Project Title: Collaborative Research: Metabolic Engineering of *E. coli* Sugar-Utilization Regulatory Systems for the Consumption of Plant Biomass Sugars.

DOE award number: DEFG3603GO13159

Principal investigator: Ramon Gonzalez.
Co-Principal investigators: J. V. Shanks and K-Y. San.

Executive Summary

The overall objective of this project is to metabolically engineer the *E. coli* sugar-utilization regulatory systems (SURS) to utilize sugar mixtures obtained from plant biomass. Of particular relevance is the implementation of a metabolic engineering cycle aided by functional genomics and systems biology tools. Our findings will help in the establishment of a platform for the efficient production of fuels and chemicals from lignocellulosic sugars. Our research has improved the understanding of the role of SURS in regulating sugar utilization and several other cellular functions. For example, we discovered that Mlc, a global regulatory protein, regulates the utilization of xylose and demonstrated the existence of an important link between catabolite repression and respiratory/fermentative metabolism. The study of SURS mutants also revealed a connection between flagellar biosynthesis and catabolite repression. Several tools were also developed as part of this project. A novel tool (Elementary Network Decomposition, END) to help elucidate the network topology of regulatory systems was developed and its utility as a discovery tool was demonstrated by applying it to the SURS in *E. coli*. A novel method (and software) to estimate metabolic fluxes that uses labeling experiments and eliminates reliance on extracellular fluxes was also developed. Although not initially considered in the scope of this project, we have developed a novel and superior method for optimization of HPLC separation and applied it to the simultaneous quantification of different functionalities (sugars, organic acids, ethanol, etc.) present in our fermentation samples. Currently under development is a genetic network driven metabolic flux analysis framework to integrate transcriptional and flux data.

Summarize project activities for the entire period of funding, including original hypotheses, approaches used, problems encountered and departure from planned methodology, and an assessment of their impact on the project results. Include, if applicable, facts, figures, analyses, and assumptions used during the life of the project to support the conclusions.

The proposed project uses all of steps generally included in the metabolic engineering cycle: (1) construction of recombinant strains with improved properties, (2) analysis of the recombinant strains, especially their performance compared with the original strain background, and (3) design of the next target for genetic engineering. The overall objective is to metabolically engineer the *E. coli* sugar-utilization regulatory systems (SURS) to utilize sugar mixtures obtained from plant biomass. The work will also improve understanding of the role of SURS in regulating sugar utilization and several other cellular functions. We have six Specific Aims, of which research for Aims 1-5 are scheduled in Years 1 and 2. The main part of the work in year 1 was devoted to specific aims 1 and 2 (construction and initial evaluation of *E. coli* strains with modified SURS). During the second year, evaluation of selected strains using tools that allow large-scale study of cellular responses to these modifications (i.e. NMR-based metabolic flux analysis, MFA, and genome-wide transcriptional analysis) was started. Due to the cyclic nature of metabolic engineering, certain overlap on time is expected between the work proposed in each specific aims.

Two opportunities exist within this project to further study some related problems. The first is the development of an optimization framework for HPLC separation, the results of which will be of great practical value in the phenotypic characterization of *E. coli* mutants in the study: i.e. for the simultaneous quantification of fermentation substrates and products (sugars, alcohols, organic acids, etc.). Also, an HPLC optimization software with built-in retention models for a widely used column will serve as a real research tool for use by other researchers. The second opportunity is characterization of the hybridization process in DNA microarray using molecular dynamics simulation. Results of this study will offer insights into a crucial but uncharacterized process in the microarray experiment, and specifically, provide a deterministic measure of non-specific hybridization in microarrays, and *in silico* hybridization as a new paradigm in design and validation of oligonucleotides for microarrays. Results in these two areas are presented on this report

Specific Aims:
1. Construct *E. coli* strains with modified SURS by introducing genetic modifications to single and multiple components of the system.
2. Evaluate the capability of these mutants to efficiently utilize glucose, xylose, and arabinose.
3. Evaluate changes in the regulation of gene expression at genomic scale resulting from engineering *E. coli* SURS using DNA microarray technology.
4. Identify functional metabolic pathways and quantify their fluxes in SURS mutants and wild type strains by using a novel flux analysis technique based on bondomer analysis of 2D NMR bond-labeling experiments.

5. Integrate these results using a novel genetic network based MFA model that combines gene expression and metabolic flux data.

6. Propose further modifications for improving the properties of SURS mutants.

Timetable

Specific Aim (Laboratory in charge[a])	Year 1	Year 2	Year 3
Construct *E. coli* strains with modified SURS (I,)	→→→→→→→→→→→	→	
Evaluate the capability of SURS mutants for utilizing sugar mixtures (I, II)	→→→→→→→→→→→→→→→→→→→		→
Evaluate genome-wide changes in gene expression in SURS mutants (I)		→→→→→→→→→→→→	→
Metabolic characterization (NMR-based MFA) of SURS mutants (II)		→→→→→→→→→→→→	→
Integrate gene expression and metabolic flux data (III)		→→→→→→→	→
Propose further modification for 2nd generation of SURS mutants (I, II, III)			→→→→→→→→

a) I, Dr. Gonzalez's Laboratory (ISU); II, Dr. Shanks' Laboratory (ISU); III, Dr. San's Laboratory (Rice).

Specific Aim 1. Construct *E. coli* strains with modified SURS by introducing genetic modifications to single and multiple components of the system.

Deletion mutants Δ*ptsG*, Δ*ptsH*, Δ*ptsI*, Δ*ptsHIcrr*, Δ*mlc*, and Δ*cra* were prepared from wild type *E. coli* W3110 (ATCC #27325) using a one step gene inactivation method according to Datsenko and Wanner [*Proc. Natl. Acad. Sci. USA*. 2000, 97, 6640-6645]. In this method, a PCR insert was first created from a pKD4 template and primers with flanking regions homologous to start and end sequences of the gene to be inactivated, resulting in an insert with kanamycin resistance gene at the center, sandwiched between 2 FRT (FLP recognition target) sites, and outer homology regions. Wild type cells expressing λ Red recombinase (transformed with pKD46, grown at 30 °C) were transformed with the PCR insert, and gene sequence between the two homology regions were replaced with the FRT::Km::FRT sequence through recombination. Positive mutants were selected by plating on LB with kanamycin plates. To eliminate kanamycin resistance gene, the mutants were transformed with pCP20, a temperature sensitive plasmid expressing FLP. FLP expressed from this plasmid cut the Km region from the FRT::Km::FRT site, leaving one FRT site behind. pCP20 was then removed by growing the cells at 43 °C.

Plasmids pPLc28 encoding CRP* D53H, G141Q, G141K, and L148R [44] were donated by James C. Lee (University of Texas, Galveston, TX). CRP* strains ET24 and ET25 were donated by Winfried Boos (University of Konstanz, Konstanz, Germany). Coding regions of either plasmid or chromosomal *crp** gene were cloned into pUC19 with restriction sites BamHI and HindIII. For plasmid bearing *cyaA**, an insert coding a 48 kDa, C-terminal truncated AC was prepared according to Crasnier *et al.* [*Mol. Gen. Genet.* **1994**, *243*, 409-416], and cloned into pUC19 with restriction sites XbaI and PstI. Transformation of viable wild type cells was done using an electroporator (Eppendorf), followed by incubation in SOC medium for 1 hour at 37 °C, and selection on LB + ampicillin plates. Primers used for knockout mutation and cloning are given below.

Specific Aim 2. Evaluate the capability of SURS mutants to efficiently utilize sugar mixtures.

Regulatory networks in Carbon Catabolite Repression (CCR) are quite complex because of the high degree of interactions between associated genes, proteins, and metabolites. This complicates the interpretation of the results obtained during the evaluation of the aforementioned mutants. With this as motivation, we developed a novel tool (Network Decomposition Table, NDT) to help elucidate the network topology of regulatory systems. Through systematic enumeration of highest level mechanistic interactions, lower level interactions can be inferred in a self-consistent manner through arithmetic operations, and hence a compact framework into which all existing knowledge about the regulatory networks can be integrated. Application of NDT to the SURS networks is presented below.

Network decomposition table

It is clear that the cellular mechanisms of SURS form a complex network topology between genes, proteins, and metabolites. Even if only the central elements are considered (IIA^{glc}, AC, CRP-cAMP, glucose), many regulatory interactions are operative:
1. $IIA^{glc} \sim P$ activates AC.
2. AC converts ATP to cAMP.
3. cAMP activates CRP (forms CRP-cAMP complex).
4. CRP-cAMP modulates *crp* transcription (both positive and negative autoregulation).
5. CRP-cAMP represses *cyaA* transcription.
6. CRP-cAMP lowers the phosphorylation state of IIA^{glc}
7. Glucose represses *crp* transcription.
8. Glucose lowers the phosphorylation state of IIA^{glc} (through PTS transport).

These interactions can be represented graphically in a diagram such as Figure 1. However, as more elements are introduced into the network, more interactions will emerge, which adds more connecting lines to the diagram. Eventually, it will be difficult to come up with an aesthetic representation of the network because of crowding and intersecting lines. Even then, correlations between elements will not be immediately obvious. For example, glucose lowers cAMP level, but this is not apparent in Figure 1, because only mechanistic interactions are depicted (glucose-cAMP correlation has to be traced back through ATP, AC, and $IIA^{glc} \sim P$).

Motivated by the limitations of schematic diagrams as discussed above, we propose a novel scheme for representation of regulatory networks, i.e., Network Decomposition Table (NDT). NDT has interesting features: (1) all knowledge about regulatory interactions in the network can be integrated/summarized within a compact representation, (2) construction of NDT requires no aesthetic effort in spatial arrangement of the network elements, (3) new elements and knowledge can be easily appended, and conversely, smaller subsets of the network can be easily extracted, (4) correlations between elements are self-consistent and explicitly presented, and (5) everything follows simple arithmetic rules of addition and multiplication. From the name, decomposition implies that the network is broken down to its smallest elements, and built again from the ground up. The following is an example of application of NDT to the regulatory networks in SURS.

Consider a network subset with only three elements; *crp* gene, CRP protein, and extracellular glucose (Table 1). Gene notation in the table (e.g., *crp*) represents extent of transcription or mRNA abundance. Notations for metabolites (e.g., glucose) and non-enzyme

proteins (e.g., CRP) represent abundance. Elements are first enumerated in the first row and first column – the vertical ones are exerting / causing interactions to the horizontal ones. The diagonal cells are empty (self-interactions are not considered / relevant), so the number of possible entries is $n(n-1)$, where n is the number of elements.
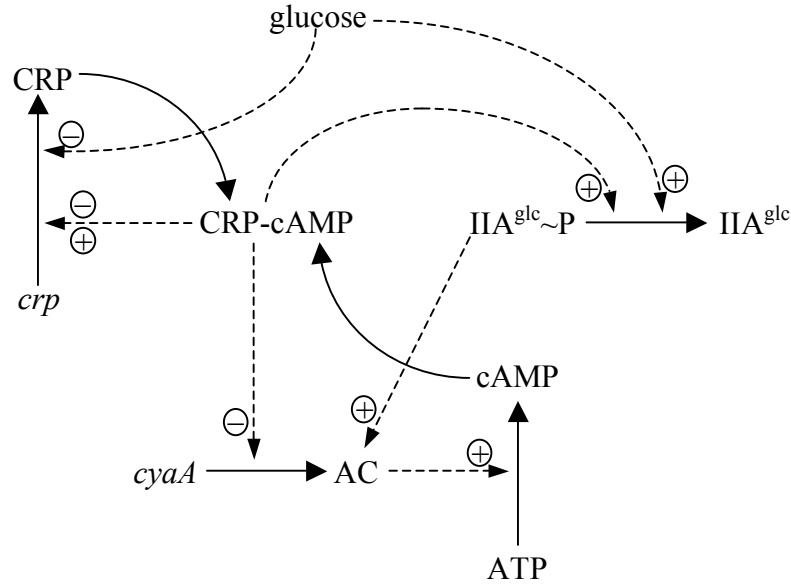

Figure 1. Network representation of central elements in CCR

Table 1. Decomposition of network elements

As *crp* encodes for CRP, the entry $\left\langle \dfrac{crp}{\text{CRP}} \right\rangle$, i.e., the influence / contribution of *crp*

transcription on CRP abundance is marked 1 (positive). Similarly, $\left\langle \dfrac{\text{glucose}_{\text{ext}}}{crp} \right\rangle$ is marked -1

because glucose represses *crp* transcription (transport of extracellular glucose is implied). These are mechanistic / highest level interactions, so the entries are in black. As will be explained later, the magnitude of these entries is less important than the direction (positive/negative). From the 2 entries, an inference can be made: glucose represses *crp* transcription, and *crp* encodes CRP; therefore, glucose represses CRP level. The inference can be represented by the notation:

$$\left\langle \frac{\text{glucose}_{\text{ext}}}{\text{CRP}} \right\rangle = \left\langle \frac{\text{glucose}_{\text{ext}}}{crp} \right\rangle \left\langle \frac{crp}{\text{CRP}} \right\rangle = (-1)(1) = -1$$

As this interaction is inferred rather than mechanistic, it is marked green in the table. Notice that the inference follows a simple arithmetic rule of multiplication, and exhibits the structure depicted by arrows on Table 1. If two entries interesect at an identity cell forming an L-shaped structure, the entry at the opposite diagonal can be calculated / inferred by multiplication. Also, note that the direction of inference (what causing what) is preserved, i.e., the inferred interaction is $\left\langle \dfrac{\text{glucose}_{\text{ext}}}{\text{CRP}} \right\rangle$, not $\left\langle \dfrac{\text{CRP}}{\text{glucose}_{\text{ext}}} \right\rangle$.

