

INTRODUCTION

One of the major problems that biologists face is the integration of multiple knowledge sources available from internet. The Mouse Atlas project at the CHUL research center requires an integrated knowledge resource of pathway information to facilitate the high throughput annotation to microarray data. The Pathway2GO project was initiated to meet this need and we report the mapping of pathway information from KEGG and Reactome to GO.

METHODOLOGY

Our approach in the Pathway2GO project involved the following steps:

- Data acquisition from de web sites in different formats (text, xml, RDF, SQL dump);
- Data conversion to RDF using PERL scripts, XSL style sheet and Sesame SeRQL queries;
- Merging data using **SESAME** based RDF store;
- Selection of the desired content with SeRQL queries;
- Pathway visualisation with the **Protégé** ontology editor employing the **TGVizTab** plugin.

Kegg and **Reactome** are the two most advanced pathway resources. While Reactome exists in RDF format, Kegg is not RDF enabled. For this reason, we developed a Kegg2RDF converter and to further integrate both of these two pathway resources in a common format: we created a customed lite version of the BioPax pathway ontology : **BioPaxLite**.

Having integrated the pathways in RDF we mapped this information to MGI annotations that link mouse genes and gene ontology terms. Our final product was an integrated pathway knowledge representation integrated with GO annotation for each mouse gene involved in a metabolic reaction.

CONCLUSION

We have shown that a semantic web approach was usefull to integrate knowledges from differents sources once converted to RDF. Protégé and Sesame open source technologies are matured tools ready to be used by the bioinformatic community. We will now integrate **Bind** and **Intact** protein interactions knowledge base to the system and use it for automatic annotation of microarray experiments of the Mouse Atlas project.

AVAILABILITY

The Pathway2GO project is registered at <http://bio2go.sourceforge.net> where perl scripts used to convert KEGG and MGI databases to RDF are located. For further information contact the author at francoisbelleau@yahoo.ca.

ACKNOWLEDGMENTS

We would like to thank the support of Genome Canada and Genome Québec.

REFERENCES

Blake JA, Richardson JE, Buit CJ, Kadin JA, Eppig JT, and the members of the Mouse Genome Database Group. 2003. MGD: The Mouse Genome Database. *Nucleic Acids Res* 31: 193-195.

BioPAX : Biological Pathways Exchange (<http://www.biopax.org/Documents.html>)

Cheung KH, Yip KY, Smith A, Deknikker R, Masiar A, Gerstein M. YeastHub: a semantic web use case for integrating data in the life sciences domain. *Bioinformatics*. 2005 Jun 1;21 Suppl 1:i85-i96

ChEBI (Chemical Entities of Biological Interest, <http://www.ebi.ac.uk/chebi/>)

Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium (2000) *Nature Genet.* 25: 25-29

