*Molecular Systems Biology* 2013, **9**: 693

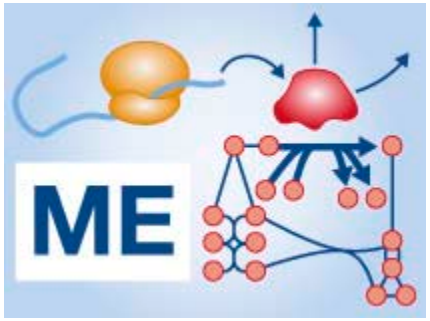
# Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction

Edward J O'Brien, Joshua A Lerman, Roger L Chang, Daniel R Hyduke & Bernhard Ø Palsson

Hai Shi

Jörg Schwender's Lab

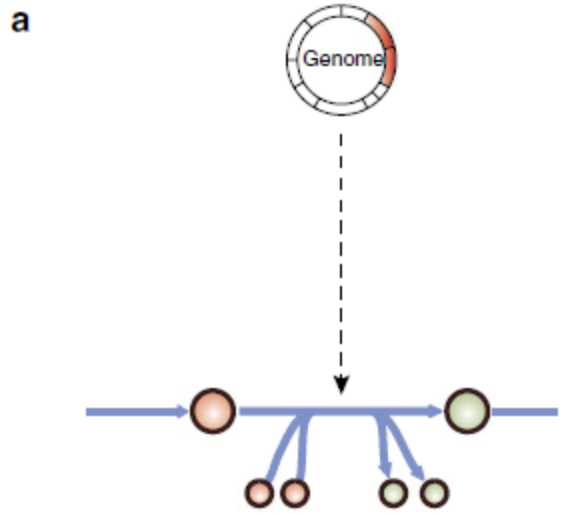
12/09/2013



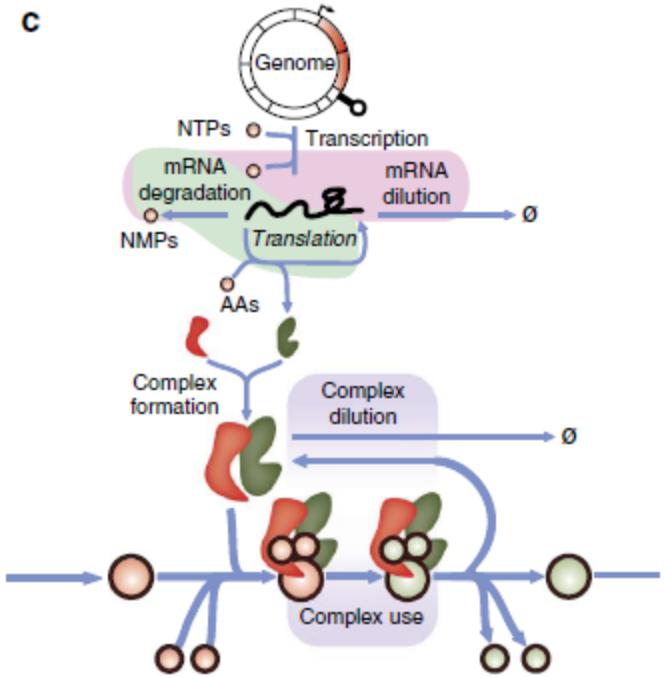
# ME-models (Metabolism and Expression): integrate metabolic and gene product expression pathways

- First ME-model for *E.coli*
- Incorporate gene express and proteomic data
- Predict several cellular phenotypes
- Provide a method that can be used to simulate growth and changes in genome scale systems
- Develop into a powerful approach to integrate transcriptomics data with metabolic networks in plant

# M-models (metabolic) *vs.* ME-models



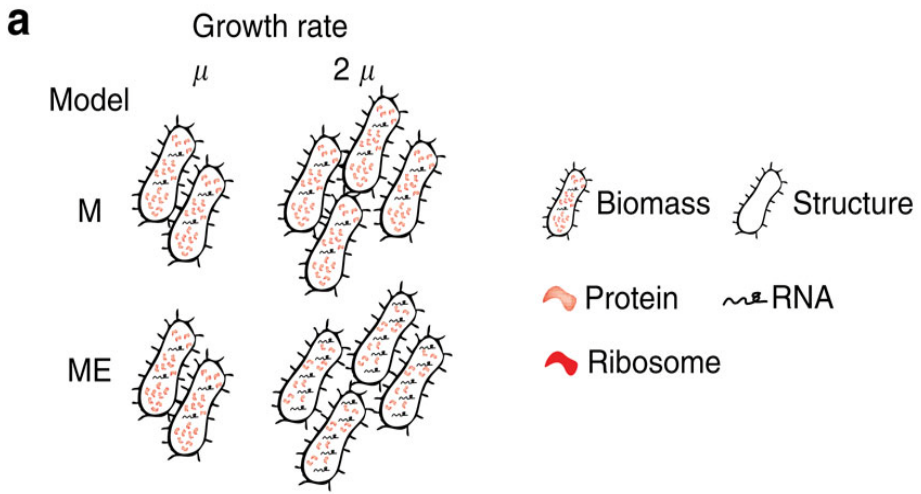

- Predict reaction flux
- Genes are either ON or OFF




- Add transcription and translation
- Account for RNA generation and degradation
- Account for peptide creation and degradation
- Gene expression and gene products explicitly modeled and predicted

Nature Communications 2012, 3: 929

# M-models *vs.* ME-models (continued)




b

| Model | Simulation objective | Measurement | Assumption |
|---|---|---|---|
| M | Maximize $\mu$ | Subject to $\frac{v_{substrate}}{v_{}}$ | Constant composition |
| ME | Minimize $v_{dilution}$ | Subject to $\mu$ | Constant composition |

## M-models

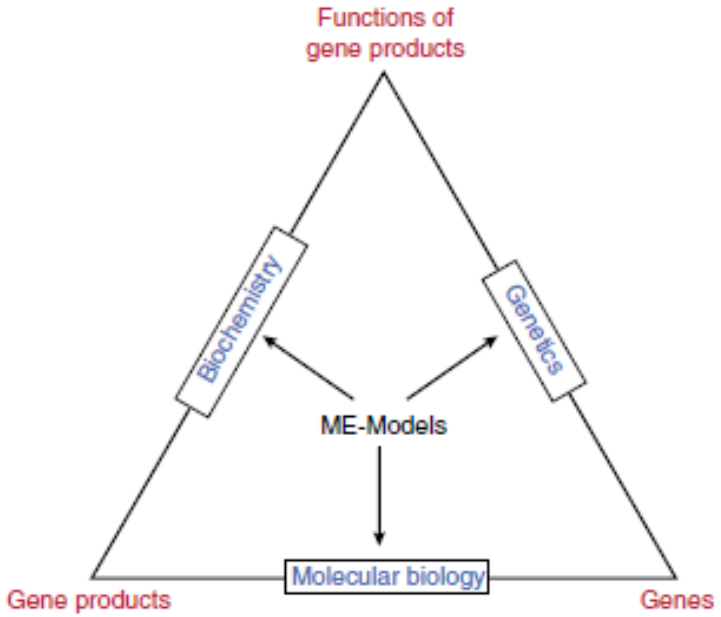
- Account for total biomass reaction
- $T_d$ was integrated to define a biomass reaction
- LP used to find max growth subject to (measured) uptake rate

## ME-models

- Account only for "constant" cell structure
  - Cofactors like Coenzyme A
  - DNA like dCTP, dGTP
  - Cell wall lipids
  - Energy necessary to create and maintain them
- Model approximates a cell whose composition is a function of environment and growth rate
- Cellular composition (mRNA, tRNA, ribosomes) taken into account as dynamic reactions
- LP used to identify the minimum ribosome production rate required to support an experimentally determined growth rate