Now consider a bigger network, i.e., the one depicted in Figure 2 (Table 2). Enzyme notation (e.g., AC) represents activity; AC activation is represented by $\left\langle \dfrac{\text{IIA}^{\text{glc}} \sim \text{P}}{\text{AC}} \right\rangle = 1$. Note that both CRP and cAMP are considered positive influence on CRP-cAMP complex, $\left\langle \dfrac{\text{CRP}}{\text{CRP - cAMP}} \right\rangle = \left\langle \dfrac{\text{cAMP}}{\text{CRP - cAMP}} \right\rangle = 1$. Repression of *cyaA* transcription and IIA$^{\text{glc}}$~P dephosphorylation by CRP-cAMP are represented by $\left\langle \dfrac{\text{CRP - cAMP}}{cyaA} \right\rangle$ and $\left\langle \dfrac{\text{CRP - cAMP}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle$. Both are assigned a negative number but with smaller magnitude (-0.1); the reason will become apparent later. As the entry for $\left\langle \dfrac{\text{CRP - cAMP}}{\text{AC}} \right\rangle$ can be achieved from more than one direction, the inference is additive / follows arithmetic addition:

$$\left\langle \frac{\text{CRP - cAMP}}{\text{AC}} \right\rangle = \left\langle \frac{\text{CRP - cAMP}}{cyaA} \right\rangle \left\langle \frac{cyaA}{\text{AC}} \right\rangle + \left\langle \frac{\text{CRP - cAMP}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle \left\langle \frac{\text{IIA}^{\text{glc}} \sim \text{P}}{\text{AC}} \right\rangle$$

$$= (-0.1)(1) + (-0.1)(1) = -0.1 - 0.1 = -0.2 \text{ (shaded yellow)}.$$

The negative control exerted by $\left\langle \dfrac{\text{CRP - cAMP}}{\text{AC}} \right\rangle$ seems to propagate through the lower-level interactions $\left\langle \dfrac{\text{CRP - cAMP}}{\text{cAMP}} \right\rangle$, $\left\langle \dfrac{\text{CRP}}{\text{cAMP}} \right\rangle$, until it reaches

$$\left\langle \frac{crp}{\text{cAMP}} \right\rangle = \left\langle \frac{crp}{cyaA} \right\rangle \left\langle \frac{cyaA}{\text{cAMP}} \right\rangle + \left\langle \frac{crp}{\text{CRP - cAMP}} \right\rangle \left\langle \frac{\text{CRP - cAMP}}{\text{cAMP}} \right\rangle + \left\langle \frac{crp}{\text{CRP}} \right\rangle \left\langle \frac{\text{CRP}}{\text{cAMP}} \right\rangle$$
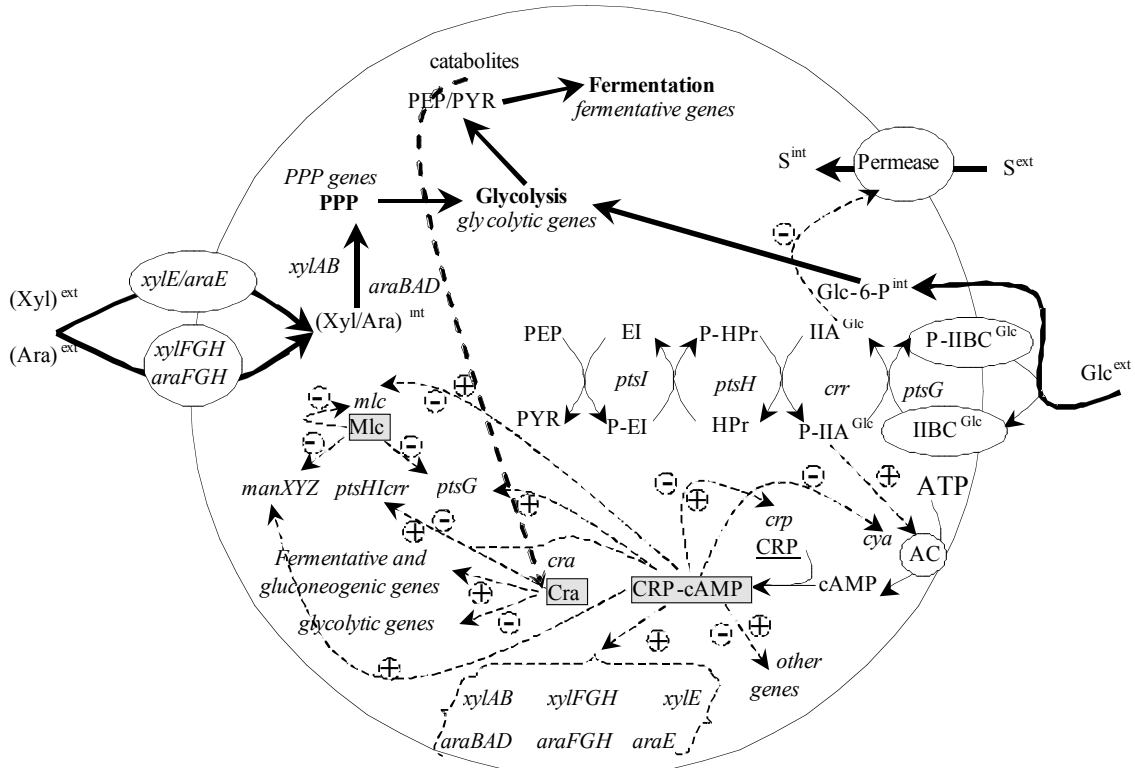
$$= -1.1$$

Figure 2. Schematic representation   of: (1) transport of sugars: glucose by PTS, xylose and arabinose by ABC and proton symport, and sugar S by an specific permease; (2) conversion of sugars into glycolytic intermediates; (3) metabolism of PEP/PYR via fermentative pathways; and (4) regulati   on of genes involved in the above pathways by PTS, and global regulators CRP   -cAMP, Cra, and Mlc. Individual genes and correspondent operons are illustrated by italics. Global regulators are shown inside shaded boxes. Broken lines indicate regulatory pathwa   ys. Thick arrows illustrate the flow of carbon from sugars into fermentation products. Thin arrows illustrate the reactions of PTS for glucose transport and phosphorylation.

Table 2. NDT of CRP, AC, IIA^{glc}, and glucose

	crp	CRP	CRP-cAMP	cya	AC	cAMP	crr	IIA^{glc}	IIA^{glc}~P	glucose^{ext}
crp		1	1	-0.2	-0.7	-1.1				
CRP			1	-0.1	-0.3	-0.5				
CRP-cAMP	-1	-1		-0.1	-0.2	-0.4			-0.1	
cya					1	1				
AC						1				
cAMP			1							
crr					1	1		1	1	
IIA^{glc}										
IIA^{glc}~P				1	1					
glucose^{ext}	-1	-1	-1	0.1	0.2	0.2				

This is consistent with Δcrp mutation: $\left\langle\dfrac{\Delta crp}{\text{cAMP}}\right\rangle = \left\langle\dfrac{\Delta crp}{crp}\right\rangle\left\langle\dfrac{crp}{\text{cAMP}}\right\rangle = (-1)(-1.1) = 1.1$, i.e., Δcrp mutants accumulate cAMP (note that Δcrp is actually not a network element, but the inference suits the notation very well). A similar propagation also happens in

$$\left\langle\frac{crp}{\text{AC}}\right\rangle = \left\langle\frac{crp}{cyaA}\right\rangle\left\langle\frac{cyaA}{\text{AC}}\right\rangle + \left\langle\frac{crp}{\text{CRP-cAMP}}\right\rangle\left\langle\frac{\text{CRP-cAMP}}{\text{AC}}\right\rangle + \left\langle\frac{crp}{\text{CRP}}\right\rangle\left\langle\frac{\text{CRP}}{\text{AC}}\right\rangle$$
$$= \text{-0.7}.$$

As $\left\langle\dfrac{cyaA}{\text{cAMP}}\right\rangle = \left\langle\dfrac{crr}{\text{cAMP}}\right\rangle = 1$, $\left\langle\dfrac{\Delta cyaA}{\text{cAMP}}\right\rangle = \left\langle\dfrac{\Delta cyaA}{cyaA}\right\rangle\left\langle\dfrac{cyaA}{\text{cAMP}}\right\rangle = (-1)(1) = -1$ and

$\left\langle\dfrac{\Delta crr}{\text{cAMP}}\right\rangle = \left\langle\dfrac{\Delta crr}{crr}\right\rangle\left\langle\dfrac{crr}{\text{cAMP}}\right\rangle = (-1)(1) = -1$, which is consistent with low cAMP levels in $\Delta cyaA$

and Δcrr mutants. One thing that is not consistent with CCR is a positive $\left\langle\dfrac{\text{glucose}^{\text{ext}}}{\text{cAMP}}\right\rangle$, which should be negative. This is because so far only positive influence of glucose on $cyaA$ transcription is taken into account; the major regulation of cAMP level comes from a negative influence by glucose through dephosphorylation of $\text{IIA}^{\text{glc}}\text{\textasciitilde}P$ via PTS transport. This part is explained in the following.

Balance of phosphorylated species in the PTS system is crucial in SURS. Therefore, the phosphoryl transfer reactions are represented by stoichiometric entries in NDT. Consider the first reaction: PEP + EI $\rightarrow$ EI\textasciitilde P + Pyruvate (Eq. 1). If EI\textasciitilde P level goes up, that means pyruvate also goes up, but PEP and EI go down, i.e., $\left\langle\dfrac{\text{EI\textasciitilde P}}{\text{Pyruvate}}\right\rangle = 1$, $\left\langle\dfrac{\text{EI\textasciitilde P}}{\text{PEP}}\right\rangle = \left\langle\dfrac{\text{EI\textasciitilde P}}{\text{EI}}\right\rangle = -1$ (third row in Table 3). The same stoichiometry is evident when pyruvate goes up (fifth row). Stoichiometric entries are considered mechanistic / highest level interactions, but are marked red to distinguish them from regulatory interactions. Entries arising from the reverse reaction may provide extra information – these are marked in blue (rows 2 and 4). With $\left\langle\dfrac{ptsI}{\text{EI}}\right\rangle = 1$, $\left\langle\dfrac{ptsI}{\text{EI\textasciitilde P}}\right\rangle$ actually can be inferred, i.e., $\left\langle\dfrac{ptsI}{\text{EI\textasciitilde P}}\right\rangle = \left\langle\dfrac{ptsI}{\text{EI}}\right\rangle\left\langle\dfrac{\text{EI}}{\text{EI\textasciitilde P}}\right\rangle = (1)(-1) = \text{-1}$. But this is not meaningful, as increase in $ptsI$ transcription does not lower EI\textasciitilde P level, unlike the stoichiometric balance of EI and EI\textasciitilde P. This is easily solved by specifying $\left\langle\dfrac{ptsI}{\text{EI\textasciitilde P}}\right\rangle = 1$ as highest-level interaction, which cannot be bypassed by inference.

Table 3. Stiochiometric representation of phosphoryl transfer reaction

	ptsI	EI	EI~P	PEP	Pyruvate
ptsI		1	1		
EI			-1	1	-1
EI~P		-1		-1	1
PEP		1	-1		-1
Pyruvate		-1	1	-1	

The complete phosphoryl trasfer reactions of the glucose PTS (Eqs. 1 – 5) are represented in Table 6. Some important inferences from this table include:

$$\left\langle \frac{ptsH}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle = \left\langle \frac{ptsH}{\text{HPr}} \right\rangle \left\langle \frac{\text{HPr}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle = (1)(1) = 1 \tag{7}$$

$$\begin{aligned}\left\langle \frac{ptsI}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle &= \left\langle \frac{ptsI}{\text{EI} \sim \text{P}} \right\rangle \left\langle \frac{\text{EI} \sim \text{P}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle = \left\langle \frac{ptsI}{\text{EI} \sim \text{P}} \right\rangle \left\langle \frac{\text{EI} \sim \text{P}}{\text{HPr} \sim \text{P}} \right\rangle \left\langle \frac{\text{HPr} \sim \text{P}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle \\ &= \left\langle \frac{ptsI}{\text{EI} \sim \text{P}} \right\rangle \left\langle \frac{\text{EI} \sim \text{P}}{\text{EI}} \right\rangle \left\langle \frac{\text{EI}}{\text{HPr} \sim \text{P}} \right\rangle \left\langle \frac{\text{HPr} \sim \text{P}}{\text{HPr}} \right\rangle \left\langle \frac{\text{HPr}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle \\ &= (1)(-1)(1)(-1)(1) = 1 \end{aligned} \tag{8}$$

$$\left\langle \frac{ptsG}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle = \left\langle \frac{ptsG}{\text{IICB}^{\text{glc}} \sim \text{P}} \right\rangle \left\langle \frac{\text{IICB}^{\text{glc}} \sim \text{P}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle = (1)(-1) = -1 \tag{9}$$

$$\left\langle \frac{\text{glucose}_{\text{ext}}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle = \left\langle \frac{\text{glucose}_{\text{ext}}}{\text{IICB}^{\text{glc}} \sim \text{P}} \right\rangle \left\langle \frac{\text{IICB}^{\text{glc}} \sim \text{P}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle = (1)(-1) = -1 \tag{10}$$

From (Eqs. 7-8) it is apparent now that both *ptsH* and *ptsI* are involved in the phosphorylation of IIA^{glc}. The result of $\Delta ptsH$ mutation is therefore $\left\langle \frac{\Delta ptsH}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle = \left\langle \frac{\Delta ptsH}{ptsH} \right\rangle \left\langle \frac{ptsH}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle = (-1)(1) = -1$, i.e., $\Delta ptsH$ mutants have low level of $\text{IIA}^{\text{glc}}\sim\text{P}$ (and thus low cAMP level). The same applies for $\Delta ptsI$ mutants. (Eq. 9) comes straight from the fact that IICB^{glc} dephosphorylates $\text{IIA}^{\text{glc}}\sim\text{P}$ – $\Delta ptsG$ mutation increases $\text{IIA}^{\text{glc}}\sim\text{P}$ level. Finally, (Eq. 10) shows that extracellular glucose, through PTS transport, lowers the phosphorylation state of $\text{IIA}^{\text{glc}}\sim\text{P}$.

Tables 2 and 4 can be combined to form a bigger network (Table 5). Now $\left\langle \dfrac{\text{glucose}^{\text{ext}}}{\text{cAMP}} \right\rangle$ is in the correct direction (negative) because both relief of *cyaA* transcription repression by CRP-cAMP and reduction of AC activity due to IIA$^{\text{glc}}$~P dephosphorylation are taken into account:

$$\left\langle \frac{\text{glucose}^{\text{ext}}}{\text{AC}} \right\rangle = \left\langle \frac{\text{glucose}^{\text{ext}}}{\text{CRP - cAMP}} \right\rangle\left\langle \frac{\text{CRP - cAMP}}{\text{AC}} \right\rangle + \left\langle \frac{\text{glucose}^{\text{ext}}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle\left\langle \frac{\text{IIA}^{\text{glc}} \sim \text{P}}{\text{AC}} \right\rangle \qquad (11)$$

$$= (-1)(-0.2) + (-1)(1) = -0.8$$

Therefore,

$$\left\langle \frac{\text{glucose}^{\text{ext}}}{\text{cAMP}} \right\rangle = \left\langle \frac{\text{glucose}^{\text{ext}}}{\text{AC}} \right\rangle\left\langle \frac{\text{AC}}{\text{cAMP}} \right\rangle + \left\langle \frac{\text{glucose}^{\text{ext}}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle\left\langle \frac{\text{IIA}^{\text{glc}} \sim \text{P}}{\text{cAMP}} \right\rangle$$

$$= (-0.8)(1) + (-1)(1) = -1.8$$

(Eq. 11) is the only additive entry in the table having opposite influences. This explains why negative control of CRP-cAMP were initially assigned a lower magnitude than the positive control of IIA$^{\text{glc}}$~P, i.e., $\left\langle \dfrac{\text{CRP - cAMP}}{cyaA} \right\rangle = \left\langle \dfrac{\text{CRP - cAMP}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle = -0.1$, $\left\langle \dfrac{\text{IIA}^{\text{glc}} \sim \text{P}}{\text{AC}} \right\rangle = 1$. Otherwise, the balance of (Eq. 11) will go in the wrong direction. This means that at this point, the NDT in Table 5 is meaningful only as far as the directions of control go, i.e. the positive or negative signs of the entries. The magnitudes of these entries become important only when an additive entry has both positive and negative influences, such as (Eq. 11).

