



# Bioproduct Production

A Continually Evolving Process

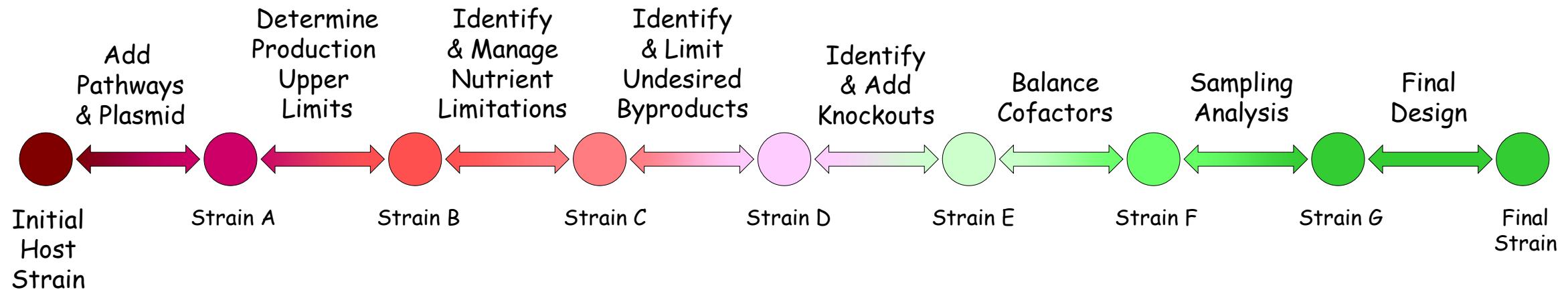


# Learning Objectives

- Explain the assumptions and limitations of strain design using constraint-based metabolic reconstructions.
- Explain the process of bioproduction identification.
- Explain the process of selecting a host strain.
- Explain the process of defining a bioproduction pathway.
- Explain the strain design process.
- Explain the purpose of understanding the bioproduction maximum production.
- Explain the purpose of carbon source selection.
- Explain the purpose of identifying nutrient & amino acid limitations.
- Explain the purpose of identifying undesired by-products.
- Explain the purpose of growth coupling.
- Explain the purpose of cofactor balancing.
- Explain the purpose of sampling analysis.



# Host Strain Design Process





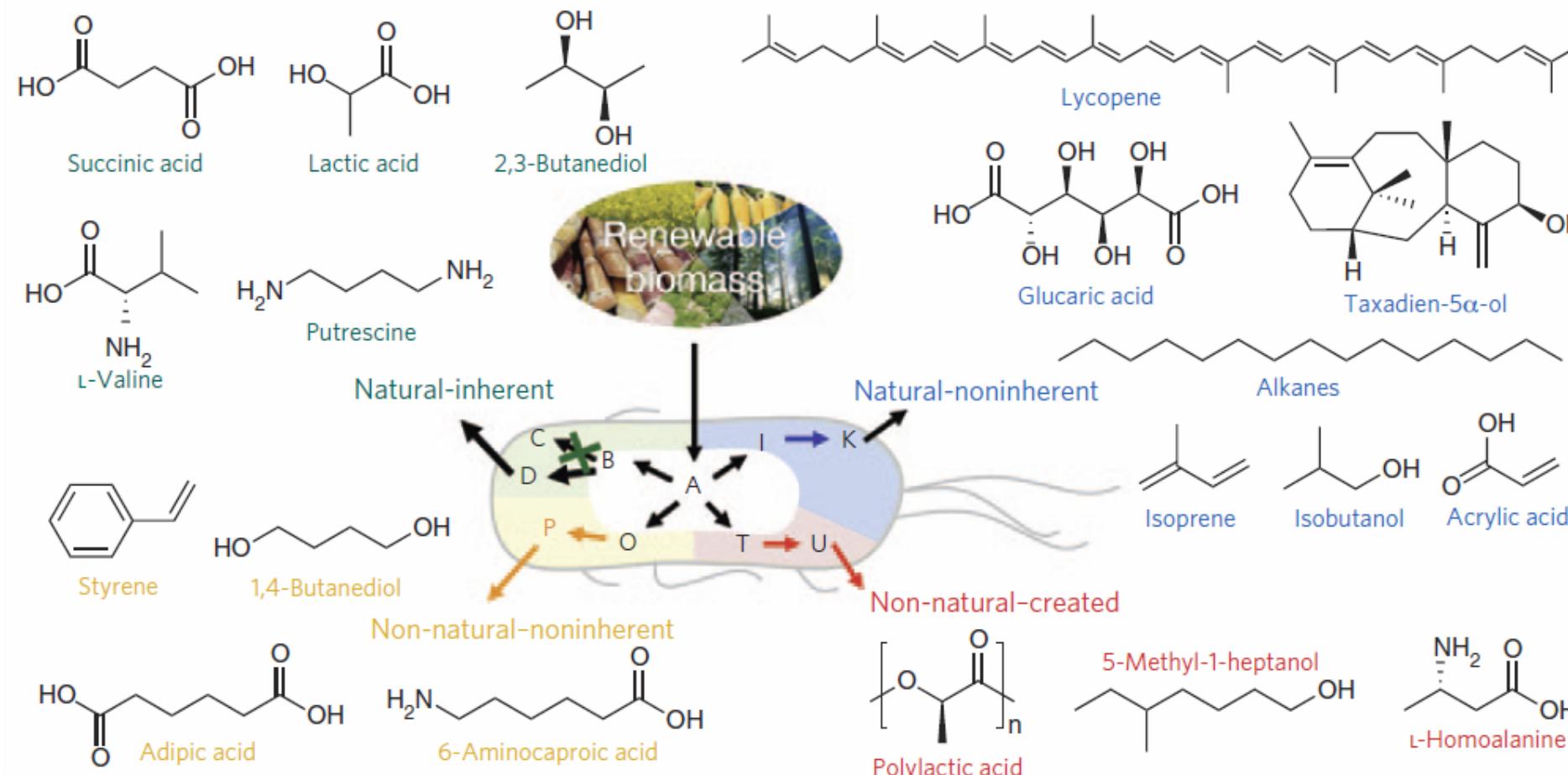
## Bioproduct Production

- Bioproduct Identification
- Select Host Strain
- Bioproduct Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduct Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





# Categories of Chemicals Produced by Microbial Cell Factories



Lee, J. W., D. Na, et al. (2012). "Systems metabolic engineering of microorganisms for natural and non-natural chemicals." *Nature chemical biology* 8(6): 536-546.



# Production of Recombinant Proteins by High Cell Density Culture of *E. coli*

Products	Host	Productivity and characteristics	Products	Host	Productivity and characteristics
Insulin-like growth factor-2 (IGF-2)	<i>E. coli</i> BL21(DE3)	9.69 g/L, inclusion body	Animolevulinate synthase	<i>E. coli</i> MG1655	5.2 g/L
Single-chain antibody variable fragment	<i>E. coli</i> RV308	1.2 g/L, PelB signal sequence	Annexin-V-hirudin chimeric protein	<i>E. coli</i> BL21(DE3)	10 mg/L/OD <sub>600</sub> (after purification)
Human interferon- $\gamma$ (hIFN-gamma)	<i>E. coli</i> BL21(DE3)	2 $\times$ 10 <sup>7</sup> U/mg protein, inclusion body	Human Tissue-type plasminogen activator	<i>E. coli</i> SF110	180 $\mu$ g/L (after purification), StII signal sequence, DsbC coexpression
Human Interleukin-7	<i>E. coli</i> HMS174 (DE3)	46% of total proteins, inclusion body	Human necrosis factor-related apoptosis-inducing ligand	<i>E. coli</i> C600	1.4 g/L soluble protein
Phytase	<i>E. coli</i> BL21(DE3)	120 U/mL, <i>kil</i> gene coexpression	Antifungal peptides	<i>E. coli</i>	40% of total proteins, PelB signal sequence, not secreted, inclusion body
Carbamoylase	<i>E. coli</i> BL21(G2)	14256 U/L, thermoregulated T7 promoter	Human necrosis factor- $\alpha$	<i>E. coli</i> BL21(DE3)	Constitutive production, PHCE promoter
Human epidermal growth factor	<i>E. coli</i> HB101	242 mg/L	Bioadhesive protein	<i>E. coli</i> AS002	5.3 g/L
Human epidermal growth factor	<i>E. coli</i> JM101	325 mg/L, OmpA signal sequence	Human leptin	<i>E. coli</i> BL21(DE3)	41% of total proteins, endoxyylanase signal sequence, DsbA coexpression
Alkaline phosphatase	<i>E. coli</i> HB101	5.2 g/L, Endoxylanase signal sequence.	Pectate lyase	<i>E. coli</i> BL21(DE3)	2200 U/mL, PelB signal sequence
Human granulocyte colony-stimulating factor (GCSF)	<i>E. coli</i> BL21(DE3)	22% of total proteins, Endoxylanase signal sequence	Human leptin	<i>E. coli</i> FMJ123	9.7 g/L, constitutive production, PHCE promoter
Protective antigen protein	<i>E. coli</i> DH5 $\alpha$	125 mg/L, inclusion body, constitutive expression	Insulin-like growth factor-1 fusion protein	<i>E. coli</i> W3110	4.3 g/L, PrsA and GlpF coexpression
Bone morphogenetic protein 2	<i>E. coli</i> TG1	8.6 g/L, inclusion body.			
Human mini-proinsulin	<i>E. coli</i> BL21(DE3)	7 g/L, two stage cyclic fed-batch culture			
Human interferon- $\alpha$	<i>E. coli</i> TG1	4 g/L, inclusion body			

Choi, J. H., K. C. Keum, et al. (2006). "Production of recombinant proteins by high cell density culture of Escherichia coli." Chemical Engineering Science 61(3): 876-885



# Metabolite Databases

- Biocyc
  - ✓ <http://biocyc.org/>
- KEGG
  - ✓ <http://www.genome.jp/kegg/>
- NCBI
  - ✓ <http://www.ncbi.nlm.nih.gov/>
- OMIM
  - ✓ <http://omim.org/>
- HGNC
  - ✓ <http://www.genenames.org/>
- HPRD
  - ✓ [http://www.hprd.org/index\\_html](http://www.hprd.org/index_html)
- e!Ensembl
  - ✓ <http://uswest.ensembl.org>
- Vega
  - ✓ <http://vega.sanger.ac.uk/index.html>
- UniProt
  - ✓ <http://www.uniprot.org/>

**KEGG** Homo sapiens (human): 3162 Help

<b>Entry</b>	3162	CDS	T01001
<b>Gene name</b>	HMOX1, HMOX1D, HO-1, HSP32, bK286B10		
<b>Definition</b>	(RefSeq) heme oxygenase 1 (EC:1.14.99.3)		
<b>KO</b>	K00510 heme oxygenase (biliverdin-producing) [EC:1.14.14.18]		
<b>Organism</b>	hsa Homo sapiens (human)		
<b>Pathway</b>	hsa00860 Porphyrin and chlorophyll metabolism hsa04066 HIF-1 signaling pathway hsa04978 Mineral absorption hsa05206 MicroRNAs in cancer		
<b>Brite</b>	KEGG Orthology (KO) [BR:hsa00001] Metabolism Metabolism of cofactors and vitamins 00860 Porphyrin and chlorophyll metabolism 3162 (HMOX1) Environmental Information Processing Signal transduction 04066 HIF-1 signaling pathway 3162 (HMOX1) Organismal Systems Digestive system 04978 Mineral absorption 3162 (HMOX1) Human Diseases Cancers 05206 MicroRNAs in cancer 3162 (HMOX1) Enzymes [BR:hsa01000] 1. Oxidoreductases 1.14 Acting on paired donors, with incorporation or reduction of one donor, and incorporation of two donors, with incorporation or reduction of one donor 1.14.14 With reduced flavin or flavoprotein as one donor, and incorporation of one other donor 1.14.14.18 heme oxygenase (biliverdin-producing) 3162 (HMOX1)		



# Bioproduction Production

- Bioproduction Identification
- • Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





# Tested Cobra Toolbox Models

<http://bigg.ucsd.edu/models>

**Model Number**

Model Number	Name
e_coli_core	Escherichia coli str. K-12 substr. MG1655
iAB_RBC_283	Homo sapiens
iAF1260	Escherichia coli str. K-12 substr. MG1655
iAF692	Methanoscarcina barkeri str. Fusaro
iAF987	Geobacter metallireducens GS-15
iAPEC01_1312	Escherichia coli APEC O1
iAT_PLT_636	Homo sapiens
iB21_1397	Escherichia coli BL21(DE3)
iBWG_1329	Escherichia coli BW2952
ic_1306	Escherichia coli CFT073
iE2348C_1286	Escherichia coli O127:H6 str. E2348/69
iEC042_1314	Escherichia coli 042
iEC55989_1330	Escherichia coli 55989
iECABU_c1320	Escherichia coli ABU 83972
iECB_1328	Escherichia coli B str. REL606
iECBD_1354	Escherichia coli 'BL21-Gold(DE3)pLysS AG'
iECD_1391	Escherichia coli BL21(DE3)
iECDH10B_1368	Escherichia coli str. K-12 substr. DH10B
iEcDH1_1363	Escherichia coli DH1
iECDH1ME8569_1439	Escherichia coli DH1
iEcE24377_1341	Escherichia coli E24377A
iECED1_1282	Escherichia coli ED1a
iECH74115_1262	Escherichia coli O157:H7 str. EC4115
iEcHS_1320	Escherichia coli HS
iECIAI1_1343	Escherichia coli IAI1
iECIAI39_1322	Escherichia coli IAI39

**Model Number**

Model Number	Name
iECNA114_1301	Escherichia coli NA114
iECO103_1326	Escherichia coli O103:H2 str. 12009
iECO111_1330	Escherichia coli O111:H- str. 11128
iECO26_1355	Escherichia coli O26:H11 str. 11368
iECOK1_1307	Escherichia coli IHE3034
iEcoliC_1368	Escherichia coli ATCC 8739
iECP_1309	Escherichia coli 536
iECs_1301	Escherichia coli O157:H7 str. Sakai
iECS88_1305	Escherichia coli S88
iECSE_1348	Escherichia coli SE11
iECSF_1327	Escherichia coli SE15
iEcSMS35_1347	Escherichia coli SMS-3-5
iECSP_1301	Escherichia coli O157:H7 str. TW14359
iECUMN_1333	Escherichia coli UMN026
iECW_1372	Escherichia coli W
iEKO11_1354	Escherichia coli KO11FL
iETEC_1333	Escherichia coli ETEC H10407
iG2583_1286	Escherichia coli O55:H7 str. CB9615
iHN637	Clostridium ljungdahlii DSM 13528
iIT341	Helicobacter pylori 26695
iJN678	Synechocystis sp. PCC 6803
iJN746	Pseudomonas putida KT2440
iJO1366	Escherichia coli str. K-12 substr. MG1655
iJR904	Escherichia coli str. K-12 substr. MG1655
iLF82_1304	Escherichia coli LF82
iLJ478	Thermotoga maritima MSB8

**Model Number**

Model Number	Name
iMM1415	Mus musculus
iMM904	Saccharomyces cerevisiae S288c
iND750	Saccharomyces cerevisiae S288c
iNJ661	Mycobacterium tuberculosis H37Rv
iNRG857_1313	Escherichia coli O83:H1 str. NRG 857C
iPC815	Yersinia pestis CO92
iRC1080	Chlamydomonas reinhardtii
iS_1188	Shigella flexneri 2a str. 2457T
iSB619	Staphylococcus aureus subsp. aureus N315
iSbBS512_1146	Shigella boydii CDC 3083-94
iSBO_1134	Shigella boydii Sb227
iSDY_1059	Shigella dysenteriae Sd197
iSF_1195	Shigella flexneri 2a str. 301
iSFV_1184	Shigella flexneri 5 str. 8401
iSFxv_1172	Shigella flexneri 2002017
iSSON_1240	Shigella sonnei Ss046
iUMN146_1321	Escherichia coli UM146
iUMNK88_1353	Escherichia coli UMNK88
iUTI89_1310	Escherichia coli UTI89
iWFL_1372	Escherichia coli W
iY75_1357	Escherichia coli str. K-12 substr. W3110
iYL1228	Klebsiella pneumoniae subsp. pneumoniae MGH 78578
iYO844	Bacillus subtilis subsp. subtilis str. 168
iZ_1308	Escherichia coli O157:H7 str. EDL933
RECON3D	Homo sapiens
STM_v1_0	Salmonella enterica subsp. enterica serovar Typhimurium str. LT2



# *E.coli* as a Host Organism

1. Unparalleled fast growth kinetics. In glucose-salts media and given the optimal environmental conditions, its doubling time is about 20 min. This means that a culture inoculated with a 1/100 dilution of a saturated starter culture may reach stationary phase in a few hours. However, it should be noted that the expression of a recombinant protein may impart a metabolic burden on the microorganism, causing a considerable decrease in generation time.
2. High cell density cultures are easily achieved. The theoretical density limit of an *E. coli* liquid culture is estimated to be about 200 gram dry cell weight/liter or roughly  $10^{13}$  viable bacteria/ml. However, exponential growth in complex media leads to densities nowhere near that number.
3. Rich complex media can be made from readily available and inexpensive components.
4. Transformation with exogenous DNA is fast and easy. Plasmid transformation of *E. coli* can be performed in as little as 5 min

Rosano, G. L. and E. A. Ceccarelli (2014). "Recombinant protein expression in Escherichia coli: advances and challenges." Frontiers in microbiology 5: 172.



# Host Selection

- Select host that naturally produces the target bioproduct
- Select heterologous host that can be engineered
  - ✓ Commonly used strains of *E.coli*
    - [http://openwetware.org/wiki/E.\\_coli\\_genotypes](http://openwetware.org/wiki/E._coli_genotypes)
- Commercial competent cell strains
  - ✓ Invitrogen - Top10 -  
<http://www.thermofisher.com/order/catalog/product/K456001>
  - ✓ Promega -  
<https://www.promega.com/products/vectors/bacterial-expression-vectors/>
  - ✓ Qiagen - <https://www.qiagen.com/us/shop/sample-technologies/protein-sample-technologies/expression-kits-and-vectors/>

## Some Commonly Used *E. coli* Strains

AG1	IJ1126	STBL4
AB1157	IJ1127	SURE (Stratagene)
B2155	JM83	SURE2 (Stratagene)
BL21	JM101	TG1 (Lucigen)
BL21(AI)	JM103	TOP10 (Invitrogen)
BL21(DE3)	JM105	Top10F' (Invitrogen)
BL21 (DE3) pLysS	JM106	W3110
BNN93	JM107	W3110 (λ857S7)
BNN97	JM108	WM3064
BW26434, CGSC Strain # 7658	JM109	XL1-Blue (Stratagene)
C600	JM109(DE3)	XL1-Blue MRF' (Stratagene)
C600 hflA150 (Y1073, BNN102)	JM110	XL2-Blue (Stratagene)
CSH50	JM2.300	XL2-Blue MRF' (Stratagene)
D1210	LE392	XL1-Red (Stratagene)
DB3.1	M15 (Qiagen)	XL10-Gold (Stratagene)
DH1	Mach1	XL10-Gold KanR (Stratagene)
DH5α	MC1061	
DH5α Turbo (NEB)	MC4100	
DH10B (Invitrogen)	MFDpir	
DH12S (Invitrogen)	MG1655	
DM1 (Invitrogen)	OmniMAX2	
E. coli(r) 5alpha (Lucigen)	OverExpress(tm)C41(DE3) (Lucigen)	
E. coli(r) 10G (Lucigen)	OverExpress(tm)C41(DE3)pLysS (Lucigen)	
E. coli(r) 10GF' (Lucigen)	OverExpress(tm)C43(DE3) (Lucigen)	
E. coli K12 ER2738 (NEB)	OverExpress(tm)C43(DE3)pLysS (Lucigen)	
ER2566 (NEB)	Rosetta(DE3)pLysS	
ER2267 (NEB)	Rosetta-gami(DE3)pLysS	
HB101	RR1	
HMS174(DE3)	RV308	
High-Control(tm) BL21(DE3) (Lucigen)	STBL2 (Invitrogen)	
High-Control(tm) 10G (Lucigen)	STBL3 (Invitrogen)	



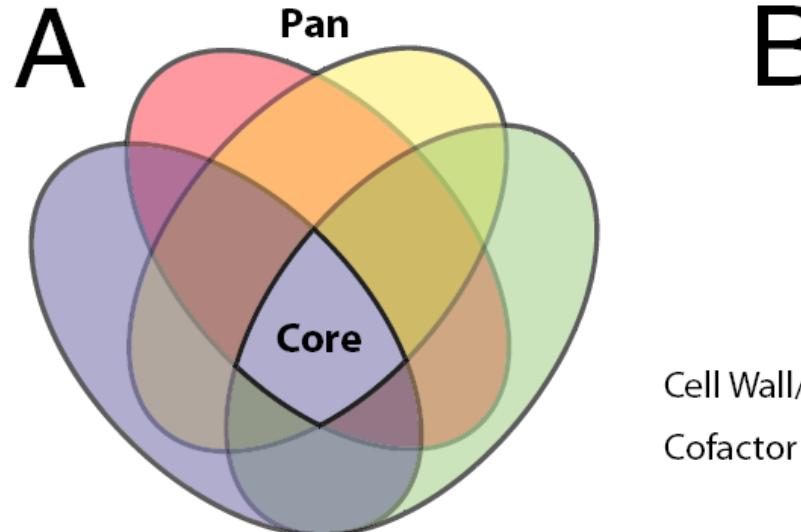
# *E. coli* Genome-scale Reconstructions

- Escherichia coli 042
- Escherichia coli 536
- Escherichia coli 55989
- Escherichia coli ABU 83972
- Escherichia coli APEC O1
- Escherichia coli ATCC 8739
- Escherichia coli B str. REL606
- Escherichia coli BL21(DE3) AM946981
- Escherichia coli BL21(DE3) BL21-Gold(DE3)pLySS AG
- **Escherichia coli BL21(DE3) CP001509**  
  
**Strain used  
at USU**
- Escherichia coli BW2952
- Escherichia coli CFT073
- Escherichia coli DH1
- Escherichia coli DH1 ME8569
- Escherichia coli E24377A
- Escherichia coli ED1a
- Escherichia coli ETEC H10407
- Escherichia coli HS
- Escherichia coli IAI1
- Escherichia coli IAI39
- Escherichia coli IHE3034
- Escherichia coli KO11FL
- Escherichia coli LF82
- Escherichia coli NA114
- Escherichia coli O103:H2 str. 12009
- Escherichia coli O111:H- str. 11128
- Escherichia coli O127:H6 str. E2348/69
- Escherichia coli O157:H7 EDL933
- Escherichia coli O157:H7 str. EC4115
- Escherichia coli O157:H7 str. Sakai
- Escherichia coli O157:H7 str. TW14359
- Escherichia coli O26:H11 str. 11368
- Escherichia coli O55:H7 str. CB9615
- Escherichia coli O83:H1 str. NRG 857C
- Escherichia coli S88
- Escherichia coli SE11
- Escherichia coli SE15
- Escherichia coli SMS-3-5
- Escherichia coli str. K-12 substr. DH10B
- Escherichia coli str. K-12 substr. MG1655
- Escherichia coli str. K-12 substr. W3110
- Escherichia coli UM146
- Escherichia coli UMN026
- Escherichia coli UMNK88
- Escherichia coli UTI89
- Escherichia coli W
- Escherichia coli W CP002185
- Escherichia coli K-12 MG1655

Monk, J. M., P. Charusanti, et al. (2013). Proceedings of the National Academy of Sciences of the United States of America 110(50): 20338-20343.

