Statistical model checking based analysis of biopathway models

P.S. Thiagarajan
Dept of Computational and Systems Biology
University of Pittsburgh
(Faeder Lab)

Classical model checking

- *Verify* if a *model* satisfies a specified *property*
- model
  - ❑ a *dynamical system*
    - ➢ hardware circuits
    - ➢ Programs
    - ➢ …..

Classical model checking

- Properties:
  - ❑ Assertions about the executions (trajectories) of the dynamical system.
    - *At some time* in the future the program will terminate *(ϕ1)*
    - Starting from now *at every time* it will be the case that x1 + x2 = 100 *(ϕ2)*

- *l1*: Input x1, x2
- *l2*: while x1 > 0:

  x1 := x1-1; x2 := x2+1
- *l3*: stop

Classical model checking

- Model satisfies a property if ***every*** *run/execution* of the model satisfies the property.
- *At some time* in the future the program will terminate *(ϕ1)*
- Starting from now *at every time* it will be the case that x1 + x2 = 100 *(ϕ2)*

- *l1*: Input x1, x2
- *l2*: while x1 > 0:

  x1 := x1-1; x2 := x2+1
- *l3*: stop

ϕ1 is satisfied by the program

ϕ2 is *not* satisfied by the program

Classical model checking

- Properties:
  - ❑ specified as *temporal logic formulas*

- *At some time* in the future the program will terminate

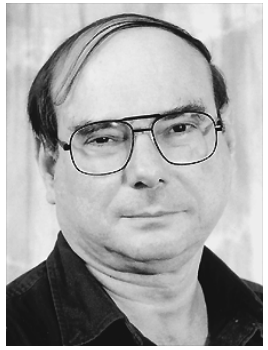
  F(l3)

- Starting from now *at every time* it will be the case that x1 + x2 = 100

  G(x1+x2 == 100)

Classical model checking

- Properties:
  - ❑ specified as *temporal logic formulas*
- future (F), always(G), until (U), next (X)
- and, or , not ….



Amir Pnueli (Turing Award 1996)

- Precise
  - ❑ (machine readable) syntax
  - ❑ mathematical semantics

Classical model checking

- Properties:
  - ❑ specified as *temporal logic formulas*
- future (F), always(G), until (U), next (X)
- and, or , not ….

- Precise
  - ❑ (machine readable) syntax
  - ❑ mathematical semantics
- The model checking problem can be solved ***automatically!***

Ed Clarke

Alan Emerson

Turing award 2007

Joseph Sifakis

Probabilistic model checking

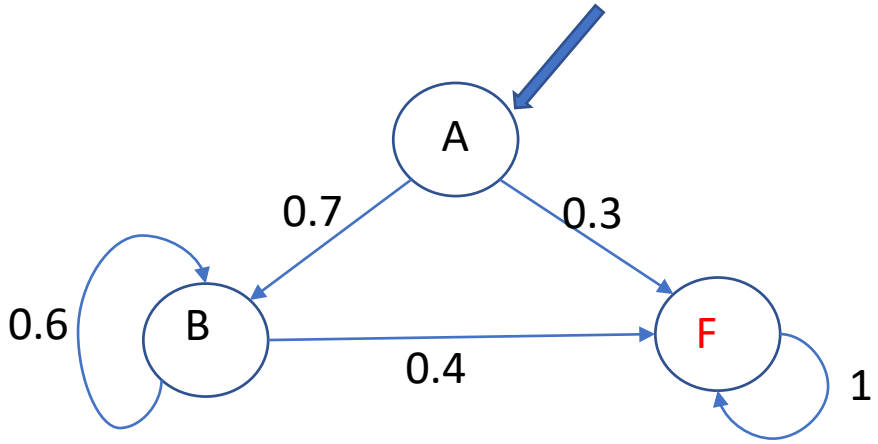
- *Verify* if a *model* satisfies a specified property *with a certain probability.*
- Models:
  - ❑ Stochastic dynamical systems
    - ➢ Discrete time Markov chain
    - ➢ Continuous time Markov chain
- Model satisfies a property with probability p if:
  - ❑ the probability of ***a randomly chosen*** *run/execution* of the model satisfying the property is p.
- This is hard problem!

Statistical model checking

- Probabilistic model checking via:
  - ❑ *sequential hypothesis testing*.
- H0: $P(\varphi) \geq r$ (null hypothesis)
- H1: $P(\varphi) < r$ (alternative hypothesis)
- r chosen by the user.
- User also fixes
  - ❑ α - false positives probability
  - ❑ β - false negatives probability
- These parameters determine the thresholds L and U

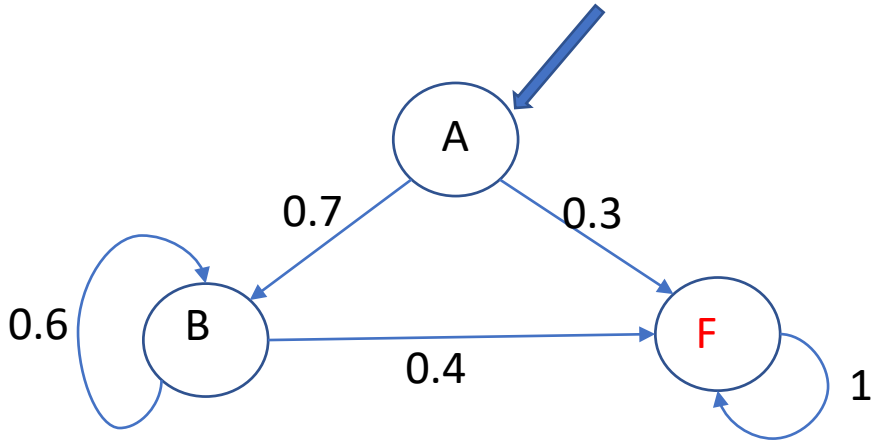
Statistical model checking

- φ – “within two steps the state F will be reached”
- H0: $P(\varphi) \geq 0.8$
- H1: $P(\varphi) < 0.8$
- $\alpha = \beta = 0.05$
- L, U



Statistical model checking

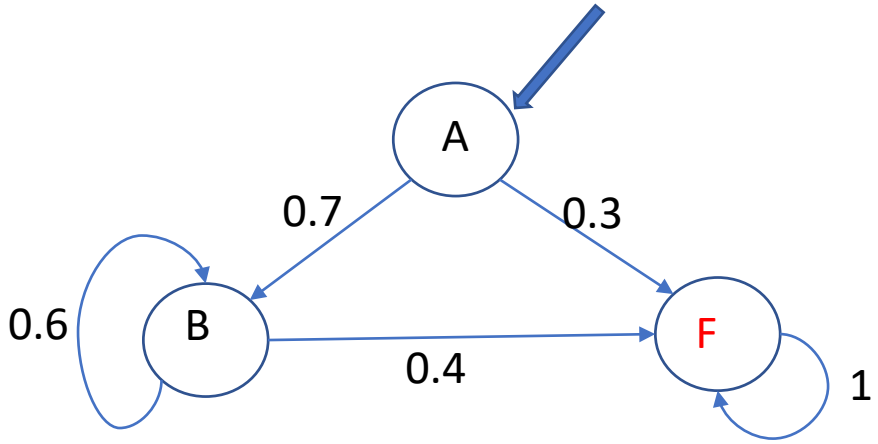
- φ – “within two steps the state F will be reached”
- Suppose m sample trajectories have been drawn so far
- and the test ratio value K_m lies between L and U
- Draw one more sample trajectory σ.



