An Interactive Modeling Environment for Systems Biology of Aging

Pat Langley

Computer Science and Engineering
Arizona State University
Tempe, Arizona

Thanks to D. Bidaye, J. Difzac, J. Furber, S. Kim, S. Racunas, N. Shah, J. Shrager, D. Stracuzzi, and M. Verdicchio for their contributions to this research, which was funded in part by a grant from Science Foundation Arizona.

Challenges in Scientific Modeling

The creation of complex scientific models raises three separate but interrelated challenges:

- The social nature of science requires the use of *communicable* formalisms that researchers can exchange.
- Science is concerned with *explaining observations* in terms of familiar concepts and mechanisms.
- Science operates in a *cumulative* manner that extends models by filling in technical details.

Software environments for scientific modeling must address each of these issues.

In this talk, I present one such environment and its responses.

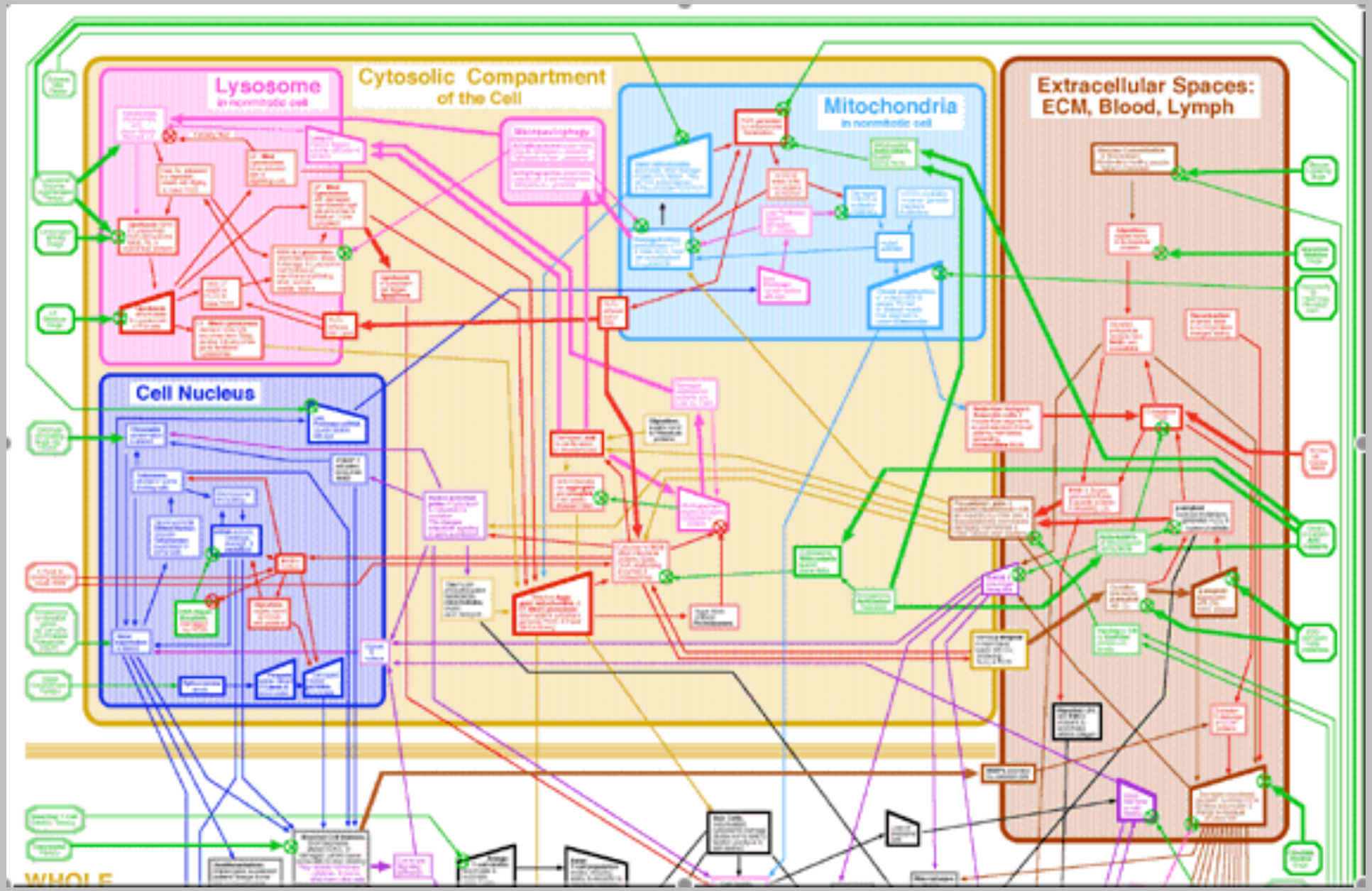
The Complexity of Human Aging

There is mounting evidence that aging involves many interacting processes, including:

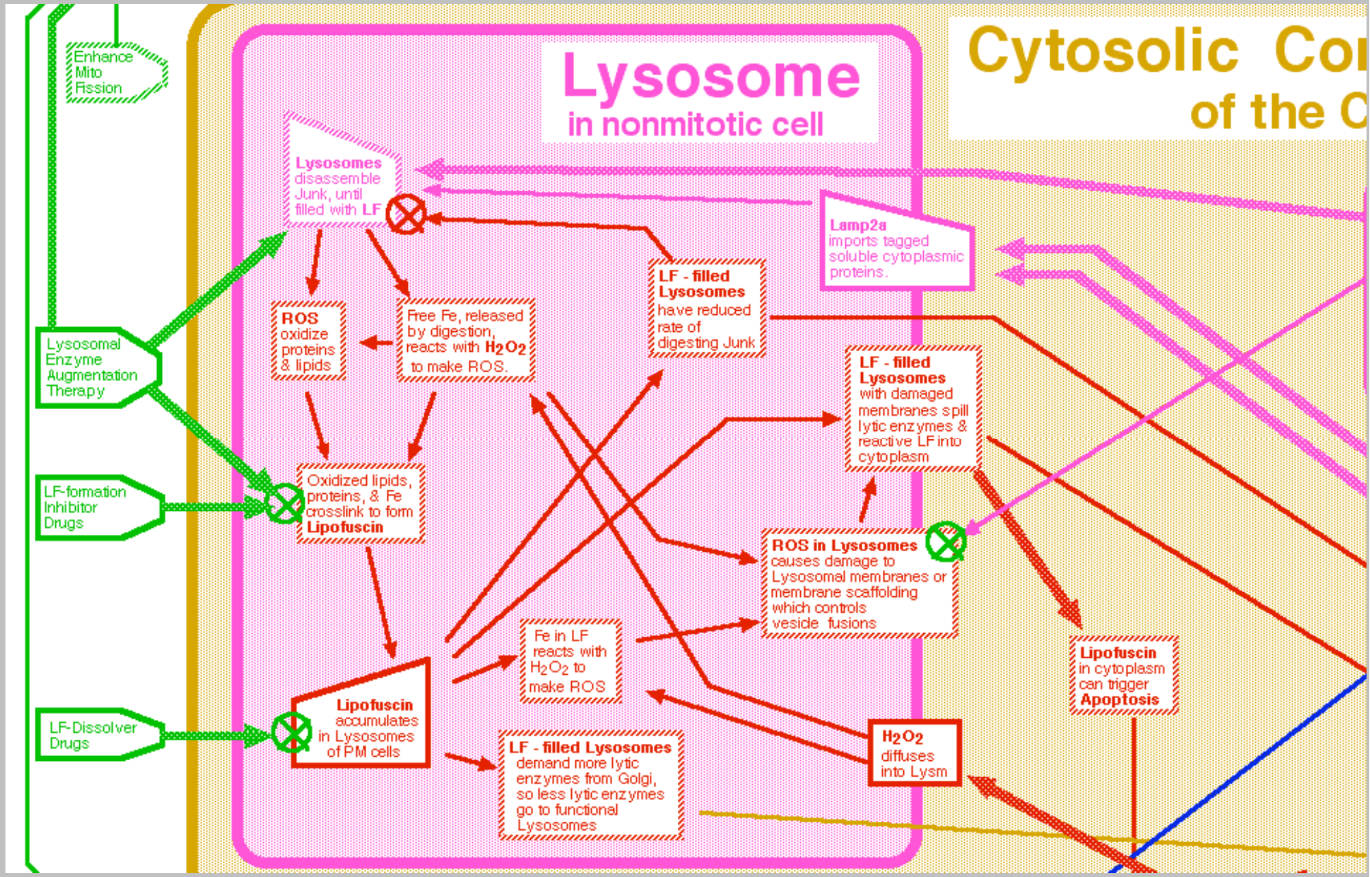
- Mutation of the mitochondrial DNA
- Accumulation of lipofuscin in the lysosomes
- Protein crosslinking in the extra-cellular matrix
- Shortening of telomeres in the nucleus

The daunting complexity of these and related processes suggests the need for computational assistance.

Furber's Network Diagram of Aging



Furber's Network Diagram of Aging



Representing Models of Aging

We want a notation for models of aging that is precise enough for a digital computer.

Computational biology is rife with candidate formalisms, but most are problematic:

- *Differential equations* require functional forms and parameters
- *Boolean networks* assume discrete, not continuous, variables
- *Bayesian networks* require arbitrary probabilistic parameters

We need a notation that makes closer contact with biologists' ideas about aging mechanisms.

Representing Models of Aging

Fortunately, one of the key insights of artificial intelligence is that computers are general symbol manipulators.

This means that we can encode in digital form any content that biologists entertain, such as:

- *Structures* that occur in organisms (e.g., organic chemicals)
- *Processes* that operate on those structures (e.g., biochemical reactions and pathways)

However, the diverse and incomplete character of results about aging make full biochemical models problematic.

We need a more abstract representation similar to that in the AI literature on qualitative reasoning (e.g., Forbus, 1984).

Qualitative Causal Models of Aging

Our modeling formalism involves a set of elements, each of which specifies:

- Places in the cell (e.g., lysosome, cytoplasm)
- Quantities that are measured in a place
  - For unstable entities (e.g., ROS, Fe)
  - For stable entities (e.g., oxidized proteins, lipofuscin)
- Causal influences between quantities
  - Increase/decrease of one quantity with another
  - Quantities may encode amounts or rates of change

Each influence describes a *qualitative* causal link between two *quantitative* variables.

We can also state this content in a constrained form of English.

