sybila

Rule-based Modelling of Biochemical Processes

Specification and Analysis

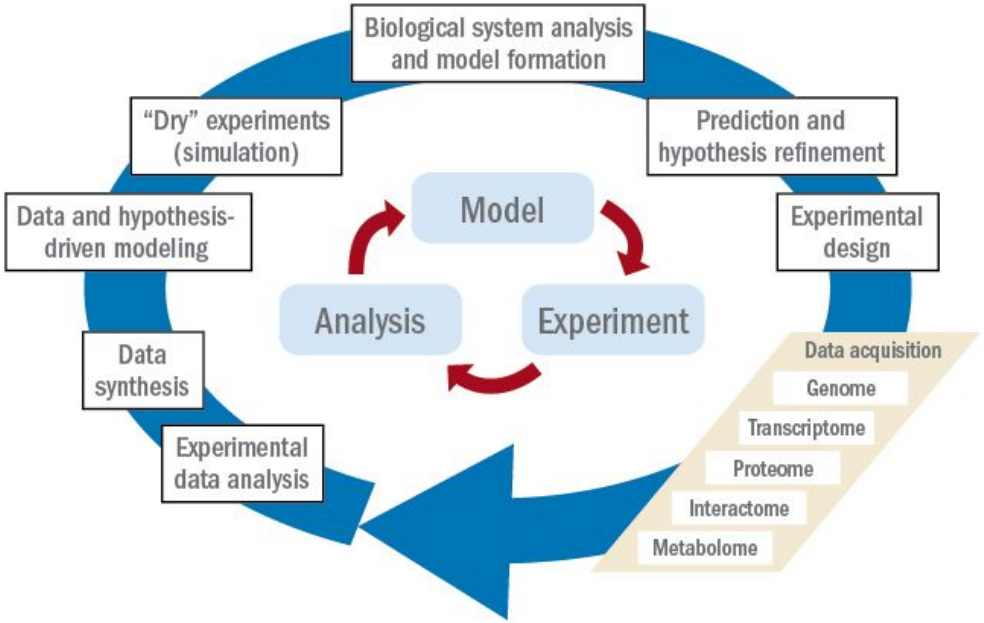
Matej Troják

PhD thesis defense

13.6.2023

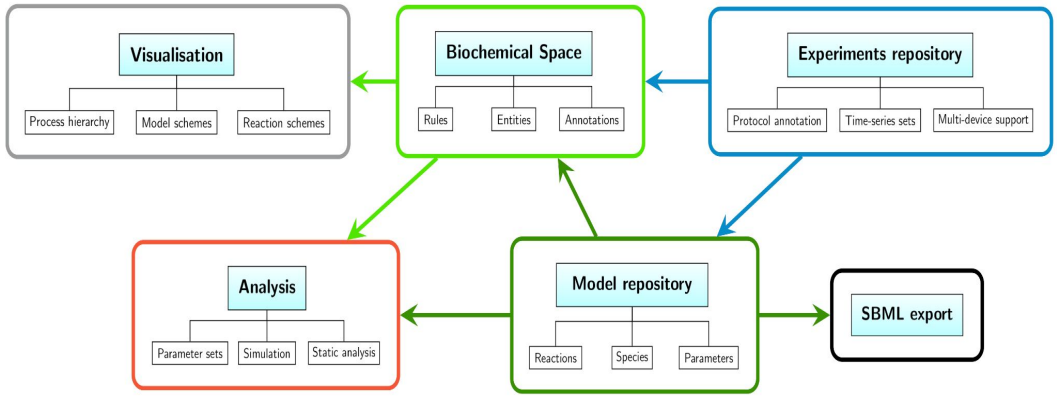
Modelling in biology

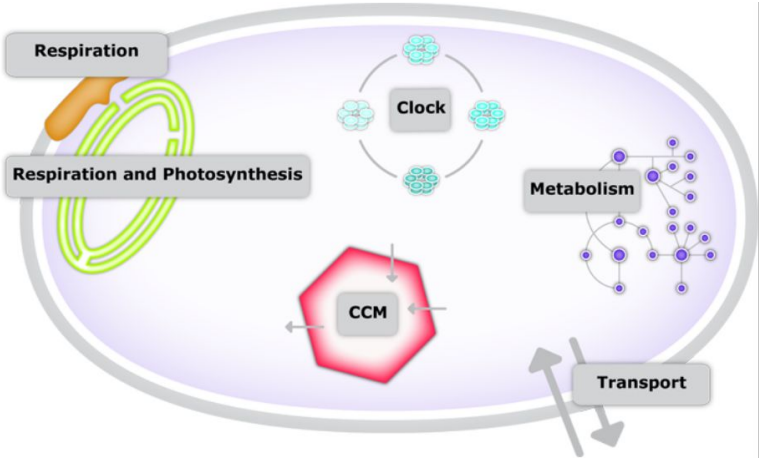
- Models can help us to decrease the number of required experiments to confirm a biological phenomenon and also to understand the investigated system
- Research done on both the experimental and modelling sides, but getting all information together is a challenging and complex process



