Immune epitope database analysis resource

Kim, Yohan; Ponomarenko, Julia; Zhu, Zhanyang; Tamang, Dorjee; Wang, Peng; Greenbaum, Jason; Lundegaard, Claus; Sette, Alessandro; Lund, Ole; Bourne, Philip E.; Nielsen, Morten; Peters, Bjoern

Published in:
Nucleic Acids Research

Link to article, DOI:
10.1093/nar/gks438

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Published online 18 May 2012 *Nucleic Acids Research, 2012, Vol. 40, Web Server issue* ***W525–W530***
doi:10.1093/nar/gks438

Immune epitope database analysis resource

Yohan Kim[1], Julia Ponomarenko[2], Zhanyang Zhu[3], Dorjee Tamang[1], Peng Wang[4], Jason Greenbaum[1], Claus Lundegaard[5], Alessandro Sette[1], Ole Lund[5], Philip E. Bourne[2], Morten Nielsen[5] and Bjoern Peters[1,*]

[1]La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, CA 92037, [2]San Diego Supercomputer Center, MC 0505, 10100 Hopkins Drive, La Jolla, CA 92093, [3]Sequenom, 3595 John Hopkins Court, San Diego, CA 92121, USA, [4]Shanghai Advanced Research Institute, No.99 Haike Road, Zhangjiang Hi-Tech Park, Pudong Shanghai, 201210, China and [5]Technical University of Denmark, Building 208, 2800, Lyngby, Denmark

Received March 2, 2012; Revised April 19, 2012; Accepted April 25, 2012

ABSTRACT

The immune epitope database analysis resource (IEDB-AR: http://tools.iedb.org) is a collection of tools for prediction and analysis of molecular targets of T- and B-cell immune responses (i.e. epitopes). Since its last publication in the NAR webserver issue in 2008, a new generation of peptide:MHC binding and T-cell epitope predictive tools have been added. As validated by different labs and in the first international competition for predicting peptide:MHC-I binding, their predictive performances have improved considerably. In addition, a new B-cell epitope prediction tool was added, and the homology mapping tool was updated to enable mapping of discontinuous epitopes onto 3D structures. Furthermore, to serve a wider range of users, the number of ways in which IEDB-AR can be accessed has been expanded. Specifically, the predictive tools can be programmatically accessed using a web interface and can also be downloaded as software packages.

INTRODUCTION

The immune system orchestrates various classes of molecules to detect an abnormal state of the host, as a first step toward the goal of protecting the host from diseases. To achieve such detection, dedicated molecules such as major histocompatibility complexes (MHCs), T-cell receptors (TCRs) and B-cell receptors (BCRs) continuously scan cellular surfaces and extra-intra-cellular spaces for non-self molecules and trigger immune responses upon their detection. Parts of such non-self molecules (e.g. proteins, carbohydrates and other macromolecules) recognized by the immune system are called epitopes.

Depending on the context of immune epitope recognition, different kinds of immune responses are initiated. For instance, TCRs binding peptide:MHC-I (class I) complexes presented on the cell surface are associated with cytotoxic lymphocyte activity (i.e. cell killing); whereas, TCRs binding peptide:MHC-II (class II) complexes are associated with recruiting helper T-cells and alerting B-cells. Thus, improved molecular-level understanding of immune epitope recognition will further enhance the development of novel vaccines, diagnostics and therapeutics in treating infectious and autoimmune diseases, allergies and cancers.

To address this need, the immune epitope database-analysis resource (IEDB-AR) aims to be the premier resource in providing web-based tools for immune epitope predictions and analyses. Since its publication in the NAR webserver issue in 2008 (1), major updates and the addition of new predictive tools have been made. Specifically, six new peptide:MHC binding, one antigen processing and one B-cell epitope predictive tool have been added. In addition, binding affinity data for MHC-I and MHC-II molecules have been significantly expanded, resulting in both increased coverage of alleles and improved accuracy for existing predictors.

