ALIGNMENT OF LINEAR BIOCHEMICAL PATHWAYS USING PROTEIN STRUCTURAL CLASSIFICATION

Authors

Sridhar Hariharaputran, Thoralf Töpel, Timm Oberwahrenbrock and Ralf Hofestädt

Address

Bielefeld University, Faculty of Technology, Bioinformatics Department
P.O.Box 10 01 31, D-33501 Bielefeld, Germany.

Corresponding author: Sridhar Hariharaputran
Email: *sharihar@techfak.uni-bielfeld.de*
Phone: +49 – (0) - 521 – 106 - 3500

Abstract: Metabolic, signaling and regulatory pathways form the basis of biological processes and are important for the analysis of cellular behavior and evolution. This paper presents an approach of aligning biochemical pathways on the basis of the structure of involved proteins and their classification. The suitable information is retrieved from integrated database system.

SIGNALIGN is available at: http://agbi.techfak.uni-bielefeld.de/signalign/index.jsp

Introduction

Metabolic, signalling and regulatory pathways are important for the analysis of cellular behaviour and evolution. Every day molecular data is generated from high-throughput techniques and experiments. Data about proteins, genes and nucleic acids and their information are integrated to create very large scale databases, pathways and networks. Thus in-silico approach helps to analyze and understand how proteins and pathways function and interact with each other.

Databases like KEGG, GO, UniPROT, DIP, BIND etc., have lots of information about proteins and pathways. But these databases give information about an individual protein, pathway or set of proteins or pathways in particular. The question is how we can utilize the valuable information to broaden our knowledge on them and about their evolutionary and inter relationships. Current and significant research addresses more on topics like prediction of pathways, network alignment and comparison of protein interaction networks across species and also about integration of databases.

This work presents a new approach to predict and align pathways based on rudimentary biochemical knowledge. The key concept of our approach is to use structural information of the proteins involved in the pathway.

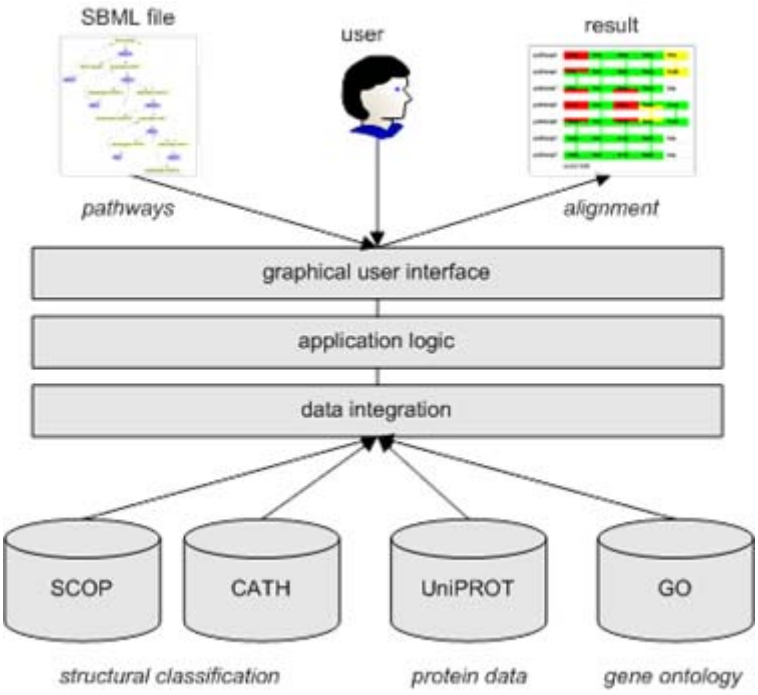
Earlier works from Dandekar et al. in 1999 for pathway alignment and later tools like PathAligner (Chen et al. 2004) and MetaPathwayHunter (Pinter et. al. 2005) detects the conserved metabolic pathways using EC numbers. PathBlast (Kelley et. al. 2004) searches for high-scoring alignments between pairs of protein interaction paths, for which

proteins of the first path are paired with putative orthologs occurring in the same order in the second path. Netalign (Liang et. al. 2006) for comparative analysis of protein interaction networks compares a query protein interaction network with the targets combining interaction topology and sequence similarity thereby identifying conserved network substructures and works by Hirsh et al. compare conserved protein complexes.

