UNIVERSITÀ
DEGLI STUDI
DI UDINE

Università degli studi di Udine

A framework for protein and membrane interactions

Original

Availability:
This version is available http://hdl.handle.net/11390/878935 since 2016-11-26T13:24:21Z

Publisher:

Published
DOI:10.4204/EPTCS.11.2

Terms of use:
The institutional repository of the University of Udine (http://air.uniud.it) is provided by ARIC services. The aim is to enable open access to all the world.

Publisher copyright

(Article begins on next page)

06 September 2026

A framework for protein and membrane interactions

Giorgio Bacci
University of Udine
giorgio.bacci@dimi.uniud.it

Davide Grohmann
University of Udine
grohmann@dimi.uniud.it

Marino Miculan
University of Udine
miculan@dimi.uniud.it

We introduce the Bioβ Framework, a meta-model for both protein-level and membrane-level interactions of living cells. This formalism aims to provide a formal setting where to encode, compare and merge models at different abstraction levels; in particular, higher-level (e.g. membrane) activities can be given a formal biological justification in terms of low-level (i.e., protein) interactions.

A Bioβ specification provides a *protein signature* together a set of *protein reactions*, in the spirit of the κ-calculus. Moreover, the specification describes when a protein configuration triggers one of the only two membrane interaction allowed, that is "pinch" and "fuse".

In this paper we define the syntax and semantics of Bioβ, analyse its properties, give it an interpretation as biobigraphical reactive systems, and discuss its expressivity by comparing with κ-calculus and modelling significant examples.

Notably, Bioβ has been designed *after* a bigraphical metamodel for the same purposes. Hence, each instance of the calculus corresponds to a bigraphical reactive system, and vice versa (almost). Therefore, we can inherith the rich theory of bigraphs, such as the automatic construction of labelled transition systems and behavioural congruences.

1 Introduction

Cardelli in [5] has convincingly argued that the various *biochemical toolkits* identified by biologists can be described as a hierarchy of *abstract machines*, each of which can be modelled using methods and techniques from concurrency theory. Like other complex situations, it seems unlikely to find a single notation covering all aspects of a whole organism; instead, several models, often presented as process calculi, have been proposed in literature, each focusing on specific aspects of the biological system, at different levels of abstractions. This arises the problem of how to *integrate* these models. Indeed, these machines operate in concert and are highly interdependent: "to understand the functioning of a cell, one must understand also how the various machines interact" [5]. To this end, we need a general *meta*model, that is, a *framework*, where these models (possibly at different abstraction levels) can be encoded, and their interactions can be formally described. Eventually "we will need a single notation in which to describe all machines, so that a whole organism can be described" [5].

As a step towards this long-term goal, in this paper we present Bioβ, a framework calculus for dealing with both protein-level and membrane-level interactions. Many specific models, with protein interactions and membrane trafficking, can be encoded in Bioβ by providing a *specification*. Notably, different specifications can be merged, which corresponds to "put together" different models and systems, often at different levels of abstraction. This allows to investigate the interactions between these models, which otherwise would be kept separate; for instance, this allows to foresee *emerging properties*, such as behaviours due to interactions of different abstraction levels, and which cannot be observed within a single machine model due to its intrinsic abstractions.

A fundamental design choice is that this framework has to be *biologically sound*, i.e., it must admit only systems and reactions which are biologically meaningful, especially at lower level machines (i.e. protein). This is important because in this way, encoding a given model in the framework provides

Gabriel Ciobanu (Ed.): Membrane Computing
and Biologically Inspired Process Calculi 2009
EPTCS 11, 2009, pp. 19–33, doi:10.4204/EPTCS.11.2

© G. Bacci & D. Grohmann & M. Miculan
This work is licensed under the Creative Commons
Attribution-Noncommercial-No Derivative Works License.

automatically a formal, biologically sound justification for the model (or "implementation") in terms of protein reactions and explains how its membrane-level interactions are realised by protein machinery. Also, the representation of a specific model is less error-prone.

To this end, Bioβ has been designed after a bigraphical metamodel, called *biobigraphs* and *biological bigraphical reactive systems (BioRS)* and presented in a companion paper [3], for dealing with both protein-level and membrane-level interactions. Actually, Bioβ can be seen as a syntactic formalism for representing precisely biobigraphical reactive systems. However, for sake of simplicity and lack of space, in this paper we will not discuss the connection between Bioβ and biobigraphs, which will be presented in forthcoming work. For the purposes of the present paper, it suffices to know that biobigraphs, and hence Bioβ, are *adequate* with respect to Danos and Laneve's κ*-calculus*, one of the most accepted formal model of protein systems: we can describe *all* and *only* protein configurations and interactions of the κ-calculus. (Is important to notice, however, that our methodology is general, and can be applied to other formal protein models.) Hence, the Bioβ Framework can be seen somehow as an extension of the κ-calculus, adding biological compartments (that is membranes); however, this is a consequence of the fact that the underlying model, the biobigraphs, are adequate w.r.t. the κ-calculus.

When membranes come into play one wants to deal also with membrane transport, hence, we have to take into consideration reactions like endocytosis, exocytosis, cellular fusion and vesiculation. Actually, as observed by biologists (see [1] and [2, Ch. 15]), membrane interactions present always a "preparation phase", where some proteins (receptors and ligands) interact to get in place, followed by the actual "membrane reconfiguration" phase, which can be either a *pinch* or a *fuse* [6]. Therefore, biobigraphs, and Bioβ, provide just two general rules for all membrane interactions, whilst the preparation phase depends on the specific proteins involved and hence left to the encoding of the specific model.

Summarizing, for encoding a given model one has just to instantiate this general schema by specifying (a) a *protein signature*, that is a set of abstract proteins; (b) a set of *protein rules*, describing protein interactions ignoring compartments; (c) a set of *protein configurations* which trigger a mobility reaction. Then, all other reactions (mobility, in particular) are automatically provided by the framework, in a sound way. Hence, modelers can focus on protein-level aspects, leaving to the framework the burden to deal with biological compartments and membrane transport.

A further motivation for our framework comes from the many general results provided by bigraphs. We mention here only the construction of compositional bisimilarities [10], allowing to prove that two systems are *observational equivalent*, that is, they can be exchanged in any organism without that the overall behaviour will change. Also for this application it is important to restrict the systems allowed by the framework to only those biologically meaningful.

Synopsis In Section 2 we present the Bioβ framework: its syntax, a type system for characterizing well-formed (i.e., biologically meaningful) terms, and general operational semantics. As an example application of this framework, in Section 3 we model the mechanism of membrane traffic. In Section 4 we provide a formal connection between κ-calculus and Bioβ. Related work is discussed in Section 5, and in Section 6 we draw some conclusion and plan future work.

2 The Bioβ Framework

In this section we present the Bioβ framework. Although one should not think of Bioβ necessarily as an extension of the κ-calculus, its connections with that model are quite strong, and hence it is convenient to reuse notations and concepts from κ.

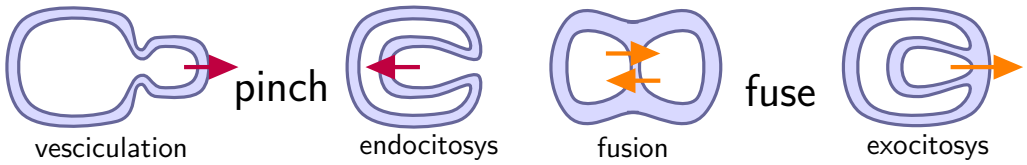

Figure 1: Vesiculation and endocytosis can be generalized by means of an inward and outward pinch, and fusion and exocytosis by a horizontal and vertical fuse, respectively.

2.1 Syntax

The language of the Bioβ framework is parametric over a *protein signature* ($\mathscr{P} = \mathscr{P}_p \cup \mathscr{P}_{ap}, s : \mathscr{P} \to \mathbb{N}$), where $\mathscr{P}$ is a finite set of (abstract) proteins, partitioned into *polar* $\mathscr{P}_p$ and *polar* $\mathscr{P}_{ap}$. $\mathscr{N}$ is an enumerable set of edge names, and $s\colon \mathscr{P} \to \mathbb{N}$ assigns to every protein the number of its *domain sites*.

A *(protein) interface* is a map from $\mathbb{N}$ to $\mathscr{N} + \{h, v\}$ (ranged over ρ, σ, ...). Given an interface ρ and a protein name $A \in \mathscr{P}$, a site (A, i) is *visible* if $\rho(i) = v$, *hidden* if $\rho(i) = h$, and *tied* if $\rho(i) \in \mathscr{N}$. A site is *free* if it is visible or hidden. In the following, we will write $\rho = 1 + \bar{2} + 3^x$ to mean $\rho(1) = v$, $\rho(2) = h$, and $\rho(3) = x$. Denote with $|\rho, x|$ the number of occurrences of x in ρ, that is, $|\rho, x| \triangleq |\{i \in s(\rho) \mid \rho(i) = x\}|$.

