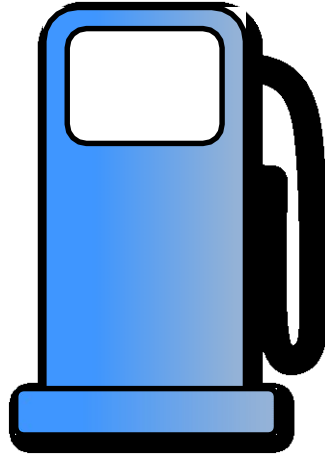


Ministério da Saúde

**FIOCRUZ**  
**Fundação Oswaldo Cruz**

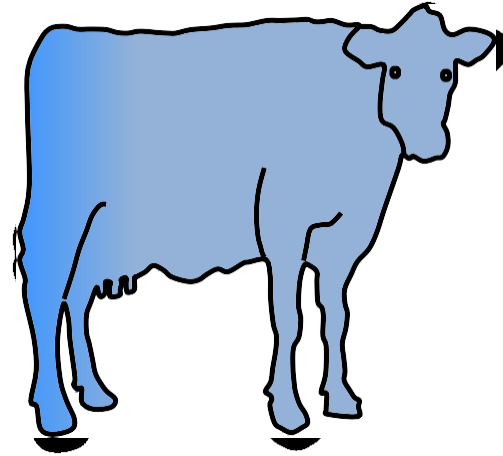
# Integração de Modelos Biológicos

Doutorando Fernando Medeiros



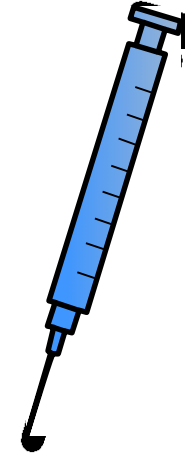
**Produtos  
Químicos e  
Combustíveis**

Otimizar o  
rendimento  
Minimizar o  
custo



**Alimento**

Otimizar o  
rendimento  
Resistir à seca  
Evitar a  
infecção

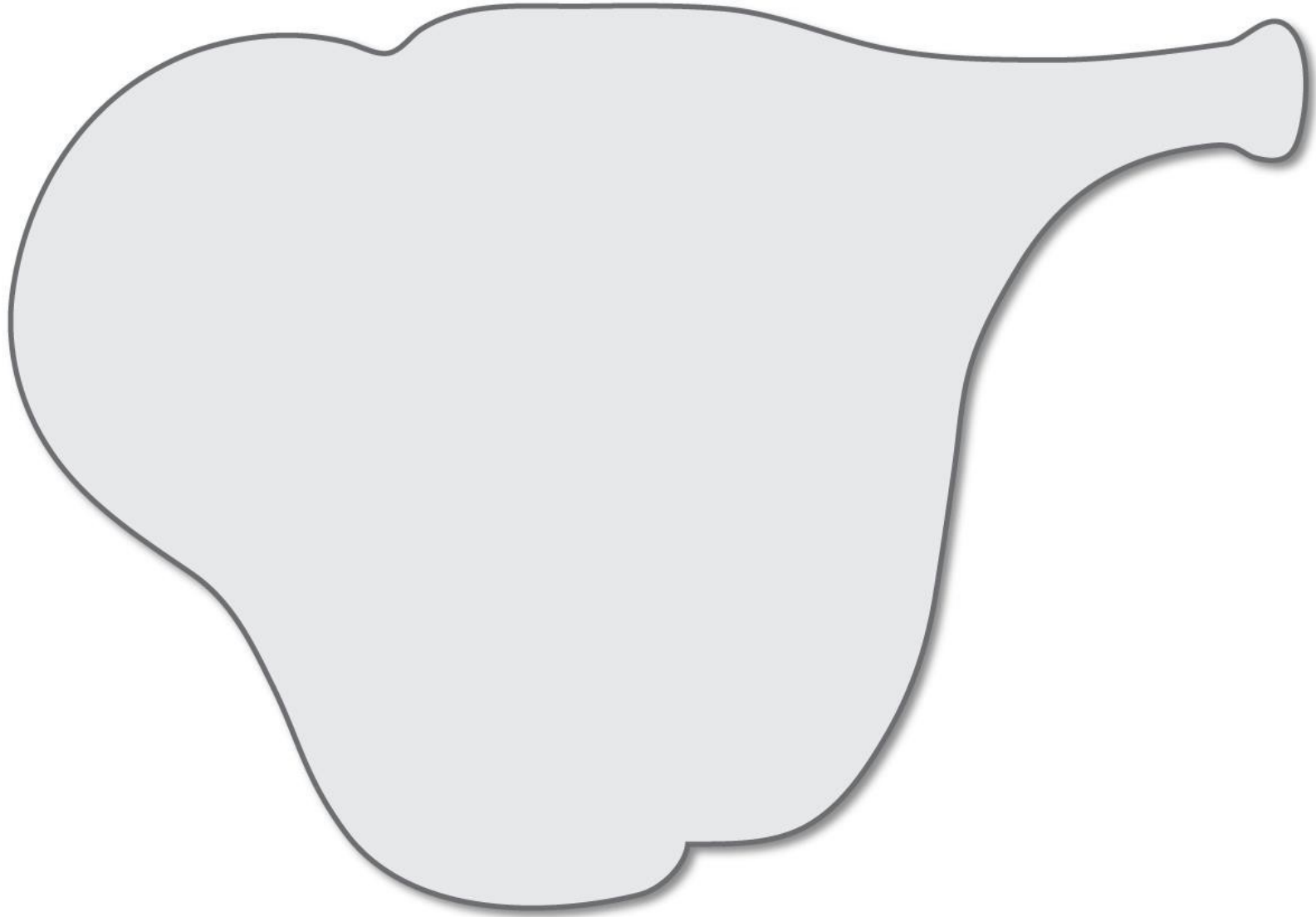


**Medicina**

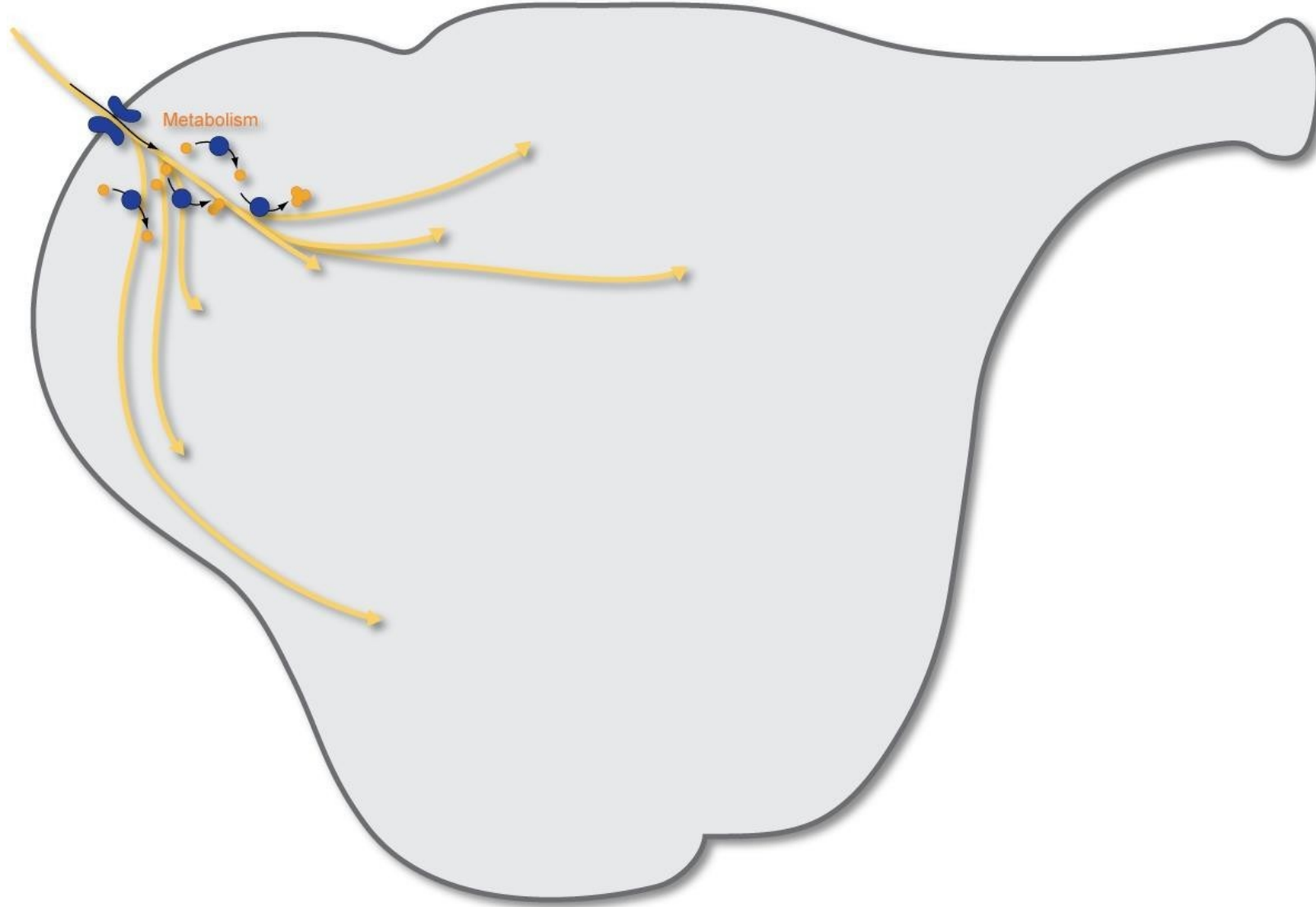
Prever prognósticos  
Otimizar a terapia  
Maximizar a qualidade  
de vida

