Temporal Logic based Framework to Model and Analyse Gene Networks with Alternative Splicing

Sohei Ito
Department of Fisheries Distribution and Management, National Fisheries University, 2-7-1 Nagata-Honmachi, Shimonoseki, Yamaguchi, Japan

Keywords: Gene Regulatory Network, Systems Biology, Alternative Splicing, Temporal Logic.

Abstract: Toward system-level understanding of biological systems, we need a formalism to model and analyse them. Due to incompleteness of knowledge about quantitative parameters and molecular mechanisms, qualitative methods have been useful alternatives. We have been working on temporal logic-based approach for qualitative modelling and analysis of gene regulatory networks. Although our framework is well-established to model several aspects of gene regulation, we still lack treatment of *alternative splicing*, which contributes to proteomic diversity of eukaryotic organisms. In this paper we extend our logic-based qualitative framework to be able to capture alternative splicing, which is crucial to model the gene regulatory networks in eukaryotic organisms. We study mechanisms of alternative splicing and propose how we model each mechanism, then demonstrate the modelling method by analysing the regulatory network of sex determination in *Drosophila* and verify that the network ensures sex determination.

1 INTRODUCTION

To understand complex activities of the cell, mathematical and computational approach is indispensable. For precise mathematical modelling, we need huge amount of *quantitative* information. Such quantitative information available, however, is unfortunately limited and not sufficient despite of recent advances in biology. Instead, a lot of *qualitative* information about biological systems has been accumulated such as schematic network representations of gene-gene interactions, protein-protein interactions, signalling pathways, and so on. Thus a *qualitative* method for modelling and analysing biological processes based on *qualitative* information is desired.

In this context, several computational formalisms in biological modelling have been proposed: Boolean network (Thomas, 1991), Petri net (Heiner et al., 2008), timed automata (Batt et al., 2007) and process algebra (Ciocchetta and Hillston, 2009), though all of them are not necessarily qualitative. In these formalisms, the possible behaviours of a system can be characterised by the traces of the model. Such computational formalisms need concrete information on molecular mechanisms and regulatory logics to construct a model. Since biological information is inherently incomplete, it is pointed out that constraint-based modelling is well-suited in biological modelling (Palsson, 2000) in which we give several constraints reflecting incomplete knowledge on the system to limit possible behaviours (solution space) of biological systems.

In accordance with the motivation of constraint-based modelling, we have been working on a logic-based qualitative approach to model and analyse behaviours of gene regulatory networks (Ito et al., 2010; Ito et al., 2013b; Ito et al., 2013a; Ito et al., 2014; Ito et al., 2015) which uses linear temporal logic (LTL) as the modelling language. In contrast to the original constraint-based modelling paradigm which intends to limit the quantitative possible behaviours, our approach aims to characterise qualitative possible behaviours using qualitative information of gene-gene interactions which is represented as gene regulatory networks. Since we only use qualitative information, the reasoning is also limited to qualitative properties. However, we can still analyse important properties of gene networks such as oscillation, stability and reachability, as it is pointed out that the overall behaviour is relatively insensitive to the exact numerical values of the kinetic constants (Palsson, 2000).

One of the difficulties in modelling and analysing gene networks is the *alternative splicing* in eukaryotic organisms. Alternative splicing of a precursor mRNA (pre-mRNA) gives rise to multiple transcrip-

Ito, S.
Temporal Logic based Framework to Model and Analyse Gene Networks with Alternative Splicing.
DOI: 10.5220/0005655001510158
In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016) - Volume 3: BIOINFORMATICS*, pages 151-158
ISBN: 978-989-758-170-0

tion products from one gene. The selection of alternative splicing at an appropriate timing is critical for cell differentiation and sex determination. In quantitative approach, the splicing process can be modelled as thermodynamical reactions (Louis et al., 2003; Wen, 2013). However, it is unclear how alternative splicing is modelled in qualitative approach. The aim of this paper is to establish a method for modelling alternative splicing in our LTL-based framework. In this paper we study mechanisms of alternative splicing and show how each mechanism can be modelled in LTL. We demonstrate our formal framework by modelling and analysing the network of sex determination in *Drosophila* (Camara et al., 2008; Salz and Erickson, 2010).

The rest of this paper is organised as follows. In section 2 we review our LTL-based framework for modelling and analysing gene networks using LTL as a baseline of this work. In section 3 we study mechanisms of alternative splicing and present how we formally model them in our framework. In section 4 we demonstrate our formal framework in analysing the network of sex determination in *Drosophila*. The final section offers conclusion and future directions.

