Hybrid Gene Regulation Models of Mammalian Circadian Cycles

Lelde Lace[a], Karlis Cerans[b], Karlis Freivalds, Gatis Melkus and Juris Viksna[c]
Institute of Mathematics and Computer Science, University of Latvia, Raina bulvaris 29, Riga, Latvia

Keywords: Gene Regulatory Networks, Hybrid System Models, Circadian Cycle Modelling, Behavioural Periodicity.

Abstract: We present hybrid system based gene regulation models of mammalian circadian cycle and the results of model behaviour analysis. The models cover genes of two recently proposed biological models with 5 and 3 gene 'core oscillators'. The advantage of the used HSM framework is limited model dependence on parameter values, which are described only at qualitative level at the extent they affect models' observable behaviour. The models represent gene regulatory networks in terms of genes, proteins, binding sites, regulatory functions, and constraints on growth rates and binding affinities. Although such models do provide limited accuracy, they are less dependent from parameter fitting and can provide predictions on some biological aspects of gene regulation that are not dependent form the choice of particular parameter values. The models can provide biologically feasible predictions about synchronised oscillation of the involved genes and functions that regulate gene activity on basis of regulatory network topology alone. The work also includes developments of new analysis methods, in particular, for analysis of available trajectories in HSM state spaces and derivation of constraints that are needed for state transition trajectories to satisfy the required specific properties.

1 INTRODUCTION

Circadian rhythm is well known process of gene activity variation during a 24 hour cycle in response to external stimulus, such as light or heat, which is usually referenced to as 'Zeitgeber'. The cyclic process of gene activity variation in general is self-sustaining, with Zeitgeber acting as synchronizer of the activity to 24 hours long period. In good level of detail the underlying gene regulatory processes of circadian rhythm have been studied for a variety of organisms, in particular for mammals, for which several consistent and partially overlapping models of gene activity regulation have been proposed. More formalised models based on differential equations have been developed (Korencic et al., 2014) as well.

The focus of our study is modelling of these mammalian circadian cycle regulatory processes using hybrid system model (HSM) based formalism (Brazma et al., 2015), with the aim to understand the suitability of HSM based approach for description of mammalian circadian cycle gene regulatory process. We have developed a number of concrete models for this purpose, and have evaluated how well these models can replicate experimentally known data, and how useful they might be for better understanding of biological mechanism driving circadian cycle rhythm.

The development of HSM framework (Brazma et al., 2013; Brazma et al., 2015) was motivated by analysis of lambda phage virus model described initially by Finite State Linear Model formalism (Brazma and Schlitt, 2003; Schlitt and Brazma, 2006). The main conclusion drawn from this analysis was relative unimportance of assumptions about concrete numerical model parameters and it was recognised that qualitative (experimentally measurable) behaviour of the model depends only on comparative relations between growth functions and binding site affinities and not on their exact form or values. Replacing these functions and parameter values with set of constraints provides the initial HSM frame of the model. One of the achievements of HSM framework was lambda phage model, for which it was shown that the only possible stable behaviours (described by attractor regions in state spaces) well correspond to biologically known *lysis* and *lysogeny* processes, and that, according to the model, no other types of behaviour can occur. A number of hypotheses regarding constraints of the model have also been derived allowing to propose experiments of virus genome rearrangements have been proposed that could allow ei-

[a] https://orcid.org/0000-0001-7650-2355
[b] https://orcid.org/0000-0002-0154-5294
[c] https://orcid.org/0000-0003-2283-2978

Lace, L., Cerans, K., Freivalds, K., Melkus, G. and Viksna, J.
Hybrid Gene Regulation Models of Mammalian Circadian Cycles.
DOI: 10.5220/0010834400003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 3: BIOINFORMATICS*, pages 130-137
ISBN: 978-989-758-552-4; ISSN: 2184-4305

ther to validate or refute these hypotheses (Ruklisa et al., 2019).

Since circadian rhythms are ones of the best studied and best understood gene regulatory processes, they are are well suited for further assessing of capabilities of the modelling framework and for development of its extensions.

2 CIRCADIAN CYCLE GENE REGULATORY PROCESSES

In this Section we briefly summarise the main details on the biologically widely acknowledged mechanism of mammalian circadian regulation.