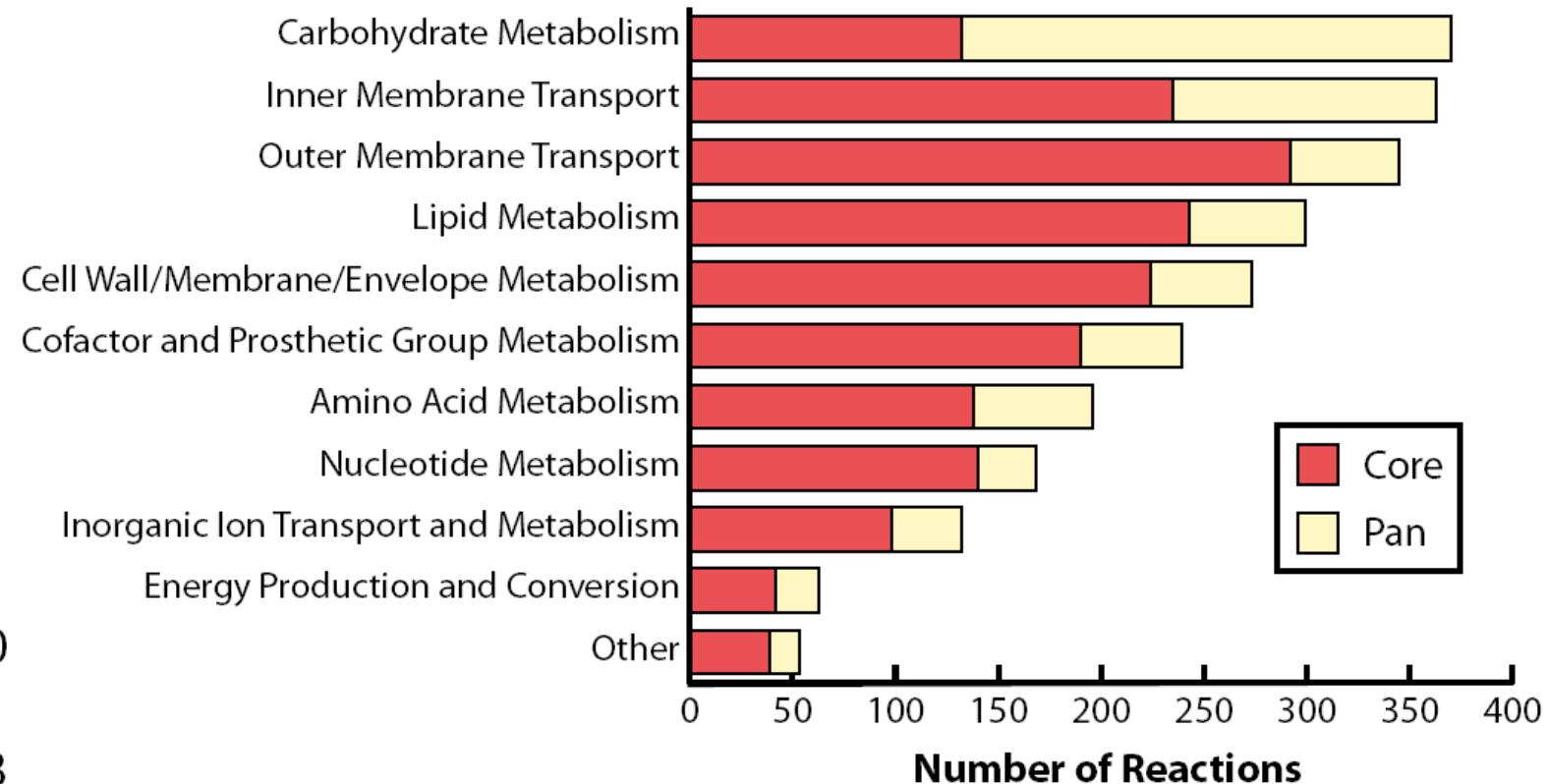


# Core and Pan Metabolic Capabilities of the *E. coli* Species.



	<b>Core:</b>	<b>Pan:</b>
<b>Genes:</b>	965	1,460
<b>Reactions:</b>	1,773	2,501
<b>Metabolites:</b>	1,665	2,043

## B Reaction Distribution by Subsystem



Monk, J. M., P. Charusanti, et al. (2013). Proceedings of the National Academy of Sciences of the United States of America 110(50): 20338-20343.



# BL21(DE3) *E.coli* Strain

- BL21 cells are deficient in the Lon protease which degrades many foreign proteins.
- Another gene missing from the genome 21 is the one coding for the outer membrane protease OmpT, whose function is to degrade extracellular proteins. The liberated amino acids are then taken up by the cell. This is problematic in the expression of a recombinant protein as, after cell lysis, OmpT may digest it.
- Plasmid loss is prevented thanks to the hsdSB mutation which disrupts DNA methylation and degradation.
- When the genes for the desired recombinant protein placed under a T7 promoter, then T7 RNA Polymerase (RNAP) needs to be provided by the host cell. In the popular BL21(DE3) strain, the λDE3 prophage was inserted in the chromosome of BL21 and contains the T7 RNAP gene under the lacUV5 promoter.
- The BL21(DE3) and its derivatives are by far the most used strains for protein expression.

Rosado, G. L. and E. A. Ceccarelli (2014). "Recombinant protein expression in Escherichia coli: advances and challenges." Frontiers in microbiology 5: 172.



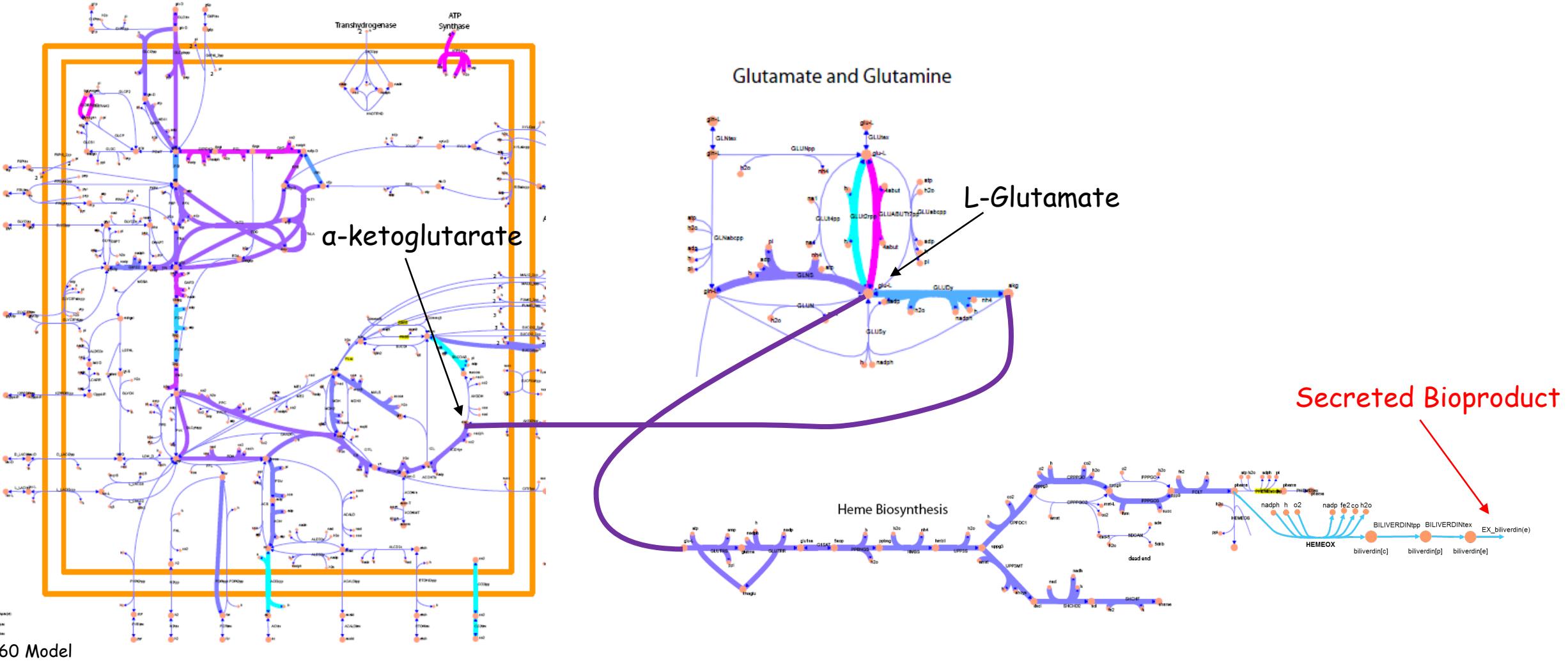
# Bioproduction Production

- Bioproduction Identification
- Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





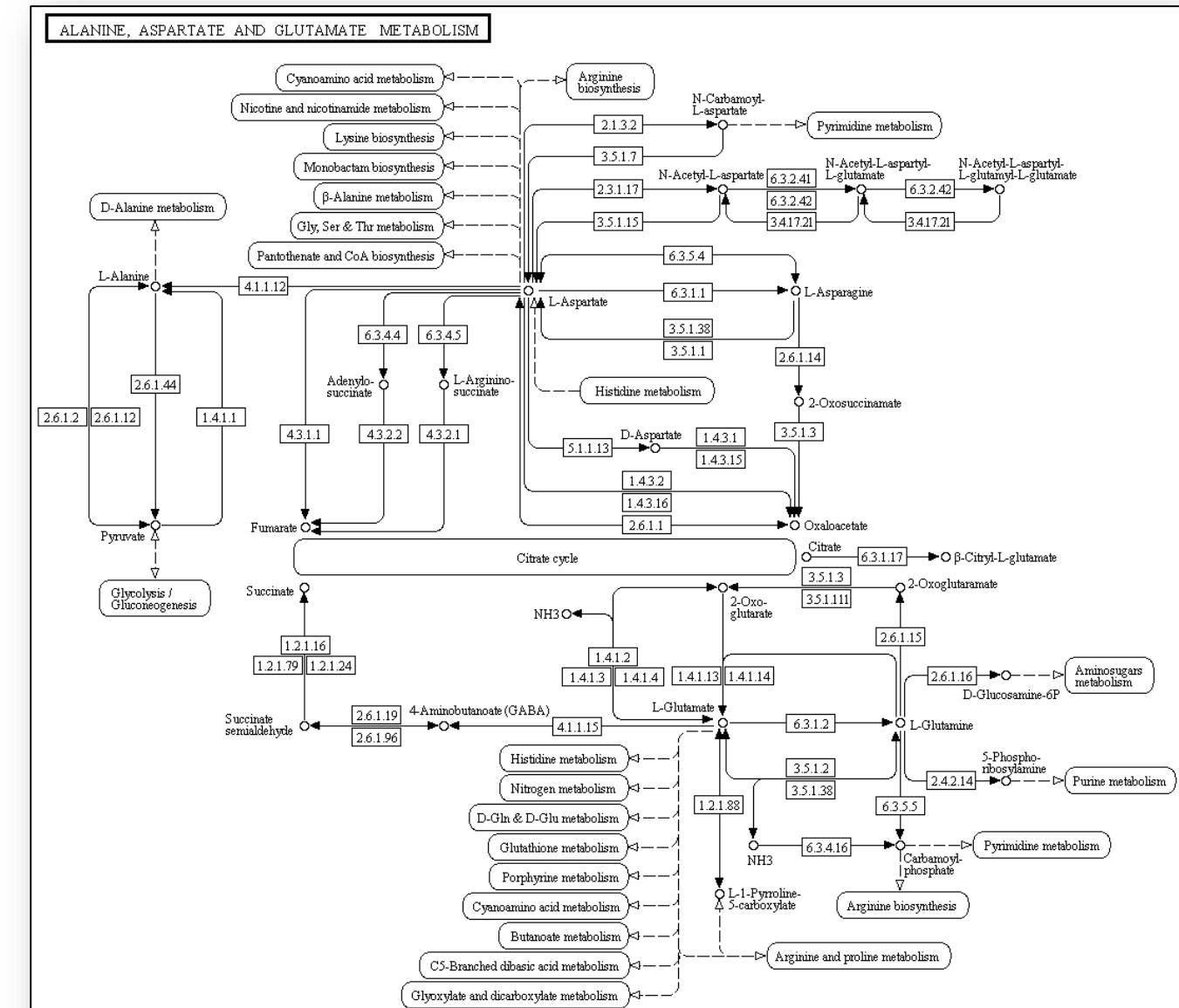
# Defining the Pathway of the Desired Bioproduct





# Pathway Databases

- Biocyc
  - ✓ <http://biocyc.org/>
- KEGG
  - ✓ <http://www.genome.jp/kegg/>
- ✓ AmiGO 2
  - ✓ <http://amigo.geneontology.org/amigo/landing>
- ✓ Reactome
  - ✓ <http://www.reactome.org/>
- ✓ WikiPathways
  - ✓ <http://www.wikipathways.org/index.php/WikiPathways>
- ✓ CPDB
  - ✓ <http://cpdb.molgen.mpg.de/>





# Bioproduction Production

- Bioproduction Identification
- Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- 
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





# Creating the Pathway in the Model: Adding Reactions

`addReaction` - Add a reaction to the model or modify an existing reaction

```
model = addReaction(model,rxnName,metaboliteList,stoichCoeffList,revFlag,lowerBound,upperBound,objCoeff,subSystem,grRule,geneNameList,systNameList,checkDuplicate)  
model = addReaction(model,rxnName,rxnFormula)
```

## INPUTS

model	COBRA model structure
rxnName	Reaction name abbreviation (i.e. 'ACALD') (Note: can also be a cell array {'abbr','name'})
metaboliteList	Cell array of metabolite names or alternatively the reaction formula for the reaction
stoichCoeffList	List of stoichiometric coefficients (reactants -ve, products +ve), empty if reaction formula is provided

## OPTIONAL INPUTS

revFlag	Reversibility flag (Default = true)
lowerBound	Lower bound (Default = 0 or -vMax)
upperBound	Upper bound (Default = vMax)
objCoeff	Objective coefficient (Default = 0)
subsystem	Subsystem (Default = '')
grRule	Gene-reaction rule in boolean format (and/or allowed) (Default = ''');
geneNameList	List of gene names (used only for translation from common gene names to systematic gene names)
systNameList	List of systematic names
checkDuplicate	Check S matrix too see if a duplicate reaction is already in the model (Default true)

## OUTPUTS

model	COBRA model structure with new reaction
rxnIDexists	Empty if the reaction did not exist previously, or if checkDuplicate is false. Otherwise it contains the ID of an identical reaction already present in the model.

## EXAMPLES

1) Add a new irreversible reaction using the formula approach

```
model = addReaction(model,'newRxn1','A -> B + 2 C')
```

2) Add the same reaction using the list approach

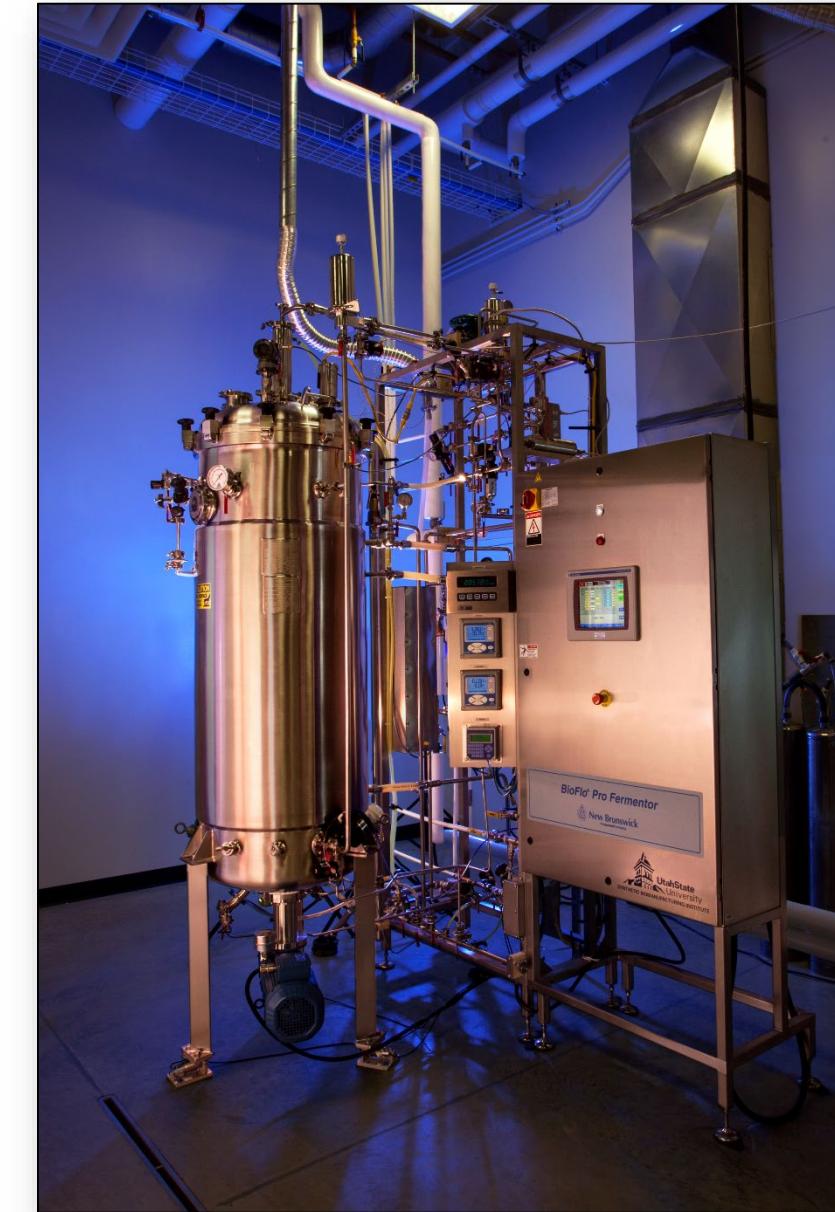
```
model = addReaction(model,'newRxn1',{'A','B','C'},[-1 1 2],false);
```

[http://opencobra.sourceforge.net/openCOBRA/opencobra\\_documentation/cobra\\_toolbox\\_2/index.html](http://opencobra.sourceforge.net/openCOBRA/opencobra_documentation/cobra_toolbox_2/index.html)



# Bioproduction Production

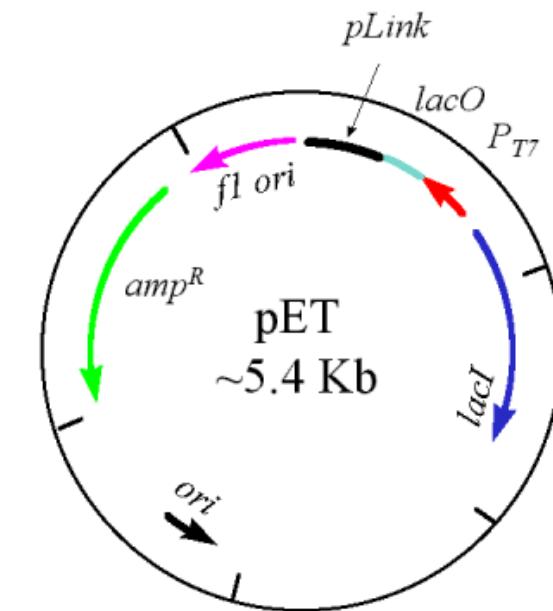
- Bioproduction Identification
- Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





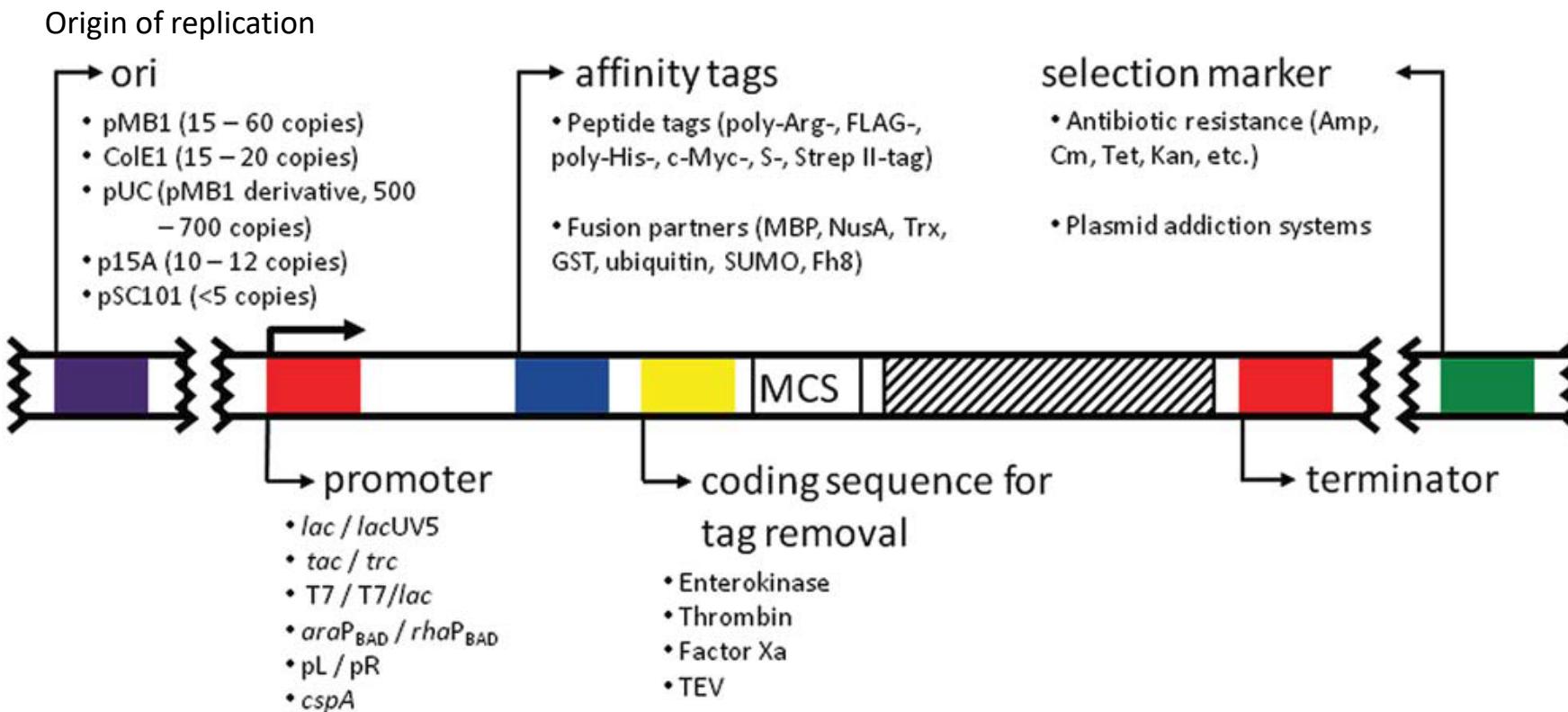
# Adding a Plasmid to a Cobra Model

- A plasmid is an expression vector designed for protein production in a cell. An expression vector must have elements necessary for protein expression. These may include a promoter, affinity tags for purification, the cloning site for the desired recombinant protein, a terminator and a selection marker (antibiotic resistance).
- The first option is to include all the plasmid components in the biomass function. This option is difficult to provide the flexibility needed to account for the variation in plasmid copy number and promoter strength.
- The second option is to create separate reactions that can be adjusted to model the plasmid load on the host cell. These separate reactions can include: nucleotide precursors, antibiotic marker production, recombinant protein production, and Lac repressor production (for Lac promoters).





# Anatomy of an Expression Vector



Rosado, G. L. and E. A. Ceccarelli (2014). "Recombinant protein expression in Escherichia coli: advances and challenges." Frontiers in microbiology 5: 172.



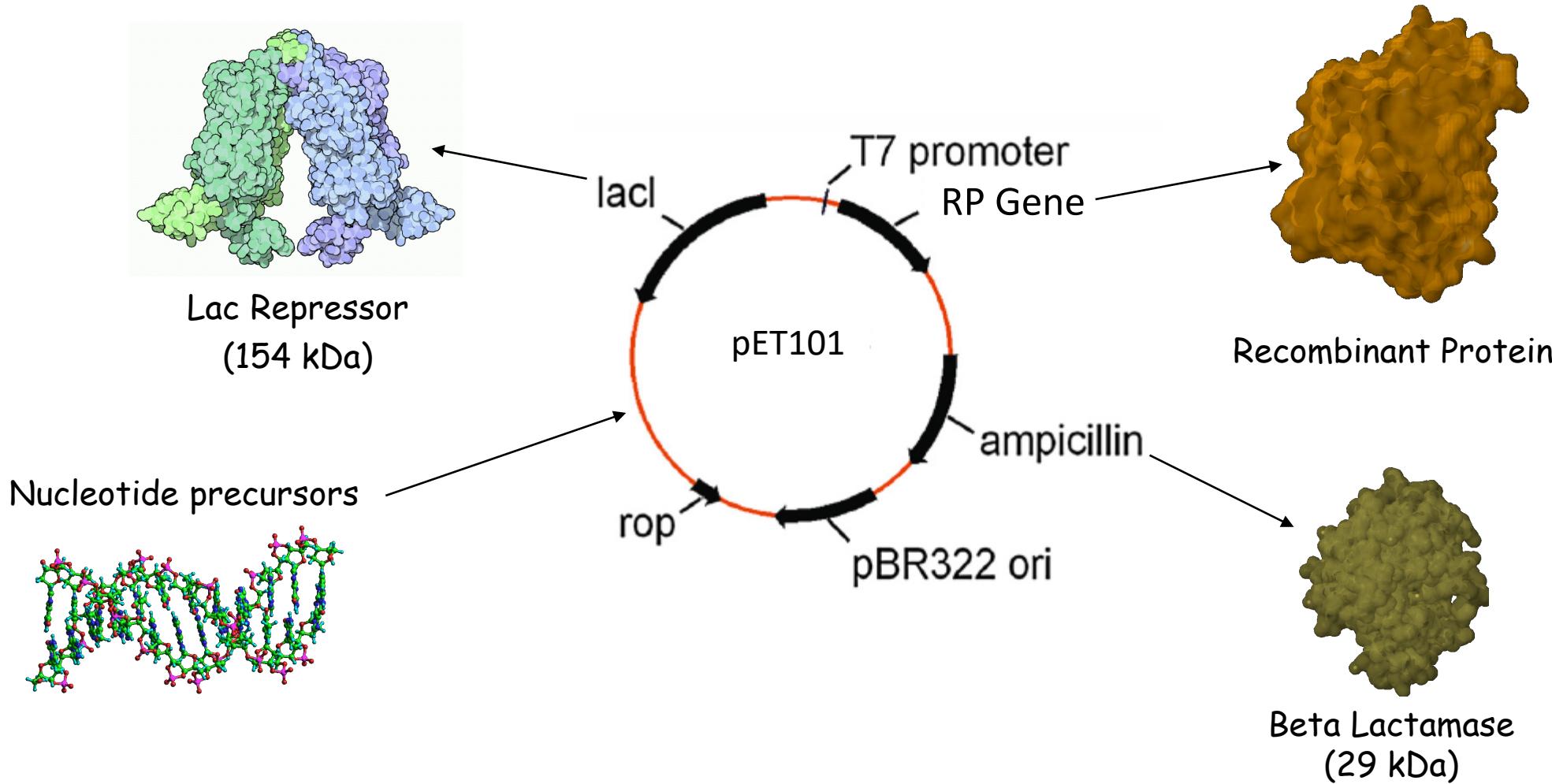
# Promoters

- **T5 Promoter:** The T5 promoter gives high protein expression levels; it is tightly repressible, and works in any *E. coli* strain. Silencing of the promoter prior to IPTG induction is achieved using symmetrical lac operators spaced around the promoter to maximize cooperativity. This operator pair ensures significantly tighter repression than regular lac operators.
- **T7 Promoter:** The T7 promoter results in high protein expression levels and is repressible. The T7 promoter only works in T7 pol expressing *E. coli* strains (eg. BL21(DE3) or T7 Express). Inducible by IPTG
- **PhoA Promoter:** The bacterial alkaline phosphatase (*phoA*) promoter is a strong promoter that is inexpensively induced when the culture is starved for inorganic phosphate. In the presence of phosphate, expression is tightly regulated making this system useful for expression of toxic proteins. Very high product yields have been obtained by use of the *phoA* system.
- **Arabinose Promoter:** The arabinose-inducible promoter *ara* is capable of high level recombinant protein expression in the presence of arabinose and is tightly regulated by glucose in the absence of arabinose. The *ara* promoter controls the genes *ara* organized in one operon. The promoter is flanked by a pair of lac operators that are recognized by the lac repressor also carried on the plasmid. IPTG binds to the repressor, inducing expression. The *ara* promoter is compatible with *E. coli* strains BL21 or DH5 $\alpha$ . In *E. coli*, glucose shuts down expression very well, even in the presence of rhamnose. This is why it works for autoinduction. You add both rhamnose and glucose, once glucose is consumed rhamnose induces.
- **Rhamnose Promoter:** Rhamnose-inducible bacterial expression vectors with different strength ribosome binding sites and origins of replication provide an excellent range of induced and uninduced expression levels. Rhamnose Expression Vectors are tightly controlled by rhamnose, enabling high expression yields, even for toxic or challenging proteins and can be used in any *E. coli* strain or other Gram-negative bacteria.

