Rule-based Modeling of Transcriptional Attenuation at the Tryptophan Operon

Céline Kuttler, Cédric Lhoussaine, Mirabelle Nebut[1,2]

[1] University of Lille 1
[2] BioComputing, LIFL (CNRS UMR8022) & IRI (CNRS USR3078)

Abstract. Transcriptional attenuation at E.coli's tryptophan operon is a prime example of RNA-mediated gene regulation. In this paper, we present a discrete stochastic model for this phenomenon based on chemical reactions. Our model is compact and intelligible, due to n-ary reactions (which preclude object-centric approaches) and to rule schemas that define finite sets of chemical reactions. Stochastic simulations with our model confirm results that were previously obtained by master equations or differential equations. In addition, our approach permits to reflect mutation experiments by simple model modifications, and to re-use model components for transcriptional attenuation in other genes and organisms.

Keywords. Systems biology, riboswitch, rule-based modeling languages, stochastic simulation, kappa.

1 Introduction

Transcriptional attenuation is a control mechanism in gene expression, that down-regulates transcription as a response to the increased speed of translation of regulatory codons, via alternative RNA conformations. Although it has been investigated within bacterial systems since the 1970 [1,2], the phenomenon attracted significantly less interest than proteins which control the first step in gene expression through DNA binding. This dramatically changed in the 2000s, with the discovery of crucial mechanisms in higher organisms that exploit properties of RNA. Quantitative investigations of regulation via RNA gained momentum for therapeutic approaches and synthetic biology [3,4,5].

E.coli's tryptophan operon is the best understood biosynthetic operon. It consists in a set of co-regulated genes through which the bacterium synthesizes the amino acid Trp, unless sufficient amounts of it are provided by the environment. The mechanism appears with variations in other bacteria and other biosynthetic operons [6,7,8].

Santillan and Zeron [9] modeled all three regulatory mechanisms of the trp operon through delay differential equations (DDE), without investigating attenuation in detail. DDEs are usually directly derived from informal and simple biochemical reactions. The main drawback with such deterministic models is that they permit only observation of average behaviours. In particular, they do

Dagstuhl Seminar Proceedings 09091
Formal Methods in Molecular Biology
http://drops.dagstuhl.de/opus/volltexte/2009/1993

not account for possible stochastic noise from which multi-modal states may arise. In that case, the average behaviour does not correspond to any of the actual states. Since regulatory systems involve few biological entities, a criterion known to increase stochastic effects, one may wonder if the deterministic assumption for modeling E.coli's trp operon is suitable. This is still an open question that calls for stochastic modeling.

We consider a stochastic model of E.coli's trp operon, adopting the same conceptional level as Elf and Ehrenberg [10], who analyzed the sensitivity of Trp attenuation through probability functions and, more generally, discrete master equations. Such an approach benefits from a rich probability theory that gives valuable insights and measurement capabilities of stochasticity. However, apart for very simple and rare regulatory systems, master equations can only be evaluated numerically rather than being solved symbolically. Moreover, each biological system requires an ad-hoc master equation or probability function that is usually hard to design from the mechanistic intuition of the system.

Discrete event models for stochastic simulation are most often described by chemical reactions which can be studied thanks to formal rule-based modeling languages [11,12,13,14]. Molecular systems are understood as multisets of molecules that are rewritten by applying chemical reactions. The reaction speed is specified by a rate constant. The stochastic semantics of chemical reactions is given in terms of continuous time Markov chains (CTMCs). Gillespie's algorithm [15,16] permits direct stochastic simulations from a given initial molecular system and a set of chemical reactions. Rule-based models are intuitive in the sense that they describe molecular interactions and, to some extend, are simpler to variate and extend than models based on classical mathematical functions.

In this work, we present the first discrete event model for the stochastic simulation of attenuation at the tryptophan operon. Our model consists of 71 chemical reactions, that are generated from 13 rule schemas. These faithfully summarize the rich narrative account in the biological literature [17,18,19]. This succinct and concise description is obtained by two ingredients, rules schemas (introduced in Section 3) and n-ary chemical reactions. Rule schemas allow to represent finite sets of chemical reactions in a compact manner, which differ only in the choice of some molecule parameters e.g. folding or phosphorilation state, location, etc. The same idea is well known from logic programming [20], unification grammars [21] or term rewriting [22].

Based on our model, we analyze the quantitative dynamics of our model within the Kappa Factory[1] [23] and Dizzy [16]. Without any probability analysis, we are able to reproduce and confirm results of Elf and Ehrenberg [10] regarding the probability of transcription as a function of the rate of trp-codon translation including the full attenuation mechanism (Section 4). We also analyze the impact of re-arrangements within the trp operon's codon region, by straightforward variations of the rule set. This flexibility is of great benefit. It allowed us to cover the transcriptional attenuator at *Salmonella typhimurium*'s histidine operon [1,24] in addition, for instance. Regarding the latter, we re-

[1] See also Kappa's web-based successor at `http://www.cellucidate.com`

strict our discussion to the model's adaptation, its quantitative analysis remains beyond the scope of the current paper.

We present a second model with stochastic simulations, that confirms the hypersensitivity of competing multi-step processes in attenuation, independently of mechanistic regulatory details (Section 5). Beyond of this, we investigate hypersensitivity related to concurrent mechanisms that our first model exhibits and that are hard, if not impossible, to cover via master equations.

Related Work. All models presented in this paper can be compiled to finite collections of chemical reactions. These are supported by rule-based modeling languages of low expressiveness [25], that are based on multi-set rewriting such as BioCham [11]. Furthermore, such collections of reactions are supported by standard tools for stochastic simulation such as Dizzy [16].

Alternative rule-based languages with higher expressiveness such as the graph rewrite language Kappa [26], BioNetGen [27,28], and Milner's Bigraphs [13] are Turing-complete. Their pattern based graph rewrite rules are like schemas, in that they can be applied to arbitrary subgraphs satisfying the pattern. In contrast to our approach, such schemas may describe infinitely many reactions. Furthermore, stochastic simulation is possible without inferring all those reactions on before hand. This generation process in uncritical in the present paper, since the overall number of reactions remains small, but is the bottleneck in other applications [23], where it grows exponentially. Generally, the price for high expressiveness is the limitation in current verification tools.

It has been argued in [25,29] that binary reactions are sufficient to represent chemical knowledge. Nevertheless, n-ary reaction are a critical factor for our models. Rewriting n-ary to binary reactions is tedious and requires sufficient expressiveness. It is possible in language where sequences of reactions can be executed within atomic transactions, so that no other interactions may intervene. This excludes object-centered modeling languages from process algebra [30] for the present case study, namely stochastic π-calculus [31,32,33] or Beta-Binders [34].

The first stochastic treatment of attenuation at the trp operon dates back to the work of Heijne et al in 1977 [35], see [10] for a discussion with respect to the coverage of later findings. Thieffry and co-authors presented a Petri net model of tryptophan biosynthesis in E.coli, however without covering the attenuation mechanism [36].

2 Transcriptional Attenuation

In this section, we recall the process of ribosome-mediated transcriptional attenuation as present in many organisms, and then review the particular case of E.coli's tryptophan (trp) operon [2,19,37] such that it becomes amenable to mechanistic modeling.

The role of transcriptional attenuation is to prematurely interrupt the transcription of a DNA sequence, in order to avoid needless protein production. In the

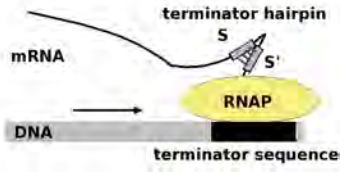

Fig. 1. Transcription termination: a hairpin forms in the most recent chunk of RNA while RNAP is paused on the following terminator DNA sequence

case of E.coli's trp operon, these proteins are enzymes for tryptophan biosynthesis. Their activity comes at a high energetic cost for the cell, and should thus be suppressed whenever sufficient amounts of Trp are available in the cell's environment. Two further mechanisms complement attenuation: repression of transcription initiation and inhibition of the enzymatic chain that yields Trp. They are not in the scope of the present paper.