The circadian system is composed of a series of molecular oscillators. These oscillators are realized through feedback loops of transcriptional regulation involving several core clock components. The primary feedback loop is based on the action of CLOCK and Bmal1 transcription factors which form a heterodimer that positively regulates several genes via the E-box cis-regulatory sequence, among which are the *Period* or *Per* and *Cryptochrome* or *Cry* genes whose products dimerize into a PER:CRY dimer that acts as the negative component of the loop, shutting off the transcription of both CLOCK and BMAL1 (and, through these, its own transcription). A secondary feedback loop is also present, and involves CLOCK:Bmal1-mediated transcriptional activation of retinoic acid-related orphan nuclear receptors $RevErb\alpha$ and $Ror\alpha$ which bind competitively to sequences known as retinoic acid-related orphan receptor response elements (ROREs or RREs), which further modulate Bmal1 expression through either Ror-mediated activation or RevErbα-mediated repression (Ko and Takahashi, 2006; Morris et al., 2020; Ueda et al., 2005). There is also an additional regulatory mechanism at play realized via transcription factor DBP which binds to the D-box cis-regulatory sequence, causing its own activation events independent of RREs or the E-box, but the physiological role of this activation remains unclear, though it is hypothesized to modulate both the output and input of the circadian system (Morris et al., 2020; Yoshitane et al., 2015). Together, with some additional functionality imparted by post-translational modifications these molecular feedback loops produce an activation/repression cycle approximately 24 hours in length: the mammalian circadian clock (Ko and Takahashi, 2006; Morris et al., 2020; Ueda et al., 2005).

Modelling this clock is non-trivial. Though simplified models can sometimes omit them, even the core regulatory circuitry of the circadian clock includes over a dozen core clock genes, and the mammalian circadian clock is known to have hundreds of molecular inputs that modulate the various states of the clock and thousands of outputs that the cycle is able to regulate (Morris et al., 2020; Ueda et al., 2005). Beyond this, the circadian clock genes also have tissue-specific effects potentially linked to hundreds of clock-controlled genes (CCGs). This creates fertile ground for explaining tissue-specific patterns of gene expression with circadian clock models of varying levels of sophistication, allowing for reasonably accurate (though by no means complete) explanations of these patterns even with relatively simple 5-gene models encompassing the interplay of some core clock genes (Korencic et al., 2014). Mathematical modelling of these patterns can reveal the principal dynamics of the gene regulatory network underpinning the circadian clock, such as the idea that while the circadian clock is ultimately complex and tissue-specific in its action, the minimal requirements for keeping its rhythmic oscillation and thus the core element of all circadian clocks specific to tissues might in fact amount to a single repressilator motif (three genes that each inhibit the next gene and are inhibited by the previous gene in a loop) comprising three genes – *Per*2, *Cry*1 and $RevErb\alpha$ (Pett et al., 2016).

The approaches described above inform our approach in this work, where we hope to model the circadian clock with a minimal array of core clock genes in order to analyse their interplay and validate our modelling approach as a useful one in approximating the properties of the circadian clock.

3 MODELLING FRAMEWORK

For describing models of gene regulatory networks we use HSM framework (for a detailed technical description see (Brazma et al., 2015)). The models are specified by 6-tuples $\mathcal{H} = \langle M, X, C, T, F, MF \rangle$, where $M = \{\mu_1, \ldots, \mu_k\}$ is a set of modes, $X = \{x_1, \ldots, x_m\}$ is a set of continuous variables with real non-negative values, $C = \{c_1, \ldots, c_r\}$ is a set of real non-negative transition thresholds, T is a set of mode transitions (in which transitions have form $\tau = \alpha \rightarrow_{x \leq c} \beta$ or $\tau = \alpha \rightarrow_{x \geq c} \beta$, where $\alpha, \beta \in M$, $x \in X$, $c \in C$), $F = \{f_1, \ldots, f_n\}$ is a set of continuous and monotonous growth/degradation functions, and $MF : M \times X \rightarrow F$ is a mapping providing *mode-function assignments* assigning to each mode $\alpha \in M$ and each variable $x \in X$ a function $g \in F$. This is consistent with other widely used hybrid system definitions, but imposes additional restrictions in order to keep the formalism as simple as possible for analysis purposes, and at the

same time still to provide sufficient modelling power for describing known biological processes.

Intuitively modes from M of HSM are meant to represent 'states' on uneventful evolution of biological system during which no observable 'events' (such as occupation or vacation of a regulatory binding site, molecular interactions that can potentially change the system's behaviour etc.) occur. Variables from X describe concentration levels of biological substances (such as proteins), and T describe allowed transitions between the modes that can be triggered by substance concentrations reaching specific thresholds or dropping below them. Functions from F describe changes of substance concentrations with time, for each of the substances these changes are mode-specific, the behaviour at each of the nodes is specified by mode-function assignment MF.

We assume that a biological system is described by a deterministic HSM model, in which substance rate changes are governed by well-defined functions from F. At the same time, we also assume that our knowledge about the modelled system is limited to what can be experimentally observed about the system's behaviour, i.e. whether concentration rates are growing or decreasing and (possibly) at what rates. This uncertainly allows to model not strictly deterministic behaviour of system (a more realistic assumption from biological perspective) and incorporate such features as time delays and randomness.

