Formal Biochemical Space with Semantics in Kappa and BNGL

David Šafránek

with T. Děd, M. Troják, M. Klement,
J. Šalagovič, L. Brim

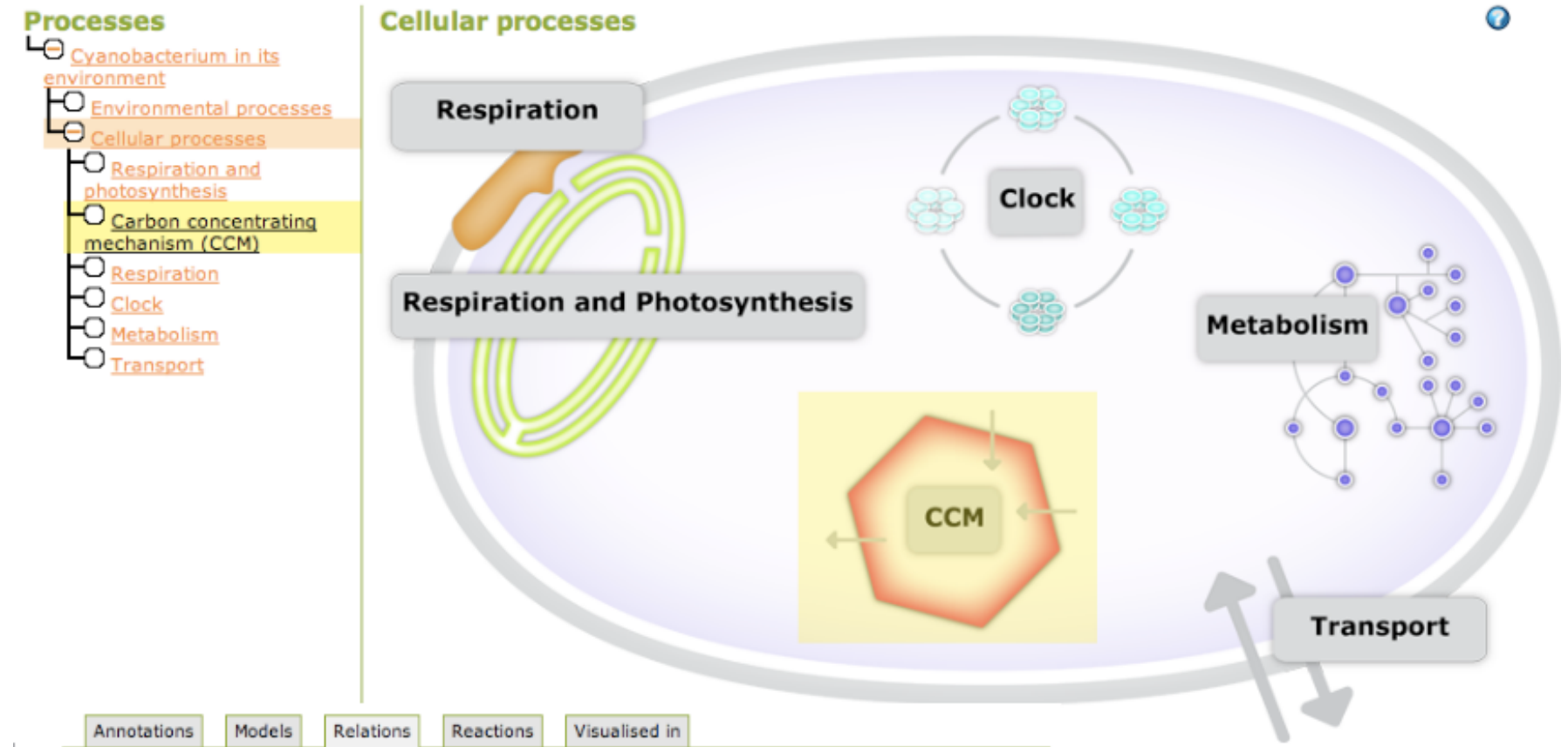
Systems Biology Laboratory
Masaryk University Brno

SASB 2015, Saint-Malo, France

8 September 2015

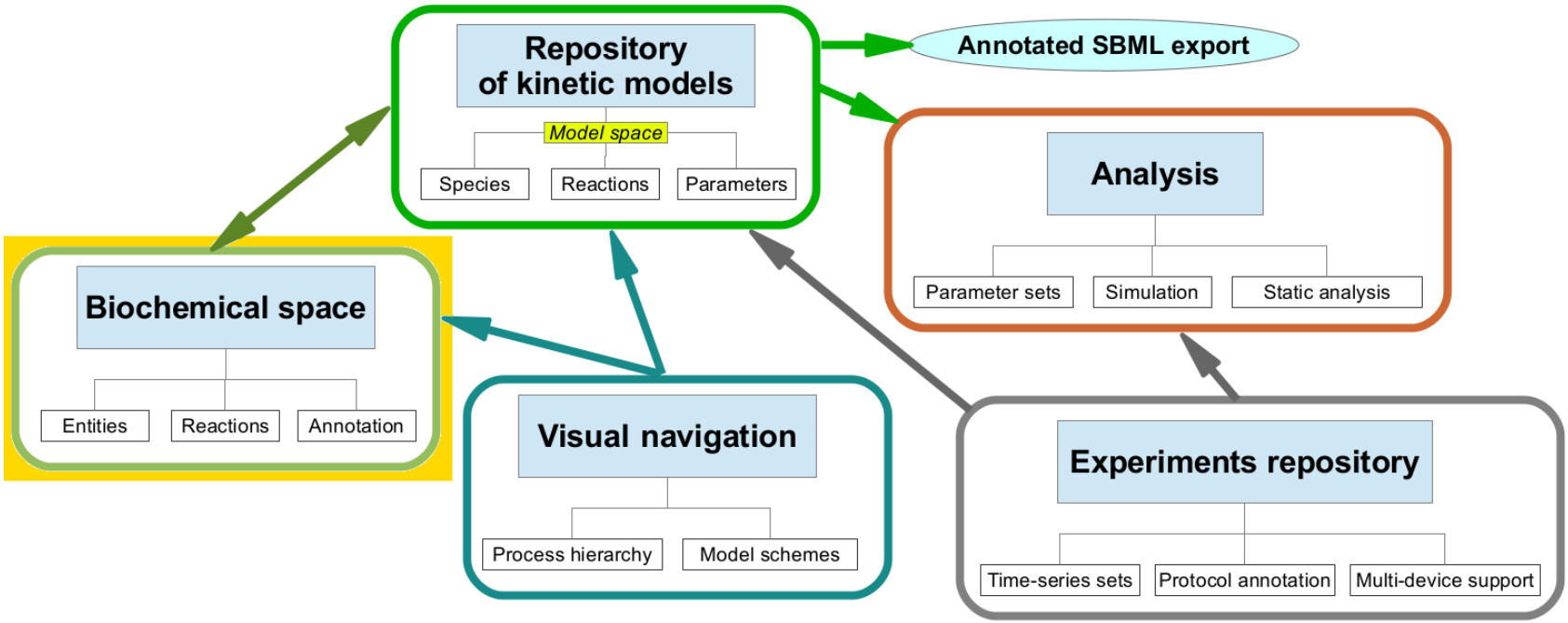
Background and Motivation

Comprehensive Modelling Platform (CMP)



Background and Motivation

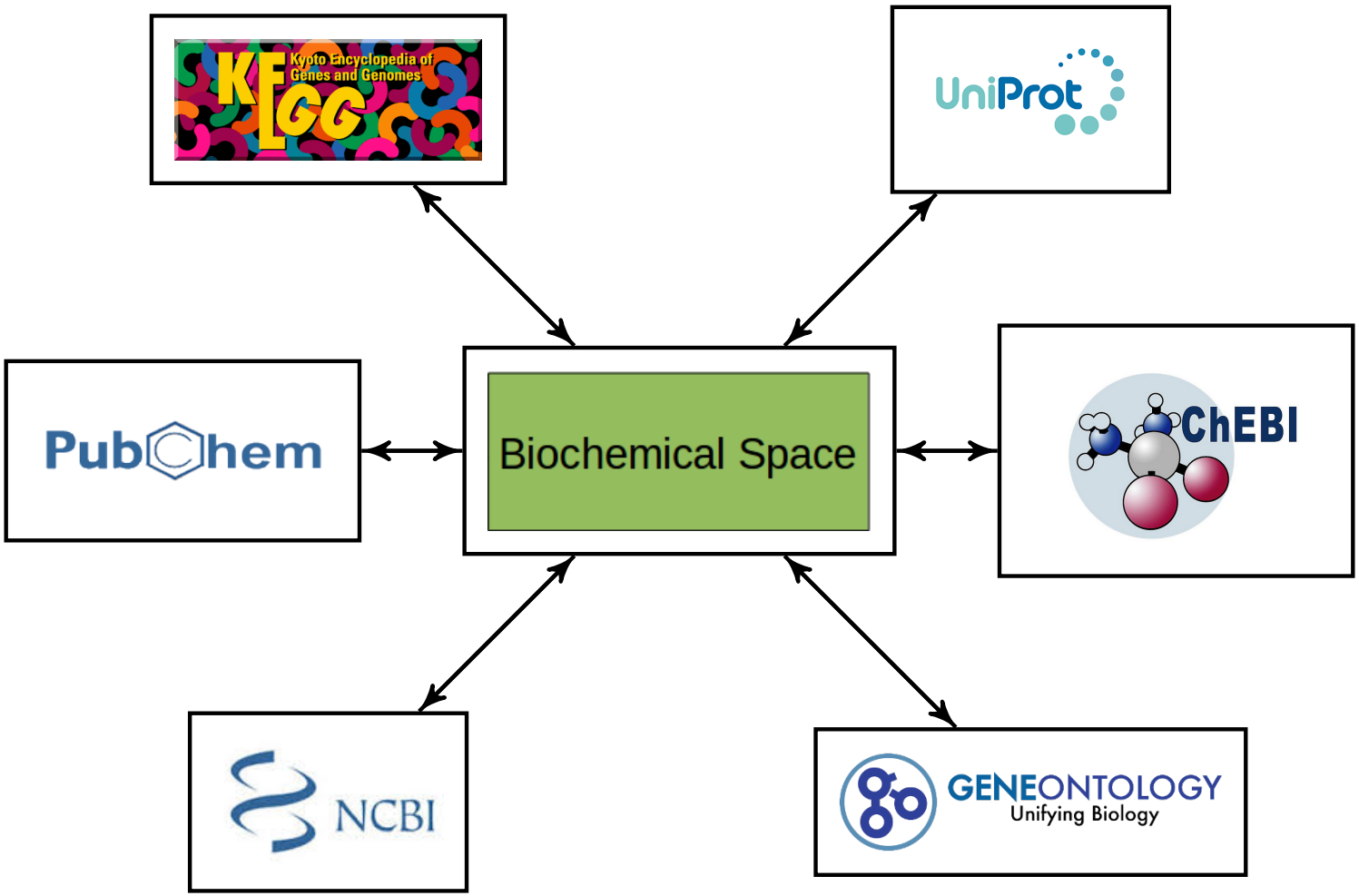
Biochemical Space in the context of CMP



- need for easy-to-understand but yet formal description of biological processes
- tackle the complexity - combinatorial explosion

Annotation Standards

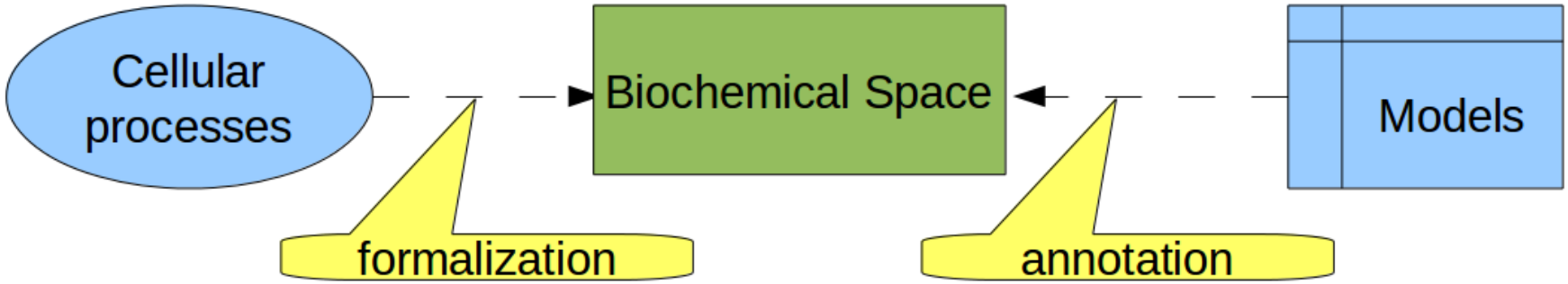
Connection to Bioinformatics



sybila
systems biology laboratory

Biochemical Space

Eliminating the Gap Between Biology and Math



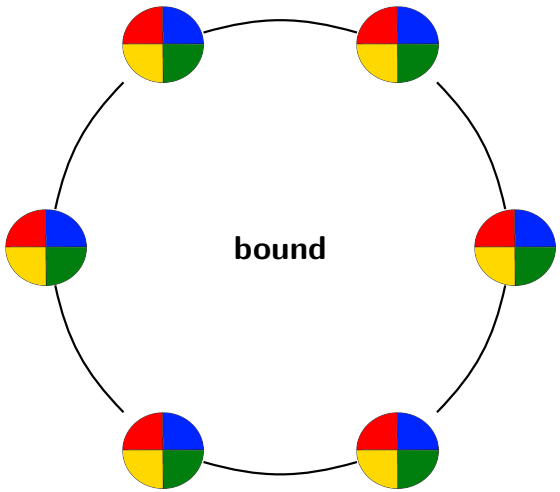
- biophysics employs a lot of indirect approximation in models
- combining rule-based approach with reaction-based approach allows compact mechanistic description
- mapping the models to such a description allows better and faster understanding

sybila systems biology laboratory

Abstraction from structural details

Biology

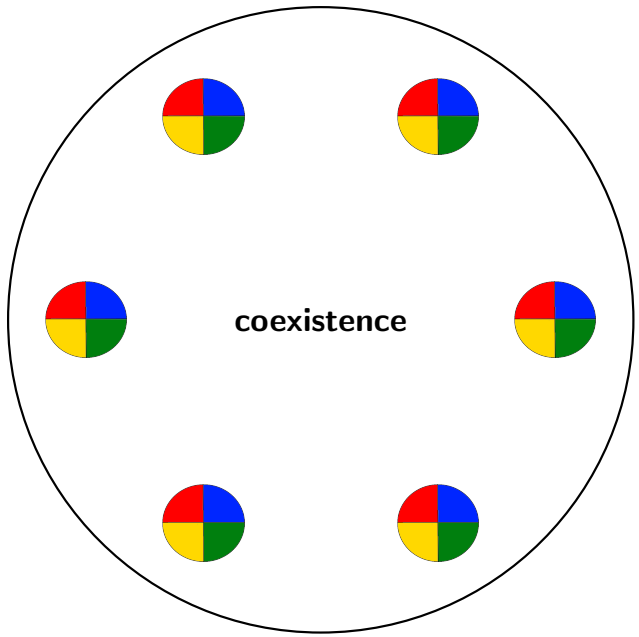
- graph "isomorphism"
- 700 different deviations



- Unphosphorylated protein
- Serine residue phosphorylated protein
- Threonine residue phosphorylated protein
- Both residues phosphorylated protein

BCS abstraction

- mixture $\rightarrow$ order not important
- 84 different deviations



sybila systems biology laboratory

Biochemical Space Language

Key Features

- **stoichiometry** $\rightarrow$ enumerated shortened forms
- **states** $\rightarrow$ encoding different forms of an entity
- **composition**
  - **full** $\rightarrow$ complexation, coexistence in a solution
  - **partial** $\rightarrow$ inner structure of interest
- **locations** $\rightarrow$ spatial organisation
- **variables** $\rightarrow$ wildcards in definitions

sybila systems biology laboratory

Differences from Other Formalisms

BioChemical **S**pace **L**anguage is a rule-based language which has:

- **no binding sites** $\rightarrow$ details of complexes formation abstracted out
  - abstract from bonds
- **annotation purpose** $\rightarrow$ specify what interacts with what without details of the interaction
- **no quantitative data** $\rightarrow$ only qualitative dynamics

sybila systems biology laboratory

Entity Declaration by Example

ENTITY ID: KaiC
ENTITY NAME: circadian clock protein kinase KaiC
STATES:
LOCATIONS: cyt
CLASSIFICATION: enzyme
DESCRIPTION: Monomer component representing a core component of the circadian clock system
LINKS: uniprot::Q79PF4, kegg::K08482, ncbi::AAM82686
ORGANISM: SYNPCC7942{Synpcc7942_1216}
NOTES:
COMPOSITION: S | T

sybila systems biology laboratory

Abstract Entity Specification

Class

KaiC.KaiC::cyt

Specialized class

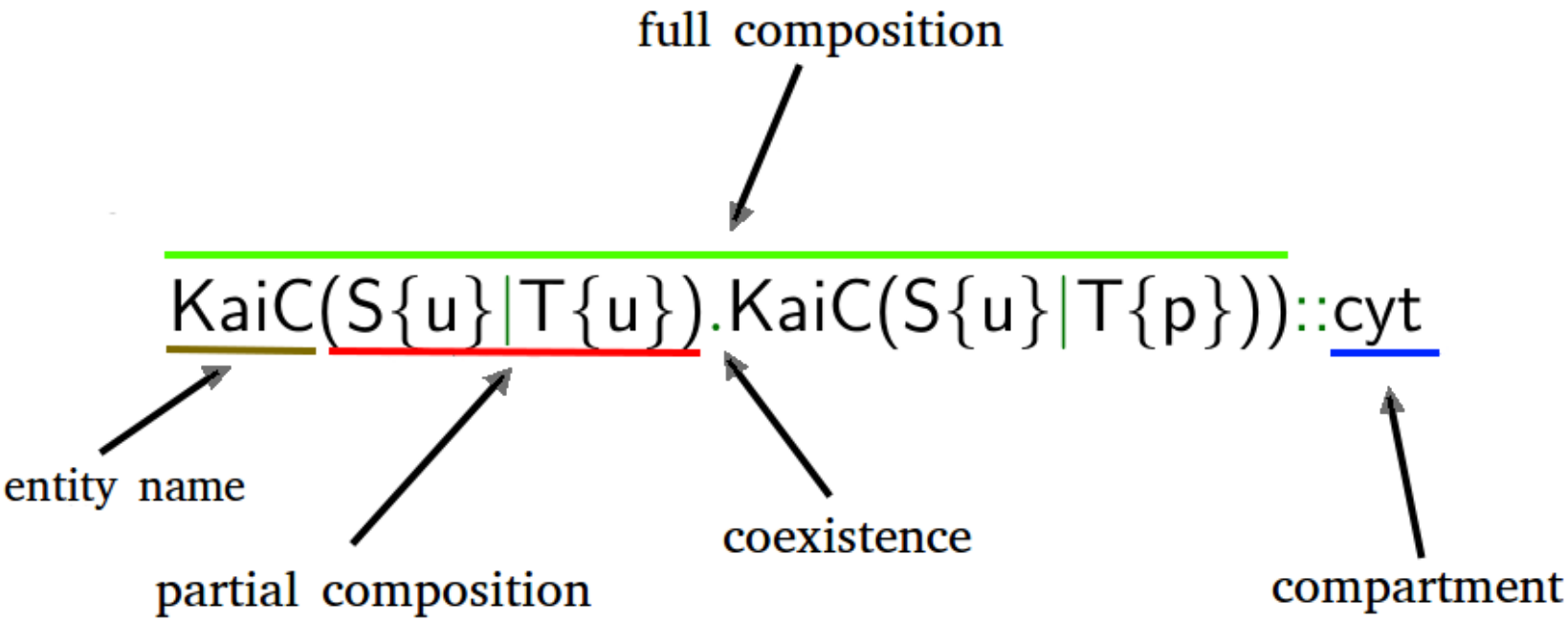
KaiC(S{u}|T{u}).KaiC::cyt
KaiC(S{u}).KaiC::cyt

