Modelization and Simulation of Nano Devices in nanoκ calculus

A. Credi[1], M. Garavelli[1], C. Laneve[2],
S. Pradalier[3], S. Silvi[1], and G. Zavattaro[2]

[1] Dipartimento di Chimica "G. Ciamician", Università di Bologna
[2] Dipartimento di Scienze dell'Informazione, Università di Bologna
[3] Ecole Polytechnique, Paris

Abstract. We develop a process calculus – the nanoκ calculus – for modeling, analyzing and validating the properties of molecular devices. The nanoκ calculus is equipped with a simple stochastic models. We then study the modelization and simulation of the behaviour of a molecular shuttle, a basic nano device currently used for building more complex systems.

1 Introduction

In 2006 the University of Bologna has funded an interdisciplinary project of its Departments of Chemistry and Computer Science – the *CompReNDe Project* (Compositional and executable Representations of Nano Devices). The project combines the expertises of two groups, one specialized in the design and construction of devices and machines of molecular size [3, 2] and the other one qualified in formal models, based on the theory of process calculi, for describing and analyzing molecular systems [7, 14]. Such expertises are joined in order to accomplish two main endeavours: (i) deliver a programming model for describing molecular machines that is also amenable to automated simulations and verifications by means of existing algorithms and (ii) apply the model for a formal analysis of complex molecular machines and possibly discover unknown behaviours to the chemical-designers.

The CompReNDe research activity started with the initial goal of formalizing a *[2]rotaxane* [20] into the κ calculus [7] and validating the experiments *in vitro* through simulations *in silico* by means of some contemporary stochastic evaluator [10, 19, 5].

[2]rotaxanes [20] (simply rotaxanes in the following) are systems composed of a molecular axle surrounded by a ring-type (macrocyclic) molecule. Bulky chemical moieties ("stoppers") are placed at the extremities of the axle to prevent the disassembly of the system. In rotaxanes containing two different recognition sites on the axle ("stations"), it is possible to switch the position of the macrocyclic ring between the two stations by an external energy input as illustrated in Figure 1. Several rotaxanes of this kind, known as *molecular shuttles*, have been already developed (see [6] and the references therein) and used for building more complex systems [13, 12, 2].

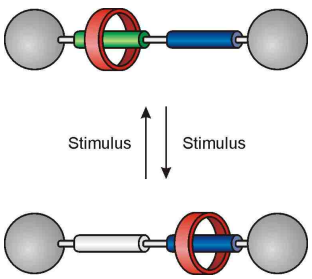

Fig. 1. Schematic representation of a two-station rotaxane and its operation as a controllable molecular shuttle.

The κ calculus is a formal language idealizing protein-protein interactions, as a particular restricted kind of graph-rewriting. Bindings are explicit: proteins are nodes with fixed numbers of sites, complexes are connected graphs built over such nodes where bonds are represented by *names*. Biological reactions are modeled by two kinds of rewriting rules: *complexations*, which create bonds, and *decomplexations*, which destroy bonds. Notably, the κ calculus has been compiled into a process calculus [16], thus making possible the simulations of molecular biology formal descriptions on uniprocessors [17] and distributed systems [21]. The compilation introduced a finer-grained concurrent model, the $\mathtt{m}\kappa$ calculus, where reactions have to be at most binary. The significant property of $\mathtt{m}\kappa$ calculus is that it may be implemented *without changing the granularity of the dynamics*: reactions exactly corresponds to process interactions. This invariance is crucial when quantitative analyses of systems must be undertaken, as it is the case for biological and chemical simulations, which are stochastic (process interactions are marked with the stochastic measure of the corresponding reactions).

We therefore undertook the formalization of a molecular shuttle in $\mathtt{m}\kappa$ calculus and we soon realized that such calculus was inadequate as well. The $\mathtt{m}\kappa$ calculus is too much verbose because it compels designers to reason in terms of bonds and complexations and decomplexations. There are reactions that are neither complexations nor decomplexations, such as the *ion exchanges*. These reactions, used in our molecular shuttle to stimulate the movement of the macrocyclic ring, might be implemented by sequences of complexations and decomplexations, thus changing the granularity of the chemical semantics. The $\mathtt{m}\kappa$ calculus model is too much abstract because it overlooks quantitative aspects. Such aspects, in particular reaction rates and the derived stochastic semantics are a must for providing meaningful simulations and validations of molecular machines.

We overcome to these inadequacies of the $\mathtt{m}\kappa$ calculus by defining a new model, the $\mathtt{nano}\kappa$ *calculus*, having three types of reactions – *creations*, *destructions*, and *updates* – and retaining a stochastic semantics. This stochastic semantics is problematic for the $\mathtt{nano}\kappa$ calculus because it uses names for representing molecular bonds. In this respect, our model is close to Milner's pi calculus [16]. However, instead of following the techniques of the stochastic pi calculus [18],

we have preferred for nanoκ calculus to extend Cardelli's language of stochastic interacting processes [4]. In facts, in this way, we get a simple model that is amenable to automated simulations and verifications by means of existing well known algorithms [9].

We then apply the nanoκ calculus to describe and analyze the "Rotaxane RaH" [15, 1], an instance of rotaxane for which the dynamic behaviour has been experimentally characterized in detail [8]. We have considered two groups of simulations. The first ones are used to validate the model checking whether the experiments reproduced *in silico* coincides with those already performed *in vitro*. The second ones simulate *in silico* the expected behaviour of the Rotaxane RaH under conditions not yet observed *in vitro*. Interestingly, we show that under extreme conditions of very low concentration of Rotaxane RaH, some of the assumptions, usually taken about the behaviour of the rotaxane in standard conditions of concentration, are no longer valid.

Structure of the paper. The next section introduces the nanoκ calculus syntax and stochastic semantics. In Section 3 we relate the nanoκ calculus semantics with Interactive Markov Chains and Continuous Time Markov Chains. In Section 4 we present our case study "[2]Rotaxanes RaH", its modelling into nanoκ calculus, and discuss some results of our simulations.

