A formal language for nano-devices.

In this Chapter we investigate the formal description of nano-devices. As detailed in the Chapter 1, the κ-calculus[26] constitutes a good candidate because it has a stochastic semantics, it is rule-based and it permits us to represent explicitly sites and internal states. Moreover it has an efficient simulator and some causality and reachability analysis techniques. However the description of nano-devices does not require the full power of the κ-calculus. So we focus on the study the nanoκ calculus. It is simple and adequate for the description of nano-devices and it retains the good properties of the κ-calculus. Moreover it can also be encoded in the stochastic π-calculus, which permits us to reuse its tools and theory, as we will see in the next Chapter.

The Chapter is organized as follows. In Section 3.1, we introduce the syntax and semantics of the nanoκ calculus. In Section 3.2, we present a modeling of the rotaxane in the nanoκ calculus and present several simulations. The Chapter is closed by a conclusion in Section 3.3 and a discussion on related works in Section 3.4.

3.1 The nanoκ calculus: syntax and semantics.

Definition 3.1.1 (nanoκ solutions, nanoκ pre-solutions) *The* nanoκ *calculus uses several sets of names:* species *ranged over by A, B, C, ...,* fields *names ranged over by r, s, t, ...,* sites *ranged over by a, b, c, ..., and* bonds *names that are* totally ordered and countable *and ranged over by x, y, z, ... In order to reflect their biochemical meaning, species, fields and sites are often addressed using strings of characters.*

We also suppose given three functions $\mathfrak{s}_f(.)$, $\mathfrak{f}(.,.)$ and $\mathfrak{s}_s(.)$: $\mathfrak{s}_f(.)$ associates to each species a set of fields, $\mathfrak{f}(.,.)$ associates to each species and fields pair a finite set of integers, and $\mathfrak{s}_s(.)$ associates to each species a set of sites.

A valuation *of a species A is a function, possibly partial, which maps the fields $r \in \mathfrak{s}_f(A)$ to a value in $\mathfrak{f}(A, r)$. Valuations are ranged over by u, v, w, ... An interface of a species A is an injective map, possibly partial, from $\mathfrak{s}_s(A)$ to either bonds or a special value ε. Interfaces are ranged over by σ, ϕ, ν, ...*

The terms defined by the following grammar:

$$\mathsf{S} \quad ::= A[u](\sigma) \;\mid\; \mathsf{S},\mathsf{S} \;\mid\; \_$$

are called solutions *when all the maps are total and* pre-solutions *otherwise. The terms $A[u](\sigma)$ are called molecules. _ is the empty solution. The operator "," is assumed to be associative, i.e. $(\mathsf{S},\mathsf{T}),\mathsf{R}$ is equal to $\mathsf{S},(\mathsf{T},\mathsf{R})$ and therefore parentheses are always omitted.*

Bonds always occur at most twice in solutions. A solution or a pre-solution is proper *if every bond therein occurs exactly twice.*

Intuitively, a molecule $A[u](\sigma)$ is determined by the species A to which it belongs, and its valuation u and its interface σ. The values of the fields in u represent the internal state of the molecule, for instance its electronic charges, or some missing or additional protons. The sites in the interface σ represent

the binding capabilities of the molecule. A site a mapped to a bond name x means that a bond, called x, is established between a and a site of some other molecule. A site mapped to the special value ε is free, it is not involved in any bond.

For example, if the species A has two fields r and *phos* and three sites nh, *bipy*, and 3 the following term is a molecule: $A[r \mapsto 0; phos \mapsto 1](nh \mapsto \varepsilon; bipy \mapsto x;)$. The fields r and *phos* have values 0 and 1, respectively; the site nh is free, the site *bipy* is bond and the bond is x and this interface does not define the state of the site 3, which may be bond or not.

Notation. In order to ease the reading, we write this molecule as $A[r^0 + phos^1](nh + bipy^x)$ (the value ε is always omitted). Let $\emptyset$ be the empty map. We write $A(\sigma)$ instead of $A[\emptyset](\sigma)$, $A[u]$ instead of $A[u](\emptyset)$, and simply A instead of $A[\emptyset](\emptyset)$. We denote by $\mathrm{ran}(\sigma)$ the range of an interface σ omitting ε and by $bonds(\mathsf{S})$ the set of bonds appearing in the solution S.

Remarque 3 *We require the set of bond names to be totally ordered in order to ease the building of a finitely branching basic transition relation (see 3.1.5). The countability of the set of bond names is used only for the encoding into the* nanoπ*-calculus (see the next Chapter and in particular the definition 4.3.1).*

Example 3.1.1 *As a running example we consider a toy chemical reaction:*

$$AB \longleftrightarrow A^+ + B^-$$

In these reactions, the complex formed by the two molecules of the species A and B can be dissociated in the two ions A^+ and B^-, and vice versa. The molecules of the two species can be in two possible states: either they have a positive charge, i.e. a missing electron, like A^+, or a negative charge, i.e. an additional electron, like B^- or they are in their standard states A and B. We model these possible states using one field e with values -1, 0 and 1 that denote respectively an additional electron, no missing or additional electron and a missing electron. Moreover we model the possible complexation using a site called bond. Formally we can use $A[e^1](bond)$ for A^+, $B[e^{-1}](bond)$ for B^- and $A[e^0](bond^x), B[e^0](bond^x)$ for AB respectively.

The structural congruence of the nanoκ calculus is given by the following definition:

Definition 3.1.2 (Structural congruence of nanoκ) *The structural equivalence between solutions, denoted $\equiv$, is the least congruence satisfying the following three rules (we recall that solutions are already quotiented by associativity of ","):*

1. $\mathsf{S}, \mathsf{T} \equiv \mathsf{T}, \mathsf{S}$;
2. $\_, \mathsf{S} \equiv \mathsf{S}$;

3. $\mathsf{S} \equiv \mathsf{T}$ *if there exists an injective renaming* $\imath$ *on bonds such that* $\mathsf{S} = \imath(\mathsf{T})$.