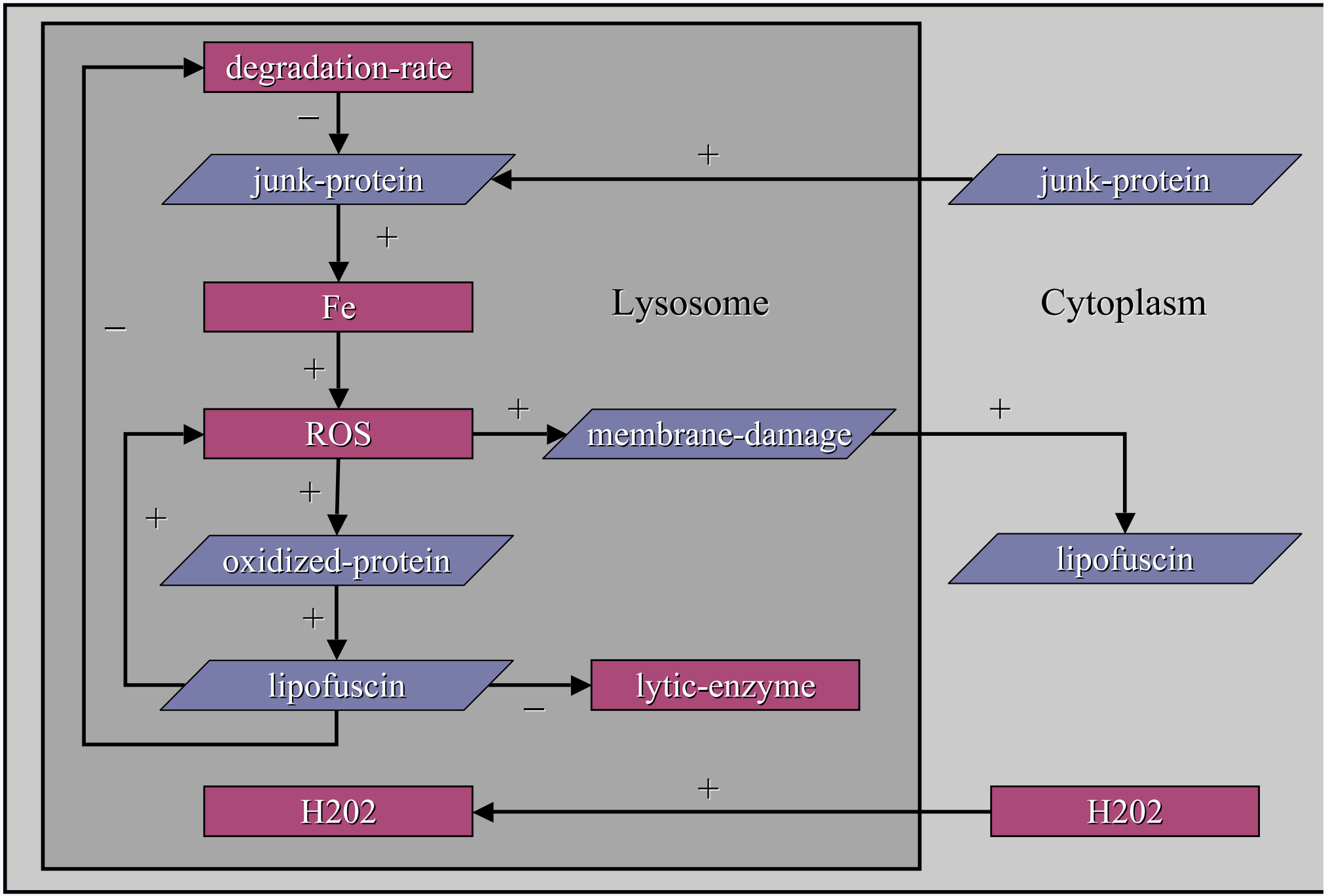
Examining a Lysosomal Model

> load “lyso.lisp”
  loaded “lyso.lisp”
> show places entities
Places:
  p1: cell
  p2: lysosome in the cell
  p3: cytoplasm in the cell
Quantities:
  q0: degradation-rate in the lysosome
  q1: junk-protein in the lysosome
  q2: junk-protein in the cytoplasm
  q3: fe in the lysosome
  q4: ros in the lysosome
  q5: oxidized-protein in the lysosome
  q6: lipofuscin in the lysosome
  q7: lipofuscin in the cytoplasm
  q8: lytic-enzyme in the lysosome
  q9: damaged-membrane in the lysosome
  q10: h2o2 in the lysosome
  q11: h2o2 in the cytoplasm

Examining a Lysosomal Model

> show claims
Hypothesized claims:
c1: junk-protein decreases with degradation-rate in the lysosome
c2: junk-protein in the lysosome increases with junk-protein in the cytoplasm
c3: fe increases with junk-protein in the lysosome
[lysosome digestion produces fe from junk proteins]
c4: ros increases with fe in the lysosome
c5: oxidized-protein increases with ros in the lysosome
[ros oxidizes proteins to produce oxidized proteins]
c6: lipofuscin increases with oxidized-protein in the lysosome
[oxidized-proteins and fe crosslink to form lipofuscin]
c7: degradation-rate decreases with lipofuscin in the lysosome
[lipofuscin reduces the rate of disassembling junk proteins]
c8: lytic-enzyme decreases with lipofuscin in the lysosome
c9: ros increases with lipofuscin in the lysosome
c10: damaged-membrane increases with ros in the lysosome
c11: lipofuscin in the cytoplasm increases with damaged-membrane in the lysosome
[damaged lysosomal membranes spill lipofuscin into cytoplasm]
c12: h2o2 in the lysosome increases with h2o2 in the cytoplasm

Depiction of the Lysosomal Model



Examining a Mitochondrial Model

> show claims
Hypothesized claims:
c1: ROS increases with ROS-rate in the mitochondria
c2: mutated-mtDNA increases with ROS in the mitochondria
c3: Lon-protease-mRNA decreases with Lon-rate in the nucleus
c4: Lon-protease in the mitochondria increases with Lon-protease-mRNA in the nucleus
c5: mutated-mtDNA decreases with Lon-Protease in the mitochondria
c6: damaged-mitochondria increases with mutated-mtDNA in the mitochondria
c7: damaged-mitochondria decreases with Lon-Protease in the mitochondria
c8: ROS decreases with antioxidants in the mitochondria
c9: ROS increases with damaged-mitochondria in the mitochondria
c10: ATP decreases with damaged-mitochondria in the mitochondria

