D1164–D1170 *Nucleic Acids Research, 2011, Vol. 39, Database issue* *Published online 28 October 2010*
doi:10.1093/nar/gkq888

IEDB-3D: structural data within the immune epitope database

Julia Ponomarenko[1,2,*], Nikitas Papangelopoulos[1], Dirk M. Zajonc[3], Bjoern Peters[3], Alessandro Sette[3] and Philip E. Bourne[1,2]

[1]San Diego Supercomputer Center, [2]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, CA 92093 and [3]La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037, USA

Received August 14, 2010; Revised September 16, 2010; Accepted September 19, 2010

ABSTRACT

IEDB-3D is the 3D structural component of the Immune Epitope Database (IEDB) available via the 'Browse by 3D Structure' page at http://www.iedb.org. IEDB-3D catalogs B- and T-cell epitopes and Major Histocompatibility Complex (MHC) ligands for which 3D structures of complexes with antibodies, T-cell receptors or MHC molecules are available in the Protein Data Bank (PDB). Journal articles that are primary citations of PDB structures and that define immune epitopes are curated within IEDB as any other reference along with accompanying functional assays and immunologically relevant information. For each curated structure, IEDB-3D provides calculated data on intermolecular contacts and interface areas and includes an application, EpitopeViewer, to visualize the structures. 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. These include queries for epitopes recognized in different pathogens, eliciting different functional immune responses, and recognized by different components of the immune system. The query results can be downloaded in Microsoft Excel format, or the entire database, together with structural data both curated and calculated, can be downloaded in either XML or MySQL formats.

INTRODUCTION

The Immune Epitope Database (IEDB) (1) catalogs experimentally identified B- and T-cell epitopes and MHC ligands through manual curation of the scientific literature. By the end of 2011, all published epitopes from infectious agents [except HIV, which are maintained in (2)], allergens and autoimmune diseases should be included in the IEDB. In addition, all epitopes characterized by 3D structures of immune receptors complexed with antigens found in the Protein Data Bank (PDB) (3) will be included, independent of the disease association of the antigen. This expansion in scope is made as to understand the general structural principles of epitope recognition a large dataset is desirable, but detailed 3D structural information on epitope complexes is rare.

The 3D structural component of IEDB, called here IEDB-3D, is embedded seamlessly within the general IEDB, thus ensuring that each reference describing 3D structure is curated within IEDB as any other journal article describing immune epitopes. The focus on curation of immunological and epitope relevant information is the major difference between IEDB and IMGT/3Dstructure-DB (4). Similar to IMGT/3Dstructure-DB (4) and other related databases, such as MPID-T (5), Epitome (6), BEID (7) and CED (8), IEDB provides calculated data on intermolecular contacts and interface areas and includes an application to visualize the structure, EpitopeViewer (9), which is a high-quality graphic and rendering tool. Among the aforementioned databases, only CED and IEDB curate epitope residues from the literature. In IEDB, antibody, MHC and T-cell receptor (TCR) residues interacting with the epitope are also curated if they are provided in the reference. Thus, within IEDB, curated and derived data on epitope and antigen–receptor interactions can be seen and compared side-by-side on the IEDB web page and also through launching EpitopeViewer.

IEDB-3D OVERVIEW

IEDB-3D is fully integrated within IEDB, with its data structure, ontology and query interface indistinguishable from those of the IEDB database. IEDB-3D is the 3D structural component of the Immune Epitope Database (IEDB) available via the 'Browse by 3D Structure' page

*To whom correspondence should be addressed. Tel: +1 858 822 5415; Fax: +1 858 822 0873; Email: jpon@sdsc.edu

