

MASARYKOVA UNIVERZITA  
FAKULTA INFORMATIKY



# Formálny popis fosforylačných stavov a súvisiacich reakcií

BAKALÁRSKA PRÁCA

Matej Troják

Brno, jar 2015

## Prehlásenie

Prehlasujem, že táto bakalárska práca je mojím pôvodným autorským dielom, ktoré som vypracoval samostatne. Všetky zdroje, pramene a literatúru, ktoré som pri vypracovaní používal alebo z nich čerpal, v práci riadne citujem s uvedením úplného odkazu na príslušný zdroj.

Matej Troják

**Vedúci práce:** RNDr. David Šafránek, Ph.D.

## Podakovanie

Ďakujem vedúcemu práce za aktívny prístup k vedeniu práce a za všetky poskytnuté rady. Ďalej by som rád poďakoval Jakubovi Šalagovičovi za pomoc pri napĺňaní spoločnej časti práce a Tadeášovi Dědovi za vecné rady a pripomienky k mojej práci.

## Zhrnutie

V prvej kapitole svojej práce v skratke vysvetľujem dôležité pojmy, s ktorými ďalej pracujem. Medzi ne patria biologické procesy, ako fosforylácia a cirkadiálny rytmus, ale tiež niektoré informatické a modelovacie prístupy, ako modelovanie založené na pravidlách.

V druhej kapitole svojej práce transformujem proteínové reakcie modelu Hertel et al. 2013 [1] a modelu Miyoshi et al. 2007 [2] do formálneho jazyka BNGL (BioNetGen Language) [3]. Tu sa vyskytlo mnoho ťažkostí spôsobených tým, že samotný modelár často zabudne, že jeho model vystihuje niečo reálne zo života – v tomto prípade biologický proces – a vníma svoj model čisto iba ako matematiku. Na konkrétne detailné otázky o abstrakcii, ktorú použil, nepozná odpoveď. Tu sa prejavuje, že každý sa na problém pozerá len zo svojho hľadiska a súvislosti zanedbáva.

V tretej kapitole využívam znalosti získané v prvých dvoch kapitolách a navrhujem spoločný biochemický priestor pre vybranú problematiku, vymedzujem vzťahy medzi modelmi a navrhnutým systémom. Tiež pomocou jasne definovanej sémantiky a vzťahov k biochemickému priestoru vyjadrujem vzťahy medzi oboma modelmi.

V poslednej kapitole overujem konzistenciu navrhnutého biochemického priestoru. To znamená, že sa snažím nájsť spôsoby, ktorými by išlo dokázať, že systém je správne navrhnutý.

## Abstract

In the first chapter I briefly explain the most important facts from field of Biology and Informatics. This includes phosphorylation, circadian clock and rule-based modeling.

In the second chapter I translate protein reactions from models Hertel et al. 2013 [1] and Miyoshi et al. 2007 [2] to the formal rule-based language BNGL (BioNetGen Language) [3]. Here I had to face up with many problems, including insufficient documentation and abstractions of the model. Problem is the modeler often considers his model from point of Mathematics, not the Biology. Here can be seen the fact that every field has its own point of view and relationships are neglected.

In the third chapter I propose biochemical space for circadian clock using rule-based approach, I define relationships between models and the space. Thanks to defined semantics I express relationships between the models themselves.

In the last chapter I check the structure consistency of proposed biochemical space. It means I am trying to prove correctness of the system.

## **Kľúčové slová**

Biologický model, formálny zápis, biochemický priestor, tvorenie komplexov, prepisovacie pravidlá

## **Key words**

Biological model, formalism, biochemical space, complexation, rewriting rules

# Obsah

1	Úvod . . . . .	2
2	Oblasť záujmu . . . . .	4
2.1	Fosforylácia . . . . .	4
2.2	Cirkadiálny rytmus . . . . .	4
2.3	Cirkadiálny rytmus v modeloch . . . . .	6
2.3.1	Model Hertel et al. . . . .	6
2.3.2	Model Miyoshi et al. . . . .	7
2.4	Modelovanie založené na pravidlách . . . . .	8
2.4.1	Reakčné pravidlá . . . . .	9
3	Formálny prepis vybraných modelov . . . . .	10
3.1	Model Hertel et al. . . . .	10
3.2	Model Miyoshi et al. . . . .	13
4	Návrh spoločného biochemického priestoru . . . . .	16
4.1	Porovnanie BCS s príslušnými modelmi . . . . .	19
4.1.1	Model Hertel et al. . . . .	19
4.1.2	Model Miyoshi et al. . . . .	21
4.2	Porovnanie modelov . . . . .	22
5	Analýza biochemického priestoru . . . . .	25
5.1	Dostupnosť stavu . . . . .	25
5.2	Reverzibilita . . . . .	26
5.3	Ohraničenosť . . . . .	28
6	Záver . . . . .	29
A	Pôvodná verzia biochemického priestoru . . . . .	34
B	Biochemický priestor v BNGL . . . . .	36
C	Zoznam elektronických príloh . . . . .	38

# 1 Úvod

Systémová biológia je vedná disciplína vyžadujúca si znalosť niekoľkých odborov, predovšetkým biológie, chémie, informatiky a matematiky. Biológia pri pohľade na daný problém používa redukcionizmus, čiže zjednodušenie. Matematika sa na rovnaký problém díva z vlastného hľadiska, ktoré je často ťažko pochopiteľné. Systémová biológia sa snaží o nájdenie spoločnej reči, ktorá je zrozumiteľná ako pre obe strany, tak i pre človeka nestranného.

Cieľom systémovej biológie je pochopiť detaily a princípy biologických systémov. K tomu je potrebný systémový pohľad od jednotlivých molekúl až po systém celkovo. Je nevyhnutné pochopiť všetky látky, aký je ich význam a zmysel, a tiež pochopiť interakcie medzi nimi. Tak, ako tvrdí Kitano vo svojej práci [4], túto oblasť čaká ešte veľa práce, najmä čo sa vývoja technológií týka, softvéru aj hardvéru, ale tiež z pohľadu skúmania.

Často sa v tejto oblasti pracuje s matematickými modelmi, ktoré sú vytvorené na základe biologických dát. Tie sú porovnávané s reálnymi nameranými dátami a vyhodnocované. Neprijemné je, že tieto modely sú väčšinou vyjadrené formou, ktorej notácia je príliš náročná, napríklad diferenciálnymi rovnicami. Preto je snahou systémovej biológie navrhnúť vyjadrovaciu formu, s ktorou by sa ľahko pracovalo a bola pochopiteľná pre každého so základným systémovým myslením.

Pre človeka bližšie vyjadrenie biologických procesov u siníc bola vytvorená webová platforma [e-cyanobacterium.org](http://e-cyanobacterium.org). Cieľom tejto platformy je vytvoriť kompletný výpočtový model siníc. Hlavnou myšlienkou je správna anotácia a mapovanie jednotlivých modelov do spoločného biochemického priestoru. To sa čiastočne dosiahlo [5], problémom sú niektoré procesy, najmä tie so stavovou explóziou.

Keďže procesy v tomto jednobunkovom organizme sú dosť komplexné a biochemický priestor má byť vyjadrený detailne a presne, zvolil sa na jeho vyjadrenie v celku málo využívaný prístup, a to zápis pomocou prepisovacích pravidiel. Tie poskytujú jasný, kompaktný a prehľadný popis biologických procesov. Komplexná reakčná sieť je v nich zahrnutá implicitne. To môže priniesť mnoho výhod, najmä fakt, že biochemický priestor bude mať svoju pevnú kostru, na ktorej bude všetko založené.



K tomuto stručnému zápisu je potrebná aj dávka opatrnosti. Implícitné vyjadrenie reakčnej siete si vyžaduje naozaj podrobné vedomosti v oblasti modelovania založeného na pravidlách a syntaxe konkrétneho použitého jazyka. Lahko sa môže stať, že pravidlá implikujú iné reakcie, než sú vyžadované. Preto treba byť obozretný, inak môže napríklad dôjsť k vzniku nekonečného stavového priestoru, či nezmyselných látok. Výhodou je, že tento prístup umožňuje ľubovoľne ohýbať prísnosť daného pravidla, od všeobecnej aplikácie až na úroveň konkrétnu, čiže ekvivalentnú s jedinou reakciou.

Cieľom mojej práce bolo vyjadriť vybrané modely cirkadiálnych rytmov a spoločný biochemický priestor skrývajúci sa za týmito procesmi práve modelovacím jazykom založeným na pravidlách. Tým dosiahnuť pomerne jednoduché a jasné vymedzenie vzťahov medzi modelmi, čo prinesie informácie pre ďalšie spracovanie modelov a vylepšenie biochemického priestoru na [e-cyanobacterium.org](http://e-cyanobacterium.org). Tiež to prinesie zistenie, nakoľko správny je doterajší zápis fosforylácie ako veľmi komplexného procesu a aj návrh riešenia ako tento proces zapísať.

Súčasťou práce je analýza konzistencie navrhnutého biochemického priestoru, najmä zo štruktúrneho hľadiska. Na to je potrebné najprv definovať vlastnosti, ktoré by mal spĺňať a overiť ich buď pomocou vhodných nástrojov alebo navrhnúť iný spôsob overovania.

## 2 Oblasť záujmu

V tejto práci sa používajú rôzne odborné pojmy z biologickej sféry. Preto sú v nasledujúcich podkapitolách vysvetlené najdôležitejšie fakty, prípadne odkazy na relevantnú literatúru.

### 2.1 Fosforylácia

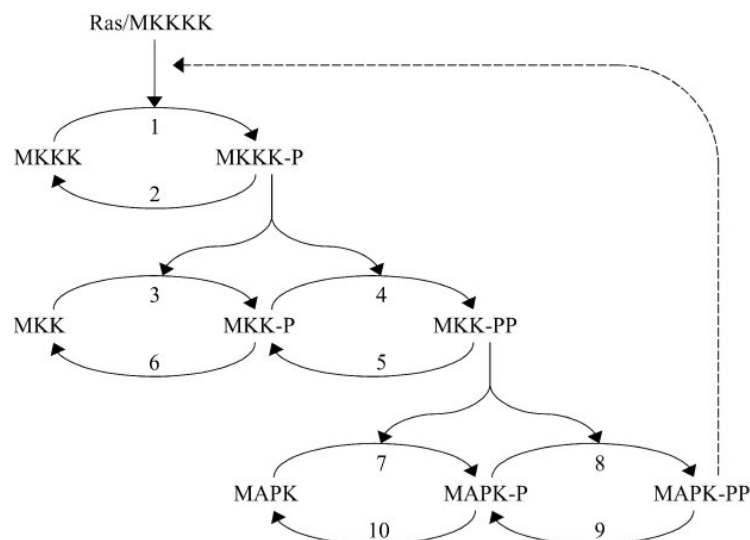
Fosforylácia [6] je jedným z najvýznamnejších mechanizmov bunky. Prebieha najčastejšie na aminokyselinách, ktoré sú začlenené do proteínovej štruktúry. Konkrétne sú to serín, treonín, tyrosín, výnimočne aj arginín, histidín a lysín. Stimulom pre tento proces sú vonkajšie vplyvy, ktoré sa prejavujú prijatím signálu pomocou receptora na bunkovej stene. Pre popis systému je to jeden z najzložitejších procesov, pretože fosforylácia je často začlenená vo veľmi zložitých mechanizmoch a v komplexoch proteínov môže spôsobovať rozmanité zmeny, ktoré vedú k obrovským kombinatorickým explóziám, čiže vzniku obrovského stavového priestoru.

Príkladom, kde sa fosforylácia vyskytuje je signálna dráha [6]. Je to spôsob, ktorým bunka komunikuje. Existujú dva základné typy: komunikácia v rámci bunky a medzi bunkami. *Intercelulárna* (medzibunková) komunikácia má tri základné kroky: vylúčenie signálu zdrojovou bunkou, prenos signálu a prijatie signálu cieľovou bunkou. Pod pojmom signál sa myslí špecifická signálna molekula o veľkosti od malej, aminokyseliny podobnej molekuly, až po rôzne proteíny. *Intracelulárna* (v rámci bunky) komunikácia je zabezpečená tzv. *kaskádou proteínov* (Obrázok 1). Šírenie je zabezpečené postupnou zmenou aktivity jednotlivých proteínov v tejto kaskáde.

Signálne dráhy sú typické najmä pre eukaryotické bunky. V prokaryotických bunkách sú typickejšie iné procesy. Príkladom sú cirkadiálne rytmy (Sekcia 2.2).

### 2.2 Cirkadiálny rytmus

Cirkadiálny rytmus [8] (ďalej už len CR) sa prejavuje oscilačným chovaním s periódou približne 24 hodín a je typický pre široké spektrum



Obr. 1: Signálna dráha MAPK [7] predstavujúca fosforylačnú kaskádu proteínov.

organizmov od cicavcov (vrátane ľudí) až po jednobunkové organizmy. Tento rytmus riadi základné procesy, od exprese génov až po metabolizmus. Cyklus má priamy súvis so striedaním dňa a noci a je s ním zosynchronizovaný. Presná funkcia CR a to, ako presne funguje, zatiaľ nie je známa u zložitých organizmov z dôvodu vysokej komplexnosti. Zistilo sa [9], že najjednoduchší CR je u siníc, teda jednobunkových baktérií schopných produkovať kyslík prostredníctvom fotosyntézy.