Object

KaiC(S{u}|T{u}).KaiC(S{u}|T{p}))::cyt

sybila systems biology laboratory

Entity Identifier



Composition

Full composition $\rightarrow$ structure of a complex

- KaiBC == KaiC.KaiB
- KaiC6 == KaiC.KaiC.KaiC.KaiC.KaiC.KaiC

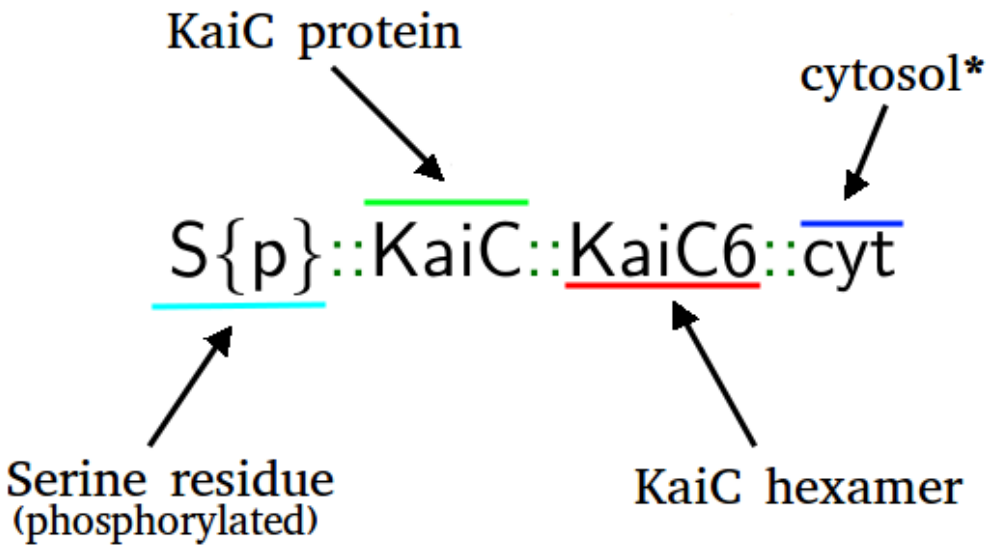
Partial composition $\rightarrow$ inner structure of an entity

- KaiC(S{u}|T{p})
- cytb6f(f{-}|bl{n}|bhc{2-})
- ps2(qb{2-}|qa{n}|chl{*}|p680{+}|pheo{-}|oec{4+}|yz{n})

sybila systems biology laboratory

Nested Entity Identifier

- another way to specify a class entity



- *compartment specification is obligatory

Examples of Entity Identifiers

- KaiC6::cyt
- KaiA.KaiA::cyt
- ATP::cyt
- CO3{2-}::liq
- KaiC(S{u}|T{p})::cyt
- CO_2::**?X** ; **X** = {lum, cell, liq, bub, pps, cyt}

Rules vs. Reactions

Reaction

$$\text{KaiC(S\{u\}|T\{u\})::cyt} + \text{KaiC(S\{p\}|T\{p\})::cyt} \Rightarrow \text{KaiC(S\{u\}|T\{u\}).KaiC(S\{p\}|T\{p\})::cyt}$$

$$\text{object} \quad + \quad \text{object} \quad \Rightarrow \quad \text{object}$$

Rule

$$\text{KaiC::cyt} + \text{KaiC::cyt} \Rightarrow \text{KaiC.KaiC::cyt}$$

$$\text{class} \quad + \quad \text{class} \quad \Rightarrow \quad \text{class}$$

sybila systems biology laboratory

Stringency of Entity Identifiers

- consider the rule:

$$S\{p\}::KaiC::KaiC6::cyt \Rightarrow S\{u\}::KaiC::KaiC6::cyt$$

- both sides identify the same object in location **cyt**
- it is a complex **KaiC6** (assume it has a given well-defined full composition)
- which contains at least one **KaiC** protein
- whose partial composition contains **S{p}**

sybila systems biology laboratory

Translation of Entities to Kappa

Assume every entity composition is lexicographically ordered.

