BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 8 2011, pages 1192–1193
doi:10.1093/bioinformatics/btr104

*Databases and ontologies*

Advance Access publication February 23, 2011

# MPID-T2: a database for sequence–structure–function analyses of pMHC and TR/pMHC structures

Javed Mohammed Khan[1], Harish Reddy Cheruku[1], Joo Chuan Tong[2,3] and Shoba Ranganathan[1,2,*]

[1]Department of Chemistry and Biomolecular Sciences and ARC Centre of Excellence in Bioinformatics, Macquarie University, Sydney, NSW, Australia, [2]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore and [3]Data Mining Department, Institute for Infocomm Research, Singapore, Singapore

Associate Editor: Jonathan Wren

**ABSTRACT**

**Summary:** Sequence–structure–function information is critical in understanding the mechanism of pMHC and TR/pMHC binding and recognition. A database for sequence–structure–function information on pMHC and TR/pMHC interactions, MHC-Peptide Interaction Database-TR version 2 (MPID-T2), is now available augmented with the latest PDB and IMGT/3Dstructure-DB data, advanced features and new parameters for the analysis of pMHC and TR/pMHC structures.

**Availability:** http://biolinfo.org/mpid-t2.

**Contact:** shoba.ranganathan@mq.edu.au

**Supplementary Information:** Supplementary data are available at *Bioinformatics* online.

Received on November 26, 2010; revised on February 15, 2011; accepted on February 18, 2011

## 1 INTRODUCTION

Major histocompatibility complexes (MHC) or human leukocyte antigens (HLAs) in human are important elements of T cell-mediated immunity. They are cell surface glycoproteins among which MHC-I proteins are ubiquitously expressed by most cells and MHC-II proteins are expressed by antigen-presenting cells (APC; Lefranc *et al*., 2005). MHC proteins bind immunogenic peptide epitopes (p) derived from antigens and present them as peptide–MHC (pMHC) complexes on the cell surface, for subsequent recognition by T-cell receptors (TR), leading to TR/pMHC complexes, which are responsible for T-cell activation and triggering the adaptive immune response cascade (Khan *et al*., 2010). Understanding the physicochemical basis for the selection of certain specific peptide epitopes by MHC alleles and the consequent recognition of pMHC ligands by TR proteins is critical for the design of T cell-based peptide vaccines (Khan *et al*., 2010).

An early collection of pMHC X-ray crystal structures in the Protein Data Bank (PDB; Berman *et al*., 2000) led to the development of MPID (Govindarajan *et al*., 2003), comprising 86 classical pMHC structures, reporting pMHC interaction parameters. With increasing pMHC and TR/pMHC structures in the PDB and in the IMGT/3Dstructure-DB (Kaas *et al*., 2004) and reports of new interaction parameters (Kaas and Lefranc, 2005), MPID-T (Tong *et al*., 2006) was developed, with 187 pMHC and 16 TR/pMHC structures along with interaction parameters for the analysis of pMHC structures alone.

With recent rise in TR/pMHC structural data in the PDB and in IMGT/3Dstructure-DB (Ehrenmann *et al*., 2010), and updated interaction parameters (Kaas *et al*., 2008), there is an increasing demand for more effective and efficient computational protocols to predict T-cell epitopes. Hence, we have updated MPID-T, augmenting it with advanced features and new parameters for the analysis of both pMHC and TR/pMHC structures, to gain an in-depth understanding of the structural determinants underlying TR/pMHC binding and recognition.

## 2 RESOURCE DESCRIPTION

MPID-T2 is a semiautomatically curated structure-derived MySQL database hosted on a Linux server, containing interaction information on all available experimental X-ray crystal structures of pMHC and TR/pMHC complexes extracted from PDB. MPID-T2 (November 2010 update) comprises 415 entries from five MHC sources (human: 282, murine: 127, rat: 3, chicken: 2 and monkey: 1), spanning 56 alleles; 353 pMHC structures, 62 TR/pMHC complexes; 352 MHC class I (MHC-I) complexes and 63 MHC class II (MHC-II) structures. Overall, 327 entries are non-redundant (MHC-I: 279 and MHC-II: 48). MPID-T2 includes non-classical structures (structures with T-cell receptor like antibodies, cluster of differentiation {CD} molecules and natural killer cell immunoglobulin like receptors {KIR} associated to the pMHC) and complexes with non-standard residues. For PDB structures with multiple molecular assemblies, the most accurate and complete structure is stored. Each structure is manually verified, classified and analyzed for pMHC and TR/pMHC interactions.