Transcriptional attenuation is best illustrated as a *race* between transcription and translation: the premature abortion of transcription shortly after initiation, depends on the speed at which a ribosome concurrently translates the nascent transcript. In the case of E.coli's trp operon, the ribosome's speed is high if tryptophan supply is high, and low otherwise. We next review the control of the race and premature termination of transcription in detail.[2]

Transcription initiates with the binding of the enzyme RNAP to a dedicated start sequence on DNA, from where RNAP starts assembling an mRNA sequence. During elongation, RNAP advances stepwise over DNA with an average speed of 50 nucleotides per second. In each step, it adds one new nucleotide to the growing mRNA sequence. The leader of the operon, which constitutes the first 150 of several thousand nucleotides, contains a *termination sequence*, on which early termination may happen, so that RNAP releases the truncated transcript already there. A second condition must hold as sketched in Fig. 1. Two segments S and S' in the most recent portion of mRNA must fold into a *terminator hairpin*.[3]

Translation is initiated by the binding of a ribosome to the free end of an mRNA transcript, the other end of which is still being elongated by RNAP. The ribosome reads words of three mRNA bases called *codons*, for each of which it adds one amino acid to a growing sequence. Ribosomes are not affected by mRNA hairpins, but simply disrupt these. While the average rate of translation is 15 codons per second, each step of the ribosome is actually limited by the

[2] The opposite of premature termination is that RNAP continues transcription beyond the mRNA leader into to so-called structural genes, which encode the enzymes for Trp synthesis, and stops after this.

[3] An alternative termination mechanism in bacteria is a small molecule that slides along mRNA and destabilizes RNAP.

Fig. 2. Control region of the trp mRNA. Adjacent pairs of the four segments can fold into alternative hairpins. The hairpin S_3:S_4 terminates transcription of the trp operon prematurely, while the anti-terminator S_2:S_3 allows transcription to continue into the protein-coding region.

abundance of the currently required amino acid. The ribosome slows down on codons for which the corresponding amino acid is in short supply in the cell.

Fig. 2 illustrates the leader segment of mRNA from E.coli's trp operon. It shows a sequence of codons, numbered increasingly from left to right starting at 0. Translating ribosomes start on the left, proceed to the right, and dissociate at codon 15 (stop signal). Codons 10 and 11 are particular, here the ribosome requires tryptophan to proceed (other codons encode less critical amino acids). The mRNA furthermore comprises distinguished segments S_1, S_2, S_3, and S_4, neighboring pairs can form hairpins. Hairpin S_1:S_2 is called (*pause*), S_2:S_3 *anti-terminator*, and S_3:S_4 *terminator*. Note that every segment can at most participate in one hairpin at the same time. Hence, the anti-terminator prohibits the formation of both the terminator and the pause loop.

Hairpins form instantaneously[4], whenever both involved segments are present *and* not covered by the ribosome's footprint. As a consequence, the pause hairpin appears first. This blocks RNAP, which is relased only once a ribosome arrives and disrupts the pause hairpin. Note that the ribosome is quite big, and covers several codons at the same time. It melts the pause loop, already when reaching the *release* codon 7.

The melting of the pause hairpin unblocks RNAP stalled at the nucleotide that we call DNA_0. This configuration starts the attenuation race and is illustrated in the left of Fig. 3, where RNAP has transcribed the leader mRNA up to and including S_2. Both RNAP and the ribosome are now synchronized for the race that has two possible outcomes. If the ribosome is slow, the *anti-terminator* hairpin forms, and RNAP continues transcription beyond the termination sequence (nucleotide DNA_{50}). The right-hand side of Fig. 3 shows this readthrough configuration in the upper part. The lower part shows the alternative conformation with *terminator* hairpin that is typically reached when the ribosome is fast.

Here are some more details. Segment S_3 becomes available as RNAP reaches nucleotide DNA_{36}, and S_4 from nucleotide DNA_{47} on. A ribosome masks segment S_2 upon reaching codon 13, for which it must have translated codons 10 and 11 that encode Trp. If translation slows down or stalls because Trp is in short supply, S_2 remains available for formation of the anti-terminator loop. If translation is fast and the ribosome reaches the stop codon 15, it remains there

[4] i.e. on a faster timescale than any of the other reactions in the system.

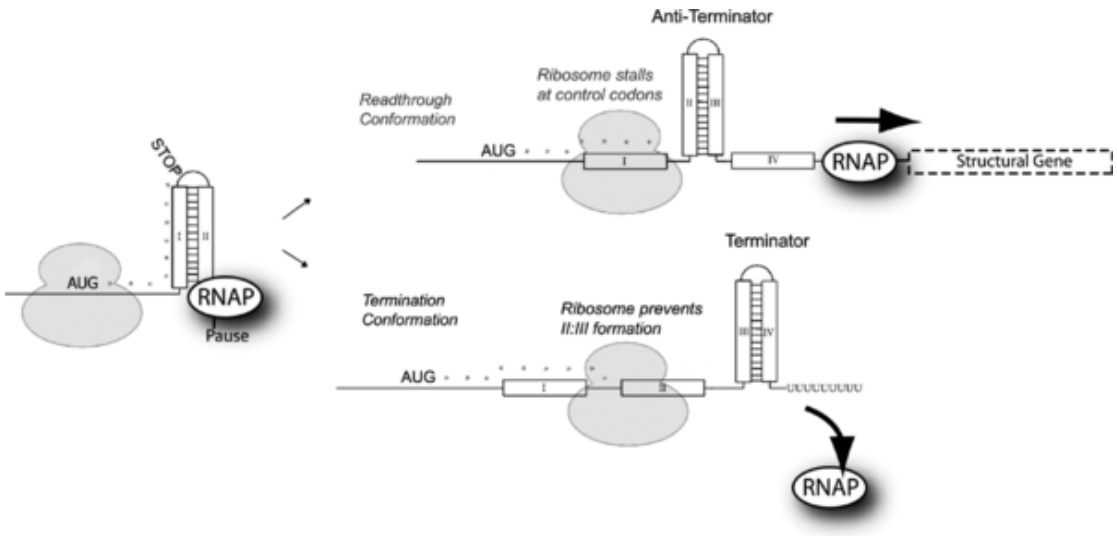

Fig. 3. Starting point and possible outcomes of the race at *E.coli* s trp operon. Left: RNAP is stalled by the pause hairpin, will be released by the ribosome's next step. Upper right: when Trp supply is low, the ribosome stalls at the control codons 10 and 11, the anti-terminator hairpin forms and transcription continues into the operon. Lower right: when Trp supply is high, the ribosome rapidly translates over the control codons. Before it unbinds from the stop codon, the terminator hairpin forms and transcription aborts. Figure reproduced with permission from [10].

for approximately one second before it unbinds, which is a long time scale in the system. Thus, the terminator hairpin forms while the ribosome masks S_2. Note that the ribosome occasionally dissociates from the stop codon after S_3 has been translated, but before S_4 is available. In this case, the first three segments are available for the alternative formation of the pause or anti-terminator hairpin. Both require segment S_2. They are mutually exclusive, and have equal probabilities.

3 Rule Schemas for Chemical Reactions

We define chemical reactions, that operate on multisets of complex molecules with attributes such as RNAP · DNA(23). The attribute value 23 here is the location of the DNA nucleotide, to which RNAP is bound. Other attributes of interest can be the compartment of a molecule, or information on its states, for instance folding or phosphorilation.

We then present a language of rules schemas allowing to define finite sets of chemical reactions in a compact manner. Rule schemas are like chemical reactions, except that attribute values are now extended to expressions with vari-

ables. Complex molecules are thus described by terms such as $\textsc{Rnap}\cdot\textsc{Dna}(x+1)$ where x is a variable with values in $\{1,\ldots,51\}$. All variables are universally quantified over finite sets, such that they define finite sets of reactions.

3.1 Chemical Reactions

In order to define the syntax of attributed molecules, we fix a possibly infinite set of attribute values $\mathcal{C}$ and a finite set $\mathcal{N}$ of molecule names. We assume that every molecule name $N \in \mathcal{N}$ has a fixed arity $ar(N) \geq 0$ which specifies the number of attributes of N.

