

Selected parts of the Kappa model for the TGF-beta extracellular matrix

Nathalie Th  ret^{1,2}, J  r  me Feret^{3,4}, Pierre Boutillier⁵, and Pierre
Vignet^{1,2}

¹Univ Rennes, Inserm, EHESP, Irset - UMR_S1085, F-35043 Rennes,
France

²Univ Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France.

³Inria Paris, Paris, France

⁴DI-  NS,   NS, CNRS, PSL University, Paris, France

⁵Harvard Medical School, Boston, USA

January 17, 2020

Modeling the extracellular matrix can hardly be done by traditional techniques, because it involves the formation of large compounds of proteins. We use the Kappa modeling environment [Boutillier et al., 2018] to summarize the knowledge of molecular interactions surrounding the activation of TGF-   in the ECM.

Complex systems of interaction between molecules are difficult to model for several reasons. Firstly, due to differing modalities of binding and post-translational changes of conformation, there exists a large (if not infinite, in the case of polymers) variety of molecular complexes, making them impossible to enumerate. Secondly, the dynamics of these systems is usually triggered by concentration- and time-scale separation; competition to share resources; complex causality chains; and non linear feedback loops. As a consequence, taking a biochemical approach, summarising reactions between molecules as generic, local patterns of interactions seems to be the only viable alternative for modeling these systems and understanding how the dynamics of these populations of molecules emerges from individual interactions at the microscopic level.

Kappa [Danos and Laneve, 2004] is a site-graph rewriting formalism, that is freely inspired by reaction schema encountered in organic chemistry. The main idea is to describe each instance of protein as a node in the graph. Each kind of protein has some interaction sites which can bind pair-wise. The interactions between molecules are formalised by the means of rewrite rules. The rules either stand for interactions that are detailed in the literature or for some fictitious interactions that exist to make assumptions about the information that is missing or to roughly simplify some parts that we do not want to detail too much. The use of rules eases frequent updates of the models, which enables the modeler to test numerous scenarios or to modify the context

of the model. A set of rules can be interpreted as a dynamical model which describe the evolution of the soup of molecules. There are several choices: when the varieties of molecular complexes is not too **great**, the set of rules may be translated into **ODEs**. **This** set of rules also induces a continuous time Markov chain; the traces of which can be sampled by simulation []. Due to the use of specific data-structures [], simulation cost does not depend on the number **or** kind of molecular complexes, which may even be infinite.

Kappa ecosystem [] offers several tools to assist the modeler during her task. **Mod-els** can be curated by **means** of static analysis [] and causal analysis []. Formal methods can also be used to identify the key elements in information propagation. The result is a model reduction which does not lose any information about the quantities that are observed [].

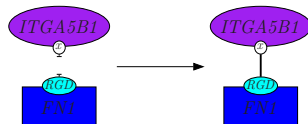
We wrote a model **for the influence on the ECM of TGF- β** , including some interactions that are documented in the literature, and some fictitious rules to stub gene activity and its interaction with TGF- β . The model is **composed of around** 300 interaction rules. Each rule is parameterised by a kinetic rate. Some of the rates are **precisely** deduced from information about the concentration of proteins at stationary distribution and their half-time periods. Some others are **chosen** approximately, in order to **best** model what is known about the time scales of each interaction.

So as to illustrate the use of Kappa, we focus on few small parts of the model.

1 Fibril complexation

We start with three rules which describe some interactions that are involved in the formation of fibril complexes. A fibril complex is made of a concatenation of several *FNI* dimers. In order to form a dimer, two *FNI* proteins shall be activated by an $\alpha 5 \beta 1$ integrin. When activated the spatial conformation of the *FNI* proteins is modified, which reveals a site which may bind to another *FNI* protein. Dimers may then assemble to each other to form double-chains of proteins.

The following rule describes the potential binding between an integrin protein and an *FNI* protein:

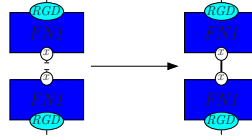


This rule is indeed a schema of reactions. The left hand side describes the conditions to be satisfied for the interaction to possibly happen, whereas the difference between both sides denotes the potential **impact** of applying the rule. Here two proteins are required; an $\alpha 5 \beta 1$ integrin protein and an *FNI* protein. These instances of protein come with a partial description of their states, which takes the form of a list of sites which may be free or bound pairwise. Here the integrin protein and the *FNI* protein shall have respectively their site x and *RGD* free (which is denoted by the symbol \neg). The proteins may have many other interaction sites. Yet they are not mentioned in the rules to **hypothesise** that their state does not matter for this interaction. When applied, the interaction binds the site x of the integrin and the site *RGD* of the protein

FNI. This rule is quite unspecific. The integrin may bind a *FNI* protein no matter **if** it is isolated, in a dimer, or even in a polymer.

