

# Pathway Analyst

Jordan Patterson\*, Luca Pireddu, Duane Szafron, Paul Lu, Russ Greiner



\*jordan@cs.ualberta.ca

http://path-a.cs.ualberta.ca



UNIVERSITY OF ALBERTA

Part of the  
**Proteome Analyst**  
Suite of Tools



## INTRODUCTION

Given:

- a number of proteins, or a proteome

Predicts:

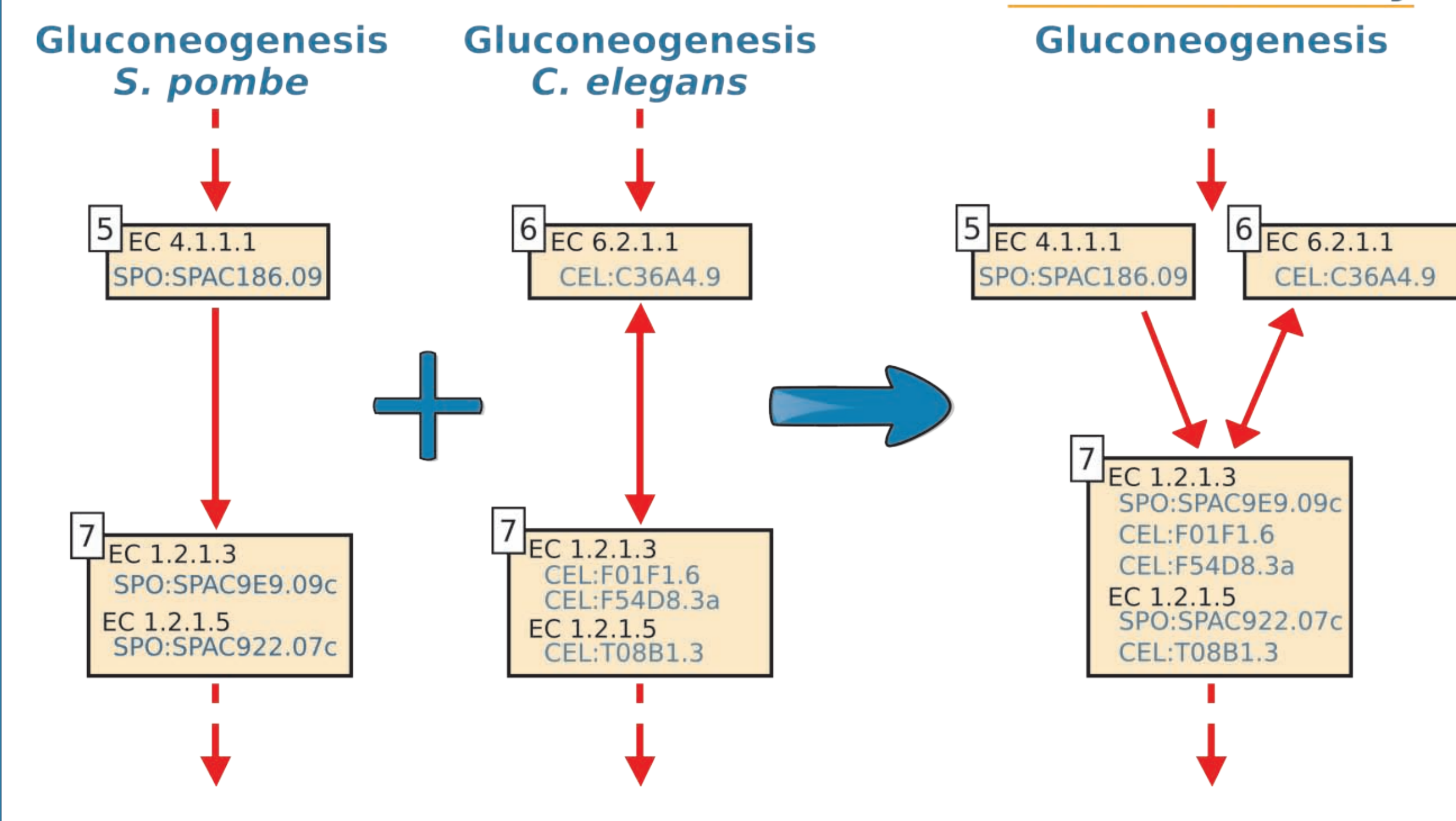
1. which pathways exist
2. which proteins catalyse each step of the pathways found

