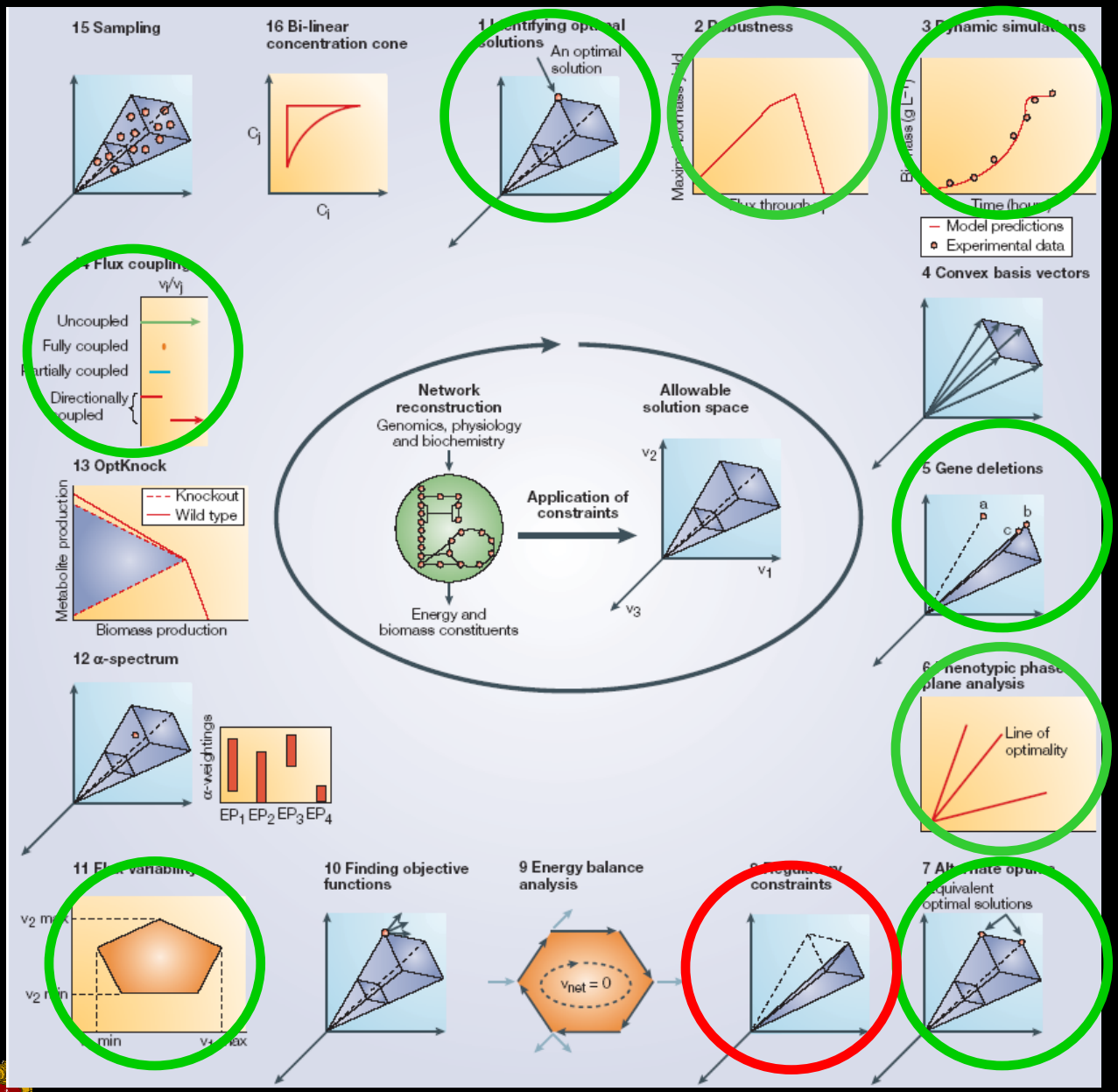
Constraint-Based Workshops

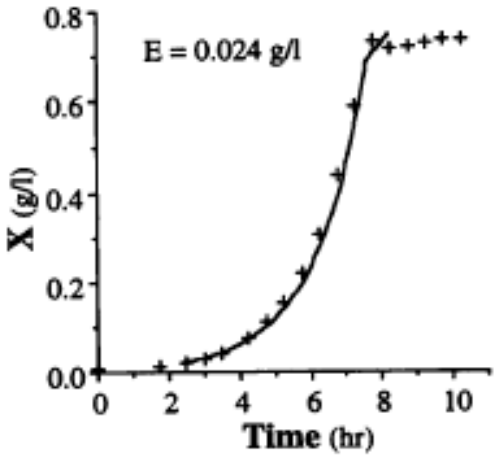
9. Transcriptional Regulatory Constraints
February 28th, 2008

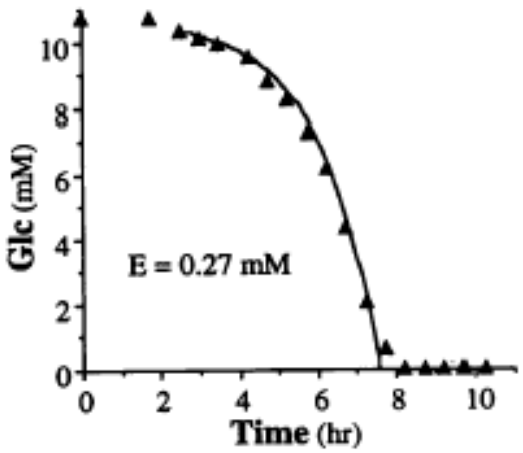
Constraint-Based Methods



- Optimal Solutions
  1. FBA
  2. Flux Variability
- Flux Dependencies
  1. Robustness
  2. Phase Planes
  3. Flux Coupling
- All Allowable Solutions
  1. Extreme Pathways
  2. Elementary Modes
  3. Sampling
- Altering Phenotypes
  1. Genetic Mutations
  2. Strain Design
- Application of Additional Constraints
  1. Regulation
  2. Energy Balance

Price, Reed, and Palsson
Nat. Reviews Microbiol.
2004





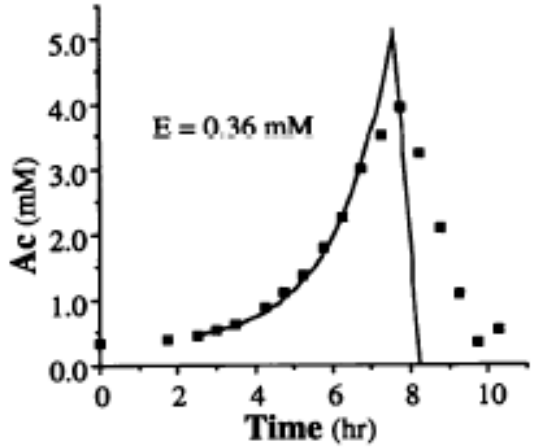


FIG. 7. Analysis of aerobic batch culture showing time profiles of cell density, glucose concentration, and acetate (Ac) concentration. The culture was not limited for minerals. The solid lines are the flux balance model predictions of the time profiles for the culture. E, average deviation between predictions of the model and experimental measurements.

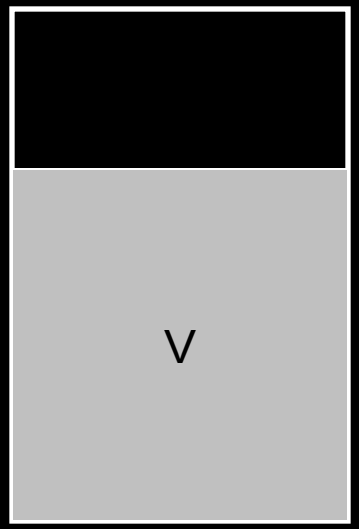
<u>In Batch Culture:</u>

V is Constant

Biomass (X) increases over time

Carbon Source (eg. glucose) decreases over time

*Acetate is secreted if not enough Oxygen is supplied



Varma and Palsson, App. Environ. Microbiol. 60(10): 3723-3731 (1994)

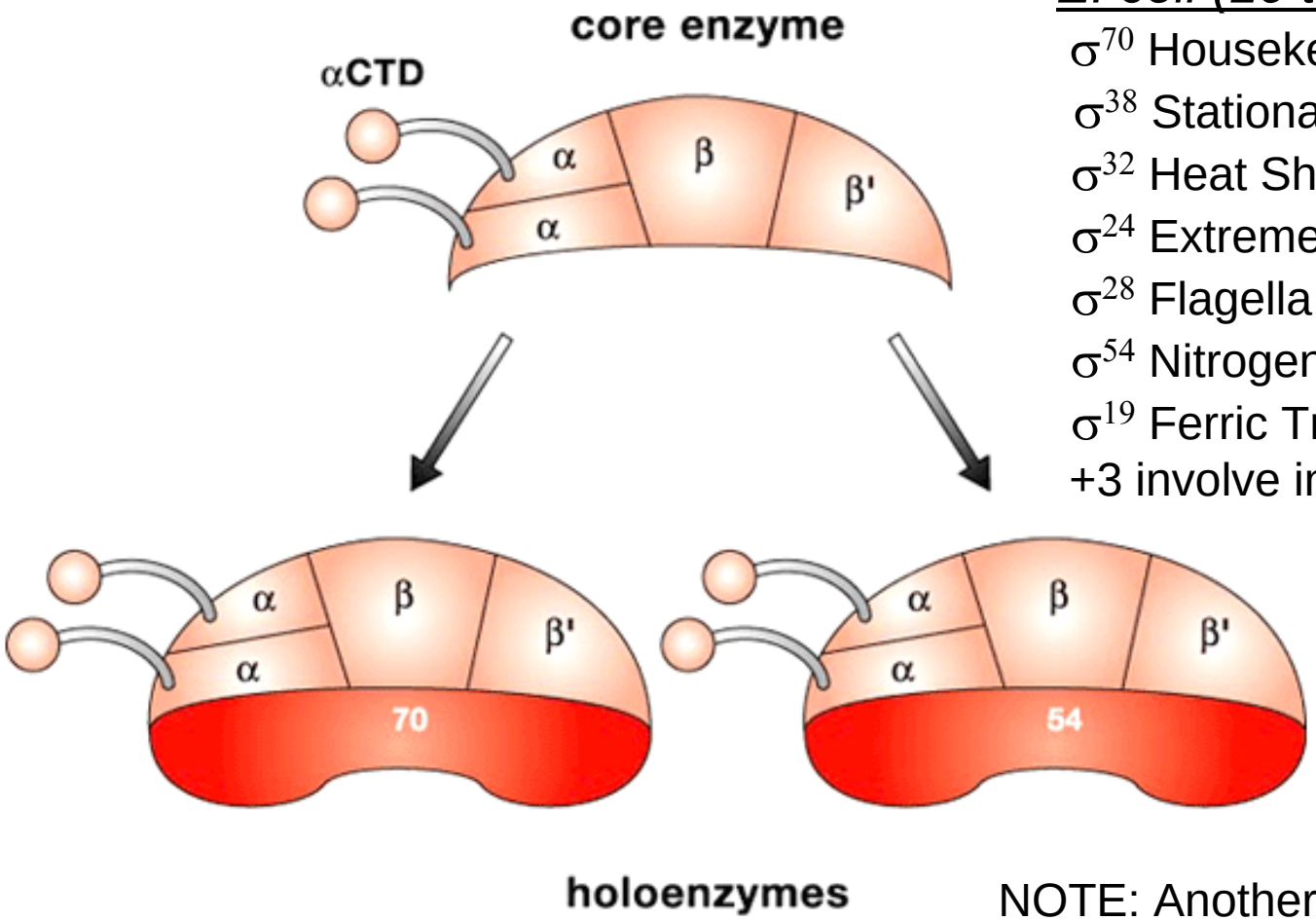
How Can You Regulate Protein Levels?

