Computational Modeling and Verification of Signaling Pathways in Cancer
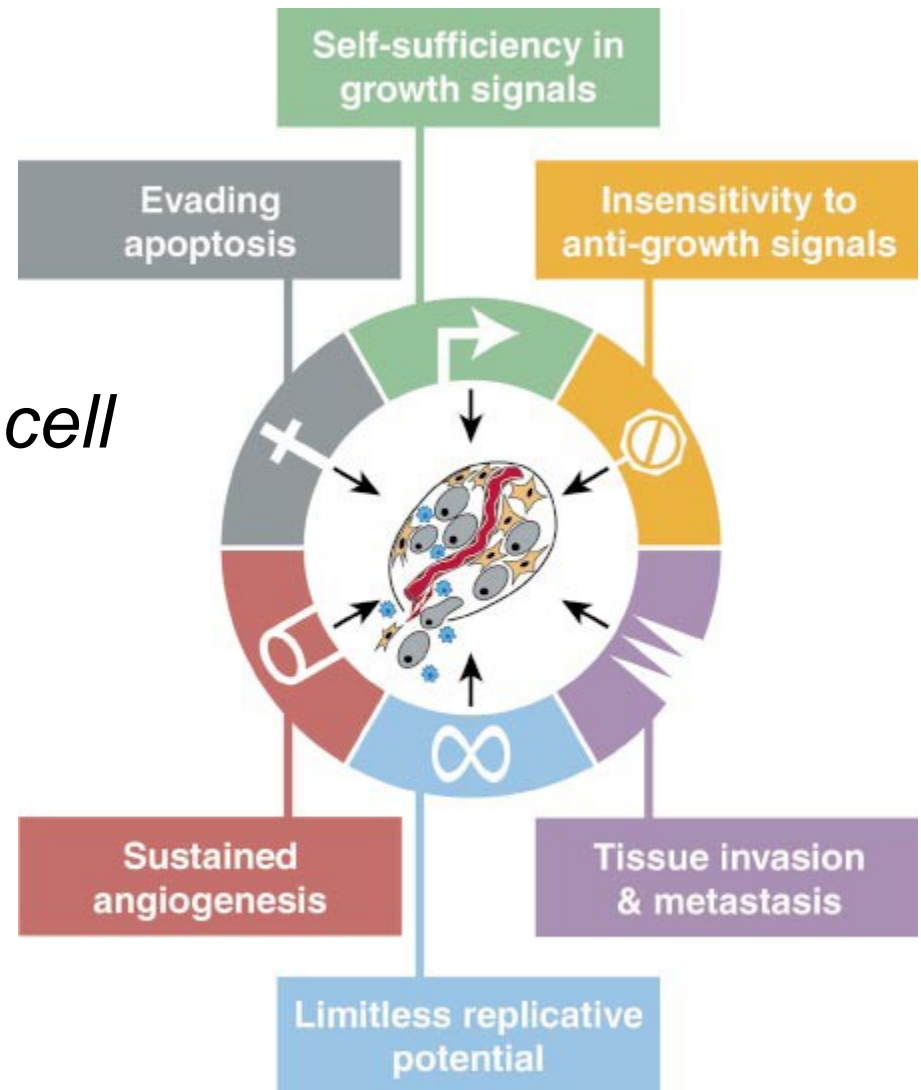
Haijun Gong*, Paolo Zuliani*, Anvesh Komuravelli*, James R. Faeder#, Edmund M. Clarke*

*Computer Science Department, Carnegie Mellon University

#School of Medicine, University of Pittsburgh

The Hallmarks of Cancer

"Six essential alterations in cell physiology that collectively dictate malignant growth."

The Hallmarks of Cancer

- All cancers share the six alterations.

- The way the alterations are acquired varies, both mechanistically and chronologically.

- Can we formalize the acquisition processes?

- Is there an "integrated circuit of the cell"?

A

Component	Acquired Capability	Example of Mechanism
	Self-sufficiency in growth signals	Activate H-Ras oncogene
	Insensitivity to anti-growth signals	Lose retinoblastoma suppressor
	Evading apoptosis	Produce IGF survival factors
	Limitless replicative potential	Turn on telomerase
	Sustained angiogenesis	Produce VEGF inducer
	Tissue invasion & metastasis	Inactivate E-cadherin

