Modelling Signalling Networks with Incomplete Information about Protein Activation States: A P System Framework of the KaiABC Oscillator

Thomas Hinze[1], Thorsten Lenser[2], Gabi Escuela[2],
Ines Heiland[1], Stefan Schuster[1]

[1] Friedrich Schiller University Jena
School of Biology and Pharmacy, Department of Bioinformatics
[2] Department of Computer Science, Bio Systems Analysis Group
Ernst-Abbe-Platz 1–4, D-07743 Jena, Germany
{thomas.hinze,thorsten.lenser,gabi.escuela}@uni-jena.de
{heiland.ines,stefan.schu}@uni-jena.de

Summary. Reconstruction of signal transduction network models based on incomplete information about network structure and dynamical behaviour is a major challenge in current systems biology. In particular, interactions within signalling networks are frequently characterised by partially unknown protein phosphorylation and dephosphorylation cascades at a submolecular description level. For prediction of promising network candidates, reverse engineering techniques typically enumerate the reaction search space. Considering an underlying amount of phosphorylation sites, this implies a potentially exponential number of individual reactions in conjunction with corresponding protein activation states. To manage the computational complexity, we extend P systems with string-objects by a subclass for protein representation able to process wild-carded together with specific information about protein binding domains and their ligands. This variety of reactants works together with assigned term-rewriting mechanisms derived from discretised reaction kinetics. We exemplify the descriptional capability and flexibility of the framework by discussing model candidates for the circadian clock formed by the KaiABC oscillator found in the cyanobacterium *Synechococcus elongatus*. A simulation study of its dynamical behaviour demonstrates effects of superpositioned protein abundance courses based on regular expressions corresponding to dedicated protein activation states.

1 Introduction

Biological signalling networks have been identified to exhibit a universal capability to process information [14, 17]. They can be viewed as complex computational devices of the cell, triggering and directing responses to external stimuli. It turns out that successive formation or decomposition of protein complexes in conjunction

with domain-specific protein binding (as during phosphorylation by kinases) plays a central role in biological signal transduction based on submolecular assembly [1]. In this context, resulting biomolecules act as information carriers of astonishing storage capacity and structural plasticity. For example, the tumor suppressor protein p53 is equipped with 27 phosphorylation sites [3]. It could theoretically assume up to $2^{27} = 34,217,728$ different activation states. Having in mind that each of these states is able to form an individual constituent of a reaction network incorporating all distinguishable states of up to several hundred interacting proteins, the potential dimension of those protein signalling networks is obvious.

In a typical scenario of exploring coupled intracellular *modules* – functional network units – the present knowledge on involved constituents and topology lacks some detailed information with regard to comprising the entirety of individual molecular interactions. Hence, an integrative setup, prediction, and reconstruction of network model candidates based on incomplete data is a challenging task in systems biology since it requires unconventional techniques to cope with the combinatorial complexity of exhaustive search within the underlying reaction space [15]. A variety of reverse engineering approaches emerged to tackle enumerative reaction network reconstruction at different levels of abstraction (cf. [10, 16]).

While the steady-state behaviour might be sufficient to characterise a metabolic network (cf. [12]), the function of a protein signalling network depends heavily on its temporal evolution [26]. Oscillators based on phosphorylation/dephosphorylation cycles represent significant examples [20, 22, 27]. Thus, the aspect of *dynamical behaviour* should be reflected in the choice of the preferred modelling approach. For that purpose, ordinary differential equations (ODEs) derived from appropriate kinetics are commonly employed. Since this method usually assumes each individual protein activation state to act as a separate species, it easily leads to an exponential growth of the number of distinct ODEs (addressed amongst others in [7]). An opportunity to temporarily unify several activation states by one dedicated species could be a keystone to overcome this insufficiency.

Inspired by this initial idea, we propose a P systems framework able to specify proteins together with relevant properties by string-objects. In contrast to species names in ODEs, phenotypic information about a protein is represented by a character string. Each individual protein property is allowed to be marked as present, absent, or arbitrary. In the latter case, placeholders known from regular expressions denote unassigned protein properties. Consequently, reaction rules may also contain placeholders processed by a matching relation for association of available particles to reactants given within rules. Furthermore, our P systems framework combines the ability to manage specific string-objects with discretised reaction kinetics. Incomplete information about protein activation states can be handled by setting placeholders if required. While they enable a unification of several activation states when specifying a protein on the one hand, placeholders contribute to trace the variety of potential effects by embedding wild-cards into reaction rules on the other hand. Thus, a bottom-up strategy for the modelling of signalling networks by successive knowledge integration can benefit from the

proposed framework. Along with intermediate results coming from simulation of a partially wild-carded system, synergies between wetlab experimental setup and model refinement considering structural dynamics might emerge. Inclusion of reaction kinetics into the formalism of P systems was explained in [18] exemplified by metabolic networks, supplemented by signalling and gene regulatory networks [13]. A previous formulation of periodic and quasi-periodic processes based on symbol objects without inner structure is given in [5]. The BioNetGen framework [6] allows handling of string pattern to constitute species. However, its expressive capability of reaction kinetics excludes stoichiometry.

The paper is organised in two main sections: Firstly, we define the P systems framework Π_{CSM} (Cell Signalling Module) with emphasis on the combination of reaction kinetics and wild-carded representation of proteins as string-objects. Matching strategies accomplish the handling of incomplete information. In order to provide formalisms to select reactants for rule-based rewriting, we adopt the strategy of loose matching [13]. It is expressed by a relation between strings forming objects and strings acting as patterns in rewriting rules. The loose matching checks whether there is at least one common wild-card free representation for both strings. So, it is intended to generate a maximal variety of potential effects. A more general matching approach able to find patterns common to a set of strings has been specified by the Angluin pattern language [2]. In order to enable detailed studies on the temporal evolution of the system, we replace the maximally parallel rewriting from the original framework [23] with a mechanism that is based on reaction kinetics. For each rewriting rule, the number of applications per turn is given by a kinetic function, depending on the current configuration of the system. This way, a deterministic system evolution is obtained. The formal system definition is followed by a comprehensive application scenario: Section 3 demonstrates the suitability of the framework for discussing model candidates of the circadian clock formed by the KaiABC oscillator found in cyanobacterium *Synechococcus elongatus*. Since the detailed mechanism of this biochemical oscillation is partially unknown, various models have been developed recently e.g. [8, 19, 29]. We show their integration into the P systems framework Π_{CSM} in terms of an intersecting superposition of consistent elements flanked by wild-carded completion. A simulation study of the dynamical system's behaviour discloses effects of superpositioned protein abundance courses based on regular expressions corresponding to dedicated protein activation states.

2 System Description

Multiset Prerequisites

Let A be an arbitrary set and $\mathbb{N}$ the set of natural numbers including zero. $\mathcal{P}(A)$ denotes the power set of A. A multiset over A is a mapping $F : A \longrightarrow \mathbb{N} \cup \{\infty\}$. $F(a)$, also denoted as $[a]_F$, specifies the multiplicity of $a \in A$ in F. Multisets can be written as an elementwise enumeration of the form $\{(a_1, F(a_1)), (a_2, F(a_2)), \ldots\}$

since $\forall(a,b_1),(a,b_2) \in F : b_1 = b_2$. The support $\mathrm{supp}(F) \subseteq A$ of F is defined by $\mathrm{supp}(F) = \{a \in A \mid F(a) > 0\}$. A multiset F over A is said to be empty iff $\forall a \in A : F(a) = 0$. The cardinality $|F|$ of F over A is $|F| = \sum_{a\in A} F(a)$. Let F_1 and F_2 be multisets over A. F_1 is a subset of F_2, denoted as $F_1 \subseteq F_2$, iff $\forall a \in A : (F_1(a) \leq F_2(a))$. Multisets F_1 and F_2 are equal iff $F_1 \subseteq F_2 \wedge F_2 \subseteq F_1$. The intersection $F_1 \cap F_2 = \{(a, F(a)) \mid a \in A \wedge F(a) = \min(F_1(a), F_2(a))\}$, the multiset sum $F_1 \uplus F_2 = \{(a, F(a)) \mid a \in A \wedge F(a) = F_1(a) + F_2(a)\}$, and the multiset difference $F_1 \ominus F_2 = \{(a, F(a)) \mid a \in A \wedge F(a) = \max(F_1(a) - F_2(a), 0)\}$ form multiset operations. Multiplication of a multiset $F = \{(a, F(a)) \mid a \in A\}$ with a scalar c, denoted $\mathrm{c} \cdot F$, is defined by $\{(a, \mathrm{c} \cdot F(a)) \mid a \in A\}$. The term $\langle A\rangle = \{F : A \longrightarrow \mathbb{N} \cup \{\infty\}\}$ describes the set of all multisets over A.

