MPDB: Molecular Pathways Brain Database

Gopal Madhana Vigneshwari[1], Sankaranarayana Ramamoorthy[1], Aarathy Muralikrishnan[1], Prerna Srivastava[1], Monisha pathania[1] & Sankaran Krishnaswamy[1,2]*

[1]Centre of Excellence in Bioinformatics, School of Biotechnology, Madurai Kamaraj University, Madurai - 625021; [2]Institute of Mathematical Sciences, Chennai - 600113, India; Sankaran Krishnaswamy – Email: mkukrishna@gmail.com; *Corresponding author

Abstract:
Molecular Pathways Brain Database (MPDB), is a novel database for molecular information of the brain pathways and is an initiative to provide an organized platform for researchers in the field of neuro-informatics. The database currently has information from 1850 molecules for three different sensory pathways namely olfactory transduction, photo transduction and long-term potentiation**.** The usefulness of the database is demonstrated by an analysis of the olfactory transduction pathway which helps understand their olfactory specifity and further indicates that some of the molecules have evolved independently among these organisms as per the need of time and function.

Availability: The database is available for free at http://pranag.physics.iisc.ernet.in/mpdb/

Keywords: gene, protein, brain pathways, database, MPDB

Back ground:
Neuroscience related databases help understand the brain better **[1,2].** However, they lack molecular level pathway information. This led to the creation of Molecular Pathways Brain Database (MPDB), which could be a resource for analysing molecular networks of the brain. Many types of databases like transcriptome database **[3]** and image databases **[4]** have been created in the recent decades. There is also a molecular database called DOR (Database of Olfactory Receptors) **[5]** which provides information only about olfactory receptor molecules in the OT pathway whereas MPDB concentrates on all the molecules involved in the OT pathway and also all the pathways in brain.

Methodology:
Data collection
The information related to the molecules involved in brain pathways were retrieved from data sources: KEGG, UniProt, PDB, HPRD, NCBI. Molecular interaction data was acquired from databases: STRING, IntAct, DIP, CGNC, HGNC. The basic information of molecules like gene name, gene length, chromosome location, protein name, protein length, protein function, protein localization, ligand information, protein structure, string interactions were collected and integrated. The collected information were used to create the database using

HTML, CSS, JavaScript for the front end, PHP for the interface and MySQL for the backend. The flowchart of the database construction is shown in **Figure 1**.

Structure of MPDB
MPDB is non-redundant database accessible through a web interface available at http://pranag.physics.iisc.ernet.in/mpdb/. MPDB is searched by selecting either pathways or structure from a pop-up menu. "Pathways" search allows the user to retrieve the details regarding the molecules involved in a particular pathway or the details of molecules from all pathways. The "Structure" search leads to a page where three options "Structure", "Model" and "Structural Genomics Initiative (SGI)" are provided. This helps the user in getting the information about the brain molecules that already have structure, molecules that do not have structure but can be modelled and molecules with neither structure information nor template to model. This is obtained through "Structure", "Model" and "Structural Genomics Initiative (SGI)" options respectively. "Structure" shows information like "protein name"," PDB ID", "structure resolution", "function", "protein size" and "Mutation". "Model" displays particulars about "molecule name", "Pathway involved", "PDB_ID of the template model"," identity of the template" and "query coverage". Finally, structural genomics initiative class shows