- Adjust the rate of mRNA synthesis.
  - TRANSCRIPTIONAL REGULATION
- Adjust the rate of mRNA translation.
  - TRANSLATIONAL REGULATION
- Adjust the activity of proteins through post-translational modifications.
  - FOLDING, CLEAVAGE, SPLICING (INTEINS), CHEMICAL MODIFICATION

Transcriptional Regulation in Prokaryotes

- RNA polymerase
- Modes of regulation
  - Regulated recruitment
  - Activation of RNA polymerase
  - Promoter activation
  - Antitermination and attenuation

E. coli RNA Polymerase



E. coli (10 total) Sigma Subunits

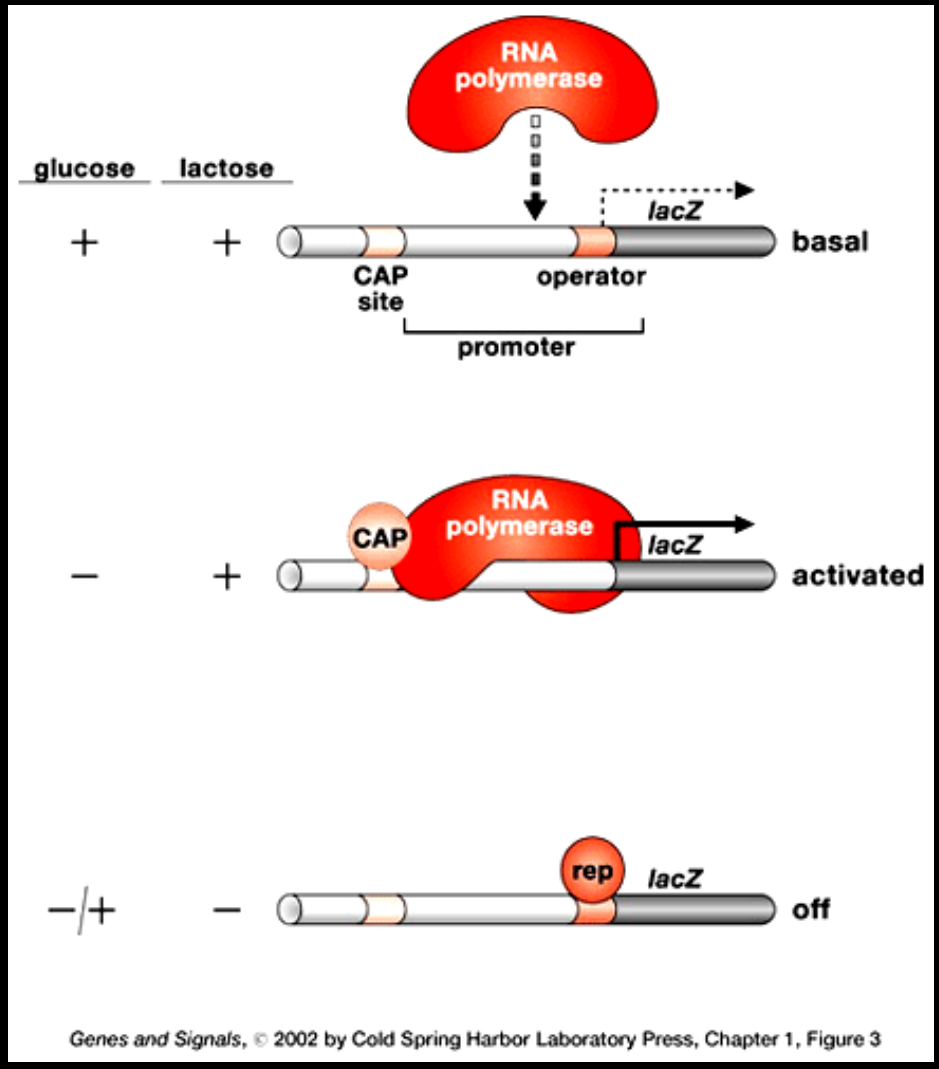
- σ^{70} Housekeeping
- σ^{38} Stationary Phase
- σ^{32} Heat Shock
- σ^{24} Extreme Heat Shock
- σ^{28} Flagella & Chemotaxis
- σ^{54} Nitrogen Regulation
- σ^{19} Ferric Transport Genes
- +3 involve in phage genes

Ptashne, Genes & Signals, Figure 1.1

NOTE: Another subunit w is associated with RNA polymerase, if deleted growth is reduced.

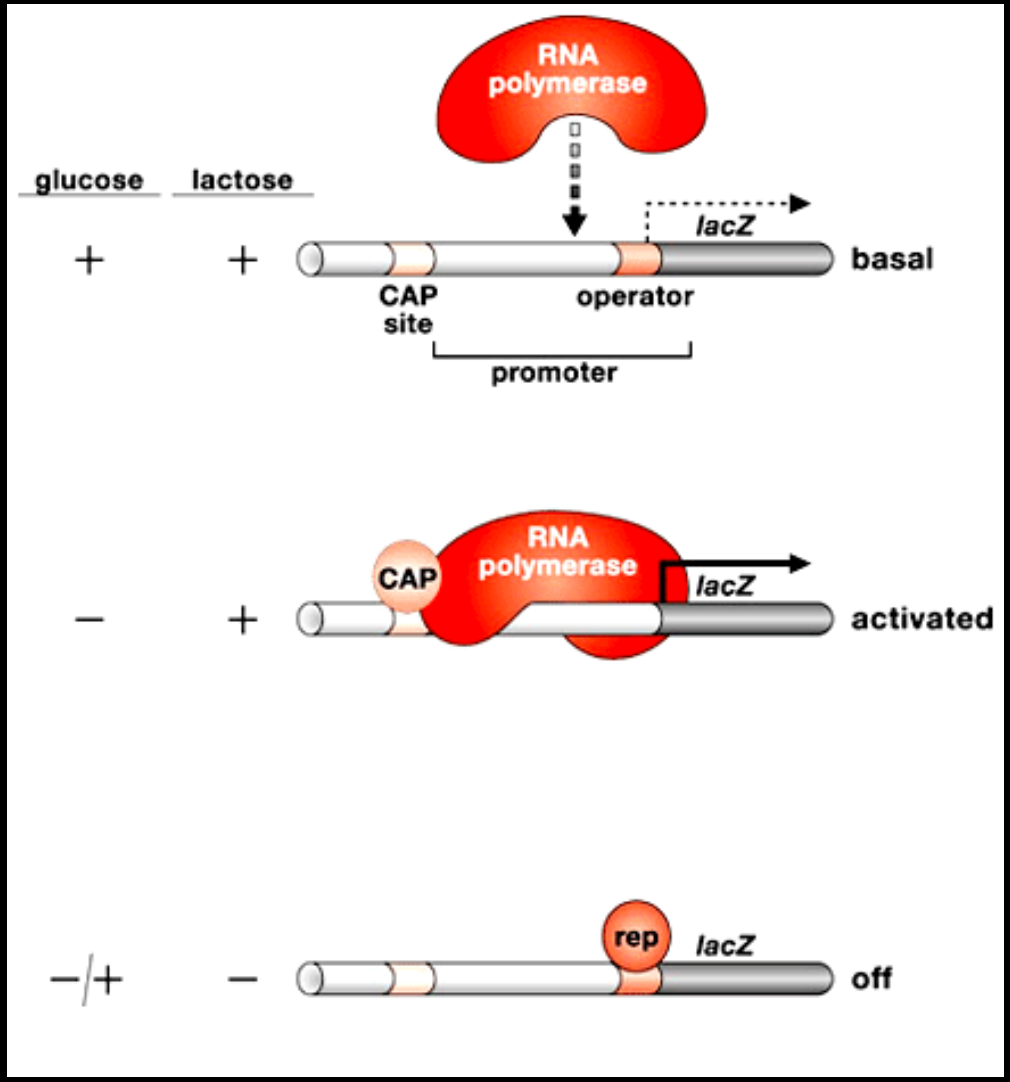
Mode 1: Regulated Recruitment

- Transcription factors (TFs) bind to specific binding sites in the promoter region of a gene
- After binding to DNA TFs either enhance (activator) or disrupt (repressor) RNA polymerase binding to DNA



