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

Nathalie Théret[1,2], Jérôme Feret[3,4], Pierre Boutillier[5], and Pierre Vignet[1,2]

[1]Univ Rennes, Inserm, EHESP, Irset - UMR_S1085, F-35043 Rennes, France
[2]Univ Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France.
[3]Inria Paris, Paris, France
[4]DI-ÉNS, ÉNS, CNRS, PSL University, Paris, France
[5]Harvard Medical School, Boston, USA

February 13, 2020

1 Kappa, a formalism adapted to model the biological component networks of the extracellular matrix

Modeling the ECM 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., 2018b] to summarize the knowledge that is available in the literature about the molecular interactions [surrounding the activation of TGF-β in the ECM.

Complex systems of interaction between [molecules are difficult to model for several reasons. Firstly, due to many potential bindings between proteins and numerous potential post-translational changes of conformation, there [exists a large (if not infinite[, in the case of polymers) [number of different kinds of molecular complexes. It is often even impossible to enumerate them. Secondly, the dynamics of these systems is usually triggered by concentration- and time-scale separation[; competition against shared-resources[; complex causality chains[; and non linear feedback loops. As a consequence, [taking a biochemical approach, which consists in [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 their populations of molecules may emerge 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 a 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 environment of the model. A set of rules can be interpreted as a dynamical system which describes the evolution of a soup of molecules. There are several choices: when the number of different kinds of molecular complexes is not too [great, the set of rules may be translated into [ODEs [Camporesi et al., 2017]. Each set of rules also induces a continuous time Markov chain the execution traces of which can be sampled by simulation [Danos et al., 2007b]. Thanks to the use of specific data-structures [Danos et al., 2007b, Boutillier et al., 2017], the computation cost of such simulation does not depend on the number [of kinds of molecular complexes, which may even be infinite.

Kappa ecosystem [Boutillier et al., 2018b] offers several tools to assist the modeler during her task. Static analysis [Danos et al., 2008, Feret and Lý, 2018, Boutillier et al., 2018a] may be used to curate models. Canonical and secondary pathways may be extracted thanks to causality analysis [Danos et al., 2007a, Danos et al., 2012]. Formal methods can also be used to identify the key elements in information propagation. The result is a model reduction which never loses any information about the quantities that are observed in the model [Feret et al., 2009, Danos et al., 2010, Camporesi et al., 2013].

We wrote a model for the influence of the ECM on TGF-β signal, including numerous extracellular interactions that are documented in the literature[, and some fictitious rules to stub gene activity and its interaction with TGF-β. The model is [made of around 300 interaction rules which are freely available on the web [Théret et al., 2020]. Each rule is parameterised by a kinetic rate. Some of the rates are deduced from precise 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. Our model comprises around 30 kinds of proteins. The potential bonds between these proteins are summarised in Fig. 1, which provides a convenient snapshot of the model, while not detailing every rule. Selected portions of the models are depicted in the following sections:

2 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 *FN1* dimers. In order to form a dimer, two *FN1* proteins shall be activated by an $\alpha 5\beta 1$ integrin. When activated the spatial conformation of the *FN1* proteins is modified, which reveals a site which may bind to another *FN1* 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 *FN1* protein:

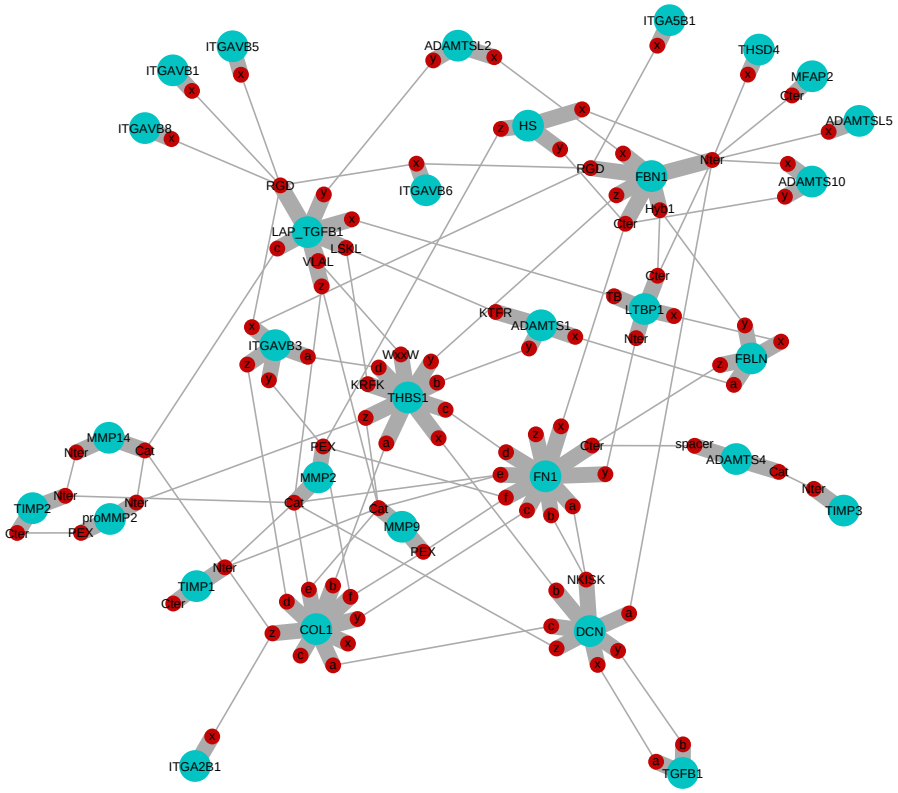
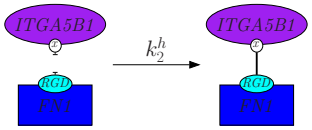


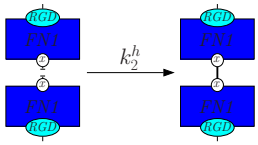
Figure 1: Contact map of the Kappa model for TGF-β activation: Projection of model describing the molecule interaction networks. Proteins and glycosaminoglycans are represented by turquoise nodes, binding sites are represented by red nodes (if the sequence involved is not known, the site is designated by a letter, x, y, z etc). The lines between the sites illustrate potential links involving these binding sites. ITGA-x-B-y, Integrin alpha-x Beta-y ; LAP-TGFB1, latent TGFB1 ; THBS1, Thrombospondin ; HS, Heparan Sulfate ; FBN1, Fibrillin 1 ; FN1, Fibronectin, FBLN, Fibulin ; THSD4, ADAMTSL-6 ; MFAP2, Microfibril Associated Protein 2 ; LTBP1, Latent Transforming Growth Factor Beta Binding Protein 1 ; MMP, Matrix Metalloprotease ; TIMP, Tissue inhibitor of MMP ; COL1, Type 1 collagen ; DCN, Decorin.



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 *FN1* 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 *FN1* protein shall have respectively their site x and *RGD* free (which is denoted by the symbol $\dashv$). 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 *FN1*. This rule is quite unspecific. The integrin may bind a *FN1* protein no matter if it is isolated, in a dimmer, or even in a polymer.

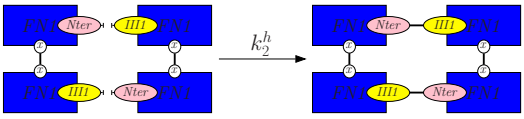
Each rule is associated with a rate constant. Each rate corresponds to a given timescale. Their precise value is left as a parameter of the model. More precisely, we assume two sets of rates depending on the arity of the rules. The rates $k_2^l << k_2^n << k_2^h$ stand for association rules, while the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

The following rule formalises the symmetric binding between two activated *FN1* proteins:



The left hand side stipulates that two *FN1* 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 *FN1* 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 *FN1* 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 *FN1* 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 copolymers:



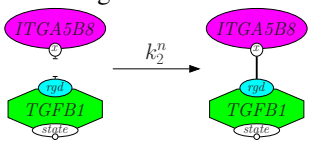
Polymer length is not mediated directly by the rules. It is a result of an equilibrium between the rules to extend the size of the polymers and the rules which may break the

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.