σ0 A F F Yes
σ1 A B B No
σ2 A B F yes

Statistical model checking

- If σ satisfies φ, increase K_m to K_{m+1}
- else decrease K_m to K_{m+1}
- If $K_{m+1} > U$ accept H0 and stop
- If $K_{m+1} < L$ accept H1 and stop
- Else draw one more sample and repeat.



σ0 A F F Yes
σ1 A B B No
σ2 A B F yes
……

Statistical model checking

- The hypothesis test is guaranteed to terminate with probability 1.
- Surprisingly few samples need to be drawn in practice
- Complexity depends on the hypothesis test parameters only
  - ❑ Cost of drawing a sample will depend on the dimension of the system.
- Amenable to parallel implementation
- Scales well

Younes, H.L.S., Simmons, R.G.: Statistical probabilistic model checking with a focus on timebounded properties. ***Inform. Comput****. 204, 1368–1409 (2006)*

Goal

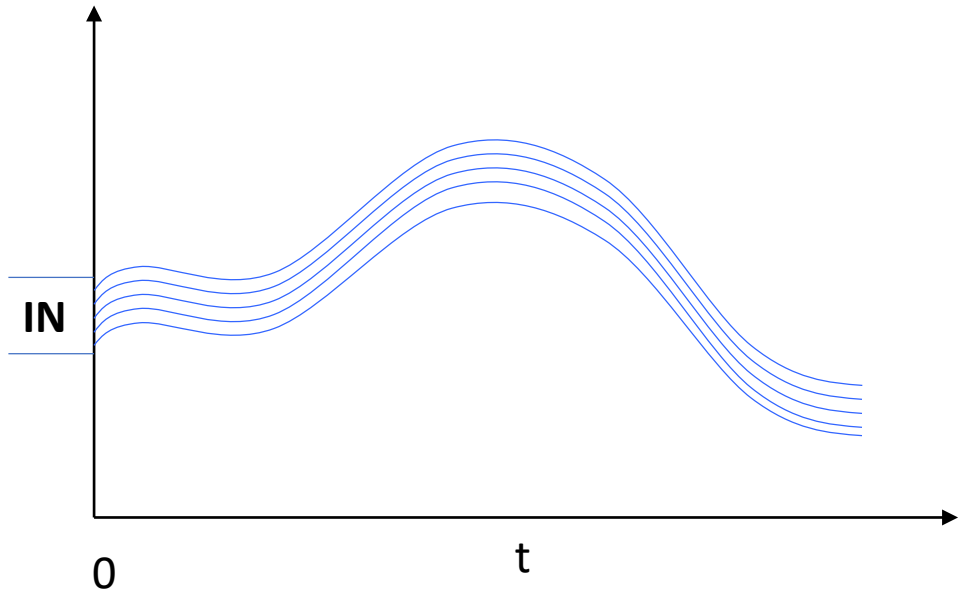
- Apply the SMC method to analyze:
  - ODEs based models of biochemical networks.
- Parameter estimation
- Sensitivity analysis
- Model check (probabilistically, approximately) for properties.

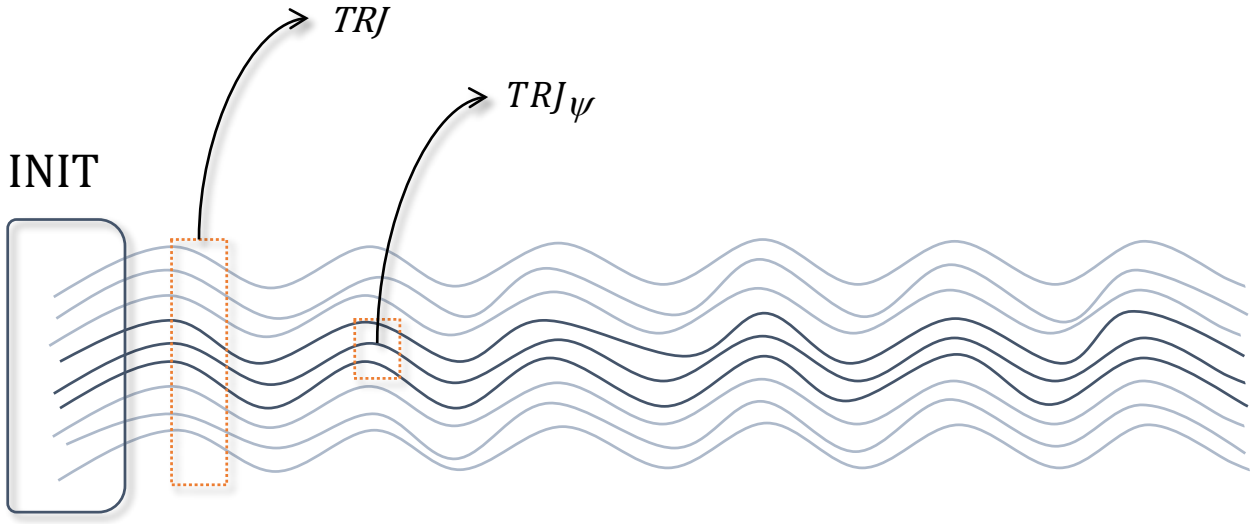
- Assume a set (interval) of initial values:
  - For the variables
- Assume distributions over these sets of initial values.

$$\frac{dS}{dt} = -k_1 \cdot [E]\,[S] + k_2 \cdot [ES]$$
$$\frac{dES}{dt} = k_1 \cdot [E]\,[S] - (k_2 + k_3) \cdot [ES]$$
$$\frac{dE}{dt} = -k_1 \cdot [E]\,[S] + (k_2 + k_3) \cdot [ES]$$
$$\frac{dP}{dt} = k_3 \cdot [ES]$$

SMC for ODEs

- Assume a set (interval) of initial values:
  - For the variables
  - Assume for now all the rate constants are known
- Assume distributions over these sets of initial values.
  - Uniform
  - Normal
  - Log uniform
  - lognormal



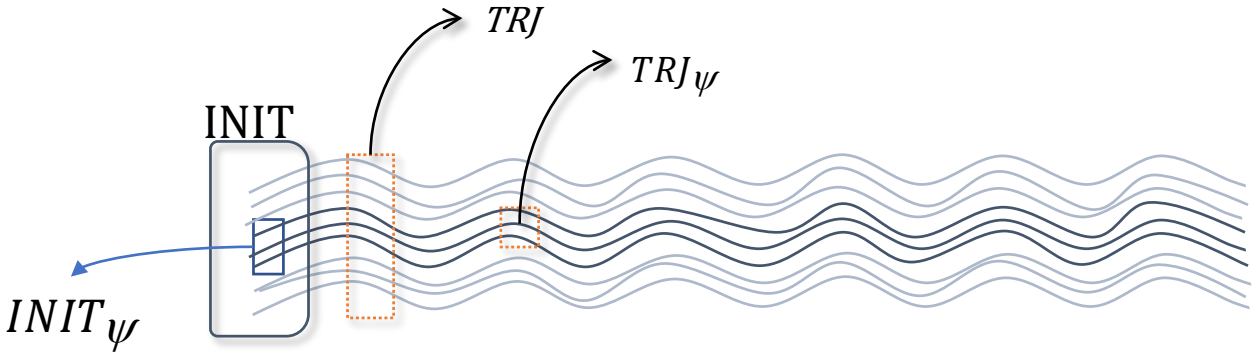


ψ *a BLTL formula*

$$TRJ_\psi = \{\tau \mid \tau \textit{ satisfies } \psi\}$$

$$P(\psi) = \frac{\#TRJ_\psi}{\#TRJ} = \frac{\mu(INIT_\psi)}{\mu(INIT)}$$

$$INIT_\psi = \{\tau(0) \mid \tau \text{ satisfies } \psi \textit{ and } \tau(0) \textit{ in } INIT\}$$



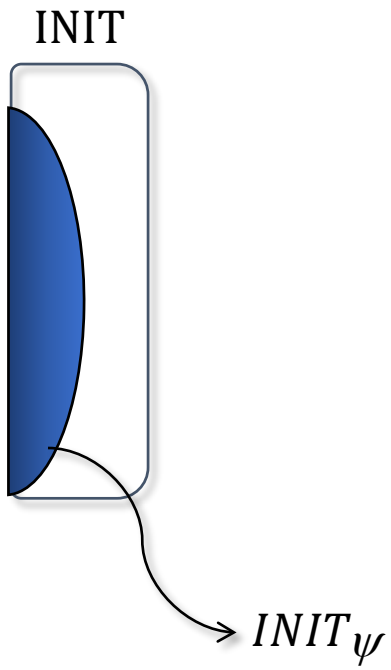
$$TRJ_\psi = \{\tau \mid \tau \textit{ satisfies } \psi\}$$

$$P(\psi) = \frac{\#TRJ_\psi}{\#TRJ} = \frac{\mu(INIT_\psi)}{\mu(INIT)}$$

$$INIT_\psi = \{\tau(0) \mid \tau \text{ satisfies } \psi \textit{ and } \tau(0) \textit{ in } INIT\}$$

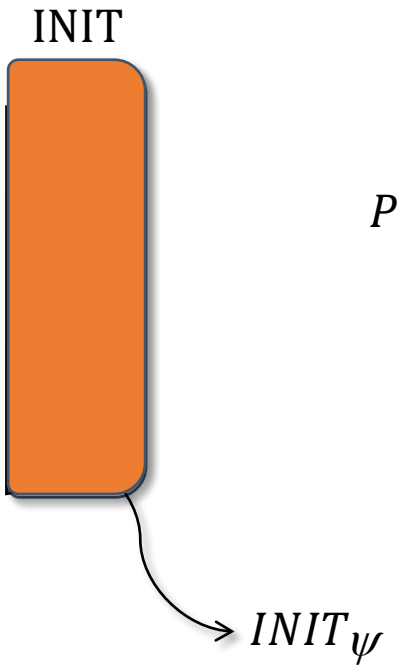
P(ψ) is well-defined because:

- *The assumed continuity properties of the ODEs system*
- *BLTL semantics*
- *Basic measure theory*



$$INIT_\psi = \{\tau(0) \mid \tau \text{ satisfies } \psi\}$$

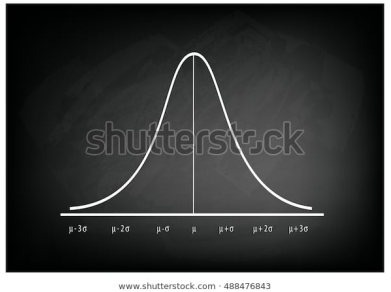
$$P(\psi) = \frac{\mu(INIT_\psi)}{\mu(INIT)}$$

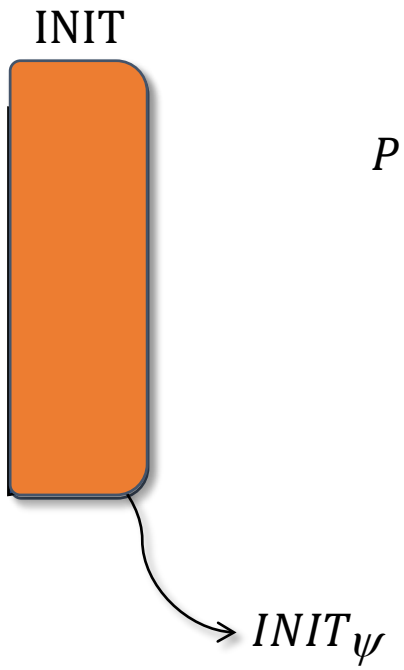


$$INIT_\psi = \{\tau(0) \mid \tau \text{ satisfies } \psi\}$$

$$P(\psi) = \frac{\mu(INIT_\psi)}{\mu(INIT)} = P(INIT_\psi)$$

- We can estimate P(ψ) by:
  - ❑ Estimating P($INIT_\psi$)
  - ❑ Using the given distribution over INIT





$$INIT_\psi = \{\tau(0) \mid \tau \text{ satisfies } \psi\}$$

$$P(\psi) = \frac{\mu(INIT\,\psi)}{\mu(INIT)} = P(INIT_\psi)$$

- Use SMC to estimate $P(INIT_\psi)$
  - ❑ Sample a point x_0 from INIT
  - ❑ Generate a trajectory σ starting from
  - ❑ Check if σ satisfies ψ
  - ❑ .....