2 The nanoκ calculus calculus: syntax and semantics

Two disjoint countable sets of names will be used: a set S of *species*, ranged over by A, B, C, $\cdots$; and a totally ordered set B of *bonds*, ranged over by x, y, z, $\cdots$. Species are sorted according to the number of *fields* and *sites* they possess. Let $\mathfrak{s}_f(\cdot)$ and $\mathfrak{s}_s(\cdot)$ be two functions returning naturals; the integers 1, 2, $\cdots$, $\mathfrak{s}_f(A)$ and 1, 2, $\cdots$ $\mathfrak{s}_s(A)$ are respectively the fields and the sites of A. ($\mathfrak{s}_f(A) = 0$ means there is no field; $\mathfrak{s}_s(A) = 0$ means there is no site). In the following, fields are ranged over by h, i, j, $\cdots$; sites are ranged over by a, b, c, $\cdots$.

Sites may be either *bound* to other sites or *unbound*, i.e. not connected to other sites. The state of sites are defined by injective maps, called *interfaces* and ranged over by σ, ρ, $\cdots$. Given a species A, its interfaces are partial functions from $\{1, \cdots, \mathfrak{s}_s(A)\}$ to the set B. A site a is bound with bond x in σ if $\sigma(a) = x$; it is unbound if $a \notin \mathrm{dom}(\sigma)$. For instance, if A is a species with three sites, $(2 \mapsto x, 3 \mapsto y)$ is an interface of its. In order to simplify the reading, we write this map as $2^x + 3^y$. In the following, when we write $\sigma + \sigma'$ we assume that the domains of σ and σ' are disjoint. Interfaces, being injective on bonds, cannot express that the endpoints of a bond belong to the same species (*cf. self complexation* in [7]).

Fields represent the internal state of a species. The values of fields are defined by maps, called *evaluations*, and ranged over by u, v, $\cdots$. For instance, if A is a species with three fields, $[1 \mapsto 5, 2 \mapsto 0, 3 \mapsto 4]$ is an evaluation of its. As before, we write this map as $1^5 + 2^0 + 3^4$. We assume there are finitely many internal states, that is every field h is mapped into values in $\{0, \cdots, n_h\}$. In the following,

we use partial evaluations and, when we write the union of evaluations $u + v$, we implicitly assume that the domains of u and v are disjoint.

Definition 1 (Molecules and Solutions). *A* molecule $A[u](\sigma)$ *is a term where* u *is a total map on the fields of* A. Solutions*, ranged over by* S*,* T*,* $\cdots$*, are defined by the following grammar*

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

The operator ",*" is assumed to be associative, so* $(S\,,S')\,,S''$ *is equal to* $S\,,(S'\,,S'')$ *(and we always omit parentheses).*

Solutions retain the property that bond names always occurs exactly twice. Let $\emptyset$ *be the empty map. We use the following shorthand notations:* $A(\sigma)$ *instead of* $A[\emptyset](\sigma)$*,* $A[u]$ *instead of* $A[u](\emptyset)$*, and simply* A *instead of* $A[\emptyset](\emptyset)$*.*

Example 1. As a running example we consider two typical reversible chemical reactions:

- $Na \;+\; Cl \;\longleftrightarrow\; Na^{+} \;+\; Cl^{-}$ (*sodium chloride*) and
- $H \;+\; H \;\longleftrightarrow\; 2H$ (*hydrogen gas*) .

In the first reaction, an ion is exchanged between two instances of species Na and Cl. The molecules of the two species can be in two possible states: either they have the extra ion Na^{+} and miss an ion Cl^{-} or they are in the states with all the ions Na and Cl. We model these two possible states using one field *ion* with values 0 and 1 respectively denoting the absence or the presence of the ion. Formally we can use $Na[ion^{0}]$ and $Na[ion^{1}]$ for Na and Na^{+}, and $Cl[ion^{0}]$ and $Cl[ion^{1}]$ for Cl^{-} and Cl, respectively.

The second chemical reaction represents the creation/destruction of a bond between two hydrogen atoms. This may be described by using a site b and bond names. For instance, the solution with $2H$ is modelled by $H(b^{x})\,,H(b^{x})$. An unbound instance of hydrogen is simply represented by H, as its evaluation and interface are both empty.)

Definition 2 (Reactions). *Reactions of* nanoκ *calculus are either* creations C*, or* destructions D*, or* exchanges E*. The format of the first two types is* $((A, a, u, u', \sigma),\ (B, b, v, v', \phi), \lambda)$*; while the format of exchanges is* $((A, u, u', \sigma),\ (B, v, v', \phi), \lambda)$*, such that:*

1. $\mathrm{dom}(u') = \mathrm{dom}(u)$ *and* u *and* u' *are partial evaluations of* A*,* $\mathrm{dom}(v') = \mathrm{dom}(v)$ *and* v *and* v' *are partial evaluations of* B*,*
2. $\mathrm{ran}(\sigma) = \mathrm{ran}(\phi)$ *and* σ *and* ϕ *are interfaces of* A *and* B*, respectively;*
3. *and* $\lambda \in \mathbb{R}^{+} \cup \{\infty\}$.

For readability's sake, we write creations as $A[u](a{+}\sigma)\,,B[v](b{+}\phi) \overset{\lambda}{\rightarrow} A[u'](a^{x}{+}\sigma)\,,B[v'](b^{x}{+}\phi)$*, destructions as* $A[u](a^{x}{+}\sigma)\,,B[v](b^{x}{+}\phi) \overset{\lambda}{\rightarrow} A[u'](a{+}\sigma)\,,B[v'](b{+}\phi)$*, and updates as* $A[u](\sigma)\,,B[v](\phi) \overset{\lambda}{\rightarrow} A[u'](\sigma)\,,B[v'](\phi)$*.*

The difference between the three kinds of rules is concerned with the modification of the interfaces: *creations* produce a new bond between the two unbound sites a and b, *destructions* remove the bond between the sites a and b, while *exchanges* leave the interfaces unchanged.[1]

Example 2. The nanoκ calculus reactions that corresponds to the two reactions of the sodium chloride are

$$Na(h^0),Cl(h^1) \overset{100}{\rightarrow} Na(h^1),Cl(h^0)$$
$$Na(h^1),Cl(h^0) \overset{10}{\rightarrow} Na(h^0),Cl(h^1)$$

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

The nanoκ calculus reactions that corresponds to the two reactions of the hydrogen gas are

$$H,H \overset{5}{\rightarrow} H[b^x],H[b^x]$$
$$H[b^x],H[b^x] \overset{0.05}{\rightarrow} H,H$$

where the right direction has been given rate 5 and the left direction has been giver rate 0.05.

