McCormick

Northwestern Engineering

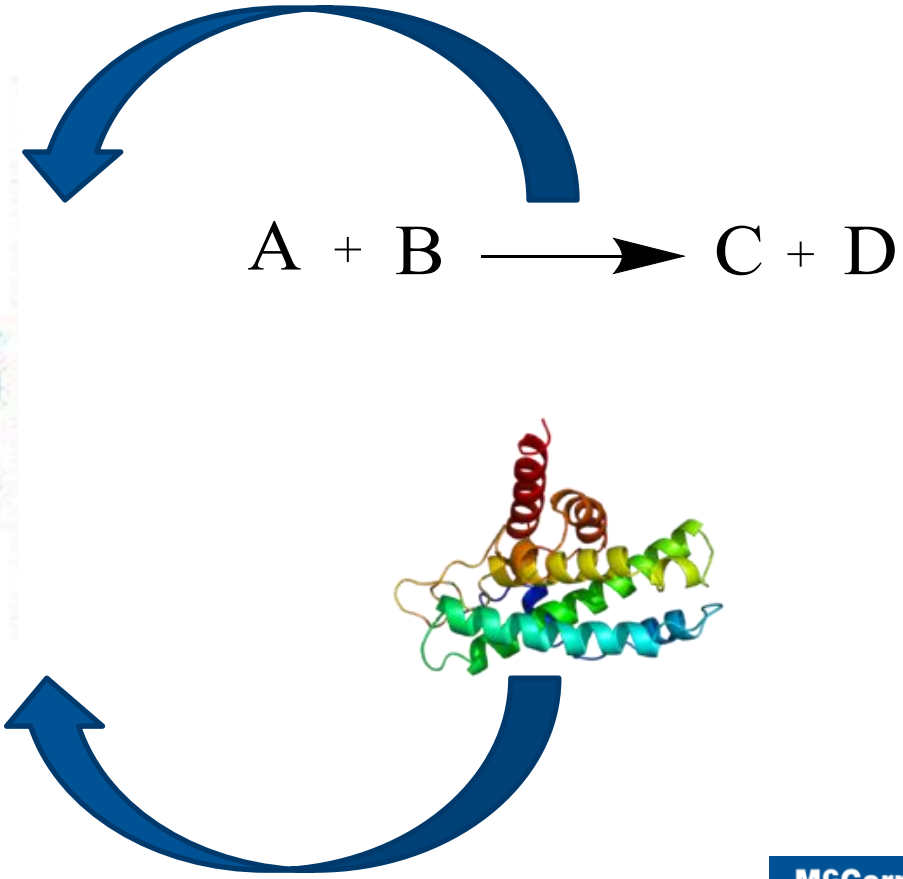
The Known Unknown: Computational Identification of Promising Enzymatic Reactions and Associated Genes

Andrew Stine

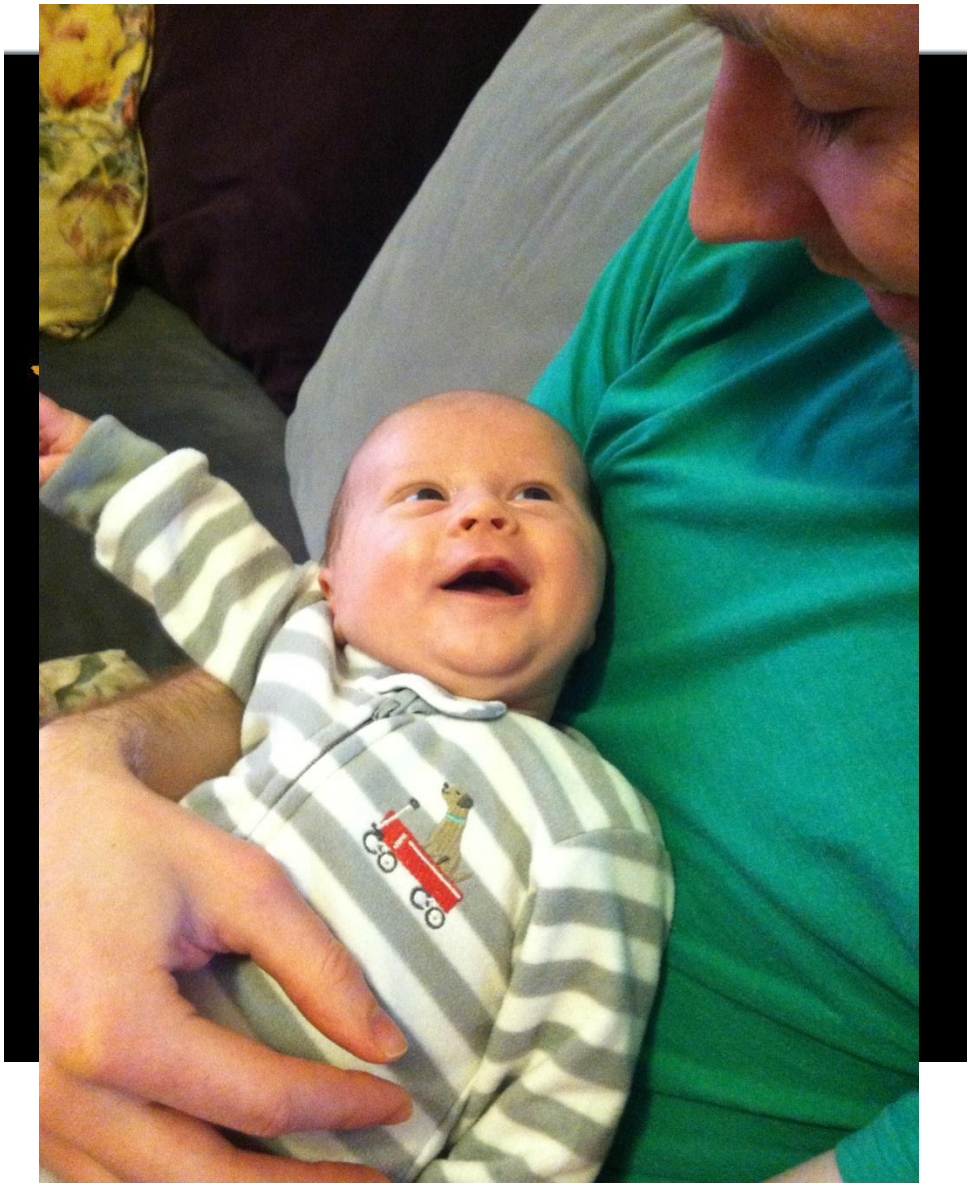
McCormick

Northwestern Engineering

What is an Enzyme?