### 2.1 Definitions of interaction parameters

*2.1.1 Predefined interaction parameters* Existing MPID-T interaction parameters namely (i) intermolecular hydrogen bonds; (ii) gap volume; (iii) gap index; and (iv) interface area have been applied to all new pMHC complexes and extended to all TR/pMHC structures (Tong *et al*., 2006, 2007).

*2.1.2 New interaction parameters* Specific new interaction parameters in MPID-T2, vital for characterizing pMHC and/or TR/pMHC binding, computed from the 3D coordinates of the crystal structures, are listed below.

*To whom correspondence should be addressed.



© The Author 2011. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

*Binding energy*: binding energy (BE) is a measure of the strength of the interaction between the ligand and the receptor in terms of binding free energy ($\Delta G$). Values for BE between peptide and MHC for all structures and between pMHC and TR for TR/pMHC structures were calculated using DCOMPLEX (Liu *et al*., 2004).

*Molecular surface electrostatic potential*: interactions between TR and pMHC depend vastly on charges displayed by TR and pMHC binding interfaces. Hence, we used webPIPSA (Richter *et al*., 2008) to calculate and ICM (Internal Coordinate Mechanics; Abagyan *et al*., 1994) to visualize molecular surface electrostatic potential (MSEP) at the binding interfaces (Supplementary Fig. S1a, b).

*TR docking angle*: TR docking angle is the angle formed by the TR interface (paratope) on the pMHC interface (epitope) with respect to the linear axis of the cognate peptide within the MHC groove. This value '$\theta$' (Supplementary Fig. S1a) was calculated by matching the respective pMHC and TR interface MSEP for complimentarity of charges, augmented by TR/pMHC interacting residues from the literature. The charged residues at the pMHC interface form an ellipse. The angle between the major axis of the ellipse and the C$\alpha$ backbone axis of the peptide was measured using ICM.

*Contact area*: contact area (CA) is the area enclosed by the interacting residues of the two molecules (Supplementary Fig. S1c), as compared to interface area, which is the interaction area at the molecular level. We have used ICM to compute CA values between peptide and MHC for all structures and between pMHC and TR for TR/pMHC structures.

## 3 IMPLEMENTATION

Each entry in MPID-T2 is given a unique identifier for ease of identification, comparison, characterization and rapid visualization. Information for each pMHC and TR/pMHC structure in MPID-T2 is classified into five major categories: (i) MHC (chain-id, allele, class and source); (ii) peptide (chain-id, sequence, source and length); (iii) computed pMHC interaction parameters (intermolecular hydrogen bonds, gap volume, gap index, interface area, BE, CA and MSEP); (iv) structural information (structure determination method, resolution, PDB release year and publication reference); and (v) hyperlinks to related external databases like PDB (for sequence–structure information), SYFPEITHI (Rammensee *et al*., 1999; for MHC ligands and peptide motifs), IMGT/HLA (Robinson *et al*., 2001; for HLA sequences) and IMGT/3Dstructure-DB (for annotations on pMHC and TR/pMHC sequences with 3D structures; Ehrenmann *et al*., 2010). However, TR/pMHC structures in MPID-T2 have additional TR/pMHC interaction parameters (BE, MSEP, TR docking angle, CA, gap volume, gap index and interface area). Search page of the database presents a web interface that allows searching for pMHC and TR/pMHC complexes based on different categories (MHC class, allele, source organism, peptide length, user-defined output required and TR type) or PDB information (PDB-ID, resolution and release year; Supplementary Fig. S2a). The search output (Supplementary Fig. S2b) shows various fields; noticeably, TR/pMHC, pMHC, MHC, peptide and TR 3-D coordinates are downloadable for structural visualization. The alignment page illustrates pMHC and TR/pMHC structural alignments based on species, MHC allele, peptide length and TR type. To portray vital pMHC and TR/pMHC interactions, precomputed schematic diagrams, generated using LIGPLOT (Wallace *et al*., 1995), are provided. Also available in the patterns page of MPID-T2 are consensus patterns, obtained using WebLogo (Crooks *et al*., 2004), showing the conservation of residues among peptides with same lengths and alleles. MPID-T2 help page lists database usability details, definitions for interaction parameters and other useful resources.