In context for our tool we utilize the protein structural and domain level classification as they are reliable compared to sequence level classification of proteins thus providing us knowledge about the homology of pathways.

Architecture and implementation

Here we propose a web-based tool, SIGNALIGN that can align pathways using integrated database information from SCOP (Murzin A et al. 1995), CATH (Orengo et al. 1997), EC number and UniProt (Apweiler et al.). The underlying concept of alignment and prediction of linear pathways applies similarity measures on the basis of structural classification of involved proteins. In our paper a pathway is defined as the set or ordered list of interacting proteins sharing similar structural and domain features with their counterparts. An alignment algorithm retrieves the corresponding structural information of the proteins from the database accompanied with appropriate visualization for clarity. Having this as the base, the difference between any two pathways can be quantified as structural similarities or identities between closely and distantly related proteins, by comparing the position of the protein domains in the classification hierarchy.



Schematic representation of SIGNALIGN system architecture: The data integration layer at the base of the system integrates suitable information from SCOP, CATH, GO and UniPROT. Based on the integrated data an application logic layer performs the alignment algorithm and controls integration of data as well as user interactions. By dint of the

graphical user interface scientists enter their pathways for alignment or select pathway elements from SBML files. The resulting alignment is visualized by a colored image.

Further this information is incorporated as a factor to the alignment scoring method along with a gap penalty, when needed. The scoring method implemented is based on the classification and amount of information content shared between the items arising from the same parent node.

The alignment result visualized with a relational coloring pattern based on the chosen classification scheme is further supported with a table and options for external information from PDB (Protein Data Bank), Gene Ontology (Ashburner et al. 2000), QSCOP (Suhrer et al. 2007) and PROCOGNATE (Bashton et al. 2008).

In addition the tool provides a search engine that mine the related inbuilt information of the proteins and about their classification schemes both in CATH and SCOP along with the relevant UniProt and PDB information. An SBML converter facility allows the import of existing SBML files directly into the system. Users can select the relevant PDB structure of the protein (s) involved in the pathway from the list projected.

Our approach enables structure based alignment of proteins and prediction of linear biochemical pathways and also to know their evolutionary relationship.

Availability

SIGNALIGN is available at: http://agbi.techfak.uni-bielefeld.de/signalign/index.jsp

Acknowledgements

Sridhar Hariharaputran would like to thank the DFG Graduate College Bioinformatics and Bioinformatics Department at Bielefeld University for funding the work.

References

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS (2004) UniProt: the Universal Protein Knowledgebase. Nucleic Acids Res. 32: D115-D119.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25-9.

Bashton M, Nobeli I, Thornton JM. PROCOGNATE: a cognate ligand domain mapping for enzymes. Nucleic Acids Res. 2008 Jan;36(Database issue):D618-22.

Chen,M. and Hofestadt,R. (2004) PathAligner: Metabolic Pathway Retrieval and Alignment. Appl. Bioinformatics, 3(4), 241-52.

Dandekar T, Schuster S, Snel B, Huynen M, Bork P (1999). Pathway alignment: application to the comparative analysis of glycolytic enzymes. Biochem J. Oct 1;343 Pt 1:115-24.

Hirsh E, Sharan R (2007) Identification of conserved protein complexes based on a model of protein network evolution. Bioinformatics Jan 15;23(2):e170-6

Kelley,B.P., Yuan,B., Lewitter,F., Sharan,R., Stockwell,B.R and Ideker,T (2004) PathBLAST: a tool for alignment of protein interaction networks. Nucleic Acids Res., 32 (Web Server issue), W83-8.

Liang Z., Xu M., Teng M and Niu L. (2006) NetAlgin: A web-based tool for comparison of protein interaction networks. Bioinformatics, 22(17), 2175-77

Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536-540.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (1997), CATH- A Hierarchic Classification of Protein Domain Structures. Structure. Vol 5. No 8. p.1093-1108.

Pinter,R.Y., Rokhlenko,O., Yeger-Lotem,E. and Ziv-Ukelson,M. (2005) Alignment of metabolic pathways. Bioinformatics, 21(16), 3401-8.

Suhrer SJ, Wiederstein M, Sippl MJ. QSCOP--SCOP quantified by structural relationships. Bioinformatics. 2007 Feb 15;23(4):513-4.

ECnumber

http://www.chem.qmul.ac.uk/iubmb/enzyme/

PDB

http://www.rcsb.org/pdb/home/home.do