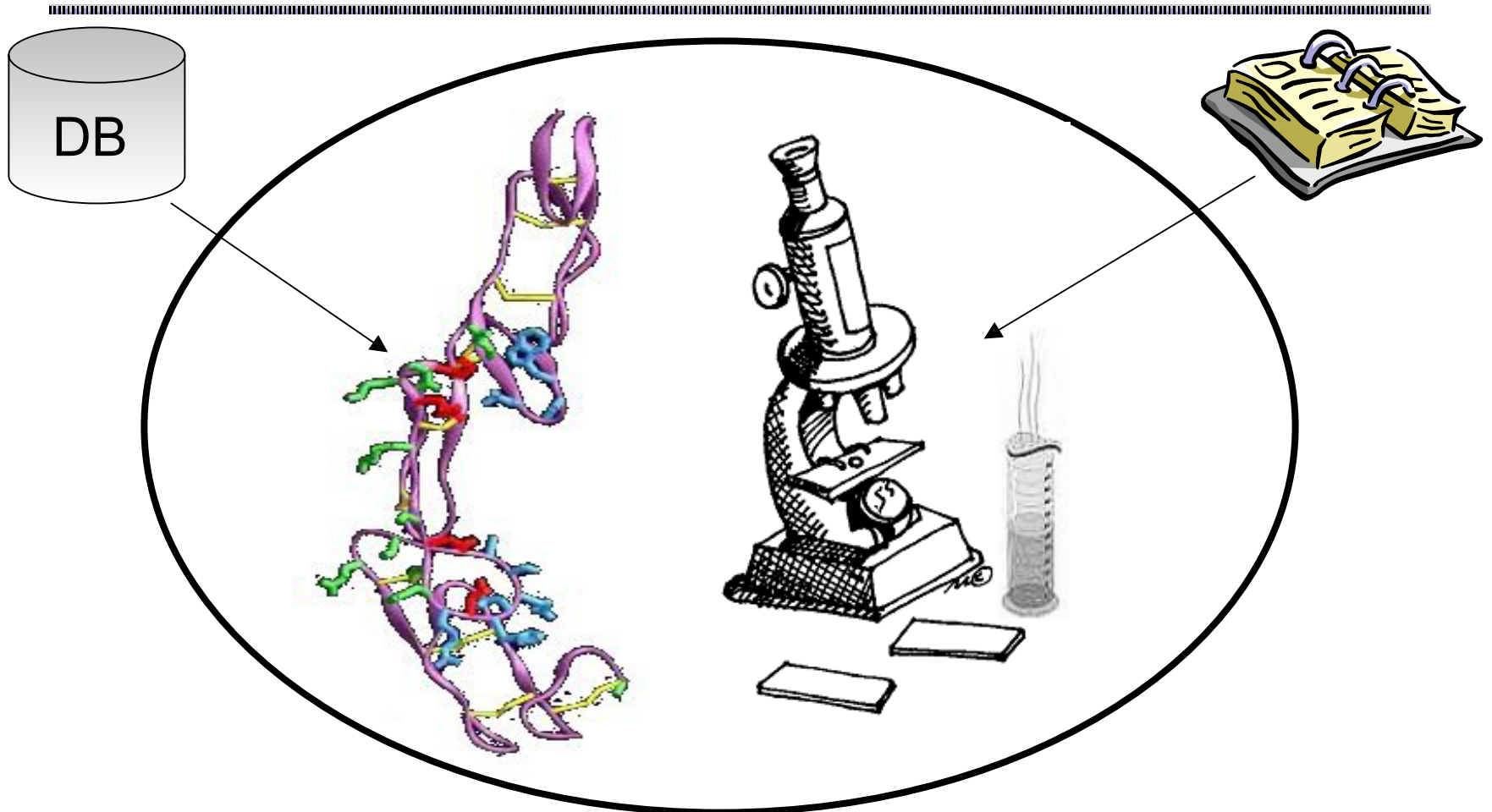


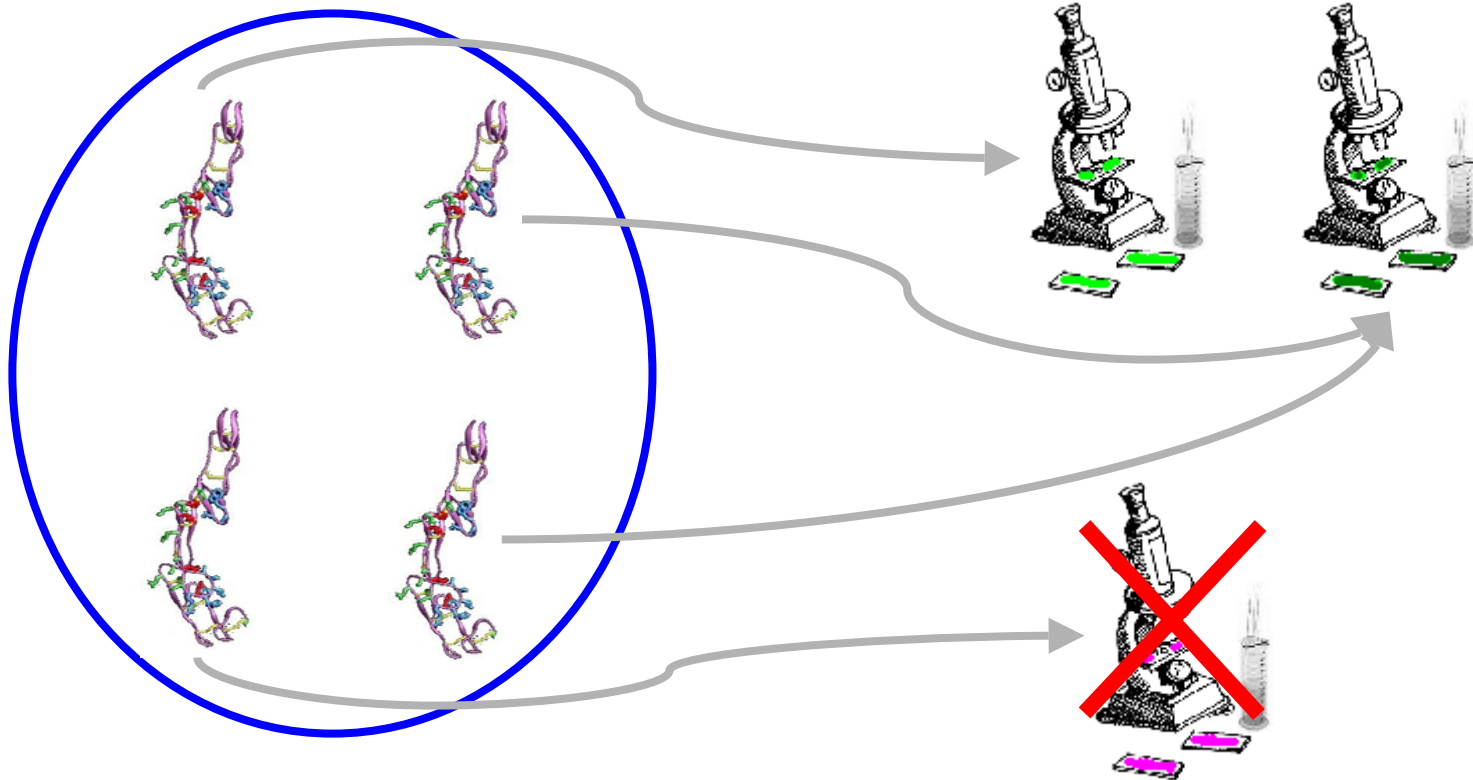
Curating Extracted Information through the Correlation between Structure and Function