B
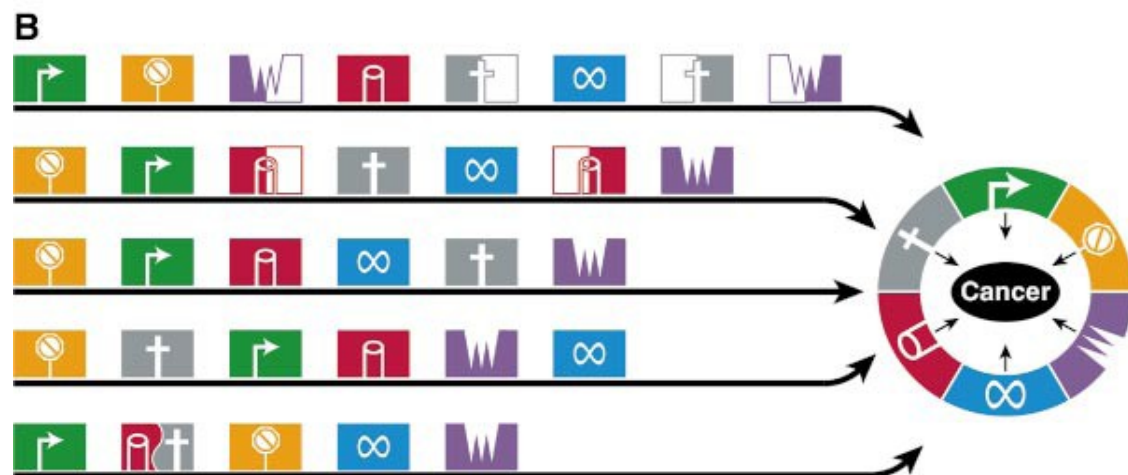
The Cell Integrated Circuit (?)
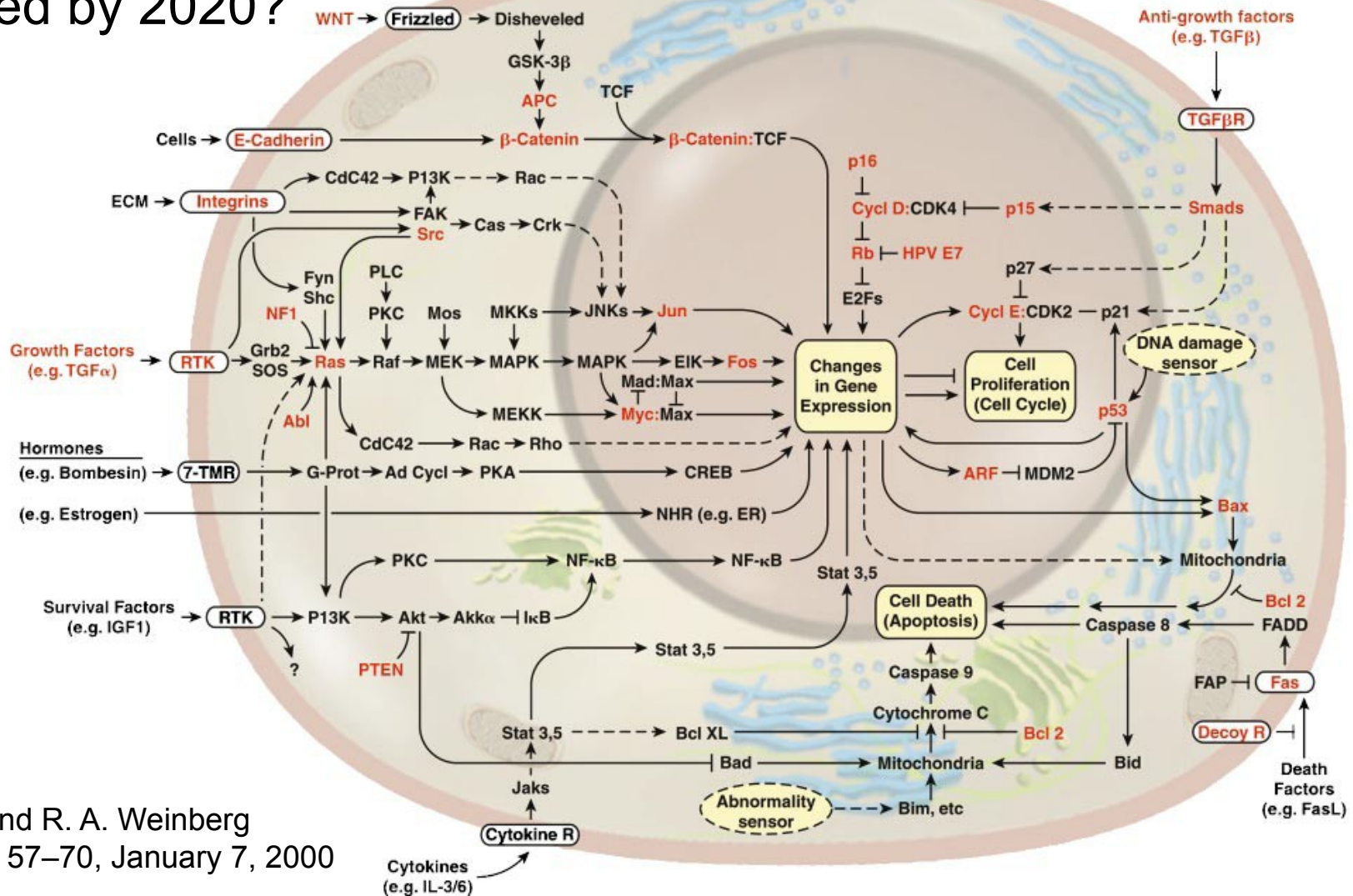
Completed by 2020?

This Work

- 2010: the "integrated circuit of the cell" still not in sight …

- But computational models can compare qualitatively well with experiments.

- We use the BioNetGen language (*http://bionetgen.org*) to describe signaling pathways important in many cancers:
  - We focus on the HMGB1 protein and the p53, NFkB, RAS and Rb signaling pathways

- We use statistical model checking to formally verify behavioral properties expressed in temporal logic:
  - Can express quantitative properties of systems
  - Scalable, can deal with large models