**Desafio central: prever o fenótipo do genótipo**

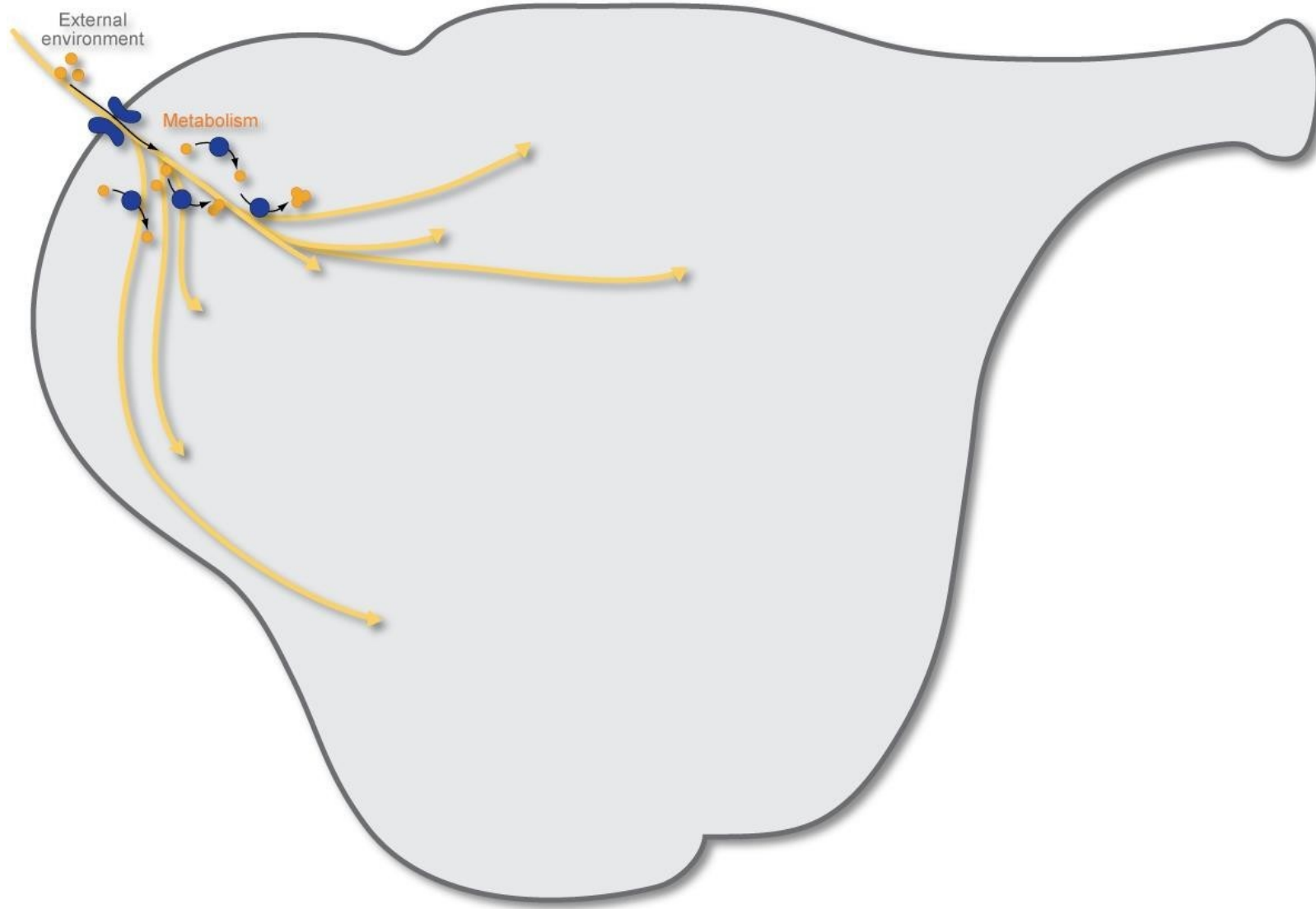
# Exemplo: biossíntese de drogas



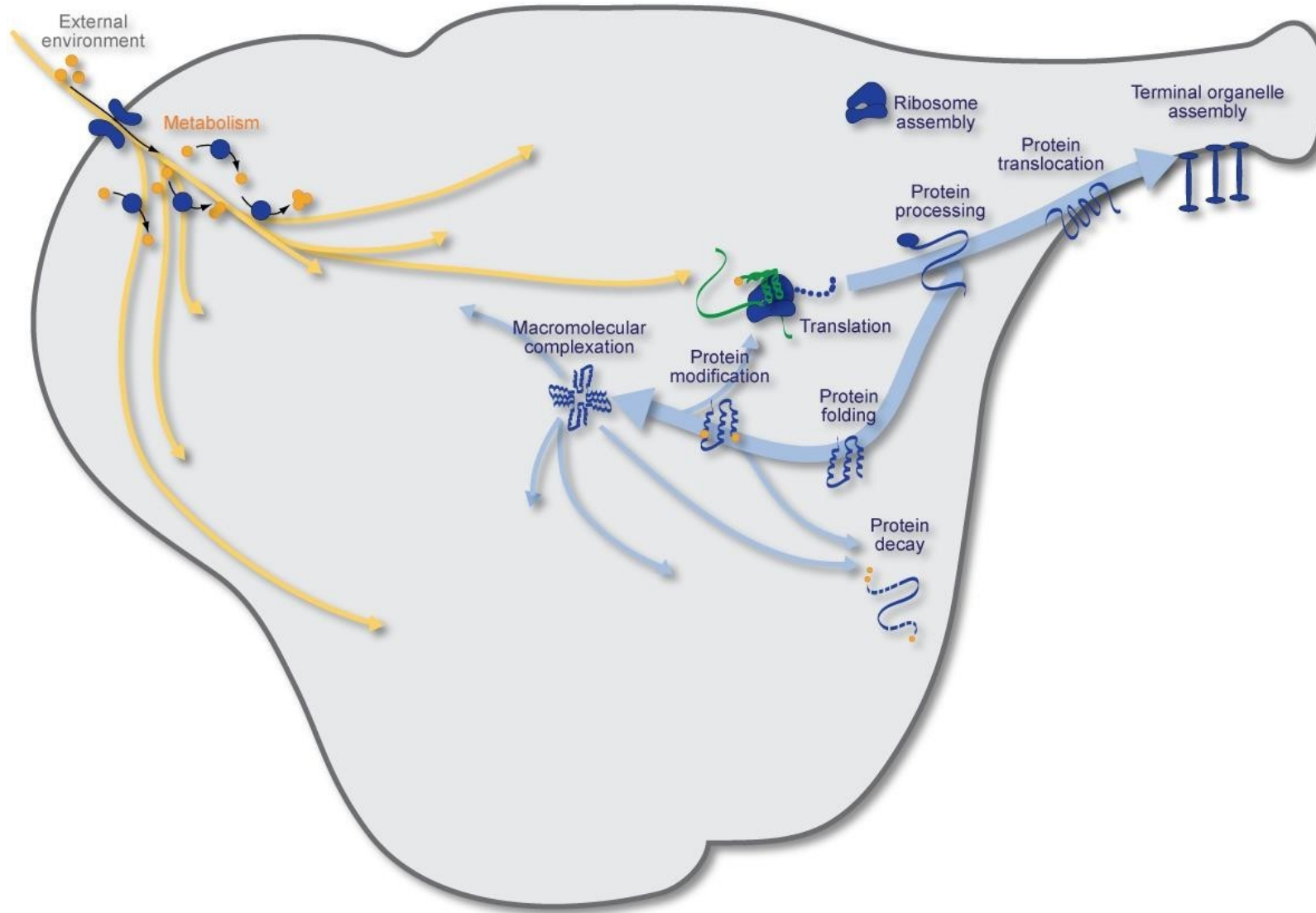
# Exemplo: biossíntese de drogas



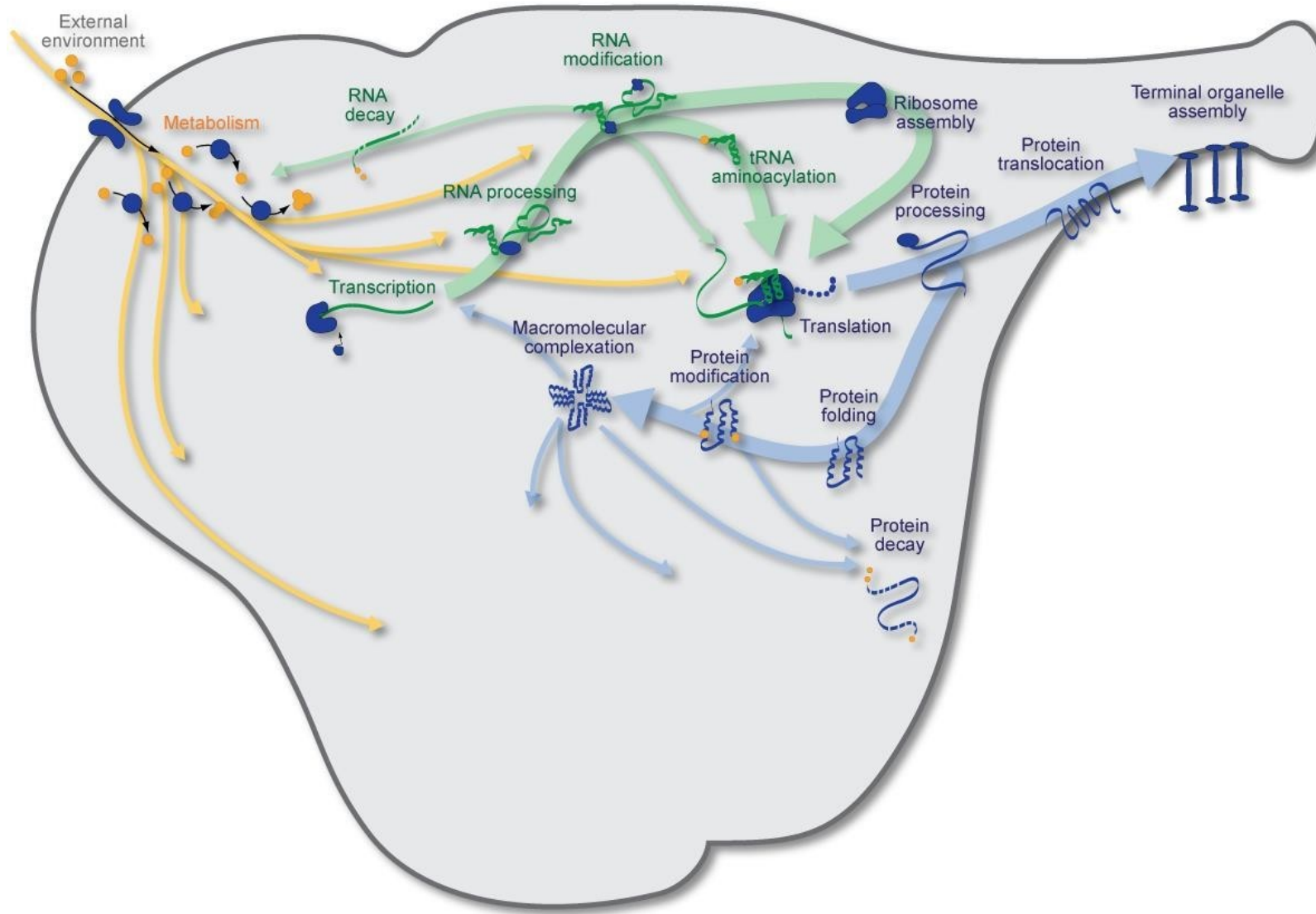
# Exemplo: biossíntese de drogas



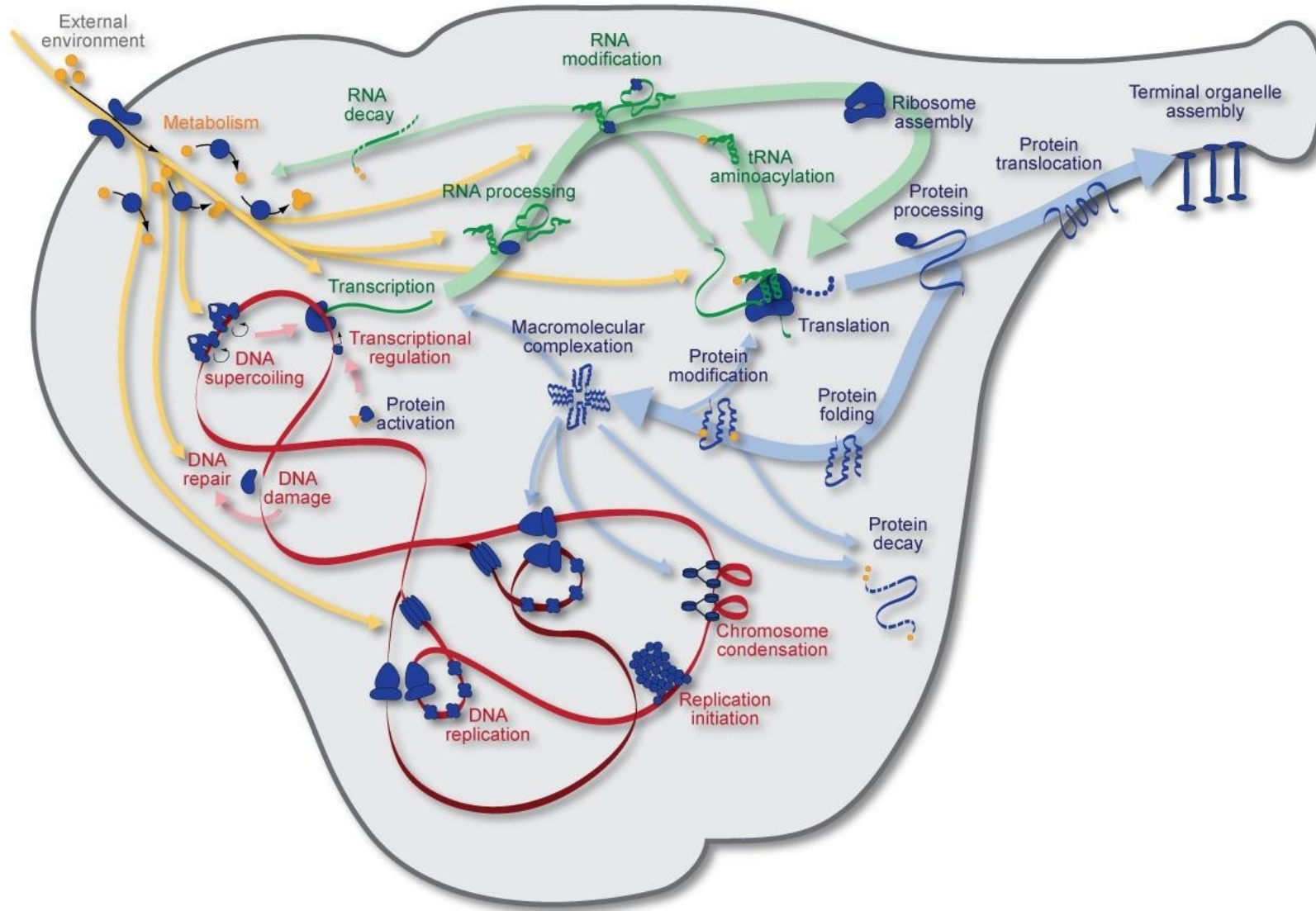
# Example: drug biosynthesis



# Exemplo: biossíntese de drogas

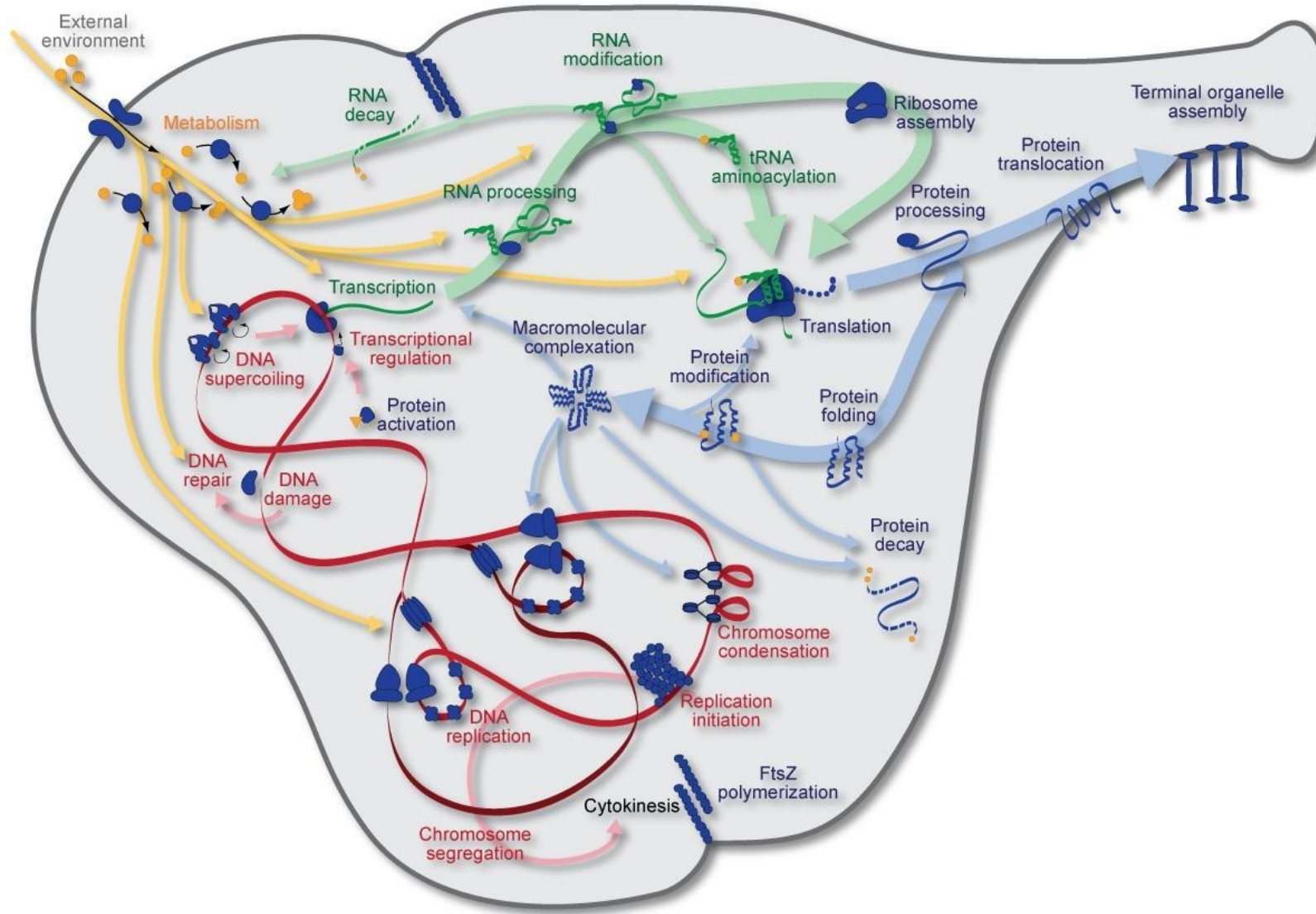


# Exemplo: biossíntese de drogas

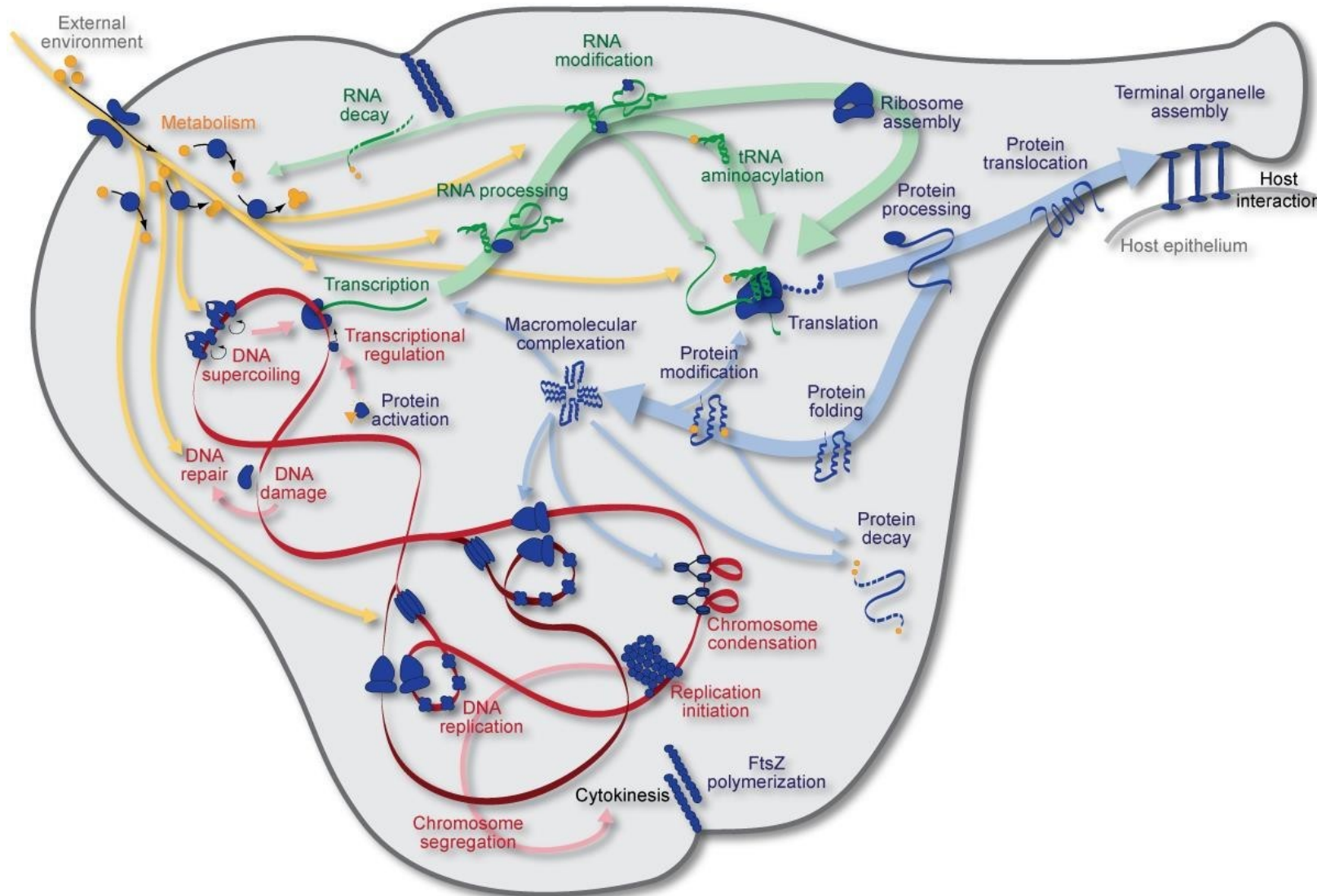




# Exemplo: biossíntese de drogas

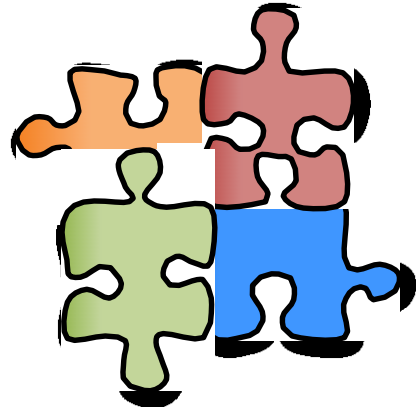


# Exemplo: biossíntese de drogas

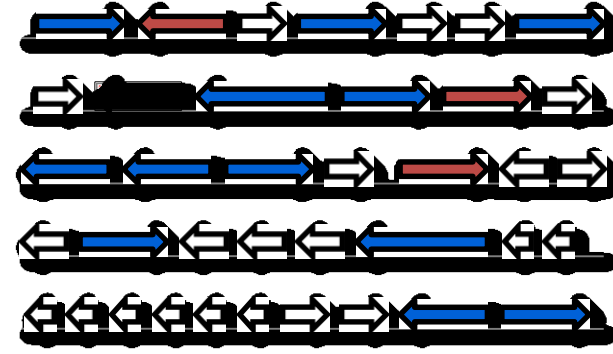


A previsão do fenótipo a partir do genótipo requer modelos “célula inteira”

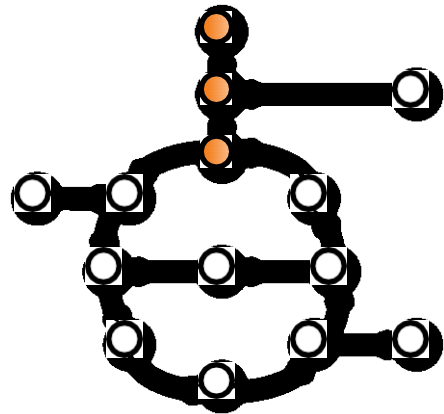
# Whole-cell modeling principles



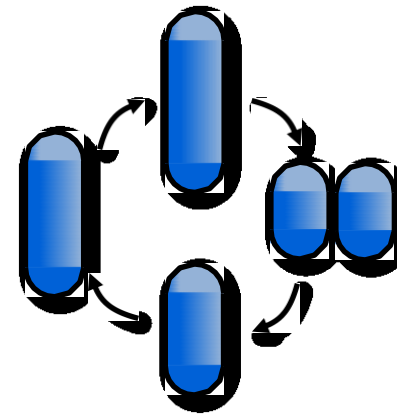
integrated



Gene-complete



Comprehensive



Dynamic



“A modelagem de células inteiras pode permitir bioengenharia racional e medicina de precisão.”



“Propomos vários princípios-chave da modelagem de células inteiras.”



“Descrevemos como construir um modelo de célula inteira, destacando as ferramentas disponíveis.”