Comprehensive modelling platform (CMP)

- Web-based platform for public sharing, annotation, analysis, and visualisation of dynamical models and wet-lab experiments
- Instances: e-cyanobacterium.org, e-photosynthesis.org
- Central role - **Biochemical Space** as a knowledge base



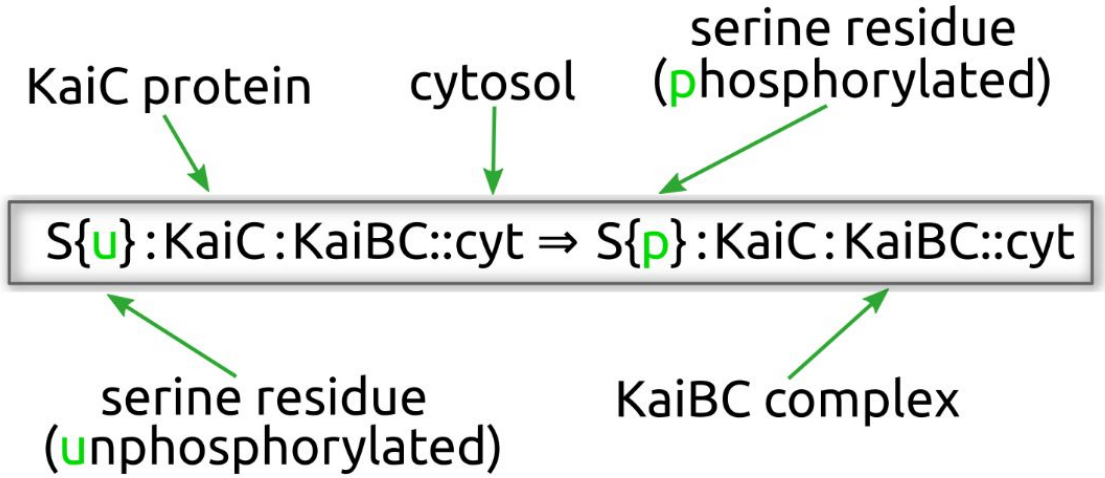


E-Cyanobacterium.org: A Web-Based Platform for Systems Biology of Cyanobacteria, CMSB 2016

BioChemical Space Language (BCSL)

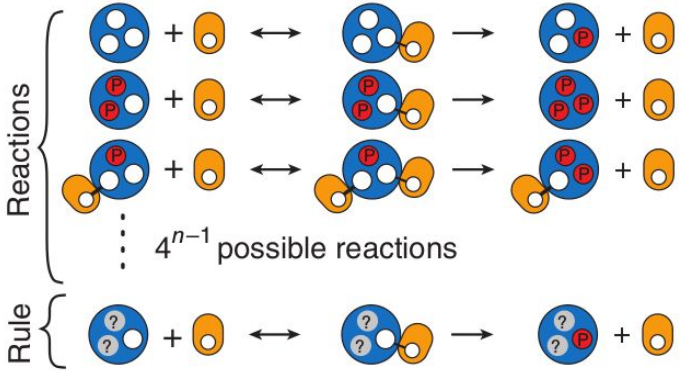
- Originally informal notation established by biologists
- A high-level **rule-based** language suitable for describing biochemical interactions on a detailed level in a concise form
- Key features: human-readability, complexes with “coexistence” meaning, support for compartments

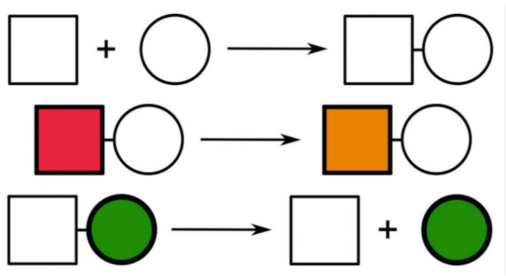
```
#! rules
r1_S ~ P(S{i})::cell => P(S{a})::cell
r1_T ~ P(T{i})::cell => P(T{a})::cell
r2 ~ P()::cell => P()::out

#! inits
1 P(S{i},T{i})::cell
```