THE WEB RESOURCE

Tools provided on the IEDB analysis resource website can be grouped into those that make predictions or carry out analyses (Table 1). Predictive tools can be further grouped based on their targeted immune recognition contexts: (i) peptide:MHC binding, (ii) antigen processing and (iii) B-cell-receptor/antibody binding. Since the last

*To whom correspondence should be addressed. Tel: +1 858 752 6914; Fax: +1 858 752 6987; Email: bpeters@liai.org

Table 1. An overview of immune-epitope related bioinformatics tools provided by the IEDB-AR

Tools category			New or updated tools	Descriptions
Prediction	T-cell	MHC-I	Consensus, SMMPMBEC, NetMHCpan	Predict peptides that bind to MHC class I molecules
		MHC-II	Consensus, NN-align, NetMHC-IIpan	Predict peptides that bind to MHC class II molecules
		Antigen processing	NetCTLpan	Predict epitopes by integrating proteasomal cleavage, TAP efficiency, peptide:MHC-I binding scores
	B-cell		ElliPro	Predict protein regions that are most likely to be bound by antibodies
Epitope analysis			Homology Mapping	Map linear and discontinuous epitopes onto 3D structures

publication on the analysis resource, major updates and additions have been made to tools in all categories. The following sections describe these updates in detail.

All T-cell epitope predictive tools take protein sequences represented by single letter amino acids as input. Once submitted, protein sequences are broken into appropriate peptide lengths as specified by the user. Predictions are made against the set of MHC molecules chosen by the user. Outputs consist of a list of peptides and their predicted scores, indicating their likelihood of binding or being epitopes.

MHC class I peptide binding predictions

For peptide:MHC-I binding prediction, a number of improvements and additions have been made since the last web-server issue on the IEDB-AR in 2008. Specifically, larger datasets have been used to re-train all existing prediction methods, adding multiple new MHC molecules in the process. Currently binding affinity data and corresponding predictors are available for 56 human, 19 non-human primate and six mouse molecules. This represents an increase of 42% in the number of molecules covered.

Furthermore, three powerful new prediction methods have been added. The first, SMM^{PMBEC} (2) is an improvement of the SMM scoring matrix-based predictor (3). SMM^{PMBEC} implements a Bayesian approach that prefers scoring matrices that are consistent with known amino acid similarities. This is particularly helpful when estimating the contribution of specific residues to peptide binding for molecules characterized by limited binding data. The second newly added method, NetMHCpan (4), is a neural-network based predictor, similar to NetMHC (5), but with the crucial difference that for NetMHCpan, a single network-ensemble is trained on all MHC molecules simultaneously, incorporating both the peptide and contact residues from the MHC sequences. This allows the method to extrapolate and estimate binding predictions for any MHC molecule, including those not included in the training set (6), by leveraging known sequence:binding data affinity relationships and extending this to those MHC molecules with no binding data.

Consensus is the third newly added method, which has not been previously described. The method was motivated by observations made by investigators in the machine-learning community that combining predictions from different predictors may yield higher predictive performance than any of the individual predictors (7). In the current implementation, predictions from various predictors are first transformed into percentile scores, thereby allowing comparisons across predictors on a uniform scale. For a given peptide and a predictor, a percentile score is defined as a percentage of random peptides sampled from naturally occurring proteins that score better than the peptide. Using the consensus approach, the final predicted binding affinity score for the peptide is a median of percentile scores from the different predictors. Notable use of an early version of a consensus method for a large-scale prediction has been described in Moutaftsi *et al.* (8).

Benchmarks for the different tools have been performed in the respective publications with average predictive performances of 0.881 AUC for the class I (NetMHCpan-2.0 against HLA A and B molecules) (4). Of special note, an MHC class I prediction competition was held recently for the first time (9). Tested on blind peptide:MHC binding datasets generated by an independent group, the consensus method hosted at the IEDB-AR has consistently ranked high (i.e. within the top 5 entries out of 20 total) among competitors. The blind datasets were generated for three molecules, each involving 9- and 10-mers. The combined dataset consisted of ~1200 measurements with 20% composed of binders, and the average predictive performance on this dataset for the IEDB consensus prediction was 0.96 AUC, notably higher than our own prediction performance estimates, only surpassed by the NetMHCcons (10), NetMHC and NetMHCpan methods.

