Automatic Schematization of Biological Mechanisms Via Alignment and Abstraction of Causal Relations Among Domain Specific Causal Pathways

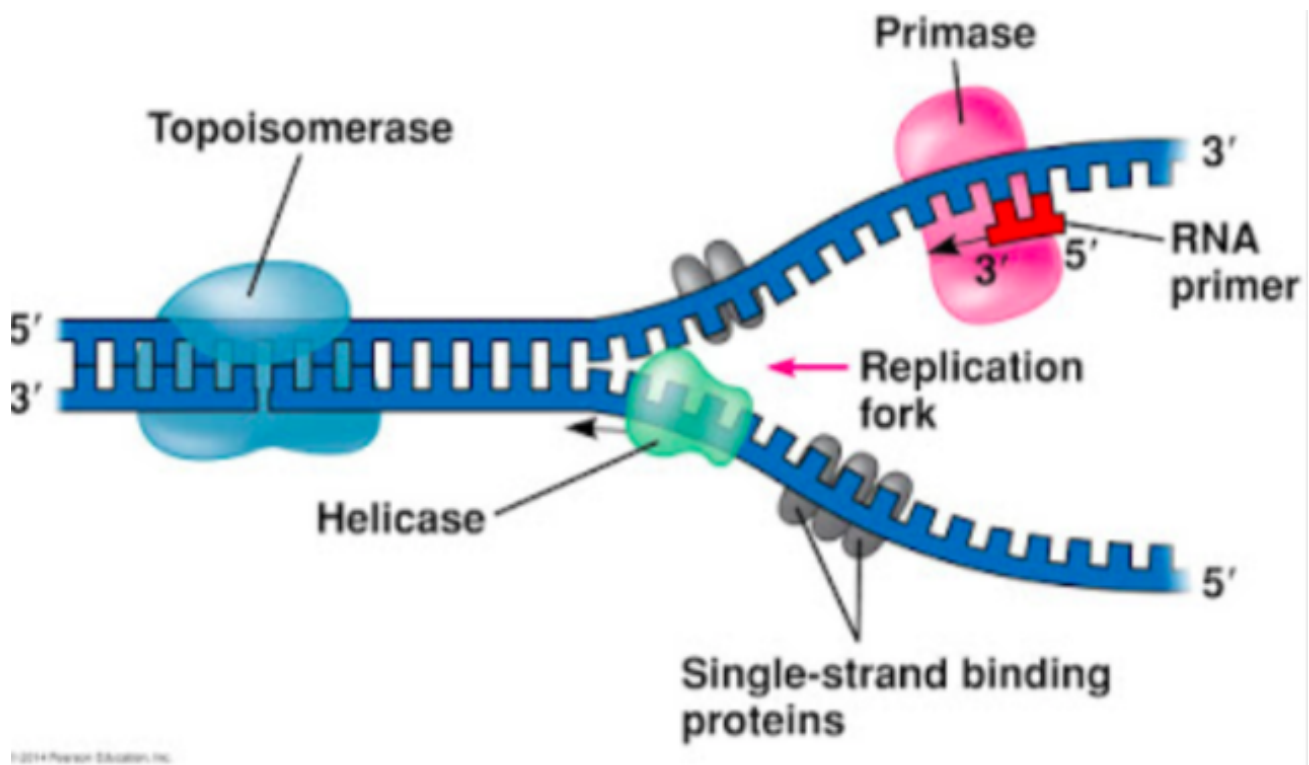Harrison Pielke-Lombardo

11/14/2020

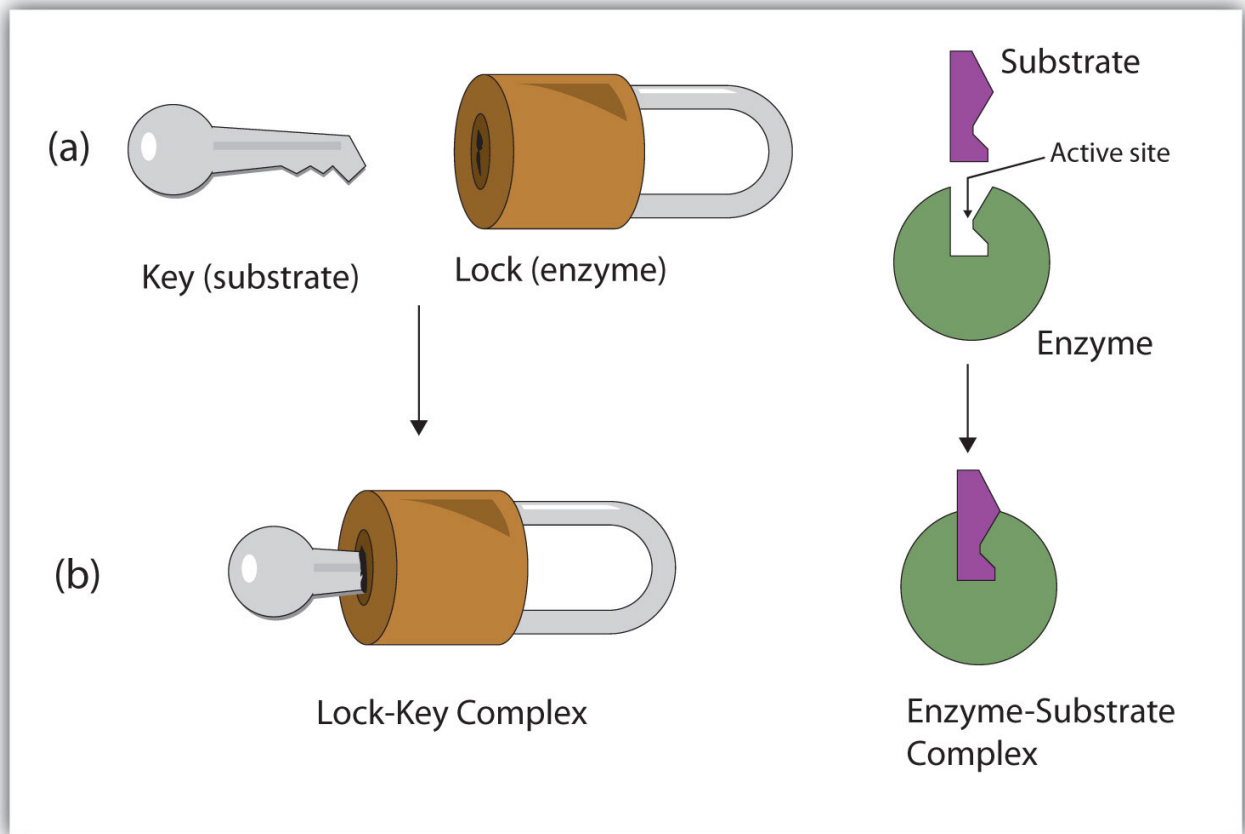## Inventing mechanism schema by refactoring biological pathways

Harrison Pielke-Lombardo
    University of Colorado Anschutz Medical Campus
    11/14/2020

Biological analogies
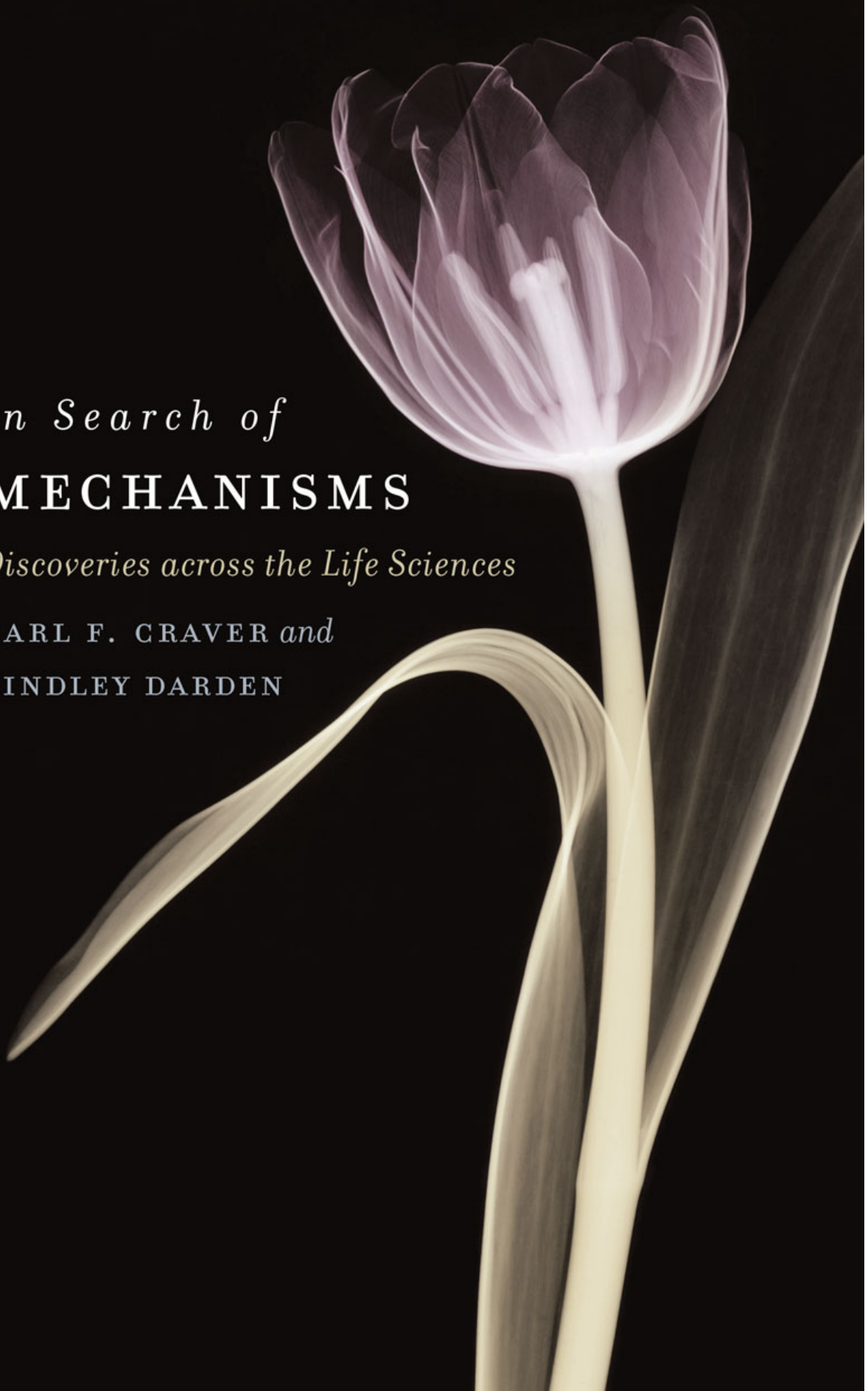

DNA replication → zipper

(a) Key (substrate) Lock (enzyme) Substrate Active site Enzyme

(b) Lock-Key Complex Enzyme-Substrate Complex

Enzyme activity → lock-and-key

In Search of

MECHANISMS

Discoveries across the Life Sciences

CARL F. CRAVER *and*

LINDLEY DARDEN

Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up conditions to finish or termination conditions.

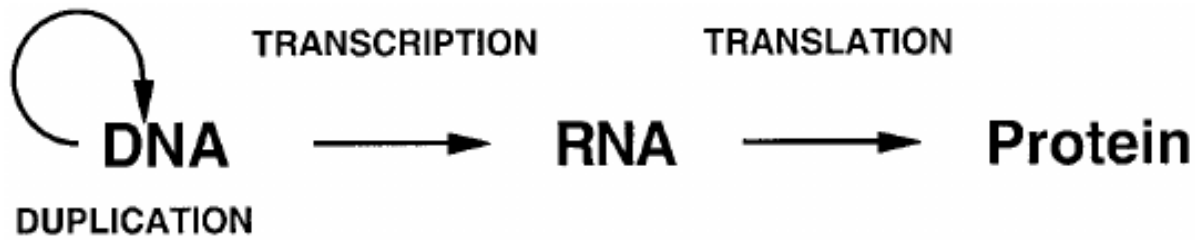## What is a mechanism schema?

Figure 1.   Schema for DNA replication and protein synthesis.