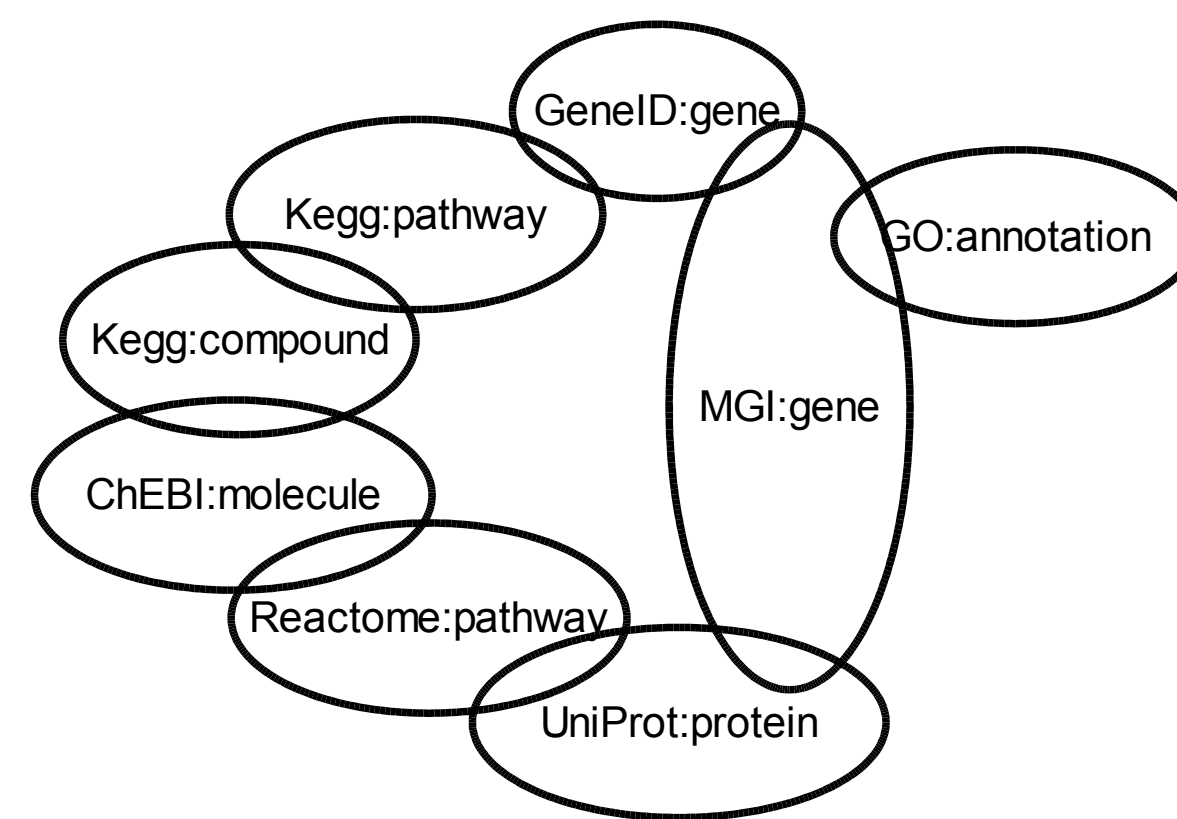
Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L. 2005. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* 2005 Jan 1;33 Database Issue:D428-32. PMID: 15608231

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

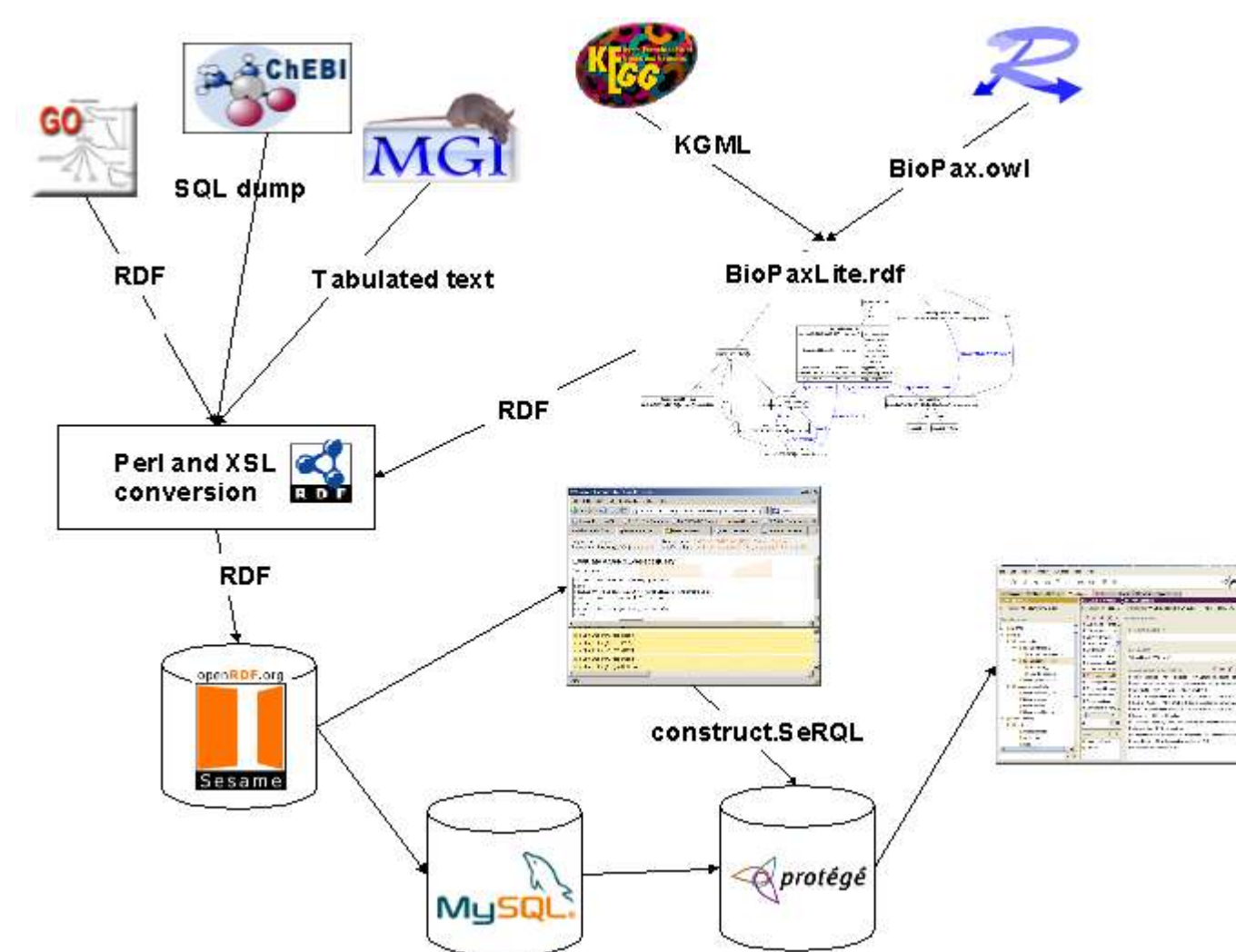
Jeen Broekstra1, Arjoh Kampman1, and Frank van Harmelen, Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema, Published at the International Semantic Web Conference 2002, Sardinia, Italy.

N. F. Noy, M. Sintek, S. Decker, M. Crubezy, R. W. Fergerson, & M. A. Musen. Creating Semantic Web Contents with Protege-2000. *IEEE Intelligent Systems* 16(2):60-71, 2001.

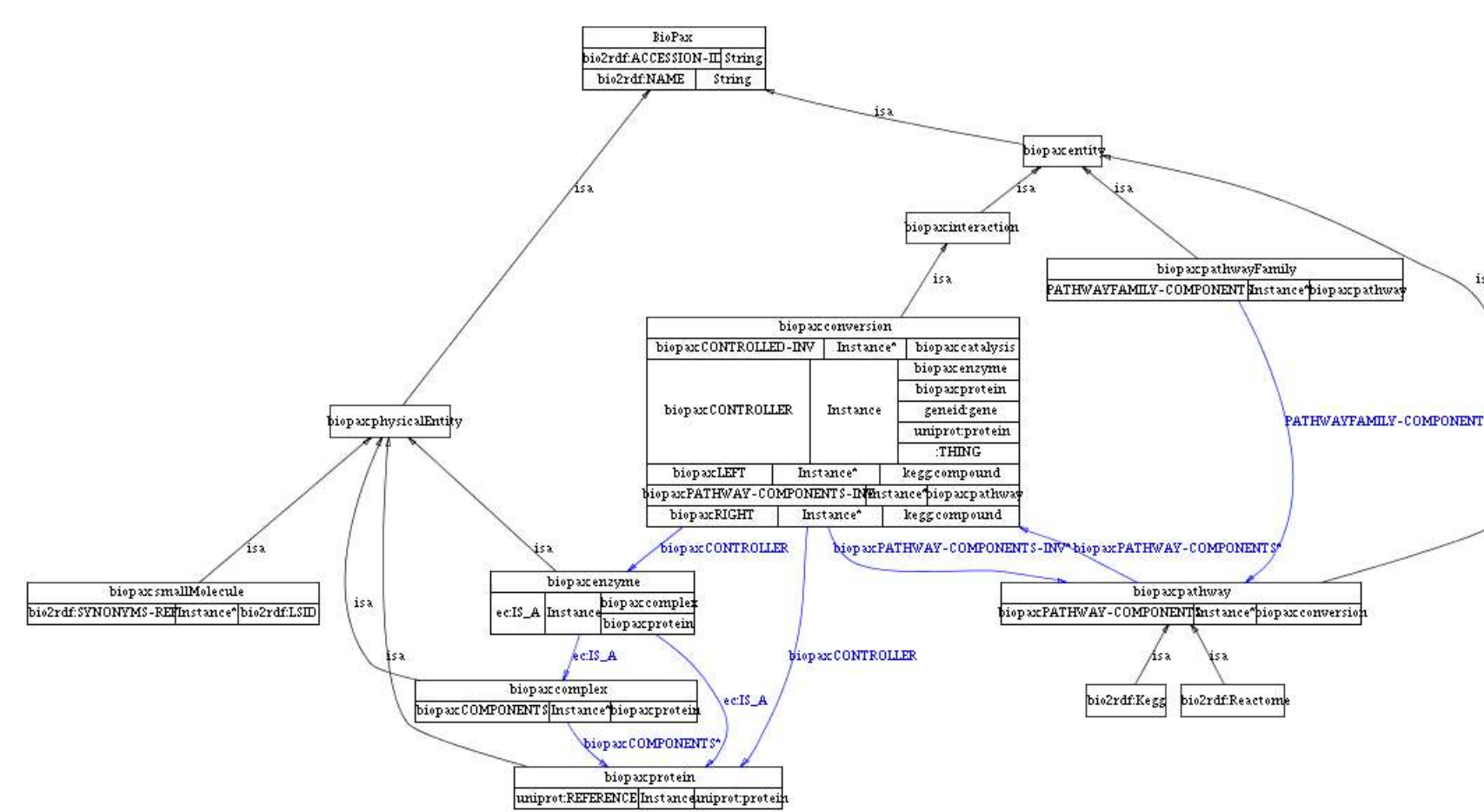
METHODOLOGY



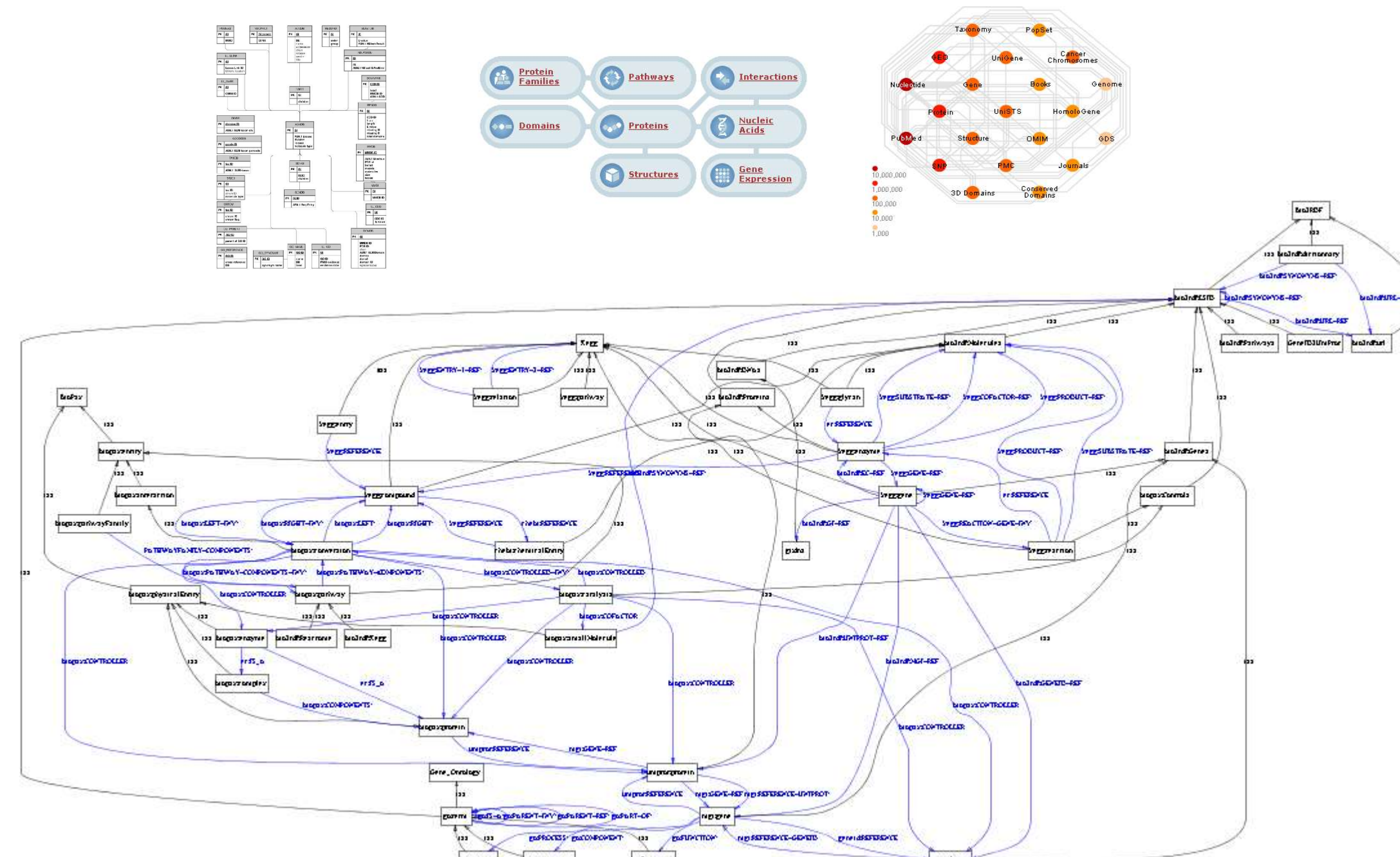
Knowledge bases integration with RDF technology
Kegg + Reactome + GO + MGI + ChEBI = Pathway2GO



Architecture of Pathway2 GO project,
Protégé Ontology editor et Sesame RDF store are used

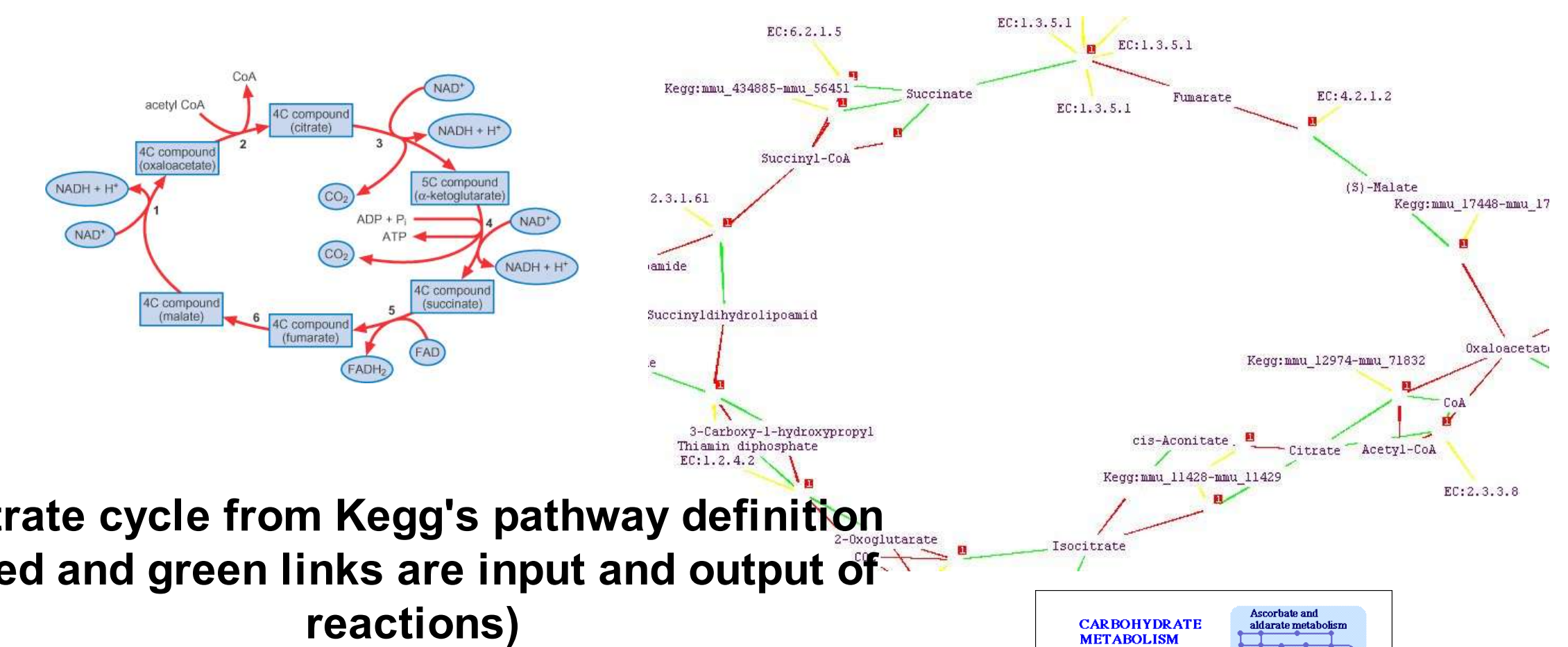


Ontology of BioPaxLite used to merge
Kegg.kgml.xml and Reactome.biopax.owl

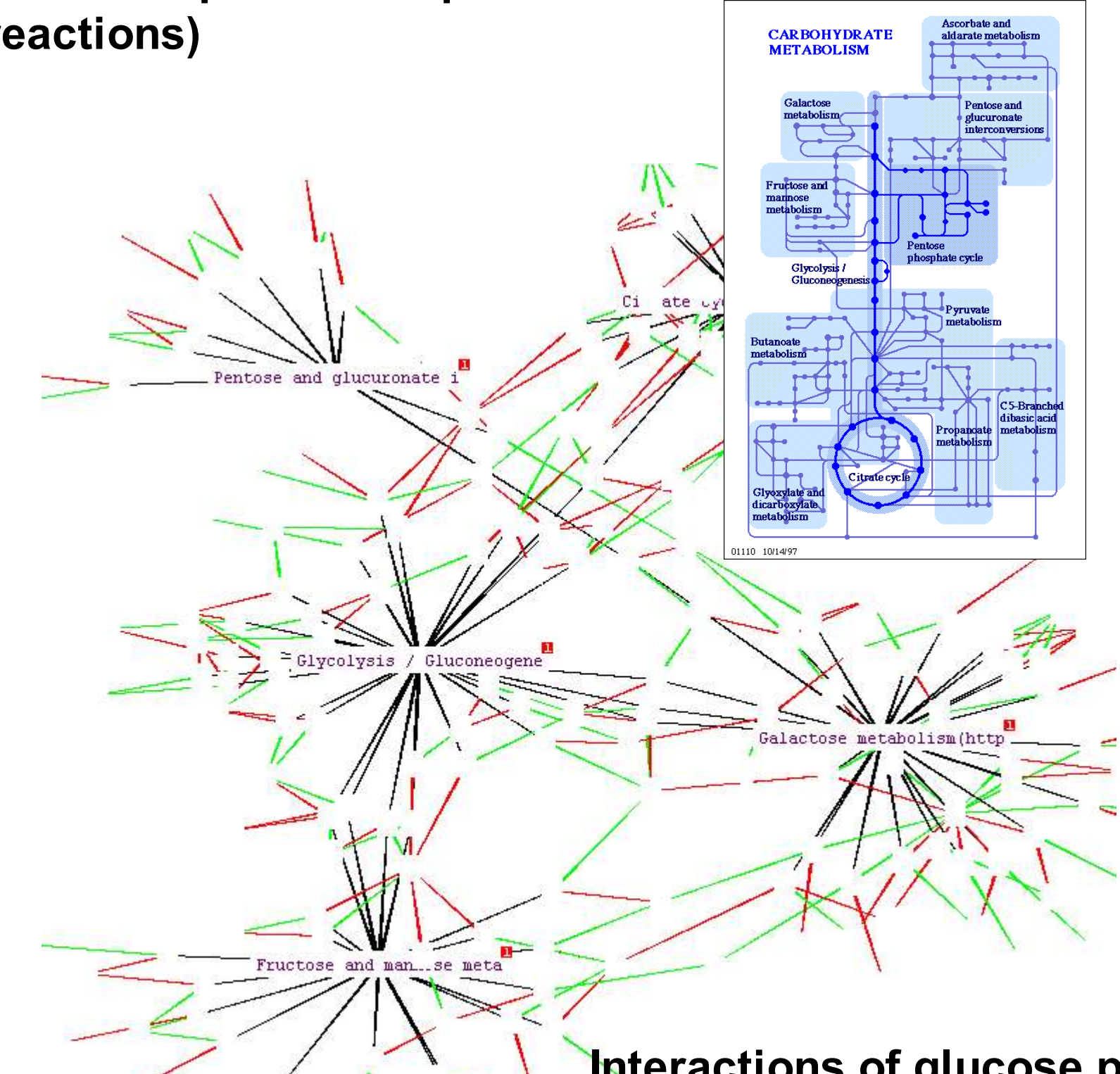