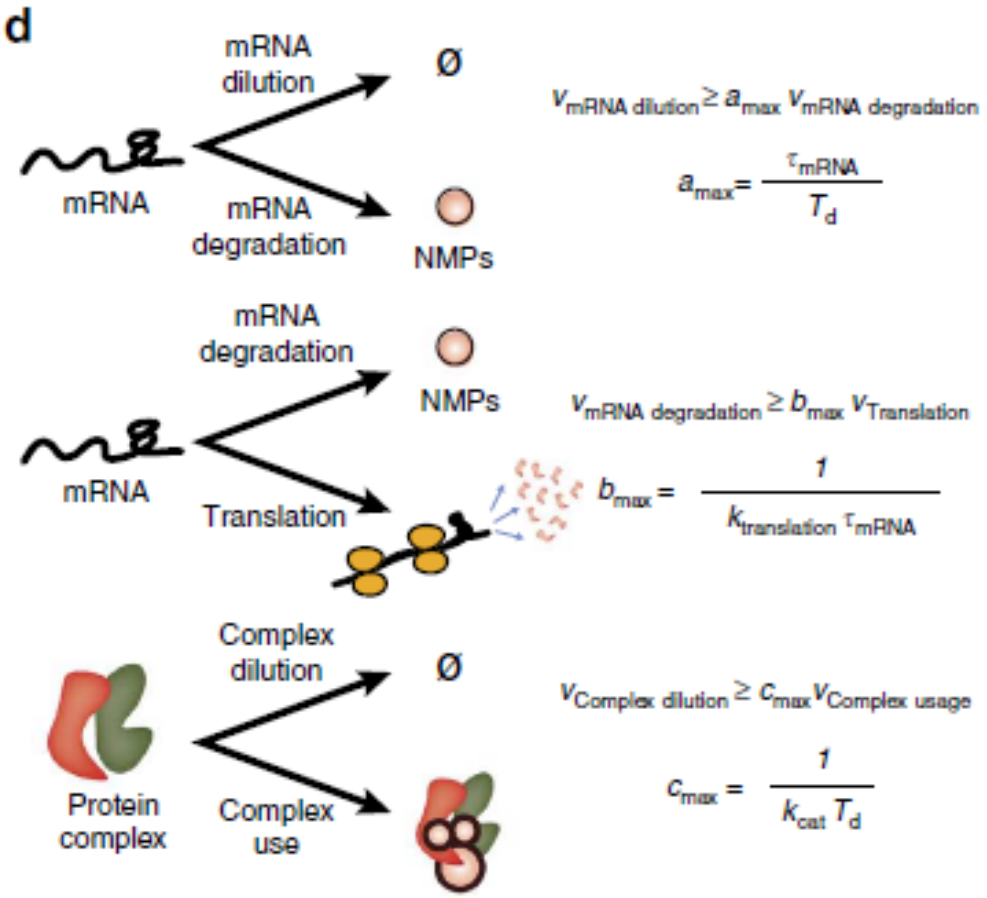
Nature Communications 2012, 3: 929

# ME-model: the details




ME-Models provide links between the biological sciences. With an integrated model of metabolism and macromolecular expression, it is possible to explore the relationships between gene products, genetic perturbations and gene functions in the context of cellular physiology.

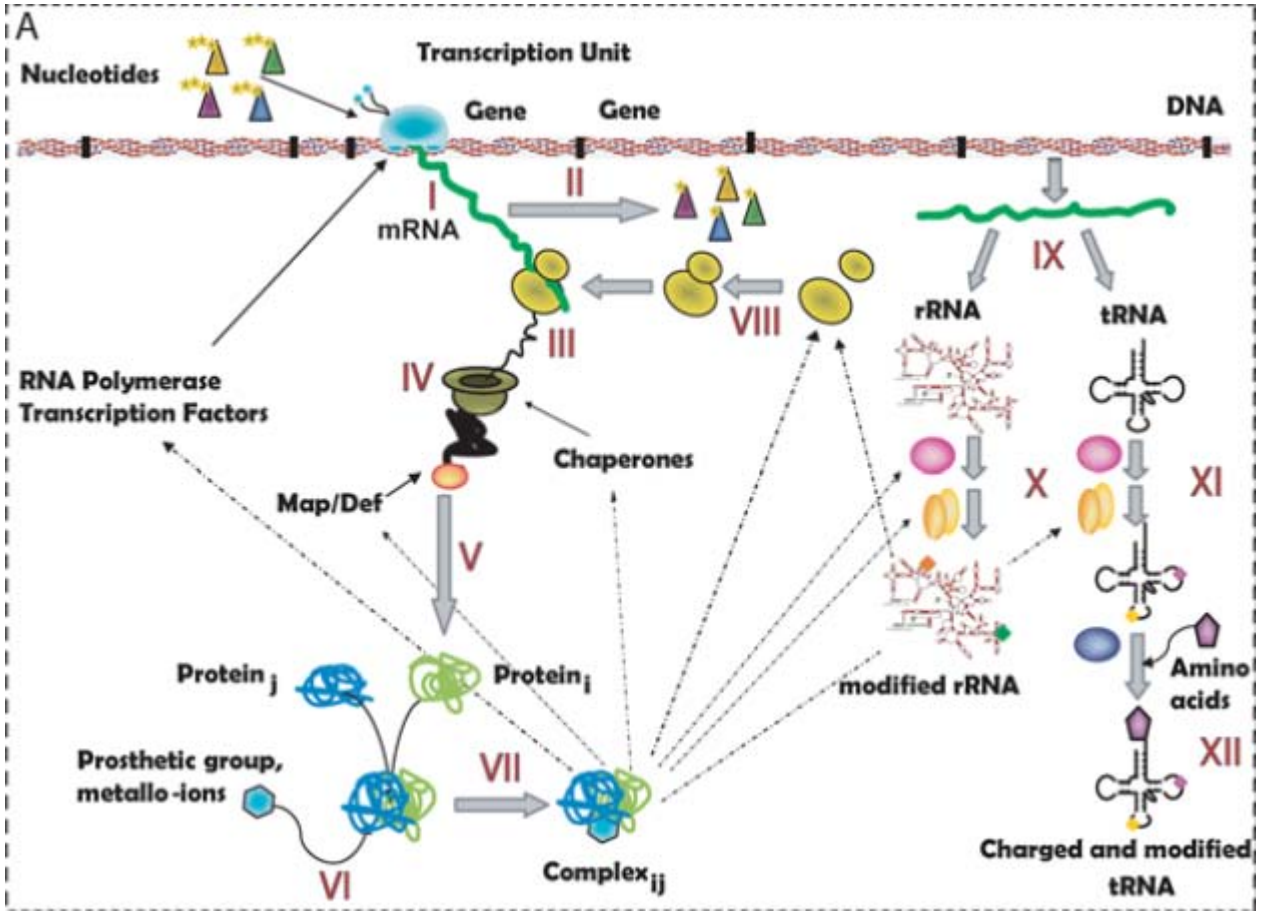
Nature Communications 2012, 3: 929




The transcriptional, translational and enzymatic activities are coupled to doubling time ($T_d$) using constraints that limit transcription and translation rates as well as enzyme efficiency.