A *molecule* M is a complex of attributed molecules with the following abstract syntax where $N \in \mathcal{N}$ and $c_1,\ldots,c_n \in \mathcal{C}$:

$$M \in \mathit{Mol} \ ::= \ N(c_1,\ldots,c_n) \mid M_1 \cdot M_2 \qquad \text{where } ar(N) = n$$

We write $M_1 \cdot M_2$ for the complex of M_1 and M_2. For instance, if $\textsc{Rnap}, \textsc{Dna} \in \mathcal{N}$ and $47 \in \mathcal{C}$ then $\textsc{Rnap} \cdot \textsc{Dna}(47)$ is molecule complex consisting of an Rnap that is bound to the Dna nucleotide at position 47. A *chemical solution* S is a sum of molecules:

$$S \in \mathit{Sol} \ ::= \ M \mid S_1 + S_2$$

A *chemical reaction* is a rule of the form

$$S_1 \rightarrow_k S_2$$

where S_1 and S_2 are solutions and $k \in \mathbb{R}^+ \cup \{\infty\}$ is a stochastic rate. For instance, the following reaction states that an Rnap bound to the Dna nucleotide at position 23 may move forwards to the Dna nucleotide at position 24. The speed of this reaction is 50 sec^{-1}:

$$\textsc{Rnap} \cdot \textsc{Dna}(23) + \textsc{Dna}(24) \ \rightarrow_{50} \ \textsc{Dna}(23) + \textsc{Rnap} \cdot \textsc{Dna}(24)$$

Of course, one needs many similar rules for the many other Dna nucleotides with different positions. We next introduce rule schemas, by which to define such sets of chemical reactions in a compact manner.

3.2 Rule Schemas

In order to define rule schemas for chemical reactions, we need variables x for attribute values and expressions such as $x+1$, in order to compute corresponding attribute values. This permits to generalize the above chemical reaction to the following rule schema:

$$\forall x \in \{1,\ldots,49\}.\ \textsc{Rnap}\cdot\textsc{Dna}(x)+\textsc{Dna}(x+1) \ \rightarrow_{50} \ \textsc{Dna}(x)+\textsc{Rnap}\cdot\textsc{Dna}(x+1)$$

We thus need a set $\mathcal{V}$ of variables that are ranged over by x, and a finite set $\mathcal{F}$ of function symbols $f \in \mathcal{F}$ with arities $ar(f) \geq 0$. Furthermore, we assume an

interpretation $[\![f]\!] : \mathcal{C}^{ar(f)} \to \mathcal{C}$ for every $f \in \mathcal{F}$. An *expression* e for values in $\mathcal{C}$ is a term e with the following abstract syntax, where $x \in \mathcal{V}$, $c \in \mathcal{C}$, and $f \in \mathcal{F}$:

$$e \in \mathit{Exp} \ ::= \ x \mid c \mid f(e_1, \ldots, e_n)$$

In our modeling case studies, we will assume that symbol $+ \in \mathcal{F}$ of arity 2 is interpreted as addition on natural numbers. We freely use infix syntax as usual, i.e., we write $e_1 + e_2$ instead of $+(e_1, e_2)$. Given a variable assignment $\alpha : \mathcal{V} \to \mathcal{C}$, every expression $e \in \mathit{Exp}$ denotes an element $[\![e]\!]_\alpha \in \mathcal{C}$ that we define as follows:

$$\begin{aligned}
&[\![c]\!]_\alpha = c \\
&[\![x]\!]_\alpha = \alpha(x) \\
&[\![f(e_1, \ldots, e_n)]\!]_\alpha = [\![f]\!]([\![e_1]\!]_\alpha, \ldots, [\![e_n]\!]_\alpha)
\end{aligned}$$

A *schematic molecule* M is like a molecule, except we now allow for expressions in attribute positions rather than attribute values only:

$$M \in \mathit{SMol} \ ::= \ N(e_1, \ldots, e_n) \mid M_1 \cdot M_2 \qquad \text{where } ar(N) = n$$

Here, we assume that $e_1, \ldots, e_n \in \mathit{Exp}$ and $N \in \mathcal{N}$. A *schematic solution* $S \in \mathit{SSol}$ is sum of schematic molecules $M \in \mathit{SMol}$:

$$S \in \mathit{SSol} \ ::= \ M \mid S_1 + S_2$$

As usual, we write $\mathcal{V}(S)$ for the set of variables that occur in molecules of S. A *rule schema* R has the form:

$$\forall x_1 \in D_1 \ldots \forall x_n \in D_n.\ S_1 \to_k S_2$$

where D_1, …, $D_n \subseteq \mathcal{C}$ are finite sets, $S_1, S_2 \in \mathit{SSol}$ with $\mathcal{V}(S_1) \cup \mathcal{V}(S_2) \subseteq \{x_1, \ldots, x_n\}$, and $k \in \mathbb{R}^+ \cup \{\infty\}$. For every variable assignment $\alpha : \mathcal{V} \to \mathcal{C}$ that maps variables to values in their domain, we can instantiate the rule schema to finitely many reactions. A schematic molecule M is mapped to a molecule $[\![M]\!]_\alpha \in \mathit{Mol}$:

$$\begin{aligned}
&[\![N(e_1, \ldots, e_n)]\!]_\alpha = N([\![e_1]\!]_\alpha, \ldots, [\![e_n]\!]_\alpha) \\
&[\![M_1 \cdot M_2]\!]_\alpha = [\![M_1]\!]_\alpha \cdot [\![M_2]\!]_\alpha
\end{aligned}$$

Similarly, schematic solutions $S \in \mathit{SSol}$ get instantiated to solutions $[\![S]\!]_\alpha \in \mathit{Sol}$:

$$[\![S_1 + S_2]\!]_\alpha = \{S_1' + S_2' \mid S_1' \in [\![S_1]\!]_\alpha, S_2' \in [\![S_2]\!]_\alpha\}$$

A rule schema R of the above form is instantiated to a set of chemical reactions, by enumerating the chemical reactions for all variables assignments licenced by the quantifiers:

$$[\![R]\!] = \{[\![S_1]\!]_\alpha \to_r [\![S_2]\!]_\alpha \mid \alpha : \mathcal{V} \to \mathcal{C},\ \alpha(x_1) \in D_1, \ldots, \alpha(x_n) \in D_n\}$$

3.3 Stochastic Semantics and Simulation

For the sake of completeness, we recall the stochastic semantics of chemical reactions and how to use them for stochastic semantics with Gillespie's algorithm. This underlines that our biological modeling case studies are indeed expressed in a formal modeling language.

The semantics of a set of chemical reactions is a continuous time Markov chain (CTMC). We restrict ourselves to the case without infinite rates $k = \infty$. Note that chemical reactions with infinite rates are always to be executed with priority and without time delay.

The states of the CTMCs are congruence classes $[S]_{\equiv}$ of chemical solutions S with respect to the least congruence relation $\equiv$ that makes complexation and summation associative and commutative:

$$\begin{array}{ll} M_1 \cdot M_2 \equiv M_2 \cdot M_1 & (M_1 \cdot M_2) + M_3 \equiv M_1 \cdot (M_2 \cdot M_3) \\ S_1 + S_2 \equiv S_2 + S_1 & (S_1 + S_2) + S_3 \equiv S_1 + (S_2 + S_3) \end{array}$$

We introduce transitions $S \xrightarrow[L]{k} S'$ stating that S can be reduced to S' by applying a chemical reaction with rate $k \in \mathbb{R}^+$ to the subset of molecules in S with positions in L. Positions are the indexes in sums such as $M_1 + \ldots + M_n$ that we also write as $\sum_{i=1}^{n} M_i$.

$$\frac{L \subseteq \{1, \ldots, n\} \qquad \sum_{i \in L} M_i \equiv S \qquad S \rightarrow_k S'}{\sum_{i=1}^{n} M_i \xrightarrow[L]{k} S' + \sum_{i \notin L} M_i}$$

We next introduce transitions $S \xrightarrow{r} S'$, where r sums up all rates of chemical reactions reducing S to S', as many times as they apply for some index set L.

$$\frac{r = \sum_{\{(L,k) \mid S \xrightarrow[L]{k} S_1 \equiv S'\}} k}{S \xrightarrow{r} S'}$$

Such translations are invariant under structural congruence, i.e. for all $S \equiv S_1$ and $S' \equiv S_1'$ it holds that $S \xrightarrow{r} S'$ if and only if $S_1 \xrightarrow{r} S_1'$. We can thus define $[S]_{\equiv} \xrightarrow{r} [S']_{\equiv}$ by $S \xrightarrow{r} S'$ as the transitions of the CTMC.

