Synthesis of Biological Models from Mutation Experiments

Ali Sinan Köksal, Saurabh Srivastava, Rastislav Bodík, UC Berkeley
Evan Pu, MIT
Jasmin Fisher, Microsoft Research Cambridge
Nir Piterman, University of Leicester

Overview

Concurrent program synthesis from examples

Programs ≡ biological explanations

Examples ≡ biological experiments

We assist natural sciences with formal methods

- Given experiments, are there other explanations?
- If so, compute a new, disambiguating experiment
- This avoids conducting superfluous experiments

This talk: how stem cells coordinate their fates

Understanding Diseases

- "Cancer is fundamentally a disease of failure of regulation of tissue growth. In order for a normal cell to transform into a cancer cell, the genes which regulate cell growth and differentiation must be altered." – from Wikipedia

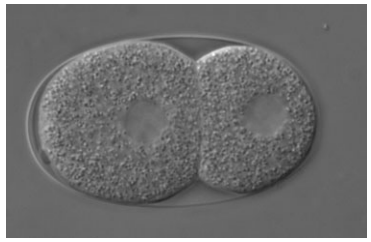
- Research on cell differentiation helps understanding diseases such as cancer.

C. elegans: A Model Organism

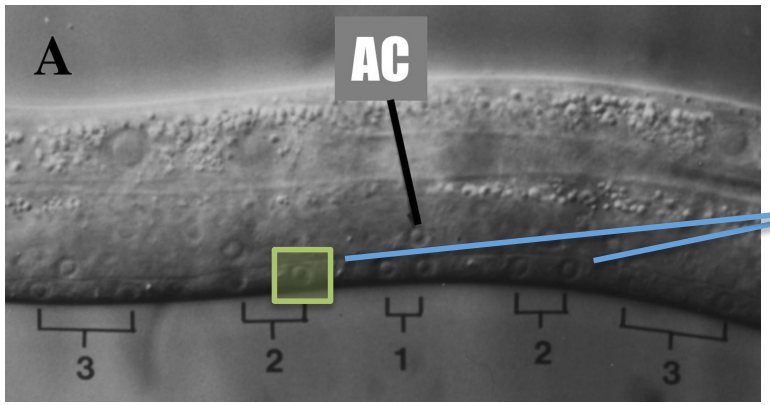
Earthworm used in developmental biology.

959 cells; its organs found in other animals.

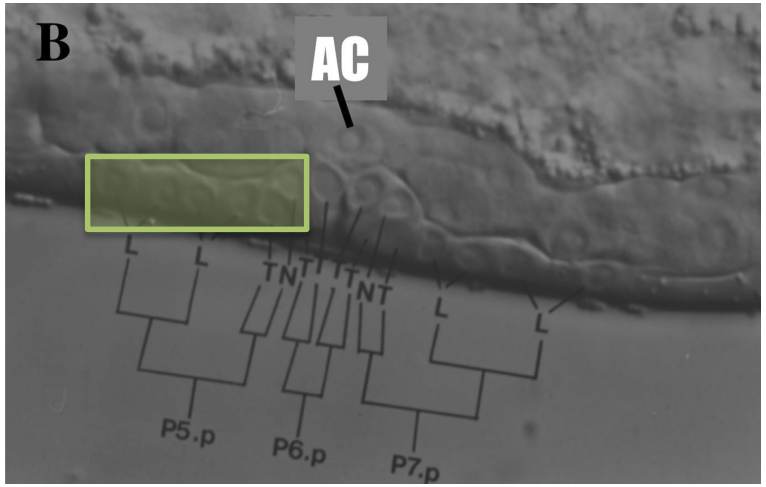
Differentiation studied on vulval development.



Initial division of embryo



Identical precursor cells collaborate to decide their fate



Differentiation and then development into organ parts

Research Goal of Biologists

What is the mechanism (program) within each cell for deciding fates through communication?