The formal definition of *reactants* and the corresponding *products* of reactions follows. We use μ to range over $\rho_L, \imath, x$ and $\rho_R, \imath, x$ and $\rho_L, \imath$ and $\rho_R, \imath$ and ρ. Let $\overline{\mu}$ be the following operation:

$$\overline{\mu} \overset{def}{=} \begin{cases} \rho_R, \imath, x & \text{if } \mu = \rho_L, \imath, x \\ \rho_L, \imath, x & \text{if } \mu = \rho_R, \imath, x \\ \rho_R, \imath & \text{if } \mu = \rho_L, \imath \\ \rho_L, \imath & \text{if } \mu = \rho_R, \imath \\ \rho & \text{if } \mu = \rho \end{cases}$$

We notice that $\overline{\overline{\mu}} = \mu$.

Definition 3 (The basic transition relation). *The basic transition relation of solutions, written* $\xrightarrow{\mu}_{\ell} \cup \xrightarrow{\mu}_{\ell,\ell'}$*, is the least relation that satisfies the following rules ($\imath$ are always injective renamings on bonds):*

- *if* $\rho = A[u](a+\sigma),B[v](b+\phi) \overset{\lambda}{\rightarrow} A[u'](a^x+\sigma),B[v'](b^x+\phi)$ *and* $z \notin \mathrm{ran}(\sigma \circ \imath + \nu)$ *then both* $A[u+w](\sigma \circ \imath + \nu) \xrightarrow{\rho_L,\imath,z}_1 A[u'+w](a^z + \sigma \circ \imath + \nu)$ *and* $B[v+w](\phi \circ \imath + \nu) \xrightarrow{\rho_R,\imath,z}_1 B[v'+w](b^z + \phi \circ \imath + \nu)$*;*
- *if* $\rho = A[u](a^x+\sigma),B[v](b^x+\phi) \overset{\lambda}{\rightarrow} A[u'](a+\sigma),B[v'](b+\phi)$ *then both* $A[u+w](a^x+\sigma\circ\imath+\nu) \xrightarrow{\rho_L,\imath,x}_1 A[u'+w](\sigma\circ\imath+\nu)$ *and* $B[v+w](b^x+\phi\circ\imath+\nu) \xrightarrow{\rho_R,\imath,x}_1 B[v'+w](\phi \circ \imath + \nu)$*;*

[1] In [K,BioK] the terms *complexation* and *decomplexation* are used instead of the terms creation and destruction that are used in nanoκ calculus because they have a more neutral chemical meaning.

- *if* $\rho = A[u](\sigma), B[v](\phi) \overset{\lambda}{\to} A[u'](\sigma), B[v'](\phi)$ *then both* $A[u+w](\sigma \circ \imath + \nu) \xrightarrow{\rho_L, \imath}_1 A[u'+w](\sigma \circ \imath + \nu)$ *and* $B[v+w](\phi \circ \imath + \nu) \xrightarrow{\rho_R, \imath}_1 B[v'+w](\phi \circ \imath + \nu)$;
- *if* $\mathsf{S} \xrightarrow{\mu}_\ell \mathsf{S}'$ *and* $(\mathrm{name}(\mathsf{S}') \backslash \mathrm{name}(\mathsf{S})) \cap \mathrm{name}(\mathsf{T}) = \emptyset$ *then both* $\mathsf{S}, \mathsf{T} \xrightarrow{\mu}_\ell \mathsf{S}', \mathsf{T}$ *and* $\mathsf{T}, \mathsf{S} \xrightarrow{\mu}_{\ell+\ell'} \mathsf{T}, \mathsf{S}'$, *where* T *has* ℓ' *molecules;*
- *if* $\mathsf{S} \xrightarrow{\mu}_\ell \mathsf{S}'$ *and* $\mathsf{T} \xrightarrow{\overline{\mu}}_{\ell'} \mathsf{T}'$ *then* $\mathsf{S}, \mathsf{T} \xrightarrow{\rho}_{\ell, \ell+\ell'} \mathsf{S}', \mathsf{T}'$, *where* ρ *is the rule of* μ. *If* ρ *is a creation, then the bond used by the reaction is the least one that is greater than those in* S, T.

The comments about the transition relation are in order:

1. whenever $\mathsf{S} \xrightarrow{\rho, \imath, x}_\ell \mathsf{T}$ and ρ is a creation, the above definition also admits $\mathsf{S} \xrightarrow{\rho, \imath, y}_\ell \mathsf{T}\{y/_x\}$, where y is fresh. However this nondeterminsm is removed when the reaction occurs because the bond has to be the least name not occurring in S;
2. 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 $\mathsf{S}', \mathsf{S}''$ such that the reactants are in S' and S''.

Example 3. We are in place for describing the (basic) dynamics of a solution. We write $k\mathsf{M}$ for $\underbrace{\mathsf{M}, \cdots, \mathsf{M}}_{k \text{ times}}$. Let ρ and ρ' be the two sodium chloride reactions and ϱ and ϱ' be the two hydrogen gas reactions. Then both

$$2Na[ion^0], Cl[ion^1], 3H \xrightarrow{\rho}_{1,3} Na[ion^1], Na[ion^0], Cl[ion^1], 3H$$

and

$$2Na[ion^0], Cl[ion^1], 3H \xrightarrow{\rho}_{2,3} Na[ion^0], Na[ion^1], Cl[ion^1], 3H$$

are possible. As regards the hydrogen gas, we have transitions $\xrightarrow{\varrho}_{4,5}$, $\xrightarrow{\varrho}_{4,6}$, and $\xrightarrow{\varrho}_{5,6}$. The reader is encouraged to write the complete transition system.

