VPDB: Viral Protein Structural Database

Om Prakash Sharma, Ankush Jadhav, Afzal Hussain, Muthuvel Suresh Kumar*

Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Pondicherry-605014, India; Muthuvel Suresh Kumar - Email: suresh.bic@pondiuni.edu.in; Phone: +91 413-2654583; Fax: +91 413 2655211; *Corresponding author

Abstract:
Viral Protein Database is an interactive database for three dimensional viral proteins. Our aim is to provide a comprehensive resource to the community of structural virology, with an emphasis on the description of derived data from structural biology. Currently, VPDB includes ~1,670 viral protein structures from >277 viruses with more than 465 virus strains. The whole database can be easily accessed through the user convenience text search. Interactivity has been enhanced by using Jmol, WebMol and Strap to visualize the viral protein molecular structure.

Availability: http://www.vpdb.bicpu.edu.in

Keywords: Database; viral proteins; virus; Jmol; STRAP; WebMol.

Background:
Viruses have always been a major cause for a large number of infectious diseases. Furthermore, recent emergences of new viral diseases have been of concern in normal human activities. Molecular knowledge of viral protein is thus seen to play an important role for developing improved protein-based vaccines, designing novel anti-viral agents and understanding the entry mechanisms of viruses. Although there exists numerous databases and resources for virus structure information, a comprehensive viral protein structural database is unavailable till the date which can provide detailed sequential as well as structural information on each viral protein (**Table 1 see Supplementary material**). ViperDB [**1**], VIDA [**2**], VirusMint [**3**], and PhEVER [**4**] are the few databases available online for viral proteins. ViperDB archives the icosahedral virus capsid structures, whereas VIDA incorporates open reading frames of animal virus proteins. The VirusMint database contains protein interactions between viral and human proteins from 490 unique viral proteins of more than 110 different strains. PhEVER is a recently developed database that aims to provide accurate evolutionary and phylogenetic information for the analysis of virus-virus and virus-host lateral gene transfer. There is currently no unique resource available till the date for viral protein structure along with their detailed annotated binding interaction information. To address this problem, we have developed VPDB as a comprehensive derived database from the analysis of a large set of viral proteins obtained from PDB [**5**], PDBsum [**6**], and UniProtKB [**7**]. VPDB comprises of derived data from viral proteins designed to serve as a single-stop solution for retrieval and analysis of viral proteins from all the viral groups and species whose 3-dimensional structure has been solved and present in PDB.

Database design:
VPDB is based on three layer structure, namely data layer, middle layer and presentation layer. Data layer is consisted of MySQL, whereas middle layer comprises server side PHP and Java scripts. User friendly web based online presentation layer was developed using PHP, Java, HTML and CSS. The database system is implemented in Red Hat (Linux). It employs Apache as a web-server.

Result and Discussion:
VPDB is a user friendly database, which allows the user to input their query text and accordingly select the query type (PDB IDs or Keywords, Author name or Title of the article and Experimental methods) in order to retrieve the related information from the database. Query related hits are displayed in a result page. User can then select the protein of interest by selecting PDB IDs (**Figure 1a**) for further detailed information.