New elements in Table 5 include LacY, which is under inducer exclusion $\left\langle \dfrac{\text{IIAglc}}{\text{LacY}} \right\rangle = -1$ and transcription repression $\left\langle \dfrac{\text{glucose}^{\text{ext}}}{\text{CRP - cAMP}} \right\rangle = -1$. Consequently, glucose inhibits lactose transport: $\left\langle \dfrac{\text{glucose}^{\text{ext}}}{\text{LacY}} \right\rangle = \left\langle \dfrac{\text{glucose}^{\text{ext}}}{lacY} \right\rangle\left\langle \dfrac{lacY}{\text{LacY}} \right\rangle + \left\langle \dfrac{\text{glucose}^{\text{ext}}}{\text{IIA}^{\text{glc}}} \right\rangle\left\langle \dfrac{\text{IIA}^{\text{glc}}}{\text{LacY}} \right\rangle = (-1)(1)+ (1)(-1) = -2$. Transcription of the xylose operon is repressed by glucose, i.e., $\left\langle \dfrac{\text{glucose}^{\text{ext}}}{xyl} \right\rangle = \left\langle \dfrac{\text{glucose}^{\text{ext}}}{\text{CRP} - \text{cAMP}} \right\rangle\left\langle \dfrac{\text{CRP} - \text{cAMP}}{xyl} \right\rangle = (-1)(1) = -1$.

ptsG transcription is downregulated by Mlc, $\left\langle \dfrac{\text{Mlc}}{ptsG} \right\rangle = -1$. *ptsG* transcription is correlated to transcription of the xylose operon:

$$\left\langle \frac{ptsG}{xyl} \right\rangle = \left\langle \frac{ptsG}{\text{IICB}^{\text{glc}} \sim \text{P}} \right\rangle\left\langle \frac{\text{IICB}^{\text{glc}} \sim \text{P}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle\left\langle \frac{\text{IIA}^{\text{glc}} \sim \text{P}}{\text{cAMP}} \right\rangle\left\langle \frac{\text{cAMP}}{\text{CRP - cAMP}} \right\rangle\left\langle \frac{\text{CRP} - \text{cAMP}}{xyl} \right\rangle$$

$$= (1)(-1)(1)(1)(1) = -1.$$

Therefore, $\Delta ptsG$ upregulates *xyl* transcription:

$$\left\langle \frac{\Delta ptsG}{xyl} \right\rangle = \left\langle \frac{\Delta ptsG}{ptsG} \right\rangle \left\langle \frac{ptsG}{xyl} \right\rangle = (-1)(-1) = 1. \tag{12}$$

This example of application to SURS shows how NDT can integrate and summarize knowledge about regulatory networks into a compact form, where correlations between network elements can be explicitly presented. Once mechanistic interactions are set as highest-level entries, other interactions can be inferred in a self-consistent manner through simple rules of arithmetic multiplication and addition. Perhaps the most interesting aspect of NDT is that it can show how exertion of regulatory control can propagate through cooperativity with lower-level interactions, such as in the case of $\left\langle \frac{crp}{cAMP} \right\rangle$.

Table 4. Phosphoryl transfer reactions of glucose PTS

	crr	IIA^{glc}	$IIA^{glc}{\sim}P$	$ptsI$	EI	$EI{\sim}P$	PEP	$Pyruvate$	$ptsH$	HPr	$HPr{\sim}P$	$ptsG$	IIB^{glc}	$IIB^{glc}{\sim}P$	$glucose^{ext}$	$G6P$
crr		1	1													
IIA^{glc}			-1										-1	1		
$IIA^{glc}{\sim}P$		-1								1	-1					
$ptsI$			1		1	1										
EI						-1				-1	1					
$EI{\sim}P$			1		-1		-1	1			-1					
PEP																
$Pyruvate$					-1	1	-1									
$ptsH$			1							1	1					
HPr		-1	1								-1					
$HPr{\sim}P$			-1	1	-1					-1						
$ptsG$			-1										1	1		
IIB^{glc}														-1	-1	1
$IIB^{glc}{\sim}P$		1	-1										-1			
$glucose^{ext}$			-1										-1	1		-1
$G6P$													1	-1	-1	

In order to elucidate the topology of a regulatory network, first it is broken down to the smallest elements; genes, proteins, metabolites, phoshorylated/unphosphorylated species, allosteric effectors, and transcriptional repressors/activators. By arrangement of these n elements into a $n \times n$ table, the network is then rebuilt from the ground up through complete enumeration of all possible $n(n-1)$ interactions. This is easily done using a spreadsheet, which also prevents circular reference in building inferences, and accomodates network expansion/reduction by simple row and column insertion/deletion.

Table 5. Network decomposition table accounting for essential elements and regulatory mechanisms in CCR

	crp	CRP	CRP-cAMP	cyaA	AC	cAMP	crr	IIA^{glc}	IIA^{glc}~P	$glucose^{ext}$	ptsI	EI	EI~P	PEP	Pyruvate	ptsH	HPr	HPr~P	ptsG	IIB^{glc}	IIB^{glc}~P	G6P	xyl	lacY	LacY	Mlc
crp		1	1	-0.2	-0.7	-1.1																				
CRP			1	-0.1	-0.3	-0.5																				
CRP-cAMP	-1	-1		-0.1	-0.2	-0.4			-0.1														1	1		
cyaA					1	1																				
AC						1																				
cAMP				1																						
crr					1	1		1	1																	
IIA^{glc}									-1											-1	1				-1	
IIA^{glc}~P					1	1	-1										1	-1								
$glucose^{ext}$	-1	-1	-1	0.1	-0.8	-1.8	1		-1											-1	1	-1	-1	-1	-2	
ptsI		1			1	1			1			1	1													
EI													-1				-1	1								
EI~P									1			-1		-1	1			-1								
PEP																										
Pyruvate												-1	1	-1												
ptsH		1			1	1			1								1	1								
HPr							-1		1									-1								
HPr~P									-1			1	-1				-1									
ptsG			-1		-1	-1			-1											1	1		-1			
IIB^{glc}									-1												-1	1				
IIB^{glc}~P							1		-1												-1					
G6P									-1												1	-1				
xyl																										
lacY																									1	
LacY																										
Mlc																		-1								

Some immediate improvements to NDT are in order. Lower-level/inferred interactions (green) are already distinguished from mechanistic/highest-level interactions (black, red, blue), but a proper hierarchy can be displayed with an extended color coding scheme, e.g., pink for second-level interactions, cyan for third-level interactions, etc. Note that negative *crp* autoregulation ($\left\langle \dfrac{\text{CRP - cAMP}}{crp} \right\rangle = -1$) is included in the example, but this rules out positive autoregulation. This can be solved by increasing the NDT resolution to account for different promoter/operator sites. As CRP-cAMP binds to different operator sites in negative and positive autoregulation [*Mol. Microbiol.* **1992**, *6*, 2489–2497], this can be represented by $\left\langle \dfrac{\text{CRP - cAMP}}{crp_1} \right\rangle = -1$ and $\left\langle \dfrac{\text{CRP - cAMP}}{crp_2} \right\rangle = 1$, with *crp₁* and *crp₂* indicating extent of *crp* transcription initiated from sites 1 and 2.

As mentioned previously, at this point the significance of NDT only goes so far as elucidating qualitative control. When an interaction has both positive and negative influences, e.g., (Eq. 11), the magnitude of the corresponding mechanistic/highest-level entries should be appropriately assigned so that the resulting interaction goes in the right direction. It may be possible to achieve a more quantitative representation of regulatory control, in which case the NDT entries will be analogous to the control coefficients in Metabolic Control Analysis [*Trends Biochem. Sci.* **1985**, *10*, 16], but this will require experimental data. For example, comparison of cAMP levels in Δ*crp* and Δ*crr* mutants will be useful in assigning values for $\left\langle \dfrac{crp}{\text{cAMP}} \right\rangle$ and $\left\langle \dfrac{crr}{\text{cAMP}} \right\rangle$. However, these are very low-level interactions in NDT, which would not be directly applicable to the highest-level mechanistic interactions, $\left\langle \dfrac{\text{CRP - cAMP}}{cyaA} \right\rangle$, $\left\langle \dfrac{\text{CRP - cAMP}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle$, and $\left\langle \dfrac{\text{IIA}^{\text{glc}} \sim \text{P}}{\text{AC}} \right\rangle$.

The ultimate goal of whole-cell modeling is to account for changes in cellular state in a dynamic setting. Kremling *et al.* described a kinetic model of *E. coli* exhibiting glucose-lactose diauxie, taking into account the glucose PTS system, lactose transport, inducer exclusion, and transcription of *ptsG*, *cyaA*, *crp* and activation of their corresponding proteins [*Metab. Eng.* **2001**, *3*, 362-379]. However, not all known mechanistic interactions are included in the model, e.g., $\left\langle \dfrac{\text{glucose}_{\text{ext}}}{crp} \right\rangle$, $\left\langle \dfrac{\text{CRP - cAMP}}{\text{IIA}^{\text{glc}} \sim \text{P}} \right\rangle$. In this respect, the value of NDT is that it serves as a complete catalog of all known regulatory mechanisms, so it will be easy to detect if an important interaction is left out. Although NDT aims to achieve different goals than dynamic cell models, vertical entries in a column represent factors influencing the element assigned to that column, analogous to the positive and negative terms in a kinetic expression. Therefore, from the kinetic point of view, NDT entries can be thought of as instantaneous representation of all factors influencing a dynamic variable.

It would be interesting to see application of NDT to other systems. The regulatory network for flagellar biosynthesis in *E. coli* has been studied well. The dynamic model of Kalir and Alon [*Cell.* **2004**, *117*, 713-720] is based on a structured diagram connecting the master

13

regulators FlhDC and FliA to transcriptional units of the flagellar genes with activation coefficients. It appears that conversion of this diagram to NDT would be particularly suitable.

Phenotypic characterization of PTS mutants

Our strategy includes characterization of the growth and sugar consumption profiles of PTS and related mutants: Δ*ptsG*, Δ*ptsH*, Δ*ptsI*, Δ*ptsHIcrr*, Δ*mlc*, and Δ*cra*, as well as CRP* and AC*. Cell culture in batch mode is advantageous because the CCR phenotype (diauxie/no diauxie) can be directly observed, and consumption/production rates can be related to industrial scale fermentation, which is commonly done in batch mode as well.

W3110, 2 g/L glucose + 2 g/L xylose ($\mu = 0.41$/h)

W3110, 5 g/L glucose + 5 g/L xylose ($\mu = 0.61$/h)

Δ*ptsG*, 2 g/L glucose + 2 g/L xylose ($\mu = 0.18$/h)

Δ*ptsG*, 5 g/L glucose + 5 g/L xylose ($\mu = 0.33$/h)
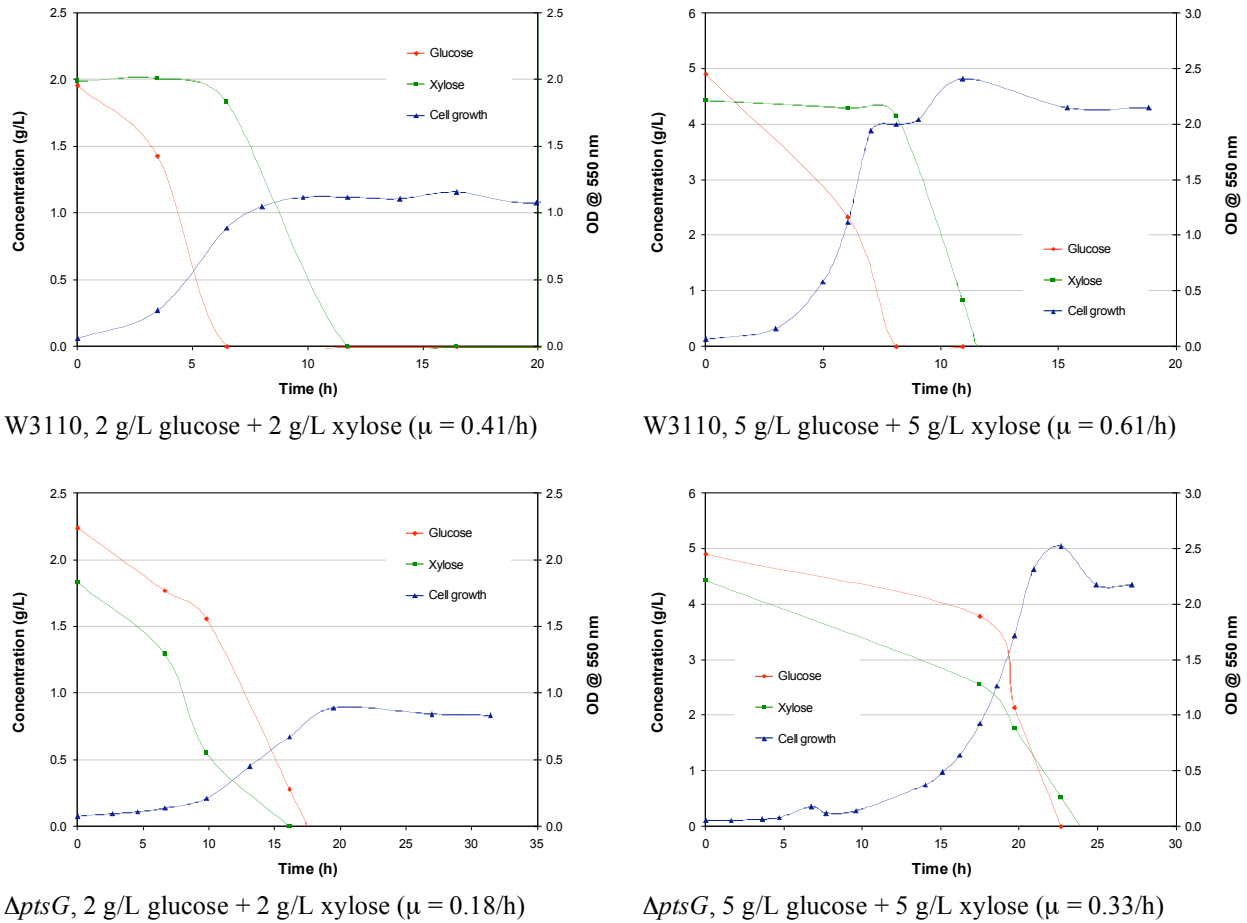
Figure 3. Growth and sugar consumption profiles of W3110 and Δ*ptsG*. CCR is relieved in Δ*ptsG* as indicated by lack of diauxie and simultaneous consumption of glucose and xylose.