Signaling Pathways
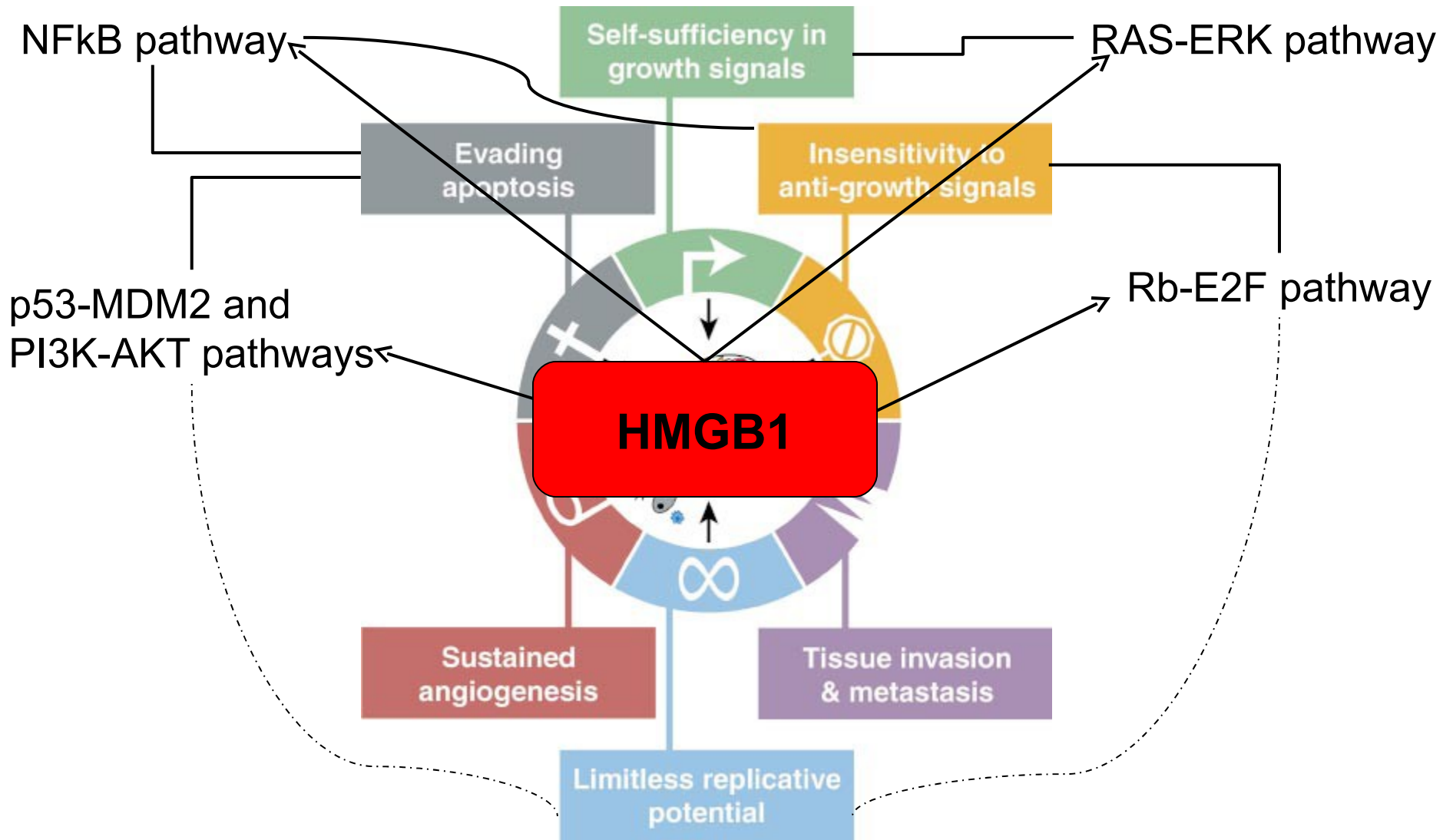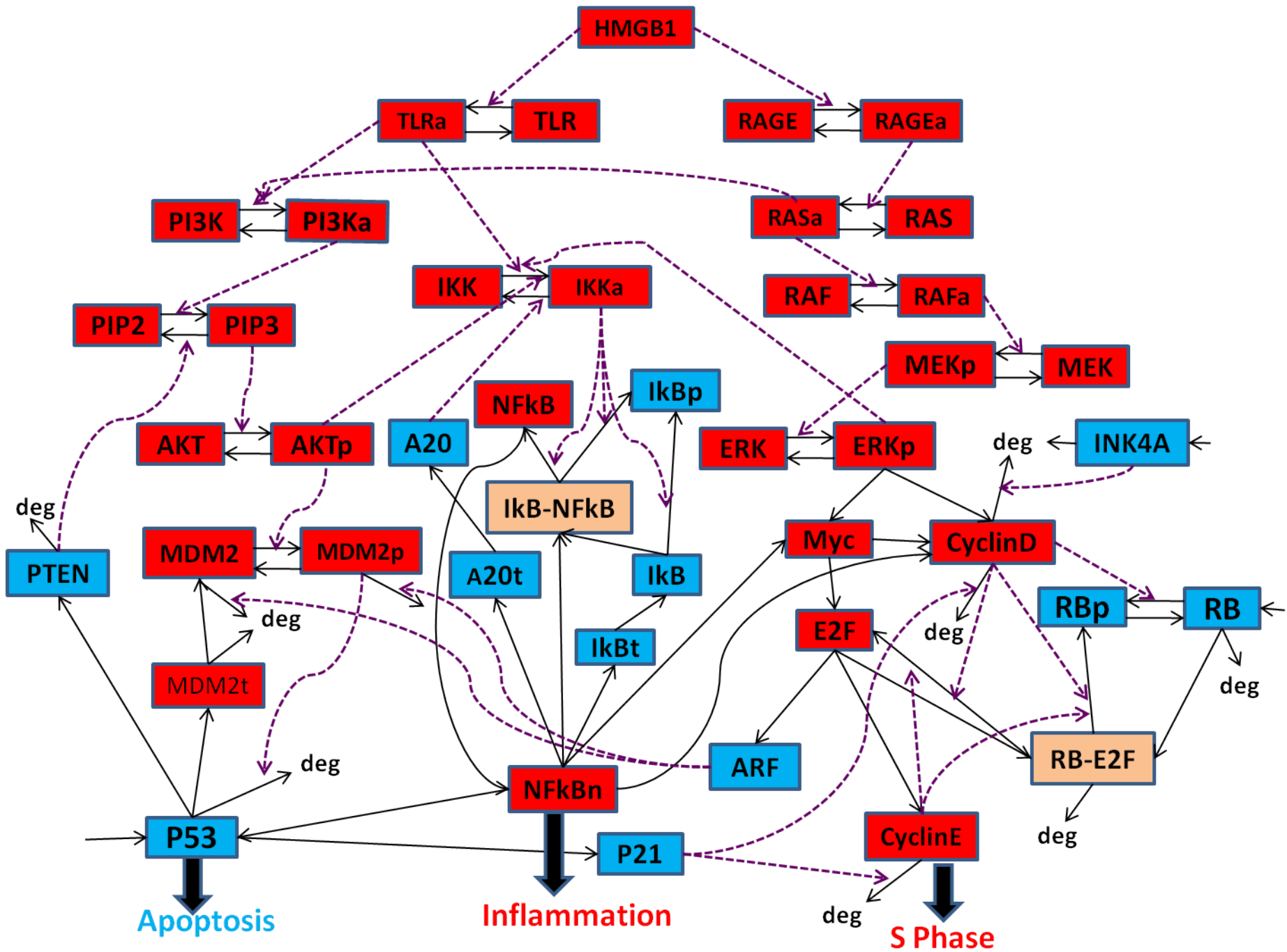