In formal terms our knowledge about HSM that is limited only to qualitative and not precise quantitative information about substance concentrations that can be represented by a notion of HSM *frame* $\mathcal{F}(\mathcal{H})$ in which concrete functions from F are replaced by values from $\{\nearrow, \searrow, \rightarrow\}$ that indicate whether concentrations are growing, decreasing, or remaining constant (usually at zero and saturation levels). Such replacement can change a strictly deterministic time evolution of the model states to very non-deterministic state evolution which will include mode transitions that contradict the known biological evidence. However, at qualitative level known experimental evidence can be incorporated in the HSM frame in form of constraints on allowed mode transitions.

At technical level this is described it terms of *constrained frames*, which additionally impose orderings on mode transitions. The general modelling workflow is as follows: 1) start with HSM model frame that is constructed based on the known biological knowledge and incorporates known ordering constraints; 2) analyse the set of all constrained state spaces that are built in accordance to different ordering constraint assumptions and partition them into equivalence classes; 3) select the set of equivalence classes in which state trajectories are consistent with experimentally known behaviour as hypothetically valid models of the biological system. The classes of state spaces that are not consistent with experimentally measured behaviour can be used to derive additional constraints on gene regulation that *in principle* can be validated by experiments (thus, HSM models potentially can provide testable predictions that can be used either to reconfirm or to refute the model validity).

Given a constrained HSM frame model for a gene regulatory network, we can simulate the network behaviour by choosing any set of appropriate functions F that satisfy the given constraints. The main advantage of hybrid models, however, is the possibility to analyse the entire range of possible dynamic behaviours of the modelled system. We are interested in the following questions about the modelled system.

Stable behaviour Regions in Model State Space. For Boolean models the regions of stable behaviour are simply described by *attractors* – simple cycles in state space graph. Cyclic attractors, however, can be obtained only for state spaces with a single outgoing transition for each vertex. The most straightforward generalisation of attractors in general digraphs can be obtained by partitioning state spaces in *strongly connected components* (SCC). In addition we can exclude SCCs for which we can prove that the state evolution cannot remain within them for infinitely long time – this will be the case if there is a gene that is active (inactive) in all SCC vertices, and for all vertices there is transition that is triggered by the protein corresponding to this gene reaching (dropping below) certain threshold. If this is the case we call SCC *transitional* and define as *attractors* all the remaining SCCs.

A given HSM model usually will provide only partial constraints on transitions and there can be a large number of different sets of full constraints and different state spaces that correspond to them. In this case the task of analysis is partitioning the state space set into equivalence classes up to isomorphism of attractor SCCs and assessing the biological merit and validity of each class separately.

Switching Conditions that Irrevocably Leads the System to reaching a Single Region of Stability. We are interested to identify sets of states with corresponding transitions such that there are fewer reachable attractors from destination vertex than there are from the source vertex. Such states can be grouped together according to attractors reachable from them and genes that trigger the discriminating transitions. In general case these vertex groups will form a decision forest.

Available Trajectories in State Spaces. An important question for assessing model validity is whether

available trajectories in state space are consistent with the known experimental data. Queries about the available trajectories can be naturally described by regular expressions and the state space itself can be regarded as Non-deterministic Finite Automaton (with given pair of vertices specifying initial and accepting states) and will satisfy the requirement on trajectories if and only if languages defined by query expression and state space automaton are equal. The problem is **NP**-complete, however, the available algorithms are sufficiently fast for the current use cases.

In the context of previously developed HSM models for phage viruses the most important part of analysis was identification of attractors (stability regions) in model state spaces – this led to dramatic reduction of number of states that merits further analysis (while the initial state spaces contained up to hundreds of thousands of nodes, the attractor regions were limited to few tens of nodes). Trajectory analysis has played a part, but only in providing suggestions whether models with different attractor topologies can be differentiated on basis of additional experimental measurements. For circadian cycle models (already by their design) most of the states are expected to belong to single attractor region (which, in this case, fortunately tend to be small). This also exclude the need for switching condition analysis, since there are none. The model assessment thus is largely based on analysis of available trajectories in model state spaces and deriving constraints that are needed to exclude specific undesirable behaviours. As noted, this task is closely related to automata equivalence problem, but for comparatively small state spaces of circadian cycle models that we have proposed the analysis still has been done semi-manually. More exact formal description of this problem and development of algorithms for automated analysis remains an interesting research challenges.