Figure 3 shows the growth and sugar consumption profiles of W3110 and Δ*ptsG*. As expected, W3110 shows exhibits CCR; diauxic growth and sequential consumption of glucose and xylose. Δ*ptsG*, shows simultaneous consumption of glucose and xylose and no diauxic growth. CCR is relieved in Δ*ptsG* because glucose transport through PTS is impaired – the cell still can transport glucose through GalP or the mannose PTS, although with lower efficiency / slower rate. This is evident in the dramatic change in specific growth rate (about 50% decrease). In terms of CCR mechanism, *ptsG* inactivation means IIA^{glc}~P is not dephosphorylated by

IICB^{glc} because it is not expressed. Therefore, cAMP level is not repressed, and thus xylose genes are expressed. In NDT notation, this is inferred as $\left\langle \dfrac{\Delta ptsG}{xyl} \right\rangle = 1$ (Eq. 12).

The CCR behavior of W3110 or $\Delta ptsG$ does not change with sugar concentration, but this is not the case in $\Delta ptsH$ and $\Delta ptsI$ (Figure 4). Both of these strains exhibit CCR when grown on 5 g/L glucose + 5 g/L xylose (high sugar), but not on 2 g/L glucose + 2 g/L xylose (low sugar). It is surprising that glucose is consumed at all. $\Delta ptsH$ and $\Delta ptsI$ are not supposed to be able to utilize PTS carbohydrates (except fructose), because phosphoryl transfer to the corresponding enzyme II cannot occur. To explain glucose consumption, it must be assumed that PTS transport somehow resumes due to substitute phosphoryl donors, bypassing the inactivated gene (direct phosphorylation of HPr for $\Delta ptsI$, or direct phosphorylation of IIA^{glc} for $\Delta ptsH$). There is no immediate explanation of how this substitution can proceed. Note that *galP* expression is repressed during growth on glucose [*Biotechnol. Bioeng.* **2003**, *83*, 687-694]. IIA^{glc} can be phosphorylated by FPr, but this occurs only during growth on fructose. Phosphorylation of IIA^{glc} by acetate kinase also requires EI and HPr [*J. Biol. Chem.* **1986**, *261*, 13498-13503]. EI can be phosphorylated by EI kinase [*J. Biol. Chem.* **1996**, *271*, 15285–15291], but this does not help in the case of $\Delta ptsI$ or $\Delta ptsH$. HPr can be phosphorylated by HPrK (HPr kinase) but only in gram-positive bacteria (*E. coli* is gram-negative) [*J. Bacteriol.* **2003**, *185*, 4003-4010].

Xylose consumption in these strains is also contrary to published reports [*Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed.; Niedhardt, F.C., Curtiss III, R., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M., Umbarger, H.E., Eds.; American Society for Microbiology: Washington, D.C., 1996; pp 1149-1174] – xylose (class II compound) requires high cAMP for its gene expression, but cAMP level is low in $\Delta ptsI$ and $\Delta ptsH$ because of lack of phosphorylation in IIA^{glc}. Assuming that the glucose PTS in fact works and there is enough residual cAMP to activate xylose genes in these strains, changes in CCR behavior with sugar level yet more difficult to explain. Repression (low cAMP) at high sugar level and derepression (high cAMP) at low sugar level occurs in wild type, but the threshold for glucose is around 0.3 mM [*Microbiology-(UK)*. **1997** *143*, 1909-1918]. This is much lower than both glucose levels studied (11 and 28 mM). There should not be a transition between high and low cAMP levels within these ranges of extracellular glucose.

Considering only xylose consumption at 5 g/L, if xylose were a class I compound (subject to inducer exclusion), the CCR behavior of $\Delta ptsG$, $\Delta ptsI$, $\Delta ptsH$, and $\Delta ptsHIcrr$ would actually make sense. $\Delta ptsG$ consumes xylose, due to the same arguments as before. $\Delta ptsI$ and $\Delta ptsH$ do not consume xylose because IIA^{glc} would mostly be in unphosphorylated state, and therefore exerts inducer exclusion on xylose. Inducer exclusion is abolished in $\Delta ptsHIcrr$ (lacking IIA^{glc}), and therefore xylose consumption is resumed. However, xylose is not a class I compound and thus the rationale does not apply.

There is a slight complication with the $\Delta ptsH$ mutation. Because *ptsH* is a small gene (258 bp), the primer design was so difficult that although a significant portion of the gene was deleted, the active site (His-15) was remaining. However, the $\Delta ptsI$ mutant bears successful deletion of the His-189 active site and exhibits similar CCR behavior, assuring that deactivation in *ptsH* in the $\Delta ptsH$ mutation was in fact successful.

The sugar consumption behavior of $\Delta ptsI$, $\Delta ptsH$, and $\Delta ptsHIcrr$ does not agree with literature. It may be necessary to repeat the mutation verification to make sure that the strains
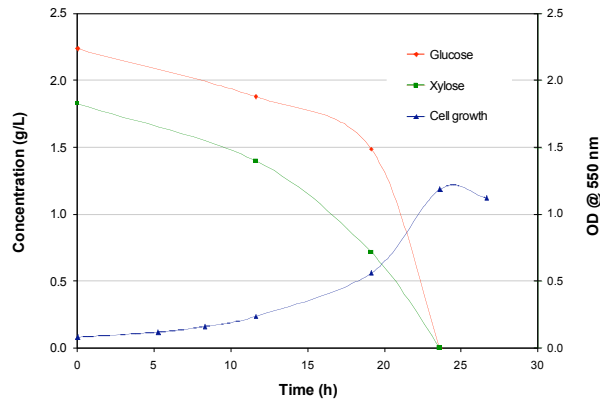
bear the correct genotype. Also, measurement of intracellular cAMP levels will provide a direct clue for assessment of CCR phenotype.
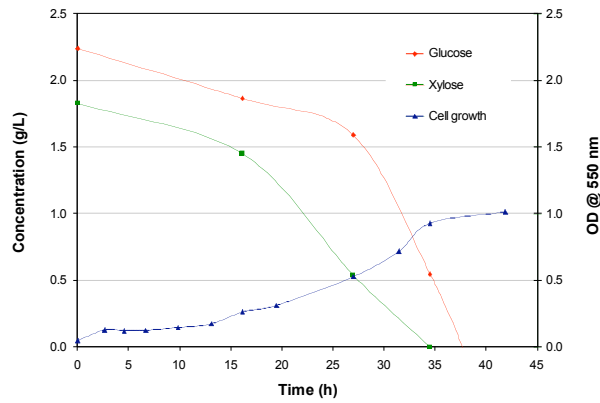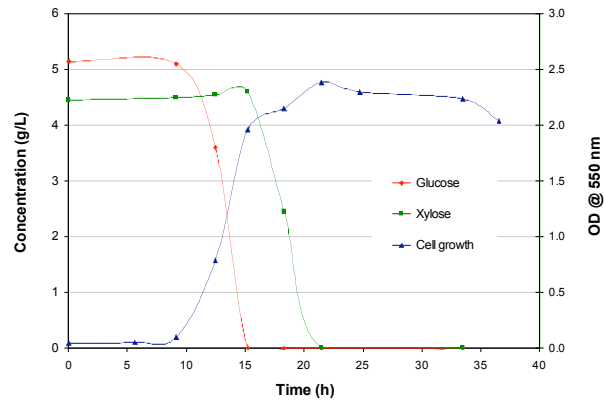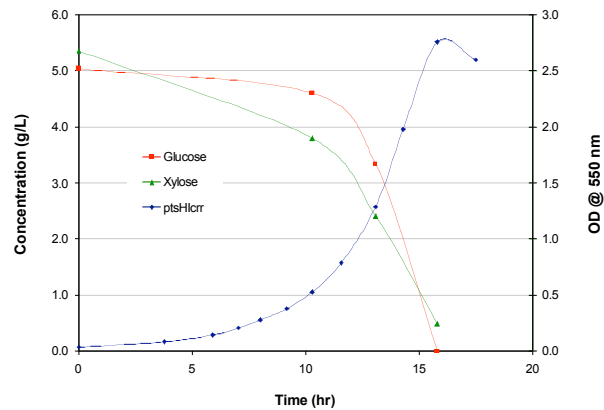

Δ*ptsH*, 2 g/L glucose + 2 g/L xylose
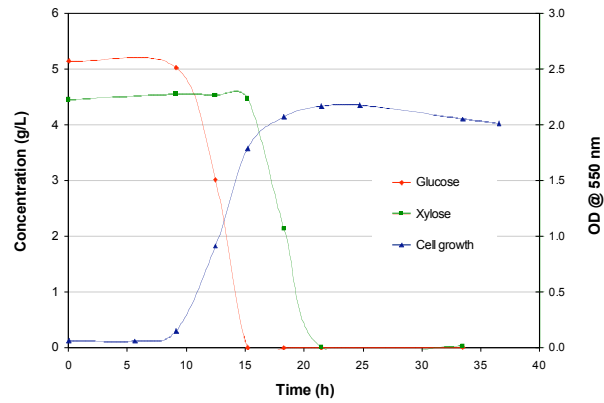
Δ*ptsH*, 5 g/L glucose + 5 g/L xylose

Δ*ptsI*, 2 g/L glucose + 2 g/L xylose

Δ*ptsI*, 5 g/L glucose + 5 g/L xylose

Δ*ptsHIcrr*, 5 g/L glucose + 5 g/L xylose

Figure 4. Growth and sugar consumption profiles of Δ*ptsH*, Δ*ptsI*, and Δ*ptsHIcrr*. Δ*ptsH* and Δ*ptsI* mutants exhibit unexpected behavior in that xylose utilization is not repressed at low sugar level, but repressed at high sugar level, which results in a diauxie. This CCR phenotype is not evident in the D*ptsHIcrr* mutant.

16
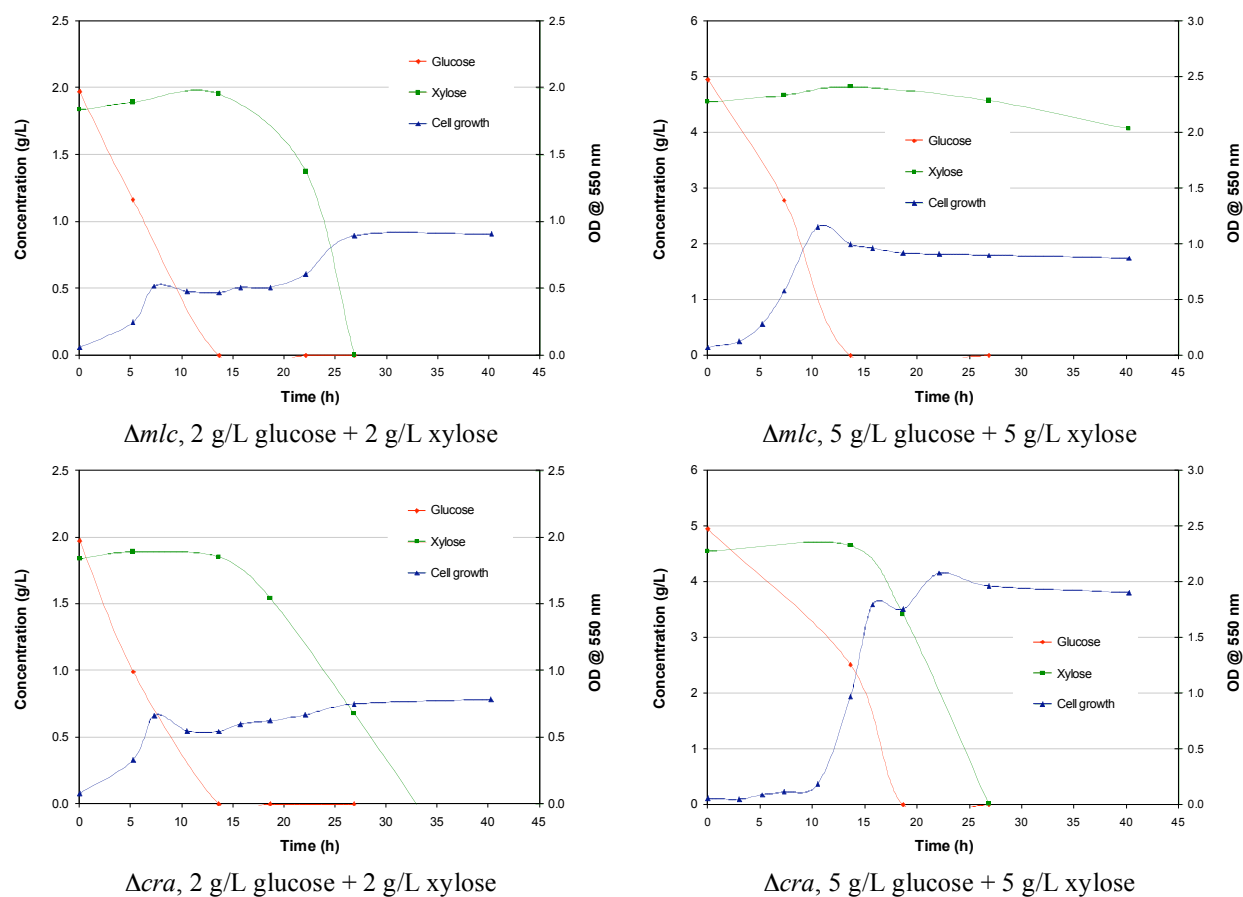
Figure 5. Growth and sugar consumption profiles of Δ*mlc* and Δ*cra*.