2 QUALITATIVE FRAMEWORK FOR MODELLING AND ANALYSING GENE NETWORKS USING LTL

A gene network is represented as a directed graph whose nodes and edges (labelled by +/-) represent genes and regulation relation (activation/inhibition) among them, respectively. In Fig. 1 we show an example of a gene regulatory network which consists of three genes. A behaviour of a gene network is represented as a time series of expression profiles of the genes in the network. Fig. 2 shows an example time series of the network depicted in Fig. 1. In this behaviour the value x_y is the threshold level of gene *x* to activate gene *y*, y_z the threshold level of gene *y* to activate gene *z* and z_y the threshold level of gene *z* to inhibit gene *y*. At the beginning, no genes are expressed. At time t_0, gene *x* begins to be expressed and its expression level begins to grow. At time t_1 gene *x* crosses the threshold x_y, thus gene *y* becomes ON due to the positive effect from gene *x*. At time t_2 gene *x* stops to be expressed and the level decreasing. At time t_3 gene *x* falls below x_y, thus gene *y* becomes OFF. This way the network changes its state over time.

This time series can be represented as a discrete

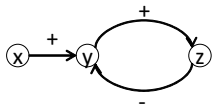

Figure 1: An example gene regulatory network. Gene *x* activates gene *y*, *y* activates *z* and *z* inhibits *y*. A plus edge represents activation and a minus edge represent inhibition.

state transition system (called *linear time structure*) depicted in Fig. 3. It consists of states (represented as circles) and transitions (represented as arrows). The configurations of the network at each state are shown by the propositions depicted below states. We have the following propositions to describe the configurations of the network:

- on_x, on_y, on_z: whether genes *x*, *y* and *z* are ON or OFF, respectively.
- x_y, y_z, z_y: whether the expression level of gene *x*, *y* and *z* are beyond the threshold x_y, y_z and z_y, respectively[1].

We easily see that state 0 represents the configuration of the network at the beginning, state 1 represents the configuration between t_0 and t_1, state 2 between t_1 and t_2, and so on.

In general, there are many behaviours which can be produced by a single network depending on the initial conditions, input scenarios, response times, and so on. Our purpose is to model (or characterise) the set of possible behaviours for a given gene network. In quantitative approach, ordinary differential equations (ODEs) are widely used. In the current setting, we do not handle a numerical time series but a symbolic time series of a behaviour (a linear time structure). To characterise and reason about such structures, linear temporal logic (LTL) is the suitable mathematical language. LTL can be seen as propositional logic equipped with temporal operators such as *G* (Globally), *F* (Future), *U* (Until) and *W* (Weak until). $G\phi$ means ϕ is always true, $F\phi$ means ϕ is eventually true, $\phi U \psi$ means ϕ is true until ψ is true and $\phi W \psi$ means ϕ is true until ψ is true or ϕ is indefinitely true (in this case ψ need not be true in any future). The formal syntax and semantics are omitted due to the space limitation.

We are to characterise the set of possible behaviours (linear time structures) for a given network. This can be done by specifying an LTL formula ϕ_N for a given network *N* such that the set of possible behaviours of a network is characterised as $\{\sigma \mid \sigma \models \phi_N\}$, i.e. all linear time structures which satisfy the behaviour specification ϕ_N. The problem of analysing

[1]Threshold values are written in Roman while propositions are written in italics.

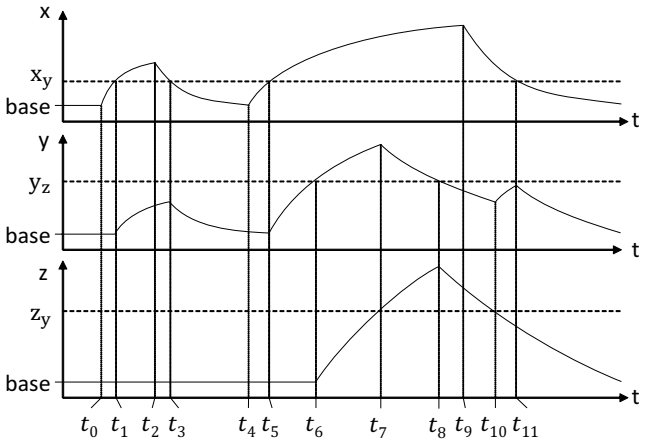

Figure 2: Time series of expression levels of gene *a*, *b* and *c* in the network Fig. 1.