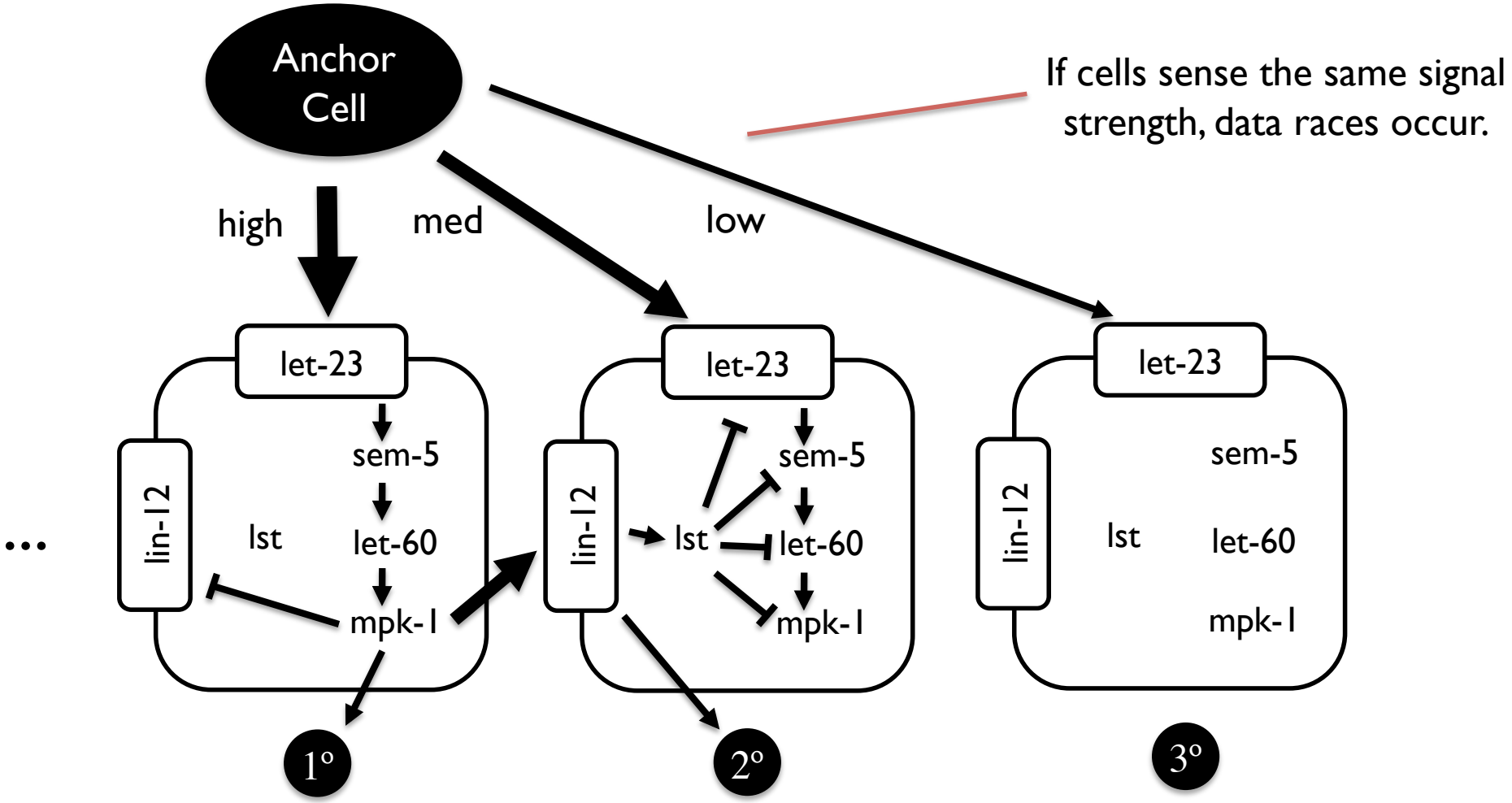
Building Blocks of these Programs

Cells contain communicating proteins.

Protein interaction: a protein senses the concentration of other proteins.

Interaction is either activation or inhibition.

