Qualitative and Quantitative Formal Modeling of Biological Systems

Paolo Milazzo

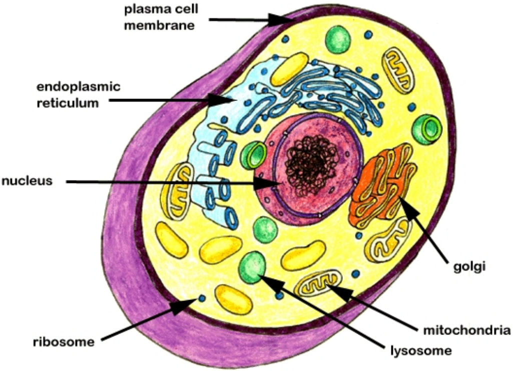
Dipartimento di Informatica, Università di Pisa, Italy

Camerino – April, 2007

Outline of the talk

1. Introduction
   - Cells are complex interactive systems
   - The EGF pathway and the *lac* operon
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - CLS as an abstraction for biomolecular systems
   - The EGF pathway and the *lac* operon in CLS
3. Bisimulations in CLS
   - A labeled semantics for CLS
   - Bisimulations in CLS
   - Bisimulations applied to the CLS model of the *lac* operon
4. CLS variants
   - Stochastic CLS
   - LCLS
5. Future Work and References

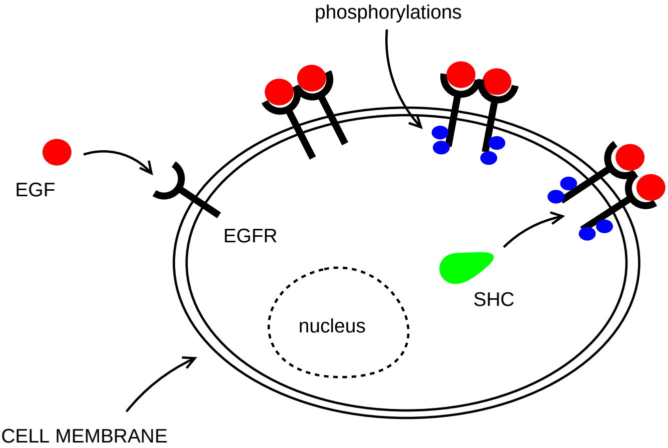
Cells: complex systems of interactive components



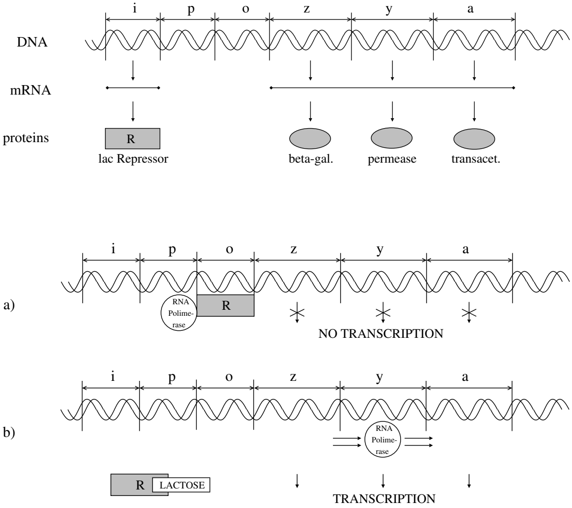
- Two classifications of cell:
  - procaryotic
  - eucaryotic
- Main actors:
  - membranes
  - proteins
  - DNA/RNA strands
- Interaction networks:
  - metabolic pathways
  - signaling pathways
  - gene regulatory networks

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

Examples of interaction networks: the EGF pathway



Examples of interaction networks: the *lac* operon



Outline of the talk

1. Introduction
   - Cells are complex interactive systems
   - The EGF pathway and the *lac* operon
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - CLS as an abstraction for biomolecular systems
   - The EGF pathway and the *lac* operon in CLS
3. Bisimulations in CLS
   - A labeled semantics for CLS
   - Bisimulations in CLS
   - Bisimulations applied to the CLS model of the *lac* operon
4. CLS variants
   - Stochastic CLS
   - LCLS
5. 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:

$$\begin{array}{rcl} T & ::= & S \quad \big| \quad \left(S\right)^L \rfloor T \quad \big| \quad T \mid T \\ S & ::= & \epsilon \quad \big| \quad a \quad \big| \quad S \cdot S \end{array}$$

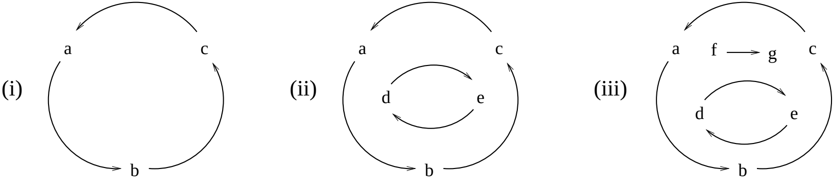
where a is a generic element of $\mathcal{E}$, and ϵ is the empty sequence.

The operators are:

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

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

Example of Terms



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

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

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

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:

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

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

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

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

and closed under structural congruence $\equiv$.