Definition of System Components

A P system for a cell signalling module (CSM) is a construct

$$\mathit{\Pi}_{\mathrm{CSM}} = (V, V', R_1, \ldots, R_r, \mathrm{f}_1, \ldots, \mathrm{f}_r, A, \mathrm{C}, \mathit{\Delta\tau})$$

where V and V' are two alphabets (not necessarily disjoint); without loss of generality $\#, \neg, * \notin V \cup V'$. The regular set

$$S = V^+ \cdot \left(\{\#\} \cdot ((V')^+ \cup \{\neg\} \cdot (V')^+ \cup \{*\})\right)^*$$

describes the syntax for string-objects. The leftmost substring from V^+ holds the protein identifier, followed by a finite number of protein property substrings from $(V')^+$ which are separated by #. For example, consider the string-object C:D#p#*#¬q identifying protein (complex) C:D with specified property p, a second arbitrary property (*), and without property q. Each protein property substring expresses a specific additional information about the protein, for instance whether it is activated by carrying a ligand at a certain binding site. Two kinds of meta symbols are allowed. The symbol ¬ excludes the subsequent property but permits all other properties at this substring position. The placeholder * stands for an arbitrary (also unknown or unspecified) protein property substring. This way, uncertainty about the properties of proteins can be explicitly expressed. String-objects can be dynamically processed by reaction rules:

$R_i \in \langle S\rangle \times \langle S\rangle$	is a reaction rule composed of two finite multisets
$\mathrm{f}_i : \langle S\rangle \longrightarrow \mathbb{N}$	is a function corresponding to kinetics of reaction R_i
$A \in \langle S\rangle$	is a multiset of axioms representing the initial molec. configuration
$\mathrm{C} \in \mathbb{R}_+$	spatial capacity of the module (vessel or compartment)
$\mathit{\Delta\tau} \in \mathbb{R}_+$	time discretisation interval

We explain the system evolution of $\mathit{\Pi}_{\mathrm{CSM}}$ within three consecutive subsections. Based on the specification of the system configuration, we define an iteration

scheme that updates this configuration from time t to time $t+1$. The update includes processing of reactions given by the rules R_i $(i = 1, \ldots, r)$. For this purpose, an appropriate matching between wild-carded strings representing reactants and those stated in the current configuration is required. Then, a reaction is executed by removing the multiset of matching reactants from the current configuration followed by adding the corresponding products. In order to consider kinetic issues, each reaction can be multiply processed. Therefore, the number of turns is provided by the function f_i.

Dynamical System Behaviour

A P system of the form Π_{CSM} evolves by successive progression of its *configuration* $L_t \in \langle S \rangle$ at discrete points in time $t \in \mathbb{N}$ for what we assume a global clock. Two consecutive dates t and $t+1$ specify a time span $\Delta\tau$ (discretisation interval). A system step at time t consists of two modification stages per reaction $1, \ldots, r$. Firstly, the multiset of reactants is determined and removed from L_t. Afterwards, the corresponding multiset of products is added. To cope with conflicts that can occur if the available amount of reactants cannot satisfy all matching reactions, we prioritise the reaction rules by their index: $R_1 > R_2 > \ldots > R_r$. Thus, we keep determinism of the system evolution and enable mass conservation.

$$
\begin{aligned}
L_0 = L_{0,0} &= A \\
L_{t,1} &= \begin{cases} L_{t,0} \ominus \mathit{Reactants}_{t,1} \uplus \mathit{Products}_{t,1} \text{ if } \mathit{Reactants}_{t,1} \subseteq L_{t,0} \\ L_{t,0} \text{ otherwise} \end{cases} \\
L_{t,2} &= \begin{cases} L_{t,1} \ominus \mathit{Reactants}_{t,2} \uplus \mathit{Products}_{t,2} \text{ if } \mathit{Reactants}_{t,2} \subseteq L_{t,1} \\ L_{t,1} \text{ otherwise} \end{cases} \\
&\vdots \\
L_{t+1} = L_{t,r} &= \begin{cases} L_{t,r-1} \ominus \mathit{Reactants}_{t,r} \uplus \mathit{Products}_{t,r} \text{ if } \mathit{Reactants}_{t,r} \subseteq L_{t,r-1} \\ L_{t,r-1} \text{ otherwise} \end{cases}
\end{aligned}
$$

Let $R_j = (A_j, B_j) \in \langle S \rangle \times \langle S \rangle$ be a reaction rule with $\mathrm{supp}(A_j) = \{a_1, \ldots, a_p\}$ and $\mathrm{supp}(B_j) = \{b_1, \ldots, b_q\}$. In terms of a chemical denotation, it can be written as

$$A_j(a_1)\ a_1 + \ldots + A_j(a_p)\ a_p \longrightarrow B_j(b_1)\ b_1 + \ldots + B_j(b_q)\ b_q$$

where $A_j(a_1), \ldots, A_j(a_p)$ represent stoichiometric factors of reactants $a_1, \ldots, a_p$, and $B_j(b_1), \ldots, B_j(b_q)$ stoichiometric factors of products $b_1, \ldots, b_q$, respectively. All reactant strings that match to the pattern a_k are provided by a dedicated relation $\mathit{Match}(a_k)$ (see next subsection for definition). A combination of reactant strings from L_t matching the left hand side of R_j forms a multiset of string-objects used to apply the reaction once. Since the kinetic law, described by the corresponding scalar function f_j, returns the number of applications of reaction rule R_j within one step, the multiset of string-objects extracted from L_t to act as reactants for R_j can be written as $\mathit{Reactants}_{t,j}$:

$$Reactants_{t,j} = \biguplus_{e_1 \in Match(a_1)} \dots \biguplus_{e_p \in Match(a_p)} \mathrm{f}_j\big(\{(e_1,\infty),\dots,(e_p,\infty)\} \cap L_{t,j-1}\big) \cdot \big\{(e_1, A_j(a_1)),\dots,(e_p, A_j(a_p))\big\}$$

Accordingly, the multiset of products resulting from reaction rule R_j is determined by the multiset $Products_j(t)$:

$$Products_{t,j} = \biguplus_{e_1 \in Match(a_1)} \dots \biguplus_{e_p \in Match(a_p)} \mathrm{f}_j\big(\{(e_1,\infty),\dots,(e_p,\infty)\} \cap L_{t,j-1}\big) \cdot \big\{(b_1, B_j(b_1)),\dots,(b_q, B_j(b_q))\big\}$$

Matching

Let the regular set S be a syntax description for string-objects. In the symmetric relation *Match*, two string-objects match iff there is at least one common representation without wild-cards. This loose strategy requires a minimum degree of similarity between objects with incomplete information. Uncertainty is interpreted as arbitrary replacements within the search space given by S.