CR u sinice je tvorený tromi proteínmi KaiA, KaiB a KaiC. Tento mechanizmus je najlepšie vysvetlený pre kmeň *Synechococcus elongatus* PCC 7942. Najdôležitejšiu úlohu má práve KaiC. V porovnaní s ostatnými dvoma proteínmi je štruktúrne najväčší. Hrá významnú rolu v aktivácii/represii prepisu (transkripcie) vlastného génu. Vo svojej štruktúre skrýva dve fosforylačné miesta – treonín a serín. Z toho vyplýva, že sa môže vyskytovať v štyroch rôznych stavoch: nefosforylovaný KaiC{u}, treonín fosforylovaný KaiC{t}, serín fosforylovaný KaiC{s} a plne fosforylovaný KaiC{d}. Tie sú prístupné, iba keď sa šesť KaiC proteínov spojí do jedného celku a vytvoria tzv. hexamér. Tým sa jednotlivé monoméry konformačne zmenia a sprístupnia tak obe fosforylačné miesta. Tieto komplexy sú následne fosforylované proteínmi KaiA počas dňa a defosforylované proteínmi KaiB počas noci.

Tu treba poznamenať dve dôležité veci. Prvou je, že KaiB spôsobuje inaktiváciu proteínov KaiA, teda nespôsobí priamo defosforyláciu, ale zabráni fosforylácii a hexamér potom samovoľne vedie k defosforylácii. To súvisí s druhým významným faktom a to, že celý fosforylačný cyklus prebieha aj samovoľne, teda bez prítomnosti ostatných proteínov, ale v oveľa menších merítkach.

Celý CR je teda založený na postupnom fosforylovaní a defosforylovaní KaiC6 hexamérov a vytváraní komplexov zo všetkých troch proteínov. Kľúčovú rolu hrá KaiC, pretože je chopné väzby na nukleotidy, čo vytvára regulačný charakter translácie. Nakoľko každý monomér môže nadobúdať štyri rôzne fosforylačné stavy, pre hexamér to vedie ku kombinatorickej explózii. Celkovo tak hexamér môže nadobúdať  $4^6$  stavov, čo je 4096. Keď sa do úvahy vezmú aj všetky rôzne štruktúry tvorené všetkými tromi proteínmi, dostávame obrovský stavový priestor.

## 2.3 Cirkadiálny rytmus v modeloch

Na portáli [e-cyanobacterium.org](http://e-cyanobacterium.org)<sup>1</sup> [10] (ďalej už len e-cyano) sú prístupné dva modely, ktoré opisujú dynamiku cirkadiálnych rytmov: Model Hertel et al. 2013 a Model Miyoshi et al. 2007. Oba v istom rozsahu reflektujú biológiu, ktorá sa za nimi skrýva. Takisto prinášajú množstvo abstrakcií, ktoré napomáhajú k jednoduchšiemu uchopeniu problému.

### 2.3.1 Model Hertel et al.

Tento model [1] kombinuje cirkadiálne transkripčno-translačné rytmy spolu s KaiABC proteínovým oscilátorom v sinici. K tomu sú využité základné proteíny KaiA, KaiB a KaiC, pričom KaiC sa môže nachádzať v štyroch rôznych fosforylačných stavoch, a ich mRNA (jedno pre KaiA a druhé spoločné pre KaiB a KaiC). Ďalej je to komplex KaiC6 tvorený zo šiestich monomérov KaiC a komplex KaiBC tvorený šiestimi monomérmi KaiB a KaiC. Oba sa môžu vyskytovať v štyroch rôznych stavoch. Dôležité je podotknúť, že KaiA netvorí komplexy. Jeho úloha je teda čisto enzymatická. Viac informácií o jednotlivých proteínoch sa nachádza v sekcii 3.1.

Model je tiež tvorený reakciami, ktoré sa dajú zaradiť do niekoľkých

1. <http://www.e-cyanobacterium.org>

kategórií. Prvou je skladanie/rozpad komplexov. Druhou je fosforylácia/defosforylácia, ktorá sa deje iba na komplexoch, nikdy nie na samostatných monoméroch. Ďalšou triedou sú degradácie. V nej je použitá zvláštna abstrakcia. Umožňujú totiž degradáciu všetkých komplexov vrátane monomérov. Mierne nezmyselné je, že majú rovnakú kinetiku. V prírode je to tak, že komplex by mal byť výrazne stabilnejší a teda aj pomalšie degradovať [6]. Predposlednou kategóriou sú transkripcie, čiže prepis proteínov z mRNA. Nakoniec sú to tzv. *exchange*, pod ktorými sa rozumie spojenie monomérov KaiB s komplexom KaiC6 za vzniku komplexu KaiBC.

Všetky reakcie sú špecificky parametrizované a výsledkom deterministickej simulácie sú oscilácie s pravidelnou periódou. Oscilácie sú po zahájení simulácie najprv nestále, ale po asi dvoch cykloch sa ustália a produkujú pravidelné kmity s rovnakou amplitúdou. Nakoľko na model nevplyvajú žiadne vonkajšie vplyvy, je tento priebeh očakávaný.

### 2.3.2 Model Miyoshi et al.

Model Miyoshi et al. [2] tiež opisuje cirkadiálne rytmy v sinici. Využíva takisto všetky tri základné proteíny KaiA, KaiB a KaiC. Proteín KaiC sa ale môže vyskytovať iba v dvoch rôznych stavoch, čo pre komplex KaiC6 znamená, že môže nadobúdať tri rôzne konformácie. Je to buď KaiC6 tvorený šiestimi nefosforylovanými monomérmi, PPKaiC6 (partially phosphorylated) tvorený tromi fosforylovanými a tromi nefosforylovanými monomérmi a nakoniec CPKaiC6 (completely phosphorylated) tvorený šiestimi fosforylovanými monomérmi. Nie je špecifikované, o akú fosforyláciu sa jedná.

Ďalšou zložkou je komplex KaiB4, ktorý môže byť vo svojej aktívnej alebo neaktívnej forme. Môže sa viazať s dimerom KaiA2 za vzniku nového komplexu KaiA2B4, ktorý má enzymatický a defosforylačný význam. Reakcie sú obdobné ako v časti 2.3.1 s tým, že pribudli reakcie na tvorbu komplexov okolo KaiB4.

Čo sa funkčnosti modelu týka, poskytuje oscilácie, ktorých amplitúda sa stále mení, ale nie výrazne. To znamená, že je do modelu zahrnutý aj určitý stupeň ruchu. Tiež poskytuje dve simulačné sady – so svetlom a bez svetla. Funkčne je to priamo zabezpečené parametrom  $L_{-}$ , ktorý má buď hodnotu jedna (svetlo svieti) alebo nula (svetlo nesvieti). To priamo ovplyvňuje, či sa jednotlivé proteíny prepisujú z mRNA, a či

degradujú. Výsledkom je, že za nedostatku svetla oscilácie prebiehajú, ale s výrazne nižšou amplitúdou.

## 2.4 Modelovanie založené na pravidlách

Každý biologický proces tvoria dve hlavné časti: reakcie a látky. Reakcia je vyjadrením vzťahu medzi látkami. Pri modelovaní takéhoto procesu si treba tento priestor rozšíriť, pretože naň vplýva prostredie a iné faktory.

Matematický model biologického procesu je charakteristický tým, že má svoje množiny látok, reakcií a parametrov. V rámci každej z týchto skupín môžeme identifikovať niekoľko vlastností. Látka je elementárna jednotka, ktorá v tomto modeli vystupuje ako reaktant, produkt alebo enzým [11]:

- **Reaktant** je látka, ktorá vstupuje do reakcie. Po uskutočnení reakcie zaniká.
- **Produkt** je výsledkom reakcie, výstup reakcie.
- **Modifikátor** má v reakcii katalytickú úlohu. Ovplyvňuje kinetiku reakcie, nevzniká ani nezaniká, rolu hrá iba jeho prítomnosť.

Každá látka je použitá v reakciách. Hrá v nej jednu z vyššie uvedených rolí. V biológii existuje množstvo rôznych typov reakcií, ktoré môžu byť veľmi komplexné a zložité. Vo všeobecnosti sa všetky dajú zapísať ako sústava tzv. elementárnych reakcií. Sú to najjednoduchšie reakcie, z ktorých sa dajú zostaviť zložitejšie reakcie.

Cieľom práce je podchytiť tieto tri množiny formálne a vyvodiť závery. Používam k tomu jazyk BNGL [3], čo je tzv. *rule-based* jazyk. Pri tejto činnosti sa manuálnemu prepisovaniu reakcie po reakcii nedá vyhnúť. Problémom je, že si treba uvedomiť, o aký typ reakcie sa jedná, či je reverzibilná (môže prebiehať aj v opačnom smere) a aké funkcie majú jednotlivé látky a parametre. Na to je potrebné dopodrobna rozumieť danému matematickému modelu a súvisiacej biológii.

Až po dôkladnom naštudovaní a pochopení celého modelu môže dôjsť k samotnému procesu prepisovania do BNGL. V tom je zahrnuté preštudovanie článku, v ktorom bol model publikovaný, diskusie

s modelárom a v neposlednom rade štúdium zodpovedajúcich biologických dejov. Štruktúra súboru BNGL je celkom jednoduchá, dôležitý je význam jednotlivých blokov [3]:

- **parameters** - všetky parametre a ich priradené hodnoty,
- **molecule types** - zoznam elementárnych látok, ktoré sa v modeli vyskytujú,
- **seed species** - počiatočné koncentrácie jednotlivých látok v modelovanom prostredí,
- **observables** - množina látok, ktoré chceme pozorovať pri simulácii,
- **functions** - definícia funkcií pre kinetické pravidlá jednotlivých reakcií,
- **reaction rules** - reakčné pravidlá 2.4.1.

#### 2.4.1 Reakčné pravidlá

Blok s najväčším významom pre moju prácu. Nejedná sa o reakcie, ale o reakčné pravidlá – prepisovacie pravidlá. Sú v tvare

$$\text{SUBSTRÁTY} \rightarrow \text{PRODUKTY} \text{ kinetic\_law}$$

kde SUBSTRÁTY je množina reaktantov oddelených znakom + a PRODUKTY je množina produktov rovnako oddelená znakom +. Platí všeobecné pravidlo, že substráty zanikajú a produkty vznikajú, t.j. substráty sa prepisujú na produkty. *Kinetic\_law* je definovaný vzťah, ktorý platí pre kinetiku daného pravidla. Zahŕňa definované parametre, koncentrácie pozorovaných látok a funkcie.

V rámci reakčných pravidiel môžu z jednotlivých látok vznikáť aj komplexy pomocou operátora "!", ktorý vyjadruje tvorenie komplexov. Spolu s operátorom "!", s ktorým vyjadrujú väzbu, tvoria kompletný zápis pre vznik komplexov. Každá väzba v rámci komplexu musí mať svoju unikátnu identifikáciu (ID).

$$\text{KaiA(m)} + \text{KaiA(m)} \rightarrow \text{KaiA(m!1).KaiA(m!1)} \text{ ksdA}$$

### 3 Formálny prepis vybraných modelov

Cieľom je získať formálny zápis matematických modelov reprezentujúcich cirkadiálne rytmy. Vďaka tomu sa tieto modely dajú ďalej využívať a hlavne to vedie k uvedomeniu si, aké procesy a reakcie daný model znázorňuje. To napomáha k rôznym analýzam a k odhaleniu súvislostí medzi modelmi.

#### 3.1 Model Hertel et al.

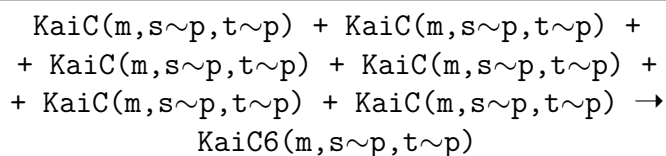
Prvá prípadová štúdia tejto práce. Model Hertel et al. [1] reprezentuje zjednodušenú funkciu cirkadiálnych rytmov (Obrázok 2). Zjednodušenú preto, lebo mnohé veci abstrahuje a vynecháva, čiže generalizuje, alebo naopak, špecializuje, t.j. proces rozoberie detailnejšie.

Prvým dôležitým faktom je, že jadro cirkadiálnych hodiniek tvoria tri proteíny KaiA, KaiB a KaiC. Napriek tomu sa v modeli KaiB neuvažuje, respektíve nepoužíva sa jeho premenné množstvo. Je považované za vždy dostatočné, pracujú s ním ako s nekonečným množstvom. Toto zjednodušenie sa dá vysvetliť tým, že KaiB a KaiC vznikajú z rovnakého génu a vždy naraz, čiže ak je dostatok KaiC, musí byť aj dostatok KaiB. Napriek tomu je to dosť vysoká miera abstrakcie.