The syntax of Bioβ over a given protein signature, consists of *systems*, which can be nested, and *membranes*. Polar proteins can freely float at the level of systems, while apolar proteins can be embedded into membranes. Let $A_t \in \mathscr{P}_t$, for $t \in \{p, ap\}$, $n \in \mathscr{N}$, and $\rho : \mathbb{N} \to \mathscr{N} + \{h, v\}$ be a protein interface.

$$P, Q ::= \diamond \mid A_p(\rho) \mid \wr S \wr P \int \int \mid P * Q \mid \nu n.P \mid \mathsf{p}_n \fatsemi P \mid \mathsf{f}_n \fatsemi \wr S \wr P \int \int \qquad \text{(Systems)}$$

$$S, T ::= \mathbf{0} \mid A_{ap}(\rho) \mid S \star T \mid \mathsf{p}_n^{\perp} \fatsemi S \mid \mathsf{f}_n^{\perp} \qquad \text{(Membranes)}$$

Systems are polar solutions (soluble in aqueous environments), which can be either the empty system $\diamond$, or a polar protein $A_p(\rho)$, or a compartment $\wr S \wr P \int \int$, or a group of systems $P * Q$, or a system prefixed by a *new* name constructor $\nu n.P$. Membranes are apolar solutions (soluble in oily environments), which can be the empty membrane $\mathbf{0}$, or an apolar protein $A_{ap}(\rho)$, or a group of apolar solutions $S \star T$. Notice that membranes are not sequences of actions (as, e.g., in Brane calculus [4]) but simply a collection of proteins. As usual the "new" operator is a binder: in $\nu n.P$, P is the scope of the binder νn. Sharing of names on interfaces represent protein domain-domain bonds. For instance $(x)(A(1^x), B(1 + 2^x)))$ is a protein solution where there is a bond between sites $(A, 1)$ and $(B, 2)$.

Finally, the special constructors *pinch* p and *fuse* f, together with their co-actions $\mathsf{p}^{\perp}, \mathsf{f}^{\perp}$, to represent when a compartment reconfiguration takes place. These actions are inspired by the two basic membrane reconfiguration processes, namely "pinch" and "fuse". A membrane can be pinched (either outward or inward, Figure 1 left), forming a bubble that detaches producing a new compartment; or two membranes can fuse, in either directions (Figure 1 right) mixing contents. Intuitively, $\mathsf{p}_n \fatsemi P$ can be seen as a system P waiting to complete, or *commit*, a pinch interaction with another system featuring the co-action $\mathsf{p}_n^{\perp}$; similarly for f. Thus, these actions are prefixes which freeze the continuation system/membrane until the compartment reconfiguration is completed. The subscripts pair-up corresponding actions and co-actions.

Let us define the *co-action* operator $(\cdot)^{\perp}$ mapping each mobility action to its dual, and such that $((\cdot)^{\perp})^{\perp} = id$; this operator is extended to sets of actions as $S^{\perp} \triangleq \{s^{\perp} \mid s \in S\}$. Moreover, for a set of names $X \subseteq \mathscr{N}$, the *restriction* of a set of actions over X is $S\restriction_X \triangleq \{s_x \in S \mid x \in X\}$. In the following, $P, Q, R, \ldots$ range over systems, $S, T, \ldots$ over membranes and $K, L, H, \ldots$ over both systems or membranes.

The set of free names $fn(P)$ and $fn(S)$ are defined as usual, covering also subscripted names on mobility actions (clearly restriction on names binds also action names). In the following $fn(K_1,\ldots,K_n)$ stands for $\bigcup_{i=1}^{n} fn(K_i)$, for every K_i either a system or a membrane.

Let $\text{ACT} \triangleq \{\mathsf{p}_n, \mathsf{p}_n^{\perp}, \mathsf{f}_n, \mathsf{f}_n^{\perp} \mid n \in \mathscr{N}\}$ be the set of all mobility actions. We define the sets of occurring actions $act(P)$, $act(S) \subseteq \text{ACT}$ on systems and membranes, respectively, as follows:

$$act(\diamond) = act(A_p(\rho)) = \emptyset \qquad act(\wr S \wr P \int\int) = act(S) \cup act(P) \qquad act(P * Q) = act(P) \cup act(Q)$$
$$act(\mathbf{0}) = act(A_{ap}(\rho)) = \emptyset \qquad act(\nu n.P) = act(P) \qquad act(S \star T) = act(S) \cup act(T)$$
$$act(\mathsf{p}_n \fatsemi P) = \{\mathsf{p}_n\} \cup act(P) \quad act(\mathsf{f}_n \fatsemi P) = \{\mathsf{f}_n\} \cup act(P) \quad act(\mathsf{p}_n^{\perp} \fatsemi S) = \{\mathsf{p}_n^{\perp}\} \cup act(S) \quad act(\mathsf{f}_n^{\perp}) = \{\mathsf{f}_n^{\perp}\}$$

Bioβ terms can be rearranged according to a *structural equivalence* $\equiv$, defined to be the least equivalence closed under α-equivalence and satisfying the rules below:

$$\frac{}{P * \diamond \equiv P} \qquad \frac{}{S \star \mathbf{0} \equiv S} \qquad \frac{\mathsf{op} \in \{*, \star\}}{K \,\mathsf{op}\, L \equiv L \,\mathsf{op}\, K} \qquad \frac{\mathsf{op} \in \{*, \star\}}{K \,\mathsf{op}\, (L \,\mathsf{op}\, H) \equiv (K \,\mathsf{op}\, L) \,\mathsf{op}\, H}$$

$$\frac{n \neq m}{\nu n.\nu m.P \equiv \nu m.\nu n.P} \qquad \frac{}{\nu n.\diamond \equiv \diamond} \qquad \frac{n \notin fn(Q)}{\nu n.(P * Q) \equiv (\nu n.P) * Q} \qquad \frac{n \notin fn(S)}{\nu n.\wr S \wr P \int\int \equiv \wr S \wr \nu n.P \int\int}$$

$$\frac{\mathsf{op} \in \{*, \star\} \quad K \equiv L}{K \,\mathsf{op}\, H \equiv L \,\mathsf{op}\, H} \qquad \frac{P \equiv Q \quad S \equiv T}{\wr S \wr P \int\int \equiv \wr T \wr Q \int\int} \qquad \frac{t \in \{\mathsf{p}, \mathsf{p}^{\perp}, \mathsf{f}\} \quad K \equiv L}{t_n \fatsemi K \equiv t_n \fatsemi L} \qquad \frac{t \in \{\mathsf{p}, \mathsf{f}\} \quad n \neq m}{\nu n.(t_m \fatsemi P) \equiv t_m \fatsemi \nu n.P}$$

Well-formedness As for other languages for graph-like structures, the syntax above is too general, since many syntactically correct terms do not have a clear biological meaning. In this subsection give an informal description of the well-formedness conditions that terms must satisfy; in the next subsection we will present a type system enforcing these conditions.

Unsurprisingly, these conditions are similar to those of related models, such as the κ-calculus, since they are motivated by biochemical properties of proteins and membranes.

First, protein bonds can form only between exactly two active sites on protein surfaces. Hence, terms like $(x)(A(1^x))$ or $(x)(A(1^x), A(1^x), B(1^x))$ have to be discarded. This requirement is extended to action names, because we want to pair-up just a couple of actions at once. Furthermore, a pinch action can be connected only to co-pinch, and similarly for fuse and co-fuse[1].

Regarding membranes, we have to ensure the *impermeability* of membrane boundaries: complexations cannot take place through membranes, hence protein bonds cannot cross them. Thus, terms like $A(1^x) * \wr \mathbf{0} \wr B(1^x) \int\int$ have to be ruled out.

Another (more technical) condition is that continuations after mobility actions cannot have other actions. The intuition is that a membrane engaged with a pinch or fuse interaction, has to complete it before initiating another one.

We summarize these conditions in the follwing definition.

Definition 2.1 (Well-formedness). A system P is *well-formed* if the following conditions hold:

Graph-likeness: free names occurs at most twice in P, and every binder ties either 0 or 2 occurrences;
Impermeability: for any occurrence of a compartment $\wr S \wr Q \int\int$ in P, $fn(Q) \subseteq fn(S)$ hold;
Action pairing: for any occurrence of an action t_n ($t \in \{\mathsf{p}, \mathsf{f}\}$), the name n can appear only in the co-action $t_n^{\perp}$.