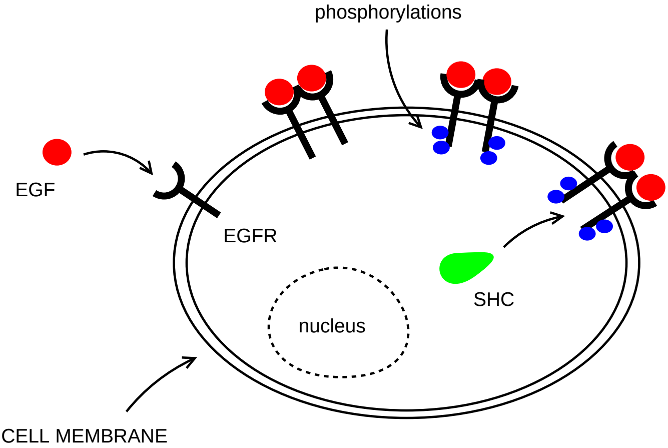
Biomolecular entities as CLS terms

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):

$$EGF \mid EGF \mid \big(EGFR \cdot EGFR \cdot EGFR \cdot EGFR\big)^L \rfloor (SHC \mid SHC)$$

Rewrite rules modeling the first steps of the pathway:

$$EGF \mid \big(EGFR \cdot \widetilde{x}\big)^L \rfloor X \mapsto \big(CMPLX \cdot \widetilde{x}\big)^L \rfloor X \tag{R1}$$

$$\big(CMPLX \cdot \widetilde{x} \cdot CMPLX \cdot \widetilde{y}\big)^L \rfloor X \mapsto \big(DIM \cdot \widetilde{x} \cdot \widetilde{y}\big)^L \rfloor X \tag{R2}$$

$$\big(DIM \cdot \widetilde{x}\big)^L \rfloor X \mapsto \big(DIMp \cdot \widetilde{x}\big)^L \rfloor X \tag{R3}$$

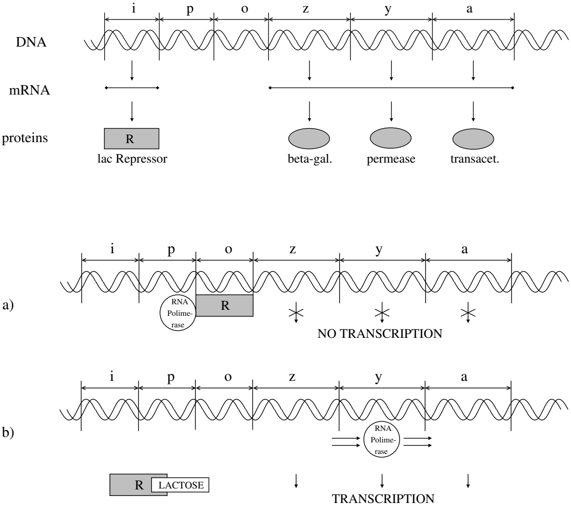
$$\big(DIMp \cdot \widetilde{x}\big)^L \rfloor (SHC \mid X) \mapsto \big(DIMpSHC \cdot \widetilde{x}\big)^L \rfloor X \tag{R4}$$

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: the *lac* operon (1)



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

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

Rules for DNA transcription/translation:

$$\begin{aligned}
lacI \cdot \widetilde{x} &\mapsto lacI' \cdot \widetilde{x} \mid repr && \text{(R1)}\\
polym \mid \widetilde{x} \cdot lacP \cdot \widetilde{y} &\mapsto \widetilde{x} \cdot PP \cdot \widetilde{y} && \text{(R2)}\\
\widetilde{x} \cdot PP \cdot lacO \cdot \widetilde{y} &\mapsto \widetilde{x} \cdot lacP \cdot PO \cdot \widetilde{y} && \text{(R3)}\\
\widetilde{x} \cdot PO \cdot lacZ \cdot \widetilde{y} &\mapsto \widetilde{x} \cdot lacO \cdot PZ \cdot \widetilde{y} && \text{(R4)}\\
\widetilde{x} \cdot PZ \cdot lacY \cdot \widetilde{y} &\mapsto \widetilde{x} \cdot lacZ \cdot PY \cdot \widetilde{y} \mid betagal && \text{(R5)}\\
\widetilde{x} \cdot PY \cdot lacA &\mapsto \widetilde{x} \cdot lacY \cdot PA \mid perm && \text{(R6)}\\
\widetilde{x} \cdot PA &\mapsto \widetilde{x} \cdot lacA \mid transac \mid polym && \text{(R7)}
\end{aligned}$$

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

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

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:

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

$$LACT \mid \left(m \cdot \widetilde{x}\right)^L \rfloor X \ \mapsto \left(m \cdot \widetilde{x}\right)^L \rfloor \left(X \mid LACT\right) \qquad \text{(R9)}$$

$$\widetilde{x} \cdot RO \cdot \widetilde{y} \mid LACT \ \mapsto \widetilde{x} \cdot lacO \cdot \widetilde{y} \mid RLACT \qquad \text{(R10)}$$

$$\left(\widetilde{x}\right)^L \rfloor \left(perm \mid X\right) \ \mapsto \left(perm \cdot \widetilde{x}\right)^L \rfloor X \qquad \text{(R11)}$$

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

$$betagal \mid LACT \ \mapsto betagal \mid GLU \mid GAL \qquad \text{(R13)}$$

CLS modeling examples: the *lac* operon (4)

