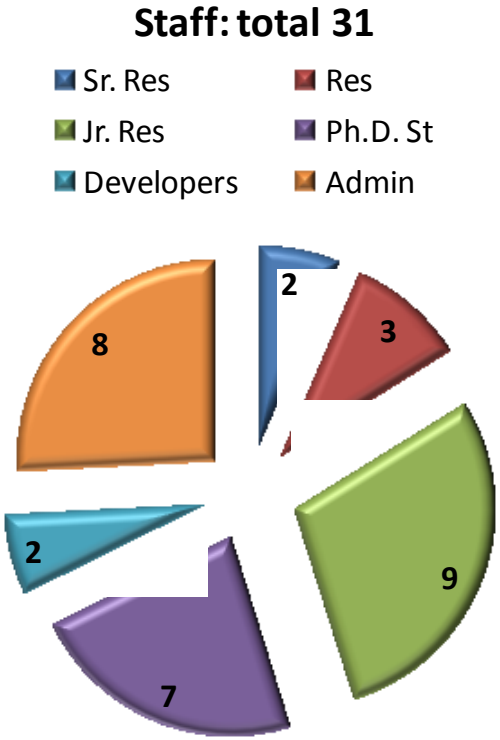
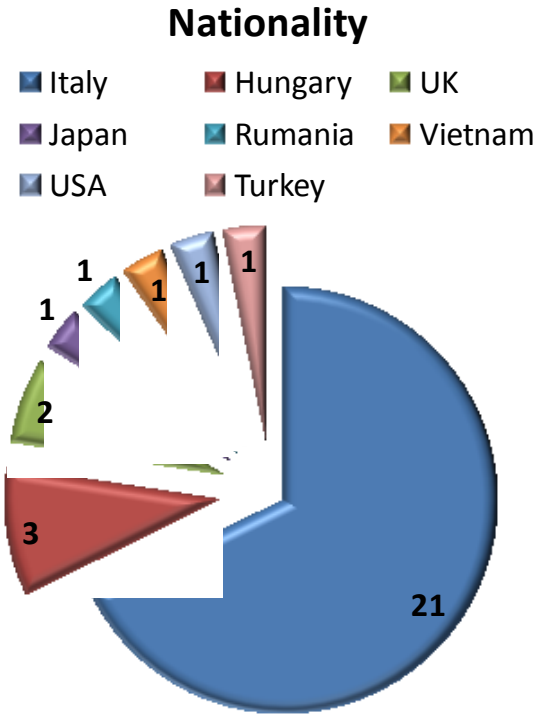
CoSBiLab: a software framework to support incremental modeling

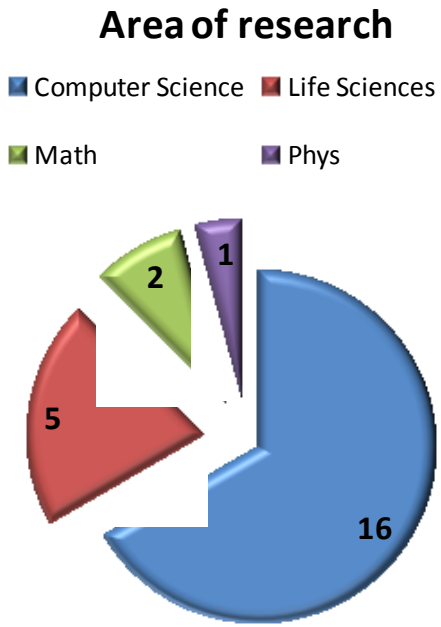
Ivan Mura
mura@cosbi.eu

Centre for Computational and Systems Biology, Trento – Italy

CoSBi in numbers







CoSBiLab

- **What we do at CoSBi**
  - In-silico lab to support modeling through refinement
- **Modeling tools at work**
  - specification
  - simulation
  - analysis/validation
  - refinement

