A Collaborative Proposal to the NSF Experimental Expeditions Program

Computational Biology of Cancer

Bud Mishra
(Courant Institute, NYU)
Our Vision

To gain fundamental new insights into the emergent behaviors of complex biological and embedded systems through the use of revolutionary, highly scalable, and fully automated modeling and analysis techniques.
Primary Challenge: Scalability

Key Scalability Issues:

- Spatial Distribution
- Stochastic Behavior
- Highly Nonlinear Behavior
- Mixed (Hybrid) Continuous-Discrete Behavior
- Vast Numbers of System State Variables & Components

Complex Biological & Embedded Systems can exhibit any combination of these features
Pancreatic Cancer

- 4th leading cause of cancer death in the US and Europe
- Five-year survival rate is only 4%
- Almost no progress in diagnosis and treatment in the past 40 years

Healthy and diseased pancreas cells

New insights into the dynamics of these deadly diseases are urgently needed!
Why Pancreatic Cancer?

- No animal model, so computational models are needed
- Signaling models from cancer experts at TGen (Translational Genomics)
- We will build new analysis and verification tools
- TGen collaborators will use tools to better understand cancer dynamics
Model Checking

The **Model Checking Problem**: Let $M$ be a state-transition graph. Let $f$ be a formula of temporal logic. 

- **Example formula**: $a \ U \ b$ means “$a$ holds true Until $b$ becomes true”

Does $f$ hold along all paths that start at the initial state of $M$?
Biological Models of Cancer

- Cancer as a disease of the genome…
- Cancer as a somatic evolutionary process…
- Cancer as a price of symbiosis (mitochondrial)…
- Cancer as a response to multi-cellularity…
- Cancer as a price of repair/regeneration (stem cells)…
- Cancer as a consequence of energy consumption (glucose metabolism)…
- Cancer as a response to external stress…
- Cancer as a response to the micro-environment (hyper- and hypo-methylation)…
Relevant Biological Processes

- Proliferation:
  - Oncogenes and TumorSuppressor Genes

- Differentiation:
  - Stem Cells…

- Signaling:
  - Kinases…

- Maintenance and Immortality:
  - Autophagy, Necrosis and Apoptosis
War on Cancer

“… as we know, there are known knowns; there are things we know we know.

“We also know there are known unknowns; that is to say we know there are some things we do not know.

“But there are also unknown unknowns – the ones we don't know we don't know.”

– Ex-US Secretary of Defense, Mr. Donald Rumsfeld, Quoted completely out of context.
Known Known Biology

- Theory: “World Where There Are Names for Everything.”
“Addicted to Death”

- Cancer is a progressive switch from apoptotic (scheduled) to necrotic (unscheduled) tumor cell death.
- The immunobiology of many intracellular factors are involved:
  - the products of purine metabolism (uric acid, ATP, and adenosine);
  - the nuclear protein HMGB1; the S100 family members; the heat shock proteins;
- Cancer is the consequence of disordered tumor cell death rather than cell growth
  - Loss of homeostasis
  - A condition called "addicted to death."
Purine Metabolism

- Purine Metabolism
  - Provides the organism with building blocks for the synthesis of DNA and RNA.

- The entire pathway is almost closed but also quite complex. It contains
  - several feedback loops,
  - cross-activations and
  - reversible reactions
Figure 10.2. Second model of purine metabolism without representation of activations and inhibitors (see text for details).
The main metabolite in purine biosynthesis is 5-phosphoribosyl-a-1-pyrophosphate (PRPP).
- A linear cascade of reactions converts PRPP into inosine monophosphate (IMP).
- IMP is transformed into AMP and GMP.
- Guanosine, adenosine and their derivatives are recycled (unless used elsewhere) into hypoxanthine (HX) and xanthine (XA).
- XA is finally oxidized into uric acid (UA).
Purine Metabolism
Variation of the initial concentration of PRPP does not change the steady state. \((\text{PRPP} = 10 \times \text{PRPP1})\) implies steady_state()

Persistent increase in the initial concentration of PRPP does cause unwanted changes in the steady state values of some metabolites.

If the increase in the level of PRPP is in the order of 70% then the system does reach a steady state, and we expect to see increases in the levels of IMP and of the hypoxanthine pool in a “comparable” order of magnitude.

Always \((\text{PRPP} = 1.7 \times \text{PRPP1})\) implies steady_state()
Consider the following statement:

Eventually
(Always (PRPP = 1.7 * PRPP1) implies steady_state())
and Eventually
Always(IMP < 2* IMP1))
and Eventually (Always (hx_pool < 10*hx_pool1)))

where IMP1 and hx_pool1 are the values observed in the unmodified trace.

The model checker determines that the above statement is false.

Counter-example: Model checker shows that the increase in IMP is about 6.5 fold while the hypoxanthine pool increase is about 60 fold.