Francisco Moreira Couto

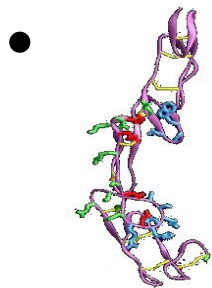
Annotation



Correlation

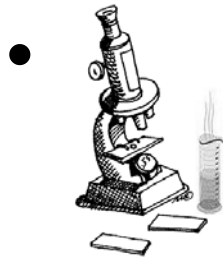


Distances



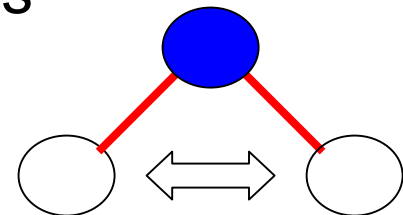
Structural

- MRRELSVRG... ~ MRRELSVRG...
- BLAST

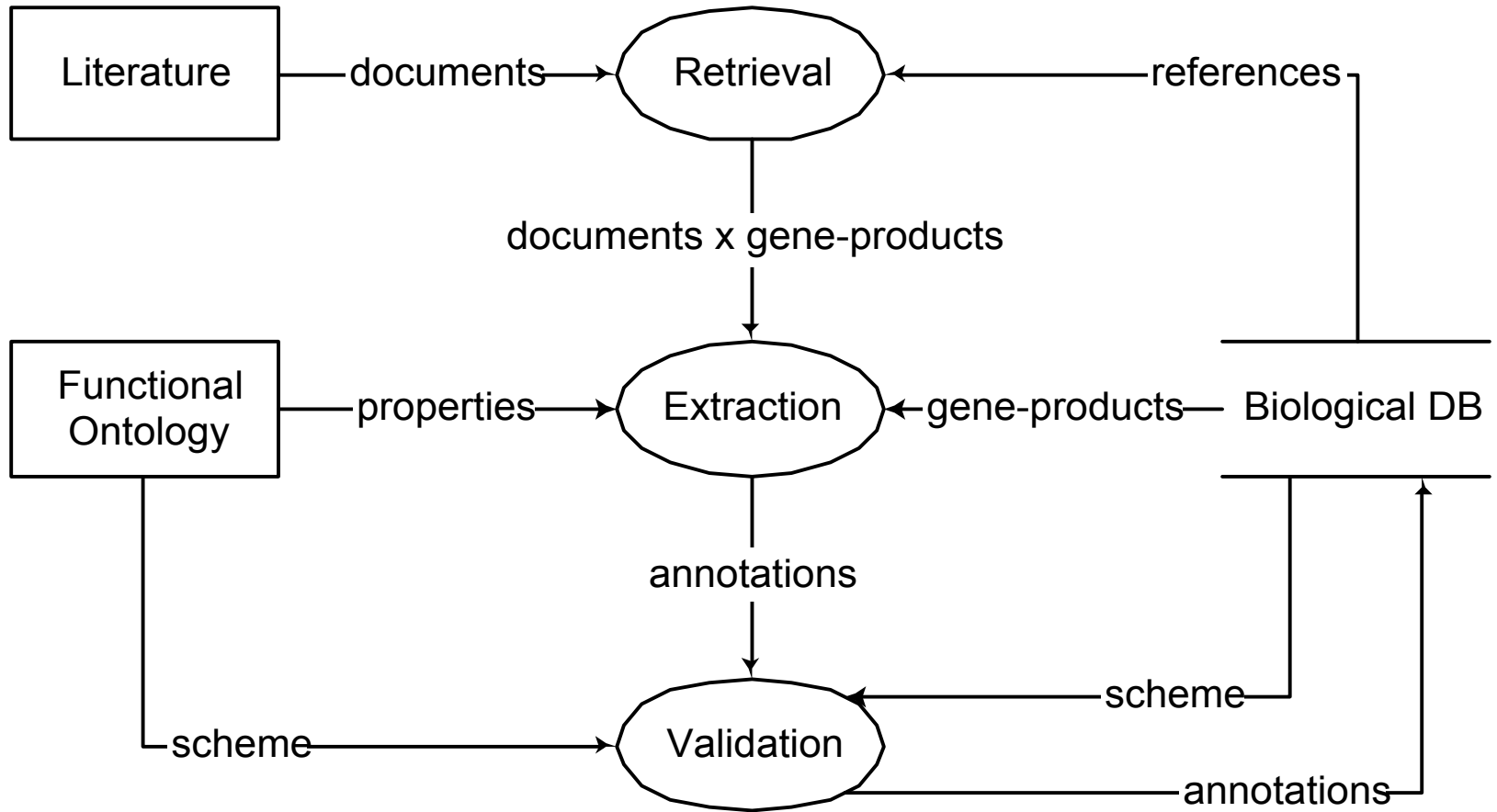


Functional

- Semantic Similarity Measures
 - » Information content
 - » Conceptual Distance



IE System



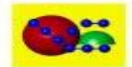
CAZy (Carbohydrate Active enZYmes)