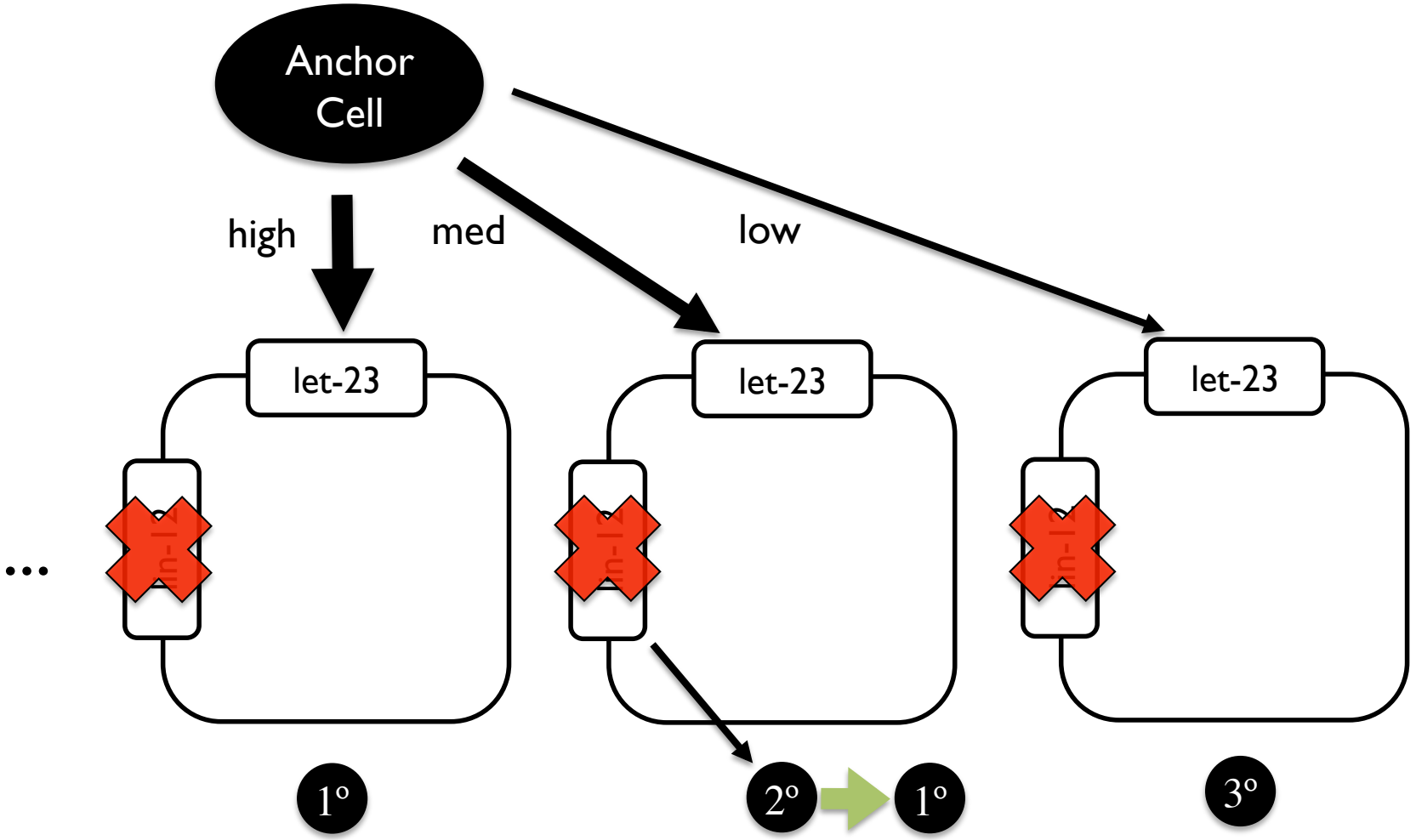
How the Vulval Cells Differentiate



How Biologists Discover Interactions

- Measuring protein levels over time is infeasible.
- If such "cell tracing" is infeasible, infer protein interaction from end-to-end experiments.
- That is, mutate cells → observe resulting fates.
- Mutation experiments change protein behavior in a controlled way:
  - Enable a protein via gene overexpression.
  - Disable a protein via gene suppression.

A Mutation Experiment



Putting Experiments Together

No protein is mutated.