For an ODEs system:
Given a distribution over the initial values sets
We can estimate/bound the probability of the system satisfying the property ψ

Parameter estimation

- Given an ODEs system:
  - ❑ Assume distributions over initial values sets
  - ❑ Assume distributions over intervals of values for unknown parameters
  - ❑ *Encode quantitative experimental data and known qualitative properties as a conjunction of BLTL formulas.*
  - ❑ Use SMC to evaluate the objective value of the current set of parameters
  - ❑ Use standard search techniques to traverse the parameter space.

Data encoding

- Quantitative experimental data
  - ❑ At time t the value of the variable x was observed to lie in the interval [l, u]
    - ➢ $F^t(l \leq x \text{ and } x \leq u)$
  - ❑ Ψ_{exp} – the conjunction of all such data point formulas.

Data encoding

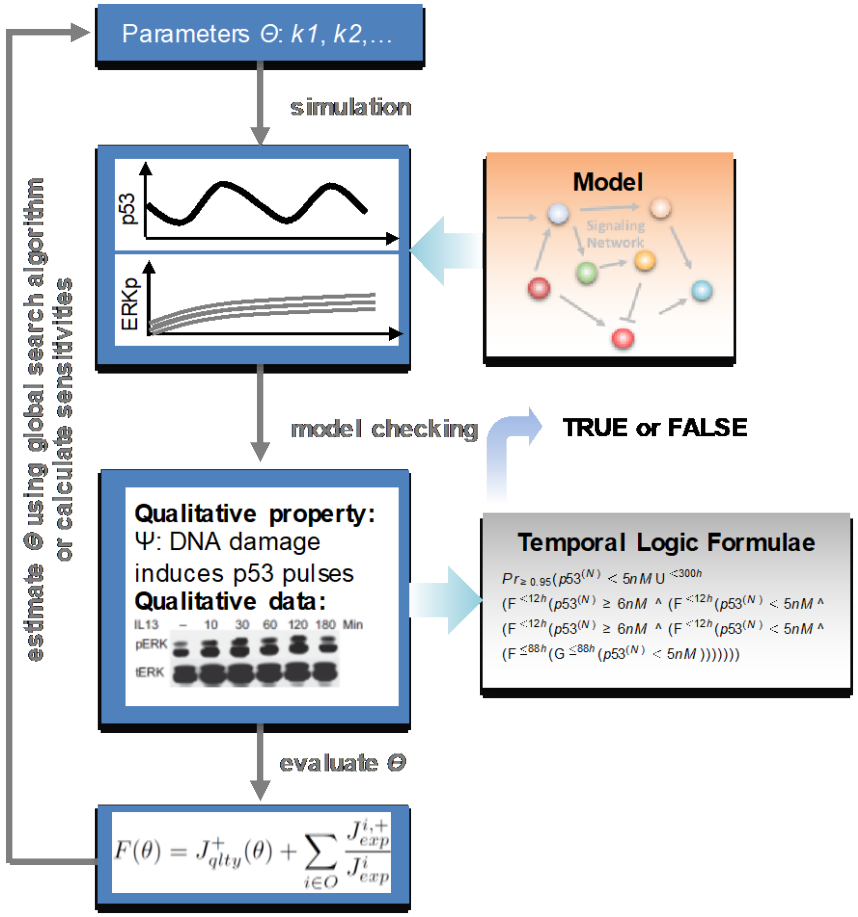
- Known Qualitative trends
  - ❑ ERK concentration reaches a peak value and then drops of to a low value for good.
    - ➢ F([ERK] > 4.8 and F (G([ERK] ≤ 0.2))
  - ❑ transient/sustained activation, oscillatory behavior, bistable, …
- Ψ_{qlt} - the conjunction of all qualitative properties.

$$\Pr_{\geq r}(\psi_{\exp} \wedge \psi_{qlty})$$

SMC based Parameter Estimation

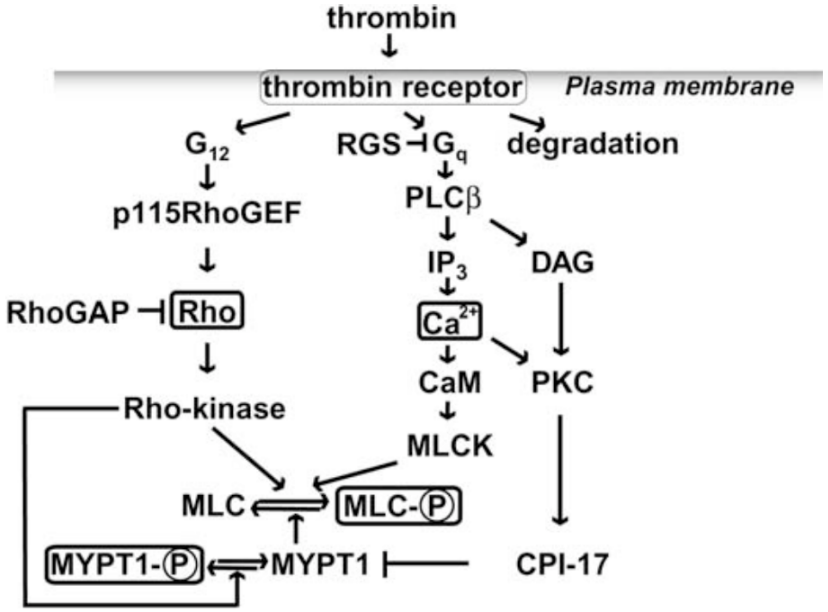
1. Guess ϑ_l
2. Verify $\psi_{\text{exp}} \wedge \psi_{qlty}$ with the chosen strength
3. Compute $F(\vartheta_l)$
4. Terminate or make a new guess (based on SRES) and repeat step 1

$$F(\theta) = J^{+}_{qlty}(\theta) + \sum_{i \in O} \frac{J^{i,+}_{exp}}{J^{i}_{exp}}$$



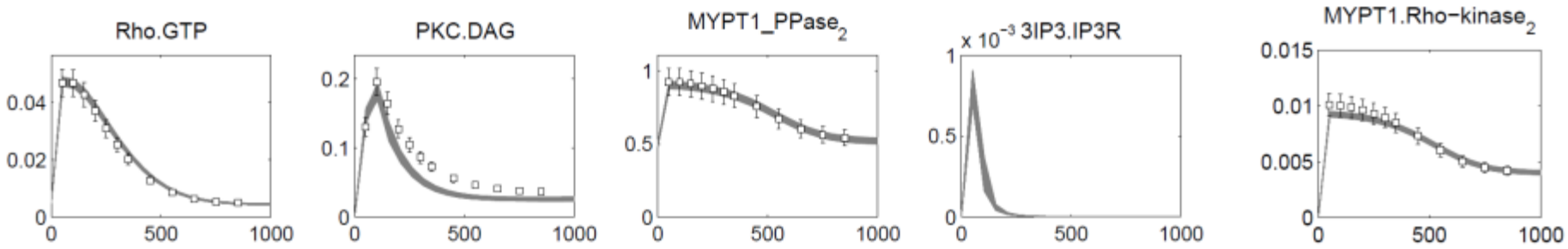
MLC Phosphorylation Pathway

- Regulates the contraction of endothelial cells
- ODE model (*Maeda et al 2006*)
  - **105** species, 197 parameters (**100** unknown parameters)
- Synthetic training data
  - Time serials: 10 species, 20 time points
  - Qualitative trend: 2 species
- Synthetic test data
  - 2 species, 12 time points



