# The Calculus of Looping Sequences

Roberto Barbuti, Giulio Caravagna, Andrea MaggioloSchettini,
Paolo Milazzo, Giovanni Pardini

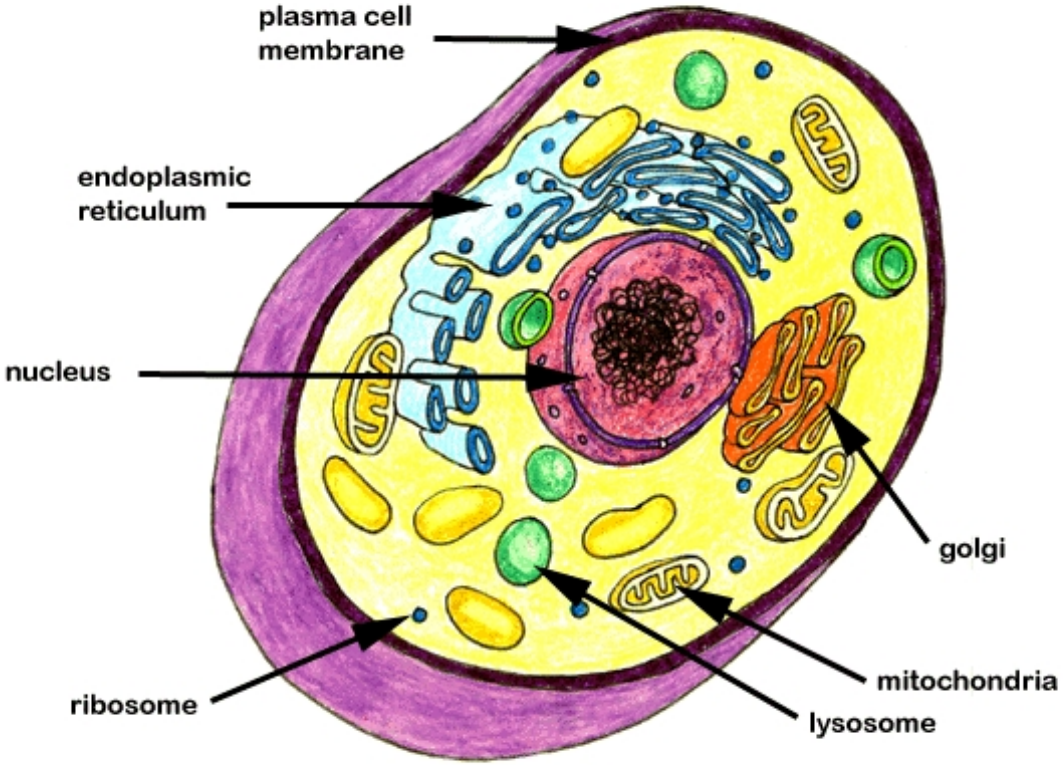
Dipartimento di Informatica, Università di Pisa, Italy

Bertinoro – June 7, 2008

# Outline of the talk


1. Introduction
   - Cells are complex interactive systems
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - The EGF pathway and the *lac* operon in CLS
3. CLS variants
   - Stochastic CLS
   - LCLS
   - Spatial CLS
4. Future Work and References

# Cells: complex systems of interactive components




- Main actors:
  - membranes
  - proteins
  - DNA/RNA
  - ions, macromolecules,...
- Interaction networks:
  - metabolic pathways
  - signaling pathways
  - gene regulatory networks

Computer Science can provide biologists with formalisms for the description of interactive systems and tools for their analysis.

# Outline of the talk


1. Introduction
   - Cells are complex interactive systems
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - The EGF pathway and the *lac* operon in CLS
3. CLS variants
   - Stochastic CLS
   - LCLS
   - Spatial CLS
4. Future Work and References

# The Calculus of Looping Sequences (CLS)

We assume an alphabet $\mathcal{E}$. **Terms** $T$ and **Sequences** $S$ of CLS are given by the following grammar:

$$\begin{array}{rcl} T & ::= & S \;\mid\; (S)^L \rfloor T \;\mid\; T \mid T \\ S & ::= & \epsilon \;\mid\; a \;\mid\; S \cdot S \end{array}$$

where $a$ is a generic element of $\mathcal{E}$, and $\epsilon$ is the empty sequence.

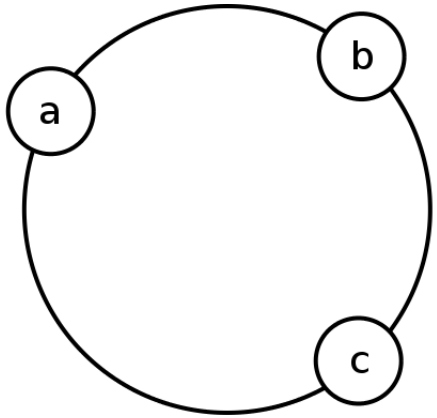
The operators are:

| | | |
|---|---|---|
| $S \cdot S$ | : | Sequencing |
| $(S)^L$ | : | Looping ($S$ is closed and it can rotate) |
| $T_1 \rfloor T_2$ | : | Containment ($T_1$ contains $T_2$) |
| $T \mid T$ | : | Parallel composition (juxtaposition) |

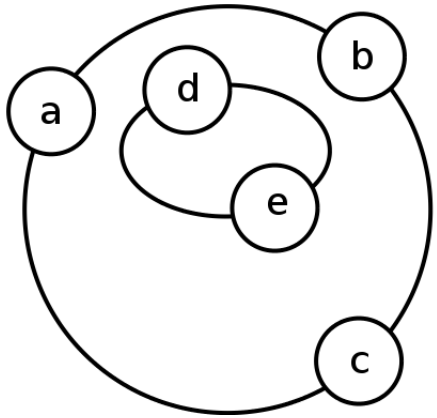
Actually, looping and containment form a single binary operator $(S)^L \rfloor T$.

# Examples of Terms

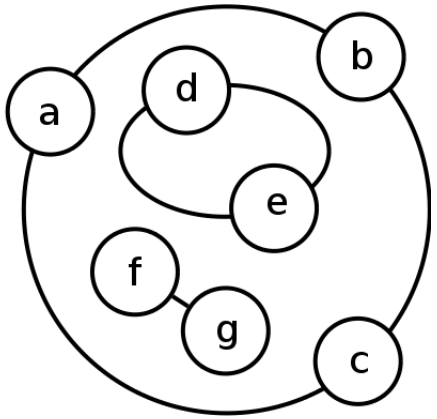
(i)




(ii)




(iii)




$$(i) \qquad \left(a \cdot b \cdot c\right)^{L} \rfloor \epsilon$$

$$(ii) \qquad \left(a \cdot b \cdot c\right)^{L} \rfloor \left(d \cdot e\right)^{L} \rfloor \epsilon$$

$$(iii) \qquad \left(a \cdot b \cdot c\right)^{L} \rfloor \left(f \cdot g \mid \left(d \cdot e\right)^{L} \rfloor \epsilon\right)$$

# Structural Congruence

The **Structural Congruence** relations $\equiv_S$ and $\equiv_T$ are the least congruence relations on sequences and on terms, respectively, satisfying the following rules:

$$S_1 \cdot (S_2 \cdot S_3) \equiv_S (S_1 \cdot S_2) \cdot S_3 \qquad S \cdot \epsilon \equiv_S \epsilon \cdot S \equiv_S S$$

$$T_1 \mid T_2 \equiv_T T_2 \mid T_1 \qquad T_1 \mid (T_2 \mid T_3) \equiv_T (T_1 \mid T_2) \mid T_3$$

$$T \mid \epsilon \equiv_T T \quad \left(\epsilon\right)^L \rfloor \epsilon \equiv_T \epsilon \quad \left(S_1 \cdot S_2\right)^L \rfloor T \equiv_T \left(S_2 \cdot S_1\right)^L \rfloor T$$

We write $\equiv$ for $\equiv_T$.

# CLS Patterns

Let us consider variables of three kinds:

- term variables $(X, Y, Z, \ldots)$
- sequence variables $(\widetilde{x}, \widetilde{y}, \widetilde{z}, \ldots)$
- element variables $(x, y, z, \ldots)$

**Patterns** $P$ and **Sequence Patterns** $SP$ of CLS extend CLS terms and sequences with variables:

$$\begin{array}{rcl} P & ::= & SP \;\mid\; (SP)^{L} \rfloor P \;\mid\; P \mid P \;\mid\; X \\ SP & ::= & \epsilon \;\mid\; a \;\mid\; SP \cdot SP \;\mid\; x \;\mid\; \widetilde{x} \end{array}$$

where $a$ is a generic element of $\mathcal{E}$, $\epsilon$ is the empty sequence, and $x, \widetilde{x}$ and $X$ are generic element, sequence and term variables

The structural congruence relation $\equiv$ extends trivially to patterns

# Rewrite Rules

$P\sigma$ denotes the term obtained by replacing any variable in $T$ with the corresponding term, sequence or element.

$\Sigma$ is the set of all possible instantiations $\sigma$

A **Rewrite Rule** is a pair $(P, P')$, denoted $P \mapsto P'$, where:

- $P, P'$ are patterns
- variables in $P'$ are a subset of those in $P$

A rule $P \mapsto P'$ can be applied to all terms $P\sigma$.

Example: $a \cdot x \cdot a \mapsto b \cdot x \cdot b$

- can be applied to $a \cdot c \cdot a$ (producing $b \cdot c \cdot b$)
- cannot be applied to $a \cdot c \cdot c \cdot a$

# Formal Semantics

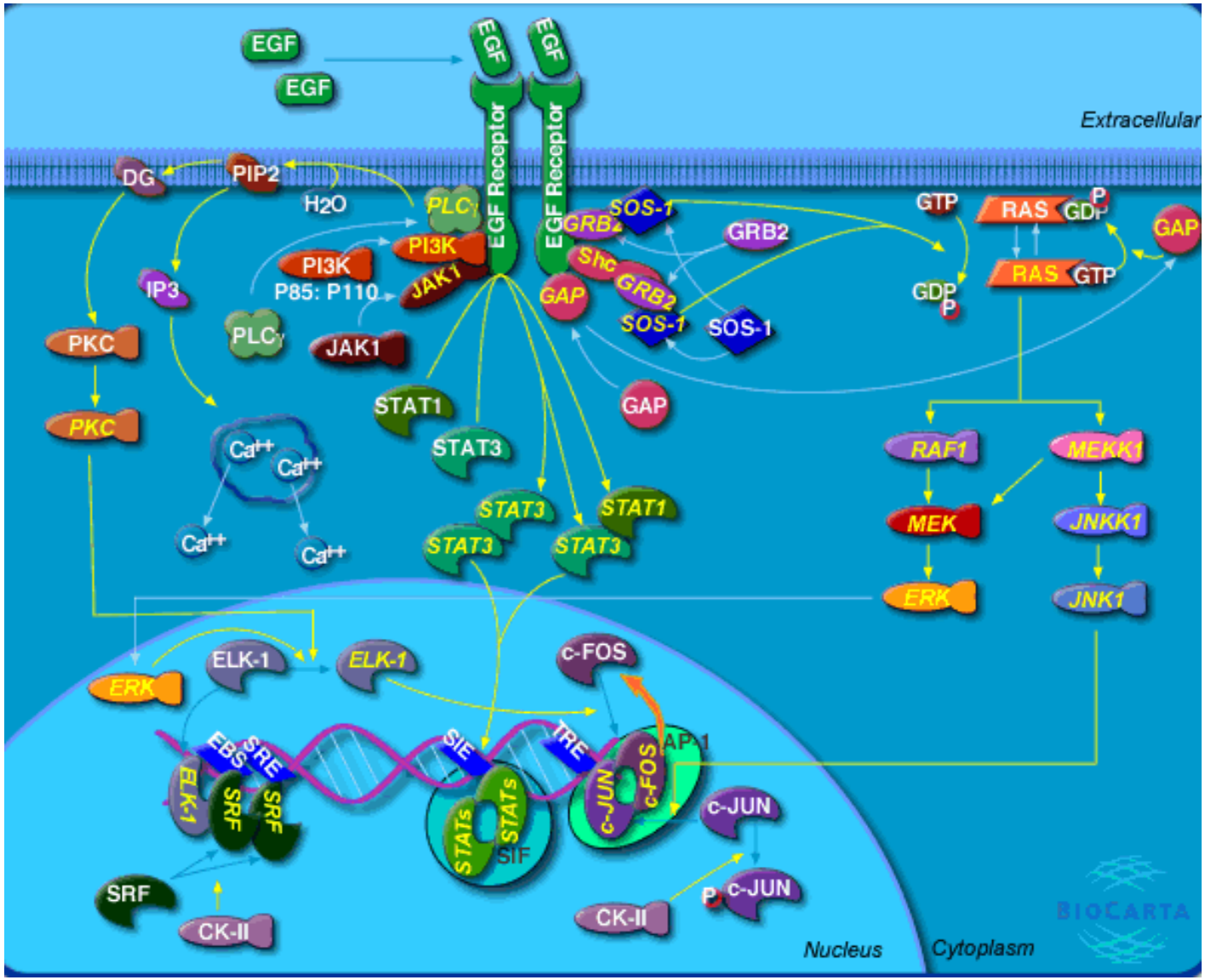
Given a set of rewrite rules $\mathcal{R}$, evolution of terms is described by the transition system given by the least relation $\rightarrow$ satisfying

$$\frac{P \mapsto P' \in \mathcal{R} \qquad P\sigma \not\equiv \epsilon}{P\sigma \rightarrow P'\sigma}$$