- **agent** ← entity name
- **agent name** ← entity name suffixed with a location
- **interface** ← partial composition (at least two internal states are required)
- **site** ← each member entity of partial composition
- **site name** ← name of a member entity
- **internal state** ← state of the entity
- **binding state** ← assign a generic structure, i.e., linear

sybila systems biology laboratory

BCSL Rules Expressed in Kappa

- left/right side **S** of a BCS rule is set of entities
  $\Rightarrow$ expressed as a set **E** of agents in Kappa
- since each entity in **S** is lexicographically ordered, rules are ensured to be unique (up-to structural equivalence of reaction complexes),
- for full compositions ('.' operator) a labeling by bound sites is created (concretisation)
  $\Rightarrow$ abstract meaning of coexistence is lost

sybila systems biology laboratory

Examples

BCS: $2\ \mathrm{KaiC(S\{p\})} \Rightarrow \mathrm{KaiC(S\{p\}).KaiC(S\{p\})}$

Kappa : $\mathrm{KaiC}(S_p),\ \mathrm{KaiC}(S_p) \rightarrow \mathrm{KaiC}(S^1_p),\ \mathrm{KaiC}(S^1_p)$

BCS: $\mathrm{S\{p\}::KaiC::KaiC6::cyt} \Rightarrow \mathrm{S\{u\}::KaiC::KaiC6::cyt}$

Kappa :

$\mathrm{KaiC}(S^1_p),\ \mathrm{KaiC}(S^1,T^2),\ \mathrm{KaiC}(S^2,T^3),\ \mathrm{KaiC}(S^3,T^4),\ \mathrm{KaiC}(S^4,T^5),\ \mathrm{KaiC}(S^5) \rightarrow$

$\rightarrow \mathrm{KaiC}(S^1_u),\ \mathrm{KaiC}(S^1,T^2),\ \mathrm{KaiC}(S^2,T^3),\ \mathrm{KaiC}(S^3,T^4),\ \mathrm{KaiC}(S^4,T^5),\ \mathrm{KaiC}(S^5)$

sybila systems biology laboratory

Case Study: Synechocystis

Metabolism

- over 1000 entities (objects) and over 500 rules
- entities concrete since there is no combinatorial explosion (rules are reactions)
- two mathematical models are mapped

Photosynthesis and Respiration

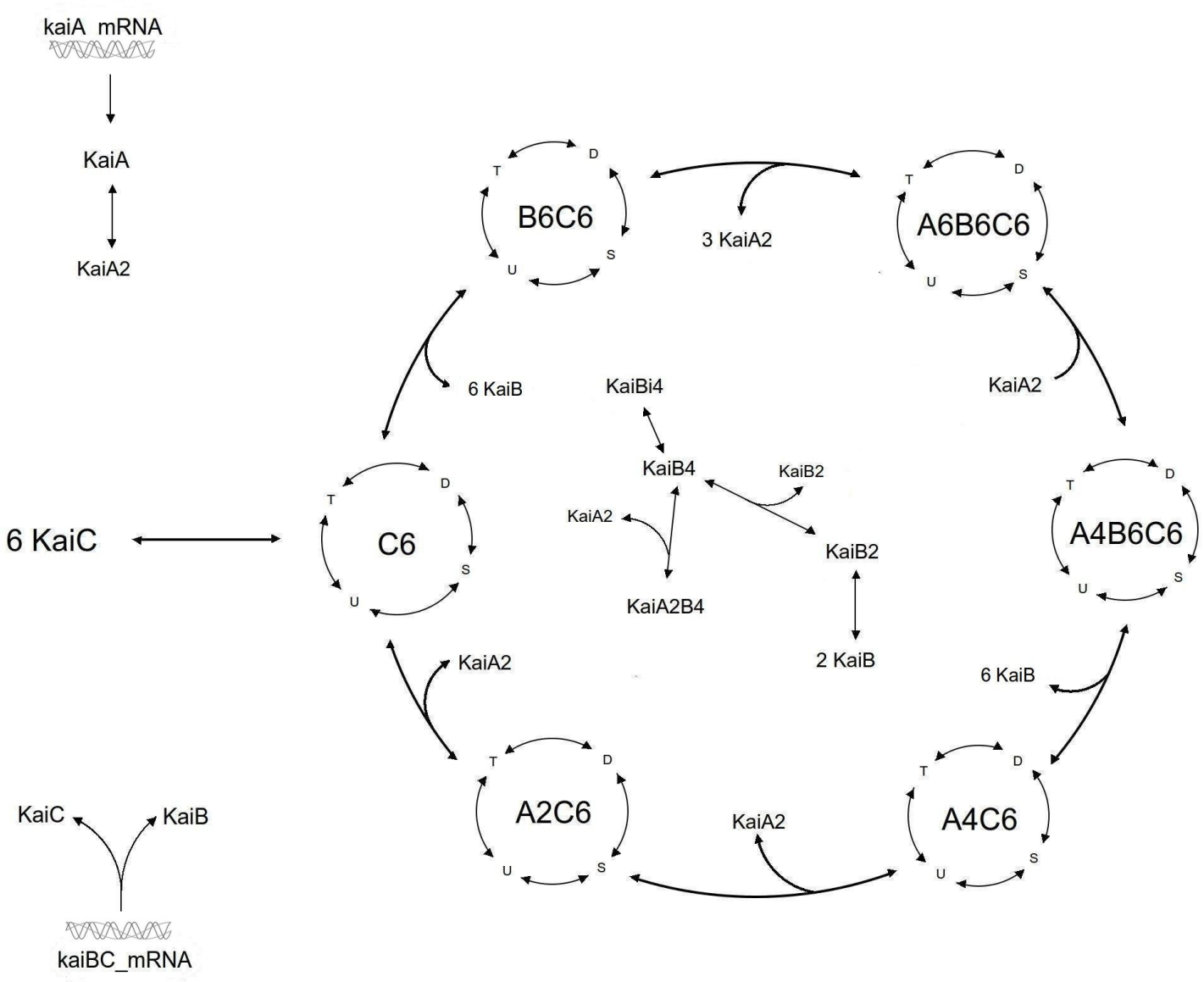
- over 100 entity classes and over 50 rules
- compaction is significant mainly for reactions
- one (lumped) mathematical model is mapped