<< phpCAZy

>>

CAZyModO



CAZyModO - Glycoside Hydrolase Classification

Family GH48

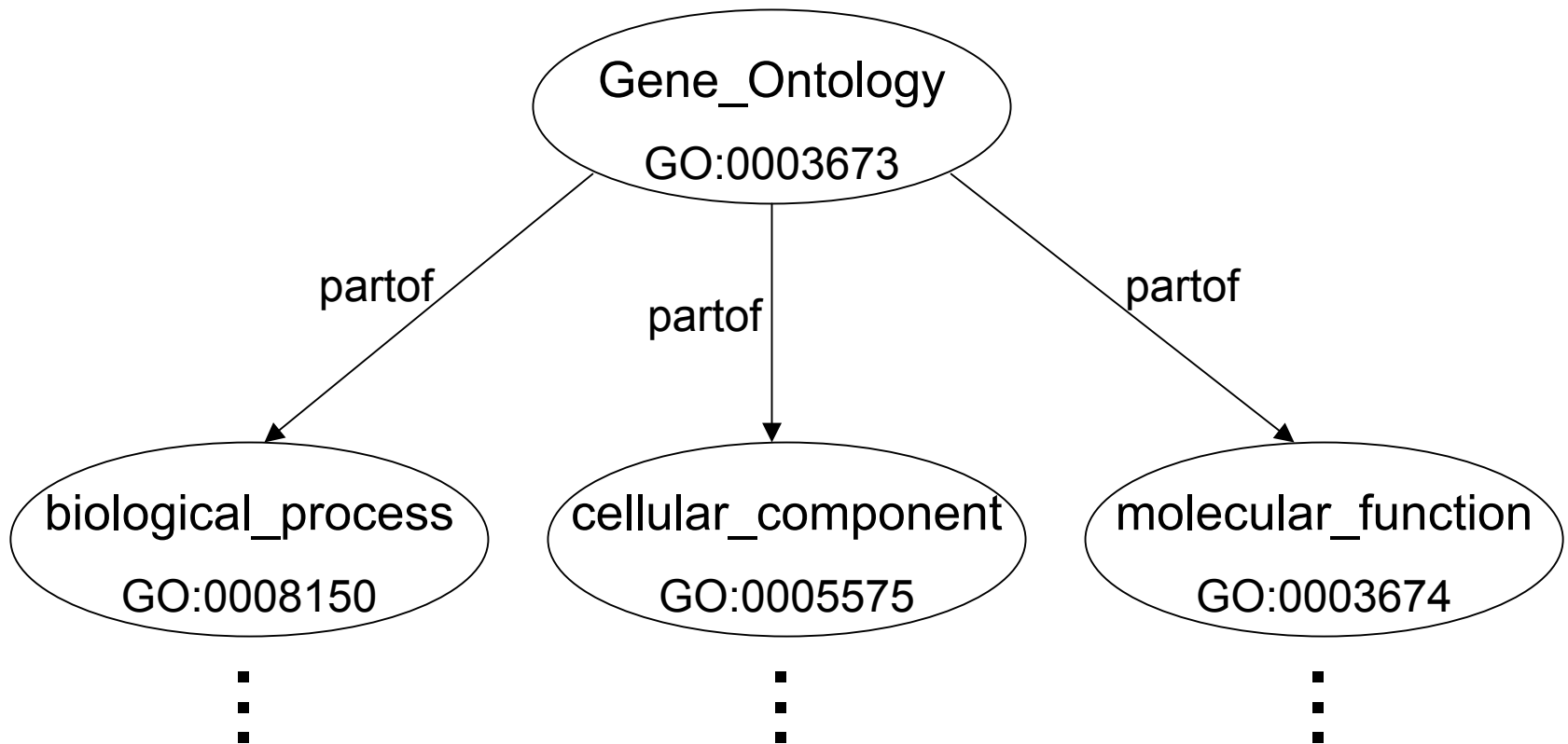
Modular Organization

Go

CAZy Family Glycoside Hydrolase Family 48
Known Activities endoglucanase (EC [3.2.1.4](#)); cellobiohydrolase (EC [3.2.1.91](#)).
Mechanism Inverting
Catalytic Nucleophile/Base Not known
Catalytic Proton Donor Not known
3D Structure Status Available (see PDB)
Note formerly known as cellulase family L.
Relevant Links [InterPro](#); [PRINTS](#)
Statistics CAZyModO(12)