[1] Actually, it suffice to disallow name sharing between actions and protein interfaces, but strengthening the condition will become useful in proving the subject reduction theorem.

$$(\mathsf{empty})\ \frac{\varepsilon \in \{\mathbf{0}, \diamond\}}{\emptyset;\emptyset \vdash \varepsilon : \emptyset} \qquad \frac{A \in \mathscr{P} \quad \forall x \in fn(\rho).\ |\rho, x| \leq 2}{\{x \in fn(\rho) \mid |\rho, x| = 1\}; \{x \in fn(\rho) \mid |\rho, x| = 2\} \vdash A(\rho) : \emptyset}\ (\mathsf{prot})$$

$$(\mathsf{action})\ \frac{t \in \{\mathsf{p}, \mathsf{p}^{\perp}, \mathsf{f}\} \quad \Gamma_1;\Gamma_2 \vdash K : \emptyset \quad act(K) = \emptyset}{\Gamma_1, x;\Gamma_2 \vdash t_x \mathbin{?} K : \{t_x\}} \qquad \frac{\Gamma_1;\Gamma_2 \vdash P : \tau \quad x \notin \Gamma_1 \quad \tau\restriction_{\{x\}} = \emptyset}{\Gamma_1;\Gamma_2 \setminus \{x\} \vdash \nu x.P : \tau}\ (\nu\text{-}\mathsf{prot})$$

$$(\mathsf{co\text{-}f})\ \frac{}{x;\emptyset \vdash \mathsf{f}_x^{\perp} : \{\mathsf{f}_x^{\perp}\}} \qquad \frac{t \in \{\mathsf{p}, \mathsf{f}\} \quad \Gamma_1;\Gamma_2, x \vdash P : \tau \cup \{t_x, t_x^{\perp}\} \quad \{t_x, t_x^{\perp}\} \cap \tau = \emptyset}{\Gamma_1;\Gamma_2 \vdash \nu x.P : \tau}\ (\nu\text{-}\mathsf{action})$$

$$(\mathsf{par})\ \frac{\mathsf{op} \in \{*, \star\} \quad \Gamma_1, \Gamma;\Gamma_2 \vdash K : \tau \quad \Delta_1, \Gamma;\Delta_2 \vdash L : \sigma \quad \mathsf{ok}_e(\Gamma_1, \Gamma_2, \Delta_1, \Delta_2) \quad \mathsf{ok}_t(\Gamma, \tau, \sigma)}{\Gamma_1, \Delta_1;\Gamma_2, \Delta_2, \Gamma \vdash K \,\mathsf{op}\, L : \tau \cup \sigma} \qquad \frac{\Gamma_1, \Gamma;\Gamma_2 \vdash S : \tau \quad \Gamma;\Delta_2 \vdash P : \sigma \quad \mathsf{ok}_e(\Gamma_1, \Gamma_2, \emptyset, \Delta_2) \quad \mathsf{ok}_t(\Gamma, \tau, \sigma)}{\Gamma_1;\Gamma_2, \Delta_2, \Gamma \vdash (\!|S|\!)\, P\, [\!]\!] : \tau \cup \sigma}\ (\mathsf{cell})$$

Figure 2: Type system for Bioβ terms.

Action prefix: for any occurrence of a prefix $t_n \mathbin{?} Q$ ($t \in \{\mathsf{p}, \mathsf{p}^{\perp}, \mathsf{f}\}$) in P, $act(Q) = \emptyset$ hold.

As expected, well-formedness is preserved by structural equivalence:

Proposition 2.1 (Well-formedness is up to $\equiv$). Let P, Q systems s.t. $P \equiv Q$, P is well-formed if it is Q.

2.2 Type system

In this subsection we present a type system for Bioβ terms, which provides a formal procedure for checking well-formedness. As usual in type theory, terms are typed by judgements w.r.t. some environment.

Definition 2.2 (Judgement). A *type* $\tau \subseteq \text{ACT}$ is a finite set of mobility actions.

A *judgement* for a Bioβ term K (a system or a membrane) is of the form $\Gamma_1;\Gamma_2 \vdash K : \tau$, where the *environment* $\Gamma_1;\Gamma_2$ is formed by two sets of names, such that $\Gamma_1 \cap \Gamma_2 = \emptyset$, each of which is written as a list of the form $x_0, \ldots, x_n$.

The two sets in the environment keep track of the free names in terms, and the number of times each name occurs: in our case, if $n \in \Gamma_1$ it occurs once, if $n \in \Gamma_2$ then it occurs twice. Types keep track of mobility actions/co-actions occurring with free names.

Moreover, let ok_e and ok_t be two predicates defined as follows:

$$\mathsf{ok}_e(\Gamma_1, \Gamma_2, \Delta_1, \Delta_2) \overset{\triangle}{\iff} (\Gamma_1 \cup \Gamma_2) \cap (\Delta_1 \cup \Delta_2) = \emptyset \qquad \mathsf{ok}_t(\Gamma, \tau, \sigma) \overset{\triangle}{\iff} (\tau\restriction_{\Gamma})^{\perp} = \sigma\restriction_{\Gamma}$$

The first predicate checks if two environments share no names. The second predicate checks if the two types τ, σ pair-up correctly, i.e., they create correct actions-coactions pairs using names in Γ.

The Bioβ type system is shown in Figure 2. Intuitively, (empty) types empty systems and membranes, which have no names and no actions. (prot) deals with proteins (possibly) having free names which fit in the environment, and no actions. The rule (co-f) types the singleton membrane containing a co-fuse, which has a name x of rank 1 and the set containing itself as type. (action) is for the remaining actions: this rule checks that the subsystem contains no actions (thus enforcing **action prefix** of Definition 2.1), adds a *fresh* name x for the action, and create a type containing the action itself. The rules (ν-prot) and (ν-action) manage the restriction, in the first case ν binds a name used by proteins, so the rule removes eventually x from names with rank 2 (remind that only names with 0 or 2 occurrences can

be tied); in the second case there is a name x of rank 2 which connects a pair action-co-action, so the rule removes x from Γ_2 as in previous case, and removes the actions from the type (such actions have no free names anymore). Last two rules are for composing systems and membranes. (par) puts side by side two subsystem, checking that the global name rank is no greater then 2 by ok_e (enforcing **graph-likeness**) and that actions pairs with respective co-actions by ok_t (**action pairing**); notice that common names (Γ) are promoted from rank 1 to 2. For (cell) is similar: we build a cell from subsystems and, beside the previous checkings, we impose that all names of rank 1 of P are contained in the ranked 1 names of S, i.e, Γ_2 is empty; this enforces **impermeability**.

Proposition 2.2 (Unicity of type). Let K be a Bioβ system or a membrane. If $\Gamma_1;\Gamma_2 \vdash K : \tau$ and $\Delta_1;\Delta_2 \vdash K : \sigma$, then $\Gamma_1 = \Delta_1$, $\Gamma_2 = \Delta_2$ and $\tau = \sigma$.

Theorem 2.1 (Well-formedness). A Bioβ system P is well-formed if and only if $\Gamma_1;\Gamma_2 \vdash P : \tau$, for some environment Γ_1, Γ_2 and type τ.

Proposition 2.3 (Subject congruence). Let P, Q be two Bioβ systems and S, T two Bioβ membranes, such that $P \equiv Q$ and $S \equiv T$, then $\Gamma_1;\Gamma_2 \vdash P : \tau$ if and only if $\Gamma_1;\Gamma_2 \vdash Q : \tau$ and $\Delta_1;\Delta_2 \vdash S : \sigma$ if and only if $\Delta_1;\Delta_2 \vdash T : \sigma$, for some Γ_1, Γ_2, Δ_1, Δ_2 and τ, σ.

From now on, we consider only well-formed terms, except when it is explicitly said.

2.3 Semantics

The operational semantics of a Bioβ system is given by a reaction relation $\rightarrow$ on systems, defined by means of a rule system parametrized over *basic protein reactions* and *mobility configurations*.

Defining a reaction semantics by means of a set of rules is the simplest way for a biologist to look at protein reactions, since it resembles usual chemical reactions. Moreover, the rule-based modeling leaves biologists a great degree of freedom in defining the behavior of a protein in solution. Such modeling freedom is somewhat limited by biochemical constraints. One of this was individuated and formalized by Danos and Laneve in [8]: the *causality* constraint.

Biochemical reactions are either complexations (i.e. where low energy bonds are formed on two complementary sites) or decomplexations, possibly co-occurring with (de)activation of sites. *Causality* does not allow simultaneous complexations *and* decomplexations on the same site. This constraint is assured by the notion of *(anti-)monotonicity*, which forces reactions not to decrease (increase) the level of connections of a solution [8]. Moreover, protein reactions should be closed on all well-formed membrane contexts, so that we are able to model reactions for proteins located in different membranes and cells.

Let us now discuss about membrane level interactions, that is, rearrangement of the membrane structure. As observed by biologists, these interactions result from complex nets of signaling pathways which induce a mechanical reshaping of biological membranes, allowing for example the formation of new vesicles, and hence new compartments. However, as noticed before, actual membrane transformations are limited and can be either a fuse or a pinch [6]; what changes from situations to situations are the proteins involved in the signaling pathway, leading to the actual fuse or pinch. In order to formalize this situation we will distinguish between the effective membrane rearrangement, and what has trigger it. More precisely, a membrane reaction is split into two steps: a "preparation phase" where mobility actions (p or f) are introduced as prefixes in the system, and a "commitment phase" where the structural reconfiguration of nested compartments is actually performed. The commitment phase is formalized by rules defined once and for all by the framework, corresponding to the pinch and fuse processes of Figure 1, but the preparation phase is specified by the modeler. In other words, one has only to describe *when* these interaction take place by indicating which are the suitable protein configurations that trigger a pinch or a fuse. In this way, each membrane interaction is given an explanation at the protein-level.

2.3.1 Protein reactions across multiple localities

When dealing with protein reactions it is useful to think at complexes as single entities which can be affected in all their protein sub-units. An example of protein reaction we want to model is the following:

$$D(1)*A(1+2^x)*\wr B(1^x+2^y)\wr C(1^y+\bar{2})\int\int \to \nu z.\big(D(1^z)*A(1^z+2^x)*\wr B(1^x+2^y)\wr C(1^y+2)\int\int\big)$$

A complexation reaction between protein D and the transmembrane ABC-complex has effect not only in the formation of a new protein bond between sites $(D,1)$ and $(A,1)$, but it toggles from hide to visible site $(C,2)$, which is not local to protein D (promotor of the reaction). This is a very common situation in protein signaling transduction, indeed transmembrane receptors (as ABC-complex) propagate extracellular signals (as protein D) into the intracellular environment according to this mechanism.

