3-D, we often project these high-dimensional cones into lower dimensions (either 2D or 3D) - particularly dimensions of interest.



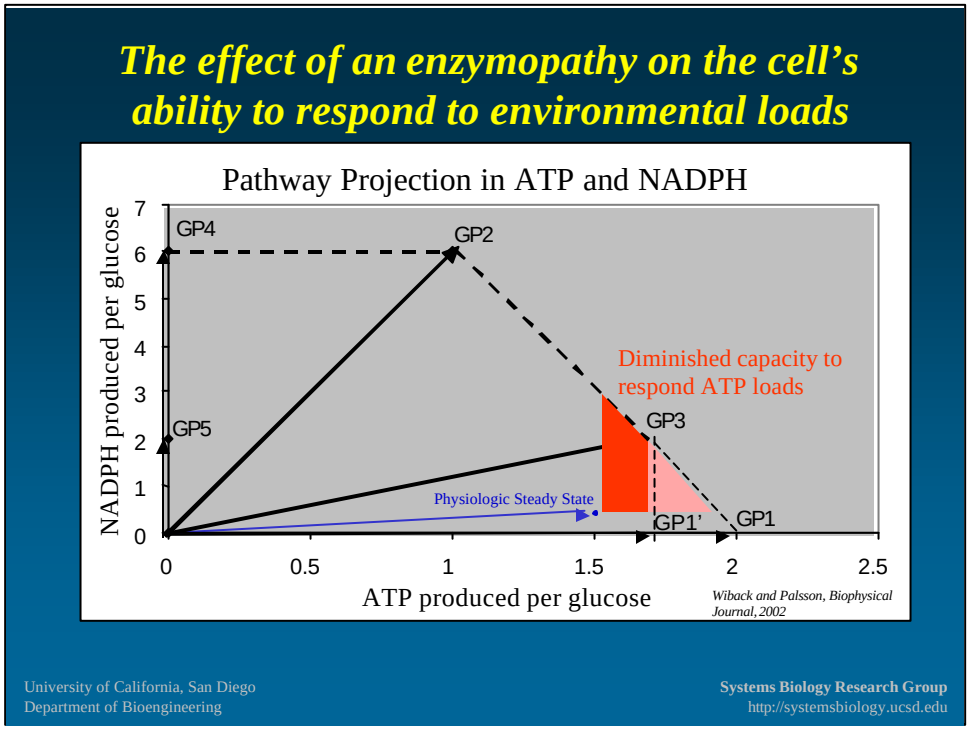
Projection of Pathways Based on Production of Key Cofactors

Here is a projection of the red cell pathways into a 2D flux space with the key cofactors ATP and NADPH production on the axis. The projected EPs show the ability of each pathway to make the given cofactors per glucose uptake. These pathways define the attainable region – outside of which there is no feasible solution for the cell. This brings us back to the kinetic red cell model that I spoke of earlier. With this model we are in the unique position of being able to find the "exact" solution point. The nominal steady state value is shown with the blue arrow. The red blood cell's capacity to respond to loads (red region) is defined as the difference between the steady state operating point (blue dot) and the edge of the solution space representing the maximum capabilities of the cell (black dotted line). Any loads outside the solution space are not attainable. The results from repeated dynamic simulation of stepwise increasing energy and oxidative loads on the red blood cell are plotted on the graph with open black circles using the Jamshidi model (Jamshidi et al., 2001). Note that the kinetic model is slightly more restrictive than the stoichiometric one.



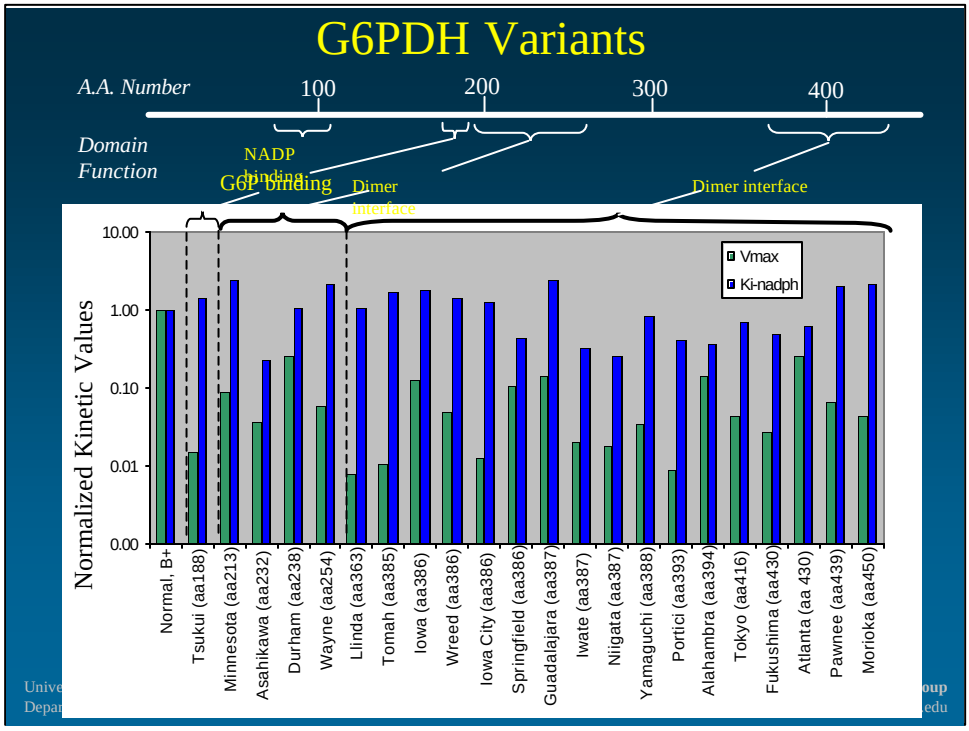
Maximal Flux Capacities

The V_{max} values of the enzymes serve to "cap off" the steady state solution cone. Changes or alterations in these V_{max} values (due to enzyme defects or SNPs) can significantly change the shape of the steady state solution space. If all the extreme pathways have high throughput, the solution space is relatively large as shown on the left (A). However, as shown on the right (B), if one of the V_{max} values is low due to some sort of defect or significant kinetic regulation, there is a shift down the $\mathbf{p_3}$ axis thus significantly shrinking the size and volume of the steady state solution cone and hence the metabolic capabilities of the system. The volume of the solution space shrinks significantly which reduces the number of steady state solutions and hence the number of homeostatic options available to the cell. Thus V_{max} values can effectively reduce the solution space and eliminate a large number of possible states of the network.

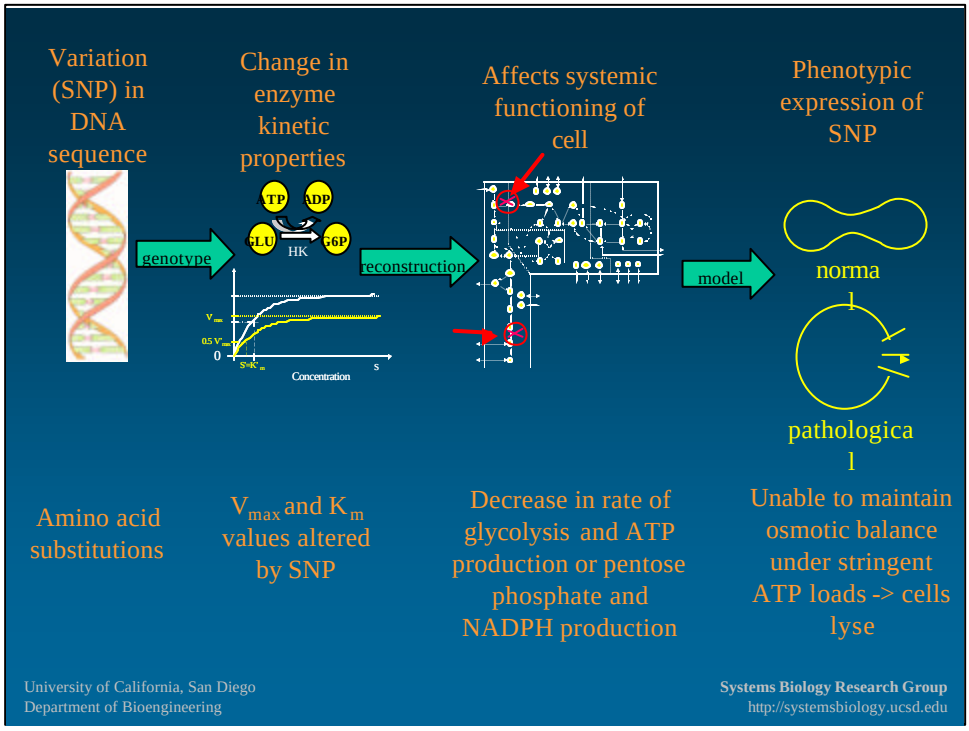


SNPs and the effect on load capacity

This theory can also be illustrated using the 2D cofactor projection from the earlier red cell pathway work. In this 2-D projection an enzymopathy has shortened GP1 to GP1' (decreased the minimum Vmax). The cell's ability to respond to energy loads is decreased as compared with the normal cell. While there may be no change under homeostatic conditions (the phys steady state is still in the feasible region) there will be problems as the cell is placed under an energy load – as is exactly the case from the results from the SNP study.