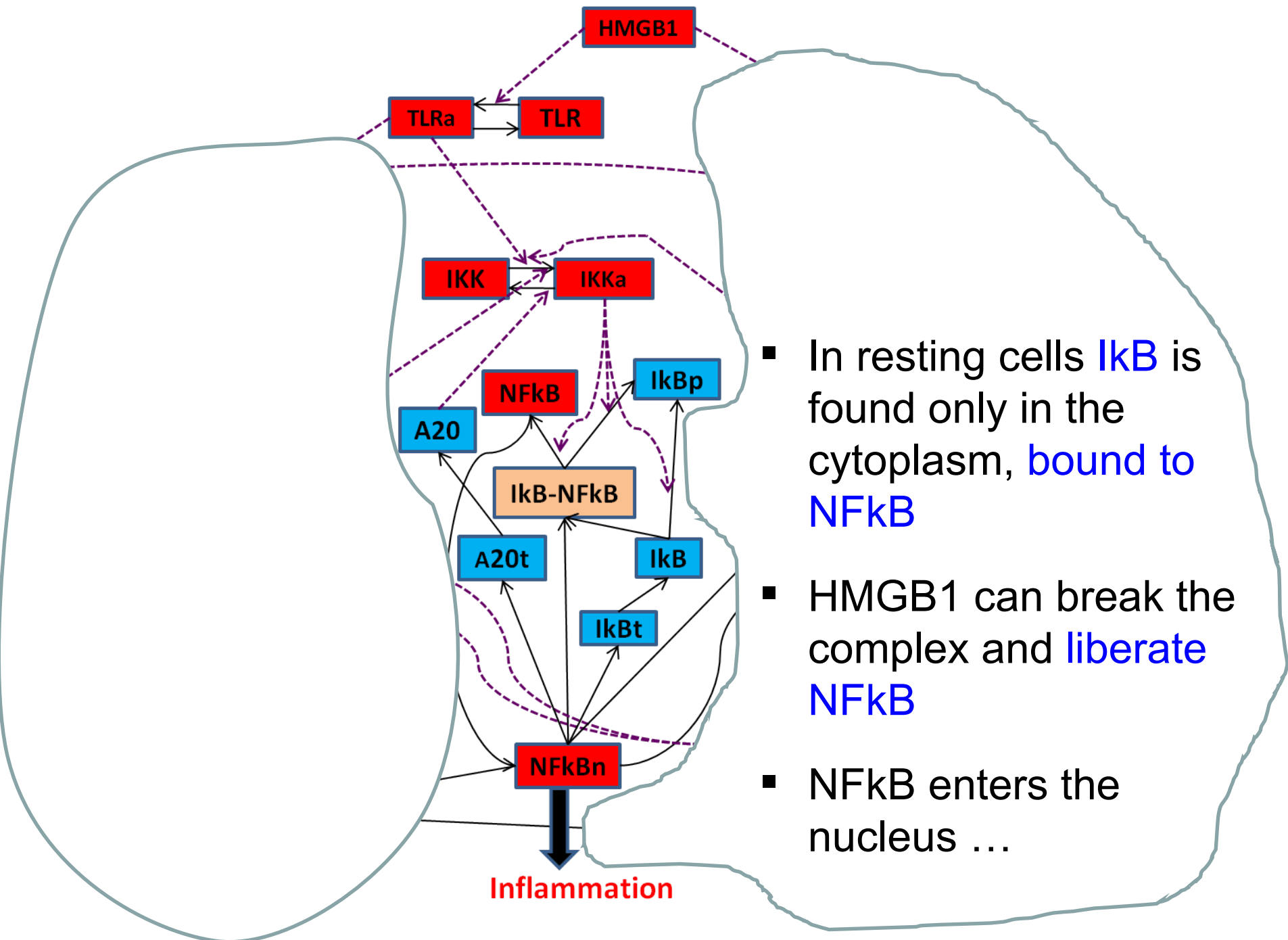
- In resting cells IkB is found only in the cytoplasm, bound to NFkB