Pri transformácii reakcií tohto modelu z webu e-cyano do množiny prepisovacích pravidiel došlo k mnohým problémom. Najväčším problémom bola nejasnosť, čo daná reakcia vyjadruje. Konkrétne komplikácie sa vyskytli pri týchto troch typoch reakcií:

- **Skladanie - assembly**

Jedná sa primárne o vytvorenie hexaméru, ktorý vzniká spojeným šiestich monomérov. Kvôli nejasnému zápisu z BCS som predpokladal, že sa reálne šesť monomérov stretne a vytvorí hexamér. Napríklad (pre zjednodušenie vynechávam *kinetic law* teraz i vo všetkých budúcich príkladoch):

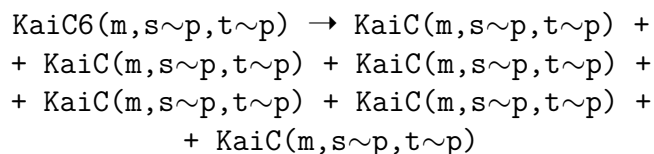




V tejto reakcii sa stretne šesť plne fosforylovaných monomérov za vzniku jedného plne fosforylovaného hexaméru.

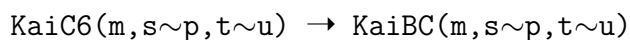
- **Rozpad - dissociation**

Obdobné nejasnosti viedli k tomu, že som mylne predpokladal rozpad hexaméru na šesť monomérov. Uvádzam príklad, ktorý je reverzibilnou reakciou k príkladu vyššie.



- **Zámena - exchange**

Tu bolo nejasné to, že KaiC6 sa *menilo* na KaiBC, teda:

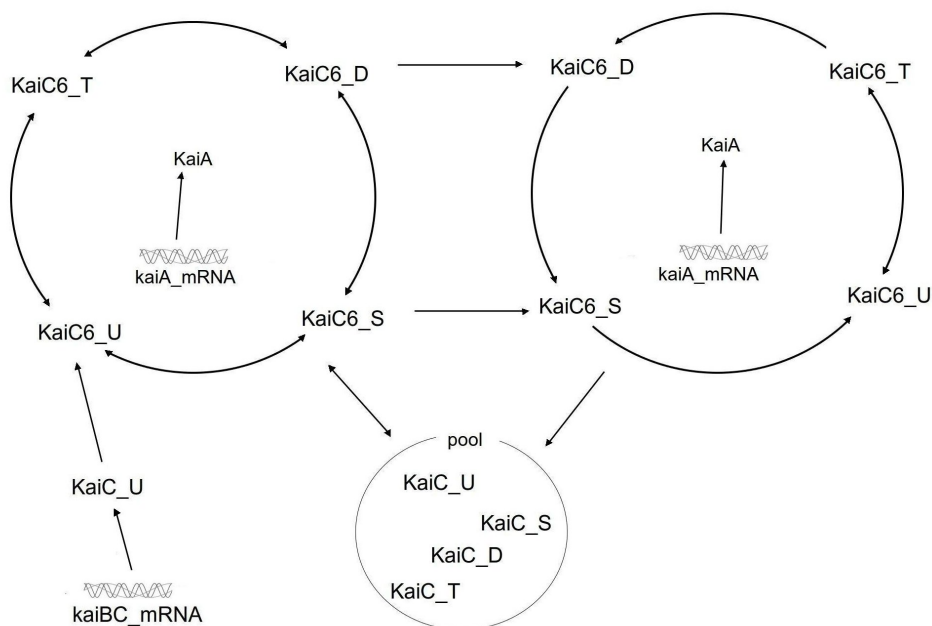


V pôvodnom význame by to nedávalo zmysel, pretože takto by sa niekde strácali látky. Nie je možné, aby z nejakého hexaméru vznikol len jeden monomér. Toto bolo základné zistenie, ktoré vyvrátilo všetky doterajšie predpoklady. Preto bolo potrebné oba modely - prepísané v BNGL aj pôvodný na e-cyano - porovnať korektným spôsobom.

K tomu som využil zabudovaný príkaz jazyka BNGL `writeMfile()`, ktorý vytvorí súbor, v ktorom sú jednotlivé látky modelu a ich dynamické rovnice závislé od času - diferenciálne rovnice. Na portáli e-cyano sa k simulácii používa *Octave solver* [12], ktorý pred simuláciou vytvorí súbor s informáciami, ktoré potrebuje. Medzi nimi sú aj diferenciálne rovnice. Problémom je, že solver nepoužíva priamo názvy látok, ale iba ich skratky, ktoré si vygeneroval. Preto som musel vytvoriť vlastný syntaktický analyzátor (príloha C), ktorý mi z daného octave súboru vyfiltroval a preložil potrebné diferenciálne rovnice do rovnakého formátu, ako `writeMfile()` v BNGL. Vďaka tomu som mohol postúpiť k samotnému porovnaniu oboch súborov.

K tomu poslužil jednoduchý program *diff*, ktorý je v základnej výbave linuxových operačných systémov. Program odhalil, že sa súbor

### 3. FORMÁLNY PREPIS VYBRANÝCH MODELOV



Obr. 2: Schéma vyjadrujúca reakčnú sieť modelu Hertel et al.

líšia v stechiometrii. V modeli z e-cyano sa vôbec nikde nevyskytoval hexamér. Celý proces sa deje iba na základe monomérov. Z tohto sa mohli vyvodiť závery, že formácia hexaméru ako takého neprebíha. Otázkou bolo, čo sa v modeli deje?

Na základe faktu, že hexaméry nevznikajú vysvitlo, že model Hertel et al. neuvažuje formáciu komplexov, ale iba zmenu stavu jednotlivého KaiC monoméru. Po aplikovaní týchto zmien boli pôvodné tri problematické typy reakcií vyriešené takto:

- **Skladanie - assembly**

Nakoľko monomér KaiC nie je reálne začlenený do hexaméru, označenie KaiC6 znamená doslovne *tento konkrétny monomér KaiC je začlenený do hexaméru*, teda iba zmena stavu.



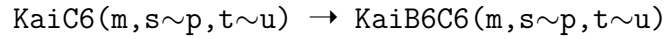
- **Rozpad - dissociation**

S rozpadom je to obdobné, KaiC monomér iba mení stav z *v hexaméri* na *voľný*.



- **Zámena - exchange**

Zámena už dáva zmysel, ale pre korektnosť som látku KaiBC premenoval na KaiB6C6, aby bolo zjavné, že sa monomér vyskytuje v stave hexaméru.

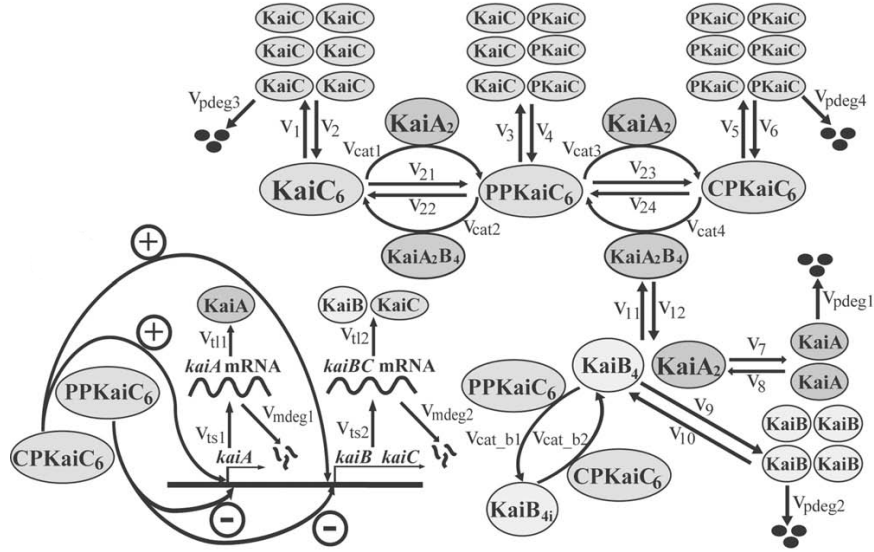


Pri správnej implementácii som dostal model vyjadrený množinou prepisovacích pravidiel v jazyku BNGL. Vďaka jasnému a formálnemu zápisu je vidieť, nakoľko model abstrahuje. Je jednoznačné, že miera abstrakcie je vysoká. Cirkadiálne rytmy ako také sú postavené na formácii hexamérov, no v tomto modeli tento fakt neuvažujú.

### 3.2 Model Miyoshi et al.

Druhým študijným príkladom tejto práce je model Miyoshi et al. [2]. Je to takisto model, ktorý reprezentuje cirkadiálne rytmy. Je výrazne starší, čo znamená, že dokumentácia ohľadne tohto modelu je lepšia a je v nej menej nejasností ako v modeli Hertel et al. (Sekcia 3.1). Na základe získaných skúseností s predchádzajúcim modelom a podrobnej formulácii modelu bolo následné prepísanie do pravidiel jednoduchšie.

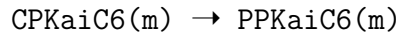
### 3. FORMÁLNY PREPIS VYBRANÝCH MODELOV



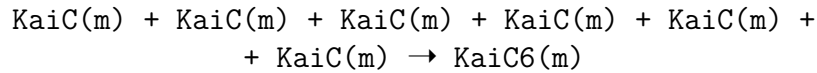
Obr. 3: Reakčná sieť modelu Miyoshi et al. 2007 prevzatá z článku [2].

Na druhej strane, približne pred siedmimi rokmi ešte biológovia nemali také podrobné informácie, a riziko intenzity abstrakcie je ešte vyššie.

Najvýraznejšia abstrakcia sa prejavuje na fosforyláciách. Na hexaméri KaiC6 sa nachádzajú dve fosforylačné miesta - treonín a serín. Miyoshi ich používa ako dve odlišné, hovorí iba o parciálnej fosforylácii. Tým zužuje stavový priestor na nefosforylovaný (-), čiastočne fosforylovaný (partially phosphorylated PP) a plne fosforylovaný (completely phosphorylated CP). Pritom nie sú používané ako stavy, ale nové látky. Napríklad:



V tomto ohľade ide o značnú abstrakciu. Na druhej strane je výhodou, že vytvára reálne hexaméry, čím viac zachytáva realitu.



Formáciou komplexov narážam na problém, ktorý je implicitne zahrnutý v jazyku BNGL. Komplikáciou je, že BNGL je natoľko formálny,

### 3. FORMÁLNY PREPIS VYBRANÝCH MODELOV

---

že berie do úvahy každý detail. Pri spojení viacerých látok dochádza ku kombinatorickým možnostiam, ako môže toto spojenie nastať. Pri šiestich reaktantoch je to  $6! = 720$  možností. Reakčná konštanta tejto reakcie je implicitne vynásobená obrátenou hodnotou tohto čísla, čo výrazne znižuje kinetiku. Táto interná záležitosť zahrnutá v jazyku BNGL sa nedá vypnúť. Preto jediným riešením bolo kinetiku tejto reakcie explicitne vynásobiť zadanou konštantou *hexamér* s priradenou hodnotou 720. Takto som to obdobne urobil pre všetky reakcie, kde sa tento problém vyskytoval.

Dôležitým aspektom tohto modelu je, že aplikuje spätnú väzbu na transkripciu. Je to vidieť na Obrázku 3. Ďalšie problémy spojené s týmto modelom sú spomenuté v sekcii 4.2.

## 4 Návrh spoločného biochemického priestoru

Na základe dát a informácií od biológov mohol vzniknúť biochemický priestor [5]. Problémom je, že zdrojom boli rovnakí biológovia, ktorí aj poskytli svoje modely, a dôveryhodnosť týchto informácií mohla byť ovplyvnená ich abstraktným pohľadom na vec. V pôvodnom BCS na e-cyano (príloha A) nie je zohľadnený základný fakt. Jedná sa o fosforyláciu hexamérov. V modeloch sa všeobecne používa abstrakcia, ktorá celý hexamér v jednom jedinom kroku posunie hneď do iného fosforylačného stavu ako celok. Napríklad:

$$\text{KaiC6}(m, s \sim u, t \sim u) \rightarrow \text{KaiC6}(m, s \sim p, t \sim u)$$

Takto to v biológii nie je a keďže BCS má odrážať realitu, tak nie je správne navrhnutý. Táto fosforylácia totiž prebieha postupne po jednotlivých monoméroch. Na komplex sa naviaže KaiA2 dimer a ofosforyluje jeden monomér na konkrétnom treonínovom alebo serínovom mieste. Takto vzniká oveľa širší stavový priestor. Z pôvodných štyroch stavov pre hexamér sa dostávame na variácie s opakovaním  $V_4(6) = 4^6 = 4096$ . Je to spôsobené tým, že každý monomér môže byť v rámci komplexu fosforylovaný samostatne a nie sú kladené žiadne obmedzenia v akom poradí. Konzistenciou a dosiahnuteľnosťou sa zaoberám v kapitole 5.