Gillespie's algorithm for stochastic simulation takes as input a finite set of chemical reactions and a chemical solution S. It then computes the overall rate of all possible transitions $R = \sum_{\{(r, [S_1]_{\equiv}) \mid S \xrightarrow{r} S_1\}} r$, selects with probability r/R a congruence class $[S_1]_{\equiv}$ with transition $S \xrightarrow{r} S_1$, and returns an arbitrary representative of this congruence class $S' \equiv S_1$ jointly with a time delay drawn randomly from the exponential distribution with rate r.

4 Modeling Transcriptional Attenuation

In this section, we present our model of ribosome-mediated attenuation of transcription at *E.coli*'s tryptophan operon. It explicits the multi-step race between

transcribing RNAP and translating ribosome, the formation of mRNA hairpins, and the translating ribosome's impact on these latter. We validate our quantitative results against those of Elf and Ehrenberg [10], and investigate the impact of structural and parameter changes to transcription attenuation at the trp operon. Finally, we discuss the adaption of our rule schemas to the his operon of *Salmonella typhimurium* [24], which underlines the flexibility of our approach.

4.1 Rules Schemas

Notation. We will use molecule names $\mathcal{N} = \{\text{RNAP}, m\text{RNA}, \text{DNA}, \text{Ribosome}, \text{S}\}$, attribute values $\mathcal{C} = \mathbb{N}_0 \cup \{\text{fr}, \text{hp}, \text{bl}\}$, function symbols $\mathcal{F} = \{+\}$, and variables $\mathcal{V} = \{i, n, m, t\}$. Since attributed molecules use only few arguments, we can denote them as upper and lower indexes, for instance, write $\text{S}^{\text{fr}}_{i+1}$ instead of $\text{S}(\text{fr}, i+1)$.

We represent the *initial solution* for the race between RNAP and the ribosome at E.coli's tryptophan operon (illustrated in Fig. 3 on page 6) as follows:

$$\begin{aligned} &\text{Ribosome} \cdot \text{mRNA}_6 + \text{mRNA}_7 + \text{mRNA}_8 + \text{mRNA}_9 + \text{mRNA}_{15} \\ &\quad + \text{S}^{\text{hp}}_1 \cdot \text{S}^{\text{hp}}_2 + \text{RNAP} \cdot \text{DNA}_0 + \text{DNA}_1 + \ldots + \text{DNA}_{50} \end{aligned} \tag{14}$$

RNAP is stalled on DNA_0 by the pause hairpin that has formed in the mRNA behind it from S_1 and S_2. The ribosome is positioned on mRNA_6, its step onward to mRNA_7 will melt the pause hairpin. The initial configuration given in solution 15 further includes the nucleotides DNA_1 to DNA_{50}, and the mRNA codons 8, 9 and 15 (the stop codon). Note that mRNA segments bear state labels fr (free), hp (hairpin), and bl (blocked), that switch upon hairpin formation. Reaction scheme 1 in Table 1 applies to indices $i \in \{1, 2, 3\}$, thus generating the reactions for the pause, anti-terminator and terminator hairpin.

Let us go through the reaction rules for transcription, equations 2-7 in Table 1. Rule scheme 2 represents RNAP's transition from one DNA nucleotide to the next, that occurs at a pace of e_1 of 50 nucleotides per second. It applies to all DNA positions with the following exceptions. Transcription *resumes* once the pause hairpin is disrupted, see the dependence on S_2's state *open* in reaction 3. As RNAP reaches DNA_{36}, the RNA segment S_3 is spawn, and S_4 at DNA_{47} (equation 4,5). Transcription terminates in the case of terminator hairpin formation, for which reaction 6 checks the presence of S^{hp}_4. Only if S_4 remains open, transcription proceeds to DNA_{48} (equation 7).

For translation, we again distinguish between a rule scheme for the majority of translation steps without side effects at rate $e_2 = 15s^{-1}$ in equation 8, and introduce specialized reaction where necessary. In our simulations, we will vary the elongation rate e_3 in reaction 9 within $]0, 15]s^{-1}$ to reflect the ribosome's slowing down. Reaction 10 for the translation step in which the ribosome disrupts the pause hairpin is n-ary: S_2 switches from the hairpinned state to fr, while S_1 expands from the abstraction level of a whole segment into the mRNA codons that constitute it. Reaction 11 triggers as the ribosome reaches DNA_{13} and its

$S_i^{fr} + S_{i+1}^{fr}$	$\rightarrow_\infty S_i^{hp} \cdot S_{i+1}^{hp} \quad \forall i \in \{1,2,3\}$	(1)
$RNAP \cdot DNA_n + DNA_{n+1}$	$\rightarrow_{e_1} DNA_n + RNAP \cdot DNA_{n+1}$ $\forall n \in \{1,\ldots,49\} \setminus \{35,46,47\}$	(2)
$RNAP \cdot DNA_0 + DNA_1 + S_2^{fr}$	$\rightarrow_{e_1} DNA_0 + RNAP \cdot DNA_1 + S_2^{fr}$	(3)
$RNAP \cdot DNA_{35} + DNA_{36}$	$\rightarrow_{e_1} DNA_{35} + RNAP \cdot DNA_{36} + S_3^{fr}$	(4)
$RNAP \cdot DNA_{46} + DNA_{47}$	$\rightarrow_{e_1} DNA_{46} + RNAP \cdot DNA_{47} + S_4^{fr}$	(5)
$RNAP \cdot DNA_{47} + DNA_{48} + S_4^{hp}$	$\rightarrow_{e_1} DNA_{47} + RNAP + DNA_{48} + S_4^{hp}$	(6)
$RNAP \cdot DNA_{47} + DNA_{48} + S_4^{fr}$	$\rightarrow_{e_1} DNA_{48} + RNAP \cdot DNA_{48} + S_4^{fr}$	(7)
$Ribosome \cdot mRNA_m + mRNA_{m+1}$	$\rightarrow_{e_2} mRNA_m + Ribosome \cdot mRNA_{m+1}$ $\forall m \in \{7,8,9,12,13,14\}$	(8)
$Ribosome \cdot mRNA_t + mRNA_{t+1}$	$\rightarrow_{e_3} mRNA_t + Ribosome \cdot mRNA_{t+1}$ $\forall t \in \{10,11\}$	(9)
$Ribosome \cdot mRNA_6 + mRNA_7$ $+S_1^{hp} \cdot S_2^{hp}$	$\rightarrow_{e_2} mRNA_6 + Ribosome \cdot mRNA_7$ $+ \sum_{i=10}^{14} mRNA_i + S_2^{fr}$	(10)
$Ribosome \cdot mRNA_{13} + S_2^{fr}$	$\rightarrow_\infty Ribosome \cdot mRNA_{13} + S_2^{bl}$	(11)
$Ribosome \cdot mRNA_{15} + \sum_{i=10}^{14} mRNA_i$	$\rightarrow_d Ribosome + mRNA_{15} + S_1^{fr}$	(12)
$Ribosome + S_2^{bl}$	$\rightarrow_\infty Ribosome \cdot mRNA_{13} + S_2^{fr}$	(13)

Table 1. Rule schemas for hairpin formation (1), transcription (2-7) with rate constant $e_1 = 50s^{-1}$, translation of non-trp codons with $e_2 = 15s^{-1}$ (8), for trp with e_3 variable in 9, masking S_2 and back (11, 13), expanding S_1 into individual mRNA codons (10), and back during ribosome release with $d = 1.0s^{-1}$ (12).

footprint covers S_2^{fr} , which remains in its blocked state and unavailable for hairpin formation. This changes after the ribosome dissociated from the stop codon $mRNA_{15}$, which also makes S_1 re-available for hairpin formation (equations 12 and 13).