3 TGF-β activation

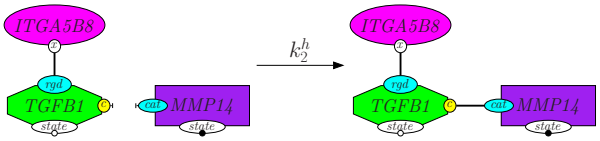
Some sites may carry an internal state, which may stand for a level of energy (such as phosphorylation, methylation, or ubiquitination level). The activation of such a site is usually done in two steps: firstly the protein shall recruit another protein which will then exchange some energy (under various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

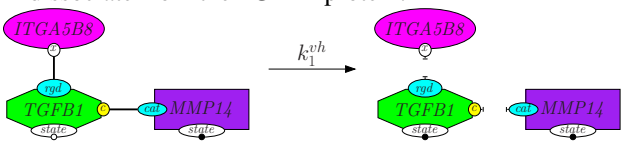


This is only possible when the *TGFB1* protein is in latent form (which is denoted by a white disk attached to the site *state*).

Then an active instance of the *MMP14* protein may bind to this instance of the *TGFB1* protein as described as follows

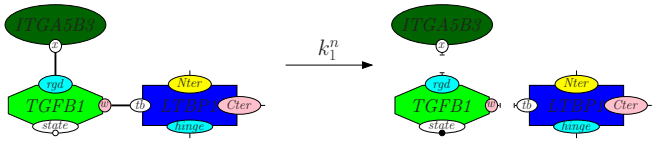


and activates it (the white disk becomes black) whereas both the integrin and the *MMP14* protein dissociate from the *TGFB1* protein.

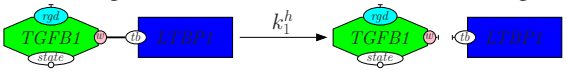


4 TGF-β releasing

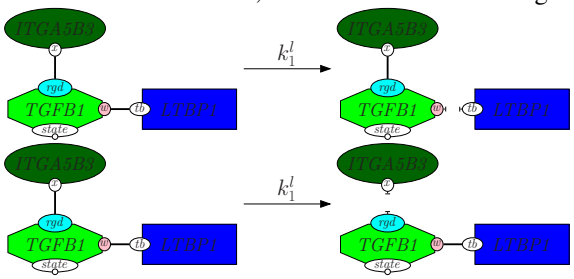
Here we describe how *TGFB1* proteins is released from the extra-cellular matrix under the action of *LTBP1* proteins. A *TGFB1* protein in its latent form (which is denoted by a white disk attached to the site called *state*) may get actived when it is bound to both an integrin and an *LTBP1* with its 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:



In order for these interactions to occur, the bond with the *LTBP1* protein shall be strong enough. Indeed, when the protein *TGFB1* is not connected to an integrin, this bond may be released at high time-scale as shown in the following rule:



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:

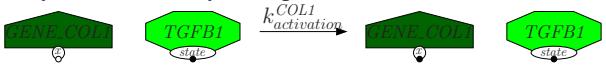


Due to time-scale separation, it is very likely that a *TGFB1* bound both to an integrin and to a *LTBP1* protein gets activated. Yet, only with low probability, one of the connection may break before the activation of the protein *TGFB1*.

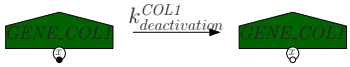
5 A black-box pattern for gene expression

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 (*TGFB1* 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., 2018a] Boutillier, P., Camporesi, F., Coquet, J., Feret, J., Lý, K. Q., Théret, N., and Vignet, P. (2018a). Kasa: A static analyzer for kappa. In Ceska, M. and Safránek, D., editors, *Computational Methods in Systems Biology - 16th International Conference, CMSB 2018, Brno, Czech Republic, September 12-14, 2018, Proceedings*, volume 11095 of *Lecture Notes in Computer Science*, pages 285–291. Springer.

[Boutillier et al., 2017] Boutillier, P., Ehrhard, T., and Krivine, J. (2017). Incremental update for graph rewriting. In Yang, H., editor, *Programming Languages and Systems - 26th European Symposium on Programming, ESOP 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings*, volume 10201 of *Lecture Notes in Computer Science*, pages 201–228. Springer.