Genes and Signals, © 2002 by Cold Spring Harbor Laboratory Press, Chapter 1, Figure 3

Three States of the *lac* Genes



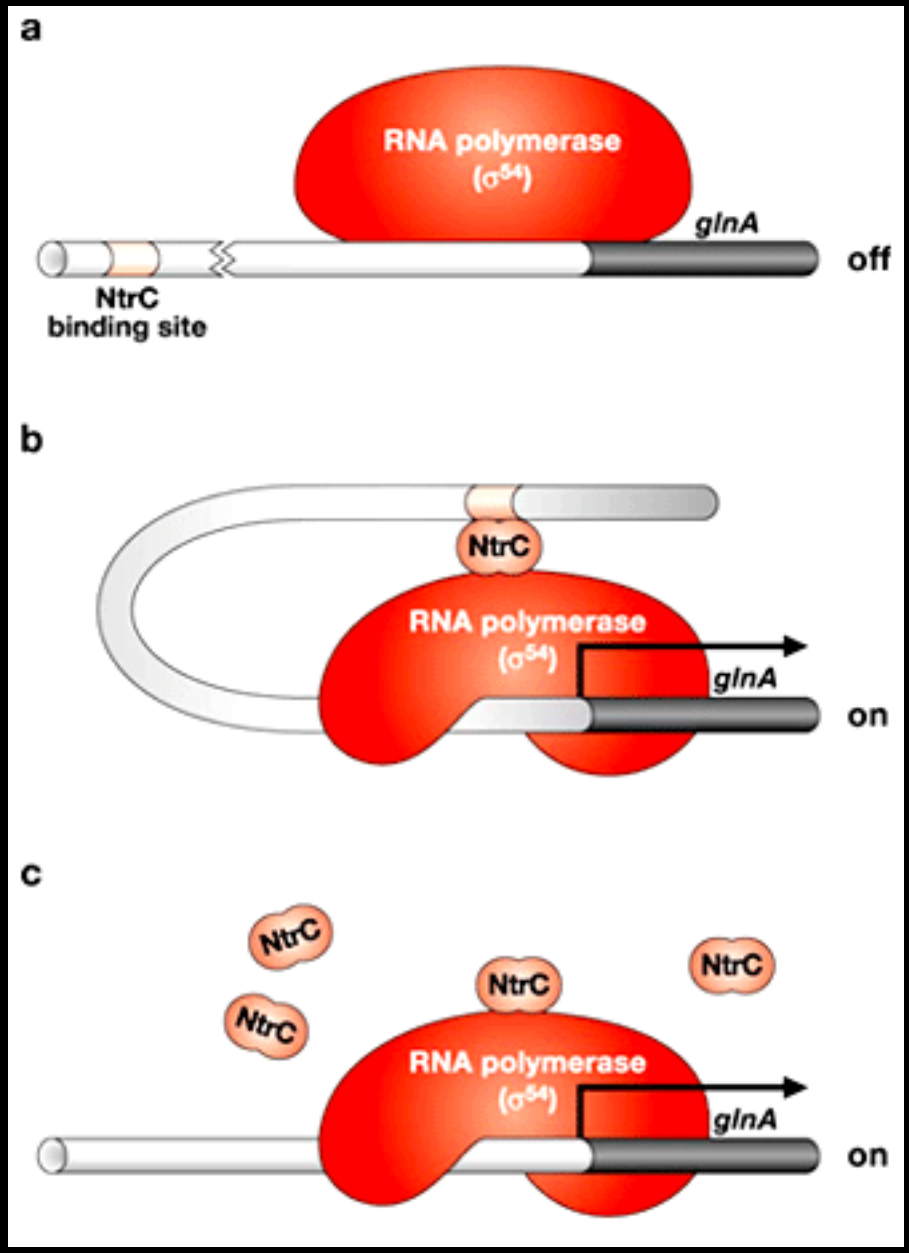
Ptashne, Genes & Signals, Figure 1.3

Mode 2: Activation of RNA Polymerase

NtrC acts on pre-bound polymerase.

Subsequent ATP hydrolysis induces conformational change.

NtrC activation is not dependent on DNA binding domain

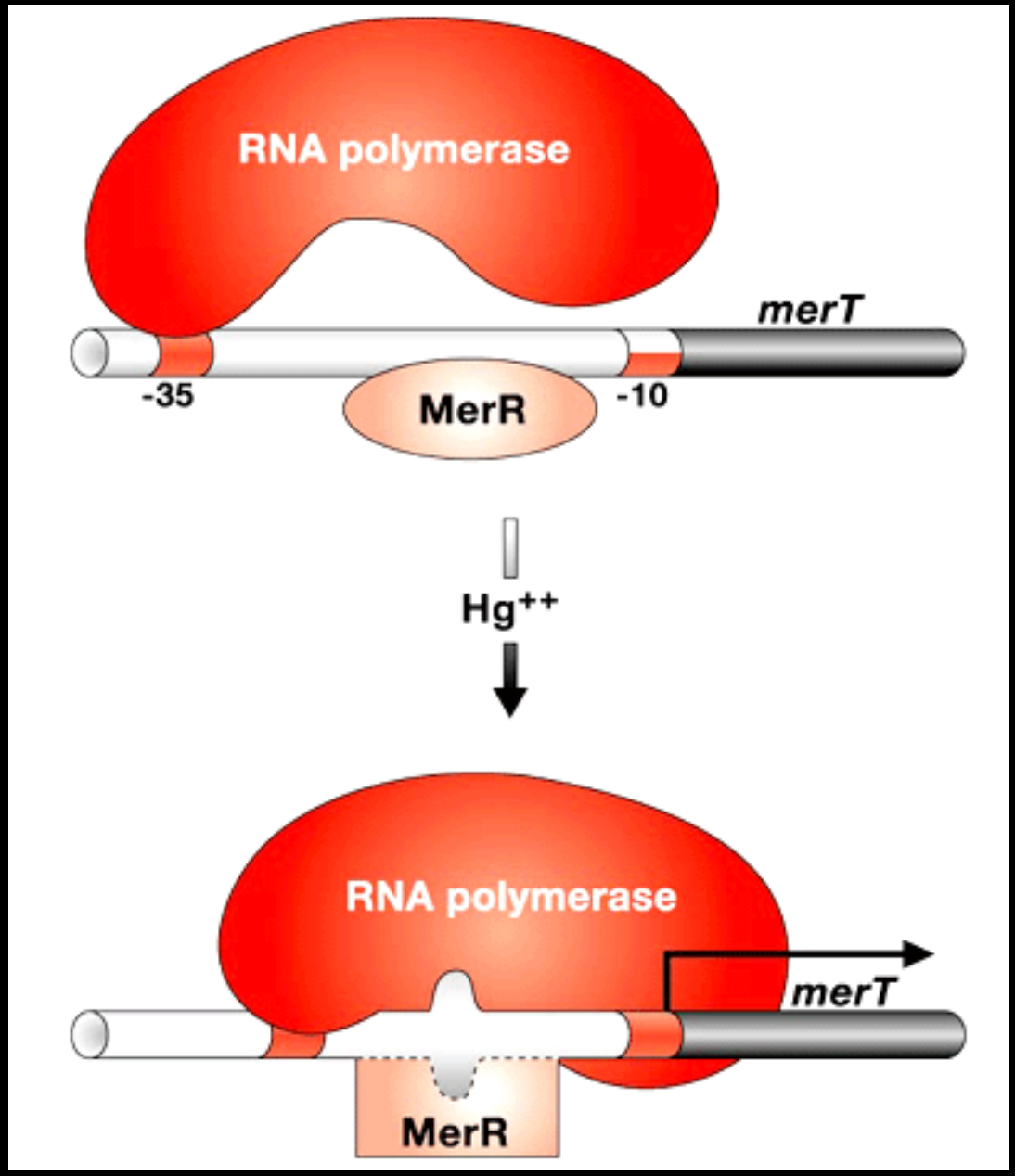


Ptashne, Genes & Signals, Figure 1.19

Mode 3: Promoter Activation

Normally, the -10 and -35 regions are separated by 17 nucleotides.

The merT promoter is separated by 19 nucleotides.