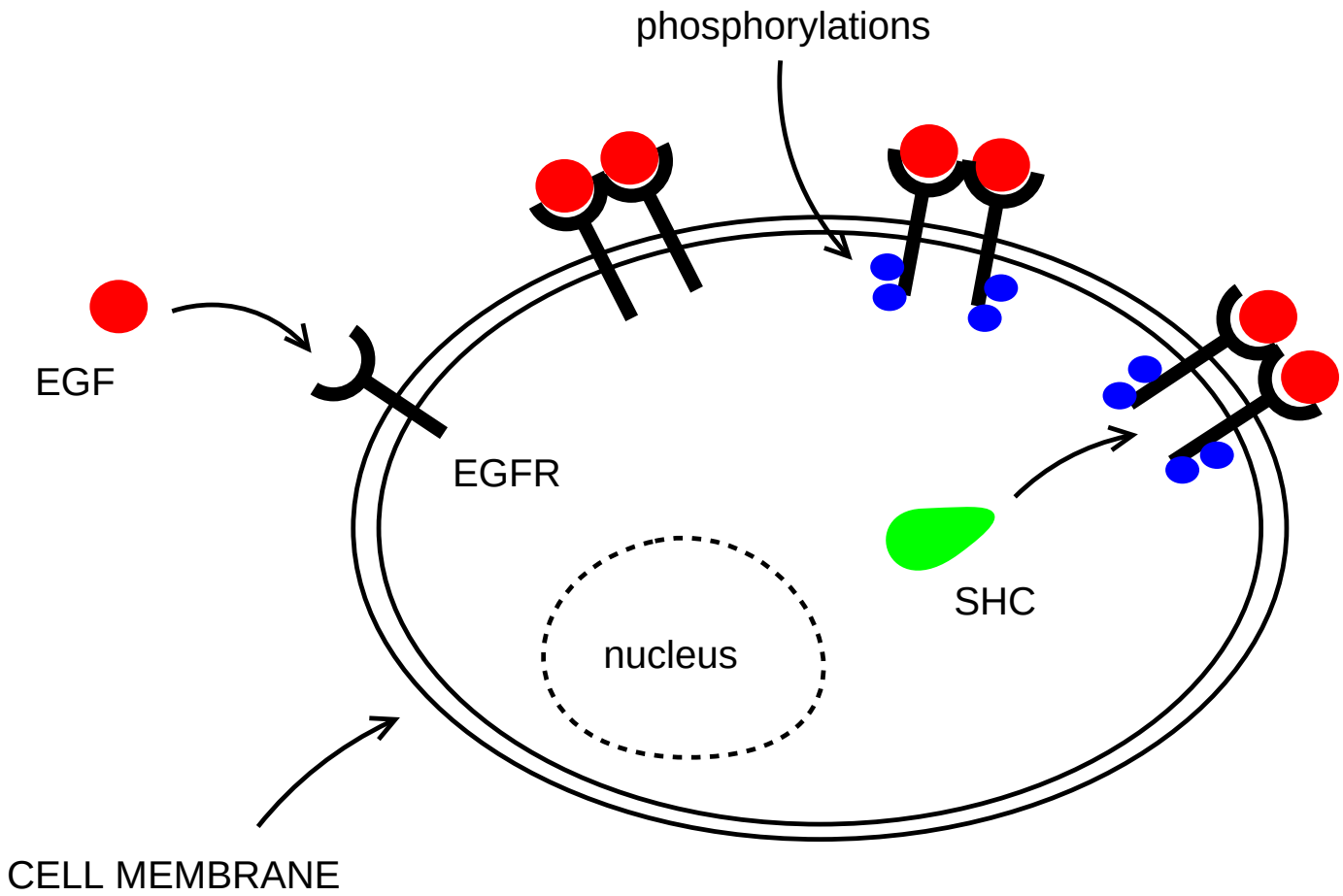
$$\frac{T \rightarrow T'}{T \mid T'' \rightarrow T' \mid T''} \qquad \frac{T \rightarrow T'}{(S)^L \rfloor T \rightarrow (S)^L \rfloor T'}$$

and closed under structural congruence $\equiv$.

# CLS modeling examples: the EGF pathway (1)



# CLS modeling examples: the EGF pathway (2)



# CLS modeling examples: the EGF pathway (3)

First steps of the EGF signaling pathway up to the binding of the signal-receptor dimer to the SHC protein

- The EGFR,EGF and SHC proteins are modeled as the alphabet symbols *EGFR*, *EGF* and *SHC*, respectively
- The cell is modeled as a looping sequence (representing its external membrane):

$$EGF \mid EGF \mid \left(EGFR \cdot EGFR \cdot EGFR \cdot EGFR\right)^{L} \rfloor \left(SHC \mid SHC\right)$$

Rewrite rules modeling the first steps of the pathway:

$$EGF \mid \left(EGFR \cdot \widetilde{x}\right)^{L} \rfloor X \;\mapsto\; \left(CMPLX \cdot \widetilde{x}\right)^{L} \rfloor X \qquad \text{(R1)}$$

$$\left(CMPLX \cdot \widetilde{x} \cdot CMPLX \cdot \widetilde{y}\right)^{L} \rfloor X \;\mapsto\; \left(DIM \cdot \widetilde{x} \cdot \widetilde{y}\right)^{L} \rfloor X \qquad \text{(R2)}$$

$$\left(DIM \cdot \widetilde{x}\right)^{L} \rfloor X \;\mapsto\; \left(DIMp \cdot \widetilde{x}\right)^{L} \rfloor X \qquad \text{(R3)}$$

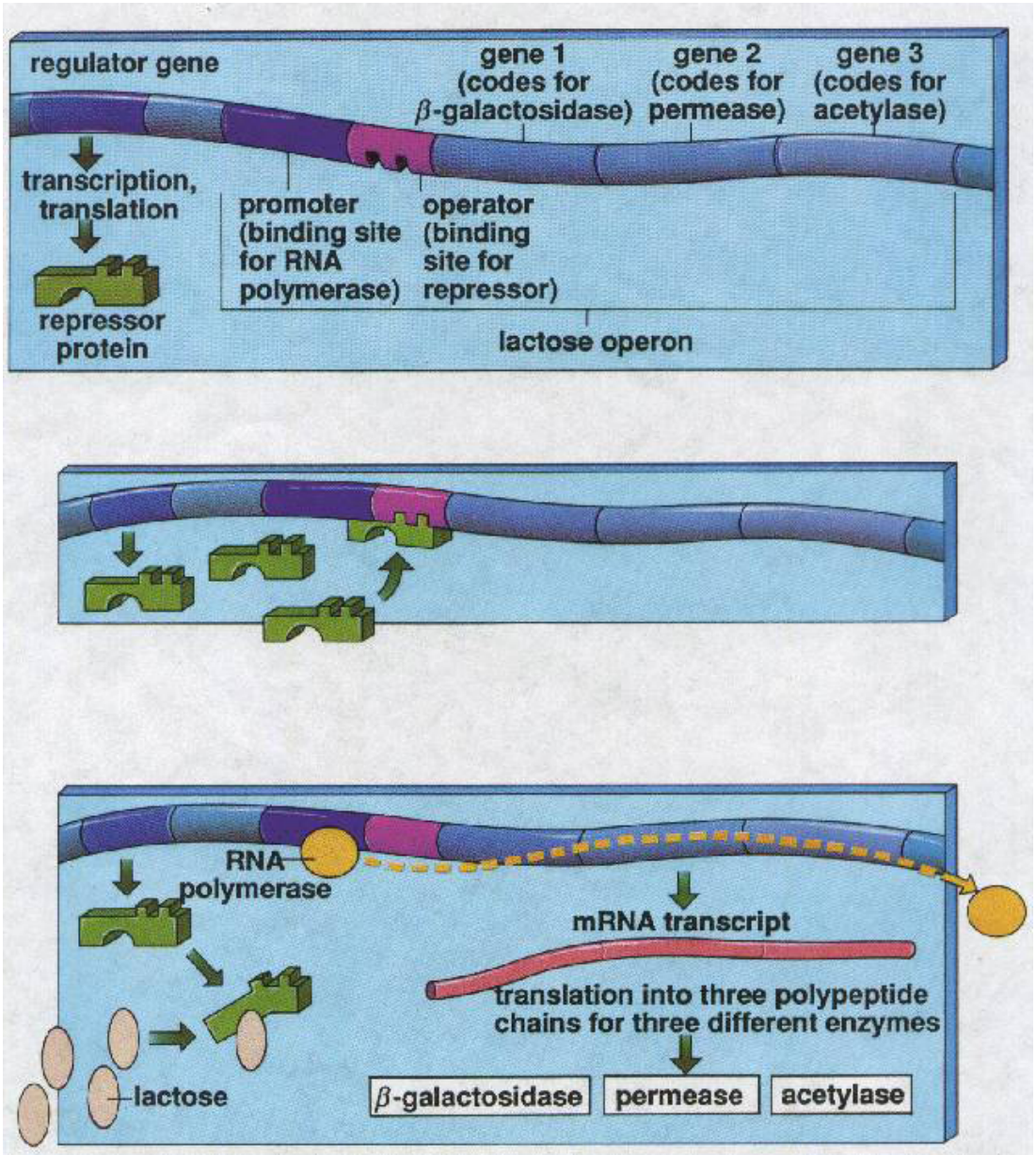
$$\left(DIMp \cdot \widetilde{x}\right)^{L} \rfloor \left(SHC \mid X\right) \;\mapsto\; \left(DIMpSHC \cdot \widetilde{x}\right)^{L} \rfloor X \qquad \text{(R4)}$$

# CLS modeling examples: the EGFR pathway (4)

A possible evolution of the system:

$$
\begin{aligned}
& EGF \mid EGF \mid \left(EGFR \cdot EGFR \cdot EGFR \cdot EGFR\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R1)} \; & EGF \mid \left(EGFR \cdot CMPLX \cdot EGFR \cdot EGFR\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R1)} \; & \left(EGFR \cdot CMPLX \cdot EGFR \cdot CMPLX\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R2)} \; & \left(EGFR \cdot DIM \cdot EGFR\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R3)} \; & \left(EGFR \cdot DIMp \cdot EGFR\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R4)} \; & \left(EGFR \cdot DIMpSHC \cdot EGFR\right)^{L} \rfloor SHC
\end{aligned}
$$

# CLS modeling examples: the *lac* operon (1)



# CLS modeling examples: the *lac* operon (2)

$$Ecoli \;::=\; (m)^{L} \rfloor (lacI \cdot lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA \mid polym)$$

Rules for DNA transcription/translation:

$$lacI \cdot \widetilde{x} \;\mapsto\; lacI' \cdot \widetilde{x} \mid repr \qquad \text{(R1)}$$

$$polym \mid \widetilde{x} \cdot lacP \cdot \widetilde{y} \;\mapsto\; \widetilde{x} \cdot PP \cdot \widetilde{y} \qquad \text{(R2)}$$

$$\widetilde{x} \cdot PP \cdot lacO \cdot \widetilde{y} \;\mapsto\; \widetilde{x} \cdot lacP \cdot PO \cdot \widetilde{y} \qquad \text{(R3)}$$

$$\widetilde{x} \cdot PO \cdot lacZ \cdot \widetilde{y} \;\mapsto\; \widetilde{x} \cdot lacO \cdot PZ \cdot \widetilde{y} \qquad \text{(R4)}$$

$$\widetilde{x} \cdot PZ \cdot lacY \cdot \widetilde{y} \;\mapsto\; \widetilde{x} \cdot lacZ \cdot PY \cdot \widetilde{y} \mid betagal \qquad \text{(R5)}$$

$$\widetilde{x} \cdot PY \cdot lacA \;\mapsto\; \widetilde{x} \cdot lacY \cdot PA \mid perm \qquad \text{(R6)}$$

$$\widetilde{x} \cdot PA \;\mapsto\; \widetilde{x} \cdot lacA \mid transac \mid polym \qquad \text{(R7)}$$

# CLS modeling examples: the *lac* operon (3)

$$Ecoli \;::=\; (m)^{L} \rfloor (lacI \cdot lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA \mid polym)$$

Rules to describe the binding of the lac Repressor to gene o, and what happens when lactose is present in the environment of the bacterium:

$$repr \mid \widetilde{x} \cdot lacO \cdot \widetilde{y} \;\mapsto\; \widetilde{x} \cdot RO \cdot \widetilde{y} \qquad \text{(R8)}$$

$$\widetilde{x} \cdot RO \cdot \widetilde{y} \;\mapsto\; repr \mid \widetilde{x} \cdot lacO \cdot \widetilde{y} \qquad \text{(R9)}$$

$$repr \mid LACT \;\mapsto\; RLACT \qquad \text{(R10)}$$

$$RLACT \;\mapsto\; repr \mid LACT \qquad \text{(R11)}$$

$$(\widetilde{x})^{L} \rfloor (perm \mid X) \;\mapsto\; (perm \cdot \widetilde{x})^{L} \rfloor X \qquad \text{(R12)}$$

$$LACT \mid (perm \cdot \widetilde{x})^{L} \rfloor X \;\mapsto\; (perm \cdot \widetilde{x})^{L} \rfloor (LACT \mid X) \qquad \text{(R13)}$$

$$betagal \mid LACT \;\mapsto\; betagal \mid GLU \mid GAL \qquad \text{(R14)}$$

# Some variants of CLS

- Full–CLS
  - The looping operator can be applied to any term
  - Terms such as $\left(a \mid \left(b\right)^{L} \rfloor c\right)^{L} \rfloor d$ are allowed
- CLS+
  - More realistic representation of the fluid nature of membranes: the looping operator can be applied to parallel compositions of sequences
  - Can be encoded into CLS
- Stochastic CLS
  - The application of a rule consumes a stochastic quantity of time
- LCLS (CLS with Links)
  - Description of protein–protein interactions at the domain level
- Spatial CLS
  - Description of physical size and position of entities

# Outline of the talk


1. Introduction
   - Cells are complex interactive systems
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - The EGF pathway and the *lac* operon in CLS
3. CLS variants
   - Stochastic CLS
   - LCLS
   - Spatial CLS
4. Future Work and References

# Background: the kinetics of chemical reactions

Usual notation for chemical reactions:

$$\ell_1 S_1 + \ldots + \ell_\rho S_\rho \underset{k_{-1}}{\overset{k}{\rightleftharpoons}} \ell'_1 P_1 + \ldots + \ell'_\gamma P_\gamma$$

where:

- $S_i, P_i$ are molecules (reactants)
- $\ell_i, \ell'_i$ are stoichiometric coefficients
- $k, k_{-1}$ are the kinetic constants

The kinetics is described by the *law of mass action*:

$$\frac{d[P_i]}{dt} = \ell'_i \underbrace{k[S_1]^{\ell_1} \cdots [S_\rho]^{\ell_\rho}}_{\textit{reaction rate}} - \ell'_i \underbrace{k_{-1}[P_1]^{\ell'_1} \cdots [P_\gamma]^{\ell'_\gamma}}_{\textit{reaction rate}}$$

# Background: Gillespie's simulation algorithm

- represents a chemical solution as a multiset of molecules
- computes the reaction rate $a_\mu$ by multiplying the kinetic constant by the number of possible combinations of reactants

Example: chemical solution with $X_1$ molecules $S_1$ and $X_2$ molecules $S_2$

reaction $R_1 : S_1 + S_2 \rightarrow 2S_1$ rate $a_1 = \binom{X_1}{1}\binom{X_2}{1}k_1 = X_1 X_2 k_1$

reaction $R_2 : 2S_1 \rightarrow S_1 + S_2$ rate $a_2 = \binom{X_1}{2}k_2 = \frac{X_1(X_1-1)}{2}k_2$

Given a set of reactions $\{R_1, \ldots R_M\}$ and a current time $t$

- The time $t + \tau$ at which the next reaction will occur is randomly chosen with $\tau$ exponentially distributed with parameter $\sum_{\nu=1}^{M} a_\nu$;
- The reaction $R_\mu$ that has to occur at time $t + \tau$ is randomly chosen with probability $\frac{a_\mu}{\sum_{\nu=1}^{M} a_\nu}$.

At each step $t$ is incremented by $\tau$ and the chemical solution is updated.

# Stochastic CLS (1)

Stochastic CLS incorporates Gillespie's stochastic framework into the semantics of CLS

Two main problems:

- What is a reactant in Stochastic CLS?
  - A *subterm* of a term $T$ is a term $T' \not\equiv \epsilon$ such that $T \equiv C[T']$ for some context $C$
  - A *reactant* is an occurence of a subterm
- What happens with variables?
  - We consider a rule $(a)^L \rfloor (b \mid X) \mapsto (c)^L \rfloor X$ as a reaction between a molecule $a$ on a membrane and *any* molecule $b$ contained in the membrane.
  - The semantics has to count how many times $b$ occurs in the instantiation of $X$

# Stochastic CLS (2)

Let us assume the syntax of Full–CLS...

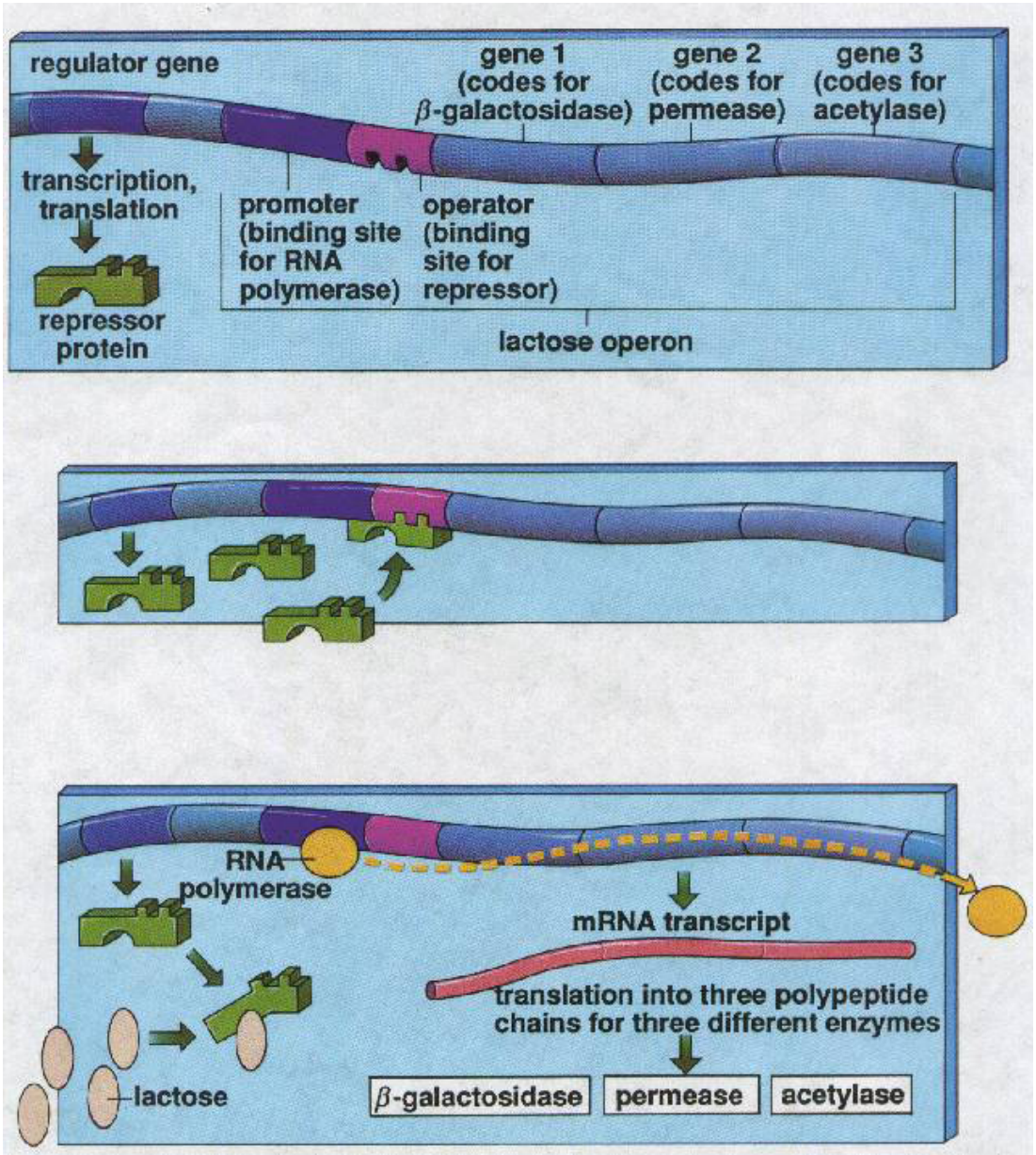
Given a finite set of stochastic rewrite rules $\mathcal{R}$, the semantics of Stochastic CLS is the least transition relation $\xrightarrow{R,T,r,b}$ closed wrt $\equiv$ and satisfying by the following inference rules:

$$\frac{R : P_L \overset{k}{\mapsto} P_R \in \mathcal{R} \quad \sigma \in \Sigma}{P_L\sigma \xrightarrow{R,P_L\sigma,k\cdot comb(P_L,\sigma),1} P_R\sigma} \qquad \frac{T_1 \xrightarrow{R,T,r,b} T_2}{T_1 \mid T_3 \xrightarrow{R,T,r,b\cdot binom(T,T_1,T_3)} T_2 \mid T_3}$$

$$\frac{T_1 \xrightarrow{R,T,r,b} T_2}{(T_1)^L \rfloor T_3 \xrightarrow{R,(T_1)^L \rfloor T_3,r\cdot b,1} (T_2)^L \rfloor T_3} \qquad \frac{T_1 \xrightarrow{R,T,r,b} T_2}{(T_3)^L \rfloor T_1 \xrightarrow{R,(T_3)^L \rfloor T_1,r\cdot b,1} (T_3)^L \rfloor T_2}$$

The transition system obtained can be easily transformed into a *Continuous Time Markov Chain*

# A Stochastic CLS model of the *lac* operon (1)



# A Stochastic CLS model of the *lac* operon (2)

Transcription of DNA, binding of lac Repressor to gene o, and interaction between lactose and lac Repressor:

$$lacI \cdot \widetilde{x} \overset{0.02}{\mapsto} lacI \cdot \widetilde{x} \mid Irna \tag{S1}$$

$$Irna \overset{0.1}{\mapsto} Irna \mid repr \tag{S2}$$

$$polym \mid \widetilde{x} \cdot lacP \cdot \widetilde{y} \overset{0.1}{\mapsto} \widetilde{x} \cdot PP \cdot \widetilde{y} \tag{S3}$$

$$\widetilde{x} \cdot PP \cdot \widetilde{y} \overset{0.01}{\mapsto} polym \mid \widetilde{x} \cdot lacP \cdot \widetilde{y} \tag{S4}$$

$$\widetilde{x} \cdot PP \cdot lacO \cdot \widetilde{y} \overset{20.0}{\mapsto} polym \mid Rna \mid \widetilde{x} \cdot lacP \cdot lacO \cdot \widetilde{y} \tag{S5}$$

$$Rna \overset{0.1}{\mapsto} Rna \mid betagal \mid perm \mid transac \tag{S6}$$

$$repr \mid \widetilde{x} \cdot lacO \cdot \widetilde{y} \overset{1.0}{\mapsto} \widetilde{x} \cdot RO \cdot \widetilde{y} \tag{S7}$$

$$\widetilde{x} \cdot RO \cdot \widetilde{y} \overset{0.01}{\mapsto} repr \mid \widetilde{x} \cdot lacO \cdot \widetilde{y} \tag{S8}$$

$$repr \mid LACT \overset{0.005}{\mapsto} RLACT \tag{S9}$$

$$RLACT \overset{0.1}{\mapsto} repr \mid LACT \tag{S10}$$

# A Stochastic CLS model of the *lac* operon (3)

The behaviour of the three enzymes for lactose degradation:

$$\left(\widetilde{x}\right)^{L} \rfloor (perm \mid X) \overset{0.1}{\mapsto} \left(perm \cdot \widetilde{x}\right)^{L} \rfloor X \qquad \text{(S11)}$$

$$LACT \mid \left(perm \cdot \widetilde{x}\right)^{L} \rfloor X \overset{0.001}{\mapsto} \left(perm \cdot \widetilde{x}\right)^{L} \rfloor (LACT \mid X) \qquad \text{(S12)}$$

$$betagal \mid LACT \overset{0.001}{\mapsto} betagal \mid GLU \mid GAL \qquad \text{(S13)}$$

Degradation of all the proteins and mRNA involved in the process:

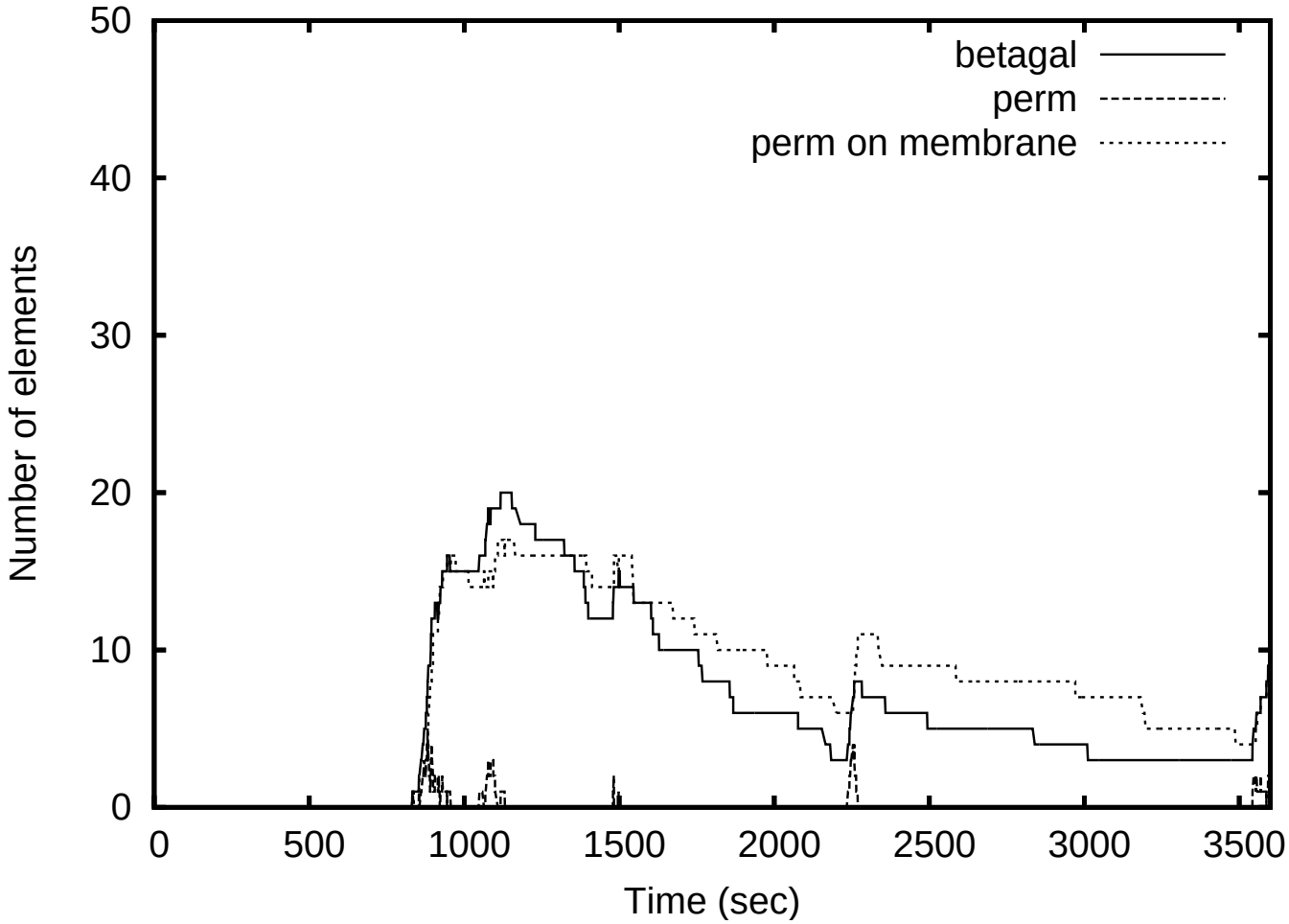
$$perm \overset{0.001}{\mapsto} \epsilon \qquad \text{(S14)} \qquad betagal \overset{0.001}{\mapsto} \epsilon \qquad \text{(S15)}$$

$$transac \overset{0.001}{\mapsto} \epsilon \qquad \text{(S16)} \qquad repr \overset{0.002}{\mapsto} \epsilon \qquad \text{(S17)}$$

$$Irna \overset{0.01}{\mapsto} \epsilon \qquad \text{(S18)} \qquad Rna \overset{0.01}{\mapsto} \epsilon \qquad \text{(S19)}$$

$$RLACT \overset{0.002}{\mapsto} LACT \qquad \text{(S20)}$$

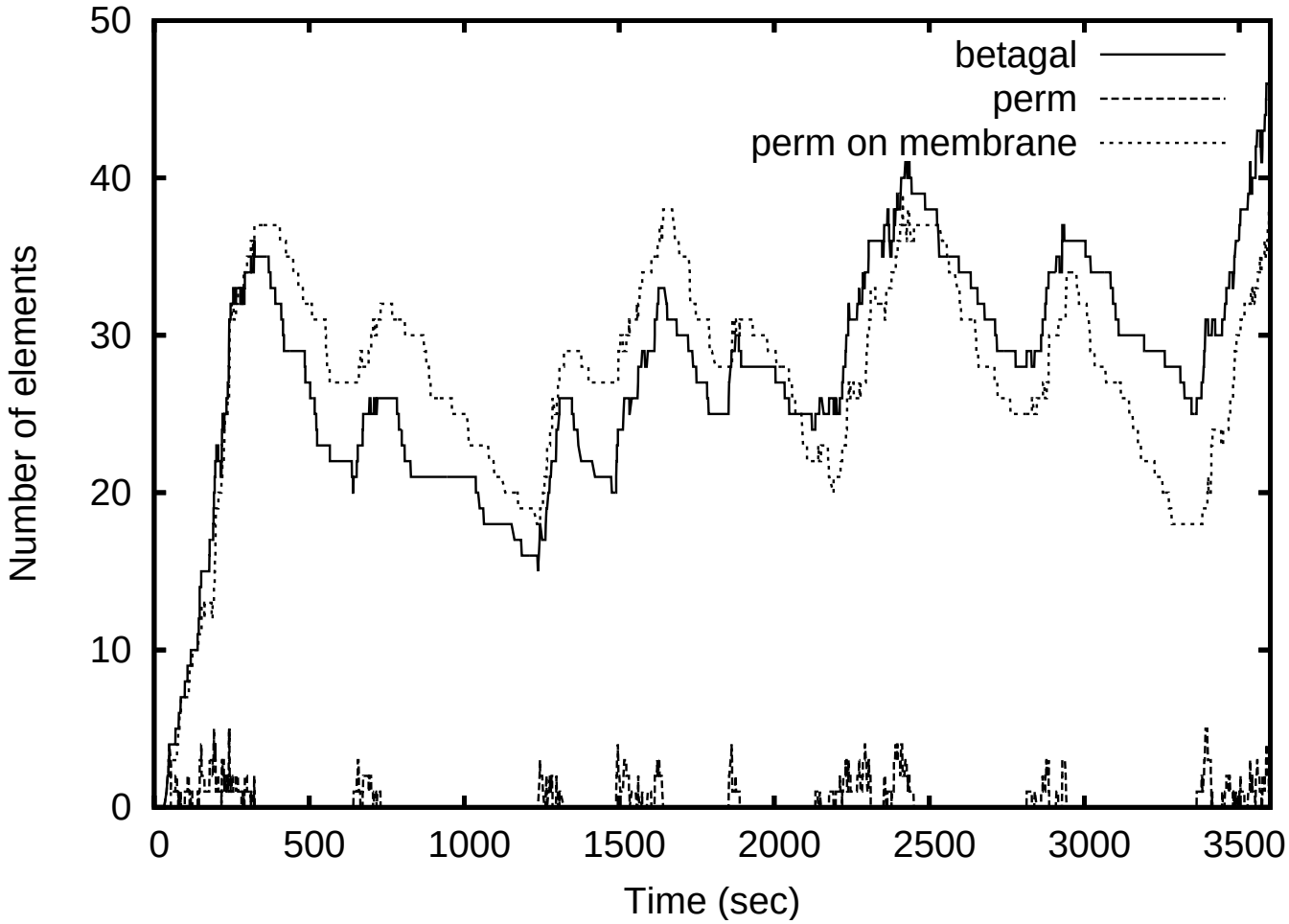
# Simulation results (1)




Production of enzymes in the absence of lactose

$$(m)^{L} \rfloor (lacI - A \mid 30 \times polym \mid 100 \times repr)$$

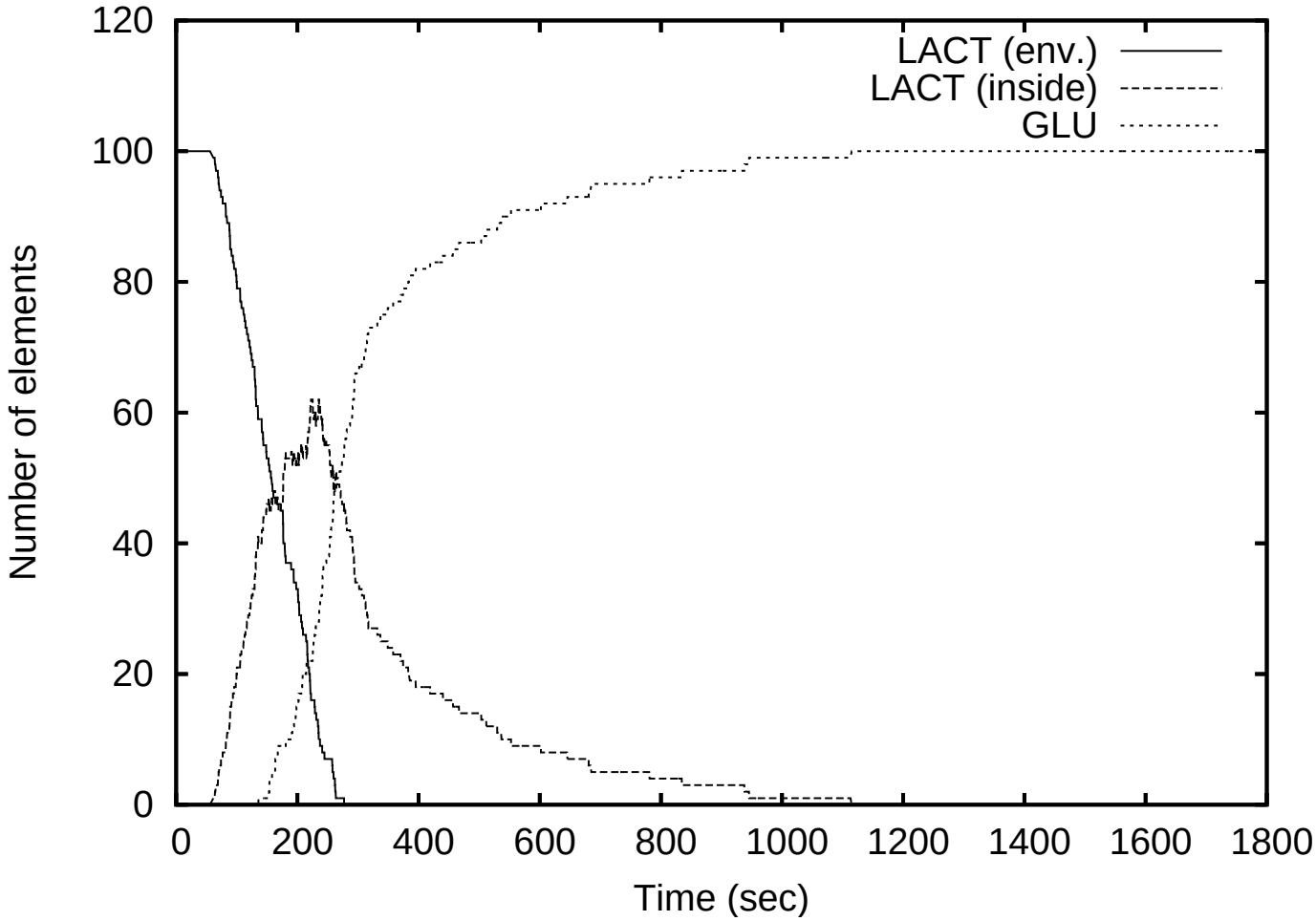
# Simulation results (2)




Production of enzymes in the presence of lactose

$$100 \times LACT \mid \left(m\right)^{L} \rfloor \left(lacI - A \mid 30 \times polym \mid 100 \times repr\right)$$

# Simulation results (3)




Degradation of lactose into glucose

$$100 \times LACT \mid \left(m\right)^{L} \rfloor \left(lacI - A \mid 30 \times polym \mid 100 \times repr\right)$$

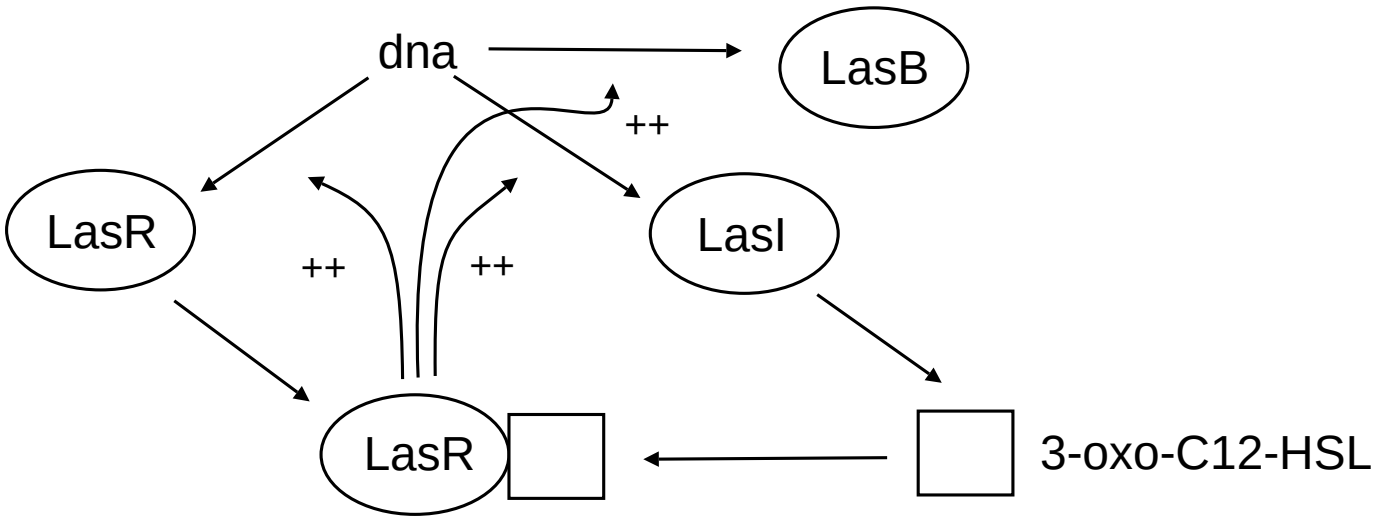
# A Stochastic CLS model of the Quorum Sensing (1)

It is recognised that many bacteria have the ability of modulating their gene expressions according to their population density. This process is called *quorum sensing*.

- A diffusible small molecules (called *autoinducers*)
- One or more transcriptional activator proteins (*R-proteins*) located within the cell
- The autoinducer can cross freely the cellular membrane
- The R-protein by itself is not active without the autoinducer. The autoinducer molecule can bind to the R-protein to form an *autoinducer/R-protein* complex.
- The *autoinducer/R-protein* complex binds to the DNA enhancing the transcription of specific genes.
- These genes regulate both the production of specific behavioural traits and the production of the autoinducer and of the R-protein.

# A Stochastic CLS model of the Quorum Sensing (2)

At low cell density, the autoinducer is synthesized at basal levels and diffuse in the environment where it is diluted. With high cell density the concentration of the autoinducer increases. Beyond a treshold the autoinducer is produced autocatalytically.
The autocatalytic production results in a dramatic increase of product concentration.



# A Stochastic CLS model of the Quorum Sensing (3)

The behaviour of a single bacterium:

$$lasO \cdot lasR \cdot lasI \overset{20}{\mapsto} lasO \cdot lasR \cdot lasI \mid LasR \qquad \text{(R1)}$$

$$lasO \cdot lasR \cdot lasI \overset{5}{\mapsto} lasO \cdot lasR \cdot lasI \mid LasI \qquad \text{(R2)}$$

$$LasI \overset{8}{\mapsto} LasI \mid 3oxo \qquad \text{(R3)}$$

$$3oxo \mid LasR \overset{0.25}{\mapsto} 3R \qquad 3R \overset{400}{\mapsto} 3oxo \mid LasR \qquad \text{(R4-R5)}$$

$$3R \mid lasO \cdot lasR \cdot lasI \overset{0.25}{\mapsto} 3RO \cdot lasR \cdot lasI \qquad \text{(R6)}$$

$$3RO \cdot lasR \cdot lasI \overset{10}{\mapsto} 3R \mid lasO \cdot lasR \cdot lasI \qquad \text{(R7)}$$

$$3RO \cdot lasR \cdot lasI \overset{1200}{\mapsto} 3RO \cdot lasR \cdot lasI \mid LasR \qquad \text{(R8)}$$

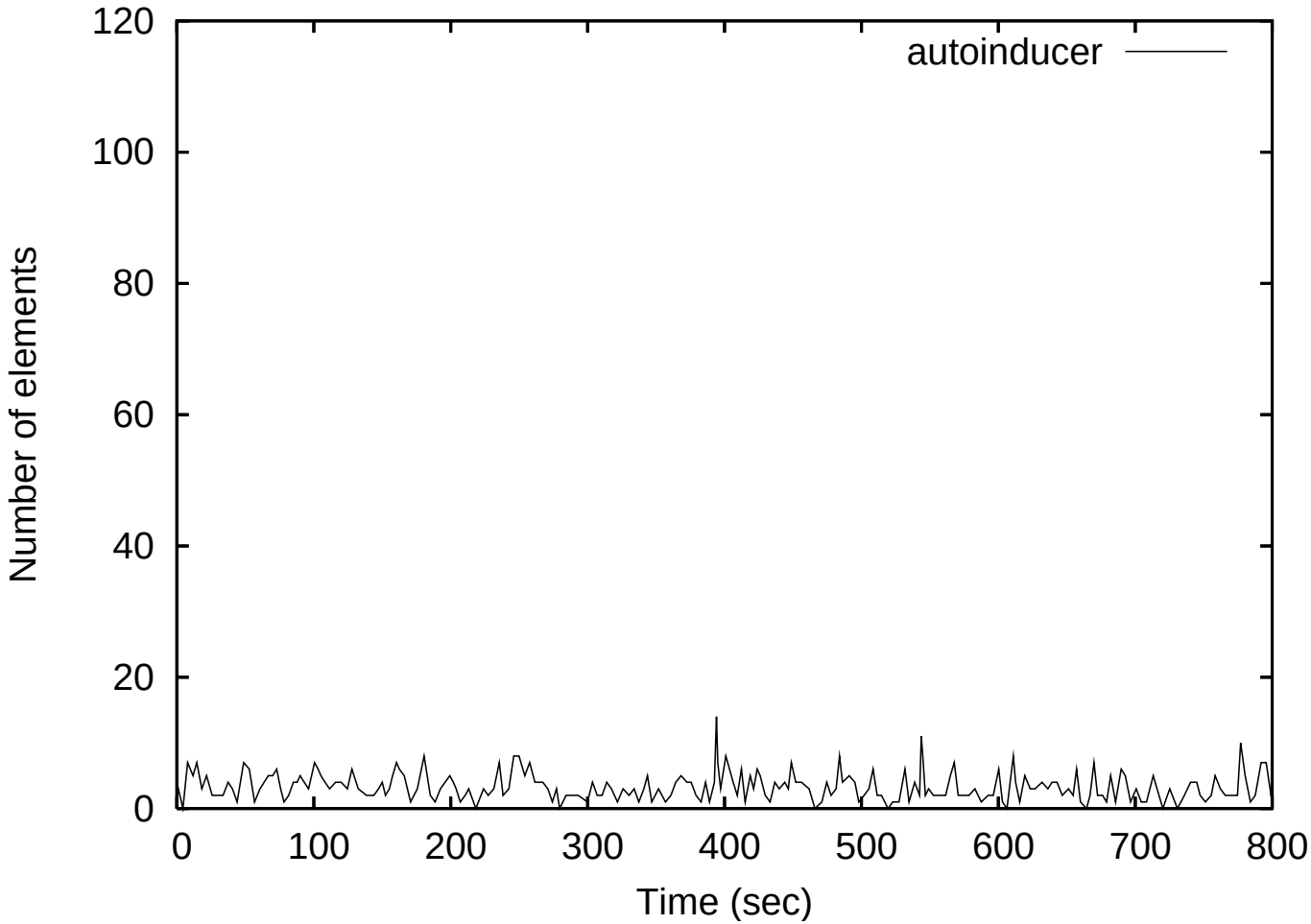
$$3RO \cdot lasR \cdot lasI \overset{300}{\mapsto} 3RO \cdot lasR \cdot lasI \mid LasI \qquad \text{(R9)}$$

$$(m)^L \rfloor (3oxo \mid X) \overset{30}{\mapsto} 3oxo \mid (m)^L \rfloor X \qquad \text{(R10)}$$

$$3oxo \mid (m)^L \rfloor X \overset{1}{\mapsto} (m)^L \rfloor (3oxo \mid X) \qquad \text{(R11)}$$

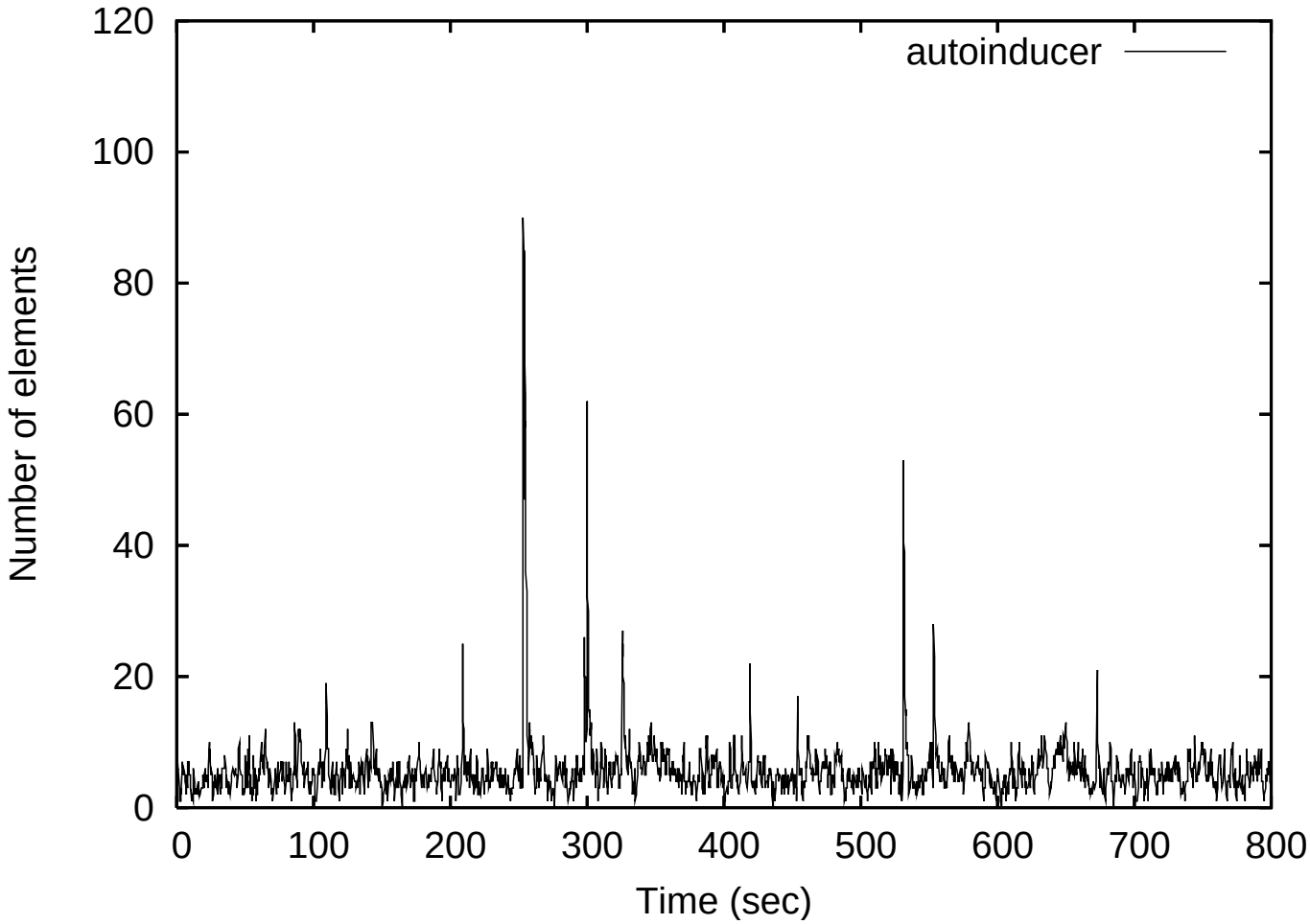
$$LasI \overset{1}{\mapsto} \epsilon \qquad LasR \overset{1}{\mapsto} \epsilon \qquad 3oxo \overset{1}{\mapsto} \epsilon \qquad \text{(R12-S14)}$$

# Simulation results (1)



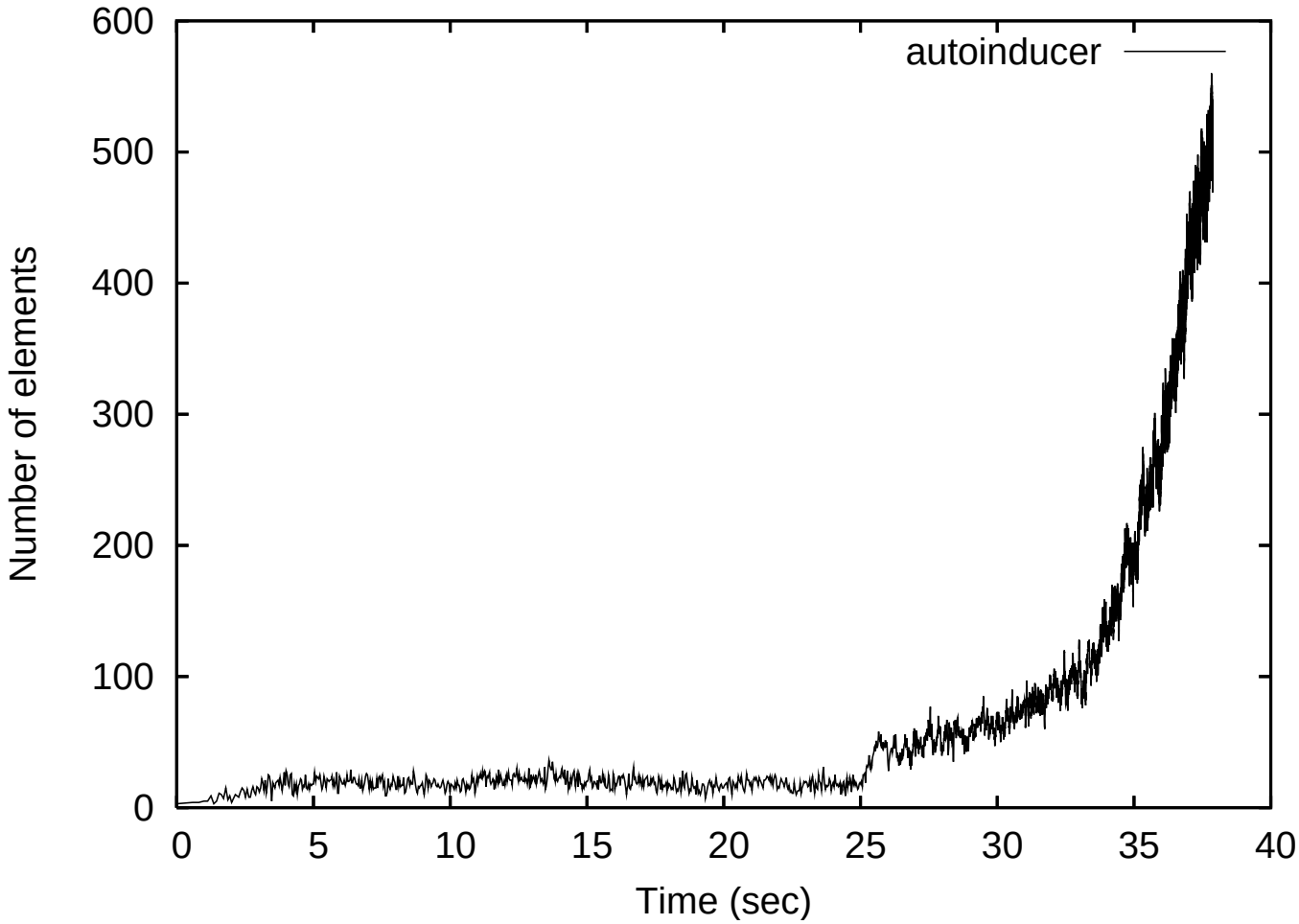

Production of the autoinducer by a single bacterium

# Simulation results (2)




Production of the autoinducer by a population of five bacteria

# Simulation results (3)




Production of the autoinducer by a population of twenty bacteria

# Outline of the talk


1. Introduction
   - Cells are complex interactive systems
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - The EGF pathway and the *lac* operon in CLS
3. CLS variants
   - Stochastic CLS
   - LCLS
   - Spatial CLS
4. Future Work and References

# Modeling proteins at the domain level

To model a protein at the domain level in CLS it would be natural to use a sequence with one symbol for each domain

The binding between two elements of two different sequences, cannot be expressed in CLS

LCLS extends CLS with labels on basic symbols

- two symbols with the same label represent domains that are bound to each other
- example: $a \cdot b^1 \cdot c \mid d \cdot e^1 \cdot f$

# Syntax of LCLS

**Terms** $T$ and **Sequences** $S$ of LCLS are given by the following grammar:

$$\begin{array}{rcl} T & ::= & S \mid (S)^L \rfloor T \mid T \mid T \\ S & ::= & \epsilon \mid a \mid a^n \mid S \cdot S \end{array}$$

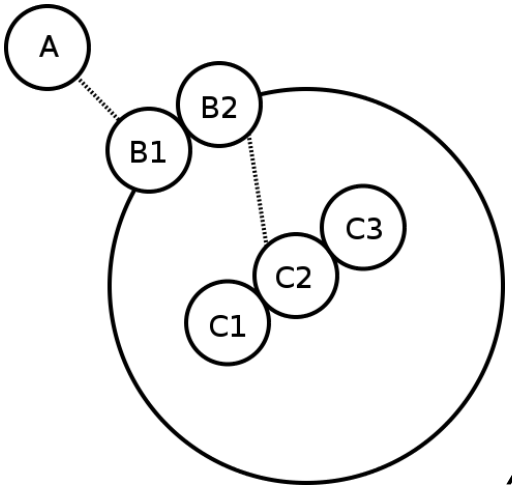
where $a$ is a generic element of $\mathcal{E}$, and $n$ is a natural number.

**Patterns** $P$ and **sequence patterns** $SP$ of LCLS are given by the following grammar:

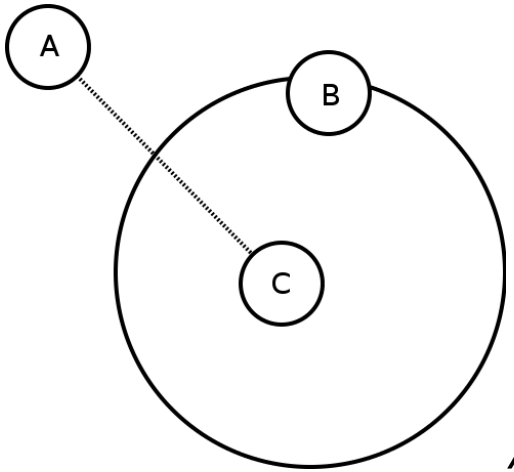
$$\begin{array}{rcl} P & ::= & SP \mid (SP)^L \rfloor P \mid P \mid P \mid X \\ SP & ::= & \epsilon \mid a \mid a^n \mid SP \cdot SP \mid \widetilde{x} \mid x \mid x^n \end{array}$$

where $a$ is an element of $\mathcal{E}$, $n$ is a natural number and $X, \widetilde{x}$ and $x$ are elements of $TV, SV$ and $\mathcal{X}$, respectively.

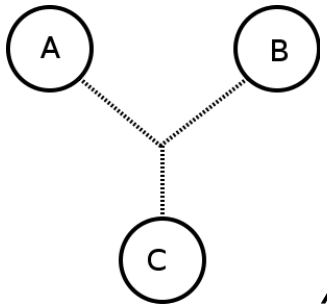
# Well–formedness of LCLS terms and patterns (1)




$$A^1 \mid \left(B1^1 \cdot B2^2\right)^L \rfloor\, C1 \cdot C2^2 \cdot C3$$ ✓




$$A^1 \mid \left(B\right)^L \rfloor\, C^1$$ ✗




$$A^1 \mid B^1 \mid C^1$$ ✗

# Well–formedness of LCLS terms and patterns (2)

An LCLS term (or pattern) is well–formed if and only if a label occurs no more than twice, and in the content of a looping sequence a label occurs either zero or two times

Type system for well–formedness:

$$1.\ (\varnothing,\varnothing) \models \epsilon \qquad 2.\ (\varnothing,\varnothing) \models a \qquad 3.\ (\varnothing,\{n\}) \models a^n$$

$$4.\ (\varnothing,\varnothing) \models x \qquad 5.\ (\varnothing,\{n\}) \models x^n \qquad 6.\ (\varnothing,\varnothing) \models \widetilde{x} \qquad 7.\ (\varnothing,\varnothing) \models X$$

$$8.\ \frac{(N_1,N_1') \models SP_1 \quad (N_2,N_2') \models SP_2 \quad N_1 \cap N_2 = N_1' \cap N_2 = N_1 \cap N_2' = \varnothing}{(N_1 \cup N_2 \cup (N_1' \cap N_2'), (N_1' \cup N_2') \setminus (N_1' \cap N_2')) \models SP_1 \cdot SP_2}$$

$$9.\ \frac{(N_1,N_1') \models P_1 \quad (N_2,N_2') \models P_2 \quad N_1 \cap N_2 = N_1' \cap N_2 = N_1 \cap N_2' = \varnothing}{(N_1 \cup N_2 \cup (N_1' \cap N_2'), (N_1' \cup N_2') \setminus (N_1' \cap N_2')) \models P_1 \mid P_2}$$

$$10.\ \frac{(N_1,N_1') \models SP \quad (N_2,N_2') \models P \quad N_1 \cap N_2 = N_1' \cap N_2 = N_1 \cap N_2' = \varnothing \quad N_2' \subseteq N_1'}{(N_1 \cup N_2', N_1' \setminus N_2') \models (SP)^L \rfloor P}$$

# Application of rewrite rules

We would like to ensure that the application of a rewrite rule to a well–formed term preserves well–formedness

- not trivial: well–formedness can be easily violated
- e.g. the rewrite rule $a \mapsto a^1$ applied to $\left(b\right)^L \rfloor a$ produces $\left(b\right)^L \rfloor a^1$

A *compartment safe* rewrite rule is such that

- it does not add/remove occurrences of variables
- it does not moves variables from one compartment (content of a looping sequence) to another one

The application of a compartment safe rewrite rule preserves well–formedness

To apply a *compartment unsafe* rewrite rule we require that

- its patterns are CLOSED
- its variables are instantiated with CLOSED terms

# The semantics of LCLS

Given a set of compartment safe rewrite rules $\mathcal{R}^{CS}$ and a set of compartemnt unsafe rewrite rules $\mathcal{R}^{CU}$, the semantics of LCLS is given by the following rules

$$\text{(appCS)} \quad \frac{P_1 \mapsto P_2 \in \mathcal{R}^{CS} \quad P_1\sigma \not\equiv \epsilon \quad \sigma \in \Sigma \quad \alpha \in \mathcal{A}}{P_1\alpha\sigma \rightarrow P_2\alpha\sigma}$$

$$\text{(appCU)} \quad \frac{P_1 \mapsto P_2 \in \mathcal{R}^{CU} \quad P_1\sigma \not\equiv \epsilon \quad \sigma \in \Sigma_{wf} \quad \alpha \in \mathcal{A}}{P_1\alpha\sigma \rightarrow P_2\alpha\sigma}$$

$$\text{(par)} \quad \frac{T_1 \rightarrow T_1' \qquad L(T_1) \cap L(T_2) = \{n_1, \ldots, n_M\} \qquad n_1', \ldots, n_M' \text{ fresh}}{T_1 \mid T_2 \rightarrow T_1'\{n_1', \ldots, n_M'/n_1, \ldots, n_M\} \mid T_2}$$

$$\text{(cont)} \quad \frac{T \rightarrow T' \qquad L(S) \cap L(T') = \{n_1, \ldots, n_M\} \qquad n_1', \ldots, n_M' \text{ fresh}}{(S)^L \rfloor T \rightarrow (S)^L \rfloor T'\{n_1', \ldots, n_M'/n_1, \ldots, n_M\}}$$

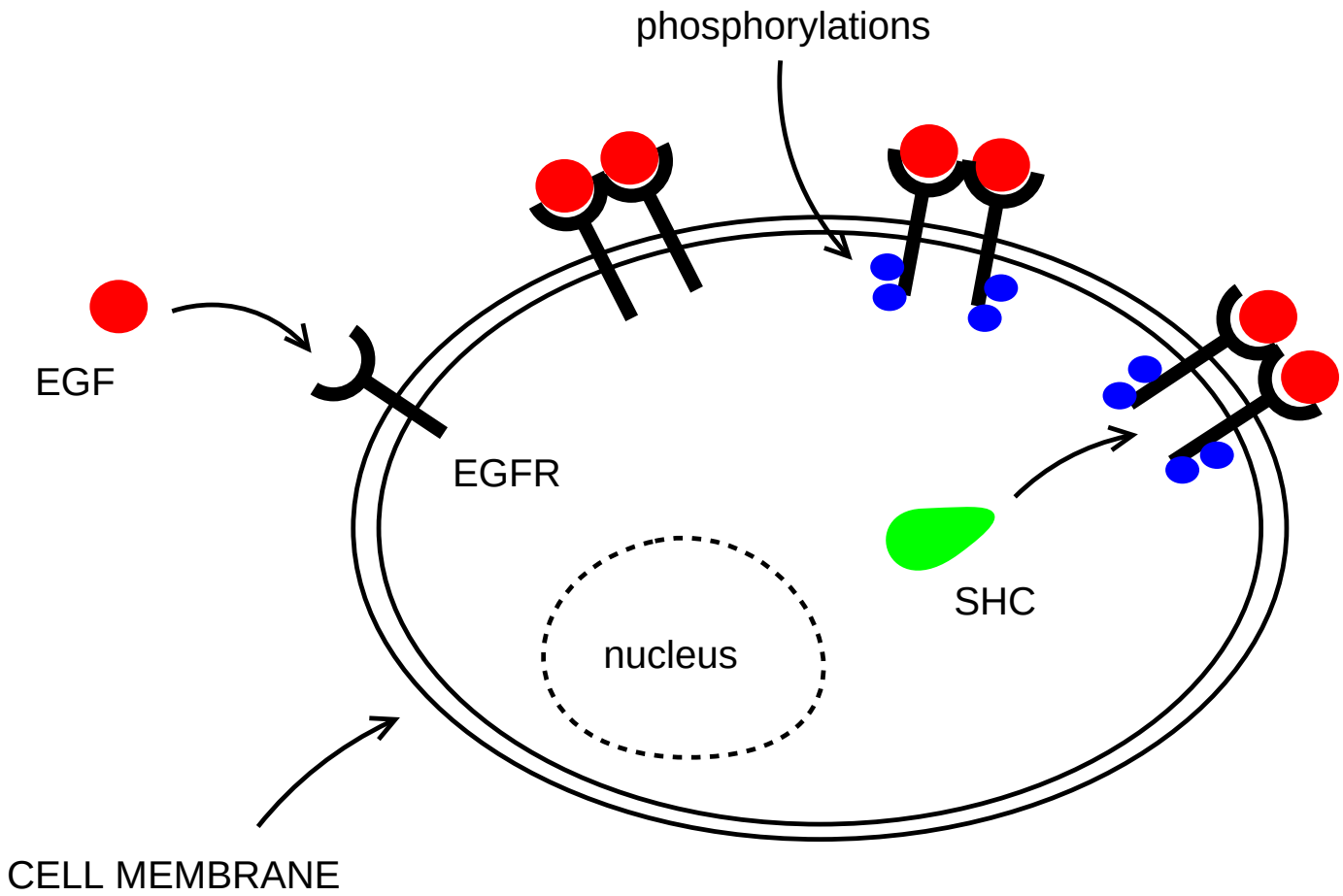
where $\alpha$ is link renaming, $L(T)$ the set of links occurring twice in the top level compartment of $T$

# Main theoretical result

**Theorem (Subject Reduction)**
Given a set of well–formed rewrite rules $\mathcal{R}$ and a well–formed term $T$

$$T \rightarrow T' \quad \Longrightarrow \quad T' \text{ well–formed}$$

# An LCLS model of the EGF pathway (1)



# An LCLS model of the EGF pathway (2)

We model the EGFR protein as the sequence $R_{E1} \cdot R_{E2} \cdot R_{I1} \cdot R_{I2}$

- $R_{E1}$ and $R_{E2}$ are two extra–cellular domains
- $R_{I1}$ and $R_{I2}$ are two intra–cellular domains

The rewrite rules of the model are

$$EGF \mid \left(R_{E1} \cdot \widetilde{x}\right)^{L} \rfloor X \;\mapsto\; EGF^{1} \mid \left(R^{1}_{E1} \cdot \widetilde{x}\right)^{L} \rfloor X \tag{R1}$$

$$\left(R^{1}_{E1} \cdot R_{E2} \cdot \widetilde{x} \cdot R^{2}_{E1} \cdot R_{E2} \cdot \widetilde{y}\right)^{L} \rfloor X \;\mapsto\; \left(R^{1}_{E1} \cdot R^{3}_{E2} \cdot \widetilde{x} \cdot R^{2}_{E1} \cdot R^{3}_{E2} \cdot \widetilde{y}\right)^{L} \rfloor X \tag{R2}$$

$$\left(R^{1}_{E2} \cdot R_{I1} \cdot \widetilde{x}\right)^{L} \rfloor X \;\mapsto\; \left(R^{1}_{E2} \cdot PR_{I1} \cdot \widetilde{x}\right)^{L} \rfloor X \tag{R3}$$

$$\left(R^{1}_{E2} \cdot PR_{I1} \cdot R_{I2} \cdot \widetilde{x} \cdot R^{1}_{E2} \cdot PR_{I1} \cdot R_{I2} \cdot \widetilde{y}\right)^{L} \rfloor (SHC \mid X) \;\mapsto$$
$$\left(R^{1}_{E2} \cdot PR_{I1} \cdot R^{2}_{I2} \cdot \widetilde{x} \cdot R^{1}_{E2} \cdot PR_{I1} \cdot R_{I2} \cdot \widetilde{y}\right)^{L} \rfloor (SHC^{2} \mid X) \tag{R4}$$