# Integration of genome-scale reaction networks of protein synthesis and metabolism

| | *i*JO1366 (this study) |
|---|---|
| *Included genes* | 1366 (32%)[a] |
| Experimentally based function | 1328 (97%) |
| Computationally predicted function | 38 (3%) |
| *Unique functional proteins* | 1254 |
| Multigene complexes | 185 |
| Genes involved in complexes | 483 |
| Instances of isozymes[b] | 380 |
| *Reactions* | 2251 |
| Metabolic reactions | 1473 |
| Unique metabolic reactions[c] | 1424 |
| Cytoplasmic | 1272 |
| Periplasmic | 193 |
| Extracellular | 8 |
| *Transport reactions* | 778 |
| Cytoplasm to periplasm | 447 |
| Periplasm to extracellular | 329 |
| Cytoplasm to extracellular | 2 |
| *Gene–protein–reaction associations* | |
| Gene associated (metabolic/transport) | 1382/706 |
| Spontaneous/diffusion reactions[d] | 21/14 |
| Total (gene associated and no association needed) | 1403/720 (94% |
| No gene association (metabolic/transport) | 70/58 (6%) |
| *Exchange reactions* | 330 |
| *Metabolites* | |
| Unique metabolites | 1136 |
| Cytoplasmic | 1039 |
| Periplasmic | 442 |
| Extracellular | 324 |




I: Transcription; II: mRNA degradation; III: translation; IV: protein maturation; V: protein folding; VI: metallo-ion binding; VII: protein complex formation; VIII: ribosome assembly; IX: RNA processing; X: rRNA modification; XI: tRNA modification; XII: tRNA charging.

Mol Syst Biol. 2011; 7: 535 PLoS Comput Biol. 2009 5(3): e1000312

# Growth demands and general constraints on molecular catalysis

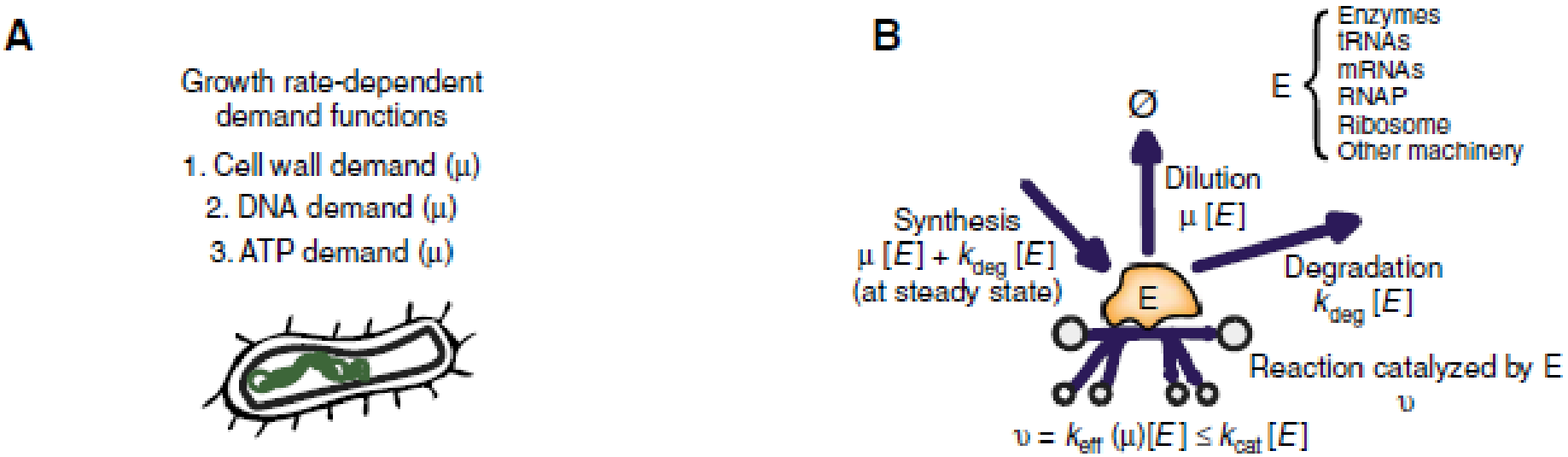
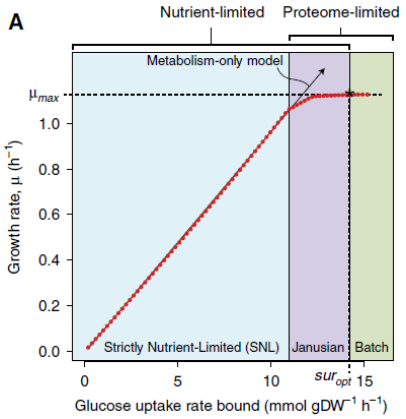
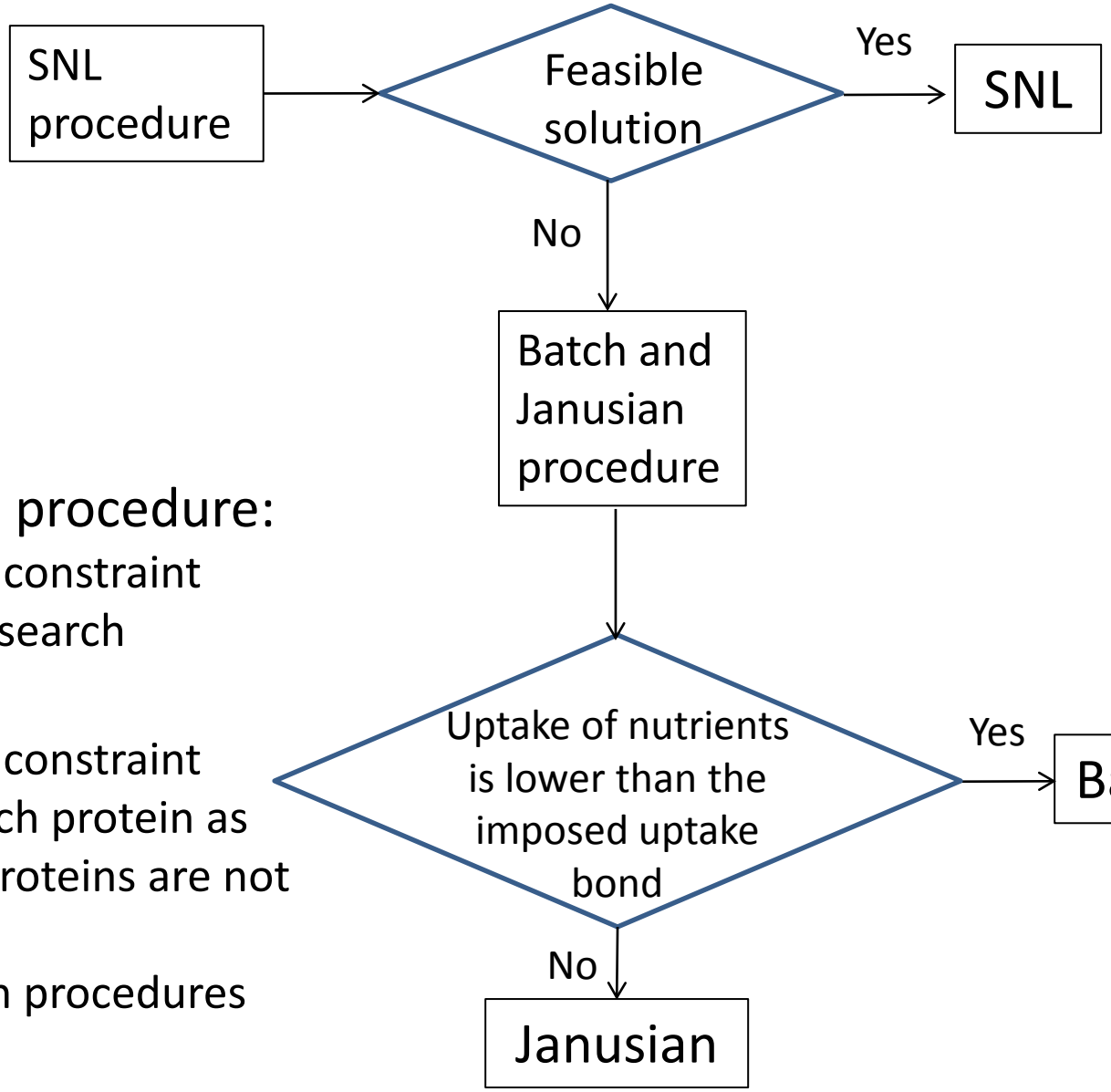



Figure 1. Growth demands and coupling constraints leading to growth rate-dependent changes in enzyme and ribosome efficiency. (**A**) Three growth rate-dependent demand functions derived from empirical observations determine the basic requirements for cell replication. (**B**) Coupling constraints link gene expression to metabolism through the dependence of reaction fluxes on enzyme concentrations.

