Journal of
Proteomics & Bioinformatics

Sharma et al., J Proteomics Bioinform 2012, 5:3
DOI: 10.4172/jpb.1000217

Research Article Open Access

Structural Epitope Database (SEDB): A Web-based Database for the Epitope, and its Intermolecular Interaction Along with the Tertiary Structure Information

Om Prakash Sharma, Arindam Atanu Das, Krishna R, Suresh Kumar M and Premendu P Mathur*

Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Pondicherry-605014, India

Abstract

An epitope is an antigenic determinant. It is a part of antigen, which is recognized by specific receptor of the immune system, exclusively by antibodies, B-cells or T-cells. Thus, exhaustive molecular knowledge of epitopes will have great implications for vaccine and drug design. Existing epitope databases are lacking complete structural information of epitopes. Therefore, we have developed an open-access database for describing the three-dimensional structure of epitopes and its interaction with antigens and antibodies. We have cataloged 619 epitope structures from Protein Data Bank and published research articles. Our database provides an extensive collection of B-cell, T-cell and MHC binding proteins. For user convenience, these epitopes have been classified based on structure (Linear epitope and Discontinuous epitope) and cell types (B-cell, T-cell and MHC binding). Our database incorporates specialized features of epitopes, including complex ID, epitope sequence, epitope length, epitope position, source information, structure determination method, taxonomic ID and antigen-antibody interacting residues. It also summarizes multi domain protein's structural information like protein name, polypeptide type, molecule name, chain ID, fragment length, sequence length and amino acid sequence. Additionally, it provides Gene-Ontology, MolProbity and Epitope visualization tool for better and clear understanding of the epitope conformation and its interaction with protein. The user interface is cross linked to other public databases like IEDB, UniprotKB, PDB, NCBI and other existing epitope database for making it more informative. The user friendly web page is freely available at URL http://sedb.bicpu.edu.in.

Keywords: Epitope; Epitope viewer; Database; DIMPLOT; Structural epitopes

Introduction

An epitope is defined as a chemical structure admitted by antigenic specific receptor of the immune system. Exhaustive knowledge about their intrinsic structure and mode of action allow us to gather and comprehend the immune response which is required to uncover, observe and fight with chronic diseases. Epitopes provides a powerful platform for designing intervention strategies and drugs, as well as new vaccine candidates [1-4]. A number of new vaccine candidates based on the T-cell and B-cell epitopes have been found, and some of them are currently under clinical trials [5-7]. For example, two structural epitopes in SEDB (1YJD and 1S78) are targets of two FDA approved drugs (Herceptin and Erbitux), which are effective in treating some cancers [8,9].

Over the last two decades, the amounts of information associated with epitopes have tremendously increased. Using "epitope" as a keyword, a query has been made against the PubMed (frequently used literature database) and Protein Data Bank (PDB) (protein tertiary structure database). PubMed search has shown 127220 hits and PDB exhibits 407 hits on 20-12-2011. As a consequence, there is a need to analyse and store these data. There are many databases available on T-cell and B-cell epitopes [10,11]. Although, none of the databases are exclusively developed for epitope three-dimensional structures except IEDB-3D [12] (Table 1). However, it lacks some of the useful and related information of epitopes, like 3-D structural information of epitopes and diffraction data of determining the arrangement of atoms within a crystal of proteins. Gene-Ontology information, Ag-Ab interaction graph, epitopes location in protein with interaction data has been not found in any available epitope databases. We took initiative to collect and add all these unique features of epitopes in a single database. A user friendly web based database has been developed to access these data publicly. We named it SEDB (Structural Epitope Database).

SEDB contains 619 epitope from B-cell, T-cell and MHC bound molecules. Each entry provides detailed information of antigen-antibody interaction data similar to IMGT/3D 3D [13] and IEDB-3D database. Additionally, we have shown Ag-Ab interaction graph, which offers more data to analyse protein-epitope complex. We have incorporated MolProbity (Ramachandran protein analysis graph) for each protein to understand the possible conformations of ψ and φ angles of a polypeptide which can help researchers to evaluate the tertiary structure of proteins.

The objective of this database is to serve a single stop platform for retrieval and analysis of all types of epitopes (T-cell, B-cell and MHC) to the research group of structural biologists and immunologist. We believe that, due to its comprehensive information of structural epitopes, SEDB will provide a valuable resource for the investigators working in the area of immunological research and vaccine design.

***Corresponding author:** Premendu P Mathur, PhD, Department of Bioinformatics, Pondicherry University, Pondicherry-605014, India, Tel: +9443368111; Fax No: +91 413 2655211; E-mail: ppmathur@gmail.com, ppmathur@yahoo.com

Received January 09, 2012; **Accepted** January 31, 2012; **Published** February 10, 2012

Citation: Sharma OP, Das AA, Krishna R, Suresh Kumar M, Mathur PP (2012) Structural Epitope Database (SEDB): A Web-based Database for the Epitope, and its Intermolecular Interaction Along with the Tertiary Structure Information. J Proteomics Bioinform 5: 084-089. doi:10.4172/jpb.1000217

Materials and Methods

Database architecture

SEDB is based on three-tier architecture [14]. The first layer or presentation layer is responsible for the presentation of the data and user interface, which enable the user to interact with the second-tier processes in a secure and intuitive manner. Second layer is known as application layer or logic layer; it manages the logic of the application and permits access to the third-tier services. Finally, the data layer protects the direct access by the client components residing within a secure network. Interaction must occur through the second-tier processes (Figure 1).

Here, data has been stored into MySQL [15] and user interface has been developed using Java, HTML and CSS. PHP have been used for both application layer as well as a presentation layer for accessing the data from the database. Apache server has been used as a web-server and the database system is implemented in Red Hat (Linux).

Database curation

All entries in this database are sourced from articles published in peer-reviewed journals and PDB. Initially, exhaustive queries were made against PDB [16], IMGT/3D, PDBsum [17] Bcipep, MHCBN [18] and IEDB-3D databases. It returns more than 1200 structures. All the PDB IDs were copied into the Microsoft-Excel. The PDB IDs were then filtered manually to exclude the redundant entries and to classify the structural epitope into B-Cell, T-Cell and MHC molecules. Based on the published article, epitope sequence has been collected from the protein sequence database (PDB, UniProt and NCBI). DIMPLOT has been run to generate the epitope interaction data. Moreover, Gene Ontology data has been retrieved from the PDBsum to complement the current database.

Resource	Purpose	Contents (Oct 2011)	Data (Oct 2011)	Limitation
SEDB	The SEDB describes the 3D-structure of epitope containing proteins and its interaction with protein. It also summarizes the source information, experimental details and additional related information of epitope such as methodology and detailed sequence information along with user convenient Jmol tool for epitope visualization. URL: **http://sedb.bicpu.edu.in**	It contains all the structure of TCR, MHC, and BCR complexes available in PDB, with emphasis on the structural characterization of these complexes. Additionally, It also includes Ramachandran Plot, DIMPLOT, and Methodology and Gene Ontology information. Reference in details and the Assay information. It complements other existing specialized 3D epitope databases.	B-Cell - 271 T-Cell - 49 MHC - 299 Extracted from the Protein Data Bank (PDB) and articles published in peer review journals.	It complements other existing specialized 3D epitope databases.
BEID	BEID describes the sequence-structure-function information on immunoglobulin (Ig)-antigen interactions URL: **http://datam.i2r.a-star.edu.sg/BEID**	The current version of the database contains 164 antigens, 126 Ig and 189 Ig-Ag.	Antigen (Ag) -164 Immunoglobulin (Ig)-126 Ig-Ag complexes-189 Extracted from the Protein Data Bank (PDB).	Comprehensive structural information of protein and epitope, Rama-chandran plot, Sequence information, Gene Ontology, Visualization tools. Developed for only B-Cell epitope. Not updated from long back.
MPID-T	MPID-T database is developed for sequence-structure-function information on T cell receptor/peptide/MHC interactions. URL:**http://variome.bic.nus.edu.sg/mpidt/index.html**	It contains all structures of TcR/pMHC and pMHC complexes, with emphasis on the structural characterization of these complexes.	pMHC - 187 TcR/pMHC - 16 Extracted from the Protein Data Bank (PDB).	Structural information of protein, Rama-chandran plot, Gene Ontology, Visualization tools, Sequence information Developed for only T-Cell epitope. Not updated from long back.
Epitome	Epitome is a database of all known antigenic residues and the antibodies that interact with them, including a detailed description of residues involved in the interaction and their sequence / structure environments. URL: **http://www.rostlab.org/services/epitome/**	It locates the CDRs in the known protein–antibody complexes through the knowledge-based approach	Ag - 142 Extracted from protein–Ab complex structures	Structural information of protein, Rama-chandran plot, Gene Ontology, Sequence information, Developed for only B-Cell epitope. Not updated from long back.
IEDB-3D	IEDB-3D is fully embedded within IEDB, thus allowing structural data, both curated and calculated, and all accompanying information to be queried using multiple search interfaces. URL: **http://www.iedb.org**	IEDB-3D include experiments describing recognition of epitopes or Ag (peptidic or nonpeptidic) by TCRs (T-cell assays describing 3D structures of antigens / epitopes in complexes with MHC and TCR), immunoglobulin or antibodies (B-cell assays describing structures of Ab–Ag complexes), and MHC molecules	B-Cell - 270 T-Cell -33 MHC -222 Extracted from the Protein Data Bank (PDB) and articles published in peer review journals.	Structural information of protein. , Rama-chandran plot, Gene Ontology, and the DIMPLOT for the interaction data.
IMGT/3D	IMGT/3Dstructure-DB provides the closest genes and alleles that are expressed in the AA sequences of the 3D structures, by aligning these sequences with the IMGT domain reference directory. URL:**www.imgt.org/IMGTindex/IMGT3Dstructure-db.html**	IMGT/3Dstructure-DB contains 3D structures of (1) AgR that comprise Ig or Ab and T CR, (2) MHC proteins of class I and class II, (3) peptide/MH (pMH) complexes (pMH1, pMH2), (4) AgR/Ag complexes (IG/Ag, TR/pMH), and (5) related proteins of the immune system (RPI)	All entries in this database are manually curated from articles published in peer review journals.	Structural information of protein, Rama- chandran plot, Gene Ontology, Sequence information
CED	CED has been developed as an information resource for investigators involved in both theoretical and applied immunology research. URL: **http://immunet.cn/ced/**	CED, provides a collection of conformational epitopes and related information including the residue make up and location of the epitope, the immunological property of the epitope, the source antigen and corresponding antibody of the epitope.	293 conformational epitopes. All entries in this database are manually curated from articles published in peer review journals.	Structural information of protein, Rama-chandran plot, Gene Ontology. Not updated from long back.

Table 1: Comparison of existing epitope databases with our database (SEDB).

Contents

SEDB provides structural and Ag-Ab interaction information, through user friendly interference. Currently, database comprises 619 entries. The statistics of the experimentally determined epitope can be surveyed in the (Figure 2). An Epitope structure, which has been experimentally proved and available in PDB, has been added in our database.

Our database includes a broad knowledge about epitopes in nine different tabs; I) Summary, II) Sequence information of proteins, III) Methodology, IV) Gene Ontology, V) Reference, VI) MolProbity, VII) Ag-Ab intPlot, IX) EpiViewer and VIII) E-Links.

Results and Discussion

SEDB is a user-friendly planned web interface, so that the users can comfortably get the desired information online. It requires Java run time environment and upgraded version of internet explorer to visualize the data.

The current version of SEDB contains 619 3D-Epitope complexes from B-cell (271), T-cell (49) and MHC binding (299). The drop-down search menu in the home page of database allows advance search options to fetch the data in four different ways of any specified field in the database. User may query the database by selecting one of the query options among (1) PDB ID or Keyword, (2) Author name or title of the article, (3) Experimental method, and (4) Sequence search (based on text match). It will catalogue all the related hits found in the database. User can click on the desired protein id to see the information in detail. For each protein ID, a result has been shown in nine different tab options (Figure 3) which have been discussed below.

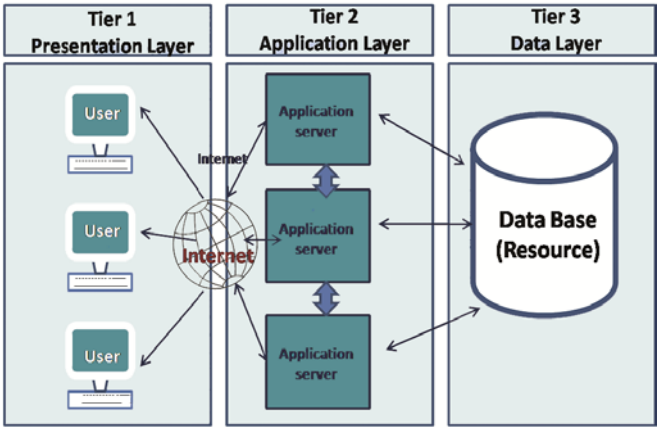

Figure 1: Architecture of the SEDB database.

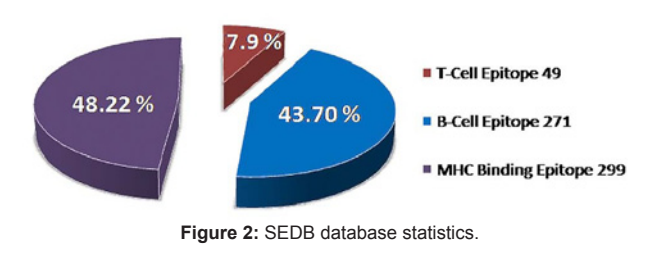

Figure 2: SEDB database statistics.

Summary tab

Summary tab includes name of the PDB complex, IEDB id (if exists), Immunization, epitope sequence with position and type of epitope under the sub heading of epitope whereas assay antigen, PDB category, Ab-Ag interacting residues, comments, and the structural determination methods have been summarized under assay information.

Sequence tab

Sequence tab contains the name of the protein, source method and name of the protein fragment; moreover, each protein sequence has been colored according to the amino acid properties [19] and related information has been added in this tab such as sequence length, chain type and UniProtKB id.

Methodology tab

It contains detailed information about the essential parameters for building three dimensional models of proteins. This tab designates various parameters that have been taken for solving the 3D structures of proteins using well known structural determination techniques such as electron microscopy, x-ray crystallography, NMR spectrometry etc (Figure 3). These data has been taken manually from the PDB database.

Gene ontology

Gene Ontology (GO) tab provides ontology of defined terms representing gene product properties. The ontology covers mainly three domains: Cellular component, Molecular function, and biological process. The GO information has been collected from PDBsum.

Reference tab

Reference tab summarizes the various literature references for detailed information obtained from PubMed. The tab includes title of the article, author name, year of publication, journal name, journal volume, first page and the last page number, PubMed id and an abstract of that particular article.

MolProbity Ramachandran analysis tab

MolProbity Ramachandran analysis tab has been included in the database to evaluate the quality of three-dimensional structures of proteins. MolProbity is used to analyze the possible conformations from ψ and φ angles for a polypeptide and shows the percentage of amino acid residues present in a favored, allowed and outlier regions [20].

Ag-Ab intPlot tab

Ag-Ab intPlot tab summarizes the antigen-antibody interaction, which has been generated using the DIMPLOT command from the LIGPLOT software [21]. DIMPLOT produces a plot of the interactions across either a dimer or a domain-domain interface. The interactions plotted to include hydrogen bonds and non-bonded contacts.

EpiViewer

EpiViewer contains the Jmol molecular visualization tool for structural analysis of epitope sequence and antigen-antibody complexes. Here we have highlighted the epitope into the Space-fill model, so that user can easily identify the epitope position in the protein. It requires java run time environment to display the molecule. User can download the antigen-antibody interaction files in PDB format; option has been given in this tab.

Citation: Sharma OP, Das AA, Krishna R, Suresh Kumar M, Mathur PP (2012) Structural Epitope Database (SEDB): A Web-based Database for the Epitope, and its Intermolecular Interaction Along with the Tertiary Structure Information. J Proteomics Bioinform 5: 084-089. doi:10.4172/jpb.1000217

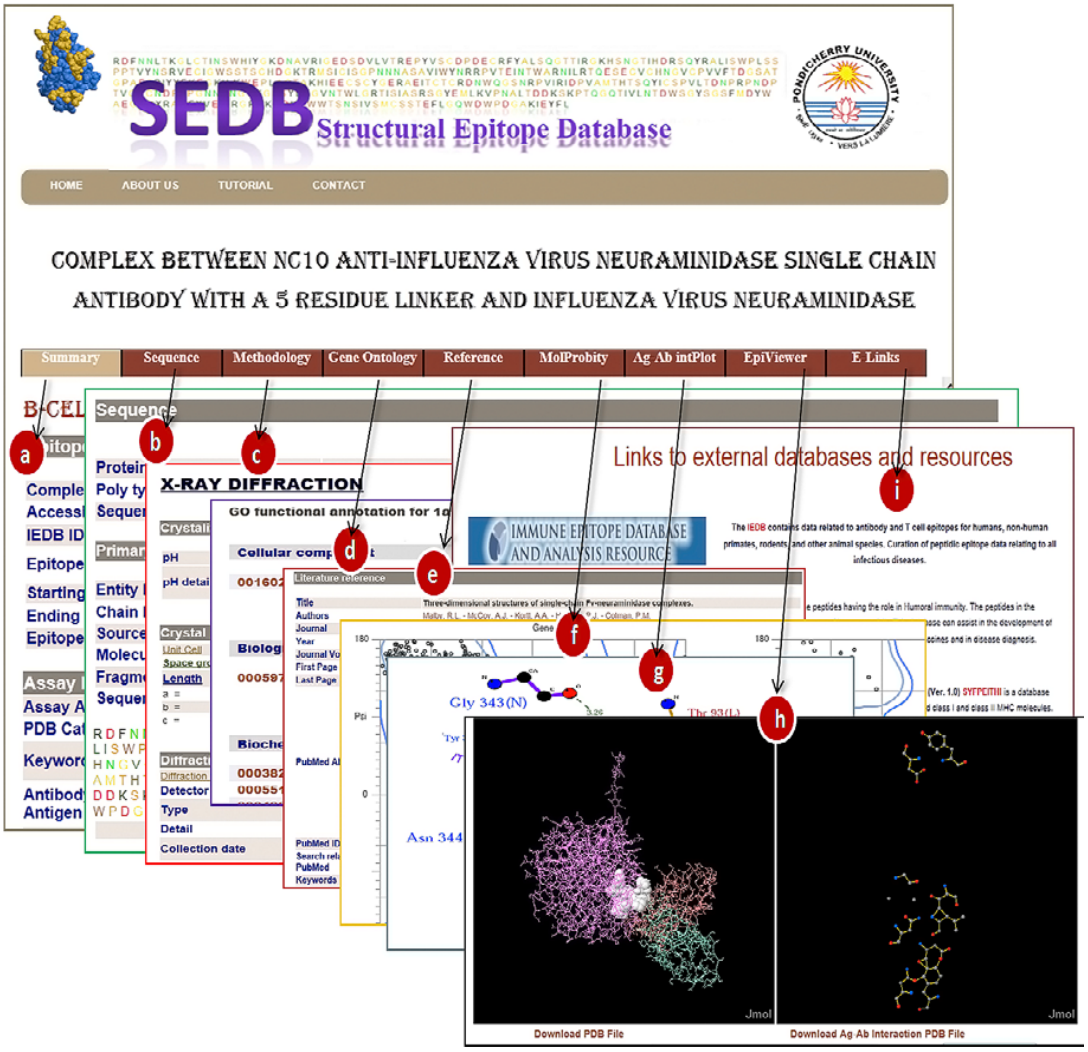

Figure 3: Snapshot of the SEDB result tab. (a) Summary tab: It contains the detailed information of the selected protein from the list, (b) Sequence tab: It contains the sequence information, (c) Methodology tab: It displays the fundamental parameters for the three dimensional protein building method, (d) GO tab provides ontology of defined terms representing gene product properties, (e) Reference tab: It contains the article information in details along, (f) MolProbity tab: it contains the Ramachandran protein structure analysis, (g) Ag-Ab intPlot for the interaction between Antigen containing protein chain and Antibody chain, (h) EpiViewer contains the Jmol molecular visualization tool for Visual analysis of epitope sequence and antigen-antibody complexes, and (i) E-links tab: It contains the quick links to the other related database.

E-Links tab

E-Links tab contains the list of epitope database. User can click on to the database link to fetch the desired information from the selected database.

EpiBLAST

We have also developed EpiBLAST. It is a stand-alone BLAST web interface tool for the amino acid sequence similarity searches. It extends the utility of BLAST to query against multiple sequence databases. It provides a user friendly output for easily parse and navigation of the BLAST results (Figure 4). User can enter their amino acid sequence in FASTA format and select the database for the similarity search against the query sequences. Currently, user can select IEP (Immune Epitope Prediction tools and database, ongoing project), Immunoglobulin sequence database from NCBI (igNCBI), PDB, and SwissProt database. Cross-reference has been given (IEP, PDB, NCBI, PIR, UniProtKB, and EMBL) in the result page of the database to make it more informative. User can use this tool to identify the epitope sequence in the selected database and can retrieve the related information, which might help them for further analysis or research.

Conclusions

We presented SEDB an online web-based structural epitope database for the analysis of 3D epitopes. Our database coupled with a web interface that, by intelligently organized information from different biological sources. It provides detailed information of epitopes and their interactions with corresponding proteins. For user convenience, these epitopes have been classified based on structure (Linear epitope and Discontinuous epitope) and cell types (B-cell, T-cell and MHC binding). It will allow researchers to select relevant datasets of epitopes (Molecular, 3D structure, literature, Gene Ontology, 3D viewer and antigen-antibody binding etc.). We hope that our effort would also facilitate epitope based vaccine designing research.

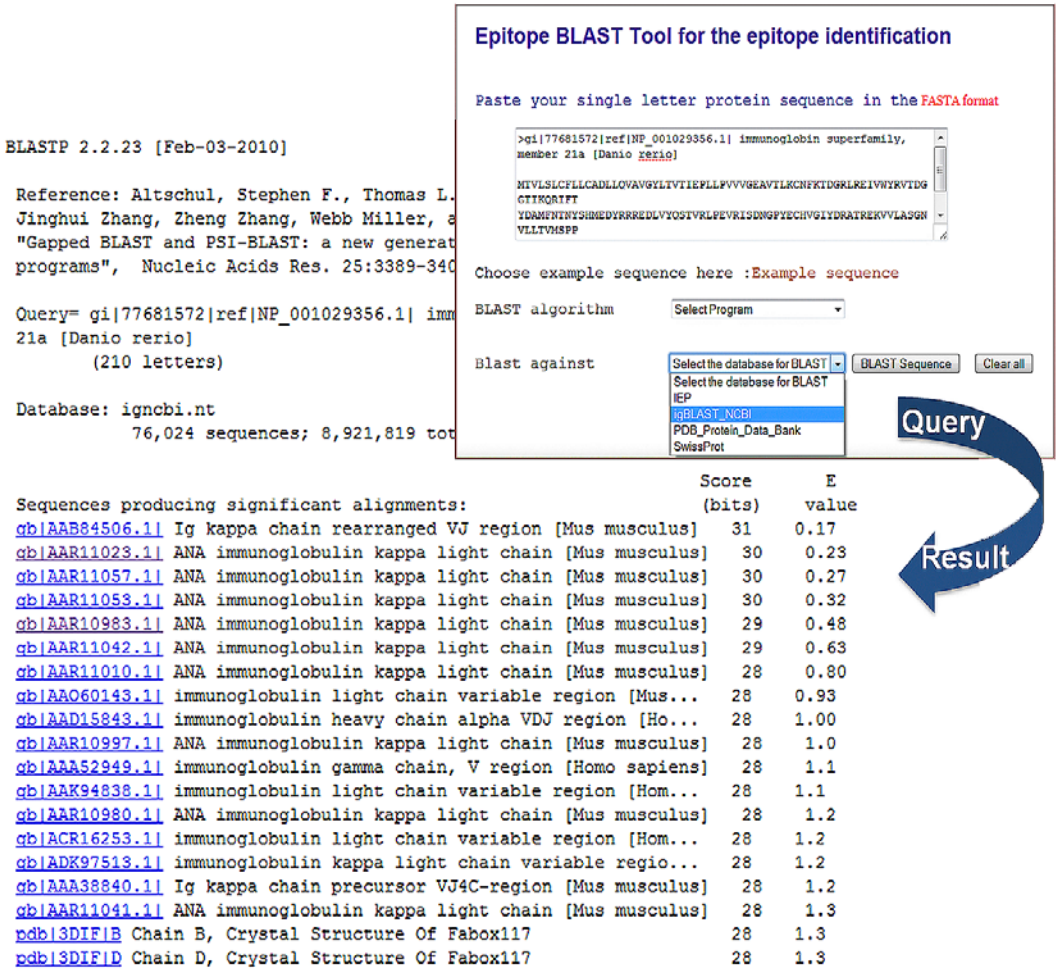

Figure 4: Snapshot of the EpiBLAST tool. Input page and result page for the epitope sequence BLAST against the igNCBI (immunoglobulin sequence from NCBI) database has been shown in the figure.

Acknowledgements

The authors thank Mr. Ankush Jadhav for the generation of DIMPLOT data and Mr. V. Saravanan for his suggestion to improve the EpiBLAST program. Authors acknowledge receipt of financial support from the Department of Biotechnology (DBT) (No. BT/BI/015/2002) and Dept of Information Technology (DIT) (No. DIT/R&D/Bio/15(9)/2007), Government of India, New Delhi, India.

References

1. Xu H, Hu C, Gong R, Chen Y, Ren N, et al. (2011) Evaluation of a Novel Chimeric B Cell Epitope-Based Vaccine against Mastitis Induced by Either Streptococcus agalactiae or Staphylococcus aureus in Mice. Clin Vaccine Immunol 18: 893-900.
2. Ben-Yedidia T, Arnon R (2007) Epitope-based vaccine against influenza. Expert Rev Vaccines 6: 939-948.
3. Adar Y, Singer Y, Levi R, Tzehoval E, Perk S, et al. (2009) A universal epitope-based influenza vaccine and its efficacy against H5N1. Vaccine 27: 2099-2107.
4. Tissot AC, Renhofa R, Schmitz N, Cielens I, Meijerink E, et al. (2010) Versatile virus-like particle carrier for epitope based vaccines. PLoS One 5: e9809.
5. El Kasmi KC, Muller CP (2001) New strategies for closing the gap of measles susceptibility in infants: towards vaccines compatible with current vaccination schedules. Vaccine 19: 2238-2244.
6. Senger T, Becker MR, Schadlich L, Waterboer T, Gissmann L (2009) Identification of B-cell epitopes on virus-like particles of cutaneous alpha-human papillomaviruses. J Virol 83: 12692-12701.
7. De Groot AS, Ardito M, McClaine EM, Moise L, Martin WD (2009) Immunoinformatic comparison of T-cell epitopes contained in novel swine-origin influenza A (H1N1) virus with epitopes in 2008-2009 conventional influenza vaccine. Vaccine 27: 5740-5747.
8. Li S, Schmitz KR, Jeffrey PD, Wiltzius JJ, Kussie P, et al. (2005) Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. Cancer Cell 7: 301-311.
9. Cho HS, Mason K, Ramyar KX, Stanley AM, Gabelli SB, et al. (2003) Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature 421: 756-760.
10. Brusic V, Rudy G, Harrison LC (1998) MHCPEP, a database of MHC-binding peptides: update 1997. Nucleic Acids Res 26: 368-371.
11. Saha S, Bhasin M, Raghava GP (2005) Bcipep: a database of B-cell epitopes. BMC Genomics 6: 79.
12. Ponomarenko J, Papangelopoulos N, Zajonc DM, Peters B, Sette A, et al. (2011) IEDB-3D: structural data within the immune epitope database. Nucleic Acids Res 39: D1164-1170.
13. Kaas Q, Ruiz M, Lefranc MP (2004) IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. Nucleic Acids Res 32: D208-210.

14. Siwicki B (1997) Three tier architecture. What role will the most advanced form of client/server play in health care? Health Data Manag 5: 65-68, 70-63.

15. Jamison DC (2003) Structured Query Language (SQL) fundamentals. Curr Protoc Bioinformatics Chapter 9: Unit9 2.

16. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucl Acids Res 28: 235-242.

17. Laskowski RA (2001) PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res 29: 221-222.

18. Bhasin M, Singh H, Raghava GP (2003) MHCBN: a comprehensive database of MHC binding and non-binding peptides. Bioinformatics 19: 665-666.

19. Ai Z, Wei Y (1993) Fast algorithm for exact rendering of space-filling molecular models with shadows. J Mol Graph 11: 200-203.

20. Kolaskar AS, Sawant S (1996) Prediction of conformational states of amino acids using a Ramachandran plot. Int J Pept Protein Res 47: 110-116.

21. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8: 127-134.