"molecules name" and "pathway" for the molecules involved in the brain which have neither structure information nor
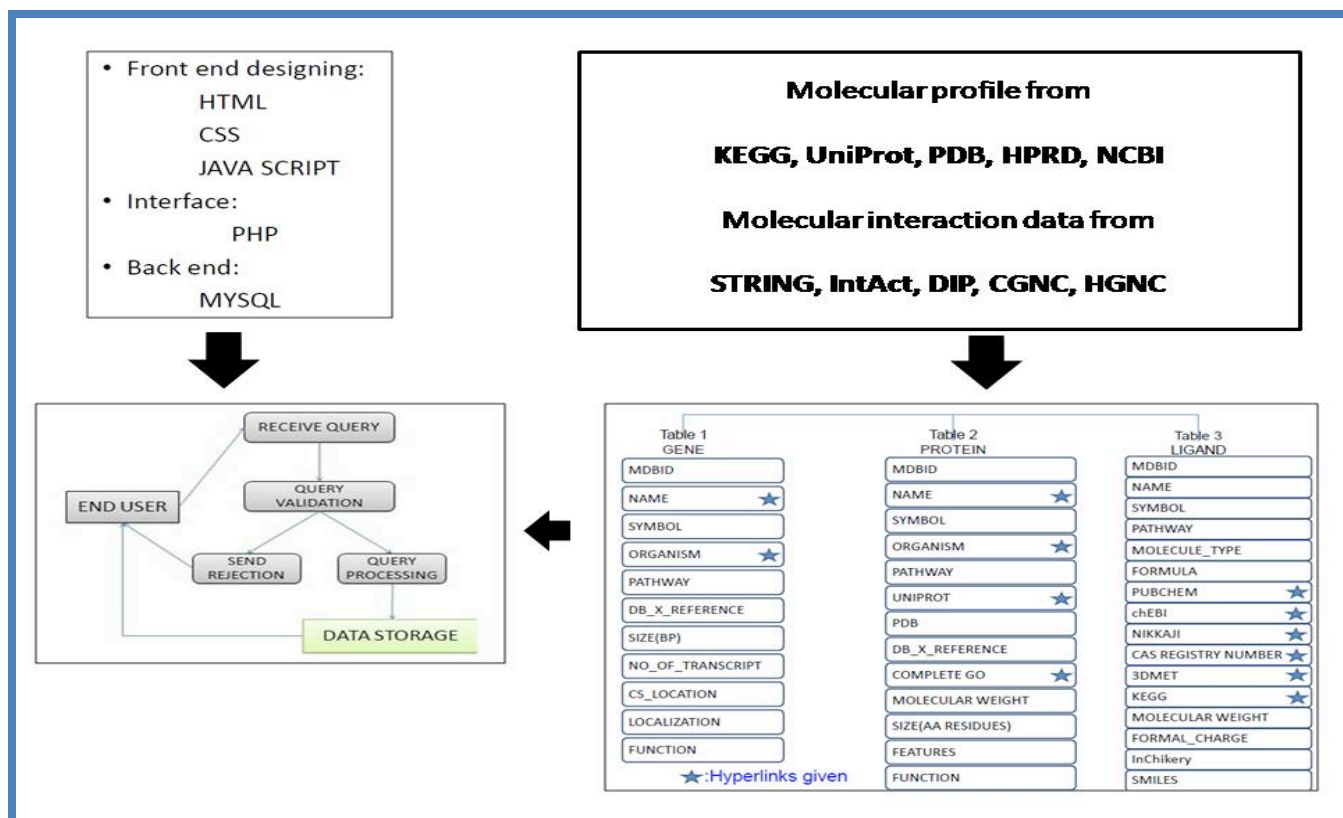
template to use for modelling.

Figure 1: Flowchart of database creation: Software tools were used for designing the front end (HTML, CSS, JavaScript), interface (PHP) and backend (MySQL). The flow of information on querying is shown. The molecular profile in terms of gene/protein ids, sequence details, structure information, ligand details were obtained from searching KEGG, UniProt, PDB, HPRD, NCBI while the interaction data were from STRING, IntAct, DIP, CGNC, HGNC. These were incorporated into tables with properties as shown for the search and retrieval.

Sequence analysis

Analysis of sequences of protein sequences of molecules from the MPDB were done with ClustalW (http:// www .ebi.ac.uk/2can/tutorials /nucleotide/ clustalw.html) and Jalview (www.jalview.org). Alignment scores for the sequences were calculated using the following formula: Alignment Score = Percentage Identity/ Alignment length. Dividing the percentage identity by alignment length ensures that, small alignments with high identity do not bias the result. Standard deviation (SD) was calculated for average percentage identity of the molecules and the range of standard deviation was calculated by the following formula: Maximum Deviation = Mean value + SD, Minimum Deviation = Mean value – SD. MEGA 5.1.4 (http:// www.megasoftware.net/) software was used for constructing phylograms. Multiple sequence alignment was implemented in MEGA using ClustalW and the phylogenetic tree was built using Neighbor-joining method. For the phylogram analysis, 100 bootstrap replications were performed.

Features

Currently, the database consists of 1850 molecules from three different sensory pathways namely olfactory transduction (OT) [6], photo transduction (PT) [7] and long term potentiation

(LTP) [8].