Fate of six neighboring cells

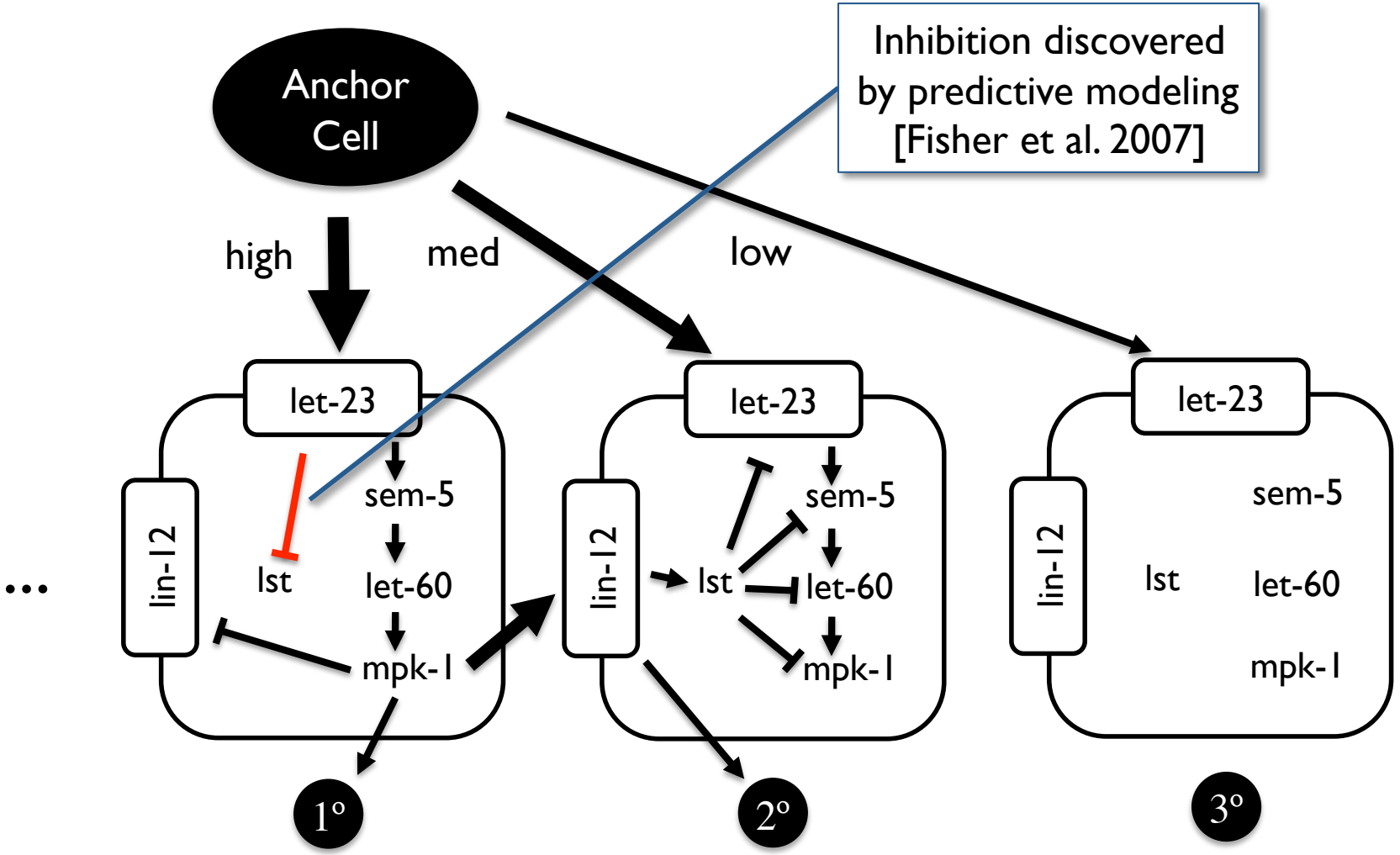
Experiment	AC	lin-12	lin-15	let-23	lst	Fate decisions
1	ON	ON	ON	ON	ON	{332123}
2	ON	OFF	ON	ON	ON	{331113}
3	ON	ON	OFF	ON	ON	{112121,122121, 212121}
...						
48	...					

lin-12 is turned off.

Multiple outcomes observed

Experiments over 35 years by 11 groups

How to Build Accurate Models?

