ABSTRACT

Title of Thesis: STRUCTURAL BASIS FOR CLONAL DIVERSITY OF THE PUBLIC T CELL RESPONSE TO DOMINANT EPITOPES FROM CYTOMEGALOVIRUS AND INFLUENZA

Xinbo Yang, Doctor of Philosophy, 2016

Thesis Directed By: Professor Roy A. Mariuzza
Department of Cell Biology and Molecular Genetics

A diverse T cell receptor (TCR) repertoire is a prerequisite for effective viral clearance. However, knowledge of human TCR repertoire to defined viral antigens is limited. Recent advances in high-throughput sequencing (HTS) and single-cell sorting have revolutionized the study of human TCR repertoires to different types of viruses. In collaboration with the laboratory of Dr. Nan-ping Weng (National Institute on Aging, NIH), we applied unique molecular identifier (UMI)-labelled HTS, single-cell paired TCR analysis, surface plasmon resonance, and X-ray crystallography to exhaustively interrogate CD8+ TCR repertoires specific for cytomegalovirus (CMV) and influenza A (Flu) in HLA-A2+ humans. Our two CMV-specific TCR-pMHC structures and two Flu-specific TCR-pMHC structures provide a plausible explanation for the much higher diversity of CMV-specific than Flu-specific TCR repertoires in humans. Our comprehensive biochemical and structural portrait of two different anti-viral T cell responses may contribute to the future development of predictors of immunity or disease at the individual level.

STRUCTURAL BASIS FOR CLONAL DIVERSITY OF THE PUBLIC T CELL RESPONSE TO DOMINANT EPITOPES FROM CYTOMEGALOVIRUS AND INFLUENZA

by

Xinbo Yang

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Roy A. Mariuzza, Chair
Professor Dorothy Beckett
Professor David Mosser
Professor John Orban
Professor Nicole A. LaRonde-LeBlanc

Dedication

I dedicate this dissertation to my parents Yin Zhang and Junxing Yang, My grandfather Kezhan Zhang, as well as my wife Yang Yang, for their complete love.

Acknowledgements

I want to thank these people that have supported me during my graduate work. Without them, this thesis would have been impossible.

To my advisor Roy Mariuzza, I have always felt fortunate and grateful to be Roy's Ph.D. student. To introduce me into structural immunology, Roy assigned me to work with Dr. Carlos A. Velikovsky and learn essential skills for structural biology, such as molecular cloning, protein overexpression and protein purification. These skills guaranteed me a smooth, efficient start and most importantly made me confident to devote myself entirely in my projects. During the later stage of my training, Roy influenced me with his broad knowledge in immunology and structural biology. He guided me with great patience and supplied me with resources and ideas. I have always enjoyed all the discussions with Roy because I can learn so much from him. Roy also taught me how to write scientific papers and grant applications, as these are essential skills to being a good scientist. Above all, Roy always encouraged me to pursue a scientific career and I deeply appreciate having him as my Ph.D. mentor.

To Drs. Dorothy Beckett, David Mosser, John Orban and Nicole LaRonde, I have taken classes from you and you are always nice and patient with me. You prepared me well with essential knowledges and theories in immunology, biochemistry and structural biology. You all are accomplished scientists in your fields and I am definitely grateful to have your constructive advice through my training.

To all the members in Roy's lab: Dr. Yili Li, Dr. Mingming Gao, Dr. Carlos A. Velikovsky, Dr. Sneha Rangarajan, Ms Qian Wang and Ms Melissa Kezic, I have learned all the

lab skills to be a structural biologist from you. Also, because of you, the lab is full of fun and I never felt bored.

To all the staffs in BISI program, thanks for helping me throughout my training. You were always available when I needed assist. A special thank goes to Dr. Michelle Brooks and Gwen Warman. When I have issues, I can always reach you.

To all my friends in College Park and back home, it has been extremely fun to team with you guys in Defense of the Ancients (DOTA). Wish all of you enjoy your career, life and DOTA.

To my parents Yin Zhang and Junxing Yang and my wife Yang Yang, you always support and encourage me to pursue a scientific career even though at the time there was only a little hope. I can only repay you with my hard work and achievements in the future. To my grandfather, you are a veterinarian and have influenced me with your persistence and optimism which are the most important characteristics for being a good scientist. Thanks to all for your endless assistance and love!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

APC	Antigen-presenting cell
CMV	Cytomegalovirus
EBV	Epstein-Barr virus
CDR	Complementarity-determining region
FPLC	Fast protein liquid chromatography
GIL	GILGFVFTL
HLA	Human leukocyte antigen
MHC	Major histocompatibility complex
MR	Molecular replacement
NLV	NLVPMVATV
PDB	Protein Data Bank
PEG	Polyethylene glycol
pMHC	Peptide-MHC
SPR	Surface plasmon resonance
TCR	T cell receptor

Chapter 1

Introduction

Our immune system is elegantly designed to fight against pathogenic invasions throughout our entire lives. Three lines of defense are built to battle with pathogens from the surface of the skin to deep inside organs. External physical barriers, including skin, mucous membranes, anti-microbial peptides, cilia and fluid flow represent our first line of immune defense. Once the pathogens breach the first line, they encounter various types of innate immune cells, including macrophages, neutrophils, natural killer cells and dendritic cells. These cells recognize microbial components such as bacteria lipoproteins, flagella or nucleic acids by different types of pattern recognition receptors (PRR). These interactions initiate a series downstream signaling events and eventually lead to inflammatory cytokine secretion and antigen-presenting cell (APC) migration to secondary lymph nodes. APC migration and cytokines awaken the adaptive immune system, which includes T cell and B cells. Compared to the rapid but much less diverse innate immune system, the adaptive immune system is slow but extremely variable. Through gene rearrangement mechanisms, T cell and B cells can recognize virtually all antigens. Meanwhile, the adaptive immune system can develop memory for antigens so it can react much more effectively if the same pathogens invade again. Whereas B cells recognize antigens directly, T cells recognize processed peptide antigens presented by major histocompatibility complex (MHC) molecules. During thymic development, T cells generate T cell receptors (TCRs) through somatic recombination of V(D)J gene segments. The theoretical T cell diversity is 10^{15} (Nikolich-Zugich et al., 2004). Thus, it is reasonable to believe that different individual should possess distinct TCR repertoires. However, when encountering certain antigens

(e.g. influenza, CMV and EBV), TCRs are heavily biased toward certain V(D)J combinations and present uniform solutions that are shared across large populations. This apparent loss of randomness and diversity is called the T cell public response (Venturil et al., 2008). Understanding the molecular basis of public TCR generation and its interaction with cognate peptide-MHC (pMHC) ligands would help in developing better vaccines for pandemic viral infections.

1.1 T cell-mediated immune response

Adaptive immune responses are mediated by both B cells and T cells. Unlike antibody secretion by B cells, T cells protect the host by killing pathogen-infected cells or cancer cells via direct contact (Zheng and Bevan 2011). Each T lymphocyte generates one TCR heterodimer that can recognize self or non-self-peptides presented by MHC molecules. During the late stages of maturation, T cells either express CD8 or CD4 molecules on their surface. CD8 corresponds to cytotoxic T cells that recognize pMHC class I, whereas CD4 corresponds to helper T cells that recognize pMHC class II. In order to fully activate a T cell response, three types of signals are required: (1) TCR-pMHC interactions; (2) costimulatory signals from B7-CD28 interaction; and (3) instructive cytokine secretion from APCs. Once T cells are fully activated, CD8+ T cells start to secrete cytotoxins, such as perforin and granules for direct killing of target cells; CD4 T cells undergo further differentiation into Th-1, Th-2, Th-17 or Tfh cells (Zhu et al., 2010). Each subtype of T helper cell secretes different kinds of cytokines to assist other types of immune cell function, such as B cell activation and maturation in germinal centers requiring Tfh-B cell

interaction, clearance of viral infection requiring Th-1 cell INFγ secretion, and anti-microbial responses requiring IL-17 secretion by Th-17 cells (Figure 1.1).

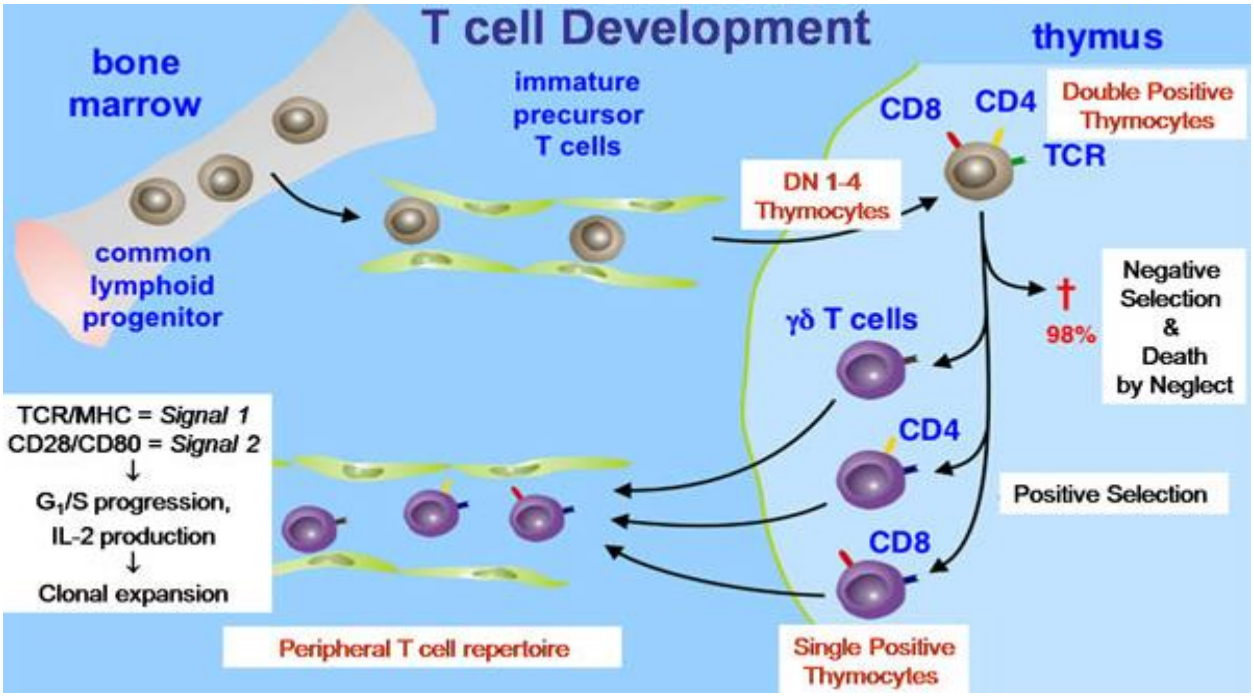


Figure 1.1: T cell development.

Immature precursor T cells derived from bone marrow migrate to the thymus and mature into CD8+ or CD4+ T cells. CD8+ T cells mediate cytotoxic killings. CD4+ T cells undergo further differentiation to various subtypes of helper T cells. Both CD8+ and CD4+ T cell activation require TCR-pMHC interactions and costimulatory signals. (www.lymphomation.org)

1.2 How T cells recognize pMHC

1.2.1 Structure of pMHC molecules

In the immune response, antigenic peptides generally are displayed to αβ T cells via class I or class II MHC molecules. Both class I and class II MHC molecules share similar structures

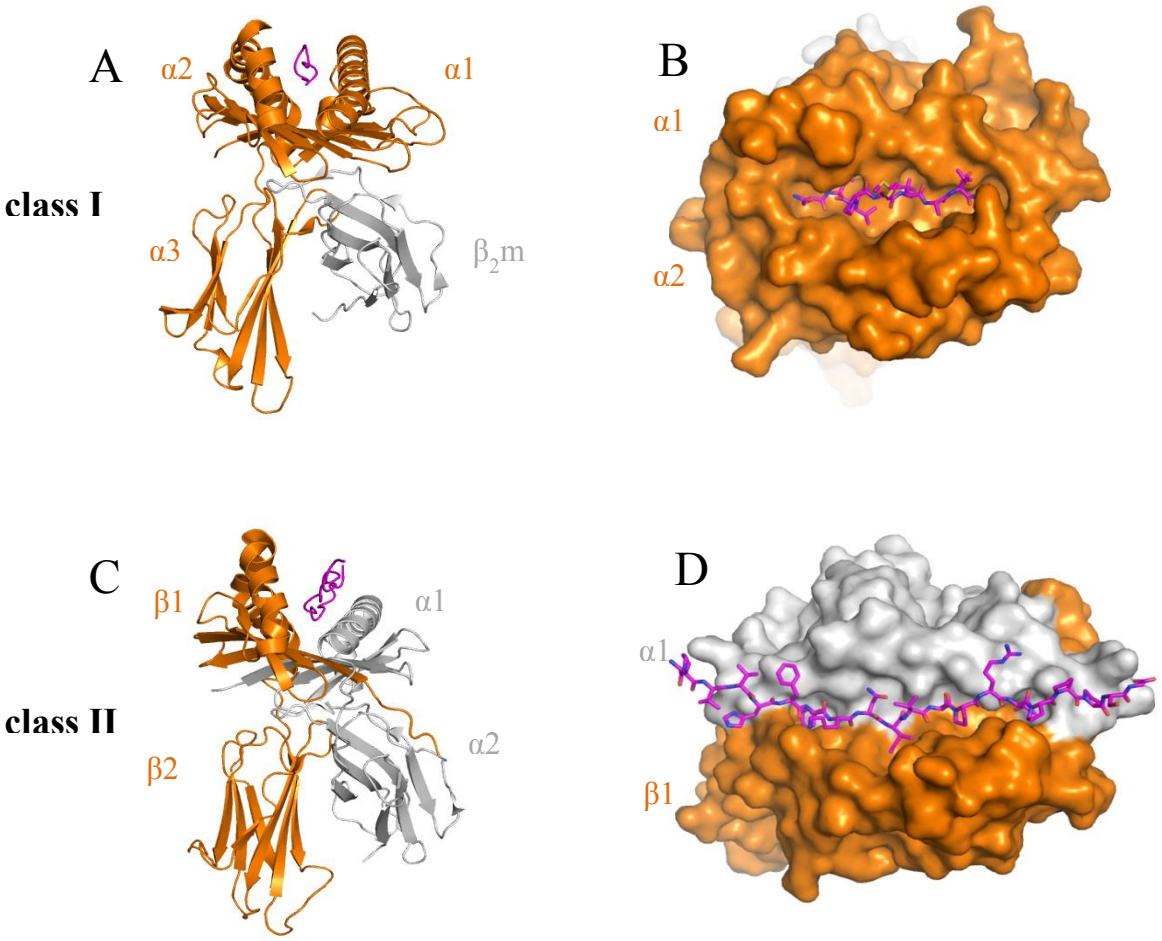


Figure 1.2: Structures of MHC class I and class II molecules.

(A) Ribbon diagram of a human MHC class I molecule, HLA-A2, bound to a 9-residue viral peptide, NLV (PDB code: 3GSO) (Gras et al., 2009). The heavy chain of HLA-A2 (shown in orange) contains three domains. β2-microglobulin is gray and the peptide is magenta.

(B) Top view of the NLV-HLA-A2 complex. The surface of HLA-A2 is orange, with the bound NLV peptide (magenta) shown as a stick model. The two ends of the peptide-binding groove are closed, which constrains the peptide length for MHC class I molecules.

(C) Ribbon diagram of a human MHC class II molecule, HLA-DR2a, bound to a 20-residue myelin basic protein (MBP) peptide (PDB code: 1FV1) (Li et al., 2000). Both the α (gray) and β (orange) chains of HLA-DR2a contain two domains.

(D) Top view of the MBP-HLA-DR2a complex. Compared to MHC class I, both ends of the peptide-binding groove in MHC class II molecules are open, which allows peptides of varying length to bind.

and are heterodimers. MHC class I molecules are expressed on almost all kinds of cells. Their peptide-binding groove is constructed from the α1and α2 domains of the heavy chain. β_2-microglobulin molecule forms a tight complex with the heavy chain α3 domain (Figure 1.2-A). In contrast to MHC class I , expression of MHC class II molecules is restricted to professional APCs, such as dendritic cells, macrophages or B cells. MHC class II proteins are assembled from two heavy chains (α1α2 and β1 β2) (Figure 1.2-C). The overall architecture is similar to that of MHC class I, whereby the binding groove is constructed by a seven-stranded β-sheet forming the floor and two long α helices forming the walls. Polymorphic residues are usually located within and around the binding groove. Sequence variation provides the structural and chemical basis for the peptide specificity of individual MHC alleles.

MHC class I molecules often present endogenous peptides produced by protease-degraded proteins. The peptide length ranges from 8 to 10 residues (Figure 1.2-B). Different MHC class I alleles offer different pockets to accommodate peptide anchor residues. Longer peptides may bind by extension at the C-terminus or adopt bulged conformations, providing additional TCR interacting hot spots. In some rare cases, MHC class I molecules are also able to present exogenous peptides and be expressed on restricted sets of APCs. This mechanism is called cross-presentation. MHC class II molecules often present exogenous peptides derived from the environment directly. The binding groove is open on both ends so that the bound peptide has no length restriction. The length of MHC class II-bound peptides can range from 13-25 residues (Figure 1.2-D) (Stern and Wiley, 1994). The peptide backbone in MHC class II sinks deeper into the binding groove than for MHC class I. Thus, the peptide presents more hot spots for TCR recognition in MHC class I due to the bulging conformation. This may explain the overall higher affinity of MHC class I-restricted TCRs than MHC class II-restricted TCRs

measured up to date. However, the MHC class II termini, particularly the N-terminal extension, may play a major role in TCR recognition.

In humans, MHC molecules are often referred to as human leukocyte antigens (HLA). Each person carries one set of HLA-A, HLA-B and HLA-C as HLA class I molecules, and HLA-DP, HLA-DR and HLA-DQ as HLA class II molecules (Lefranc et al., 2009). The products of individual HLA alleles can differ from one another by up to 20 amino acids. This polymorphic nature allows HLA molecules to present virtually any kind of peptide in their pocket and eventually drive extremely diverse T cell responses (Rudolph et al., 2006).

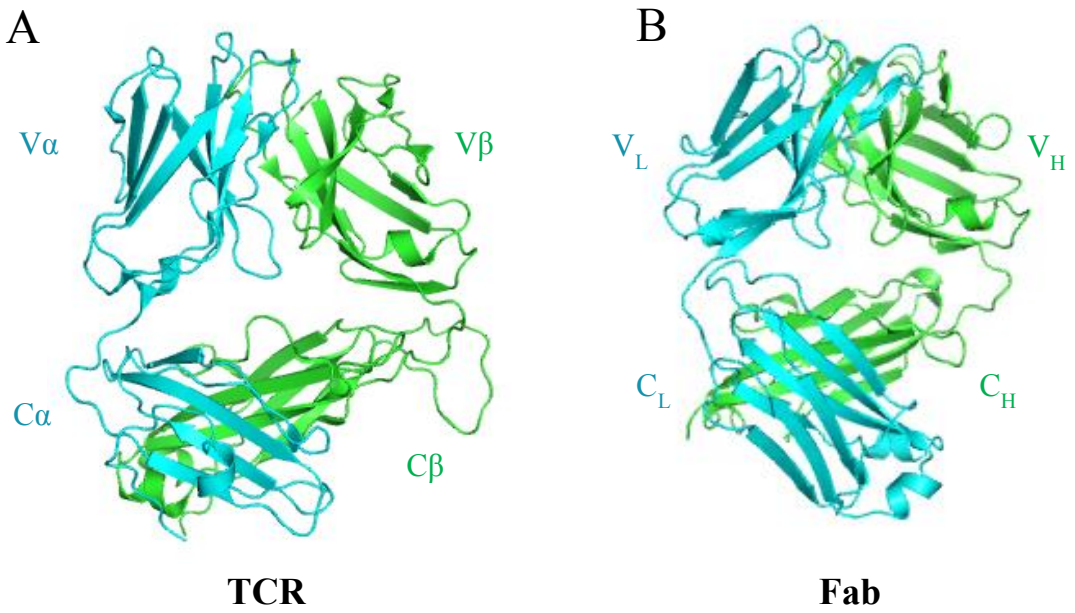


Figure 1.3 Structures of TCR and Fab fragment of an IgG.

(A) Ribbon diagram of the first human TCR structure, TCR A6 (PDB code: 1QSF) (Ding et al., 1999). The α- and β-chains are cyan and green, respectively. The individual domains are labeled.
(B) Structure of the Fab fragment of an anti-HCV antibody, HC33.4 (PDB code: 5FGB) (Keck et al., 2015). The light chain is cyan and the heavy chain is green. The individual domains are labeled.

1.2.2 Structure of the T cell receptor

The very first event of T cell activation requires TCR-pMHC engagement. Crystal structures of TCRs in the Protein Data Bank (PDB) have revealed great structural similarity between TCRs and antibody Fab fragments (Figure 1.3). Each TCR is composed of an α chain and a β chain linked by a disulfide bond in the constant domain (Rudolph et al., 2006). In the human genome, 70 TRAVs and 61 TRAJs are encoded in chromosome 14, and 52 TRBVs, 2 TRBDs and 13 TRBJs are encoded in chromosome 7 (Lefranc et al., 2009). During T cell development in the thymus, αβ TCR generation requires random gene rearrangements of variable (V) and joining (J), and random gene rearrangement of V, D and J segments from the TCRα and TCRβ loci, respectively (Figure 4). As in the case of antibodies, random V(D)J combinations, as well as random insertion, deletion and mutation in the joining regions, generate three complementary-determine regions (CDRs) in each α chain and β chain. Gene rearrangements are carried out by the lymphoid specific proteins RAG1 and RAG2, as well as by non-lymphoid specific proteins, such as DNA repair enzymes (Swanson, 2004). The random addition, mutation or deletion of nucleotides further increases repertoire size (Matthews and Oettinger, 2009; Tillman et al., 2004). The processed β chain is first paired with a pre-Tα chain and expressed on the cell surface. Pre-TCR signaling is important for T cell maturation and potentially determines the size of T cell repertoires. In later stages of T cell maturation, a rearranged α chain is paired with the β chain and the αβ heterodimer eventually associates with CD3 molecules to produce a fully functional TCR-CD3 complex.

In the canonical orthogonal docking mode, TCR engages MHC molecule with CDR1s and CDR2s, while peptide specificity is mediated by CDR3s. Interestingly, antibody CDRs are able to recognize virtually any type of antigen, whereas TCRs are solely designed for pMHC

interactions (Rudolph et al., 2006). This pMHC restriction is partially explained by conserved germline encoded-interactions of CDR1 and CDR2 with the MHC α helices. However, recent studies have demonstrated that germline-encoded interactions are not always energetically favored (Piepenbrink et al., 2013). Moreover, a quadruple mouse

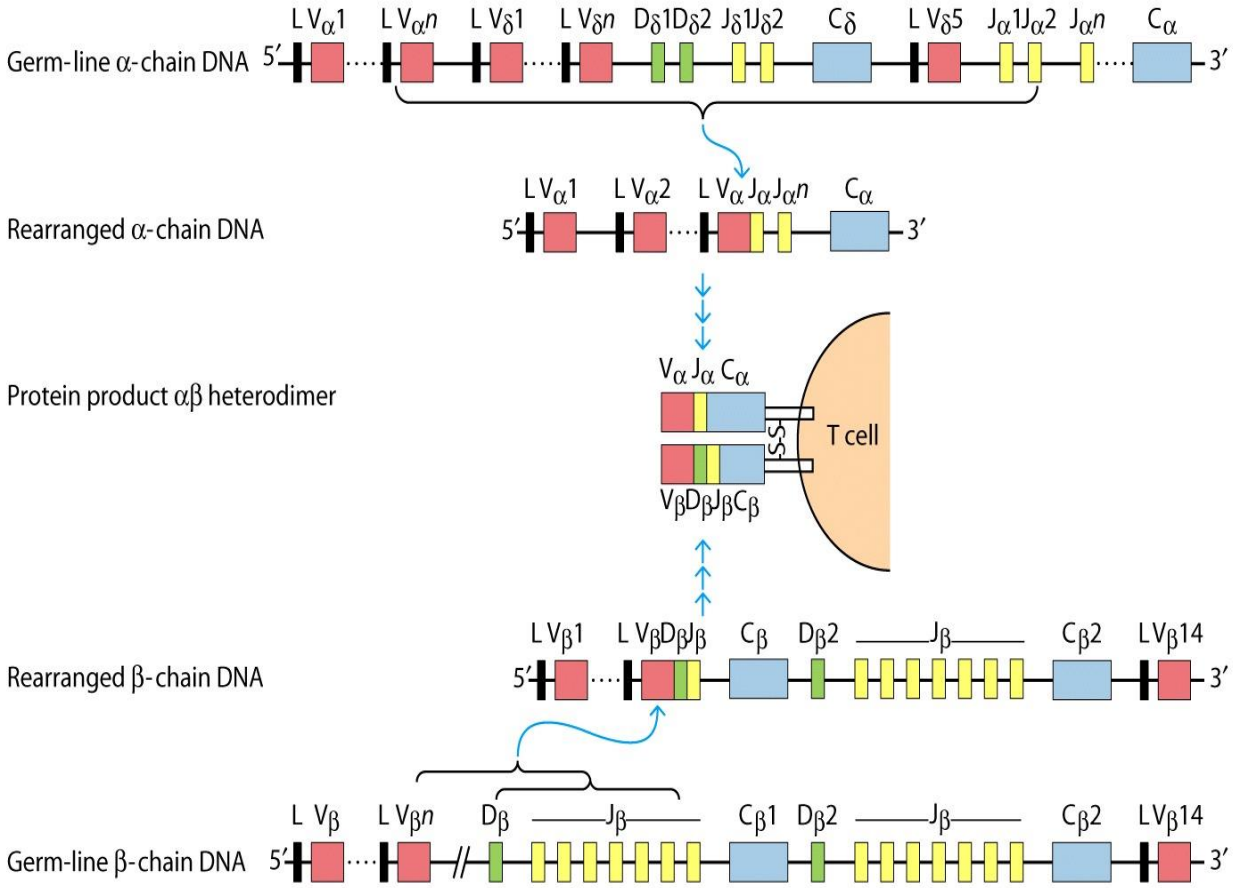


Figure 1.4 Gene rearrangements of TCR α- and β-chains.

TCR gene rearrangements are DNA recombination events that occur in early T cell development stages. After the V regions are formed by rearranged V(D)J segments, the selected genes are transcribe into mRNA, followed by splicing the V regions to connect with the C regions. It is notable that α and β gene rearrangements take place at different stages of T cell development. TCR β gene rearrangement occurs in the "double-negative" stage when each TCR β chain pairs with an invariant α chain (pre-Tα). TCR α gene rearrangement occurs in a later "double-positive" stage. The matured α chains then replace the pre-Tα to assemble fully functional αβTCR surface molecules. (Janeway's immunobiology 8th edition)

model (MHCI-/-, MHCII-/-. CD8-/- and CD4-/-) is able to generate TCRs without pMHC bias that directly bind antigenic proteins (Tikhonova et al., 2012).

1.3 Public and private T cell responses

The total theoretical diversity of TCRs is ~10^{15}, but each human only has ~10^{8} T cells (Li et al., 2015; Qi et al., 2014; Robins et al., 2010). Thus, in theory, any two individuals are unlikely to share the same TCRs. The unique T cell clones only present in one individual are called private T cell responses. However, in some cases, a T cell has lost randomness and diversity, with the response biased toward certain V(D)J combination. Such clones are highly represented in multiple individuals. This apparent loss of randomness and diversity is called the public T cell response. Public TCRs that dominantly use certain amino acid sequence are considered an unusual phenomenon. Sequence conservation has suggested selective pressure during thymic selection and exposure to the environment. Studies have revealed that public T cell responses are important in various types of disease, such as acute viral infections, persistent viral infections, and autoimmune and allo-reactive responses (Venturil et al., 2008).

The phenomenon of public T cell responses has been partially explained by structural studies of different TCR-pMHC complexes (Rossjohn et al., 2015). Initial insights into T cell public responses came from studies of HLA-B*08:01 individuals presenting an Epstein Barr virus (EBV) dominant epitope (named FLR). In a corresponding TCR-pMHC structure, P7-Tyr protruded into a pocket formed by the CDR3 loops (Kjer-Nielsen et al., 2003). In another case, HLA-B3508 presented the EBV BZLF peptide with 12 amino acids (Tynan et al., 2007). The unusually long peptide presented by HLA class I resulted in a “bulged” conformation and limited

the structural solutions for a TCR to bind the pMHC. In addition, HLA-A2 presenting the influenza virus GIL peptide has only one amino acid substantially exposed to solvent (Stewart-Jones et al., 2003). This featureless peptide has also limited structural solutions for TCRs to recognize pMHC.

1.4 Virus-specific T cell responses

1.4.1 Cytomegalovirus-specific T cell response

Human cytomegalovirus (CMV) is a ubiquitous and persistent virus that infects 60%-90% of the population worldwide and has co-evolved with its mammalian hosts over a million years. Usually, CMV infections are well controlled in healthy individuals but can cause life-threatening cases when one's immune system is compromised. Moreover, even in healthy individuals, persistent CMV infection may impact host innate and adaptive immune responses. For instance, individuals carrying CMV have substantial CMV-specific CD8+ T cell memory pools through aging (La Gruta and Turner, 2014). The most dominant CMV peptides are derived from HCMV 65kDa phosphoprotein (pp65) and 55 kDa immediate-early protein 1 (IE1). Individuals expressing the HLA class I molecules HLA-A2 (*0201) and HLA-B7 produce CD8+ T cells that recognize the immunodominant pp65 epitope NLVPMVATV (NLV) and the immunodominant IE1 epitope KARAKKDEL, respectively (Wills et al., 1996). The total population of CMV-specific T cell response in different individuals range from 5% to 30%. The reasons underlying different magnitudes of CD8+ T cell responses in different individuals are not well understood. Studies also have shown that elderly individuals tend to have higher CD8+ T cell responses than younger individuals.

The NLV-specific T cell repertoire is characterized by a high prevalence of TCRs that are frequently observed in multiple unrelated individuals. These public TCRs feature identical, or nearly identical, CDR3α and/or CDR3β sequences. TCRs expressing public CDR3α motifs may pair with different CDR3β motifs (and the reverse), giving rise to highly diverse NLV-specific TCR repertoires. These CDR3s, in addition, can pair with different variable regions. Thus, this interchangeability of TCR V regions and CDR3 motifs permits multiple structural solutions to binding an identical pMCH ligand, and thereby the generation of a clonally diverse public T cell response to CMV (Day et al., 2007; Koning et al., 2014; Neller et al., 2015; Nguyen et al., 2014; Peggs et al., 2002; Trautmann et al., 2005).

1.4.2 Influenza-specific T cell response

Influenza virus is also a ubiquitous virus and infects hosts seasonally. The typical symptoms of flu infections can be annoying, including coughing, sneezing, sore throat or fevers. Usually, influenza virus is cleared by the host immune system within several weeks. However, in 1918-1919, a pandemic influenza outbreak called “Spanish flu” took millions of lives worldwide. During influenza virus invasions, respiratory tract epithelial cells are not only infected but also serve as the first line of defense. Infected cells are able to secrete type I interferons (IFN), inflammatory cytokines, and chemokines to recruit innate and adaptive immune cells. CD8+ T cells are able to destroy infected cells by perforin/granzyme secretion or FasL-Fas mediated apoptosis. CD8+ T cells also cooperate with innate immune cells by secreting various proinflammatory cytokines, such as, IFN-γ and TNF-α. Regulatory cytokines such as IL-10 are

secreted by various sources to prevent excess CD8+ T cell-mediated cell destruction (Kim et al., 2011).

HLA-A2+ individuals are able to present the dominant M1 peptide GILGFVFTL (GIL) that is recognized by CD8+ T cells bearing public TCRs (Gotch et al., 1987; Moss et al., 1991; Lehner et al., 1995; Griffiths et al., 2015). Almost 80% of M1-specific TCRs use TRBV19 and associate with CDR3β containing the conserved IRSS motif (Gil et al., 2015; Gotch et al., 1987; Moss et al., 1991). In a study analyzing primary CD8+ T cell responses to GIL-HLA-A2 in very young children, T cells showed no signs of bias to TRBV19 but this β chain gradually become dominant during the first few years of life (Lawson et al., 2001). Thus, it is believed that in HLA-A2+ humans TRBV19 TCR is selected during the first influenza infection and reinforced during subsequent infections. In addition, GIL-specific TCR α chains are more diverse than β chains, indicating the importance of the β chain for interacting with GIL-HLA-A2.

1.4.3 High-throughput DNA sequencing and single T cell sorting isolate novel NLV/GIL-specific TCRs

Recent advances in high-throughput DNA sequencing (HTS) and single-cell sorting have revolutionized the study of human TCR repertoires in response to infection with CMV and influenza virus (Shugay et al., 2014). To date, several hundred distinct TCRβ and TCRα sequences have been reported for both CMV and influenza virus T cell response. However, it remains to be determined whether these numbers adequately represent the full diversity of these antigen-specific TCR repertoires in individuals and in populations. Dr. Nan-ping Weng (National Institute on Aging, NIH), with whom we collaborated on this project, has applied unique

molecular identifier (UMI)-labelled high-throughput sequencing (HTS) and single-cell paired TCR analysis to exhaustively interrogate virus-specific $CD8^+$ T cell repertoires in humans. His laboratory identified thousands of new NLV- and GIL-specific TCRα and TCRβ sequences, as well as dozens of distinct CDR3α and CDR3β consensus motifs. This diversity is substantially greater than previously described for T cell responses to single viral epitopes, both for private and public TCR clonotypes. In collaboration with the Weng laboratory, we obtained both NLV- (C7, C25, D12, C31, C32, C34) and GIL- (F5, F6, F8, F22, F26, F50) specific paired αβ TCR sequences from sorted single T cells. These TCRs, which include public α with private β, private α with public β, and dual public α and β chains, provide unique molecular insights into the clonal diversity of the public T cell response to CMV and Flu viral epitopes.

1.5 Biophysical and structural studies of MHC class I-restricted antigen recognition by TCRs

1.5.1 Peptide binding by MHC class I molecules

Peptides presented by MHC class I molecules are stabilized at both ends of the peptide-binding groove. The free N- and C-termini of the peptide provide the major anchor points for peptide-MHC class I complexes. Synthetic peptide analogs lacking N- or C-terminal amino acids fail to bind MHC class I molecules. Peptides presented by MHC class I molecules are usually 8-10 amino acids long (Rudolph et al., 2006). Longer peptide may also be accommodated in the peptide-binding pocket via bulged conformations. This is different from peptide binding to MHC class II molecules, which impose has no length restriction (Rudolph et al., 2006). The interaction between peptide and MHC class I molecules provides a broad peptide-binding

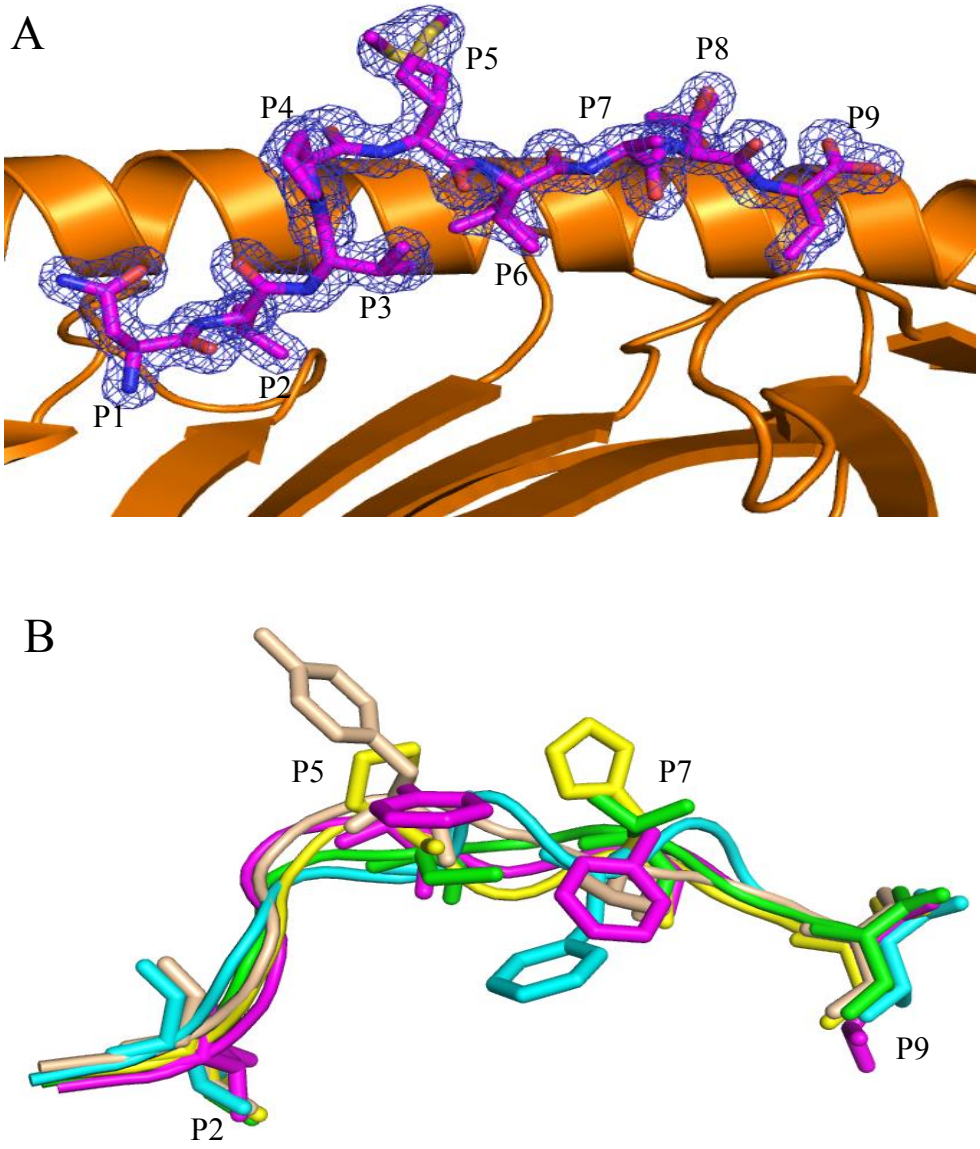


Figure 1.5 Peptide binding by MHC class I molecules.

(A) NLV peptide electron density in the NLV-HLA-A2 complex interface (PDB code: 3GSO). Density from the final $2F_o$-F_c map at 1.6 Å resolution is contoured at 1σ. Each peptide residue is labeled from P1 to P9. Magenta, NLV peptide (stick diagram); orange, HLA-A2.
(B) Five viral peptide presented by HLA-A2 molecules. Green, HIV RTase peptide (PDB code: 1HHG); cyan, HTLV Tax peptide (PDB code: 1HHH); magenta, influenza A virus M1 peptide (PDB code: 1HHI); yellow, HIV gp120 peptide (PDB code: 1HHJ); tints, HBV nucleocapsid peptide (PDB code: 1HHK). The MHC anchoring residue P2 and P9 are labeled. The TCR interacting residue P5 and P7 are labeled.

specificity. In addition, MHC molecules are highly polymorphic and some of polymorphism is located in the peptide-binding groove or TCR interacting α helices. Thus, different MHC molecules have different peptide preferences that ultimately shape the T cell repertoire.

Among all human HLA class I molecules, HLA-A2 is the most common. The structural and biophysical properties of different peptides presented by HLA-A2 molecules have been studied extensively (Bjorkman et al., 1987; Madden et al., 1993; Rudolph et al., 2006). The P2 and P9 residues are the most important anchor points; tight peptide-MHC complex formation is further reinforced by the P1, P3 and P7 residues (Figure 1.5-A). A structural study of five distinct viral peptides presented by HLA-A2 has revealed the overall similarity among all complexes, with nearly identical HLA-A2 conformations (Madden et al., 1993). However, although peptide residues P2 and P9 are bound similarly in all five cases, the main chain and side chain conformations of each peptide are strikingly different in the center of binding sites; these differences are accessible to direct TCR recognition (Figure 1.5-B). Thus, the distinct T cell responses against different viruses are likely attributable to the viral peptides rather than MHC molecules.

1.5.2 Structural basis of TCR-pMHC-I recognition

The first crystal structures of an αβ TCR-pMHC class I complex, solved twenty years ago, provided enormous information on the T cell recognition event (Figure 1.6-A) (Garcia et al., 1996). The TCR sat above the long axis of the peptide binding groove in a diagonal mode (Figure 1.6-B). The CDR3 loops interacted with peptide, whereas the germline-encoded CDR1 and CDR2 loops mediated MHC interactions. The complex interface featured low shape

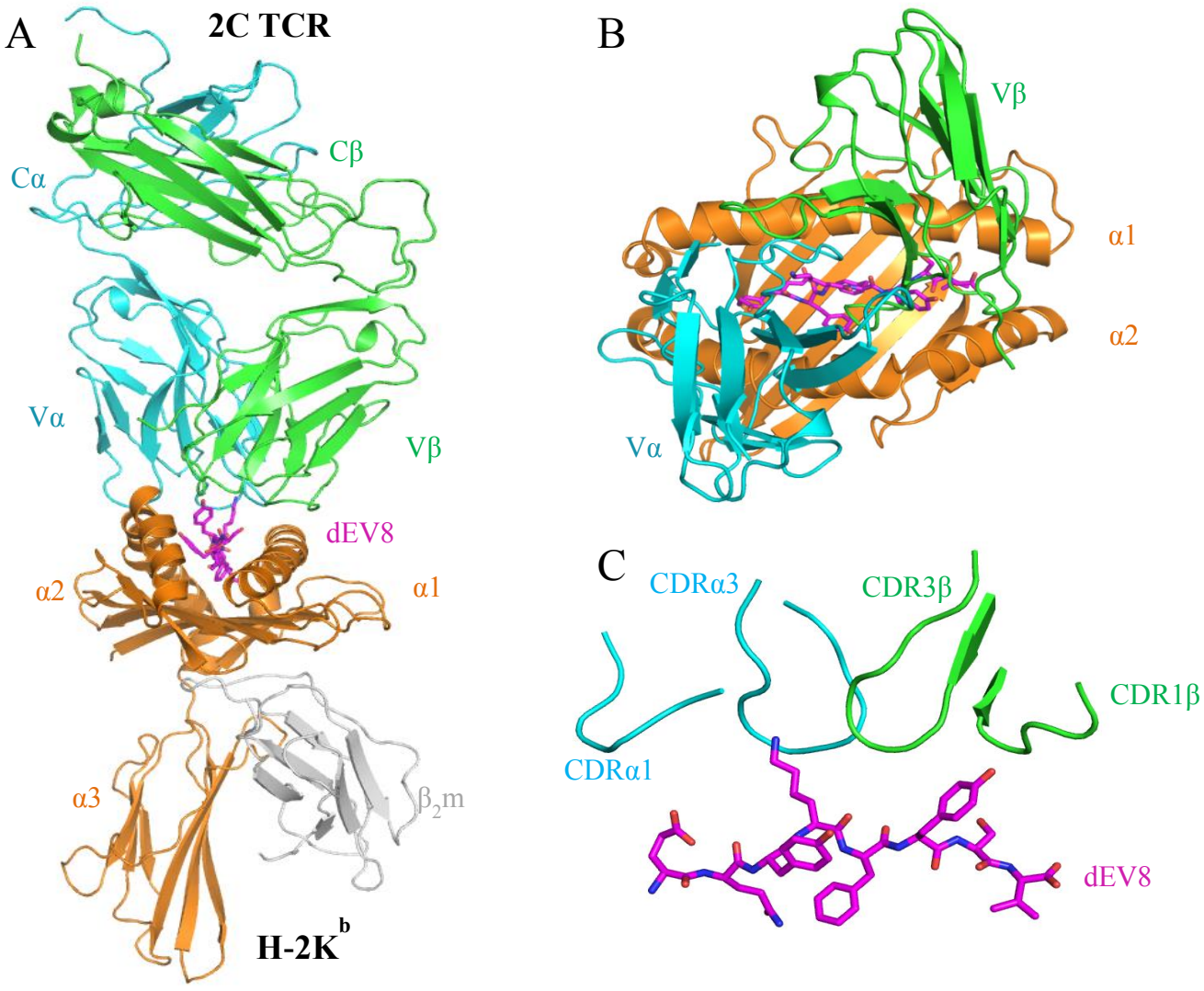


Figure 1.6 Structure of TCR-pMHC class I complex.

(A) Overview of the complex of 2C TCR and H-2K^b bound to a self-peptide, dEV8 (PDB code: 2CKB) (Garcia et al., 1996).
(B) Top view of the structure. The Cα and Cβ domains of 2C TCR, as well as the α3 and β_2-microglobulin domains of H-2K^b, are removed for clarity. The dEV8 peptide is shown as magenta sticks.
(C) Side view of the interaction of two CDR1 loops and two CDR3 loops with the HA peptide.

complementarity, which corresponds well to the relatively low affinity of the interaction. The CDR3 loops showed considerable flexibility and underwent conformational change upon pMHC binding. The flexible CDR3 loops provided the first clue of inherent TCR cross-reactivity (Figure 1.6-C).

After the first structures, many TCR-pMHC complexes have revealed the structural basis for key immunological concepts, such as TCR cross-reactivity, TCR alloreactivity, autoimmune reaction, allergies, T cell development and maturation, MHC restriction, MHC class I polymorphism and MHC class I self-tolerance (Rossjohn et al., 2015). To date, more than 30 TCR-pMHC-I complex structures have been determined. All these structures present an overall similar docking topology in which Vα interacts with the MHC α2 helix and Vβ interacts with the MHC α1 helix. In TCR-pMHC-II complexes, Vα is positioned over the MHC β1 helix and the Vβ is positioned over the MHC α1 helix. This canonical diagonal docking solution is attributed to different factors by two theories: 1. The co-evolution theory of germline-encoded TCR V residues with MHC residues. 2. The co-receptor theory of bringing CD8 or CD4 molecules proximate to CD3 molecules for effective TCR signaling. However, one study recently has shown that TCRs isolated from regulatory T cells are able to engage pMHC-II in a reversed docking orientation (Beringer et al., 2015). Whether this reverse TCR docking applies more generally or these TCRs simply are outliers remains unknown. Further investigations to find more reverse-docking TCRs, as well as fully characterizing the function of T cells bearing reverse TCRs, are essential.

1.5.3 Structural basis for RA14 TCR recognition of NLV-HLA-A2

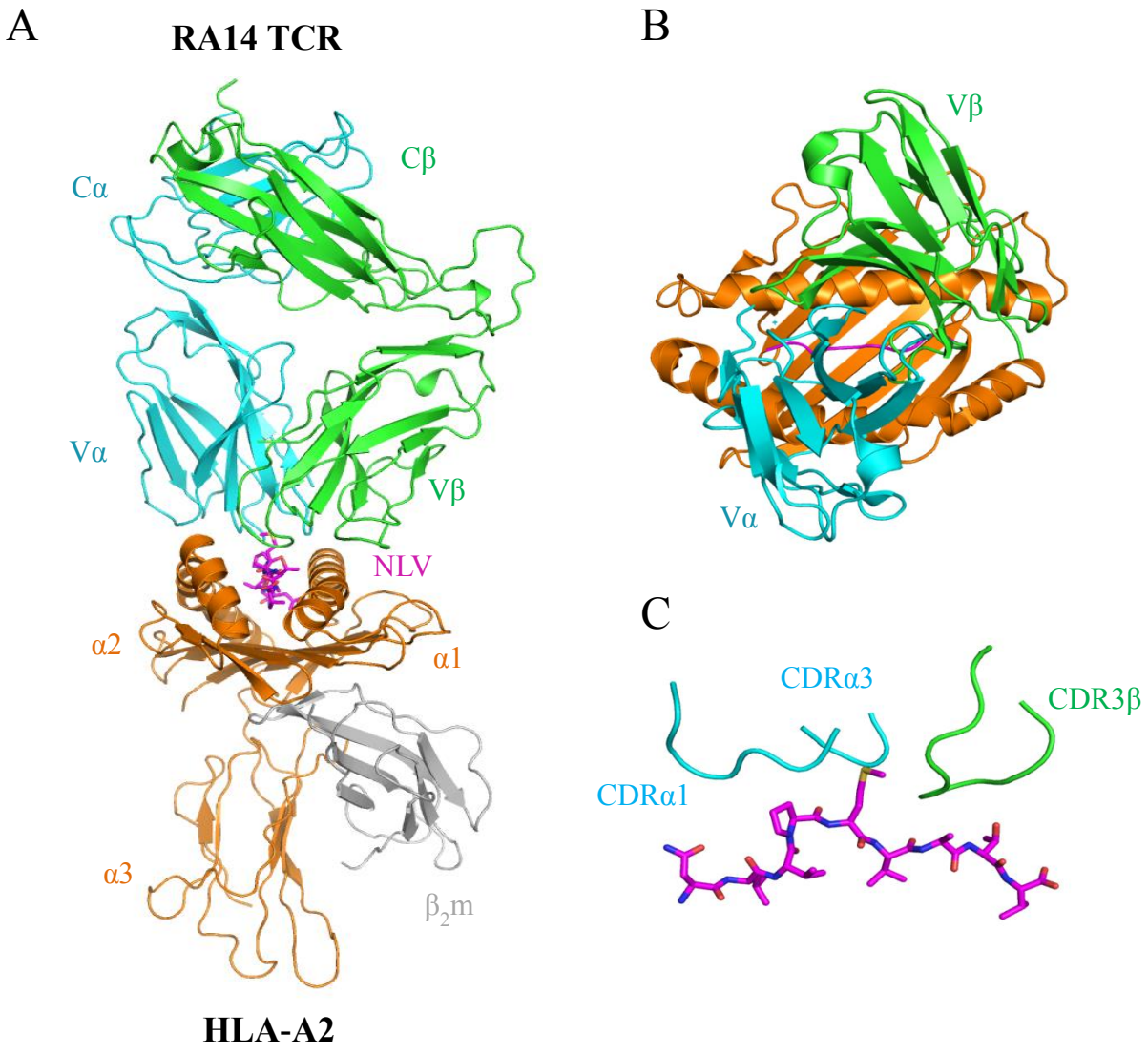


Figure 1.7 Structure of TCR RA14-NLV-HLA-A2 complex.

(A) Overview of the complex of RA14 TCR and HLA-A2 bound to a CMV pp65 peptide, NLV (PDB code: 3GSN) (Gras et al., 2009).
(B) Top view of the structure. The Cα and Cβ domains of RA14 TCR, as well as the α3 and $\beta_2$microglobulin domains of HLA-A2, are removed for clarity. The NLV peptide is shown as magenta cartoon representation.
(C) Side view of the interaction between the CDR1α, CDR3α and CDR3β loops and the NLV peptide. The NLV peptide is shown as magenta sticks.

Prior to this study, the only TCR-NLV-HLA-A2 structure available in the PDB was RA14 TCR complexed with NLV-HLA-A2 (Figure 1.7-A) (Gras et al., 2009). RA14 was originally isolated from CMV-seropositive healthy donors as well as immunosuppressed organ transplant recipients (Trautmann et al., 2005). The specific αβ pair (TRAV24-TRAJ49/TRBV6-5-TRBJ2-7) was one of the most abundant TCR pairs isolated from multiple donors. RA14 features extensive peptide readout via docking over NLV-HLA-A2 with a relatively acute angle (29°) (Figure 1.7-B). Different from most TCR-pMHC class-I structures, TCR RA14 mediates peptide binding not only with somatically-generated CDR3s, but also with germline-encoded CDR1s. In addition, The RA14-HLA-A2 germline-encoded interactions are distinct from previous TCR-pMHC structures. The particular focusing of Tyr48β and Asp56β on the Gln72H is different from previous structures in which these two well-conserved residues interact with Arg65H. This slight shift is partially explained by the acute docking angle made by RA14 and indicates that TCR-pMHC germline interactions are not necessarily hardwired by limited residues on TCR and MHC. The NLV-HLA-A2 structure in the RA14-bound and unbound states are overall the same. Upon RA14 binding, NLV peptide residues Met P5 and Thr P8 lose some mobility, as evident by weak side chain electron density in the RA14 non-bound structures.

The affinity of WT RA14 bound to NLV-HLA-A2 is slightly weaker for a typical TCR-microbial peptide-MHC interaction (K_D = 28 μM). NLV peptide point mutations have been made at the P4, P5 and P8 positions. The mutations M5S, T8A and T8V led to undetectable responses. The affinity results correspond well with RA14 T cell clone activation experiments, in which non-interacting peptides require highest peptide concentration to trigger T cell response. The RA14-NLV-HLA-A2 structure has revealed several compelling features of how a public TCR engages a viral dominant peptide. The most striking feature of the NLV-HLA-A2-specific T cell

response is its large clonal diversity. This is different from the restricted T cell repertoire against several other viruses (e.g. Flu and EBV). Thus, the RA14-NLV-HLA-A2 complex alone is unlikely to provide a full picture of the structural basis for the diverse NLV-HLA-A2-specific T cell repertoire.

1.5.4 Structural basis of JM22 TCR recognizing GIL-HLA-A2

The JM22-GIL-HLA-A2 complex represents the only available Flu-specific TCR-pMHC structure to date in the PDB (Stewart-Jones et al., 2003). Notably, JM22-GIL-HLA-A2 is also the highest resolution (1.4 Å) TCR-pMHC structure reported to date. The resolution of TCR-pMHC complexes rarely exceeds 2.5 Å. TCR-pMHC interfacial contacts are dominated by the JM22 β chain, which contains a unique CDR3β motif "XRSS". The side chain of Arg98β is inserted into a notch between the peptide and HLA-A2 complex, forming 3 hydrogen bonds with Ala150H and Q155H. These hydrogen bonds are further reinforced by 14 van der Waals contacts and 2 water-mediated bridges. The importance of Arg98β was confirmed by surface plasmon resonance (SPR) affinity measurements, as an Arg-to-Ala mutation completely abrogated the interactions between JM22 and GIL-HLA-A2. Thus, this unique solution has enabled JM22 to interact with GIL-HLA-A2 with high affinity (K_D = 3 μM), even though the GIL peptide is essentially featureless. The structure along with the biophysical experiments provided an explanation for the restricted T cell response against influenza virus.

Recently, high-throughput sequencing and single-cell αβ paired analysis have been applied to interrogate virus-specific CD8+ T cell repertoires. Hundreds of new GIL-specific TCR α and β chains, as well as dozens of distinct CDR3α and CDR3β consensus motifs, have been

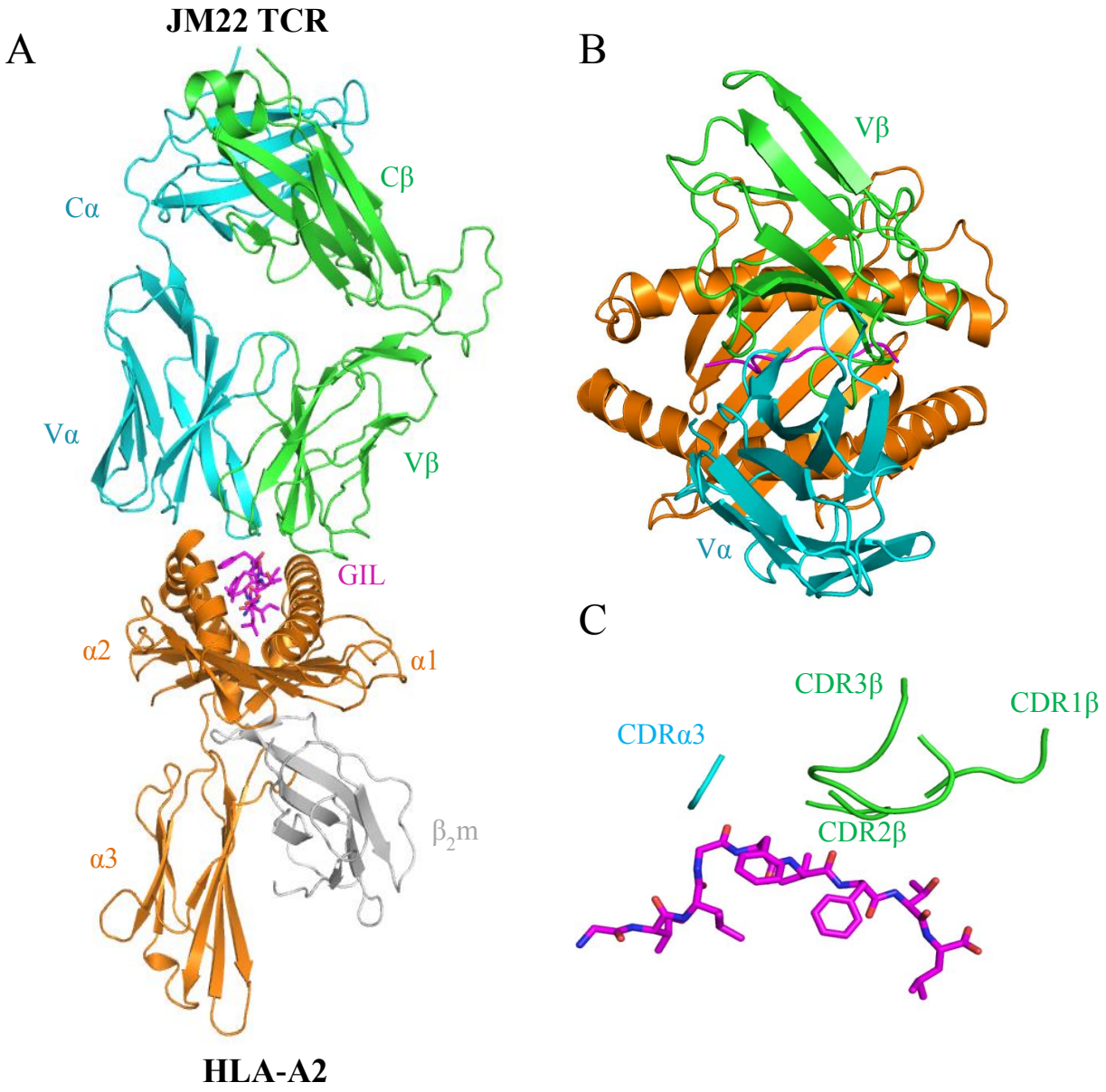


Figure 1.8 Structure of TCR JM22-GIL-HLA-A2 complex.

(A) Overview of the complex of TCR JM22 and HLA-A2 bound to a flu matrix protein peptide, GIL (PDB code: 1OGA) (Stewart-Jones et al., 2003).
(B) Top view of the structure. The Cα and Cβ domains of JM22 TCR, as well as the α3 and β_2-micrglobulin domains of HLA-A2, are removed for clarity. The GIL peptide is shown as magenta cartoon representation.
(C) Side view of the interaction between the CDR3α, CDR1β, CDR2β and CDR3β loops and the GIL peptide. The GIL peptide is shown as magenta sticks.

identified. Along with the sequencing results, several questions are brought to the fore: 1. How are TRBV19 TCRs able to associate with multiple TRAV chains? 2. How do different CDR3α loops affect TCR-GIL-HLA-A2 interactions? 3. How do non-TRBV19 TCRs engage GIL-HLA-A2? 4. Why are non-TRBV19 TCR populations small compared to TRBV19 TCRs? To answer these and related questions from a structural perspective, structures of non-TRBV19 GIL-specific TCRs, and of TRBV19 GIL-specific TCRs paired with different α chains, need to be determined in complex with GIL-HLA-A2.

In order to reveal the structural mechanisms of distinct public T cell response for NLV and GIL epitope in HLA-A2+ population, I crystallized and determined two NLV-specific TCR-NLV-HLA-A2 and two GIL-specific TCR-GIL-HLA-A2 complex structures. In this thesis, I will discuss NLV/GIL-specific TCR and NLV/GIL-HLA-A2 expression and purification protocols in chapter 2. The surface plasmon resonance experiments to study NLV/GIL-specific TCR and NLV/GIL-HLA-A2 interactions are discussed in chapter 3. The crystallization and structure determination of two NLV-specific TCR-NLV-HLA-A2 complexes and two GIL-specific TCR-GIL-HLA-A2 complexes in chapter 4 and chapter 5, respectively.

Chapter 2

Protein expression and purification of NLV-specific TCRs, GIL-specific TCRs, NLV-HLA-A2, and GIL-HLA-A2

Summary

Various methods and protein expression systems can be used to express large quantities of high quality proteins for macromolecule crystallographic studies. In my protocols, we apply *in vitro* folding to produce NLV-specific TCRs, GIL-specific TCRs, NLV-HLA-A2, and GIL-HLA-A2. For TCR production, we engineered an interchain disulfide bond in the Cα and Cβ domains to promote α/β chain association. For NLV/GIL-HLA-A2, the initial protocols which used a rapid dilution method for *in vitro* folding failed to yield sufficient amounts of proteins. An improved protocol via dialysis produced milligrams of NLV/GIL-HLA-A2. In summary, both TCRs and pMHC can be produced in large quantity and high quality.

2.1 Background

Macromolecule crystallography usually requires large quantities of proteins to obtain crystals suitable for X-ray diffraction experiments. To set up a 96 well crystal screen, at least 0.2 mg of protein is required. Therefore, the ability to produce proteins at the milligram level is absolutely prerequisite for carrying out hundreds of crystal screens. For our NLV/GIL-specific TCRs and NLV/GIL-HLA-A2, we produced these proteins in soluble form.

Due to a variety of applications, pMHC complexes are not just produced for crystallographic studies but also can be used for biophysical characterization, immunofluorescence, or antigen-specific T cell sorting by flow cytometry. pMHC-I can be produced from various sources ranging from prokaryotic to eukaryotic cells (Bjorkman et al., 1987; Fremont et al., 1992; Garboczi et al., 1992). The most common way to produce homogenous pMHC-I molecules is *in vitro* folding from bacterial inclusion bodies, with yields of about 0.5-1 mg/L. Compared to pMHC-I, pMHC-II molecules are generally more challenging to produce due to relatively weak α/β chain association. pMHC-II can be produced via refolding from bacterial inclusion bodies (Arimilli et al., 1995), as for pMHC-I, or from eukaryotic cells, such as Sf9 or Hi5 insect cells (Dessen et al., 1997; Bolin et al., 2000). The first peptide-HLA-A2 crystals were produced by releasing membrane bound peptide-HLA-A2 by papain from homozygous human lymphoblastoid cell lines (Bjorkman et al., 1987). This glycoprotein led to a 3.5 Å dataset and revealed a tremendous amount of information regarding the bound peptide and TCR recognition. At the same time, a peptide-HLA-A28 structure also was determined (Bjorkman et al., 1987). Polymorphic residues were located within the binding groove and provided insight into how MHC polymorphism could mediate peptide specificity. However, the heterogeneity of peptide loading and differential glycosylation of HLA-A2 prevented crystals

from diffracting to higher resolution. Many residues were omitted or truncated to alanine due to the lack of clear electron density.

Later protocols employed prokaryotic expression systems and produced the HLA-A2 heavy chain and β2-microgolubilin separately as bacterial inclusion bodies (Garboczi et al., 1992). The inclusion bodies were dissolved in 8 M urea and rapidly diluted into refolding mixture. Unlike recombinant pMHC-II molecules in which the peptide is covalently linked to the β subunit via a flexible hydrophilic linker, MHC-I antigenic peptide are chemically synthesized and directly added to the refolding mixture. The folding mixture was concentrated after 48 hours incubation at 4 ^{o}C and subjected to gel filtration and ion exchange purification. The purified protein reached 99% purity, as judged by SDS-PAGE. Homogenous peptide loading and lack of glycosylation led to formation of high quality peptide-HLA-A2 crystals. Nowadays, the majority of pMHC-I molecules are produced similarly to this protocol. The typical resolution for pMHC-I structures ranges from 1.5-2.5 Å. The vast improvement of resolution greatly expanded our knowledge of various immunological concepts, such as MHC polymorphism and MHC peptide specificity.

TCR molecules also can be produced using either prokaryotic or eukaryotic systems. The first αβTCR-pMHC complex crystals were grown using a TCR from insect cells (Garcia et al., 1996). This protocol is still widely used by different labs today for various purposes. However, the TCR αβ heterodimer may gradually fall apart within a week. This instability and heavy glycosylation of the protein from insect cells are unattractive characteristics for macromolecule crystallization. However, insect cell production is still a viable approach for γδTCRs that are challenging to refold (Adams et al., 2006). To overcome the unstable nature of αβ heterodimer association, various strategies have been developed. For example, a single

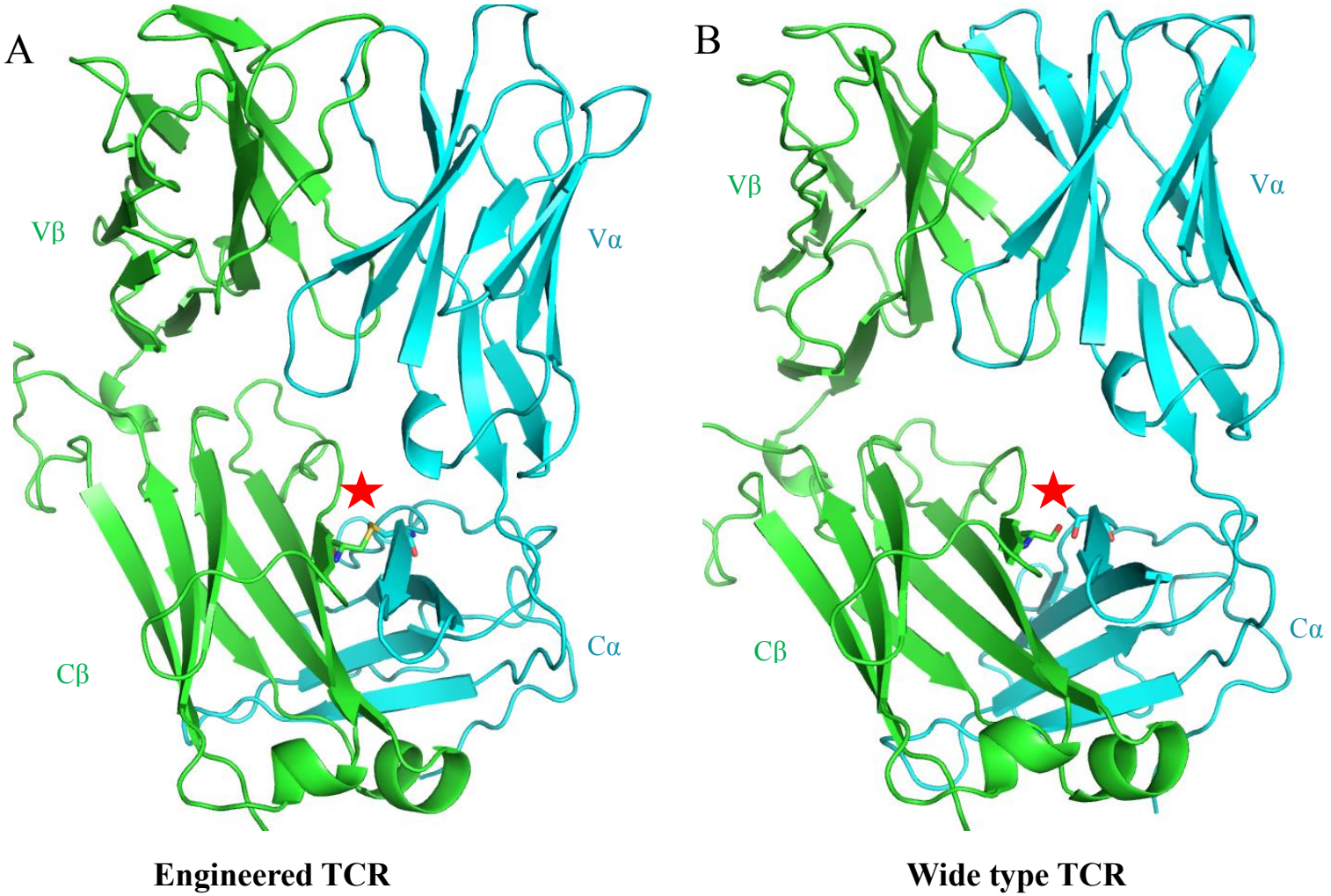


Figure 2.1 Structure of interchain disulfide bonded engineered TCR and wide type TCR.

(A) Structure of the interchain disulfide bonded engineered C25 TCR (PDB code: 5D2N) (Yang et al., 2015). The red star marks the engineered disulfide bond in C25 TCR Cα and Cβ region.
(B) Structure of wide type JM22 TCR (PDB code: 1OGA) (Stewart-Jones et al., 2003). The red star marks the original Cα Thr and Cβ Ser amino acids.

chain form of TCR (scTCR) was designed by connecting the Vα and Vβ domains via a flexible linker (Schodin et al., 1996). However, by eliminating the Cα and Cβ domains, the solubility of the scTCR is greatly reduced. Aggregates generally form throughout the purification process and the final yields are usually not high, which limits the usefulness of scTCRs for potential applications.

Another method involves covalently linking the TCR α and β ectodomain via an interchain disulfide bond formed by a pair of cysteine residues engineered in the middle of the C domains (Figure 2.1) (Boutler et al., 2003). This artificial disulfide bond has dramatically enhanced the stability of αβ TCRs and improved overall yields. In addition, this introduced disulfide bond does not change the biochemical properties of the TCR. Currently, many labs have adopted this engineered version of the TCR and most TCRs produced from *in vitro* folding contain this engineered disulfide bond.

In order to produce NLV/GIL-HLA-A2 and NLV/GIL-specific TCRs in large quantity with great purity, we applied *in vitro* folding for both pMHC and TCRs. For TCRs, we used the constructs contain the engineered disulfide bond to promote efficient αβ chain association.

2.2 Results

2.2.1 Production of NLV-HLA-A2

CMV-specific cytotoxic T lymphocytes (CTLs) recognize the immunodominant epitope pp65 495—503 (NLVPMVATV) presented by HLA-A2. Initially, the magnitude of the pp65 CTL response was compared with responses to Immediate Early gene product I (IE1) and

glycoprotein B (gB). The pp65 CTL response is substantially higher than to the other two antigens and a 15-mer (ARNLVPMVATVQGQN) was identified in HLA-A2+ donors as the antigenic peptide. The 15-mer was further simplified as a 10-mer containing P2 leucine and P9 valine (NLVPMVATVQ). The CTL responses to the 15-mer and 10-mer were essentially identical. Subsequent studies identified a 9-mer (NLVPMVATV) as the immunodominant epitope. We directly synthesized this 9-mer peptide (referred to here as NLV) and refolded it with HLA-A2 heavy chain and β2-microglobulin inclusion bodies.

i. Low yield of NLV/GIL-HLA-A2 via rapid dilution

The chemically synthesized NLV and GIL peptides were dissolved in deionized water and dimethyl sulfoxide (DMSO), respectively. The peptides were added to a 1 L refolding mixture containing 100 mM Tris (pH 8.0), 400 mM L-arginine-HCl, 2 mM EDTA, 5 mM reduced glutathione, and 0.5 mM oxidized glutathione. Subsequently, 21 mg HLA-A2 heavy chain and 10 mg β2-microglobulin inclusion bodies were added to the refolding mixture and incubated in 4 C° for 3 days. Correctly folded NLV/GIL-HLA-A2 was purified by three consecutive purification steps and the protein was concentrated to 10 mg/ml. However, the overall yield ranged from 0.5-1.0 mg/L, which is lower than the expected yield based on previous reports.

ii. Satisfactory yield of NLV/GIL-HLA-A2 via dialysis

To produce NLV-HLA-A2 in larger quantity, we needed to modify our refolding protocol. The rationale is that HLA-A2 heavy chain inclusion bodies aggregated before they could fold correctly. Therefore, we switched to a protocol that could slow down the refolding process. We

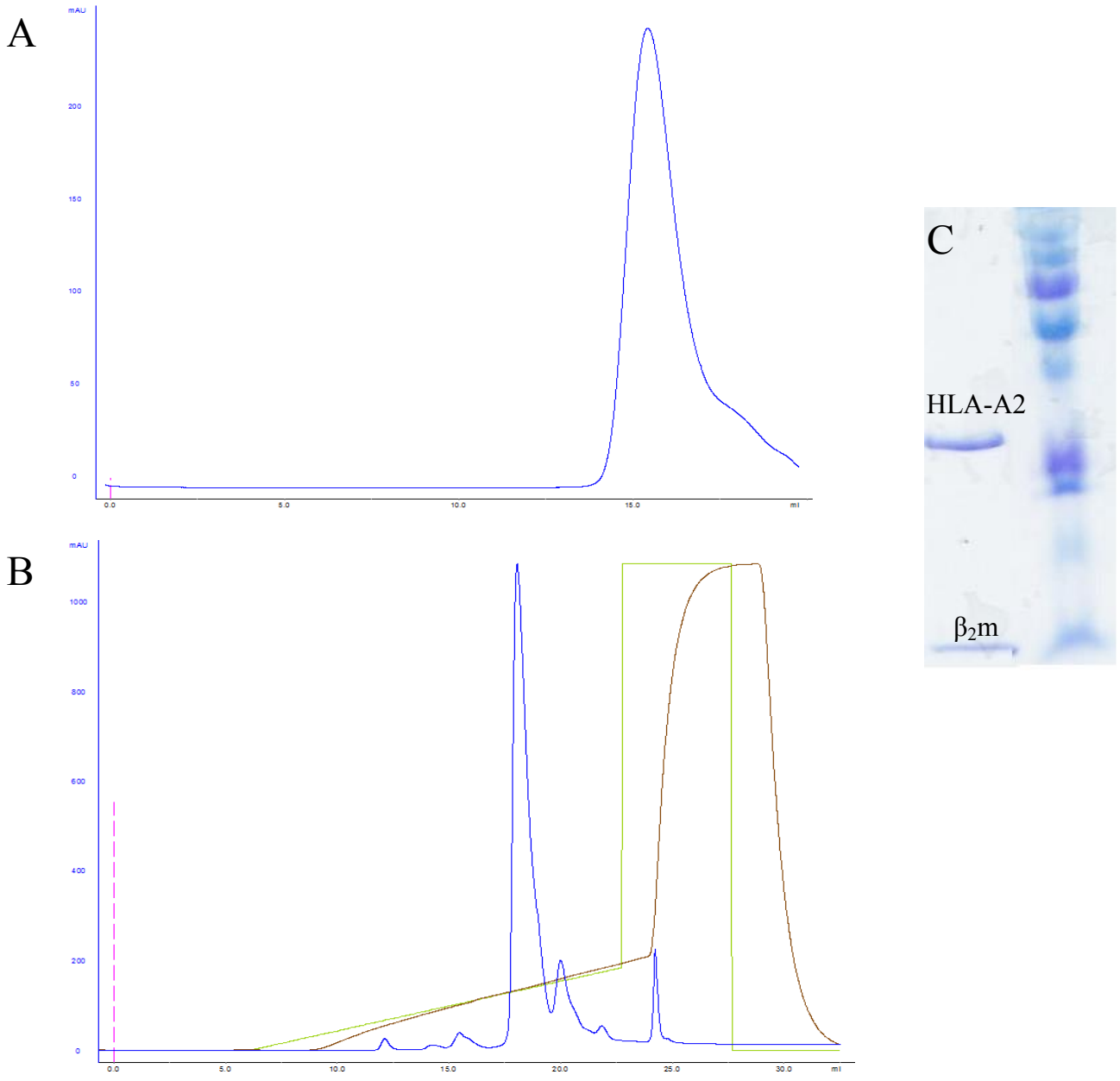


Figure 2.2 Protein purification profile of NLV/GIL-HLA-A2.

(A) Size exclusion chromatography of NLV/GIL-HLA-A2 using a Superdex-200 column. The arrow indicates the fraction containing the folded protein.
(B) Anion exchange chromatography of NLV/GIL-HLA-A2 using a MonoQ column. The arrow indicates the fraction containing the folded protein.
(C) SDS-PAGE electrophoresis of the fraction containing the folded protein. The higher molecule weight band is the HLA-A2 heavy chain and the lower molecule weight band is β_2-microgolubulin.

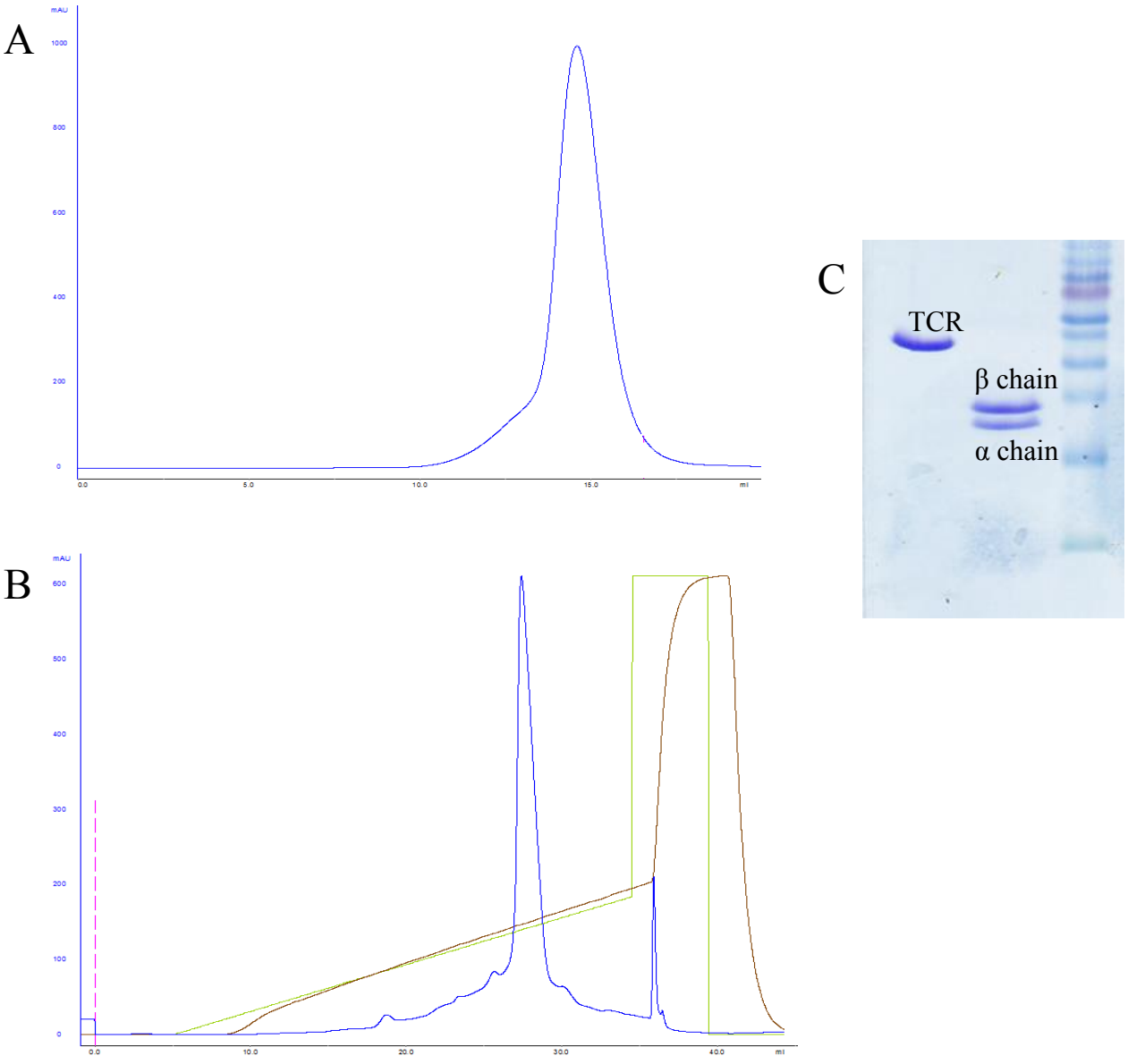


Figure 2.3 Purification of NLV/GIL-specific TCR.

(A) Size exclusion chromatography of NLV/GIL-specific TCR using a Superdex-200 column. The arrow indicates the fraction containing the folded protein.
(B) Anion exchange chromatography of NLV/GIL-specific TCR using a MonoQ column. The arrow indicates the fraction containing the folded protein.
(C) SDS-PAGE electrophoresis of the fraction containing the folded protein. Lane 1 is the engineered TCR in non-reducing conditions. Lane 2 is the engineered TCR in reducing conditions.

added 5 M urea to the 1 L refolding buffer before adding our dissolved inclusion bodies and then dialyzed the 1 L refolding buffer against 10 L of deionized water. After 3 days of incubation, we concentrated our folding mixture and subjected it to size exclusion and anion exchange chromatography. The final product was concentrated to 10 mg/ml and the overall yield was 2-3 mg/L. Typical NLV/GIL-HLA-A2 Hitrap Q, S200 gel filtration, Mono Q anion exchange chromatography, and SDS-PAGE profiles are shown in Figure 2.2.

2.2.2 Production of NLV/GIL-specific TCRs

NLV/GIL specific TCRs were also produced via *in vitro* folding. All the TCRs contained the engineered artificial disulfide bond in the C domains. The overall yield of the TCRs ranged from 2-9 mg/L. The high quality of these recombinant molecules was demonstrated by SDS-PAGE. Briefly, 45 mg of α chain and 35 mg of β chain inclusion bodies were mixed prior to refolding. The 1 L refolding solution contained 5 M urea, 100 mM Tris (pH 8.0), 400 mM arginine-HCl, 2 mM EDTA, 5 mM reduced glutathione, and 0.5 mM oxidized glutathione. The refolding solution was further dialyzed into 10 L of deionized water for 3 days and concentrated for gel filtration and anion exchange chromatography purification. The final product was concentrated to 10 mg/ml and stored at -80 C°. Typical S200 gel filtration, MonoQ anion exchange chromatography, and SDS-PAGE profiles are shown in Figure 2.3.

2.3 Discussion

There are many approaches to producing large quantities of pMHC and TCR molecules from eukaryotic and prokaryotic systems. For crystallization purposes, we prefer to use prokaryotic systems for two reasons: 1. Both pMHC and TCR are glycoproteins and eukaryotic systems carry out glycosylation during protein expression. All glycosylation sites are distal from TCR-pMHC interface thus are unlikely to affect the biochemical and biophysical nature of TCR-pMHC interactions. Therefore, the heterogeneity of glycosylation may impede us from getting well-diffracting complex crystals. 2. Overall protein purity is very high after extensive inclusion bodies wash cycles. As result, no purification tags need to be added to the proteins and no subsequent protease digestion steps are required to remove these tags.

The initial low yield of peptide-HLA-A2 molecules produced via rapid dilution could be attributed to HLA-A2 heavy chain aggregation. Indeed, we observed white aggregates within 30 min after adding inclusion bodies, as well as large quantities of β_2-microgolobulin monomers during purification. This excess β_2-microgolobulin suggested that the majority of refolded β_2-microgolobulin molecules were not able to associate with the HLA-A2 heavy chain. In the improved protocol, we included 5 M urea in the 1 L refolding buffer to prevent rapid folding of the denatured protein. Dialysis against 10 L deionized water gradually reduced the high concentration of urea in the refolding mixture. We added 10 mM Tris (pH 8.0) every 12 hours to maintain the optimal pH, as well as to dilute excess of urea and L-arginine. After 3 days, we were not able to observe white protein aggregates, indicating that the efficiency of HLA-A2 heavy chain folding and β_2-microgolobulin association had greatly improved. MHC-I molecules have wide applications for studying immunological questions. Therefore, being able to produce high quality MHC-I molecules in large quantity is highly desirable. We believe our improved

protocol could potentially be applied to other pMHC-I complexes, as well as to non-classical MHC-I molecules.

TCR production exhibits a wide range of yields, ranging from 2-9 mg/L. The different yields are likely attributable to specific TCR αβ sequences. The length of CDR3s and the amino acid composition may affect the pairing of α chain and β chain. Several NLV/GIL-specific TCRs did not associate as αβ heterodimers, instead only ββ homodimers were purified even though we tried using different α:β molar ratios. Therefore, eukaryotic expression systems may be required to produce these problematic TCRs instead of *in vitro* folding. Nevertheless, *in vitro* folding combined with an artificial disulfide bond represents a robust way to produce large quantities of homogeneous TCR molecules. Although there are more than 30 TCR-pMHC-I structures in the PDB, the structural basis of public T cell response to dominant viral epitope was only revealed for EBV-specific responses (Rudolph et al., 2006; Rossjohn et al., 2015). For the more clinically relevant pathogens such as HCMV and influenza A, only one TCR-pMHC-I complex is currently available for each viral epitope. Thus, it is necessary to expand the structural database of public TCR-NLV/GIL-HLA-A2 complexes. Such structural information may help us refine vaccine design strategies for these viruses.

Chapter 3

NLV/GIL-specific TCRs exhibit wide ranges of binding affinities for NLV/GIL-HLA-A2

Summary

TCR-pMHC affinities are typically weak with K_Ds ranging from 1-200 μM. Anti-microbial TCRs expressed on CD8+ T cells generally bind to their cognate ligands at the higher end of this range. The relatively tight binding of anti-microbial TCR is likely attributable to the antigenic peptide. It is evident that the tighter binding facilitates better killing of infected cells and also promotes clonal expansion. To characterize interactions between TCR-NLV-HLA-A2 and TCR-GIL-HLA-A2, we performed steady state binding assays using surface plasmon resonance (SPR). We immobilized our biotinylated NLV/GIL-HLA-A2 ligand on streptavidin (SA) chips and flowed different concentrations of TCR sequentially. The results showed that NLV/GIL TCRs are able to bind their ligands with a wide range of affinities that generally corresponded well with T cell population sizes in peripheral blood.

3.1 Background

The initiation of TCR signaling is a key step that triggers T cell activation and proliferation (Brownlie and Zamoyska 2013). Early events in TCR signaling can differentiate between invading agonist and self-ligands with remarkable selectivity (Weiss 2014). The mechanism of the TCR's extraordinary sensitivity to extremely low concentration of antigenic peptide is still a mystery and many studies are going-on. Currently, there are several ways to directly quantify TCR-pMHC engagement, including: 1. Affinity calculations when TCR-pMHC interactions reach steady state equilibrium; 2. Using kinetic parameters (k_{on} and k_{off}) to derive association constants. Previous investigations showed the K_Ds of TCR-pMHC interactions to fall in the 1-200 μM range via the steady state equilibrium approach. Kinetic studies have revealed relatively short half-lives of TCR-pMHC interactions.

Whether the affinity or the dissociation rate is better correlated to T cell triggering is still a hot topic and there are lines of evidence supporting both parameters. On one hand, it has been demonstrated that T cell stimulation and magnitude of expansion are closely related to binding affinity (Lyons et al., 1996). On the other hand, kinetic studies demonstrated that TCR-pMHC affinity is mainly derived from k_{off} values (Kersh et al., 1998), suggesting that longer half-life may prolong TCR-pMHC interactions thus triggering downstream signaling events. In addition, studies have shown that TCR-pMHC complexes with similar affinities but disparate half-lives were able to differentiate TCR signaling. Whether TCR-pMHC complexes with similar half-lives but different affinities can trigger similar TCR signaling remains to be tested. Nevertheless, the affinity and half-life of a given TCR to its cognate pMHC are important criteria for characterizing TCR-pMHC interaction. Such information is particularly useful for guiding TCR-pMHC complex crystallization.

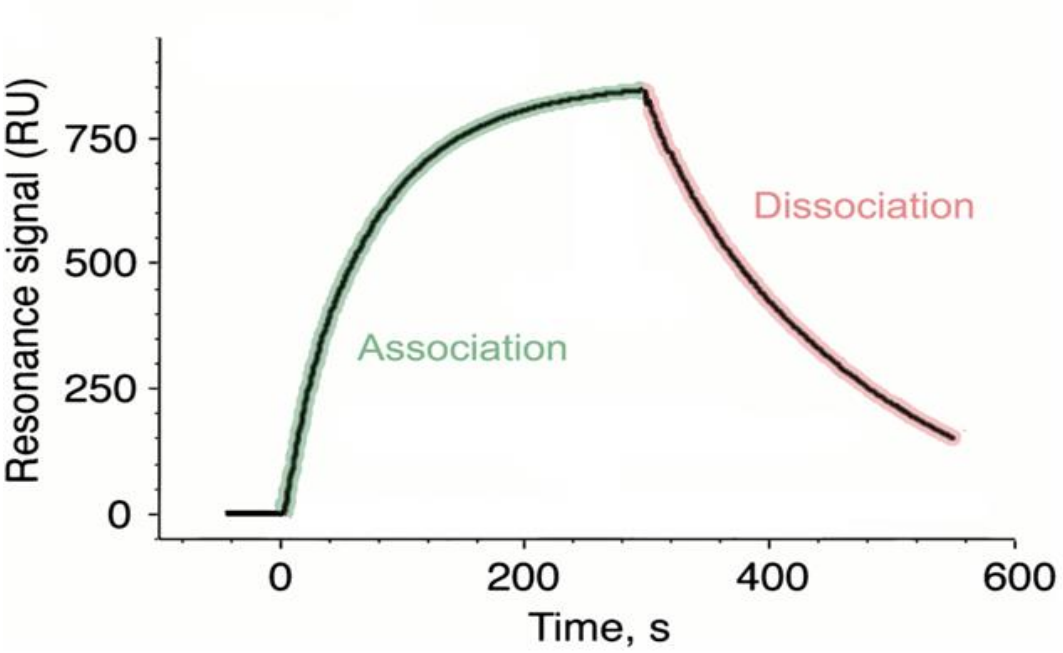


Figure 3.1 Example of an SPR sensorgram.

In this experiment, the injection of analyte was started at time 0. Association of the analyte to the ligand on the chip surface caused an increase in the resonance signal. Injection was stopped at 300 sec, which marked the start of the dissociation phase. The association curve is highlighted in green, and the dissociation curve is highlighted in red.

SPR, usually carried out with a Biacore instrument, is one of the most common techniques to characterize protein-protein interactions. There are several advantages in using SPR to studying TCR-pMHC interactions compared to other techniques such as isothermal titration calorimetry (ITC), analytical ultracentrifugation (AUC), and fluorescence stopped-flow. First, SPR can provide both kinetic and thermodynamic parameters, whereas ITC and AUC cannot obtain kinetic results. Second, SPR usually only requires nanogram amounts of protein to immobilize on a sensor chip and microgram amounts of protein to flow over the chip. This is

particularly suitable for analyzing TCR-pMHC interactions, whose K_Ds are usually in the 1-100 μM range. To characterize such low affinity interactions by ITC or AUC, 10 mg of proteins for each partner are usually required. Third, in order to characterize several TCRs at once, pMHC ligands are usually immobilized on the chip and TCRs with different concentrations are flowed over the chip. The chip surface can be regenerated after the runs and be fully capable of another experiment. In contrast, materials from ITC and AUC cannot be regenerated and always need to replenished for new experiments. This is very important for characterizing public TCRs-pMHC interactions since multiple TCRs all recognize the same pMHC antigen.

In a typical SPR experiment, one molecule (pMHC) is immobilized onto a sensor chip surface and another molecule (TCR) flows over the chip (Figure 3.1). During the association phase, TCR binding will increase the SPR signal, which is expressed as Response Units (RU). After a period of contact time, buffer will flow over the surface and decrease the SPR signal. This is called the dissociation phase. For a 1:1 binding reaction like TCR:pMHC, the association and dissociation phases can be described by equations. Here, the concentration of TCR-pMHC complex can be replaced by RU. The total concentration of pMHC is equivalent to the maximum concentration of TCR-pMHC and can be represented as RU_{max}. These two equations are used to represent the association and dissociation curves for Biacore sensorgram. The concentration of TCR is known and input by the operator. Thus, k_{on}, k_{off} and RU_{max} are fitted according to the equation (Eq 3.1).

Due to instrumental limitations, measureable k_{on} and k_{off} rates are within 10^3-10^8 $M^{-1}s^{-1}$ and 10^{-5}-10^{-1} s^{-1}, respectively (Cole et al., 2007). The available kinetic parameters for TCR-pMHC interactions feature fast k_{on} (10^4-10^5 $M^{-1}s^{-1}$) and k_{off} (10^{-2}-10^{-1} s^{-1}). In addition, these kinetic parameters were obtained for relatively tight TCR-pMHC binding (K_D =1-10 μM); for

weaker binding, these parameters were often unmeasurable. To screen large numbers of TCR-pMHC interactions, obtaining affinity values from equilibrium analysis is much more practical. By flowing serial concentrations of TCR over a sensor chip immobilized with pMHC ligand, SPR signals at the plateau are recorded. By using the equation, RU and TCR concentrations were fitted as two variables to obtain K_Ds.

$$RU = \frac{[TCR]}{K_D + [TCR]} RU_{max} \quad \text{(Eq 3.1)}$$

$$TCR + pMHC \rightleftharpoons TCR \cdot pMHC \quad \text{(Eq 3.2)}$$

$$K_D = k_{off}/k_{on} \quad \text{(Eq 3.3)}$$

$$Halflife\left(t_{1/2}\right) = ln2/k_{off} \quad \text{(Eq 3.4)}$$

$$\frac{dRU}{dt} = k_{on}[TCR](RU_{max} - RU) - k_{off}[TCR \cdot pMHC] \quad \text{(Eq 3.5)}$$

$$\frac{dRU}{dt} = -k_{off} RU \quad \text{(Eq 3.6)}$$

The relatively tighter binding of public TCRs to pMHC ligand is likely attributable to the greater expansion of public T cells. Thus, we are interested in characterizing affinities of our newly isolated public TCRs for comparison with previously reported public TCRs. TCR RA14, (specific for NLV-HLA-A2) and TCR JM22 (specific for GIL-HLA-A2) both bind their cognate ligand strongly. We hypothesized that our newly identified NLV-specific TCRs should have comparable affinities with RA14 and that less expanded GIL-specific TCRs should have weaker affinities compared to JM22.

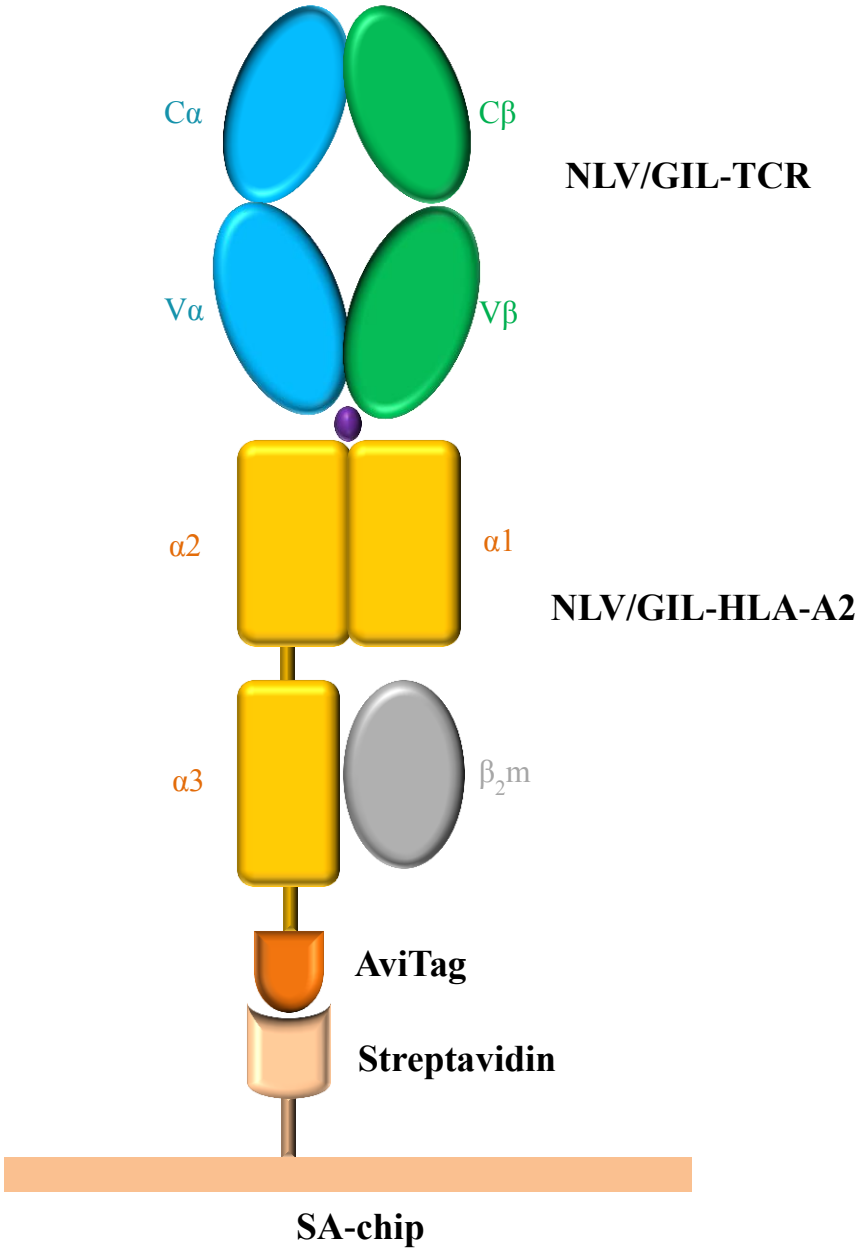


Figure 3.2 Experimental design for SPR analysis of the binding between NLV/GIL-specific TCRs and NLV/GIL-HLA-A2.

Biotinylated NLV/GIL-HLA-A2 was directly immobilized onto a streptavidin SA-chip. In this way, the NLV/GIL-HLA-A2 ligand was directionally immobilized, which effectively reduces signal noise caused by heterogeneous immobilization. NLV/GIL-specific TCRs are then flowed in different concentrations to generate SPR sensograms for equilibrium analysis.

3.2 Results

3.2.1 Immobilization of NLV/GIL-HLA-A2 onto a biosensor chip

To achieve directional immobilization of NLV/GIL-HLA-A2, we obtained biotinylated NLV/GIL-HLA-A2 monomers from the NIH Tetramer Core facility located at Emory University. The Avitag peptide sequence was fused to the C-terminus of the HLA-A2 heavy chain. The final biotinylated products were concentrated to ~2 mg/ml and shipped to our lab. An SA chip pre-immobilized with streptavidin was docked into a Biacore T100 instrument; 2-4 μg of pMHC were used to produce 1000-1500 RU of immobilized NLV/GIL-HLA-A2. Such moderate RU can maximize SPR signal upon TCR binding, as well as prevent steric hindrance due to large amounts of immobilized ligand (Figure 3.2).

3.2.2 Equilibrium measurements of six NLV-specific TCRs and six GIL-specific TCRs

A series of concentrations of TCRs were flowed over immobilized NLV-HLA-A2 for equilibrium analysis. Under equilibrium conditions, K_Ds were obtained and listed in figures. The NLV-specific TCRs C7, C25 and D12 bound NLV-HLA-A2 with K_Ds of 4.7±0.3 μM, 5.2±0.2 μM and 10±1 μM, respectively (Figure 3.3). These affinities are at the high end of the range for TCR-pMHC interactions, indicating that these TCRs are equally effective regardless their distinct V(D)J usage and α/β pairing (Cole et al., 2007; Rudolph et al., 2006; van der Merwe and Davis, 2003). TCRs C31, C32 and C34 showed no detectable binding even at 200 μM

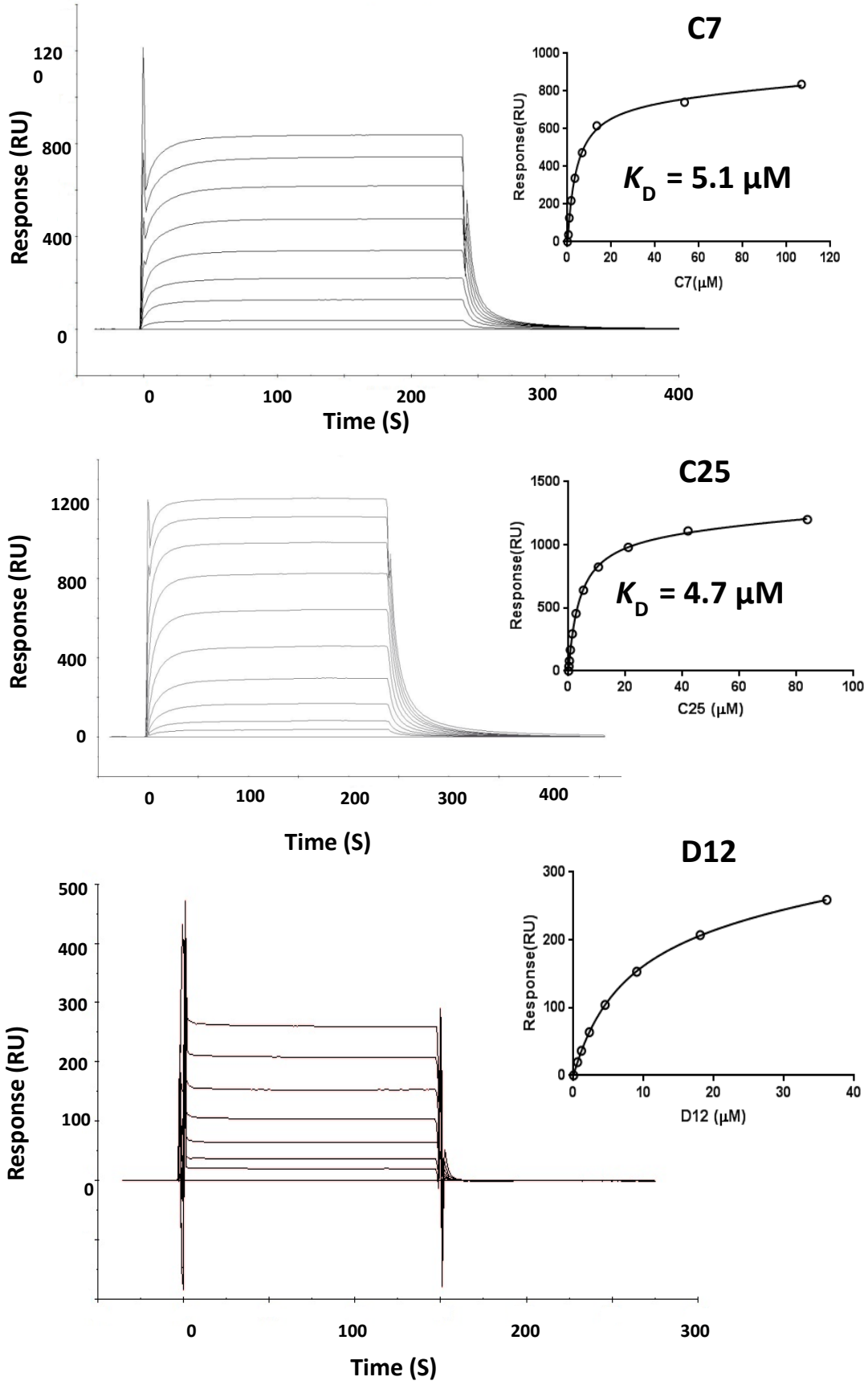
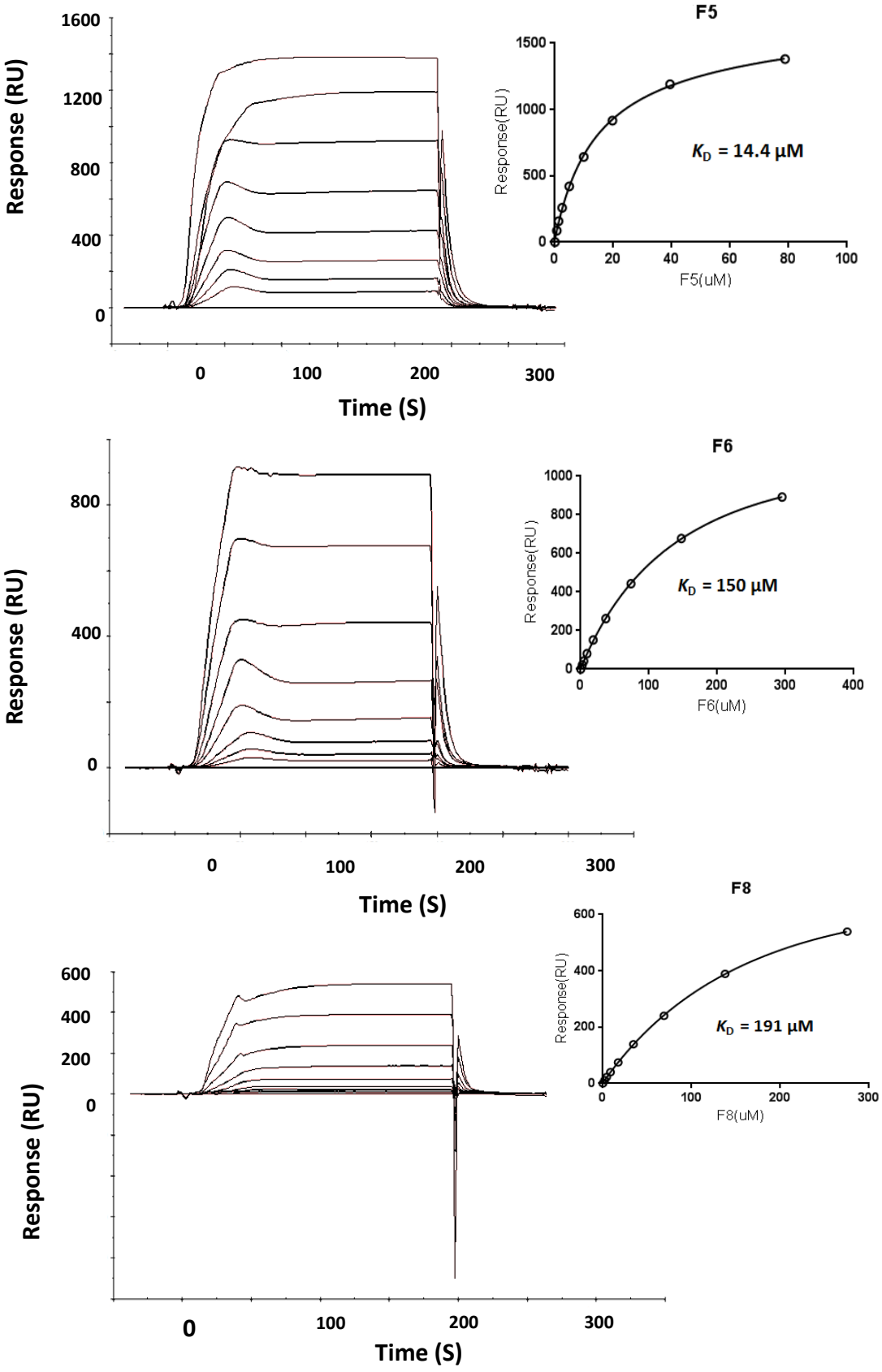


Figure 3.3 SPR analysis of TCR C7, C25 and D12 binding to NLV-HLA-A2.

concentration. Several reasons likely account for the lack of binding by these TCRs: 1. The TCRs were sorted by NLV/GIL-HLA-A2 tetramers. This avidity enhancement may promote non-specific binding of pMHC tetramers. 2. Sequencing errors could be introduced from RT PCR because reverse transcriptase fidelity is not as high as for DNA polymerase. 3. Obtaining correct TCR α/β pairing information from single T cells is extremely challenging. During cell lysis and mRNA extraction, contamination or RNA degradation may occur. In addition, one T cell may bear two α chains and lead to incorrect of TCR α/β pairings.

A similar protocol was applied to GIL-specific TCRs for equilibrium measurements. F5, F22 and F26 bound GIL-HLA-A2 with K_Ds of 15±1.5 μM, 3.2±0.2 μM and 1.9±0.2 μM, respectively. Notably, these TCRs all express identical or almost identical Vβ chains as well as identical CDR3α motifs. The strongly biased TCR usage and high affinities of these TCRs suggest that restricted Vβ usage is probably due to a limited number of structural solutions to recognizing GIL-HLA-A2. We also tested two non-TRBV19 GIL-specific TCRs (F8 and F50). Both exhibited much weaker binding to GIL-HLA-A2 (200±17 μM and 70±6 μM, respectively). Indeed, T cells bearing F8 or F50 TCRs are much less expanded. Finally, we tested a TCR (F6) which is almost identical to F5 (JM22) except the CDR3α and Jα usage. Interestingly, F6 exhibited weak binding (150±15 μM), suggesting that CDR3α may play a critical role in recognition of GIL-HLA-A2 (Figure 3.4).


F5
Response (RU)
1600
1200
800
400
0
0
100
200
300
Time (S)
1500
1000
500
0
Response(RU)
K_D = 14.4 µM
0
20
40
60
80
100
F5(uM)
F6
Response (RU)
800
400
0
0
100
200
300
Time (S)
1000
800
600
400
200
0
Response(RU)
K_D = 150 µM
0
100
200
300
400
F6(uM)
F8
Response (RU)
600
400
200
0
0
100
200
300
Time (S)
600
400
200
0
Response(RU)
K_D = 191 µM
0
100
200
300
F8(uM)

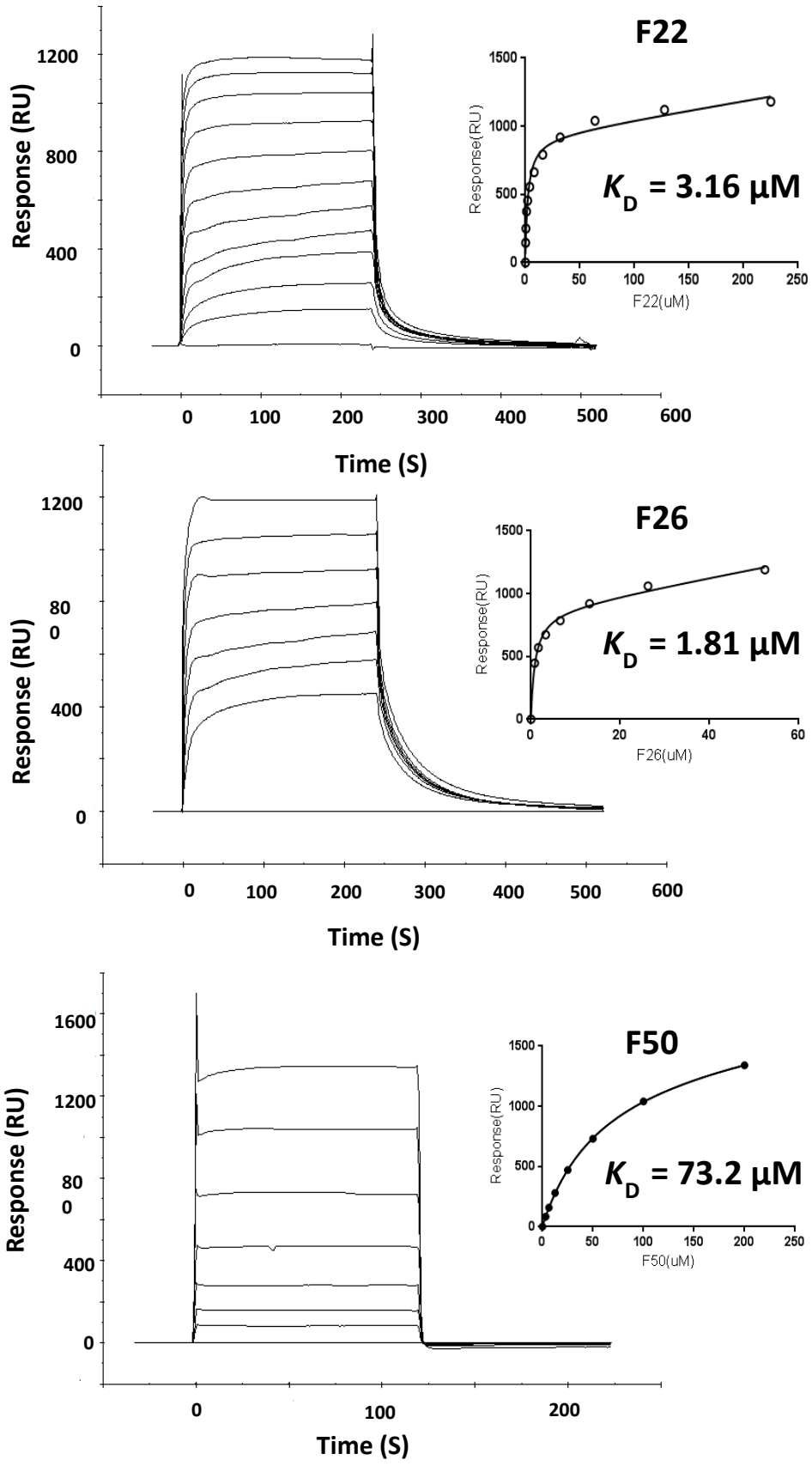


Figure 3.4 SPR analysis of TCR F5, F6, F8, F22, F26 and F50 binding to GIL-HLA-A2.

3.3 Discussion

Public T cell responses have long been a focus of immune repertoire studies. However, the exact mechanism by which public T cells are selected and expanded in different individuals is still elusive. A structural explanation suggests that public TCR are selected based on the 'shape' of peptide-MHC complexes. On one hand, 'bulged' peptide conformations have been observed; such protruding conformations allow limited TCR access (Tynan et al., 2005). Limited TCR access results in limited structural solutions, leading to biased public TCR expansion. On the other hand, a 'featureless' peptide also can restrict TCR access and result in few public TCRs being preferentially expanded (Turner et al., 2003 Stewart-Jones et al., 2003). In support this notion, different nucleotide sequences usually give rise to identical or nearly identical amino acid sequences in the public TCRs, suggesting that selection pressure more likely operates on a structural rather than genetic level.

The widely accepted definition of a successful public TCR structural solution is usually based on how well the TCR binds its cognate ligand. The relatively tight binding of sequence unrelated TCRs C7, C25 and D12 to NLV-HLA-A2 suggests that multiple structural solutions are equally effective. Sequence analysis of T cell repertoires also revealed the dominant α and β chain usage of C7, C25 and D12 in NLV-specific response. Thus, the high affinities of these TCR agree well with their dominant presence in human T cell repertoires. For TCRs C31, C32 and C34, undetectable SPR signals could be attributed to the technological challenge of sorting single T cells. Most single cell sorting studies reported to date are not coupled with biochemical or biophysical validation of TCR-pMHC interactions (Wang et al., 2013; Nguyen et al., 2014). Thus, our affinity experiments revealed the necessity to confirm actual TCR-pMHC complex

formation and should be included in future single T cell sorting experiments and antigen-specific TCR repertoire analyses.

GIL-specific TCRs feature highly restricted usage of TRBV19-TRBD2-TRBJ2-7, as well as a conserved CDR3 'RSS' motif. Previous structural study of JM22, a GIL-specific TCR, has revealed the dominant role of Vβ in engaging GIL-HLA-A2. F5, F22 and F26 all share the identical or nearly identical Vβ regions as JM22 and all bind to GIL-HLA-A2 strongly. This probably suggests that the TRBV19-TRBD2-TRBJ2-7 combination is the best structural solution our immune system can come up with because GIL-HLA-A2 lacks protruding peptide side chains for TCR binding. JM22 TCR employs Arg residues in the CDR3β RSS motif to occupy a notch between HLA-A2 and the GIL peptide. Mutation of this Arg into Ala completely abrogates the interaction between JM22 and GIL-HLA-A2 (Ishizuka et al., 2008). It is reasonable to expect TCRs F5, F22 and F26 to interact with GIL-HLA-A2 similarly to JM22. However, F5, F22 and F26 have totally different TRAV usage. This reflects the minimal contribution of CDR1α and CDR2α to interacting with GIL-HLA-A2 and allows GIL-specific TCR α chain diversity. The hypothesis that TRBV19-TRBD2-TRBJ2-7 may be the best solution for recognizing GIL-HLA-A2 is further supported by two non-TRBV19 TCRs binding to GIL-HLA-A2 with much weak affinity. In addition, the tight binding of TRBV19 GIL-specific TCRs, such as F6, may also be affected by the TCRα chain, particularly by CDR3α length. Notably, although F6 binds to GIL-HLA-A2 with much lower affinity than JM22, F6 TCRα (TRAV27-TRAJ37) usage in deep sequencing analyses reveals a dominant role in GIL-specific T cell repertoires. This discrepancy can be explained by the CDR3α of F6 being different from other TRAV27-TRAJ37 TCR CDR3α. Also, it is important to keep in mind that SPR, although a powerful technique to study protein-protein interactions, may not reflect physiologically relevant conditions for membrane

proteins. Indeed, it has been reported that TCRs with very different SPR affinities showed no differences in more physiological 2D affinity measurements (Huang et al., 2010). In the end, we selected TCRs that showed binding to their cognate pMHC ligands for crystallization trials.

Chapter 4

Structural studies of NLV-specific TCRs recognizing NLV-HLA-A2

Summary

Crystal structures of the C7-NLV-HLA-A2 and C25-NLV-HLA-A2 complexes were determined to resolutions of 2.1 Å and 3.5 Å, respectively. The completely distinct V(D)J combination and CDR3s of C7 and C25 lead to totally different structural solutions of NLV-HLA-A2 recognition by these TCRs. C25 sits on NLV-HLA-A2 with an orthogonal docking mode and peptide specificity is heavily biased on the center of NLV peptide. By contrast, C7 docks on NLV-HLA-A2 with a more diagonal mode and allows the TCR to interact with the entire NLV peptide. NLV P5 Met protrudes from the peptide-binding cleft of HLA-A2 and serves as a hot spot for recognition by both TCRs. Interestingly, although C25 and C7 CDR3s adopt different conformations to accommodate NLV P5 Met, both TCRs are equally effective in binding NLV-HLA-A2.

4.1 Background

As discussed in previous chapters, CMV is a ubiquitous and persistent human pathogen that is kept in check by CD8+ cytotoxic T lymphocytes. Individuals expressing the MHC class I molecule HLA-A2 produce cytotoxic T lymphocytes bearing TCRs that recognize the immunodominant CMV epitope NLVPMVATV (NLV). The NLV-specific T cell repertoire is characterized by a high prevalence of TCRs that are frequently observed in multiple unrelated individuals (Trautmann et al., 2005; Wang et al., 2013; Nguyen et al., 2014). These public TCRs feature identical, or nearly identical, CDR3α and/or CDR3β sequences. The TCRs may express public CDR3α motifs alone, public CDR3β motifs alone, or dual public CDR3αβ motifs. In addition, the same public CDR3α motif may pair with different CDR3β motifs (and the reverse), giving rise to highly diverse NLV-specific TCR repertoires.

Despite the great diversity of NLV-specific T cell repertoire, the structural database is still too limited to understand the molecular basis for such a diverse T cell response. A previous study of the NLV-specific RA14 TCR bound to NLV-HLA-A2 revealed important structural features that are distinct from other public T cell responses (Gras et al., 2009). For instance, RA14 docks over NLV-HLA-A2 with a crossing angle of 35°. This diagonal docking mode differs from that of other public TCRs which feature a more orthogonal docking mode. The diagonal docking position of RA14 allows the CDR1 and CDR2 loops to participate in peptide recognition and maximizes the peptide readout. P4 Pro, P5 Met and P8 Thr from NLV peptide are hot spots for RA14 binding. Mutating these residues, especially P5 Met, destabilized the RA14-NLV-HLA-A2 complex and rendered the peptide ineffective in activating RA14 T cells. The RA14-NLV-HLA-A2 structure also highlighted distinct RA14 germline-encoded residues interacting with HLA-A2. For instance, Asn29α and Tyr31α are exclusively expressed in

TRAV24 or TRAV21. In addition, RA14 TRBV6-5 features a hydrophobic CDR2β sequence that makes multiple van der Waals contacts with the ligand. Notably, CDR2 β residues Tyr48β and Asp56β are found in other NLV-specific β chains, suggesting a similar hydrophobic interaction pattern in these TCRs. RA14-NLV-HLA-A2 structural characteristics explained in part the dominance of RA14 in the response to NLV-HLA-A2. Thus, the relatively high affinity and maximum peptide readout of RA14 enables this TCR to recognize CMV-infected cells. However, RA14-NLV-HLA-A2 structure alone could not account for the diverse nature of the NLV-specific T cell response.

To investigate the structural underpinnings of NLV-specific T cell clonal diversity, we determined crystal structures of two public TCRs (C7 and C25) in complex with NLV-HLA-A2. These TCRs utilize completely different CDR3α and CDR3β motifs that, in addition, can associate with multiple variable α and variable β regions in NLV-specific T cell repertoires. Through detailed structural analysis and comparison of the C7-NLV-HLA-A2 and C25-NLV-HLA-A2 complexes, we found that these two TCRs exhibit divergent footprints on pMHC such that C25 is more focused on the central portion of the NLV peptide than is C7. These structures explain in part how the public CDR3α motif of C25 may associate with different variable α regions and how the public CDR3α motif of C7 may pair with different CDR3β motifs. This interchangeability of TCR V regions and CDR3 motifs permits multiple structural solutions to binding an identical peptide-MHC ligand and thereby the generation of a clonally diverse public T cell response to CMV (Yang et al., 2015).

4.2 Results

4.2.1 Crystallization of TCRs C7 and C25 bound to NLV-HLA-A2

The C7 and C25 expression constructs contain an artificial disulfide bond in the TCR C region. HLA-A2 heavy chain and β_2-microglobulin are wild-type constructs and the NLV peptide was synthesized by GenScript. Both TCRs and NLV-HLA-A2 were produced via *in vitro* folding from bacterial inclusion bodies. Initial crystallization screenings of the C7-NLV-HLA-A2 and C25-NLV-HLA-A2 complexes were carried out using Rigaku Wizard I, II, III and IV kits. Crystals of the C7-NLV-HLA-A2 complex grew in drops containing 30% PEG 400 (w/v), 0.1 M Tris-HCl (pH 8.0), and 0.2 M $MgCl_2$. The crystals were cryoprotected with 35% PEG 400 (w/v) solution prior to flash-cooling. Crystals of the C25-NLV-HLA-A2 complex grew initially in drops containing 20% PEG 3000 (w/v), 0.1 M Hepes (pH 7.5), and 0.2 M calcium acetate. However, these crystals diffracted poorly. Further improvement was carried out by screening with a matrix of PEG 3000 combined with 12 salts and 8 buffer systems. Eventually, crystals that diffracted to satisfactory resolution grew in 10-15% PEG 3000 (w/v), 0.1 M imidazole, and 0.2 M calcium acetate. The crystals were cryoprotected with 30% of glycerol (v/v) and flash-cooled. The D12-NLV-HLA-A2 complex has not yet produced well-diffracting crystals.

4.2.2 X-ray crystallographic data collection, structure determination and refinement

Diffraction datasets were collected at beamline 22ID of the Advanced Photon Source, Argonne National Laboratory with a MAR 300 CCD detector. The data were indexed, integrated, and scaled with the HKL2000 program. Data collection statistics are presented in Table 4.1 The C7-NLV-HLA-A2 complex dataset was processed to 3.5 Å resolution with R_{merge}= 20%. The

overall completeness value for the dataset is 99.5% (95.0% for the highest resolution shell), which facilitated molecular replacement (MR). There are an estimated five complexes in one asymmetric unit. However, we were never able to locate the fifth C7 TCR or NLV-HLA-A2 molecule. We thus concluded that only four C7-NLV-HLA-A2 complexes were in the asymmetric unit. The high solvent content is consistent with the relatively low resolution of the X-ray dataset. We also tried crystal dehydration at different time point but the diffraction quality was either not improved or the crystals deteriorated after a 12 h dehydration period.

To determine the C7-NLV-HLA-A2 complex structure via MR, a gliadin-specific TCR (PDB accession code: 4OZF) and NLV-HLA-A2 (PDB accession code: 3GSN) were used as initial search models. The CDRs of the TCR and the NLV peptide were removed. Three complex molecules were located first; the fourth was found according to non-crystallographic symmetry. Structural refinement was performed using rigid body and simulated annealing via the Phenix program. The model was further refinement by manual model building with Coot based on $2F_o$-F_c and F_o-F_c maps with the NLV peptide omitted in the initial refinement. The final R_{work} and R_{free} values for the C7-NLV-HLA-A2 complex are 27.0% and 35.5%, respectively.

The C25-NLV-HLA-A2 complex dataset was processed to 2.1 Å resolution with an overall R_{merge}= 8.4%. The resolution of C25-NLV-HLA-A2 complex is one of the highest among all the TCR-pMHC-I and TCR-pMHC-II structures reported to date. The overall completeness value for the dataset is 99.2% (97.1% for the highest-resolution shell). To determine the C25-NLV-HLA- A2 complex structure via MR, an EBV-specific TCR LC13 (PDB accession code: 1MI5) and NLV-HLA-A2 (PDB accession code: 3GSO) were used as initial search models. Two complex molecules were located in one asymmetric unit based on solvent content. Initially, one of the C25 TCR model was located in a different asymmetric unit.

TABLE 4.1
Data collection and structure refinement statistics

	C7–NLV–HLA-A2	C25–NLV–HLA-A2
Data Collection		
Space group	$C222_1$	$P2_12_12_1$
Cell dimensions		
a (Å)	151.8	84.3
b (Å)	366.6	124.9
c (Å)	152.0	193.8
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution range (Å)[a]	49.5–3.51 (3.64–3.51)	40.7–2.10 (2.17–2.10)
Unique reflections[a]	52,979 (4,991)	119,101 (11,500)
R_{merge} (%)[a,b]	20.3 (89.0)	8.4 (59.4)
Mean $I/\sigma(I)$[a]	10.7 (2.4)	22.6 (3.6)
Completeness (%)[a]	99.5 (95.0)	99.2 (97.1)
Refinement		
Resolution range (Å)[a]	49.5–3.51	40.7–2.10
R_{work}(%)/R_{free}(%)[a,c]	26.8 (30.9)/35.5 (37.4)	20.1 (26.2)/25.4 (29.6)
No. of protein atoms	25,304	13,250
No. of water molecules	54	814
Rms deviations from ideality		
Bond lengths (Å)	0.009	0.016
Bond angles (°)	1.59	1.15
Ramachandran statistics (%)		
Most favored	87.3	96.0
Allowed	12.2	3.9
Disallowed	0.5	0.1

[a]Values in parentheses are statistics for the highest resolution shell.
[b]$R_{merge} = \sum|I_j - \langle I \rangle|/\sum I_j$, where I_j is the intensity of an individual reflection and $\langle I \rangle$ is the average intensity of that reflection.
[c]$R_{work} = \sum||F_o| - |F_c||/\sum|F_o|$, where F_c is the calculated structure factor. R_{free} is as for R_{work} but calculated for a randomly selected 5.0% of reflections not included in the refinement.

A new search model was generated by deleting that TCR and placing a TCR opposite the located NLV-HLA-A2 molecule. Two C25-NLV-HLA-A2 complexes were then found by Phaser. Structure refinement was performed using rigid body and simulated annealing via the Phenix program. The model was further refinement by manual model building with Coot based on $2F_o$-F_c and F_o-F_c maps with NLV peptide omitted in the initial refinement. The final R_{work} and R_{free} values for the C7-NLV-HLA-A2 complex are 20.1% and 25.4%, respectively. Refinement statistics are summarized in Table 4.1. Stereochemical parameters were evaluated by PROCHECK.

4.2.3 Overview of the C25-NLV-HLA-A2 and C7-NLV-HLA-A2 complexes

Both the C25-NLV-HLA-A2 and C7-NLV-HLA-A2 complexes showed unambiguous electron density for two (C25-NLV-HLA-A2) (Figure 4.1-A and B) or four (C7-NLV-HLA-A2) (Figure 4.1-C and D) complex molecules in the asymmetric unit. The root mean square difference (r.m.s.d) in α-carbon positions for TCR VαVβ and MHC α1α2, including the peptide is 0.4 Å for two C25-NLV-HLA-A2 complex molecules. The r.m.s.d of four C7-NLV-HLA-A2 complex molecules ranges from 0.5-0.83 Å. Based on such close similarities, the following description applies to all C25-NLV-HLA-A2 or C7-NLV-HLA-A2 complex molecules in asymmetric units.

Both C25 and C7 sit symmetrically over NLV-HLA-A2 in a canonical diagonal orientation. The crossing angles of C25 TCR and C7 to pMHC are 61° and 29°, respectively.

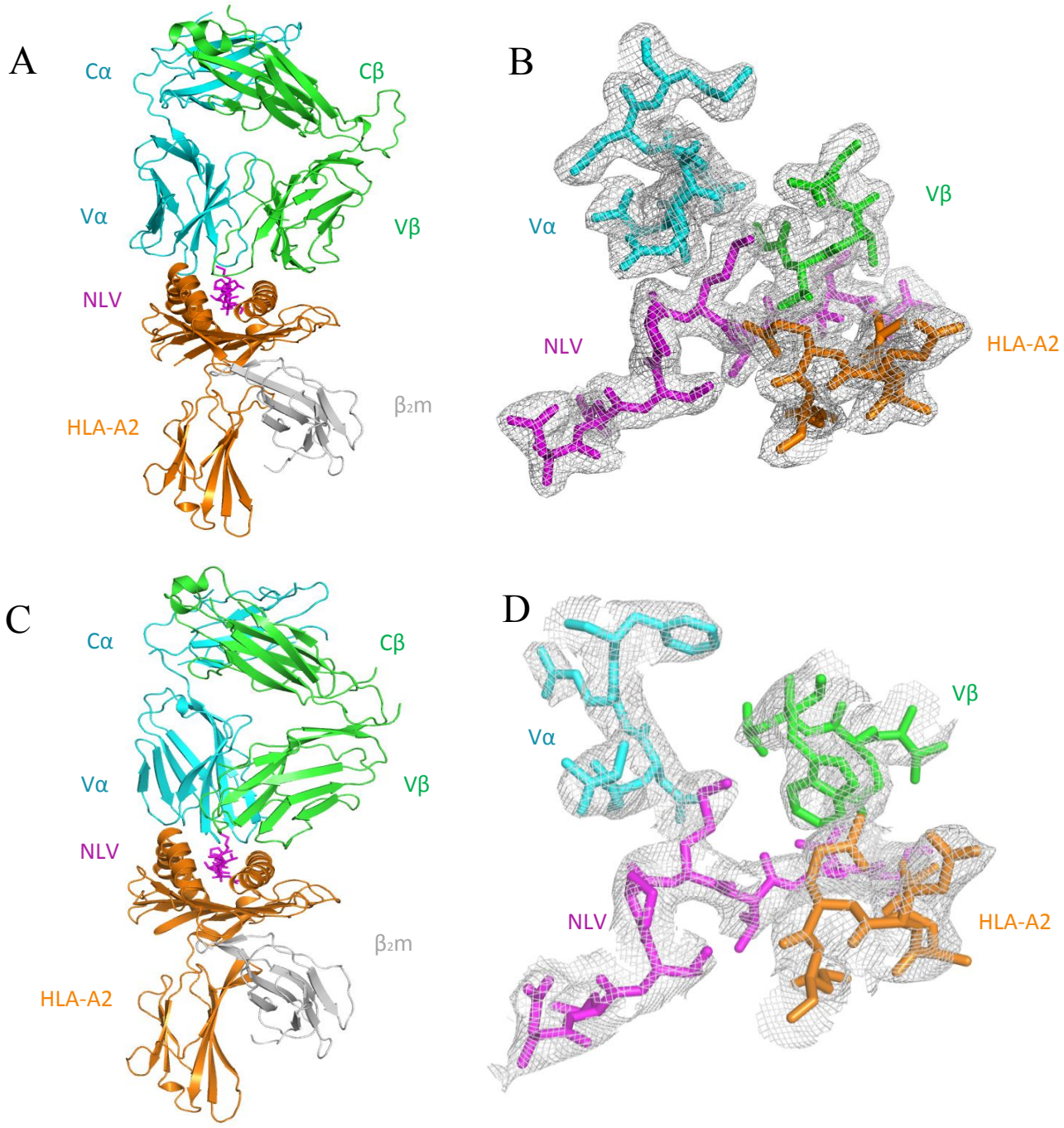


Figure 4.1. Structure of TCR–NLV–HLA-A2 complexes.

(A) Side view of the C25–NLV–HLA-A2 complex (ribbon diagram). TCR α chain, cyan; TCR β chain, green; HLA-A2 heavy chain, orange; β_2-microglobulin (β_2m), gray; NLV peptide, magenta.
(B) Electron density in the interface of the C25–NLV–HLA-A2 complex. Density from the final $2F_o - F_c$ map at 2.1 Å resolution is contoured at 1σ.
(C) Side view of the C7–NLV–HLA-A2 complex.
(D) Electron density in the interface of the C7–NLV–HLA-A2 complex. Density from the final $2F_o - F_c$ map at 3.5 Å resolution is contoured at 1σ.

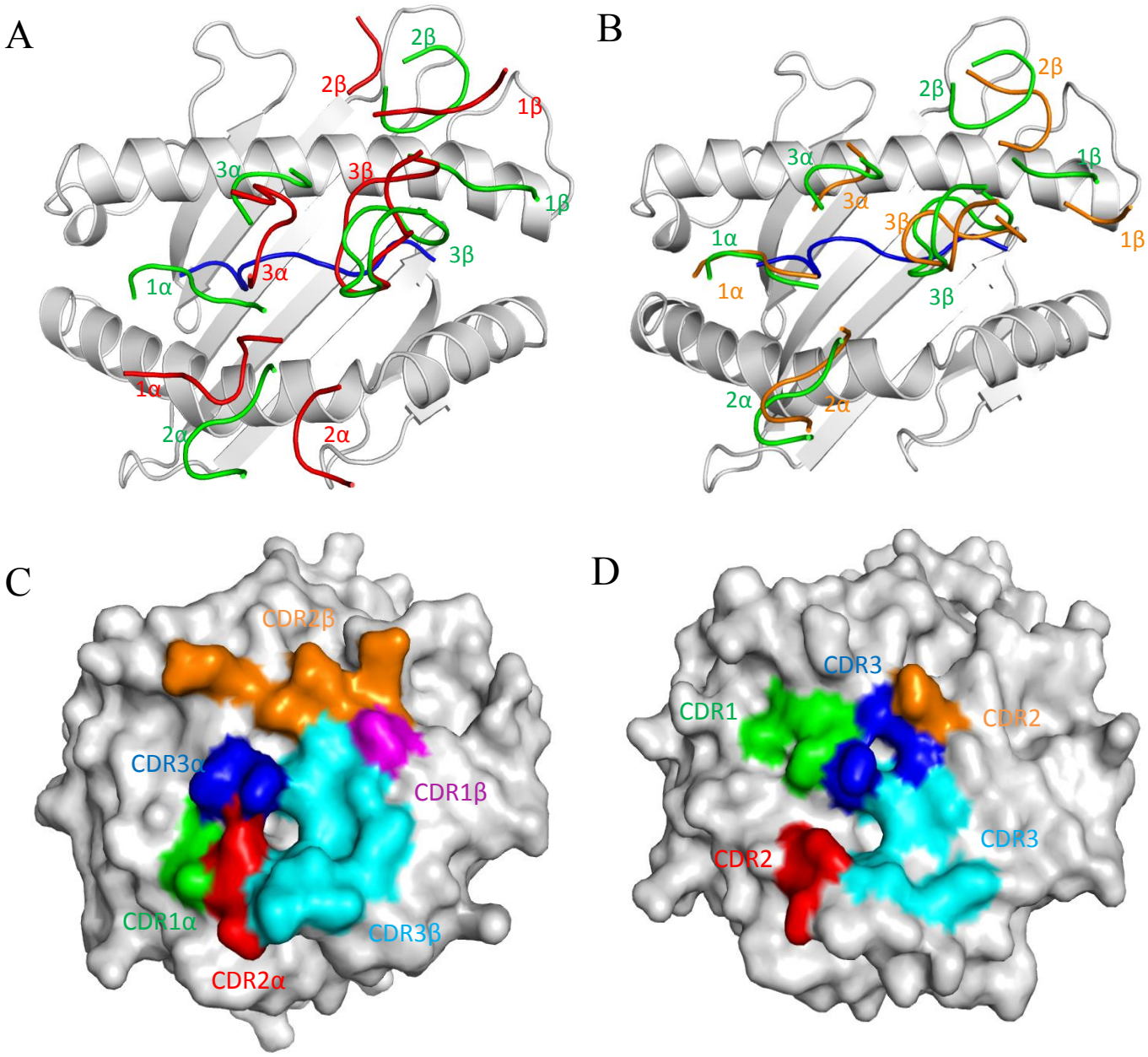


Figure 4.2 Comparison of TCR footprints on NLV–HLA-A2.

(A) Positions of CDR loops of TCRs C25 and RA14 (PDB accession code: 3GSN) (27) on NLV–HLA-A2 (top view). CDRs of C25 are shown as numbered red loops. CDRs of RA14 are green. HLA-A2 is gray. The NLV peptide is blue.
(B) Positions of CDR loops of TCRs C7 and RA14 on NLV–HLA-A2. CRDs of C7 are orange. CDRs of RA14 are green.
(C) Footprint of TCR C25 on NLV–HLA-A2. The top of the MHC molecule is depicted as a gray surface. The areas contacted by individual CDR loops are color-coded: CDR1α, green; CDR2α, red; CDR3α, blue; CDR1β, magenta; CDR2β, orange; CDR3β, cyan.
(D) Footprint of TCR C7 on NLV–HLA-A2.

Upon binding NLV-HLA-A2, C25 and C7 bury 89% (272 Å^2) and 86% (314 Å^2) of the peptide solvent-accessible surface, respectively. These percentages are at the higher end of the range for TCR-pMHC-I complexes, which varies from 60 to 91% in other structures. Extensive peptide burial enables C25 and C7 to maximize readout of the NLV peptide. Such extensive peptide burial is also a salient feature of the RA14-NLV-HLA-A2 complex. However, C25 and C7 recognize NLV-HLA-A2 in distinct fashion.

The footprints of C25 and C7 on the pMHC surface showed that both TCRs contact the N-terminal half of the NLV peptide mainly via the CDR1α and CDR3α loops, whereas the CDR3β loop mainly contacts the C-terminal half (Figure 4.2-A and B). The germline-encoded residues of C25 CDR1α and CDR2α mainly interact with HLA-A2 α2 helix, whereas all three CDRs from Vβ interact with the HLA-A2 α1 helix. Most contacts between Vβ and the HLA-A2 α1 helix are mediated by CDR2β (57 out of 74 total) (Figure 4.2 C). In contrast to C25, C7 interacts with HLA-A2 in a more Vα-dominant fashion, such that Vα CDRs mediate 74% of interactions with HLA-A2 (Figure 4.2 D).

4.2.4 Interaction of TCR C25 with HLA-A2

The total buried solvent-accessible surface for the C25-NLV-HLA-A2 complex is 1857 Å^2, comparable to that in other TCR-pMHC complexes. Vβ buries considerably greater surface area (516 Å^2, 60%) than Vα (333 Å^2, 40%). Such dominance by Vβ is unusual among all the TCR-pMHC-I complex structures reported to date, in which Vα and Vβ typically bury equal amount of surface area. Indeed, only three other TCR-pMHC-I complexes displaying a

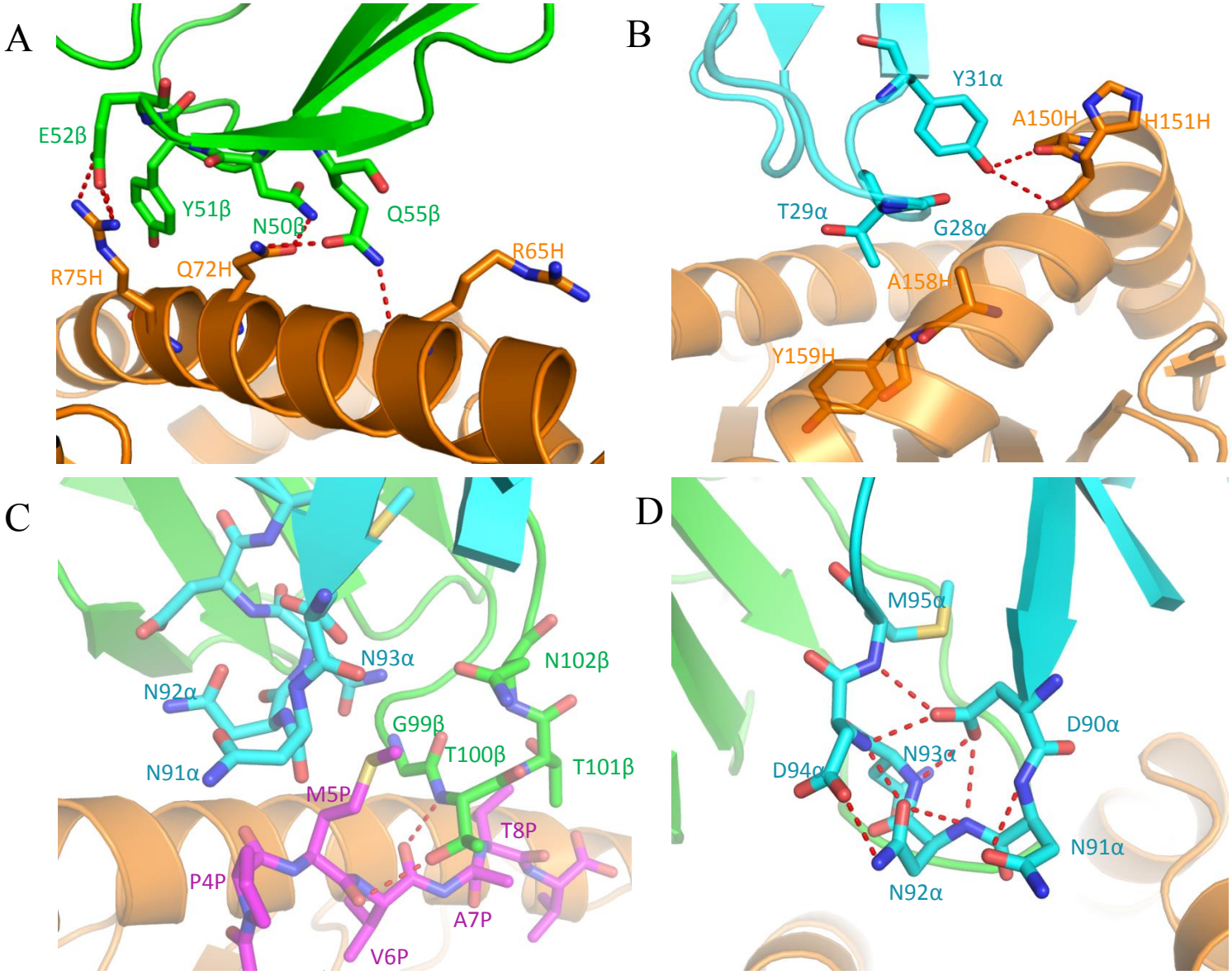


Figure 4.3. Interactions of TCR C25 with HLA-A2 and the NLV peptide.

(A) Interactions between CDR2β (green) of C25 and the HLA-A2 α1 helix (orange). The side chains of contacting residues are drawn in stick representation with carbon atoms in green (CDR2β) or orange (HLA-A2), nitrogen atoms in blue, and oxygen atoms in red. Hydrogen bonds are indicated by red dashed lines.

(B) Interactions between CDR1α (cyan) of C25 and the HLA-A2 α2 helix (orange).

(C) Interactions between C25 and the NLV peptide (magenta). Peptide residues are identified by a one-letter amino acid designation followed by position (P) number. CDR3α (cyan) and CDR3β (green) form a pocket that accommodates the side chain of P5 Met. The sulfur atom of P5 Met is yellow.

(D) Conformational stabilization of CDR3α of C25 by a dense network of eight intraloop hydrogen bonds.

similar degree of Vβ dominance as C25-NLV-HLA-A2 have been reported, involving the HLA-A2 restricted TCR JM22 (67%), the H-2K^b restricted TCR BM3.3 (63%), and the HLA-E-restricted TCR KK50.4 (61%). Overall, 69 van der Waals contacts are made by Vβ to HLA-A2 and these contacts were further reinforced by 5 hydrogen bonds. Consistent with smaller solvent-accessible surface burial, only 15 van der Waals contacts and 3 hydrogen bonds are made by Vα. All these contacts are mediated by 8 Vβ and 4 Vα residues and involve 15 MHC residues, of which 10 are contacted by RA14 and 7 by C7.

Excluding the NLV peptide, CDR1α, CDR2α and CDR3α contribute 18%, 13% and 6% total buried surface on HLA-A2, compared with 1%, 37% and 28%, respectively for CDR1β, CDR2β and CDR3β. Hence, CDR2β of TCR C25 contributes more to the binding interface with MHC than any other CDRs. The extraordinarily large contribution of CDR2β to the C25-HLA-A2 interface (37%) is highlighted by a comparison with 34 other TCR-pMHC-I structures, in which CDR2β accounts for only 12% of the buried surface. Residues Asn50β, Glu52β, and Gln55β of CDR2β form a dense network of five side-chain-side-chain hydrogen bonds linking C25 to residues Arg65, Gln72 and Arg75 on the HLA-A2 α1 helix. These hydrogen bonds are further reinforced by 56 van der Waals contacts that stabilize the interaction of CDR2β with the α1 helix (Figure 4.2- A and Table 4.2).

Notably, the HLA-B-restricted EBV-specific TCR LC13 expresses nearly the same Vα/Vβ gene pair (TRAV26-2/TRBV7-8) as C25 (TRAV26-2/TRBV7-6), resulting in the same CDR1α, CDR2α, and CDR1β, and a very similar CDR2β. In addition, LC13 mediates similar germline-encoded interactions with MHC as C25. This is in agreement with the hypothesis that the canonical diagonal docking orientation of TCR on MHC observed in TCR-pMHC complexes is at least partly the result of co-evolution of TCR and MHC molecules. Tyr31α, Gln50β and

TABLE 4.2
Interactions between TCR and MHC molecules in the C25–NLV–HLA-A2, RA14–NLV–HLA-A2 complexes

	Hydrogen bond	van der Waals contacts	Hydrogen bond	Van der Waals contacts
HLA-A2	C25	C25	RA14	RA14
E63H				
R65H	R65H(O) Q55β($N^{\varepsilon 2}$)	Q55β		T94α
K66H				N29α
A69H		Q55β	A69H(O) N96α($N^{\delta 2}$)	G95α N96α
Q72H	Q72H($O^{\varepsilon 1}$) N50β($N^{\delta 2}$) Q72H($N^{\varepsilon 2}$) Q55β($O^{\varepsilon 1}$)	N50β Q55β	Q72H($O^{\varepsilon 1}$) N96α($N^{\delta 2}$) Q72H($N^{\varepsilon 2}$) Y48β(O^{η}) Q72H($N^{\varepsilon 2}$) D56β($O^{\varepsilon 1}$)	Y48β V50β
T73H		P98β		
R75H	R75H($N^{\eta 1}$) E52β($O^{\varepsilon 1}$) R75H($N^{\eta 2}$) E52β($O^{\varepsilon 2}$)	Y51β E52β		I54β
V76H		V30β		V50β I54β
K146H		T101β	K146H(N^{ζ}) E30β($O^{\varepsilon 2}$)	V96β
W147H		T101β		
A149H			A149H(O) Y101β(O^{η})	Y101β
A150H	A150H(O) Y31α(OH)	Y31α T100β T101β		Y101β
H151H	H151H(O) Y31α(OH)	Y31α		
V152H		T100β		
E154H	E154H($O^{\varepsilon 1}$) T51α($O^{\gamma 1}$)	L50α		
Q155H		L50α T100β		Y31α T51α I100β
A158H		G28α		L52α
Y159H		T29α		

Glu52β of LC13 make hydrogen bonds with Arg 151, Gln 72 and Arg 75, respectively, of HLA-B. Structurally equivalent hydrogen bonds in the C25-NLV-HLA-A2 complex are: C25 Tyr31α OH-O His151H HLA-A2, C25 Asn50β Nδ2-Oε1 Gln72H HLA-A2, and C25 Glu52β Oε2-Nη1 Arg75H HLA-A2. However, LC13 and C25 have unrelated CDR3 sequences, which explains their different specificities.

C25 contacts the HLA-A2 α2 helix through CDR1α and CDR2α. In particular, the side chain of CDR1α Tyr31 binds to a site formed by HLA-A2 α2 residues Ala150H and His151H, in a manner resembling that observed for other MHC-I-restricted TCRs bearing a CDR1α Tyr/Phe31 motif (Figure 4.3-B and Table 4.2). The CDR3 loops of C25 do not engage MHC, except for some minor contacts involving CDR3β. Thus, MHC recognition by C25 is almost exclusively germline-encoded.

4.2.5 Peptide recognition by TCR C25

Except for a few contacts between CDR1α Thr30 and P4 Pro of NLV peptide, all interactions between C25 and the NLV peptide are mediated by the somatically generated CDR3 loops, with CDR3α and CDR3β accounting for 16 and 29 contacts, respectively. Peptide specificity is conferred mainly by shape complementarity, since the C25-NLV interface includes only two hydrogen bonds: C25 Thr100β Oγ1-O P5 Met and C25 Thr100β N-O P6 Val. Nearly all solvent-exposed NLV residues (P4 Pro, P5 Met, P6 Val, P7 Ala, P8 Thr) are engaged by C25 with the primary focus on P5 Met (Table 4.3). The CDR3 loops of C25 accommodate P5 Met by forming a hydrophobic pocket and account for 50% of all contacts with the NLV peptide (Figure 4.3-C and Figure 4.4-A). The conformation of CDR3α is locked by eight main-chain-side-chain

hydrogen bonds within the Asp 91-Asn92-Asn 93-Asn 94-Asp 95-Met 96 (DNNNDM) motif at the tip of this loop, suggesting a very restricted CDR3α conformation (Figure 4.3-D). Nonpolar interactions dominate the C25-NLV interface and only one van der Waals contact is made with P8 Thr.

TABLE 4.3
Interactions between TCR and NLV peptide in the C25–NLV–HLA-A2, RA14–NLV–HLA-A2 complexes

	Hydrogen bond	Van der Waals contacts	Hydrogen bond	Van der Waals contacts
NLV	C25	C25	RA14	RA14
V3P				Y31α
P4P		T30α N91α N92α		N29α F30α Y31α
M5P	M5P(O) T100β($O^{\gamma 1}$)	N91α N92α N93α G99β T100β N102β	M5P(S^{δ}) N96α(N)	Y31α G95α G98β G99β I100β
V6P	V6P(O) T100β(N)	G99β T100β		G98β
A7P		T100β T101β		T97β G98β G99β
T8P		T101β	T8P(N) T97β(O) T8P($O^{\gamma 1}$) E30β($O^{\varepsilon 1}$) T8P($O^{\gamma 1}$) T97β(O) T8P($O^{\gamma 1}$) T97β(N)	E30β T97β

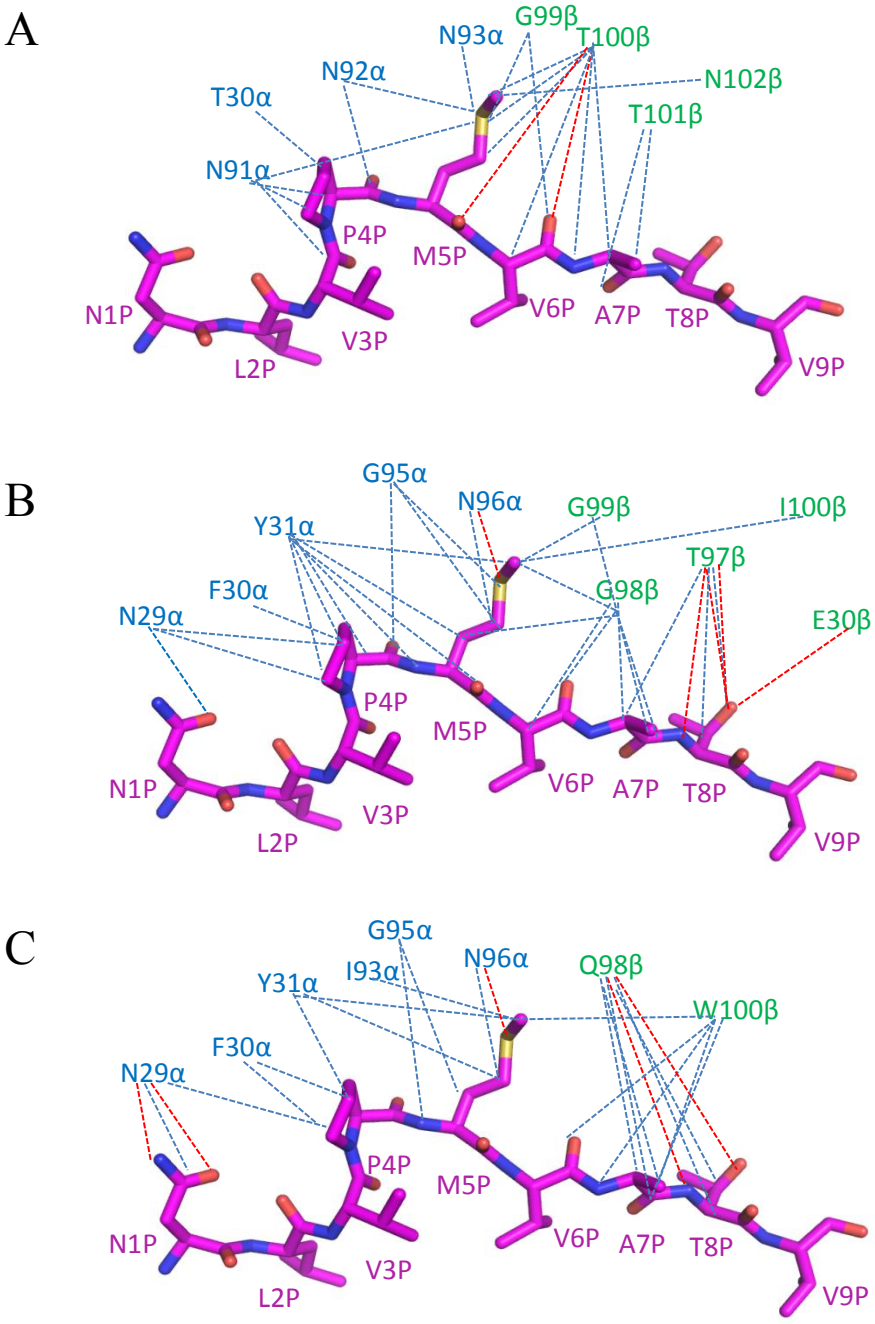


Figure 4.4 Comparison of interactions between TCRs and the NLV peptide.

(A) Interactions between TCR C25 and NLV. Hydrogen bonds are red dotted lines; van der Waals contacts are blue dotted lines.
(B) Interactions between TCR RA14 (27) and NLV.
(C) Interactions between TCR C7 and NLV.

4.2.6 Interaction of TCR C7 with HLA-A2

The C7-NLV-HLA-A2 complex buries a total solvent-accessible surface of 2103 Å^2, significantly more than the C25-NLV-HLA-A2 complex (1857 Å^2). Unlike the C25-NLV-HLA-A2 complex, in which Vβ dominates NLV-HLA-A2 surface burial, Vα (559 Å^2, 55%) of C7 contributes more than Vβ (463 Å^2, 45%). C7 expresses the same Vα region as RA14 (TRAV24) and has a nearly identical CDR3α sequence, ITGNQF, compared with NTGNQF for RA14, a public CDR3α motif. However, these two TCRs utilize unrelated Vβ regions (TRBV7-2 for C7; TRBV6-5 for RA14) and CDR3β sequences (SQTQLWETQ for C7; SPVTGGIYGY for RA14). Because the CDR3β sequences of C7 has not been identified as a public CDR3β motif, C7 belongs to the category of NLV-specific TCRs, comprising 34% of the total repertoire characterized so far, that uses either CDR3α or CDR3β public motifs, but not both. Because of the same Vα region usage, the overall docking mode of C7-NLV-HLA-A2 and RA14-NLV-HLA-A2 is quite similar, with crossing angles of TCR to pMHC of 29° and 39°, respectively. However, the detailed interactions with HLA-A2 made by C7 are considerably different from RA14, even for the shared Vα chain.

All six C7 TCR CDRs except CDR1β are involved in interacting with HLA-A2. Similar to RA14, C7 utilizes CDR1α, CDR3α and CDR2β to recognize the HLA-A2 α1 helix. Vα contribute many more contacts than Vβ, as well as three out of four hydrogen bonds: C7 Asn29α Nδ2-Oδ2 Glu63H HLA-A2, C7 Asn29α Oδ1-Nζ Lys 66H HLA-A2, and C7 Asn96 Nδ2-Oε1 Gln72H HLA-A2 (Figure 4.5 A and Table 4.4). Although the first two of these hydrogen bonds are not present in the RA14-NLV-HLA-A2 structure, both C7 and RA14 utilize CDR3α Gly95 and Asn96 from the XTGNQF motif to interact extensively with the HLA-A2 α1 helix.

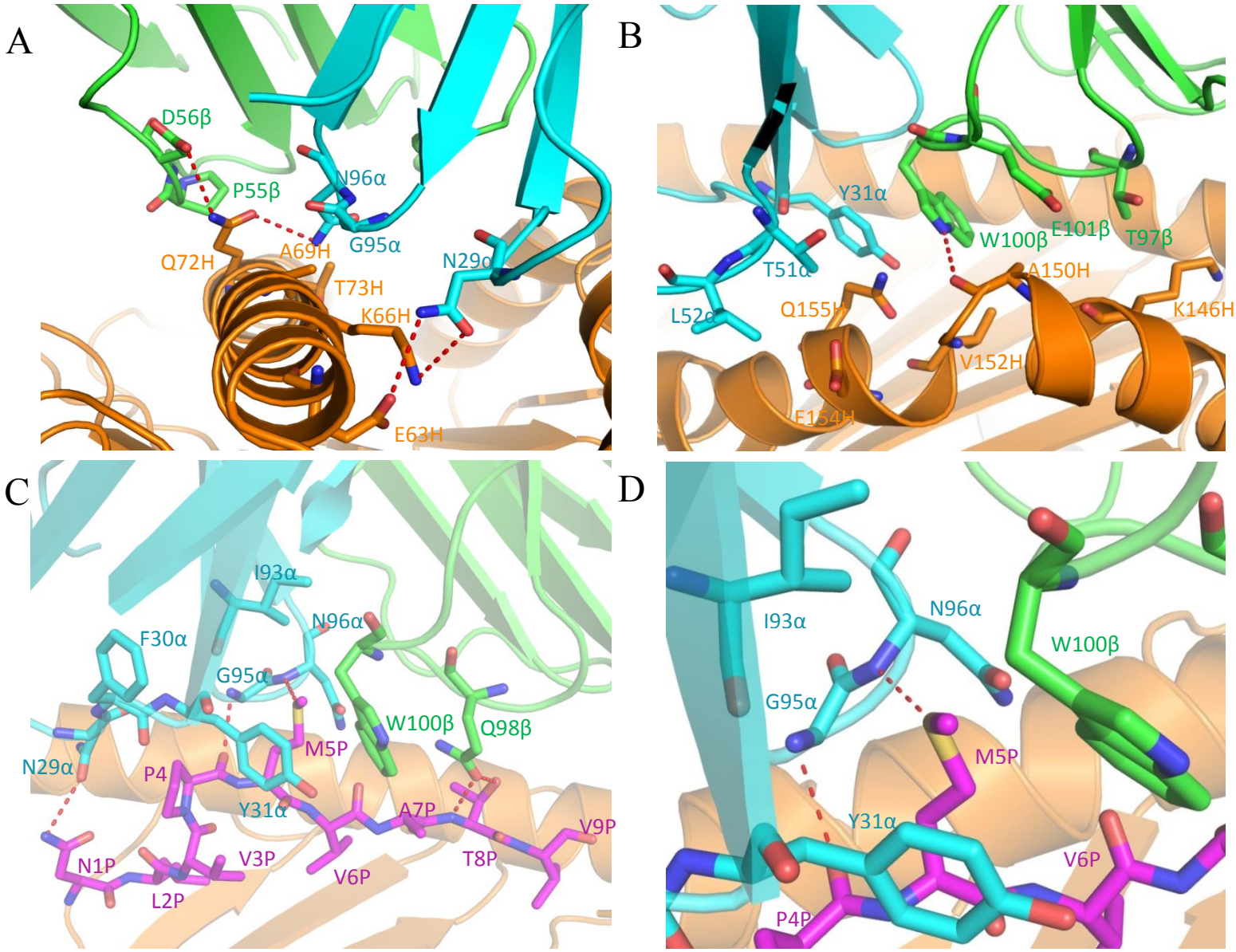


Figure 4.5 Interactions of TCR C7 with HLA-A2 and the NLV peptide.

(A) Interactions of CDR1α, CDR3α and CDR2β with the HLA-A2 α1 helix. The side chains of contacting residues are drawn in stick representation with carbon atoms in cyan (CDR1α and CDR3α), green (CDR2β) or orange (HLA-A2), nitrogen atoms in blue, and oxygen atoms in red. Hydrogen bonds are indicated by red dashed lines.

(B) Interactions of CDR1α, CDR2α and CDR3β with the HLA-A2 α2 helix. The side chains of contacting residues are drawn with carbon atoms in cyan (CDR1α and CDR2α), green (CDR3β) or orange (HLA-A2).

(C) Interactions of CDR1α, CDR3α and CDR3β with the NLV peptide. The side chains of contacting residues are drawn with carbon atoms in cyan (CDR1α and CDR3α), green (CDR3β) or magenta (NLV).

(D) Close-up of interactions between C7 and P5 Met.

TABLE 4.4
Interactions between TCR and MHC molecules in the C7-NLV-HLA-A2, RA14–NLV–HLA-A2 complexes

	Hydrogen bond	Van der Waals contacts	Hydrogen bond	Van der Waals contacts
HLA-A2	RA14	RA14	C7	C7
E63H			E63H($O^{\delta2}$) N29α($N^{\delta2}$)	
R65H		T94α		
K66H		N29α	K66H(N^{ζ}) N29α($O^{\delta1}$)	N29α
A69H	A69H(O) N96α($N^{\delta2}$)	G95α N96α		G95α N96α
Q72H	Q72H($O^{\varepsilon1}$) N96α($N^{\delta2}$) Q72H($N^{\varepsilon2}$) Y48β(O^{η}) Q72H($N^{\varepsilon2}$) D56β($O^{\varepsilon1}$)	Y48β V50β	Q72H($O^{\varepsilon1}$) N96α($N^{\delta2}$) Q72H($N^{\varepsilon2}$) D56β($O^{\varepsilon1}$)	N96α P55β
T73H				N96α
R75H		I54β		
V76H		V50β I54β		
K146H	K146H(N^{ζ}) E30β($O^{\varepsilon2}$)	V96β		T97β
W147H				
A149H	A149H(O) Y101β(O^{η})	Y101β		
A150H		Y101β	A150H(O) W100β($N^{\varepsilon1}$)	E101β
H151H				
V152H				W100β
E154H				L52α
Q155H		Y31α T51α I100β		Y31α T51α W100β
A158H		L52α		

(Figure 4.5-B and Table 4.4). However, the specific interactions made by these two residues in the C7-NLV-HLA-A2 and RA14-NLV-HLA-A2 complexes are considerably different due to the different CDR3α conformations. C7 interacts with the HLA-A2 α2 helix through CDR1α, CDR2α and CDR3β. A side-chain-main-chain hydrogen bond (C7 Trp100β Nε1-O Ala150H HLA-A2), not present in the RA14-NLV-HLA-A2 complex, provides additional stabilization.

4.2.7 Peptide recognition by TCR C7

TCR C7 mediates peptide recognition through CDR1α, CDR3α, and CDR3β via five hydrogen bonds. Similar to RA14, C7 engages nearly all solvent-exposed NLV residues (P1 Asn, P4 Pro, P5 Met, P6 Val, P7 Ala, P8 Thr), thereby burying 331 $Å^2$ of surface at the C7-NLV interface and enabling maximum readout of the peptide sequence. In contrast to C25, C7 engages the NLV peptide extensively at both the N- and C-termini, especially P8 Thr. P4 Pro is wedged between the side chains of CDR1α Asn29 and Tyr31, with which it establishes multiple hydrophobic contacts. The side chain of P5 Met alone accounts for 36% of all contacts with C7, mainly through CDR1α and CDR3α. In addition to extensive hydrophobic interactions with CDR1α Tyr31, CDR3α Asn96, and CDR3β Trp100, P5 Met forms a hydrogen bond through its sulfur atom with the main-chain nitrogen of CDR3α Asn96 (Figure 4.5-C and D). This hydrogen bond is also visible in the RA14-NLV-HLA-A2 complex. Four additional hydrogen bonds further stabilize C7-NLV interaction: C7 Asn29α Oδ1-Nδ2 P1 Asn, C7 Gly95α N-O P4 Pro, C7 Gln98β Oε1-N P8 Thr, and C7 Gln98β Oε1-Oγ1 P8 Thr. Thus, P5 Met appears to be the most critical peptide residue for TCR recognition, with P1 Asn, P4 Pro and P8 Thr also serving as interaction hot spots (Figure 4.4 and Table 4.3).

TABLE 4.5
Interactions between TCR and NLV peptide in the C7-NLV-HLA-A2, RA14–NLV–HLA-A2 complexes

	Hydrogen bond	Van der Waals contact	Hydrogen bond	Van der Waals contact
NLV	RA14	RA14	C7	C7
N1P			N1P($N^{\delta 2}$) N29α($O^{\delta 1}$)	N29α
V3P		Y31α		
P4P		N29α F30α Y31α	P4P(O) G95α(N)	N29α F30α Y31α G95α
M5P	M5P(S^{δ}) N96α(N)	Y31α G95α G98β G99β I100β	M5P(S^{δ}) N96α(N)	Y31α I93α G95α N96α W100β
V6P		G98β		W100β
A7P		T97β G98β G99β		Q98β W100β
T8P	T8P(N) T97β(O) T8P($O^{\gamma 1}$) E30β($O^{\varepsilon 1}$) T8P($O^{\gamma 1}$) T97β(O) T8P($O^{\gamma 1}$) T97β(N)	E30β T97β	T8P(N) Q98β($O^{\varepsilon 1}$) T8P($O^{\gamma 1}$) Q98β($O^{\varepsilon 1}$)	Q98β

4.2.8 Influence of CDR3β on CDR3α loop conformation in TCR C7

Since TCRs C7 and RA14 use almost identical Vα regions to bind NLV-HLA-A2, the CDRα conformations are expected to be the same, or at least very similar. Indeed, CDR1α and CDR2α display nearly identical conformations in the two complexes: r.m.s.d. in α-carbon positions of 1.0 Å and 1.3 Å for residues SSNFY of CDR1α and TLNGD of CDR2α, respectively. However, the CDR3α conformations of C7 and RA14 are substantially different, with an r.m.s.d. in α-cabon positions of 2.3 Å for residues TGNQ. As a consequence, CDR3α engages pMHC through a somewhat different set of contacts in the C7-NLV-HLA-A2 and RA14-NLV-HLA-A2 complexes.

The different conformations of CDR3α observed in C7 and RA14 are induced by the CDR3β loops of these TCRs, which differ in both sequence and length (SQTQLWETQ for C7; SPVTGGIYGY for RA14). These structural differences in CDR3β are transmitted to CDR3α via interactions between these loops in the TCR binding site. In RA14, the tip of CDR3β points toward CDR3α, with which it makes several van der Waals contacts and a main-chain-side-chain hydrogen bond (CDR3β Gly98 N-Oδ1 Asn96 CDR3α) (Figure 4.6). These interactions, which are absent in C7 due to an unrelated CDR3β structure, effectively draw CDR3α toward CDR3β in RA14, resulting in a displacement of 2.8 Å in the α-carbon position of CDR3α Asn96 relative to its position in C7.

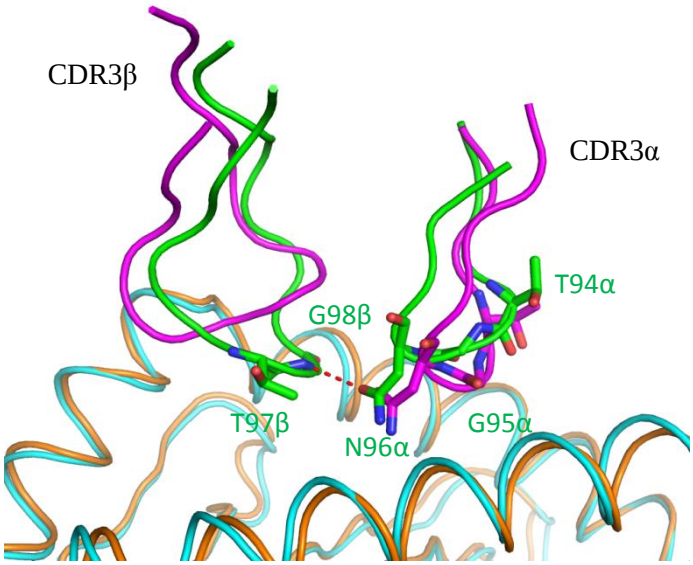


FIGURE 4.6 Influence of CDR3β on the conformation of CDR3α in TCR C7.

Conformation of CDR3α and CDR3β loops in superposed C7–NLV–HLA-A2 and RA14–NLV–HLA-A2 complexes (TCR C7, pink; C7-bound NLV–HLA-A2, cyan; TCR RA14, green; RA14-bound NLV–HLA-A2, orange). In RA14, but not in C7, the tip of CDR3β points toward CDR3α, resulting in interactions that draw CDR3α close to CDR3β in RA14. These interactions are absent in C7, whose CDR3β sequence (SQTQLWETQ) is unrelated to that of RA14 (SPVTGGIYGY).

4.3 Discussion

Previous structural studies of TCR recognition of immunodominant viral peptides presented by MHC-I molecules have focused mainly on EBV (Kjer-Nielsen et al., 2003; Tynan et al., 2005). Interestingly, comparison of three public TCRs in complex with a bulged EBV peptide presented by HLA-B8 revealed two distinct binding modes: one in which the TCR accommodates the bulged peptide but makes few contacts with MHC, and one in which the TCR focuses on the N-terminus of the peptide and leaves the bulged part untouched. In contrast to EBV, knowledge of TCR recognition of CMV has so far been limited to the RA14-NLV-HLA-A2 complex. The structural mechanism of how different TCRs are able to bind the same pMHC ligand is particularly relevant in light of a growing appreciation for the surprising diversity of public TCR response to certain viral epitopes revealed by advanced new approaches to T cell repertoire analysis.

In the case of the CMV NLV-specific TCR response, seven public CDR3α and seven public CDR3β motifs have so far been identified, including one additional public CDR3β motif reported here (Wang et al., 2013). Although public CDR3α motifs often pair with public CDR3β motifs, pairings between public and private CDR3α/CDR3β motifs occur with equal frequency. In addition, even among NLV-specific TCRs expressing dual public CDR3α/CDR3β motifs, the same public CDR3α motif may pair with different public CDR3β motifs (and the reverse). Importantly, this striking flexibility of CDR3α/CDR3β pairing is not unique to NLV-specific TCRs, as it has now also been documented among TCRs specific for the influenza NP_{366} epitope, which had previously been thought to elicit a narrow TCR repertoire comprising only a few clonotypes.

A comparison of the C7-NLV-HLA-A2 and RA14-NLV-HLA-A2 structures reveals how the same public CDR3α motif (XTGNQF) can pair with multiple unrelated CDR3β motifs, one private (SQTQLWETQ for C7) and the other public (SPVTGGIYGY for RA14) yet still remain equally effective in binding NLV-HLA-A2 (K_D=5.1 μM for C7; 28 μM for RA14). We have shown that CDR3α adopts different conformations in C7 and RA14 to accommodate large structural differences in CDR3β, which abuts CDR3α in the TCR binding site. Nevertheless, the core of the XTGNQF CDR3α motif maintains key interactions with the HLA-A2 α1 helix and P5 Met of NLV via Gly95α and Asn96α. As the result, the common Vα domain of C7 and RA14 are able to docking over NLV-HLA-A2 in the almost same way despite their pairing with totally unrelated Vβ domains in the TCR heterodimer. Based on our structural comparison of C7-NLV-HLA-A2 and RA14-NLV-HLA-A2, we predict other CDR3β sequences that have been found to pair with the XTGNQF CDR3α motif should dock over NLV-HLA-A2 in a similar fashion. More generally, our study reveals how the malleability of protein-protein interfaces permits preservation of function (in this case, pMHC specificity) through accommodation of structural changes in the binding partners.

The C25 TCR provides a different structural solution to binding NLV-HLA-A2 than C7, yet is as effective as C7 with an essentially identical K_D: 4.7 μM for C25 versus 5.1 μM for C7. These relatively high affinities support an antigen-driven selection process for both public TCRs. However, unlike C7, whose Vα domain contributes more buried surface to the interface with pMHC than Vβ (55 and 45%, respectively), the opposite is true for C25 (Vα, 40%; Vβ, 60%). In addition, the C25-NLV-HLA-A2 and C7-NLV-HLA-A2 complexes exhibit divergent TCR footprints on pMHC due to distinct crossing angles of 61° and 29°, respectively. The more acute

crossing angle of C7 enables this TCR to contact both N- and C-termini of the NLV peptide, whereas C25 is decidedly more focused on the peptide center, primarily P5 Met.

The public CDR3α motif of TCR C25 (XNNNDM) has been shown to pair with multiple public and private CDR3β motifs that vary in both sequence and length, including SISDLAKNIQ, QLQGHTEA, SVSDVANTEA, SLEGYTEA, and SLAPGATNEKL. The CDR3α structure is rigidified by eight intraloop hydrogen bonds in the C25-NLV-HLA-A2 complex, making it difficult to alter loop conformation. Thus, these CDR3β loops must likely adapt to the CDR3α XNNNDM motif in much the same way as C25 CDR3β.

CMV was recently shown to boost the immune response of young, healthy individuals to influenza. Similarly, mice infected with CMV were found to be resistant to infection with the bacterial pathogens *Listeria monocytogenes* and *Yersinia pestis*. These and related observations have led to the hypothesis that the ubiquity of CMV infection in human and many other species might be beneficial to the host. Although the underlying mechanisms of CMV-mediated cross-protection are elusive, one possibility is that CMV-specific TCRs may cross-react with epitopes from other pathogens. Indeed, a degree of cross-reactivity of CD8+ T cell epitopes between CMV and influenza has been reported. The promiscuity of TCRs, whereby a single receptor can recognize many different peptides, coupled with structural diversity of CMV NLV-specific TCRs described here, further support the notion of cross-reactivity as a possible mechanism to help explain CMV-mediated heterologous immunity to influenza and other microbial pathogens.

Chapter 5

Structural studies of GIL-specific TCRs recognizing GIL-HLA-A2

Summary

The crystal structures of the F6-GIL-HLA-A2 and F50-GIL-HLA-A2 complexes were solved to resolutions of 2.1 Å and 1.7 Å, respectively. Due to the overall sequence similarity between JM22 and F6, these two TCRs dock over GIL-HLA-A2 similarly. Both TCRs exhibit Vβ dominance in recognizing GIL-HLA-A2, as well as conserved key interactions by the CDR3β 'RSS' motif. However, F50 utilizes a completely different α/β pair thus binds GIL-HLA-A2 in a manner distinct from F22 and F6. Our results revealed the structural basis for the restricted clonal diversity of the GIL-specific CD8+ T cell response.

5.1 Background

As discussed in previous chapters, influenza virus is a ubiquitous human pathogen responsible for seasonal outbreaks. Individuals expressing the HLA-A2 molecule produce cytotoxic T lymphocytes bearing TCRs that recognize the immunodominant M1 epitope GILGFVFTL (GIL). GIL-specific T cell responses are characterized by highly restricted TRBV19 gene usage, as well as by a highly conserved 'IRSS' CDR3β motif (Moss et al., 1991; Lehner et al., 1995). Of these residues, arginine at position 98 (R98) and serine at position 99 (S99) are most conserved among TRBV19 TCRs. The residues encoding CDR3s are encoded by the joining region of V(D)J gene rearrangements with random addition or deletion of nucleotides. Thus, it is puzzling why GIL-specific responses are heavily biased to TRBV19 TCRs with extremely conserved CDR3β sequences. In an analysis of young children, GIL-specific TCRs showed no bias toward TRBV19 at the beginning, but TRBV19 gradually become dominant over the years. This suggests that TRBV19 is selected during influenza virus infections in HLA-A2+ populations. The dominance of T cell clones possessing TRBV19 also indicates that Vβ is more important than Vα in recognizing GIL-HLA-A2. Indeed, deep sequencing of HLA-A2+ GIL-specific TCR repertoires revealed that Vα usage is not as restricted as Vβ usage (Gil et al. 2015).

A previous structural study of the JM22 TCR-GIL-HLA-A2 complex revealed distinct mechanisms for recognizing the featureless GIL peptide (Stewart-Jones et al., 2013). The resolution of JM22-GIL-HLA-A2 structure (1.4 Å) is the highest among all the TCR-pMHC complexes reported to date. The JM22 TCR sits over GIL-HLA-A2 in canonical docking mode. Interestingly, the Vβ domain of JM22 buries 67% of the total solvent-accessible surface and mediates more than 70% of total interactions. In contrast, the Vα domain shows only limited involvement in contacting GIL-HLA-A2. The extremely conserved CDR3β 'RSS' motif is

critical for linking the GIL peptide and HLA-A2 to the TCR. The arginine at position 98 (R98) of JM22 Vβ occupies a notch between the GIL peptide and HLA-A2. This allows JM22 to efficiently contact the GIL peptide, which lacks solvent-exposed residues. Mutating R98 to other residues completely abrogated the interaction between JM22 and GIL-HLA-A2. By contrast, mutating S99 to alanine had no impact on JM22-GIL-HLA-A2 complex formation. Considering that both arginine and serine can be generated from multiple codons, the conserved CDR3β 'RSS' motif may be selected under both structural and nucleotide pressure. The JM22-GIL-HLA-A2 complex recruits 10 water molecules to mediate water bridges that reinforce tight binding between TCR and pMHC. Such interfacial waters are rarely observed in other TCR-pMHC complexes. This could be an effective strategy to promote TCR binding to featureless peptides. Alternatively, these other TCR-pMHC structures may simply lack interfacial waters due to relatively low resolutions. To address this question, additional TCR-pMHC structures involving featureless peptides need to be determined to high resolution.

Recent advances in high-throughput sequencing and single-cell paired analysis have allowed immunologists to exhaustively interrogate GIL-specific T cell repertoires. Many new GIL-specific TCR α and β clonotypes have been identified, as well as new CDR3α and CDR3βs motifs. Although TRBV19 TCRs are still the largest population, other α/β pairs are observed. Whether these non-TRBV19 TCRs are as effective as TRBV19 TCR in recognizing GIL-HLA-A2 remains to be determined. In addition, the TRBV19 β chain can either pair with non-TRAV27 α chains or with TRAV27 α chains with different CDR3αs. Such α chain swapping increases the total diversity of GIL-specific TCR repertoires. It remains unclear whether swapping different TRAVs or different CDR3αs can impact TCR-pMHC complex

formation. To address these questions, we crystallized three GIL-specific TCR (F22, F6 and F50) in complex with GIL-HLA-A2.

5.2 Results

5.2.1 Crystallization of TCRs F6 and F50 bound to GIL-HLA-A2

The constructs for expressing F6 and F50 contained an artificial disulfide bond in the TCR constant region. The HLA-A2 heavy chain and β_2-microglobulin are wide-type constructs. The GIL peptide was chemically synthesized (GenScript). Both TCRs and GIL-HLA-A2 are produced via *in vitro* folding from bacterial inclusion bodies. Initial crystallization screenings of the F6-GIL-HLA-A2 and F50-GIL-HLA-A2 complexes were carried out using Rigaku Wizard I, II, III and IV kits. Because no crystals were obtained with these kits, we developed an in-house crystallization screen. The F6-GIL-HLA-A2 grew plate-like crystals in 10-15 % (w/v) of PEG 3350, 0.1 M imidazole (pH 8.0). The crystals were cryoprotected with 25% glycerol (w/v) solution prior to flash-cooling. The F50-GIL-HLA-A2 grew plate-like crystals in 10-15% (w/v) of PEG 3350, 0.1 M imidazole (pH 8.0), and 0.2 M sodium malonate. The crystals were cryoprotected with 25% glycerol (w/v) solution prior to flash-cooling.

5.2.2 X-ray crystallographic data collection, structure determination and refinement

Diffraction datasets were collected at beamline 24ID-E of the Advanced Photon Source, Argonne National Laboratory with an ADSC Q315 CCD detector. Data were indexed, integrated, and scaled with the HKL2000 program. Data collection statistics are presented in Table 5.1. The

overall completeness of the F6-GIL-HLA-A2 and F50-GIL-HLA-A2 datasets are 99.9% and 95.6%, respectively. The F6-GIL-HLA-A2 dataset was processed to 2.1 Å resolution; the F50-GIL-HLA-A2 dataset was processed to 1.7 Å resolution.

TABLE 5.1
Data collection and structure refinement statistics

	F6–GIL–HLA-A2	F50–GIL–HLA-A2
Data collection		
Space group	$P12_11$	$P12_11$
Cell dimensions		
a (Å)	134.7	66.3
b (Å)	54.1	71.1
c (Å)	149.3	100.7
α, β, γ (°)	90, 116.6, 90	90, 96.2, 90
Resolution range (Å)[a]	120.5–2.10 (2.18–2.10)	37.9–1.7 (1.76–1.7)
Unique reflections[a]	113,074 (11,231)	100320 (9832)
R_{merge}[a,b]	0.144 (1.678)	0.055 (0.647)
Mean $I/\sigma(I)$[a]	11.7 (1.3)	7.52 (1.2)
Completeness (%)[a]	99.9 (99.8)	98.1 (96.6)
Refinement		
Resolution range (Å)[a]	120.5–2.10	37.9–1.7
R_{work} (%)/R_{free} (%)[a,c]	22.9 (34.2)/28.1 (38.5)	19.1 (28.9)/21.3 (28.0)
No. of protein atoms	13,037	6662
No. of water molecules	397	808
r.m.s.d from ideality		
Bond lengths (Å)	0.010	0.013
Bond angles (°)	1.29	1.18
Ramachandran statistics (%)		
Most favored	95.0	98.0
Allowed	4.9	2.0
Disallowed	0.1	0.0

[a]Values in parentheses are statistics for the highest resolution shell.
[b]$R_{merge} = \sum|I_j - \langle I \rangle|/\sum I_j$, where I_j is the intensity of an individual reflection and $\langle I \rangle$ is the average intensity of that reflection.
[c]$R_{work} = \sum||F_o| - |F_c||/\sum|F_o|$, where F_c is the calculated structure factor. R_{free} is as for R_{work} but calculated for a randomly selected 5.0% of reflections not included in the refinement.

To determine the F6-GIL-HLA-A2 complex structure via MR, JM22 TCR (PDB accession code: 1OGA) and GIL-HLA-A2 (PDB accession code: 1OGA) were used as initial search models. The CDR3α of JM22 TCR and the GIL peptide were removed. Initially, one of the F6 TCR models was located in a different asymmetric unit. A new search model was generated by deleting that TCR and placing a TCR opposite the located GIL-HLA-A2 molecule. Two F6-GIL-HLA-A2 complexes were then found by Phaser. Structural refinement was performed using rigid body and simulated annealing via the Phenix program. The model was further refined by manual model building with Coot based on $2F_o$-F_c and F_o-F_c maps with the NLV peptide omitted in the initial refinement. The final R_{work} and R_{free} values for F6-GIL-HLA-A2 complex are 22.9% and 28.1%, respectively (Table 5.1).

The F50-GIL-HLA-A2 complex crystal dataset was processed to 1.7 Å resolution with an overall R_{merge}= 5.3%. To determine the F50-GIL-HLA-A2 complex structure via MR, the HIV-specific TCR T36-5 (PDB accession code: 3VXU) and GIL-HLA-A2 (PDB accession code: 1OGA) were used as initial search models. Only one complex molecule was in one asymmetric unit based on solvent content. Indeed, one F50-GIL-HLA-A2 complex was found by Phaser. Structural refinement was performed using rigid body and simulated annealing via Phenix. The model was further refined by manual model building with Coot based on $2F_o$-F_c and F_o-F_c maps with the NLV peptide omitted in the initial refinement. The final R_{work} and R_{free} values for F50-GIL-HLA-A2 complex are 19.0% and 25.4%, respectively (Table 5.1). Refinement statistics are summarized in Table 5.1. Stereochemical parameters were evaluated by PROCHECK.

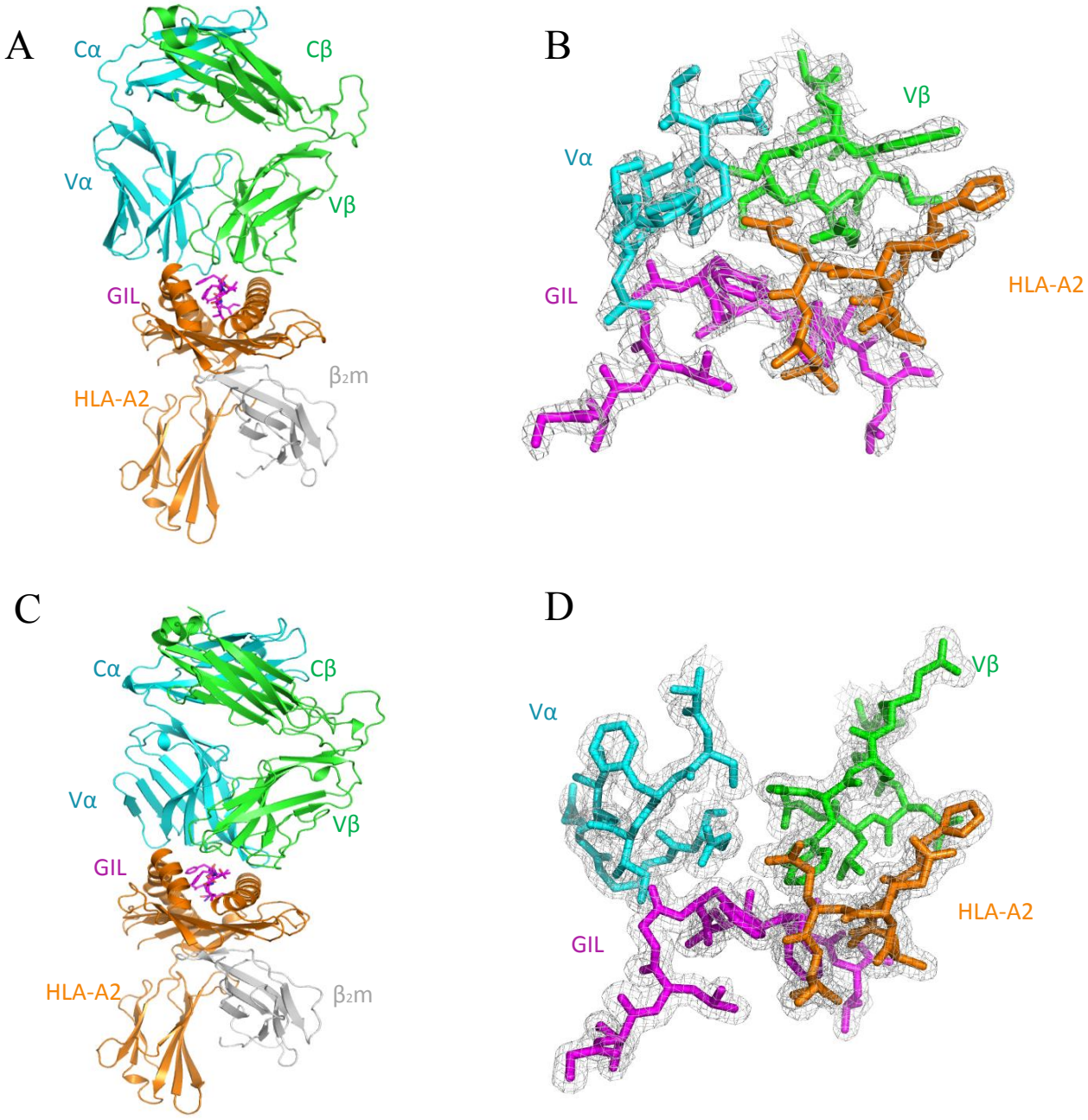


Figure 5.1. Structure of TCR–GIL–HLA-A2 complexes.

(A) Side view of the F6–GIL–HLA-A2 complex (ribbon diagram). TCR α chain, cyan; TCR β chain, green; HLA-A2 heavy chain, orange; β_2-microglobulin (β_2m), gray; GIL peptide, magenta.
(B) Electron density in the interface of the F6–GIL–HLA-A2 complex. Density from the final $2F_o - F_c$ map at 2.1 Å resolution is contoured at 1σ.
(C) Side view of the F50–GIL–HLA-A2 complex.
(D) Electron density in the interface of the F50–GIL–HLA-A2 complex. Density from the final $2F_o - F_c$ map at 1.7 Å resolution is contoured at 1σ.

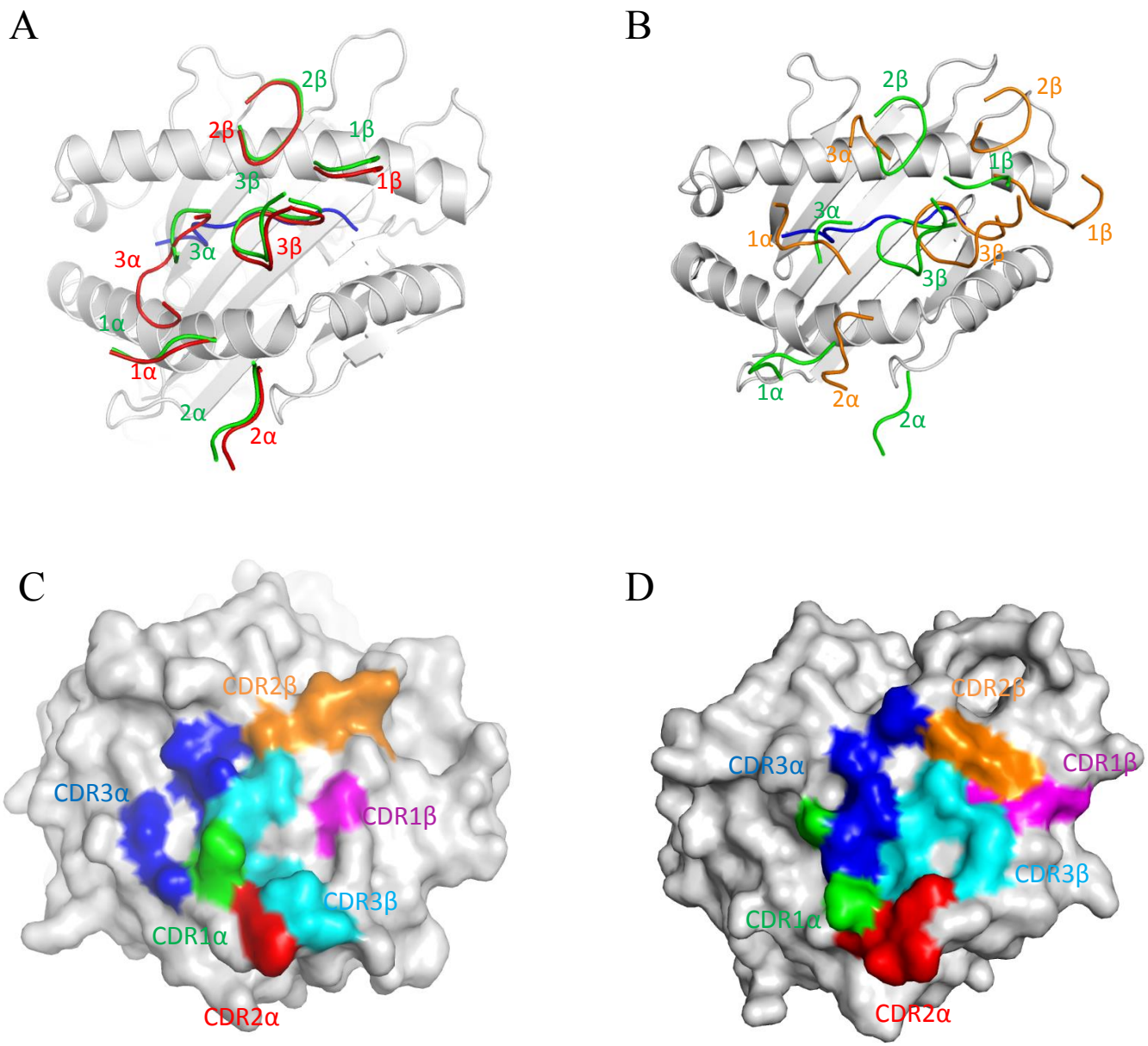


Figure 5.2 Comparison of TCR footprints on NLV–HLA-A2.

(A) Positions of CDR loops of TCRs F6 and JM22 (PDB accession code: 1OGA) (27) on GIL–HLA-A2 (top view). CDRs of F6 are shown as numbered red loops. CDRs of JM22 are green. HLA-A2 is gray. The GIL peptide is blue.
(B) Positions of CDR loops of TCRs F50 and JM22 on GIL–HLA-A2. CRDs of F50 are orange. CDRs of JM22 are green.
(C) Footprint of TCR F6 on GIL–HLA-A2. The top of the MHC molecule is depicted as a gray surface. The areas contacted by individual CDR loops are color-coded: CDR1α, green; CDR2α, red; CDR3α, blue; CDR1β, magenta; CDR2β, orange; CDR3β, cyan.
(D) Footprint of TCR F50 on GIL–HLA-A2.

5.2.3 Overview of the F6-GIL-HLA-A2 and F50-GIL-HLA-A2 complexes

Both the F6-GIL-HLA-A2 (Figure 5.1-A and B) and F50-GIL-HLA-A2 (Figure 5.1-C and D) complexes showed unambiguous electron density for two (F6-GIL-HLA-A2) or one (F50-GIL-HLA-A2) complex molecules in the asymmetric unit. The r.m.s.d. in α-carbon positions for the TCR VαVβ and MHC α1α2 modules, including the GIL peptide, is 0.2 Å for the two F6-GIL-HLA-A2 complexes. Of note, the resolution of the F50-GIL-HLA-A2 complex (1.7 Å) is the second highest reported for any TCR-pMHC-I or TCR-pMHC-II complex, the highest (1.4 Å) being for the JM22-GIL-HLA-A2 complex.

Both F6 and F50 sit symmetrically over NLV-HLA-A2 in a canonical diagonal orientation. The crossing angles of F6 TCR and F50 to pMHC are 69° and 29°, respectively. Upon binding GIL-HLA-A2, F6 and F50 bury 85% (258 $Å^2$) and 73% (215 $Å^2$) of peptide solvent-accessible surface. The percentage for F6-GIL-HLA-A2 is at the higher end of the range for TCR-pMHC class I complexes, which varies from 60 to 91% in other structures. Extensive peptide burial enables F6 to maximize readout of the GIL peptide. Such extensive peptide burial is also a salient feature of the JM22-GIL-HLA-A2 complex.

5.2.4 Interaction of TCR F6 with HLA-A2

The F6-GIL-HLA-A2 complex buries a total solvent-accessible surface of 1475 $Å^2$, which is at the lower end of the range for TCR–pMHC complexes (1240– 2400 $Å^2$), but comparable to that in the JM22–GIL–HLA-A2 complex (1560 $Å^2$). In the F6–GIL–HLA-A2

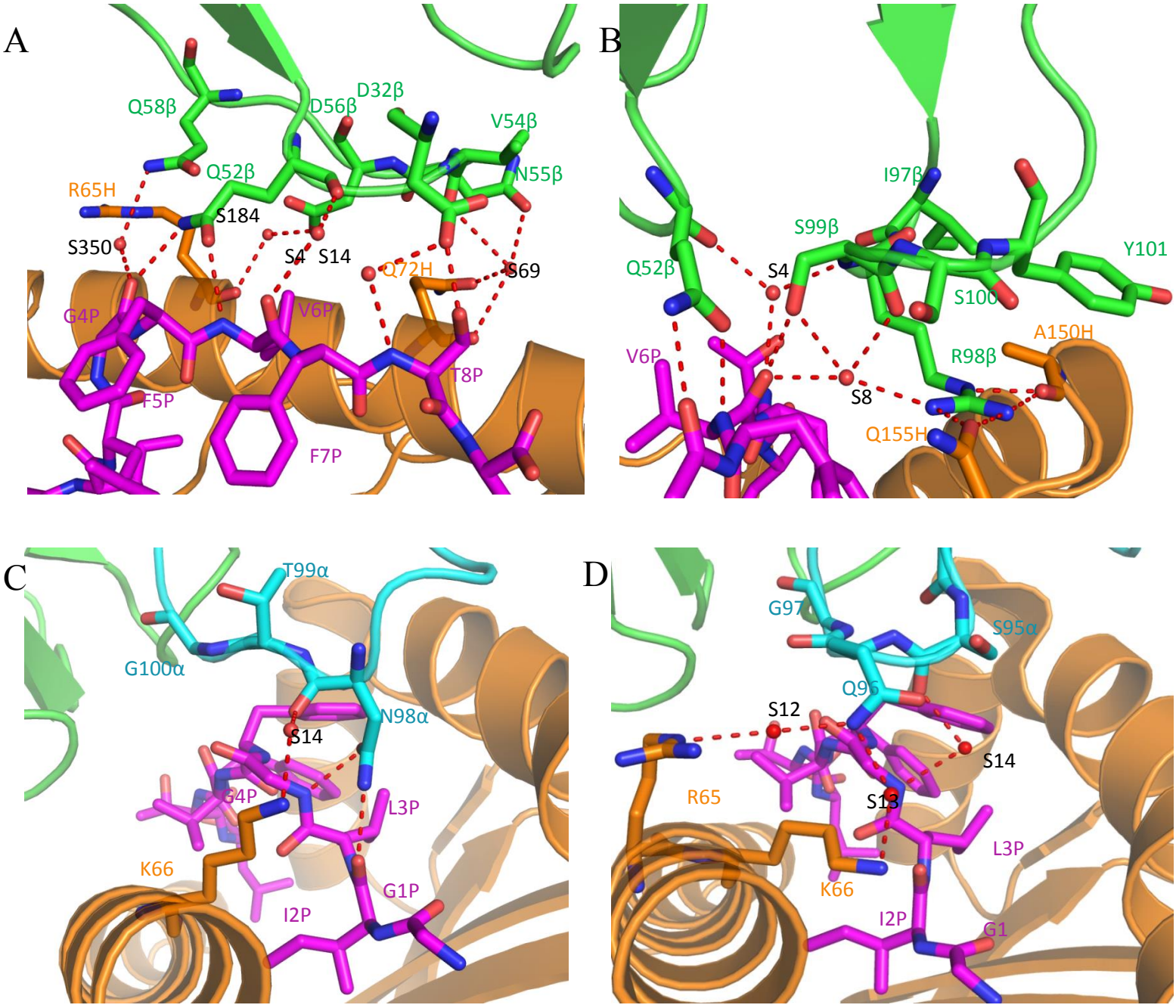


Figure 5.3 Interactions of TCRs F6 with the GIL peptide and HLA-A2.

(A) Direct and water-mediated interactions between TCR F6 (green) and the GIL peptide (magenta). The HLA-A2 α1 helix is orange. The side chains of contacting residues are drawn in stick representation with carbon atoms in green (CDR1β and CDR2β) or magenta (GIL), nitrogen atoms in blue, and oxygen atoms in red. Water molecules are depicted as red spheres. Hydrogen bonds are indicated by red dashed lines. Peptide residues are identified by a one-letter amino acid designation followed by position (P).

(B) Interactions between CDR3β (green) of F6 and GIL–HLA-A2, showing hydrogen bond network mediated by Arg98β and Ser99β. The guanidinium group of Arg98β inserts into a shallow pocket between the GIL peptide (magenta) and the HLA-A2 α2 helix (orange).

(C) Interactions between CDR3α (cyan) of F6 and GIL–HLA-A2.

(D) Interactions between CDR3α (cyan) of JM22 (Stewart-Jones et al., 2003) and GIL–HLA-A2.

complex, the buried surface area on Vβ (424 $Å^2$, 58%) is considerably greater than on Vα (314 $Å^2$, 42%). This Vβ dominance is also a salient feature of the JM22–GIL–HLA-A2 complex. As discussed in a previous chapter, such dominance by Vβ is unusual among TCR-pMHC class I complexes, in which Vα and Vβ typically contribute roughly equal buried surfaces. Indeed, only four other TCR-pMHC class I complexes displaying a similar degree of Vβ dominance as F6-GIL-HLA-A2 have been reported, involving the HLA-A2 restricted TCR JM22 (67%), the H-$2K^b$ restricted TCR BM3.3 (63%), HLA-E-restricted TCR KK50.4 (61%), and HLA-A2 restricted TCR C25 (60%).

Of the total buried surface on HLA-A2, excluding GIL peptide, CDR1α, CDR2α and CDR3α contribute 7%, 10% and 24%, respectively, compared with 2%, 28% and 27%, respectively, for CDR1β, CDR2β and CDR3β. Compared to JM22 TCR, which utilizes the TRAV27-TRAJ42 and TRBV19-TRBD2-TRBJ2-7 combination, F6 utilizes the TRAV27-TRAJ37 and TRBV19-TRBD2-TRBJ2-7 pair. Thus, TCRs JM22 and F6 share nearly the same sequences, except for CDR3α. The JM22 CDR3α (AGAGSQGNLI) is three amino acids shorter than the F6 CDR3α (AGAIGSSNTGKLI). The shorter CDR3α of JM22 results in a considerably smaller buried surface area (11%) compared to F6 (24%), given that both complexes have similar total buried surfaces. As expected from the larger buried surface area on Vβ, Vβ made significantly more contacts (44 contacts in total) than Vα (19 contacts in total) to HLA-A2. Because of F6 uses the same TRBV19-TRBD2-TRBJ2-7 combination as JM22, almost all contacts made by JM22 Vβ to HLA-A2 are maintained in the F6-GIL-HLA-A2 complex, including hydrogen bonds involving both germline-encoded (Asn51β-Arg75H) and somatically-generated (Arg98β-Ala150H and Arg94β-Gln155H) TCR residues. Moreover, water-mediated

contacts between JM22 Vβ and HLA-A2 are also preserved in the F6-GIL-HLA-A2 complex, including Asp52βO-H_2O-Arg65HO, V50βO-H_2O-Q72HO, and

Table 5.2
Interactions between TCR and MHC in the F6–GIL–HLA-A2 and JM22–GIL–HLA-A2 complexes

	Hydrogen bonds	Van der Waals contacts	Hydrogen bonds	Van der Waals contacts
HLA-A2	F6	F6	JM22	JM22
R65H				Q58β
K66H		N98α		
K68H				D56β
A69H		D56β		D56β
Q72H		V54β N55β		I53β N55β
T73H				I53β
R75H		N55β		N55β
V76H				I53β
A149		Y101β		Y101β
A150H	A150H(O) R98β(N^{ε}) A150H(O) R98β($N^{\eta 2}$)	I97β R98β Y101β	A150H(O) R98β(N^{ε}) A150H(O) R98β($N^{\eta 2}$)	R98β Y101β
H151H		V51α R98β Y101β		V51α R98β Y101β
V152H		R98β		R98β
E154H				S31α V51α
Q155H	Q155H($N^{\varepsilon 2}$) T99α(O) Q155H($O^{\varepsilon 1}$) R98β($N^{\eta 1}$) Q155H($O^{\varepsilon 1}$) R98β($N^{\eta 2}$)	S31α R98β S100β	Q155H($O^{\varepsilon 1}$) R98β($N^{\eta 1}$) Q155H($O^{\varepsilon 1}$) R98β($N^{\eta 2}$)	S31α G94α R98β S100β
A158H		G95α		
Y159H		N98α		
T163H		P96α N98α		

Table 5.3
Water bridges between TCR and MHC in the F6–GIL–HLA-A2 and JM22–GIL–HLA-A2 complexes

HLA-A2	Water	F6	*B* factor	Water	JM22	*B* factor
E154H(O)	S117	R70α($N^{\eta 1}$)	37.1	S60	R70α($N^{\eta 1}$)	25.2
R65H(N^{ε})				S123	Q96α($N^{\varepsilon 2}$)	33.3
K66H(N^{ζ})	S149	N98α(N)	37.0	S135	Q96α($N^{\varepsilon 2}$)	27.5
R65H(O)	S184	D56β($O^{\delta 1}$)	33.2	S150	D56β($O^{\delta 1}$) D56β($O^{\delta 2}$)	33.4
Q72H($O^{\varepsilon 1}$) Q72H(O)	S69	V54β(O) N55β($O^{\delta 1}$)	37.6	S601	V54β(O) N55β($O^{\delta 1}$)	28.7

N51βN-H_2O-Q72HO (Figure 5.3-A and B, Table 5.2 and 5.3). The largest structural change in HLA-A2 upon F6 binding involves a 2.2 Å shift in the position of the Gln155H side chain. This reorientation, which was also identified in JM22-GIL-HLA-A2 structure, opens a notch that accommodates the Arg98β side chain. Remarkably, Gln155H in F50-GIL-HLA-A2 structure is reoriented to face the TCR and completely open the notch between HLA-A2 and the GIL peptide. Thus, it is evident that Gln155H has great flexibility to adapt to different TCRs, which may permit a more diverse T cell response. Unlike JM22, whose CDR3α only makes one direct contact with HLA-A2, F6 CDR3α makes 9 contacts with HLA-A2, including one hydrogen bond between Thr99αO and Gln155HN and one salt bridge between Asn98αN and Lys66HN (Figure 5.3-C and D, Table 5.2 and 5.3). Thus, the TRAV27-TRAJ37 combination is able to provide more interactions with HLA-A2 than TRAV27-TRAJ42. This may be advantageous in cases where TRAV27-TRAJ37 pairs with Vβs other than TRBV19-TRBD2-TRBJ2-7.

5.2.5 Interaction of TCR F6 with the GIL peptide

Upon binding to GIL-HLA-A2, F6 buries 85% (258 $Å^2$) of accessible surface of the GIL peptide, indicating maximum readout of the peptide. Peptide recognition is evenly distributed between Vα (23 contacts) and Vβ (21 contacts). All the contacts made by Vα are solely contributed by F6 CDR3α. In contrast, the 21 contacts made by Vβ are distributed among all CDRβs, but mostly CDR2β (12 contacts). Thus, peptide recognition by F6 is both germline and somatically governed. Although JM22 also makes 21 contacts to GIL via Vβ, only 15 contacts are made by Vα. The fewer contacts made by JM22 Vα are mainly due to its shorter CDR3α comparing to F6 CDR3α. Five polar contacts to the GIL peptide (Asn98αOD1-N P4Gly,

Asn98αND2-N P2Ile, Asp28βO-OG1 P8Thr, Gln48βOE1-O P6Val and Ser95βOG-O P6Val) are made upon F6 TCR binding. Additionally, Asn98α makes two hydrogen bonds with P2Ile and P4Gly backbone nitrogens. These interactions are absent in the JM22-GIL interface due to the shorter CDR3α. The five polar contacts are further reinforced by 39 van der Waals contacts and 6 water-mediated hydrogen bonds (Figure 5.3 C and D; Table 5.4 and 5.5). Thus, although the GIL peptide only has P8Thr substantially exposed in solvent, F6 TCR combines several strategies to achieve maximum peptide readout.

The high resolution of the F6–GIL–HLA-A2 complex permitted the inclusion of many ordered water molecules in the structure, including ones in the interface between TCR and pMHC (Figure 5.3; Table 5.3 and 5.5). In particular, eight water molecules are mostly or completely buried in the interface, where they mediate hydrogen bonding interactions between F6 and the GIL peptide. In addition to forming an intricate solvent network linking TCR and pMHC, these and other bound waters contribute significantly to interfacial shape complementarity, based on calculations of the shape correlation statistic (S_c). Thus, the S_c value for the F6–GIL–HLA-A2 complex with interfacial waters is 0.61 (S_c = 1.0 for interfaces with perfect geometrical fits), but only 0.54 without interfacial waters. The corresponding S_c values for the JM22–GIL–HLA-A2 complex are 0.77 with interfacial waters versus 0.64 without such waters. The substantially greater shape complementarity of the JM22–GIL–HLA-A2 interface likely contributes to the higher affinity of TCR JM22 compared to F6, as discussed below. Notably, most bridging water molecules are conserved in the F6–GIL–HLA-A2 and JM22–GIL–HLA-A2 complexes, even though they crystallized under different conditions and in different space groups. This conservation underscores the intrinsic importance of interfacial waters to complex stabilization. Besides the lower shape complementarity of the F6-GIL-HLA-A2 than the

Table 5.4

Interactions between TCR and GIL peptide in the F6–GIL–HLA-A2 and JM22–GIL–HLA-A2 complexes

	Hydrogen bonds	Van der Waals contacts	Hydrogen bonds	Van der Waals contacts
GIL	F6	F6	JM22	JM22
I2P	I2P(O) N98α($N^{\delta 2}$)	N98α		
L3P		N98α		
G4P	G4P(N) N98α($O^{\delta 1}$) G4P(O) Q52β($N^{\varepsilon 2}$)	N98α G100α Q52β	G4P(O) Q52β($N^{\varepsilon 2}$)	S95α Q96α Q52β
F5P		T99α G100α Q52β R98β S99β		S95α G97α Q52β R98β S100β
V6P	V6P(N) Q52β($O^{\varepsilon 1}$) V6P(O) S99β(O^{γ})	Q52β S99β	V6P(N) Q52β($O^{\varepsilon 1}$) V6P(O) S99β(O^{γ})	Q52β S99β
T8P	T8P($O^{\gamma 1}$) D32β($O^{\delta 2}$)	D32β I53β	T8P($O^{\gamma 1}$) D32β($O^{\delta 2}$)	D32β I53β

Table 5.5

Water bridges between TCR and GIL peptide in the F6–GIL–HLA-A2 and JM22–GIL–HLA-A2 complexes

GIL	Water	F6	*B* factor	Water	JM22	*B* factor
V6(O)	S4	Q52β(O) S99β(N)	21.6	S1	Q52β(O) S99β(N)	17.2
V6(O)	S8	R98β($N^{\eta 1}$) S99β(O^{γ})	22.0	S6	R98β($N^{\eta 1}$) S99β(O^{γ})	18.4
G4(O)	S350	Q58β($N^{\varepsilon 2}$)	40.5	S10	Q58β($N^{\varepsilon 2}$)	20.5
G4(N)				S14	S95α(O)	19.0
T8(N)	S14	D32β($O^{\delta 1}$)	21.6	S19	D32β($O^{\delta 1}$)	18.7

JM22-GIL-HLA-A2 interface, another factor that could contribute to the lower affinity of F6 may be the entropic cost of reducing the conformational mobility of the F6 VαCDR3 loop upon complex formation, which would at least partially offset any enthalpy gained from contacts made by this loop. This entropic cost may be higher for F6 than for JM22 due to the greater length, and presumably greater flexibility, of F6 VαCDR3.

5.2.6 Interaction of the TCR F50 with HLA-A2

The F50-GIL-HLA-A2 complex buries a total solvent-accessible surface of 1760 $Å^2$, comparable to other TCR-pMHC complexes. Vα and Vβ bury 46% (333 $Å^2$) and 54% (391 $Å^2$) of HLA-A2 surface area, respectively. This roughly equal contribution by the Vα and Vβ chains is distinct from JM22, which features a dominant Vβ (67%) in engaging HLA-A2. Of the total buried surface on HLA-A2, excluding GIL, CDR1α, CDR2α and CDR3α contribute 6%, 22% and 19%, compared with 9%, 20% and 28% for CDR1β, CDR2β and CDR3β, respectively. Both the CDR2α and CDR2β buried surfaces are significantly higher than average for TCR-pMHC complexes (11% and 12%, respectively), highlighting the extensive germline-encoded interactions made by F50 to HLA-A2. In particular, CDR2α of JM22 only contributes 7% of the buried surface on HLA-A2. Unlike JM22 and F6, F50 does not recruit water molecules to the interface with HLA-A2. Of 52 total contacts, 27 are made by Vα and 25 by Vβ. Such equal distribution of TCR-MHC interactions is distinct from that of TCRs expressing TRBV19. Residues Asn32α, Ser53α, and Asn54α of F50 Vα form three side- chain-side-chain hydrogen bonds to residues Glu154H and Gln155H on the HLA-A2 α2 helix (Figure 5.4-B).

Table 5.6
Interactions between TCR and MHC in the F50–GIL–HLA-A2 and JM22–GIL–HLA-A2 complexes

	Hydrogen bonds	Van der Waals contacts	Hydrogen bonds	Van der Waals contacts
HLA-A2	F50	F50	JM22	JM22
R65H				Q58β
K66H		N98α		
K68H				D56β
A69H		I96β G99β		D56β
Q72H		M50β G99β		I53β N55β
T73H				I53β
R75H				N55β
V76H		M50β		I53β
T80H	A80H($O^{\gamma 1}$)N51β($N^{\delta 2}$)	N51β		
Y84H	Y84H(OH)E30β($O^{\varepsilon 2}$)	E30β		
K146H		L95β		
A149H				Y101β
A150H		R52α L96β W99β	A150H(O) R98β(N^{ε}) A150H(O) R98β($N^{\eta 2}$)	R98β Y101β
H151H		R52α		V51α R98β Y101β
V152H		W99β		R98β
E154H	E154H($O^{\varepsilon 2}$)N54α($N^{\delta 2}$)	N54α		S31α V51α
Q155H	Q155H($N^{\varepsilon 2}$) N32α($N^{\delta 2}$) Q155H($O^{\varepsilon 1}$) S53α(O^{γ})	N32α R52α S53α N54α W99β	Q155H($O^{\varepsilon 1}$) R98β($N^{\eta 1}$) Q155H($O^{\varepsilon 1}$) R98β($N^{\eta 2}$)	S31α G94α R98β S100β

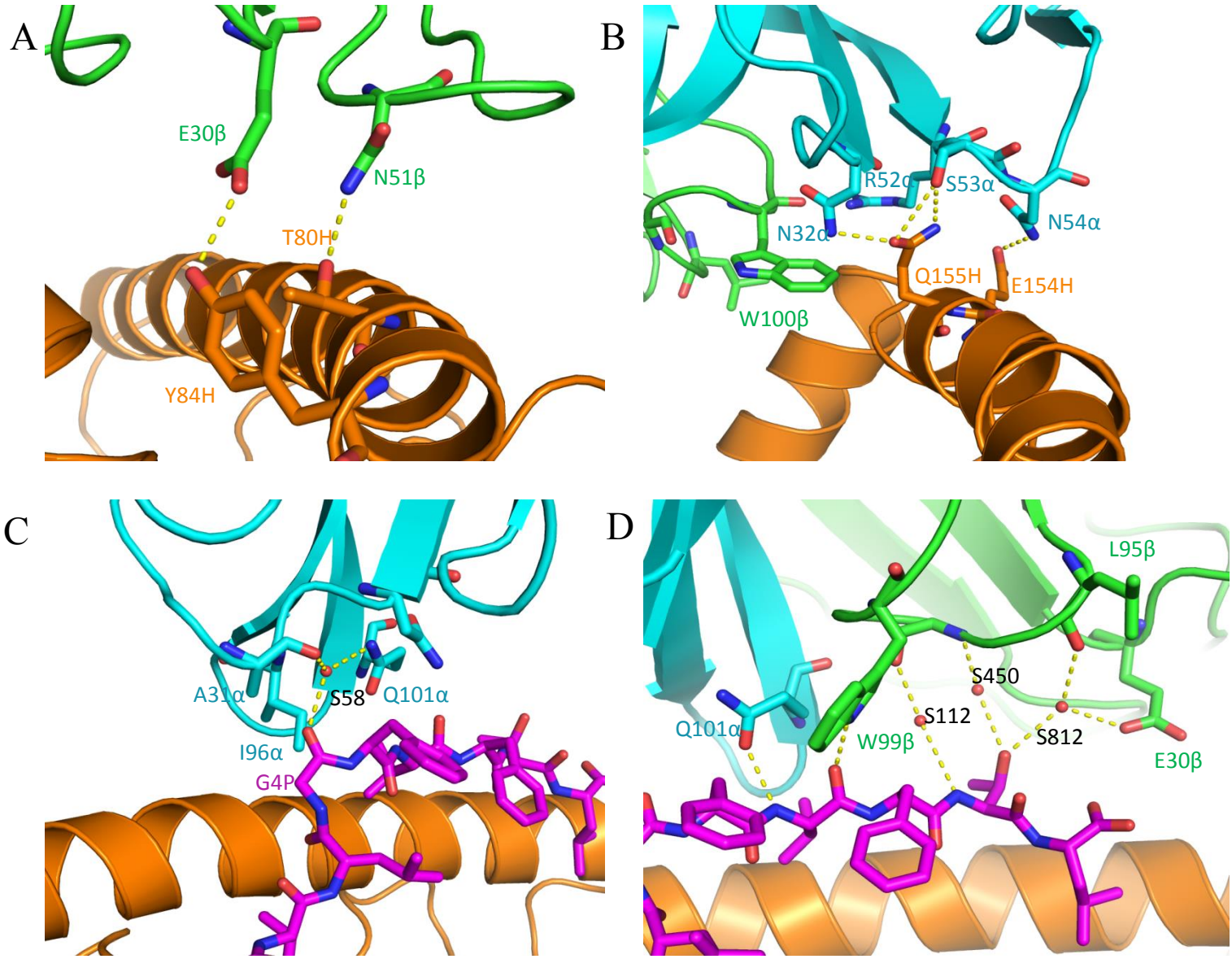


Figure 5.4 Interactions of TCR F50 with the GIL peptide and HLA-A2.

(A) Direct interactions between TCR F50 (green) and the HLA-A2 (orange). The side chains of contacting residues are drawn in stick representation with carbon atoms in green (CDR1β and CDR2β), nitrogen atoms in blue, and oxygen atoms in red. Hydrogen bonds are indicated by yellow dashed lines. Water molecules are depicted as red spheres. Peptide residues are identified by a one-letter amino acid designation followed by position (P).
(B) Interactions between CDR2α (cyan), CDR3β (green) of F50 and GIL–HLA-A2, showing hydrogen bond network mediated by Asn32α, Ser53α and Asn54α.
(C) Interactions between Vα (cyan) of F50 and GIL peptide.
(D) Interactions between Vβ (green) of F50 and GIL peptide.

These hydrogen bonds are further reinforced by 16 germline-encoded van der Waals contacts from CDR1α and CDR2α. Due to an acute crossing angle (29^0), the F50 CDR3α loop makes 8 contacts with the HLA-A2 α1 helix instead of interacting with HLA-A2 α2 helix, as do JM22 and F6. All of these contacts are van der Waals contacts due to the very hydrophobic nature of the CDR3α motif (FIIQGA). TCR F50 contacts the HLA-A2 α1 helix through CDR1β and CDR2β. In particular, the side chains of Glu30β, Met50β and Asn51β make germline-encoded contacts with the HLA-A2 α1 helix. Glu30β and Asn51β form three side-chain-side-chain hydrogen bonds to Tyr84H and Thr80H on the HLA-A2 α1 helix (Figure 5.4-A). These interactions focus on the C-terminus of the α1 helix, which is rarely targeted by other TCRs. For instance, the HLA-A24-restricted T36-5 TCR has utilizes the same TRBV27 as F50 but Asn51β forms two hydrogen bonds with Ala69 and Thr73 of HLA-A24. Similar to F50 CDR3α, the somatically-generated interactions between F50 CDR3β and the HLA-A2 α2 helix are exclusively van der Waals contacts, with the majority (9 out of 15) of interactions focusing on Trp99β. The bulky Trp99β side chain occupies a notch between HLA-A2 and the GIL peptide, in a way that somewhat mimics Arg98β of JM22 and F6 TCRs (Figure 5.4-D). However, Trp99β cannot form hydrogen bonds with HLA-A2 Ala150H or Gln155H. Mutating Trp99β to Arg completely abolished the interaction between F50 and GIL-HLA-A2, confirming the importance of this residue. Interestingly, Gln155H has different rotamer conformations in JM22/F6-GIL-HLA-A2 and F50-GIL-HLA-A2 complexes. In JM22/F6-GIL-HLA-A2 complexes, the side chain of Gln155H points toward Arg98β, with which it makes two hydrogen bonds. By contrast, the Gln155H side chain is pushed away from Trp99β in the F50-GIL-HLA-A2 complex probably due to steric hindrance (Figure 5.5).

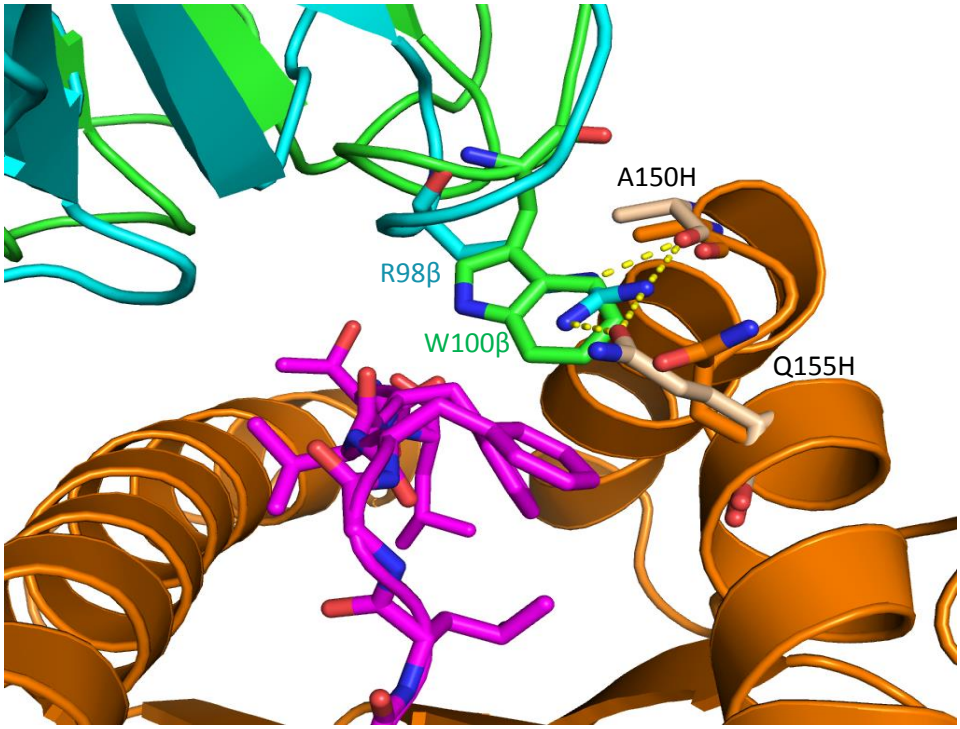


Figure 5.5 Different Q155H rotamer conformations in the F50-GIL-HLA-A2 and JM22-GIL-HLA-A2 complexes.

A150H and Q155H in the JM22-GIL-HLA-A2 complex are colored beige. A150H and Q155H in the F50-GIL-HLA-A2 complex are colored orange. TCR JM22 is colored cyan; TCR F50 is colored green.

5.2.6 Interaction of TCR F50 with the GIL peptide

Except for a few contacts (5 out of 36) made by F50 CDR1α and CDR1β, all interactions between F50 and the GIL peptide are mediated by the somatically-generated CDR3β loops, with CDR3α and CDR3β accounting for 18 and 13 contacts, respectively. Peptide specificity is conferred mainly by shape complementarity, since the F50-GIL interface features only two hydrogen bonds: F50 Gln101α Oε1-N P6 Val and F50 Trp99β Nε1-O P6 Val. F50 engages central and C-terminal portions of the GIL peptide (P4 Gly, P5 Phe, P6 Val, P7 Phe, P8 Thr and P9 Leu) with the principal focus on P5 Phe and P6 Val (Figure 5.4-C and D). The N-terminal P1-P3 residues of GIL made no contact with F50, which explains why F50 buries only 73% (215 $Å^2$) of the peptide solvent-accessible surface upon binding GIL-HLA-A2 compared to 85% (258 $Å^2$)

by F6. The longer CDR3α of F6 enables this TCR to make several contacts with Ile P2 and Leu P3. Residues CDR3α Gln101α and CDR3β Trp99β of F50 protrude from the tips of the CDR3 loops and grasp the center of peptide like two fingers. Indeed, these two residues alone account for 23 out of 36 contacts made by TCR F50 to the GIL peptide. Whereas F50 makes 88 total contacts to GIL-HLA-A2, JM22 makes 104. This difference likely contributes to the 20-fold weaker affinity of F50 than F6 for GIL-HLA-A2.

Table 5.7
Interactions between TCR and GIL peptide in the F50–GIL–HLA-A2 and JM22–GIL–HLA-A2 complexes

	Hydrogen bonds	Van der Waals contacts	Hydrogen bonds	Van der Waals contacts
GIL	F50	F50	JM22	JM22
I2P				
L3P				
G4P		A30α I96α	G4P(O) Q52β($N^{\varepsilon 2}$)	S95α Q96α Q52β
F5P		N32α I96α Q101α W99β		S95α G97α Q52β R98β S100β
V6P	V6P(N) Q101α($O^{\varepsilon 1}$) V6P(O) W99β($N^{\varepsilon 1}$)	Q101α W99β	V6P(N) Q52β($O^{\varepsilon 1}$) V6P(O) S99β(O^{γ})	Q52β S99β
T8P		L96β	T8P($O^{\gamma 1}$) D32β($O^{\delta 2}$)	D32β I53β
L9P		E30β		

Table 5.8
Water bridges between TCR and GIL peptide in the F50–GIL–HLA-A2 and JM22–GIL–HLA-A2 complexes

GIL	Water	F50	*B* factor	Water	JM22	*B* factor
G4(O)	S58	A31α(O) Q101α($N^{\varepsilon 2}$)	22.3	S1	Q52β(O) S99β(N)	17.2
T8(N)	S122	G97β(O)	23.9	S6	R98β($N^{\eta 1}$) S99β(O^{γ})	18.4
T8($O^{\gamma 1}$)	S450	L95β(O)	30.1	S10	Q58β($N^{\varepsilon 2}$)	20.5
T8($O^{\gamma 1}$)	S812	G97β(N)	35.9	S14	S95α(O)	19.0
				S19	D32β($O^{\delta 1}$)	18.7

5.3 Discussion

The diversity of TCR repertoires has been linked to high-avidity recognition of pMHC ligands, effective viral clearance, and prevention of viral escape. Previous structural studies of public T cell responses to immunodominant viral epitope have been largely restricted to EBV-specific TCRs (Liu et al., 2013; Tynan et al., 2005; Kijer-Nielsen et al., 2003; Rossjohn et al., 2015). The two TCR-NLV-HLA-A2 complexes described here provide insights into the structural basis for the diverse T cell response against CMV in HLA-A2+ individuals (Yang et al., 2015). EBV-specific TCRs recognize a bulged peptide presented by HLA-B8 using two distinct binding modes: one in which the TCR straddles the bulged peptide but makes few contacts with MHC (Tynan et al., 2005), and one in which the TCR is positioned toward the N-terminal end of the HLA-B8 peptide-binding groove, thereby bypassing the bulged peptide (Liu et al., 2013). In contrast to EBV, the dominant CMV epitope NLV presented by HLA-A2 features multiple solvent-exposed residues, thereby providing multiple TCR anchor points. Such feature has resulted much more diverse CMV-specific T cell response than EBV T cell response regardless both virus are ubiquitous and persistent pathogens.

Influenza A virus cause acute infections and is rapidly cleared by host immune system. The T cell response to the dominant GIL epitope is heavily biased toward certain V(D)J combinations and features a highly conserved CDR3β motif (Moss et al., 1991; Lehner et al., 1995). Our structural studies of two GIL-specific TCRs in complex with GIL-HLA-A2 revealed the structural basis for the narrow GIL-specific repertoire. The previously determined JM22-GIL-HLA-A2 structure showed several advantages of the TRBV19 Vβ region and 'RSS' CDR3β motif in recognizing the featureless GIL-peptide: 1. Arg98 in Vβ inserts into a notch between the GIL peptide and HLA-A2. 2. TRBV19 TCRs are able to recruit water molecules to TCR-pMHC interface to stabilize TCR-pMHC complex. 3. Vα has very limited contacts with GIL-HLA-A2, thus allowing TRBV19 pairing with multiple Vα to expand the GIL-specific T cell response.

Our structural study of the low-affinity F6 TCR in complex with GIL-HLA-A2 revealed how restriction of CDR3α length and sequence may impact TCR binding. Interestingly, the Vα of F6 is dominantly present in GIL-specific TCR repertoires, suggesting that F6-like TCRs are effective in killing influenza-infected cells. The apparent discrepancy between the low affinity of F6 and its prominence in GIL-specific TCR repertoires can perhaps be explained by the way we measured its affinity. We measured the 3D affinity of the TCR-pMHC interaction by immobilizing pMHC ligand on a chip and flowing the TCRs. In a more physiologically setting, both TCR and pMHC should be immobilized on membranes (e.g. on cells transfected with specific TCR and pMHC). Indeed, in such experiments, TCRs with similar 3D affinities can display surprisingly different 2D affinities and TCRs with different 3D affinities may have with similar 2D affinities. Nevertheless, our F6-GIL-HLA-A2 complex structure emphasizes the effectiveness of TRBV19 Vβ in recognizing the featureless GIL peptide. Moreover, TRBV19 Vβ

is able to pair with multiple Vα, as well as different CDR3α motifs, to expand T cell repertoire diversity.

The F50-GIL-HLA-A2 structure provides a wealth of information on how a non-TRBV19 TCR can engage GIL-HLA-A2. The relatively low affinity of F50 compared to JM22 is likely attributable, at least in part, to fewer contacts, in particular polar contacts, across the TCR-pMHC interface. Trp99β of F50 TCR mimics the function of Arg98 β of JM22, but not as effectively. The F50 CDR3α and CDR3β motifs were newly identified by our deep sequencing analysis and may therefore be considered as minor public clonotypes. Nevertheless, F50 TCR serves as a possible candidate to further expand the GIL-specific T cell repertoire and such expansion may improve viral clearance.

Chapter 6

Conclusions and future perspectives

CD8+ T lymphocytes specific for antigenic peptides derived from viral proteins protect us against different viruses (e.g. EBV, CMV, influenza A virus). The diversity of the T cell repertoire for specific viral peptides is critical for effective viral clearance. Previous studies used combinatorial peptide libraries to examine TCR cross-reactivity (Hemmer et al., 1997; Ishizuka et al., 2009; Wooldridge et al., 2012; Birnbaum et al., 2014; Adams et al., 2016). However, a comprehensive analysis of how many TCRs can recognize a single pMHC ligand requires extensive TCR sequence information. Recent advances of deep sequencing and single-cell analysis have revolutionized our understanding of the actual size of T cell repertoires for specific antigenic peptides (Shugay et al., 2014).

CMV and Flu cause chronic and acute infections in human, respectively. The $CD8^+$ T cell response to CMV and Flu has been studied extensively. The CMV matrix protein pp65 accounts for 70–90% of the $CD8^+$ cytotoxic T cell response to this ubiquitous herpesvirus . The dominant epitope in HLA-$A2^+$ subjects corresponds to residues 495–503 of pp65 (NLVPMVATV). In Flu, the dominant epitope for cytotoxic T cells in HLA-$A2^+$ subjects corresponds to residues 58–66 of matrix protein M1 (GILGFVFTL). Characterization of the TCR repertoires elicited by these two dominant viral epitopes has revealed several important features. For NLV-specific TCRs, preferential usage of certain Vβ gene segments is observed in some individuals, but such bias does not seem to be shared by different individuals, suggesting that the NLV-specific TCR repertoire is large and functionally redundant. In contrast, GIL-specific TCRs exhibit more

restricted in V gene usage, with high percentage representations of TRBV27, TRAV12 and TRBV19.

The much higher diversity of NLV- than GIL-specific TCR repertoires observed in multiple studies can likewise be explained in structural terms. We determined structures of two public TCRs in complex with NLV–HLA-A2 (C25-NLV-HLA-A2 and C7-NLV-HLA-A2). These TCRs utilize completely different CDR3α and CDR3β motifs that, in addition, can associate with multiple Vα and Vβ regions in NLV-specific T cell repertoires. This interchangeability of TCR V regions and CDR3 motifs is made possible, at least in part, by the nature of the NLV peptide, which contains several residues, mainly P4 Pro and P5 Met, with solvent-exposed side chains that contribute to the TCR recognition surface. These protruding side chains offer multiple anchor points for TCR attachment, thereby permitting multiple structural solutions to binding the NLV–HLA-A2 ligand and generation of a clonally diverse T cell repertoire. By comparison, the relatively featureless GIL peptide, with its lack of protruding side chains, presents a difficult target for TCR recognition, as illustrated by comparing the JM22/F6-GIL-HLA-A2 and F50–GIL–HLA-A2 structures. The rarity of the TRBV27 β chain expressed by F50 (1.5% of GIL-specific TCRβ sequences) contrasts sharply with the prevalence of the TRBV19 β chain expressed by F6 (80% of sequences), which enables insertion of a highly conserved CDR3β Arg98 residue into a notch on the surface of GIL–HLA-A2. Indeed, the diversity of the GIL-specific TCR repertoire arises in large measure from the pairing of the TRBV27 Vβ region with a variety of Vα regions. This interchangeability is made possible by the dominance of TRBV27 in interactions with pMHC, as seen in the F50–GIL–HLA-A2 complex. The weak affinity of F50 is consistent with the underepresentation of TRBV27 β in the total GIL-specific T cell repertoire. Since most immunodominant viral epitopes are likely to resemble

NLV rather than GIL in terms of side-chain exposure to TCRs, we predict that most such epitopes will elicit highly diverse T cell responses, similar to NLV.

Appendix: Methods and materials

1. List of constructs and *E.coli* stocks.

Name	C7α
Description	Ectodomain of C7 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MILNVEQSPQSLHVQEGDSTNFTCSFPSSNFYALHWYRWETAKSPE ALFVMTLNGDEKKKGRISATLNTKEGYSYLYIKGSQPEDSATYLCA FITGNQFYFGTGTSLTVIPNIQNPDPAVYQLRDSKSSDKSVCLFTDF DSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNKSDFA CANAFNN SIIPEDTFFPSPESS
Protein MW (kDa)	22.9
Protein PI	4.7

Name	C7β
Description	Ectodomain of C7 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MGAGVSQSPSNKVTEKGKDVELRCDPISGHTALYWYRQRLGQGLE FLIYFQGNSAPDKSGLPSDRFSAERTGESVSTLTIQRTQQEDSAVYL CASSQTQLWETQYFGPGTRLLVLEDLKNVFPPEVAVFEPSEAEISHT QKATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQPA LNDSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDR AKPVTQIVSAEAWGRAD
Protein MW (kDa)	27.5
Protein PI	5.3

Name	C25α
Description	Ectodomain of C25 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MDAKTTQPNSMESNEEEPVHLPCNHSTISGTDYIHWYRQLPSQGPE YVIHGLTSNVNNRMASLAIAEDRKSSTLILHRATLRDAAVYYCILD NNNDMRFGAGTRLTVKPNIQNPDPAVYQLRDSKSSDKSVCLFTDF DSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNKSDFA CANAFNNSIIPEDTFFPSPESS
Protein MW (kDa)	22.7
Protein PI	5.0

Name	C25β
Description	Ectodomain of C25 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MGAGVSQSPRYKVTKRGQDVALRCDPISGHVSLYWYRQALGQGP EFLTYFNYEAQQDKSGLPNDRFSAERPEGSISTLTIQRTEQRDSAMY RCASSLAPGTTNEKLFFGSGTQLSVLEDLNKVFPPEVAVFEPSEAEIS HTQKATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQ PALNDSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQ DRAKPVTQIVSAEAWGRAD
Protein MW (kDa)	27.8
Protein PI	5.8

Name	D12α
Description	Ectodomain of D12 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MAQSVAQPEDQVNVAEGNPLTVKCTYSVSGNPYLFWYVQYPNRG LQFLLYITGDNLVKGSYGFEAEFNKSQTSFHLKKPSALVSDSALYFC AVRDISARLMFGDGTQLVVKPNIQNPDPAVYQLRDSKSSDKSVCLF TDFDSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNKSD FACANAF NNSIIPEDTFFPSPESS
Protein MW (kDa)	22.9
Protein PI	4.7

Name	D12β
Description	Ectodomain of D12 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MDAGVIQSPRHEVTEMGQEVTLRCKPISGHDYLFWYRQTMMRGL ELLIYFNNNVPIDDSGMPEDRFSAKMPNASFSTLKIQPSEPRDSAVY FCASSSVNEQFFGPGTRLTVLEDLKNVFPPEVAVFEPSEAEISHTQK ATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQPALN DSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAK PVTQIVSAEAWGRAD
Protein MW (kDa)	27.8
Protein PI	5.8

Name	C31α
Description	Ectodomain of C31 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MGQQLNQSPQSMFIQEGEDVSMNCTSSSIFNTWLWYKQEPGEGPV LLIALYKAGELTSNGRLTAQFGITRKDSFLNISASIPSDVGIYFCAGP MKTSYDKVIFGPGTSLSVIPNIQNPDPAVYQLRDSKSSDKSVCLFTD FDSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNKSDFA CANAFNNSIIPEDTFFPSPESS
Protein MW (kDa)	22.8
Protein PI	4.5

Name	C31β
Description	Ectodomain of C31 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MDAGVIQSPRHEVTEMGQEVTLRCKPISGHNSLFWYRQTMMRGLE LLIYFNNNVPIDDSGMPEDRFSAKMPNASFSTLKIQPSEPRDSAVYF CASSSANYGYTFGSGTRLTVVEDLKNVFPPEVAVFEPSEAEISHTQK ATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQPALN DSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAK PVTQIVSAEAWGRAD
Protein MW (kDa)	27.4
Protein PI	5.4

Name	C32α
Description	Ectodomain of C31 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MGQQLNQSPQSMFIQEGEDVSMNCTSSSIFNTWLWYKQEPGEGPV LLIALYKAGELTSNGRLTAQFGITRKDSFLNISASIPSDVGIYFCAGP MKTSYDKVIFGPGTSLSVIPNIQNPDPAVYQLRDSKSSDKSVCLFTD FDSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNKSDFA CANAFNNSIIPEDTFFPSPESS
Protein MW (kDa)	22.8
Protein PI	4.5

Name	C32β
Description	Ectodomain of C32 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MNAGVTQTPKFQVLKTGQSMTLQCAQDMNHEYMSWYRQDPGM GLRLIHYSVGAGITDQGEVPNGYNVSRSTTEDFPLRLLSAAPSQTSV YFCASSYSGNSGYTFGSGTRLTVVEDLNKVFPPEVAVFEPSEAEISH TQKATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQP ALNDSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQD RAKPVTQIVSAEAWGRAD
Protein MW (kDa)	27.1
Protein PI	5.5

Name	C34α
Description	Ectodomain of C34 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MILNVEQSPQSLHVQEGDSTNFTCSFPSSNFYALHWYRWETAKSPE ALFVMTLNGDEKKKGRISATLNTKEGYSYLYIKGSQPEDSATYLCA FPYNNNDMRFGAGTRLTVKPNIQNPDPAVYQLRDSKSSDKSVCLFT DFDSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNKSDF ACANAFNNSIIPEDTFFPSPESS
Protein MW (kDa)	23.1
Protein PI	4.9

Name	C34β
Description	Ectodomain of C34 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MEAQVTQNPRYLITVTGKKLTVTCSQNMNHEYMSWYRQDPGLGL RQIYYSMNVEVTDKGDVPEGYKVSRKEKRNFPLILESPSPNQTSLY FCASSLEGYTEAFGQGTRLTVVEDLKNVFPPEVAVFEPSEAEISHTQ KATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQPAL NDSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRA KPVTQIVSAEAWGRAD
Protein MW (kDa)	27.1
Protein PI	5.5

Name	F5α
Description	Ectodomain of F5 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MQLLEQSPQFLSIQEGENLTVYCNSSSVFSSLQWYRQEPGEGPVLL VTVVTGGEVKKLKRLTFQFGDARKDSSLHITAAQPGDTGLYLCAG AGSQGNLIFGKGTKLSVKPNIQNPDPAVYQLRDSKSSDKSVCLFTD FDSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNKSDFA CANAFNNSIIPEDTFFPSPESS
Protein MW (kDa)	22.3
Protein PI	4.9

Name	F5β
Description	Ectodomain of F5 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MVDGGITQSPKYLFRKEGQNVTLSCEQNLNHDAMYWYRQDPGQG LRLIYYSQIVNDFQKGDIAEGYSVSREKKESFPLTVTSAQKNPTAFY LCASSSRSSYEQYFGPGTRLTVTEDLKNVFPPEVAVFEPSEAEISHT QKATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQPA LNDSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDR AKPVTQIVSAEAWGRAD
Protein MW (kDa)	27.1
Protein PI	5.5

Name	F6α
Description	Ectodomain of F6 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MQLLEQSPQFLSIQEGENLTVYCNSSSVFSSLQWYRQEPGEGPVLL VTVVTGGEVKKLKRLTFQFGDARKDSSLHITAAQPGDTGLYLC AGAIGSSNTGKLIFGKGTKLSVKPNIQNPDPAVYQLRDSKSSDKSVC LFTDFDSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNK SDFACANAFNNSIIPEDTFFPSPESS
Protein MW (kDa)	22.6
Protein PI	5.0

Name	F6β
Description	Ectodomain of F6 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MVDGGITQSPKYLFRKEGQNVTLSCEQNLNHDAMYWYRQDPGQG LRLIYYSQIVNDFQKGDIAEGYSVSREKKESFPLTVTSAQKNPTAFY LCASSIRSSYEQYFGPGTRLTVTEDLKNVFPPEVAVFEPSEAEISHTQ KATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQPAL NDSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRA KPVTQIVSAEAWGRAD
Protein MW (kDa)	27.8
Protein PI	5.6

Name	F8α
Description	Ectodomain of F8 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MQLLEQSPQFLSIQEGENLTVYCNSSSVFSSLQWYRQEPGEGPVLL VTVVTGGEVKKLKRLTFQFGDARKDSSLHITAAQPGDTGLYLC AGAIGSSNTGKLIFGKGTKLSVKPNIQNPDPAVYQLRDSKSSDKSVC LFTDFDSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNK SDFACANAFNNSIIPEDTFFPSPESS
Protein MW (kDa)	22.6
Protein PI	5.0

Name	F8β
Description	Ectodomain of F8 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MVDGGITQSPKYLFRKEGQNVTLSCEQNLNHDAMYWYRQDPGQG LRLIYYSQIVNDFQKGDIAEGYSVSREKKESFPLTVTSAQKNPTAFY LCASSIRSSYEQYFGPGTRLTVTEDLKNVFPPEVAVFEPSEAEISHTQ KATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQPAL NDSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRA KPVTQIVSAEAWGRAD
Protein MW (kDa)	27.8
Protein PI	5.6

Name	F22α
Description	Ectodomain of F22 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	METNVEQHPSTLSVQEGDSAVIKCTYSDSASNYFPWYKQELGKRP QLIIDIRSNVGEKKDQRIAVTLNKTAKHFSLHITETQPEDSAVYF CAAQGSQGNLIFGKGTKLSVKPNIQNPDPAVYQLRDSKSSDKSVCL FTDFDSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNK SDFACANAFNNSIIPEDTFFPSPESS
Protein MW (kDa)	22.9
Protein PI	5.2

Name	F22β
Description	Ectodomain of F22 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MDGGITQSPKYLFRKEGQNVTLSCEQNLNHDAMYWYRQDPGQGL RLIYYSQIVNDFQKGDIAEGYSVSREKKESFPLTVTSAQKNPTAFYL CASSIRSSYEQYFGPGTRLTVTEDLKNVFPPEVAVFEPSEAEISHTQK ATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQPALN DSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAK PVTQIVSAEAWGRAD
Protein MW (kDa)	27.7
Protein PI	5.6

Name	F26α
Description	Ectodomain of F26 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MPQSVTQLDSQVPVFEEAPVELRCNYSSSVSVYLFWYVQYPNQGL QLLLKYLSGSTLVKGINGFEAEFNKSQTSFHLRKP SVHISDTAEY FCAVGGSQGNLIFGKGTKLSVKPNIQNPDPAVYQLRDSKSSDKSVC LFTDFDSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNK SDFACANAFNNSIIPEDTF FPSPESS
Protein MW (kDa)	23.0
Protein PI	5.0

Name	F26β
Description	Ectodomain of F26 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MDGGITQSPKYLFRKEGQNVTLSCEQNLNHDAMYWYRQDPGQGL RLIYYSQIVNDFQKGDIAEGYSVSREKKESFPLTVTSAQKNPTAFYL CASSIRSSYEQYFGPGTRLTVTEDLKNVFPPEVAVFEPSEAEISHTQK ATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQPALN DSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAK PVTQIVSAEAWGRAD
Protein MW (kDa)	27.7
Protein PI	5.6

Name	F50α
Description	Ectodomain of F50 TCR α chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MGENVEQHPSTLSVQEGDSAVIKCTYSDSASNYFPWYKQELGKRP QLIIDIRSNVGEKKDQRIAVTLNKTAKHFSLHITETQPEDSAVYFCA ASFIIQGAQKLVFGQGTRLTINPNIQNPDPAVYQLRDSKSSDKSVCL FTDFDSQTNVSQSKDSDVYITDKCVLDMRSMDFKSNSAVAWSNKS DFACANAFNNSIIPEDTFFPSPESS
Protein MW (kDa)	23.3
Protein PI	5.1

Name	F50β
Description	Ectodomain of F50 TCR β chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MEAQVTQNPRYLITVTGKKLTVTCSQNMNHEYMSWYRQDPGLGL RQIYYSMNVEVTDKGDVPEGYKVSRKEKRNFPLILESPSPNQTSLY FCASSLLGGWSEAFFGQGTRLTVVEDLKNVFPPEVAVFEPSEAEISH TQKATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQP ALNDSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQD RAKPVTQIVSAEAWGRAD
Protein MW (kDa)	27.8
Protein PI	5.8

Name	HLA-A2 heavy chain
Description	Ectodomain of HLA-A2 heavy chain
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MGSHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQR MEPRAPWIEQEGPEYWDGETRKVKAHSQTHRVDLGTLRGYYNQS EAGSHTVQRMYGCDVGSDWRFLRGYHQYAYDGKDYIALKEDLRS WTAADMAAQTTKHKWEAAHVAEQLRAYLEGTCVEWLRRYLENG KETLQRTDAPKTHMTHHAVSDHEATLRCWALSFYPAEITLTWQRD GEDQTQDTELVETRPAGDGTFQKWAAVVVPSGQEQRYTCHVQHE GLPKPLTLRWE
Protein MW (kDa)	31.9
Protein PI	6.0

Name	β_2-microglobulin
Description	Intact β_2-microglobulin
Plasmid	pET-26b(+)
Expression strain	BL21(DE3)
Antibiotics	Kan^+
Protein sequence	MIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERI EKVEHSDLSFSKDWSFYLLYYTEFTPTEKDEYACRVNHVTLSQPKI VKWDRDM
Protein MW (kDa)	11.9
Protein PI	6.0

Name	NLV peptide
Description	CMV virus dominant epitope
Method	Chemically synthesized
Company	GenScript
Purity	90%
Protein sequence	NLVPMVATV
Protein MW (kDa)	0.94
Protein PI	5.5

Name	GIL peptide
Description	Flu virus dominant epitope
Method	Chemically synthesized
Company	GenScript
Purity	90%
Protein sequence	GILGFVFTL
Protein MW (kDa)	0.97
Protein PI	5.5

2. Molecular cloning

PCR and sub-cloning were performed according to *Molecular Cloning* or product manuals. The genes encoding the ectodomain of all TCR αβ, HLA-A2 heavy chain and β2-microglobulin were amplified via standard PCR protocols and sub-cloned into pET-26b(+) vector via NdeI and XhoI restriction sites. All constructs were verified by DNA sequencing before transforming into BL-21(DE3) strain for protein expression.

3. Freezing bacteria

Single colony of BL-21(DE3) containing the corresponding construct was pre-cultured in 100 mL of LB culture with correct antibiotics for 12 hours at 30 °C. The pre-culture was supplemented to 20% glycerol before storing in a -80°C freezer.

4. Growth of E. coli cultures for expression of inclusion bodies

1) Grow BL-21(DE3) cells with TCR α plasmid in 1 L LB culture until OD reaches 0.6-0.8.
2) Add 1 ml 1M IPTG to the culture to induce recombinant protein expression.
3) Shake the culture 3 or 3.5 hours, then harvest cells by centrifugation (7000 rpm for 15 min) in a Beckman Coulter JLA-8.1000 rotor. The harvested cells can be stored at -80°C.

5. Inclusion body processing

Solutions:

Wash buffer I	*Wash buffer II*	*Urea buffer*
5% Triton X-100	50 mM Tris.HCl, pH 8.0	50 mM Tris.HCl, pH 8.0
50 mM Tris.HCl, pH 8.0	100 mM NaCl	8 M Urea
100 mM NaCl	2 mM EDTA	10 mM EDTA
2 mM EDTA		10 mM DTT

1) Resuspend the cells in 50 mL inclusion wash buffer 1.
2) Sonicatate (2 min sonication and 2min rest for cooling) the suspended cells for 4 cycles and centrifuge at 7000 rpm for 15min. Discard supernatant.
3) Repeat steps 1 and 2 for 2 more times.
4) Resuspend the cells in inclusion wash buffer 2.
5) Sonicate (2 min sonication and 2 min rest for cooling) the suspended cells for 4 cycles and centrifuge at 7000 rpm for 15min. Discard supernatant.
6) Repeat steps 4 and 5 for 1 more time.
7) Dissolve the inclusion bodies in urea buffer overnight at 4 ^{0}C.
8) The dissolved inclusion bodies can be stored at -80 ^{0}C.

6. Refolding TCR and peptide-HLA-A2

Refolding TCR:

1) α and β inclusion bodies were mixed first and then added to the refolding buffer. drop by drop. The final concentration of inclusion bodies is 60-80 mg/L (α 45 mg and β 35mg).
2) Refolding mixture dialyzed against 10 L of DI water for 48 hours.

3) Dialyzed against 10 L of 10mM Tris pH 8 for 24 hours.
4) Concentrate to 50 mL.
5) Dialyze against 2L 50 mM MES pH 6 for 24 hours.
6) Dialyze against 2L 50mM Tris pH 8 for 24 hours.

Refolding peptide-HLA-A2

1) Add 15mg of peptide dissolved in solvent to 1L refolding buffer.
2) HLA-A2 heavy chain and β2-microglobulin inclusion bodies were mixed first and then added to the refolding buffer drop by drop to a final protein concentration of 40 mg/L (HLA-A2 heavy chain 20 mg and β2-microglobulin 20 mg).
3) Refolding mixture dialyzed against 10 L of 10 mM Tris pH 8 for 72 hours with buffer exchange every 12 hours.
4) Concentrate to 50mL.

Refolding buffer

5 M urea

0.4 M Arginine-HCl

100 mM Tris-HCl pH 8.0

3.7 mM Cystamine

6.6 mM Cysteamine

2 mM EDTA

7. Purification of *in vitro* folded TCR

1) Remove aggregates by ultracentrifugation for 30 min at 20,000 rpm using a JA25.5 rotor.

2) Further concentrate TCR solutions to ~5 mL and filter the solution by 0.2 μM membrane.
3) Run gel filtration with Superdex 200 column on an FPLC system. Inject 1 mL of sample each time. Run with Tris-HCl buffer at pH 8. The flow rate was set at 0.5 mL/min. The fractions contained correct proteins were pooled together.
4) Run anion-exchange chromatography with MonoQ column on an FPLC system. Inject 5 mL of sample each time. Run with Tris-HCl buffer at pH 8. Use a NaCl gradient of 1%-per-mL increase to elute proteins. The fractions containing target proteins were pooled together.

8. Purification of *in vitro* folded peptide-HLA-A2

1) Remove aggregates by ultracentrifugation for 30min at 20,000 rpm using a JA25.5 rotor.
2) Further concentrate TCR solutions to ~5 mL and filter the solution by 0.2 μM membrane.
3) Run gel filtration with Superdex 200 column on an FPLC system. Inject 1 mL of sample each time. Run with Tris-HCl buffer at pH 8. The flow rate was set at 0.5 mL/min. The fractions contained correct proteins were pooled together.
4) Run anion-exchange chromatography with MonoQ column on an FPLC system. Inject 5 mL of sample each time. Run with Tris-HCl buffer at pH 8. Use a NaCl gradient of 1%-per-mL increase to elute proteins. The fractions containing target proteins were pooled together.

9. SPR analysis of NLV/GIL-specific TCR binding to NLV/GIL-HLA-A2 using Biacore T100 instrument

Running buffer:

1x PBS

50 mM PBS, pH 7.4

150 mM NaCl

0.05% Surfactant P-20

1) Insert a SA Biacore sensor chip into the chip cassette.
2) Prime the system with the running buffer.
3) Immobilization of NLV/GIL-HLA-A2: dilute the 2 mg/mL NLV/GIL-HLA-A2 solution to 2 μg/mL using 1x PBS buffer. Open manual control and set the flow rate at 10 μL/min. Stop injection when ~1000 RU NLV/GIL-HLA-A2 are immobilized on flow cell 2 or 3. Flow cell 1 is left for blank.
4) Inject of NLV/GIL-specific TCRs: set flow rate at 10 μL/min. For a binding cycle of equilibrium measurement, inject each concentration of TCR for 120-180 sec. Dissociation time is set to 60 sec. The highest concentration should be more than 10 times the estimated K_D. Do 1:2 serial dilutions with PBS for 6-8 more concentrations.
5) Fit the equilibrium RU with concentration and obtain K_D via BiaEvaluation 3000 software.

10. Protein analysis

10.1 SDS-PAGE to check the purity of protein samples

1) Prepare non-reduced samples by mixing 2 μL protein and 2 μL 2x Laemmli sample buffer. For preparing reduced samples, add an extra 0.5 μL 1 M DTT.
2) Heat samples at 100 ^{0}C for 5 min.
3) Carefully load 4 μL sample into each well of the sample applicator.
4) Set up a 20% homogeneous PhastGel on the PhastSystem unit. Insert the buffer strips. Slide in the sample applicator.
5) Close the lid and run the preset program for 20% PhastGel.

Ethanol-based fast Coomassie staning:

Coomassie blue stain

50% Ethanol	50 mL	
0.25% Coomassie R250	0.31 g	
40% H_2O	40 mL	Total
10% Acetic acid	10 mL	100 mL

Mix everything else before adding acetic acid. Filter to remove undissolved chemicals.

Fixing solution

50% Ethanol	250 mL	
7% Acetic acid	35 mL	Total
43% H_2O	215 mL	500 mL

Stain/Destain solution

5% Ethanol	25 mL	
7.5% Acetic acid	37.5 mL	Total
87.5% H_2O	447.5 mL	500 mL

Preserving solution

5% Ethanol	50 mL	
10% Acetic acid	100 mL	Total
85% H_2O	850 mL	1 L

1) Make the staining solution by adding 1 mL Coomassie blue stain to 100 mL Stain/Destain solution.
2) Add 100 mL fixing solution to cover gel.
3) Microwave both for 1min.
4) Transfer the gel into the staining solution and put on a rocking shelf.
5) Bands can be seen after 10 min and will become darker if staining is continued.
6) Destain if needed with the Stain/Destain solution. Or directly transfer into preserving solution for destaining and preserving.

10.2 UV spectrometry to determine protein concentration

1) Use the buffer of the protein as blank.
2) Dilute the protein samples X times with same buffer until the measured A280 is in the range of 0.1 to 1.
3) Check the value of absorbance parameters for each protein and calculate the concentration of a protein in mg/mL or M using the following two equations:

$$[protein](mg/mL) = A280 \times X \times (mg/mL\ per\ OD280)$$
$$[protein](\mu M) = A280 \times X \times 106 \div Extinction\ Coefficient$$

11. Co-crystallization of TCR and peptide-HLA-A2

1) Concentrate both TCR and peptide-HLA-A2 to 10 mg/mL.
2) Mix TCR and peptide-HLA-A2 at a 1:1 molar ratio.

3) Set up sitting-drop crystallization trials using the available commercial crystal screening kits, including Wizard I, II, III, IV (Rigaku system), Index and PEG/Ion screens. Each well from a 96 well plate received 80 μL of crystallization solution; 0.2 μL of protein solution was mixed with 0.2 μL crystallization solution using a Mosquito robot.
4) Seal the wells using cover slips and store crystal plates at room temperature in a disturbance-free place.
5) Check the plates for crystal formation every 2 days.

12. Optimization of crystallization conditions

Since only the C7-NLV-HLA-A2 complex crystallized using commercial screens, we developed in-house crystal screens based on previously reported TCR-pMHC crystallization conditions. To date, most TCR-pMHC complex crystals have been grown from PEG 3000-8000 and pH 6-8. Thus, we set matrices to exhaustively array one PEG at 20% concentration (v/v) with eight pH conditions from 6-8.5 and eleven additives (salts). Specifically, C25-NLV-HLA-A2 crystals were grown in conditions containing 20% PEG 3000, imidazole (pH8.0) 0.1 M and 0.2M $Ca(CH_3COO)_2$. F6-GIL-HLA-A2 crystals were grown in 20% PEG 3350 and imidazole (pH8.0) 0.1 M. F50-GIL-HLA-A2 crystals were grown in 20% PEG 3350, imidazole (pH8.0) 0.1 M and 0.2M $C_3H_2O_4Na_2$.

PEG 3000 20% (v/v)

Additives (0.2 M)

Buffer pH (0.1M)		$Ca(CH_3COO)_2$	KHCOO	KBr	NH_4 CH_3COO	$C_3H_2O_4Na_2$	$MgCl_2$	$CaCl_2$	Li_2SO_4	NH_4SO_4	NaCl	LiCl
Cacodylate (pH6.0)	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Cacodylate (pH6.5)	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
MES (pH 6.5)	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
HEPES (pH7.0)	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
HEPES (pH7.5)	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
Imidazole (pH 8.0)	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Tris-HCl (pH 8.0)	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
Tris-HCl (pH 8.5)	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12

PEG 3350 20% (v/v)

Additives (0.2 M)

Buffer pH (0.1M)		$Ca(CH_3COO)_2$	KHCOO	KBr	NH_4 CH_3COO	$C_3H_2O_4Na_2$	$MgCl_2$	$CaCl_2$	Li_2SO_4	NH_4SO_4	NaCl	LiCl
Cacodylate (pH6.0)	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Cacodylate (pH6.5)	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
MES (pH 6.5)	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
HEPES (pH7.0)	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
HEPES (pH7.5)	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
Imidazole (pH 8.0)	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Tris-HCl (pH 8.0)	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
Tris-HCl (pH 8.5)	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12

PEG 4000 20% (v/v)

Additives (0.2 M)

Buffer pH (0.1M)		$Ca(CH_3COO)_2$	KHCOO	KBr	NH_4 CH_3COO	$C_3H_2O_4Na_2$	$MgCl_2$	$CaCl_2$	Li_2SO_4	NH_4SO_4	NaCl	LiCl
Cacodylate (pH6.0)	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Cacodylate (pH6.5)	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
MES (pH 6.5)	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
HEPES (pH7.0)	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
HEPES (pH7.5)	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
Imidazole (pH 8.0)	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Tris-HCl (pH 8.0)	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
Tris-HCl (pH 8.5)	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12

PEG 6000 20% (v/v)

Additives (0.2 M)

Buffer pH (0.1M)		$Ca(CH_3COO)_2$	KHCOO	KBr	NH_4 CH_3COO	$C_3H_2O_4Na_2$	$MgCl_2$	$CaCl_2$	Li_2SO_4	NH_4SO_4	NaCl	LiCl
Cacodylate (pH6.0)	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Cacodylate (pH6.5)	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
MES (pH 6.5)	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
HEPES (pH7.0)	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
HEPES (pH7.5)	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
Imidazole (pH 8.0)	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Tris-HCl (pH 8.0)	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
Tris-HCl (pH 8.5)	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12

PEG 8000 20% (v/v)

Additives (0.2 M)

Buffer pH (0.1M)		$Ca(CH_3COO)_2$	KHCOO	KBr	NH_4 CH_3COO	$C_3H_2O_4Na_2$	$MgCl_2$	$CaCl_2$	Li_2SO_4	NH_4SO_4	NaCl	LiCl
Cacodylate (pH6.0)	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Cacodylate (pH6.5)	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
MES (pH 6.5)	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
HEPES (pH7.0)	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
HEPES (pH7.5)	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
Imidazole (pH 8.0)	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Tris-HCl (pH 8.0)	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
Tris-HCl (pH 8.5)	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12

13.Flash freezing and storage of protein crystals

1) Make 100 μL of cryoprotectant solution containing all the ingredients of crystallization solution plus 20% glycerol.
2) Fill a dewar with liquid nitrogen and immerse a CryoPuck in liquid nitrogen slowly.
3) Select a right size of crystal loop to fish out target crystals and transfer crystals into cryoprotectant solutions for a few seconds.
4) Transfer the cryo-protected crystals into liquid nitrogen immersed CryoPuck.
5) After filling the Puck space with 16 crystals, cap the Puck with a Puck Wand.
6) Insert the CryoPuck into a CryoPuck Carrier and put the carrier back in a dry shipper filled with liquid nitrogen.
7) The dry shipper is then ready to ship to synchrotron facilities.

14. X-ray data collection and processing

All X-ray diffraction data were collected at synchrotron facilities. C25-NLV-HLA-A2 and C7-NLV-HLA-A2 datasets were collected at beamline 22ID of the Advanced Photon Source (APS), Argonne National Laboratory with a MAR 300 CCD detector. F6-GIL-HLA-A2 diffraction datasets were collected at beamline 24ID-E of the Advanced Photon Source, Argonne National Laboratory with an ADSC Q315 CCD detector. F50-GIL-HLA-A2 diffraction datasets were collected at beamline 12-2 of the Stanford Synchrotron Radiation Lightsource (SSRL) with an MARmosaic 325 CCD detector. Data processing was performed with the program HKL-2000 and scaled to 2.1 Å, 3.5 Å, 2.1 Å and 1.7 Å resolutions.

15. Molecular replacement and structure refinement

Molecular replacement was performed using the program Phenix. Details of search steps are discussed in previous chapters. A correct solution should lead to an R_{free} of less than 50%. Structure refinement was carried out by Phenix initially, followed by manual adjustment. Details and statistics of refinement are discussed in previous chapters. All complex structures gave reasonable R_{work} and R_{free} in the end.

16. Structure analysis and figure preparation

Solvent-accessible surface areas were calculated using the program PISA with a probe radius of 1.4 Å. Contacts were identified by CONTACT using a cut-off distance of 4.0 Å. All structure figures were prepared using PyMOL, a powerful and versatile molecular graphics program written by Warren L. DeLano.

Bibliography

Adams, E.J., Strop, P., Shin, S., Chien, Y.H., Garcia, K.C. (2008). An autonomous CDR3delta is sufficient for recognition of the nonclassical MHC class I molecules T10 and T22 by gammadelta T cells. Nat. Immunol. *9*, 777-784.

Adams, J.J., Narayanan, S., Birnbaum, M.E., Sidhu, S.S., Blevins, S.J., Gee, M.H., Sibener, L.V., Baker, B.M., Kranz, D.M., and Garcia, K.C. (2016). Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. Nat. Immunol. *17*, 87-94.

Arimilli, S., Cardoso, C., Mukku, P., Baichwal, V., Nag, B. (1995). Refolding and reconstitution of functionally active complexes of human leukocyte antigen DR2 and myelin basic protein peptide from recombinant alpha and beta polypeptide chains. J. Biol. Chem. *270*, 971-977.

Beringer, D.X., Kleijwegt, F.S., Wiede, F., van der Slik, A.R., Loh, K.L., Petersen, J., Dudek, N.L., Duinkerken, G., Laban, S., Joosten, A., Vivian, J.P., Chen, Z.,Uldrich, A.P., Godfrey, D.I., McCluskey, J., Price, D.A., Radford, K.J., Purcell, A.W., Nikolic T, Reid HH, Tiganis T, Roep BO, Rossjohn J. (2015). T cell receptor reversed polarity recognition of a self-antigen major histocompatibility complex. *16*, 1153-1161.

Birnbaum, M.E., Mendoza, J.L., Sethi, D.K., Dong, S., Glanville, J., Dobbins, J., Ozkan, E., Davis, M.M., Wucherpfennig, K.W., Garcia, K.C. (2014). Deconstructing the peptide-MHC specificity of T cell recognition. Cell *157*, 1073-1087.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., Wiley, D.C. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. Nature *329*, 506-512.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., Wiley, D.C. (1987). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature *329*, 512-518.

Bolin, D.R., Swain, A.L., Sarabu, R., Berthel, S.J., Gillespie, P., Huby, N.J., Makofske, R., Orzechowski, L., Perrotta, A., Toth, K., Cooper, J.P., Jiang, N., Falcioni, F., Campbell, R., Cox, D., Gaizband, D., Belunis, C.J., Vidovic, D., Ito, K., Crowther, R., Kammlott, U., Zhang, X., Palermo, R., Weber, D., Guenot, J., Nagy, Z., Olson, G.L. (2000). Peptide and peptide mimetic inhibitors of antigen presentation by HLADR class II MHC molecul es. Design,structure-activity relationships, and X-ray crystal structures. J. Med. Chem. *43*, 2135-2148.

Boulter, J.M., Glick, M., Todorov, P.T., Baston, E., Sami, M., Rizkallah, P., Jakobsen, B.K. (2003). Stable, soluble T-cell receptor molecules for crystallization and therapeutics. Protein. Eng. *16*, 707-711.

Brownlie, R.J., Zamoyska, R. (2013). T cell receptor signalling networks: branched, diversified and bounded. Nat. Rev. Immunol. *13*, 257-269.

Cole, D.K., Pumphrey, N.J., Boulter, J.M., Sami, M., Bell, J.I., Gostick, E., Price, D.A., Gao, G.F., Sewell, A.K., and Jakobsen, B.K. (2007). Human TCR-binding affinity is governed by MHC class restriction. J. Immunol. *178*, 5727-5734.

Day, E.K., Carmichael, A.J., ten Berge, I.J., Waller, E.C., Sissons, J.G., and Wills, M.R. (2007). Rapid $CD8^+$ T cell repertoire focusing and selection of high-affinity clones into memory following primary infection with a persistent human virus: human cytomegalovirus. J. Immunol. *179*, 3203-3213.

Ding, Y.H., Baker, B.M., Garboczi, D.N., Biddson, W.E., Wiely, D.C. (1999). Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. Immunity *11*, 45-56.

Fremont, D.H., Hendrickson, W.A., Marrack, P., Kappler, J. (1996).Structures of an MHC class II molecule with covalently bound single peptides. Science *272*, 1001-1004.

Garboczi, D.N., Hung, D.T., Wiley, D.C. (1992). HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with single antigenic peptides. Proc. Natl. Acad. Sci. *89*, 3429-3433.

Garcia, K.C., Degano, M., Stanfield, R.L., Brunmark, A., Jackson, M.R., Peterson, P.A., Teyton, L, Wilson, I.A. (1996). An αβ T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. Science *274*, 209–219.

Gil, A., Yassai, M.B., Naumov, Y.N., and Selin, L.K. (2015). Narrowing of human influenza A virus-specific T cell receptor α and β repertoires with increasing age. J. Virol. *89*, 4102-4116.

Gotch, F., Rothbard, J., Howland, K., Townsend, A., and McMichael, A. (1987). Cytotoxic T lymphocytes recognize a fragment of influenza virus matrix protein in association with HLA-A2. Nature *326*, 881-882.

Gras, S., Saulquin, X., Reiser, J.B., Debeaupuis, E., Echasserieau, K., Kissenpfennig, A., Legoux, F., Chouquet, A., Le Gorrec, M., Machillot, P., et al. (2009). Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. J. Immunol. *183*, 430-437.

Griffiths, P., Baraniak, I., and Reeves, M. (2015). The pathogenesis of human cytomegalovirus. J. Pathol. *235*, 288-297.

Ishizuka, J., Stewart-Jones, G.B., van der Merwe, A., Bell, J.I., McMichael, A.J., and Jones, E.Y. (2008). The structural dynamics and energetics of an immunodominant T cell receptor are programmed by its Vβ domain. Immunity *28*, 171-182.

Kjer-Nielsen, L., Clements, C.S., Purcell, A.W., Brooks, A.G., Whisstock, J.C., Burrows, S.R., McCluskey, J., Rossjohn, J. (2003). A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. Immunity *18*, 53-64.

Keck, Z.Y., Girard-Blanc, C., Wang, W., Lau, P., Zuiani, A., Rey, F.A., Krey, T., Diamond, M.S., Foung, S.K. (2016). Antibody response to hypervariable region 1 interferes with broadly neutralizing antibodies to hepatitis C virus. J. Virol. *90*, 3112-3122.

Koning, D., Costa, A.I., Hasrat, R., Grady, B.P., Spijkers, S., Nanlohy, N., Kesmir, C., and van Baarle, D. (2014). In vitro expansion of antigen-specific $CD8^+$ T cells distorts the T-cell ρεπερτοιρε. Θ. Ιμμυνολ. Μετηοδσ *405*, 199–203.

La Gruta, N.L., and Turner, S.J. (2014). T cell mediated immunity to influenza: mechanisms of viral control. Trends Immunol. *35*, 396-402.

Lawrence, M.C., and Colman, P.M. (1993). Shape complementarity at protein/protein interfaces. J. Mol. Biol. *234*, 946-950.

Lawson, T.M., Man, S., Williams, S., Boon, A.C., Zambon, M., Borysiewicz, L.K. (2001). Influenza A antigen exposure selects dominant Vβ17+ TCR in human CD8+ cytotoxic T cell responses. Int. Immunol. *13*, 1373–1381.

Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Jabado-Michaloud, J., Folch, G., Bellahcene, F., Wu, Y., Gemrot, E., Brochet, X., Lane, J., Regnier, L., Ehrenmann, F., Lefranc, G., Duroux, P., 2009. IMGT®, the international ImMunoGeneTics information system®. Nucleic Acids Research. *37*, 1006-1012.

Lehner, P.J., Wang, E.C., Moss, P.A., Williams, S., Platt, K., Friedman, S.M., Bell, J.I., Borysiewicz, L.K. (1995). Human HLA-A0201-restricted cytotoxic T lymphocyte recognition of influenza A is dominated by T cells bearing the Vβ17 gene segment. J. Exp. Med. *181*, 79–91.

Li, H.M., Hiroi, T., Zhang, Y., Shi, A., Chen, G., De, S., Metter, E.J., Wood, W.H., III, Sharov, A., Milner, J.D., et al. (2015). TCRβ repertoire of $CD4^+$ and $CD8^+$ T cells is distinct in richness, distribution, and CDR3 amino acid composition. J. Leukoc. Biol. *99*, 505-513.

Li, Y., Li, H., Martin, R., Mariuzza, R.A. (2000). Structural basis for the binding of an immunodominant peptide from myelin basic protein in different registers by two HLA-DR2 proteins. J. Mol. Biol. *304*, 177-188

Madden, D.R., Garboczi, D.N., Wiley, D.C. (1993).The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. *75*, 693-708.

Matthews, A.G., Oettinger, M.A. (2009). RAG: a recombinase diversified. Nat. Immunol. *10*, 817-821.

Moss, P.A., Moots, R.J., Rosenberg, W.M., Rowland-Jones, S.J., Bodmer, H.C., McMichael, A.J., and Bell, J.I. (1991). Extensive conservation of α and β chains of the human T-cell antigen receptor recognizing HLA-A2 and influenza A matrix peptide. Proc. Natl. Acad. Sci. USA *88*, 8987-8990.

Neller, M.A., Ladell, K., McLaren, J.E., Matthews, K.K., Gostick, E., Pentier, J.M., Dolton, G., Schauenburg, A.J., Koning, D., Fontaine Costa, A.I., et al. (2015). Naive CD8 T-cell precursors display structured TCR repertoires and composite antigen-driven selection dynamics. Immunol. Cell Biol. *93*, 625-633.

Nguyen, T.H., Rowntree, L.C., Pellicci, D.G., Bird, N.L., Handel, A., Kjer-Nielsen, L., Kedzierska, K., Kotsimbos, T.C., and Mifsud, N.A. (2014). Recognition of distinct cross-reactive virus-specific $CD8^+$ T cells reveals a unique TCR signature in a clinical setting. J. Immunol. *192*, 5039-5049.

Nikolich-Zugich, J., Slifka, M.K., and Messaoudi, I. (2004). The many important facets of T-cell repertoire diversity. Nat. Rev. Immunol. *4*, 123-132.

Peggs, K., Verfuerth, S., Pizzey, A., Ainsworth, J., Moss, P., and Mackinnon, S. (2002). Characterization of human cytomegalovirus peptide-specific $CD8^+$ T-cell repertoire diversity following in vitro restimulation by antigen-pulsed dendritic cells. Blood *99*, 213-223.

Piepenbrink, K.H., Blevins, S.J., Scott, D.R., Baker, B.M. (2013). The basis for limited specificity and MHC restriction in a T cell receptor interface. Nat. Commun. *4*, 1-9.

Qi, Q., Liu, Y., Cheng, Y., Glanville, J., Zhang, D., Lee, J.Y., Olshen, R.A., Weyand, C.M., Boyd, S.D., and Goronzy, J.J. (2014). Diversity and clonal selection in the human T-cell repertoire. Proc. Natl. Acad. Sci. USA *111*, 13139-13144.

Robins, H.S., Srivastava, S.K., Campregher, P.V., Turtle, C.J., Andriesen, J., Riddell, S.R., Carlson, C.S., and Warren, E.H. (2010). Overlap and effective size of the human $CD8^+$ T cell receptor repertoire. Sci. Transl. Med. *2*, 47-64.

Rossjohn, J., Gras, S., Miles, J.J., Turner, S.J., Godfrey, D.I., and McCluskey, J. (2015). T cell antigen receptor recognition of antigen-presenting molecules. Annu. Rev. Immunol. *33*, 169-200.

Rudolph, M.G., Stanfield, R.L., and Wilson, I.A. (2006). How TCRs bind MHCs, peptides, and coreceptors. Annu. Rev. Immunol. *24*, 419-466.

Schodin BA, Schlueter CJ, Kranz DM. (1996). Binding properties and solubility of single-chain T cell receptors expressed in *E. coli*. Mol. Immunol. *33*, 819-829.

Shugay, M., Britanova, O.V., Merzlyak, E.M., Turchaninova, M.A., Mamedov, I.Z., Tuganbaev, T.R., Bolotin, D.A., Staroverov, D.B., Putintseva, E.V., Plevova, K., et al. (2014). Towards error-free profiling of immune repertoires. Nat. Methods *11*, 653-655.

Stern, L.J., Wiley, D.C. (1994). Antigenic peptide binding by class I and class II histocompatibility proteins. Structure *2*, 245-251.

Stewart-Jones, G.B., McMichael, A.J., Bell, J.I., Stuart, D.I., and Jones, E.Y. (2003). A structural basis for immunodominant human T cell receptor recognition. Nat. Immunol. *4*, 657-663.

Swanson, P.C. 2004. The bounty of RAGs: recombination signal complexes and reaction outcomes. Immunol Rev. *200*, 90-114

Tikhonova, A.N., Van Laethem, F., Hanada, K., Lu, J., Pobezinsky, L.A., Hong, C., Guinter, T.I., Jeurling, S.K., Bernhardt, G., Park, J.H., Yang, J.C., Sun, P.D., and Singer A. (2012). αβ T cell receptors that do not undergo major histocompatibility complex-specific thymic selection possess antibody-like recognition specificities. Immunity *36*, 79-91.

Trautmann, L., Rimbert, M., Echasserieau, K., Saulquin, X., Neveu, B., Dechanet, J., Cerundolo, V., and Bonneville, M. (2005). Selection of T cell clones expressing high-affinity public TCRs within human cytomegalovirus-specific CD8 T cell responses. J. Immunol. *175*, 6123-6132.

Tillman, R.E., Wooley, A.L., Hughes, M.M., Khor, B., Sleckman, B.P. 2004. Regulation of T-cell receptor beta-chain gene assembly by recombination signals: the beyond 12/23 restriction. Immunol. Rev. *200*, 36-43.

Tynan, F.E., Reid, H.H., Kjer-Nielsen, L., Miles, J.J., Wilce, M.C., Kostenko, L., Borg, N.A., Williamson, N.A., Beddoe ,T., Purcell, A.W., Burrows, S.R., McCluskey, J., Rossjohn, J. (2007). A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. *8*, 268-276.

Venturi, V., Price, D.A., Douek, D.C., and Davenport, M.P. (2008). The molecular basis for public T-cell responses? Nat. Rev. Immunol *8*, 231-238.

Wang, G.C., Dash, P., McCullers, J.A., Doherty, P.C., and Thomas, P.G. (2012). T cell receptor αβ diversity inversely correlates with pathogen-specific antibody levels in human cytomegalovirus infection. Sci. Transl. Med. *4*, 128ra142.

Wills, M.R., Carmichael, A.J., Mynard, K., Jin, X., Weekes, M.P., Plachter, B., and Sissons, J.G. (1996). The human cytotoxic T-lymphocyte (CTL) response to cytomegalovirus is dominated by structural protein pp65: frequency, specificity, and T-cell receptor usage of pp65-specific CTL. J. Virol. *70*, 7569-7579.

Wooldridge, L., Ekeruche-Makinde, J., van den Berg, H.A., Skowera, A., Miles, J.J., Tan, M.P., Dolton, G., Clement, M., Llewellyn-Lacey, S., Price, D.A., et al. (2012). A single autoimmune T cell receptor recognizes more than a million different peptides. J. Biol. Chem. *287*, 1168-1177.

Yang, X., Gao, M., Chen, G., Pierce, B.G., Lu, J., Weng, N.P., and Mariuzza, R.A. (2015). Structural basis for clonal diversity of the public T cell response to a dominant human cytomegalovirus epitope. J. Biol. Chem. *290*, 29106-29119.

Zhang, N., and Bevan, M.J. (2011). $CD8^+$ T cells: foot soldiers of the immune system. Immunity *35*, 161-168.

Zhu, J., Yamane, H., Paul, W.E. (2009). Differentiation of effector CD4 T cell populations. Annu. Rev. Immunol. *28*, 445-489