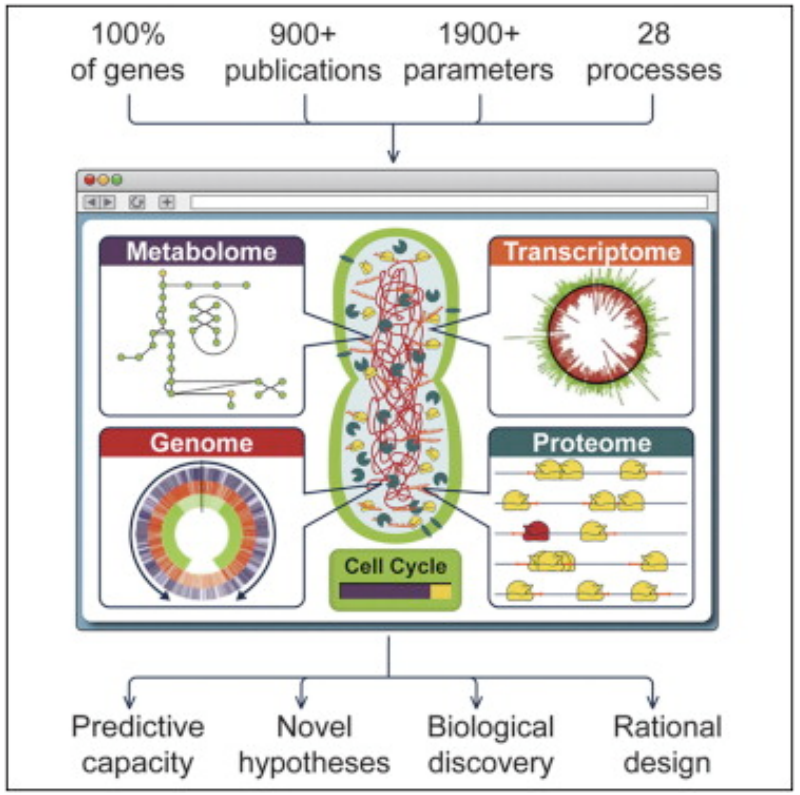
A whole-cell model for the life cycle of the human pathogen *Mycoplasma genitalium*

Theory

Cell

A Whole-Cell Computational Model Predicts Phenotype from Genotype

Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[2] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]
[1]Graduate Program in Biophysics
[2]Department of Bioengineering
Stanford University, Stanford, CA 94305, USA
[3]J. Craig Venter Institute, Rockville, MD 20850, USA
[4]These authors contributed equally to this work
*Correspondence: mcovert@stanford.edu
http://dx.doi.org/10.1016/j.cell.2012.05.044



Cell 150, 389-401 (2012)

Main aims

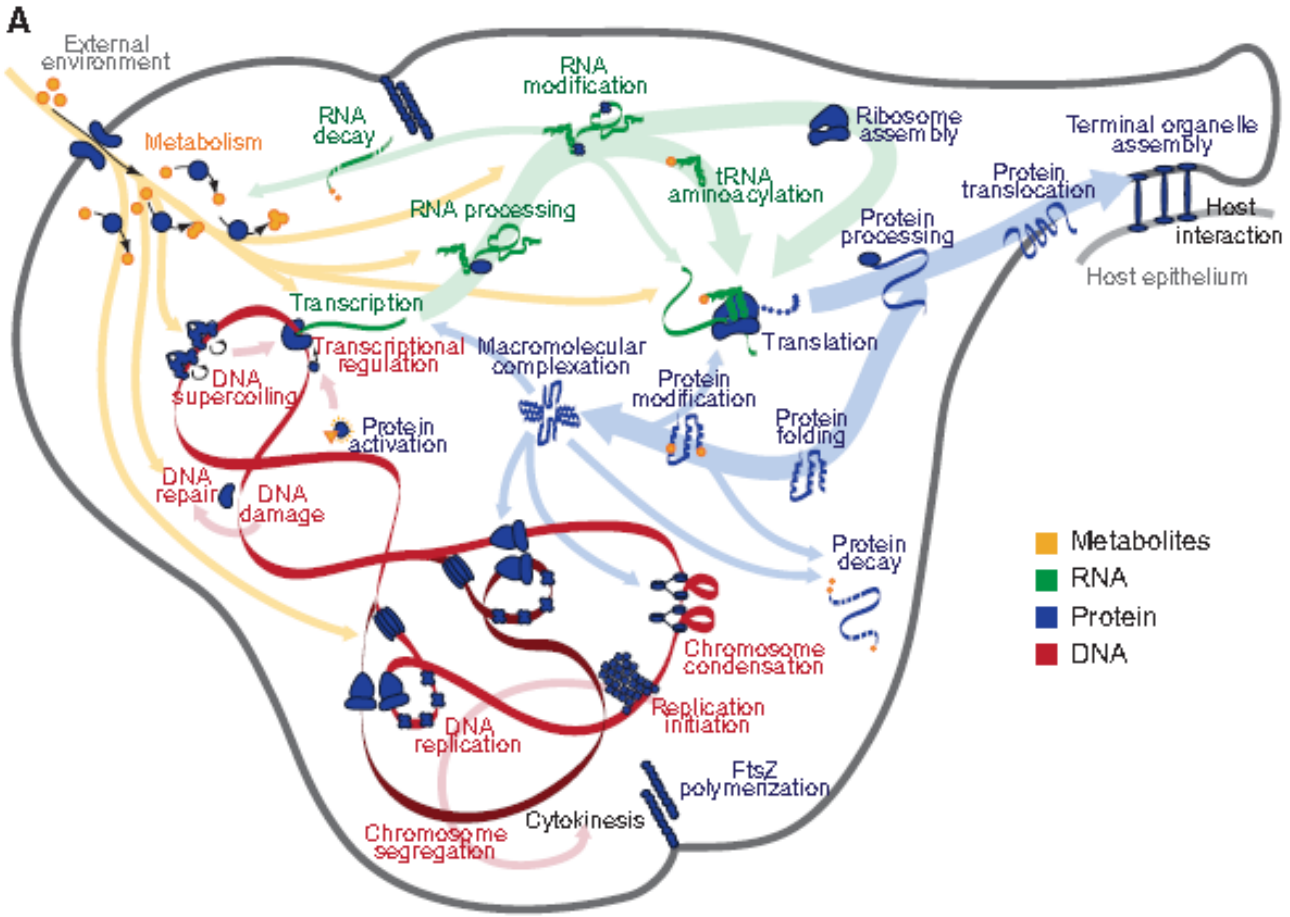
(1) Describe the life cycle of a single cell from the level of individual molecules and their interactions

(2) Account for the specific function of every annotated gene product

(3) accurately predict a wide range of observable cellular behaviors

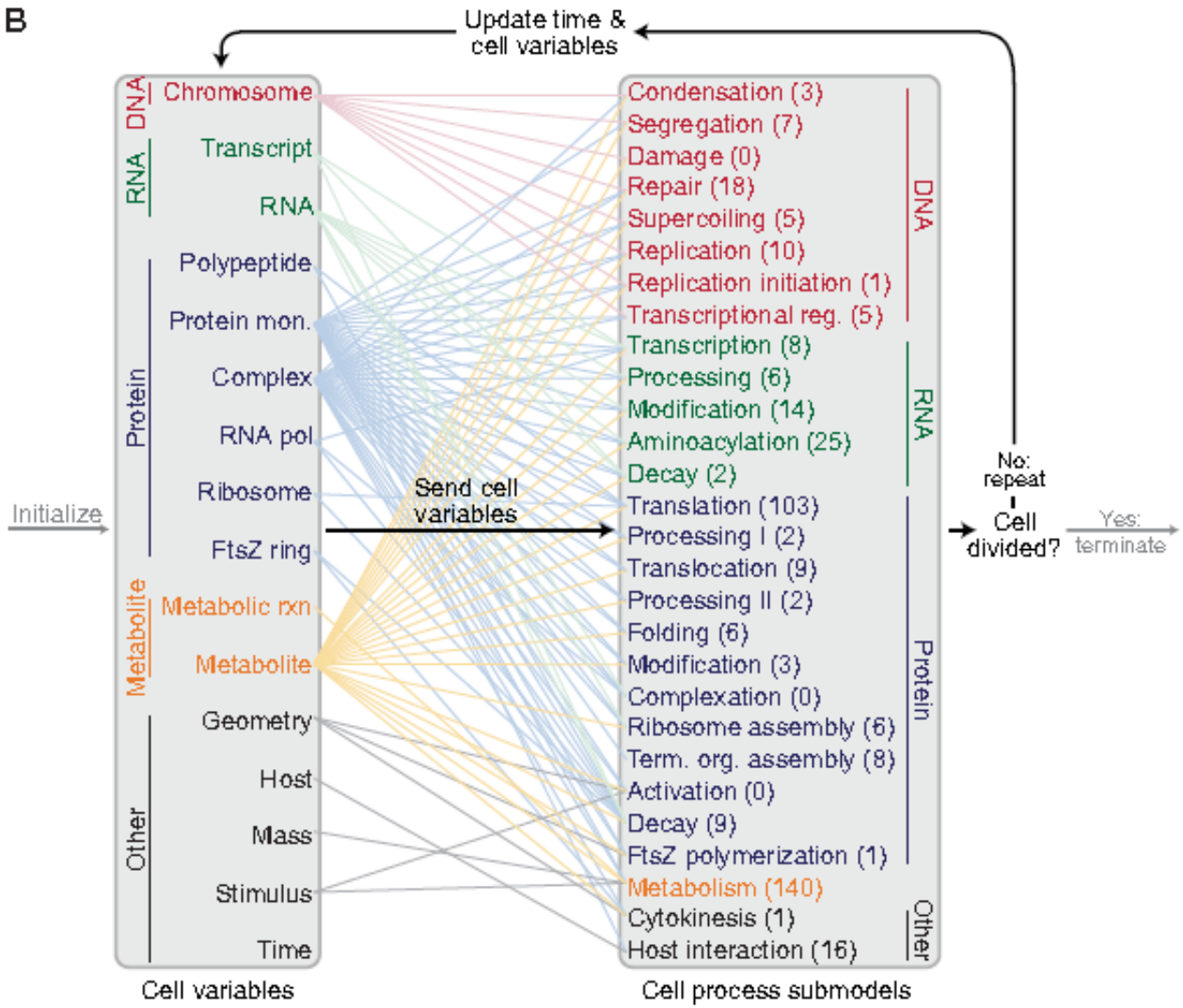
Problem: In a cell, everything seems connected to everything.

Divide and conquer approach (Caesar): split whole-cell model into 28 independent submodels



28 submodels are built / parametrized / iterated independently

Cell variables

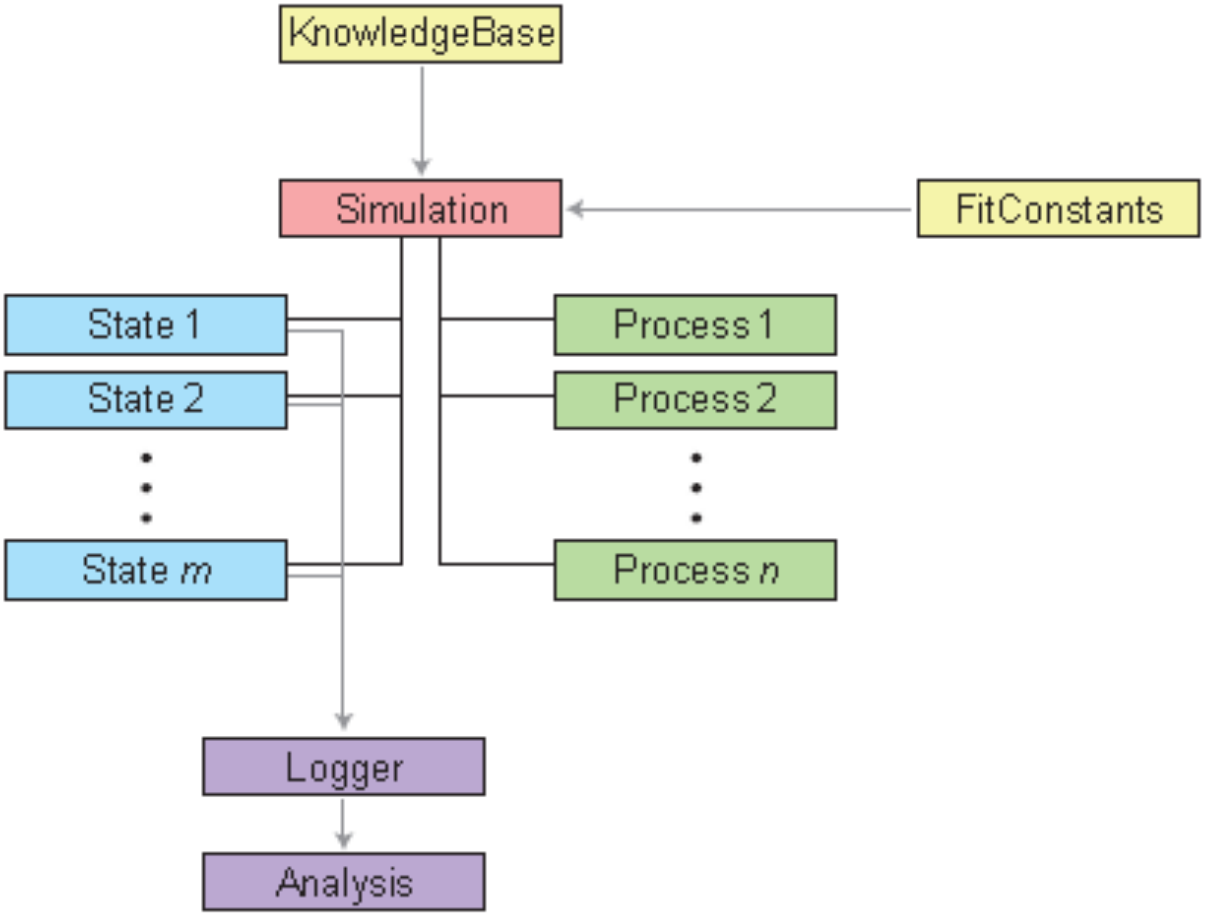


System state is described by 16 cell variables

Colored lines: cell variables affected by individual submodels

Mathematical tools:
-Differential equations
-Stochastic simulations
-Flux balance analysis