Growth and sugar consumption profiles for Δ*mlc* and Δ*cra* are shown in Figure 5. Both strains exhibit CCR at low and high sugar levels. Mlc is a transcriptional repressor of *ptsG*, so Δ*mlc* mutation will result in upregulation of IICB^{glc} expression, which translates to higher dephosphorylation of IIA^{glc}, low cAMP, and therefore CCR. The second phase of the diauxic growth is not apparent in Δ*mlc* at high sugar, because xylose transport is apparently impaired. This could be caused by the repressive effect of constitutive *ptsG* expression as discussed above, although the phenotype may be leaky. Out of four trials, two cultures exhibit this behavior, but two exhibit normal behavior with second diauxic growth phase and complete xylose consumption. Cra regulates downstream carbon utilization by acting as transcriptional activator of gluconeogenic genes and repressor of fermentative genes. A Δ*cra* mutation therefore would result in upregulation of fermentative pathways. This change does not seem to affect CCR behavior though, because the phenotype is essentially that of wild type.

17

Specific Aim 3. Evaluate changes in the regulation of gene expression at genomic scale resulting from engineering *E. coli* SURS using DNA microarray technology.

Transcriptional profiling of ΔptsG vs. W3110

The use of DNA microarray offers a system-wide view into the transcriptional profile of the cell. Characterization of PTS mutants with DNA microarray may highlight some important changes arising from pleiotropic effects. Knowledge gained from the transcriptional analysis will serve as a feedback in the overall metabolic engineering strategy. As a start, comparison of Δ*ptsG* vs. W3110 is chosen because of the obvious difference in the CCR phenotypes.

Total RNA extraction from Δ*ptsG* and W3110 cells yielded good quality RNA; $A_{260/280}$ values were 2.1 for both samples. Figure 6 shows the gel electrophoresis image of the samples, with sharp bands for the 15S and 23S RNA. Figure 7 shows the overlay image of Cy3 and Cy5 channels for array set 1.
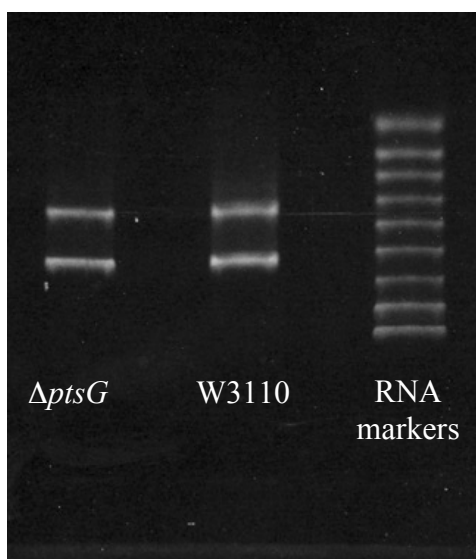

Figure 6. Gel electrophoresis image of total RNA samples

Upregulation of xylose-related genes is evident in Δ*ptsG* vs. W3110 (Table 6). The two array sets serve as two data points of $\log_2$-transformed ratios, from which the mean, standard deviation, and coefficient of variation can be calculated. The mean fold change is calculated as 2^μ. Although xylose transport and degradation genes are significantly upregulated, the genes for the non-oxidative branch of pentose phosphate pathway further downstream of xylose metabolism (*rpe*, *rpiA*, *rpiB*, *talA*, *talB*, *tktA*, *tktB*) are not. They are either only slightly upregulated (fold change < 2), have opposing signs in set 1 and 2, or even slightly downregulated.
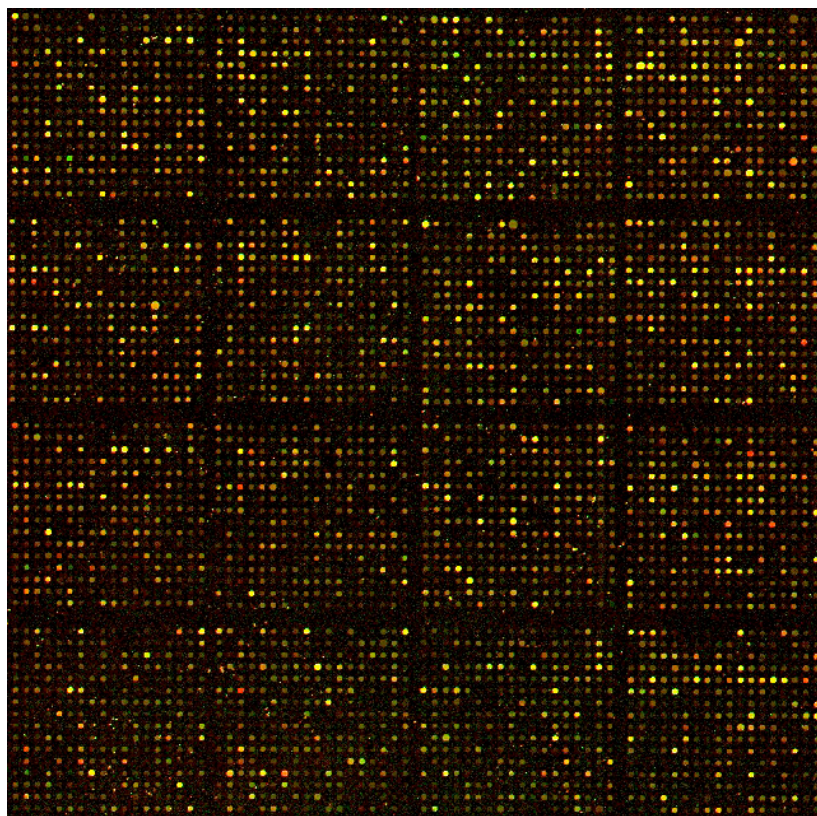
Figure 7. Overlay image of *ptsG* **(Cy5) and W3110 (Cy3) channels for array set 1**

Table 6. Upregulation of xylose genes in Δ*ptsG* vs. W3110

Gene	Description	Set 1 (log2)	Set 2 (log2)	μ (log2)	σ	σ/μ	Mean Fold change
xylA	D-xylose isomerase	4.430	4.654	4.542	0.158	0.03	23.3
xylB	xylulokinase	2.925	3.003	2.964	0.055	0.02	7.8
xylE	xylose-proton symport	5.613	4.826	5.219	0.557	0.11	37.3
xylF	xylose binding protein transport system	5.602	5.556	5.579	0.032	0.01	47.8
xylG	putative ATP-binding protein of xylose transport system	4.822	5.301	5.062	0.339	0.07	33.4
xylH	putative xylose transport, membrane component	4.908	4.774	4.841	0.095	0.02	28.7
xylR	putative regulator of *xyl* operon	4.733	3.534	4.134	0.848	0.21	17.6

Gonzalez *et al.* reported a similar unresponsive transcriptional behavior for an ethanologenic *E. coli* strain [*Biotechnol. Prog.* **2002**, *18*, 6-20]. During growth on xylose vs. glucose, carbon flux through PPP increased, but the associated transcripts did not show significant upregulation. From the literature, it looks like transcriptional regulation of the PPP genes in *E.coli* has not been studied extensively. However, our data suggests that CRP-cAMP mediated regulation is decoupled from transcriptional regulation in PPP, i.e., PPP genes do not belong to CRP-cAMP regulon.

Interestingly, flagellar biosynthesis genes are greatly downregulated in Δ*ptsG* vs. W3110 (Table 7). It has been reported that acetyl phosphate, the intermediate in AcCoA – acetate conversion is a signaling molecule in flagellar biosynthesis / biofilm development (acetyl phosphate downregulates expression of flagellar genes) [*J. Bacteriol.* **1995**, *177*, 4696-4702]. However, genes for the first and second step in the AcCoA → acetyl phosphate → acetate reaction (*pta* and *ack*) are both downregulated to the same extent (-2.1 and -2.0), so it is hard to infer something about the accumulation of acetyl phosphate in Δ*ptsG*. Nevertheless, our data suggests an interesting link between CCR and flagellar biosynthesis, and serves as an example of how a single mutation can lead to system-wide effects, including seemingly unrelated cellular functions.

Table 7. Downregulation of flagellar genes in Δ*ptsG* vs. W3110.

Gene	Mean fold change	Gene	Mean fold change	Gene	Mean fold change
flgA	-11.4	*flgK*	-6.3	*fliH*	-2.7
flgC	-47.4	*flgL*	-2.6	*fliI*	-1.6
flgD	-21.3	*flgM*	-4.6	*fliJ*	-14.4
flgE	-21.9	*flhB*	-2.3	*fliK*	-5.8
flgF	-4.9	*fliA*	-57.2	*fliL*	-8.6
flgG	-10.6	*fliD*	-3.7	*fliM*	-26.9
flgH	-5.8	*fliE*	-10.6	*fliN*	-18.6
flgI	-4.0	*fliF*	-7.3	*fliO*	-9.8
flgJ	-7.8	*fliG*	-1.9	*fliQ*	-3.9
flhD	-1.1	*flhC*	-1.3	*fliS*	-16.6

Specific Aim 4. Identify functional metabolic pathways and quantify their fluxes in SURS mutants and wild type strains by using a novel flux analysis technique based on bondomer analysis of 2D NMR bond-labeling experiments.

The activities and findings relevant to this Specific Aim are presented in three parts as follows: (1) *E. coli* reaction network development, (2) Calculation of metabolic fluxes using conventional Metabolic Flux Analysis (c-MFA), and (3) Labeling experiment and analysis of isotopomer distribution. Our initial efforts (and therefore the results presented here) focused on the implementation of the method for wild-type W3110 during growth on glucose as the only sugar (carbon and energy source).

E. coli reaction network development

For purpose of developing the bioreaction network model of fermentative metabolism in *E. coli*, several sources were referred and all the metabolic reactions were considered which are known to be active under anaerobic conditions. The use of minimum media necessitates

considering biosynthetic reaction since the sugar (glucose or xylose) is the sole source of biosynthetic carbon. Thus, the aim was to develop the comprehensive reaction network model including all the known biosynthetic reactions for building blocks in addition to the central metabolic pathways. The model thus developed contained 273 reactions (central metabolic pathways, biosynthetic pathways, transport reactions) and 242 metabolites (substrates, intermediates, cofactors and products). The biosynthetic reactions consisted of reactions for synthesis of the building blocks and after that average macromolecule composition of building blocks were taken from literature. A maintenance reaction was included whose flux would be able to tell how much ATP is consumed for general maintenance of cell. In addition to these reactions, transhydrogenase reaction was also included in the model. Using this model, a stoichiometric matrix was developed, by taking balances around intermediate metabolites, for the purpose of flux calculation via conventional flux analysis. In addition to the balances on intermediate metabolites, balances were taken for certain reducing and energy cofactors (NADH, NADPH and ATP). However, the model ran into the problem of having several linearly dependent reactions. Several reactions were lumped and some of the metabolites (of similar function) were pooled in order to remove the linearly dependent reaction. The resulting model consisted of 222 reactions and 205 metabolites. This model met the criteria of linearly independent reactions for c-MFA. However, usage of this model for flux calculation led to practically unrealizable fluxes (negative fluxes for irreversible reactions). Due to complexity of the model, it was difficult to analyze the problem. However, it was found that the stoichiometric matrix had a condition number (degree of precision required for the measurements) of 1037 which is very high for the fermentation measurements. It is stated by Stephanopoulos et. al. ["*Metabolic Engineering: principles and methodology.*" Academic Press, California, 1998] that if the condition number is above 100, the matrix is ill-conditioned for flux calculations using extracellular measurements. Thus, the approach to solve this problem was to further reduce the model by lumping the biosynthetic reactions. This was done by taking single step reactions from precursor metabolites to the building blocks, the model for this being constructed from information available in literature. The stoichiometric matrix constructed for this model had a condition number of 73 and thus it seemed to be a reasonable model to work with. This model had 82 reactions and 68 metabolites and the reaction network for this model is listed in appendix A. To test this model, a synthetic data (manually generated data which had good carbon and redox balance) was used for flux calculation and the resulting fluxes had a good carbon closure (error <0.1%) verifying validity of the model.

The model developed for purpose of flux calculation via c-MFA could not be used as such for the purpose of n-MFA because:
- For reversible reactions, c-MFA model considered net forward reaction.
- The c-MFA model did not include information for carbon-carbon rearrangement information.

A new model was developed by incorporating these information into the c-MFA model and verified by calculating fluxes from synthetic data and the carbon balance closure for resulting fluxes was found to be good (<1%). Also, flux identifiablity analysis was done for this model and it was found that the fluxes were identifiable.

Calculation of metabolic fluxes using conventional Metabolic Flux Analysis (c-MFA)

The HPLC data generated by analysis of fermentation samples needs to be transformed into fluxes in order to use it as an input for intracellular flux calculation. While evaluating

different methods to calculate fluxes from the fermentation data, it was found that the fluxes were constant during the exponential phase as depicted in Fig. 8. This shows that cells grow in a state of balanced growth during the exponential phase and flux calculation at any point during the exponential phase should yield essentially the same result. However, due to errors associated with rate calculations at a point, it is advisable to calculate the fluxes for a duration in which there is significant change in the concentration of various products. Another aspect developed in regard to extracellular flux calculation was to divide the rate of formation of product/ consumption of substrate by time-averaged cell density instead of average cell density. Also, for the purpose of rate calculations, Figure 8 shows that method using a polynomial fit for concentration profile is as good as taking linear profile during exponential phase.

