

Simulation of Kohn's Molecular Interaction Maps Through Translation into Stochastic CLS+

Roberto Barbuti¹ Daniela Lepri² Andrea Maggiolo-Schettini¹
Paolo Milazzo¹ Giovanni Pardini¹ Aureliano Rama¹

¹Dipartimento di Informatica
Università di Pisa, Italy

²Institutt for Informatikk
Universitetet i Oslo, Norway

PSI'09

Novosibirsk, Akademgorodok – June 18, 2009

Outline

- 1 Introduction
- 2 Molecular Interaction Maps
- 3 The Calculus of Looping Sequences
- 4 Translating MIMs to CLS
- 5 Conclusions

- 1 Introduction
- 2 Molecular Interaction Maps
- 3 The Calculus of Looping Sequences
- 4 Translating MIMs to CLS
- 5 Conclusions

Biological systems

Biology is “the science that studies living organisms”. It includes:

- the analysis of molecular interactions at the level of proteins, enzymes, etc;
- the study of cells and tissues;
- the study of the origin, development and distribution of animals and plants.

Biological systems are composed by a huge number of interactive elements, exhibiting complex behaviours

- ▶ Formalisms for concurrent software systems can applied to Biology

Systems Biology focuses on studying “*biological systems*” as a whole

- Formal methods in Systems Biology
 - precise and unambiguous descriptions
 - compositionality
 - automatic analysis techniques
 - development of simulators
- ▶ π -calculus, *Calculus of Looping Sequences* CLS, Bio-PEPA, κ -calculus, Petri nets, Pathway logic, . . .
- ▶ Analysis methods for studying qualitative/quantitative properties
 - based on different models (e.g. ODEs, Markov chains)
 - static analysis
 - model checking
 - simulation (*Gillespie*)

Molecular Interaction Maps (MIMs)

- graphical notation for bioregulatory networks
- no formal semantics

In this work, we provide:

- a **formalization** of the structure of MIM diagrams
- a **translation** of MIMs to executable models in the **Calculus of Looping Sequences**
 - giving an unambiguous interpretation of MIMs

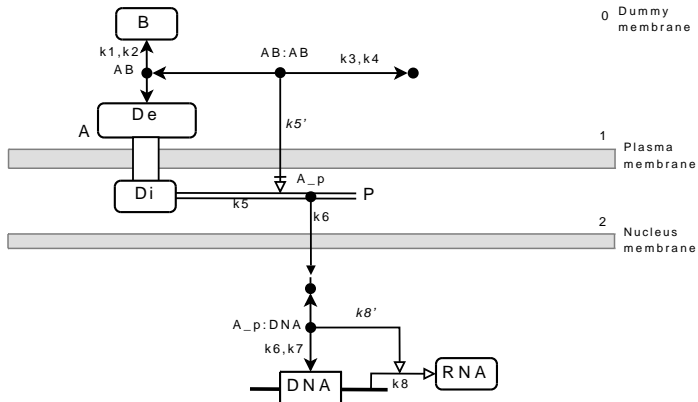
- ▶ By this work we provide Molecular Interaction Maps with all the tools developed for CLS
 - Analysers, simulators, ...

Outline

- 1 Introduction
- 2 Molecular Interaction Maps**
- 3 The Calculus of Looping Sequences
- 4 Translating MIMs to CLS
- 5 Conclusions

Molecular Interaction Maps

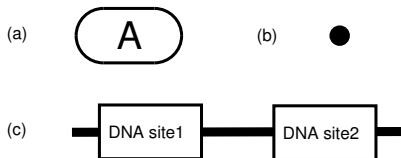
- graphical notation for bioregulatory networks, developed by *Kohn*
- depict possible interactions among species
 - different symbols for denoting different reactions
- but they lack formality



MIM symbols

Species

Species

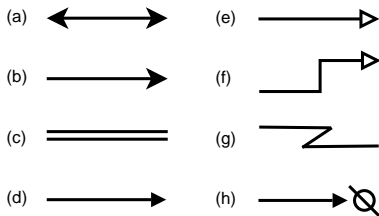


- (a) named molecular species
- (b) bullet for compound species, resulting from reactions
- (c) DNA sites