$$\begin{aligned} &Match \subseteq S \times S \\ &Match = \bigcup_{m \in \mathbb{N}} \{(p\#p_1\#p_2 \dots \#p_m,\ s\#s_1\#s_2 \dots \#s_m) \mid (p = s) \wedge \\ &\qquad \forall j \in \{1,\dots,m\} : [(p_j = s_j) \vee (p_j = *) \vee (s_j = *) \vee \\ &\qquad\quad ((p_j = \neg q) \wedge (s_j \neq q)) \vee ((s_j = \neg q) \wedge (p_j \neq q))]\} \end{aligned}$$

Matching of a single string-object $w \in S$ to the entire set S is defined by

$$Match(w) = \{s \in S \mid (w,s) \in Match\}$$

Consequently, we define the matching of a language $L \subseteq S$ by the function $Match : \mathcal{P}(S) \longrightarrow \mathcal{P}(S)$ with

$$Match(L) = \bigcup_{w \in L} Match(w).$$

Discrete Reaction Kinetics

Within the P systems framework Π_{CSM}, we formulate reaction kinetics by specification of scalar functions f_j attached to corresponding reactions R_j $(j = 1,\dots,r)$. Each scalar function converts the current configuration L_t, a multiset of string-objects, into the number of turns for application of rewriting rule R_j:

$$\mathrm{f}_j(L_t) = \left\lfloor k_j \prod_{\forall \alpha \in Match(A_j) \cap Match(L_t) : (R_j = (A_j, B_j))} \hat{\mathrm{f}}(L_t(\alpha))^{|Match(A_j) \cap \{(\alpha,\infty)\}|} \right\rfloor \quad (1)$$

whereas the auxiliary term α passes through all string-objects present in L_t which also form reactants in R_j. The multiplicity $L_t(\alpha)$ of occurrences of α acts as argument for a kinetic law $\hat{\mathrm{f}}(L_t(\alpha))$. Examples adopted from mass-action, Michaelis-Menten, and Hill kinetics are shown in Figure 1.

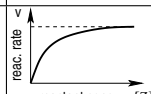
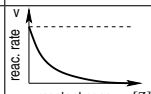
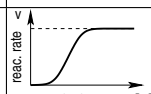
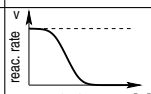
Kinetics	Activation	Repression
Mass-Action (no saturation)	 $\hat{\mathrm{f}}([Z]) = [Z]$	–
Michaelis-Menten (saturation)	 $\hat{\mathrm{f}}([Z]) = \frac{[Z]}{\Theta+[Z]}$	 $\hat{\mathrm{f}}([Z]) = \left(1 - \frac{[Z]}{\Theta+[Z]}\right)$
Higher-Order Hill (saturation)	 $\hat{\mathrm{f}}([Z]) = \frac{[Z]^n}{\Theta+[Z]^n}$	 $\hat{\mathrm{f}}([Z]) = \left(1 - \frac{[Z]^n}{\Theta+[Z]^n}\right)$

Fig. 1. Overview of several widely used kinetic laws $\hat{\mathrm{f}}([Z])$ dependent on reactant concentration $[Z]$. Parameters: threshold $\Theta \in \mathbb{R}_+$, Hill coefficient $n \in \mathbb{N}_+$

Relations to ODE-Based Reaction Kinetics

For a reaction system with a total number of n species ($i = 1, \ldots, n$) and r reactions ($j = 1, \ldots, r$)

$$a_{1,j}Z_1 + a_{2,j}Z_2 + \ldots + a_{n,j}Z_n \xrightarrow{\hat{k}_j} b_{1,j}Z_1 + b_{2,j}Z_2 + \ldots + b_{n,j}Z_n$$

the corresponding ODEs

$$\frac{\mathrm{d}\,[Z_i]}{\mathrm{d}\,t} = \sum_{j=1}^{r} \left(\hat{k}_j \cdot (b_{i,j} - a_{i,j}) \cdot \prod_{l=1}^{n} \hat{\mathrm{f}}_j([Z_l])^{a_{l,j}} \right) \quad \text{with} \quad i = 1, \ldots, n. \tag{2}$$

describe the temporal systems behaviour by consideration of stoichiometric coefficients $a_{i,j} \in \mathbb{N}$ (reactants) and $b_{i,j} \in \mathbb{N}$ (products) as well as a kinetic law $\hat{\mathrm{f}}_j([Z_i]) : \mathbb{R}_+ \to \mathbb{R}_+$ that maps a species concentration $[Z_i]$ into an effective reaction rate [9]. All initial concentrations $[Z_i](0) \in \mathbb{R}_+$, $i = 1, \ldots, n$ are allowed to be set according to the needs of the reaction system.

A species concentration $[Z_i] := \frac{z_i}{\mathrm{C}}$ is defined as fraction of its molecular amount $z_i = \mathrm{supp}(\{(Z_i, z_i)\})$ with respect to the spatial system capacity $\mathrm{C} \in \mathbb{R}_+$.

A correspondence between the reaction rate k_j (employed in Π_{CSM} by function f_j attached to reaction R_j) and the kinetic constant $\hat{k}_j$ utilised in ODE (2) can be obtained by the Euler method of integrating differential equations. Discretisation of (2) with respect to time and concentration value results in:

$$\frac{\frac{z_{i,t+1}-z_{i,t}}{\mathrm{C}}}{\Delta\tau} = \sum_{j=1}^{r}\left(\hat{k}_j \cdot (b_{i,j}-a_{i,j}) \cdot \prod_{l=1}^{n} \hat{\mathrm{f}}_j([Z_l])^{a_{l,j}}\right)$$

$$z_{i,t+1} - z_{i,t} = \mathrm{C} \cdot \Delta\tau \cdot \sum_{j=1}^{r}\left(\hat{k}_j \cdot (b_{i,j}-a_{i,j}) \cdot \prod_{l=1}^{n} \hat{\mathrm{f}}_j([Z_l])^{a_{l,j}}\right)$$

By setting $k_j = \hat{k}_j \cdot \mathrm{C} \cdot \Delta\tau$, we obtain:

$$z_{i,t+1} - z_{i,t} = k_1(b_{i,1}-a_{i,1}) \prod_{l=1}^{n} \hat{\mathrm{f}}_1([Z_l])^{a_{l,1}} + \ldots + k_r(b_{i,r}-a_{i,r}) \prod_{l=1}^{n} \hat{\mathrm{f}}_r([Z_l])^{a_{l,r}}$$

Replacing $k_j \cdot \hat{\mathrm{f}}_j([Z_l])^{a_{l,j}}$ by the discretised (and hence approximated) scalar function $\mathrm{f}_j(L_t)$ from Equation (1) leads to:

$$z_{i,t+1} - z_{i,t} \approx (b_{i,1}-a_{i,1}) \cdot \mathrm{f}_1(L_t) + \ldots + (b_{i,r}-a_{i,r}) \cdot \mathrm{f}_r(L_t)$$

Since the stoichiometric coefficients $a_{i,j}$ and $b_{i,j}$ of each reaction $R_j = (A_j, B_j)$ in Π_{CSM} are expressed by multisets A_j (reactants) and B_j (products), we write:

$$z_{i,t+1} - z_{i,t} = (B_1(b_i) - A_1(a_i)) \cdot \mathrm{f}_1(L_t) + \ldots + (B_r(b_i) - A_r(a_i)) \cdot \mathrm{f}_r(L_t)$$

From that, we achieve the update scheme for species Z_i present in L_t with $z_{i,t}$ copies at time t by processing reaction R_j:

$$z_{i,t+1} = z_{i,t} - A_j(Z_i) \cdot \mathrm{f}_j(L_t) + B_j(Z_i) \cdot \mathrm{f}_j(L_t)$$

By extension from a single species to the entire configuration along with inclusion of matching, we finally yield

$$L_{t+1,j} = L_{t,j} \ominus \mathit{Reactants}_{t,j} \uplus \mathit{Products}_{t,j}$$

in accordance to the iteration scheme for Π_{CSM} evolution. The conversion of thresholds Θ occurring in Michaelis-Menten or Hill terms from the ODE approach into the Π_{CSM} framework can be done by parameter fitting or regression that maps the concentration-based gradient into an amount-based counterpart.