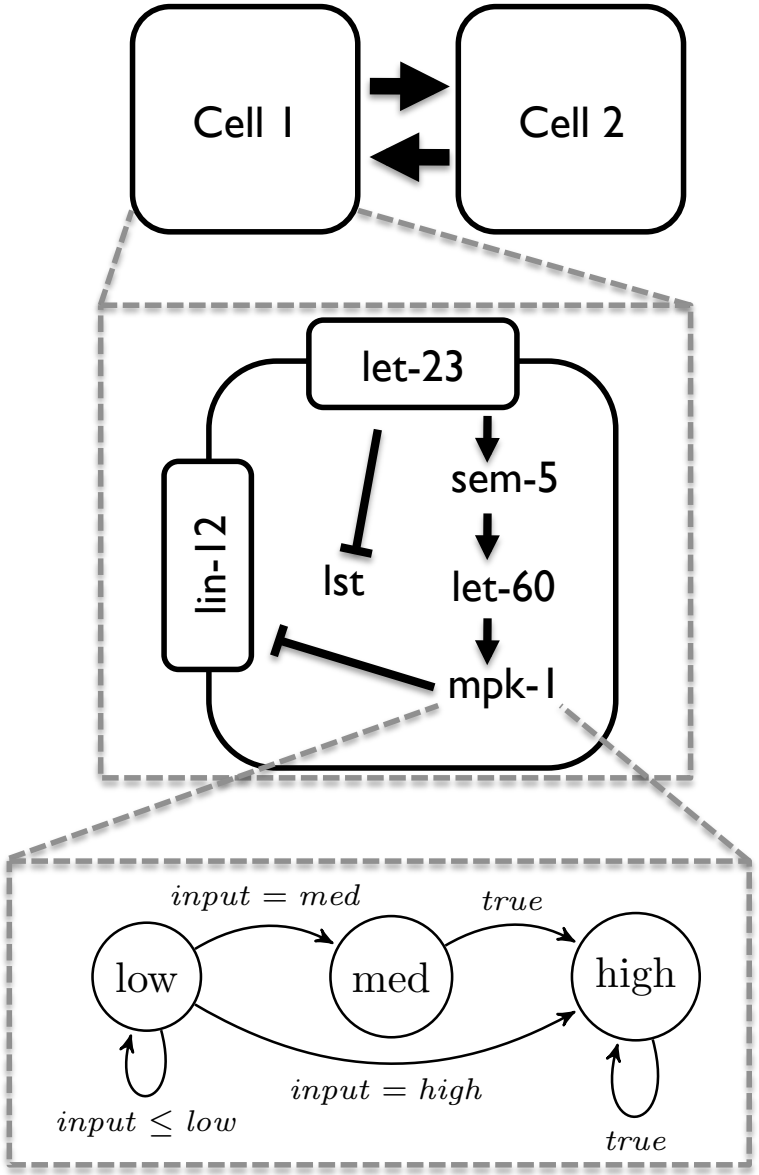


Semantics of the Modeling Language

- Program has cells
- Non-deterministic outcomes via **schedule** interleaving

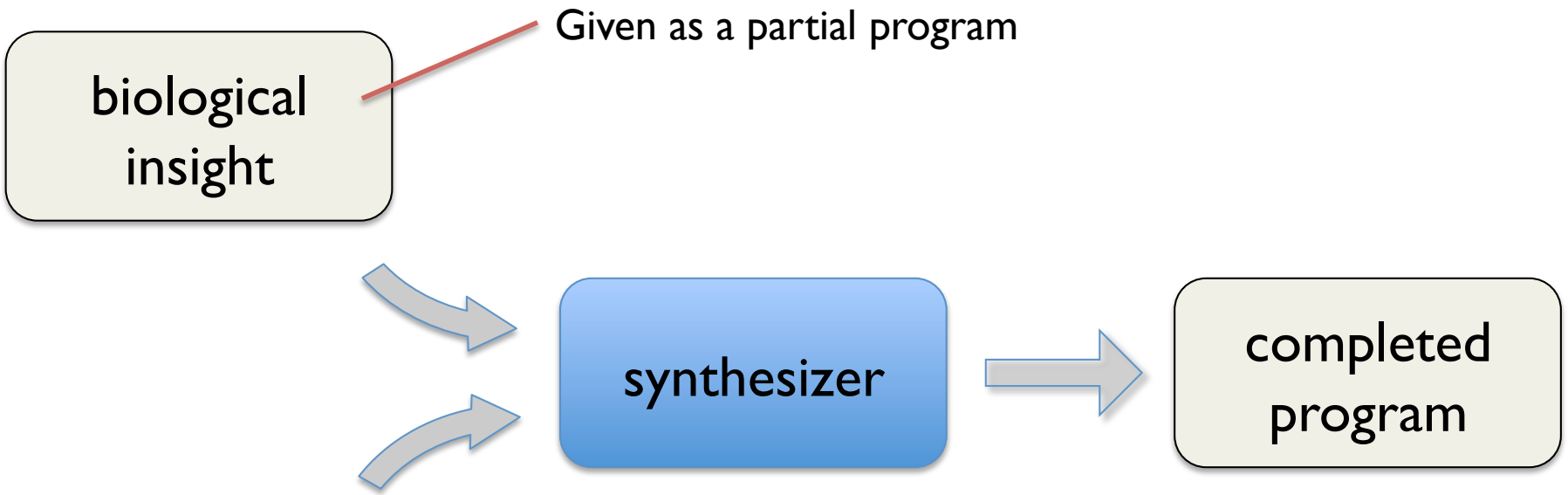
- Cell has proteins
- All proteins advance synchronously

- Proteins have discrete state and update functions.