A mechanism schema is a truncated abstract description of a mechanism that can be filled with more specific descriptions of component entities and activities.
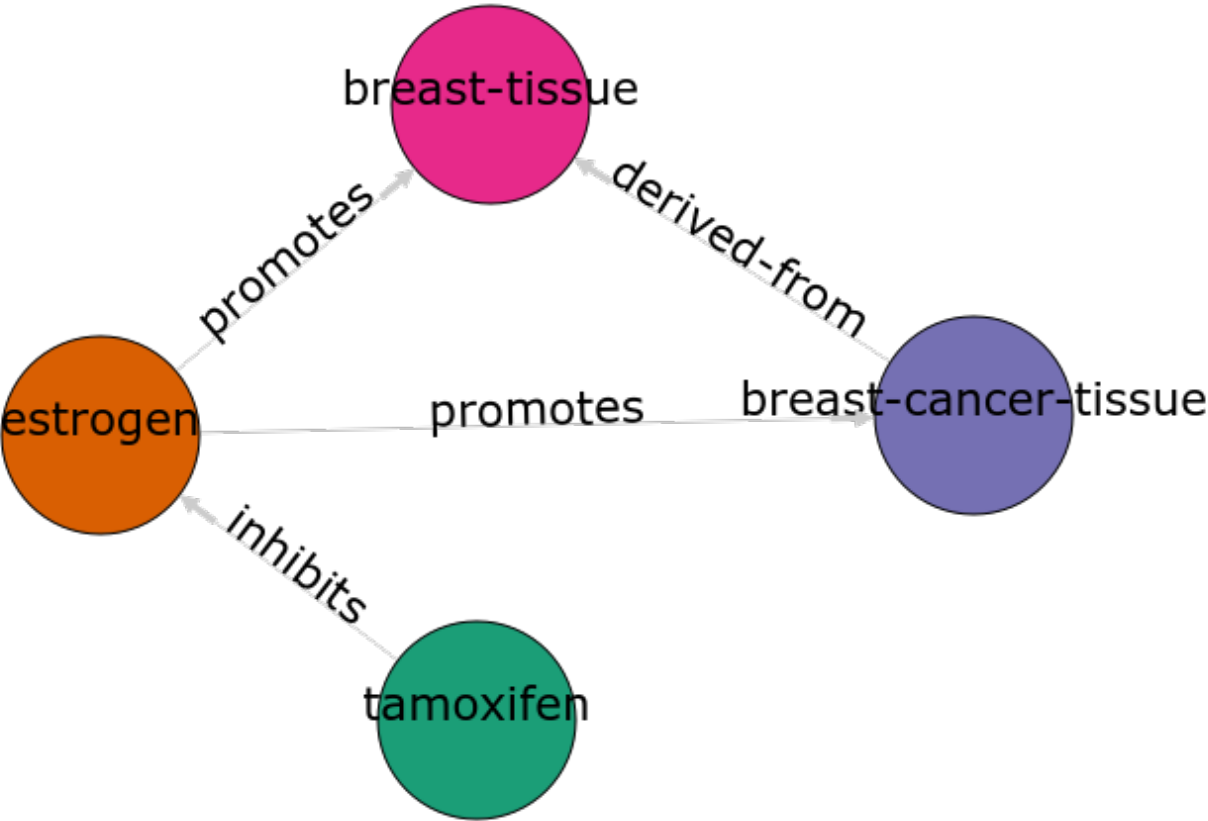
[2]

How does biological mechanism schematization relate to computational analogy?