Whole-cell model architecture

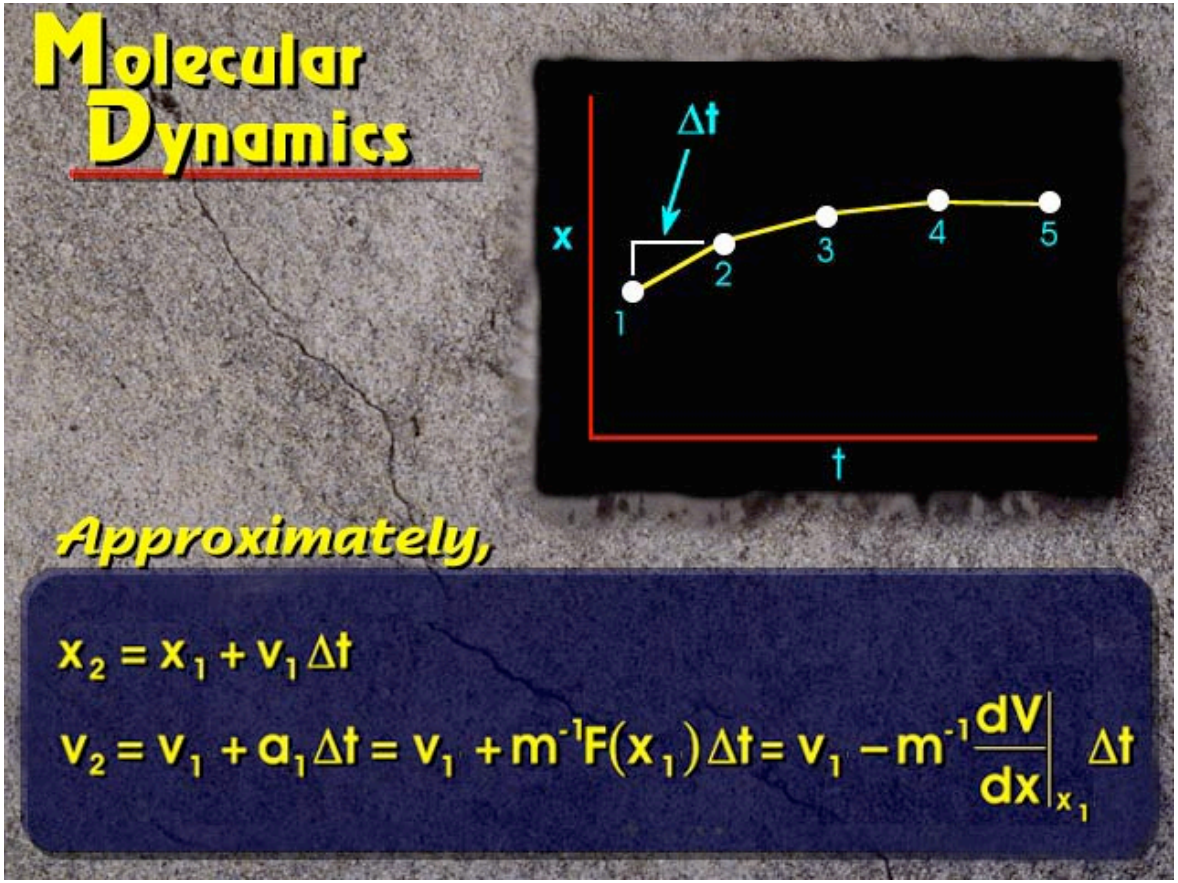


Schematic S17. Whole-cell model architecture.

Show movie.

Q: Can treatment as independent processes lead to problems?

Time step in molecular atomistic dynamics simulations



Too long time step leads to numerical instabilities

-> atoms move „too far“ along the direction of attraction

Similar problems may arise in whole-cell simulation.
-particles of one type accumulate

Whole-cell dynamic simulation algorithm

Algorithm S1 | Whole-cell dynamic simulation algorithm.

Construct whole-cell simulation objects using the `KnowledgeBase` classes
Computationally align processes and fit parameters
Identify initial conditions variance control parameters using Algorithm S2
Initialize cell state using Algorithm S3 and the fit values of the cell state variance control parameters
repeat
  Increment the time by **1** s
  Set the external conditions based on Table S3F and Table S3H
  Allocate shared resources:
  foreach *metabolite* i *in compartment* j **do**
    foreach *process* k **do** Calculate the demand, d_{ijk}, of process k for metabolite i in compartment j
    Divide the total count, m_{ij}, of metabolite i in compartment j into temporary dedicated pools, m_{ijk}, for each process proportional to demand, $m_{ijk} \leftarrow m_{ij} \frac{d_{ijk}}{\sum_k d_{ijk}}$
  Compute temporal evolution:
  foreach *process* i **do**
    Retrieve the current values of cell state variables and the counts of metabolites allocated to process i
    Compute the contribution of process i to the temporal evolution of the cell state
    Update the values of the cell state variables
until *cell divided or time* $> 1.5\times$ *average mass doubling time*

Algo to identify initial conditions

Algorithm S2 | Initial conditions identification algorithm.

Initialize the initial cell state variance control parameters: $\sigma_m \leftarrow 0$, $\eta_r \leftarrow 0$, $\eta_p \leftarrow 0$

repeat

- Simulate the life cycle of a population of wild type cells using Algorithm S3 to initialize the value of each cell state variable
- Randomly segregate the cellular content into two daughter cells
- Calculate the variances of the total cell mass, RNA copy number, and protein copy number states
- Set the values of the initial distribution control parameters of each state equal to that of the final distribution
- $\sigma_m \leftarrow$ standard deviation of the final cell mass distribution
- $\eta_r \leftarrow \sigma_r^2/N_r$, where N_r and σ_r^2 are the mean and variance of the final RNA copy number distribution
- $\eta_p \leftarrow \sigma_p^2/N_p$, where N_p and σ_p^2 are the mean and variance of the final RNA copy number distribution

until *the initial variance control parameters (σ_m, η_r, and η_p) converge*

Initialize cell state

Algorithm S3 | Cell state initialization procedure.

Input: $\sigma_m \leftarrow$ standard deviation of the initial total cell mass distribution
Input: $\eta_r, \eta_p \leftarrow$ RNA and protein copy number distribution initial variance control parameters
Input: $f_r, f_p \leftarrow$ reconstructed fractional cell RNA and protein composition
Input: $e_r, e_p \leftarrow$ expected relative expression of each RNA and protein species
Input: $w_r, w_p \leftarrow$ molecular weight of each RNA and protein species
Input: $N_i\,(f_i, e_i, w_i) \leftarrow f_i m / (e'_i * w_i)$ total initial RNA (i =r) and protein (i =p) copy number functions
Set time $\leftarrow$ 0 s
Set values of external stimuli and metabolites according to Table S3F and S3H
Set total cell mass, $m \leftarrow\sim N(\mu, \sigma_m)$, and calculate cell volume and shape
Set the metabolite counts according to the total cell mass and reconstructed cell composition (see Table S3I)
Initialize the `Chromosome` state with one methylated chromosome; decrement dNMP counts to maintain cell mass
Set mature RNA copy numbers according to $\texttt{multinomialRand}(\eta_r N_r, e_r)/\eta_r$; decrement NMP counts
Set mature protein monomer copy numbers according to $\texttt{multinomialRand}(\eta_p N_p, e_p)/\eta_p$; decrement amino acids
Form macromolecules by calculating the steady-state of the `Macromolecular Complexation` process
Set the `RNA Polymerase` and `Transcript` states to a steady-state of the `Transcription` sub-model
Set the `Ribosome` and `Polypeptide` states to a steady-state of the `Translation` sub-model
Set the FtsZ state to a steady-state of the `FtsZ Polymerization` sub-model with no septal rings
Set the growth rate and metabolic reaction fluxes to a steady-state of the `Metabolism` sub-model
Set the `Host` state to a steady-state of the `Host Interaction` sub-model
Set the chromosome protein occupancy to a steady-state the chromosome-interacting sub-models

List S1. Primary sources of the *M. genitalium* reconstruction.

Data source	Content
Bernstein *et al.*, 2002[602]	mRNA half-lives
BioCyc[6]	Genome annotation, metabolic reactions
BRENDA[570]	Reaction kinetics
CMR[168]	Genome annotation
Deuerling *et al.*, 2003[388]	Chaperone substrates
DrugBank[847]	Antibiotics
Eisen *et al.*, 1999[891]	DNA repair
Endo *et al.*, 2007[391]	Chaperone substrates
Feist *et al.*, 2007[558]	Metabolic reactions
Glass *et al.*, 2006[193]	Gene essentiality
Güell *et al.*, 2009[418]	Transcription unit structure
Gupta *et al.*, 2007[280]	N-terminal methionine cleavage
KEGG[113]	Genome annotation, orthology
Kerner *et al.*, 2005[389]	Chaperone substrates
Krause *et al.*, 2004[409]	Terminal organelle assembly
Lindahl *et al.*, 2000[462]	DNA damage
Morowitz *et al.*, 1962[870]	Cell chemical composition
NCBI Gene[61,777]	Genome annotation
Neidhardt *et al.*, 1990[393]	Cell chemical composition
Peil, 2009[105]	RNA modification
PubChem[587]	Metabolite structures
SABIO-RK[100]	Reaction kinetics
Solabia[754–759]	Media chemical composition
Suthers *et al.*, 2009[610]	Metabolic reactions
UniProt[96]	Genome annotation
Weiner *et al.*, 2000[411]	Promoters
Weiner *et al.*, 2003[569]	mRNA expression

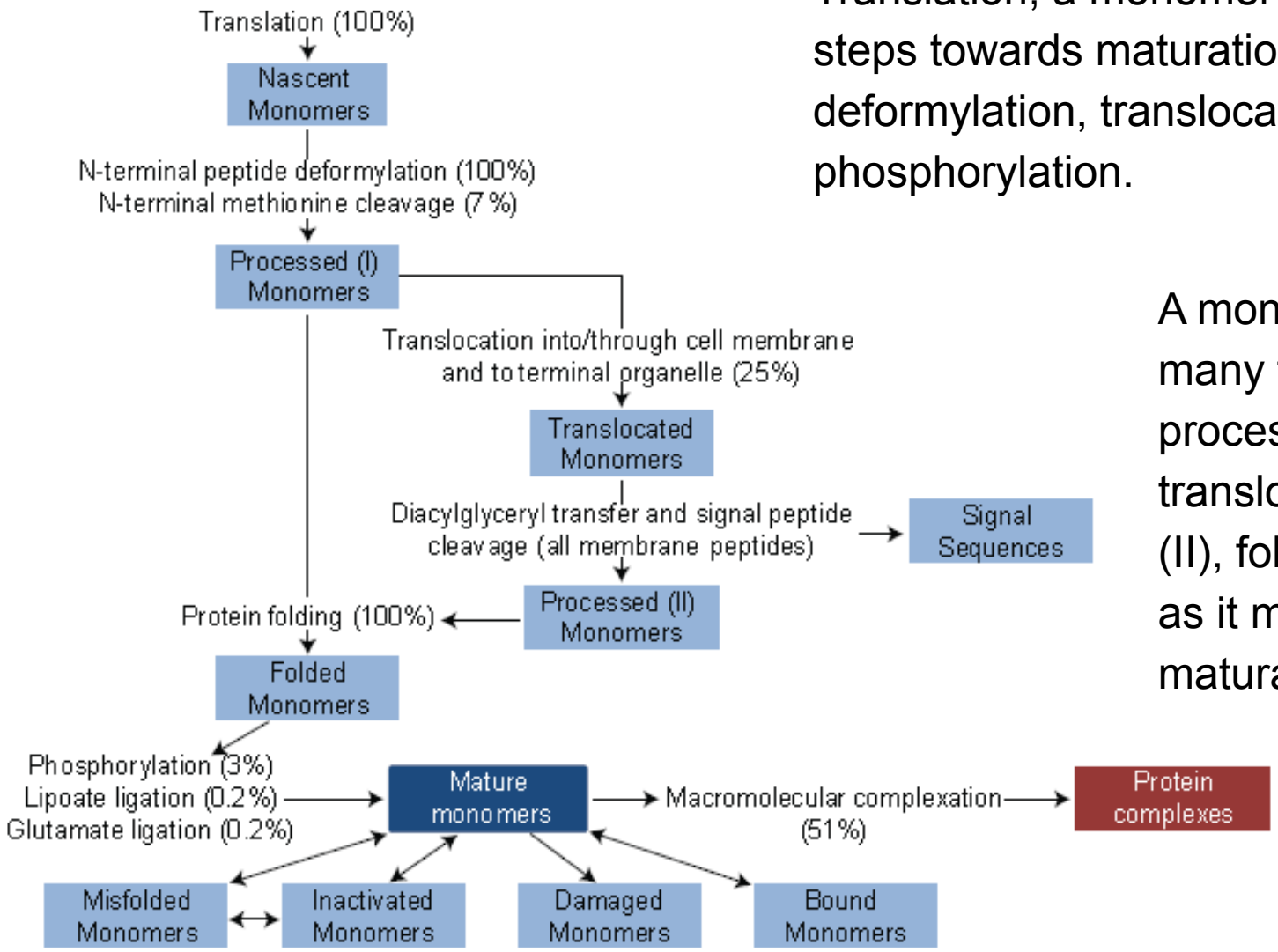
Nucleotide states

This state represents the polymerization, winding, modification, and protein occupancy of each nucleotide of each strand of each copy of the *M. genitalium* chromosome, and the (de)catenation status of the two sister chromosomes following replication.

List S2. Mathematical representation of nucleotide $i = \{1..L\}$ of strand $j = \{1..2\}$ of chromosome copy $k = \{1..2\}$.

Physical Property	Symbol	Size	Type
Polymerization	p_{ijk}	$L \times 2 \times 2$	Boolean
Winding	w_{ijk}	$L \times 2 \times 2$	Real
Modification			
Gap site	m^g_{ijk}	$L \times 2 \times 2$	Boolean
Abasic site	m^a_{ijk}	$L \times 2 \times 2$	Boolean
Sugar-phosphate	m^p_{ijkl}	$L \times 2 \times 2 \times M$	Boolean
Base	m^b_{ijkl}	$L \times 2 \times 2 \times M$	Boolean
Intrastrand cross link	m^c_{ijk}	$L \times 2 \times 2$	Boolean
Strand break	m^s_{ijk}	$L \times 2 \times 2$	Boolean
Holliday junction	m^h_{ijk}	$L \times 2 \times 2$	Boolean
Protein occupancy			
Monomer	b^m_{ijkl}	$L \times 2 \times 2 \times B^m$	Boolean
Complex	b^c_{ijkl}	$L \times 2 \times 2 \times B^c$	Boolean
Catenation	s	1×1	Boolean

Protein Monomers are the direct result of successful translation events. Upon Translation, a monomer undergoes various steps towards maturation including deformylation, translocation, folding, and phosphorylation.



Schematic S3. Protein monomer forms diagrammed in the context of the maturity pipeline.

A monomer can exist in many forms (nascent, processed (I), translocated, processed (II), folded, and mature) as it moves through the maturation pipeline.

Ribosome

Ribosomes are large ribonucleoproteins which synthesize polypeptides.

The *M. genitalium* 70S ribosome is composed of two subunits – the 30S and 50S ribosomal subunits – which assemble on mRNA with assistance from initiation factors 1-3 (MG173, MG142, MG196).

The 30S subunit is composed of 1 RNA and 20 protein monomer subunits.

The 50S subunit is composed of 2 RNA and 32 protein monomer subunits.

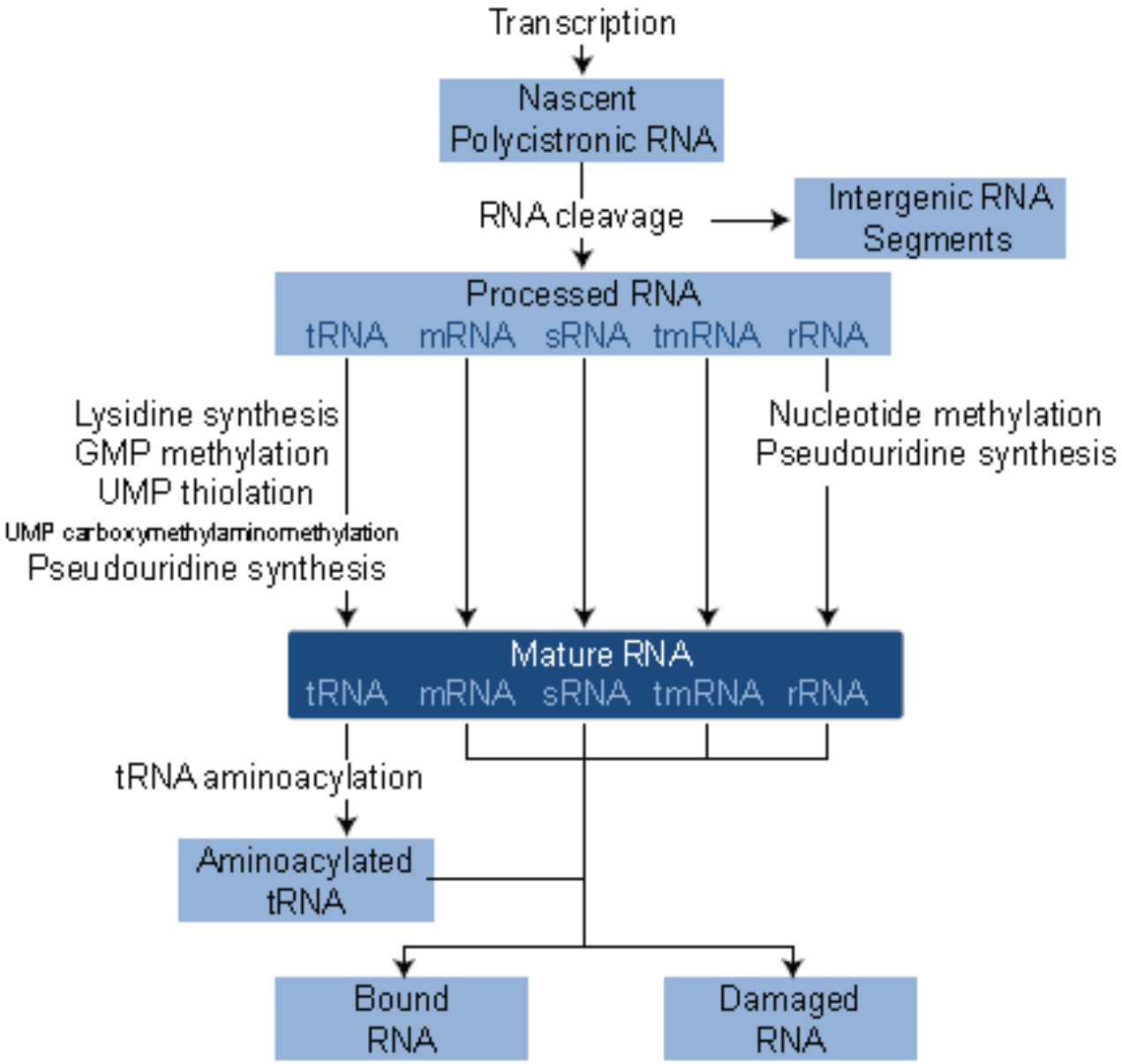
The 30S and 50S ribosomal subunits are believed to assemble in stereotyped patterns, and six GTPases – EngA, EngB, Era, Obg, RbfA, and RbgA – have been associated with ribosomal subunit assembly.

The exact functions of the six GTPases are unknown.

Ribosome

Algorithm S4 | `Ribosome` **and** `Polypeptide` **state initialization.**

Free 70S ribosomes ← `min`(*free 30S, 50S ribosomal particles*)
Decrement the copy numbers of free 30S and 50S ribosomal particles
foreach *70S ribosome* i **do**
    Select the mRNA of 70S ribosome i weighted by the product of mRNA copy number and length
    Set the bound mRNA species of 70S ribosome i
    Select the position of 70S ribosome i along the bound mRNA with uniform probability
    Set the status of 70S ribosome i to actively translating
    Decrement the copy number of free 70S ribosomes
    Increment the copy number of bound 70S ribosomes
    Set the sequence of the nascent polypeptide corresponding to 70S ribosome i



Schematic S4. RNA forms diagrammed in the context of RNA maturation.

Basic outline of the direct method of Gillespie

(Step i) generate a list of the components/species and define the initial distribution at time $t = 0$.

(Step ii) generate a list of possible events E_i (chemical reactions as well as physical processes).

(Step iii) using the current component/species distribution, prepare a probability table $P(E_i)$ of all the events that can take place.
Compute the total probability

$$P_{tot} = \sum P(E_i)$$

$P(E_i)$: probability of event E_i .

(Step iv) Pick two random numbers r_1 and $r_2 \in [0...1]$ to decide which event E_μ will occur next and the amount of time τ by which E_μ occurs later since the most recent event.

Resat et al., J.Phys.Chem. B 105, 11026 (2001)

Basic outline of the direct method of Gillespie

Using the random number r_1 and the probability table,
the event E_μ is determined by finding the event that satisfies the relation

$$\sum_{i=1}^{\mu-1} P(E_i) < r_1 P_{tot} \leq \sum_{i=1}^{\mu} P(E_i)$$

The second random number r_2 is used to obtain the amount of time τ between the reactions

$$\tau = -\frac{1}{P_{tot}} \ln(r_2)$$

As the total probability of the events changes in time, the time step between occurring steps varies.

Steps (iii) and (iv) are repeated at each step of the simulation.

The necessary number of runs depends on the inherent noise
of the system and on the desired statistical accuracy.

Resat et al., J.Phys.Chem. B 105, 11026 (2001)

Transcriptional regulation

Algorithm S26 | Transcriptional regulation simulation.

Input: n_{ik} is true if promoter i is expressed in chromosome k
Input: $p_{a,i}$ free cytosolic copy number of transcriptional regulator i
Input: $p_{b,i}$ DNA-bound copy number of transcriptional regulator i
Input: x_{ij} Binding site of transcriptional regulator i at promoter j
Input: F_{ij} fold-change effect of transcriptional regulator i on promoter j
Input: b^m_{ijkl}, b^c_{ijkl} chromosomal protein occupancy as defined in List S2
Output: f_i fold-change effect of transcriptional regulation on RNA polymerase affinity for promoter i

Calculate the relative rate, r_{ijk}, transcriptional regulator i binds promoter j of chromosome k:
foreach *DNA-binding transcriptional regulator* i *in promoter* j *of chromosome* $k = \{1..2\}$ **do**
  $r_{ijk} \leftarrow n_{jk} p_{a,i} F_{ij}$

Bind transcriptional regulators to the chromosome:
repeat
  Select regulator i, promoter j, and chromosome $k \sim \texttt{multinomialRand}(1, r_{ijk} / \sum_{ijk} r_{ijk})$
  if *regulator* i *expressed* $(p_{a,i} > 0)$ *and* `isRegionAccessible`(*promoter* j *of chromosome* k *to regulator* i) **then**
    Bind protein to chromosome: $b^z_{y \bullet ki} = 1 \; \forall \; y \in \{x_{ij}..x_{ij} + l_i - 1\}$, where $z = m$ for monomers and c for complexes
    Update free and bound copy numbers: $p_{a,i} \leftarrow p_{a,i} - 1$, $p_{b,i} \leftarrow p_{b,i} + 1$
  Update binding rate: $r_{ijk} \leftarrow 0$
until *no additional transcriptional regulator can bind DNA* $(r_{ij} = 0 \; \forall \; i, \; j)$

Calculate the fold-change effect of transcriptional regulators on the affinity of RNA polymerase for each promoter
Initialize fold-change effects: $f_i \leftarrow 1 \; \forall \; i$
foreach *promoter* j *of chromosome* $k = \{1..2\}$ *bound by DNA-binding transcriptional regulator* i **do**
  Add fold-change effects multiplicatively: $f_j \leftarrow f_j F_{ij}$
foreach *promoter* j *regulated by an expressed non-DNA-binding transcriptional regulator* i $(p_{a,i} > 0)$ **do**
  $f_j \leftarrow f_j F_{ij}$

Algorithm S5 | DNA damage simulation. See the Mathematical Model section above and List S2 for definition of the mathematical notation.

Input: m_i copy number of metabolite i
Input: M_{ij} stoichiometry of metabolite i in reaction j
Input: z_i^g, z_i^a, $Z_{\bullet i}^p$, $Z_{\bullet i}^b$, z_i^e, and z_i^s: final base configuration resulting from reaction i

foreach *DNA modification reaction* i **do**
- Calculate base modification rate
- **switch** *trigger of reaction* i **do**
  - **case** *spontaneous*
    - $r_i \leftarrow k_i$
  - **case** *radiation*
    - $j \leftarrow$ index of radiation trigger
    - $r_i \leftarrow k_i s_j$
- **foreach** *base* j *in strand* k *of chromosome* l *susceptible to reaction* i *in a random order* **do**
  - **if** *insufficient metabolic resources to support reaction* i *(*$\exists\ j$ *s.t.* $m_j < -M_{ji}$*)* **then**
    - break
  - **if** $\texttt{poissonRand}(r_i) \geq 1$ **then**
    - Update the configuration of base j of strand k of chromosome k
    - $m_{jkl}^g \leftarrow z_i^g$
    - $m_{jkl}^a \leftarrow z_i^a$
    - $m_{jkl\bullet}^p \leftarrow Z_{\bullet i}^p$
    - $m_{jkl\bullet}^b \leftarrow Z_{\bullet i}^b$
    - $m_{jkl}^e \leftarrow z_i^e$
    - $m_{jkl}^s \leftarrow z_i^s$
    - Update metabolite copy numbers: $m \leftarrow m + M_{\bullet i}$

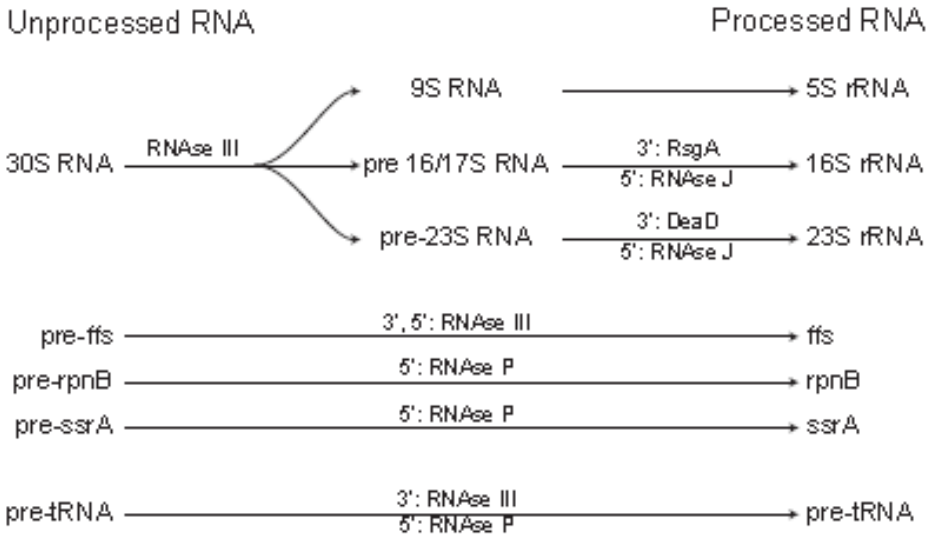
DNA repair simulation

Algorithm S6 | DNA repair simulation.

Repair, methylate, and restrict DNA
foreach *reaction* i *in a random order* **do**
    foreach *base* j *of strand* k *of chromosome* l *susceptible to reaction* i *in a random order* **do**
        if *sufficient enzymatic and metabolic resources to support reaction* i **then**
            Execute reaction i
            Update DNA configuration
            Update metabolite copy numbers
            Decrement enzymatic capacity

Bind DisA to damaged DNA
while *there is free DisA and at least 1 DisA-accessible DNA lesion* **do**
    Let $i, j, k \leftarrow$ represent the base, strand, and chromosome of a DisA-accessible DNA lesion
    Bind DisA to base i of strand j of chromosome k

Non-coding RNA cleavage



Schematic S14. Non-coding RNA cleavage.

Algorithm S24 | RNA processing simulation. See Mathematical Model section above for definition of the mathematical notation.

Input: r_i^p copy number of processed RNA species i
Input: r_i^i copy number of intercistronic fragment i
Input: R_{ji}^g is one if operonic RNA i contains gene j, and zero otherwise
Input: R_{ji}^i is one if operonic RNA i contains intercistronic fragment j, and zero otherwise

Let $k_i \leftarrow e_i \Delta t$ be the capacity of enzyme i for RNA processing
repeat
  Calculate cleavage rates
  foreach *operonic non-coding RNA species* i **do**
    Calculate v_i according to Eq. S29
  Select operonic non-coding RNA species $i \sim$ `multinomialRand(1,` $v_i / \sum_j v_j$`)`
  Update RNA copy numbers: $r_i^u \leftarrow r_i^u - 1$, $r^p \leftarrow r^p + R_{\bullet i}^g$, $r^i \leftarrow r^i + R_{\bullet i}^i$
  Update metabolites: $m \leftarrow m - M_{\bullet i}$
  Update enzyme catalytic capacity: $k \leftarrow k - K_{\bullet i}$
until *no further cleavage possible* ($v_i = 0 \; \forall \; i$)

Model FtsZ polymerization

Cell division in many bacterial species requires the assembly of an FtsZ ring at the cell membrane around the midplane of the cell.

FtsZ is a homologue of eukaryotic tubulin that assembles into long polymers.

These polymers are typically localized to the center of the cell, forming a membrane-bound ring.

FtsZ is a GTPase, and GTP hydrolysis to GDP causes the FtsZ filaments to bend.

This bending serves as one of the forces enabling cell division.

Model FtsZ polymerization

FtsZ can exist in one of multiple states: inactivated monomer, activated monomer (GTP bound), nucleated (dimer of two activated FtsZ molecules), elongated polymer of three or more GTP bound FtsZ molecules.

The following differential equation model is evaluated at each timestep:

$$\frac{dF}{dt} = k_{\mathrm{act2}} F_T - k_{\mathrm{act1}} F$$

$$\frac{dF_D}{dt} = k_{\mathrm{ex2}} F_T \left[\mathrm{GDP}\right] - k_{\mathrm{ex1}} F_D \left[\mathrm{GTP}\right]$$

$$\frac{dF_T}{dt} = k_{\mathrm{act1}} F - k_{\mathrm{act2}} F_T + k_{\mathrm{ex1}} F_D \left[\mathrm{GTP}\right] - k_{\mathrm{ex2}} F_T \left[\mathrm{GDP}\right] - 2k_{\mathrm{nuc1}} F_T^2 + \cdots$$
$$+ \cdots 2k_{\mathrm{nuc2}} F_{T2} - k_{\mathrm{el1}} F_T \left(\sum_{i=2}^{8} F_{Ti}\right) + k_{\mathrm{el2}} \left(\sum_{i=3}^{9} F_{Ti}\right)$$

$$\frac{dF_{T2}}{dt} = k_{\mathrm{nuc1}} F_T^2 - k_{\mathrm{nuc2}} F_{T2} - k_{\mathrm{el1}} F_T F_{T2} + k_{\mathrm{el2}} F_{T3}$$

$$\frac{dF_{Ti}}{dt} = k_{\mathrm{el1}} F_T F_{Ti-1} - k_{\mathrm{el2}} F_{Ti} - k_{\mathrm{el1}} F_T F_{Ti} + k_{\mathrm{el2}} F_{Ti+1}, \text{ for } i \in 3..8$$

$$\frac{dF_{T9}}{dt} = k_{\mathrm{el1}} F_T F_{T8} - k_{\mathrm{el2}} F_{T9}$$

Formation of macromolecular complexes

The relative formation rate, r_i of each complex, i, is described by mass-action kinetics,

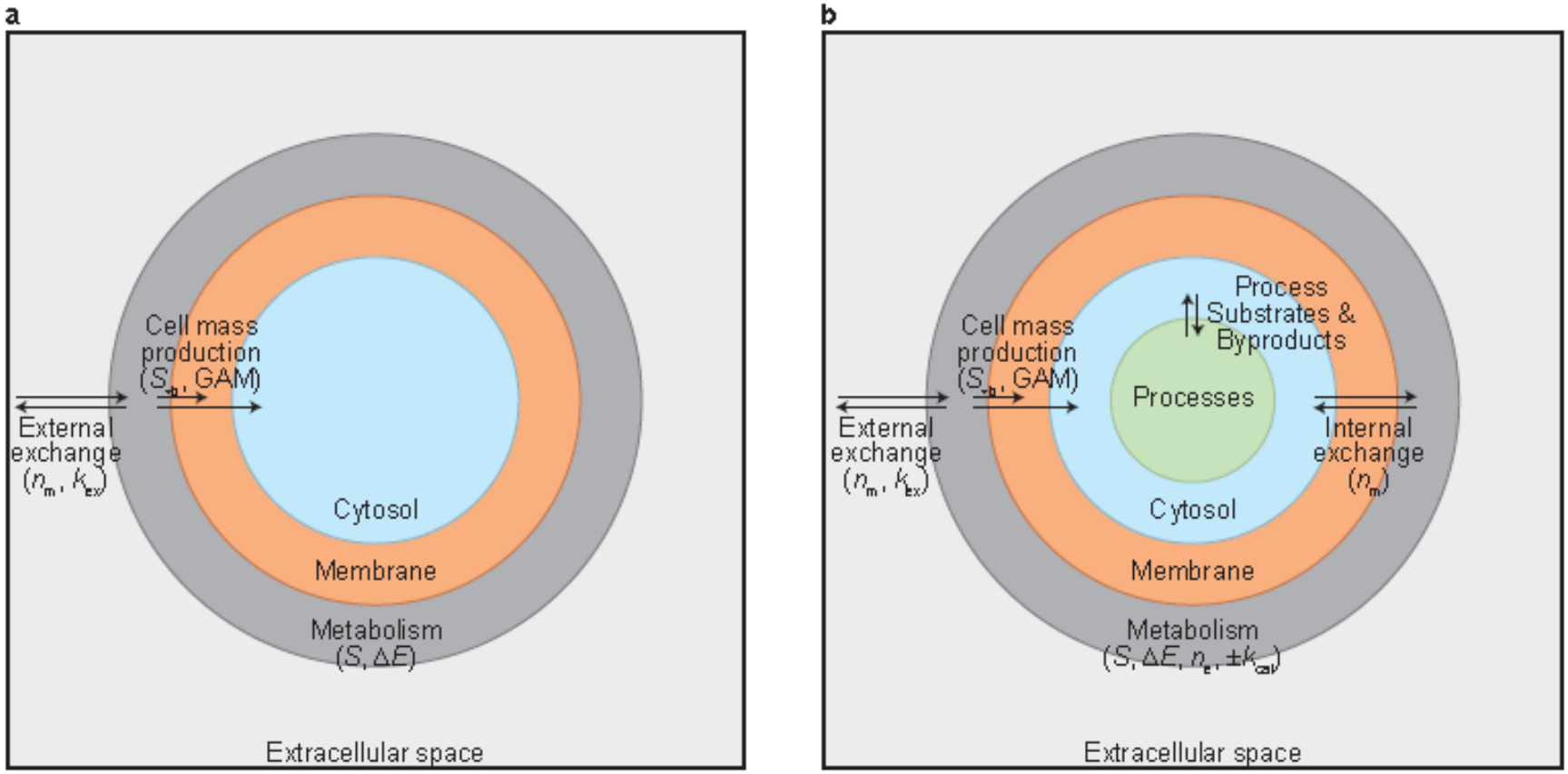
$$r_i = \prod_j \left(\frac{m_j}{V}\right)^{S_{ij}}$$

m_j : copy number of gene product j,
V is the cell volume,
s_{ij} : stoichiometry of subunit j in complex i.

Algorithm S8 | Macromolecular complexation simulation. See Mathematical Model above for mathematical notation.

repeat
  foreach *protein complex* i **do**
    Calculate relative formation rate, $r_i \leftarrow \prod_j \left(\frac{m_j}{V}\right)^{S_{ij}}$
  Select a complex k to form according to $\texttt{multinomialRand}(1, r_i / \sum_i r_i)$
  Increment copy number of complex k, $c_k \leftarrow c_k + 1$
  Decrement copy numbers of complex k subunits, $m_j \leftarrow m_j - S_{k,j}$
until *Insufficient subunits to form additional complexes*

Model small-molecule metabolism by FBA



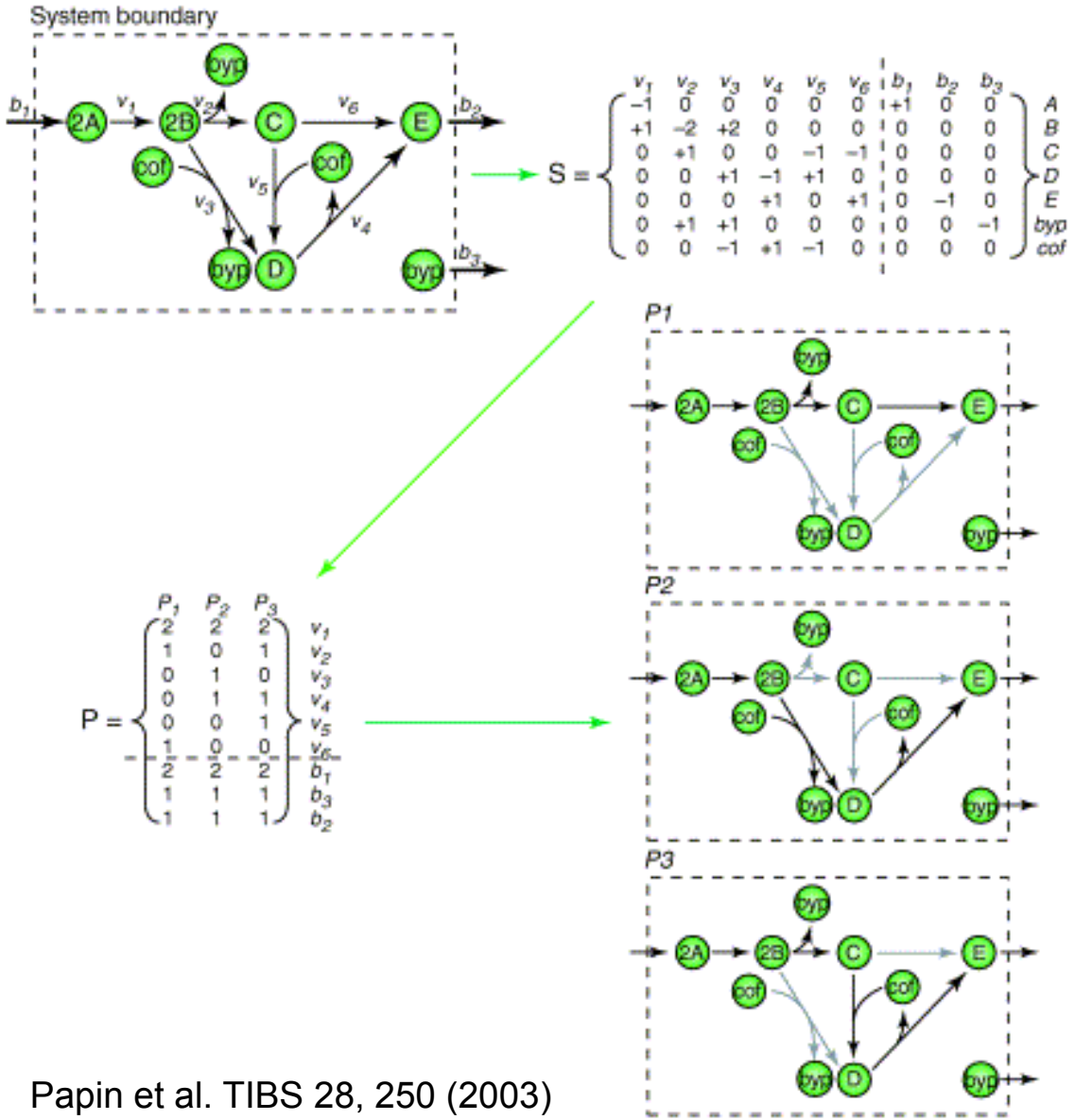
Schematic S10. Metabolite perspective of the flux-balance analysis (FBA) metabolic model. a, Conventional FBA metabolic model. b, Integrated FBA metabolic model.

V15 Flux Balance Analysis – Extreme Pathways

Stoichiometric matrix S:
m × *n* matrix with stochiometries of the *n* reactions as columns and participations of *m* metabolites as rows.

The stochiometric matrix is an important part of the *in silico* model.

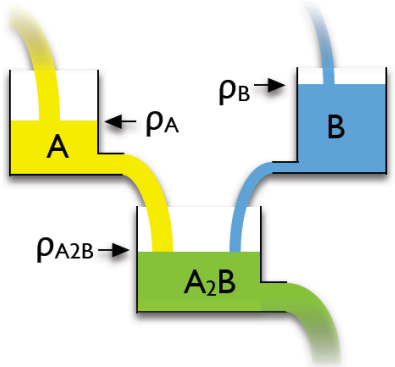
With the matrix, the methods of extreme pathway and elementary mode analyses can be used to generate a unique set of pathways P1, P2, and P3 that allow to express all steady-state fluxes as linear combinations of P1 – P3.



Papin et al. TIBS 28, 250 (2003)

Flux balancing

Any chemical reaction requires **mass conservation**. Therefore one may analyze metabolic systems by requiring mass conservation. Only required: knowledge about stoichiometry of metabolic pathways.



Steady state: concentrations are constant
=> flux in = flux out

$$\frac{dA_2B(t)}{dt} = G_{A_2B} - L_{A_2B} = 0$$

For each metabolite X_i :

$$dX_i/dt = V_{synthesized} - V_{used} + V_{transported\_in} - V_{transported\_out}$$

Under **steady-state conditions**, the mass balance constraints in a metabolic network can be represented mathematically by the matrix equation:

$$\mathbf{S} \cdot v = 0$$

where the matrix **S** is the stoichiometric matrix and the vector *v* represents all fluxes in the metabolic network, including the internal fluxes, transport fluxes and the growth flux.

Flux balance analysis

Since the number of metabolites is generally smaller than the number of reactions ($m < n$) the flux-balance equation is typically **underdetermined**.

Therefore there are generally multiple feasible flux distributions that satisfy the mass balance constraints.

The set of solutions are confined to the **nullspace** of matrix **S**.