Synthesizing Cellular Programs

Synthesis of Programs



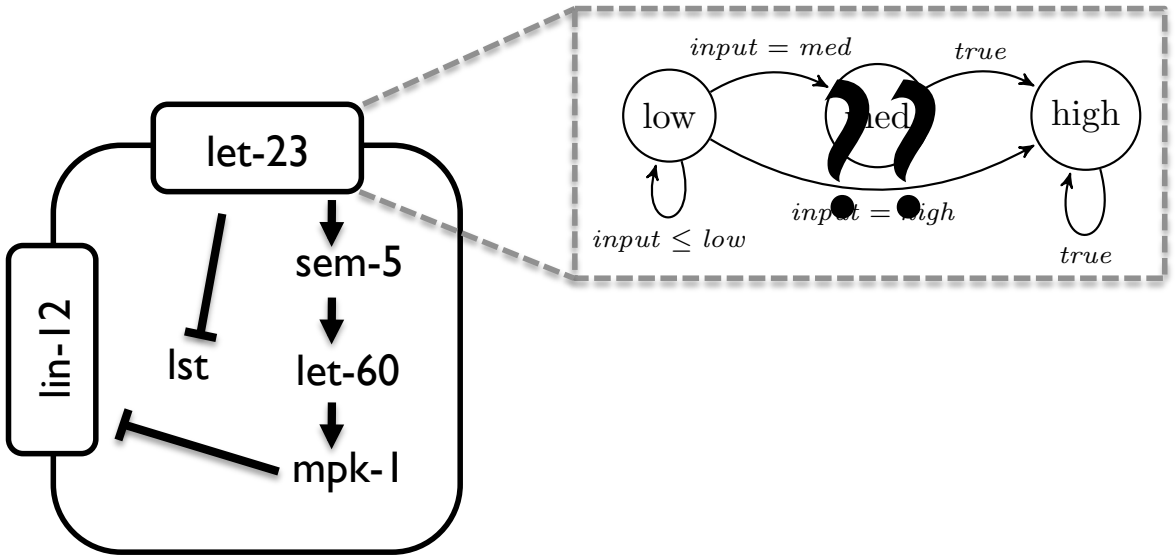
Experiment	AC	lin-12	lin-15	let-23	lst	Fate decisions
1	ON	ON	ON	ON	ON	{332123}
2	ON	OFF	ON	ON	ON	{331113}
3	ON	ON	OFF	ON	ON	{112121,122121, 212121}
...						

Partial Programs

Partial programs express biological insight:

- Which proteins are in the cell
- Which proteins may interact

Update functions can be unknown.



Synthesis Algorithm

Correctness Condition

Experiment	AC	lin-12	lin-15	let-23	lst	Fate decisions
1	ON	ON	ON	ON	ON	{332123}
2	ON	OFF	ON	ON	ON	{331113}
3	ON	ON	OFF	ON	ON	{112121, 122121, 212121}
...						