Specification

Fuzzy pairing in biology

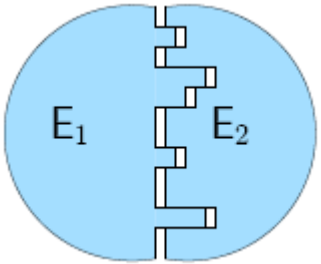
Computer science

- interactions occur as communications on a channel

☞ *exact matching name/co-name*

Key-Lock model

1 key opens 1 lock



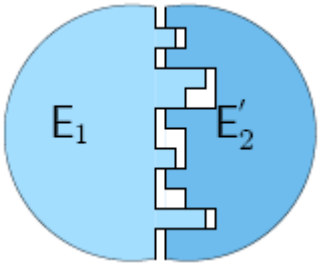
Observed biological phenomena

- protein interactions
- DNA binding

☞ *fuzzy matching, variable affinity*

Master-Key model

1 key opens a set of locks

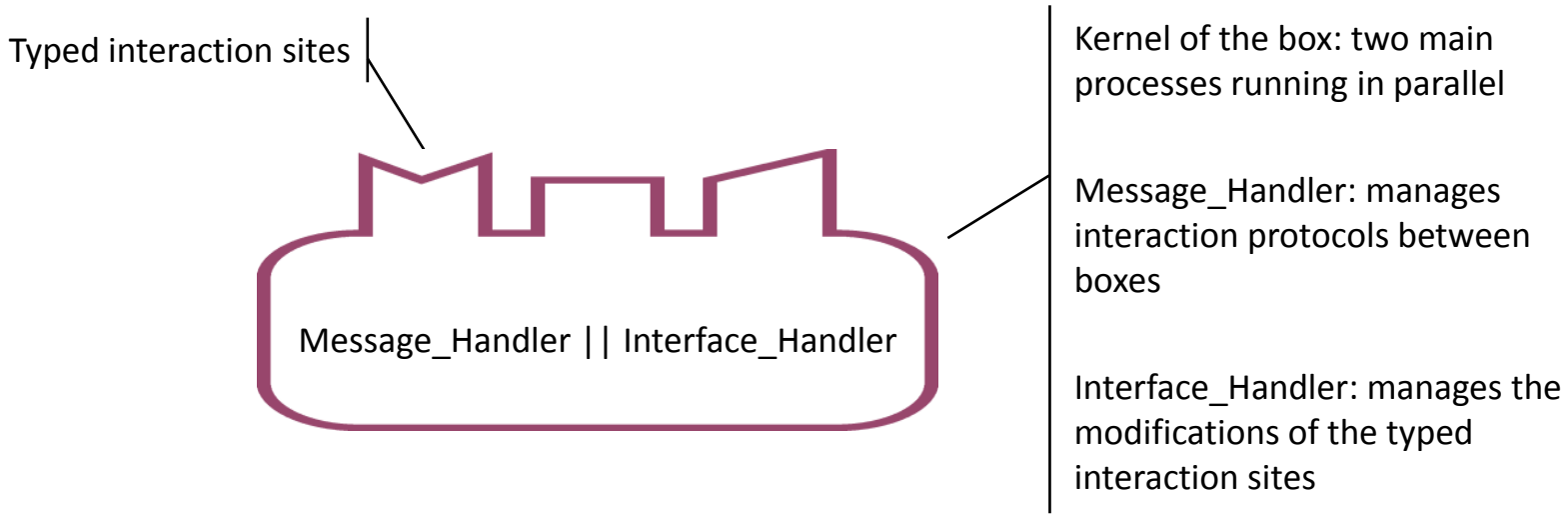


BlenX

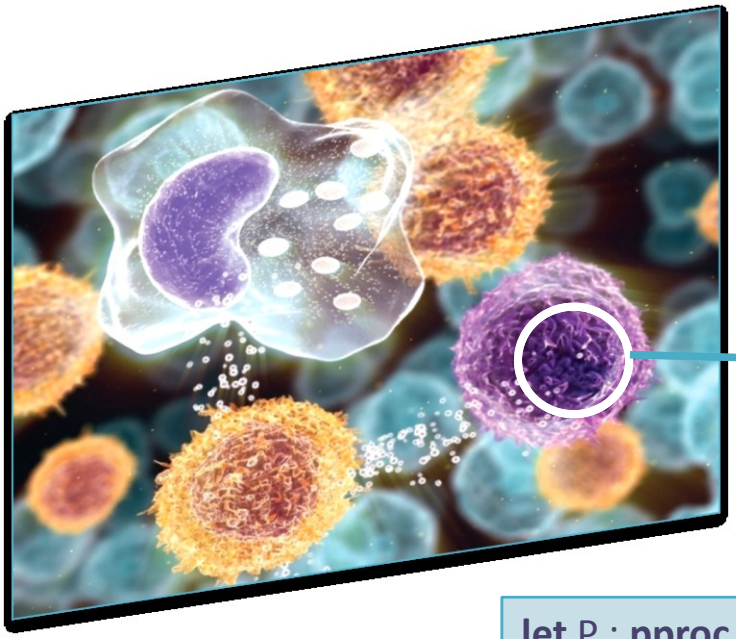
Biological transformations as interactions of ***bio-processes*** equipped with affine ***binders***

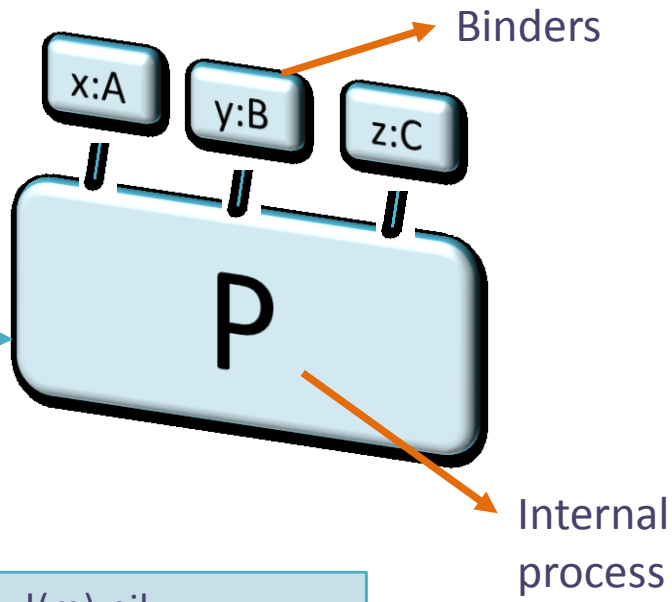
- bio-processes inspired to pi-calculus
- typed binders, affinity defined between type pairs