Examining a Nuclear Model

> show claims
Hypothesized claims:
c1: ROS increases with ROS-rate in the nucleus
c2: mutated-nDNA increases with ROS in the nucleus
c3: DNA-repair-enzymes decreases with ROS in the nucleus
c4: mutated-nDNA decreases with DNA-repair-enzymes in the nucleus
c5: altered-gene-expression increases with mutated-nDNA in the nucleus
c6: shortened-telomeres increases with altered-gene-expression in the nucleus
c7: shortened-telomeres increases with ROS in the nucleus
c8: oxidized-nuclear-pore-proteins increases with ROS in the nucleus
c9: cytoplasmic-tubulin increases with oxidized-nuclear-pore-proteins in the nucleus
c10: mutated-nDNA increases with cytoplasmic-tubulin in the nucleus
c11: damaged-nuclear-proteins increases with cytoplasmic-tubulin in the nucleus
c12: damaged-nuclear-proteins increases with Progerin in the nucleus
c13: damaged-nuclear-proteins increases with ROS in the nucleus
c14: Lon-protease-mRNA increases with altered-gene-expression in the nucleus
c15: altered-chromatin-conformation increases with PARP-1 in the nucleus
c16: altered-gene-expression increases with altered-chromatin-conformation in the nucleus
c17: altered-chromatin-conformation increases with shortened-telomeres in the nucleus

Reasoning About Aging

A scientific model of aging is lacking unless one can relate it to observable phenomena.

Such an account should let one answer questions like:

- What biological effects does the model predict?
- What observations/experiments does the model match?
- How does the model explain a given phenomenon?
- How would changes to the model alter its predictions?

Model complexity can make these difficult to do manually, but we can provide computational support for such reasoning.

Encoding Predictions and Observations

We cannot relate a model's predictions to observations until we represent them; we use a notation similar to model hypotheses:

- A quantity increases (decreases) with another quantity
  - E.g., lipofuscin in the lysosome increases with time
  - E.g., lytic-enzyme decreases with ROS in the lysosome
- A quantity does not change with another quantity
  - E.g., H2O2 does not vary with ROS in the lysosome

Each statement specifies, in constrained English, a pairwise relation between two quantities.

Note: *These describe phenomena the model does or should predict; they are not part of the model themselves.*

Examining Facts, Queries, and Predictions

> show facts
Empirical facts:
  f1: lipofuscin in the lysosome increases with time
  f2: lipofuscin in the cytoplasm increases with time
  f3: lytic-enzyme decreases with ros in the lysosome
  f4: h2o2 does-not-change with ros in the lysosome
> does lytic-enzyme change with ros in the lysosome ?
   p1: lytic-enzyme decreases with ros in the lysosome
> does oxidized-protein change with h2o2 in the lysosome ?
   p2: oxidized-protein does-not-change with h2o2 in the lysosome
> predict f1 f2 f3 f4
  f1: lipofuscin in the lysosome increases with time
       the model makes ambiguous predictions.
  f2: lipofuscin in the cytoplasm increases with time
       the model makes ambiguous predictions.
  f3: lytic-enzyme decreases with ros in the lysosome
  p3: the model predicts the same relation.
  f4: h2o2 does-not-change with ros in the lysosome
  p4: the model predicts the same relation.

Simplifying Assumptions

Our approach to prediction relies on some important modeling assumptions:

- All quantities are positive, including those describing rates.
- The system being modeled exists in a single qualitative state.
- If nothing influences a quantity, then it has a constant value.
- Stable quantities accumulate effects on them over time.
- Transient quantities do not accumulate effects over time.

Taken together, these assumptions simplify the task of reasoning over models substantially.

They seem appropriate for aging and age-related disease, but not for all biological problems.

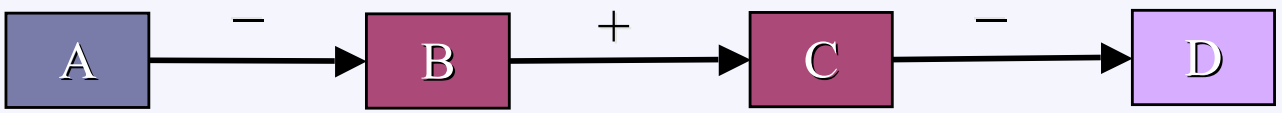
Generating Experimental Predictions

Given a qualitative query about the relation between two model quantities, the environment:

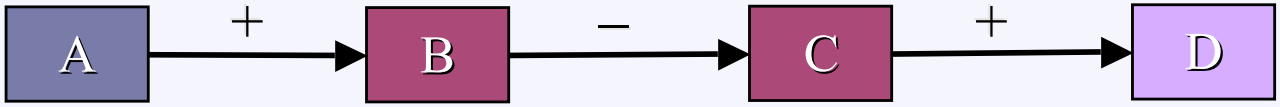
- Chains backward from dependent term D to find all paths that connect it with independent term I.
- If no paths exist, then it predicts no empirical relationship.
- If a path has an even number of *decreases* links, it predicts D *increases* with I, else that D *decreases* with I.
- Provides a definite relationship if all path predictions agree.
- Notes an ambiguous relationship if path predictions disagree.

This form of qualitative reasoning predicts changes in response to experimental manipulation.

