The Calculus of Looping Sequences

Paolo Milazzo

Dipartimento di Informatica, Università di Pisa, Italy

Bertinoro – September, 2008

Our aim...

At the beginning of our work our aim was to try to apply formal methods to models of biological systems

We were looking for a formalism

- based on term rewriting
- with a simple semantics
- very general

As a consequence, we defined the Calculus of Looping Sequences (CLS)...

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:

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

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

The operators are:

$S \cdot S$: Sequencing

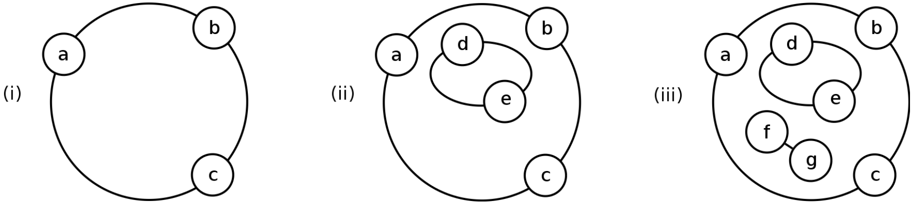
$(S)^L$: Looping (S is closed and it can rotate)

$T_1 \rfloor T_2$: Containment (T_1 contains T_2)

$T|T$: Parallel composition (juxtaposition)

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

Examples of Terms



$$(i) \qquad (a \cdot b \cdot c)^L \rfloor \epsilon$$

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

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

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:

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

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

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

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:

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

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

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 that are instantiations of P.

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 $a \cdot c \cdot c \cdot a$

Example: $\left(a \cdot \widetilde{x}\right)^L \rfloor (b \mid X) \mapsto \left(c \cdot \widetilde{x}\right)^L \rfloor X$

- can be applied to $\left(a \cdot a \cdot a\right)^L \rfloor (b \mid b \mid \left(a\right)^L \rfloor b)$
- the result is either $\left(c \cdot a \cdot a\right)^L \rfloor (b \mid \left(a\right)^L \rfloor b)$ or $\left(a \cdot a \cdot a\right)^L \rfloor (b \mid b \mid \left(c\right)^L \rfloor \epsilon)$

Formal Semantics

$P\sigma$ 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

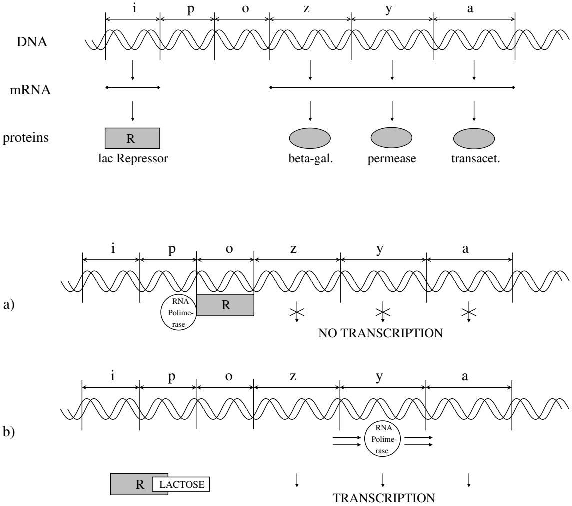
Given a set of rewrite rules $\mathcal{R}$, evolution of terms is described by the transition system given by the least relation $\rightarrow$ satisfying

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

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

and closed under structural congruence $\equiv$.

CLS modeling examples: the *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:

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

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

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

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

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

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

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

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:

$$
\begin{aligned}
repr \mid \widetilde{x} \cdot lacO \cdot \widetilde{y} &\mapsto \widetilde{x} \cdot RO \cdot \widetilde{y} && \text{(R8)}\\
LACT \mid \left(m \cdot \widetilde{x}\right)^{L} \rfloor X &\mapsto \left(m \cdot \widetilde{x}\right)^{L} \rfloor \left(X \mid LACT\right) && \text{(R9)}\\
\widetilde{x} \cdot RO \cdot \widetilde{y} \mid LACT &\mapsto \widetilde{x} \cdot lacO \cdot \widetilde{y} \mid RLACT && \text{(R10)}\\
\left(\widetilde{x}\right)^{L} \rfloor \left(perm \mid X\right) &\mapsto \left(perm \cdot \widetilde{x}\right)^{L} \rfloor X && \text{(R11)}\\
LACT \mid \left(perm \cdot \widetilde{x}\right)^{L} \rfloor X &\mapsto \left(perm \cdot \widetilde{x}\right)^{L} \rfloor \left(LACT \mid X\right) && \text{(R12)}\\
betagal \mid LACT &\mapsto betagal \mid GLU \mid GAL && \text{(R13)}
\end{aligned}
$$