4 CIRCADIAN CYCLE MODELS

For representation in HSM framework we have chosen five gene 'core oscillator' formed by genes *Bmal*1, *Per*2, *RevErb*, *Cry*1 and *Dbp* proposed and analysed in very detailed level in (Korencic et al., 2014). The model is based on numerous semi-formal mammalian circadian cycle models developed and published earlier (Relogio et al., 2011; Ukai and Ueda, 2010; Leloup and Goldbetter, 2003). In general these models are in good agreement about the underlying mechanism behind the circadian cycle, but usually consider a larger set of involved genes. At the same time, tissue specificity of expression of genes involved in circadian rhythm is also well known, and the proposed core oscillator was designed to cover the genes whose oscillatory behaviour is observed among all the tissues, and to be self-sufficient to dictate expression patterns of other genes involved in circadian regulatory networks. The model was further fine-tuned (Pett et al., 2016; Schmal et al., 2019), including identification and analysis of smaller three gene *Bmal*1, *Per*2 and *RevErb* sub-network.

These models (genes and their regulatory interactions) are shown in Figure 1. The external Zeitgeber regulator acts as a synchronisation factor, but is not required for oscillatory behaviour.

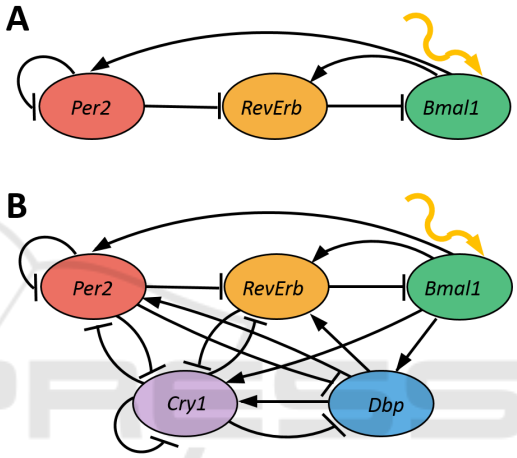

Figure 1: Core oscillators of mammalian circadian cycles with 3 genes (A) and 5 genes (B). Activation or inhibition regulatory properties of specific genes are well established, however, from modelling perspective, the exact form of regulatory functions can be a subject of interpretation.

In order to represent the model in HSM framework two important things that have to be decided are: 1) level of details that will be represented by the model, and 2) concrete forms of regulatory functions for each of the genes.

Regarding the level of details that should be included, for bacteriophage models important role has differences in same protein binding affinities at different sites. For circadian networks, however, there is no evidence about notably distinct affinities of sites binding the same proteins. Moreover, at core oscillator model's abstraction level several biologically similar genes are grouped together, thus binding site affinities may not have clear biological meaning. High level of connectivity in circadian regulatory network will also allow many very different affinity assignments to match a specific desired model behaviour, thus severely limiting the usefulness and credibility of such type of models. For these reasons our proposed HSM models include just a single binding site

for each protein.

Regarding the choice of exact form of regulatory functions, a natural constraint is that biologically acknowledged positive and negative regulatory actions of specific genes have to be respected – thus a regulatory formula for gene G with activating regulators $A_1,\ldots,A_k$ and repressing regulators $R_1,\ldots,R_l$ should be given as a Boolean expression of conjunctions $\wedge$ and disjunctions $\vee$ with variables $A_i, i = 1\ldots k$ and $\neg R_j, j = 1\ldots l$. This still leaves many candidates for multi-variable formulas, although without additional assumptions a conjunction of all involved variables appears to be the most widely used choice. There is, however, a convincing evidence for grouping certain regulatory factors together in what are known as E-boxes, D-boxes and RRE-s. From five gene core oscillator model two genes *Per*2 and *Cry*1 are included in a single E-box, thus it is useful also to consider the case where for transcription repression both of these genes have to be active.

Based on these considerations we have constructed two different 5 gene HSM models **Circadian5A** and **Circadian5B**. **Circadian5A** model contains 5 genes *Bmal*1, *Per*2, *RevErb*, *Cry*1 and *Dbp*, a single binding site for each of the genes, and uses conjunctions for all gene regulatory functions:
$Bmal1 = \neg RevErb$,
$Per2 = Bmal1 \wedge Dbp \wedge \neg Per2 \wedge \neg Cry1$,
$RevErb = Bmal1 \wedge Dbp \wedge \neg Per2 \wedge \neg Cry1$,
$Dbp = Bmal1 \wedge \neg Per \wedge \neg Cry1$,
$Cry1 = Bmal1 \wedge Dbp \wedge \neg RevErb \wedge \neg Per2 \wedge \neg Cry1$.

