

Structural bioinformatics

TCRBuilder: Multi-state T-cell receptor structure prediction

Wing Ki Wong¹, Claire Marks¹, Jinwoo Leem^{1,†}, Alan P. Lewis², Jiye Shi³ and Charlotte M. Deane^{1,*}

¹Department of Statistics, University of Oxford, Oxford OX1 3LB, UK, ²Data and Computational Sciences, GlaxoSmithKline Research and Development, Gunnels Wood Road, Stevenage, SG1 2NY, UK and ³Chemistry Department, UCB Pharma, 216 Bath Road, Slough, SL1 3WE, UK.

*To whom correspondence should be addressed.

†Present address: BenevolentAI, London, W1T 5HD, UK.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: T-cell receptors (TCRs) are immune proteins that primarily target peptide antigens presented by the major histocompatibility complex. They tend to have lower specificity and affinity than their antibody counterparts, and their binding sites have been shown to adopt multiple conformations, which is potentially an important factor for their polyspecificity. None of the current TCR modelling tools predict this variability which limits our ability to accurately predict TCR binding.

Results: We present TCRBuilder, a multi-state TCR structure prediction tool. Given a paired $\alpha\beta$ TCR sequence, TCRBuilder returns a model or an ensemble of models covering the potential conformations of the binding site. This enables the analysis of structurally-driven polyspecificity in TCRs, which is not possible with existing tools.

Availability: <http://opig.stats.ox.ac.uk/resources>

Contact: deane@stats.ox.ac.uk

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

1 Introduction

In the immune system, antigens are foreign molecules recognised by two types of receptors with similar architectures: antibodies and T-cell receptors (TCRs). In humans, paired $\alpha\beta$ TCRs dominate the TCR population. They consist of one α and one β chain, analogous to the light-heavy chain dimer in antibodies. These four chains are encoded by the repertoire V, D and J genes of the respective TCR or antibody loci, and share a similar β -sandwich fold (Dunbar *et al.*, 2014). Despite the similar genetic formation and architecture, their biological functions differ: TCRs bind to peptide antigens presented by the major histocompatibility complex (MHC), and their binding is known to be polyspecific (Rudolph *et al.*, 2006). Antibodies on the other hand can bind to virtually any antigen with high specificity and affinity (Sela-Culang *et al.*, 2013). This distinction in binding preferences may be related to differences in their binding site structures (Wong *et al.*, 2019). For both TCRs and antibodies, the majority

of the binding site is made up of six loops known as the complementarity-determining regions (CDRs). It has been shown that when multiple solved structures of an identical sequence TCR CDR are available, 20% adopt distinct conformations, which is far higher than the ~8% observed in antibodies despite the far larger number of antibody structures (Wong *et al.*, 2019). This structural variability may allow alternative binding configurations for a given TCR, leading to its polyspecificity (Garcia *et al.*, 1998; Morris and Allen, 2012; Wong *et al.*, 2019).

Several packages exist to model the structures of TCRs including LYRA, TCRModel and Repertoire Builder (Klaussen *et al.*, 2015; Gowthaman and Pierce, 2018; Schmitt *et al.*, 2019). However, none of these tools produce predictions that reflect the multiple conformations of TCR CDRs. Here, we present TCRBuilder, a multi-state TCR modelling server. TCRBuilder is based on the ABodyBuilder pipeline for antibody modelling (Leem *et al.*, 2016) and predicts TCR models with multiple CDR conformations where applicable.

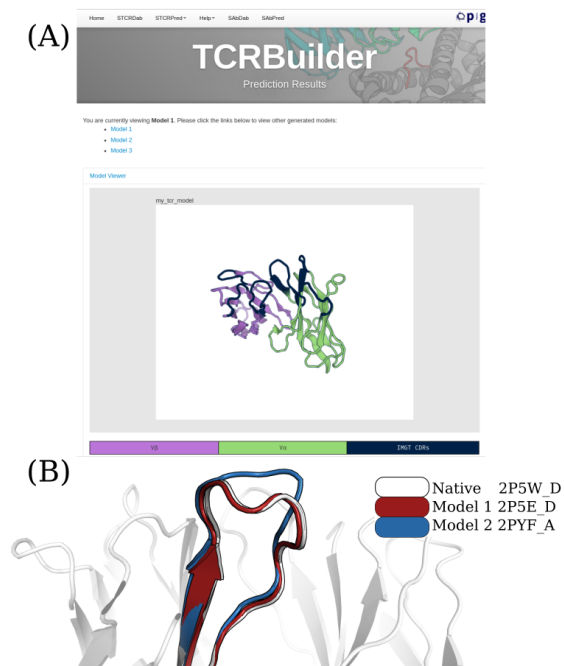


Fig. 1. (A) An example result page of the TCRBuilder web server. (B) An example case where a CDR α 3 sequence (AVRPLLDGTYIPT) with multiple distinct conformations was modelled. One of the conformations (2P5E chain D in red) was nearly identical to the native 2P5W chain D structure, but an alternative conformation of this loop (2PYF chain A in blue) is also predicted as a possible shape.

2 Webserver

TCRBuilder is available as a web server (Fig. 1A). It takes as input a paired $\alpha\beta$ TCR sequence, and returns model structures with simple biophysical property annotations similar to those in ABodyBuilder (Leem *et al.*, 2016). The detailed pipeline is described in the Supplementary Materials.

2.1 Performance analysis

To evaluate TCRBuilder, we carried out a leave-one-out cross-validation on 169 non-redundant TCR sequences in STCRDab (Leem *et al.*, 2018), barring the algorithm from using full sequence-identical template structures. The performance was assessed by superimposing the predicted model onto the native structure, and calculating the root-mean-square-distance (RMSD) over the backbone atoms of the given region. RMSDs over the individual CDRs were obtained after aligning two anchor residues before and two after the CDR. If there were multiple crystal structures or models for a given sequence, all measurements were recorded. The average RMSD for the complete model was 1.49 Å. In the models, CDRs 1 and 2 on both the α and β chains are well predicted, but CDR3 less so. Of the 169 models generated, 106 required template-free modelling (Marks *et al.*, 2017) for at least one of their CDRs. This is a far larger percentage than the <4% of antibody cases that required template-free modelling in a large test (Leem *et al.*, 2016). These results highlight the lack of structural data available for TCRs.

To test whether TCRBuilder is comparable to current state-of-the-art, freely-available TCR webserver, we constructed a blind test set of the 10 (9 non-redundant) $\alpha\beta$ TCR structures released between 24 July 2019 and 22 October 2019. We compared the modelling performance of LYRA, TCRModel and Repertoire Builder with TCRBuilder (Klausen *et al.*, 2015; Gowthaman and Pierce, 2018; Schmitt *et al.*, 2019), preventing them from using any structures released since 24 July 2019. TCRBuilder

has comparable performance with existing TCR modelling tools, on this set of unseen structures (Supplementary Materials).

2.2 Predicting multiple conformations

As TCR CDR sequences are structurally variable, TCRBuilder is designed to capture the distinct conformations of CDRs that are known to exist (Wong *et al.*, 2019). Eighty-nine of the 169 evaluated models are predicted to have more than one conformation of their binding sites. Figure 1B shows an example of a CDR α 3 sequence that can have multiple distinct conformations. As the number of TCR structures increases, it is likely that we will find many more CDR sequences with variable conformations.

3 Conclusion

TCRBuilder is the first TCR modelling server that captures the multiple stable conformations in TCR CDRs. The multi-state models it generates better represent the possible binding configurations a TCR can adopt, paving the way to understand polyspecificity in TCRs.

Funding

This work was supported by funding from the Engineering and Physical Sciences Research Council and Medical Research Council [grant number EP/L016044/1].

References

- Dunbar, J. *et al.* (2014). Examining variable domain orientations in antigen receptors gives insight into tcr-like antibody design. *PLoS Computational Biology*, **10**(9), 1–10.
- Garcia, K. C. *et al.* (1998). Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science*, **279**(5354), 1166–1172.
- Gowthaman, R. and Pierce, B. G. (2018). TCRmodel: high resolution modeling of T cell receptors from sequence. *Nucleic Acids Research*, **46**(W1), W396–W401.
- Klausen, M. S. *et al.* (2015). LYRA, a webserver for lymphocyte receptor structural modeling. *Nucleic Acids Research*, **43**(W1), W349–W355.
- Leem, J. *et al.* (2016). ABodyBuilder: Automated antibody structure prediction with data-driven accuracy estimation. *MAbs*, **8**(7), 1259–1268.
- Leem, J. *et al.* (2018). STCRDab: the structural T-cell receptor database. *Nucleic Acids Research*, **46**(D1), D406–D412.
- Marks, C. *et al.* (2017). Sphinx: merging knowledge-based and *ab initio* approaches to improve protein loop prediction. *Bioinformatics*, **33**(9), 1346–1353.
- Morris, G. P. and Allen, P. M. (2012). How the TCR balances sensitivity and specificity for the recognition of self and pathogens. *Nature Immunology*, **13**(2), 121.
- Rudolph, M. G. *et al.* (2006). How TCRs bind MHCs, peptides, and coreceptors. *Annual Review of Immunology*, **24**(1), 419–466.
- Schmitt, D. *et al.* (2019). Repertoire Builder: high-throughput structural modeling of B and T cell receptors. *Molecular Systems Design & Engineering*, **4**, 761–768.
- Sela-Culang, I. *et al.* (2013). The Structural Basis of Antibody–Antigen Recognition. *Frontiers in Immunology*, **4**, 302.
- Wong, W. K. *et al.* (2019). Comparative analysis of the CDR loops of antigen receptors. *Frontiers in Immunology*, **10**, 2454.