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

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

Example:

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

Outline of the talk

1. Introduction

2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - 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. Stochastic CLS

5. Ongoing work
   - Spatial CLS
   - Translation of Stochastic CLS into Maude

6. 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.

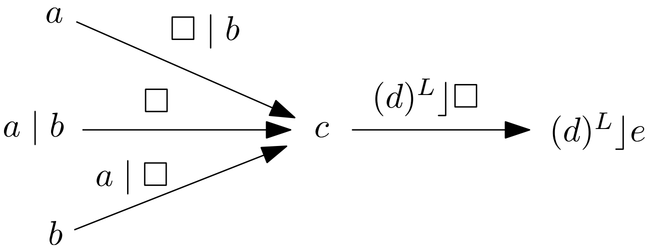
- In process calculi, transitions are usually labeled with actions.
- In CLS labels are contexts in which rules can be applied.

Labeled semantics

The idea: There is a (labeled) transition between terms T and T' if there exists a context C such that a rewrite rule can be applied to $C[T]$ with T' as result.

- C is used as transition label
- C must not provide the whole left hand side of the applied rewrite rule

An example: Let $\mathcal{R} = \{ \; a \mid b \mapsto c \;\; , \;\; (d)^L \rfloor c \mapsto (d)^L \rfloor e \; \}$



Labeled semantics

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

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

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

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:

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

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

where $\mathcal{C}_P$ are contexts that do not include $\left(S\right)^{L} \rfloor \mathcal{C}$ and the dual version of the *(par)* rule is omitted.

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

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

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

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

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

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

Applying bisimulations to the *lac* operon (1)

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

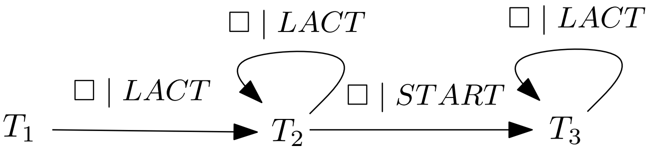
$$\begin{aligned}(\widetilde{w})^{L} \rfloor (\widetilde{x} \cdot RO \cdot \widetilde{y} \mid LACT \mid X) \mid START \;\mapsto&\\ (\widetilde{w})^{L} \rfloor (\widetilde{x} \cdot lacO \cdot \widetilde{y} \mid RLACT \mid X)& \qquad \text{(R10bis)}\end{aligned}$$

Applying bisimulations to the *lac* operon (2)

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

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

that is a system satisfying the wanted property.



Outline of the talk

1. Introduction
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - 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. Stochastic CLS
5. Ongoing work
   - Spatial CLS
   - Translation of Stochastic CLS into Maude
6. References

Background: Gillespie's simulation algorithm

- represents a chemical solution as a multiset of molecules
- each chemical reaction is associated with a kinetic constant
- 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 $a_1 = \binom{X_1}{1}\binom{X_2}{1}k_1 = X_1 X_2 k_1$

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 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

- Rewrite rules are enriched with kinetic constants

What is a reactant in Stochastic CLS?