Database description:
Result page has been further classified into nine different tabs. Each tab has its own web page that provides a larger set of data. Summary tab contains the summary of the five sub categories under the five sub headings such as molecular summary, structural summary, experimental details, general annotation and citation (**Figure 1b**). Sequence tab contains information about the protein sequence, sequence length, chain type, source method, protein sequence in fasta format and UniProtKB reference. Furthermore, secondary structure of protein has been computed using DSSP server [**8**] for the secondary structural analysis of proteins. Knowledge of the binding pocket is also well known to play an important role for drug designing and existing drug refinement. To make VPDB more comprehensive, the largest binding pocket is identified from the CASTp server based upon the area-volume, although the result shows high but not absolute correlation between pocket volume, and binding affinities. The MolProbity tab contains the MolProbity Ramachandran analysis for the three-dimensional protein. It is a web based tool offering quality validation for three-dimensional structure of proteins. The LigPlot tab contains the name of ligand, LigPlot and Nucplot interaction data (**Figure 1e**). Three java based open-source tools Jmol, WebMol and STRAP has been incorporated in the database for tertiary structure analysis. These tools require a java-enabled browser. Reference tab summarizes the detailed literature information obtained from PubMed. Moreover, Quick links tabs and download tab has been provided for other related database links and to download the database files for viral protein respectively.
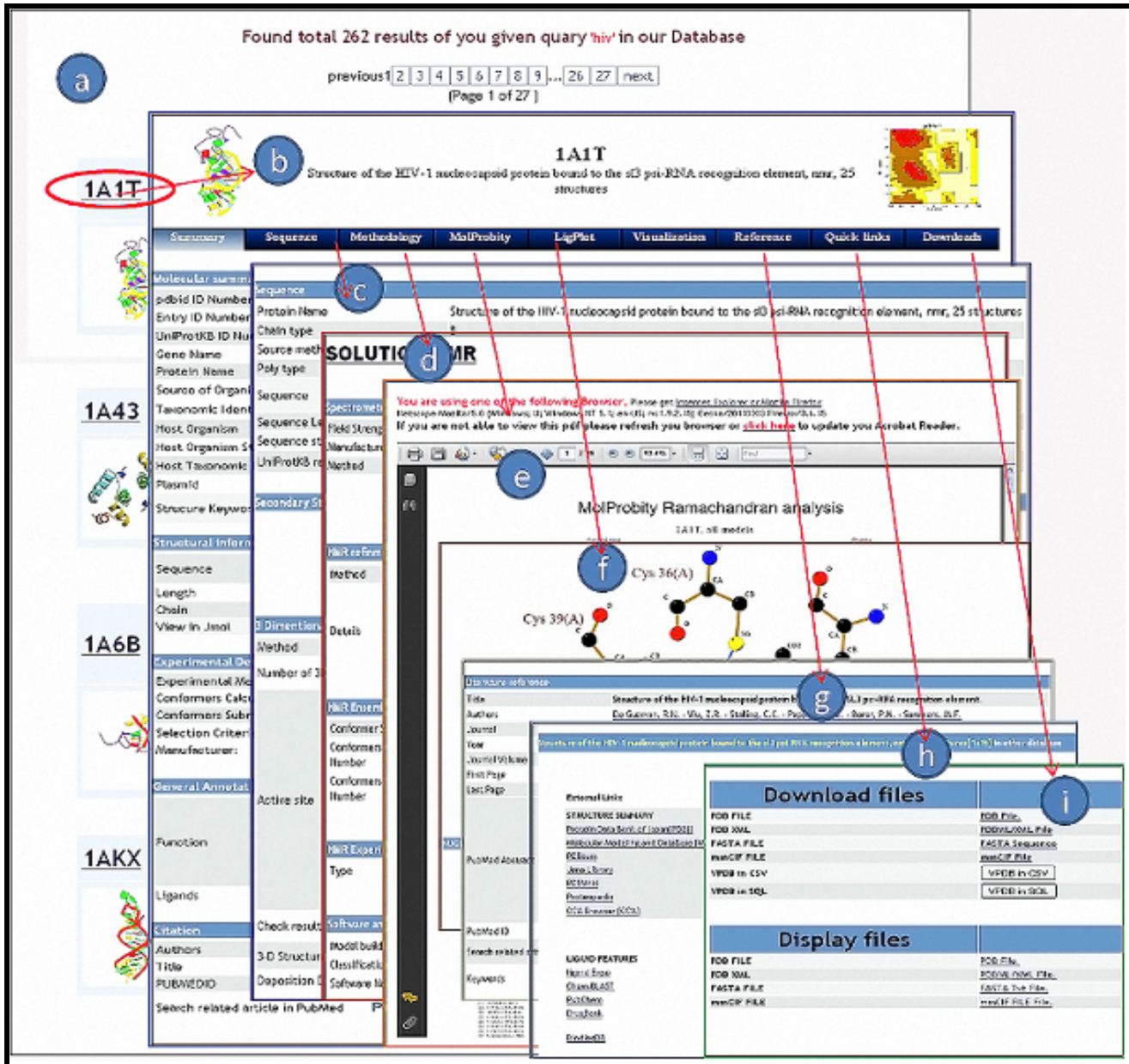
Figure 1: A snapshot of VPDB web interface, showing detailed results with various tabs and options. (a) the list of the hits as a result of a given query for HIV virus, (b) Summary tab: It contains the detailed information of the selected protein from the list (1A1T), (c) Sequence tab: It contains the sequence information, secondary structure and 3D model, (d) Methodology tab: Parameters for the 3 Dimensional protein building method, (e) MolProbity tab,(f) Ligplot tab, (g) Reference tab: It contains the journal article in detail, (h) Quick links tab, and (i) Download tab.