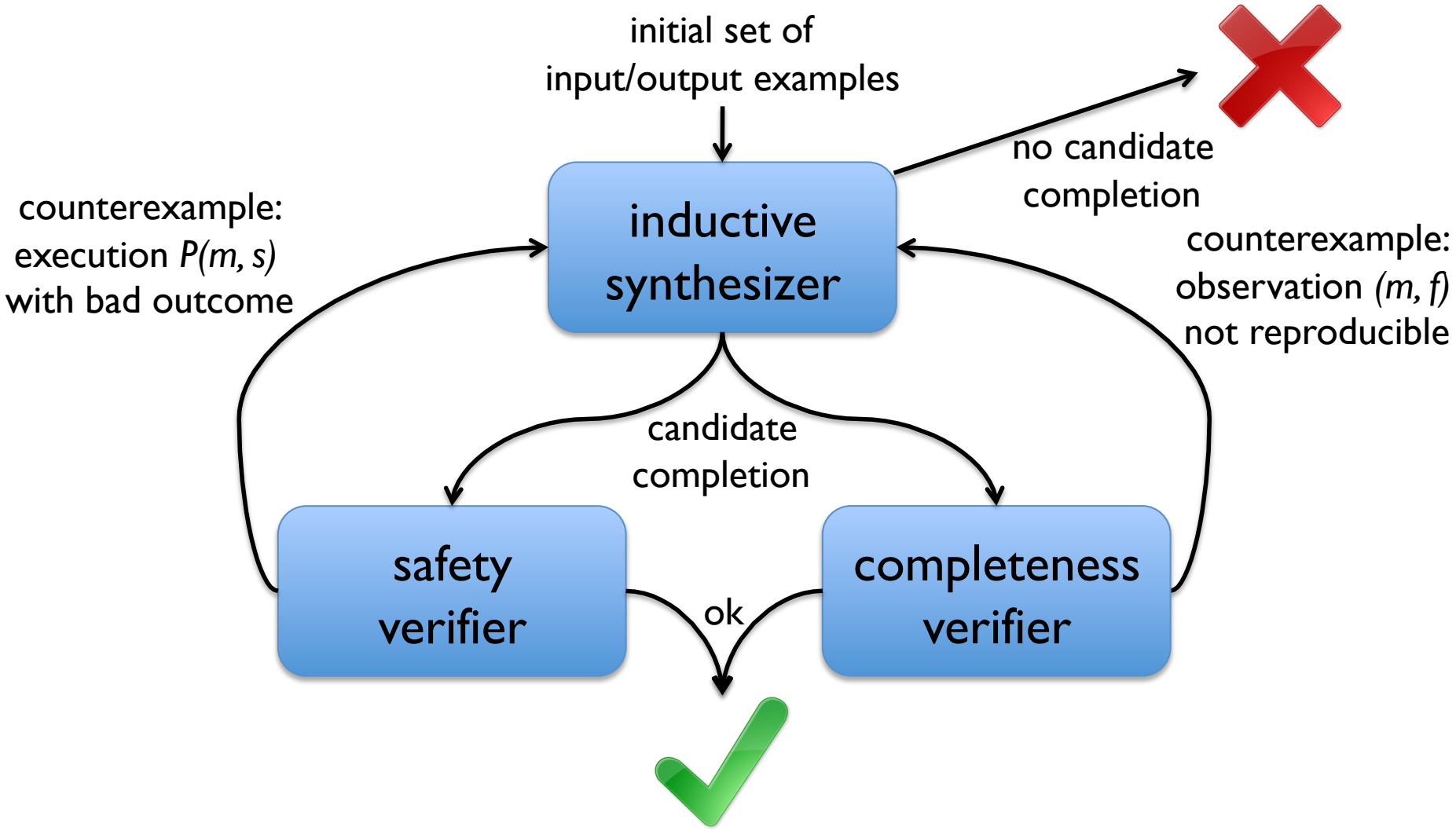
Safety: all schedules must lead the program to produce experiment outcomes observed in the wet lab.

$$\forall \text{ mutation } m.\ \forall \text{ schedule } s.\ P(m, s) \in E(m)$$

Completeness: each observed experiment outcome must be reproducible by the program for some schedule.

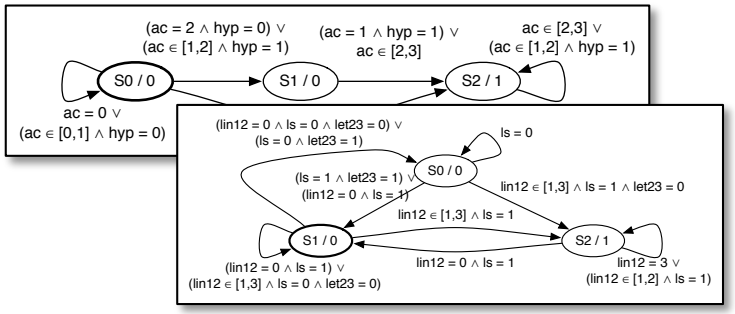
$$\forall \text{ mutation } m.\ \forall \text{ fate } f \in E(m).\ \exists \text{ schedule } s.\ P(m, s) = f$$