© The Author(s) 2010. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

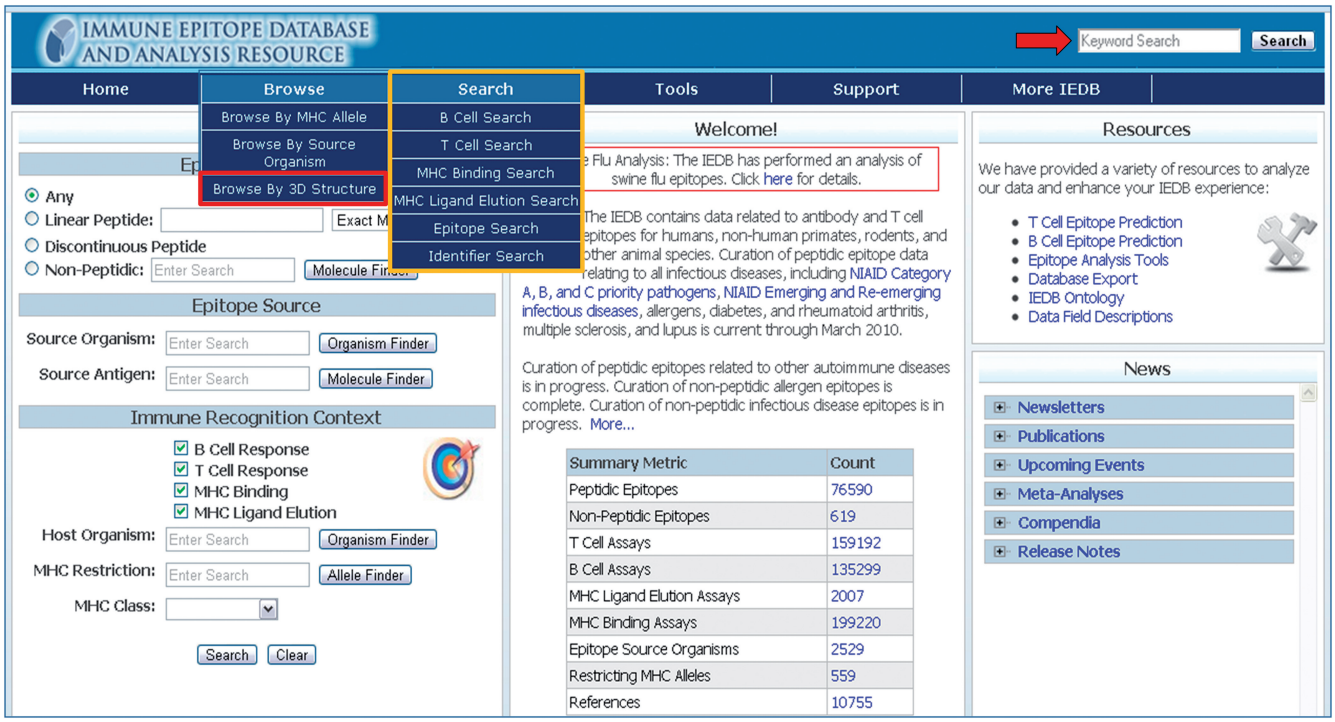


Figure 1. The screen shot of IEDB web page.

at http://www.iedb.org (Figure 1). Similar to IEDB, the types of curated data include experiments describing recognition of epitopes or antigens (peptidic or non-peptidic) by TCRs (T-cell assays describing 3D structures of antigens/epitopes in complexes with MHC and TCR), immunoglobulins or antibodies (B-cell assays describing structures of antibody–antigen complexes), and MHC molecules (MHC binding assays describing structures of antigens/epitopes in complex with MHC) (Figure 2).

Curation of 3D complex data is handled like the remaining data in the IEDB, except that all epitopes are considered in scope, including those from HIV or cancer. IEDB does not handle HIV data, if the paper provides only information on immunological assay involving HIV antigen explicitly, since these data are curated in HIV database (2). However, if 3D structural data are available for HIV epitopes, they are curated, but with a lower priority compared to other 3D structural data. In IEDB-3D, in addition to the immunological context curated for all epitopes in IEDB, information on 3D structure is provided in the section '3D Structure of Complex' (Figure 3).

An overview of IEDB-3D content is given in Table 1. Current statistics on the number of distinct epitope structures curated by assay type and source organism can be accessed from the 'Browse by 3D Structure' page (Figure 1, red box).

