

# **Engineering TCRs For Improved Stability By Targeting The Non-CDR Framework Regions**

Catriona McMurrán

A thesis submitted to Cardiff University  
in candidature for the degree of  
Doctor of Philosophy

January 2022



# Summary

---

T-cell receptors (TCRs) form a key part of the immune system by recognising peptides presented on the cell surface by the major histocompatibility complex. Recombination events between different gene segments create a highly diverse TCR repertoire to bind diverse targets, with sequence diversity occurring not only at the binding interface but across entire TCR variable domains. Framework regions that do not make direct binding contacts are still variable in sequence and may influence function. Framework diversity is also a vital consideration for the use of TCRs in pharmaceutical contexts, where different TCRs may have different developability challenges such as stability and aggregation propensity.

This thesis examines the framework regions of the TCR through multiple different approaches. A comparison of structures with different framework sequences highlighted key residues at inter-domain interfaces and identified specific amino acids that are best placed to make these inter-domain contacts. These key residues reflect those previously identified in published stabilising strategies and offers a new model for predicting and evaluating stabilising mutagenesis. A novel dataset of experimentally determined melting points for over 200 soluble TCRs was collated, revealing a broad range of intrinsic protein stabilities. Examination for sequence features that could determine overall protein stability did not support theories that certain genes or combinations of genes were inherently more stable, or that the hydrophobicity of surface-exposed residues was the determining factor; instead TCR thermal stability is highly variable even between similar sequences. Alongside these broader analyses, a detailed study of stabilising mutagenesis was carried out on two related TCRs to identify point mutations in framework regions that improve thermal stability. The most successful of these mutations were shown to improve stability in not only closely related molecules but also a panel of sequence-diverse TCRs, representing a more general approach to framework engineering for more stable molecules.

# Acknowledgements

---

I am grateful for the opportunity to carry out this PhD research in collaboration between Immunocore and Cardiff University, and for the generous funding of Immunocore Ltd. Andrew Sewell and David Cole provided useful and incisive comments and feedback on my research and writing, and I'm particularly indebted to Stephen Harper for the invaluable help and guidance he has offered throughout my PhD and my career at Immunocore.

I'd like to thank all my colleagues at Immunocore for their helpful comments and practical support throughout my PhD, in particular Sunir Malla, Nikolai Lissin, Bibiana Macedo and all members of Protein Science Research past and present.

I can't thank my friends enough for all their support throughout my PhD, from cakes in the post to long chats on the phone, and for providing sympathy and Sunday dinners in equal measure. Finally, I will forever be grateful to my parents for their love and encouragement.

# Table of Contents

---

<b>Declaration</b>	<b>i</b>
<b>Summary</b>	<b>ii</b>
<b>Acknowledgements</b> .....	<b>iii</b>
<b>Publications</b>	<b>iii</b>
<b>Table of Contents</b> .....	<b>iv</b>
<b>List Of Figures</b> .....	<b>vii</b>
<b>List Of Tables</b>	<b>viii</b>
<b>Abbreviations used</b> .....	<b>ix</b>
<b>Chapter One Introduction</b> .....	<b>1</b>
1.1 T cells and their role in the immune system .....	1
1.1.1 T cell subtypes .....	1
1.1.2 T -cell antigen binding and activation .....	2
1.2 TCR target binding.....	6
1.2.1 Contacts to peptide and MHC.....	6
1.3 T cell receptor sequence and structural diversity .....	10
1.3.2 VDJ recombination .....	12
1.3.3 Overview of T cell receptor structure.....	14
1.3.4 T cell receptor structure compared to antibodies.....	15
1.3.5 Amino acid diversity exists across the entire variable domain of TCRs .....	19
1.3.6 TCR interdomain interface.....	20
1.3.7 Preferential pairing between V regions.....	21
1.4 Use of T cell receptors as therapeutic agents .....	22
1.4.1 Adoptive T cell therapies .....	22
1.4.2 TCRs as part of a soluble protein therapeutic agent.....	24
1.5 Design requirements for protein-based drugs .....	25
1.5.1 Developability of protein therapeutic agents.....	25
1.5.2 Lessons from the field of antibody engineering .....	31
1.6 TCR stability is an engineering challenge .....	34
1.6.1 Creating a stable and soluble TCR molecule.....	34
1.6.2 Stabilising TCR constant domains .....	35
1.6.3 Directed evolution for TCR stability .....	38
1.6.4 Stabilising TCRs to increase cell surface expression .....	40
1.7 Summary of research aims .....	41
<b>Chapter Two Materials and Methods</b> .....	<b>44</b>
2.1 Protein expression and purification .....	44
2.1.1 Bacterial culture media .....	44
2.1.2 Bacterial strains used .....	44
2.1.3 Vectors and cloning.....	45
2.1.4 Inclusion body expression.....	45
2.1.5 Refold and purification .....	46

2.2 Protein analysis.....	47
2.2.1 Surface Plasmon Resonance .....	47
2.2.2 Thermal stability measurement .....	47
2.2.3 Serum stability measurement .....	48
2.3 Modelling and structural analysis .....	50
2.3.1 Structure preparation.....	50
2.3.2 Protein contact analysis .....	51
2.3.3 Residue exposure, hydrophobicity, and diversity scoring.....	52
2.3.4 Modelling of framework mutations .....	53
2.4 Phage Display.....	55
2.4.1 Phage library construction .....	56
2.4.2 Electroporation of phagemid constructs .....	57
2.4.3 Phage preparation .....	58
2.4.4 Optimising selective pressure for stability.....	58
2.4.5 Phage panning .....	59
2.4.6 Next-generation sequencing of phage outputs. ....	60
<b>Chapter Three Mapping structural diversity at TCR interfaces .....</b>	<b>62</b>
3.1 Introduction.....	62
3.1.1 Aims of chapter.....	63
3.2 Results.....	63
3.2.1 Structural analysis of a representative subset of TCRs .....	63
3.2.2 Interactions at the V $\alpha$ :C $\alpha$ interface.....	76
3.2.3 Interactions at the V $\beta$ :C $\beta$ interface .....	77
3.2.4 Overview of the V $\alpha$ :V $\beta$ interface .....	79
3.2.5 Conserved interactions at the V $\alpha$ :V $\beta$ interface .....	81
3.2.6 Diversity at the V $\alpha$ :V $\beta$ interface.....	82
3.3 Conclusion.....	84
<b>Chapter Four Trends in TCR thermostability.....</b>	<b>86</b>
4.1 Introduction.....	86
4.1.1 Aims of chapter.....	87
4.2 Results.....	88
4.2.1 Thermostability analysis of refolded TCRs.....	88
4.2.2 Influence of V-genes on thermostability of soluble TCR proteins .....	89
4.2.3 CDR chemistry is not predictive of thermostability .....	91
4.2.4 Structural features of TCRs which may alter stability .....	93
4.2.5 Sequence features that dominate in more thermally stable molecules.....	95
4.2.6 Thermostability changes during affinity maturation of TCRs .....	98
4.3 Discussion .....	100
<b>Chapter Five Optimising TCR stability through directed evolution ....</b>	<b>106</b>
5.1 Introduction.....	106
5.1.1 Historical approaches to TCR stability enhancement .....	106

5.1.2 Aims of chapter.....	110
5.2 Results.....	111
5.2.1 Choice of TCRs for study.....	111
5.2.2 Modelling of potential thermostability mutants.....	112
5.2.3 Establishing selection for thermal stability of TCRs displayed on a phage particle .....	113
5.2.4 Building and panning deep mutational scan libraries.....	117
5.2.5 Enrichment of stability enhancing mutations .....	119
5.2.6 Enrichment patterns in different structural features .....	121
5.2.7 Biochemical characterisation of enriched mutations .....	126
5.2.8 Hydrophobic core residues enriched post-selection .....	128
5.2.9 Solvent exposed residues enriched post-selection .....	129
5.2.10 Variable-constant interface mutations enriched post-selection .....	130
5.3 Combining point mutations.....	131
5.4 Discussion .....	133
<b>Chapter Six Global approaches to stabilise TCRs .....</b>	<b>136</b>
6.1 Introduction.....	136
6.1.1 Aims of chapter.....	138
6.2 Results.....	139
6.2.1 Translating stability enhancing mutations onto related molecules .....	139
6.2.2 Stabilising mutations can be translated onto unrelated TCR sequences ..	142
6.2.3 Combining the best alpha and beta mutations .....	144
6.2.4 Comparison of V-domain framework mutations to published stabilising strategies.....	145
6.2.5 Mutations that increase melting point also improve stability in serum ...	147
6.3 Discussion .....	152
<b>Chapter Seven.....General discussion</b>	<b>156</b>
7.1 Summary of research .....	156
7.1.1 Structural classification.....	156
7.1.2 Stability of TCRs .....	157
7.1.3 Screening for stability enhancing mutations .....	158
7.2 Future directions.....	163
7.2.1 Understanding TCR surface expression and protein stability .....	163
7.2.2 Using existing stability information to guide future engineering .....	165
7.3 Concluding remarks.....	168
<b>List of References .....</b>	<b>169</b>

## List Of Figures

---

Figure 1.1 Overview of T cell receptors and signalling pathway.....	3
Figure 1.2 Diagram of canonical docking alignment.....	8
Figure 1.3 Overview of TCR structure and genes .....	13
Figure 1.5 Comparison of antibody and T cell receptor structures.....	15
Figure 3.1 Overview of TCR:pHLA binding geometries.....	71
Figure 3.2 Overview of inter-domain geometries for TCR structures.....	73
Figure 3.3. Solvent exposure and residue amino acid diversity .....	76
Figure 3.4 : Overview of V $\alpha$ :C $\alpha$ contacts.....	77
Figure 3.5: Overview of V $\beta$ :C $\beta$ contacts.....	79
Figure 3.6: Overview of V $\alpha$ :V $\beta$ interface contacts. ....	80
Figure 3.7: Interactions at the V $\alpha$ :V $\beta$ interface .....	82
Figure 3.8: Heat map showing TRAV:TRBV inter-domain contacts. ....	83
Figure 4.1. Thermal stability of refolded wild-type TCRs .....	91
Figure 4.2. CDR features and TCR stability.....	93
Figure 4.3. Hydrophobicity score and structural features of the TCR. ....	95
Figure 4.4: Analysis of overall melting point (T <sub>m</sub> ) values.....	97
Figure 4.5. Melting point values after affinity maturation.....	100
Figure 5.1 Schematic of the three TCR capture strategies trialed .....	114
Figure 5.2. Establishing selection for thermal stability of TCRs.....	116
Figure 5.3. Illustration of regions mutated in library design .....	118
Figure 5.4. Heat map of V $\alpha$ domain of two TCRs.....	124
Figure 5.5. Heat map of V $\beta$ domain of two TCRs .....	126
Figure 6.1 Location of stability enhancing mutations.....	140
Figure 6.2 Serum stability and DSF melt curves for TCR-scFV fusion proteins .....	152

## List Of Tables

---

Table 1.1 Overview of different stability measurements.....	26
Table 3.1. List of V and J genes represented in the dataset .....	64
Table 3.2: List of human V and J genes which did not have an available crystal structure. ....	67
Table 3.3 Breakdown of TCR interfaces. ....	71
Table 5.1. Enrichment score for all mutations present .....	120
Table 5.2. Properties of refolded mutants .....	126
Table 5.3. Biochemical properties of TCRs with multiple mutations .....	131
Table 6.1. Mutations and improvement to thermal stability .....	141
Table 6.2. Mutations and improvement to thermal stability .....	143
Table 6.3. Mutations and improvement to thermal stability. ....	145
Table 6.4 Comparison of thermal stability values. ....	147
Table 7.1: Comparison of different display systems .....	160

## Abbreviations used

---

$\alpha\beta$ TCR	Alpha-Beta T cell receptor
$\beta$ 2m	Beta2-microglobulin
$\gamma\delta$ TCR	Gamma-Delta T cell receptor
Abs	Absorbance
AIEX	Anion Exchange
BCR	B cell receptor
BiTE	Bispecific T-cell engager
bp	Base pair
BSA	Bovine serum albumin
C	Celsius
CD (number)	Cluster of differentiation (number)
CDR	Complementarity determining region
CIEX	Cation Exchange
CM5	Carboxymethylated dextran
cfu	Colony Forming Units
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DSC	Differential Scanning Calorimetry
DSF	Differential Scanning Fluorimetry
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-Linked Immunosorbent Assay
FACS	Fluorescence Activated Cell Sorting
Fv	Variable fragment
Fab	Fragment of antigen binding
GF	Gel filtration
HLA	Human Leukocyte Antigen
HRP	Horseradish peroxidase
HSA	Human serum albumin
IEX	Ion exchange
Ig	Immunoglobulin
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
ITAM	Immunoreceptor Tyrosine-based Activation Motif
$K_D$	Dissociation constant
$K_{off}$	Dissociation off-rate
$K_{on}$	Dissociation on-rate
LB	Lysogeny broth
mAB	Monoclonal antibody
MAIT	Mucosal-Associated Invariant T-cell
MHC	Major Histocompatibility complex
mRNA	Messenger Ribonucleic Acid
mTCR	Monoclonal TCR
MW	Molecular Weight
NK T cell	Natural Killer T cell
NGS	Next Generation Sequencing
OD	Optical density

PBS	Phosphate Buffered Saline
PBST	Phosphate Buffered Saline Tween
PCR	Polymerase Chain Reaction
pHLA	Peptide-Human Leukocyte Antigen
pMHC	Peptide-Major Histocompatibility complex
RACE	Rapid Amplification of cDNA Ends
RNA	Ribonucleic Acid
RPM	Revolutions Per Minute
RU	Response Unit
scFV	Single chain antibody fragment variable domains
scTCR	Single chain T cell receptor
scTV	Single chain T cell receptor variable domains
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SHM	Somatic hypermutation
SPR	Surface Plasmon Resonance
$T_{1/2}$	Dissociation half-life
TCR	T cell receptor
$T_m$	Melting temperature
VDJ	Variable diversity and joining region
VH	Variable Heavy chain (Antibody)
VL	Variable Light chain (Antibody)
UV	Ultraviolet

All units of measurement and their abbreviations follow the International System of Units unless otherwise specified. Amino acids are designated by their single letter or three letter nomenclature.

# Chapter One

## Introduction

---

### **1.1 T cells and their role in the immune system**

The immune system has evolved to protect against pathogens through both the fixed innate immune system (which includes physical barriers like skin as well as cells which respond to standard pathogen signals) and the adaptive immune system, which produces a repertoire of cells to recognise different pathogen antigenic markers after exposure and either produce antibodies (B cells) or attack infected cells directly (T cells). Cytotoxic T cells form a key part of the adaptive immune system, identifying foreign antigens by using their membrane bound T-cell receptor (TCR) to recognise and bind non-self peptides presented by the major histocompatibility complex (MHC). The MHC presents short peptides from proteosomal degradation of cytoplasmic peptides onto the surface of almost all cell types, providing an external snapshot of the internal proteome. In this chapter I will give a brief overview of the function and structure of T cell receptors, and address the engineering challenges that arise when modifying this class of molecules for use as soluble protein based therapeutics.

#### **1.1.1 T cell subtypes**

T cells are split into different subcategories based on function and expression of different cell surface proteins. The majority of human T cells (approx. 90%) express a T-cell receptor encoded by alpha ( $\alpha$ ) and beta ( $\beta$ ) V genes, but a smaller and less well understood subtype instead use gamma ( $\gamma$ ) and delta ( $\delta$ ) genes; these  $\gamma\delta$  T cells are amongst those considered as “unconventional” T cells (Pellicci, Koay and Berzins, 2020; Willcox, Mohammed and Willcox, 2020). This thesis will focus mostly on conventional pMHC binding  $\alpha\beta$  T cells.

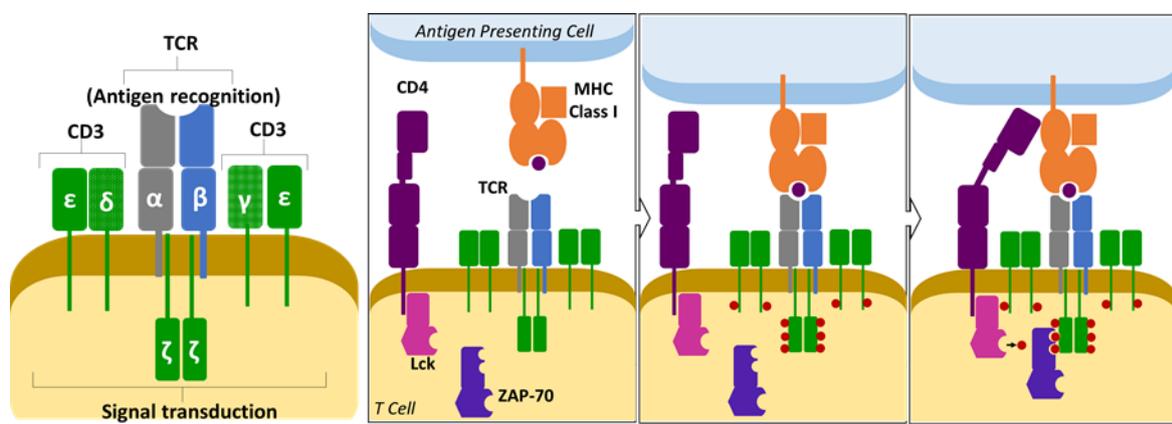
Within the conventional  $\alpha\beta$  T cells, there are multiple subtypes with different functions to provide immunity; these are often separated based on the presence of

key surface proteins, either CD4+ (a lineage that includes helper T cells) or CD8+ (cytotoxic T cells). These two lineages generally recognise the two main classes of MHC molecule, MHC class I (bound by CD8+) and II (bound by CD4+ cells). The MHC classes both present short peptides in a central groove between two alpha helices: this groove is closed at both ends in the class I type, restricting it to shorter peptides in the 8-10 amino acid length range, whilst the MHC class II molecule presents longer peptides due to the open ends of the peptide binding groove. MHC class II molecules are presented on specialised antigen presenting cells such as dendritic cells, macrophages, and activated B cells, and these peptides are often taken from exogenous proteins that have been internalised by the antigen presenting cell. MHC class I molecules are displayed on almost all cells in the body, and present peptides from degraded cytosolic proteins. Both classes of MHC are highly polymorphic and also polygenic; there are multiple different allelic variants of each gene within the MHC classes across the human population, and each cell will present MHC molecules encoded by multiple different alleles. The variation in MHC creates significant diversity; both in the peptide sequences that are preferentially bound and presented by different variants of the complex, and in the MHC molecule surrounding the groove that presents the peptide.

### **1.1.2 T-cell antigen binding and activation**

T cells require a complex assembly of proteins in order to translate extracellular target binding to an internal cellular response. The structure of the full TCR-CD3 signalling complex has recently been solved through cryogenic electron microscopy, and represented only a single static snapshot of what is likely to be a dynamic complex (Dong *et al.*, 2019). The T-cell receptor complex comprises the membrane anchored heterodimeric  $\alpha\beta$ TCR, which recognises and binds to pMHC targets; the CD3 co-receptor complex which in humans comprises of heterodimeric CD3 $\epsilon\gamma$  and CD3 $\epsilon\delta$  proteins; and the homodimeric CD3  $\zeta$ -chain, which together are required to translate pMHC binding into cytoplasmic signalling. TCR:pMHC binding results in phosphorylation of sites on the CD3 cytoplasmic domains known as ITAMs

(immunoreceptor tyrosine-based activation motifs) and the colocalisation of the CD4 or CD8 coreceptors to the immune synapse. This ITAM phosphorylation creates binding sites for the Zap70 kinase, which is activated by the Lck kinase and triggers a signalling cascade (Courtney, Lo and Weiss, 2018; J.-R. Hwang *et al.*, 2020). Figure 1.1 shows a simplified schematic of the T cell signal transduction pathway for CD4+ T cells binding to peptides presented by MHC class II. A similar pathway is mediated by the CD8 co receptor when CD8+ T cells bind to peptides presented by MHC class I molecules.



*Figure 1.1 Overview of T cell receptors and signalling pathway. Schematic representation of an  $\alpha\beta$  T cell receptor is shown in blue; it is a heterodimeric protein with four extracellular domains and short transmembrane domains, and forms a complex with the  $\zeta$ ,  $\epsilon$ ,  $\delta$  and  $\gamma$  CD3 subunits as shown. The extracellular domains of the TCR bind to peptides presented by the MHC (in this example, MHC class I), and binding triggers phosphorylation of the cytoplasmic regions of CD3 subunits (red circles represent added phosphoryl group ( $PO_3^-$ )). Phosphorylation creates binding sites for the Zap70 kinase, which is activated by Lck kinase recruited to the binding synapse by association with CD8 (that binds to MHC Class I, as shown in this schematic. (Adapted from P. Parham, *The Immune System*. Garland Science, 2 ed., 2004)*

### 1.1.2.1 Other proteins have a stabilising influence on T cell surface expression

The TCR-CD3 complex is not a static feature of the T-cell plasma membrane, but instead is repeatedly recycled between the extracellular surface and internal endosomes. Membrane bound  $\alpha\beta$  T cell receptors are not stably present on the cell surface in isolation but instead form part of a complex with CD3; in the absence of the CD3 $\zeta$  subunit the whole complex is rapidly degraded at the cell surface and recycled back into the T cell (Alcover, Alarcón and Di Bartolo, 2018). This dynamism

allows for rapid recruitment of TCR signalling complexes towards immune synapses as the T cell interacts with other cells and allows for up- or down-regulation of TCRs in response to other stimuli such as binding. There must be a balance maintained between sufficient TCRs on the cell surface to allow for target recognition and signalling and preventing off-target activation from transient TCR:pMHC interactions. TCR stability on the surface can be influenced by coreceptor protein interactions but the structural features of specific TCRs may also impact stability on the surface; different TCR sequences show higher or lower levels of surface expression when artificially introduced into T cells (Heemskerk *et al.*, 2007).

#### 1.1.2.2 Possible mechanisms for TCR signal transduction

The mechanism by which the extracellular event of TCR binding to pMHC is transmitted into the cytoplasm for signalling is still under debate (van der Merwe and Dushek, 2011; Mariuzza, Agnihotri and Orban, 2020). I will give a brief overview of some of the signal transduction models proposed, particularly those where variation in framework regions of the TCR could plausibly exert an influence on signalling behaviour.

TCR binding to pMHC could induce signalling through a kinetic segregation model, in which inhibitory phosphatases with large extracellular domains on the TCR surface (CD45) are constantly suppressing activation by dephosphorylating ITAM domains. Binding between the TCR and pMHC of the target cell creates a narrow synapse that excludes the larger phosphatases, allowing the constitutive phosphorylation of ITAM domains by kinases such as Lck (Figure 1.2) to build up to a level that induces TCR triggering. In this model, differences in the framework domains of the TCR would only impact signalling through their effect on TCR:pMHC binding. In support of the kinetic segregation model, elongating pMHC molecules with extra domains so the TCR:pMHC binding interface is further from the antigen presenting cell surface has been shown to reduce T cell activation (Choudhuri *et al.*, 2005). This supported the idea that the large inhibitory molecules are normally excluded from the binding synapse by the short distance between T cell and target

cell surfaces. However, this elongation could also have disrupted mechanical signalling mechanisms by reducing the amount of physical force transmitted between the pMHC presenting cell and the T cell.

The kinetic segregation model would suggest that details of the TCR structure are only relevant for signalling in terms of the physical distance between the two membranes at the immune synapse. However, alternative models of signal transduction have been proposed where conformational change to the TCR is a vital step in the initiation of T cell signalling.

Comparison of bound and unbound TCR structures has not shown any significant difference in TCR structure that might indicate a straightforward signalling pathway through induced large scale conformational change on binding. Some NMR evidence has suggested that pMHC binding results in the movement of residues in the beta constant domain in the same region where CD3 extracellular subunits interact, which could engage the coreceptor to initiate the phosphorylation cascade (Natarajan *et al.*, 2017). Alternatively (or additionally) the CA loop of C $\alpha$  domain has also been proposed as a site of conformational change upon binding, perhaps followed by dimerization of TCRs to activate signalling. Both these models involve allosteric changes in the TCR constant domains, but as TCRs bind pMHC with loops at the opposite end of their V domains the binding signal would have to be transferred through the V domain in order to enact these conformational shifts. The interface between variable and constant domains for both alpha and beta chains is formed by interactions between residues that vary in amino acid use between different V genes. This may modulate signalling behaviour across the V:C interface depending on the sequence of the TCR V domain; I will examine the variations in these interfaces in more detail in Chapter Three.

The proposal of allosteric regions in the constant domain(s) is not the only TCR conformational change model for signal transduction on pMHC binding. A mechanosensor model has been proposed where there are “catch bonds” within the structure of the TCR that are unstable in a resting state, but when under mechanical

load stress induced by pMHC binding these bonds increase in their strength and permit longer binding to allow for coreceptor recruitment even with a weak affinity interaction (Feng, Reinherz and Lang, 2018). Molecular dynamics modelling suggested that the V $\alpha$ :V $\beta$  interface dynamics are important in allowing signal transduction through this type of whole-protein movement (W. Hwang *et al.*, 2020). Motion between the two domains is suppressed on pMHC binding and application of physical load, where they are stabilised by contacts to the MHC allowing correct positioning of the CDR loops to make peptide contacts. This model suggests a key role for V $\alpha$ :V $\beta$  interface residues in modulating TCR signal transduction; although the authors discuss this in terms of a conserved mechanism of relative motion between V $\alpha$  and V $\beta$  in all TCRs there exists significant amino acid variation at the V $\alpha$ :V $\beta$  interface (discussed in more detail later in the introduction) that could result in TCR sequence-dependent variation in signalling under this hypothetical model. All these models of TCR signal transduction are not necessarily mutually exclusive, and multiple mechanisms could occur in concert in order to transduce the extracellular signal of pMHC binding into the cytoplasm to begin the T cell signalling cascade.

## **1.2 TCR target binding**

The class I major histocompatibility complex (MHC) presents short peptides from proteosomal degradation of cytoplasmic peptides onto the surface of almost all cell types. Cytotoxic T cells use a membrane bound receptor (TCR) to recognise and bind non-self peptides presented by the MHC. The alpha-beta heterodimeric T cell receptor binds to specific peptide-MHC (pMHC) with three sequence diverse loops per chain (shown in Figure 1.3 and discussed in more detail later), creating a highly specific binding interface between these complementarity determining regions (CDRs) and the pMHC.

### **1.2.1 Contacts to peptide and MHC**

Alpha beta TCRs evolved to bind the pMHC using six flexible loops, first identified by Kabat, Wu and Bilofsky, (1977) from sequence information alone due to high levels of amino acid diversity in those regions compared to the overall TCR. Residues within these CDR loops form the interface between TCR and pMHC in a generally consistent fashion regardless of TCR or peptide sequence, with a canonically diagonal binding mode across the groove of the pMHC. Figure 1.2 shows this canonical binding footprint of the TCR; it positions the CDR3 from both alpha and beta chains across the centre of the peptide in the MHC groove, whilst the CDR1 and CDR2 loops on both chains are more distant from the peptide and tend to make contacts mostly to the MHC helices. Peptide specificity is vital to allow targeted binding and signalling of T-cells, rather than binding to the MHC alone without discrimination between self and non-self peptides. However, TCR binding contacts have been shown to spread evenly across the peptide and the MHC helices in terms of binding energetics, not focused only on making peptide contacts. This contrasts with the docking mode of antibodies that have been engineered to target specific pMHCs, where modelling suggested that the antigen binding was focused to a few energetic hotspots (Holland *et al.*, 2020). The differences between TCRs and antibodies (even when binding the same pMHC target) will be discussed in more detail in subsequent sections.

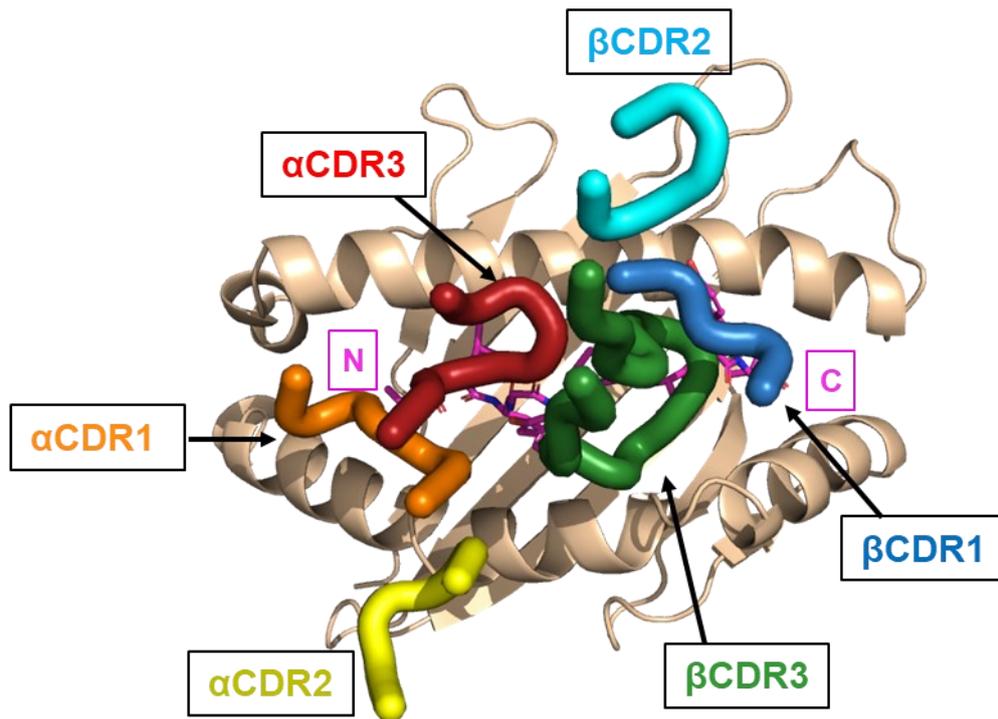


Figure 1.2 Diagram of TCR CDR loops binding to pMHC in a canonical docking alignment. MHC Class I protein shown in a top-down view looking into the groove made up by two alpha helices (light brown cartoon). The peptide (shown in pink with backbone and sidechains represented as sticks) sits within this groove with the N and C termini positioned as labelled. For clarity, only the CDR loops of the TCR are shown, with the rest of the variable and constant domains hidden. Loops are coloured and labelled to show where they interact in this canonical docking mode; the alpha chain loops sit closer to the N terminus and the beta chain loops closer to the C terminus, with CDR2 of both chains making contact to the MHC helices and the CDR3 loops positioned over the centre of the peptide. Structural image taken from crystal structure of the Tax-A6 TCR (PDB code 1AO7).

TCRs have generally much weaker affinity to their pMHC targets than the binding affinities observed for antibodies (an analogous molecule within the adaptive immune system) to their cognate antigen. Although, as seen in Figure 1.2, specific contacts are made across the whole pMHC surface by the TCR CDR loops, molecular dynamics modelling of TCR:pMHC interactions suggested that there were no specific interaction hotspots that drove binding but instead the energetics of binding were spread over a broad surface (Holland *et al.*, 2020). This large but weakly interacting surface may be an outcome of the thymic selection process, in which T cells expressing a TCR that is incapable of binding MHC undergo apoptosis. All TCRs presented on mature T cells must therefore make at least some contacts to the MHC helices to pass through positive selection. It may also be that the binding footprint is

widely spread across the whole pMHC surface to allow for mismatches and cross-reactivity to related peptide sequences; in order to cover the theoretical diversity of peptides that could be presented on MHC in the body TCRs must be cross-reactive (Mason, 1998).

The factors which determine the canonical docking mode shown in Figure 1.2 are a matter of debate; CDR1 and CDR2 loops are encoded in the germline of the TRAV and TRBV repertoire (the use of different V genes is discussed in more detail later) and this has led to an argument that they have evolved to make hard wired contacts to the MHC helices, leaving the more sequence diverse CDR3 loops to create peptide specificity (Adams *et al.*, 2015). This would explain why, although some non-canonical “reverse binders” have been identified through X-ray crystallography (Szeto *et al.*, 2021), the majority of TCRs where crystal structures have been obtained show a consistent binding footprint. Canonical binding has been shown to be essential for TCR signalling (Zareie *et al.*, 2021), with the implication that coreceptor assembly and signalling was significantly disrupted by reverse binding.

The MHC family of proteins are encoded by polymorphic loci across the human population and have different amino acids at the surface of the alpha helices that make up the peptide binding groove, and different germline TRBV genes seem to be preferentially expressed in T cells depending on the MHC haplotype present (Sharon *et al.*, 2016) (a finding also observed in naïve T cells isolated from umbilical cord blood, which will have passed through thymic selection but not been exposed to any antigens (Gao *et al.*, 2019)). This suggests that different germline CDR1 and CDR2 loops may be better suited to making contacts to different MHC allelic variants based on intrinsic compatibility between the amino acid sidechains used.

However, the canonical docking mode to MHC may instead be enforced by the presence of other proteins; the full complex of TCR and coreceptor molecules is required for productive signalling and the requirement to form this signalling complex when binding pMHC may impose steric requirements on how the TCR can interact with pMHC. Mice models where MHC class I and II and the CD4 and CD8

coreceptor genes were all knocked out were able to produce TCRs that bound non-MHC targets (Van Laethem *et al.*, 2007). MHC restriction may well be created by a combination of innate compatibility and enforcement by thymic selection (La Gruta *et al.*, 2018).

### **1.3 T cell receptor sequence and structural diversity**

The extracellular domain of the T cell receptor is a heterodimer that has a modular structure of four domains ( $V\alpha C\alpha$ ,  $V\beta C\beta$ ) that position the six CDR loops (discussed above) to contact peptide-MHC in the manner shown in Figure 1.2. Figure 1.3 shows an illustration of an example membrane-bound TCR structure with the loops highlighted. The genes that encode these loops are also shown; the CDR1 and 2 loops are encoded by different V genes from the *tra* and *trb* gene loci, and have a high level of sequence diversity between different genes despite the fact that CDR2 loops in particular typically contact only the MHC helices rather than the more diverse peptides presented within the groove. The CDR3 loops (which mostly contact peptide) are hyper-variable; the exact sequence of the loop is not germline encoded but instead is created by recombination events between different genes during the development of mature T-cells.

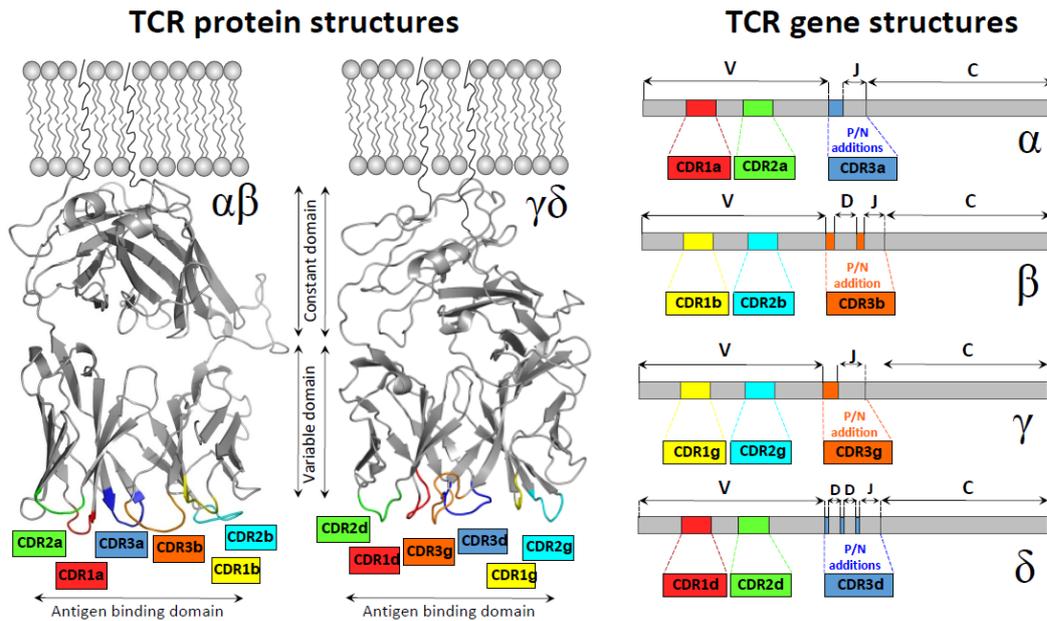


Figure 1.3 Overview of TCR structure and the genes that encode TCRs. On the left two example crystal structures of  $\alpha\beta$  and  $\gamma\delta$  are shown with a schematic representation of the T cell membrane. Variable domains (furthest from the membrane) have the six CDR loops that make up the antigen binding domain labelled according to chain and loop number. On the right, a schematic of the DNA sequence for each chain in these proteins is shown using the same colour scheme. The gene segments V (variable), D (diversity), J (joining) and C (constant) which encode each section of the protein are labelled. (Adapted from Attaf et al. (2015))

V- and J- gene encoded sequence diversity in the variable domains of the TCR extends beyond the CDR positions that bind pMHC into the whole architecture of the protein, which could indicate an evolutionary advantage in maintaining sequence diversity across the entire domain. Both the variable and constant domains are made up of immunoglobulin folds, a common protein motif.

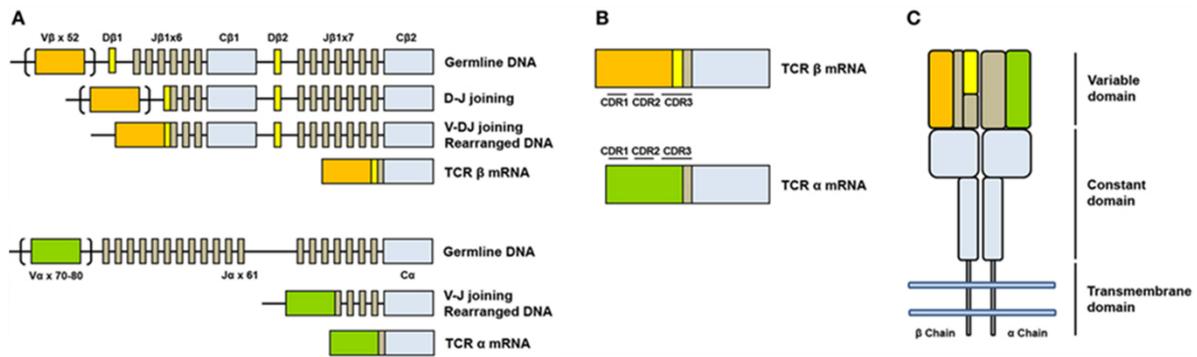
### 1.3.1.1 Immunoglobulin-like domains

As discussed above, the extracellular domains of TCRs are made up of two chains (alpha and beta, or delta and gamma for that subclass of T cells) of two domains each (the variable and constant). These four domains share a common architecture based on an immunoglobulin (Ig) fold. The Ig-like protein domain has been classified as a motif made up of antiparallel beta strands arranged to form two stacked beta sheets (which are often linked with a disulphide bond). This structure lends itself well to

modular fusions as the N and C termini of the motif are at opposite ends and can easily be daisy-chained together (Bork, Holm and Sander, 1994). Multiple evolutionary gene duplication events have led to large loci of related genes encoding Ig-like folds as a common component of immune systems (Antonacci *et al.*, 2020), with subgroups of related gene sequences that have diverged to provide functional sequence diversity whilst maintaining the modular Ig-like architecture. As seen in Figure 1.3, the overall architecture of the four domains of the TCR is encoded by multiple different genes, with the J gene encoding one of the strands that makes up the full Ig-like fold of the V domains whilst the rest of the domain is encoded by a V gene. These genes are polymorphic in the genome, and different V and J genes can be assembled randomly and still create a functional Ig-like fold for both alpha and beta variable domains. The nature of this assembly of gene segments requires a process called VDJ recombination, a somatic rearrangement of genes unique to immune system receptors.

### **1.3.2 VDJ recombination**

The gene loci for T cell receptors evolved from multiple duplication events, creating a cluster of related genes for each of the alpha and beta V, D and J segments (TRAV/TRBV, TRBD and TRAJ/TRBJ) that are randomly recombined then joined to constant domains (TRAC/TRBC) to create a full-length alpha or beta chain. The cleavage and joining of each segment is an imperfect process that introduces more diversity through random addition and subtraction of nucleotides. This junctional diversity combined with the recombinational diversity from different V and J gene usage could theoretically create between  $10^{15}$  and  $10^{20}$  unique clonotypes, orders of magnitude greater than the total number of cells in the human body. The actual diversity of a human T cell repertoire can vary depending on many factors such as age and prior infections/immune exposure, and some sequences dominate both in individual repertoires and between unrelated people because they are statistically more likely to occur than sequences which require multiple insertion/deletion events (Laydon, Bangham and Asquith, 2015).



*Figure 1.4 Diagram of VDJ recombination. A) shows the process of recombination events that occur in the beta (*trb*) and alpha (*tra*) loci during thymic selection of T cells. First the beta locus rearranges to join either the D $\beta$ 1 gene segment (*TRBD1*) to one of six J $\beta$ 1 (*TRBJ*) segments or to join the D $\beta$ 2 (*TRB2*) gene segment to one of seven J $\beta$ 2 (*TRBJ*) segments. This combined segment is then joined to one of the 52 V $\beta$  segments (*TRBV*) to make fully rearranged DNA. During transcription this segment is spliced to C $\beta$ 1 or C $\beta$ 2 to make up fully rearranged mRNA. For the alpha chain the process is similar but without the D segment; one of the V $\alpha$  (*TRAV*) genes is joined to one of the J $\alpha$  (*TRAJ*) segments. B) shows the fully rearranged alpha and beta mRNA, and C) is a schematic representation of the full TCR structure, using the same colour scheme as in A) and B). The V domain is made up of regions encoded by V, J and D genes. (Adapted from De Simone, Rossetti and Pagani (2018))*

Figure 1.4 shows a simplified schematic of the TCR VDJ recombination event, and where the regions encoded by different genes are positioned upon the final molecule. The regions of highest diversity are created by additions and deletions at gene segment junctions that occur in the region encoding CDR3 loops, which make the most direct contacts to the peptide. This matches the binding mode discussed above in section 1.2.1 where the CDR1 and CDR2 loops typically make contacts to the more conserved MHC helices, whilst the peptide sequences presented by the different MHC subtypes are highly diverse. In comparison, antibodies have evolved to bind much more sequence and structurally varied antigens; the presence of somatic hypermutation during affinity maturation of antibodies introduces extra sequence diversity (on top of that generated by V-(D)-J recombination) across all the CDR loops and also the non-CDR framework, which allows for optimisation of binding to more structurally-varied target antigens than the TCR:pMHC surface.

### **1.3.3 Overview of T cell receptor structure**

In the above sections I discussed how  $\alpha\beta$  T cell receptors vary in sequence between different T cell clones and bind to their cognate peptide-MHC (pMHC) with six sequence diverse loop regions, creating a highly specific binding interface between these complementarity determining regions (CDRs) and the pMHC. However, TCRs have a high degree of sequence diversity that is not restricted to the pMHC contacting CDR loops. This is created by random combining of germline encoded variable (V), joining (J) and (for beta chains) diversity (D) genes during the development of the T cell to create the full-length heterodimeric protein. The germline genes vary in amino acid sequence outside the regions that encode CDR loops, and the interface between  $V\alpha$  and  $V\beta$  domains of the TCR is made up of interactions between V and J gene encoded regions that will be different for each combination of chains produced. V-(D)-J recombination not only produces a wide variety of amino acids and chemical properties at the CDR loops for pMHC contacts, but also results in a wide range of interface angles (as the alpha-beta interface is also sequence diverse), a diverse network of inter- and intra-domain interactions around the CDR loops and overall variation in the stability of the TCR due to amino acid variance in the framework.

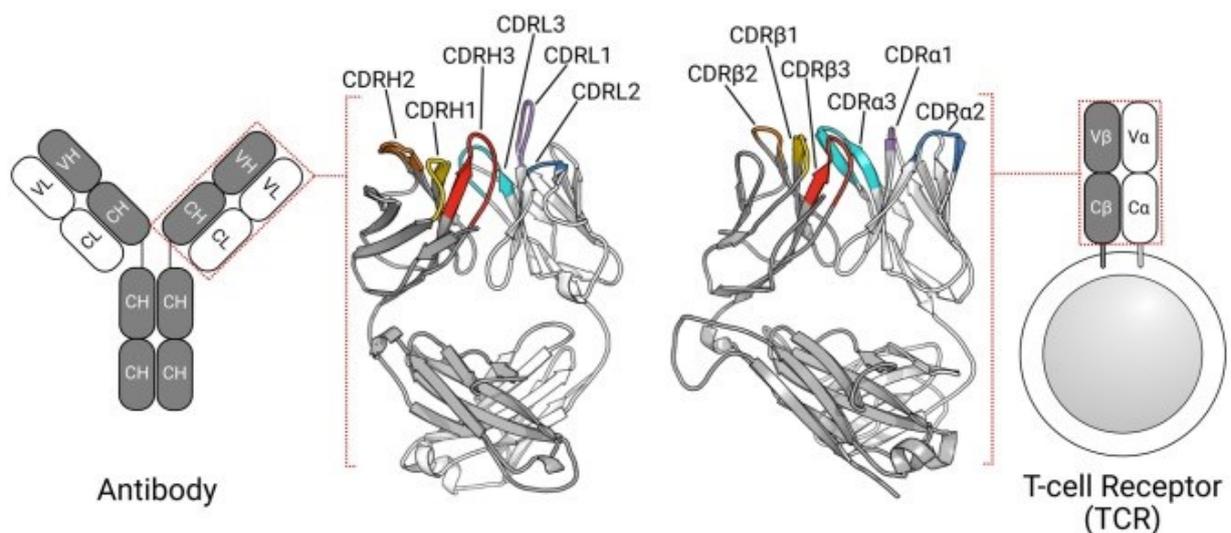
Structural and sequence diversity of TCRs is important not only for the antigen recognition within the context of a T cell, but also for the pharmaceutical applications for TCRs. The evolved specificity for binding peptides presented by MHC makes TCRs an attractive scaffold for targeting specific peptide markers of disease, but protein drugs require solubility and stability characteristics that the native membrane-bound TCR does not possess. I will address the difficulties in TCR stability in later sections. TCR engineering to overcome these issues must also overcome the variation in stability seen across different molecules in order to improve development of this class of protein-based therapeutics; previous

engineering strategies to overcome this poor intrinsic stability are discussed in more detail in later sections.

### 1.3.4 T cell receptor structure compared to antibodies

T cells and B cells have complementary roles in the adaptive immune system, and the gene loci for both TCRs and antibodies likely evolved from a common progenitor system (Flajnik and Kasahara, 2010). Both TCR and antibody structures are built up from immunoglobulin folds and the Fab (fragment of antigen binding) portion of antibodies and the extracellular portion of TCRs share a common architecture, as seen in Figure 1.5.

Antibodies are a class of molecule that evolved to function as both a membrane-bound receptor (BCR) and also as a freely soluble protein, a role that contrasts with that of TCRs which are only ever naturally present on the surface of T cells in the presence of coreceptor proteins. This may have implications for their overall inherent stability, as soluble antibodies in the bloodstream will require greater resistance to degradation and aggregation in order to remain functional than a cell surface protein that undergoes constant cycling between surface and endoplasmic reticulum (and therefore can be replaced if it unfolds or degrades).



*Figure 1.5 Comparison of antibody and T cell receptor structures. Schematic images show the antibody as a soluble IgG molecule, with the Fab heterodimer dimer portion (VH CH chain and VL CL chain, outlined in red box) shown as a*

*cartoon of the 3D structure. The T cell receptor schematic shows the TCR bound to the membrane of a T cell, with a 3D structure of the extracellular portion of the Va Ca Vβ Cβ heterodimer. The CDR loops in both structures are labelled and coloured as follows; CDRH/β1 yellow, CDRH/β2 orange, CDRH/β3 red, CDRL/α1 purple, CDRL/α2 blue, CDRL/α3 turquoise. (Adapted from Wong, Leem and Deane (2019))*

Although the overall architecture of TCRs and the antibody Fab portion is similar, there are important differences particularly in the antigen binding CDR loops.

Comparisons of crystal structures for antibodies and TCRs showed that there was a notable difference in the loop length of corresponding heavy chain/beta-chain and light chain/alpha chain CDRs (Wong, Leem and Deane, 2019). In particular, antibody CDRL3 loops studied were much shorter than TCR CDR3α (most were 9 amino acids long compared to 10-15 amino acids for TCRs), whilst previous analysis has noted that antibody CDRH3 loops are much longer than TCRβ3 (Rock *et al.*, 1994), giving them a more balanced loop length profile between the alpha and beta chains. As well as differing in overall CDR length, TCRs and antibodies did not show any convergence onto similar CDR loop conformations (so called “canonical forms” observed in multiple different structures (Nowak *et al.*, 2016)) to those forms adopted by antibody loops with similar sequence motifs. It is particularly interesting to note that TCRs appeared to possess more flexible loops than antibodies, as multiple different CDRs of identical sequence had been crystallised in different loop conformations; there appeared to be less sequence-encoded restriction on CDR conformation than that observed for antibodies (Wong, Leem and Deane, 2019). This variation in loop forms was hypothesised to reflect the different nature of the antigens they bind; the flat surface of a pMHC molecule compared to the more conformationally diverse antigens recognised by antibodies. I will discuss the implications of conformational flexibility on binding in more detail later in this introduction.