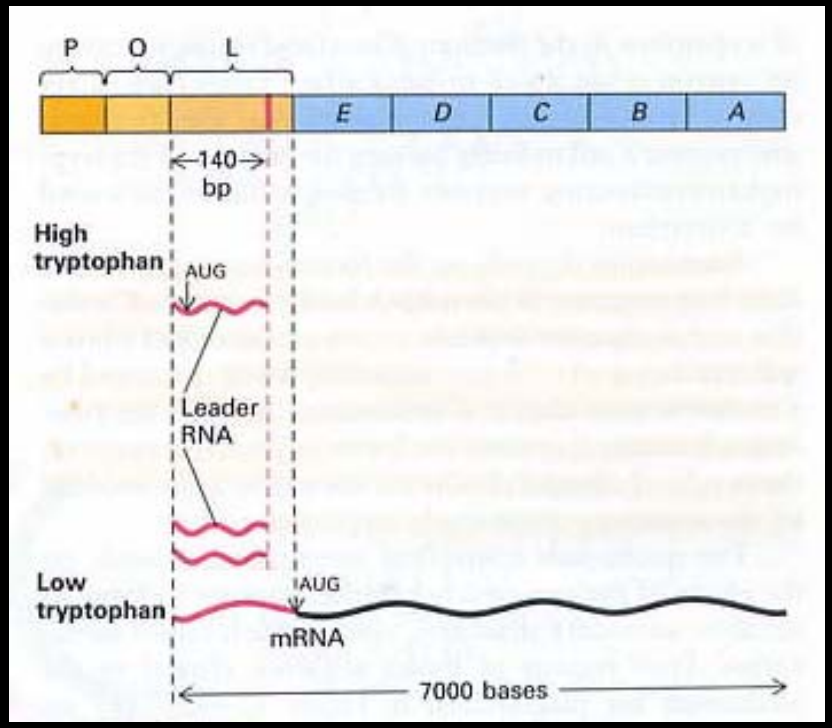


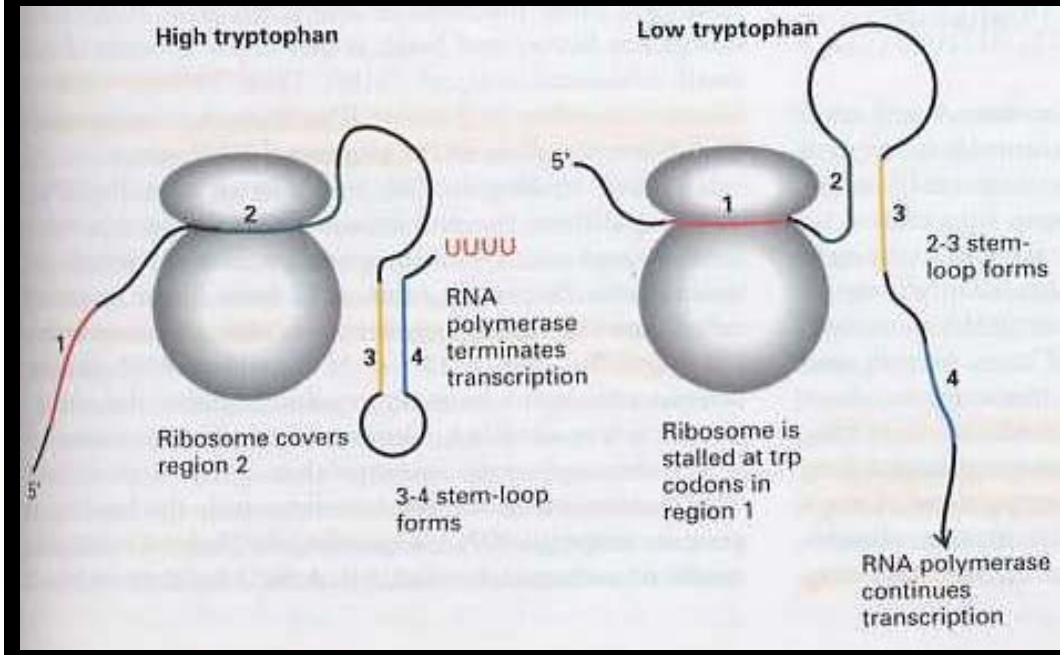
Ptashne, Genes & Signals, Figure 1.20

Mode 4: Regulate Termination by Attenuation

- Transcription and translation are coupled in prokaryotes and the rate of translation can affect termination.

Tryptophan Leader Sequence (20 AA): MKAIFVLKGWWRTS





Lodish, Mol. Cell Biology, Figure 12.3 and 12.4

Histidine Leader: 7 His AA in a row
Phenylalanine Leader: 2 sets of 3 Phe AA in a row

Mode 5: Regulate Termination by Antitermination

- RNA polymerase can stall during translation of genes
- Protein binding at various sites can allow the RNA polymerase to continue translation and prevent termination.
- Proteins “modify” the RNA polymerase and allow for it to resist terminators.
- Examples include: phage λ (nusABEG, N and Q proteins)

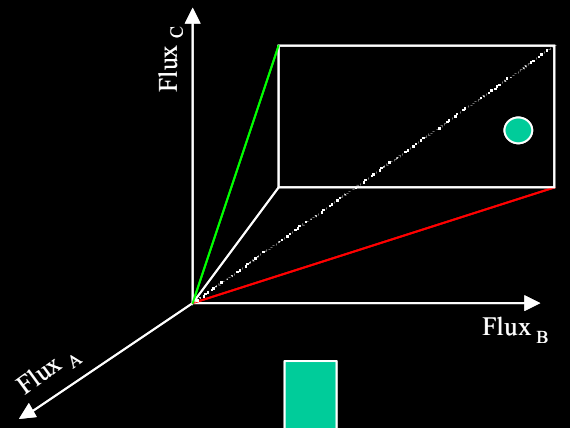
How Can We Incorporate Regulatory Constraints into Constraint-Based Models?

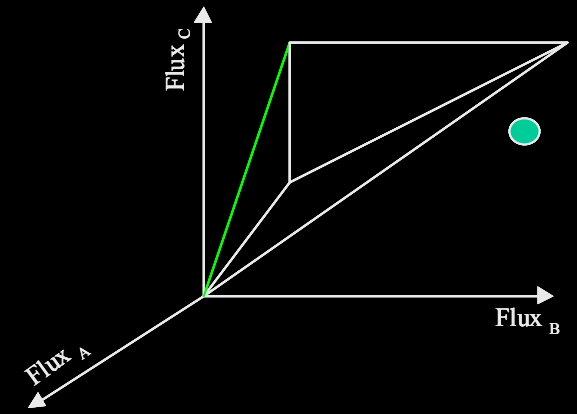
- We need to be able model how 100s of genes are regulated.
- We need to be able to account for how gene expression affects fluxes in constraint-based models.

Boolean Formalism

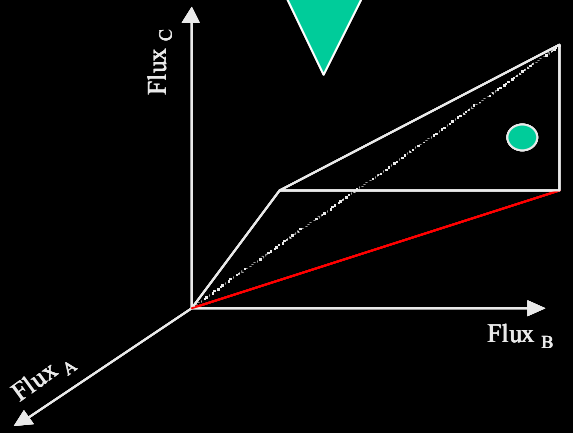
- Assume that genes are in one of two states:
  - Gene is Expressed
  - Gene is Not Expressed
- If genes needed for a reaction are not expressed then the reaction can not occur.
- With FBA and other methods, we assume that all enzymes & reactions are available for a cell to use.
- NOTE: Regulation of essential genes is not amenable to a Boolean approach

Regulation: Shaping the Space





No ‘red’: solution not achievable



No ‘green’: solution achievable but solution space smaller

1. Temporary, adjustable constraints
2. Flux(es) constrained to zero
3. Extreme vector(s) removed
4. Dimension or volume of space is reduced

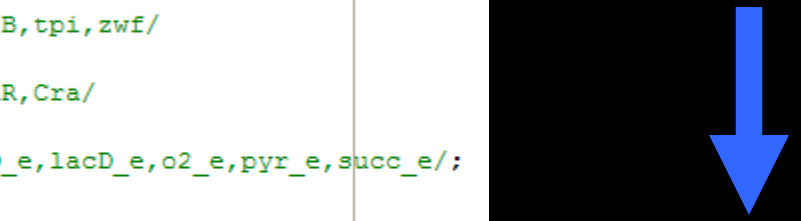
Covert, Schilling and Palsson, JTB 2001

Regulatory Network File

- Medium Components (uptake & secretion rates)
- Transcription Factor Activity Status
- Metabolic Gene Expression Status
- Reaction Status (if all genes needed are expressed reaction is allowed)

```
sets genes /aceA,aceB,ackA,acnA,acnB,adhE,adk,aceE,aceF
atpA,atpB,atpC,atpD,atpE,atpF,atpG,atpH,atpI
crr,cydA,cydB,dctA,dcuC,eno,fba,fbp,focA,frdA,frdB,frdC,frdD
fumA,fumB,fumC,gapA,gltA,gnd,icdA,ldhA,lpdA,maeB,mdh
nuoA,nuoB,nuoE,nuoF,nuoG,nuoH,nuoI,nuoJ,nuoK,nuoL,nuoM,nuoN
pckA,pfkA,pfkB,pflA,pflB,pflC,pflD,pgi,pgk,pgl,pgm,pitA,pitB
pntA,pntB,ppc,ppsA,pta,ptsG,ptsH,ptsI,pykA,pykF,rpe,rpi
sdhA,sdhB,sdhC,sdhD,sfcA,sucA,sucB,sucC,sucD,talA,tktA,tktB,tpi,zwf/

transcriptionfactors /Fnr,ArcA,DcuS,DcuR,FadR,IclR,Mlc,PdhR,Cra/

components(i) /ac_e,akg_e,etoh_e,formate_e,fumarate_e,glcD_e,lacD_e,o2_e,pyr_e,succ_e/;

parameters reactionstatus(j)
           genestatus(genes)
           TFstatus(transcriptionfactors)
           medium(components)
           temp1
           temp2;

genestatus(genes)=1;
reactionstatus(j)=1;
```

INPUT:
v from FBA



OUTPUT:
ReactionStatus
(0 or 1)

```
**********************************************************************************
*****Determine What is Present in the Medium (Substrate & By-Produces**********
**********************************************************************************
if((LowerLimits('EX_ac_e')<0) or (v.l('EX_ac_e')>0), medium('ac_e')=1;
        else medium('ac_e')=0;);
if((LowerLimits('EX_akg_e')<0) or (v.l('EX_akg_e')>0), medium('akg_e')=1;
        else medium('akg_e')=0;);
if((LowerLimits('EX_etoh_e')<0) or (v.l('EX_
        else medium('etoh_e')=0;);
if((LowerLimits('EX_for_e')<0) or (v.l('EX_
        else medium('formate_e')=0;);
if((LowerLimits('EX_fum_e')<0) or (v.l('EX_
        else medium('fumarate_e')=0;);
if((LowerLimits('EX_glc_e')<0) or (v.l('EX_
        else medium('glcD_e')=0;);
if((LowerLimits('EX_lacD_e')<0) or (v.l('EX_
        else medium('lacD_e')=0;);
if((LowerLimits('EX_o2_e')<0) or (v.l('EX_o
        else medium('o2_e')=0;);
if((LowerLimits('EX_pyr_e')<0) or (v.l('EX_
        else medium('pyr_e')=0;);
if((LowerLimits('EX_succ_e')<0) or (v.l('EX_succ_e')>0), medium('succ_e')=1;
        else medium('succ_e')=0;);

**********************************************************************************
*****Determine Transcription Factor Activities (0=Active, 1=Inactive)**********
**********************************************************************************
TFstatus('Fnr')=1-medium('o2_e');
TFstatus('ArcA')=1-medium('o2_e');
TFstatus('DcuS')=medium('succ_e');
TFstatus('DcuR')=TFstatus('DcuS');
TFstatus('FadR')=max(medium('glcD_e'),1-medi
TFstatus('IclR')=TFstatus('FadR');
TFstatus('Mlc')=1-medium('glcD_e');
```

If the lower limit is negative OR if the FBA solution predicts secretion, then the metabolite would be present in the medium.