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

Example:

$$\begin{aligned}
& Ecoli|LACT|LACT \\
\to^* & \, (m)^L \rfloor (lacI' \cdot lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA \mid polym \mid repr)|LACT|LACT \\
\to^* & \, (m)^L \rfloor (lacI' \cdot lacP \cdot RO \cdot lacZ \cdot lacY \cdot lacA \mid polym)|LACT|LACT \\
\to^* & \, (m)^L \rfloor (lacI' \cdot lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA|polym|RLACT)|LACT \\
\to^* & \, (perm \cdot m)^L \rfloor (lacI'{-}A|betagal|transac|polym|RLACT)|LACT \\
\to^* & \, (perm \cdot m)^L \rfloor (lacI'{-}A|betagal|transac|polym|RLACT|GLU|GAL)
\end{aligned}$$

Outline of the talk

1. Introduction
   - Cells are complex interactive systems
   - The EGF pathway and the *lac* operon
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - CLS as an abstraction for biomolecular systems
   - The EGF pathway and the *lac* operon in CLS
3. Bisimulations in CLS
   - A labeled semantics for CLS
   - Bisimulations in CLS
   - Bisimulations applied to the CLS model of the *lac* operon
4. CLS variants
   - Stochastic CLS
   - LCLS
5. Future Work and References

Bisimulations

Bisimilarity is widely accepted as the finest extensional behavioral equivalence one may impose on systems.

- Two systems are bisimilar if they can perform step by step the same interactions with the environment.
- Properties of a system can be verified by assessing the bisimilarity with a system known to enjoy them.

Bisimilarities need semantics based on labeled transition relations capturing the potential interactions with the environment.

- In process calculi, transitions are usually labeled with actions.
- In CLS labels are contexts in which rules can be applied.

Labeled semantics (1)

Contexts $\mathcal{C}$ are given by the following grammar:

$$\mathcal{C} ::= \Box \quad \big| \quad \mathcal{C} \mid T \quad \big| \quad T \mid \mathcal{C} \quad \big| \quad (S)^{L} \rfloor \mathcal{C}$$

where $T \in \mathcal{T}$ and $S \in \mathcal{S}$. Context $\Box$ is called the *empty context*.

Parallel Contexts $\mathcal{C}_P$ are given by the following grammar:

$$\mathcal{C}_P ::= \Box \quad \big| \quad \mathcal{C}_P \mid T \quad \big| \quad T \mid \mathcal{C}_P .$$

where $T \in \mathcal{T}$.

$C[T]$ is context application and $C[C']$ is context composition.

Labeled semantics (2)

Given a set of rewrite rules $\mathcal{R} \subseteq \Re$, the **labeled semantics** of CLS is the labeled transition system given by the following inference rules:

$$\text{(rule\_appl)}\ \frac{P \mapsto P' \in \mathcal{R} \quad C[T''] \equiv P\sigma \quad T'' \not\equiv \epsilon \quad \sigma \in \Sigma \quad C \in \mathcal{C}}{T'' \xrightarrow{C} T'\sigma}$$

$$\text{(cont)}\ \frac{T \xrightarrow{\square} T'}{(S)^L \rfloor T \xrightarrow{\square} (S)^L \rfloor T'} \qquad \text{(par)}\ \frac{T \xrightarrow{C} T' \quad C \in \mathcal{C}_P}{T \mid T'' \xrightarrow{C} T' \mid T''}$$

where the dual version of the *(par)* rule is omitted.

Rule (rule_appl) describes the (potential) application of a rule.

- $T'' \not\equiv \epsilon$ in the premise implies that C cannot provide completely the left hand side of the rewrite rule.
- Example: let $R = a \mid b \mapsto c$, we have $a \xrightarrow{\square \mid b} c$, but $\epsilon \overset{a|b}{\not\longrightarrow}$.

Labeled semantics (3)

Given a set of rewrite rules $\mathcal{R} \subseteq \Re$, the **labeled semantics** of CLS is the labeled transition system given by the following inference rules:

$$(\text{rule\_appl}) \; \frac{P \mapsto P' \in \mathcal{R} \quad C[T''] \equiv T\sigma \quad T'' \not\equiv \epsilon \quad \sigma \in \Sigma \quad C \in \mathcal{C}}{T'' \xrightarrow{C} T'\sigma}$$

$$(\text{cont}) \; \frac{T \xrightarrow{\square} T'}{(S)^L \rfloor T \xrightarrow{\square} (S)^L \rfloor T'} \qquad (\text{par}) \; \frac{T \xrightarrow{C} T' \quad C \in \mathcal{C}_P}{T \mid T'' \xrightarrow{C} T' \mid T''}$$

where the dual version of the *(par)* rule is omitted.

Rule (cont) propagates □–labeled transitions from the inside to the outside of a looping sequence.

- Transition labeled with a non–empty context cannot be propagated.
- Example: let $R = a \mid b \mapsto c$, we have $a \xrightarrow{\square \mid b} c$, but $(d)^L \rfloor a \not\xrightarrow{\square | b}$.

