

Pacific Symposium on Biocomputing
Kauai, Hawaii
January 7, 2003

Integrating GO into UMLS



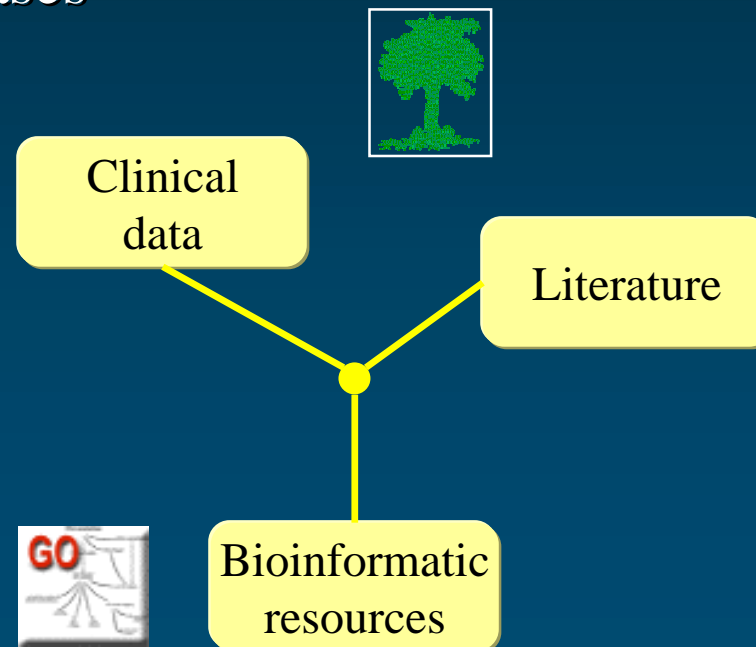
Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA

Introduction

- ◆ Bioinformatic landscape
 - Complex, dynamic data
 - Large amount on information
 - Heterogeneous databases

- ◆ Interoperability



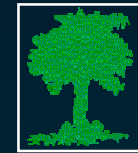
UMLS

- ◆ National Library of Medicine
- ◆ Content
 - 1.5 M unique English strings
 - 775,000 concepts
 - From over 60 biomedical vocabularies
 - Core (e.g., anatomy, drugs)
 - Clinical (e.g., SNOMED, ICD)
 - Literature (e.g., MeSH)
 - 11 M relationships