- A *reactant combination* is an occurrence (up to $\equiv$) of a left hand side of a rewrite rule

Example: The application rate of $a \mid b \overset{k}{\mapsto} c$ to $a \mid a \mid a \mid b \mid b$ is $6k$

Example: The application rate of $(a \cdot \widetilde{x})^L \rfloor (b \mid X) \overset{k}{\mapsto} (c \cdot \widetilde{x})^L \rfloor X$ to $(a \cdot a \cdot a)^L \rfloor (b \mid b) \mid (a \cdot a)^L \rfloor b$ is

- $6k$, with $(c \cdot a \cdot a)^L \rfloor b \mid (a \cdot a)^L \rfloor b$ as result
- $+\ 2k$, with $(a \cdot a \cdot a)^L \rfloor (b \mid b) \mid (c \cdot a)^L \rfloor \epsilon$ as result
- $=\ 8k$

Stochastic CLS (2)

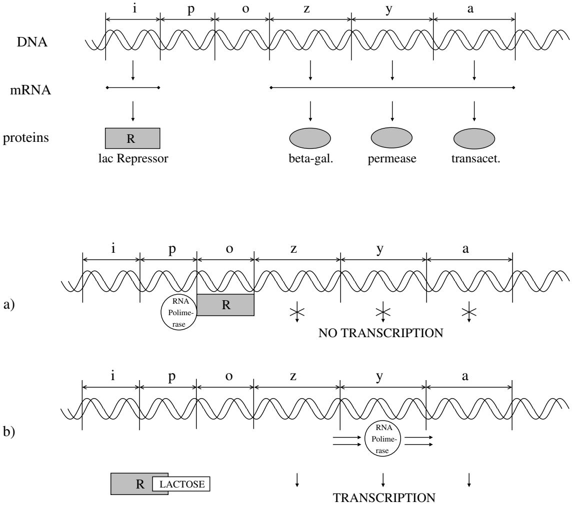
Given a finite set of stochastic rewrite rules $\mathcal{R}$, the semantics of Stochastic CLS is the least transition relation $\xrightarrow{R,T,r,b}$ closed wrt $\equiv$ and satisfying by the following inference rules:

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

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

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:

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

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

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

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

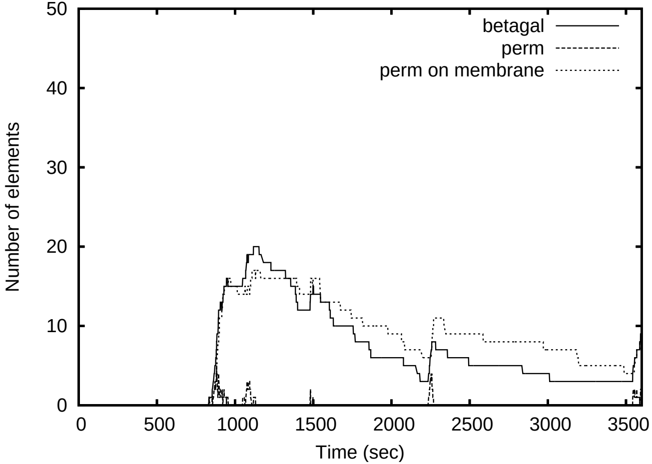
$$perm \overset{0.001}{\mapsto} \epsilon \quad \text{(S14)} \qquad betagal \overset{0.001}{\mapsto} \epsilon \quad \text{(S15)}$$

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

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

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

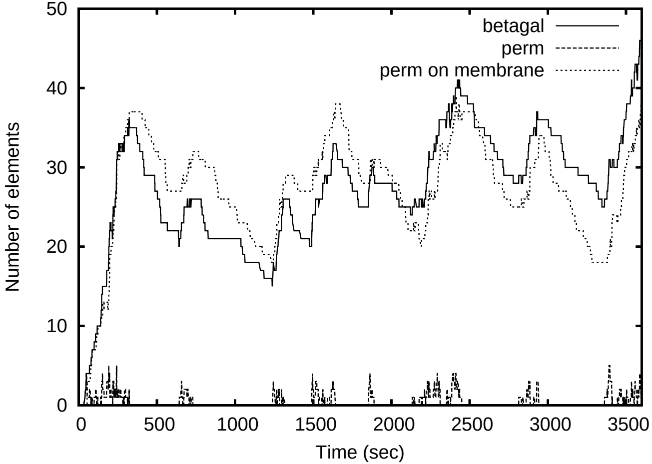
Simulation results (1)