One box, intuitively



Language flavor



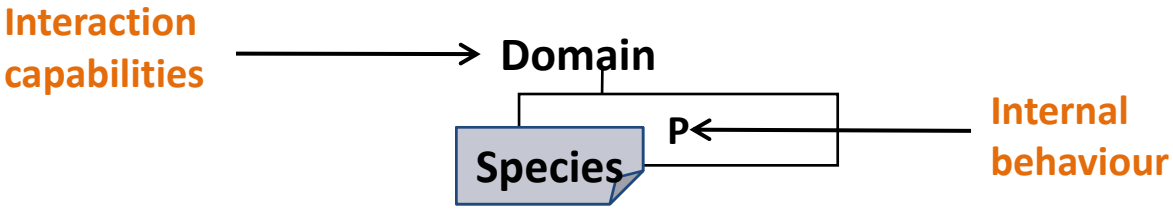


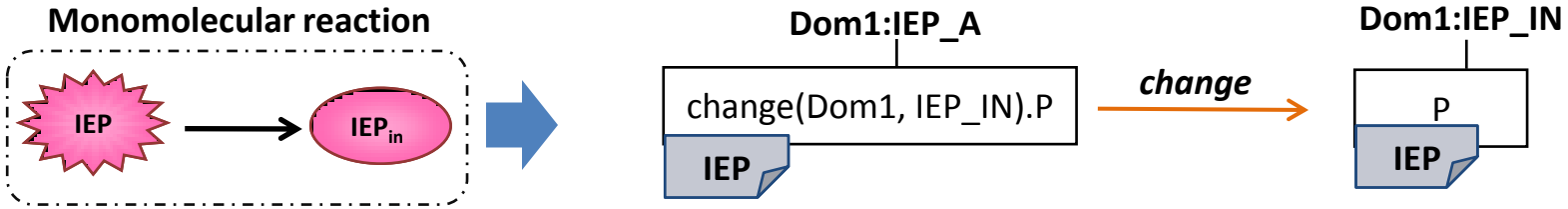

```
let P : pproc = x?(m).y!(m).z!(m).nil;

let Box : bproc = #(x,A),#(y,B),#(z,C)[ P ];
```

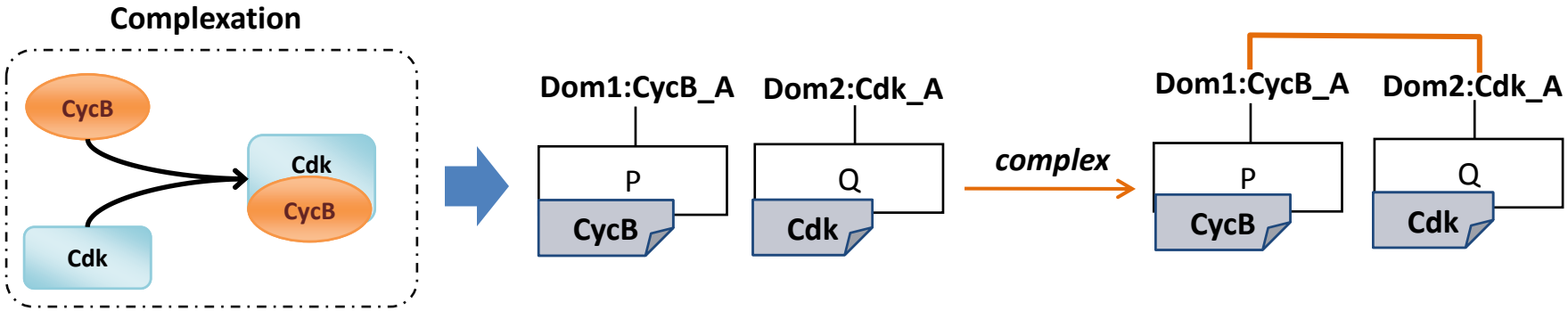
Stochastic

- each interaction happens in a random time
- the average time of interaction is determined by the affinity

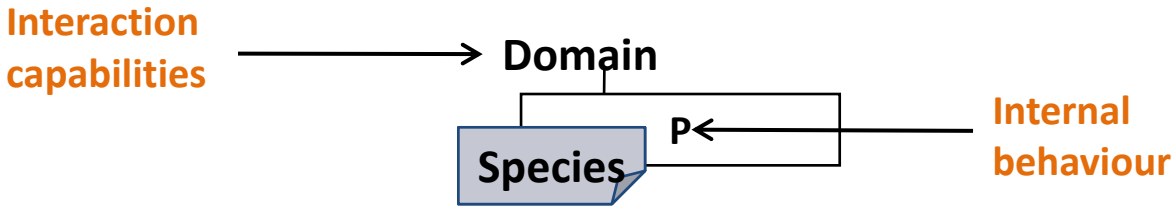
The BlenX language: elementary reactions

