The Calculus of Looping Sequences for Modeling Biological Membranes

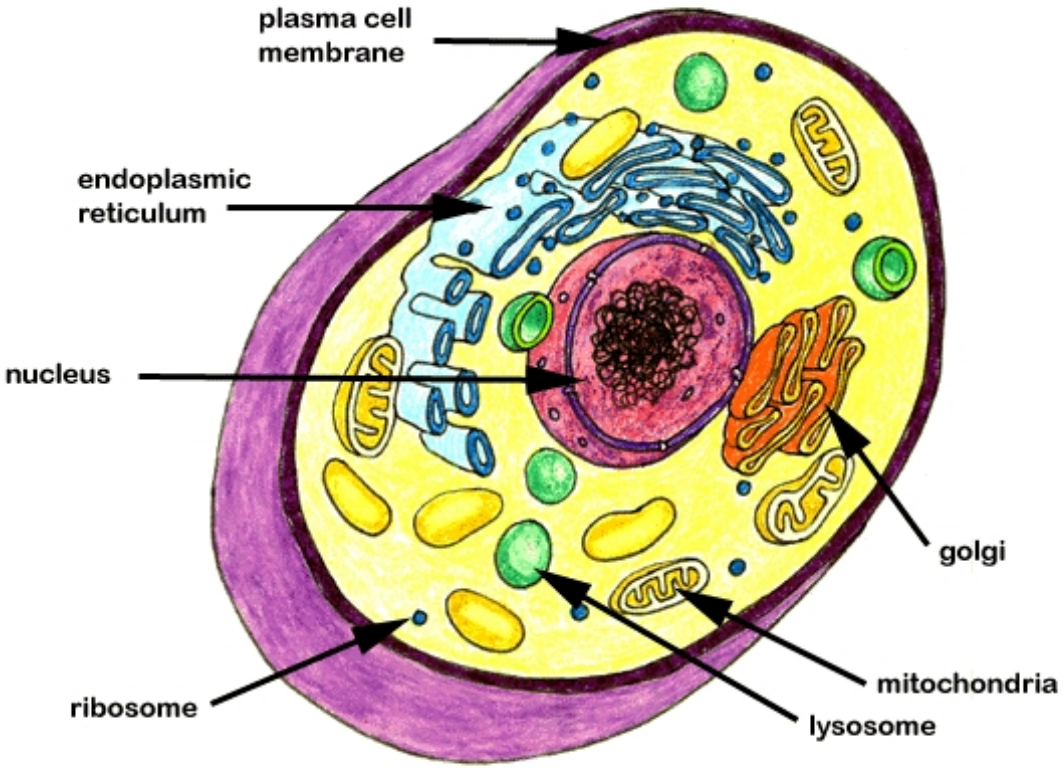
Roberto Barbuti Andrea Maggiolo–Schettini
Paolo Milazzo Angelo Troina