- Tasks
    - Retrieval
    - Mapping
    - **Abstraction/generalization**
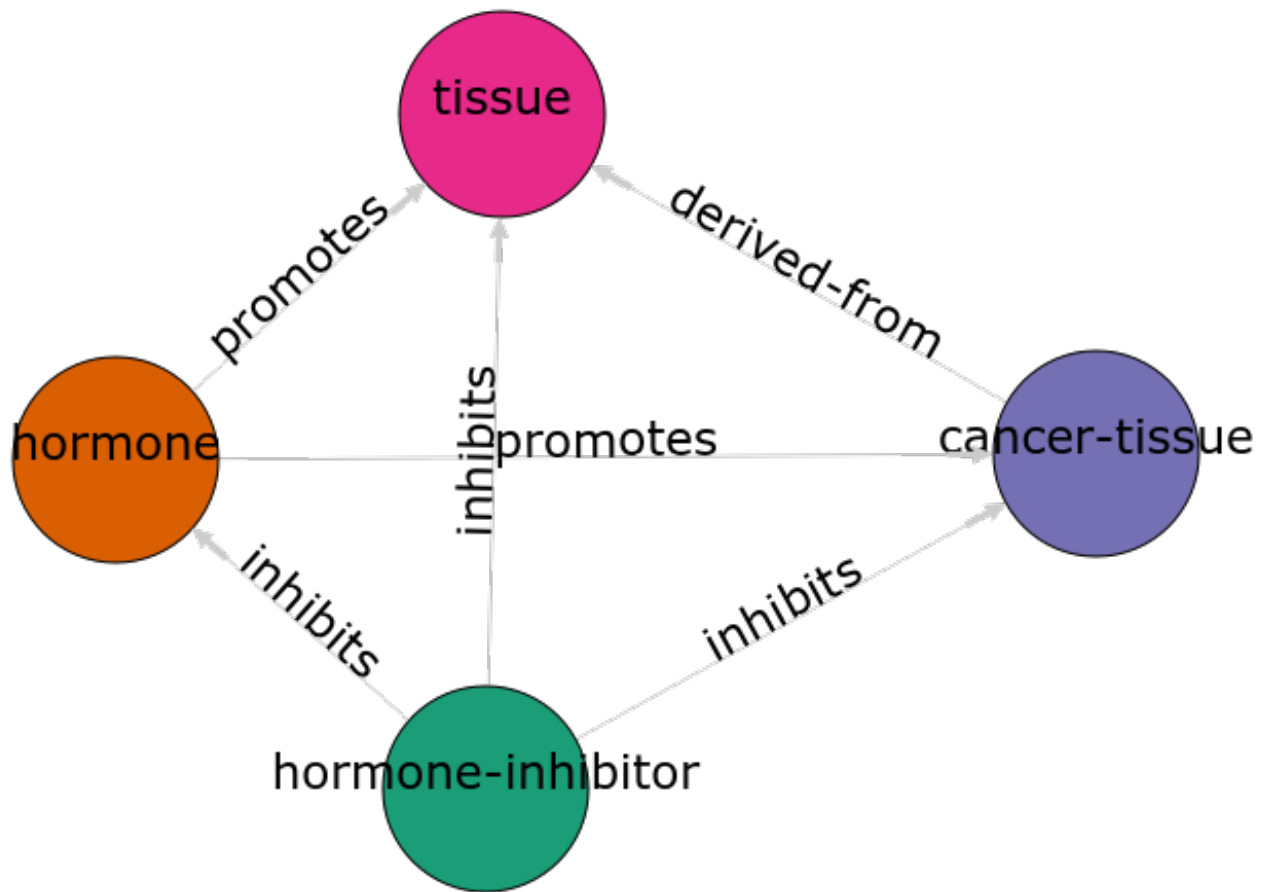   - Rerepresentation