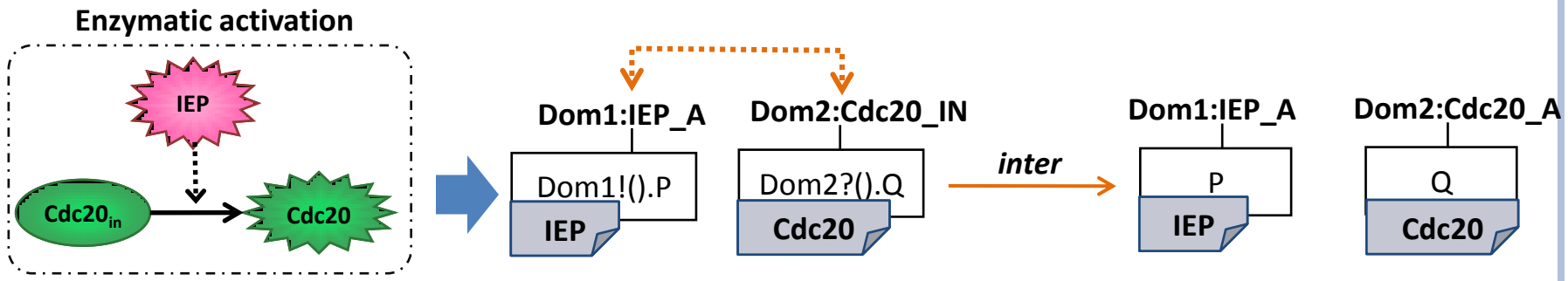


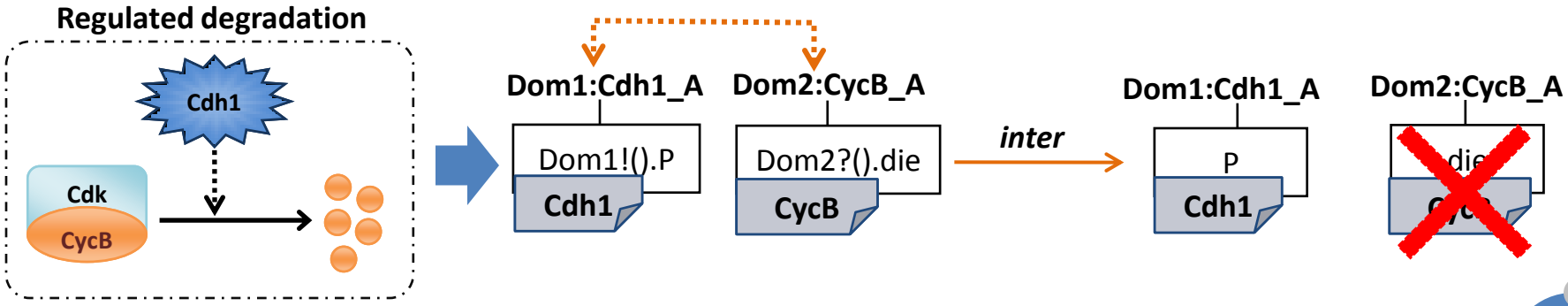




The BlenX language: abstract reactions



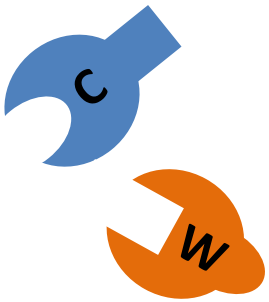


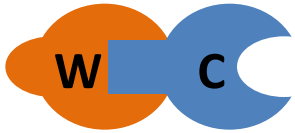


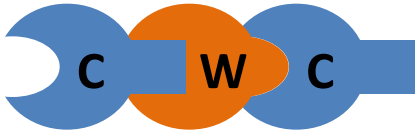
Dynamic creation of entities

Biological compounds have *sites* of interaction

- multiple sites can be present in the same entity
- bindings occur reversibly between **2** affine sites
- complexes of biological components can assembly without a precise order and can result in different topological structures
- example: protein C has 2 sites, both affine to 2 sites of protein W







How many structures can form?

- It may be cumbersome or even impossible to specify such a behaviors in many modeling formalisms

Simulation

Intrinsic discreteness

- **The truly molecular nature of biological interaction was considered hardly tractable**
  - tracking single molecules state, location and movement is indeed quite heavy from a computational point of view
- **In 1976, D. T. Gillespie**
  - proved that the evolution of a well-stirred biochemical system can be represented as a Markov process
  - provided a very simple and extremely efficient simulation algorithm for computing realizations of such process
- **Gillespie's algorithm (SSA) paved the way for a number of discrete modeling approaches**

Direct method (1976)

- **Model**
  - N species, M reactions
  - X(t) represent the state of the system at time t, t≥0 (a stochastic process)

- **Given X(t)=x , the probability that the next reaction happens in the infinitesimal time interval [t+τ,t+τ+dt] and is a reaction of type j is**

$$a_j(x) \cdot exp(-a_0(x)\,\tau)$$

  - the time t to the next reaction is an exponential random variable of mean $1/a_0(\mathbf{x})$
  - the probability that next reaction is of type j is $a_j(\mathbf{x})\,/a_0(\mathbf{x})$

- **At each simulation step, 2 uniform r.n. u and v are drawn**
  - t is chosen to be $\ln(u^{-1})/a_0(\mathbf{x})$
  - j is chosen as the smallest integer satisfying $\sum_{i=1}^{j} a_i(x) > v \cdot a_0(x)$

Reformulations of the method

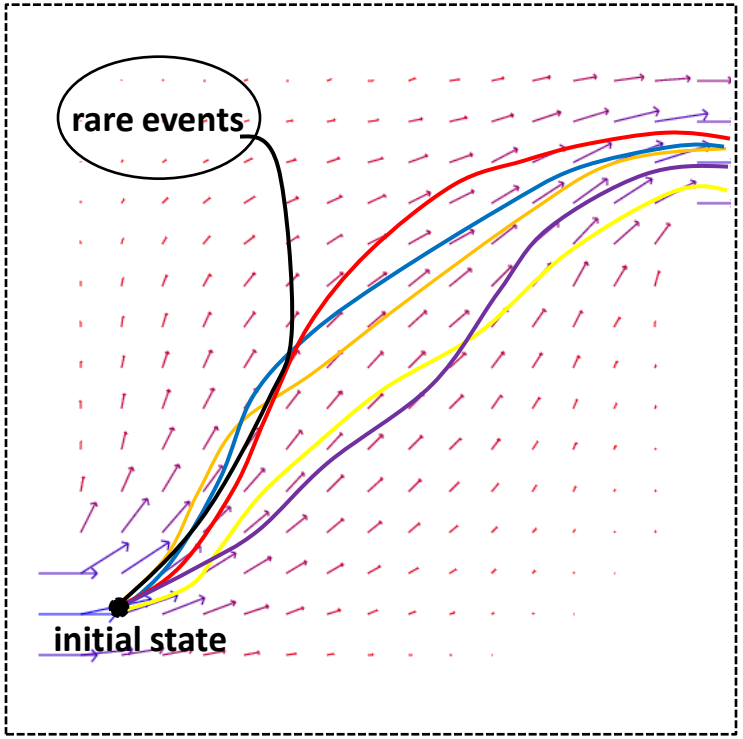
- **First reaction method (1976)**
  - at each simulation step, draw m uniform r.n. and compute t_1, t_2,…, t_m, the putative time of all reactions
  - choose t as the min(t_1, t_2,…, t_m)
  - choose *j* as the index of the minimum above
- **Next reaction (2000)**
  - same as the above one, but the putative times are saved in an indexed binary tree so that the minimum is always at the top
  - a dependency graph is used to keep track of coupling among reactions to determine when putative times in the tree have to be resampled
- **Modified direct method (2004)**
  - a pre-run to determine a suitable order of reactions to minimize cost of step 2)
- **Sorting direct method (2006)**
  - self-adaptive version of the one above, no pre-run