Maeda A[1] et.al. ***Ca2+ -independent phospholipase A2-dependent sustained Rho-kinase activation exhibits all-or-none response.*** ***Genes Cells.*** *2006 Sep;11(9):1071-83*

MLC Phosphorylation Pathway

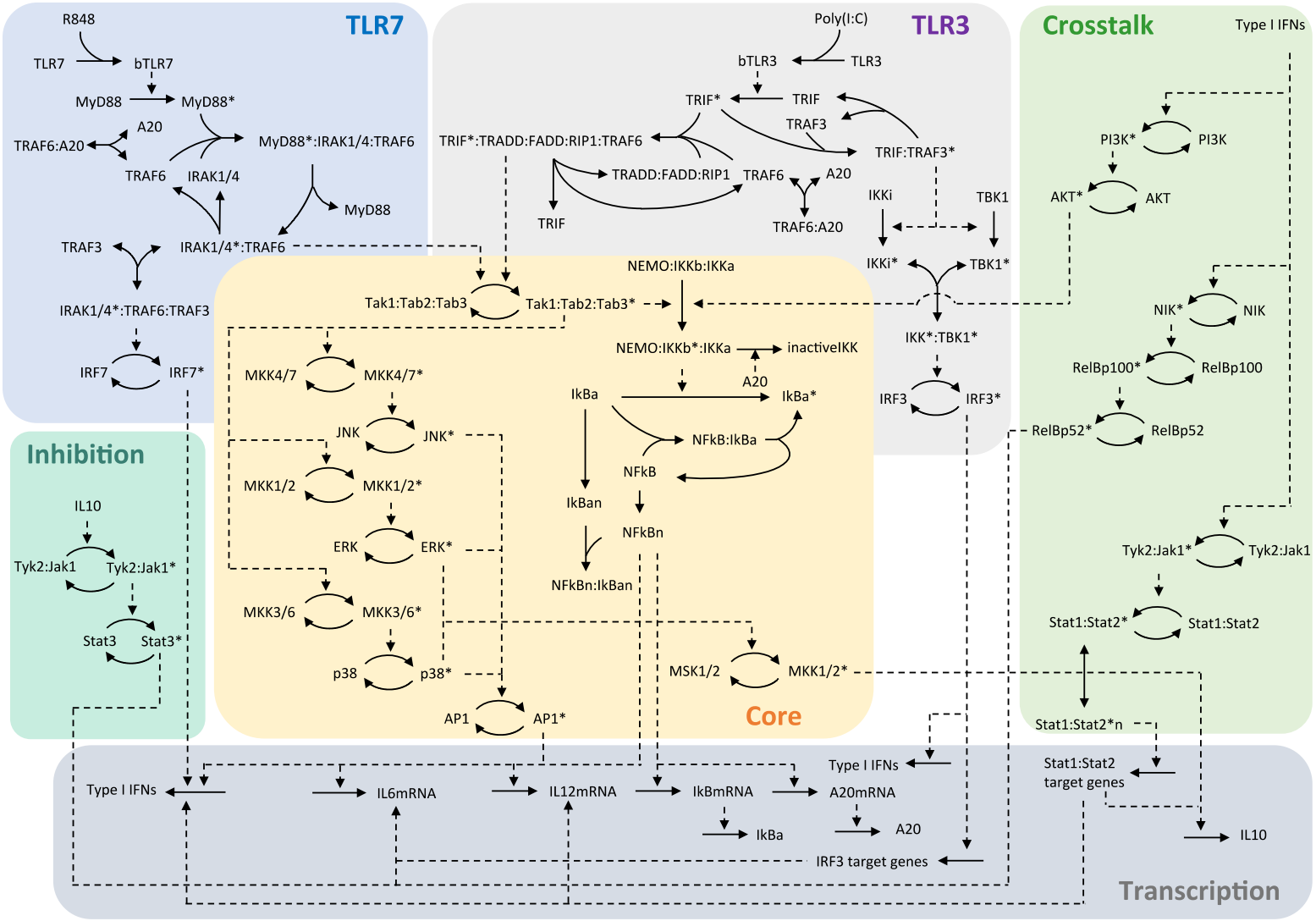


Training data

Test data

TLR3-TLR7 Pathways Modeling

- TLR3 activation followed by TLR7 activation leads to synergistic production of cytokines
- Investigated the cross talk mechanism causing this synergy