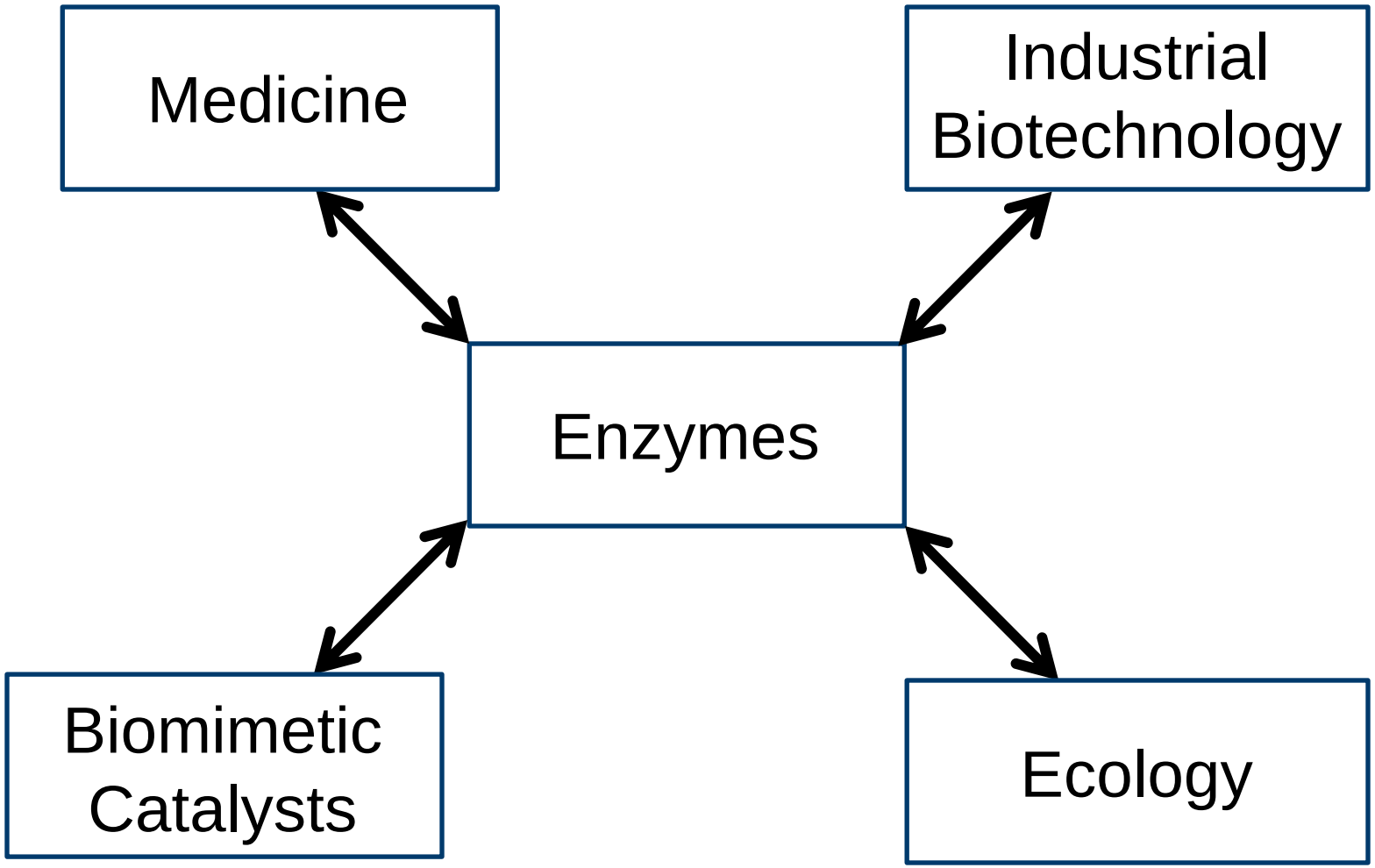


Importance of Enzymes



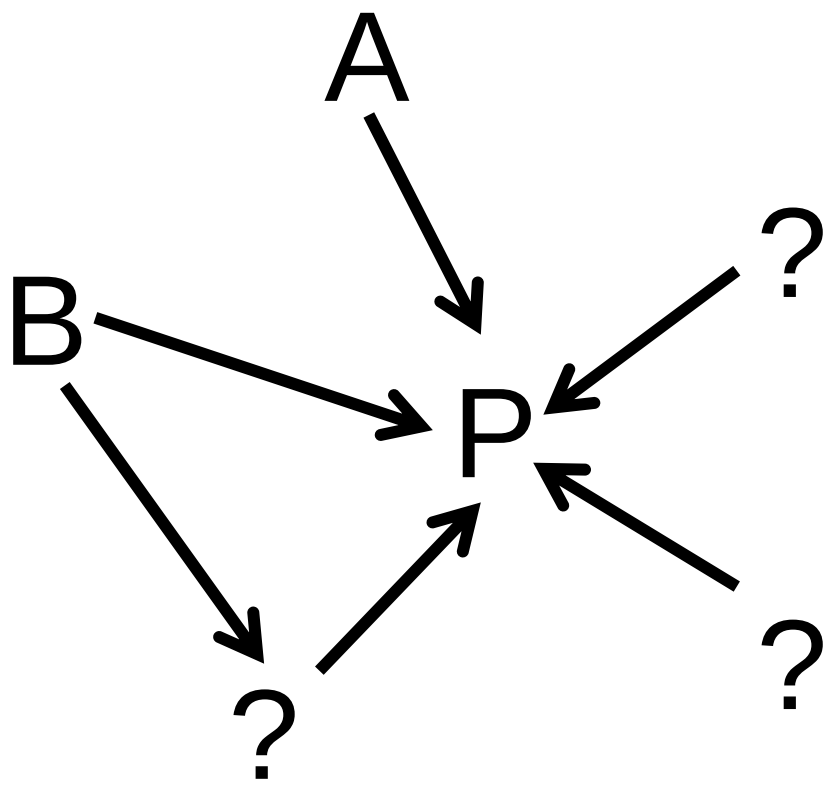
McCormick
Northwestern Engineering

Applications of Enzymes

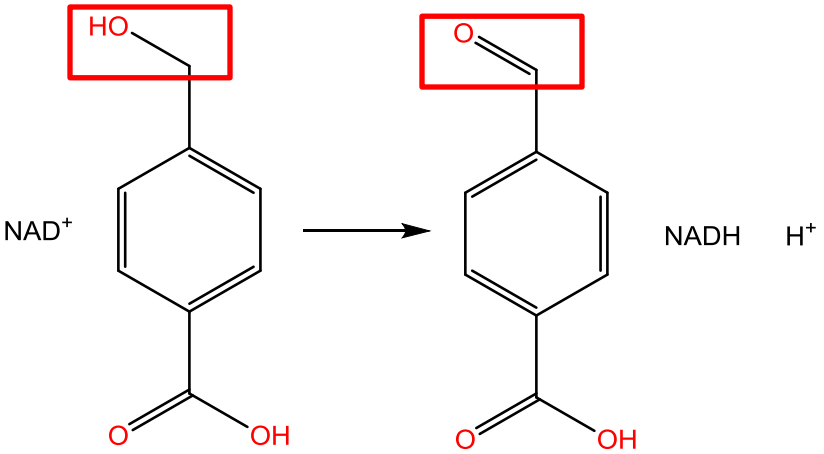


Enzyme Research: An Unknown Unknown

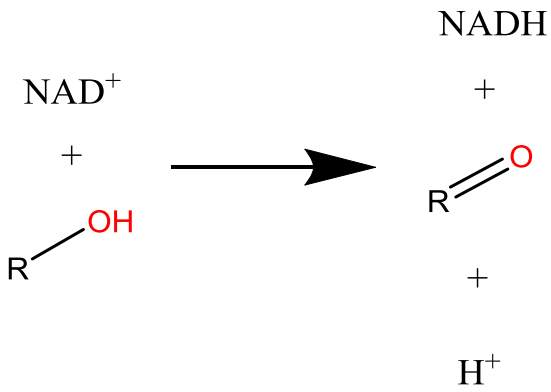
- What makes enzymatic research difficult?
- An unknown unknown
- Still only know a small fraction of enzymatic chemistry
- Difficult to know what reactions are possible but just have not been observed
- Need some way to make predictions about enzymatic chemistry we have yet to observe

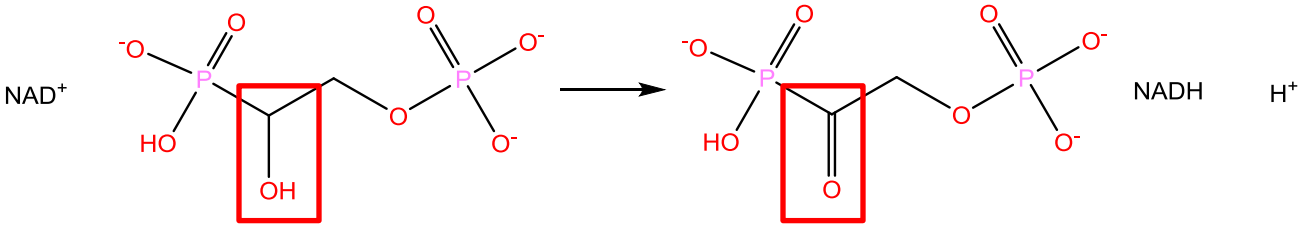


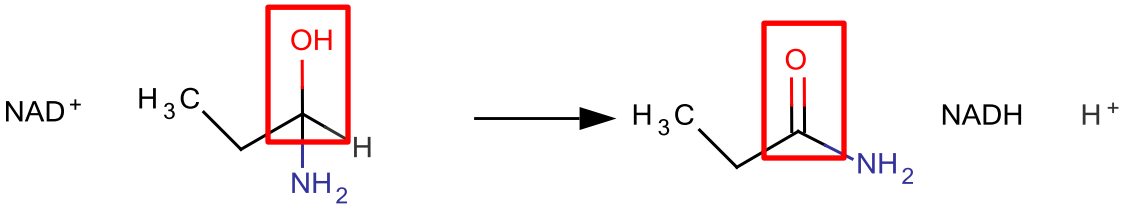
A Known Unknown



Reaction Rule (Operator)