Mechanism Schematization of Cancer: Breast vs Prostate
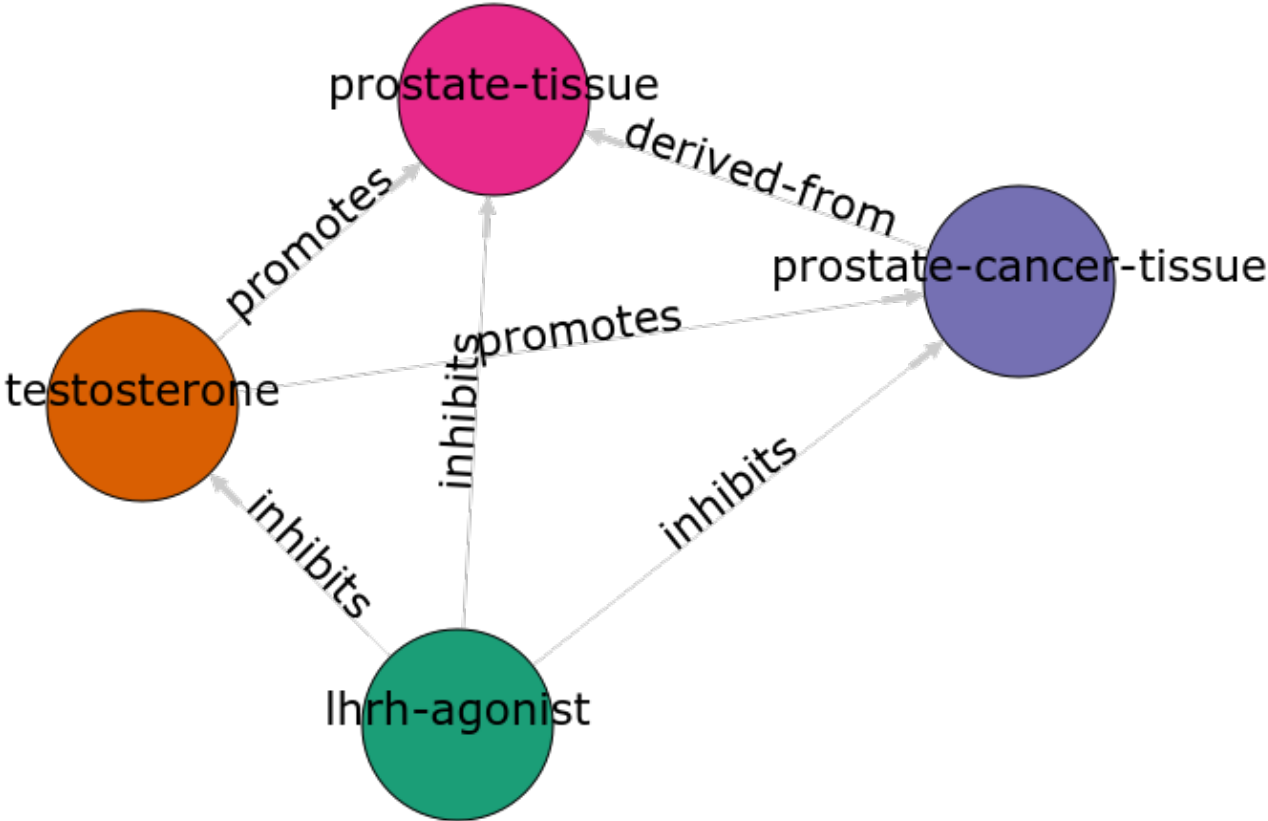
Prostate Cancer