Another important difference between the two proteins is the inter-domain angles adopted by TCRs and antibodies. A systematic comparison of these angles was carried out by Dunbar *et al.* (2014), in which the main differences between TCRs and antibodies was found to be a twist between the two V domains where TCRs had a

wider angle between  $V\alpha$  and  $V\beta$  than the equivalent angle in antibody VH and VL domains. This may better accommodate the longer CDR3 $\alpha$  in TCRs, whereas the shorter CDRL3 allows for the VL domain to twist closer to the VH domain.

Antibodies which have been selected or engineered to bind pMHC tended to adopt a more TCR-like VH:VL interface angle. A particularly interesting observation from this paper was that although antibody VH:VL interdomain angles could be clustered based on the specific amino acid present at a single interface position (IMGT position 50 on the light chain), an equivalent position for TCRs could not be determined. This may be a consequence of a TCR interface that is made up of many relatively weak contacts, or it could imply greater flexibility around the  $V\alpha$   $V\beta$  interface that is not fully captured by static crystal structures. I will discuss this interface in more detail elsewhere in the introduction.

These differences in structure are useful to note, as subtle differences in domain angles or in flexibility may be responsible for the notably weaker affinity of TCRs than antibodies. However, the overall similarity between the antibodies and TCRs is a useful observation, as antibodies are a well established therapeutic platform. Engineering techniques developed in order to improve antibody function may be applicable to TCR engineering; I will discuss this in more detail in subsequent sections.

#### *1.3.4.1 Role of flexibility in pMHC binding*

The extent to which conformational flexibility is beneficial or detrimental to forming specific peptide-MHC contacts is an open question; Scott *et al.* (2011) argue that the more rigid CDR loops in the DMF5 TCR drive cross reactivity by allowing strong MHC contacts to occur in a peptide independent manner, whilst the hyperflexible loops of the A6 TCR (as measured by fluorescence anisotropy) allow more specific induced-fit binding. In contrast Stadinski *et al.* (2014) present a TCR in which binding affinity and specificity is linked to stabilising the CDR3 in the optimum binding configuration. Analysis of the murine J809.B5 TCR in complex with MHC class II IAb-3k revealed a germline encoded tyrosine on the alpha chain ( $\alpha$ CDR1

Y31) that interacts with the backbone of the bCDR3 loop. An alanine mutant at this position (Y31A) was much more cross reactive; T-cell hybridomas expressing Y31A became reactive to splenocytes presenting self-peptides, and Y31A bound to non-target peptides identified from a randomised peptide MHCII library with similar KD to the target. Loss of this CDR3 stabilising interaction was also detrimental to target binding, with a 7.5-fold increase in off-rate. Crystal structures confirmed that the Y31A mutation does not change  $\beta$ CDR3 conformation when bound to the target pMHC, but the loss of the stabilising interaction was hypothesised to drive cross-reactivity by permitting greater  $\beta$ CDR3 flexibility. It is possible that flexibility or rigidity of loops is a property unique to individual TCR molecules; molecular dynamics modelling of two different TCRs (A6 and DMF5) suggested different amounts of motion occur in unbound DMF5 compared with the more flexible A6 CDRs, with the difference most pronounced for the hypervariable CDR3s (Ayres *et al.*, 2016). The A6 TCR has also been captured as a crystal structure with a different angle at the interface between V $\alpha$  and V $\beta$  when bound to an artificial peptide ligand compared to its natural target peptide (Gagnon *et al.*, 2006), which suggested that flexibility at the interface may allow for greater degeneracy of binding targets by accommodating a wider range of docking footprints and CDR positions.

Sequence diversity results in differing degrees of conformational flexibility both in the overall interface and pivot angle between the TCR alpha and beta chains, and also in rigidity of the CDR loops themselves. This flexibility may have advantages for binding in the context of T cell function, where low affinity interactions with multiple different peptides allows the TCR to recognise many different peptides and therefore cover a wider range of possible antigens.

In the analogous antibody system, the introduction of affinity enhancing mutations during B cell maturation (via somatic hypermutation) tends to reduce the flexibility of the structure in order to enhance affinity through minimising entropic cost. Even when antibodies target a flexible glycan loop (on the viral envelope of HIV-1), SHM-induced affinity enhancing modifications occurred in residues that did not directly

contact the antigen but that were shown to reduce conformational flexibility as assessed by hydrogen/deuterium exchange (Davenport *et al.*, 2016).

### **1.3.5 Amino acid diversity exists across the entire variable domain of TCRs**

TCRs have more variability at the V gene sequence level compared to antibodies (Actor, 2012), potentially to compensate for the lack of somatic hypermutation which adds greater diversity to antibody repertoires through multiple rounds of single base alterations by activation-induced cytidine deaminase (AID) followed by clonal selection to enrich clones that bind to the antigen with better affinity.

As discussed above, for TCRs the CDR1 and CDR2 loops are encoded by germline V genes with no further mutations introduced, which may reflect less structural and sequence variation in the MHC alpha helices they bind compared with the peptides presented in MHC. As the TCR locus underwent duplication and divergence during its evolution, there would have been selective pressure for a wide range of germline-encoded loops to match the polymorphism and polygenic nature of the MHC and increase the repertoire of pMHC contacts possible by TCRs. On top of these diverse germline encoded loops, V-D-J recombination introduces changes at the regions where these segments join to maximise diversity at the peptide-contacting CDR3 loop. However, the families of TRAV and TRBV genes are highly diverse across the full length of the gene, not just at the CDR regions. This may be merely a function of evolutionary divergence between the V genes, with many mutations introduced that were not deleterious to the overall protein function and therefore had no selective disadvantage. However, the role of the framework (non-CDR) regions of the TCR in allowing binding and signalling would imply that there is a functional reason behind the overall high levels of diversity. For example, as discussed above, TCRs have been observed with a wide range of  $V\alpha:V\beta$  interface angles, which may allow accommodation of different pMHC ligands. These interface angles are not set by a

few key residues as in antibodies, but instead multiple different interface amino acids are involved in determining the interface angles (Dunbar *et al.*, 2014). Diversity at the interface can have consequences for both TCR binding and for TCR signalling, if the mechanosensor model involving changes in V $\alpha$ :V $\beta$  interface dynamics results in altered signalling properties depending on the specific interface for different TCRs. More generally, the diversity across the entire structure of the TCR has significant implications for biotechnological applications of TCRs. Variation in V and J gene sequences produces TCRs with different protein sequences and therefore different stabilities and affinities, which will require different extents of engineering work in order to make them viable therapeutic agents.

### **1.3.6 TCR interdomain interface**

Figure 3.6 also showed that the V $\alpha$ :V $\beta$  interface is made up of amino acids encoded by both V and J genes; somatic recombination gives rise to both diversity at the pHLA binding interface but also the interdomain interface. The presence of amino acid diversity at the interdomain interface is particularly interesting as TCRs have been observed to possess a wide range of interdomain angles (Dunbar *et al.*, 2014; Hoffmann, Krackhardt and Antes, 2015). Hoffmann, Krackhardt and Antes (2015) compared the relative angle of V $\alpha$ :V $\beta$  domains in a panel of 85 different structures and found a difference of up to 30°. The superposition of all the V $\alpha$  domains allowed identification of the “centre of rotation” (the pivot point on which all V $\alpha$ :V $\beta$  angles varied) as the highly conserved H-bond between  $\alpha$ Q44 and  $\beta$ Q44 (IMGT numbering). In particular they highlighted the example of the murine 2C TCR, which has been crystallised in two different conformations: a “wild type” interdomain angle when binding to the H2-K1b MHC and a different angle when the mutated T7 version of the 2C TCR was crystallised bound to H2-Ld. This variation in interdomain angle may be due to the different pMHC target bound by the TCR, but may be more likely due to the multiple mutations in the T7 variant; no other TCRs with structures bound to different ligands were shown to adopt significantly different angles. Interestingly the mutations in the T7 variant include two at the V $\alpha$ :V $\beta$

surface ( $\alpha$ L50P,  $\beta$ G49E) and one at the V $\alpha$ :C $\alpha$  interface ( $\alpha$ W96R), all of which may disrupt the native interdomain angles. Deep mutational scanning of an antibody has identified a single amino acid mutation at the variable:constant interface that improved both overall protein thermal stability and affinity for the antigen through a mechanism that shifted the interdomain angle between the variable domains of the light and heavy chains (Koenig *et al.*, 2017), indicating that even mutations distant from the interface can have allosteric effects that alter interdomain angles.

### **1.3.7 Preferential pairing between V regions**

Deep sequencing analysis of the repertoire of TCR sequences expressed in T cells after thymic selection does not show any overall bias in pairing between different genes (Howie *et al.*, 2015; Carter *et al.*, 2019); as a modular system it seems at all the V and J genes present in the immune loci can be rearranged together to form a full four-domain TCR without constraint.

The existence of preferential pairings between V gene for antibodies has also been proposed. Small studies of ~2000 known antibody sequences suggested some preferences for certain germline VH:VL pairings (Jayaram, Bhowmick and Martin, 2012) but larger single cell sequencing studies of ~170000 naïve and mature B-cells have not identified any statistically significant bias in V gene pairings (DeKosky *et al.*, 2016).

Paired repertoire sequencing is more limited in sampling depth than bulk single chain RNA sequencing, due to the need to either manipulate single T-cells directly into a high-throughput microfluidics based sequencing system (Spindler *et al.*, 2020) or carry out a combinatorial analysis of parallel alpha and beta bulk sequencing (Howie *et al.*, 2015); it is therefore possible that insufficient numbers of pairs of V (and J) genes have been sampled to statistically determine more subtle overall pairing preferences. It is also relevant to note that high throughput comparison of TCR surface expression levels with V and J gene pairing has not yet been established;

although there may be no detectable pairing bias based on mRNA expression, some pairs may still be better represented as functional cell-surface TCRs than others.

The only significant patterns in paired sequencing data would indicate that TRAV and TRBV pairings are correlated with the T-cell lineage (ie CD8+ or CD4+, which bind MHC class I and II respectively) implying that instead of certain germline CDRs being better suited for making contacts to different MHC classes, instead different pairings create an interface that better creates the optimal binding footprint for different MHC classes. Generally, it seems that the ability to form an interface between alpha and beta chains is a common feature shared amongst all the diverse V and J gene sequences encoded in the genome (Shcherbinin, Belousov and Shugay, 2020).

However, some TCRs show clear dominant expression at the cell surface compared to others (Heemskerk *et al.*, 2007; Motozono *et al.*, 2015), a trait also seen for TCRs which have been engineered for greater stability. This suggests that although all pairings of TRAV and TRBV genes can form a TCR which can bind and be expressed on the cell surface sufficiently to pass through thymic selection, there is a large variation in the stability and surface expression levels of different membrane-bound TCRs based on diversity in their amino acid sequence. Analysis of TCR pairing in the immune repertoire is based on the sequencing of RNA transcripts and therefore may not detect hypothetical bias in the actual surface level presentation of different combinations of V $\alpha$  and V $\beta$  sequences.

## **1.4 Use of T cell receptors as therapeutic agents**

The high specificity of T cells for different targets has attracted significant commercial interest as therapeutic agents, and many different approaches have been used to harness them.

### **1.4.1 Adoptive T cell therapies**

One class of therapeutics involves the use of T cells as a cell based therapy, an approach which can take many forms. In patients with solid tumours, isolation of tumour-infiltrating lymphocytes from biopsies can allow identification of endogenous T cells which recognise cancer markers and are capable of attacking cancerous cells. By simply expanding these cells *in vitro* and then reinfusing them into the patient, endogenous immune responses can be amplified. This technique can also identify TCRs capable of recognising neoantigens or other important cancer markers (Deniger *et al.*, 2018). Such TCRs are a good candidate for adoptive T cell therapy approaches, where a non-native TCR is introduced into patient T cells in order to redirect them towards a disease-related peptide (Goebeler and Bargou, 2020). These adoptive therapies can involve highly engineered TCR molecules which have been modified to improve affinity and specificity to a target of interest.

A more involved engineering approach is the use of chimeric antigen receptor T cells (CAR-T) as therapeutic agents, which is based upon replacing the extracellular T cell receptor with an antibody engineered to bind a particular target, therefore allowing redirection of the T cell killing to any cell surface marker an antibody can recognise. Antibody engineering techniques for affinity and specificity are more well established than those for TCRs, and the antibody based extracellular targeting domain is not restricted to binding pMHC but instead can target a wider range of cell surface markers.

Although therapeutic use of T cells is a highly promising area of drug development, there are some limitations to this approach; it typically requires patient-specific T cells to be isolated and expanded for each person being treated, and in the absence of additional engineering to introduce “suicide genes” there is no way to titrate the drug after administration in response to toxicity issues. The variation in TCR sequence and structure discussed above can produce TCRs that require further engineering in order to be presented on the surface at sufficient density for signalling, and there can be risks of mis-pairing between chains from introduced TCRs and those

endogenously expressed by patient T cells (strategies to avoid this will be discussed in subsequent sections).

T-cell receptors evolved to recognise MHC-presented peptides at very low levels with high specificity, making them a promising scaffold for a number of therapeutic strategies that rely on targeting disease-associated proteins; alongside their native T cell surface expression, there are alternative modes such as soluble drugs where TCRs can have significant therapeutic potential.

### **1.4.2 TCRs as part of a soluble protein therapeutic agent**

Many protein-based drugs are designed as a modular system, where one component of the molecule binds a target on a cell of therapeutic interest and another component delivers a therapeutic effect; examples may be a cytotoxic drug, a cytokine that redirects immune activity, or a second targeting arm that recruits T cells to the target cell for killing. The advantage of this modularity in bispecific proteins is that the subcomponents can be engineered separately, allowing for established effector functions to be combined with new targeting subcomponents (and vice versa) without requiring a novel platform to be developed each time.

Recruitment of T cells to a target using an antibody specific to a T-cell surface protein allows for redirection of the immune system towards novel target cells without requiring T cell engineering; instead, the cytotoxic T cells are brought into close enough contact with the target to form an immune synapse without the need for TCR-binding and activation by the double-ended nature of a drug bispecific to both target and T-cell surface proteins (Goebeler and Bargou, 2020). This approach has been trialled in multiple different antibody formats but can be limited by the difficulty in identifying highly expressed disease-specific surface antigens. Attempts have also been made to create antibodies that bind disease-specific pMHC in a TCR-like manner (reviewed by Yixiang Xu *et al.* (2019)). However, the binding of antibodies to pMHC targets has been shown to be less peptide-focused than that of

TCRs and can be more cross reactive (Holland et al., 2020). Instead, solubilised T cell receptors can be used as the targeting arm of a bispecific protein, linked to a T-cell recruiting antibody that acts as the effector function for the drug. ImmTAC molecules (Oates and Jakobsen, 2013) use the full length TCR including constant domains stabilised by an introduced disulphide bond as the targeting domain of a bispecific, whereas the T cell engaging receptor (TCER) molecule produced by Immatics relies on a single chain V $\alpha$  V $\beta$  construct (Goebeler and Bargou, 2020).

TCRs tend to bind with much weaker affinities than antibodies (dissociation constant in the micromolar range compared to nanomolar) and therefore require significant affinity enhancement in order to act in an antibody-like manner as a targeting module. Some approaches to affinity engineering are reviewed in more detail by Robinson *et al.* (2021); briefly, phage display can produce TCRs with picomolar affinity through mutations at the CDR loops that improve contacts to the pMHC surface (Li *et al.*, 2005). Alternative rational and computationally aided engineering approaches have also been successfully applied to enhance the weak native affinity of TCRs. Haidar *et al.* (2009) built an iterative structurally informed model to score and predict affinity enhancing mutations, with the best of the single point mutations improving affinity 6 fold. Zoete *et al.* (2013) used molecular dynamics simulations to predict mutations *in silico* that would enhance affinity of a NY-ESO specific TCR (using a structure model based on the crystal structure of a TCR which differed by only four amino acids), with 13 out of the 24 predicted point mutations shown to have enhanced affinity for the pMHC target when expressed and tested using ELISA titration.

## **1.5 Design requirements for protein-based drugs**

### **1.5.1 Developability of protein therapeutic agents**

The pathway to turn a promising candidate into a successfully produced therapeutic agent is long and has a high attrition rate – approximately 1 in 10 therapeutics fail even at the final stages between first-in-human dosing and clinical approval (Kola and Landis, 2004). Optimising the manufacturing and storage of therapeutics can be as difficult a challenge as the identification of a viable drug target or of a candidate which binds to the target with sufficient specificity. Developability is an umbrella term for many therapeutically and commercially relevant drug characteristics; aggregation propensity, as large aggregates can induce an adverse immune response (Rosenberg, 2006); degradation and modification during storage, which alters the biophysical properties of the drug; yield, which affects manufacturing costs.

Recent reviews (Jain *et al.*, 2017; Yingda Xu *et al.*, 2019; Bailly *et al.*, 2020) have encouraged the use of constant screening for long term stability and aggregation potential throughout the pipeline of developing protein-based therapeutics. Historically screening for biologic drugs focused on measuring and improving affinity and specificity for the target of interest, with assessment of production and storage characteristics only occurring at a much later stage in the pipeline. The development of higher throughput assays for stability allows for much earlier identification of molecules which will be overly expensive to produce in commercially relevant quantities or which will have a debilitatingly short shelf-life due to instability. Different methods for assessing stability are outlined in Table 1.1.

*Table 1.1 Overview of different stability measurements and experimental techniques commonly used in pharmaceutical settings.*

Measure of stability	Techniques
Melting temperature ( $T_m$ )	Differential scanning calorimetry (DSC) Differential scanning fluorimetry (DSF and nanoDSF) Circular dichroism (CD)

Large aggregate formation	Dynamic light scattering (DLS) Gel electrophoresis
Self-association	Surface plasmon resonance (SPR) Affinity-capture self-interaction nanoparticle spectroscopy (AC-SINS)
Activity retention in storage	Forced degradation Functional activity measurements (SPR, ELISA, other assays)
Post translational modification	Mass spectrometry peptide mapping Ion Exchange/ Capillary isoelectric focusing (cIEF) for charge variation

Protein stability in therapeutic drug contexts is a multifaceted description, covering both physical and chemical stability of the molecule. Physical stability properties include the conformational stability of a protein; the Gibbs free energy difference between the folded and unfolded states, which exist in a state of dynamic equilibrium as the protein is in constant motion in solution.

#### *1.5.1.1 Conformational stability*

Conformational stability can be enhanced by either stabilising the fully folded state or destabilising the unfolded state in order to shift the equilibrium between the two states towards fully folded and functional conformation (Kuroda and Tsumoto, 2020). Another highly relevant consideration is colloidal stability; the propensity of a protein to clump together and form aggregates. Hydrophobic surface patches, or hydrophobic regions of the protein interior exposed from full or partial unfolding events, are thought to drive irreversible assembly into larger oligomeric states (Meric, Robinson and Roberts, 2017). This colloidal instability is problematic in therapeutic molecules as aggregation disrupts the native fold and therefore function

of the protein, as well as potentially exposing immunogenic regions of the therapeutic that would normally be buried within the folded core (Rosenberg, 2006). It also presents a manufacturing issue in lowering yield of functional drug as aggregates will form both during protein purification and also during soluble expression in recombinant cell systems, as well as requiring more stringent storage conditions in order to minimise risk of protein aggregation before the drug can be administered.

#### *1.5.1.2 Chemical stability*

The chemical stability of proteins is another important factor for biotherapeutics; amino acid side chains are vulnerable to degradation and modification, such as deamidation of asparagine side chains, aspartate isomerization, oxidation of methionines and tryptophans, and glycation of lysines. These modifications can be problematic in multiple contexts; they may introduce an immunogenicity risk, or alter the protein surface properties (such as charge) in a way that increases physical instability and aggregation risk. Most importantly, post-translational modifications can occur to amino acids at or near functional binding sites and interrupt therapeutic function. Chemical stability can be assessed by peptide mapping using liquid chromatography mass spectrometry in order to experimentally determine which side chains are undergoing post translational modifications, and how abundant these modifications are in a protein sample after different storage conditions. However, this is a sample- and labour- intensive analysis approach; more commonly, computational methods are used to predict potential risk residues based on sequence features, structural predictions of solvent exposure, and past experimental data of degradation hotspots. Amino acids with high predicted or observed risk of chemical modification can be targeted for site-directed mutagenesis to replace the side chain (if this does not alter binding functionality) or adjacent amino acids in order to reduce the likelihood of modification. If mutagenesis is not a viable strategy due to location and function of the risk residues, post translational modifications that occur during manufacturing and storage can be minimised through use of different buffer

conditions and additives; this however does not address the risk of modifications occurring *in vivo*, so engineering approaches are generally favoured (Gupta *et al.*, 2021). Chemical and physical stability of proteins are interlinked; amino acid post translational modifications can alter the conformational stability of a protein or introduce an aggregation prone region (Gupta *et al.*, 2021), and low conformational stability can expose risk residues that would normally be buried and inaccessible.

### 1.5.1.3 Stability assessment techniques

Table 1.1 summarises different approaches to assess the stability of clinically relevant proteins. The developability criteria for a protein based drug depend somewhat on the therapeutic use and storage, as those drugs which are less stable in long-term can achieve clinically required levels of sample quality by optimising buffer formulation or requiring frozen or cold chain storage and shipping.

Forced degradation assays are a particularly common approach to identify and characterise stability issues, where a protein of interest is incubated in a damaging environment (for example, extreme pH or high temperature) in order to more rapidly induce degradation events that may occur during storage or *in vivo* (Nowak *et al.*, 2017). This allows for characterisation of residues at risk for chemical modification, as well as inducing aggregation due to local protein unfolding and modifications that increase protein self-association. The results of a forced degradation study can be analysed in multiple assays depending on the particular developability issue being interrogated. A general screening approach to assess binding function after degradation (for example, using techniques such as SPR or ELISA to determine binding affinity) is often used to highlight particular proteins that exhibit significant loss of function, which can then be subject to more detailed analysis to determine the exact degradation pathway that drove instability and potentially reveal specific risk residues for subsequent mutagenesis.

Forced degradations studies are designed to reveal instabilities in a protein based drug that may occur *in vivo* as an easier and higher throughput methodology. However, *in vivo* analysis is also a vital part of stability assessments in order to

identify stability issues that occur in the stress-inducing environment of the body. A useful proxy for in vivo stability assays is the use of in vitro serum incubation as a forced degradation step.

#### *1.5.1.4 Thermal stability*

Thermally induced unfolding of proteins in order to determine melting point (midpoint transition temperature,  $T_m$ ) is often used as a proxy measure for overall protein physical stability. This can be determined through direct measurement of the changes in heat flux as a protein sample is heated compared with a buffer-only reference in order to detect phase transitions as the protein unfolds (Differential Scanning Calorimetry, or DSC). DSC analysis of proteins is typically low to medium throughput and can require high concentrations of protein, with detailed analysis of melting curves requiring scanning data generated over the course of multiple days (Johnson, 2013). Circular dichromism (CD) is another dye-free method for monitoring changes in protein phase as it unfolds, in this case indicated by changes in absorbance of polarised light as highly ordered structures fall apart on heating (Greenfield, 2006). Differential scanning fluorimetry (DSF) also involves heating a protein until it fully unfolds, but in this instance unfolding is monitored either from changes in intrinsic fluorescence of tryptophan residues as their local environment changes (Alexander *et al.*, 2014) or by using a dye probe which fluoresces in the presence of exposed hydrophobic groups as the protein core unfolds to the solvent (Gao, Oerlemans and Groves, 2020). DSF is a much higher throughput technique than DSC and CD as experiments can be carried out within an hour on a high number of samples using pre-existing filter sets for RT-PCR machines and a low sample concentration requirement (Shi *et al.*, 2013), and  $T_m$  values obtained from this method correlate well with those obtained from the “gold standard” of DSC (Lang and Cole, 2017).

$T_m$  values measure the conformational stability of a protein, but this is not an independent property to colloidal stability and aggregation propensity. Sakhnini *et al.* (2019) screened a library of antibodies (in which one to three residues had been

mutated to aspartate) for predicted conformational stability improvements and characterised a panel of 26 mutants which showed strong correlation between reduced formation of higher molecular weight aggregates and improved  $T_m$  values. Zhang *et al.* (2018) modelled mutations which would reduce overall conformational flexibility in an antibody that was already highly thermostable (72 °C, close to the highest range observed in clinical stage antibodies by Jain *et al.* (2017)). Only small gains in melting point were achieved (expected given the high starting point) but those mutations which lowered  $T_m$  also resulted in faster loss of monomers and appearance of multimeric aggregates. However, for a panel of mutations to a humanised antibody Fab  $T_m$  was only predictive of aggregation rates when the protein was incubated at 65 °C and therefore already beginning to unfold, rather than at lower temperatures (23 °C and 4 °C) more relevant for storage conditions (Chakroun *et al.*, 2016).

#### *1.5.1.5 Summary of stability considerations*

Assessing the stability of a protein-based drug requires a multifaceted approach, as stability is a broad concept. It can be used to refer to the intrinsic characteristic of a protein fold and the energy required to break existing bonds and denature the three-dimensional structure, or the more functional definitions of how prone to aggregation and degradation a protein is, or how long it can retain functional properties during storage or in deleterious conditions. All these considerations are interlinked, and different stabilisation engineering approaches can increase some or all these stability metrics, improving the developability of a protein-based drug.

### **1.5.2 Lessons from the field of antibody engineering**

Antibody based drugs are a huge and well-established field, with at least 98 therapeutic monoclonal antibodies currently approved by the United States Food and Drug Administration (US FDA) at the time of writing (The Antibody Society, accessed 12/21). As discussed above, the basic architecture of antibody Fab regions

has a high level of structural homology to TCRs, and the development of antibody-based drugs offers many insights into future strategies for harnessing T cell receptor proteins for therapeutic purposes.

Of particular interest for the scope of this thesis is antibody engineering strategies that involving structurally informed stability mutagenesis. As discussed above there is a lot of structural similarity between the Fab region of antibodies and TCRs, and some structural insights may be transferrable between the two systems.

Antibodies of clinical relevance have often undergone somatic hypermutation, a process that introduces mutations into germline encoded regions during B cell maturation (and does not occur in T cells), instead of or alongside artificial mutagenesis in order to enhance affinity for desired targets. There is sometimes a trade off between affinity enhancement and stability, as mutagenesis can introduce amino acids which destabilise the overall TCR even as they improve antigen contacts (Julian *et al.*, 2017; Rabia *et al.*, 2018). Schwaigerlehner *et al.* (2019) examined the stability of four different monoclonal antibodies and their germline-encoded parent molecule, revealing that even for the germline versions there was significant difference in thermal stability and expression yield in CHO cell lines, whilst the somatically mutated versions were not necessarily more stable. This implied that stability and manufacturability can vary widely between different molecules of the same class, matching earlier observations that identified different V genes that were more or less thermally stable when produced as isolated domains (Ewert *et al.*, 2003).

#### *1.5.2.1 Importance of inter-domain interfaces*

The Fab portion of antibodies shares a common architecture with TCRs, including a large interface between the two variable domains (section 1.3.4). This is a frequent site of mutagenesis in antibody engineering (Masuda *et al.*, 2006; Nakanishi *et al.*, 2008; McConnell *et al.*, 2013), and plays an important role in positioning the CDRs for correct high affinity binding.

As discussed above, there is little evidence to support intrinsic bias in the human immune repertoire for certain combinations of V genes over others, both for antibodies and for TCRs, indicating that V domain interfaces are largely interchangeable in the modular design of both protein classes. However, in the artificial conditions of display libraries differences in expression (especially in prokaryotic systems) and stability of the displayed molecule may drive bias towards some V genes and combinations of V genes. Ponsel *et al.* (2011) observed that VH1 and VH3 in particular were highly overrepresented and VH4 underrepresented after selection by phage display, and Lloyd *et al.* (2009) screened a large scFv library against 28 different antigens and identified strong preferential enrichment of a VH1:Vk1 pairing regardless of antigen. Interestingly, an mRNA display library of different single chain Fv format antibodies also showed bias towards VH3 and VH1 after selection for a range of different antigens but did not produce any enrichment for the VH1:Vk1 chain pairing seen in the phage display study of Lloyd *et al.* Eukaryotic yeast display libraries have been shown to exhibit lower levels of V gene usage bias and better sample the immune repertoire (Bowley *et al.*, 2007), so this may represent a feature of prokaryotic systems. No equivalent study has yet been carried out to determine if certain TCR V genes or combinations of TRAV and TRBV genes have greater or less stability when considered as soluble proteins or on an artificial display system.

More generally, the importance of the VH:VL interface for both antibody stability and affinity has been repeatedly demonstrated. Herold *et al.* (2017) carried out an alanine scan to mutate residues in the VH:VL interface, finding that many mutations particularly on the light chain were relatively well tolerated and did not disrupt the fold, but loss of a few highly conserved contact pairs between VH and VL resulted in significantly impaired association between the two domains as measured by ELISA. Warszawski *et al.* (2019) used a deep mutational scanning approach that identified mutations at the VH:VL interface with enhanced affinity for the antigen, and used these hotspot positions to generate mutations (for two unrelated antibodies) that

were modelled to improve VH:VL contacts. This approach produced mutated antibodies with enhanced affinity for their target antigens and improved mammalian expression yield, but, as there were few commonalities between the VH:VL mutations produced by this approach for two different antibodies, this study also indicated that optimal inter-domain interfaces were likely to be unique to each protein.

## **1.6 TCR stability is an engineering challenge**

### **1.6.1 Creating a stable and soluble TCR molecule**

As the T-cell receptor evolved in the context of membrane bound signalling, it initially proved problematic to study as a soluble protein due to low stability when the transmembrane domains were removed. Biophysical analysis of TCR structure and binding properties generally require the ability to produce the pHLA binding moiety in a form that is soluble in aqueous buffers and can be made in sufficient quantity and purity.

Low expression yields and high aggregation and misfolding of the protein hampered initial attempts to generate an scFV-style soluble TCR (Novotny *et al.*, 1991) where the alpha and beta domains are joined in a single chain by a linker sequence, as this approach had been successful in producing soluble antibodies capable of binding target antigen. However, the unmodified TCR required significant engineering to replace ten solvent-exposed hydrophobic amino acids (as predicted by modelling the structure based on homology to antibodies) in order to achieve acceptable solubility after expression as inclusion bodies in *E. coli* and subsequent refolding and purification. Further work on expressing TCRs as a single chain format for display on the surface of yeast (Shusta *et al.*, 2000; Richman *et al.*, 2009; Aggen *et al.*, 2011) or phage (Gunnarsen *et al.*, 2013) particles has also required identification of beneficial mutations by a process of randomised mutagenesis across the molecule and selection for changes that permit greater (or any) surface expression or for enhanced thermostability. Different successful strategies have emerged from these

processes, including the need to replace surface exposed hydrophobic residues primarily at what would normally be V $\alpha$ -C $\alpha$  or V $\beta$ -C $\beta$  interfaces, introducing mutations at the V $\alpha$ -V $\beta$  interface to strengthen connections between the two domains, surface exposed mutations at the V $\beta$  hypervariable loop 4, and alterations to the linker sequence.

All this indicates that TCRs are intrinsically unstable when expressed as a single chain linked V domain format; and that whilst this can be overcome with subsequent engineering work to introduce beneficial mutations, they are generally specific to that particular TCR due to the high sequence diversity of variable domains.

### **1.6.2 Stabilising TCR constant domains**

A more universal approach to stability engineering would therefore include all four of the extracellular domains (V $\alpha$  C $\alpha$  V $\beta$  C $\beta$ ) and target the invariant constant domains for engineering strategies to boost solubility and stability. Initial solubilisation strategies targeted replacing of the hydrophobic transmembrane regions at the C terminus of the constant domains with alternative modules to improve the pairing of the extracellular alpha and beta TCR domains through a number of strategies.

Fusing jun-fos leucine zipper domains to the C terminus of both C $\alpha$  and C $\beta$  to promote dimerization between the two chains resulted in homogenous expression of the TCR in an insect cell line, significantly improving yield compared to the same TCR expressed without this dimerization motif (Chang *et al.*, 1994). Leucine zipper fusion versions of two TCRs were successfully refolded from E. coli expressed inclusion bodies and their binding kinetics to target pHLA characterised using surface plasmon resonance and isothermal calorimetry (Willcox *et al.*, 1999). However, as this relied on fusing the leucine zipper domains to the TCR with an intrinsically unstructured linker sequence, it proved difficult to obtain crystal structures of the TCRs linked to this dimerization motif. Limited success in soluble production was also achieved by fusing antibody kappa domains to both C $\alpha$  and C $\beta$

(creating a six-domain structure) but this approach was not widely adopted (Grégoire *et al.*, 1991).

The native disulphide bond at the start of the transmembrane domains of the TCR in theory should add stability to the solubilised TCR, but including these residues disrupted folding when TCRs were expressed as inclusion bodies whilst truncation of the c terminus before this disulphide allowed for refolding of sufficient material for X-ray crystallisation study albeit at low yield with poor stability (Garboczi *et al.*, 1996). By instead introducing a novel disulphide bond between the constant domains of the TCR, Boulter *et al.* (2003) were able to refold soluble TCRs from inclusion bodies of alpha and beta chains that had sufficient stability for crystallisation studies, without the difficulties encountered from the addition of unstructured leucine zippers. This inter-chain disulphide approach was successfully adopted for higher throughput crystallisation screening (van Boxel *et al.*, 2009) and has been included in soluble TCR-based therapeutics (Oates and Jakobsen, 2013).

More recent work on TCR stability has involved a combination of computational modelling and screening of large panels of predicted mutations. Sádio *et al.* (2020) built on the existing Boulter disulphide and used a crystal structure of a TCR to algorithmically suggest sites where extra novel disulphides could be introduced, either within the C domain to stabilise the fold or at the V $\alpha$ -C $\alpha$  and V $\beta$ -C $\beta$  interfaces to strengthen interdomain interactions. However, despite the favourable modelling, introduction of these bonds increased the thermal stability of the TCR only slightly (an increase of +1-3 °C in melting point as determined by differential scanning calorimetry for multiple additional disulphides), whereas most introduced disulphides improve thermal stability by at least 5 °C (reviewed in Liu *et al.*, 2016), and improvements of up to 15 °C have been obtained by introducing two engineered disulphide bonds between domains of the human IgG1 Fc fragment (Wozniak-Knopp, Stadlmann and Rüker, 2012).

Another approach to stabilise constant domains in the absence of the engineered disulphide bond involved the computational screening of the whole C $\alpha$  and C $\beta$

domains to predict individual mutations which would stabilise the TCR. Favourably ranked mutations were then screened on a single chain TCR construct lacking the variable domains (C $\alpha$ -linker-C $\beta$ ) to identify those with increased mammalian expression titres and greater resistance to thermal denaturation. Those stabilising mutations that could be experimentally verified to improve stability as single mutations were combined into a novel constant region with five mutations compared to the wild type amino acid sequence. Although separate single mutations did not offer significant improvements alone, the overall combination of seven mutations gave a 20 °C boost to thermal stability of the C $\alpha$ -C $\beta$  single chain construct, with a  $T_m$  of 73.4 °C. However, when expressed as part of a full length four-domain TCR the best improvement these modified constant domains gained was only an 8.5 °C increase in  $T_m$ , implying that some of the stability gain from these modified constants is unique to the C $\alpha$ -C $\beta$  single chain construct.

This is not dissimilar to the studies into V $\alpha$ -V $\beta$  single chain constructs discussed above, where the loss of half of the TCR structure is a significant setback to stability and requires significant engineering work to recover. Interestingly the  $T_m$  for the C $\alpha$ -C $\beta$  format (which also included both the Boulter C $\alpha$ :C $\beta$  disulphide and the native C $\alpha$ :C $\beta$  C-terminal disulphide) and all seven mutations was 73.4 °C, whilst the best  $T_m$  value for the four-domain equivalents was between 65-69 °C. This demonstrated the importance of screening mutations on as close a construct to the final use case as possible, as in this instance it seems the optimised constant domains are destabilised somewhat when fused to their respective V domains.

It is also relevant to note that, when in the presence of the Boulter interchain disulphide (alongside the endogenous C $\alpha$ :C $\beta$  C-terminal disulphide), only two of the TCRs showed any improvement in melting point and the overall melt curves became broad and biphasic. This indicates that the stabilising effects of the constant domain mutations did not necessarily translate to all TCRs and, although distant from the introduced disulphide, incompatible with this stabilising strategy in some cases.

Similarly, introduction of additional internal disulphide bonds (both within the constant domains and between V $\alpha$  and C $\alpha$ ) resulted in a 1-3 °C increase in  $T_m$  for the A6 and DMF5 TCRs but also introduced this multi-transition melting behaviour (Sáudio *et al.*, 2020). This suggested that although the constant domains may be invariant, stabilising approaches which modify them do not always improve overall molecule stability depending on the specific V domains which make up the full length TCR. Biphasic melting curves are typically associated with multiple independent domains of a protein unfolding separately (Björk and Pol, 1992; Yoshida *et al.*, 2019) and this may reflect less interdependence between the four immunoglobulin-like folds of the TCR than previously thought. It is possible that the stabilised constant domains in these studies were more resistant to thermal denaturation but the pMHC binding V domains remain less stable, making this strategy less useful for increasing the thermal stability of protein-based drugs which rely on TCR:pMHC binding interactions.

### **1.6.3 Directed evolution for TCR stability**

In addition to the structurally informed mutagenesis discussed above, unbiased approaches have been carried out to screen for changes to TCR sequences that improve stability and display. The use of directed evolution techniques involves large numbers of amino acid mutations inserted into the protein sequence to create a library of variants, followed by a selection step that enriches the library for variants which improve the desired selection characteristic and deplete mutations which are deleterious. Variants can be created through the use of error-prone PCR techniques to randomly insert changes (although these will typically be limited to amino acid changes that can be encoded by altering a single nucleotide), or can be introduced in a more targeted approach via the use of degenerate codons (eg NNK, NNS, NNN, where N is any base, K is G or T and S is C or G) at certain positions. However, there is a limit to the number of positions that can be randomised simultaneously in libraries of this design, so alternative library designs which sample all possible mutations only as single amino acid changes can allow for a useful compromise

between mutational depth and coverage of mutational space (Harris *et al.*, 2016; Sharma and Kranz, 2018; Nisthal *et al.*, 2019). The use of randomised mutagenesis followed by selection allows for the enrichment of mutations that offer favourable characteristics that may not have been predicted from a rational approach.

Initial experiments in randomised mutagenesis identified several mutations to the 2C murine TCR in single chain V $\beta$ :linker:V $\alpha$  format that improved stability when displayed on the surface of yeast and also increased resistance to thermal degradation (Kieke *et al.*, 1999; Shusta *et al.*, 1999). A more systematic screen of a different human TCR identified a key mutation at the V $\alpha$ :V $\beta$  interface that both stabilised and improved affinity for the target pMHC (Sharma and Kranz, 2018). However, although this screening was done using the systematic deep mutational scanning approach to test all possible single mutations, not all the TCR framework was included in the library; residues modelled to occur near the surface were deliberately chosen to maximise the chances of identifying a beneficial interface mutation, but some enrichment and depletion of mutations was still seen in the “control” regions of surface exposed TCR framework that was also screened that may indicate selective advantage for mutations to solvent-exposed residues.

Phage display screening has also successfully identified stabilising mutations in the 4B2A1 murine TCR, with a hydrophobic leucine residue mutated to polar serine resulting in increased periplasmic expression yields as well as improved resistance to denaturation (Gunnarsen *et al.*, 2013).

The mutations identified from the above studies often occur at the V $\alpha$ :V $\beta$  interface, which as discussed above is highly diverse between different TCRs, or at the surface of the protein. Surface exposed residues are often implicated in poor stability of proteins as they can drive aggregation, and mutations to a surface exposed position are more likely to still produce a functional protein as it will not disrupt the more interconnected hydrophobic core (Strickler *et al.*, 2006; Raybould *et al.*, 2019).

## **1.6.4 Stabilising TCRs to increase cell surface expression**

### *1.6.4.1 Defining TCR stability*

In the context of soluble proteins, stability typically reflects the physical properties of the protein such as aggregation propensity and ability to maintain folded conformation. However, TCR stability is also often considered in the context of the full signalling complex of the T cell. Strategies that have been demonstrated to enhance the stability of soluble TCRs (such as introducing a disulphide between the C $\alpha$  and C $\beta$  domains, (Boulter *et al.*, 2003)) also increase expression of TCRs on the T cell surface (Cohen *et al.*, 2007). However, there is limited evidence to directly compare surface expression levels and conformational stability of the TCR, as other confounding factors (such as association with coreceptor proteins) may also influence surface expression.

### *1.6.4.2 T cell based therapeutics and the need for stable TCRs*

Alongside (and complementary to) the development of soluble TCR molecules as part of an antibody-like bispecific, a great deal of engineering has been carried out to optimise T cell receptors for cell-based therapeutics (recently summarised by Rath & Arber (Rath and Arber, 2020)). One stability challenge facing T-cell adoptive immunotherapies is the presence of endogenous TCRs which may either outcompete transduced receptors, or mis-pair between endogenous and transduced chains to create a hybrid receptor. Several engineering approaches have been pursued to overcome these obstacles by promoting correct dimerization of the two chains. As the novel inter-chain disulphide bond identified by Boulter *et al* successfully stabilised soluble TCRs this was also adopted for membrane bound receptors, exhibiting enhanced expression on the lymphocyte surface and reduced mispairing for TCRs engineered to contain the additional TRAC-TRBC disulphide (Cohen *et al.*, 2007; Kuball *et al.*, 2007). Swapping a charge pair of residues at the C $\alpha$ :C $\beta$  interface in both mouse & human TCRs also minimised mispairing (Voss *et al.*, 2008). The

modular nature of the four-domain TCR heterodimer also allowed for whole-domain swapping of the constant domains, to create hybrid V $\alpha$ C $\beta$ -V $\beta$ C $\alpha$  format receptors or replace alpha-beta constant domains with the more stable gamma-delta equivalents in order to prevent compatibility with endogenous chains (Bethune *et al.*, 2016). Murine T cell receptors were shown to express at higher levels on the surface of lymphocytes (Sommermeyer *et al.*, 2006; Voss *et al.*, 2006), allowing for higher avidity binding and increased antitumour activity, and Cohen *et al.* (2006) demonstrated that transplanting murine constant domains onto low avidity human TCRs was able to replicate this favourable property. Further work on maximising TCR surface expression revealed that minimal murinisation of 9 constant domain residues was sufficient to promote correct pairing of an unstable human TCR (Bialer *et al.*, 2010). Replacing positively charged residues with hydrophobics in the transmembrane domain improved surface expression & anti-tumour activity in multiple TCRs, including those already stabilised by murinised constant domains (Haga-Friedman, Horovitz-Fried and Cohen, 2012).

Improved cell surface expression levels and increased stability of soluble TCRs are often reached by similar engineering strategies, particularly those focusing on improving contacts between the alpha and beta chains. Similarly, improvements in stability can often boost avidity and affinity in soluble and membrane bound formats alike.

## **1.7 Summary of research aims**

This thesis focuses on understanding the impact of non-CDR residues on the stability of soluble alpha-beta TCRs. Interrogating the role of the non-CDR regions will help to improve the development of TCR based drugs and will also offer greater understanding into the importance of framework region diversity in endogenous TCR chain pairing and in TCR binding and recognition.

In Chapter Three, a systematic analysis of existing crystal structures is carried out to identify underlying patterns common to most TCRs. The key interactions between

the different domains of the TCR are mapped and quantified across a representative panel of structures, with these surfaces further broken down depending on the sequence diversity at each position.

Building on this analysis of structural features, a database of experimentally determined thermal stability measurements for TCRs of known structurally aligned sequences was built to allow a complementary approach which does not require crystallographic data. These data were analysed in Chapter Four to identify sequence features that correlate with high or low stability measurements and used to identify which features of the TCR are key determinants of overall stability. This analysis was also broken down based on the gene usage to investigate the importance of V and J gene variability within the T cell repertoire.

The above chapters use large and diverse datasets to identify underlying sequence and structural features based on existing TCR diversity. In Chapter Five an alternative engineering approach is explored to systematically screen all possible amino acid changes to a specific TCR, allowing for unbiased identification of residues that are key to TCR stability. A directed evolution approach was used to select for single mutations which improve resistance to thermal degradation, identifying a number of beneficial mutant variants which are then further investigated to determine their biophysical properties.

In Chapter Six the stabilising mutations identified from the work in previous chapters are tested on a wider panel of TCR-based bispecific molecules which are composed of a range of different V and J gene sequences. Previous studies to improve the stability of TCRs have either focused on alterations to the constant domains or to screening for mutations at a specific TCR level. In this chapter, the most successful V $\alpha$  and V $\beta$  mutations produce stabilising effects that translate well onto different TCR scaffolds with diverse amino acid usage, demonstrating that some stabilising mutations to V domains can be universally beneficial to protein stability.

Finally, the implications of these results for both future engineering of TCR-based therapeutics and more generally for explaining the evolutionary pressure that

produces amino acid diversity throughout the TCR variable domains are discussed in Chapter Seven.

# Chapter Two

## Materials and Methods

---

### **2.1 Protein expression and purification**

#### **2.1.1 Bacterial culture media**

Media	Ingredients
Lysogeny Broth Medium (LB)	Tryptone 10 g/l (Oxoid) Yeast Extract 5 g/l (Sigma) NaCl 10 g/l (Sigma)
TYP Medium	Tryptone 16 g/l Yeast Extract 16 g/l NaCl 5 g/l K <sub>2</sub> HPO <sub>4</sub> 1 g/l (Sigma)
2xYeast Extract Tryptone Medium (2xYT)	Tryptone 16 g/l Yeast Extract 10 g/l NaCl 5 g/l
LB Agar	LB medium + BactoAgar 16 g/l (Appleton Woods)
YTE Agar	2xYT medium + BactoAgar 16 g/l

---

#### **2.1.2 Bacterial strains used**

Strain	Genotype
--------	----------

---

---

<i>E. coli</i> DH5 alpha (NEB)	fhuA2 $\Delta$ (argF-lacZ)U169 phoA glnV44 $\Phi$ 80 $\Delta$ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17
<i>E. coli</i> TG1 electrocompetent cells (Lucigen)	F' traD36 proAB lacIqZ $\Delta$ M15] supE thi-1 $\Delta$ (lac-proAB) $\Delta$ (mcrB-hsdSM)5(rK - mK -)thi-1 $\Delta$ (lac-proAB) $\Delta$ (mcrB-hsdSM)5(rK-mK-)
<i>E. coli</i> Rosetta (DE3) pLysS (Novagen)	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE (CamR)

---

### **2.1.3 Vectors and cloning**

All alpha, beta, and beta-scFV fusion expression plasmids were based on the pGMT7 vector, plasmids and sequences kindly provided by my colleagues at Immunocore. All constant domain sequences contained the introduced mutant cysteine residues at TRAC48 and TRBC57, as described in Boulter *et al.*, (2003). Any point mutations were introduced into the relevant sequence using one-step site directed mutagenesis with the modified QuickChange protocol of Liu and Naismith, (2008).

### **2.1.4 Inclusion body expression**

Alpha and beta TCR chains were expressed separately in Rosetta DE3 *E. coli* cells (Novagen). Cells were transformed with plasmids described above and plates on LB -agar medium supplemented with 50  $\mu$ g/ml ampicillin and grown overnight at 37 °C. Single colonies were picked and used to inoculate an overnight starter culture of 10 mL LB media supplemented with 34  $\mu$ g/ml chloramphenicol and 100  $\mu$ g/ml ampicillin, grown at 30 °C with shaking at 220 rpm. Starter culture was used to inoculate 1L of TYP media supplemented with 34  $\mu$ g/ml chloramphenicol and 100  $\mu$ g/ml ampicillin, grown for 16-18 hours at 37 °C with shaking at 220 rpm until an OD<sub>600</sub> of 1.0-2.0 was reached. Expression was then induced with 0.5 mM IPTG, and

cultures incubated for a further 2 hours before harvesting of cells by centrifugation for 30 min at 11,200 xg.

For a cell pellet harvested from 1L of culture, cells were lysed using 40 mL BugBuster protein extraction reagent (Novagen #70584-4) supplemented with 0.5 M MgCl<sub>2</sub>, 20 mg/ml DNase I, followed by centrifugation for 30 min at 11,200 xg to pellet inclusion bodies. IBs were then resuspended in a Triton detergent-based buffer to remove contaminants (0.5 % Triton-X 100, 50 mM Tris pH 8.1, 100 mM NaCl, 10 mM EDTA), pelleted and resuspended a further two times in the same buffer to wash inclusion bodies, and then washed and resuspended in 50 mM Tris pH 8.1, 100 mM NaCl, 10 mM EDTA. Inclusion body size and purity was confirmed by SDS-PAGE analysis using BioRad Gel Doc and ImageLab software (version 5.2.1).