Labeled semantics (4)

Given a set of rewrite rules $\mathcal{R} \subseteq \Re$, the **labeled semantics** of CLS is the labeled transition system given by the following inference rules:

$$\text{(rule\_appl)}\ \frac{P \mapsto P' \in \mathcal{R} \quad C[T''] \equiv T\sigma \quad T'' \not\equiv \epsilon \quad \sigma \in \Sigma \quad C \in \mathcal{C}}{T'' \xrightarrow{C} T'\sigma}$$

$$\text{(cont)}\ \frac{T \xrightarrow{\Box} T'}{(S)^L \rfloor T \xrightarrow{\Box} (S)^L \rfloor T'} \qquad \text{(par)}\ \frac{T \xrightarrow{C} T' \quad C \in \mathcal{C}_P}{T \mid T'' \xrightarrow{C} T' \mid T''}$$

where the dual version of the *(par)* rule is omitted.

Rule (par) propagates transitions labeled with parallel contexts in parallel components.

- Example: let $R = (a)^L \rfloor b \mapsto c$, we have $b \xrightarrow{(a)^L \rfloor \Box} c$, but $b \mid d \not\xrightarrow{(a)^L \rfloor \Box}$ because R cannot be applied $(a)^L \rfloor (b \mid d)$

Bisimulations in CLS (1)

A binary relation R on terms is a **strong bisimulation** if, given T_1, T_2 such that $T_1 R T_2$, the two following conditions hold:

- $T_1 \xrightarrow{C} T_1' \implies \exists T_2'$ s.t. $T_2 \xrightarrow{C} T_2'$and $T_1' R T_2'$
- $T_2 \xrightarrow{C} T_2' \implies \exists T_1'$ s.t. $T_1 \xrightarrow{C} T_1'$ and $T_2' R T_1'$.

The *strong bisimilarity* $\sim$ is the largest of such relations.

A binary relation R on terms is a **weak bisimulation** if, given T_1, T_2 such that $T_1 R T_2$, the two following conditions hold:

- $T_1 \xrightarrow{C} T_1' \implies \exists T_2'$ s.t. $T_2 \overset{C}{\Longrightarrow} T_2'$and $T_1' R T_2'$
- $T_2 \xrightarrow{C} T_2' \implies \exists T_1'$ s.t. $T_1 \overset{C}{\Longrightarrow} T_1'$ and $T_2' R T_1'$.

The *weak bisimilarity* $\approx$ is the largest of such relations.

Theorem: Strong and weak bisimilarities are congruences.

Bisimulations in CLS (2)

Consider the following set of rewrite rules:

$$\mathcal{R} = \{ \quad a \mid b \mapsto c \quad , \quad d \mid b \mapsto e \quad , \quad e \mapsto e \quad , \quad c \mapsto e \quad , \quad f \mapsto a \quad \}$$

We have that $a \sim d$, because

$$a \xrightarrow{\square \mid b} c \xrightarrow{\square} e \xrightarrow{\square} e \xrightarrow{\square} \ldots$$

$$d \xrightarrow{\square \mid b} e \xrightarrow{\square} e \xrightarrow{\square} \ldots$$

and $f \approx d$, because

$$f \xrightarrow{\square} a \xrightarrow{\square \mid b} c \xrightarrow{\square} e \xrightarrow{\square} e \xrightarrow{\square} \ldots$$

On the other hand, $f \not\sim e$ and $f \not\approx e$.

$$e \xrightarrow{\square} e \xrightarrow{\square} e \xrightarrow{\square} \ldots$$

Bisimulations in CLS (3)

Let us consider systems $(T, \mathcal{R})$...

A binary relation R is a **strong bisimulation on systems** if, given $(T_1, \mathcal{R}_1)$ and $(T_2, \mathcal{R}_2)$ such that $(T_1, \mathcal{R}_1) R (T_2, \mathcal{R}_2)$:

- $\mathcal{R}_1 : T_1 \xrightarrow{C} T_1' \implies \exists T_2'$ s.t. $\mathcal{R}_2 : T_2 \xrightarrow{C} T_2'$ and $(T_1', \mathcal{R}_1) R (T_2', \mathcal{R}_2)$
- $\mathcal{R}_2 : T_2 \xrightarrow{C} T_2' \implies \exists T_1'$ s.t. $\mathcal{R}_1 : T_1 \xrightarrow{C} T_1'$ and $(R_2, T_2') R (\mathcal{R}_1, T_1')$.

The *strong bisimilarity on systems* $\sim$ is the largest of such relations.

A binary relation R is a **weak bisimulation on systems** if, given $(T_1, \mathcal{R}_1)$ and $(T_2, \mathcal{R}_2)$ such that $(T_1, \mathcal{R}_1) R (T_2, \mathcal{R}_2)$:

- $\mathcal{R}_1 : T_1 \xrightarrow{C} T_1' \implies \exists T_2'$ s.t. $\mathcal{R}_2 : T_2 \overset{C}{\Longrightarrow} T_2'$ and $(T_1', \mathcal{R}_1) R (T_2', \mathcal{R}_2)$
- $\mathcal{R}_2 : T_2 \xrightarrow{C} T_2' \implies \exists T_1'$ s.t. $\mathcal{R}_1 : T_1 \overset{C}{\Longrightarrow} T_1'$ and $(T_2', \mathcal{R}_2) R (T_1', \mathcal{R}_1)$

The *weak bisimilarity on systems* $\approx$ is the largest of such relations.

Strong and weak bisimilarities on systems are NOT congruences.