- HMGB1 can break the complex and liberate NFkB

- NFkB enters the nucleus …

- The Rb-E2F pathway is important in the cell cycle

- It regulates the G1-S transition

- Rb keeps E2F in a complex

- HMGB1 can break it and liberate E2F

- E2F activates the transcription of CyclinE …
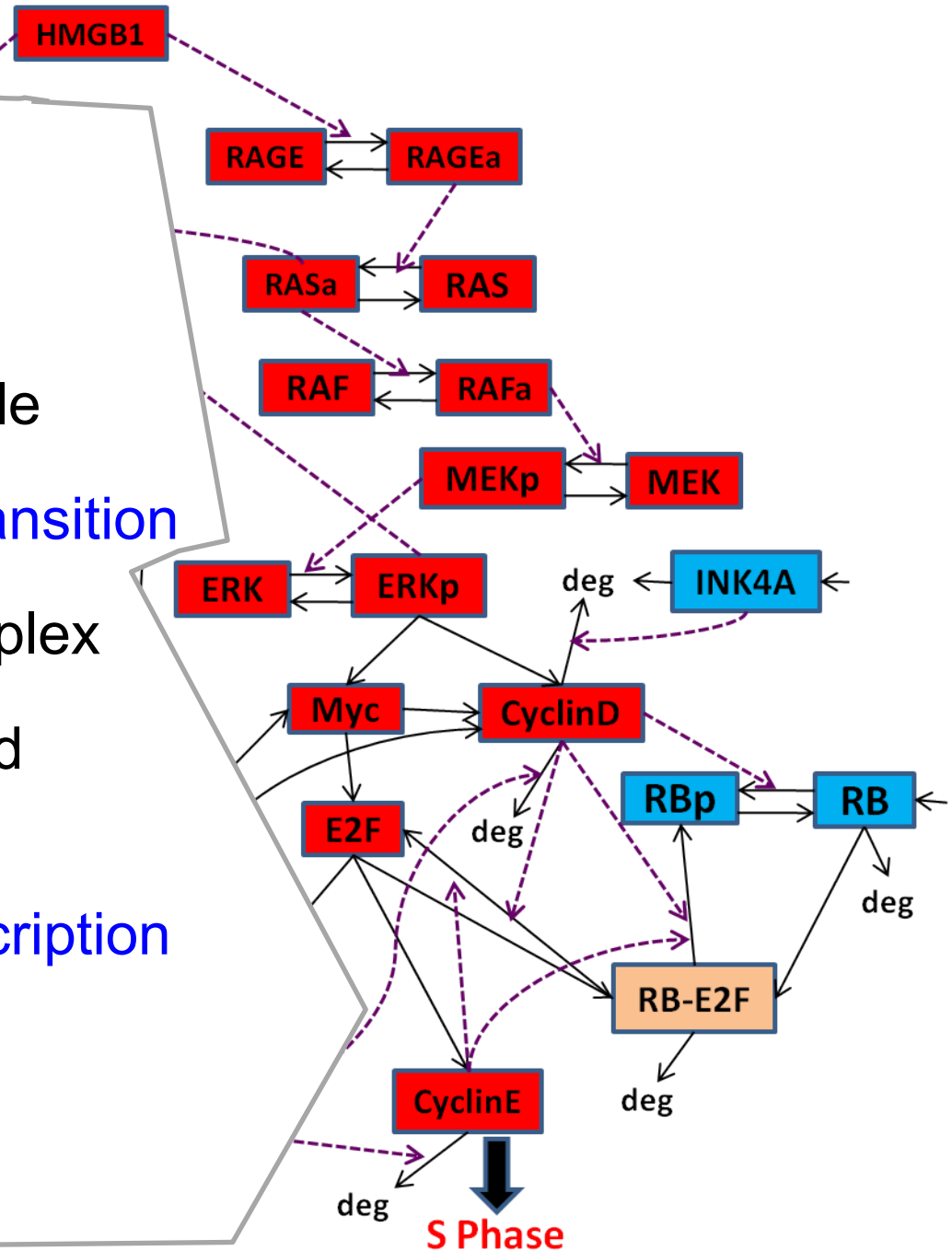
44 molecular species
82 reactions
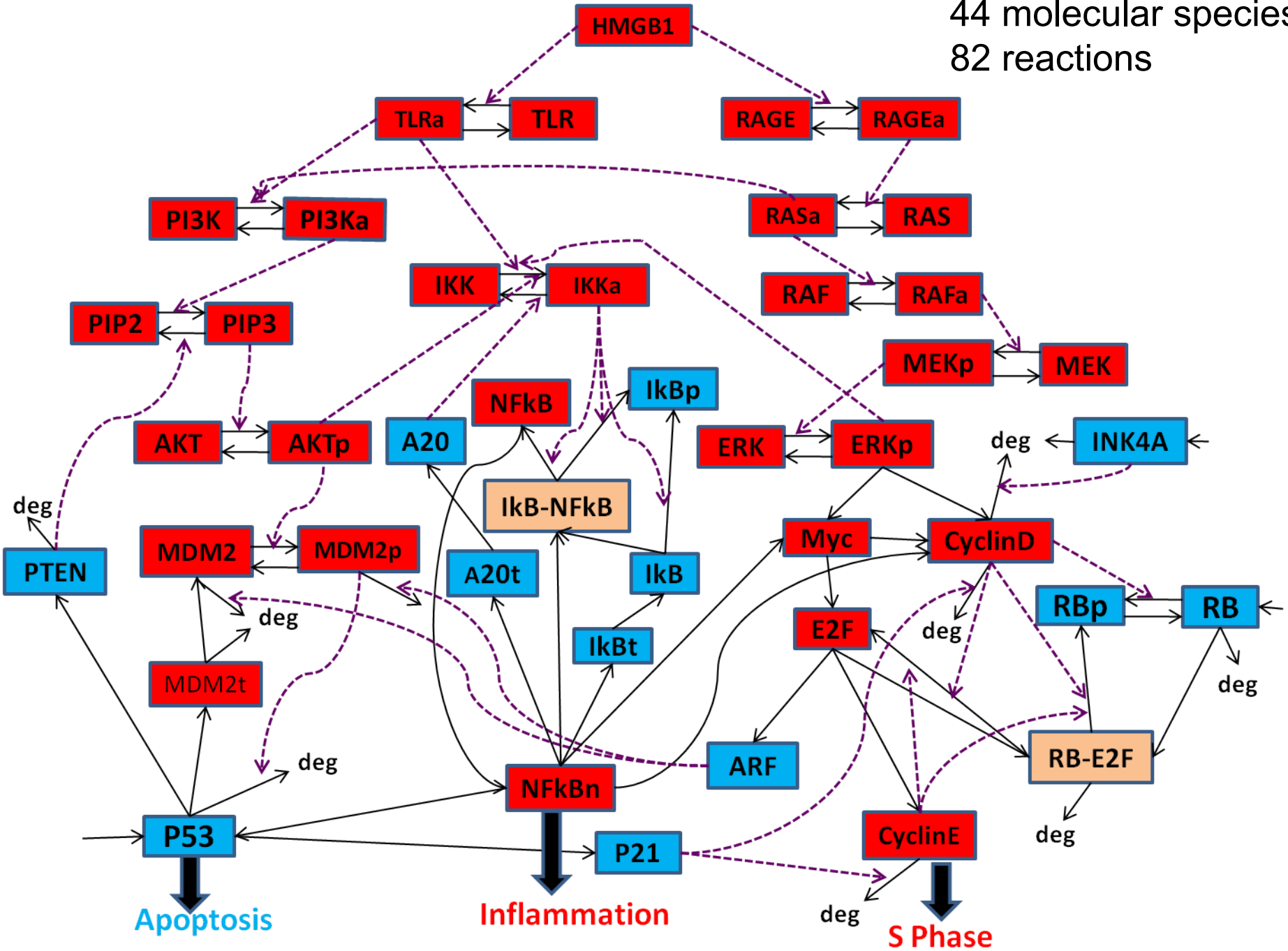
BioNetGen.org

- Rule-based modeling for biochemical systems

- Ordinary Differential Equations and Stochastic simulation (Gillespie's algorithm)

- *Example*: AKT has a component named d which can be labeled as U (unphosphorylated) or p (phosphorylated)

begin species

```
AKT(d~U)    1e5
AKT(d~p)    0
```

end species

begin parameters

```
k    1.2e-7
d    1.2e-2
```

end parameters

Faeder JR, Blinov ML, Hlavacek WS **Rule-Based Modeling of Biochemical Systems with BioNetGen.** In *Methods in Molecular Biology: Systems Biology*, (2009).

BioNetGen.org

- PIP3 can phosphorylate AKT, and dephosphorylation of AKT

begin reaction_rules

```
PIP(c~p) + AKT(d~U) → PIP(c~p) + AKT(d~p)     k

AKT(d~p) → AKT(d~U)                           d
```

end reaction_rules

- The corresponding ODE (assuming AKT+AKTp=const) is:

$$AKTp(t)' = k·PIP3(t)·AKT(t) - d·AKTp(t)$$

- The propensity functions for Gillespie's algorithm are:

$$k·[PIP(c~p)]·[AKT(d~U)]$$
$$d·[AKT(d~p)]$$

Verification of BioNetGen Models

- Temporal properties over the model's stochastic evolution

- For example: "does AKTp reach 4,000 within 20 minutes, with probability at least 0.99?"

- In our formalism, we write:

$$P_{\geq 0.99} (\mathbf{F}^{20} (AKTp \geq 4,000))$$

- For a property Φ and a fixed $0<\theta<1$, we ask whether

$$P_{\geq \theta} (\Phi) \quad \text{or} \quad P_{<\theta} (\Phi)$$

Equivalently

- A biased coin (Bernoulli random variable):
  - Prob (Head) = p Prob (Tail) = $1-p$
  - p is unknown

- Question: Is $p \geq \theta$? (for a fixed $0 < \theta < 1$)

- A solution: flip the coin a number of times, collect the outcomes, and use:
  - Statistical hypothesis testing: returns *yes/no*
  - Statistical estimation: returns "p in (a,b)" (and compare a with θ)

Statistical Model Checking

Key idea

- Suppose system behavior w.r.t. a (fixed) property Φ can be modeled by a Bernoulli random variable of parameter p:

  - System satisfies Φ with (unknown) probability p

- Question: $P_{\geq\theta}(\Phi)$? (for a fixed $0<\theta<1$)