Production of enzymes in the absence of lactose

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

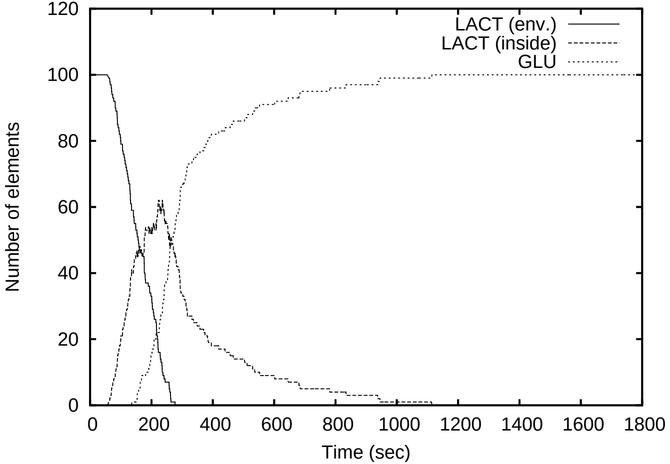
Simulation results (2)



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
2. The Calculus of Looping Sequences (CLS)
    - Definition of CLS
    - 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. Stochastic CLS
5. Ongoing work
    - Spatial CLS
    - Translation of Stochastic CLS into Maude
6. References

Spatial CLS

The spatial organization of elements may affect system dynamics

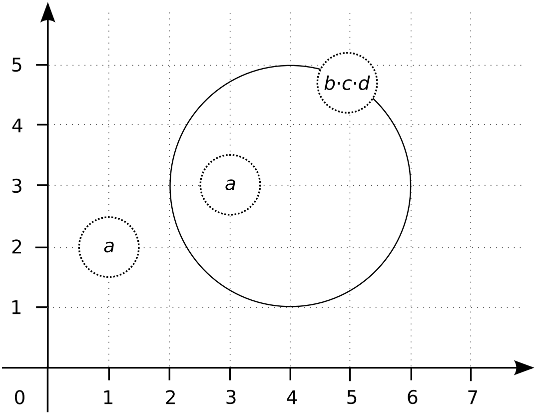
- reaction-diffusion system
- molecular crowding

We developed Spatial CLS by extending the Calculus of Looping Sequeces

Elements of Spatial CLS are spheres in a continuous space

- the containment hierarchy is reflected in the spheres
- elements can move autonomously
- interactions can depend on the spatial information of elements (position, radius, ecc.)
- rewrite rules are endowed with rates

Example of Spatial CLS term



$$T = \big(a\big)_{[(1,2),m_1],0.5} \mid \Big(\big(b \cdot c \cdot d\big)_{\cdot,0.5}\Big)^{L}_{[(4,3),m_2],2} \rfloor \big(a\big)_{[(-1,0),m_3],0.5}$$

Rewrite rules

$$R \ : \ [\, f_c \,] \quad P_L \overset{k}{\mapsto} P_R$$

k : reaction rate

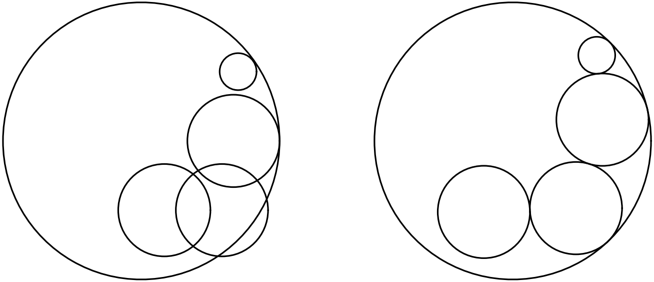
f_c : application constraints

- takes into account the spatial information of involved elements (eg. position, radius, ecc.)