4.2 Simulation

Fig. 4 shows the dependency of relative transcription frequency on e_3, the rate of trp-codon translation.[5] Two distinct pathways can lead to formation of the

[5] We varied e_3 from 0 to 15 s^{-1} in steps of one, performed 5000 simulations per value within the kappa factory, and extracted the statistics reported via kappa *stories*.

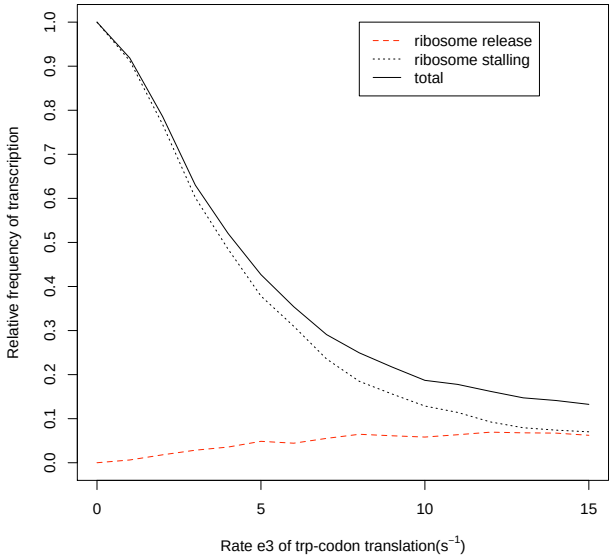

Fig. 4. Relative frequency of transcription as a function of the rate of the Trp codon translation in the leader mRNA, distinguishing between anti-terminator formation during ribosome stalling, and after ribosome release from the stop codon.

terminator loop. Formation of the anti-terminator is most frequent during *ribosome stalling* on the trp-codons 10 or 11, and becomes rarer when increasing the rate of trp-codon translation.

The second pathway to anti-terminator loop formation is *ribosome release*, which may occur between formation of the third and fourth mRNA segment if trp-codon translation is efficient. It thus shows the opposite dependence on the rate of trp-codon translation. It is responsible for the high basal level of transcription into the genes that encode the enzymes for Trp synthesis, despite high Trp level in E.coli's environment.

Our results on *ribosome release* confirm those of Elf and Ehrenberg [10], while those on *ribosome stalling* drop less significantly than theirs. However, our experiments predict a rate of basal transcription of 13% when trp-codon translation is efficient, which is within the experimental estimate $(10 - 15\%)$ [38], where Elf and Ehrenberg predicted 8%.

Increasing basal read-through level by structural changes to the level. We investigated the basal read-through level's dependency on structural changes to our model. For the data shown in Fig. 5, we advanced the position where RNAP

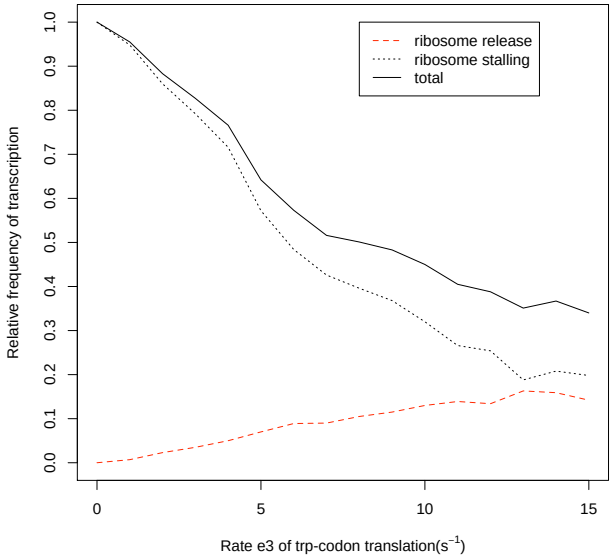

Fig. 5. Relative frequency of transcription as a function of the rate of the Trp codons in the leader mRNA, with two modifications to the model: S_3 spawn 8 codons earlier. Pause hairpin formation after ribosome release disabled.

spawns the mRNA segment S_3 by 8, and disabled formation of the pause hairpin after ribosome release from the stop codon. That is, we applied reaction 4 to DNA_{28}, reaction 2 to DNA_{36}, and reaction schema 1 to all $i \in \{2, 3\}$ instead of previously $\{1, 3\}$. The effect is a doubling in the chance for anti-terminator formation, compared to the original setting, and a decreased overall sensitivity.

4.3 His Operon

The bacterium *Salmonella typhimurium* regulates the expression of its operon for the amino acid histine with an attenuation mechanism that, although mediated by the ribosome, differs from *E.coli*'s trp operon in several points [1,24]. The differences lead to increased sensitivity, which is important to this organism where Trp biosynthesis is solely inhibited by attenuation. The main difference is that the basal read-through level is decreased, compared to *E.coli*'s trp operon.

Fig. 6 shows the architecture of *S. typhimurium*'s his leader mRNA. First, the number of control codons in the leader increases from 2 out of 9 (Trp) to 7 out of 11 (His). Second, the his operon features two additional segments for hairpin formation between the anti-terminator S_2:S_3 and the terminator S_5:S_6 hairpin. The extra hairpin S_3:S_4 forms rapidly when transcription resumes after melting

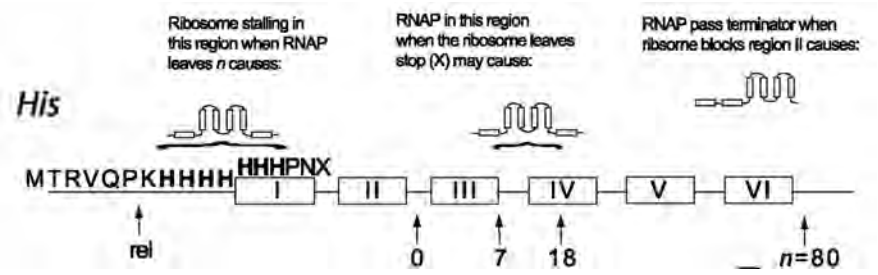

Fig. 6. Architecture of the mRNA control region of *Salmonella typhimurium*'s his operon. Reproduced from [10].

of the pause structure S_1:S_2. It promotes terminator formation by preventing formation of the anti-terminator, thus lowering transcriptional read-through in the case of efficient translation of the control region and early ribosome release.

$$\text{Ribosome} \cdot \text{mRNA}_6 + \text{mRNA}_7 + \text{mRNA}_8 + \text{mRNA}_9 + \text{mRNA}_{10} \quad (15)$$
$$+\text{mRNA}_{11} + S_1^{\text{hp}} \cdot S_2^{\text{hp}} + \text{mRNA}_{17} + \text{RNAP} \cdot \text{DNA}_0 + \text{DNA}_1 + \ldots + \text{DNA}_{80}$$

The initial solution 15 reflects the differences with respect to *E.coli*'s trp operon less, than the adjustment of our rule schemas as follows.

The *transcript elongation* rule schema 2 from Table 1 applies for RNAP's transition to 2-100, except those positions with further control conditions. RNAP's steps to following positions lead to special events: S_3 is spawn at DNA_{15}, S_4 at DNA_{26} , [6] S_5 at DNA_{62}, S_6 at DNA_{78}, while transcription terminates at DNA_{80} unless S_6 is *free*.

Translation occurs with distinct rates for his-codons, which are positions two to eight after melting of the pause loop and release of RNAP, and codons other than his. For this we adapt rule schemas 9 and 8 from Table 1, respectively. Furthermore, we re-use the rule schema for disruption of the pause hairpin by the ribosome as in rule 12 for the case of Trp, and adapt the indices of rules for (un)blocking S_2 and re-assembling S_1. In order to allow formation of the five different mRNA hairpins, rule 1 applies to a larger index set larger than for Trp: $n \in \{i, \ldots, 5\}$. While quantitative evaluation of this model remains future work, in the next section we discuss the impact of the number m of control codons, translated in race with the transcription n nucleotides on DNA.