Further, for the comparison of viral proteins with the other database, VSE (Virus Search Engines) offers search tools to retrieve the data instantly from source databases. All these information embedded in a single database makes VPDB a better resource for viral protein analysis in comparison to other available virus databases. We believe that VPDB, due to its rich information content will be a useful and novel resource for researchers in the field of structural virology providing an insight into the molecular mechanisms brought about by different viral proteins.

References:
[1] Shepherd CM *et al*. *Nucleic Acids Res*. 2006 **34**: D386 [PMID: 16381893]
[2] Alba MM *et al*. *Nucleic Acids Res*. 2001 **29**: 133 [PMID: 11125070]
[3] Chatr-aryamontri A *et al*. *Nucleic Acids Res*. 2009 **37**: D669 [PMID: 18974184]
[4] Palmeira L *et al*. *Nucleic Acids Res*. 2011 **39**: D569 [PMID: 21081560]
[5] www.pdb.org
[6] Laskowski RA. *Nucleic Acids Res*. 2001 **29**: 221 [PMID: 11125097]
[7] Boutet E *et al*. *Methods Mol Biol*. 2007 **406**: 89 [PMID: 18287689]
[8] Kabsch W & Sander C. *Biopolymers*. 1983 **22**: 2577 [PMID: 6667333]

Edited by P Kangueane
Citation: Sharma *et al*. Bioinformation 6(8): 324-326 (2011)

BIOINFORMATION

Supplementary material:

Table 1: Comparison of VPDB with other Viral Protein databases.

Database	About the database	Contents	Proteins	Search options	3d -viewer
VPDB	VPDB is database of comprehensive set (molecular structure, sequence, molecular interaction etc.) of virus protein. It includes all kind of virus proteins. E.g. – plant virus, human virus etc. http://vpdb.bicpu.edu.in	It includes all kinds of viral proteins e.g.- plant virus, human virus, animal virus etc. Stores Viral Protein Structural information.	1,750	PDB IDs, Keywords (protein name, virus name, family, gene name etc.), Author name, Title of the journal, Experimental methods.	Java based Jmol, WebMol and STRAP.
VIPERdb	VIPERdB is a database for icosahedral virus capsid structure. http://viperdb.scripps.edu/	It includes animal viral proteins with their structural information.	298	PDB IDs, virus name, capsid name, family based, x-ray entries based.	3D I.AU.,φ-ψ Explorer.
VirusMINT	VirusMINT is a database which shows interaction between human and viral proteins. http://mint.bio.uniroma2.it/virusmint/Welcome.do	It includes human viruses like influenza virus, hepatitis virus with general bibliographic information and their interactions with host protein.	428	Protein accession number, protein name, domain name, gene name.	VirusMint viewer.
VIDA	VIDA database contains a complete collection of homologous protein families derived from open reading frames from complete and partial virus genome. We can get conserved region, function and taxonomy information. http://www.biochem.ucl.ac.uk/bsm/virus_database/VIDA.html.	VIDA database contains conserved region, function and taxonomic information.	5477	Virus name, function class, homologous protein family number, gene bank protein number	Not available
ViralZone	ViralZone is a database for viral genus and families, providing general molecular information. http://expasy.org/viralzone/	ViralZone includes all kinds of virus with their taxonomic and molecular information.	15,700	Sequence search, families search and BLAST search.	Not available