

# STCRDab: the structural T-cell receptor database

Jinwoo Leem, Saulo H.P. de Oliveira, Konrad Krawczyk and Charlotte M. Deane\*

Department of Statistics, University of Oxford, 24–29 St Giles, Oxford, OX1 3LB, UK

Received August 14, 2017; Revised September 22, 2017; Editorial Decision October 07, 2017; Accepted October 09, 2017

## ABSTRACT

**The Structural T-cell Receptor Database (STCRDab; <http://opig.stats.ox.ac.uk/webapps/stcrdab>) is an online resource that automatically collects and curates TCR structural data from the Protein Data Bank. For each entry, the database provides annotations, such as the  $\alpha/\beta$  or  $\gamma/\delta$  chain pairings, major histocompatibility complex details, and where available, antigen binding affinities. In addition, the orientation between the variable domains and the canonical forms of the complementarity-determining region loops are also provided. Users can select, view, and download individual or bulk sets of structures based on these criteria. Where available, STCRDab also finds antibody structures that are similar to TCRs, helping users explore the relationship between TCRs and antibodies.**

## INTRODUCTION

T-cell receptors (TCRs) are proteins of the adaptive immune response. They are expressed on the surfaces of T-cells and typically recognise peptides that are presented by major histocompatibility complex (MHC) molecules. Despite their micromolar binding affinity and potential cross-reactivity, TCRs are selective for foreign peptide-MHC complexes on antigen presenting cells (APCs; 1–3). Upon binding, TCRs can activate the T-cell for direct killing of APCs, or stimulate other components of the adaptive immune system, such as B-cells (4–6). The clinical relevance of TCRs has attracted interest in understanding the structural basis of a TCR's activity (7,8), and exploring the possibility of designing TCRs as novel biotherapeutics (9).

Given the sensitivity of TCR-MHC interactions and the extreme diversity of the TCR repertoire (10,11), computational methods are increasingly being used for rational TCR design (10,12–15). TCR structural data is an invaluable resource for designing and developing computational tools, for example, template-based modelling pipelines (12).

A small number of publicly available databases focus on delivering TCR-specific data (16–18). McPAS-TCR (18) is a manually curated database that maps  $\alpha\beta$  TCR sequences to pathogens or epitopes (18). The database does not contain structural information, making it difficult to determine

the importance of specific residues in MHC and antigen binding. There are two databases that contain some TCR structural information: ATLAS (16) and IMGT (17). ATLAS is a manually curated database, containing a large volume of affinity data; users can view and download one of 87 experimental structures, and retrieve summaries of individual queries. The bulk of the structural data in ATLAS is comprised of homology models of variants of experimental structures. These structures lack annotations that can be useful for further analyses (e.g. numbering; 16). Once again like McPAS-TCR, only  $\alpha\beta$  TCRs are annotated. IMGT (17) has a richer (308 experimental structures) and more diverse set of structural data (e.g.  $\gamma\delta$  TCRs). However, it is only possible to search based on a limited set of attributes; for example, it is not possible to specify the peptide sequence of the antigen. In addition, IMGT does not allow users to generate bespoke datasets for analysis (17).