Medium values are 0 if absent & 1 if present

TF Status are Determined by Medium Components & Some Intracellular Fluxes

Translating Boolean Rules in GAMS:

****Here X,Y,Z are binary (0/1), and represent genes, TFs, media components**

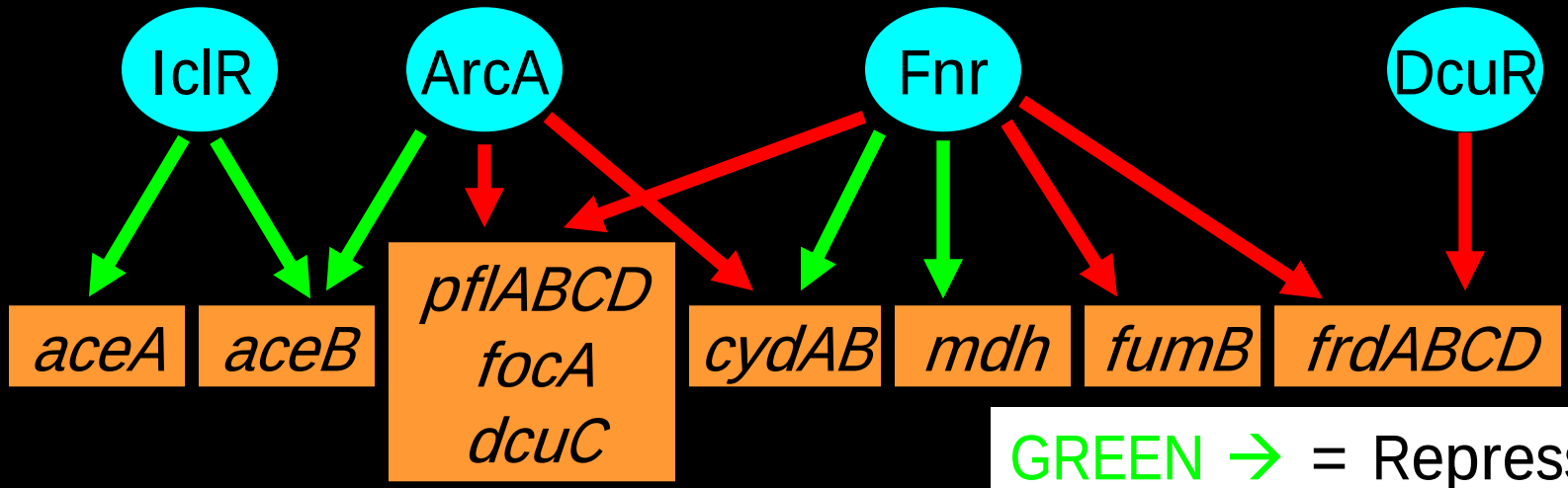
- Z active if X
- Z active if NOT X
- Z active if X AND Y
- Z active if X OR Y

- Z = X
- Z = 1-X
- Z = min(X,Y)
- Z = max(X,Y)

- Eg. Fnr is active when O_2 is NOT in the medium
- Eg. FadR is active Glucose is present OR acetate is absent from medium.

- Fnr=1-medium(O_2)
- FadR=max[medium(glc) ,1-medium(ac)]

TF-Target Regulatory Rules



GREEN → = Repression
RED → = Activation

```
**********************************************************
**********Determine Gene Expression States for Metabolic Gene
**********************************************************
genestatus('aceA')=1-TFstatus('IclR');
genestatus('aceB')=1-max(TFstatus('IclR'),TFstatus('ArcA'));
genestatus('adhE')=max(1-medium('o2_e'),1-min(medium('o2_e'),
genestatus('frdA')=max(TFstatus('Fnr'),TFstatus('DcuR'));
genestatus('frdB')=max(TFstatus('Fnr'),TFstatus('DcuR'));
genestatus('frdC')=max(TFstatus('Fnr'),TFstatus('DcuR'));
genestatus('frdD')=max(TFstatus('Fnr'),TFstatus('DcuR'));
genestatus('cydA')=max(1-TFstatus('Fnr'),TFstatus('ArcA'));
genestatus('cydB')=max(1-TFstatus('Fnr'),TFstatus('ArcA'));
genestatus('fumB')=TFstatus('Fnr');
genestatus('pflA')=max(TFstatus('Fnr'),TFstatus('ArcA'));
genestatus('pflB')=max(TFstatus('Fnr'),TFstatus('ArcA'));
genestatus('pflC')=max(TFstatus('Fnr'),TFstatus('ArcA'));
genestatus('pflD')=max(TFstatus('Fnr'),TFstatus('ArcA'));
genestatus('focA')=max(TFstatus('Fnr'),TFstatus('ArcA'));
genestatus('dcuC')=max(TFstatus('Fnr'),TFstatus('ArcA'));
genestatus('mdh')=1-TFstatus('Fnr');
```

aceA is expressed if NOT IclR is active
→ IclR represses *aceA*
aceB is expressed if NOT (IclR OR ArcA are active)
→ Either IclR or ArcA represses *aceB*
frdABCD are expressed if Fnr OR DcuR are active
→ Either Fnr or DcuR activates *frdABCD*
fumB is expressed if Fnr is active
→ Fnr activates *fumB*
cydAB are expressed if ArcA or NOT Fnr are active
→ ArcA activates and Fnr represses *cydAB*
mdh is expressed if NOT Fnr is active
→ Fnr represses *mdh*
pflABCD, focA, dcuC are expressed if Fnr OR ArcA are active → Either Fnr or ArcA activates these genes

Gene to Reaction Associations

- If necessary genes are expressed (ie. genestatus =1) then the reaction status is 1, meaning the reaction can happen.
- Otherwise, if the necessary genes aren't all expressed then the reaction status is 0, meaning reaction can not occur.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%Change Reaction Status to 0 When Genes Needed Aren't Expressed%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
reactionstatus('ICL')=genestatus('aceA');
reactionstatus('MALS')=genestatus('aceB');
reactionstatus('ACKr')=genestatus('ackA');
reactionstatus('ADHEr')=genestatus('adhE');
reactionstatus('ADK1')=genestatus('adk');
reactionstatus('FUMt2_2')=genestatus('dctA');
reactionstatus('SUCCt2_2')=genestatus('dctA');
reactionstatus('SUCCt2b')=genestatus('dcuC');
```

Only One Gene

```
*These reactions have isozymes so if at least one is present the reaction can occur
reactionstatus('ACONT')=max(genestatus('acnA'),genestatus('acnB'));
reactionstatus('FUM')=max(genestatus('fumA'),genestatus('fumB'),genestatus('fumC'));
reactionstatus('PFK')=max(genestatus('pfkA'),genestatus('pfkB'));
reactionstatus('PFL')=max(min(genestatus('pflA'),genestatus('pflB'))
reactionstatus('PYK')=max(genestatus('pykA'),genestatus('pykF'));
reactionstatus('TKT1')=max(genestatus('tktA'),genestatus('tktB'));
reactionstatus('TKT2')=max(genestatus('tktA'),genestatus('tktB'));

*These reactions are carried out by multiple gene products, all have t
reactionstatus('SUCOAS')=min(genestatus('sucC'),genestatus('sucD'));
reactionstatus('AKGDH')=min(genestatus('sucA'),genestatus('sucB'),genes
reactionstatus('PDH')=min(genestatus('aceE'),genestatus('aceF'),genesta
reactionstatus('PIt')=min(genestatus('pitA'),genestatus('pitB'));
reactionstatus('NADTRHD')=min(genestatus('pntA'),genestatus('pntB'));
```

Multiple Isozymes: 'pflD'))); "max"

o occur

Multiple Subunits "min"

rFBA Main File

```
**THIS CODE WAS WRITTEN FOR THE CONSTRAINT-BASED MODELING WORKSHOP J.REE

*Read in the appropriate S matrix
$include CoreTextbookModel.gms

*Place limits on the exchange fluxes based on the minimal media
*for a negative flux through the exchange reactions implies that
*the metabolites are being taken up or consumed by the cell.
*By default the upperlimits of the exchange fluxes are all set to
*the Vmax, indicating that the cell can secrete any of the extracellular
*metabolites
UpperLimits(j)=Vmax;
*CARBON SOURCE: select upper and lower limits for exchange flux
LowerLimits('EX_glc_e')=-5;
*allow co2,pi,o2,h,h2o to be taken up by the cell
LowerLimits('EX_co2_e')=-Vmax;
LowerLimits('EX_h2o_e')=-Vmax;
LowerLimits('EX_h_e')=-Vmax;
LowerLimits('EX_o2_e')=-Vmax;
LowerLimits('EX_pi_e')=-Vmax;

Parameter  c(j) used to define the obje
rFBA_fluxes(j) used to store regulated
FBA_fluxes(j)  used to store unregulate

c('Biomass')=1;

Variables
v(j) flux values through reaction in ne
Obj  this is the value of the objective