Dipartimento di Informatica, Università di Pisa, Italy

Thessaloniki – June, 2007

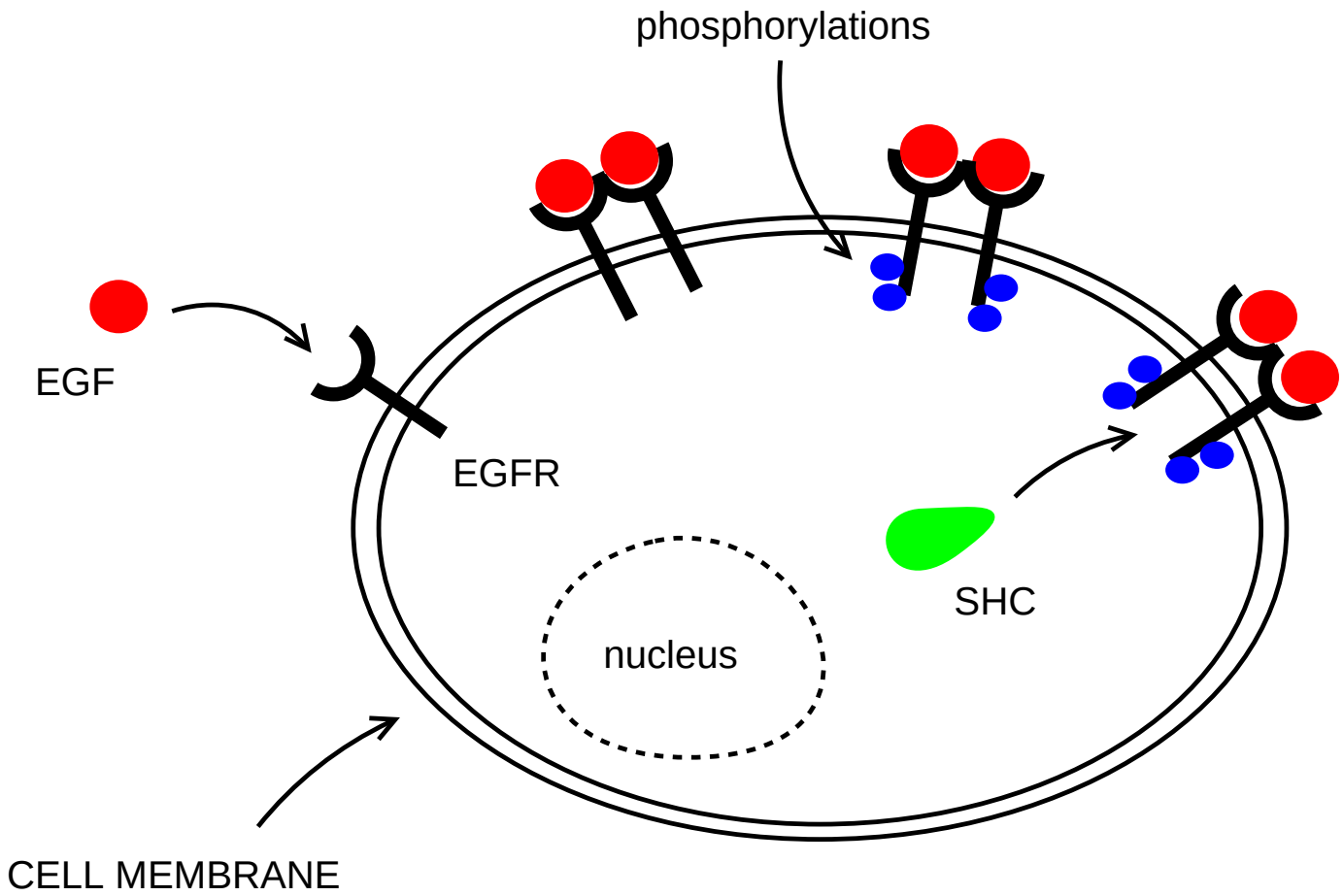
Outline of the talk

Cells: complex systems of interactive components

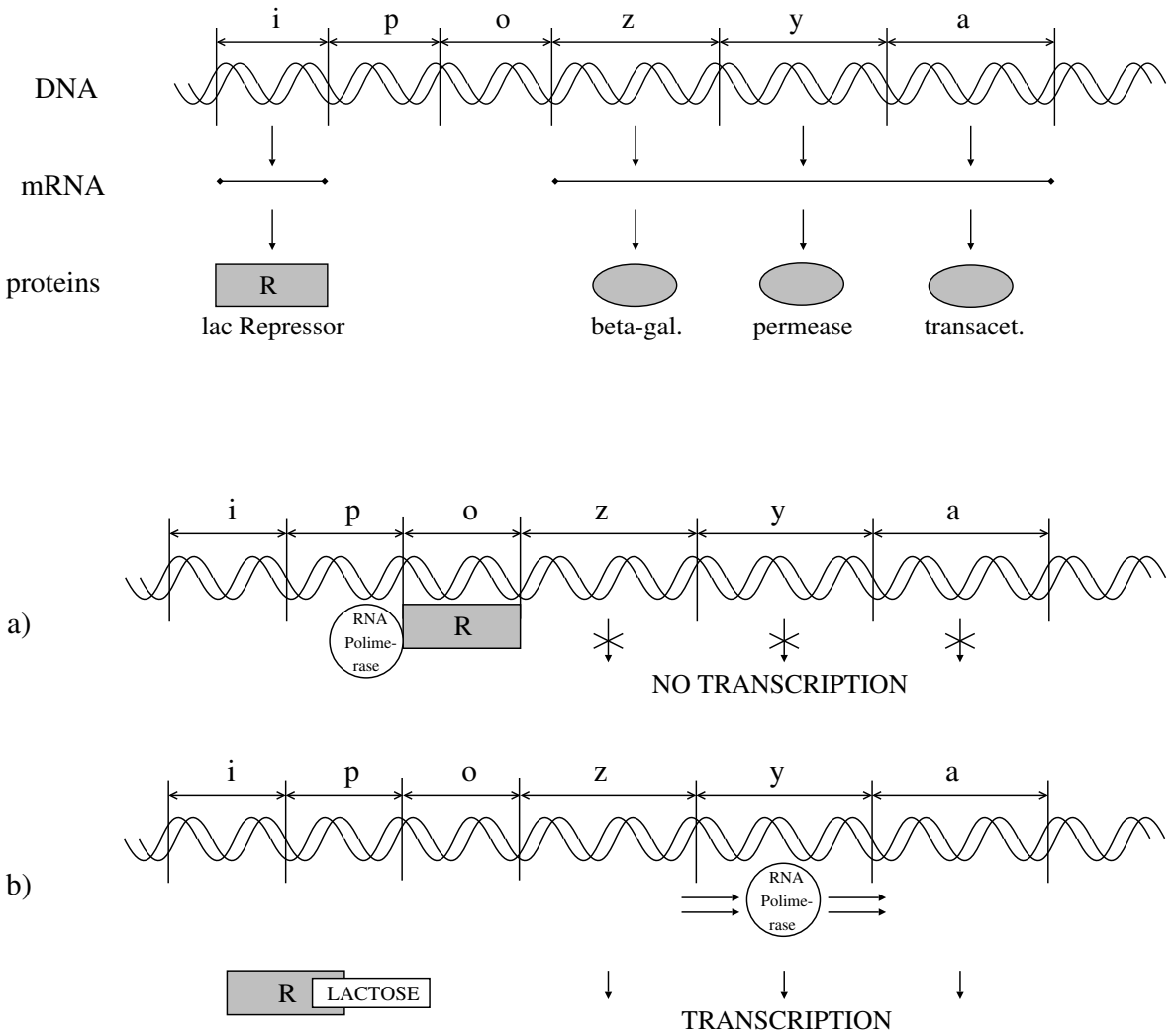


- Two classifications of cell:
  - procaryotic
  - eucaryotic
- Main actors:
  - membranes
  - proteins
  - DNA/RNA strands
  - other molecules
- Interaction networks:
  - metabolic pathways
  - signaling pathways
  - gene regulatory networks

Example of interaction networks: the EGF pathway



Example of interaction networks: the *lac* operon



Outline of the talk

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 ϵ is the empty sequence.

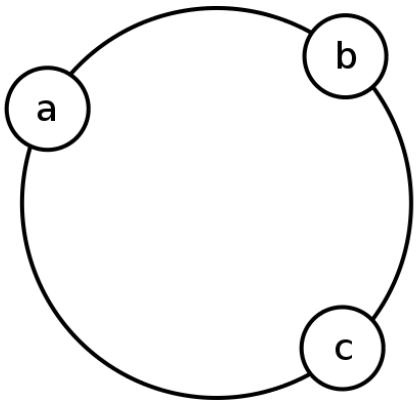
The operators are:

$S \cdot S$	:	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)

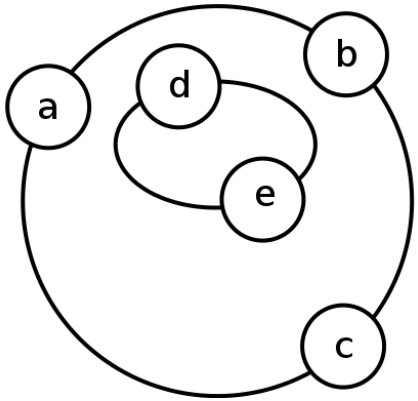
Actually, looping and containment form a single binary operator $(S)^L \rfloor T$.

