FungalWeb

A Semantic Web for Exploring
Knowledge-Based Bioinformatics

Greg Butler
Volker Haarslev, Chris Baker, Sabine Bergler
Leila Kosseim, Doina Precup, Justin Powlowski
Nematollah Shiri, Adrian Tsang

Centre for Structural and Functional Genomics
Dept of Computer Science & Software Engineering
Concordia University, Montreal, Canada

http://www.cs.concordia.ca/FungalWeb

Kingdom: Eumycota
  Phyla:
    Chytridiomycota
    Glomeromycota
    Zygomycota
    Dikaryomycota
        Ascomycotina
        Basidiomycotina

- Introduction to Knowledge-Based Bioinformatics

- Introduction to FungalWeb

  - Fungi, Enzymes, Industry

  - FungalWeb Ontology

  - Application Scenarios

- Conclusion

Knowledge-Based Bioinformatics Aim:
  Provide an automated Research Assistant to a bio-scientist
  [ie Make a human Research Assistant's life more interesting]

Find the data that … answers a question…
Compute a … phylogenetic tree … of …

Find all papers relevant to …

What is the answer to …?
How confident are you in the answer?
On what evidence is the answer based?
How did you arrive at the answer?

What hypothesis best matches the evidence?
What experiment should I perform to answer this question?

We all know how to *create knowledge* …

Information *retrieval*, data collection, ….

Information *extraction*, data access and analysis, …

Organize …
    integrate data and knowledge from multiple sources
    classify examples into categories
    note *relationships* between examples and categories
    note patterns, rules, constraints, …

Observe… correlations, trends, exceptions, …

But how to (semi-)automate?

Transparent Access to Knowledge	Tip of IceBerg
Workflow Coordination	Hidden
Knowledge Representation Reasoning	Hidden
Storage, Access, Analysis	Hidden
Scientific Literature — Data Collections — Algorithms	Hidden

Typical Workbench for Knowledge-Based Bioinformatics

The vision is to turn data into knowledge
  … how best can the computer assist human knowledge workers

