

Predicting Metabolic Pathways: A Bioinformatics Approach

Imran Shah

pathway.uchsc.edu/Shah
School of Medicine
University of Colorado

Project Goal

Metabolic prediction to aid in "targeted and purposeful alteration of metabolism"

Example problems:

Elucidating the metabolic pathways of microbes from high throughput biological data

Understanding the fate of toxic substances released into the environment.

Harnessing biocatalysis to synthesize complex compounds

What is Metabolic Prediction ?

Elucidation of metabolic pathway(s), from input compound to output, using partial data on genes, proteins, metabolites and known pathways.

Hypothesis: Metabolic routes can be predicted from data using biological knowledge

Examples of biological knowledge

The genome encodes genes

Biocatalytic functions can be assigned to putative proteins

Compounds undergo a variety of enzyme-catalyzed transformations

Metabolic pathways consist of a series of enzyme-catalyzed transformations

Important Challenges in Metabolic Prediction

High throughput biology is yielding unprecedented amount of data but powerful inference tools are necessary for prediction

Important challenges to metabolic prediction:

Organizing metabolic information: Metabolic data is disparate and complex -- spanning substructure details of compounds to pathways of enzyme-catalyzed transformations

Classifying biocatalysis: To infer with biocatalytic data it must be represented formally

Correlating catalytic function with protein properties: Assigning biocatalytic function(s) to putative proteins is difficult -- requires human intervention and hard to automate

Metabolic pathway synthesis: "Pathway reconstruction" is generally used; novel metabolic routes are difficult to infer

Bioinformatics approach involves four interacting components:

Metabolic Knowledgebase (MKB) to semantically integrate data from chemical structures to complete pathways; incorporate data from public domain sources

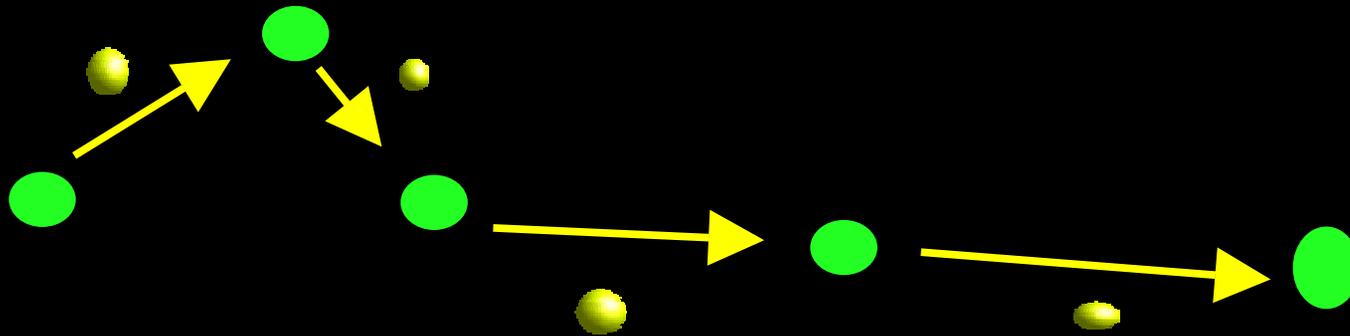
Biocatalysis classification scheme (BCS) by induction over existing enzymatic knowledge to predict transformations

Biocatalysis assignment tool (BAT), using machine learning methods, to accurately assign biocatalytic functions to putative proteins.

Metabolic pathway synthesis (MPS) to generate pathways using heuristic search

Abstracting Metabolic Pathways ...

- Compounds connected by transformations, catalyzed by specific enzymes
- Pathways abstracted as a graph consisting of nodes (compounds) and arcs (transformations).



I. Metabolic Knowledgebase (MKB)

Goal: Integrating large amounts of biomolecular data into a semantically coherent system is necessary for inference.

Hypothesis: Ontologies can formally express metabolic information, data sources, learners and induced knowledge and be implemented as a KB

Strategy:

Use ontology to formalize metabolic knowledge and implement large flexible and scalable metabolic knowledgebase to capture multiple levels of chemical/biological information

Develop inference tools to support complex metabolic queries

II. Biocatalysis Classification Scheme

Goal: Formalizing biocatalysis is necessary to assign functions to putative proteins and to predict metabolic pathways

Hypothesis: Biocatalysis can be formalized on the basis of compound substructural features and induced from known enzyme-catalyzed transformations

Strategy:

Use MKB to identify subset of biocatalytic functions based on pathways of interest (Purine synthesis, Pentose phosphate pathway)

For each biocatalytic function use MKB to find relevant compounds and infer their substructural features

Use patterns of compound substructural features to induce BCS; Integrate BCS into MKB

III. Biocatalysis Assignment Tool (BAT)

Goal: Accurately and automatically assign biocatalytic functions to putative proteins

Hypothesis: Biocatalytic functions of proteins can be correlated to physico-chemical features conserved among family members

Strategy:

- Use MKB to identify relevant sets of proteins and their biocatalytic functions

- Identify conserved features among proteins

- Induce correlations between functions and conserved features (classifiers)

- Classifiers form the basis for BAT

- Integrate BAT into MKB

IV. Metabolic Pathway Synthesis (MPS)

Goal: Elucidate a metabolic route from input compound to output compound

Hypothesis: Plausible metabolic pathways between two compounds can be inferred by heuristic search

Strategy:

Use BCS/MKB to find the possible transformations from input compound to intermediate

Apply heuristics for limiting intermediates (uses BCS, BAT)

Using input \leftarrow intermediate, repeat above until output is equal to intermediate

Project Management Plan

| Months | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 24 | 27 | 30 | 33 | 36 |
|--------|---|---|---|----|----|-----|----|----|----|-----|----|-----|
| MKB | D | P | I | I | T | Doc | | | | | | |
| BCS | | | D | P | P | I | I | T | T | Doc | | |
| BAT | | | | D | P | I | I | T | T | Doc | | |
| MPS | | | | | | | | D | P | I | T | Doc |

D = Design

P = Prototype

I = Implement

T = Test

Progress

MKB is designed, partially implemented and being applied to manage various data sources

BCS is in development

BAT is partially implemented for protein subsequence features

MPS is in the design phase

Summary

- Ontologies can integrate multiple levels of biochemical data to support complex metabolic inferences
- Machine learning methods can induce protein sequence properties that are related to biocatalytic functions
- Biocatalytic functions can be formalized computationally to make inferences about transformations and pathways
- Novel biocatalytic functions and novel pathways can be predicted by Metabolism, *in silico*

Acknowledgements

Project members

Ashok Kolaskar, Co-PI (University of Pune, India)

Ronald Taylor, Computer Scientist (CU, SOM)

Scott Richard, Graduate Student, Bioninformatics (CU, SOM)

Shilpa Rao, Research Associate, Biochemistry (CU, SOM)

Domain Experts

Lawrence Hunter, Bioinformatics (CU, SOM)

Dhinakar Kompala, Metabolic Engineering (CU, Boulder)

Shelley Copley, Mechanistic Enzymology (CU, Boulder)

Funding

National Science Foundation (NSF)

Department of Energy (DoE)

Office of Naval Research (ONR)