Rare events

- **Events that lead to abnormal states are rare but have devastating effects**
  - rare epigenetic modifications play crucial roles in cancer development
  - failure of DNA repair mechanisms occurs randomly with a very low probability per replication
- **SSA computational requirements for the analysis of rare events may be substantial**
- **Example**
  - the spontaneous switch from the lysogenic to the lytic state in phage λ-infected *E. coli* is experimentally estimated to be in the order of 10^{-7} per cell per generation
  - SSA would generate sample trajectories of this rare event once every 10^{7} runs
    - it would require more than 10^{11} simulation runs to generate an estimated probability within 1% relative half-width of a confidence interval (at a 95% confidence level)

Biasing SSA trajectories: wSSA

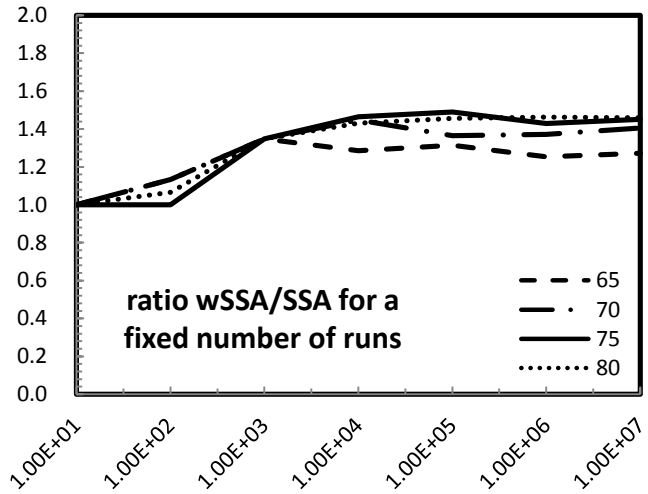
- **Trajectories generated by the SSA are determined by the reaction propensities** $a_j(x)$, $j=1,2,\ldots,M$
  - vector field in the phase portrait
- **Rare event ⇨ most trajectories do not reach it**
- **Biasing propensities (I.S. technique) can increase the likelihood of finding it**
  - more samples in lesser simulation runs
  - increased precision with the same computation time
- **There is a simple way to**
  - bias the propensities
  - remove the bias from results

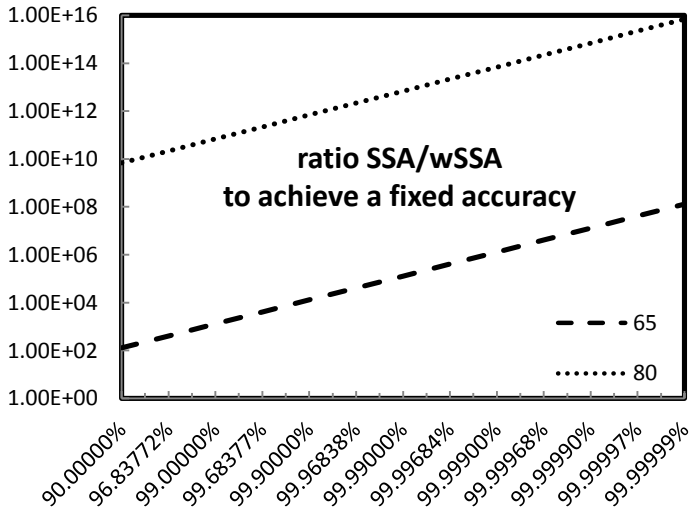


Kuwahara, Mura. **An efficient and exact stochastic simulation method to analyze rare events in biochemical systems**, JCP, 129(16), 2008

Computational saving

- **Example: simple synthesis/degradation system**
  - $\varnothing \rightarrow A$, with rate $k_1=1$
  - $A \rightarrow \varnothing$, with rate $k_2=0.025$
  - analytical solution, steady state distribution of A is Poisson, average A=40
- **We estimate the probability that, starting at t=0 with A=40, the systems reaches a state with A={65,70,75,80} within the interval [0,100]**