MIM symbols

Reaction symbols

Reaction symbols, connecting species nodes

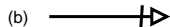


- ▶ Each symbol represents a different kind of interaction

MIM symbols

Contingency symbols

Contingency symbols, pointing to other reactions



(a) stimulation (rate increase)

(b) necessity

(c) inhibition

(d) catalysis

Outline

- 1 Introduction
- 2 Molecular Interaction Maps
- 3 The Calculus of Looping Sequences**
- 4 Translating MIMs to CLS
- 5 Conclusions

Syntax of CLS+

Definition

Terms \mathcal{T} , **branes** \mathcal{B} and **sequences** \mathcal{S} of CLS+ are defined as:

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

where a is a generic symbol of an alphabet \mathcal{E} .

The operators are:

$S \cdot S$: sequencing

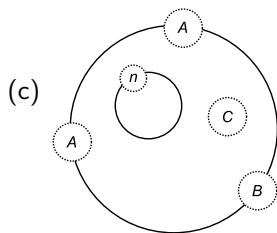
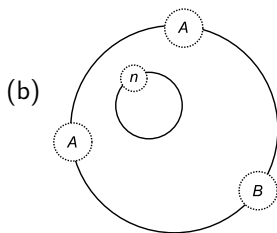
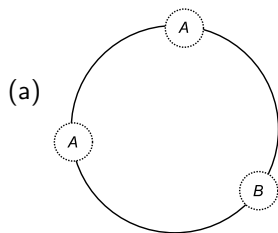
$(B)^L$: looping

$T_1 \rfloor T_2$: containment

$T_1 \mid T_2$: parallel composition

- Structural congruence

Examples of CLS+ terms



(a) $(A|A|B)^L \rfloor \epsilon$

(b) $(A|A|B)^L \rfloor (n)^L \rfloor \epsilon$

(c) $(A|A|B)^L \rfloor (C|(n)^L \rfloor \epsilon)$

Definition

A **pattern** is a **term with variables** of the following types:

- sequence variables $\tilde{x}, \tilde{y}, \dots \in SV$
- symbol variables $x, y, \dots \in \mathcal{X}$
- term variables $X, Y, \dots \in TV$
- brane variables $\bar{X}, \bar{Y}, \dots \in BV$

The set of all variables is denoted by \mathcal{V}

- ▶ A pattern P may be instantiated into a ground term $P\sigma$
 - the **instantiation function** $\sigma : \mathcal{V} \rightarrow \mathcal{T}$ associates a value to each variable in P

Definition

A **rewrite rule** $P' \xrightarrow{f} P''$ is such that:

- P' and P'' are patterns
- $f : \Sigma \rightarrow \mathbb{R}^+$ is the **rate function** of the rewrite rule
 - gives the **rate value** for instantiations of the rule

- ▶ A rewrite rule $P' \xrightarrow{f} P''$ can transform a subterm $P'\sigma$ into $P''\sigma$
- ▶ The semantic of the calculus is a **Continuous Time Markov Chain**
 - state** : CLS term
 - transition** : application of one rewrite rule
 - transition rate = rate value \times occurrence count

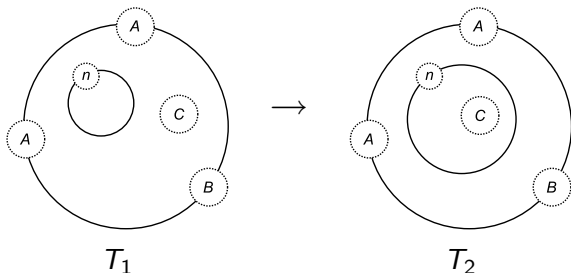
Example of rewrite rules

Rewrite rule R can be applied to T_1 yielding T_2

$$R : (B|\overline{X})^L \rfloor (C|(n)^L \rfloor Y) \xrightarrow{f_1} (B|\overline{X})^L \rfloor (n)^L \rfloor (C|Y)$$