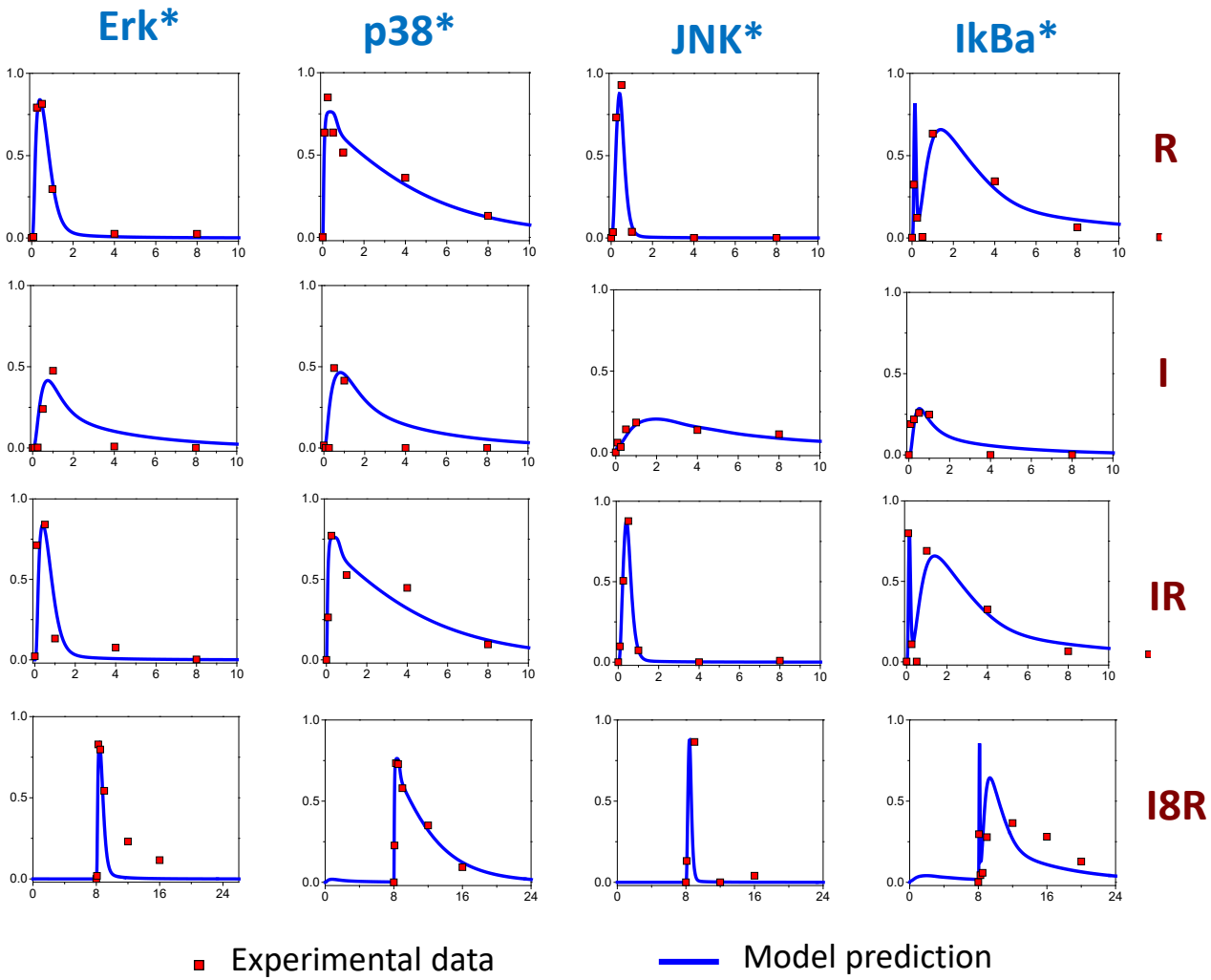


Liu, B., Liu, Q., Palaniappan, S., Bahar, I., Thiagarajan, P.S., Ding, J.L.: Innate Immune Memory and Homeostasis May Be Conferred Through TLR3-TLR7 Pathway Crosstalk. ***Sci. Signal.*** *9(436), ra70 (2016)*

Model Calibration using Training Data

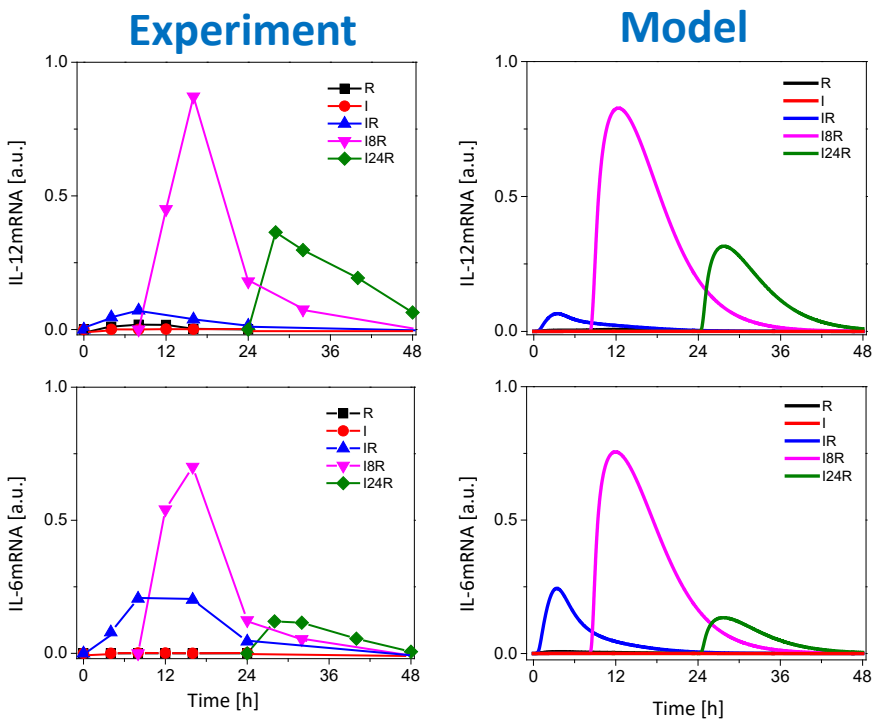
112 ODEs

129 unknown parameters



Model Calibration and Validation

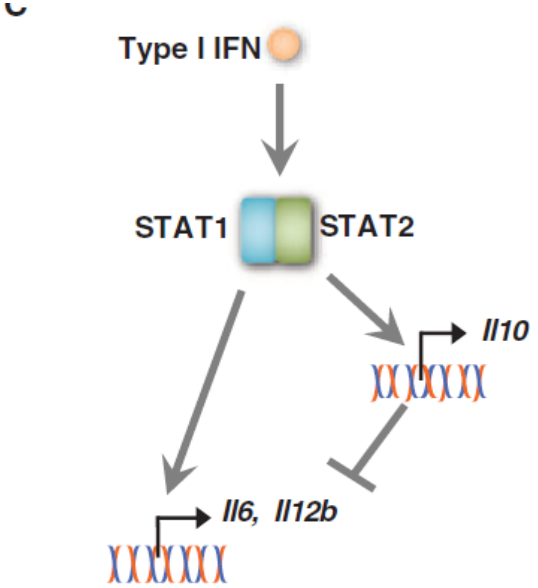
- Test data: [IL6mRNA], [IL12mRNA] at {0, 4, 8, 12, 16, 24, 28, 32, 40, 48 h}