To date, top-performing methods have been entirely sequence based, despite potential advantages of structure-based methods. It is, however, expected that as structural modeling techniques improve, predictive methods based on 3D structures with comparable accuracy will emerge (11, 12).

MHC class II peptide-binding predictions

Similar to the class I tools, all class II tools have been re-trained using newly available binding data. Importantly, a large set of data have become available that covers a set of prevalent HLA-DP and DQ molecules, for which previously there were very little data available. This was not due to lack of importance of these molecules but rather due to the significantly greater experimental effort involved in characterizing them when compared to

the HLA DR molecules. With the new data available, for the first time, the prediction methods cover a large fraction of human MHC class II molecules (13). The molecules covered (Table 2) were selected for experimental characterization based on their high frequency in the worldwide human population.

Two new methods for class II binding prediction [NN-align (8) and NetMHCIIpan (9)] were added to the resource website. Both methods are artificial neural network based and are trained using a concurrent alignment and weight optimization neural network training procedure described in (14–16). NN-align is molecule-specific (i.e. one neural network method is trained for each MHC class II molecule), and NetMHCIIpan is HLA-DR pan-specific (i.e. one neural network method is trained covering all HLA-DR molecules). Both methods include encoding of peptide flanking regions (PFR), PFR length and the peptide length to boost the predictive performance. The pan specificity of NetMHCIIpan is achieved (as was the case for NetMHCpan) by including both the peptide and contact residues from the MHC sequences in the network training. Both novel methods have demonstrated superior predictive performance compared to the earlier methods included in the resource website with average AUC performance values of 0.821 (NN-align) and 0.846 (NetMHCIIpan) when benchmarked against a large set of HLA-DR molecules (15).

Table 2. HLA-DP, DQ, DR molecules chosen based on their high frequency in the human population. Allele frequency data are provided by dbMHC

Allelic variant	Allele frequency
HLA-DPA1*0201-DPB1*0101	16
HLA-DPA1*0103-DPB1*0201	17.5
HLA-DPA1*01-DPB1*0401	36.2
HLA-DPA1*0301-DPB1*0402	41.6
HLA-DPA1*0201-DPB1*0501	21.7
HLA-DQA1*0501-DQB1*0201	11.3
HLA-DQA1*0501-DQB1*0301	35.1
HLA-DQA1*0301-DQB1*0302	19
HLA-DQA1*0401-DQB1*0402	12.8
HLA-DQA1*0101-DQB1*0501	14.6
HLA-DQA1*0102-DQB1*0602	14.6
HLA-DRB1*0101	5.4
HLA-DRB1*0301	13.7
HLA-DRB1*0401	4.6
HLA-DRB1*0404	3.6
HLA-DRB1*0405	6.2
HLA-DRB1*0701	13.5
HLA-DRB1*0802	4.9
HLA-DRB1*0901	6.2
HLA-DRB1*1101	11.8
HLA-DRB1*1302	7.7
HLA-DRB1*1501	12.2
HLA-DRB3*0101	26.1
HLA-DRB4*0101	41.8
HLA-DRB5*0101	16