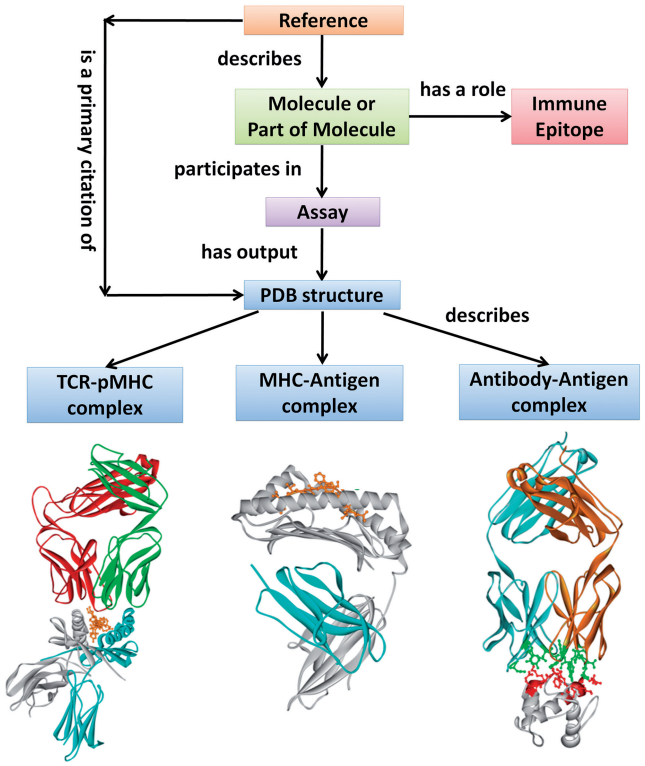


Figure 2. Objects, processes and roles as represented in IEDB-3D.

CURATION OF 3D STRUCTURAL DATA

The process of curating epitopes from 3D structures is shown in Figure 2. To be curated, the paper describing the structure (primary citation in PDB) should be published; likewise the paper describing the structure cannot be curated if the structure has not been deposited in PDB, unless this article describes the epitope in the

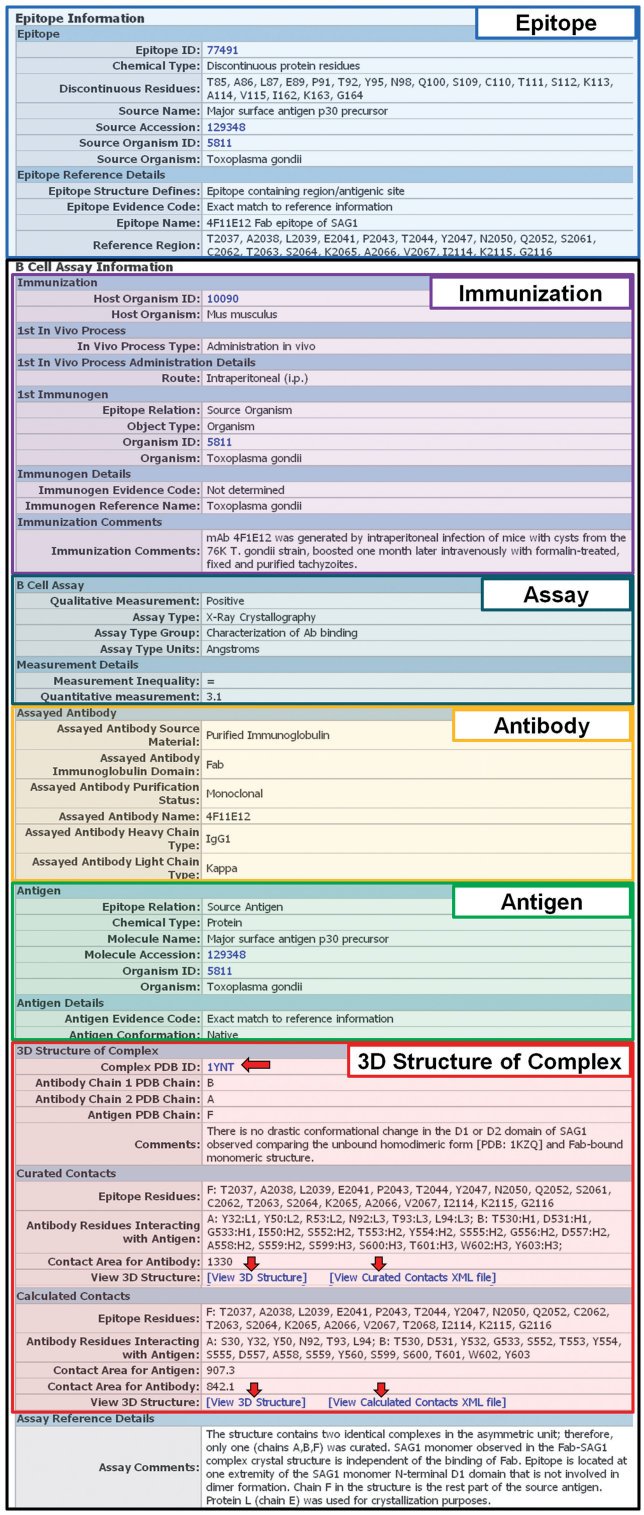


Figure 3. Example of the B-cell response page. Detailed information is provided for the epitope (blue box) and B-cell response, including data on the immunization (purple box), assay (teal box), antibody (yellow box), antigen used in the X-ray crystallography experiment (green box) and the 3D structure of antigen-antibody complex (red box). The antigen-antibody interactions were curated from (12) and calculated from the coordinates provided in the PDB file [PDB: 1YNT]. The results of clicking on the links 'View 3D Structure' and 'View Contacts XML file' (red arrows) are shown in Figure 4.

Table 1. IEDB-3D content

	Epitope type	Total number of references[a]	Curated references[b]
IEDB category			
Infectious diseases	Peptidic	131	131
	Non-peptidic	11	4
Allergy	Peptidic	8	8
	Non-peptidic	5	3
Autoimmunity	Peptidic	39	30
	Non-peptidic	11	11
Cancer	Peptidic	40	16
	Non-peptidic	8	5
Transplant	Peptidic	22	18
	Non-peptidic	2	1
Other	Peptidic	241	130
	Non-peptidic	82	2
HIV		55	4
Total		655	365
Structural category			
B-cell epitopes	Peptidic	311	168
	Non-peptidic	108	20
T-cell epitopes	Peptidic	217	177
	Non-peptidic	19	0
Total		655	365

[a]PDB and PubMed search as of January 7, 2010.
[b]Curation status as of January 9, 2010.

context of immunological assays. Upon curation, both a PDB structure and journal article, which is a primary reference of the structure are considered as sources of information for curation.

Similar to other assays, for 3D structures, data on epitope molecular structure, epitope source, immunization that led to recognition of the epitope, assay, antigen, antibody, TCR or MHC are curated according to IEDB general procedure (1). An example of a B-cell response is given in Figure 3. The difference of IEDB-3D data compared to regular IEDB curations is the information captured under the '3D Structure of Complex' heading.

In general, studies often do not provide the exact epitope structure recognized by an antibody/B-cell receptor, especially in the case of discontinuous B-cell epitopes. Therefore, discontinuous epitopes are often curated as partial epitopes, with only a few key residues captured. In cases where the epitope is curated from the structure this can be avoided; if the authors of the paper stated that only residues for part of the epitope were provided, we calculate the epitope residues from the PDB structure as the antigen residues separated from the antibody by 4 Å atomic distance (T-cell epitopes are curated as whole peptides or non-peptidic structures). The mapping between the numbers of the epitope residues provided in the paper and the external database is done manually, using the mapping between the PDB and UniProt numberings provided on the PDB website. For example, the epitope shown in Figure 3 had different residue numbers in the paper ('Reference Region' field at the top in Figure 3) compared to the numbering that matches the source

antigen sequence stored in the external database ('Discontinuous Residues' field).

Information on the 3D structure is provided in the section '3D Structure of Complex'. There are three types of tables with different fields for B-cell, T-cell and MHC-binding, respectively. The fields curated for B-cell responses, or structures of antibody–antigen complexes, are shown in Figure 3. The following structural information is curated: PDB ID, antibody and antigen chain IDs, epitope and antibody residues in PDB numbering, as well as CDR loops for the antibody residues and contact surface areas for the antibody and antigen if they are provided in the paper.

Interacting residues and contact areas are also calculated. The former are defined based on 4-Å atomic distance, according to the definition from (10), the latter are calculated using the NACCESS program (11). Pairwise atomic contacts are provided in the XML files available via hyperlinks (Figure 3) 'View Curated Contacts XML file' and 'View Calculated Contacts XML file'. The first link is provided only if the contacts were specified in the journal article and curated, the latter link allows the user to view and download the file providing calculated antibody–antigen (antigen-MHC, or antigen-TCR and MHC-TCR) inter-molecular contacts (Figure 4C). The following types of contacts are calculated: hydrogen bonds, salt bridges, van der Waals, hydrophobic and 4-Å interactions (interactions are defined in the EpitopeViewer tutorial at http://spdc.sdsc.edu/iedb/epitopeViewer/EpitopeViewerTutorial_v2.0.htm; this link is also available inside the EpitopeViewer application). The calculation is done during curation, using a php program that takes as an input curated data and PDB file and outputs two XML files with the interacting residues, contact areas and pairwise contacts. These files are also used as input for EpitopeViewer (Figure 4). The curators use the php tool and EpitopeViewer as part of the curation process and to check for errors.

EpitopeViewer allows the user to visualize, render and analyze the structure and save structural and contact views as high-quality pictures for publication (9). Figur 4 shows an example of how the EpitopeViewer can be used to analyze specific inter-molecular contacts. The residue Arg53 of the antibody light chain (Figure 3) was curated as part of the paratope; however, it can be seen that it is located relatively far from the nearest epitope atom (5.4 Å) (Figure 4A). At the same time, another light chain residue, Ser30, which was not curated as part of the paratope, contacts the epitope residues through van der Waals and 4-Å interactions (Figure 4B).

Additional detailed information on the 3D structure is captured in free text form in the 'Assay Comments' field (bottom of Figure 3). This can, for example, provide information on antigen and receptor conformational changes and comparison with other relevant structures if this information is mentioned in the paper. Also, the user can get further details on the PDB website via the PDB ID hyperlink in the 'Complex PDB ID' field.

QUERYING IEDB-3D

Since IEDB-3D is fully integrated within IEDB, structural data, both curated and calculated, and all accompanying information can be queried using the multiple search capabilities implemented as part of the IEDB web site and described in this section.

On the IEDB home page (Figure 1), epitopes can be searched by epitope sequence, epitope source organism, source antigen and the immune recognition context, including the type of response (B-cell, T-cell or MHC-binding), host organism and MHC allele. If epitopes curated from 3D structures satisfy the search criteria, they will be retrieved together with all other epitopes curated in IEDB.

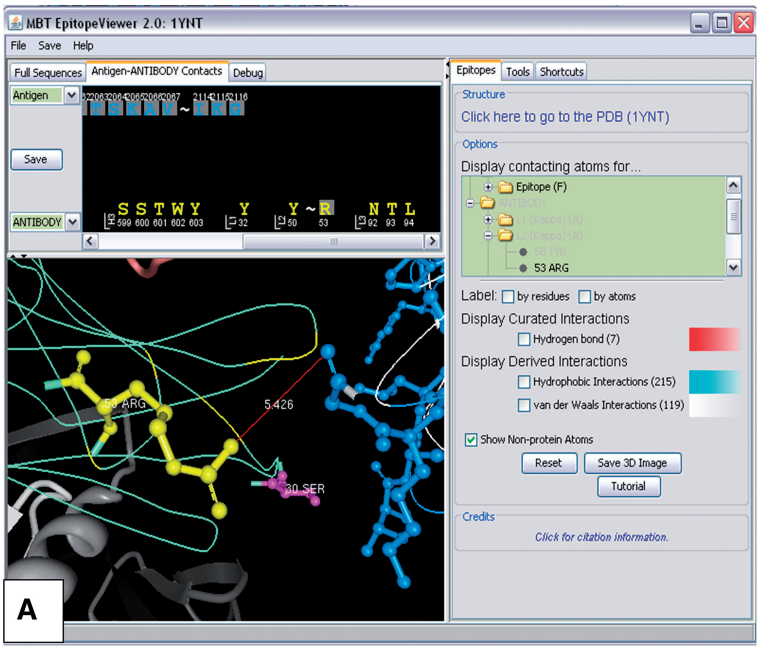
Using the keyword search option on the IEDB hom page (the box is located in the right top corner of the page, Figure 1, red arrow), data are retrieved by keywords, PDB ID or SwissProt/GenBank ID; the keyword search runs throughout all fields in the database. In addition, the user can explicitly qualify what identifiers to search for when using the option 'Search by Identifier' from the 'Search' dropdown menu on the home page (Figure 1, yellow box). This type of search can be done, using IEDB internal identifiers, PDB ID, PubMed ID and CHEBI ID.

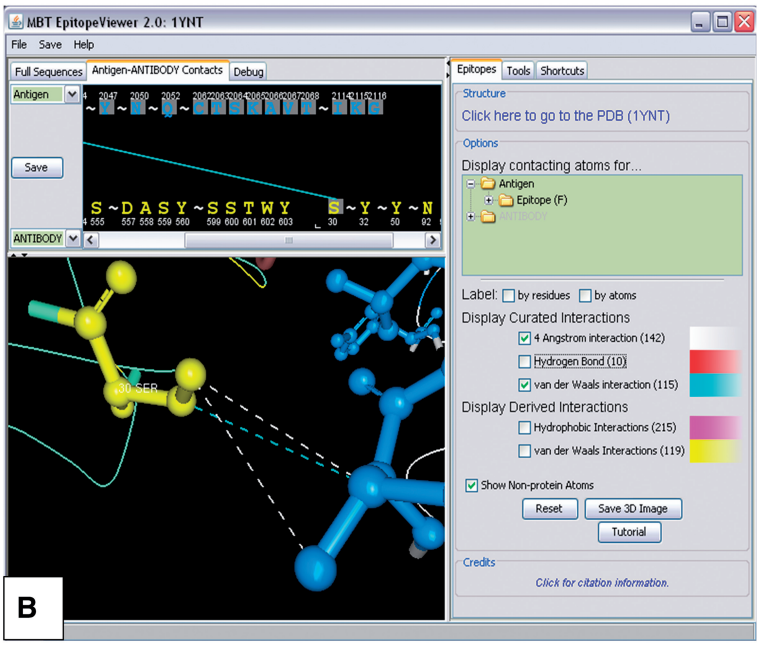
The 'Search' dropdown menu on the home page allows advanced search of any specified field in the database, including the fields related to 3D structure. For example, on the 'B Cell Search' webpage, the search fields related to 3D structure are made visible by clicking the '+' sign next to '3D Structure of Complex' within the 'B Cell Assay' subsection. Figure 5A shows how to search on the 'B Cell Search' page for the structures of antigens in complex with antibodies obtained via *in vivo* administration/immunization with *Toxoplasma gondii* and containing tryptophan (W) in the paratope (residues in the antibody interacting with the antigen). The result of this query is a single epitope that was curated from the structure with the PDB ID 1YNT (Figure 3).

Alternatively, epitopes for which 3D structural information is available can also be searched for by specifying a particular type of assay. For example, the epitope shown in Figure 3 can be found by specifying the assay type 'X-ray crystallography' using the 'Assay Finder' on the 'B Cell Search' form. Additionally, the resolution for the 3D structure can be specified.

In addition to the keyword, simple and advanced searches, epitopes curated from 3D structures can be accessed via the 'Browse by 3D Structure' page provided in the 'Browse' dropdown menu on the IEDB home page (Figure 1, red box). By expanding the 'B Cell', 'T Cell' and 'MHC Binding' trees the epitopes can be browsed by the organism that is the source of the antibody, T cell, and MHC molecule, respectively. This page also provides an up to date overview on the number of distinct epitope structures curated in IEDB-3D.

When clicking on an epitope ID (Figure 5C), all assays in which that epitope is characterized are returned.





C

```
- <dataroot generated="Mar 25, 2010">
  - <contacts_xxxx>
      <AtomPairID>1</AtomPairID>
      <BindingID>1</BindingID>
      <PDB_ID>1YNT</PDB_ID>
      <Antigen_Chain>F</Antigen_Chain>
      <Antigen_Residue>E (GLU)</Antigen_Residue>
      <Antigen_Residue_Position>2041</Antigen_Residue_Position>
      <Antigen_Residue_Atom>OE1</Antigen_Residue_Atom>
      <Immunoglobulin_Chain>B</Immunoglobulin_Chain>
      <Immunoglobulin_Residue>W (TRP)</Immunoglobulin_Residue>
      <Immunoglobulin_Residue_Position>602</Immunoglobulin_Residue_Position>
      <Immunoglobulin_Residue_Atom>NE1</Immunoglobulin_Residue_Atom>
      <Contact_Type>Hydrogen Bond</Contact_Type>
    </contacts_xxxx>
      ..........
  - <contacts_xxxx>
      <AtomPairID>267</AtomPairID>
      <BindingID>1</BindingID>
      <PDB_ID>1YNT</PDB_ID>
      <Antigen_Chain>F</Antigen_Chain>
      <Antigen_Residue>T (THR)</Antigen_Residue>
      <Antigen_Residue_Position>2063</Antigen_Residue_Position>
      <Antigen_Residue_Atom>CB</Antigen_Residue_Atom>
      <Immunoglobulin_Chain>B</Immunoglobulin_Chain>
      <Immunoglobulin_Residue>S (SER)</Immunoglobulin_Residue>
      <Immunoglobulin_Residue_Position>599</Immunoglobulin_Residue_Position>
      <Immunoglobulin_Residue_Atom>O</Immunoglobulin_Residue_Atom>
      <Contact_Type>4 Angstrom interaction</Contact_Type>
    </contacts_xxxx>
  </dataroot>
```

Figure 4. Visualization of 3D structures and contacts. EpitopeViewer screen-shots for the curated (**A**) and calculated (**B**) data for the epitope shown in Figure 3. Epitope residues are in blue. (A) The residue Arg53 of the antibody light chain, which was curated as part of the paratope, is shown in an all-atom presentation and colored in yellow; the red line shows the distance between this residue and the nearest epitope atom. Another light chain residue Ser30, which was not curated as part of the paratope, is shown in magenta. This residue is highlighted in yellow in (B) and its calculated contacts with the epitope are shown as broken lines (cyan, van der Waals interactions; white, 4 Å interactions). (**C**) The fragment of the XML file providing calculated contacts between the antibody and antigen [PDB: 1YNT].

For example, the query shown in Figure 5A returns one epitope (Figure 5B and C), for which two B-cell response assays were curated (Figure 5D), one of which describes the 3D structure (Figure 3) and another, the immunological assay (not shown).

The query results obtained using either simple or advanced search options can be exported as a Microsoft Excel formatted file and downloaded in either full (all fields are present except pairwise contacts) or compact form (Figure 5C and D).

DATA ACCESS

The entire database, together with structural data both curated and calculated (including pairwise contacts), can be downloaded in XML and MySQL formats.

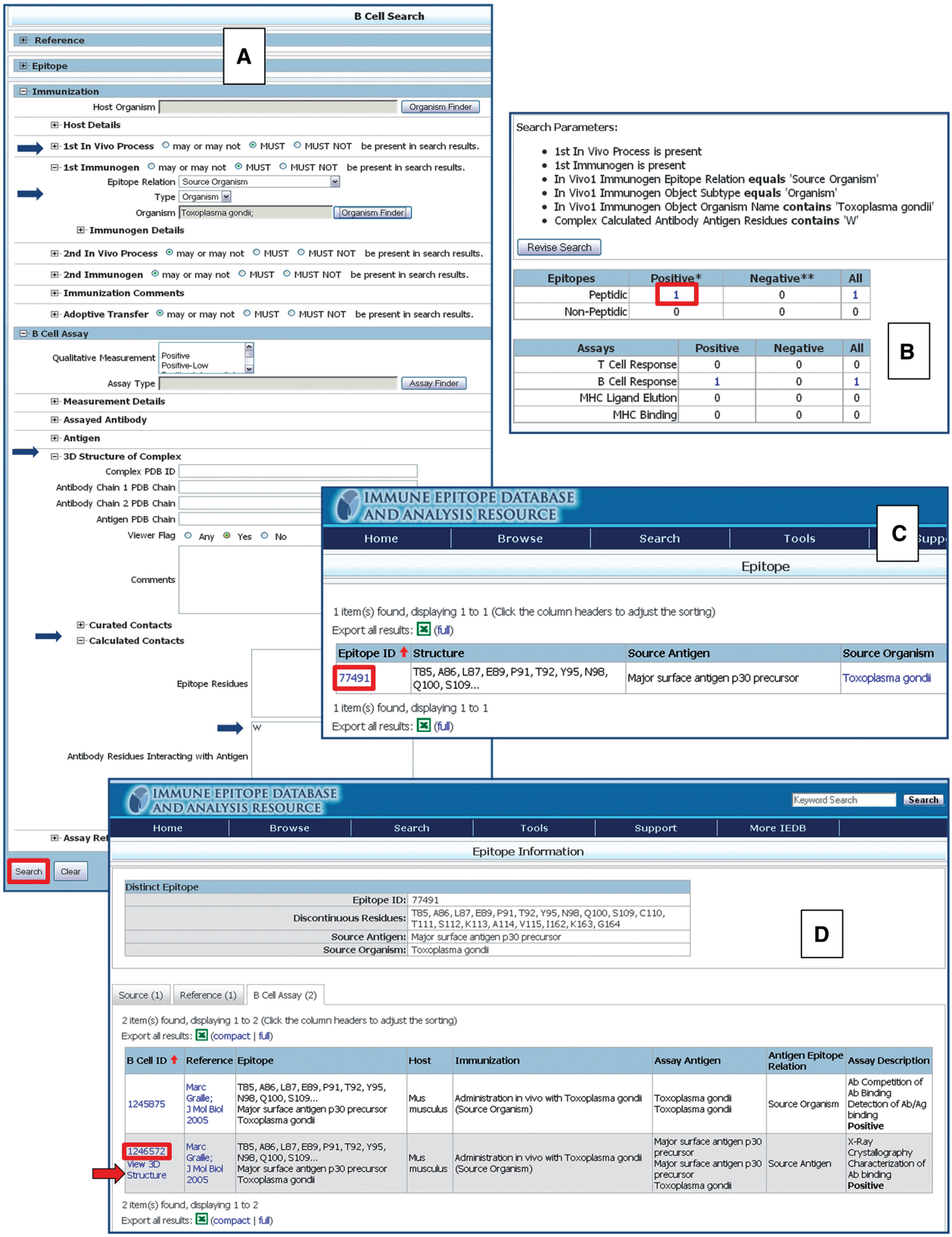


Figure 5. Example of the IEDB-3D query. (**A**) The advanced B-cell search page showing the query for 3D structures of antigens in complex with antibodies obtained via *in vivo* administration/immunization with *Toxoplasma gondii* and containing tryptophan (W) in the paratope calculated as the antibody residues interacting with the antigen at <4Å atomic distance. The result of clicking on the button 'Search' (red box) is shown in B. (**B**) The result of the search shown in A. The result of clicking on the number of epitopes (red box) is shown in C. (**C**) The list of epitopes as the result of the search shown in A. The result of clicking on the epitope ID (red box) is shown in (D). (**D**) The list of assays for the epitope found using the query shown in A. The button 'View 3D Structure' (red arrow) launches the EpitopeViewer for curated data (Figure 4A). The result of clicking on the assay ID (red box) is shown in Figure 3.

FUNDING

Funding for open access charge: National Institute of Health/National Institute of Allergy and Infectious Diseases (contract number HHSN2662004000 0 6C).

Conflict of interest statement. None declared.

REFERENCES

1. Vita,R., Zarebski,L., Greenbaum,J.A., Emami,H., Hoof,I., Salimi,N., Damle,R., Sette,A. and Peters,B. (2010) The immune epitope database 2.0. *Nucleic Acids Res.*, **38**, D854–D862.
2. Korber,B.T.M., Brander,C., Haynes,B.F., Koup,R., Moore,J.P., Walker,B.D. and Watkins,D.I. (2007) HIV Molecular Immunology. *Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico*, LA-UR 07-4752.
3. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
4. Ehrenmann,F., Kaas,Q. and Lefranc,M.P. (2010) IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res.*, **38**, D301–D307.
5. Tong,J.C., Kong,L., Tan,T.W. and Ranganathan,S. (2006) MPID-T: database for sequence-structure-function information on T-cell receptor/peptide/MHC interactions. *Appl. Bioinformatics*, **5**, 111–114.
6. Schlessinger,A., Ofran,Y., Yachdav,G. and Rost,B. (2006) Epitome: database of structure-inferred antigenic epitopes. *Nucleic Acids Res.*, **34**, D777–D780.
7. Tong,J.C., Song,C.M., Tan,P.T., Ren,E.C. and Sinha,A.A. (2008) BEID: Database for sequence-structure-function information on antigen-antibody interactions. *Bioinformation*, **3**, 58–60.
8. Huang,J. and Honda,W. (2006) CED: a conformational epitope database. *BMC Immunol.*, **7**, 7.
9. Beaver,J.E., Bourne,P.E. and Ponomarenko,J.V. (2007) EpitopeViewer: a Java application for the visualization and analysis of immune epitopes in the Immune Epitope Database and Analysis Resource (IEDB). *Immunome Res.*, **3**, 3.
10. Greenbaum,J.A., Andersen,P.H., Blythe,M., Bui,H.H., Cachau,R.E., Crowe,J., Davies,M., Kolaskar,A.S., Lund,O., Morrison,S. *et al.* (2007) Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J. Mol. Recognit.*, **20**, 75–82.
11. Hubbard,S.J. and Thornton,J.M. (1993) NACCESS. *Department of Biochemistry and Molecular Biology*, University College, London.
12. Graille,M., Stura,E.A., Bossus,M., Muller,B.H., Letourneur,O., Battail-Poirot,N., Sibai,G., Gauthier,M., Rolland,D., Le Du,M.H. *et al.* (2005) Crystal structure of the complex between the monomeric form of Toxoplasma gondii surface antigen 1 (SAG1) and a monoclonal antibody that mimics the human immune response. *J. Mol. Biol.*, **354**, 447–458.