Hypothesis: *Use ontologies and the semantic web*

Web provides access, autonomy, diversity, …

Ontologies *organize* knowledge: instances, concepts, relations, rules
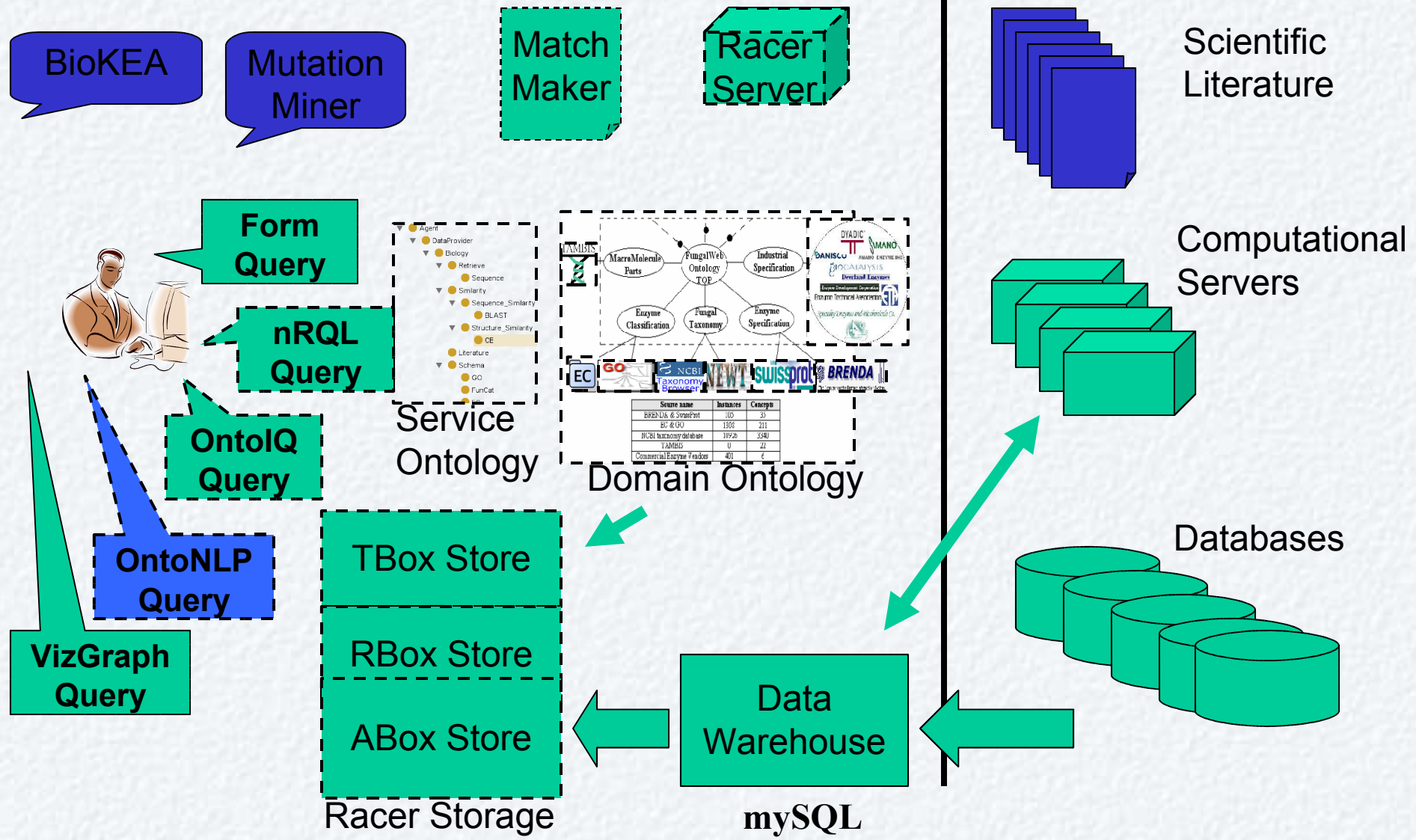Ontologies *integrate* knowledge … bridge sites across web