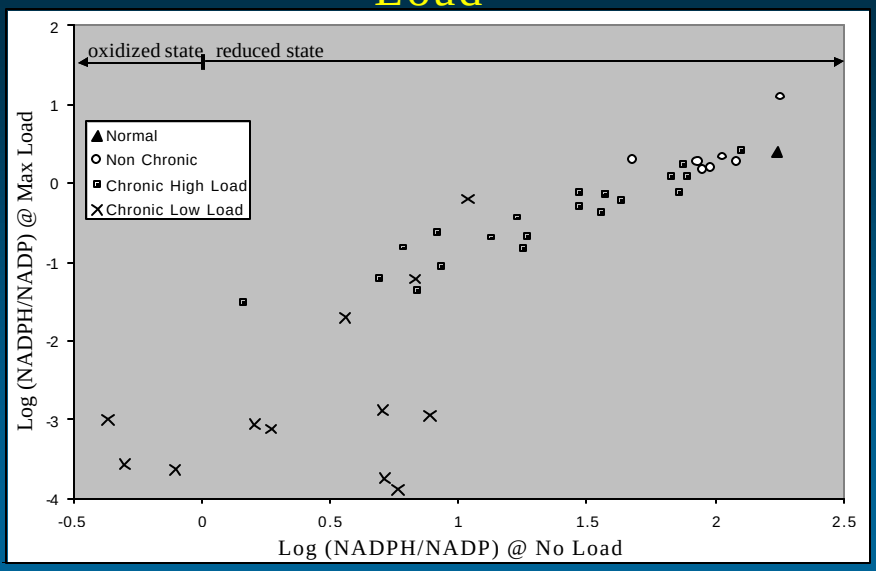


Here is an example of the data for the G6PDH variants in which the Vmax and inhibition constant for NADPH are highlighted. Their location along the protein is also highlighted as it often corresponds to a key active site in the protein. As you can see, there is a wide variation of the constants with no clear pattern emerging.



These altered parameters can then be put into the model and simulated to see their systemic effect. In some cases (the non-chronic cases), the SNPs did not affect the homostatic state and only presented a problem when the cell was put under a load – this is often the case in the clinical setting as well where an enzyme defect in the red cells is not diagnosed until the patient is put under some sort of stress such as a medicine which gives off oxygen free radicals. The SNPs could then be tested under loads to see the phenotypic consequences.

G6PDH Variant Response to Oxidative Load



University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Here are some results from the G6PDH case in which there appears to be a distinct difference between the chronic and non-chronic cases of SNPs. NADPH levels were used to gauge the state of the cell as this is a defect in the oxidative branch pentose phosphate reactions. Under normal conditions (i.e. oxidative load, $v_{ox} = 0$) there are differences between the chronic and non-chronic groups with the chronic group having a somewhat lower homeostatic steady state NADPH/NADP ratio than the non-chronic group. When subjected to an oxidative load ($v_{ox} > 0$), noticeable differences between the two groups (chronic and non-chronic) appear. The NADPH/NADP ratio at the maximum tolerated oxidative load (v_{ox} = max value) correlates with this ratio in the un-stressed situation ($v_{ox} = 0$). The group of chronic hemolytic anemia patients are clearly separated from the normal and non-chronic group. A number of the chronic cases can only withstand a very modest oxidative load.

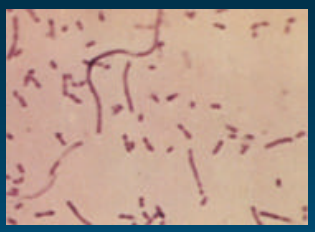
Extreme pathway analysis of *H. influenzae*

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

H. influenzae in a bronchiole

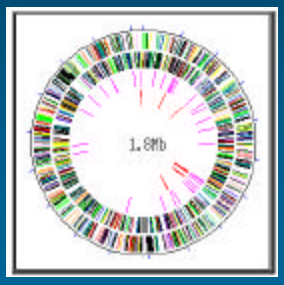
Pathology
- Gram-negative pathogen
- Colonizes the upper respiratory mucosa
- Most common cause of bacterial meningitis in children < 5 yrs



http://www.bact.wisc.edu/Bact330/lectureHflu

Statistics
- 380,000-600,000 deaths/yr worldwide

Genome Characteristics
- First fully-sequenced genome
- 1.83 Mbp genome length
- 1,703 estimated genes



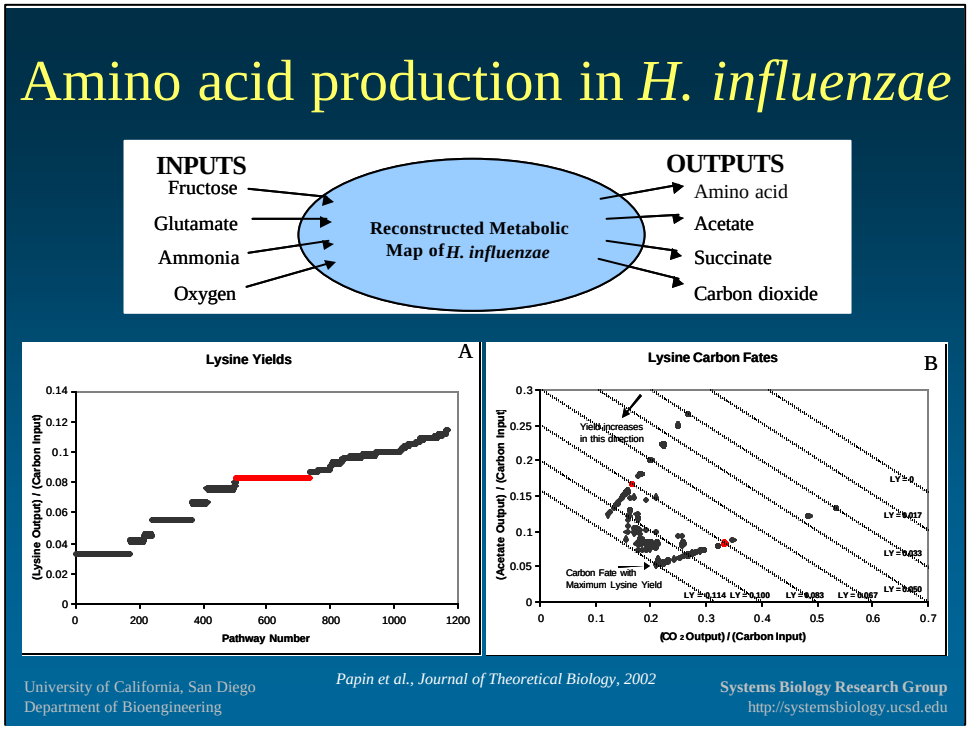
http://www.tigr.org/

Model Characteristics
- 461 Reactions (412 from the genome)
- 367 Metabolites

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

H. influenzae is a Gram-negative respiratory pathogen which causes around half a million deaths each year, many from bacterial meningitis. It was the first genome sequenced and as mentioned is somewhat smaller than *E. coli*. Its current metabolic model includes 461 reactions with 367 metabolites.