5 Hyper-Sensitivity of Multi-Step Races

Ribosome-mediated transcriptional attenuation boils down to a *race* between RNAP and ribosome, transcription of n nucleotides on DNA and translation of m

[6] The indices vary from those in the figure, reflecting that the nascent mRNA remains temporarily masked by the RNAP's footprint.

$$\text{RNAP} \cdot \text{DNA}_i^{\text{nil}} + \text{DNA}_{i+1}^{x} \rightarrow_k \text{DNA}_i^{\text{nil}} + \text{RNAP} \cdot \text{DNA}_{i+1}^{x} \quad (16)$$
$$\forall i \in \{0, \ldots, n\} \forall x \in \{\text{codon}, \text{nil}, \text{stop}\}$$

$$\text{RNAP} \cdot \text{DNA}_i^{\text{codon}} + \text{DNA}_{i+1}^{x} \rightarrow_k \text{DNA}_i^{\text{codon}} + \text{RNAP} \cdot \text{DNA}_{i+1}^{x} + m\text{RNA}_{m+(i+1) mod 3}^{\text{cont}} \quad (17)$$
$$\forall i \in \{0, \ldots, n\} \forall x \in \{\text{codon}, \text{nil}, \text{stop}\}$$

$$\text{RNAP} \cdot \text{DNA}_i^{\text{stop}} + \text{DNA}_{i+1}^{x} \rightarrow_k \text{DNA}_i^{\text{stop}} + \text{RNAP} \cdot \text{DNA}_{i+1}^{x} + m\text{RNA}_{m+(i+1) mod 3}^{\text{stop}} \quad (18)$$
$$\forall i \in \{0, \ldots, n\} \forall x \in \{\text{codon}, \text{nil}, \text{stop}\}$$

$${}^{\text{stop}}\text{DNA}_n \cdot \text{RNAP} \rightarrow_k {}^{\text{stop}}\text{DNA}_n + \text{RNAP} \quad (19)$$

$$\text{Ribosome} \cdot m\text{RNA}_i^{\text{cont}} + m\text{RNA}_{i+1}^{x} \rightarrow_{k'} m\text{RNA}_i^{\text{cont}} + \text{Ribosome} \cdot m\text{RNA}_{i+1}^{x} \quad (20)$$
$$\forall i \in \{0, \ldots, m\} \forall x \in \{\text{cont}, \text{stop}\}$$

$$\text{Ribosome} \cdot m\text{RNA}_m^{\text{stop}} \rightarrow_{k'} m\text{RNA}_m^{\text{stop}} + \text{Ribosome} \quad (21)$$

Table 2. For every $m \geq 0$ and $n \geq 0$, we define a model with the following rule schemas, that applies to the initial solution with m control codons (schemas 17, 18, and 21) and n DNA nucleotides (19).

codons on mRNA. Transcription aborts if the ribosome reaches its destination, and if RNAP it continues to transcribe into the structural genes. Thus, as the translation rate increases, the transcription probability decreases. Elf and Ehrenberg [10] demonstrate that the competition between the multi-step mechanisms of transcription and translation leads to a sharp probability transition between 1 and 0. In other words, as the race involves more steps, the decrease of transcription probability becomes sharper. The authors identify this as a source of *hyper-sensitivity* in attenuation, as observed between *E.coli*'s trp operon versus *Salmonella typhimurium* his operon.

In the following, we define a rule-based model of transcription and translation elongation, without the details of mRNA folding from the previous section. Its primary goal it to observe the winner of the race between RNAP and ribosome by simulations, i.e. to observe the respective delays after which of RNAP (resp, the ribosome) reach a given target on DNA (resp. mRNA). As opposed to the previous section, the growing transcript is however elongated codon by codon, this requires to distinguish three rule schemas for transcript elongation where one alone would allow to reach the simulation goal. However, this model is interesting for its simplicity and generality, because it copes with concurrent issues and permits further investigations.

5.1 Model

We assume an initial solution with $m \geq 0$ codons and $n \geq 0$ DNA nucleotides.

$$\begin{aligned} &\text{Ribosome} \cdot m\text{RNA}_0^{\text{cont}} + \cdots + m\text{RNA}_m^{\text{stop}} + \\ &\text{RNAP} \cdot \text{DNA}_0^{\text{nil}} + \text{DNA}_1^{\text{nil}} + \text{DNA}_2^{\text{codon}} + \cdots +^{\text{stop}} \text{DNA}_n \end{aligned} \tag{22}$$

Notation. Our vocabulary is chosen to support the following attributed molecules:

- $m\text{RNA}_j^x$ denotes an mRNA codon at position j, and x a state that distinguishes codons where translation proceeds (cont), from the stop codon where it terminates.
- RNAP_j denotes an RNAP carrying the index j of the next codon appended to the growing mRNA,
- ${}^y\text{DNA}_i^x$ denotes a DNA nucleotide at position i, the state x indicates whether nothing is appended to the transcript (nil) at this position, if a codon at which translation proceeds is appended (codon), or finally a stop codon (stop). If set, the flag y marks a position as the end point of transcription (stop), we omit it unless this is the case. To ensure that RNAP extends the transcript with a new codon once every three nucleotides, we impose the following constraint between DNA and their state x in our initial configurations: for indices i such that $i \bmod 3 = 0, x \in \{\text{codon}, \text{stop}\}$, and for $i \bmod 3 \in \{1, 2\}, x = \text{cont}$.

Table 5 provides the rule schemas. The variables n and m are assigned values by the initial solution of each of our simulations, and are also used when expanding the rule schemas (fixing the end points of transcription and translation, respectively).

Rule schemas 16 to 18 represent RNAP's steps over DNA in elongation, while transcription termination is represented by rule schema 19. For instance, rule schema 17 applies when RNAP is bound to DNA at position i in a state that indicates transcription of a regular codon and nucleotide DNA_{i+1} (in any state x) is free for binding. This results in the transcription of a mRNA codon at position j (as recorded by RNAP), RNAP's translocation to the next nucleotide. Translation elongation steps are represented by rule schema 20, and termination of translation by 21.

5.2 Simulation

Let m and n be the lengths of the mRNA and DNA, that ribosome and RNAP respectively have step over to win the race. Elf and Ehrenberg [10] study the multistep mechanism through three stochastic models for different combinations of m and n ($m = 1$ and $n = 1$, $m = 1$ and $n = 50$, $m = 10$ and $n = 50$). Here, we obtain three rule-based models by instantiation of our generic model to those m and n values. Fig. 7 shows the result of our simulations.[7] The transcription rate

[7] Carried out with Cellucidate http://www.cellucidate.com.

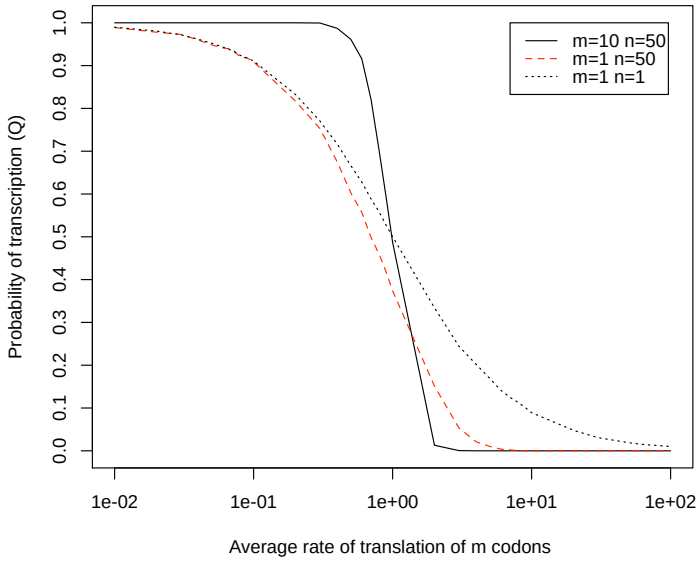

Fig. 7. Probability of transcription as a function of the average rate of translation of m regulatory codons, in race with the transcription of n DNA nucleotides.

k for individual nucleotides is chosen such that the average time to transcribe n nucleotides is always 1 second. The translation rate k' for individual codons is $x \times m$ where x varies logarithmically between 0.01 and 100. Therefore, for a given x, the average time to transcribe m codons is also 1 second. For comparable results with [10], in the initial configurations the length of mRNA transcript is m (or more, m being a stop codon). The curves we obtain coincide with those of [10], confirming the hyper-sensibility induced by the multistep mechanism.