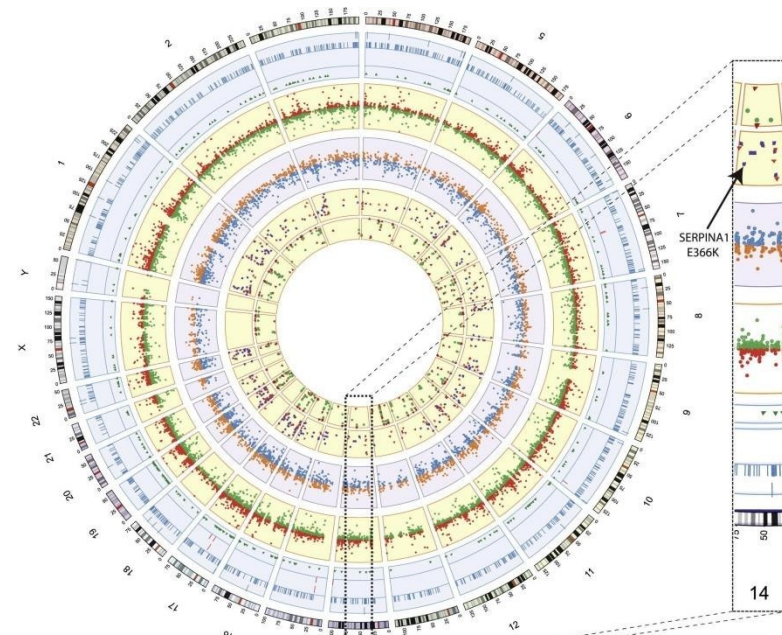


“Descrevemos os principais desafios de modelagem computacional e experimental de células inteiras em código aberto”

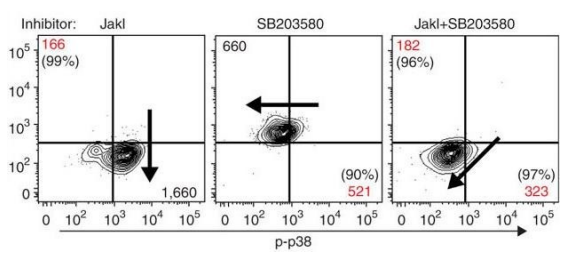
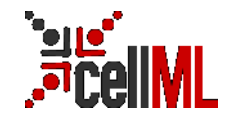
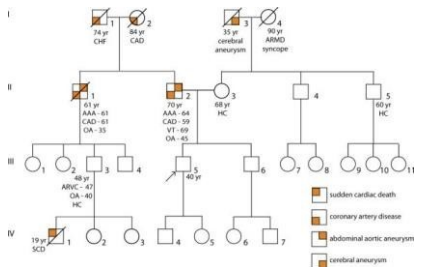
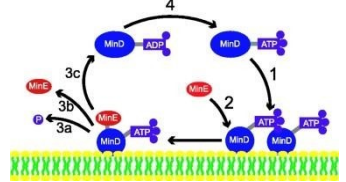
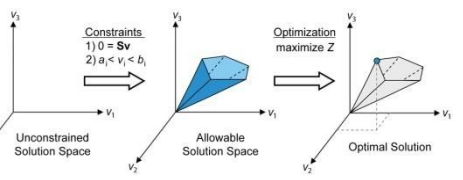
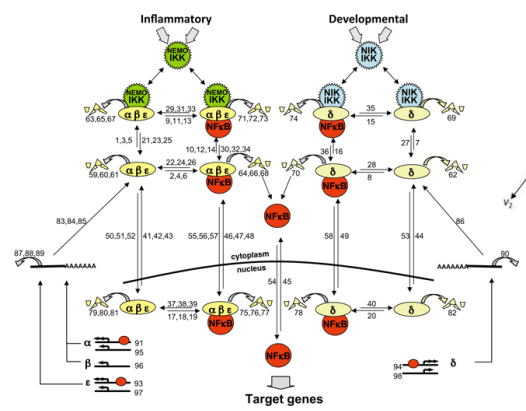
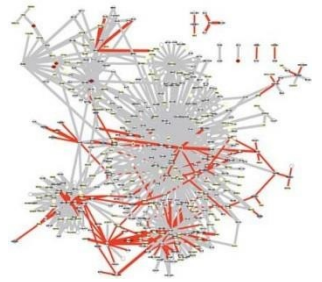
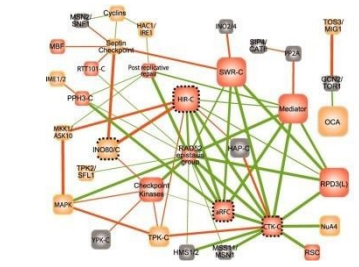
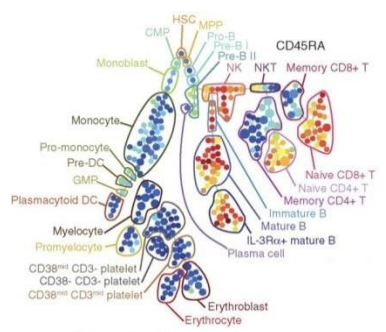
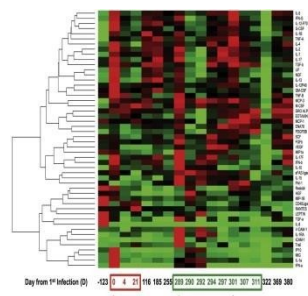
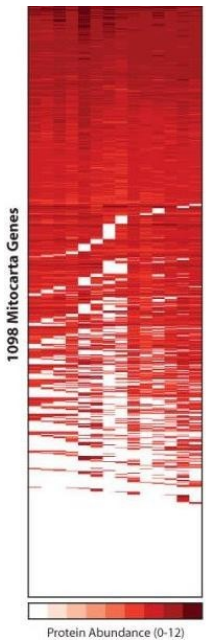


“Alcançar modelos completos de células inteiras requer uma forte comunidade interdisciplinar.”

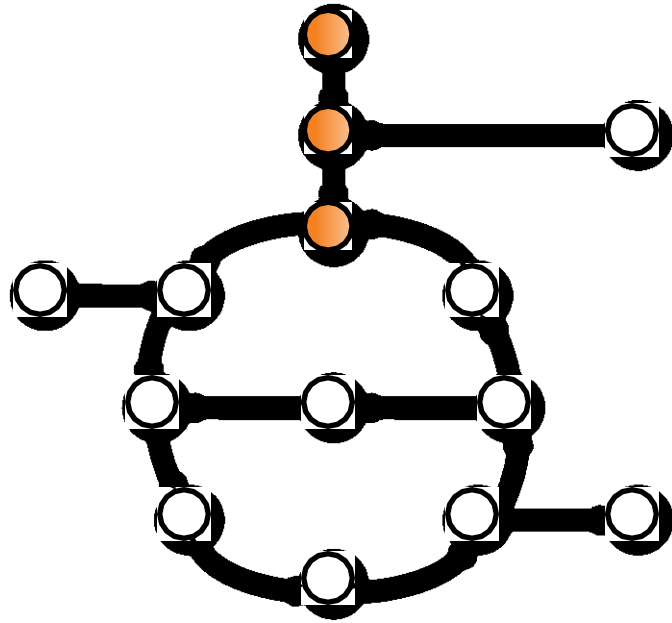
# Disponibilidade dos Dados



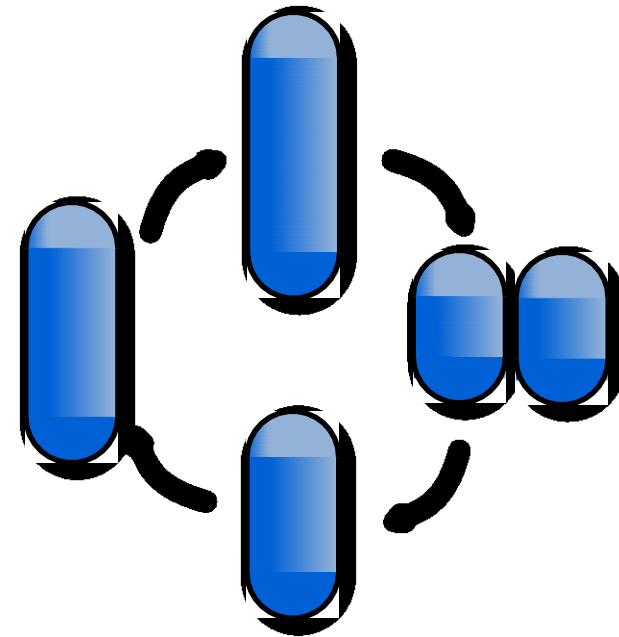
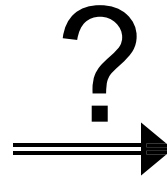
- RNA Edits
- Heteroallelic SNVs
- Protein-Downregulated (HRV vs Healthy)
- Protein-Upregulated (HRV vs Healthy)
- RNA-Downregulated (HRV vs Healthy)
- RNA-Upregulated (HRV vs Healthy)
- Indels
- SV-Duplications
- SV-Deletions
- Chr. Ideogram
- Chr. Number



# Objetivos do Whole-cell model



Data



Knowledge

# Whole-cell modeling

A grand challenge of the 21<sup>st</sup> century

– *Masaru Tomita*

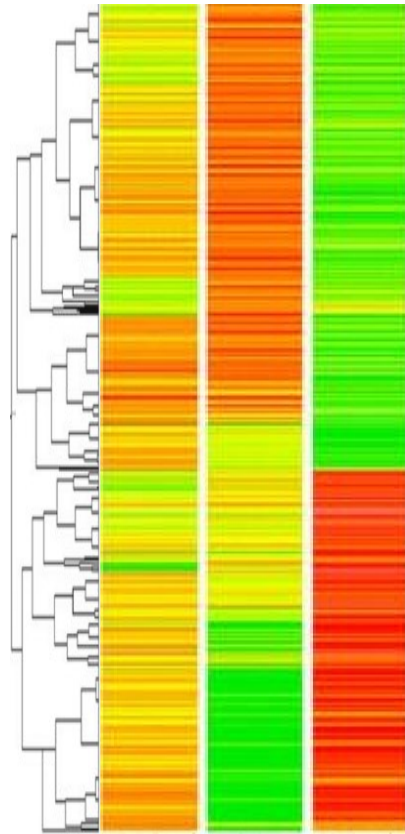
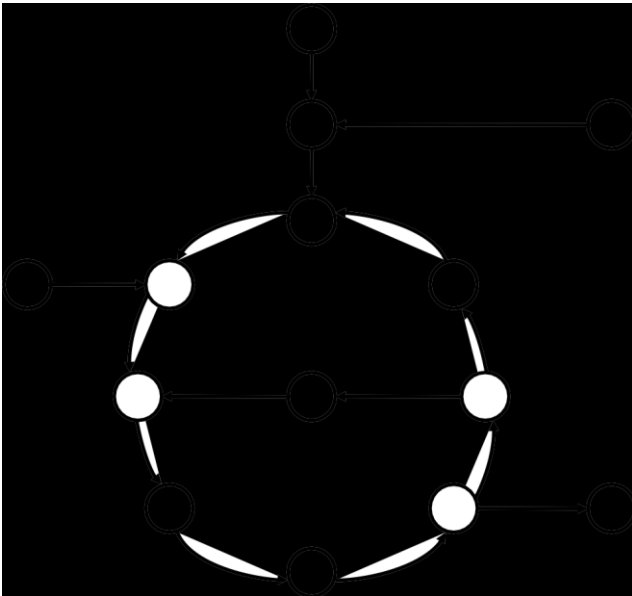
A biologia precisa urgentemente de uma base teórica para unificá-la

– *Sydney Brenner*

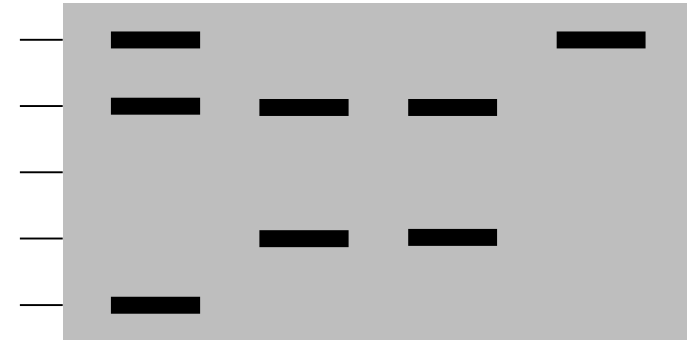
O teste final de entender uma célula simples, mais do que ser capaz de construir uma, seria construir um modelo computacional da célula.

– *Clyde Hutchison*

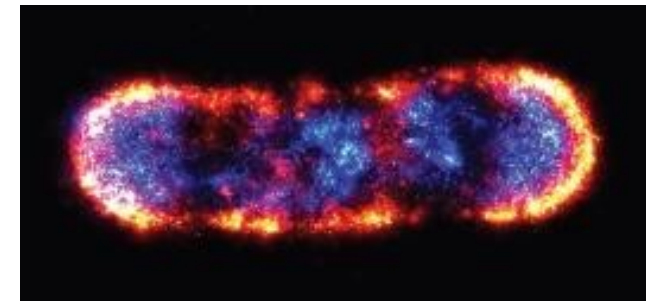
# Desafio de modelagem de dados



Transcription  
*RNA-seq*



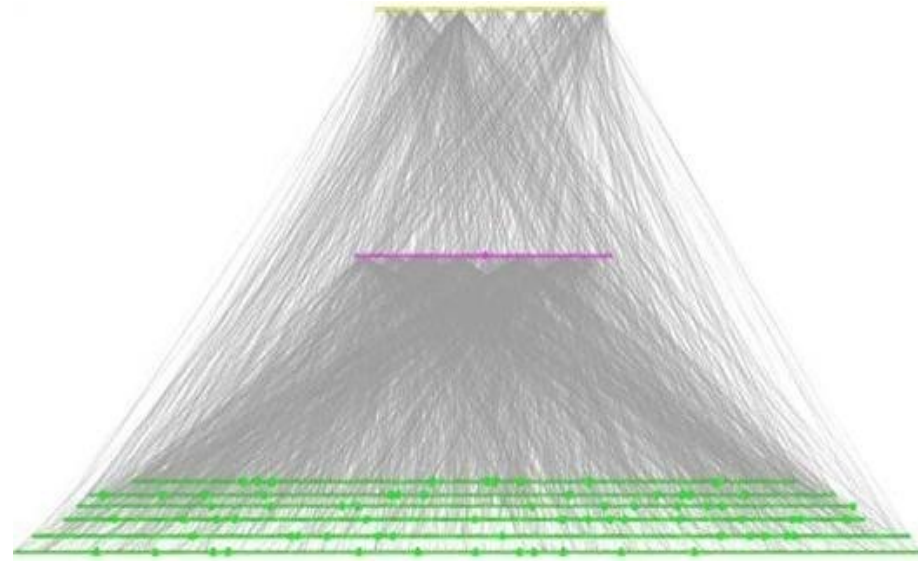
Protein expression  
*Mass-spec, Western blot*