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$$

Consider

$$\begin{pmatrix} 0 & 2 & 1 \\ 3 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Corresponds to

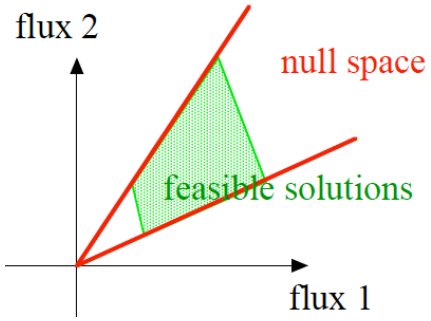
$$\begin{aligned} 2x_2 + x_3 &= 0 \\ 3x_1 - x_2 + x_3 &= 0 \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} 2x_2 &= -x_3 \\ 2x_1 &= -x_3 \end{aligned}$$

=> one free parameter: x_3

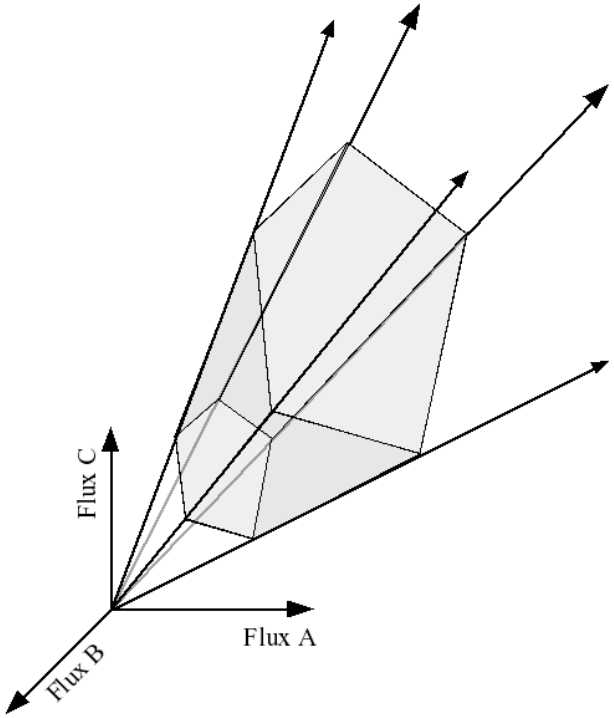
null space: $\vec{x} = \begin{pmatrix} -a \\ -a \\ 2a \end{pmatrix}$

Add inequalities for external fluxes
(here, e.g.: $x_3 \geq 0$)
=> **feasible** solutions for $a \geq 0$

Generally: null space is a cone,
constraints select part of it



Feasible solution set for a metabolic reaction network



The steady-state operation of the metabolic network is restricted to the region within a **pointed cone**, defined as the feasible set.

The feasible set contains all flux vectors that satisfy the physicochemical constrains.

Thus, the feasible set defines the capabilities of the metabolic network. All feasible metabolic flux distributions lie within the feasible set.

Edwards & Palsson PNAS 97, 5528 (2000)

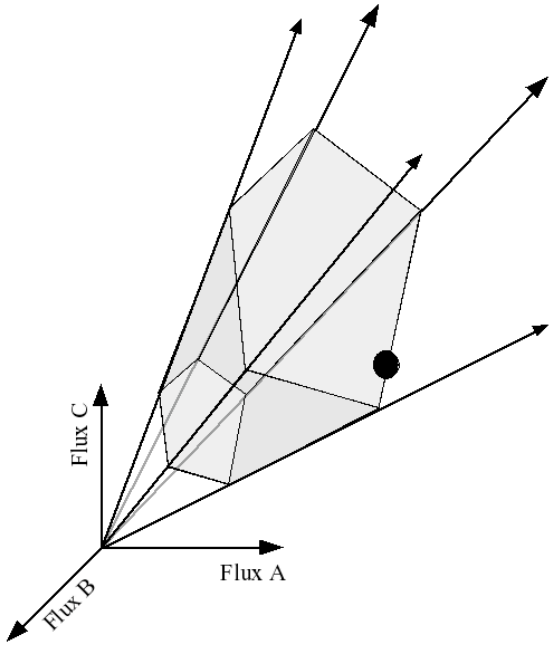
True biological flux

To find the „true“ biological flux in cells (→ e.g. Heinzle, UdS) one needs additional (experimental) information,
or one may impose constraints

$$\alpha_i \leq v_i \leq \beta_i$$

on the magnitude of each individual metabolic flux.

The intersection of the nullspace and the region defined by those linear inequalities defines a region in flux space = the **feasible set of fluxes**.



In the limiting case, where all constraints on the metabolic network are known, such as the enzyme kinetics and gene regulation, the feasible set may be reduced to a single point. This single point must lie within the feasible set.

Metabolism FBA simulation

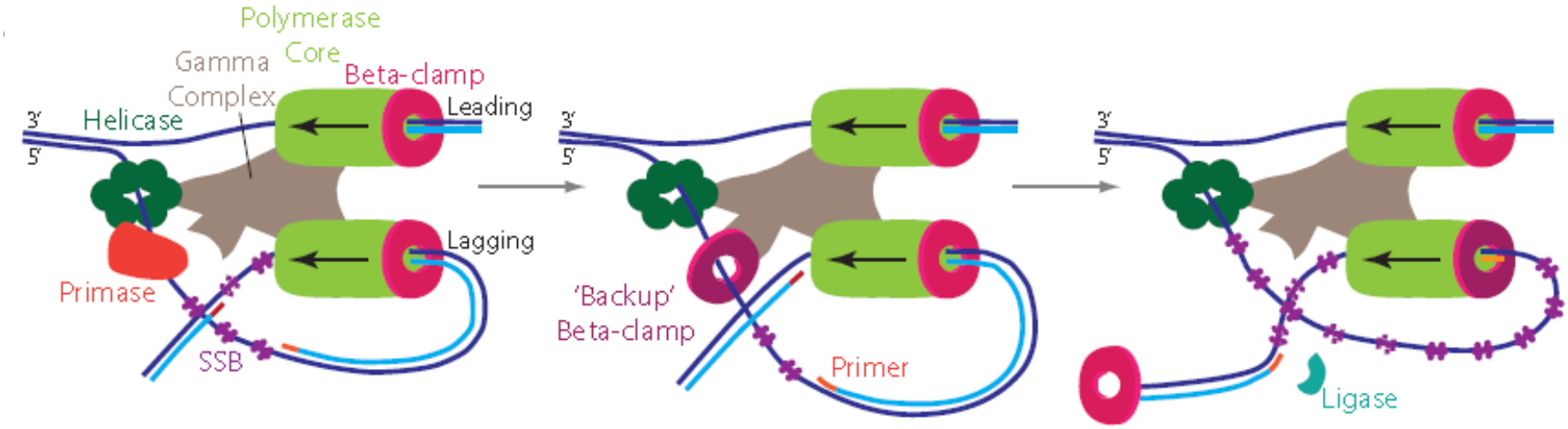
Algorithm S9 | Metabolism FBA simulation. See Mathematical Model section above for definition of the mathematical notation.

Calculate flux bounds

begin

- Initialize bounds: $v_{l,i} \leftarrow -\inf$, $v_{u,i} \leftarrow +\inf$
- **foreach** *thermodynamically irreversible reaction* i **do**
  - Constrain reverse flux to zero: $v_{l,i} \leftarrow 0$
- **foreach** *chemically catalyzed reaction* i **do**
  - $j \leftarrow$ index of enzyme which catalyzes reaction i
  - **if** $k_{-,i}$ *is known* **then**
    - bound flux by enzyme kinetics and expression: $v_{l,i} \leftarrow \max\left(v_{l_i}, k_{-cat,i} n_{e,j}\right)$
  - **else** bound flux by enzyme expression: $v_{l,i} \leftarrow v_{l,i} \cdot (n_{e,j} > 0)$
  - **if** $k_{+,i}$ *is known* **then**
    - bound flux by enzyme kinetics and expression: $v_{u,i} \leftarrow \min\left(v_{u_i}, k_{+cat,i} n_{e,j}\right)$
  - **else** bound flux by enzyme expression: $v_{u,i} \leftarrow v_{u,i} \cdot (n_{e,j} > 0)$
- **foreach** *chemical reaction* i **do**
  - **foreach** *protein substrate* j *of reaction* i **do**
    - **if** *protein substrate* j *is not expressed* **then** constrain flux to zero: $v_{l,i} \leftarrow 0$, $v_{u,i} \leftarrow 0$
- **foreach** *internal exchange reaction* i **do**
  - $j \leftarrow$ index of metabolite exchanged by reaction i
  - Bound internal metabolite exchange by copy number: $v_{l,i} \leftarrow \max\left(v_{l_i}, -n_{m,j}\right)$
- **foreach** *external exchange reaction* i **do**
  - $j \leftarrow$ index of metabolite exchanged by reaction i
  - Bound external metabolite exchange by copy number and maximum exchange rate: $v_{l,i} \leftarrow \max\left(v_{l_i}, mk_{-ex,i}\right)$, $v_{u,i} \leftarrow \min\left(v_{u_i}, mk_{+ex,i}, n_{m,j}\right)$

DNA replication



Schematic S11. Schematic of DNA replication.

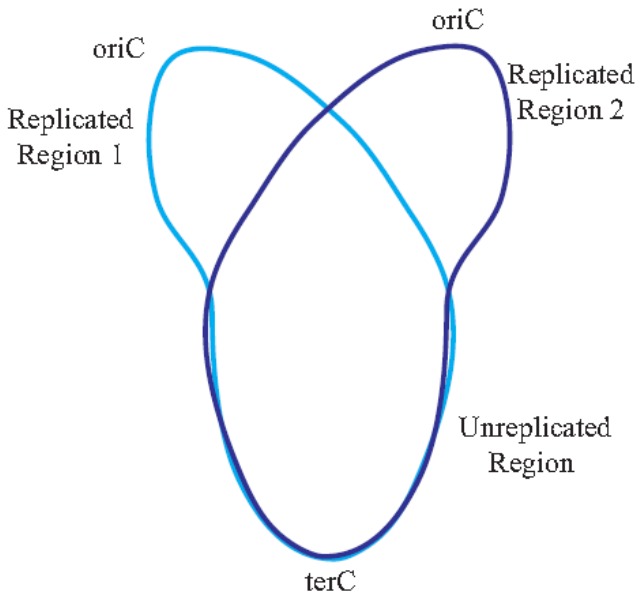
List S15. Enzymes and complexes used in the `Replication` process class.

Enzymes/Complexes	Composition	Gene Name(s)	DNA Footprints (nt)
DNA helicase	(6) MG094	dnaB	20
DNA primase	(1) MG250	dnaG	14
β-clamp	(2) MG001	dnaN	25
DNA polymerase core	(1) MG031, (1) MG261	polC, polC-2	24
γ-complex	(1) MG007, (1) MG351, (4) MG419	holB, holA, dnaX	26
DNA ligase	(1) MG254	ligA	19
Single stranded binding protein (SSB) 8mer	(8) MG091	ssb	145

DNA replication

DNA naturally exists at a certain level of helicity, and this level of helical density is important for the DNA's stability, its ability to fit in the cell, and its ability to bind proteins.

M. genitalium has 3 topoisomerase proteins: DNA gyrase, topoisomerase I, and topoisomerase IV.
These proteins transiently break a DNA strand to wind (topoisomerase I) or unwind (topoisomerase IV, gyrase) the DNA.



Schematic S7. Regions of varying superhelical density on the replicating chromosome.

Which states are affected by replication?

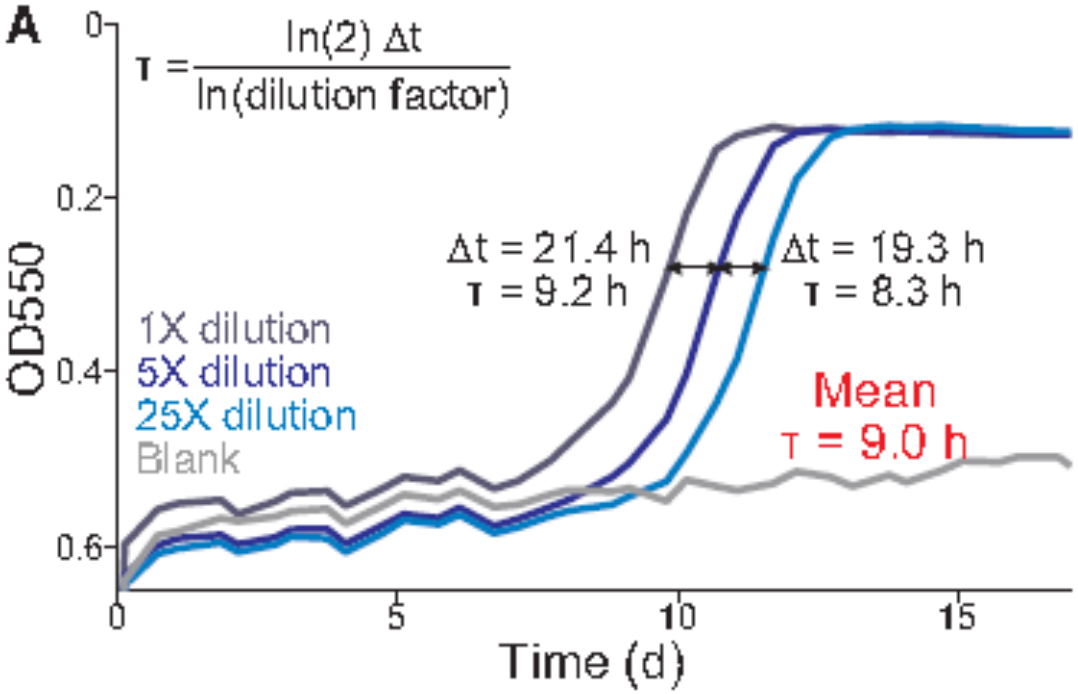
Upon replication initiation (the binding of 29 DnaA-ATP molecules near the oriC by the Replication Initiation process class), the Replication process class tracks the progression of the replication proteins on the known chromosome sequences.

List S17. State classes connected to the `Replication` process class.

Connected States	Read from state	Written to state
`Chromosome`	▪ Whether a DnaA complex has formed at oriC ▪ DNA-bound protein locations ▪ Superhelicity ▪ DNA strand breaks to be ligated ▪ DNA sequence ▪ DNA footprints of proteins ▪ Chromosome regions accessible for protein binding ▪ Damaged DNA bases ▪ OriC position ▪ TerC position ▪ Sequence Length	▪ Polymerized regions of DNA ▪ DNA-bound protein locations ▪ Unwound bases (Effect on superhelicity) ▪ DNA strand breaks to be ligated

Results from simulations

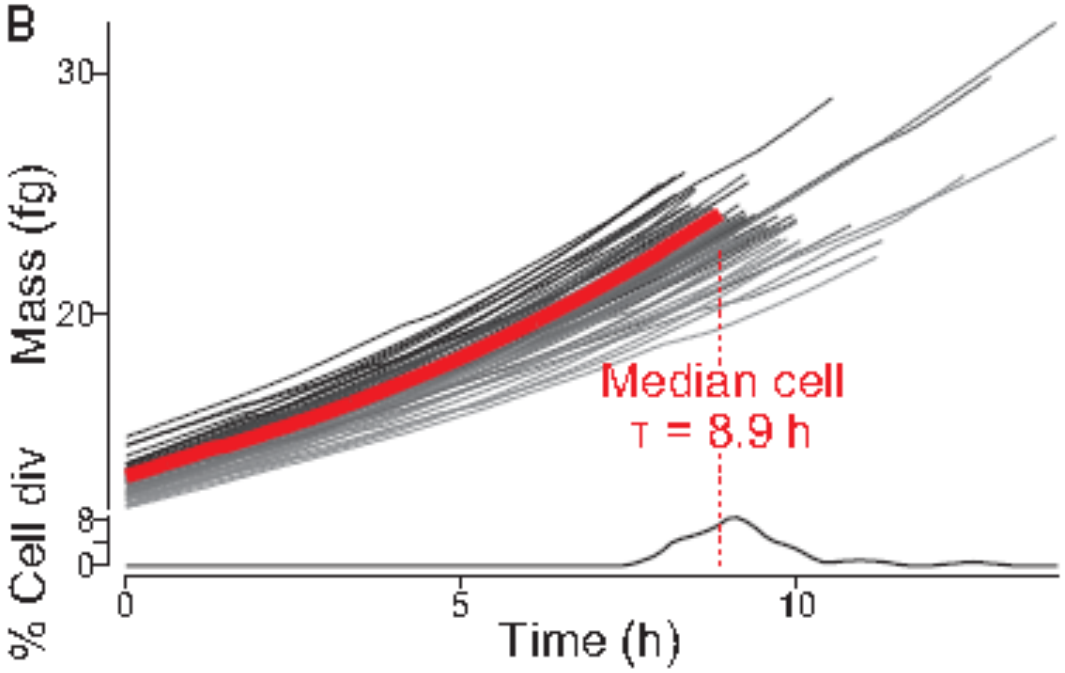
Growth of virtual cell culture



Growth of three cultures (dilutions indicated by shade of blue) and a blank control measured by OD550 of the pH indicator phenol red. The doubling time, t, was calculated using the equation at the top left from the additional time required by more dilute cultures to reach the same OD550 (black lines).

The model calculations were consistent with the observed doubling time!

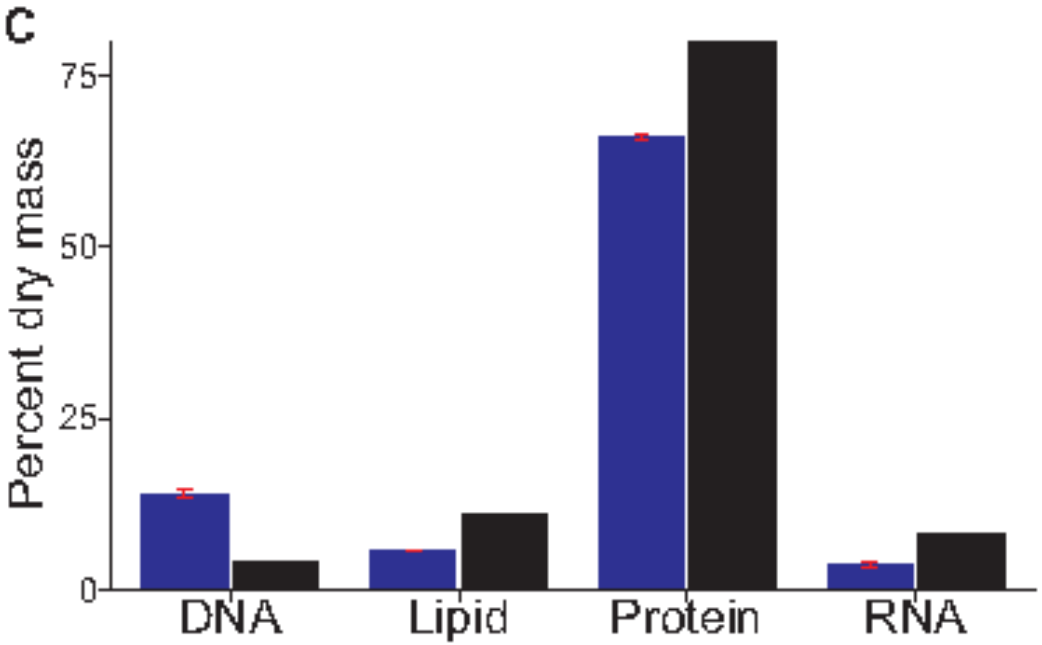
Individual simulations



Predicted growth dynamics of one life cycle of a population of 64 in silico cells

Q: what is the source of the variability of the length of the cell cycle? (later)

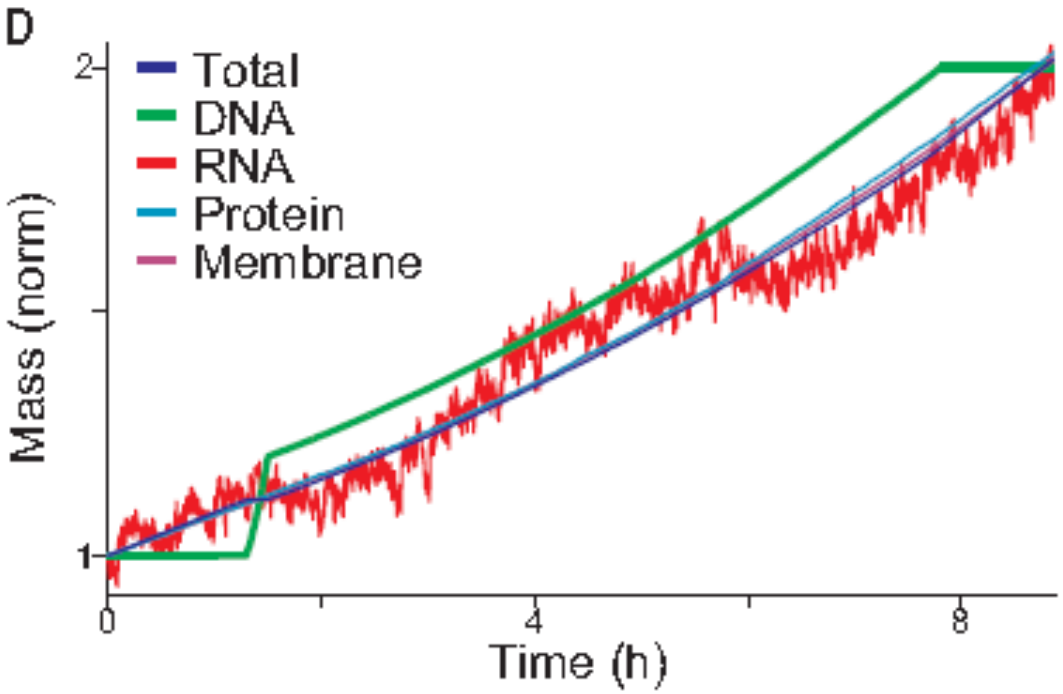
Chemical composition



Comparison of the predicted and experimentally observed cellular chemical compositions

Model calculations were consistent with the observed **cellular chemical composition**!

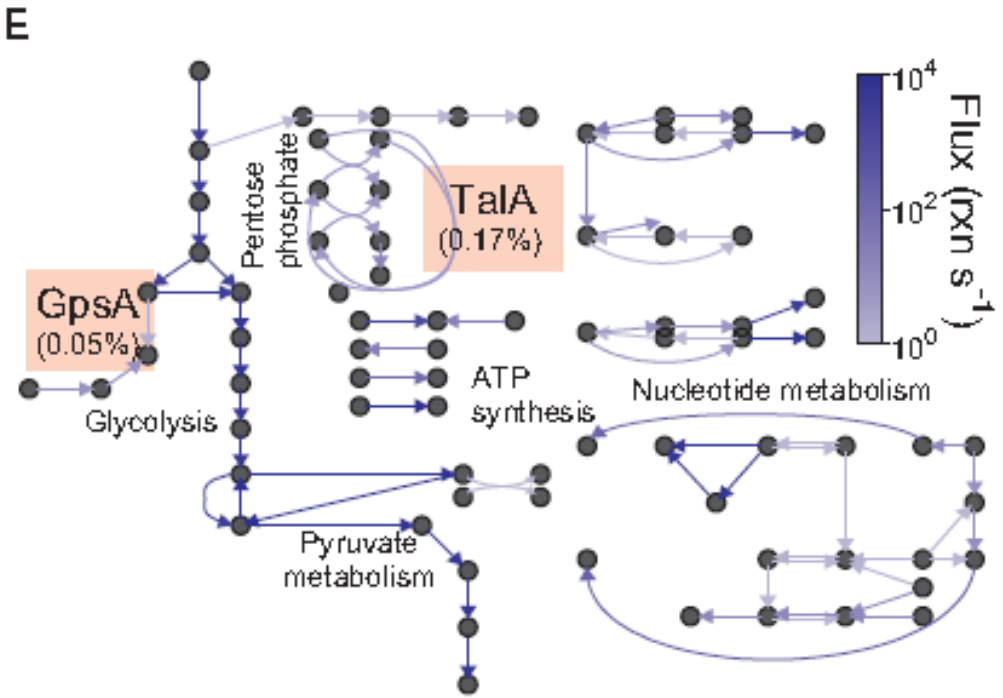
Increase of cell mass



Temporal dynamics of the total cell mass and four cell mass fractions of a representative in silico cell.

Model calculations were consistent with the observed replication of **major cell mass fractions.**

Metabolic flux rates

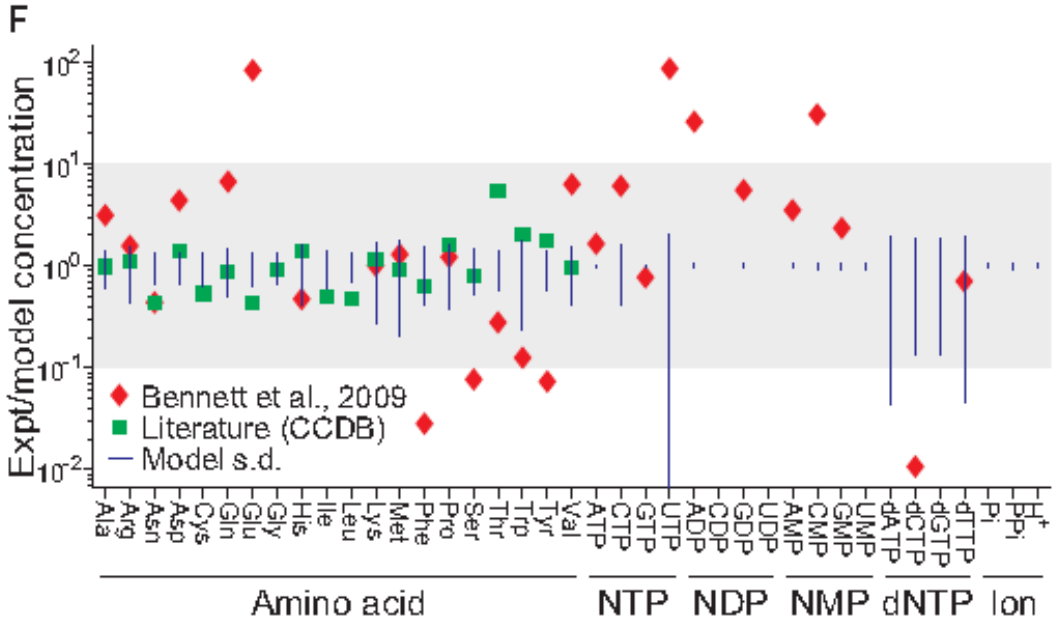


Average predicted metabolic fluxes (from FBA modeling).

Arrow brightness indicates flux magnitude.

In agreement with exp data, the model predicts that the flux through glycolysis is >100-fold more than that through the pentose phosphate and lipid biosynthesis pathways.

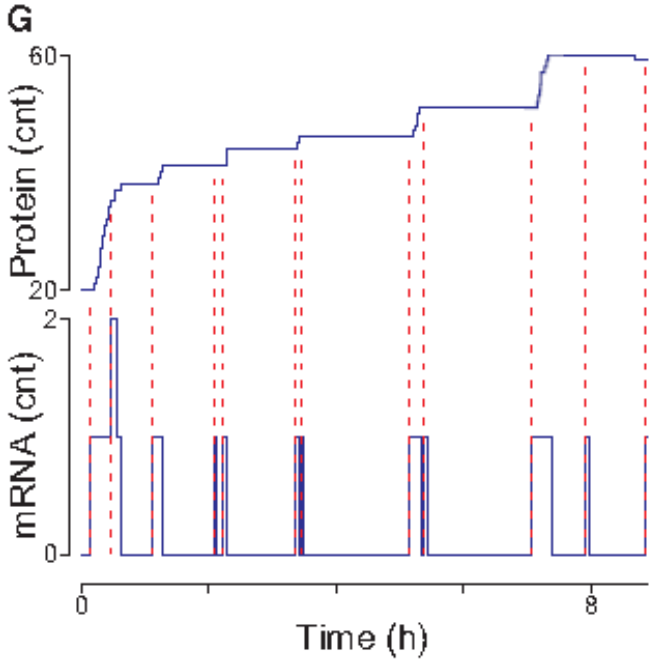
Metabolite concentrations



Ratios of observed and average predicted concentrations of 39 metabolites.

The predicted metabolite concentrations are within an order of magnitude of concentrations measured in Escherichia coli for 100% of the metabolites in one compilation of data and for 70% in a more recent high-throughput study.

mRNA and protein synthesis events



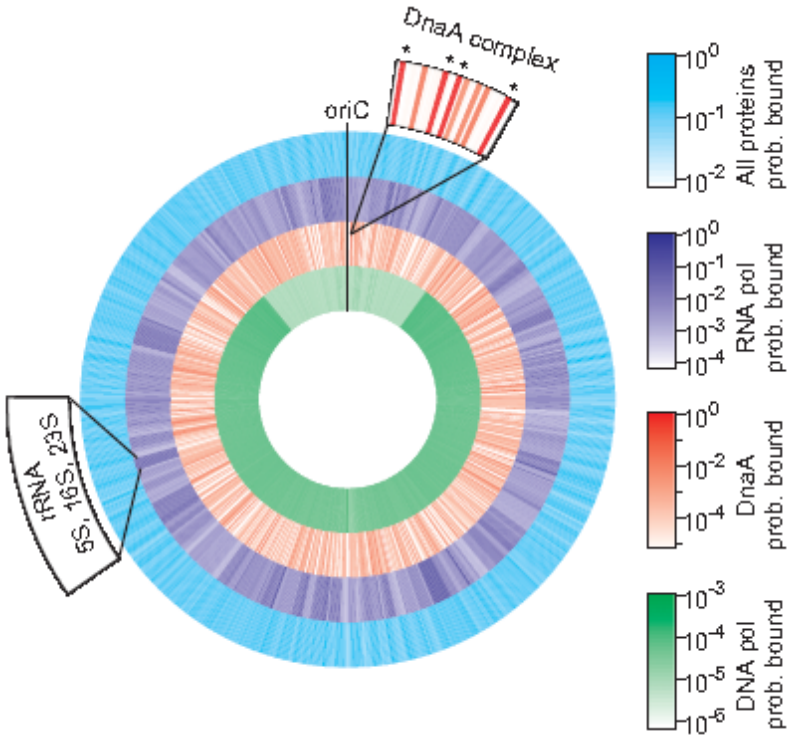
Temporal dynamics of cytadherence high-molecular-weight protein 2 (HMW2, MG218) mRNA and protein expression of one in silico cell.

Red dashed lines indicate the direct link between mRNA synthesis and subsequent bursts in protein synthesis.

Due to (a) the local effect of intermittent messenger RNA (mRNA) expression and (b) the global effect of stochastic protein degradation on the availability of free amino acids for translation,
model predicts ‘‘burst-like’’ protein synthesis.

This is comparable to exp. observations!

Density of DNA-bound proteins

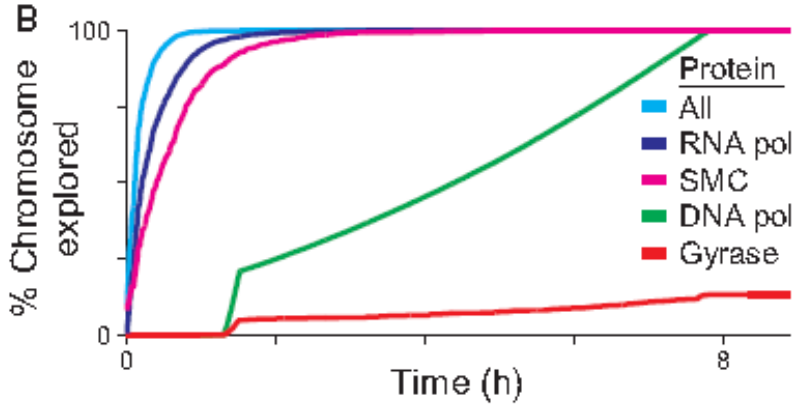


Average density of all DNA-bound proteins and of the replication initiation protein DnaA and DNA and RNA polymerase of a population of 128 in silico cells.
Top magnification : average density of DnaA at several sites near the oriC;
DnaA forms a large multimeric complex at the sites indicated with asterisks, recruiting DNA polymerase to the oriC to initiate replication.
Bottom left : location of the highly expressed rRNA genes..

Consistent with recent experimental data, the predicted high-occupancy RNA polymerase regions correspond to highly transcribed rRNAs and tRNAs.

In contrast, the predicted DNA polymerase chromosomal occupancy is significantly lower and biased toward the terC.

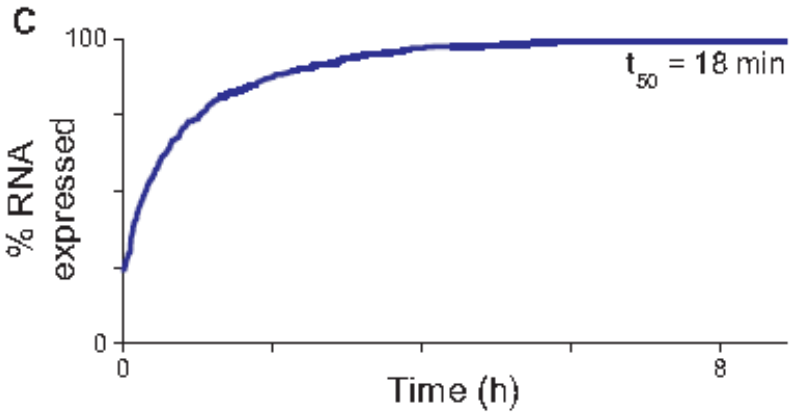
Exploration of genome



Percentage of the chromosome that is predicted to have been bound (B) as functions of time.
SMC is an abbreviation for the name of the chromosome partition protein (MG298).

The model further predicts that the chromosome is explored very rapidly, with 50% of the chromosome having been bound by at least one protein within the first 6 min of the cell cycle and 90% within the first 20 min

Dynamics of genome expression



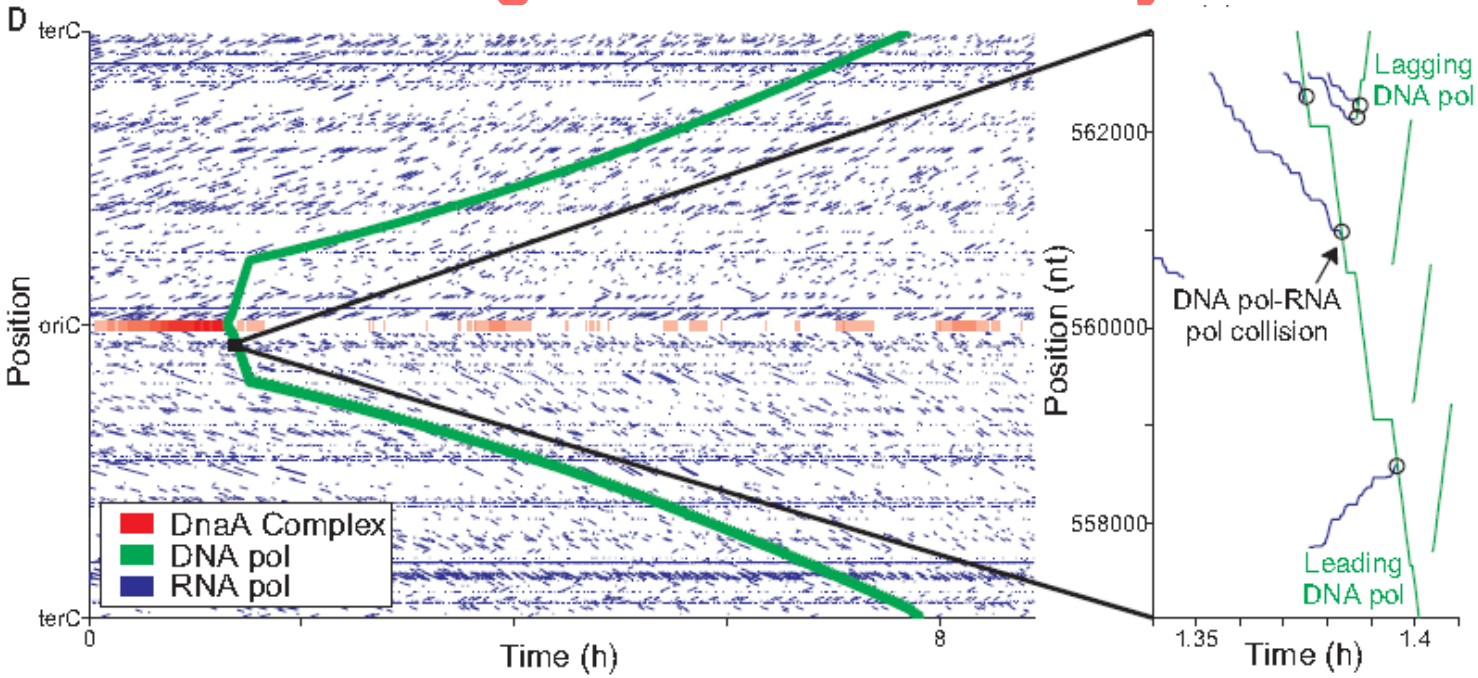
Percentage of the number of genes that are predicted to have been expressed (C) as functions of time.

RNA polymerase contributes the most to chromosomal exploration,

It binds 90% of the chromosome within the first 49 min of the cell cycle.

On average, this results in expression of 90% of genes within the first 143 min.

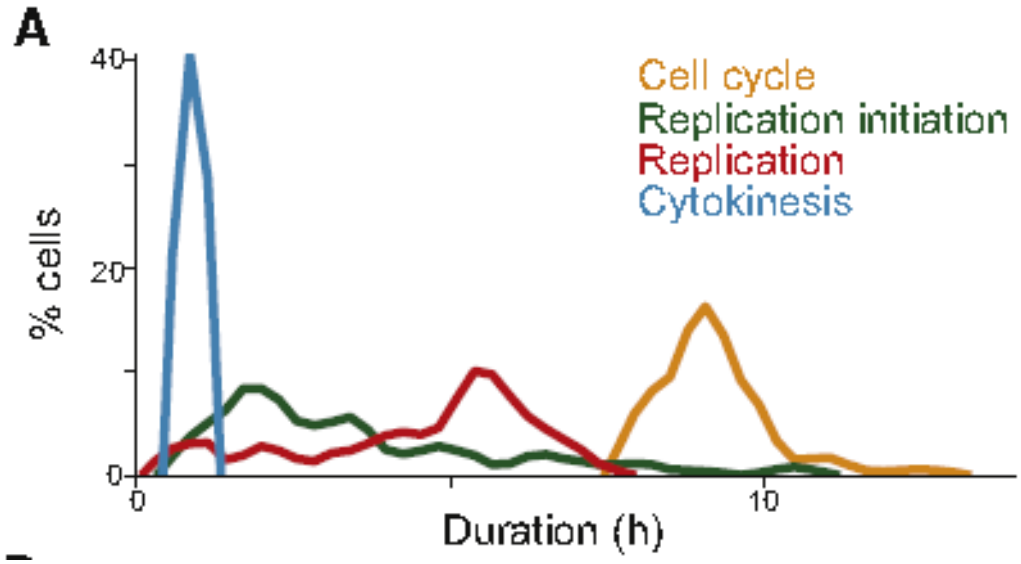
DNA-binding and dissociation dynamics



DNA-binding and dissociation dynamics of the oriC DnaA complex (red) and of RNA (blue) and DNA (green) polymerases for one in silico cell. The oriC DnaA complex recruits DNA polymerase to the oriC to initiate replication, which in turn dissolves the oriC DnaA complex. RNA polymerase traces (blue line segments) indicate individual transcription events. The height, length, and slope of each trace represent the transcript length, transcription duration, and transcript elongation rate, respectively.

Inset : several predicted collisions between DNA and RNA polymerases that lead to the displacement of RNA polymerases and incomplete transcripts.

Predictions for cell-cycle regulation



Distributions of the duration of three cell-cycle phases, as well as that of the total cell-cycle length, across 128 simulations.

There was relatively more cell-to-cell variation in the durations of the replication initiation (64.3%) and replication (38.5%) stages than in cytokinesis (4.4%) or the overall cell cycle (9.4%).

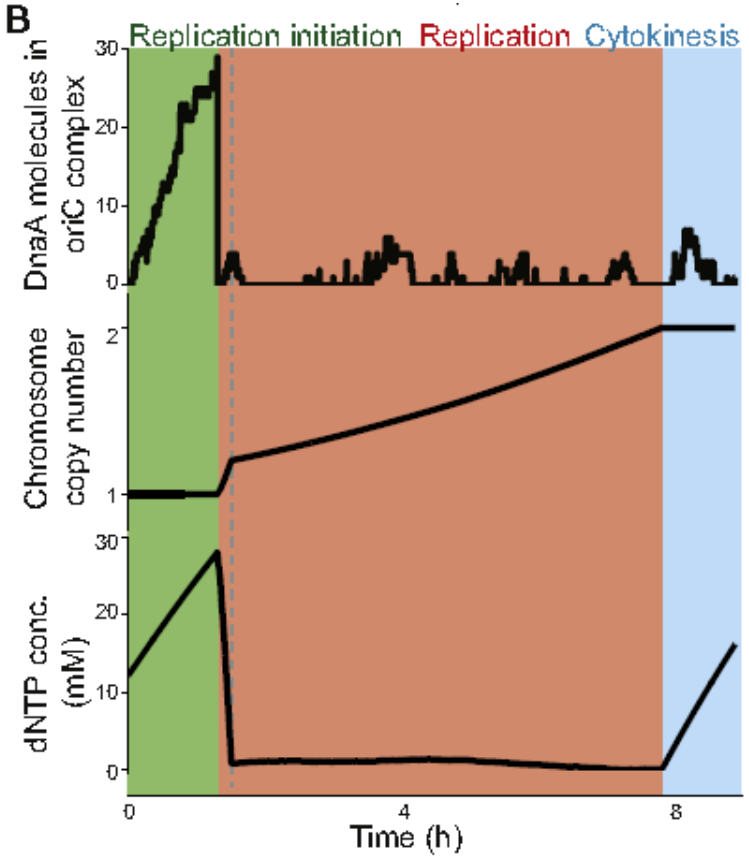
This data raised two questions:
(1)what is the source of duration variability in the initiation and replication phases; and
(2) why is the overall cell-cycle duration less varied than either of these phases?

Replication initiation

Replication initiation occurs as DnaA protein monomers bind or unbind stochastically and cooperatively to form a multimeric complex at the replication origin.

When the complex is complete, DNA polymerase gains access to the origin, and the complex is displaced.