<https://www.dna20.com/products/expression-vectors/bacterial#2>

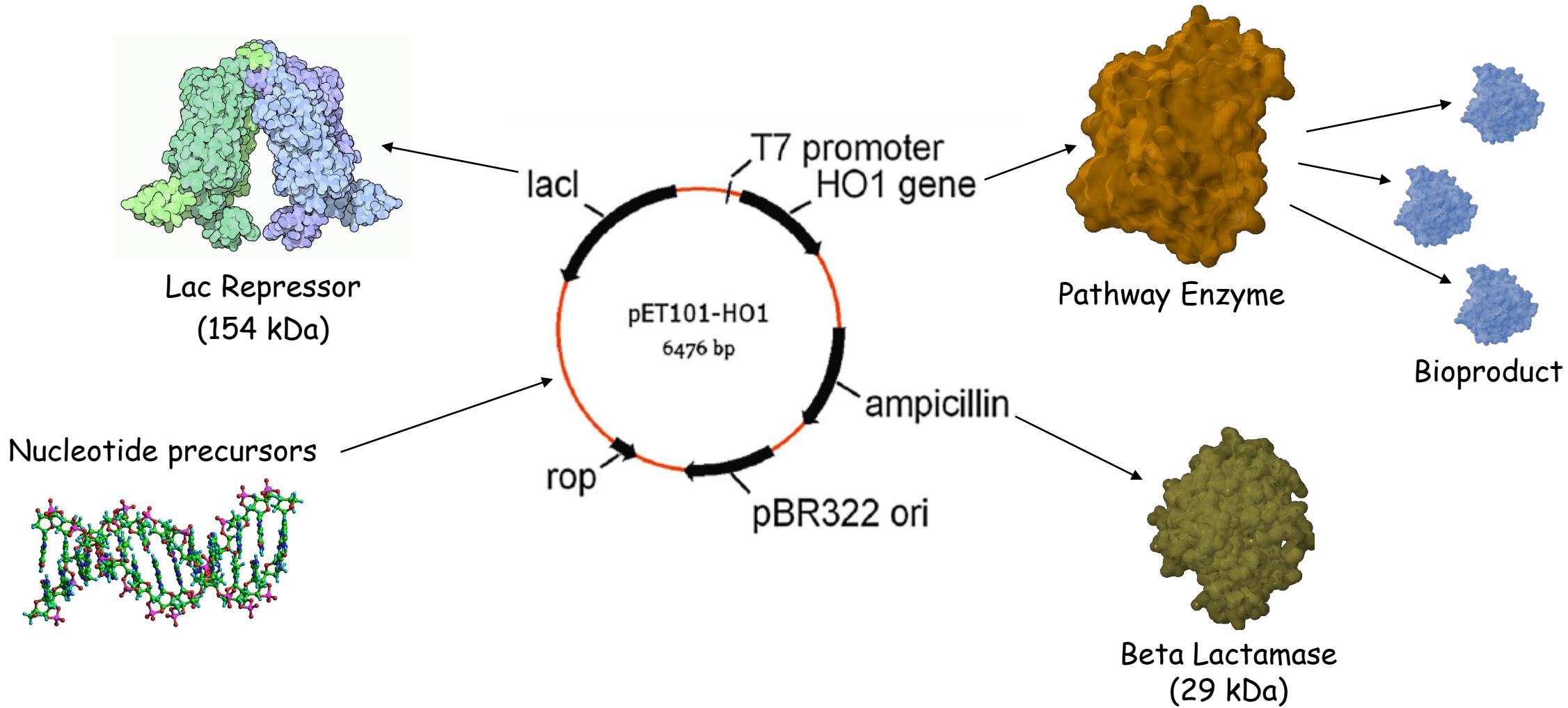


# Components Produced by Plasmid





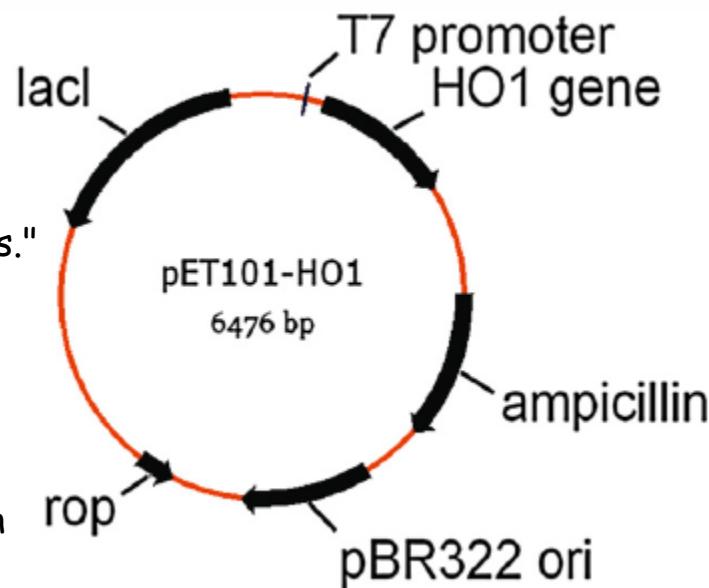
# Components Produced by Plasmid





# Plasmid Production Load

- Nucleotide precursors
  - ✓ The plasmid requires four nucleotide precursors (dGTP, dCTP, dATP, dTTP) for its replication based on an approximated *E. coli* GC content of 50.6%.
    - Ow, D. S., D. Y. Lee, et al. (2009). "Identification of cellular objective for elucidating the physiological state of plasmid-bearing *Escherichia coli* using genome-scale in silico analysis." *Biotechnology progress* 25(1): 61-67.
- Total antibiotic marker production
  - ✓ Can be up to 3% of the total protein mass
    - Ow, D. S., D. Y. Lee, et al. (2009). "Identification of cellular objective for elucidating the physiological state of plasmid-bearing *Escherichia coli* using genome-scale in silico analysis." *Biotechnology progress* 25(1): 61-67.
- Recombinant protein production
  - ✓ With high copy, strong promoter vectors the recombinant protein concentration can be as high as 50% of the total protein mass
    - Baneyx, F. (1999). "Recombinant protein expression in *Escherichia coli*." *Current opinion in biotechnology* 10(5): 411-421.
- Lac repressor production
  - ✓ There must be a low concentration of repressors or gene expression could not be induced
  - ✓ Typically less than 0.1% of the total protein; should be able to ignored





# Creating the Nucleotide Precursors Reactions

The biosynthetic precursors and energetic requirements for plasmid DNA replication

- Each plasmid (number of base pairs) requires four nucleotide precursors (dGTP, dCTP, dATP, & dTTP) for its DNA replication based on an approximated hosts GC content.
- For the case of the pOri2, which is composed of 4,575 bp, the plasmid requires four nucleotide precursors for its replication based on an approximated E coli K-12 GC content of 50.6%:
  - ✓ dGTP coefficient =  $4575 * 0.506 = 2315.0$
  - ✓ dCTP coefficient =  $4575 * 0.506 = 2315.0$
  - ✓ dATP coefficient =  $4575 * 0.494 = 2260.1$
  - ✓ dTTP coefficient =  $4575 * 0.494 = 2260.1$
- Thus
  - ✓ `model=addReaction(model,'PLASMID','2315.0 dgtp[c] + 2315.0 dctp[c] + 2260.1 datp[c] + 2260.1 dtpp[c] -> plasmid[c]);`
- A demand reaction for the plasmid is required to set the production rate of the nucleotide precursors .
  - ✓ `model = addDemandReaction(model,'plasmid[c]');`
  - ✓ `model = changeRxnBounds(model,'DM_plasmid[c]',0.63e-9,'l');` % Measured flux rates 0.64 pmol/gDW-hr (copy number ~600)

Ow, D. S., D. Y. Lee, et al. (2009). "Identification of cellular objective for elucidating the physiological state of plasmid-bearing Escherichia coli using genome-scale in silico analysis." *Biotechnology progress* 25(1): 61-67.



# Creating the Antibiotic Marker Reactions

The biosynthetic precursors and energetic requirements for antibiotic marker protein

- The biosynthetic precursor balance for the sole plasmid encoded antibiotic marker protein combined with energetic requirements of 4.306 mol ATP/mol amino acids.
- The reaction includes the number of each amino acids that makes up the protein plus the energetic requirements of ATP, ADP and Pi which is the sum of the number of amino acids multiplied by 4.306 mol ATP/mol amino acids.
- For the case of the pOri2, where  $\beta$ -lactamase is used for resistance to penicillins, cephalosporins, cephemycins and carbapenems, the total number of amino acids is 286 ( $286 * 4.306 = 1,231.52$ )
  - ✓ `model = addReaction(model, 'b-lactamase', '28 ala-L[c] + 3 cys-L[c] + 16 asp-L[c] + 20 glu-L[c] + 9 phe-L[c] + 21 gly[c] + 7 his-L[c] + 17 ile-L[c] + 11 lys-L[c] + 33 leu-L[c] + 10 met-L[c] + 8 asn-L[c] + 14 pro-L[c] + 9 gln-L[c] + 19 arg-L[c] + 17 ser-L[c] + 20 thr-L[c] + 16 val-L[c] + 4 trp-L[c] + 4 tyr-L[c] + 1,231.52 atp[c] -> b-lactamase[c] + 1,231.52 adp[c] + 1,231.52 pi[c']);`
- A demand reaction for the  $\beta$ -lactamase gene assumes 3% of total cellular protein (copy number ~600).
  - ✓ `model = addDemandReaction(model, 'b-lactamase[c]');`
  - ✓ `model = changeRxnBounds(model, 'DM_b-lactamase[c]', 0.000569, 'l');`
- The model does not include reactions for the synthesis of stress-response metabolites and proteins.

Ow, D. S., D. Y. Lee, et al. (2009). "Identification of cellular objective for elucidating the physiological state of plasmid-bearing Escherichia coli using genome-scale in silico analysis." *Biotechnology progress* 25(1): 61-67.



# Creating Recombinant Protein Enzyme

- The biosynthetic precursor balance for the sole plasmid encoded recombinant protein combined with energetic requirements of 4.306 mol ATP/mol amino acids.
- The reaction includes the number of each amino acids that makes up the protein plus the energetic requirements of ATP, ADP and Pi which is the sum of the number of amino acids multiplied by 4.306 mol ATP/mol amino acids.
- For example, the heme oxygenase gene is represented by
  - ✓ model = addReaction(model,'HOprotein','35 ala-L[c] + 3 cys-L[c] + 14 asp-L[c] + 36 glu-L[c] + 17 phe-L[c] + 19 gly[c] + 6 his-L[c] + 6 ile-L[c] + 26 lys-L[c] + 32 leu-L[c] + 13 met-L[c] + 14 asn-L[c] + 10 pro-L[c] + 16 gln-L[c] + 13 arg-L[c] + 12 ser-L[c] + 14 thr-L[c] + 13 val-L[c] + 2 trp-L[c] + 14 tyr-L[c] + 1,356.39 atp[c] -> HO[c] + 1,356.39 adp[c] + 1,356.39 pi[c']);
- A demand reaction for the heme oxygenase gene is required to set can be as high as 50% of the total protein mass.
  - ✓ model = addDemandReaction(model,'HO[c]');
  - ✓ model = changeRxnBounds(model,'DM\_HO[c]',0.0102,'l');
- Remember that Cobra biomass includes the required components (amino acids, etc.) of the host cell reactions in the biomass function, thus the need to add the separate load of the recombinant protein enzyme to the model.



```
% Biliverdin_Production_PPP_BL21_Plasmid.m
clear;
% Input the E.coli core model
model=readCbModel('iECD_1391');

% Add plasmid nucleotide precursors
model=addReaction(model,'PLASMID','3276.86 dgtp[c] + 3276.86 dctp[c] + 3199.14 datp[c] + 3199.14 dtpp[c] -> plasmid[c]');

% Add demand reaction for plasmid DNA and set flux rate
model = addDemandReaction(model,'plasmid[c]');
model = changeRxnBounds(model,'DM_plasmid[c]',0.63e-9,'l');

% Add beta-lactamase
model=addReaction(model,'b-lactamase','28 ala-L[c] + 3 cys-L[c] + 16 asp-L[c] + 20 glu-L[c] + 9 phe-L[c] + 21 gly[c] + 7 his-L[c] + 17 ile-L[c] + 11 lys-L[c] + 33 leu-L[c] + 10 met-L[c] + 8 asn-L[c] + 14 pro-L[c] + 9 gln-L[c] + 19 arg-L[c] + 17 ser-L[c] + 20 thr-L[c] + 16 val-L[c] + 4 trp-L[c] + 4 tyr-L[c] + 1,231.52 atp[c] -> b-lactamase[c] + 1,231.52 adp[c] + 1,231.52 pi[c]');

% Add demand reaction for beta-lactamase
model = addDemandReaction(model,'b-lactamase[c]');
model = changeRxnBounds(model,'DM_b-lactamase[c]',0.000569,'l');

% Add heme oxygenase enzyme
model=addReaction(model,'HOprotein','35 ala-L[c] + 3 cys-L[c] + 14 asp-L[c] + 36 glu-L[c] + 17 phe-L[c] + 19 gly[c] + 6 his-L[c] + 6 ile-L[c] + 26 lys-L[c] + 32 leu-L[c] + 13 met-L[c] + 14 asn-L[c] + 10 pro-L[c] + 16 gln-L[c] + 13 arg-L[c] + 12 ser-L[c] + 14 thr-L[c] + 13 val-L[c] + 2 trp-L[c] + 14 tyr-L[c] + 1,356.39 atp[c] -> HO[c] + 1,356.39 adp[c] + 1,356.39 pi[c]');

% Add demand reaction for heme oxygenase enzyme
model = addDemandReaction(model,'HO[c]');
model = changeRxnBounds(model,'DM_HO[c]',0.0102,'l'); % 50% of total protein

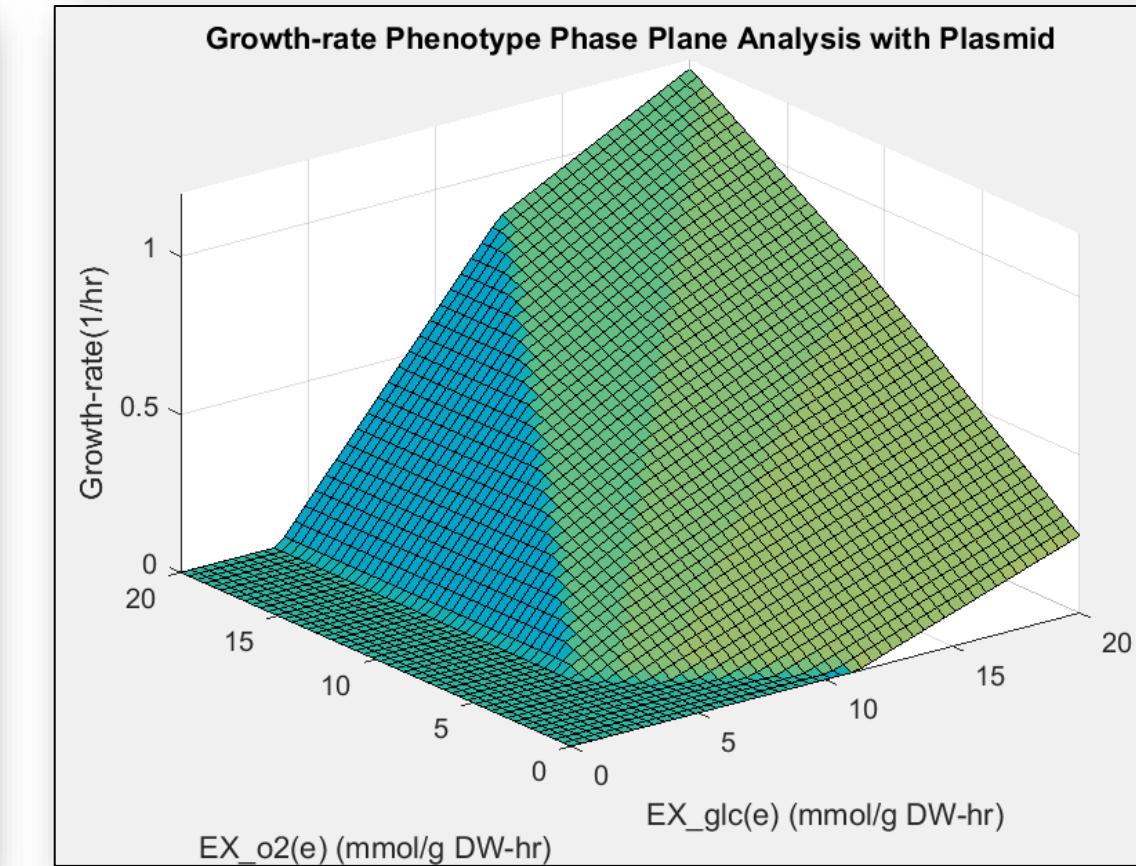
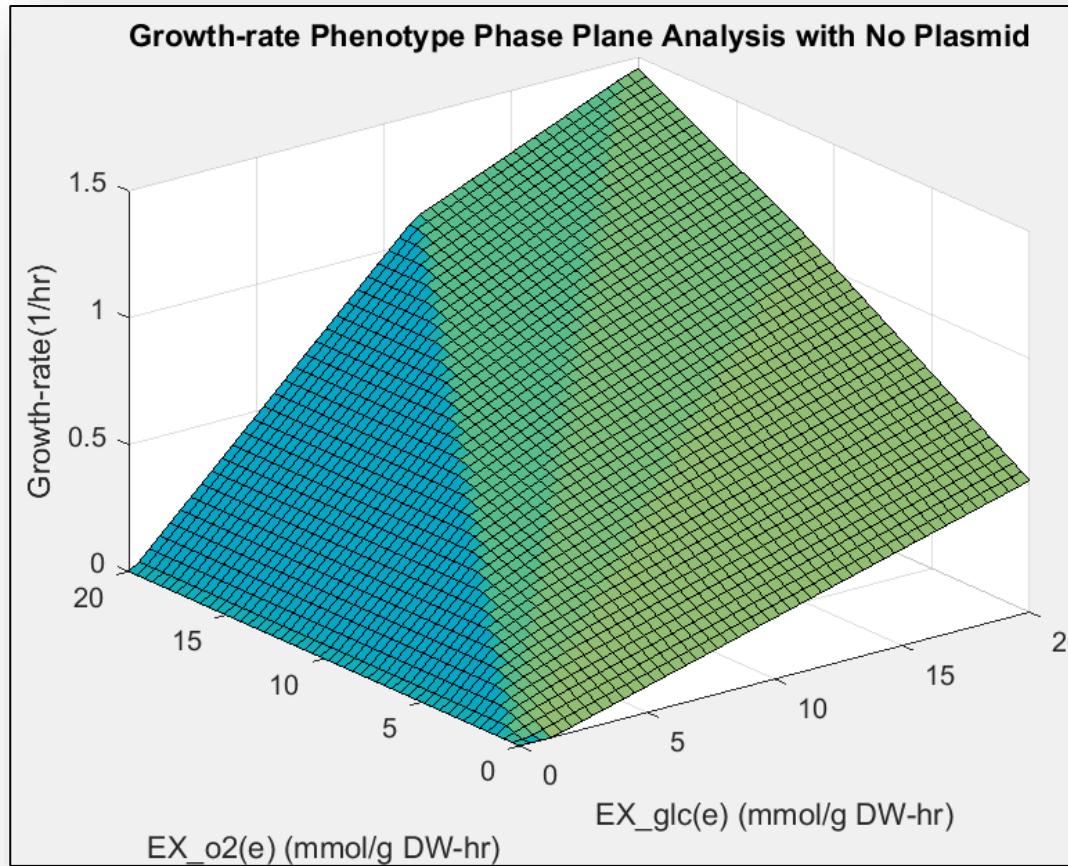
% Add heme oxygenase reaction
model=addReaction(model,'HEMEOX','pheme[c] + 3 nadph[c] + 5 h[c] + 3 o2[c] -> biliverdin[c] + fe2[c] + co[c] + 3 nadp[c] + 3 h2o[c]');
```

# Plasmid Code



# Metabolic Load of Plasmids - Phenotype Phase Plane

(Without Bioproduct Production [Spider Silk])

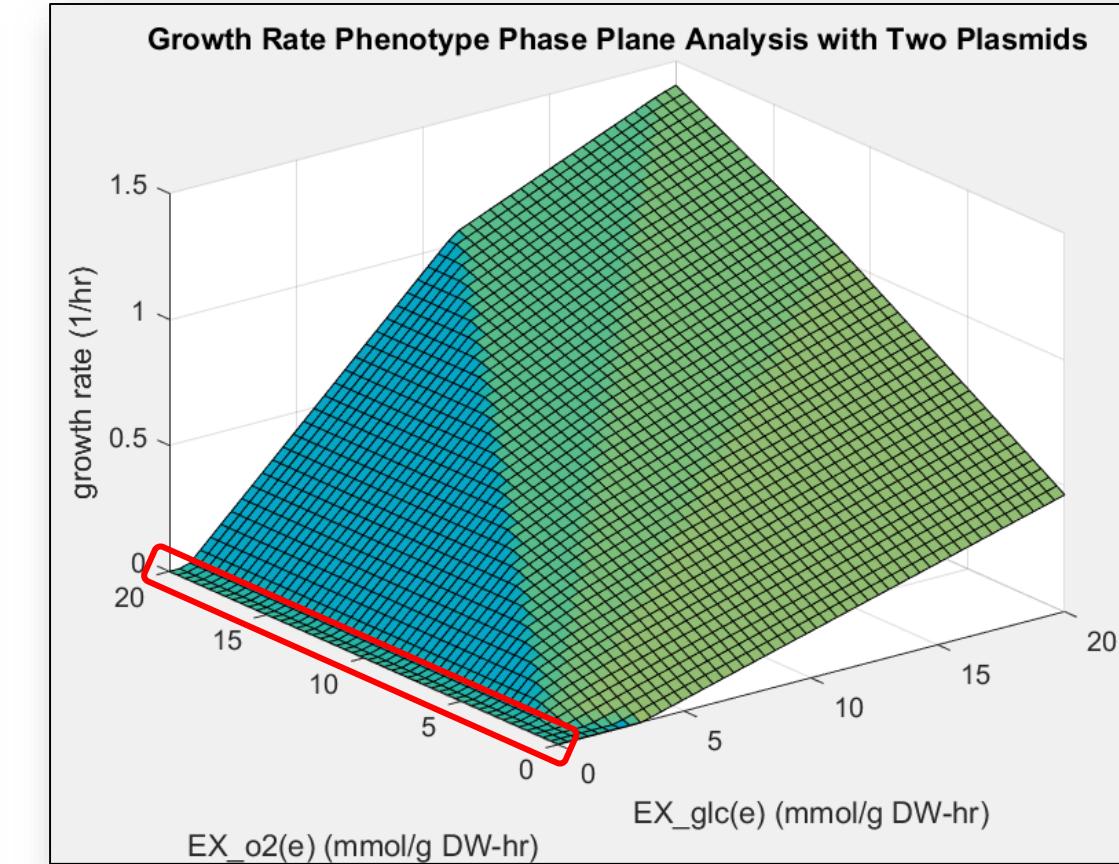
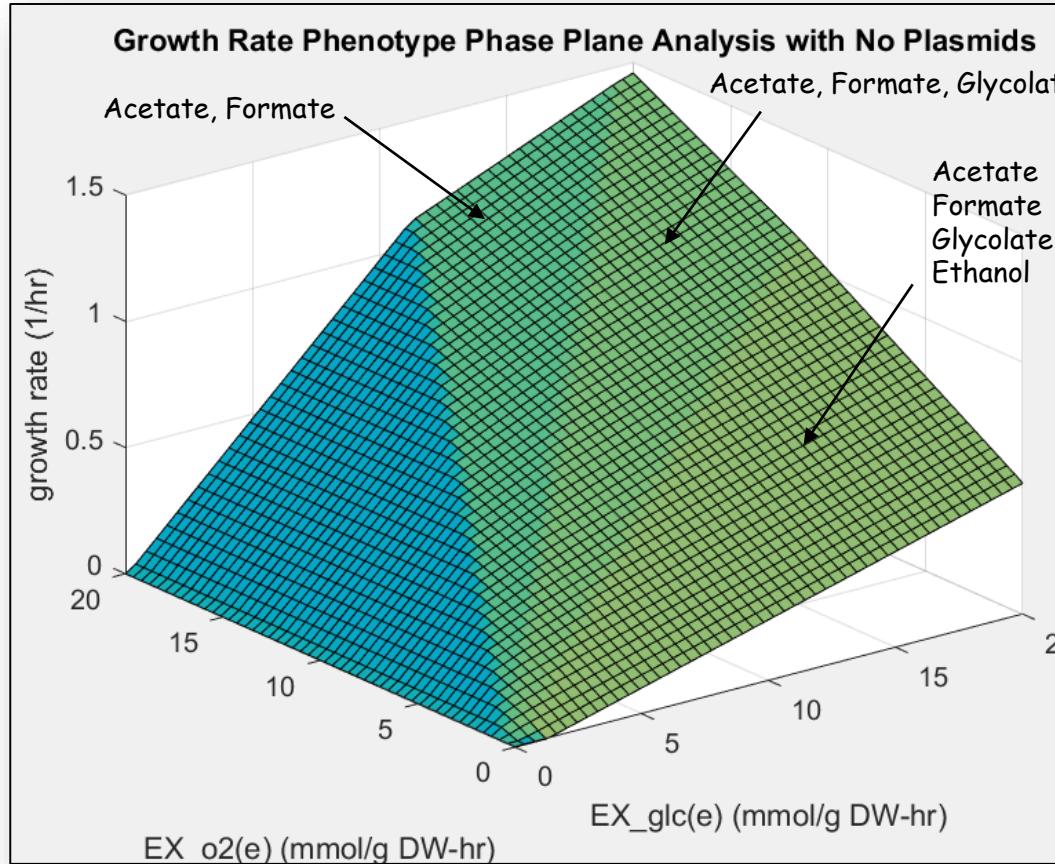