We believe that, in addition of being formally well-defined, our rule-based generic model is simpler to define and provides more information than a stochastic model defined by means of probability density functions. For instance, and as opposed to the model in [10], we can study additional concurrent issues in the race implied by an mRNA target that is not initially transcribed. In that case, a fast ribosome may be queueing behind a slow RNAP that delays the transcription of the mRNA's target. In future work, we plan to carry out simulations of this kind of scenario, and to investigate the relationship to hyper-sensitivity.

6 Conclusion

We have shown that rule-based modeling provides concise models for transcriptional attenuation, a problem left open by previous work on discrete event modeling of the tryptophan operon [36]. The core ingredients for our model are rule schemas for n-ary chemical reactions. The need for n-ary reactions renders representations in object-centered languages such as the stochastic pi-calculus [31,32,33] unappropriate, in practice.

We used the Kappa factory for stochastic simulation [23], which provides convenient analysis tools such as stories, not available for Dizzy [16]. At the time being, we don't know whether all rules schema as presented here can be directly expressed in Kappa without enumerating all chemical reactions they abstract. So far, we found an enumeration-free solution of the rule schemas in Section 5, but not for our first model in Section 4 (the Kappa code of which is included in the appendix).

As modeling technique, we identified nodes of graphs by numbers and addressed successors by simple arithmetics. This technique has its limitations, when graphs become more complex than simple lists. Alternatively, we could assign names to edges, and memorize edge names in attribute values of adjacent nodes, similarly to Kappa.

In future work, we plan a more detailed analysis of our quantitative predictions for the trp operon, to simulate the dynamics of the his operon, and to investigate how rule-based modeling could cover transcriptional attenuation mediated by others than the ribosome [6,7,8]. We also wish to elucidate the discrepancies between our predictions and those of Elf and Ehrenberg [10], regarding anti-terminator formation due to ribosomal stalling at high rates of trp-codon translation in Section 4.

Acknowledgements. The Master's project with Valerio Passini at the Microsoft Research - University of Trento Center for Computational and Systems Biology sharpened our view of the system from a biological perspective and lead us to a rule-based approach. The previous Master's project of Gil Payet (co-supervised with Denys Duchier) had confronted us with the limitations of object-based approaches to the representation of the complex dependencies of transcriptional attenuation. We thank Maude Pupin, who was the first to point us at the intricate regulatory mechanisms at *E.coli*'s tryptophan operon, and Joachim Niehren for his valuable coaching. Finally, we thank the CNRS for a sabbatical to Cédric Lhoussaine, and the Agence Nationale de Recherche for funding this work through a *Jeunes Chercheurs* grant (ANR BioSpace, 2009-2011).

References

1. Kasai, T.: Regulation of the expression of the histidine operon in Salmonella typhimurium. Nature **249** (1974) 523–527
2. Yanofsky, C.: Attenuation in the control of expression of bacterial operons. Nature **289** (1981) 751–758
3. Barboric, M., Peterlin, B.M.: A new paradigm in eukaryotic biology: HIV Tat and the control of transcriptional elongation. PLoS Biology **3** (2005) 0200–02003
4. A Saunders, L.C., Lis, J.: Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol **7** (2006) 556–567
5. Beisel, C.L., Smolke, C.D.: Design principles for riboswitch function. PLoS Computational Biology **5** (2009) e1000363
6. Gollnick, P., Babitzke, P., Antson, A., Yanofsky, C.: Complexity in regulation of tryptophan biosynthesis in Bacillus subtilis. Annual Review of Genetics **39** (2005) 47–68
7. Yanofsky, C.: RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. RNA - A publication of the RNA Society **13** (2007) 1141–1154
8. Gutierrez-Preciado, A., Jensen, R., Yanofsky, C., Merino, E.: New insights into regulation of the tryptophan biosynthetic operon in Gram-positive bacteria. Trends in Genetics **21** (2005) 432–436
9. Santillan, M., Zeron, E.S.: Dynamic influence of feedback enzyme inhibition and transcription attenuation on the tryptophan operon response to nutritional shifts. Journal of Theoretical Biology **231** (2004) 287–298
10. Elf, J., Ehrenberg, M.: What makes ribosome-mediated trascriptional attenuation sensitive to amino acid limitation? PLoS Computational Biology **1** (2005) 14–23
11. Chabrier-Rivier, N., Fages, F., Soliman, S.: The biochemical abstract machine BioCham. In: Proceedings of CMSB 2004. Volume 3082 of Lecture Notes in Bioinformatics. (2005) 172–191
12. Danos, V., Laneve, C.: Formal molecular biology. Theoretical Computer Science **325** (2004) 69–110
13. Krivine, J., Milner, R., Troina, A.: Stochastic bigraphs. In: 24th Conference on the Mathematical Foundations of Programming Semantics. Volume 218 of Electronical notes in theoretical computer science., Elsevier (2008) 73–96
14. Ciocchetta, F., Hillston, J.: Bio-PEPA: a framework for modelling and analysis of biological systems. Theoretical Computer Science (2008) To appear.
15. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics **22** (1976) 403–434
16. Ramsey, S., Orrell, D., Bolouri, H.: Dizzy: stochastic simulation of large-scale genetic regulatory networks. Journal of Bioinformatics and Computational Biology **3** (2005) 415–436
17. Trun, N., Trempy, J.: Fundamental bacterial genetics. Blackwell (2003)
18. Konan, K.V., Yanofsky, C.: Role of ribosome release in regulation of tna operon expression in Escherichia coli. J. Bacteriol. **181** (1999) 1530–1536
19. Gollnick, P.: Trp operon and attenuation. In Lennarz, W.J., Lane, M.D., eds.: Encyclopedia of Biological Chemistry. Elsevier, New York (2004) 267 – 271
20. Lloyd, J.W.: Foundations of Logic Programming, 2nd Edition. Springer Verlag (1987)
21. Shieber, S.M.: An Introduction to Unification-Based Approaches to Grammar. Volume 4. CLSI Publications (1986)

22. Baader, F., Nipkow, T.: Term rewriting and all that. Cambridge University Press, New York, NY, USA (1998)
23. Danos, V., Feret, J., Fontana, W., Krivine, J.: Scalable simulation of cellular signaling networks. In: Programming Languages and Systems, 5th Asian Symposium. Volume 4807 of Lecture Notes in Computer Science., Springer Verlag (2007) 139–157
24. Alifano, P., Fani, R., Lio, P., Lazcano, A., Bazzicalupo, M., Carlomagno, M., Bruni, C.: Histidine biosynthetic pathway and genes: Structure, regulation, and evolution. Microbiological Reviews **60** (1996) 44–69
25. Cardelli, L., Zavattaro, G.: On the computational power of biochemistry. In: Proceedings of the 3rd international conference on Algebraic Biology, Berlin, Heidelberg, Springer-Verlag (2008) 65–80
26. Danos, V., Feret, J., Fontana, W., Harmer, R., Krivine, J.: Rule-based modelling of cellular signalling. In: CONCUR - Concurrency Theory, 18th International Conference. Volume 4703 of Lecture Notes in Computer Science., Springer Verlag (2007) 17–41
27. Hlavacek, W.S., Faeder, J.R., Blinov, M.L., Posner, R.G., Hucka, M., Fontana, W.: Rules for modeling signal-transduction systems. Science STKE **re6** (2006)
28. Faeder, J.R., Blinov, M.L., Goldstein, B., Hlavacek, W.S.: Rule-based modeling of biochemical networks. Complexity **10** (2005) 22–41
29. Pradalier, S., Credi, A., Garavelli, M., Laneve, C., Zavattaro, G.: Modelization and simulation of nano devices in the nano-kappa calculus. In: Computational Methods in Systems Biology, International Conference CMSB 2007. (2007)
30. Hillston, J.: Process algebras for quantitative analysis. In: 20th IEEE Symposium on Logic in Computer Science (LICS 2005), 26-29 June 2005, Chicago, IL, USA, Proceedings, IEEE Computer Society (2005) 239–248
31. Regev, A.: Computational Systems Biology: A Calculus for Biomolecular Knowledge. Tel Aviv University (2002) PhD thesis.
32. Phillips, A., Cardelli, L.: Efficient, correct simulation of biological processes in the stochastic pi-calculus. In: Computational Methods in Systems Biology, International Conference. Volume 4695 of Lecture Notes in Computer Science., Springer Verlag (2007) 184–199
33. Kuttler, C., Lhoussaine, C., Niehren, J.: A stochastic pi calculus for concurrent objects. In: Second International Conference on Algebraic Biology. Volume 4545 of Lecture Notes in Computer Science., Springer Verlag (2007) 232–246
34. Dematté, L., Priami, C., Romanel, A.: The beta workbench: A tool to study the dynamics of biological systems. Briefings in Bioinformatics **9** (2008) 437–449
35. von Heijne, G., Nilsson, L., Blomberg, C.: Translation and messenger RNA secondary structure. Journal of Theoretical Biology **68** (1977) 321–329
36. Simão, E., Remy, E., Thieffry, D., Chaouiya, C.: Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in E .coli. In: ECCB/JBI. (2005) 190–196
37. Yanofsky, C.: Transcription attenuation: once viewed as a novel regulatory strategy. Journal of Bacteriology **182** (2000) 1–8
38. JR Roesser, Y.N., Yanofsky, C.: Regulation of basal level expression of the tryptophan operon of Escherichia coli. J Biol Chem **264** (1989) 12284–8