Let us now consider the following reaction:

$$C(1^y+\bar{2})*\wr B(1^x+2^y)\wr A(1+2^x)*D(1)\int\int \to \nu z.\big(C(1^y+2)*\wr B(1^x+2^y)\wr A(1^z+2^x)*D(1^z)\int\int\big)$$

This reaction differs from the previous one only for the direction in which the "signal" is propagated. There is no biological reason to distinguish between these two forms of reactions, indeed they are exactly the same protein reaction. In this sense, membranes have to be taken into consideration only in the way they partition solutions into distinct locations.

In order to define protein interactions in nested systems we introduce *(linear) contexts*. Contexts generalize the Bioβ syntax introducing variables, which are sorted into two categories, systems variables X and membrane variables Y, used as "holes" for systems and membrane, respectively. Sequences $K_1,\ldots,K_n$ are denoted by $\vec{K}$, and concatenation by $\vec{K},\vec{K'}=K_1,\ldots,K_n,K'_1,\ldots,K'_m$.

$$\begin{aligned} C[\vec{X};\vec{Y}],D[\vec{X};\vec{Y}] ::= {} & \diamond \mid A_p(\rho) \mid X \mid \mathsf{p}_n\wr P \mid \mathsf{f}_n\wr\wr S\wr P\int\int \mid \\ & C[\vec{X'};\vec{Y'}]*D[\vec{X''};\vec{Y''}] \mid \nu n.C[\vec{X};\vec{Y}] \mid \wr E[\vec{Y'}]\wr C[\vec{X};\vec{Y''}]\int\int && \text{(System contexts)} \\ E[\vec{Y}],F[\vec{Y}] ::= {} & \mathbf{0} \mid A_{ap}(\rho) \mid \mathsf{p}_n^{\perp}\wr S \mid \mathsf{f}_n^{\perp} \mid E[\vec{Y'}]\star F[\vec{Y''}] && \text{(Membrane contexts)} \end{aligned}$$

where $X\in\vec{X}$, $Y\in\vec{Y}$, and let π,π' denote permutations, $\vec{X}=\pi(\vec{X'},\vec{X''})$, $\vec{Y}=\pi'(\vec{Y'},\vec{Y''})$. Context application $C[\vec{P};\vec{S}]$ is defined as expected, for $\vec{P}$ and $\vec{S}$ two sequences of systems and membranes, respectively.

As in the case of the κ-calculus, protein reactions can be of two kinds: monotone and anti-monotone. Monotonicity is defined on a growing relation ($\rhd$) over Bioβ protein solutions. Differently from κ-calculus, the notion of "protein solution" in the Bioβ Framework have to take into account system and membrane proteins, and the fact that they are in different locations. Thus, we define protein solutions as sequences of system groups and membrane groups of proteins, as follows:

Definition 2.3 (Wide protein solutions). A *group of system proteins* is a Bioβ system P of the form $A_1(\rho_1)*\ldots*A_n(\rho_n)$, where $A_i\in\mathscr{P}_p$ $(1\le i\le n)$; similarly, a *group of membrane proteins* is a Bioβ membrane S of the form $B_1(\sigma_1)\star\ldots\star B_m(\sigma_m)$, where $B_j\in\mathscr{P}_{ap}$ $(1\le j\le m)$.

A *wide protein solution* is a couple of sequences of system groups and membrane groups of proteins, which can be restricted on some names $\vec{x}$, denoted as $\nu\vec{x}.\langle P_1,\ldots,P_k \mid S_1,\ldots,S_h\rangle$.

Informally, a group of proteins is a *local* protein solution, that is, all proteins in that group reside in the same location. Wide protein solutions are non-local and are formed as sequences of protein groups of two kinds according with the syntactic separation of membranes and systems. Actually, wide protein solutions are not Bioβ terms, but we use them as specifications of how complexes are formed abstracting on localities. Often, we will shorten $\nu\vec{x}.\langle\vec{P}\mid\vec{S}\rangle$ as $\langle\vec{P}\mid\vec{S}\rangle$ when $|\vec{x}|=0$. Free names are defined as $fn(\nu x_1,\ldots x_n.\langle\vec{P}\mid\vec{S}\rangle)=fn(\vec{P},\vec{S})\setminus\{x_1,\ldots x_n\}$.

We are interested in connected wide protein solutions, that is, wide solutions where exists a path of bonds linking all the proteins in the solution. Closed connected wide protein solutions are actually *protein complexes*. Formally:

Definition 2.4 (Connectedness). The *connected* wide protein solutions are defined inductively as:

$$\frac{}{\langle A(\rho) \mid \rangle} \qquad \frac{}{\langle \mid B(\sigma) \rangle} \qquad \frac{\nu\vec{x}.\langle \vec{P} \mid \vec{S} \rangle}{\nu x\vec{x}.\langle \vec{P} \mid \vec{S} \rangle} \qquad \frac{\langle \vec{P} \mid \vec{S} \rangle \qquad \langle \vec{Q} \mid \vec{T} \rangle \quad fn(\langle \vec{P} \mid \vec{S} \rangle) \cap fn(\langle \vec{Q} \mid \vec{T} \rangle) \neq \emptyset}{\langle \vec{P}, \vec{Q} \mid \vec{S}, \vec{T} \rangle}$$

$$\frac{\langle \vec{P}_1, P, \vec{P}_2 \mid \vec{S} \rangle \qquad \langle Q \mid \rangle \quad fn(\langle \vec{P}_1, P, \vec{P}_2 \mid \vec{S} \rangle) \cap fn(\langle Q \mid \rangle) \neq \emptyset}{\langle \vec{P}_1, P * Q, \vec{P}_2 \mid \vec{S} \rangle} \qquad \frac{\langle \vec{P} \mid \vec{S}_1, S, \vec{S}_2 \rangle \qquad \langle \mid T \rangle \quad fn(\langle \vec{P} \mid \vec{S}_1, S, \vec{S}_2 \rangle) \cap fn(\langle \mid T \rangle) \neq \emptyset}{\langle \vec{P} \mid \vec{S}_1, S * T, \vec{S}_2 \rangle}$$

A wide protein solution $\nu\vec{x}.\langle \vec{P} \mid \vec{S} \rangle$ is a *wide complex* if it is connected and $fn(\nu\vec{x}.\langle \vec{P} \mid \vec{S} \rangle) = \emptyset$.

Definition 2.5 (Growing relation $\rhd$). Let $\tilde{x}$ be a set of fresh names. A growing relation $\rhd$ over interfaces is defined inductively as follows:

$$\frac{}{\tilde{x} \vdash \bar{i} \rhd i} \qquad \frac{}{\tilde{x} \vdash i \rhd \bar{i}} \qquad \frac{x \in \tilde{x}}{\tilde{x} \vdash i \rhd i^x} \qquad \frac{\tilde{x} \cap fn(\rho) = \emptyset}{\tilde{x} \vdash \rho \rhd \rho} \qquad \frac{\tilde{x} \vdash \rho \rhd \sigma \quad \tilde{x} \vdash \rho' \rhd \sigma'}{\tilde{x} \vdash \rho + \rho' \rhd \sigma + \sigma'}$$

A growing relation $\rhd$ over wide protein solutions is defined inductively as follows:

$$\frac{}{\vec{x} \vdash \langle \diamond \mid \rangle \rhd \langle \diamond \mid \rangle} \qquad \frac{}{\vec{x} \vdash \langle \mid \mathbf{0} \rangle \rhd \langle \mid \mathbf{0} \rangle} \qquad \frac{\vec{x} \vdash \langle \vec{P} \mid \vec{S} \rangle \rhd \langle \vec{P}' \mid \vec{S}' \rangle \qquad \vec{x} \vdash \langle \vec{Q} \mid \vec{T} \rangle \rhd \langle \vec{Q}' \mid \vec{T}' \rangle}{\vec{x} \vdash \langle \vec{P}, \vec{Q} \mid \vec{S}, \vec{T} \rangle \rhd \langle \vec{P}', \vec{Q}' \mid \vec{S}', \vec{T}' \rangle}$$

$$\frac{\vec{x} \vdash \langle \vec{P}, P \mid \vec{S} \rangle \rhd \langle \vec{Q}, Q \mid \vec{T} \rangle \qquad \vec{x} \vdash \rho \rhd \sigma}{\vec{x} \vdash \langle \vec{P}, P * A(\rho) \mid \vec{S} \rangle \rhd \langle \vec{Q}, Q * A(\sigma) \mid \vec{T} \rangle} \qquad \frac{\vec{x} \vdash \langle \vec{P} \mid \vec{S}, S \rangle \rhd \langle \vec{Q} \mid \vec{T}, T \rangle \qquad \vec{x} \vdash \rho \rhd \sigma}{\vec{x} \vdash \langle \vec{P} \mid \vec{S}, S \star B(\rho) \rangle \rhd \langle \vec{Q} \mid \vec{T}, T \star B(\sigma) \rangle}$$

$$\frac{\vec{x} \vdash \langle \vec{P} \mid \vec{S} \rangle \rhd \langle \vec{Q}, Q \mid \vec{T} \rangle \qquad fn(\rho) \subseteq \vec{x}}{\vec{x} \vdash \langle \vec{P} \mid \vec{S} \rangle \rhd \langle \vec{Q}, Q * A(\rho) \mid \vec{T} \rangle} \qquad \frac{\vec{x} \vdash \langle \vec{P} \mid \vec{S} \rangle \rhd \langle \vec{Q} \mid \vec{T}, T \rangle \qquad fn(\sigma) \subseteq \vec{x}}{\vec{x} \vdash \langle \vec{P} \mid \vec{S} \rangle \rhd \langle \vec{Q} \mid \vec{T}, T \star B(\sigma) \rangle}$$

Here, growing means possibly creating "new" bonds and/or proteins, and toggling sites from visible to hide or vice versa. Note that new bonds use names from $\vec{x}$, that is, the following proposition holds:

Proposition 2.4. If $\vec{x} \vdash \langle \vec{P} \mid \vec{S} \rangle \rhd \langle \vec{Q} \mid \vec{T} \rangle$, then $fn(\langle \vec{Q} \mid \vec{T} \rangle) = fn(\langle \vec{P} \mid \vec{S} \rangle) \cup \vec{x}$ and $\vec{x} \cap fn(\langle \vec{P} \mid \vec{S} \rangle) = \emptyset$.

On the concept of growing we define monotonicity, which is very important to guarantee that the protein reactions respect the principle of *causality*.

Definition 2.6 (Monotonicity and Anti-monotonicity). Let $\langle \vec{P} \mid \vec{S} \rangle$ and $\langle \vec{Q} \mid \vec{T} \rangle$ two wide protein solutions, we say that the pair $(\langle \vec{P} \mid \vec{S} \rangle, \nu\vec{x}.\langle \vec{Q} \mid \vec{T} \rangle)$ is *monotone* if $\vec{x} \vdash \langle \vec{P} \mid \vec{S} \rangle \rhd \langle \vec{Q} \mid \vec{T} \rangle$ and $\nu\vec{x}.\langle \vec{Q} \mid \vec{T} \rangle$ is connected. $(\nu\vec{x}.\langle \vec{Q} \mid \vec{T} \rangle, \langle \vec{P} \mid \vec{S} \rangle)$ is *anti-monotone* if $(\langle \vec{P} \mid \vec{S} \rangle, \nu\vec{x}.\langle \vec{Q} \mid \vec{T} \rangle)$ is monotone.

We can now define what is a *protein reaction specification*, and the induced protein reactions.

Definition 2.7 (Protein reaction specifications). A *protein reaction specification* is a set $\mathscr{R}$ of monotone and/or anti-monotone pairs of wide protein solutions.

Given a protein reaction specification $\mathscr{R}$, the corresponding *protein reactions* are defined as follows:

$$\frac{(\langle \vec{P} \mid \vec{S} \rangle, \nu\vec{x}.\langle \vec{Q} \mid \vec{T} \rangle) \in \mathscr{R}}{C[\vec{P}; \vec{S}] \to \nu\vec{x}.C[\vec{Q}; \vec{T}]}\ (\text{mon}) \qquad \frac{(\nu\vec{x}.\langle \vec{Q} \mid \vec{T} \rangle, \langle \vec{P} \mid \vec{S} \rangle) \in \mathscr{R}}{\nu\vec{x}.C[\vec{Q}; \vec{T}] \to C[\vec{P}; \vec{S}]}\ (\text{anti-mon})$$

where $C[\vec{X}; \vec{Y}]$ is a context such that $\nu\vec{x}.C[\vec{Q}; \vec{T}]$ is well-formed (i.e., well-typed).

In virtue of the generality of context $C[\vec{X}; \vec{Y}]$, we can adopt a simplified notation for reactions: $\langle \vec{P} \mid \vec{S} \rangle \to \nu\vec{x}.\langle \vec{Q} \mid \vec{T} \rangle$ for monotone reactions and, similarly, $\nu\vec{x}.\langle \vec{Q} \mid \vec{T} \rangle \to \langle \vec{P} \mid \vec{S} \rangle$ for anti-monotone ones.

2.3.2 Membrane transport

Membrane transformations are sorted into two groups: pinch (p) and fuse (f). They occur only with the interaction of an action t_n and a co-action $t_n^\perp$ ($t \in \{\mathsf{p},\mathsf{f}\}$). Pinch and fuse can be applied with different orientations, depending on where actions constructs are placed. Formally, the *commitment rule* are

$$\mathsf{p}_n \fatsemi P * \wr \mathsf{p}_n^\perp \fatsemi S \star T \wr Q \int\int \to \wr T \wr \wr S \wr P \int\int * Q \int\int \qquad \text{(pinch-in)}$$

$$\wr \mathsf{p}_n^\perp \fatsemi S \star T \wr \mathsf{p}_n \fatsemi P * Q \int\int \to \wr S \wr P \int\int * \wr T \wr Q \int\int \qquad \text{(pinch-out)}$$

$$\mathsf{f}_n \fatsemi \wr S \wr P \int\int * \wr \mathsf{f}_n^\perp \star T \wr Q \int\int \to \wr S \star T \wr P * Q \int\int \qquad \text{(fuse-hor)}$$

$$\wr \mathsf{f}_n^\perp \star T \wr \mathsf{f}_n \fatsemi \wr S \wr P \int\int * Q \int\int \to P * \wr S \star T \wr Q \int\int \qquad \text{(fuse-ver)}$$

We can pinch *inward* or *outward* a sub-system depending on the direction where p and $\mathsf{p}^\perp$ are placed relatively to each other. Note that the continuations of p and $\mathsf{p}^\perp$ are used to form the new compartment, either in (pinch-in) or (pinch-out). Fusions can be either *horizontal* (fuse-hor) or *vertical* (fuse-ver) according to the relative positions of compartments to be merged. Notice that, even in these reactions, f and $\mathsf{f}^\perp$ continuations are used to select which part of the system has to be merged.

Mobility reactions respect *bitonality* in the sense of [4, 6], in fact the parity of nesting of P and Q is preserved in all these reactions, hence they preserve the bitonal coloring of those subsystems.

In nature, membrane transport occurs when a certain protein configuration, allowing the mechanical reshaping of the membrane, is reached. Thus, in order to specify when a pinch or a fuse takes place, it is sufficient to specify which are the protein configurations leading to the "appearing" of mobility controls.

Definition 2.8 (Membrane reaction specifications). A *pinching configuration* is a 5-tuple $(P,P',S',S',Q) \in \Pi \times \Pi \times \Sigma \times \Sigma \times \Pi$, where Π, Σ are the sets of action-free systems and action-free membranes, respectively. A *fusing configurations* is 5-tuple $(P,S,R,T,Q) \in \Pi \times \Sigma \times \Pi \times \Sigma \times \Pi$.

A *membrane reaction specification* is a pair $(\mathscr{R}_\mathsf{p},\mathscr{R}_\mathsf{f})$ where $\mathscr{R}_\mathsf{p}$ is a set of pinching configurations, and $\mathscr{R}_\mathsf{f}$ is a set of fusing configurations Given a membrane reaction specification $(\mathscr{R}_\mathsf{p},\mathscr{R}_\mathsf{f})$, the corresponding *mobility introduction reactions* are defined as follows:

$$\frac{(P,P',S,S',Q) \in \mathscr{R}_\mathsf{p} \qquad fn(P) \subseteq fn(S) \qquad fn(S) \subseteq fn(P,Q) \qquad n \text{ fresh}}{P * P' * \wr S \star S' \star T \wr Q * R \int\int \to \nu n.(\mathsf{p}_n \fatsemi P * P' * \wr \mathsf{p}_n^\perp \fatsemi S \star S' \star T \wr Q * R \int\int)} \qquad \text{(intro-p-in)}$$

$$\frac{(P,P',S,S',Q) \in \mathscr{R}_\mathsf{p} \qquad fn(P) \subseteq fn(S) \qquad fn(S) \subseteq fn(P,Q) \qquad n \text{ fresh}}{Q * \wr S \star S' \star T \wr P * P' * R \int\int \to \nu n.(Q * \wr \mathsf{p}_n^\perp \fatsemi S \star S' \star T \wr \mathsf{p}_n \fatsemi P * P' * R \int\int)} \qquad \text{(intro-p-out)}$$

$$\frac{(P,S,R,T,Q) \in \mathscr{R}_\mathsf{f} \qquad n \text{ fresh}}{\wr S \star S' \wr P * P' \int\int * R * \wr T \star T' \wr Q * Q' \int\int \to \nu n.(\mathsf{f}_n \fatsemi \wr S \star S' \wr P * P' \int\int * R * \wr \mathsf{f}^\perp \star T \star T' \wr Q * Q' \int\int)} \qquad \text{(intro-f-hor)}$$

$$\frac{(P,S,R,T,Q) \in \mathscr{R}_\mathsf{f} \qquad n \text{ fresh}}{\wr T \star T' \wr R * \wr S \star S' \wr P * P' \int\int * Q * Q' \int\int \to \nu n.(\wr \mathsf{f}_n^\perp \star T \star T' \wr R * \mathsf{f}_n \fatsemi \wr S \star S' \wr P * P' \int\int * Q * Q' \int\int)} \qquad \text{(intro-f-ver)}$$

Intuitively, the 5-tuples in $\mathscr{R}_\mathsf{p}$ and $\mathscr{R}_\mathsf{f}$ describe the systems configurations which trigger a mobility action. Notice that in the two pinch introduction rules, the first two pairs of terms (P,P') and (S,S') in the tuple correspond to two splits of sub-terms in the left hand side of the reaction rule: this is for distinguishing the part that is actually pinched (P and S) from the part that just helps triggering reaction but do not change its position (P' and S', which may be enzymes). The same holds for fuse introduction

rules, where P and S will be fused with the compartment where T and Q occur, whereas R just helps the fusion but will not change its position. In practice, actions are used for "freezing" the state of the subsystem involved in the reconfiguration, until the reconfiguration actually takes place.

Informally, $\mathscr{R}_{\mathsf{p}}$ and $\mathscr{R}_{\mathsf{f}}$ provide a template of suitable conditions, cause of which membrane reconfiguration can happen. Such conditions are to be defined by the modeler, who knows which are the requisites to make possible a pinch (either for vesiculation or endocytosis) or a fusion (either for fusion or endocytosis). This way of describing a model in the Bioβ framework gives biologists (in general, modelers) the possibility to justify membrane reconfigurations at the protein-level, providing only the essential characteristics that a sub-system must satisfy to perform a mobility reaction.

Summing up, to define a Bioβ model, users have to provide the protein signature, the set of protein rules $\mathscr{R}$, ignoring compartments, and two sets of configurations $\mathscr{R}_{\mathsf{p}}$ and $\mathscr{R}_{\mathsf{f}}$ describing the situations for which a mobility reaction can occur. All other reactions are automatically provided by the framework. Hence, modelers can focus on protein-level aspects, leaving the framework to deal with biological compartments and membrane transport.

Definition 2.9 (Bioβ specification). A Bioβ specification is a quadruple $(\mathscr{P},\mathscr{R},\mathscr{R}_{\mathsf{p}},\mathscr{R}_{\mathsf{f}})$ where $\mathscr{P}$ is a protein signature, $\mathscr{R}$ is a protein reaction specication, and $(\mathscr{R}_{\mathsf{p}},\mathscr{R}_{\mathsf{f}})$ is a membrane reaction specication.

Definition 2.10 (Bioβ reactive system). Let $(\mathscr{P},\mathscr{R},\mathscr{R}_{\mathsf{p}},\mathscr{R}_{\mathsf{f}})$ be a specification. The associated *Bioβ reactive system* is a pair $(\Pi^{act},\rightarrow)$, where Π^{act} is the set of well-formed Bioβ systems over $\mathscr{P}$, and $\rightarrow$, called the *reaction relation*, is the least binary relation over Π^{act} containing all protein reactions over $\mathscr{R}$, mobility reactions pinch-in, pinch-out, fuse-hor, fuse-ver, and all introduction reactions over $\mathscr{R}_{\mathsf{p}}$ and $\mathscr{R}_{\mathsf{f}}$, and it is closed under structural equivalence, system composition, restriction of names, and compartment nesting.

Closure of $\rightarrow$ under system composition, restriction and compartment constructs, allow to focus on the actual "reacting parts" of the system. Clearly, $\rightarrow$ is not closed under action prefixes, because action prefixes "freeze" their continuation so reactions cannot happen in them.

We can prove that for Bioβ reactive systems subject reduction holds, that is, well-formedness for systems is preserved by reaction steps.

Proposition 2.5 (Subject reduction). Let P, Q be two Bioβ systems. If $\Gamma_1;\Gamma_2 \vdash P : \tau$ and $P \rightarrow Q$, then $\Gamma_1;\Delta_2 \vdash Q : \sigma$ where either $\Gamma_2 = \Delta_2$ and $\tau = \sigma$, or $\Gamma_2 = \Delta_2, n$ and $\tau = \sigma + \{t_n, t_n^{\perp}\}$ for some $t \in \{\mathsf{p},\mathsf{f}\}$.

3 An example: Membrane Traffic

In this section we give a formal Bioβ description of how membrane traffic works. Here we take two important processes into account, which are fundamental in many transport mechanisms in cells: vesicle formation and vesicle docking. Vesicles form by budding from membranes. Each bud has distinctive coat protein on cytosol surface, which shapes the membrane to form a bubble. The bud captures the correct molecules for outward transport by a selective bond with a cargo receptor attached through the double layer to the coat protein complex, which is lost after the budding completes (see Figure 3(a)). Vesicle contents are released in the target membrane by fusing vesicle with it.

To ensure that membrane traffic proceeds in an orderly way, transport vesicles must be highly selective in recognizing the correct target membrane with which to fuse. Because of the diversity of membrane systems, a vesicle is likely to encounter many potential target membranes before it finds the correct one. Specificity in targeting is ensured displaying protein surface markers that identify them according to their origin and type of cargo (*SNAREs*). Vesicles are marked by v-SNAREs (vector SNAREs) which

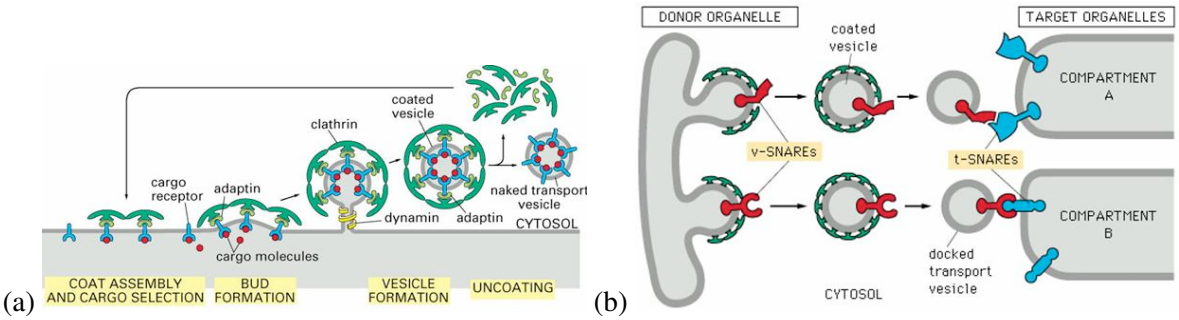

Figure 3: (a) Vesicle formation, (b) Vesicle docking. (Pictures taken from [2])

selectively bind with some corresponding t-SNAREs (target SNAREs) located on target membrane (see Figure 3(b)). SNARE proteins have a central role both in providing specificity and in catalyzing the fusion of vesicles with the target membrane.

Let $\mathscr{P}_p = \{R_i^e, R_i^c, C_i, Ad_i, Cl, Sn_i^e, Sn_i^c, tSn_i^e, tSn_i^c\}$ and $\mathscr{P}_{ap} = \{R_i^m, Sn_i^m, tSn_i^m\}$, for $i \in \{0,1\}$. Trans-membrane proteins are represented by three sub-units in order to specify the direction they exhibit with respect to membranes into which are embedded. We denote with $R_i^c(1+2)$, $R_i^m(1+2)$, $R_i^e(1+2)$ the three cargo receptor sub-units, respectively for cytosolic, trans-membrane, and extra-cytosolic; SNAREs (Sn_i^* and tSn_i^*) are defined in a similar way. $C_i(1)$ denotes a cargo molecule, $Ad_i(1+2)$ adaptin, and $Cl(1)$ clathrin. Proteins are sub-scripted with an index $i \in \{0,1\}$ to distinguish different families of proteins.

We describe the Bioβ model for membrane traffic giving the set of reaction rules for proteins and introduction of mobility. Protein reactions are defined as follows:

$$\langle C_i(1) * R_i^e(1+2^x), R_i^c(1^y+\bar{2}) \mid R_i^m(1^x+2^y)\rangle \xrightarrow{\mathsf{rec}} \nu z.\langle C_i(1^z) * R_i^e(1^z+2^x), R_i^c(1^y+2) \mid R_i^m(1^x+2^y)\rangle$$

$$\langle R_i^c(1^x+2) * Ad_i(1+\bar{2}) \mid\rangle \xrightarrow{\mathsf{adpt}} \nu y.\langle R_i^c(1^x+2^y) * Ad_i(1^y+2) \mid\rangle$$

$$\langle Ad_i(1^x+2) * Cl(1) \mid\rangle \xrightarrow{\mathsf{coat}} \nu y.\langle Ad_i(1^x+2^y) * Cl(1^y) \mid\rangle$$

$$\nu y.\langle R_i^c(1^x+2^y) * Ad_i(1^y+2^z) * Cl(1^z) \mid\rangle \xrightarrow{\mathsf{uncoat}} \langle R_i^c(1^x+2) * Ad_i(1+2^z) * Cl(1^z) \mid\rangle$$

$$\langle Sn_i^c(1^x+2) * tSn_i^c(1+2^z) \mid\rangle \xrightarrow{\mathsf{snare}} \nu y.\langle Sn_i^c(1^x+2^y) * tSn_i^c(1^y+2^z) \mid\rangle$$

Intuitively, (rec) describes the formation of a bond between the cargo molecule and the corresponding cargo receptor, which toggles to visible the hidden site of the receptor cytosolic sub-unit. (adpt) and (coat) allow for the creation of the clathrin coat which leads to the formation of the bud. (uncoat) frees the vesicle from the adaptin-clathrin coat. Finally, (snare) deals with the snare-mediated docking of the naked vesicle, by forming a bond between the v-SNARE and t-SNARE of the same family. Figure 3 gives an informal idea of how protein reactions work.

Introduction rules are defined by the two sets of pinching and fusing configurations: $\mathscr{R}_\mathsf{p}$ and $\mathscr{R}_\mathsf{f}$.

$$\mathscr{R}_\mathsf{p} = \{(P, P', S, S', Q)\} \quad \text{where} \quad \begin{array}{ll} P = C_i(1^x) * R_i^e(1^x+2^y) * Sn_i^e(1^z) & P' = \diamond \\ S = R_i^m(1^y+2^w) \star Sn_i^m(1^z+2^u) & S' = \mathbf{0} \\ Q = R_i^c(1^w+2^a) * Ad_i(1^a+2^b) * Cl(1^b) * Sn_i^c(1^u+2) & \end{array}$$

The configuration (P, P', S, S', Q) allows the vesicle to detach when the adaptin-clathrin coat completely cover the forming bud. For sake of simplicity we consider coats of size one, i.e. with only one adaptin-clathrin complex (generalization to coat of size n is straightforward).

$$\mathscr{R}_{\mathsf{f}} = \{(P,S,R,T,Q)\} \quad \text{where} \quad \begin{array}{lll} P = C_i(1^x) * R_i^e(1^x + 2^y) * Sn_i^e(1^a) & & T = tSn_i^m(1^d + 2^e) \\ S = R_i^m(1^y + 2^z) \star Sn_i^m(1^a + 2^b) & & Q = tSn_i^e(1^e) \\ R = R_i^c(1^z + 2) * Sn_i^c(1^b + 2^c) * tSn_i^c(1^c + 2^d) & & \end{array}$$

The configuration (P,S,R,T,Q) describe the situation when the vesicle is docked to the target membrane, hence the fusion can take place.

One possible initial configuration can be defined as follows:

$$\begin{aligned} Main &= DOrg * Targ_0 * Targ_1 \\ DOrg &= Cl(1) * Cl(1) * Ad_0(1+\bar{2}) * Ad_1(1+\bar{2}) * R_0^c(1^{y_0}+\bar{2}) * R_1^c(1^{y_1}+\bar{2}) * Sn_0^c(1^{b_0}+2) * Sn_1^c(1^{b_1}+2) * \\ &\quad \wr R_0^m(1^{x_0}+2^{y_0}) \star R_1^m(1^{x_0}+2^{y_1}) \star Sn_0^m(1^{a_0}+2^{b_0}) \star Sn_1^m(1^{a_1}+2^{b_1}) \\ &\qquad \wr R_0^e(1+2^{x_0}) * R_1^e(1+2^{x_1}) * Sn_0^e(1^{a_0}) * Sn_1^e(1^{a_1}) * C_0(1) * C_1(1) \int \int \\ Targ_0 &= tSn_0^c(1+2^{d_0}) * \wr tSn_0^m(1^{d_0}+2^{e_0}) \wr tSn_0^e(1^{e_0}) \int \int \\ Targ_1 &= tSn_1^c(1+2^{d_0}) * \wr tSn_1^m(1^{d_1}+2^{e_1}) \wr tSn_1^e(1^{e_1}) \int \int \end{aligned}$$

In this case we have a single donor organelle which is able to pinch out two vesicles (using two different type of SNAREs). Such vesicle can approach the corresponding docking membrane. In Figure 3(b) is depicted a possible run of computation.

4 Comparing with κ-calculus

In this section we give a formal connection between Bioβ and Danos and Laneve's κ-calculus [8]. This calculus has been taken as the formal model protein interactions in the definition of biobigraphs, the bigraphical model underlying the Bioβ Framework.

The syntax of κ-solutions is the following:

$$S,T ::= \mathbf{0} \mid A(\rho) \mid S,T \mid (x)(S)$$

A κ-solution can be either the empty solution $\mathbf{0}$, or a protein $A(\rho)$, or a group of solutions S,T, or a solution prefixed by name restriction $(x)(S)$ with $x \in \mathscr{N}$.

As usual in process calculi, a structural equivalence ($\equiv$) over processes is introduced, which is the least structural containing α-equivalence and satisfying the abelian monoid on composition and the scope extension and extrusion laws. Moreover we define the sets of free edge names $fn(\rho)$ and $fn(S)$, on interfaces and solutions, defined in the common way, where the new name constructor $(x)(S)$ is the only name binder. We say a κ-solution S is *closed* iff $fn(S) = \emptyset$.

Connectedness of a κ-solutions is particular a simper case of the one seen before.

Definition 4.1 (Connectedness and complexes). *Connected κ-solutions* are defined inductively as:

$$\frac{}{A(\rho)} \qquad \frac{S}{(x)(S)} \qquad \frac{S \quad T \quad fn(S) \cap fn(T) \neq \emptyset}{S,T} \qquad \frac{S \quad S \equiv T}{T}$$

A κ-solution S is a *complex* if it is closed, graph-like, and connected.

The semantic of κ-calculus is defined by means of a transition system built up by a user-defined set of monotone and anti-monotone rules $\mathscr{R}_\kappa$.

Definition 4.2 (Growing relation). A growing relation $\rhd$ over κ-solutions is defined inductively as:

$$\frac{}{\tilde{x} \vdash \mathbf{0} \rhd \mathbf{0}} \qquad \frac{\tilde{x} \vdash S \rhd T \quad \tilde{x} \vdash \rho \rhd \sigma}{\tilde{x} \vdash S, A(\rho) \rhd T, A(\sigma)} \qquad \frac{\tilde{x} \vdash S \rhd T \quad fn(\sigma) \subseteq \tilde{x}}{\tilde{x} \vdash S \rhd T, A(\sigma)}$$

Again, monotonicity is a simplification of the one shown for Bioβ.

Definition 4.3 (Monotone reactions). Let L, R be two κ-solutions. $L \to (\tilde{x})R$ is a *monotone reaction* if $\tilde{x} \vdash L \rhd R$, where $L, (\tilde{x})R$ graph-like and R connected. $(\tilde{x})L \to R$ is *antimonotone* if $R \to (\tilde{x})L$ is monotone.

Finally, we are able to characterize a protein transition system:

Definition 4.4 (Protein transition systems). Given a set $\mathscr{R}_\kappa$ of monotone and antimonotone reactions, we define a *protein transition system* (PTS) as a pair $\langle \mathscr{S}, \rightsquigarrow \rangle$, such that $\mathscr{S}$ is a set of κ-solutions, and $\rightsquigarrow$ is the least binary relation over $\mathscr{S}$ such that $\mathscr{R}_\kappa \subseteq \rightsquigarrow$, closed w.r.t. $\equiv$, composition and name restriction.

Now we are able to establish formally a connection between the Bioβ framework and the κ-calculus, that is all protein reactions that can be take place in our systems are justifiably by means of protein reactions in a κ-solution. In order to do this, first we show a way to translate Bioβ systems and reactions into κ solutions and reactions ($[\![-]\!]$). $[\![-]\!]$ is defined as follows:

$$\begin{array}{cccc} [\![\diamond]\!] = \mathbf{0} & [\![A_p(\rho)]\!] = A_p(\rho) & [\![P * Q]\!] = [\![P]\!], [\![Q]\!] & [\![\wr S \wr P \int \int]\!] = [\![S]\!], [\![P]\!] \\ [\![\mathbf{0}]\!] = \mathbf{0} & [\![A_{ap}(\rho)]\!] = A_{ap}(\rho) & [\![S \star T]\!] = [\![S]\!], [\![T]\!] & [\![\nu n.P]\!] = (n)([\![P]\!]) \\ [\![\mathsf{p}_n \wr P]\!] = [\![P]\!] & [\![\mathsf{f}_n \wr P]\!] = [\![P]\!] & [\![\mathsf{p}_n^\perp \wr S]\!] = [\![S]\!] & [\![\mathsf{f}_n^\perp]\!] = \mathbf{0} \end{array}$$

Such encoding induces a type system over the κ-solution for checking if the solutions are graph-like. It can be given as follows:

$$(\mathsf{zero})\ \frac{}{\emptyset;\emptyset \vdash \mathbf{0}} \qquad \frac{A \in \mathscr{P} \quad \forall x \in fn(\rho).|\rho, x| < 2}{\{x \in fn(\rho) \mid |\rho, x| = 1\}; \{x \in fn(\rho) \mid |\rho, x| = 2\} \vdash A(\rho)}\ (\mathsf{prot})$$

$$(\mathsf{res})\ \frac{\Gamma_1; \Gamma_2 \vdash S \quad x \notin \Gamma_1}{\Gamma_1; \Gamma_2 \setminus \{x\} \vdash (x)S} \qquad \frac{\Gamma_1, \Gamma; \Gamma_2 \vdash S \quad \Delta_1, \Gamma; \Delta_2 \vdash T \quad (\Gamma_1 \cup \Gamma_2) \cap (\Delta_1 \cup \Delta_2) = \emptyset}{\Gamma_1, \Delta_1; \Gamma_2, \Delta_2, \Gamma \vdash S, T}\ (\mathsf{par})$$

The following properties stating a statical relation between Bioβ and κ hold.