$$T_1 = (A|A|B)^L \rfloor (C|(n)^L \rfloor \epsilon)$$

$$T_2 = (A|A|B)^L \rfloor (n)^L \rfloor C$$



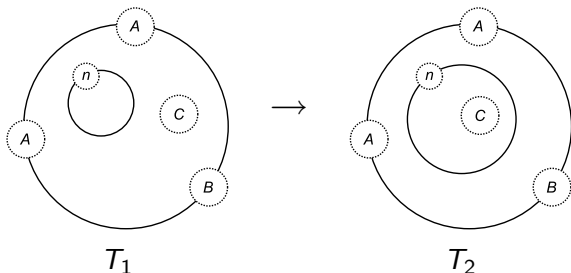
Example of rewrite rules

Rewrite rule R can be applied to T_1 yielding T_2

$$R : (B|\bar{X})^L \rfloor (C|(n)^L \rfloor Y) \xrightarrow{f_1} (B|\bar{X})^L \rfloor (n)^L \rfloor (C|Y)$$

$$T_1 = (A|A|B)^L \rfloor (C|(n)^L \rfloor \epsilon)$$

$$T_2 = (A|A|B)^L \rfloor (n)^L \rfloor C$$



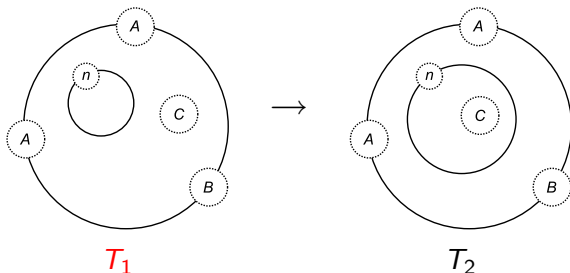
Example of rewrite rules

Rewrite rule R can be applied to T_1 yielding T_2

$$R : (B|\overline{X})^L \rfloor (C|(n)^L \rfloor Y) \xrightarrow{f_1} (B|\overline{X})^L \rfloor (n)^L \rfloor (C|Y)$$

$$T_1 = (A|A|B)^L \rfloor (C|(n)^L \rfloor \epsilon)$$

$$T_2 = (A|A|B)^L \rfloor (n)^L \rfloor C$$



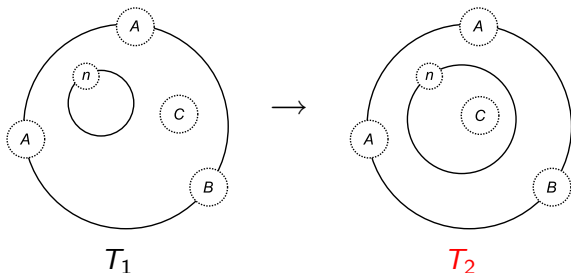
Example of rewrite rules

Rewrite rule R can be applied to T_1 yielding T_2

$$R : (B|\overline{X})^L \rfloor (C|(n)^L \rfloor Y) \xrightarrow{f_1} (B|\overline{X})^L \rfloor (n)^L \rfloor (C|Y)$$

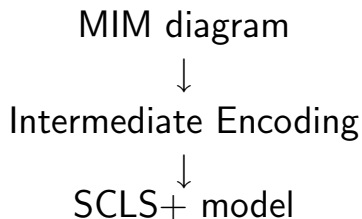
$$T_1 = (A|A|B)^L \rfloor (C|(n)^L \rfloor \epsilon)$$

$$T_2 = (A|A|B)^L \rfloor (n)^L \rfloor C$$



Outline

- 1 Introduction
- 2 Molecular Interaction Maps
- 3 The Calculus of Looping Sequences
- 4 Translating MIMs to CLS**
- 5 Conclusions



- ▶ The *Intermediate Encoding* is constructed by inspecting the MIM diagram

Intermediate Encoding

The Intermediate Encoding of the MIM formally specifies:

- the membrane structure of the MIM diagram
- the species appearing in the MIM diagram, and their **initial quantity** in each compartment
- the possible reactions among species, and their contingencies

Intermediate Encoding

Membrane Structure \mathcal{MS} : set of tuples $\langle i, I \rangle \in \mathbb{N} \times \mathcal{P}(\mathbb{N})$