Example

$$[\, dist(p,q) \leq 5 \,] \quad \big(a\big)_{[p,f_1],r_1} \mid \big(b\big)_{[q,f_2],r_2} \ \overset{0.8}{\mapsto} \ \big(c\big)_{[\frac{p+q}{2},m],r_3}$$

Resolving space conflicts



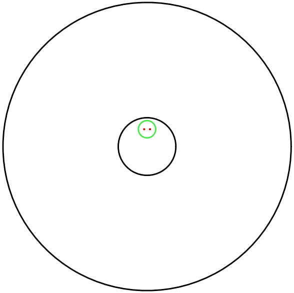
Elements push each other

- the pushing effect is modeled with a system of differential equations
- the rearranged state corresponds to its equilibrium state

Modeling cell proliferation

Initial state of the system:

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



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

Rewrite rules modeling the behavior

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

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

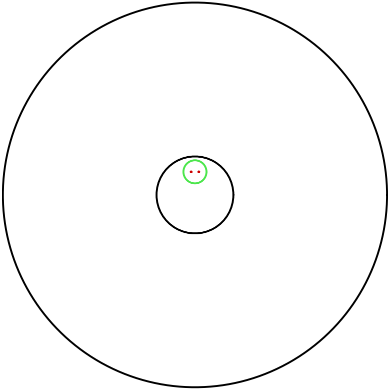
$$R_3 : [\, r = 14\,] \quad (m)^{L}_{[p,f],r} \rfloor \Big((n)^{L} \rfloor X \Big) \overset{0.5}{\mapsto} (m)^{L}_{[p,f],r} \rfloor \Big((n_{\mathrm{dup}})^{L} \rfloor X \Big)$$

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

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

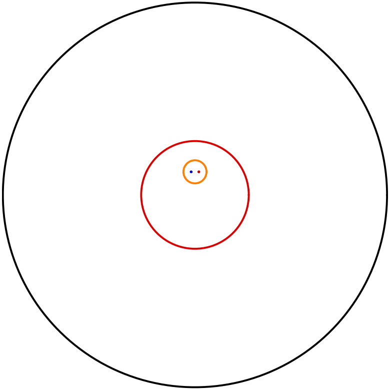
$$R_6 : (m)^{L}_{[(x,y),f],r} \rfloor \Big((n)^{L} \rfloor X \mid (n)^{L} \rfloor Y \Big) \overset{1}{\mapsto}$$
$$(m)^{L}_{[(x-5,y),f],7} \rfloor (n)^{L} \rfloor X \mid (m)^{L}_{[(x+5,y),f],7} \rfloor (n)^{L} \rfloor Y$$