The model “over-predicts” the increases in products by amounts that are physiologically impossible…

The model should therefore be amended
XS-Systems:
(AAMC M. et al. 2001-2009)

Canonical Form:

\[
\begin{align*}
\dot{X}_i &= \alpha_i \prod_{j=1}^{n-m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \quad i = 1 \ldots n \\
C_l(X_1(t), \ldots, X_{n+m}(t)) &= \sum (\gamma_l \prod_{j=1}^{n+m} X_j^{f_{ij}}) = 0
\end{align*}
\]

Characteristics:
- Predefined Modular Structure
- Automated Translation from Graphical to Mathematical Model
- Scalability
Rule-based modeling protocol

1. Define components as *structured objects* and interactions as *rules*.

2. Determine **concentrations** and **rate constants**

3. Generate and simulate the model.
The activation of Casp9 needs APAF1 and cytochrome c.

DNA damage → nucleus

Mitochondria

Cytochrome c

dATP

Casp9

Active holoenzyme

Pro-casp3

Active casp3

Cleave downstream cellular proteins to kill cell

DEVD-Afc

Active site

nucleus damage
xS-System Model
Simpathica recapitulate the holoenzyme formation process

The Simulation of Apoptotic Holoenzyme Kinetics

Rodriguez and Lazebnik (1999)
Decreasing [APAF-1] Kill Caspase Activity

- Mix RNAi(APAF-1) treated and untreated E1A cell extract.

- Measure the caspase 3 activity of the mixed cell extract.

![Graph showing the relationship between DEVD-afc rate and Percentage of IMR 90/E1A.](image)

**Expected**

- APAF1 extract with little APAF1.

**IMR90 primary**

- IMR90/E1A siRNA(APAF1) and IMR90/E1A mix

**αAPAF-1**

**αCasp9**
Where to modify the model in Simpathica?

Reversibility?

Linear

Non-Linear?

Linear
- **Recombinant system:**
  - cytochrome c, caspase-9, APAF1

- **Purification of endogenous APAF1/cytc oligomer**

[Diagram showing the interaction between APAF1, cytochrome c, and caspase-9]

Cell extract with little casp9

Adding back

\[[\text{APAF1/cytc/dATP}]\rightarrow \text{caspase3 activity}\]

Linear dependence?
**XS-Systems:**
(AAMC M. et al. 2001-2009)

**Canonical Form:**

\[
\begin{align*}
\dot{X}_i &= \alpha_i \prod_{j=1}^{n+m} X_j^{\gamma_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{\delta_{ij}} \quad i = 1 \ldots n \\
C_i(X_1(t), \ldots, X_{n+m}(t)) &= \sum (\gamma_i \prod_{j=1}^{n+m} X_j^{f_{ij}}) = 0
\end{align*}
\]

**Characteristics:**

- Predefined Modular Structure
- Automated Translation from Graphical to Mathematical Model
- Scalability

**Figure 1:** Representation of an unmodified and of a reversible reaction.

**Figure 2:** Representation of a divergence and of a convergence branch point (the two processes in each reaction are independent of each other).

**Figure 3:** Representation of a single splitting reaction generating two products, \(X_2\) and \(X_3\), in stoichiometric proportions and of a single synthetic reaction involving two source components, \(X_1\) and \(X_2\) always in stoichiometric proportions.

**Figure 4:** The conversion of \(X_1\) into \(X_2\) is modulated (stimulation or inhibition is represented by the sign of the arrow) by \(X_3\). The reaction between \(X_1\) and \(X_2\) requires coenzyme \(X_3\), which in the process is converted into \(X_4\).
Definition 1 (S-system). An S-system is a quadruple $S = (DV, IV, DE, C)$ where:

- $DV = \{X_1, \ldots, X_n\}$ is a finite non empty set of dependent variables ranging over the domains $D_1, \ldots, D_n$, respectively;
- $IV = \{X_{n+1}, \ldots, X_{n+m}\}$ is a finite set of independent variables ranging over the domains $D_{n+1}, \ldots, D_{n+m}$, respectively;
- $DE$ is a set of differential equations, one for each dependent variable, of the form
  \[
  \dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}}
  \]
  with $\alpha_i, \beta_i \geq 0$ called rate constants;
- $C$ is a set of algebraic constraints of the form
  \[
  C_j(X_1, \ldots, X_{n+m}) = \sum (\gamma_j \prod_{k=1}^{n+m} X_k^{f_{jk}}) = 0
  \]
  with $\gamma_j$ called rate constraints.
Verifying temporal properties of a reactive system

Step 1. Formally encode the behavior of the system as a semi-algebraic hybrid automaton

Step 2. Formally encode the properties of interest in TCTL

Step 3. Automate the process of checking if the formal model of the system satisfies the formally encoded properties using quantifier elimination
Solution

- Bounded Model Checking
- Constrained Systems
  - Linear Systems
  - O-minimal
  - SACoRe (Semi algebraic Constrained Reset)
  - IDA
Subway Map of Cancer
Is this View of Cancer Necessarily Accurate?