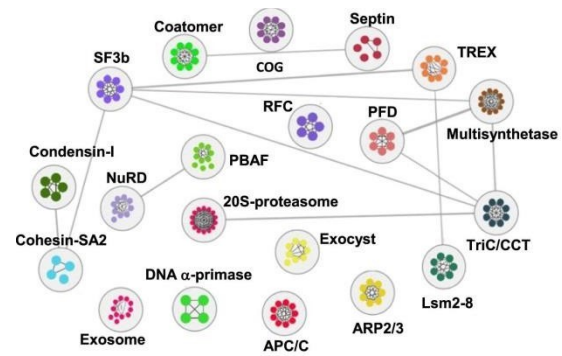
Single-cell variation  
*Microscopy*



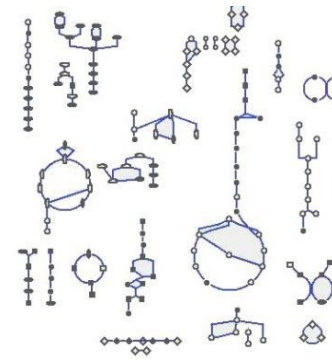
# Modelling challenge



Transcriptional regulatory

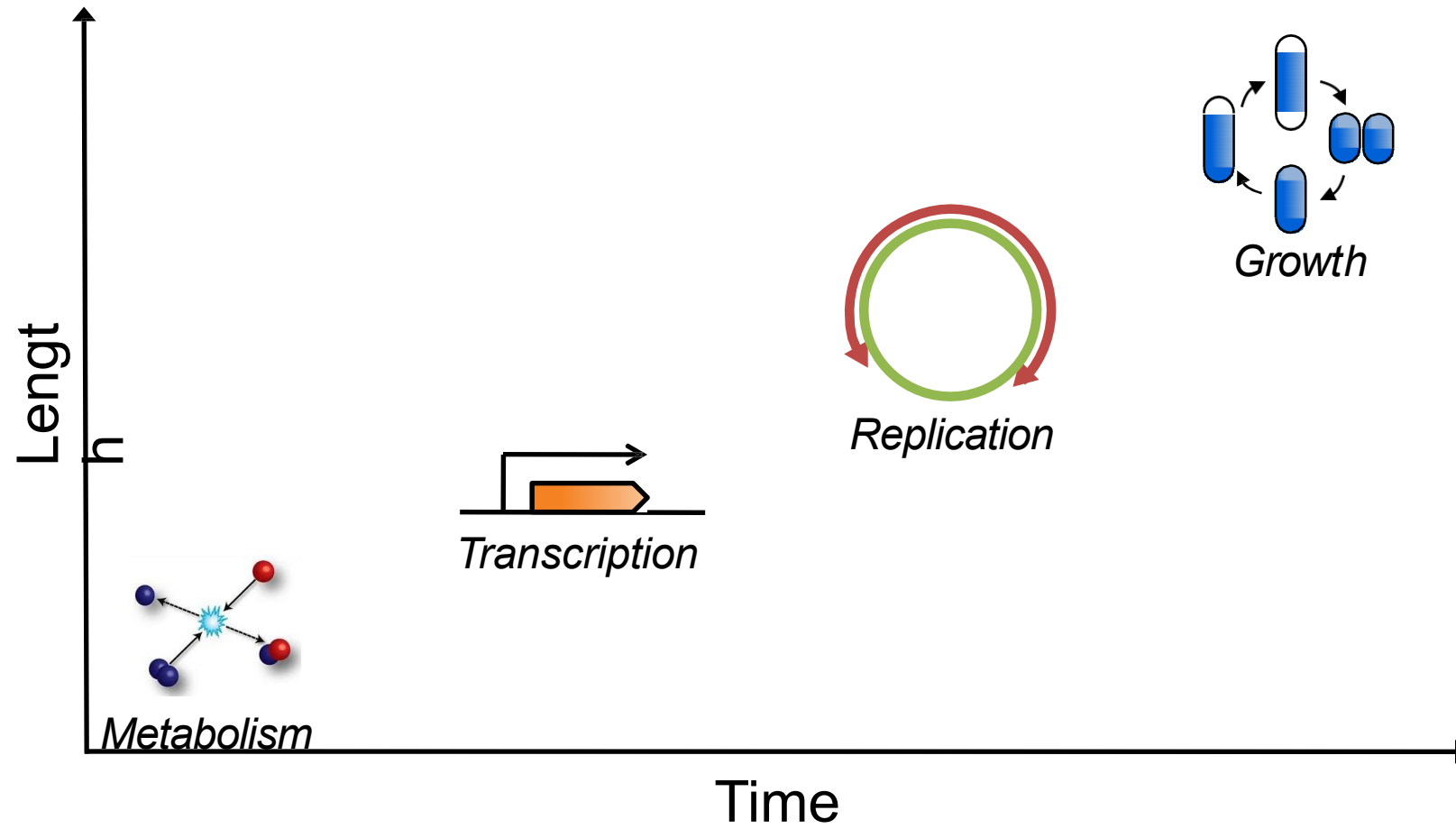


Signaling

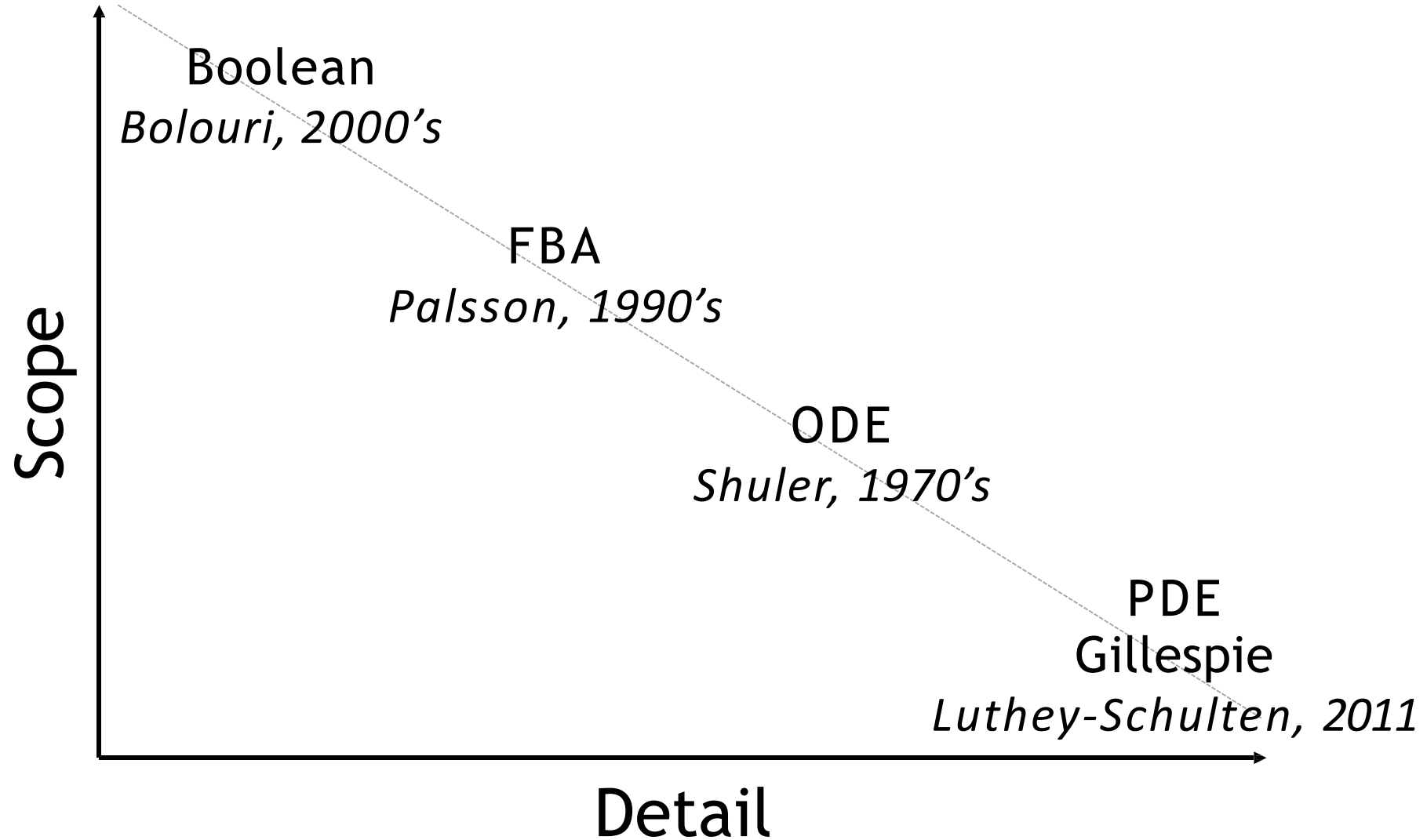


Metabolic

# Modeling challenge

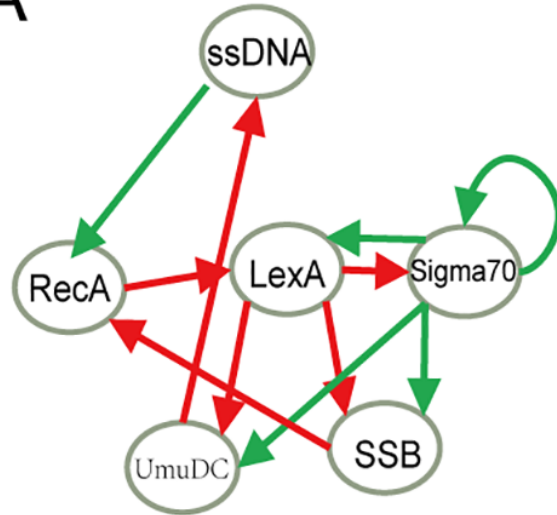


# Predictive modeling methodologies



# Predictive modeling methodologies

A

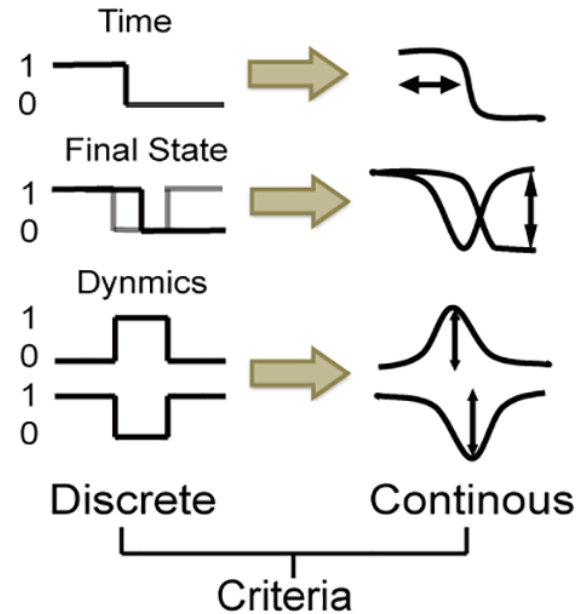


B

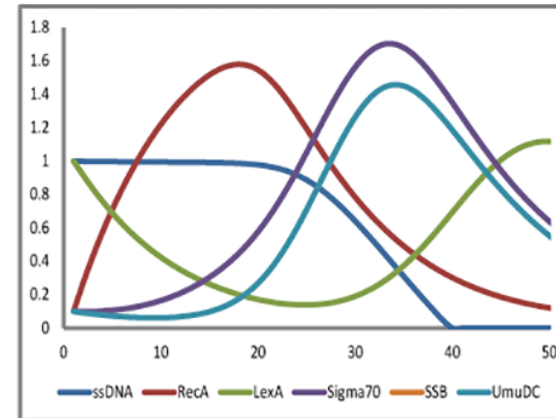
time	ssDNA	RecA	LexA	Sigma70	UmuDC	SSB
1	1	0	1	0	0	0
2	1	1	1	0	0	0
3	1	1	0	0	0	0
4	1	1	0	1	0	0
5	1	1	0	1	1	1
6	0	0	0	1	1	1
7	0	0	1	1	1	1
8	0	0	1	0	0	0

Boolean network model

C

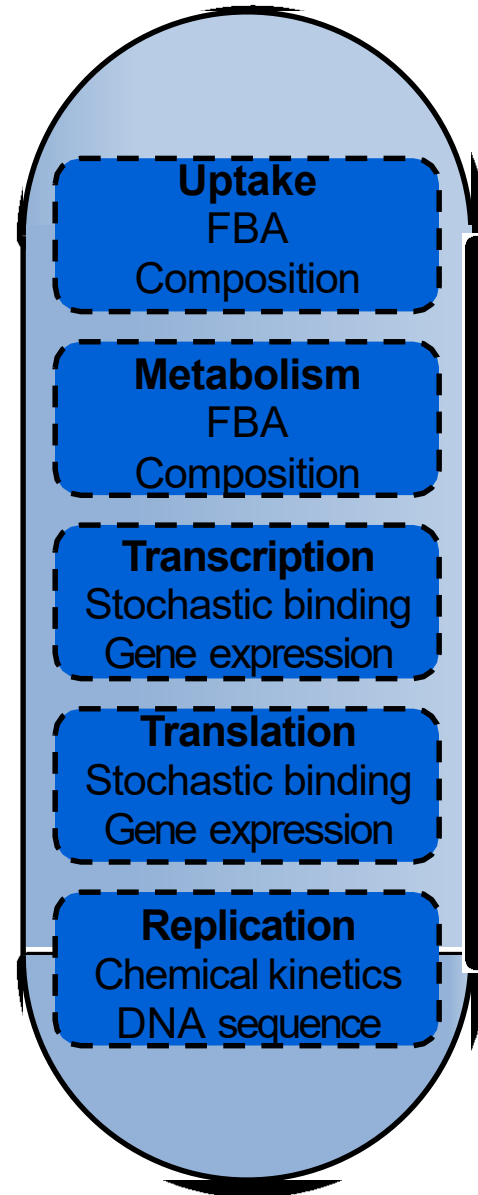


D

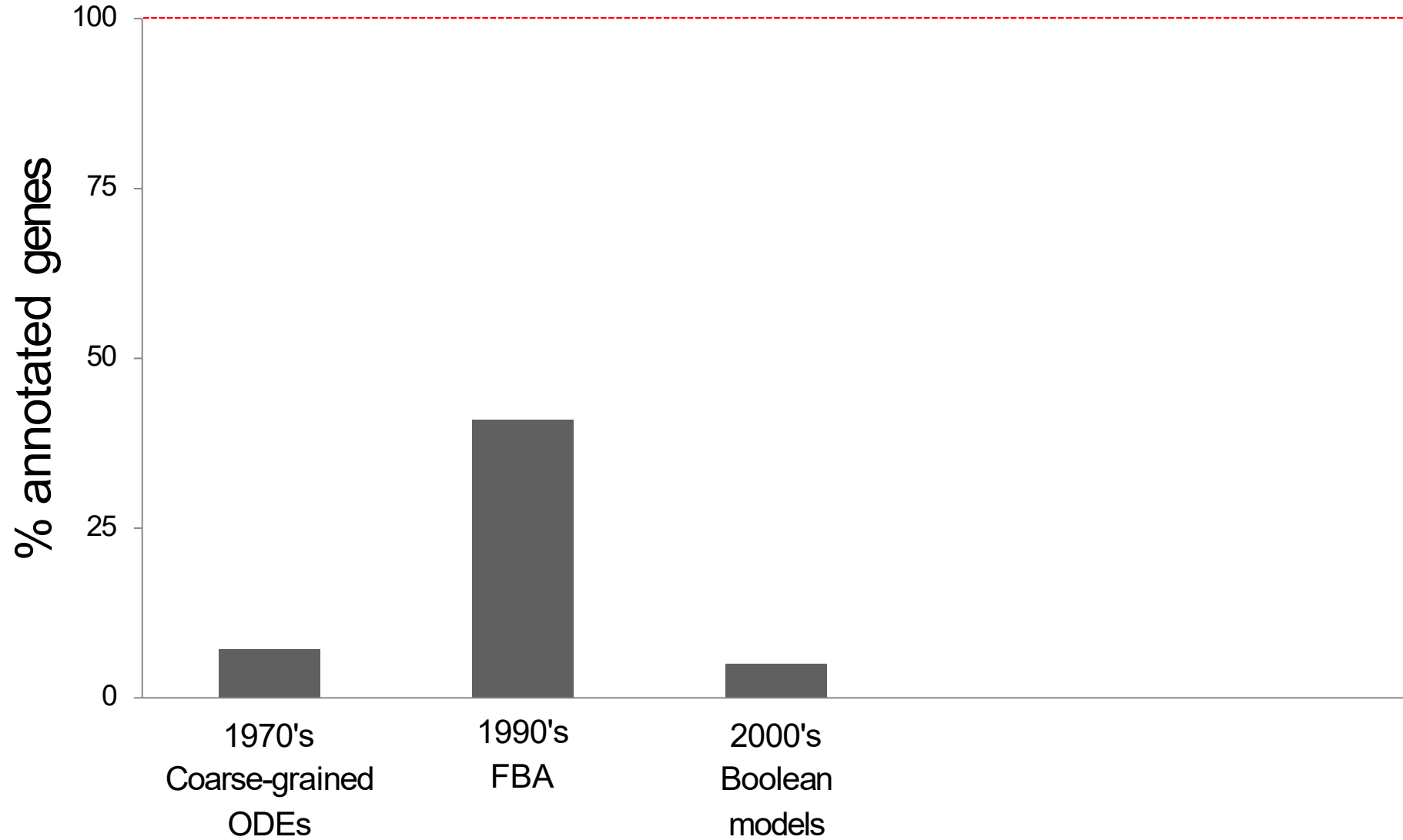


ODE model

# Solution: integrated models

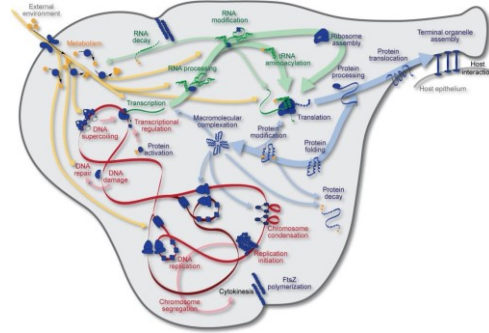


# Whole-cell modeling progress

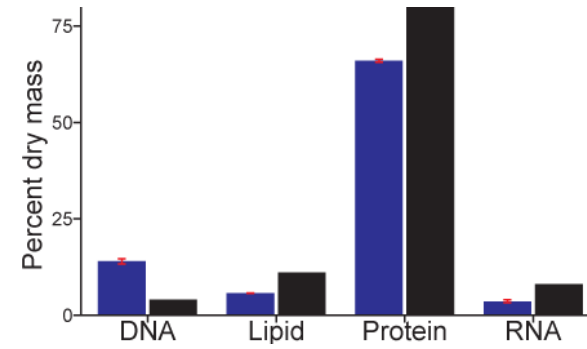


# Whole-cell modeling

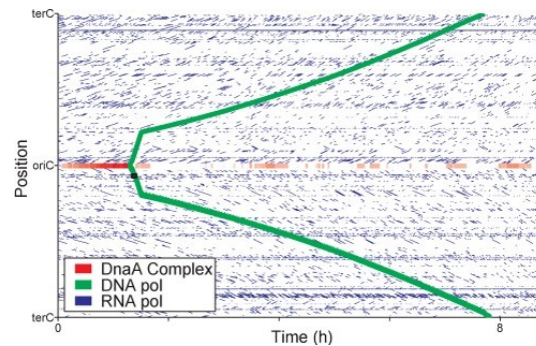
## Model



## Validate

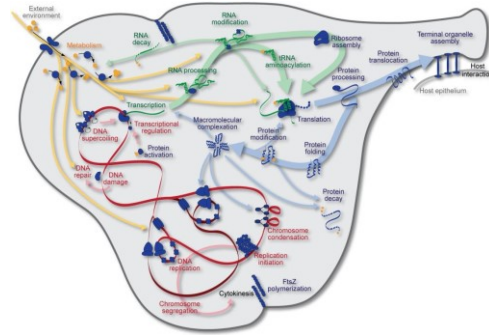


## Engineer

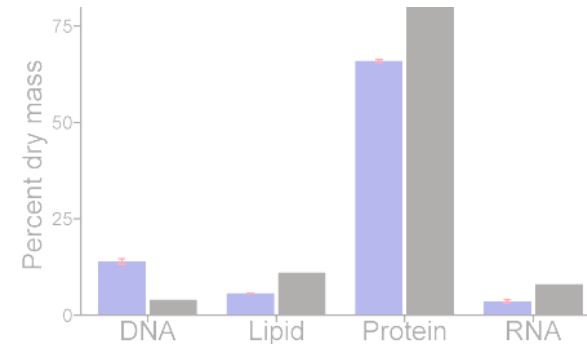


# Whole-cell modeling

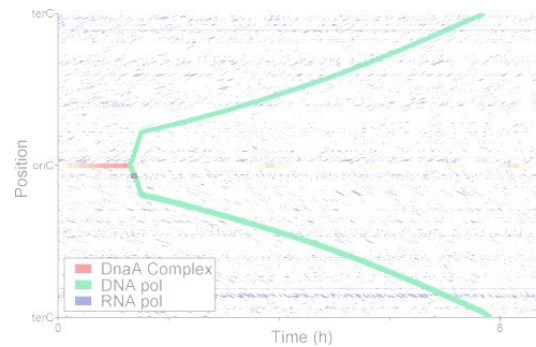
Model



Validate



Engineer

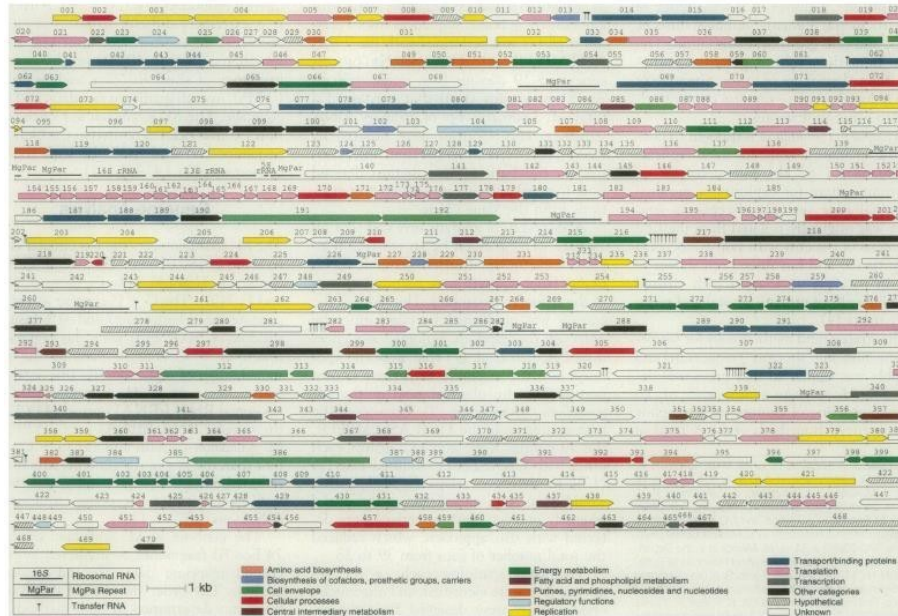




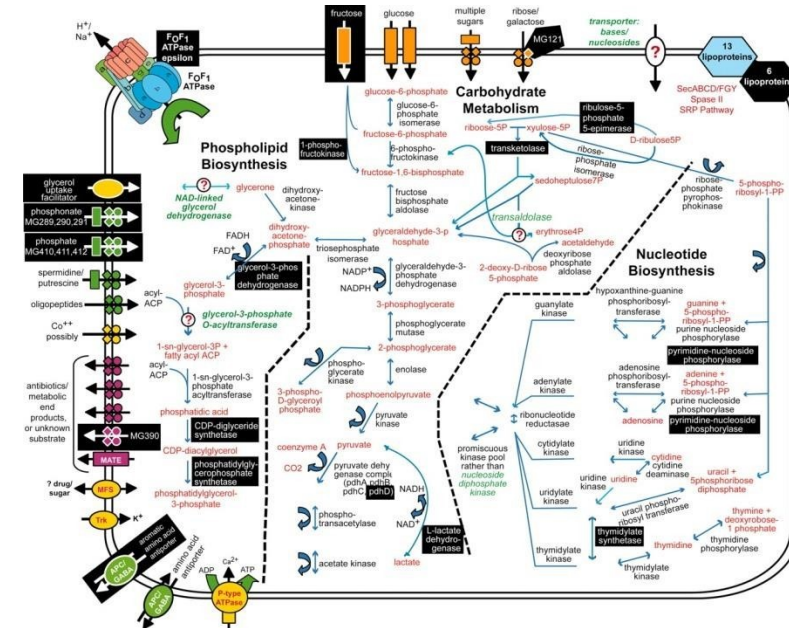
# Model construction

1. Definir sistema
2. Definir escopo
3. Curar dados
4. Escolha a representação
5. Identifique parâmetros
6. Testar previsões