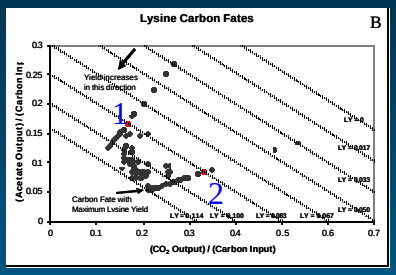
For the H. influenzae metabolic network, the inputs were constrained to minimal medium conditions. This is the minimal set of substrates in which the organism can survive (which, as a side note, is a difficult feature to characterize, but has been experimentally verified in some organisms, and which is an important feature and advantage of extreme pathway analysis). As we can see here, from the minimal medium, in order to synthesize amino acids, H. influenzae uses fructose, glutamate, ammonia, and oxygen. As outputs, the metabolic network was only allowed to produce the amino acid, acetate, succinate, and carbon dioxide. These constrained exchanges with the environment allowed for precise characterizations of the extreme pathway structure.

Here we see two types of data that which are of interest in this type of analysis. First, yield is defined as the amount of product per amount of substrate uptake. To normalize for varied carbon sources (glutamate and fructose), the lysine output was normalized to the total number of carbon inputted into the system. We can see the distinct yield values that are achievable by the metabolic network. Of note are these regions in which there are multiple pathways with identical yield values. It is important to remember that each of these pathways (or points) is a systemically independent vector.

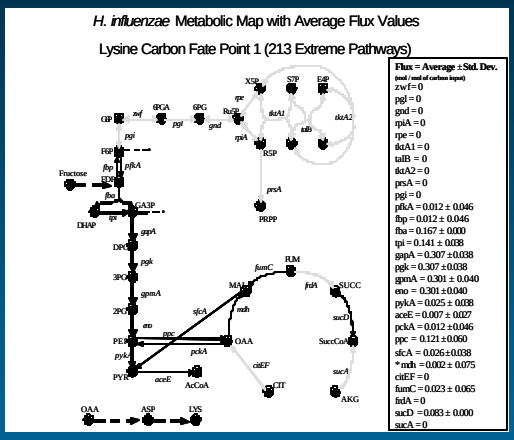
In order to further characterize these extreme pathways, a plot was generated to look at how carbon flowed through the system. Here, each point represents a ratio of acetate output to carbon dioxide output (both normalized to the carbon input). The only other sink for the carbon is in lysine. Hence, as we move closer to the origin, the represented pathway has an increased yield. OK was that dense enough for you?

One interesting characteristic of this carbon fate plot involves the constraining lines seen here. Another important point that can be seen in the above figure is that we can have points where the yield is equal but which lie on opposite sides of the space. The largest yield group in Figure A corresponds to the two red points in Figure B. These two points represent 231 pathways.

Carbon Fate Point 1



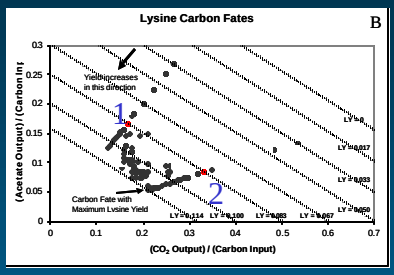
- Points 1 and 2 have the same lysine yield and yet each lies on a distinct extreme line that surrounds the possible values for the carbon fates in the system.



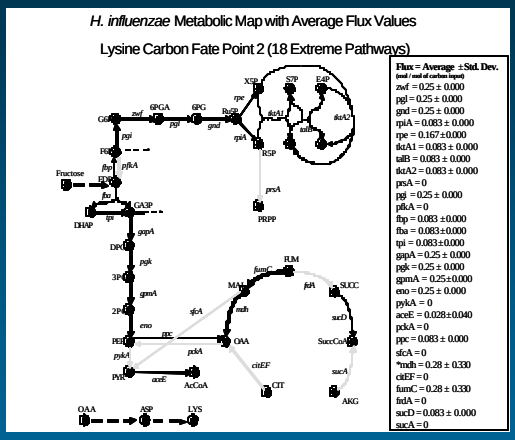
Papin et al., Journal of Theoretical Biology, 2002

At point 1 (from the previous slide) there are 213 pathways. Here we see an average flux map for the reactions in central metabolism. Note that the pentose phosphate reactions are not used.

Carbon Fate Point 2



- Note that between points 1 and 2, there are big differences in central metabolism (let alone other components of the network).
- Pentose phosphate reactions are used in one and not the other.
- Some reactions associated with the TCA cycle are used in one and not the other.

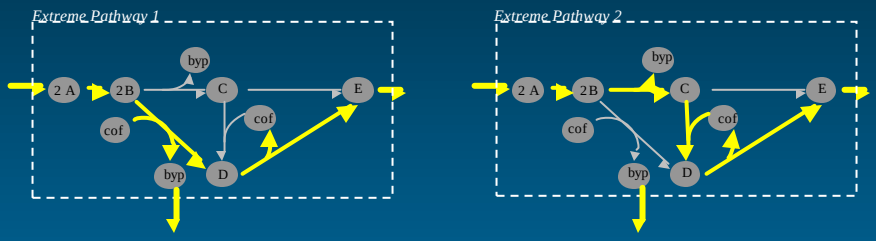


Papin et al., Journal of Theoretical Biology, 2002

Here we see the average flux map for reactions in central metabolism corresponding to the other carbon fate point. The width of the arrows represents the average flux value. Note the difference between the two maps. There is a totally different distribution between the two pathways that have equivalent yields.

Pathway Redundancy

- The diversity in metabolic maps at nearly equivalent carbon fate points and yield values motivated a quantitative evaluation of "pathway redundancy."



- For example, these extreme pathways have the same "external state"
  - input: 2 A
  - output: 1 E and 1 byp
- However, the internal flux distribution is very different in the two pathways

➔ Pathway Redundancy = 2

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

These patterns and the diversity of the underlying internal flux maps, motivated a quantitative evaluation of this property. Here we introduce the idea of pathway redundancy. Here is a sample system and these are two of the three extreme pathways which characterize it. The yellow lines represent fluxes that are used in the pathway. The light gray lines represent fluxes which are inactive in the corresponding extreme pathway. Note that both of the extreme pathways have equivalent "external states"… they both input 2 moles of metabolite A and output 1 mole of metabolite E and 1 mole of metabolite byp. However, both of the pathways have systemically independent ways of achieving this objective.

Pathway redundancy can then be calculated as the number of "internal states" per unique "external state." In this sample network, the pathway redundancy would be 2.

Summary of redundancy values in *H. influenzae*

- The highlighted redundancy values to the right demonstrate the wide range of values seen in *H. influenzae* under tested conditions.
- Case 1 only allowed acetate as a carbon byproduct.
- Case 2 allowed acetate and succinate as carbon byproducts.

Case 1

Amino Acid	PN	EYG	YR	MaxY	MinY	YLG	PWLG	PWLG / EYG (%)	UExV	PR
Alanine	1739	35	50	0.333	0.111	0.167	370	21%	62	28
Asparigine	445	14	32	0.167	0.083	0.167	104	23%	31	14
Aspartic Acid	690	8	86	0.167	0.083	0.167	366	53%	18	38
Glutamine	690	35	20	0.168	0.091	0.125	200	29%	66	10
Glycine	456	1	456	0.167	0.167	0.167	456	100%	3	152
Histidine	1507	48	31	0.077	0.030	0.067	248	16%	115	13
Isoleucine	1480	97	15	0.106	0.028	0.067	306	21%	135	11
Leucine	3884	27	144	0.111	0.048	0.111	1198	31%	55	71
Lysine	1168	78	15	0.114	0.033	0.083	231	20%	120	10
Methionine	1343	27	50	0.083	0.030	0.074	306	23%	55	24
Phenylalanine	1758	10	176	0.050	0.028	0.042	486	28%	28	63
Proline	2624	95	28	0.138	0.050	0.091	648	25%	154	17
Serine	690	8	86	0.167	0.083	0.167	366	53%	19	36
Threonine	1318	45	29	0.151	0.042	0.111	450	34%	74	18
Tryptophan	3540	7	506	0.033	0.017	0.030	972	27%	15	236
						0.028	972	27%		
Tyrosine	1758	10	176	0.050	0.028	0.042	486	28%	28	63
Valine	1739	35	50	0.167	0.056	0.083	370	21%	59	29
Average:	**1578**	**34**	**115**	**0.132**	**0.059**	**0.099**	**474**	**32%**	**61**	**49**
Average(without Leu,His,Trp):	**1278**	**36**	**91**	**0.145**	**0.065**	**0.111**	**368**	**34%**		**37**