- Draw a sample of system simulations and use:

  - Statistical hypothesis testing: Null vs. Alternative hypothesis

    $$H_0 : \mathcal{M} \models P_{\geqslant\theta}(\phi) \qquad H_1 : \mathcal{M} \models P_{<\theta}(\phi)$$

  - Statistical estimation: returns "p in (a,b)" (and compare a with θ)

Motivation

- Pros: Simulation is feasible for **many** systems

  - Often easier to simulate a complex system than to build the transition relation for it
  - Easier to parallelize

- Cons: answers may be wrong

  - But error probability can be bounded

Our Approach

Statistical Model Checking of biochemical models: $M \models P_{\geq \theta}(\Phi)$?
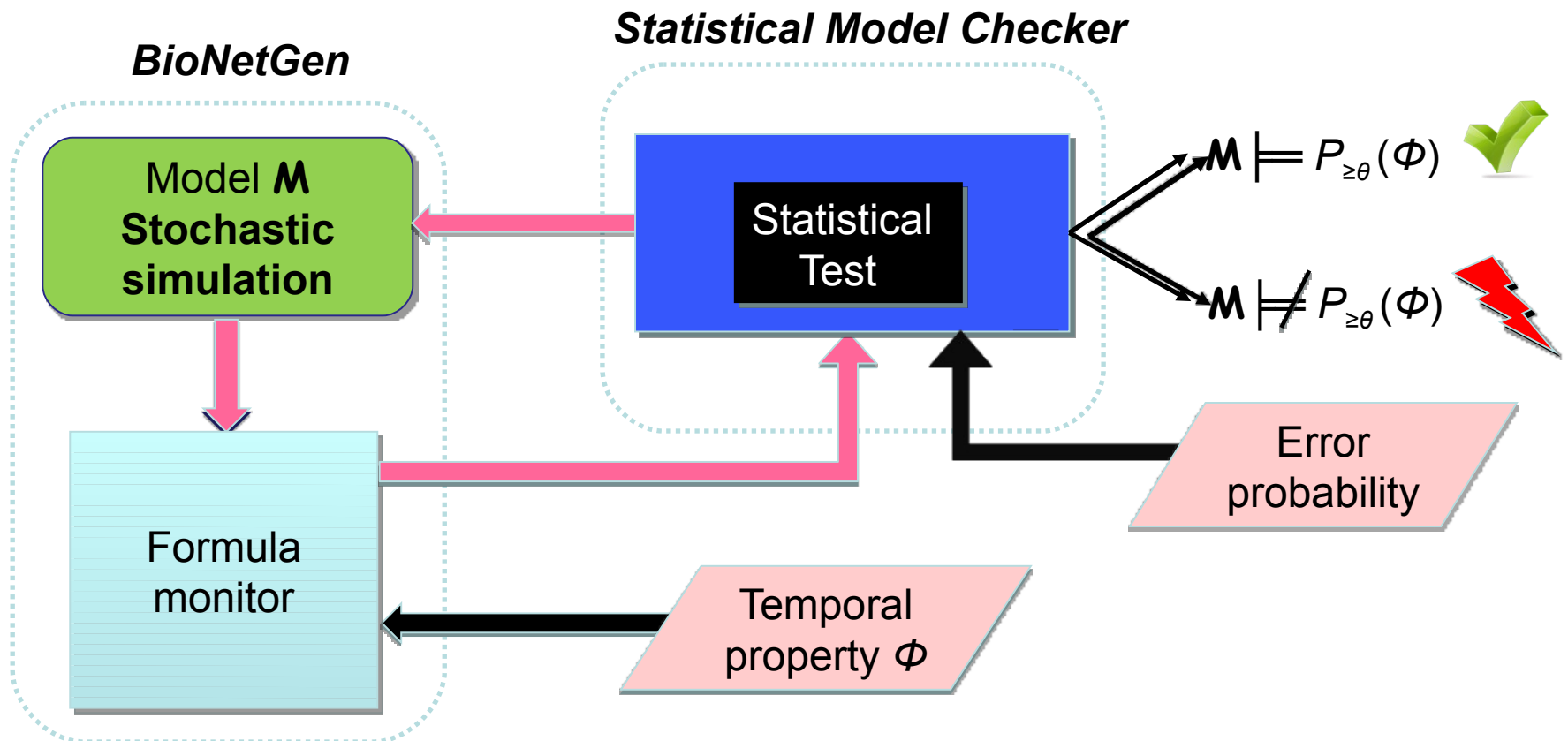
Sequential Bayesian Statistical MC - I

- $X = (x_1, \ldots, x_n)$ a sample of Bernoulli random variables

- Prior probabilities $P(H_0)$, $P(H_1)$ strictly positive, sum to 1

- Posterior probability (Bayes Theorem [1763])

$$P(H_0|X) = \frac{P(X|H_0)P(H_0)}{P(X)}$$

for $P(X) > 0$

- Ratio of Posterior Probabilities:

$$\frac{P(H_0|X)}{P(H_1|X)} = \frac{P(X|H_0)}{P(X|H_1)} \cdot \frac{P(H_0)}{P(H_1)}$$

Bayes Factor

Sequential Bayesian Statistical MC - II

- Recall the Bayes factor $B = \dfrac{P(X|H_0)}{P(X|H_1)}$

- Jeffreys' [1960s] suggested the Bayes factor as a statistic:
  - For fixed sample sizes
  - For example, a Bayes factor greater than 100 "strongly supports" H_0

- We introduce a sequential version of Jeffrey's test

- Fix threshold $T \geq 1$ and prior probability. Continue sampling until
    - Bayes Factor $> T$: Accept H_0
    - Bayes Factor $< 1/T$: Reject H_0

Sequential Bayesian Statistical MC - III

Require: *Property* $P_{\geq\theta}(\Phi)$*, Threshold* $T \geq 1$*, Prior density* g

$n := 0$ *{number of traces drawn so far}*

$x := 0$ *{number of traces satisfying Φ so far}*

repeat

     σ := draw a sample trace from BioNetGen (iid)

     $n := n + 1$

     if $\sigma \models \Phi$ **then**

       $x := x + 1$

     endif

     Β := *BayesFactor(n, x, θ, g)*

until (Β > T ∨ Β < $1/T$)

if (Β > T) **then**

            return "*H_0 accepted*"

else

            return "*H_0 rejected*"

endif

Correctness

Theorem (Termination). The Sequential Bayesian Statistical MC algorithm terminates with probability one.

Theorem (Error bounds). When the Bayesian algorithm – using threshold T – stops, the following holds:

Prob ("accept H_0" | H_1) $\leq 1/T$

Prob ("reject H_0" | H_0) $\leq 1/T$

Note: bounds independent from the prior distribution.

[Zuliani, Platzer, Clarke – HSCC 2010]

Bounded Linear Temporal Logic

- Bounded Linear Temporal Logic (BLTL): Extension of LTL with time bounds on temporal operators.

- Let $\sigma = (s_0, t_0), (s_1, t_1), \ldots$ be an execution of the model

    - along states $s_0, s_1, \ldots$

    - the system stays in state s_i *for time t_i*

    - divergence of time: $\Sigma_i\, t_i$ diverges (i.e., non-zeno)

- σ^i: Execution trace starting at state i.

- A model for BioNetGen simulation traces

Semantics of BLTL

The semantics of BLTL for a trace σ^k:

- $\sigma^k \models ap$ iff atomic proposition ap true in state s_k

- $\sigma^k \models \Phi_1 \vee \Phi_2$ iff $\sigma^k \models \Phi_1$ or $\sigma^k \models \Phi_2$

- $\sigma^k \models \neg\Phi$ iff $\sigma^k \models \Phi$ does not hold

- $\sigma^k \models \Phi_1 \cup^t \Phi_2$ iff there exists natural i such that

  1) $\sigma^{k+i} \models \Phi_2$

  2) $\Sigma_{j<i}\ t_{k+j} \leq t$

  3) for each $0 \leq j < i,\ \sigma^{k+j} \models \Phi_1$

  "within time $t,\ \Phi_2$ will be true and Φ_1 will hold until then"

- In particular, $F^t \Phi = true \cup^t \Phi$ $G^t \Phi = \neg F^t \neg\Phi$

Simulations

- Oscillations of NFkB and IKK in response to HMGB1 release: ODE vs stochastic simulation
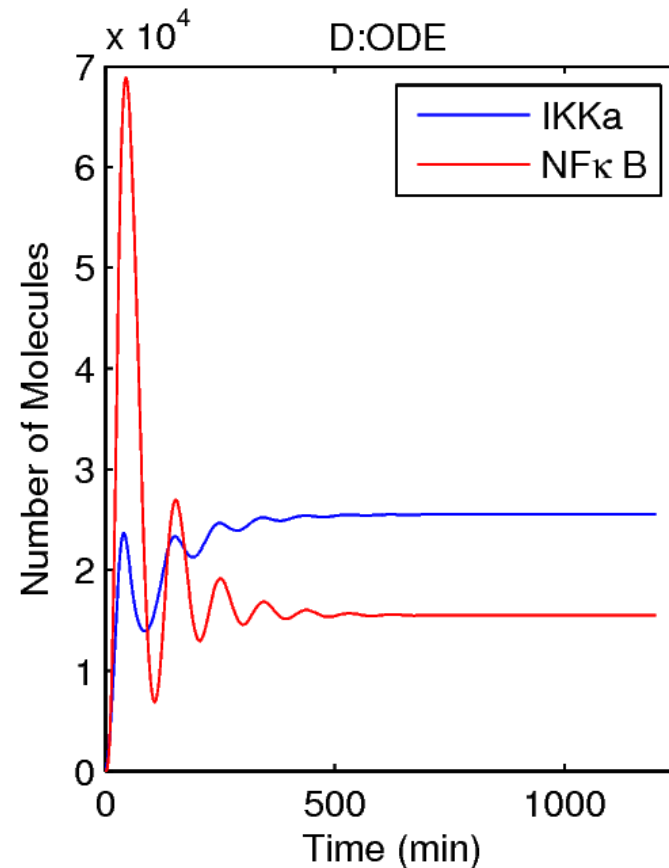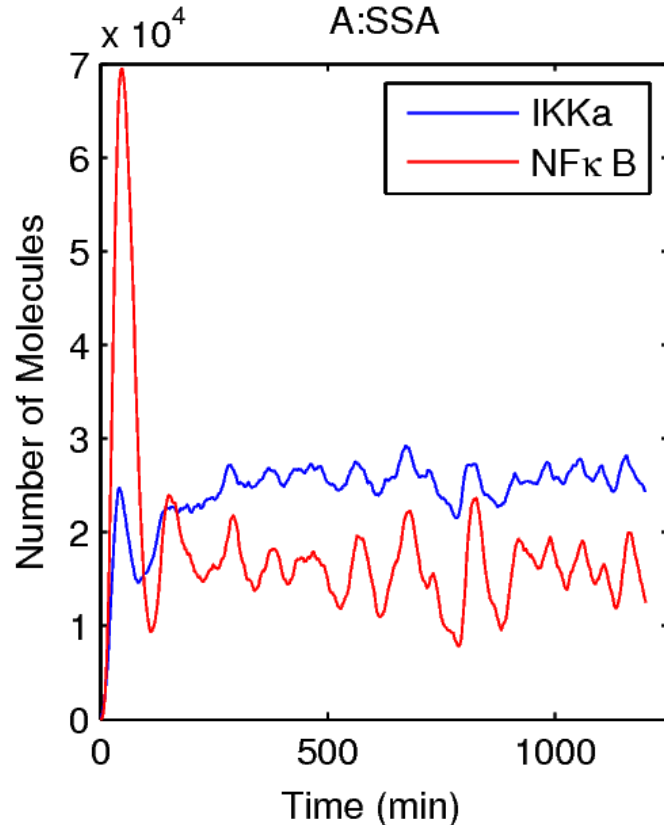
Verification

- Coding oscillations of NFkB in temporal logic

- Let R be the fraction of NFkB molecules in the nucleus

- We model checked the formula

$$P_{\geq 0.9} \; \mathbf{F}^t (R \geq 0.65 \; \& \; \mathbf{F}^t (R < 0.2 \; \& \; \mathbf{F}^t (R \geq 0.2 \; \& \; \mathbf{F}^t (R < 0.2))))$$

- The formula codes four changes in the value of R, which must happen in consecutive time intervals of maximum length t

- Note: the intervals need not be of the same length

Verification

- Statistical model checking

- T=1000, uniform prior, Intel Xeon 3.2GHz

$P_{\geq 0.9}$ $\mathbf{F}^t$ (R ≥ 0.65 & $\mathbf{F}^t$ (R < 0.2 & $\mathbf{F}^t$ (R ≥ 0.2 & $\mathbf{F}^t$ (R <0.2))))

HMGB1	t (min)	Samples	Result	Time (s)
10^2	45	13	False	76.77
10^2	60	22	True	111.76
10^2	75	104	True	728.65
10^5	30	4	False	5.76