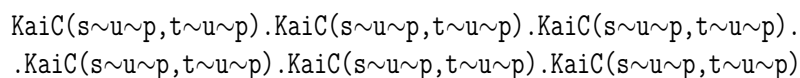
K tomu je potrebné zmeniť zápis hexamérov. S doterajšími možnosťami boli možné iba vyššie spomenuté štyri stavy. Ako príklad budem v ďalšej časti používať KaiC6 hexamér, ostatné komplexy sú schematicky zaznamenané v prílohe B. Ako riešenie som zvažoval dve možnosti. Prvou je, že komplex KaiC6 bude mať šesť väzbových miest a každé sa môže vyskytovať v štyroch stavoch:

$$\text{KaiC6}(m1 \sim u \sim s \sim t \sim d, m2 \sim u \sim s \sim t \sim d, m3 \sim u \sim s \sim t \sim d, \\ m4 \sim u \sim s \sim t \sim d, m5 \sim u \sim s \sim t \sim d, m6 \sim u \sim s \sim t \sim d)$$

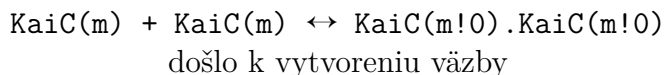
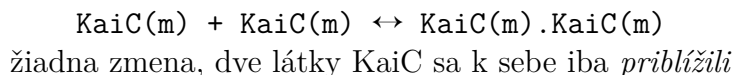
Napriek tomu, že takýto zápis vyhovuje požiadavkám a všetky stavy sú dosiahnuteľné, má dve veľké nevýhody: neprehľadnosť a fakt, že takýto zápis je pre jazyk BNGL kombinatoricky príliš náročný.

#### 4. NÁVRH SPOLOČNÉHO BIOCHEMICKÉHO PRIESTORU

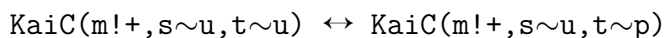
Lepší zápis je pomocou jednotlivých monomérov a vytváraním komplexov medzi nimi, ktorý so sebou prináša mnoho výhod. V prvom rade je to vyššia prehľadnosť. Každý monomér je na oboch fosforylačných miestach buď forforylovaný alebo nie:



Tento zápis je zjednodušený. V skutočnosti BNGL potrebuje vždy špecifikovať, medzi ktorými dvoma látkami je väzba utvorená. Operátorom ”.” totiž dôjde len k priblíženiu dvoch látok, ale k spojeniu dôjde až po označení väzby. Zjednodušenie som použil pre väčšiu prehľadnosť zápisu, nakoľko pre potreby BCS špecifikácia väzby nie je dôležitá.

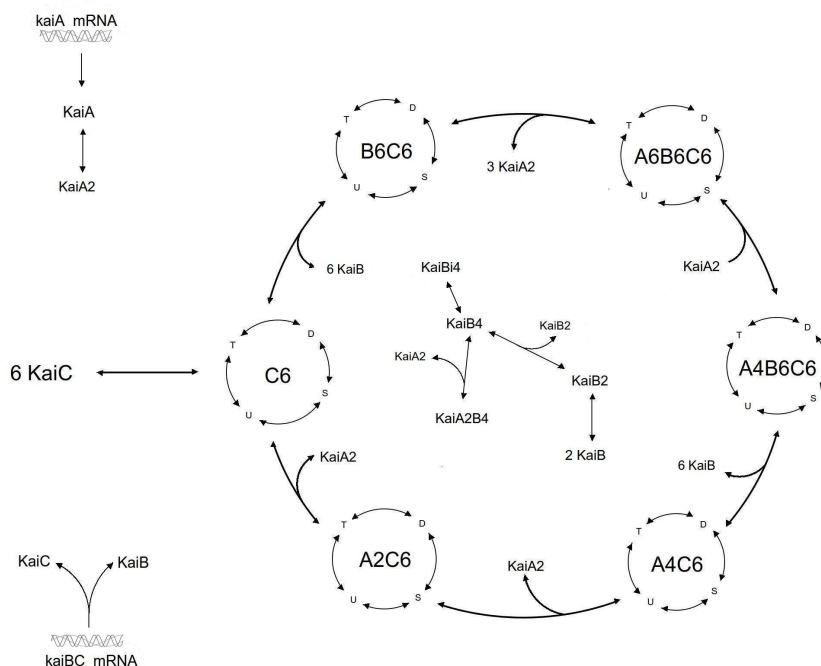


Takýto spôsob zápisu je síce menej prehľadný oproti pôvodnému BCS, ale je korektný. Navyše, vďaka tomu sa dá oveľa kompaktnejšie zapísať fosforylácia. Keďže pôvodne sa každý komplex nazýval inak, bolo nutné pre každý definovať jeho vlastný fosforylačný cyklus. Teraz je každý komplex zapísaný pomocou najmenších zmysluplných jednotiek – monomérov – vďaka čomu je možné fosforyláciu ako takú vyjadriť komplexne. K tomu slúži vlastnosť jazyka BNGL, ktorá umožňuje výber podmnožín. Takže namiesto zdĺhavého vypisovania stačí napísať:



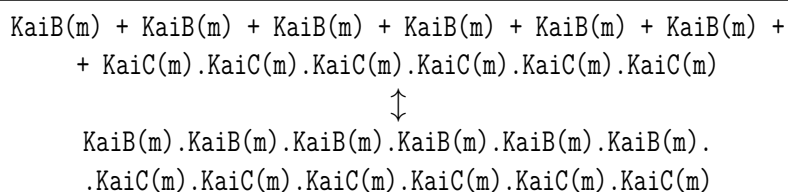
Zápis vyjadruje, že ak je monomér KaiC v plne nefosforylovanom stave naviazaný na akúkoľvek inú látku, tak prebehne fosforylácia na treonínovom mieste. Toto pravidlo sa aplikuje na každý komplex, v ktorom je KaiC(m!+, s ~ u, t ~ u) podmnožinou, takže na všetky hexaméry KaiC6 a všetky ostatné komplexy (KaiB6C6, KaiA6B6C6...). Vďaka tejto vlastnosti sa počet prepisovacích pravidiel výrazne znížil. Všetky prepisovacie pravidlá môžete nájsť v prílohe B.

#### 4. NÁVRH SPOLOČNÉHO BIOCHEMICKÉHO PRIESTORU



Obr. 4: Schematický zápis BCS. Veľký cyklus zjednodušene vyjadruje, že prechod medzi jednotlivými komplexmi je možný v akomkoľvek fosforylačnom stave. V rámci malých cyklov pre jednotlivé komplexy je možný každý stav zo stavového priestoru, nie len význačné U, S, T a D.

Vďaka komplexnému zápisu fosforylácii sa zvyšok BCS dá zhrnúť do pravidiel formujúcich, respektíve rozpadových, transkripčných a de-gradačných. Pre čiastočne prehľadný zápis v nasledujúcej časti vynechám špecifikáciu väzby medzi látkami. Príkladom formujúceho pravidla je formácia komplexu KaiB6C6:



O stavoch tu nie je ani zmienka. Nie je to potrebné, takýto komplex, tak ako aj ostatné, sa môže tvoriť nezávisle na stavoch jednotlivých látok. O dosiahnutie ostatných konkrétnych stavov sa potom starajú



fosforylačné pravidlá, ktoré sú jediné typu *zmena stavu*. Opäť všetky pravidlá formujúce komplexy sú v prílohe B. Všetky pravidlá sú ale reverzibilné a spolu s fosforylačnými vytvárajú schému na Obrázku 4.

Pravidlá týkajúce sa prepisu proteínov a ich degradácia sa nezmenili. Sú to pravidlá, ktoré do systému pridávajú, respektíve z neho odoberajú prvky.

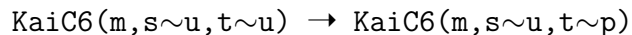
## 4.1 Porovnanie BCS s príslušnými modelmi

Po správnom návrhu biochemického priestoru by malo byť možné priradiť chemickým reakciám z modelov príslušné reakcie (pravidlá) z BCS. Tejto činnosti sa hovorí tzv. *mapovanie*. Zmyslom tejto činnosti je umožniť nájsť prepojenia medzi modelmi. Ak oba predstavujú inú časť BCS, potom mapovaním na BCS dôjde k objasneniu súvislostí medzi nimi.

V prípade modelov, ktoré predstavujú rovnakú časť BCS je cieľom nájsť súvis medzi nimi, v čom sa zhodujú a v čom rozchádzajú. To môže viesť k ďalším možnostiam využitia modelov, najmä k uvedomeniu si, čo daná časť modelu reprezentuje v biológii. Celý zmysel biochemického priestoru spočíva v tom, aby z matematických modelov, ktoré biológiu často výrazne abstrahujú, bol možný návrat opäť k procesu, ktorý má daný model opisovať. Význam BCS je teda dvojsečný: prináša súvislosti aj anotáciu.

### 4.1.1 Model Hertel et al.

V tomto prípade bol postup v podstate priamočiary. V prvom rade bolo dôležité si uvedomiť, aký typ abstrakcie model využíva. Najdôležitejším aspektom je to, že nie sú uvažované hexaméry (Sekcia 3.1), čo vedie k trochu inému vyloženiu reakcií. Napríklad táto reakcia:



má reprezentovať zmenu stavu na hexaméri z nefosforylovaného na fosforylovaný na treonínovom mieste. Na základe zistení v sekcii 3.1 v skutočnosti k formovaniu KaiC6 hexaméru nedochádza. Jedná sa len o špecifické vyjadrenie zmeny stavu monoméru, keď hovoríme, že je zastúpený v hexaméri. Po tomto uvedomení môžeme tvrdiť, že takisto ako

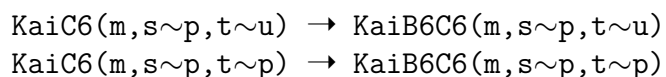
#### 4. NÁVRH SPOLOČNÉHO BIOCHEMICKÉHO PRIESTORU

táto, tak aj všetky obdobné komplexy sú vyjadrené abstraktne a je možné im priradiť príslušné pravidlá z BCS.

V modeli sú uvažované dva komplexy KaiBC a KaiC6, na ktorých prebieha fosforylačný cyklus. To znamená, že každý z ich monomérov môže meniť stav treonín (resp. serín) fosforylovaný (P) /nefosforylovaný (U). Na to slúžia v BCS štyri reverzibilné pravidlá, ktoré hovoria, že keď je monomér naviazaný (nezáleží na tom na čo), tak môže zmeniť treonín alebo serín stav. To vedie ku kombinatorickej explózii, ktorá vyústi k obsiahnutiu všetkých možností, ktoré hexamér alebo iný komplex môže nadobúdať. Vďaka tomu je možné každý z vyššie spomínaných komplexov popísať.

Formačné reakcie týchto dvoch komplexov sú po odmyslení si abstrakcie, ktorú model využíva, celkom jednoducho zapísateľné v BCS. Stačia na to dve pravidlá, podľa ktorých sa šesť ľubovoľných KaiC monomérov (t.j. nezáleží na stave) môže stretnúť a vytvoriť hexamér KaiC6. Podobne je to s druhým pravidlom, podľa ktorého šesť KaiB monomérov a jeden KaiC6 hexamér môžu vytvoriť komplex KaiBC. Obe tieto pravidlá sú reverzibilné, čím je zabezpečený rozpad komplexov.

Reakcie z modelu sa nedajú úplne jednoducho mapovať. Problémom je práve abstrakcia, ktorú používajú. Faktom je, že KaiB monoméry sa v modeli fyzicky vôbec nevyskytujú. Preto treba s ich neprítomnosťou narábať opatrne. V modeli KaiBC vzniká tzv. *exchange* reakciou. Znamená to, že komplex vznikne iba z KaiC6, KaiB sú prítomné implicitne. Tiež je dôležité to, že táto reakcia sa nemôže udiť v ktoromkoľvek stave, ale iba v dvoch prípadoch:



Na základe všetkých týchto faktov je možné mapovanie reakcií na vybrané reakcie z BCS.

Transkripčné reakcie ponúkajú tiež istú formu abstrakcie. Rozdiel modelu od BCS je v tom, že prepisom z *kaiBC\_mRNA* vzniká iba KaiC, pritom ale KaiC zdieľa gén spoločný s KaiB, vznikajú teda súčasne. Táto abstrakcia je vyvolaná tou predchádzajúcou, čiže KaiB nie je uvažované v premennom množstve, ale konštantnom a všadeprítomnom.

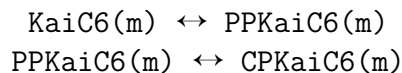
Degradačné reakcie boli ľahko mapovateľné, pre každý z komplexov či monomérov stačilo pravidlo, ktoré znamená degradáciu a nešpecifikuje stav, takže sa môže aplikovať na všetky možné stavy.

Po správnom pochopení reakcií v modeli a ich správnom uchopení bolo možné všetky reakcie jednoznačne namapovať na BCS. Kompletné mapovanie je rozpísané v prílohe B.

#### 4.1.2 Model Miyoshi et al.

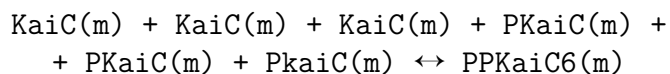
Model poskytuje odlišnú abstrakciu ako model Hertel et al. V prvom rade je to monomér KaiC, ktorý nemá dve fosforylačné miesta, ale iba jedno. Môže nadobúdať dva stavy KaiC (nefosforylovaný) a PKaiC (fosforylovaný). Pre hexamér KaiC6 sú následne uvažované iba tri rôzne stavy (Sekcia 2.3.2).

To pre mapovanie fosforylačných reakcií týkajúcich sa KaiC6 znamená generalizáciu, čiže na jednu reakciu z modelu sa mapuje viacero reakcií z BCS. Konkrétnym problémom je PKaiC, ktorý môže vyjadrovať fosforyláciu na treonín alebo serín mieste.



Tieto dve reverzibilné reakcie môžu predstavovať celkovo štyri reverzibilné reakcie z BCS, nakoľko PPKaiC6 môže predstavovať čiastočne serín alebo treonín fosforylovaný komplex.

Formačné reakcie sú obdobné. Opäť dochádza k tomu, že nie je jednoznačne určené, o aký typ fosforylácie sa jedná. Preto je nutné použiť ďalšiu generalizáciu, kde v reakcii



môžu byť tri PKaiC reprezentované dvojakým spôsobom.

Problémom boli reakcie, ktoré formovali tetramer KaiB4 a s tým súvisiace reakcie. Tento komplex v spolupráci s KaiA2 za vzniku KaiB4A2 má v modeli enzymatickú úlohu. Konkrétne KaiA2 pôsobí pozitívne na fosforyláciu a KaiB4A2 negatívne. V mojom doterajšom návrhu BCS sa takéto komplexy netvorila.

Tvorba KaiB4 teramerov bola načrtnutá ako vedľajší efekt pri bioluminiscencii [13], dokázaná bola v pomerne novej práci zaoberajúcou sa štruktúrou jednotlivých proteínov cirkadiálneho rytmu sinice [14]. Zo získaných informácií vyplýva, že KaiB monomér môže tvoriť dimer a dva dimery tetramer. Preto bol BCS obohatený o tvorbu tohto komplexu. Existencia KaiB4A2 podložená nie je. Pridanie pravidla formujúceho tento komplex je len na základe tohto modelu. Takisto je v modeli uvažovaná aj neaktívna forma KaiB4i, čo pre BCS znamenalo pridanie buď stavu pre KaiB(m $\leadsto$ a $\leadsto$ i) alebo pridanie látky KaiBi, čo vyplynulo ako jednoduchšie riešenie.

To znamená, že po týchto zmenách bol BCS kompletný a bolo možné domapovať potrebné reakcie. Čo sa týka degradačných a transkripčných reakcií, v tých sa problém nevyskytol a mapovanie bolo 1:1. Celé mapovanie je zhrnuté v prílohe B.

## 4.2 Porovnanie modelov

Na základe mapovania modelov na BCS a ujasnenia si významu všetkých častí, môže dôjsť k porovnaniu oboch modelov a vymedzeniu vzťahov medzi nimi. V prvom rade je potrebné si uvedomiť rozdiel v premenných, teda látkach používaných v modeloch. Zatiaľ čo model Hertel et al. (ďalej už len Hertel) umožňuje nadobúdanie stavov jednotlivých látok, v modeli Miyoshi et al. (ďalej už len Miyoshi) každý stav reprezentuje nová látka. Konkrétne, ak v Hertel chceme povedať, že nefosforylovaný KaiC6 sa fosforyluje na treonín mieste, tak to interpretujeme zmenou stavu:



Pre Miyoshi to ale znamená, že nefosforylovaný KaiC6 zanikne a vznikne nový, čiastočne fosforylovaný PPKaiC6:



Celkovo sa dá zhrnúť, že v Miyoshi je 13 látok, zatiaľ čo v Hertel ich je 15. Z toho vyplýva, že model Miyoshi je o niečo jednoduchší. Treba podotknúť, že je tu použitých niekoľko látok týkajúcich sa KaiB4

komplexu. Celkovo teda Hertel odráža biológiu presnejšie. Je to spôsobené vekom modelov, takmer sedemročný rozdiel posunul modelovanie bližšie k interpretácii reálnej biológie.

Ďalším rozdielom je využitie spätnej väzby na mRNA. Práve táto väzba prináša do systému oscilácie. Rozdiel je v tom, že kým v Miyoshim je vyjadrená explicitne, teda priamo v *reaction rate* transkripcie mRNA, u Hertel je to zahrnuté implicitne v reakčnej konštante.

Napriek výrazným rozdielom (Sekcie 3.2, 3.1) sa modely zhodujú v mnohých veciach. Je zachovaný oscilačný charakter ich simulácií. Ten je odlišný vo viacerých aspektoch. Prvým je kvantitatívny, ten nie je predmetom mojej štúdie. Z kvalitatívneho hľadiska je potrebné sa zamerať na kľúčové látky a vzájomne ich porovnávať. Zo všeobecného hľadiska sú totiž modely svojim chovaním veľmi podobné. Za rovnaký čas oba produkujú približne rovnaký počet lokálnych maxím, tzv. *peak* a tikajú s približne 24 hodinovou periódou.

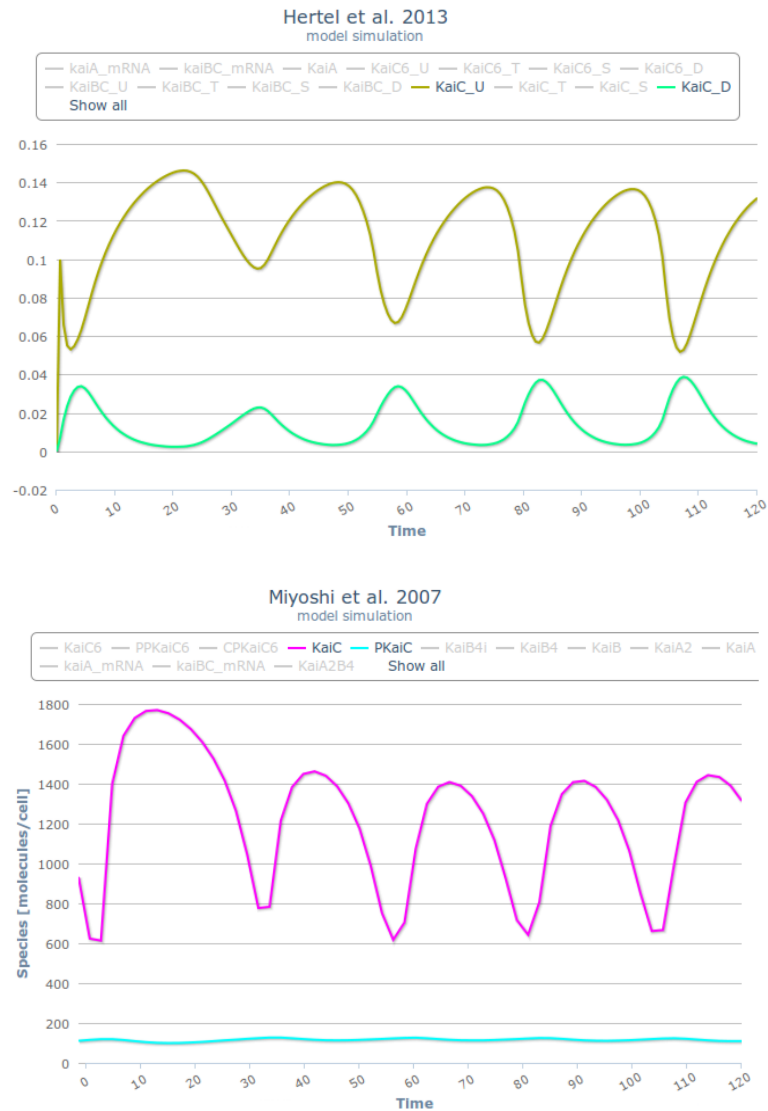
Pohľad z vnútra prináša množstvo rozdielov. Príklad rozdielu je na Obrázku 5. Zmeny koncentrácií monomérov KaiC sú v oboch modeloch odlišné.

Ďalším výrazným rozdielom je vzťah medzi nefosforylovaným a plne fosforylovaným KaiC6. V modeloch majú opačný charakter. V modeli Miyoshi je hladina fosforylovaného KaiC6 okrem iniciálneho kroku vždy vyššia. U Hertel je to naopak. Zaujímavosťou je tiež monomér KaiA. V modeli Hertel je totiž používaný len ako enzým, teda látka ovplyvňujúca *reaction rate* fosforylačných reakcií. Nie je priamo spotrebúvaný. Keďže kinetika translácie tejto látky je vyššia ako kinetika degradácie, celková koncentrácia KaiA konverguje k istej hodnote. V Miyoshim je to úplne inak. monomér je spotrebúvaný na tvorbu komplexu KaiA2B4 a z toho dôvodu koncentrácia osciluje.

Ostatné látky prinášajú v modeloch rozdiely, ktoré sú ťažké na porovnanie. Ide predovšetkým o látky týkajúce sa monoméru KaiB a jeho komplexov. Problémom je, že v Hertel nie je uvažované premenné množstvo tohto monoméru, no napriek tomu sa využíva na tvorbu komplexov KaiBC. Tie v modeli Miyoshi nie sú vôbec uvažované, napriek tomu, že KaiB má premenlivú koncentráciu. Navyše, v ňom vzniká komplex KaiB4, ktorý plní čisto enzymatickú úlohu.

Všetky vyššie spomenuté fakty znamenajú, že štruktúrne sú modely dosť odlišné, no napriek tomu podávajú veľmi podobný výsledok. Pri zanedbaní detailov si kvalitatívne odpovedajú. To vypovedá o mno-

#### 4. NÁVRH SPOLOČNÉHO BIOCHEMICKÉHO PRIESTORU



Obr. 5: Simulácia zmeny koncentrácií monoméru KaiC. Je viditeľný rozdielny oscilačný charakter. V modeli Hertel je koncentrácia fosforylovaného KaiC variabilnejšia. U nefosforylovaného je v modeloch opačný charakter zvyšovania a následného znižovania koncentrácie.

hých možnostiach abstrakcie. Dá sa povedať, že model Miyoshi je o niečo jednoduchší, čiže sa viac približuje minimálnemu modelu simulujúcemu cirkadiálne rytmy v sinici. Na druhej strane, používa viac abstrakcií, preto Hertel je bližší reálnej biológii.

## 5 Analýza biochemického priestoru

V rámci návrhu spoločného biochemického priestoru je možné vytvoriť hypotézy, ktoré by mal BCS spĺňať. Sú najmä štruktúrneho charakteru, hovoria o konzistencii celého priestoru. Typické sú otázky dosiahnuteľnosti stavov, ohraničenosti systému či reverzibility reakcií.

Vygenerovaný stavový priestor z BCS sa dá chápať a interpretovať ako orientovaný graf  $G = (S, R)$ , kde

- **S (States)** - neprázdna konečná množina reprezentujúca vrcholy grafu (stavy BCS)
- **R (Reactions)** - množina usporiadaných dvojíc typu  $(u, v)$  takých, že  $u \in S$ ,  $v \in S$  a  $u \neq v$ , reprezentujúca orientované hrany grafu (reakcie BCS)

Na takto vytvorený graf sa dá aplikovať overovanie modelov, tzv. *model checking*. K tomu slúži stromová výpočtová logika CTL [15], ktorou sa dajú vyjadriť požadované vlastnosti. Tie môžu byť pomocou vhodného nástroja overované.

### 5.1 Dostupnosť stavu

Dostupnosťou stavu sa rozumie určiť tzv. *izolované* látky:

- **Init** - množina všetkých látok, ktoré sú vstupom do systému, hrana vedie z nich, ale nie do nich. Tzv. *vstupný stupeň vrchola* je nulový. K tomu je potrebné vytvoriť obrátený graf  $G' = (S', R')$ , pre ktorý platí, že  $S = S'$  a všetky hrany z  $R'$  majú opačný smer ako v grafe  $G$  (sú invertované).

$$\forall (u, v) \in R : (v, u) \in R'$$

Potom nad vytvoreným grafom  $G'$  túto vlastnosť vyjadruje formula, kde **Init** je maximálna množina splňujúca danú vlastnosť:

$$\mathbf{Init} \subseteq S' \models AX (\text{false})$$

- **Sink** - množina všetkých látok, ktoré sú výstupom (produktom) systému, hrana vedie iba do nich. Tzv. *výstupný stupeň vrchola* je nulový. Množina Sink je maximálna množina splňujúca túto vlastnosť:

$$\text{Sink} \subseteq \mathbf{S} \models \text{AX} (\text{false})$$

- **Unreachable dead species (UDS)** - množina všetkých látok, z ktorých a do ktorých nevedie hrana. Také by sa v BCS vyskytovať nemali.

$$\text{UDS} = \text{Init} \cap \text{Sink}$$

N základe vyššie vytvorených množín a invertovaného grafu  $G'$  je možné overiť, či vybraná časť BCS tvorí jeden súvislý graf. Na to sú potrebné dve množiny stavov  $M$  a  $N$ , ktoré sú vždy maximálne množiny splňujúce danú vlastnosť:

1. je z nich dosiahnuteľná množina *Sink*

$$\mathbf{M} \subseteq \mathbf{S} \models \text{EF} (\text{Sink})$$

2. sú dosiahnuteľné z množiny *Init*

$$\mathbf{N} \subseteq \mathbf{S}' \models \text{EF} (\text{Init})$$

Ak je prienik  $M \cap N = \emptyset$ , potom sa jedná o oddelené grafy.

## 5.2 Reverzibilita

Reverzibilita prevedená na grafový problém znamená, že ak existuje cesta z jedného stavu do druhého, tak musí existovať aj cesta späť. Ťažko sa overuje, či sa jedná presne o tú istú cestu, na to by bolo potrebné overiť osobitne stav za stavom na danej ceste. Pre potreby BCS stačí overiť, či je cesta späť vôbec možná.

Nech  $\forall A, B \in \mathbf{S}; A \neq B$ . Takto definované makrá sa dajú v CTL logike interpretovať:

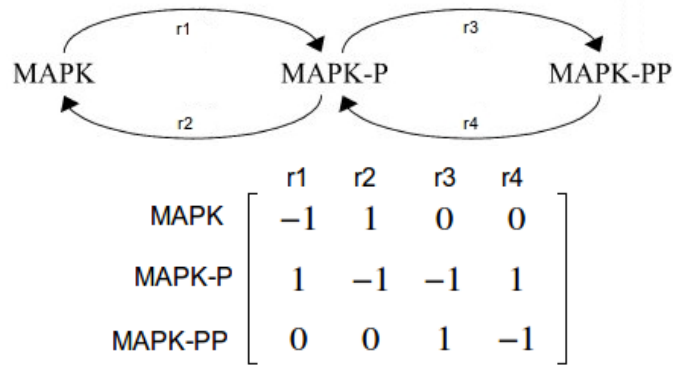


$$\mathbf{S} \models (\mathbf{A} \Rightarrow \text{EF}(\mathbf{B})) \wedge (\mathbf{B} \Rightarrow \text{EF}(\mathbf{A}))$$

čo znamená, že pre každé dve rôzne látky  $\mathbf{A}$ ,  $\mathbf{B}$  musí platiť, že ak je  $\mathbf{B}$  dosiahnuteľné z  $\mathbf{A}$ , potom je aj  $\mathbf{A}$  dosiahnuteľné z  $\mathbf{B}$ . Táto formula hovorí o globálnej reverzibilitate, teda odpovedá aj na otázku izolovaných látok. Ak je celý systém reverzibilný, znamená to, že neexistuje látka bez vstupnej aj výstupnej hrany.

Iný spôsob ako reprezentovať tento problém je pomocou incidenčnej matice [16]. Je to iná forma reprezentácie grafu. Stĺpce reprezentujú hrany (reakcie), riadky zase stavy (látky). Hodnoty sú priradené podľa toho, koľko látok sa v danej reakcii odoberá a koľko pridáva. Látkam, ktoré sa v reakcii nevyskytujú je priradená hodnota nula. Príklad takejto matice je na Obrázku 6.

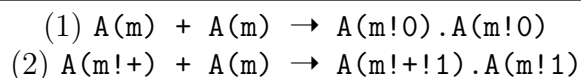
Nech matica incidencie BCS je  $\mathbf{M}$ , potom pre rovnicu  $\mathbf{M} \cdot \vec{x} = 0$  platí, že každé riešenie  $\vec{x}$  tejto lineárnej sústavy je *vyvážená* distribúcia reakčného toku (implikujúca rovnovážny stav charakterizovaný nulovou zmenou koncentrácií látok). Inak povedané, vektor  $\vec{x}$  reprezentuje reakcie systému. Každá reakcia vo vektore  $\vec{x}$  má buď hodnotu jedna, čiže tvorí súčasť reverzibilného pod systému, alebo hodnotu nula, čiže nepatrí do reverzibilného pod systému. Pre BCS by malo platiť, že existuje vektor  $\vec{x}$  tvorený samými jednotkami, pretože v systéme sa, ak neberieme v úvahu mRNA, nevyskytujú žiadne izolované látky (Sekcia 5.1).



Obr. 6: Matica incidencie pre časť kaskády MAPK [7].

### 5.3 Ohraničenosť

Tento problém čiastočne súvisí s vyššie spomenutými problémami. Ide o to rozhodnúť, či je BCS ohraničený, teda konečný. Takáto situácia by mohla nastať, keby sa v systéme nachádzala tzv. *reťazcovo-polymerizačná* reakcia (z jednotlivých molekúl vytvára dlhý reťazec). V modelovaní založenom na pravidlách by takáto reakcia bola interpretovaná dvoma pravidlami:



Pravidlo (1) je prípravným pre aplikáciu ďalšieho. Pravidlo (2) hovorí, že akákoľvek viazaná látka  $A$  môže tvoriť väzbu s ďalšou voľnou látkou  $A$ . Z toho vyplýva, že toto pravidlo bude tvoriť teoreticky nekonečný reťazec. Obmedzenie prichádza iba zo strany počtu látok v systéme. Pri budovaní stavového priestoru sa počiatočné podmienky neberú do úvahy, takže takýto systém by bol nekonečný.

V BCS sa žiadne pravidlo, ktoré by malo na ľavej strane obdobnú podmienku, nevyskytuje. Z tohto dôvodu je stavový priestor BCS konečný, teda ohraničený.

## 6 Záver

Pri formálnom prepise vybraných modelov do jazyka založeného na prepisovacích pravidlách bolo zistené, že sa môže objaviť mnoho problémov, ktoré túto činnosť spomalia alebo znemožnia. Najväčší vplyv má slabá anotácia modelov, nedostatočne vysvetlené časti, či priveľká miera abstrakcie. Napriek tomu sa tento prepis podaril pre oba vybrané modely. Manuálny prepis ponúka prácu do budúcnosti – softvér, ktorý by správne anotovaný model automaticky previedol do prepisovacích pravidiel.

Vďaka jasnému zápisu modelov došlo k vytvoreniu spoločného biochemického priestoru (BCS) v rovnakom jazyku, čo prinieslo mnoho výhod. V pôvodnej verzii bolo obtiažne porovnávanie a mapovanie modelov a zápis biochemického priestoru nebol presný. Zápis v prepisovacích pravidlách priniesol ľahké porovnávanie modelov s BCS a modelov vzájomne. Tým bolo možné jasne vymedziť vzťahy medzi nimi a nájsť miesta, kde sa zhodujú alebo naopak rozchádzajú. Táto časť pomôže k vylepšeniu zápisu BCS na e-cyanobacterium.org tým, že umožní jednoduchšie mapovanie modelov v systéme.

Na testovanie konzistencie tejto časti BCS bolo navrhnutých niekoľko metód, ktoré slúžia na automatické overovanie hypotéz a vlastností. K tomu je možné využiť mnoho prístupov, práca sa zameriava na vyjadrenie vlastností pomocou stromovej výpočtovej logiky CTL a na operácie pomocou lineárnej algebry. Nad rámec zadania boli realizované pokusy na automatické overovanie CTL formulí pomocou nástroja Biocham [17]. K tomu bolo najprv potrebné vytvoriť súbor vo formáte, ktorý vyhovuje Biochamu na vstupe. Na to som vytvoril jednoduchý syntaktický analyzátor v jazyku GNU Bash (príloha C), ktorý prevádza dáta do požadovaného formátu. Biocham mal problém s príliš veľkým objemom dát, ktorý nebol schopný spracovať.

Z tohto dôvodu bude v budúcnosti potrebné vytvoriť program, ktorý by dokázal efektívnejšie pracovať s veľkými objemami dát, pretože ako sa ukázalo, pri komplexnosti biochemických procesov je niečo také nevyhnutnosťou.

Nad rámec zadania boli využité dostupné softvéry (Charlie [18], Snoopy [19]) k vytvoreniu incidenčnej matice. Nevýhoda je, že spomenuté nástroje najprv potrebujú Petriho sieť [20], čo je pri danom objeme

dát výpočtovo a pamäťovo náročné. Preto bude v budúcnosti potrebné navrhnúť pamäťovo efektívnejšie riešenie, ktoré nepotrebuje explicitné vytváranie Petriho siete.

## Literatúra

- [1] S. Hertel, C. Brettschneider, I. M. Axmann, Revealing a two-loop transcriptional feedback mechanism in the cyanobacterial circadian clock, *PLoS Comput Biol* 9 (3) (2013) e1002966. doi:10.1371/journal.pcbi.1002966.
- [2] 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, *J Biol Rhythms* 22 (1) (2007) 69–80. doi:10.1177/0748730406295749.
- [3] J. Faeder, M. Blinov, W. Hlavacek, Rule-based modeling of biochemical systems with bionetgen., *Methods Mol Biol* 500 (2009) 113–67. doi:10.1007/978-1-59745-525-1\_5.
- [4] H. Kitano, Systems biology: toward system-level understanding of biological systems, *Foundations of systems biology* (2001) 19–54.
- [5] M. Klement, T. Děd, D. Šafránek, J. Červený, S. Mueller, R. Steuer, Biochemical space: A framework for systemic annotation of biological models, in: *Proceedings of the 5th International Workshop on Interactions between Computer Science and Biology (CS2Bio'14)*, Elsevier, Amsterdam, 2014, pp. 31–44. doi:10.1016/j.entcs.2014.06.013.
- [6] E. Dr. Klipp, R. Dr. Herwig, A. Dr. Kowald, C. Wierling, H. Prof. Dr. Lehrach, *Systems Biology in Practice: Concepts, Implementation and Application*, Wiley-VCH Verlag GmbH & Co. KGaA, 2005. doi:10.1002/3527603603.
- [7] B. N. Kholodenko, Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades, *European Journal of Biochemistry* 267 (2000) 1583–1588. doi:10.1046/j.1432-1327.2000.01197.x.
- [8] D. Bell-Pedersen, V. M. Cassone, D. J. Earnest, S. S. Golden, P. E. Hardin, T. L. Thomas, M. J. Zoran, Circadian rhythms from multiple oscillators: lessons from diverse organisms, *Nature Reviews Genetics* 6 (2005) 544–556. doi:10.1038/nrg1633.

- 
- [9] S. S. Golden, M. Ishiura, C. H. Johnson, T. Kondo, Cyanobacterial circadian rhythms, *Annual Review of Plant Physiology and Plant Molecular Biology* 48 (1997) 327–354. doi:10.1146/annurev.arplant.48.1.327.
- [10] M. Klement, D. Šafránek, T. Děd, A. Pejznoch, L. Nedbal, R. Steuer, J. Červený, S. Mueller, A comprehensive web-based platform for domain-specific biological models, in: *Proceedings of the fourth International Workshop on Interactions between Computer Science and Biology (CS2Bio'13)*, Elsevier, 2013, pp. 61–67. doi:10.1016/j.entcs.2013.11.006.
- [11] R. Fontes, J. Ribeiro, A. Sillero, Inhibition and activation of enzymes. The effect of a modifier on the reaction rate and on kinetic parameters, *Acta Biochim Pol.* 47 (1) (2000) 233–257.
- [12] J. W. Eaton, D. Bateman, S. Hauberg, GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations, CreateSpace Independent Publishing Platform, 2009, ISBN 1441413006.
- [13] Q. Xie, M. Soutto, X. Xu, Y. Zhang, C. Johnson, Bioluminescence resonance energy transfer (bret) imaging in plant seedlings and mammalian cells, *Methods Mol Biol* 680 (2011) 3–28. doi:10.1007/978-1-60761-901-7\_1.
- [14] J. Snijder, R. Burnley, A. Wiegard, A. Melquiond, A. Bonvin, I. Axmann, A. Heck, Insight into cyanobacterial circadian timing from structural details of the kaiB-kaiC interaction, *Proc Natl Acad Sci U S A.* 111 (4) (2014) 1379–1384. doi:10.1073/pnas.1314326111.
- [15] E. Emerson, J. Y. Halpern, Decision procedures and expressiveness in the temporal logic of branching time, *Journal of Computer and System Sciences* 30 (1) (1985) 1–24. doi:10.1016/0022-0000(85)90001-7.
- [16] R. Diestel, *Graph Theory*, Vol. 173, Graduate Texts in Mathematics, ISBN 978-3-642-14278-9, 2005.

- [17] F. Fages, S. Soliman, Formal cell biology in biocham, in: Formal Methods for Computational Systems Biology, Vol. 5016 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2008, pp. 54–80. doi:10.1007/978-3-540-68894-5\_3.
- [18] M. Heiner, M. Schwarick, J. Wegener, Charlie – an extensible Petri net analysis tool, in: Proc. PETRI NETS 2015, Vol. 9115, Springer, 2015, pp. 200–211. doi:10.1007/978-3-319-19488-2\_10.
- [19] M. Heiner, M. Herajy, F. Liu, C. Rohr, M. Schwarick, Snoopy – a unifying petri net tool, in: Application and Theory of Petri Nets, Vol. 7347, Springer Berlin Heidelberg, 2012, pp. 398–407. doi:10.1007/978-3-642-31131-4\_22.
- [20] C. A. Petri, W. Reisig, Petri net, Scholarpedia 3 (4) (2008) 6477. doi:10.4249/scholarpedia.6477.

# A Pôvodná verzia biochemického priestoru

Hertel	Biochemical Space	Miyoshi
	<b>KaiC6 phosphorylation</b>	
KaiC6(m,s~u,t~u) → KaiC6(m,s~p,t~u)	KaiC6(m,s~u,t~u) → KaiC6(m,s~p,t~u)	KaiC6(m) → PPKaiC6(m)
KaiC6(m,s~u,t~u) → KaiC6(m,s~u,t~p)	KaiC6(m,s~u,t~u) → KaiC6(m,s~u,t~p)	
KaiC6(m,s~p,t~u) → KaiC6(m,s~u,t~u)	KaiC6(m,s~p,t~u) → KaiC6(m,s~u,t~u)	PPKaiC6(m) → KaiC6(m)
KaiC6(m,s~u,t~p) → KaiC6(m,s~u,t~u)	KaiC6(m,s~u,t~p) → KaiC6(m,s~u,t~u)	
KaiC6(m,s~u,t~p) → KaiC6(m,s~p,t~p)	KaiC6(m,s~u,t~p) → KaiC6(m,s~p,t~p)	PPKaiC6(m) → CPKaiC6(m)
KaiC6(m,s~p,t~u) → KaiC6(m,s~p,t~p)	KaiC6(m,s~p,t~u) → KaiC6(m,s~p,t~p)	
KaiC6(m,s~p,t~p) → KaiC6(m,s~p,t~u)	KaiC6(m,s~p,t~p) → KaiC6(m,s~p,t~u)	CPKaiC6(m) → PPKaiC6(m)
KaiC6(m,s~p,t~p) → KaiC6(m,s~u,t~p)	KaiC6(m,s~p,t~p) → KaiC6(m,s~u,t~p)	
	<b>KaiBC66 phosphorylation</b>	
KaiB6C6(m,s~u,t~u) → KaiB6C6(m,s~p,t~u)	KaiB6C6(m,s~u,t~u) → KaiB6C6(m,s~p,t~u)	--
KaiB6C6(m,s~p,t~u) → KaiB6C6(m,s~u,t~u)	KaiB6C6(m,s~p,t~u) → KaiB6C6(m,s~u,t~u)	--
KaiB6C6(m,s~u,t~u) → KaiB6C6(m,s~u,t~p)	KaiB6C6(m,s~u,t~u) → KaiB6C6(m,s~u,t~p)	--
KaiB6C6(m,s~u,t~p) → KaiB6C6(m,s~u,t~u)	KaiB6C6(m,s~u,t~p) → KaiB6C6(m,s~u,t~u)	--
KaiB6C6(m,s~u,t~p) → KaiB6C6(m,s~p,t~p)	KaiB6C6(m,s~u,t~p) → KaiB6C6(m,s~p,t~p)	--
KaiB6C6(m,s~p,t~p) → KaiB6C6(m,s~u,t~p)	KaiB6C6(m,s~p,t~p) → KaiB6C6(m,s~u,t~p)	--
KaiB6C6(m,s~p,t~u) → KaiB6C6(m,s~p,t~p)	KaiB6C6(m,s~p,t~u) → KaiB6C6(m,s~p,t~p)	--
KaiB6C6(m,s~p,t~p) → KaiB6C6(m,s~p,t~u)	KaiB6C6(m,s~p,t~p) → KaiB6C6(m,s~p,t~u)	--
	<b>KaiC6 formation</b>	
KaiC(m,s~u,t~u) ↔ KaiC6(m,s~u,t~u)	6 KaiC(m,s~u,t~u) ↔ KaiC6(m,s~u,t~u)	KaiC(m) + KaiC(m) + KaiC(m) + KaiC(m) + KaiC(m) + KaiC(m) ↔ KaiC6(m)
KaiC(m,s~p,t~u) ↔ KaiC6(m,s~p,t~u)	6 KaiC(m,s~p,t~u) ↔ KaiC6(m,s~p,t~u)	KaiC(m) + KaiC(m) + KaiC(m) + PKaiC(m) + PKaiC(m) + PKaiC(m) ↔ PPKaiC6(m)
KaiC(m,s~u,t~p) ↔ KaiC6(m,s~u,t~p)	6 KaiC(m,s~u,t~p) ↔ KaiC6(m,s~u,t~p)	
KaiC(m,s~p,t~p) ↔ KaiC6(m,s~p,t~p)	6 KaiC(m,s~p,t~p) ↔ KaiC6(m,s~p,t~p)	PKaiC(m) + PKaiC(m) + PKaiC(m) + PKaiC(m) + PKaiC(m) + PKaiC(m) ↔ CPKaiC6(m)
	<b>KaiB6C6 formation</b>	
KaiB6C6(m,s~u,t~u) → KaiC(m,s~u,t~u)	6 KaiB + KaiC6(m,s~u,t~u) ↔ KaiB6_KaiC6(m,s~u,t~u)	--
KaiC6(m,s~p,t~u) → KaiB6C6(m,s~p,t~u)	6 KaiB + KaiC6(m,s~p,t~u) ↔ KaiB6_KaiC6(m,s~p,t~u)	--
KaiB6C6(m,s~p,t~u) → KaiC(m,s~p,t~u)	6 KaiB + KaiC6(m,s~p,t~u) ↔ KaiB6_KaiC6(m,s~p,t~u)	--
KaiB6C6(m,s~u,t~p) → KaiC(m,s~u,t~p)	6 KaiB + KaiC6(m,s~u,t~p) ↔ KaiB6_KaiC6(m,s~u,t~p)	--
KaiB6C6(m,s~p,t~p) → KaiC(m,s~p,t~p)	6 KaiB + KaiC6(m,s~p,t~p) ↔ KaiB6_KaiC6(m,s~p,t~p)	--
KaiC6(m,s~p,t~p) → KaiB6C6(m,s~p,t~p)	6 KaiB + KaiC6(m,s~p,t~p) ↔ KaiB6_KaiC6(m,s~p,t~p)	--
	<b>KaiA_mRNA</b>	
--	KaiA(m) + KaiA(m) ↔ KaiA2(m)	KaiA(m) + KaiA(m) ↔ KaiA2(m)
KaiA(m) → 0	KaiA(m) → 0	KaiA(m) → 0
0 → kaiA_mRNA(m)	0 → kaiA_mRNA(m)	0 → kaiA_mRNA(m)
0 → KaiA(m)	0 → KaiA(m)	0 → KaiA(m)
kaiA_mRNA(m) → 0	kaiA_mRNA(m) → 0	kaiA_mRNA(m) → 0
	<b>KaiBC_mRNA</b>	
0 → kaiBC_mRNA(m)	0 → kaiBC_mRNA(m)	0 → kaiBC_mRNA(m)
kaiBC_mRNA(m) → 0	kaiBC_mRNA(m) → 0	kaiBC_mRNA(m) → 0
0 → KaiC(m,s~u,t~u)	0 → KaiC(m,s~u,t~u) + KaiB(m)	0 → KaiB(m) + KaiC(m)
--	KaiB(m) → 0	KaiB(m) → 0
	<b>KaiC6 degradation</b>	
KaiC6(m,s~u,t~u) → 0	KaiC6(m,s~u,t~u) → 0	--
KaiC6(m,s~u,t~p) → 0	KaiC6(m,s~u,t~p) → 0	--
KaiC6(m,s~p,t~u) → 0	KaiC6(m,s~p,t~u) → 0	--
KaiC6(m,s~p,t~p) → 0	KaiC6(m,s~p,t~p) → 0	--
	<b>KaiB6C6 degradation</b>	
KaiB6C6(m,s~u,t~u) → 0	KaiB6C6(m,s~u,t~u) → 0	--
KaiB6C6(m,s~u,t~p) → 0	KaiB6C6(m,s~u,t~p) → 0	--
KaiB6C6(m,s~p,t~u) → 0	KaiB6C6(m,s~p,t~u) → 0	--
KaiB6C6(m,s~p,t~p) → 0	KaiB6C6(m,s~p,t~p) → 0	--
	<b>KaiC degradation</b>	
KaiC(m,s~u,t~u) → 0	KaiC(m,s~u,t~u) → 0	KaiC(m) → 0
KaiC(m,s~u,t~p) → 0	KaiC(m,s~u,t~p) → 0	
KaiC(m,s~p,t~u) → 0	KaiC(m,s~p,t~u) → 0	PKaiC(m) → 0
KaiC(m,s~p,t~p) → 0	KaiC(m,s~p,t~p) → 0	



## A. PÔVODNÁ VERZIA BIOCHEMICKÉHO PRIESTORU

	<b>KaiA2C6 formation</b>	
	$\text{KaiA2} + \text{KaiC6(m,s-u,t-u)} \leftrightarrow \text{KaiA2\_KaiC6(m,s-u,t-u)}$ $\text{KaiA2} + \text{KaiC6(m,s-u,t-p)} \leftrightarrow \text{KaiA2\_KaiC6(m,s-u,t-p)}$ $\text{KaiA2} + \text{KaiC6(m,s-p,t-u)} \leftrightarrow \text{KaiA2\_KaiC6(m,s-p,t-u)}$ $\text{KaiA2} + \text{KaiC6(m,s-p,t-p)} \leftrightarrow \text{KaiA2\_KaiC6(m,s-p,t-p)}$	
	<b>KaiA4C6 phosphorylation</b>	
	$\text{KaiA4\_KaiC6(m,s-u,t-u)} \leftrightarrow \text{KaiA4\_KaiC6(m,s-p,t-u)}$ $\text{KaiA4\_KaiC6(m,s-u,t-u)} \leftrightarrow \text{KaiA4\_KaiC6(m,s-u,t-p)}$ $\text{KaiA4\_KaiC6(m,s-p,t-u)} \leftrightarrow \text{KaiA4\_KaiC6(m,s-p,t-p)}$ $\text{KaiA4\_KaiC6(m,s-u,t-p)} \leftrightarrow \text{KaiA4\_KaiC6(m,s-p,t-p)}$	
	<b>KaiA4B6C6 phosphorylation</b>	
	$\text{KaiA4\_KaiB6\_KaiC6(m,s-u,t-u)} \leftrightarrow \text{KaiA4\_KaiB6\_KaiC6(m,s-u,t-p)}$ $\text{KaiA4\_KaiB6\_KaiC6(m,s-u,t-u)} \leftrightarrow \text{KaiA4\_KaiB6\_KaiC6(m,s-p,t-u)}$ $\text{KaiA4\_KaiB6\_KaiC6(m,s-p,t-u)} \leftrightarrow \text{KaiA4\_KaiB6\_KaiC6(m,s-p,t-p)}$ $\text{KaiA4\_KaiB6\_KaiC6(m,s-u,t-p)} \leftrightarrow \text{KaiA4\_KaiB6\_KaiC6(m,s-p,t-p)}$	
	<b>KaiA6B6C6 phosphorylation</b>	
	$\text{KaiA6\_KaiB6\_KaiC6(m,s-u,t-u)} \leftrightarrow \text{KaiA6\_KaiB6\_KaiC6(m,s-u,t-p)}$ $\text{KaiA6\_KaiB6\_KaiC6(m,s-u,t-u)} \leftrightarrow \text{KaiA6\_KaiB6\_KaiC6(m,s-p,t-u)}$ $\text{KaiA6\_KaiB6\_KaiC6(m,s-p,t-u)} \leftrightarrow \text{KaiA6\_KaiB6\_KaiC6(m,s-p,t-p)}$ $\text{KaiA6\_KaiB6\_KaiC6(m,s-u,t-p)} \leftrightarrow \text{KaiA6\_KaiB6\_KaiC6(m,s-p,t-p)}$	
	<b>KaiA4C6 formation</b>	
	$\text{KaiA2} + \text{KaiA2\_KaiC6(m,s-u,t-u)} \leftrightarrow \text{KaiA4\_KaiC6(m,s-u,t-u)}$ $\text{KaiA2} + \text{KaiA2\_KaiC6(m,s-p,t-u)} \leftrightarrow \text{KaiA4\_KaiC6(m,s-p,t-u)}$ $\text{KaiA2} + \text{KaiA2\_KaiC6(m,s-u,t-p)} \leftrightarrow \text{KaiA4\_KaiC6(m,s-u,t-p)}$ $\text{KaiA2} + \text{KaiA2\_KaiC6(m,s-p,t-p)} \leftrightarrow \text{KaiA4\_KaiC6(m,s-p,t-p)}$	
	<b>KaiA4B6C6 formation</b>	
	$\text{KaiA4\_KaiC6(m,s-u,t-u)} + 6 \text{ KaiB::cyt} \leftrightarrow \text{KaiA4\_KaiB6\_KaiC6(m,s-u,t-u)}$ $\text{KaiA4\_KaiC6(m,s-p,t-u)} + 6 \text{ KaiB::cyt} \leftrightarrow \text{KaiA4\_KaiB6\_KaiC6(m,s-p,t-u)}$ $\text{KaiA4\_KaiC6(m,s-u,t-p)} + 6 \text{ KaiB::cyt} \leftrightarrow \text{KaiA4\_KaiB6\_KaiC6(m,s-u,t-p)}$ $\text{KaiA4\_KaiC6(m,s-p,t-p)} + 6 \text{ KaiB::cyt} \leftrightarrow \text{KaiA4\_KaiB6\_KaiC6(m,s-p,t-p)}$	
	<b>KaiA6B6C6 formation KaiA2</b>	
	$\text{KaiA4\_KaiB6\_KaiC6(m,s-u,t-u)} + \text{KaiA2} \leftrightarrow \text{KaiA6\_KaiB6\_KaiC6(m,s-u,t-u)}$ $\text{KaiA4\_KaiB6\_KaiC6(m,s-u,t-p)} + \text{KaiA2} \leftrightarrow \text{KaiA6\_KaiB6\_KaiC6(m,s-u,t-p)}$ $\text{KaiA4\_KaiB6\_KaiC6(m,s-p,t-u)} + \text{KaiA2} \leftrightarrow \text{KaiA6\_KaiB6\_KaiC6(m,s-p,t-u)}$ $\text{KaiA4\_KaiB6\_KaiC6(m,s-p,t-p)} + \text{KaiA2} \leftrightarrow \text{KaiA6\_KaiB6\_KaiC6(m,s-p,t-p)}$	
		$\text{KaiB4i(m)} \leftrightarrow \text{KaiB4(m)}$
		$\text{KaiB4(m)} \leftrightarrow \text{KaiB(m)} + \text{KaiB(m)} + \text{KaiB(m)}$
		$\text{KaiB4(m)} + \text{KaiA2(m)} \leftrightarrow \text{KaiA2B4(m)}$