Case 2

Amino Acid	PN	EYG	YR	MaxY	MinY	YLG	PWLG	PWLG / EYG (%)	UExV	PR
Alanine	8818	101	87	0.333	0.050	0.091	1662	19%	217	41
Asparagine	4419	21	210	0.167	0.040	0.046	2142	48%	56	79
Aspartic Acid	1470	11	134	0.167	0.071	0.091	684	47%	24	61
Glutamine	4750	105	45	0.168	0.053	0.063	2406	51%	233	20
Glycine	2436	7	348	0.167	0.031	0.091	1296	53%	18	135
Isoleucine	6133	238	26	0.106	0.019	0.029	576	9%	396	15
Lysine	5257	170	31	0.114	0.024	0.035	456	9%	315	17
Methionine	1748	34	51	0.083	0.030	0.074	330	19%	76	23
Phenylalanine	5034	20	252	0.050	0.017	0.035	1020	20%	55	92
Proline	4356	149	29	0.138	0.031	0.059	830	19%	259	17
Serine	1470	11	134	0.167	0.071	0.091	684	47%	25	59
Threonine	3683	80	46	0.151	0.042	0.091	698	19%	160	23
Tyrosine	5034	20	252	0.050	0.017	0.035	1020	20%	55	92
Valine	10356	110	94	0.167	0.024	0.059	1854	18%	206	50
Average(without Leu,His,Trp):	**4640**	**77**	**124**	**0.145**	**0.037**	**0.063**	**1118**	**28%**		**52**

University of California, San Diego
Department of Bioengineering

Papin et al., Journal of Theoretical Biology, 2002

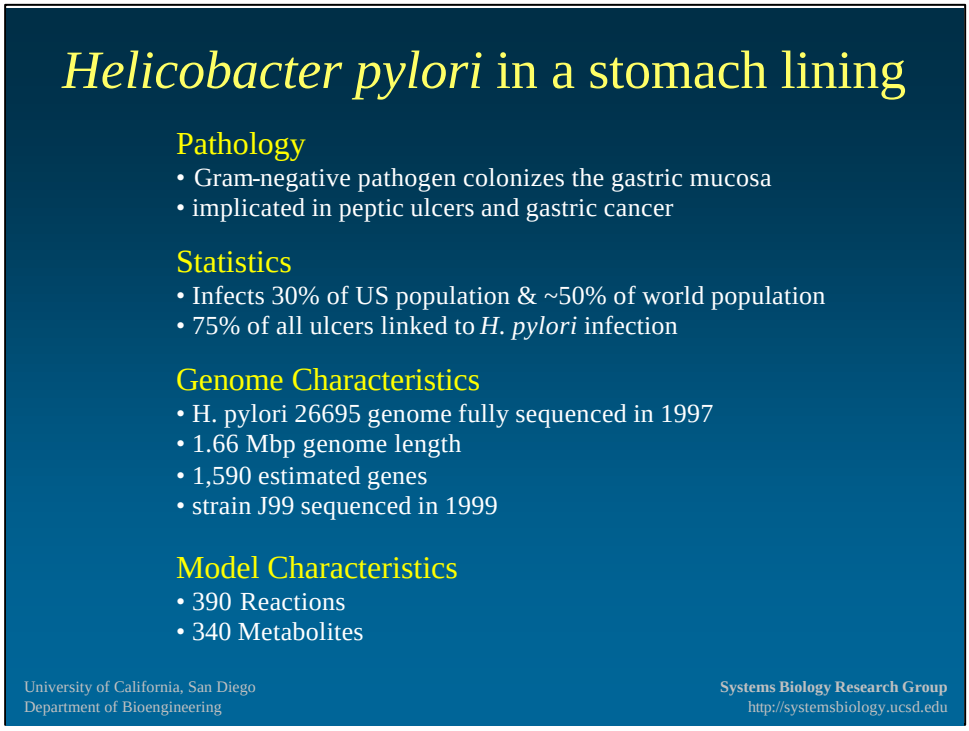
Systems Biology Research Group
http://systemsbiology.ucsd.edu

The pathway redundancy for amino acid production in H. influenzae under the conditions described, were calculated. As we can see, there is approximately an average of 50 internal states per unique external state. In other words, there are 50 systemically independent ways to achieve this same objective for amino acid synthesis in H. influenzae.

Extreme pathway analysis of *H. pylori*

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu



H. pylori is primarily known for its link to 75% of peptic ulcers and potentially to cases of gastric cancer. It was thought that the stomach was aseptic because of its acidity, but it is now known that *H. pylori* is able to live in the stomach lining where it avoids the immune response and where the environment is less acidic. Furthermore, the metabolism of the organism is such that it secretes ammonia, releasing a basic cloud into its local environment which protects it. The interesting features of this organism's metabolism, combined with the release of its annotated genome sequence in 1997, motivated us to construct a metabolic model.

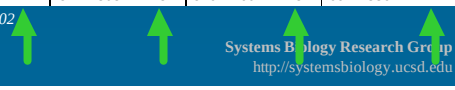
The metabolic model for H. pylori is slightly smaller than that of H. influenzae, with 390 reactions involving 340 metabolites. It's genome is also of comparable size to that of H. influenzae.

Summary of redundancy in *H. pylori*

- A similar redundancy analysis was done with the metabolic model of *H. pylori.*
- One interesting difference was in the lack of variation among the amino acids and under the various environmental conditions.

	Case 1			Case 2			Case 3			Case 4		
AMINO ACID	EP	ES	EP / ES	EP	ES	EP / ES	EP	ES	EP / ES	EP	ES	EP / ES
Asparigine	217	105	2.1	295	154	1.9	295	154	1.9	340	193	1.8
Aspartic Acid	360	102	3.5	466	142	3.3	466	142	3.3	491	163	3.0
Cysteine	473	232	2.0	822	420	2.0	822	420	2.0	1022	612	1.7
Glutamine	249	101	2.5	290	140	2.1	290	140	2.1	315	164	1.9
Glutamic Acid	441	148	3.0	473	178	2.7	473	178	2.7	493	198	2.5
Glycine	0			0			348	149	2.3	377	173	2.2
Lysine	474	167	2.8	587	213	2.8	587	213	2.8	611	237	2.6
Proline	479	149	3.2	621	209	3.0	621	209	3.0	867	326	2.7
Serine	212	101	2.1	326	158	2.1	326	158	2.1	355	186	1.9
Threonine	275	136	2.0	432	206	2.1	432	206	2.1	469	242	1.9
Tryptophan	936	615	1.5	1431	1034	1.4	1431	1034	1.4	1958	1486	1.3
Tyrosine	584	302	1.9	825	477	1.7	825	477	1.7	1008	649	1.6
AVERAGE	**392**	**196**	**2.4**	**547**	**303**	**2.3**	**576**	**290**	**2.3**	**692**	**386**	**2.1**

Price et al., Genome Research, 2002



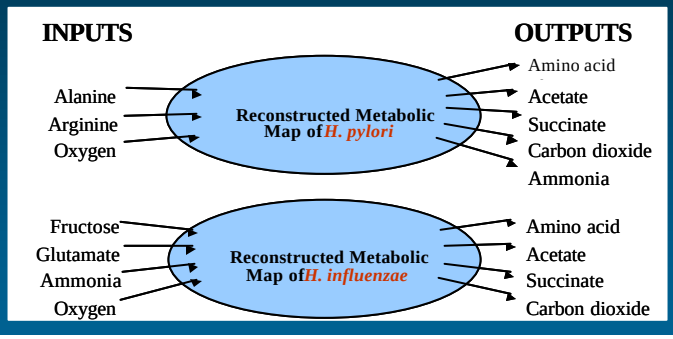
University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Similar to the analysis of H. influenzae, the pathway redundancies for amino acid synthesis in H. pylori was calculated. As you can already see, there is an order of magnitude difference between the degree of redundancy in the two organisms. Another interesting difference is in the variation of the values. In H. influenzae, we saw large changes for the different amino acids and under different environmental conditions. That same fluctuation is not observed here.

Comparative pathway redundancy

Even with similarly sized genomes, *H. influenzae* and *H. pylori* had an order of magnitude difference in the degree of pathway redundancy for amino acid production. This serves as an important example of "emergent properties" that can only be seen with such genome-scale analyses.



