

Proxiris, a Browser Companion for Biocurators of Fungal Genomics Literature



Vahé Chahinian, Marie-Jean Meurs, David Mason, Erin McDonnell, Ingo Morgenstern, Greg Butler,

Adrian Tsang

Centre for Structural and Functional Genomics, Concordia University, Canada

Literature Mining and Curation

• Reading, interpreting, curating bio-literature • Labor-intensive, error-prone and expensive task

- \Rightarrow Natural Language Processing (NLP) techniques: ✓ extract knowledge from papers
- \Rightarrow Semantic techniques:

 \checkmark connect information from various sources

Proxiris > what you see

• Augmented browsing for literature curation

- Support genomics-based lignocellulose research [2]
- Content processing with mycoMINE [1]
- Highlight facts and entities of interest





• Provide additional content from external DBs:

BRENDA identifier, recommended and systematic name, SwissProt identifier, NCBI Taxonomy identifier, CAZy family

• Link to external knowledge sources:

BRENDA, SwissProt, NCBI, Google, Wikipedia.

Sign in to NCBI Sign in to NCBI				
Public gov PubMed Search US National Library of Medicine National Institutes of Health Advanced Help				
Display Settings: 🕑 Abstract	SpringerLink			
Biochemistry (Mosc). 2012 May;77(5):492-501. doi: 10.1134/S0006297912050094.				
Isolation and properties of	Save items			
Volkov PV, Sinitsyna OA, Fedorova E	🕆 Add to Favorites 🔻			
Sinitsyn AP.	Didahar			
Bach Institute of Biochemistry, Russian A	Sidebar			
Abstract The genes inuA and inu1, encoding two inulinases (32nd glycosyl hydrolase family) from filamentous fungi Aspergillus niger and A. awamori, were cloned into Penicillium canescens recombinant strain. Using chromatographic techniques, endoinulinase InuA (56 kDa, pl 3) and exoinulinase Inu1 (60 kDa, pl 4.3) were purified to homogeneity from the enzymatic complexes of P. canescens new transformants. The properties, such as substrate specificity, pH- and T-optima of activity, stability at different temperatures, influence of cations and anions on the catalytic activity, etc., of both recombinant inulinases were studied. PMID: 22813590 [PubMed - indexed for MEDLINE]		 inulinase hydrolase nulinase hydrolase endoinulinase exoinulinase Family hydrolase family Fungus Aspergillus niger A. awamori Penicillium canescens 		
enzyme alias	inulinase	···· 🚛 P. canescens		
brenda webpade	http://www.brenda-enzymes.org/php/result_flat.php42ecpo=3.2.1	.80		
brenda_systematicname	beta-D-fructan fructohydrolase	See reviews See all		
brenda_ecnumber	3.2.1.80			
google_search	http://www.google.com/search?g=inulinase			
brenda recommendedname	fructan beta-fructosidase			
swissprot id	Q03174			
wikipedia_search	http://en.wikipedia.org/wiki/inulinase			

features (left click, highlight toggle sidebar functionality)

• JsTree builds the sidebar

• CSS provides all the HTML markup descriptions

A successfull approach > how and why

• Evaluation: 2 curators, 114 PubMed abstracts, Proxiris prototype with limited features \Rightarrow triage time reduced by 21% ● **flexible** and **generic** ► easy integration of text mining and semantic services Proxy approach: • browser independance > no limitation in browser selection, no custom code • preserves format of original documents > pictures, tables, embedded services protected • circumvents same origin policy

• systematic processing of publications from selected origins > quickly available for users • service available only on selected web sites > high level quality of service

Acknowledgmen	t	References
Funding: Genome Canada, Génome Québec CONCORDIA		• [1] Meurs et al., Semantic text mining support for lignocellulose research, BMC MIDM, 2012
<i>Technical support:</i> Andrei Wasylyk	GEN	• [2] Murphy et al., <i>Curation of characterized glycoside hydrolases of fungal origin</i> , Database, 2011