MaSp2\_PPP\_2Plasmid\_iECD\_1391.m



# Metabolic Load of Plasmids - Phenotype Phase Plane

(Without Spider Silk [MaSp2] Production)



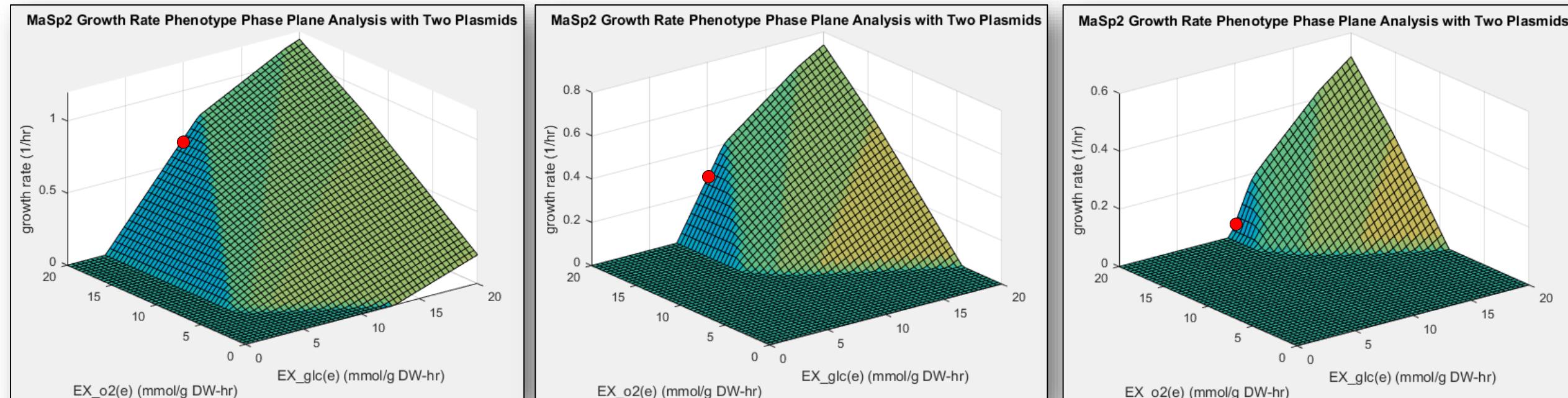
MaSp2\_PPP\_2Plasmid\_iECD\_1391.m



# Metabolic Load - Cell Growth Rate

(2 Plasmid System with Glucose)

Note the impact of increased flux through the MaSp2 reaction (larger copy number and stronger promoters)



MaSp2 > 0.00233 mmol/gDW·h  
(275 mg/gDW·h)\*

MaSp2 > 0.006 mmol/gDW·h  
(708 mg/gDW·h)

MaSp2 > 0.008 mmol/gDW·h  
(944 mg/gDW·h)

\*Protein production = Flux-rate · MW

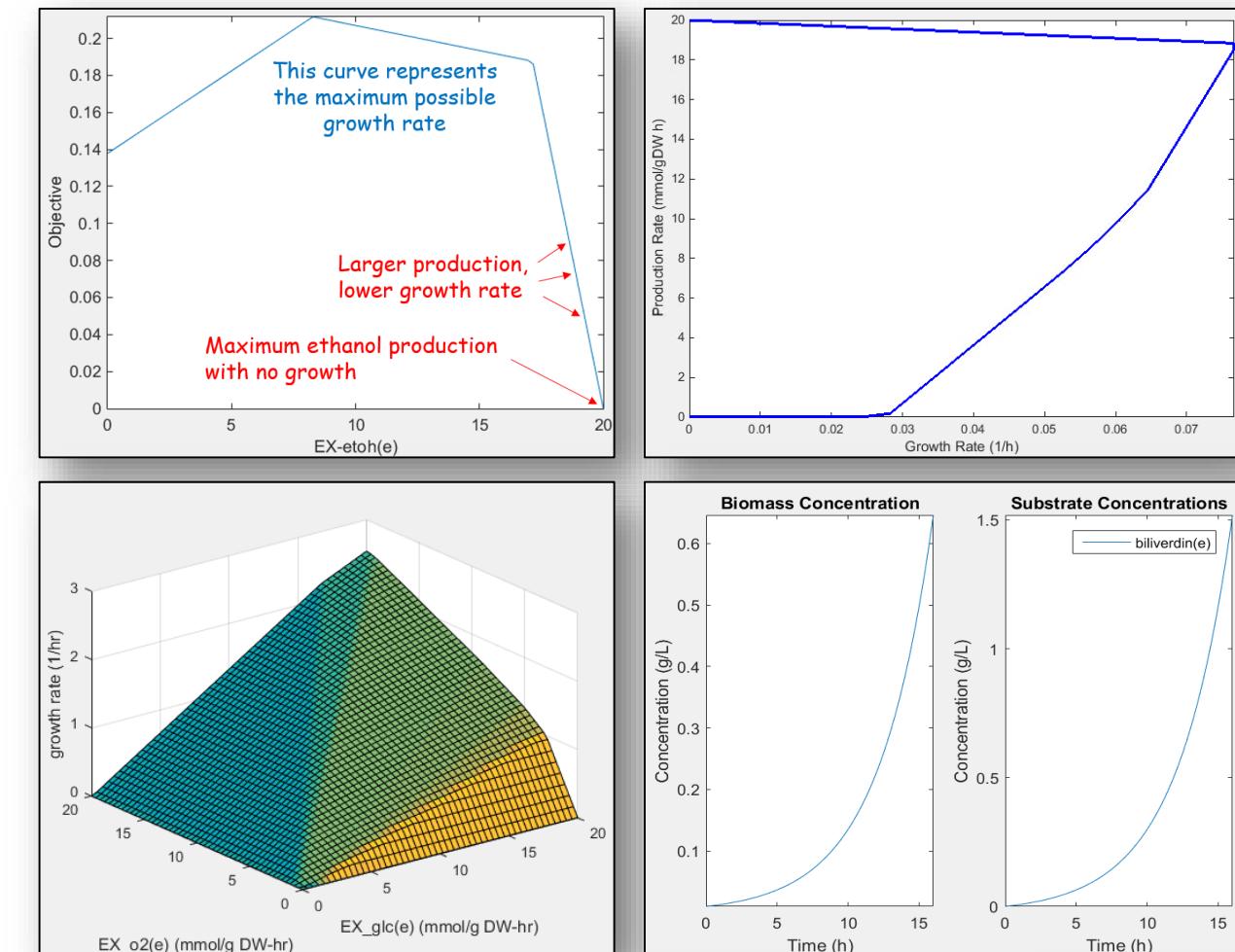
MaSp2\_PPP\_2Plasmid\_iECD\_1391.m

EX\_glc(e) ≥ -10; EX\_o2(e) ≥ -20;



# Cobra Tools for Analyzing the Plasmid Impact

- Robustness Analysis - Determines the change in fluxes for any reaction in both the plasmid and plasmid-free strains.
- Production Envelopes - Determines the bioproduction capabilities of a strain that includes the plasmid.
- Phenotype Phase Plane Analysis - Determines the phenotype phase planes for the plasmid included strain.
- Dynamic Flux Balance Analysis - Determines the upper limit of bioproduction product production (g/L) for the plasmid included strain.





# Bioproduction Production

- Bioproduction Identification
- Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





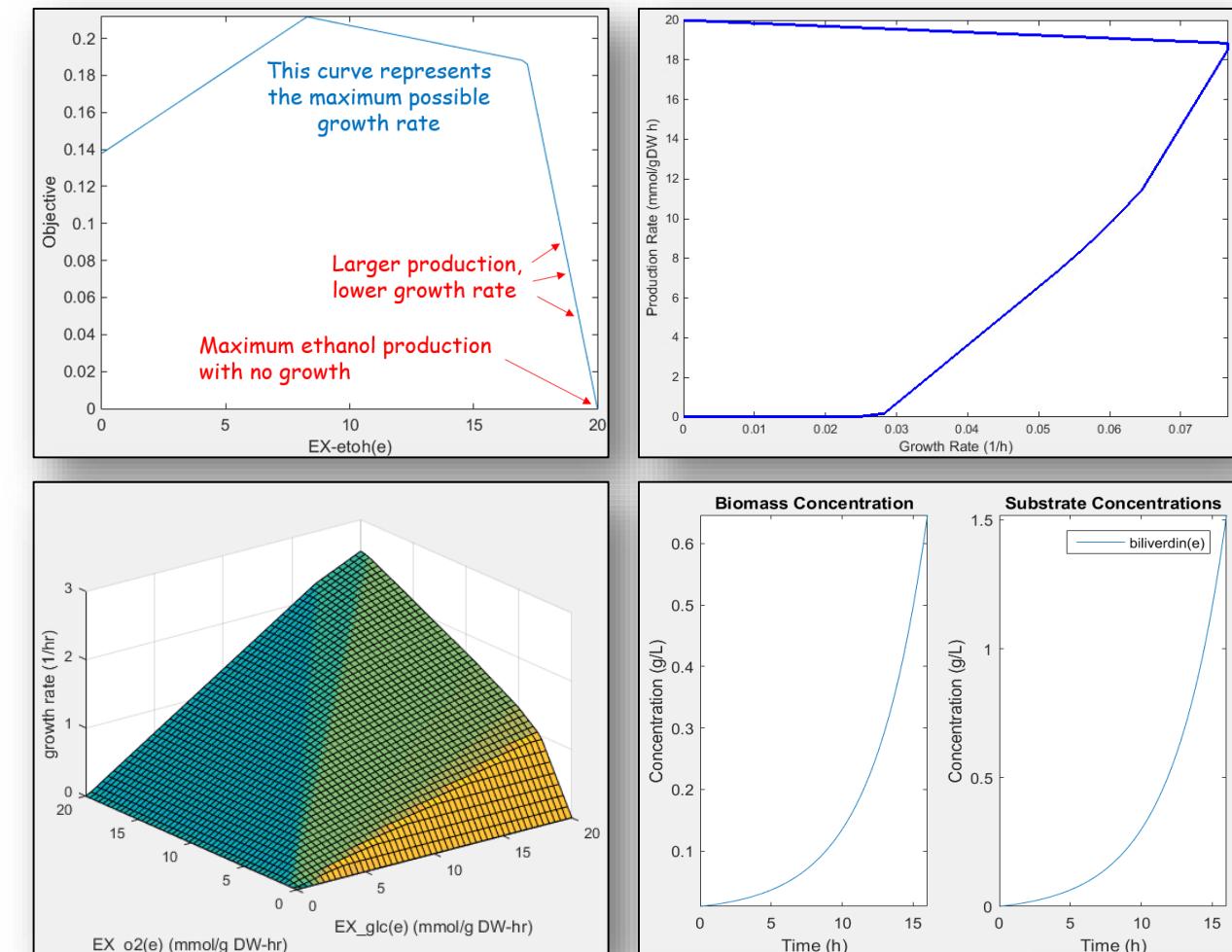
# Practical Production Limits

- Maximum biomass density
  - ✓ An upper limit for cell concentration in a large scale fermenter is less than 100 gDW/L (OD<sub>600</sub> = 47)
    - Choi, J. H., K. C. Keum, et al. (2006). "Production of recombinant proteins by high cell density culture of *Escherichia coli*." *Chemical Engineering Science* 61(3): 876-885
    - Shiloach, J. and R. Fass (2005). "Growing *E. coli* to high cell density--a historical perspective on method development." *Biotechnol Adv* 23(5): 345-357
- Protein production limit (Total protein = 550 mg per gram dry weight biomass)
  - ✓ An upper limit for the production of a recombinant protein is less than 50% of the total protein (275 mg/gDW)
    - Rosano, G. L. and E. A. Ceccarelli (2014). "Recombinant protein expression in *Escherichia coli*: advances and challenges." *Front Microbiol* 5: 172.
- *E. coli* growth is inhibited when the following nutrients are present above certain concentrations (shown in brackets)
  - ✓ Glucose (50 g/L), ammonia (3 g/L), iron (1.15 g/L), magnesium (8.7 g/L), phosphorous (10 g/L) and zinc (0.038 g/L), acetate (5 g/L), ethanol (47.3 g/L)
    - Lee, S. Y. (1996). "High cell-density culture of *Escherichia coli*." *Trends Biotechnol* 14(3): 98-105.
    - Riesenbergs, D. (1991). "High-cell-density cultivation of *Escherichia coli*." *Current opinion in biotechnology* 2(3): 380-384.
    - Goodarzi, H. (2010). "Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*." *Molecular Systems Biology*, 6, 378



# Cobra Tools for Analyzing Maximum Production

- Robustness Analysis - Determines range of fluxes for any reaction in the designed strain, including the bioproduct pathway reactions, as a function of growth rate.
- Production Envelopes - Determines the bioproduct production capabilities of a designed strain as a function of growth rate.
- Phenotype Phase Plane Analysis - Determines the phenotype phase planes for the designed strain.
- Dynamic Flux Balance Analysis - Determines the upper limit of bioproduct product production (g/L) over time.





# Bioproduction Production

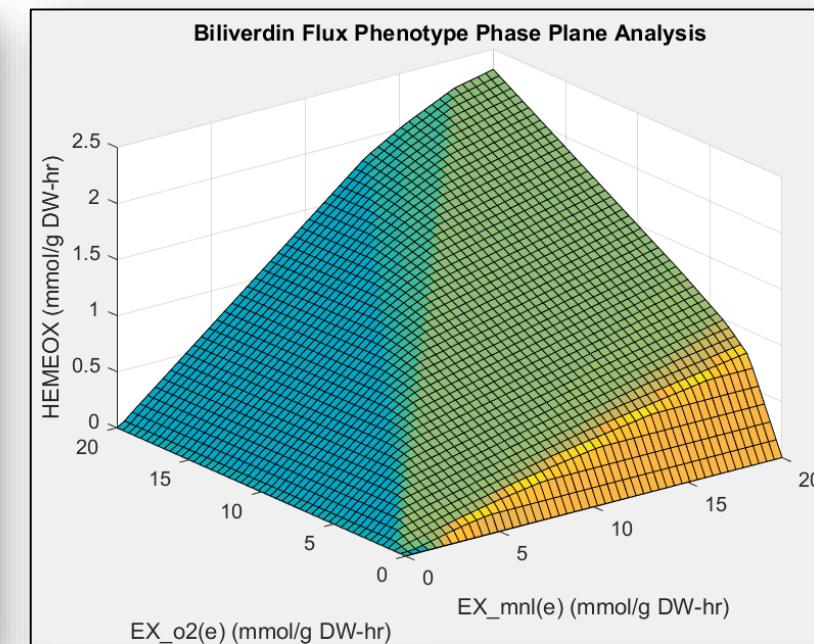
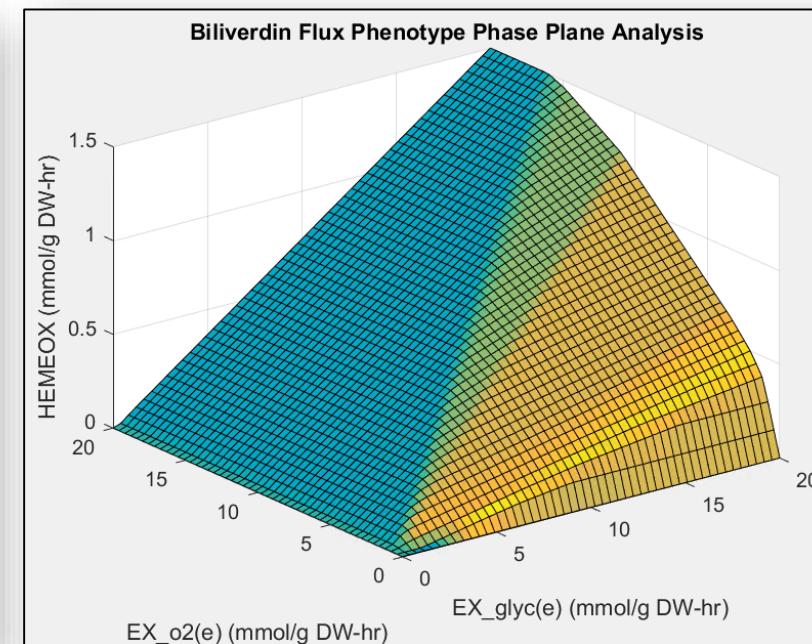
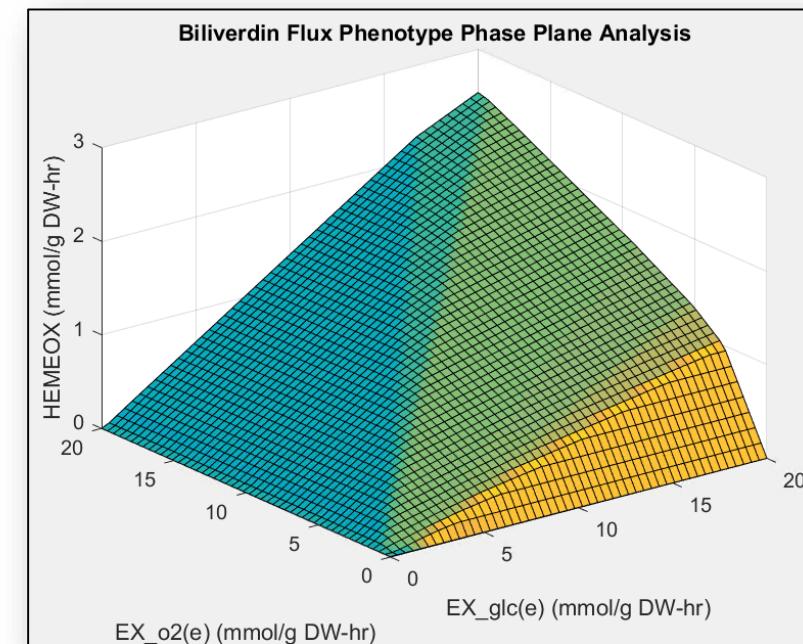
- Bioproduction Identification
- Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ➡ ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





# Maximum Biliverdin Production by Carbon Source

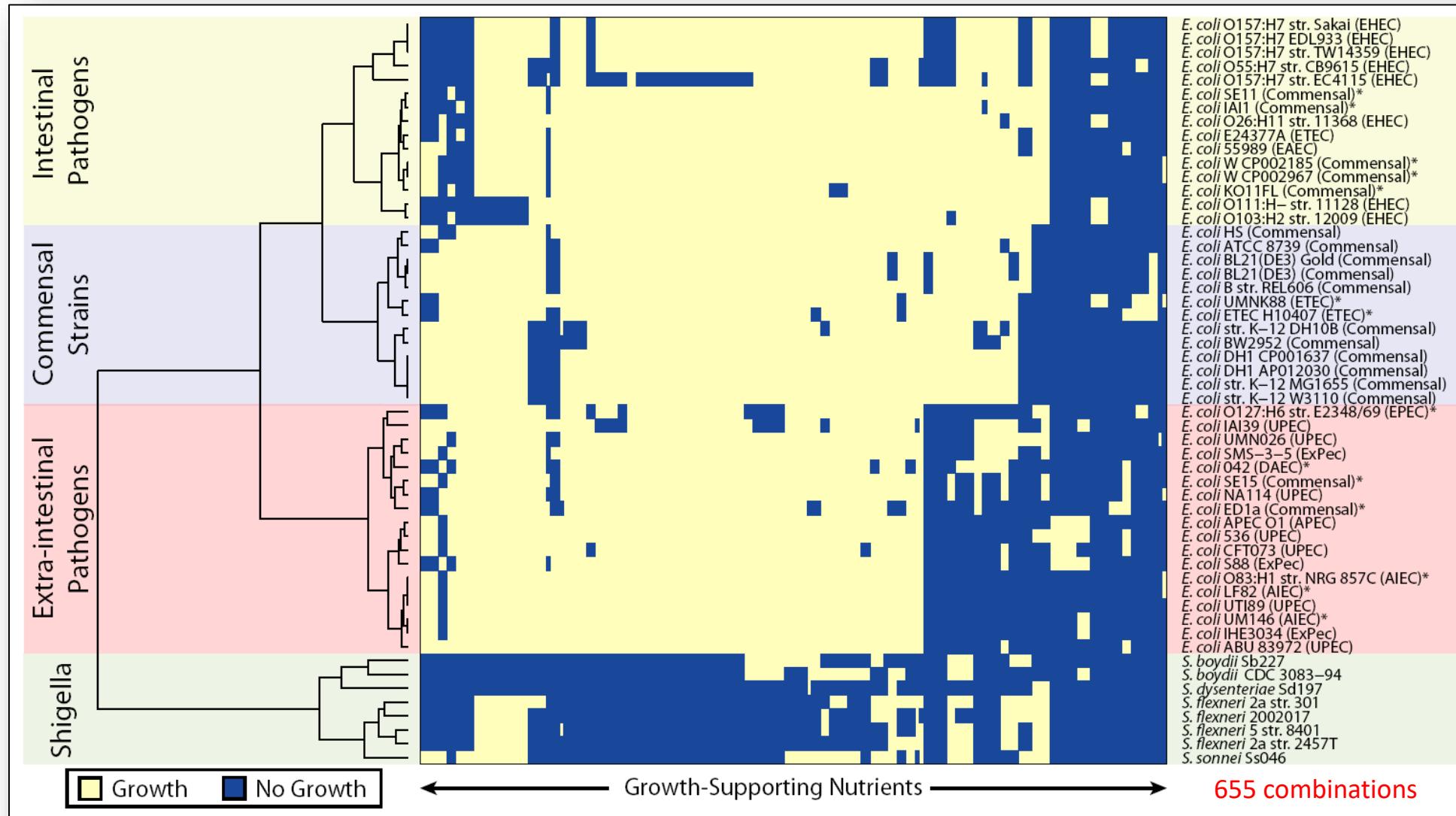
$\text{EX\_Carbon\_Source} = 20 \text{ mmol/gDW-h}$ ,  $\text{EX\_o2(e)} = 20 \text{ mmol/gDW-h}$



Biliverdin\_Production\_PPP\_BL21.m



# Clustering of Species by Unique Growth-supporting Conditions



Monk, J. M., P. Charusanti, et al. (2013). Proceedings of the National Academy of Sciences of the United States of America 110(50): 20338-20343.