- i : membrane id
- I : ids of children membranes

Set of Species \mathcal{SS} : $\langle S, (i, j), q \rangle \in \mathcal{S} \times (\mathbb{N} \times \{0, 1\}) \times \mathbb{N}$

- S : CLS+ sequence
- (i, j) : membrane id and position (on the surface or inside the membrane)
- q : quantity

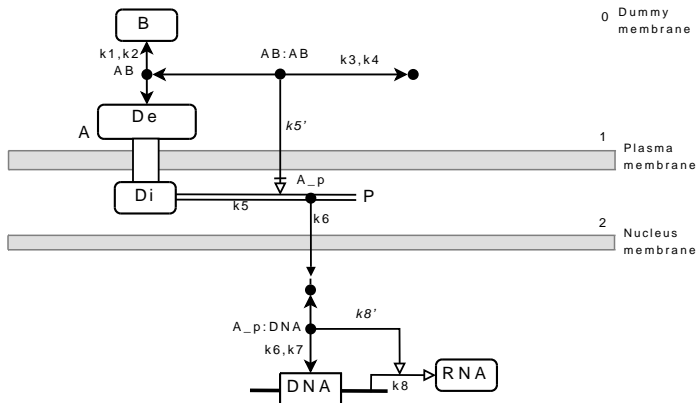
Set of Contingecies \mathcal{CS} : $\langle C, k \rangle \in \mathcal{SS} \times \mathbb{R}$

- C : required species
- k : corresponding rate of reaction

Set of Reactions \mathcal{RS} : $\langle CS, R, P \rangle \in \mathcal{CS} \times \mathcal{SS} \times \mathcal{SS}$

- CS : contingencies
- R : reactants
- P : products

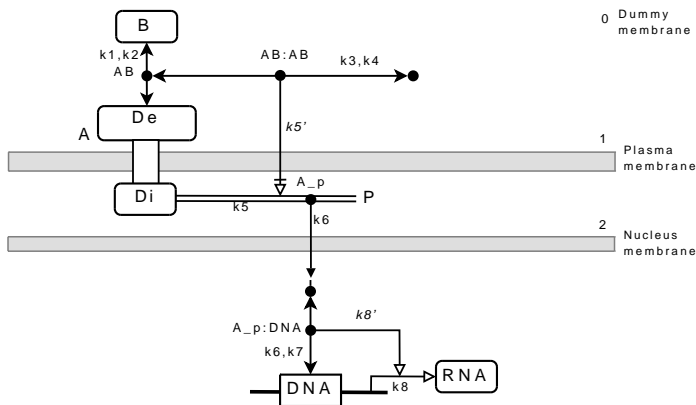
Membrane Structure



In this MIM, this is the encoding of the Membrane Structure:

$$MS = \{\langle 0, \{1\} \rangle, \langle 1, \{2\} \rangle, \langle 2, \emptyset \rangle\}$$

MS and Initial State

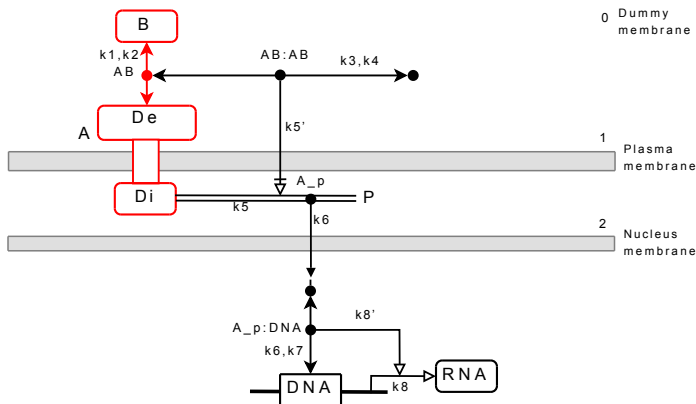


$$MS = \{\langle 0, \{1\} \rangle, \langle 1, \{2\} \rangle, \langle 2, \emptyset \rangle\}$$

A possible initial state is the following:

$$S_0 = \{\langle B, (0, 1), 2 \rangle, \langle A, (1, 0), 3 \rangle, \langle DNA, (2, 1), 1 \rangle\}$$

MS and Reactions 1

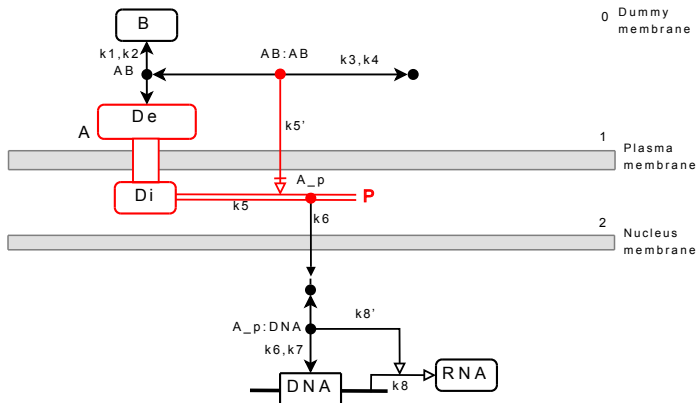


$$MS = \{\langle 0, \{1\} \rangle, \langle 1, \{2\} \rangle, \langle 2, \emptyset \rangle\}$$

$$R_1 = \{\langle \emptyset, k_1 \rangle, \langle \langle A, (1, 0), 1 \rangle, \langle B, (0, 1), 1 \rangle \rangle, \langle \langle AB, (1, 0), 1 \rangle \rangle\}$$

$$R_2 = \{\langle \emptyset, k_2 \rangle, \langle \langle AB, (1, 0), 1 \rangle \rangle, \langle \langle A, (1, 0), 1 \rangle, \langle B, (0, 1), 1 \rangle \rangle\}$$

MS and Reaction 2



$$MS = \{\langle 0, \{1\} \rangle, \langle 1, \{2\} \rangle, \langle 2, \emptyset \rangle\}$$

$$R_3 = \{\langle \emptyset, k_5 \rangle, \langle \langle AB:AB, (0, 1), 1 \rangle, k_5' \rangle, \langle A, (1, 0), 1 \rangle, \langle pA, (1, 0), 1 \rangle \}$$

Producing the SCLS+ Term from Intermediate Encoding

The Intermediate Encoding Sets we just prepared have all the information needed to produce the SCLS+ Terms that can simulate our original MIM.

Translate the Initial State

- Set of Species Translation Function ϕ
- Term Translation Function $[[\cdot]]$

Translate all the Interactions

- Parallel Pattern Builder $[[\cdot]]_{pp}$
- Reaction Translation Function $[[\cdot]]_r$

Given the MS and Initial State of our example:

$$MS = \{\langle 0, \{1\} \rangle, \langle 1, \{2\} \rangle, \langle 2, \emptyset \rangle\}$$

$$S_0 = \{\langle B, (0, 1), 2 \rangle, \langle A, (1, 0), 3 \rangle, \langle DNA, (2, 1), 1 \rangle\}$$

the translation produces this SCLS+ term:

$$B | B | (1 | A | A | A)^L \rfloor \left((2)^L \rfloor DNA \right)$$

Interactions 1

Given the first reactions of our example:

$$MS = \{\langle 0, \{1\} \rangle, \langle 1, \{2\} \rangle, \langle 2, \emptyset \rangle\}$$

$$R_1 = \{\langle \emptyset, k_1 \rangle, \langle A, (1, 0), 1 \rangle, \langle B, (0, 1), 1 \rangle, \langle AB, (1, 0), 1 \rangle\}$$

$$R_2 = \{\langle \emptyset, k_2 \rangle, \langle AB, (1, 0), 1 \rangle, \langle A, (1, 0), 1 \rangle, \langle B, (0, 1), 1 \rangle\}$$

the translation produces these rewrite rules:

$$(R_1) \quad (0 | X_0)^L \rfloor \left(B | \left((1 | A | X_2)^L \rfloor X_3 \right) | X_1 \right) \xrightarrow{f_1} \\ (0 | X_0)^L \rfloor \left(\left((1 | AB | X_2)^L \rfloor X_3 \right) | X_1 \right)$$

$$(R_2) \quad (0 | X_0)^L \rfloor \left(\left((1 | AB | X_2)^L \rfloor X_3 \right) | X_1 \right) \xrightarrow{f_2} \\ (0 | X_0)^L \rfloor \left(B | \left((1 | A | X_2)^L \rfloor X_3 \right) | X_1 \right)$$