A Kappa Rules for Attenuation Model in Section 4

In the following we list the Kappa code of our model in Section 4, as we implemented it in Kappa Factory v 12.2.0. in order to run stochastic simulations.

Some comments on the syntax of Kappa seem appropriate. For instance consider the following reaction produced by rule schema 1:

```
'LoopS2S3 (anti-terminator)' S2(s~fr),S3(s~fr)
  -> S2(s~hp!1),S3(s~hp!1) @ $INF
```

This reaction is given the name `'LoopS2S3(anti-terminator)'` and an infinite rate `@ $INF`. The reactants `S2(s~fr)` and `S3(s~fr)` have an attribute `s` both with values `fr`. The produced the molecules `S2(s~hp!1)` and `S3(s~hp!1)`, where attribute `s` has value `hp`. The modifier `!1` means that resulting molecules form a complex that is linked by edge `1`.

```
# rule schema 1
'LoopS2S3 (anti-terminator)' S2(s~fr),S3(s~fr)
  -> S2(s~hp!1),S3(s~hp!1) @ $INF
'LoopS3S4 (terminator)' S4(s~fr),S3(s~fr) -> S4(s~hp!1),S3(s~hp!1) @ $INF
'LoopS1S2 (pause)' S1(s~fr),S2(s~fr) -> S1(s~hp!1),S2(s~hp!1) @ $INF
# rule schema 2
'RNAPto2' RNAP(d!1),DNA1(t!1),DNA2(t) -> RNAP(d!2),DNA1(t),DNA2(t!2) @ 50.0
...
'RNAPto35' RNAP(d!1),DNA34(t!1),DNA35(t) -> RNAP(d!2),DNA34(t),DNA35(t!2) @ 50.0
'RNAPto37' RNAP(d!1),DNA36(t!1),DNA37(t) -> RNAP(d!2),DNA36(t),DNA37(t!2) @ 50.0
...
'RNAPto46' RNAP(d!1),DNA45(t!1),DNA46(t) -> RNAP(d!2),DNA45(t),DNA46(t!2) @ 50.0
'RNAPto49' RNAP(d!1),DNA48(t!1),DNA49(t) -> RNAP(d!2),DNA48(t),DNA49(t!2) @ 50.0
'RNAPto50' RNAP(d!1),DNA49(t!1),DNA50(t) -> RNAP(d!2),DNA49(t),DNA50(t!2) @ 50.0
# rule schema 3
'RNAPresumes_S1S2broken' RNAP(d!1),DNA0(t!1),DNA1(t),S2(s~fr) ->
 RNAP(d!2),DNA0(t),DNA1(t!2),S2(s~fr) @ 50.0
# rule schema 4
'RNAPto36_spawnS3' RNAP(d!1),DNA35(t!1),DNA36(t)
   -> RNAP(d!2),DNA35(t),DNA36(t!2),S3(s~fr) @ 50.0
# rule schema 5
'RNAPto47_spawnS4' RNAP(d!1),DNA46(t!1),DNA47(t)
   -> RNAP(d!2),DNA46(t),DNA47(t!2),S4(s~fr) @ 50.0
# rule schema 6
'RNAPto48_whenS4hp' RNAP(d!1),DNA47(t!1),DNA48(t),S4(s~hp)
      -> RNAP(d),DNA47(t),DNA48(t),S4(s~hp) @ 50.0
# rule schema 7
'RNAPto48_whenS4fr' RNAP(d!1),DNA47(t!1),DNA48(t),S4(s~fr)
      -> RNAP(d!2),DNA47(t),DNA48(t!2),S4(s~fr) @ 50.0
# rule schema 8
'RiboTo8' Ribo(m!1),mRNA7(t!1),mRNA8(t) -> Ribo(m!2),mRNA7(t),mRNA8(t!2) @ 15.0
```

```
...
'RiboTo15' Ribo(m!1),mRNA14(t!1),mRNA15(t)
   -> Ribo(m!2),mRNA14(t),mRNA15(t!2) @ 15.0
# rule schema 9
'RiboTo11_Trp' Ribo(m!1),mRNA10(t!1),mRNA11(t)
       -> Ribo(m!2),mRNA10(t),mRNA11(t!2) @ 1.0
'RiboTo12_Trp' Ribo(m!1),mRNA11(t!1),mRNA12(t)
       -> Ribo(m!2),mRNA11(t),mRNA12(t!2) @ 1.0
# rule schema 10
'RiboTo7_MeltS1S2' Ribo(m!1),mRNA6(t!1),mRNA7(t),S1(s~hp!2),S2(s~hp!2)
    -> Ribo(m!2),mRNA6(t),mRNA7(t!2),mRNA10(t),mRNA15(t),mRNA11(t),
       S2(s~fr),mRNA14(t),mRNA13(t),mRNA12(t) @ 15.0
# rule schema 11
'blockS2_ribo@13' Ribo(m!1),mRNA13(t!1),S2(s~fr)
  -> Ribo(m!1),mRNA13(t!1),S2(s~bl) @ $INF
# rule schema 12
'RiboRelease@15_joinS1' mRNA15(t!1),Ribo(m!1),mRNA10(t),mRNA11(t),mRNA12(t),
mRNA13(t),mRNA14(t) -> Ribo(m),S1(s~fr) @ 1.0
# rule schema 13
'Ribo_free_unblockS2' S2(s~bl),Ribo(m) -> S2(s~fr),Ribo(m) @ $INF
```

All quantitative information reported in Figures 4 and 5 is obtained from Kappa *stories* for the application of rules `'LoopS2S3 (anti-terminator)'` and `'LoopS3S4 (terminator)'`. We performed 5000 simulation runs, for each rate of trp-codon translation in the range between 0 and 15, in steps of one.

Note that the order of the first three rules must be maintained as listed above to reproduce our results, due to a bug in the implementation of infinite rates: if in a solution, several rules with infinite are applicable, Kappa v 12.2.0 always selects the one among these stated first in the model.[8] In order to obtain a proper stochastic semantics in terms of CTMCs for chemical reactions with infinite rates, one needs to assume that selection among several next reactions with infinite rates is equally probable [33]. Due to this same reason, the probabilities for obtained for stories leading to `'LoopS2S3 (anti-terminator)'` need to be reduced by half: whenever S_2:S_3 is formed, S_1 is also available should have an equal chance to pair with S_2.

[8] Infinite rates are not mentionned in Kappa's semantics [23].