MHC class I antigen processing predictions

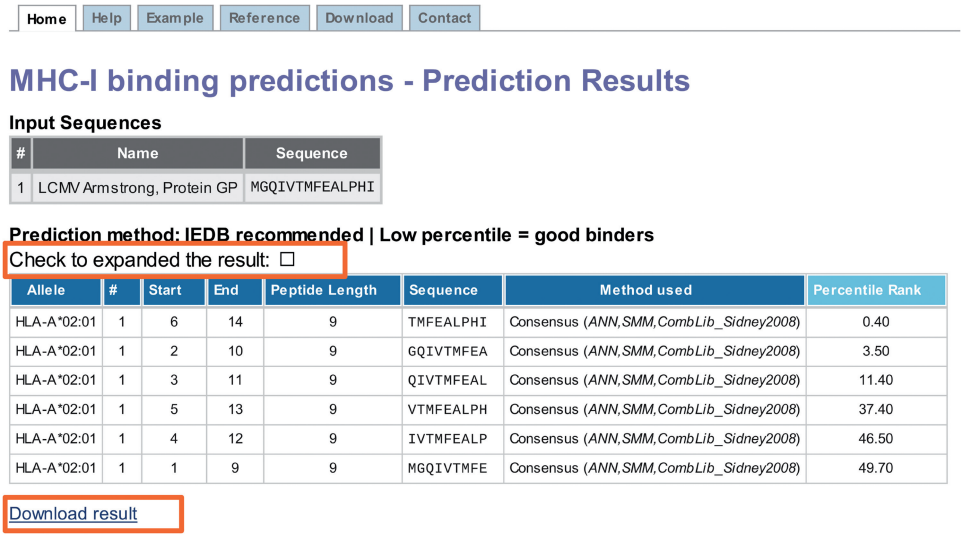
In addition to binding of peptides to MHC molecules, there are additional steps in the MHC class I pathway that a peptide has to pass in order to be recognized by the immune system (17). This includes proteasomal cleavage (18, 19) and TAP transport (20), which have been utilized in combination with MHC binding predictions to identify T-cell epitopes (21). A new such integrative predictive approach has recently been developed called NetCTLpan (22), distinguished by the use of the NetMHCpan method for the peptide:MHC binding step. It has been demonstrated that for cases where a low false positive rate is desired, (high specificity predictions) proteasome cleavage and transport efficiency by TAP contribute to improved predictive performances (22).

User interface updates for MHC class I/II peptide binding tools

The web interfaces to the MHC class I and II binding predictive tools have been updated based on user feedback. One example was that the selection of MHC molecules from a drop-down list used to make predictions had become cumbersome, especially with the addition of the NetMHCpan tools. To address this, there is now a checkbox that is selected by default that limits the MHC molecules included for selection to those that occur in at least 1% of the human population. The vast majority of users are focusing on such molecules, and the smaller list is much easier to navigate, whereas the entirety of alleles is still available by simply unselecting the checkbox. This feature is currently available only for MHC class I tools.

In addition, the user can select different combinations of MHC molecules and lengths, so that a different group of predictions can be run in one iteration rather than repeatedly selecting prediction method and retrieving results. Also with the addition of NetMHCpan, one can upload any MHC molecule of interest to allow predictions for MHC molecules outside of those provided by the IEDB-AR. Once a table of predictions is generated, it can be 'expanded' to show greater detail of how individual components of the scores contribute to the final scores (Figure 1).

Finally, another user request was for the IEDB team to spell out, which prediction method is recommended for a given task. Therefore, a default choice is provided, named 'IEDB Recommended'. Based on availability of predictors and previously observed predictive performance, this selection tries to use the best possible method for a given MHC molecule. Currently, for peptide:MHC-I binding prediction, for a given MHC molecule, 'IEDB Recommended' uses the consensus method consisting of NetMHC, SMM and CombLib if a trained predictor is available for the molecule. Otherwise, NetMHCpan is used. This choice was motivated by the expected predictive performance of the methods in decreasing order: Consensus > NetMHC > SMM > NetMHCpan > CombLib. For peptide:MHC-II binding prediction, 'IEDB Recommended' again uses the 'consensus' approach, combining NN-align, SMM-align and CombLib. The expected predicted performance for MHC-II binding methods in decreasing order are Consensus > NetMHCIIpan > NN-align > SMM-align > CombLib. Of note, we fully expect the IEDB recommendation to



Home | Help | Example | Reference | Download | Contact

MHC-I binding predictions - Prediction Results

Input Sequences

#	Name	Sequence
1	LCMV Armstrong, Protein GP	MGQIVTMFEALPHI

Prediction method: IEDB recommended | Low percentile = good binders

Check to expanded the result: ☐

Allele	#	Start	End	Peptide Length	Sequence	Method used	Percentile Rank
HLA-A*02:01	1	6	14	9	TMFEALPHI	Consensus (*ANN,SMM,CombLib_Sidney2008*)	0.40
HLA-A*02:01	1	2	10	9	GQIVTMFEA	Consensus (*ANN,SMM,CombLib_Sidney2008*)	3.50
HLA-A*02:01	1	3	11	9	QIVTMFEAL	Consensus (*ANN,SMM,CombLib_Sidney2008*)	11.40
HLA-A*02:01	1	5	13	9	VTMFEALPH	Consensus (*ANN,SMM,CombLib_Sidney2008*)	37.40
HLA-A*02:01	1	4	12	9	IVTMFEALP	Consensus (*ANN,SMM,CombLib_Sidney2008*)	46.50
HLA-A*02:01	1	1	9	9	MGQIVTMFE	Consensus (*ANN,SMM,CombLib_Sidney2008*)	49.70

Download result

Figure 1. Screenshot of the peptide:MHC-I binding predictive tool results page generated using the 'IEDB recommended' option. The first highlighted area at the top indicates a checkbox with which the user can expand the table to display method-specific predictions. The second highlighted area at the bottom allows the user to download the prediction results as a text file.

change as we perform larger benchmarks of newly developed methods on blind datasets to determine an accurate assessment of prediction quality. For example, recent evaluations suggested that NetMHCpan is actually superior in performance to all allele-specific predictors unless there is a very large amount of data available for the particular allele (10). If this result can be confirmed on new binding datasets, the IEDB recommendation will change.

B-cell epitope prediction

A new addition to IEDB-AR since its last publication is the ElliPro tool (23). ElliPro predicts linear and discontinuous epitopes for a protein structure or sequence provided by the user; for sequences, the structure is modeled using MODELLER (24). ElliPro is based on the geometrical properties of protein structure and does not require training. Tested on a benchmark dataset of discontinuous epitopes inferred from 3D structures of antibody–protein complexes (25), ElliPro has an AUC of 0.73 when the most significant prediction was considered for each protein. Since the rank of the best prediction was at most in the top three for >70% of proteins and never exceeded five, ElliPro can be considered a useful research tool for identifying antibody epitopes in protein antigens. Details on the comparison of ElliPro with other structure-based epitope prediction methods can be found in (23). It would also be interesting to compare the method against sequence-based approaches as well (26).

Epitope analysis: Homology mapping tool

The homology mapping tool enables analysis of an epitope's location in the 3D structure of its source protein. For a given epitope, linear or discontinuous, from IEDB or submitted by a user, the tool searches for known 3D protein structures in the Protein Data Bank (PDB) (27) that are homologous to the epitope source sequence. The output page (Figure 2) provides mapping of the epitope to the source sequence, the PDB hits that contain the epitope regions, secondary structures and solvent accessibilities for each residue, presented in the format of either pairwise or multiple sequence alignment for the selected PDB hits, obtained using ClustalW2 (28,29). Residues in the alignment are colored by relative solvent accessibility; coloring can be modified according to the user-specified cutoffs on relative solvent accessibility of all atoms or side chain atoms only. PDB structures with mapped epitopes can be visualized using EpitopeViewer (30). The 3D viewer feature uses java webstart technology, which does not appear to be smoothly integrated into Google chrome browsers. To use the feature, we recommend using Firefox.

In addition to maintaining the homology mapping tool, IEDB-AR is open to incorporating tools developed by external groups. One such type of tools that is of much interest would address the problem of choosing an optimal set of epitopes given various constraints for vaccine design (31,32).

IEDB application programming interface (IEDB-API) for peptide:MHC binding predictive tools

A frequent user request has been to enable integration of IEDB-AR tools into external applications. For example, the Los Alamos HIV Immunology Database (33) wanted to utilize the IEDB-AR MHC binding prediction for their Epitope Location Finder tool (ELF: www.hiv.lanl.gov/content/sequence/ELF/epitope_analyzer.html). To address this general need, an application programming interface (IEDB-API) for tools predictive of MHC binding has been implemented.