Equations
massbalance(i) mass balance equations f
calcobj calculates the dot product of t
massbalance(i).. sum( j,S(i,j)*v(j) )=e
calcobj.. Obj=e=sum( j,c(j)*v(j) );
```

Read in S matrix

Select Carbon Source

Aerobic / Anaerobic Constraint

```
*****************************************************************************
***************Solve the regular FBA problem First*************************
*****************************************************************************
v.lo(j)=LowerLimits(j);
v.up(j)=UpperLimits(j);
solve FBA using lp maximizing Obj;
FBA_fluxes(j)=v.l(j);

*****************************************************************************
*Evaluate TF Activity & Gene Expression States to Constrain Certain Fluxes**
*to Equal 0. Use FBA to again to Solve for the Fluxes**********************
*****************************************************************************
$include CoreRegulation.gms
v.lo(j)=reactionstatus(j)*LowerLimits(j);
v.up(j)=reactionstatus(j)*UpperLimits(j);
solve FBA using lp maximizing Obj
rFBA_fluxes(j)=v.l(j);

display TFstatus,genestatus;
```

Solve Unregulated Model

Evaluate TF,Gene,Reaction States

Constrain Fluxes & Solve

Output Text File

"Model status: "	1.00		
"Solver status: "	1.00		
"Rxn Abbr"	"FBA Solution"	"Reaction Status (0=Off, 1=On)"	"rFBA Solution"
"EX_ac_e"	0.0000	1	1.5639
"EX_akg_e"	0.0000	1	0.0000
"EX_co2_e"	9.1599	1	7.7815
"EX_etoh_e"	0.0000	1	0.0000
"EX_for_e"	0.0000	1	0.0000
"EX_fum_e"	0.0000	1	0.0000
"EX_glc_e"	-5.0000	1	-5.0000
"EX_h_e"	5.1807	1	6.3098
"EX_h2o_e"	10.0569	1	8.6032
"EX_lacD_e"	0.0000	1	0.0000
"EX_o2_e"	-8.5938	1	-7.2629
"EX_pi_e"	-1.8015	1	-1.6503
"EX_pyr_e"	0.0000	1	0.0000
"EX_succ_e"	0.0000	1	0.0000
"ACKr"	0.0000	1	-1.5639
"ACONT"	1.9175	1	0.4840
"ACt2r"	0.0000	1	-1.5639
"ADHEr"	0.0000	1	0.0000
"ADK1"	0.0000	1	0.0000
"AKGt2r"			
"ATPM"	0.000		
"ATPS4r"			.3393
"Biomass"	0.4897	1	0.4486
"CO2t"	-9.1599		-7.7815
"CS"	1.9175	1	
"CYTBD"	17.1		14.5257
"DLACt2"	0.0		0.0000

μ(FBA)

μ(rFBA)

Reaction Status

"Transcription Factor Status (0=Inctive, 1=Active)"	
"Fnr"	0
"ArcA"	0
"DcuS"	0
"DcuR"	0
"FadR"	1
"IclR"	1
"Mlc"	0
"PdhR"	1
"Cra"	0
"Gene Expression Status (0=Off, 1=On)"	
"succ"	1
"ENO"	1
"FBA"	1
"FBP"	1
"GND"	1
"MDH"	1
"PGI"	1
"PGK"	1
"PGL"	1
"PGM"	1
"PPC"	1
"RPE"	1
"RPI"	1
"TALA"	1
"TPI"	1
"aceA"	0
"aceB"	0

TFs FadR, IclR, and PdhR are Active

Here *aceA* and *aceB* are not Expresses

Regulatory Model Calculations Part 1

- What are the growth rate predictions for glucose aerobic growth using FBA & rFBA (glucose uptake = 5)?
- What reactions are not available during glucose aerobic growth? What about during glucose anaerobic growth?
- What are the growth rate predictions if you remove PPC (phosphoenolpyruvate carboxylase: pep to oaa) for glucose aerobic growth?

Regulatory Model Calculations Part 1 Ans

- What are the growth rate predictions for glucose aerobic growth using FBA & rFBA?
  - ANS: FBA = 0.49 and rFBA = 0.45
- What reactions are not available during glucose aerobic growth? What about during glucose anaerobic growth?
  - Aerobic: FORt, FRD, ICL, MALS, PFL, PPS, AKGDH, SUCCt2b,
  - Anaerobic: ICL, MALS, MDH, PDH, PPS, AKGDH, SUCD1i, SUCD4,
- What are the growth rate predictions if you remove PPC (phosphoenolpyruvate carboxylase) for glucose aerobic growth?
  - ANS: FBA = 0.49 and rFBA = 0

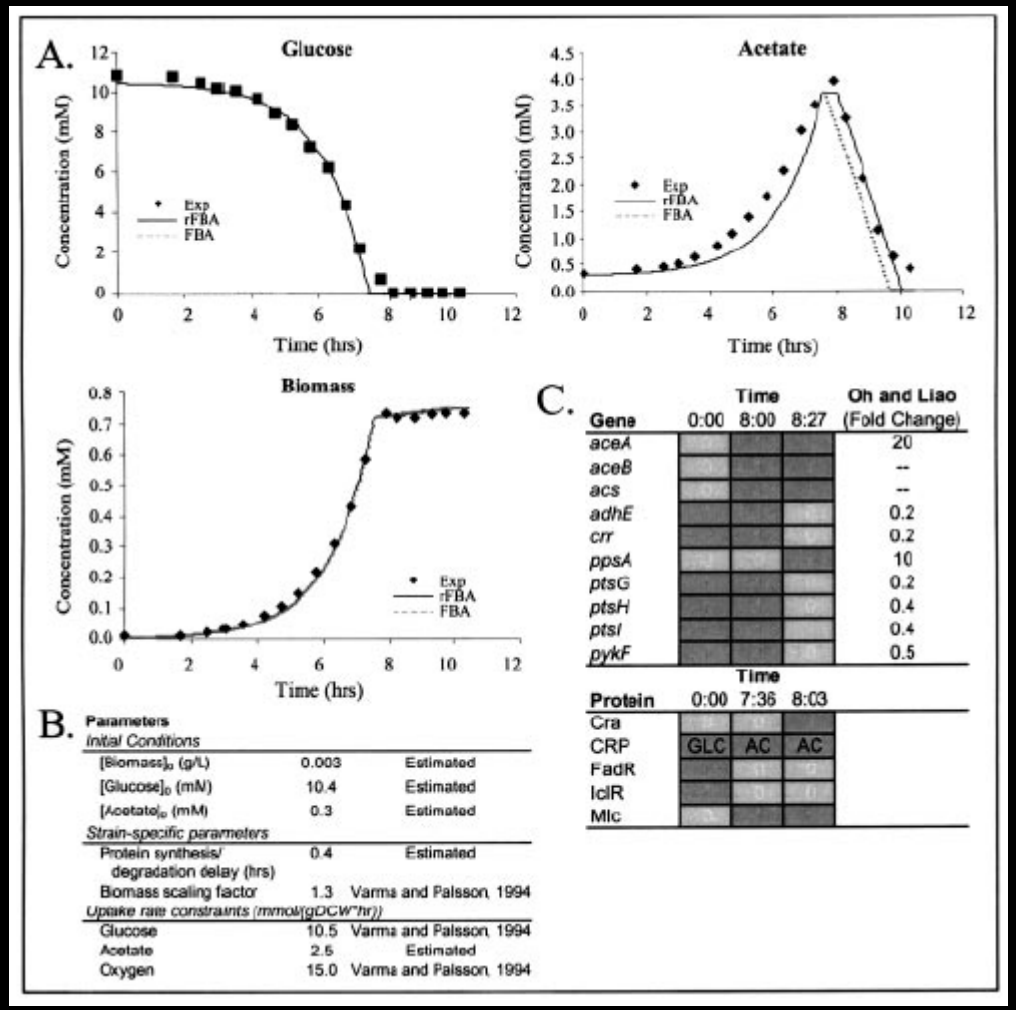
Regulatory Model Calculations Part 2