In silico organism	Genome Size	ORFs in Genome	Reactions in Model	# of PW / UExV
H. influenzae	1.83 Mb	1740	461	46
H. pylori	1.67 Mb	1590	390	2

University of California, San Diego
Department of Bioengineering

Price et al., Genome Research, 2002

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Since the minimal medium conditions for H. pylori are different than those of H. influenzae (they have different minimal medium requirements), direct comparisons between the organisms are difficult to make. However, accounting for these differences, shows the degree of variability. The average number of pathways per unique external state (for the same set of amino acids and with the same external carbon sinks), there were 46 internal states per external state in H. influenzae, and 2 internal states per external state in H. pylori.

This suggests and gives a quantitative measure for the amount of robustness in the organism. It suggests that H. pylori perhaps fits into a much more defined environment and very specifically fits into its environmental niche.

Pathway redundancy in *H. pylori*

- In *H. pylori*, it was possible to calculate the extreme pathways for the set of amino acids or for the set of ribonucleotides.

Product	# of PW	# of UExV	# of PW / UExV
Equimolar Amino Acids	6032	2825	2.1
E. coli Ratio Amino Acids	5553	2481	2.2
Equimolar Ribonucleotides	1325	1164	1.1

Price et al., Genome Research, 2002

- The redundancy values for the simultaneous production of the amino acids was of the same order of magnitude as that of the individual amino acids.
- The redundancy values for ribonucleotide synthesis was approximately half of that seen in amino acid synthesis.

University of California, San Diego
Department of Bioengineering

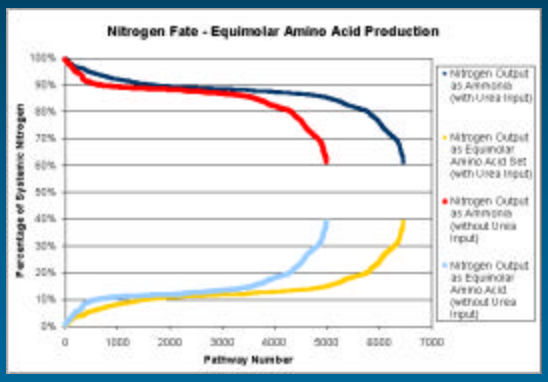
Systems Biology Research Group
http://systemsbiology.ucsd.edu

In H. pylori, it was also possible to look at a few small linked outputs. Interestingly, the redundancy values for the set of amino acids in 1:1 proportions was still about 2. The ratio of the composition of amino acids in the set was also inspected. The relative amount of each of amino acid was approximated to that of E. coli (since such data is unavailable in H. pylori), and still the redundancy was about 2.

Interestingly, the redundancy for the production of ribonucleotides was nearly half…a little more than one, which implies that there are more unique routes for nucleotide synthesis.

Nitrogen production built into structure of *H. pylori* metabolic network

- Another interesting feature from the ExPA of *H. pylori* amino acid production was in the flow of nitrogen in the system. Important due to the habitat of the microorganism.
- Over 80% of the extreme pathways directed more than 80% of the input nitrogen into the production of ammonia…potentially important for pH regulation in its acidic habitat.



Price et al., Genome Research, 2002

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Since an interesting feature of H. pylori is its ability to survive in an intensely acidic environment, an analysis of the flow of nitrogen in the system is also of importance. Here we are looking at a representation of the more than 6000 extreme pathways associated with the synthesis of the set of amino acids. Here, the set of amino acids and ammonia are the only nitrogen sinks in the network. We also allowed urea as an input in the system (which has been hypothesized as a critical metabolite…urease is a mass-produced enzyme in H. pylori. The breakdown of urea results in the production of ammonia). In all of the cases, more than 80% of the extreme pathways direct more than 80% of the nitrogen inputted to the synthesis of ammonia…it cannot use the nitrogen to synthesize amino acids. In addition, in this case, no more than 40% of the inputted nitrogen can be incorporated into amino acid synthesis. Since amino acids represent a significant demand of systemic nitrogen, it has been hypothesized that the flow of nitrogen to the synthesis of ammonia is “built in” for the H. pylori metabolic network.

Methods for analyzing large sets of Extreme Pathways

University of California, San Diego
Department of Bioengineering

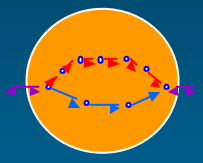
Systems Biology Research Group
http://systemsbiology.ucsd.edu

The results up until now have illustrated the need for being able to pick out important information from the vast amounts of data that extreme pathway analysis generates. To date, some work has been done on developing methods for doing this very thing.

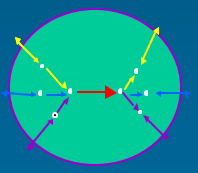
Methods for parsing out salient information

$$\mathbf{P} = \begin{pmatrix} \mathbf{v}_{11} & \mathbf{v}_{12} & \cdots & \mathbf{v}_{1n} \\ \mathbf{v}_{21} & \mathbf{v}_{22} & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \mathbf{v}_{m1} & \cdot & \cdots & \mathbf{v}_{mn} \end{pmatrix} \begin{matrix} P_1 & P_2 & \cdots & P_n \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \end{matrix} \begin{matrix} \\ v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_m \end{matrix}$$

PT·P → Pathway Length Matrix
the number of reactions in an extreme pathway



P·PT → Reaction Participation Matrix
the number of pathways that go through a particular reactions



University of California, San Diego
Department of Bioengineering

Papin et al., in review, 2002

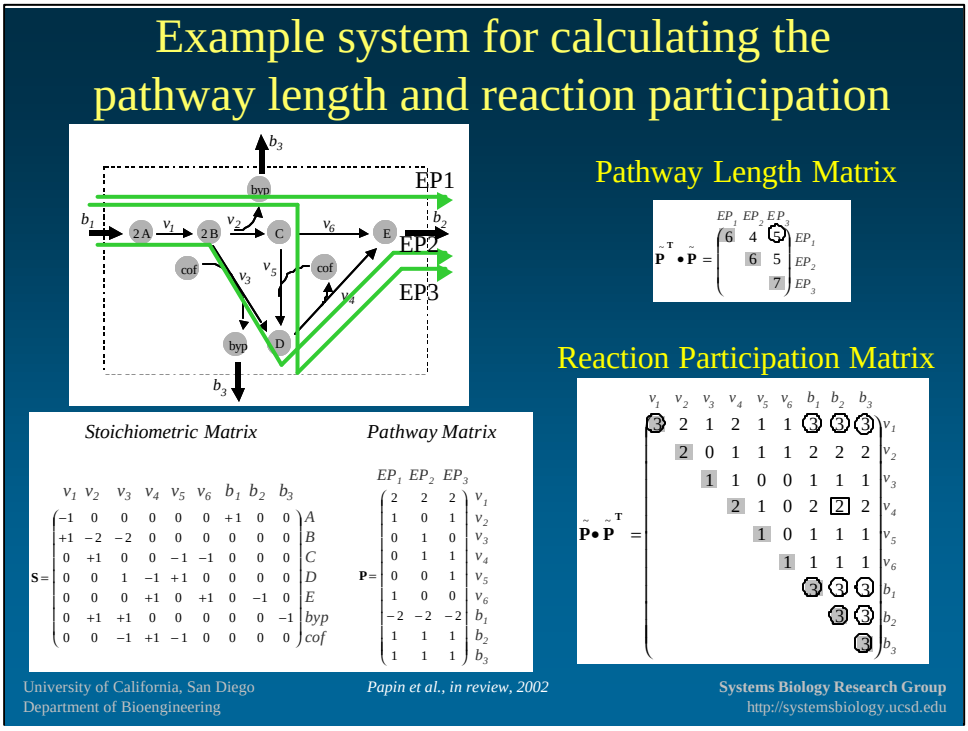
Systems Biology Research Group
http://systemsbiology.ucsd.edu

As an example, one such method involves first rewriting the pathway matrix into a binary form (i.e. 1 if it is used and 0 if it is not used). Then, this matrix is pre- and post-multiplied by its transpose. The two resultant matrices we will call the reaction participation matrix and the pathway length matrix.