## B Biochemický priestor v BNGL

Hertel	Biochemical Space	Miyoshi
	<b>KaiC6 phosphorylation</b>	
KaiC6(m,s-u,t-u) → KaiC6(m,s-p,t-u)	KaiC(ml+,s-u,t-u) ↔ KaiC(ml+,s-u,t-p)	KaiC6(m) → PPKaiC6(m)
KaiC6(m,s-u,t-u) → KaiC6(m,s-u,t-p)		
KaiC6(m,s-p,t-u) → KaiC6(m,s-u,t-u)	KaiC(ml+,s-u,t-p) ↔ KaiC(ml+,s-p,t-p)	PPKaiC6(m) → KaiC6(m)
KaiC6(m,s-u,t-p) → KaiC6(m,s-u,t-u)		
KaiC6(m,s-u,t-p) → KaiC6(m,s-p,t-p)	KaiC(ml+,s-p,t-p) ↔ KaiC(ml+,s-p,t-u)	PPKaiC6(m) → CPKaiC6(m)
KaiC6(m,s-p,t-u) → KaiC6(m,s-p,t-p)		
KaiC6(m,s-p,t-p) → KaiC6(m,s-p,t-u)	KaiC(ml+,s-p,t-u) ↔ KaiC(ml+,s-u,t-u)	CPKaiC6(m) → PPKaiC6(m)
KaiC6(m,s-p,t-p) → KaiC6(m,s-u,t-p)		
	<b>KaiB6C6 phosphorylation</b>	
KaiB6C6(m,s-u,t-u) → KaiB6C6(m,s-p,t-u)	KaiC(ml+,s-u,t-u) ↔ KaiC(ml+,s-u,t-p)	--
KaiB6C6(m,s-p,t-u) → KaiB6C6(m,s-u,t-u)		--
KaiB6C6(m,s-u,t-u) → KaiB6C6(m,s-u,t-p)	KaiC(ml+,s-u,t-p) ↔ KaiC(ml+,s-p,t-p)	--
KaiB6C6(m,s-u,t-p) → KaiB6C6(m,s-u,t-u)		--
KaiB6C6(m,s-u,t-p) → KaiB6C6(m,s-p,t-p)	KaiC(ml+,s-p,t-p) ↔ KaiC(ml+,s-p,t-u)	--
KaiB6C6(m,s-p,t-p) → KaiB6C6(m,s-u,t-p)		--
KaiB6C6(m,s-p,t-u) → KaiB6C6(m,s-p,t-p)	KaiC(ml+,s-p,t-u) ↔ KaiC(ml+,s-u,t-u)	--
KaiB6C6(m,s-p,t-p) → KaiB6C6(m,s-p,t-u)		--
	<b>KaiC6 formation</b>	
KaiC(m,s-u,t-u) ↔ KaiC6(m,s-u,t-u)		KaiC(m) + KaiC(m) + KaiC(m) + KaiC(m) + KaiC(m) + KaiC(m) ↔ KaiC6(m)
KaiC(m,s-p,t-u) ↔ KaiC6(m,s-p,t-u)	KaiC(m) + KaiC(m) + KaiC(m) + KaiC(m) + KaiC(m) + KaiC(m) ↔ KaiC(m).KaiC(m).KaiC(m).KaiC(m).KaiC(m).KaiC(m)	KaiC(m) + KaiC(m) + KaiC(m) + PkaiC(m) + PkaiC(m) + PkaiC(m) ↔ PPKaiC6(m)
KaiC(m,s-u,t-p) ↔ KaiC6(m,s-u,t-p)		
KaiC(m,s-p,t-p) ↔ KaiC6(m,s-p,t-p)		PKaiC(m) + PKaiC(m) + PKaiC(m) + PKaiC(m) + PKaiC(m) + PKaiC(m) ↔ CPKaiC6(m)
	<b>KaiB6C6 formation</b>	
KaiB6C6(m,s-u,t-u) → KaiC(m,s-u,t-u)		--
KaiB6C6(m,s-p,t-u) → KaiC(m,s-p,t-u)		--
KaiB6C6(m,s-u,t-p) → KaiC(m,s-u,t-p)	KaiB(m) + KaiB(m) + KaiB(m) + KaiB(m) + KaiB(m) + KaiB(m) + KaiC(m).KaiC(m).KaiC(m).KaiC(m).KaiC(m).KaiC(m)	--
KaiB6C6(m,s-p,t-p) → KaiC(m,s-p,t-p)	KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m)	--
KaiC6(m,s-p,t-u) → KaiB6C6(m,s-p,t-u)		--
KaiC6(m,s-p,t-p) → KaiB6C6(m,s-p,t-p)		--
	<b>KaiA mRNA</b>	
--	KaiA(m) + KaiA(m) ↔ KaiA2(m)	KaiA(m) + KaiA(m) ↔ KaiA2(m)
KaiA(m) → 0	KaiA(m) → 0	KaiA(m) → 0
0 → kaiA_mRNA(m)	0 → kaiA_mRNA(m)	0 → kaiA_mRNA(m)
0 → KaiA(m)	0 → KaiA(m)	0 → KaiA(m)
kaiA_mRNA(m) → 0	kaiA_mRNA(m) → 0	kaiA_mRNA(m) → 0
	<b>KaiBC_mRNA</b>	
0 → kaiBC_mRNA(m)	0 → kaiBC_mRNA(m)	0 → kaiBC_mRNA(m)
kaiBC_mRNA(m) → 0	kaiBC_mRNA(m) → 0	kaiBC_mRNA(m) → 0
0 → KaiC(m,s-u,t-u)	0 → KaiC(m,s-u,t-u) + KaiB(m)	0 → KaiB(m) + KaiC(m)
--	KaiB(m) → 0	KaiB(m) → 0
	<b>KaiC6 degradation</b>	
KaiC6(m,s-u,t-u) → 0		--
KaiC6(m,s-u,t-p) → 0		--
KaiC6(m,s-p,t-u) → 0	KaiC(m).KaiC(m).KaiC(m).KaiC(m).KaiC(m).KaiC(m) → 0	--
KaiC6(m,s-p,t-p) → 0		--
	<b>KaiB6C6 degradation</b>	
KaiB6C6(m,s-u,t-u) → 0		--
KaiB6C6(m,s-u,t-p) → 0	KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m)	--
KaiB6C6(m,s-p,t-u) → 0		--
KaiB6C6(m,s-p,t-p) → 0		--
	<b>KaiC degradation</b>	
KaiC(m,s-u,t-u) → 0		KaiC(m) → 0
KaiC(m,s-u,t-p) → 0		
KaiC(m,s-p,t-u) → 0	KaiC(m) → 0	
KaiC(m,s-p,t-p) → 0		PKaiC(m) → 0

## B. BIOCHEMICKÝ PRIESTOR V BNGL

	<b>KaiA2C6 formation</b>	
	$\text{KaiC(m).KaiC(m).KaiC(m).KaiC(m).KaiC(m).KaiC(m)} + \text{KaiA2(m,n)} \leftrightarrow \text{KaiC(m).KaiA2(m,n).KaiC(m).KaiC(m).KaiC(m).KaiC(m).KaiC(m)}$	
	<b>KaiA4C6 phosphorylation</b>	
	$\text{KaiC(ml+,s-u,t-u)} \leftrightarrow \text{KaiC(ml+,s-u,t-p)}$	
	$\text{KaiC(ml+,s-u,t-p)} \leftrightarrow \text{KaiC(ml+,s-p,t-p)}$	
	$\text{KaiC(ml+,s-p,t-p)} \leftrightarrow \text{KaiC(ml+,s-p,t-u)}$	
	$\text{KaiC(ml+,s-p,t-u)} \leftrightarrow \text{KaiC(ml+,s-u,t-u)}$	
	<b>KaiA4B6C6 phosphorylation</b>	
	$\text{KaiC(ml+,s-u,t-u)} \leftrightarrow \text{KaiC(ml+,s-u,t-p)}$	
	$\text{KaiC(ml+,s-u,t-p)} \leftrightarrow \text{KaiC(ml+,s-p,t-p)}$	
	$\text{KaiC(ml+,s-p,t-p)} \leftrightarrow \text{KaiC(ml+,s-p,t-u)}$	
	$\text{KaiC(ml+,s-p,t-u)} \leftrightarrow \text{KaiC(ml+,s-u,t-u)}$	
	<b>Kai64B6C6 phosphorylation</b>	
	$\text{KaiC(ml+,s-u,t-u)} \leftrightarrow \text{KaiC(ml+,s-u,t-p)}$	
	$\text{KaiC(ml+,s-u,t-p)} \leftrightarrow \text{KaiC(ml+,s-p,t-p)}$	
	$\text{KaiC(ml+,s-p,t-p)} \leftrightarrow \text{KaiC(ml+,s-p,t-u)}$	
	$\text{KaiC(ml+,s-p,t-u)} \leftrightarrow \text{KaiC(ml+,s-u,t-u)}$	
	<b>KaiA4C6 formation</b>	
	$\text{KaiC(m).KaiA2(m,n).KaiC(m).KaiC(m).KaiC(m).KaiC(m).KaiC(m)} + \text{KaiA2(m,n)} \leftrightarrow \text{KaiC(m).KaiA2(m,n).KaiC(m).KaiC(m).KaiA2(m,n).KaiC(m).KaiC(m).KaiC(m)}$	
	<b>KaiA4B6C6 formation</b>	
	$\text{KaiC(m).KaiA2(m,n).KaiC(m).KaiC(m).KaiA2(m,n).KaiC(m).KaiC(m).KaiC(m)} + \text{KaiB(m)} + \text{KaiB(m)} + \text{KaiB(m)} + \text{KaiB(m)} + \text{KaiB(m)} + \text{KaiB(m)} \leftrightarrow \text{KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m)}$	
	<b>KaiA6B6C6 formation KaiA2</b>	
	$\text{KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m)} + \text{KaiA2(m)} \leftrightarrow \text{KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m)}$	
	<b>KaiA6B6C6 formation 3*KaiA2</b>	
	$\text{KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiB(m).KaiC(m)} + \text{KaiA2(m)} + \text{KaiA2(m)} + \text{KaiA2(m)} \leftrightarrow \text{KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m).KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m).KaiA2(m,n).KaiB(m).KaiC(m)}$	
	<b>KaiB2 formation</b>	
	$\text{KaiB(m)} + \text{KaiB(m)} \leftrightarrow \text{KaiB(m).KaiB(m)}$	
	<b>KaiB4 formation</b>	
	$\text{KaiB(m).KaiB(m)} + \text{KaiB(m).KaiB(m)} \leftrightarrow \text{KaiB(m).KaiB(m).KaiB(m).KaiB(m)}$	$\text{KaiB4(m)} \leftrightarrow \text{KaiB(m)} + \text{KaiB(m)} + \text{KaiB(m)} + \text{KaiB(m)}$
	<b>KaiB4 inactivation</b>	
	$\text{KaiB(m).KaiB(m).KaiB(m).KaiB(m)} \leftrightarrow \text{KaiBi(m).KaiBi(m).KaiBi(m).KaiBi(m)}$	$\text{KaiB4i(m)} \leftrightarrow \text{KaiB4(m)}$
	<b>KaiA2B4 formation</b>	
	$\text{KaiB(m).KaiB(m).KaiB(m).KaiB(m)} + \text{KaiA2(m)} \leftrightarrow \text{KaiA2(m).KaiB(m).KaiB(m).KaiB(m).KaiB(m)}$	$\text{KaiB4(m)} + \text{KaiA2(m)} \leftrightarrow \text{KaiA2B4(m)}$

## C Zoznam elektronických príloh

- táto práca vo formáte PDF,
- LaTeX zdrojový kód tejto práce, citačný zdroj a použité obrázky,
- modely a biochemický priestor zapísané v jazyku BNGL,
- porovnanie pôvodnej verzie biochemického priestoru s modelmi,
- porovnanie biochemického priestoru v jazyku BNGL s modelmi,
- syntaktický analyzátor v jazyku GNU Bash,
- prevodný skript v jazyku GNU Bash do formátu Biocham.