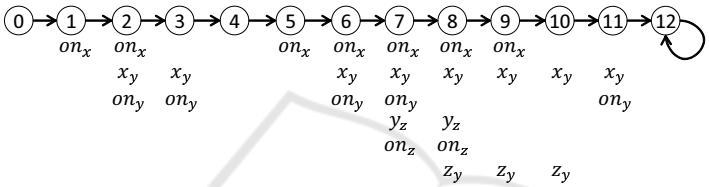

Figure 3: Symbolic representation of the time series of Fig. 2.

network behaviours, e.g. checking whether there is a behaviour which satisfies a certain property ψ (also written in LTL), can be solved by finding σ such that $\sigma \models \phi_N \wedge \psi$, i.e. checking *satisfiability* of the formula $\phi_N \wedge \psi$. Thus analysing a gene network is reduced to satisfiability checking of LTL. Once we have a formula $\phi_N \wedge \psi$, the analysis can be automatically done by LTL satisfiability checkers. LTL satisfiability checkers construct a Büchi automaton of a given LTL formula which precisely accepts the linear time structures in which the formula is true (Vardi and Wolper, 1994). Hence if the language accepted by the automaton is empty, the formula is not satisfiable. The non-emptiness problem of Büchi automata is solved by checking the existence of a maximal strongly connected component containing accepting states of the automata.

Due to the space limitation, we do not show in detail how to specify ϕ_N which characterises possible behaviours of a given network N. Interested readers would like to consult our previous work (Ito et al., 2015). The key idea is the following qualitative principles of gene network behaviours:

- A gene is ON when its activators are expressed beyond some thresholds.
- A gene is OFF when its inhibitors are expressed beyond some thresholds.
- If a gene is ON, its expression level increases.
- If a gene is OFF, its expression level decreases.

By expressing these principles in LTL, we have a characterisation of possible behaviours of the network. We only show an example characterisation of the possible behaviours of the network in Fig. 1. For this network we introduce the set of propositions $\{on_x, on_y, on_z, x_y, y_z, z_y\}$. Using these propositions, we have the following behaviour specification.

$$
\begin{aligned}
&G(x_y \wedge \neg z_y \rightarrow on_y) \wedge G(z_y \rightarrow \neg on_y) \wedge G(y_z \leftrightarrow on_z) \wedge \\
&G(on_x \rightarrow F(x_y \vee \neg on_x)) \wedge \\
&G(on_x \wedge x_y \rightarrow (x_y W \neg on_x)) \wedge \\
&G(\neg on_x \rightarrow F(\neg x_y \vee on_x)) \wedge \\
&G(\neg on_x \wedge \neg x_y \rightarrow (\neg x_y W on_x)) \wedge \ldots .
\end{aligned}
$$

In this specification we assumed that gene *y* is OFF if gene *z* is inhibiting *y* nevertheless gene *x* is activating *y*. For this network let us check the bistability of the expression of gene *y*, which is written in LTL as:

$$(Gon_x \rightarrow FGon_y) \wedge (G\neg on_x \rightarrow FG\neg on_y)$$

We check the satisfiability of the conjunction of the above formulae. We used T^3-builder (Aoshima, 2003) to check it and had the answer 'Yes'. This means that the network of Fig. 1 produces two opposite behaviours: a behaviour in which gene y is always ON after some time point and another behaviour in which gene y is always OFF after some time point, which is determined by whether gene x is ON.

3 MODELLING ALTERNATIVE SPLICING BY LTL

In the formal framework described in the previous section, we did not take alternative splicing into consideration. This section discusses how we model the alternative splicing in our framework.

In most eukaryotic organisms, the process of gene expression consists of three steps: (i) a DNA region which encodes a gene is transcribed into a precursor messenger RNA (pre-mRNA), (ii) introns (and some exons) in a pre-mRNA are removed, and (iii) a processed mRNA is transported outside of a nucleus and translated into a protein. Alternative splicing happens in the step (ii) which causes the diversity of processed mRNA from a single pre-mRNA by removing some exons selectively as well as introns (Fig. 4). Due to alternative splicing several isoforms of a protein are obtained from one gene.

A natural solution to handle alternative splicing is that we regard each isoform as being produced by different (virtual) genes. However, this treatment causes blow-up of the number of propositions and the size of a behaviour specification, which deteriorates the performance of analysis.

In this section we study the molecular mechanisms of alternative splicing (David and Manley, 2008; Hertel, 2008; Kornblihtt, 2005; Matlin et al., 2005) and propose how to model these mechanisms in LTL without introducing extra genes.