#ac	Protein	Organism	Modular Organisation	Status
154	cellulase CelA	<i>Anaerocellum thermophilum</i>	1 GH9 443 445CBM3612 613 681 682CBM3834 835 880 881CBM31034 1035 1080 1061 GH48 1071	
1497	CelA	<i>Caldicellulosiruptor saccharolyticus</i>	123 24 GH9 466 467CBM3647 643 708 701CBM3857 858 903 904CBM31060 1061 1112 1035 GH48 1068	
1591	cellobiohydrolase B	<i>Cel48A Cellulomonas fimi</i>	133 34 35 GH 62GH48 665 699FN376 700FN3696 891FN3974 989CBM21090	
10828	ORF CAC0911	<i>Clostridium acetobutylicum ATCC 824</i>	66 GH48 660 661DOC1726	
1716	CelF	<i>Cel48A Clostridium cellulolyticum</i>	129 31 GH48 660 665DOC1722	3D
1730	exoglucanase S	<i>Clostridium cellulovorans</i>	132 33 GH48 66 644DOC1703	
1733	CelD	<i>Clostridium josui</i>	129 31 GH48 660 651DOC1719	

GO (Gene Ontology)

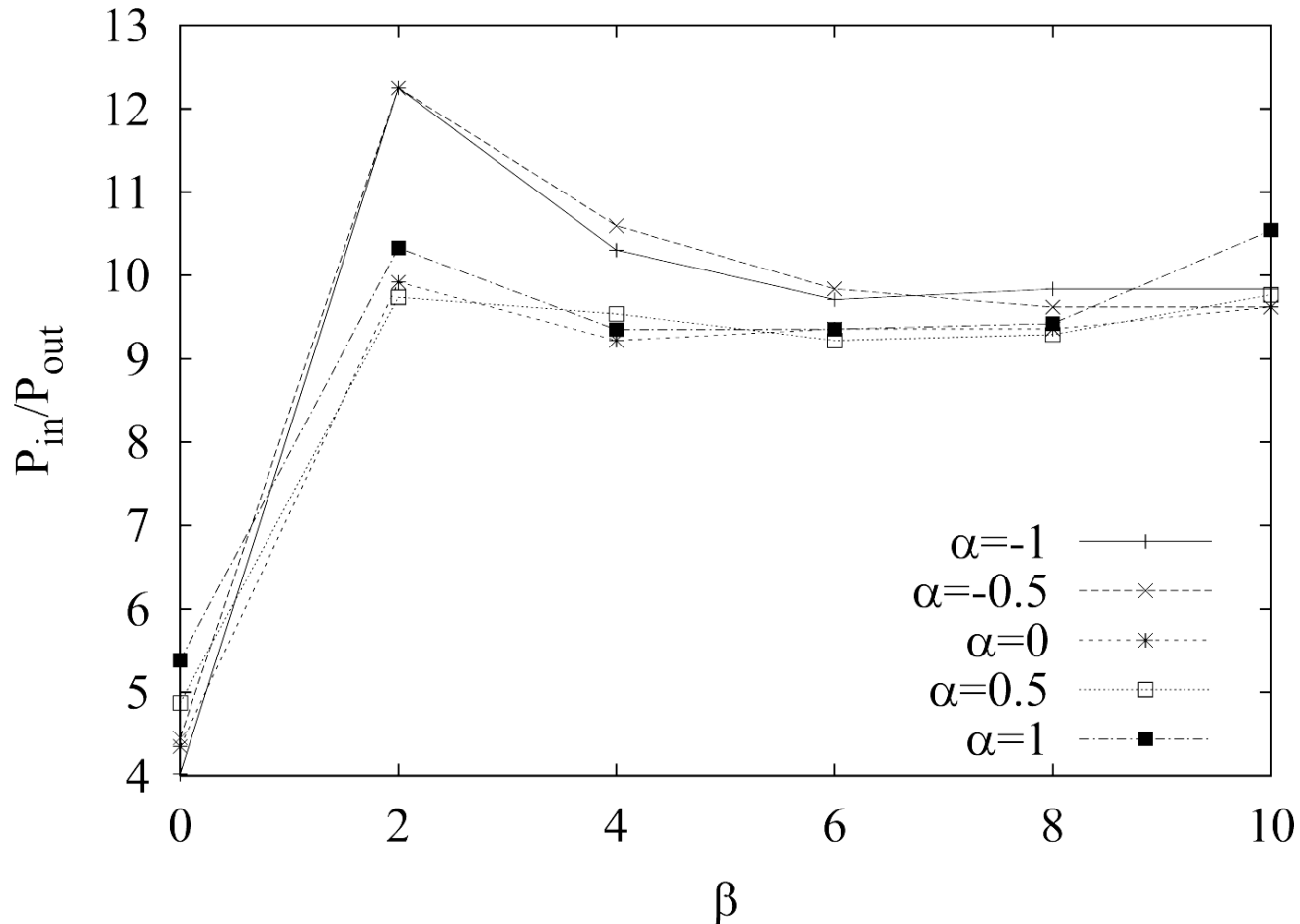


Results

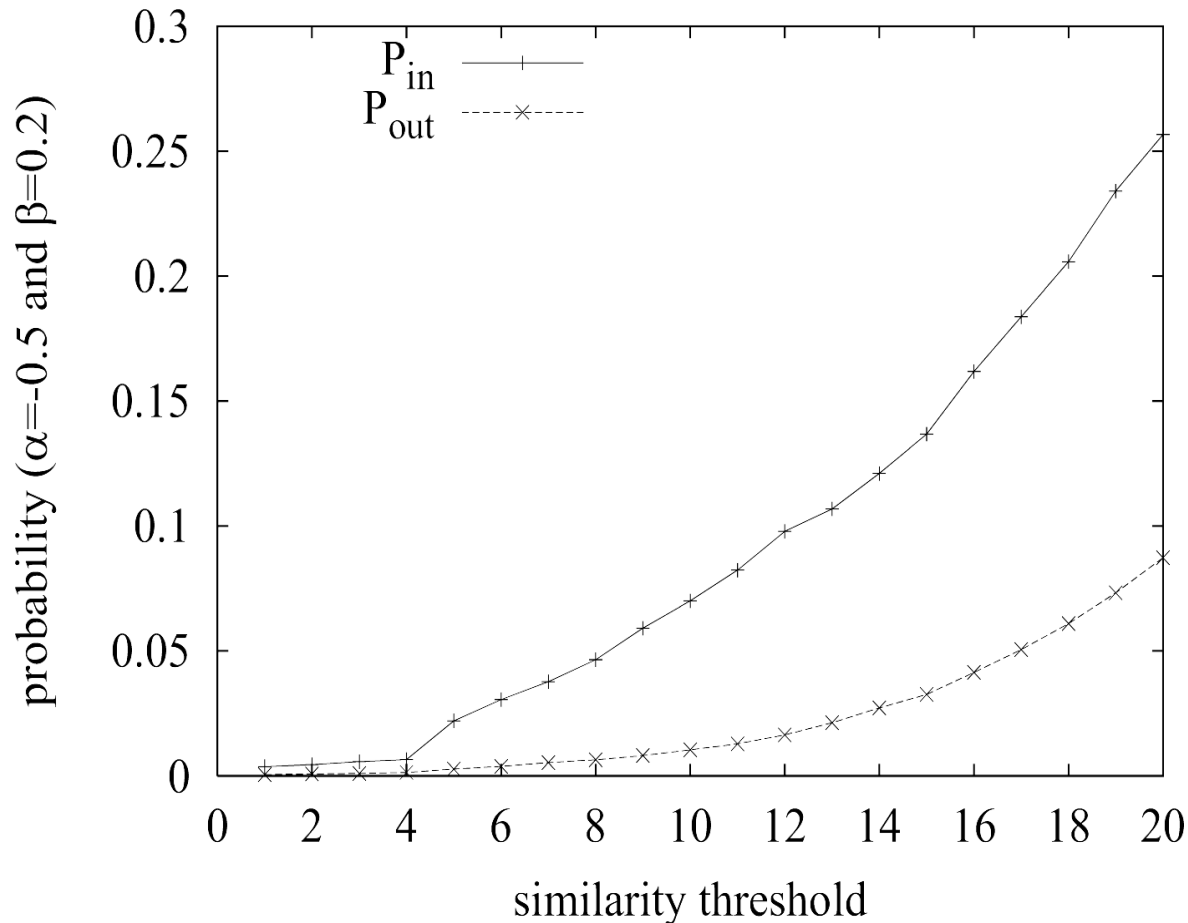
	References	Documents
GenBank(GenPept)	22849	4575
SwissProt	8998	4006
PDB	3561	785
Total		6377

13869 annotations; 6918 enzymes; 1342 terms

Family's Similar Terms (1)



Family's Similar Terms (2)



Conclusions

- Correlation between structure and function was computed from automatically extracted annotations
- The correlation metric provides an effective tool for curating extracted information