# Carbon Sources for iJO1366 Model

180/285 are Growth Supporting

12ppd-R[e]	LalaDgluMdapDala[e]	amp[e]	cpgn-un[e]	dump[e]	fusa[e]	gln-L[e]	ile-L[e]	manglyc[e]	pnto-R[e]	sulfac[e]	ump[e]
12ppd-S[e]	LalaLglu[e]	anhgm[e]	crn[e]	duri[e]	g1p[e]	glu-L[e]	imp[e]	melib[e]	ppa[e]	tartr-D[e]	ura[e]
14glucan[e]	ac[e]	arab-L[e]	crn-D[e]	eca4colipa[e]	g3pc[e]	gly[e]	indole[e]	meoh[e]	ppal[e]	tartr-L[e]	urea[e]
15dap[e]	acac[e]	arbte[e]	csn[e]	enlipa[e]	g3pe[e]	glyald[e]	inost[e]	met-D[e]	pppn[e]	taur[e]	uri[e]
23camp[e]	acald[e]	arbtn[e]	cyan[e]	enter[e]	g3pg[e]	glyb[e]	ins[e]	met-L[e]	pro-L[e]	tcynt[e]	val-L[e]
23ccmp[e]	acgal[e]	arbtn-fe3[e]	cynt[e]	etha[e]	g3pi[e]	glyc[e]	isetac[e]	metsox-R-L[e]	progly[e]	thm[e]	xan[e]
23cgmp[e]	acgal1p[e]	arg-L[e]	cys-D[e]	ethso3[e]	g3ps[e]	glyc-R[e]	kdo2lipid4[e]	metsox-S-L[e]	psclys[e]	thr-L[e]	xmp[e]
23cump[e]	acgam[e]	ascb-L[e]	cys-L[e]	etoh[e]	g6p[e]	glyc2p[e]	lac-D[e]	mincyc[e]	pser-L[e]	thrp[e]	xtsn[e]
23dappa[e]	acgam1p[e]	asn-L[e]	cytd[e]	f6p[e]	gal[e]	glyc3p[e]	lac-L[e]	minohp[e]	ptrc[e]	thym[e]	xyl-D[e]
26dap-M[e]	acmana[e]	asp-L[e]	dad-2[e]	fald[e]	gal-bD[e]	glyclt[e]	lcts[e]	mmet[e]	pydam[e]	thymd[e]	xylu-L[e]
2ddglcn[e]	acmum[e]	btn[e]	damp[e]	fe3dcit[e]	gal1p[e]	gmp[e]	leu-L[e]	mnl[e]	pydx[e]	tma[e]	
34dhpac[e]	acnam[e]	but[e]	dca[e]	fe3dhbz[s]e	galct-D[e]	gsn[e]	lipa[e]	mso3[e]	pydxn[e]	tmao[e]	
3amp[e]	acolipa[e]	butso3[e]	dcmp[e]	fe3hox[e]	galctn-D[e]	gthox[e]	lipa_cold[e]	nac[e]	pyr[e]	tre[e]	
3cmp[e]	acser[e]	cbi[e]	dcyt[e]	fe3hox-un[e]	galctn-L[e]	gthrd[e]	lipoate[e]	nmn[e]	quin[e]	trp-L[e]	
3gmp[e]	ade[e]	cbl1[e]	ddca[e]	fecrm[e]	galt[e]	gtp[e]	lys-L[e]	novbcn[e]	r5p[e]	ttdca[e]	
3hcinnm[e]	adn[e]	cgly[e]	dgmp[e]	fecrm-un[e]	galur[e]	gua[e]	lyx-L[e]	o16a4colipa[e]	rfamp[e]	ttdcea[e]	
3hpp[e]	adocbl[e]	chol[e]	dgsn[e]	feenter[e]	gam[e]	hacolipa[e]	mal-D[e]	ocdca[e]	rib-D[e]	ttrcyc[e]	
3hpppn[e]	agm[e]	chtbs[e]	dha[e]	feoxam[e]	gam6p[e]	halipa[e]	mal-L[e]	ocdcea[e]	rmn[e]	tym[e]	
3ump[e]	akg[e]	cit[e]	dimp[e]	feoxam-un[e]	gbbtn[e]	hdca[e]	malt[e]	octa[e]	sbt-D[e]	tyr-L[e]	
4abut[e]	ala-B[e]	cm[e]	din[e]	for[e]	gdp[e]	hdcea[e]	malthx[e]	orn[e]	ser-D[e]	tyrp[e]	
4hoxpacd[e]	ala-D[e]	cmp[e]	dms[e]	fru[e]	glc-D[e]	his-L[e]	maltp[e]	orot[e]	ser-L[e]	uacgam[e]	
5dglcn[e]	ala-L[e]	co2[e]	dmsso[e]	frulys[e]	glcn[e]	hom-L[e]	malttr[e]	pacald[e]	skm[e]	udpacgal[e]	
5mtr[e]	alaala[e]	colipa[e]	dopa[e]	fruur[e]	glcr[e]	hx[e]	malttr[e]	peamn[e]	spmd[e]	udpg[e]	
LalaDglu[e]	all-D[e]	colipap[e]	doxrbcn[e]	fuc-L[e]	glcur[e]	hxan[e]	man[e]	phe-L[e]	succ[e]	udpgal[e]	
LalaDgluMdap[e]	alltn[e]	cpgn[e]	dtmp[e]	fum[e]	glcur1p[e]	idon-L[e]	man6p[e]	pheme[e]	sucr[e]	udpglcur[e]	

Orth, J. D., T. M. Conrad, et al. (2011). "A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011." Molecular Systems Biology 7: 535.



# Nitrogen Sources for iJO1366 Model

94/178 are Growth Supporting

15dap[e]	ade[e]	colipap[e]	enlipa[e]	gthox[e]	nac[e]	spmd[e]	xan[e]
23camp[e]	adn[e]	cpgn[e]	enter[e]	gthrd[e]	<b>nh4[e]</b>	taur[e]	xmp[e]
23ccmp[e]	adocbl[e]	cpgn-un[e]	etha[e]	gtp[e]	nmn[e]	tcynt[e]	xtsn[e]
23cgmp[e]	agm[e]	crn[e]	fe3dhbzs[e]	gua[e]	no[e]	thm[e]	
23cump[e]	ala-B[e]	crn-D[e]	fe3hox[e]	hacolipa[e]	no2[e]	thr-L[e]	
23dappa[e]	ala-D[e]	csn[e]	fe3hox-un[e]	halipa[e]	no3[e]	thrp[e]	
26dap-M[e]	ala-L[e]	cyan[e]	fecrm[e]	his-L[e]	novbcn[e]	thym[e]	
3amp[e]	alaala[e]	cynt[e]	fecrm-un[e]	hom-L[e]	o16a4colipa[e]	thymd[e]	
3cmp[e]	alltn[e]	cys-D[e]	feenter[e]	hxan[e]	orn[e]	tma[e]	
3gmp[e]	amp[e]	cys-L[e]	feoxam[e]	ile-L[e]	orot[e]	tmao[e]	
3ump[e]	anhgm[e]	cytd[e]	feoxam-un[e]	imp[e]	peamn[e]	trp-L[e]	
4abut[e]	arbtn[e]	dad-2[e]	frulys[e]	indole[e]	phe-L[e]	ttrcyc[e]	
LalaDglu[e]	arbtn-fe3[e]	damp[e]	g3pc[e]	ins[e]	pheme[e]	tym[e]	
LalaDgluMdap[e]	arg-L[e]	dcmp[e]	g3pe[e]	kdo2lipid4[e]	pnto-R[e]	tyr-L[e]	
LalaDgluMdapDala[e]	asn-L[e]	dcyt[e]	g3ps[e]	leu-L[e]	pro-L[e]	tyrp[e]	
LalaLglu[e]	asp-L[e]	dgmp[e]	gam[e]	lipa[e]	progly[e]	uacgam[e]	
acgal[e]	btn[e]	dgsn[e]	gam6p[e]	lipa_cold[e]	psclys[e]	udpacgal[e]	
acgal1p[e]	cbi[e]	dimp[e]	gbbtn[e]	lys-L[e]	pser-L[e]	udpg[e]	
acgam[e]	cbl1[e]	din[e]	gdp[e]	met-D[e]	ptrc[e]	udpgal[e]	
acgam1p[e]	cgly[e]	dopa[e]	gln-L[e]	met-L[e]	pydam[e]	udpglcur[e]	
acmana[e]	chol[e]	doxrbcn[e]	glu-L[e]	metsox-R-L[e]	pydx[e]	ump[e]	
acmum[e]	chtbs[e]	dtmp[e]	gly[e]	metsox-S-L[e]	pydxn[e]	ura[e]	
acnam[e]	cm[e]	dump[e]	glyb[e]	mincyc[e]	rfamp[e]	urea[e]	
acolipa[e]	cmp[e]	duri[e]	gmp[e]	mmet[e]	ser-D[e]	uri[e]	
acs[e]	colipa[e]	eca4colipa[e]	gsn[e]	n2o[e]	ser-L[e]	val-L[e]	

Orth, J. D., T. M. Conrad, et al. (2011). "A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011." Molecular Systems Biology 7: 535.



# Phosphorus & Sulfur Sources for iJO1366 Model

Phosphorous Sources			Sulfur Sources	
acgal1p[e]	gal1p[e]	udpacgal[e]	5mtr[e]	tcynt[e]
acgam1p[e]	gam6p[e]	udpg[e]	btn[e]	thm[e]
acolipa[e]	gdp[e]	udpgal[e]	butso3[e]	tsul[e]
adocbl[e]	glcur1p[e]	udpglcru[e]	cgly[e]	
amp[e]	glyc2p[e]	ump[e]	cys-D[e]	
cbl1[e]	glyc3p[e]	xmp[e]	cys-L[e]	
cmp[e]	gmp[e]		dms[e]	
colipa[e]	gtp[e]		dmso[e]	
colipap[e]	hacolipa[e]		ethso3[e]	
damp[e]	halipa[e]		gthox[e]	
dcmp[e]	imp[e]		gthrd[e]	
dgmp[e]	kdo2lipid4[e]		h2s[e]	
dimp[e]	lipa[e]		isetac[e]	
dtmp[e]	lipa_cold[e]		lipoate[e]	
dump[e]	man6p[e]		met-D[e]	
eca4colipa[e]	minohp[e]		met-L[e]	
enlipa[e]	nmn[e]		metsox-R-L[e]	
f6p[e]	o16a4colipa[e]		metsox-S-L[e]	
g1p[e]	pi[e]		mmet[e]	
g3pc[e]	ppt[e]		mso3[e]	
g3pe[e]	pser-L[e]		so2[e]	
g3pg[e]	r5p[e]		so3[e]	
g3pi[e]	thrp[e]		so4[e]	
g3ps[e]	tyrp[e]		sulfac[e]	
g6p[e]	uacgam[e]		taur[e]	

Orth, J. D., T. M. Conrad, et al. (2011). "A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011." Molecular Systems Biology 7: 535.



# Catabolite Repression

- If the carbon source is glucose in an aerobic environment a more accurate model would block the reactions that would naturally be turned off as a result catabolite repression.
- For the iaf1260 model the reactions blocked are given to the right.

## iaf1260 reactions block due to catabolite repression

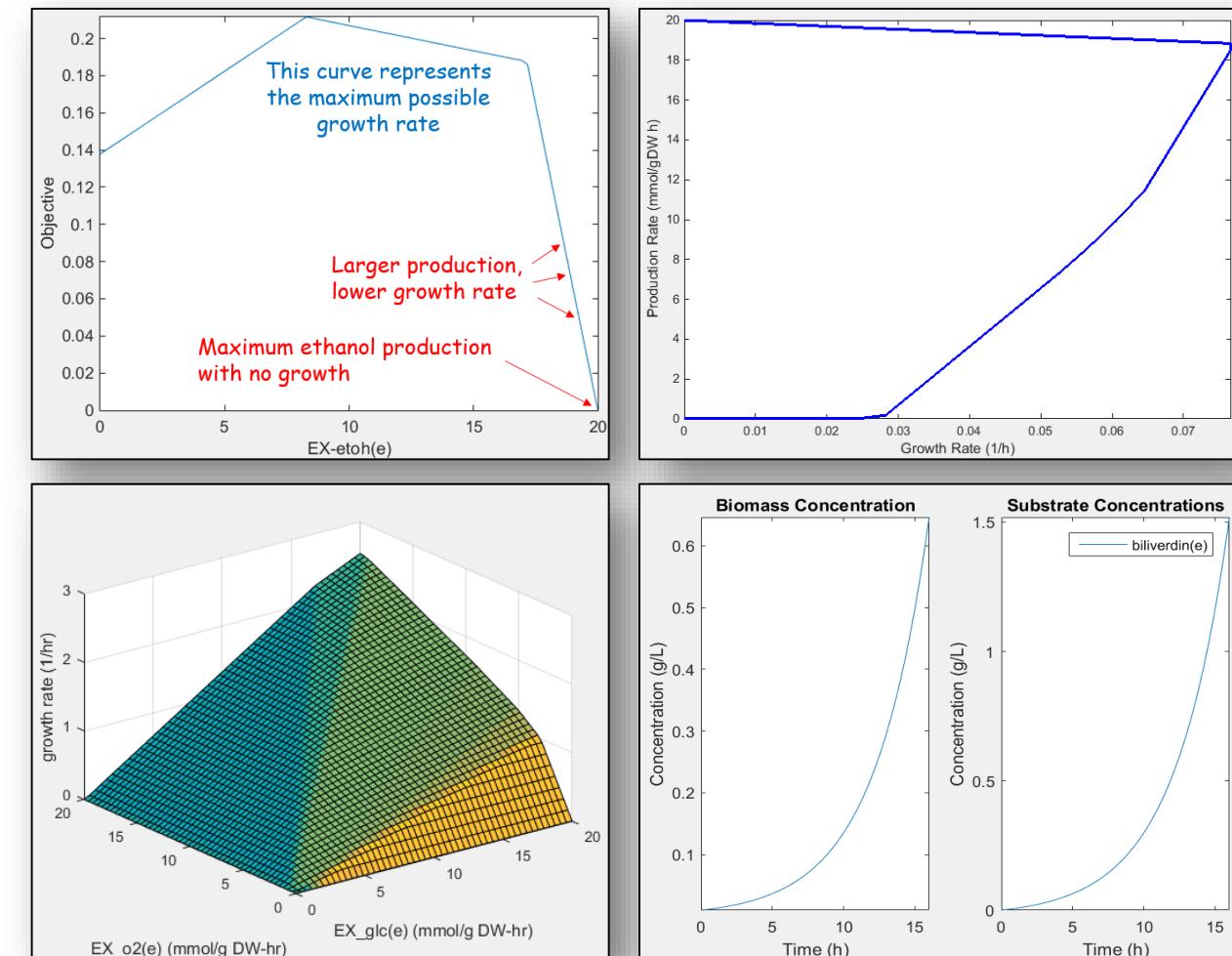
14GLUCANabcpp	ARBt2rpp	DCYTD	FCLK	G3PSabcpp	GPDDA5pp	KAT3	MALTHXabcpp	OROTt2_2pp	RNTR3c	XYL12
14GLUCANTexi	ASNNpp	DMSOR1	FORt2pp	G6Pt6_2pp	HACD1i	KAT4	MALTHXtexi	P5CD	RNTR4c	XYLK
ACOAD1f	ASPT	DMSOR2	FORppi	GALabcpp	HACD2i	KAT5	MALPTAbcpp	PFL	SBTPD	
ACOAD2f	ASPt2_2pp	DOGULNR	FRD2	GALS3	HACD3i	KAT6	MALTptspp	PPAt4pp	SBTptspp	
ACOAD3f	CADVtpp	DXYLK	FRD3	GAM6Pt6_2pp	HACD4i	KAT7	MALPTtexi	PROD2	SERD_D	
ACOAD4f	CRNBCT	ECOAH1	FRUUrt2rpp	GLCabcpp	HACD5i	KAT8	MALTtexi	PROt4pp	SUCCt2_2pp	
ACOAD5f	CRNCAR	ECOAH2	FUCtpp	GLCtexi	HACD6i	LACZ	MALTRRabcpp	RBK	TMAOR1	
ACOAD6f	CRNCBCT	ECOAH3	FUMt2_2pp	GLUNpp	HACD7i	LCARS	MALTRRtexi	RBK_L1	TMAOR2	
ACOAD7f	CRNCDH	ECOAH4	G3PCabcpp	GLYC3Pabcpp	HACD8i	LCTStpp	MALTTTRabcpp	RBP4E	TMDPP	
ACOAD8f	CRNt7pp	ECOAH5	G3PD5	GLYC3Pt6pp	HYD1pp	LYXI	MALTTTRtexi	RMI	TRE6PH	
ACS	CRNt8pp	ECOAH6	G3PD6	GLYK	HYD2pp	LYXt2pp	MAN6Pt6_2pp	RMK	TREptspp	
ALDD2y	CYANST	ECOAH7	G3PD7	GPDDA1pp	HYD3pp	MALDt2_2pp	MELIBt2pp	RMNtpp	TRPAS2	
ALDD3y	CYSDS	ECOAH8	G3PEabcpp	GPDDA2pp	ICL	MALS	NO2t2rpp	RMPA	TRPt2rpp	
ARAI	CYTD	F6Pt6_2pp	G3PGabcpp	GPDDA3pp	KAT1	MALT2_2pp	NTRIR2x	RNTR1c	XYLabcpp	
ARBabcpp	DAAD	FCI	G3Plabcpp	GPDDA4pp	KAT2	MALTabcpp	OBTFL	RNTR2c	XYL11	

Feist, A. M., C. S. Henry, et al. (2007). "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." Molecular Systems Biology 3: 121.



# Cobra Tools for Analyzing Carbon Sources

- Robustness Analysis - Determines range of fluxes for any reaction in the designed strain, including the bioproduct pathway reactions, as a function of growth rate.
- Production Envelopes - Determines the bioproduct production capabilities of a designed strain as a function of growth rate.
- Phenotype Phase Plane Analysis - Determines the phenotype phase planes, shadow prices, and reduced costs for the designed strain.
- Dynamic Flux Balance Analysis - Determines the upper limit of bioproduct product production (g/L) over time.





# Bioproduction Production

- Bioproduction Identification
- Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





# Media Comparison

- Minimal media are those that contain the minimum nutrients possible for colony growth, generally without the presence of amino acids, and are often used to grow "wild type" microorganisms. Minimal medium typically contains: a carbon source for bacterial growth, which may be a sugar such as glucose, water, and various salts which may vary among bacteria species and growing conditions; these generally provide essential elements such as magnesium, nitrogen, phosphorus, and sulfur to allow the bacteria to synthesize protein and nucleic acids.
- LB (Lysogeny Broth) is the most commonly used medium for culturing *E. coli*. It is easy to make, it has rich nutrient contents and its osmolarity is optimal for growth at early log phase. All these features make it adequate for protein production and compensate for the fact that it is not the best option for achieving high cell density cultures. Despite being a rich broth, cell growth stops at a relatively low density.
- An undefined medium (also known as a basal or complex medium) includes a carbon source such as glucose for bacterial growth, water, various salts needed for bacterial growth, a source of amino acids and nitrogen (e.g., beef, yeast extract). This is an undefined medium because the amino acid source contains a variety of compounds with the exact composition being unknown.
- In autoinduction media, a mixture of glucose, lactose, and glycerol is used in an optimized blend. Glucose is the preferred carbon source and is metabolized preferentially during growth, which prevents uptake of lactose until glucose is depleted, usually in mid to late log phase. Consumption of glycerol and lactose follows, the latter being also the inducer of lac-controlled protein expression. In this way, biomass monitoring for timely inducer addition is avoided, as well as culture manipulation

Rosado, G. L. and E. A. Ceccarelli (2014). "Recombinant protein expression in *Escherichia coli*: advances and challenges." Frontiers in microbiology 5: 172.



# M9 Minimal Medium

- One liter of M9 medium (Sigma catalog no. 6030) contains:
  - ✓  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$  (6.8g),  $\text{KH}_2\text{PO}_4$  (3g),  $\text{NaCl}$  (0.5g),  $\text{NH}_4\text{Cl}$  (1g),  $\text{MgSO}_4$  (2 mM),  $\text{CaCl}_2$  (0.1 mM)
- Growth on minimal medium was simulated by maximizing flux through a defined biomass objective function and allowing the uptake of
  - ✓  $\text{NH}_4$ ,  $\text{SO}_4$ ,  $\text{O}_2$ , and  $\text{P}_i$  and the free exchange of  $\text{H}^+$ ,  $\text{H}_2\text{O}$ , and  $\text{CO}_2$
- All exchange reaction lower constraints, except the following, should be greater than zero
  - ✓  $-1000 \leq \text{NH}_4, \text{SO}_4, \text{O}_2, \text{and P}_i \leq 0$
  - ✓  $-1000 \leq \text{H}^+, \text{H}_2\text{O}, \text{and CO}_2 \leq 1000$
  - ✓  $-1000 \leq \text{Carbon source} \leq 0$
  - ✓ Use the following commands to change the constraints
    - `model = changeRxnBounds(model, 'EX_xxx(e)', -1000, 'l')`
    - `model = changeRxnBounds(model, 'EX_xxx(e)', 1000, 'u')`
- ✓ Verify that no other metabolites are allowed to be uptaken
  - No other metabolites should have a negative lower constraint
  - Check using the "printConstraints(model, -1001, 1)" command



# Recipe for M-9 Minimal Media

- 5X M9 basis
  - $\text{Na}_2\text{HPO}_4 \cdot 12 \text{H}_2\text{O}$  85.7 g
  - $\text{KH}_2\text{PO}_4$  15.0 g
  - $\text{NaCl}$  2.5 g
  - Dissolve above components in 1000 ml of milli-Q and autoclave
- 5 g  $(\text{NH}_4)_2\text{SO}_4$  in 15 ml of H<sub>2</sub>O
- Trace elements
  - 1 g EDTA
  - 29 mg  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$
  - 198 mg  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$
  - 254 mg  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$
  - 13.4 mg  $\text{CuCl}_2$
  - 147 mg  $\text{CaCl}_2$
  - Dissolve in 100 ml of milli-Q and autoclave
- 20% (w/v) glucose: 25 g in 100 ml of milliQ and filter with 0.22 micron filter
- 0.1 M  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ : 1.47 g in 100 ml milliQ and filter with 0.22 micron filter
- 1M  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ : 24.65 g in 100 ml milliQ and filter with 0.22 micron filter
- 10 mM  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ : 140 mg in 50 ml of milliQ (prepare fresh)
- 1% thiamine: 500mg in 10 ml of milliQ (prepare fresh)
- Proportions for 1 liter M-9 media
  - 200 ml of M-9 basis; 3 ml of  $(\text{NH}_4)_2\text{SO}_4$ ; 1 ml of  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ ; 1 ml trace elements; 20 ml glucose; 1ml  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ ; 1 ml  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$  2ml thiamine; 1ml antibiotic (standard conc.)



# *in silico* M9 Minimal Media

This *in silico* media assumes the cell can uptake all the minerals wanted/needed from the media. It does not allow amino acid uptake.

Monk, J. M., P. Charusanti, et al. (2013). "Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments." Proc Natl Acad Sci USA 110(50): 20338-20343.

Reaction Abbreviation	Reaction Name	Formula	Lower Bound	Upper Bound
EX_ca2(e)	Calcium exchange	ca2[e] <=>	-1000	1000
EX_cl(e)	Chloride exchange	cl[e] <=>	-1000	1000
EX_co2(e)	CO2 exchange	co2[e] <=>	-1000	1000
EX_cobalt2(e)	Co2+ exchange	cobalt2[e] <=>	-1000	1000
EX_cu2(e)	Cu2+ exchange	cu2[e] <=>	-1000	1000
EX_fe2(e)	Fe2+ exchange	fe2[e] <=>	-1000	1000
EX_fe3(e)	Fe3+ exchange	fe3[e] <=>	-1000	1000
EX_h(e)	H+ exchange	h[e] <=>	-1000	1000
EX_h2o(e)	H2O exchange	h2o[e] <=>	-1000	1000
EX_k(e)	K+ exchange	k[e] <=>	-1000	1000
EX_mg2(e)	Mg exchange	mg2[e] <=>	-1000	1000
EX_mn2(e)	Mn2+ exchange	mn2[e] <=>	-1000	1000
EX_mobd(e)	Molybdate exchange	mobd[e] <=>	-1000	1000
EX_na1(e)	Sodium exchange	na1[e] <=>	-1000	1000
EX_tungs(e)	tungstate exchange	tungs[e] <=>	-1000	1000
EX_zn2(e)	Zinc exchange	zn2[e] <=>	-1000	1000
EX_ni2(e)	Ni2+ exchange	ni2[e] <=>	-1000	1000
EX_sel(e)	Selenate exchange	sel[e] <=>	-1000	1000
EX_slnt(e)	selenite exchange	slnt[e] <=>	-1000	1000
EX_so4(e)	Sulfate exchange	so4[e] <=>	-1000	1000
EX_nh4(e)	Ammonia exchange	nh4[e] <=>	-1000	1000
EX_pi(e)	Phosphate exchange	pi[e] <=>	-1000	1000
EX_cbl1(e)	Cob(I)alamin exchange	cbl1[e] <=>	-0.01	1000



# Amino Acid Exchange Reactions

Note: Amino acids are only allowed to be secreted in the basic model ( $LB = 0$ ).  
The model can be modified to allow amino acid uptake.