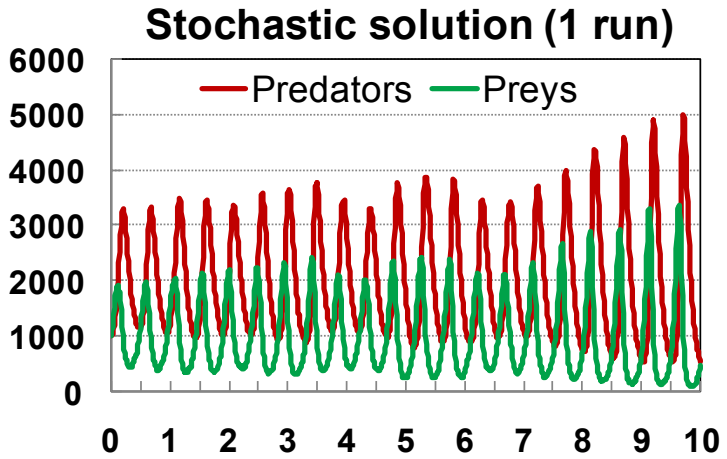


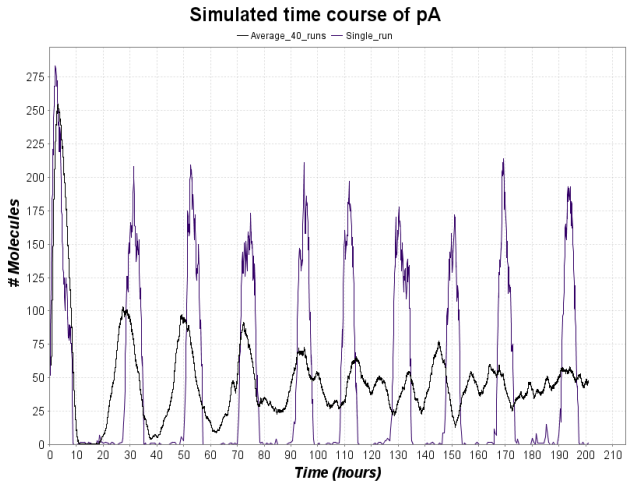
- **When the rare event is defined as "reaching the state when A=80 within [0,100]", wSSA can compute an estimate that is 99.9999% accurate in $5.6 \cdot 10^4$ secs, whereas SSA would require $2.3 \cdot 10^5$ years**

Analysis/validation

Oscillatory behaviors

- **Many biological systems achieve equilibrium conditions that are not commonly found in artificial systems**
  - living systems keep oscillating
- **Many systems have transient oscillation that stop abruptly**
  - dead
- **This poses issues in**
  - defining adequate measures that can characterize cyclic system behavior
  - comparing variants of models of systems



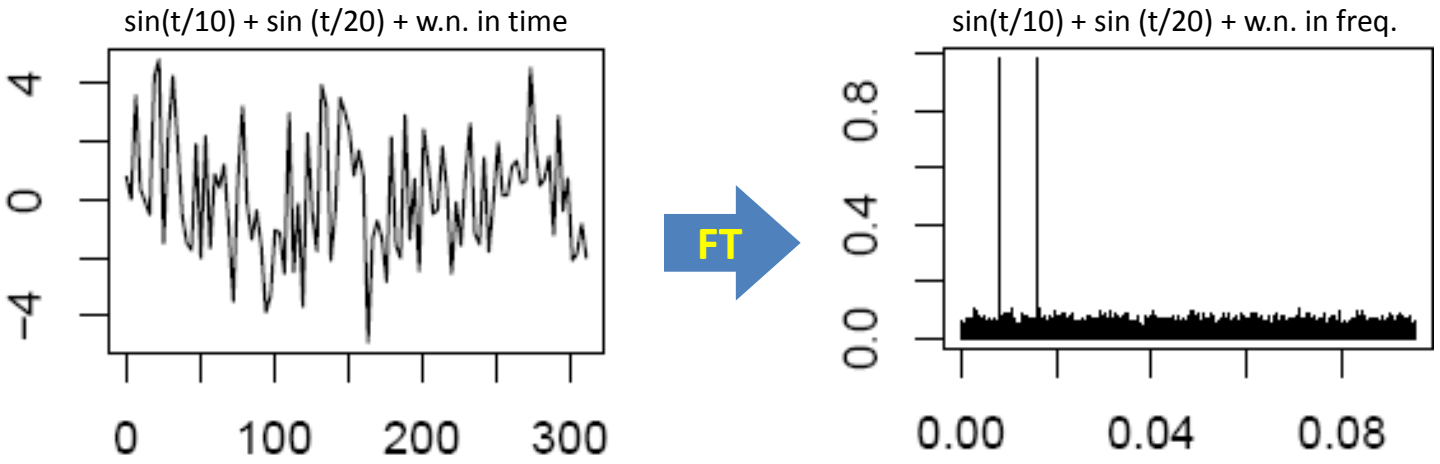


Fourier analysis

Analysis of stochastic simulations for oscillatory phenomena

- identification of oscillation
- determination of period from noisy traces

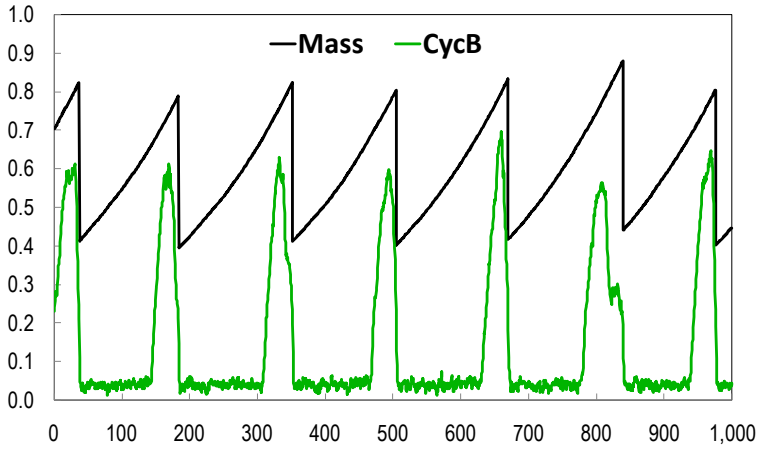
A time to frequency transformation can facilitate the job