The basic transition relation is too much intensional. Consider a solution containing hundreds of molecules of the species A and B that could react with ρ. The relation $\xrightarrow{\mu}_{\ell, \ell'}$ distinguishes the two pairs of reactants, and this is not possible in practice. A more sensible transition relation should represent *collectively* all the possible combinations of one molecule of species A with one molecule of species B. For instance, the solution A, A, B transits with $\xrightarrow{\rho}_{1,3}$ and $\xrightarrow{\rho}_{2,3}$, with rates are $rate(\rho)$. Abstracting out the order of the molecules, we obtain a unique transition whose rate is $2 * rate(\rho)$. However quotienting the solutions with commutativity axioms of "$,$" does not yield an adequate extensionality. In facts, when ρ is a destruction, between A and B, the solution $A(a^x), A(a^y), B(a^x), B(a^y)$ transits with $\xrightarrow{\rho}_{1,3}$ and $\xrightarrow{\rho}_{2,4}$ into two solutions that cannot be equated by permutations of the molecules in the solution. In these cases one has to use injective renamings of bonds.

Definition 4 (Structural equivalence). *The structural equivalence between solutions, noted $\equiv$, is the least equivalence satisfying the following two rules (we remind that solutions are already quotiented by associativity of ","):*

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

Example 4. Commutativity of the structural equivalence, permits to prove the following equivalence

$$Na[h^0],Cl[h^1] \;\equiv\; Cl[h^1],Na[h^0]$$

while injective renaming permits to prove that

$$H(b^x),H(b^x) \;\equiv\; H(b^y),H(b^y)$$

Combining both commutativity and injective renaming we can prove that

$$H(b^x),H(b^x),H(b^z),H(b^z) \;\equiv\; H(b^y),H(b^k),H(b^k),H(b^y)$$

Proposition 1. *Let* $\mathsf{S} \equiv \mathsf{S}'$.

1. *If* $\mathsf{S} \xrightarrow{\mu}_{\ell} \mathsf{T}$ *then there is* T' *and a renaming* $\imath$ *such that* $\mathsf{S}' \xrightarrow{\imath(\mu)}_{\ell'} \mathsf{T}'$ *and* $\mathsf{T}' \equiv \mathsf{S}'$;
2. *if* $\mathsf{S} \xrightarrow{\rho}_{\ell,\ell'} \mathsf{T}$ *then there is* T' *such that* $\mathsf{S}' \xrightarrow{\rho}_{\ell'',\ell'''} \mathsf{T}'$ *and* $\mathsf{T}' \equiv \mathsf{S}'$.

The following notations are relevant for the definition of the collective transition relation:

- $rate(\rho)$ returns the rate of the reaction ρ;
- $next(\mathsf{S}) = \{(\rho_{\ell,\ell'},\mathsf{T}) \mid \mathsf{S} \xrightarrow{\rho}_{\ell,\ell'} \mathsf{T}\}$; $next_\infty(\mathsf{S}) = \{(\rho_{\ell,\ell'},\mathsf{T}) \mid \mathsf{S} \xrightarrow{\rho}_{\ell,\ell'} \mathsf{T}$ and $rate(\rho) = \infty\}$;
- $\mathcal{S}$ *has finite rates* if, for every $(\rho_{\ell,\ell'},\mathsf{T}) \in \mathcal{S}$, $rate(\rho)$ is not ∞;
- let $\mathcal{S}$ be a set of pairs (X,T') (the second element is a solution; the first one is not specified), $[\mathcal{S}]_{\mathsf{T}}$ is the subset of $\mathcal{S}$ of those pairs (X,T') such that $\mathsf{T}' \equiv \mathsf{T}$;
- $can(\mathcal{S})$ is defined over sets of pairs (X,T) such that the solutions occurring as second element of the pairs are all structurally equivalent. It returns a solution S such that there is X with $(X,\mathcal{S}) \in \mathsf{S}$.

Definition 5 (The collective transition relation). *The* nanoκ *calculus collecive transition relation* $\xmapsto{\lambda}$, *where* $\lambda \in \mathbb{R}^+ \cup \{\infty\}$, *is the least relation satisfying the following rules:*

- *if* $\mathsf{S} \xrightarrow{\rho}_{\ell,\ell'} \mathsf{T}$ *and* $rate(\rho) = \infty$ *then* $\mathsf{S} \xmapsto{\infty} can([next_\infty(\mathsf{S})]_{\mathsf{T}})$;
- *if* $\mathsf{S} \xrightarrow{\rho}_{\ell,\ell'} \mathsf{T}$ *and* $next(\mathsf{S})$ *has finite rates then* $\mathsf{S} \xmapsto{\lambda} can([next(\mathsf{S})]_{\mathsf{T}})$, *where*

$$\lambda \;=\; \sum_{(\rho_{\ell,\ell'},\mathsf{T}') \in [next(\mathsf{S})]_{\mathsf{T}}} rate(\rho)$$

We notice that, by definition, the nanoκ calculus collective transition system is such that there is no state with outgoing $\stackrel{\infty}{\longmapsto}$ *and* $\stackrel{\lambda}{\longmapsto}$ (λ finite) transitions. In the following, the states with $\stackrel{\infty}{\longmapsto}$ outgoing transitions are called *transient states*, the other ones are called *markovian states*.

The interrelation between basic and collective transition relations is as follows: the collective one partitions the products of a solution (according to the basic transition relation) into equivalence classes, takes a canonical representative of the class, and defines a transition whose label is the sum of the rates of the reactions in the basic one that yield solutions in the equivalence class.

Example 5. As an example of collective transitions we consider the basic transitions of sodium chloride discussed in the Example 3. Since the products are structural equivalent, we obtain a unique IMC transition:

$$2Na[ion^0], Cl[ion^1], 3H \stackrel{200}{\longmapsto} Na[ion^1], Na[ion^0], Cl[ion^1], 3H$$

We also observe that there is a unique IMC transition $\stackrel{15}{\longmapsto}$ outgoing the initial solution and corresponding to $\stackrel{\varrho}{\longrightarrow}_{4,5}$, $\stackrel{\varrho}{\longrightarrow}_{4,6}$, and $\stackrel{\varrho}{\longrightarrow}_{5,6}$.

3 Stochastic semantics of nanoκ calculus

The collective transition relation of nanoκ calculus corresponds to an *Interactive Markov Chain (IMC) transition system* with only *silent* interactive transitions [11]. In particular, the $\stackrel{\infty}{\longmapsto}$-transitions are silent interactive transitions that are executed in the IMC model instantaneously and under the *maximal progress assumption*. That is, the so-called *sojourn time* in a transient state is 0, which amounts to favour silent transitions to those labelled with finite rates (called *markovian* transitions). On the contrary, in a markovian state with n outgoing markovian transitions labelled $\lambda_1, \cdots, \lambda_n$, the probability that the sojourn time is less than t is exponentially distributed with rate $\lambda = \lambda_1 + \cdots + \lambda_n$, i.e. $\mathsf{Prob}\{delay < t\} = 1 - e^{-t \sum_i \lambda_i}$, and the probability that the j-th transition is taken is $\lambda_j / (\sum_i \lambda_i)$.

However the models underlying traditional simulation algorithms such as [9] are *Continuos Time Markov transition systems (CTMC)* that do not comprise interactive transitions. Having a CTMC is therefore primary to run automatic analysis tools for experimenting *in silico* the dynamics of nano-machines specifications in nanoκ calculus.

The mismatch between IMC with only silent actions and CTMC systems is due to two main reasons: (i) the nondeterminism and (ii) the persistency of the silent transitions. As regards (i), consider two silent actions that apply to the same reactants and give two different products. If these products have only markovian transitions it is not possible to collect them in a unique solution. As regards (ii), if an infinite sequence of silent transitions exists then the simulation time of the CTMC system will not advance anymore. Therefore collapsing all

these transitions, by identifying the initial and final solutions of the sequence, is again not possible.

However, there are cases where the downgrading of an IMC system to a CTMC one is possible without modifying the semantics. This is *when all silent actions may be partitioned into confluent finite directed acyclic graphs*. In fact, when the silent transitions are partitioned into confluent directed acyclic graphs, there are no loops (there is no infinite sequence of silent transitions), and all sequences of silent transitions starting from the same state share the same final state, to which the initial state may be safely collapsed. The meaning of this collapse is that we are removing a finite amount of work which is performed in zero time.

The formal definition of downgrading of IMC to CTMC systems follows. We first introduce the auxiliary function *next markovian state* defined on solutions and yielding sets:

– $nextm(\mathsf{S}) = \{((\lambda, \mathsf{T}'), \mathsf{T}) \mid \mathsf{S} \stackrel{\lambda}{\longmapsto} \mathsf{T}' \stackrel{\infty}{\longmapsto}^{*} \mathsf{T} \text{ and } \lambda \in \mathbb{R}^{+} \text{ and } \mathsf{T} \stackrel{\infty}{\not\longmapsto}\}$

We notice that $nextm(\mathsf{S})$ is undefined when S is transient.

Definition 6 (Downgrading of an IMC system). *An IMC system* $(\mathcal{S}, \stackrel{\lambda}{\longmapsto})$ *is* strictly-markovian *if*

1. *states are either transient or markovian and*
2. *every subsystem consisting of silent transitions is a confluent and finite direct acyclic graph.*

Let $(\mathcal{S}, \stackrel{\lambda}{\longmapsto})$ *be strictly-markovian; the transition relation* $\stackrel{\nu}{\Longmapsto}$*, where* $\nu \in \mathbb{R}^{+}$*, is the least one such that:*

– *if* S *is markovian then* $\mathsf{S} \stackrel{\nu}{\Longmapsto} can([nextm(\mathsf{S})]_{\mathsf{T}})$ *with*

$$\nu = \sum_{((\lambda,\mathsf{T}'),\mathsf{T}'') \in [nextm(\mathsf{S})]_{\mathsf{T}}} \lambda$$

It is easy to verify that the relation $\stackrel{\nu}{\Longmapsto}$ defines a CTMC system. Moreover, the properties below are direct consequences of the construction:

– the probability distribution of the sojourn time in a markovian state is the same in the IMC and in the downgraded CTMC and
– the probability that one of the paths $\mathsf{S} \stackrel{\lambda}{\longmapsto} \stackrel{\infty}{\longmapsto}^{*} \mathsf{T}'$ with $\mathsf{T}' \equiv \mathsf{T}$ is taken in the IMC corresponds to the probability the unique transition $\mathsf{S} \stackrel{\lambda'}{\Longmapsto} \mathsf{T}''$, with $\mathsf{T}'' \equiv \mathsf{T}$, is taken in the downgraded CTMC.

Actually the correspondence between strictly-markovian IMC and the associated CTMC is much stronger: the IMC semantics, the *markovian bisimulation* [11], is still a markovian bisimulation on the CTMC when restricted to its states. We will detail the correspondence in the full paper.

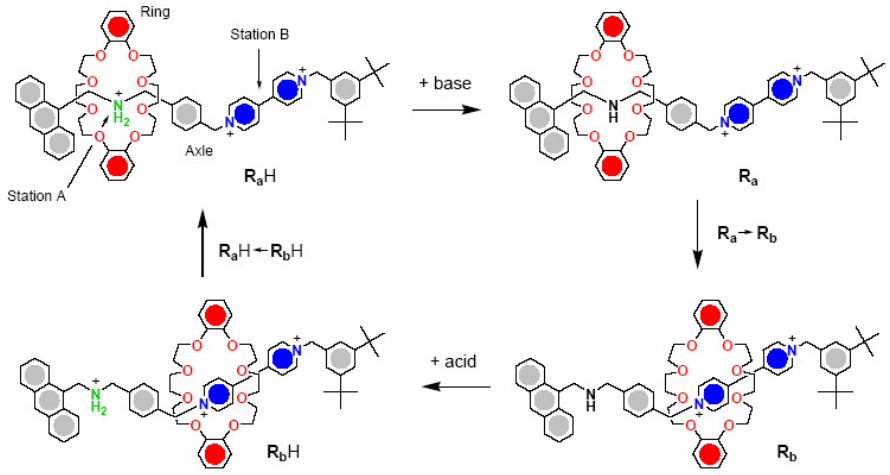

Fig. 2. Schematic representation of the shuttling processes of the molecular ring.

4 nanoκ calculus at work: the rotaxane case study

The investigated Rotaxane RaH (Figure 2) is made of a dumbbell component containing an ammonium (A) and an electron acceptor bipyridinium (B) units 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 (Figure 2). 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 unit. 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 [1] that the acid-base controlled switching, which is fully reversible and relatively fast, exhibits a clear-cut on-off behaviour.

The Rotaxane RaH is particularly appropriate to test the modeling approach described in the present paper because it is one of the very few cases wherein not only the thermodynamic properties, but also the dynamic behavior of the system has 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 [8]. The rate constants for the "forward" (Ra→Rb) and "backward" (RbH→RaH) shuttling motions (horizontal processes in Figure 2) of the molecular ring which occur, respectively, upon deprotonation and reprotonation of the ammonium/amine recognition site on the axle (vertical processes in Figure 2), were found to be $0.72s^{-1}$ and $40s^{-1}$ at $293K$, respectively.

4.1 Modeling the Rotaxane RaH in nanoκ calculus

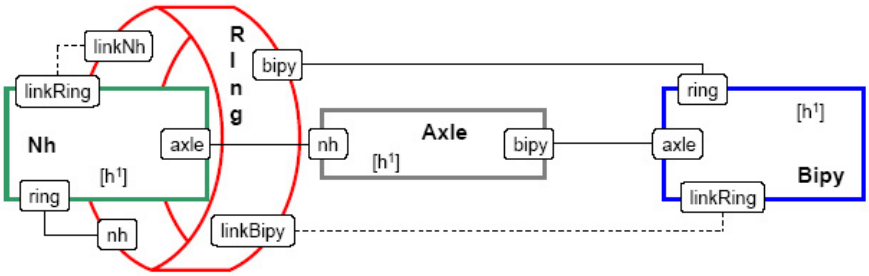

Fig. 3. Initial state of the Rotaxane RaH in nanoκ calculus.

The nanoκ calculus molecules. Figure 3 reports the nanoκ calculus modeling of the Rotaxane RaH. We use four nanoκ calculus species:

- `Nh` models the amonium/amine station of the rotaxane: it has one field h and three sites *ring, axle* and *linkRing*;
- `Axle` models the axle between the two stations: it has one field h and two sites *nh* and *bipy*;
- `Bipy` models the bipyridinium station: it has one field h and three sites *ring, axle* and *linkRing*;
- `Ring` models the crown ether ring: it has no field and four sites *nh, bipy, linkNh* and *linkBipy*.

Bonds of much different chemical nature appears in RaH: covalent bonds, hydrogen bonds and charge-transfer interactions. Yet we can simply model all of them as nanoκ calculus bonds because here it is sufficient to know the speed of the interactions while their nature is not important.

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 structure of the rotaxane. The pairs of sites *nh* of `Ring` and *ring* of `Nh`, and *bipy* of `Ring` and *ring* of `Bipy`, are also permanently linked. They model the fact that the ring molecule cannot "escape" from the rotaxane. This implicitly models the stoppers. Having two links, one to `Nh` and one to `Bipy`, permits reactions to test the belonging of two molecules `Nh` (or `Bipy`) and `Ring` to the same rotaxane (see reactions 7 and 8 below).

The pairs of sites *linkRing* of `Nh` and *linkNh* of `Ring`, and *linkRing* of `Bipy` and *linkBipy* of `Ring` can also be linked. They model the position of the ring: when the interaction between the ammonium station and the ring is the strongest, the former bond exists and not the latter, and vice versa if the strongest interaction is the one with the bipyridinium station.

Ammonium and amine functions have different chemical nature but can be seen as protonated and deprotonated version 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.

As the `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 consider the species `Acid` and `Base` both with one field h having value 1 in case the acid/base molecule holds the proton to be exchanged, 0 otherwise (for instance $\texttt{Acid}[h^1]$ and $\texttt{Base}[h^0]$ are respectively an acid molecule ready to give a proton and a base molecule ready to receive a proton).

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

$$\begin{array}{l}\texttt{Nh}[h^1](ring^{r_1} + axle^{s_1} + linkRing^x)\ ,\\ \texttt{Axle}[h^1](nh^{s_1} + bipy^{s_2})\ ,\\ \texttt{Bipy}[h^1](ring^{r_2} + axle^{s_2} + linkRing)\ ,\\ \texttt{Ring}(nh^{r_1} + bipy^{r_2} + linkNh^x + linkBipy)\end{array}$$

graphically depicted in Figure 3.

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

The nanoκ calculus reactions. We can now present the reactions used in our modeling. Reactions 1, 2, 7 and 8 are presented with a double arrow; such reactions are called *reversible*. Formally they correspond to two nanoκ calculus reactions, one achieved reading the reaction from left to right considering the rate over the arrow, and another one 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 4.2.

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 models this phenomena. The systems corresponding to the left-hand side and right-hand side coexist, even if one can be much predominant according to the ratio $nh\_base/base\_nh$ (and $acid\_nh/nh\_acid$).

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

$$\texttt{Nh}[h^0], \texttt{Acid}[h^1] \underset{nh\_acid}{\overset{acid\_nh}{\rightleftarrows}} \texttt{Nh}[h^1], \texttt{Acid}[h^0] \quad (2)$$

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. It is crucial that these reactions are immediate, since in chemistry that are no such reactions, the information is instantaneously known by the whole rotaxane.

if $(\alpha \neq \beta)$

$$\texttt{Nh}[h^{\alpha}](axle^{x}), \texttt{Axle}[h^{\beta}](nh^{x}) \overset{\infty}{\rightarrow} \texttt{Nh}[h^{\alpha}](axle^{x}), \texttt{Axle}[h^{\alpha}](nh^{x}) \qquad (3)$$

and:

$$\texttt{Axle}[h^{\alpha}](bipy^{x}), \texttt{Bipy}[h^{\beta}](axle^{x}) \overset{\infty}{\rightarrow} \texttt{Axle}[h^{\alpha}](bipy^{x}), \texttt{Bipy}[h^{\alpha}](axle^{x}) \qquad (4)$$

We achieve the modeling of Ring movements in two steps. Firstly the bond between Ring and the station to which he is currently linked is destroyed (reactions 5 and 6), and secondly the Ring move to the other station (reaction 7 and 8). This two-steps mechanism permits to model not only the movement of the ring but also the transient state. Reactions 5 and 6 have infinite rates since they represent immediate consequences of protonation or deprotonation of Nh. Reactions 7 and 8 are reversible, because due to the Brownian motion the Ring is susceptible to return to the previous station.