No essential amino acids for *E.coli*

Rxn name	Rxn description	Formula	LB	UB
EX_ala_L(e)	L-Alanine exchange	ala-L[e] <=>	0	1000
EX_arg_L(e)	L-Arginine exchange	arg-L[e] <=>	0	1000
EX_asn_L(e)	L-Asparagine exchange	asn-L[e] <=>	0	1000
EX_asp_L(e)	L-Aspartate exchange	asp-L[e] <=>	0	1000
EX_cys_L(e)	L-Cysteine exchange	cys-L[e] <=>	0	1000
EX_gln_L(e)	L-Glutamine exchange	gln-L[e] <=>	0	1000
EX_glu_L(e)	L-Glutamate exchange	glu-L[e] <=>	0	1000
EX_gly(e)	Glycine exchange	gly[e] <=>	0	1000
EX_his_L(e)	L-Histidine exchange	his-L[e] <=>	0	1000
EX_ile_L(e)	L-Isoleucine exchange	ile-L[e] <=>	0	1000
EX_leu_L(e)	L-Leucine exchange	leu-L[e] <=>	0	1000
EX_lys_L(e)	L-Lysine exchange	lys-L[e] <=>	0	1000
EX_met_L(e)	L-Methionine exchange	met-L[e] <=>	0	1000
EX_phe_L(e)	L-Phenylalanine exchange	phe-L[e] <=>	0	1000
EX_pro_L(e)	L-Proline exchange	pro-L[e] <=>	0	1000
EX_ser_L(e)	L-Serine exchange	ser-L[e] <=>	0	1000
EX_thr_L(e)	L-Threonine exchange	thr-L[e] <=>	0	1000
EX_trp_L(e)	L-Tryptophan exchange	trp-L[e] <=>	0	1000
EX_tyr_L(e)	L-Tyrosine exchange	tyr-L[e] <=>	0	1000
EX_val_L(e)	L-Valine exchange	val-L[e] <=>	0	1000



# K-12 Undefined Media

- K-12 is an undefined media based on yeast extract and phosphates.
- Growth in K-12 media was simulated by adjusting lower bounds of exchange reactions to correspond to media conditions

	Chemical	Concentration
K12 Medium:	$\text{KH}_2\text{PO}_4$	2 g/L
	$\text{K}_2\text{HPO}_4 \cdot 3\text{H}_2\text{O}$	4 g/L
	$(\text{NH}_4)_2\text{HPO}_4$	5 g/L
	Yeast Extract	5 g/L
	Glucose	25 g/L
	$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	0.5 g/L
	Thiamine	2.5 mg/L
	K12 trace metal	5 ml/L
K12 trace metal solution:	NaCl	5 g/L
	$\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$	1 g/L
	$\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$	4 g/L
	$\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$	4.75 g/L
	$\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$	0.4 g/L
	$\text{H}_3\text{BO}_3$	0.575 g/L
	$\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$	0.5 g/L
	6N $\text{H}_2\text{SO}_4$	12.5 ml/L



# Assigned Metabolite Uptake Rates for K-12 Media

(Metabolite lower bound determined by initial concentration of metabolite times the ratio of the initial glucose concentration/lower bound)

Metabolite	MW (g/mol)	g/L in media	mmol/L in media	Lower bound (mmol gDW <sup>-1</sup> h <sup>-1</sup> )
Glucose	180.16	25	138.7655417	-11
Ammonium	18.03851	1.365829484	75.71742258	-6.002150375
Phosphate	94.9714	6.655736264	70.08147994	-5.555386948
Potassium	39.0983	1.945099309	49.74894838	-3.943619038
Sulfate	96.07	0.203323529	2.11641021	-0.167768684
Chloride	35.453	0.062050364	1.750214773	-0.138740225
Copper	63.546	0.000509009	0.008010093	-0.000634963
Iron (III)	55.845	0.00490684	0.087865335	-0.00696512
Magnesium	24.305	0.049304203	2.028562155	-0.160804933
Manganese	54.938044	0.005551956	0.101058487	-0.008010947
Molybdate	95.95	0.001826052	0.019031286	-0.001508618
Sodium	22.98976928	0.029775544	1.295164976	-0.102668246
Thiamine	265.35	0.0025	0.009421519	-0.000746848
Zinc	65.38	0.001136847	0.017388304	-0.001378378
Alanine	89.09	0.225	2.525535975	-0.200200247
Arginine	174.2	0.145	0.832376579	-0.065982824
Asparagine	132.12	0.16	1.211020285	-0.095998062

For more accurate results these uptake rates need to be measured

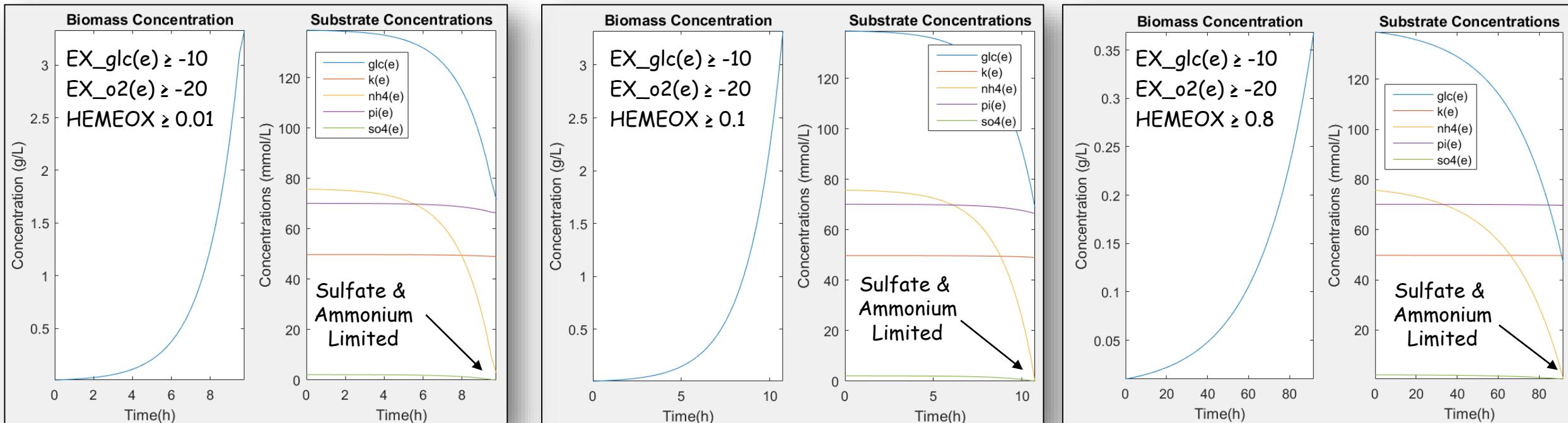
Metabolite	MW (g/mol)	g/L in media	mmol/L in media	Lower bound (mmol gDW <sup>-1</sup> h <sup>-1</sup> )
Aspartic acid	133.1	0.16	1.202103681	-0.09529124
Cysteine	121.16	0.02	0.165070981	-0.013085243
Glutamine	146.14	0.345	2.360749966	-0.187137594
Glutamic Acid	147.13	0.345	2.344865085	-0.185878393
Glycine	75.07	0.135	1.798321567	-0.14255367
Histidine	155.15	0.06	0.386722527	-0.030655649
Isoleucine	131.17	0.145	1.105435694	-0.08762833
Leucine	131.17	0.21	1.600975833	-0.126909995
Lysine	146.19	0.22	1.504890895	-0.119293303
Methionine	149.21	0.045	0.301588365	-0.02390703
Phenylalanine	165.19	0.12	0.726436225	-0.05758489
Proline	115.13	0.115	0.998870842	-0.079180891
Serine	105.09	0.13	1.237034922	-0.098060253
Threonine	119.12	0.135	1.133310947	-0.089838012
Tyrosine	181.19	0.09	0.496716154	-0.039374888
Valine	117.15	0.165	1.408450704	-0.111648451

Sarah Allred, "Metabolic Modeling of Spider Silk Production in *E. coli*," MS Thesis, USU, 2014



# Example: Impact of Essential Nutrients in the Media on Cell Growth

How are the essential nutrients impact the production process? They include: 'EX\_cl(e)', 'EX\_cu2(e)', 'EX\_fe3(e)', 'EX\_k(e)', 'EX\_mg2(e)', 'EX\_mn2(e)', 'EX\_mobd(e)', 'EX\_na1(e)', 'EX\_nh4(e)', 'EX\_pi(e)', 'EX\_so4(e)', 'EX\_zn2(e)'



DynamicBiliverdinProduction\_Nutrient\_Uptake\_BL21\_Plasmid.m



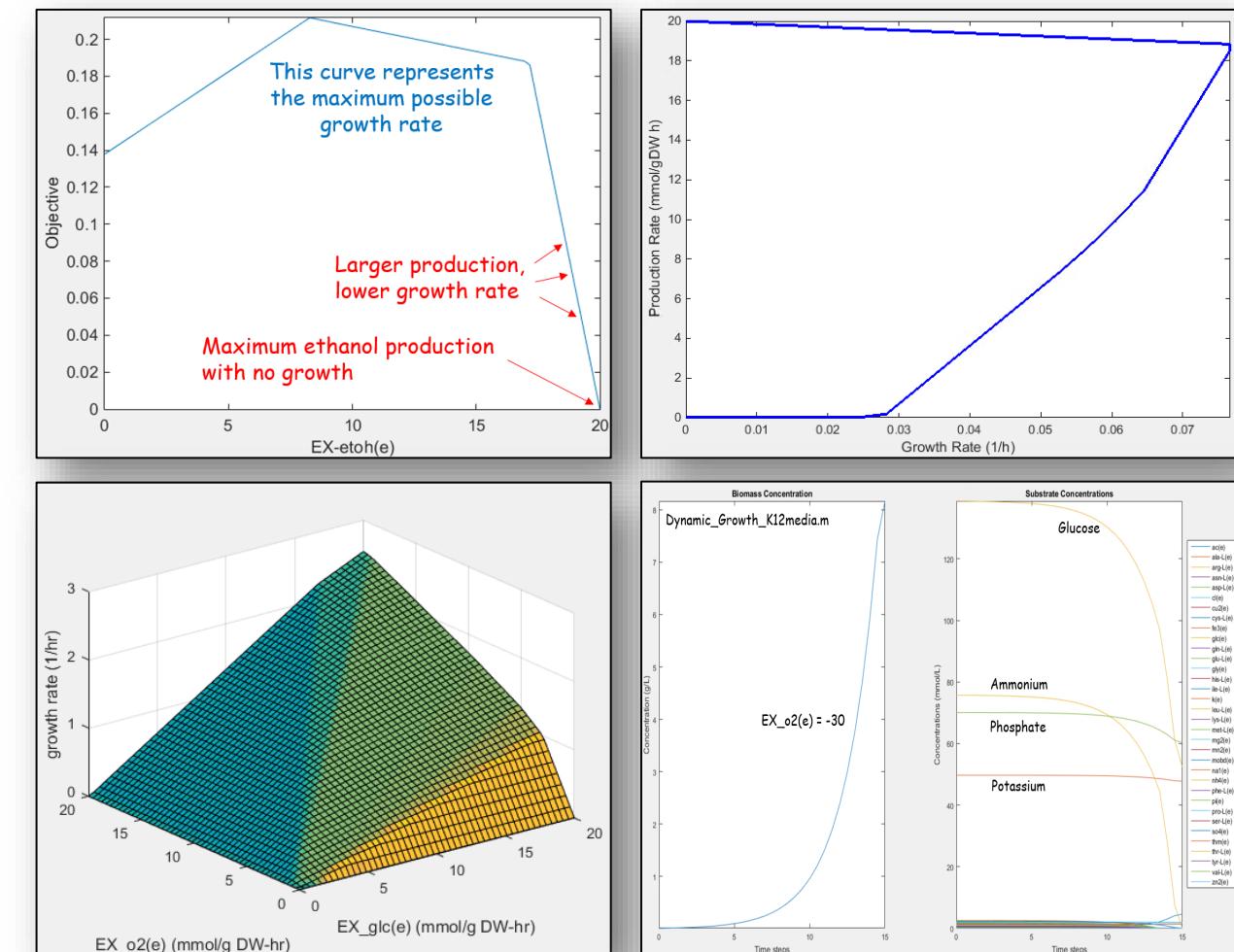
# Nutrient Limits in Production

- *E. coli* growth is inhibited when the following nutrients are present above certain concentrations (shown in brackets)
  - ✓ glucose (50 g/L),
  - ✓ ammonia (3 g/L),
  - ✓ iron (1.15 g/L),
  - ✓ magnesium (8.7 g/L),
  - ✓ phosphorous (10 g/L) and
  - ✓ zinc (0.038 g/L)
- References
  - ✓ Lee, S. Y. (1996). "High cell-density culture of *Escherichia coli*." *Trends Biotechnol* 14(3): 98-105.
  - ✓ Riesenbergs, D. (1991). "High-cell-density cultivation of *Escherichia coli*." *Current opinion in biotechnology* 2(3): 380-384.



# Cobra Tools for Analyzing Nutrient & Amino Acid Limitations

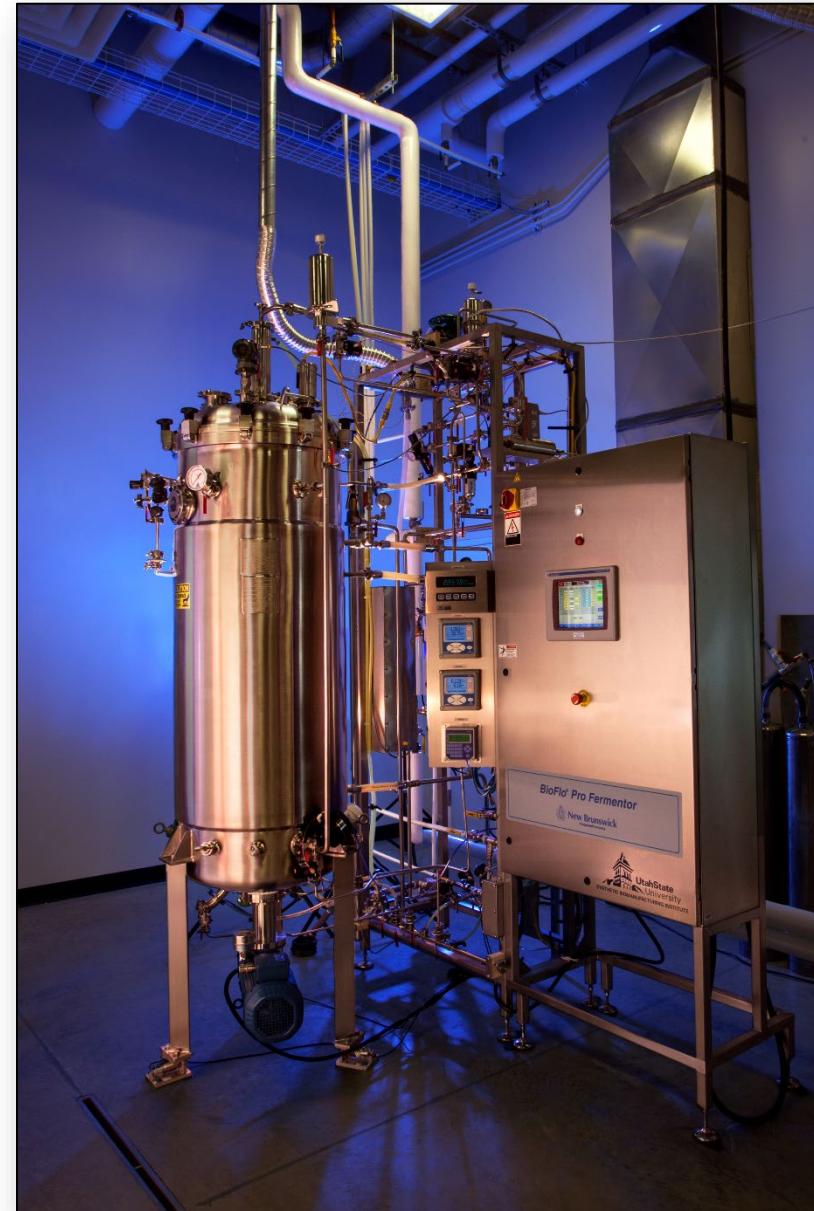
- Robustness Analysis - Determines range of fluxes for any reaction, including nutrients and amino acids, in the designed strain, including the bioproduction pathway reactions, as a function of growth rate.
- Production Envelopes - Determines the bioproduction production capabilities of a designed strain as a function of growth rate for different nutrient and amino acids compositions .
- Phenotype Phase Plane Analysis - Determines the phenotype phase planes for the designed strain. The potential impact of different nutrients and amino acids on the designed strain can be determined through shadow prices and reduced costs.
- Dynamic Flux Balance Analysis - Determines the impact of nutrients and amino acids on the upper limit of bioproduction product production (g/L) over time.





# Bioproduction Production

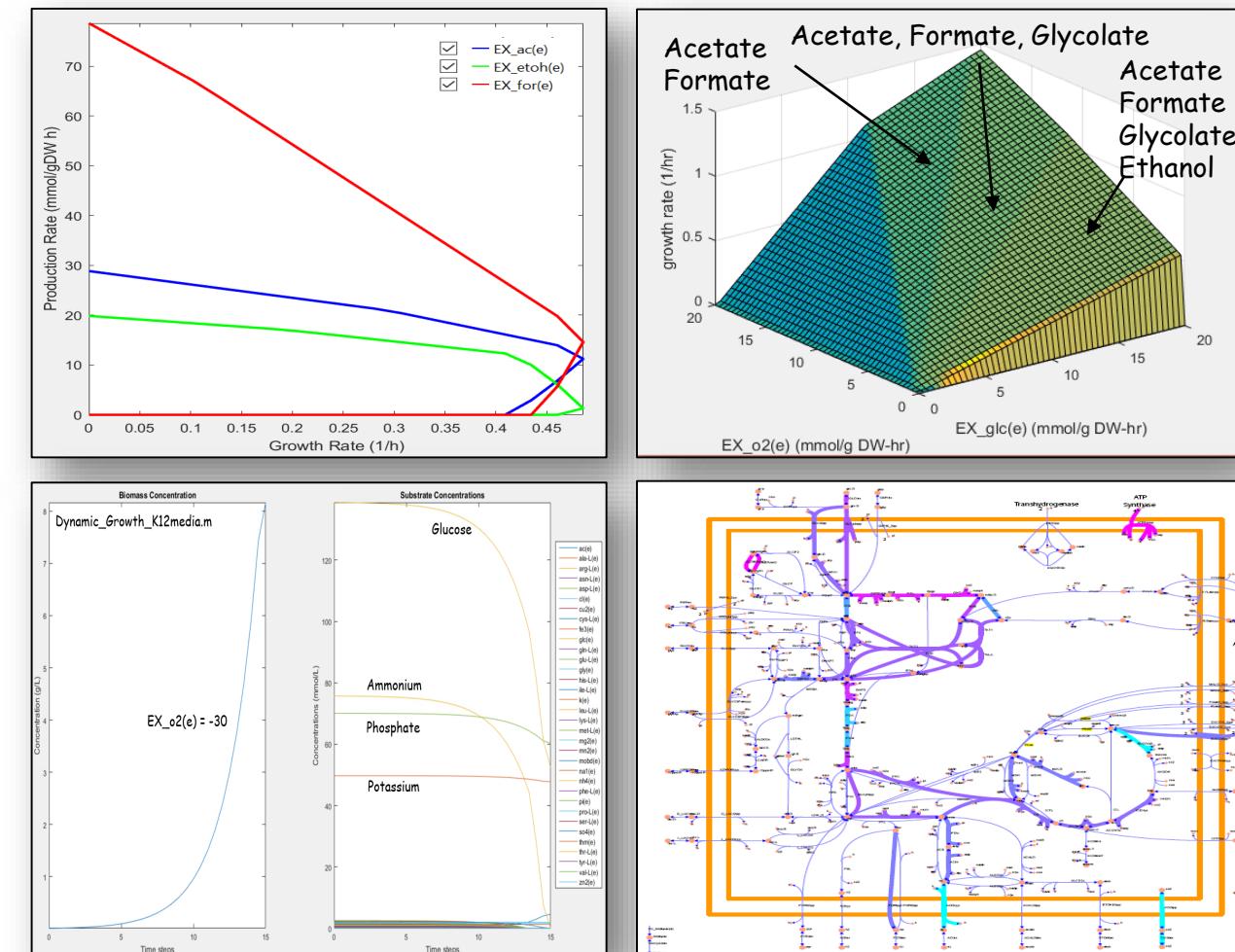
- Bioproduction Identification
- Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





# Cobra Tools for Analyzing and Reducing Undesired Byproducts

- Multi-Production Envelopes - Determines the undesired byproducts produced by a designed strain in a defined media.
- Phenotype Phase Plane Analysis - Determines the phenotype phase planes for the designed strain. The different planes are the result of phenotypes that could be the result of secreted byproducts.
- Dynamic Flux Balance Analysis - Determines the impact of byproducts on the upper limit of bioproduction product production (g/L) since some of the secreted byproducts could become carbon sources for the design strain.
- Flux Balance Analysis - Determines value of fluxes for given environmental conditions of the designed strain. These flux values can be used to understand the magnitude of all the secreted byproducts.





# Bioproduction Production

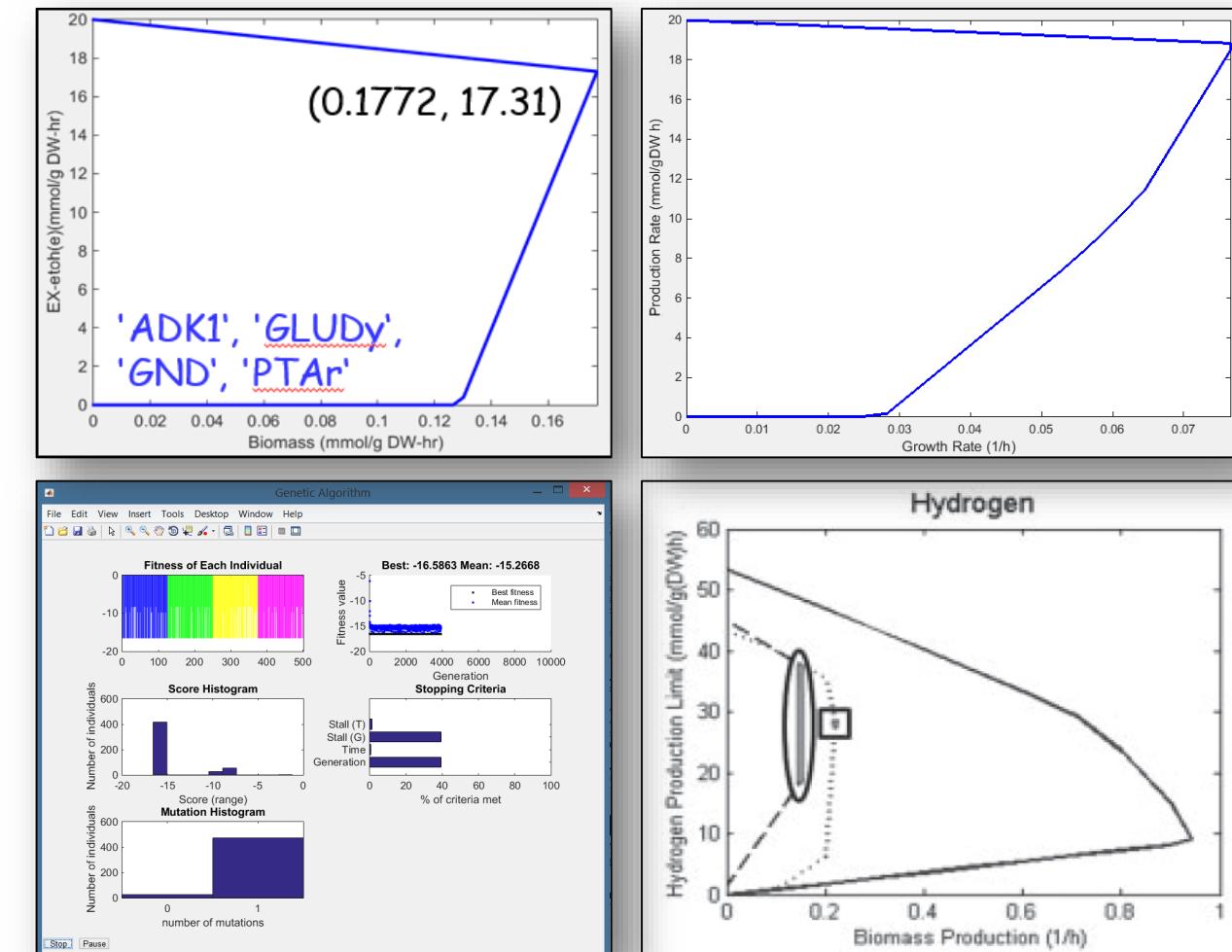
- Bioproduction Identification
- Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





# Cobra Tools for Determining Knockouts

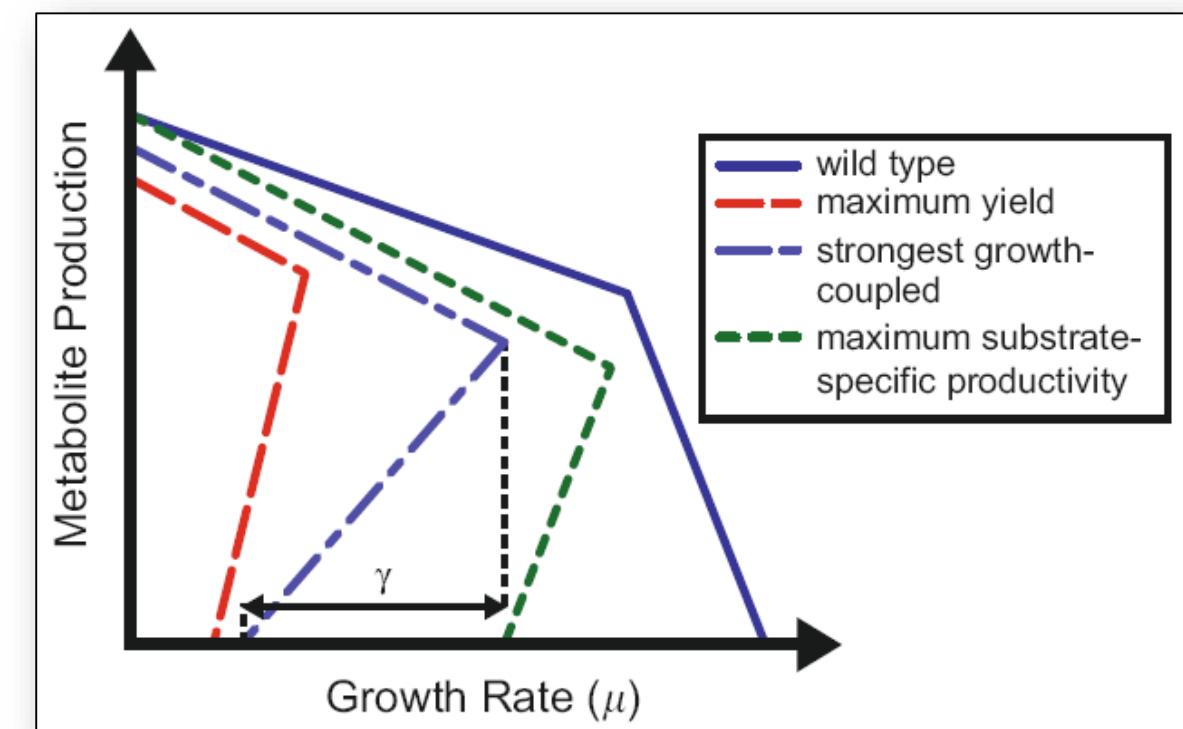
- With growth-coupling, the desired phenotype will show an increase in biomass yield coupled to an increase in the production rate of a desired by-product.
- OptKnock - Though a global search process identifies and subsequently removes metabolic reactions that are capable of uncoupling cellular growth from chemical production.
- GDLS - Using a genetic design local search algorithm to run faster than the global search performed by OptKnock, however, it is not guaranteed to identify the global optima.
- OptGene - An evolutionary programming-based method to determine gene knockout strategies for overproduction of a specific product.
- RobustKnock - A constraint-based method that predicts gene deletion strategies that lead to the over-production of chemicals of interest, by accounting for the presence of competing pathways in the network. RobustKnock provides more robust predictions than those obtained via current state-of-the-art methods (requires the Tomlab CPLEX solver).