The values in the diagonals of the pathway length matrix are the lengths of the respective extreme pathways, and are the number of reactions which participate in the matrix. The values in the off-diagonals of the matrix are the numbers of reactions in which a given pair of pathways participates. It is important to remember that extreme pathways are not simply linear chains of reactions (contrary to the schematic shown above). Rather, extreme pathways can have multiple inputs and multiple outputs. Consequently, the "pathway lengths" in the Pathway Length matrix, are perhaps more precisely characterized as the "size" of the extreme pathway. These extreme pathway sizes have been analyzed and interesting characterizations have been made.

In the reaction participation matrix, the diagonal values correspond to the number of pathways in which a reaction participates. The off-diagonal values correspond to the number of pathways in which a given pair of reactions appear together.

As the first schematic tries to demonstrate, one pathway can be "longer" than another. As the other schematic tries to demonstrate, one reaction may participate in more pathways than another.



Just to review briefly, and perhaps to clarify a little bit, the generation of the pathway length matrix and the reaction participation matrix.

First, here is a reaction network with 7 metabolites and 6 internal reactions and 3 exchange fluxes. And here is the extreme pathway matrix, P, with 3 extreme pathways.

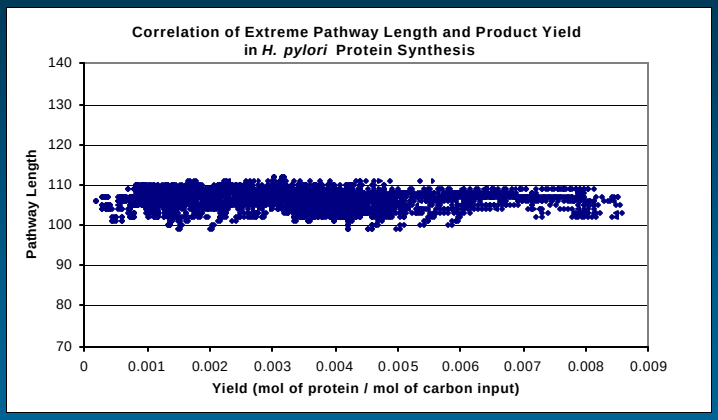
Now, if we pre- and post-multiply the binary pathway matrix by its transpose, we generate the pathway length and reaction participation matrices. Note that they are symmetric, so only one half of the values are shown.

Note that there are 3 extreme pathways. EP1 and EP2 have 6 participating reactions. EP3 has 7 participating reactions. Also note that extreme pathways 1 and 3 have 5 reactions in common.

In the reaction participation matrix, note that if we look at the reaction corresponding to flux v1, there are three extreme pathways in which it participates. Similar characterizations can be seen with the other reactions. As an example of an off-diagonal value, look at the value corresponding to fluxes v4 and b2. There are 2 extreme pathways in which both of the corresponding reactions participates. In the pathway matrix we can see that v4 and b2 appear together in 2 of the 3 extreme pathways.

Although this demonstration may seem straightforward, the simplicity of its interpretation for larger networks is powerful and very applicable. Let's now look at a few applications to the data sets from H. influenzae and H. pylori that we have already been looking at.

Pathway Length – Product Yield Correlation



Papin et al., in review, 2002

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

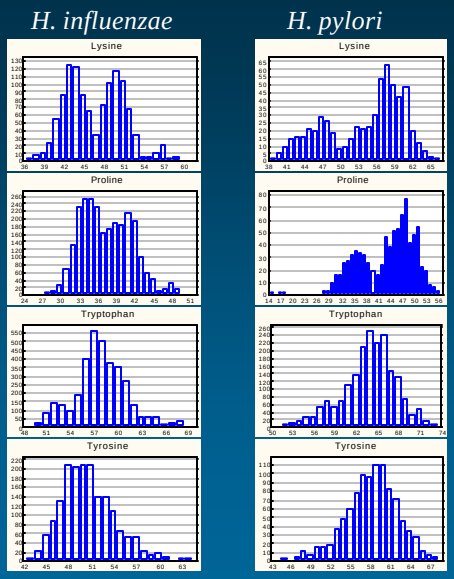
Here we looked at the extreme pathways for the synthesis of a linked output… the synthesis of the set of amino acids in H. pylori. The pathway length and reaction participation matrices were generated for this data set (note that there are over 6000 extreme pathways in this data set). The pathway length values ranged from just under 100 to just over 110. These values were correlated with the amino acid yield of the respective pathways.

Surprisingly, there was a very poor correlation between these two variables. This is another important example of an emergent property that can be observed from a genome-scale analysis like ExPA. This has very important implications. A pathway of optimal yield, perhaps very important for metabolic engineering purposes, can not be identified by a simple visual inspection. Perhaps one might think that a pathway with a smaller size might correspond to one in which less carbon or nitrogen is lost to a byproduct and hence produces a given output optimally.

The correlations between pathway length and other variables of interest (for example, the number of carbons in the target product) were also evaluated. Although some measures had slightly better correlations, they were all fairly weak correlations.

Pathway Length Distributions

- Interestingly, pathway length distributions exhibited particular statistical trends.
- There were distinct trends in the skewness, number of modes in the distributions, and other such properties.



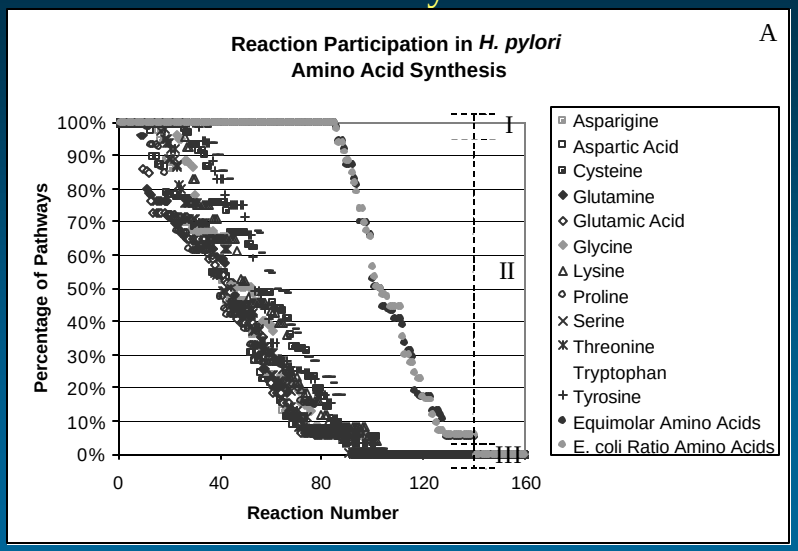
University of California, San Diego
Department of Bioengineering

Papin et al., in review, 2002

Systems Biology Research Group
http://systemsbiology.ucsd.edu

The pathway length matrices were generated for all of the data sets that we have previously described (amino acid synthesis in H. influenzae; amino acid and ribonucleotide synthesis in H. pylori). Here we see the distributions of pathway lengths for the extreme pathways of a few representative amino acids. You can see that distinct differences in the statistical measures between organisms can be seen. As we see here, there is a general skewness (to the right in the pathway lengths in H. pylori and to the left in the pathway lengths in H. influenzae). We also see some bimodal distributions. These types of statistical measures provide for important investigations to see why such features exist.

Reaction participation in *H. pylori* amino acid synthesis

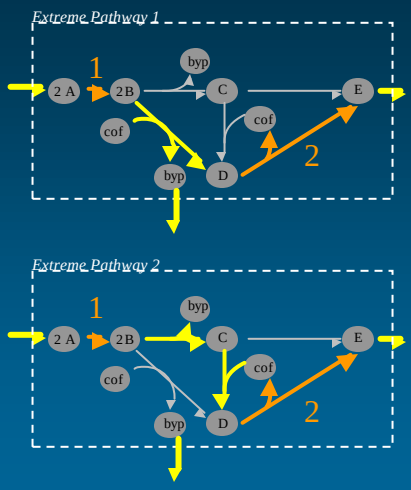


University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Non-obvious, systemically correlated reactions

- Reactions that always appear together for the synthesis of a particular product (for example, reactions 1 → and 2 → that produce E from substrate A in the sample system to the right).
- These reactions may indicate regulatory structure.
- At the least, the correlated groups provide for interesting hypotheses for further inspection.