The extracellular and biosynthetic fluxes calculated from product concentration profiles and cell growth profiles were used to calculate the intracellular fluxes, which are shown in Figure 9. Among the most relevant features of this flux map are the very low flux through the Pentose Phosphate Pathway (PPP), the activity of the enzyme pyruvate dehydrogenase (PDH), and the activity of the transhydrogenases (interconverting NAD(H) and NADP(H)).
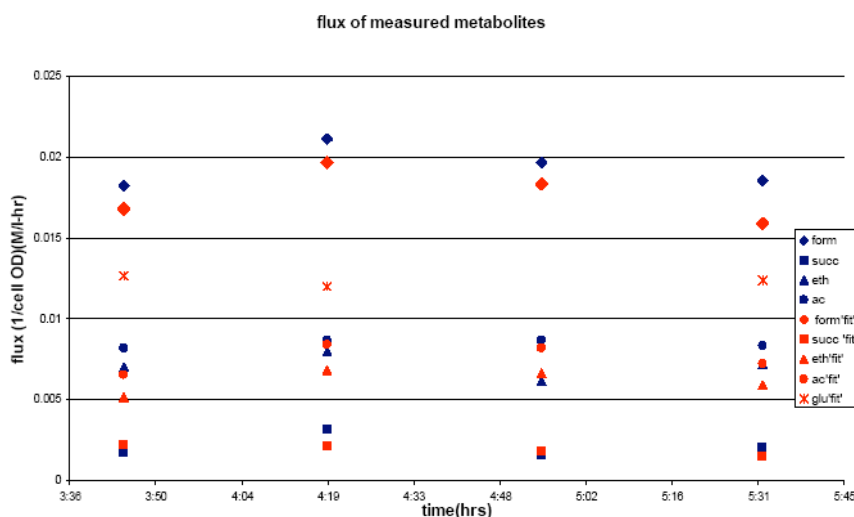

Figure 8. Fluxes at different time points during exponential phase. The fluxes denoted by blue symbols are calculated by considering linear profile. The fluxes represented by red symbols are calculated by fitting the profile into a polynomial equation and getting an analytical solution.
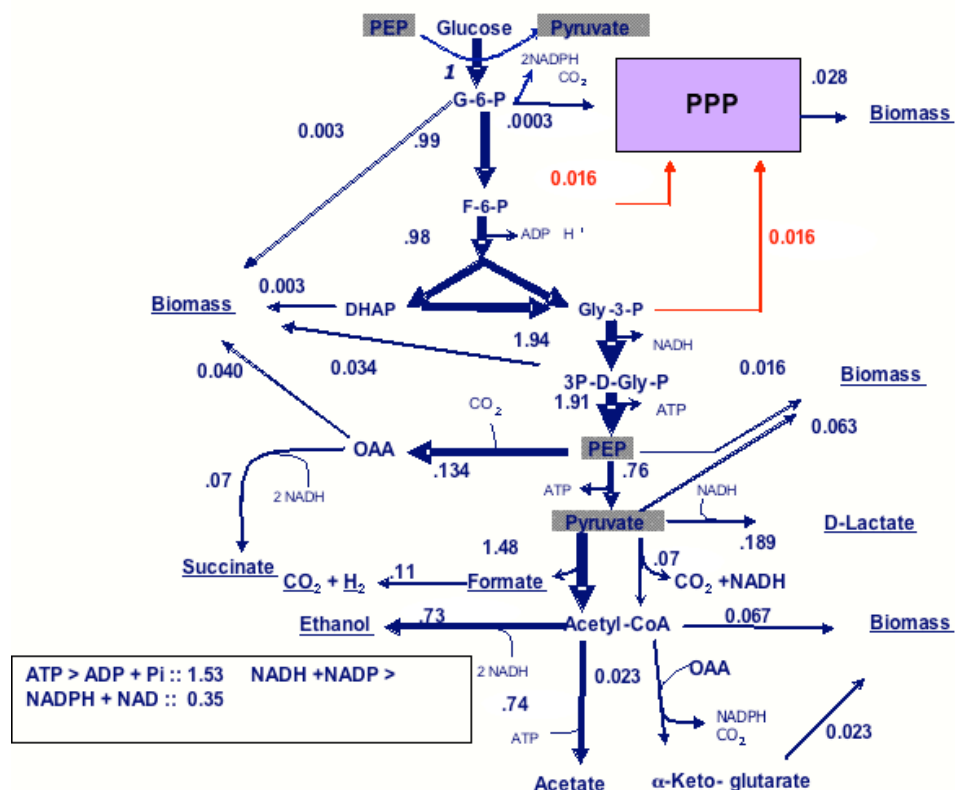
Figure 9. Metabolic flux distribution in E. coli W3110 during the anaerobic fermentation of glucose in minimum medium. Units are mmoles of substrate/g cell/h.

Labeling experiment and analysis of isotopomer distribution

To gain further evidence regarding the observed distribution of metabolic fluxes, due to the power of NMR to elucidate internal fluxes to a high degree of resolution and accuracy as compared to conventional methods (c-MFA), a fermentation was performed with uniformly labeled ^{13}C glucose. Cell biomass was harvested during exponential phase, hydrolyzed and analyzed for the labeling pattern of amino acids via 2D HSQC NMR spectroscopy. The spectrum obtained from NMR spectroscopy was analyzed by software named "NMR view" and Figure 10 shows the peak assignment for most of the identifiable peaks. The intensities of these peaks were calculated using the same software. For some of the carbon atoms of amino acids, the peaks were not well resolved and spectral analysis was done to get the intensities. The intensities were then converted to fractional isotopomer distribution.

Information about carbon-carbon bond connectivities in pyruvate could be obtained from six aliphatic carbon positions: α-Ala and α-Val which are derived from C_2 of pyruvate and, β-Val, γ^1-Val, γ^2-Ile and δ^1-Leu which are derived from C_3 of pyruvate. In case of PEP, α-Tyr and α-Phe are derived from C_2 and, β-Tyr and β-Phe are derived from C_3 of PEP. For these carbon positions, 'f' values are calculated which give information about carbon-carbon bond connectivities [*Eur. J. Biochem.* **232**, 433-448, 1995]. For the terminal carbon, $f^{(2)}$ denotes fraction having intact bond with neighboring carbon and $f^{(1)}$ denotes fraction having a neighboring carbon arising from different source molecule. For central carbon, $f^{(3)}$ denotes

fraction with both neighbors intact, $f^{(2a)}$ denotes fraction with intact neighboring carbon showing smaller scalar coupling, $f^{(2b)}$ denotes fraction with intact neighbor showing larger scalar coupling and $f^{(1)}$ denotes fraction having both neighbors originating from different source molecule. The analytical method to calculate these fractions was developed by Szyperski [*Eur. J. Biochem.* **232**, 433-448, 1995]. Looking at the $f^{(2)}$ values of terminal carbon atoms in Table 8, it could be seen that the fraction of intact C2-C3 carbon bonds is very high (close to 1) implying that these bonds were not broken via PPP. This implies low flux through PPP, although it is not a conclusive evidence. The $f^{(3)}$ fraction of α-Ala, is 0.28 and its comparison to the conservation of C3-C2 connectivity implies that there is a significant cleavage of C1-C2 bond in pyruvate. Pyruvate is derived form PEP and thus it is important to investigate C1-C2 connectivity in PEP. From the 'f' values of α-Tyr and α-Phe, it could be calculated the 95% of C1-C2 connectivity in PEP is conserved, thus providing evidence that most of the PEP and Pyruvate originate from single source glucose. This suggests that the flux through PPP is very low as compared to the flux through EMP and that the cleavage of C1-C2 connectivity in pyruvate is due to the action of pyruvate formate lyase which is similar to that found by Szyperski [*Eur. J. Biochem.* **232**, 433-448, 1995]. Also, since conversion of PEP to pyruvate is reversible, even the small amount of C1-C2 bond cleavage is not entirely due to action of PPP, thereby suggesting that flux through PPP is very low. This finding supports our c-MFA result that flux through PPP is very low as compared to flux through EMP in the wild type W3110.
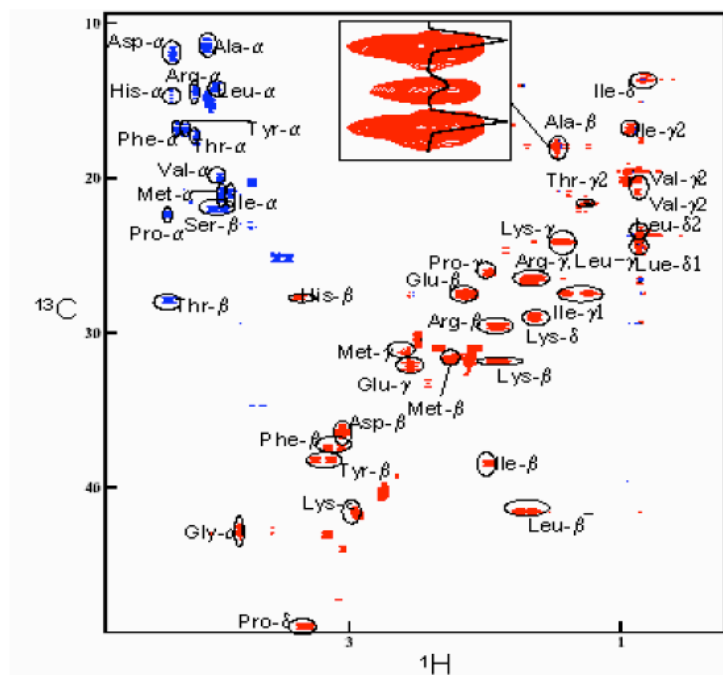

Figure 10. 2D HSQC NMR spectrum of the hydrolyzed sample for W3110. Each spot represent the peak for one carbon atom of an amino acid. The spot for Ala-b is blown up in the inset to demonstrate the peaks.

Table 8. Relative intensities of ^{13}C multiplet components and derived relative abundances of intact C_2 and C_3 fragments in amino acids.

^{13}C fine structuer	In precursor	Observed relative intensity				Relative abundance of intact carbon fragments			
Terminal		I_s	I_d			$f^{(1)}$	$f^{(2)}$		
β-Ala	C3 of Pyr	.119	.88			.03	.97		
$γ^1$-val	C3 of Pyr	.146	.854			.07	.93		
$γ^2$-Ile	C3 of Pyr	.118	.882			.03	.97		
$δ^1$-Leu	C3 of Pyr	.146	.854			.07	.93		
Central in C_3		I_s	I_{da}	I_{db}	I_{dd}	$f^{(1)}$	$f^{(2a)}$	$f^{(2b)}$	$f^{(3)}$
α-Ala	C2 of Pyr	.099	.014	.556	.33	.016	.006	.684	.287
α-Val	C2 of Pyr	.593	.315	.068	.024	.718	.302	-.007	-.015
α-Tyr	C2 of PEP	.112	.011	.156	.72	.027	.008	.188	.77
α-Phe	C2 of PEP	.117	.01	.122	.751	.034	.005	.145	.81
β-Phe	C3 of PEP	.123	.82	0	.055	.054	1.00	-.01	-.04

NMR-based Metabolic Flux Analysis (NMR-MFA)

The intracellular fluxes were solved by minimizing the deviation between real and simulated data. Extra cellular fluxes of Ethanol/Acetate and formate and all available measurements of relative multiplet intensities were used in the formulation of error criterion. The objective function to be minimized was defined as

$$Chi - square = \frac{(I_{sim} - I_{exp})^2}{N_{exp}^2} + \frac{(F_{mes} - F_{sim})^2}{N^2}$$

There are 41 net reactions rates and 13 exchange rates i.e. total of 54 parameters to be estimated. 22 intracellular metabolite pools contribute linear constraint and L=17 extra cellular fluxes are directly measured but two of them (Formate and Acetate/Ethanol) are not used. Hence, labeling data have to supply the information for the remaining 54-15-22=17 degree of freedom. TCA cycle, Pentose and Glyoxalytae fluxes were taken as free fluxes.

NMR2Flux which uses Simulated Annealing (SA) with Powell method to find global optimum was used to estimate the intracellular fluxes.
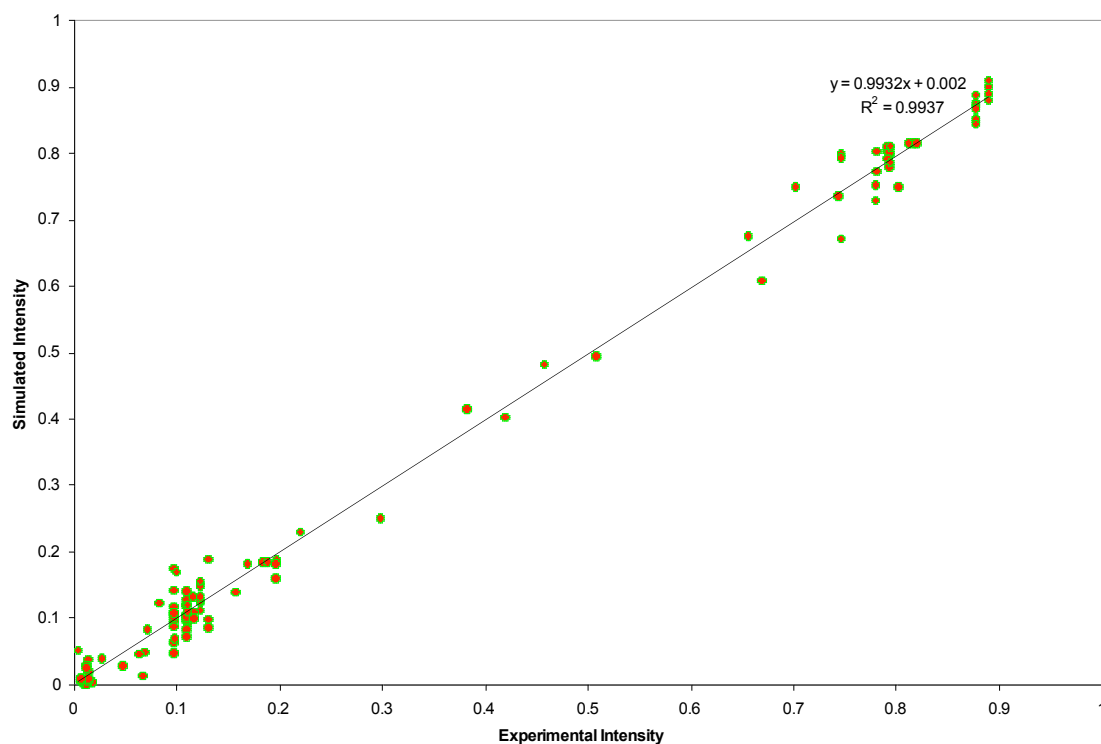
Figure 11. Comparison of experimental and simulated NMR intensities of amino acids for 1st experiment.