Bisimulations in CLS (4)

Consider the following sets of rewrite rules

$$\mathcal{R}_1 = \{ a \mid b \mapsto c \} \qquad \mathcal{R}_2 = \{ a \mid d \mapsto c \, , \;\; b \mid e \mapsto c \}$$

We have that $\langle a, \mathcal{R}_1 \rangle \approx \langle e, \mathcal{R}_2 \rangle$ because

$$\mathcal{R}_1 : a \xrightarrow{\square | b} c \qquad \mathcal{R}_2 : e \xrightarrow{\square | b} c$$

and $\langle b, \mathcal{R}_1 \rangle \approx \langle d, \mathcal{R}_2 \rangle$, because

$$\mathcal{R}_1 : b \xrightarrow{\square | a} c \qquad \mathcal{R}_2 : d \xrightarrow{\square | a} c$$

but $\langle a \mid b, \mathcal{R}_1 \rangle \not\approx \langle e \mid d, \mathcal{R}_2 \rangle$, because

$$\mathcal{R}_1 : a \mid b \xrightarrow{\square} c \qquad \mathcal{R}_2 : c \mid d \not\xrightarrow{\square}$$

Applying bisimulations to the *lac* operon (1)

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

It can be easily proved that

$$\begin{gathered} lacI \cdot lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA \\ \approx \\ lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA \mid repr \end{gathered}$$

and since weak bisimularity is a congruence the former can be replaced by the latter in the model.

Applying bisimulations to the *lac* operon (2)

By using the weak bisimilarity on systems we can prove that from the state in which the repressor is bound to the DNA we can reach a state in which the enzymes are synthesized only if lactose appears in the environment.

We replace rule

$$\widetilde{x} \cdot RO \cdot \widetilde{y} \mid LACT \mapsto \widetilde{x} \cdot lacO \cdot \widetilde{y} \mid RLACT \qquad \text{(R10)}$$

with

$$(\widetilde{w})^{L} \rfloor (\widetilde{x} \cdot RO \cdot \widetilde{y} \mid LACT \mid X) \mid START \mapsto \\ (\widetilde{w})^{L} \rfloor (\widetilde{x} \cdot lacO \cdot \widetilde{y} \mid RLACT \mid X) \qquad \text{(R10bis)}$$

The obtained model is bisimilar to $(T_1, \mathcal{R})$ where $\mathcal{R}$ is

$$\begin{array}{llllllll} T_1 \mid LACT & \mapsto & T_2 & \text{(R1')} \qquad & T_2 \mid START & \mapsto & T_3 & \text{(R3')} \\ T_2 \mid LACT & \mapsto & T_2 & \text{(R2')} \qquad & T_3 \mid LACT & \mapsto & T_3 & \text{(R4')} \end{array}$$

that is a system satisfying the property.

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

Bisimilarities of Brane Calculi are preserved after translation into CLS

Some variants of CLS

- Full–CLS
  - The looping operator can be applied to any term
  - Rule $a \mid b \mapsto c$ can be applied to $b \mid \big(a \cdot a \cdot a \cdot a\big)^{L} \rfloor d$

- 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

Outline of the talk

1. Introduction
   - Cells are complex interactive systems
   - The EGF pathway and the *lac* operon
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - CLS as an abstraction for biomolecular systems
   - The EGF pathway and the *lac* operon in CLS
3. Bisimulations in CLS
   - A labeled semantics for CLS
   - Bisimulations in CLS
   - Bisimulations applied to the CLS model of the *lac* operon
4. CLS variants
   - Stochastic CLS
   - LCLS
5. 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)
- ℓ_i, ℓ'_i are stoichiometric coefficients
- k, k_{-1} are the kinetic constants

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

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

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 $a_1 = X_1 X_2 c_1$

reaction $R_2 : 2S_1 \rightarrow S_1 + S_2$ rate $a_2 = \frac{X_1(X_1-1)}{2} c_2$

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.

Stochastic CLS (1)

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

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

Example: given $T = a \mid a \mid b \mid b$

- the set of subterms of T is

$$\{a\ ,\ b\ ,\ a \mid a\ ,\ a \mid b\ ,\ b \mid b\ ,\ a \mid a \mid b\ ,\ a \mid b \mid b\ ,\ T\}$$

- the multiset of reactants in T is

$$\{a\ ,\ a\ ,\ b\ ,\ b\ ,\ a \mid a\ ,\ a \mid b\ ,\ a \mid b\ ,\ a \mid b\ ,\ a \mid b\ ,\ b \mid b\ , \\ a \mid a \mid b\ ,\ a \mid a \mid b\ ,\ a \mid b \mid b\ ,\ a \mid b \mid b\ ,\ T\}$$

Stochastic CLS (2)

- Given $T = a \mid a \mid b \mid b$ the multiset of reactants in T is

$$\{a\,,\ a\,,\ b\,,\ b\,,\ a \mid a\,,\ a \mid b\,,\ a \mid b\,,\ a \mid b\,,\ a \mid b\,,\ b \mid b\,,\\ a \mid a \mid b\,,\ a \mid a \mid b\,,\ a \mid b \mid b\,,\ a \mid b \mid b\,,\ T\}$$

Defining the stochastic semantics would be easy for rules without variables:

- Rewrite rules could be extended with kinetic constants (e.g. $a \mid b \overset{k}{\mapsto} c$)
- The number of possible combinations of molecules involved in the reactions corresponds to the number of reactants equivalent to the left-hand side of the rule (e.g. $2 \times 2 = 4$ corresponds to 4 occurrences of $a \mid b$)