# M. genitalium is well-characterized

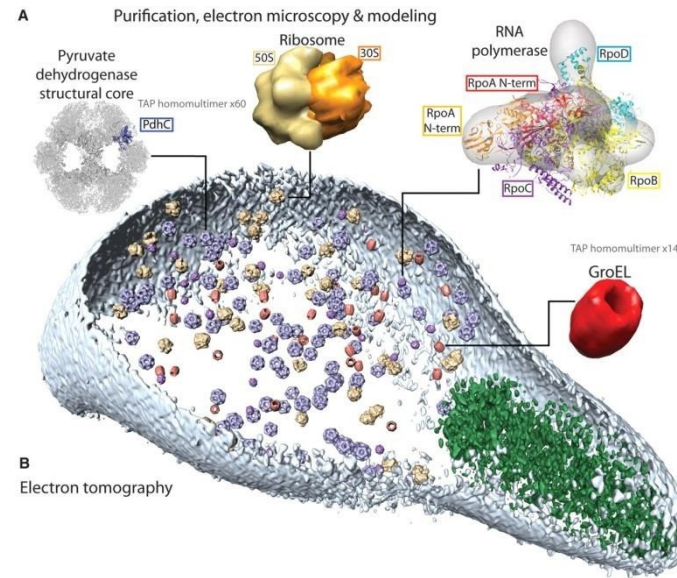


**Comparative genomics**  
Fraiser et. al, 1995



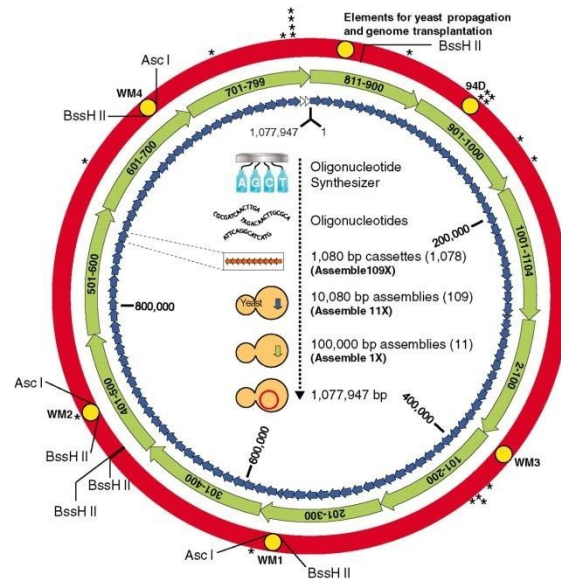
**Genome-wide essentiality**  
Glass et. al, 1999

# *M. genitalium* is well-characterized

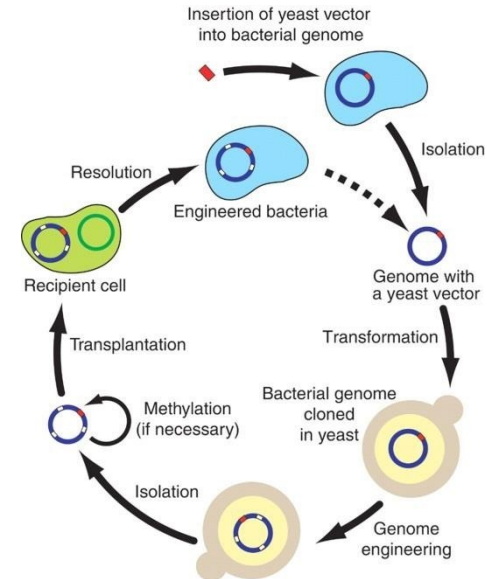


**Genomic-scale data**  
Kühner et. al, 2009

# *M. genitalium* has unique engineering tools

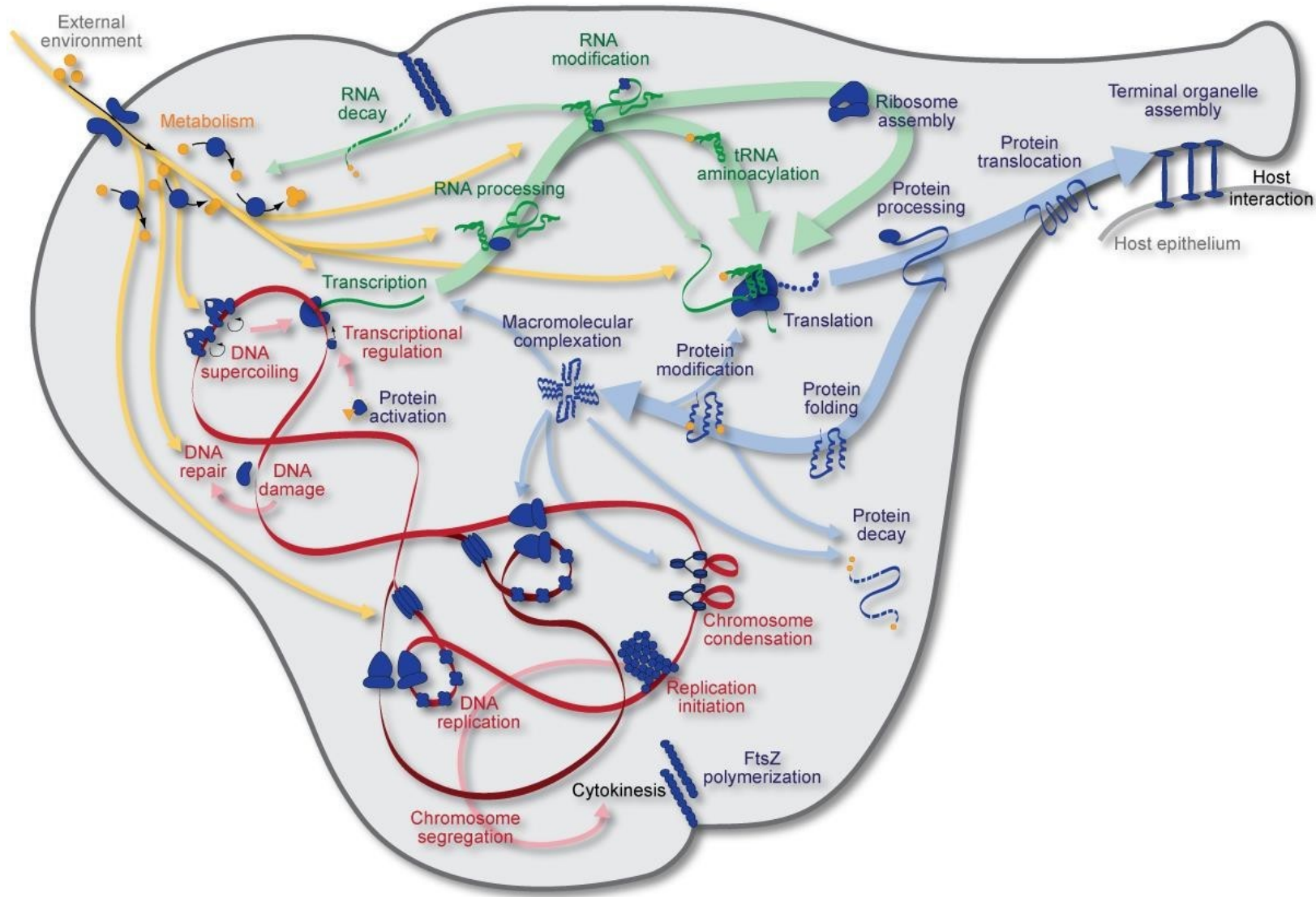


**Genomic synthesis**  
Gibson et. al, 2009



**Genomic transplantation**  
Lartigue et. al, 2009

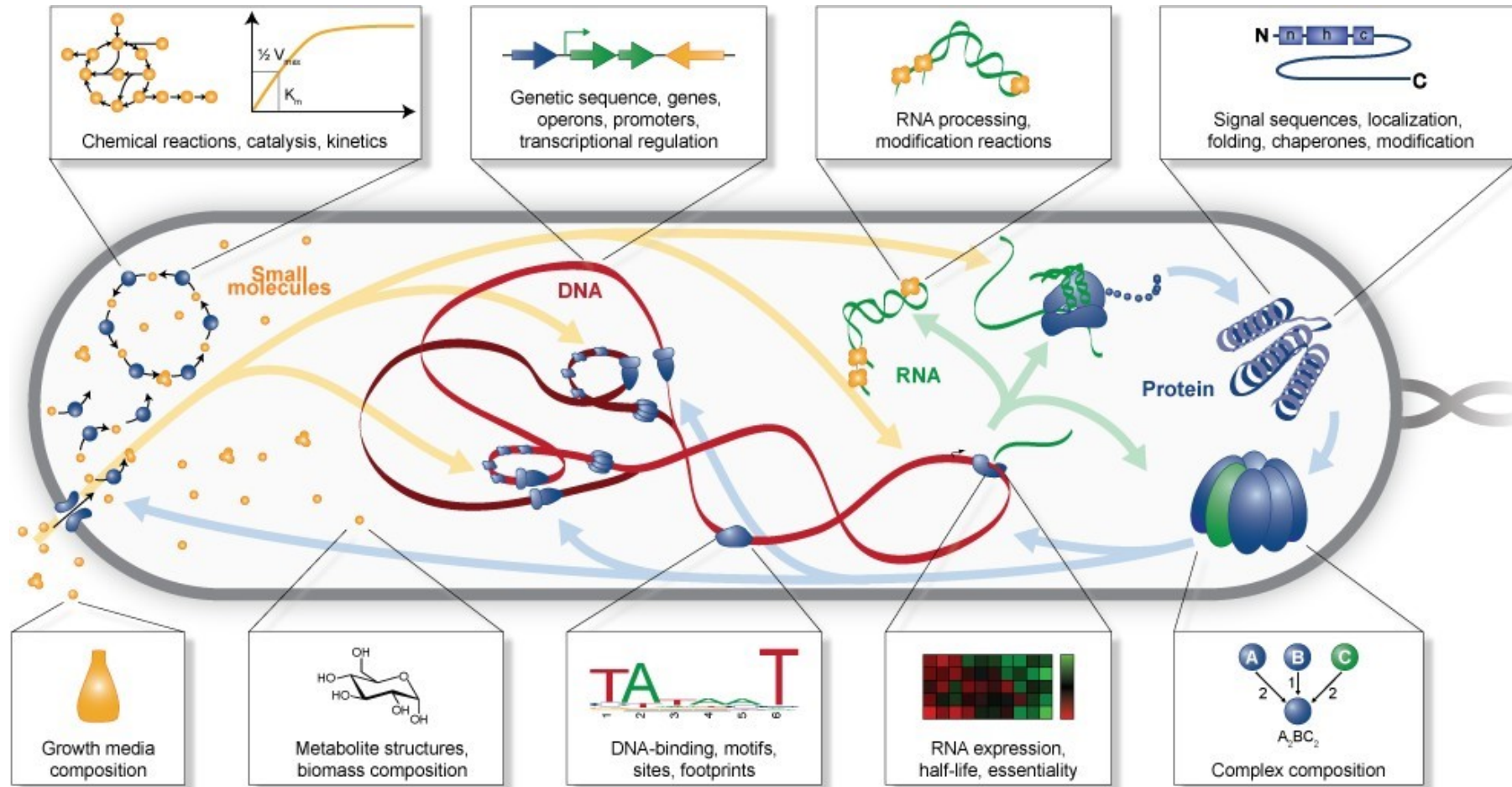
## 2. Escolher o escopo do modelo



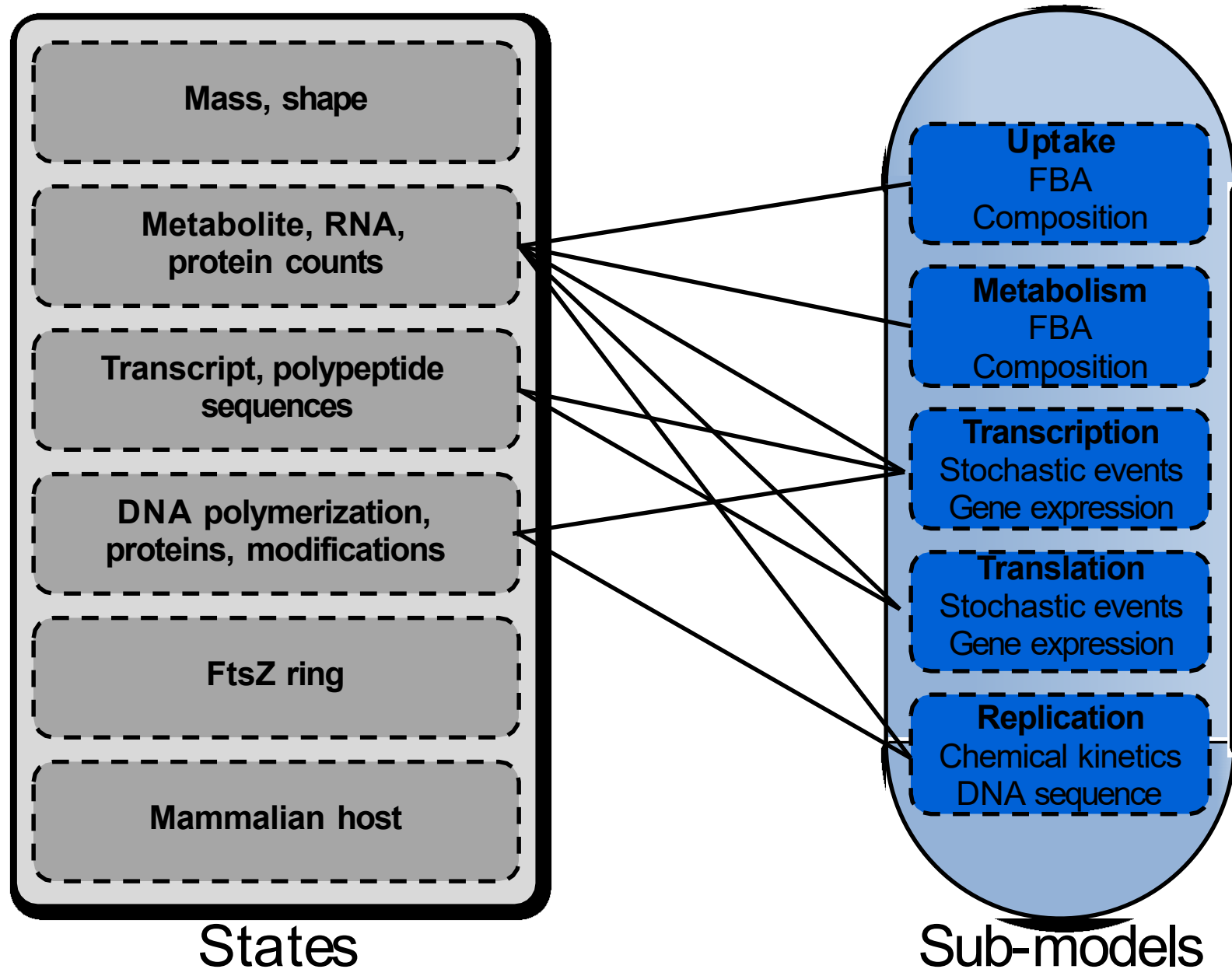
## 2. Escolher o escopo do modelo

- Representam explicitamente cada metabólito, gene, RNA e proteína;
- Modelam explicitamente a função de cada produto gênico caracterizado;
- Explica o custo metabólico de todos os produtos gênicos não caracterizados
- Representa moléculas importantes e bem caracterizadas individualmente

# 3. Fazer curadoria dos dados

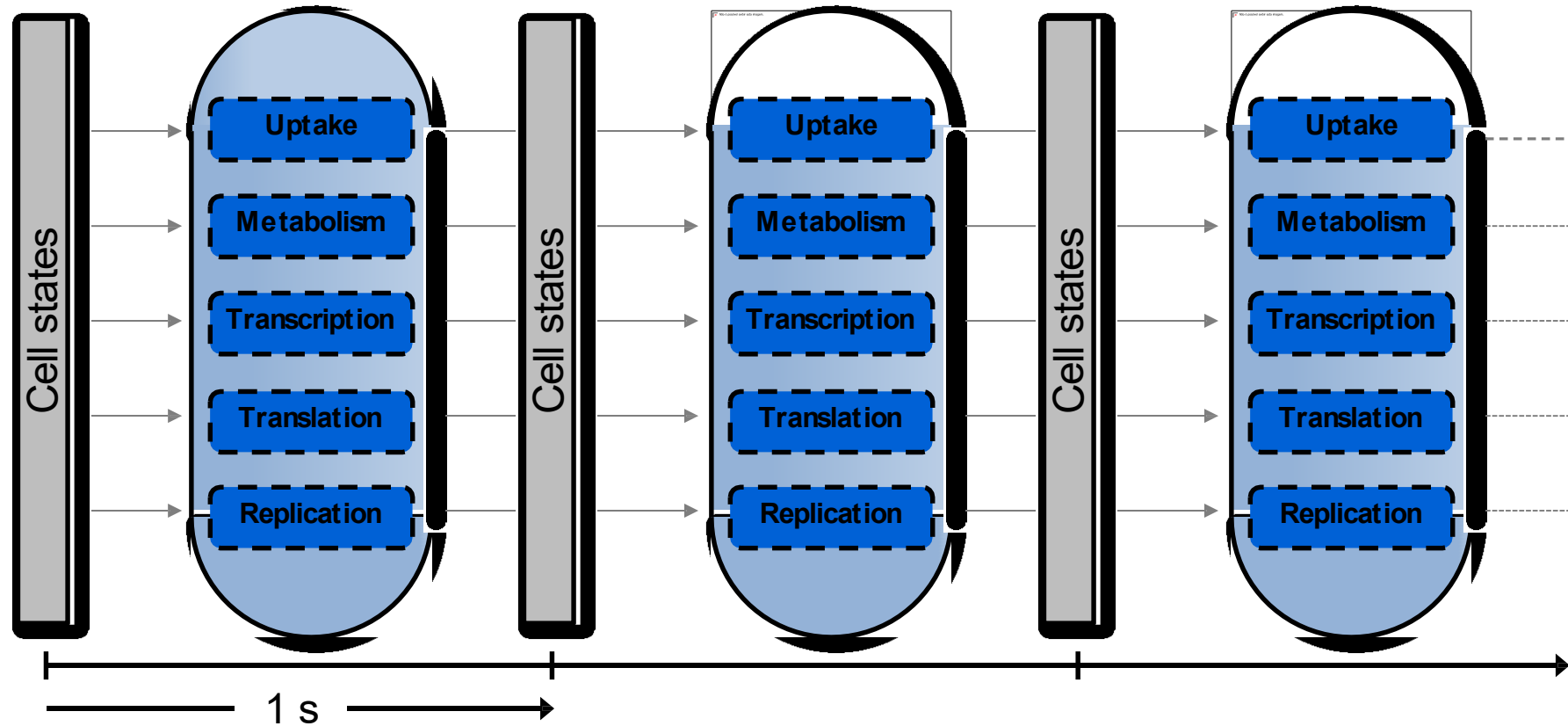


## 4. Selecionar a representação matemática mais adequada

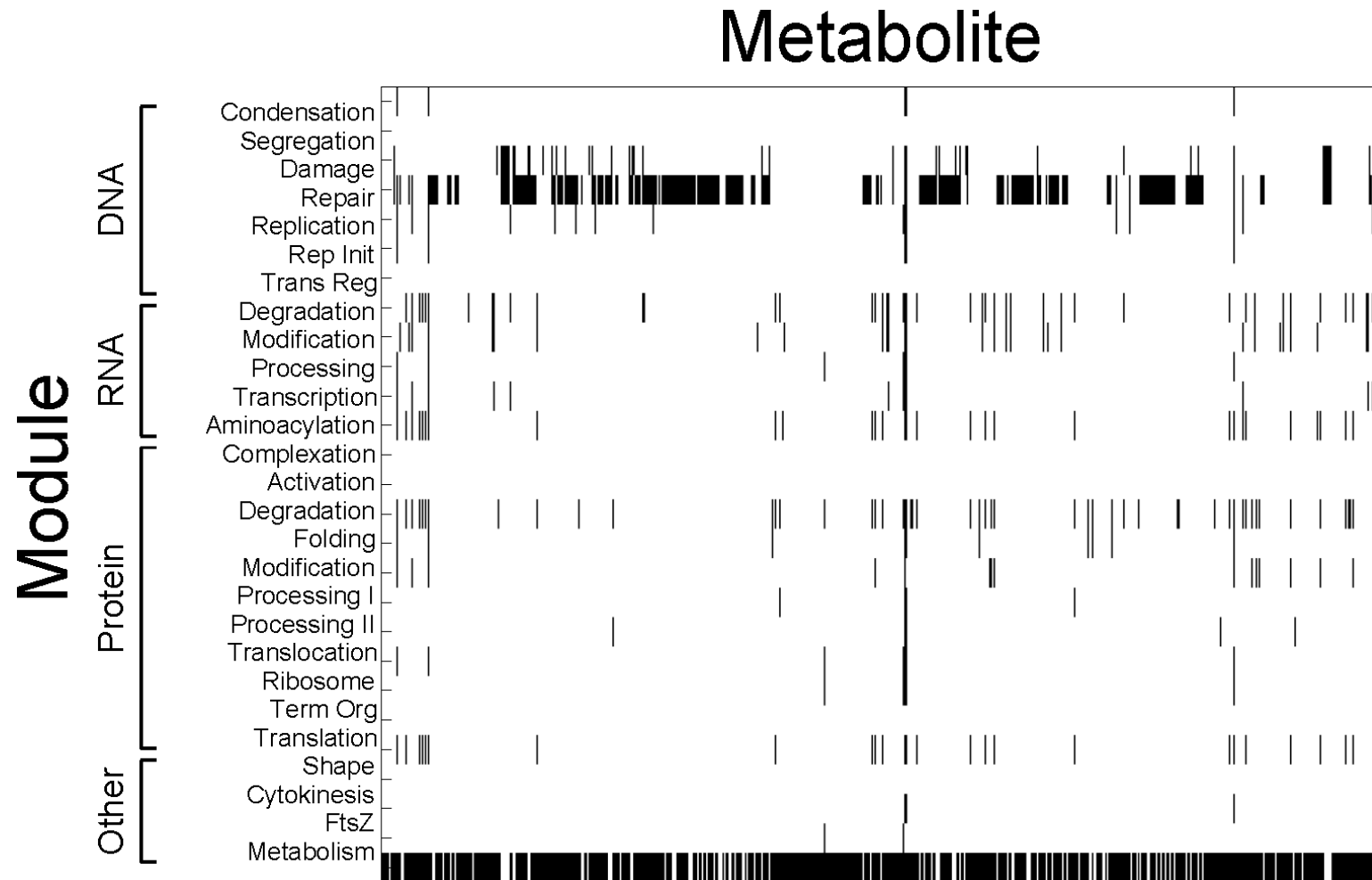




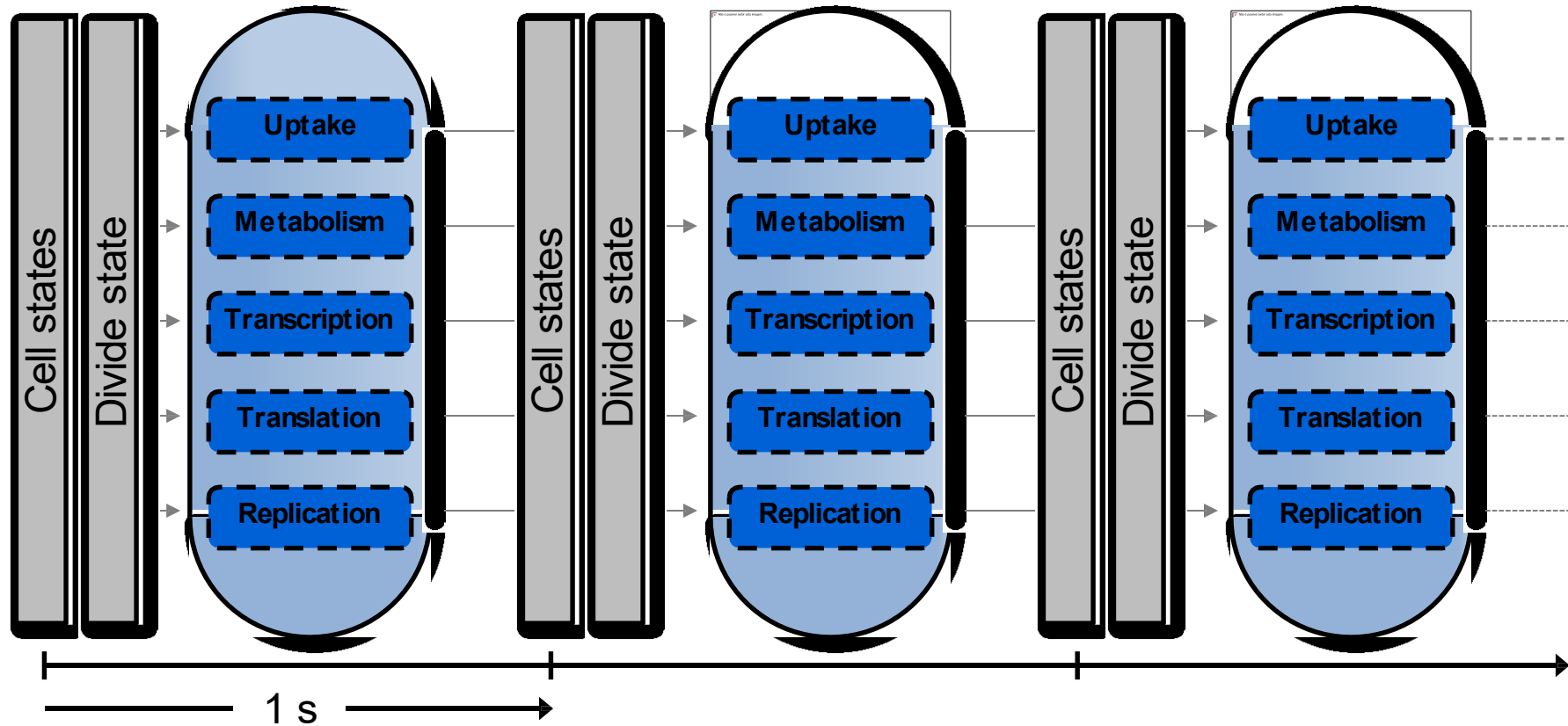
# Simular o algoritmo



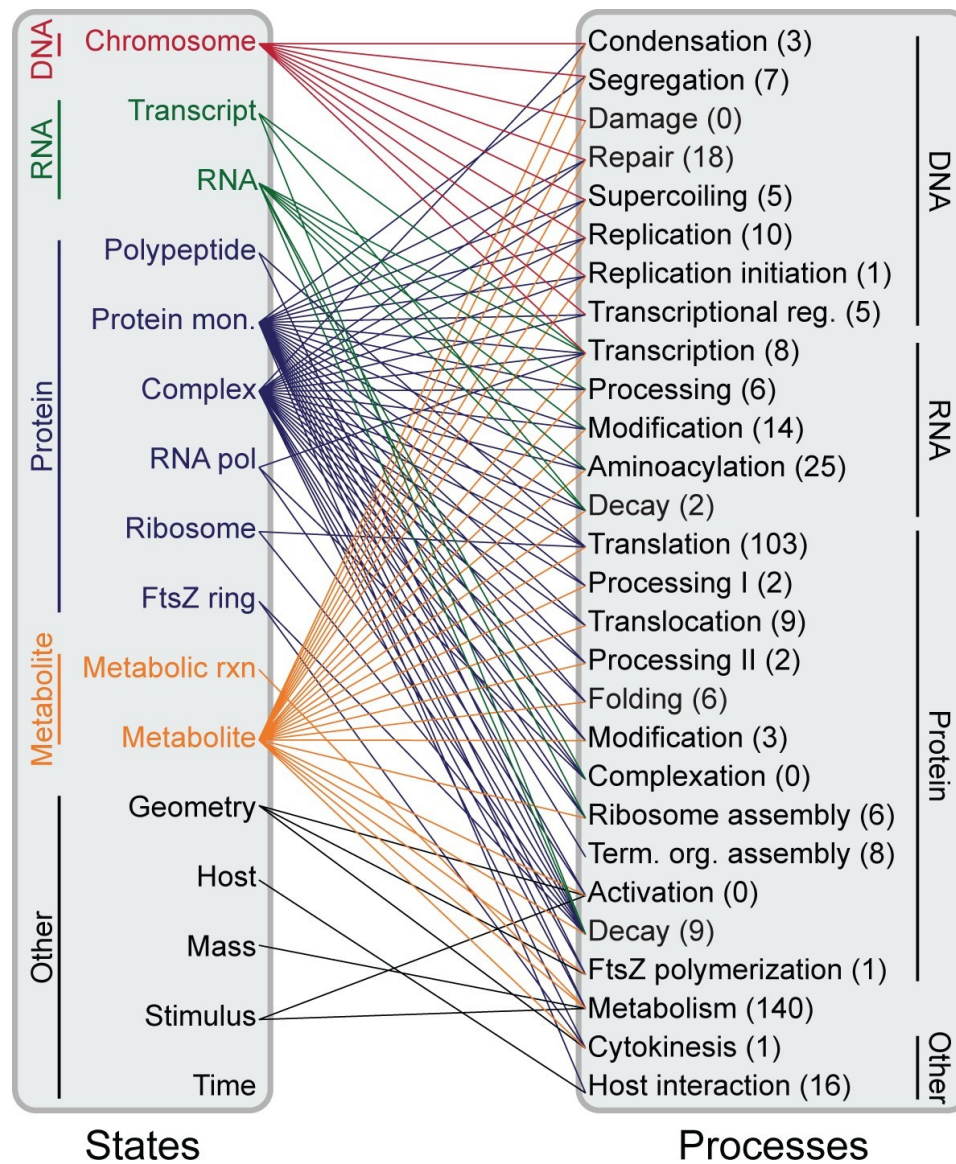
# Muitos recursos são compartilhados



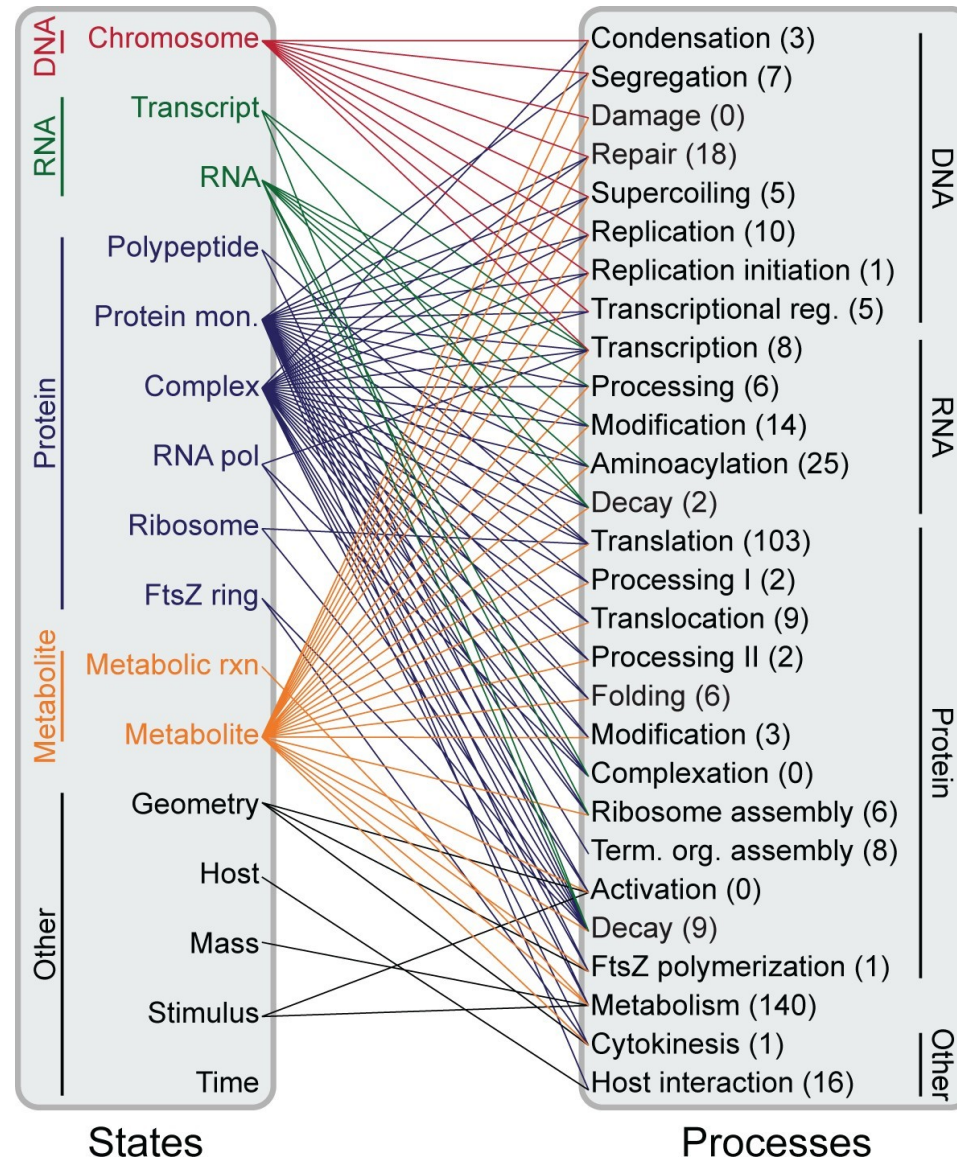
# Simular o algoritmo



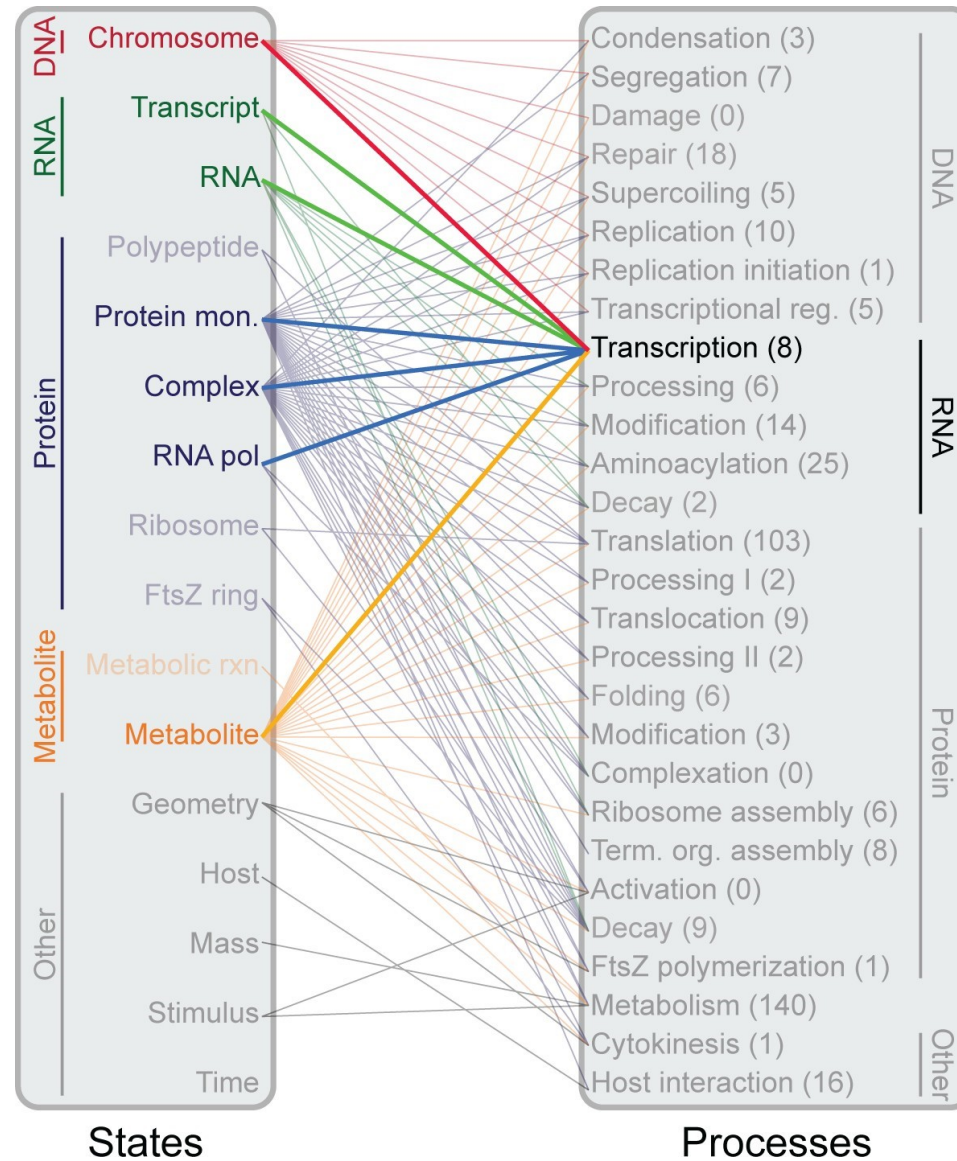
# O modelo da Mycoplasma contém 28 sub-models



# Example sub-model: Transcription

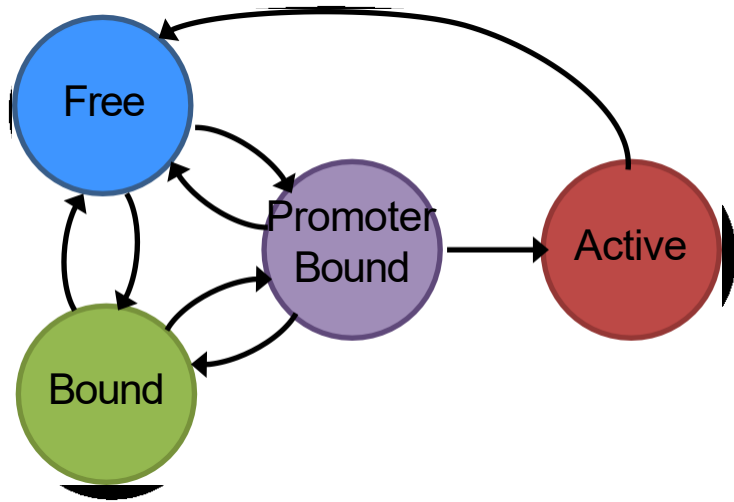


# Example sub-model: Transcription

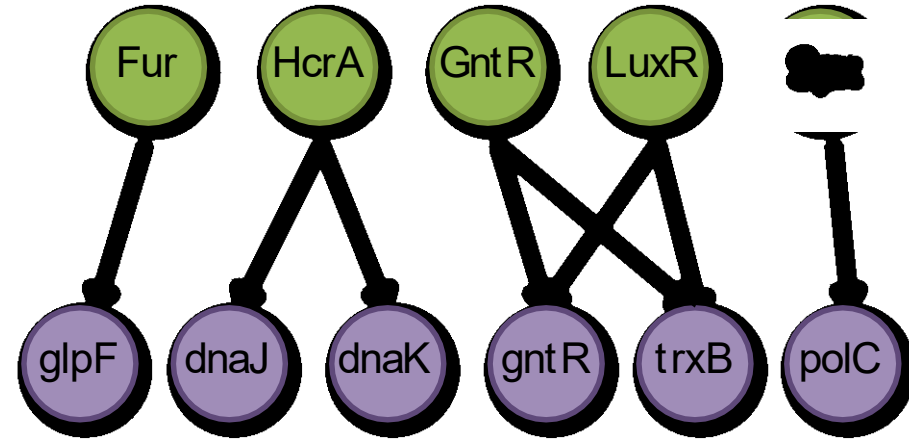


# Example sub-model: Transcrição

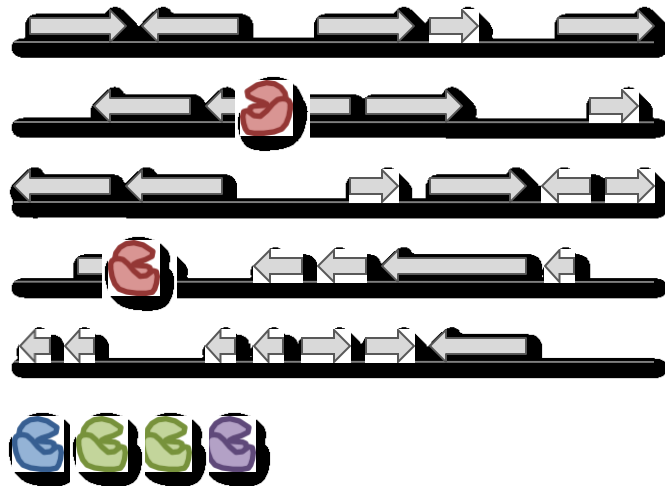
1. Update RNA polymerase states



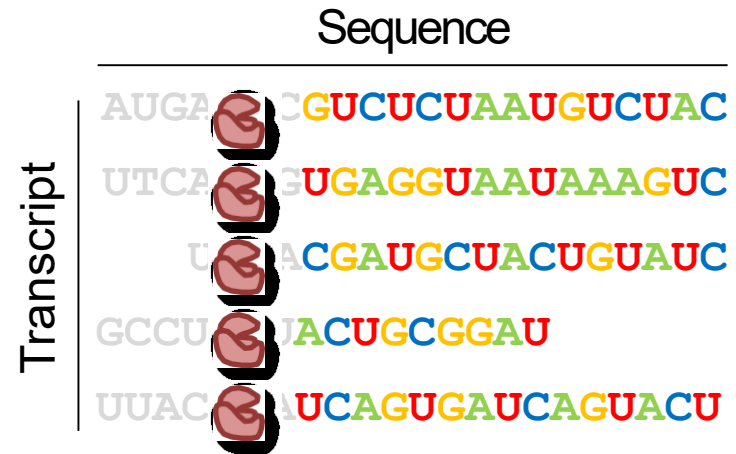
2. Calculate promoter affinities



3. Bind RNA polymerase

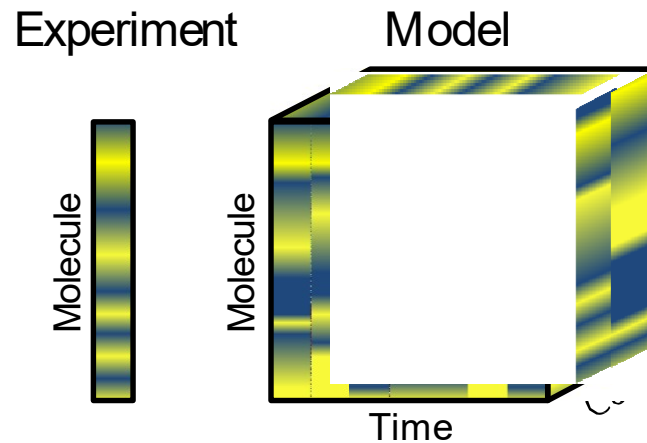


4. Elongate and terminate transcripts



# A redução do modelo permite a identificação de parâmetros

## 1. Reduce model



## 2. Identify reduced model parameters using traditional methods

## 3. Manually tune parameters using full model



# Software: wholecell.org

**=cell**

Volume 150 Number 2 July 20, 2012 www.cell.com

**TO INCLUDE??**

- DNA replication, binding
- Collision free
- Protein translation, Ribosome pausing, cell cycle, cleavage, compartment, regulation
- RNA Transcription, cleavage, regulation, tRNA, rRNAs, Small molecules, metabolites, regulation, activation, membrane capacitance, etc.

**BUILD Modules (Method)**

```

INIT to X0
EVAL to F(t, X)
UPDATA to X+1, t+1
    
```

\*SIMILAR TO NUMERICAL INT. (Euler, RK4, etc.)

- Build individual models
- Characterize & test
- Assume independence at small times

RNA pol (Markov process) choosing data from data

location? rich gene probability

Assuming no regulation for now

$$\frac{dmRNA}{dt} = k_{base} - k_{deg} \cdot mRNA$$

@ St. St.

$$mRNA_{SS} = \frac{k_{base}}{k_{deg}}$$

prob(gene) =  $\frac{k_{base}}{\sum k_{base}}$

**A Whole-Cell Mathematical Model**

4 states:

- Free - not transcribing or bound
- NS - non-specifically bound
- SP - specifically bound but not transcribing

Chromosomal position

Tells us gene relative loc. \* relevant even at 0.01

including in vivo rates of protein-DNA association ing the necessary model parameters. Subsequently, alternative

## A Whole-Cell Computational Model Predicts Phenotype from Genotype

Jonathan R. Karr,<sup>1,4</sup> Jayodita C. Sanghvi,<sup>2,4</sup> Derek N. Macklin,<sup>2</sup> Miriam V. Gutschow,<sup>2</sup> Jared M. Jacobs,<sup>2</sup> Benjamin Bolival, Jr.,<sup>2</sup> Nacyra Assad-Garcia,<sup>3</sup> John I. Glass,<sup>3</sup> and Markus W. Covert<sup>2,\*</sup>

<sup>1</sup>Graduate Program in Biophysics  
<sup>2</sup>Department of Bioengineering  
 Stanford University, Stanford, CA 94305, USA  
<sup>3</sup>J. Craig Venter Institute, Rockville, MD 20850, USA  
<sup>4</sup>These authors contributed equally to this work  
 \*Correspondence: mcovert@stanford.edu  
<http://dx.doi.org/10.1016/j.cell.2012.05.044>

**SUMMARY**

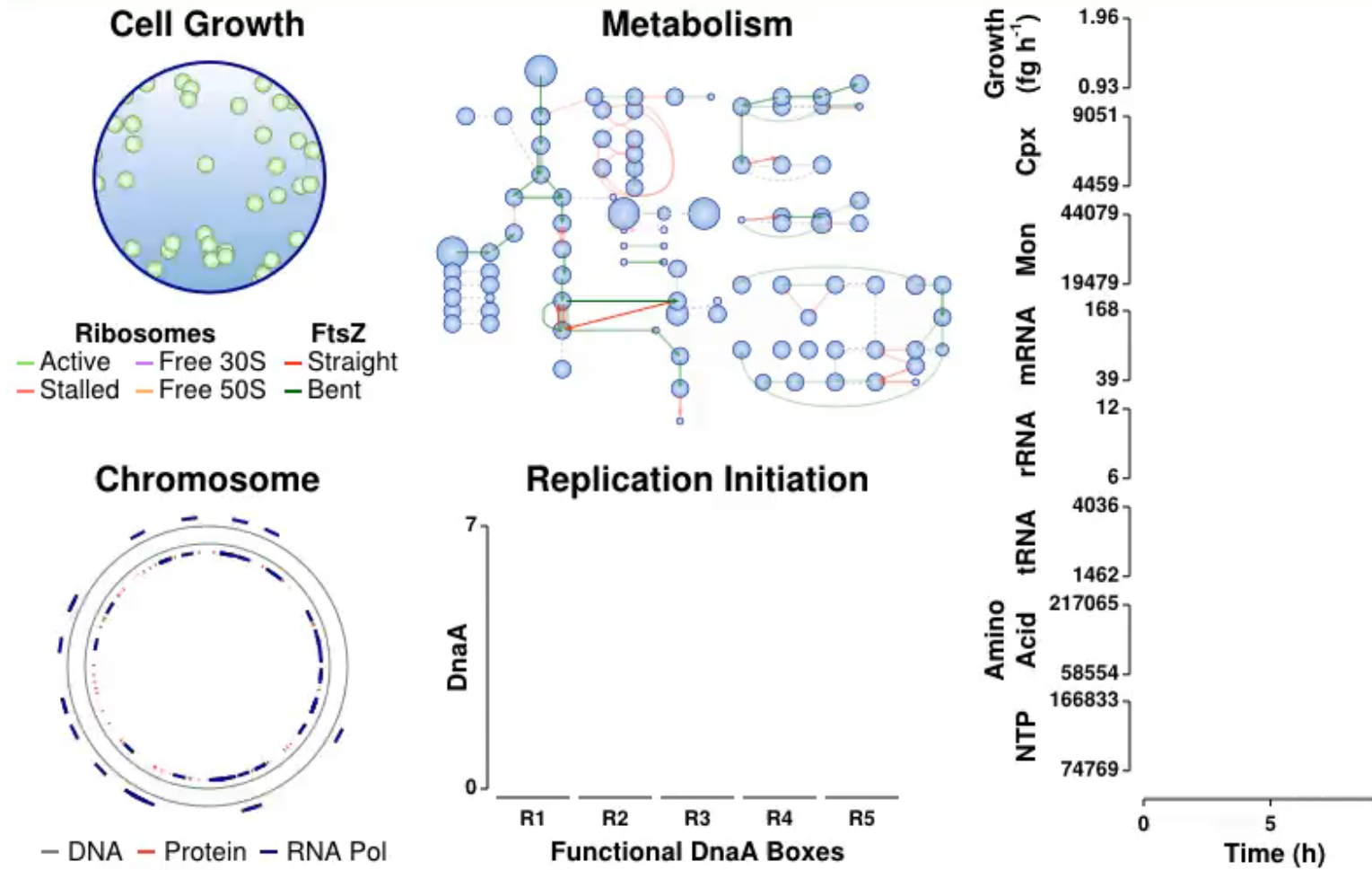
Understanding how complex phenotypes arise from individual molecules and their interactions is a primary challenge in biology that computational approaches are poised to tackle. We report a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. An integrative approach to modeling that combines diverse mathematics enabled the simultaneous inclusion of fundamentally different cellular processes and experimental measurements. Our whole-cell model accounts for all annotated gene functions and was validated against a broad range of data. The model provides insights into many previously unexplained cellular behaviors, including the necessity for a minimum number of genes for growth, the effect of gene deletions on growth rate, and the impact of gene duplications on growth rate. The model also predicts the phenotypic consequences of gene deletions and duplications, including the necessary model parameters. Subsequently, alternative

First, until recently, not enough has been known about the individual molecules and their interactions to completely model any one organism. The advent of genomics and other high-throughput measurement techniques has accelerated the characterization of some organisms to the extent that comprehensive modeling is now possible. For example, the mycoplasmas, a genus of bacteria with relatively small genomes that includes several pathogens, have recently been the subject of an exhaustive experimental effort by a European consortium to determine the transcriptome (Güell et al., 2009), proteome (Kühner et al., 2009), and metabolome (Yus et al., 2009) of these organisms. The second limiting factor has been that no single computational method is sufficient to explain complex phenotypes in terms of molecular components and their interactions. The first approaches to modeling cellular physiology, based on ordinary differential equations (ODEs) (Atlas et al., 2008; Browning et al., 2004; Castellanos et al., 2004, 2007; Domach et al., 1984; Tomita et al., 2003), were limited by the difficulty in obtaining the necessary model parameters. Subsequently, alternative

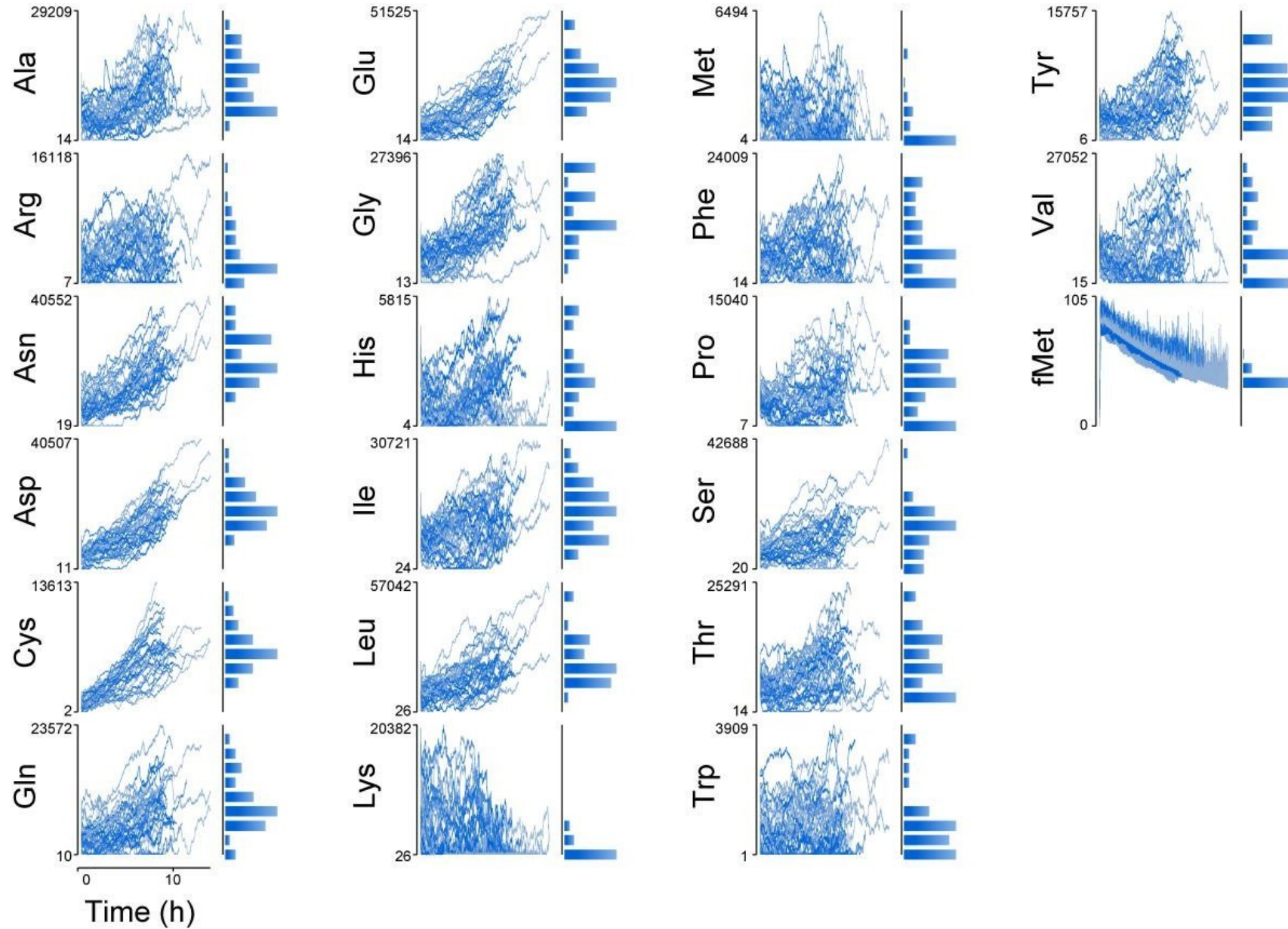
# Software

- ODE models
  - COPASI: [copasi.org](http://copasi.org)
  - V-Cell: [nrcam.uchc.edu](http://nrcam.uchc.edu)
  - Systems biology toolbox
- Boolean models
  - CellNOpt
- Flux-balance analysis
  - openCOBRA: [opencobra.sourceforge.net](http://opencobra.sourceforge.net)
  - RAVEN
- Integrative models
  - E-Cell: [e-cell.org](http://e-cell.org)
  - Whole-cell: [wholecell.org](http://wholecell.org)
- Standards
  - SBML: [sbml.org](http://sbml.org)
  - CellML: [cellml.org](http://cellml.org)

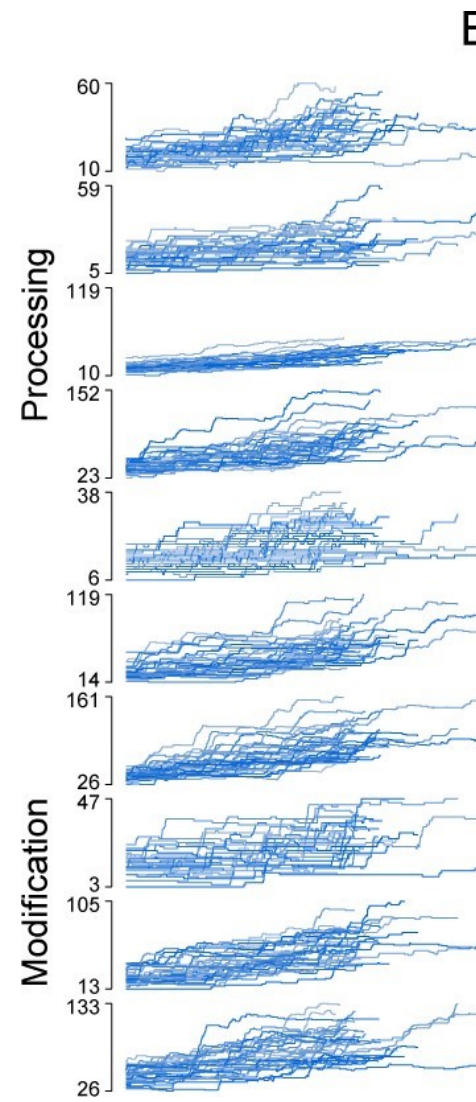
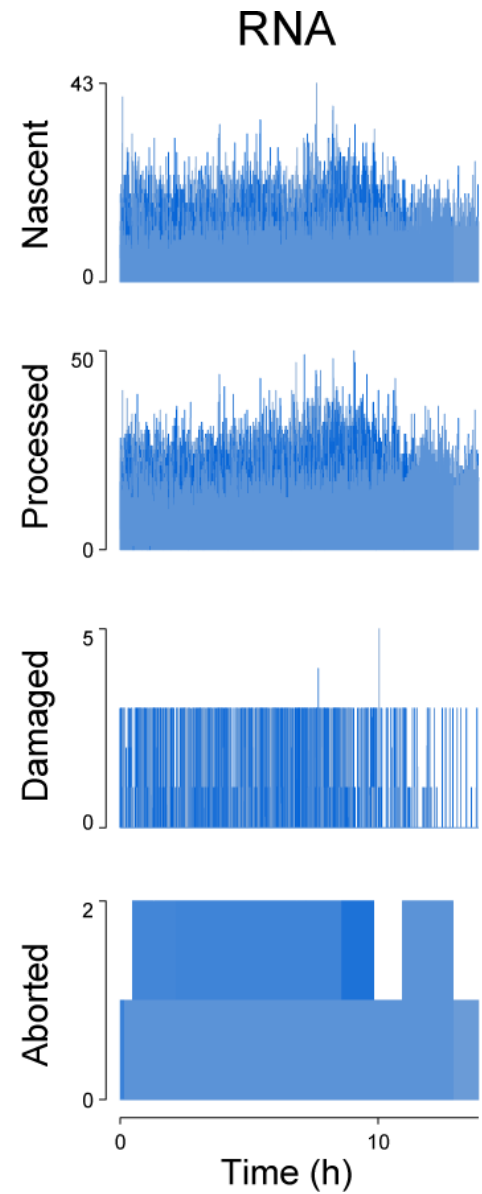
# A Whole-Cell Model Predicts Phenotype from Genotype



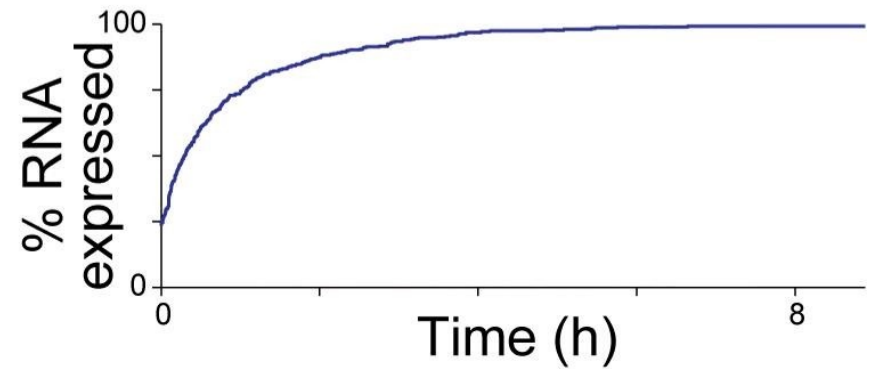
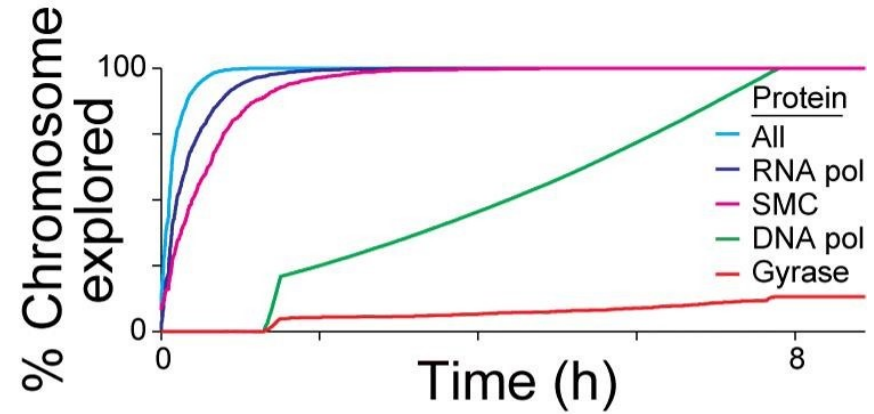
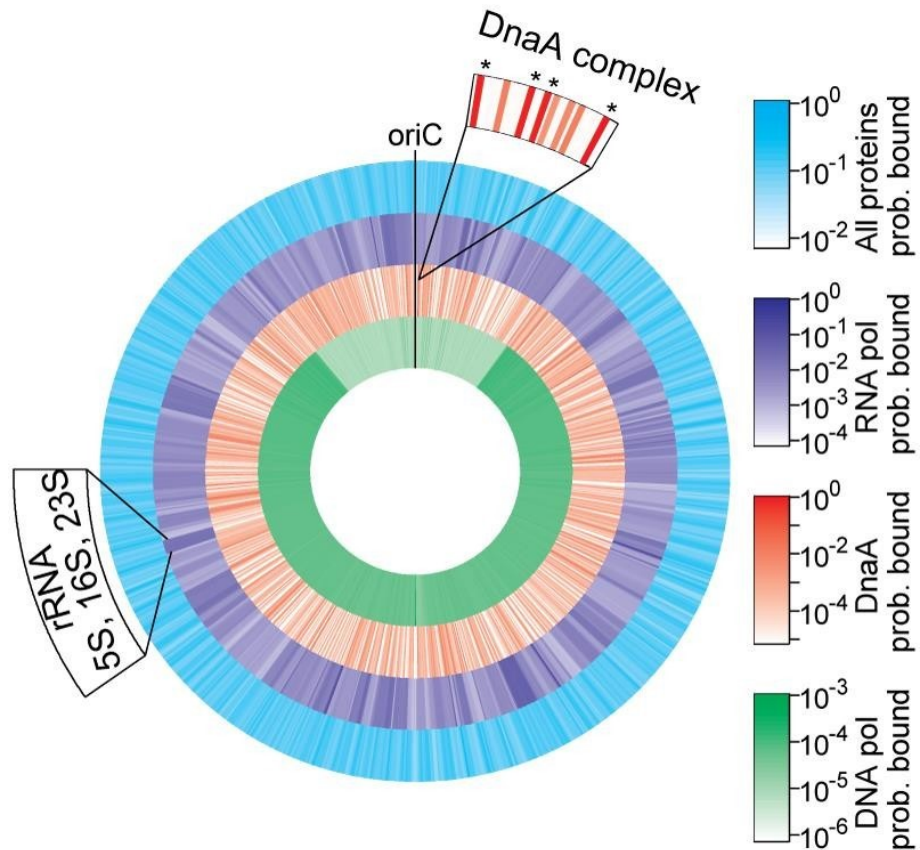
# Metabolite concentrations



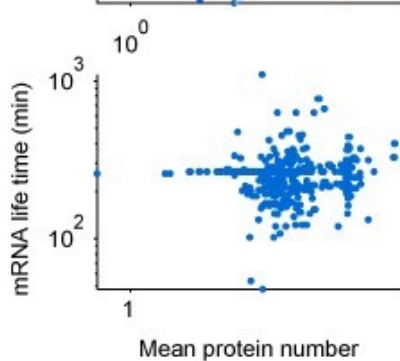
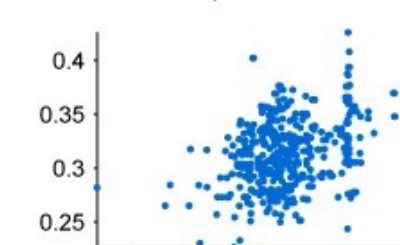
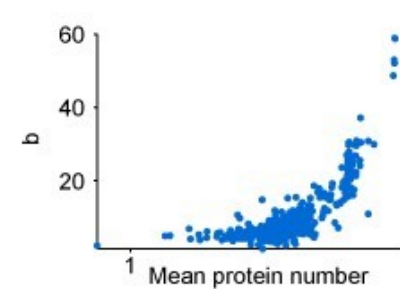
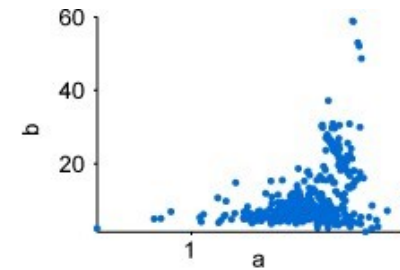
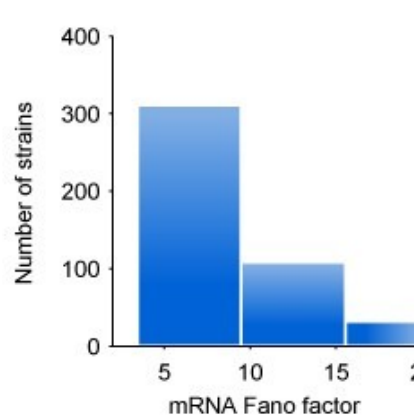
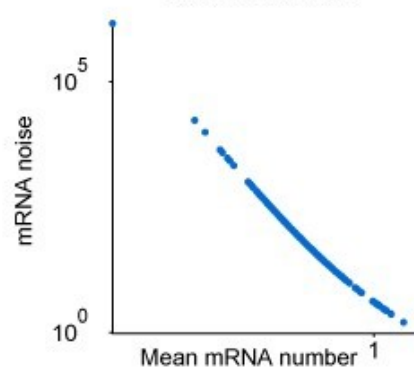
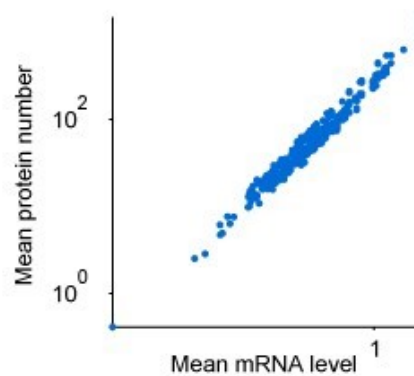
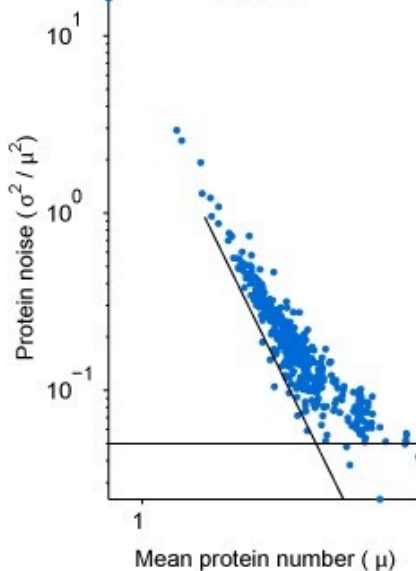
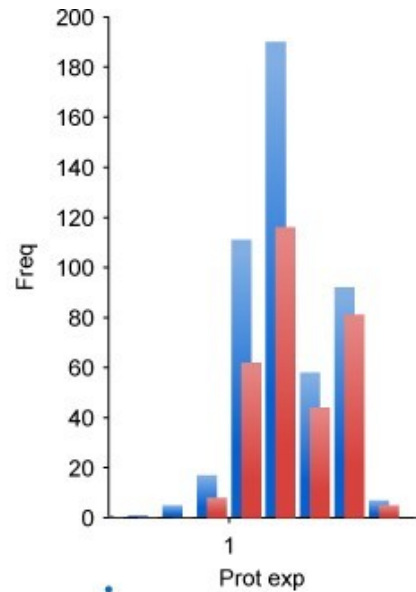
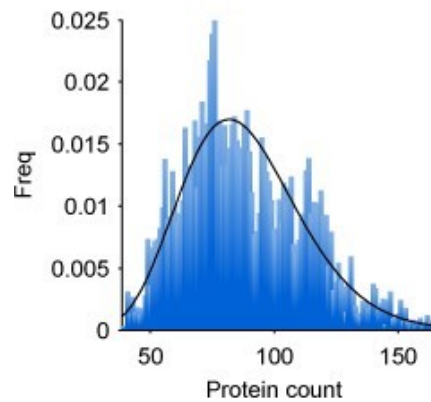
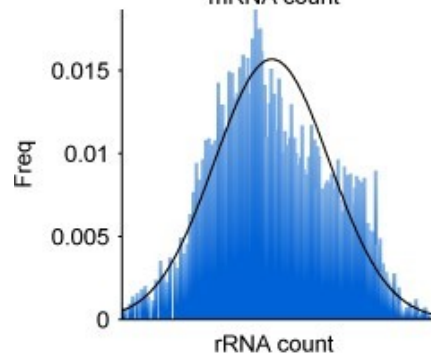
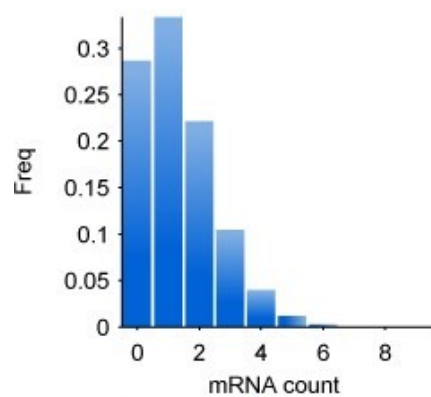
# RNA synthesis rates



# DNA binding

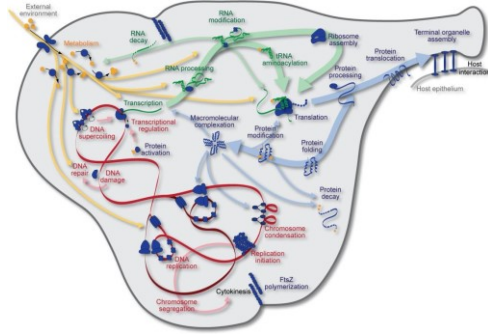


# Translation

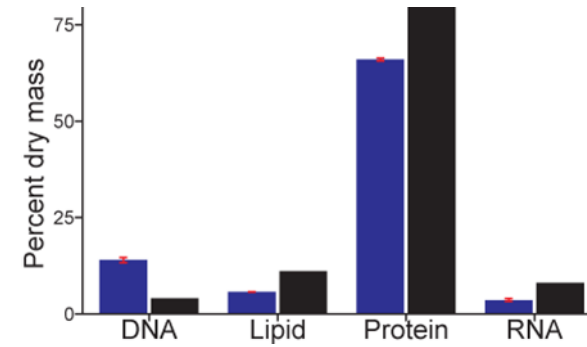


# Whole-cell modeling

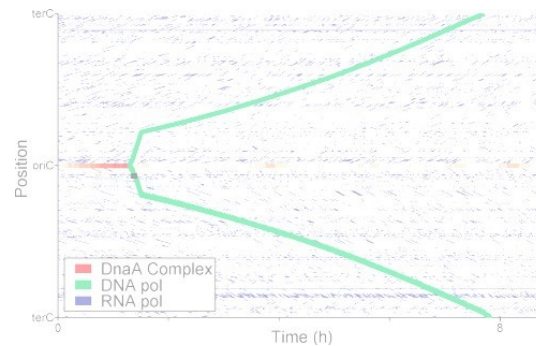
Model



Validate



Engineer





# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

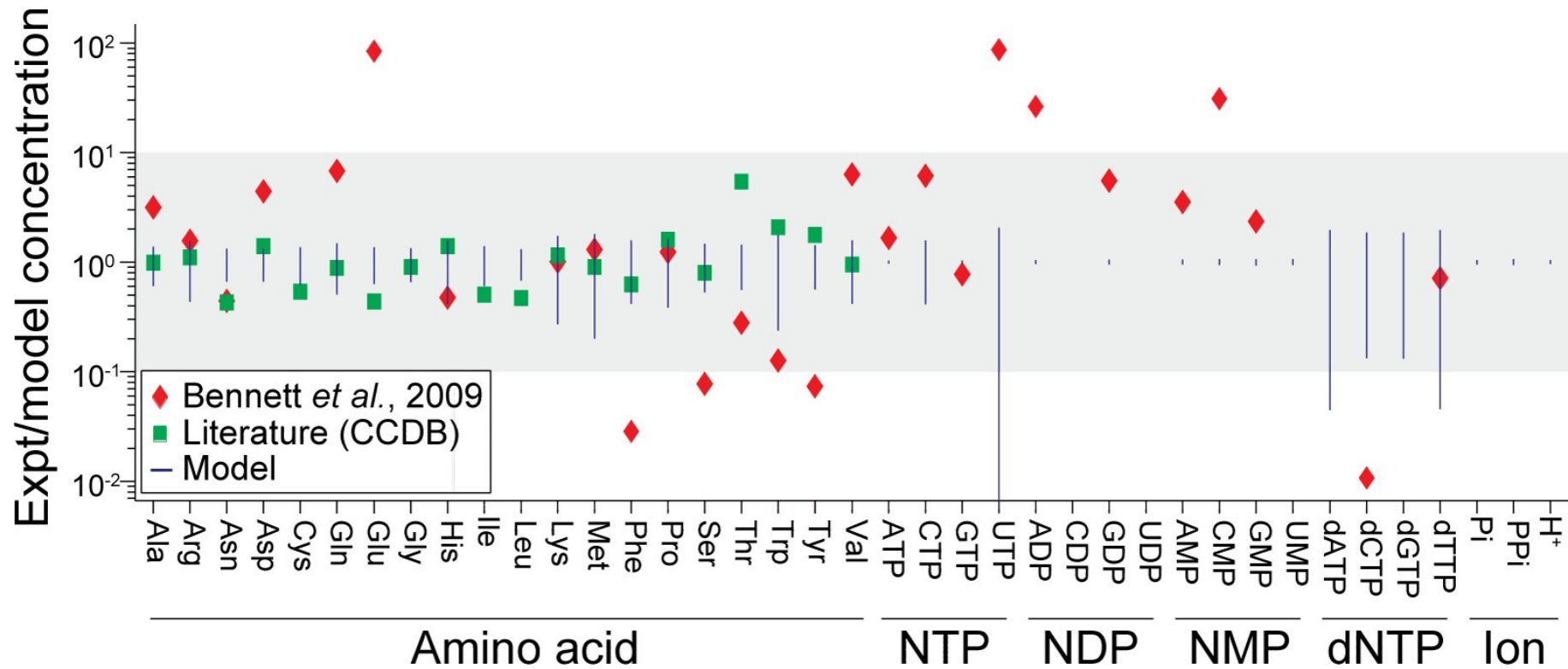
## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Model reproduces observed metabolomics



# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing



# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing



# Validate model against experiments and theory

## Matches training data

- Cell mass, volume
- Biomass composition
- RNA, protein expression, half-lives
- Superhelicity

## Matches published data

- Metabolite concentrations
- DNA-bound protein density
- Gene essentiality

## Matches new data

- Wild-type growth rate
- Disruption strain growth rates

## Matches theory

- Mass conservation
- Central dogma
- Cell theory
- Evolution

## No obvious errors

- Plot model predictions
- Manually inspect data
- Compare to known biology

## Software stable

- Simulation code is stable
- Tests passing

# Desafios

- Como podemos modelar fisiologia mais complexa?
- Como podemos modelar organismos mais complexos?
- Larger bacteria
  - Eukaryotes
  - Multicellularity
  - Humans
- Como podemos usar modelos para direcionar a engenharia?

# Leituras recomendadas

- Karr JR et al. (2012) A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*, **150**, 389-401.
- Macklin DN, Ruggero NA, Covert MW (2014) The future of whole-cell modeling. *Curr Opin Biotechnol*, **28C**, 111-115.
- Shuler ML, Foley P, Atlas J (2012). Modeling a minimal cell. *Methods Mol Biol*, **881**, 573-610.
- Joyce AR, Palsson BØ (2007). Toward whole cell modeling and simulation: comprehensive functional genomics through the constraint-based approach. *Prog Drug Res* **64**, 267-309.
- Tomita M (2001). Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol* **6**, 205-10.
- Surovtsev IV et al. (2009) Mathematical modeling of a minimal protocell with coordinated growth and division. *J Theor Biol*, **260**, 422-9.

# Leituras recomendadas: FBA

- Thiele I et al. (2009). Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput Biol.* **5**, e1000312.
- Orth JD, Thiele I, Palsson BØ (2010). What is flux balance analysis? *Nat Biotechnol*, **28**, 245-8.
- Covert MW et al (2008). Integrated Flux Balance Analysis Model of *Escherichia coli*. *Bioinformatics* **24**, 2044–50.
- Covert MW et al (2004). Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, **429**, 92-6.