Figure 11 shows the simulated and experimental intensities. Figure 12 shows a metabolic flux map of two labeling experiments. Pentose flux was found to be 4.48 and 12. Table 9 lists literature value for pentose fluxes in E. coli under anaerobic condition

Table 9

Pentose Flux		
Strain	**Source**	**Value**
JM101	Uwe Sauer 1999 J bact	less 25
E.coli B	Neilsen ,1999,J.Biotech	77
K-12 W3110	cMFA	0.003

It has been reported that flux through TCA cycle(*suc*) to be small under anaerobic condition. Similarly, Glyoxalyate shunt flux is expected to be small under anaerobic condition
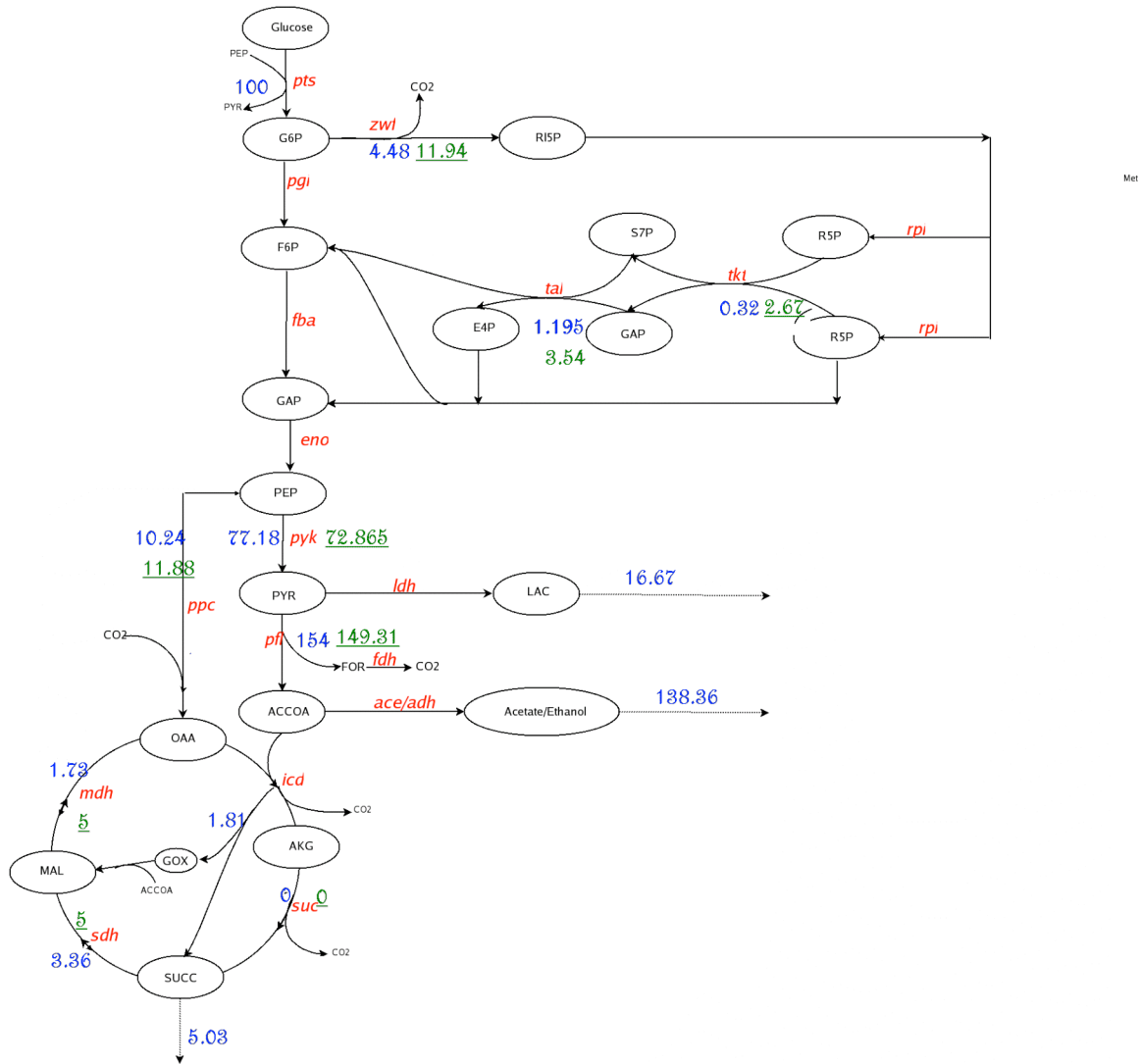
Figure 12. Metabolic Flux Map for two labeling experiment

NADH, NADPH and ATP Balances

Table 10: 1ˢᵗ Experiment

	Net Formed	Required for Biomass synthesis	Balance
NADH	23.35	8.5	14.85
NADPH	11.54	43.68	-32.14
ATP	169.45	44.31	125.14

NADH and NADPH balance were found to be not closed. Even NADH+NADPH balance is not closed. This suggests *pdh* may be active. If flux thorough *fdh* is assumed to be zero, and *pdh* is assumed to be active, flux through *pdh* is 13.44. ATP maintenance was found to be 125 compared to 153 in the conventional MFA.

Identifiability

Table 11 shows pentose and TCA fluxes estimated by the NMFA without extracellular measurement, CMFA and NMFA with extra cellular fluxes. Clearly, amino acids intensities without extra cellular measurements are not sufficient to estimate metabolic fluxes.

Table 11

	CMFA	NMFA+extra cellular	NMFA without extra cellular meas
Pentose	2.5 - 12	3.85 - 6.4	0 - 50
TCA		0 - 0	0 - 50
Glyoxylate		1.13 - 1.78	0 - 2.6

Simulations were done to find whether fluxes indentifiablity can be improved in case when extra cellular fluxes are not available or they are associated with large error. Figure 13 shows the result. The maximum improvement (2 times) was found to be for 100% 1- labeled glucose. To verify these results, synthetic intensities were used as experimental intensities and NMR2Flux was used to estimate fluxes. Mixture of 10% U-labeled and 50% 1-labeled was able to estimate fluxes correctly whereas 10% U –labeled and 20% 1-labeled was not able to do so.
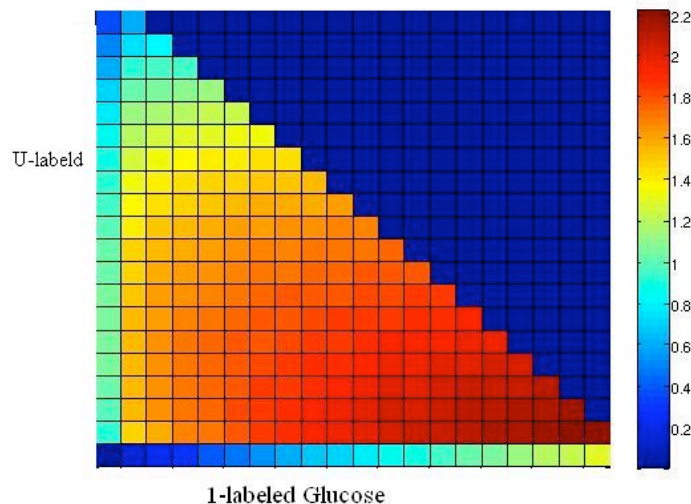

Figure 13: Optimal Experiment design for *E. coli* grown on mixture of 1-labeled and U-labeled Glucose

With the mixture of 1-labeled and U-labeled glucose, *mez* and ED pathway fluxes are identifiable which not the case is in the U-labeled experiment
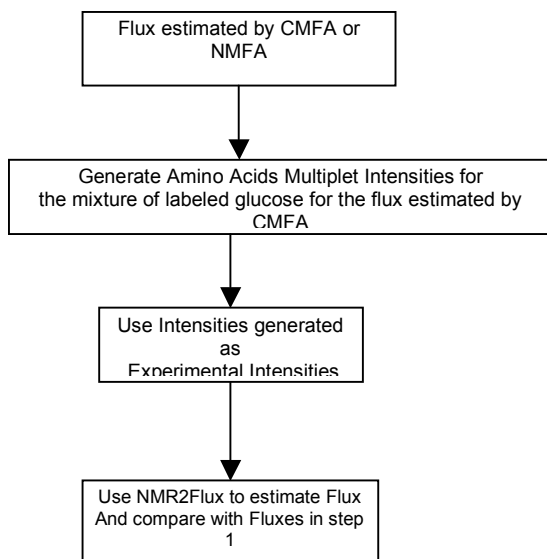
Figure 14: The algorithm used to verify the identifiability results.

Specific Aim 5. Integrate these results using a novel genetic network based MFA model that combines gene expression and metabolic flux data.

We are developing a framework for the development of an algorithm that is capable of capturing the inner working of sugar-utilization regulatory systems in response to the availability of various carbon sources. We are currently embarking in two major efforts:

1. In the first project, we are constructing a mechanistic based model to describe the xylose utilization system under the control of a global transcription factor Crp. As a result, the gene expression and enzymatic levels of those genes related to xylose uptake and utilization will be regulated by Crp.

2. Concurrently, we are constructing a mechanistic based glucose utilization model which involves several regulatory molecules, such as cAMP, Mlc, and Crp.

An important feature of this modeling approach is that the output from the glucose utilization model will provide pertinent information about the transcription factor Crp which will in turns regulates the xylose utilization system. We expect that the model will be able to describe the dynamics of simultaneous presence of both carbon sources once the two sub-models are integrated together. The current model structure was built on exist knowledge about the regulatory mechanisms of the glucose and xylose systems. The parameter values used in the model are based on data reported in the literature.

Results on two additional problems studied in this project: (a) HPLC optimization and (b) characterization of the hybridization process in DNA microarray using molecular dynamics simulation

(a) Computer-assisted optimization of HPLC separation for simultaneous quantification of substrates and products in microbial fermentation

Monitoring of *E. coli* fermentation requires quantification of substrates and products in the culture medium. Simultaneous separation, identification, and quantification of medium components by HPLC provide a fast and convenient assay, but is not always possible due to peak overlaps. This problem can be solved by optimizing the operating conditions to achieve the best peak separation. HPLC optimization strategy consists of four elements; experimental design, retention modeling, quality criteria function, and optimum search method. In this part of the work we present a simple, superior alternative to general classes of classical resolution functions (*S* function) and a novel optimum search algorithm (ISS, iterative stochastic search) for HPLC optimization.

The alternative global resolution function *S* for HPLC optimization proposed in this work shows superior performance when compared to general classes of quality criteria functions (R_s, R_p, R_{min}), including correct assessment of favorable separation conditions, preservation of individual peak pair contributions, elimination of arbitrary cut-off values, and a unique capability to interpret absolute significance of function values through a simple inequality. The novel global optimization algorithm (iterative stochastic search, ISS) also developed in this work shows clear advantages over existing algorithms for HPLC optimization (grid search, simplex search, simulated annealing) in its ability to correctly identify the global optimum (instead of local optimum), with higher precision, with more efficient use of computation cycles, and with easier implementation. Based on a case study using a hyperdimensional function with many optima, robust performance of ISS also suggests its possible application in simultaneous, higher-dimensional HPLC optimizations. Successful application of *S* and ISS to HPLC optimization was demonstrated in the separation of ten representative substrates and products found in microbial fermentation. Excellent agreement was found between actual and predicted values for the optimized (best) and pathological (worst) conditions. The overall optimization strategy successfully implemented in this work can be generalized to any HPLC optimization problem. For future work, the retention model will be expanded to include more compounds related to *E. coli* fermentation: lactose, isocitrate, 2,3-butanediol, propionic acid, butyric acid, and MOPS, and the OPTIMIZE software will be updated.

(b) Mesoscopic simulation of the hybridization process in oligonucleotide microarray by molecular dynamics

DNA microarray is established as a powerful tool in global genomic analysis. However, the hybridization process, in which target cDNA molecules specifically bind to their complementary probes on the array, is still largely uncharacterized. We conducted a molecular dynamics simulation of the hybridization of small oligonucleotide target and probes (8 bases), using a simple bead and spring model to represent the DNA molecules, and an implicit treatment of the solvent (water) using a distance-dependent dielectric function. A score function is defined to monitor the extent of hybridization of target molecules during simulation. To compensate for the lack of dihedral orientation in the beads, an algorithm is implemented to ensure that hydrogen bonding occurs in one-to-one correspondence. The significance of the simple model in designing oligonucleotide probes is discussed, and many improvements to the model are proposed.

With the motivation of characterizing the hybridization process in a microarray experiment, the model can be applied particularly to the design of oligonucleotides in microarrays. Currently, the oligos are designed through a bioinformatics search to find the most unique patterns in the genes of interest. A fast, row-resolution simulation such as the one described in this work could be utilized as a validation tool.

Many simplifications are used in the present model, but improvements are forthcoming. Easy modifications can readily make the model more realistic. The restriction that prevents H-bonds formation within the same molecule can easily be lifted. Also, higher bond-energy in G-C base pairs can be described in the model by insertion of a few lines in the code. The Berendsen thermostat [*J. Chem. Phys.* **1984**, *81*, 3684-3690] has been used in many MD simulations of DNA, and the implementation is simple enough to be incorporated in the model with minimal modification.

For a more realistic representation of the oligonuleotides, it is crucial that the model be reworked using real units. Levitt *et al*. [*Comput. Phys. Comm.* **1995**, *91*, 215-231] tabulated model parameter values for bond length, stretching, as well as bond and dihedral angles for MD simulations of DNA. Although these values are reported at the atomic resolution, approximation for the bead model can be made considering the C4'-C5'-PO4-C3' connection along the backbone of the oligonucleotides. Finally, a larger system (both in oligo size and number probe and target molecules) should be studied to more accurately capture the kinetic behavior of the system.

6. Identify products developed under the award and technology transfer activities, such as:

a. Publications (list journal name, volume, issue), conference papers, or other public releases of results. If not provided previously, attach or send copies of any public releases to the DOE Project Officer identified in Block 11 of the Notice of Financial Assistance Award;

The reported work has been featured in the following presentations and papers:

Publications
- Dharmadi, Y., and R. Gonzalez. (2005). A Better Global Resolution Function and a Novel Iterative Stochastic Search Method for Optimization of HPLC Separation. *J. Chromatogr. A* 1070: 89-101
- Dharmadi, Y., and R. Gonzalez. (2004). DNA Microarrays: Experimental Issues, Data Analysis, and Application to Bacterial Systems. *Biotechnol. Prog.* 20 (5): 1309-1324.

Presentations:
Invited Presentations
- Understanding and Manipulating the Anaerobic Metabolism of *E. coli*: A Systems Biology Approach. Presented at: (1) Ninth Annual Symposium on Industrial and Fermentation Microbiology, Friday, April 29, 2005. The Radisson Center, La Crosse, WI; (2) Innovations at the Interface of Polymers and Biology Symposium, Polytechnic University, New York, May 10-12, 2005; and (3) Department of Veterinary Microbiology and Preventive Medicine, Iowa Sate University, December 7, 2004.

Contributed Presentations
- Dharmadi,Y., and Gonzalez, R. A network decomposition framework for integration of knowledge on regulatory networks in biological systems. AIChE 2005 Annual Meeting, October 30-November 4, Cincinnati, OH.