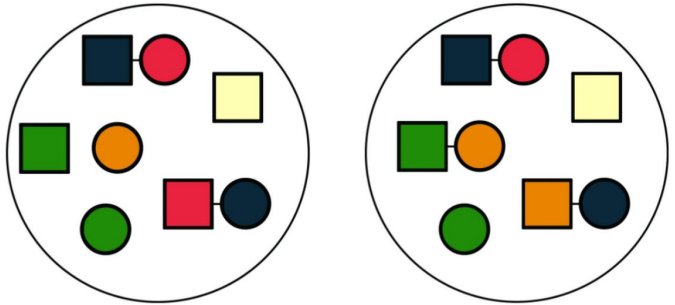


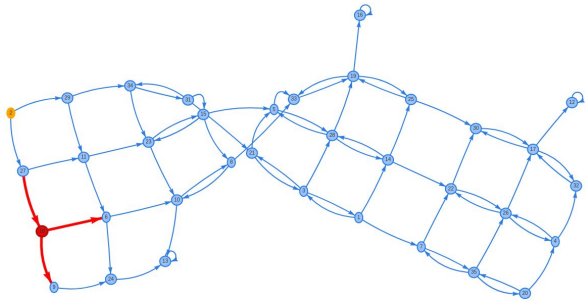
Rule-based modelling

- An abstraction to describe multiple reactions using a single pattern
- Typically allows us to change the properties of molecules (colours) and connect molecules to groups (binding)
- “Don’t care, don’t write” policy









The first definition

- Definition via translating BCSL to other languages (Kappa and BNGL)
- Indirect rewriting semantics (allowed to translate and then simulate models)
- Presented the first case studies from e-cyanobacterium.org (circadian clock, photosynthesis, metabolism)

Algorithm 1. Transform a BCS model $\mathtt{M}$ to a *kappa*$_s$ model $\mathcal{M}$.

function toKappa($\mathtt{M} = ((\Sigma_\tau, \Sigma_x), \mathtt{R})$)
  $\Sigma_\kappa, I, \mathcal{R}, \mathtt{A} := \emptyset$ # *global kappa*$_s$ *signature, rules and agent names*
  for all $r \in \mathtt{R}$ **do**
    for all $\Gamma \in \{\Gamma_l, \Gamma_r\}$ such that $r = \Gamma_l \Rightarrow \Gamma_r$ **do** # *for both rule sides*
      $\mathtt{E} := \emptyset$
      for all $\rho\ \epsilon :: \mathrm{c} \in \Gamma$ **do** # *repeat* ρ*-times*
        if ϵ has the form a{s} **then**
          $\mathtt{E} \leftarrow$ translateAtom($\epsilon :: \mathrm{c}$)
        if ϵ has the form $\tau(\gamma_p)$ **then**
          $\mathtt{E} \leftarrow$ translateStructure($\epsilon :: \mathrm{c}$)
        if ϵ has the form X **then**
          $\mathtt{E} :=$ translateComplex($\epsilon :: \mathrm{c}, \mathtt{E}$)
    construct a *kappa*$_s$ rule r_κ from the two resulting sets $\mathtt{E}$ obtained for Γ_l, Γ_r
    $\mathcal{R} \leftarrow r_\kappa$ # *extend kappa*$_s$ *rules*
  return $\mathcal{M} = (\Sigma_\kappa, I, \mathcal{R})$

RULE ID:	PSII oxidation
RULE EQUATION:	$ps2(oec\{3+\} \mid yz\{+\}) :: tlm \Leftrightarrow ps2(oec\{4+\} \mid yz\{n\}) :: tlm$
MODIFIER:	
RULE NAME:	oxidation from S3 to S4 of oxygen evolving complex
CLASSIFICATION:	oxidation
DESCRIPTION:	Oxidation occurring on photosystem II. Electron is transferred from oxygen evolving complex *oec* to active tyrosine *yz*.

Formal Biochemical Space with Semantics in Kappa and BNGL, SASB 2015

Direct definition

- Encoding of complexes was not suitable (too strict)
- Rule-based potential in semantics limited by the translation
  - assign various mathematical meanings to the model
- “Direct” definition based on low-level mathematical objects (multisets)
  - allowed us to be more flexible
  - define multiple model variants with their semantics (with/without kinetics, parametrised),
  - separated core and syntactic extensions - react to language changes and extensions
- Static analysis
  - the notion of *compatibility* (specificity comparison)
  - detect redundant rules, context reductions, non-reachability analysis

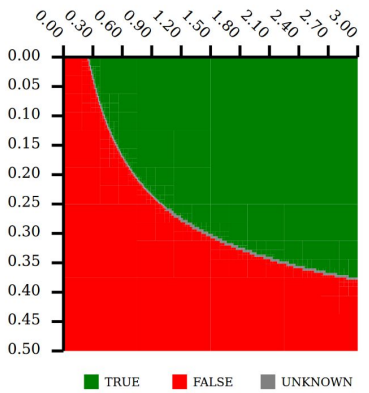
Executable Biochemical Space for Specification and Analysis of Biochemical Systems, SASB 2018 & Plos ONE 2020