### **2.1.5 Refold and purification**

TCRs were refolded from inclusion bodies as described in (Dunn *et al.*, 2006). Briefly, for a 1L refold 1.5 µmol of alpha chain inclusion body and 1.0 µmol of beta chain (a molar excess of alpha chain has been shown to improve refold yield (Immunocore, unpublished)) were mixed with a denaturation buffer (6.0 M Guanidine hydrochloride, 50 mM Tris pH 8.1, 100 mM NaCl, 10 mM EDTA) to a final protein concentration of 5mg/ml and incubated for at least 30 minutes at 37 °C to fully denature the inclusion bodies. Denatured inclusion bodies were then added to 1L of refold buffer (4.0 M urea, 0.4 M L-Arginine, 100 mM Tris pH 8.1, 2mM EDTA) to which the redox couple was added just before refolding (1.9 mM Cystamine (oxidant), 6.5 mM Cysteamine (reductant)). Refold mixture added to dialysis tubing (Visking) and was dialysed at 4 °C against 10 L of 10 mM Tris pH 8.1 with three changes of buffer over three days before filtering through 0.45 µM filter (Whatman GE) to remove any aggregates.

Refolded proteins purified by anion exchange (PorosHQ resin, Applied Biosystems) using 20 mM Tris pH 8.1, eluted with 20 mM Tris pH 8.1 + 1 M NaCl, followed by size exclusion chromatography (Superdex S200, GE lifescience) in PBS (Sigma).

Protein purity and size were assessed by SDS-PAGE using Biorad GelDoc and ImageLab software (version 5.2.1). All proteins used for further analysis were at least 95% pure of other molecular weight species.

ImmTAC molecules were refolded and purified from inclusion bodies using a similar protocol (details of modifications to protocol redacted due to commercial sensitivity).

pMHC molecules were expressed, purified and biotinylated by members of the HLA group (Immunocore), following protocols as described in Dunn *et al.* (2006); I am very grateful for their assistance.

## **2.2 Protein analysis**

### **2.2.1 Surface Plasmon Resonance**

Affinity measurements carried out using Biacore 8k (GE Lifesciences) according to manufacturer recommended protocol. Running buffer used was PBS buffer (Sigma) supplemented with 0.005 % P20 surfactant. Approx. 50 response units of relevant biotinylated pMHC complex loaded onto CM5 chip. Five serial dilutions of TCR loaded sequentially with a flow rate of 50  $\mu$ l/min. Data fitted using BiaEvaluation software single cycle kinetics model to determine kinetic parameters.

### **2.2.2 Thermal stability measurement**

Thermal stability was assessed using differential scanning fluorimetry as described by Lo *et al.*, 2004. Differential scanning fluorimetry (DSF) is a dye based method for determining melting point, typically using a RT-PCR machine to both control the temperature gradient and measure fluorescence changes. SyproOrange dye fluoresces when bound to exposed hydrophobic groups. As protein unfolds due to increasing temperature more hydrophobic groups are exposed, leading to increased fluorescence up to a maximum signal (fully unfolded sample), the strength of which is determined by both dye and protein concentration. Aggregation of the unfolded protein leads to reduction of signal after peak fluorescence reached. The midpoint of

sigmoidal curve produced from changes in fluorescence during unfolding is assumed to represent a 1:1 ratio of folded:unfolded material, giving the melting temperature ( $T_m$ ) for the sample.  $T_m$  values obtained through this technique are comparable to those obtained from circular dichroism with a much lower sample requirement (Lavinder *et al.*, 2009).

Proteins were diluted to 0.25 mg/ml in PBS buffer (Sigma) with the addition of SYPRO Orange dye (Sigma) to a working concentration of 4x. 50  $\mu$ L total of protein-dye mix added to each well of optical grade RT-PCR plate (MicroAmp Fast 96 well plates, Thermofisher), mixed well and sealed using optical adhesive film (MicroAmp Optical Adhesive Film Sealers, Thermofisher). Plates were spun for 5 minutes at 100 g to ensure sample remained at bottom of well before transfer to RT-PCR machine.

StepOnePlus RT-PCR machine (Applied Biosystems) used to measure fluorescence with the ROX filter set (excitation and emission wavelengths 490 nm and 580 nm) over a temperature gradient from 25 to 95 °C with a step size of 1 °C/min. Melt curve data were analysed using Protein Thermal Shift Software v1.3 (Life Technologies), with the maxima of the first derivative of the melt curve taken as an estimate for  $T_m$  for each sample, as in Gao, Oerlemans and Groves, (2020). All protein samples were measured in triplicate and reported as a mean  $T_m$  value, with outlier or poor quality data (as flagged by analysis software) omitted from the mean calculations.

### **2.2.3 Serum stability measurement**

Incubation in human serum has been shown to be a useful model system to identify stability issues in monoclonal antibodies (Yang *et al.*, 2018). This method was adapted to determine the stability of ImmTAC format TCR fusion proteins based on their ability to retain binding function after serum incubation.

#### **2.2.3.1 Serum incubation**

Purified TCRs in ImmTAC format were diluted to 1  $\mu$ M in PBS (Sigma) and then mixed with human serum (BioIVT pooled gender human serum, product code

HUMANSRMP2N) to a final concentration of 50 nM protein in 500  $\mu$ L of human serum. 150  $\mu$ L of protein:serum mixture immediately frozen with dry ice and stored at -80  $^{\circ}$ C, to provide T=0 timepoint for subsequent comparison. Remaining protein:serum mixture incubated at 37  $^{\circ}$ C in static incubator for 10-14 days, with at least two aliquots taken and frozen as above at day 6-8 and day 10-14.

### 2.2.3.2 Bi-functional activity determination by ELISA

Protein activity after incubation in human serum was assessed by a sandwich ELISA (enzyme linked immunosorbent assay). The cognate pHLA for each ImmTAC was used as a capture reagent on the ELISA plate, to provide a direct capture interaction that requires the TCR portion of the ImmTAC to retain normal binding function.

Detection was carried out using a CD3 molecule conjugated to horseradish peroxidase (HRP), the binding of which requires normal binding function of the anti-CD3 scFv portion of the ImmTAC. Both binding functions must therefore be present in order to allow for colorimetric development of ELISA signal through the action of HRP on a detection reagent.

Streptavidin F96 ELISA plates (Nunc 436014) were washed three times with PBS + 0.05% Tween 20 using a BioTek 405 Microplate washer, then coated with 3  $\mu$ g/ml of biotinylated cognate pHLA and incubated at room temperature for 30 minutes with shaking at 600 rpm (Titramax 100). Plate was then washed three times with PBS + 0.05% Tween 20 then blocked with Dulbecco's PBS pH 7.4 supplemented with 1% w/v skimmed milk powder (Marvel) for a 60 minute incubation at room temperature with shaking at 600 rpm. Plate was then washed three times in PBS+Tween as above.

Protein samples from the serum incubation described above were then diluted to 150 ng/ml in Dulbecco's PBS pH 7.4 supplemented with 1% w/v skimmed milk powder, and 100  $\mu$ L of each sample added to the prepared plate. All samples were added in duplicate. Plate was sealed to prevent evaporation and incubated at room

temperature for 60 minutes with shaking at 600 rpm, followed by three wash steps in PBS+Tween as above.

CD3 conjugated to HRP was generously supplied by Pranav Bheamadu (Immunocore). It was diluted in PBS +1% milk and 100  $\mu$ L added to each well. Plate was sealed to prevent evaporation and incubated at room temperature for 60 minutes with shaking at 600 rpm, followed by three wash steps in PBS+Tween as above.

1 step Turbo TMB-ELISA colorimetric reagent (Thermo-Fisher) was used to develop ELISA plate. 100  $\mu$ L was added to each well, and sample incubated in the dark at room temperature for 20 minutes. Plate was then read using 96 well spectrophotometer; absorbance at 650 nM was recorded after 5 seconds of shaking. Mean absorbance was calculated for duplicate wells and relative loss of function was calculated by comparing signal strength of post-incubation samples to the T=0 starting timepoint sample.

## **2.3 Modelling and structural analysis**

All modelling carried out using Molecular Operating Environment (Chemical Computing Group ULC, 2019), using experimentally determined structures taken from both publicly available submissions to the protein data bank (Berman *et al.*, 2002) and from structures that have been solved in house by the Immunocore Structural Group (unpublished). I manually assembled a list of structures to maximise the number of V and J genes represented whilst removing any duplicated structures and any mutated structures where a wild type structure was available. In the absence of a wild type structure, I chose the structure with the fewest introduced mutations compared with wild type sequence.

### **2.3.1 Structure preparation**

All crystal structures were loaded into a MOE database and prepared for modelling using the default QuickPrep function, as all structures used lacked sufficient

resolution to map the hydrogen bond network and contained regions of poor resolution where the backbone could not be resolved or other crystallographic artefacts. Briefly, this workflow removed distant ( $>8\text{\AA}$  from protein) solvent atoms from the structure. N and C termini (and any termini introduced by breaks in the protein chain where regions of amino acids could not be resolved in the experimental structure) were capped either with an ACE (acetyl) or NME (formyl) residue (for residues either side of a chain break) or capped as charged termini. Empty residues preceding or following a chain terminus were automatically deleted from the structure (but sequence information retained to allow sequence-based alignments). Explicit hydrogens were then assigned to all unprotonated groups using the Pronate3D application (Labute, 2009) to identify protonated states with optimised titration free energy; hydrogen orientations are chosen so as to maximize H-bond networks and minimize the overall self-energy. The overall structure is then energy minimised using the default settings for protein domains (Amber10:EHT forcefield, Case *et al.*, 2005).

TCR, HLA and peptide ligands were then automatically identified and annotated within in each structure based on structural and sequence homology to a reference set of structures built into the MOE software. For files with multiple copies of the TCR:pMHC complex within the unit cell, the complex with the highest average resolution was retained and all other duplicate chains deleted in order to simplify analysis by creating a database of single complexes for each unique TCR:pMHC structure.

All structures were then assigned a human V gene and numbered according to the IMGT classification system (Lefranc *et al.*, 2003) based on sequence analysis, using a sequence alignment script kindly supplied by CCG. Constant domains were numbered as IMGT + 1000 (ie, starting as residue number 1001) in order to distinguish from variable domain numbering.

### **2.3.2 Protein contact analysis**

Variable and constant domains were defined based on IMGT numbering definition of sequence (as discussed above). Inter-domain contacts were identified using the MOE Protein Contacts tool with the two relevant domains selected to define the interacting atom sets. Protein contacts were defined as follows: Hydrogen bonds (Hbond), Metal, Ionic, Arene, Covalent, and Van der Waals distance interactions (Distance). These contacts were detected using the following default cutoff values: 4.5 Å distance threshold, -0.5 kcal/mol energy minimum for predicted H-bond, H-pi and metal/ion interactions. The energy was summed for multiple atoms within the same residue forming a bond, and the average distance was calculated for multiple atoms within the same residue forming a Van der Waals interaction.

I carried out this analysis on all of the 200 structures in the dataset. Results from the protein contact analysis were grouped according to residue number, gathered into a spreadsheet and manually separated into each TCR:pMHC complex.

Protein interaction maps were created based on the IMGT aligned sequences of each structure in the dataset. Only residues where more than 5% of the structures in the database formed an inter-domain interaction were reported as a contact.

### **2.3.3 Residue exposure, hydrophobicity, and diversity scoring**

I classified residues as buried or solvent exposed based on the same structural dataset described above. Relative Solvent Accessibility for each residue in each TCR structure was calculated as the accessible surface area in angstrom divided by the ideal surface area as determined from Gly\_X\_Gly triplets (Rost and Sander, 1994). Residues were classified as buried if they had a relative solvent accessibility score of less than 30% of the maximum possible for that amino acid.

Relative hydrophobicity was calculated based on the sequence, with each amino acid matched to the corresponding hydrophobicity score defined by Sweet and Eisenberg, (1983).

Residue diversity was calculated based on the IMGT repertoire of V genes defined by Lefranc *et al.* (2003). All alpha and beta V genes were aligned separately based on numbering, and the diversity score was calculated as Shannon entropy of each position in the alignment using the online Entropy Two tool provided by the Los Alamos National Laboratory HIV Sequence Database ([hiv.lanl.gov/content/sequence/ENTROPY/entropy\\_two.html](http://hiv.lanl.gov/content/sequence/ENTROPY/entropy_two.html), accessed December 2021). All entropy values were then divided by 4.32 to allow them to be presented as a percentage of the maximum theoretical diversity possible if all 20 amino acids were to be present in equal proportions.

## **2.3.4 Modelling of framework mutations**

### **2.3.4.1 Creating model structures to match phage-displayed protein sequences**

TCR structures (1A07, 4FTV, unpublished Kif-related structure) were prepared for mutation modelling experiments using the same workflow described above. In order to create a complete model structure for both the Tax-A6 and Kif-B5 TCRs some modifications needed to be made to existing crystal structures. The alpha constant domain is not resolved in PDB structure 1A07, so the structure was aligned with that of the high affinity Tax-A6 variant (4FTV). The last strand of the 4FTV V domain and the full constant domain (starting at IMGT residue number 122) was copied to replace the equivalent positions on the 1A07 structure, and this replacement C $\alpha$  domain structure was joined to that of 1A07 using the MOE Protein Builder “Join Chain” tool.

Five amino acid changes in the CDR loops were made to the Kif-related structure in order to model the correct CDR sequences for the Kif-B5 variant of the protein. These side chain alterations were introduced into the structure using the protein builder tool in MOE, and the new side chains were automatically screened based on the default rotamer library in MOE v2019.0104 to find the lowest energy rotamer.

Side chain and backbone energy minimisation was then carried out using the repacking function.

#### 2.3.4.2 Identifying stabilising mutations from a computational residue scan.

All 20 natural amino acids were modelled at each position in the V domains (defined as IMGT positions 1- 97 inclusive) of the alpha and beta chains for both the Tax-A6 and Kif-B5 structure. Only single mutations were modelled in order to minimise the combinatorial search space.

Mutations were modelled using the Residue Scan tool in MOE v2019.0104.

Mutations were carried out with the following parameters:

- System pH was set to 7 and no alternate protonation states were considered for mutations (in order to minimise search space required).
- Local environment (backbone and side chain of all residues within 4.5Å) was allowed to be repacked during conformational search for lowest energy conformation.
- Potential energy minimisation was scored with an RMS gradient of 0.5. By default, during energy minimisation all atoms in the mutated residue were permitted to move freely, neighbouring atoms within 8 Å were tethered with a weight of 10 kcal/Å and a deviation of 0.25 Å, and all other atoms in the structure were fixed in place.

Thermostability was calculated for each mutation (including mutations to the same amino acid as present in the structure) using the stability scoring function. This calculates the free energy of folding ( $\Delta G$ ) as the difference between the stability of the folded and unfolded states of the protein.

In the case of the folded state, this was calculated using the following equation, applying the ideal theoretical linear interaction energy values of 1 for  $\Delta E_{vdw}$  and 0.5 for  $\Delta E_{ele}$ .

$$\Delta G_s^{WT \rightarrow Mut} = \alpha[\Delta E_{vdw} + 0.5(\Delta E_{coul} + \Delta E_{sol})] + \beta\Delta E_{SS} + \gamma\cdot\Delta SA_{sc} + \varepsilon\cdot\Delta SA_{pol}$$

where  $\alpha$  is a scaling factor accounting for configurational entropy effects,  $\Delta E_{vdw}$  is the AMBER van der Waals interaction energy (Case *et al.*, 2005),  $\Delta E_{coul}$  is the AMBER Coulomb interaction energy,  $\Delta E_{sol}$  is the change in solvation energy calculated using a forcefield energy approach, and  $\Delta E_{SS}$  is the change in energy due to the presence of a disulphide bond (Chemical Computing Group ULC, 2019).

The  $\Delta$ thermostability of each mutation was calculated relative to the modelled WT residue (which had been energy minimised following the same workflow as each mutation in order to allow for a fair comparison).

## 2.4 Phage Display

Directed evolution of proteins is a technique where random changes are made to the DNA sequence, followed by selection of the proteins that exhibit a desired characteristic. In order to identify the nature of the change that improved a protein there must be a link between the phenotype and the modified genotype that encodes it; the protein should be displayed on some form of genotype-containing particle, such as a cell or a virus. Phage display is a technique for displaying proteins linked to a bacteriophage particle that has been developed for the display and directed evolution of peptides (Smith, 1985) and antibody proteins (McCafferty *et al.*, 1990).

I carried out directed evolution experiments using a filamentous bacteriophage (m13), a well established system for display of proteins (Rakonjac *et al.*, 2011). The m13 phage infects bacterial hosts such as *E. coli* through the F pilus and hijacks the endogenous cell machinery to amplify the viral genome and express viral proteins. This phage particle has a protein coat made up of five different proteins (pIII, pVI, pVII, pVIII and pIX); linking of an exogenous protein is typically done by insertion of the exogenous protein gene into the gene for either pIII or pVIII. As pIII has been

shown to accommodate display of larger molecules (Rakonjac *et al.*, 2011) I fused full length TCR sequences to this gene.

In order to minimise the risk of introducing multiple copies of the TCR-pIII fusion onto the same phage particle (which would have an avidity advantage in binding over monovalent display), I used the well-established phagemid and helper phage system (Rakonjac, Jovanovic and Model, 1997). The phagemid vector encoded TCR-pIII fusion protein gene under the control of a weak promoter and included both bacterial and phage origins of replication (so it could be amplified during standard E coli cloning techniques as well as amplified during phage replication) and an ampicillin resistance gene. Phage particles were created by co-infection of both phagemid and a helper phage particle (M13K07). The helper phage contained all the required genes for replication and assembly of new phage particles, including the wild-type pIII with no fusion protein attached, as well as a kanamycin resistance gene to allow for double-antibiotic selection that ensured both phagemid and helper phage were present in *E. coli* culture. Phage particles assembled from this phagemid – helper phage system mostly lack the pIII-fusion protein, but those that did express it were likely to do so monovalently. The phage origin of replication was impaired in the helper phage, so all phage particles contained the phagemid genotype to maintain the genotype-phenotype link.

In order to carry out directed evolution, a library of possible variants of the protein of interest must be assembled. I built libraries of different mutated versions of TCRs using a phagemid that encoded the full length TCR (including the exogenous disulphide bond first identified by Boulter *et al*) fused to the pIII coat protein of m13 phage.

### **2.4.1 Phage library construction**

Library design was based on the phagemid construct described in Li *et al.*, 2005. TRAV12-2 and TRBV6-5 variable regions were synthesised as randomised single position libraries by Twist Bioscience. Positions 1 – 99 (IMGT numbering) for both

alpha and beta chains were pooled (including or excluding the CDR1 and CDR2 amino acids as defined by IMGT, to give two versions of the library) and amplified using primers outside the regions of randomisation.

CDR3 and constant regions were amplified from plasmids encoding the wild type A6 TCR (including the artificial disulphide bond between TRAC and TRBC (Boulter *et al.*, 2003)). Overlap extension PCR was used to stitch CDR fragments onto the degenerate variable region, followed by restriction digest and ligation into pEX backbone. DNA purified with Machery-Nagel NucleoSpin Gel and PCR Clean-up kit according to manufacturer protocol and concentration of eluted product determined using Nanodrop ND-1000 spectrophotometer.

#### **2.4.2 Electroporation of phagemid constructs**

100-200 ng of purified DNA (as described above) transformed into TG1 electrocompetent *E. coli* (Lucigen) using standard electroporation kit (Biorad Micropulser). 40  $\mu$ L thawed *E. coli* TG1 cells mixed with 3  $\mu$ L of purified DNA and transferred to a chilled electroporation cuvette, where a 1.8 kV was applied to the cell suspension. 0.96 mL of SOC media (pre-warmed to 37 °C) was added to the cuvette to collect the transformed cells, and the sample transferred to a 1.5 mL microcentrifuge tube for a 1 hour recovery incubation at 37 °C.

Following recovery in SOC media at 37 °C, the 1 mL full volume of cell culture was spread onto large plates (Corning square BioAssay dish) of YTE agar supplemented with 100 g/ml ampicillin and 2% glucose (w/v). Plates were grown overnight at 30 °C and then the lawn of phage-infected *E. coli* was scraped into a solution of 15% (v/v) glycerol-supplemented 2xTY media to create a library stock. All bacterial library stocks were snap frozen on dry-ice and the cell stocks were stored at -80°C.

Library size was assessed from serial dilutions of the original 1 mL transformation volume onto smaller plates which were grown overnight at 30 °C and individual colonies counted by eye.

### **2.4.3 Phage preparation**

Phage was generated for panning from the libraries described above using the protocol from Dunn *et al.* (2006). 500 ml 2xYT media (EzMix Sigma Y2627) supplemented with 2% glucose and 100 µg/ml ampicillin inoculated from bacterial glycerol stocks of library. Culture grown at 37 °C with shaking to an OD600 of 0.5 before infection with KM13 helper phage at a ratio of 20:1 for 30 min at 37°C. Cells then resuspended into 500 ml 2xYT media supplemented with 100 µg/ml ampicillin and 50 µg/ml kanamycin, and harvested after overnight growth at 26 °C with shaking. Phage was precipitated from the supernatant using 20 % PEG-8000 (Sigma 5413), 2.5 M NaCl and resuspended in PBS.

### **2.4.4 Optimising selective pressure for stability**

Phagemid templates and glycerol stocks for high and low stability TCRs (B5 and B7) kindly provided by Elena Galfre (Immunocore). Separate single-sequence libraries for these two phage sequences were prepared as above from 10 mL cultures inoculated from glycerol stock and diluted to a total volume of 5 mL in Dulbecco's Phosphate Buffered Saline (Sigma).

70 µL purified phage was incubated for 30 min at a range of temperatures from 20 °C to 95 °C using a Bio-Rad T100 Thermal Cycler PCR machine. Phage samples were then centrifuged for 5 minutes at 16,200 x g in order to remove large aggregates that precipitated and would settle during centrifugation. 60 µL of the phage solution was then incubated with 60 µL PBS supplemented with 6% w/v skimmed milk powder (Marvel) at room temperature for 1 hour.

ELISA plates (Maxisorp NUNC-IMMUNO plate, Thermo Scientific) were coated with 100 µL of a 5 µg/ml solution of streptavidin (Sigma) in PBS for the pHLA binding assay, incubated for 37 °C for one hour, and then washed three times with PBS using a BioTek 405 Microplate washer. 100 µL of cognate biotinylated pHLA diluted to 5 µg/ml in PBS was then added to each well and plate incubated at room temperature for 30 minutes with shaking (600 rpm, Titramax 1000 plate shaker).

For the antibody capture assay the plates were coated in 100  $\mu\text{L}$  of a 2  $\mu\text{g}/\text{ml}$  dilution in PBS of anti-C $\alpha$  or anti-C $\beta$  antibody (Human TCR  $\alpha\text{F1}$  MAb clone 3A8 (Thermo Scientific) and Anti-TCR  $\beta$  Antibody G-11 (Santa Cruz Biotechnology)), then plate was incubated at room temperature for 30 minutes with shaking (600 rpm).

All plates were then washed twice with PBS and blocked with 370  $\mu\text{L}$  PBS supplemented with 3% w/v skimmed milk powder (Marvel) and incubated at room temperature for 1 hour. Plates were washed three times with PBS and 50  $\mu\text{L}$  of blocked phage solution added to each well and incubated for 1 hour with shaking. Plates were washed five times with PBS supplemented with 0.1% v/v TWEEN-20 detergent to fully remove unbound phage, then 100  $\mu\text{L}$  HRP/Anti M13 Monoclonal Conjugate (GE Healthcare, cat number 27-9421-01) diluted 1:5000 in PBS supplemented with 0.5 % BSA was added to each of the wells. Plates were sealed and incubated at room temperature for 1 hour with shaking at 600 rpm before being washed five times with PBS supplemented with 0.1% v/v TWEEN-20. 150  $\mu\text{L}$  of TMB Microwell Peroxidase Substrate (KPL) was added to each well to develop the colorimetric assay, and plates incubated for 10 minutes in the dark. Plate was then read using 96 well spectrophotometer; absorbance at 650 nM was recorded after 5 seconds of shaking to give signal strength for each well.

### **2.4.5 Phage panning**

Purified phage heated for 30 minutes at 55  $^{\circ}\text{C}$  in order to introduce selective pressure for enhanced thermal stability by unfolding and aggregation of less stable TCRs. This thermal challenge step was then followed by addition of blocking buffer (PBS + 3% Marvel milk powder) in order to prevent thermally-induced protein aggregates that may still have aggregation-promoting patches exposed from initiating further aggregation of TCRs.

Phage selection continued as described in Dunn et al. (2006). Briefly, the thermally challenged library was incubated with 100 nM biotinylated cognate pMHC bound to M-280 Streptavidin-coated magnetic beads (Life Technologies). Non/weakly binding

TCRs displayed on phage were removed through multiple wash steps using PBS + 0.1 % Tween20. Phage particles presenting TCR that bound to peptide (and therefore was not aggregated during the thermal challenge step) was eluted from the magnetic beads using trypsin cleavage at an introduced site in the PIII gene. Eluted phage particles were then used to infect TG1 E. coli culture grown to OD600 of 0.5, and spread on YTE-agar plates as above. This allowed rescue of phage particles expressing TCR variants that had been enriched by this panning process.

Post-selection phage libraries were purified as described above, and the purified phage was used as a template for next generation sequencing reactions.

#### **2.4.6 Next-generation sequencing of phage outputs.**

Library preparation and next generation sequencing was carried out with the assistance and guidance of Sunir Malla. Separate V $\alpha$  and V $\beta$  sequencing libraries were prepared with molecular indexing based on a method described in Turchaninova et al. 2016, with some adaptation. A primer containing unique molecular index (UMI) sequence was annealed to a region upstream of the alpha or beta sequence in the phagemid construct and single primer extension reaction was carried out with Kapa HiFi DNA polymerase (Roche Diagnostics, 07958935001). Following reaction cleanup with ExoProStar (GE Healthcare, US78210) and column purification (Macherey-Nagel, 740650), second PCR reaction was carried out with primers specific to the primer containing molecular index and a reverse primer designed to the alpha or beta constant gene. Sequencing libraries were prepared from purified PCR products (Ampure XP beads, Beckman coulter) using NebNext Ultra II DNA library prep kit (NEB, E7645S) according to manufacturer's instructions. Library QC was done with Agilent bioanalyser HS kit (Agilent biosystems, 5067-4626). Library DNA concentrations were measured with Qubit HS dsDNA kit (Life technologies, Q33230). Libraries were pooled and quantified with qPCR (Kapa biosystems, KK4824). Libraries were sequenced using Illumina V3 SBS chemistry on the MiSeq sequencer by Sunir Malla, who also carried out the subsequent data processing. Basecalling and sample demultiplexing was carried out using the MiSeq

reporter to generate fastq files. Fastq files were further processed using the MIGEC pipeline to trim barcode sequences and assemble consensus reads based on UMI. Assembled TCR sequences were aligned based on V gene reference sequences using the MiXCR pipeline (Bolotin *et al.*, 2015) and exported as assembled clonotype sequences. Assembled clonotype sequences were manually curated by myself to remove sequences identified as truncated or non functional by the MiXCR pipeline.

# Chapter Three

## Mapping structural diversity at TCR interfaces

---

### **3.1 Introduction**

The basic architecture of a TCR is a modular assembly of immunoglobulin domains, as discussed in the introduction. Our understanding of TCR structural diversity has been aided by the growth in available crystal structures, but structural reviews have tended to concentrate on the interactions between TCR and pMHC rather than the internal interactions that govern the relative positions of the four ( $V\alpha$ ,  $C\alpha$ ,  $V\beta$ ,  $C\beta$ ) domains. However, the sequence and structural variation of TCRs at these inter-domain interfaces is an area of key interest. Whole domain shifts in relative orientation between domains are likely to define the architecture of higher-order assemblies between TCRs and coreceptor proteins. The varying chemistry of amino acid side chains making these inter-domain interfaces will define the strength of docking between domain faces, which has obvious implications for TCR stability. The relative angles of alpha and beta variable domains will dictate the positioning of CDR loops, so the specificity of a T cell receptor is not only defined by the sequence of the loops but by how they are positioned relative to each other. In the analysis of Dunbar *et al.* (2014) of the relative angles between TCR domains as compared to antibodies, the main difference found was an increase in twist between the  $V\alpha$  and  $V\beta$  TCR domains compared to  $V_L V_H$  domains of most antibodies, allowing the accommodation of the generally longer  $V\alpha$  CDR3 in TCRs. Although the distributions differed between these two classes of molecule, the overall range of angles seen for TCRs was as diverse as that seen for antibodies. This implies that the TCR can occupy a wide range of interdomain angles and still maintain the geometry required for successful pHLA engagement, explaining the high sequence diversity seen in the  $V\alpha$ : $V\beta$  interface region. Variability in interface sequence and in inter-

domain angles may in fact be an evolutionary advantage for the TCR, permitting a wider range of possible pHLA binding opportunities for the CDRs.

### **3.1.1 Aims of chapter**

In this chapter I will compare existing crystal structures of T cell receptors and build a representative non-redundant panel of TCRs to cover different V gene encoded regions. Previous studies have looked at individual TCR structures, or at a panel of structures and their pMHC binding interfaces (Szeto *et al.*, 2021), but the structures of the non-CDR framework regions have not been compared in detail at the inter-domain interfaces within the TCR.

As discussed in the introduction, due to the large number of different V and J genes in the human genome, TCRs exhibit sequence diversity not just in the pMHC binding CDR loops but across the entire V domain. I will identify which positions are most and least conserved amongst the V gene repertoire, and map that sequence diversity onto structural function. I will then model possible inter-domain contacts between all the subunits of the TCR and identify which amino acid side chains are involved in mediating these contacts. This mapping of contact regions will allow me to highlight framework residues involved in inter-domain interfaces and any differences that occur between different framework sequences.

Analysis of these contact maps against the aligned sequences of all the TCRs studied will allow me to identify a structurally informed schema to characterise the role of non-CDR IMGT-aligned positions in the V domains of TCRs, which I will refer to throughout this thesis in the context of other stabilising mutations. I will also review previously published stabilising mutations in the context of my analysis, and identify if they fit patterns of improving inter-domain contacts.

## **3.2 Results**

### **3.2.1 Structural analysis of a representative subset of TCRs**

### 3.2.1.1 Assembling a structural dataset

In order to interrogate these inter-domain interfaces in greater detail, I assembled a database of pHLA-bound human  $\alpha\beta$  TCR structures from both publicly available structures from the Protein Data Bank (Berman *et al.*, 2002) and from internally obtained structures from Immunocore (unpublished). All structures represent a unique TCR:pMHC complex and were chosen to maximise the number of different V and J genes covered within the dataset. (see Table 3.1 for a list of these genes and number of each present in the database). However, no crystal structures were available for some V or J genes in either published or unpublished structures (see Table 3.2 for a list of the sequences not covered in this study).

*Table 3.1. List of V and J genes represented by at least one structure in the dataset. Gene usage identified from alignment of amino acids against the IMGT reference sequences. Numbers refer to unique TCR sequences selected for the dataset, and consist of both published and unpublished structures.*

TRAV genes	No. of structures	TRAJ genes	No. of structures	TRBV genes	No. of structures	TRBJ genes	No. of structures
TRAV1-2	3	TRAJ4	2	TRBV2	2	TRBJ1-1	18
TRAV3	6	TRAJ5	7	TRBV4-1	1	TRBJ1-2	29
TRAV4	6	TRAJ6	7	TRBV4-2	1	TRBJ1-3	4
TRAV5	3	TRAJ7	2	TRBV4-3	1	TRBJ1-4	3
TRAV8-2	2	TRAJ8	1	TRBV5-1	9	TRBJ1-5	16
TRAV8-3	6	TRAJ9	1	TRBV5-4	1	TRBJ1-6	8
TRAV8-4	5	TRAJ1 1	3	TRBV5-5	2	TRBJ2-1	36
TRAV9-2	7	TRAJ1 2	7	TRBV5-6	4	TRBJ2-2	24

TRAV10	1	TRAJ1 3	8	TRBV5-8	2	TRBJ2-3	16
TRAV12-1	1	TRAJ1 5	3	TRBV6-1	14	TRBJ2-4	2
TRAV12-2	55	TRAJ1 6	2	TRBV6-2	13	TRBJ2-5	8
TRAV12-3	5	TRAJ1 7	3	TRBV6-4	1	TRBJ2-6	2
TRAV13-1	3	TRAJ1 8	2	TRBV6-5	32	TRBJ2-7	33
TRAV14/DV 4	6	TRAJ2 0	1	TRBV6-6	12		
TRAV17	5	TRAJ2 1	2	TRBV7-2	4		
TRAV19	4	TRAJ2 2	5	TRBV7-3	3		
TRAV20	4	TRAJ2 3	14	TRBV7-6	3		
TRAV21	37	TRAJ2 4	10	TRBV7-8	1		
TRAV22	5	TRAJ2 6	2	TRBV7-9	21		
TRAV24	8	TRAJ2 7	3	TRBV9	22		
TRAV26-1	4	TRAJ2 8	13	TRBV10-3	3		
TRAV26-2	4	TRAJ2 9	5	TRBV11-1	2		

TRAV27	4	TRAJ3 0	7	TRBV11-2	8		
TRAV29/DV 5	4	TRAJ3 1	5	TRBV12-3	1		
TRAV30	3	TRAJ3 2	2	TRBV12-4	1		
TRAV35	4	TRAJ3 3	6	TRBV13	2		
TRAV36/DV 7	2	TRAJ3 4	6	TRBV14	1		
TRAV38- 2/DV8	1	TRAJ3 6	1	TRBV15	1		
TRAV39	1	TRAJ3 7	8	TRBV16	1		
		TRAJ3 9	4	TRBV18	2		
		TRAJ4 0	2	TRBV19	10		
		TRAJ4 1	4	TRBV24-1	1		
		TRAJ4 2	5	TRBV25-1	1		
		TRAJ4 3	6	TRBV27	7		
		TRAJ4 5	7	TRBV28	6		
		TRAJ4 8	8	TRBV29-1	2		

		TRAJ4 9	9	TRBV30	1		
		TRAJ5 0	1				
		TRAJ5 2	6				
		TRAJ5 3	3				
		TRAJ5 4	2				
		TRAJ5 7	1				
		TRAJ5 8	3				

*Table 3.2: List of human V and J genes which did not have an available crystal structure.*

TRAV genes unrepresented (17/47)	TRAJ genes unrepresented (18/61)	TRBV genes unrepresented (23/60)
TRAV1-1	TRAJ1	TRBV1
TRAV2	TRAJ2	TRBV3-1
TRAV6	TRAJ3	TRBV3-2
TRAV7	TRAJ10	TRBV5-3
TRAV8-1	TRAJ14	TRBV5-7
TRAV8-6	TRAJ19	TRBV6-3
TRAV8-7	TRAJ25	TRBV6-7
TRAV9-1	TRAJ35	TRBV6-8

TRAV11	TRAJ38	TRBV6-9
TRAV13-2	TRAJ44	TRBV7-1
TRAV16	TRAJ46	TRBV7-4
TRAV18	TRAJ47	TRBV7-7
TRAV25	TRAJ51	TRBV10-1
TRAV34	TRAJ55	TRBV10-2
TRAV38-1	TRAJ56	TRBV11-3
TRAV40	TRAJ59	TRBV12-1
TRAV41	TRAJ60	TRBV12-2
	TRAJ61	TRBV12-5
		TRBV17
		TRBV20-1
		TRBV21-1
		TRBV23-1
		TRBV26

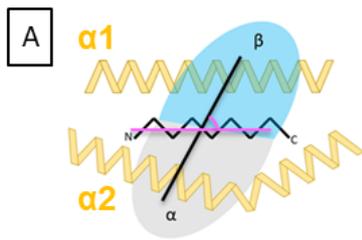
### *3.2.1.2 Building contact maps for all structures*

To interrogate interactions between TCR domains, contact maps for all interfaces were created using the Molecular Operating Environment (MOE) software to list arene, ionic, hydrogen bonds and Van der Waals predicted interactions of any residues within 4.5 Å of another TCR domain. Due to the limited size of the dataset, all such interactions are considered a positive hit for the purposes of further analysis and were considered in aggregate. IMGT definitions of V and C domains (residues 1-128 and 1001-1128 respectively, as all C domain residue numbers were increased by 1000 to distinguish them from the V domain) will be used throughout, including residues encoded by V D and J genes within the variable subunit. The key residues that interact at each interface will be discussed throughout this chapter.

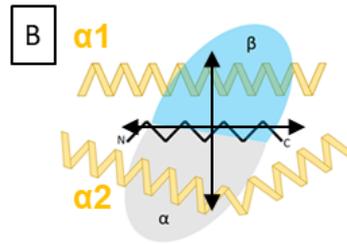
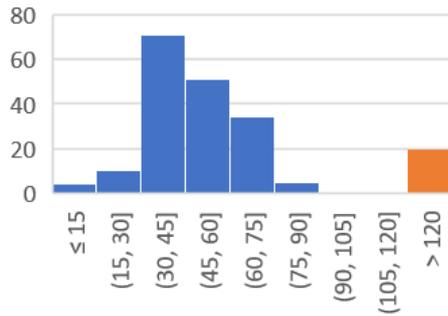
### 3.2.1.3 Geometry of TCR:pMHC binding footprints

Figure 3.1 shows an overview of the structure binding footprints and binding geometries, demonstrating that a range of docking positions are present within this dataset. The crossing angle of TCRs relative to the pMHC is calculated as the angle between the peptide binding groove and the centre of the two conserved disulphide bonds in the TCR V domains, which canonically lie diagonally across the groove (Rudolph, Stanfield and Wilson, 2006). Figure 3.1A highlights the 20 structures which bind with a docking angle greater than  $120^\circ$ , indicating a reverse binding orientation. This non-canonical binding position is associated with failure to form a functional T-cell signalling complex (Zareie *et al.*, 2021), but these structures will not be excluded from further analysis of their internal domain contacts as they still represent fully folded and pHLA bound TCRs. Figure 3.1B shows that despite this reverse docking, the centre of their binding footprint to pHLA overlaps with the range made by canonically-docked TCRs.

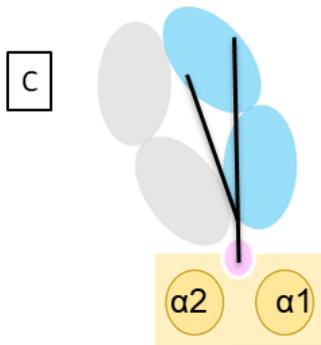
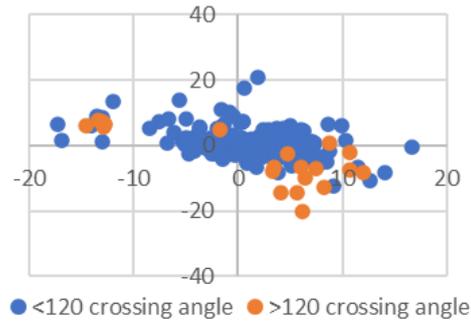
The column chart in Figure 3.1E shows that the distribution of contacts between CDRs and the pHLA for all structures fits with reported TCR docking profiles (Szeto *et al.*, 2021), in which the longer CDR3 loops of both chains make up the majority of the TCR:pHLA interface. There is limited contribution to peptide contacts by the  $\alpha$ CDR2 loops, which instead mostly make contact to the HLA helices, whereas the  $\beta$ CDR2 loops occasionally make peptide contacts as well as HLA. The majority of peptide contacts are made by the CDR3 loops of both  $V\alpha$  and  $V\beta$ .



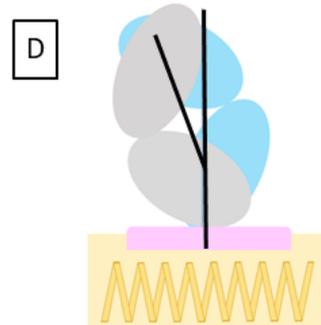
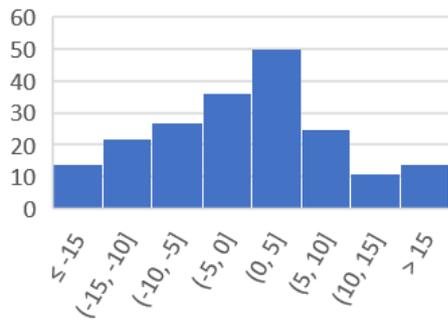
TCR/MHC crossing angle



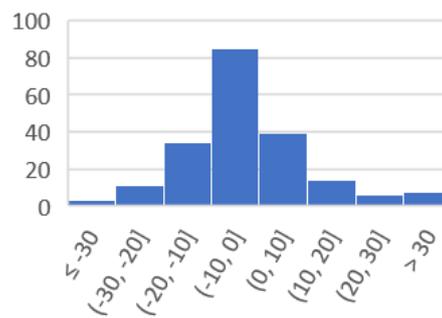
TCR/MHC docking



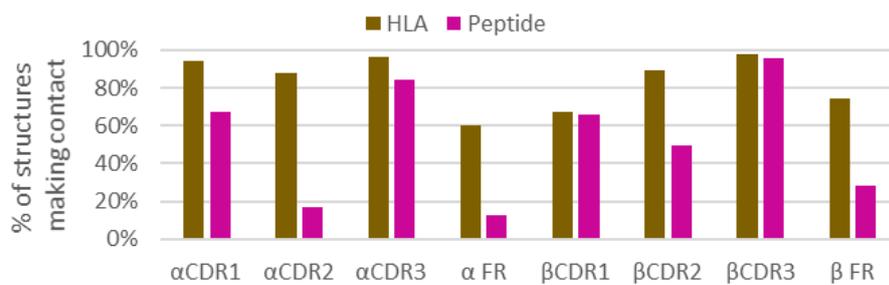
TCR/MHC tilt angle



TCR/MHC roll



**E** pHLA contacts



*Figure 3.1 Overview of TCR:pHLA binding geometries for structures studied. Schematic representations show TCR as ovals (grey and light blue for alpha and beta chains), peptide as black zigzag line, HLA as yellow helices. A) shows a schematic of the TCR:pMHC crossing angle as a top-down view of the pHLA groove, and a histogram of the distribution of angles in the structural dataset. Orange bar indicates “reverse binders”, where the crossing angle is greater than 120°. B) shows a schematic of the TCR:pMHC docking position as a top-down view of the pHLA groove, and an XY coordinate plot of the centre of mass of the TCR overlaid on the pMHC. Reverse binders are again shown in orange; they fit within the existing docking positions. C) Shows the tilt angle of TCR docking onto the pMHC as the tilt away from perpendicular to the peptide groove, with a histogram to show the distribution of tilt angles in the dataset. D) shows the roll angle of TCR docking onto the pMHC as the tilt away from a perpendicular angle to the line of the peptide, with a histogram to show the distribution of tilt angles in the dataset. E) shows a column chart of the percentage of structures in the dataset that make contact to the pHLA (brown) or the peptide (magenta) using different CDR loops or non-CDR framework residues (FR).*

### 3.2.1.4 Inter-domain interfaces within TCR structures

TCRs are a four-domain modular structure, with interfaces occurring within chain (V $\alpha$ :C $\alpha$ , V $\beta$ :C $\beta$ ) and between the two chains (V $\alpha$ :V $\beta$ , C $\alpha$ :C $\beta$ ). As the focus of this study is on variation in the framework regions of the variable domains, I will not cover the invariant C $\alpha$ :C $\beta$  interface. Table 3.3 shows a comparison of these internal interfaces with that of the TCR:pMHC binding interfaces. All internal interfaces are smaller than the TCR:pMHC binding interface, in particular the V $\alpha$ :C $\alpha$  interface where only four residues form contacts between the two domains. Although the pMHC binding interface is mediated by the highly sequence-diverse CDRs, there is still reasonable diversity in inter-domain contacting residues.

*Table 3.3 Breakdown of TCR interfaces. BSA is shown as the mean result across all 200 structures. The number of residues given indicates the number of positions which make an interface contact in more than 5% of my dataset of 200 structures, taken from the contact map discussed in the text. Diversity is calculated as the mean of the Shannon entropy of each germline-encoded residue position involved in the interface, as calculated from an alignment of all TRAV and TRBV genes. Higher Shannon entropy represents higher diversity, where a position which encoded all 20 amino acids in equal proportions would score 4.32.*

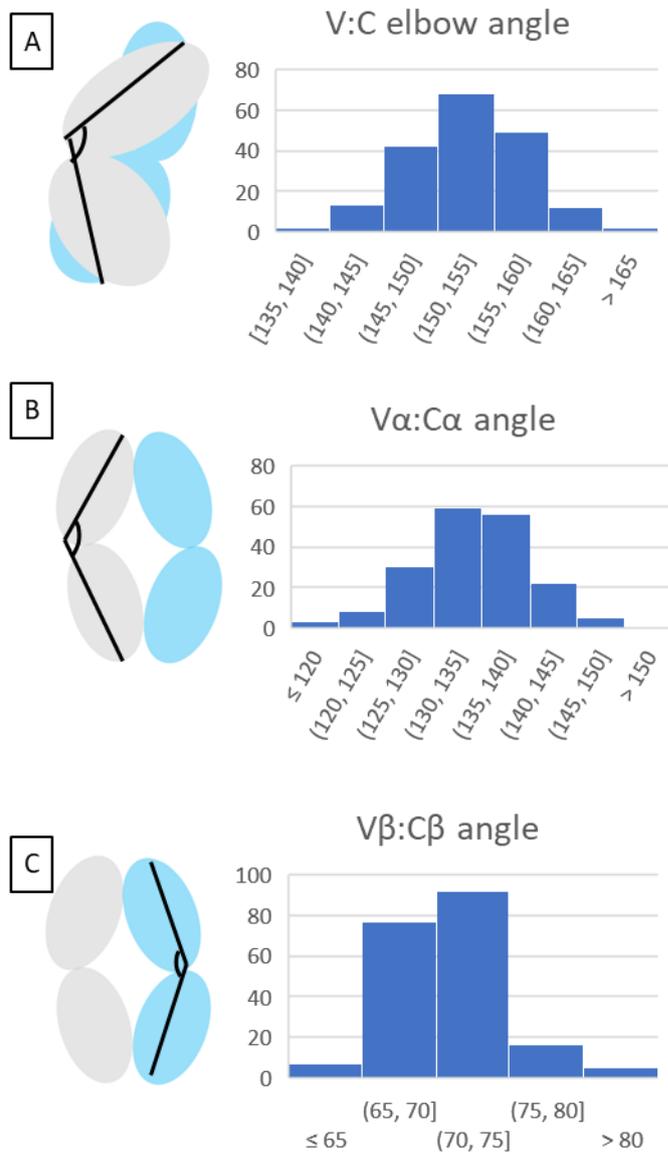
Interface	Average buried surface area	Average no. of V residues	Diversity (Shannon entropy)
TCR:HLA	1430 Å	52	1.85

V $\alpha$ :C $\alpha$	283 Å	4	1.4
V $\beta$ :C $\beta$	591 Å	9	1.5
V $\alpha$ :V $\beta$	881 Å	V $\alpha$ :30, V $\beta$ :37	1.4

Figure 3.2 shows an overview of the internal geometries of the TCRs in my dataset. The overall “elbow angle” of TCRs (Figure 3.2A) was determined based on the pseudo-dyad method from Stanfield *et al.* (2006), with the IMGT numbered residues 127 and 128 used to define the end of the V $\alpha$  and V $\beta$  domains respectively. This represents the angle between the two variable and the two constant domains of the structure, which in antibodies can vary between 115 and 225 for antibodies with lambda chains (a slightly narrower range of 125-195 was observed for kappa-chain antibodies (Stanfield *et al.*, 2006)). For TCRs, this elbow angle is much more restricted – as seen in Figure 3.1A 98% of the structures in this representative subset fall within a 25° range (140° to 165°). This likely reflects the evolutionary requirement for functional TCRs to correctly associate with coreceptors to form a signalling complex, which presumably imposes steric restrictions on the elbow angle (Xu, Li and Xu, 2020).

As the range of overall V:C elbow angles is relatively narrow for TCRs, it may obscure subtler differences between the interdomain angles for the V and C domains of each TCR chain considered separately. The angle between three structurally conserved residues (cysteine 23 in the V domain, position 127 at the end of the J gene, cysteine 104 in the constant domain), adapted from equivalent antibody residues (Al Qaraghuli *et al.*, 2020)) was measured for each alpha and beta chain. As seen in Figure 3.2C, the range of V $\beta$ :C $\beta$  angles for the structures studied is very narrow, with 188 of the 200 structures falling within a 15° arc. The rigidity of this interface has been proposed as a possible mechanism for transmitting allosteric changes on

TCR:pHLA binding during TCR signalling (Mariuzza, Agnihotri and Orban, 2020), which is consistent with the narrow range of angles observed relative to the  $V\alpha:C\alpha$  interface (Figure 3.2B). The more unconstrained  $V\alpha:C\alpha$  interface is mediated through fewer inter-domain contacts, discussed in more detail later in this chapter.

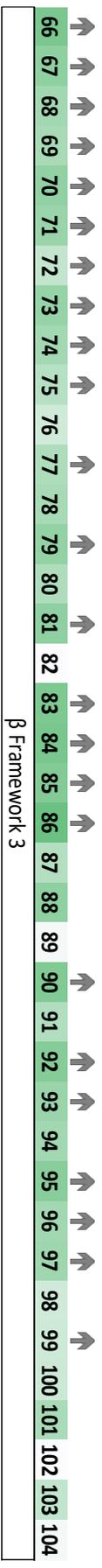
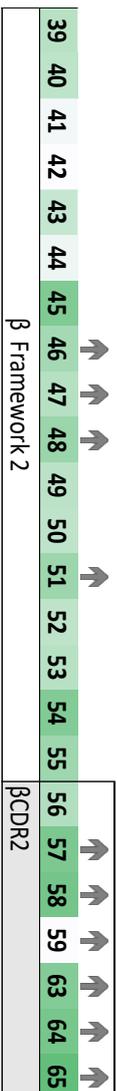
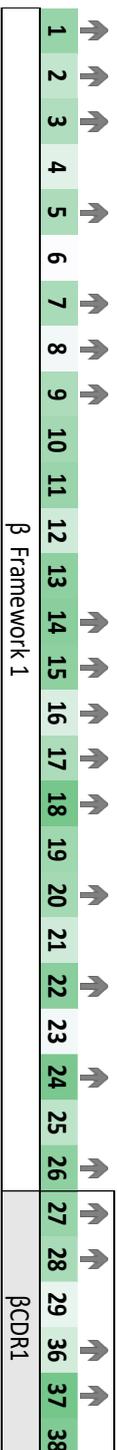
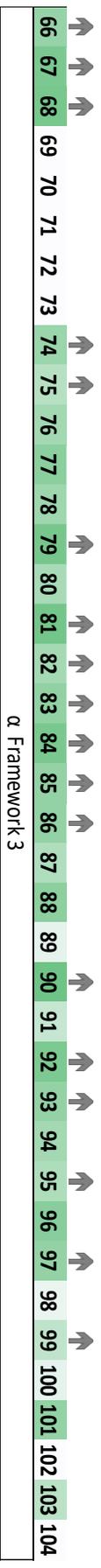
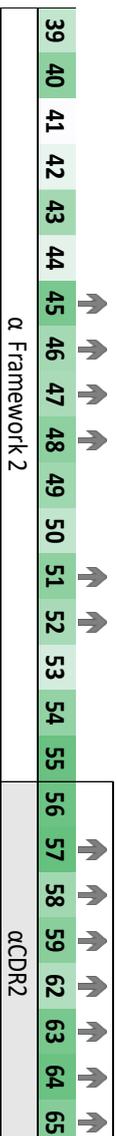
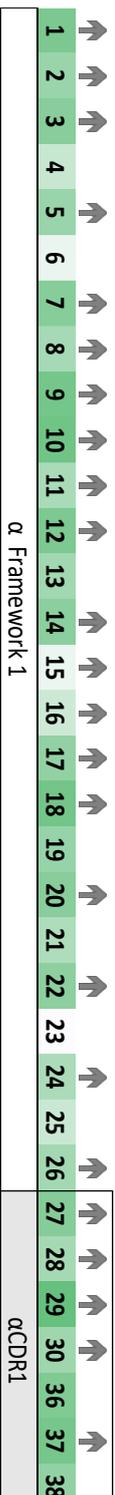


*Figure 3.2 Overview of inter-domain geometries for TCR structures in this dataset. Schematic representations show TCR as ovals (grey and light blue for alpha and beta chains). A) shows the elbow angle between the variable and constant domains of the TCR, calculated using the pseudo-dyad method from Stanfield et al. (2006). The distribution of angles observed in my TCR dataset is shown as a histogram B&C) show the separate  $V\alpha:C\alpha$  and  $V\beta:C\beta$  inter-domain angles, calculated as the geometric angle between cysteine 23 in the V domain, position 127 at the end of the*

*J gene and cysteine 104 in the constant domain. Distribution of these angles in the structures studied is shown as a histogram.*

### 3.2.1.5 Solvent exposure and amino acid diversity for V gene residues

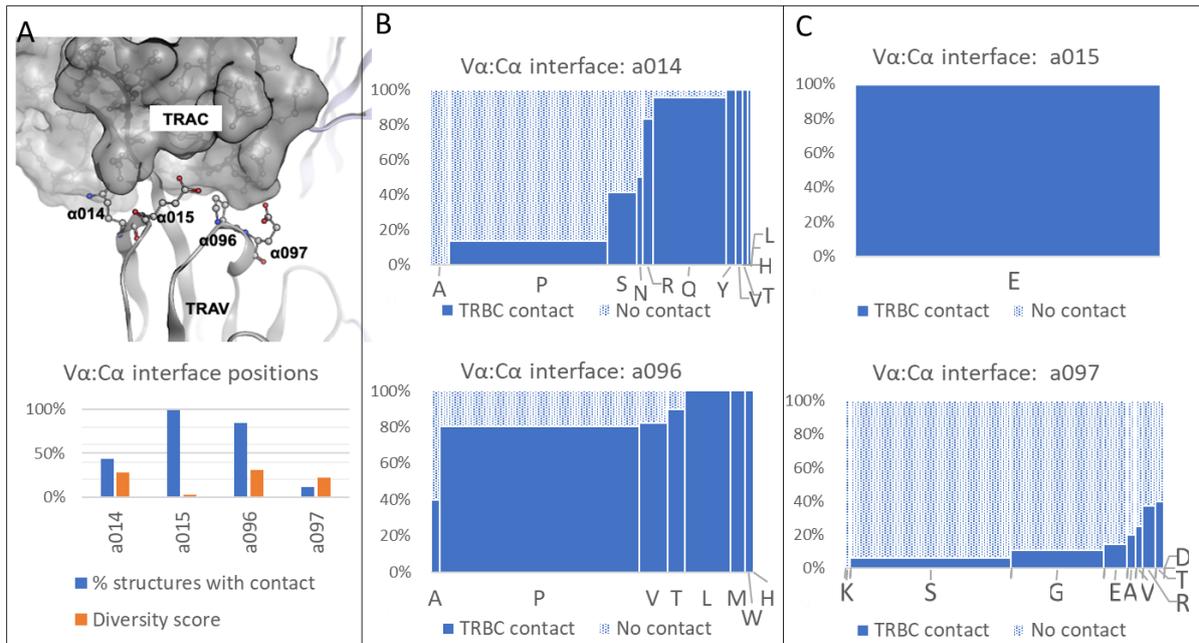
Figure 3.3 shows the solvent exposure and amino acid diversity for residues across the TRAV and TRBV encoded regions of the V domain. Relative Solvent Accessibility for each residue in each TCR structure was calculated as the accessible surface area in angstrom divided by the ideal surface area as determined from Gly\_X\_Gly triplets (Rost and Sander, 1994). Residues were classified as buried if they had a relative solvent accessibility score of less than 30% of the maximum possible for that amino acid. Amino acid diversity (shown using a heat-map approach) was calculated as the Shannon entropy of the TRAV or TRBV repertoire aligned according to IMGT number. Amino acid diversity was shown to exist across the framework regions as well as the germline CDRs, with most highly conserved residues being buried within the protein core. This distribution of diversity implied that different V genes presented different protein surfaces depending on the side chain properties present, which may have implications for TCR stability as hydrophobic patches on the surface can influence aggregation potential. I will examine the diversity of residues that make contacts at different inter-domain interfaces in subsequent sections.



*Figure 3.3. Solvent exposure and residue amino acid diversity across the TRAV and TRBV encoded regions of the V domains. Residues are numbered according to IMGT schema. Green colouring is a heat map to indicate Shannon entropy at germline-encoded residue position, as calculated from an alignment of all TRAV and TRBV genes. Higher Shannon entropy represents higher diversity, and is coloured in a stronger green. Grey arrows indicate surface exposed positions, calculated based on the mean relative solvent accessibility of that position across the 200 structures studied. Residues were classified as buried if they had a relative solvent accessibility score of less than 30% of the maximum possible for that amino acid. Relative Solvent Accessibility for each residue in each TCR structure was calculated as the accessible surface area in angstrom divided by the ideal surface area as determined from Gly\_X\_Gly triplets (Rost and Sander, 1994).*

### **3.2.2 Interactions at the V $\alpha$ :C $\alpha$ interface**

Figure 3.4 shows the key positions involved in the interface contacts between the constant and variable domains of the alpha chain (as defined by IMGT). Interface contact positions were identified as such if the side chain was involved in a contact in more than 10% of the structures in the database, as discussed above. The interface between the V $\alpha$  and C $\alpha$  domains is the smallest such inter-domain interface discussed in this chapter (an average buried surface area of only 283 Å<sup>2</sup> in the structures studied) and consists of four key contact residues from the V $\alpha$  domain, as seen in Figure 3.4A. These residues can be separated into those where the amino acid usage at that position notably impacts on the percentage of structures where a contact is found (alpha 014 and 096, Figure 3.4B) and those where either all structures use the same conserved residue to form an inter-domain contact (alpha 015) or where the residue is reasonably diverse but most structures do not make a contact regardless of side chain chemistry (alpha 097). The first group represents obvious targets for structure-guided mutagenesis if the existing side chain is not making a possible contact.

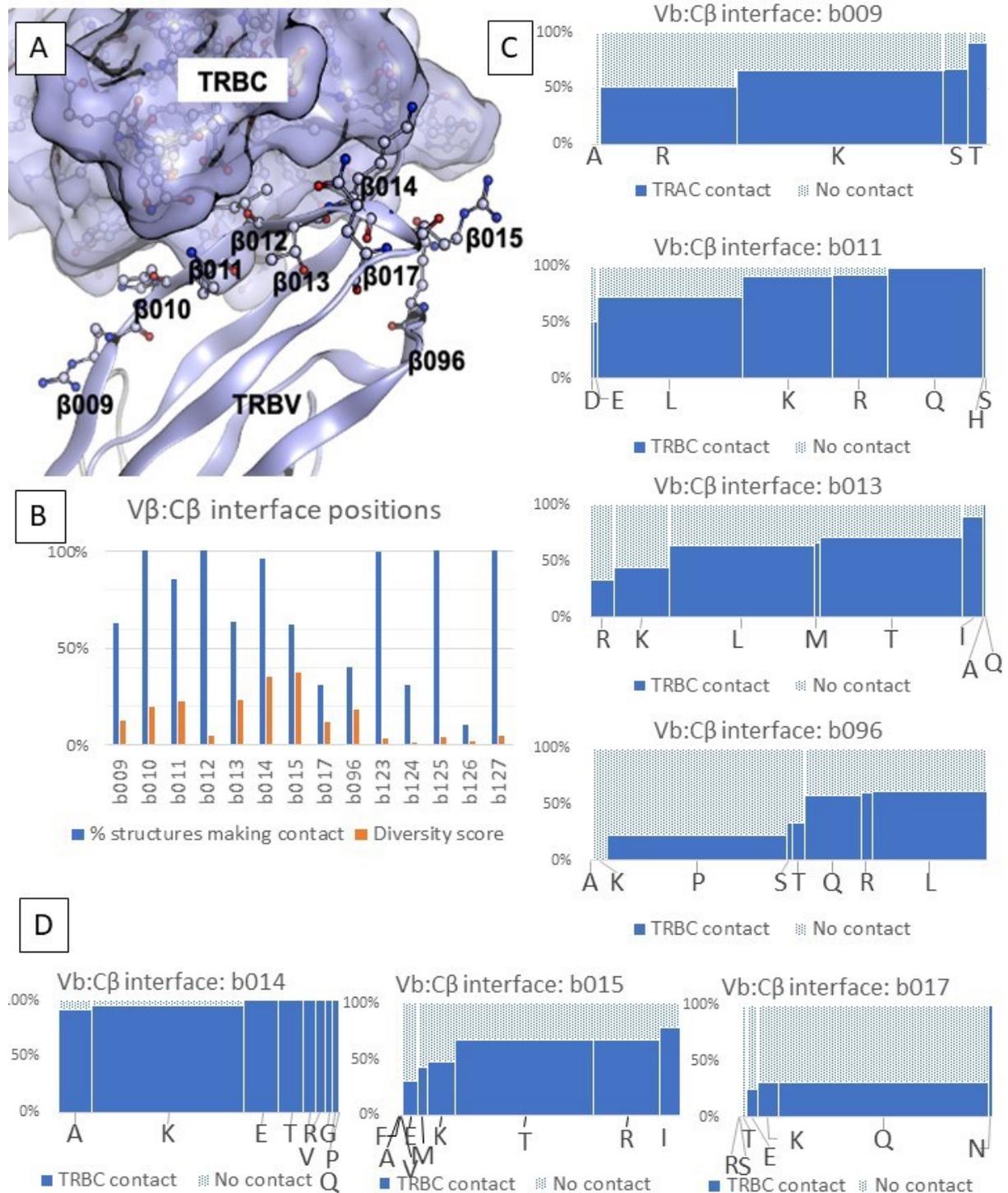


**Figure 3.4 : Overview of Va:Ca contacts.** A) Inset image shows an example crystal structure with the relevant residues labelled, and the Ca domain indicated by a rendered molecular surface. Column chart showing which IMGT positions make contacts between the Va and Ca domains in more than 10% of the structures studied. X axis represents % of structures in dataset where this position is involved in a contact. Orange columns indicate sequence variability score amongst V genes at each position, calculated as Shannon Entropy at that position for an alignment of all V genes by IMGT number, where a score of 100% would indicate all 20 amino acids represented equally at that position. (B & C) Detailed contact analysis at specific structural positions. Variation in side chain contact propensity shown by use of variable width column chart, where column height represents the percentage of structures where the labelled amino acid is involved (solid colour) or not involved (dot fill) in an inter-domain contact. Column width corresponds to number of structures where a particular amino acid is present at the position of interest, to visualise the size of the dataset for that residue. Positions have been split into b) contacts that are side chain dependent and C) contacts which show limited variation based on side chain chemistry.

### 3.2.3 Interactions at the V $\beta$ :C $\beta$ interface

Figure 3.5 shows an overview of contacts at the V $\beta$ :C $\beta$  interface. As discussed in Table 3.3, this is a larger interface than Va:Ca, with more residues that make inter-domain contacts. Residues 009, 011, 013 and 096 showed notable amino-acid variation in contact propensity; for example, less than 40% of structures with an arginine at position 013 made a V $\beta$ :C $\beta$  contact, whereas almost 80% of structures where this residue is an isoleucine instead had an inter-domain contact at that site.

From Figure 3.5B it can be seen that residue diversity alone was not a useful predictor of the presence of amino-acid specific contacts; the most diverse position ( $V\beta$  014) showed nearly uniform formation of an interface contact regardless of the amino acid present at that structural position.

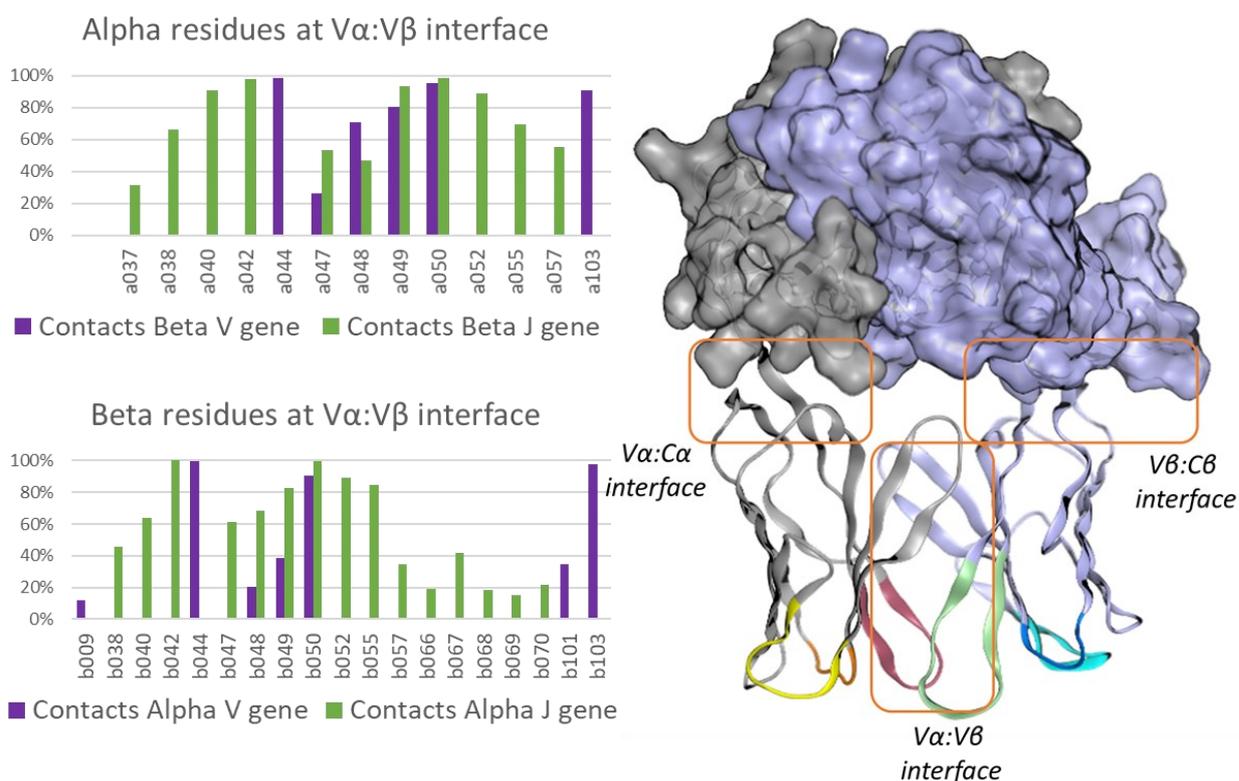


*Figure 3.5: Overview of V $\beta$ :C $\beta$  contacts. A) Inset image shows an example crystal structure with the relevant residues labelled, and the C $\beta$  domain indicated by a rendered molecular surface. Column chart showing which IMGT positions make contacts between the V $\beta$  and C $\beta$  domains in more than 10% of the structures studied. X axis represents % of structures in dataset where this position is involved in a contact. Orange columns indicate sequence variability score amongst V genes at each position, calculated based on the Shannon Entropy at that position for an alignment of all V genes by IMGT number, where a score of 100% would indicate all 20 amino acids represented equally at that position. B & C: Detailed contact analysis at specific structural positions. Variation in side chain contact propensity shown by use of variable width column chart, where column height represents the percentage of structures where the labelled amino acid is involved (solid colour) or not involved (dot fill) in an inter-domain contact. Column width corresponds to number of structures where a particular amino acid is present at the position of interest, to visualise the size of the dataset for that residue. Positions have been split into B) contacts that are side chain dependent and C) contacts which show limited variation based on side chain chemistry.*

### **3.2.4 Overview of the V $\alpha$ :V $\beta$ interface**

The V $\alpha$ :V $\beta$  interface is the largest inter-domain interface within the four domain structure of the TCR (average of 881 Å buried surface area, compared with 283 Å and 591 Å for the V $\alpha$ -C $\alpha$  and V $\beta$ -C $\beta$  interfaces respectively). Previous work examining this interface has been focused around the concept of chain pairs; the use of different TRAV and TRBV genes to define the overall TCR subtype. This is a logical shorthand when discussing conserved chain pairs for target binding (such as the chain restricted MR1 specific TCR classes) but the frequent reference to chain pairings driving alterations in TCR stability, T-cell surface presentation or soluble expression levels is an oversimplification of the complex nature of this interface. As seen in Figure 3.6A, the majority of residues encoded by TRAV genes are involved in contacts with TRBJ encoded residues, not those from the TRBV gene (and the equivalent is true for TRBV encoded residues, due to the symmetrical nature of the interface). The complex interaction network between the two chains would suggest that if hard-wired pairing preferences exist, they could as easily be dependent on V-J interactions as the V $\alpha$ -V $\beta$  pairing most often cited. Analysis of existing single-cell sequencing datasets show essentially unbiased pairing of TRAV/TRAJ and TRBV/TRBJ genes within T-cell repertoires, so no combination of chain pairs would be incompatible with surface expression and target binding (Shcherbinin, Belousov

and Shugay, 2020). However, in the context of soluble TCR production there has been reported chain-specific variation in the refold yield of different molecules (internal data, unpublished) which would indicate that although unbiased pairing can occur in a cellular context, certain interfaces are more optimal for refolding in mTCR format.



**Figure 3.6: Overview of Va:Vβ interface contacts.** Bar charts show the percentage of structures making an inter-domain contact at each position, with bars coloured to indicate if they interact with residues of the opposing domain encoded by the J gene (Green) or V gene (purple).

Stabilisation of the analogous antibody system has successfully targeted increasing contacts between the heavy and light variable regions for increased stability.

(Warszawski *et al.*, 2019). To further investigate the existing interfaces and their potential for engineering to maximise contacts, Va:Vβ contacts will be analysed as either sequence conserved contacts involving sequence-conserved residues or side-chain specific interactions that only occur between subsets of V and J genes.

### **3.2.5 Conserved interactions at the V $\alpha$ :V $\beta$ interface**

The V $\alpha$ :V $\beta$  interface is the most complex of the three inter-domain interfaces discussed in this chapter. Based on the diversity of amino acids present at each of the contact positions, I will discuss the conserved core interactions present in the majority of structures and the more TCR-specific interactions (which occur between residues that are highly sequence diverse) separately. Conserved interactions at the interface are shown in Figure 3.7, with the limited interface of TRAV to TRBV encoded contacts highlighted as the most conserved core. Interaction heat maps show that most structures studied make contacts between the domains as part of a symmetrical core of hydrophobic residues. Of particular interest is the H-bond between residues  $\alpha$ Q44 and  $\beta$ Q44; Hoffmann, Krackhardt and Antes (2015) identified this as the fulcrum point around which the V $\alpha$ :V $\beta$  interdomain angles of published structures rotate. The interface is otherwise mediated through hydrophobic contacts, and the conserved interface (illustrated in inset image of Figure 3.7) between all 10 positions of this symmetrical core is fully present in the contact maps for 70% of structures in this study.

The symmetrical nature of the V $\alpha$  V $\beta$  interface is also observed further out from this core, where three highly sequence conserved TRAV or TRBV encoded residues make contacts to J-gene encoded amino acids at the end of the CDR3 of the opposite domain (V $\alpha$ -J $\beta$  and V $\beta$ -J $\alpha$  paired interactions). These contacts most frequently occur between position 42 on one domain and the conserved phenylalanine residue (118) that ends the CDR3 loop as part of the FGxG motif of the opposing domain, again driven by hydrophobic interactions and aromatic stacking of the sidechains. Despite occurring within the hypervariable CDR3 loop (as defined by IMGT) position 116 on both chains is relatively sequence conserved so is also included in this subset of interface residue, although the majority of contacts to this residue are contacting backbone atoms rather than the sidechain as seen in both inset example crystal structures. In both the alpha and beta chains the sequence conserved positions 49

and 52 contact a conserved residue of the opposing chain in fewer than 50% of the structures studied. However, these residues can also interact with positions of high sequence diversity (as discussed in the following section).

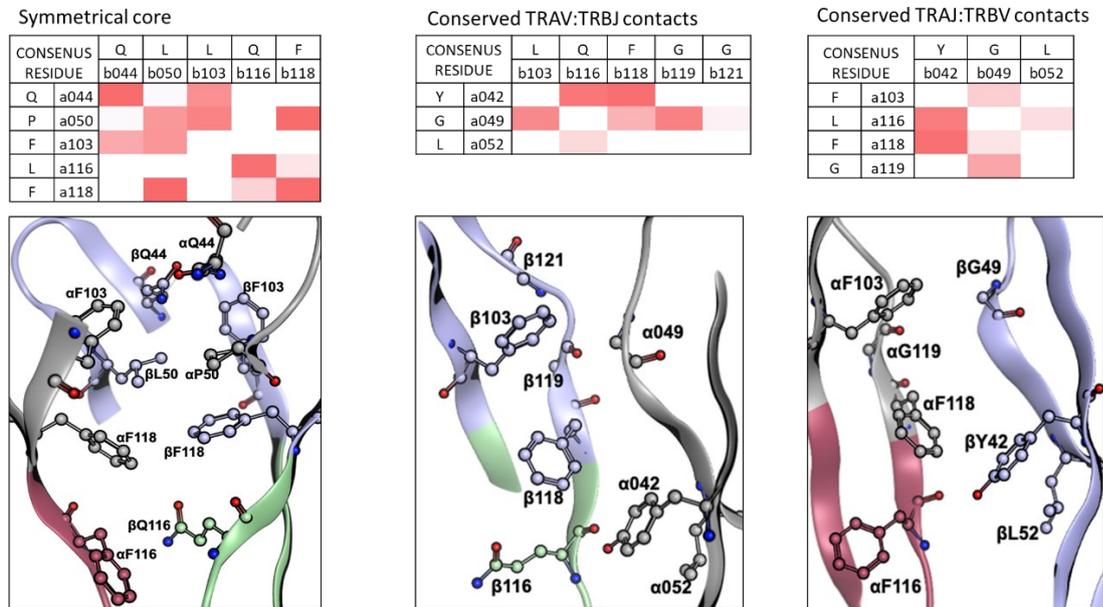


Figure 3.7: conserved interactions at the  $V\alpha:V\beta$  interface can be broken down into three groupings based on the inter-domain contacts. The most conserved interactions happen at the symmetrical core of this interface where structurally equivalent positions interact with their partner on the opposing chain. Further to each side of this core, sequence-conserved TRAV-encoded residues of the  $V\alpha$  interact with  $V\beta$  residues encoded by TRBJ genes, and sequence-conserved TRBV-encoded residues of the  $V\beta$  interact with  $V\alpha$  residues encoded by TRAJ.

### 3.2.6 Diversity at the $V\alpha:V\beta$ interface

The more conserved contacts described above occur in most structures studied using a relatively restricted subset of amino acid contacts. However, there are also many  $V\alpha:V\beta$  contacts that occur frequently but are mediated using a wider range of potential amino acid sidechains, as they occur between positions that have high sequence diversity across all V and J genes encoded in the human repertoire. Figure 3.8 shows the heat map for these interactions, which are exclusively between TRAV:TRBJ or TRBV:TRAJ encoded residues; the few direct interactions between TRAV and TRBV positions all involve sequence-conserved residues and are discussed in the previous section. The inter-domain contacts that occur between diverse

residues are highly side-chain dependent, as seen in Figure 3.8 where the majority of the residues involved in  $V\alpha:V\beta$  contacts are highly sequence diverse across the repertoire.

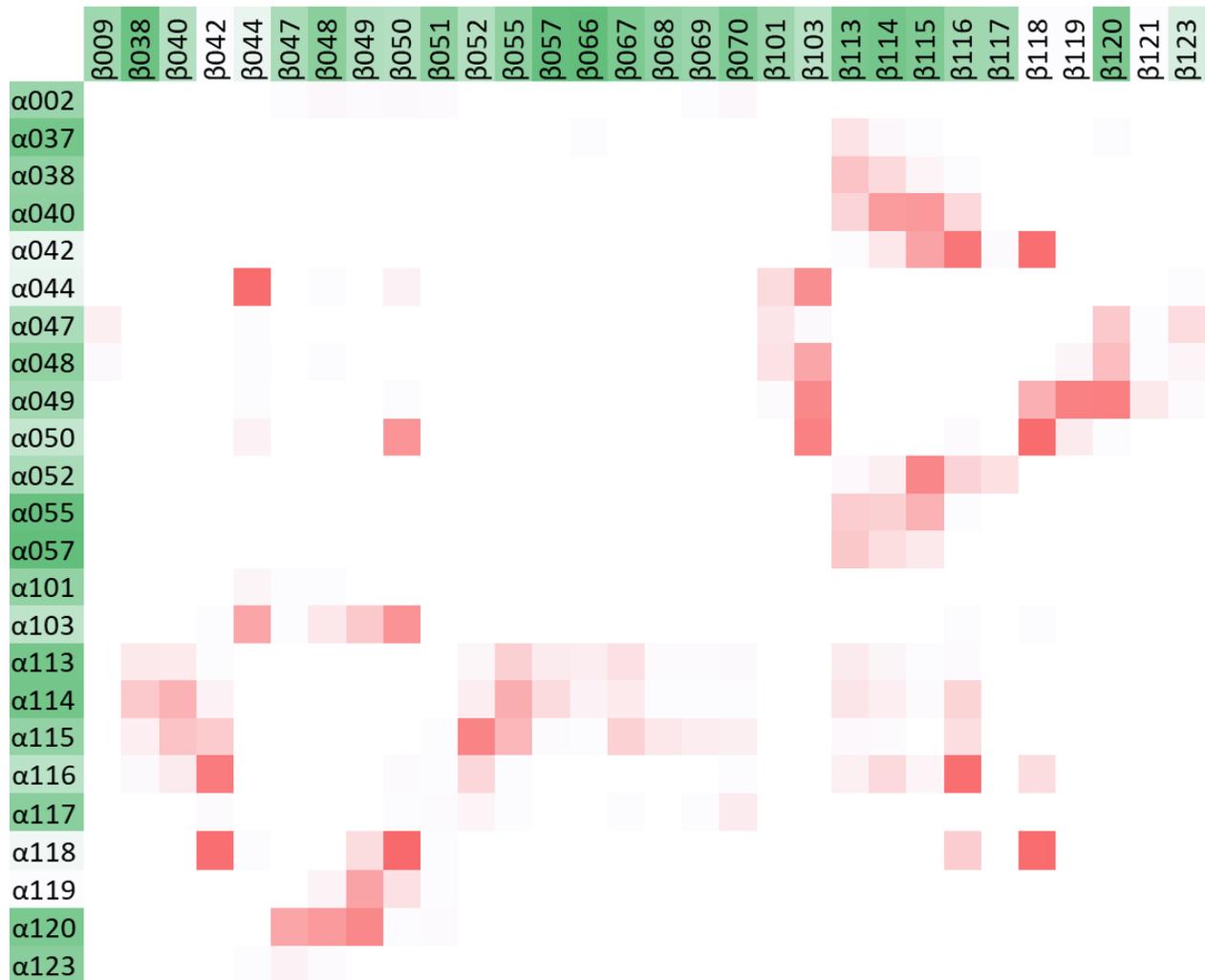


Figure 3.8: Heat map showing TRAV:TRBV inter-domain contacts. Stronger red used to indicate that more structures exist where the two residues in the matrix create a contact (as defined in text). Green colour on the residue numbering is stronger for more sequence-diverse positions (calculated as in Figure 3.3) Boxed residue numbers are part of the IMGT-defined CDR3 loops. Residues 107-112 were removed from diagram as they made up the hyper-variable core of the CDR3 loops. Only positions where more than 5% of the structures in dataset made an inter-domain contact are shown.

### **3.3 Conclusion**

This systematic analysis of a representative subset of TCR crystal structures allows the identification of positions that create contacts between the various modular domains of the receptor. This framework for defining interface contact positions will be used throughout subsequent chapters.

The main limitation of this approach is the limited dataset available. Even with the addition of in-house unpublished structures, the TCR structural dataset does not cover the full diversity of V and J (see Table 3.1 for a list of these genes and number of each present in the database). The absence of some V and J genes obviously has implications for predicting mutations based on sequences that are not present in the available structural dataset (see Table 3.2 for a list of the sequences not covered in this study), or conversely risks bias towards those chains that are dominant in the pool of existing structural data. However, although suboptimal for a truly unbiased approach, my database will predominantly cover those TCRs which are easily refolded and have therefore been found amenable to crystallisation – this may be beneficial for suggesting mutations to TCRs using better studied chains and could indicate that over-represented chains are better for soluble production and worthy of selection in their own right.

TCRs have evolved in the context of acting as signalling systems embedded in T cell membranes. Translating this system into an effective soluble drug is a challenge. Notably, my investigation of available TCR structures has revealed no structures where all most optimal residues at the conserved portions are present. This suggests that either the TCRs do not need to make all contacts possible (and presumably a very stable interface) in their evolved function, or perhaps that the best residues are sterically hindered from existing in the same molecule.

The novelty of this approach is the ability to generate suggestions for structure-guided mutagenesis based on an IMGT-numbered sequence alone. Consensus mutagenesis has been proposed as a stabilising strategy for multiple classes of

proteins, but generally solely relies upon sequence alignments to related proteins to identify possible substitutions for the target molecule (Lehmann *et al.*, 2000; Amin *et al.*, 2004; Sternke, Tripp and Barrick, 2019). Mutations to the consensus residue at a given position in the aligned sequences identified in this fashion can be beneficial in published examples, but required systematic experimental screening to identify successful hits. By also considering the structural importance of side chains for making contacts, this approach allows the identification of possible substitutions using an amino acid that is both evolutionarily tolerated at a given structural site and also more predisposed to make an interface contact.

# Chapter Four

## Trends in TCR thermostability

---

### 4.1 Introduction

As described in the introduction, melting point ( $T_m$ ) has long been used as a measurement of overall protein stability, particularly as a rapid assay for high throughput screening during the development of therapeutic agents (Ma, Ó'Fágáin and O'Kennedy, 2020). Given that TCRs have highly variable amino acid sequences and show sequence-dependent variation in their surface presentation levels when expressed in T cells (Hoogenboom *et al.*, 1999; Hart *et al.*, 2008; Thomas *et al.*, 2019), it is expected that there will be variation in the thermal stability of soluble disulphide-stabilised TCR proteins. The modular combination of sequence diverse variable domains (V $\alpha$  and V $\beta$ ) also offers the possibility that certain combinations of alpha and beta chains may be more stable as refolded proteins than other combinations, despite apparent unrestricted chain pairing occurring *in vivo* (Howie *et al.*, 2015; Shcherbinin, Belousov and Shugay, 2020). Identifying sequence features which correlate with thermostability has great potential value for improving TCRs as soluble therapeutics; both by aiding initial choice of molecules based on what is known to be a stable and therefore an attractive manufacturing candidate, and by suggesting mutations which may be used to improve poorly thermostable molecules that have other positive characteristics. More broadly, understanding the role of TCR V domains in modulating overall molecular stability offers insight into the evolutionary advantage of V domain amino acid diversity away from the peptide-MHC contacting residues, and may explain why certain V genes seem to be preferentially expressed on the T cell surface (Heemskerk *et al.*, 2007; Thomas *et al.*, 2019).

Within the overall T cell repertoire in peripheral blood some bias in V gene usage exists depending on MHC subtype, indicating that the germline CDR1 and CDR2 loops of these chains may be better suited to making contacts to different MHC

surfaces (Sharon *et al.*, 2016). This is also matched by the observation that pairing of  $V\alpha$  and  $V\beta$  genes shows some bias depending on the T cell lineage and the target MHC molecule (Carter *et al.*, 2019). However, there seems to be no binding-agnostic bias within the repertoire to imply that different combinations of V and J genes are better suited to pairing together (Shcherbinin, Belousov and Shugay, 2020). Alternatively, as discussed in Chapter Three, the modular nature of TCR V domain structures allows for the two domains to pair in an unrestricted fashion with both conserved and sequence diverse contacts being made across the  $V\alpha V\beta$  interface. TCR stability may vary independently of V or J gene usage, as the overall structure has regions of high diversity and CDR3 sequences in particular are highly diverse due to V-D-J recombination.

Examining the relative stability of TCRs in different formats has only been carried out in a limited manner on small panels of single chain variable-domain constructs (scTV) (Richman *et al.*, 2009; Aggen *et al.*, 2011) or to assess the ability of endogenous TCRs to outcompete an artificially stabilised TCR for T-cell surface expression (Thomas *et al.*, 2019). However, to date there has been no systematic study comparing the stabilities of different soluble  $\alpha\beta$  TCRs to determine the range of inherent stabilities possible in this sequence-diverse class of molecules, or to identify common sequence or structural trends that may drive stability differences between TCRs.

### **4.1.1 Aims of chapter**

In this chapter I will build a dataset of unique TCRs of known sequence and thermostability information in order to interrogate the role of different sequence features in determining protein stability. As different TCRs have been known to refold from inclusion bodies with greater or lesser efficacy (Immunocore, unpublished), or present on the surface of yeast cells at higher or lower levels (Richman *et al.*, 2009), it is reasonable to assume that there will be a range of melting points represented from a mixed group of TCRs with different sequences.

Possible underlying reasons behind such variability in stability will be investigated. TCRs using the same V genes are often assumed to share similar properties due to their high sequence similarity, but variation at the unique CDR3 loops may instead be more important features for thermostability; I will investigate any link between gene usage and overall melting point to identify which V genes tend to encode TCRs of greater or lesser thermostability.

Protein thermostability can also be influenced by more general structural and sequence features. The common architecture of TCRs allowed me to build a standard structural designation for IMGT aligned sequences based on my work in Chapter Three; I will use this to look for structurally informed sequence properties that may influence overall protein stability. In particular, the hydrophobicity of surface residues has been implicated as a frequent site for thermostabilising mutagenesis of proteins (Strub *et al.*, 2004; Güler *et al.*, 2020), so I will seek to identify any differences in native hydrophobicity of surface exposed residues between strongly and weakly thermostable TCRs. The hydrophobic core of proteins can be poorly packed or contain charge in weakly thermostable examples (Borgo and Havranek, 2012); I will also assess if difference in hydrophobicity of buried side chains could explain variation in overall TCR thermostability.

Finally I will investigate the impact of CDR mutagenesis on wild type TCR stability. Affinity enhancing mutations are often engineered purely on the basis of improved binding to pMHC targets, but TCR engineering for successful applications as a soluble biotherapeutic agent requires a balance between affinity and developability.

## **4.2 Results**

### **4.2.1 Thermostability analysis of refolded TCRs**

The melting points ( $T_m$ ) of various soluble TCRs and their amino acid sequences were available thanks to the experimental work carried out by multiple members of the Protein Science department at Immunocore. I assembled a dataset based on this information in order to investigate the relative stabilities of different TCRs and

identify trends in gene usage and other features of the protein that may explain variation in melting points. The  $T_m$  values for these 267 unique wild-type TCRs were obtained from differential scanning fluorimetry (DSF), and all sequences were aligned by IMGT numbering to enable direct comparison of structurally significant residues. The constant domains ( $C\alpha$  and  $C\beta$ ) do not have the sequence diversity of the variable domains so are presumed to not impact differences in overall TCR stability. As seen Figure 4.1A, a wide variety of  $T_m$  values is observed for these wild type molecules, with a mean of 58 °C but a range spanning more than 30 degrees. As the melting temperatures vary so widely between different TCRs that all share the same Boulter-disulphide stabilised constant domains, further investigation into the differences between variable domains that may determine greater or lesser thermostability was carried out.

#### **4.2.2 Influence of V-genes on thermostability of soluble TCR proteins**

Different TRAV or TRBV genes are sometimes associated with strong expression levels as native receptors or as engineered single-chain constructs, but the association between V gene usage and protein stability has not previously been studied directly.  $T_m$  values for the panel of TCRs studied were assigned IMGT TRAV and TRBV gene definitions based on amino acid sequence and the median value & interquartile ranges for each subgroup plotted separately in Figure 4.1. Some TRAV and TRBV genes were not present or only represented by a single molecule which limits the opportunity to draw more general conclusions about chain usage, so for the purposes of this study only those genes where  $T_m$  values for more than five unique wild type TCRs have been measured were included. Although the median  $T_m$  for different V genes varies, there is very wide within-group variation for TCRs which share a common TRAV or TRBV gene (Figure 4.1C&D), indicating that TRAV or TRBV usage is not the sole determinant of overall stability. Interestingly, TRBV9 and 7-9 were amongst the chains identified in a previous study as more likely to occur in T cells expressing “weak TCRs” which could not compete for expression

against an artificially stabilised exogenous TCR (Thomas *et al.*, 2019). However, in this study the mean  $T_m$  value of TCRs with these TRBV genes was 60.4 °C and 59.8 °C respectively, compared to an overall mean of 57.7 °C for all TCRs, indicating that the lack of dominance for these TRBV genes at the T cell surface does not translate into reduced thermostability. Other V genes associated with weak or dominant cell surface level expression in the study by Thomas *et al* were not present in this dataset in sufficient numbers to determine if their thermal stability as refolded TCRs matched T-cell surface expression behaviour. Another V gene noted to possess high stability (in this case for display in the scTV format) is TRAV12-2 (Aggen *et al.*, 2011), both subtypes of which have mean  $T_m$  values 2-3 °C higher than the mean for all TCRs. These alpha chains are present in 84 of the 267 TCRs measured in this study and this subgroup has a range of over 20 °C, indicating that although on average they are more stable than other TRAVs, TRAV12-2 does not necessarily ensure high thermal stability.

As grouping based on TRAV or TRBV alone will only define less than half of the V domain diversity, the ranges of melting points observed for chain pairings were also considered (Figure 4.1B). Again, there is wide within-group variation suggesting that the difference in CDR3 and J gene sequence between TCRs that have otherwise identical V domain amino acid sequences must drive variance in overall protein stability. It is also interesting to note that, when possible to assess due to sample size, the median  $T_m$  values for paired sequences follow the trend of the individual V types; there does not appear to be any combinatorial effects that make specific chain pairings notably more stable compared to the TRAV and TRBV rankings.

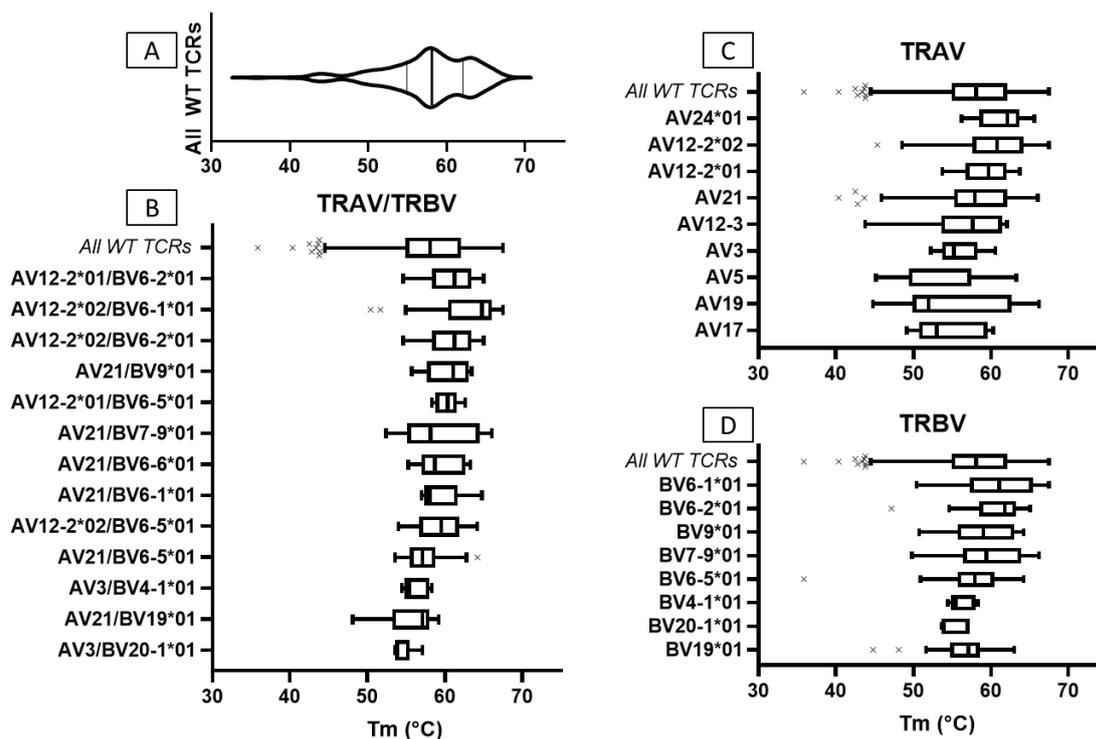


Figure 4.1. Thermal stability of refolded wild-type TCRs, with melting point ( $T_m$ ) measured using differential scanning calorimetry. A) Violin plot for all 267 values, with median  $T_m$  as thick solid line & top and bottom quartile values as thin lines. B-D) Box plots showing distribution of  $T_m$  values for different subgroups within this dataset, with full dataset included at top for comparison. Median value shown as thick line within box, with upper and lower quartile values defining box edges. Range outliers (as calculated using Tukey plot definition) shown as crosses. Only TRAV or TRBV genes/ gene pairings represented by at least five unique TCRs shown. A) shows  $T_m$  values separated by TRAV/TRBV pairing, C&D) show TRAV and TRBV separately, grouping TCRs with unique partner chain and CDR3/J gene sequences together based on gene usage.

### 4.2.3 CDR chemistry is not predictive of thermostability

As TRAV or TRBV gene usage did not reveal significant patterns in thermostability and TCRs with the same TRAV and TRBV chain usage can have a wide range of melting points, the variation in  $T_m$  observed may be more due to differences in CDR3 loop chemistry that affect stability. Hydrophobicity at surface exposed regions like CDR loops is commonly associated with poor developability characteristics (Dudgeon, Famm and Christ, 2008; Henry *et al.*, 2017; Jetha *et al.*,

2018; Du *et al.*, 2019; Raybould *et al.*, 2019) and may drive aggregation or otherwise destabilise soluble proteins. The overall hydrophobicity score for each of the six CDR loops for all TCRs in the dataset was calculated based on IMGT alignment of sequences to identify loops (residues 27-38, 56-65 and 105-117) and calculating the mean of the standard hydrophobicity score (Sweet and Eisenberg, 1983) for each amino acid in the loop. CDR hydrophobicity is often linked to poor stability profiles for antibodies (Raybould *et al.*, 2019; Shehata *et al.*, 2019) so, the hydrophobicity of germline-encoded CDR1 and 2 was also considered, as different V genes may encode CDR loops with similar hydrophobicity scores. The melting point data for all 267 TCRs in this dataset showed no significant correlation to the hydrophobicity score of any of the CDR loops (Figure 4.2). Antibody CDRH3 loop length (generally longer than those of TCRs) has also been considered indicative of poor stability (Rabia *et al.*, 2018; Wong, Leem and Deane, 2019), so the impact of both alpha and beta CDR3 length was also investigated. As both loops sit within the  $V\alpha:V\beta$  interface, the combined length and ratio of loop lengths (data not shown) were also considered as possible predictive factors for overall TCR stability; long loops may disrupt packing at the interdomain interface, or short loops be unable to make all interdomain contacts and still maintain pMHC binding. However, no trends between these factors and TCR melting point were significant indicating that the range of CDR3 lengths present within this dataset were all equally tolerated within the TCR structure.

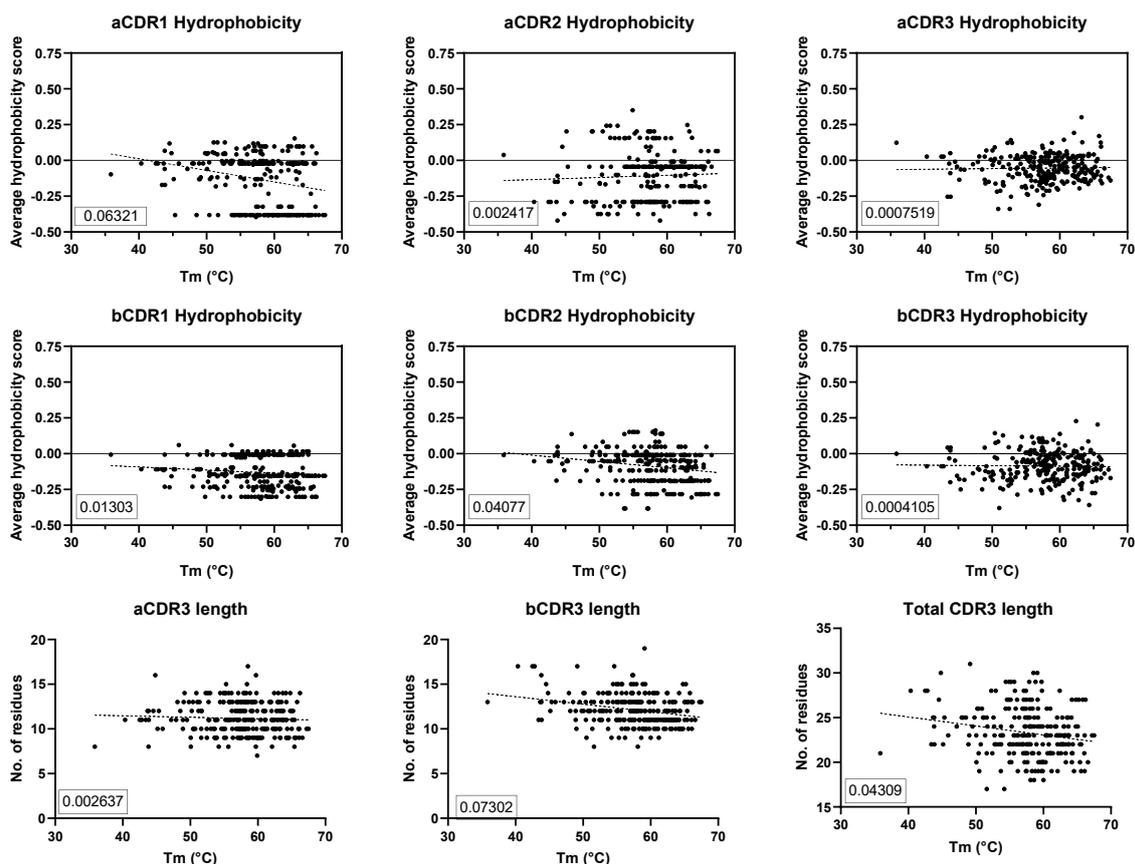


Figure 4.2. CDR features and TCR stability. Top two rows show melting point ( $T_m$ ) for 267 TCRs plotted against average CDR loop hydrophobicity as calculated by the summed hydrophobic index score for each amino acid (according to the values assigned by Sweet et al (Sweet and Eisenberg, 1983), where a more positive number indicates a more hydrophobic amino acid). Bottom row shows  $T_m$  plotted against number of residues in the variable CDR3 loops (as defined by IMGT positions 105 to 117). Dashed lines are simple linear regression fits for all data points, with square of Pearson correlation coefficient ( $R^2$ ) value for each fit as boxed insert.

#### 4.2.4 Structural features of TCRs which may alter stability

CDR hydrophobicity did not show any significant impact on overall thermal stability measurements, but the six CDR loops are only a small portion of the overall exposed surface of a TCR molecule; other surface exposed hydrophobic residues could be involved in driving aggregation and lowering stability. The buried core residues of the TCR are much more hydrophobic than those which are exposed to solvent, but as discussed in Chapter Three there is still some diversity across all V genes for residues

structurally classified as buried. Disruption of hydrophobic core packing in the immunoglobulin-like fold may also disrupt overall protein stability, or a highly hydrophobic core may increase the melting point for a TCR. Surface exposed and buried residues of the protein were identified based on the structural classification of IMGT-aligned residue numbers discussed in Chapter Three, and mean hydrophobicity of each class of residue calculated as above. As seen in Figure 4.3, there is very little variation amongst all TCRs for the hydrophobicity score of surface exposed or buried core residues, despite the sequence-diverse nature of surface exposed residues. No correlation was observed between  $T_m$  and hydrophobicity at the protein surface or within the buried core.

As discussed in Chapter Three, the interface between  $V\alpha$  and  $V\beta$  TCRs is composed of both conserved (mostly hydrophobic) interactions and interactions between sequence diverse positions of the V domains. Contrary to the two residue classes discussed above, the residues involved in  $V\alpha:V\beta$  interface contacts have a wide range of average hydrophobicity across the TCRs included in this study. Increased hydrophobicity at  $V\alpha:V\beta$  interface residues was hypothesised to aid stability by improving protein:protein interactions (Young, Jernigan and Covell, 1994), but again, the range of hydrophobicity scores observed does not show any correlation with melting point.

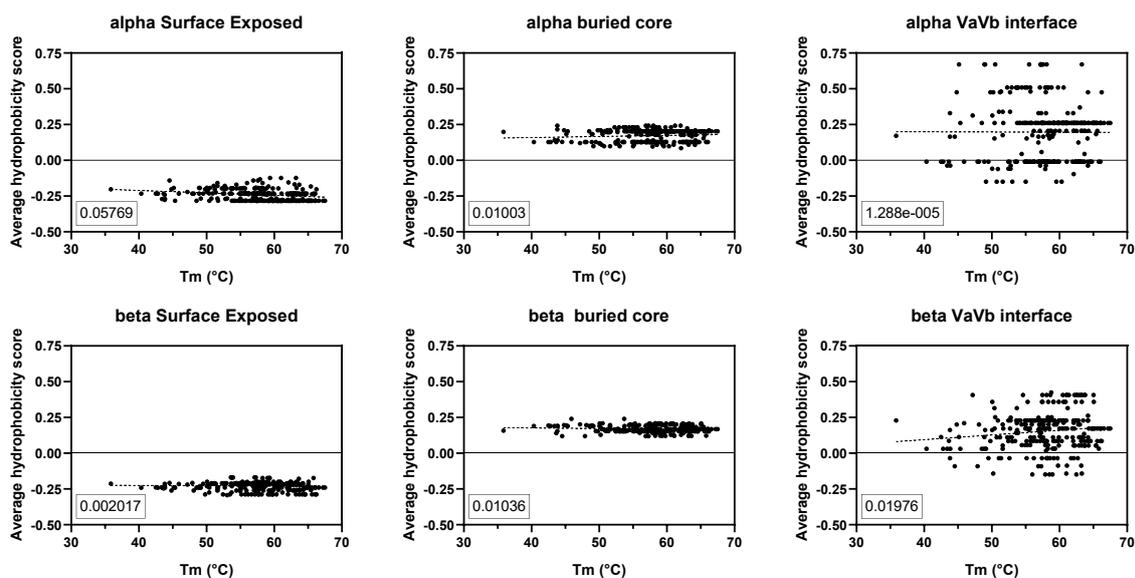


Figure 4.3. Hydrophobicity score and structural features of the TCR. All graphs show melting point ( $T_m$ ) of 267 TCRs plotted against average hydrophobicity of different residue classes (identified in graph title) as calculated by the mean of hydrophobic index score for each amino acid (according to the values assigned by Sweet et al (Sweet and Eisenberg, 1983), where a more positive number indicates a more hydrophobic amino acid). Structural classes assigned based on the structurally informed schema for labelling IMGT numbered residues discussed in Chapter Three. Dashed lines are simple linear regression fits for all data points, with square of Pearson correlation coefficient ( $R^2$ ) value for each fit as boxed insert. CDR3 residues excluded from this analysis as their greater structural variability made structural classification based on residue numbering less straightforward in the absence of structural information.

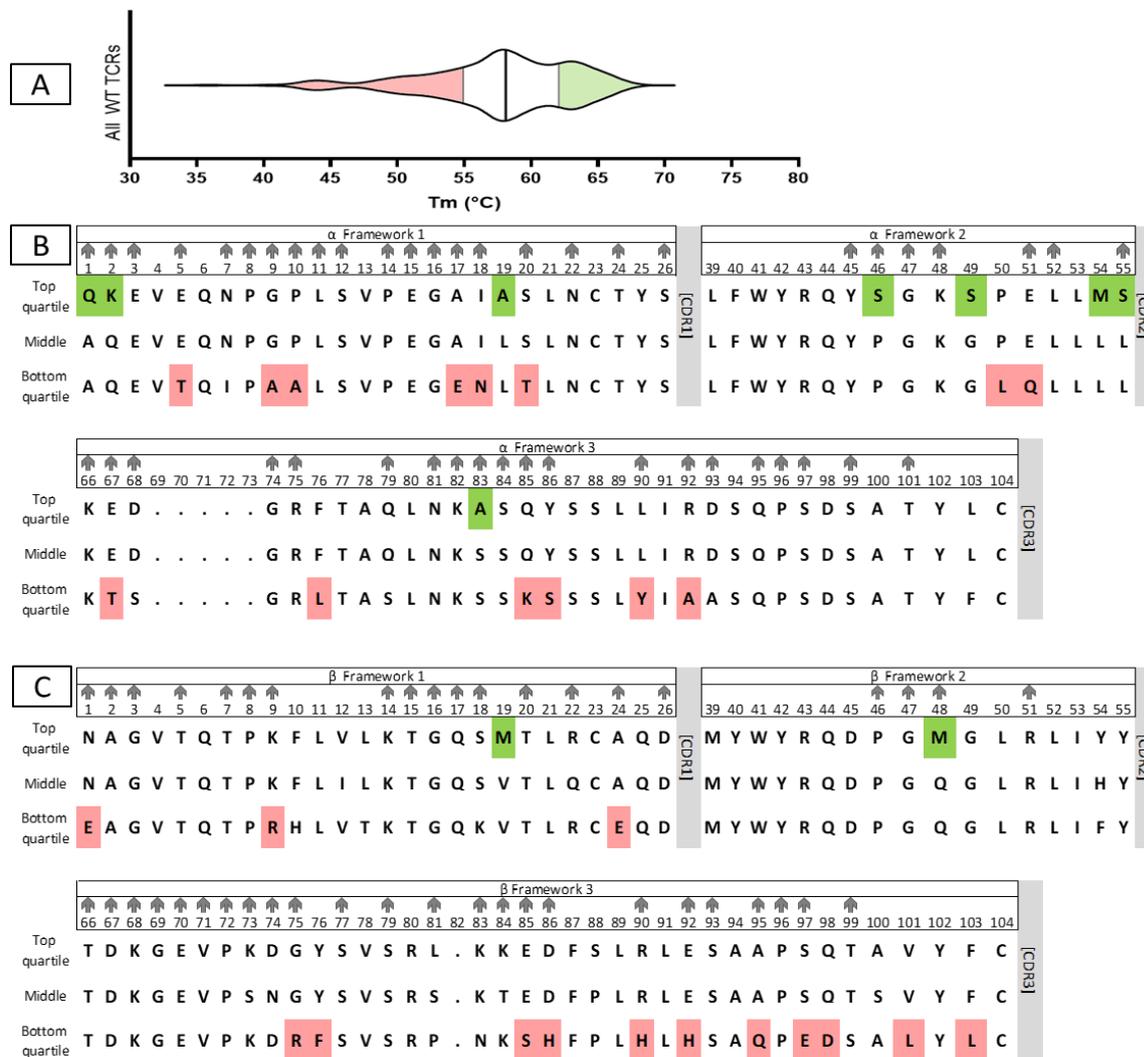
#### **4.2.5 Sequence features that dominate in more thermally stable molecules**

Although wild type TCRs show a wide range of thermal stability, it does not appear as though the TRAV/TRBV usage or the hydrophobicity of different structural features is a strong indicator of stability. The role of amino acid variation across the V domains in driving differences in TCR melting points may be less specifically linked to these structural classes, or may not be fully captured by grouping by V gene usage as some positions are more sequence diverse across all V genes than others.

Therefore, alongside examination of known features of TCR structure, a comparison

of the overall amino acid sequences of all 267 TCRs in the dataset was carried out to identify any differences at the amino acid level that are linked to high or low melting points. CDRs were excluded from this analysis to identify patterns in the diversity that exists outside of the pMHC binding loops.

Wild type TCRs were split into three groups based on the range of melting temperatures observed – those belonging to the top and bottom quartile of the overall range to represent particularly stable and unstable molecules, and the 50% of moderately stable TCRs that make up the middle of the range (as indicated in Figure 4.4A). IMGT aligned sequences were compared within these groups to generate a consensus sequence of the most abundant amino acid for each position amongst all TCRs of that stability group (B&C). Positions where the dominant amino acid varied between these classes were highlighted as significant only when either the most stable quartile or the least stable quartile differed from the remaining 75% (as more likely to indicate influence on stability, as opposed to positions which are hypervariable amongst all TCRs).



**Figure 4.4: Analysis of overall melting point ( $T_m$ ) values and the key sequence differences between high and low  $T_m$  TCRs. (A) Violin plot showing  $T_m$  range for all WT TCRs discussed in this chapter, with bottom and top quartile highlighted in red and green respectively, median  $T_m$  value shown as solid line. **B&C**) comparison of IMGT-numbered consensus sequences for the alpha (**B**) and beta (**C**) framework regions of WT TCRs with  $T_m$  values in the top & bottom quartiles of the dataset, along with the consensus for TCRs in the 25-75% range (middle). Green highlights show amino acids where the consensus is different for only the most thermally stable TCRs, red highlights show where the consensus residue is different only for the least stable. CDRs excluded from analysis. Grey arrows indicate surface exposed residues, as classified from structural analysis in Chapter Three.**

The highlighted residues were evenly split between surface exposed and buried residues (based on the structural classification of positions discussed in Chapter Three). Of particular interest are some residues which occur at the  $V\alpha:V\beta$  interface. One example ( $\alpha 55$ ) is predominantly serine in the highest stability quartile, whilst

leucine is the consensus for the lower 75% of melting points. A  $\alpha$ F55S mutation was observed to improve resistance to thermal degradation for the 868 TCR in scTV format (Aggen *et al.*, 2011), and  $\alpha$ 55S is present in TRAV12-2\*01 but not TRAV12-2\*02 subtype. However, as seen in Figure 4.1, the difference between the two subtypes is not statistically significant; in fact, the median  $T_m$  for the  $\alpha$ 55F TRAV12-2\*02 TCRs is higher than for the  $\alpha$ 55S TRAV12-2\*01 group. Also, at the interdomain interface,  $\beta$ 103L is the dominant amino acid in the lowest stability quartile: the consensus residue  $\beta$ 103F dominates the more stable groups and was identified in Chapter Three as forming part of the highly conserved core of the  $V\alpha:V\beta$  interface, whilst the smaller leucine side chain may not make as strong an inter-domain contacts in the hydrophobic packing interactions between the two chains. No residues at the  $V\alpha:C\alpha$  interface showed any difference in consensus amino acid between the high and low stability groups, and only one residue ( $\beta$ 009) at the  $V\beta:C\beta$  interface. Structural analysis in Chapter Three indicates that  $\beta$ 009R makes a  $V\beta:C\beta$  interface contact less frequently than  $\beta$ 009K, and this may explain the dominance of arginine in the lowest stability quartile.

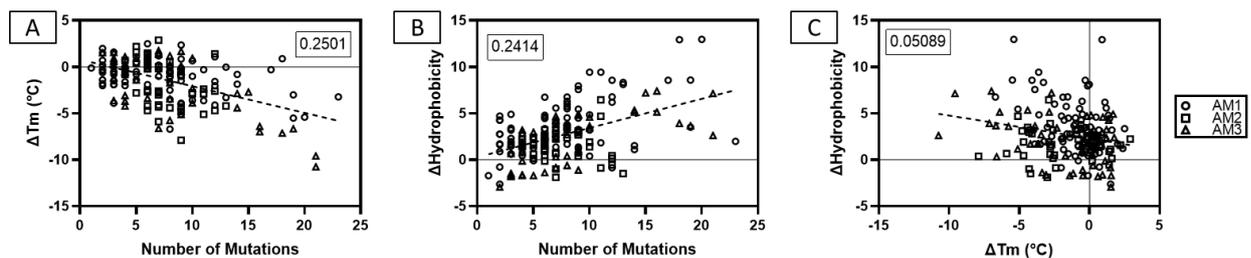
#### **4.2.6 Thermostability changes during affinity maturation of TCRs**

Alongside the wild type sequences discussed above, an additional dataset of affinity matured TCR variants and their melting point change relative to the parent wild type TCR was also assembled. The higher affinity variants of these proteins come from phage display affinity maturation (again carried out by multiple members of the Protein Science department at Immunocore), where CDR loops have been randomised by use of degenerate codons followed by selection for improved affinity, as described by Li *et al.* (Li *et al.*, 2005). High affinity CDRs were then combined and the TCRs refolded for further testing, leading to picomolar affinity but a high mutational load. Successful binding to pHLA requires that the TCR on the phage

surface is correctly folded enough to remain bound during wash steps, so any mutations that completely destabilise the molecule to the point of losing binding function will not be carried through selection. However, less extreme destabilising effects will not necessarily be removed from libraries during this style of affinity maturation panning. Mutations which greatly increase binding to the target pHLA will have a strong selective advantage that may be sufficient to overcome reduced stability, especially as thermal melting point is only an imperfect proxy measurement for aggregation propensity at room temperature. After successful affinity enhancing mutations are identified from these screening approaches, they are combined without further phage selection, so there is no additional screening step to bias the output towards improved stability as well as affinity.

To track the impact of these cumulative mutations (all of which resulted in stronger affinity to pHLA target than the parent wild type molecule), the relative change in melting point ( $T_m$ ) was measured for three different wild type TCRs and at least 20 of their daughter variants. As seen in Figure 4.5A, the majority (63%) of mutated daughter variants have lower  $T_m$  relative to the parent molecule, and this tendency is worse for those TCRs where more mutations have been introduced to the sequence. On average, each mutation away from the wildtype sequence has  $-0.8$  °C cost to  $T_m$ , although a linear correlation between the two variables only explains 25% of the  $T_m$  variability. One explanation for this reduction in thermal stability is that a high proportion of the mutations increase the hydrophobicity of the TCR (Figure 4.5B) as measured by the summed change in hydrophobicity score for each mutation away from the wildtype sequence. Hydrophobic residues at the pHLA interface are likely to be overrepresented in phage display outputs as they can make energetically favourable (although potentially non-specific) contacts with the HLA helices (Kaleli, Karadag and Kalyoncu, 2019), but surface exposed hydrophobic residues are also often implicated in self aggregation which can lower thermal stability (Du *et al.*, 2019). However, Figure 4.5C shows that the most destabilised TCRs in this study were not those with the greatest increase in hydrophobicity index compared to their parent molecule. The overall loss in thermal stability upon introduction of affinity

enhancing mutations does not show significant correlation with the increase in hydrophobicity, but must instead be due to other destabilising effects from mutations. This is consistent with the results from Figure 4.2, where even in the highly sequence diverse CDR3 loops, average hydrophobicity index was not correlated with the melting point of wild type TCRs.



*Figure 4.5. Results from the affinity maturation of three unrelated TCRs (AM1, AM2 and AM3), representing 50-80 unique variants produced from each parent TCR. Figure A shows the difference in melting point for each of these variants compared to the number of amino acid changes from the wildtype parent sequence. Figure B shows the same variants with the overall  $\Delta$  hydrophobicity calculated as the summed change for each amino acid substitution in hydrophobic index (according to the values assigned by Sweet et al (Sweet and Eisenberg, 1983), where a more positive number indicates a mutation to a more hydrophobic amino acid). Fig C shows the change in hydrophobicity index plotted against the change in thermal stability for each mutation. Dashed lines are simple linear regression fits for all data points, with square of Pearson correlation coefficient ( $R^2$ ) value for each fit as boxed insert.*

### 4.3 Discussion

The demonstration that TCRs show a greater than thirty degree range of melting points when refolded as soluble  $C\alpha:C\beta$  disulphide-stabilised molecules reveals that amino acid diversity between TCRs has a significant impact on their overall thermostability. As a class of molecule, some receptors are inherently much more stable than others and this is likely to be reflected in surface expression levels across the TCR repertoire. The high sequence diversity that exists away from any pMHC binding sites may indicate another route for tuning the reactivity of T-cells against

target antigens; well-presented TCRs have an avidity advantage, whilst poorly presented TCRs which are mildly self-reactive may survive thymic selection.

This study sought to identify features of TCR structure or sequence that could be linked to high or low stability by assessing melting points of a panel of soluble TCRs that have diverse sequences. Most commonly proposed methods for predicting TCR stability based on sequence information (V gene usage, CDR hydrophobicity) do not show strong correlation with measured  $T_m$  values in this dataset. This may be a limitation of the small and biased pool of TCRs that were measured as part of this study, as a larger and more diverse dataset may capture V genes not present in this study or reveal statistically significant correlations that were not identified due to low sample size. The lack of correlation between surface hydrophobicity and thermal stability reflects an ongoing dispute in the wider field of protein stability engineering as to the importance of solvent exposed hydrophobic residues. Studies by Schwehm *et al.*, (1998) showed that introduction of hydrophobic residues at the surface of a model protein (staphylococcal nuclease) showed some destabilising effects, but not all such mutations reduced the ability of the protein to withstand chemical denaturation. Later studies showed that hydrophobic side chains introduced into a surface exposed loop position can drive improvement in thermal stability (Islam *et al.*, 2019).

Certain V genes are overrepresented in studies of TCRs; as of December 2020, 108 out of 355 *H. sapiens* TCR structures with assigned IMGT V genes were TCRs encoded by TRAV12-2, with a further 84 of 355 encoded by TRAV8-4 (TCR3d database, Gowthaman and Pierce, 2019), and as discussed in Chapter Three (Table 3.2) I was unable to obtain experimentally determined structures for 17 of the 47 human TRAV genes. This is partly due to the dominance of a small number of highly studied TCRs in published crystal structures, but there may be certain V genes that are more or less well-suited to soluble refolding and production. It is possible that some V genes not studied here do encode regions of notably poor stability and may consistently show low thermal stability independently of other features such as CDR3

sequence, but due to this low stability they have not been refolded in sufficient quantities or with sufficient purity to allow for thermostability measurements; a larger and more balanced dataset would be required to identify if any V genes were strongly correlated with poor thermostability.

However, it is still notable that even within this small study a wide range of melting points can be observed for TCRs which share the same TRAV and TRBV genes, indicating that more subtle and potentially cooperative effects of amino acid diversity within CDR3 and J-gene encoded regions of the molecule are driving significant variation in overall TCR stability. It is also the case that none of the V genes studied here show consistently higher melting point compared to the rest of the dataset, as again there is significant variation within each subset of V gene groupings. This limits the potential utility of identifying “stable V genes” for future engineering, and demonstrates that TCR inherent stability has evolved to exist across a broad range. Related evidence from antibody expression studies also suggests that molecules encoded by the same germline V genes do not necessarily share stability or yield characteristics.

Of particular interest is the lack of support for particular chain pairings being more stable than others. Chain pairing preferences have not been detected yet from analysis of T cell expression repertoires, but instead all paired sequencing results seem to indicate that all TRAV genes can pair equally well with all TRBV genes. Paired repertoire sequencing is more limited in sampling depth than bulk single chain RNA sequencing, due to the need to either manipulate single T-cells directly into a high-throughput microfluidics based sequencing system (Spindler *et al.*, 2020) or carry out a combinatorial analysis of parallel alpha and beta bulk sequencing (Howie *et al.*, 2015); it is therefore possible that insufficient numbers of pairs of V (and J) genes have been sampled to statistically determine more subtle overall pairing preferences. It is also relevant to note that high throughput comparison of TCR surface expression levels with V and J gene pairing has not yet been established; although there may be no detectable pairing bias based on mRNA expression, some

pairs may still be better represented as functional cell-surface TCRs than others. If overall surface expression levels tend to correlate with TCR stability, variation in stability amongst different chain pairs may fine tune the sensitivity of the immune system to different targets based on a germline-encoded binding preference. As discussed in Chapter One, some pairings of antibody chains seem to express at higher levels than others in scFV format which may indicate a chain-pairing stability advantage (Lloyd *et al.*, 2009; Chen *et al.*, 2015).

However, the stabilities of different chain pairings observed here do not seem to show any pairings of significantly higher or lower stability. Although the available data do not cover all possible chain combinations, the wide range of  $T_m$  values observed for soluble TCRs which share the same chain pair implies that even if any pairing preference exists it is not the main determining factor of TCR thermal stability.

The comparison of wildtype TCRs with varying degrees of protein stability (Figure 4.4) revealed critical residues within the non-CDR framework that offer highly plausible targets for stabilising mutagenesis. Research by Thomas *et al.* (2019) suggests that transferring key amino acids from well-presented TCRs to less dominant chains improved T-cell surface expression, and this is a potential future avenue for stability engineering. As all six CDR loops are frequent targets of mutagenesis to increase affinity, identifying indicators of high or low stability in non-CDR regions also has the potential to be more widely applicable for high-affinity soluble TCR based drugs, as it will allow for stability engineering which is less likely to impact affinity. The trade-off between affinity and thermal stability is a known issue in the engineering of higher affinity antibodies (Julian *et al.*, 2017; Rabia *et al.*, 2018) and the data presented in Figure 4.5 for TCR affinity engineering reflects a similar trend. The nature of phage display affinity maturation techniques has been more widely discussed elsewhere (Ledsgaard *et al.*, 2018; Alfaleh *et al.*, 2020) but an acknowledged limitation of building libraries with multiple degenerate codons is that mutations which provide a selection advantage often occur in clones with multiple

different mutations, most of which do not contribute to the increased affinity of the TCR variant. These incidental carrier mutations are introduced into the CDR loops which, in theory, are more permissive of mutations than more structurally constrained regions of the immunoglobulin fold. Although CDRs 1 and 2 of TCRs are not natively subject to any extra mutations away from germline sequences (unlike the somatic mutations that occur in antibodies), loops tend to be flexible and able to take on multiple conformations (Armstrong, Piepenbrink and Baker, 2008; Ayres *et al.*, 2016; Wong, Leem and Deane, 2019). Mutations in the central core of a protein tend to be more likely to disrupt the fold and be deleterious to function (Bowie *et al.*, 1990; Nisthal *et al.*, 2019) whereas mutations in flexible loops would be more likely to be tolerated. The overall trend of increased hydrophobicity in CDR loops after the introduction of affinity enhancing mutations is the expected output from a screening process focusing on improving pMHC contacts, as non-specific hydrophobic interactions may increase affinity between the TCR and the target pMHC surface. It is interesting to note that natural antibody B-cell somatic hypermutation has been reported to follow a similar trajectory of initial increase in hydrophobic residues (particularly tryptophan), followed by later mutations that tend to introduce charged or polar amino acids (Clark *et al.*, 2006). This process is hypothesised to reflect the move from generic hydrophobic interactions to more specific target binding. It is also interesting to note that generally mutations away from the wildtype TCR sequence are deleterious to stability. Does this suggest that the CDR loops encoded in germline sequences are optimal for TCR folding? The relatively lower stability of solubilised TCRs relative to structurally similar antibodies may indicate that this is not true; Jain *et al.*, (2017) showed that the majority of clinically relevant antibodies measured in their study had  $T_m$  values that fall within the 60-80 °C range, much higher than that observed for TCRs here. However, as increased mutational load shows a weak but statistically significant correlation with reduced thermostability it seemed that changes in loops can reduce the overall melting point of a TCR. This is not due to increased hydrophobicity but instead appeared to result from more subtle destabilising effects. In contrast, a review of mAb sequences and structure based on

clinical stage therapeutics compared with the natural human antibody repertoire has highlighted low CDR hydrophobicity as one of five key measures for predicting stable antibodies with good developability metrics (Raybould *et al.*, 2019).

The lack of correlation between CDR hydrophobicity and  $T_m$  has implications for future TCR affinity engineering, as minimising disruption to the native TCR structure and stability cannot easily be predicted from sequence features alone. However, antibody mutation and affinity selection strategies that also included selection for improved stability can generate molecules which can compensate for destabilising but high affinity loops (Julian *et al.*, 2017), and in subsequent chapters I will address strategies for mutating TCRs for greater thermal stability.

# Chapter Five

## Optimising TCR stability through directed evolution

---

### **5.1 Introduction**

As discussed in Chapter One, T cell receptor protein stability is typically poor when made as a soluble molecule and modifications such as introduction of disulphide bonds between the constant domains (Boulter *et al.*, 2003), or the fusion of leucine zipper domains to the C termini (Willcox *et al.*, 1999), were found to be necessary to stabilise the overall structure.. However, as demonstrated in Chapter Four there is still a high degree of variation in stability between different TCRs, with up to a 30 °C different in melting point (a measure of conformational stability, as discussed in Chapter One it has often been found to be predictive of other important developability metrics). As no framework regions or features showed obvious stability advantages over others, it seems that TCR stability is a feature of multiple interrelated aspects of sequence and structure that cannot be easily predicted. In addition to work in Chapter Four that attempted to identify any framework sequences that have intrinsically higher stability than others, I will also investigate the possibility of engineering framework sequences to introduce stability enhancing mutations. This TCR-specific approach allows for stability enhancements that work in the context of a known TCR of interest, which is an attractive methodology for targeting molecules that have therapeutically relevant affinity or specificity properties and allowing them to be engineered into a more stable form better suited to therapeutic use.

#### **5.1.1 Historical approaches to TCR stability enhancement**

As discussed in Chapter One, multiple approaches to TCR engineering have been successfully exploited in previous studies to identify mutations which improve

affinity, specificity, or soluble expression of TCRs in varying formats. In particular, several studies have focused on the use of directed evolution, by first building a library of different mutant variants of a given TCR displayed in a system that allows linking the genotype and phenotype of the molecule, and then panning that library to select for mutations that enhance desirable protein features. Directed evolution to select or screen for mutations with desirable characteristics allows for an unbiased approach, which may produce mutations that could not have been predicted by a more structure-based rational design of amino acid changes. This is possible by randomly introducing mutations across regions of the protein when assembling a library of variants. Previous studies have successfully used this technique for TCRs displayed on yeast (Shusta *et al.*, 2000), mammalian (Wagner *et al.*, 2019) and phage display (Gunnarsen *et al.*, 2013) platforms. Due to the lack of post translational modifications (which may introduce confounding stability effects) I will use a phage display system to express and screen my TCR libraries; this system is well established for affinity maturation of full length TCRs but may require some adaptation to select for enhanced stability.

#### *5.1.1.1 Deep Mutational Scanning*

Deep mutational scanning is a natural extension of functional mapping techniques such as alanine scans, which explore the impact of mutating individual amino acids in a protein. Across the whole protein (or a subregion of interest), each amino acid in turn is mutated to all nineteen possible alternatives from the standard genetically-encoded amino acids. By randomising each amino acid in this stepwise fashion, the full sequence-fitness landscape can be sampled in a relatively small library, covering all possible point mutations in the region of interest. Deep sequencing of libraries built in this manner and subjected to selection for desirable characteristics reveals not only which specific amino acids are enriched in response to selective pressure, but also allows identification of positions that can tolerate high sequence variability. Mapping this sequence fitness landscape is particularly of interest in engineering a TCR for use as a high-affinity soluble bispecific drug, as the wildtype structure has

evolved to function as a membrane-bound protein typically binding target peptides with micromolar affinity. Improving the affinity of the TCR by purely optimising the contacts to pMHC ignores the potential lack of fitness in the framework and inter-domain interface regions for the strong-affinity, high-stability characteristics that are desirable for a biologic drug.

Harris *et al.* (2016) used deep mutational scans of CDRs to analyse the binding interface of a TCR and identified some highly enriched mutations that improved affinity. Sharma and Kranz, (2018) expanded this technique to some regions of the alpha-beta interface (a total of 44 positions within the single chain V $\beta$ -linker-V $\alpha$  structure), identifying an enriched framework mutation which improved pMHC binding affinity as well as giving some thermostability improvement. In the analogous antibody system, Koenig *et al.* (2017) carried out a deep mutational scan across the entire IGHV and IGLV domains of an antibody, identifying enrichment for an antigen-distal mutation which increased both antigen affinity and overall stability of the Fab. Less systematic mutational approaches (such as error prone PCR) have previously been applied to the non-CDR region of the TCR with the aim of improving TCR stability or surface expression in various display formats (Richman *et al.*, 2009; Aggen *et al.*, 2011). Deep mutational scanning has the advantage over error prone PCR of covering all possible single amino acid mutations without the bias towards mutations encoded by a codon that is only a single base pair different to the endogenous amino acid, as is the case of error prone PCR (Neylon, 2004). It will also only produce single point mutations without the possibility of enriching TCR sequences with multiple mutations, some of which may be “carrier mutations” that do not contribute to enhanced stability but merely occur in the same clone as other relevant mutations. Although single mutations are generally more limited in their effects than introducing several mutations at once, point mutations enriched in selections that have demonstrated improved stability can later be combined.

### 5.1.1.2 Selective pressure for stability enhancement

Once a library of potential mutations has been assembled and transferred into a display system such as filamentous M13 phage, the next step is to subject all the molecules in the library to conditions that will select for enrichment of variants that have improved characteristics (whatever the desired selection criteria may be). For stability engineering, early experiments in antibody phage display incorporated incubation steps at elevated temperatures or in the presence of proteases, followed by selection for antigen binding. These incubation steps thermally or enzymatically degrade the displayed antibodies, with the most stable variants most likely to retain their correct fold and therefore binding function. Another selection strategy attempted was incubation with low concentration of guanidine hydrochloride, a chemical denaturant that will disrupt the folding of proteins, on the assumption that more stable variants will have greater resistance to this denaturing effect. However, Jung, Honegger and Plückthun, (1999) noted that this approach was less stringent in selection for improved stability as the unfolding effect of guanidine hydrochloride was reversible, allowing unstable antibodies to refold and bind their antigen during the next selection step.

Thermal stress as a method for enriching mutations which improve stability is a common approach in antibody phage display (Jermutus *et al.*, 2001; Dudgeon *et al.*, 2012; Enever, Pupecka-Swider and Sepp, 2015). Gunnarsen *et al.* (2013) investigated thermostability selection for single-chain format TCRs in a phage display format using incubation of the phage at high temperatures followed by selection using capture by a CDR conformation-specific monoclonal antibody unique to their murine TCR.

Although phage display followed by affinity selection requires that the protein is at minimum stable enough to bind the target and withstand any washing steps, other eukaryotic display systems have been shown to have a greater bias to preferentially display more stable molecules at higher levels; Shusta *et al.* (1999) demonstrated that higher levels of yeast display corresponded to increased thermal stability and

increased soluble secretion of mutated single chain TCRs. Sharma and Kranz (2018) expanded on this via exposure of their yeast-displayed deep mutational scan library of a TCR to 45 °C incubation step to thermally denature less stable variants. They then attempted two different post-incubation selection experiments, both using the target MART-1 pHLA and also a conformationally specific antibody that bound the TRBV14 beta chain of the TCR of interest, and identified enrichment for several mutations that improved thermal stability.

The stability panning approaches described above often involved a selection step that is antigen-independent, in order to prioritise identification of mutations that improve stability without the confounding effect of affinity changes; mutations which greatly enhance affinity would have a selective advantage in selections against the target antigen and might out compete more stable variants. Selection for improved stability would be best achieved with the ability to discriminate between folded and unfolded protein, which has previously been achieved through the use of conformation-specific antibodies (as in the case of the two TCR studies discussed above). However, conformation specific antibodies are not available for all TRAV and TRBV domains so I will explore alternative stability selection strategies.

### **5.1.2 Aims of chapter**

My experimental approach expands upon the work carried out in the previous TCR stabilisation studies described above, in order to identify framework mutations that enhance the stability of TCRs.

As discussed above, establishing selective pressure for enhanced stability is not necessarily straightforward. I will test multiple different selection strategies to establish a protocol that can separate TCRs of higher and lower thermostability.

I will then build deep mutational scan libraries covering the full V domains of two TCRs and use a phage display system to subject these libraries to selective pressure for enhanced thermal stability based on my optimised protocol. By screening two TCRs which share some framework sequence features I hope to find mutations

common to both that enhance stability in a pMHC target- and CDR3 sequence-independent manner.

Mutations identified through the deep mutational scanning library post-selection outputs will be refolded as soluble T cell receptors and further biochemically characterised to identify molecules with greater resistance to thermal denaturation than their parent sequences.

Alongside this phage display work, I will model the impact of all possible point mutation on the conformational stability of the TCR, in order to help identify mutations that may improve stability and to aid in short listing mutations identified from libraries post-selection.

## **5.2 Results**

### **5.2.1 Choice of TCRs for study**

For the full deep mutational scan phage libraries and selections, two TCRs were investigated – Tax-A6 and Kif-B5. The Tax-A6 TCR is a well-studied TCR (Garboczi *et al.*, 1996; Cole *et al.*, 2013; Piepenbrink *et al.*, 2013; Rangarajan *et al.*, 2018) that binds to the 111-19 peptide from the human T cell lymphotropic virus presented by HLA A\*0201. It is known to express and retain pMHC binding function as a full length TCR fused to the pIII coat protein of M13 phage (Li *et al.*, 2005). The V genes used in the A6 TCR (TRAV12-2/TRBV6-5) are highly abundant in the natural TCR repertoire (Howie *et al.*, 2015) and have been used successfully for phage display experiments across multiple different TCRs (Immunocore, unpublished data) . The Kif-B5 TCR, which recognises a disease-specific epitope peptide in the context of HLA A\*0201 (Immunocore, unpublished data), shares the V gene pairing of the A6 TCR, and a crystal structure of a closely related variant is also available (Immunocore, unpublished data). Screening two TCRs concurrently allows identification of any shared enrichment patterns between the two that may be more universal for this subset of V domains.

Additionally, as discussed in Chapter Three, a large proportion of the variable alpha-beta interface consists of J gene encoded amino acids. The two TCRs in this study were encoded by different J genes (Tax-A6 TRAJ24/TRBJ2, Kif-B5 TRAJ23/TRBJ1-2). I will use these TCRs with identical V genes (TRAV12-2 and TRBV 6-5) but different J genes to reveal any J-region dependent framework mutations, particularly at the interface between the alpha and beta chains. These TCRs also bind different peptide sequences, and the docking angles required for optimum peptide contacts also vary between the two.

### **5.2.2 Modelling of potential thermostability mutants**

All modelling of the A6 TCR structure was based on the wild type structure 1AO7. Due to poor density of the TRAC in this structure, a model was built using the high affinity A6 TCR structure 4FTV. Residues C-terminal of V126 (IMGT numbering) were removed from the 1AO7 structure and the corresponding TRAJ and TRAC domain was grafted from the 4FTV structure, followed by energy minimisation of backbone and sidechain of amino acids proximal to this modelled peptide bond (within 8 Å radius) using the default Amber10:EHT forcefield in MOE v2019.0104 (Chemical Computing Group ULC, 2019).

The most similar available structure to the KIF-B5 TCR (internal Immunocore structure solved by Vijaykumar Karuppiah, unpublished) differed from the sequence of Kif-B5 by five amino acids. These amino acid changes were introduced into the structure with the side chains modelled using the protein builder tool in MOE, and the side chain conformations were automatically energy minimised based on the default rotamer library in MOE v2019.0104.

I carried out a saturation mutagenesis experiment *in silico* using these structures to screen for mutations that were predicted to improve protein stability. Residues 1-97 (IMGT numbering) of the V $\alpha$  and V $\beta$  for both TCRs were mutated *in silico* with each mutated structure undergoing energy minimisation to find the lowest energy state of

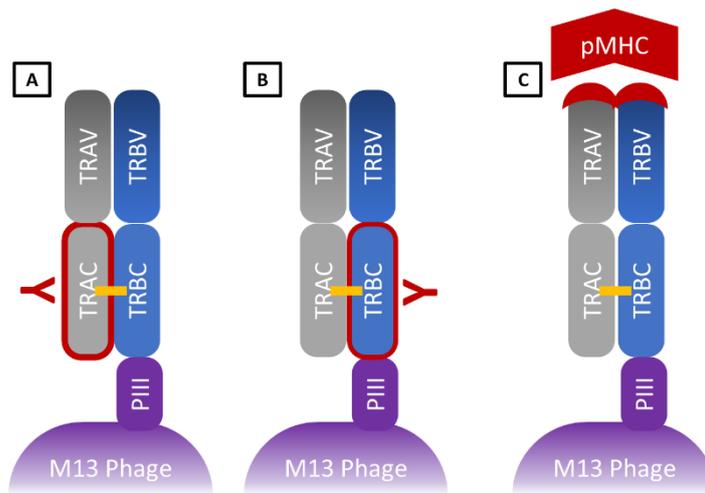
the model. All natural 20 amino acids were modelled for each position, as in the deep mutational scan described above. Mutations were then scored based on the predicted difference in  $\Delta G$  between the wild type and mutant model protein. For both the Tax-A6 and the Kif-B5 TCRs only one position was identified that was predicted to improve the stability of the molecule when mutated. The Tax-A6 modelling suggested that  $\alpha P96L$  and  $\alpha P96V$  would improve stability ( $\Delta\Delta G_{\text{pred.}}$  of -0.41 and -0.21 kcal/mol respectively) whereas for the Kif-B5 structure only  $\alpha P96L$  was predicted to stabilise the TCR ( $\Delta\Delta G_{\text{pred.}}$  of -1.25 kcal/mol).

The low number of mutations predicted to improve the stability of either TCR may have reflected the limitations in *in silico* screening to identify stabilising mutations. I went on to experimentally screen for mutations in the TCR framework that could improve thermal stability using a phage display approach.

### **5.2.3 Establishing selection for thermal stability of TCRs displayed on a phage particle**

Preliminary investigations were carried out to determine optimal conditions to introduce selective pressure for improved thermal stability by identifying a temperature at which known stable and unstable TCRs can be discriminated, as the less stable TCRs will aggregate and lose their normal structure and functional binding ability at a lower temperature.

Multiple TCR capture strategies were also trialled; Figure 5.1 shows a schematic representation of a TCR presented by M13 filamentous phage, with the different possible sites where a “capturing” molecule could bind during panning highlighted in red. Cognate pMHC, or antibodies that bind the phage protein-fused TRBC or the disulphide-fused TRAC domains, could be used to capture phage particles displaying functional TCRs after a thermal challenge step.



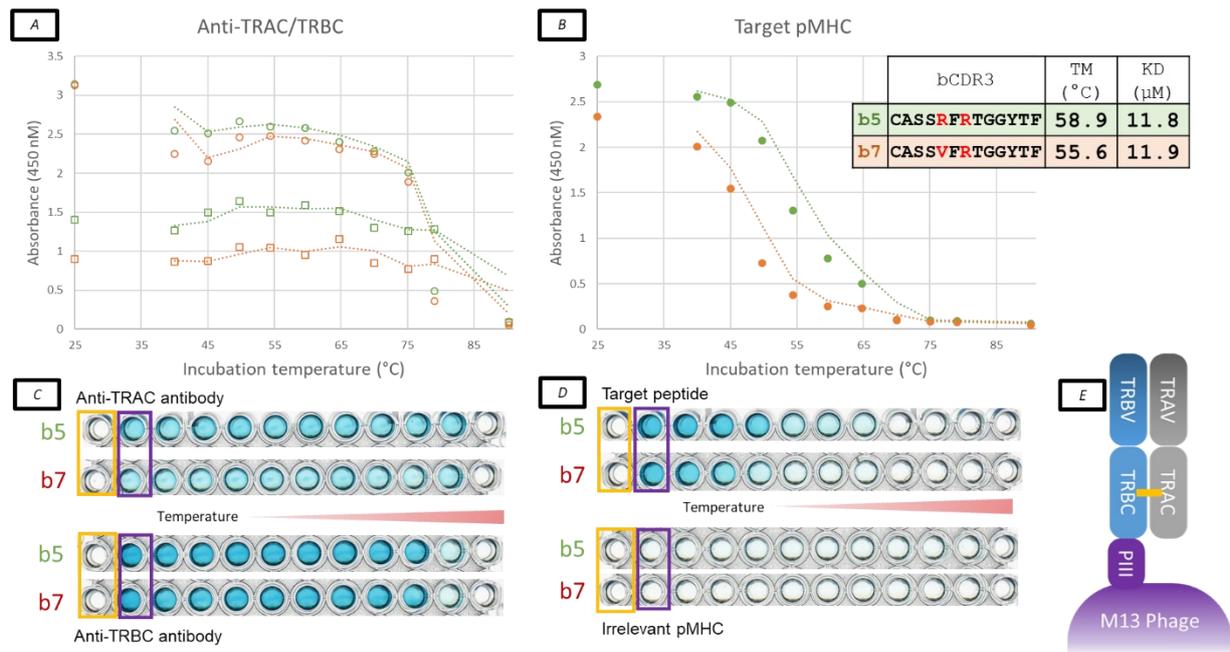
*Figure 5.1 Schematic of the three TCR capture strategies trialled. TCR is displayed on filamentous M13 phage via the TRBC domain, which fused to pIII of phagemid coat, whilst alpha chain associates with the beta chain and is attached by exogenous inter-chain disulphide (yellow bar). A) illustrates use of a TRAC specific antibody, selecting for the presence of TRAC (red outline), B) illustrates use of a TRBC specific antibody, selecting for the presence of TRBC (red outline) and C) illustrates use of cognate pMHC, selecting for presence of CDR loops in proper binding-compatible conformation (red crescents).*

Two high affinity variants (named B5 and B7) of the KIF-specific TCR discussed above had previously been identified (from an affinity maturation experiment carried out by E. Galfre, unpublished). These TCRs bind their cognate pMHC with similar affinity but exhibit different thermal stability as assessed by differential scanning fluorimetry (DSF); B5 has a melting point ( $T_m$ ) 3.3 °C higher than that of B7. I expressed both TCRs separately as pIII fusion proteins and purified phage particles were incubated for 30 minutes at a range of temperatures. The phage particles were then added to an enzyme linked immunosorbent assay (ELISA) where the wells were coated with different possible capture molecules (cognate pMHC, anti-TRAC antibody, anti-TRBC antibody). ELISA signal was developed using an anti-M13 phage antibody; signal reflected amount of phage-presenting TCR that had withstood incubation at different temperatures and been bound by the different capture molecules.

This thermal stress was intended to cause unfolding and aggregation of the TCR displayed on the phage surface, resulting in loss of functional TCR. The more

thermostable B5 variant was expected to denature at a higher temperature than the less stable B7, based on the melting points of both proteins when they were refolded as soluble molecules. However, both TCRs show limited reduction in ELISA signal when measuring binding to TRAC-specific antibodies after phage heating, even at much higher temperatures than the protein melting points of 56 and 59 °C (Figure 5.2A). This indicates that no loss of alpha chain from the surface of the phage occurs during thermal denaturation, probably due to the presence of the disulphide bond between alpha and beta constant domains tethering this domain to the phage particle even after the TCR unfolds. Similarly, binding to the anti-TRBC antibody shows that the beta chain remains fused to the phage pIII protein even after incubation at 80 degrees. In addition, no difference in ELISA signal was seen between the two TCRs across the temperature gradient despite the 3 °C difference in their melting temperature when measured as refolded dimers.

Conversely, functional binding to the target pMHC shows a clear temperature dependence. Figure 5.2B shows that the more thermally stable B5 TCR retains ELISA signal for pMHC binding at a higher temperature than the less stable B7 variant. Based on this ELISA data 55 °C was chosen as a relatively stringent selection temperature, as roughly half of the signal from the more stable b5 TCR was lost after incubation at 55 °C.



*Figure 5.2. Establishing selection for thermal stability of TCRs displayed on phage particle. [A] and [B] show loss of ELISA signal after incubation at multiple temperatures for two mutants of differing thermostability (inset showing sequence and melting temperature). Binding measured to either [A] anti-TCR alpha constant antibody (open squares), anti-TCR beta constant antibody (open circles) or [B] the target KIF peptide (closed circles). [C] and [D] show the original ELISA plates for [A] and [B] respectively, with negative control wells (coated with milk powder only) boxed in yellow and positive control (unheated phage) in purple. All ELISA plates developed with an anti-M13 phage antibody linked to HRP, confirming no loss of phage particle after thermal challenge. [E] shows schematic of phage display format, as in Figure 5.1.*

I then carried out a preliminary phage panning experiment based on the results of this comparison of screening approaches. Phage particles that presented the stable b5 and less stable b7 were mixed at a 1:1 ratio and then thermally challenged for 30 minutes at 55 °C, followed by a pulldown using the KIF pMHC complex linked to magnetic beads, and infection into *E. coli*. The initial 1:1 ratio was confirmed by Sanger sequencing of the pre-selection library (13/24 colonies b5). After thermal challenge the less stable b7 phage population was reduced to 12.5% of the output. In the absence of a thermal challenge, b7 made up 37.5 % of the post-pulldown output, possibly indicating a display advantage for the more stable b5 variant. This demonstrated the efficacy of thermal challenge at 55 °C to deplete less thermostable mutations in a phage library context, allowing selective pressure for thermal stability to be enforced.

The results of the phage selection experiment summarised in Figure 5.2 indicated that selective pressure for thermal stability in a phage context required a functional pMHC binding screen as part of the panning design. Selection of phage based solely on the presence of TCR alpha chain did not discriminate between fully folded functional TCRs and partially unfolded non-binders, limiting the efficacy of the selection strategy. Therefore, all stability selections from this point will contain an element of affinity selective pressure, with the advantage of preventing selection of stability-enhancing mutations that dramatically reduce binding function, but with the risk that mutations that greatly increase affinity will be enriched independently of their impact on thermal stability. To minimise this risk, a high concentration (100 nM) of target pMHC will be used to pull down phage after thermal challenge, reducing the selective advantage of affinity-enhancing mutations

Having established that I could select for thermostability of TCRs expressed on the surface of phage using a model system I next wanted to extend the approach to look for stabilising framework mutations. This required building deep mutational scanning libraries to introduce mutations across the V $\alpha$  and V $\beta$  domains of a TCR.

## **5.2.4 Building and panning deep mutational scan libraries**

### *5.2.4.1 Library design*

I designed single-site saturation mutagenesis libraries (in which all possible single amino acid variants are present) to cover most of the V gene encoded sequence of both the TRAV and TRBV, with residues downstream of IMGT position 98 left unmutated to aid cloning of different CDR3 regions for the two TCRs in this study. Libraries were built both including and excluding CDR1 and CDR2 residues, and alpha and beta libraries were built separately to minimise risk of multiple-mutated sequences that would confound analysis by deep sequencing (as the V $\alpha$  and V $\beta$  are sequenced separately). An illustration of which amino acids were mutated and their location in the overall TCR structure is shown in Figure 5.3.

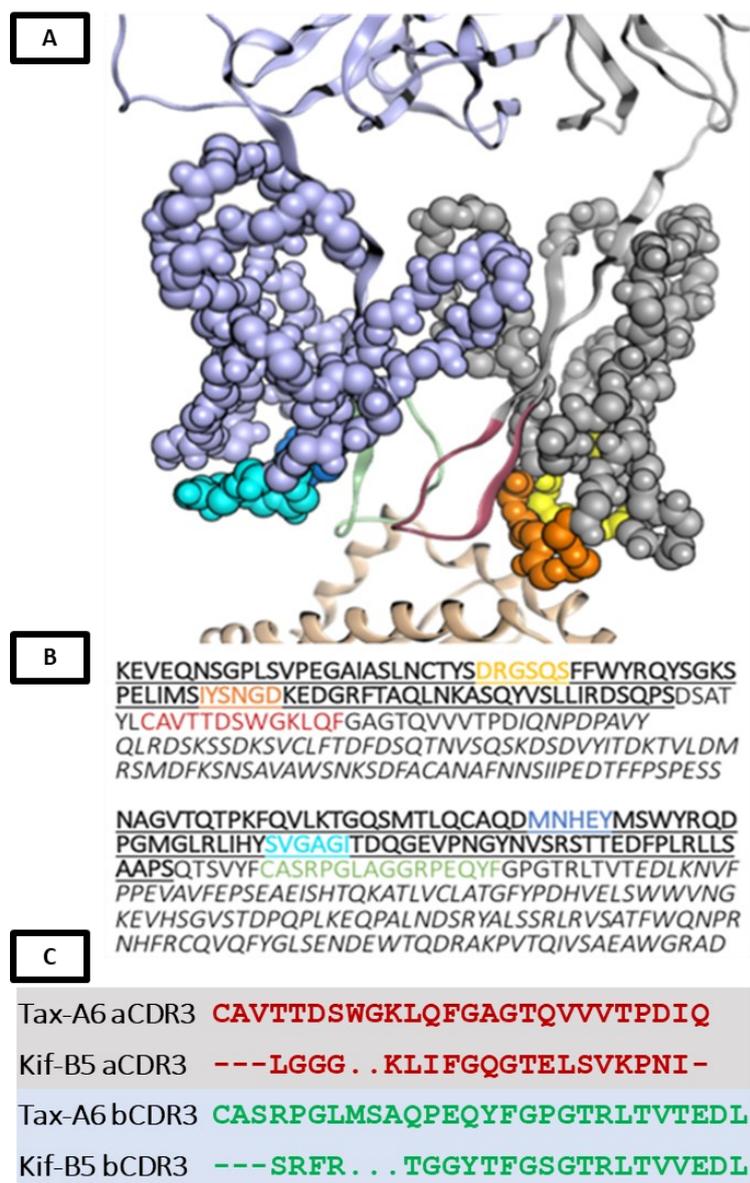


Figure 5.3. Illustration of regions mutated in library design. A) crystal structure of Tax-A6 (based on the composite structure described in the text) showing location of regions of TCR mutated in the library. Spheres represent positions that were mutated in the separate alpha (grey chain) and beta (blue) libraries. CDRs are coloured as follows: aCDR1 yellow, aCDR2 orange, aCDR3 red, bCDR1 blue, bCDR2 teal, bCDR3 green. B) sequence of Tax-A6 shows mutated region underlined and in bold, with same colouring scheme. [C] Alignment of CDR3 and J-region sequences of the two TCRs studied in this chapter. Conserved residues indicated with a dash, dots indicate gaps due to varying CDR loop length. As in structure, CDR3 $\alpha$  is in dark red and CDR3 $\beta$  is in green.

Phage based libraries were constructed as described in Dunn *et al.* (2006) for the two TCRs, with a library size of over  $5 \times 10^6$  as assessed by colony count. The theoretical library diversity for single position saturation library of alpha or beta V genes is  $20 \times$

(number of residues), so these libraries represented sufficient coverage to fully sample the theoretical diversity for alpha ( $1.6 \times 10^3$ ) and beta ( $1.7 \times 10^3$ ) libraries.

### **5.2.5 Enrichment of stability enhancing mutations**

The selective pressure used to enrich for mutations with increased thermal stability was chosen based on the preliminary experiments described in section 4.2.1. Briefly, purified phage libraries were incubated at 55 °C for 30 minutes, followed by pull down of TCRs which had resisted thermal denaturation and retained functionality by using magnetic beads coated in 100 nM target pMHC for two rounds of selection. Library output was analysed by NGS after the second round of selective pressure and assembled and aligned with IMGT numbering using the MiXCR pipeline (Bolotin *et al.*, 2015). Sequencing, alignment and analysis carried out with the assistance of S. Malla.

A more stringent selection regime (incubation at 75 °C) was also carried out but resulted in >50% loss of functionality due to strong enrichment of truncated sequences (as assessed by sanger sequencing of libraries after second round of panning). Based on this result, no further investigation was carried out on these outputs; all subsequent discussion concerns the libraries panned at 55 °C.

Table 5.1 shows the overall mutational enrichment for both TCRs as the percentage of the total library that contained a mutation at each position. This enrichment score was summed for all residues in each of the structural subcategories listed and divided by the number of residues in each category. Residues were classified by structural role based on the analysis developed in Chapter Three. Figure 5.4 and Figure 5.5 show the full mutational landscape after selection for thermostability, with the % sequence count for each mutation represented as a heat map of enrichment. The concurrent panning of two TCRs has not resulted in similar enrichment patterns in either the alpha or beta V domains, indicating that individual protein characteristics

may influence which mutations give the greatest selective advantage in thermostability panning.

*Table 5.1. Enrichment score for all mutations present after two rounds of selection for increased thermostability, separated according to structural classification of each position. Scoring is weighted to adjust for the different number of residues in each grouping. Colouring is a heat scale where stronger red indicates greater enrichment for mutations in residues that fall into that structural subgroup.*

Structural subgroup	Summed mutational enrichment			
	Tax-A6 V $\alpha$	Kif-B5 V $\alpha$	Tax-A6 V $\beta$	Kif-B5 V $\beta$
V $\alpha$ -V $\beta$ interface	0.16	0.07	0.38	0.10
pMHC contact	0.11	0.32	0.65	1.79
Buried	0.33	0.20	0.31	0.16
V-C interface	0.29	0.48	0.39	0.30
Solvent exposed	1.54	0.76	0.50	0.93
Hydrophobic surface patch	0.24	0.46	0.44	0.47
+ve surface patch	0.07	0.00	0.30	0.69
-ve surface patch	2.12	0.36	0.50	0.58

### 5.2.5.1 Solvent exposed surface residues

Surface exposed hydrophobic patches in antibodies are correlated with poor developability characteristics as these may drive aggregation (Jetha *et al.*, 2018), and regions of charged residues are also linked to problematic stability issues (Mahler *et al.*, 2009). Therefore, the overlap between enrichment of alternative residues on the surface and regions of significant hydrophobicity or charge (as predicted from the structures using the MOE Protein Patch tool) is unsurprising, as the stability

selection pressure described above will bias the output towards mutations which mitigate these aggregation prone regions. In particular the alpha chain of the Tax-A6 TCR shows greatest mutational burden at negative surface patches; however as seen in Figure 5.4 the high mutational load seen at position D65 provides the majority of this enrichment.

## **5.2.6 Enrichment patterns in different structural features**

### *5.2.6.1 Solvent exposed surface residues*

Surface exposed hydrophobic patches in antibodies are correlated with poor developability characteristics as these may drive aggregation (Jetha *et al.*, 2018), and regions of charged residues are also linked to problematic stability issues (Mahler *et al.*, 2009). Therefore, the overlap between enrichment of alternative residues on the surface and regions of significant hydrophobicity or charge (as predicted from the structures using the MOE Protein Patch tool) in Table 5.1 was unsurprising, as the stability selection pressure described above would bias the output towards mutations which mitigate these aggregation prone regions. In particular the alpha chain of the Tax-A6 TCR showed greatest mutational burden at negative surface patches; however as seen in Figure 5.4 the high mutational load seen at position D65 provides the majority of this enrichment so this may not reflect a general trend towards mutations at all negative surface patches.

### *5.2.6.2 pMHC contacting residues*

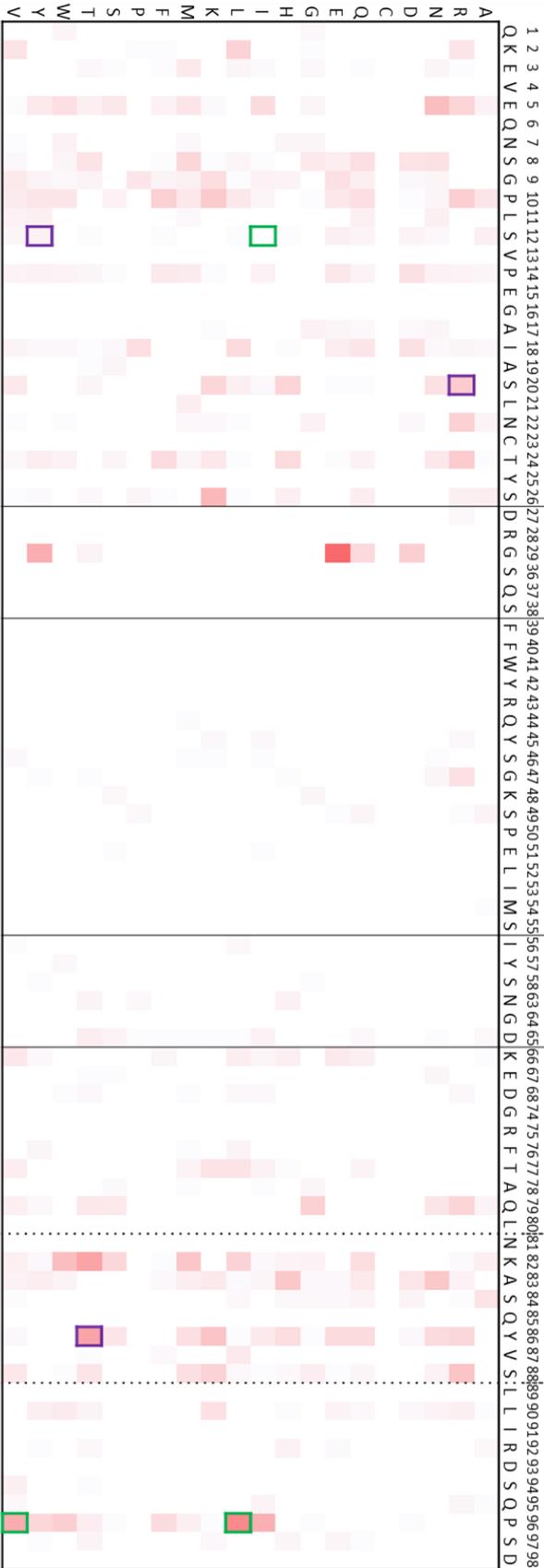
As discussed above, creating effective selective pressure for thermostability required simultaneous selection for retention of pMHC binding. Although the pMHC concentration used for these rounds of selection was not stringent, the selective advantage of improved pMHC binding will still enrich for residues with a higher affinity for the target as well as those with improved thermal stability. Mutations to residues that have been classified as likely to be directly involved in pMHC binding

(as characterised in Chapter Three) were presumed to have been enriched due to enhanced pMHC affinity. This included positions in the CDR1 and CDR2 loops (enriched mutations were noted in these loops, as seen in the heat maps of both Figure 5.4 and Figure 5.5), as well as in the pMHC proximal DE loop.

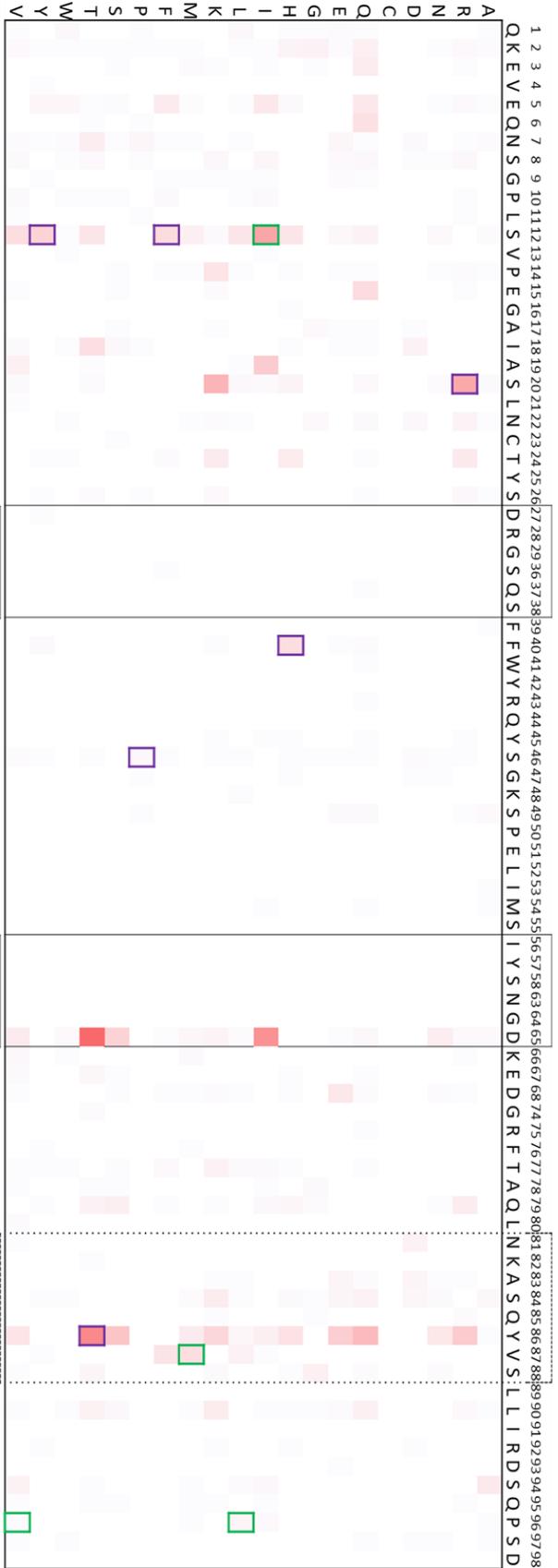
### 5.2.6.3 Enrichment of mutations in the DE loop

Figure 5.4 shows both V $\alpha$  in both the Tax-A6 and Kif-B5 TCR showed enrichment for multiple mutations in the loop between the D and E strands (IMGT residues 81-88, dashed line box), often referred to as the hypervariable region 4 loop (HV4). This loop was shown to be involved in target binding for the TCR of iNKT cells (Paletta *et al.*, 2015), and lysine at position  $\alpha$ 82 in this region has previously been shown in alanine scan experiments to make significant contribution to pMHC binding for the Tax-A6 TCR (Piepenbrink *et al.*, 2013). As seen in Figure 5.4, the Kif-B5 TCR showed enrichment for multiple alternative residues at position  $\alpha$ 82, notably the bulky residues tryptophan and methionine. These mutations likely resulted from the selective pressure for binding to target pMHC rather than for stability. Further supporting this implication, there was limited enrichment seen at this position for the Tax-A6 libraries. From the crystal structure of the Tax-A6 TCR the A6 alpha DE loop is close to the pMHC and the wild type K82 is already contacting the critical HLA-A2 residue E116 (Piepenbrink *et al.*, 2013), potentially minimising the affinity improvements that were possible from a single mutation in this region.

### KIF-B5 alpha



### TAX-A6 alpha



*Figure 5.4. Heat map of Va domain of two TCRs showing mutations enriched after two rounds of selection for improved stability, with wild type sequence and IMGT numbering of each residue in header. Colour indicates percentage of library output for each point mutation seen in chart – stronger red indicates higher enrichment. CDR 1 and 2 residue numbers enclosed in solid black boxes, with pHLA adjacent DE loop in dashed lines. Boxes indicate specific mutations which have been refolded for further analysis, with green boxes indicating mutations which showed improved thermal stability relative to wild type TCR (see Table 2)*



Figure 5.5. Heat map of V $\beta$  domain of two TCRs showing mutations enriched after two rounds of selection for improved stability, with wild type sequence and IMGT numbering of each residue in header. Colour indicates percentage of library output for each point mutation seen in chart – stronger red indicates higher enrichment. CDR 1 and 2 residue numbers enclosed in solid black boxes, with pHLA adjacent DE loop in dashed lines. Boxes indicate specific mutations which have been refolded for further analysis, with green boxes indicating mutations which showed improved thermal stability relative to wild type TCR

## **5.2.7 Biochemical characterisation of enriched mutations**

Enriched residues from the above subcategories were selected for further investigation. Table 5.2 shows the thermal stability of mutated soluble TCRs as assessed by differential scanning fluorimetry (DSF). The enrichment score for a particular mutation did not appear to be strongly predictive of a significant thermal stability improvement.

Table 5.2 also shows the results of the *in silico* modelling described in section 5.2.2 . The most striking increases in measured melting point temperature ( $\beta$ T84L,  $\beta$ R90T) were not predicted to have any stabilising impact, but instead to be deleterious to the conformational stability of both TCRs. This demonstrates the value of obtaining *in vitro* experimental data to validate the output from modelling experiments and indicates that *in silico* ranking of mutations for stability enhancement was not the best approach for analysis of deep mutational scan outputs in the absence of DSF experimental data.

*Table 5.2. Properties of refolded mutants enriched during selection for thermal stability, separated by residue class. Enrichment refers to % of the library that contained that specific mutation after two rounds of selection (coloured using the same heat scale as in Figure 5.4 & Figure 5.5)  $\Delta$ Stability predicted from modelling of mutation as described in text, where a negative number indicated a predicted improvement in stability ( $\Delta\Delta G$ , units (kcal/mol).  $T_m$  measurement experimentally determined using differential scanning fluorimetry (DSF), with  $\Delta T_m$  values calculated relative to the parent molecule.  $K_D$  values determined experimentally using SPR to measure affinity for cognate pHLA. Values for mutant variants shown as a percentage of the  $K_D$  value measured for the parent molecule. Mutations which*

resulted in improved thermostability are bold and underlined. (Asterix indicates data not generated)

Tax-A6					
Mutation	Enrichment	$\Delta$ Stab (pred.)	$T_m$ (°C)	$\Delta T_m$ (°C)	% WT $K_D$
Hydrophobic core					
$\alpha$ V87M	2.55	5.27	53.4	-1.0	232%
Solvent exposed					
$\alpha$ S20R	1.17	1.38	53.7	-0.7	81%
$\alpha$ Y86T	2.05	1.10	*	*	*
$\beta$ N77T	0.11	+0.06	56.8	+2.4	85%
V-C interface					
$\alpha$ P96L	0.22	-0.41	56.5	+2.1	
$\alpha$ P96V	0.06	-0.21	55.4	+1.0	106%
$\alpha$ S12I	0.80	0.35	55.7	+1.3	110%
$\alpha$ S12F	0.38	0.83	54.0	-0.4	73%
$\alpha$ S12Y	0.74	0.66	53.8	-0.6	68%
V $\alpha$ -V $\beta$ interface					
$\alpha$ F40H	0.46	2.05	53.8	-0.6	25%
$\alpha$ S46P	0.72	0.66	53.5	-0.9	96%
$\alpha$ S49E	0.08	3.13	53.1	-1.3	99%
$\alpha$ S49P	0.04	2.16	52.8	-1.6	76%

$\beta$ L50F	0.74	0.04	*	*	*
Kif-B5					
Mutation	Enrichment	$\Delta$ Stab (pred.)	T <sub>m</sub> (°C)	$\Delta$ T <sub>m</sub> (°C)	% WT KD
Hydrophobic core					
$\beta$ V04P	1.11	1.56	61.2	+2.3	*
Solvent exposed					
$\beta$ A24R	6.61	0.77	55.8	-3.1	126%
$\beta$ T84L	0.65	1.63	62.1	+3.2	96%
$\beta$ R90T	0.75	0.97	61.7	+2.8	120%
V-C interface					
$\alpha$ P96L	0.61	-1.25	60.3	+1.4	87%
V $\alpha$ -V $\beta$ interface					
$\alpha$ S49E	0.03	1.73	57.0	-1.9	80%
$\alpha$ S49Q	0.06	0.93	59.1	+0.7	104%
$\beta$ S40N	0.27	1.96	57.0	-1.9	283%

## 5.2.8 Hydrophobic core residues enriched post-selection

### 5.2.8.1 $\alpha$ V87

Tax-A6 showed strong enrichment for an  $\alpha$ V87M mutation, but this had a deleterious effect on the thermostability of the refolded protein by -1.0 °C reduction

in  $T_m$ . Residue 87 protrudes into the hydrophobic core between the DE and CDR2 loops, which may create the 2-fold improvement in affinity observed through optimising loop positioning for pMHC binding. This affinity improvement was likely to be the driving force behind enrichment for the  $\alpha$ V84M mutation in this library. However, a minor  $T_m$  improvement (+0.9 °C) was seen when this mutation was translated onto the Kif-B5 TCR, possibly due to improved packing of the hydrophobic core in this region.

## **5.2.9 Solvent exposed residues enriched post-selection**

### *5.2.9.1 $\beta$ A24*

The Kif-B5 TCR showed very pronounced enrichment for replacement of beta residue A24 with arginine (the highest enrichment for a single mutation seen across all libraries screened in this study, Table 5.2). An arginine residue is very uncommon in the native V gene repertoire at that structural position (only occurring once, in TRBV20-1). Analysis of the crystal structure shows that A24 is adjacent to a region of negative charge created by glutamate 85 and aspartate 86 residues. Limited enrichment was seen for mutations to either of these residues, which implied that they are conserved for functional importance particularly as they are part of the MHC adjacent DE loop discussed above. Given the nature of the single mutation libraries, disrupting larger charged patches may be more efficiently carried out by introducing opposing mutations adjacent to these aggregation prone patches, which was presumed to be the function of the  $\beta$ A24R mutation.

However, despite the strong enrichment observed for the  $\beta$ A24R mutation following thermal challenge of the phage libraries, it can be seen from Table 5.2 that when refolded the  $\beta$ A24R mutant form of Kif-B5 was over three degrees less thermally stable than the wildtype. There was no significant improvement in affinity which may explain the high levels of selection in favour of this mutation. It is possible that

this alteration to surface charge had a beneficial impact on phage display levels or otherwise provided a selective advantage during the library panning process that is not obvious from assessment of the refolded protein.

#### 5.2.9.2 $\beta$ T84L

Panning outputs from both TCRs showed enrichment for hydrophobic residues at this position, with a marked enrichment for leucine seen in Kif-B5 (Figure 5.5) This enrichment of  $\beta$ T84L is counterintuitive, given the presumed aggregation cost from the introduction of hydrophobic residues on solvent-exposed protein surface. No TRBV genes in the TCR repertoire have any hydrophobic amino acids at position  $\beta$ 84; there is instead a strong dominance of charged or hydrophilic side chains. However, when refolded the improvement in thermal stability of T84L mutation was the most significant of all tested; a +3.2 degree shift relative to the wild type Kif-B5 TCR, and a similarly significant +2.1 shift seen for Tax-A6. The mechanism by which this stability enhancement occurs was not clear from analysis of the structure.

### 5.2.10 Variable-constant interface mutations enriched post-selection

#### 5.2.10.1 $\alpha$ P96

Both TCRs show enrichment for hydrophobic residues at position  $\alpha$ P96, a key residue at the  $V\alpha$ - $C\alpha$  inter-domain interface. This residue was the only site where *in silico* modelling suggested that mutations could improve the stability of the two TCRs tested in this study (as discussed in Section 5.2.2 ). In TRAV12-2 chains, residue 96 is already the most highly conserved amino acid (proline) for this position in the alpha variable gene repertoire, but both alpha chain libraries show enrichment for the next most frequent residues leucine (and valine). Mutation of IMGT position 96 to leucine has been previously shown to greatly improve levels of TCR surface

expression in an engineered T cell transduced with multiple mutated TCRs (Thomas *et al.*, 2019).

From examination of the structures of both Tax-A6 and Kif-B5, mutation to leucine could allow the side chain of residue  $\alpha 96$  to come within 3 Å of V1084 on the TRAC domain. Proximity of residues  $\alpha 96$  and  $\alpha 1084$  may be optimising an interaction between the EF loop of the variable domain and the D strand of the constant domain, creating a stabilising hydrophobic interface that led to the thermal stability improvements shown of 1-2 °C for both Kif-B5 and Tax-A6 TCRs.

#### 5.2.10.2 $\alpha S12$

Alpha S12, also proximal to the  $V\alpha-C\alpha$  inter-domain interface, was very tolerant of multiple mutations in the Tax-A6 TCR (Figure 5.4). Selective pressure for thermostability improvement enriched for hydrophobic residues – most frequently isoleucine or valine – or for tyrosine or histidine at this position. However, only S12I mutation showed measurable improvement in stability as measured by DSF. Stability enhancement of S12I mutation was shared by both TCRs.

### 5.3 Combining point mutations

Mutations from Table 5.2 that resulted in an improvement in thermal stability were selected for further study. Multiple studies (Shusta *et al.*, 2000; Richman *et al.*, 2009; Aggen *et al.*, 2011; Koenig *et al.*, 2017; Thomas *et al.*, 2019) show that combining stability enhancing mutations can give an improvement in TCR stability greater than that from an individual mutation. Selected stability enhancing mutations from Table 5.2 were combined in a pairwise fashion and refolded for further analysis as above.

*Table 5.3. Biochemical properties of TCRs with multiple mutations.  $T_m$  measured by DSF, and  $\Delta T_m$  given as the difference in melting point between the parent molecule and the mutated variant, where a positive number indicates an increase in*

*thermostability. Distance between the alpha carbon of the two residues mutated was measured using crystal structure of Kif-B5 or Tax-A6 TCR respectively.*

TCR	Mutations	$T_m$ (°C)	$\Delta T_m$ (°C)	Distance (Å)
Kif-B5	$\alpha$ P96L $\alpha$ S12I	59.23	+0.3	13
	$\alpha$ P96L $\alpha$ V87M	59.78	+0.9	26
	$\alpha$ V87M $\alpha$ S12I	58.33	-0.6	22
	$\alpha$ P96L $\beta$ R90T	61.60	+2.7	42
	$\alpha$ S12I $\beta$ V04P	60.88	+2.0	34
	$\alpha$ P96L $\beta$ V04P	61.79	+2.9	29
	$\alpha$ P96L $\beta$ T84L	61.39	+2.5	39
Tax-A6	$\alpha$ P96V $\alpha$ V87M	55.52	+1.1	27
	$\alpha$ P96V $\alpha$ S12I	55.65	+1.3	14
	$\alpha$ P96L $\beta$ N77T	57.10	+2.7	45
	$\alpha$ V87M $\beta$ N77T	56.00	+1.6	36

All double mutants were screened for thermal stability and affinity changes as above. Table 5.3 shows the results of these combinations; few pairings of mutants tested showed any significant improvement in thermal stability relative to the single mutations shown in Table 5.2. The highest  $T_m$  value achieved for any single mutation to Kif-B5 was 62.2 °C for  $\beta$ T84L, but combining it with the most stable V $\alpha$  variant ( $\alpha$ P96L, 60.3 °C) did not increase stability. Combinations of the  $\alpha$ P96V mutant of Tax-A6 (single mutant  $T_m$  of 55.4 °C) showed some improvement, particularly when combined with the most stabilising beta mutation  $\beta$ N77T.

However, the Tax-A6  $\alpha$ P96L  $\beta$ N77T double mutant only gain 0.1 °C improvement in thermal stability greater than the single beta mutant version.

Generally the impact of these stability enhancing mutations in combination was very modest. Table 5.3 also shows the distance in Å between these mutations (based on the crystal structure of Tax A6, PDB ID: 1AO7). As seen from these measurements all paired mutations tested were sufficiently spatially distant that the likelihood of incompatible clashes is minimal. Improvements in thermal stability based on reduced aggregation propensity could plausibly be combined by minimising aggregation prone patches on more than one area of the protein solvent exposed surface. Increasing stability by the creation of favourable inter- or intra- domain contacts again could plausibly act in an additive fashion to increase overall protein thermal stability. However, no significant such combinatorial effect was seen in the combinations shown here.

## **5.4 Discussion**

In this chapter I have demonstrated the utility of directed evolution experiments to identify multiple point mutations in the framework regions of two TRAV12-2/TRBV6-5 T-cell receptors which give improved thermal stability. I have established a protocol for selectively panning a phage-display library for thermostability enhancing mutations and used this protocol to successfully select single mutations which give up to 3 °C increase in overall TCR melting temperature.

It is interesting to note that the majority of significant improvements to thermal stability arose from mutations in the beta chain rather than the alpha. This may reflect fewer aggregation prone regions in the alpha chain, or fewer such patches dominated by the impact of a single residue that could be targeted by point mutations. It may also be significant that the alpha chain of classical TCRs has limited stability in the constant region, lacking one of the canonical beta sheets of the immunoglobulin fold (Bork, Holm and Sander, 1994). It is possible that this instability drives thermal denaturation of the alpha chain, with the result that few

mutations in the variable domain can outweigh this reduced stability. The beta chain in contrast may thermally denature without this domain bias, allowing multiple mutations on the surface of the beta variable domain to increase overall chain stability. Potentially also of significance, the thermal denaturation curves of the WT and all mutated TCRs show a single denaturation event, indicative of cooperative unfolding of the entire four domain protein.

As discussed in Froning *et al.* (2020), the constant domains of TCRs can be successfully engineered for improved stability. It is perhaps significant that the highest stability achieved from these constant region mutations was a melting point of approximately 61 °C for full length TCRs (as opposed to 73 °C for a truncated C $\alpha$ C $\beta$  only construct). This may represent an upper threshold of stability for the overall full length TCR, as I also have not observed TCR  $T_M$  values in excess of ~65 °C (as discussed in Chapter 4), and the mutated  $\beta$ T84L Kif-B5 TCR described above had a melting point of 62 °C, the highest observed in this study. However, as not all mutations reached the same level of thermal stability, this threshold effect does not fully explain the lack of combinatorial benefit from multiple stabilising mutations. The relatively modest improvements in stability gained from these novel point mutations may simply reflect the limited scope for improvement possible from a single amino acid change.

This limited improvement seen from most paired mutant combinations compared to the single mutations may reflect the library design, in which only single mutations were enriched under selective pressure for enhanced stability so there was no selection for combinatorial effects. The failure of most of these initial combinations of mutation pairs to give a significant improvement in stability compared with single mutations provided limited rationale for further investigations into larger combinatorial effects by introducing more than two point mutations at a time. Instead, in the subsequent chapter I will assess if the beneficial impact of point mutations can be translated to TCRs with either closely related or sequence-dissimilar V domain sequences.



# Chapter Six

## Global approaches to stabilise TCRs

---

### 6.1 Introduction

TCRs evolved as membrane bound receptors that act as part of complex with coreceptor proteins on a dynamic cell surface, so attempts to create soluble versions of the pMHC binding domains required significant protein engineering to achieve sufficient yield and stability (Robinson *et al.*, 2021). As discussed in Chapter One, there are multiple possible strategies for engineering greater stability in soluble TCR formats. Most successful engineering approaches have involved introducing changes to the constant domains (Boulter *et al.*, 2003; Froning *et al.*, 2020; Sádio *et al.*, 2020) which, as they are common to all full length TCR structures, in theory improve stability in a V-domain sequence agnostic fashion. The inter-chain disulphide bond linking C $\alpha$  and C $\beta$  introduced by Boulter *et al.*, (2003) has been widely adopted to stabilise a wide panel of TCRs for further engineering or crystallisation studies. An alternative C-domain stabilising approach was undertaken by unbiased screening for stabilising mutations across both C $\alpha$  and C $\beta$ , resulting in a seven-mutation combination that improved thermal stability and mammalian expression yields for four different molecules (Froning *et al.*, 2020). However, this stabilising benefit varied between different TCRs with different V domain sequences, and as discussed in Chapter Four there is still a wide range of thermal stabilities for TCRs which all have the same disulphide stabilised constant domains. The overall stability of TCRs is therefore not solely driven by constant domains which require stabilisation after cleavage of the transmembrane domains, but instead the diversity across different alpha and beta variable domain sequences results in greater or lesser stability in a TCR-specific manner which may require TCR-specific engineering to overcome. In addition, some TCR formats (such as single chain V $\alpha$ V $\beta$  constructs) lack constant domains entirely and will require alternative stabilising approaches.

Other studies have improved TCR stability using randomised mutagenesis approaches across the V domain and selection for better stability or presentation, but this engineering is necessarily limited to a small number of TCRs which can be screened for TCR-specific mutations. While generally successful in increasing resistance to thermal degradation (Gunnarsen *et al.*, 2013) or the surface expression of different TCRs (Sharma and Kranz, 2018), there is little convergence on common mutational solutions across all the TCRs which have undergone V domain stabilisation engineering. For example, two TCRs sharing a TRAV12-2 encoded V domain were tested for mutations that improved expression on a yeast display system, resulting in mutations at the V $\alpha$  V $\beta$  interface and on the surface of the V domains (Aggen *et al.*, 2011). However, no mutations were enriched that were common to both TCRs, despite the high degree of sequence similarity between them. Direct translation of mutations identified in one TCR to another has been infrequently attempted. Mutations previously identified from scTV yeast display screens of the mouse 2C TCR (Kieke *et al.*, 1999; Shusta *et al.*, 2000) were transferred to two different TCRs that shared either a TRAV or TRBV gene with 2C, resulting in overall improved expression levels for the TCR that shared the stabilised V $\beta$  chain regardless of fusion protein format (Lunde *et al.*, 2010). The other TCR showed improved expression from the 2C-derived V $\alpha$  mutations when expressed as TCR V domains fused to antibody constant domains, but did not stabilise full length V $\alpha$ C $\alpha$ :V $\beta$ C $\beta$  formats either with or without the inter-chain Boulter disulphide. The V $\alpha$  mutations included a mutation at position  $\alpha$ 96 ( $\alpha$ W96R) at the V $\alpha$  C $\alpha$  interface, which in the originating study (Kieke *et al.*, 1999) was presumed to stabilise the newly exposed surface where the constant domain was removed: this mutation is not compatible with stabilising TCRs where the endogenous constant domain is present. Although the ability to outcompete artificially stabilised TCRs on the T cell surface does not necessarily directly correlate to improved stability, a subset of amino acids at key V:C interface positions that were identified from dominantly expressing TCRs were able to increase the surface expression of poorly presented example TCRs

(Thomas *et al.*, 2019). In this study by Thomas *et al.*, some of the mutations designed to make weak expressers more sequence-similar to the dominant TCRs resulted in reduced surface levels, indicating that not all stabilising features are transferrable between different TRAV and TRBV contexts.

TCR-specific stability engineering is a bottleneck in current approaches to creating stable and soluble TCRs for use as protein-based therapeutics. The ability to identify stabilising mutations with limited screening rather than the resource intensive selection screens of large mutant libraries would allow for more efficient use of the broad diversity of TCR sequences without relying on a small number of stable scaffolds. As discussed in Chapter Four, there is a wide range of stabilities possible for TCRs when produced as a soluble molecule, and the developability issues associated with poor stability would limit the pharmaceutical development of some TCRs that might otherwise have positive affinity and specificity attributes. Increased yield and presentation of TCRs will also allow easier biochemical characterisation of novel TCRs, as current structural and binding datasets are still not representative of the diversity present in the immune system.

### **6.1.1 Aims of chapter**

In this study, I will assess the universality of stabilising mutations beyond the specific TCR in which they were first identified. As discussed in Chapter Five, despite the high levels of sequence similarity between the two TRAV12-2 TRBV6-5 TCRs screened for improved thermostability, there was little convergence between the TCRs for which stabilising mutations were enriched after selection. I will attempt to transfer mutations which gave some improvement in melting point on other closely related TCRs which share the same V gene usage, in order to determine if some stabilising effects are independent of CDR or J region sequence or if stability engineering will require a TCR-specific approach for each new molecule. To further expand on this work I will introduce the best stability enhancing mutations onto a larger panel of unrelated TCRs in a pharmaceutically relevant TCR-antibody fusion format (ImmTAC, (Oates and Jakobsen, 2013)). This will hopefully reveal some

stabilising mutations that are generally beneficial independent of the V gene usage, as prior TCR engineering has failed to identify general solutions to improve the stability of V domains.

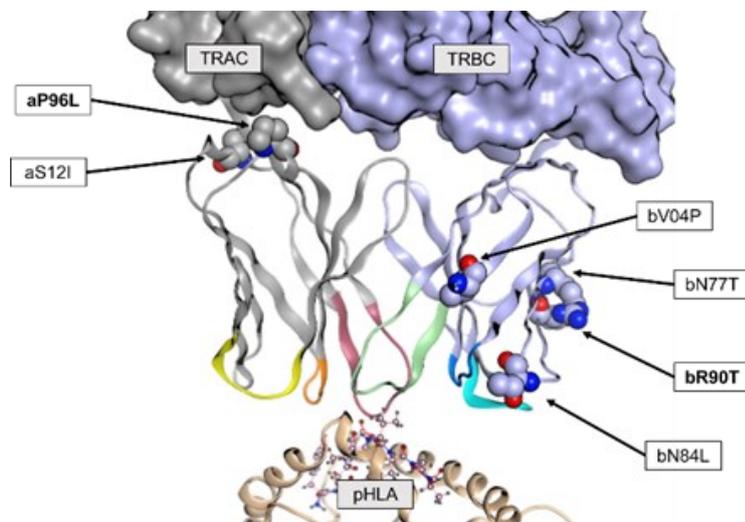
The mutations identified in the previous chapter were only assessed on the basis of improved melting point for the mTCR format, a useful proxy assay for overall stability (as discussed in Chapter One). However, a more relevant measure of stability for TCRs that will be used as soluble drugs is their ability to retain pMHC binding functionality in the complex and degrading environment of human serum at 37 °C. It is also important to ensure that thermostability improvements in TCR format will also be beneficial in pharmaceutically relevant formats, so I will use an *in vitro* serum stability assay in order to identify stabilising mutations that translate well to improved retention of binding activity for an ImmTAC format TCR fusion protein. *In vitro* incubation assays have been shown to be an effective model for *in vivo* antibody stability and allow for identification of stability issues that affect protein functionality (Yang *et al.*, 2018).

## **6.2 Results**

### **6.2.1 Translating stability enhancing mutations onto related molecules**

In Chapter Five I demonstrated that unbiased screening of a saturation mutagenesis library can generate positive hits: point mutations that act to stabilise the TCR as assessed by improved thermal stability. However, the results from this unbiased phage display screen showed limited commonality of enriched mutations between two TCRs which shared the same TRAV and TRBV gene sequences. This is surprising given the location of enriched mutations (as labelled in Figure 6.1); often both distant from the peptide-HLA interfaces (where the two TCRs of interest bound to two different peptides, both presented by HLA-A2\*01) and distant from the V $\alpha$ :V $\beta$

interface (where there is some amino acid differences between the two TCRs due to different CDR3 and J gene sequences present). As discussed in Chapter Three, this interface is predominantly mediated through interactions involving at least one J-gene encoded residue, and as the J-genes used for the two TCRs undergoing phage panning were different it can be assumed that stabilising mutations in this area are more likely to be specific to that particular TCR structure. Most residues at the solvent exposed surface or at the  $V\alpha:C\alpha$  interface would be assumed to be stabilising the molecule through a mechanism which should translate onto other TCRs which share the same amino acid sequence in the area surrounding the mutation site.



*Figure 6.1 Location of stability enhancing mutations shown on structure of KIF-B5 TCR.  $V\alpha$  and  $V\beta$  domains shown as ribbons in grey and light blue respectively, with  $C\alpha$  and  $C\beta$  domains shown as surfaces. CDRs are coloured as follows: aCDR1 yellow, aCDR2 orange, aCDR3 red, bCDR1 dark blue, bCDR2 teal, bCDR3 green. The location of residues where I have identified stabilising mutations are shown as spheres, with the specific mutations labelled in text boxes. Mutations are either surface exposed (bN77T, bR90T, bN84L), buried (bV04P), or near the  $V\alpha$   $C\alpha$  interface (aP96L, aS12I).*

The mutations (identified from my phage display selections discussed in Chapter Five) that showed improvement in thermal stability were then chosen for further testing on closely related molecules that all shared the same TRAV12-2/TRBV6-5 chain pairing. This panel consisted of both the Tax-A6 and KIF-B5 TCRs used as templates for the deep mutational scan approach alongside a third TRAV12-

2:TRBV6-5 TCR (ImmA) specific for an HLA A\*0201-restricted disease-specific epitope (Immucore, unpublished) which was refolded in the ImmTAC format as this allows mutations to the TCR component of a fusion protein to be assessed in the context of a pharmaceutically relevant bispecific drug. The ImmTAC format consists of a fusion of the TCR beta chain to an anti-CD3 antibody in scFV format, allowing for therapeutic use of TCRs as protein drugs that act to redirect T cells for killing.

*Table 6.1: Mutations and improvement to thermal stability for TRAV12-2/TRBV6-4 molecules at different position of the TCR. Asterix indicates mutation was identified using that TCR in an unbiased single mutation library screening for improved thermal stability. All  $T_m$  values are the mean of three replicate measurements, with  $T_m$  change for mutants shown relative to the unmodified parent TCR in each case.*

<b>TCR/ImmTAC</b>	<b>TaxA6 TCR</b>	<b>KifB5 TCR</b>	<b>ImmA ImmTAC</b>
<b>TRAV</b>	AV12-2	AV12-2	AV12-2
<b>TRBV</b>	BV6-5	BV6-5	BV6-5
<b><math>T_m</math> (DSF)</b>	54.4°C	58.9 °C	52.8°C
<b><math>\alpha</math>12</b>	*S→I +1.3°C	S→I +1.1°C	S→I +1.5°C
<b><math>\alpha</math>40</b>	*F→H -0.6 °C	F→H +1.5 °C	F→H (not tested)
<b><math>\alpha</math>87</b>	*V→M -1.0 °C	V→M +0.9 °C	V→M (not tested)
<b><math>\alpha</math>96</b>	P→L +2.1°C	*P→L +1.4°C	P→L +2.9°C
<b><math>\beta</math>04</b>	V→P (not tested)	*V→P +2.3°C	V→P +0.9°C
<b><math>\beta</math>77</b>	*N→T +2.4°C	N→T (not tested)	N→T +0.3°C
<b><math>\beta</math>84</b>	T→L +2.1°C	*T→L +3.2°C	T→L +1.8°C
<b><math>\beta</math>90</b>	R→T +2.6°C	*R→T +2.8°C	R→T +2.1°C

The high levels of sequence similarity between the three AV12-2/TRBV6-5 TCRs tested above make it less surprising that most of the mutations tested were beneficial in more than one TCR. Unusually, the  $\alpha$ F40H mutation was enriched in stability panning for the Tax-A6 TCR but was mildly deleterious to stability for that TCR ( $\Delta T_m$

-0.6 °C) but did improve the melting point of the Kif-B5 TCR ( $\Delta T_m +1.5$  °C). However, this mutation also reduced binding to the target pMHC for both TCRs (25% of the parent TCR KD in Tax-A6 and 89% for Kif-B5) and was not chosen for further studies due to this deleterious impact on binding; as it forms part of the V $\alpha$  V $\beta$  interface the risk of disrupting CDR3 positioning is greater than the more surface exposed mutations. Mutation  $\alpha$ V87M in the hydrophobic core was enriched in selections for Tax-A6 due to the improvement in pMHC binding (232% of parent KD) but again showed a reduction in thermal stability ( $\Delta T_m -1.0$  °C). Although this mutation was slightly stabilising when tested in the Kif-B5 TCR ( $\Delta T_m +0.9$  °C), it did not greatly improve binding (112% of the parent KD) and was not selected for further tests. All other mutations resulted in an increase in melting point for at least two of the three TRAV12-2/TRBV6-5 TCRs tested and exhibited KD values between 90-110% of the parent TCR so were carried forward for further testing.

## **6.2.2 Stabilising mutations can be translated onto unrelated TCR sequences**

The six mutations which improved stability in at least two of the three TCRs were selected for further testing on a panel of ImmTAC bispecific molecules that do not use TRAV12-2 or TRBV6-5, to assess if any of these improvements can be translated into a different context. These molecules (ImmB-H) are all fusion proteins with a CD3-binding scFV fused to  $\alpha\beta$  TCRs specific for different disease-related peptides presented by HLA-A\*0201. The results shown in Table 6.2 indicate that at least some of the mutations can improve  $T_m$  in a V-domain sequence independent fashion. As only one TCR gave a greater than 1 °C improvement in  $T_m$  for V $\beta$  mutations  $\beta$ X04P and  $\beta$ X77T, these mutations were not investigated further.

Table 6.2. Mutations and improvement to thermal stability for unrelated TCR-fusion molecules at different position of the TCR. All  $T_m$  values are the mean of three replicate measurements, with  $T_m$  change for mutants shown relative to the unmodified parent TCR in each case.

TCR	TRAV	TRBV	$T_m$ (°C)	aX12	aX96	bX04	bN77	bX84	bX90
TaxA6 TCR	AV12-2	BV6-5	54.4	S→I +1.3°C	P→L +2.1°C	V	N→T +2.4°C	T→L +2.1°C	R→T +2.6°C
KifB5 TCR	AV12-2	BV6-5	58.9	S→I +1.1°C	P→L +1.4°C	V→P +2.3°C	N	T→L +3.2°C	R→T +2.8°C
ImmA	AV12-2	BV6-5	52.8	S→I +1.5°C	P→L +2.9°C	V→P +0.9°C	N→T +0.3°C	T→L +1.8°C	R→T +2.1°C
ImmB	AV3	BV29-1	63.0	N→I +1.2°C	V→L +1.4°C	I→P -1.07°C	P→T -7.8°C	L	T
ImmC	AV17	BV19	51.1	S→I +1.2°C	A→L +5.2°C	I→P -1.7°C	S	K→L -1.2°C	T
ImmD	AV26-2	BV19	57.3	E→I +0.1°C	L	I	S	K→L -0.8°C	T
ImmE	AV19	BV6-2	55.0	S	V→L +2.1°C	V	N	K→L -1.0°C	G→T +4.2
ImmF	AV5	BV27	59.7		T→L +2.3°C			K→L -1.1°C	I→T +0.3°C
ImmG	AV12-3	BV10-3	59.5		P→L +1.6°C			T→L +0.7°C	T
ImmH	AV24	BV5-8	66.5		P→L +2.8			N→L -1.8	

The most broadly beneficial mutation was at the V $\alpha$  C $\alpha$  interface position ( $\alpha$ 96L) which improved thermal stability in 9 of the 10 molecules tested (the wild type sequence of ImmD already has L at this structural position) with an average  $T_m$  improvement of 2.4 °C. This leucine mutation was already discussed in Chapter

Three as one of the amino acids most frequently making a contact at the V $\alpha$ :C $\alpha$  interface, as one of four key positions that can interact between the two domains.

Mutations at the surface exposed position 90 on the beta chain universally stabilised TRBV6-5 TCRs by 2.1-2.8 °C and also gave the overall highest improvement in melting point for single point mutation in a TRBV6-2 TCR ( $\Delta T_m$  +4.2 °C). This result is particularly interesting as it resulted from a glycine to threonine mutation (rather than arginine to threonine, as in the case of TRBV6-5). The introduction of a threonine side chain at this surface exposed position may be acting to better stabilise surrounding side chains, rather than the stabilising effect being solely due to the loss of a highly charged arginine residue that may be driving charge-based aggregation as was hypothesised for the enrichment of  $\beta$ R90T in Chapter Five. For ImmF the  $\Delta T_m$  of  $\beta$ I90T was only +0.3 °C, much lower than the results from the TRBV6 family TCRs; the local environment for this TRBV27 chain may be less stabilised by the introduction of a threonine side chain.

### **6.2.3 Combining the best alpha and beta mutations**

The improvements in melting point were most consistent for the V $\alpha$  X96L and V $\beta$  X90T mutations, so for those molecules where both gave a positive  $\Delta T_m$  the two mutations were combined. Table 6.3 shows the result of these double mutations on protein stability; for the two TCRs which showed >2 °C improvement from both single mutations (ImmA and ImmE) the double mutant form was +4 and +7 degrees more stable respectively. ImmF had a  $\Delta T_m$  of +2.3 °C for the  $\alpha$ T96L mutation, but only +0.3 °C for  $\beta$ I90T; this smaller improvement in stability may not be sufficient to provide a combinatorial boost as the double mutant exhibited  $\Delta T_m$  of only +1.9 °C, lower than the best single mutation. A  $T_m$  value within 0.3 °C of the parent may be an experimental artifact rather than a substantial stabilising effect, so only stabilising mutations which produce greater than 2 °C increases in  $T_m$  should be considered for combination.

*Table 6.3. Mutations and improvement to thermal stability for unrelated TCR-fusion molecules at different position of the TCR. All  $T_m$  values are the mean of three replicate measurements, with  $T_m$  change for mutants shown relative to the unmodified parent TCR in each case.  $T_m$  improvement for single and combined mutations shown in same row for ease of comparison.*

TCR	TRAV	TRBV	$T_m$ (°C)	aX96L	bX90T	aX96L/bX90T
ImmA	AV12-2	BV6-5	52.8	P→L +2.9°C	R→T +2.1°C	+4.0 °C
ImmE	AV19	BV6-2	55.0	V→L +2.1°C	G→T +4.2	+7.0 °C
ImmF	AV5	BV27	59.7	T→L +2.3°C	I→T +0.3°C	+1.9 °C

#### **6.2.4 Comparison of V-domain framework mutations to published stabilising strategies**

Froning *et al.* (2020) demonstrated improvement in stability from stabilised TCR constant domains, where a total of seven mutations spread across the surface of TRAC and TRBC were identified from a computational screen for stabilising mutations. These stabilised C domains were demonstrated to improve thermal stability and mammalian expression yield when grafted onto four different TCRs with different V gene usage. However, these results were tested on a format that included the endogenous C terminal disulphide between TRAC and TRBC (which in early studies limited expression and solubility of refolded TCR (Garboczi *et al.*, 1996)) as well as the engineered Boulter disulphide. To compare the stabilising effect of these mutated constant domains to the V-domain mutations identified from randomised phage screening, they were grafted onto the TCR bispecifics used above. As seen from Table 6.4, the  $T_m$  improvements generated from introducing seven mutations into the constant domain vary depending on the specific TCR V domains

onto which the mutant constants were grafted ( $T_m$  increases from 3 to 7 °C for four examples shown). Although surprising given that the TCRs in this study all share the same constant domain sequence (so the improvement in stability from a stabilised constant should be consistent), this variation in improvement was also observed in the original study for the four different TCRs tested in that instance by Froning *et al.*

The melting curves for my ImmTAC molecules with stabilised constant domains were biphasic with two overlapping peaks visible in the TCR melt curve (

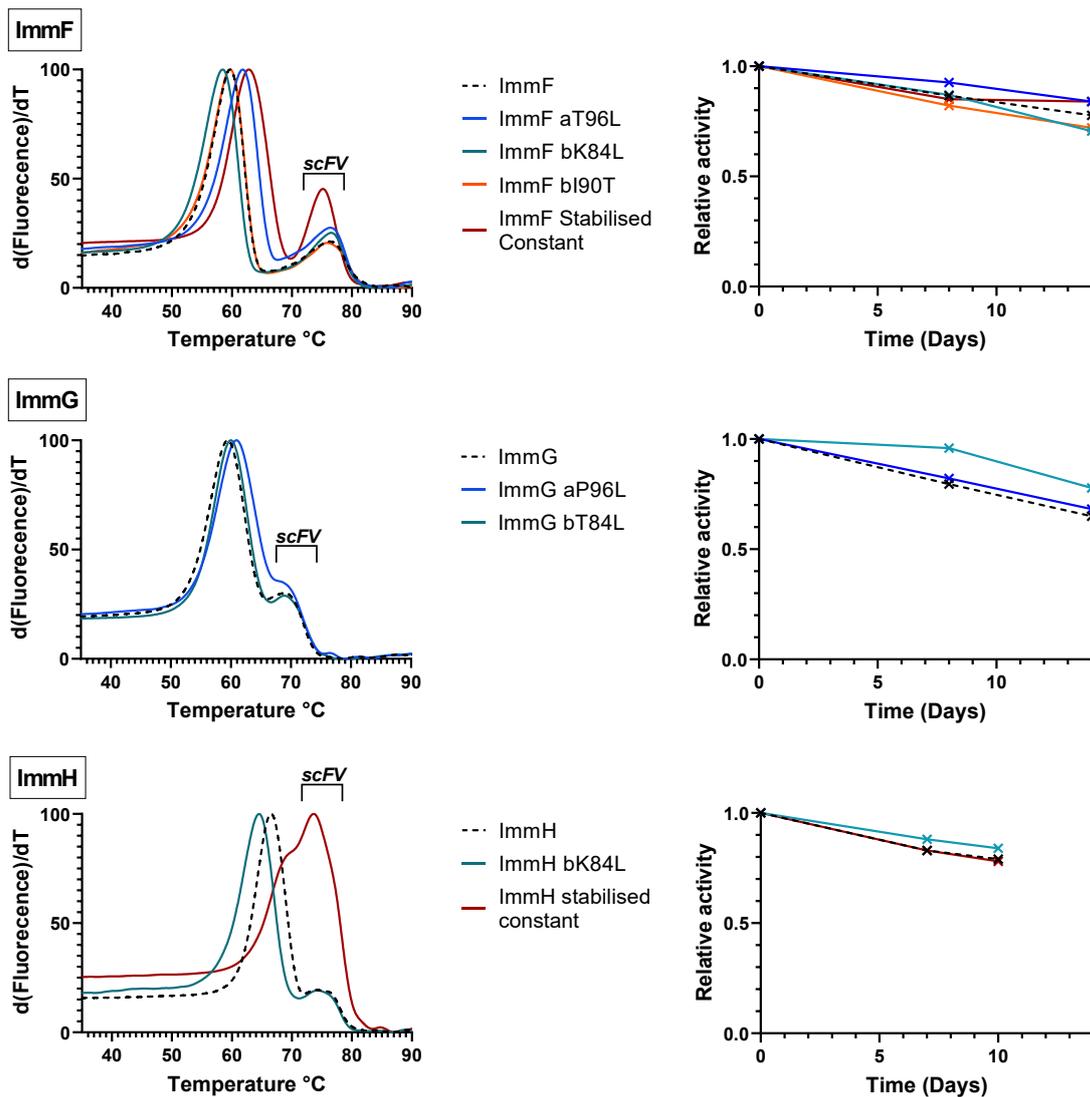


Figure 6.2), indicating that the mutations introduced into the constant domains may not fully stabilise the entire four-domain TCR portion of the ImmTAC, but instead only increase the  $T_m$  for the  $C\alpha C\beta$  portion. This biphasic behaviour was also seen in

the original paper (Froning *et al.*, 2020), and seems to be characteristic of TCRs with these stabilised constant domains.

The greatest improvement in thermal stability observed was for ImmH, where the melting point of the seven-mutation constant form was 7.2 °C higher than the unmodified form. This compares favourably to the  $T_m$  increase achieved from introducing only two V domain framework mutations to ImmE ( $\Delta T_m +7.0$ , Table 6.3). For the ImmA ImmTAC, the introduction of two variable-domain framework mutations was more stabilising than seven mutations to the constant domain ( $\Delta T_m +4.0$  vs.  $+3.1$ ). In addition, the biphasic behaviour indicative of domains unfolding separately was not observed for variable domain mutations.

*Table 6.4 Comparison of thermal stability values obtained from introduction of 1 or 2 V-domain mutations, or 7 C domain mutations. All  $T_m$  values obtained by DSF, carried out on TCRs fused to scFV in ImmTAC format (only  $T_m$  value for TCR portion of fusion protein reported). V domain framework mutations are those discussed in the text. Constant domain mutations are those identified by Froning *et al* (2020).*

TCR	$T_m$ (°C)	TRAV	TRBV	Best single mutation	Best double mutation	X7 Constant domain mutations
ImmA	52.8	AV12-2	BV6-5	aP96L +2.9°C	aP96L R90T +4 °C	+3.2
ImmD	57.3	AV26-2	BV19	E12I +0.1°C	N/A	+4.2°C
ImmF	59.7	AV5	BV27	aT96L +2.3°C	aP96L R90T +1.9 °C	+3.1
ImmH	66.5	AV24	BV5-8	aP96L +2.8	N/A	+7.3

### **6.2.5 Mutations that increase melting point also improve stability in serum**

However, the difference in thermal stability observed for the TCR-scTV bispecifics with stabilising framework mutations did not always confer a dramatic improvement to retention of activity in *in vitro* serum studies (

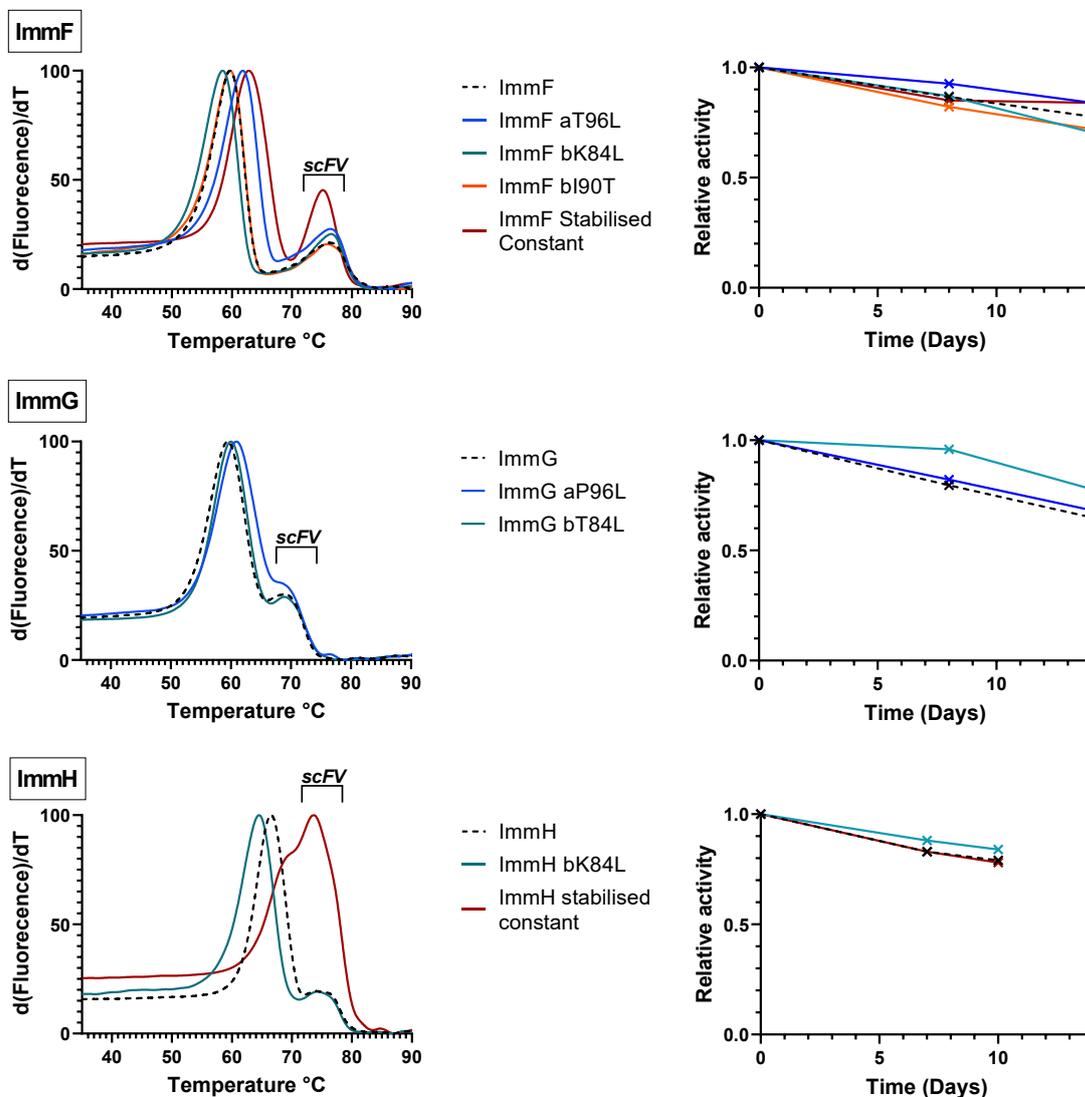


Figure 6.2). Incubation of mutated TCRs in human serum at 37 °C is a pharmacologically relevant measure of overall protein stability and ability to withstand degradation and retain functional binding behaviour in a complex environment. As these molecules were tested in clinically relevant bispecific formats, the activity was assessed bifunctionally to ensure that any global stabilising or destabilising effects from the TCR framework mutations would be captured in the assay. Bifunctional activity was therefore measured by coating ELISA plates with the specific pMHC target of each TCR bispecific and detecting binding of the scTV

portion to a labelled target protein. At least two time points were taken (at 7 and 10-14 days of incubation) for all samples to capture the extent of loss of activity.

As seen in

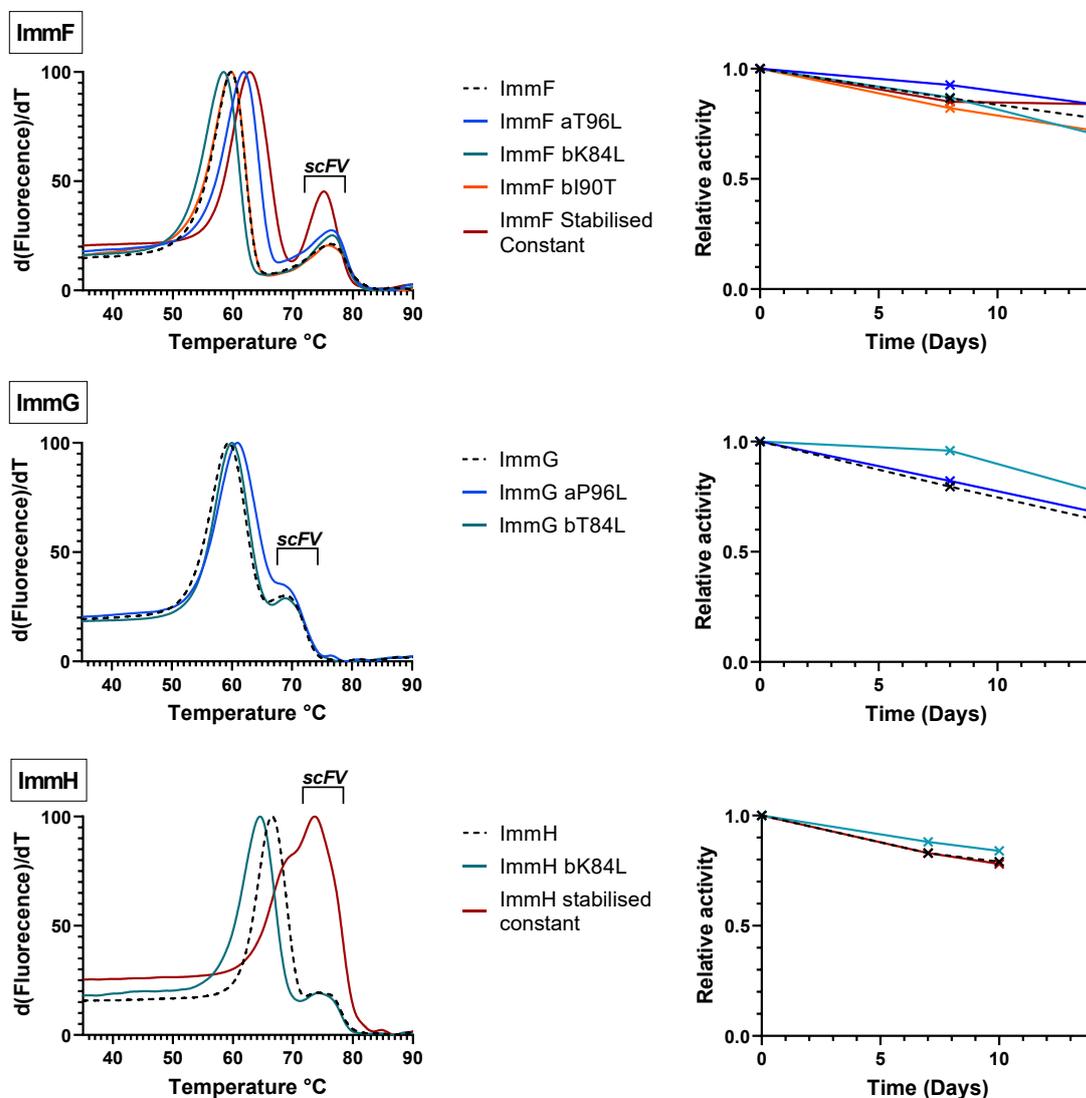
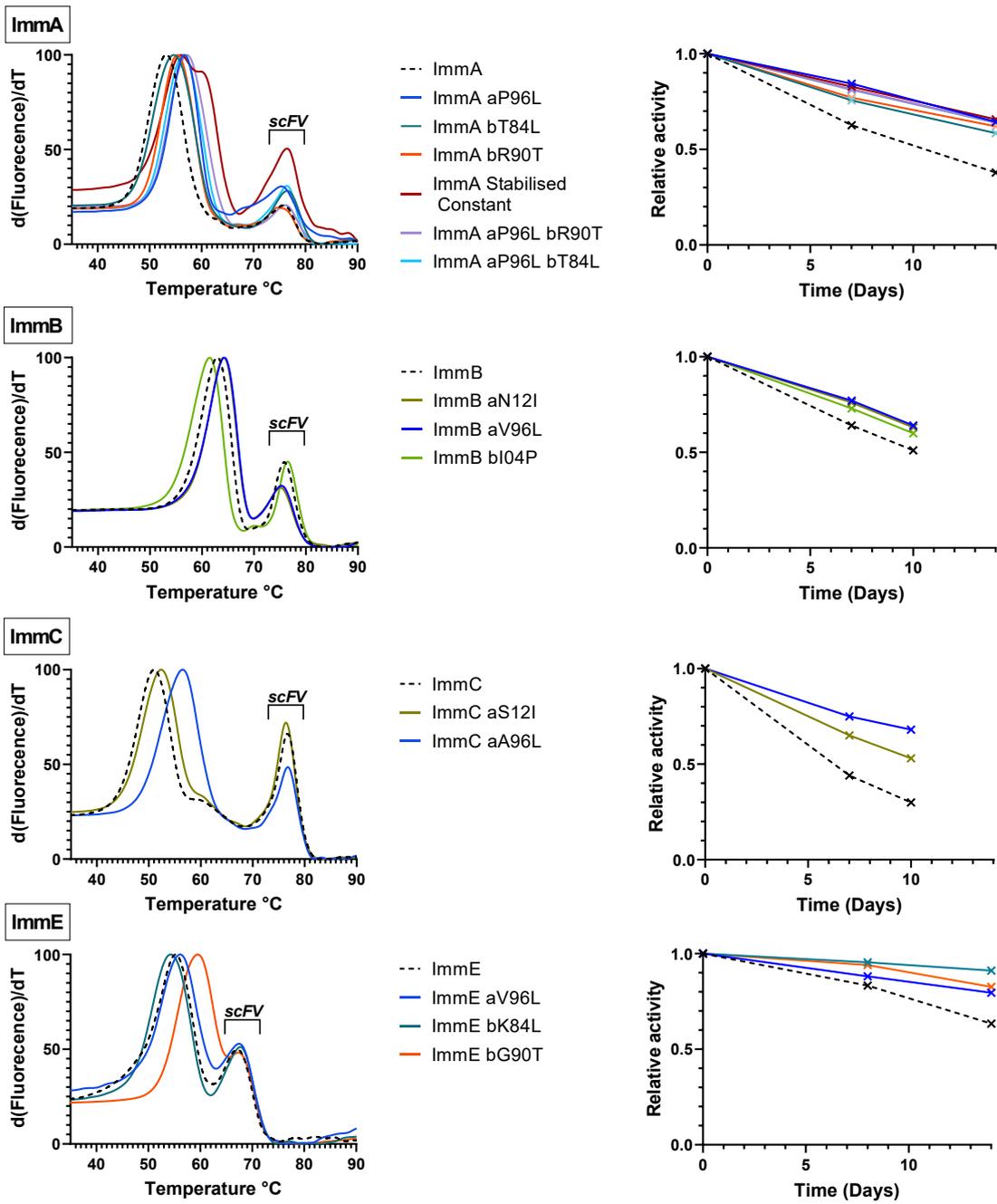


Figure 6.2, all TCRs stabilised with mutations show higher levels of binding activity after serum incubation than the unmodified forms. The stabilising impact is greatest for ImmA and ImmC, which were the two TCRs with the lowest  $T_m$  in their unmodified form (52.8 °C and 51.1 °C respectively). This suggests that functional stabilisation is more effective in those TCRs which exhibit inherently poor stability, whereas those which lose less than 50% of binding activity during serum incubations will show a smaller impact from additional stability engineering.

For ImmF, the  $\beta$ K84L ( $\Delta T_m -1.1$ ) and  $\beta$ I90T ( $\Delta T_m +0.3$ ) mutations were mildly destabilising, retaining only 70% of activity rather than 77% for the unmodified parent, whilst both  $\alpha$ T96L ( $\Delta T_m +2.3$ ) and the mutated constants ( $\Delta T_m +3.1$ ) increased this to 84%. This is consistent with the impact of these mutations on melting point; the minor improvement from  $\beta$ I90T mutation does not seem to translate to functional stabilisation, whereas  $T_m$  increases greater than 1 °C tend to be more predictive of serum stability.



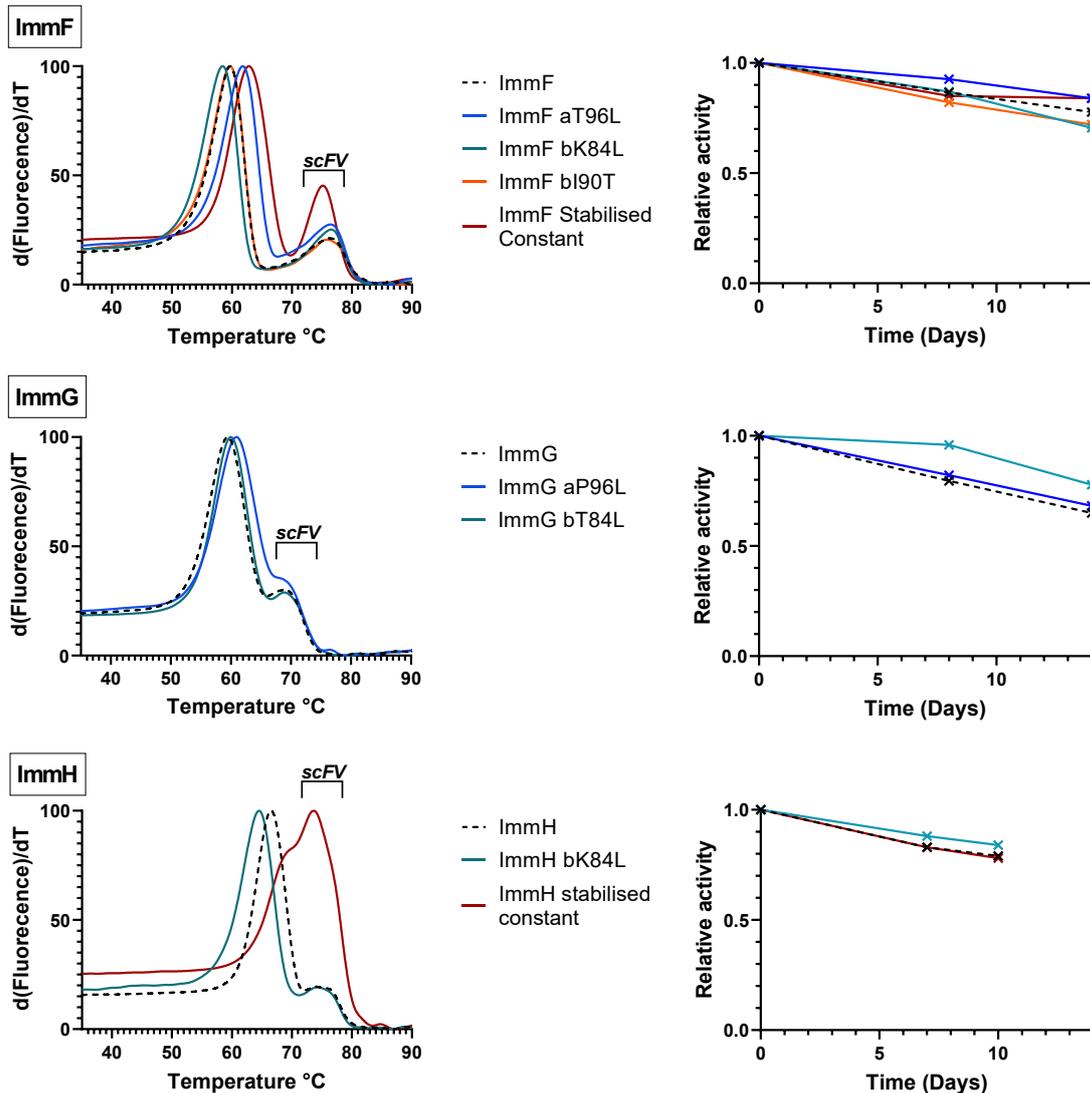


Figure 6.2 Serum stability and differential scanning fluorimetry ( $^{\circ}\text{C}$ ) melt curves for TCR-scFV fusion proteins. Dashed lines indicate unmodified ImmTAC, with stabilised versions labelled individually. Line colouring indicates different mutant versions of the ImmTAC and are consistently coloured for both graphs. Left: DSF curves for ImmTACs and mutant variants are shown as first derivative of raw melt curve data, normalised to set maximal fluorescence at each well as 100%. Brackets indicate the melt curve for the scFV fusion proteins, which are not modified as apt of this study. Right: corresponding serum stability graphs show loss of ImmTAC activity after incubation in human serum at  $37^{\circ}\text{C}$  relative to starting (pre-incubation) activity. Crosses indicate time points during incubation; at least two time points were taken over the 10-14 day period.

### 6.3 Discussion

Unlike antibodies, TCR frameworks did not evolve to act as a soluble protein and require significant engineering in order to overcome poor stability and developability characteristics to make them viable candidates for protein therapeutics. Historically this engineering work has targeted both the sequence-conserved TCR constant

domains as well as the highly diverse variable domains, but TCRs with stabilised constant domains still exhibit a very broad range of stabilities (as discussed in Chapter Three), and, although targeted and often library-based screens for stabilising mutations for specific TCRs have successfully isolated more stable variants, this work requires bespoke engineering for each new molecule. Structural variability exists across the entire  $V\alpha$  and  $V\beta$  domain of TCRs due to the different side chain chemistries of amino acids encoded by different TRAV and TRBV genes, and specific stabilising mutations in V domains have not previously been shown to improve stability for refolded soluble TCRs with unrelated sequences. Limited studies have been published that explore the universality of stabilising mutations away from the sequence conserved constant domains. In this chapter I demonstrate that stabilising mutations identified from a randomised screen of specific TCRs can be successfully translated onto other TCRs, including those with low levels of sequence homology.

The stabilising mutations which translated well into other molecules were located at different structural positions of the TCR. The best broadly beneficial mutation was alpha P96L, a variable:constant interface mutation predicted by my structural analysis in Chapter Three to increase the contacts between  $V\alpha$  and  $C\alpha$ . This structural position was also identified in a study of surface expression on human T cells as common to dominant and well-expressed chains (Thomas *et al.*, 2019). For the single weakly expressed TCR tested by Thomas *et al.* there was a threefold improvement in expression levels when an  $\alpha X96L$  mutation was introduced, with two additional unrelated TCRs also shown to be stabilised by this mutation in combination with two beta chain  $V\beta$   $C\beta$  interface mutants. In my study of soluble TCR-based bispecifics the  $\alpha X96L$  mutation gave an average 2.5 °C increase in  $T_m$  and reduced loss of function in a serum stability assay, both highly relevant metrics for the production and use of protein based therapeutic agents.

The  $T_m$  improvements obtained from introducing single amino acid changes to a TCR ( $\Delta T_m +0.3$  to  $+5.2$ ) were consistent with other reported values for single point mutations – for a single-domain enzyme the majority of stabilising mutations resulted in  $\Delta T_m$  of less than  $3\text{ }^\circ\text{C}$  (Huang *et al.*, 2020), and in a deep mutational scan of an antibody (Koenig *et al.*, 2017) only one mutation was identified with a  $\Delta T_m$  of  $+5\text{ }^\circ\text{C}$  (all other stabilising mutations reported  $\Delta T_m$  between  $0.5$  and  $2\text{ }^\circ\text{C}$ ). I also directly compared the impact of variable framework mutations to published constant-domain stabilising mutations, with the improvements from two of my framework V domain mutations giving comparable stability improvement to those seen from seven mutations to C $\alpha$  and C $\beta$ . Although melting curves for TCRs are typically monophasic, indicative of a single unfolding event rather than independent denaturation of the four domains, this does not necessarily imply that the TCR domains will be evenly stabilised. The melting curves of TCRs stabilised in the constant domain exhibit biphasic peaks which indicate that the improved thermal stability of the constant domains does not translate to overall protein stability. It may be that, for all the TCRs tested both in this study and that of Froning *et al.* (2020), the constant domains unfold independently of the variable domains and that in the absence of stabilised constant domains this multi-stage melting is not obvious from the overall melting profile as the two denaturation events occur at the same temperature. However, two of the constant domain mutations occur at the V $\beta$ :C $\beta$  interface and all seven were isolated from screens of a C $\alpha$ -linker-C $\beta$  construct; these mutations may have disrupted the interface between variable and constant domains and allowed the constant domains to retain their fold as the variable domains unfolded. In contrast, my stabilising framework mutations retained the original TCR monophasic melt curve. This suggests that this stabilisation strategy of V domain mutations seems to improve stability for the TCR as a single unit, which is more advantageous for use of TCRs as a soluble drug. Instability of the variable domains is more likely to disrupt binding function, as the pMHC-contacting CDR loops require

the scaffolding of the V domain framework for correct positioning, so stabilising strategies should focus on improving these domains. Although some of the stabilised constant domain TCRs gave a larger increase in  $T_m$  than the variable domain mutations (+7.3 vs +2.8 °C for ImmH), the performance of both stabilised molecules in functional serum stability assays were comparable despite the lower  $T_m$  of the V $\alpha$  mutant. Stabilising the constant domains is an advantageous strategy in that it can easily be transferred to different unrelated TCRs, but introducing stabilising mutations into the variable domains may be better suited to enhancing functional stability, as well as theoretically compatible with a wider range of protein formats such as single chain V domain only constructs (scTV).

The success of stabilising mutagenesis reflects the fact that wildtype TCRs are inherently unstable when expressed as a soluble protein without the endogenous transmembrane domains or the presence of coreceptor proteins. Random mutations to less thermostable proteins are demonstrated to be more likely to be stabilising than those which occur in already highly thermostable proteins (Huang *et al.*, 2020), as the potential sequence space for stabilising mutagenesis is much smaller and there are fewer gains to be had; all mutations to a theoretically perfectly stable molecule would be deleterious to stability. Translating point mutations identified from a systematic screen of an individual TCR onto closely related molecules demonstrates that some stability enhancing effects act independently of CDR3 differences, and a subset of these mutations were shown to be stabilising across a panel of TCRs with low sequence similarity. Although the lower stability of soluble TCRs compared to antibodies may pose an engineering challenge for the creation of soluble protein therapeutics, the results obtained from this study suggest that there are many valid mutational strategies to overcome this issue. TCR-specific screens for stabilising mutations can be generalised to other molecules without the need to repeat labour-intensive randomisation and library screening techniques.

# Chapter Seven

## General discussion

---

### **7.1 Summary of research**

Understanding T cell receptor stability is important both in the context of examining variability and function as part of the immune system, and also as a key property for harnessing T cell receptors as protein-based drugs. The variability of T cell receptors is often discussed only in the context of the peptide-MHC contacting loops, but in fact there exists high amino acid variation across the entire V domain. This may have evolved as a side product of gene drift that resulted in germline CDR sequence variation (which varied the amino acid side chain chemistries at the peptide-MHC binding interface, allowing for specific affinity for different pMHCs), but it seems much more likely that the framework of the V domain has functions beyond merely positioning the CDR loops for binding, and that the sequence of the framework region has significant impact on overall protein stability and function.

#### **7.1.1 Structural classification**

I began by establishing a structurally informed schema to characterise the role of non-CDR IMGT-aligned positions in the V domain based on a large dataset of crystal structures, allowing me to functionally assign a structural role to residues discussed later in the absence of specific crystal or modelled structures. Classifying residues based on average side chain exposure in all structures, I identified a disparity in sequence variability within the V gene repertoire between residues that are predominantly buried (generally conserved) and those which are exposed to solvent (exhibiting more variation between different V genes). This is unsurprising given the evolutionary disadvantage of mutations within the core of a protein; they are more likely to be destabilising than changes made at the protein surface because their local environment includes contacts to many other residues that may be disrupted by mutagenesis. The higher variability of surface exposed residues is particularly

interesting in the context of stability analysis, as side chain chemistry at the surface of proteins has been demonstrated to influence overall protein stability and aggregation propensity.

I also took a systematic approach to identifying residues that were involved in inter-domain contact in more than 10% of the structures in my panel of 200 TCRs where there was an available structure. This allows for direct comparison of the different amino acids present at key interface positions in different structures. TCRs are sequence-diverse at most interface positions, and through this breakdown of interface contacts by amino acid I was able to identify which side chains most frequently make inter-domain contacts. Based on this information novel mutations could be proposed that mutate a poorly-contacting amino acid to one which is better suited to making interdomain contacts, offering a promising avenue for further research.

This approach also has potential for future affinity enhancement. Redefining important pMHC-binding residues based on structural information has the potential to be a useful strategy for affinity engineering, as it would allow more focused libraries to be built to maximise the chances of improving pMHC contacts.

### **7.1.2 Stability of TCRs**

The T cell receptor evolved as a membrane bound protein forming part of a signalling network with many different protein complexes. Thymic selection requires TCRs to be capable of binding pMHC and therefore will remove any TCRs that are so unstable they cannot form a pMHC binding interface, but there has not been the same evolutionary pressure for TCR stability as a soluble molecule compared with antibodies that have evolved to act both as B cell receptors and soluble immunoglobulin.

The relative stability of different soluble T cell receptors has not previously been assessed on a large scale. Chapter Four discussed the variation in thermal stability seen for a panel of 267 unique wild-type TCRs, which revealed a broad (~30 °C)

range of  $T_m$  values between the most and least stable examples. In general, TCRs exhibited lower thermostability than comparable published data for antibody Fab and scFV molecules, which may reflect their evolutionary niche as surface-bound receptor. As there was such a wide range of melting points observed for different TCRs, I went on to characterise different sequence and structural features that could influence thermostability. However, no single factor was found that correlated well with the observed  $T_m$  measurements. This may be due to the limited nature of the dataset, which did not cover all V and J genes in the human repertoire and was biased towards those TCRs that had been successfully refolded for characterisation. It may also reflect TCR stability as a multi-faceted problem; it is possible that more sophisticated multi-factorial analysis could have identified some underlying rules that explain variation in TCR thermal stability. From the experimental evidence shown in Chapter Four it does not appear that V gene usage or V gene pairing is enough to predict the stability of a TCR. However, as TCRs seem to vary in stability this suggests that there is room to engineer different TCRs to improve their stability, an approach I discussed in Chapters Five and Six.

### **7.1.3 Screening for stability enhancing mutations**

#### *7.1.3.1 Library design*

The library screening approach discussed in Chapter Five addresses a current gap in published screens of TCRs for stabilising mutations. The historical approach to building libraries of variants typically involved either the use of error prone PCR or targeted degenerate codons. The advantage of targeted libraries is that specific amino acids (for example, at the CDR loops or at domain interfaces) can be randomised in combination to maximise the diversity at areas of interest, but this approach is limited in the number of positions that can be investigated at one time before the theoretical combinatorial diversity of the library becomes too large to be presented by any display system (see Table 7.1). However, the use of single point saturation mutagenesis libraries allows for avoidance of creating unwieldy combinatorial library sizes whilst still sampling a wide proportion of the available

sequence space. Alternative methods for designing a more functional library of mutants have been proposed which rely on introducing only amino acids that are found at that position in an alignment of homologous genes, therefore biasing the library towards mutations to amino acids that have been selected by evolution as functional (Porebski and Buckle, 2016; Sternke, Tripp and Barrick, 2019). However, this consensus mutagenesis approach does assume that the gene family used to build the consensus library is made up of functional variants; in the case of T-cell receptors and their known poor stability when converted to soluble proteins, it is less clear that consensus mutagenesis would enrich the library with more stable variants than a randomising approach. Comparison of consensus libraries compared to straightforward randomisation has found that it does not necessarily produce more functional outputs (Jäckel *et al.*, 2010). Interestingly, the most successful stabilising mutations produced by my randomised library did introduce amino acids that exist at those positions in other V genes.

The screening system was carried out in M13 filamentous phage, a well-established system for presenting full length TCR libraries when screening for stronger affinity (Li *et al.*, 2005) or for stability enhancement (Gunnarsen *et al.*, 2013). However, other screening tools are available with different advantages and disadvantages, as summarised in Table 7.1. One particular disadvantage of the phage display approach is the lack of a screening option to directly monitor presentation of TCR variants; FACS staining of libraries displayed on yeast with an antibody against the TCR can be used to directly select for mutations which improve surface expression levels, a property strongly correlated with soluble protein stability (Shusta *et al.*, 1999). To select solely for increase resistance to thermal degradation in my phage display library I initially trialled a panning approach using antibodies against the alpha or beta constant domains, but this did not allow for discrimination between two known stable and unstable variants of a TCR; it is likely that, as unfolded alpha and beta chains will still be linked by the introduced disulphide between the constant domains, selection only for the presence of either constant domain did not select for correctly folded TCR after a thermal challenge. Although this disulphide bond has

been shown to improve presentation of TCRs on phage particles as assessed through increased functional output from selections (Li *et al.*, 2005; Løset *et al.*, 2015), removing it would potentially destabilise the TCR structure and therefore increase the selective advantage for stabilising mutations.

Alternative library formats include those which would be compatible with fluorescence-assisted cell sorting (FACS), allowing selection for both affinity and stability at the same by staining for both pMHC binding and also for the surface expression level of the TCR (rather than the sequential thermal and affinity challenges in my phage display studies).

*Table 7.1: Comparison of different display systems (adapted from Robinson et al (Robinson et al., 2021))*

Display system	Maximum number of cells that can be screened	Genotype-phenotype link	Display of full length TCR format	FACS sorting possible for affinity and expression levels
Phage(Li <i>et al.</i> , 2005; Gunnarsen <i>et al.</i> , 2013; Alfaleh <i>et al.</i> , 2020)	$10^{12}$ - $10^{13}$	Yes	Yes	No
Yeast (Kieke <i>et al.</i> , 1999; Chao <i>et al.</i> , 2006)	$10^9$ - $10^{10}$	Yes	No, only single chain formats	Yes
Mammalian (Beerli <i>et</i>	$\sim 10^6$	Weak; transient	Yes	Yes

<i>al.</i> , 2008; Ho and Pastan, 2009; Wagner <i>et al.</i> , 2019)		transfection is not robust (Standard viral transduction links genotype- phenotype, but can result in heterogenous expression levels due to varying sites of integration )		
----------------------------------------------------------------------------------	--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	--

It is also interesting to note that the majority of mutations enriched in this library approach were at the surface of the TCR. As discussed above, the overall V gene repertoire has evolved greater sequence diversity at surface exposed side chains than at those buried in the core, due to the more constrained local environment within the protein core making mutations more likely to destabilise the fold. The nature of the library design (as discussed above) restricts all possible mutations to single amino acid changes, which may also influence this bias in the structural location of the mutations enriched after selection. Alternative library designs which allow covariance could potentially return different results, as this may allow for rearrangements of core interactions by introducing multiple mutations at once. However, the limitations of library size in a covariant library discussed above

restricts the scope for isolating a novel more stable arrangement of core residues. In addition, the predominance of highly destabilised TCRs in library designed to rearrange the buried core residues could drive aggregation of all the library TCRs and make isolating any successfully folded variants more difficult. No strongly enriched stabilising mutations were found in the  $V\alpha:V\beta$  interface in my libraries, which again may reflect the limitations of a non-combinatorial library design.

The principle of introducing mutations to a TCR sequence in some senses resembles the affinity maturation process of antibodies, where a weakly binding antibody with a sequence obtained from V-(D)-J recombination is further altered by the process of somatic hypermutation. This does not just introduce mutations in the epitope binding CDR loops but throughout the structure of the antibody. Mutations that can occur across the entire antibody sequence are then selected for improved antigen binding. Synthetic affinity maturation also tends to introduce mutations only at the CDR loops, whereas studies of naturally introduced mutations show they are much more widely spread across the structure of the antibody; for example in directed evolution panning experiments of a library generated in mutation-prone *E. coli* strain, only one of the ten mutations which boost affinity were directly contacting the epitope (Boder, Midelfort and Wittrup, 2000). It is interesting to note the work of Sheng *et al.* (2017) in analysing antibody repertoires indicates that there are mutations that occur at high frequency in antibodies encoded by the same chain independently of the target antigen. This suggests that some mutations may be generally advantageous, and perhaps are offering a stability improvement that aids the function of the antibody – much like the speculative reason for low intrinsic TCR stability discussed above, unmodified antibodies could have lower stability in order to allow for more flexibility to bind a wider range of potential antigens.

The real novelty of my framework mutagenesis approach is examined in Chapter Six. where these mutations are transferred onto alternative scaffolds. While mutations to the constant domains have been demonstrated to translate onto multiple different TCRs and still stabilise, previous studies using mutations in the variable domains

have tended to be restricted to only the specific TCR being studied. Although some published mutations showed improved expression levels (strongly correlated with overall stability) across multiple closely related TCRs, such as those with the same murine TRBV13-2 gene encoding part of the V domain but different J regions(Lunde *et al.*, 2010), no published examples currently exist of mutations which can be introduced into different TCR V genes and consistently improve stability of soluble protein. Chapter Six demonstrates that introduction of a single point mutation at the V $\alpha$ :C $\alpha$  interface ( $\alpha$ X96L) consistently gives an improvement to the overall thermal stability of different TCRs with diverse sequences without altering the binding kinetics, and that a surface exposed mutation on the E strand of the beta chain ( $\beta$ X90T) can do the same on a smaller number of different TCRs.

## **7.2 Future directions**

### **7.2.1 Understanding TCR surface expression and protein stability**

The scope of this research has focused on the TCR as a soluble molecule, rather than in the natural context of the T-cell membrane and the presence of coreceptor molecules. As discussed in Chapter One, there is significant commercial interest in harnessing the pMHC recognition of TCRs for therapeutic purposes in both a soluble and membrane bound context. The level of surface expression for TCRs in the context of a T-cell membrane does seem to vary between different TCR sequences (Heemskerk *et al.*, 2007) and can be improved with different stabilising strategies, such as the use of murine constant domains or C $\alpha$ :C $\beta$  disulphide bonds (Cohen *et al.*, 2006, 2007). Thomas *et al.* (2019) identified trends in V gene usage that seem to predispose TCRs towards strong or weak presentation, as measured by their ability to compete with a high-stability engineered TCR on the surface of Jurkat cells (engineered to have no endogenous TCRs), and were able to isolate key residues that

could be translated onto weakly expressing TCRs to improve their surface expression levels in this assay. As noted above, these key residues include the  $\alpha$ X96L mutation that was also isolated from my phage screening approach of soluble TCRs and gave consistent improvement in thermal stability for other non-related TCRs (Chapter Six); this strongly implies that the improved stability of the molecule in isolation is linked to the stability of a TCR on the T cell surface. However, there has been limited direct comparison between TCR stability in terms of improved surface level expression compared with protein-specific measurements of stability such as melting point or aggregation propensity. The overall stability of TCRs on the cell surface is part of a highly dynamic system and requires correct association with other proteins such as CD3 (Alcover, Alarcón and Di Bartolo, 2018), so it is unclear how much impact the overall protein thermal stability has in defining surface levels overall. An obvious extension of the research here would be to examine if the stabilising framework mutations I identified have an impact on the surface expression levels of TCRs, to determine if the improvement in thermal stability observed will directly correspond to an increase in presentation on T-cells. As these successfully stabilising mutations include residues at the surface of the molecule as well as at the variable:constant interfaces, it would be particularly interesting to note if the location of the mutation has more impact on T-cell surface presentation than on soluble molecule stability. Many successful protein engineering strategies for soluble proteins target surface-exposed residues, particularly disrupting hydrophobic patches, in order to minimise aggregation in solution. This may not be a relevant consideration in the dynamic T-cell membrane where interactions occur in a 2D surface and TCRs are part of a larger complex with coreceptor proteins.

The existence of mis-pairing between the alpha and beta chains of introduced and endogenous TCRs is a known issue in T-cell engineering (Rath and Arber, 2020) and has been addressed with strategies including swapping the alpha and beta constant domains, introducing pairs of charged residues, and disulphide bonds between the constant domains. Mis-pairing may require a different stabilisation strategy, perhaps

focusing more on finding mutations at the  $V\alpha:V\beta$  interface residues that I defined in Chapter Three to promote stronger binding between the two chains. However, T cell engineering approaches to remove the endogenous TCR entirely have successfully been demonstrated (Legut *et al.*, 2018; Stadtmauer *et al.*, 2020), which removes the need for TCR engineering to reduce mis-pairing.

It would also be interesting to investigate the impact of different stabilisation mutations on T cell signalling. In a bispecific therapeutic such as the ImmTAC format the need for endogenous signal transduction upon TCR:pMHC binding is bypassed (instead T-cells are activated by the molecule binding to both the pMHC and the CD3 coreceptor of a T-cell and physically redirecting the T-cell to form a synapse and trigger killing). As long as affinity is not affected by the stabilising mutations no difference in T-cell activation is expected between a stabilised and wild type framework in this case, but in the case of endogenous T-cell function the transduction of the signal produced on TCR binding to pMHC must pass through the full length of the molecule into the transmembrane domains. The nature of signal transduction is still under debate, but hypotheses such as the mechanosensor model imply that whole-domain shifts between variable and constant domains are required to produce T cell signalling on binding. The stabilising mutations at the  $V\alpha:C\alpha$  interface I have introduced involve introducing residues that are present at that position in other V genes so are less likely to disrupt any evolved signalling pathways than a non-consensus mutation, but it is still untested if stabilising interfaces between the four domains of the TCR heterodimer will be deleterious to T-cell signalling.

### **7.2.2 Using existing stability information to guide future engineering**

The stability of protein biotherapeutics is a key commercial property that must be considered during engineering of TCRs as soluble drugs. The use of constant screening for long term stability and aggregation throughout the lifetime of

developing a protein biotherapeutic is a common theme in antibody drug development (Jain et al., 2017; Yingda Xu et al., 2019; Bailly et al., 2020).

Monitoring TCR stability via experimental determination of  $T_m$  is an important step towards building a better understanding of how changes to protein sequence can impact developability characteristics.

When discussing the relative stability of different molecules as identified in Chapter Four, it is tempting to identify the most stable TCR as a platform which should be adapted to bind other target pMHC. As the work of Smith *et al.* (2014) shows, it can be possible to engineer a TCR away from one target peptide to another (in this instance, from Tax to MART1, both presented by HLA A\*02) by randomising key pHLA-contacting residues and panning these yeast display libraries for variants which bind the novel peptide and do not bind the original cognate. However, this approach may be limited to a subset of pMHC complexes – in the same study, another novel HLA A\*02 peptide (from WT-1) was also chosen for selection, but multiple rounds of panning failed to identify any variants which had switched affinity away from the cognate peptide. In antibodies, the grafting of CDR loops from a discovery library onto a human scaffold has been discussed in order to both stabilise and humanise the overall structure. Although this principle of grafting onto a more stable scaffold has been successful, it is frequently noted that further framework mutations are necessary in order to reproduce the binding affinity of the original CDRs (Apgar *et al.*, 2016; Lu *et al.*, 2020). Despite the modular nature of the Ig fold system, subtle differences in the so called “Vernier zone” of residues surrounding the CDR loops can have a dramatic impact on binding kinetics. Even mutations at the opposite end of the fold can impact on binding, as seen in the whole-domain angle shifts observed by Koenig et al when a single amino acid alteration at the V:C interface led to an improvement not only in stability but also in affinity (Koenig *et al.*, 2017). The complexity of the system limits the utility of a standardised stable scaffold approach as, based on the precedent set by antibodies, further bespoke engineering will be required in the majority of cases in order to maintain binding activity.

The existence of structural and sequence diversity within the TCR repertoire (as discussed in Chapter Three) also implies an evolutionary advantage for amino acid variation outside the pMHC-contacting regions of the molecule. By allowing a range of docking and inter-domain angles, alongside greater or lesser flexibility of the CDR loops on binding, the naturally highly diverse side chains found throughout the overall architecture of the TCR allow the immune system to produce receptors capable of binding a wide range of peptide-MHC complexes. A stability strategy that removes this diversity by using a single highly stable scaffold which can be altered at the CDR3 loops has the risk of reducing the overall range of different peptides that could be bound effectively. In particular, the unique advantage of the TCR system is the highly specific binding footprint to pMHC that can distinguish between single amino acid differences in the peptide; it could be that this specificity requires the range and flexibility of different scaffolds and that restricting the repertoire to only the most stable chains will prevent binding to some potentially clinically relevant pMHCs.

