THE MECHANOBIOLOGY OF THE T-CELL RECEPTOR: STRUCTURALLY CONNECTING CATCH BONDS WITH LIGAND BINDING AND TRIGGERING

A Thesis
Presented to
The Academic Faculty

by

Paul Antonio Cárdenas Lizana

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Wallace H. Coulter Department of Biomedical Engineering

Georgia Institute of Technology & Emory University
December 2017

THE MECHANOBIOLOGY OF THE T-CELL RECEPTOR: STRUCTURALLY CONNECTING CATCH BONDS WITH LIGAND BINDING AND TRIGGERING

VOLUME I

by

Paul Antonio Cárdenas Lizana

THE MECHANOBIOLOGY OF THE T-CELL RECEPTOR: STRUCTURALLY CONNECTING CATCH BONDS WITH LIGAND BINDING AND TRIGGERING

Approved by:

Andrés J. García, Committee Chair
School of Mechanical Engineering
Georgia Institute of Technology

Cheng Zhu, Advisor
Wallace H. Coulter Department of
Biomedical Engineering
Georgia Institute of Technology

Arash Grakoui
Department of Medicine, Division of
Infectious Diseases
Emory University

James Gumbart
School of Physics
Georgia Institute of Technology

Jizhong Lou
Institute of Biophysics
Chinese Academy of Sciences

Date Approved: August 17, 2017

Dedicado a Carmen y Alejandro ...

ACKNOWLEDGEMENTS

Paul Cárdenas Lizana.

Georgia Institute of Technology.

Georgia, Atlanta, USA, July 2017.

TABLE OF CONTENTS

VOLUME I

LIST OF TABLES

LIST OF FIGURES

SUMMARY

There are continually strong demands in understanding the principles for molecular recognition of antigens and activation of T cells because of their broad implications in immunology and the benefits in cancer immunotherapy. Antigen processing and presentation was extensively studied; however, how TCR recognizes pMHC and then activates the signaling machinery is not understood. This question has puzzled immunologist for decades and understanding if there exists a mechanism genetically encoded to propagate information read from the TCR distal-membrane binding site to inside the cell is considered the "holy grail of molecular immunology". It is believed that TCRs mediate T cell activation by connecting their dimeric $\alpha\beta$ chains with CD3$\epsilon\gamma$, CD3$\epsilon\delta$, and CD3$\zeta\zeta$ subunits using an "unknown mechanism" and the Cβ FG loop in TCRs is regarded as very important since its removal severely affects T cell activation and development.

The first part of this thesis studies the force response of a TCR reading out a pMHC in the context of viral infection and explained how information encoded in the peptide is decoded by the TCR. For the first time, it is demonstrated that TCRs are capable of forming *"in silico"* catch bonds as they do in experiments. It is explained the characteristics and requirements of catch bonds in antigen recognition and how mutant epitopes abolish it. It is shown that TCR modulates the peptide conformation by using highly conserved residues in MHCs. Finally, it is proposed the molecular mechanism of how the TCR reads a viral featureless pMHCs: "RA14 uses a molecular lever to form an *in silico* catch bond and recognize antigenic peptides".

The second part established the principles for molecular recognition of antigens and the mechanical activation of signaling domains by the TCR. It is proposed that the decoding process and early intracellular signaling are connected by conformational changes in

the TCR. The TCR interaction with either the pMHC and CD3 domains are hardwired on its structural dynamics. TCRs are not rigid entities, as currently thought, but they are deformable proteins and mechanical forces modulate their conformations. The ability of the TCR to deform without releasing the pMHC and to transform linear force into rotational torque is the key to understand this molecular mechanism. Lastly, it is described the molecular mechanism of how TCRs use the Cβ FG loop in contacting and potentially activating the CD3 signaling domains.

CHAPTER I

SPECIFIC AIMS AND HYPOTHESES

1.1 *Objetive*

we proposed to study at the molecular level the mechanobiology of the T-cell receptor (TCRs) by using *"in silico"* biology. The TCR is an integral part of the immune system and is a complex protein that has multiple domains. The TCR extracellular domain or ectodomain binds multiple peptide ligands in the context of major histocompatibility complex (MHC) molecules. Free molecular dynamics (MD), targeted (TMD), and steered (SMD) simulations are employed to investigate the TCR conformational changes and its ligand dissociation in the presence of mechanical forces. The simulation is driven and motivated by the need to understand the intricate experimental studies.

TCR interacts with peptide-MHC (pMHC) to discriminate pathogens from self-antigens and trigger adaptive immune responses. The sustained interest in TCR-pMHC interaction stems from a fundamental hypothesis that their interaction parameters have a central role in determining the subsequent T-cell response. Using computer simulations, the TCR-pMHC interaction is modeled and the role of catch bond and epitope recognition are investigated in the context of viral infection and autoimmune disease.

1.2 *Aims*

1.2.1 Aim 1: Describe the molecular mechanism of how TCRs recognize featureless viral pMHC landscapes and how peptide mutations decrease immune recognition in the context of human infection.

The scientific question is how TCRs decode information in featureless pMHC landscape presenting viral epitopes and produce a functional binding. The first hypothesis is that "Force induces exposure of the buried residues in the MHC to enable contact with the TCR

and generate new long-lived hydrogen bonds". The goal is to understand how the TCR exposes these residues and modifies/modulates the conformational states of the epitope. Is force necessary during TCR scanning in the context of viral infection? We will also study how force modulates the peptide flexibility and conformational states, which could potentially regulate the binding affinity of TCR and pMHC. Viral persistence is linked in part to the rise of "viral mutations" that have decreased functional effects compared to those of the wild type (WT), thereby decreasing recognition by the immune system.

The second hypothesis is that viral persistence is characterized by converting "catch-slip" into "slip-only" bond by mutant epitopes of antigen specific TCR-pMHC interactions. We will investigate atomistically the role of converting catch into slip-only bonds and its role in peptide recognition for the escaping mutations in the context of viral infection. We will describe the molecular mechanisms for this change in bond type and correlate the catch-slip conversion to the loss of function of epitopes previously recognized by the TCR. This Aim explains how persistent viral infection employ mechanically-regulated pathways to escape immune recognition.

1.2.2 Aim 2: Identify the molecular pathway of how peptide ligand recognition by TCRs is propagated to the CD3 signaling domains to induce T-cell activation

The question that has puzzled immunologist for decades is that if there exists a preconceived traveling pathway genetically encoded in the TCR genes to propagate information read from its distal-membrane binding site to activate the CD3 signaling domains. First, the general rules for molecular recognition of antigens and activation of CD3 signaling domains by the TCR are investigated. Different TCRs are used to study how catch bonds relate to TCR specificity and sensitivity and how TCRs employ catch bonds for molecular recognition of antigens and activation. The first hypothesis is that the weak interaction between the TCR and pMHC becomes strong in the presence of force to produce an *"in silico"* catch bond and activate the T cells.

Second, we will study how the decoding process and intra-cellular signaling are connected and if they involve conformational changes in the TCR domains. The second hypothesis is that "in the presence of force, a cognate peptide presented to TCR induces conformational changes in the TCR domains that are away from its membrane-distal binding site, and includes the opening/closing of the FG loop and the extension (or unfolding) of connecting peptides (CP) –the membrane-proximal stalks. The rationale is that an open C_β FG loop contacts the upper surface of CD3$\epsilon\gamma$ and push it down; then it is pull up the CD3 domains that in turns expose its ITAMs for phosphorylation and the activation of subsequent signaling cascade. In addition, the extension of the CPs expose a highly conserved sequence of residues located on the α-chain stalk, that is importantly functional.

Currently, the rules for recognition and activation are not well understood; there are many parameters when defining a successful engagement of TCR-pMHC, *i.,e.;* specificity, sensitivity, etc. The question is what TCR property governs them. Is the TCR dynamical structure? Sensitivity and specificity refer to the TCR's ability to correctly detect an antigenic peptide in a sea of unrelated pMHC.

CHAPTER II

BACKGROUND AND SIGNIFICANCE

The immune system is composed of innate and adaptive responses and their fundamental role is to protect the body and fight against infectious organisms and tumor cells. The first line of defense is mediated by the innate immune system between hours while the adaptive response develops over a longer period of time, usually days [1]. The adaptive response is divided into antibody-mediated (also called humoral) and cell-mediated immunity with the major players being the lymphocytes. Lymphocytes originate in the bone marrow from lymphoid progenitor [2] and either remain there and mature into B cells, or migrate to the thymus, where they undergo development to become T lymphocytes.

Lymphocytes residing in the thymus are called thymocytes and they undertake multiple stages to fully develop a T-cell antigen receptor (TCR), bind weakly self-antigens to be positively selected, and finally commit their lineage to either be CD4$^+$ or CD8$^+$ naïve T cells [3, 4]. Thus, T-cell recognition of antigens is the central event of the adaptive immune response [5, 6], where each mature T cell expresses a unique TCR, which is the only molecule on the T-cell surface that senses micro-organism-derived peptides conjugated with the major histocompatibility (MHC) molecules expressed on the antigen presenting cells (APC) [7].

T cells are distinguished from other lymphocytes by the presence of TCR on their membrane surface and they can be broadly divided into Helper, Cytotoxic, Memory, Regulatory, $\gamma\delta$, and Natural Killer T cells. Fully maturated, naïve T cells circulate in the lymph nodes or spleen to patrol for antigens expressed on APCs, survival signals (*i.e.*, immunological tolerance), and activation cues. When T cells encounter a foreign antigen, they become effector T cells and undergo clonal expansion. The effector T cells will function as either releasing

4

cytotoxic granule (CD8$^+$ T cells) to kill the specific cells directly or release cytokines to trigger other immune cells like B cells for help (CD4$^+$ T cells). After their activity, some T cells acquire memory, they migrate and reside in the body, and they are able to remember and respond strongly if they meet the same invader. TCRs have exquisite sensitivity and specificity and their recognition capabilities are very important in both thymic development and T cell activation [8, 2].

Cytotoxic T Lymphocytes (CTLs, CD8$^+$ T cells) are an important type of T cells that express on their surface the CD8$^+$ coreceptor, are capable of killing infected cells, and inter-act with a peptide bound in major histocompatibility (pMHC) molecule at the cell surface on APCs. Similarly, T helper cells also play a fundamental role in the adaptive responses as they help to activate other immune cells by releasing cytokines [8, 2]. They express on their surface the CD4 coreceptor and thus are commonly called CD4$^+$ T cells. CD4$^+$ T cells assist in suppressing or down regulating immune responses. They are essential in B cell antibody class switching, in the activation and growth of cytotoxic T cells, and in max-imizing the bactericidal activity of phagocytes such as macrophages [8, 2]. In general, the T cell activation and effector functions rely mainly on the recognition of a pMHC molecule at the cell surface by TCR and the engagements of its coreceptors.

Understanding how the TCR in the immune system accomplishes its biological func-tion is very critical for organ transplants and treatment of immunodeficiency, autoimmune diseases, persistent viral infection, and cancer since it is shown that they are the result of malfunctioning and misunderstanding of T cell immunity [9, 10, 11].