# Optimization procedure




Batch and Janusian procedure:

- Biomass capacity constraint
- Perform a binary search

SNL procedure:

- Biomass capacity constraint
- Cells make as much protein as possible and the proteins are not saturated
- Two binary search procedures

# Derivation of constraints on molecular catalytic rates (1)

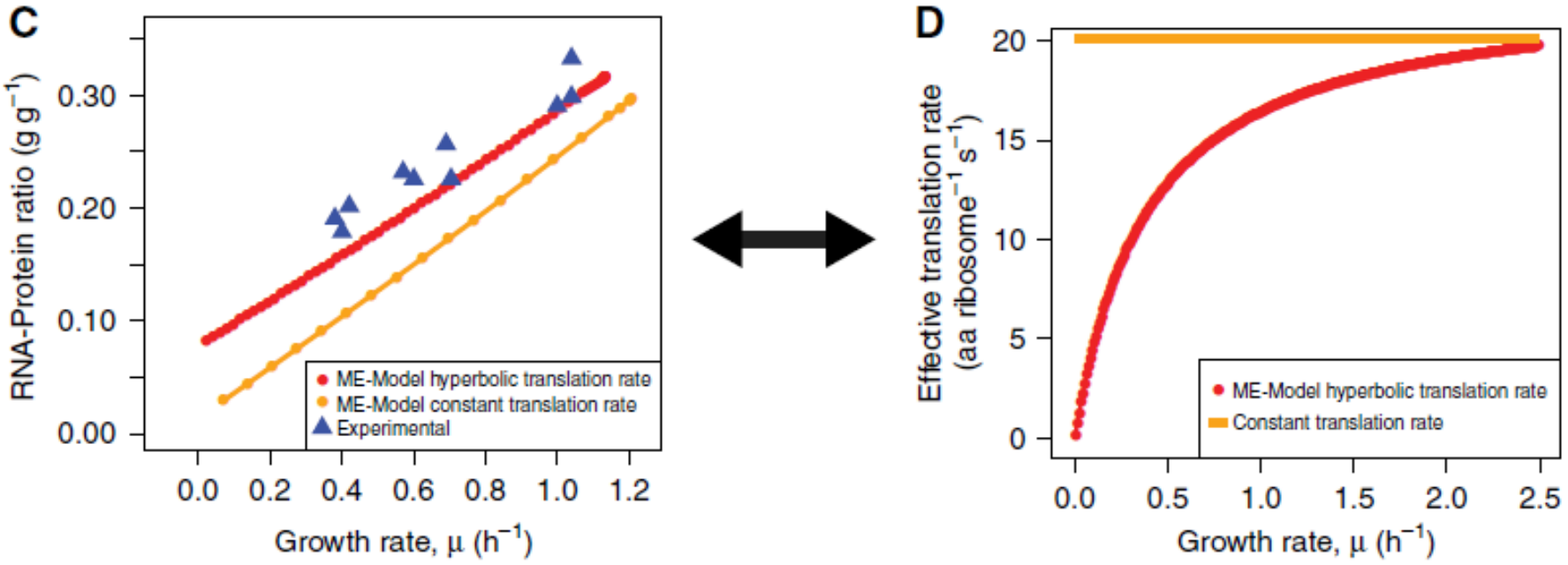



Figure 1. (**C**, **D**) RNA:protein ratio predicted by the ME-Model with two different coupling constraint scenarios, one for variable translation rate versus growth rate (red lines) and one for constant translation rate (orange lines). Experimental data in (**C**) obtained from Scott*et al* (2010).

# Derivation of constraints on molecular catalytic rates (2)

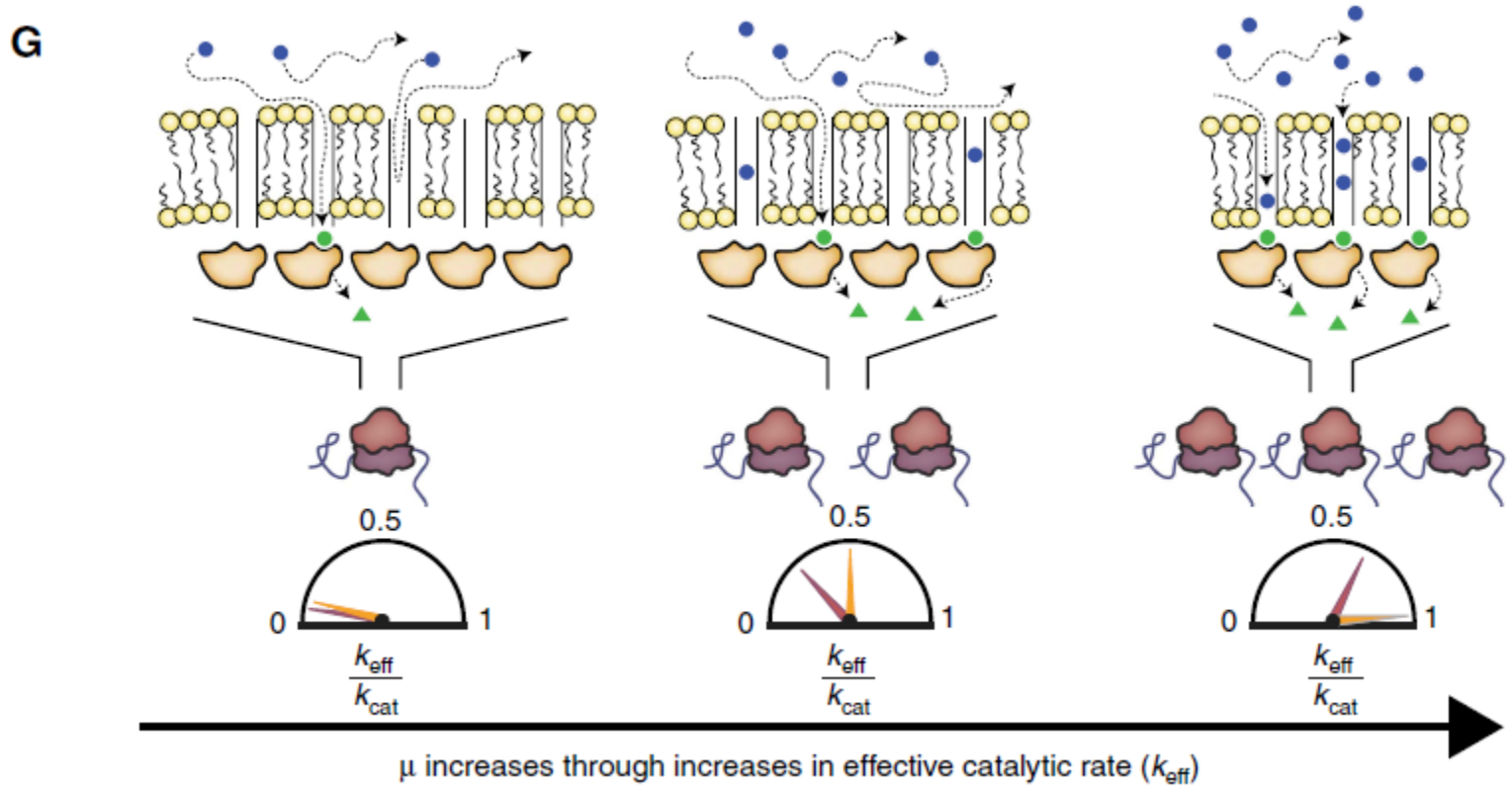



Figure 1. (**G**) The cartoon depicts changes in extra- (blue) and intra- (green) cellular substrate (circle) and product (triangle) concentrations and metabolic enzyme (orange) and ribosome (purple/maroon) levels as the concentration of a growth-limiting nutrient (and growth rate) increases. The dials show $k_{eff}/k_{cat}$, the effective catalytic rate over the maximum for metabolic enzymes (orange) and ribosomes (purple/maroon).

# Derivation of constraints on molecular catalytic rates (3)

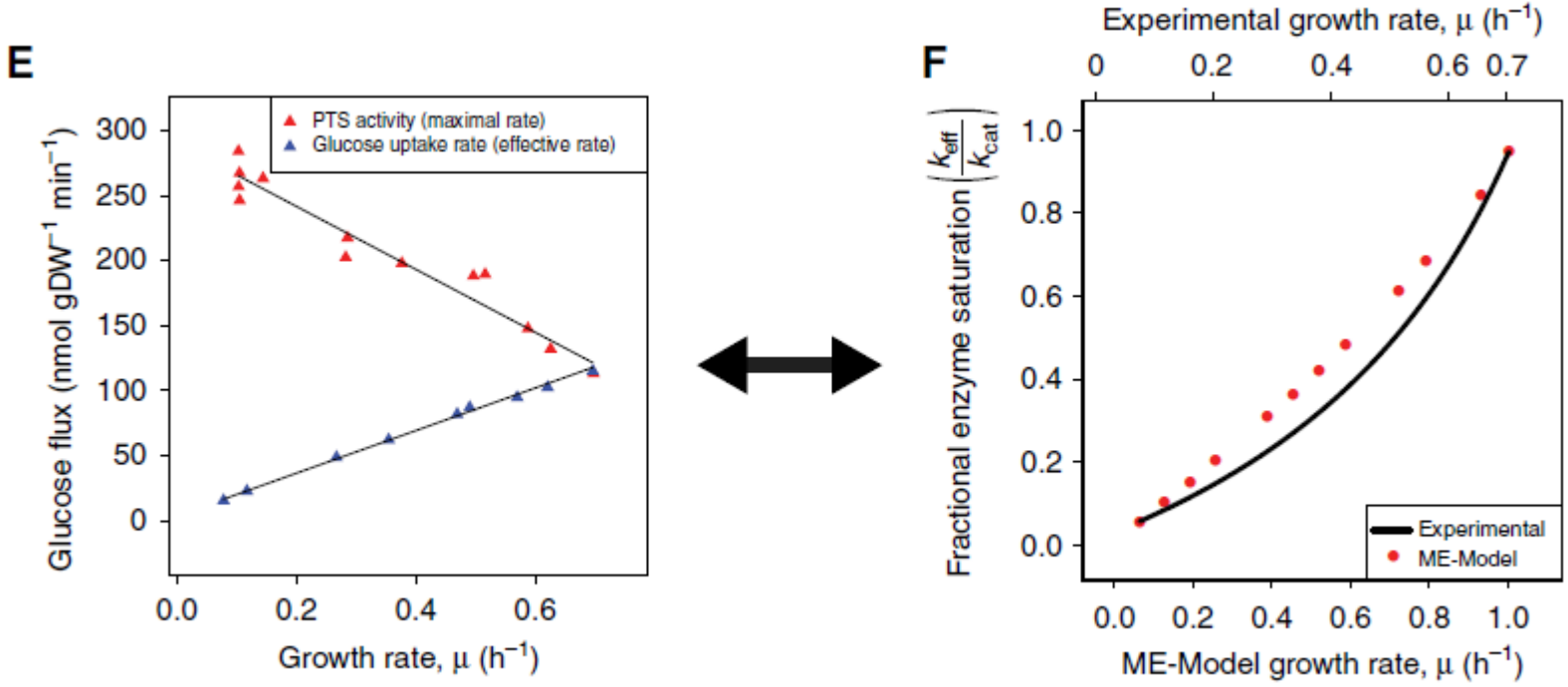



Figure 1. (**E**) Phosphotransferase system (PTS) transient activity following a glucose pulse in a glucose-limited chemostat culture (red) and glucose uptake before the glucose pulse (blue) is plotted as a function of growth rate. The data shown were obtained from O'Brien *et al* (1980)). (**F**) Data from (**E**) are used to plot glucose uptake as a fraction of PTS activity. The resulting value is the fractional enzyme saturation (black line). The fractional enzyme saturation predicted by the ME-Model is plotted as a function of growth rate under carbon limitation (red dots).

# Growth regions under varying nutrient availability

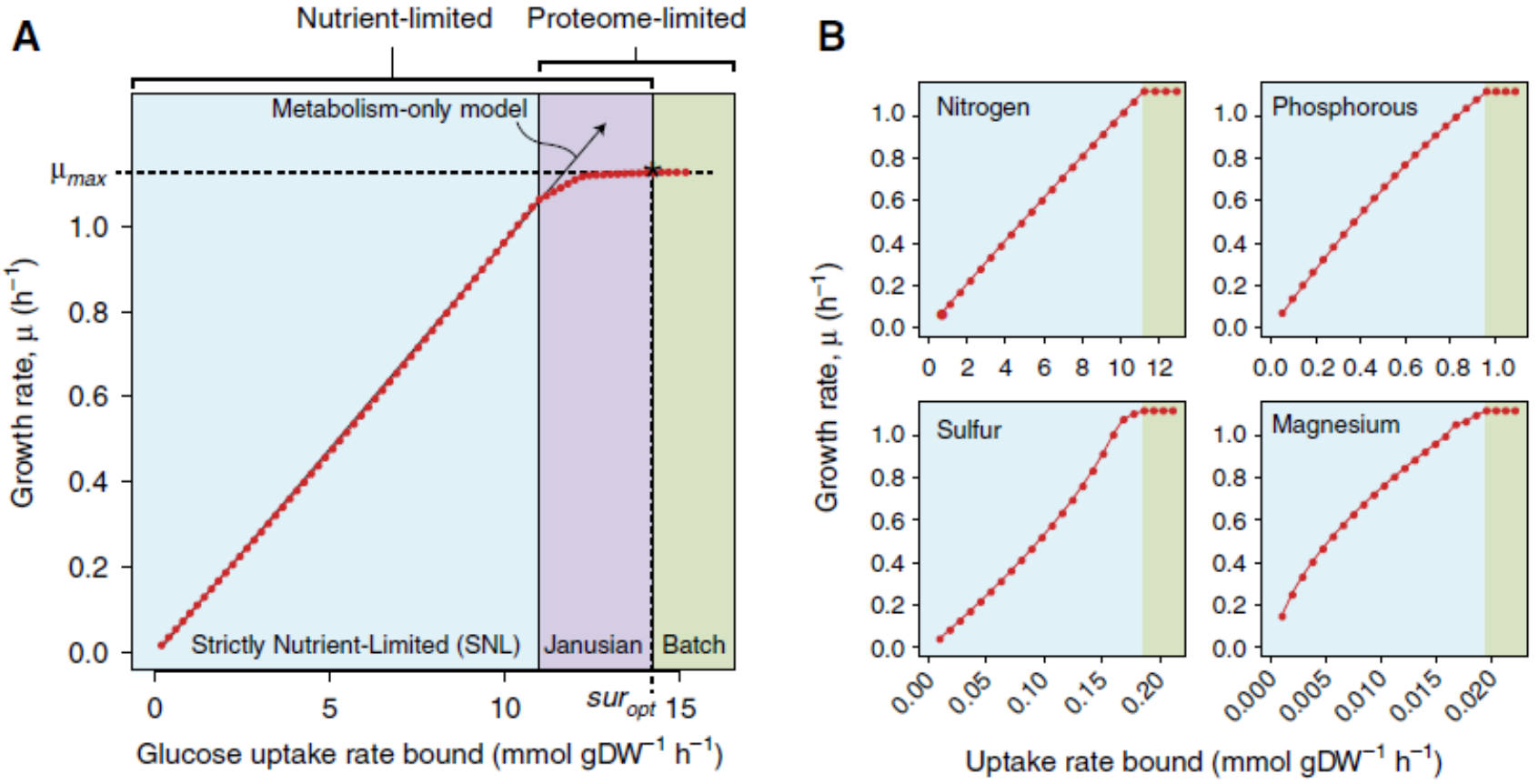



Figure 2. Predicted growth, yield, and secretion. (**A**) Predicted growth rate is plotted as a function of the glucose uptake rate bound imposed in glucose minimal media (**B**) Predicted growth rates as a function of uptake of a limiting nutrient with glucose in excess. The shaded regions correspond to those as labeled in (**A**).