Example 3.1.2 *Commutativity and injective renaming of the structural equivalence make it possible to prove:*

$$A[e^0](bond^x), B[e^0](bond^x) \;\equiv\; B[e^0](bond^z), A[e^0](bond^z)$$

The dynamics of the nanoκ calculus is governed by means of reaction rules. These rules correspond closely to the biochemical reactions we wish to model. Intuitively a nanoκ term can perform a transition when it contains an instance of the left hand side of a rule. Before formally presenting the rules of the nanoκ calculus a few preliminary definitions are in order:

- we write $\sigma \leq \sigma'$ if $\mathrm{dom}(\sigma) = \mathrm{dom}(\sigma')$ and, for every i, if $\sigma(i) \neq \varepsilon$ then $\sigma(i) = \sigma'(i)$ (intuitively all the bonds present in σ appear also in σ');
- when we write $u + u'$ and $\sigma + \sigma'$ we assume that $\mathrm{dom}(u) \cap \mathrm{dom}(u') = \emptyset$ and $\mathrm{dom}(\sigma) \cap \mathrm{dom}(\sigma') = \emptyset$.

We can now define the different kinds of rules for the nanoκ calculus:

Definition 3.1.3 *Reactions of* nanoκ *calculus are either* creations, destructions, *or* exchanges *and they are labelled by a* rate, *which is a positive real number or* ∞. *Creations have the format:*

$$A[u](\sigma), B[v](\tau) \overset{\lambda}{\rightarrow} \quad A[u'](\sigma'), B[v'](\tau'), C_1[w_1](\eta_1), \cdots, C_n[w_n](\eta_n)$$

where both hand sides are proper pre-solutions and where $\sigma \leq \sigma'$, $\tau \leq \tau'$, $\mathrm{dom}(u) = \mathrm{dom}(u')$, $\mathrm{dom}(v) = \mathrm{dom}(v')$, *and* w_i *and* η_i *are total. Destructions have one of the formats:*

$$A[u](\sigma), B[v](\tau) \overset{\lambda}{\rightarrow} A[u'](\sigma'), B[v'](\tau')$$
$$A[u](\sigma), B[v](\tau) \overset{\lambda}{\rightarrow} A[u'](\sigma')$$

where both hand sides are proper pre-solutions and where $\sigma \geq \sigma'$, $\mathrm{dom}(u) = \mathrm{dom}(u')$, *and, in the first case,* $\tau \geq \tau'$, $\mathrm{dom}(v) = \mathrm{dom}(v')$ *and, in the second case,* τ *has to be total. Exchanges have one of the formats:*

$$A[u](\sigma), B[v](\tau) \overset{\lambda}{\rightarrow} A[u'](\sigma), B[v'](\tau)$$
$$A[u](a^x + \sigma), B[v](b + \tau) \overset{\lambda}{\rightarrow} A[u'](a + \sigma), B[v'](b^x + \tau)$$

where the pre-solutions $A[u](\sigma), B[v](\tau)$ *and* $A[u](a+\sigma), B[v](b+\tau)$ *are proper and* $\mathrm{dom}(u) = \mathrm{dom}(u')$ *and* $\mathrm{dom}(v) = \mathrm{dom}(v')$.

In the rest of the thesis we assume that reactants share at most one bond, i.e. $\mathrm{ran}(\sigma) \cap \mathrm{ran}(\rho)$ *is either an empty set or a singleton.*

Creations produce new bonds between two unbound sites and/or synthesize new molecules. Destructions behave in the other way around. Exchanges either leave the interfaces unchanged or move one bond from a reactant to the other, which we call bond-flipping exchange.[1]

It is worthwhile to remark that reactions do not address every field and site of the reactants (both hand sides of a rule are pre-solutions). The intended meaning is that two molecules react if they are *instances* of the left-hand side of a reaction. We will formalize this notion later on in the basic transition relation (see definition 3.1.5).

Example 3.1.3 *The* nanoκ *calculus reactions that corresponds to the two reactions of our toy example are:*

$$A[e^0](bond^x), B[e^0](bond^x) \overset{100}{\rightarrow} A[e^1](bond), B[e^{-1}](bond)$$
$$A[e^1](bond), B[e^{-1}](bond) \overset{10}{\rightarrow} A[e^0](bond^x), B[e^0](bond^x)$$

where we have considered a rate 100 *for the left to right direction and* 10 *for the right to left direction.*

Stochastic semantics of the nanoκ calculus. We can now present the stochastic semantics of nanoκ. As we anticipated in the Chapter 2, it is achieved in several steps: first we build the basic transition relation, then the collective transition relation is derived from the basic one and finally the resulting IMC is downgraded into a CTMC, assuming that our system meets the strictly Markovian property.

The semantics depends strongly on the sets of species and of reactions considered. We formalize this with the notion of nanoκ *system*:

Definition 3.1.4 (nanoκ systems) *A* nanoκ *system is a tuple determined by* $\mathcal{S}$ *a set of species names,* $\mathcal{N}$ *a set of fields and sites names,* $\mathcal{B}$ *a totally ordered and countable set of bond names,* $\mathfrak{s}_f(.)$ *a map yielding the fields of a species,* $\mathfrak{f}(.,.)$ *a map yielding the set of possible values of a field of a species,* $\mathfrak{s}_s(.)$ *a map yielding the sites of a species and* $\mathcal{R}$ *a set of reactions.*

Notation. *We refer to a* nanoκ *system as* $(\mathcal{S},\mathcal{R})$ *and keep the other elements implicit in* $(\mathcal{S},\mathcal{R})$*.*

We now present the basic transition relation of the nanoκ calculus. In this case it is not necessary to follow the general methods presented in the Chapter 2, there exists a more efficient ad hoc method. Indeed since a nanoκ solution can be seen as a sequence of molecules, the redex of a rule is uniquely identified by the position of the two reactants inside this sequence. Therefore we only use pairs of integers as identifiers. Note however that this process is much eased

[1] The terms creation and destruction have been preferred to *complexation* and *decomplexation* used in [26, 52] because they have a more neutral chemical meaning.

by postponing the structural congruence to the step of the collective transition relation.

The definition of the basic transition relation of the nanoκ calculus requires some notation. Let μ range over $\rho_L, \imath$ and $\rho_R, \imath$ and let $\overline{\rho_L, \imath} = \rho_R, \imath$ and $\overline{\rho_R, \imath} = \rho_L, \imath$ where $\imath$ is an injective renaming (notice that $\overline{\overline{\mu}} = \mu$). The nano$\kappa$ reactions may be addressed by:

$$A[u](\sigma), B[v](\rho) \overset{\lambda}{\twoheadrightarrow} A[u'](\sigma'), \mathsf{S}$$

where S may also be $\_$. With an abuse of notation we lift a renaming $\imath$ to a solution by applying it pointwise. Finally we denote the set of names present in a solution $\mathtt{S}$ with $\mathrm{name}(S)$.