Mechanism 1. One of the mechanisms of alternative splicing is the usage of multiple promoters. The mouse α-amylase gene is known to have this mechanism. For the purpose of illustration, let us consider a gene u which has two promoters X and Y (Fig. 5). Gene u has two splicing patterns depending on promoters. The choice of promoters is made by a transcription complex. To model this mechanism in LTL, we introduce propositions TC_u^X and TC_u^Y to represent whether the levels of transcription complexes for promoter X and Y are sufficient, respectively. We write $R^+(u)$ and $R^-(u)$ for LTL terms representing conditions for activation(+) and inhibition(-) of gene u, respectively[2]. For example if gene v activates u and gene w inhibits u, $R^+(u)$ will be $v_u \wedge \neg w_u$ and $R^-(u)$ will be $\neg v_u \wedge w_u$. Then conditions for activating/inhibiting gene u can be described as:

$$G(R^+(u) \wedge TC_u^X \wedge \neg TC_u^Y \rightarrow on_u^X \wedge \neg on_u^Y), \quad (1)$$

$$G(R^+(u) \wedge TC_u^Y \wedge \neg TC_u^X \rightarrow on_u^Y \wedge \neg on_u^X), \quad (2)$$

$$G(R^-(u) \rightarrow \neg on_u^X \wedge \neg on_u^Y), \quad (3)$$

[2]In general we have several conditions for activating/inhibiting gene u, but treatment is the same.

where the propositions on_u^X and on_u^Y represent gene u is expressed from promoter X and Y, respectively. Formula (1) says that if gene u is activated and the transcription complex for promoter X is sufficient and that for promoter Y is not sufficient, gene u is expressed from promoter X, not from promoter Y. Formula (2) describes the case that gene u is expressed from promoter Y. Formula (3) says that if gene u is not activated, gene u is expressed neither from promoter X nor from promoter Y. Note that we can use $\leftrightarrow$ instead of $\rightarrow$ in the above formulae depending on our assumptions for a system to be modelled. (The same argument applies to the other mechanisms.)

The problem is that how we describe the case when both TC_u^X and TC_u^Y are true, i.e. $R^+(u) \wedge TC_u^X \wedge TC_u^Y \rightarrow$?. The situation is almost the same as the case when both activators and inhibitors are active for a gene, which is discussed in our previous work. The solution depends on the knowledge or assumption we have on a given network or a problem. We may write $on_u^X \wedge on_u^Y$ which means both on_u^X and on_u^Y are true, or write $on_u^X \vee on_u^Y$ which means on_u^X or on_u^Y are true (both are true is allowed). If we do not add any clause, the values of on_u^X and on_u^Y are free in this case. Another choice is to assume that both TC_u^X and TC_u^Y cannot be true simultaneously by adding the clause $G\neg(TC_u^X \wedge TC_u^Y)$.

For each gene that is regulated by the translated products of gene u from promoter X/Y, we introduce propositions of the expression levels for u^X and u^Y such as $u_1^X, u_2^X, \ldots, u_1^Y, u_2^Y, \ldots$, and clauses for the changes of the expression levels of them. These clauses are the same as those we have for normal genes.

Readers might wonder what is the difference from the modelling manner in that we split gene u into the (virtual) genes u^X and u^Y. If we do so, we must duplicate regulating terms $R^+(u)$ and $R^-(u)$ into $R^+(u^X)$, $R^+(u^Y)$, $R^-(u^X)$ and $R^-(u^Y)$. For example, if we assume that gene u has two regulators v and w, we need to introduce the threshold levels of v and w for both gene u^X and u^Y: v_u^X, v_u^Y, w_u^X and w_u^Y. This blows up the number of clauses for the changes of the expression levels of gene u and w. In the above specification, however, such blow-ups of the number of propositions and clauses are avoided.

Mechanism 2. The next mechanism of alternative splicing is the presence/absence of splicing factors (SFs). SFs bind to introns or exons and change the splice sites of an transcribed pre-mRNA of a gene. SFs can activate or inhibit a certain splice sites, but the important fact is that splicing is determined by whether SFs are binding or not. We assume that a gene u produces two isoforms u^A (when the SF is

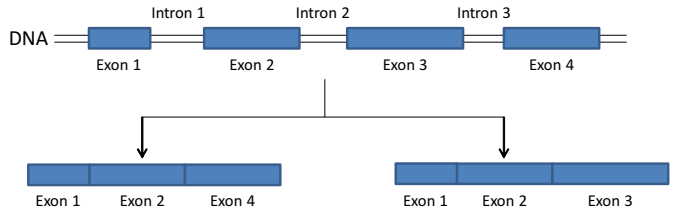

Figure 4: Alternative splicing produces different mRNAs from a single pre-mRNA.

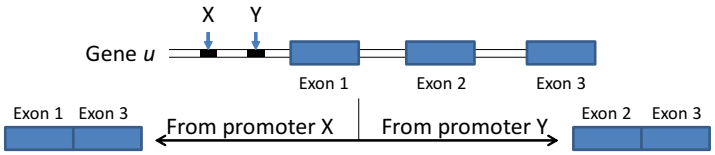

Figure 5: Gene u has two promoters X and Y.

binding) and u^B (when the SF is not binding). We introduce a proposition SF_u which represents the level of the SF exceeds the threshold SF_u upon which the SF affects on splicing. Then conditions for activating/inhibiting gene u can be described as:

$$G(R^+(u) \wedge SF_u \rightarrow on_u^A \wedge \neg on_u^B), \quad (4)$$

$$G(R^+(u) \wedge \neg SF_u \rightarrow on_u^B \wedge \neg on_u^A), \quad (5)$$

$$G(R^-(u) \rightarrow \neg on_u^A \wedge \neg on_u^B). \quad (6)$$

Formula (4) says that if gene u is activated and the level of the SF is beyond the threshold SF_u, gene u produces the isoform u^A. Formula (5) describes the case when the level of the SF is not enough to bind pre-mRNAs of gene u. In this case gene u produces the isoform u^B. Formula (6) says that if gene u is not activated, u does not produce any isoform.

In general, multiple SFs (SF1, SF2, ...) involve the splicing of a gene u which results in many isoforms $u^A, u^B, u^C, \ldots$. By generalising the above formulae we can easily model such complex splicing. For each combination of effective SFs, we specify that the corresponding isoform is expressed (ON).

4 DEMONSTRATION

In this section we apply our method for modelling alternative splicing described in section 3 to analyse the network of sex determination in *Drosophila* (Camara et al., 2008; Salz and Erickson, 2010).

Genes involved in this sex determination process are *Sxl*, *tra*, *tra-2* and *dsx*. *Sxl*, *tra* and *dsx* have both male-specific and female-specific splicing. Moreover, *Sxl* has two promoters – the early promoter and the late promoter. *Sxl* is known to have two female-specific splicing – one from the early promoter and the other from the late promoter. Male-specific splicing of *Sxl* occurs only from the late promoter. Thus we have three isoforms from *Sxl*. We represent S^e (from early promoter), S^f (female-specific splicing from the late promoter) and S^m (male-specific splicing) for each isoform. We similarly write t^f (female-specific) and t^m (male-specific) for *tra*, and d^f (female-specific) and d^m (male-specific) for *dsx*.

The network controlling sex determination in *Drosophila* is illustrated in Fig. 6. First the isoform S^e is produced from *Sxl* by the early promoter. S^e activates female-specific splicing of *Sxl* itself and produces the isoform S^f, which inhibits male-specific splicing of *Sxl* and *tra*. As a result *tra* produces female-specific isoform t^f. This t^f with *tra-2* activates female-specific splicing of *dsx*.

To model this network in LTL we introduce the following propositions.

- $on_S^m, on_S^f, on_S^e, on_t^m, on_t^f, on_d^m, on_d^f$: representing whether the isoforms S^m, S^f, S^e, t^m, t^f, d^m and d^f are expressed, respectively.
- $S_S^e, S_S^f, S_t^f, t_d^f, t2$: these propositions correspond to whether each isoform is expressed beyond the threshold level for each activation/inhibition between genes. S_S^e corresponds to $S^e \xrightarrow{+} S^f$, S_S^f to $S^f \xrightarrow{-} S^m$, S_t^f to $S^f \xrightarrow{-} t^m$, t_d^f to $t^f \xrightarrow{+} d^f$ and $t2$ to *tra-2* $\xrightarrow{+} d^f$ (see Fig. 6). We consider S_S^f and t_d^f as splicing factors for *tra* and *dsx*, respectively.
- TC_S^E, TC_S^L: representing whether the levels of transcription complexes of *Sxl* for the early(E) and late(L) promoters are sufficient, respectively.

Here we show the essential part, i.e. how the splicing is controlled, of behaviour specification of the network.

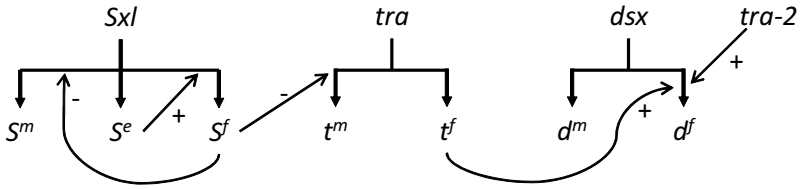

Figure 6: The network controlling sex determination in *Drosophila*.

$$G(TC_S^E \wedge \neg TC_S^L \leftrightarrow on_S^e \wedge \neg on_S^f \wedge \neg on_S^m) \wedge \quad (7)$$

$$G(S_S^e \wedge TC_S^L \wedge \neg TC_S^E \leftrightarrow on_S^f \wedge \neg on_S^m) \wedge \quad (8)$$

$$G(\neg S_S^f \wedge TC_S^L \wedge \neg TC_S^E \leftrightarrow on_S^m \wedge \neg on_S^f) \wedge \quad (9)$$

$$G(TC_S^L \leftrightarrow (on_S^m \vee on_S^f)) \wedge \quad (10)$$

$$G(S_S^f \rightarrow \neg on_S^m) \wedge \quad (11)$$

$$G(S_t^f \rightarrow (\neg on_t^m \wedge on_t^f)) \wedge \quad (12)$$

$$G(\neg S_t^f \rightarrow (on_t^m \wedge \neg on_t^f)) \wedge \quad (13)$$

$$G(t2 \wedge t_d^f \rightarrow on_d^f \wedge \neg on_d^m) \wedge \quad (14)$$

$$G(t2 \wedge \neg t_d^f \rightarrow \neg on_d^f \wedge on_d^m) \wedge \quad (15)$$

$$G(\neg t2 \rightarrow \neg on_d^f \wedge \neg on_d^m) \wedge \ldots \quad (16)$$

Formulae (7)-(9) are directly derived from the mechanism 1 in the previous section. Since we do not have an explicit regulator for *Sxl* in the network, the regulating condition ($R^+(\cdot)$ in section 3) is empty in formula (7). Formula (10) reflects the assumption that isoforms S^m and S^f need to be expressed from the late promoter. Formula (11) stipulates the negative effect of S^f to the expression of S^m. Formulae (12)-(16) are derived from the mechanism 2 where S_t^f and t_d^f as splicing factors.

For this network we check the bistability – female-specific stability and male-specific stability. The critical switch to determine this is the female-specific transcription complex for early promoter of *Sxl*. If its intracellular level is sufficient at the initial time, the cell eventually reaches female-specific stability, otherwise the cell eventually reaches male-specific stability. This property is described in LTL as:

$$(TC_S^E \rightarrow FG(on_S^f \wedge on_t^f \wedge on_d^f)) \wedge$$
$$(\neg TC_S^E \rightarrow FG(on_S^m \wedge on_t^m \wedge on_d^m))$$

in which female(male)-specific stability is written as that female(male)-specific splicing of the three genes *Sxl*, *tra* and *dsx* are maintained. Using the LTL satisfiability checker we have the result 'Yes', which means that this network surely satisfies the bistability.

We analysed the Büchi automaton constructed by the LTL formula by the LTL satisfiability checker and investigate the witnesses of the satisfiability[3]. Since there are many possible behaviours from all possible initial state, we extracted a behaviour from the initial state where gene *Sxl* is OFF (other genes are arbitrary) and transcription complex of *Sxl* for the early promoter is present/absent. We depict the behaviours (linear structures) obtained from the automaton as witnesses in Fig. 7. The states are represented as vectors of propositional values for $(TC_S^E, TC_S^L, on_S^m, on_S^f, on_t^m, on_t^f, on_d^m, on_d^f)$ in Fig. 7, where 1 represents true and 0 represents false. For simplicity the other propositions are omitted, thus the edge in the figure does not necessarily corresponds the one atomic step of the original behaviour because the states in which only the values of the omitted propositions differ are identified.

In both Fig. 7(a) and (b), after some initial perturbation, the network reaches the stable state in which all the sex-specific splicing is maintained. Note that this is just instances of the possible behaviours happened to be produced by the automaton as witnesses. The behaviours include some interesting features: the transcription for the late promoter of *Sxl* is sometimes cut off in the female-specific behaviour (a) and the transcription factor for the early promoter of *Sxl*, which is known to be a female-specific transcription complex, is once produced in the male-specific behaviour (b), nevertheless, the network reaches the final desirable states (only sex-specific splicing are maintained). This can be interpreted that the network has *homeostasis* against the perturbation on the transcription complexes for *Sxl*.

Compared to Fig. 7 (b), the interpretation of (a) might be a bit difficult: especially the step from the state 00001000 to the last state 01010101 looks somewhat mysterious. The key to understand this behaviour is the previous two states (1111<u>1</u>1000 and 0101<u>1</u>1000) where gene *Sxl* is expressed in the female-specific isoform (underlined). Therefore the isoform is sufficiently stored while the supply of the transcription complex for the late promoter is instantaneously suspended (at the state 00001000). Thus the female-specific splicing of *Sxl* is re-started once the supply of the transcription complex for the late promoter is resumed and maintained (at the last state 01010101).