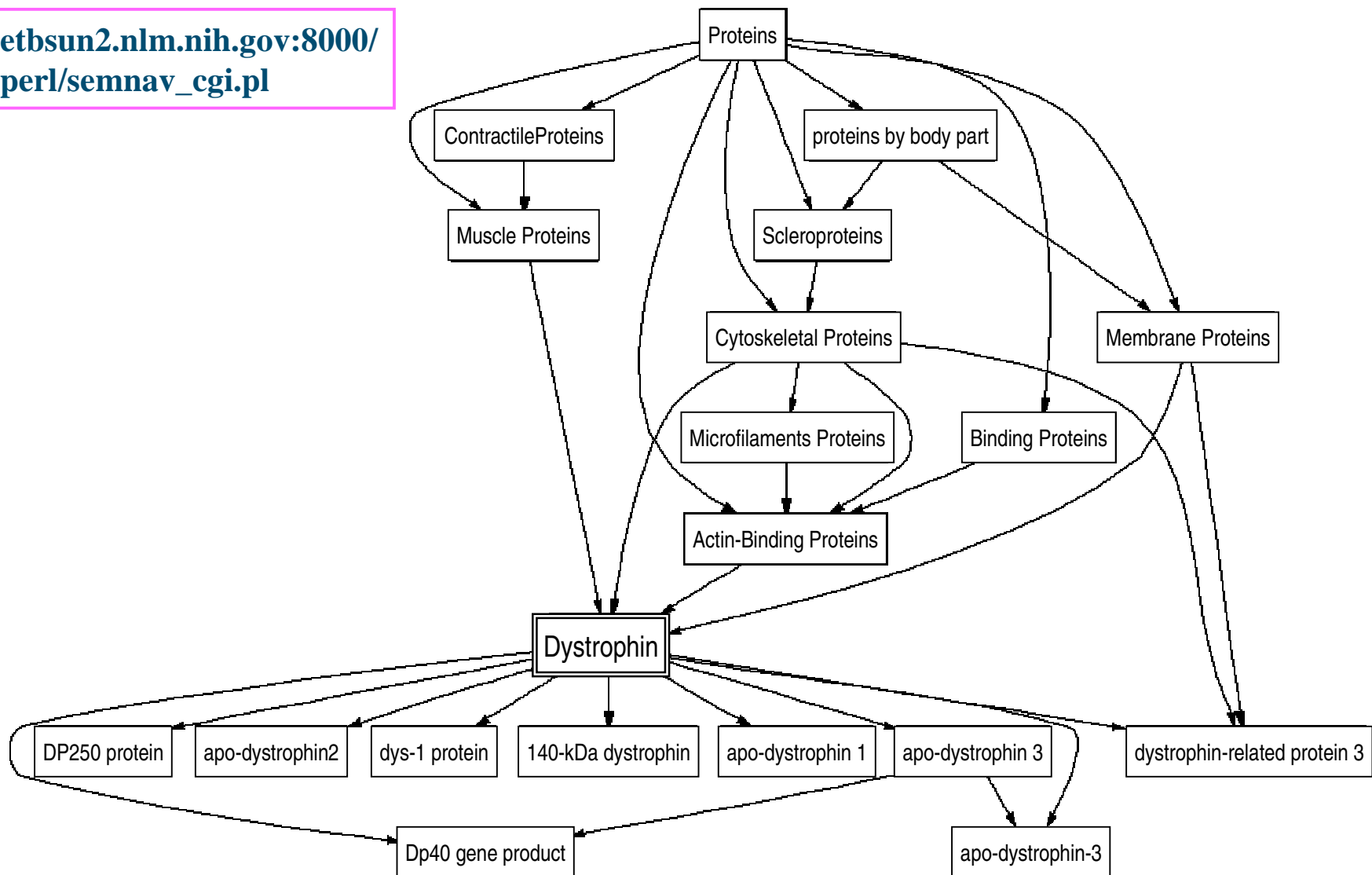


umlsinfo.nlm.nih.gov

Dystrophin in UMLS



[etbsun2.nlm.nih.gov:8000/
perl/semnav.cgi.pl](http://etbsun2.nlm.nih.gov:8000/perl/semnav.cgi.pl)



Gene Ontology

- ◆ Gene Ontology Consortium
- ◆ Controlled vocabulary used to annotate gene products
- ◆ 3 separate hierarchies
 - Molecular function (~5600 names)
 - Cellular component (~4700 names)
 - Biological process (~1100 names)