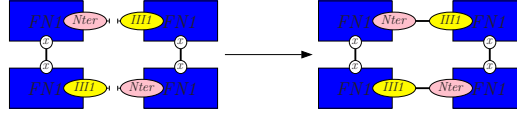
Each rule is associated with a rate constant. Each rate corresponds to a given time-scale. Their precise value is left as a parameter of the model. More precisely, we assume two sets of rates depending on the arity of the rules. The rates $k_2^l \ll k_2^n \ll k_2^h$ stand for association rules, while the rates $k_1^l \ll k_1^n \ll k_1^h$ stand for dissociation and modification rules.

The following rule formalises the symmetric binding between two activated *FNI* proteins:



The left hand side stipulates that two *FNI* proteins are required. Moreover their sites *RGD* must both be bound. Yet the sites to which they are bound are not specified (which is denoted by the symbol $-$). This way, the *FNI* protein may be activated by any kind of integrin. The sites that are used to establish symmetric bonds in dimers are not known precisely, we call them x and requires both of them to be free. Under these conditions, both *FNI* may bind to each other as specified in the right hand side of the rule. The states of the other sites do not matter for this interaction. Interestingly, we notice that the change of conformation of the protein, when activated by an integrin is not explicitly described. It is encoded implicitly by the fact that the dimerisation rules requires the *FNI* to be bound on their sites *RGD* to form dimers.

This third rule specifies that dimers may bind to each **other**, in order to form **co-polymers**:

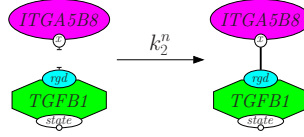


Polymer length is not mediated directly by the rules. It is a result of an equilibrium between the rules to extend the size of the polymers and the rules which may break the bonds within polymers (which are omitted here for the sake of brevity). The longer a polymer is, the more bonds it contains and the more likely it is to be split into several parts.

2 TGF- β activation

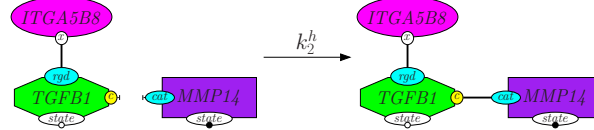
Some sites may **carry an** internal state, which may stand for a level of energy (such as phosphorylation, methylation, or ubiquitination level). The activation of such a site is usually done in two steps: firstly the protein shall recruit another protein which will then exchange some energy (under various means).

Here we describe the activation of *TGFB1* proteins under the action of *MMP14* proteins. In order, to bind to *MMP14*, *TGFB1* proteins shall be bound to an $\alpha 5\beta 8$ integrin, by the means of the following rule:

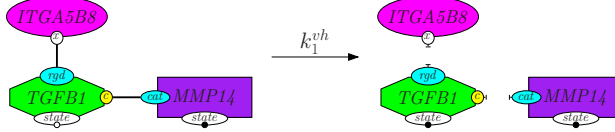


This is only possible when the *TGFBI* protein is in latent form (which is denoted by a white disk attached to the site *state*).

Then an active instance of the *MMP14* protein may bind to this instance of the *TGFBI* protein as described as follows



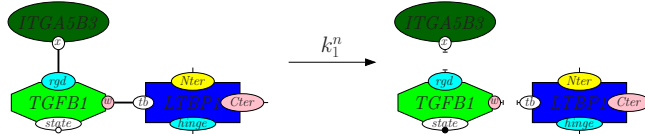
and activates it (the white disk becomes black) whereas both the integrin and the *MMP14* protein dissociate from the *TGFBI* protein.



3 TGF- β activation

Some sites may carry out an internal state, which may stand for a level of energy (such as phosphorylation, methylation, or ubiquitination level). The activation of such a site is usually done in two steps: firstly the protein shall recruit another protein which will then exchange some energy (under various means).