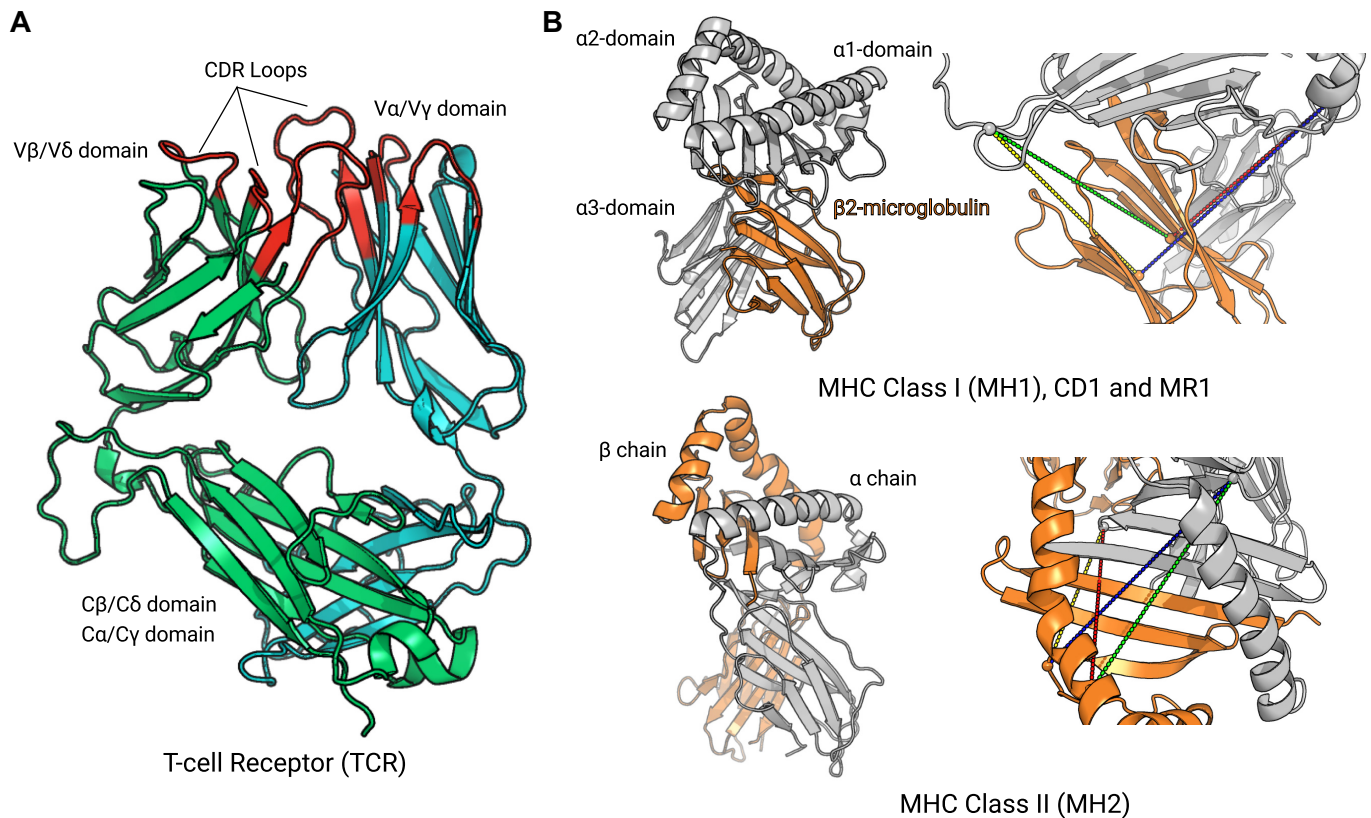
We have developed the Structural TCR Database (STCRDab), building on our Structural Antibody Database (SAbDab; 19). STCRDab is a TCR database that automatically collects and curates data on a weekly basis. Users can browse and select both  $\alpha\beta$  and  $\gamma\delta$  TCRs based on a wide range of criteria, such as the sequence of the TCR's complementarity-determining region (CDR) loops, the resolution of the structure, and the type of MHC molecule bound by the TCR. Users can also search by structural annotations, such as the orientation between the TCR's variable domains (20). STCRDab is linked to SAbDab, so that users can find antibody structures that are similar to TCRs, providing insight into designing TCR-like antibodies and chimaeric antigen receptors. Following a query, users can inspect and download individual or sets of TCR structures. Each search generates a unique zip file, containing a summary of the search and Protein Data Bank (PDB) format files of structures that match the query (21,22).

## Structure nomenclature

STCRDab is primarily focussed on consistently annotating TCR structural data, but also numbers MHC molecules consistently. The terminology for both types of structures is shown in Figure 1 and described below.

*TCR structures.* The majority of available TCR structures are  $\alpha\beta$  TCRs, which are formed of TCR $\alpha$  and TCR $\beta$  chains.

\*To whom correspondence should be addressed. Tel: +44 1865 281301; Email: deane@stats.ox.ac.uk



**Figure 1.** Nomenclature and colouring scheme used in STCRab. (A) T-cell receptors (TCRs) are formed from two chains: TCR $\beta$ /TCR $\alpha$  (to form  $\alpha\beta$  TCRs, as shown), or TCR $\delta$ /TCR $\gamma$  (to form  $\gamma\delta$  TCRs). The residues coloured in red indicate the IMGT-defined CDR loops. This colouring scheme is also used on the website. (B) Major histocompatibility complex (MHC) molecules can be divided into classical and nonclassical MHCs. MH1 and MH2 are considered ‘classical’ MHCs, while CD1 and MR1 are ‘nonclassical’. However, CD1 and MR1 are structurally similar to MH1, whereas MH2 is structurally distinct. To pair MH1, we use the following distance constraints:  $\alpha 15$ – $\beta 23$  (green; 32 Å),  $\alpha 15$ – $\beta 104$  (yellow; 32 Å),  $\alpha 51$ – $\beta 23$  (red; 32 Å),  $\alpha 51$ – $\beta 104$  (blue; 37 Å). To pair MH2, the following distance constraints are used:  $\alpha 29$ – $\beta 64$  (green; 34 Å),  $\alpha 29$ – $\beta 39$  (yellow; 22 Å),  $\alpha 37$ – $\beta 64$  (red; 32 Å),  $\alpha 37$ – $\beta 39$  (blue; 28 Å).

A small number of TCRs are  $\gamma\delta$  TCRs, consisting of TCR $\gamma$  and TCR $\delta$  chains. The TCR $\beta$  and TCR $\delta$  chains are considered to be analogous to antibody heavy chains while the TCR $\alpha$  and TCR $\gamma$  chains are considered to be analogous to antibody light chains (23).

Each TCR chain is characterised by two immunoglobulin domains: a variable domain (*V*) and a constant (*C*). Both variable and constant domains have a conserved  $\beta$ -sandwich structure (Figure 1), making it possible to number and compare variable domains from different TCRs (24). In STCRDab, we use the IMGT numbering as it provides consistent numbering for the CDR loops (21), and has been used on other occasions for structural analysis of TCRs (8,20). On each variable domain, there are three hypervariable loops that have the highest degree of sequence and structural variation, known as the CDRs. Flanking the CDRs, the remaining portions of the TCR structure are collectively known as the TCR’s ‘framework’.

**MHC structures.** APCs use either the ‘classical’ MHC to present peptide antigens, or the ‘nonclassical’ MHC-like molecules to present lipid molecules or vitamin B precursors (25). The classical MHCs can be subdivided into MHC class I (MH1) and MHC class II (MH2), while the nonclassical MHC-like molecules include cluster of differentia-

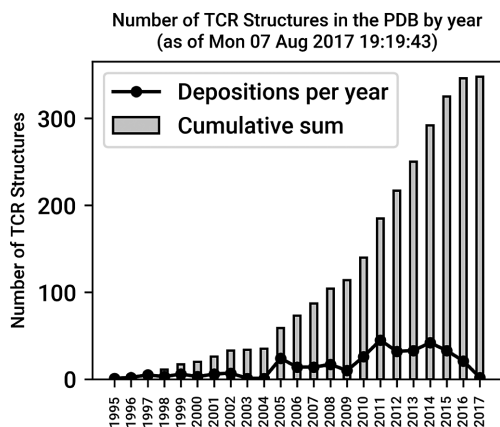
tion 1 (CD1) and MHC class I-related protein (MR1). Both classical and nonclassical MHCs have an antigen binding groove formed by a  $\beta$ -sheet, flanked by two  $\alpha$  helices (Figure 1). MH1, CD1 and MR1 are formed by the pairing of the MHC chain and a  $\beta_2$  microglobulin, while MH2 is formed by the MHC $\alpha$  and MHC $\beta$  chains. As with the TCR structures, the IMGT numbering is used for MHCs (22).

## DATA SOURCES AND CONTENTS

### TCR structures

As of 7 August 2017, STCRDab contains 348 entries with at least one TCR chain. On average, two TCR structures have been deposited in the PDB per month since 2007 (Figure 2). STCRDab is automatically updated weekly, in line with the PDB updating schedule (Figure 3). Paired  $\alpha\beta$  TCRs form the majority of the data, followed by single TCR chains, e.g. V $\beta$  only structures, then  $\gamma\delta$  TCRs (Table 1). There are also structures that fit none of these categories – for instance, an engineered TCR $\delta$ /TCR $\alpha$  receptor (PDB: 4wo4).

STCRDab uses a modified version of ANARCI (24) to detect and number any human and mouse TCR and MHC sequences in newly-released structures from the PDB. Briefly, ANARCI aligns sequences to a database of hidden Markov models (HMMs). The HMMs were built



**Figure 2.** Growth of TCR structures in the PDB. As of 7 August 2017, there are 348 entries of human and mouse TCR structures in the PDB. On average, two new structures have been deposited per month since 2007.

**Table 1.** Number of TCR structures in STCRDab

TCR type	Unbound	Classical MHC	Nonclassical MHC	Total
$\alpha\beta$ TCR	91	273	118	482
$\gamma\delta$ TCR	12	0	5	17
Unpaired TCR	80	0	0	80
Other	0	7	2	9

An entry in STCRDab can have multiple TCR structures, e.g. PDB: 2vlr.

using multiple sequence alignments of human and mouse TCR and MHC sequences from IMGT (17). We only pursue further annotation if there is at least one TCR chain in the PDB entry.

Similar cutoffs to those in SAbDab (19) are used for pairing  $\alpha\beta$  TCRs and  $\gamma\delta$  TCRs. We consider two TCR chains to pair if the distance between the C $\alpha$  atoms of the conserved cysteines at IMGT positions 104 in each domain is  $<22\text{\AA}$ . MHC chains are paired by using four distance constraints (Figure 1). For any MHC molecule that pairs with a  $\beta_2$  microglobulin, i.e. MH1, CD1, and MR1, we calculate the distance between the C $\alpha$  atoms of IMGT positions 15, 51 on the MH1/CD1/MR1 chain, and the C $\alpha$  atoms of IMGT positions 23 and 104 in the  $\beta_2$  microglobulin. For MH2 molecules, the distances between the C $\alpha$  atoms of IMGT positions 29 and 37 in the  $\alpha$  chain and IMGT positions 39 and 64 in the  $\beta$  chain are used.

The paired TCR and MHC molecules are matched together if there is at least one C $\beta$  atom from the TCR's CDR3 loops that is within  $8\text{\AA}$  of any of the C $\beta$  atoms in the helix regions of the MHC (8). The putative TCR-MHC pair with the highest number of C $\beta$ -C $\beta$  contacts is set as the TCR-MHC complex. Potential antigens are then identified by searching for proteins, peptides, and other non-polymeric ligands (haptens). Peptide and protein antigens are matched if a C $\beta$  atom is within  $8\text{\AA}$  of the TCR-MHC complex. For haptens, we apply a  $3.5\text{\AA}$  cutoff between its atoms and the MHC, and an  $8\text{\AA}$  cutoff with C $\beta$  atoms of the TCR's CDR3 loops. Unconventional structures, such as PDB: 2icw (which features a protein that is between the MHC and TCR), are flagged for manual inspection.

## V $\alpha$ -V $\beta$ orientation, docking angle

In order to describe the TCR binding mode with the MHC, we use a TCR-specific version of ABangle (TRangle; 20), and calculate the docking angle between the TCR and the MHC (23). TRangle describes the relative orientation between the V $\alpha$  and V $\beta$  domains using six parameters. The effect of V $\alpha$ -V $\beta$  orientation on MHC binding is not yet clear, though it can provide the basis for engineering TCR-like antibodies, or antibody-like TCRs (20). STCRDab automatically calculates the TRangles for  $\alpha\beta$  TCRs. Due to the small amount of data, the TRangle method is currently not used for  $\gamma\delta$  TCRs; however, as data increases, this will become possible.

The docking angle describes how the TCR engages with the MHC. Here, we implement a previously established formula to calculate the docking angle (23).

## Complementarity-determining region loops and clustering

In STCRDab, the CDR loops are identified using the IMGT definition (21): CDR1 (IMGT 27–38), CDR2 (IMGT 56–65) and CDR3 (IMGT 105–117). The CDR $\alpha$ 1, CDR $\alpha$ 2, CDR $\alpha$ 3, CDR $\beta$ 1, CDR $\beta$ 2 and CDR $\beta$ 3 loops have been clustered into canonical forms (12,26), as has been done for the CDR loops of antibody structures (27–29).

We have clustered the CDR loops of TCRs using a length-independent density-based clustering (DBSCAN) method, as first proposed for antibodies (29). Briefly, we took the CDR loops of all TCR $\beta$  and TCR $\alpha$  structures with resolution  $\leq 2.8\text{\AA}$ ; we removed loops with missing residues, or those that have at least one backbone atom with a B-factor of 80 or higher. We then calculated the length-independent root-mean square deviation (RMSD) between CDR loops using a dynamic time warp algorithm. The RMSD matrix is then clustered using DBSCAN. To compare our newly identified clusters, we map them to canonical forms from previous studies (12,26). The canonical forms for the CDR $\alpha$ 3 and CDR $\beta$ 3 loops are currently early-stage observations and are not yet useful for modelling (12). However with more data, we should have more accurate definitions of the canonical forms in TCRs, as is the case in antibodies.

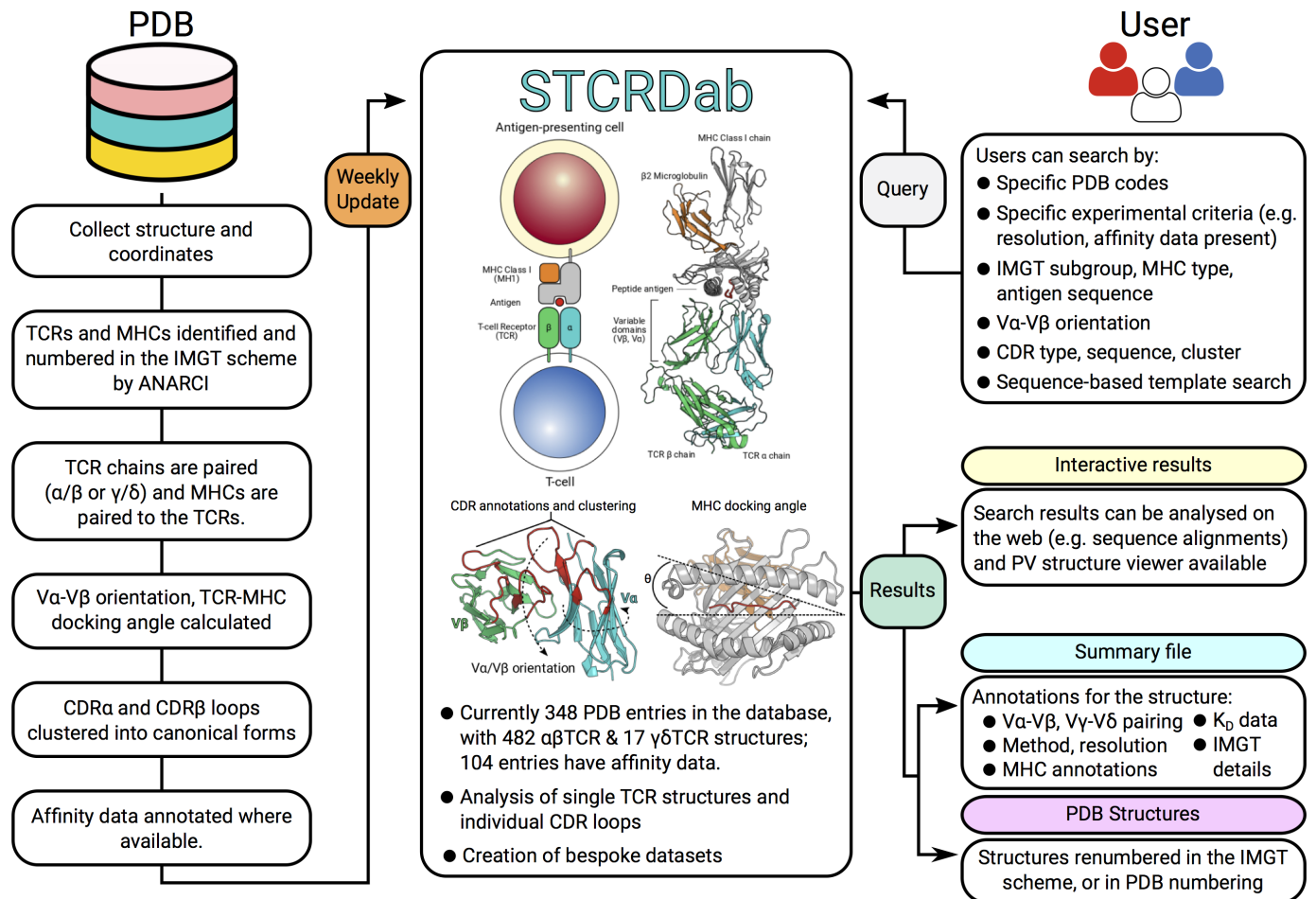
## TCR binding affinity

The binding affinities of TCR-MHC complexes were manually curated from PDBBind (30) and ATLAS (16). Where possible, experimental details describing how the affinity was measured (e.g. surface plasmon resonance) were also annotated. For cases where the affinity of a TCR-MHC complex was measured in multiple studies (e.g. PDB: 3qdj), the values from the authors that determined the TCR structure are cited. There are currently 104 entries in STCRDab with a  $K_D$  value. These values should serve as a useful resource for those interested in TCR docking and design.

## DATA ACCESS

### Download options

STCRDab provides a tab-separated file that summarises the results of a particular query with annotations for each



**Figure 3.** Workflow for STCRDab. Every week, STCRDab automatically detects and numbers newly released TCR structures from the PDB using ANARCI (24). Any MHC or MHC-like molecules are also numbered by ANARCI. Each structure is automatically annotated with several structural properties, such as its TCR-MHC docking angle (23). Users can submit a variety of queries to STCRDab to retrieve structures. Users are given their results for online analysis, and custom datasets are dynamically generated for download.

TCR structure. STCRDab also provides two sets of structure files: either the raw file directly from the PDB (31), or a structure file in re-numbered in the IMGT scheme (21,22) via ANARCI (24).

In the re-numbered PDB structure file, TCR variable domains, MHC G-domains, and the  $\beta 2$  microglobulin are numbered in the IMGT scheme (21,22). All non-TCR and non-MHC chains retain their numbering as in the original PDB file. The header of the renumbered PDB file contains TCR pairing information, along with their paired MHC and antigen in a REMARK field. For instance, in the entry 2vlr, there are two TCR structures, formed between chains E and D, and between chains J and I. The TCR E-D binds to the peptide antigen (chain C) presented by the MH1 molecule (A-B); likewise, TCR J-I binds to the antigen on chain H, presented by F-G. Thus, the header shows

```
REMARK 5 PAIRED_ABTCR BCHAIN=E ACHAIN=D MHCCHAINS=AB AGCHAIN=C AGTYPE=PEPTIDE
REMARK 5 PAIRED_ABTCR BCHAIN=J ACHAIN=I MHCCHAINS=FG AGCHAIN=H AGTYPE=PEPTIDE
```

The tab-separated summary file contains more detailed information about each entry, and can be used by most spreadsheet applications. Each column of the tab-separated

file contains more information for the TCR, such as the pairing information, and the paired MHC type.

pdb	Bchain	Achain	...	mhc_type	mhc_chain1	mhc_chain2	docking_angle	...
2vlr	E	D	...	MH1	A	B	61.224	...
2vlr	J	I	...	MH1	F	G	61.782	...

The summary file is highly flexible as it can be generated for one particular entry (as shown), or for a collection of entries that satisfy a user's search criteria. For every search, users can download these files individually per entry, or as a zip file that contains the collection of PDB files and the summary file.

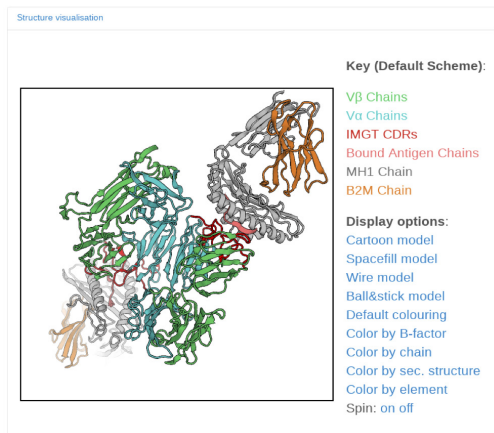
### Analysis of individual structures

An individual entry can be viewed interactively using its PDB accession code (e.g. 2vlr). Users will be directed to the summary page as shown in Figure 4A. The structure is visualised by BioPV (32); by default, the colouring scheme from Figure 1 is applied. It is also possible to use different representations (e.g. ball-and-stick model) and colouring schemes (e.g. colour by B-factor).

Below the viewer, STCRDab shows details of the PDB entry, including the organism information (as listed in the

**A** Details for 2vlr

Click on any of the tabs below to see the detailed information about the structure.

**B** 3. Apply an advanced search of structures.

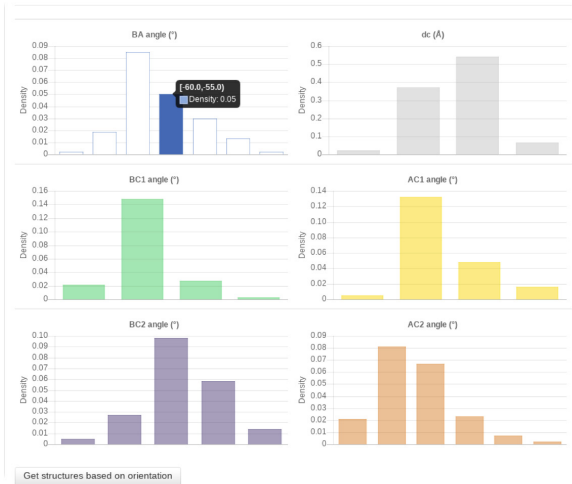
TCR Type (?):

Species (?):  Subgroup (?):

MHCtype (?):  Antigen type (?):  Antigen sequence (?):

Resolution cutoff (?):  R-factor cutoff (?):

Affinity present? (?):

**C** 1. Search for TCR structures by selecting each TRangle.**D** Pairwise alignments

Sequences have been IMGT-numbered and aligned.

CDR positions are indicated in RED.

Matching positions: green; mismatching positions: purple.

5mer: E/D

Alpha chain

Position	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15
Query	K	Q	E	V	T	Q	I	P	A	A	L	S	V	P	E
Template	-	I	Q	V	E	Q	S	P	P	D	L	I	L	Q	E

Position	A16	A17	A18	A19	A20	A21	A22	A23	A24	A25	A26	A27	A28	A29	A36
Query	G	E	N	L	V	L	N	C	S	F	T	D	S	A	I
Template	G	A	N	S	T	L	R	C	N	F	S	D	S	V	-

Position	A37	A38	A39	A40	A41	A42	A43	A44	A45	A46	A47	A48	A49	A50	A51
Query	Y	N	L	Q	W	F	R	Q	D	P	G	K	G	L	T
Template	N	N	L	Q	W	F	H	Q	N	P	W	G	Q	L	I

**Figure 4.** Screenshots of STCRDab. (A) Summary page for a single TCR structure in STCRDab. Each page has the IMGT-numbered TCR structure visualised in PV (32), followed by additional details and links. (B) The advanced ‘TCR search’ page gives users the option to specify a range of criteria to retrieve a subset of structures from STCRDab. (C) The orientation search page allows users to retrieve a subset of structures based on the TRangle parameters. (D) The sequence search results page shows how the query sequence is aligned with those in STCRDab.

PDB), the method for structure determination, and the method used for measuring the affinity. Next, for each detected TCR in the PDB entry, STCRDab provides additional annotations, such as the IMGT subgroup, the species of the IMGT subgroup, and the IMGT-numbered sequence, along with a FASTA file for the TCR sequence. Where relevant, STCRDab provides the orientation, docking angles, and links to view individual CDR loops.

At the end of the summary page, users are given links to download PDB files, either raw or re-numbered, along with the summary file.

**Analysis of individual CDR loops**

Each CDR loop has a unique CDR viewer page, which is both loop and chain-specific (e.g. CDRβ3 loop on chain E). STCRDab assigns the canonical class for the CDR loop where possible. In addition, if the CDR loop has a length-matched backbone RMSD of less than 1.5Å to

an analogous antibody CDR loop (i.e. CDRα1/CDRL1), STCRDab provides links to antibody structures from SAbDab (19).

**Search options**

Users can browse and search for TCR structures using a variety of options. For every query, a table is dynamically generated, summarising the results and providing links to summary pages and downloads.

*Advanced search.* The advanced search tool can help users filter TCR structures by TCR type (e.g. αβ TCR/γδ TCR), the IMGT species (human or mouse) and subgroup (e.g. TRAV10) of the TCR. It is also possible to apply filters based on the MHC type, or the antigen’s type (e.g. peptide versus hapten) and sequence. Finally, users can select structures based on quality (resolution, *R*-factor), and if there is affinity data available for the TCR. Upon submission, users

will be given a table listing the TCR structures that satisfy the specified criteria.

**CDR Search.** The CDR search engine provides similar criteria to the advanced search tool, such as antigen type, MHC type, and defining the IMGT subgroup of the TCR. Additional attributes, such as the type of CDR, loop length and canonical class are also available. Users are provided with a table showing the PDB codes and CDR sequences of every hit. The sequences are hyperlinked to their respective entries. The data can be downloaded as described in the ‘Download options’ section above.

**Orientation search.** The orientation search tool allows users to select TCRs by choosing specific bins of TRangle parameters from the interactive online graphs. It is not necessary to choose the bins for all six parameters, as unselected parameters act as a ‘wild card’. Users can also search for antibodies based on a particular TCR structure in a specific entry, e.g. 2vlr:ED. Since orientation calculations are only performed for  $\alpha\beta$  TCRs, this search tool will only retrieve  $\alpha\beta$  TCR structures.

Following selection, STCRDab returns a list of TCR structures that are within the user-defined  $V\alpha$ - $V\beta$  orientation space. Where possible, an antibody with a similar orientation to a TCR structure is also retrieved from SAbDab (19). To define orientation similarity, we use the ABangle distance measure  $d_{\text{ABangle}}$ , defined as

$$d_{\text{ABangle}} = \sqrt{\sum (\theta_{i,a} - \theta_{i,t})^2}. \quad (1)$$

$d_{\text{ABangle}}$  represents the Euclidean distance between the  $i$ th ABangle parameter,  $\theta_{i,a}$  and its analogous TRangle parameter  $\theta_{i,t}$ , e.g. ‘HL’ angle of antibodies vs. the ‘BA’ angle of TCRs. A TCR structure  $t$  and an antibody structure  $a$  were considered to be similar if  $d_{\text{ABangle}}$  is less than or equal to 10. We only list the closest antibody structure, i.e. lowest  $d_{\text{ABangle}}$  in the orientation search results page.

**Sequence search.** The sequence search engine allows users to submit TCR sequences and retrieve TCR structures that can be used as templates for template-based modelling tools. Users can submit the sequences of the TCR $\beta$  and/or TCR $\alpha$  chains to find templates based on sequence identity across the entire variable domain, the framework, or the CDR loops. STCRDab uses ANARCI to detect and number variable domains in the query sequence. STCRDab returns  $N$  structures in decreasing order of matched sequence identity. Given the number of structures that are available, this search method is only enabled for TCR $\beta$  and TCR $\alpha$  chains.

## CONCLUSION

STCRDab automatically collects and curates TCR structural data from the PDB. STCRDab builds upon the foundations of our antibody database, SAbDab, in order to provide consistent annotations, and open a gateway for users to easily access, view, and download custom datasets for analysis. The database aims to act as a resource for the emerging field of computational TCR design, and to help uncover the

unique structural properties of TCRs. STCRDab also provides a bridge to the extensive knowledge base of antibody structures in SAbDab, which can potentially be used to inform TCR-like antibody design or antibody-like TCR design. The database is entirely open-access and available at <http://opig.stats.ox.ac.uk/webapps/stcrdab>.

## FUNDING

Engineering and Physical Sciences Research Council (EP-SRC) [EP/G037280/1, EP/L016044/1]; Medical Research Council (MRC) [EP/L016044/1]. Funding for open access charge: RCUK OA Block Grant.

*Conflict of interest statement.* None declared.

## REFERENCES

- van der Merwe, P.A. and Dushek, O. (2011) Mechanisms for T cell receptor triggering. *Nat. Rev. Immunol.*, **11**, 47–55.
- Birnbaum, M.E., Mendoza, J.L., Sethi, D.K., Dong, S., Glanville, J., Dobbins, J., Özkan, E., Davis, M.M., Wucherpfennig, K.W. and Garcia, K.C. (2014) Deconstructing the peptide-MHC specificity of T cell recognition. *Cell*, **157**, 1073–1087.
- Morris, G.P. and Allen, P.M. (2012) How the TCR balances sensitivity and specificity for the recognition of self and pathogens. *Nat. Immunol.*, **13**, 121–128.
- Janeway, C.A., Travers, P., Walport, M. and Shlomchik, M.J. (2001) *Immunobiology: The Immune System in Health and Disease*, 5th edn, Garland Science.
- Beringer, D.X., Kleijwegt, F.S., Wiede, F., van der Slik, A.R., Loh, K.L., Petersen, J., Dudek, N.L., Duinkerken, G., Laban, S., Joosten, A. *et al.* (2015) T cell receptor reversed polarity recognition of a self-antigen major histocompatibility complex. *Nat. Immunol.*, **16**, 1153–1161.
- Knapp, B., Demharter, S., Esmailbeiki, R. and Deane, C.M. (2015) Current status and future challenges in T-cell receptor/peptide/MHC molecular dynamics simulations. *Brief. Bioinform.*, **16**, 1035–1044.
- Song, I., Gil, A., Mishra, R., Ghersi, D., Selin, L.K. and Stern, L.J. (2017) Broad TCR repertoire and diverse structural solutions for recognition of an immunodominant CD8+ T cell epitope. *Nat. Struct. Mol. Biol.*, **24**, 395–406.
- Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L.E., Rubelt, F., Ji, X., Han, A., Krams, S.M., Pettus, C. *et al.* (2017) Identifying specificity groups in the T cell receptor repertoire. *Nature*, **547**, 94–98.
- Fesnak, A.D., June, C.H. and Levine, B.L. (2016) Engineered T cells: the promise and challenges of cancer immunotherapy. *Nat. Rev. Cancer*, **16**, 566–581.
- Riley, T.P., Ayres, C.M., Hellman, L.M., Singh, N.K., Cosiano, M., Cimini, J.M., Anderson, M.J., Piepenbrink, K.H., Pierce, B.G., Weng, Z. *et al.* (2016) A generalized framework for computational design and mutational scanning of T-cell receptor binding interfaces. *Protein Eng. Des. Sel.*, **29**, 595–606.
- Attaf, M., Legut, M., Cole, D.K. and Sewell, A.K. (2015) The T cell antigen receptor: the Swiss army knife of the immune system. *Clin. Exp. Immunol.*, **181**, 1–18.
- Klausen, M.S., Anderson, M.V., Jespersen, M.C., Nielsen, M. and Marcatili, P. (2015) LYRA, a webserver for lymphocyte receptor structural modeling. *Nucleic Acids Res.*, **43**, W349–W355.
- Haidar, J.N., Pierce, B., Yu, Y., Tong, W., Li, M. and Weng, Z. (2009) Structure-based design of a T-cell receptor leads to nearly 100-fold improvement in binding affinity for pepMHC. *Proteins*, **74**, 948–960.
- Zoete, V., Irving, M., Ferber, M., Cuendet, M. and Michielin, O. (2013) Structure-based, rational design of T cell receptors. *Front. Immunol.*, **4**, 268.
- Malecek, K., Grigoryan, A., Zhong, S., Gu, W.J., Johnson, L.A., Rosenberg, S.A., Cardozo, T. and Krogsgaard, M. (2014) Specific increase in potency via structure-based design of a TCR. *J. Immunol.*, **193**, 2587–2599.
- Borrman, T., Cimini, J., Cosiano, M., Purcaro, M., Pierce, B.G., Baker, B.M. and Weng, Z. (2017) ATLAS: a database linking binding affinities with structures for wild-type and mutant TCR-pMHC complexes. *Proteins*, **85**, 908–916.

17. Kaas,Q., Ruiz,M. and Lefranc,M. (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.
18. Tickotsky,N., Sagiv,T., Prilusky,J., Shifrut,E. and Friedman,N. (2017) McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics*,**33**, 2924–2929.
19. Dunbar,J., Krawczyk,K., Leem,J., Baker,T., Fuchs,A., Georges,G., Shi,J. and Deane,C.M. (2014) SAbDab: the structural antibody database. *Nucleic Acids Res.*, **42**, D1140–D1146.
20. Dunbar,J., Knapp,B., Fuchs,A., Shi,J. and Deane,C.M. (2014) Examining variable domain orientations in antigen receptors gives insight into TCR-like antibody design. *PLOS Comput. Biol.*, **10**, 1–10.
21. Lefranc,M.P., Pommie,C., Kaas,Q., Duprat,E., Bosc,N., Guiraudou,D., Jean,C., Ruiz,M., Da Piedade,I., Rouard,M. *et al.* (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev. Comp. Immunol.*, **29**, 185–203.
22. Lefranc,M.-P., Duprat,E., Kaas,Q., Tranne,M., Thiriot,A. and Lefranc,G. (2005) IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev. Comp. Immunol.*, **29**, 917–938.
23. Rudolph,M.G., Stanfield,R.L. and Wilson,I.A. (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.*, **24**, 419–466.
24. Dunbar,J. and Deane,C.M. (2016) ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics*, **32**, 298–300.
25. Rossjohn,J., Gras,S., Miles,J.J., Turner,S.J., Godfrey,D.I. and McCluskey,J. (2015) T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol.*, **33**, 169–200.
26. Al-Lazikani,B., Lesk,A.M. and Chothia,C. (2000) Canonical structures for the hypervariable regions of T cell  $\alpha\beta$  receptors. *J. Mol. Biol.*, **295**, 979–995.
27. Chothia,C. and Lesk,A. (1987) Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.*, **196**, 901–917.
28. North,B., Lehmann,A. and Dunbrack,R.L. (2011) A new clustering of antibody CDR loop conformations. *J. Mol. Biol.*, **406**, 228–256.
29. Nowak,J., Baker,T., Georges,G., Kelm,S., Klostermann,S., Shi,J., Sridharan,S. and Deane,C.M. (2016) Length-independent structural similarities enrich the antibody CDR canonical class model. *mAbs*, **8**, 751–760.
30. Wang,R., Fang,X., Lu,Y., Yang,C.-Y. and Wang,S. (2005) The PDBbind database: methodologies and updates. *J. Med. Chem.*, **48**, 4111–4119.
31. 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.
32. Biasini,M. (2015) pv: v1.8.1.