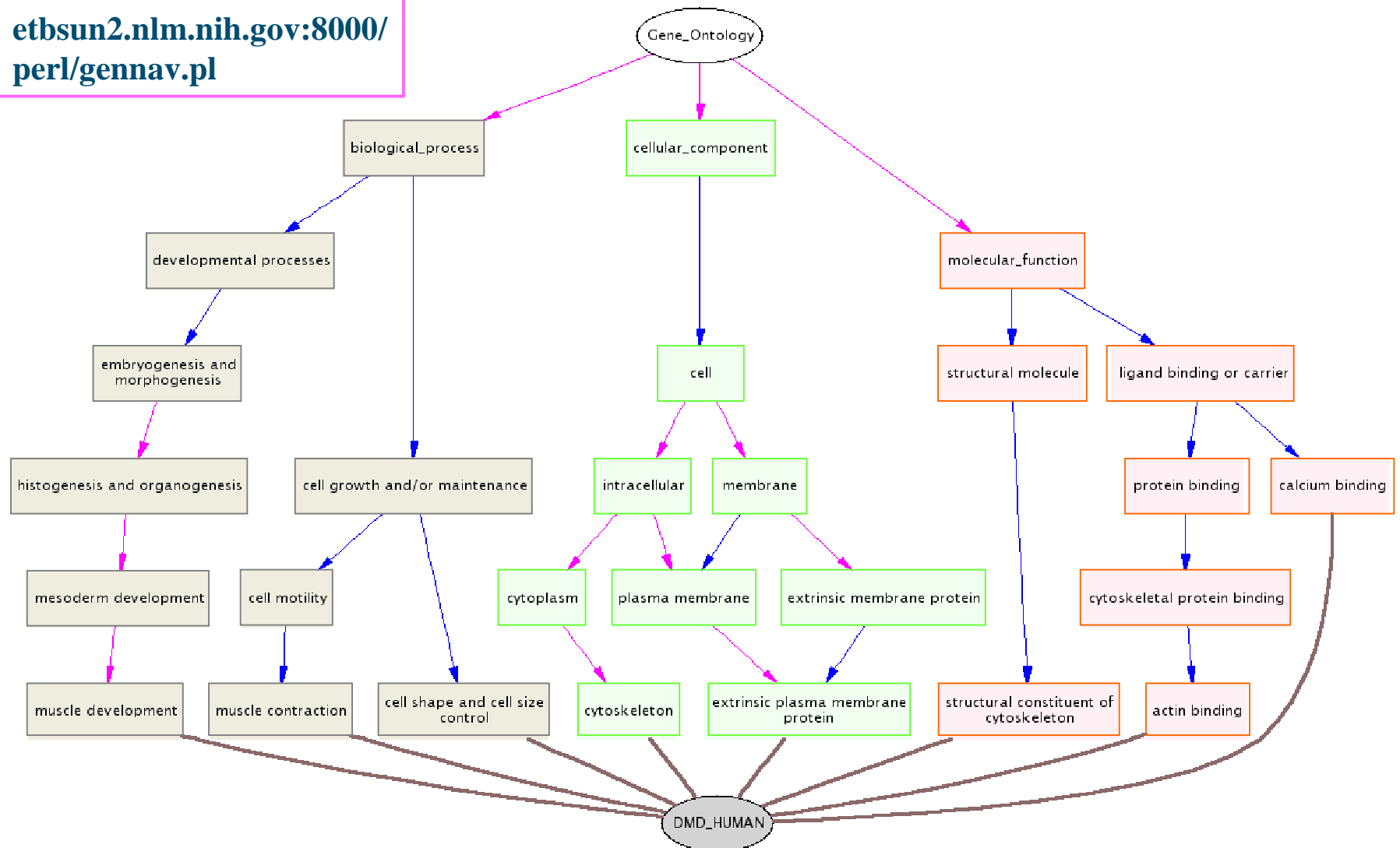


geneontology.org

Dystrophin in Gene Ontology

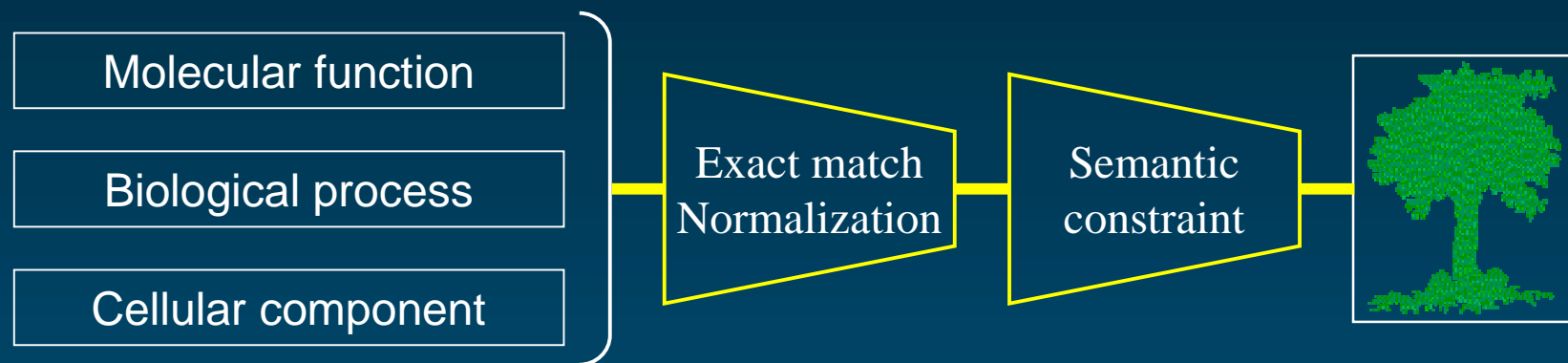


[etbsun2.nlm.nih.gov:8000/
perl/gennav.pl](http://etbsun2.nlm.nih.gov:8000/perl/gennav.pl)