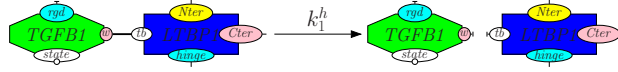
Here we describe the activation of *TGFBI* proteins under the action of *LTBP1* proteins. A *TGFBI* protein in its latent form (which is denoted by a white disk attached to the site called *state*) may get activated when it is bound to both an integrin and an *LTBP1* with its site *Nter*, *Cter*, and *hinge* bound. At an intermediary time scale, the protein may get activated (the white disk becomes black) while breaking its connection to these proteins as depicted in the following rule:



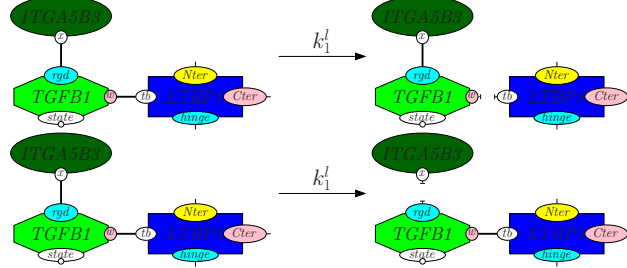
Est-ce que la liaison avec l'intégrin est vraiment nécessaire ? Avec le jeu des constantes cinétiques, si *TGFBI* n'est pas lié à une intégrin, la liaison avec *LTBP1* va vraisemblablement se casser avant que *TGFBI* ne s'active.

In order for these interactions to occur, the bond with the *LTBP1* protein shall be strong enough. Indeed, when the protein *TGFBI* is not connected to an integrin, this bond may be released at high time-scale as shown in the following rule:

La pour moi, le fait que *Nter*, *Cter*, et *hinge* doivent être liées est assez suspect.



Even, when connected to an integrin, the connection with the integrin or with the *LTBP1* may be broken at low time-scale, as written in the following rules:

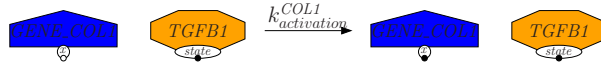


Due to time-scale separation, it is very likely that a *TGFBI* bound both to an integrin and to a *LTBP1* protein gets activated. Yet, only with low probability, one of the connection may break before the activation of the protein *TGFBI*.

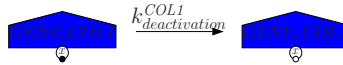
4 A black-box pattern for gene expression

When modeling complex systems of interactions between molecules, it often happens that not enough knowledge about one part of the model is available. **Even when this knowledge is** accurate enough, it **is often** too complicated to encode the information in a comprehensive way. In such cases, some parts of the model may be abstracted into **simpler** ones, provided that the **latter approximately guarantee** the expected behavior.

Here we describe a generic pattern that we use to model at a very rough level gene activity. Our simplification has two main folds. Firstly we describe neither the distributions of genes within DNA strands nor the gliding of RNA polymerases. Secondly the production of proteins is abstracted into a single step (we abstract away RNA). In this simplistic view, genes are described as sets of instances of abstract agents. Each of these instances may be visible (black disk), or hidden (white disk). Gene activity results **from** an equilibrium between gene activation and deactivation. Promoters enhance the proportion **of** visible genes. In the following rule, we show how instances of collagen genes may be activated by their promoter (*TGFBI* in its active form):



Conversely, visible instances of genes may spontaneously become hidden, as specified by the following rule:



We assume, that in the absence of visible genes, the concentration of collagen results from an equilibrium between synthesis and degradation, which is modeled by the following rules:



Visible instances of collagen genes boost the production of collagen as depicted in the following rule:



The rates of rules are chosen in order to satisfy what is known about the the concentration of collagen at low and high gene activity, as well as the half-time period of each instance of collagen protein and the relaxation time of genes. These **quantities** are usually available in the literature.

References

- [Boutillier et al., 2018] Boutillier, P., Maasha, M., Li, X., Medina-Abarca, H. F., Krivine, J., Feret, J., Cristescu, I., Forbes, A. G., and Fontana, W. (2018). The kappa platform for rule-based modeling. *Bioinformatics*, 34(13):i583–i592.
- [Danos and Laneve, 2004] Danos, V. and Laneve, C. (2004). Formal molecular biology. *Theor. Comput. Sci.*, 325(1):69–110.