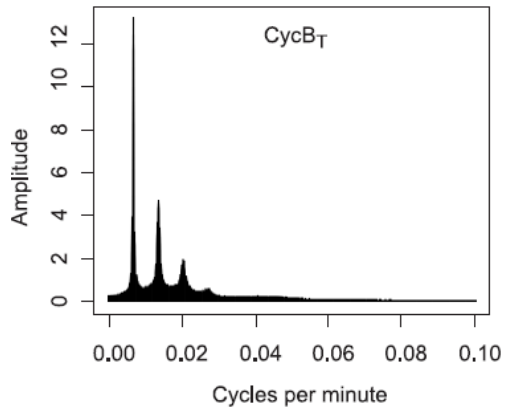


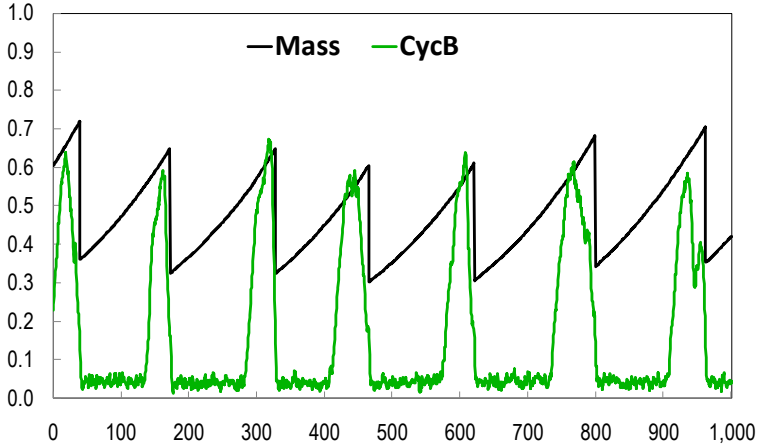
Comparison between models

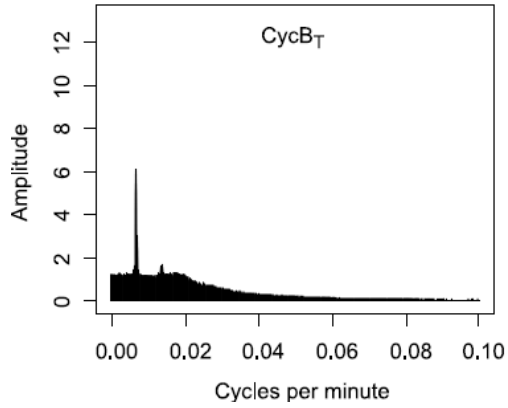
Wild Type and cln1Δ cln2Δ cln3Δ sic1Δ S. cerevisiae mutant