Mapping GO terms to UMLS concepts

- ◆ Automatic mapping
- ◆ Quantitative evaluation

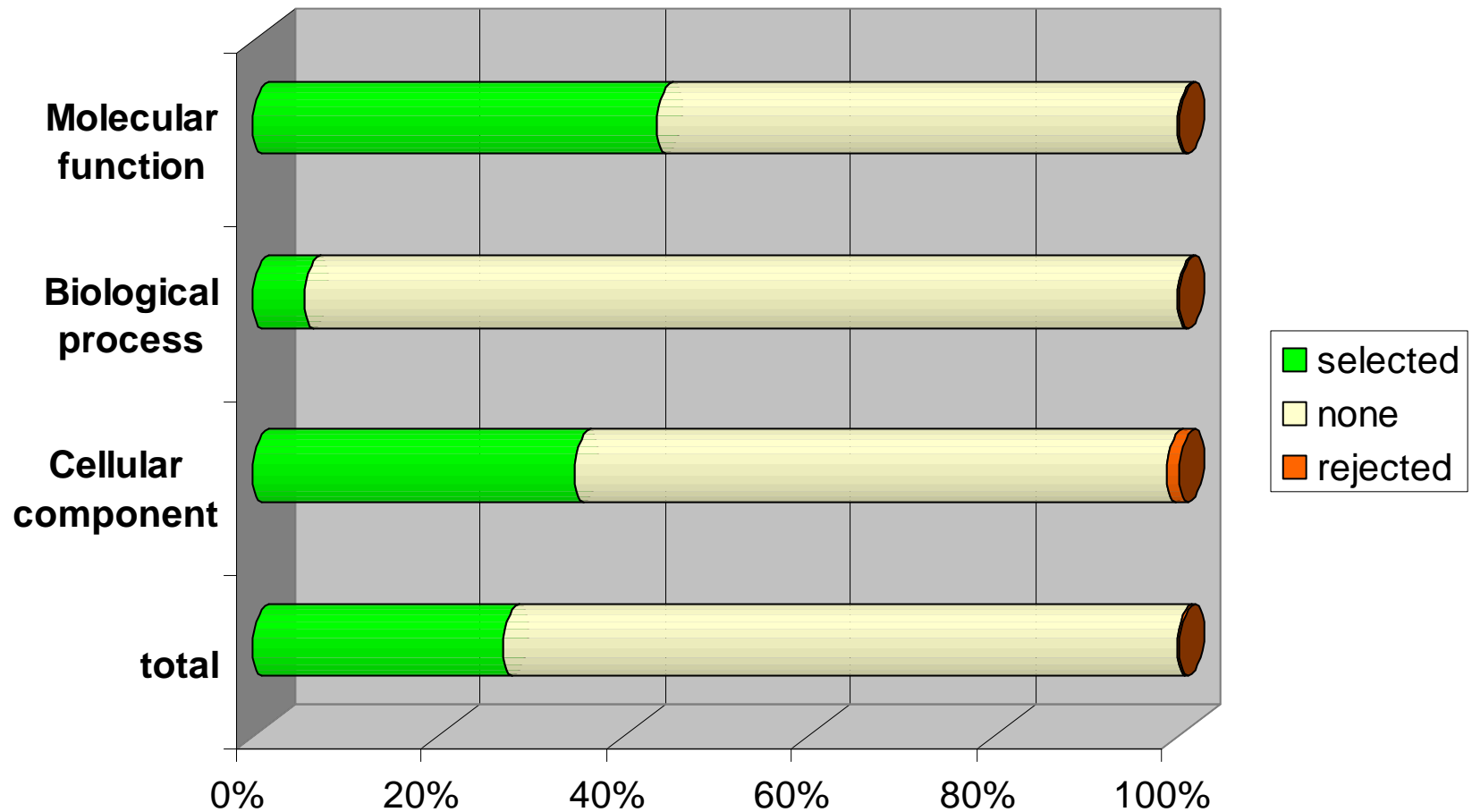


Bodenreider O., Mitchell J.A., McCray A.T.
*Evaluation of the UMLS as a terminology and knowledge resource
for biomedical informatics.*

Proc AMIA Fall Symp. 2002:61-65



Results



3,062 mappings selected

Curated mapping

- ◆ Started last fall
- ◆ In collaboration with GO editors
(Jane Lomax)
- ◆ Manual validation of the mappings

Lomax J.

The Gene Ontology and its insertion into UMLS.

Proc SOFG 2002

www.ebi.ac.uk/microarray/General/Events/SOFG/SOFG.html



Preliminary results

- ◆ Consistent with the algorithmic mapping
- ◆ Overlap with existing sources
 - 11% with SNOMED
 - 20% with MeSH
 - 23% overall
- ◆ Some issues
 - Synonymy in GO
 - Structure vs. Function (e.g., enzymes)

Conclusions

- ◆ Having GO in UMLS should enable interoperability among clinical, literature, and bioinformatic resources
- ◆ 73% of GO concepts are specific to GO (at least finer-grained than what is currently available in UMLS)
- ◆ Availability of GO in UMLS: mid-2003

Contact: olivier@nlm.nih.gov



Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA