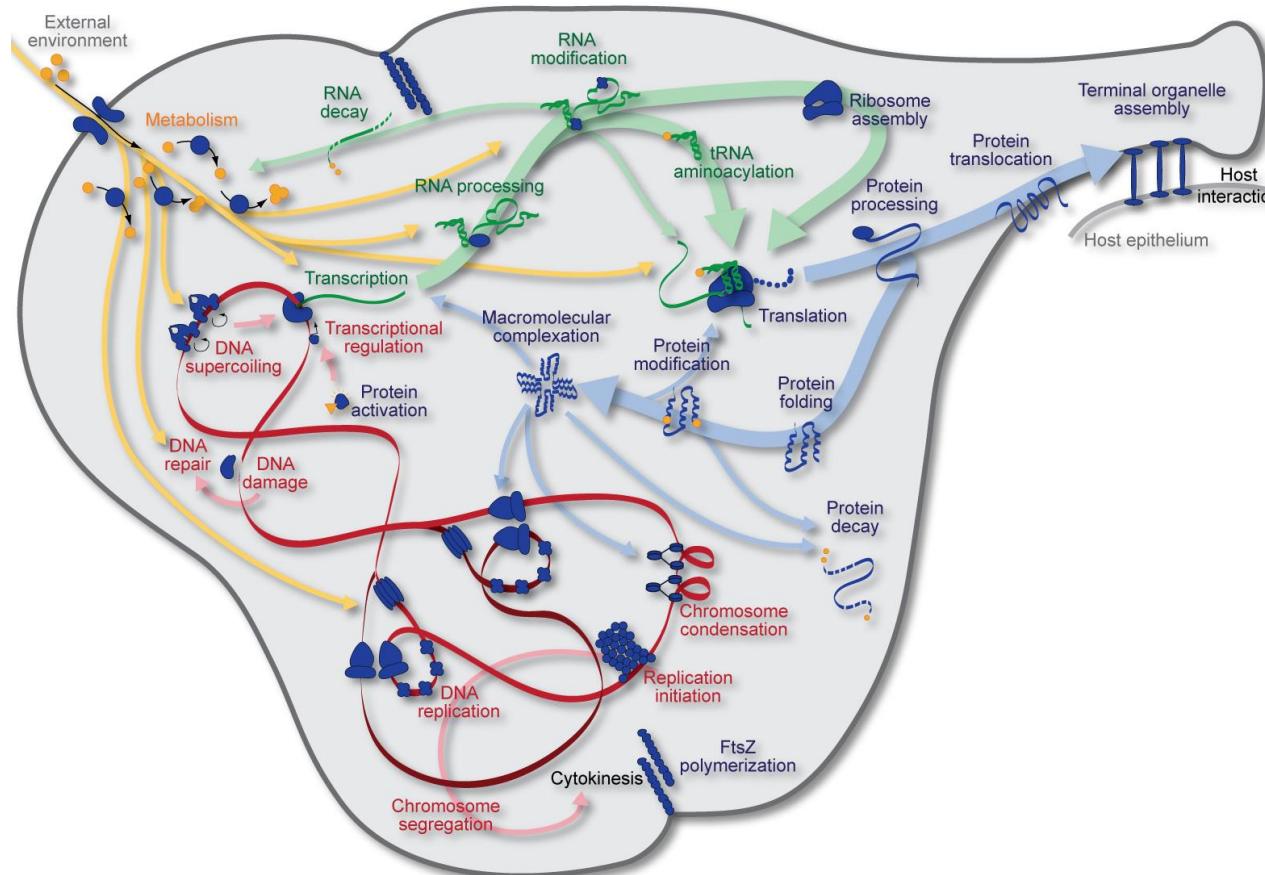


Introduction to whole-cell modeling



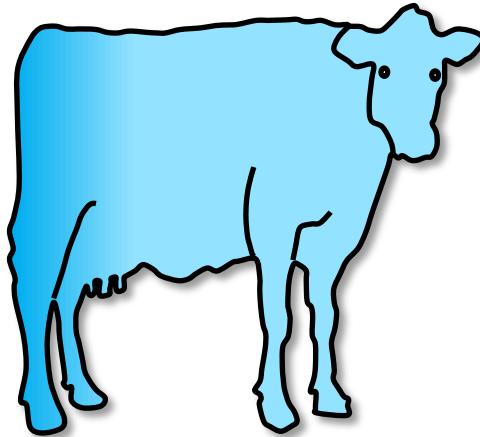
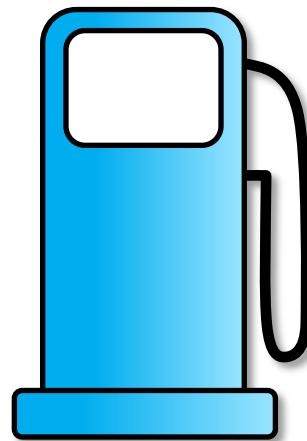
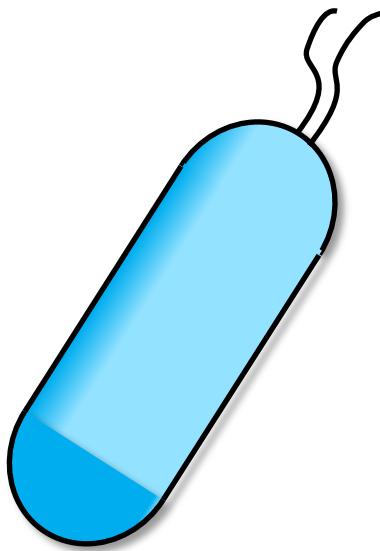
Outline

- Motivation
- What is a WC model?
- Challenges
- Model building
- State-of-the-art
- Software tools
- Open challenges

Outline

- Motivation
- What is a WC model?
- Challenges
- Model building
- State-of-the-art
- Software tools
- Open challenges

Motivation



Basic science

What is life?

How does life emerge?

Industrial manufacturing

Optimize yield

Minimize cost

Agriculture

Optimize yield

Resist drought

Prevent infection

Medicine

Predict prognoses

Optimize therapy

Maximize quality of life

Central challenges

- Understand how genotype determines phenotype
- Design genomes which desired phenotypes

Barriers to bioengineering and medicine

Current approaches rely on trial-and-error

- No complete understanding of biology
- Cannot predict behavior

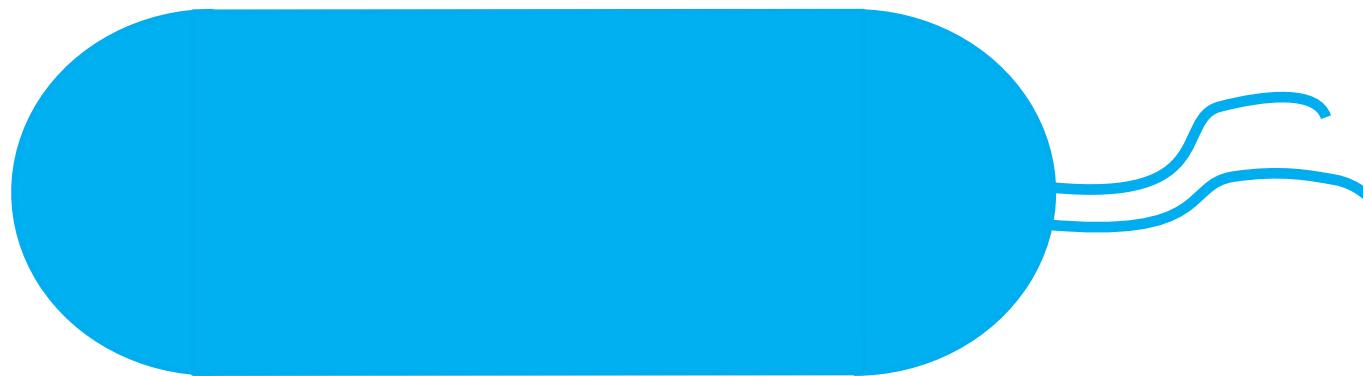
Statistical models rely on observations

- Impossible to observe all mutations

Physics-based models enable rational engineering

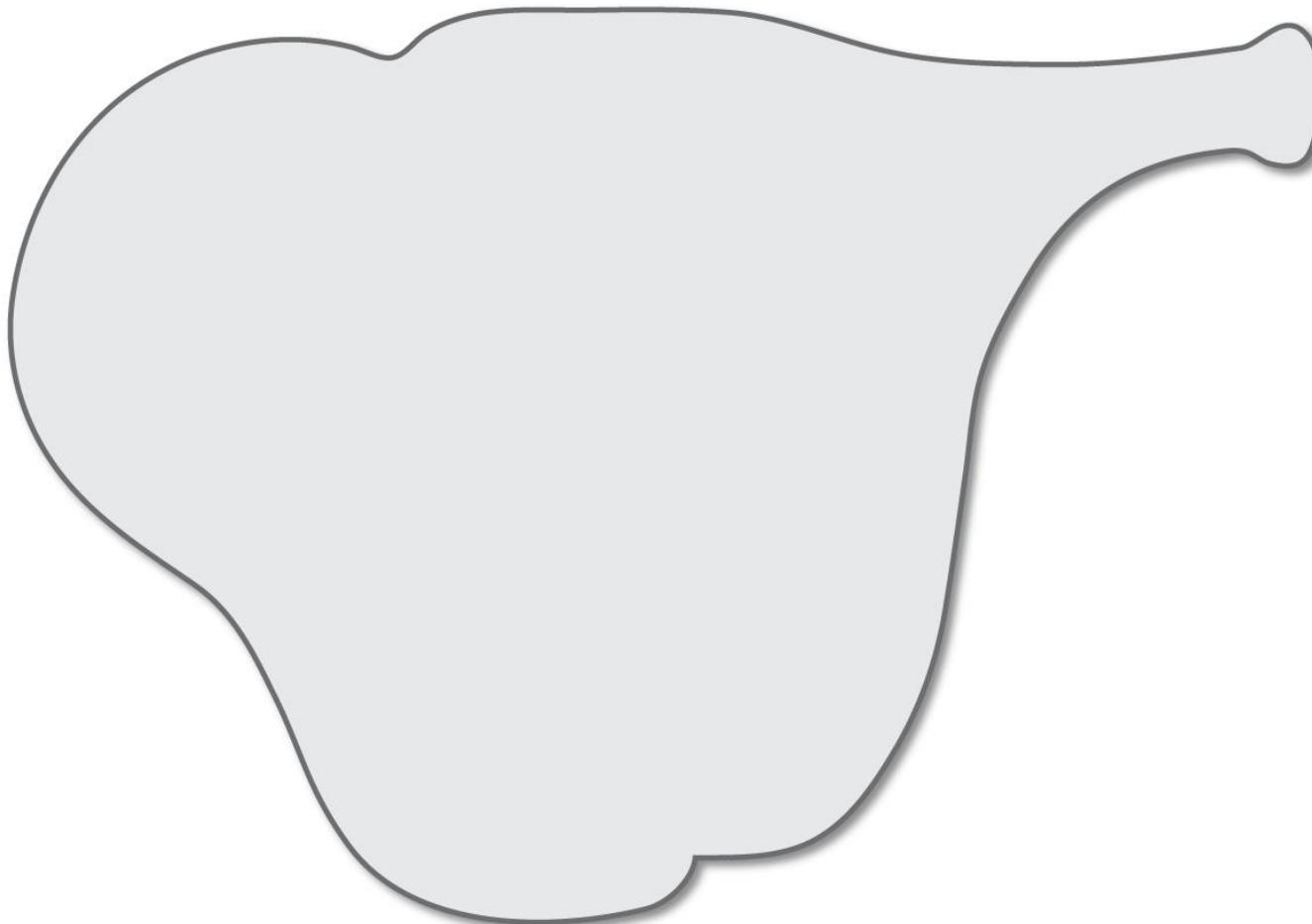


Planes are designed with CAD and physics-based models

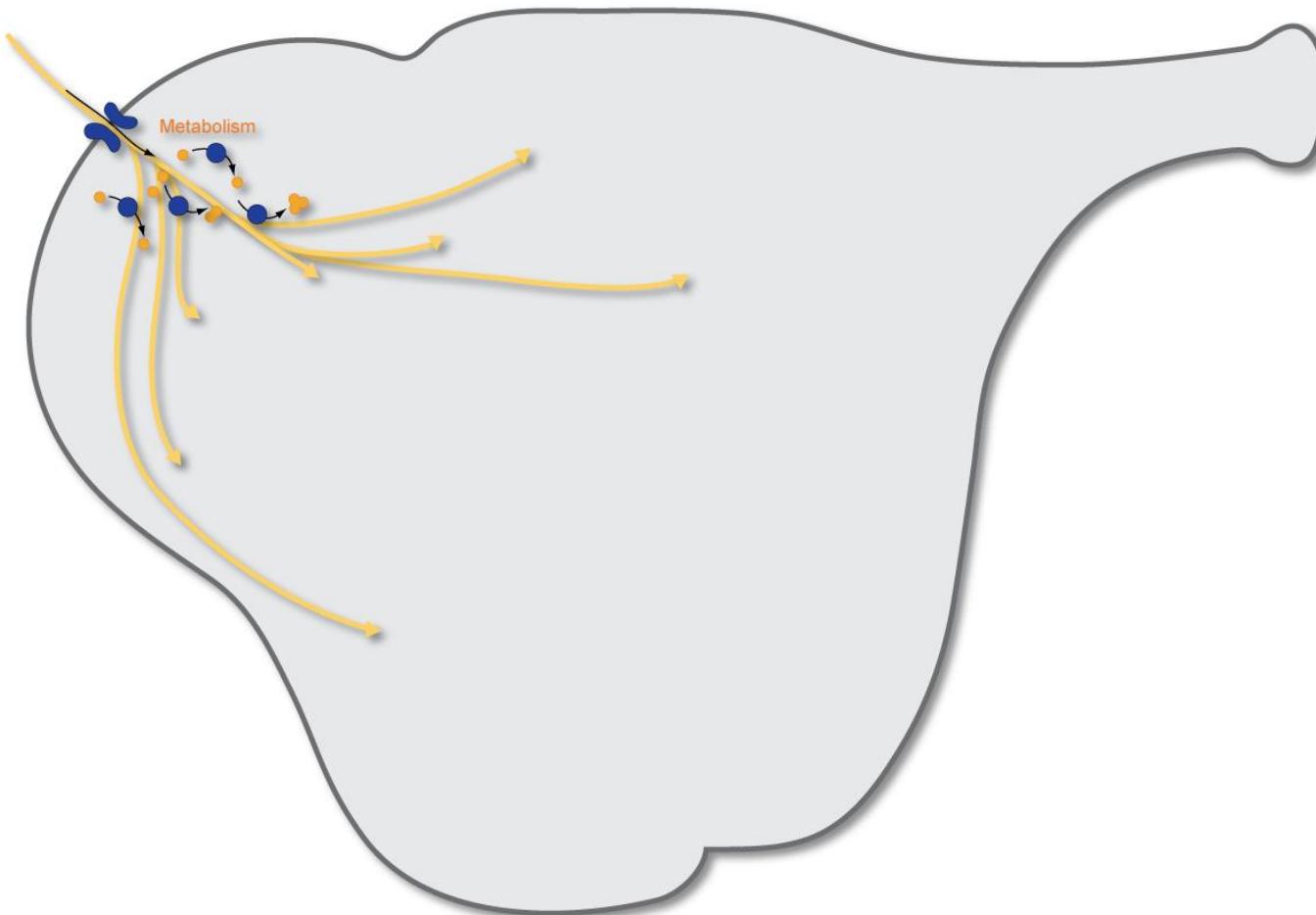


Cells could be designed with WC models

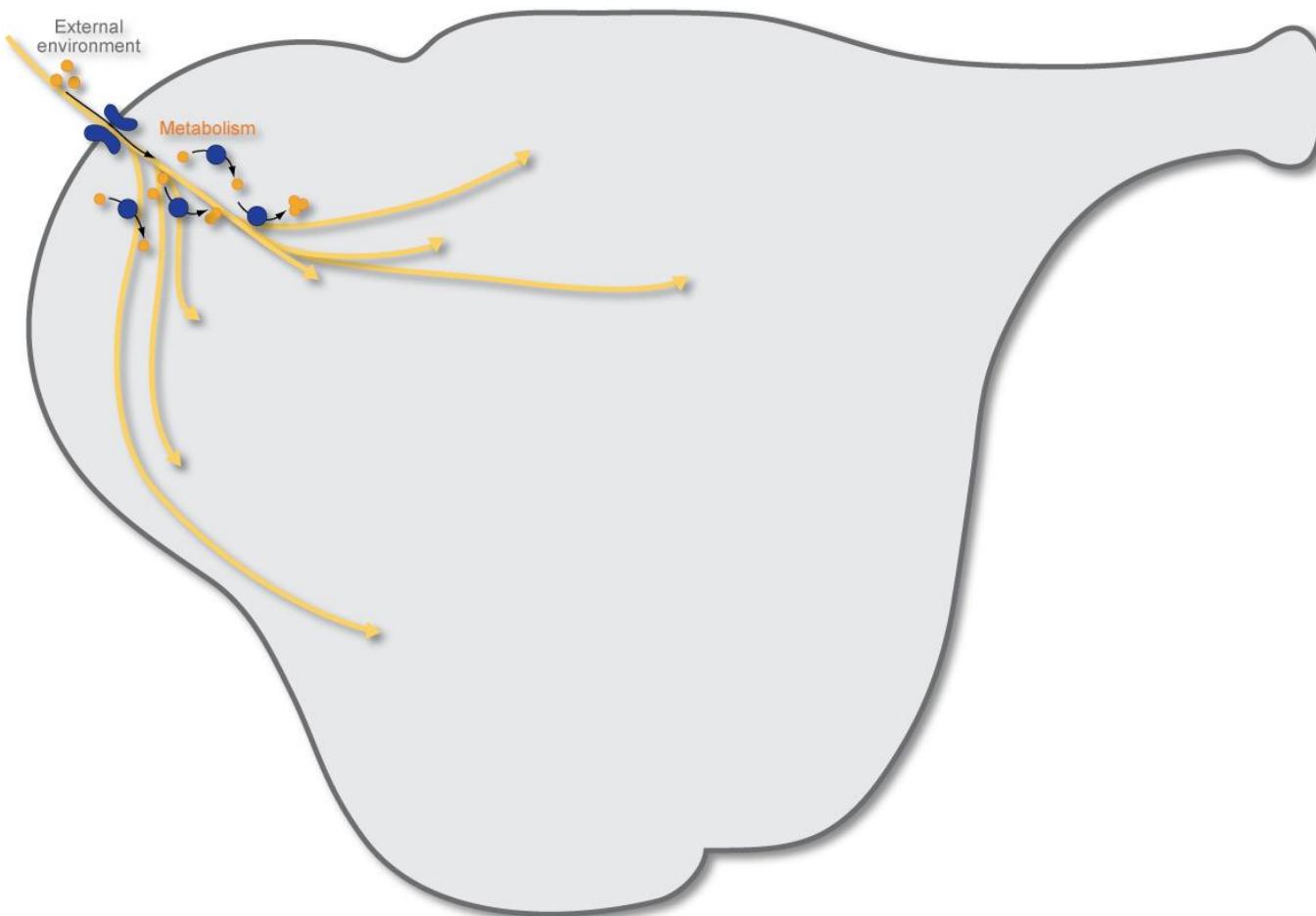
Example: drug biosynthesis



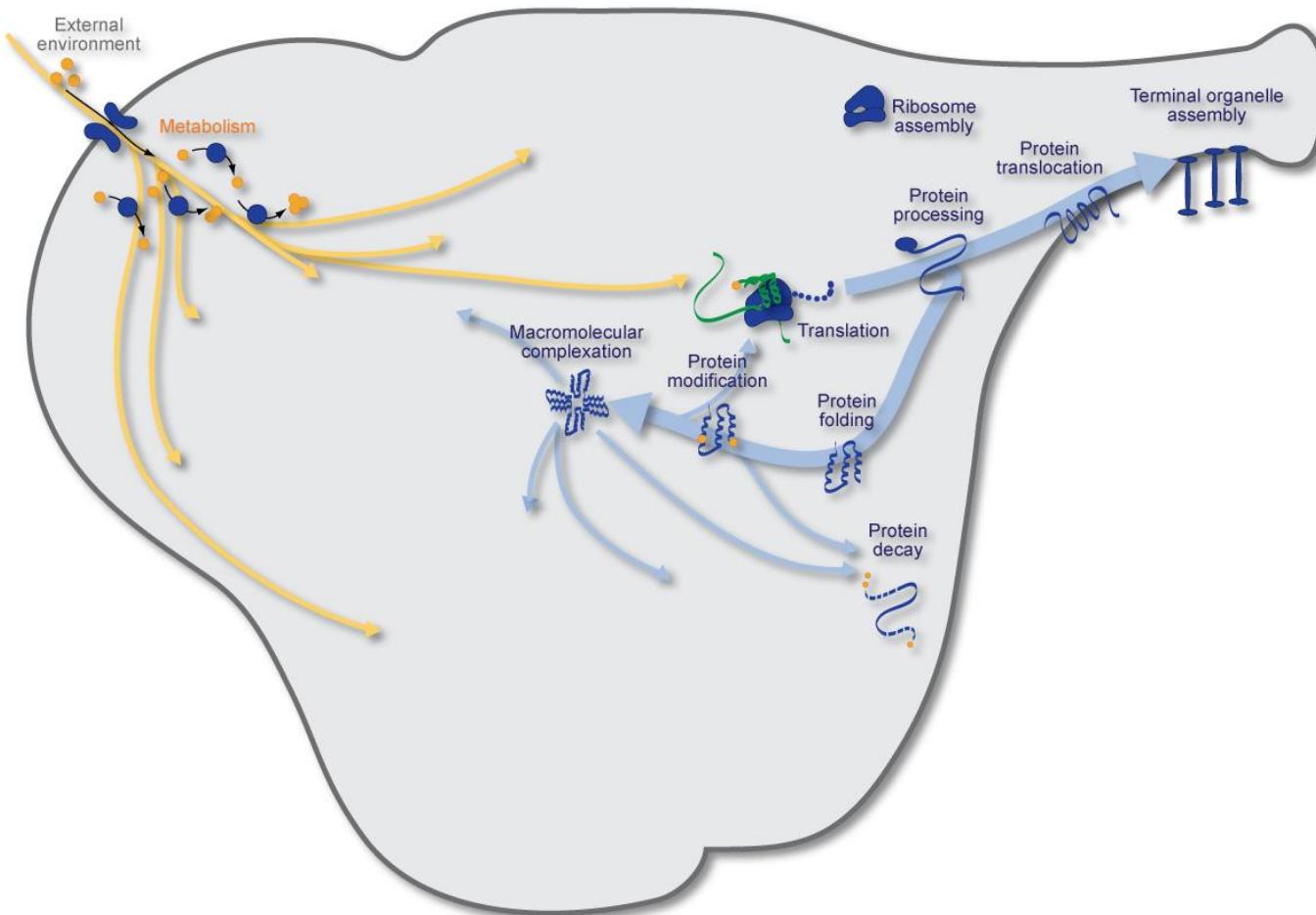
Example: drug biosynthesis



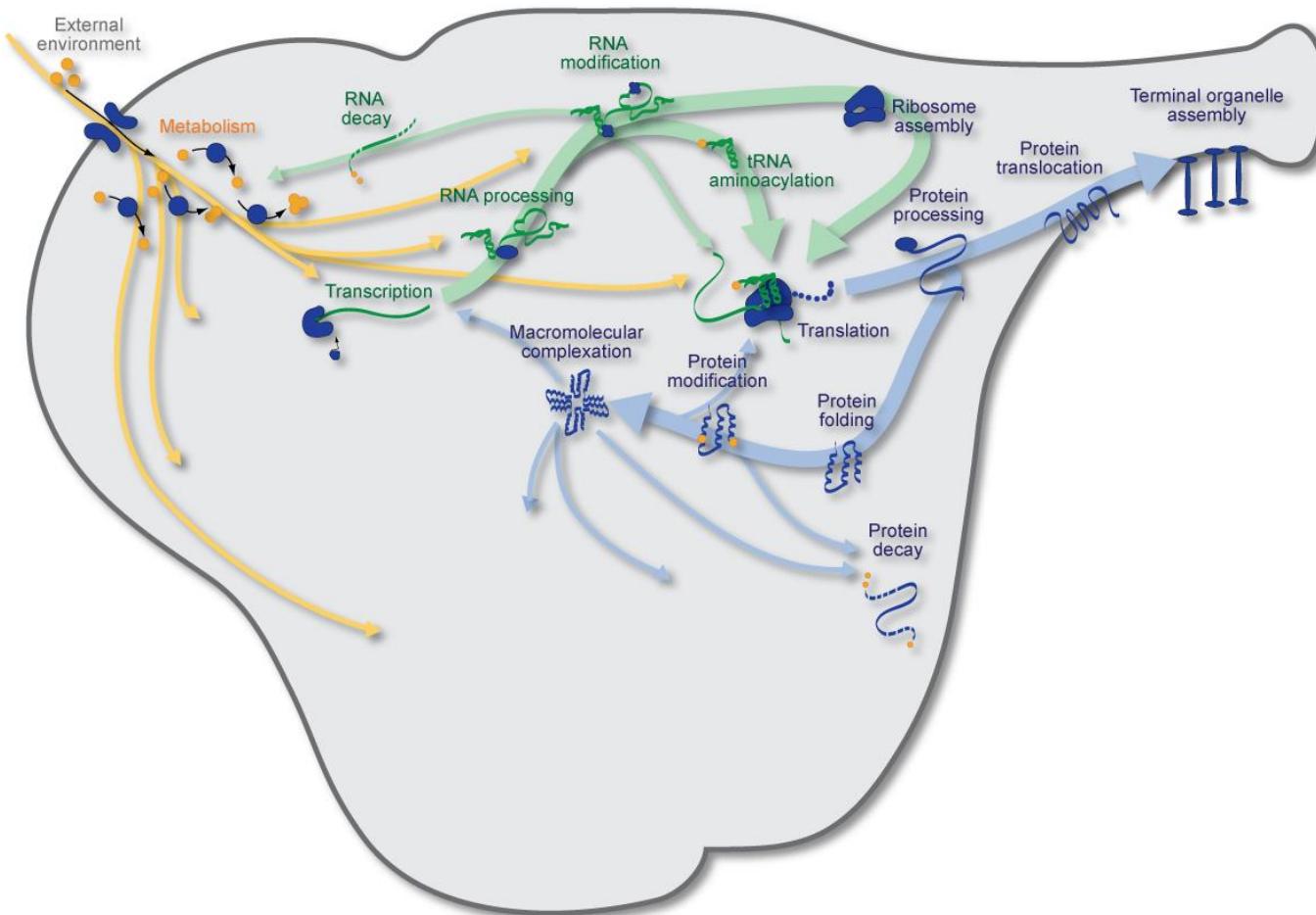
Example: drug biosynthesis



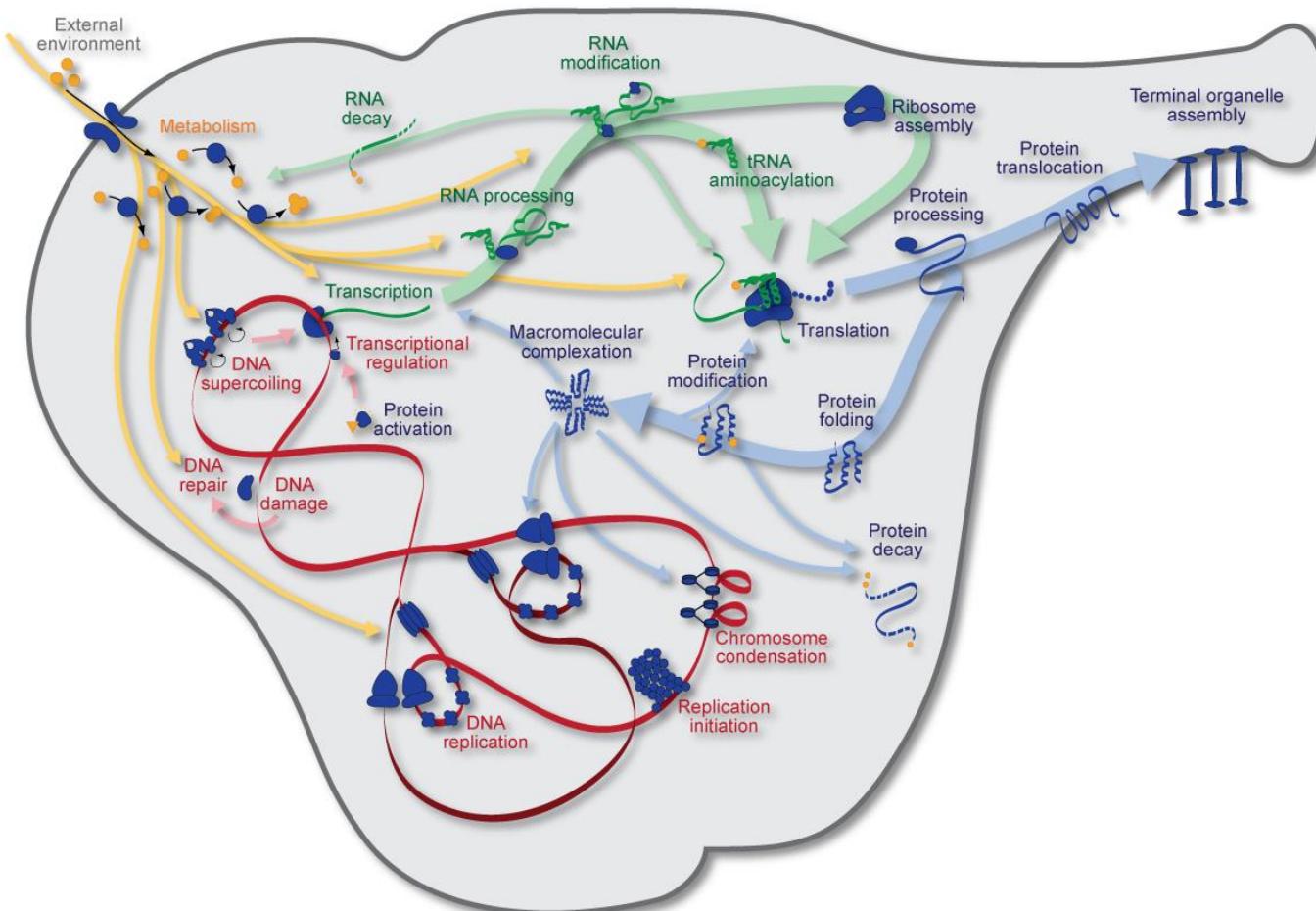
Example: drug biosynthesis



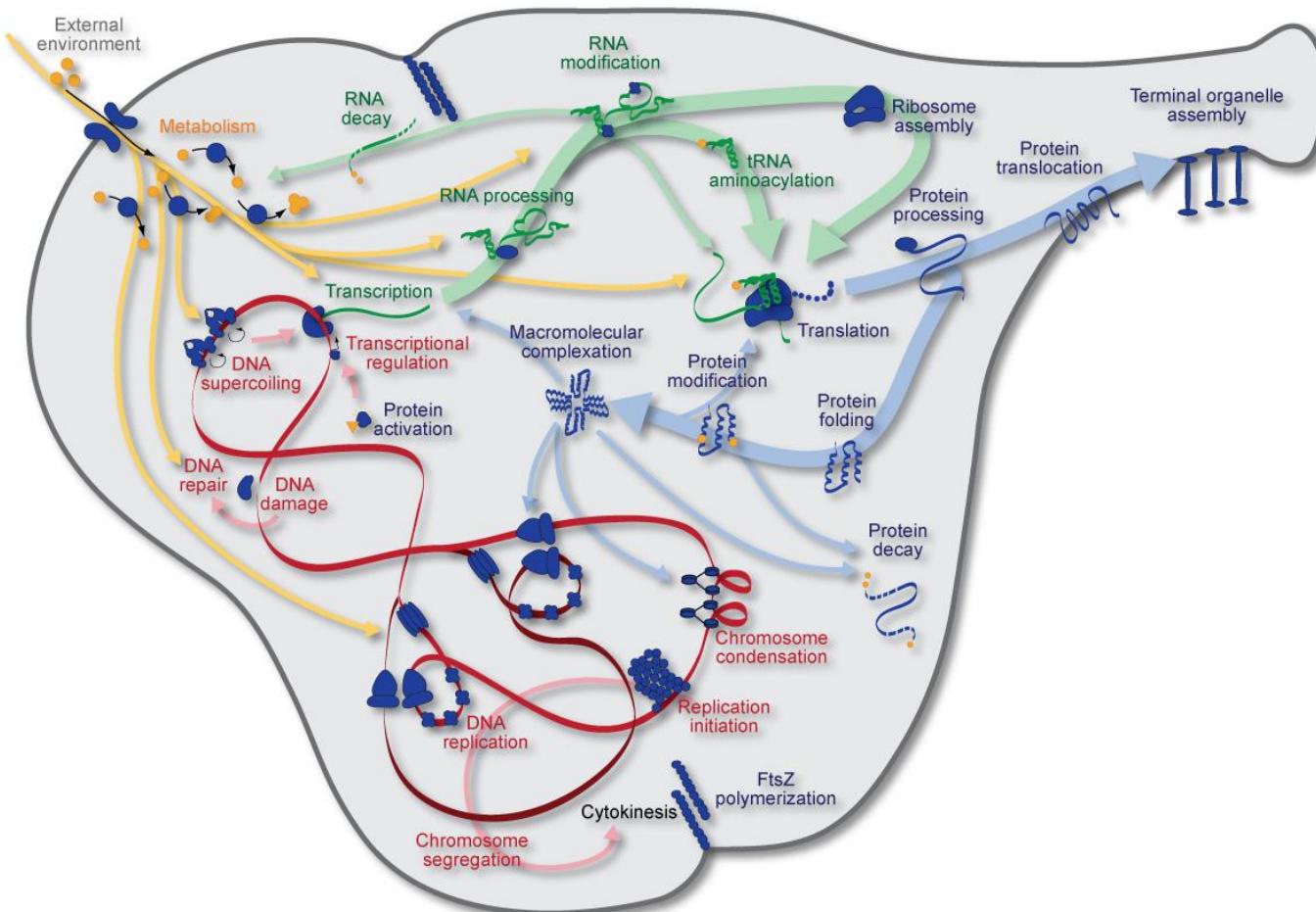
Example: drug biosynthesis



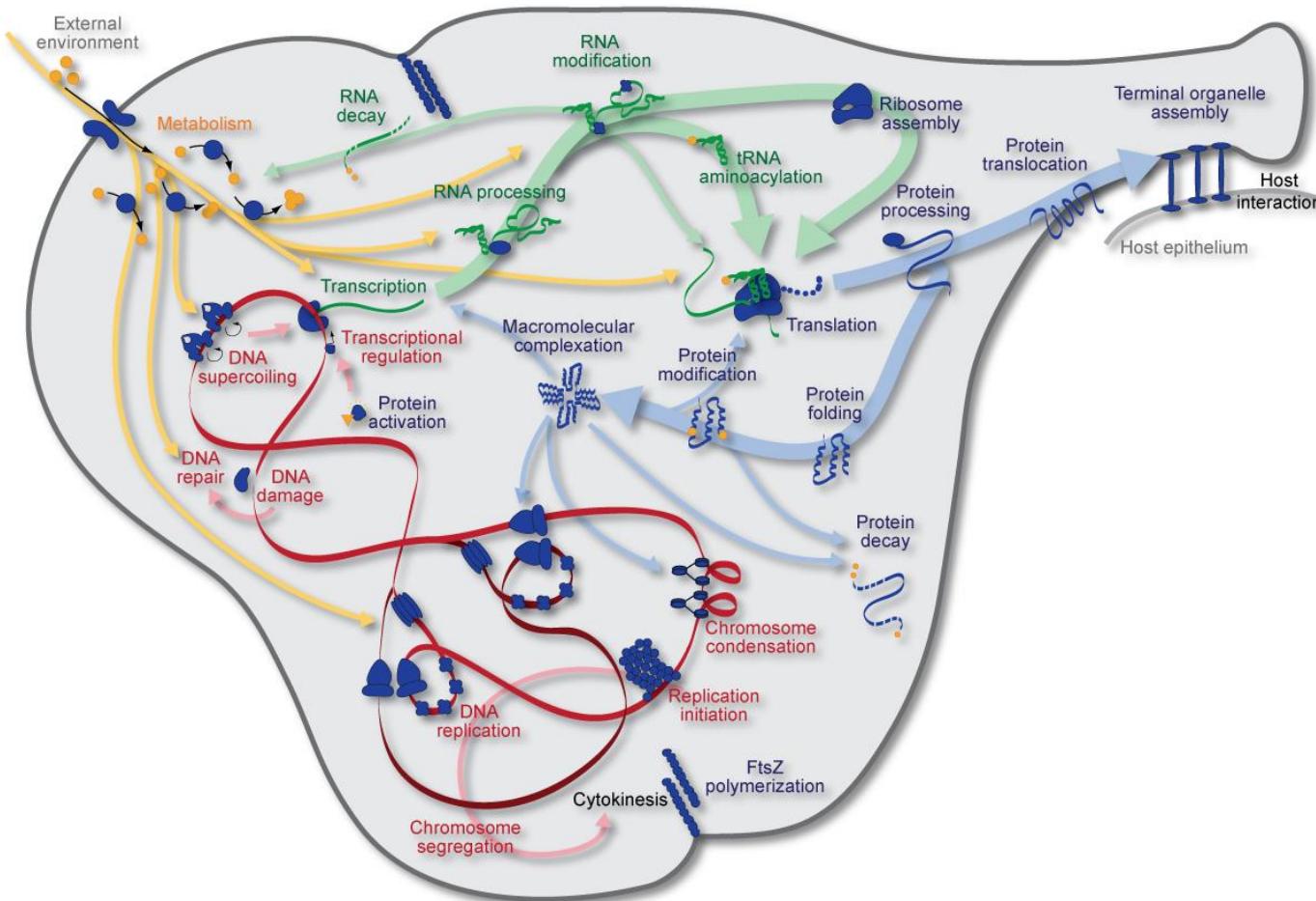
Example: drug biosynthesis



Example: drug biosynthesis



Example: drug biosynthesis



Designing genomes with specific phenotypes requires comprehensive WC models

Outline

- Motivation
- What is a WC model?
- Challenges
- Model building
- State-of-the-art
- Software tools
- Open challenges

WC modeling principles



Single-cell



Temporally complete

GATCCA

Species-specific



Mechanistic/dynamic



Genetically complete



Stochastic



Molecularly precise



Accurate

WC modeling

The ultimate test of understanding a simple cell,
more than being able to build one, would be to build
a computer model of the cell

– *Clyde Hutchison, 1999*

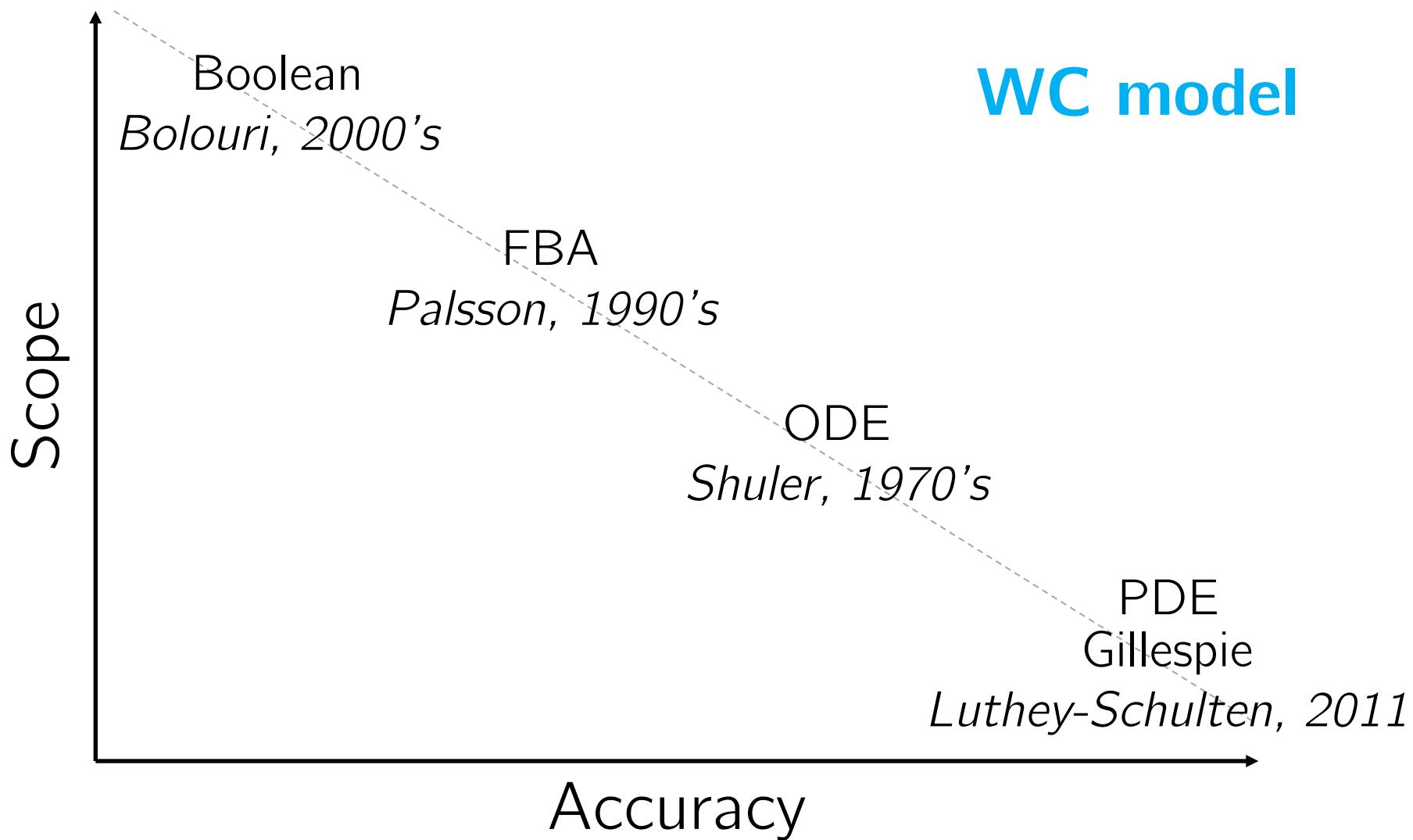
A grand challenge of the 21st century

– *Masaru Tomita, 2001*

Biology urgently needs a theoretical basis to unify it

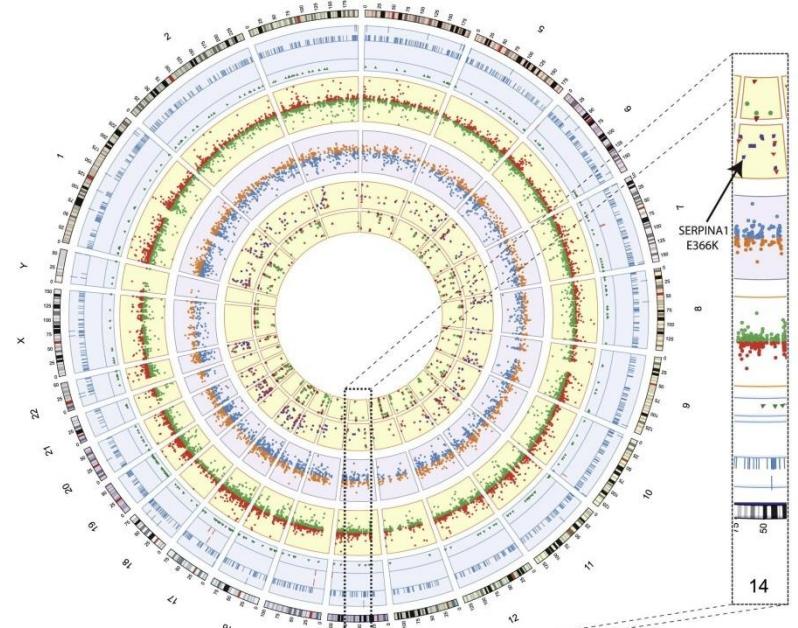
– *Sydney Brenner, 2010*

Predictive modeling methodologies

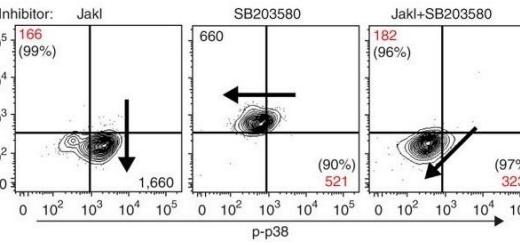
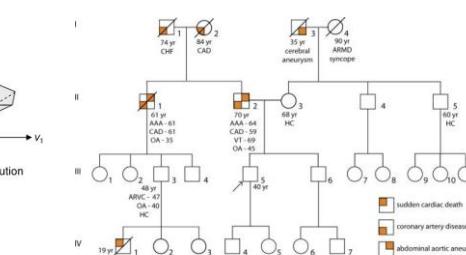
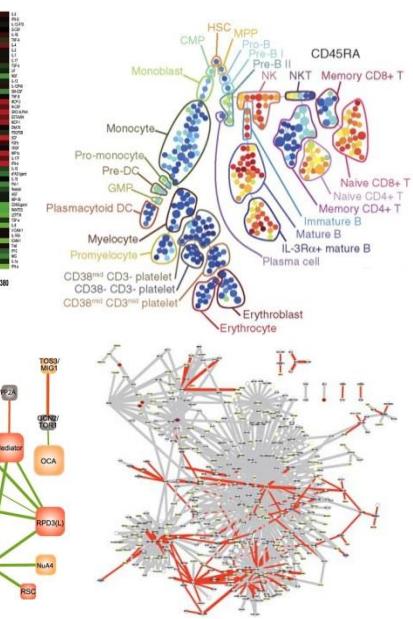
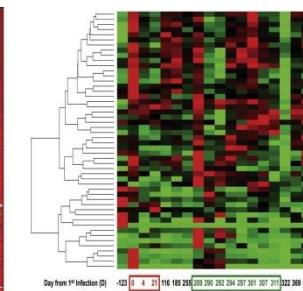
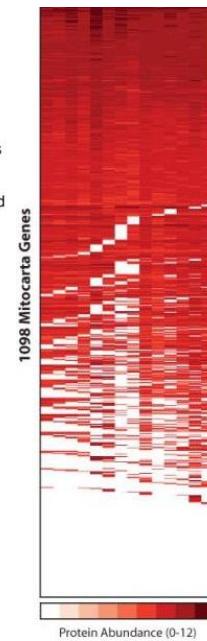
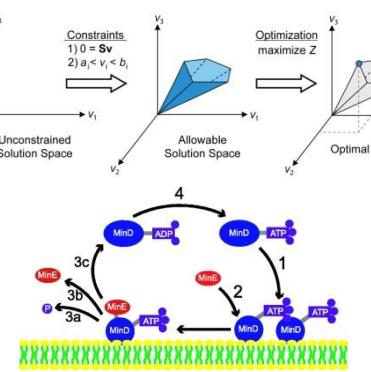
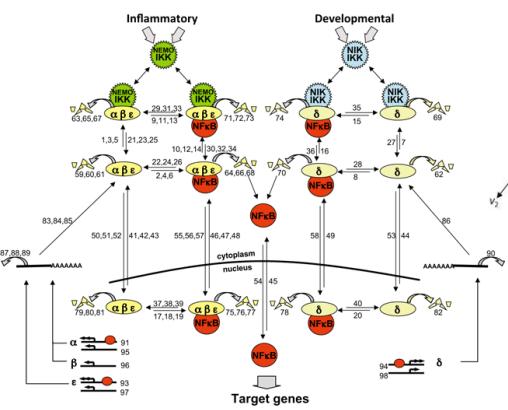


WC modeling was limited by data and simulation algorithms

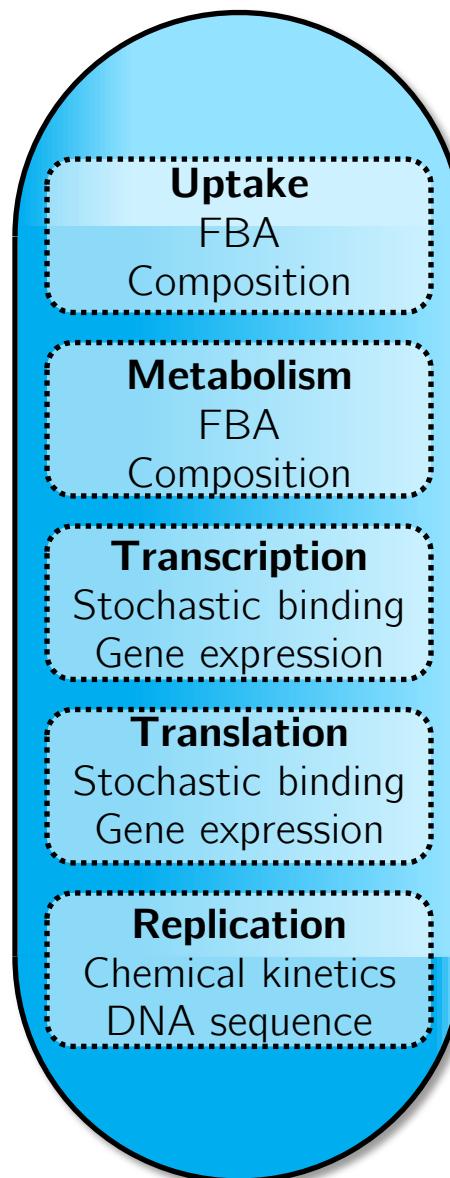
Molecular data is now available



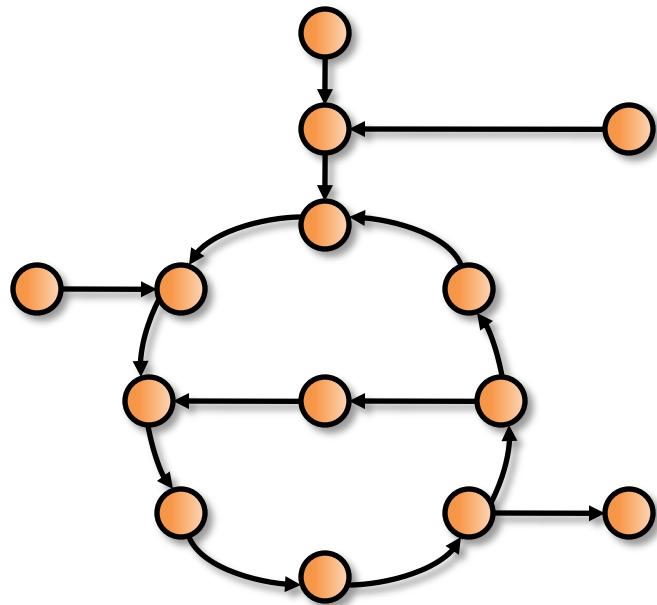
RNA Edits
Heteroallelic SNVs
Protein-Downregulated (HRV vs Healthy)
Protein-Upregulated (HRV vs Healthy)
RNA-Downregulated (HRV vs Healthy)
RNA-Upregulated (HRV vs Healthy)
Indels
SV-Duplications
SV-Deletions
Chr. Ideogram
Chr. Number



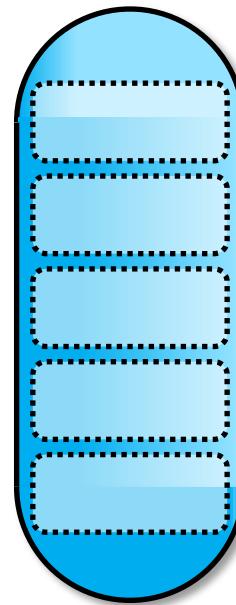
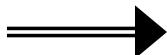
Integrative modeling methods are now available



WC modeling is now feasible

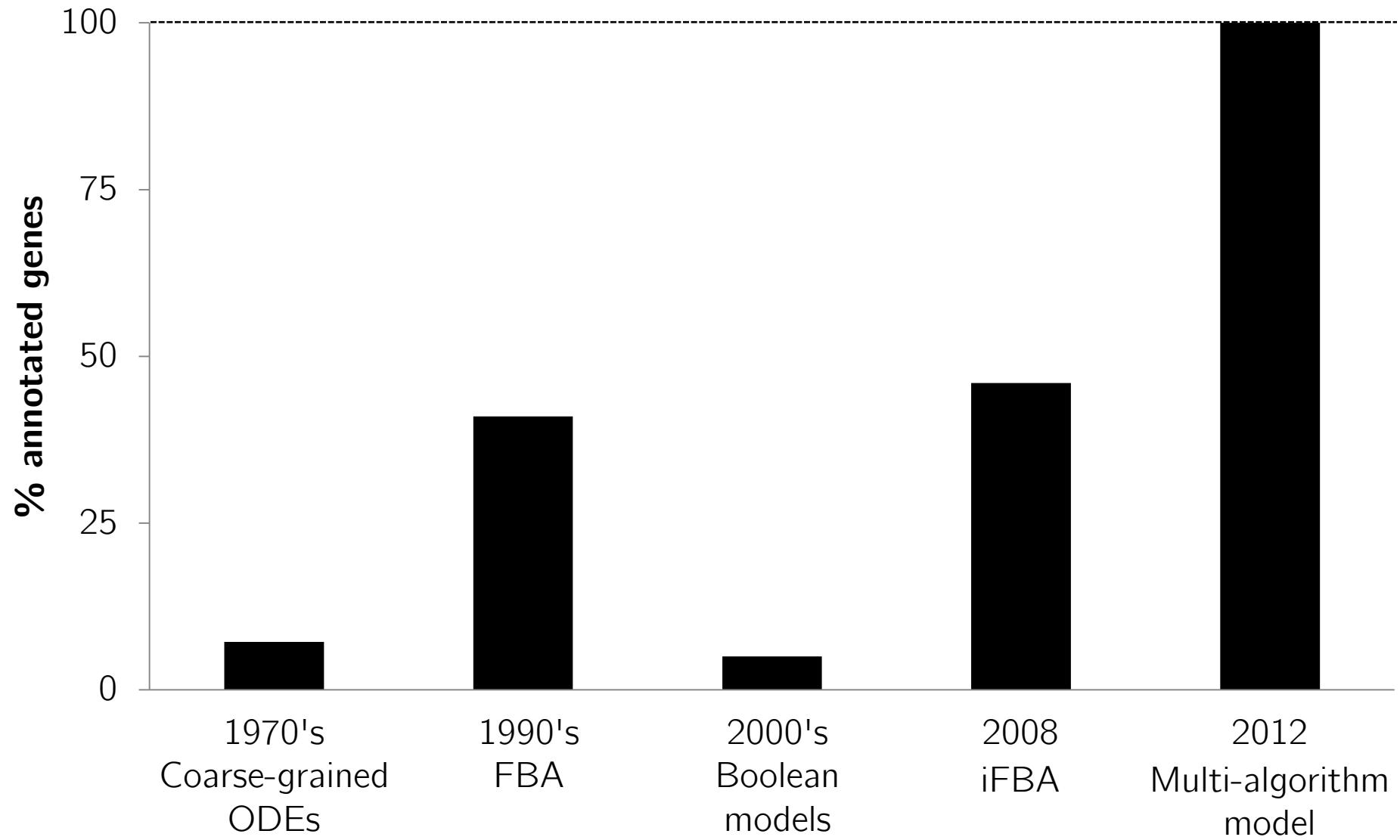


Molecular data



Integrative model

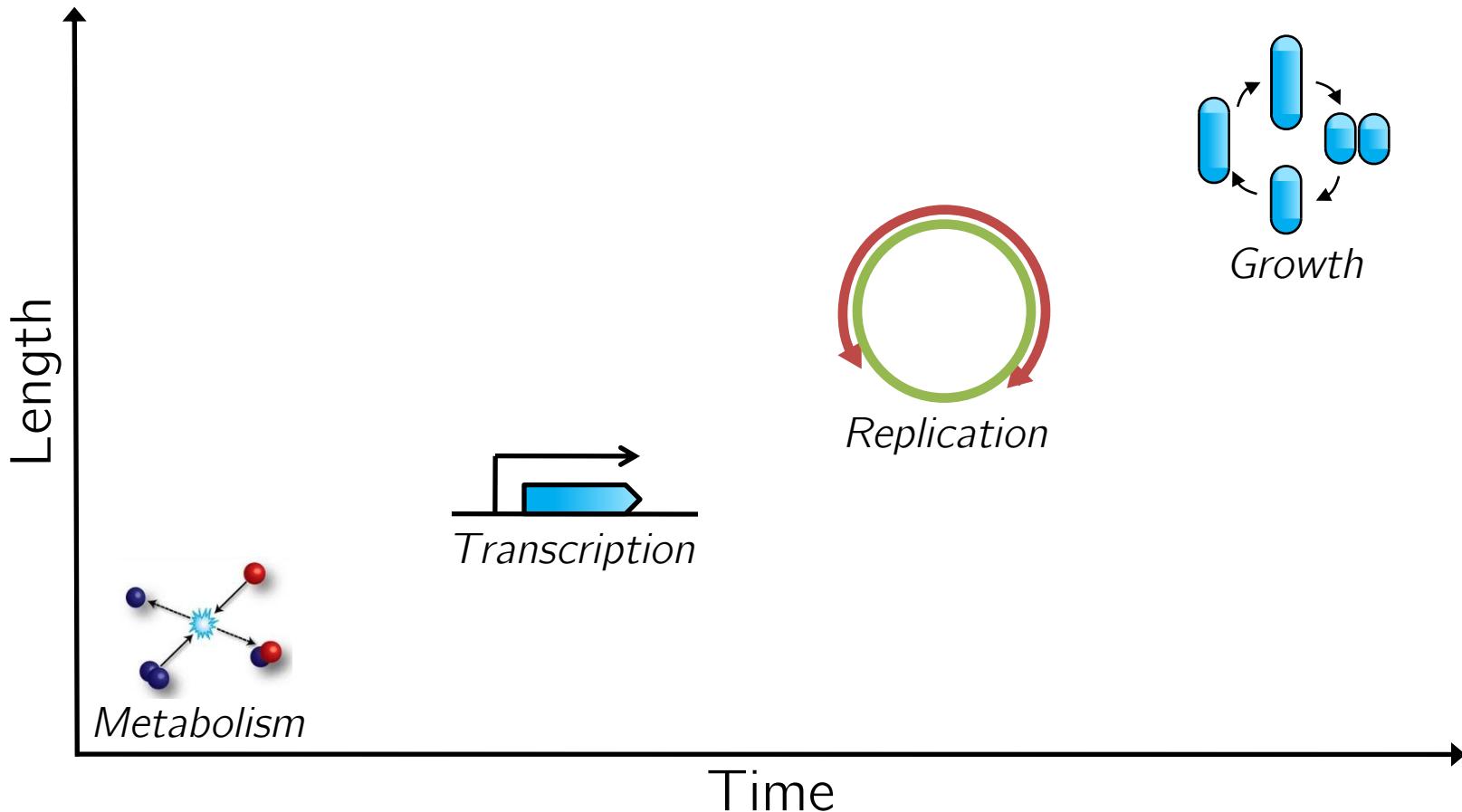
WC modeling progress



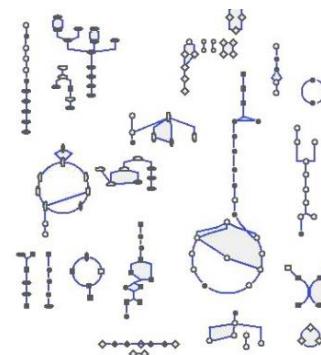
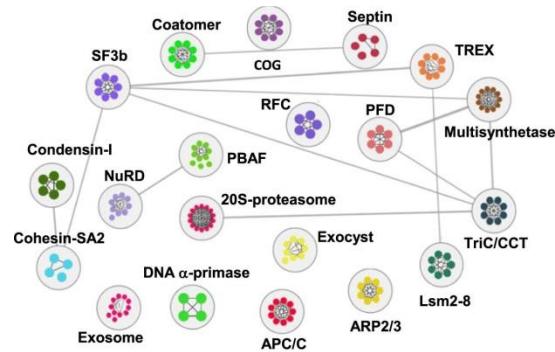
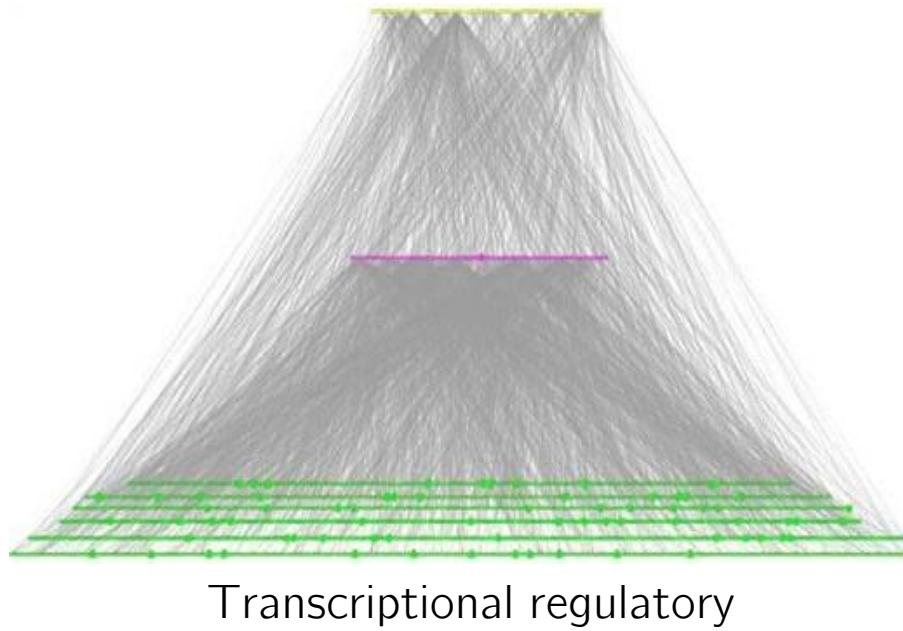
Outline

- Motivation
- What is a WC model?
- Challenges
- Model building
- State-of-the-art
- Software tools
- Open challenges

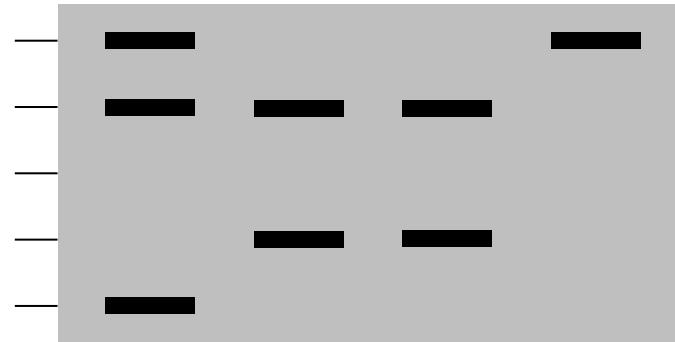
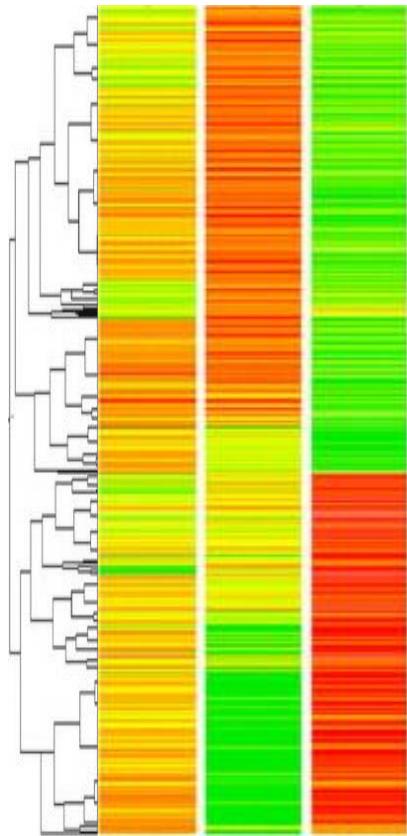
Challenge: multiple time and length scales



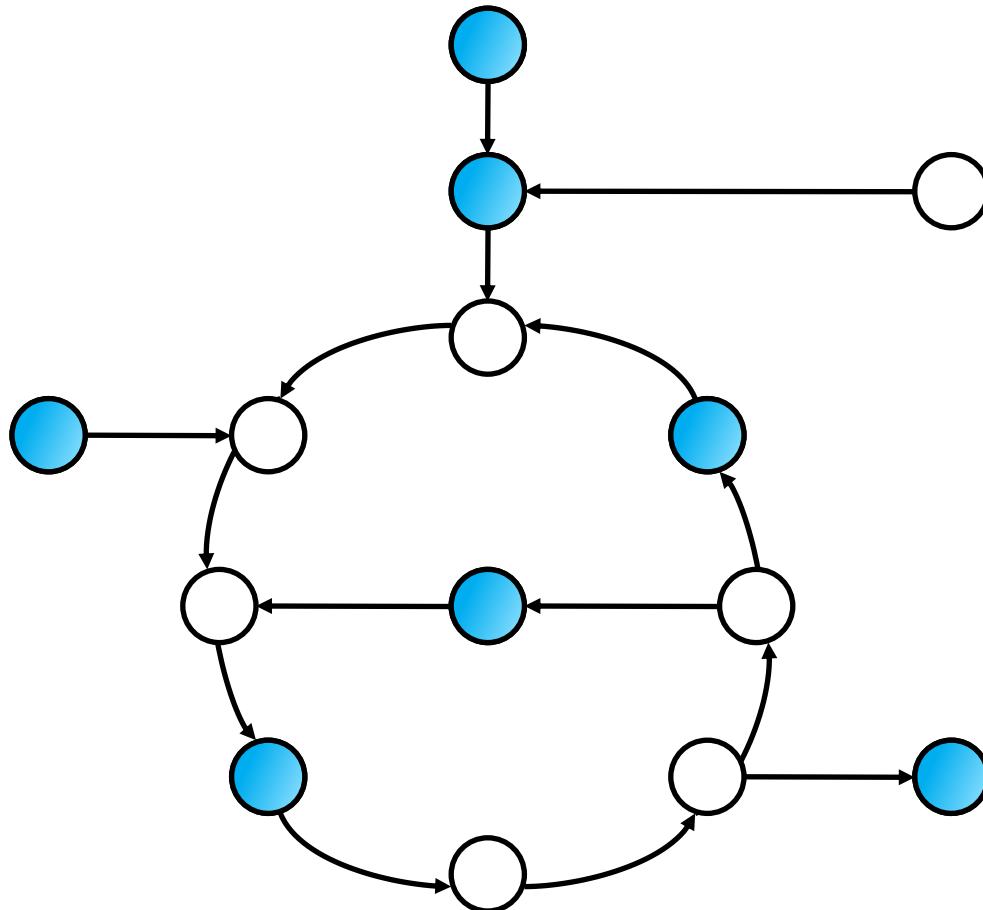
Challenge: heterogeneous networks



Challenge: heterogeneous data



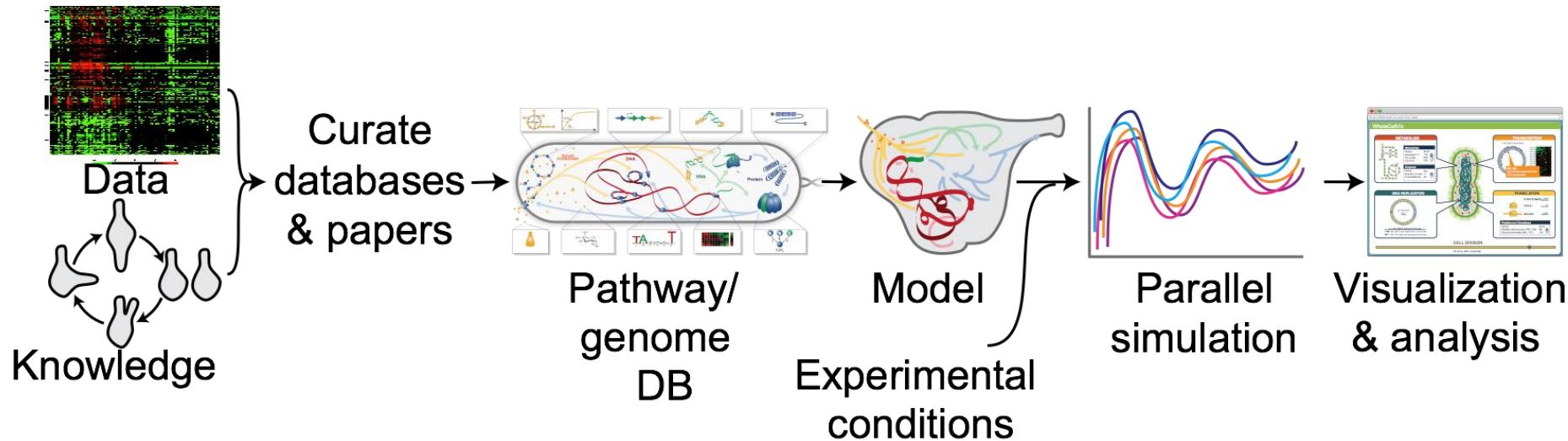
Challenge: sparse data



Outline

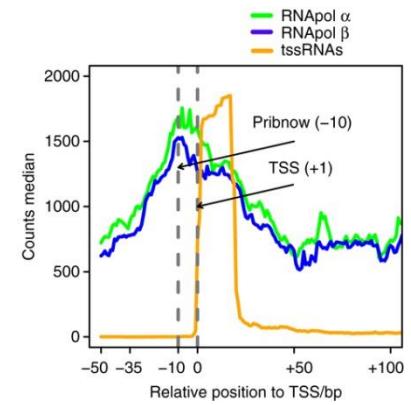
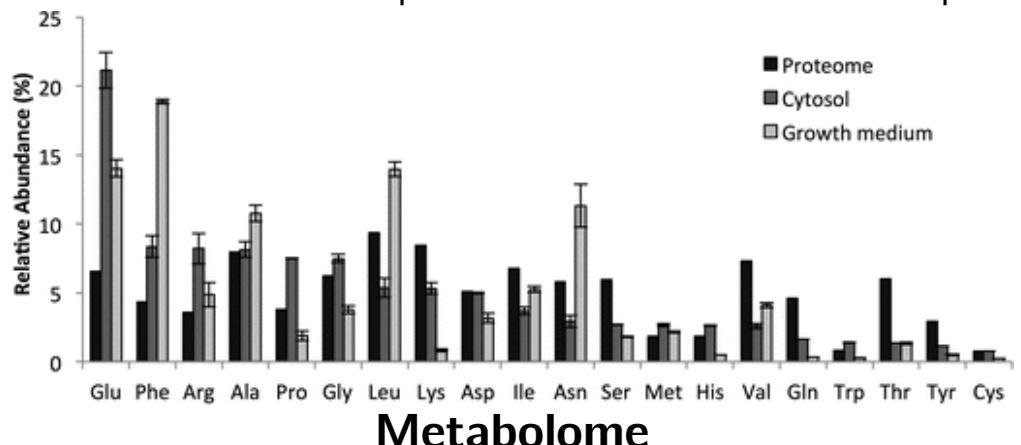
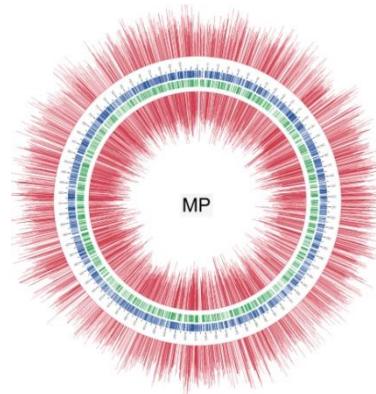
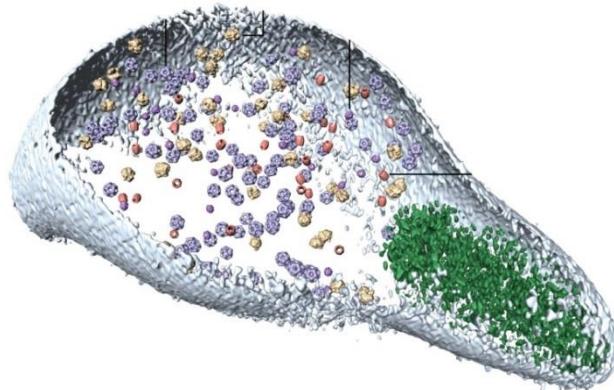
- Motivation
- What is a WC model?
- Challenges
- Model building
- State-of-the-art
- Software tools
- Open challenges

WC modeling process



1. Characterize organism
2. Aggregate data
3. Organize data
4. Design submodels
5. Merge submodels
6. Simulate model
7. Estimate parameters
8. Verify model
9. Validate model
10. Visualize/analyze predictions
11. Applications: Engineering, medicine

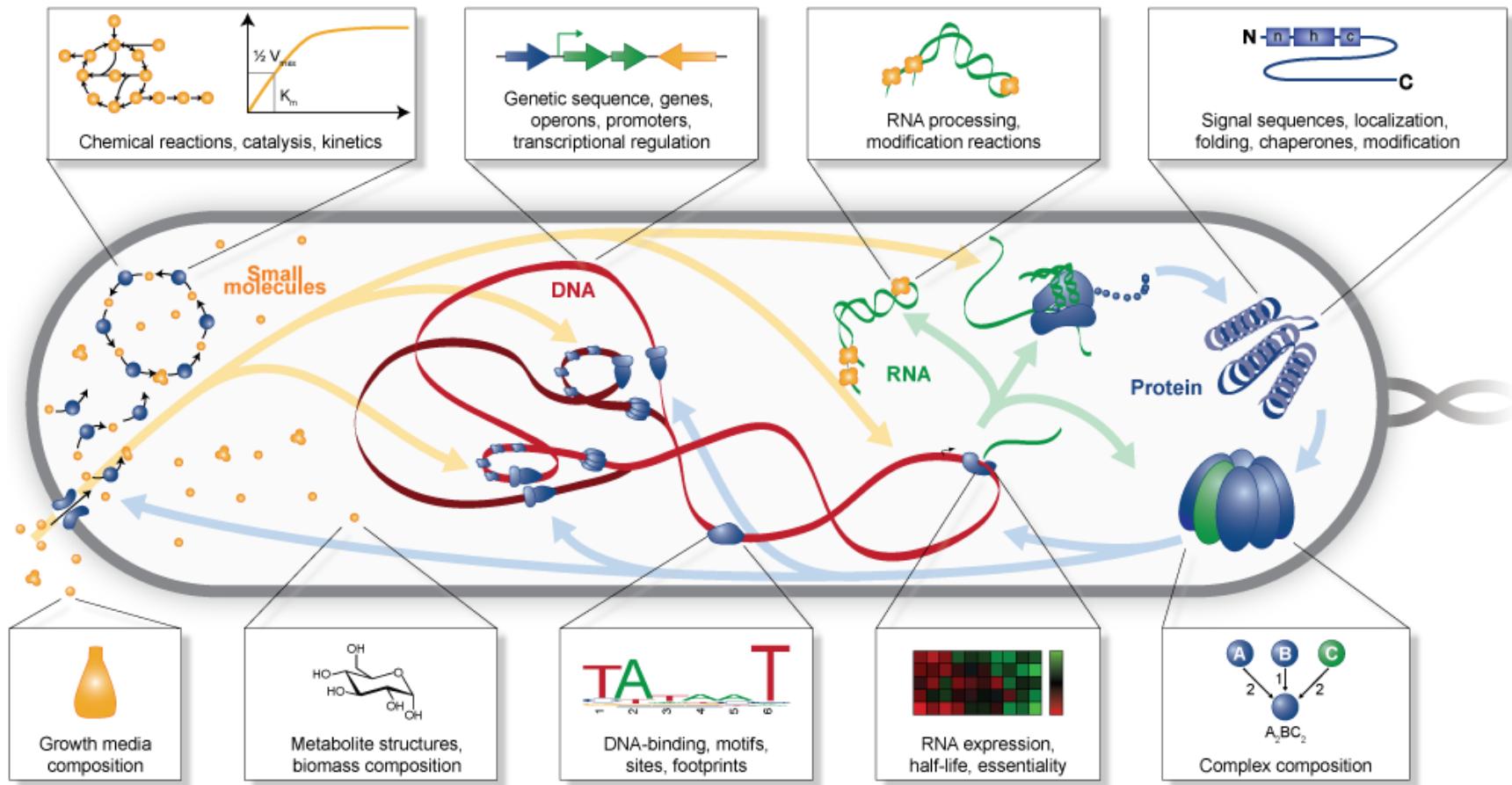
Step 1: Characterize organism



Lectures: Genomics (Serrano, Lluch),
proteomics (Gavin), automation (Gardner)

Fraser et al., 1995; Kühner et al., 2009; Lluch-Senar et al., 2013; Maier et al., 2013; Yus et al. 2012

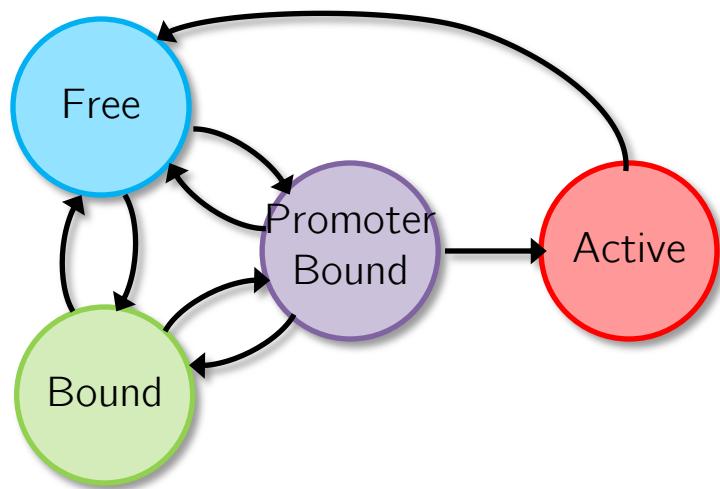
Step 2-3: Aggregate, organize data



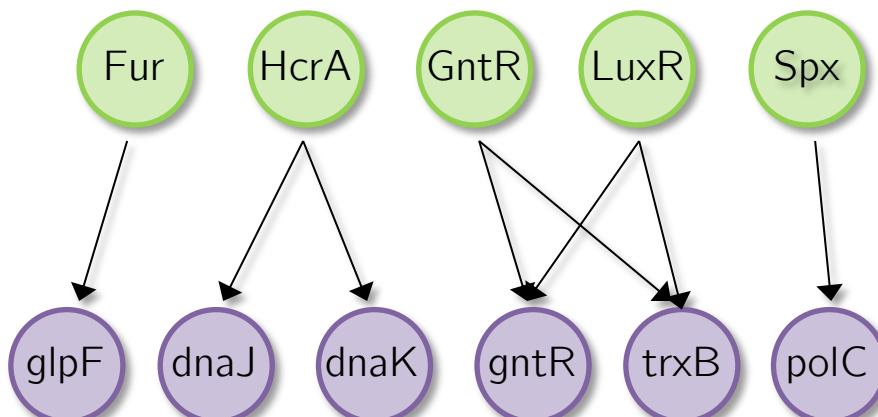
Lectures: PGDBs, Pathway Tools, BioCyc (Peter Karp)

Step 4: Design submodels

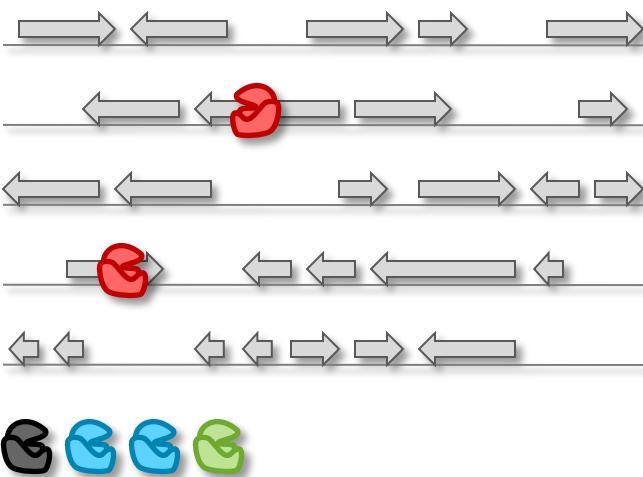
1. Update RNA polymerase states



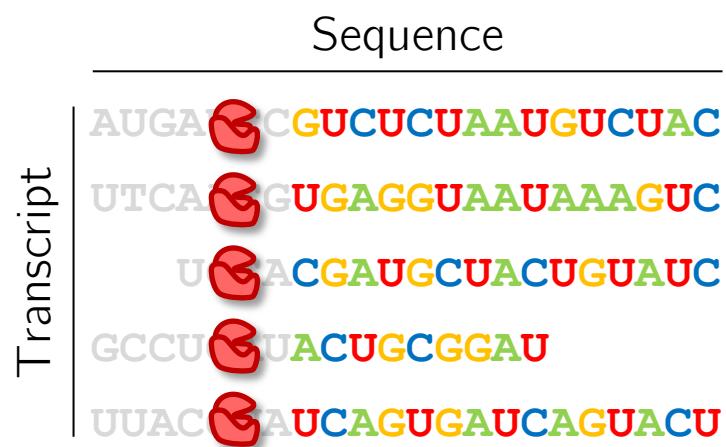
2. Calculate promoter affinities



3. Bind RNA polymerase



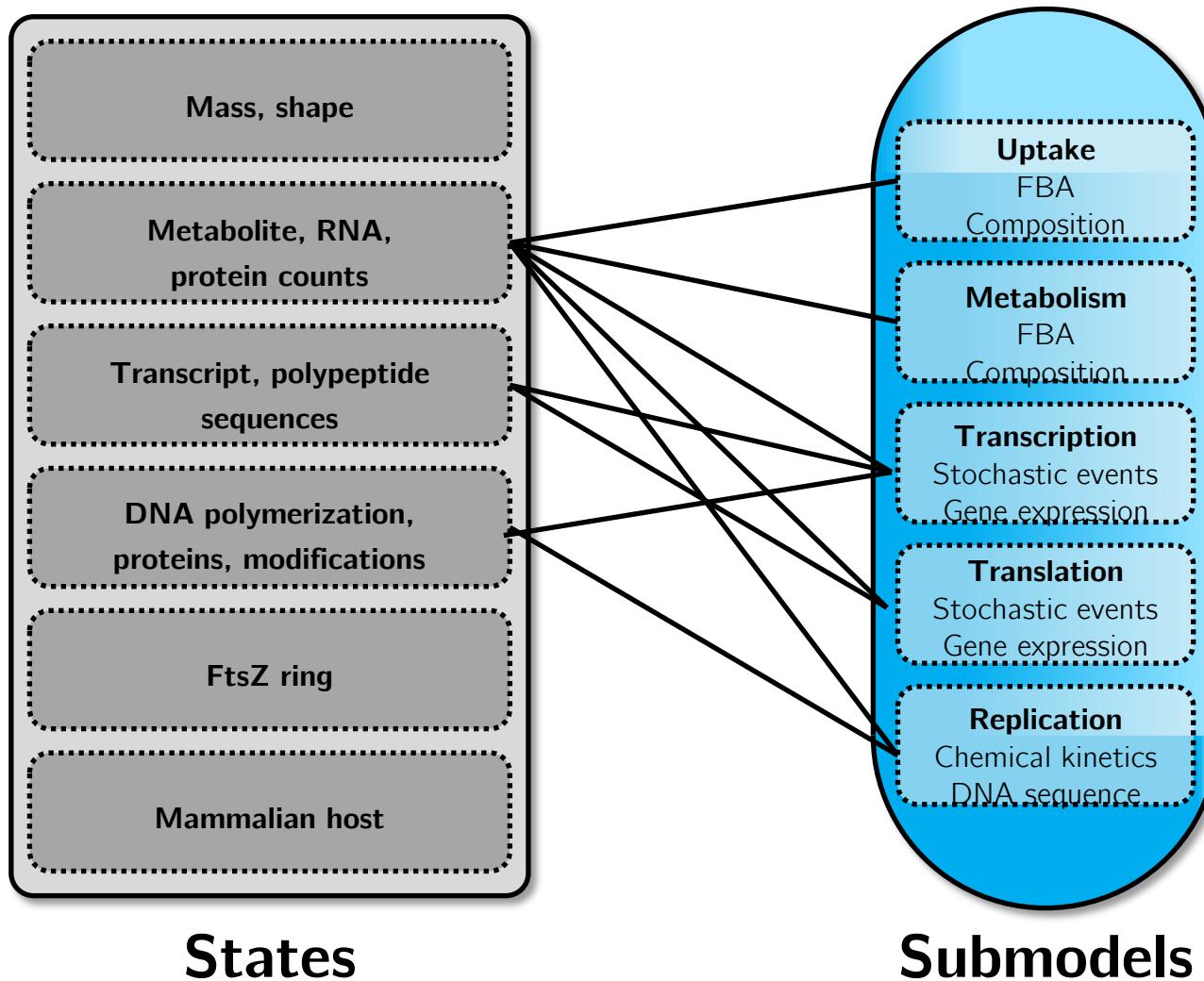
4. Elongate and terminate transcripts



Step 4: Design submodels

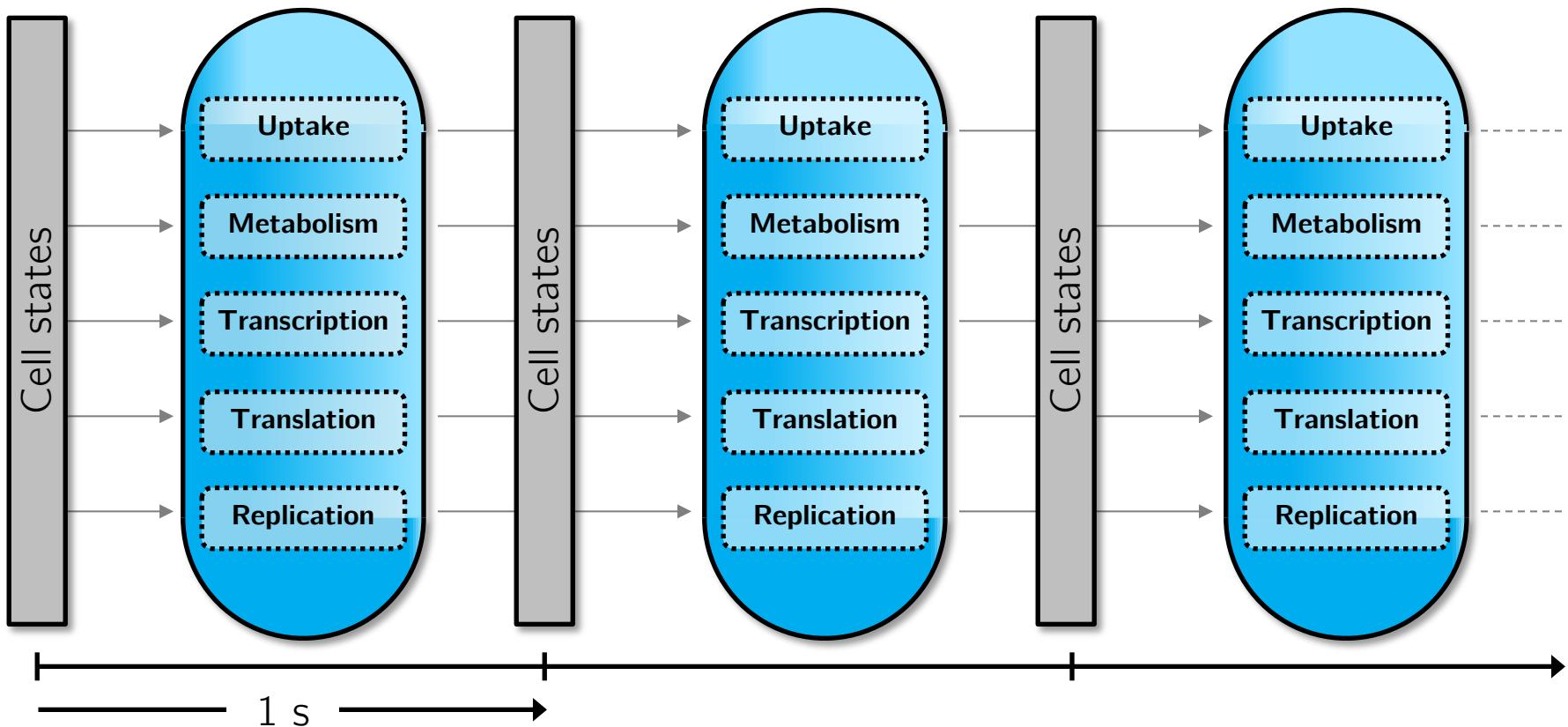
- **Biology:** Metabolism, translation, signaling, cell division, and cell cycle;
- **Modeling formalisms:** ODEs, FBA, stochastic, spatial, and logical modeling
- **Large-scale modeling techniques:** rule-based modeling, standards

Step 5: Map submodels onto common state



Lectures: Rule-based modeling (James Faeder),
model composition (Yin Hoon Chew)

Step 6: Concurrently integrate submodels



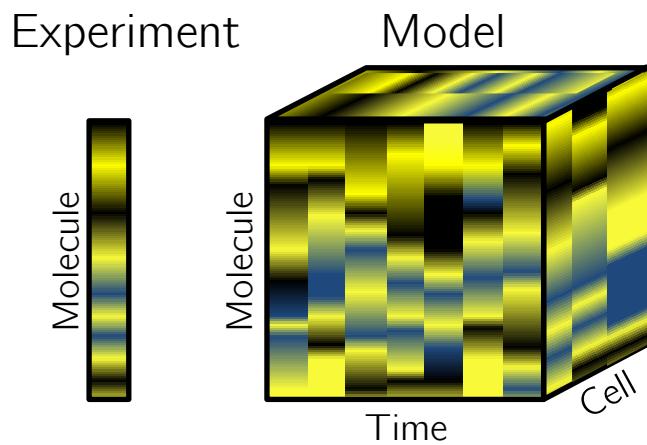
Lectures: Multi-algorithm modeling (Jonathan Karr)

Step 7: Estimate parameters

- Compare the model's predictions to data, y_i
- Numerically minimizing prediction error
- Challenges
 - High-dimensional expensive model
 - Little dynamic, single cell data
 - Noisy, heterogeneous data

Step 7: Estimate parameters

1. Reduce model



2. Identify reduced model parameters using traditional methods

3. Manually tune parameters using full model

Steps 8-9: Verify and validate model

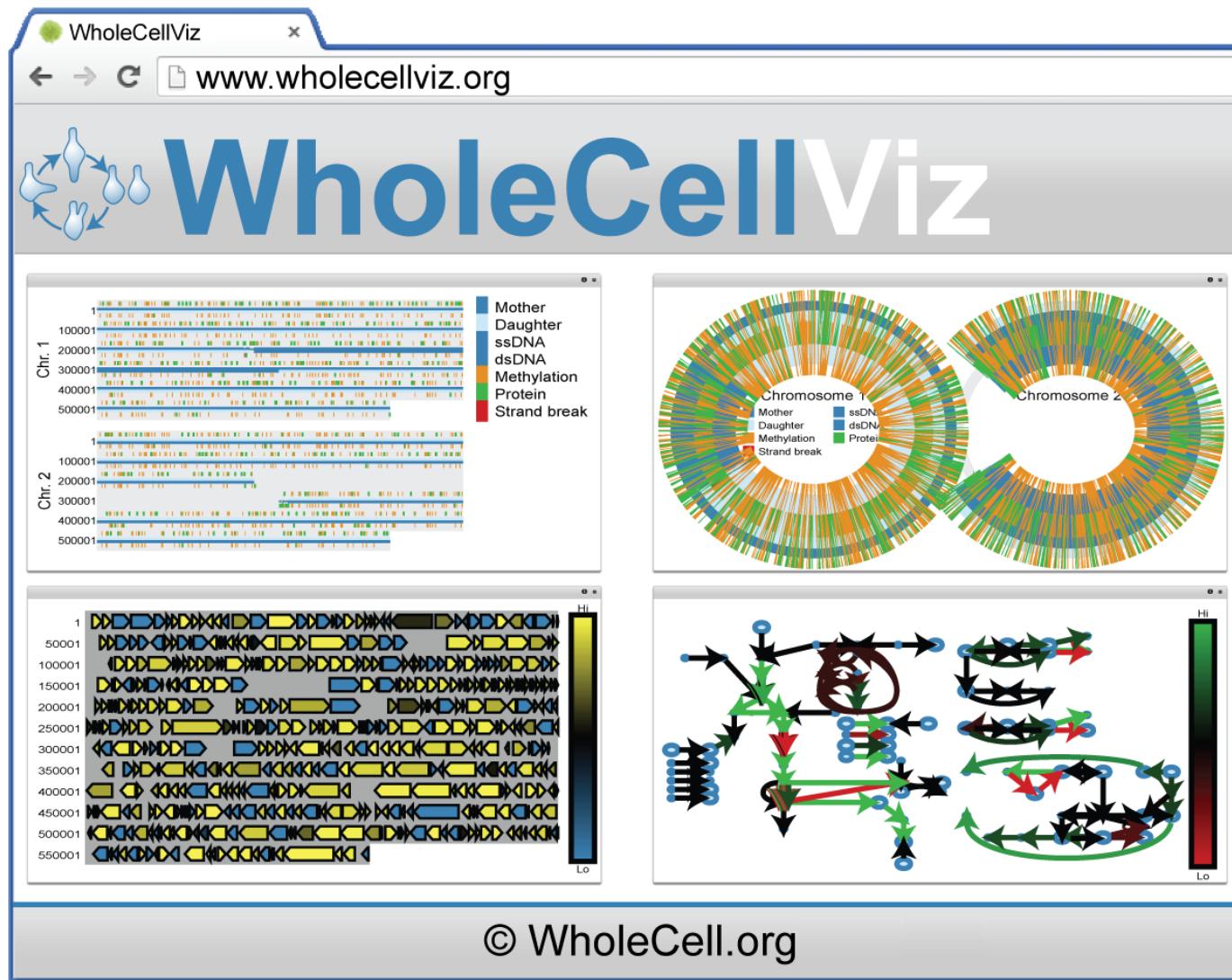
Verify: known biology

- Matches training data
- Matches published data
- Matches theory
- No obvious errors

Validate: true biology

- Matches new data

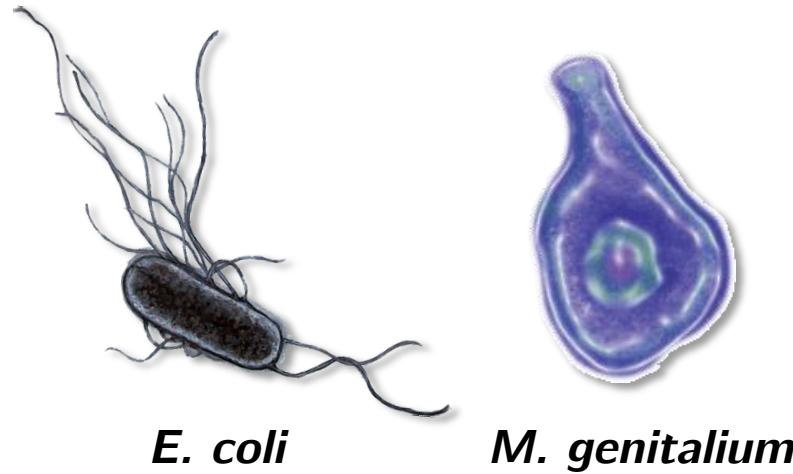
Steps 10-11: Visualize, analyze, discover & engineer



Outline

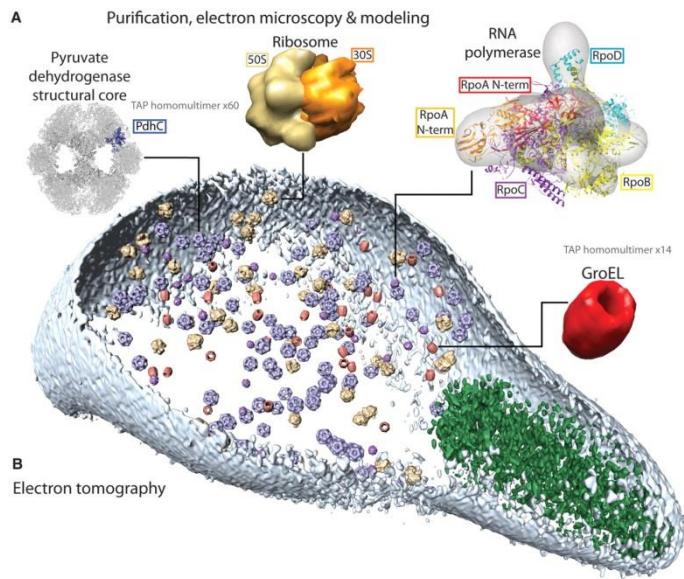
- Motivation
- What is a WC model?
- Challenges
- Model building
- State-of-the-art
- Software tools
- Open challenges

Example: *Mycoplasma genitalium* WC model



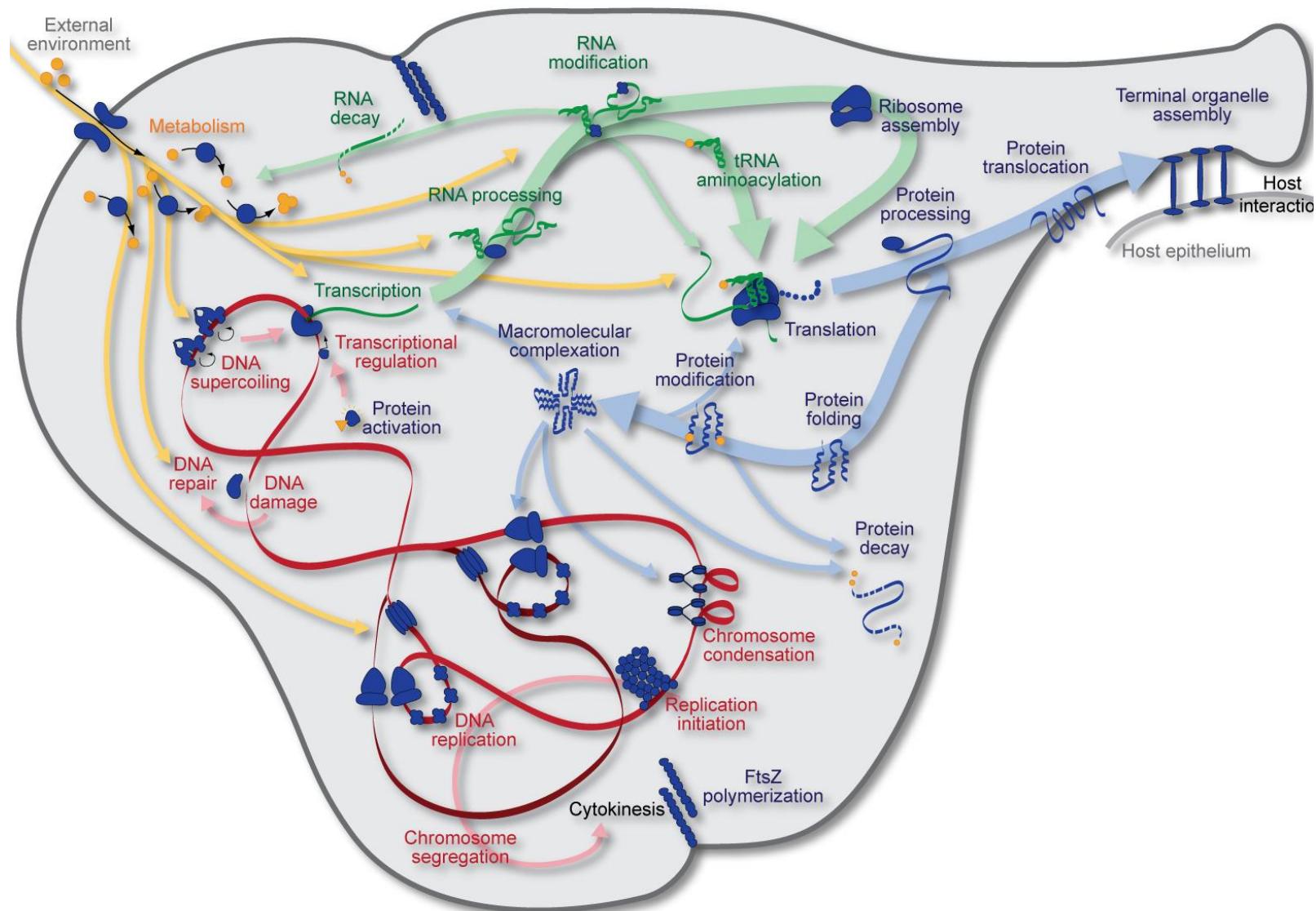
Genome	4700 kb	580 kb
Genes	4461	525
Size	2 μm × 0.5 μm	0.2-0.3 μm

Mycoplasmas are well-characterized

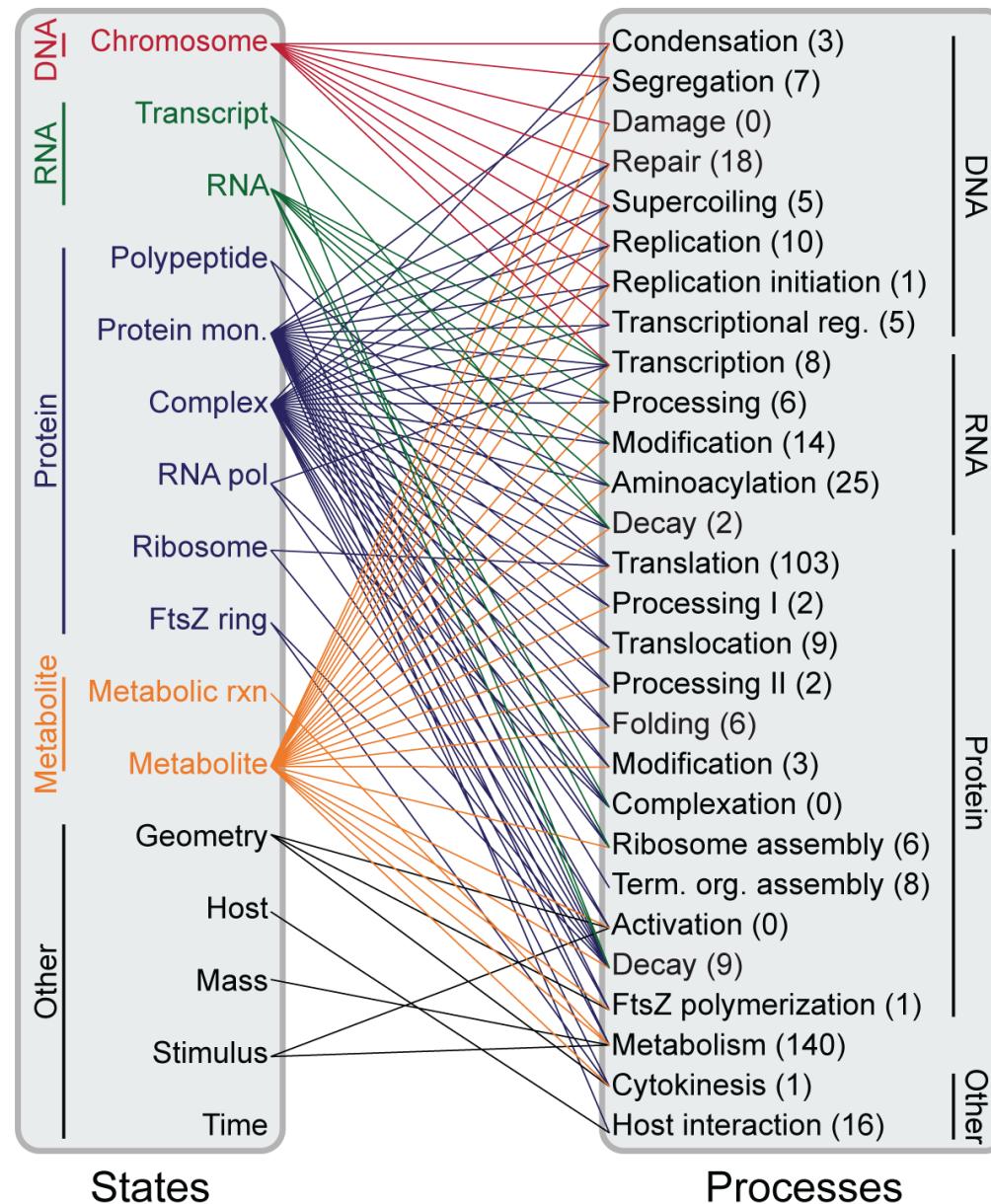


Genomic-scale data
Kühner et. al, 2009

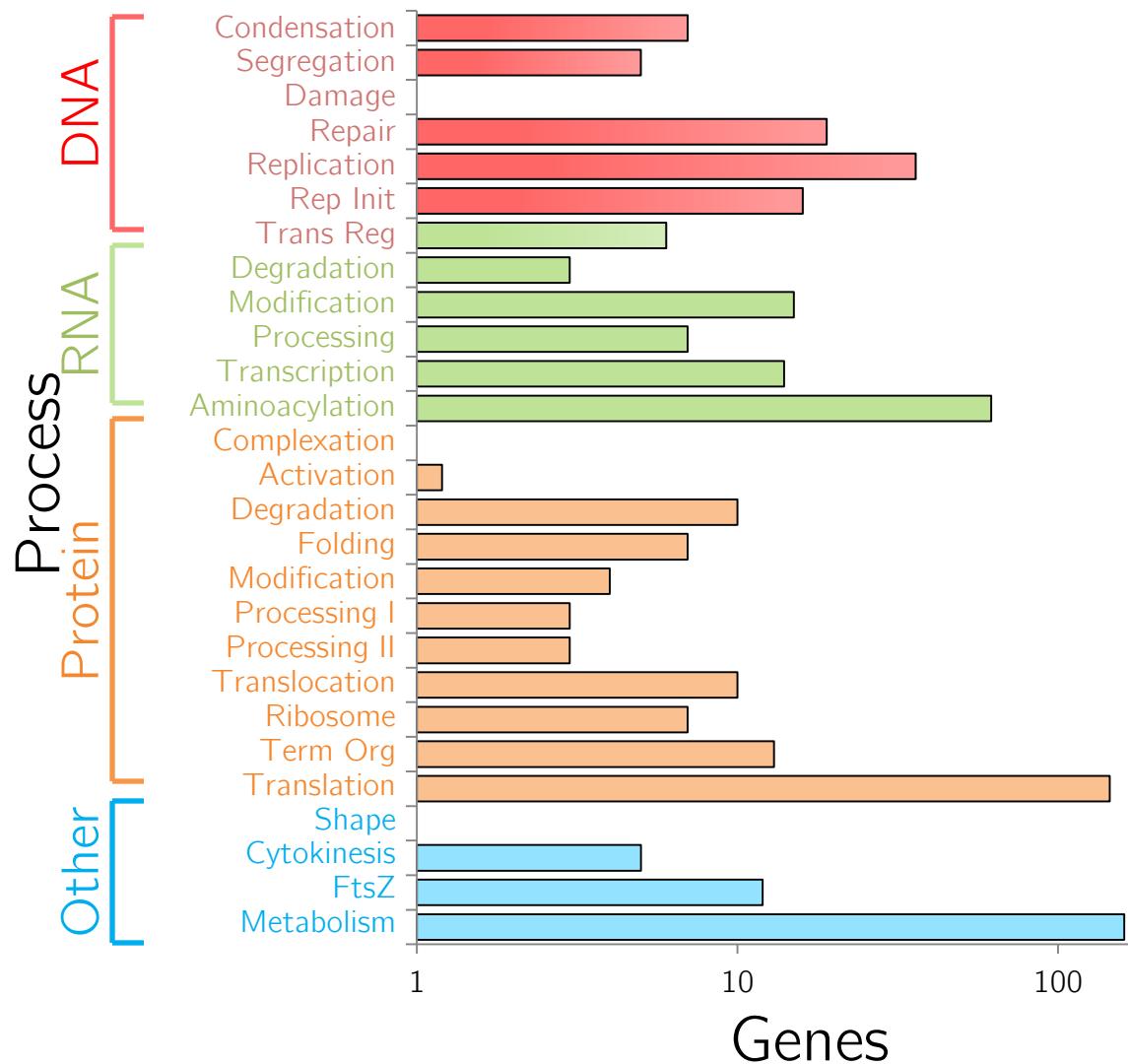
Model includes 28 pathway submodels



Submodels are mapped onto 16 common states



Model represents 76% of genes



Model statistics

Model

- 525 genes
- 700 proteins
- 600 metabolites
- 1,800 reactions
- 1,900 parameters
- 930 citations

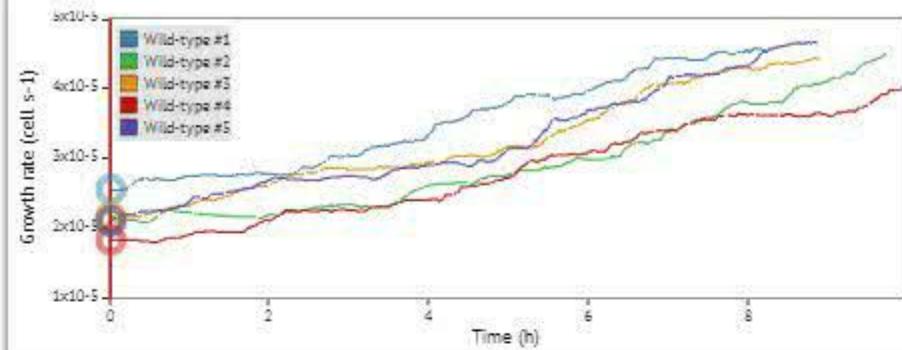
Code

- 163,590 lines (34,380 model)
- 2,172 revisions
- 3,151 published simulations
- 1.5 TB compressed
- 24+ cpu yr to fit model

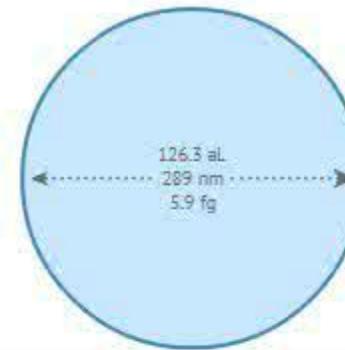
Languages

- MATLAB
- Python, JavaScript, HTML, MySQL
- Perl, Shell

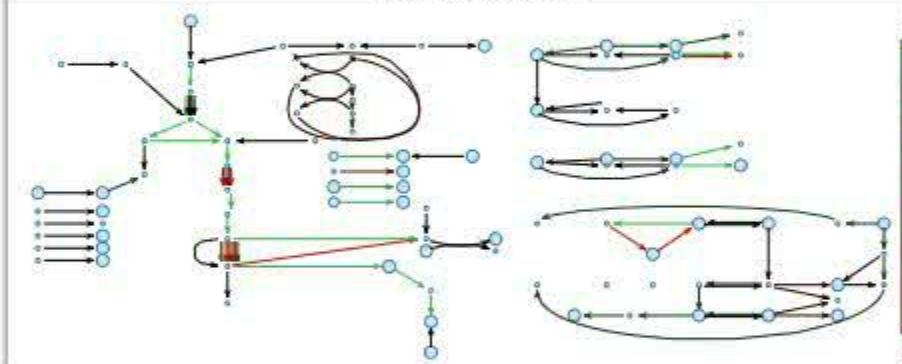
Growth rate
Simulation: multiple selected



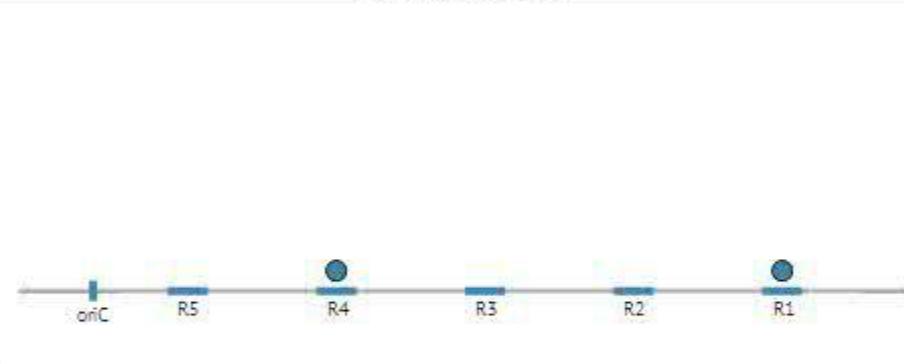
Cell shape
Simulation: Wild-type #1



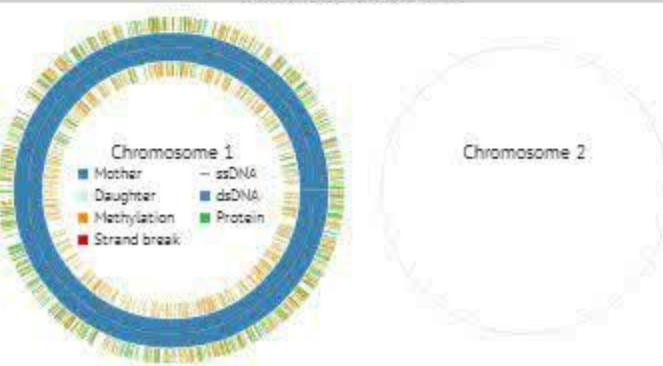
Metabolism
Simulation: Wild-type #1



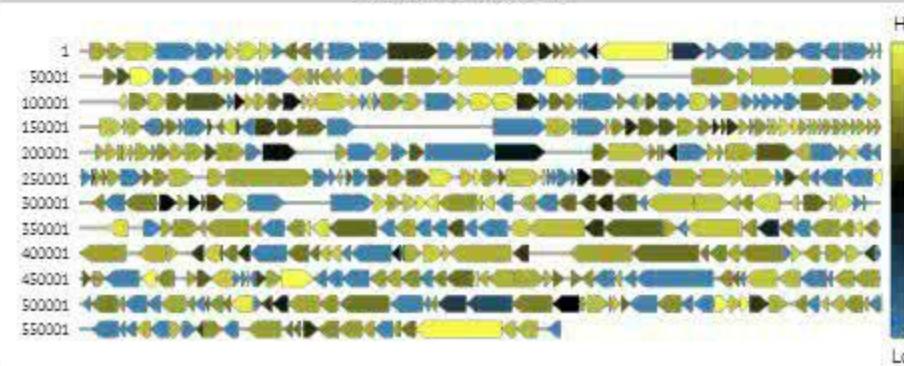
Replication initiation – oriC DnaA Boxes
Simulation: Wild-type #1



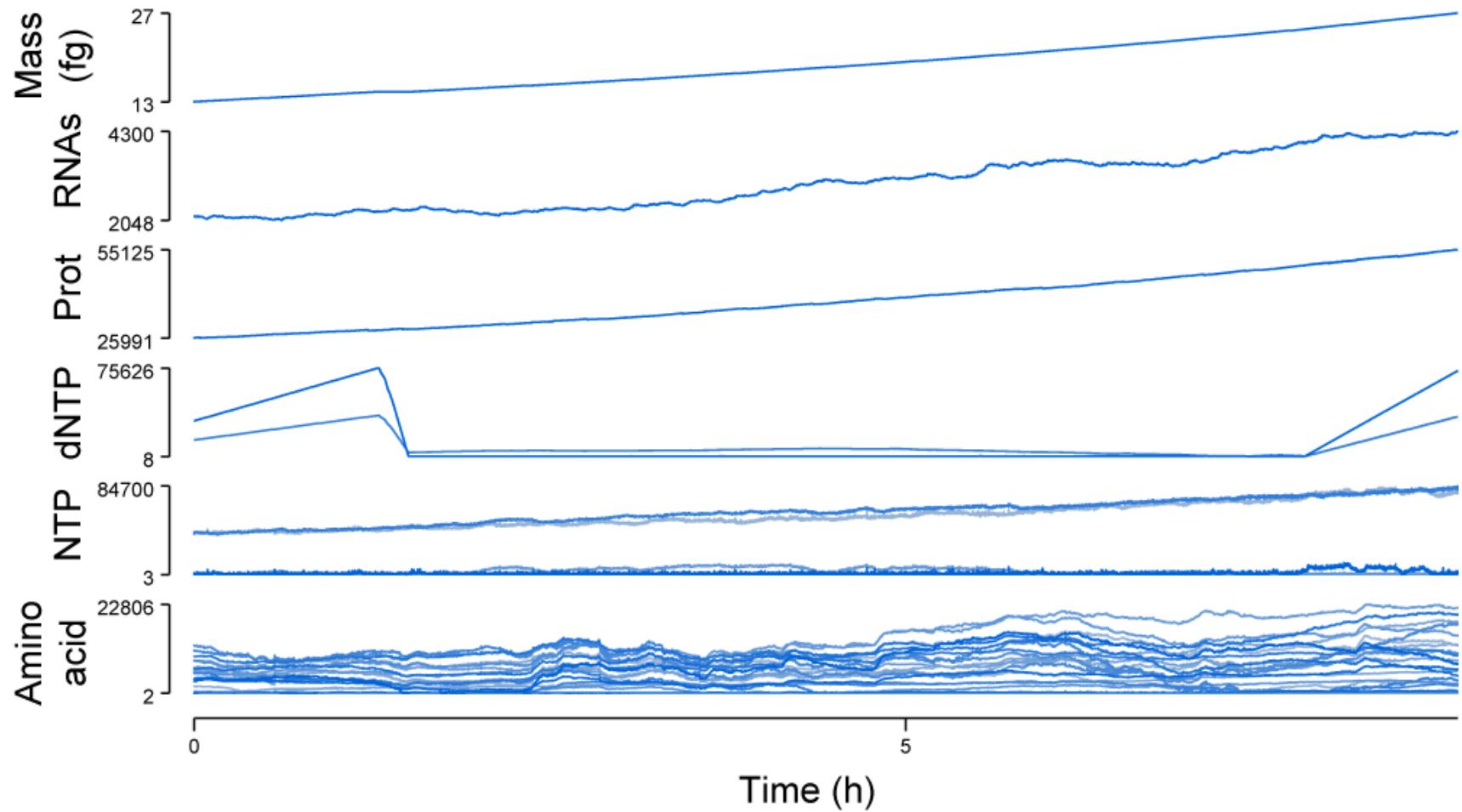
DNA replication, protein occupancy, methylation, & damage
Simulation: Wild-type #1



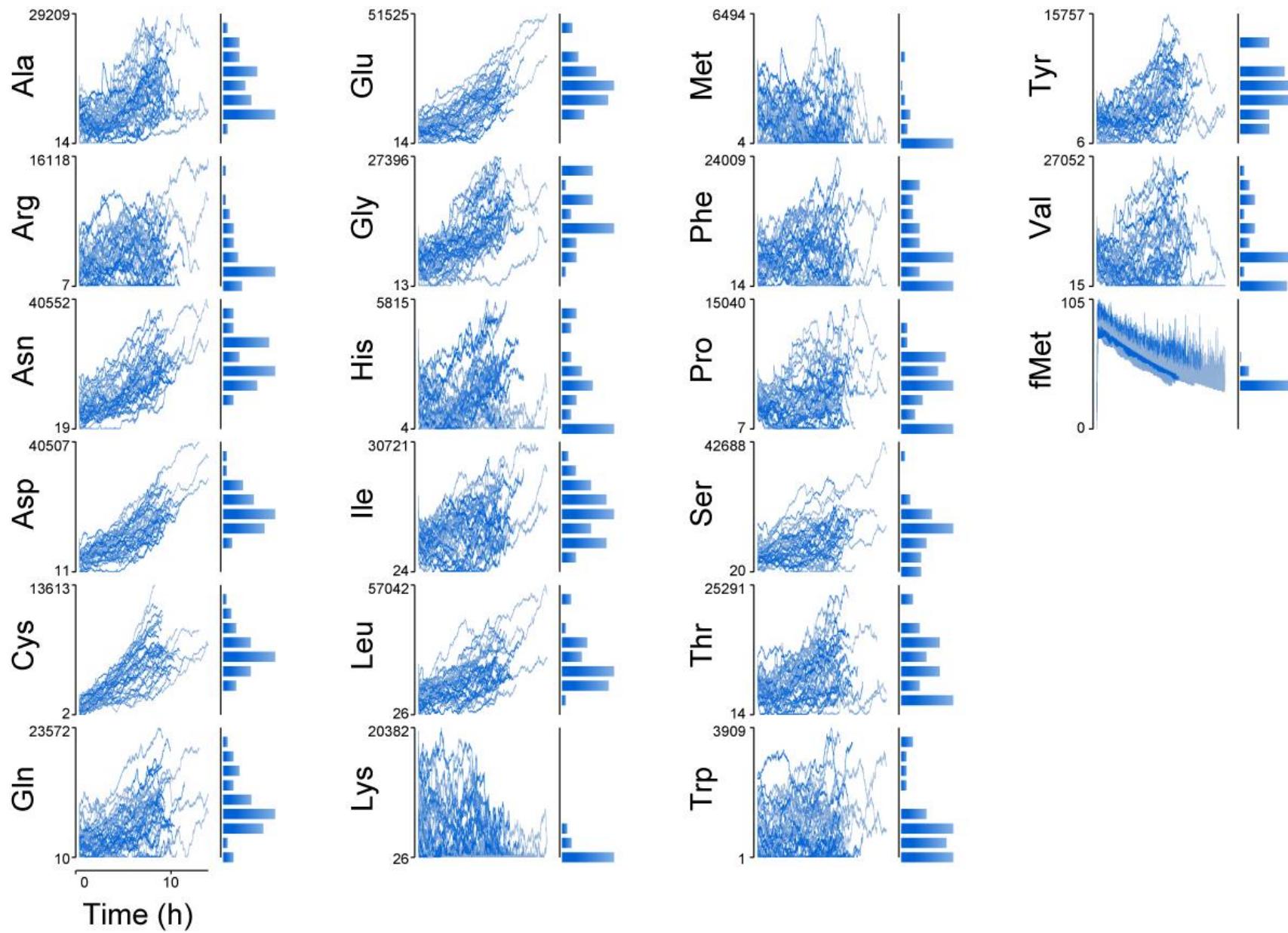
Mature protein monomer expression
Simulation: Wild-type #1



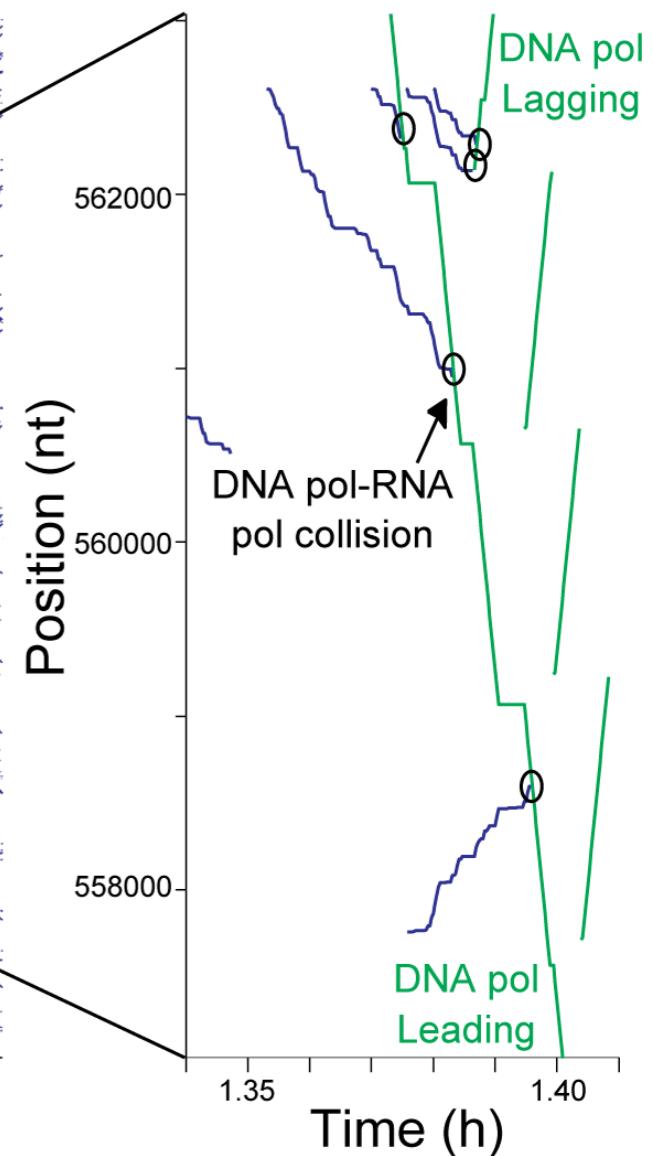
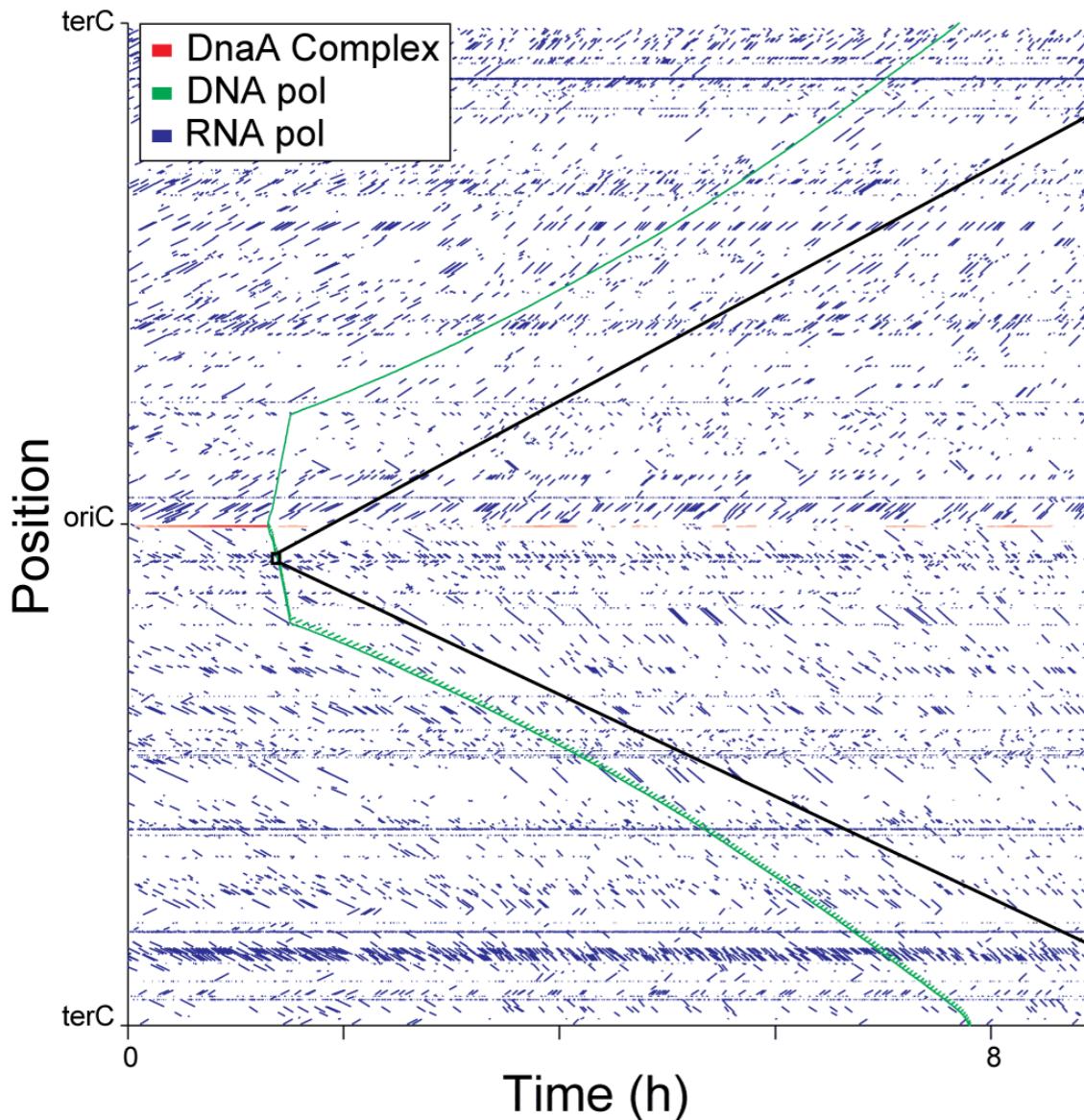
Predicted cell composition



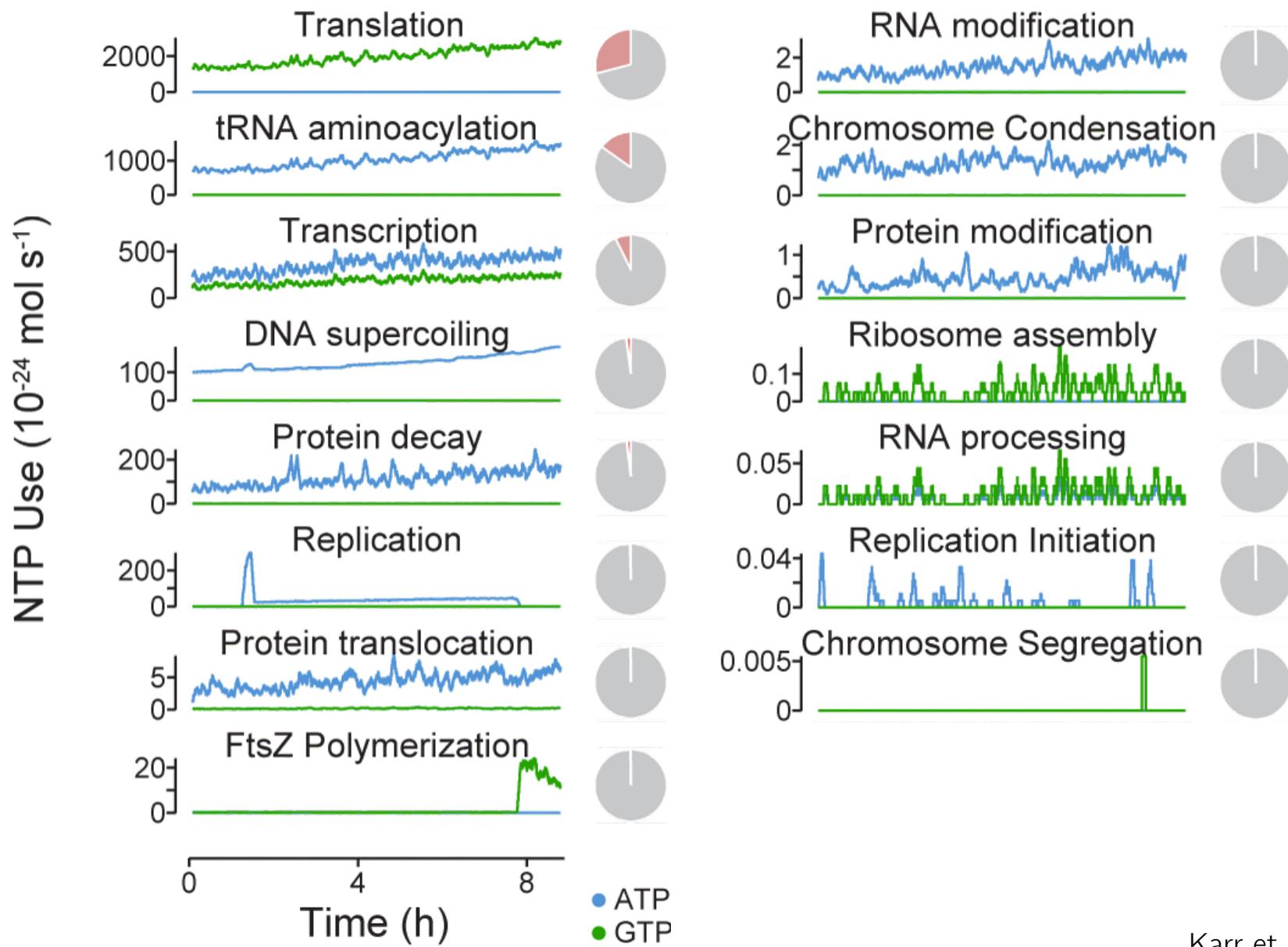
Predicted metabolite concentrations



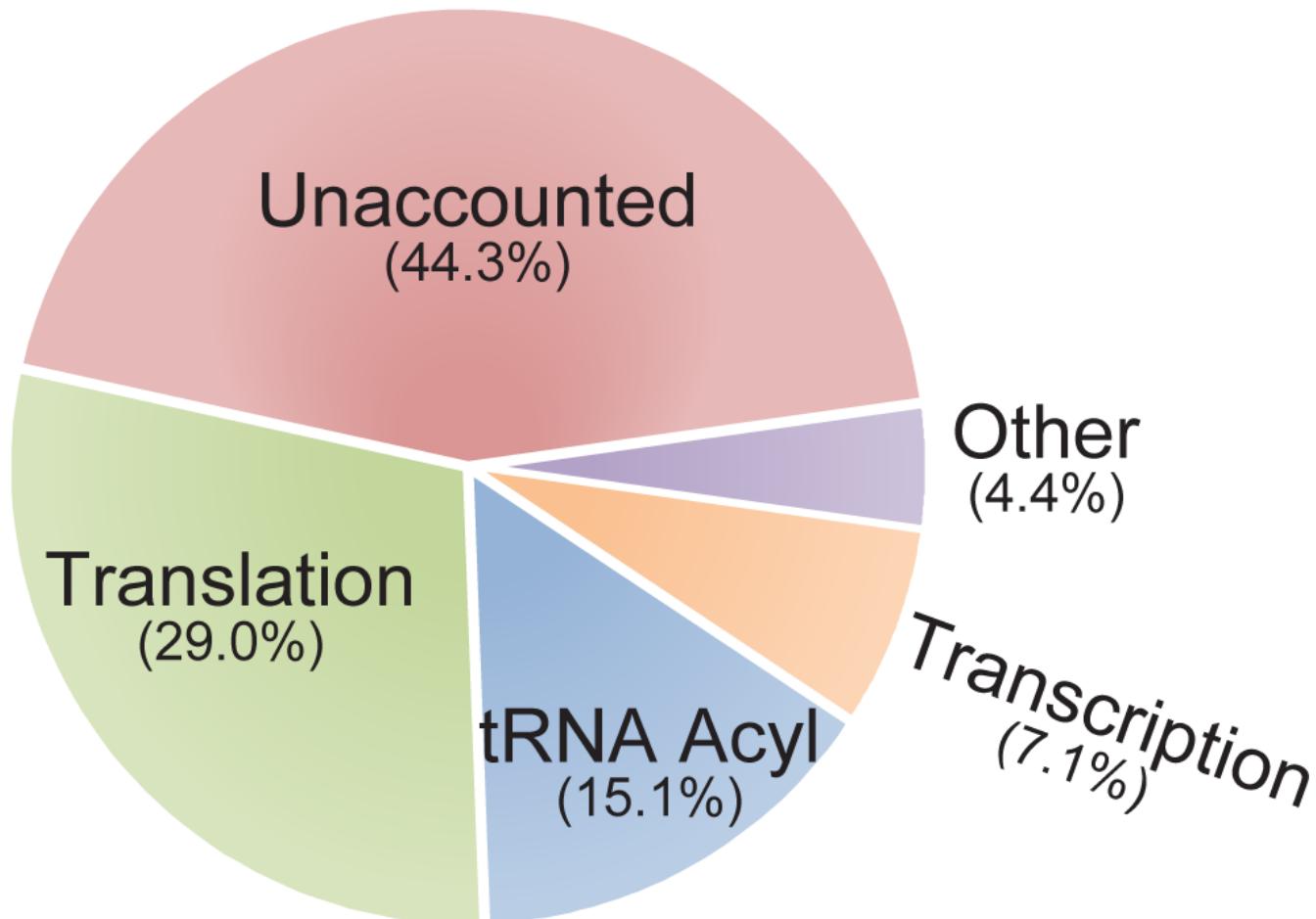
Predicted DNA binding protein collisions