Example of Terms

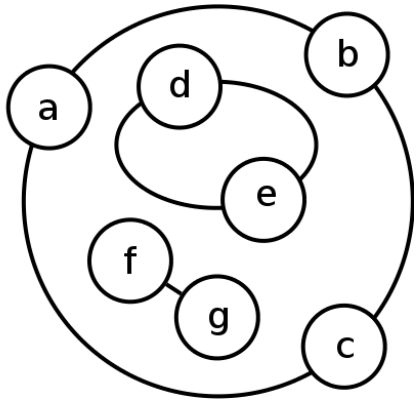
(i)



(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 $\mathcal{E}$, ϵ 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.

Σ is the set of all possible instantiations σ

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

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

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

and closed under structural congruence $\equiv$.

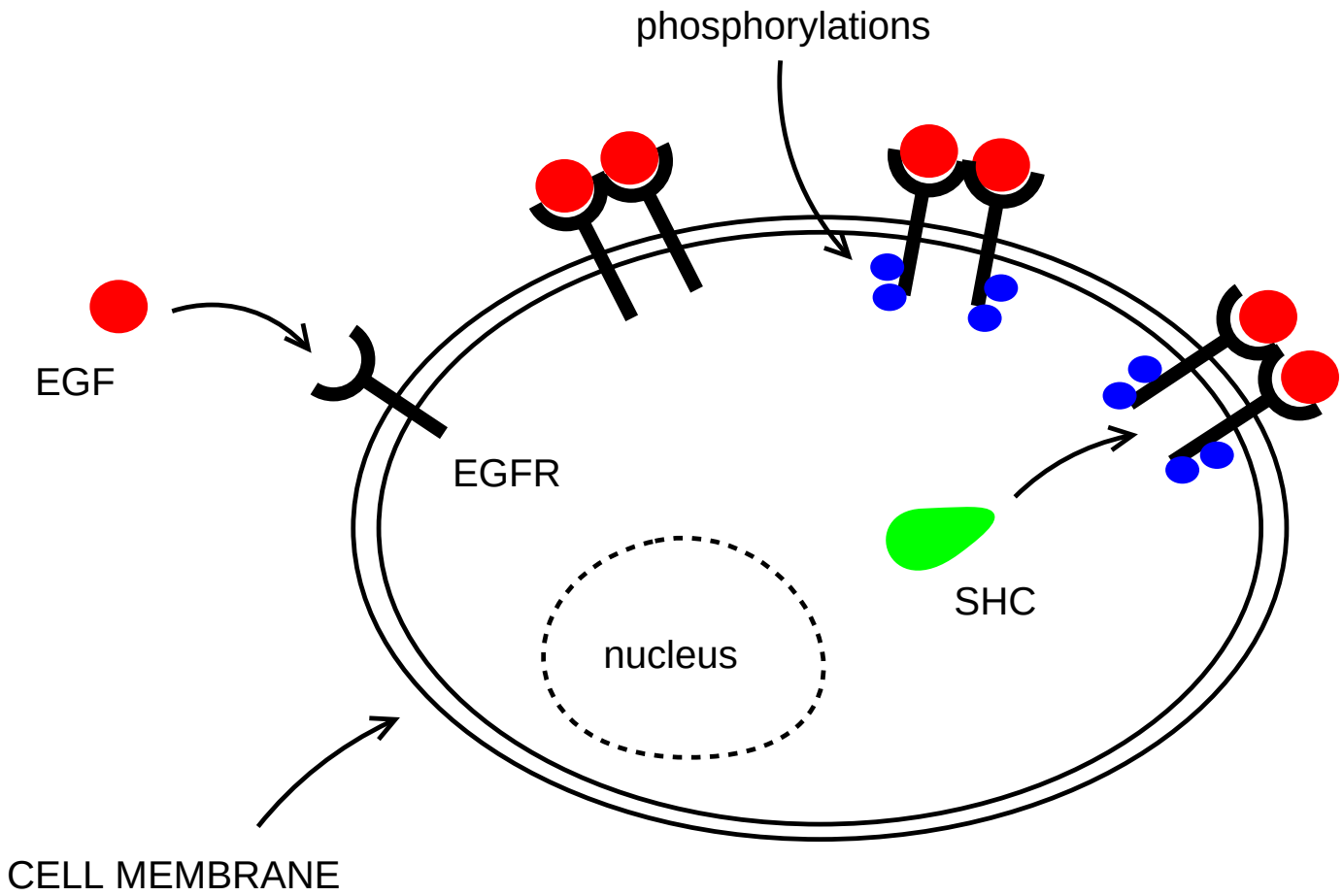
Biomolecular entities as CLS terms

Biomolecular Entity	**CLS Term**
Gene, protein domain, macro molecule, ...	Alphabet symbol
DNA strand	Sequence of elements representing genes
RNA strand	Sequence of elements representing transcribed genes
Protein	Single alphabet symbols or sequence of elements representing domains
Molecular population	Parallel composition of molecules
Membrane	Looping sequence

Biomolecular events as CLS rewrite rules

Biomolecular Event	**Examples of CLS Rewrite Rule**
Complexation	$a \mid b \mapsto c \qquad \widetilde{x} \cdot a \cdot \widetilde{y} \mid b \mapsto \widetilde{x} \cdot c \cdot \widetilde{y}$
Catalysis	$c \mid P_1 \mapsto c \mid P_2$ where $P_1 \mapsto P_2$ is the catalyzed event
Complexation on membrane	$\left(a \cdot \widetilde{x} \cdot b \cdot \widetilde{y}\right)^L \rfloor X \mapsto \left(c \cdot \widetilde{x} \cdot \widetilde{y}\right)^L \rfloor X$ $a \mid \left(b \cdot \widetilde{x}\right)^L \rfloor X \mapsto \left(c \cdot \widetilde{x}\right)^L \rfloor X$
Membrane crossing	$a \mid \left(\widetilde{x}\right)^L \rfloor X \mapsto \left(\widetilde{x}\right)^L \rfloor (a \mid X)$ $\widetilde{x} \cdot a \cdot \widetilde{y} \mid \left(\widetilde{z}\right)^L \rfloor X \mapsto \left(\widetilde{z}\right)^L \rfloor (\widetilde{x} \cdot a \cdot \widetilde{y} \mid X)$
Membrane joining	$\left(\widetilde{x}\right)^L \rfloor (a \mid X) \mapsto \left(a \cdot \widetilde{x}\right)^L \rfloor X$ $\left(\widetilde{x}\right)^L \rfloor (\widetilde{y} \cdot a \cdot \widetilde{z} \mid X) \mapsto \left(\widetilde{y} \cdot a \cdot \widetilde{z} \cdot \widetilde{x}\right)^L \rfloor X$
Catalyzed membrane fusion	$\left(a \cdot \widetilde{x}\right)^L \rfloor (X) \mid \left(b \cdot \widetilde{y}\right)^L \rfloor (Y) \mapsto$ $\left(a \cdot \widetilde{x} \cdot b \cdot \widetilde{y}\right)^L \rfloor (X \mid Y)$
Catalyzed membrane division	$\left(a \cdot \widetilde{x} \cdot b \cdot \widetilde{y}\right)^L \rfloor (X \mid Y) \mapsto$ $\left(a \cdot \widetilde{x}\right)^L \rfloor (X) \mid \left(b \cdot \widetilde{y}\right)^L \rfloor (Y)$

CLS modeling examples: the EGF pathway (1)



CLS modeling examples: the EGF pathway (2)

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:

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

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

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

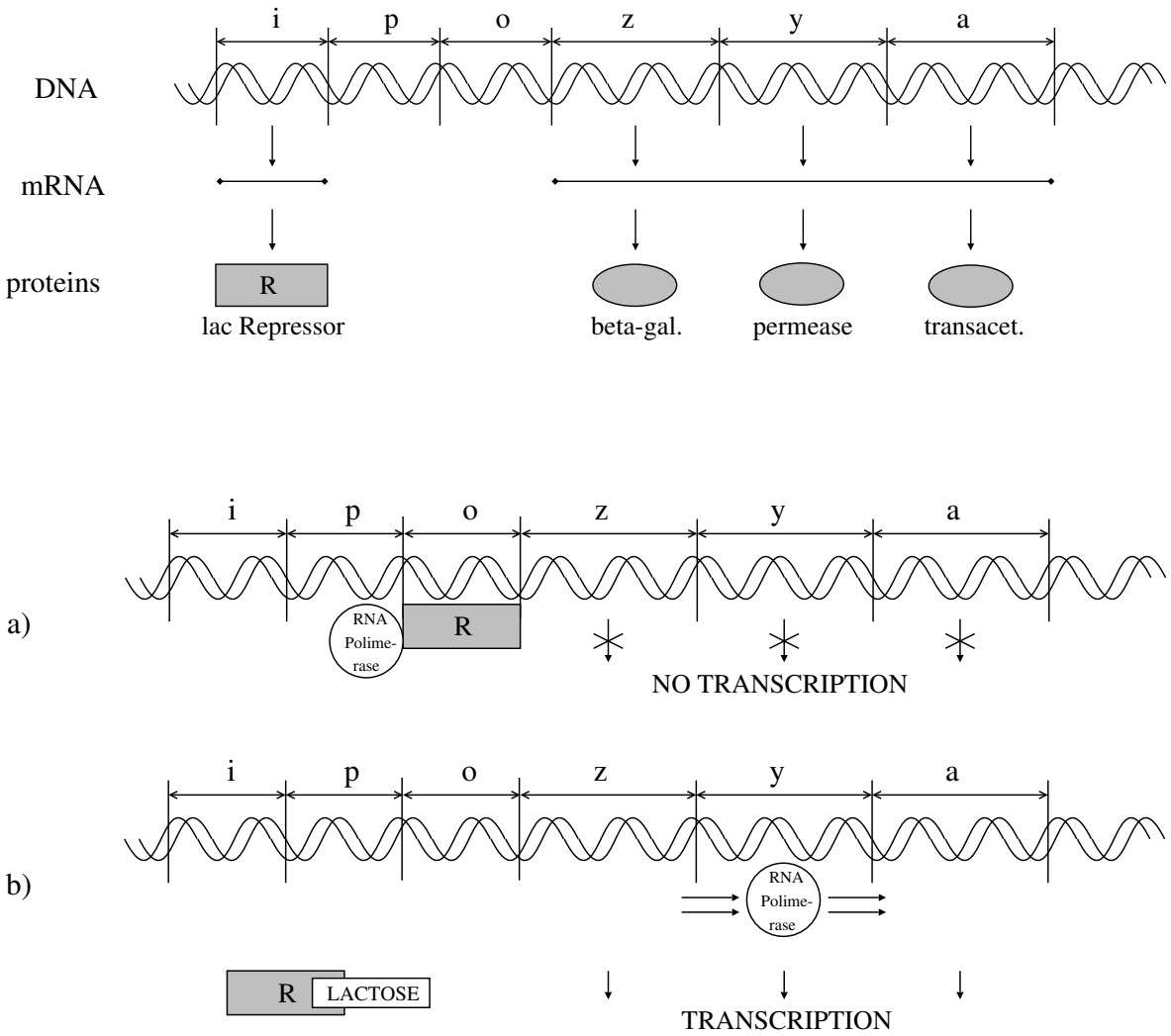
$$\left(DIMp \cdot \widetilde{x}\right)^{L} \rfloor \left(SHC \mid X\right) \;\mapsto\; \left(DIMpSHC \cdot \widetilde{x}\right)^{L} \rfloor X \qquad \text{(R4)}$$

CLS modeling examples: the EGFR pathway (2)

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: gene regulation (1)



CLS modeling examples: gene regulation (2)

Rules for DNA transcription/regulation:

$$Polym \mid p \cdot \widetilde{x} \;\mapsto Pp \cdot \widetilde{x} \tag{R1}$$

$$Repr \mid \widetilde{x} \cdot o \cdot \widetilde{y} \;\mapsto \widetilde{x} \cdot Ro \cdot \widetilde{y} \tag{R2}$$

$$Pp \cdot o \cdot \widetilde{x} \;\mapsto p \cdot Po \cdot \widetilde{x} \tag{R3}$$

$$\widetilde{x} \cdot Po \cdot zya \;\mapsto \widetilde{x} \cdot o \cdot Pzya \tag{R4}$$

$$\widetilde{x} \cdot Pzya \;\mapsto Polym \mid \widetilde{x} \cdot zya \mid rna \tag{R5}$$

CLS modeling examples: gene regulation (3)

The only possible evolution of a term representing an initial situation in which no repressor is present is

$$Polym \mid p \cdot o \cdot zya \xrightarrow{(R1)} Pp \cdot o \cdot zya \xrightarrow{(R3)} p \cdot Po \cdot zya$$
$$\xrightarrow{(R4)} p \cdot o \cdot Pzya \xrightarrow{(R5)} Polym \mid p \cdot o \cdot zya \mid rna$$

When the repressor is present, instead, a possible evolution is

$$Repr \mid Polym \mid p \cdot o \cdot zya \xrightarrow{(R1)} Repr \mid Pp \cdot o \cdot zya \xrightarrow{(R2)} Pp \cdot Ro \cdot zya$$

and it corresponds to a situation in which the repressor stops the transcription of the gene by hampering the activity of the RNA polymerase.

Some theoretical results

CLS is Turing complete

- A Turing machine encoded into a CLS term and a single rewrite rule

Formalisms capable of describing membranes can be encoded into CLS

- Brane Calculi
- P Systems

A labeled semantics of CLS is given

- Strong and weak bisimulations are defined and proved to be congruences for CLS terms

Some variants of CLS

- CLS+
  - More realistic representation of the fluid nature of membranes: the looping operator can be applied to a parallel composition 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

Outline of the talk

CLS+

The fluid nature of a membrane can be expressed representing the objects on its surface as a multiset

In CLS this corresponds to allowing the application of the looping operator to a parallel composition of sequences

The Syntax of CLS+

Terms T, **Branes** B and **Sequences** S of CLS+ are given by the following grammar:

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

where a is a generic element of $\mathcal{E}$.

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

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

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

The semantics of CLS+

The semantics of CLS+ ensures that on membrane we may have multisets of objects which cannot be membranes themselves

Given a set of rewrite rules $\mathcal{R} \subseteq \Re$ the *semantics* of CLS is given by the following rules

$$\frac{(P_1, P_2) \in \mathcal{R} \quad P_1\sigma \not\equiv \epsilon \quad \sigma \in \Sigma}{P_1\sigma \longrightarrow P_2\sigma}$$

$$\frac{T_1 \longrightarrow T_2}{T \mid T_1 \longrightarrow T \mid T_2} \qquad \frac{T_1 \longrightarrow T_2}{(B)^L \rfloor T_1 \longrightarrow (B)^L \rfloor T_2}$$

$$\frac{(BP_1, BP_2) \in \mathcal{R}_{\mathcal{B}} \quad BP_1\sigma \not\equiv \epsilon \quad \sigma \in \Sigma}{BP_1\sigma \longrightarrow_{\mathcal{B}} BP_2\sigma}$$

$$\frac{B_1 \longrightarrow_{\mathcal{B}} B_2}{B \mid B_1 \longrightarrow_{\mathcal{B}} B \mid B_2} \qquad \frac{B_1 \longrightarrow_{\mathcal{B}} B_2}{(B_1)^L \rfloor T \longrightarrow (B_2)^L \rfloor T}$$

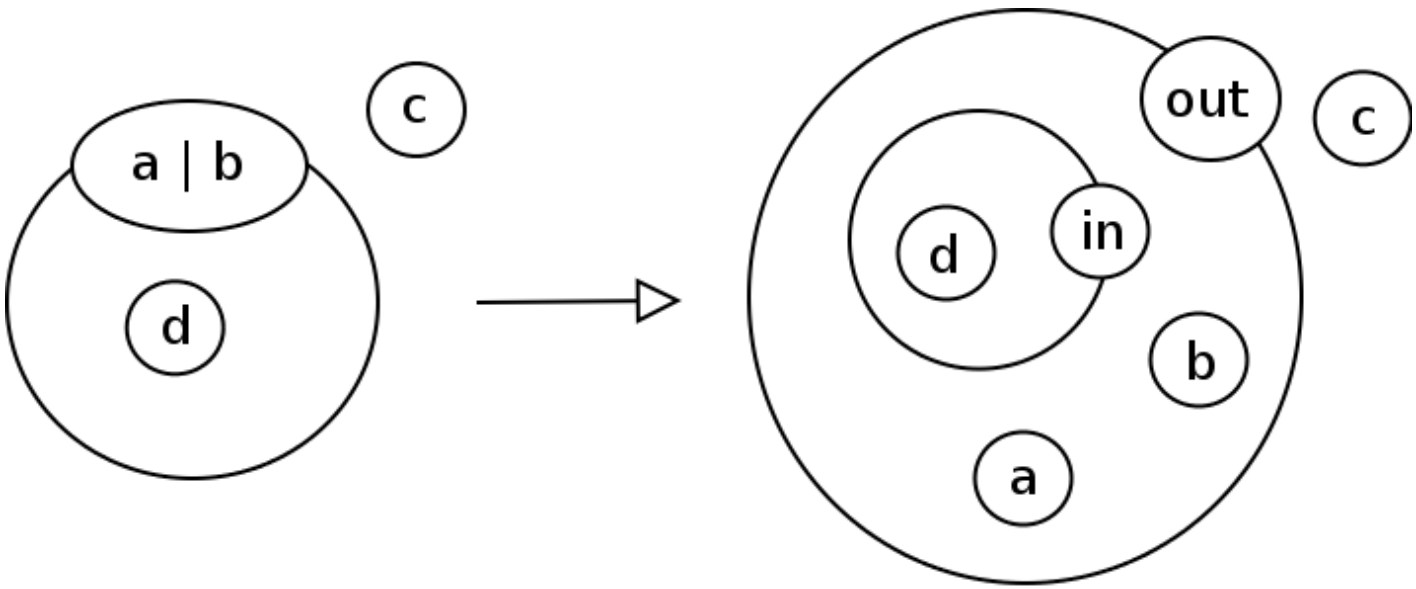
Translation of CLS+ into CLS

The CLS+ term

$$c \mid (a \mid b)^{L} \rfloor d$$

becomes the CLS term

$$c \mid (out)^{L} \rfloor (a \mid b \mid (in)^{L} \rfloor d)$$



Outline of the talk

Stochastic CLS (1)

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

Rewrite rules are extended with kinetic constants (e.g. $a \mid b \overset{k}{\mapsto} c$) to be multiplied by the number of reactants

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 occurrence of a subterm
- What happens with variables?
  - At each step we compute the set of ground rules that can be applied among those obtained by instantiating variables of the rewrite rule
  - We reduce the problem of defining the semantics to the simpler problem of defining the semantics with ground rules only

Stochastic CLS (2)

Given a finite set of rewrite rules $\mathcal{R}$, the semantics of Stochastic CLS is given by the following inference rule

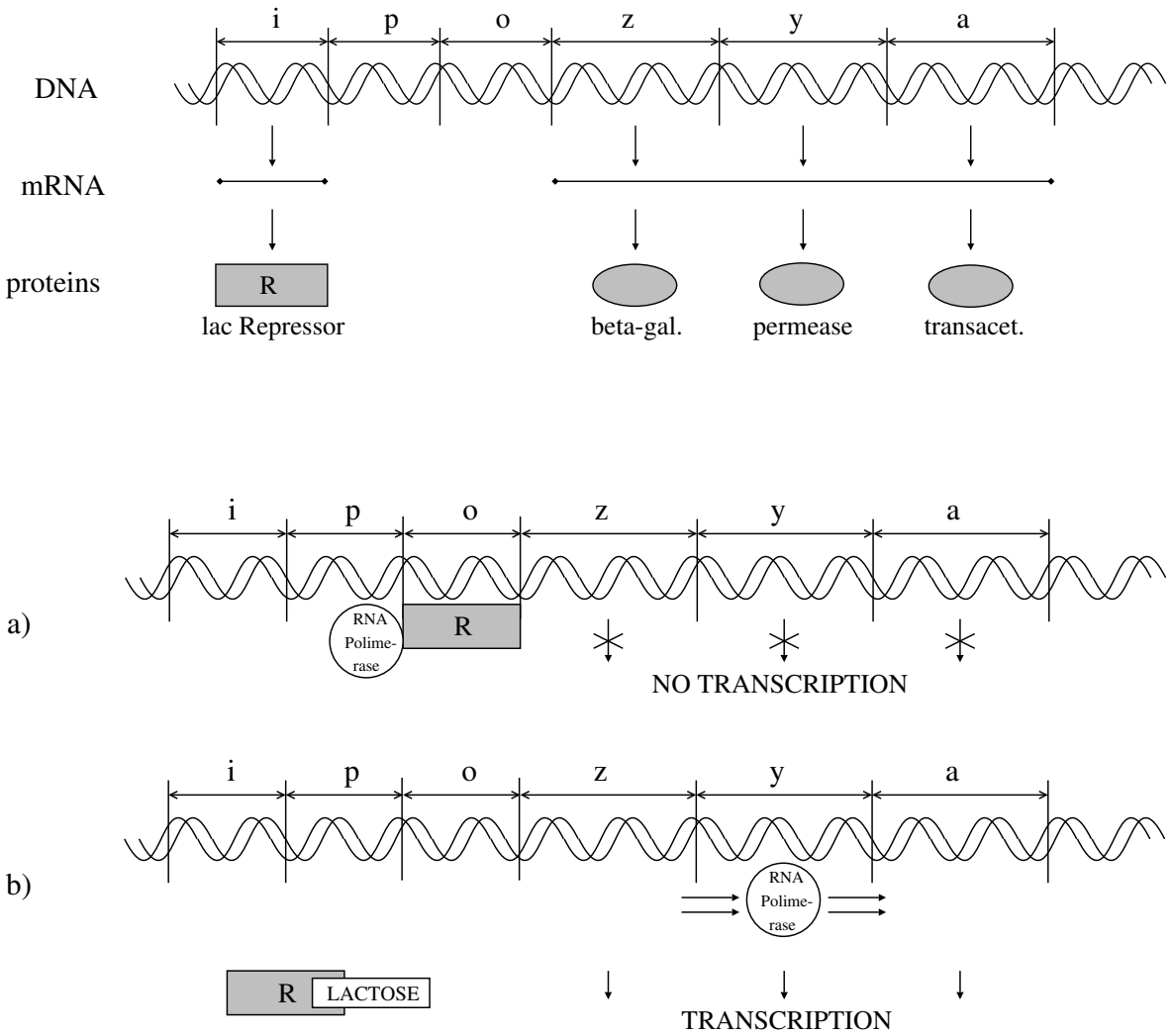
$$\frac{R = T_1 \overset{k}{\mapsto} T_2 \in AR(\mathcal{R}, T) \qquad T \equiv C[T_1] \qquad T' \equiv C[T_2]}{T \xrightarrow{R, k \cdot AC(R,T,T')} T'}$$

where:

- $AR(\mathcal{R}, T)$ is the set of ground rewrite rules obtained from rules in $\mathcal{R}$ and applicable to T
- $AC(R, T, T')$ is the number of reactants in T equivalent to the left–hand side of the ground rule R and that allows obtaining term T' after the application of R

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

A Stochastic CLS model of the gene regulation (1)



A Stochastic CLS model of the gene regulation (2)

The Stochastic CLS model extends the CLS model by including

- a kinetic constant in each rewrite rule
- two rewrite rules describing the unbinding of the RNA polymerase and of the repressor from the DNA (to obtain a more realistic model).

$$Polym \mid p \cdot \widetilde{x} \xrightarrow{0.1} Pp \cdot \widetilde{x} \qquad (R1)$$

$$Pp \cdot \widetilde{x} \xrightarrow{2} Polym \mid p \cdot \widetilde{x} \qquad (R1')$$

$$Repr \mid \widetilde{x} \cdot o \cdot \widetilde{y} \xrightarrow{1} \widetilde{x} \cdot Ro \cdot \widetilde{y} \qquad (R2)$$

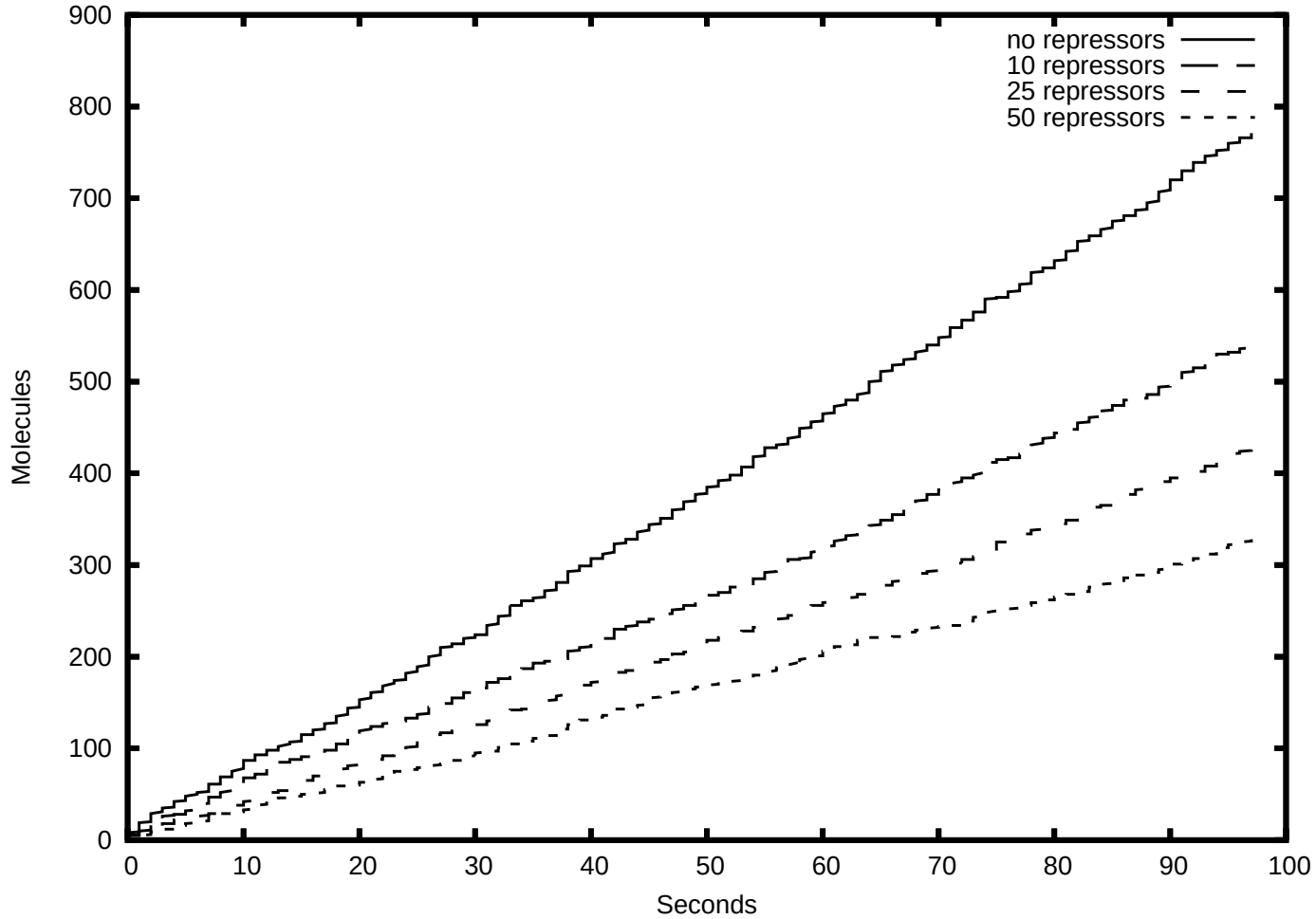
$$\widetilde{x} \cdot Ro \cdot \widetilde{y} \xrightarrow{10} Repr \mid \widetilde{x} \cdot o \cdot \widetilde{y} \qquad (R2')$$

$$Pp \cdot o \cdot \widetilde{x} \xrightarrow{100} p \cdot Po \cdot \widetilde{x} \qquad (R3)$$

$$\widetilde{x} \cdot Po \cdot zya \xrightarrow{100} \widetilde{x} \cdot o \cdot Pzya \qquad (R4)$$

$$\widetilde{x} \cdot Pg \xrightarrow{30} Polym \mid rna \mid \widetilde{x} \cdot zya \qquad (R5)$$

Simulation results



RNA produced over time by varying the number of repressors in the system

Outline of the talk

Modeling proteins at the domain level

In proteins there are places (domains) where bindings to other molecules can occur

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:

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

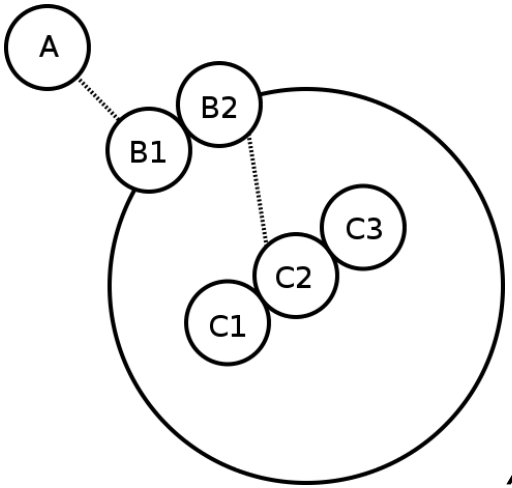
where a is a generic element of $\mathcal{E}$, and n is a natural number.

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

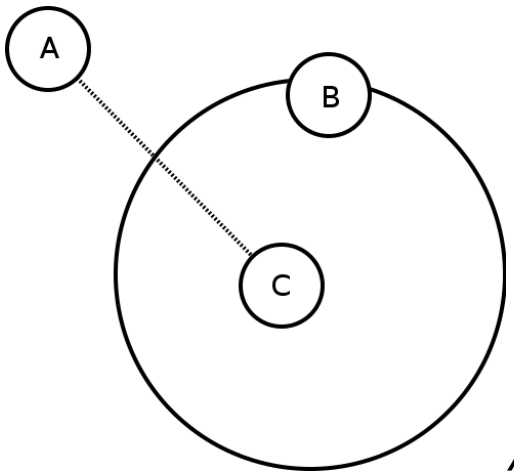
$$\begin{array}{rcl} P & ::= & SP \mid (SP)^L \rfloor P \mid P \mid P \mid X \\ SP & ::= & \epsilon \mid a \mid a^n \mid SP \cdot SP \mid \widetilde{x} \mid x \mid x^n \end{array}$$

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.

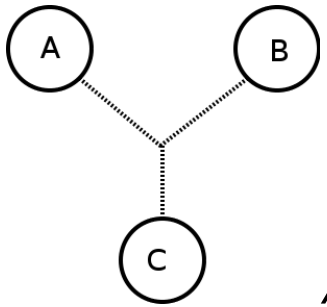
Well–formed LCLS terms and patterns



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



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



$$A^1 \mid B^1 \mid C^1$$ ✗

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 and single labels
- 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 compartment unsafe rewrite rules $\mathcal{R}^{CU}$, the semantics of LCLS is given by the following rules

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

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

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

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

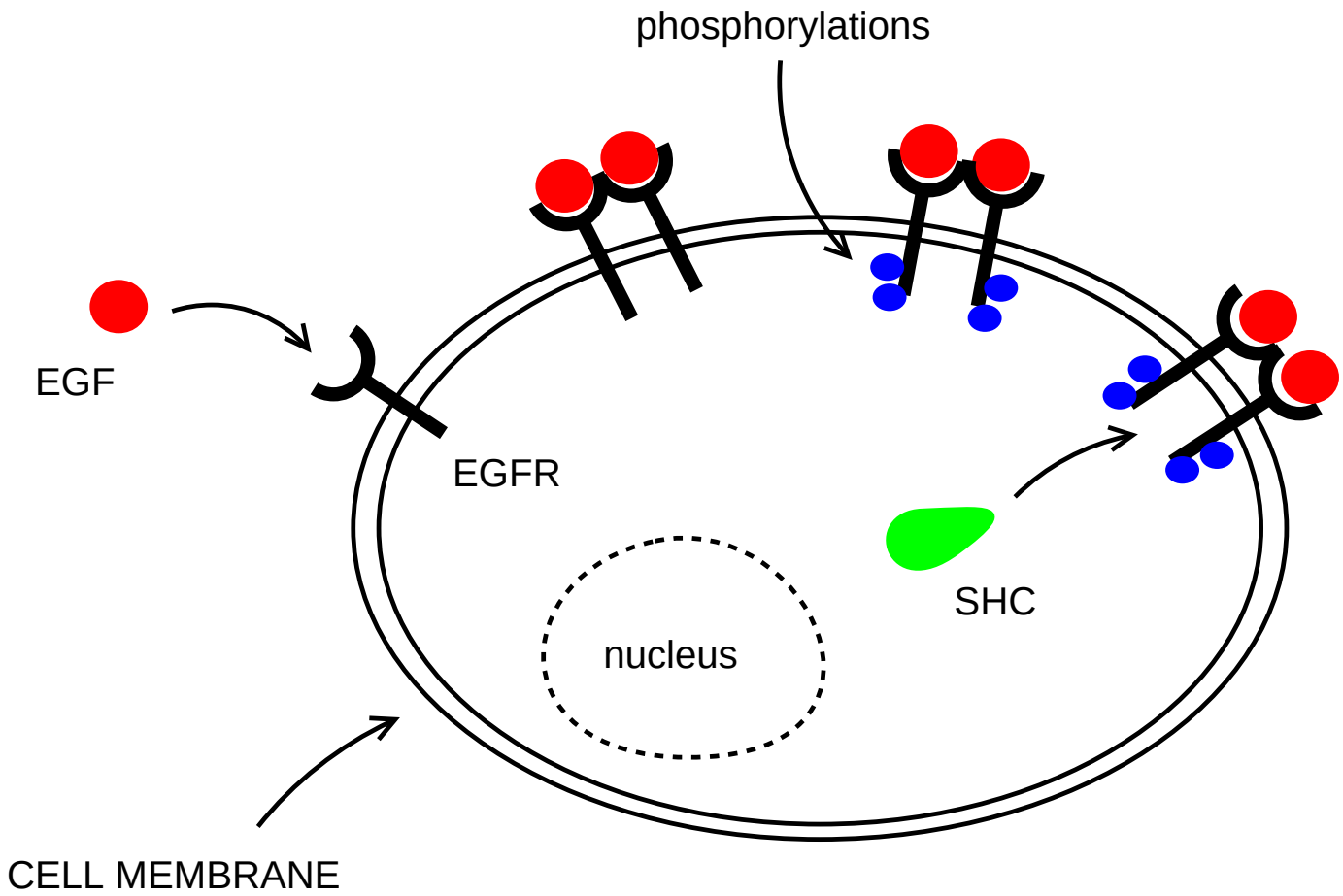
where α is link renaming, $L(T)$ the set of links occurring twice in the top level compartment of T

Main theoretical result

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

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

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

Observations

The language does not assume a fixed set of operators on membranes. Quite general operations can be modeled by rewrite rules

Rewrite rules are applied sequentially both in the non–deterministic and in the stochastic case. Forms of parallelism in rewrite rule application could be studied

Verification and Simulation

We have defined strong and weak bisimulations for CLS and properties such as causality of events can be proved

We have developed a simulator for Stochastic CLS

Current and future work

In order to model cell divisions and differentiations, tissues, etc...

- we are developing a spatial extension of CLS in which terms are placed and can move in a 2D/3D space

Moreover,

- we are developing a translation of Kohn Molecular Interaction Maps into CLS

As future work:

- we plan to use CLS to study (in collaboration with biologists) retinal cell develpment and differentiation

References

R. Barbuti, A. Maggiolo-Schettini, P. Milazzo, P. Tiberi and A. Troina. *Stochastic CLS for the Modeling and Simulation of Biological Systems*. Submitted for publication.

R. Barbuti, A. Maggiolo-Schettini and P. Milazzo. *Extending the Calculus of Looping Sequences to Model Protein Interaction at the Domain Level*. Int. Symposium on Bioinformatics Research and Applications (ISBRA'07), LNBI 4463, pages 638–649, Springer, 2006.

R. Barbuti, A. Maggiolo-Schettini, P. Milazzo and A. Troina. *Bisimulation Congruences in the Calculus of Looping Sequences*. Int. Colloquium on Theoretical Aspects of Computing (ICTAC'06), LNCS 4281, pages 93–107, 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, pages 21–35, 2006.

P. Milazzo. *Qualitative and Quantitative Formal Modeling of Biological Systems*, PhD Thesis, Università di Pisa.

APPENDIX

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)
- ℓ_i, ℓ'_i are stoichiometric coefficients
- k, k_{-1} are the kinetic constants

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

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

Background: Gillespie's simulation algorithm

- represents a chemical solution as a multiset of molecules
- computes the reaction rate a_μ 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 $a_1 = X_1 X_2 k_1$

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

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 τ exponentially distributed with parameter $\sum_{\nu=1}^{M} a_\nu$;
- The reaction R_μ 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 τ and the chemical solution is updated.

Well–formedness of LCLS terms and patterns in LCLS

An LCLS term (or pattern) is well–formed if and only if a label occurs no more than twice, and two occurrences of a label are always in the same compartment

Type system for well–formedness:

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

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

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

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

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