University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

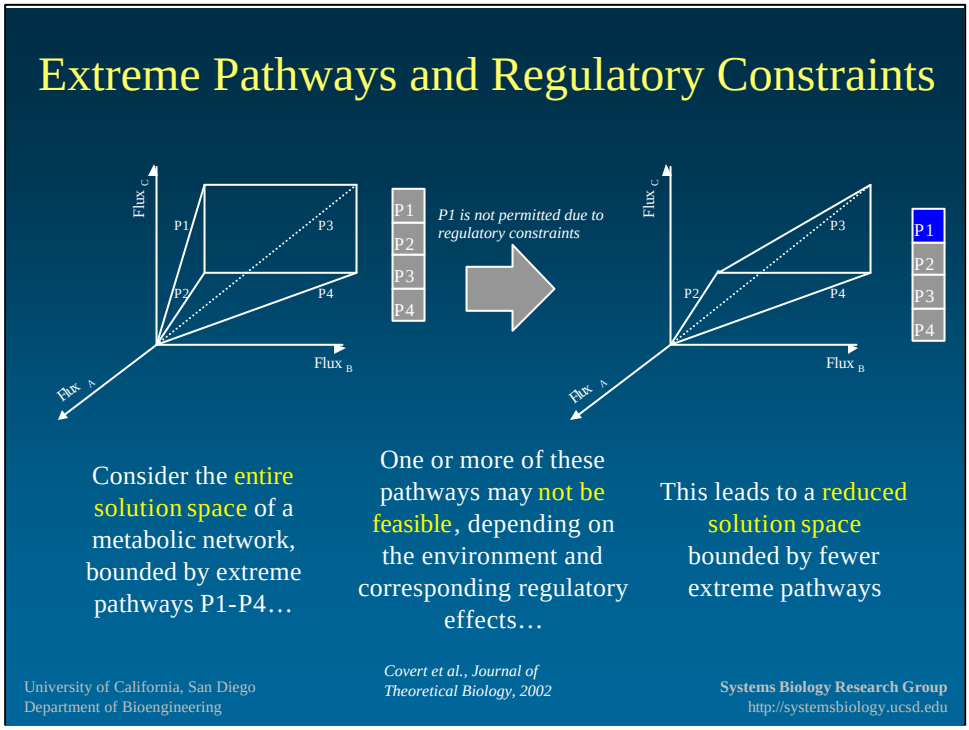
Now to move on to some of the characterizations that were made with the reaction participation matrix. As we saw earlier, the reaction participation matrix indicates how many extreme pathways a given reaction (or pair of reactions) participates in. This information can be readily used to determine which reactions always appear together. For example, in the sample system to the right, reactions 1 and 2 are active in both of the extreme pathways shown while the other internal reactions change. These reactions can be thought of as systemically correlated. While previous work (Shilling and Palsson, 2000) has used such an approach to look at correlated reactions in the extreme pathways for subsystems, this genome-scale analysis takes into account reactions that might otherwise be separated into subsystems.

Such systemically correlated reactions could be an indication for regulatory structure. At the least, correlated groups provide for interesting hypotheses of network objectives and should be further inspected.

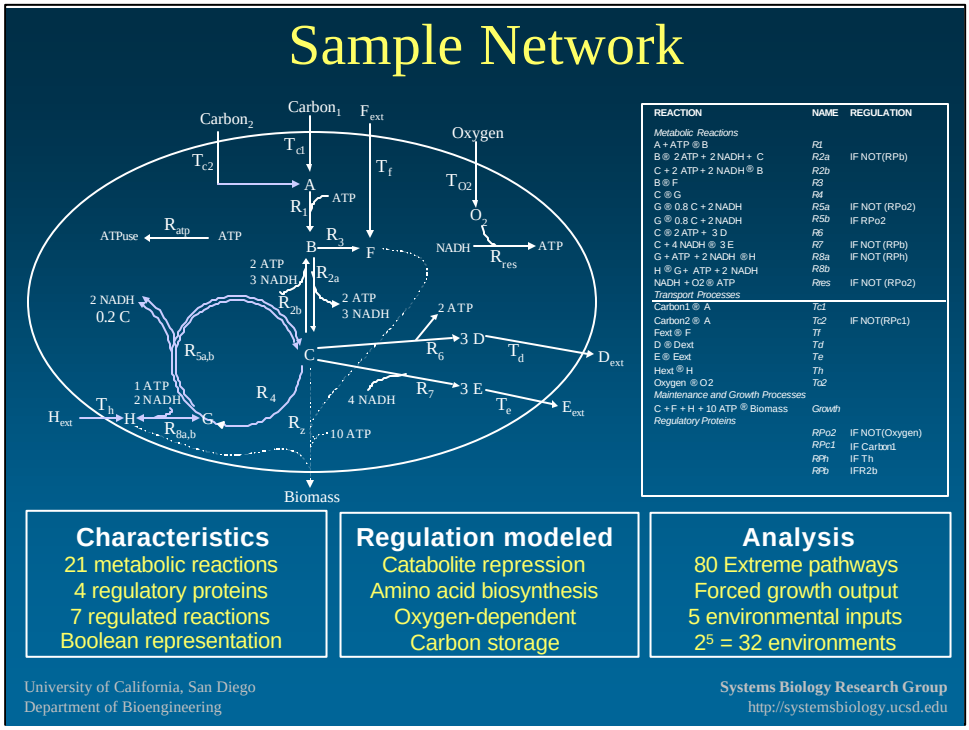
Interpretation of transcriptional regulation with extreme pathway analysis

University of California, San Diego
Department of Bioengineering

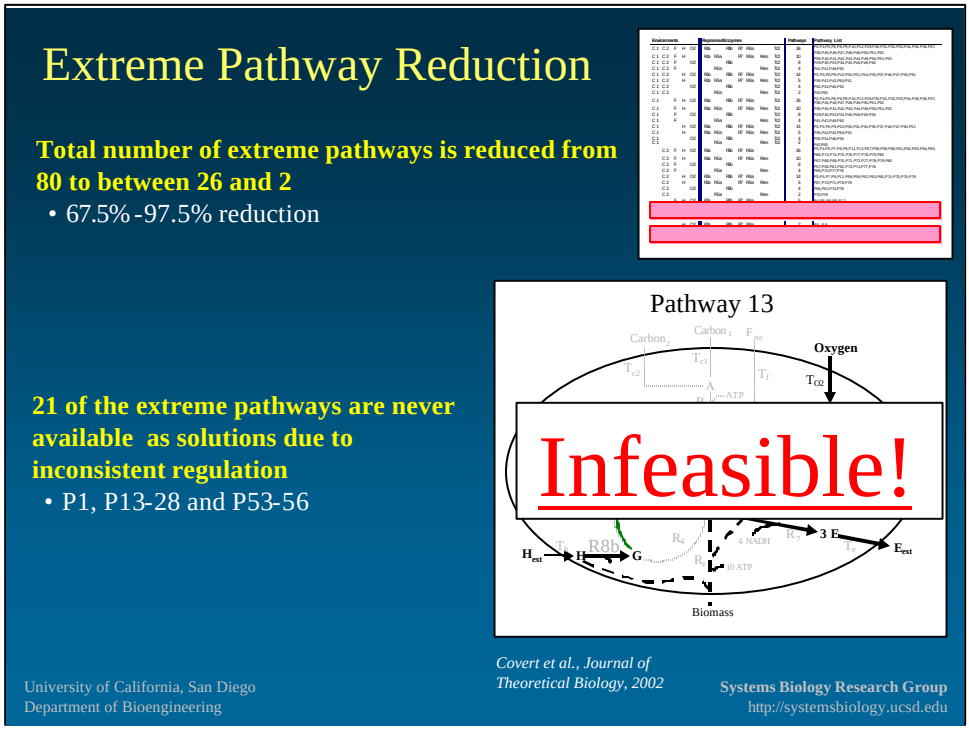
Systems Biology Research Group
http://systemsbiology.ucsd.edu



This slide shows how regulatory constraints reduces the number of active extreme pathways in a system. Let's assume that transcriptional regulation is modeled as a Boolean network (see Thomas 1978 in the References for more detail), and that pathways are considered "ON" or "OFF" depending on these rules. This basically means that a certain pathway may or may not be feasible under given conditions. So if we consider the entire solution space of a metabolic network, bounded by extreme pathways P1-P4, one or more of these pathways may not be feasible, depending on the environment and corresponding regulatory effects and is therefore eliminated from the boundaries of the space. In the case shown here, P1 is infeasible and therefore the solution space is reduced and bounded only by P2, P3 and P4.