Counterexample-Guided Inductive Synthesis

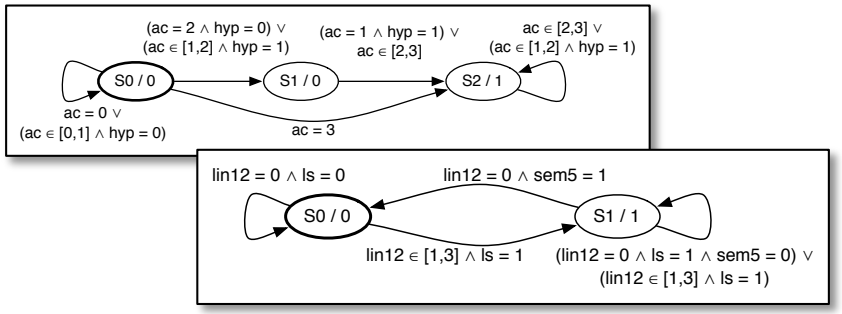


Synthesized Models

- We synthesized two models of VPCs.
- Input: Partial model that specifies known, simple protein behaviors.
- Output: Synthesized update functions for two key proteins.



model 1



model 2

Additional Algorithms for Going Beyond Synthesis to Assist Scientists

Querying Spaces of Models

- Assume a scientist obtains a formal model that agrees with all performed experiments.
- How can he make sure that a future mutation experiment won't invalidate this model?
- We can search for an alternative model that differs on a future experiment.
- Performing the new experiment will disambiguate between the two models.

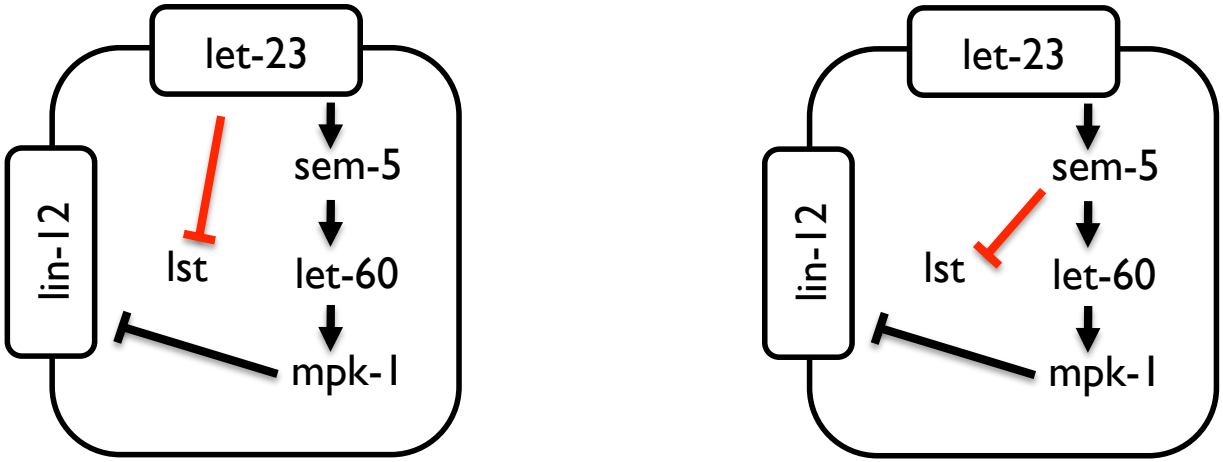
Finding Disambiguating Experiments

Simulation of future experiments using partial data:

- Assuming we didn't have the experiments from Sternberg and Horvitz 1989, we can synthesize four hypothesis models.
- Our tool suggests experiments from this paper to invalidate two of them.

Differentiating Plausible Models

- Can we differentiate the two plausible models that we synthesized?



- Mutating the modeled proteins will not suffice to disambiguate them, which suggests other methods (e.g. gene marking).

Avoiding Superfluous Experiments

- Can the scientist avoid performing superfluous experiments when revalidating results?
- We can search for a minimal, non-ambiguous subset of a set of experiments.
- Out of 48 VPC experiments, 4 suffice to yield a unique model from a given partial program.

Conclusion

Biological experiments as specification for synthesis

A synthesis algorithm with three solvers

Explore spaces of alternative models

Avoid conducting superfluous experiments