# Effect of proteome limitation on secretion phenotypes

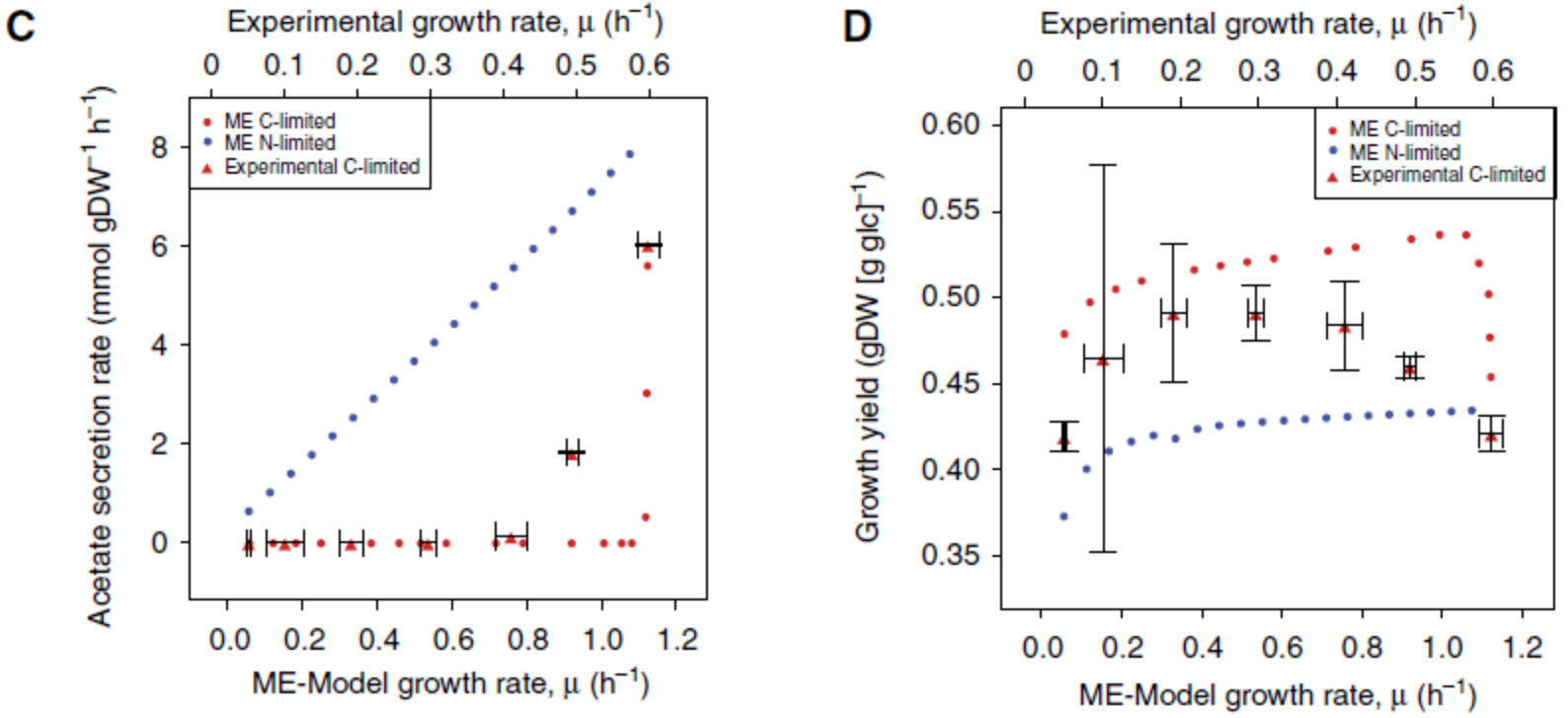



Figure 2. (**C**) Experimental (triangle) and ME-Model-predicted (circle) acetate secretion in Nitrogen- (blue) and Carbon- (red) limited glucose minimal medium are plotted as a function of growth rate. (**D**) Experimental (triangle) and ME-Model-predicted (circle) carbon yield (gDW Biomass/g Glucose) in Carbon- (red) and Nitrogen- (blue) limited glucose minimal medium are plotted as a function of growth rate

# The general behavior in the Janusian growth region

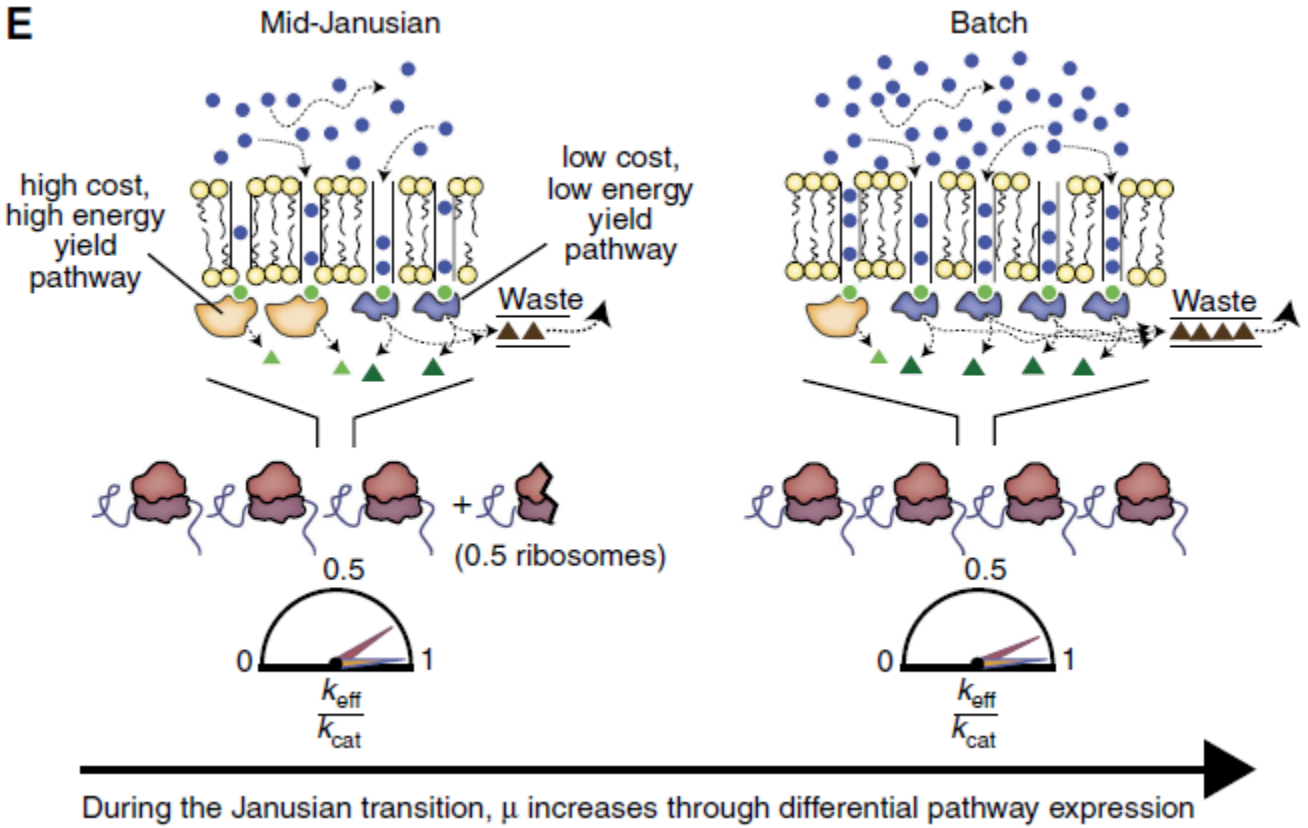



Figure 2. (**E**) The cartoon depicts changes in extra- (blue) and intra- (green) cellular substrate (circle) and product (triangle) concentrations and metabolic enzyme (blue/orange) and ribosome (purple/maroon) levels during the Janusian region. The dials show $k_{eff}/k_{cat}$, the effective catalytic rate over the maximum for metabolic enzymes (blue/orange) and ribosomes (purple/maroon).