Software agents *carry out* plans, tasks, workflow, …reasoning,…

But … is this enough, is it *buildable*, is it *usable* by bio-scientists

FungalWeb

The External World

BioKEA

Mutation Miner

Match Maker

Racer Server

Scientific Literature

Form Query

Service Ontology

Domain Ontology

Computational Servers

nRQL Query

OntoIQ Query

OntoNLP Query

VizGraph Query

TBox Store

RBox Store
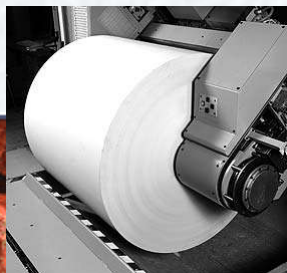
ABox Store

Racer Storage

Data Warehouse

mySQL

Databases

Concordia UNIVERSITY

The Kingdom of Fungi includes over 1.5 million species

Pulp and paper

Brewing

pharmaceutical

Personal care
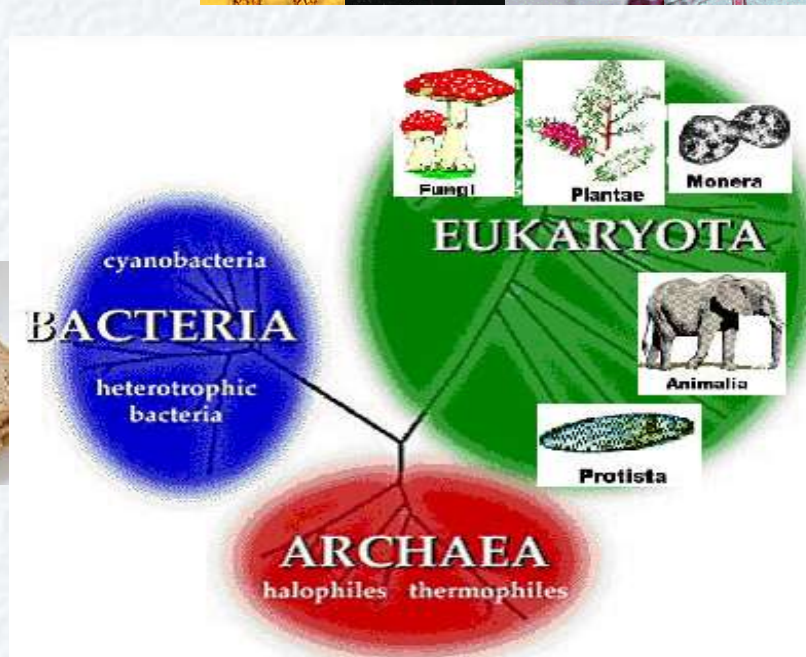
Five kingdoms of life

Concordia
UNIVERSITY
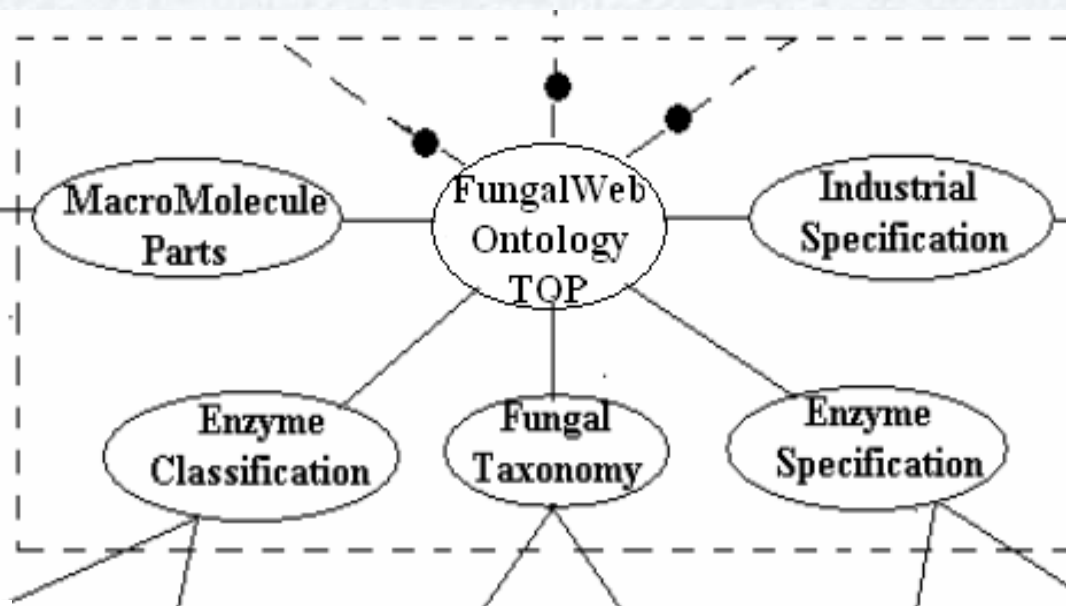
Source name	Instances	Concepts
BRENDA & SwissProt	105	35
EC & GO	1308	211
NCBI taxonomy database	10926	3340
TAMBIS	0	22
Commercial Enzyme Vendors	401	6

ISWC05 2nd prize (Semantic Web Challenge)

Scenario 1: Enzymes acting on substrates

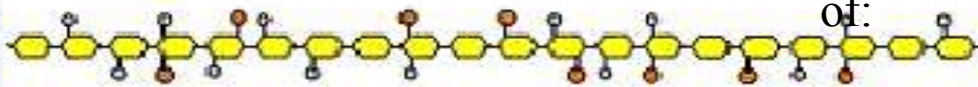Scenario 2: Enzyme taxonomic provenance

Scenario 3: Enzyme benchmark testing

Scenario 4: Enzyme improvement

Could an enzyme be used to degrade this novel chemical substrate ?