The main findings

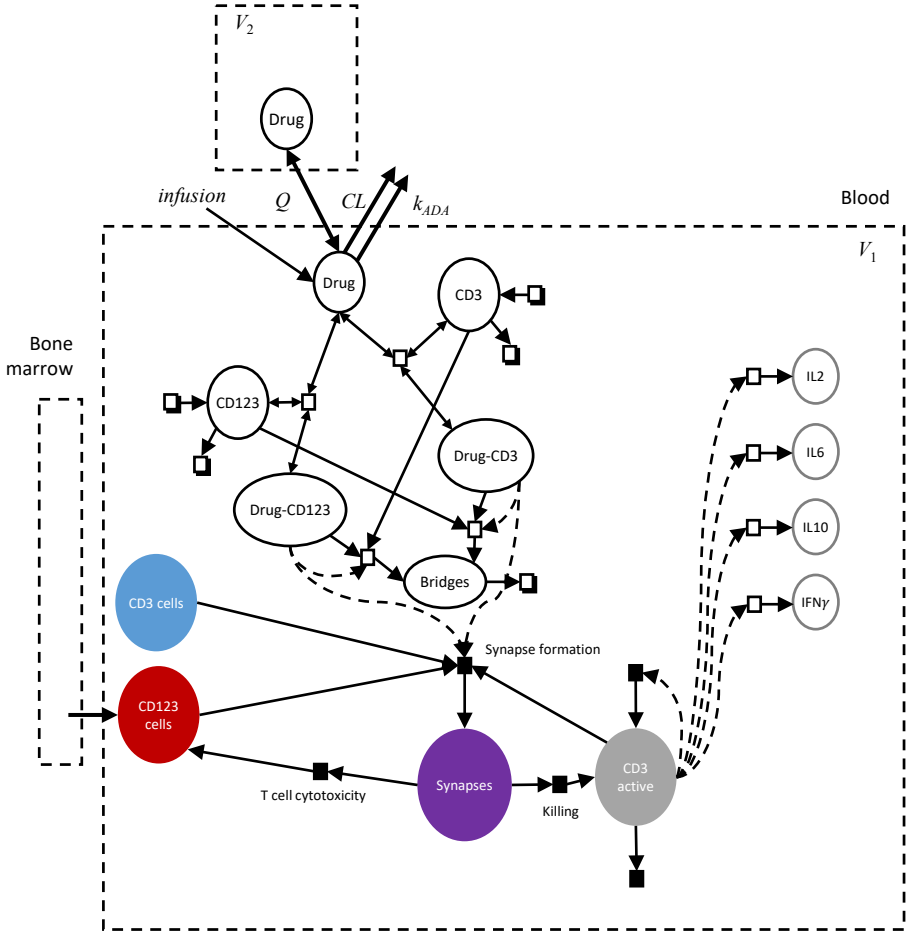
The JAK-STAT1/2 pathway is the main mechanism responsible for the induction of synergistic cytokine production

The cytokine response is biphasic due to an incoherent type I feedforward loop



QSP model of Sanofi's bispecific antibody

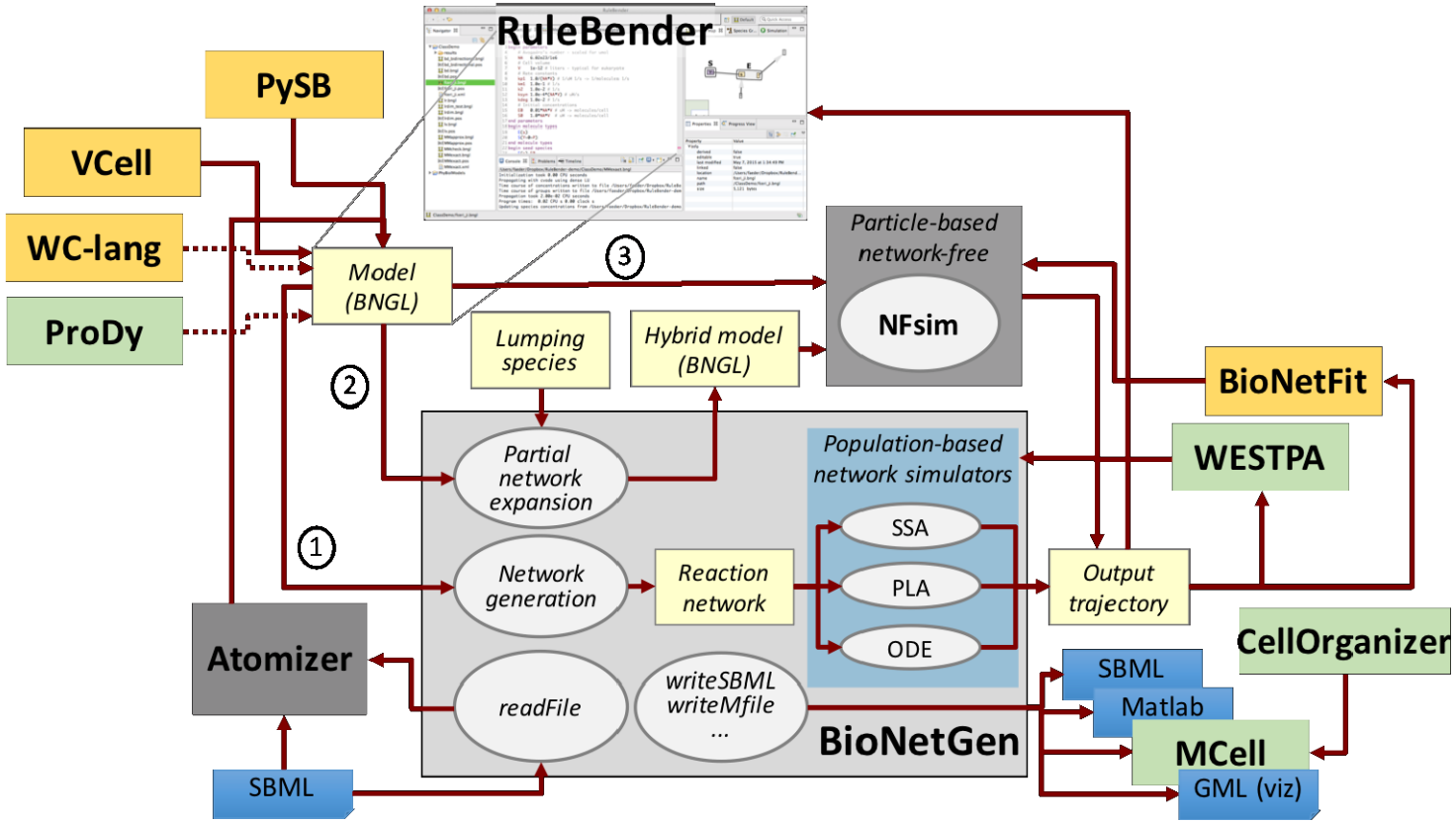
- SAR440234 is a bispecific antibody
  - capable of co-engaging the CD3 receptor on T cells and
  - the CD123 receptor
    - highly expressed on AML blasts
- Two level model to capture:
  - PK dynamics
  - Synapse formation
  - Killing of Cd123+ cells (AML blasts)
  - Cytokines release