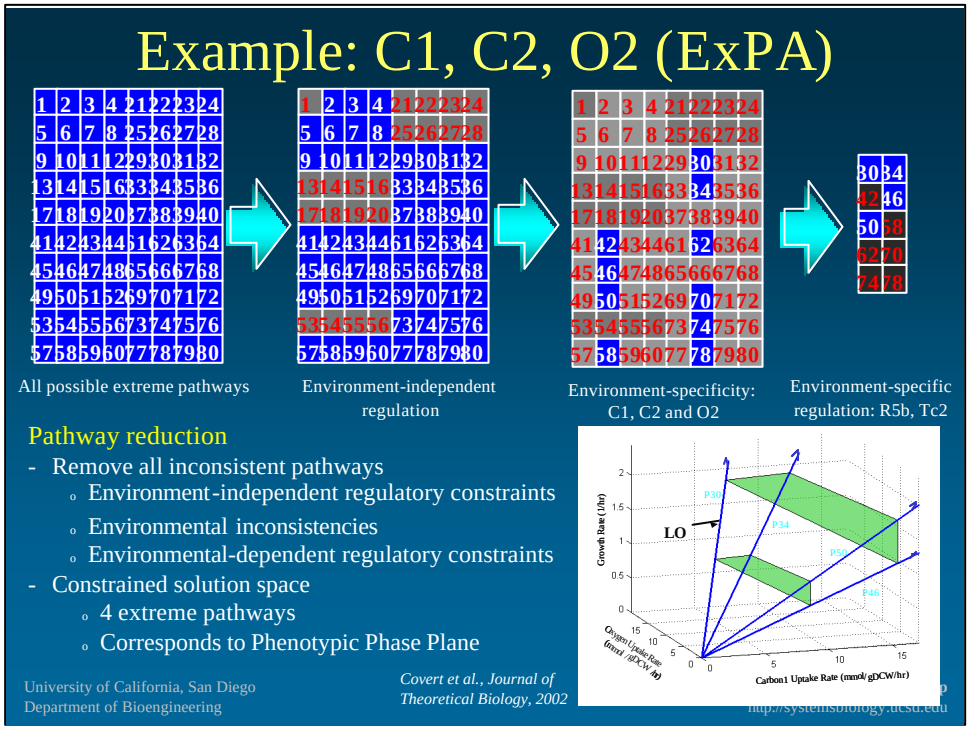


Here is the sample network we will use to illustrate solution space reduction due to transcriptional regulation. It is supposed to represent central metabolism in a “typical” cell, together with some of the corresponding regulation. Some characteristics, as well as different types of regulation which can be modeled with this system, are shown. Overall, with a forced growth output we obtain 80 extreme pathways which characterize this system. Given the 5 environmental inputs (Carbon1, Carbon2, F, H and Oxygen) and considering each as either “present” or “absent” in the extracellular medium, we have 32 possible environments which may affect this system.



The first thing we notice is that none of the conditions have all 80 pathways available to them. In fact, the largest number of pathways available to the system under any condition is 26 and the smallest is 2 (not counting those environments incapable of sustaining growth). The incorporation of regulatory constraints can therefore greatly reduce the number of feasible extreme pathways. If you look carefully at the list you will also notice that certain environments have identical pathway lists or lists which are subsets of other lists.

Another interesting observation which can be made from these results is that 21 of the extreme pathways are never available as feasible solutions. This is due to inconsistent regulation in the extreme pathway's flux distribution. Pathway 13, for example, has a flux distribution as shown, where the active fluxes are red if active, or green to indicate an anaerobic isozyme (aerobic isozymes are generally shown in purple in these diagrams as you'll see later). Now in this case, the oxidative reactions are active, which are only expressed under aerobic conditions, but also isozymes which make up this simplified TCA cycle are also expressed. These isozymes are only active under anaerobic conditions! Therefore, this flux distribution is always infeasible due to regulatory constraints.

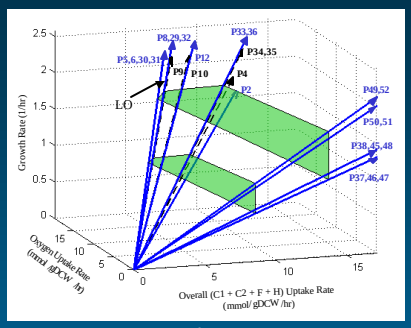


Let's examine one of these environments in greater detail – the Carbon1, Carbon2, Oxygen environment. Starting with 80 pathways, we first remove the pathways which are infeasible due to inconsistent regulation – *environment-independent* regulation as discussed in the last slide. Next we remove any pathways which use H or F (only C1, C2 and Oxygen are available). Finally, the regulatory constraints for the environment indicate that R5b and Tc2 may not be used, so we remove any pathways which require those reactions. We are left with 4 pathways. Once these pathways have been identified, we may define the solution cone (shown here as a 3-dimensional projection on the Carbon1 Uptake Rate – Oxygen Uptake Rate – Growth Rate axis). P30 is the line of optimality. You will see later in the course and briefly on the next slide that this space will correspond to the phenotypic phase plane of this system.

Complex medium: Regulation of pathways

Environment-independent regulation

Environment-specific regulation: R2a, R5b, R7, R8a, Tc2



Covert et al., Journal of Theoretical Biology, 2002

- Number of extreme pathways is only reduced to 26
- More flexibility in the system

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Here is the most complex environment – all 5 of the inputs are present. In this case the reactions R2a, R5b, R7, R8a and Tc2 are repressed, reducing the number of extreme pathways to 6, as we mentioned earlier. Here you can also see a three and two-dimensional projection of the solution space, both of which show more flexibility in the system.

Summary

- Genome-scale metabolic networks are being reconstructed.
- Extreme pathway analysis of representative systems have illustrated important features.
- Extreme pathway analysis can be used to elucidate the genotype-phenotype relationship
- Extreme pathway analysis of genome-scale models have found new emergent properties.
- Incorporation of transcriptional regulation will provide for an even greater physiological characterization.

References

- Schilling and Palsson. Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective, *J. Theor. Biol.*, 2000; 203: 249-283.
- Schuster, Pfeiffer, Moldenhauer, Koch, and Dandekar. Bioinformatics, 2002 Feb;18(2):351-61.
- Papin, Price, Edwards, and Palsson. The Genome-Scale Metabolic Extreme Pathway Structure in Haemophilus influenzae Shows Significant Network Redundancy. *J. Theor. Biol.*, 2002; in press.
- Price, Papin, and Palsson. Determination of Redundancy and Systems Properties of Helicobacter pylori's Metabolic Network Using Genome-Scale Extreme Pathway Analysis. *Genome Res.*, 2002; 12:760-769.
- Papin, J.A., Price, N.D., and Palsson, B.O. "Extreme Pathway Lengths and Reaction Participation in Genome-Scale Metabolic Networks." *Genome Research*, **12**: pp. 1889-1900 (2002).
- Samatova et al. Parallel Out-of-core Algorithm for Genome-Scale Enumeration of Metabolic Systematic Pathways. Proceedings of the First IEEE Workshop on High Performance Computational Biology (HiCOMB2002). April 2002.
- Covert M.C. and Palsson B.O., "Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*," *JBC*, Papers in Press (2002).

References

- Beutler, E. 1986. Red cell metabolism. Churchill Livingstone, Edinburgh ; New York.
- Covert, M. W., C. H. Schilling, I. Famili, J. S. Edwards, I. I. Goryanin, E. Selkov, and B. O. Palsson. 2001a. Metabolic modeling of microbial strains *in silico*. *TIBS*. 26:179-186.
- Grimes, A. J. 1980. Human Red Cell Metabolism. Blackwell Scientific Publications, Oxford.
- Jamshidi, N., J. S. Edwards, T. Fahland, G. M. Church, and B. O. Palsson. 2001. Dynamic simulation of the human red blood cell metabolic network. *Bioinformatics*. 17:286-7.
- Joshi, A., and B. O. Palsson. 1989. Metabolic dynamics in the human red cell. Part I--A comprehensive kinetic model. *Journal of Theoretical Biology*. 141:515-28.
- Covert, M., Schilling, C. H., Palsson, B.O., "Regulation of Gene Expression in Flux Balance Models of Metabolism," *Journal of Theoretical Biology*, 213(1): pp. 73-88 (2001).