

# Bisimulation Congruences in the Calculus of Looping Sequences

Roberto Barbuti    Andrea Maggiolo Schettini  
Paolo Milazzo    Angelo Troina

Dipartimento di Informatica, Università di Pisa, Italy

Tunis – ICTAC 2006

# Introduction

Formal models for systems of interactive components can be easily used or adapted for the modelling of biological phenomena

- Examples: Petri Nets,  $\pi$ -calculus, Mobile Ambients

The modelling of biological systems allows:

- 1 the development of simulators
- 2 the verification of properties

We defined the Calculus of Looping Sequences (CLS): a formalism to describe biochemical systems in cells

In this talk:

- 1 we recall the definition of CLS
- 2 we present bisimulation relations for CLS
- 3 we show the CLS model of a gene regulation process in E. Coli

# 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:

$$\begin{aligned} T & ::= S \mid (S)^L \mid T \mid T \\ S & ::= \epsilon \mid a \mid S \cdot S \end{aligned}$$

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

The operators are:

$S \cdot S$  : Sequencing

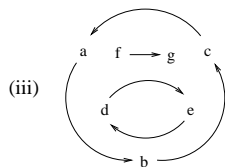
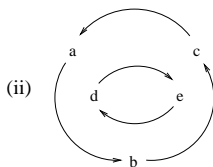
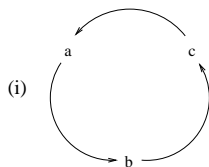
$(S)^L$  : Looping ( $S$  is closed and it can rotate)

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

$T \mid T$  : Parallel composition (juxtaposition)

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

## Example of Terms



$$(i) \quad (a \cdot b \cdot c)^L \mid \epsilon$$

$$(ii) \quad (a \cdot b \cdot c)^L \mid (d \cdot e)^L \mid \epsilon$$

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

# 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:

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

$$T_1 \mid T_2 \equiv_T T_2 \mid T_1 \quad T_1 \mid (T_2 \mid T_3) \equiv_T (T_1 \mid T_2) \mid T_3$$
$$T \mid \epsilon \equiv_T T \quad (\epsilon)^L \rfloor \epsilon \equiv_T \epsilon \quad (S_1 \cdot S_2)^L \rfloor T \equiv_T (S_2 \cdot S_1)^L \rfloor T$$

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

# Dynamics of the Calculus (1)

Let  $\mathcal{T}_{\mathcal{V}}$  be the set of terms which may contain variables of three kinds:

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

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

A **Rewrite Rule** is a pair  $(T, T')$ , denoted  $T \mapsto T'$ , where:

- $T, T' \in \mathcal{T}_{\mathcal{V}}$
- variables in  $T'$  are a subset of those in  $T$

A rule  $T \mapsto T'$  can be applied to all terms  $T\sigma$ .

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

# 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:

$$\mathcal{C} ::= \square \mid \mathcal{C} \mid T \mid T \mid \mathcal{C} \mid (S)^L \mid \mathcal{C}$$

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

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

$$\mathcal{C}_P ::= \square \mid \mathcal{C}_P \mid T \mid T \mid \mathcal{C}_P.$$

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  $\mathcal{R} \subseteq \mathfrak{R}$ , the **labeled semantics** of CLS is the labeled transition system given by the following inference rules:

$$\begin{array}{c} \text{(rule\_appl)} \quad \frac{T \mapsto T' \in \mathcal{R} \quad C[T''] \equiv T\sigma \quad T'' \neq \epsilon \quad \sigma \in \Sigma \quad C \in \mathcal{C}}{T'' \xrightarrow{C} T'\sigma} \\ \\ \text{(cont)} \quad \frac{T \xrightarrow{\square} T'}{(S)^L \rfloor T \xrightarrow{\square} (S)^L \rfloor T'} \quad \text{(par)} \quad \frac{T \xrightarrow{C} T' \quad C \in \mathcal{C}_P}{T \mid T'' \xrightarrow{C} T' \mid T''} \end{array}$$

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

Rule *(rule\_appl)* describes the (potential) application of a rule.

- $T'' \neq \epsilon$  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  $\epsilon \not\xrightarrow{a \mid b}$ .

## Labeled Semantics (3)

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

$$\begin{array}{c} \text{(rule\_appl)} \frac{T \mapsto T' \in \mathcal{R} \quad C[T''] \equiv T\sigma \quad T'' \neq \epsilon \quad \sigma \in \Sigma \quad C \in \mathcal{C}}{T'' \xrightarrow{C} T'\sigma} \\ \\ \text{(cont)} \frac{T \xrightarrow{\square} T'}{(S)^L \mid T \xrightarrow{\square} (S)^L \mid T'} \quad \text{(par)} \frac{T \xrightarrow{C} T' \quad C \in \mathcal{C}_P}{T \mid T'' \xrightarrow{C} T' \mid T''} \end{array}$$

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

Rule (cont) propagates  $\square$ -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  $(d)^L \mid a \not\xrightarrow{\square \mid b}$ .

## Labeled Semantics (4)

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

$$\begin{array}{c}
 \text{(rule\_appl)} \quad \frac{T \mapsto T' \in \mathcal{R} \quad C[T''] \equiv T\sigma \quad T'' \neq \epsilon \quad \sigma \in \Sigma \quad C \in \mathcal{C}}{T'' \xrightarrow{C} T'\sigma} \\
 \\
 \text{(cont)} \quad \frac{T \xrightarrow{\square} T'}{(S)^L \rfloor T \xrightarrow{\square} (S)^L \rfloor T'} \quad \quad \text{(par)} \quad \frac{T \xrightarrow{C} T' \quad C \in \mathcal{C}_P}{T \mid T'' \xrightarrow{C} T' \mid T''}
 \end{array}$$

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  $b \xrightarrow{(a)^L \rfloor \square} c$ , but  $b \mid d \not\xrightarrow{(a)^L \rfloor \square}$  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.  $T_2 \xRightarrow{C} T'_2$  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 *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:

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

We have that  $a \sim d$ , because

$$\begin{aligned} a &\xrightarrow{\square|b} c \xrightarrow{\square} e \xrightarrow{\square} e \xrightarrow{\square} \dots \\ d &\xrightarrow{\square|b} e \xrightarrow{\square} e \xrightarrow{\square} \dots \end{aligned}$$

and  $f \approx d$ , because

$$f \xrightarrow{\square} a \xrightarrow{\square|b} c \xrightarrow{\square} e \xrightarrow{\square} e \xrightarrow{\square} \dots$$

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

$$e \xrightarrow{\square} e \xrightarrow{\square} e \xrightarrow{\square} \dots$$

## 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  $(\mathcal{R}_2, T'_2)R(\mathcal{R}_1, T'_1)$ .

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.  $\mathcal{R}_2 : T_2 \xRightarrow{C} T'_2$  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  $(T'_2, \mathcal{R}_2)R(T'_1, \mathcal{R}_1)$

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

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

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

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

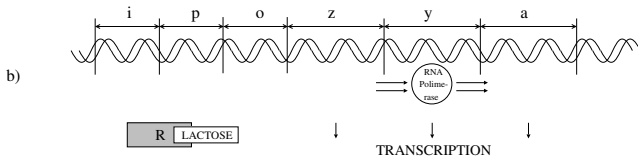
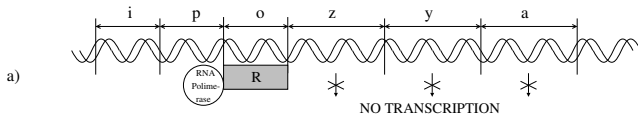
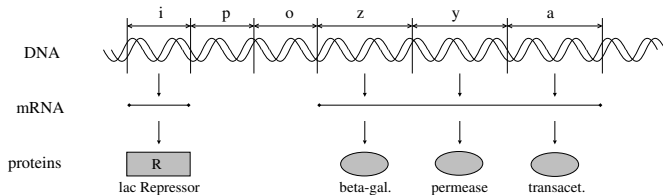
and  $\langle b, \mathcal{R}_1 \rangle \approx \langle d, \mathcal{R}_2 \rangle$ , because

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

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

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

# The Lactose Operon in E.coli (1)



## The Lactose Operon in E.coli (2)

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

Rules for DNA transcription/translation:

$$lacI \cdot \tilde{x} \longrightarrow lacI' \cdot \tilde{x} \mid repr \quad (R1)$$

$$polym \mid \tilde{x} \cdot lacP \cdot \tilde{y} \longrightarrow \tilde{x} \cdot PP \cdot \tilde{y} \quad (R2)$$

$$\tilde{x} \cdot PP \cdot lacO \cdot \tilde{y} \longrightarrow \tilde{x} \cdot lacP \cdot PO \cdot \tilde{y} \quad (R3)$$

$$\tilde{x} \cdot PO \cdot lacZ \cdot \tilde{y} \longrightarrow \tilde{x} \cdot lacO \cdot PZ \cdot \tilde{y} \quad (R4)$$

$$\tilde{x} \cdot PZ \cdot lacY \cdot \tilde{y} \longrightarrow \tilde{x} \cdot lacZ \cdot PY \cdot \tilde{y} \mid betagal \quad (R5)$$

$$\tilde{x} \cdot PY \cdot lacA \longrightarrow \tilde{x} \cdot lacY \cdot PA \mid perm \quad (R6)$$

$$\tilde{x} \cdot PA \longrightarrow \tilde{x} \cdot lacA \mid transac \mid polym \quad (R7)$$

## The Lactose Operon in E.coli (3)

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

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

$$repr \mid \tilde{x} \cdot lacO \cdot \tilde{y} \longrightarrow \tilde{x} \cdot RO \cdot \tilde{y} \quad (R8)$$

$$LACT \mid (m \cdot \tilde{x})^L \rfloor X \longrightarrow (m \cdot \tilde{x})^L \rfloor (X \mid LACT) \quad (R9)$$

$$\tilde{x} \cdot RO \cdot \tilde{y} \mid LACT \longrightarrow \tilde{x} \cdot lacO \cdot \tilde{y} \mid RLACT \quad (R10)$$

$$(\tilde{x})^L \rfloor (perm \mid X) \longrightarrow (perm \cdot \tilde{x})^L \rfloor X \quad (R11)$$

$$LACT \mid (perm \cdot \tilde{x})^L \rfloor X \longrightarrow (perm \cdot \tilde{x})^L \rfloor (LACT \mid X) \quad (R12)$$

$$betagal \mid LACT \longrightarrow betagal \mid GLU \mid GAL \quad (R13)$$

# The Lactose Operon in E.coli (4)

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

Example:

$Ecoli \mid LACT \mid LACT$

$\rightarrow^* (m)^L \rfloor (lacI' \cdot lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA \mid polym \mid repr) \mid LACT \mid LACT$

$\rightarrow^* (m)^L \rfloor (lacI' \cdot lacP \cdot RO \cdot lacZ \cdot lacY \cdot lacA \mid polym) \mid LACT \mid LACT$

$\rightarrow^* (m)^L \rfloor (lacI' \cdot lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA \mid polym \mid RLACT) \mid LACT$

$\rightarrow^* (perm \cdot m)^L \rfloor (lacI' - A \mid betagal \mid transac \mid polym \mid RLACT) \mid LACT$

$\rightarrow^* (perm \cdot m)^L \rfloor (lacI' - A \mid betagal \mid transac \mid polym \mid RLACT \mid GLU \mid GAL)$

# Applying Bisimulations (1)

It can be easily proved that

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

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

## Applying Bisimulations (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

$$\tilde{x} \cdot RO \cdot \tilde{y} \mid LACT \longrightarrow \tilde{x} \cdot lacO \cdot \tilde{y} \mid RLACT \quad (R10)$$

with

$$\begin{aligned} (\tilde{w})^L \rfloor (\tilde{x} \cdot RO \cdot \tilde{y} \mid LACT \mid X) \mid START &\longrightarrow \\ (\tilde{w})^L \rfloor (\tilde{x} \cdot lacO \cdot \tilde{y} \mid RLACT \mid X) &\quad (R10bis) \end{aligned}$$

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

$$\begin{array}{ll} T_1 \mid LACT \longrightarrow T_2 & (R1') & T_2 \mid START \longrightarrow T_3 & (R3') \\ T_2 \mid LACT \longrightarrow T_2 & (R2') & T_3 \mid LACT \longrightarrow T_3 & (R4') \end{array}$$

that is a system satisfying the property.

# Conclusions

The Calculus of Looping Sequences can be used to describe biological systems

The bisimulation relations we have defined can be used

- to find equivalent reduced models
- to verify properties

If we consider models in which the same set of rewrite rules is used, strong and weak bisimulations are congruences.

We used bisimulations on a model of a real biological phenomenon:

- to find an equivalent reduced model
- to verify a causality property