- Dharmadi,Y., and Gonzalez, R. Computer-assisted optimization of HPLC separation for simultaneous quantification of substrates and products in microbial fermentation. AIChE 2005 Annual Meeting, October 30-November 4, Cincinnati, OH.
- Peacock, R., Zhu, J., Shanks, J.V., Gonzalez, R. and San, K-Y. Development of a mechanistic model for sugar-utilization regulatory systems. AIChE 2005 Annual Meeting, October 30-November 4, Cincinnati, OH.
- Dharmadi, Y., Moon, Y., Shanks, J. V., San, K-Y., and Gonzalez, R. A Systems Biology Approach for Metabolic Engineering of *Escherichia coli* Sugar-Utilization Regulatory Systems. 2004 AIChE Annual Meeting, November 7 - 12, Austin, TX.
- Dharmadi, Y., Lamm, M., and Gonzalez, R. Mesoscopic Simulation of the Hybridization Process in Oligonucleotide Microarray by Molecular Dynamics. 2004 AIChE Annual Meeting, November 7 - 12, Austin, TX.
- Dharmadi, Y., and Gonzalez, R. A Better Global Resolution Function and a Novel Iterative Stochastic Search Method for Optimization of HPLC Separation. 2004 AIChE Annual Meeting, November 7 - 12, Austin, TX.
- San, K-Y., Gonzalez, R., Shanks, J. V., and Peacock, R. Development of a genetic network driven model for sugar-utilization regulatory systems. 2004 AIChE Annual Meeting, November 7 - 12, Austin, TX.
- Dharmadi, Y., Moon, Y., Shanks, J. V., San, K-Y., and Gonzalez, R. Metabolic Engineering of *E. coli* Sugar-Utilization Regulatory Systems for the Consumption of Sugar Mixtures. Metabolic Engineering V: Genome to Product. September 19-23, 2004, Lake Tahoe, CA.

b. Web site or other Internet sites that reflect the results of this project;

None.

c. Networks or collaborations fostered;

None.

d. Other products, such as data or databases, physical collections, audio or video, software or netware, models, educational aid or curricula, instruments or equipment.

NMR2Flux software has been developed by Dr. Shanks group to estimate the metabolic fluxes in plants. The original program was specific to the plant system (developing soybean embryos). Hence, it cannot be used for the *E. coli* system without modification. The program has been modified so that it can be used for any organism without any changes. The user has to just provide the program with metabolic network model of the organism in the form of input files along with experimental conditions and data.

7. For projects involving computer modeling, provide the following information with the final report:

In recent years, metabolic flux analysis (MFA) has become an important tool in metabolic engineering. The results of MFA are metabolic flux maps which can be used for the

systematic study of cellular responses to genetic perturbations. In some cases, however, intracellular metabolic fluxes can be estimated by doing balance for intracellular metabolites, and solving the resulting set of linear equations. This is called stoichiometric flux analysis or conventional flux analysis. The stoichiometric flux analysis fails in the case of parallel reactions and metabolic cycles. In such cases, ^{13}C labeling based MFA can be used to estimate fluxes.

The key assumptions of ^{13}C based MFA are that the organism does not distinguish between ^{12}C and ^{13}C labeled substrate and it is at metabolic and isotopic steady state i.e intracellular metabolite concentration and its isotopic distribution does not change with time. In such an experiment, a mixture of ^{12}C and ^{13}C substrate is fed to the biological system. The labeled carbon atoms are then distributed all over the metabolic network Finally the isotopic distribution in the amino acids pools can be measured by NMR or MS instruments. The resulting data provide a large amount of additional information to quantify the intracellular fluxes.

a. **Model description, key assumptions, version, source and intended use;**

The calculation of metabolic fluxes from labeling data requires a model of the metabolic network. The metabolic model for *E.coli* under anaerobic condition has been developed is shown in Figure 1. It includes all principal pathways of primary metabolism (PTS transport of glucose, glycolysis, pentose phosphate pathway, TCA cycle, glyoxylate shunt, and fermentative reactions) and the biosynthetic pathways that convert primary metabolic precursors to sink metabolites (not shown). The TCA cycle is not complete under anaerobic conditions but has been assumed complete to make the model comprehensive and account for the effect of leftover oxygen present.

The reactions in the model were assumed reversible unless information on irreversibility was available. All reversible reactions were modeled as two fluxes. The reaction from succinate to malate (Mal) in the TCA cycle can lead to an inversion of the labeling pattern, owing to the fact that succinate is a symmetrical molecule while Mal is not . To account for this fact, this reaction was modeled as two parallel fluxes, one that conserves the carbon skeleton and another that inverts the same. The few reactions have been lumped to reduce the computation time.
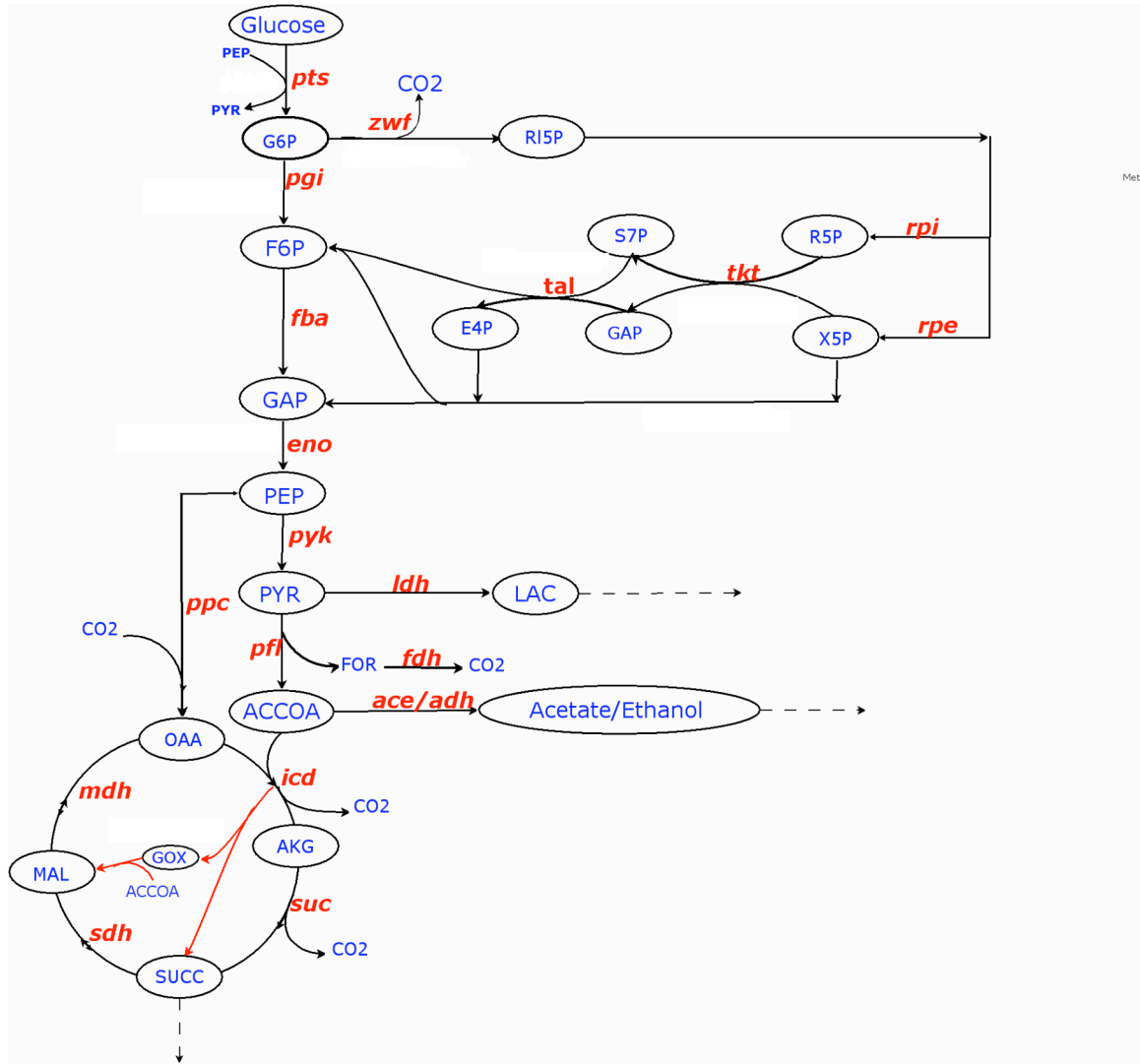
Figure 1: Metabolic Network model for E.coli under anaerobic condition

The key assumption of the model is that it accounts for the reactions active in the central carbon metabolism under the experimental conditions and also accounts for the carbon rearrangements due to these reactions. The model can be used for evaluating intracellular fluxes for *E. coli* under anaerobic conditions.

The sources of information for the primary metabolic and biosynthetic pathways in the model were the recent literature on *E. coli* biochemistry (Neidhardt et al, 2005; Szyperski, 1995) and the online source Ecocyc (Keseler et al,2005). These sources also provided information on the precursors of the sink metabolites and carbon rearrangement.

b. Performance criteria for the model related to the intended use;

NMR2Flux requires the model to evaluate the fluxes for simulating the labeling experiment. The model has to account for the observed labeling data and there has to be good fit between the experimental and simulated data.

c. Test results to demonstrate the model performance criteria were met (e.g., code verification/validation, sensitivity analyses, history matching with lab or field data, as appropriate);

The model has been used to estimate fluxes for the experiment data reported by Szyperski (1995) and there was a good fit between experimental and simulated data. Schmidt et al (1999) used same experimental data to estimate the metabolic fluxes. However, our flux map was a little different from those obtained by them and we had a more optimal chi-squared value – this difference can be attributed to difference in the algorithm used.

A ^{13}C labeling experiment was performed with wild type *E.coli* K-12 (w3110) under anaerobic conditions with a mixture of naturally labeled and 10% ^{13}C U-labeled glucose. The model developed has been used to estimate metabolic fluxes. There was a good agreement between experimental and simulated data, as shown in Figure.
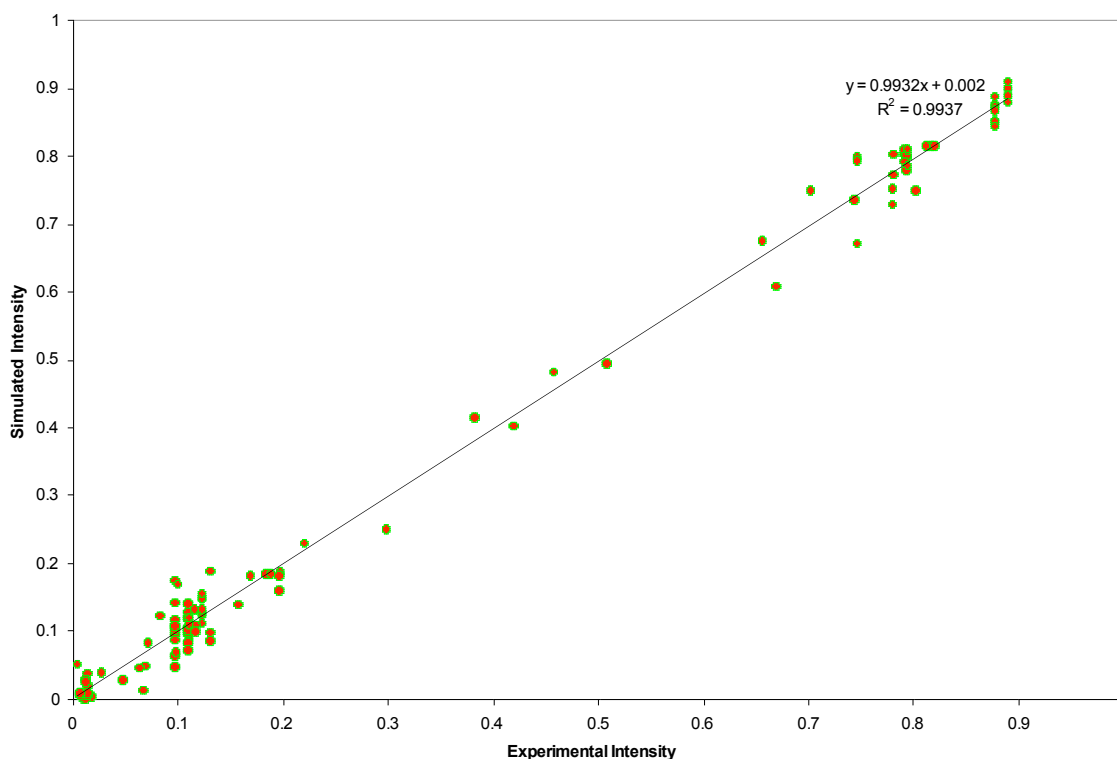

Figure 2 :Comparison of experimental and simulated NMR intensities of amino acids

d.   Theory behind the model, expressed in non-mathematical terms;

The model consists of central carbon metabolism reactions and biosynthetic pathways leading to biomass formation. It is assumed that model accounts for all reactions active in the central carbon metabolism under the experimental conditions and can be used to simulate the carbon labeling experiment.

e. Mathematics to be used, including formulas and calculation methods;

The relationship between labeling data and metabolic fluxes is non-linear. Thus, fluxes are estimated in iterative method. The intracellular fluxes are assumed and   carbon labeling

experiment is simulated using (cumulative) isotopomer balances for the fluxes. The set of guessed fluxes which give the best estimate of simulated intensities are taken as true fluxes. NMR2Flux uses optimization algorithm (Simulated annealing and Powell method) to minimize the chi-square, the deviation between the simulated and experimental data. Finally, Monte Carlo simulation is done to obtain statistical distribution of fluxes.

f. **Whether or not the theory and mathematical algorithms were peer reviewed, and, if so, include a summary of theoretical strengths and weaknesses;**

The tool NMR2Flux has been used to estimate metabolic fluxes in the plant system and the algorithm has been published in the *Plant Physiology* (Sriram et al, 2004).

NMR2Flux uses innovative techniques like "boolean function mapping" which makes the algorithm faster. It uses cumulative isotopomer balances which make it accurate. The algorithm is computationally intensive and it is challenging to develop metabolic network model that gives good fit of simulated and experimental labeling intensities. *The algorithm cannot evaluate fluxes without extracellular measurements and it true for all ^{13}C based MFA estimation algorithms.*

g. **Hardware requirements; and**

The computer program, NMR2Flux, is implemented in the programming language C, and can be run on computer with the Red Hat Linux operating system with any Pentium/Celeron family or compatible   processor.

References:
1. I.M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I.T. Paulsen, M. Peralta-Gil and P.D. Karp. EcoCyc: A comprehensive database resource for *Escherichia coli*, *Nucleic Acids Research* 33:D334-7 2005.

2. Neidhardt, F.C., Curtiss, R., Ingraham, J.L., Liu, E.C.C., Low, K.B., Magasamik, B., Rezinkoff, W.G., Riley, M., Schaechter, M., Umbrager, H.E., (Eds) (1996). "*Escherichia coli* and *Salmonella typhimurium,*" 2nd ed., American Society for Microbiology, Washington.

3. Sriram G, Fulton DB, Iyer VV, Peterson JM, Zhou R, Westgate ME, Spalding MH, Shanks JV. Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional (13)C labeling, two-dimensional [(13)C, (1)H] nuclear magnetic resonance, and comprehensive isotopomer balancing. *Plant Physiol* 2004; 136**:**3043–57

4. Schmidt et al., 1997 K. Schmidt, M. Carlsen, J. Nielsen and J. Villadsen, Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices, *Biotechnol. Bioeng.* 55 (1997), pp. 831–840.