- “If I said yes, that would then suggest that that might be the only place where it might be done which would not be accurate, necessarily accurate.

- “It might also not be inaccurate, but I'm disinclined to mislead anyone.”

  – Ex-US Secretary of Defense, Mr. Donald Rumsfeld, Once again quoted completely out of context.
Known Unknown Biology

- Reality: “World Where There Are No Names of Anything.”
The New Synthesis

Genome Evolution

Genotype

DNA

RNA

Protein

Transcription

Translation

Phenotype

Selection

MCAI 2.0
Cancer Initiation and Progression

Mutations, Translocations, Amplifications, Deletions
Epigenomics (Hyper & Hypomethylation)
Alternate Splicing

Proliferation, Motility, Immortality, Metastasis, Signaling, Microenvironment (autophagy)
Amplifications & Deletions

Tumor suppressor genes: APC, DCC, p53
Oncogene: ras

Mutation in a TSG
Epigenomics
Conversion of a Proto-Oncogene
Deletion of a TSG
Deletion of a TSG

Normal colon cells
APC gene loss
Increased cell growth
DNA hypomethylation
Adenoma class I
ras gene mutation
Adenoma class II
DCC gene loss
Adenoma class III
p53 gene loss
Carcinoma
Other gene losses
Metastasis
Karyotyping
Microarray Analysis of Cancer Genome

- Representations are reproducible samplings of DNA populations in which the resulting DNA has a reduced complexity.
  - Array probes derived from low complexity representations of the normal genome
  - We measure differences in gene copy number between normal and tumor samples ratiometrically
Daruwala et al. (PNAS, 2004)
Allelic Frequencies: Cancer & Normal
(Anantharaman et al. unpublished)
Some tumor-specific conditions (e.g., hypoxia, low pH and low level of glucose) commonly cause the glucose-regulated stress response of cancer cells.

One can induce various stress responses in cancer cells artificially, and study them experimentally.

For example, Tunicamycin induces (glycosylation) stress:

– It blocks the synthesis of all N-linked glycoproteins (N-glycans)
– And causes cell cycle arrest in G1 phase.
Proprietary experimental results removed.
PolyA cDNAs in solution

Shear Flow

Restriction Enzyme ex. ‘Rsa I’

cDNAs ‘coded’
ex. ‘GTAC’

0010100000100010

AFM Imaging

Image Processing, Pattern Matching

Reed, et al., Nanotechnology, 2007
Single Molecule Restriction Map

Sample Input

Cell isolation
RNA extraction
2 hours

Reverse Transcription
2 hours

Fix to cDNA Surface
Enzyme digestion
30 min

Scan Molecules
<12 hours

Readout

Microfluidic Device + Fast AFM
AFM vs Sequence

337 bp
2,366 bp
1,320 bp
357 bp

870 x 1,500 nm

AFM vs Seq.

MCAI 2.0
Identify and Count

10 x 10 micron
Histogram of Transcript Abundance

Transcript Counts

Counts

Transcript
Models that are Concepty

- “I’m not into this detail stuff.
- “I'm more concepty.”
  – Ex-US Secretary of Defense, Mr. Donald Rumsfeld, Once again quoted completely out of context.
GOALIE: GO Algorithmic Logic for Invariant Extraction

GO categories describing genes in "source" cluster

GO categories describing "destination" cluster but not "source"

GO categories describing "source" cluster but not "destination"
Unknown Unknown Biology
Pathologist’s View

Healthy and diseased pancreas cells
A Challenge

“At present, description of a recently diagnosed tumor in terms of its underlying genetic lesions remains a distant prospect. Nonetheless, we look ahead 10 or 20 years to the time when the diagnosis of all somatically acquired lesions present in a tumor cell genome will become a routine procedure.”

– Douglas Hanahan and Robert Weinberg

“I would not say that the future is necessarily less predictable than the past. I think the past was not predictable when it started.”

– Ex-US Secretary of Defense, Mr. Donald Rumsfeld.
Foci

- Measurements
  - Single Cell Single Molecule Experiments
- Modeling & Model Checking
  - Phenomenological & Mechanistic Models
- Mining
  - Hypotheses
- Manipulation
  - Diagnostics and Therapeutics
Translational Systems Biology


Models of Apoptosis

Model Checking in Biology


Algorithmic Algebraic Model Checking


Optical Mapping

Copy Number Fluctuations

Single Molecule/Single Cell Nanotechnology


Answer to Cancer

“If I know the answer I'll tell you the answer, and if I don't, I'll just respond, cleverly.”

– Ex-US Secretary of Defense, Mr. Donald Rumsfeld.
The end