Bioinformatics, 35(24), 2019, 5323–5325
doi: 10.1093/bioinformatics/btz517
Advance Access Publication Date: 26 June 2019
Applications Note

OXFORD

Structural bioinformatics

TCR3d: The T cell receptor structural repertoire database

Ragul Gowthaman[1,2,3] and Brian G. Pierce[1,2,3,*]

[1]University of Maryland Institute for Bioscience and Biotechnology Research, Rockville, MD 20850, USA, [2]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA and [3]University of Maryland Marlene and Stewart Greenebaum Comprehensive Cancer Center, Baltimore, MD 21201, USA

*To whom correspondence should be addressed.
Associate Editor: Jonathan Wren

Received on March 1, 2019; revised on May 31, 2019; editorial decision on June 16, 2019; accepted on June 20, 2019

Abstract

Summary: T cell receptors (TCRs) are critical molecules of the adaptive immune system, capable of recognizing diverse antigens, including peptides, lipids and small molecules, and represent a rapidly growing class of therapeutics. Determining the structural and mechanistic basis of TCR targeting of antigens is a major challenge, as each individual has a vast and diverse repertoire of TCRs. Despite shared general recognition modes, diversity in TCR sequence and recognition represents a challenge to predictive modeling and computational techniques being developed to predict antigen specificity and mechanistic basis of TCR targeting. To this end, we have developed the TCR3d database, a resource containing all known TCR structures, with a particular focus on antigen recognition. TCR3d provides key information on antigen binding mode, interface features, loop sequences and germline gene usage. Users can interactively view TCR complex structures, search sequences of interest against known structures and sequences, and download curated datasets of structurally characterized TCR complexes. This database is updated on a weekly basis, and can serve the community as a centralized resource for those studying T cell receptors and their recognition.
Availability and implementation: The TCR3d database is available at https://tcr3d.ibbr.umd.edu/.
Contact: pierce@umd.edu

1 Introduction

T cell receptors (TCRs) are a critical component of the adaptive immune response, mediating the specific recognition of foreign peptide and small molecule antigens presented on the surfaces of cells. To enable their recognition of an immense array of foreign antigens, a large and diverse repertoire of unique TCRs is present in each individual. Accurate and potent TCR targeting of viral and tumor antigens is critical for TCR-based therapeutics and vaccine design, while aberrant TCR targeting is linked to autoimmune diseases such as diabetes, celiac disease and multiple sclerosis (Yin *et al.*, 2012). The 3D structures of these interfaces provide unparalleled views of these interactions, and can lead to better therapeutics, vaccines, and understanding of the molecular basis of many diseases. As noted in several reviews (Rossjohn *et al.*, 2015; Rudolph *et al.*, 2006), a number of experimentally determined TCR structures are available that collectively provide insights into the basis of TCR structure and targeting. This has led to identification of key residues and TCR features underlying specific binding of antigens and antigen-presenting molecules (Blevins *et al.*, 2016; DeWitt *et al.*, 2018; Stadinski *et al.*, 2016). To provide a central resource and interface for all known TCR structures, we have developed the TCR3d database. This complements recently developed databases focused on TCR sequences (Shugay *et al.*, 2018; Tickotsky *et al.*, 2017), and with a focus on TCR targeting and antigen interactions, in addition to other notable features described below, it is distinguished from a previously developed database, STCRDab (Leem *et al.*, 2018).

2 Materials and methods

TCR complex structures are automatically identified from the Protein Data Bank (PDB) (https://www.rcsb.org) on a weekly basis, through searching of all protein sequences from the PDB using hidden Markov models representing TCR variable domain sequences. A custom C++ program was developed to calculate TCR docking angles, based on the approach of Rudolph et al. (Rudolph *et al.*, 2006) (available at https://github.com/piercelab/tcr_docking_angle), while the NACCESS (http://wolf.bms.umist.ac.uk/naccess/) and S_c (from CCP4 suite; http://www.ccp4.ac.uk) programs are used to calculate interface buried surface area and shape complementarity, respectively. TCR variable germline genes are identified based on comparison with IMGT reference protein sequences (http://www.imgt.org). TCR complementarity determining region (CDR) loop clustering is performed using backbone $\varphi\psi$ conformational distances and affinity propagation, as previously described for the classification of antibody CDR loops (North *et al.*, 2011). The database and interface are implemented in SQLite and the Flask web framework, data plots are generated using plotly (https//plot.ly), and JSmol is used for structural visualization (http://www.jmol.org). Sequence searches against the database are performed using ungapped pairwise global or semi-global alignment (CDR3 and peptide sequences) or BLAST (Camacho *et al.*, 2009) (variable domain sequences).

3 Features and usage

TCR3d is designed to provide comprehensive and up-to-date information on all known TCR structures in the PDB, and a reference for key features of their antigen recognition. This includes:

- **TCR complex sets**: comprehensive tables of all known TCR complexes, organized by Major Histocompatibility Complex (MHC) class or MHC-like molecule, complex visualization
- **TCR complex analysis**: TCR docking angles, binding interface parameters, unbound structures and binding conformational changes
- **TCR structures**: germline gene coverage, CDR loop sequences, CDR loop structure clustering
- **TCR sequences**: sets of cancer-specific and virus-specific TCR sequences of interest, search sequences against known structures and an external database of TCR sequences (Shugay *et al.*, 2018)

A primary interface of the TCR3d database consists of browsable tables of all TCR-antigen complexes classified by TCR restriction (Class I MHC, Class II MHC, CD1d, MR1), as well as $\gamma\delta$ TCRs, containing key features of the TCRs and targeting. Links to measured wild-type and mutant binding affinities in the ATLAS database (Borrman *et al.*, 2017) are provided for each complex, if values are available. Clicking the PDB codes in these tables gives a dedicated viewing window of complex structures and binding axes (Fig. 1A). Interactive plots provide summaries of distribution of TCR docking angles and interface parameters (Fig. 1B), as well as germline gene representation from experimentally determined TCR structures (Fig. 1C). For those studying or predicting TCR binding, TCR3d includes a curated set of unbound TCR and peptide-MHC structures, corresponding to an updated version of a previously reported benchmark (Pierce and Weng, 2013). Users can download complex structures individually from the complex viewing page, and can download sets of all TCR complex structures from the Downloads page. CDR loop sequences from all known TCR structures, and clusters of CDR loop structures, are also available to users.