Circadian5B model assumes that *Per*2 and *Cry*1 can act as repressors only when both of these genes are active (i.e. all genes from a single E-box must be present for regulatory activity:
$Bmal1 = \neg RevErb$,
$Per2 = Bmal1 \wedge Dbp \wedge (\neg Per2 \vee \neg Cry1)$,
$RevErb = Bmal1 \wedge Dbp \wedge (\neg Per2 \vee \neg Cry1)$,
$Dbp = Bmal1 \wedge (\neg Per2 \vee \neg Cry1)$,
$Cry1 = Bmal1 \wedge Dbp \wedge \neg RevErb \wedge (\neg Per2 \vee \neg Cry1)$.

Model **Circadian3** contains only 3 genes *Bmal*1, *Per*2, *RevErb*. With at most two regulators per gene there is only a single natural choice for each of regulatory functions:
$Bmal1 = \neg RevErb$,
$Per2 = Bmal1 \wedge \neg Per2$,
$RevErb = Bmal1 \wedge \neg Per2$.

5 RESULTS

The main conclusions that have been obtained from the proposed differential equation models (Korencic et al., 2014; Pett et al., 2016) are that they allow to model gene activity oscillation changes consistently with experimentally measured values. The models focus on peak concentrations of oscillation phases, that according to models are observed at the following sequence: *Bmal*1, *RevErb*, *Dbp*, *Per*2, *Cry*1. The exact peak ordering, however, seems not very strongly defined and can depend from the values of parameters (which include explicit timing delays) that have been fitted for specific version of the model. In particular, the order of *Per*2 and *RevErb* concentration peak offsets in relation to *Bmal*1 can differ between the models and there appears to be lack of experimental evidence for preference of one phase ordering versus another for these two genes (Schmal et al., 2019). Moreover, model simulation results indicate that phase order of concentration minimums can be different from the order of concentration peaks (e.g. can be swapped for *Bmal*1 and *Cry*1), and there is also experimental evidence that phase order of concentration peaks for some genes varies between different cell types. There is very strong evidence, however that concentrations growths alternate for *Bmal*1 and *Per*2/*RevErb*, with timing of *Bmal*1 concentration peak approximately coinciding with *Per*2 and *RevErb* concentration minimums.

In the context of analysis of our hybrid system based models it should be noted, that the HSM formalism is actually designed for assessing the modelled system's behaviour as independently from the exact values of protein concentrations as it possibly can, thus the models are not necessarily expected to provide information about the exact timing of concentration maximum and minimum values. The explicitly modelled, however, are state transitions triggered by association and dissociation of proteins from their binding sites. Thus the minimal expectations for validity and usefulness of the models are the following: 1) all involved genes must participate in regulatory activities; 2) the model state space should contain cyclic trajectories that for each of the genes involve exactly one transition triggered by its growth and exactly one transition triggered by its decrease. A particular transition trajectory will place also certain constraints on phase orders on concentration peaks and drops, but generally several alternatives for such order shall be possible.

The initial step of model analysis consists of identification of attractors. This has been a crucial step for phage virus models that we have analysed previously, reducing the sizes of state spaces from tens of thousands to few tens of nodes. For circadian cycle models, however, we are interested in properties of cyclic trajectories in model state spaces, which by definition will belong to attractor regions, thus the re-

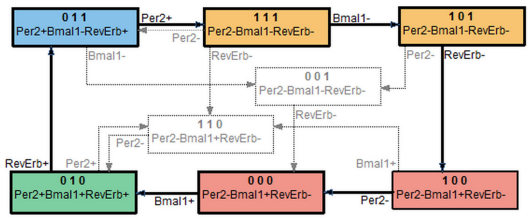

Figure 2: The stable region of state space of 3 gene circadian cycle model. The dotted transitions and states can excluded by priority constraints, leaving as the only possibility a single cyclic trajectory of 6 states. This is consistent with the observed periodic activity oscillations.

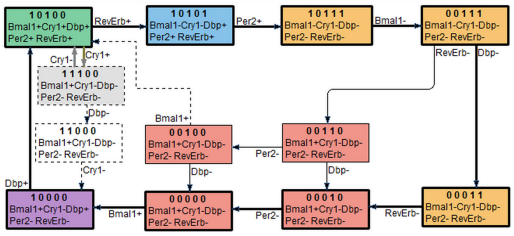

Figure 3: The stable region and available trajectories in 5 gene model with independent repressory activity of *Per* and *Cry1* genes. The periodic activity oscillation of all 5 genes in principle is feasible according to the model, but requires synchronised traversing of two interlocked cycles.

duction of state space sizes, whilst useful, is less significant. The state spaces consist of 8 nodes for **Circadian3** model and 32 nodes for **Circadian5A** and **Circadian5B** models, with single attractor regions of sizes 8, 18 and 32 correspondingly.

The next step involves applying constraints to exclude some transitions and nodes that during this step become disconnected. Here use two priority constraints: 1) transitions triggered by *RevErb* are prioritised over transitions triggered by *Per*2; and 2) transitions triggered by a gene that have been growing or decreasing in two or more consecutive states are prioritised over transitions that are triggered by gene growing or decreasing only at transition origin's state.

The second of these constraints follows from a natural assumption about transition delays – a protein concentration needs some time either to grow or to decrease to make a change in the modelled system's state. The first constraint can be interpreted either as higher concentration change rate or lower binding threshold for *RevErb* compared to *Per*2. In combination both these priority constraints are well consistent with explicit conditions on transition delays that are imposed on ODE models.

With attractor regions further restricted by these constraints the **Circadian3** model state space retains 6 nodes linked in a simple cycle with the following transition triggering order: $Bmal1+$, $RevErb+$, $Per2+$, $Bmal1-$, $RevErb-$, $Per2-$ (Figure 2). Here $''+''$ and $''-''$ signs after the gene names indicate whether the transition is triggered correspondingly either by growth or decrease of the product of the particular gene. The model predicts alternating and synchronised oscillation patterns of genes *Bmal*1 and *Per*2/*RevErb*, consistent with known experimental data.

In five gene model **Circadian5A** with independent repressory activities of *Per*2 and *Cry*1 genes the attractor part of state space consists of 18 nodes and is further reduced to 12 nodes by the applied reduction rules (Figure 3). The model as one of behavioural trajectories allows the same 6 state transition cycle that is present in **Circadian3** with genes *Dpb* and *Cry*1 not being involved in regulatory processes. There are, however, several alternative 8-transition cycles that additionally include regulatory activity by *Dbp*. A more problematic is inclusion of *Cry*1 in regulatory activities – this can be achieved only by requiring transitions to follow in alternating order one of the 8 transition cycles and a single 2 state cycle triggered by *Cry* regulatory activity. Although such behaviour in principle is possible according to the model, the lack of any synchronisation between *Cry*1 and other gene regulatory activities makes its biological accuracy very questionable.

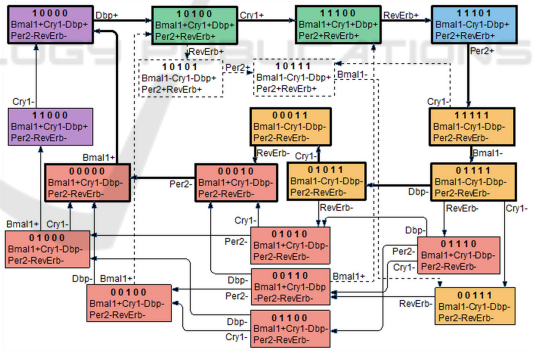

Figure 4: The stable region and available trajectories in 5 gene model with repressory action requiring co-activity *Per* and *Cry1* genes. The model allows only cyclic trajectories with periodic activity oscillation of all 5 involved genes, consistently with biologically observed behaviour.

The situation is significantly more promising for **Circadian5B** model. The attractor contains all 32 state space nodes, application of reducing rules removes some of transitions, but not the nodes (Figure 4). The model allows 18 different types of 10 state transition cycles (Table 1). Here transitions are referenced by the first letter of the gene triggering them, in upper case if triggered by growth and in lower case if triggered by decrease of the gene's product – e.g.

B stands for *Bmal*1+ and *b* for *Bmal*1−). The model does not give preference for each of these cycle types, but of particular interest is cycle in the 4th row of the table: *Bmal*1+, *Dbp*+, *Cry*1+, *RevErb*+, *Per*2+, *Bmal*1−, *Dbp*−, *Cry*1−, *RevErb*−, *Per*2−, as it is the only cycle that includes identical phase offsets for activation and deactivation transitions triggered by each of the genes.

Table 1: All the possible sequences of gene activity state transitions according to **Circadian5B** model.

	1	2	3	4	5	6	7	8	9	10
1	*B*	*D*	*C*	*R*	*P*	*b*	*c*	*d*	*r*	*p*
2	*B*	*D*	*C*	*R*	*P*	*b*	*c*	*r*	*d*	*p*
3	*B*	*D*	*C*	*R*	*P*	*b*	*c*	*r*	*p*	*d*
4	*B*	*D*	*C*	*R*	*P*	*b*	*d*	*c*	*r*	*p*
5	*B*	*D*	*C*	*R*	*P*	*b*	*d*	*r*	*p*	*c*
6	*B*	*D*	*C*	*R*	*P*	*b*	*d*	*r*	*c*	*p*
7	*B*	*D*	*C*	*R*	*P*	*b*	*r*	*c*	*d*	*p*
8	*B*	*D*	*C*	*R*	*P*	*b*	*r*	*c*	*p*	*d*
9	*B*	*D*	*C*	*R*	*P*	*b*	*r*	*p*	*d*	*c*
10	*B*	*D*	*C*	*R*	*P*	*b*	*r*	*p*	*c*	*d*
11	*B*	*D*	*C*	*R*	*P*	*b*	*r*	*d*	*p*	*c*
12	*B*	*D*	*C*	*R*	*P*	*b*	*r*	*d*	*c*	*p*
13	*B*	*D*	*C*	*R*	*c*	*P*	*b*	*d*	*r*	*p*
14	*B*	*D*	*C*	*R*	*c*	*P*	*b*	*r*	*d*	*p*
15	*B*	*D*	*C*	*R*	*c*	*P*	*b*	*r*	*p*	*d*
16	*B*	*c*	*D*	*C*	*R*	*P*	*b*	*r*	*p*	*d*
17	*B*	*c*	*D*	*C*	*R*	*P*	*b*	*d*	*r*	*p*
18	*B*	*c*	*D*	*C*	*R*	*P*	*b*	*r*	*d*	*p*

In general this behaviour can be considered reasonably consistent with predictions obtained from ODE models, taking into account that in HSM framework the focus is on the order of transitions between gene regulatory states rather than offsets of concentration peak phases (some of which appear to be more the result of fitted parameter values rather than based on experimental observations). As a strength of our **Circadian5B** model should be emphasised the fact that it predicts synchronised oscillation of all 5 genes of core oscillator as the only behaviour that is possible according to the model, independently from the values of any numerical parameters. Another strength is the fact that comparison of both 5 gene circadian cycle models indicates strong preference for collaborative repressory action of *Per*2 and *Cry*1 genes, which seems to be well supported by experimental evidence.

The changes of gene concentrations that are consistent with *Bmal*1+, *Dbp*+, *Cry*1+, *RevErb*+, *Per*2+, *Bmal*1−, *Dbp*−, *Cry*1−, *RevErb*−, *Per*2− transition cycle are shown in Figure 5. Growth and degradation of protein concentrations are shown as linear functions for illustrative purpose only – according to the model these can be any monotone functions reaching their maximal and minimal values at the shown order (and not providing more detailed predictions about the order at points in which depicted maximums and minimums for several genes coincide). The figure also does not provide any indications about the absolute concentration values of proteins nor any comparison of concentrations rates between different proteins. Since the model uses the same regulatory functions for *Per*2 and *RevErb* genes, their concentration changes are shown to be identical.

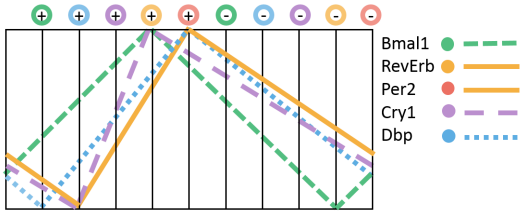

Figure 5: Oscillations of gene expression activities that are consistent with available trajectories in HSM model.

The predictions that can be obtained about protein concentration peak at minimum phases, assuming that this particular 10 state transition cycle persistently holds, are the following. For maximal values the peaks of proteins *Bmal*1 and *Cry*1 must strictly alternate with peaks of proteins *RevErb*, *Per*2 and *Dbp* (without giving specific precedence for order of peak phases in these two groups – these are also allowed to vary between different cycles). The predicted sequence of minimal value phases is *Bmal*1, followed by *Dpb* and then the group *RevErb*, *Per*2 and *Cry*1. This is largely consistent with ODE simulation results as shown in (Pett et al., 2016). In particular, the HSM model allows the same ordering of maximal peak phases, and (quite interestingly) also suggests lower degradation rate for *Cry*1 in comparison to *Bmal*1, which also seems to be implied by ODE simulations. An outlier in this aspect is phase of minimal value of *Dbp*, which has different offset in comparison to ODE model predictions. As we already have noted, however, the strengths of HSM based models is prediction of state transition sequences, and limited information can be derived about behaviour of functions describing protein concentration changes.

6 DISCUSSION

The aim of this work was to assess the suitability of HSM modelling framework for description of mammalian circadian cycle gene regulatory process, to develop a number of concrete models for this purpose, and to evaluate how well these models perform to

replicate experimentally known data, and how useful they might be for better understanding of biological mechanism driving circadian cycle rhythm.