IEDB-API has been implemented using a simple (RESTful) interface that allows a user to send a prediction request via a simple HTTP POST request to the IEDB-AR

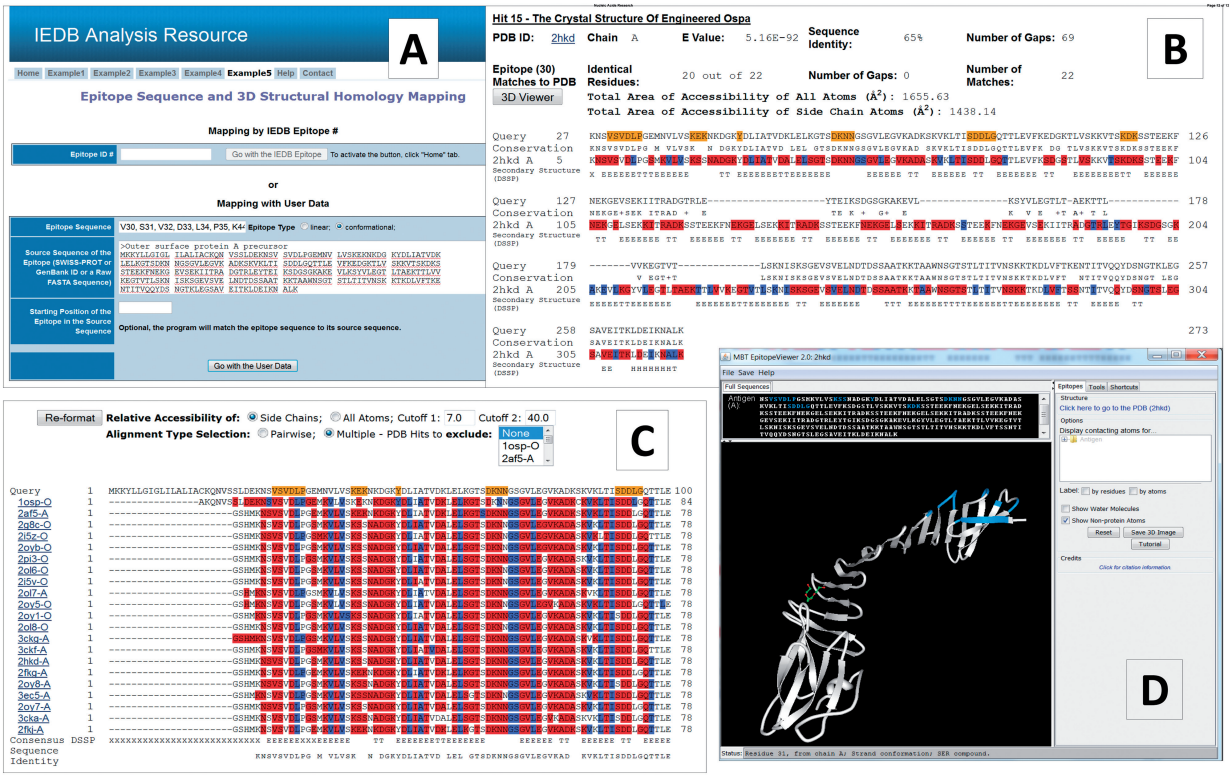


Figure 2. Screenshots of the homology modeling tool. (**A**) The input page. (**B**) The output page: a pair-wise sequence alignment of the source protein and one of the PDB hits. Epitope residues are shown in orange. Solvent exposed residues (with a relative solvent accessibility of side chain atoms, RSA, above 40%) are shown in red and buried (RSA below 7%), in blue (these cut-offs can be changed as shown in C). In the annotation for secondary structures (34), 'H' denotes an alpha-helix; 'G', a 3-10 helix; 'E', a beta-strand; 'T', a turn; 'X', no structure. (**C**) The output page: a fragment of a multiple sequence alignment of the source protein and all PDB hits (at the Blast *E*-value < 1.0E-3). (**D**) Default view of the protein source and epitope (colored in blue) in EpitopeViewer. The view can be changed using the EpitopeViewer's tools and shortcuts accessible on the right top panel.

server, which in turn generates a page with the prediction results. Such programmatic calls can be integrated into external applications, and they not only allow the user to run prediction jobs in batch mode but also ensure that the most up-to-date predictions will continue to be used. Finally, these calls do not require any local installation. Details on the use of IEDB-API as well as examples are provided at the IEDB-AR website (http://tools.iedb.org/main/html/iedb_api.html).

SOFTWARE DISTRIBUTION PACKAGES OF PREDICTIVE TOOLS

In addition to the web services described above, downloadable versions of the predictive tools are available freely for educational use, and with a licensing fee for commercial use. The MHC class I and class II binding prediction tools are available as a standalone package in compressed tarball and Ubuntu packages. These packages consist of Python scripts that are directly callable from the command line, making it convenient for batch processing.

Additionally, a virtual machine image of the IEDB Tools server is available for download and includes all tools on the website, with the user-friendly web interface to which users are accustomed. The image can be easily imported into an existing VMware ESX/ESXi deployment for site-wide access and should also run on a local installation of any virtualization software that supports OVF format. Both versions are updated on a 6-month cycle with the release of the website.

SUMMARY

A comprehensive set of immune epitope prediction and analysis tools are provided in the IEDB-AR. All components of the resource have been updated since the last publication in 2008, including retraining of prediction methods and addition of new methods. Furthermore, with the ability to use the IEDB-AR tools through a web-based API, it is becoming easier to integrate the tools into external applications. In the future, we plan to continue to make improvements to existing tools and add new ones. In addition to these efforts, we very much encourage people to notify us of predictive tools they are willing to contribute to the IEDB.

FUNDING

The IEDB is funded by National Institutes of Health [contract number HHSN272201200010C]. Funding for open access charge: National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