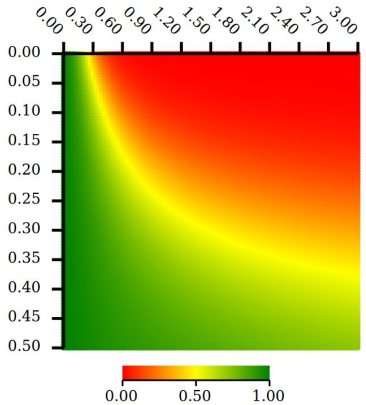
Quantitative semantics

- Dynamics of chemical reactions needed to be reflected
- Extension by adding *rate expressions* to the rules
  - rate can be interpreted in several ways, typically assigning a speed or a probability

T()::cyt => T()::ext @ k1 x [T()::cyt]

- This consequently means possible unknown parameters and the need to investigate them
- PCTL model checking and parameter synthesis
  - performed on the underlying DTMC using Storm/PRISM tool (normalisation of evaluated rates)



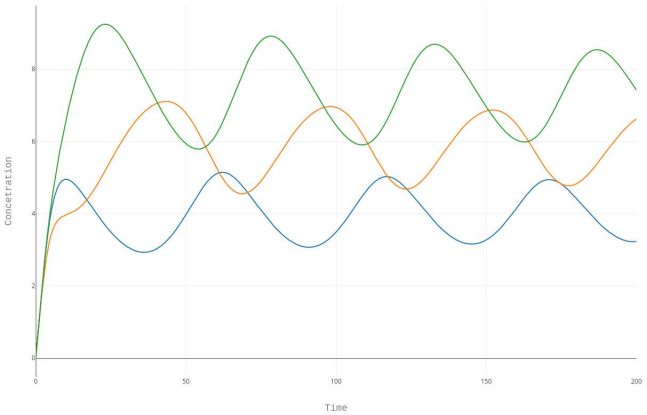


property P >= 0.5 [F T()::cyt > 8]

Parameter Synthesis and Robustness Analysis of Rule-Based Models, NFM 2020

Software tool

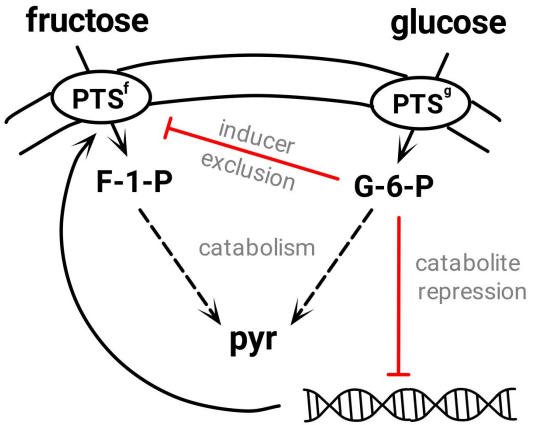
- Tool eBCSgen developed in parallel with the language
- Originally a desktop tool, then extracted to a command line tool
- First attempt to wrap within Galaxy environment
  - simple pipelining of tools and sharing results
- Supported features
  - interactive editor
  - simulations (deterministic and stochastic)
  - PCTL model checking and parameter synthesis
  - static analysis methods
  - visualisations
    - simulation
    - transition system exploration
    - satisfiability of parameter space exploration
- Tutorial, documentation



eBCSgen: A Software Tool for Biochemical Space Language, CMSB 2020

Regulation mechanisms

- Limited/sketched knowledge about modelled system's parts
  - unknown details on qualitative level
  - e.g. case study of carbon transport in E. Coli
- Regulations
  - additional conditions on the rewriting process
  - allow to capture uncertain details on the abstract level
  - several introduced types
    - Regular
    - Ordered
    - Programmed
    - Conditional
    - Concurrent-free
- Theoretical results
  - extend the expressive power of the language
  - comparison of individual classes of regulated systems



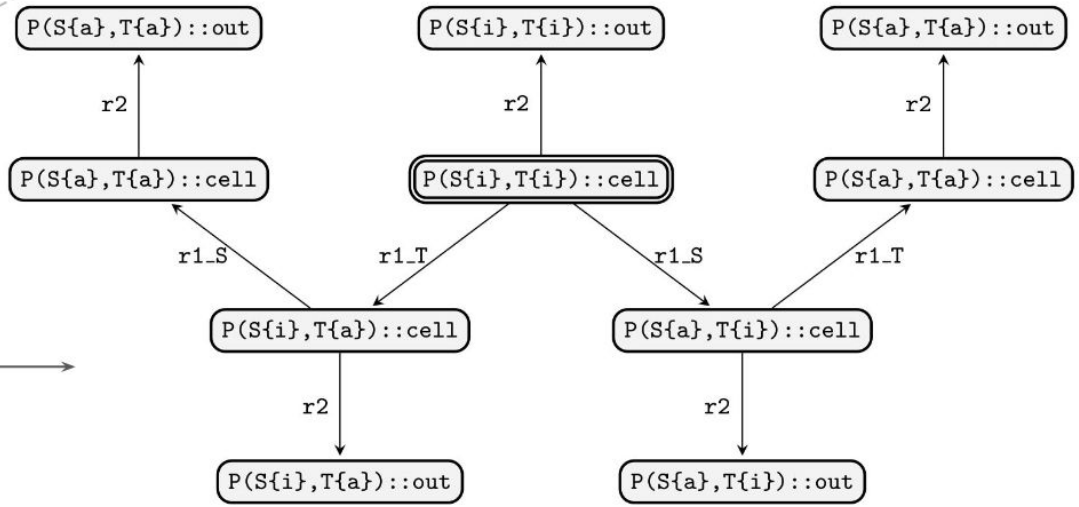
Rule-based Modelling of Biological Systems Using Regulated Rewriting, BioSystems 2023

Regulation effects

```
#! rules
r1_S ~ P(S{i})::cell ⇒ P(S{a})::cell
r1_T ~ P(T{i})::cell ⇒ P(T{a})::cell
r2   ~ P()::cell     ⇒ P()::out

#! inits
1 P(S{i},T{i})::cell
```

Model

Transition system



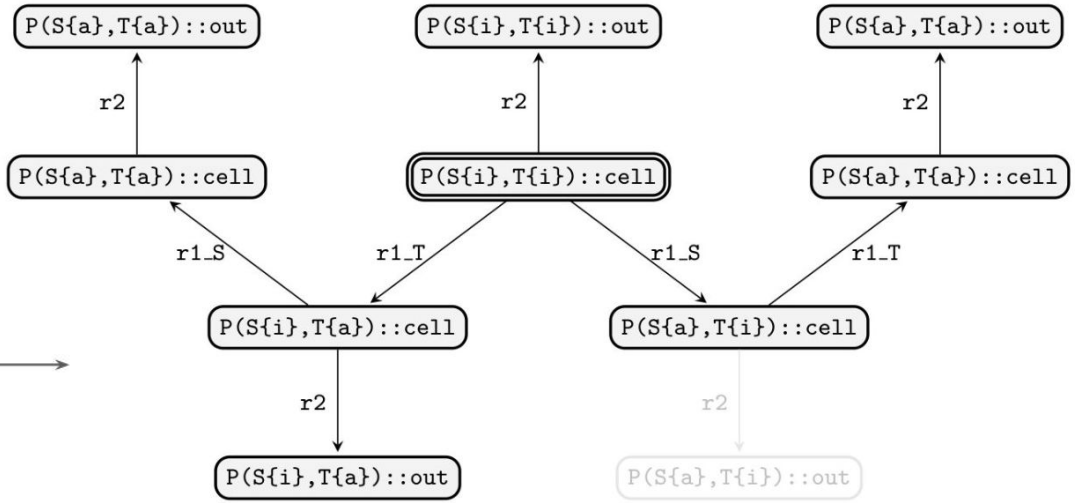
Regulation effects

+ Conditional regulation
r2 -> { P(S{a},T{i})::cell }

```
#! rules
r1_S ~ P(S{i})::cell ⇒ P(S{a})::cell
r1_T ~ P(T{i})::cell ⇒ P(T{a})::cell
r2   ~ P()::cell     ⇒ P()::out

#! inits
1 P(S{i},T{i})::cell
```

Model

Transition system



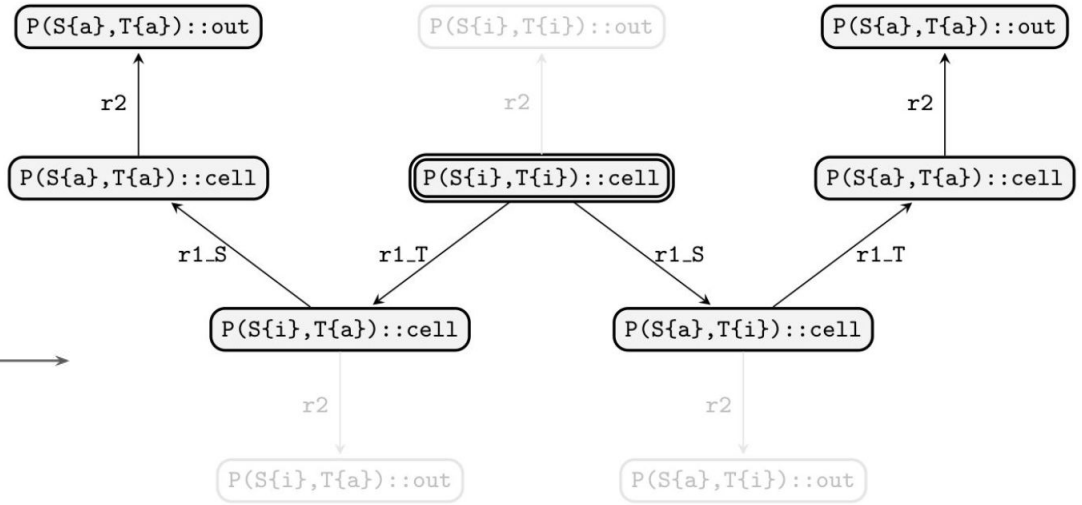
Regulation effects

+ Concurrent-free regulation
  r1_S > r2
  r1_T > r2

```
#! rules
r1_S ~ P(S{i})::cell ⇒ P(S{a})::cell
r1_T ~ P(T{i})::cell ⇒ P(T{a})::cell
r2   ~ P()::cell     ⇒ P()::out

#! inits
1 P(S{i},T{i})::cell
```

Model

Transition system

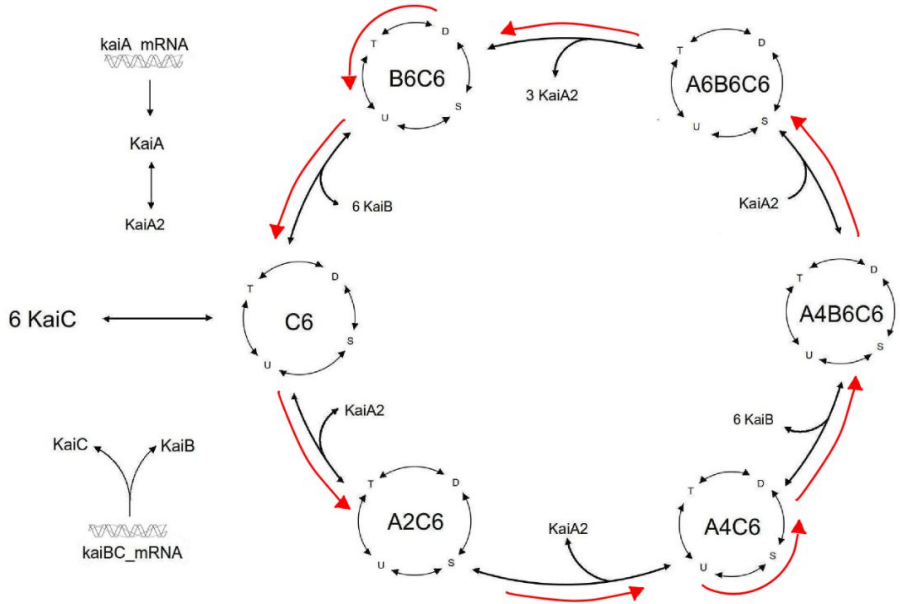


Fine-tuning the software tool

- Distribution
  - automatic bioconda package and docker container on release
  - properly shared Galaxy wrappers via official channels (including some of the visualisations)
- New features
  - support for CTL model checking
  - SBML-multi standard export
    - definition of the relation
    - discussions with the standardisation community
  - support for all defined regulations

Case studies - circadian clock

- The simplest known circadian oscillator (cyanobacteria)
- 3 molecules interacting in 10 rules, forming complexes and changing states of their phosphorylation sites
- Used regular regulation to capture a typical behavioural scenario




```
rpS ~ S{u}:KaiC():?::cyt ⇔ S{p}:KaiC():?::cyt ;
    ? = {KaiC6,KaiA2C6,KaiB6C6,KaiA4C6,KaiA4B6C6,KaiA6B6C6}
```

Case studies - tumour growth

- Division of tumour cells, four rules with parametrised rates
- Investigation of unknown parameters for unlimited division, specified by the

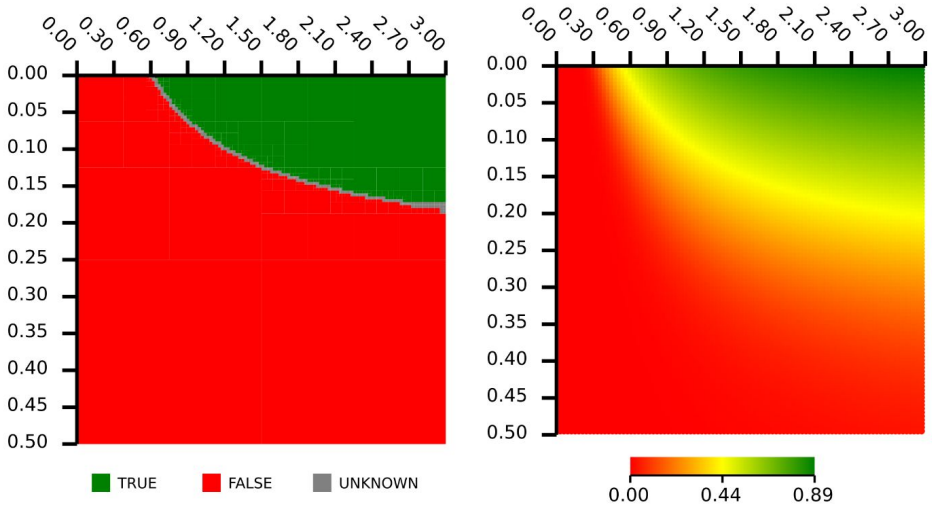
$$\phi = \Pi_{\geqslant 0.5}(\mathbf{F}\ \mathsf{T}() > 8)$$

$$\mathsf{T}(\mathsf{P}\{\mathsf{i}\}) \Rightarrow \mathsf{T}(\mathsf{P}\{\mathsf{m}\})\ @\ a_1 \times [\mathsf{T}(\mathsf{P}\{\mathsf{i}\})]$$

$$\mathsf{T}(\mathsf{P}\{\mathsf{m}\}) \Rightarrow 2\ \mathsf{T}(\mathsf{P}\{\mathsf{i}\})\ @\ a_2 \times [\mathsf{T}(\mathsf{P}\{\mathsf{m}\})]$$

$$\mathsf{T}(\mathsf{P}\{\mathsf{i}\}) \Rightarrow @\ d_1 \times [\mathsf{T}(\mathsf{P}\{\mathsf{i}\})]$$

$$\mathsf{T}(\mathsf{P}\{\mathsf{m}\}) \Rightarrow @\ d_2 \times [\mathsf{T}(\mathsf{P}\{\mathsf{m}\})]$$



Case studies - methane combustion

- A model of (methane) combustion where the role of regulation is crucial in capturing the targeted phenomena
- Complete combustion is preferred before incomplete incomplete (without any quantitative information)
  - i.e. we need to assign priority to the rule **complete** over the rule **incomplete**

```
#! rules
complete   ~ 2 CH4() + 4 O2() ⇒ 2 CO2() + 4 H2O()
incomplete ~ 2 CH4() + 3 O2() ⇒ 2 CO()  + 4 H2O()

#! regulation
type concurrent-free
(complete, incomplete)
```

Future work

- Syntactic usability improvements (e.g. robust rational expressions)
- SBML-multi relationship and export
- Finding more users and applications (FGFR pathway case study)
- Use-case based development of new features (e.g. replication in cell cycle)
- Versioning of the language itself (separate the interpreter core from the tool)
- Efficient (symbolic) representation and computation (challenge)

Summary of publications

- **Formal Biochemical Space with Semantics in Kappa and BNGL**. SASB 2015
  T. Děd, D. Šafránek, M. Troják, M. Klement, J. Šalagovič, L. Brim.
- **E-Cyanobacterium.org: A Web-Based Platform for Systems Biology of Cyanobacteria**. CMSB (Rank B) 2016
  M. Troják, D. Šafránek, J. Hrabec, J. Šalagovič, F. Romanovská, J. Červený.
- **Executable Biochemical Space for Specification and Analysis of Biochemical Systems**. SASB 2018
  M. Troják, D. Šafránek, L. Brim, J. Šalagovič, J. Červený. (technical report)
- **Fully Automated Attractor Analysis of Cyanobacteria Models**. ICSTCC (Rank B) 2018
  N. Beneš, L. Brim, J. Červený, S. Pastva, D. Šafránek, J. Šalagovič, M. Troják. (best paper award)
- **Executable Biochemical Space for Specification and Analysis of Biochemical Systems**. Plos ONE (IF 3.24) 2020
  M. Troják, D. Šafránek, L. Mertová, L. Brim.
- **eBCSgen: A Software Tool for Biochemical Space Language**. CMSB (Rank B) 2020
  M. Troják, D. Šafránek, L. Mertová, L. Brim.
- **Parameter Synthesis and Robustness Analysis of Rule-Based Models**. NFM (rank B) 2020
  M. Troják, D. Šafránek, L. Mertová, L. Brim.
- **MSMetaEnhancer: A Python Package for Mass Spectra Metadata Annotation**. JOSS 2022
  M. Troják, H. Hecht, M, Čech, E. J. Price.
- **eBCSgen 2.0: Modelling and Analysis of Regulated Rule-Based Systems**. CMSB (rank B) 2022
  M. Troják, D. Šafránek, B. Brozmann, L. Brim.
- **Rule-based Modelling of Biological Systems Using Regulated Rewriting**. BioSystems (IF 1.96) 2023
  M. Troják, D. Šafránek, S. Pastva, L. Brim.

Conferences and mentoring

Talks

- SASB - 2017 (New York), 2018 (Freiburg)
- COMBINE - 2017 (Milan)
- HSB - 2020 (online)
- NFM - 2020 (online)
- GCC - 2022 (Minneapolis)
- CMSB - 2022 (Bucharest)

Posters

- MEMICS - 2016 (Telč), 2017 (Telč)
- ISMB/ECCB - 2017 (Prague)
- Biomania - 2017 (Brno), 2022 (Brno)
- CMSB - 2018 (Brno), 2019 (Trieste)
- HSB - 2019 (Prague)
- COMBINE - 2022 (Berlin)

Teaching & mentoring

- seminar tutor of Foundations of programming (2018 and 2019)
- 4x supervisor of bachelor thesis
- 14x reader of bachelor/master thesis

Artifact evaluation committee

- CMSB - 2022, 2023
- QEST - 2021, 2022

Summary

- **BioChemical Space language** - a novel rule-based language suitable for representing objects and processes in biology on a detailed level
- **eBCSgen** - a tool to support modelling in BCSL and analysis of models
- Regulations - an extension to capture uncertain or abstract knowledge about the modelled system
- Case studies of BCSL application to a variety of biological phenomena as well as application of analysis techniques and regulations

Thank you for your attention!

Q&A of reviewers

1. The unusual treatment of complexes is noted as a distinction of BCSL. Did any of the case studies benefit from its looser conception of complexes, compared to other languages?

The circadian clock of cyanobacteria

2. [this is intended as a light-hearted question] Galaxy is introduced as a means to reach an intended audience but who is that audience? For example, if I run a biology lab with busy experimental scientists and expensive equipment to quantify molecules, should I try to find time to learn eBCSGen so I can adapt and run a model in Galaxy? Or should I hire you, to enjoy some years working in Edinburgh and use the native python software?

Q&A of reviewers

- Chapter 1 "Preliminaries": on page 12, you introduce a special empty rule $\epsilon = (\emptyset, \emptyset)$ in the aim of getting infinite sequences. Is that made to ease subsequent treatments, for instance model-checking, or does that correspond to biological phenomena?

Just a technical aspect, introduced primarily for regulations, as these often need a notion of "skipping" a rule. The epsilon rules were also used to ensure infiniteness to simplify proofs of some theoretical results.

- Chapter 2 "State of the art": many work are cited but not Bio-Pepa (developed at the University of Edinburgh). Could you comment on it and relate it to your work?

Bio-Pepa has syntax more distant from plain chemical reactions (more algebraical), no rule-based features, allows more general rate functions.

[CH08] Federica Ciocchetta and Jane Hillston. "Bio-PEPA: an Extension of the Process Algebra PEPA for Biochemical Networks". In: *Electronic Notes in Theoretical Computer Science* 194.3 (2008), pp. 103–117.

Q&A of reviewers

- Chapter 3 “Biochemical Space Language”: what about considering the rules as atomic actions and including them as the basic constructions of “classical” process algebras, such as CCS?

Definitely possible. We originally chose similar approach (translation to Kappa/BNGL) and later used MRS as basis for regulations. However, we tried to avoid any higher-level algebras/languages in the definition to maintain freedom in the BCSL development.

- Chapter 4 “Regulations”: regulations are defined in a static way and thus apply to all the runs. Would it be interesting to change regulations dynamically and would that correspond to biological phenomena?

One of the possible future direction of regulations research (such as combinations of regulations). For example, changing different modes of behaviour (“hybrid” system) could be useful (e.g. change of environment conditions).

Q&A of reviewers

- Chapter 5 "Analysis techniques": you introduce interesting analysis techniques without discussing the feasibility of implementing them efficiently. Could you comment on that?

  Not the primary focus of the research. Static analysis straightforward, others done on explicit representation (except PCTL analysis via Storm).

- Chapter 6 "Software tool": could you comment on the use of the tool by biologists?

  Already covered.

- Chapter "Conclusion": can you comment on future work?

  Already covered.

Timeline

- Research at **Systems biology laboratory** (Sybila), Faculty of Informatics, Masaryk University
  - with David Šafránek and Luboš Brim
- Collaboration with Global Change Research Institute of the Czech Academy of Sciences (CzechGlobe)
  - with Jan Červený
- Employment at Research Centre for Toxic Compounds in the Environment (RECETOX)
  - with Helge Hecht and Elliott James Price