# Feist (2010)-Production Potential for Growth-Coupled Bioproducts

- An integrated approach through a systematic model-driven evaluation of the production potential for *E. coli* to produce multiple native products from different representative feedstocks through coupling metabolite production to growth rate.
- Optimal strain designs were based on designs which possess maximum yield, substrate-specific productivity, and strength of growth-coupling for up to 10 reaction eliminations (knockouts).
- The method introduced a new concept of objective function tilting for strain design



Feist, A. M., D. C. Zielinski, et al. (2010). "Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*." *Metabolic engineering* 12(3): 173-186.



# Desirable Production Metrics

- All designs had to be growth coupled. Each equation examines a different desirable production phenotype.
- Product yield ( $Y_{p/s}$ ): Maximum amount of product that can be generated per unit of substrate.

$$Y_{p/s} = \frac{\text{production rate}_{\text{product}}}{\text{consumption rate}_{\text{substrate}}} \left( \frac{\text{mmol product}}{\text{mmol substrate}} \right)$$

- Substrate-specific productivity (SSP): Product yield per unit substrate multiplied by the growth rate

$$SSP = \frac{\text{production rate}_{\text{product}} \cdot \text{growth rate}}{\text{consumption rate}_{\text{substrate}}} \left( \frac{\text{mmol product}}{\text{mmol substrate} \cdot \text{hr}} \right)$$

- Strength of growth coupling (SOC): Product yield per unit substrate divided by the slope of the lower edge of the production curve

$$SOC = \frac{(\text{product yield})^2}{\text{slope}} \left( \frac{\text{mmol product}}{\text{mmol substrate} \cdot \text{hr}} \right)$$

- The slope in this function is the slope of the line between the point of minimum production rate at maximum growth and the point of maximum growth at zero production on a production envelope plot. When this slope is high, it is possible for a strain to grow at very close to the maximum growth rate with only a small production rate, which is undesirable. Therefore, optimizing for maximum production rate is the same as optimizing for maximum product yield. Maximizing for substrate specific productivity introduces a non-linear objective function, which can be handled by OptGene but not OptKnock. Similarly, the strength of coupling is also a non-linear objective function and can only be handled by OptGene. Additionally, a penalty can be added to the scoring function in OptGene by multiplying the objective function with the following penalty function:

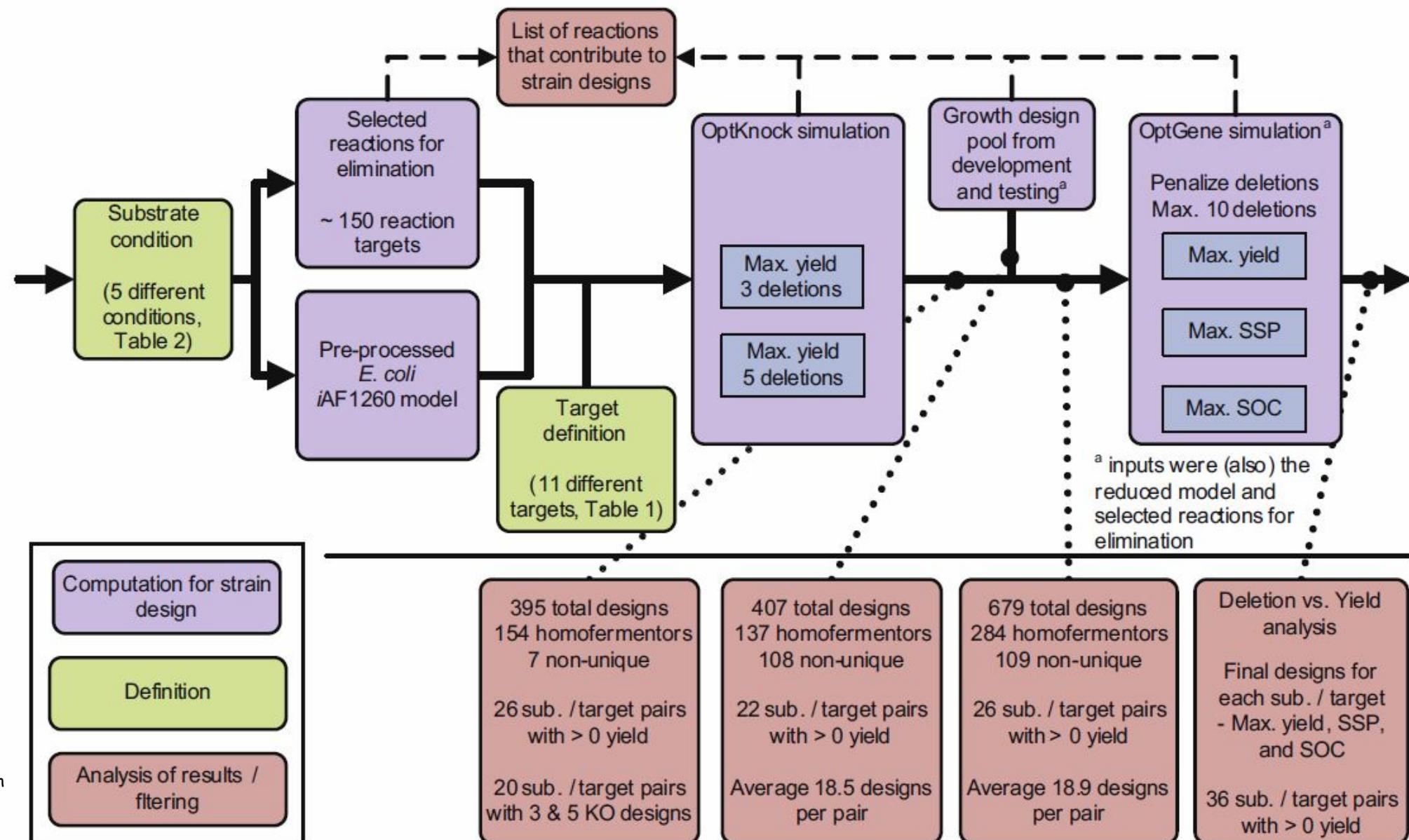
$$\text{objective\_new} = \text{objective\_original} * \text{delPenalty}^{\text{numDels}}$$

where `objective_new` is the new score of the objective function, `objective_original` is the original objective function (e.g., product yield), `delPenalty` is the deletion penalty, and `numDels` is the number of knockout reactions. This penalty ensures that designs with fewer knockouts will be selected over designs with similar phenotypes, but more knockouts. Fewer knockouts are desirable for ease of strain construction.

Feist, A. M., D. C. Zielinski, et al. (2010). "Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli." *Metabolic engineering* 12(3): 173-186.



# Strain Design



Feist, A. M., D. C. Zielinski, et al. (2010). "Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*." *Metabolic engineering* 12(3): 173-186.



# Substrate Conditions

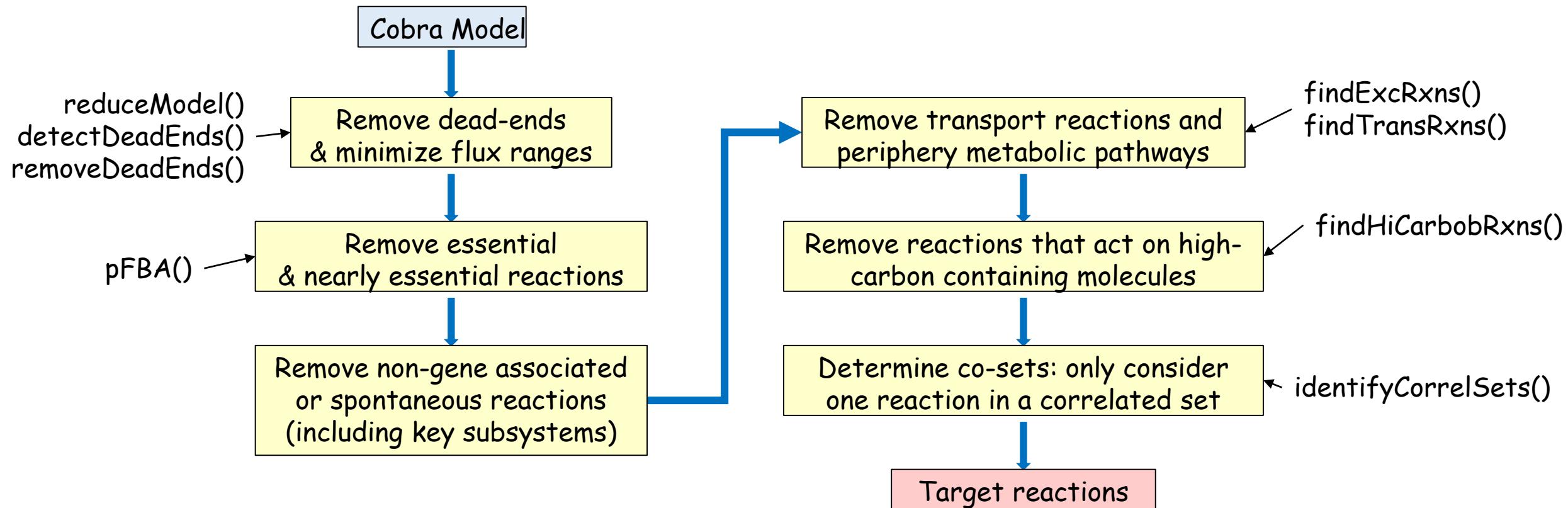
Carbon substrate(s)	Aerobicity	Wild type growth rate	Target reactions after reduction of scope
Glucose	Anaerobic	0.459	142
Xylose	Anaerobic	0.319	141
Glycerol	Anaerobic	0.119	140
Glucose	Aerobic	1.276	170
Xylose	Aerobic	1.131	165
Glycerol	Aerobic	0.983	166

Maximum uptake rates for primary carbon sources were set to  $20 \text{ mmol gDW}^{-1} \text{ h}^{-1}$ . In aerobic simulations, maximum oxygen uptake rate was set to  $20 \text{ mmol gDW}^{-1} \text{ h}^{-1}$ .

Feist, A. M., D. C. Zielinski, et al. (2010). "Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*." *Metabolic engineering* 12(3): 173-186.



# Problem Formulation: Reduction of Model and Selection of Targeted Reactions



Feist, A. M., D. C. Zielinski, et al. (2010). "Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli." *Metabolic engineering* 12(3): 173-186.



# Theoretical maximum production $Y_{p/s}$ (%) analysis.

$$Y_{p/s} = \frac{\text{production rate}_{\text{product}}}{\text{consumption rate}_{\text{substrate}}} \left( \frac{\text{mmol product}}{\text{mmol substrate}} \right)$$

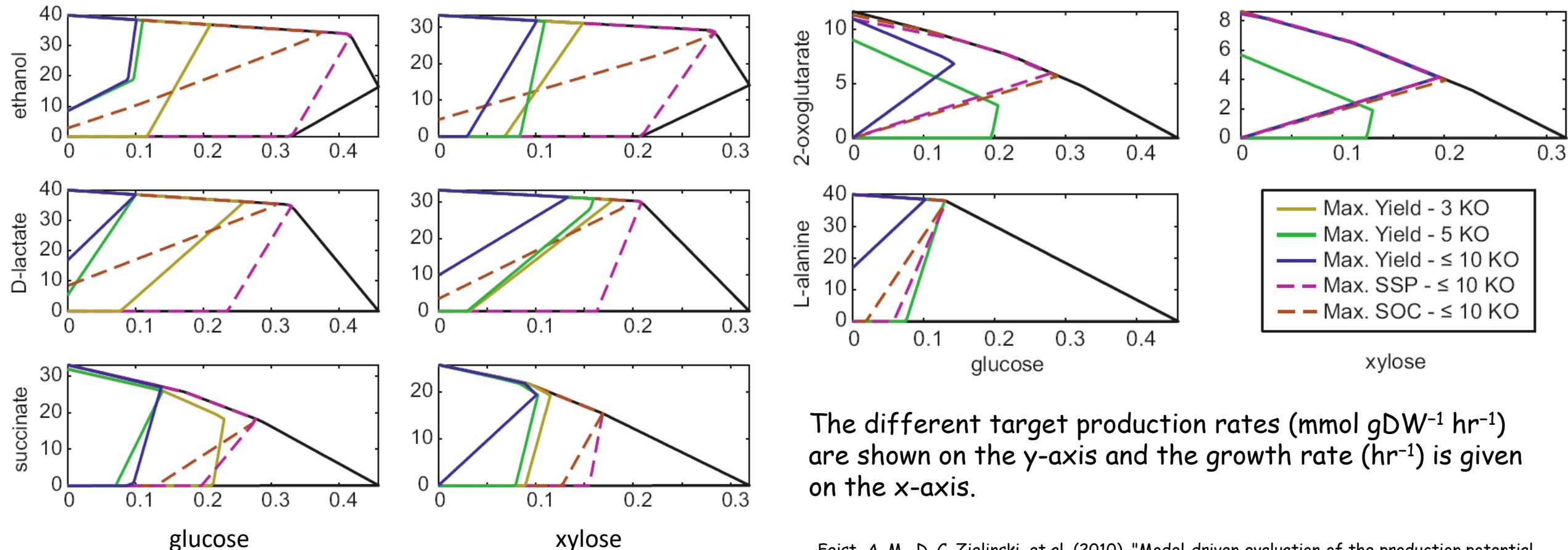
Product	No. of carbons	Anaerobic			Aerobic		
		Glucose	Xylose	Glycerol	Glucose	Xylose	Glycerol
Ethanol	2	49 <sup>a</sup>	49 <sup>a</sup>	49	49	49	54
D-Lactate	3	95 <sup>a</sup>	95 <sup>a</sup>	13	95	95	97
Glycerol	3	37	27		75	74	
L-Alanine	3	95 <sup>a</sup>	76	13	95	93	96
L-Serine	3	47	35	6	115	114	116
Pyruvate	3	71	60	6	100	99	100
Fumarate	4	54	40	5	110	108	117
L-Malate	4	63	46	5	127	125	135
Succinate	4	93	81	12	104	101	111
2-Oxoglutarate	5	40	32	3	98	96	101
L-Glutamate	5	44	36	3	92	90	97

<sup>a</sup> Indicates anaerobic condition where homofermentation of product is possible (<2 wt% other carbon products, CO<sub>2</sub> exempt), all aerobic conditions have homofermentation potential.

Feist, A. M., D. C. Zielinski, et al. (2010). "Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli." *Metabolic engineering* 12(3): 173-186.



# The Strain Designs Generated for Five Different Targets from Glucose and Xylose Anaerobically



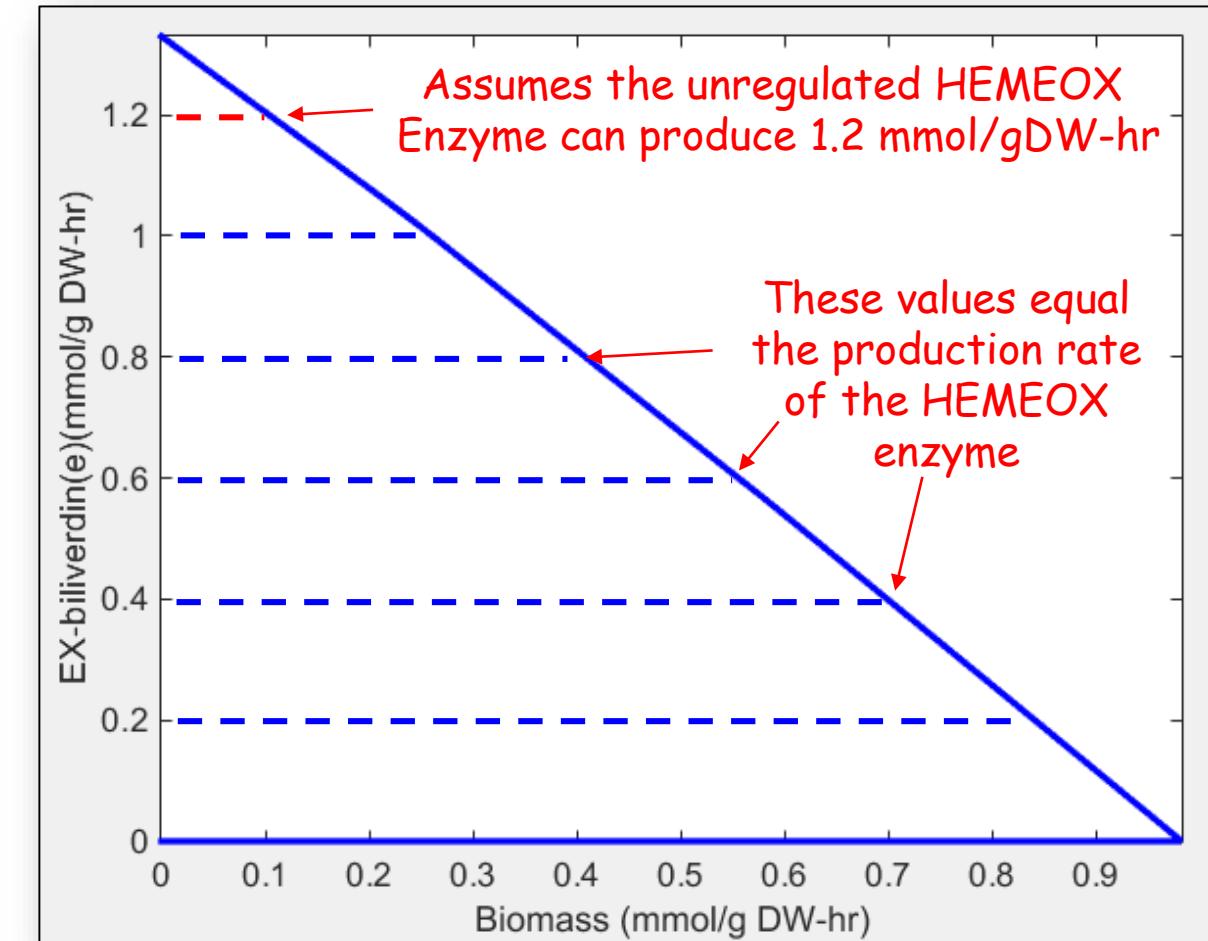
The different target production rates ( $\text{mmol gDW}^{-1} \text{ hr}^{-1}$ ) are shown on the y-axis and the growth rate ( $\text{hr}^{-1}$ ) is given on the x-axis.

Feist, A. M., D. C. Zielinski, et al. (2010). "Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*." *Metabolic engineering* 12(3): 173-186.



# Uncoupled Bioproducts

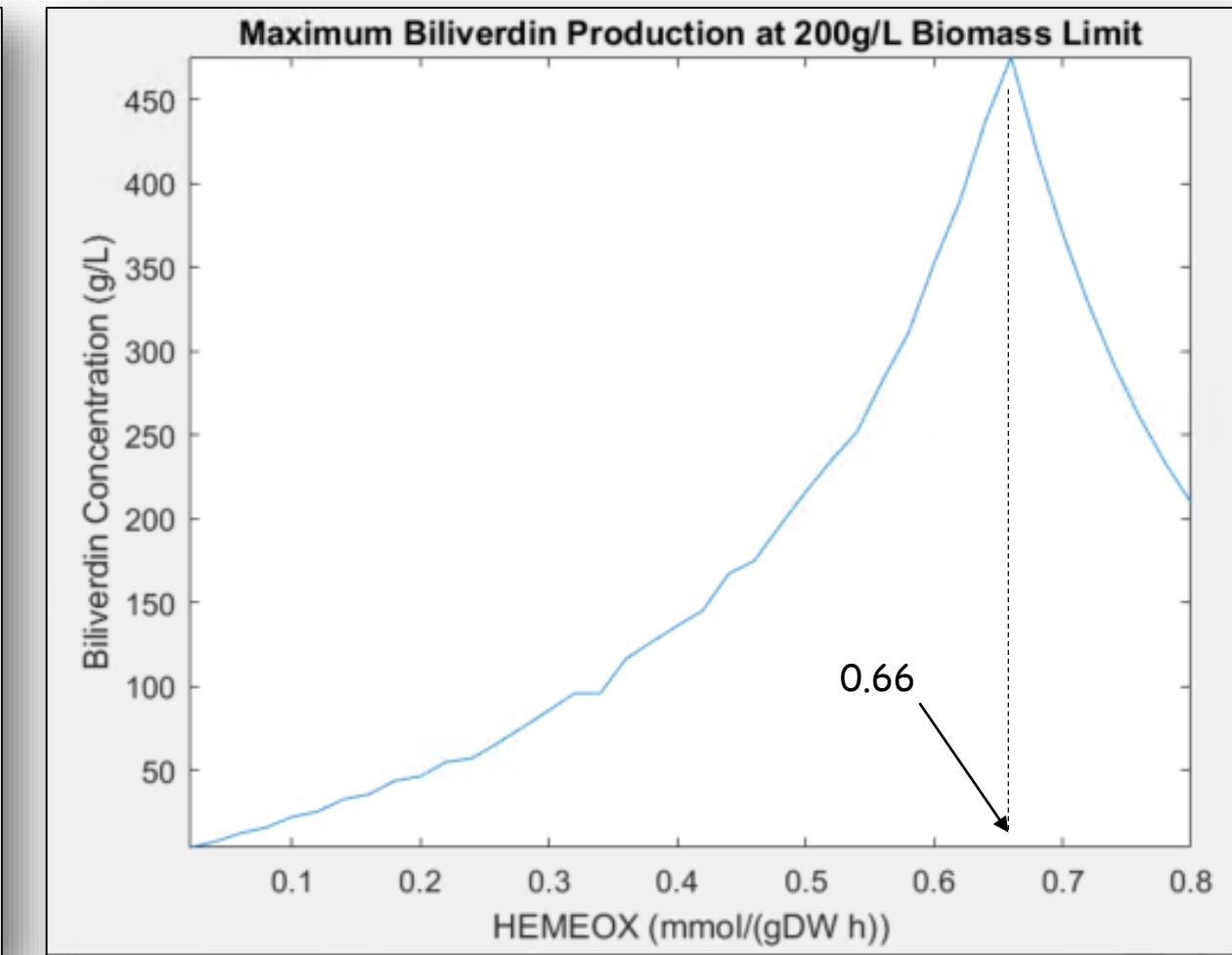
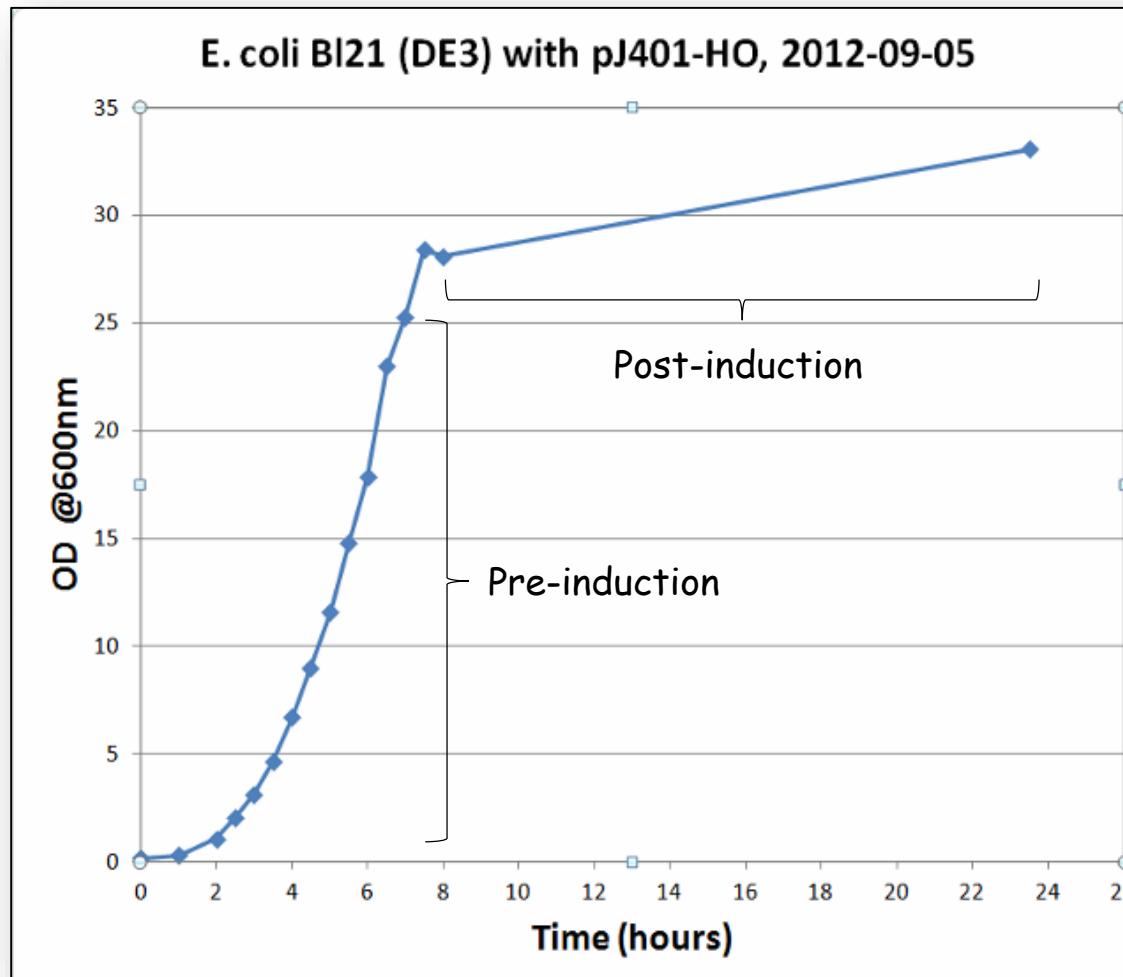
- In some cases it is desirable to avoid coupling the desired bioproduct to the growth of the cell.
- Through the use of an unregulated plasmid the bioproduct pathway can use nearly all of the cells resources to create the desired bioproduct. The result is virtually no cell growth but with large bioproduct production.
- The key enzymes in the desired pathway are included in the plasmid.
- The flux supported by the key enzymes is a function of the plasmids copy number and promoter strength. The more copies of the plasmid and stronger promoter strength lead to larger flux.