Chemical Analysis describes it as a a polymer of:

| **IUBMB Enzyme Nomenclature**	EC 3.2.1.67
Common name:	galacturan 1,4-a-**galacturon**idase
Reaction:	$(1,4$-a-D-**galacturon**ide$)n$ + H2O = $(1,4$-a-D-**galacturon**ide$)n$-1 + D-**galacturon**ate
Other name(s):	exopoly**galacturon**ase; poly(**galacturon**ate) hydrolase; exo-D-**galacturon**ase; exo-D-**galacturon**anase; exopoly-D-**galacturon**ase
Systematic name:	poly(1,4-a-D-**galacturon**ide) **galacturon**ohydrolase

NLP Semantic word stem summary: '**GALACTURON**'
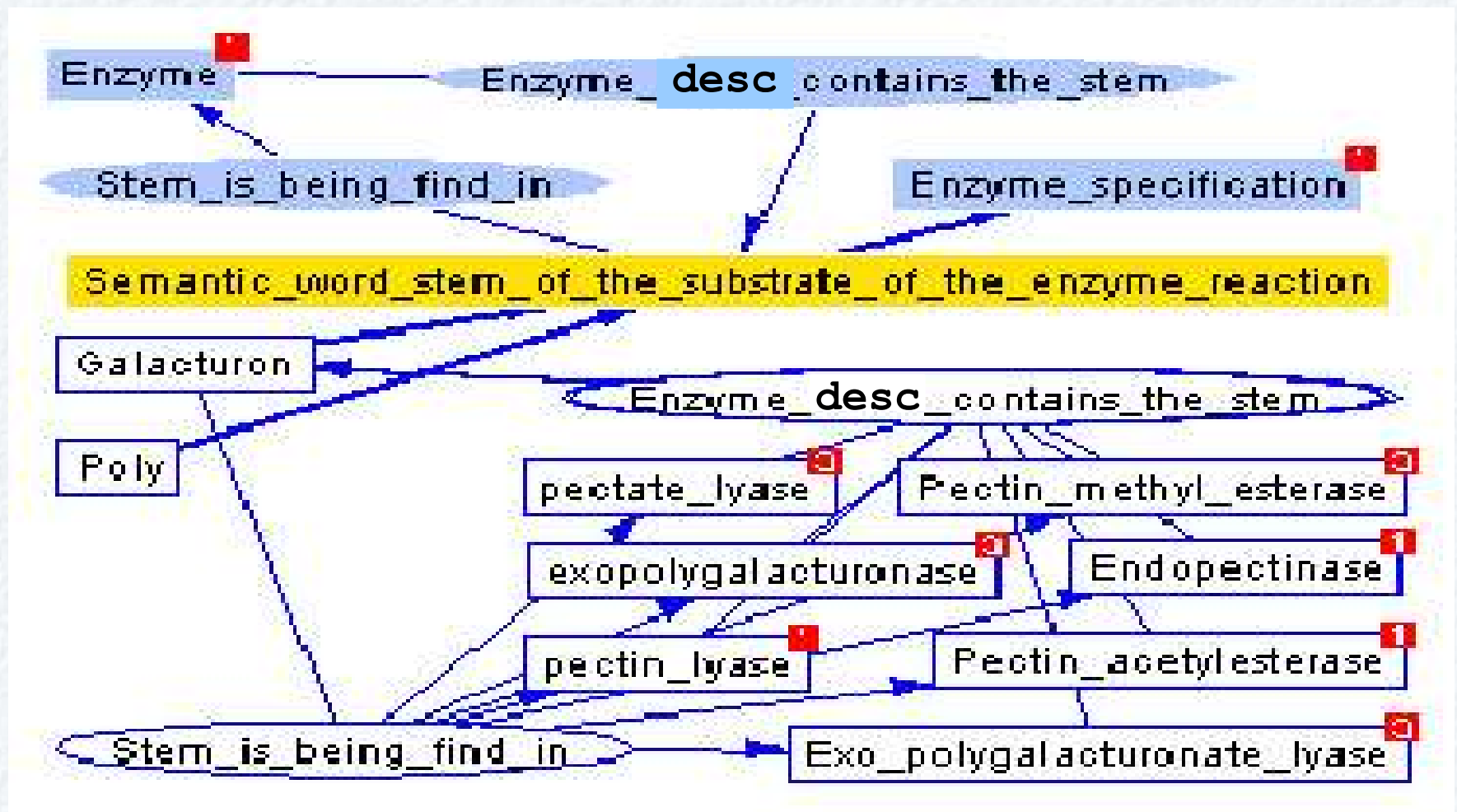
Conceptual frame supporting the identification of pectinase enzymes using substrate word stems.