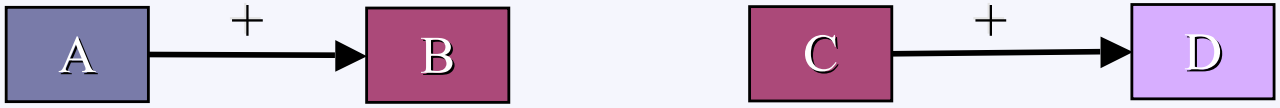
Generating Experimental Predictions



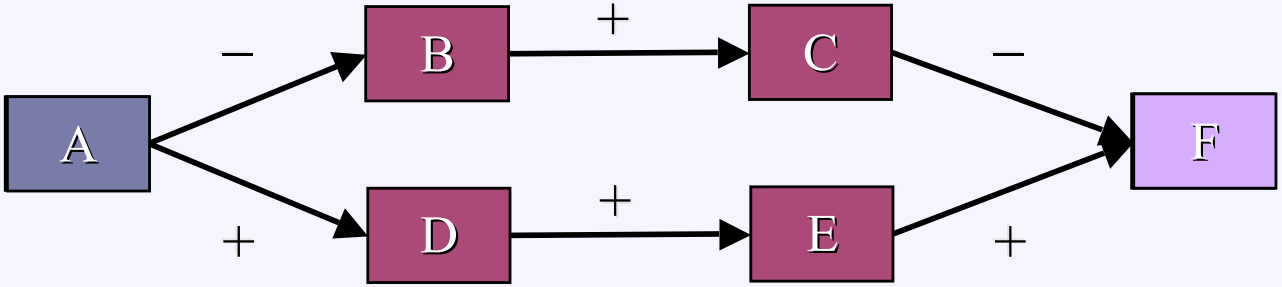
D increases with A



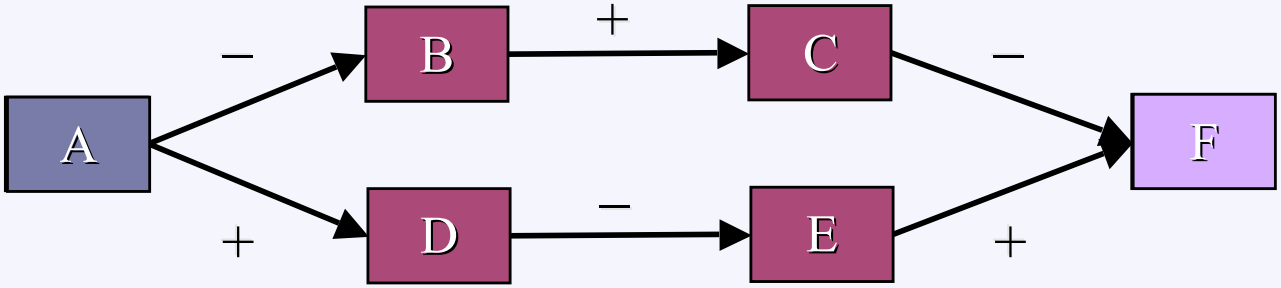
D decreases with A



D does not vary with A

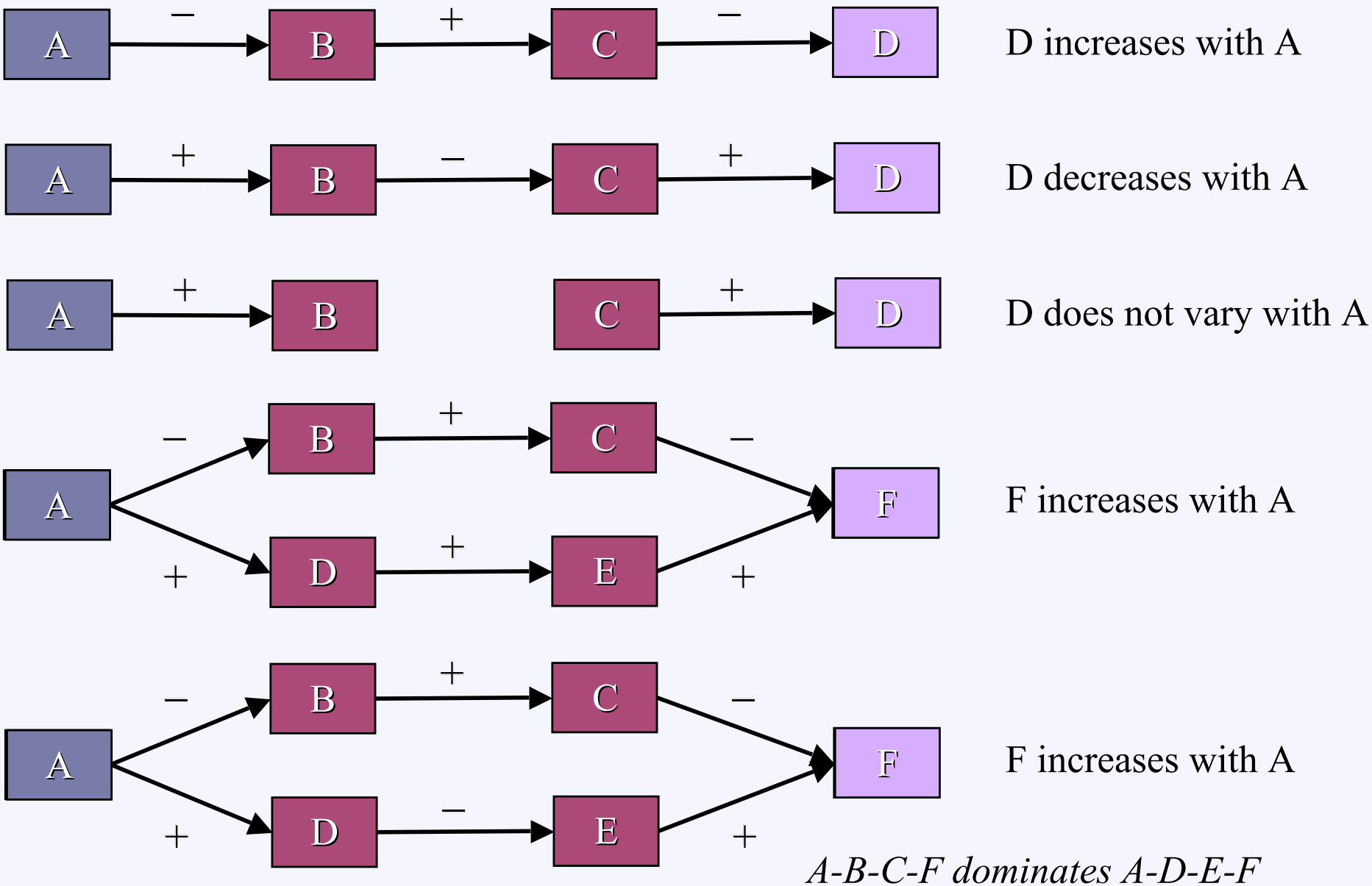


F increases with A



ambiguous relation

Avoiding Ambiguity with Dominance Relations



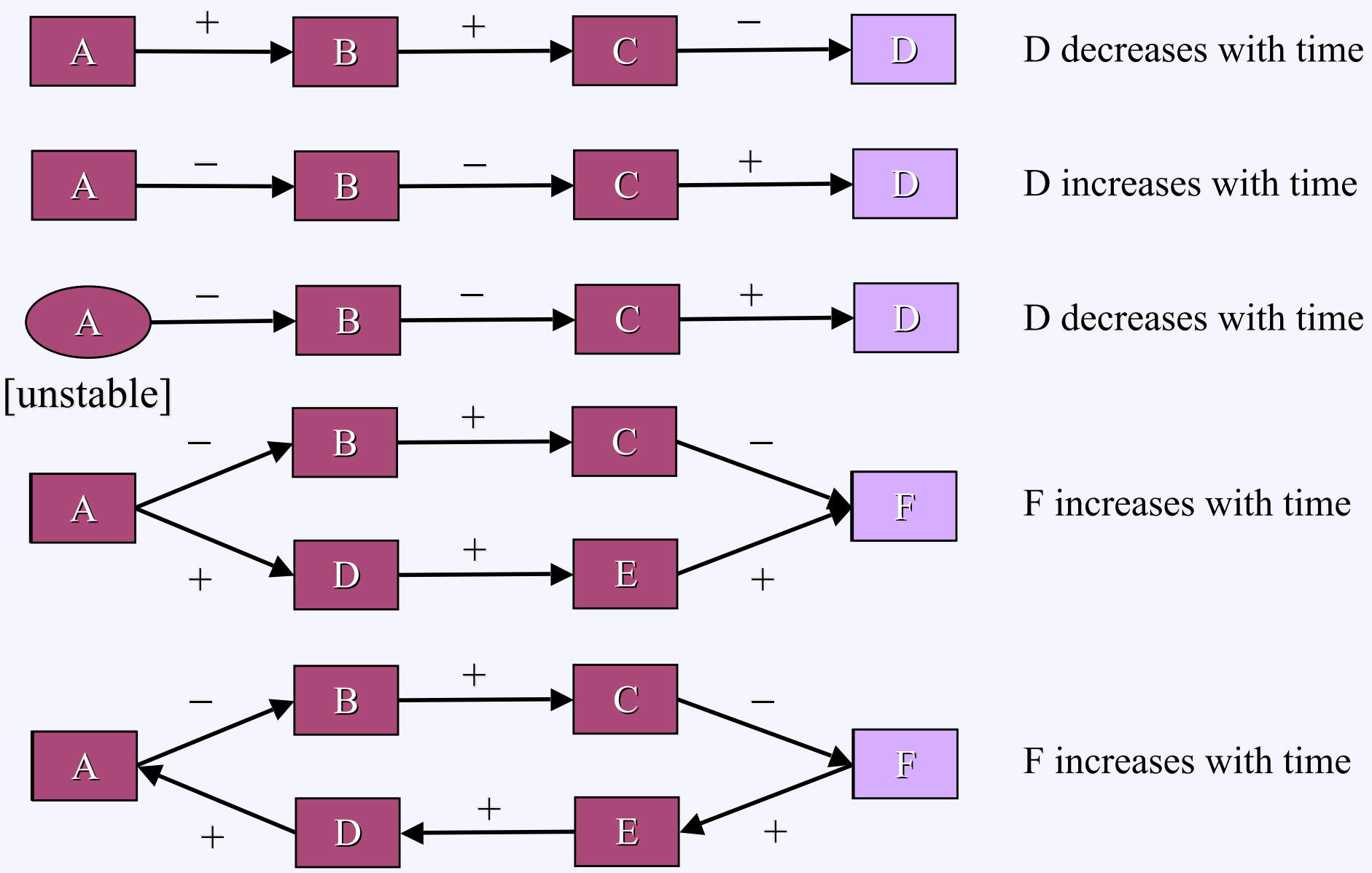
Generating Temporal Predictions

Given a qualitative query about the relation between a model quantity and time, the environment:

- Chains backward from dependent term D to find the start of all paths that lead to D.
- If no paths exist, then it predicts no change over time.
- Assumes that any transient quantities not preceded in a path by a stable quantity are constant.
- Assumes the first stable quantity in a path increases/decreases depending on its predecessor's influence.

Otherwise, it follows the same strategy as used for experimental settings to determine how D changes over time.

Generating Temporal Predictions



Examining Explanations of Predictions

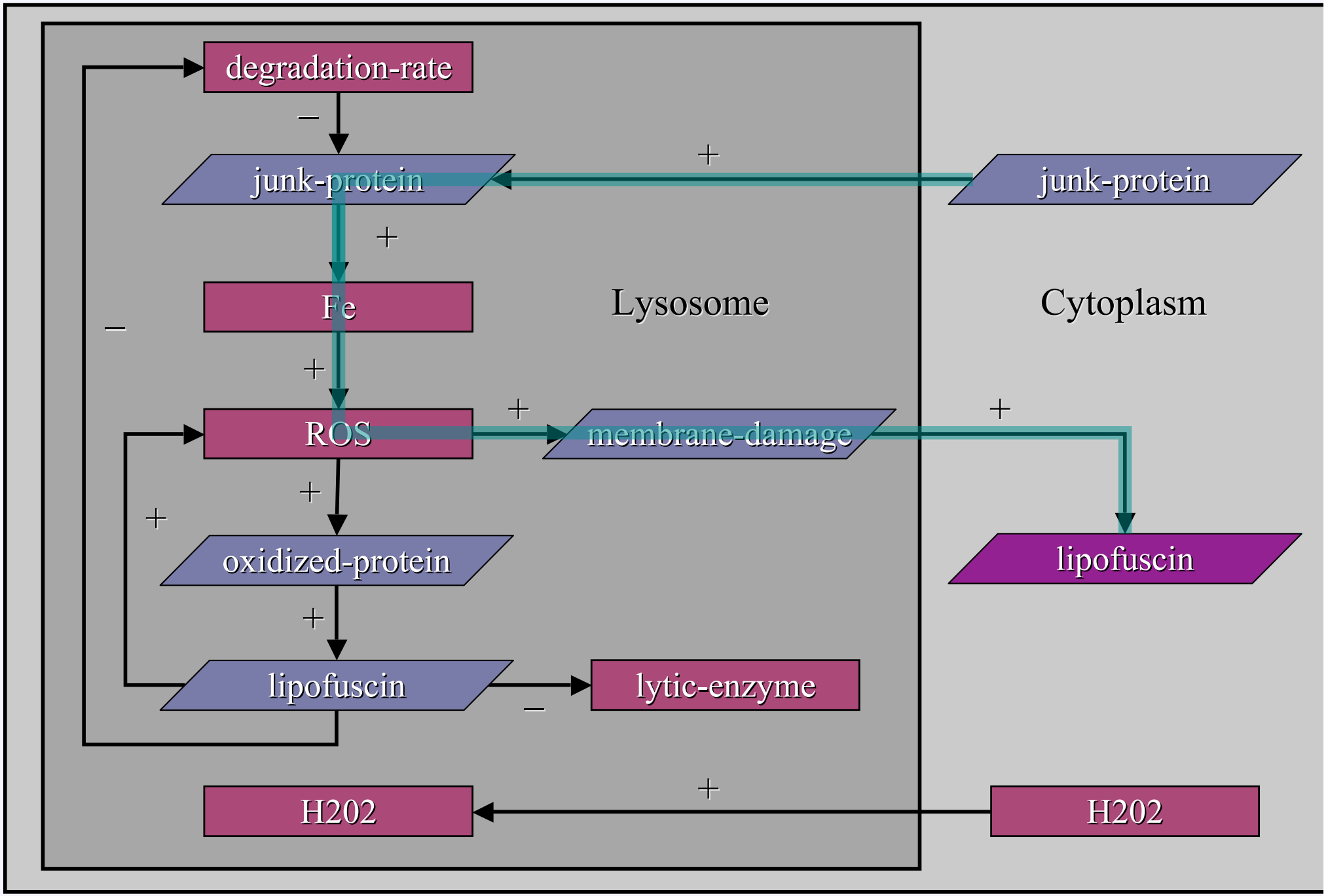
> explain p1
P1: LIPOFUSCIN in the CYTOPLASM INCREASES with TIME because:

Predicted effect: INCREASES
LIPOFUSCIN in the CYTOPLASM INCREASES with DAMAGED-MEMBRANE in the LYSOSOME
DAMAGED-MEMBRANE INCREASES with ROS in the LYSOSOME
ROS INCREASES with LIPOFUSCIN in the LYSOSOME
LIPOFUSCIN INCREASES with OXIDIZED-PROTEIN in the LYSOSOME
OXIDIZED-PROTEIN INCREASES with ROS in the LYSOSOME

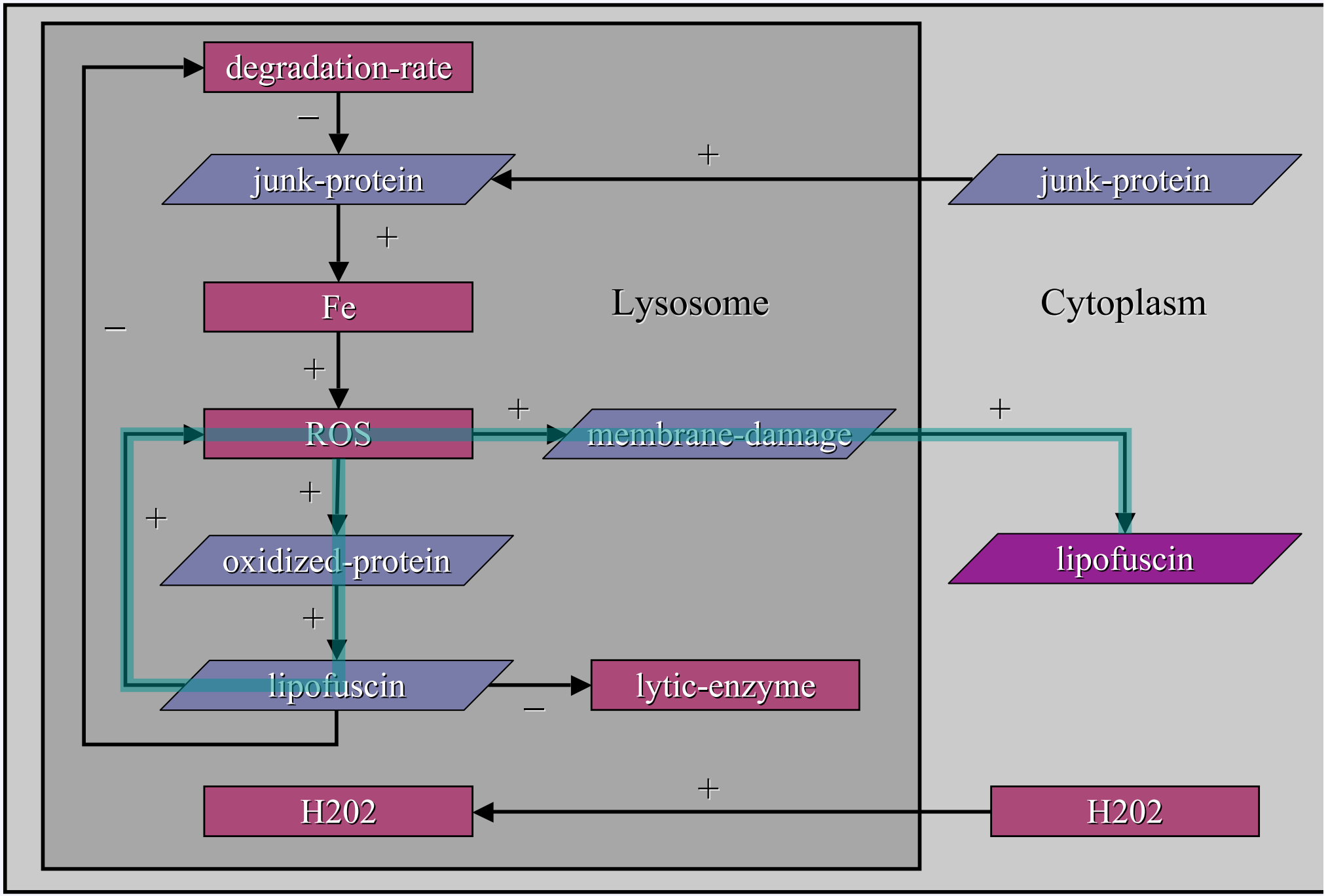
Predicted effect: INCREASES
LIPOFUSCIN in the CYTOPLASM INCREASES with DAMAGED-MEMBRANE in the LYSOSOME
DAMAGED-MEMBRANE INCREASES with ROS in the LYSOSOME
ROS INCREASES with FE in the LYSOSOME
FE INCREASES with JUNK-PROTEIN in the LYSOSOME
JUNK-PROTEIN in the LYSOSOME INCREASES with JUNK-PROTEIN in the CYTOPLASM