Metabolic enzymes are saturated throughout the entire Janusian region. To increase the growth rate, the cell expresses metabolic pathways that have lower operating costs. (Pathways with the smaller blue proteins taken to be 0.25 the cost of the pathways with larger orange proteins.) A higher glucose uptake and turnover results, but energy yield is lower and some carbon is 'wasted' and secreted (brown triangles).

# Central carbon fluxes reflect growth optimization subject to catalytic constraints

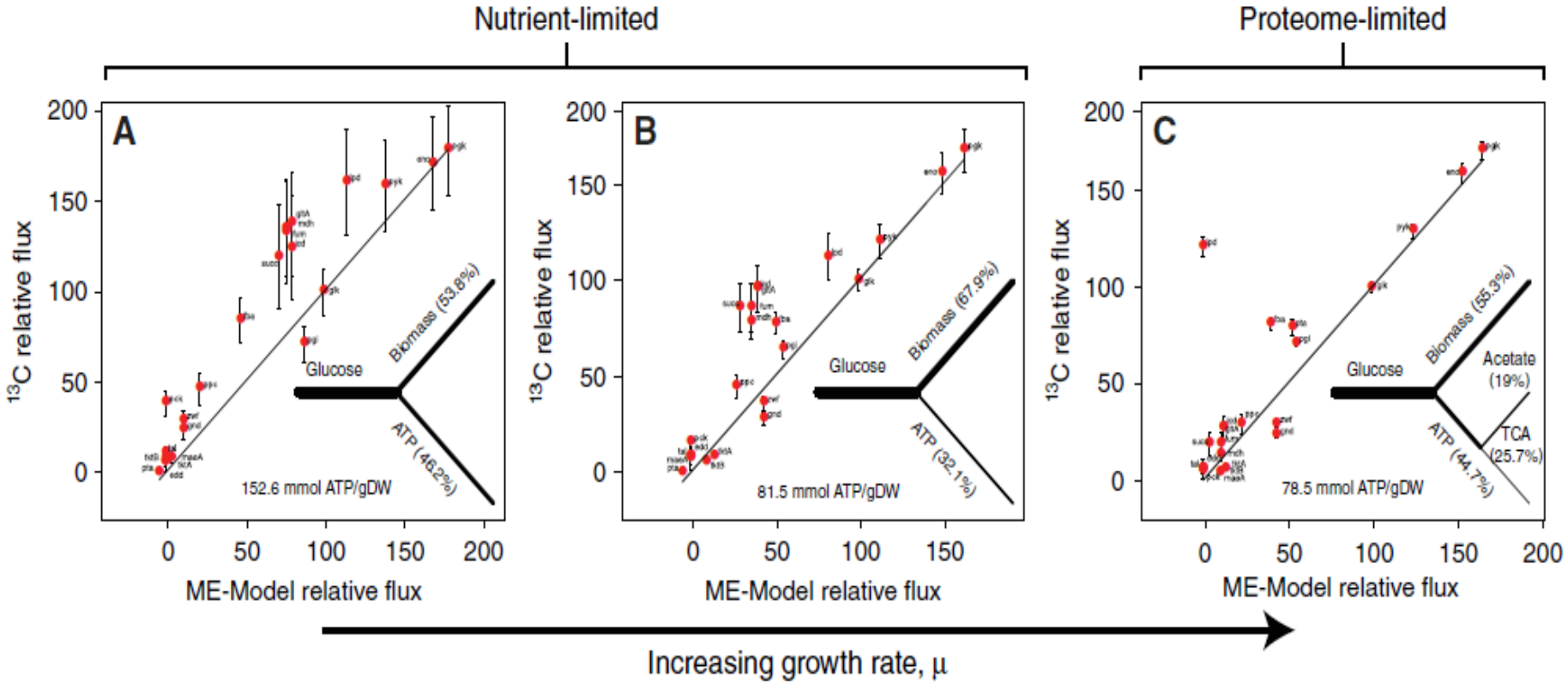



Figure 3. Central carbon metabolic flux patterns under glucose-limited and glucose-excess conditions. (**A–C**) Relative fluxes from $^{13}$C experiments are plotted versus the fluxes predicted by the ME-Model. (**A**, **B**) Comparison of nutrient-limited model solutions with chemostat culture conditions and (**C**) comparison of the batch ME-Model solution with batch culture data.