1-Is Galacturon an instance for Semantic_word_stem_of_the_substrate_of_the_enzyme_reaction?
Retrieve ( ) (||http://a.com/ontology#Galacturon|
|http://a.com/ontology#Semantic_word_stem_of_the_substrate_of_the_enzyme_reaction|))

True

2-Find all Enzyme names which contain semantic word stem of the substrate of the enzyme reaction that matches with Galacturon

Retrieve (?x) (AND (?x |http://a.com/ontology#Enzyme|)(?x |http://a.com/ontology#Galacturon| |
http://a.com/ontology#Enzyme_description_contains_the_stem|)))

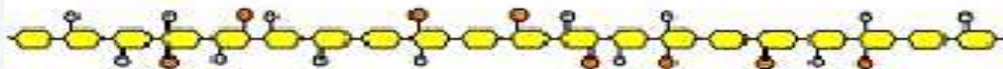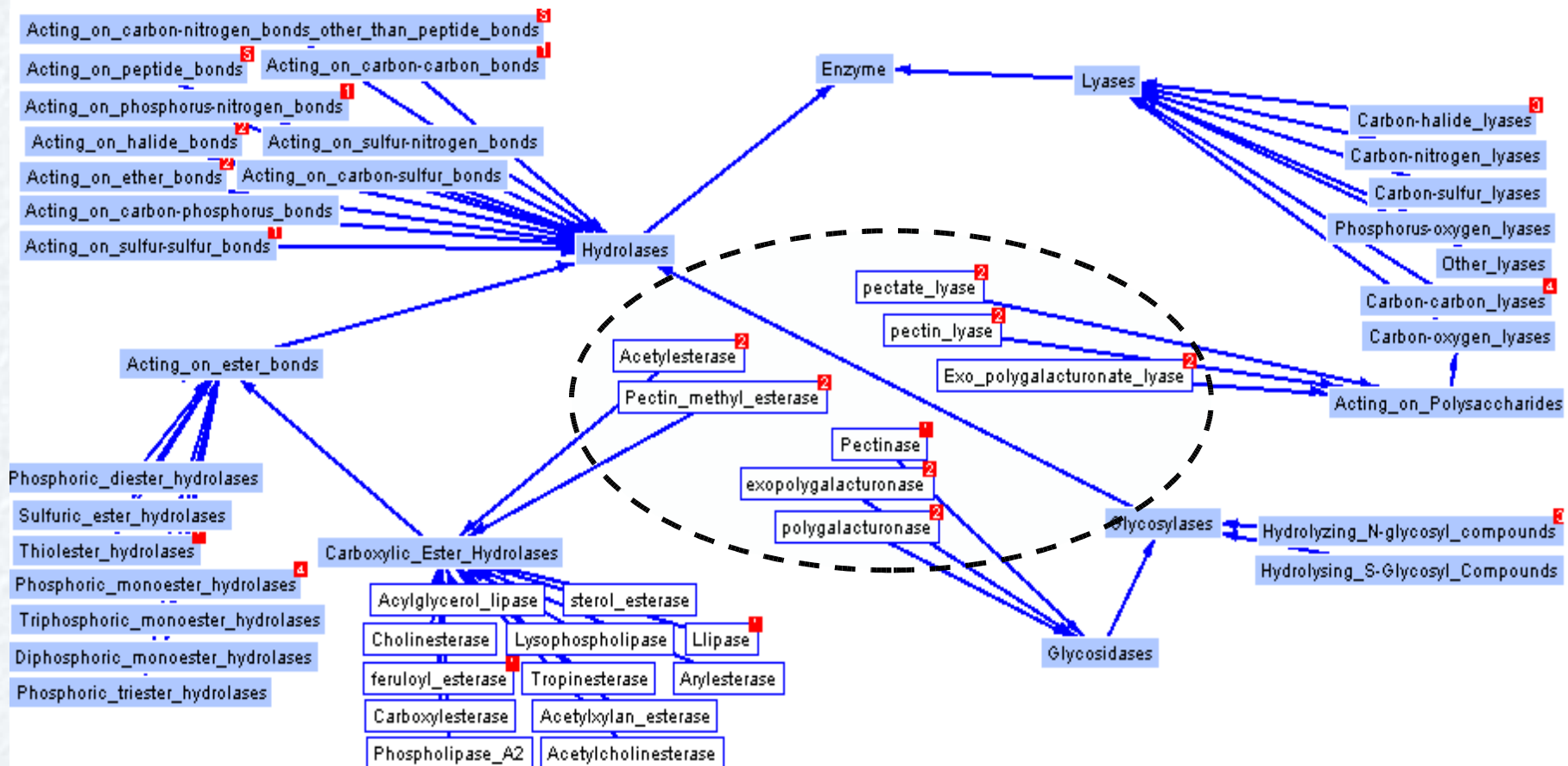 <<<?X :http://a.com/ontology#exopolygalacturonase:>>
 <<?X :http://a.com/ontology#pectin_lyase:>>
 <<?X: http://a.com/ontology#Pectin_methyl_esterase:>>
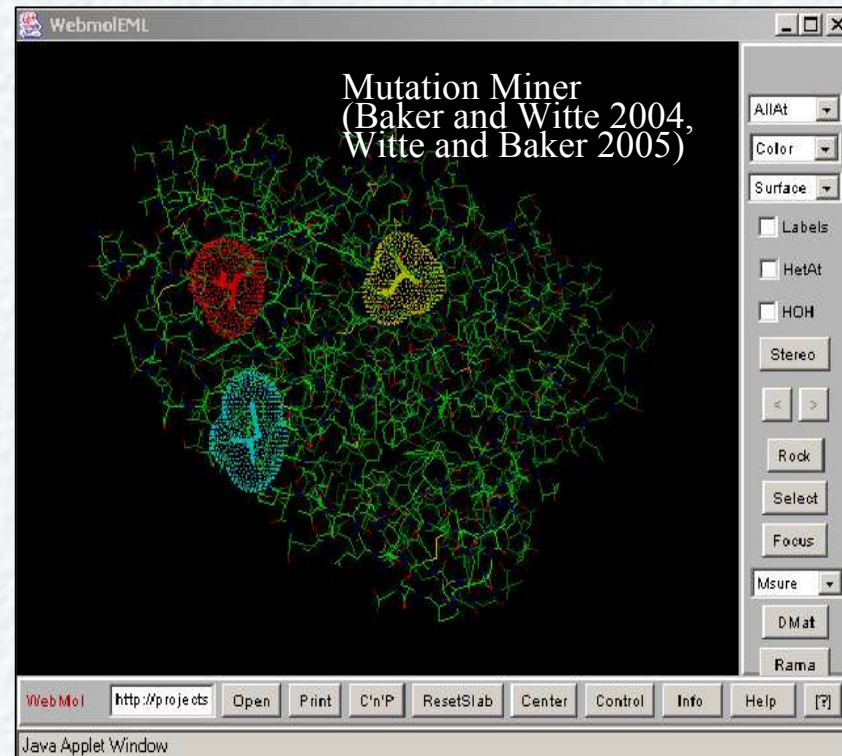 <<?X:http://a.com/ontology#Exo_polygalacturonate_lyase:>>
 <<?X :http://a.com/ontology#Endopectinase:>>
 <<?X :http://a.com/ontology#pectate_lyase:>>
 <<?X :http://a.com/ontology#Pectin_acetylesterase:>>>