- spectra obtained from multiple stochastic simulation traces

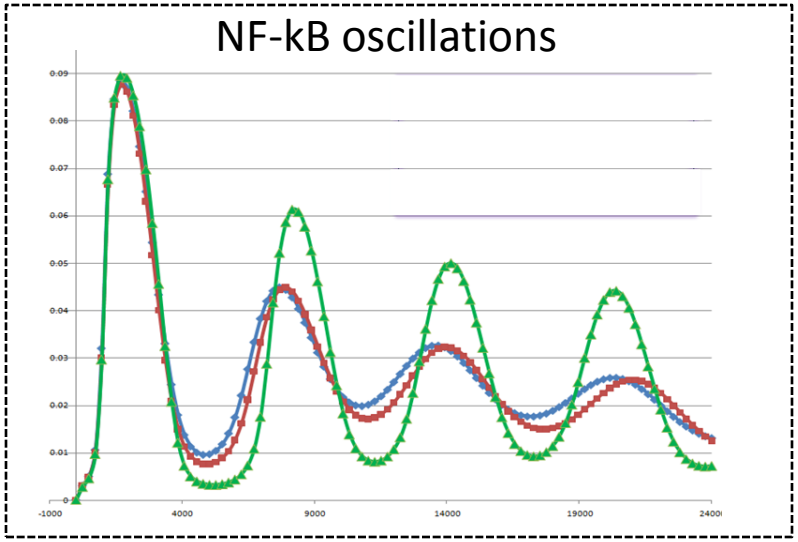


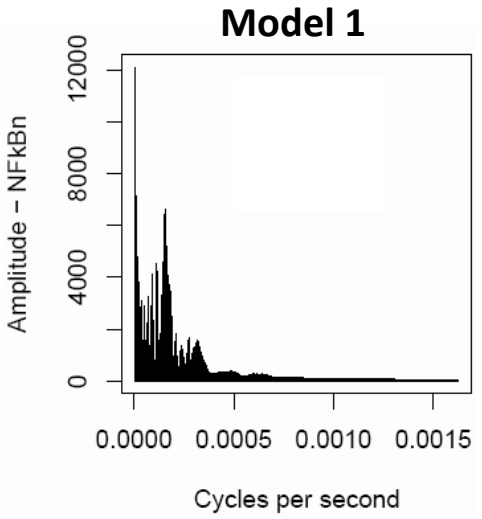


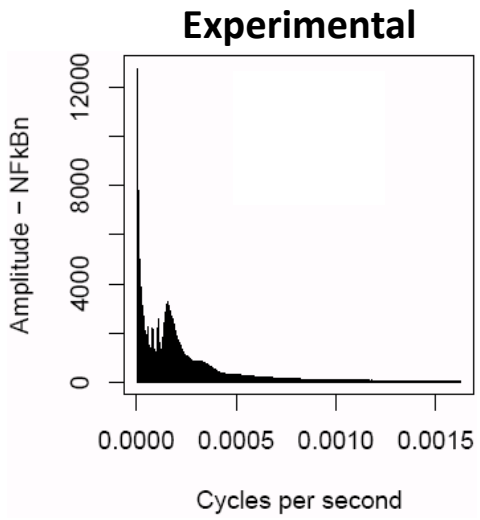




Comparison of models and experiments

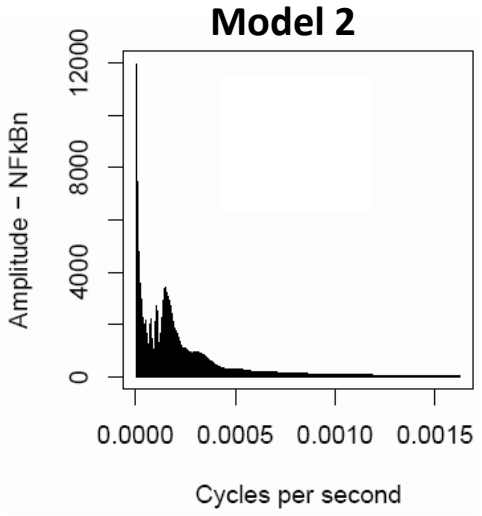






$\rho = 0.0625$

$\rho = 0.0352$



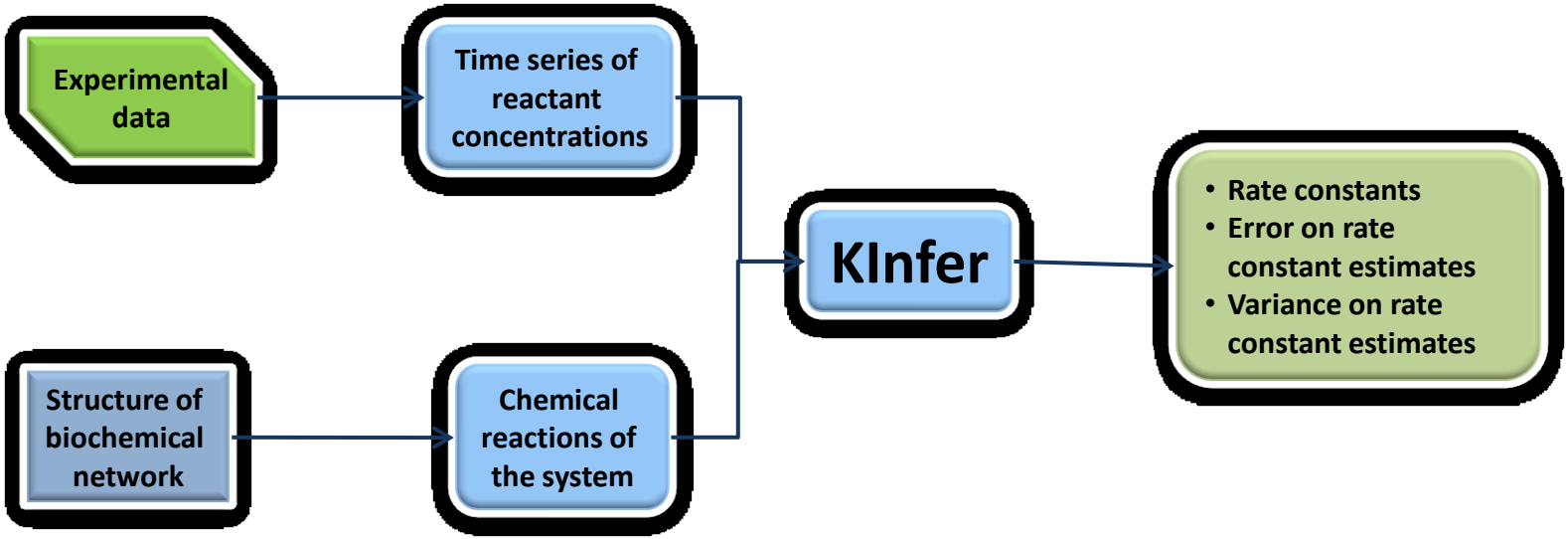
Refinement

Including new experimental data

A model gets refined by inclusion of new experimental evidence

- additional species
- detailed reactions

We are working on a tool for estimating the kinetic parameters of models from wet-lab data samples



ML approach

The values of the model's parameters have to be the most likely values giving the observed time course of the concentration.

The Maximum likelihood (ML) approach can be used to achieve this goal. The main steps are:

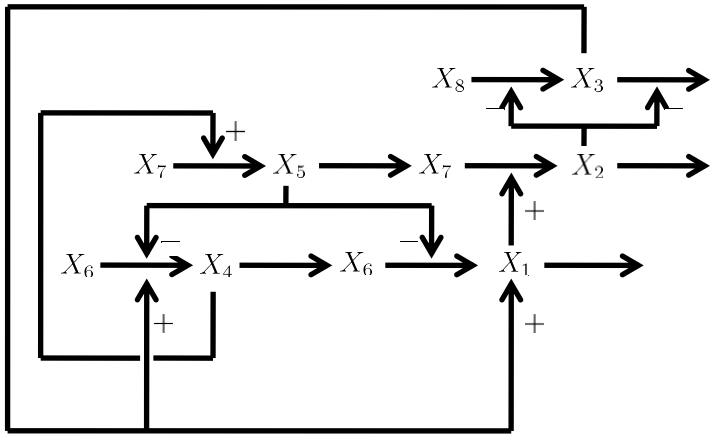
- to build a suitable expression of the joint transitional density for expressing the probability density function of the observed outcomes in terms of measured system variables and parameters;
- and to optimize this function to determine unknown parameters.

Lecca, Palmisano, Priami, Sanguinetti, A new probabilistic generative model of parameter inference in biochemical networks, ACM Symposium on Applied Computing 2009

Kikuchi model, benchmarking rate inference

Kikuchi et. al, "*Dynamic modeling of genetic networks using genetic algorithm and S-system*", BIOINFORMATICS, 19(5) 2003

A gene regulation network with complex feedbacks



Our results

- simulations with inferred rates