IOWA STATE UNIVERSITY

OAK RIDGE National Laboratory

Myra B. Cohen[1] (mcohen@iastate.edu), Blessing Ngara[1], Robert W. Cottingham[2]
[1]Iowa State University, [2]Oak Ridge National Laboratory

KBase PREDICTIVE BIOLOGY

Dependable, Explainable, Reusable, AI-Driven Computational Biology

ABSTRACT

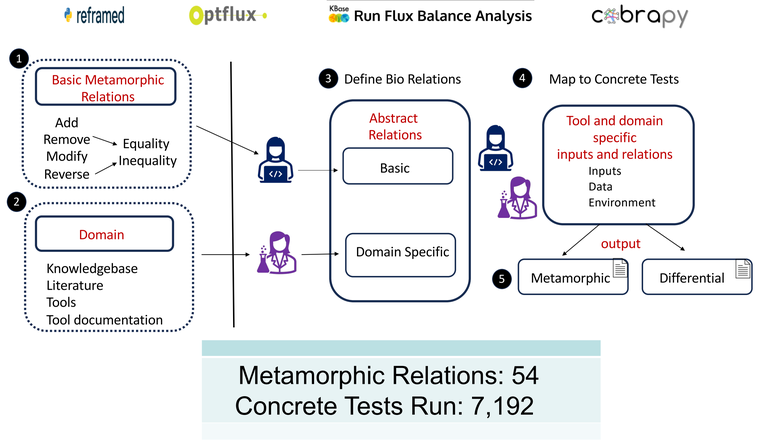
Biological research has become dependent on computational tools to help guide hypothesis generation and experimentation. However, many tools may be difficult to use, and their results may be hard to reproduce, explain, or validate. This project proposes to build a bridge between state-of-the-art software testing techniques, biologists, and bioinformaticists within the context of a widely used, DOE-supported open software platform, the Department of Energy Systems Biology Knowledgebase (KBase). We are designing meta-computational approaches that are accessible to different types of users. Our goals are to (1) enable and empower biologists to explore tools that expose reasoning and parameter settings, increasing confidence and trust in their results; (2) build embedded software testing techniques that help identify faults in predictive biological programs to improve their reliability and prediction accuracy; (3) provide tools to help automatically repair faulty code for the tool developer; and (4) design families of scientific workflows as reusable building blocks.

METAMORPHIC TESTING

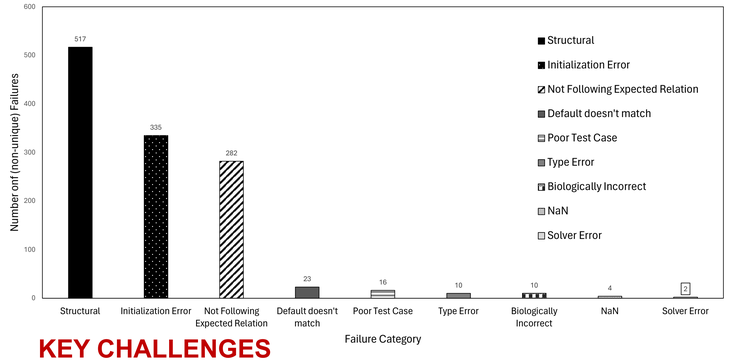
Base test case: Flux balance analysis with Carbon-D-Glucose media
Follow on test: Remove carbon sources from media
Evaluate: Follow on test should have lower (or no growth)

MOTIVATION

CASE STUDY on FOUR FLUX BALANCE ANALYSIS TOOLS [1]



Metamorphic Relations: 54
Concrete Tests Run: 7,192



KEY CHALLENGES

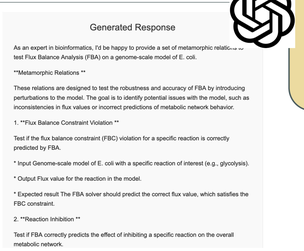
1. Automation is Difficult
2. Domain Knowledge is Needed to Build Useful Tests
3. Explainability and Exploration Approaches are Needed

WORK IN PROGRESS

1. Guided Metamorphic Testing
2. Automated Program Repair
3. Test Exploration/Explanation

1. Guided Metamorphic Testing

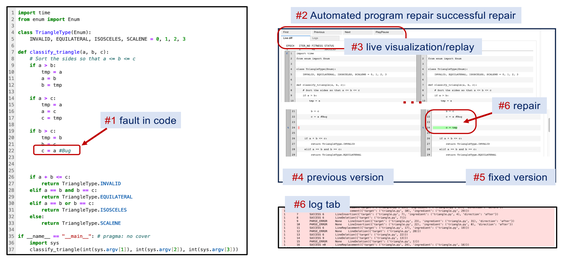
PROMPT "Suppose you are an expert in bioinformatics. Please provide a set of metamorphic test relations for flux balance analysis on a genome scale model of *E. coli*. Keep going until done"



Retrieval Augmented Generation (RAG)

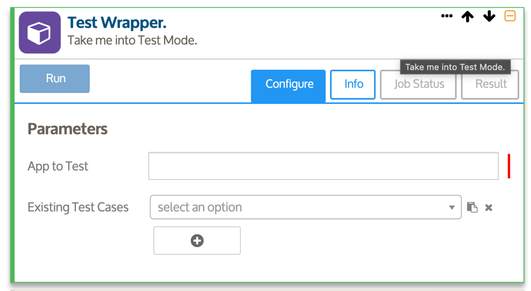
Combine domain specific documents with general prompting to generate test cases

2. Automated Program Repair



Automated program repair and visualization in notebooks

3. Test Exploration/Explanation



Wrapper for apps with testing framework and explanation breadcrumbs

REFERENCES
1. A. Marsh, M.B. Cohen, R.W. Cottingham, A Tale from the Trenches: Applying Metamorphic and Differential Testing to Bioinformatics Software, *IEEE International Conference on Software Testing (ICST)* Industry Track, April 2025.

U.S. DEPARTMENT of ENERGY

Acknowledgments: This work was supported by the Office of Biological and Environmental Research (BER) in the U.S. Department of Energy (DOE) Office of Science award DE-SC0025510 and by BER support of the DOE Systems Biology Knowledgebase (kbase.us) at Oak Ridge National Laboratory, managed by UT-Battelle, LLC, under contract DE-AC05-00OR22725. We thank Alexis Marsh for research on Metamorphic testing of Flux Balance Analysis Tools[1] and early discussions of this work.