$$\texttt{Nh}[h^{0}](linkRing^{x}), \texttt{Ring}(linkNh^{x}) \overset{\infty}{\rightarrow} \texttt{Nh}[h^{0}](linkRing), \texttt{Ring}(linkNh) \qquad (5)$$

$$\texttt{BiAx}[h^{1}](linkRing^{x}), \texttt{Ring}(linkBiax^{x}) \overset{\infty}{\rightarrow} \texttt{BiAx}[h^{1}](linkRing), \texttt{Ring}(linkBiax) \qquad (6)$$

$$\texttt{Ring}(bipy^{x} + linkNh + linkBiax), \texttt{Bipy}[h^{0}](ring^{x} + linkRing) \underset{unlink\_bipy}{\overset{link\_bipy}{\rightleftarrows}} \texttt{Ring}(bipy^{x} + linkNh + linkBiax^{z}), \texttt{Bipy}[h^{0}](ring^{x} + linkRing^{z}) \qquad (7)$$

$$\texttt{Ring}(nh^{x} + linkNh + linkBiax), \texttt{Nh}[h^{1}](ring^{x} + linkRing) \underset{unlink\_nh}{\overset{link\_nh}{\rightleftarrows}} \texttt{Ring}(nh^{x} + linkNh^{z} + linkBiax), \texttt{Nh}[h^{1}](ring^{x} + linkRing^{z}) \qquad (8)$$

4.2 Simulation results

It is not difficult to see that the modeling of Rotaxane RaH in nanoκ calculus reported above satisfies the properties described in the Definition 6 (all sequences of ∞ labelled transitions rooted at the same state are finite and convergent); thus we can achieve, using the construction reported in the definition, a CTMC semantics that we use to simulate *in silico* the behaviour of the Rotaxane RaH.

As previously discussed the rates for the ring movements are respectively $link\_bipy = 0.72$ and $link\_nh = 40$. The rates for the reverse reactions are quantified two orders of magnitude smaller, i.e. $unlink\_bipy = 0.0072$ and $unlink\_nh = 0.4$.

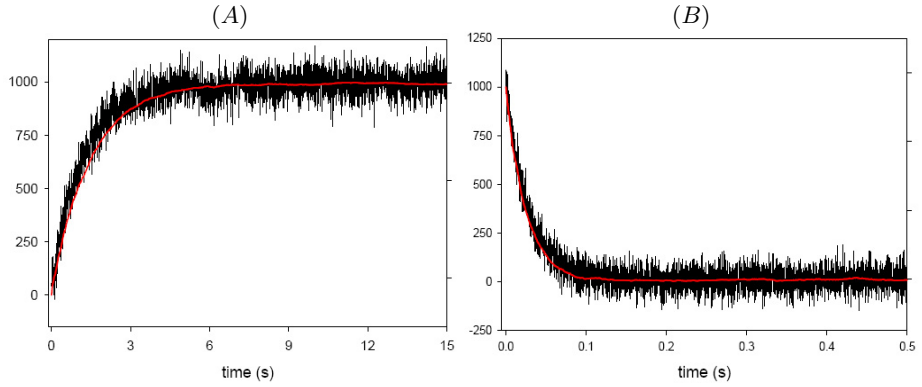

Fig. 4. Comparing the *in vitro* –black line– and the *in silico* –red line– experiments: number of `Ring`s linked to `Bipy`s during the "forward" Ra→Rb (part A) and the "backward" RbH→RaH (part B) shuttlings.

The aim of the first two simulations depicted in Figure 4 is to check whether the experimentation *in silico* corresponds to the experimentation in *in vitro* [8]. The techniques used for the *in vitro* experimentation did not permit to 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. $nh\_base = acid\_nh = \infty$ and $base\_nh = nh\_acid = 0$. 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 strike 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 2, 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 acid/base molecules and rotaxanes. In fact, this is a bimolecular reaction with rate is influenced by

the concentration of the reactants. For instance, at the concentration considered in [8], i.e. $10^{-5}M$, a plausible rate for deprotonation/reprotonation is $2 \times 10^{5} s^{-1}$ (with reverse reaction with rate $2 \times 10^{-2} s^{-1}$) while at the concentration $10^{-8}M$ it is $2 \times 10^{2} s^{-1}$ (with reverse reaction with rate $2 \times 10^{-5} s^{-1}$).

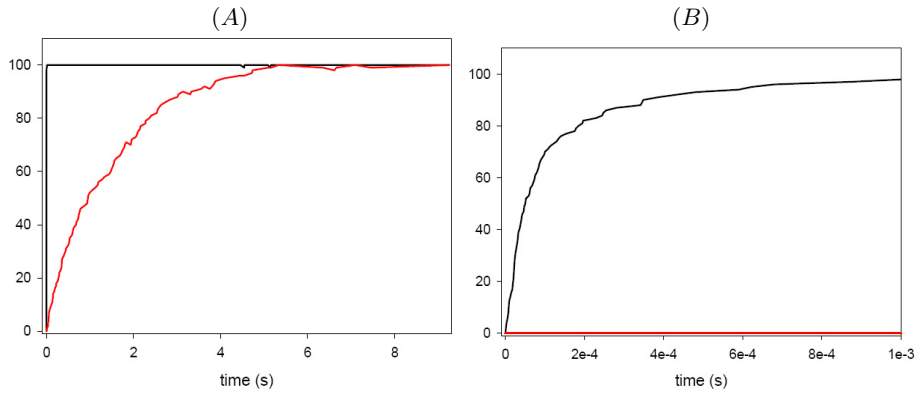

Fig. 5. Comparing the number of deprotonated rotaxanes –black line– with the number of `Ring`s linked to `Bipy`s –red line– during Ra→Rb shuttling at concentration $10^{-8}M$.