Timothy R. Lezon, James R. Faeder

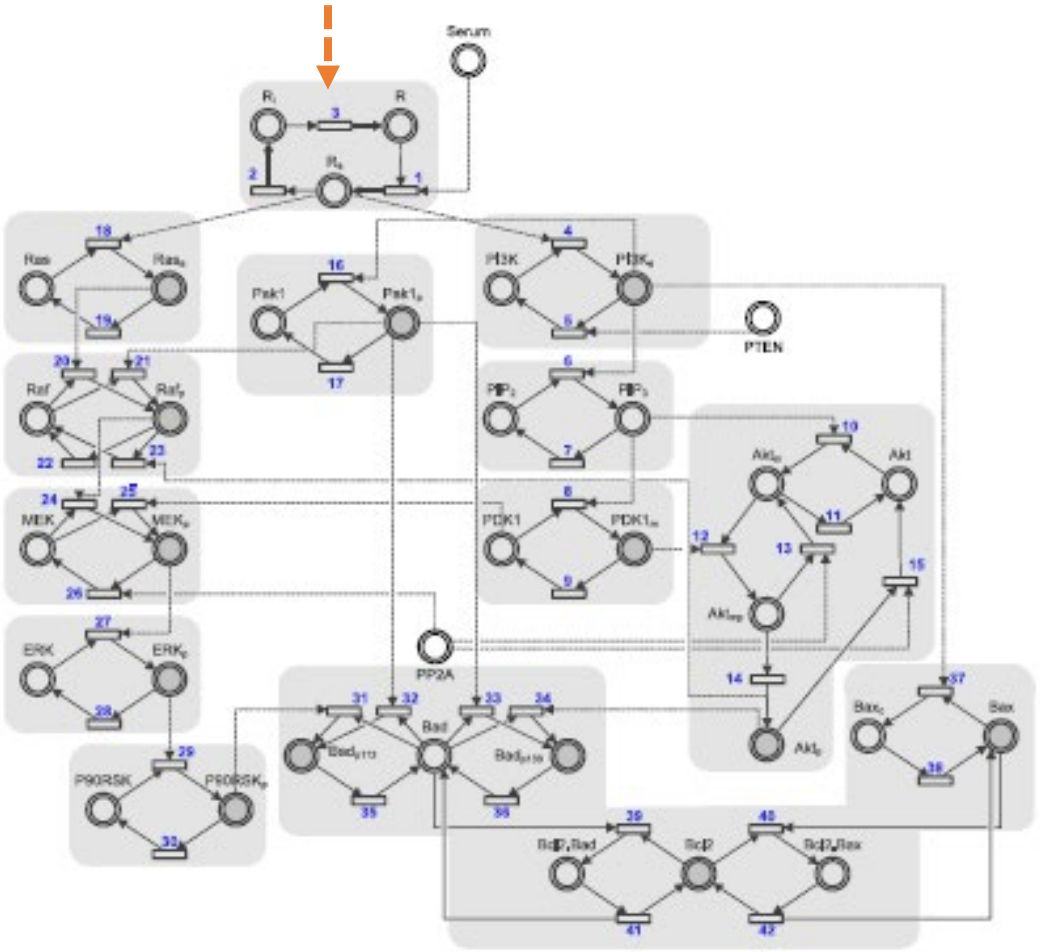
Going forward

- Add the SMC based method to the BioNetGen toolkit
- Current solutions: PTEMPEST, BioNetFit, SBML tools



JR Faeder, unpublished

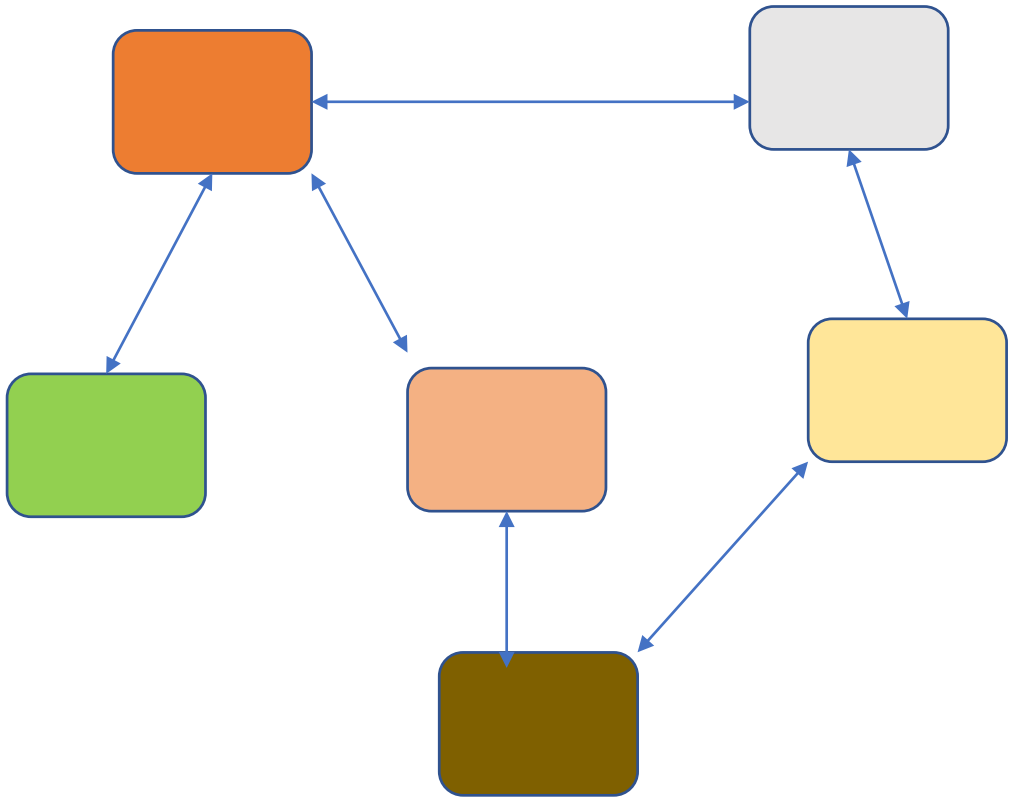
Going forward



The decomposed AKT-MAPK signaling pathway

- Decompositions based parameter estimation
- Decompose the model into its maximal strongly connected components
- Use the resulting DAG to guide the parameter estimation procedure.
- Estimate the parameters of the upstream components first
- Complications:
  - ❑ Distribution of experimental data
  - ❑ Computing consistent global estimates from local ones.
    - ➢ *Belief propagation*

Going forward

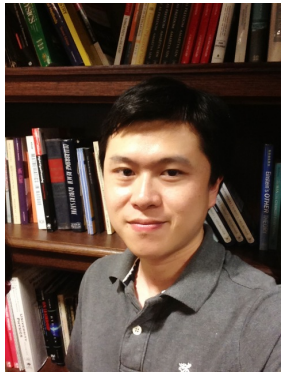


- Network based based parameter estimation
- Estimate the parameters of the components individually
- Compute consistent global estimates from local ones.
  - *Belief propagation*

Acknowledgements

Jim Faeder

Tim Lezon



Liu Bing

Benjamin Gyori
(LSP, Harvard Medical School)