- What are the growth rate predictions if you remove PDH (pyruvate dehydrogenase: pyr to accoa) for glucose aerobic growth?
- What enzymes would you need to express if you wanted the cells to grow aerobically on glucose without PDH? (hint: look at the reduced costs for removed reactions)
- Is it possible for rFBA to find a non-zero growth rate if FBA finds a zero maximum growth rate?

Regulatory Model Calculations Part 2 Ans

- What are the growth rate predictions if you remove PDH for glucose aerobic growth?
  - ANS: FBA = 0.46 and rFBA = 0
- What enzymes would you need to express if you wanted the cells to grow cells to grow aerobically on glucose without PDH? (hint: look at the reduced costs for removed reactions)
  - ANS: PFL (and PDH but this was deleted)
- Is it possible for rFBA to find a non-zero growth rate if FBA finds a zero maximum growth rate?
  - ANS: No, rFBA should always find a smaller flux through your objective function than FBA. Regulated solution space is a subspace of unregulation solution space.

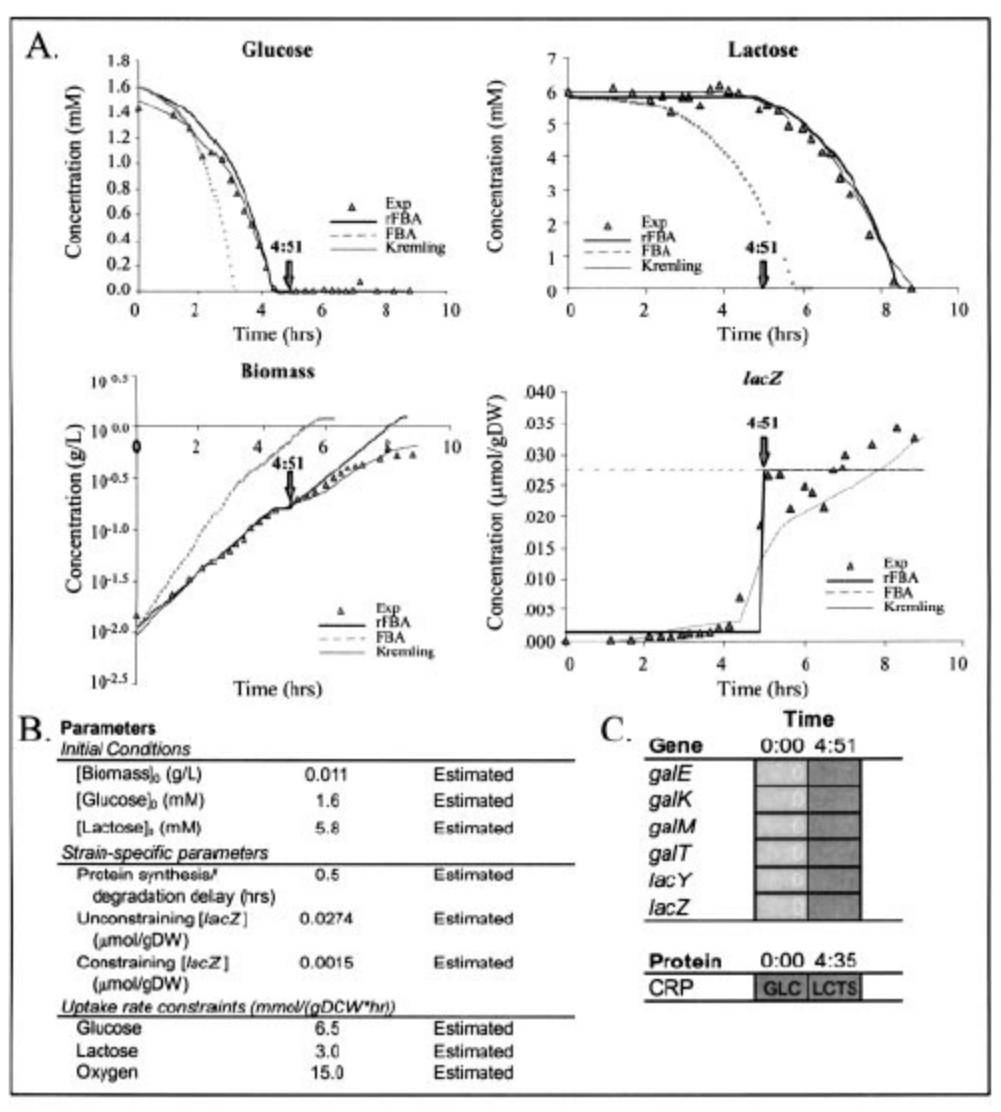
Glucose Aerobic Growth



- rFBA correctly predicts the time delay before acetate is reconsumed.
- rFBA can be used to predict gene expression measurements.

MW Covert & BO Palsson. J Biol Chem. 277(31): 28058-64 (2002).

Diauxic Growth: Glucose vs. Lactose



- FBA incorrectly predicts that glucose and lactate will be consumed simultaneously.
- rFBA can be used to predict gene expression measurements.

MW Covert & BO Palsson. J Biol Chem. 277(31): 28058-64 (2002).

rFBA Can Correct FBA Over Predictions

- rFBA correctly predicts that some gene deletions are lethal.
  - Pyruvate dehydrogenase for glucose or succinate aerobic growth.
  - Phosphoenolpyruvate carboxylase for glucose or glycerol aerobic growth
  - Can predict TF knockout effects
- Overall, rFBA correctly predicted 106/116 cases (different mutants under different conditions)

MW Covert & BO Palsson. J Biol Chem. 277(31): 28058-64 (2002).

TABLE I

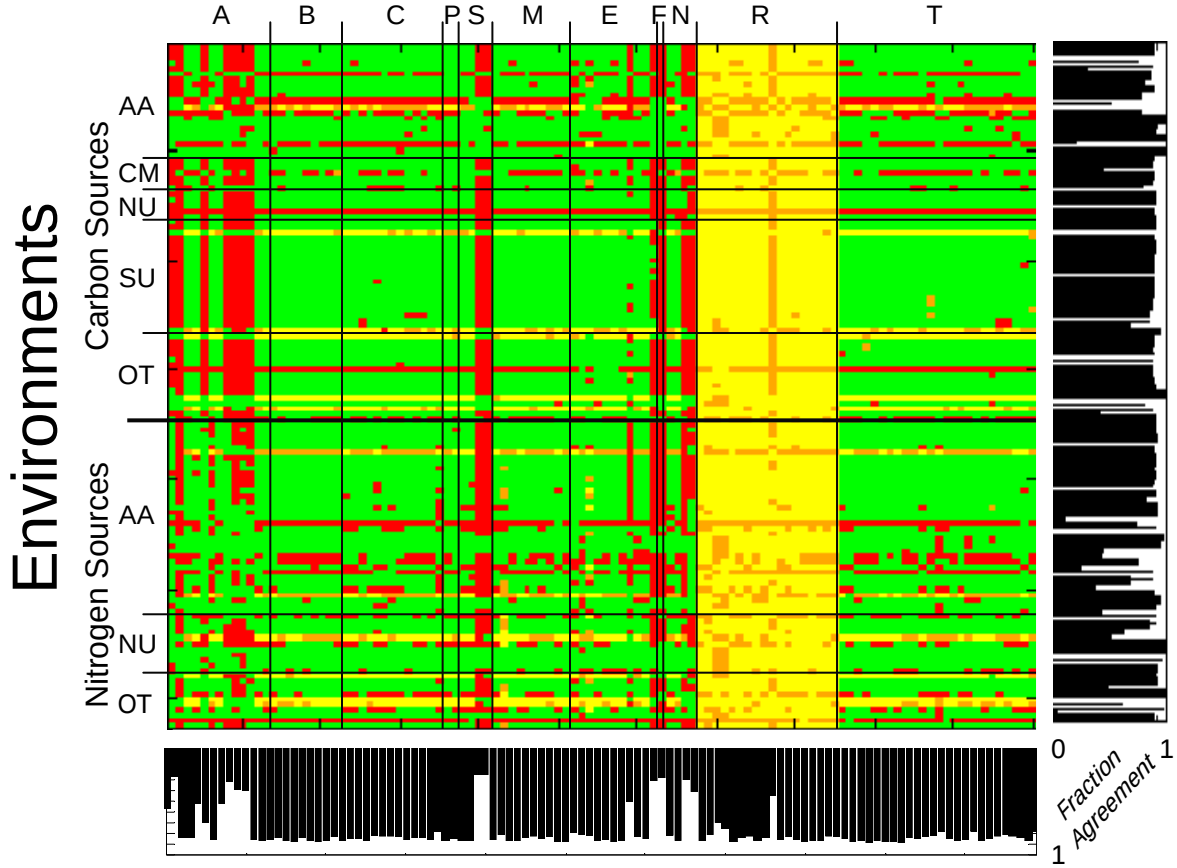
Comparison of experimental mutant studies with in silico predictions of the FBA and rFBA models under various conditions

Results are scored as + or − meaning growth or no growth determined from *in vivo*/FBA/rFBA data. An *N* blank indicates that data were not available for these conditions. Cases where rFBA makes a correct prediction either unpredicted or incorrectly predicted by FBA alone are denoted by a shaded box. In 106 of 116 cases, the rFBA prediction matches the experimentally observed behavior. gl, glycerol; suc, succinate; ($-O_2$), anaerobic conditions.

	glc	gl	suc	ac	rib	glc ($-O_2$)	Dual Substrates	Ref
aceA	+/+/+		+/+/+	-/-/-		+/+/+		21
aceB				-/-/-				22
aceEF	-/+/-		-/+/-	+/+/+		+/+/+	(glc-ac) +/+/+	23
ackA				+/+/+				24
ackA + pta + acs				-/-/-				24
acnA	+/+/+	+/+/+	+/+/+	+/+/+		+/+/+		22,25
*acnB**	+/+/+	+/+/+	+/+/+	-/+/+		+/+/+		25
pgk	-/-/-	-/-/-	-/-/-				(gl-suc) +/+/+	28
pgl	+/+/+							29
ppc	-/+/-	-/+/-	+/+/+				(gl-suc) +/+/+ (glc-suc) +/+/+	29,31

High-Throughput Phenotyping Data

Knockout Strains



HT Phenotyping/Model Comparison

exp/met/reg		Percent
Reg and Met models predict correctly		
+/+/+	6222	45.3%
-/-/-	2094	15.2%
Reg model predicts correctly		
-/+/-	657	4.8%
+/-/+	0	0.0%
+/n/+	1350	9.8%
-/n/-	505	3.7%
Met model only predicts correctly		
+/n/-	242	1.8%
-/n/+	153	1.1%
-/-/+	0	0.0%
+/+/-	257	1.9%