# An LCLS model of the EGF pathway (3)

Let us write $EGFR$ for $R_{E1} \cdot R_{E2} \cdot R_{I1} \cdot R_{I2}$

A possible evolution of the system is

$$
\begin{aligned}
& EGF \mid EGF \mid \left(EGFR \cdot EGFR \cdot EGFR\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R1)} \quad & EGF^{1} \mid EGF \mid \left(R^{1}_{E1} \cdot R_{E2} \cdot R_{I1} \cdot R_{I2} \cdot EGFR \cdot EGFR\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R1)} \quad & EGF^{1} \mid EGF^{2} \mid \left(R^{1}_{E1} \cdot R_{E2} \cdot R_{I1} \cdot R_{I2} \cdot EGFR \cdot R^{2}_{E1} \cdot R_{E2} \cdot R_{I1} \cdot R_{I2}\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R2)} \quad & EGF^{1} \mid EGF^{2} \mid \left(R^{1}_{E1} \cdot R^{3}_{E2} \cdot R_{I1} \cdot R_{I2} \cdot EGFR \cdot R^{2}_{E1} \cdot R^{3}_{E2} \cdot R_{I1} \cdot R_{I2}\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R3)} \quad & EGF^{1} \mid EGF^{2} \mid \left(R^{1}_{E1} \cdot R^{3}_{E2} \cdot PR_{I1} \cdot R_{I2} \cdot EGFR \cdot R^{2}_{E1} \cdot R^{3}_{E2} \cdot R_{I1} \cdot R_{I2}\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R3)} \quad & EGF^{1} \mid EGF^{2} \mid \left(R^{1}_{E1} \cdot R^{3}_{E2} \cdot PR_{I1} \cdot R_{I2} \cdot EGFR \cdot R^{2}_{E1} \cdot R^{3}_{E2} \cdot PR_{I1} \cdot R_{I2}\right)^{L} \rfloor (SHC \mid SHC) \\
\xrightarrow{(R4)} \quad & EGF^{1} \mid EGF^{2} \mid \left(R^{1}_{E1} \cdot R^{3}_{E2} \cdot PR_{I1} \cdot R^{4}_{I2} \cdot EGFR \cdot R^{2}_{E1} \cdot R^{3}_{E2} \cdot PR_{I1} \cdot R_{I2}\right)^{L} \rfloor (SHC^{4} \mid SHC)
\end{aligned}
$$

# Outline of the talk


1. Introduction
   - Cells are complex interactive systems
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - The EGF pathway and the *lac* operon in CLS
3. CLS variants
   - Stochastic CLS
   - LCLS
   - Spatial CLS
4. Future Work and References

# Spatial CLS

Spatial CLS extends the CLS with space and time:

- elements can be associated with spheres in a 2D/3D space, and each sphere represents:
  - the space occupied by the element
  - the space available inside membrane
- elements can move autonomously during the passage of time
- the applicability of reactions can be constrained to the position of the elements involved

Spatial CLS can be especially useful to describe biological processes where the behaviour depends on the exact position of the elements, in order to obtain faithful representation of their evolution.

# Syntax of Spatial CLS

**Terms** $T$, **branes** $B$ and **sequences** $S$ of Spatial CLS are defined as:

$$\begin{array}{lcl} T & ::= & \lambda \;\mid\; (S)_d \;\mid\; (B)_d^L \rfloor T \;\mid\; T \mid T \\ B & ::= & (S)_d \;\mid\; B \mid B \\ S & ::= & \epsilon \;\mid\; a \;\mid\; S \cdot S \end{array}$$

where $a$ is a generic element of the alphabet $\mathcal{E}$, $\epsilon$ is the empty sequence, and $\lambda$ is the empty term.

Parameter $d$ describes the spatial information of elements.
Elements can be:

- positional, when $d = \langle [p, m], r \rangle$
- non-positional, when $d = \langle \cdot, r \rangle$

where $p$ is the center of the sphere, $r$ the radius, and $m$ denotes a movement function.

# Representing the movement

A movement function gives new position $p'$ of the element, reached after a time interval $\delta t$ from the current time $t$:

$$p' = m_{\text{fun}}(p, r, x, l, t, \delta t)$$

where

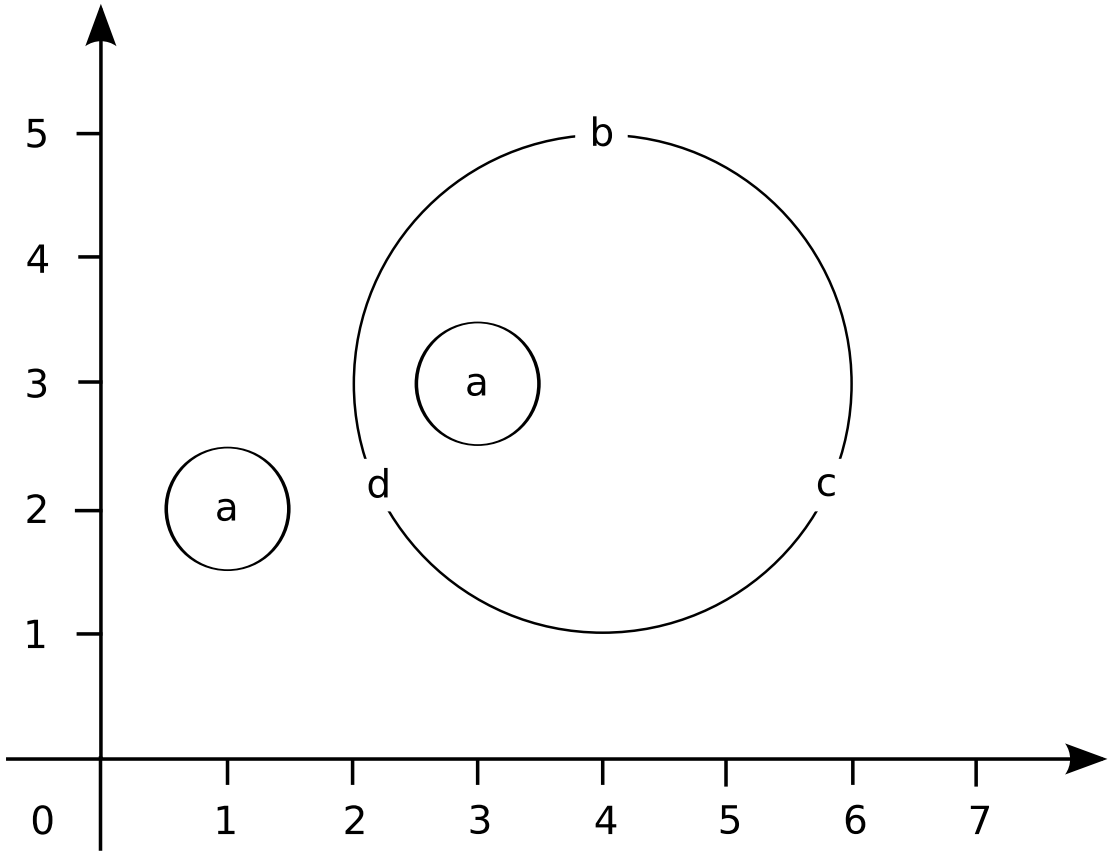
- $p$: current position of the element (at time $t$)
- $r$: radius of the element
- $x$: where the element appears: on the surface or inside a membrane
- $l$: radius of the containing membrane or $\infty$

**Example**

Linear motion: $m_{\text{fun}}(p, r, x, l, t, \delta t) = \vec{q} + \vec{v}t$

# Example of term

Term $T_1 = (a)_{[(1,2),m_1],0.5} \mid (b \cdot c \cdot d)^{L}_{[(4,3),m_2],2} \rfloor (a)_{[(-1,0),m_3],0.5}$ can be represented graphically as:



# Semantics

Rewrite rules are of the form $[\, f_c \,] \quad P_L \overset{k}{\longmapsto} P_R$ where

- $k$ is the reaction rate parameter (like in Stochastic CLS)
- $f_c$ are the application constraints, which express the applicability of the rule according to the position of the elements

A system evolves by performing a sequence of steps, where each step is composed of two phases:

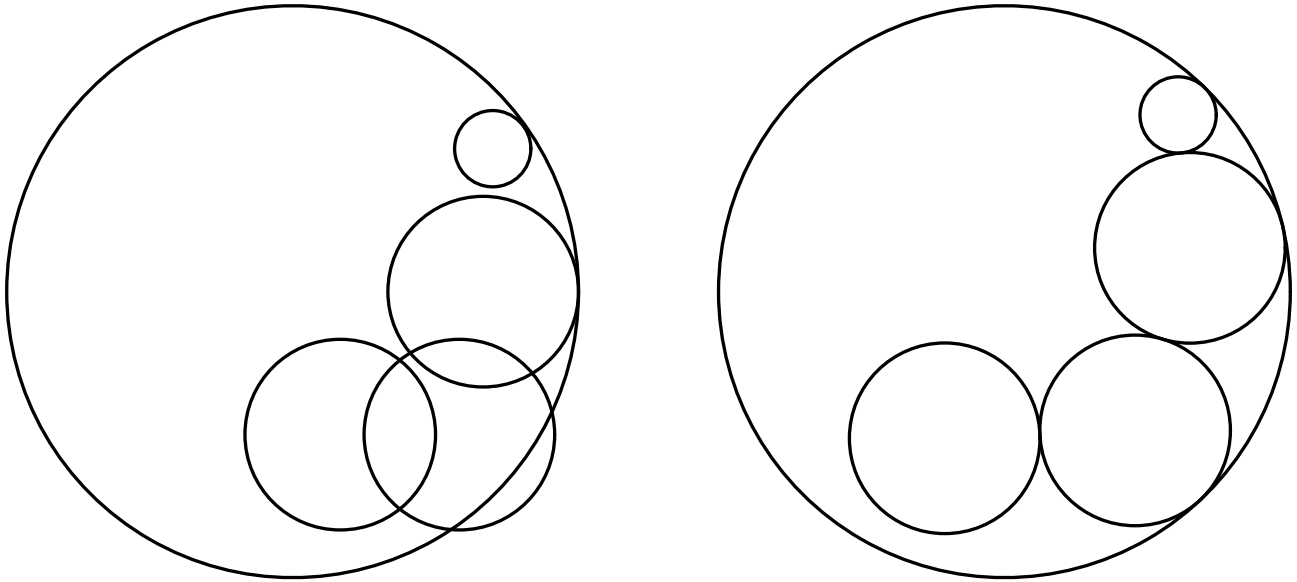
1. at most one rewrite rule is applied, among those enable in the state
2. the elements are moved according to their movement functions

*Non-positional* elements are assumed homogeneously distributed, thus the rate of reactions involving them are in accordance to the *Law of Mass Action*.

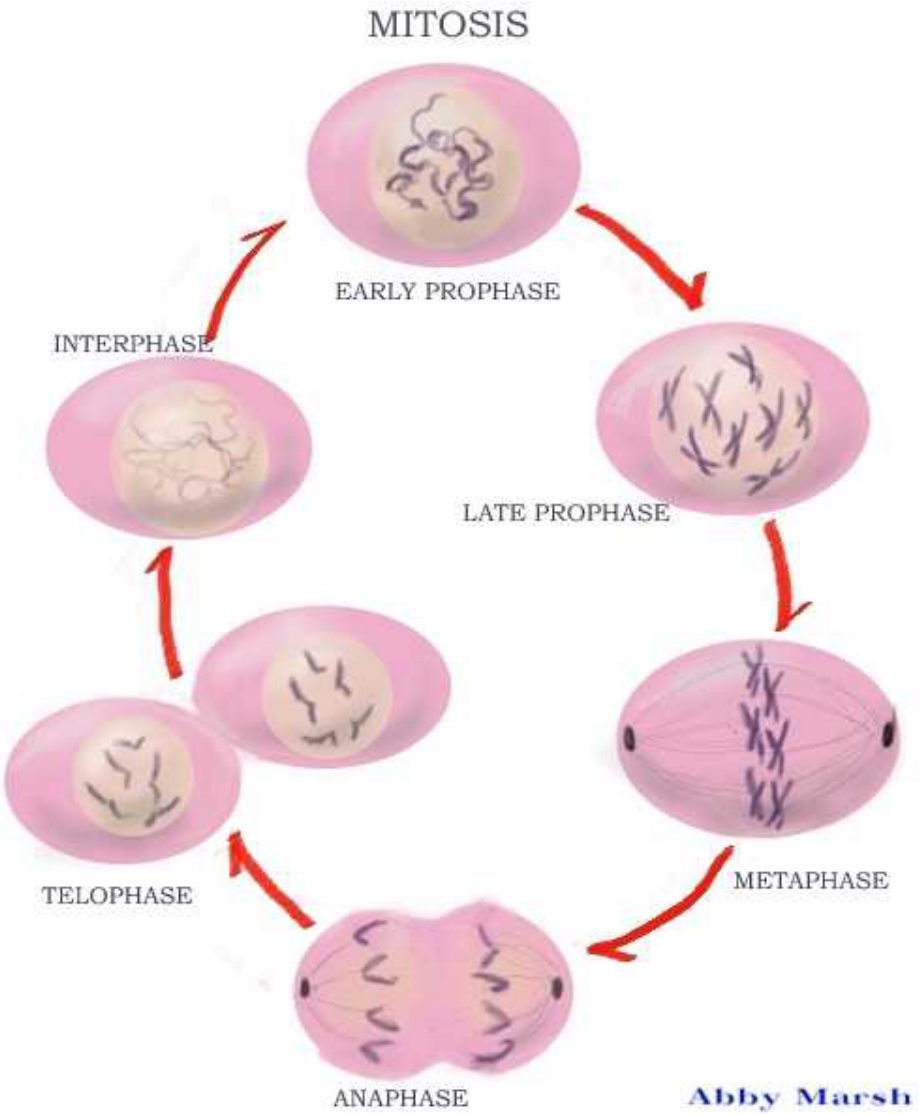
# Resolving space conflicts

During the evolution of the system, conflicts between the space occupied by different elements may arise.
In such cases, the elements are rearranged, as if they push each other.

**Example**



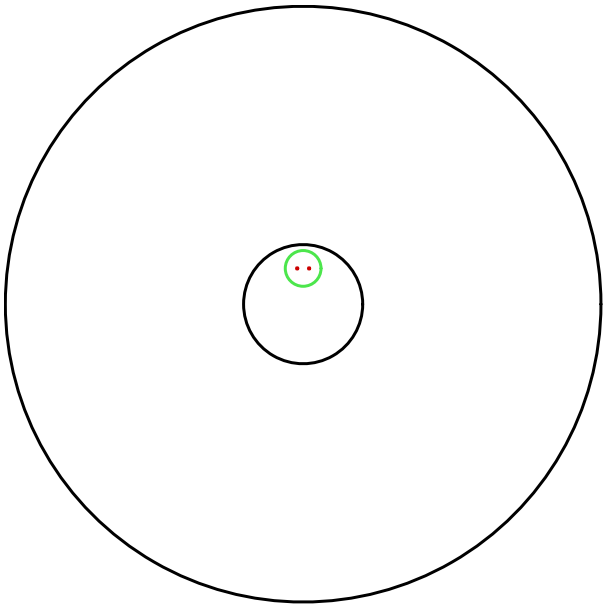
# Cell proliferation



# A model of cell proliferation

The initial state of the system is described by the following term:

$$T = (b)^{L}_{\cdot,50} \rfloor (m)^{L}_{[(0,0),m_1],10} \rfloor (n)^{L} \rfloor (cr \cdot g_1 \cdot g_2 \cdot g_3 \mid cr \cdot g_4 \cdot g_5)$$



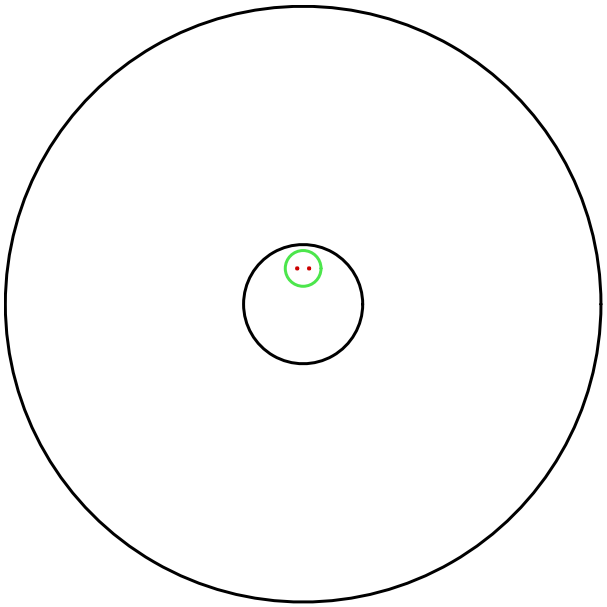
The subterms appearing in $T$ represent:

- $(b)^{L}_{\cdot,50}$: the space available
- $(m)^{L}_{[(0,0),m_1],10}$: the membrane of the cell
- $(n)^{L}$: the nucleus
- $cr \cdot \ldots$: the chromosomes

# A model of cell proliferation

The initial state of the system is described by the following term:

$$T = (b)^{L}_{\cdot,50} \rfloor (m)^{L}_{[(0,0),m_1],10} \rfloor (n)^{L} \rfloor (cr \cdot g_1 \cdot g_2 \cdot g_3 \mid cr \cdot g_4 \cdot g_5)$$



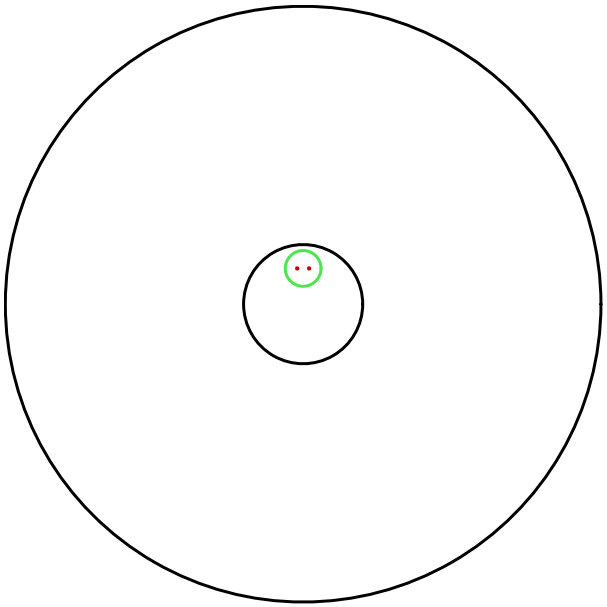
The subterms appearing in $T$ represent:

- $(b)^{L}_{\cdot,50}$: the space available
- $(m)^{L}_{[(0,0),m_1],10}$: the membrane of the cell
- $(n)^{L}$: the nucleus
- $cr \cdot \ldots$: the chromosomes

# A model of cell proliferation

The initial state of the system is described by the following term:

$$T = (b)^{L}_{\cdot,50} \rfloor (m)^{L}_{[(0,0),m_1],10} \rfloor (n)^{L} \rfloor (cr \cdot g_1 \cdot g_2 \cdot g_3 \mid cr \cdot g_4 \cdot g_5)$$



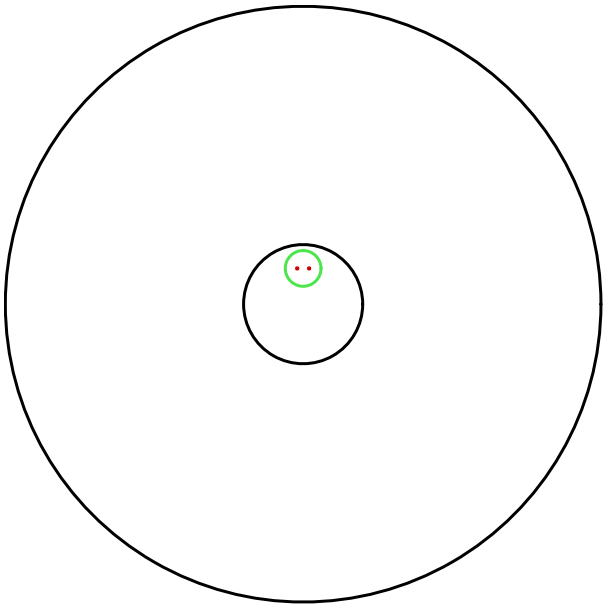
The subterms appearing in $T$ represent:

- $(b)^{L}_{\cdot,50}$: the space available
- $(m)^{L}_{[(0,0),m_1],10}$: the membrane of the cell
- $(n)^{L}$: the nucleus
- $cr \cdot \ldots$: the chromosomes

# A model of cell proliferation

The initial state of the system is described by the following term:

$$T = (b)^{L}_{\cdot,50} \rfloor (m)^{L}_{[(0,0),m_1],10} \rfloor (n)^{L} \rfloor (cr \cdot g_1 \cdot g_2 \cdot g_3 \mid cr \cdot g_4 \cdot g_5)$$



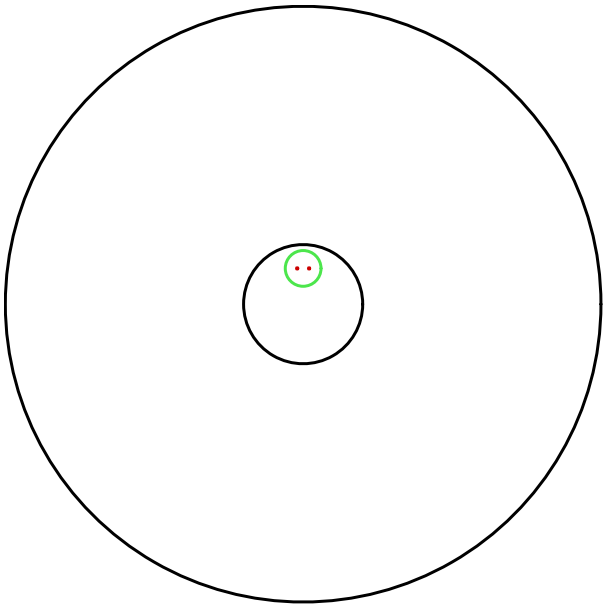
The subterms appearing in $T$ represent:

- $(b)^{L}_{\cdot,50}$: the space available
- $(m)^{L}_{[(0,0),m_1],10}$: the membrane of the cell
- $(n)^{L}$: the nucleus
- $cr \cdot \ldots$: the chromosomes

# A model of cell proliferation

The initial state of the system is described by the following term:

$$T = (b)^{L}_{\cdot,50} \rfloor (m)^{L}_{[(0,0),m_1],10} \rfloor (n)^{L} \rfloor (cr \cdot g_1 \cdot g_2 \cdot g_3 \mid cr \cdot g_4 \cdot g_5)$$



The subterms appearing in $T$ represent:

- $(b)^{L}_{\cdot,50}$: the space available
- $(m)^{L}_{[(0,0),m_1],10}$: the membrane of the cell
- $(n)^{L}$: the nucleus
- $cr \cdot \ldots$: the chromosomes

# Rewrite rules

The evolution of the system is modeled by the following rewrite rules:

$$R_1 : [\, r = 7 \,] \quad (m)^L_{[p,f],r} \rfloor X \overset{0.33}{\mapsto} (m)^L_{[p,f],10} \rfloor X$$

$$R_2 : [\, r = 10 \,] \quad (m)^L_{[p,f],r} \rfloor X \overset{0.25}{\mapsto} (m)^L_{[p,f],14} \rfloor X$$

$$R_3 : [\, r = 14 \,] \quad (m)^L_u \rfloor \left( (n)^L \rfloor X \right) \overset{0.5}{\mapsto} (m)^L_u \rfloor \left( (n_{\mathrm{dup}})^L \rfloor X \right)$$

$$R_4 : (n_{\mathrm{dup}})^L \rfloor (cr \cdot \widetilde{x} \mid X) \overset{0.125}{\mapsto} (n_{\mathrm{dup}})^L \rfloor (2cr \cdot \widetilde{x} \mid X)$$

$$R_5 : (n_{\mathrm{dup}})^L \rfloor (2cr \cdot \widetilde{x} \mid 2cr \cdot \widetilde{y}) \overset{0.17}{\mapsto} (n)^L \rfloor (cr \cdot \widetilde{x} \mid cr \cdot \widetilde{y}) \mid (n)^L \rfloor (cr \cdot \widetilde{x} \mid cr \cdot \widetilde{y})$$

$$R_6 : (m)^L_{[(x,y),f],r} \rfloor \left( (n)^L \rfloor X \mid (n)^L \rfloor Y \right) \overset{1}{\mapsto} (m)^L_{[(x-5,y),f],7} \rfloor (n)^L \rfloor X \mid (m)^L_{[(x+5,y),f],7} \rfloor (n)^L \rfloor Y$$

The rates are expressed with respect to a time unit of 1 hour.

# Outline of the talk


1. Introduction
   - Cells are complex interactive systems
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - The EGF pathway and the *lac* operon in CLS
3. CLS variants
   - Stochastic CLS
   - LCLS
   - Spatial CLS
4. Future Work and References

# Current and future work

We developed a stochastic simulator based on Stochastic CLS

We have defined an intermediate language for stochastic simulation of biological systems (sSMSR)

- high level formalisms (Stochastic CLS, $\pi$–calculus, etc...) can be translated into sSMSR
- we plan to develop analysis and verification techniques for sSMSR

We are translating Kohn's Molecular Interaction Maps (MIM) into CLS

We plan to define the evolution of CLS terms by means of operators based on MIM.

# References


- R. Barbuti, A. Maggiolo-Schettini, P. Milazzo and G. Pardini. **Spatial Calculus of Looping Sequences**. Submitted.
- R. Barbuti, A. Maggiolo-Schettini, P. Milazzo, P. Tiberi and A. Troina. **Stochastic CLS for the Modeling and Simulation of Biological Systems**. Transactions on Computational Systems Biology, in press.
- R. Barbuti, G. Caravagna, A. Maggiolo-Schettini and P. Milazzo. **An Intermediate Language for the Stochastic Simulation of Biological Systems**. Theoretical Computer Science, in press.
- R. Barbuti, A. Maggiolo-Schettini, P. Milazzo and A. Troina. **Bisimulations in Calculi Modelling Membranes**. Formal Aspects of Computing, in press.
- R. Barbuti, A. Maggiolo-Schettini, P. Milazzo and A. Troina. **The Calculus of Looping Sequences for Modeling Biological Membranes**. WMC8, LNCS 4860, 54–76, Springer, 2007.
- R. Barbuti, A. Maggiolo-Schettini and P. Milazzo. **Extending the Calculus of Looping Sequences to Model Protein Interaction at the Domain Level**. ISBRA'07, LNBI 4463, 638–649, Springer, 2006.
- R. Barbuti, A. Maggiolo-Schettini, P. Milazzo and A. Troina. **A Calculus of Looping Sequences for Modelling Microbiological Systems**. Fundamenta Informaticae, volume 72, 21–35, 2006.