Breast Cancer

Hormone-driven Cancer Schema

Goal

Develop and evaluate an automatic mechanism schematization method using existing biomedical ontologies, databases, and knowledge-bases and use the schema to hypothesize new relationships and disease mechanisms.

Hypothesis: Analogical alignment of concrete mechanisms will produce predictive and externally coherent mechanism schema.

Starting materials

- Input

  - Causal knowledge → concrete mechanism graphs (REACTOME/GO-CAMs/MecCog) [4, 7, 3]
  - Other knowledge → knowledge graph (KaBOB) [5]

- Output

  - Mechanism schema

- Optimization

  - Maximize coverage of concrete mechanisms while minimizing information loss
  - Minimum common subgraph

Algorithm overview

Given a set of concrete mechanisms and a knowledge graph, find a mechanism schema.
  Step = causal relationship between two entities.

1. Consider every possible step between two entities in the set of concrete mechanisms.

2. Score each possible step.

3. Assemble high scoring steps into the mechanism schema.

Final step score equation

$$S_{GO}(A, B) = \frac{\sum_{t \in T_A \cap T_B} (S_A(t) + S_B(t))}{SV(A) + SV(B)}$$

(1)

[8]

$$C(a, b, m) = len(shortestpath(a, b, m))$$

$$P(a, b, M) = \frac{1}{|M|} \sum_{m \in M} S_{GO}(a, a_m) S_{GO}(b, b_m) k^{-C(a_m, b_m, m)}$$

Future directions

- Case study: oncogenesis via developmental pathways

- Explore hyper-parameters:
  - score threshold
  - proxy-entity inclusion threshold
  - causal path distance weight
  - semantic similarity weights

Questions?

Evaluating mechanism schemas: virtues and vices

Virtues	Vices
Internal coherence	Superficiality
Testability	Incompleteness
Generates useful hypotheses	Incorrectness
Conservatism	
Simplicity	
Elegance	
Empirical adequacy	
Prediction	
Explanation	
External coherence	
Generality	
Unification	