Instead of a generic stabilising platform, it is perhaps better to identify a generic system for stabilising structurally diverse TCRs. Strategies exist to improve stability by engineering the constant domains, either through introduction of a disulphide bond (Boulter *et al.*, 2003) or through combining mutations identified from a computational screen of the entire domain (Froning *et al.*, 2020). However, these strategies are not compatible with the increasing range of alternative TCR formats that do not include constant domains, such as single-chain V-domain only formats (Kieke *et al.*, 1999). More generally, given that TCRs vary in their melting point even when expressed with identically stabilised constant domains (a phenomenon seen with both Boulter-disulphide stabilised TCRs in Chapter Four and also observed for the disulphide-free stabilised constant variants tested by Froning *et al.* (2020) and the mutations I screened in Chapter Six), the V domains are clearly determining the overall stability of the molecule. Library based screening of a specific TCR with desired affinity and/or specificity to identify stability enhancing mutations is a valid approach to overcome these stability issues but is highly labour intensive. My

research shows that stability enhancing mutations can be translated to unrelated molecules without the need to build individual libraries for each case, vastly simplifying the process of stabilising larger panels of TCRs. The poor stability of soluble TCRs (even when stabilised by an introduced disulphide bond) compared with that of antibodies is a significant challenge for drug development, but this intrinsic low stability may provide more “low hanging fruit” for generic stability enhancing mutations.

### **7.3 Concluding remarks**

The importance of non-CDR regions of the TCR has often been overlooked. They are important sites of variability that may have significant impact on the stability of soluble TCRs, and I have demonstrated that engineering framework mutations for enhanced stability is a promising avenue to improve the developability of TCR-based therapeutics. More generally, TCR non-CDR regions and inter-domain interfaces are highly sequence diverse. For the membrane bound T cell receptor, differences between different TCR sequences could have implications for surface expression, binding and signalling that have not yet been established.

## List of References

---

- Actor, J. K. (2012) '4 - T-Cell Immunity', in Actor, J. K. B. T.-E. I. R. I. and M. (Second E. (ed.). Philadelphia: W.B. Saunders, pp. 25–32. doi: <https://doi.org/10.1016/B978-0-323-07447-6.00004-1>.
- Adams, J. J. *et al.* (2015) 'Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity', *nature immunology*, 17(1), pp. 87–94. doi: 10.1038/ni.3310.
- Aggen, D. H. *et al.* (2011) 'Identification and engineering of human variable regions that allow expression of stable single-chain T cell receptors.', *Protein engineering, design & selection : PEDS*. 2010/12/17, 24(4), pp. 361–72. doi: 10.1093/protein/gzq113.
- Alcover, A., Alarcón, B. and Di Bartolo, V. (2018) 'Cell Biology of T Cell Receptor Expression and Regulation', *Annual Review of Immunology*, 36, pp. 103–125. doi: 10.1146/annurev-immunol-042617-053429.
- Alexander, C. G. *et al.* (2014) 'Novel microscale approaches for easy, rapid determination of protein stability in academic and commercial settings.', *Biochimica et biophysica acta*, 1844(12), pp. 2241–2250. doi: 10.1016/j.bbapap.2014.09.016.
- Alfaleh, M. A. *et al.* (2020) 'Phage Display Derived Monoclonal Antibodies: From Bench to Bedside ', *Frontiers in Immunology* , p. 1986. Available at: <https://www.frontiersin.org/article/10.3389/fimmu.2020.01986>.
- Amin, N. *et al.* (2004) 'Construction of stabilized proteins by combinatorial consensus mutagenesis', *Protein Engineering, Design and Selection*, 17(11), pp. 787–793. doi: 10.1093/protein/gzh091.
- Antonacci, R. *et al.* (2020) 'Evolution of the T-cell receptor (TR) Loci in the adaptive immune response: The tale of the TRG locus in mammals', *Genes*, 11(6). doi: 10.3390/genes11060624.

Apgar, J. R. *et al.* (2016) 'Beyond CDR-grafting: Structure-guided humanization of framework and CDR regions of an anti-myostatin antibody', *mAbs*, 2016/09/13, 8(7), pp. 1302–1318. doi: 10.1080/19420862.2016.1215786.

Armstrong, K. M., Piepenbrink, K. H. and Baker, B. M. (2008) 'Conformational changes and flexibility in T-cell receptor recognition of peptide - MHC complexes', *Biochemical Journal*, pp. 183–196. doi: 10.1042/BJ20080850.

Attaf, M. *et al.* (2015) 'The T cell antigen receptor: the Swiss army knife of the immune system', *Clinical and Experimental Immunology*, 181(1), pp. 1–18. doi: 10.1111/cei.12622.

Ayres, C. M. *et al.* (2016) 'Differential utilization of binding loop flexibility in T cell receptor ligand selection and cross-reactivity', *Scientific Reports*, 6. doi: 10.1038/srep25070.

Bailly, M. *et al.* (2020) 'Predicting Antibody Developability Profiles Through Early Stage Discovery Screening', *mAbs*, 12(1), p. 1743053. doi: 10.1080/19420862.2020.1743053.

Beerli, R. R. *et al.* (2008) 'Isolation of human monoclonal antibodies by mammalian cell display', *Proceedings of the National Academy of Sciences*, 105(38), pp. 14336 LP – 14341. doi: 10.1073/pnas.0805942105.

Berman, H. M. *et al.* (2002) 'The protein data bank', *Acta Crystallographica Section D: Biological Crystallography*, 58(6 I), pp. 899–907. doi: 10.1107/S0907444902003451.

Bethune, M. T. *et al.* (2016) 'Domain-swapped T cell receptors improve the safety of TCR gene therapy.', *eLife*, 5. doi: 10.7554/eLife.19095.

Bialer, G. *et al.* (2010) 'Selected Murine Residues Endow Human TCR with Enhanced Tumor Recognition', *The Journal of Immunology*, 184(11), pp. 6232 LP – 6241. doi: 10.4049/jimmunol.0902047.

Björk, I. and Pol, E. (1992) 'Biphasic transition curve on denaturation of chicken

- cystatin by guanidinium chloride. Evidence for an independently unfolding structural region.', *FEBS letters*, 299(1), pp. 66–68. doi: 10.1016/0014-5793(92)80102-m.
- Boder, E. T., Midelfort, K. S. and Wittrup, K. D. (2000) 'Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity.', *Proceedings of the National Academy of Sciences of the United States of America*, 97(20), pp. 10701–10705. doi: 10.1073/pnas.170297297.
- Bolotin, D. A. *et al.* (2015) 'MiXCR: software for comprehensive adaptive immunity profiling', *Nature methods*. United States, pp. 380–381. doi: 10.1038/nmeth.3364.
- Borgo, B. and Havranek, J. J. (2012) 'Automated selection of stabilizing mutations in designed and natural proteins', *Proceedings of the National Academy of Sciences of the United States of America*. 2012/01/17, 109(5), pp. 1494–1499. doi: 10.1073/pnas.1115172109.
- Bork, P., Holm, L. and Sander, C. (1994) 'The immunoglobulin fold: Structural classification, sequence patterns and common core', *Journal of Molecular Biology*. J Mol Biol, pp. 309–320. doi: 10.1006/jmbi.1994.1582.
- Boulter, J. M. *et al.* (2003) 'Stable, soluble T-cell receptor molecules for crystallization and therapeutics', *Protein Engineering, Design and Selection*, 16(9), pp. 707–711. doi: 10.1093/protein/gzg087.
- Bowie, J. U. *et al.* (1990) 'Deciphering the message in protein sequences: tolerance to amino acid substitutions', *Science*, 247(4948), pp. 1306 LP – 1310. doi: 10.1126/science.2315699.
- van Boxel, G. I. *et al.* (2009) 'Some lessons from the systematic production and structural analysis of soluble  $\alpha\beta$  T-cell receptors', *Journal of Immunological Methods*, pp. 14–21. doi: 10.1016/j.jim.2009.08.008.
- Carter, J. A. *et al.* (2019) 'Single T Cell Sequencing Demonstrates the Functional Role of  $\alpha\beta$  TCR Pairing in Cell Lineage and Antigen Specificity', *Frontiers in*

*immunology*, 10, p. 1516. doi: 10.3389/fimmu.2019.01516.

Case, D. A. *et al.* (2005) 'The Amber biomolecular simulation programs.', *Journal of computational chemistry*, 26(16), pp. 1668–1688. doi: 10.1002/jcc.20290.

Chakroun, N. *et al.* (2016) 'Mapping the Aggregation Kinetics of a Therapeutic Antibody Fragment', *Molecular Pharmaceutics*, 13(2), pp. 307–319. doi: 10.1021/acs.molpharmaceut.5b00387.

Chang, H. C. *et al.* (1994) 'A general method for facilitating heterodimeric pairing between two proteins: Application to expression of  $\alpha$  and  $\beta$  T-cell receptor extracellular segments', *Proceedings of the National Academy of Sciences of the United States of America*, 91(24), pp. 11408–11412. doi: 10.1073/pnas.91.24.11408.

Chao, G. *et al.* (2006) 'Isolating and engineering human antibodies using yeast surface display.', *Nature protocols*, 1(2), pp. 755–768. doi: 10.1038/nprot.2006.94.

Chemical Computing Group ULC (2019) 'Molecular Operating Environment (MOE)'. Montreal, QC, Canada.

Chen, L. *et al.* (2015) 'Preferential germline usage and VH/VL pairing observed in human antibodies selected by mRNA display', *Protein Engineering, Design and Selection*, 28(10), pp. 427–435. doi: 10.1093/protein/gzv042.

Choudhuri, K. *et al.* (2005) 'T-cell receptor triggering is critically dependent on the dimensions of its peptide-MHC ligand.', *Nature*, 436(7050), pp. 578–582. doi: 10.1038/nature03843.

Clark, L. A. *et al.* (2006) 'Trends in Antibody Sequence Changes during the Somatic Hypermutation Process', *The Journal of Immunology*, 177(1), pp. 333–340. doi: 10.4049/jimmunol.177.1.333.

Cohen, C. J. *et al.* (2006) 'Enhanced Antitumor Activity of Murine-Human Hybrid T-Cell Receptor (TCR) in Human Lymphocytes Is Associated with Improved Pairing and TCR/CD3 Stability', *Cancer Research*, 66(17), pp. 8878 LP – 8886. doi:

10.1158/0008-5472.CAN-06-1450.

Cohen, C. J. *et al.* (2007) 'Enhanced antitumor activity of T cells engineered to express T-cell receptors with a second disulfide bond', *Cancer research*, 67(8), pp. 3898–3903. doi: 10.1158/0008-5472.CAN-06-3986.

Cole, D. K. *et al.* (2013) 'Increased Peptide Contacts Govern High Affinity Binding of a Modified TCR Whilst Maintaining a Native pMHC Docking Mode', *Frontiers in immunology*, 4, p. 168. doi: 10.3389/fimmu.2013.00168.

Courtney, A. H., Lo, W.-L. and Weiss, A. (2018) 'TCR Signaling: Mechanisms of Initiation and Propagation', *Trends in biochemical sciences*. 2017/12/18, 43(2), pp. 108–123. doi: 10.1016/j.tibs.2017.11.008.

Davenport, T. M. *et al.* (2016) 'Somatic Hypermutation-Induced Changes in the Structure and Dynamics of HIV-1 Broadly Neutralizing Antibodies.', *Structure (London, England : 1993)*, 24(8), pp. 1346–1357. doi: 10.1016/j.str.2016.06.012.

DeKosky, B. J. *et al.* (2016) 'Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires', *Proceedings of the National Academy of Sciences*, 113(19), p. E2636 LP-E2645. doi: 10.1073/pnas.1525510113.

Deniger, D. C. *et al.* (2018) 'T-cell Responses to TP53 "Hotspot" Mutations and Unique Neoantigens Expressed by Human Ovarian Cancers', *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2018/05/31, 24(22), pp. 5562–5573. doi: 10.1158/1078-0432.CCR-18-0573.

Dong, D. *et al.* (2019) 'Structural basis of assembly of the human T cell receptor–CD3 complex', *Nature*, 573(7775), pp. 546–552. doi: 10.1038/s41586-019-1537-0.

Du, F. *et al.* (2019) 'Engineering an EGFR-binding Gp2 domain for increased hydrophilicity', *Biotechnology and bioengineering*. 2018/12/30, 116(3), pp. 526–535. doi: 10.1002/bit.26893.

Dudgeon, K. *et al.* (2012) 'General strategy for the generation of human antibody variable domains with increased aggregation resistance', *Proceedings of the National Academy of Sciences*, 109(27), pp. 10879–10884. doi: 10.1073/pnas.1202866109.

Dudgeon, K., Famm, K. and Christ, D. (2008) 'Sequence determinants of protein aggregation in human VH domains', *Protein Engineering, Design and Selection*, 22(3), pp. 217–220. doi: 10.1093/protein/gzn059.

Dunbar, J. *et al.* (2014) 'Examining Variable Domain Orientations in Antigen Receptors Gives Insight into TCR-Like Antibody Design', *PLOS Computational Biology*, 10(9), p. e1003852. Available at: <https://doi.org/10.1371/journal.pcbi.1003852>.

Dunn, S. M. *et al.* (2006) 'Directed evolution of human T cell receptor CDR2 residues by phage display dramatically enhances affinity for cognate peptide-MHC without increasing apparent cross-reactivity', *Protein Sci*, 15(4), pp. 710–721. doi: 10.1110/ps.051936406.

Enever, C., Pupecka-Swider, M. and Sepp, A. (2015) 'Stress selections on domain antibodies: "What doesn't kill you makes you stronger"', *Protein Engineering Design and Selection*, 28(3), pp. 59–66. doi: 10.1093/protein/gzu057.

Ewert, S. *et al.* (2003) 'Biophysical properties of human antibody variable domains.', *Journal of molecular biology*, 325(3), pp. 531–553. doi: 10.1016/s0022-2836(02)01237-8.

Feng, Y., Reinherz, E. L. and Lang, M. J. (2018) 'αβ T Cell Receptor Mechanosensing Forces out Serial Engagement', *Trends in Immunology*, 39(8), pp. 596–609. doi: 10.1016/j.it.2018.05.005.

Flajnik, M. F. and Kasahara, M. (2010) 'Origin and evolution of the adaptive immune system: genetic events and selective pressures', *Nature reviews. Genetics*. 2009/12/08, 11(1), pp. 47–59. doi: 10.1038/nrg2703.

Froning, K. *et al.* (2020) 'Computational stabilization of T cell receptors allows

pairing with antibodies to form bispecifics', *Nature Communications*, 11(1), pp. 1–14. doi: 10.1038/s41467-020-16231-7.

Gagnon, S. J. *et al.* (2006) 'T Cell Receptor Recognition via Cooperative Conformational Plasticity', *Journal of Molecular Biology*, 363(1), pp. 228–243. doi: <https://doi.org/10.1016/j.jmb.2006.08.045>.

Gao, K. *et al.* (2019) 'Germline-Encoded TCR-MHC Contacts Promote TCR V Gene Bias in Umbilical Cord Blood T Cell Repertoire', *Frontiers in Immunology*, p. 2064. Available at: <https://www.frontiersin.org/article/10.3389/fimmu.2019.02064>.

Gao, K., Oerlemans, R. and Groves, M. R. (2020) 'Theory and applications of differential scanning fluorimetry in early-stage drug discovery', *Biophysical Reviews*, 12(1), pp. 85–104. doi: 10.1007/s12551-020-00619-2.

Garboczi, D. N. *et al.* (1996) 'Assembly, specific binding, and crystallization of a human TCR- $\alpha$  with an antigenic Tax peptide from human T lymphotropic virus type 1 and the class I MHC molecule HLA-A2.', *Journal of immunology (Baltimore, Md. : 1950)*, 157(12), pp. 5403–10. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8955188>.

Goebeler, M.-E. and Bargou, R. C. (2020) 'T cell-engaging therapies — BiTEs and beyond', *Nature Reviews Clinical Oncology*, 17(7), pp. 418–434. doi: 10.1038/s41571-020-0347-5.

Gowthaman, R. and Pierce, B. G. (2019) 'TCR3d: The T cell receptor structural repertoire database', *Bioinformatics (Oxford, England)*, 35(24), pp. 5323–5325. doi: 10.1093/bioinformatics/btz517.

Greenfield, N. J. (2006) 'Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions', *Nature protocols*, 1(6), pp. 2527–2535. doi: 10.1038/nprot.2006.204.

Grégoire, C. *et al.* (1991) 'Engineered secreted T-cell receptor alpha beta heterodimers', *Proceedings of the National Academy of Sciences*, 88(18).

La Gruta, N. L. *et al.* (2018) 'Understanding the drivers of MHC restriction of T cell receptors', *Nature Reviews Immunology*, 18(7), pp. 467–478. doi: 10.1038/s41577-018-0007-5.

Güler, R. *et al.* (2020) 'Increasing thermal stability and improving biodistribution of VEGFR2-binding affibody molecules by a combination of in silico and directed evolution approaches', *Scientific Reports*, 10(1), p. 18148. doi: 10.1038/s41598-020-74560-5.

Gunnarsen, K. S. *et al.* (2013) 'Chaperone-assisted thermostability engineering of a soluble T cell receptor using phage display', *Sci Rep*, 3, p. 1162. doi: 10.1038/srep01162.

Gupta, S. *et al.* (2021) 'Oxidation and Deamidation of Monoclonal Antibody Products: Potential Impact on Stability, Biological Activity, and Efficacy', *Journal of Pharmaceutical Sciences*. doi: <https://doi.org/10.1016/j.xphs.2021.11.024>.

Haga-Friedman, A., Horovitz-Fried, M. and Cohen, C. J. (2012) 'Incorporation of transmembrane hydrophobic mutations in the TCR enhance its surface expression and T cell functional avidity.', *Journal of immunology (Baltimore, Md. : 1950)*, 188(11), pp. 5538–5546. doi: 10.4049/jimmunol.1103020.

Haidar, J. N. *et al.* (2009) 'Structure-based design of a T-cell receptor leads to nearly 100-fold improvement in binding affinity for pepMHC', *Proteins*, 74(4), pp. 948–960. doi: 10.1002/prot.22203.

Harris, D. T. *et al.* (2016) 'Deep mutational scans as a guide to engineering high affinity t cell receptor interactions with peptide-bound major histocompatibility complex', *Journal of Biological Chemistry*, 291(47), pp. 24566–24578. doi: 10.1074/jbc.M116.748681.

Hart, D. P. *et al.* (2008) 'Retroviral transfer of a dominant TCR prevents surface expression of a large proportion of the endogenous TCR repertoire in human T cells', *Gene Therapy*, 15(8), pp. 625–631. doi: 10.1038/sj.gt.3303078.

Heemskerk, M. H. M. *et al.* (2007) 'Efficiency of T-cell receptor expression in dual-specific T cells is controlled by the intrinsic qualities of the TCR chains within the TCR-CD3 complex', *Blood*, 109(1), pp. 235–243. doi: 10.1182/blood-2006-03-013318.

Henry, K. A. *et al.* (2017) 'Stability-Diversity Tradeoffs Impose Fundamental Constraints on Selection of Synthetic Human VH/VL Single-Domain Antibodies from In Vitro Display Libraries', *Frontiers in Immunology*, p. 1759. Available at: <https://www.frontiersin.org/article/10.3389/fimmu.2017.01759>.

Herold, E. M. *et al.* (2017) 'Determinants of the assembly and function of antibody variable domains', *Scientific reports*, 7(1), p. 12276. doi: 10.1038/s41598-017-12519-9.

Ho, M. and Pastan, I. (2009) 'Mammalian cell display for antibody engineering.', *Methods in molecular biology (Clifton, N.J.)*, 525, pp. 337–52, xiv. doi: 10.1007/978-1-59745-554-1\_18.

Hoffmann, T., Krackhardt, A. M. and Antes, I. (2015) 'Quantitative Analysis of the Association Angle between T-cell Receptor V $\alpha$ /V $\beta$  Domains Reveals Important Features for Epitope Recognition', *PLOS Computational Biology*, 11(7), p. e1004244. Available at: <https://doi.org/10.1371/journal.pcbi.1004244>.

Holland, C. J. *et al.* (2020) 'Specificity of bispecific T cell receptors and antibodies targeting peptide-HLA', *The Journal of Clinical Investigation*, 130(5), pp. 2673–2688. doi: 10.1172/JCI130562.

Hoogenboom, H. R. *et al.* (1999) 'Selection-dominant and nonaccessible epitopes on cell-surface receptors revealed by cell-panning with a large phage antibody library', *European Journal of Biochemistry*, 260(3), pp. 774–784. doi: 10.1046/j.1432-1327.1999.00214.x.

Howie, B. *et al.* (2015) 'High-throughput pairing of T cell receptor  $\alpha$  and  $\beta$  sequences', *Science Translational Medicine*. 2015/08/21, 7(301), p. 301ra131. doi:

10.1126/scitranslmed.aac5624.

Huang, P. *et al.* (2020) 'Evaluating Protein Engineering Thermostability Prediction Tools Using an Independently Generated Dataset', *ACS Omega*, 5(12), pp. 6487–6493. doi: 10.1021/acsomega.9b04105.

Hwang, J.-R. *et al.* (2020) 'Recent insights of T cell receptor-mediated signaling pathways for T cell activation and development', *Experimental & Molecular Medicine*, 52(5), pp. 750–761. doi: 10.1038/s12276-020-0435-8.

Hwang, W. *et al.* (2020) 'The  $\alpha\beta$ TCR mechanosensor exploits dynamic ectodomain allostery to optimize its ligand recognition site', *Proceedings of the National Academy of Sciences of the United States of America*, 117(35), pp. 21336–21345. doi: 10.1073/pnas.2005899117.

Islam, M. M. *et al.* (2019) 'Hydrophobic surface residues can stabilize a protein through improved water–protein interactions', *The FEBS Journal*, 286(20), pp. 4122–4134. doi: <https://doi.org/10.1111/febs.14941>.

Jäckel, C. *et al.* (2010) 'Consensus Protein Design without Phylogenetic Bias', *Journal of Molecular Biology*, 399(4), pp. 541–546. doi: <https://doi.org/10.1016/j.jmb.2010.04.039>.

Jain, T. *et al.* (2017) 'Biophysical properties of the clinical-stage antibody landscape', *Proceedings of the National Academy of Sciences*, 114(5), pp. 944–949. doi: 10.1073/PNAS.1616408114.

Jayaram, N., Bhowmick, P. and Martin, A. C. R. (2012) 'Germline VH/VL pairing in antibodies', *Protein Engineering, Design and Selection*, 25(10), pp. 523–530. doi: 10.1093/protein/gzs043.

Jermutus, L. *et al.* (2001) 'Tailoring in vitro evolution for protein affinity or stability', *Proc Natl Acad Sci U S A*, 98(1), pp. 75–80. doi: 10.1073/pnas.011311398.

Jetha, A. *et al.* (2018) 'Homology modeling and structure-based design improve hydrophobic interaction chromatography behavior of integrin binding antibodies',

*mAbs*, 10(6), pp. 890–900. doi: 10.1080/19420862.2018.1475871.

Johnson, C. M. (2013) ‘Differential scanning calorimetry as a tool for protein folding and stability.’, *Archives of biochemistry and biophysics*, 531(1–2), pp. 100–109. doi: 10.1016/j.abb.2012.09.008.

Julian, M. C. *et al.* (2017) ‘Efficient affinity maturation of antibody variable domains requires co-selection of compensatory mutations to maintain thermodynamic stability’, *Scientific Reports*, 7(1), p. 45259. doi: 10.1038/srep45259.

Jung, S., Honegger, A. and Plückthun, A. (1999) ‘Selection for improved protein stability by phage display’ Edited by J. A. Wells’, *Journal of Molecular Biology*, 294(1), pp. 163–180. doi: <https://doi.org/10.1006/jmbi.1999.3196>.

Kabat, E. A., Wu, T. TE and Bilofsky, H. (1977) ‘Unusual distributions of amino acids in complementarity determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites’, *Journal of Biological Chemistry*, 252(19), pp. 6609–6616. doi: 10.1016/s0021-9258(17)39891-5.

Kaleli, N. E., Karadag, M. and Kalyoncu, S. (2019) ‘Phage display derived therapeutic antibodies have enriched aliphatic content: Insights for developability issues’, *Proteins: Structure, Function, and Bioinformatics*, 87(7), pp. 607–618. doi: <https://doi.org/10.1002/prot.25685>.

Kieke, M. C. *et al.* (1999) ‘Selection of functional T cell receptor mutants from a yeast surface-display library.’, *Proceedings of the National Academy of Sciences of the United States of America*, 96(10), pp. 5651–5656. doi: 10.1073/pnas.96.10.5651.

Koenig, P. *et al.* (2017) ‘Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding’, *Proceedings of the National Academy of Sciences*, 114(4), pp. E486–E495. doi: 10.1073/pnas.1613231114.

Kola, I. and Landis, J. (2004) ‘Can the pharmaceutical industry reduce attrition

rates?', *Nature Reviews Drug Discovery*, 3(8), pp. 711–716. doi: 10.1038/nrd1470.

Kuball, J. *et al.* (2007) 'Facilitating matched pairing and expression of TCR chains introduced into human T cells', *Blood*. 2006/11/02, 109(6), pp. 2331–2338. doi: 10.1182/blood-2006-05-023069.

Kuroda, D. and Tsumoto, K. (2020) 'Engineering Stability, Viscosity, and Immunogenicity of Antibodies by Computational Design', *Journal of Pharmaceutical Sciences*, 109(5), pp. 1631–1651. doi: <https://doi.org/10.1016/j.xphs.2020.01.011>.

Labute, P. (2009) 'Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures', *Proteins: Structure, Function, and Bioinformatics*, 75(1), pp. 187–205. doi: <https://doi.org/10.1002/prot.22234>.

Van Laethem, F. *et al.* (2007) 'Deletion of CD4 and CD8 coreceptors permits generation of alphabetaT cells that recognize antigens independently of the MHC.', *Immunity*, 27(5), pp. 735–750. doi: 10.1016/j.immuni.2007.10.007.

Lang, B. E. and Cole, K. D. (2017) 'Differential scanning calorimetry and fluorimetry measurements of monoclonal antibodies and reference proteins: Effect of scanning rate and dye selection', *Biotechnology Progress*, 33(3), pp. 677–686. doi: <https://doi.org/10.1002/btpr.2464>.

Lavinder, J. J. *et al.* (2009) 'High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering', *Journal of the American Chemical Society*, 131(11), pp. 3794–3795. doi: 10.1021/ja8049063.

Laydon, D. J., Bangham, C. R. M. and Asquith, B. (2015) 'Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1675), p. 20140291. doi: 10.1098/rstb.2014.0291.

Ledsgaard, L. *et al.* (2018) 'Basics of Antibody Phage Display Technology.', *Toxins*, 10(6). doi: 10.3390/toxins10060236.

Lefranc, M.-P. *et al.* (2003) 'IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains', *Developmental & Comparative Immunology*, 27(1), pp. 55–77. doi: [https://doi.org/10.1016/S0145-305X\(02\)00039-3](https://doi.org/10.1016/S0145-305X(02)00039-3).

Legut, M. *et al.* (2018) 'CRISPR-mediated TCR replacement generates superior anticancer transgenic T cells', *Blood*, 131(3), pp. 311–322. doi: [10.1182/blood-2017-05-787598](https://doi.org/10.1182/blood-2017-05-787598).

Lehmann, M. *et al.* (2000) 'From DNA sequence to improved functionality: Using protein sequence comparisons to rapidly design a thermostable consensus phytase', *Protein Engineering*, 13(1), pp. 49–57. doi: [10.1093/protein/13.1.49](https://doi.org/10.1093/protein/13.1.49).

Li, Y. *et al.* (2005) 'Directed evolution of human T-cell receptors with picomolar affinities by phage display', *Nat Biotechnol*, 23(3), pp. 349–354. doi: [10.1038/nbt1070](https://doi.org/10.1038/nbt1070).

Liu, H. and Naismith, J. H. (2008) 'An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol', *BMC Biotechnology*, 8(1), p. 91. doi: [10.1186/1472-6750-8-91](https://doi.org/10.1186/1472-6750-8-91).

Liu, T. *et al.* (2016) 'Enhancing protein stability with extended disulfide bonds', *Proceedings of the National Academy of Sciences*, 113(21), pp. 5910 LP – 5915. doi: [10.1073/pnas.1605363113](https://doi.org/10.1073/pnas.1605363113).

Lloyd, C. *et al.* (2009) 'Modelling the human immune response: performance of a 1011 human antibody repertoire against a broad panel of therapeutically relevant antigens.', *Protein engineering, design & selection : PEDS*, 22(3), pp. 159–168. doi: [10.1093/protein/gzn058](https://doi.org/10.1093/protein/gzn058).

Lo, M.-C. *et al.* (2004) 'Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery.', *Analytical biochemistry*, 332(1), pp. 153–159. doi: [10.1016/j.ab.2004.04.031](https://doi.org/10.1016/j.ab.2004.04.031).

Løset, G. Å. *et al.* (2015) 'Phage Display Engineered T Cell Receptors as Tools for the

Study of Tumor Peptide-MHC Interactions', *Frontiers in oncology*, 4, p. 378. doi: 10.3389/fonc.2014.00378.

Lu, R.-M. *et al.* (2020) 'Development of therapeutic antibodies for the treatment of diseases', *Journal of Biomedical Science*, 27(1), p. 1. doi: 10.1186/s12929-019-0592-z.

Lunde, E. *et al.* (2010) 'Stabilizing mutations increase secretion of functional soluble TCR-Ig fusion proteins', *BMC biotechnology*, 10(1), p. 61. doi: 10.1186/1472-6750-10-61.

Ma, H., Ó'Fágáin, C. and O'Kennedy, R. (2020) 'Antibody stability: A key to performance - Analysis, influences and improvement', *Biochimie*, 177, pp. 213–225. doi: <https://doi.org/10.1016/j.biochi.2020.08.019>.

Mahler, H.-C. *et al.* (2009) 'Protein aggregation: Pathways, induction factors and analysis', *journal of pharmaceutical sciences*, 98(9), pp. 2909–2934. doi: 10.1002/jps.21566.

Mariuzza, R. A., Agnihotri, P. and Orban, J. (2020) 'The structural basis of T-cell receptor (TCR) activation: An enduring enigma', *The Journal of biological chemistry*. 2019/12/17, 295(4), pp. 914–925. doi: 10.1074/jbc.REV119.009411.

Mason, D. (1998) 'A very high level of crossreactivity is an essential feature of the T-cell receptor', *Immunology Today*, 19(9), pp. 395–404. doi: 10.1016/S0167-5699(98)01299-7.

Masuda, K. *et al.* (2006) 'The role of interface framework residues in determining antibody VH/VL interaction strength and antigen-binding affinity', *The FEBS Journal*, 273(10), pp. 2184–2194. doi: <https://doi.org/10.1111/j.1742-4658.2006.05232.x>.

McCafferty, J. *et al.* (1990) 'Phage antibodies: filamentous phage displaying antibody variable domains.', *Nature*, 348(6301), pp. 552–554. doi: 10.1038/348552a0.

McConnell, A. D. *et al.* (2013) 'An integrated approach to extreme

thermostabilization and affinity maturation of an antibody', *Protein Engineering, Design and Selection*, 26(2), pp. 151–163. doi: 10.1093/protein/gzs090.

Meric, G., Robinson, A. S. and Roberts, C. J. (2017) 'Driving forces for nonnative protein aggregation and approaches to predict aggregation-prone regions', *Annual Review of Chemical and Biomolecular Engineering*, 8, pp. 139–159. doi: 10.1146/annurev-chembioeng-060816-101404.

van der Merwe, P. A. and Dushek, O. (2011) 'Mechanisms for T cell receptor triggering', *Nature Reviews Immunology*, 11(1), pp. 47–55. doi: 10.1038/nri2887.

Motozono, C. *et al.* (2015) 'Clonotypically similar hybrid alphabeta T cell receptors can exhibit markedly different surface expression, antigen specificity and cross-reactivity', *Clin Exp Immunol*, 180(3), pp. 560–570. doi: 10.1111/cei.12610.

Nakanishi, T. *et al.* (2008) 'Critical contribution of VH-VL interaction to reshaping of an antibody: the case of humanization of anti-lysozyme antibody, HyHEL-10', *Protein science : a publication of the Protein Society*, 17(2), pp. 261–270. doi: 10.1110/ps.073156708.

Natarajan, K. *et al.* (2017) 'An allosteric site in the T-cell receptor C $\beta$  domain plays a critical signalling role.', *Nature communications*, 8, p. 15260. doi: 10.1038/ncomms15260.

Neylon, C. (2004) 'Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution', *Nucleic acids research*, 32(4), pp. 1448–1459. doi: 10.1093/nar/gkh315.

Nisthal, A. *et al.* (2019) 'Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis', *Proceedings of the National Academy of Sciences*, 116(33), pp. 16367 LP – 16377. doi: 10.1073/pnas.1903888116.

Novotny, J. *et al.* (1991) 'A soluble, single-chain T-cell receptor fragment endowed with antigen-combining properties.', *Proceedings of the National Academy of Sciences of the United States of America*, 88(19), pp. 8646–50. doi:

10.1073/pnas.88.19.8646.

Nowak, C. *et al.* (2017) 'Forced degradation of recombinant monoclonal antibodies: A practical guide', *mAbs*, 2017/08/30, 9(8), pp. 1217–1230. doi: 10.1080/19420862.2017.1368602.

Nowak, J. *et al.* (2016) 'Length-independent structural similarities enrich the antibody CDR canonical class model.', *mAbs*, 8(4), pp. 751–760. doi: 10.1080/19420862.2016.1158370.

Oates, J. and Jakobsen, B. K. (2013) 'ImmTACs: Novel bi-specific agents for targeted cancer therapy', *Oncoimmunology*, 2(2), pp. e22891–e22891. doi: 10.4161/onci.22891.

Paletta, D. *et al.* (2015) 'The hypervariable region 4 (HV4) and position 93 of the alpha chain modulate CD1d-glycolipid binding of iNKT TCRs', *Eur J Immunol*, 45(7), pp. 2122–2133. doi: 10.1002/eji.201545534.

Pellicci, D. G., Koay, H. F. and Berzins, S. P. (2020) 'Thymic development of unconventional T cells: how NKT cells, MAIT cells and  $\gamma\delta$  T cells emerge', *Nature Reviews Immunology*, 20(12), pp. 756–770. doi: 10.1038/s41577-020-0345-y.

Piepenbrink, K. H. *et al.* (2013) 'The basis for limited specificity and MHC restriction in a T cell receptor interface', *Nature communications*, 4, p. 1948. doi: 10.1038/ncomms2948.

Ponsel, D. *et al.* (2011) 'High affinity, developability and functional size: the holy grail of combinatorial antibody library generation', *Molecules (Basel, Switzerland)*, 16(5), pp. 3675–3700. doi: 10.3390/molecules16053675.

Porebski, B. T. and Buckle, A. M. (2016) 'Consensus protein design', *Protein Engineering, Design and Selection*, 29(7), pp. 245–251. doi: 10.1093/protein/gzw015.

'Protein Thermal Shift™ | Life Technologies' (no date). Available at: <http://www.lifetechnologies.com/us/en/home/life-science/pcr/real-time-pcr/real-time-pcr>

time-pcr-applications/real-time-pcr-protein-analysis/protein-thermal-shift.html.

Al Qaraghuli, M. M. *et al.* (2020) 'Antibody-protein binding and conformational changes: identifying allosteric signalling pathways to engineer a better effector response', *Scientific Reports*, 10(1), pp. 1–10. doi: 10.1038/s41598-020-70680-0.

Rabia, L. A. *et al.* (2018) 'Understanding and overcoming trade-offs between antibody affinity, specificity, stability and solubility', *Biochemical engineering journal*. 2018/06/05, 137, pp. 365–374. doi: 10.1016/j.bej.2018.06.003.

Rakonjac, J. *et al.* (2011) 'Filamentous bacteriophage: biology, phage display and nanotechnology applications.', *Current issues in molecular biology*, 13(2), pp. 51–76.

Rakonjac, J., Jovanovic, G. and Model, P. (1997) 'Filamentous phage infection-mediated gene expression: construction and propagation of the gIII deletion mutant helper phage R408d3.', *Gene*, 198(1–2), pp. 99–103. doi: 10.1016/s0378-1119(97)00298-9.

Rangarajan, S. *et al.* (2018) 'Peptide–MHC (pMHC) binding to a human antiviral T cell receptor induces long-range allosteric communication between pMHC- and CD3-binding sites', *Journal of Biological Chemistry*, 293(41), pp. 15991–16005. doi: 10.1074/jbc.RA118.003832.

Rath, J. A. and Arber, C. (2020) 'Engineering Strategies to Enhance TCR-Based Adoptive T Cell Therapy', *Cells*, 9(6), p. 1485. doi: 10.3390/cells9061485.

Raybould, M. I. J. J. *et al.* (2019) 'Five computational developability guidelines for therapeutic antibody profiling', *Proceedings of the National Academy of Sciences of the United States of America*, 116(10), pp. 4025–4030. doi: 10.1073/pnas.1810576116.

Richman, S. A. *et al.* (2009) 'Structural features of T cell receptor variable regions that enhance domain stability and enable expression as single-chain V $\alpha$ V $\beta$  fragments', *Molecular Immunology*. 2008/10/31, 46(5), pp. 902–916. doi:

10.1016/j.molimm.2008.09.021.

Robinson, R. A. *et al.* (2021) 'Engineering soluble T-cell receptors for therapy', *The FEBS Journal*, 288(21), pp. 6159–6173. doi: <https://doi.org/10.1111/febs.15780>.

Rock, E. P. *et al.* (1994) 'CDR3 length in antigen-specific immune receptors.', *The Journal of experimental medicine*, 179(1), pp. 323–328. doi: 10.1084/jem.179.1.323.

Rosenberg, A. S. (2006) 'Effects of protein aggregates: an immunologic perspective.', *The AAPS journal*, 8(3), pp. E501-7. doi: 10.1208/aapsj080359.

Rost, B. and Sander, C. (1994) 'Conservation and prediction of solvent accessibility in protein families', *Proteins: Structure, Function, and Bioinformatics*, 20(3), pp. 216–226. doi: <https://doi.org/10.1002/prot.340200303>.

Rudolph, M. G., Stanfield, R. L. and Wilson, I. A. (2006) 'How TCRs bind MHCs, peptides, and coreceptors.', *Annual review of immunology*, 24, pp. 419–466. doi: 10.1146/annurev.immunol.23.021704.115658.

Sádio, F. *et al.* (2020) 'Stabilization of soluble high-affinity T-cell receptor with de novo disulfide bonds', *FEBS letters*. 2019/10/08, 594(3), pp. 477–490. doi: 10.1002/1873-3468.13616.

Sakhnini, L. I. *et al.* (2019) 'Improving the Developability of an Antigen Binding Fragment by Aspartate Substitutions.', *Biochemistry*, 58(24), pp. 2750–2759. doi: 10.1021/acs.biochem.9b00251.

Schwaigerlehner, L. *et al.* (2019) 'Germinality does not necessarily define mAb expression and thermal stability', *Applied microbiology and biotechnology*. 2019/07/26, 103(18), pp. 7505–7518. doi: 10.1007/s00253-019-09998-3.

Schwehm, J. M. *et al.* (1998) 'Stability effects of increasing the hydrophobicity of solvent-exposed side chains in staphylococcal nuclease.', *Biochemistry*, 37(19), pp. 6939–6948. doi: 10.1021/bi9725069.

Scott, D. R. *et al.* (2011) 'Disparate degrees of hypervariable loop flexibility control T-cell receptor cross-reactivity, specificity, and binding mechanism', *Journal of Molecular Biology*, 414(3), pp. 385–400. doi: 10.1016/j.jmb.2011.10.006.

Sharma, P. and Kranz, D. M. (2018) 'Subtle changes at the variable domain interface of the T-cell receptor can strongly increase affinity', *Journal of Biological Chemistry*, 293(5), pp. 1820–1834. doi: 10.1074/jbc.M117.814152.

Sharon, E. *et al.* (2016) 'Genetic variation in MHC proteins is associated with T cell receptor expression biases', *Nature genetics*, 48(9), pp. 995–1002. doi: 10.1038/ng.3625.

Shcherbinin, D. S., Belousov, V. A. and Shugay, M. (2020) 'Comprehensive analysis of structural and sequencing data reveals almost unconstrained chain pairing in TCR $\alpha\beta$  complex', *PLoS Computational Biology*, 16(3), pp. 1–17. doi: 10.1371/journal.pcbi.1007714.

Shehata, L. *et al.* (2019) 'Affinity Maturation Enhances Antibody Specificity but Compromises Conformational Stability Report Affinity Maturation Enhances Antibody Specificity but Compromises Conformational Stability', *CellReports*, 28(13), pp. 3300-3308.e4. doi: 10.1016/j.celrep.2019.08.056.

Sheng, Z. *et al.* (2017) 'Gene-Specific Substitution Profiles Describe the Types and Frequencies of Amino Acid Changes during Antibody Somatic Hypermutation.', *Frontiers in immunology*, 8, p. 537. doi: 10.3389/fimmu.2017.00537.

Shi, S. *et al.* (2013) 'DSF method optimization and its application in predicting protein thermal aggregation kinetics', *journal of pharmaceutical sciences*, 102(8), pp. 2471–2483. doi: 10.1002/jps.23633.

Shusta, E. V. *et al.* (2000) 'Directed evolution of a stable scaffold for T-cell receptor engineering', *Nature Biotechnology*, 18(7), pp. 754–759. doi: 10.1038/77325.

Shusta, E. V *et al.* (1999) 'Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency.', *Journal of molecular biology*,

292(5), pp. 949–56. doi: 10.1006/jmbi.1999.3130.

De Simone, M., Rossetti, G. and Pagani, M. (2018) ‘Single Cell T Cell Receptor Sequencing: Techniques and Future Challenges’, *Frontiers in Immunology*. Available at: <https://www.frontiersin.org/article/10.3389/fimmu.2018.01638>.

Smith, G. P. (1985) ‘Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface’, *Science*, 228(4705), pp. 1315–1317. doi: 10.1126/science.4001944.

Smith, S. N. *et al.* (2014) ‘Changing the peptide specificity of a human T-cell receptor by directed evolution’, *Nature Communications*, 5. doi: 10.1038/ncomms6223.

Sommermeier, D. *et al.* (2006) ‘Designer T cells by T cell receptor replacement.’, *European journal of immunology*, 36(11), pp. 3052–3059. doi: 10.1002/eji.200636539.

Spindler, M. J. *et al.* (2020) ‘Massively parallel interrogation and mining of natively paired human TCR $\alpha\beta$  repertoires’, *Nature Biotechnology*, 38(5), pp. 609–619. doi: 10.1038/s41587-020-0438-y.

Stadinski, B. D. *et al.* (2014) ‘Effect of CDR3 Sequences and Distal V Gene Residues in Regulating TCR–MHC Contacts and Ligand Specificity’, *The Journal of Immunology*, 192(12), pp. 6071–6082. doi: 10.4049/jimmunol.1303209.

Stadtmauer, E. A. *et al.* (2020) ‘CRISPR-engineered T cells in patients with refractory cancer’, *Science*, 367(6481), p. eaba7365. doi: 10.1126/science.aba7365.

Stanfield, R. L. *et al.* (2006) ‘Antibody Elbow Angles are Influenced by their Light Chain Class’, *Journal of Molecular Biology*, 357(5), pp. 1566–1574. doi: <https://doi.org/10.1016/j.jmb.2006.01.023>.

Sternke, M., Tripp, K. W. and Barrick, D. (2019) ‘Consensus sequence design as a general strategy to create hyperstable, biologically active proteins’, *Proceedings of the National Academy of Sciences*, 116(23), pp. 11275–11284. doi: 10.1073/PNAS.1816707116.

- Strickler, S. S. *et al.* (2006) 'Protein Stability and Surface Electrostatics: A Charged Relationship', *Biochemistry*, 45(9), pp. 2761–2766. doi: 10.1021/bi0600143.
- Strub, C. *et al.* (2004) 'Mutation of exposed hydrophobic amino acids to arginine to increase protein stability', *BMC Biochemistry*, 5(1), p. 9. doi: 10.1186/1471-2091-5-9.
- Sweet, R. M. and Eisenberg, D. (1983) 'Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure.', *Journal of molecular biology*, 171(4), pp. 479–488. doi: 10.1016/0022-2836(83)90041-4.
- Szeto, C. *et al.* (2021) 'TCR recognition of peptide–MHC-I: Rule makers and breakers', *International Journal of Molecular Sciences*, 22(1), pp. 1–26. doi: 10.3390/ijms22010068.
- The Antibody Society (no date) *Therapeutic monoclonal antibodies approved or in review in the EU or US (Accessed 08/07/21)*. Available at: [www.antibodysociety.org](http://www.antibodysociety.org) (Accessed: 8 July 2021).
- Thomas, S. *et al.* (2019) 'Framework engineering to produce dominant T cell receptors with enhanced antigen-specific function', *Nature Communications*, 10(1), pp. 1–15. doi: 10.1038/s41467-019-12441-w.
- Voss, R.-H. *et al.* (2006) 'Redirection of T cells by delivering a transgenic mouse-derived MDM2 tumor antigen-specific TCR and its humanized derivative is governed by the CD8 coreceptor and affects natural human TCR expression', *Immunologic Research*, 34(1), pp. 67–87. doi: 10.1385/IR:34:1:67.
- Voss, R.-H. *et al.* (2008) 'Molecular Design of the Ca $\beta$  Interface Favors Specific Pairing of Introduced TCRA $\beta$  in Human T Cells', *The Journal of Immunology*, 180(1), pp. 391 LP – 401. doi: 10.4049/jimmunol.180.1.391.
- Wagner, E. K. *et al.* (2019) 'Human cytomegalovirus-specific T-cell receptor engineered for high affinity and soluble expression using mammalian cell display', *Journal of Biological Chemistry*, 294(15), pp. 5790–5804. doi:

10.1074/jbc.RA118.007187.

Warszawski, S. *et al.* (2019) 'Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces', *PLoS Computational Biology*, 15(8), pp. 1–24. doi: 10.1371/journal.pcbi.1007207.

Willcox, B. E. *et al.* (1999) 'Production of soluble alphabeta T-cell receptor heterodimers suitable for biophysical analysis of ligand binding.', *Protein science : a publication of the Protein Society*, 8(11), pp. 2418–2423. doi: 10.1110/ps.8.11.2418.

Willcox, C. R., Mohammed, F. and Willcox, B. E. (2020) 'The distinct MHC-unrestricted immunobiology of innate-like and adaptive-like human  $\gamma\delta$  T cell subsets—Nature's CAR-T cells', *Immunological Reviews*, 298(1), pp. 25–46. doi: <https://doi.org/10.1111/imr.12928>.

Wong, W. K., Leem, J. and Deane, C. M. (2019) 'Comparative Analysis of the CDR Loops of Antigen Receptors', *Frontiers in Immunology*, 10, p. 2454. doi: 10.3389/fimmu.2019.02454.

Wozniak-Knopp, G., Stadlmann, J. and R ker, F. (2012) 'Stabilisation of the Fc fragment of human IgG1 by engineered intradomain disulfide bonds.', *PloS one*, 7(1), p. e30083. doi: 10.1371/journal.pone.0030083.

Xu, X., Li, H. and Xu, C. (2020) 'Structural understanding of T cell receptor triggering', *Cellular & Molecular Immunology*, 17(3), pp. 193–202. doi: 10.1038/s41423-020-0367-1.

Xu, Yingda *et al.* (2019) 'Structure, heterogeneity and developability assessment of therapeutic antibodies', *mAbs*. 2018/12/17, 11(2), pp. 239–264. doi: 10.1080/19420862.2018.1553476.

Xu, Yixiang *et al.* (2019) 'T-cell receptor mimic (TCRm) antibody therapeutics against intracellular proteins', *Antibody Therapeutics*, 2(1), pp. 22–32. doi: 10.1093/abt/tbz001.

Yang, N. *et al.* (2018) 'Use of In Vitro Systems To Model In Vivo Degradation of Therapeutic Monoclonal Antibodies.', *Analytical chemistry*, 90(13), pp. 7896–7902. doi: 10.1021/acs.analchem.8b00183.

Yoshida, T. *et al.* (2019) 'Differential scanning fluorimetric analysis of the amino-acid binding to taste receptor using a model receptor protein, the ligand-binding domain of fish T1r2a/T1r3', *PLOS ONE*, 14(10), p. e0218909. Available at: <https://doi.org/10.1371/journal.pone.0218909>.

Young, L., Jernigan, R. L. and Covell, D. G. (1994) 'A role for surface hydrophobicity in protein-protein recognition.', *Protein science : a publication of the Protein Society*, 3(5), pp. 717–729. doi: 10.1002/pro.5560030501.

Zareie, P. *et al.* (2021) 'Canonical T cell receptor docking on peptide-MHC is essential for T cell signaling.', *Science (New York, N.Y.)*, 372(6546). doi: 10.1126/science.abe9124.

Zhang, C. *et al.* (2018) 'Computational Design To Reduce Conformational Flexibility and Aggregation Rates of an Antibody Fab Fragment', *Molecular Pharmaceutics*, 15(8), pp. 3079–3092. doi: 10.1021/acs.molpharmaceut.8b00186.

Zoete, V. *et al.* (2013) 'Structure-Based, Rational Design of T Cell Receptors ', *Frontiers in Immunology* , p. 268. Available at: <https://www.frontiersin.org/article/10.3389/fimmu.2013.00268>.