1. Zhang,Q., Wang,P., Kim,Y., Haste-Andersen,P., Beaver,J., Bourne,P.E., Bui,H.-H., Buus,S., Frankild,S., Greenbaum,J. *et al.* (2008) Immune epitope database analysis resource (IEDB-AR). *Nucleic. Acids Res.*, **36**, W513–W518.
2. Kim,Y., Sidney,J., Pinilla,C., Sette,A. and Peters,B. (2009) Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. *BMC Bioinformatics*, **10**, 394.
3. Peters,B. and Sette,A. (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*, **6**, 132.
4. Hoof,I., Peters,B., Sidney,J., Pedersen,L., Sette,A., Lund,O., Buus,S. and Nielsen,M. (2009) NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*, **61**, 1–13.
5. Nielsen,M., Lundegaard,C., Worning,P., Lauemoller,S.L., Lamberth,K., Buus,S., Brunak,S. and Lund,O. (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.*, **12**, 1007–1017.
6. Nielsen,M., Lundegaard,C., Blicher,T., Lamberth,K., Harndahl,M., Justesen,S., Roder,G., Peters,B., Sette,A., Lund,O. *et al.* (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One*, **2**, e796.
7. Lam,L. and Suen,S.Y. (1997) Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Trans. Syst. Man Cybern.*, **27A**, 553–568.
8. Moutaftsi,M., Peters,B., Pasquetto,V., Tscharke,D.C., Sidney,J., Bui,H.-H., Grey,H. and Sette,A. (2006) A consensus epitope prediction approach identifies the breadth of murine TCD8+-cell responses to vaccinia virus. *Nat. Biotech.*, **24**, 817–819.
9. Zhang,G.L., Ansari,H.R., Bradley,P., Cawley,G.C., Hertz,T., Hu,X., Jojic,N., Kim,Y., Kohlbacher,O. and Lund,O. (2011) Machine learning competition in immunology, prediction of HLA class I molecules. *J. Immunol. Methods*, **374**, 1–4.
10. Karosiene,E., Lundegaard,C., Lund,O. and Nielsen,M. (2012) NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics*, **64**, 177–186.
11. Bordner,A.J. (2010) Towards universal structure-based prediction of Class II MHC epitopes for diverse allotypes. *PLoS One*, **5**, e14383.
12. Yanover,C. and Bradley,P. (2011) Large-scale characterization of peptide-MHC binding landscapes with structural simulations. *Proc. Natl Acad. Sci. USA*, **108**, 6981–6986.
13. Wang,P., Sidney,J., Kim,Y., Sette,A., Lund,O., Nielsen,M. and Peters,B. (2010) Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*, **11**, 12.
14. Nielsen,M. and Lund,O. (2009) NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics*, **10**, 296.
15. Nielsen,M., Justesen,S., Lund,O., Lundegaard,C. and Buus,S. (2010) NetMHCIIpan-2.0 – improved pan-specific HLA-DR predictions using a novel concurrent alignment and weight optimization training procedure. *Immunome Res.*, **6**, 9.
16. Andreatta,M., Schafer-Nielsen,C., Lund,O., Buus,S. and Nielsen,M. (2011) NNAlign: a web-based prediction method allowing non-expert end-user discovery of sequence motifs in quantitative peptide data. *PLoS One*, **6**, e26781.
17. Shastri,N., Schwab,S. and Serwold,T. (2002) Producing nature's gene-chips: the generation of peptides for display by MHC Class I molecules. *Annu. Rev. Immunol.*, **20**, 463–493.
18. Paz,P., Brouwenstijn,N., Perry,R. and Shastri,N. (1999) Discrete proteolytic intermediates in the MHC Class I antigen processing pathway and MHC I dependent peptide trimming in the ER. *Immunity*, **11**, 241–251.
19. Nielsen,M., Lundegaard,C., Lund,O. and Keşmir,C. (2005) The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*, **57**, 33–41.
20. van Endert,P., Riganelli,D., Greco,G., Fleischhauer,K., Sidney,J., Sette,A. and JF,B. (1995) The peptide-binding motif for the human transporter associated with antigen processing. *J. Exp. Med.*, **182**
21. Burgevin,A., Saveanu,L., Kim,Y., Barilleau,E., Kotturi,M., Sette,A., van Endert,P. and Peters,B. (2008) A detailed analysis of the murine TAP transporter substrate specificity. *PLoS ONE*, **3**, e2402.
22. Stranzl,T., Larsen,M., Lundegaard,C. and Nielsen,M. (2010) NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics*, **62**, 357–368.
23. Ponomarenko,J., Bui,H.-H., Li,W., Fusseder,N., Bourne,P.E., Sette,A. and Peters,B. (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, **9**, 514.
24. Eswar,N., Webb,B., Marti-Renom,M.A., Madhusudhan,M.S., Eramian,D., Shen,M.-Y., Pieper,U. and Sali,A. (2006) Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics*, **15**, 5.6.1–5.6.30.
25. Ponomarenko,J. and Bourne,P.E. (2007) Antibody–protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct. Biol.*, **7**, 64.
26. El-Manzalawy,Y., Dobbs,D. and Honavar,V. (2008) Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.*, **21**, 243–255.
27. 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.
28. Thompson,J.D., Gibson,T.J. and Higgins,D.G. (2002) Multiple Sequence Alignment Using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, **2**, 2.3.1–2.3.22.
29. Goujon,M., McWilliam,H., Li,W., Valentin,F., Squizzato,S., Paern,J. and Lopez,R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.*, **38**, W695–W699.
30. Beaver,J., Bourne,P.E. and Ponomarenko,J. (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**, 6.
31. Toussaint,N.C. and Kohlbacher,O. (2009) OptiTope, a web server for the selection of an optimal set of peptides for epitope-based vaccines. *Nucleic Acids Res.*, **37**, W617–W622.
32. Vider-Shalit,T., Raffaeli,S. and Louzoun,Y. (2007) Virus-epitope vaccine design: informatic matching the HLA-I polymorphism to the virus genome. *Mol. Immunol.*, **44**, 1253–1261.
33. Yusim,K., Korber,B.T.M., Brander,C., Haynes,B.F., Koup,R., Moore,J.P., Walker,B.D. and Watkins,D.I. (eds), (2009) *HIV Molecular Immunology*. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico, LA-UR 09-05941.
34. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **12**, 2577–2637.