Algorithm

### Scoring a step for each specific mechanism
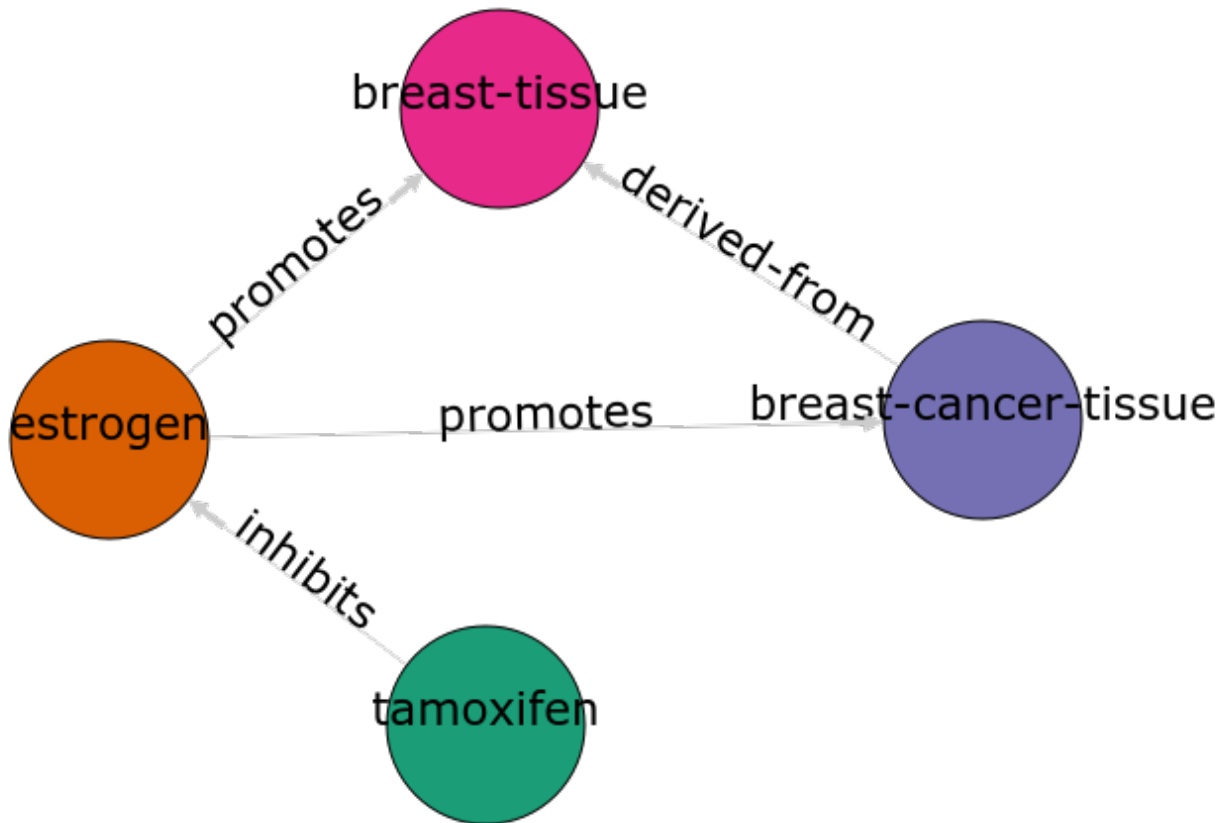
Is there an analogous step in this mechanism?

1. Find proxy entities using semantic similarity → analogous entities

2. Causal path distance between proxy entities → analogous activity or mechanism module

What are proxy entities?

Step: tamoxifen → testosterone
   What is an equivalent step between tamoxifen and testosterone in the prostate cancer mechanism?
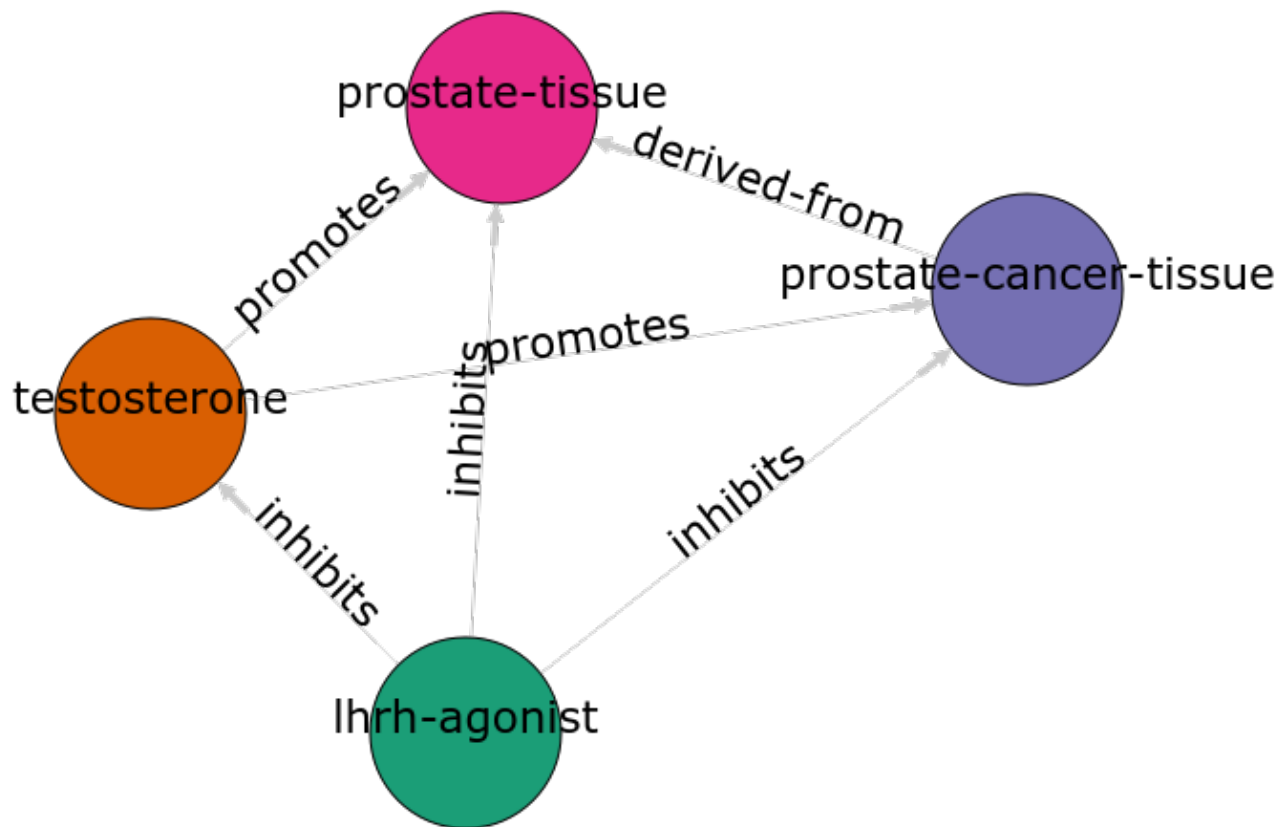
Evaluation