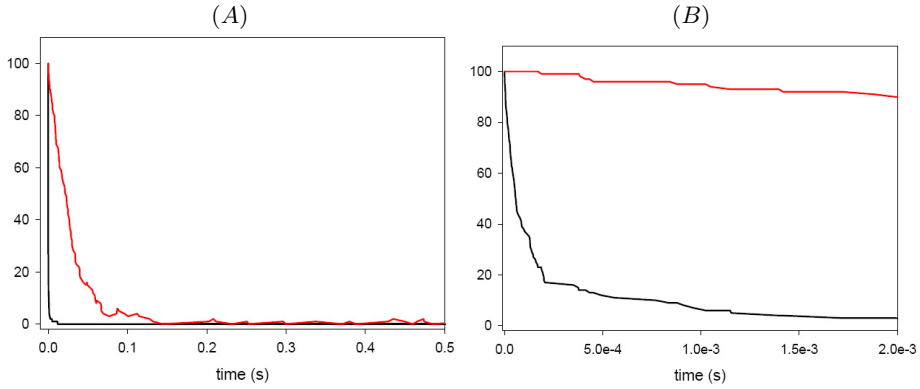

Fig. 6. Comparing the number of deprotonated rotaxanes –black line– with the number of `Ring`s linked to `Bipy`s –red line– during RbH→RaH shuttling at concentration $10^{-8}M$.

We have repeated the two simulations: deprotonation with subsequent "forward" shuttling and reprotonation with subsequent "backward" shuttling considering the two different concentrations. We had to reduce the number of rotaxanes and acid/base molecules to 100 because the simulation required too much time

with 1000 molecules. The increment of the simulation time is due to the presence of reverse deprotonation/reprotonation reactions that were considered absent in the simulation reported Figure 4, i.e. these reactions had rate equal to 0.

The results at concentration $10^{-5}M$ are not reported in the paper as they essentially confirm the validity of the "instantaneous" deprotonation/reprotonation assumption. We report only the results at concentration $10^{-8}M$. In Figure 5 part (A) the curves of deprotonation and consequent "forward" shuttling are reported for the whole duration of the simulation (8 seconds) while in part (B) we focus on the the deprotonation phase that took only around 10^{-3} seconds. Figure 6, reporting the simulation results for the reprotonation and consequent "forward" shuttling, is definitely more interesting; it shows that the rings start moving before the reprotonation phase completes. This proves that in the Rotaxane RaH the stimulus and the subsequent shuttling could interplay. This opens interesting scenarios that requires further investigation; for instance, it could be the case that using weak acid/base molecules (for which the ratio between the deprotonation/reprotonation rate and the reverse rate is smaller) the interplay between the stimulus and the shuttling could give rise to currently unknown emerging behaviours.

References

1. P. R. Ashton, R. Ballardini, V. Balzani, A. Credi I. Baxter, M. C. T. Fyfe, M. T. Gandolfi, M. Gómez-López, M.-V. Martínez-Díaz, A. Piersanti, N. Spencer, J. F. Stoddart, M. Venturi, A. J. P. White, and D. J. Williams. Acid-base controllable molecular shuttles. *Journal of Am. Chem. Soc.*, 120:11932–11942, 1998.
2. J. D. Badjic, V. Balzani, A. Credi, S. Silvi, and J. F. Stoddart. A molecular elevator. *Science*, 303:1845–1849, 2004.
3. Vincenzo Balzani, Alberto Credi, and Margherita Venturi. Molecular devices and machines – concepts and perspectives for the nano world. In *Wiley-VCH*, Weinheim, 2007.
4. Luca Cardelli. On process rate semantics. 2007.
5. Luca Cardelli and Andrew Phillips. Spim homepage. At `research.microsoft.com/~aphillip/spim/`, 2006.
6. Benoit Champin, Pierre Mobian, and Jean-Pierre Sauvage. Transition metal complexes as molecular machine prototypes. *Chemical Society Reviews*, 36(2):358–366, 2006.
7. Vincent Danos and Cosimo Laneve. Formal molecular biology. *Theoritical Computer Science*, 325(1):69–110, 2004.
8. S. Garaudée, S. Silvi, M. Venturi, A. Credi, A.H. Flood, and J. F. Stoddart. Shuttling dynamics in an acid-base-switchable [2]rotaxane. *ChemPhysChem*, 6:2145, 2005.
9. Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem*, 81:2340–2361, 1977.
10. Stephen Gilmore and Jane Hillston. The pepa workbench: a tool to support a process algebra-based approach to performance modelling. In *Proceedings of the 7th international conference on Computer performance evaluation : modelling techniques and tools*, pages 353–368, Secaucus, NJ, USA, 1994. Springer-Verlag New York, Inc.

11. H. Hermanns. *Interactive Markov Chains: The quest for quantitative quality*, volume 2428 of *LNCS*. Springer-Verlag, Amsterdam, 2002.
12. T.-J. Huang, B. Brough, C.-M. Ho, Y. Liu, A.H. Flood, P.A. Bonvallet, H.-R. Tseng, J.F. Stoddart, M. Baller, and S. Magonov. A nanomechanical device based on linear molecular motors. *Applied Physics Letters*, 85(22):5391–5393, 2004.
13. M.C. Jimenez, C. Dietrich-Buchecker, and J.-P. Sauvage. Towards synthetic molecular muscles: Contraction and stretching of a linear rotaxane dimer. *Angew. Chem. Int. Ed.*, 39(18):3284–3287, 2000.
14. Cosimo Laneve and Fabien Tarissan. A simple calculus of protein and cells. In *Proceeding of MeCBIC'06*, volume 1677 of *ENTCS*, 2007.
15. M.-V. Martínez-Díaz, N. Spencer, and J. F. Stoddart. The self-assembly of a switchable [2]rotaxane. *Angew. Chem. Int. Ed. Engl.*, 36:1904–1907, 1997.
16. Robin Milner. *Communicating and mobile systems: the π-calculus*. Cambridge University Press, Cambridge, 1999.
17. Benjamin C. Pierce and David N. Turner. Pict: A programming language based on the pi-calculus. In Gordon Plotkin, Colin Stirling, and Mads Tofte, editors, *Proof, Language and Interaction: Essays in Honour of Robin Milner*, pages 455–494. MIT Press, 2000.
18. Corrado Priami. Stochastic pi-calculus. *Computer Journal*, 38(7):578–589, 1995.
19. Corrado Priami, Aviv Regev, Ehud Shapiro, and William Silverman. Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Information Processing Letters*, 80:25–31, 2001.
20. J.-P. Sauvage and C. O. Dietrich-Bucheker (eds.). Molecular catenanes, rotaxanes and knots. In *Wiley-VCH*, Weinheim, 1999.
21. Pawel Tomasz Wojciechowski. *Nomadic Pict: Language and Infrastructure Design for Mobile Computation*. PhD thesis, Computer Laboratory, University of Cambridge, 2000.