- ▶ Rate functions f_1, f_2 are derived from reactions rates k_1, k_2

Interactions 2

The second interaction shows us an example of contingency (two different k associated with the same reaction)

$$MS = \{\langle 0, \{1\} \rangle, \langle 1, \{2\} \rangle, \langle 2, \emptyset \rangle\}$$

$$R_3 = \{\langle \emptyset, k_5 \rangle, \langle \langle AB:AB, (0, 1), 1 \rangle, k'_5 \rangle, \langle \langle A, (1, 0), 1 \rangle, \langle pA, (1, 0), 1 \rangle \rangle\}$$

Translating this intermediate encoding we obtain this rewrite rule:

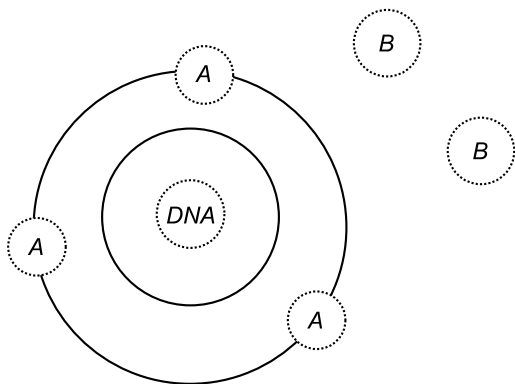
$$(R_3) \quad (0 | X_0)^L \rfloor ((1 | A | X_2)^L \rfloor X_3 | X_1) \xrightarrow{f} (0 | X_0)^L \rfloor ((1 | pA | X_2)^L \rfloor X_3 | X_1)$$

$$basalRate(\sigma) = \begin{cases} k'_5 & \text{if } \sigma(X_1) \equiv AB:AB \mid T \\ k_5 & \text{otherwise} \end{cases}$$

where

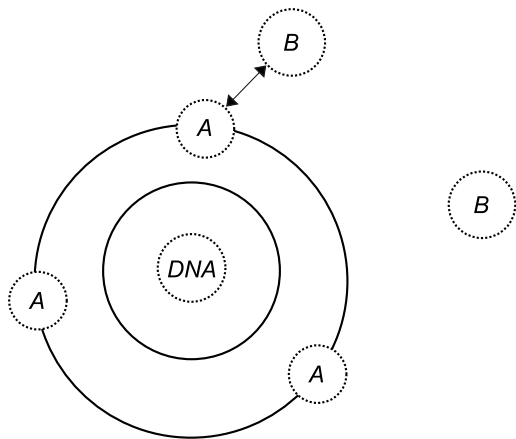
$$f(P\sigma) = basalRate(\sigma) \cdot \#(\text{reactant combinations})$$

Term Evolution



$$(1 | A | A | A)^L \rfloor \left((2)^L \rfloor DNA \right) | B | B$$

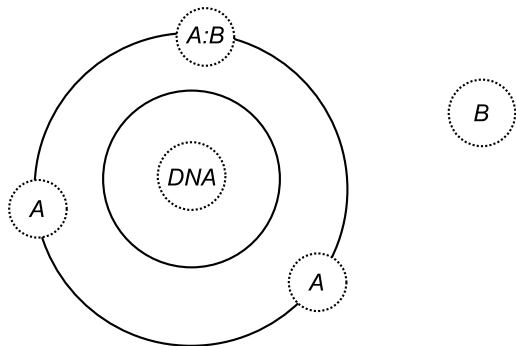
Term Evolution



$$(1 | A | A | A)^L \rfloor \left((2)^L \rfloor DNA \right) | B | B$$

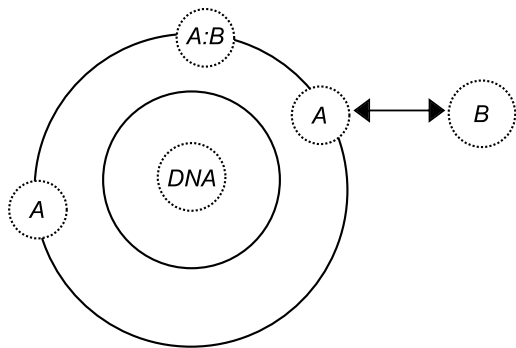
$$\text{reaction rate} = k_1 \cdot \binom{3}{1} \cdot \binom{2}{1}$$

Term Evolution



$$(1 \mid \mathbf{A:B} \mid \mathbf{A} \mid \mathbf{A})^L \mid ((2)^L \mid \mathbf{DNA}) \mid B$$

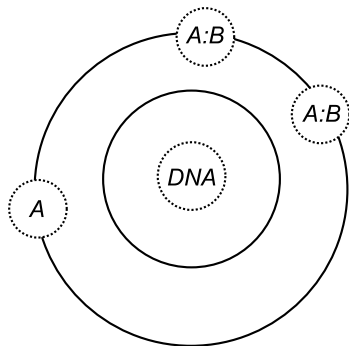
Term Evolution



$$(1 | A : B | A | A)^L \left| \left((2)^L \right| DNA \right) | B$$

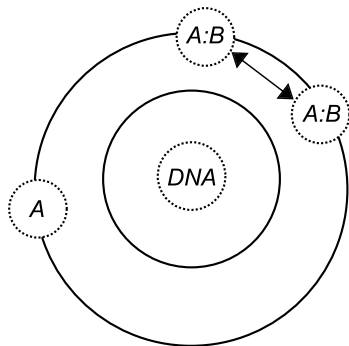
$$\text{reaction rate} = k_1 \cdot \binom{2}{1} \cdot \binom{1}{1}$$

Term Evolution



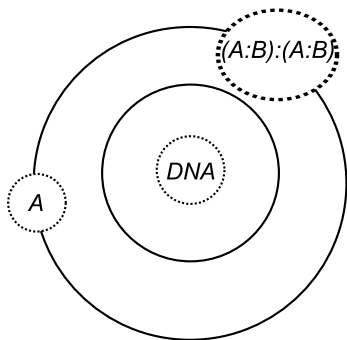
$$(1 \mid A:B \mid A:B \mid A)^L \mid \left((2)^L \mid DNA \right)$$

Term Evolution



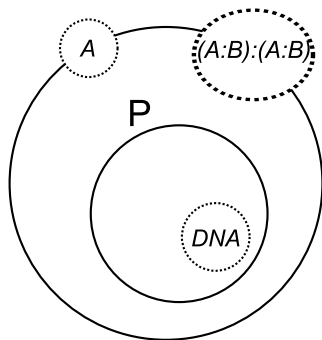
$$(1 \mid A:B \mid A:B \mid A)^L \mid ((2)^L \mid DNA)$$

Term Evolution



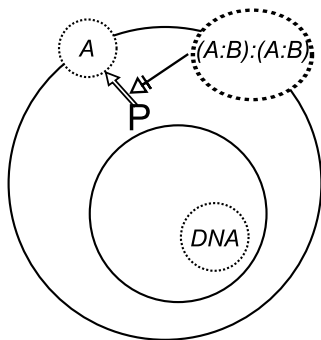
$$(1 \mid (A:B):(A:B) \mid A)^L \mid ((2)^L \mid DNA)$$

Term Evolution



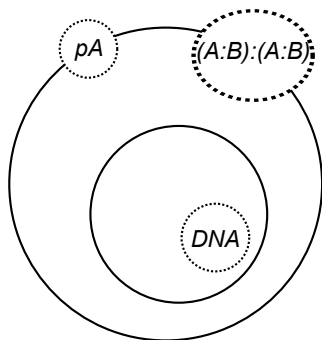
$$(1 \mid (A:B):(A:B) \mid A)^L \mid ((2)^L \mid DNA)$$

Term Evolution



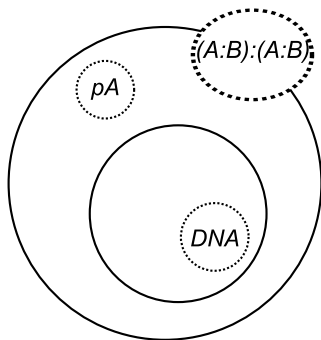
$$(1 \mid (A:B):(A:B) \mid A)^L \mid ((2)^L \mid DNA)$$

Term Evolution



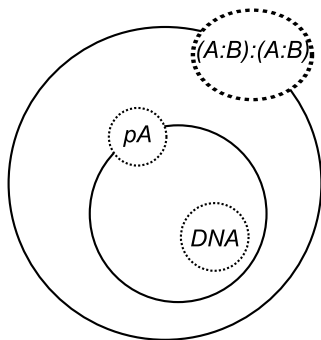
$$(1 \mid (A:B):(A:B) \mid \overline{pA})^L \mid ((2)^L \mid DNA)$$

Term Evolution



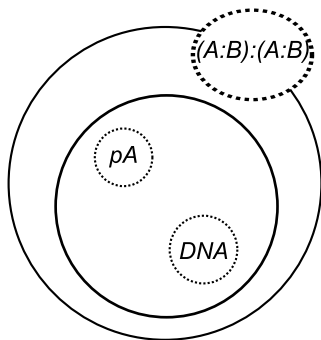
$$(1 \mid (A:B):(A:B))^L \mid (\overline{pA} \mid (2)^L \mid DNA)$$

Term Evolution



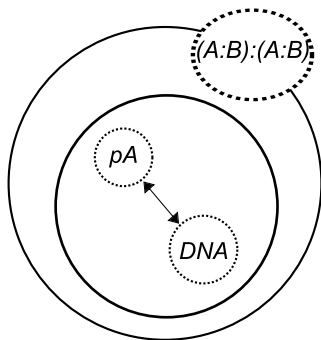
$$(1 \mid (A:B):(A:B))^L \mid \left((2 \mid \overline{pA})^L \mid DNA \right)$$

Term Evolution



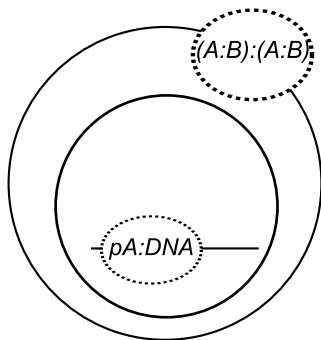
$$(1 \mid (A:B):(A:B))^L \mid \left((2)^L \mid (\overline{pA} \mid DNA) \right)$$

Term Evolution



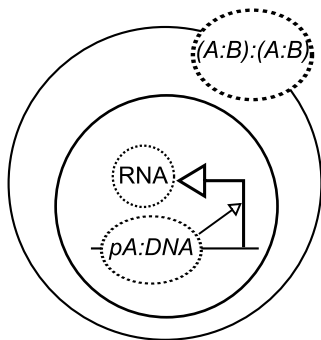
$$(1 \mid (A:B):(A:B))^L \mid \left((2)^L \mid (\overline{pA} \mid DNA) \right)$$

Term Evolution



$$(1 \mid (A:B):(A:B))^L \mid ((2)^L \mid (\overline{pA}:DNA))$$

Term Evolution



$$(1 \mid (A:B):(A:B))^L \mid \left((2)^L \mid (\overline{pA:DNA} \mid \mathbf{RNA}) \right)$$

Outline

- 1 Introduction
- 2 Molecular Interaction Maps
- 3 The Calculus of Looping Sequences
- 4 Translating MIMs to CLS
- 5 Conclusions**

Conclusions

We have given a systematic translation of MIM diagrams into Stochastic CLS+ models, thus:

- providing Molecular Interaction Maps with a formal semantics
- allowing unambiguous understanding and reasoning on MIMs
- enabling the use of existing SCLS+ simulators and formal analysis tools on MIMs

Future work

- a graphical interface for specifying MIMs, producing the equivalent intermediate encoding
- an automatic translator from the intermediate encoding to SCLS+
- integration with existing simulators and tools

Thank you.