2.1 General Structure of the $\alpha\beta$ T cell receptor

Although the $\alpha\beta$ TCR is the most widely studied and commonly referred in the literature as the TCR, there is a very small set of T cells expressing a similar TCR formed by γ and δ chains, termed as the $\gamma\delta$ T cells with their biological relevance still highly debatable and not established yet [12, 13]. Most T cells have $\alpha\beta$ TCRs (95%); however, 5% of

5

them express only the $\gamma\delta$ TCR [2]. T cells recognize peptides derived from cell-associated pathogens of viral, bacterial or fungal origin, as well as tumor antigens bound in major histocompatibility (MHC) molecule at the cell surface on antigen presenting cells (APC) through the T cell antigen receptor (TCR) [8, 2]. The antigen-binding TCR is similar in topology and size (height and width) to an Ig Fab fragment. The TCR is a disulfide-linked, membrane-anchored heterodimer consisting of α and β chain [14, 15, 16, 17].

Each α and β chain are composed of one Ig-like variable (V) domain, one Ig-like constant (C) domain, a short connecting peptide just proximal to the membrane, a hydrophobic transmembrane region, and a short cytoplasmic region as shown in Fig 1 (B). The V domains provide clonotypic antigen specificity and vary from one to another T cell, while the C domains are constant or invariant. Ig-like domains consist of two β-sheets formed by antiparallel β-strands each containing an average of 6 to 8 amino acids. The two β-sheets are arranged in a sandwich-like fold to form a hydrophobic core. The β-sheets are further stabilized by a conserved disulfide bond between two cysteines, and an "invariant" hydrophobic residue (W). The Ig fold for V_α and V_β domains have each 115 amino acid residues in average that form four anti-parallel β-strands per β-sheet plus a short β-strand segment across the top of the domain.

The Ig fold for C_β domains is smaller and has about 105 residues, yielding sheets with 4 and 3 β-strands. The V_α and V_β fold consists of A, B, C, C', C", D, E, F, G β-strands, while C_β domain is composed of A, B, C, C', D, E, F, G. The A, B, E strands are on one sheet and G, F, C in the other [15, 18, 14]. The C_α domain diverges from typical Ig C domains sequences [14], having only about 15% sequence identities plus missing many conserved residues; *e.g.*, an "invariant" hydrophobic residue (W) that occurs in the C strand is not present. The C_α domain shows that the A, B, E, and D strands from the back a sheet and are highly similar to the corresponding back sheet of the standard Ig-fold. However, the C, F, and G strands do not form the top β-sheet; they are far from each other and loosely packed against the bottom sheet since the hydrophobic residue (W) used to attach the C strand to

6

the bottom sheet is missing. The C' strand is removed from the classical C-like Ig domain. These differences complicate the classification of the C_α domain into typical Ig-like folds.

The flexible loops in the membrane-distal area of the V_α and V_β domains form the antigen-binding site and are termed the complementary-determining regions (CDRs). They are analogous to the loops used by antibodies to capture antigens; however, as opposed to the concave antigen-binding surface of antibodies, they form a relatively flat binding surface to interact with the pMHC molecule on the APC. There are three CDRs in each V_α and V_β domain. CDR1 and CDR2 are germ line-encoded while CDR3 is somatically rearranged. The CDR3s is the most diverse portion of the TCR and usually contact directly the peptide presented by the MHC. On the other hand, the CDR1 and CDR2 loops mostly recognize parts of MHC that form the peptide-binding groove [12].

2.2   Structure of pMHC molecules

Antigenic peptides are presented to T cells via the heterodimers MHC class I or class II molecules and remarkably they share common structural features. MHC class I molecules are expressed on almost all kinds of cells, while MHC class II molecules are restricted to professional APCs, such as dendritic cells, macrophages or B cells. Class I MHC forms it peptide-binding site from folding the α_1, α_2, α_3 domains, while the β_2 μicroglobulin provides stability to the α domains by forming tight dining with the α_3 domain. The overall architecture is shown in Fig 1 (B). Polymorphic residues are located around the binding site and they provide sequence variation for the peptide specificity of individual MHC alleles. MHC class I molecules present endogenous peptides produced by protease-degraded proteins, while MHC class II molecules present exogenous peptides derived from the environment. The peptide length ranges from 8 to 10 and from 13 to 25 residues for MHC class I and MHC class II, respectively [19]. In MHC class II, the peptide is located deep inside the binding groove compared to MHC class I. The MHC binding groove is closed at both ends so longer peptides have to adopt bulged conformations, providing specific hot spots.

In MHC class II, the binding groove is open so longer peptides can be outside the ends of MHC binding site. Humans carry 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. Human leukocyte antigens are abbreviated as HLA. Each individual HLA alleles from the same person can differ up to 20 amino acids. This polymorphic nature permits HLA proteins to express and present any kind of epitope in their binding groove to be easily scanned by the TCR [20].
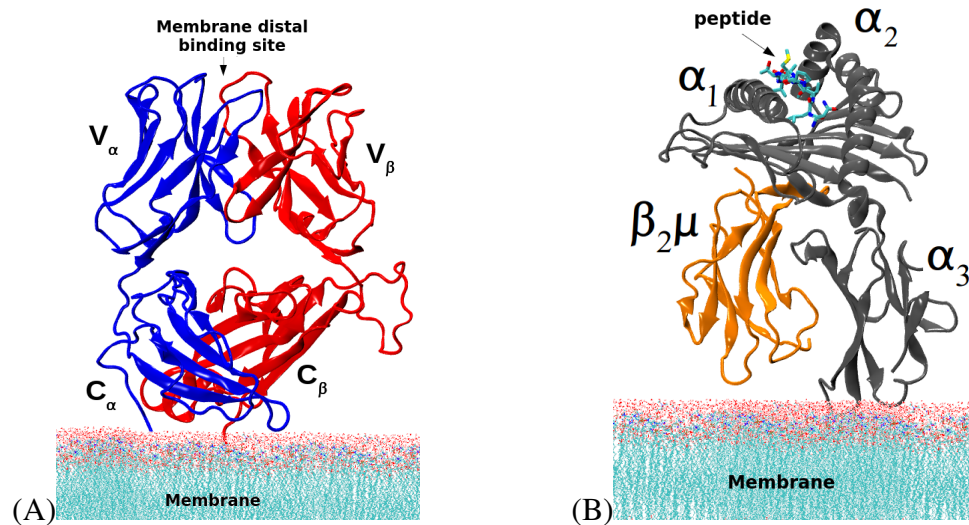
Figure 1: Ribbon diagram of the structure of human RA14 TCR and the MHC displaying an NLV peptide using the PDB code: 3GSN [21]. (A) Crystal structure of the human T-cell receptor, (TCR). TCR is an immunoreceptor found on the surface of every T lymphocytes and it is responsible for molecular recognition of antigenic peptides, also called epitopes. The TCR α and β chains are represented in ribbon diagram and are colored blue and red, respectively. The constant and variable domains are labeled. The membrane-distal binding site is about 80Å away from the membrane. Importantly TCRs do not have intrinsic signaling properties. (B) The structure of HLA-A2 bound to a viral NLV peptide colored in green. The major histocompatibility complex, also called MHC, is a cell surface molecule and its principal function is to bind to antigenic determinants derived from pathogens, display them on the cell surface, and present them to T-cells for recognition. The heavy and light chain of HLA-A2 are shown and they are labeled. β_2 microglobulin is colored in orange. The bound TCR or pMHC are not shown in both structures for clarity.

2.3   CD3 Signaling domains

TCRs do not have intrinsic signaling properties since it has a short cytoplasmic tail that lack signaling motif, and as a consequence they are required to transduce signals through the associated CD3 subdomains [22]; thus the $\alpha\beta$ TCR associates noncovalently with the CD3 $\epsilon\gamma$, $\epsilon\delta$, and $\zeta\zeta$ dimers to form a multisubunit complex across the membrane [7, 23]. TCR/CD3 complex plays an important role in differentiation, survival, and function of T cells by engaging with pMHC. In the TCR complex, the TCR $\alpha\beta$ chains are responsible for antigen recognition whereas the invariant CD3 chains are responsible for the signal transduction. The CD3 contains a total of 10 immunoreceptor tyrosine-based activation motifs (ITAMs). Binding of the extracellular domains of TCR to pMHC results in changes in the cytoplasmic region of the CD3 complex become phosphorylated and it triggers downstream signaling cascades leading to eventual responses.

The association is stabilized by polar interactions between CD3 molecules and TCR that have negatively charged and positively charged transmembrane region, respectively. For instance, sequence analysis in the transmembrane domains indicates that for each chain of the TCR complex, at least one charged residue is located within this transmembrane region and published studies indicate that these charged residues are responsible for the assembly of the transmembrane domains of the TCR complex [25, 26, 27, 24]. As shown in Fig. 1, each of the three basic residues on TCR $\alpha\beta$ chain interacts with a pair of acidic residues of the CD3 signaling dimers (CD3$\epsilon\delta$, CD3$\epsilon\gamma$ and $\zeta\zeta$). Structural studies provide insight on the arrangement of the TCR transmembrane domains [25], which reveal the basic residue motif that interacts with two acidic residues.

Even though the structure of the transmembrane domain of $\zeta\zeta$ chain [26] are solved, it is still a mystery how recognition of peptide by the TCR is propagated to CD3 domains and to initiate activation since the detail molecular assembly between $\alpha\beta$ TCR with CD3 signaling machinery is not available. Once propagated, the immunoreceptor tyrosine-based activation motif (ITAMs) of the CD3 signaling dimers are exposed and phosphorylated [28,

Figure 2: Assembly of the TCR-pMHC with the CD3 complex. (A) The TCR-pMHC associates noncovalently with the CD3 $\epsilon\gamma$, $\epsilon\delta$, and $\zeta\zeta$ dimers to form a multisubunit complex across the membrane. The TCR-CD3 complex plays an important role in differentiation, survival, and function of T cells by engaging with pMHC. In the TCR complex, the TCR $\alpha\beta$ chains are responsible for antigen recognition whereas the invariant CD3 chains are responsible for the signal transduction. The CD3 has a total of 10 immunoreceptor tyrosine-based activation motifs (ITAMs). (B) Each of the three basic residues in the TCR Transmembrane(TM) regions serves as a critical contact for one of the three signaling dimers that associate with TCR. TCR-CD3 polypeptides are shown in a representation of the lipid bilayer. (C) Simplified helical wheel projection depicting interactions among TM domains in the same relative positions as above. Images adapted from references [2, 24].

29], which are required for activation of the downstream process. Several models, including a "safety" model [30], have been proposed to explain how engagement of the TCR by agonist pMHC may lead to the phosphorylation of the ITAMs in the associated CD3 dimers. The "safety" model hypothesizes that the cytoplasmic domains of the CD3 ζ and ϵ chains are associated with the plasma membrane, which buries the tyrosine residues inside the membrane; thus preventing them from being phosphorylated. TCR-pMHC engagement destabilizes their associations and releases those tyrosines from protection and allows their phosphorylation. NMR structures of the CD3ϵ domain [27] supports this model.

Figure 3: The TCR safety model: The TCR subunits including its extracellular portion are depicted and colored as TCR α (gray), TCR β (purple), CD3γ (dark green), CD3δ (light green), and CD3ϵ (orange), the cytoplasmic domains, including CD3 ζ (yellow). The phenyl-rings indicate the critical tyrosines located inside the ITAM motifs of CD3ζ cytoplasmic domains while the blue dots represent the basic residues that interact with the acidic residues located in the inner leaf of the membrane. (A) Prior to activation, basic residues in the cytoplasmic domain of the CD3 subunits associate with the plasma membrane such that the key signaling tyrosines are sequestered in the plasma membrane; thus the TCR safety is "on" and the ITAMs are embedded. (B) The safety is "off" due to pMHC binding. The critical tyrosine residues in the ITAMs are accessible to cytoplasmic Src kinases and become phosphorylated in the "off" state. However, this model does not answer what dislodges the ITAM motifs from the membrane and allow phosphorylation?. The Figure is taken form ref [30].

2.4 TCR C_β FG Loop

Currently, it is widely accepted that TCRs mediate T cell activation by connecting their dimeric $\alpha\beta$ chains with CD3$\epsilon\gamma$, CD3$\epsilon\delta$, and CD3$\zeta\zeta$ subunits using an **"unknown mechanism"**. The TCR β chain contains the C_β FG loop insertion that protrudes between V_β and C_β domains. The 14-residue C_β FG loop is a mammalian adaptation and it is highly conserved. Contemporaneous with this insertion event was molecular speciation of a CD3 precursor into CD3γ, CD3δ genes. The FG loop is critical in the mechanosensing operation of the TCR since it greatly increases the sensitivity of T cell-based recognition. A number of features point to its potential role in mechanosensing. First, the positioning of the C_β FG loop between the C_β and VC_β domains physically rigidifies and stabilizes these structures. Second, the FG loop contains several residues that are conserved across mammals. Third,

the FG loop region is adapted to CD3$\epsilon\gamma$ but not CD3$\epsilon\delta$ to promote interaction. Last, sensitivity measurements suggest that the FG loop is significantly influential in TCR triggering. T cells expressing TCR mutants that lack the FG loop require 100 to 1,000 fold higher molar concentrations of antigenic peptides for stimulation than T cells expressing wild-type TCR counterparts. Thus, the TCR C_β FG loop is regarded as very important in this process since its removal also affects T cell activation and development [31]. For instance, FG-loop deleted T cells: induce less tyrosine phosphorylation; manifest reduced proliferation and cytokine production abilities upon T cell stimulation; and impair negative selection [32]. The FG loop is internally well-structured and is composed of 12 to 14 amino acid long chain that protrudes out of C_β domain [33, 34], as shown in the inset of Fig 4. The FG loop has a conserved W223 that forms hydrogen bonds with the backbone of Q225 and R227; W223 also makes extensive hydrophobic contacts with residues L217 and P230. W223 is held fixed in space by those interactions that in turn lend rigidity to the FG loop. This closely packs the FG loop to the β chain and integrates the C_β and V_β domains by forming a rigid structural entity. It has been long asked how recognition of peptide by TCR evokes activity on T cells since TCR antigen-specific binding site is located distant from the FG loop and CD3 subunits. Therefore understanding how recognition of epitopes presented by MHC is transmitted to the CD3 domains to induce T-cell activation is still regarded as the **"holy grail"** of molecular immunology.

2.5 *The role of CD4/CD8 co-receptors*

Along with TCR/CD3 complex, the role of coreceptors is still controversial in T cell recognition and activation. T cells have two co-receptors, CD4 and CD8– that play important roles in the binding of TCR-pMHC. Co-receptors bind the MHC molecule mainly to bring Src family tyrosine kinase p56 Lck or Lck to the TCR/CD3 complex and trigger the signaling cascades [35, 36].

CD4 is composed of 4 immunoglobulin domains (D1, D2, D3, and D4) that are linearly
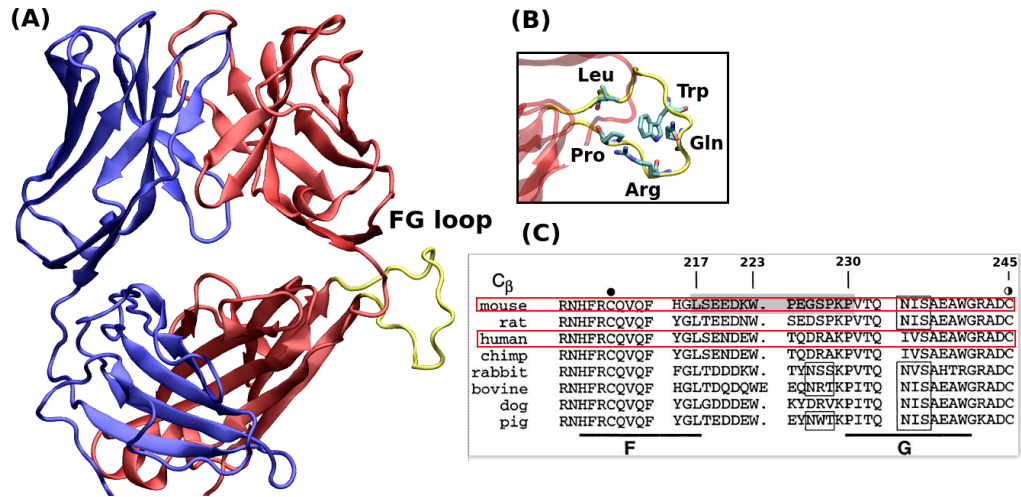
Figure 4: The structure and sequence TCR Cβ FG loop. (A) Crystal structure of the αβ TCR coloring in yellow the FG loop. (B) Molecular detail of the FG loop, highlighting key residues. (C)Sequence comparison among various species. The FG loop insertion contains well-conserved key residues. The F strand cysteine contributing to the Cβ intra-chain disulfide bond is indicated by a dot. Overall mammals have elongated Cβ FG loops and correspondingly distinct CD3γ and CD3ϵ subunits. The TCR Cβ FG loop is very important since its removal affects T cell activation and development. FG-loop deleted T cells: 1) induce less tyrosine phosphorylation; 2) manifest reduced proliferation and cytokine production abilities upon T cell stimulation, and 3) impair negative selection. The PDB code: 3GSN [21] is employed to depict the TCR.

joint as a glycoprotein chain. The D domain is flexible to change conformation and interact with the MHC. D1 interacts with the β_2 domain of an MHC II molecule [37, 38] while cytoplasmic C-terminal end interacts recruits Lck molecules. CD8 is a αβ heterodimer transmembrane glycoprotein formed by two IgV-like domain located away from and connected to the plasma membrane by a thin stalk. The extracellular IgV α chain interacts with the α3 portions of a Class I MHC molecule and the cytoplasmic tail also recruits Lck molecules to the TCR/CD3 complex. CD4 and CD8 are expressed on different subpopulations of T cells that have different functions. CD4 is restricted to an MHC class II expressed on T helper cells, and CD8 is restricted to an MHC class I expressed on cytotoxic T lymphocytes (CTL) [39]. CD8 and CD4 interacting with the TCR-pMHC complexes are shown in Fig 5.

Currently how the TCR and its (CD8/CD4) co-receptors interact to initiate signals with defined cellular responses is not entirely understood. The TCR α-chain connecting peptide contains an conserved motif called α-CPM. Mutations α-CPM promote unresponsiveness to antigenic stimuli and defects in positive selection; interestingly, negative selection with an agonist ligand was unaffected [40].
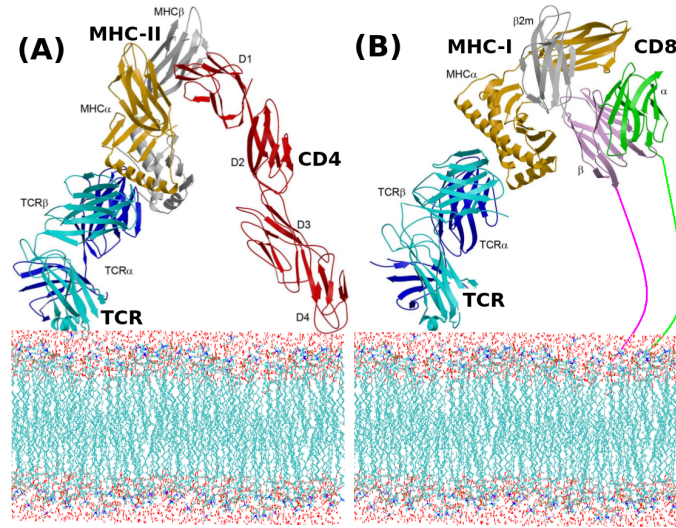
Figure 5: (A) The Crystal structure of the ternary complex of TCR-pMHC-CD4 (MS2-3C8-MBP-DR4-CD4) using the PDB code 3T0E [41]. (B) Model of the ternary complex of TCR-pMHC-CD8. It is constructed by superposing the H2D^d-CD8$\alpha\beta$ complex using PDB code 3DMM [42] onto a TCR-H2D^b complex using the MHC class I molecule as a reference, (3PQY) [43]. The CD8$\alpha\beta$ stalk regions are drawn manually and they are inserted in the cell membrane. All domains are labeled. Mature T cells express CD4 or CD8 molecules and function as important co-receptors by binding to invariant portions of the MHC I (CD8) or MHC II (CD4) and increasing the interaction of the pMHC. CD4 and CD8 play important roles is to bring the Src family tyrosine kinase p56 Lck or Lck to the TCR/CD3 complex and trigger the signaling cascades. CD4 is composed of 4 immunoglobulin domains that are linearly joint as a glycoprotein chain. CD8 is a $\alpha\beta$ heterodimer transmembrane glycoprotein formed by two IgV-like domain located away from and connected to the plasma membrane by two thin stalks. The plasma membrane is created using VMD.

2.6 The current models of interactions between the TCR and CD3 signaling subunits

The contacting interactions between the TCR and CD3 signaling subunits and how they are docked one respect to others are still highly debatable and controversial in the scientific

community. It is shown that the FG loop in TCR Cβ interacts with CD3$\epsilon\gamma$ while the DE loop in TCR Cα interacts with CD3$\epsilon\delta$; and the CC' loop in TCR Cβ was found to interact with both CD3$\epsilon\gamma$ and CD3$\epsilon\delta$ [44, 45, 46]. It is suggested that CD3ϵ domain lies in close proximity to the Cβ FG loop and it is potentially accommodated in a small cavity formed below the Cβ FG loop and the Cα CD and EF loops [47] and the Cβ FG loop facilitates signal initiation [44]. It is suggested that AB loop located on the TCR Cα domain experiences conformational changes when interacting with the CD3ϵ [48, 49].

More sophisticated NMR studies have independently reported that the helix-3 and helix 4-F strand on the TCR β interacts with the CD3$\epsilon\gamma$ [50, 51] while F and C strands on the TCR α chain interacts with CD3$\epsilon\delta$ [51]. It is proposed that CD3$\epsilon\delta$ is placed below rather than alongside the TCR α-chain [52]. A previous study also supports this model of the TCR binding to CD3 signaling subunits [40]. TCRs with a substituted α-chain connecting peptide motif exhibit a reduced association with the CD3δ subunit and this correlates with the finding that CD3δ-deficient thymocytes are defective in undergoing positive selection [53]. This implies a role for CD3δ being placed below the TCR and interacting with the connecting peptides.

Based on these experiments, many competing models of how the CD3 signalling subunits re-arrange around the TCR have been proposed. Importantly, these models are constructed by using indirect experiments such as mutagenesis, MNR, small-angle x-ray scattering, electron microscopy images, and T-cell functional assays since it is very complicated to crystallize the TCR bound to CD3 signaling subunits. The molecular models can be broadly classified into 3 categories: the one-sided model [50, 54, 55, 56], the two-sided model [57, 33], and the sitting-underneath model [52, 51, 40].

It is reported recently that the buried residues of TCR α/β-interface undergo significant conformational transitions upon cognate ligand binding, with most of these residues located at the Cα/Cβ interface [48, 58, 59, 60]. This proposed a systematic reorganization of the

TCR α/β interface. In addition, upon binding, a residue (V116$^\beta$) located at end of the 3_{10}-helix connector is indicated to become mobile based on the broadening of the NMR spectra line [60]. These new conformational changes are not included in the previous models.

Importantly, the structural basis of the TCR conformational changes, their impact on interactions with the CD3 signaling subunits, and how they can be related to the models are not investigated. The proposed models are not mutually exclusive if the TCR is thought to be a dynamic structure.

2.7 Immune response against viruses and their mutant variations

In human, CTLs play a critical in controlling infections as humans are constantly fighting persistent viral infections. CTLs recognize viral determinants in the context of HLA class I molecules. Once the CTLs are activated, they eliminate infected cell by secreting perforin/granzyme lytic granule or antiviral cytokines [61, 62]. The Epstein Barr Virus (EBV) and the cytomegalovirus (CMV) infect asymptomatically humans with more than 70% and 80% of adult populations, respectively [63, 64]. Hepatitis C (HCV; infecting globally about ~175 million people) and Human Immunodeficiency Virus (HIV; infecting globally about ~36.7 million people) are also viral infection; however, they are very severe to the body and they pose challenge to the immune system with inhibitor molecules to impair the TCR activation [65, 66, 67].

Another alternative used by viruses to escape immune recognition is via mutations that increase the viral fitness [68]. The killing capability of the immune response is severely disrupt by chronic and persistent infections [69, 70] bearing epitope mutations. In such case, if some viral variants survive the attack of the immune response, they are select based on their viral fitness to allows the virus to rebounds displaying persistent viral infection. At this stage, if the system is incapable of adapting and killing the variants a chronic infection is established and it will persist in the host.

The cause of persistent infections can be attributed to the viral mutations. A single

amino acid substitution can take place at anchoring residues and this results in the complete loss of an exposed residue at the MHC groove that initially interacts with the TCR; hence, the infected cell escapes the immune response. If the single amino acid substitution takes place at non-anchoring residues, it could potentially lead to the TCR forming a loose binding with the pMHC; thus this interaction produces an improper immune response to the infected cell [71, 72, 73]. In addition, viruses utilize the machinery of the infected cell to express components of viral proteins that disrupt the presentation of the epitope and dramatically impacting TCR recognition.

However, T cells are expected to adapt to the mutant variations as it is shown in the case of the influenza viral infection (NP418-426 the epitope of HLA-B*3501) [71], where the TCR shows signs of antigenic drift and emphasizes its adaptability. TCRs adapting to variations have the ability to recognize viral mutations and induce a proper CTL response [74, 75]. The TCR variants are shown to be compatible inducing cross-reactivity against multiple viral strains [76].

2.8 Cytomegalovirus-specific T cell response

Antigens produce an immune response by the host since they are considered foreign materials by the body and they are usually recognized by the immune system. Antigens are made of multiple linear antigenic determinants, known as epitopes, and the epitopes bind to a specific TCR in the context of MHC or antibody. Although Viruses could infect the host by multiple means, they usually enter into the cell's body by two pathways. The virions predominantly bind to the cell surface by forming an interaction between its glycoproteins– located at the surface– and the respective proteins found on the host cell surface. Alternatively, virions break into the cell by receptor-mediated endocytosis. The result of these pathways primarily guarantees that the viral genome enters the cytoplasm and fuses to the native genome. After entering the host, the proteasome degrades the virus-associated proteins and other unneeded or damaged proteins by proteolysis into small and linear peptides.

The peptides are usually smaller than 10 to 13 amino acids and they are easily shuttled into the endoplasmic reticulum (ER). The complex pMHC is formed and is delivered to the cell surface in this process. The peptide is very important in forming and stabilizing the complex pMHC where antigen specific CD8$^+$ T cells recognize and act upon after recognition [77, 78]. This process is depicted in Fig 6.
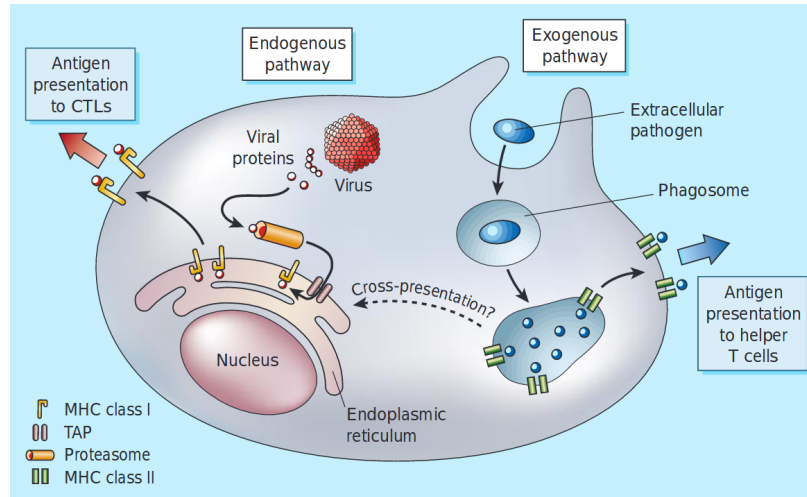
Figure 6: Proteins from intracellular pathogens, such as viruses, are degraded by the proteasome into small peptides that are shuttled into the ER by the transporter associated with antigen processing proteins (Endogenous pathway). These peptides are loaded onto MHC class I molecules and the complexes are presented to cytotoxic T cells that recognize viral peptides and kill the infected cells. In the exogenous pathway, extracellular pathogens are engulfed by phagosomes where the pathogen-derived peptides are loaded directly onto MHC class II molecules. This pathway usually activates helper T cells that stimulate the production of antibodies. But some peptides from extracellular antigens can also be "presented" on MHC class I molecules. The figure is adapted from Ref [78].

Cytomegalovirus (CMV) is classified as a member of the Betaherpesvirinae subfamily of Herpesviridae; although the infection is persistent, it is asymptomatic. It infects 50-85% of adults by the age of 40 and the virus has evolved to live in equilibrium with some cells and is typically in a latent state of infection [79]. However, immunocompromised individuals with CMV infection can lead to morbidity and mortality[7]. Hepatitis C virus (HCV) belongs to the Hepacivirus. Unlike CMV, HCV poses various features to evade the immune detection causing even an immunocompetent host the possibility of developing cirrhosis or hepatocellular carcinoma following HCV infection. CMV does not pose a

threat to an immune competent subject in terms of developing a life-threatening disease; however infected subject with HCV evolves to a chronic state during the course of infection. The development of a chronic condition is linked in part to the rise of "viral mutations" that have decreased functional effects as those of the wild type thereby escaping recognition from the immune system. Nevertheless, the true extent of TCR-pMHC interaction, its role in the molecular mechanism of antigen recognition, and the tolerance of leukocytes to viral epitopes upon infection, still remain a speculative issue, largely due to the lack of experimental approaches that could definitively address this question [80, 81].

Even though CD8[+] T cells specific for 55 kDa immediate-early protein 1 (IE1) are found at higher frequencies in some populations, the immune response is mounted against a CMV peptide derived 65kDa viral tegument protein (pp65) also called phosphoprotein 65. The epitope, NLVPMVATV, is extracted from the linear sequence of residues from 495-503 of pp65, and it is refer referred as NLV for its 3 first letter sequence. Humans expressing the HLA class I molecules HLA-A2 (*0201) and HLA-B7 reacts against the epitope, NLVPMVATV (495-503), from the pp65, referred to as for its first 3 letter. The pp65 is responsible for modulating the host cell immune response during CMV infection and it also serves as the immunodominant antigen for CD8[+] T cells [82, 83].

The virus has a double DNA inside an icosahedral capsid. Between the capsid and the outer lipid envelope contains tegument protein. The virus infects epithelial cells, B cells, and monocytes. The virus binds to the cell surface by forming an interaction between glycoproteins on the virion envelope and the heparin sulfate proteoglycans found on the cell surface. Followed by the fusion that takes place involving the envelope of the virion and the plasma membrane of the cell, the capsid, and associated tegument proteins are released into the cytoplasm. The dynein inside the cytoplasm mediates the movement of these viral proteins to the nucleus allowing the viral DNA enters the nucleus through the nuclear pore. The viral genomic DNA is encapsulated inside an icosahedral protein capsid and between this capsid and the outer lipid envelope contains proteinaceous tegument. Some of these

tegument proteins are responsible for modulating the infected cell that disrupts the immune response during a viral immune response of the host [84].

2.8.1 Private vs public T cell responses

Multiple crystal structural of NLV-specific TCRs in complex with NLV-HLA-A2 are generated [21, 85] to study private vs public T cell response and understand the rise of a high clonal diversity. Private T cell response is when the TCR repertoires elicited by a particular antigenic epitope are distinct between individuals and public T cell response is when the repertoires contain TCRs that are frequently observed in multiple unrelated people. Public TCRs have been described in immune responses to a variety of human viruses, including human CMV, human immunodeficiency virus (HIV), and Epstein-Barr virus (EBV) [86, 87].

2.9 T-Cell-Mediated Autoimmunity of Diabetes

Autoimmunity occurs when the body's immune system attacks and destroys its own healthy cells and tissues. It is estimated that there are >80 types of autoimmune disease. Prominent examples include diabetes mellitus type1, multiple sclerosis, celiac disease, rheumatoid arthritis, dermatomyositis, sarcoidosis, systemic lupus erythematosus, Hashimoto's thyroiditis, and Graves' disease. Autoreactive T lymphocytes are key players in these autoimmune diseases by mistakenly recognizing self-pMHC as foreign antigen via their TCRs [88]. Many studies of autoimmune TCRs have demonstrated that their interaction with autoantigen pMHC is very important to T cell activity [89, 90, 91, 92]. However, the structural database for human autoantigen-reactive TCRs in complex with their self-pMHC is very limited. Indeed, there exist only 5 complexes: Four are restricted to MHC class II and specific for multiple sclerosis [93, 94, 95, 96], while only one is restricted to MHC class I and specific for Diabetes mellitus type 1–a human, preproinsulin reactive, CD8$^+$ T cell clone [97]. The latter is also shown that it is capable of recognizing over 1 million different peptides. High-resolution structures of the TCR bound to 7 altered peptide

ligands shows that binding was stabilized through a conserved lock-and-key-like minimal binding footprint; this enables 1E6 TCR to tolerate vast numbers of substitutions outside of this so-called hotspot [98].

Diabetes mellitus type 1 is commonly known as type 1 diabetes (T1D). It affects >1.25 million Americans and the incidence is increasing at an alarming rate of 4% per year [99]. The precise cause of T1D is unknown but may be attributed to genetic factors, environmental triggers, and exposure to viruses and drugs. T1D results from the autoimmune destruction of the insulin-producing β-cells in the pancreas [97] that leads to lifelong insulin dependence. T1D is currently not preventable and involve the expansion of autoreactive CD4[+] T helper cells and CD8[+] T cells, autoantibody-producing B cells, and activation of the innate immune system [100, 101]. Recently, the crystal structure has been solved for an autoreactive TCR (1E6) on CD8[+] T-cells liganded with its cognate autoantigen peptide from the major β-cell preproinsulin (PPI) protein complexed with HLA-A2 MHC [97]. The study suggests that the TCR-pMHC interaction is very important in determining the fate of T cell. "Autoreactive T cells could evade negative selection due to a light-touch MHC contact footprint and weak TCR binding". However, it is still unclear what molecular nature of TCR-pMHC interaction leads to T cells escape thymic selection.

2.10 Specficity vs Sensitivity

The TCR differs from other cell-surface receptors by its unusual features of antigen recognition. Firstly, each unique TCR recognizes only a subset of pMHC molecules. Since the high-affinity agonist pMHC molecules on APCs are of very low abundance, TCR recognition must be very sensitive. Indeed, it has been shown that a single antigenic pMHC can trigger TCR activation [102, 103, 104]. Secondly, generally MHC molecules on APCs are conjugated with self-peptides, so the TCR has to be able to recognize a tiny subpopulation of agonist from a huge background of self-pMHCs. During the maturation process,

however, T cells are positively selected for survival by the ability of their TCRs to bind self-pMHCs [105]. So the TCR has to be able to bind self-pMHC but this binding cannot trigger TCR activation; otherwise, it would cause autoimmunity. Interestingly, the concurrent presence of self-pMHCs can increase the sensitivity of TCR to agonist pMHC [106, 107, 108]. Thirdly, the same TCR can recognize multiple ligands with different binding affinities, resulting in different immune responses [89]. Lastly, TCRs can recognize foreign peptide ligands that they have never encountered before, exemplifying the diverse nature of TCR-pMHC interactions. T-cells also expresses coreceptors CD8 or CD4, which enhances TCR signaling for binding with MHC class I or class II molecules respectively [6]. The precise roles of the coreceptors are not fully understood.

The triggering mechanism of TCR and the effect of CD8 or CD4 coreceptors are major focuses of current immunological studies. Various TCR triggering models have been suggested. It has been reported that the application of mechanical forces to the TCR induces or enhances TCR triggering [44, 109]. It has also been proposed that pMHC binding could push or twist the TCR [110, 33]. A piece of circumstantial evidence is that agents that disrupt the actin cytoskeleton abrogate TCR triggering [111], presumably by inhibiting the force transmission.

2.11 Receptor-ligand signatures: Catch, slip, and ideal bonds

Force spectroscopy experiments manipulate single molecules and are capable of producing forces in the range of pico Newton that are necessary to dissociate a receptor from its ligand. At the same time, these experiments sensitively measure the dissociation time with a high precision. If the effects of mechanical forces on single receptor-ligand bonds is measured under different forces, it would produce a specific molecular signature shown in a bond lifetime vs force profile. Experimentally, the force between a single bond interaction is kept at a constant level until the bond dissociates.

A bond behavior is called "**slip**" bond (Fig 7 A) if the average bond lifetime decreases as

the force increases; thus this bond becomes shorter lived in the presence of force. "**Catch**"-bond (Fig 7 B) behavior exhibits a unique phenomenon where under a certain regime of force, the lifetime increases as the force increases. Then the catch bond would transition to slip bond behavior as this is due to the forces overpowering the bonds formed. It is said that the interacting proteins catch and their bond becomes longer lived when they are mechanically pulled. Catch bonds are nonlinear responses that prolong the protein-protein interactions in the presence of mechanical stress and it provides a way for proteins to grip tightly and stabilize their attachments. "**Ideal**" bonds (Fig 7 C) are independent or insensitive to mechanical forces. Slip bonds have basically a constant bond lifetime under different forces. Slip bonds are commonly observed in nature while catch bonds have been observed with motor proteins, with adhesive proteins like selectins, integrins, and cadherins, and with immune receptors like TCRs interacting with pMHCs. Ideal bonds are currently observed only for cadherins [112, 113].
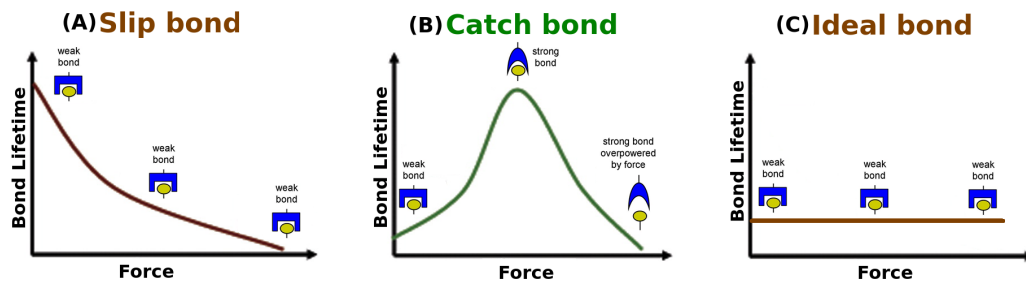
Figure 7: Catch, slip, and ideal bonds: When a mechanical force is applied to a receptor-ligand interaction, it responds as: (A) Slip bonds when the interacting molecules slip apart and the bond becomes shorter lived in the presence of tensile force, (B) Catch bonds when they catch and the bond becomes longer lived when they are mechanically pulled or (C) Ideal bonds when they are insensitive to mechanical stress and the bond is basically constant under different forces.

2.12   Significance

T cell activation is an important feature of the immune response of the host to effectively clear viral pathogen and cancerous cells. It is also important in establishing self-tolerance and avoid autoimmunity. The TCR is a mechanosensor and there is a huge effort from the

scientific community in understanding how peptide ligand recognition is translated to TCR activation and how it is influenced by peptide potency and force. It has been more than 3 decades of extensive study and the rules of TCR recognition and activation of T cell is still not understood. This thesis studies how the structural dynamics of the TCR are relate to the recognition of pMHC, and triggering of the CD3 signaling domains and it provides insight in how to structurally manipulating the interaction of the TCR ligated to pMHC. It is described 1) how the information encoded in the peptide is decoded mechanically by the TCR and what is the structural role of catch bonds; 2) how information read from the TCR binding site is propagated to the CD3 signaling domains. This can be exploited to develop and engineering proteins for a medical and biotechnological application. For instance, the ImmTAC technology for cancer immunotherapy combines old primitive structural knowledge of TCR and CD3 for producing vaccines.

This study is also relevant because it employs human TCRs. All previous models are performed in mice TCR that have been shown previously to map poorly when used to understand human infections, immune disease, and cancer. This study elucidates the molecular mechanism driving the human TCR-pMHC interaction and how it leads to activation in the context of viral persistent infections and autoimmune diseases. Viral persistent infections pose problems to the host by making viral mutations, which occurs to increase viral fitness by evading immune responses. Autoimmunity is a condition where the TCR recognizes its self-peptide ligand and induces the body's immune system attacks the healthy cells. This study also aims to investigate the molecular details of such interactions by characterizing their molecular footprint in the presence of force and correlating the findings with a functional assay. More importantly, it describes how the contribution of force increases the number of H-bonds (bond-lifetime) at the atomic level in the context of viral and autoimmune peptide ligands with the capability to lower cytolytic responses of CTLs.

This study will contribute to understand how altered peptide ligands impact on recognition to activate or induce tolerance in T cells by using computer simulations. This study

24

is also significant because it provides evidence that TCRs are capable to experience large deformations without releasing the peptide ligands and be still functional and describes the molecular mechanism of how TCRs uses their C_β FG loop to propagate information to CD3 signaling domains. This approach is also innovative since it allow to study the mechanical regulation of TCR-pMHC interactions at the atomic level and provides a unique perspective to the field of Immuno-engineering. Overall, the resulting insights of this study provide evidence to understand how to manipulate and optimize the interactions between TCR and pMHC to effectively trigger cytolytic function and improve emerging technologies, such as, chimeric antigen receptors, ImmTAC, etc.

CHAPTER III

METHODS: THE COMPUTATIONAL MICROSCOPE FROM STATIC STRUCTURES TO LIVING PROTEINS

3.1 Background

Optical microscope has played a fundamental role in understanding and capturing biological samples using the single and compound lens. The first microscopes were developed to see the inner working of biological materials beyond the limit of a naked eye and explore their biophysical properties. Since the seventeenth century, science and technological developments have led to a manufacturing breakthroughs of microscopes that significantly advanced their designs improving dramatically the image quality. However, despite the manufacturing breakthroughs, optical microscopes are still hampered by an optical limit that is imposed by the diffraction of visible light. For instance, to see the atomic world, the wavelength of radiation normally needs to be similar to the dimensions of the atoms to be observed (0.05 nm). This wavelength of light (500 nm) is unable to accomplish it since it is several orders of magnitude bigger. To achieve atomic resolution and overcome the diffraction barrier, nowadays structural biologists commonly use three main experimental techniques to determine and observe the 3D structure of macromolecules. These experimental techniques have been developed and complement each other since they have their own limitations [114].

"X-ray crystallography" is used to determine the positions of atoms in biomolecules. The goal of X-ray experiments is to calculate an "electron-density" map of a molecule from its diffraction pattern. The obstacles are that X-rays cannot be focused as it is done in conventional microscopes, where lenses are used to focus and magnify light. In an X-ray diffraction film, spots convey only information about amplitudes of the waves that caused

them but not their phases; thus important information is lost. This "phase" problem was solved and determined by introducing heavy atoms into protein crystals and comparing the diffraction patterns obtained from different "isomorphous" crystals [115].

"Electron microscopy" (EM) has higher magnifications and greater resolving power as it uses a beam of electrons rather than light to produce an image of the sample, allowing smaller objects to be observed in finer detail. There are two main types of EMs: Transmission Electron Microscope (TEM) and Scanning Electron Microscope (SEM). TEM is a high-resolution tool based on transmitted electrons while SEM is based on scattered electrons. SEM focuses on the surface of the sample and its composition whereas TEM provides great details about internal composition. SEM is used to characterize the surface morphology, micro particle size, cross-sections of coatings deposit, and thickness. TEM is capable of measuring nano particles size, crystallite size, atomic arrangement in materials at nano resolution. Thus, TEM is used to obtain 3D structure of proteins at low resolution (3.5 to 7.0 Å) [116, 117, 118].

"Nuclear magnetic resonance" (MNR) is a physical phenomenon in which nuclei are made to resonate in an applied magnetic field such as the spin polarization is transferred from one population of nuclear spins to another. Nuclear spin that is coupled through space due to spatial proximity give rise to nuclear Overhauser effect (NOE). NOE is the change in intensity of one NMR resonance that occurs when another is saturated. To determine protein structures, observed resonances have to be assigned to a particular chemical group. The distance information that is provided by the observed NOEs is the main basis for structure determination. NMR experiments are constrained by protein size since the width of resonances increases with the protein size. The advantage of NMR is that the macromolecule is studied in aqueous solution where no crystals are required or heavy atom derivatives are introduced [119].

X-ray crystallography is the most powerful and efficient structural method, however, it requires the production of good crystals. NMR only deal with macromolecules that are

smaller than about 50 kDa. However, it works in aqueous solution and is able to capture multiple modes of flexible regions of proteins. Although TEM experiments are performed in harsh conditions, they are a good fit for obtaining the overall structure of large systems and their multiple conformational states at relatively low-resolution. About 89% of solved protein (129,367 entries) in the Protein Data Bank (PDB) were obtained using X-ray crystallography, while 1% and 9% are coming from EM and MNR, respectively.

Unfortunately, those techniques only provide static pictures without any information about the dynamic nature of macromolecules. In biological processes, protein-protein interactions are likely to involve structure fluctuations and small-to-large conformational changes. For instance, kinetic properties of molecular interactions, such as catch or slip bonds, are modified by applying a force; thus its structural mechanism may include forced-induced conformational changes between states not observed in the static structures in which the crystal is formed in the absence of force. To overcome those challenges and understand the inner working of proteins, single-molecule force spectroscopy has been largely developed. Although these experiments are somehow successful sensing their properties and describing their molecular feature, they do not provide a clear representation of their dynamic process. Proteins are not static structure having a single unique conformation, but rather they are most likely undergoing multiple conformational changes locally (tenths of Å) or globally (tenths of nm). These conformational changes are functionally relevant and span a variety of time scales (from ns to s). It is known that such conformational changes are involved in ligand-receptor recognition, signaling, and many other biological processes. A microscope with an atomic resolution is required to capture the dynamic molecular detail of proteins and how their changing structure relates to their function. The development of high-performance computing in capacity and speed has opened new avenues of research. A well-established approach is molecular dynamics simulations that have elucidated the microscopic interactions between macromolecules that play an important roles in folding, binding, ligand recognition, and catalytic mechanisms.

Molecular dynamics simulation is a powerful technique and has become indispensable for understanding the structure and function relationship of biomolecules. Nowadays computer is just as important a tool for scientist, if not more, as the test tubes since simulations are becoming so realistic that they predict the outcome of traditional experiments. The scientific community has recognized it by awarding the 2013 Nobel Prize in Chemistry to Martin Karplus, Michael Levitt, and Arieh Warshel for "the development of multiscale models for complex chemical systems". These scientist have dramatically advanced the field of structural biology by developing sophisticated computer algorithms to build models of complex biological molecules. This Nobel Prize is the first given to work in computational biology, indicating that the field has matured and is on a par with experimental biology.

3.2   Molecular dynamics

Molecular dynamics (MD) simulation is a computational technique used to study the physical movements of atoms from small molecules to relatively large proteins. MD was initially developed in 1955 within the field of theoretical physics [120, 121, 122] but now it is virtually applied to most fields. In biology, it is widely employed since it provides fundamental tools for understanding the molecular relationship between structure and function of proteins. Proteins are not viewed as rigid entities anymore, they are dynamic structures that constantly modify their internal motions. These motions not only change the conformational state of proteins but also plays an essential role in determining their function. MD simulation creates a way to do "statistical mechanics by numbers" [123] and it magically implements the Laplace's demon–a deterministic vision of predicting the future by using Newtonian mechanics [124]. If a demon determines at a specific time the position and momentum of every particle in the universe, the past and future of all particles for any time can be calculated from the laws of classical mechanics. The demon not only tracks the sequence of events of all atoms but also provides insight into molecular motion on an atomic

scale by animating nature's forces.

In a protein, the atoms are allowed to interact for a period of time and their trajectories are determined by solving numerically Newton's second law of motion, where forces between atoms and their potential energies are calculated using interatomic potentials or molecular mechanics force fields as shown in Equation (1)

$$m_i \frac{\mathrm{d}^2 \vec{r}_i}{\mathrm{d}t^2} = -\frac{\partial}{\partial \vec{r}_i} U_{total}(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N), \quad i = 1, 2, ..., N \tag{1}$$

Where the m_i is the mass of atom i and U_{total} is the total potential energy. The trajectory of an MD simulation is employed to determine macroscopic thermodynamic properties of macromolecules, only if, the systems comply with the "ergodic hypothesis" in which the time averages of the system corresponds to microcanonical ensemble averages; $\langle A_{ensemble} \rangle = \langle A_{time} \rangle$.

Molecular dynamics simulation is usually performed by controlling the temperature, T [125, 126]. The velocity Langevin thermostat is used for this purpose and it generates the Boltzmann distribution for canonical (NVT) ensemble simulations. Molecular dynamics with the Langevin thermostat becomes a stochastic differential equation in which two additional force terms are being added to the equation of motion. The effects of solvent molecules are approximated by these two terms. The first term represents a frictional force which takes into account the frictional drag on the solute and the second one corresponds to random kicks associated with the thermal motions of the solvent molecules. It is assumed that friction force is proportional to the particle's velocity and opposite to the direction of the particle motion. The equation of motion is therefore written as a Langevin equation

$$m_i \frac{\mathrm{d}^2 \vec{r}_i}{\mathrm{d}t^2} = \underbrace{- m_i \zeta \frac{\mathrm{d}\vec{r}_i}{\mathrm{d}t}}_{friction} \underbrace{- \frac{\partial}{\partial \vec{r}_i} U_{total}}_{force} + \vec{\eta}_i \tag{2}$$

where ζ is the damping constant, U_{total} is the total potential energy, and $\vec{\eta}_i$ is the vectorial random force acting on the particle i, which satisfies the fluctuation-dissipation theorem

$$\langle \vec{\eta}_i \rangle = \vec{0} \quad and \quad \langle \vec{\eta}_i \cdot \vec{\eta}_j \rangle = 6k_B T \zeta \delta_{ij} \delta(t - t') \tag{3}$$

30

where δ_{ij} and $\delta(t - t')$ are the Kronecker delta and Dirac delta function, respectively. For constant pressure simulations (NPT ensembles), the "Langevin-Hoover" method is used and it is shown to generate the correct ensemble distribution [125, 126, 127, 128, 129]. The "Langevin-Hoover" equations are demostrated and found in Ref [126].

The total energy of the system is the summation of the following terms

$$U_{total} = U_{vdW} + U_{Coulomb} + U_{bond} + U_{angle} + U_{dihedral} \tag{4}$$

where U_{vdW} is the Lennard-Jones potential, $U_{Coulomb}$ is the Coulomb potential, U_{bond} is the harmonic potential, U_{angle} is the bond-angle potential, and $U_{dihedral}$ is the dihedral angle potential. These potentials are explained in the following sections.

3.3   Force fields

The force fields refer to the interactions that one particle exerts on another particle or a collection of other particles. Generally speaking, the force field can be divided into two categories, a bonded and nonbonded interaction. In this study, bonded interactions act between particles which have a common bond or bond-angle, and they are divided into bond potential and bond-angle potential. They are calculated on the basis of a fixed list. Nonbonded interactions act between atoms in the same molecule and those outside the molecule. In this study, nonbonded interactions are separated into Coulomb interaction and excluded-volume interaction. They are computed on the basis of a neighbor list.

3.3.1   Nonbonded interactions

3.3.1.1   Excluded-volume potential

The excluded volumes of particles, including the monomers and the counterions, are described by Lennard-Jones (LJ) potential, which is represented by an attractive and a repulsive term. The attractive term represents the van der Waals interaction due to induced dipole-dipole interaction. The repulsive term arises from the nonbonded overlap between electron clouds and has an arbitrary form. The LJ potential repels at close range, then

attracts, and at long range vanishes. It is defined by

$$U^{ij}_{vdW}(r_{ij}) = 4\epsilon_{vdW}\left[\left(\frac{\sigma_{vdW}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{vdW}}{r_{ij}}\right)^{6}\right] \tag{5}$$

where r_{ij} is the separation distance between particle i and j, ϵ_{vdW} is the strength of LJ interaction, which describes the hardness of explicit particles, and σ_{vdW} is the collision diameter. The Lorentz- Berthelot mixing rule is applied for the interaction between different kinds of particle, $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$.

3.3.1.2 Coulomb potential

The particles also interact with each other via electrostatic interaction. The electrostatic interaction is defined by the Coulomb's law, which states that the magnitude of the electrostatic force between two point charges is directly proportional to the magnitudes of each charge and inversely proportional to the square of the distance between the charges. The Coulomb potential is thus written as

$$U^{ij}_{Coulomb}(r_{ij}) = \frac{z_i z_j}{4\pi\epsilon_0}\frac{1}{r_{ij}} \tag{6}$$

Here z_i , z_j are the valances of particles i and j, respectively. The Bjerrum length, λ_B, can be defined as the distance at which the Coulomb potential of two unit charges "e" is equal to the thermal energy $k_B T$. k_B is the Boltzmann constant. λ_B is expressed as $e^2/(4\pi\epsilon_o\epsilon)k_B T$ where ϵ is the dielectric constant of the continuum and ϵ_o is the vacuum permittivity. Particle-particle particle-mesh (PPPM) algorithm is used to calculate the long-range Coulomb interaction, in which interactions with periodic images are taken into account. The PPPM method, which is an alternative approach to the Ewald sum and was proposed by Hockney and Eastwood [130], is an accurate and com- putationally efficient method for calculating interactions in molecular simulations. The PPPM method is based on separating the long-range inter-particle force into the sum of rapidly-varying short-range interactions and slowly-varying long-range interactions. The PPPM method is computed by two steps, first, by a PP method, and then by a PM method. The PP method is used to

find the total short-range contribution to the force on each particle and is only computed by direct particle- particle summation within some cutoff radius. Then, the PM method is used to compute the total slowly-varying, long-range interaction contributions which are calculated in the reciprocal space, approximated on a grid.

3.3.2 Bonded interactions

3.3.2.1 Harmonic potential

Neighboring atoms on the protein are connected by a spring, described by a harmonic potential. The harmonic potential, associated with small bond stretches about the equilibrium bond length, can be approximated by a parabolic equation, and it is defined by

$$U_{bond}^{ij}(r_{ij}) = k(r_{ij} - b_o)^2 \tag{7}$$

where k is the spring constant that gives the stiffness of the bond and b_o is the equilibrium distance.

3.3.2.2 Bond-angle potential

The bond-angle potential provides an intrinsic stiffness to the protein backbone. It is modeled by a three-body potential acting among adjacent atoms triplets on the protein chain, and it is defined by

$$U_{angle}^{ij}(\theta_{ijk}) = k_\theta(\theta_{ijk} - \theta_o)^2 \tag{8}$$

where θ_{ijk} is the bond angle between three consecutive atoms, $(i, j, k = i - 1, i, i + 1)$, and θ_o is the equilibrium angle, whose value is π rad. k_θ is the spring constant.

3.3.2.3 Dihedral-angle potential

The dihedral angle potential is also known as torsion potentials and they describe the interaction arising from torsional forces. A dihedral angle is formed between two intersecting planes through two sets of three atoms, having two atoms in a specific plane. The dihedral describes atom pairs separated by exactly three covalent bonds with the central bond

subject to the torsion angle ϕ

$$U_{dihedral}^{ijkl}(\phi_{ijkl}) = K_i^{\phi}(1 + cos(n_i\phi_{ijkl} - \gamma_{ijkl})) \tag{9}$$

where θ_{ijk} is the dihedral angle between four consecutive atoms, $(i, \; j, \; k, \; l \; = \; i-1, \; i, \; i+$ $1, \; i+2)$, and γ_{ijkl} is the equilibrium angle. k_i^{ϕ} is the spring constant. (Fig. 2). An "improper" dihedral term governing the geometry of four planar covalently bonded atoms is also included.

Nowadays, multiple force fields for MD simulations are available. Basically, there are three major force fields dominating most MD simulations. They are CHARMM, AMBER, OPLS-AA, etc, they also have different versions. Those force fields have similar functional forms for the potential energy as it was proposed in the study of "Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules" by Lifson [131].

3.3.3   Integration algorithm

The velocity-Verlet algorithm is used to integrate the equation of motion. There are various versions of Verlet algorithm and even new numerical integration schemes. However, the velocity-Verlet algorithm is, at the same time, simple, accurate and stable. The velocity-Verlet algorithm generates a phase-space trajectory which is a sequence of "snapshots" for the particle coordinates and velocities at time t, by the following procedure

$$\vec{r}(t + d\tau) = \vec{r}(t) + \vec{v}(t)d\tau + \frac{1}{2}d\tau^2\frac{\vec{f}(t)}{m} \tag{10}$$

$$\vec{v}(t + d\tau) = \vec{v}(t) + \frac{\vec{f}(t + d\tau) + \vec{f}(t)}{2m}d\tau \tag{11}$$

where $d\tau$ is a small time increment and $\vec{f}(t)$ is the total force acting on a particle at time t. If the initial conditions $\vec{r}(0)$ and $\vec{v}(0)$ are given, it is possible to compute $\vec{r}(t)$ and $\vec{v}(t)$, sequentially by applying Equations (10) and (11).

3.3.4 Steered molecular dynamics simulation

It was proposed in MD simulations [132] that a group of atoms is constrained and another is chosen as pulling atoms. The proteins are pulled apart by external forces following a specific direction and it can be applied to one or more atoms in the proteins. This protocol is termed steered molecular dynamics (SMD) simulations. The application of an external force was initially used to study the unfolding process of proteins. Here it is used to induce small-to-large conformational changes the protein since MD simulations do not effectively access or sample all possible configurations of the protein during nanoscale simulations. SMD simulations closely represent the manipulation of proteins through a force probe. The addition of an external force is a computational technique used to reduce the energy barriers and move across different conformations. The equilibrium state is usually an ensemble of structures as it is observed in the crystal structure. The other states are either rare events or steady conformers. SMD simulations usually include the proteins without the cytoplasmic tail and plasma membrane; however, in SMD it is assumed that the ectodomains are attached to an imaginary cell membrane and the proteins are constrained at their C-terminus.

The external force can be constant or variable over time, and this is accomplished by using constant-force or constant-velocity pulling, respectively. In the former, the SMD atoms are undergoing a constant force. In the latter, the center of mass (COM) of the SMD atoms is connected to a dummy atom via a spring with a force constant, k and the COM of SMD atoms is then moved at a constant velocity, v. The force is measured following a harmonic potential, U_{SMD}, defined as

$$\vec{F} = -\nabla U_{SMD} \tag{12}$$

$$U_{SMD} = \frac{1}{2}k[vt - (\vec{r} - \vec{r}_0) \cdot \vec{n}]^2 \tag{13}$$

$\vec{r}$ and $\vec{r}_0$ are the current and initial positions of the COM of the SMD atoms, respectively. $\vec{n}$ is the direction of pulling.

3.3.5 Targeted molecular dynamics simulation

Targeted molecular dynamics (TMD) is a computational technique to produce conformational change that leads to a known target structure at ordinary temperature by using purely a time-dependent geometrical constraint while the dynamics of the molecule is only minimally influenced by the applied constraint. TMD simulations enforce the transition independently of the height of barriers. Generally speaking, TMD simulations move initially a subset of atoms toward to a final target reference by using steering forces. Similarly to SMD, a harmonic potential is added to the total potential energy and the force on the target atoms is calculated by the gradient of the potential, U_{TMD},

$$U_{TMD} = \frac{1}{2}\frac{k}{N}\left[RMS(t) - RMS^{Target}(t)\right]^2 \qquad (14)$$

The spring constant k is scaled down by the number of selected atoms, N. The RMS distance between the current coordinates and the target structure is computed after first aligning the target to the current structure at each timestep. $RMS(t)$ is the instantaneous best-fit RMS distance of the current from the target coordinates. $RMS^{Target}(t)$ is changing linearly from the initial RMSD to the final RMSD.

3.4 Nanoscale Molecular Dynamics (NAMD)

NAMD– also formerly known as Not Another Molecular Dynamics Program– is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems, written using the Charm++ programming model– an object-oriented portable parallel language built on top of C++ [133]. NAMD is famous for its parallel efficiency and is often used to simulate systems containing millions of atoms. It scales well to 10s

and 100s of CPUs on small commodity clusters and high-end parallel supercomputers, respectively. NAMD is capable of using multiple force fields such as AMBER, CHARMM, and GROMACS. NAMD has been developed by the joint collaboration of the Theoretical and Computational Biophysics Group (TCB) directed by Prof Schulten, and the Parallel Programming Laboratory (PPL) directed by Prof Kale, at the University of Illinois at Urbana-Champaign. It was first introduced by Nelson et al. in 1995 as a parallel molecular dynamics code enabling interactive simulation by linking to the visualization code Visual Molecular Dynamics (VMD). NAMD has since matured, adding many features and scaling beyond ~ 0.5 millions of CPUs.
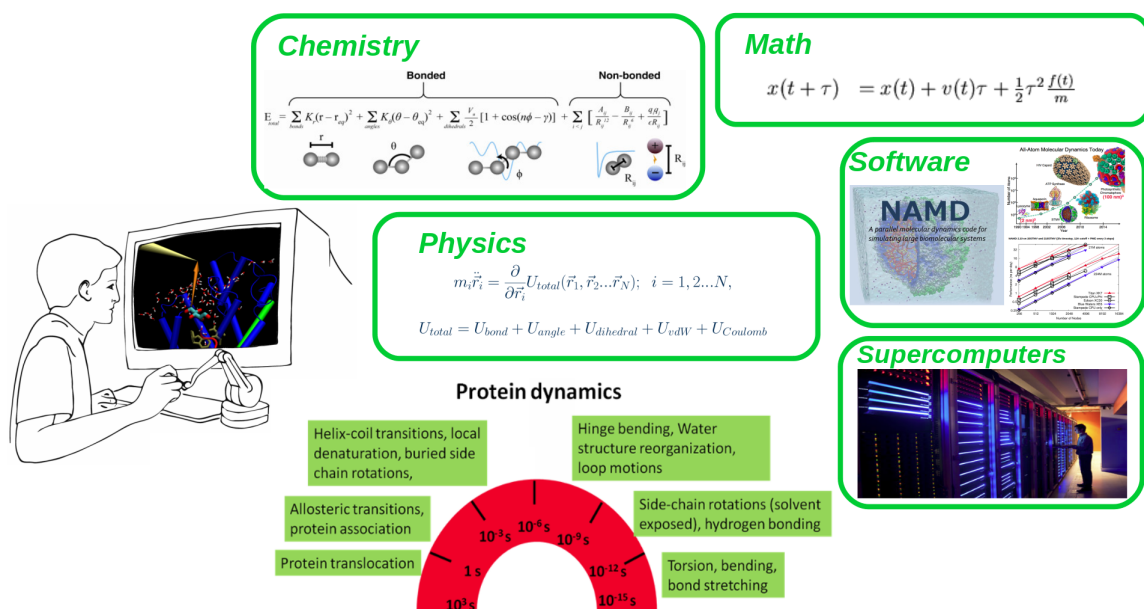
Figure 8: An Illustration of the performing molecular dynamics simulation and analysis. Molecular dynamics is a method that uses Newton's equations of motion to simulate the time evolution of a set of interacting particles. MD simulations are dependent on a description of how the particles interact using a force field and are popular in materials chemistry, biochemistry and biophysics.

Multiple molecular dynamics software packages are also available, for instance, CHARMM, AMBER, NAMD, GROMACS, Desmond, and OpenMM. They have a similar functionality to NAMD.

3.5   NAMD Code Performance and Scaling

In this thesis, the simulations are run on "Stampede" and "Comet" located at the Texas Advanced Computing Center (TACC) and at the San Diego Supercomputer Center (SDSC), respectively. NAMD is employed and it is benchmarked on these high-performance supercomputer using systems with sizes varying from 0.05 to 1.16 million of atoms under a different number of CPU cores. The CHARMM36 force field is used and it is observed that systems in that range can be linearly scaled up to ≈ 2000 cores. For this scaling, the systems can be roughly categorized into 3 sizes: 0.1, 0.2 and 2 millions of atoms.

Figure 9 (a) and 10 (a) show that Performance scales linearly with the number of nodes when they are smaller than 60 nodes for the two smaller systems. Performance is reported as simulation time in ns/day. The average simulation time is 100 ns. In Stampede, 60 nodes are able to complete such simulation in 1 and 1.5 days for 0.1 and 0.2 M, respectively. In Comet, 50 nodes are capable to finish it in 2 and 2.5 days for the two smaller systems. Figure 9 (b) and 10 (b) report that Speedup is at an optimal point around 50 and 60 nodes for Comet and Stampede, respectively. Figure 9 (c) and 10 (c) similarly show that Efficiency is optimal around that range. The requested SUs are estimated according to equation

$$S Us = (walltime)(\# \, node)(\# \, cores) \tag{15}$$

The wall time is set to 24 hours. Stampede and Comet have 16 and 24 cores per node, respectively. For the smaller systems,Figure 9 (d) (Stampede) and Figure 10 (d) (Comet) shows that the SUs per nanosecond of simulation per millions of atoms are economically optimal when 50 and 60 nodes are employed, respectively, because those use less SUs and yet they require less time for completing the simulation. For bigger systems, Figure 11 (b) shows that Performance for Stampede is invariant for systems ranging from 0.5 to 1.3 million of atoms under different of nodes, while it decreases for Comet in Figure 11 (a). Figure 11 (d) (Stampede) shows that the SUs per nanosecond of simulation per millions of atoms are best at 120 nodes in term of performance. Similarly, Figure 11 (c) (Comet)
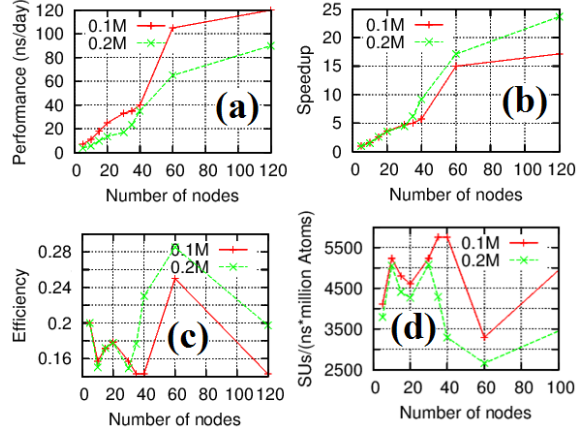
Figure 9: NAMD Benchmarks on TACC Stampede. (a) Performance, (b) Speedup, (c) Efficiency, and (d) SUs/(ns x millions of atoms) vs Number of nodes for 0.1 and 0.2 M atoms. Figure (d) is used to estimate the SUs usage in the present study. The optimal SU is about 2600 and is located at 60 nodes.
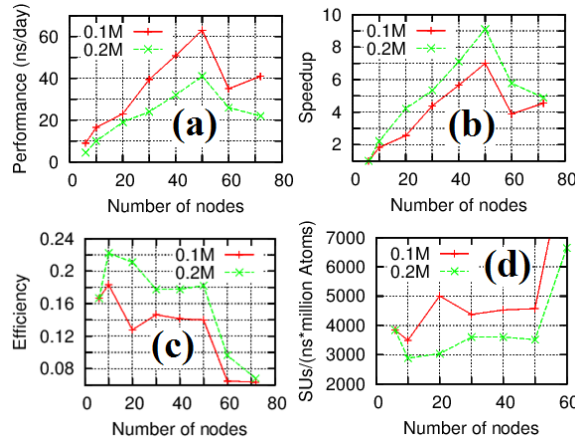
Figure 10: NAMD Benchmarks on SDSC Comet. (a) Performance, (b) Speedup, (c) Efficiency, and (d) SUs/(ns x millions of atoms) vs Number of nodes for 0.1 and 0.2 M atoms. Figure (d) is used to estimate the SUs usage in the present study. The average is about 3600 in the range of 30 to 50 nodes for 0.1 and 0.2 M atoms.

shows that it is roughly constant for 50 nodes but inefficient and expensive. These suggest that the output rate on Stampede is significantly better than Comet; thus our bigger system will be best run at Stampede. NAMD 2.10 were employed for this benchmark. The canonical ensemble (NVT) with periodic boundary conditions are used and the temperature is kept constant at 310 K using Langevin dynamics. The CHARMM36 force field was used with SETTLE (non-iterative) and SHAKE (iterative) algorithms to constrain the lengths
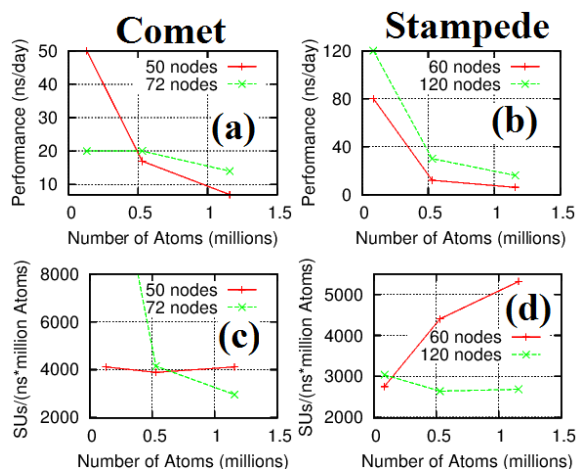
Figure 11: NAMD Benchmarks on SDSC Comet and TACC Stampede for bigger systems. Performance vs Number of atoms for Comet (a) and Stampede (b) at two number of nodes. SUs/(ns x millions of atoms) vs Number of atoms for Comet (c) and Stampede (d) at two number of nodes. Figure (d) shows Stampede provides an optimal SUs usage for bigger systems and the average is about 2600.

of bonds of water and other hydrogens, respectively. Water is represented by the standard TIP3P models. Electrostatic interactions are treated using Particle Mesh Ewald with a grid spacing of 1.0 Å. 2 fs is the timestep employed in all simulation with long-range non-bounded interactions evaluated every two steps.