Stochastic CLS (3)

We consider rewrite rules containing variables as *rewrite rule schemata*

- at step we compute the set of ground rules that can be applied among those obtained by instantiating variables of the rewrite rule schama
- we reduce the problem of defining the semantics with rule schemata to the simpler problem of defining the semantics with ground rules only

Example: given $T = a \cdot b \mid a \cdot c$

From rule schema $a \cdot \widetilde{x} \mid a \cdot \widetilde{y} \overset{k}{\mapsto} d$ we can derive only $a \cdot b \mid a \cdot c \overset{k}{\mapsto} d$

From rule schema $a \cdot \widetilde{x} \mid a \cdot \widetilde{y} \overset{k}{\mapsto} \widetilde{y}$ we can derive both $a \cdot b \mid a \cdot c \overset{k}{\mapsto} b$ and $a \cdot b \mid a \cdot c \overset{k}{\mapsto} c$

Problem: the kinetic constant could be different for different instantiations

- We enrich rewrite rules with rate functions $f : \Sigma \rightarrow \mathbb{R}$ rather than k

Stochastic CLS (4)

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

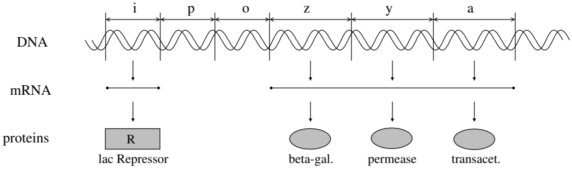
$$\frac{R = T_1 \overset{k}{\mapsto} T_2 \in AR(\mathcal{R}, T) \qquad T \equiv C[T_1]}{T \xrightarrow{R,k\cdot AC(R,T,C[T_2])} C[T_2]}$$

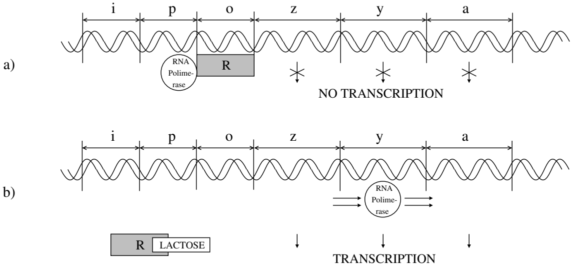
where:

- $AR(\mathcal{R}, T)$ is the set of ground rewrite rules obtained by schemata 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 *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:

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

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

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

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

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

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

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

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

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

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

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

The behaviour of the three enzymes for lactose degradation:

$$(\widetilde{x})^{L} \rfloor (perm \mid X) \overset{0.1 \cdot f_1}{\longmapsto} (perm \cdot \widetilde{x})^{L} \rfloor X \qquad \text{(S11)}$$

$$LACT \mid (perm \cdot \widetilde{x})^{L} \rfloor X \overset{0.001 \cdot f_2}{\longmapsto} (perm \cdot \widetilde{x})^{L} \rfloor (LACT|X) \qquad \text{(S12)}$$

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

where $f_1(\sigma) = occ(perm, \sigma(X)) + 1,\ f_2(\sigma) = occ(perm, \sigma(\widetilde{x})) + 1.$

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

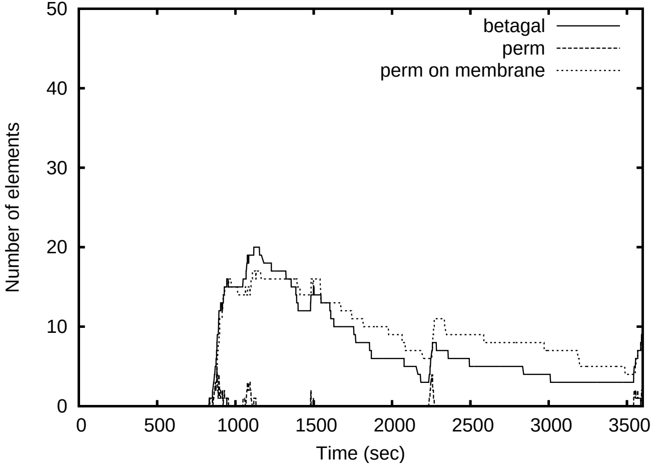
$$perm \overset{0.001}{\longmapsto} \epsilon \qquad \text{(S14)} \qquad\qquad betagal \overset{0.001}{\longmapsto} \epsilon \qquad \text{(S15)}$$

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

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

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

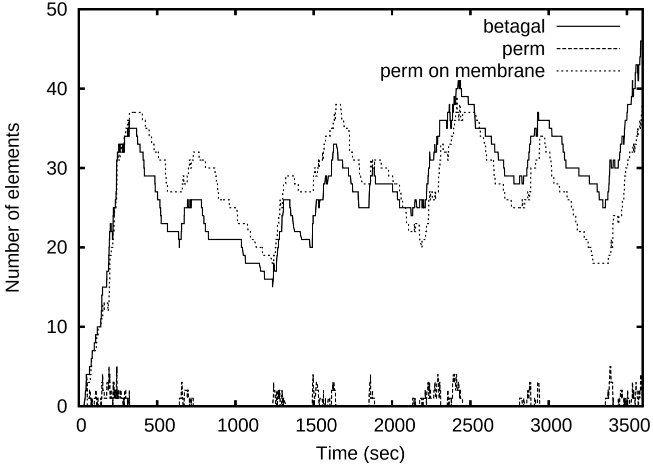
Simulation results (1)



Production of enzymes in the absence of lactose

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

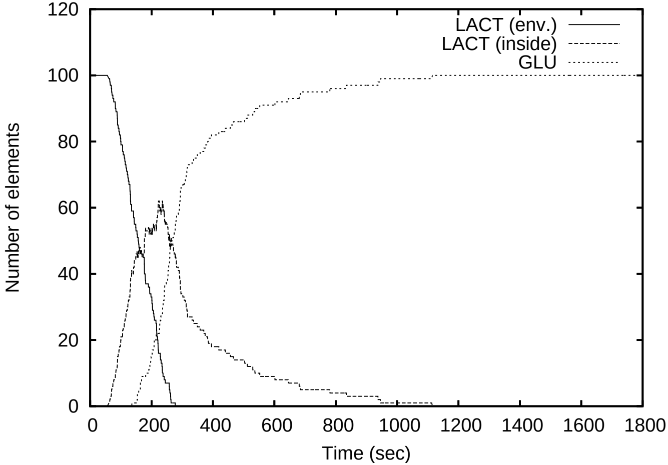
Simulation results (2)



Production of enzymes in the presence of lactose

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

Simulation results (3)



Degradation of lactose into glucose

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

Outline of the talk

1. Introduction
   - Cells are complex interactive systems
   - The EGF pathway and the *lac* operon
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - CLS as an abstraction for biomolecular systems
   - The EGF pathway and the *lac* operon in CLS
3. Bisimulations in CLS
   - A labeled semantics for CLS
   - Bisimulations in CLS
   - Bisimulations applied to the CLS model of the *lac* operon
4. CLS variants
   - Stochastic CLS
   - LCLS
5. 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:

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

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:

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

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

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:

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

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

$$8.\ \frac{(N_1, N'_1) \models SP_1 \quad (N_2, N'_2) \models SP_2 \quad N_1 \cap N_2 = N'_1 \cap N_2 = N_1 \cap N'_2 = \emptyset}{(N_1 \cup N_2 \cup (N'_1 \cap N'_2), (N'_1 \cup N'_2) \setminus (N'_1 \cap N'_2)) \models SP_1 \cdot SP_2}$$

$$9.\ \frac{(N_1, N'_1) \models P_1 \quad (N_2, N'_2) \models P_2 \quad N_1 \cap N_2 = N'_1 \cap N_2 = N_1 \cap N'_2 = \emptyset}{(N_1 \cup N_2 \cup (N'_1 \cap N'_2), (N'_1 \cup N'_2) \setminus (N'_1 \cap N'_2)) \models P_1 \mid P_2}$$

$$10.\ \frac{(N_1, N'_1) \models SP \quad (N_2, N'_2) \models P \quad N_1 \cap N_2 = N'_1 \cap N_2 = N_1 \cap N'_2 = \emptyset \quad N'_2 \subseteq N'_1}{(N_1 \cup N'_2, N'_1 \setminus N'_2) \models (SP)^L \rfloor P}$$

Instantiation of LCLS patterns

An instantiation function σ is well–formed if it maps variables into well–formed CLOSED terms and sequences

- otherwise the w.f. pattern $\left(a\right)^{L} \rfloor X$ could be instantiated to the non–w.f. term $\left(a\right)^{L} \rfloor b^1$

The definition of pattern instantiation AVOIDS this kind situations:

- $P = a \cdot \widetilde{x} \mid X$, $\sigma(\widetilde{x}) = b^1 \cdot c^1$, $\sigma(X) = d^1 \cdot e^1$ are all w.f.
- $P\sigma = a \cdot b^1 \cdot c^1 \mid d^1 \cdot e^1$ is non–w.f.

Clashing labels are renamed during pattern instatiation

Compartment safe rewrite rules

By applying a rewrite rule composed by w.f. patterns to a w.f. term by using a w.f. instantiation function we obtain a w.f. term

W.f. instantiations are closed

- Rule $(a)^L \rfloor (\widetilde{x} \mid \widetilde{y}) \mapsto \widetilde{x} \mid (a)^L \rfloor \widetilde{y}$ cannot be applied to $(a)^L \rfloor (b^1 \mid c^1)$ (so to obtain $b^1 \mid (a)^L \rfloor c^1$)

BUT

- Rule $\widetilde{x} \cdot a \mapsto \widetilde{x} \cdot b$ cannot be applied to $c^1 \mid d^1 \cdot a$ (so to obtain $c^1 \mid d^1 \cdot b$)

To allow application of the second kind of rules

- we relax the constraint on instantiations
- we add a constraint on rewrite rules

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

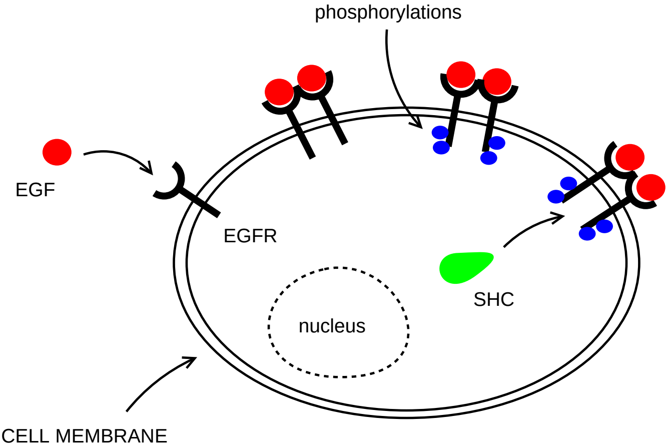
Main theoretical result

Theorem (Subject Reduction)

