Multiscale Modeling of Alternative Splicing Regulation

Damien Eveillard[1*], Delphine Ropers[2], Hidde de Jong[3], Christiane Branlant[2], and Alexander Bockmayr[1**]

[1] LORIA, Université Henri Poincaré, BP 239, 54506 Vandœuvre-lès-Nancy, France
[2] Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP, BP 239, 54506 Vandoeuvre-lès-Nancy, France
[3] INRIA Rhône-Alpes, Helix Project, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier, France

Abstract. Alternative splicing is a key process in post-transcriptional regulation, by which several kinds of mature RNA can be obtained from the same premessenger RNA. Using a constraint programming approach, we model the alternative splicing regulation at different scales (single site vs. multiple sites), thus exploiting different types of available experimental data.

1 Introduction

Alternative splicing is a biological process occurring in post-transcriptional regulation of RNA. Through the elimination of selected introns, alternative splicing allows generating several kinds of mRNA from the same premessenger RNA. The combinatorial effect of splicing contributes to biological diversity, especially in the case of the human immunodeficiency virus (HIV-1). Recent biological studies show the impact that SR proteins have on the dynamics of post-transcriptional regulation via the control of the splicing process [8]. SR proteins can be divided into two functional classes: they may either activate or inhibit splicing. Due to the complexity of alternative splicing regulation, the knowledge that can be gained from experiments is limited. Each experiment focus on one splicing site. In a first approach, we model SR regulation in this restricted context. Using differential equations, we develop a qualitative model for the A3 splicing site in HIV-1. The qualitative behavior depends on the values of the reaction kinetic parameters. Experimental results available to us validate this first approach in the equilibrium phase. Our second approach aims at validation on a higher scale. The ultimate goal is to obtain a model that can be validated qualitatively both on the scale of a single splicing site and on the scale of the whole HIV-1, in order to represent the global effect of alternative splicing in the HIV-1 cycle.

* Corresponding author: `Damien.Eveillard@loria.fr`

** Part of this work was done within the ARC INRIA "Process Calculi and Biology of Molecular Networks", `http://contraintes.inria.fr/cpbio`

Our models are developed in a constraint programming framework [2, 3]. Constraint programming seems well-suited for modeling biological systems because it allows one to handle partial or incomplete information on the system state. Each constraint gives one piece of information on the system. The overall knowledge is accumulated in the constraint store. The constraint engine available in constraint programming systems operates on the constraint store. It may add new information to the store or check whether some property is entailed by the information already available.

While a constraint model may be refined whenever additional biological knowledge becomes available, it allows one to make useful inferences even from partial and incomplete information. Therefore, constraint programming seems to be a natural computational approach to face the current situation in systems biology as it is described by B. Palsson [16]: "Because biological information is incomplete, it is necessary to take into account the fact that cells are subject to certain constraints that limit their possible behaviors. By imposing these constraints in a model, one can then determine what is possible and what is not, and determine how a cell is likely to behave, but never predict its behavior precisely."

The organisation of the paper is as follows: we start in Sect. 2 with a description of the biological process of alternative splicing regulation. Based on a number of biological hypotheses, we develop a continuous model of the regulation at one splicing site. This model includes competition and compensation of different proteins on two binding sites, ESE and ESS2. The single-site model is validated in a qualitative way by extracting from the model a splice efficiency function, which can be measured in experiments. In Sect. 3, we first simulate the single-site continuous model in the hybrid concurrent constraint programming language `Hybrid cc` [9, 10]. Then we derive a more global model involving two generic splicing sites, which may be generalized to multiple sites. This means that we model at two different scales, using the splice efficiency as an abstraction of the local model of one site in the more global context of different sites. The two-site model uses the constraint solving and default reasoning facilities of `Hybrid cc`. This allows us to make predictions on the global behavior even in absence of detailed local information on some of the splicing sites.

2 The biological problem of alternative splicing regulation

2.1 Biological process

Eukaryotic and virus gene expression is based on production of RNA containing intronic sequences. The process of splicing allows for intron elimination and junction of exonic sequences [14], see Fig. 1. Splicing is a major biological process in the HIV-1 life cycle: the viral RNA either remains unchanged to serve as genomic RNA for new virions, or it is alternatively spliced to allow for the production of virion proteins [26]. The viral genome, initially RNA, is integrated into the host genome. In the HIV-1 case, splicing regulation is a complex phenomenon involving 4 donors sites (SD) and 8 acceptors sites (SA), which may yield 40 proteins [17].

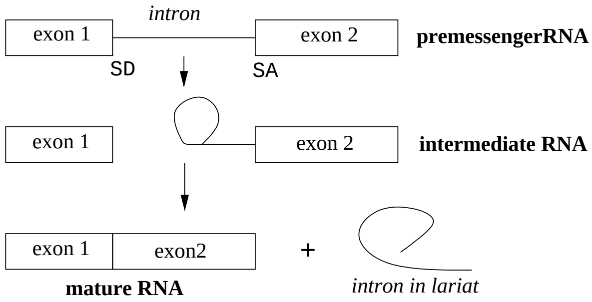

Fig. 1. Representation of the splicing process

This combinatorial complexity is achieved by regulating the selection of the acceptor site [15, 17]. Protein factors control the regulation via binding sites. We focus in our study on the acceptor site A3, where splicing can be repressed by hnRNP A/B via the ESS2 binding site [4, 6], see Fig. 2. Recent experimental studies carried out in our group [18] show that an ESE sequence can activate splicing at the A3 site when SC35 and ASF/SF2 proteins bind to it. More specifically, the ratio of hnRNP A/B and SR proteins determines the splice efficiency at the A3 site.

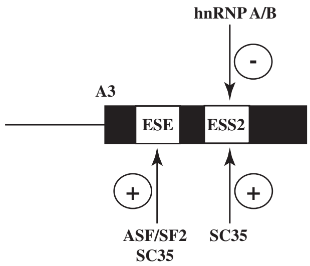

Fig. 2. Regulatory elements of the A3 splicing site

2.2 Biological hypotheses

We model the regulation by SR proteins in the restricted context of the A3 splicing site (see Fig. 2) under the following hypotheses:

– We study only one splicing site. Thus, we consider regulation at the scale corresponding to our experimental results. These yield the splice efficiency as the ratio of the mature RNA over premessenger RNA.
– We suppose that the splicing process involves two steps, relating three functional classes of RNA: immature, intermediate, and mature RNA, see Fig. 1.

Intermediate RNA corresponds to immature RNA activated by proteins. Mature RNA corresponds to mature RNA and introns in lariat.
– The protein concentration in experiments is saturated. Therefore, we assume that it is constant, despite the binding of proteins to the RNA during regulation.
– SR proteins have two functions. They regulate the splicing process, and they initialize the splicing machinery.
– Regulation is controlled by the ESE and ESS2 binding sites, which are independent. Thus, only indirect interaction is possible between ESE and ESS2.
– The SR proteins ASF/SF2 and SC35 may activate the first splicing reaction by binding to the site ESE. We assume that these two proteins compensate each other.
– The hnRNP proteins may inhibit the first splicing reaction by binding to the site ESS2. On the other hand, if the SC35 proteins bind to ESS2, this activates the first splicing reaction. Therefore we have a competition between hnRNP and SC35.
– Due to a lack of experimental results, we assume Michaelis-Menten kinetics for all proteins. Our goal is a qualitative model validated at equilibrium. Thus, we do not consider transient states.

Our biological hypotheses are summarized in Fig. 3

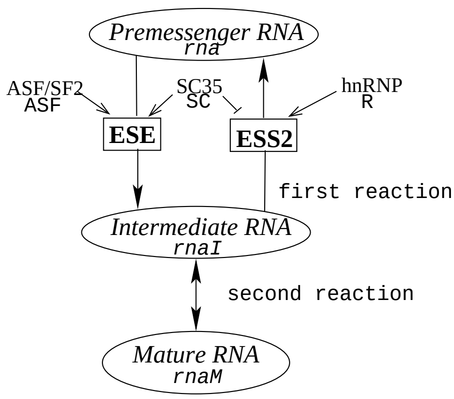

Fig. 3. Schematic representation of the splicing site regulation

2.3 Formal model

The biological hypotheses can be translated into a mathematical formalism, which leads to a system of ordinary differential equations based on the Michaelis-Menten relation [22]. The single-site model will later be integrated into a larger multi-site model, see Sect. 3. We describe the splicing process by three ordinary differential equations (ODEs) corresponding to the three functional classes of RNA. Two terms represent the first splicing reaction. The first term represents the cooperation between ASF/SF2 and SC35 in the regulation of ESE. Since we assume compensation, only the sum of activator proteins is important. We represent this interaction by a Michaelis-Menten function depending on the quantity of immature RNA and controlled by the sum of the proteins ASF/SF2 and SC35 [13]:

$$\frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna$$

The symbols used are given in Tab. 1. The second term captures the antagonistic function of hnRNP and SC35 proteins on the site ESS2. In this case, the Michaelis-Menten function represents the inhibitive competition between two proteins : hnRNP and SC35 [27]. It depends on the quantity of intermediary RNA:

$$\frac{\varphi_R \times R}{k_R(1 + \frac{SC}{k_{SC}+R})} rnaI$$

The second splicing reaction is modeled by a simple first order kinetic with constant parameters κ and κ'. Different RNAs decrease proportional to the same degradation factor λ. We formalize the biological process by the system of differential equations:

$$\begin{aligned}
\frac{d(rna)}{dt} &= \frac{\varphi_R \times R}{k_R(1 + \frac{SC}{k_{SC}+R})} rnaI - \frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna - \lambda \cdot rna \\
\frac{d(rnaI)}{dt} &= \frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna - \frac{\varphi_R \cdot R}{k_R(1 + \frac{SC}{k_{SC}+R})} rnaI - \kappa \cdot rnaI \\
&\quad + \kappa' \cdot rnaM - \lambda \cdot rnaI \\
\frac{d(rnaM)}{dt} &= \kappa \cdot rnaI - \kappa' \cdot rnaM - \lambda \cdot rnaM
\end{aligned}$$

2.4 Validation of the regulatory system

The formal model of regulation at a single-site can be directly simulated in the constraint programming language `Hybrid cc`, see Sect. 3.1. However, before doing this, it should first be validated with respect to existing biological knowledge.

A mathematical analysis of the ODE system in Sect. 2.3 shows that the partial derivatives in the Jacobian matrix have two characteristic properties:

- the partial derivatives on the diagonal elements are negative.
- the partial derivatives on the extra diagonal elements are positive.

Table 1. Symbols and units for the biological variables and parameters

Symbol	Variables and Parameters	unit
rna	Immature RNA	μM
$rnaI$	Intermediary RNA	μM
$rnaM$	Mature RNA	μM
ASF	Protein ASF/SF2	μM
SC	Protein SC35	μM
R	Protein hnRNP	μM
φ_{ESE}	Maximal affinity for the enhancer	s^{-1}
φ_R	Maximal affinity of hnRNP	s^{-1}
k_{ESE}	Half saturation coefficient for the enhancer	μM
k_{SC}	Half saturation coefficient for SC35	μM
k_R	Half saturation coefficient for hnRNP	μM
κ	Reaction rate	s^{-1}
κ'	Reaction rate	s^{-1}
λ	Degradation coefficient	s^{-1}

In our model, the RNA concentrations do not reach an equilibrium, but continue to decrease until total degradation of RNA. However, the above two properties imply that the *splice efficiency* defined by

$$\textit{efficiency}\ (t) = \frac{rnaM(t)}{rna(t)}$$

reaches an equilibrium [1]. From the condition $d\,\textit{efficiency}/dt = 0$, we may derive the following formula for the splice efficiency in the equilibrium phase:

$$\textit{efficiency} = \frac{\kappa \cdot \varphi_{ESE}(ASF + SC)(k_R \cdot k_{SC} + k_R \cdot SC + R \cdot k_{SC})}{\kappa'(k_{ESE} + ASF + SC) \cdot \varphi_R \cdot R \cdot k_{SC}}$$

It is reasonable to assume that this equilibrium is reached after a short transient phase, so that it can be measured by experiments. According to our formula, the splice efficiency is

- an increasing function of the activators SC and ASF.
- a decreasing function of the inhibitor R.

Experimental results show that

- $rnaM/rna$ increases with an increase of activator proteins.
- $rnaM/rna$ decreases with an increase of inhibitor proteins.

These results of our model correlate with available experimental data. In summary, the model may be considered as qualitatively validated under the hypotheses described in Sect. 2.2. We next consider simulation in the concurrent constraint language `Hybrid cc`.

3 Multiscale modeling and simulation with Hybrid cc

3.1 Hybrid concurrent constraint programming

In constraint programming, the user specifies constraints on the behavior of the system that is being studied. Each constraint expresses some partial information on the system state. The constraint solver may check constraints for consistency or infer new constraints from the given ones. In concurrent constraint programming (`cc`), different computation processes may run concurrently. Interaction is possible via the *constraint store*. The store contains all the constraints currently known about the system. A process may *tell* the store a new constraint, or *ask* the store whether some constraint is entailed by the information currently available, in which case further action is taken [19]. One major difficulty in the original `cc` framework is that `cc` programs can detect only the presence of information, not its absence. To overcome this problem, [20] proposed to add to the `cc` paradigm a sequence of phases of execution. At each phase, a `cc` program is executed. At the end, absence of information is detected, and used in the next phase. This results in a synchronous reactive programming language, `Timed cc`. But, the question remains how to detect negative information instantaneously. `Default cc` extends `cc` by a negative ask combinator `if` c `else` A, which imposes the constraints of A unless the rest of the system imposes the constraint c. Logically, this can be seen as a default. Introducing phases as in `Timed cc` leads to `Timed Default cc` [21]. Only one additional construct is needed: `hence` A, which starts a copy of A in each phase after the current one.

Agents	Propositions
c	c holds now
`if` c `then` A	if c holds now, then A holds now
`if` c `else` A	if c will not hold now, then A holds now
`new` X `in` A	there is an instance $A[T/X]$ that holds now
A, B	both A and B hold now
`hence` A	A holds at every instant after now
`always` A	same as (A, `hence` A)
`when(`c`)` A	same as `(if` c `then` A`, hence (if` c `then` A`))`
`unless(`c`)` A `else` B	same as `(if` c `then` B`, if` c `else` A`)`

Table 2. Combinators of `Hybrid cc`

`Hybrid cc` [9, 10] is an extension of `Default cc` over continuous time. First continuous constraint systems are allowed, i.e., constraints may involve differential equations that express initial value problems. Second, the `hence` operator is interpreted over continuous time. It imposes the constraints of A at every real time instant after the current one. The evolution of a system in `Hybrid cc` is piecewise continuous, with a sequence of alternating point and interval phases. All discrete changes take place in a point phase, where a simple `Default cc` program is executed. In a continuous phase, computation proceeds only through the evolution of time. The interval phase, whose duration is determined in the previous point phase, is exited as soon as the status of a conditional changes [10]. Tab. 2 summarizes the basic combinators of `Hybrid cc`.

`Hybrid cc` is well-suited for modeling dynamic biological systems, as argued in [2, 3].

3.2 Single-site model: local modeling

The single-site model from Sect. 2.3 with experimental values can be expressed directly in `Hybrid cc`.

```
interval t, rna, rnaI, rnaM;
t=0; rna = 0.06; rnaI = 0; rnaM = 0;
always{ t' = 200;
  rna'  = (Pr*R*rnaI)/(kr*(1+(SC/ksc))+R)
       -(Pese*(ASF+SC)*rna)/(kese+ASF+SC)-delta*rna;
  rnaI' = (Pese*(ASF+SC)*rna)/(kese+ASF+SC)
       -(Pr*R*rnaI)/(kr*(1+(SC/ksc))+R)-k*rnaI+kk*rnaM-delta*rnaI;
  rnaM' = k*rnaI-kk*rnaM-delta*rnaM;
}
sample(rna, rnaI, rnaM);
```

During the simulation, we obtain the predicted equilibrium for the splice efficiency, see Fig. 4. Under our hypotheses, which include protein competition and compensation, the model correctly simulates the alternative splicing activity. This supports the hypotheses made in the model such as the role of the ESE and ESS2 binding sites.

3.3 Two site model: global modeling

A realistic model of alternative splicing has to reflect the combinatorial complexity discussed in Sect. 2.1. Assuming that regulation is modular [12], the single-site model may be seen as one module inside a larger framework. The qualitative validation given in Sect. 2.4 justifies the introduction of the single-site model into a larger scale model involving several splicing sites. To illustrate this, we consider the generic example of two splice acceptor sites (SA) associated with one donor site (SD), see Fig. 5.

The behavior at one splicing site can be captured by a single function, the splice efficiency, which depends on the protein concentrations. This function is used in a larger-scale global model that describes the choice between two acceptor sites A3 and A4. In the HIV-1 case, the A3 site is the default splicing site. Only if the splice efficiency `eff1` gets smaller than `eff2`, regulation switches to the other state.

Recent work [5] shows the linearity of the splicing kinetics. Thus, on the larger scale, we may consider splicing as a linear process described by two ordinary

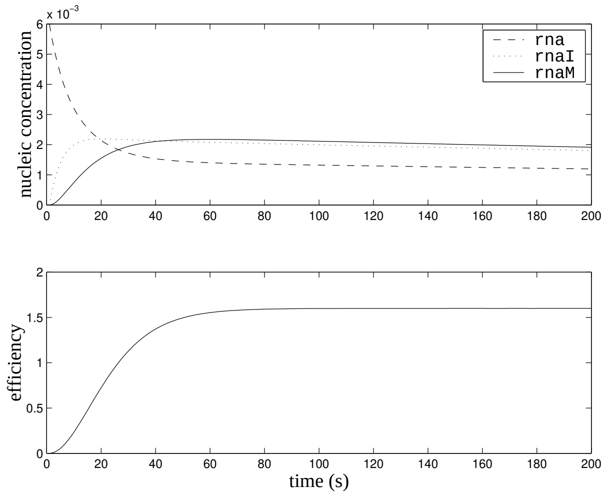

Fig. 4. Variation of the pool of RNA and the splice efficiency in the splicing reaction

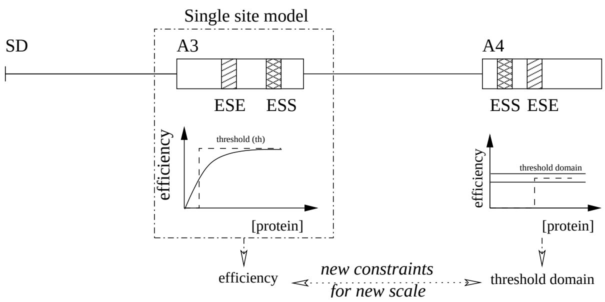

Fig. 5. Single-site model inside a general splicing regulation model

differential equations involving some kinetic constant. The first system represents the default behavior characterized by the constant Ka4. Following local changes on a single-site, the model may exhibit a different behavior, characterized by the kinetic constant Ka3. Thus, in this example, only the kinetic constants change, while the overall structure of the ODE system remains the same.

This default behavior can be naturally expressed in `Hybrid cc` using the combinator `unless`(c) A.

```
always {
    if (eff1 <= eff2)  {rna' = -Ka3*rna;
                        rnam1' = 0;
                        rnam2' = Ka3*rna;};
    unless ((eff1 <= eff2))  {rna' = - Ka4*rna;
                              rnam1' = Ka4*rna;
                              rnam2' = 0;};
}
```

Note that `unless`(c) A is not equivalent to `if` $\neg c$ `then` A. The second alternative will be chosen if the solver cannot verify that `(eff1 <= eff2)`. This may have *two* reasons:

- `(eff1 <= eff2)` is false, i.e. `(eff1 > eff2)`, or
- `(eff1 <= eff2)` is unknown (default behavior).

Simulation in `Hybrid cc` yields the behavior illustrated in Fig.6. We observe the switch from the first system of ODEs to the second when `eff2` passes the upper threshold for `eff1`.

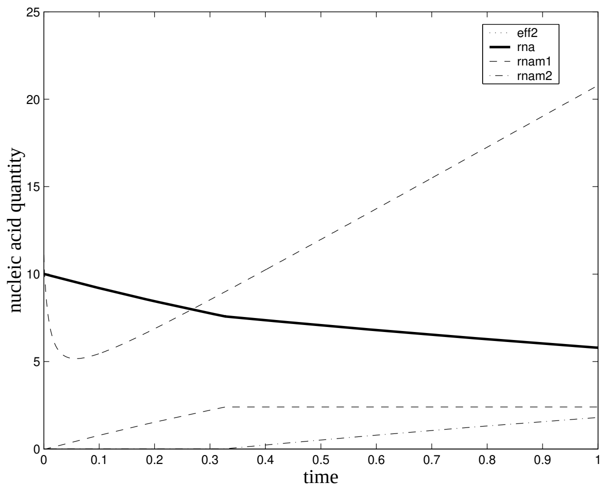

Fig. 6. Variation of RNA with a variation of splice efficiency

4 Conclusion and further research

Our approach combines mathematical and computational methods. Mathematical analysis allows us to validate the single-site model in a qualitative way. This is possible using the experimental data obtained in our group. The validation shows the consistency of our biological hypotheses. Based on this, we can extract the splice efficiency as a suitable abstraction of the local behavior at one site inside a more global model involving different sites. For the experimental biologist, the single-site model may serve as a computational tool to evaluate his knowledge on a fine-grained biological process.

On the computational side, the constraint solving and default reasoning capabilities of `Hybrid cc` allow us to exploit as much as possible the incomplete knowledge that is typical for ongoing biological research. Indeed, default behavior may compensate the lack of experimental data. Thus, using constraint programming, we can delimit with our model the possible splicing behavior. Similar to mathematical analysis, constraint reasoning therefore may provide a powerful qualitative validation.

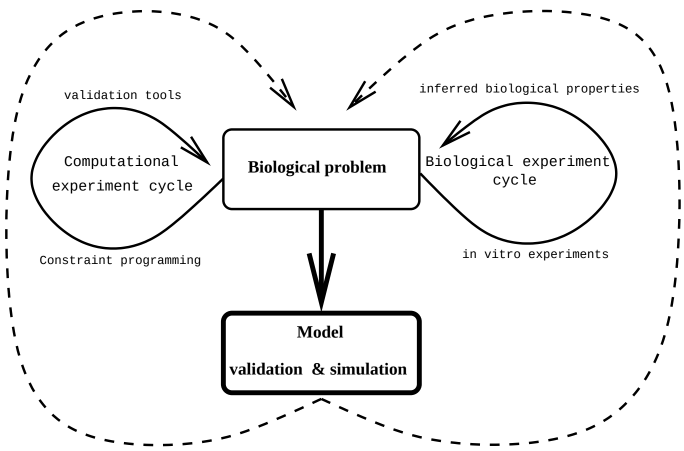

Fig. 7. Representation of ideal interactions for modeling biology

The combination of mathematical analysis and computational methods is also the key to the multiscale modeling developed in this paper. It leads to the qualitative validation represented by the extraction of the splice efficiency function. The splice efficiency characterizes the modularity of the regulation. Thus, the smaller-scale behavior is represented in the larger-scale model, based on the single-site splice efficiency. The extraction of a suitable criterion on the smaller scale is crucial to understanding an experimental process from a systems biology perspective. Furthermore, constraints can be used to handle the problem of missing data in a multiscale model. Different scales usually correspond to biological experiments yielding different types of results. Despite the variety of possible experiments, these must be integrated into a global model in order to better understand the biological process.

Multiscale modeling requires a close interaction between biological and computational approaches, as illustrated by Fig.7.

In the context of alternative splicing regulation, we are currently working on new experimental data for the quantitative validation of our models. On the computational side, we have integrated our model into a general HIV-1 model [11]. Preliminary results show that the modification of a splice constant may induce different behaviors in the HIV-1 life cycle model. Using the extended model, we may validate several biological hypotheses on the global effect of alternative splicing in the full HIV-1 life cycle.

References

1. Bernard, O., and Gouzé J.-L.: Transient behavior of biological models as a tool of qualitative validation - Application to the Droop model and to a N-P-Z model. J. Biol. Syst, 4(3) (1996) 303–314
2. Bockmayr, A., Courtois, A.: Modeling biological systems in hybrid concurrent constraint programming (Abstract). In 2nd Int. Conf. Systems Biology, ICSB'01, Pasadena, CA (2001)
3. Bockmayr, A., Courtois, A.: Using hybrid concurrent constraint programming to model dynamic biological systems. In 18th International Conference on Logic Programming, ICLP'02, Copenhagen Springer, LNCS 2401 (2002)
4. Caputi, M.A., Mayeda, M.A., Krainer, A.R., and Zahler, A.M.: hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. EMBO 18(14) (1999) 4060–7
5. Dautry, F., Weill, D.: Kinetic analysis of mRNA metabolism. In Interdisciplinary School on imaging, modelling and manipulating transcriptional regulatory networks, Ambleteuse (2002) 20/92
6. Del Gatto-Konczak, F., Olive, M., Gesnel, M.C., and Breathnach, R.: hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. Mol. Cell. Biol. 19 (1) (1999) 251–60
7. Graveley, B.R., Hertel, K.J., Maniatis, T.: A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. EMBO 17(22) (1998) 6747–6756
8. Graveley, B.R.: Sorting out the complexity of SR protein functions. RNA 6 (2000) 1197–1211
9. Gupta, V., Jagadeesan, R., Saraswat V.: Computing with continuous change. Science of computer programming 30(1-2) (1998) 3–49
10. Gupta, V., Jagadeesan, R., Saraswat V., Bobrow, D.G.: Programming in hybrid constraint languages. In *Hybrid Systems II*, 226–251. Springer, LNCS 999 (1995)
11. Hammond, B.J.: Quantitative Study of the Control of HIV-1 Gene Expression. J. Theor. Biol. 163 (1993) 199–221
12. Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W.: From molecular to modular cell biology. Nature 402 (1999) C47–C52

13. Heinrich, R., Schuster, S.: *The regulation of cellular systems.* Int'l Thomson Publishing, New York (1996)
14. Moore, M.J., Query, C.C., Sharp, P.A.: Splicing of precursors to mRNA by the spliceosome. In *The RNA World*, Cold Spring Harbor Laboratory Press (1993) 303–357
15. O'Reilly, M., McNally, M.T., Beemon, K.L.: Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. Virology 213(2) (1995) 373–85
16. Palsson, B.: The challenges of in silico biology. Nature Biotechnology 18 (2000) 1147 – 1150
17. Purcell, D.F., Martin, M.A.: Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J. Virol. 67(11)(1993) 6365–6378
18. Ropers, D., Ayadi, L., Jacquenet, S., Méreau, A., Thomas, D., Mougin, A., Bilodeau, P., Stoltzfus, C.M., Gattoni, R., Stévenin, J., Branlant, C.: A complex regulation of the central A3 to A5 acceptor sites in HIV-1 RNA. In Eukaryotic mRNA Processing (2001)
19. Saraswat, V.A.: *Concurrent constraint programming.* ACM Doctoral Dissertation Awards. MIT Press (1993)
20. Saraswat, V.A., Jagadeesan, R., Gupta, V.: Foundations of timed concurrent constraint programming. In 9th Symp. Logic in Computer Science, LICS'94, Paris IEEE (1994) 71 – 80
21. Saraswat, V.A., Jagadeesan, R., Gupta, V.: Timed default concurrent constraint programming. Journal of Symbolic Computation 22(5/6) (1996) 475–520
22. Segel, L.A.: *Modelling dynamic phenomena in molecular and cellular biology.* Cambridge University Press (1984)
23. Si, Z.H., Amendt, B.A., Stoltzfus, C.M.: Splicing efficiency of human immunodeficiency virus type 1 tat RNA is determined by both a suboptimal 3' splice site and a 10 nucleotide exon splicing silencer element located within tat exon 2. N. A. R. 25(4) (1997) 861–867
24. Si, Z.H., Rauch, D., Stoltzfus, C.M.: The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. Mol. Cell. Biol 18(9) (1998) 5404–13
25. Staffa, A., Cochrane, A.: The tat/rev intron of human immunodeficiency virus type 1 is inefficiently spliced because of suboptimal signals in the 3' splice site. J. Virol. 68(5) (1994) 3071–9
26. Tang, H., Kuhen, K.L., Wong-Staal, F.: Lentivirus replication and regulation. Annu. Rev. Genet 33 (1999) 133–170
27. Voit, E.: *Computational Analysis of Biochemical Systems.* Cambridge University Press (2000)