Though its primary focus is TCR structures and interfaces, TCR3d also contains sets of published TCR sequences, particularly those that target neoantigens and viral antigens. Users have the option to search any of those sequences against sequences of experimentally determined TCR structures, and a link is provided to submit sequences directly to the TCRmodel server (Gowthaman and Pierce, 2018) to generate a 3D model. TCR3d also allows users to search their own CDR3, peptide, or variable domain sequences against known structures, with the option to search using subsequence motifs for CDR3 or peptide sequences. Recent studies have demonstrated that sequence-based analysis can lead to insights into TCR targeting (DeWitt *et al.*, 2018; Glanville *et al.*, 2017), and by providing a simple and easily accessible interface on TCRs and their interactions through TCR3d, we hope to enable further advances incorporating structural and molecular data, leading to a predictive and mechanistic understanding of TCR structures, antigen recognition and specificity.

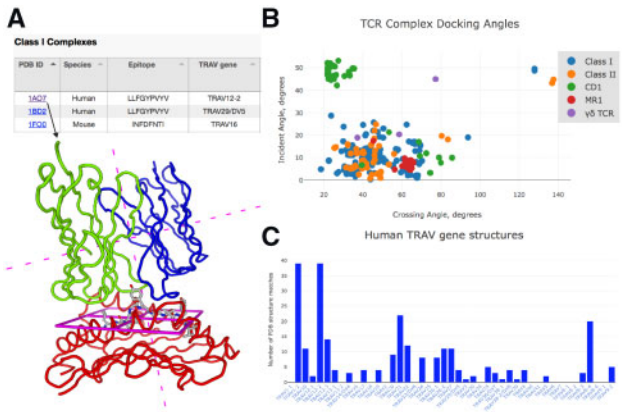

Fig. 1. TCR structural and complex data in TCR3d. (**A**) The tables of TCR complexes (top) link to interactive TCR complex structure viewers, which show the structure of the selected TCR (green, blue) bound to peptide or antigen (gray sticks) and MHC or MHC-like molecule (red), along with MHC-peptide plane (magenta grid) and TCR inter-domain axis and vector (dashed magenta lines), to indicate twist (crossing angle) and tilt (incident angle). (**B**) The set of calculated docking orientation angles (incident angle, crossing angle) for TCRs, classified according to TCR restriction and type. (**C**) Human TRAV gene representation in experimentally determined TCR structures (Color version of this figure is available at *Bioinformatics* online.)

Acknowledgements

We thank Johnathan Guest and Philip Bradley for helpful input, as well as Zhiping Weng, Sydney Blevins and Brian Baker, whose input and discussions spurred the initial development of the database. Tyler Borrman provided assistance with linking TCR complex entries to the ATLAS database, while Mikhail Shugay provided assistance with the VDJdb API. Christian Presley and the IBBR computing services team supported server and web site setup. We also express our thanks and appreciation to the x-ray crystallographers whose efforts produced the structures that provided the basis for this database.

Funding

This work was supported by the National Institutes of Health [GM126299 to B.G.P.].

Conflict of Interest: none declared.

References

Blevins,S.J. *et al.* (2016) How structural adaptability exists alongside HLA-A2 bias in the human alphabeta TCR repertoire. *Proc. Natl. Acad. Sci. USA*, **113**, E1276–1285.

Borrman,T. *et al.* (2017) ATLAS: a database linking binding affinities with structures for wild-type and mutant TCR-pMHC complexes. *Proteins*, **85**, 908–916.

Camacho,C. *et al*. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

DeWitt,W.S., 3rd. *et al*. (2018) Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *eLife*, **7**, e38358.

Glanville,J. *et al*. (2017) Identifying specificity groups in the T cell receptor repertoire. *Nature*, **547**, 94–98.

Gowthaman,R. and Pierce,B.G. (2018) TCRmodel: high resolution modeling of T cell receptors from sequence. *Nucleic Acids Res*., **46**, W396–W401.

Leem,J. *et al*. (2018) STCRDab: the structural T-cell receptor database. *Nucleic Acids Res*., **46**, D406–D412.

North,B.,Jr. *et al*. (2011) A new clustering of antibody CDR loop conformations. *J. Mol. Biol*., **406**, 228–256.

Pierce,B.G. and Weng,Z. (2013) A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes. *Protein Sci*., **22**, 35–46.

Rossjohn,J. *et al*. (2015) T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol*., **33**, 169–200.

Rudolph,M.G. *et al*. (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol*., **24**, 419–466.

Shugay,M. *et al*. (2018) VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res*., **46**, D419–D427.

Stadinski,B.D. *et al*. (2016) Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nat. Immunol*., **17**, 946–955.

Tickotsky,N. *et al*. (2017) McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics*, **33**, 2924–2929.

Yin,Y. *et al*. (2012) Structural basis for self-recognition by autoimmune T-cell receptors. *Immunol. Rev*., **250**, 32–48.