The proposed models can be regarded as successful at these aspects. The models 3 gene **Circadian3** and 5 gene **Circadian5B** core oscillator models demonstrate cyclic activity switching for all the included genes as the only type of behaviour that is allowed under assumptions of these models. Moreover, the models allow only strictly synchronised activity oscillations for all the involved genes. Such result might be less surprising for **Circadian3** model, where the potential of cyclic oscillations of all the 3 genes can be deduced from gene regulation 'wiring diagram' (Figure 1), still the fact that oscillations involving only 1 or 2 genes can be excluded by simple priority constraints was not immediately obvious.

For 5 gene model we can derive a concrete prediction regarding gene regulatory functions – namely, that to obtain biologically feasible model behaviour genes *Per*2 and *Cry*1 should act in collaboration for their repressory activity. Such collaborative action of these genes is well supported by biologically known facts about circadian cycle gene regulatory mechanism. The corresponding model **Circadian5B** also predicts synchronised oscillation of all 5 genes of core oscillator as the only behaviour that is possible according to the model. There are several different sequences of transition orders that can lead to this behaviour, with one of them in reasonably good agreement with simulation results obtained from ODE based models. There is a also good potential to refine this model further (i.e. by explicitly including in the model several different binding site affinities for the same protein, introducing more complex regulatory functions, or additional priority constraints), provided that there are some experimental/biological data supporting specific modifications.

In a broader context HSM models for circadian cycle that we describe here are the first attempt to apply HSM framework for modelling gene regulation in more complex organisms, since all the previous models have been designed for viruses. As a use case of modelling more complex organisms our proposed models appear to be successful. They also allow to identify challenges and research priorities that need to be considered for development of such type of models.

REFERENCES

Brazma, A., Cerans, K., Ruklisa, D., Schlitt, T., and Viksna, J. (2015). Modeling and analysis of qualitative behavior of gene regulatory networks. *Lecture Notes in Computer Science*, 7699:51–66.

Brazma, A. and Schlitt, T. (2003). Reverse engineering of gene regulatory networks: a finite state linear model. *Genome Biol.*, 4:P5:1–31.

Brazma, R., Cerans, K., Ruklisa, D., Schlitt, T., and Viksna, J. (2013). HSM – a hybrid system based approach for modelling intracellular networks. *Gene*, 518:70–77.

Ko, C. and Takahashi, J. (2006). Molecular components of the mammalian circadian clock. *Human Molecular Genetics*, 15(S2):R271–R277.

Korencic, A., Kasir, R., Bordyugov, G., Lehmann, R., Rozman, D., and Herzel, H. (2014). Timing of circadian genes in mammalian tissues. *Scientific Reports*, 4:5782.

Leloup, C. and Goldbetter, A. (2003). Toward a detailed computational model for the mammalian circadian clock. *PNAS*, 100:7051–7056.

Morris, A., Stanton, D., Roman, D., and Liu, A. (2020). Systems level understanding of circadian integration with cell physiology. *J Mol. Biol.*, 432(12):3547–3564.

Pett, J., Korencic, A., Wessener, F., and Herzel, H. (2016). Feedback loops of the mammalian circadian clock constitute repressilator. *PLOS Comp. Biol.*, 12(12).

Relogio, A., Westermark, P., Wallach, T., Schellenberg, K., Kramer, A., and Herzel, H. (2011). Tuning the mammalian circadian clock: robust synergy of two loops. *PLOS Comp. Biol.*, 7(12).

Ruklisa, D., Brazma, A., Cerans, K., T., S., and Viksna, J. (2019). Dynamics of gene regulatory networks and their dependence on network topology and quantitative parameters – the case of phage lambda. *BMC Bioinformatics*, 20:296.

Schlitt, T. and Brazma, A. (2006). Modelling in molecular biology: describing transcription regulatory networks at different scales. *Philos. Trans. R. Soc. Lond. B*, 361:483–494.

Schmal, C., Ono, D., Myung, J., Pett, J., Honma, S., and Herzel, H. (2019). Weak coupling between intracellular feedback loops explains dissociation of clock gene dynamics. *PLOS Comp. Biol.*, 15(9):e1007330.

Ueda, H., Hayashi, S., Chen, W., Sano, M., Machida, M., Shigeyoshi, Y., Iino, M., and Hashimoto, S. (2005). System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nature Genetics*, 37(2):187–192.

Ukai, H. and Ueda, H. (2010). Systems biology of mammalian circadian clocks. *Ann. Rev. Physiology*, 72(1):579–603.

Yoshitane, H., Asano, Y., Sagami, A., Sakai, S., Suzuki, Y., Okamura, H., Iwasaki, W., Ozaki, H., and Fukada, Y. (2015). Functional D-box sequences reset the circadian clock and drive mRNA rhythms. *Communications Biology*, 2(1):300.