Knowledge model of Pathway2GO
ontology vs SeqHound, Biozon and NCBI

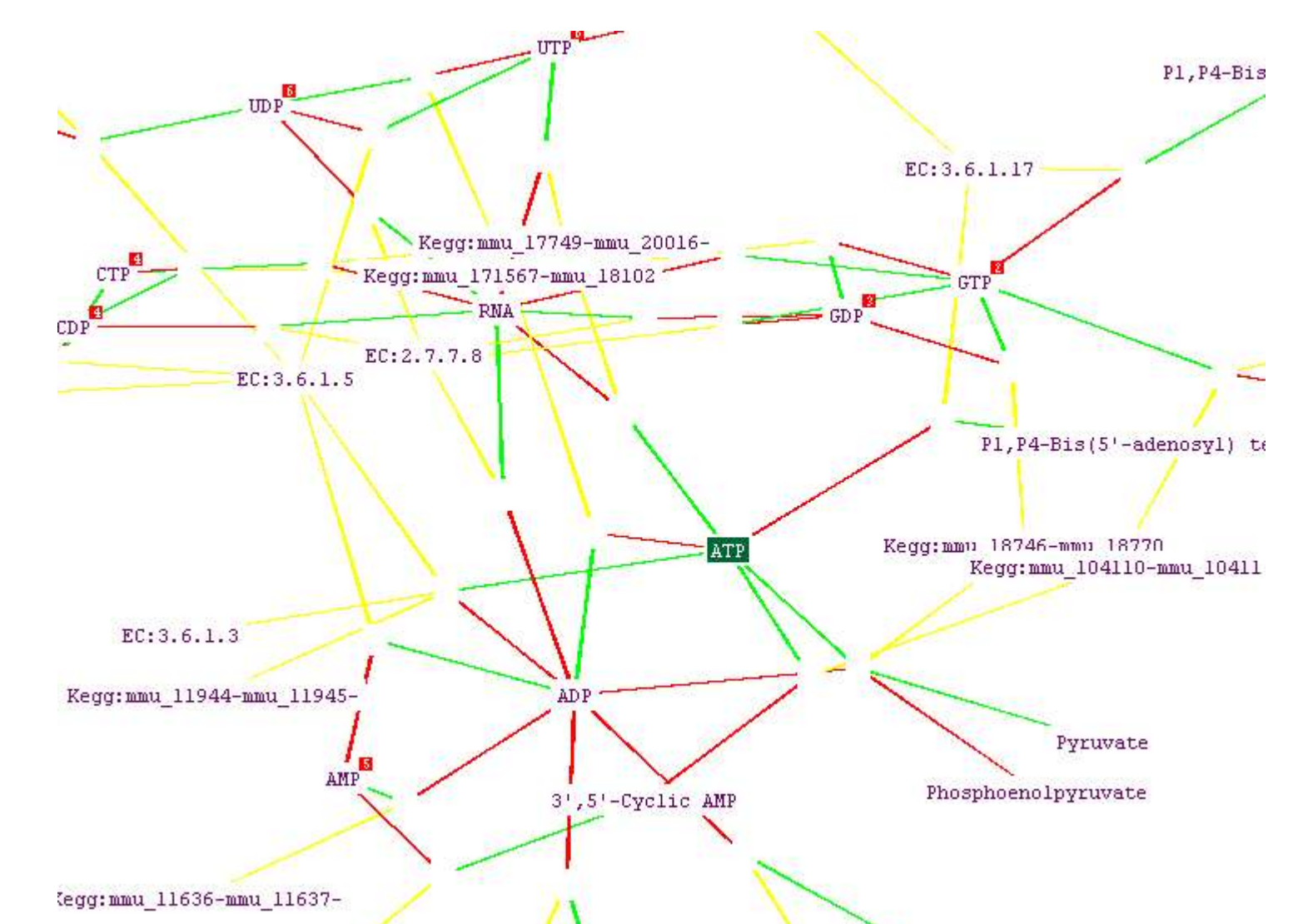
RESULTS



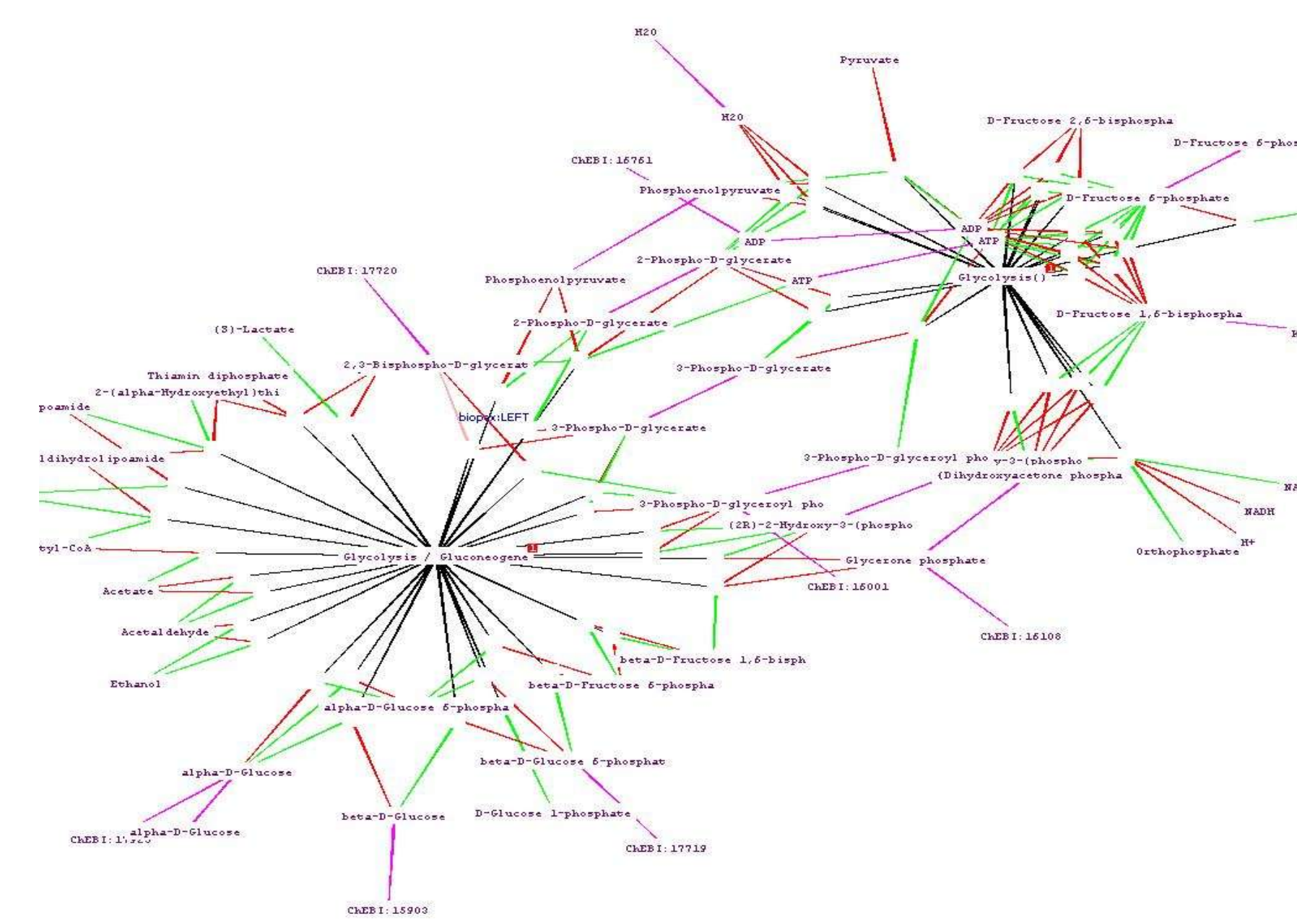
Citrate cycle from Kegg's pathway definition
(red and green links are input and output of reactions)



Interactions of glucose pathways
with the reactions involved



Biochemical reactions involving ATP
with the implied enzymes in yellow



Comparison of Kegg and Reactome
glycolysis pathways with common
molecules in purple