Definition 3.1.5 *Given a* nanoκ *system whose set of reactions is $\mathcal{R}$, its basic transition relation, written either $\xrightarrow{\rho,\imath}_{\ell,\ell'}$ or $\xrightarrow{\mu,\imath}_{\ell}$, is the least relation that satisfies the following rules:*

- *(init) If $\rho = A[u](\sigma), B[v](\phi) \overset{\lambda}{\twoheadrightarrow} A[u'](\sigma'), \mathsf{S} \in \mathcal{R}$, then for all ν we have both:*
  - $A[u + w](\imath \circ \sigma + \nu) \xrightarrow{\rho_L,\imath}_1 A[u' + w](\imath \circ \sigma' + \nu)$ *and*
  - $B[v + w](\imath \circ \phi + \nu) \xrightarrow{\rho_R,\imath}_1 \mathsf{T}$

  where T is either $B[v' + w](\imath \circ \phi' + \nu), \imath(\mathsf{S}')$ if $\mathsf{S} = B[v'](\phi'), \mathsf{S}'$ or $\imath(\mathsf{S})$ otherwise and where $\imath$ is an order-preserving injective renaming with $\mathrm{ran}(\imath) \cap \mathrm{ran}(\nu) = \emptyset$;

- *(lift) if $\mathsf{S} \xrightarrow{\mu,\imath}_{\ell} \mathsf{S}'$ then both:*
  - $\mathsf{S}, \mathsf{T} \xrightarrow{\mu,\imath}_{\ell} \mathsf{S}', \mathsf{T}$ *and*
  - $\mathsf{T}, \mathsf{S} \xrightarrow{\mu,\imath}_{\ell' + \ell} \mathsf{T}, \mathsf{S}'$

  where T has ℓ' molecules and where $(\mathrm{name}(\mathsf{S}') \setminus \mathrm{name}(\mathsf{S})) \cap \mathrm{name}(\mathsf{T}) = \emptyset$ if the rule of μ is a creation;

- *(communication) if $\mathsf{S} \xrightarrow{\mu,\imath}_{\ell} \mathsf{S}'$ and $\mathsf{T} \xrightarrow{\overline{\mu},\imath}_{\ell'} \mathsf{T}'$, let ρ be the rule of μ and let $\jmath$ be an order-preserving injective renaming which maps $\mathrm{name}(\mathsf{S}', \mathsf{T}') \setminus \mathrm{name}(\mathsf{S}, \mathsf{T})$ (i.e. the created names) into the least bonds not belonging to $\mathrm{name}(\mathsf{S}, \mathsf{T})$ then:*
  - $\mathsf{S}, \mathsf{T} \xrightarrow{\rho,\jmath}_{\ell,\ell'' + \ell'} \jmath(\mathsf{S}', \mathsf{T}')$

  where S has ℓ'' molecules.

The tags of the basic transition relation of the nanoκ calculus are integers or pairs of integers. According to the general approach presented in the preliminaries one would have to label each molecule with a unique identifier which would be used as subscript when the molecule is part of a redex. However in our case it is sufficient to record the position of the molecule inside the sequence of molecules. For instance supposing that the reaction:

$$A[e^1](bond), B[e^{-1}](bond) \twoheadrightarrow A[e^0](bond^y), B[e^0](bond^y)$$

is labelled ρ, then the solution $A[e^1](bond), A[e^1](bond), A[e^1](bond), B[e^{-1}](bond)$ has three outgoing basic transitions:

$$\begin{array}{l} A[e^1](bond), A[e^1](bond), A[e^1](bond), B[e^{-1}](bond) \xrightarrow{\rho}_{1,4} \\ \quad A[e^0](bond^y), A[e^1](bond), A[e^1](bond), B[e^0](bond^y) \\ A[e^1](bond), A[e^1](bond), A[e^1](bond), B[e^{-1}](bond) \xrightarrow{\rho}_{2,4} \\ \quad A[e^1](bond), A[e^0](bond^y), A[e^1](bond), B[e^0](bond^y) \\ A[e^1](bond), A[e^1](bond), A[e^1](bond), B[e^{-1}](bond) \xrightarrow{\rho}_{3,4} \\ \quad A[e^1](bond), A[e^1](bond), A[e^0](bond^y), B[e^0](bond^y) \end{array}$$

The basic transition relation uses also finite injective renamings. We first present the role of the $\imath$ renaming in the (init) rule and then the role of the $\jmath$ renaming in the (communication) rule.

The role of the renaming of the (init) rule is to allow the instantiation of the bond names of a rule in a given solution. To clarify this point, consider the creation $\varrho' = \ C(1^x + 2), C(1^x + 2) \overset{10}{\twoheadrightarrow} C(1^x + 2^y), C(1^x + 2^y)$ (a bond is created between two C molecules provided they are already bond). Then take the solution $C(1^z+2), C(1^v+2), C(1^z+2), \ C(1^v+2)$. We derive the expected transition

$$\begin{array}{l} C(1^z + 2), C(1^v + 2), C(1^z + 2), C(1^v + 2) \\ \xrightarrow{\varrho'}_{1,3} \ C(1^z + 2^w), C(1^v + 2), C(1^z + 2^w), \ C(1^v + 2) \end{array}$$

following a structured operational semantics approach [62]. Namely, we focus on the single reactants and lift the transitions to "," -contexts. This is correct to the extent that one records the instantiation of bonds in the left-hand sides of reactions with the actual names of the molecules: the two reactants must instantiate bonds in the same way. This is the reason why the first two molecules of the above solution cannot react with ϱ. More precisely, $C(1^z + 2) \xrightarrow{\varrho'_L, \imath}_1 C(1^z + 2^w)$, where $\imath = [x \mapsto z, y \mapsto w]$, and $C(1^v + 2) \not\xrightarrow{\varrho'_R, \imath}_1$.

The role of the renaming in the (communication) rule is to ensure that for a given a reaction and a pair of molecules of a given solution, one can derive at most one basic transition corresponding to these molecules and this reaction. If we do not require that the renaming is injective and order-preserving we would be able to derive a transition $C(1^x+2), C(1^x+2) \xrightarrow{\varrho'}_{1,2} \ C(1^x+2^y), C(1^x+2^y)$ for any free name y and so one occurrence of one redex would yield infinitely many basic transitions.

Thus we need to choose one transition among these possibilities. By asking that the created are the least ones we prevent the infinite number of possible transitions. However since several bonds can be created by a reaction this only ensures that the number of possible transitions is finite but not equal to 1. So we also ask that the renaming is order preserving. This permit us to choose one transition: the one where the name of the least created bond is mapped to the

least name not present in the solution, the name of second least created bond to the second least name not present in the solution, ...

It is also worthwhile to notice that there is no rule lifting a transition $\xrightarrow{\mu}_{\ell,\ell'}$ to a context ",": we use the associativity of , to partition a solution S into S',S'' such that the reactants are in S' and S''.

Remarque 4 *One might wish to derive transition constituted of the firing of several reactions. This approach seems relevant since all the reactions happen in parallel. However the Gillespie algorithm [37], which is the standard simulation method for stochastic process algebra, is not compatible with this approach. Indeed the Gillespie algorithm simulates systems of biochemical reactions by probabilistically selecting the next reaction to happen and the time spent before it happens.*

The compatibility of the structural congruence with respect to the basic transition relation is stated in the following proposition:

Proposition 3.1.1 *Let* $\mathsf{S} \equiv \mathsf{S}'$.

1. *If* $\mathsf{S} \xrightarrow{\mu}_{\ell} \mathsf{T}$ *then there exists a* T' *and a renaming* $\imath$ *such that* $\mathsf{S}' \xrightarrow{\imath(\mu)}_{\ell'} \mathsf{T}'$ *and* $\mathsf{T} \equiv \mathsf{T}'$ *(with* $\imath(\mu)$ *we denote the extension of the renaming* $\imath$ *to the label* μ*);*

2. *if* $\mathsf{S} \xrightarrow{\rho}_{\ell,\ell'} \mathsf{T}$ *then there exists* T' *such that* $\mathsf{S}' \xrightarrow{\rho}_{\ell'',\ell'''} \mathsf{T}'$ *and* $\mathsf{T} \equiv \mathsf{T}'$.

Proof

1. The proof is a straightforward induction on the derivation tree of $\mathsf{S} \xrightarrow{\mu}_{l} \mathsf{T}$.

2. The result is a direct consequence of the first item and of the (communication) rule.

□

Now that the basic transition relation is defined, we can derive the collective transition relation according to the definition 2.1.2. It is illustrated in the following example.

Example 3.1.4 *As we have seen above the solution* $A[e^1](bond), A[e^1](bond), A[e^1](bond), B[e^{-1}](bond)$ *has three outgoing transitions labelled* $\xrightarrow{\rho}_{1,4}$, $\xrightarrow{\rho}_{2,4}$ *and* $\xrightarrow{\rho}_{3,4}$ *to structurally congruent states. Therefore we obtain an unique collective transition:*

$$A[e^1](bond), A[e^1](bond), A[e^1](bond), B[e^{-1}](bond) \overset{300}{\longmapsto} A[e^0](bond^x), A[e^1](bond), A[e^1](bond), B[e^0](bond^x)$$

Finally the downgrading of a nanoκ collective transition relation can be performed according to the definition 2.2.3.

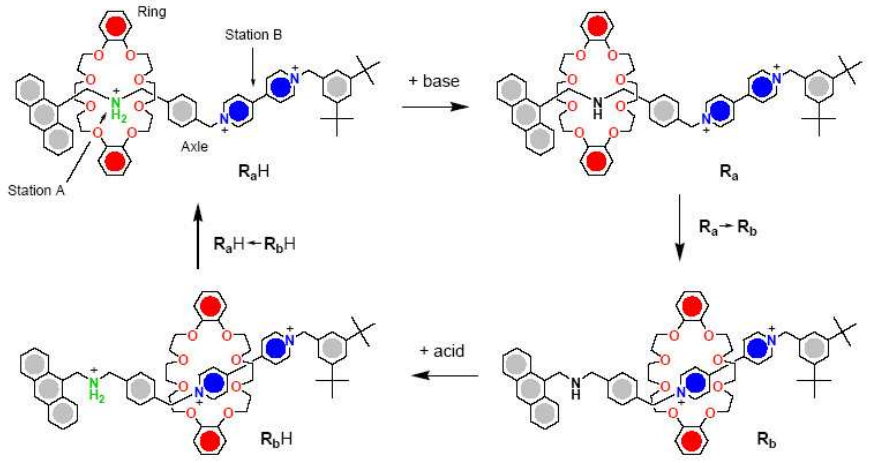

Figure 3.1: Schematic representation of the shuttling processes of the molecular ring in the examined rotaxane.

3.2 The nanoκ calculus at work: the rotaxane case study

The investigated rotaxane RaH (Figure 3.1) [54, 1] is made of a stoppered axle containing an ammonium (A) and an electron acceptor bipyridinium (B) stations that can establish hydrogen-bonding and charge-transfer interactions, respectively, with the ring component, which is a crown ether with electron donor properties. An anthracene moiety is used as a stopper because its absorption, luminescence, and redox properties are useful to monitor the state of the system. Since the hydrogen bonding interactions between the macrocyclic ring and the ammonium center are much stronger than the charge-transfer interactions of the ring with the bipyridinium unit, the rotaxane exists as only one of the two possible translational isomers, denoted as RaH in Figure 3.1. In solution, addition of a base (e.g., tributylamine) converts the ammonium center into an amine function, giving the transient state Ra that is transformed into the stable state Rb as a consequence of the displacement of the macrocycle onto the B station. The process can be reversed by addition of acid (e.g., trifluoroacetic acid) and the initial state is restored, passing through the transient state denoted as RbH. Nuclear magnetic resonance, absorption and luminescence spectroscopic experiments, together with electrochemical measurements, indicate that the acid-base controlled switching, which is fully reversible and relatively fast, exhibits a clear-cut on-off behaviour [1].

The Rotaxane RaH is particularly appropriate to test the modeling approach of the nanoκ calculus because it is one of the very few cases wherein not only the thermodynamic properties, but also the dynamic behaviour of the system

have been experimentally characterized in detail. Specifically, the macrocycle's shuttling process between the ammonium/amine and bipyridinium stations in this rotaxane, driven by the successive addition of base and acid, have been investigated in solution [36]. The rate constants for the "forward" (Ra→Rb) and "backward" (RbH→RaH) shuttling motions (vertical processes in Figure 3.1) of the molecular ring, which occur, respectively, upon deprotonation and reprotonation (that is upon loss or gain of a proton respectively) of the ammonium/amine recognition site on the axle (horizontal processes in Figure 3.1), were found to be $0.72s^{-1}$ and $40s^{-1}$ at $293°K$, respectively.

3.2.1 Modeling the rotaxane RaH in the nanoκ calculus.

The nanoκ calculus molecules. Figure 3.2 illustrates the nanoκ calculus modeling of the rotaxane RaH. We use four species:

- *Nh* models the ammonium/amine station of the rotaxane: it has one field h and two sites *ring* and *axle*;
- *Axle* models the spacer between the two stations: it has two fields h and s and three sites *nh*, *bipy*, and *ring*;
- *Bipy* models the bipyridinium station: it has one field h and two sites *ring* and *axle*;
- *Ring* models the crown ether ring: it has no field and one site *link*;
- *AcidBase* models the acid-base couple used to trigger the motion of the rotaxane: it has one field h and no site.

The pairs of sites *axle* of *Nh* and *nh* of *Axle*, and *axle* of *Bipy* and *bipy* of *Axle* are always linked in our modeling. They model the covalent bonds maintaining the structural integrity of the axle. Exactly one site *ring* of *Nh*, *Bipy*, and *Axle* is linked at a given moment at *link* of *Ring*. The first two cases respectively model the "stable" RaH and Rb states of Figure 3.1 in which the ring is steadily located around the *Nh* or the *Bipy* molecules, respectively. The last case models the "unstable" states; these are the Ra and RbH states of Figure 3.1 in which the ring is not steadily located.

Ammonium and amine functions have different chemical nature but can be seen as protonated and deprotonated versions of the same species. Thus we model both by the same nanoκ calculus species *Nh*. Its field h is used to record the presence or absence of a proton on *Nh*: its value is 1 if it is protonated, and 0 otherwise. We also need to distinguish between the two transient states where the *Ring* is on the *Axle*: does it come from the *Nh* station of from the *Bipy* one? In order to store this information we use the field s: its value is 1 if the *Ring* comes from the *Nh* station and 0 otherwise.

As *Ring*'s movements are triggered by protonations and deprotonations due to acid-base reactions, we also need to have acid and base molecules in our modeling. We choose to model an acid-base couple with only one species since

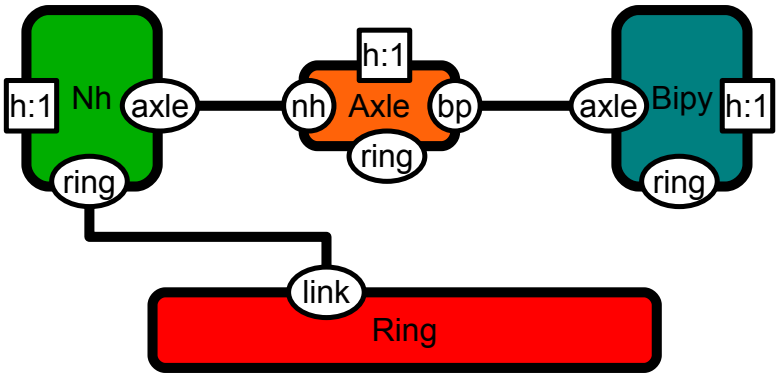

Figure 3.2: Initial state of the Rotaxane RaH in nanoκ calculus.

an acid and a base of the same couple only differ by a proton. We consider the species *AcidBase* with no site and one field h having value 1 in case the acid/base molecule holds the proton to be exchanged, 0 otherwise (for instance $AcidBase[h^1]$ and $AcidBase[h^0]$ are respectively an acid molecule ready to give a proton and a base molecule ready to receive a proton). If a different acid-base were to be considered it would be modeled similarly by a species $AcidBase_2$ with one field h and no site.

The initial state for rotaxane RaH is thus modeled by the term:

$$Nh[h^1](axle^s + ring^x) \ , \quad Axle[h^1 + s^1](nh^s + bipy^r + ring) \ ,$$
$$Bipy[h^1](axle^r + ring) \ , \quad Ring(link^x)$$

graphically depicted in Figure 3.2.

Note that the *Nh* is initially protonated (and this information is present also in the *Axle* and the *Bipy*), the *Axle* is bond to the *Nh* and the *Bipy*, and the *Ring* is bond to the *Nh*.

The nanoκ calculus reactions. We now present the reactions used in our modeling. Reactions 1, 2, 9 and 10 are presented with a double arrow (these are *reversible* reactions). Formally they correspond to two nanoκ calculus reactions, one achieved reading the reaction from left to right considering the rate over the arrow, and the ones achieved reading it from right to left considering the rate below. In this section we do not consider numerical values of rates, this is detailed in part 3.2.2.

A base can get the proton of a protonated *Nh*, and a *Nh* can get a proton from an acid. These acid-base reactions are reversible. Reactions 1 and 2 model

this phenomenon. The systems corresponding to the left-hand side and right-hand side coexist, even if one can be more predominant according to the ratio $nh\_base/base\_nh$.

$$Nh[h^1], AcidBase[h^0] \underset{base\_nh}{\overset{nh\_base}{\rightleftarrows}} Nh[h^0], AcidBase[h^1] \quad (1)$$

The protonation state of the molecule *Nh* needs to be known by *Bipy* because it affects its interaction with *Ring*. Reactions 3 and 4 achieve this by passing information from *Nh* to *Bipy* through *Axle*. These updates are instantaneous because they represent an immediate consequence of the protonation or deprotonation of the Nh station.

if $(\alpha \neq \beta)$
$Nh[h^\alpha](axle^s), Axle[h^\beta](nh^s) \overset{\infty}{\rightarrow} Nh[h^\alpha](axle^s), Axle[h^\alpha](nh^s) \qquad (3_{\alpha,\beta})$
and:
$Axle[h^\alpha](bipy^r), Bipy[h^\beta](axle^r) \overset{\infty}{\rightarrow} Axle[h^\alpha](bipy^r), Bipy[h^\alpha](axle^r) \qquad (4_{\alpha,\beta})$

The above rule correspond actually to many rules, one for each possible value of α and β. We gather them in two rules for the sake of the clarity. We achieve the modeling of *Ring* movements in two steps. Firstly the instantaneous reactions to deprotonation/reprotonation (reactions 5–8), and secondly the actual *Ring* shuttling (reactions 9 and 10). The reactions (5) and (6) are used to enter in "unstable" states when the *Nh* is deprotonated while the *Ring* is around the *Nh* (reaction (5)), or protonated while the *Ring* is around the *Bipy* (reaction (6)). On the other hand, the reactions (7) and (8) are used to re-enter in a "stable" state in the case the *Nh* returns to its previous (de)protonated state before the *Ring* actually binds to its new station. All these events are immediate consequences of deprotonation or reprotonation of *Nh*; for this reason, they have infinite rates. When a field contains a $*$, it means that there is a rule for each possible value of the field.

$$Nh[h^0](axle^s + ring^x), Axle[s^*](nh^s + ring) \overset{\infty}{\rightarrow} Nh[h^0](axle^s + ring), Axle[s^1](nh^s + ring^x) \quad (5)$$
$$Bipy[h^1](axle^r + ring^x), Axle[s^*](biax^r + ring) \overset{\infty}{\rightarrow} Bipy[h^1](axle^r + ring), Axle[s^0](biax^r + ring^x) \quad (6)$$
$$Axle[s^1](nh^s + ring^x), Nh[h^1](axle^s + ring) \overset{\infty}{\rightarrow} Axle[s^1](nh^s + ring), Nh[h^1](axle^s + ring^x) \quad (7)$$
$$Axle[s^0](nh^s + ring^x), Bipy[h^0](axle^s + ring) \overset{\infty}{\rightarrow} Axle[s^0](nh^s + ring), Bipy[h^0](axle^s + ring^x) \quad (8)$$

We now complete our modeling with reactions 9 and 10 representing the completion of the *Ring* movement. These reactions are reversible because the *Ring*

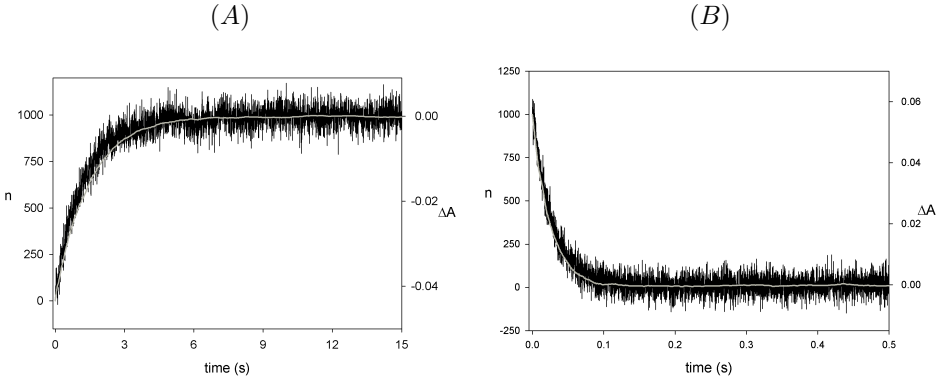

Figure 3.3: Comparing the simulations *in silico* with the experiments *in vitro*. Grey traces: number of *Ring*s located around *Bipy*s during the "forward" Ra→Rb (part A) and the "backward" RbH→RaH (part B). Black traces: UV absorbance changes observed upon the occurrence of the same respective shuttling processes.

is susceptible to leave its "stable" station due to the Brownian motion.

$$Axle[s^*](bipy^r + ring^x)\,,Bipy[h^0](axle^r + ring) \underset{unlink\_bipy}{\overset{link\_bipy}{\rightleftarrows}} Axle[s^0](bipy^r + ring)\,,Bipy[h^0](axle^r + ring^x) \tag{9}$$

$$Axle[s^*](nh^s + ring^x)\,,Nh[h^1](axle^s + ring) \underset{unlink\_nh}{\overset{link\_nh}{\rightleftarrows}} Axle[s^1](nh^s + ring), Nh[h^1](axle^s + ring^x) \tag{10}$$

3.2.2 Simulation results.

The above modeling of rotaxane RaH in `nano`κ calculus yields an IMC system that is strictly markovian and so it can be downgraded to an equivalent CTMC. Therefore we obtain a CTMC system that we use to simulate *in silico* the behaviour of the rotaxane RaH. The simulations are performed using the SPiM tool [19] using the encoding from the `nano`κ calculus to the stochastic π-calculus of the Chapter 4. We did not use the κ-factory because at the time we performed the simulations it were not able to handle infinite rates.

As previously discussed the rates for the ring movements are respectively $link\_bipy = 0.72s^{-1}$ and $link\_nh = 40s^{-1}$. On the basis of the estimated equilibrium constants, the rates for the reverse reactions are quantified two orders of magnitude smaller, i.e. $unlink\_bipy = 0.0072s^{-1}$ and $unlink\_nh = 0.4s^{-1}$.

The aim of the first two simulations depicted in Figure 3.3 is to check whether the experimentation *in silico* can reproduce the results observed in *in vitro* [36]. The techniques used for the *in vitro* experimentation did not make it possible to

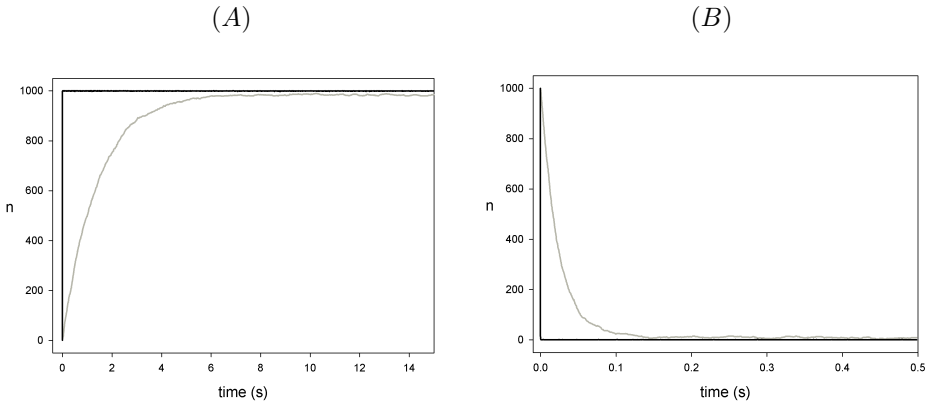

Figure 3.4: Number of *Ring*s located around *Bipy*s (grey trace) and number of deprotonated rotaxanes (black trace) during the "forward" shuttling in the presence of base molecules (part A) and the "backward" shuttling in the presence of acid molecules (part B) at concentration $10^{-4}M$.

observe and quantify the deprotonation/reprotonation rates (this is not surprising as these are very fast acid-base reactions). Thus, in the simulation we have considered instantaneous deprotonation/reprotonation, i.e. either $nh\_base = \infty$ and $base\_nh = 0$ for protonation, or $nh\_base = 0$ and $base\_nh = \infty$ for deprotonation. In both simulations, we have considered 1000 rotaxanes: in the first one we have simulated deprotonation and "forward" (Ra→Rb) shuttling, in the second one reprotonation and "backward" (RbH→RaH) shuttling. In the first simulation the shuttling phase is completed in around 6 seconds, while in the second one in 0.1 seconds; this is a consequence of the different rates of the two directions of shuttling. Very remarkably, simulated data are in striking agreement with the experimental results.

After these initial encouraging results, we have decided to use the *in silico* simulation techniques to provide a comprehensive view of the overall reactions depicted in Figure 3.1, simulating also the deprotonation/reprotonation phases not observed in the *in silico* experimentation. More precisely, the aim of this second group of simulations was to either validate or invalidate the assumption according to which deprotonation/reprotonation can be considered "instantaneous" with respect to the shuttling time. To this aim, we have simulated deprotonation/reprotonation under two different concentrations of rotaxanes. In fact, this is a bimolecular reaction whose rate is influenced by the concentration of the reactants. For instance, at a concentration close to those considered in [36], e.g. $10^{-4}M$, assuming 1000 instances of rotaxane and base/acid, a plausible rate for deprotonation/reprotonation is $2 \times 10^{3}s^{-1}$ (with reverse reaction rate of the order of $2 \times 10^{-4}s^{-1}$) while at the concentration $10^{-8}M$ it is $0.2s^{-1}$ (with reverse reaction on the order of $0.2 \times 10^{-7}s^{-1}$).

We have performed the two simulations, namely deprotonation with subse-

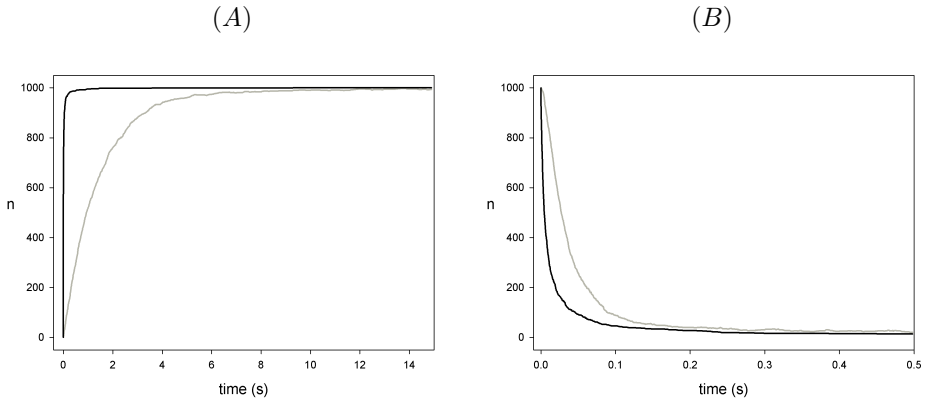

Figure 3.5: Number of *Ring*s located around *Bipy*s (grey trace) and number of deprotonated rotaxanes (black trace) during the "forward" shuttling in the presence of base molecules (part A) and the "backward" shuttling in the presence of acid molecules (part B) at concentration $10^{-8}M$.

quent "forward" shuttling and reprotonation with subsequent "backward" shuttling, considering the two different concentrations.

The results at concentration $10^{-4}M$ are reported in Figure 3.4; they essentially confirm the validity of the "instantaneous" deprotonation/reprotonation assumption at this concentration level. We report in Figure 3.5 the results for concentration $10^{-8}M$; in this case the rings start moving before the deprotonation/reprotonation phase is over. This proves that in the rotaxane RaH the stimulus and the subsequent shuttling could interplay.

In the light of this observation, we have decided to investigate some additional scenarios not yet considered in the *in vitro* experimentations. In particular, we have decided to analyze the interplay between shuttling and a stimulus given by *weaker* acid/base molecules, that is, for which the ratio between the deprotonation/reprotonation rate and the reverse rate is smaller. In fact, the ratio considered in the previously discussed simulations is on the order of 10^7; a smaller reasonable ratio could be on the order of 10^3. Considering this new ratio, assuming 1000 instances of rotaxane and base/acid, at the concentration $10^{-4}M$ the new rates for deprotonation/reprotonation is $2\times10^3 s^{-1}$ with reverse reaction rate on the order of $2s^{-1}$, while at the concentration $10^{-8}M$ it is $0.2s^{-1}$ with reverse reaction on the order of $0.2 \times 10^{-3}s^{-1}$. Using these new rates, we have simulated the "forward" and "backward" shuttling at both concentrations, $10^{-4}M$ in Figure 3.6 and $10^{-8}M$ in Figure 3.7.

Interestingly, we found out that the "forward" shuttling is no longer guaranteed. In fact, only in some of the deprotonated rotaxanes the *Ring* actually moves around the *Bipy*. In other terms, the efficiency of the rotaxane is no longer close to 100% (as was the case in the *in vitro* experimentations and in the other *in silico* simulations) but it is around 35% for concentration $10^{-4}M$,

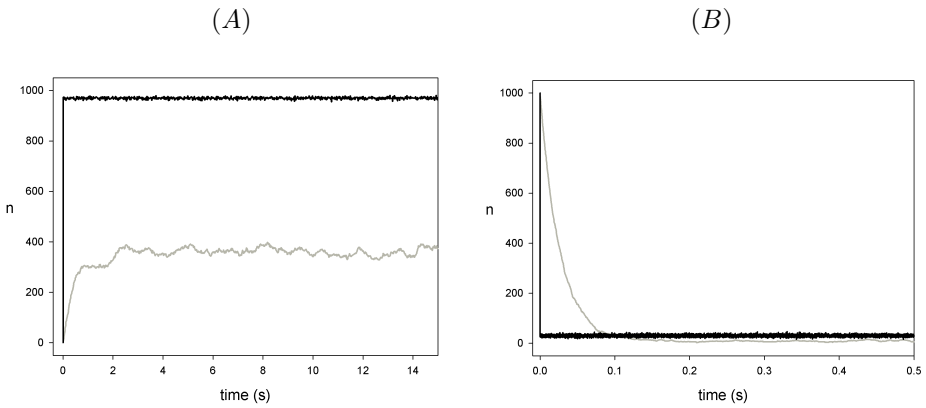

Figure 3.6: Number of *Ring*s located around *Bipy*s (grey trace) and number of deprotonated rotaxanes (black trace) during the "forward" shuttling in the presence of weak base molecules (part A) and the "backward" shuttling in the presence of weak acid molecules (part B) at concentration $10^{-4}M$.

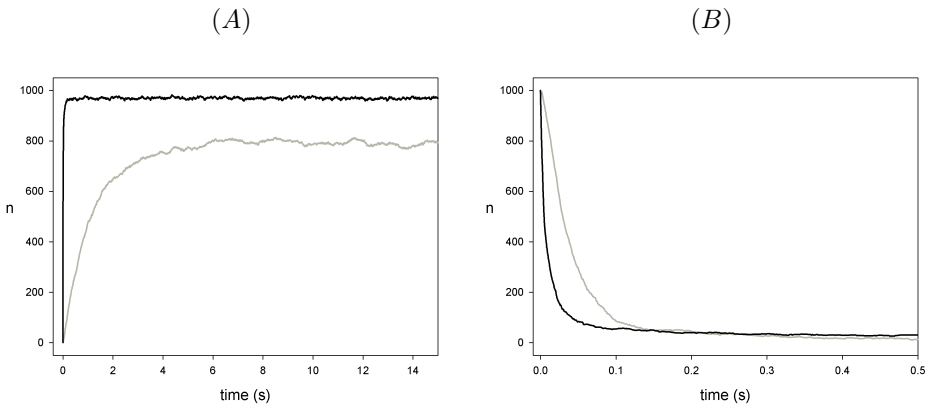

Figure 3.7: Number of *Ring*s located around *Bipy*s (grey trace) and number of deprotonated rotaxanes (black trace) during the "forward" shuttling in the presence of weak base molecules (part A) and the "backward" shuttling in the presence of weak acid molecules (part B) at concentration $10^{-8}M$.

or 75% for concentration $10^{-8}M$. After an analysis of this initially unexpected results, we can conclude that the inefficiency of the rotaxane is justified by the fact that the reverse reaction of deprotonation (i.e. re-protonation) can activate a chain of reactions that allows an already deprotonated rotaxane, with the *Ring* around the *Bipy*, to return in the initial state (protonated with the *Ring* around the *Nh*). This chain of reactions, under these particular circumstances, plays an important role in the equilibrium between the number of deprotonated rotaxanes with the *Ring* around the *Nh* and the number of deprotonated rotaxanes with the *Ring* around the *Bipy*.

3.3 Conclusion.

We have introduced `nano`κ, a calculus designed on purpose for the modeling of nano-devices. The calculus is equipped with a stochastic semantics (defined in terms of a CTMC) that can be used to simulate the evolution of the behaviour of nano-devices using stochastic simulation techniques such as, for instance, the Gillespie algorithm [37]. We have applied the `nano`κ calculus to the modeling and simulation of the RaH rotaxane [54, 1], a nano-device that attracted a lot of attention inside the nano science and technology community, because it proved very useful for building more complex nano-devices [48, 46, 2]. We have used the `nano`κ calculus model of the RaH rotaxane to simulate its behaviour under conditions that were not yet considered in the *in vitro* experimentations. We found out that under particular circumstances the nano-device is not as efficient as expected. In particular, even if almost all the rotaxanes in a solution are stimulated, only some of them change their internal structure according to the stimulus.

As future work, we intend to use the `nano`κ calculus to model and simulate also more complex nano-devices, such as the nano-elevator [2]. As we detailed in Chapter 1, nano elevators are composed of a platform and of three rotaxanes that, once appropriately stimulated, move the platform up or down. We expect to reuse the modeling of the rotaxane presented in this paper. In fact, one of the most important peculiarities of the `nano`κ calculus is that it supports compositional modeling: the reactions describing the behaviour of the molecules that are part of a nano-device, are still valid reactions also when the nano-device is itself considered as a part of a more complex system.

We have already discussed in the Introduction the origins of the `nano`κ calculus, and its strong relationship with the κ-calculus[26]. Here we simply recall that the `nano`κ calculus can be seen as a member of the κ-family. The κ-calculus benefits from efficient techniques of simulation and analysis [25, 24, 47, 27]. In contrast to our `nano`κ calculus this formalism allows reactions involving an arbitrary number of molecules, but there are no exchanges rules, edges can only be created or destroyed, not moved. These differences are explained by our field of application. Dealing with the behaviour of nano-complexes, the relevant reaction we met involve barely more than two molecules, but edges are often exchanged and moved between molecules.

3.4 Related works.

The nanoκ calculus has been influenced also by Cardelli's language of stochastic interacting processes [14, 17] that has been put in correspondence with Ordinary Differential Equations. The stochastic semantics of the nanoκ calculus, indeed, has been given following these lines.

Another process calculus for the modeling of biochemical systems is Bio-PEPA [22, 7]. Differently from the Cardelli's approach, there is no one-to-one correspondence between processes and molecules, but one process is used to represent the concentration of one species. In Bio-PEPA the rates are associated to the actions by means of "functional rates": these are functions that are evaluated at the moment of the reduction of the systems. The idea of functional rates is particularly useful when different kinetic laws are considered in the same unifying framework. The possibility of considering different kinetic laws is also proposed in BIOCHAM [13], a programming environment for modeling biochemical systems, making simulations and querying the model in temporal logic. Our approach is different from both Bio-PEPA and BIOCHAM because we follow the Cardelli's one-to-one correspondence between molecules and processes. In fact, we have found this approach appropriate for a compositional model of discrete state systems (in which we count the number of molecules instead of considering their concentrations).

The beta-binders [63] which evolved recently into the BlenX language [30] are another formalism that can represent complexing molecules. It is based on a π-calculus where the usual communication discipline is relaxed to better represent the complementarity of molecular binding sites. It is achieved by means of a wrapping operator associating an interface to a group of π-processes.

Finally, in the *calculus of looping sequences* [6, 4] a different paradigm is taken. Molecules are represented simply by a name rather than by a π-process and they can be assembled in sequences. Closed chains of molecules are used to represent membranes, while dynamics is governed by rewriting rules on names.