Simulation



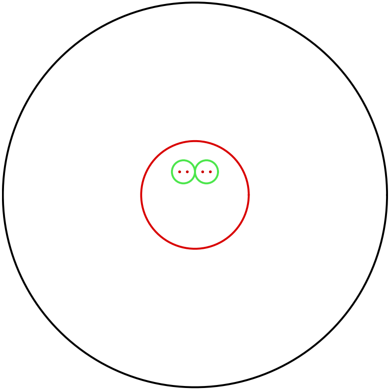
$t = 0h$

Simulation



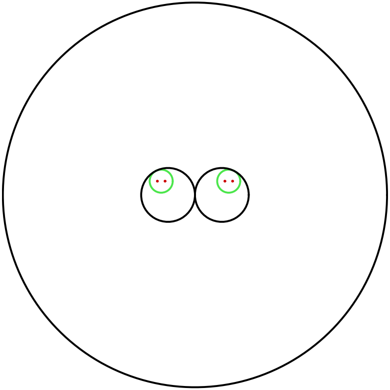
$t = 6h$

Simulation



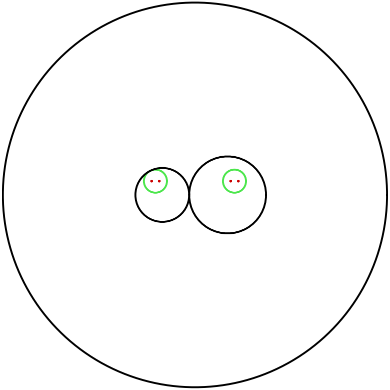
$t = 15h$

Simulation



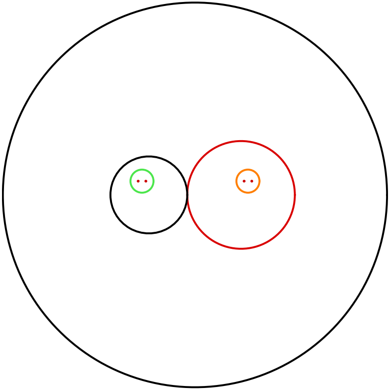
$t = 16h$

Simulation



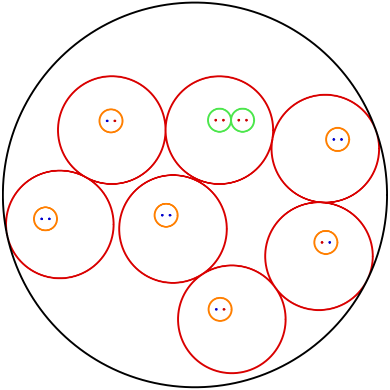
$t = 20h$

Simulation



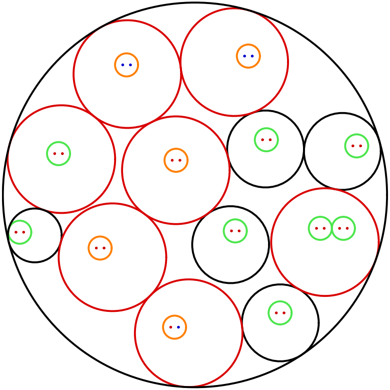
$t = 25.66h$

Simulation



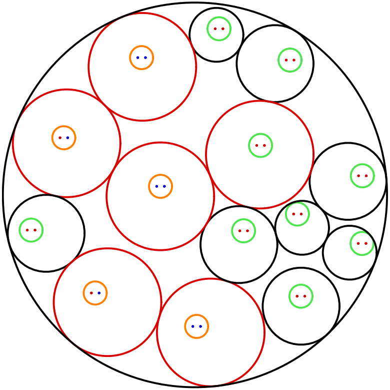
$t = 88h$

Simulation



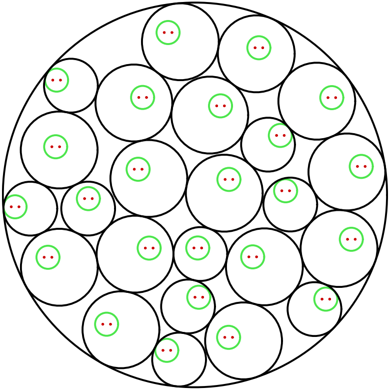
$t = 98h$

Simulation



$t = 102h$

Simulation



$t = 141h$

Outline of the talk

1. Introduction
2. The Calculus of Looping Sequences (CLS)
   - Definition of CLS
   - 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. Stochastic CLS
5. Ongoing work
   - Spatial CLS
   - Translation of Stochastic CLS into Maude
6. References

A model checker for Stochastic CLS

As candidate model checkers we have considered:

- PRISM
- Murphi
- PMaude

All of them are probabilistic/stochastic model checkers

PMaude is the most suitable

- It uses a language based on rewrite rules (rewrite logic) that eases the translation of Stochastic CLS rules

Unfortunately, the model checking module of PMaude seems not to be available

- a possible alternative: Real-Time Maude

Real-Time Maude

Maude is a specification language equipped with efficient analysis tools, which supports three modelling paradigms:

- algebraic style (via equations)
- rewrite logic (via rewrite rules)
- object oriented (via classes and messages)

Real-Time Maude extends Maude with a notion of time

- rewrite rule applications might consume (a fixed amount of) time

Real-Time Maude has two kinds of rules

- istantaneous rules:
  `crl [l] :  t => t' if cond`
- tick rules:
  `crl [l] :  t => t' in time τ if cond`

Translation of Stochastic CLS into Real-Time Maude

Real-Time Maude is not stochastic

- we will include Gillespie's simulation algorithm (slightly changed) in the translation of Stochastic CLS models
- it will be used to generate single executions of the model
- Real-Time Maude analysis tools will be applied to the simulation results

This is *statistical model checking*

- we loose exhaustivity (properties are checked on a number of runs)
- huge systems could be handled

Translation of Stochastic CLS into Real-Time Maude

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

```
(omod CLS is
  pr NAT
  sorts Elem Seq Term Loop
  subsorts Elem < Seq < Term

  op empty :  -> Seq [ctor]
  op _._ :  Seq Seq -> Seq
        [assoc gather (E e) id:  empty ctor]
  op `[_`]LContains`[_`] :  Seq Term -> Term
        [prec 41 gather (& &) ctor]
  op _|_ :  Term Term -> Term
        [assoc comm prec 45 gather (E e) id:  empty ctor]
endom)
```

Translation of Stochastic CLS into Real-Time Maude

Lotka reactions as Stochastic CLS rules

$$S_1 \xrightarrow{10} S_1|S_1 \qquad S_1|S_2 \xrightarrow{0.01} S_2|S_2 \qquad S_2 \xrightarrow{10} \epsilon$$

```
rl [ S1 ] :
    < 0 : CLSTerm | term :  (T | S1), mu :  1, step :  4 >
=>
    < 0 : CLSTerm | term :  (T | S1 | S1), step :  5 >

rl [ S2 ] :
    < 0 : CLSTerm | term :  (T | S1 | S2), mu :  2, step :  4 >
=>
    < 0 : CLSTerm | term :  (T | S2 | S2), step :  5 >

rl [ S3 ] :
    < 0 : CLSTerm | term :  (T | S2), mu :  3, step :  4 >
=>
    < 0 : CLSTerm | term :  T, step :  5 >
```

Analysis example: statistical model checking

Initialisation of 100 stochastic simulations

```
rl [ initialise1 ] :
  < step :  0 >
=>
  < seed :  random(1), step :  1 >
⋮


⋮
rl [ initialise100 ] :
  < step :  0 >
=>
  < seed :  random(100), step :  1 >
```

Analysis example: statistical model checking

Verification of properties espressed as LTL formulas. Some state formulas:

vanished(T) indicates that term T has vanished from the system,

IsLessThan(T,T') indicates that the occurences of term T are less than the occurences of T'.

Starting with 4 x S_1 and 4 x S_2 we prove

- that S_2 will eventually disappear (i.e. $\Diamond$vanished(S_2))
- that the amount of S_2 will eventually become less than the amount of S_1 (i.e. $\Diamond$IsLessThan(S_2,S_1))

```
(mc INIT({S1}4 | {S2}4) |=t <> vanished(S2) in time<=1 .)

Result Bool :  true

(mc INIT({S1} 4 | {S2} 4) |=t <> IsLessThan(S2,S1) in time<=1 .)

Result Bool :  true
```

References

- P. Milazzo, *Formal Modeling in Systems Biology. A Theoretical Computer Science Approach*, VDM Verlag Dr. Muller, Saarbruecken, 2008.
- P. Milazzo. *Qualitative and Quantitative Formal Modeling of Biological Systems*, PhD Thesis, Università di Pisa, 2007.
- R. Barbuti, A. Maggiolo-Schettini, P. Milazzo and A. Troina. *A Calculus of Looping Sequences for Modelling Microbiological Systems*. Fundamenta Informaticae 72, 21-35, 2006.
- R. Barbuti, A. Maggiolo-Schettini and P. Milazzo. *Extending the Calculus of Looping Sequences to Model Protein Interaction at the Domain Level*. Symp. on Bioinf. Research and Applic. (ISBRA'07), LNBI 4463, 638-649, Springer, 2007.
- R. Barbuti, A. Maggiolo-Schettini, P. Milazzo and A. Troina. *The Calculus of Looping Sequences for Modeling Biological Membranes*. Invited paper at the 8th Work. on Membrane Computing (WMC8), LNCS 4860, 54-76, Springer, 2007.
- R. Barbuti, A. Maggiolo-Schettini, P. Milazzo and A. Troina. *Bisimulations in Calculi Modelling Membranes*. Formal Aspects of Computing 20, 351-377, 2008.
- T.A. Basuki, A. Cerone, P. Milazzo. *Translating Stochastic CLS into Maude*. Work. on Membr. Computing and Biol. Inspired Calculi (MeCBIC'08), 2008.
- R. Barbuti, A. Maggiolo-Schettini, P. Milazzo, P. Tiberi and A. Troina. *Stochastic CLS for the Modeling and Simulation of Biological Systems*. Transaction on Computational Systems Biology, in press.