[3]We used GOAL (Tsay et al., 2007) to analyse Büchi automata.

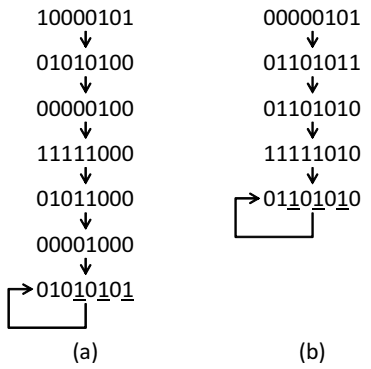

Figure 7: Example behaviours of the sex determination network in *Drosophila*. For simplicity the vector only consists of the values of the propositions $(TC_S^E, TC_S^L, on_S^m, on_S^f, on_t^m, on_t^f, on_d^m, on_d^f)$. (a) A female-specific behaviour. Underlined are female-specific isoforms. (b) A male-specific behaviour. Underlined are male-specific isoforms.

Next we check another property: this network is not able to keep female- and male-specific splicing simultaneously. This property is formally written as:

$$FG(on_S^f \wedge on_t^f \wedge on_d^f \wedge on_S^m \wedge on_t^m \wedge on_d^m)$$

The result is 'No', as we expected. There is no such behaviour in the possible behaviours of the network controlling sex determination. Both verifications show that this network ensures the sex determination.

The sex determination of *Drosophila* is also modelled and analysed in (Louis et al., 2003). In that work, the authors model the behaviour of the expression of gene *Sxl* and analyse how the bistable female-specific and male-specific differences arise. They use ODE models for the transcription of *Sxl* from the early promoter and probabilistic models for that of the late promoter. The two models are combined to employ the overall analysis. Their model is elaborated and requires deep knowledge in molecular mechanisms. In addition, the mathematical inference of unavailable kinetic parameters is required. For simplicity they abstracted the downstream genes of *Sxl* such as *tra* and *dsx*. We guess the reason of this simplification in the modelling and analysis as that the cascading the quantitative model makes the entire model very sensitive to the changes of parameter values in each stage and to 'correct' inference of the uncertain parameters is more crucial to reproduce the sex determination. In contrast, our qualitative framework allows us to model the network concisely with the same conclusion (bistability of the network). The downstream genes of *Sxl* are also included in the analysis without difficulty. We must note, however, that their *quantitative* model enables the robustness analysis such that how the sex determination works against the modulation of gene doses. Due to qualitativeness of our formalism, such analysis is infeasible in our framework. We, however, showed that both female-specific and male-specific gene expressions cannot be maintained, i.e. the possible behaviours of the network does not contain such behaviours. To prove this with quantitative model is very difficult since we need to test all combinations of possible quantitative parameters and confirm that such behaviours are never produced.

(Fear et al., 2015) modelled the sex determination regulatory network of *Drosophila* using structural equation models. The purpose of their work is to infer statistically likely links between genes in the known network or to find new genes which can likely be included in the network, rather than to investigate how the sex determination is ensured by the network. Although the aim of modelling the network is different from ours, the prediction of plausible extension of the known network is interesting aspect of systems biology. This line of research will be future work in our formalism. Whereas their method for finding plausible extension is brute force: they enumerate all possible interactions and insertion of new genes in all possible locations in the graph, in our framework some logical inference may be helpful to find plausible extensions of the network instead of using brute force.

5 CONCLUSION

In this paper we presented a qualitative framework to model and analyse gene networks using linear temporal logic. We studied molecular mechanisms of alternative splicing and showed how such mechanisms are modelled in our framework. As a demonstration, we modelled the network of sex determination in *Drosophila* and checked the sex-specific bistability of the network.

Since this work is still at a theoretical stage, we are investigating applications of our framework to real biological problems. For this we are to develop (semi-)automatic modelling method from gene regulation information. Compared to quantitative approaches which need manual parameter inference or tuning, a (semi-)automatic model construction of a gene regulatory network is more feasible in our framework. The start point will be to devise a machine readable uniform presentation of splicing networks. One promising approach is to extend SBML Qualitative Models Package (Chaouiya et al., 2013).

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number 26730153.

REFERENCES

Aoshima, T. (2003). *On a verification System for Reactive System Specifications*. PhD thesis, Tokyo Institute of Technology.