Biliverdin\_ProductionEnvelope\_BL21.m



# Maximum Production: Post-induction



Data provided by Dr. Dong Chen at Utah State University

DynamicBiliverdinProduction\_BL21\_maxProd\_200Limit\_Plasmid.m



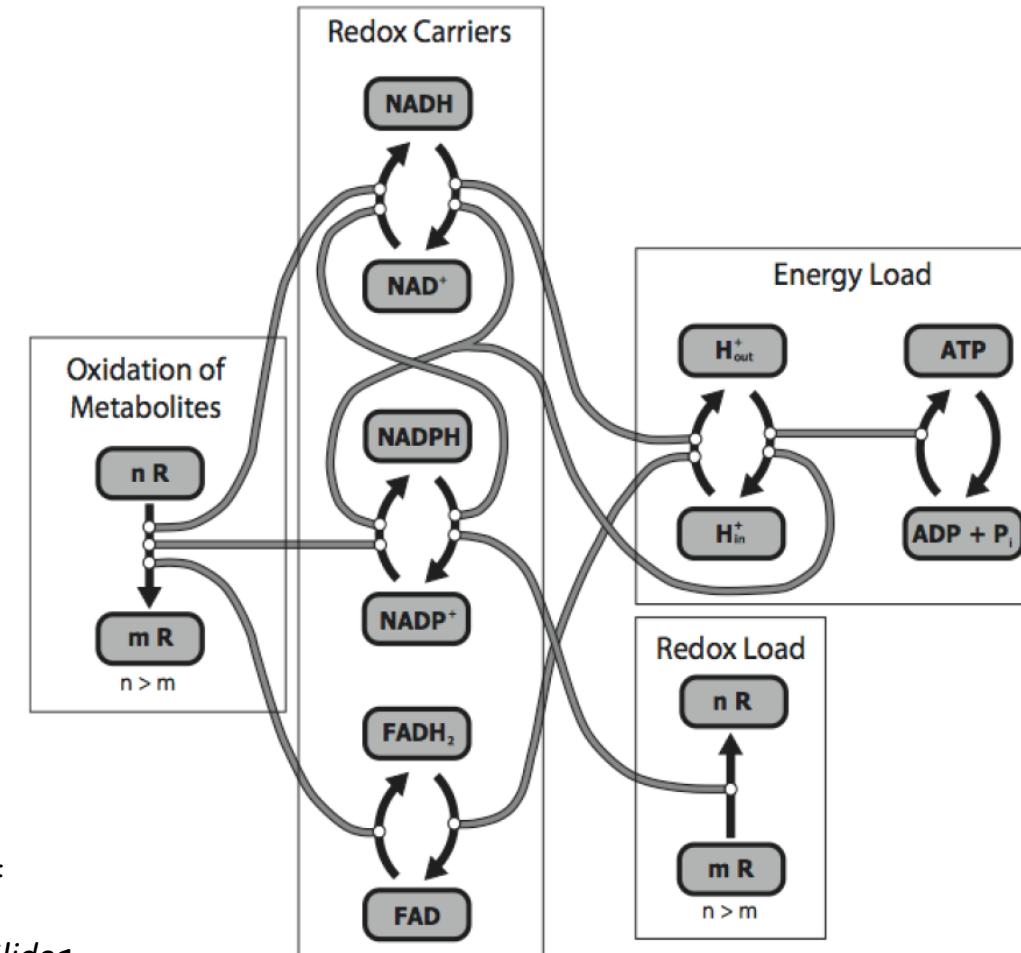
# Bioproduction Production

- Bioproduction Identification
- Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





# Redox Trafficking in the Core Metabolic Pathways: Cofactor View



B. O. Palsson lectures on Systems Biology: Simulation of Dynamic Network States, Lecture #8.  
<http://sbrg.ucsd.edu/Publications/Books/SB2LectureSlides>

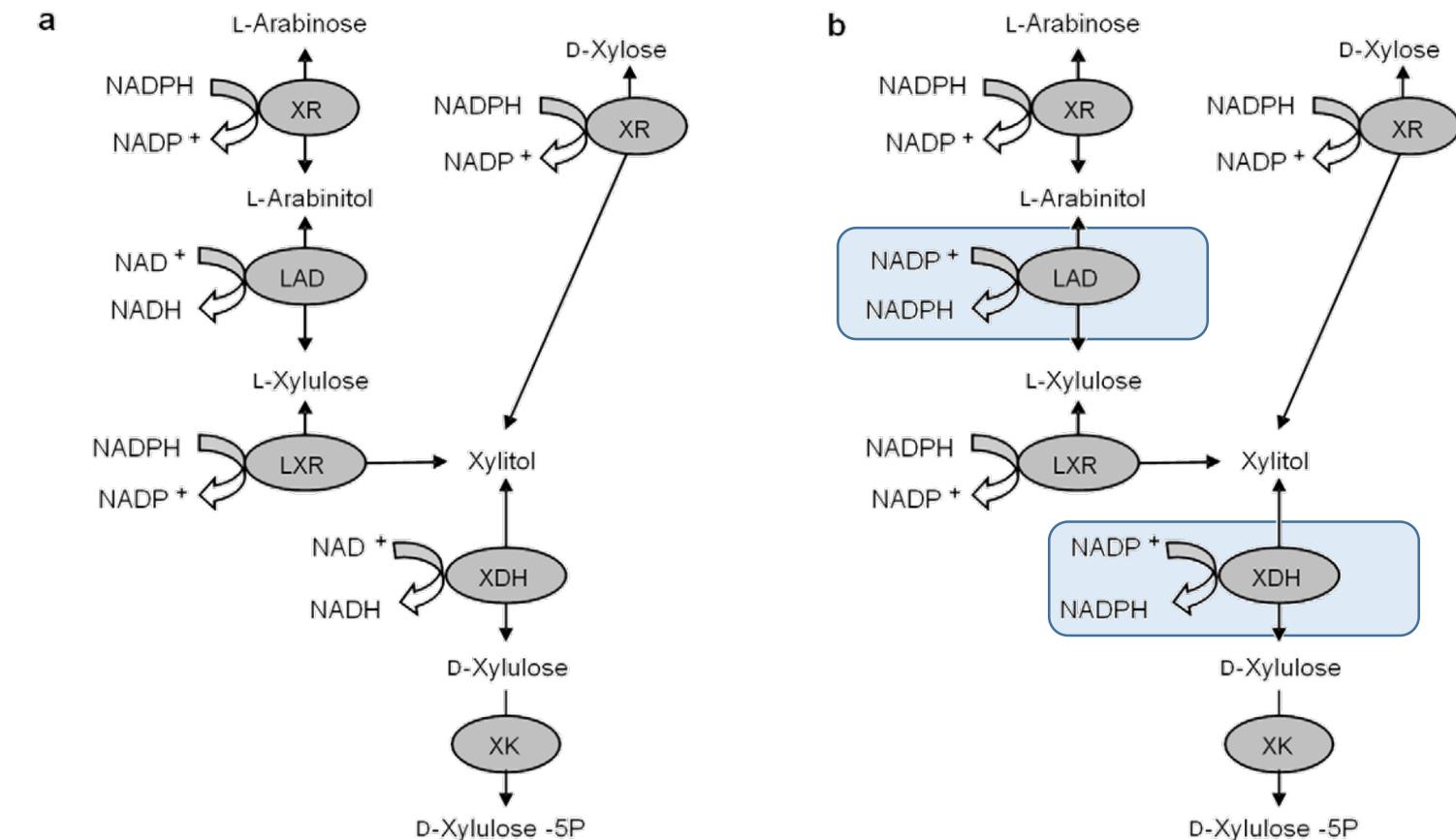


# Cofactor Balance

- Cofactor imbalance - Changes that will cause an imbalance in cofactor utilization that would need to be compensated for elsewhere in the network.
- The primary role of the reduced respiratory cofactor NADH is to transfer electrons to oxygen via the electron transport chain, generating the proton gradient that is used for oxidative phosphorylation of adenosine diphosphate (ADP) to adenosine triphosphate (ATP). Concurrently, the reduced cofactor NADPH donates electrons to anabolic reactions and drives the biosynthetic pathways in the cell.
- Solutions include
  - ✓ Altering genes/reaction cofactor specificity to balance cofactors (NADH vs NADPH)
  - ✓ Add heterogeneous genes (e.g. ferredoxins) that recycle cofactors



Wang, Y., K. Y. San, et al. (2013). "Cofactor engineering for advancing chemical biotechnology." *Current opinion in biotechnology* 24(6): 994-999.



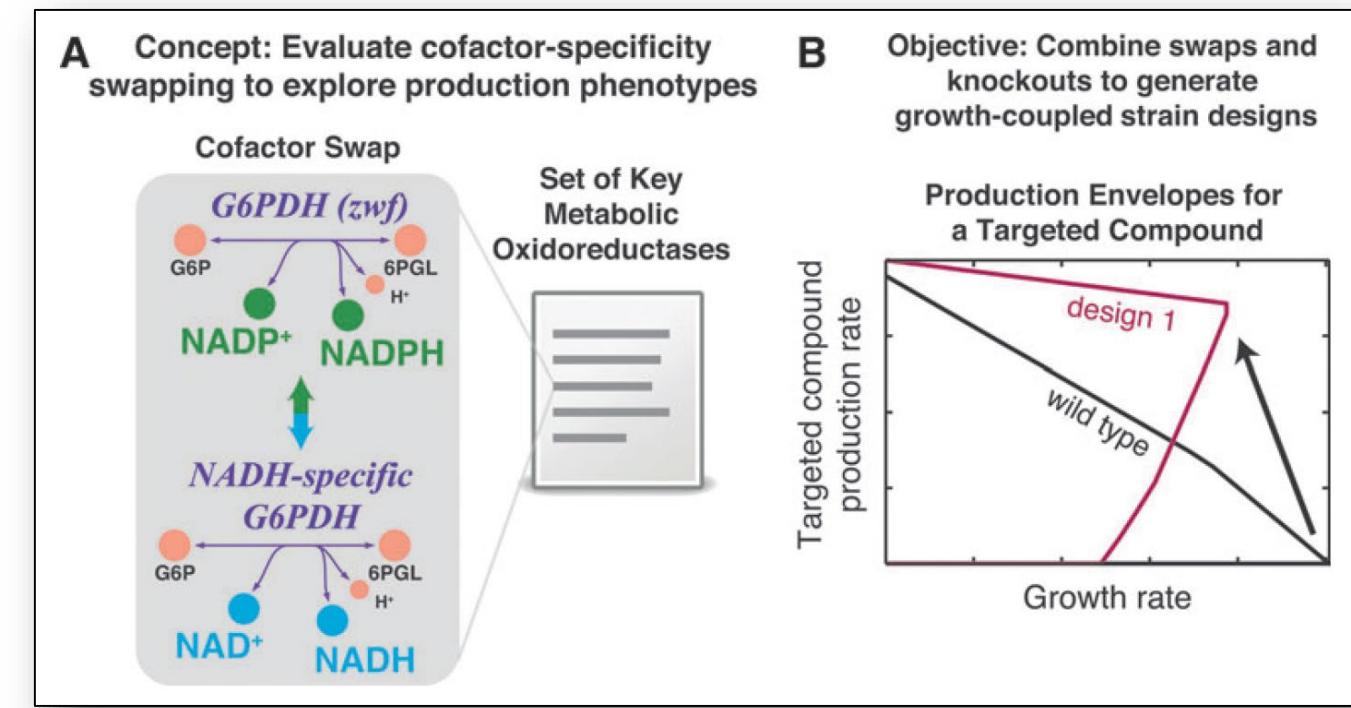
(a) Engineered cofactor imbalanced L-arabinose and D-xylose pathways, (b) Engineered cofactor balanced L-arabinose and D-xylose pathways. XR: Xylose reductase; LAD: L-arabinitol dehydrogenase; LXR: L-xylulose reductase; XDH: xylitol dehydrogenase; XK: xylulokinase.

Ghosh A, Zhao H, Price ND (2011) Genome-Scale Consequences of Cofactor Balancing in Engineered Pentose Utilization Pathways in *Saccharomyces cerevisiae*. PLoS ONE 6(11): e27316. doi:10.1371/journal.pone.0027316



# OptSwap

- Central oxidoreductase enzymes (eg, dehydrogenases, reductases) in microbial metabolism often have preferential binding specificity for one of the two major currency metabolites NAD(H) and NADP(H). These enzyme specificities result in a division of the metabolic functionality of the currency metabolites: enzymes reducing NAD<sup>+</sup> to NADH drive oxidative phosphorylation, and enzymes reducing NADP<sup>+</sup> to NADPH drive anabolic reactions.
- OptSwap predicts bioprocessing strain designs by identifying optimal modifications of the cofactor binding specificities of oxidoreductase enzymes and complementary reaction knockouts.
- Independent of the Cobra Toolbox and requires the Tomlab CPLEX solver

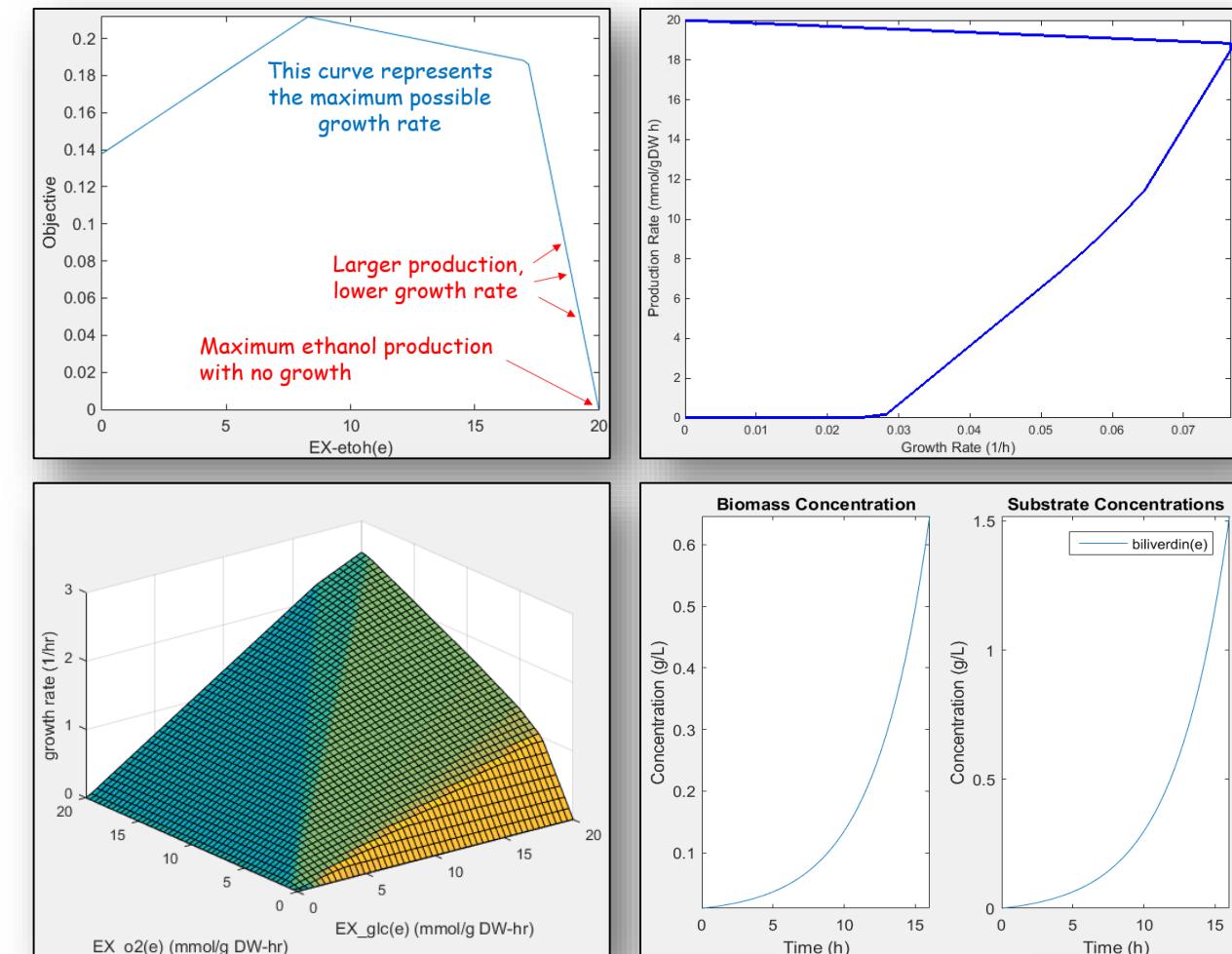


Z. A. King and A. M. Feist, Optimizing cofactor specificity of oxidoreductase enzymes for the generation of microbial production strains—OptSwap. *Ind. Biotechnol.* 9, 236-246



# Cobra Tools for Balancing Cofactors

- Robustness Analysis - Used to determine pathway reaction improvements resulting from cofactor balancing.
- Production Envelopes - Determines the bioproduct production capabilities from the cofactor balanced strain.
- Phenotype Phase Plane Analysis - Determines the phenotype phase planes for the cofactor balanced strain.
- Dynamic Flux Balance Analysis - Determines the upper limit of bioproduct product production (g/L) over time for the cofactor balanced strain.
- OptSwap - Predicts bioprocessing strain designs by identifying optimal modifications of the cofactor binding specificities of oxidoreductase enzymes and complementary reaction knockouts (requires the Tomlab CPLEX solver)





# Bioproduction Production

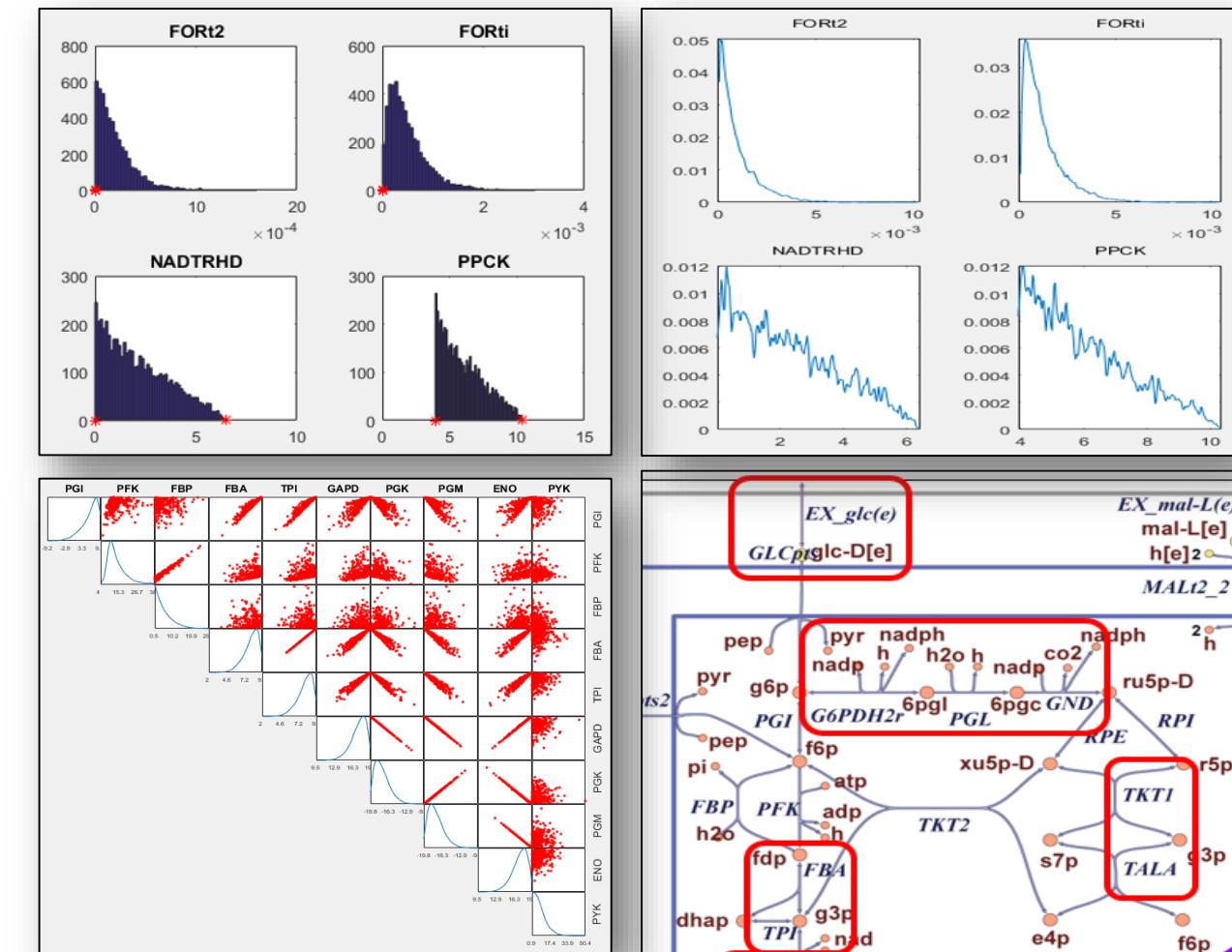
- Bioproduction Identification
- Select Host Strain
- Bioproduction Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduction Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





# Cobra Tools for Sampling Analysis

- gpSampler - Samples the solution space. Identifies the distribution of flux activity possible through a reaction for a given solution space.
- plotSampleHist - Compare flux histograms for one or more samples for one or more reactions
- sampleScatterMatrix - A scatter plot allows the visualization of the interaction between two network reactions. It can also display correlation coefficients associated with two reactions.
- identifyingCorrelSets - Two reactions are part of the same "correlated reaction set" if their fluxes are linearly correlated. If a knockout reaction has been identified then any of the correlated reaction set can be removed with the same impact.





# Bioproduct Production

- Bioproduct Identification
- Select Host Strain
- Bioproduct Pathway
  - ✓ Defining pathway
  - ✓ Adding pathway to the model
  - ✓ Adding the plasmid to the model
- Strain Design
  - ✓ Bioproduct Maximum Production
  - ✓ Carbon Sources
  - ✓ Nutrient & Amino Acid Limitations
  - ✓ Undesired By-products
  - ✓ Growth Coupling
  - ✓ Cofactor Balancing
  - ✓ Sampling Analysis





# Reflective Questions

1. What are the main Cobra tools for determining the knockouts that can improve bioproduct production?
2. What is the Cobra function for adding a reaction?
3. Can a host cell be growth-coupled to all new pathways created by adding reactions?
4. Why can a cell produce a bioproduct when a gene is added to a host cell even though the host is not growth-coupled to the bioproduct?
5. What governs the maximum production of a bioproduct in a given host cell?
6. What role does the media play in the maximum production of a bioproduct?
7. What is cofactor balancing?
8. In addition to the designated carbon source in a growth media what other components of the media can act as carbon sources?
9. What components in a growth media can impact the maximum bioproduct production?
10. How is the M9 minimal media modeled?
11. In the standard E.coli models (textbook, iAF1260, and iJO1366) what are the typical default lower bounds for most amino acids and minerals? What are the default upper bounds?
12. What relationship do the amino acids have with the biomass function?
13. How can amino acid and mineral uptake rates impact growth rate and bioproduct production?
14. How can reduced costs be used to understand the impact of an amino acid or mineral on bioproduct production?
15. Are regulatory constraints included in the standard E.coli constraint-based models?
16. What impacts can be observed by adding a plasmid to a host cell?
17. How can plasmids be modeled?
18. After a gene has been knockout or a new gene has been added to a host cell does the maximum theoretical performance match the laboratory results?
19. What are the components of the host strain design process?
20. What are the assumptions and limitations of bioproduction modeling?
21. What are the differences between natural-inherent chemicals, natural-noninherent chemicals, non-natural noninherent chemicals, and nonnatural-created chemicals?
22. Why is E.coli a good host organism?
23. What are uncoupled bioproducts?