Cyanobacterial circadian clock

- 18 class entities interacting in 18 rules
- creates over 500 object entities
- two (lumped) mathematical models are mapped

sybila systems biology laboratory

Cyanobacteria Circadian Clock BCS



Conclusions

- in the paper: formal syntax and semantics of BCS
- implementation: operational semantics in BNGL (then translated to DiVinE)
- currently we work on direct SOS to employ CTL model checking directly on BCS
- it will allow to study equivalences and reduction directly at the syntactic level of BCS

sybila systems biology laboratory

The End

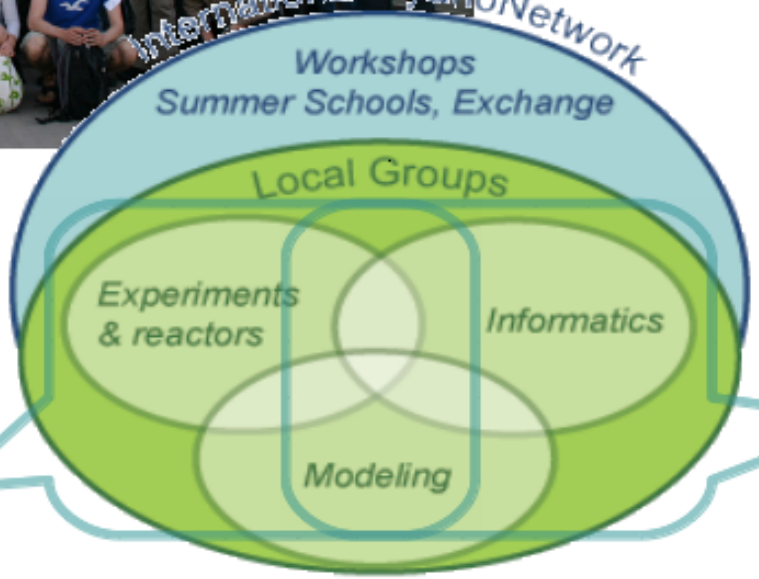
Thank You for your attention.

sybila
systems biology laboratory

The Project Team

CyanoNetwork consorcium

Humboldt University Berlin
University of Vienna
University of Turku
University of Oxford ...



www.cyanoteam.org

CzechGlobe
Centre for Global Climate Change Impacts Studies

SB study programme

Faculty of Informatics MU

sybila
systems biology laboratory