Proposition 4.1 (Syntax).
1. For S a κ-solution, if $\Gamma_1; \Gamma_2 \vdash S$ and $\Delta_1; \Delta_2 \vdash S$ then $\Gamma_1 = \Delta_1$ and $\Gamma_2 = \Delta_2$.
2. For S a κ-solution, S is graph-like iff $\Gamma_1; \Gamma_2 \vdash S$ for some Γ_1, Γ_2.
3. For S, T two κ-solutions, if $S \equiv T$, then $\Gamma_1; \Gamma_2 \vdash S$ iff $\Gamma_1; \Gamma_2 \vdash T$.
4. For P a Bioβ system, if $\Gamma_1; \Gamma_2 \vdash P : \tau$, then $\Gamma_1; \Gamma_2 \vdash [\![P]\!]$.

Finally, we can state and prove the following semantics/dynamic relation between Bioβ and κ:

Proposition 4.2 (Semantics).
1. For S, T two κ-solutions, if $\Gamma_1; \Gamma_2 \vdash S$ and $S \rightsquigarrow T$, then $\Gamma_1; \Gamma_2 \vdash T$.
2. For a wide monotone protein reaction $(\langle \vec{P}, \vec{S} \rangle, \nu \vec{x}.\langle \vec{P}', \vec{S}' \rangle) \in \mathscr{R}$, and for every $C[\vec{X}, \vec{Y}]$ such that $C[\vec{P}, \vec{S}]$ is well-formed, then $[\![C[\vec{P}, \vec{S}]]\!] \rightsquigarrow [\![\nu \vec{x}.C[\vec{P}', \vec{S}']]\!]$.
3. For a wide anti-monotone protein reaction $(\nu \vec{x}.\langle \vec{P}, \vec{S} \rangle, \langle \vec{P}', \vec{S}' \rangle) \in \mathscr{R}$, for every $C[\vec{X}, \vec{Y}]$ such that $\nu \vec{x}.C[\vec{P}, \vec{S}]$ is well-formed, then $[\![\nu \vec{x}.C[\vec{P}, \vec{S}]]\!] \rightsquigarrow [\![C[\vec{P}', \vec{S}']]\!]$.

Intuitively, this proposition states that every protein transition performed in the Bioβ framework is justified by a corresponding transition in the κ-calculus. The converse does not hold, due to the compartment confinements: not all reactions of the translation of a Bioβ system which are possible in κ-calculus, correspond effectively to some wide (anti)monotone protein reactions in the Bioβ framework.

5 Related works

In literature several other calculi, covering both protein reactions and membrane interactions, have been proposed. Two particularly interesting cases are the bioκ-calculus [9] and the $\mathscr{C}$-calculus [7]. In this section we compare our framework with these calculi analyzing analogies and differences of the approaches.

bioκ-calculus can be seen as a version of the κ-calculus extended with compartments. Compared to Bioβ, the most remarkable similarities are at the syntactic level. Both calculi have a compartment constructor which embeds proteins into membranes, and similar well-formedness conditions to restrict to only biological relevant processes. Both calculi are sound at protein level in the sense of the κ-calculus. In both cases preservation of well-formedness during reactions is proved, although bioκ-calculus does not have a formal type system as Bioβ.

Several differences emerge at the semantic level, though. An important design choice distingushing bioκ and Bioβ concerns how the behaviour of a compartment is specified. In bioκ, the mobility capabilites of a compartment are indicated by its *name*, which appears explicitly in the rearrangement rules which can be applied. The result of a membrane interaction is one or more new membranes whose names have to be specified in the rearrangement rule. This means that the rule defines also the future (mobility) behaviour of the resulting systems; hence it is up to the modeller to specify correctly the evolutions of names of membranes.

On the other hand, in Bioβ compartments have no names: membranes are just containers and their behaviour is completely determined by the interactions of proteins floating in the aqueous solutions and lipid bi-layers. Therefore, a modeller has only to provide protein-level rules and define the configurations which trigger a mobility reaction. One does not have to declare how a membrane will evolve (i.e., "which is the next name" in bioκ), because the behaviour is defined by the rules of the framework.

Due to these radically different design choices, a formal comparison at the semantic level between bioκ-calculus and Bioβ is not easy. Moreover, the bioκ semantics is given by means of an LTS, instead Bioβ has a reaction (reduction) relation. A possible future work is to derive an LTS with associated congruential bisimilarity, taking advantage of the theory of bigraphs.

$\mathscr{C}$-calculus Both $\mathscr{C}$-calculus and Bioβ frameworks are founded on bigraphical reactive systems, which does not only confer a formal graphical representation but provides also a powerful categorical theory for automatically constructing labelled transition systems (LTSs) whose bisimilarities are always congruences. Even though $\mathscr{C}$-calculus and Bioβ share common sources of inspiration, they differs in many aspects, one of which is the way compartment reconfiguration are performed.

$\mathscr{C}$-calculus adopts meta-biological "gates" to represent a semi-fused intermediate state in membrane fusion or fission. Gates act like channels, allowing protein diffusion between compartments, and it does not allow for phagocytosis. Moreover, one of the aim of Bioβ framework is to characterize higher-level membrane interactions with lower-level protein reactions, and to be adequate with respect to the κ-calculus; a formal connection between $\mathscr{C}$-calculus and κ-calculus has not been investigated yet. Finally, in the $\mathscr{C}$-calculus, mobility and protein reactions are more intertwined, since membrane interactions are in many steps: introduction of gates, some protein diffusions, and deletion of gates. This mechanism does not provide an explicit, logical separation between compartment reconfiguration and protein reactions. On the other hand, the Bioβ framework aims to an explicit distinction between protein reactions and mobility actions, as implemented by means of introduction rules, which freeze the subsystems involved in the rearrangement until the structural reconfiguration is actually performed.

6 Conclusion

In this paper we have presented the Bioβ Framework, a language for both (lower-level) protein and (higher-level) membrane interactions of living cells. A typed language has been given for describing only biologically meaningful compartments and proteins.

In the Bioβ framework membrane reconfigurations are logically different from protein reactions, but the connection between the two aspects is formally specified. In fact, membrane activities can be motivated by a formal biological justication in terms of protein interactions. Notably, protein interactions are adequate with respect to the κ-calculus.

As an example application of this framework, we have modelled the mechanism of membrane traffic, which consists in formation, transportation and docking of vesicles.

Related and future work. Due to lack of space, we have omitted a formal connection between the proposed framework and its inspiring bigraphical model. Another interesting aspect to investigate is the bisimilarity definable via the IPO construction on bigraphs [10], which is always a congruence. This can be useful e.g. in *synthetic biology*, for instance, to verify if a synthetic protein (or even a subsystem to be implanted) behaves as the natural one. We plan to describe such results in a forthcoming work.

Another interesting investigation is to give a formal comparison of Bioβ with the other two framework present in literature, such as bioκ and $\mathscr{C}$-calculus. In particular, it could be useful to understand the differences of $\mathscr{C}$-calculus rule generators and our introduction rules, even because they are both inspired by bigraphical wide reactions.

As a next step, a stochastic version of the Bioβ framework should be analyzed for simulation purposes. One way could be trying to adapt the extended Gillespie simulation algorithm [11], that works for calculi with multi-compartments.

References

[1] Alan Aderem and David M. Underhill. Mechanisms of phagocytosis in macrophages. *Annual Review of Immunology*, 17(1):593–623, 1999. PMID: 10358769.

[2] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Essential Cell Biology, 2nd edition*. Garland, 2003.

[3] Giorgio Bacci, Davide Grohmann, and Marino Miculan. Bigraphical models for protein and membrane interactions. In Gabriel Ciobanu, editor, *Proc. MeCBIC'09*, EPTCS, 2009.

[4] Luca Cardelli. Brane calculi. In Vincent Danos and Vincent Schächter, editors, *Proc. CMSB*, volume 3082 of *Lecture Notes in Computer Science*, pages 257–278. Springer, 2004.

[5] Luca Cardelli. Abstract machines of systems biology. *T. Comp. Sys. Biology*, 3737:145–168, 2005.

[6] Luca Cardelli. Bitonal membrane systems: Interactions of biological membranes. *Theoretical Computer Science*, 404(1-2):5 – 18, 2008. Membrane Computing and Biologically Inspired Process Calculi.

[7] Troels Christoffer Damgaard, Vincent Danos, and Jean Krivine. A language for the cell. Technical Report TR-2008-116, IT University of Copenhagen, December 2008.

[8] Vincent Danos and Cosimo Laneve. Formal molecular biology. *Theoretical Computer Science*, 325, 2004.

[9] Cosimo Laneve and Fabien Tarissan. A simple calculus for proteins and cells. *Theor. Comput. Sci.*, 404(1-2):127–141, 2008.

[10] Robin Milner. Pure bigraphs: Structure and dynamics. *Information and Computation*, 204(1):60–122, 2006.

[11] Cristian Versari and Nadia Busi. Stochastic simulation of biological systems with dynamical compartment structure. In Muffy Calder and Stephen Gilmore, editors, *Proc. CMSB*, volume 4695 of *Lecture Notes in Computer Science*, pages 80–95. Springer, 2007.