[Boutillier et al., 2018b] Boutillier, P., Maasha, M., Li, X., Medina-Abarca, H. F., Krivine, J., Feret, J., Cristescu, I., Forbes, A. G., and Fontana, W. (2018b). The kappa platform for rule-based modeling. *Bioinformatics*, 34(13):i583–i592.

[Camporesi et al., 2013] Camporesi, F., Feret, J., and Hayman, J. M. (2013). Context-sensitive flow analyses: A hierarchy of model reductions. In Gupta, A. and Henzinger, T. A., editors, *Computational Methods in Systems Biology - 11th International Conference, CMSB 2013, Klosterneuburg, Austria, September 22-24, 2013. Proceedings*, volume 8130 of *Lecture Notes in Computer Science*, pages 220–233. Springer.

[Camporesi et al., 2017] Camporesi, F., Feret, J., and Lý, K. Q. (2017). Kade: A tool to compile kappa rules into (reduced) ODE models. In Feret, J. and Koeppl, H., editors, *Computational Methods in Systems Biology - 15th International Conference,*

CMSB 2017, Darmstadt, Germany, September 27-29, 2017, Proceedings, volume 10545 of *Lecture Notes in Computer Science*, pages 291–299. Springer.

[Danos et al., 2012] Danos, V., Feret, J., Fontana, W., Harmer, R., Hayman, J. M., Krivine, J., Thompson-Walsh, C. D., and Winskel, G. (2012). Graphs, rewriting and pathway reconstruction for rule-based models. In D'Souza, D., Kavitha, T., and Radhakrishnan, J., editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2012, December 15-17, 2012, Hyderabad, India*, volume 18 of *LIPIcs*, pages 276–288. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik.

[Danos et al., 2007a] Danos, V., Feret, J., Fontana, W., Harmer, R., and Krivine, J. (2007a). Rule-based modelling of cellular signalling. In Caires, L. and Vasconcelos, V. T., editors, *CONCUR 2007 - Concurrency Theory, 18th International Conference, CONCUR 2007, Lisbon, Portugal, September 3-8, 2007, Proceedings*, volume 4703 of *Lecture Notes in Computer Science*, pages 17–41. Springer.

[Danos et al., 2010] Danos, V., Feret, J., Fontana, W., Harmer, R., and Krivine, J. (2010). Abstracting the differential semantics of rule-based models: Exact and automated model reduction. In *Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science, LICS 2010, 11-14 July 2010, Edinburgh, United Kingdom*, pages 362–381. IEEE Computer Society.

[Danos et al., 2007b] Danos, V., Feret, J., Fontana, W., and Krivine, J. (2007b). Scalable simulation of cellular signaling networks. In Shao, Z., editor, *Programming Languages and Systems, 5th Asian Symposium, APLAS 2007, Singapore, November 29-December 1, 2007, Proceedings*, volume 4807 of *Lecture Notes in Computer Science*, pages 139–157. Springer.

[Danos et al., 2008] Danos, V., Feret, J., Fontana, W., and Krivine, J. (2008). Abstract interpretation of cellular signalling networks. In Logozzo, F., Peled, D. A., and Zuck, L. D., editors, *Verification, Model Checking, and Abstract Interpretation, 9th International Conference, VMCAI 2008, San Francisco, USA, January 7-9, 2008, Proceedings*, volume 4905 of *Lecture Notes in Computer Science*, pages 83–97. Springer.

[Danos and Laneve, 2004] Danos, V. and Laneve, C. (2004). Formal molecular biology. *Theor. Comput. Sci.*, 325(1):69–110.

[Feret et al., 2009] Feret, J., Danos, V., Krivine, J., Harmer, R., and Fontana, W. (2009). Internal coarse-graining of molecular systems. *Proceedings of the National Academy of Sciences of the United States of America*.

[Feret and Lý, 2018] Feret, J. and Lý, K. Q. (2018). Reachability analysis via orthogonal sets of patterns. *Electr. Notes Theor. Comput. Sci.*, 335:27–48.

[Théret et al., 2020] Théret, N., Feret, J., Cocquet, J., Vignet, P., Boutillier, P., and Camporesi, F. (2020). `https://github.com/feret/TGF-Kappa/tree/master/model`.