Neither model predicts correctly		
-/+/+	702	5.1%
+/-/-	1568	11.4%
Met	8968	65.2%
Reg	10828	78.7%
Total	13750	100.0%

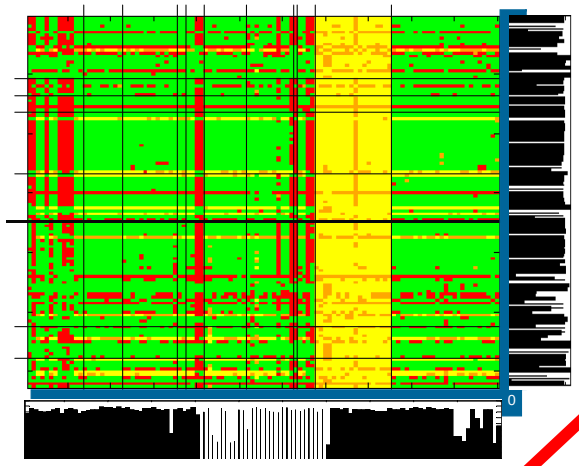
ASAP Database (Glasner et al., 2003)

- 125 Environments, 110 Strains
  - 13,750 Cases Overall

Total Consistency:
10,828 (78.7%)

Covert, et al. Nature, **429**: 92-96 (2004).

Network Elucidation



Carbon Sources	Nitrogen Sources
● Acetoacetic Acid	● Adenine
● Formic Acid	○ Alanine-Leucine
○ g-Amino Butyric Acid	Guanosine
● Glycine	◍ L-Lysine
○ L-Arginine	◍ L-Methionine
L-Glutamic Acid	◍ L-Phenylalanine
○ L-Ornithine	● N-Acetyl-D-Mannosamine
○ Putrescine	● Putrescine
◍ Thymidine	◍ Xanthine

Knockout Strains	
argB-	◍ *ilvY-*
argC-	◍ *metA-*
argD-	*pgi-*
argE-	◍ *pls-*
argG-	◍ *purD-*
◍ *glgA-*	◍ *purH-*
◍ *glgC-*	*tpiA-*
◍ *ilvD-*	

○ Missing regulatory interactions
◍ Missing metabolic enzymes
● Conflicts with subsequent observations

Reconciling the model with measured phenotypes leads to specific hypotheses

- Thymidine as a carbon source: not predicted by model
  - Thymine transporter is missing
- Arginine biosynthesis: essential to model, knockout strain grows
  - Reversibility of certain enzymes (YgjG, SpeC or SpeR, AldH or Prr) enables the *in silico* strain to produce arginine
- *ilvY-* strain: lethal in model due to essentiality of IlvC (15-fold shift)
  - Intervention of another factor
  - Sufficient basal level of IlvC activity
  - Uncharacterized IlvC isozyme

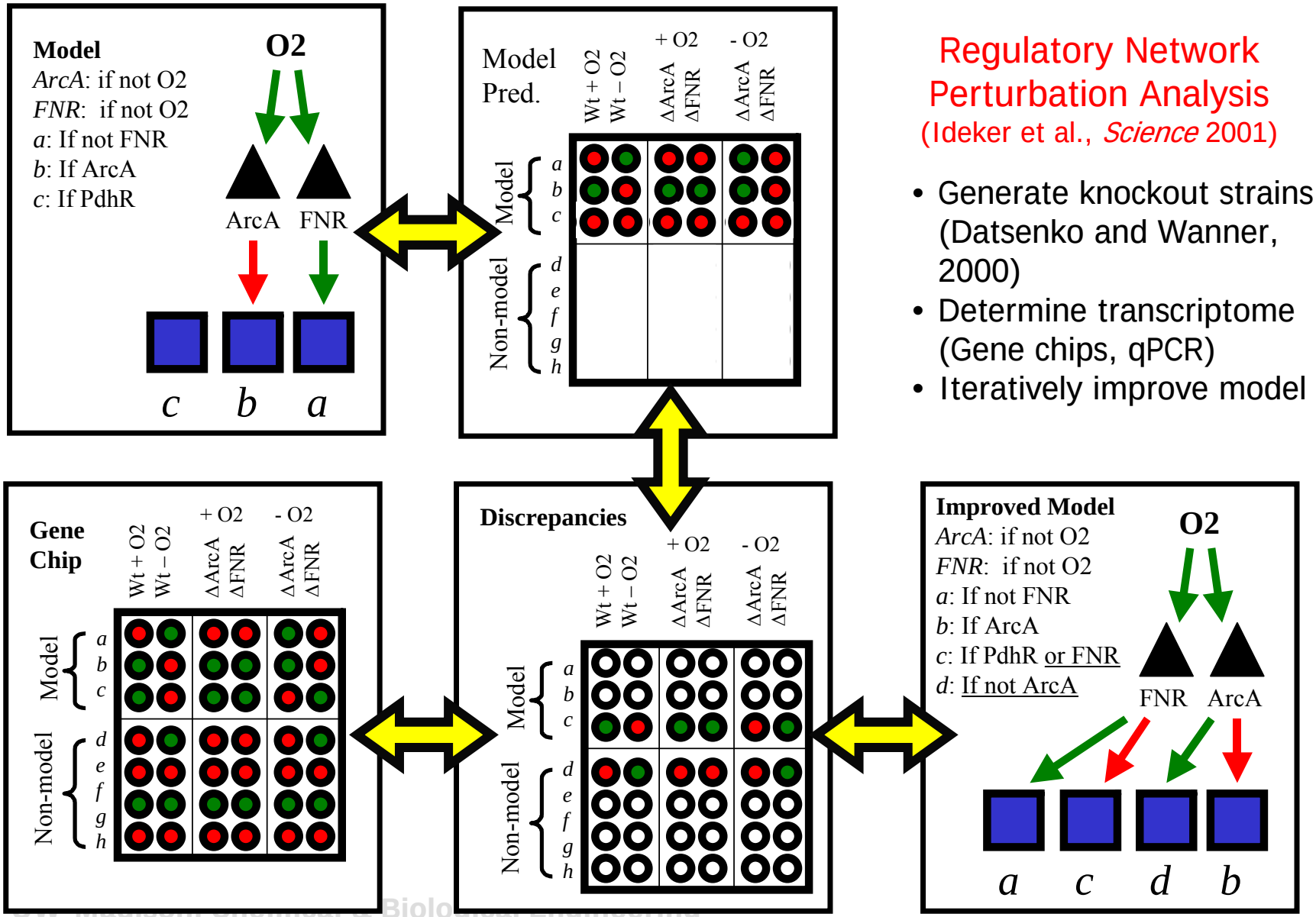
Covert, et al. Nature, **429**: 92-96 (2004).

Gene Expression Data

- Aerobic/Anaerobic Shift
  - Affymetrix gene chips/qPCR for +/- O2
  - 437 Significant Shifts (P <0.007, FDR = 5%)
  - 151 altered genes included in the model
- Original model
  - Model predictions of all conditions for comparison
  - 75 predicted shifts (28:NA)
  - Model predicts 23 shifts correctly
  - 23/151 = 15% coverage
  - 23/(75-28) = 49% accuracy

Shift Comparisons (predicted: experimental)		
Pred Total		75
Exp Total		437
Exp in Model		151
1	(P:E)	23
2	(0:0)	606
3	(P:-E)	1
4	(P:0)	23
5	(0:E)	127
No comparison possible		
(P:NA)		28
(0:NA)		197
(NM:E)		286
(NM:0)		2067
(NM:NA)		989

Regulatory Network Interrogation

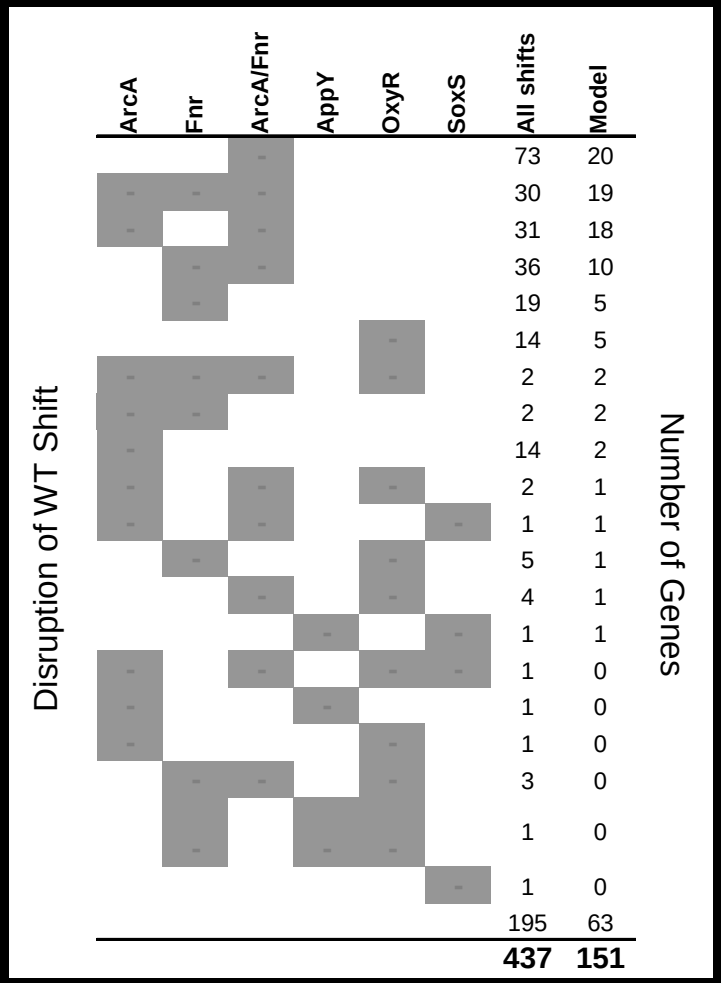


Regulatory Network Perturbation Analysis
(Ideker et al., *Science* 2001)

- Generate knockout strains (Datsenko and Wanner, 2000)
- Determine transcriptome (Gene chips, qPCR)
- Iteratively improve model

Model-Centric Hypothesis Generation

- Systematic network perturbation
  - Locate significant shifts in WT
  - Δ*arcA*⁻, Δ*fnr*⁻, Δ*arcA*⁻*fnr*⁻, Δ*appY*⁻, Δ*oxyR*⁻, Δ*soxS*⁻
  - ANOVA to identify rules
- Observations
  - Most rules unaffected by all KO strains
    - Still TFs/effects to identify
  - Complex rules (Δ*arcA*⁻ or Δ*fnr*⁻ but not Δ*arcA*⁻*fnr*⁻)
    - Microaerobic as in *cydAB*
    - Indirect metabolic/regulatory effects (Cra)
  - TF expression and activity correlation
    - arcA: positive correlation
    - *fnr, betI, fur*: negative correlation



Recommended Reading

- Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. Nature. 2004 May 6;429(6987):92-6.

- Covert MW, Palsson BØ. Transcriptional regulation in constraints-based metabolic models of Escherichia coli. J Biol Chem. 2002 Aug 2;277(31):28058-64.

- Covert MW, Schilling CH, Palsson B. Regulation of gene expression in flux balance models of metabolism. J Theor Biol. 2001 Nov 7;213(1):73-88.