## 4 DISCUSSION

MPID-T2 aims to facilitate mining of fundamental relationships and structural descriptors hidden within TR/pMHC and pMHC interactions for in-depth characterization. Inclusion of structural descriptors like BE, MSEP, TR docking angle and CA have facilitated in understanding the principles underlying TR/pMHC binding (Khan and Ranganathan, unpublished results). These descriptors can be used as parameters defining pMHC and TR/pMHC interactions, thereby facilitating rational development of methods to identify strong MHC binding T-cell epitopes with greater propensity to activate T cells. This highlights the utility of MPID-T2 in vaccine research. We have now enabled TR-specific searches by classifying TR/pMHC structures based on TR types. Future enhancements will include listing post-translational modifications (PTM) for peptides to help understand the effect of PTM on TR/pMHC binding and interaction. MPID-T2 will be updated on a quarterly basis.

*Conflict of Interest*: none declared.

## REFERENCES


Abagyan,R.A. *et al*. (1994) ICM: a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem*., **15**, 488–506.

Berman,H.M. *et al*. (2000) The Protein Data Bank. *Nucleic Acids Res*., **28**, 235–242.

Crooks,G.E. *et al*. (2004) WebLogo: a sequence logo generator. *Genome Res*., **14**, 1188–1190.

Ehrenmann,F. *et al*. (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.

Govindarajan,K.R. *et al*. (2003) MPID: MHC-Peptide Interaction Database for sequence-structure-function information on peptides binding to MHC molecules. *Bioinformatics*, **19**, 309–310.

Kaas,Q. and Lefranc,M.P. (2005) T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. *In Silico Biol.*, **5**, 505–528.

Kaas,Q. *et al*. (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–D210.

Kaas,Q. *et al*. (2008) IMGT standardization for molecular characterization of the T cell receptor/peptide/MHC complexes. In Schoenbach,C. *et al.* (eds) *Immunoinformatics*, *Immunomics Reviews Series*. Springer, New York, pp. 19–49.

Khan,J.M. *et al.* (2010) Structural immunoinformatics: understanding MHC-peptide-TR binding. In Davies,M.N. *et al.* (eds) *Bioinformatics for Immunomics*, Vol. 3, Springer, New York, pp. 77–94.

Lefranc,M.P. *et al*. (2005) IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev. Comput. Immunol.*, **29**, 917–938.

Liu,S. *et al*. (2004) A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. *Proteins*, **56**, 93–101.

Rammensee,H. *et al*. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, **50**, 213–219.

Richter,S. *et al*. (2008) webPIPSA: a web server for the comparison of protein interaction properties. *Nucleic Acids Res.*, **36**, W276–W280.

Robinson,J. *et al*. (2001) IMGT/HLA Database-a sequence database for the human major histocompatibility complex. *Nucleic Acids Res*., **29**, 210–213.

Tong,J.C. *et al*. (2006) MPID-T: database for sequence-structure-function information on T-cell receptor/peptide/MHC interactions. *Appl. Bioinformatics*, **5**, 111–114.

Tong,J.C. *et al*. (2007) *In silico* grouping of peptide/HLA class I complexes using structural interaction characteristics. *Bioinformatics,* **23**, 177–183.

Wallace,A.C. *et al*. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*., **8**, 127–134.


Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016