Predicted energy consumption



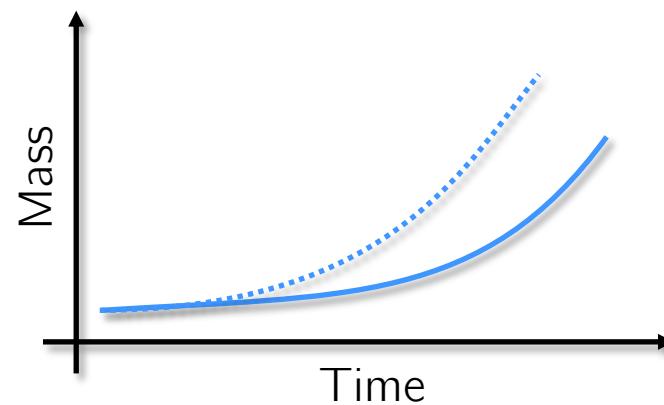
Predicted energy consumption



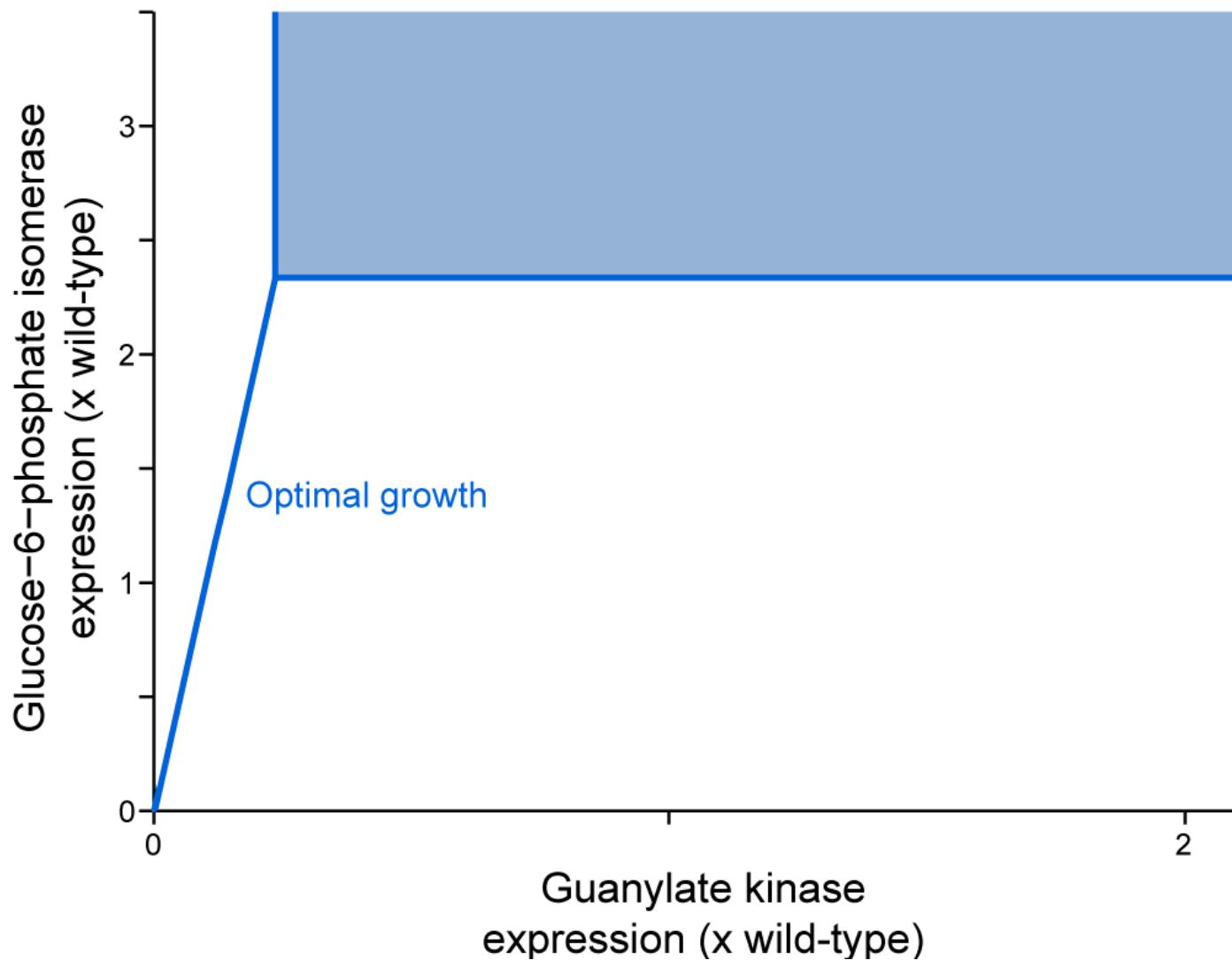
60 m mol ATP / gDCW
80 a mol ATP / cell

Example: growth optimization

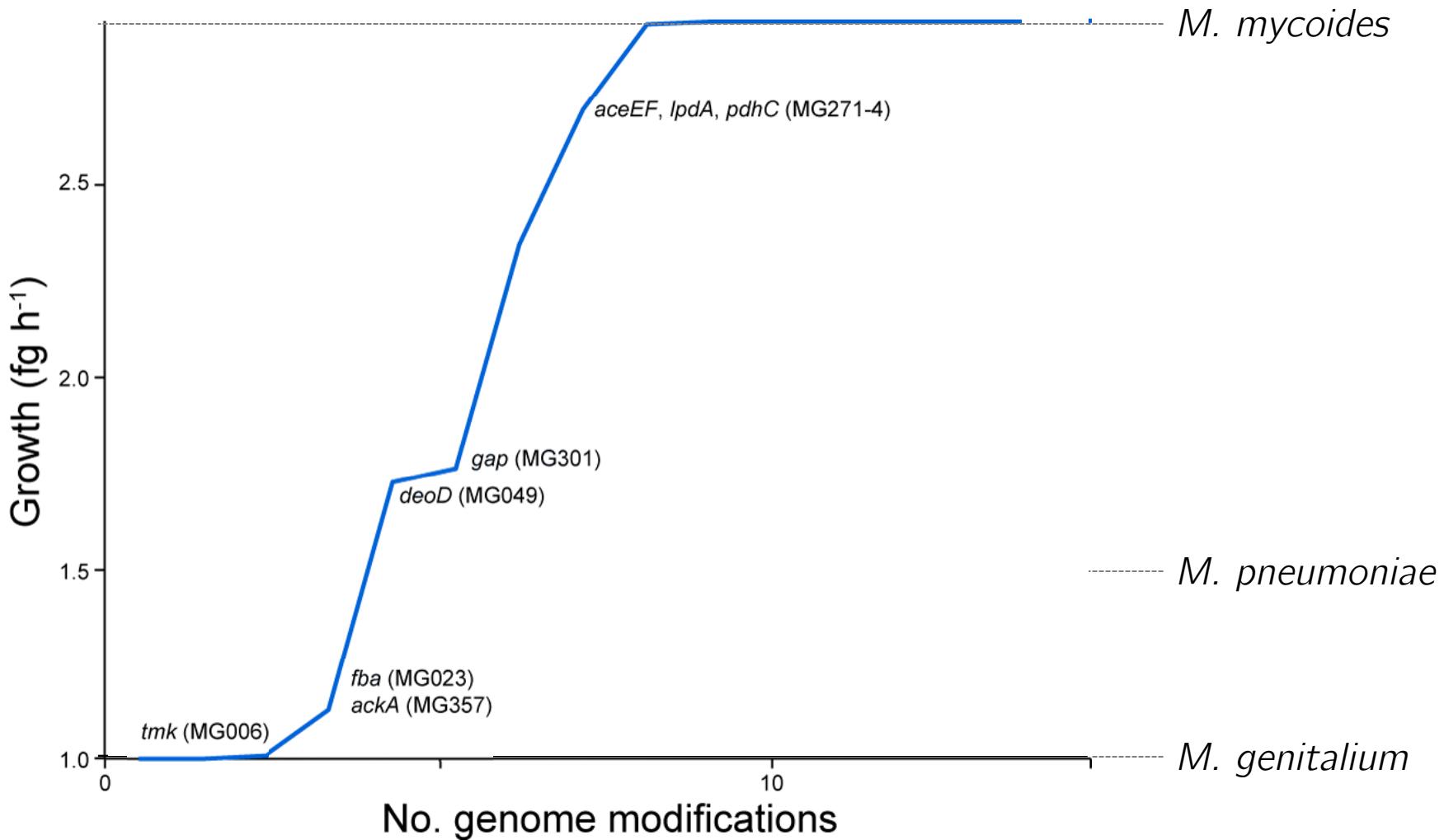
What genomic modifications maximize growth?



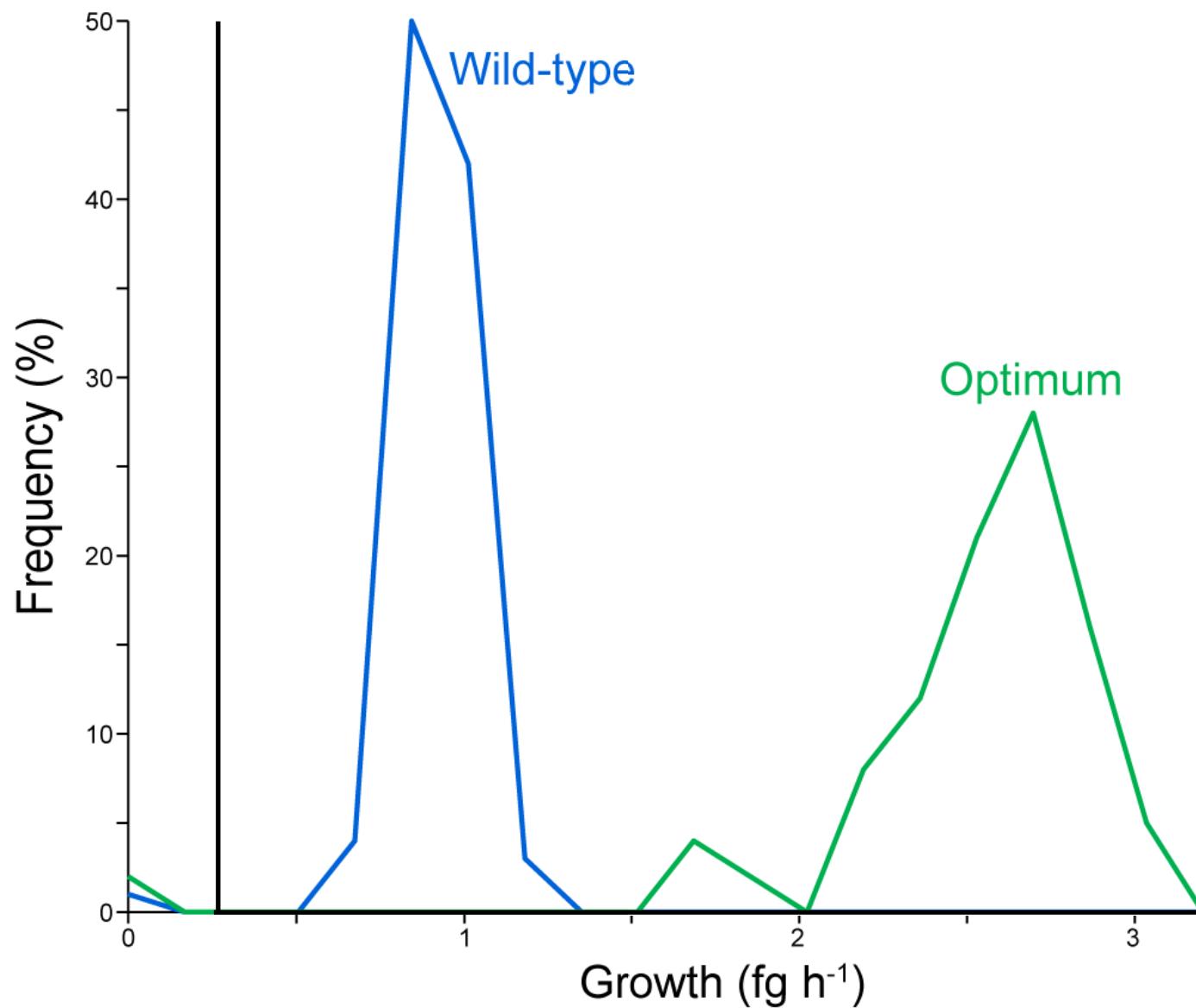
Optimal gene expression



Optimal gene expression



Optimal gene expression retains robustness



Drug repositioning

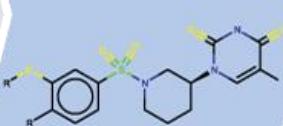


M. genitalium
525 genes

Identify fragile genes
Systems modeling



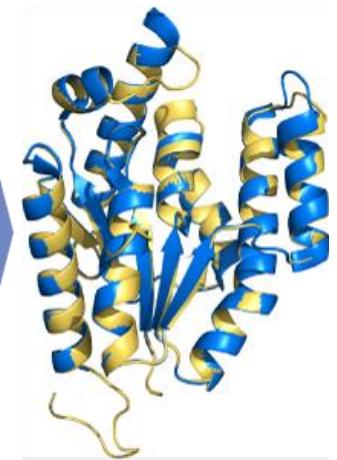
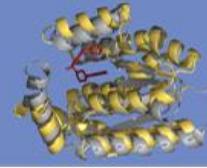
Identify inhibitors
Bioinformatics



Assess active & binding sites
homology
Bioinformatics

```
--TMEPLTEAYLFAAARTEHIS  
--DMDIRTEAMLFAAASRREHLV  
DEVITDKAEVLMFYAARVQLVE  
--DVEDHSVHLLFSANRWEQVP
```

Assess affinity
Structural modeling



Tmk
1 gene

Outline

- Motivation
- What is a WC model?
- Challenges
- Model building
- State-of-the-art
- Software tools
- Open challenges

WC modeling software

Pathway/genome databases

- Pathway Tools: biocyc.org
- WC-KB: wholecellkb.org

Modeling languages

- SBML: sbml.org
- CellML: cellml.org

Integrative modeling

- E-Cell: e-cell.org
- Whole-cell: wholecell.org

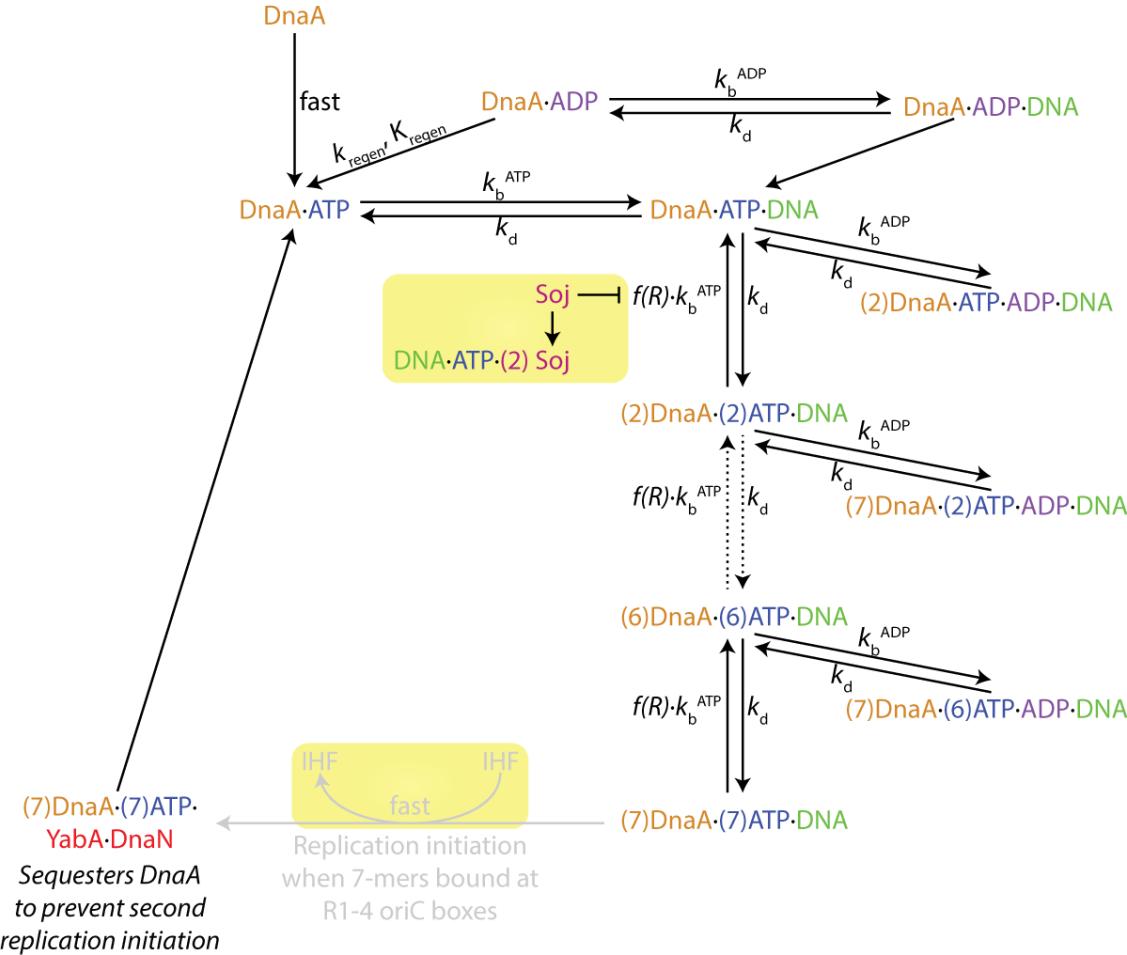
Simulation results visualization

- WC-SimDB: wholecellsimdb.org
- WC-Viz: wholecellviz.org

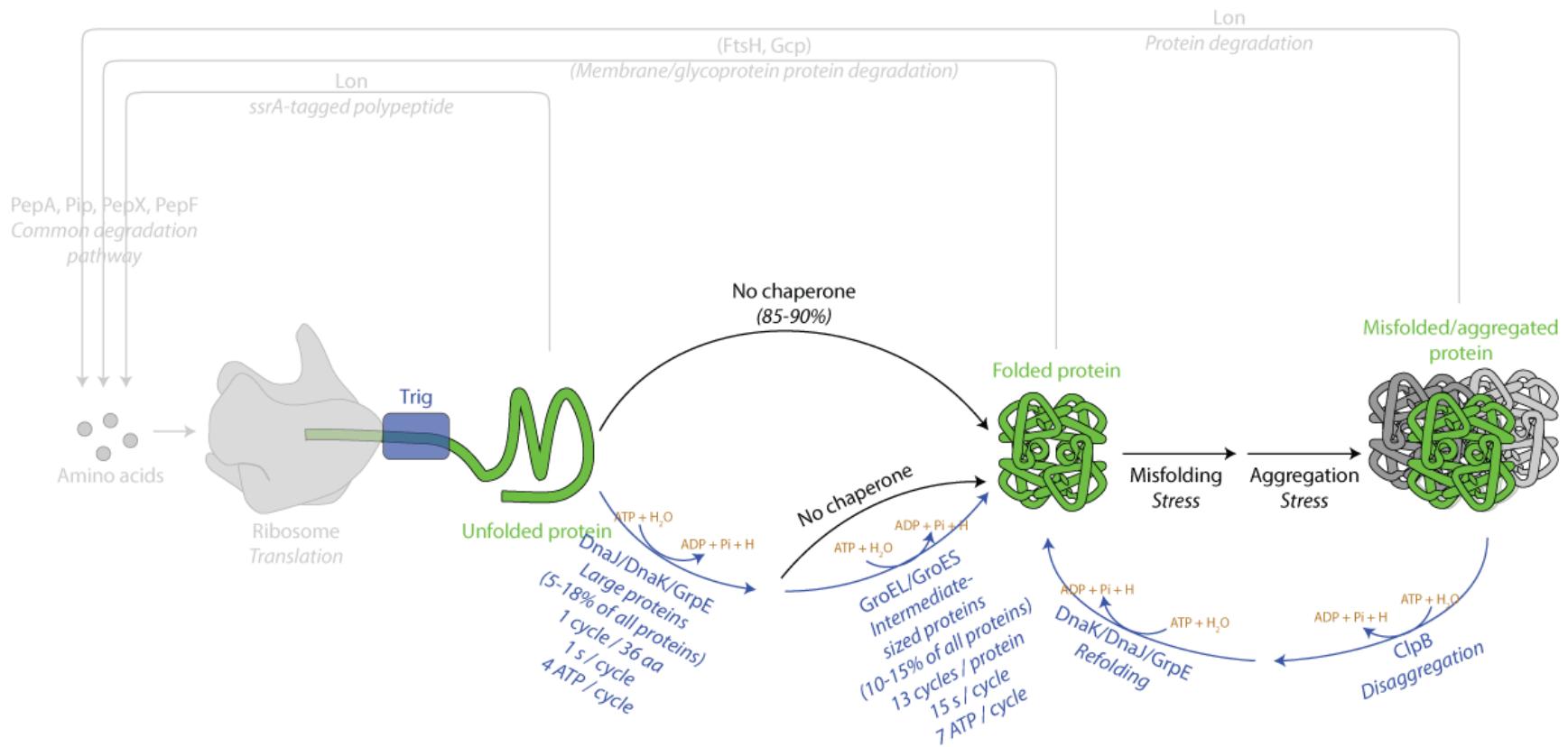
Outline

- Motivation
- What is a WC model?
- Challenges
- Model building
- State-of-the-art
- Software tools
- Open challenges

Limitations of existing methods



Limitations of existing methods



Limitations of existing methods

Limited scope and accuracy

- Don't represent several cell functions
- Don't predict several phenotypes

Methods are not rigorous

- Multi-algorithm simulation
- Parameter estimation
- Verification
- Model reduction

Time-consuming to construct

- Curate data
- Design model
- Verify model

Hard to understand, reuse, reproduce

Open challenges

Experimental characterization and data aggregation

- Comprehensive metabolomic, proteomic, kinetic measurements
- Data annotation, deposition

Systemizing and accelerating WC modeling

- Data curation tools
- WC modeling language
- Formal verification
- High-performance simulator

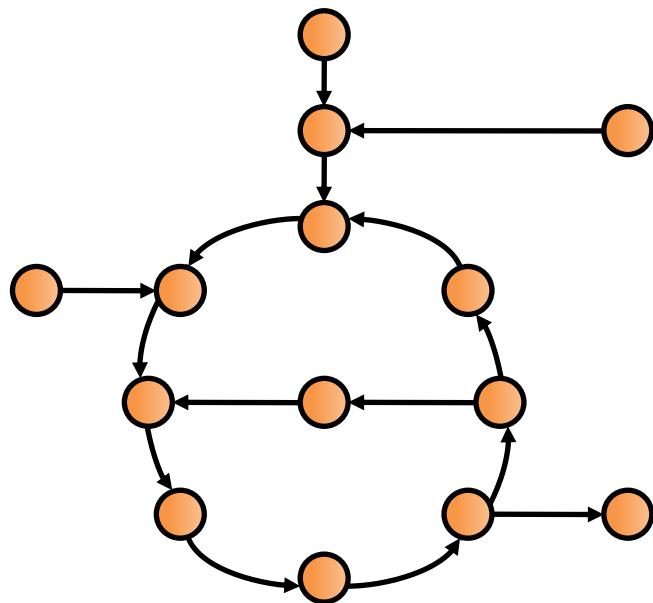
Modeling more complex physiology

- Regulatory pathways
- Larger volumes, genomes
- Spatial organization
- Multicellularity
- Longer time scales: aging, evolution

Using models to direct engineering and medicine

- Algorithms to optimize phenotypes
- Large-scale genome engineering methods
- Medical decision support

Summary



Integrative model

Broadly **predicts** cell physiology

Integrates heterogeneous data and models

Guides **bioengineering** and **medicine**

Opportunities to **develop improved methods**

Acknowledgements



Anne Marie Barrette



Yin Hoon Chew



Arthur Goldberg



Graeme Gossel



Pablo Meyer
IBM/Sinai

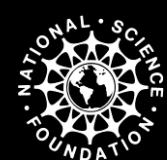


Roger Rodriguez
UNAM



Center for Regulatory Genomics

Luis Serrano
Maria Lluch-Senar
Veronica Llorens



JAMES S.
MCDONNELL
FOUNDATION

Mount
Sinai