Verification

- HMGB1 can activate PI3K, RAS and AKT in large quantities

- Let PI3Kr, RASr, and IKKr be the fraction of activated molecules of PI3K, RAS, and IKK, respectively

- We model checked the formula:

$$P_{\geq 0.9} \, \mathbf{F}^t \, \mathbf{G}^{180} \, (PI3Kr > 0.9 \ \& \ RASr > 0.8 \ \& \ IKKr > 0.6 \,)$$

t (min)	samples	result	time (s)
90	9	False	21.27
110	38	True	362.19
120	22	True	214.38

Conclusions

- Computational modeling is feasible for large models

- Temporal logic can be used to express behavioral properties

- Statistical Model Checking allows efficient and automatic verification of behavioral properties

- Modeling compares qualitatively well with experiments

- Further work:
  - parameter estimation
  - importance sampling
 - multi-scale systems

Acknowledgments

- This work supported by the NSF Expeditions in Computing program

- Thanks to Michael T. Lotze (University of Pittsburgh) for calling our attention to HMGB1

- Thanks to Marco E. Bianchi (Università San Raffaele) for discussions on HMGB1

Backup slides

The Cell Cycle

- G_0: resting, non-proliferating state

- G_1: cell is active and continuously growing, but no DNA replication

- S (synthesis): DNA replication

- G_2: continue cell growth and synthesize proteins

- M (mitosis): cell divides into two cells

The Biology of Cancer. R. A. Weinberg, 2006.
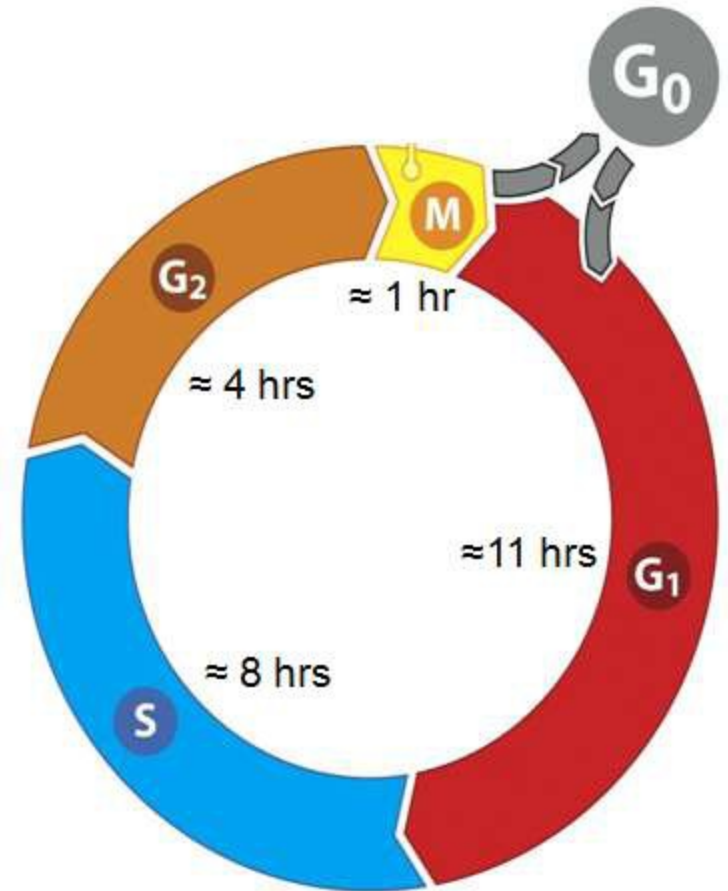
Bayesian Statistics

Three ingredients:

1. Prior probability
   - Models our initial (a priori) uncertainty/belief about parameters (what is Prob($p \geq \theta$) ?)

1. Likelihood function
   - Describes the distribution of data (*e.g.*, a sequence of heads/tails), given a specific parameter value

1. Bayes Theorem
   - Revises uncertainty upon experimental data - compute Prob($p \geq \theta \mid data$)

Sequential Bayesian Statistical MC

- Model Checking $H_0 : \mathcal{M} \models P_{\geqslant \theta}(\phi)$ $H_1 : \mathcal{M} \models P_{< \theta}(\phi)$

- Suppose $\mathcal{M}$ satisfies ϕ with (unknown) probability p
  - p is given by a random variable (defined on [0,1]) with density g
  - g represents the prior belief that $\mathcal{M}$ satisfies ϕ

- Generate independent and identically distributed (iid) sample traces.

- x_i: the ith sample trace σ satisfies ϕ
  - $x_i = 1$ iff $\sigma_i \models \phi$
  - $x_i = 0$ iff $\sigma_i \not\models \phi$

- Then, x_i will be a Bernoulli trial with conditional density (likelihood function)
$$f(x_i|u) = u^{x_i}(1 - u)^{1-x_i}$$

Computing the Bayes Factor - I

Definition: Bayes Factor of sample X and hypotheses H_0, H_1 is

joint (conditional) density of independent samples

$$\frac{P(H_0|X)}{P(H_1|X)} \cdot \frac{P(H_1)}{P(H_0)} = \frac{\int_\theta^1 f(x_1|u)\cdots f(x_n|u)\cdot g(u)\ du}{\int_0^\theta f(x_1|u)\cdots f(x_n|u)\cdot g(u)\ du} \cdot \frac{1-\pi_0}{\pi_0}$$

- $\pi_0 = P(H_0) = \int_\theta^1 g(u)du$ prior g is Beta of parameters $\alpha>0$, $\beta>0$

$$g(u) = \frac{1}{B(\alpha,\beta)} u^{\alpha-1}(1-u)^{\beta-1}$$

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\ dt$$

Computing the Bayes Factor - II

Proposition

The Bayes factor of $H_0 : \mathbf{M} \models P_{\geq \theta}(\Phi)$ vs $H_1 : \mathbf{M} \models P_{< \theta}(\Phi)$ for n Bernoulli samples (with $x \leq n$ successes) and prior Beta(α,β)

$$B = \frac{1 - \pi_0}{\pi_0} \cdot \left(\frac{1}{F_{(x+\alpha,\, n-x+\beta)}(\theta)} - 1 \right)$$

where $F_{(\cdot,\cdot)}(\cdot)$ is the Beta distribution function.

$$F_{(x+\alpha,\, n-x+\beta)}(\theta) = \frac{1}{B(x+\alpha,\, n-x+\beta)} \int_0^{\theta} u^{x+\alpha-1} (1-u)^{n-x+\beta-1} \, du$$

- No need of integration when computing the Bayes factor