Analysis of olfactory transduction (OT) pathway

Olfaction is a chemosensory process of sensing smell in which volatile odorant molecules bind to specific receptors present on olfactory epithelium in nose to activate a series of chemical reactions that finally transduces the signal to the brain. About 16 molecules participate in the OT pathway cascade to achieve the function [9].

In order to find the evolutionary relationship of the OT pathway 23 molecules **(Figure 2)** were aligned and their phylograms were compared with the standard taxonomic tree. There is no reasonable comparison between the clustering of molecules based on their sequences and the standard taxonomic clustering implying differential evolutionary pressures at the macroscopic and molecular level.
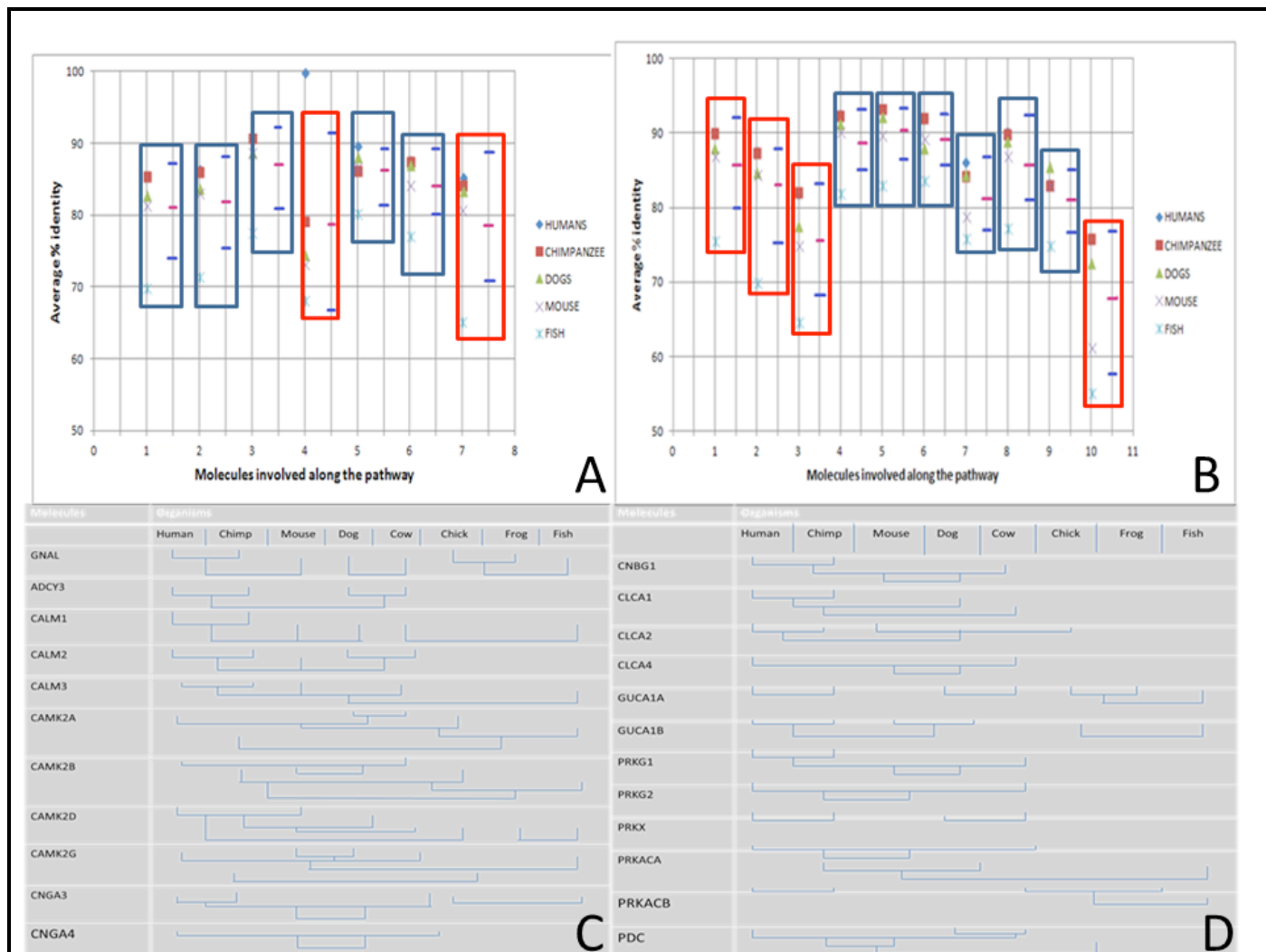
BIOMEDICAL

INFORMATICS

©2016

Figure 2: Analysis of Olfactory Transduction (OT) pathway based on data from MPDB: **A)** Plot showing the conservation and variation of molecules along the First Secondary messenger (cAMP) pathway. The numbering on X-axis indicates the molecules involved along the pathway (1-GNAL, 2-ADCY3, 3-CNGA3, 4-CALM1, 5-CALM2, 6-CALM3, 7-CAMK2A, 8-CAMK2D, 9-CAMK2G, and 10-CLCA1) and Y-axis denotes the average percentage identity; **B)** Plot showing the conservation and variation of molecules along the Second Secondary messenger (cGMP) pathway. The numbering on X-axis indicates the molecules involved along the pathway (1-GUCA1A, 2-GUCA1B, 3-PRKG1, 4-PRKX, 5-PRKACA, 6-PRKACB and 7-PDC) and Y-axis denotes the average percentage identity. The pink line indicates the Mean value and the blue line indicate the range of standard deviation (SD), i.e. Maximum deviation (Mean + SD) and Minimum deviation (Mean – SD). The blue box indicates the molecules that are conserved along the pathway of the organisms and red boxes indicate molecules that show variation among the organisms; **C)** and **D)** Clustering of organisms based on the molecular information from Molecular Pathways Brain Database.

As OT pathway is a dual secondary messenger pathway **[10]**, 7 molecules from first secondary messenger pathway and 10 molecules from second secondary messenger pathway for five different organisms (for which molecule information is available) were taken for the analysis. It was found that in first secondary- messenger pathway the initial and final molecules like CALM1, CALM2, CALM3, CAMK2A, CAMK2D, CAMK2G show variations among organisms while the molecules like GNAL, ADCY3, CNGA3 and CLCA1 in the middle of the pathway are more conserved. However, the second secondary messenger pathway does not show this pattern and only molecules like PRKX and PDC showed significant variation among organisms. Though the same molecules are involved in

the pathway among different vertebrates, the number of components of these molecules varies from organism to organism. Amphibians have only second secondary messenger pathway and the first secondary messenger pathway is not developed in them. Molecules like CLCA2 and CLCA4 are only present in terrestrial vertebrates and are not necessary for olfaction in aquatic life.

The analysis of the OT pathway using the MPDB helps in understanding about how individual molecules evolved during species divergence. The variation in number of components in the pathway among organisms perhaps confers specificity for their ability to smell. The molecules seem to have evolved

independently among organisms as per their functional requirements.

Further development:
Pathways such as taste transduction, long term depression, cholinergic synapse, Dopaminergic synapse, GABAergic synapse **[11]** need to be incorporated with specific reference to neurodegenerative diseases.

References:
[1] Van Essen DC, Curr Opin Neurobiol. 2002 **12**: 574 [PMID: 12367638]
[2] Skoufos E *et al. Nucleic Acids Res*. 2000 **28**: 341 [PMID: 10592268]
[3] Sun J *et al. Mar Biotechnol(NY)*. 2013 **15**: 244 [PMID: 22875536]
[4] Rosano C *et al. Neuroepidemiology.* 2005 **24**: 221 [PMID: 15832060]
[5] Nagarathnam B *et al. Bioinform Biol Insights*. 2014 **8**: 147 [PMID: 25002814]
[6] Hildebrand JG, *Proc Natl Acad Sci U S A*. 1995 **92**: 67 [PMID: 7816849]
[7] Jindrova H, *Physiol Res*. 1998 **47**: 155 [PMID: 9803480]
[8] Malenka RC & Nicoll RA, *Science*. 1999 **285**: 1870 [PMID: 10489359]
[9] Paysan J & Breer H, *Pflugers Arch*. 2001 **441:** 579 [PMID: 11294238]
[10] Ache BW & Zhainazarov A, *Curr Opin Neurobiol*. 1995 **5:** 461 [PMID: 7488847]
[11] Kanehisa M *et al. Nucleic Acids Res*. 2016 **44**: D457 [PMID: 26476454]

©2016