Mutation Miner
(Baker and Witte 2004,
Witte and Baker 2005)
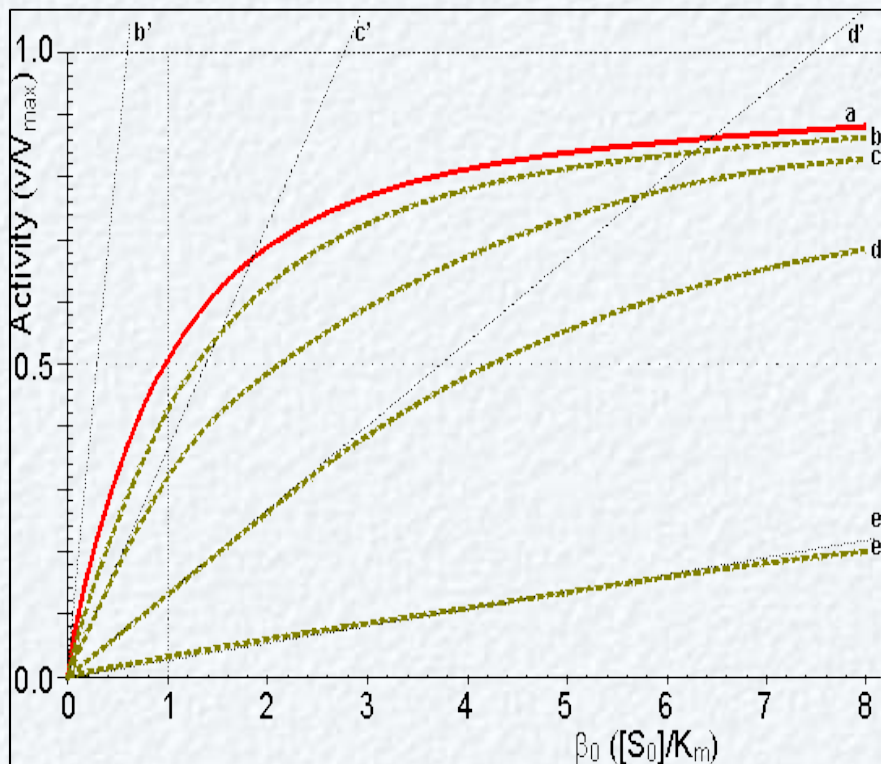
MOL_ID: 1; MOLECULE: EXOCELLOBIOHYDROLASE I; CHAIN: A; FRAGMENT: CATALYTIC DOMAIN 1-434;

☑ Xylanase

The mutations N11D and N38E did not have any significant effect:N11D increased the half life scarcely 1.5 times at 55?C, and N38E about 1.5

Q162H, Q162Y,Q162L and Q162K, increased the half life of XYNII at 55?C (pH 5) to 36,39, 26 and 11 min, respectively. Q162H, Q162Y and Q

The glutamic acid residue at position 210, which is part of the active center in this family of enzymes, was changed to either aspartic acid (E210

RESET COLORS

Java Applet Window

Designed to:

- Extract from full-text papers,

- …sentences that describe impacts of mutations, and

- …legitimately map them to protein structures
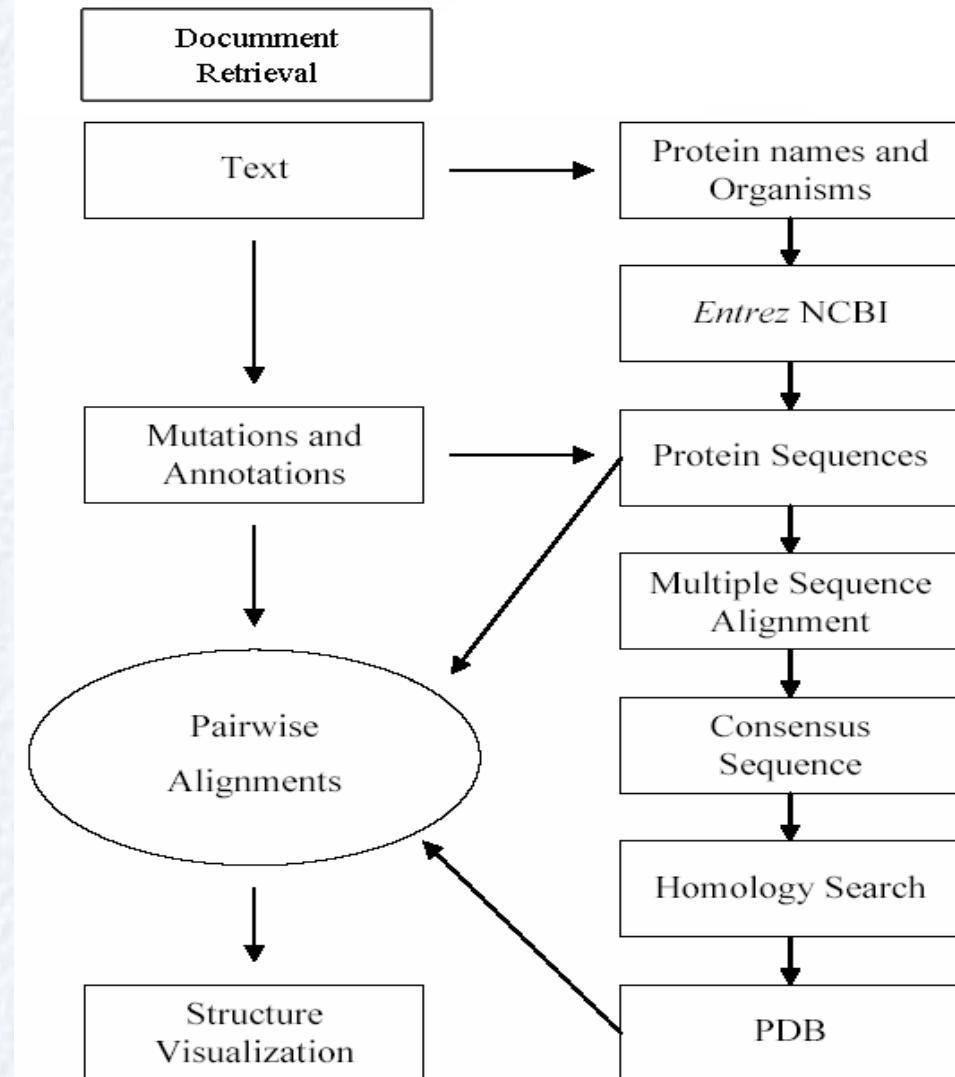
Data and knowledge integration works:

- •Fungal Web Ontology can support *real* biological questions not easily queryable from bioinformatics databases

- • Ontologies are difficult to build, evaluate, …

- • RACER nRQL syntax is expressive enough,

    but is unreadable to scientists

Powerful approach to integrate

   ontologies, NLP, computation, and visualization

    eg Mutation Miner

Better user interfaces to access data

- OntoIQ form-based pattern-based interface for nRQL
- OntoNLP natural language interface for nRQL
- Visual graph-based queries

FungalWeb data warehouse

- A web of data for experimentation with DB, agents, and FungalWeb Ontology
- A benchmark for genomics databases

Ongoing validation of

- PRM tools and application scenarios
- NLP tools: Mutation Miner, BioKea, BioRAT

Technology will always need organization to create knowledge!

- **IT** being web services, semantic web, data, …
- Ontologies offer a way to organize
- Ontologies *evolve* through community use, review, …
- This takes *people*: expert knowledge workers

Remember human interaction steps

- Data entry, Manual curation
- Review, feedback, corrections, evolution,…of data and knowledge

Remember *science evolves* through theories, evidence, refutation

- What assumptions/theories are your computations based upon?
- How do differing assumptions affect results?
- Does your system accommodate competing conflicting theories?
- Can you undo/refute all results based on a discredited theory/assumption?

Acknowledgements

Volker Haarslev, Chris Baker, and all the FungalWeb team: see http://www.cs.concordia.ca/FungalWeb

Adrian Tsang, Reg Storms, Justin Powlwoski, and the bioinformatics team on the fungal genomics project

My graduate students: Farzad Kohantorabi, Ju Wang, Yue Wang, Michel Nathan