Prediction evaluation using leave-one out link prediction

- Input: GO-CAMs as concrete mechanisms
- Gold standard: Left out mechanism steps
- Success metric: F1-score

External coherence evaluation using hierarchical clustering

- Input: REACTOME pathways as concrete mechanisms
- Gold standard: Gene Ontology Biological Process (GO-BP) hierarchy
- Success metric: Tree edit distance [6]

Unification vs superficiality evaluation optimizing coverage vs information loss

- Input: GO-CAMs or REACTOME pathways
- Gold standard: None exists. Manual inspection
- Success metric: Comparison of coverage to information loss

What do we expect to see during mechanism schematization?

- Abstraction of entities

- Modularization of activities

- Generalization of steps

Use of mechanism schemas

- Indexing/retrieval

- Comparison

- Transformation

- Induction

Role of statistics

- Simulate generality by compiling lots of specific examples and averaging over them

- Use knowledge to constrain statistical analysis

Acknowledgments

References

[1] Carl F. Craver and Lindley Darden. *In Search of Mechanisms: Discoveries across the Life Sciences*, volume 51. University of Chicago Press, Chicago and London, 2014.

[2] Lindley Darden. Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining. *Philosophy of Science*, 69(S3):S354–S365, September 2002.

[3] Lindley Darden, Kunal Kundu, Lipika R. Pal, and John Moult. Harnessing formal concepts of biological mechanism to analyze human disease. *PLoS Computational Biology*, 14(12):1–10, 2018.

[4] Antonio Fabregat, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, Bijay Jassal, Florian Korninger, Bruce May, Marija Milacic, Corina Duenas Roca, Karen Rothfels, Cristoffer Sevilla, Veronica Shamovsky, Solomon Shorser, Thawfeek Varusai, Guilherme Viteri, Joel Weiser, Guanming Wu, Lincoln Stein, Henning Hermjakob, and Peter D'Eustachio. The Reactome Pathway Knowledgebase. *Nucleic Acids Research*, 46(D1):D649–D655, 2018.

[5] Kevin M. Livingston, Michael Bada, William A. Baumgartner, and Lawrence E. Hunter. KaBOB: Ontology-based semantic integration of biomedical databases. *BMC Bioinformatics*, 16(1):126, 2015.

[6] Mateusz Pawlik and Nikolaus Augsten. RTED: A Robust Algorithm for the Tree Edit Distance. *arXiv:1201.0230 [cs]*, December 2011.

[7] Paul D. Thomas, David P. Hill, Huaiyu Mi, David Osumi-Sutherland, Kimberly Van Auken, Seth Carbon, James P. Balhoff, Laurent Philippe Albou, Benjamin Good, Pascale Gaudet, Suzanna E. Lewis, and Christopher J. Mungall. Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems. *Nature Genetics*, 51(10):1429–1433, October 2019.

[8] James Z. Wang, Zhidian Du, Rapeeporn Payattakool, Philip S. Yu, and Chin-Fu Chen. A new method to measure the semantic similarity of GO terms. *Bioinformatics*, 23(10):1274–1281, May 2007.