3 The KaiABC Oscillator – A Circadian Clock

Biological Background

Circadian rhythms embody an interesting biological phenomenon that can be seen as a widespread property of life. The coordination of biological activities into daily

cycles provides an important advantage for the fitness of diverse organisms [4, 25]. Based on self-sustained biochemical oscillations, circadian clocks are characterised by a period of approximately 24h that persists under constant conditions (like constant darkness or constant light). Their ability for compensation of temperature in the physiological range enables then to maintain the period in case of environmental changes. Furthermore, circadian clocks can be entrained. This property allows a gradual reset of the underlying oscillatory system for adjustment by exposure to external stimuli like light/dark or temperature cycles. A variety of metabolic, cell signalling, and gene regulatory processes is synchronised or controlled by circadian clocks. Chemically, they utilise an individual cycling reaction scheme including one or more feedback loops. Most of the circadian clocks comprise gene transcription and translation feedback loops [24].

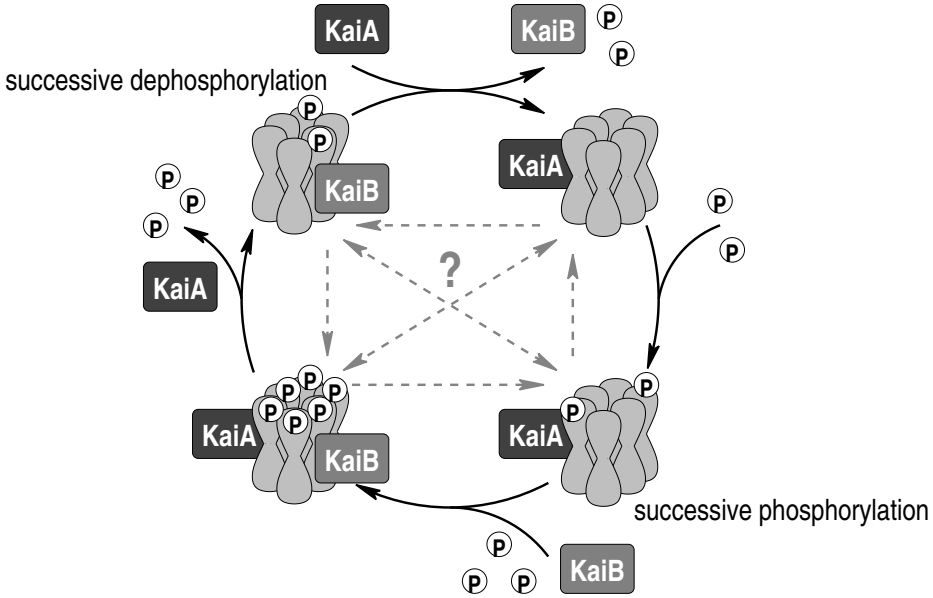


Fig. 2. Reaction cycle of the KaiABC oscillator characterised by four phases and incomplete information about interphase feedback loops, arranged from descriptions of the oscillatory mechanism given in [11, 20]. A corresponding minimal model of the four-phase cycle has been proposed in [4].

Surprisingly, the prokaryotic cyanobacterium *Synechococcus elongatus* was discovered to carry a post-translational circadian clock even functioning *in vitro* [27]. Three key clock proteins KaiA, KaiB, and KaiC with known atomic structure could be identified [21]. KaiC as the focal protein rhythmically oscillates between hypophosphorylated and hyperphosphorylated forms [22]. The spatial structure of KaiC represents a homohexamer shaped as a "double doughnut" with 6 phosphory-

lation twin sites at the interfaces between monomeric subunits. Presence of the supplementary protein KaiA specifically enhances KaiC phosphorylation while KaiBC complex formation activates KaiC dephosphorylation [20]. The KaiABC circadian oscillator appears as a reaction cycle consisting of four consecutive phases [11], see Figure 2: KaiAC complex formation releasing KaiB, successive KaiAC phosphorylation, KaiABC complex formation, and successive KaiABC dephosphorylation in conjunction with KaiA dissociation. Each of these phases takes approximately 6h. There is some evidence for further interactions between the aforementioned protein complexes and intermediate products in terms of negative feedback loops stabilising the oscillation. However, the detailed mechanism is still unclear and gives room for hypotheses reflected in a couple of model candidates [4]. A current study raises the question whether clock-protein expression could still be involved in its general function [28].

Review of Modelling Approaches

In this section, we briefly compare three current model candidates [8, 19, 29] beyond a minimal model [4] able to capture the dynamical behaviour of the KaiABC oscillator in accordance with wetlab experimental data. Assumptions on unknown parts of the oscillator mechanism result from empirical studies. Here, an underlying reaction network topology is hypothesised and afterwards filled with appropriate parameter values obtained by fitting using an exhaustive search.

KaiA sequestration has been suggested in [8]. The resulting model identifies a total number of 15 interacting species where $C^0, \ldots, C^6$ correspond to the amount of phosphorylated monomeric subunits within KaiC. Accordingly, $BC^0, \ldots, BC^6$ are species names for complex KaiBC. B indicates KaiB. KaiA is assumed to be sequestered by the KaiC/KaiBC complexes and hence not modelled explicitly. Instead, it is interpreted as an inhibiting factor causing negative feedback loops. See Figure 3 **A** for the reaction network topology.

Following the idea of a quick KaiC monomer shuffle, in [29] a network topology containing 54 dedicated species is proposed. There are two categories of species marked as "tense" (T) for those employed in the phosphorylation phase and "relaxed" (R) for the dephosphorylation phase. Indexes attached to T and R ranging from 0 to 6 comprise the number of currently phosphorylated monomeric subunits while association of KaiA and/or KaiB complexes is denoted by concatenation of A or B to the species names. Figure 3 **B** illustrates the network topology by usage of dashed arrows for monomer shuffle.

A different description has been introduced in [19] managing on 7 species (by neglecting intermediate products of protein degradation). Inspired by the insight that distinction of two states is sufficient to obtain robust oscillations of KaiC phosphorylation, a cascade of elementary cell signalling motifs is proposed. In this two-stage scenario, three phosphates from species $KaiC$ can be added and removed per stage by catalysts $KaiA$ and $KaiAB$, respectively. Additionally, the model formulates the complex formation of $KaiAB$ which is catalysed by the threefold phosphorylated protein $PKaiC$. Vice versa, its decomposition is supported by

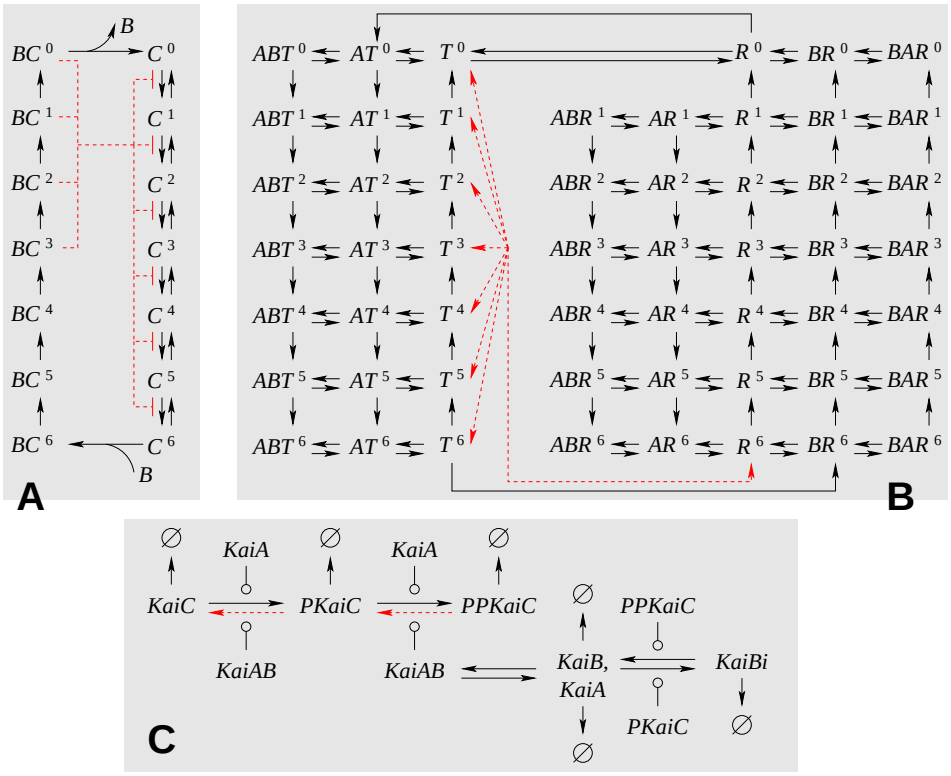


Fig. 3. Comparison of KaiABC oscillator network topologies adapted from [8] (**A**), [29] (**B**), and [19] (**C**). Dashed lines indicate relevant feedback loops for sustained oscillation.

the six-fold phosphorylated protein *PPKaiC*. Decay reactions for each protein complete the model candidate's network topology, see Figure 3 **C**.

Conversion to the Π_{CSM} Framework

We demonstrate a conversion of the core oscillator extracted from different model candidates into the P systems framework Π_{CSM}. The capability of this algebraic approach is to cope with a potential combinatorial complexity of protein states, shown by formulating reaction and transduction rules using placeholders ($*$) for arbitrary or unknown molecular constituents.

Each of the six KaiC monomeric subunits is said to be phosphorylated iff both phosphorylation sites are saturated. Theoretically, the KaiABC protein complex could induce a maximum of $2^8 = 256$ potential states. This amount results from the general assumption that each monomeric subunit is able to be individually phosphorylated or dephosphorylated in combination with present or absent association of KaiA and KaiB, respectively. In terms of a distinction of 8 binary digits from

these molecular configurations, a full network of $2 \cdot \binom{256}{2} = 65,280$ bi-molecular reactions could be spanned. Since KaiC turns out to be a highly symmetric homo-hexamer, the individual monomeric subunits cannot be distinguished in practice. Instead, the number of attached phosphates is utilised that varies in a seven-stage range from 0 up to 6. In addition to the combinatorial variety caused by present or absent association of KaiA and KaiB, KaiABC possess $7 \cdot 4 = 28$ states from a biochemical point of view.

For the P systems description, we identify a module for the cycling reaction scheme sketched in Figure 2. Key proteins KaiA, KaiB, and KaiC resulting from expression of corresponding genes are assumed to be present in the module *ab initio*. Considering the core oscillator, 17 reaction rules along with loose matching correspond to the four-phase reaction cycle. Successive KaiC phosphorylation in the presence of KaiA is expressed by rules R_1 to R_6 followed by successive de-phosphorylation in the presence of KaiB within rules R_7 to R_{12}. Finally, R_{13} and R_{14} formulate inhibiting KaiA/KaiB exchange acting as negative feedback loops, and R_{15} up to R_{17} reflect protein degradation. A kinetic function f is attached to each reaction rule that follows from discretised Michaelis-Menten kinetic laws in concert with linear mass-action kinetics for protein degradation.

$$\Pi_{KaiABC} = (V, V', R_1, \ldots, R_{17}, \mathrm{f}_1, \ldots, \mathrm{f}_{17}, A, \mathrm{C}, \Delta\tau)$$

$$\begin{aligned} V &= \{\mathrm{C}\} \cup \ldots\ldots\ldots\ldots\ldots\text{identifier of the focal protein KaiC} \\ &\quad \{\mathrm{A}, \mathrm{B}\}\ldots\ldots\ldots\ldots\ldots\text{identifiers of proteins KaiA and KaiB} \\ V' &= \{\mathrm{A}, \mathrm{B}\} \cup \ldots\ldots\ldots\ldots\text{KaiA, KaiB within a complex associated to KaiC} \\ &\quad \{0, 1, 2, 3, 4, 5, 6\}\ldots\ldots\text{number of attached phosphates} \end{aligned}$$

$$\begin{aligned} R_1 &= \mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#0 + \mathrm{A} \longrightarrow \mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#1 + \mathrm{B} \\ R_2 &= \mathrm{C}\#\mathrm{A}\#*\#1 + \mathrm{A} \longrightarrow \mathrm{C}\#\mathrm{A}\#*\#2 + \mathrm{A} \\ R_3 &= \mathrm{C}\#\mathrm{A}\#*\#2 + \mathrm{A} \longrightarrow \mathrm{C}\#\mathrm{A}\#*\#3 + \mathrm{A} \\ R_4 &= \mathrm{C}\#\mathrm{A}\#*\#3 + \mathrm{A} \longrightarrow \mathrm{C}\#\mathrm{A}\#*\#4 + \mathrm{A} \\ R_5 &= \mathrm{C}\#\mathrm{A}\#*\#4 + \mathrm{A} \longrightarrow \mathrm{C}\#\mathrm{A}\#*\#5 + \mathrm{A} \\ R_6 &= \mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#5 + \mathrm{B} \longrightarrow \mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#6 + \mathrm{A} \\ R_7 &= \mathrm{C}\#*\#\mathrm{B}\#6 + \mathrm{B} \longrightarrow \mathrm{C}\#*\#\mathrm{B}\#5 + \mathrm{B} \\ R_8 &= \mathrm{C}\#*\#\mathrm{B}\#5 + \mathrm{B} \longrightarrow \mathrm{C}\#*\#\mathrm{B}\#4 + \mathrm{B} \\ R_9 &= \mathrm{C}\#*\#\mathrm{B}\#4 + \mathrm{B} \longrightarrow \mathrm{C}\#*\#\mathrm{B}\#3 + \mathrm{B} \\ R_{10} &= \mathrm{C}\#*\#\mathrm{B}\#3 + \mathrm{B} \longrightarrow \mathrm{C}\#*\#\mathrm{B}\#2 + \mathrm{B} \\ R_{11} &= \mathrm{C}\#*\#\mathrm{B}\#2 + \mathrm{B} \longrightarrow \mathrm{C}\#*\#\mathrm{B}\#1 + \mathrm{B} \\ R_{12} &= \mathrm{C}\#*\#\mathrm{B}\#1 + \mathrm{B} \longrightarrow \mathrm{C}\#*\#\mathrm{B}\#0 + \mathrm{B} \\ R_{13} &= \mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#* + \mathrm{A} \longrightarrow \mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#* + \mathrm{B} \\ R_{14} &= \mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#* + \mathrm{B} \longrightarrow \mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#* + \mathrm{A} \\ R_{15} &= \mathrm{A} \longrightarrow \emptyset \end{aligned}$$

$$
\begin{aligned}
R_{16} &= \mathrm{B} \longrightarrow \emptyset \\
R_{17} &= \mathrm{C}\#*\#*\#* \longrightarrow \emptyset \\
\mathrm{f}_1(L_t) &= \left\lfloor k_1 \cdot \frac{L_t(\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#0)}{\Theta_{1,1} + L_t(\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#0)} \cdot \frac{L_t(\mathrm{A})}{\Theta_{1,2} + L_t(\mathrm{A})} \right\rfloor \\
\mathrm{f}_2(L_t) &= \left\lfloor k_2 \cdot \frac{L_t(\mathrm{C}\#\mathrm{A}\#*\#1)}{\Theta_{2,1} + L_t(\mathrm{C}\#\mathrm{A}\#*\#1)} \cdot \frac{L_t(\mathrm{A})}{\Theta_{2,2} + L_t(\mathrm{A})} \right\rfloor \\
\mathrm{f}_3(L_t) &= \left\lfloor k_3 \cdot \frac{L_t(\mathrm{C}\#\mathrm{A}\#*\#2)}{\Theta_{3,1} + L_t(\mathrm{C}\#\mathrm{A}\#*\#2)} \cdot \frac{L_t(\mathrm{A})}{\Theta_{3,2} + L_t(\mathrm{A})} \right\rfloor \\
\mathrm{f}_4(L_t) &= \left\lfloor k_4 \cdot \frac{L_t(\mathrm{C}\#\mathrm{A}\#*\#3)}{\Theta_{4,1} + L_t(\mathrm{C}\#\mathrm{A}\#*\#3)} \cdot \frac{L_t(\mathrm{A})}{\Theta_{4,2} + L_t(\mathrm{A})} \right\rfloor \\
\mathrm{f}_5(L_t) &= \left\lfloor k_5 \cdot \frac{L_t(\mathrm{C}\#\mathrm{A}\#*\#4)}{\Theta_{5,1} + L_t(\mathrm{C}\#\mathrm{A}\#*\#4)} \cdot \frac{L_t(\mathrm{A})}{\Theta_{5,2} + L_t(\mathrm{A})} \right\rfloor \\
\mathrm{f}_6(L_t) &= \left\lfloor k_6 \cdot \frac{L_t(\mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#5)}{\Theta_{6,1} + L_t(\mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#5)} \cdot \frac{L_t(\mathrm{B})}{\Theta_{6,2} + L_t(\mathrm{B})} \right\rfloor \\
\mathrm{f}_7(L_t) &= \left\lfloor k_7 \cdot \frac{L_t(\mathrm{C}\#*\#\mathrm{B}\#6)}{\Theta_{7,1} + L_t(\mathrm{C}\#*\#\mathrm{B}\#6)} \cdot \frac{L_t(\mathrm{B})}{\Theta_{7,2} + L_t(\mathrm{B})} \right\rfloor \\
\mathrm{f}_8(L_t) &= \left\lfloor k_8 \cdot \frac{L_t(\mathrm{C}\#*\#\mathrm{B}\#5)}{\Theta_{8,1} + L_t(\mathrm{C}\#*\#\mathrm{B}\#5)} \cdot \frac{L_t(\mathrm{B})}{\Theta_{8,2} + L_t(\mathrm{B})} \right\rfloor \\
\mathrm{f}_9(L_t) &= \left\lfloor k_9 \cdot \frac{L_t(\mathrm{C}\#*\#\mathrm{B}\#4)}{\Theta_{9,1} + L_t(\mathrm{C}\#*\#\mathrm{B}\#4)} \cdot \frac{L_t(\mathrm{B})}{\Theta_{9,2} + L_t(\mathrm{B})} \right\rfloor \\
\mathrm{f}_{10}(L_t) &= \left\lfloor k_{10} \cdot \frac{L_t(\mathrm{C}\#*\#\mathrm{B}\#3)}{\Theta_{10,1} + L_t(\mathrm{C}\#*\#\mathrm{B}\#3)} \cdot \frac{L_t(\mathrm{B})}{\Theta_{10,2} + L_t(\mathrm{B})} \right\rfloor \\
\mathrm{f}_{11}(L_t) &= \left\lfloor k_{11} \cdot \frac{L_t(\mathrm{C}\#*\#\mathrm{B}\#2)}{\Theta_{11,1} + L_t(\mathrm{C}\#*\#\mathrm{B}\#2)} \cdot \frac{L_t(\mathrm{B})}{\Theta_{11,2} + L_t(\mathrm{B})} \right\rfloor \\
\mathrm{f}_{12}(L_t) &= \left\lfloor k_{12} \cdot \frac{L_t(\mathrm{C}\#*\#\mathrm{B}\#1)}{\Theta_{12,1} + L_t(\mathrm{C}\#*\#\mathrm{B}\#1)} \cdot \frac{L_t(\mathrm{B})}{\Theta_{12,2} + L_t(\mathrm{B})} \right\rfloor \\
\mathrm{f}_{13}(L_t) &= \left\lfloor k_{13} \cdot \left(1 - \frac{L_t(\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#*)}{\Theta_{13,1} + L_t(\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#*)}\right) \cdot \left(1 - \frac{L_t(\mathrm{A})}{\Theta_{13,2} + L_t(\mathrm{A})}\right) \right\rfloor \\
\mathrm{f}_{14}(L_t) &= \left\lfloor k_{14} \cdot \left(1 - \frac{L_t(\mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#*)}{\Theta_{14,1} + L_t(\mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#*)}\right) \cdot \left(1 - \frac{L_t(\mathrm{B})}{\Theta_{14,2} + L_t(\mathrm{B})}\right) \right\rfloor \\
\mathrm{f}_{15}(L_t) &= k_{15} \cdot L_t(\mathrm{A}) \\
\mathrm{f}_{16}(L_t) &= k_{16} \cdot L_t(\mathrm{B}) \\
\mathrm{f}_{17}(L_t) &= k_{17} \cdot L_t(\mathrm{C}\#*\#*\#*) \\
A &\in \langle \{\mathrm{C}\#*\#*\#*\} \rangle
\end{aligned}
$$

Simulation Case Study

Using the KaiABC circadian oscillator we conducted a simulation case study to demonstrate the practicability of the modelling approach addressed before. The

reaction scheme formulated by the P system Π_{KaiABC} exhibits a high degree of symmetry among its constituents. The main reaction cycle is composed of 12 consecutive feedforward reactions flanked by widespread negative feedback loops. They affect each intermediate product within the reaction cycle following the intention of an inhibiting KaiA/KaiB exchange independent of the phosphorylation state.

For simulation of the dynamical behaviour of Π_{KaiABC}, we empirically parameterise and initialise the system in a symmetric way to obtain phase-shifted protein abundance courses which stably oscillate with a period of approximately 24 hours. To avoid a transient oscillation phase, the initial amounts of protein constituents were set directly at the discrete limit cycle. This constraint is reflected in the following multiset of axioms:

$$\begin{aligned} A = \{&(\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#0, 470), (\mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#1, 351), (\mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#2, 198),\\ &(\mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#3, 135), (\mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#4, 148), (\mathrm{C}\#\mathrm{A}\#\neg\mathrm{B}\#5, 210),\\ &(\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#6, 282), (\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#5, 364), (\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#4, 463),\\ &(\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#3, 541), (\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#2, 586), (\mathrm{C}\#\neg\mathrm{A}\#\mathrm{B}\#1, 571),\\ &(\mathrm{A}, 2520), (\mathrm{B}, 2520)\} \end{aligned}$$

Each KaiC protein within the pattern $\mathrm{C}\# * \# * \#*$ keeps an average amount of 360 copies (arbitrarily chosen).

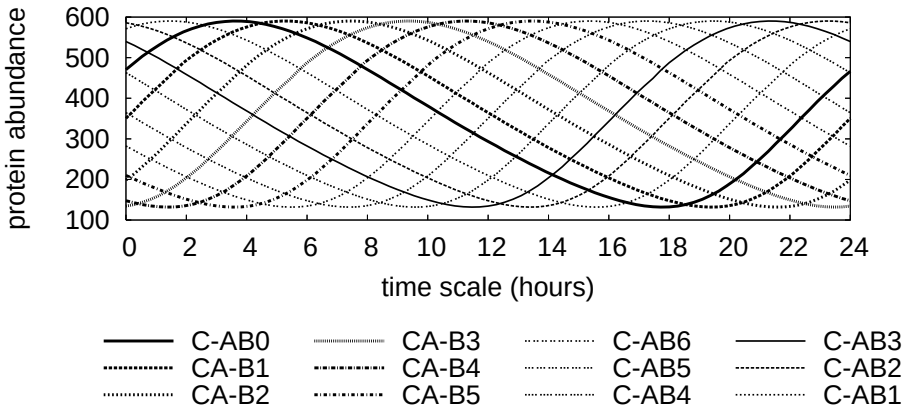


Fig. 4. Temporal courses of 12 specific KaiABC subproducts representing the process status of the reaction cycle. Kinetic parameters and initial amounts adjusted in a way to obtain a period of ≈ 24 hours and symmetry among individual oscillations.

Figure 4 shows the corresponding individual protein abundance courses resulting from following parameter setting for the discrete iteration scheme: $\Theta_{i,1} = 79.2, \Theta_{i,2} = 554.4, \hat{k}_i = 360.0$ for $i \in \{1, \ldots, 12\}$; $\Theta_{i,1} = 64.8, \Theta_{i,2} = 453.6, \hat{k}_i =$

412.8 for $i \in \{13, 14\}$, and $\hat{k}_{15} = \hat{k}_{16} = 508.1$, $\hat{k}_{17} = 254.6$; C = 1.2, $\Delta\tau = 0.05$. The iteration scheme for system evolution was implemented in the programming language C to obtain the course data.

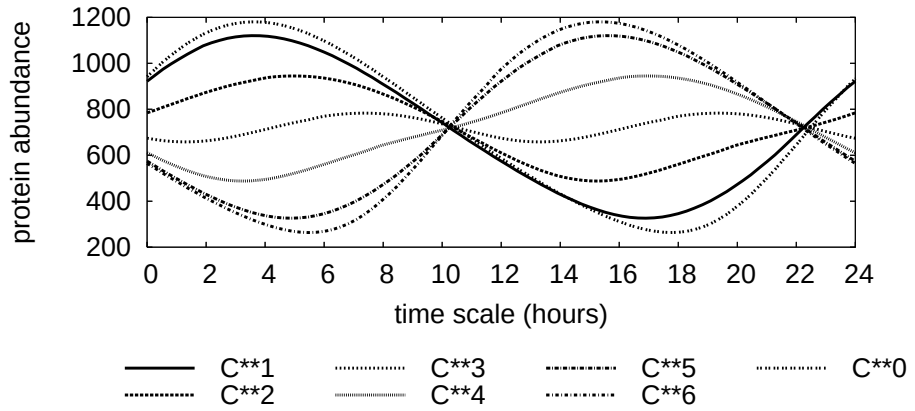


Fig. 5. Temporal courses of KaiABC subproducts subsumed by their level of phosphorylation ranging from 0 to 6. Kinetic parameters and initial amounts adjusted in a way to obtain a period of ≈ 24 hours and symmetry among individual oscillations.

Based on the individual protein abundance courses depicted in Figure 4, Figure 5 illustrates the effect of subsuming KaiABC subproducts according to their number of attached phosphates ranging from 0 to 6. Association of KaiA and KaiB is neglected here resulting in consideration of regular expressions $\mathrm{C}\# * \# * \#i$ for $i = 0, \ldots, 6$. The simulation shows that medium phosphorylation levels possess smaller amplitudes than minor or major phosphorylation levels. Due to symmetry reasons, KaiABC subproducts carrying three phosphates double the frequency of oscillation. Hence, the reaction system is able to act as a scaler. This feature could be useful to control downstream processes at a subcircadian granularity.

Classification of KaiABC subproducts with regard to association of KaiA and KaiB leads to simulation results depicted in Figure 6. As expected, both courses proceed in opposite direction emphasising the mutually exclusive association of KaiA and KaiB to KaiC.

Further simulation studies could explore the effects of different temperatures to the network behaviour. To this end, modified forms of Arrhenius terms based on the Boltzmann constant instead of the universal gas constant might be utilised to replace each reaction parameter k_j. In this way, a possible capability of temperature compensation or entrainment is investigable and can be applied to fine-tuning of the model.

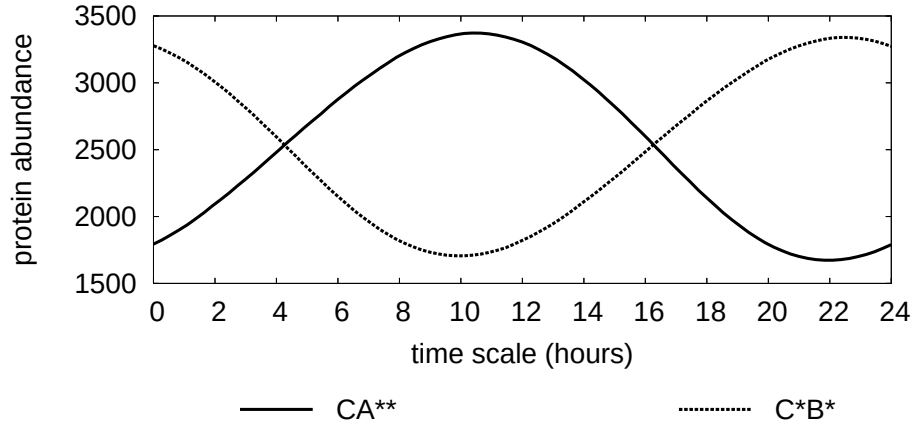


Fig. 6. Temporal courses of KaiABC subproducts separated into two groups by association of KaiA resp. KaiB to KaiC. Kinetic parameters and initial amounts adjusted in a way to obtain a period of ≈ 24 hours and symmetry among individual oscillations.

Extensions of the System

In this section, we address specialties of the different modelling approaches [8, 19, 29] in the context of their conversion into the P systems framework by additional wild-carded reactions. Each of these reactions subsumes a variety of individually interacting components that form feedback loops capable of stabilising or destabilising the oscillating behaviour of the whole system. Kinetic laws within system extensions also employ discretised Michaelis-Menten kinetics for enzymatic processes and linear mass-action kinetics for protein degradation.

Premature dissociation or association of KaiA or KaiB can destabilise the oscillatory behaviour by damping effects. In contrast, spontaneous dephosphorylation and monomer shuffle amplify the influence of feedbacks within the reaction system. This makes the network behaviour more sensitive to slight parameter changes. Toggling KaiB between an active and an inactive form as well as inhibition of KaiC phosphorylation catalysed by KaiB is able to break the symmetry among the reaction cycle.

Premature KaiA association [29]:

$$\mathrm{A} + \mathrm{C}\#\neg\mathrm{A}\# * \#* \longrightarrow \mathrm{C}\#\mathrm{A}\# * \#*$$

Premature KaiA dissociation [29]:

$$\mathrm{C}\#\mathrm{A}\# * \#* \longrightarrow \mathrm{A} + \mathrm{C}\#\neg\mathrm{A}\# * \#*$$

Premature KaiB association [29]:

$$\mathrm{B} + \mathrm{C}\# * \#\neg\mathrm{B}\#* \longrightarrow \mathrm{C}\# * \#\mathrm{B}\#*$$

Premature KaiB dissociation [29]:

$$\mathrm{C}\#*\#\mathrm{B}\#* \longrightarrow \mathrm{B} + \mathrm{C}\#*\#\neg\mathrm{B}\#*$$

Spontaneous dephosphorylation [8, 29]:

$$\begin{aligned}
\mathrm{C}\#*\#*\#6 &\longrightarrow \mathrm{C}\#*\#*\#5\\
\mathrm{C}\#*\#*\#5 &\longrightarrow \mathrm{C}\#*\#*\#4\\
\mathrm{C}\#*\#*\#4 &\longrightarrow \mathrm{C}\#*\#*\#3\\
\mathrm{C}\#*\#*\#3 &\longrightarrow \mathrm{C}\#*\#*\#2\\
\mathrm{C}\#*\#*\#2 &\longrightarrow \mathrm{C}\#*\#*\#1\\
\mathrm{C}\#*\#*\#1 &\longrightarrow \mathrm{C}\#*\#*\#0
\end{aligned}$$

Monomer shuffle in absence of KaiA and KaiB [29]:

$$\mathrm{C}\#\neg\mathrm{A}\#\neg\mathrm{B}\#* \longrightarrow \mathrm{C}\#\neg\mathrm{A}\#\neg\mathrm{B}\#*$$

Toggling KaiB between active and inactive form [19]: A new species Bi is introduced that denotes KaiB in its inactive form. KaiC in its partial or complete phosphorylated state then catalyses the toggling reactions.

$$\begin{aligned}
\mathrm{B} + \mathrm{C}\#*\#*\#3 &\longrightarrow \mathrm{Bi} + \mathrm{C}\#*\#*\#3\\
\mathrm{Bi} + \mathrm{C}\#*\#*\#6 &\longrightarrow \mathrm{B} + \mathrm{C}\#*\#*\#6
\end{aligned}$$

Inhibition of KaiC phosphorylation [8]: Here, the additional string-object $\mathrm{C}\#*\#\mathrm{B}\#i$, $i \in \{0,\ldots,3\}$ acts as an inhibiting factor for phosphorylating reactions $R_1,\ldots,R_6$.

4 Conclusions

Coping with incomplete information about protein activation states can be seen as a challenging task in systems biology. Particularly, the number of individual protein interactions that can potentially occur grows exponentially with regard to the number of binding sites for activation. In order to conduct exhaustive studies about the variety of potential behavioural scenarios of an entire network that includes unknown parts, all corresponding subnetworks covering these unknown parts have to be considered. Incorporation of regular expressions for representation of proteins and their activation states enables usage of placeholder symbols to express arbitraryness or uncertainty about components within those states. In this way, a wild-carded representation may subsume a combinatorial variety of individual activation states.

Accordingly, the proposed P systems framework Π_{CSM} intends to combine advantages of processing regular expressions that represent molecular entities with the corresponding dynamical behaviour of an entire reaction network resulting

from superpositioning of individual molecular abundance courses. To this end, we have integrated string-objects into a deterministic framework able to emulate discretised forms of reaction kinetics in concert with dedicated matching strategies in order to identify reactants from the current system configuration. A simulation study of the KaiABC oscillator demonstrates the practicability of this approach.

From an algebraic point of view, oscillations that occur in structural or configural dynamics of P systems can be detected using a backtracking mechanism along with the temporal system evolution: By monitoring the overall configurations over time, a derivation tree is obtained. Stable oscillations appear as recurring, but non-adjacent overall configurations along a path through the derivation tree. Equipping P systems analysis tools with such a backtracking mechanism is a promising idea for futural work.

Acknowledgements

We gratefully acknowledge funding from the German Federal Ministry of Education and Research (BMBF, project no. 0315260A) within the Research Initiative in Systems Biology (FORSYS).

References

1. U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits.* Chapman & Hall, 2006
2. D. Angluin. *Finding Patterns Common to a Set of Strings.* Journal of Computer and System Sciences 21:46-62, 1980
3. A.P. Arkin. *Synthetic Cell Biology.* Current Opinion in Biotechnology 12(6):638-644, 2001
4. I.M. Axmann, S. Legewie, H. Herzel. *A Minimal Circadian Clock Model.* Genome Inform 18:54-64, 2007
5. F. Bernardini, V. Manca. *Dynamical aspects of P systems.* BioSystems 70:85-93, 2003
6. M.L. Blinov, J.R. Faeder, B. Goldstein, W.S. Hlavacek. *BioNetGen: Software for Rule-Based Modeling of Signal Transduction Based on the Interactions of Molecular Domains.* Bioinformatics 20:3289-3292, 2004
7. M.L. Blinov, J.R. Faeder, B. Goldstein, W.S. Hlavacek. *A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity.* BioSystems 83:136-151, 2006
8. S. Clodong, U. Dühring, L. Kronk, A. Wilde, I.M. Axmann, H. Herzel, M. Kollmann. *Functioning and robustness of a bacterial circadian clock.* Molecular Systems Biology 90(3):1-9, 2007
9. K.A. Connors. *Chemical Kinetics.* VCH Publishers, 1990
10. R. Eils, A. Kriebe (Eds.), *Computational Systems Biology.* Academic Press, 2005
11. S.S. Golden, V.M. Cassone, A. LiWang. *Shifting Nanoscopic Clock Gears.* Nature Structural and Molecular Biology 14:362-363, 2007
12. R. Heinrich, S. Schuster. *The Regulation of Cellular Systems.* Springer-Verlag, 2006

13. T. Hinze, T. Lenser, P. Dittrich. *A Protein Substructure Based P System for Description and Analysis of Cell Signalling Networks.* In H.J. Hoogeboom, G. Păun, G. Rozenberg, A. Salomaa (Eds.), Membrane Computing, Series Lecture Notes in Computer Science 4361:409-423, 2006
14. T. Hinze, R. Fassler, T. Lenser, P. Dittrich. *Register Machine Computations on Binary Numbers by Oscillating and Catalytic Chemical Reactions Modelled using Mass-Action Kinetics.* International Journal of Foundations of Computer Science 20(3):411-426, 2009
15. E. Klipp, R. Herwig, A. Kowald, C. Wierling, H. Lehrach. *Systems Biology in Practice: Concepts, Implementation, and Application.* Wiley-VCH, 2006
16. T. Lenser, T. Hinze, B. Ibrahim, P. Dittrich. *Towards Evolutionary Network Reconstruction Tools for Systems Biology.* In E. Marchiori, J.H. Moore, J.C. Rajapakse (Eds.), Proceedings Fifth European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics (EvoBIO2007), Series Lecture Notes in Computer Science 4447:132-142, 2007
17. M.O. Magnasco. *Chemical Kinetics is Turing Universal.* Physical Review Letters 78(6):1190-1193, 1997
18. V. Manca, L. Bianco, F. Fontana. *Evolution and oscillation in P systems: Applications to biological phenomena.* Lecture Notes in Computer Science 3365:63-84, 2005
19. F. Miyoshi, Y. Nakayama, K. Kaizu, H. Iwasaki, M. Tomita. *A Mathematical Model for the Kai-Protein-Based Chemical Oscillator and Clock Gene Expression Rhythms in Cyanobacteria.* Journal of Biological Rhythms 22(1):69-80, 2007
20. T. Mori, D.R. Williams, M.O. Byrne, X. Qin, M. Egli, H.S. Mchaourab, P.L. Stewart, C.H. Johnson. *Elucidating the Ticking of an In Vitro Circadian Clockwork.* PLoS Biology 5(4):841-853, 2007
21. M. Nakajima, K. Imai, H. Ito, T. Nishiwaki, Y. Murayama. *Reconstitution of Circadian Oscillation of Cyanobacterial KaiC Phosphorylation in vitro.* Science 308:414-415, 2005
22. D.A. Paranjpe, V.K. Sharma. *Evolution of Temporal Order in Living Organisms.* Journal of Circadian Rhythms. 3:7, 2005
23. G. Păun. *Computing with Membranes.* Journal of Computer and System Sciences 61(1):108-143, 2000
24. E. Rosato. *Circadian Rhythms: Methods and Protocols.* Springer-Verlag, 2007
25. M.R. Roussel, D. Gonze, A. Goldbeter. *Modeling the differential fitness of cyanobacterial strains whose circadian oscillators have different free-running periods.* J Theor Biol 205(2):321-340, 2000
26. S. Schuster, I. Zevedei-Oancea. *A Theoretical Framework for Detecting Signal Transfer Routes in Signalling Networks.* Comput. Chem. Eng. 29:597-617, 2005
27. J. Tomita, M. Nakajima, T. Kondo, H. Iwasaki. *No transcription-translation feedback in circadian rhythm of KaiC phosphorylation.* Science 307:251-254, 2005
28. Y. Xu, T. Mori, C.H. Johnson. *Circadian clock-protein expression in cyanobacteria: rhythms and phase-setting.* EMBO Journal 19:3349-3357, 2007
29. M. Yoda, K. Eguchi, T.P. Terada, M. Sasai. *Monomer-Shuffling and Allosteric Transition in KaiC Circadian Oscillation.* PLoS ONE 5:1-7, 2007