Predicted effect: INCREASES
LIPOFUSCIN in the CYTOPLASM INCREASES with DAMAGED-MEMBRANE in the LYSOSOME
DAMAGED-MEMBRANE INCREASES with ROS in the LYSOSOME
ROS INCREASES with FE in the LYSOSOME
FE INCREASES with JUNK-PROTEIN in the LYSOSOME
JUNK-PROTEIN DECREASES with DEGRADATION-RATE in the LYSOSOME
DEGRADATION-RATE DECREASES with LIPOFUSCIN in the LYSOSOME
LIPOFUSCIN INCREASES with OXIDIZED-PROTEIN in the LYSOSOME
OXIDIZED-PROTEIN INCREASES with ROS in the LYSOSOME

An Explanatory Pathway



Another Explanatory Pathway



Updating a Qualitative Model of Aging

Biologists should also be able to extend and revise models of aging easily and efficiently.

Our environment lets users alter the current model interactively in five basic ways:

- Adding a new place, quantity, claim, or fact
- Adding a note (e.g., a URL) to a claim or fact
- Removing a place, quantity, claim, or fact
- Disabling or enabling a claim or fact
- Saving revisions to a file that can be loaded later

The modular character of models makes them straightforward to update and revise.

Examining the Effects of Model Revision

> predict
  f1: lipofuscin in the lysosome increases with time
      the model makes ambiguous predictions.
  f2: lipofuscin in the cytoplasm increases with time
      the model makes ambiguous predictions.
  f3: lytic-enzyme decreases with ros in the lysosome
  p3: the model predicts the same relation.
  f4: h2o2 does-not-change with ros in the lysosome
  p4: the model predicts the same relation.
> disable c1 [junk-protein decreases with degradation-rate in the lysosome]
> predict
  f1: lipofuscin in the lysosome increases with time
  p5: the model predicts the same relation.
  f2: lipofuscin in the cytoplasm increases with time
  p6: the model predicts the same relation.
  f3: lytic-enzyme decreases with ros in the lysosome
  p7: the model predicts the same relation.
  f4: h2o2 does-not-change with ros in the lysosome
  p8: the model predicts the same relation.

Status of the Interactive System

Our modeling environment is still under development, but the system already lets users:

- Load, examine, and alter a qualitative model of aging
- Answer queries about how one quantity affects another
- Compare the model's predictions to observed facts
- Determine effects of model revisions on predictions

Also, we have initial encodings for the lysosomal, nuclear, mitochondrial, and extracellular portion of Furber's diagram.

We have not yet implemented pathway dominance, but we have the basic machinery to support models of aging.

Generality of the Approach

Our framework for biological modeling should prove useful in any domain for which:

- Diverse results require abstract, qualitative models that still provide explanations;
- Scientists have at least informal hypotheses about causes of known phenomena; and
- Experimental and temporal influences on quantities are primarily monotonic in character.

Candidates for other applications include age-related diseases like Alzheimer's, Parkinson's, and many forms of cancer.

Intellectual Influences

Our approach to computational biological aides incorporates ideas from many traditions:

- formalizations of biological knowledge (e.g., EcoCyc, 2003)
- qualitative reasoning and simulation (e.g., Forbus, 1984)
- languages for scientific simulation (e.g., STELLA, PROMETHEUS)
- Web-based tools for biological visualization (e.g., KEGG)
- interactive for biological processing (e.g., BioBike, 2007)

However, it combines these ideas in novel ways to assist in the construction of system-level models of aging.

Directions for Future Research

Our effort is still in its early stages, and we need further work to:

- provide a *graphical interface* for viewing and using models
- augment system to group causal influences into *processes*
- extend notation to include *contextual* and *arithmetic conditions*
- evaluate the software's actual usability for biologists
- apply the approach to a variety of age-related diseases
- make the system *accessible remotely* over the Web
- support *community-based development* of models

Together, these changes should make our interactive system a powerful computational aid for biological researchers.

Key Contributions

In summary, we are developing an interactive environment for biological modeling that supports:

- Abstract yet *communicable* models of aging that specify clear relationships among biological quantities;
- Interpret these models in ways that answer queries, generate predictions, and *explain phenomena*; and
- *Cumulative extension* of models to improve their coverage with little training and effort.

We are still developing the system, but a more advanced version would offer many benefits to the biology community.