Dynamics of macromolecule abundance

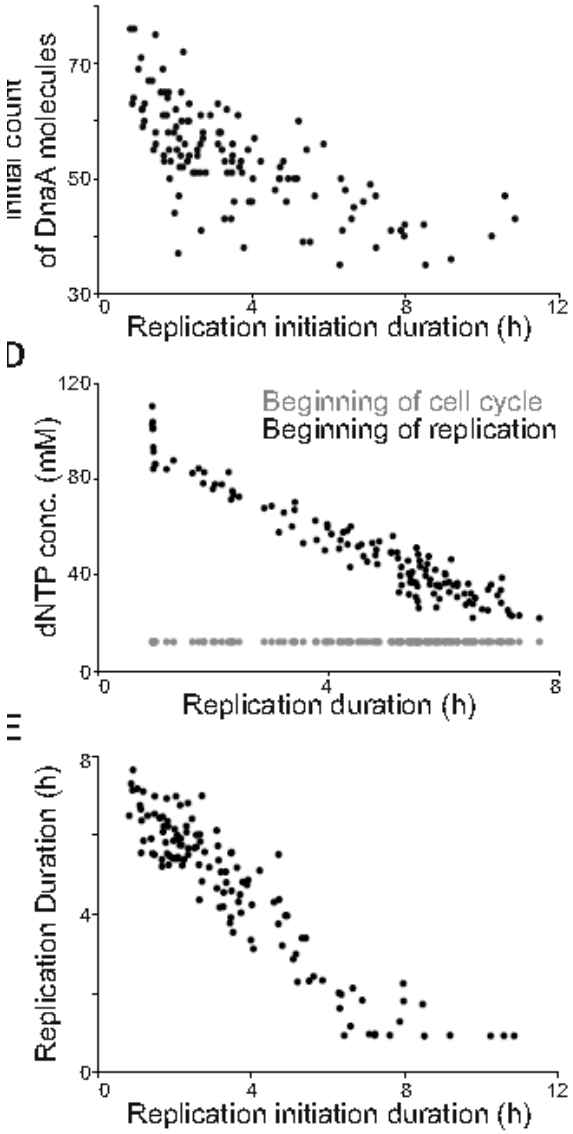


Top : the size of the DnaA complex assembling at the oriC (in monomers of DnaA);

middle, the copy number of the chromosome;

Bottom : cytosolic dNTP concentration. The quantities of these macromolecules correlate strongly with the timing of key cell-cycle stages.

What determines replication duration?



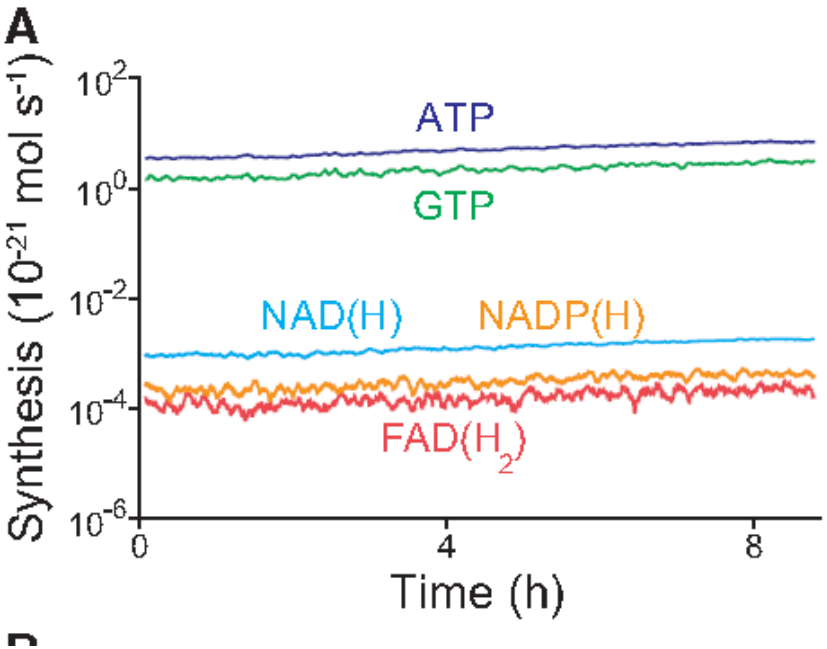
We found a correlation ($R^2 = 0.49$) between the predicted duration of replication initiation and the initial number of free DnaA monomers.

The duration of the replication phase in individual cells is more closely related to the free dNTP content at the start of replication than to the dNTP content at the start of the cell cycle

The durations of the initiation and replication phases are inversely related to each other in single cells.

Cells that require extra time to initiate replication also build up a large dNTP surplus, leading to faster replication.

Cellular energy sources

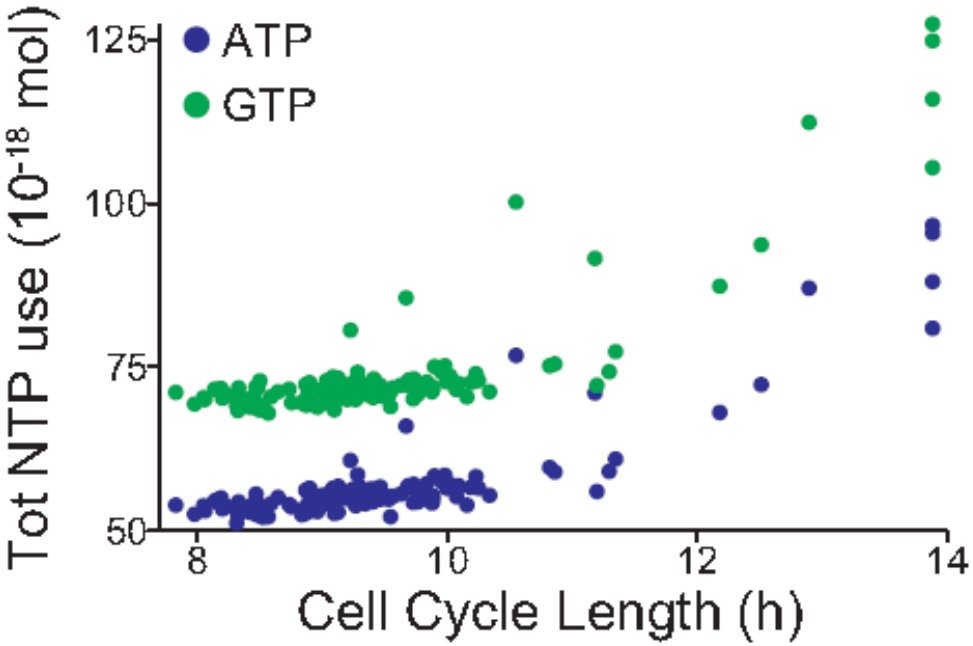


Intracellular concentrations of the cellular energy carriers over time for one in silico cell.

ATP and GTP are synthesized more than 1000-fold faster than other High-energy intermediates.

Increases over time look small due to logarithmic scale.

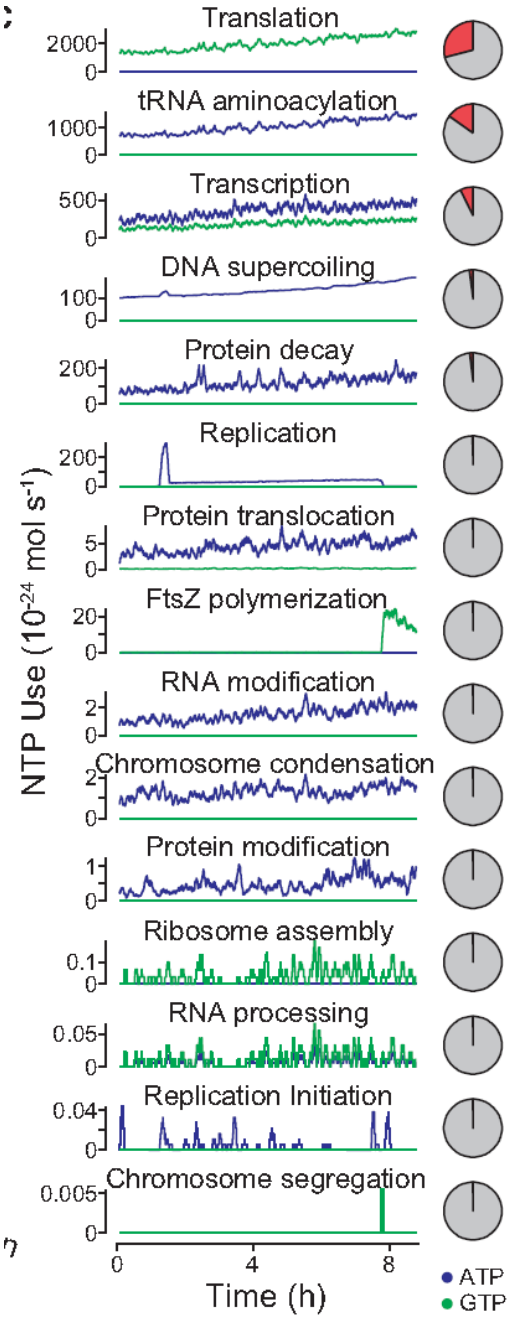
Usage of ATP and GTP



Overall of ATP and GTP did not vary considerably in all simulations.

Exception: very slow cells consume about twiche as much energy.

Usage of ATP and GTP

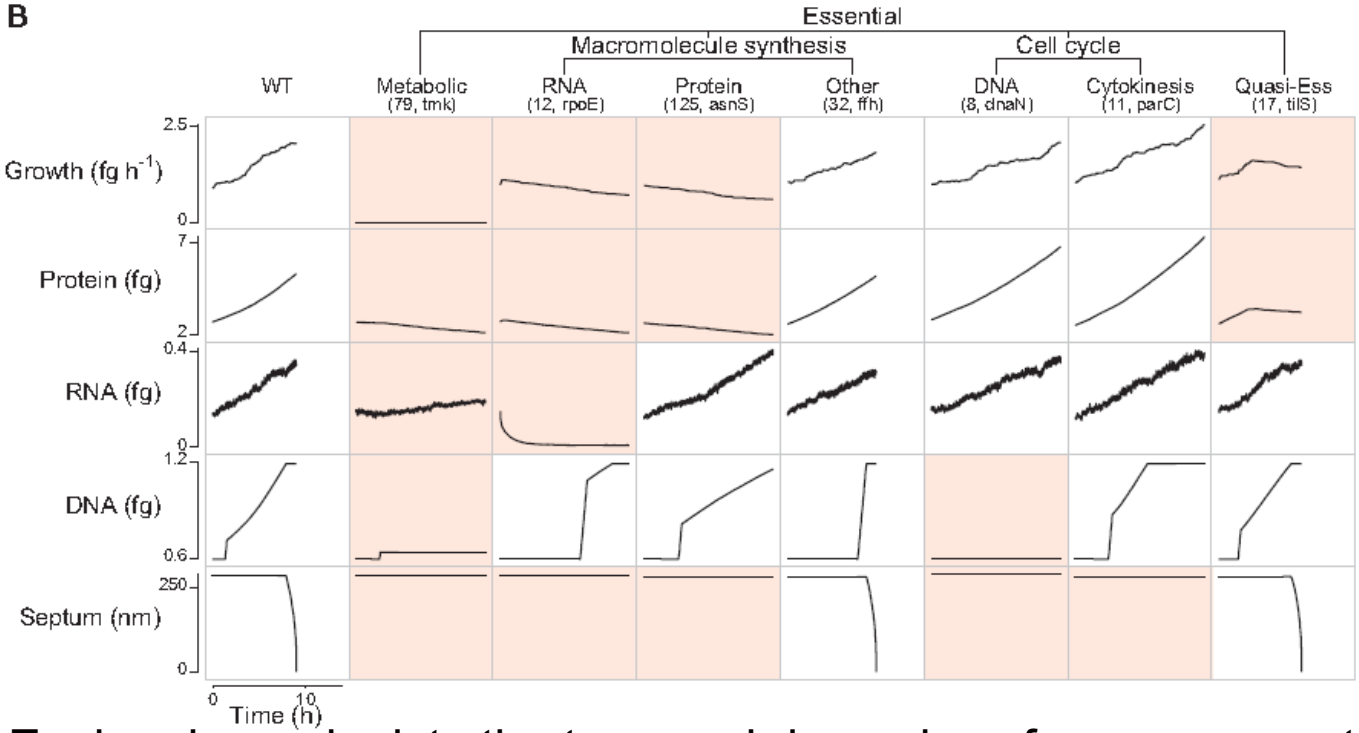


ATP (blue) and GTP (green) usage of 15 cellular processes throughout life cycle of one in silico cell.

Pie charts: percent used as a fraction of the total usage.

Usage is dominated by production of mRNA and protein.

Single-gene knockouts : essential vs. non-essential genes



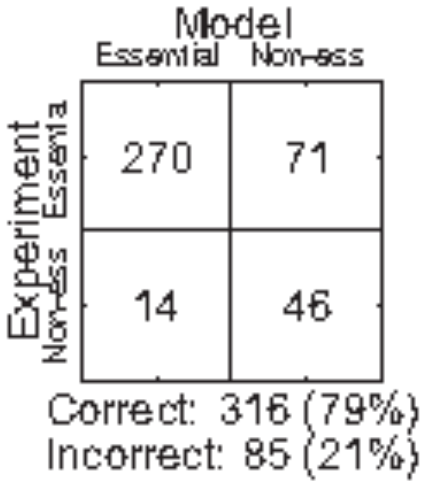
Single-gene disruption strains grouped into phenotypic classes (columns) according to their capacity to grow, synthesize protein, RNA, and DNA, and divide (indicated by septum length).

Each column depicts the temporal dynamics of one representative in silico cell of each essential disruption strain class.

Dynamics significantly different from wild-type are highlighted in red.

The identity of the representative cell and the number of disruption strains in each category are indicated in parenthesis.

Construct in-silico mutants to predict essentiality



Construct all possible 525 single-gene deletions.

Comparison of predicted and observed gene essentiality.

Then make predictions for new biological insights (important to be able to publish in Cell, not discussed here because too detailed.

Summary

Comprehensive whole-cell model accounts for all annotated gene functions identified in M. genitalium and explains a variety of emergent behaviors in terms of molecular interactions.

This is still a first draft.

Whole-cell models may accelerate biological discovery and bioengineering by facilitating experimental design and interpretation (?).

Combined with recent de novo synthesis of the *M. genitalium* chromosome and successful genome transplantation of Mycoplasma genomes to produce a synthetic cell (Craig Venter), this raise the exciting possibility of using whole-cell models to enable computer-aided rational design of novel microorganisms.

The construction of whole-cell models and the iterative testing of them against experimental information will enable the scientific community to assess how well we understand integrated cellular systems.