# Growth rate-dependent gene expression under glucose limitation

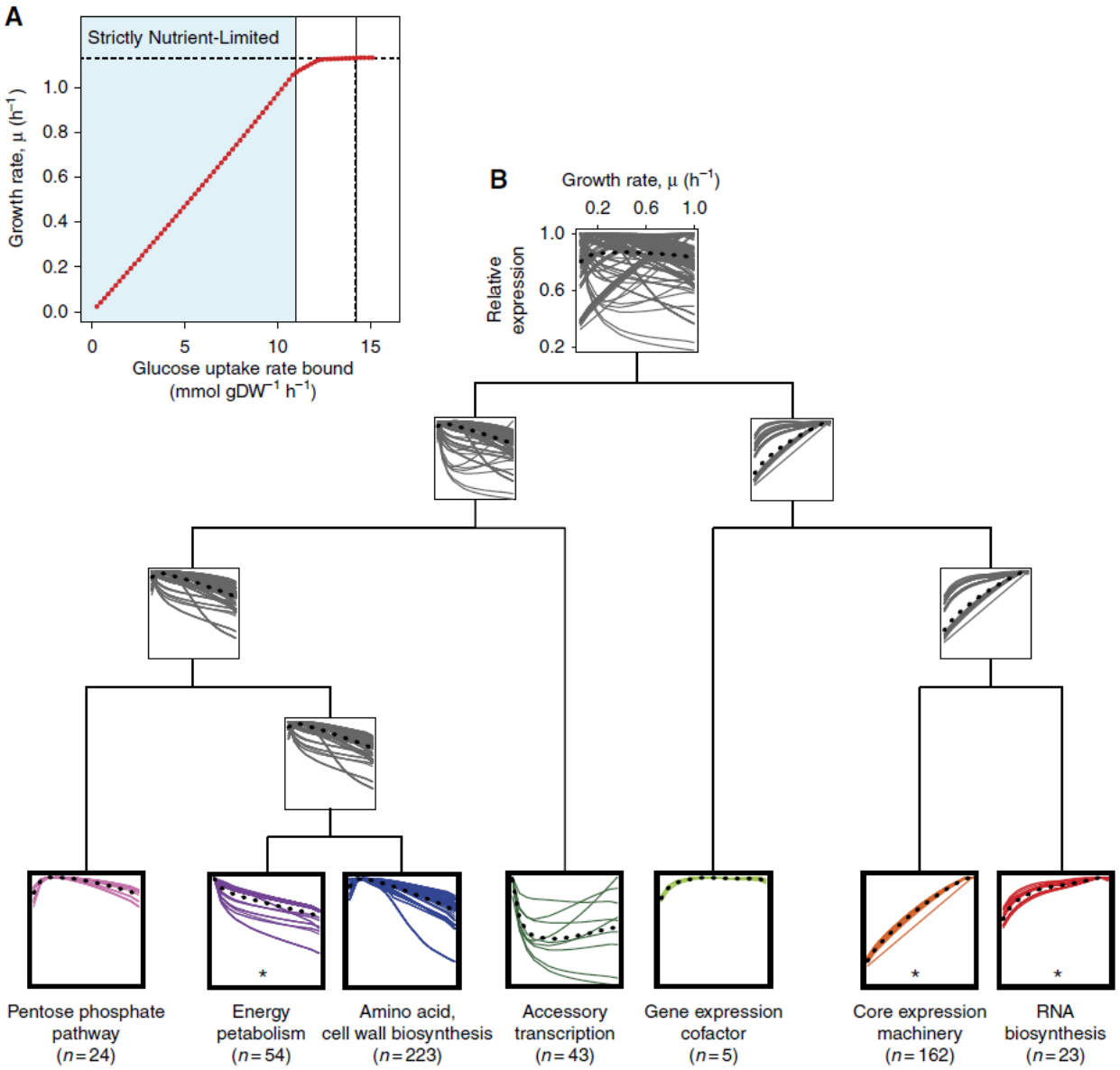
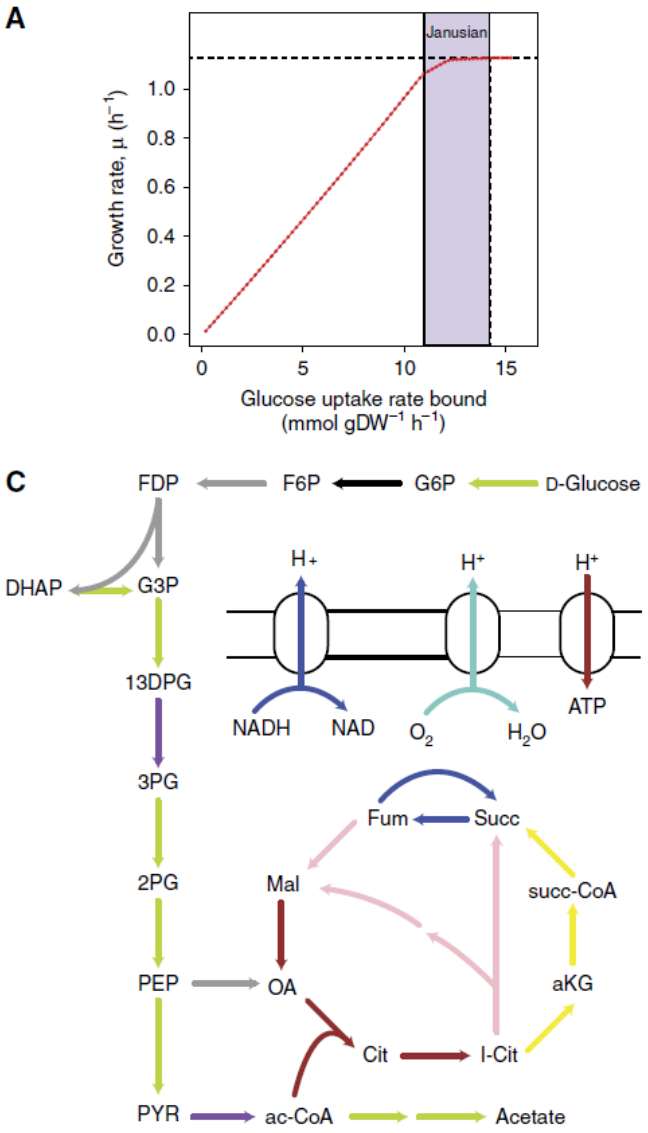



Figure 4. (**A**) Gene expression changes predicted by the ME-Model to occur in the Strictly Nutrient-Limited (SNL) growth region indicated in light blue under glucose limitation in minimal media are analyzed. (**B**) ME-Model-computed relative gene–enzyme pair expression is plotted as a function of growth rate; the normalized *in silico* expression profiles are clustered hierarchically.

# Gene expression during the Janusian region



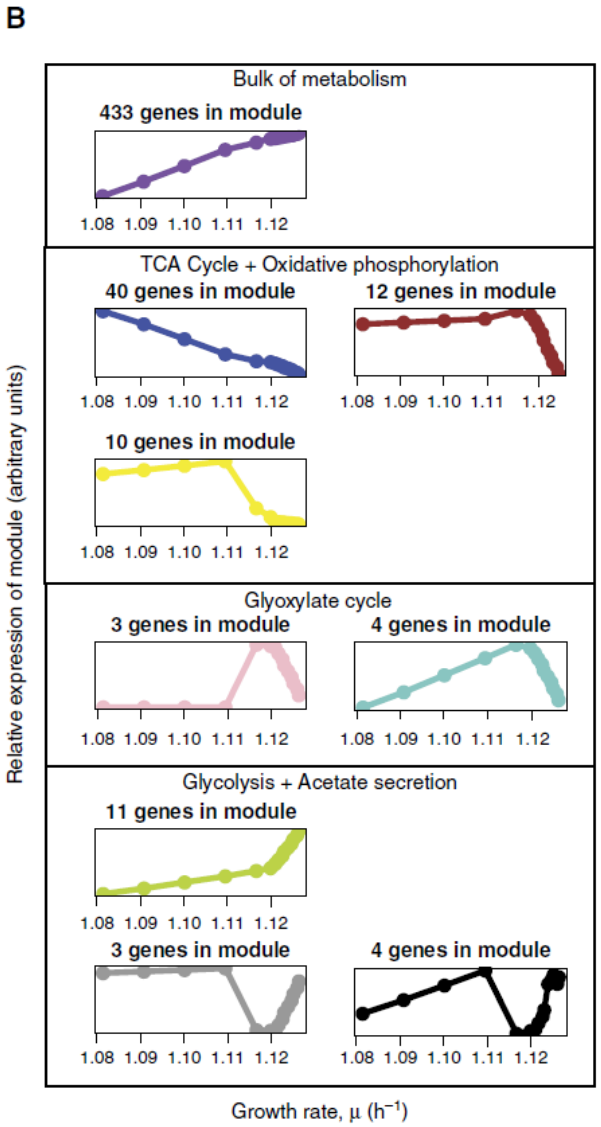




Figure 5. (**A**) Gene expression changes predicted by the ME-Model to occur in the Janusian growth region indicated in purple under glucose limitation in minimal media are analyzed. (**B**) Simulated expression profiles are clustered using signed power (β=25) correlation similarity and average agglomeration (**C**) Many of the expression modules correspond to genes of central carbon energy metabolism. Reactions are colored according to the module color in (**B**).

# Conclusion

- An integrated network of metabolic and gene expression pathways is built for *E. coli*.
- A growth model is developed by adding demands and constraints on molecular catalysis.
- Model yields accurate predictions of growth phenotypes from molecules to whole cell:
  1. the cell's maximum growth rate (μ*) in the specified environment
  2. substrate uptake/by-product secretion rates at μ*
  3. metabolic fluxes at μ*
  4. gene product expression levels at μ*
- A few basic principles underlie growth rate optimization at the systems level.
  - The model predicts three distinct regions of microbial growth, defined by the factors (nutrient and/or proteome) limiting growth.
  - A growth rate-dependent Michaelis–Menten-type model for polymerization speed
  - Proteomic constraints improve predictions of metabolism itself
  - Gene expression changes as the cell transitions through and between the different growth regions

# Take-Home Message

- Because ME-Models explicitly represent gene expression, directly investigating omics data in the context of the whole is now feasible
- ME-model is more intricate then M-model, more room for unknown/incomplete knowledge
  - Lack of specific translation efficacy for each protein
  - Lack of specific degradation rates for each mRNA
  - lack of signaling
  - Lack of regulatory circuitry
- A big challenge for Plant system