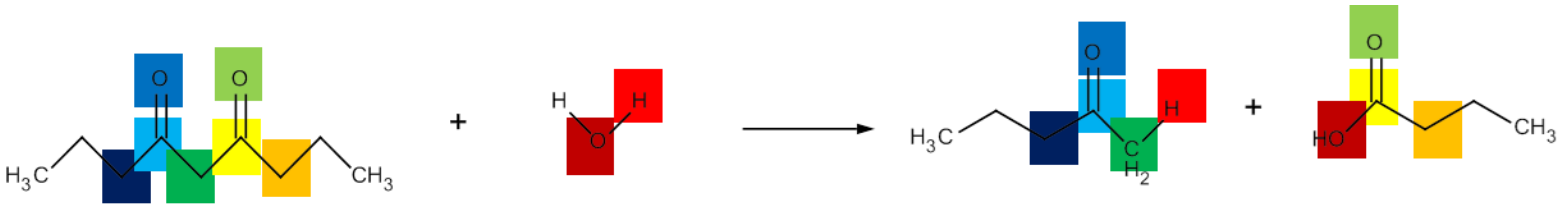






McCormick
Northwestern Engineering

Translating Rules for a Computer



Site BEM

	O	H	C	C	O	C	C	O	C
O	0	1	0	0	0	0	0	0	0
H	1	0	0	0	0	0	0	0	0
C	0	0	0	1	0	0	0	0	0
C	0	0	1	0	2	1	0	0	0
O	0	0	0	2	0	0	0	0	0
C	0	0	0	1	0	0	1	0	0
C	0	0	0	0	0	1	0	2	1
O	0	0	0	0	0	0	2	0	0
C	0	0	0	0	0	0	1	0	0

+

Operator Matrix

	O	H	C	C	O	C	C	O	C
O	0	-1	0	1	0	0	0	0	0
H	-1	0	0	0	0	1	0	0	0
C	0	0	0	0	0	0	0	0	0
C	1	0	0	0	0	-1	0	0	0
O	0	0	0	0	0	0	0	0	0
C	0	1	0	-1	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0

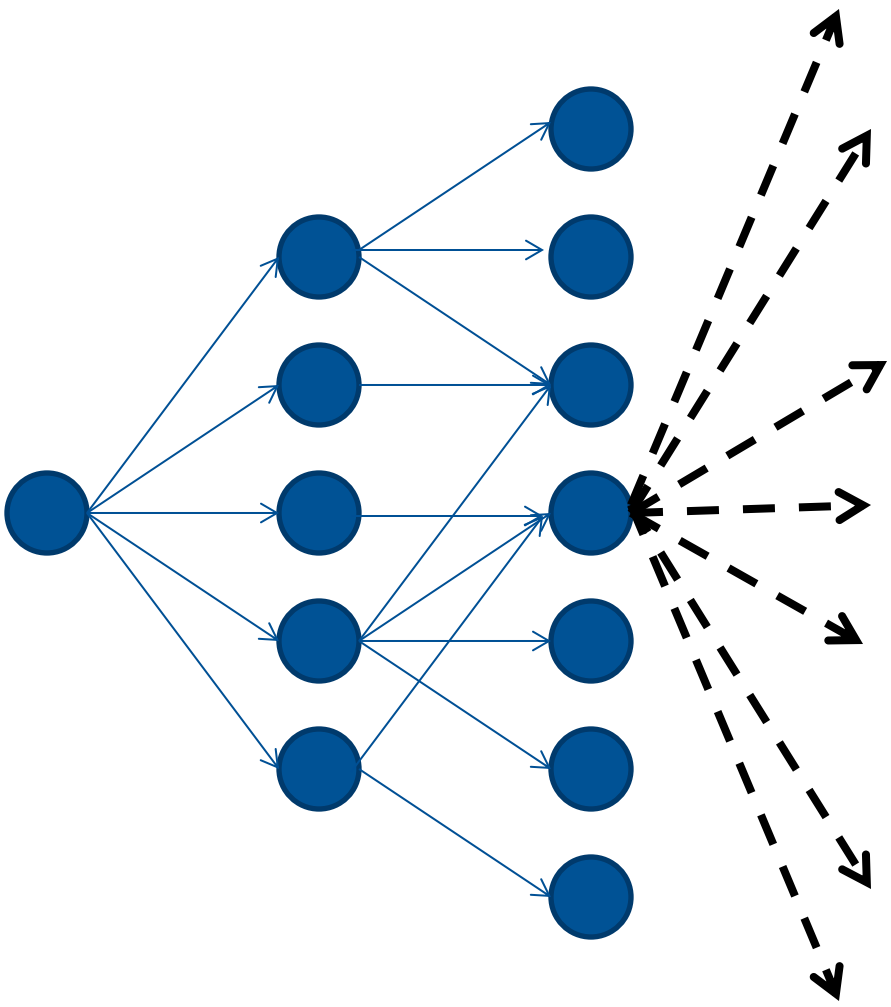
=

Product Matrix

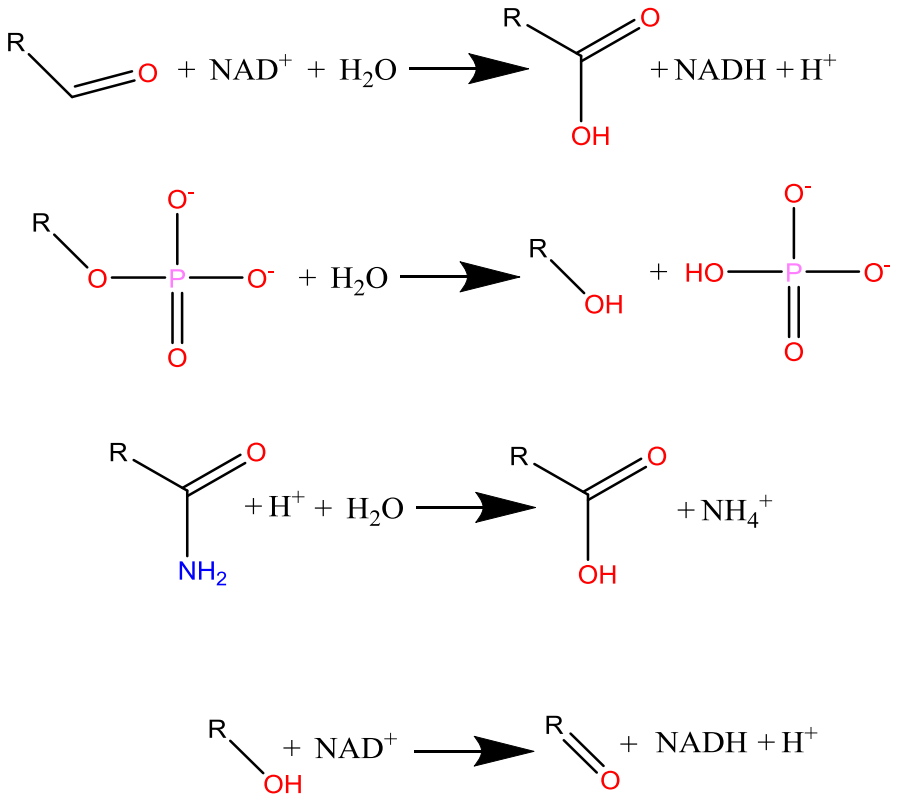
	O	H	C	C	O	C	C	O	C
O	0	0	0	1	0	0	0	0	0
H	0	0	0	0	0	1	0	0	0
C	0	0	0	1	0	0	0	0	0
C	1	0	1	0	2	0	0	0	0
O	0	0	0	2	0	0	0	0	0
C	0	1	0	0	0	0	1	0	0
C	0	0	0	0	0	1	0	2	1
O	0	0	0	0	0	0	2	0	0
C	0	0	0	0	0	0	1	0	0