Using:

- DB of model pathways
- prediction algorithm
- specialized predictors

## Example: creating a model pathway

### Model Pathway



## THE MODEL PATHWAY

Combines multiple organism-specific variants, or instances, of a pathway. It brings together:

- the reactions that make up the instances
- the proteins that catalyze each reaction

Model pathway includes

1. all reactions
2. all catalysts from original instances.

## ALGORITHM

Input:

- model pathway (computed on-demand)
- proteome of the target organism (e.g. *A. thaliana*)

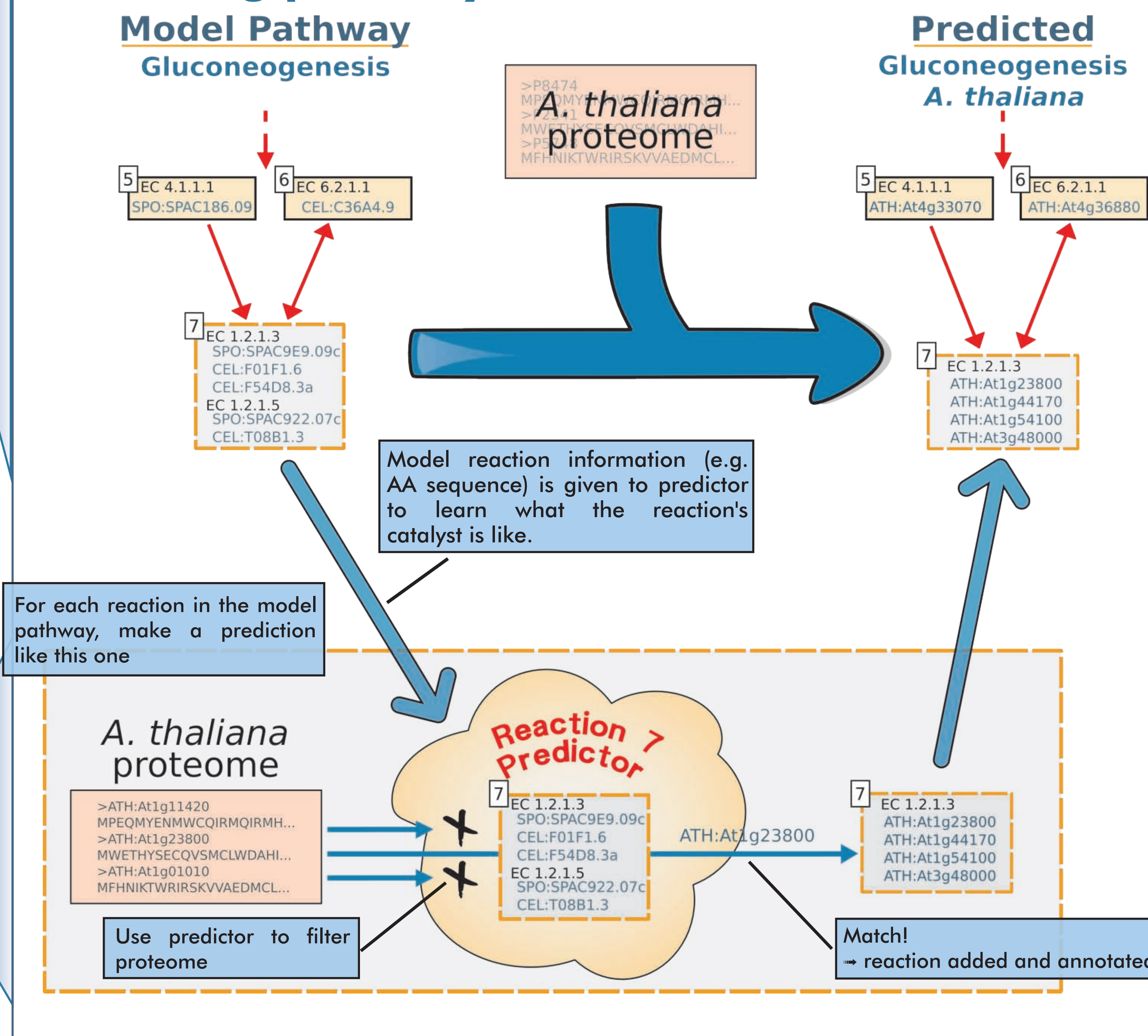
Output:

- predicted pathway

Algorithm:

For each of the pathway's reactions:  
Search for catalysts in the proteome  
Were any found?  
If yes, reaction exists in organism  
→ add reaction to predicted pathway  
→ annotate with catalysts found

## Predicting pathways



## What are those reactions?

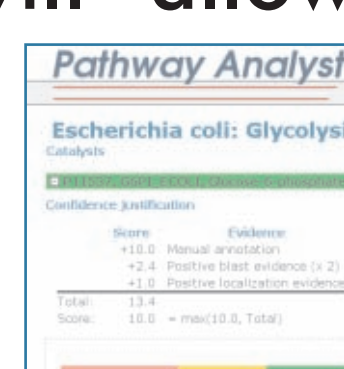
Here is a listing of the enzymatic reactions that appear in the examples.

EC	Official enzyme name
EC 1.2.1.3	Aldehyde dehydrogenase (NAD(+))
EC 1.2.1.5	Aldehyde dehydrogenase (NAD(P)(+))
EC 4.2.1.1	Carbonate dehydratase
EC 6.2.1.1	Acetate--CoA ligase

## WEB INTERFACE

We are developing a web interface for Pathway Analyst. It will allow users to:

- submit analyses
- examine their results
- easily search our DB of precomputed analysis results



## EXPERIMENTAL DESIGN

Data set:

- a subset of KEGG PATHWAYS DB
  - 10 pathways
  - 13 organisms
- 125 pathway instances (some organisms do not have all pathways)

Cross-validation for each pathway:

- withhold one organism (one pathway instance)
- predict the withheld instance using the others as training data
- aggregate scores from all folds to give results

## SPECIALIZING PREDICTORS

F-measure peaks at different thresholds

Example: performance of BLAST

- two reactions
- same parameters
- different peaks

e-value	F-measure EC 1.8.1.4	F-measure EC 2.3.1.12
1e-3	0.250	0.750
1e-10	0.333	<b>0.857</b>
1e-20	0.400	<b>0.857</b>
1e-50	0.667	0.800
1e-100	<b>1.000</b>	0.500
1e-160	<b>1.000</b>	0.000

Therefore, specialize parameters for each reaction

- Opt BLAST
- Opt HMM

## RESULTS

Catalyst prediction scores

Classifier	F-measure	Precision	Recall
Opt BLAST	0.803	0.715	0.915
Opt HMM	0.767	0.706	0.839
BLAST-HMM	0.667	0.657	0.677
HMM	0.650	0.627	0.674
BLAST Thresh	0.639	0.527	0.810
Motif SVM	0.635	0.586	0.683
BLAST NN	0.468	0.453	0.484

Structure prediction scores

Classifier	F-measure	Precision	Recall
Opt BLAST	0.889	0.857	0.924
Opt HMM	0.864	0.847	0.881
Motif SVM	0.862	0.837	0.889
BLAST-HMM	0.860	0.850	0.871
BLAST Thresh	0.860	0.840	0.880
HMM	0.831	0.880	0.787
BLAST NN	0.665	0.506	0.970

## PREDICTOR

Predicts whether a protein has the same function as the model catalysts

- Trained from catalysts of model pathway

Various techniques tested:

- BLAST
- Profile HMM
- Combined BLAST and HMM
- SVM using Pfam motifs as features

Specializing parameters for each reaction yields better results

- Opt BLAST
- Opt HMM

## CONCLUSION

- Pathway Analyst accurately predicts metabolic pathways
- specialized predictors perform better
- empirically tested
- soon to be available via Internet