Batt, G., Salah, R. B., and Maler, O. (2007). On timed models of gene networks. In *FORMATS 2007*, volume 4763 of *LNCS*, pages 38–52.

Camara, N., Whitworth, C., and Van Doren, M. (2008). The creation of sexual dimorphism in the Drosophila soma. *Curr. Top. Dev. Biol.*, 83:65–107.

Chaouiya, C., Bérenguier, D., Keating, S. M., Naldi, A., van Iersel, M. P., Rodriguez, N., Dräger, A., Büchel, F., Cokelaer, T., Kowal, B., Wicks, B., Gonçalves, E. J. V., Dorier, J., Page, M., Monteiro, P. T., von Kamp, A., Xenarios, I., de Jong, H., Hucka, M., Klamt, S., Thieffry, D., Novère, N. L., Saez-Rodriguez, J., and Helikar, T. (2013). SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools. *BMC Systems Biology*, 7:135.

Ciocchetta, F. and Hillston, J. (2009). Bio-PEPA: A framework for the modelling and analysis of biological systems. *Theor. Comput. Sci.*, 410:3065–3084.

David, C. and Manley, J. (2008). The search for alternative splicing regulators: new approaches offer a path to a splicing code. *Genes Dev.*, 22(3):279–285.

Fear, J., Arbeitman, M. N., Salomon, M., Dalton, J., Tower, J., Nuzhdin, S. V., and McIntyre, L. M. (2015). The wright stuff: reimagining path analysis reveals novel components of the sex determination hierarchy in drosophila melanogaster. *BMC Systems Biology*, 9:53.

Heiner, M., Gilbert, D. R., and Donaldson, R. (2008). Petri nets for systems and synthetic biology. In *SFM 2008*, volume 5016 of *LNCS*, pages 215–264.

Hertel, K. (2008). Combinatorial control of exon recognition. *J. Biol. Chem.*, 283(3):1211–1215.

Ito, S., Hagihara, S., and Yonezaki, N. (2014). A qualitative framework for analysing homeostasis in gene networks. In *Proceedings of BIOINFORMATICS 2014*, pages 5–16.

Ito, S., Ichinose, T., Shimakawa, M., Izumi, N., Hagihara, S., and Yonezaki, N. (2013a). Modular analysis of gene networks by linear temporal logic. *J. Integrative Bioinformatics*, 10(2).

Ito, S., Ichinose, T., Shimakawa, M., Izumi, N., Hagihara, S., and Yonezaki, N. (2013b). Qualitative analysis of gene regulatory networks using network motifs. In *Proceedings of BIOINFORMATICS 2013*, pages 15–24.

Ito, S., Ichinose, T., Shimakawa, M., Izumi, N., Hagihara, S., and Yonezaki, N. (2015). Qualitative analysis of gene regulatory networks by temporal logic. *Theor. Comput. Sci.*, 594(23):151–179.

Ito, S., Izumi, N., Hagihara, S., and Yonezaki, N. (2010). Qualitative analysis of gene regulatory networks by satisfiability checking of linear temporal logic. In *Proceedings of BIBE 2010*, pages 232–237.

Kornblihtt, A. (2005). Promoter usage and alternative splicing. *Curr. Opin. Cell Biol.*, 17(3):262–268.

Louis, M., Holm, L., Sánchez, L., and Kaufman, M. (2003). A theoretical model for the regulation of sex-lethal, a gene that controls sex determination and dosage compensation in drosophila melanogaster. *Genetics*, 165(3):1355–84.

Matlin, A., Clark, F., and Smith, C. (2005). Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.*, 6(5):386–398.

Palsson, B. (2000). The challenges of in silico biology. *Nat. Biotechnol.*, 18:1147–50.

Salz, H. K. and Erickson, J. W. (2010). Sex determination in Drosophila: The view from the top. *Fly*, 4:60–70.

Thomas, R. (1991). Regulatory networks seen as asynchronous automata: A logical description. *J. Theor. Biol.*, 153(1):1–23.

Tsay, Y.-K., Chen, Y.-F., Tsai, M.-H., Wu, K.-N., and Chan, W.-C. (2007). GOAL: a graphical tool for manipulating Büchi automata and temporal formulae. In *Proceedings of the 13th international conference on Tools and algorithms for the construction and analysis of systems*, volume 4424 of *TACAS'07*, pages 466–471, Berlin, Heidelberg. Springer-Verlag.

Vardi, M. Y. and Wolper, P. (1994). Reasoning about infinite computations. *Inf. Comput.*, 115:1–37.

Wen, J. (2013). *Computational modeling and inference of alternative splicing regulation*. PhD thesis, University of Miami.