Given a set of well–formed rewrite rules $\mathcal{R}$ and a well–formed term T

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

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

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

$$\begin{gathered}\left(SR_{E1}\cdot R_{E2}\cdot R_{I1}\cdot R_{I2}\cdot\widetilde{x}\cdot SR_{E1}\cdot R_{E2}\cdot R_{I1}\cdot R_{I2}\cdot\widetilde{y}\right)^{L} \rfloor X \;\mapsto \\ \left(SR_{E1}\cdot R^{1}_{E2}\cdot R_{I1}\cdot R_{I2}\cdot SR_{E1}\cdot R^{1}_{E2}\cdot R_{I1}\cdot R_{I2}\cdot\widetilde{x}\cdot\widetilde{y}\right)^{L} \rfloor X\end{gathered} \tag{R2}$$

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

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

$$\begin{gathered}\left(R^{1}_{E2}\cdot PR_{I1}\cdot R_{I2}\cdot\widetilde{x}\cdot R^{1}_{E2}\cdot PR_{I1}\cdot R_{I2}\cdot\widetilde{y}\right)^{L} \rfloor (SHC \mid X) \;\mapsto \\ \left(R^{1}_{E2}\cdot PR_{I1}\cdot R^{2}_{I2}\cdot\widetilde{x}\cdot R^{1}_{E2}\cdot PR_{I1}\cdot R_{I2}\cdot\widetilde{y}\right)^{L} \rfloor (SHC^{2} \mid X)\end{gathered} \tag{R5}$$

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

$$
\begin{array}{rl}
 & EGF \mid EGF \mid \left(EGFR \cdot EGFR \cdot EGFR \cdot EGFR\right)^{L} \rfloor \left(SHC \mid SHC\right) \\
\xrightarrow{(R1)} & EGF \mid \left(SR_{E1} \cdot R_{E2} \cdot R_{I1} \cdot R_{I2} \cdot EGFR \cdot EGFR \cdot EGFR\right)^{L} \rfloor \left(SHC \mid SHC\right) \\
\xrightarrow{(R1)} & \left(SR_{E1} \cdot R_{E2} \cdot R_{I1} \cdot R_{I2} \cdot EGFR \cdot SR_{E1} \cdot R_{E2} \cdot R_{I1} \cdot R_{I2} \cdot EGFR\right)^{L} \rfloor \left(SHC \mid SHC\right) \\
\xrightarrow{(R2)} & \left(SR_{E1} \cdot R^{1}_{E2} \cdot R_{I1} \cdot R_{I2} \cdot SR_{E1} \cdot R^{1}_{E2} \cdot R_{I1} \cdot R_{I2} \cdot EGFR \cdot EGFR\right)^{L} \rfloor \left(SHC \mid SHC\right) \\
\xrightarrow{(R3)} & \left(SR_{E1} \cdot R^{1}_{E2} \cdot PR_{I1} \cdot R_{I2} \cdot SR_{E1} \cdot R^{1}_{E2} \cdot R_{I1} \cdot R_{I2} \cdot EGFR \cdot EGFR\right)^{L} \rfloor \left(SHC \mid SHC\right) \\
\xrightarrow{(R4)} & \left(SR_{E1} \cdot R^{1}_{E2} \cdot PR_{I1} \cdot R_{I2} \cdot SR_{E1} \cdot R^{1}_{E2} \cdot PR_{I1} \cdot R_{I2} \cdot EGFR \cdot EGFR\right)^{L} \rfloor \left(SHC \mid SHC\right) \\
\xrightarrow{(R5)} & \left(SR_{E1} \cdot R^{1}_{E2} \cdot PR_{I1} \cdot R^{2}_{I2} \cdot SR_{E1} \cdot R^{1}_{E2} \cdot PR_{I1} \cdot R_{I2} \cdot EGFR \cdot EGFR\right)^{L} \rfloor \left(SHC^{2} \mid SHC\right)
\end{array}
$$

Outline of the talk

1. Introduction
    - Cells are complex interactive systems
    - The EGF pathway and the *lac* operon
2. The Calculus of Looping Sequences (CLS)
    - Definition of CLS
    - CLS as an abstraction for biomolecular systems
    - The EGF pathway and the *lac* operon in CLS
3. Bisimulations in CLS
    - A labeled semantics for CLS
    - Bisimulations in CLS
    - Bisimulations applied to the CLS model of the *lac* operon
4. CLS variants
    - Stochastic CLS
    - LCLS
5. Future Work and References

Current and future work

We developed a prototype simulator based on Stochastic CLS to run the *lac* operon example

- currently, we are developing a complete and efficient simulator

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 develop a symbolic semantics of CLS, and a symbolic bisimulation relation to allow the development of a verification tool
- we plan to use CLS to study (in collaboration with biologists) retinal cell develpment and differentiation

References

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

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.