BNICE Reaction Network Generation

- Biological Network Integrated Computational Explorer[1]

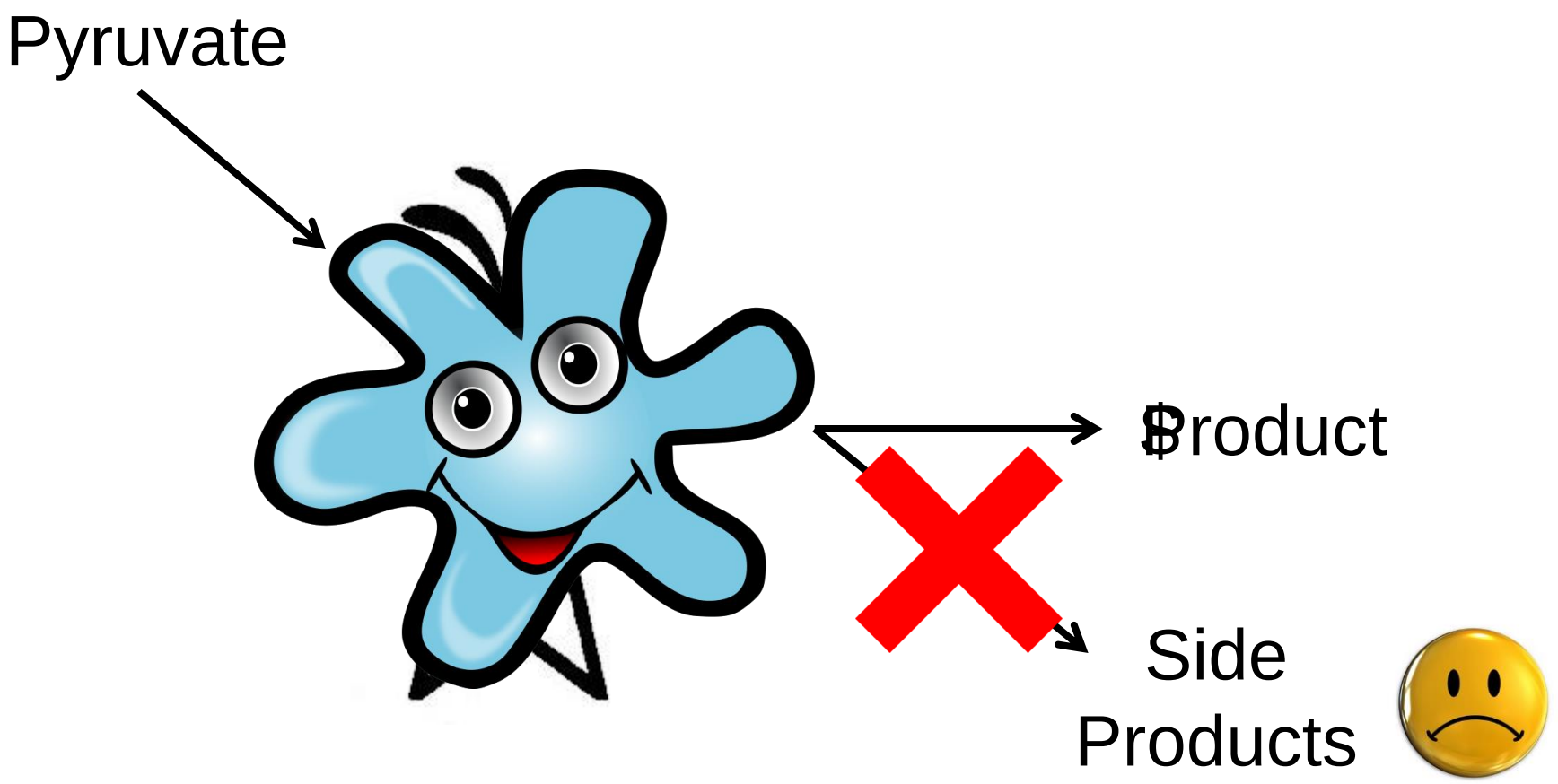


Enzymatic Reaction Rules



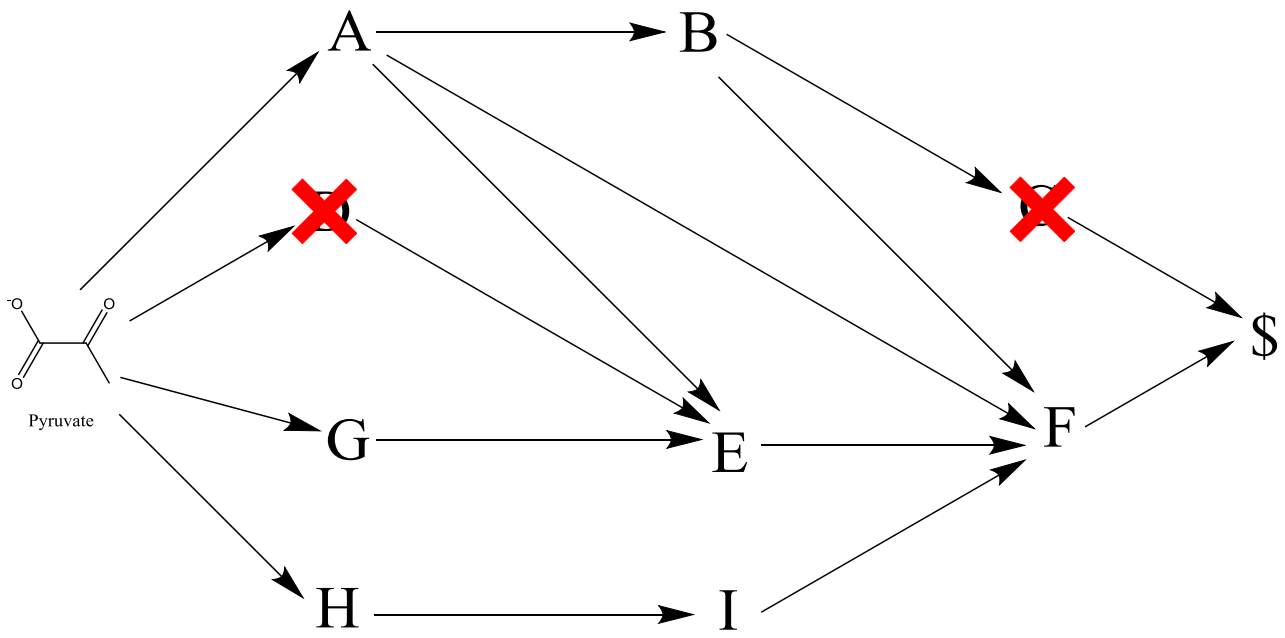
McCormick
Northwestern Engineering

Example Application



Analysis of Longer Pyruvate Pathways

4 Reactions

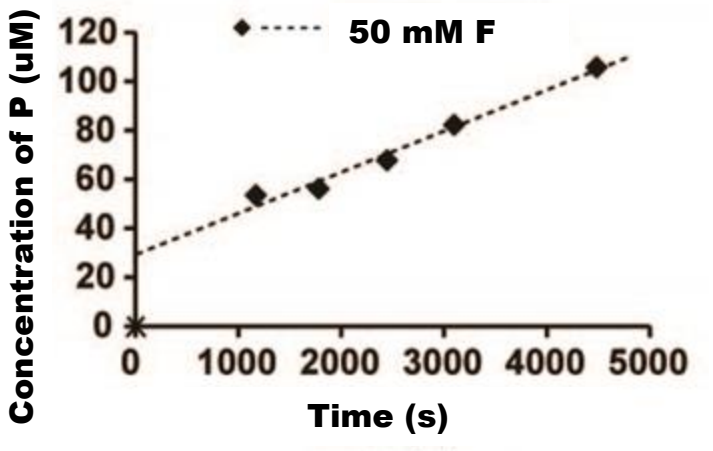


5 Reactions

Reactants of the final step	# of pathways
F	1259
Y	36
Z	32
C	24
Total Pathways	1410

Experimental Confirmation

F → $ → Enzyme 3 →



Unpublished work by Miaomin Zhang, Soo Ro, and Keith Tyo

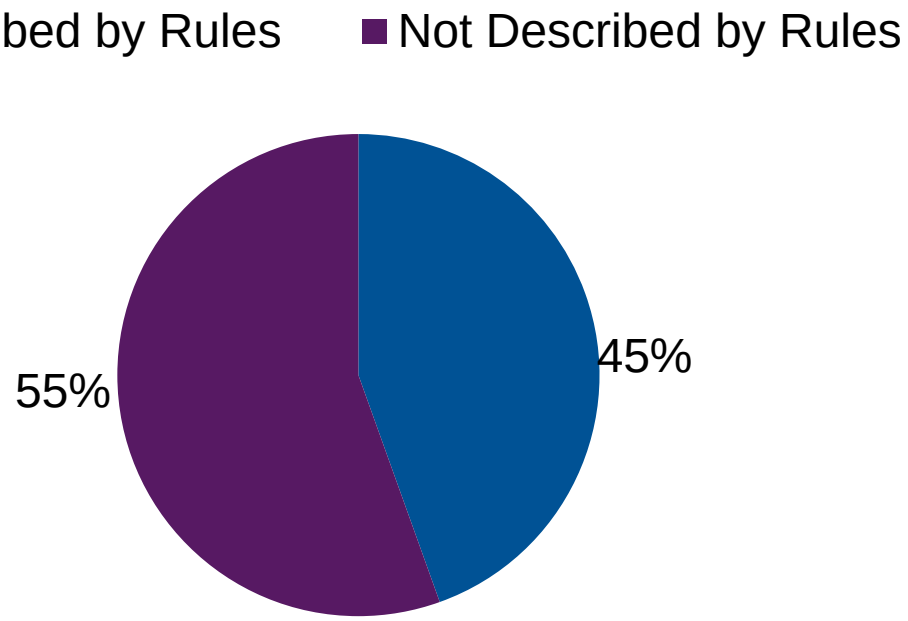
McCormick Northwestern Engineering

Creating Biochemical Reaction Rules

- Problems with this method:
  1. Potential for human error
  2. Rules may be designed for a different application
  3. **Slow**

Results for Manual Reaction Rules

- Applying reaction rules to KEGG[1]:



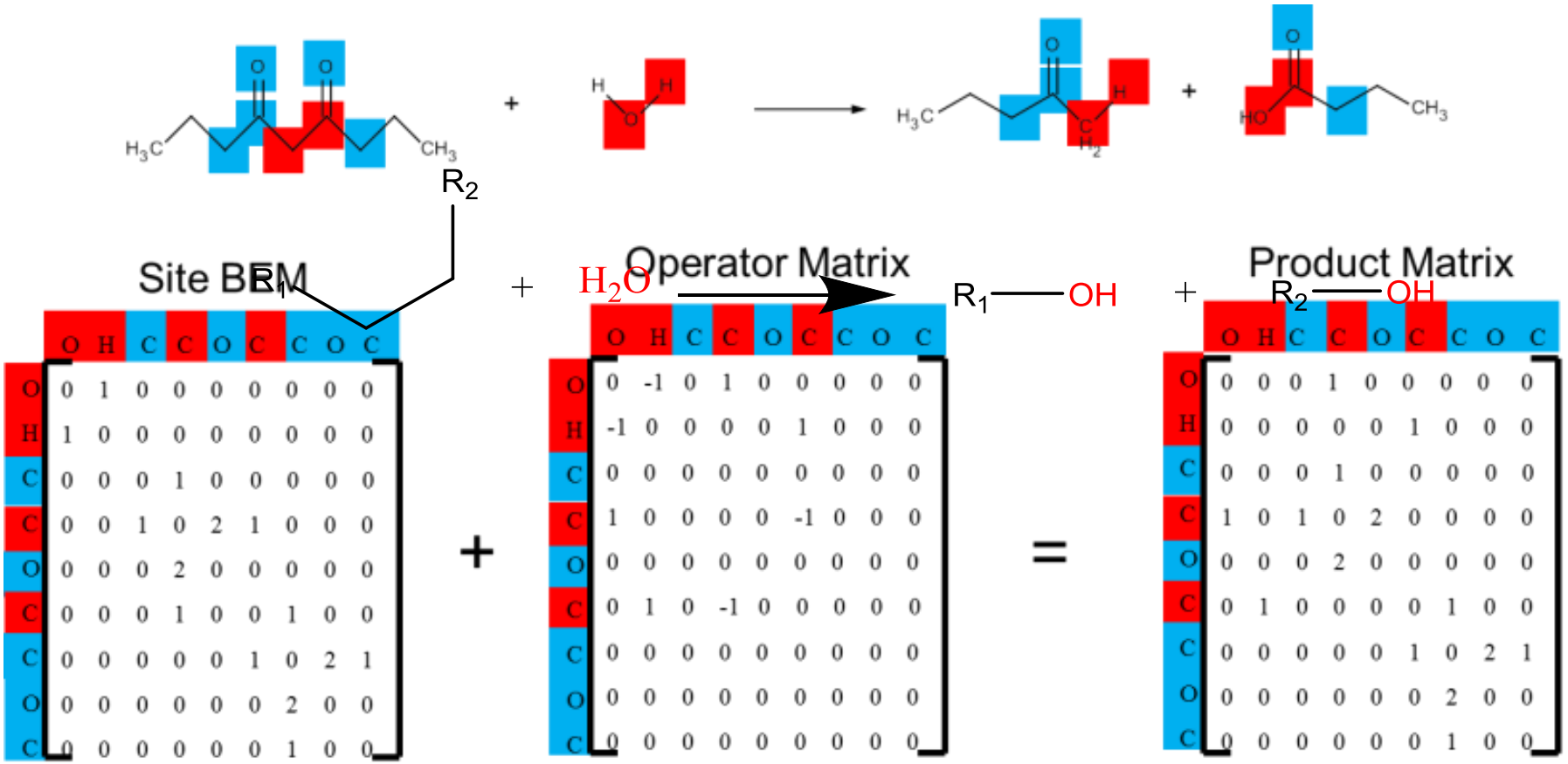
- Can we do this computationally?

Yes!

1. Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research, 28*(1), 27-30.

Automatic Reaction Rule Generation

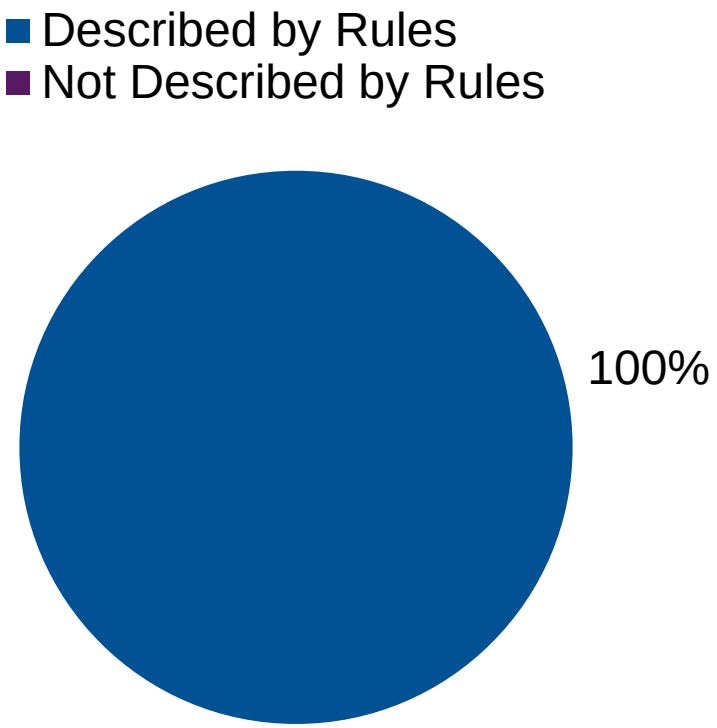
- Simplest rule consists of bonds broken across reaction
- Can obtain from our matrix representations



McCormick
Northwestern Engineering

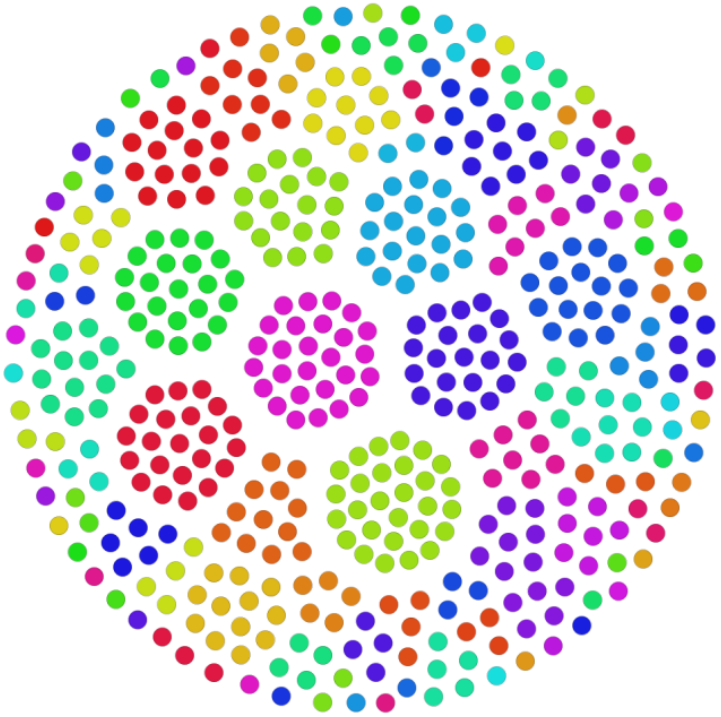
Automatic Operator Results

- Applying automatic reaction rules to MetaCyc[1]:



1. Karp, P. D.; Riley, M.; Paley, S. M.; Pellegrini-Toole, A., The MetaCyc database. *Nucleic Acids Res* **2002,** *30* (1), 59-61.

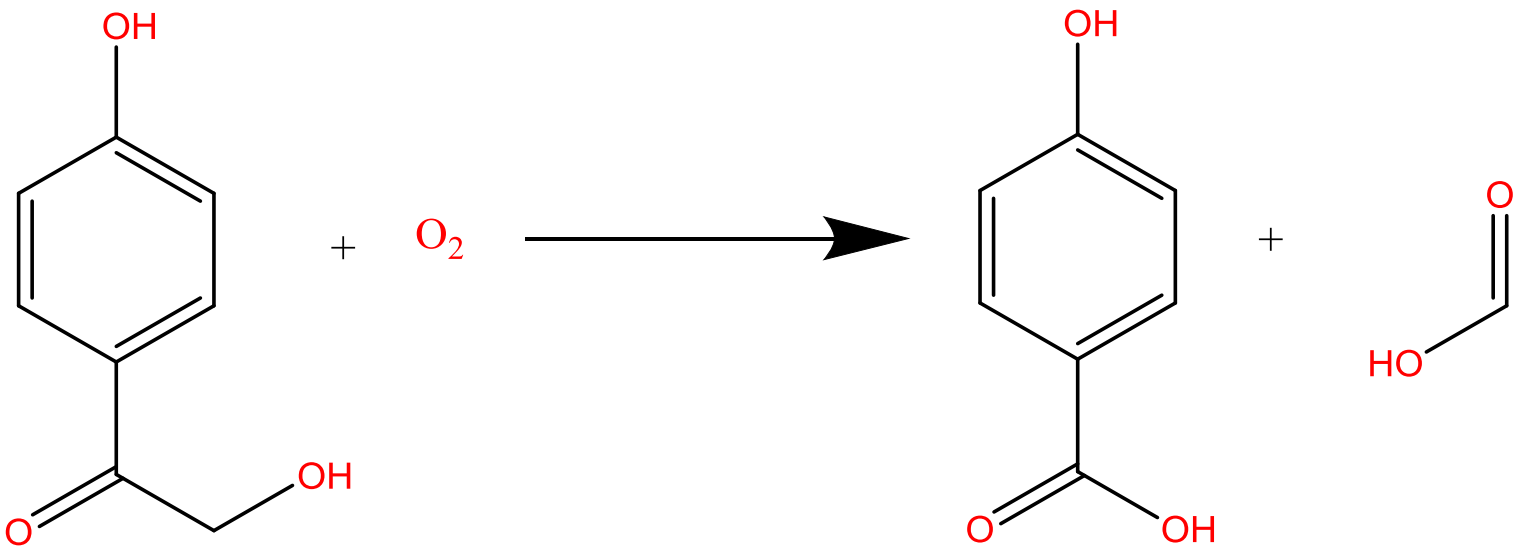
Distribution of Reaction Rules



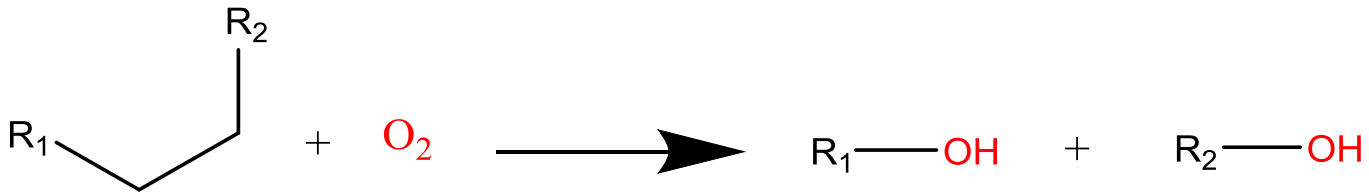
Visualization of MetaCyc. Each dot represents ten reactions, each color represent a rule

Number of Reaction Rules	Reactions Described (%)
5	14.4%
9	22.6%
13	28.5%
28	40.0%
49	49.4%
110	59.6%
215	67.5%
375	73.7%
661	80.4%
2329	100.0%

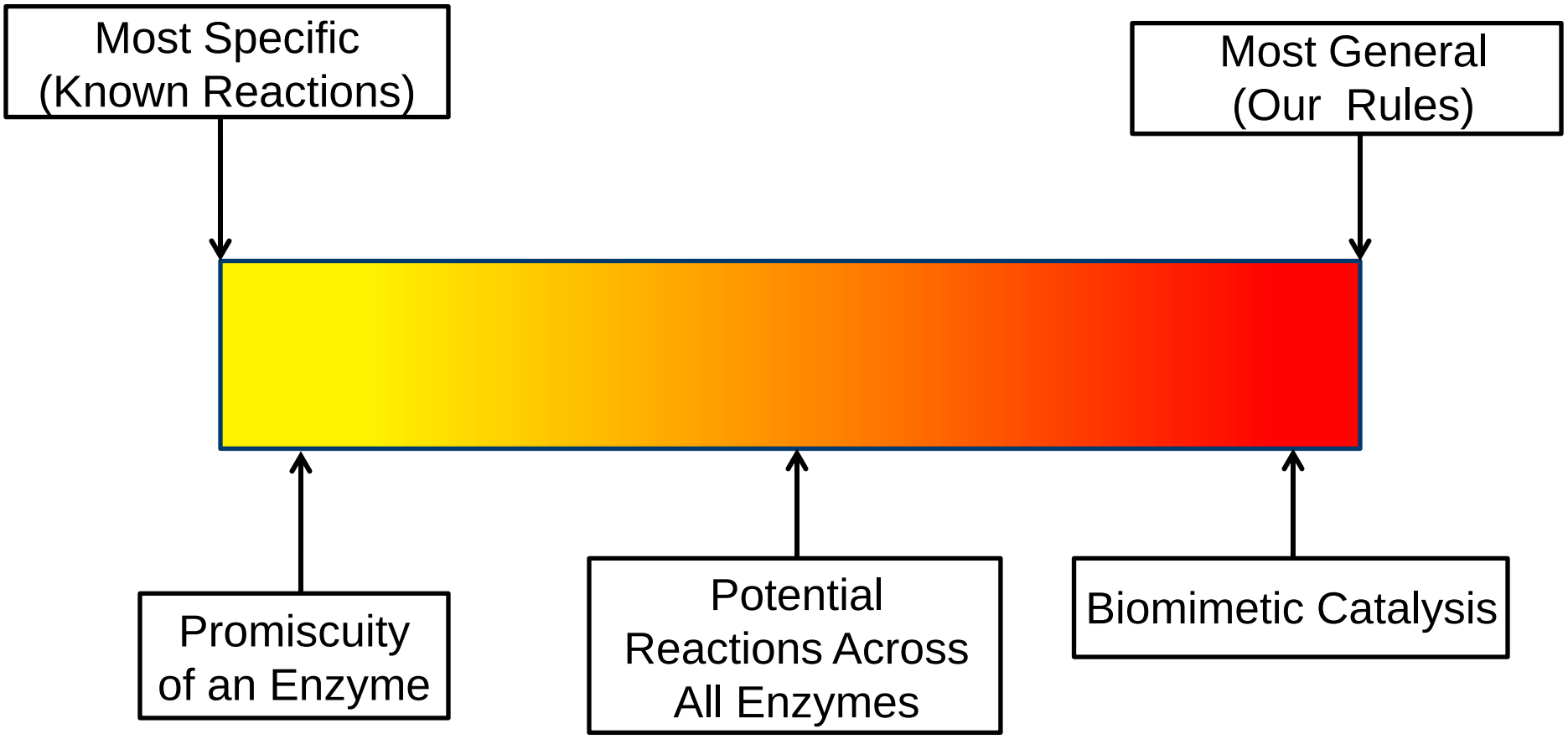
Problem with This Approach



Reaction Rule



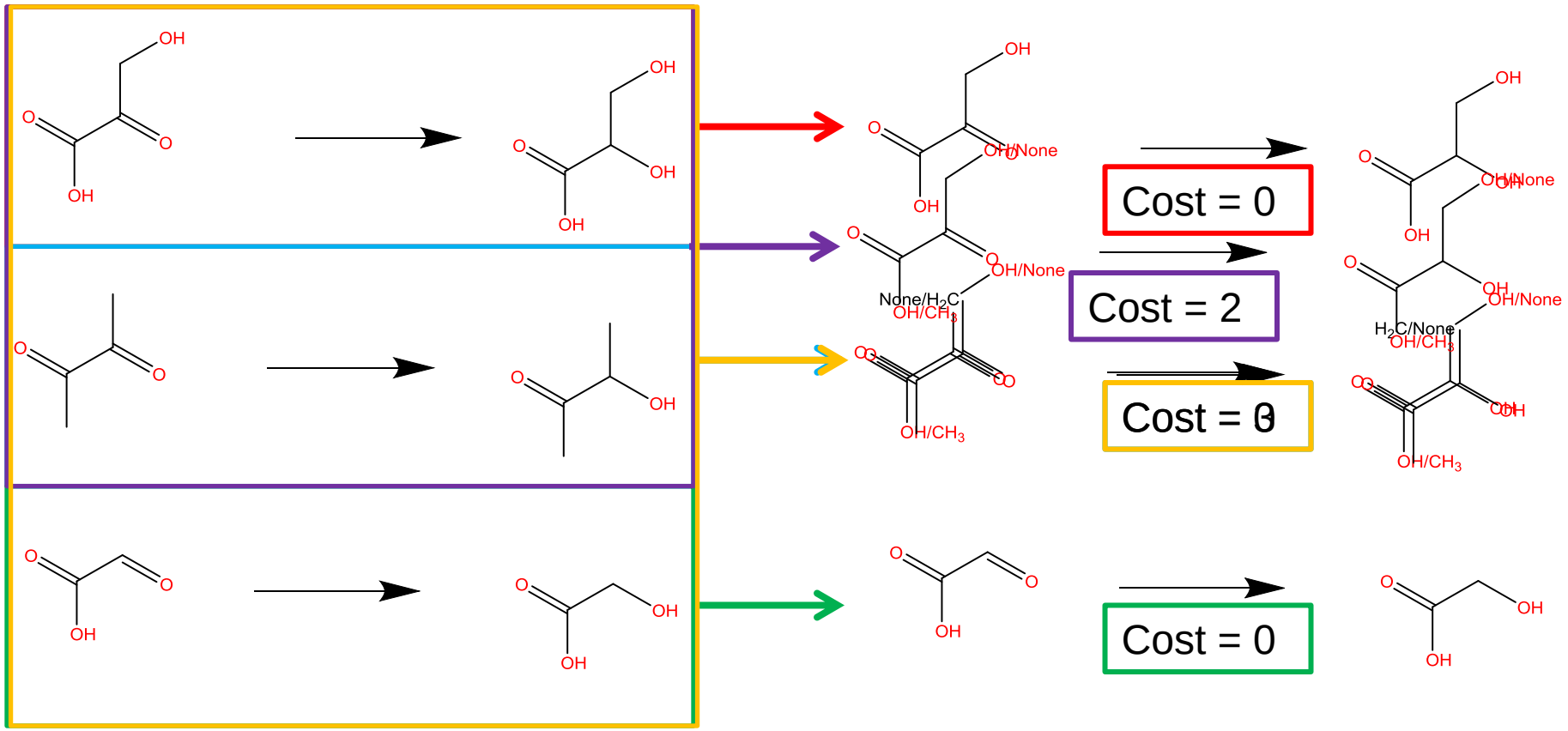
Variable Specificity



Grouping Reactions

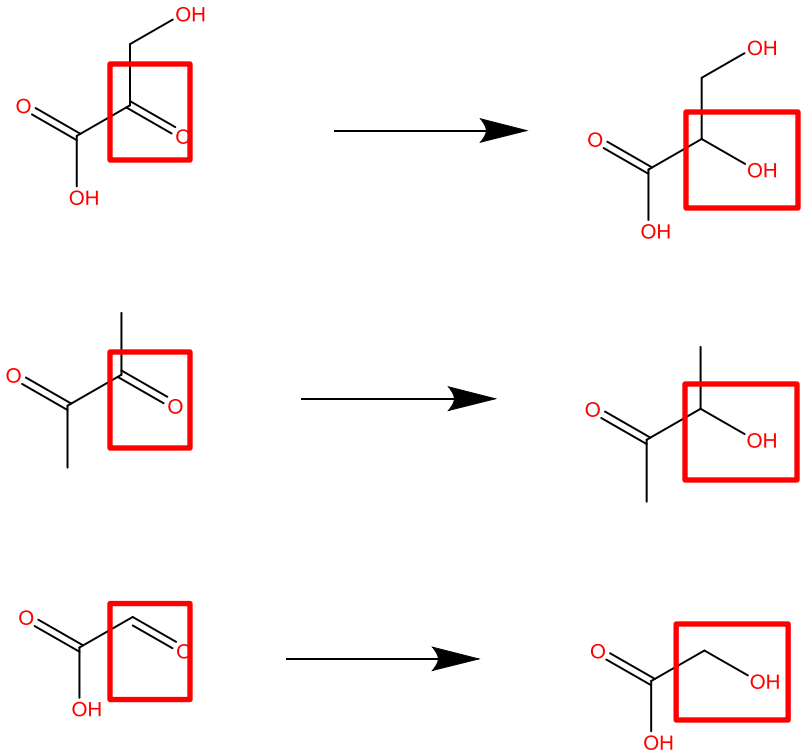
Find the Lowest Cost Rule that Describes all the Reactions in the Group

Cost = Number of Atoms that must be generalized



Grouping Reactions

Find the Lowest Cost Rule that Describes all the Reactions in the Group



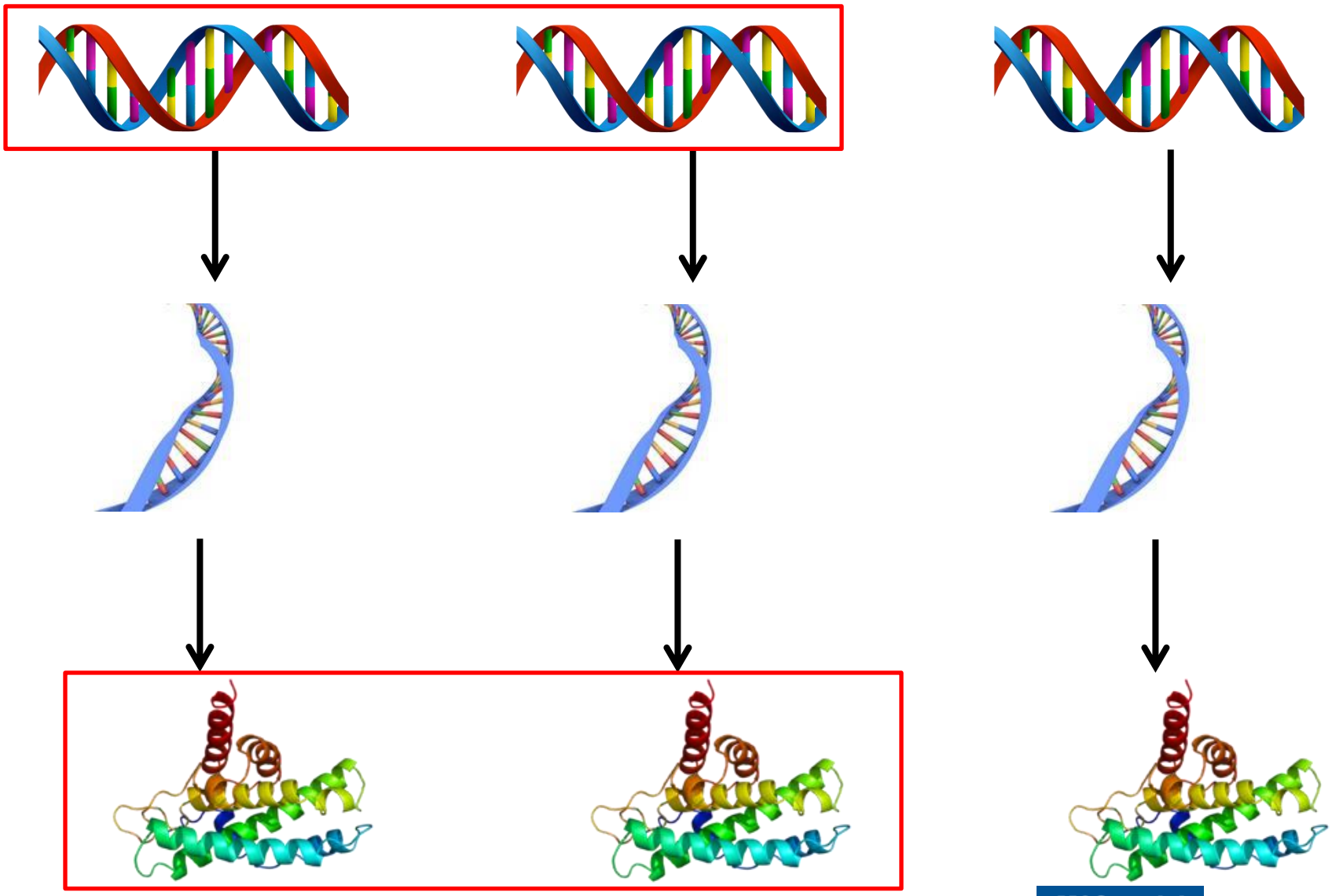
Cost = Number of Atoms that must be generalized

Simplifying Assumptions:

1. Atoms whose bonds change are the same across all reactions
2. Atoms are only allowed to be generalized with atoms that are the same distance from the atoms whose bonds change

Simplifications allowed us to develop a dynamic programing algorithm to quickly solve this problem

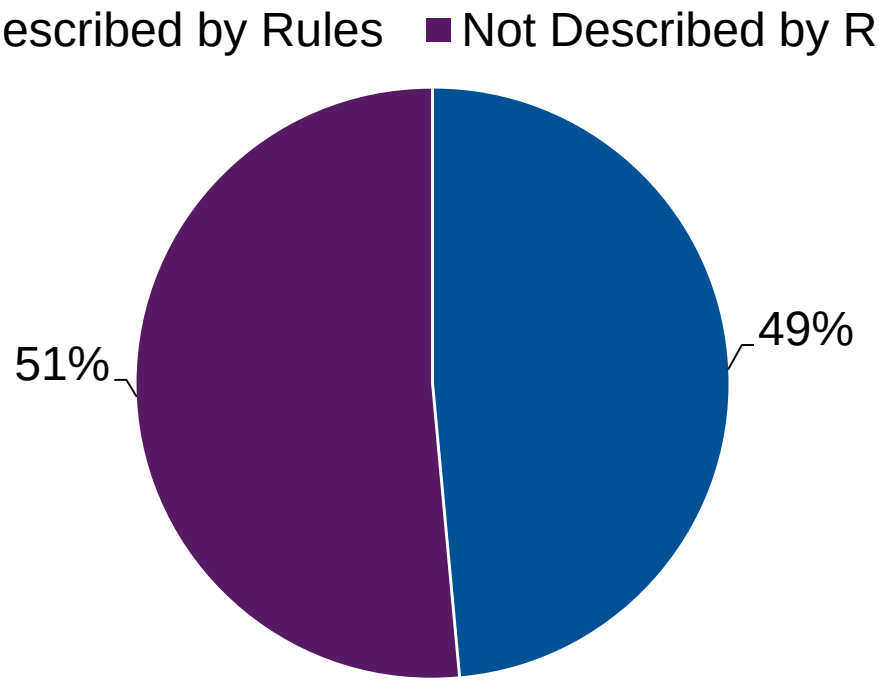
Genetic Groupings



McCormick
Northwestern Engineering

Genetic Grouping Results

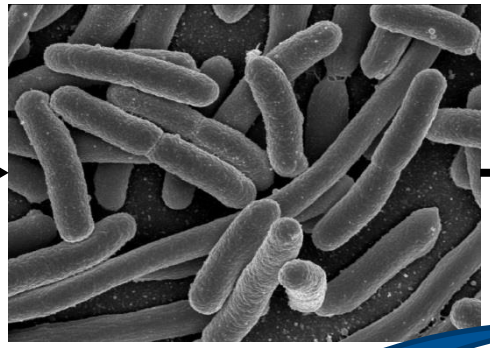
- Used Conserved Domain Database[1] Superfamilies to group reactions by genetic similarity



1. Marchler-Bauer A, Derbyshire MK, Gonzales NR, et. al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015 Jan 28;43(Database issue):D222-2. doi: 10.1093/nar/gku1221. Epub 2014 Nov 20. [PubMed PMID: 25414356]

Future Work

Genetically Grouped Reaction Rules



List of Probable Reactions, Compounds, and Genetic Properties

Industrial Biologists

Investigate enzymatic chemistry to compounds of interest

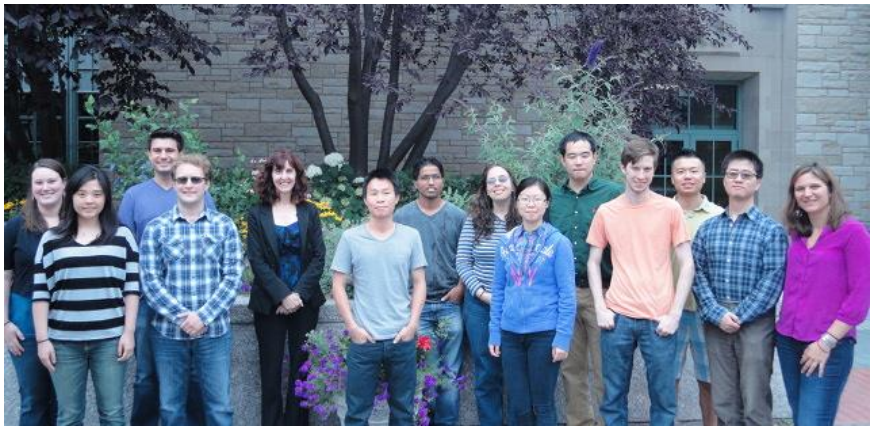
1. Rocky Mountain Laboratories, NIAID, NIH

Summary

- Enzymatic chemistry is an important but difficult topic
- Utilizing generalized reaction rules we can generate intelligent hypothesis about probable enzymatic chemistry
- We have developed an algorithm for the automatic generation of rules based upon user defined reaction groupings
- Using a genetic based grouping method has the potential to not only predict reactions but also properties of genes associated with those reactions

Acknowledgements

- Linda Broadbelt (Northwestern)
- Keith Tyo (Northwestern)
- Chris Henry (Argonne)
- Broadbelt Research Group



DOE CSGF

McCormick
Northwestern Engineering