3.6 Anton 2 supercomputer increases speed and size of micro molecular dynamics simulations

Anton 2 machine located at the Pittsburgh Supercomputing Center (PSC) is used in Aim 2 of this thesis. Anton 2 machine is a second-generation, special-purpose, massively-parallel supercomputer built by D. E. Shaw Research [134] for performing Microsecond MD simulations of proteins. Anton 2 machine is ten times faster than its predecessor, Anton 1, and it greatly expands the reach of molecular simulations. Anton 2 is the first platform to achieve MD simulation of multiple microseconds of physical time per day for systems with millions of atoms. For instance, Anton 2 easily simulates a 24K particle system at a rate of 85 μs/day. This is 200 times faster than any other general-purpose supercomputer.

For this thesis, a total of 128 cores are available on Anton 2 machine and used to

perform the simulations of Aim 2. Based on the benchmarks provided by Anton 2, it is estimated that the TCR-pMHC complex requires 4.53 machine-day to complete the simulations; $24us * 1/5.3$(machine-day/us). This makes a total of 230000 MD simulation units on Anton 2 machine. The simulations are performed using NPT ensemble with position restrains and SHAKE constrains. The CHARMM36 force field is used and the system is composed of protein, ions, and water. Water is modeled with standard TIP3P models. When fully solvated and counterion balanced, the system has systems have about 210,000 atoms. 2.5 femtosecond time step is used in the simulations with long-range non-bonded interactions evaluated every two steps and with thermostat applied every 100 steps and save the coordinate files each 250 ps.

3.7   *Visual Molecular Dynamics*

Visual Molecular Dynamics (VMD) [135] is developed at the Theoretical and Computational Biophysics Group at the University of Illinois and it is a program for modeling, visualization, and analyzing MD simulations of biomolecules. VMD is capable of displaying the structure of standard Protein Data Bank (PDB) files, it does not have limits on the number of particles or number of simulation frames, and it offers support for multicore processors and GPU accelerated computers. VMD provides a wide variety of methods for rendering and coloring a protein such as the space-filling spheres (CPK model), licorice bonds, ribbons, (new-) cartoon drawings, and many others. VMD also produce quality image and movies by using a number of popular ray tracing and image rendering packages, including POV-Ray and Tachyon. It usually acts as a graphical front end for an external MD program by animating a molecule through the simulation and it also offers stereo display capabilities. VMD supports multiple sequence alignment and it has a unified bioinformatics analysis environment that integrates and interact well with many MD package.

The images and movies on this thesis are produced using VMD [135] with its tachyon parallel ray tracing system [136] and the secondary structures in the proteins are computed

from the atomic coordinates using the STRIDE program incorporated into VMD [137]

3.8   Limitations of molecular dynamics simulations

The timescale is one of the limitations of MD simulations; structural changes in proteins span easily from nanoseconds to milliseconds or even seconds. MD is computationally constrained to the range of ns-to-micro seconds since it requires a time step of 2 fs for numerical stability and this time step cannot achieve milliseconds yet using current computing power. Million-to-quadrillion operations are needed to reach longer simulation time and this requires a substantial amount of computer power not available currently. Non-bonded interactions are very expensive and they influence most of the calculation since they act between every single pair of atoms. For instance, in a system consisting of N atoms, the number of non-bonded calculations is proportional to N^2. Atoms separated by more than a cutoff distance could be only ignored for van der Waals interactions since these forces fall off quickly with distance. However, for electrostatics interactions, the forces fall off very slowly with distance and they require a special treatment since they are longer-range interactions. MD simulation timescales still remain a big challenge even though recent advances in computer power is enabling microsecond MD simulations. The ultimate goal is going from microsecond to longer timescale simulations (seconds) and this is still an active research area that must involve a collaborative effort in algorithmic improvements, parallel computing, and specialized hardware.

It is almost impossible to solve the equation (1) analytically for proteins since they are large systems that typically have a vast number of atoms. MD simulations use numerical integration to solve Newton's second law of motion; however using the numerical techniques produce problems if they are employed over a long period of timesteps. The results become mathematically ill-conditioned since the numerical integration schemes generate cumulative errors. The error can be minimized if a proper algorithm and parameters are selected but they can not be eliminated entirely.

Force field accuracy is another limitation since molecular mechanics force fields are inherently created by many approximations. Although they have been improved substantially over time, they still have limitations and no all force fields are experimentally tested and validated. In addition, during standard MD simulations, the force fields cannot describe critical events such as the formation or breakage of covalent bonds; proteins, in MD simulations, have all their covalent bonds well-defined form using a typical function. This is different from proteins that some of their covalent bonds do form and break more frequently during the cell life, for instance, disulfide bonds between cysteines are usually formed. Electron transfer events cannot be described with the MD force fields such as acidic or basic amino acid residues experience the loss or gain of protons.

MD simulations are improving and reducing the amount of computation per timestep by using faster algorithms. PPPM, for instance, is a fast approximate method to compute electrostatic interactions and other methods allow to evaluate some force field terms every other time step. Reducing the number of time steps would require freezing some very fast motions of atoms such as certain bond lengths. Another important research area to reduce the simulation time is making events of interest to occur rapidly or making the simulation reach all conformational states more quickly. For example, one might apply steered forces to pull a protein from its ligand or push the simulation away from states it has already sampled.

In SMD simulation, the main drawback is that the pulling velocity is still several orders of magnitude from experiments (10^2 to 10^3). Thus, a direct comparison between simulations and single-molecule experiments was not possible because of the large difference in pulling velocity exerted on the protein. However, the gap between experiments and computer simulation was recently closed using high-speed force spectroscopy [138]. It was demonstrated that manipulating proteins experimentally using similar speed as in simulations provides a dissociation and unfolding pathway consistent with the prediction of SMD simulations, hence allowing a direct comparison.

Parallelizing the simulation across multiple computers is still the main technique to reduce simulation time. The computation is divided among CPU nodes and each node is associated with atoms in one spatial region. Multiple CPUs also require efficient communication protocol and improving the message passing library and the algorithms can reduce the communication requirements. Another more expensive and time-consuming route is to redesign and create computer chips to make this computation run faster. Graphics processor units are now used for MD simulations. They pack more arithmetic logic units on a chip than CPUs, and give a substantial speedup, although parallelizing across multiple GPUs is still a challenge. Several projects, such as Anton machine, have designed chips, especially for MD simulation.

Another limitation of MD simulation of biological molecules is that it requires a crystal or NMR structure as the initial configuration for the simulation. This external limitation does not arise from the method itself but it depends on having the structure of the protein or complex of interest. This is an important limitation if the structure is not available since the folding of large proteins from their linear sequence will be almost impossible to be accomplished by using MD simulation. Although there is some success generating initial structures of proteins using homology models, they are unreliable in predicting the new conformational changes a protein would undergo if a sequence of residues that do not align with the template is inserted or deleted. In addition, Homology model or docking techniques are not very useful in predicting the structure of a protein binding its ligand even though both receptor and ligand structures are available. Thus, a crystal or NMR structure is initially required for performing the MD simulation of biomolecules.

CHAPTER IV

THE MOLECULAR BASIS OF MUTANT EPITOPES DECREASING THE ACTIVITY OF CYTOTOXIC T CELL

4.1 Abstract

The main goal of this chapter is to understand how mutations modulate the TCR-pMHC interaction that leads to a decrease in the cytolytic activity of T cells. Thus it is proposed to study atomistically the force-induced dissociation of TCR-pMHC interaction by means of computer simulations. Adaptive immune receptors binding and recognizing their ligand remains as one of the most interesting mechanisms and hot topics in immunology and at the same time, they are still highly debatable and controversial. This chapter studies the force response of a TCR reading out a pMHC in the context of molecular-antigen recognition of viral epitopes. It is demonstrated for the first time that catch bonds formed experimentally between a TCR and a pMHC can be predicted a priori by using *"in silico"* biology. The best predictor of catch bonds for the TCR-pMHC interaction is shown to be the total number of H-bonds at the binding interface and this is in good agreement with the experimental studies. It is described the molecular characteristics and requirements for a TCR to form catch bonds in the context of antigen recognition and how mutant peptides abolish it. It is then explained the mechanism of how the TCR changing its initial docking orientation modulates the peptide conformation inside the MHC binding groove. Finally, it is proposed a molecular mechanism of how TCRs read featureless pMHCs in the context of viral infection: RA14 uses a molecular lever to form an *"in silico"* catch bond to uncover the peptide from the featureless pMHC and then the RA14 is able to read and recognize the antigenic peptide.

The following abbreviations are used in this chapter: HCMV, human CMV; HLA2, HLA-A*0201 allele; NLV, NLVPMVATV; RA14-NLV-A2, RA14 TCR-NLV-A2 complex; Hydrogen bonds, H-bonds; Protein-protein interactions, PPI.

4.2 Background

The topology of pMHC class I complex not only influences TCR repertoire diversity but also play a pivotal role in TCR recognition. The structural accessibility of the pMHC class I complexes presenting peptides can be classified as "featureless", "featured" or "bulged" [86]. Fig. 12 shows a simplified cartoon of the structural accessibility and reports the types of specific TCR repertoire that is selected after infection. Featureless peptides are 8-10 residues but have little to no solvent-exposed amino acids with prominent side-chains; thus the peptides tend to look more similar to self and the T cells that are capable of recognizing these self-looking peptides are deleted from the repertoire by thymic negative selection [86, 139, 140, 141]. Featured peptides are also 8-10 amino-acid long but with 1 to 3 prominent side-chains residues being exposed for recognition by specific TCRs. The positive and negative selection in the thymus results in a diverse array of T cells in the naive repertoire that can recognize such pMHC complexes. Bulged peptides, on the other hand, are longer peptides (>11 residues) and due to structural constraints, the peptides are bent in the center to allow binding to the MHC class I. This leads to a prominent bulge that is protruding out of the MHC and limits TCR ligation as the TCR contacts with the MHC α-helices are geometrically constrained and impaired. Few T cells are capable of overcoming these structural constraints, resulting in a T-cell repertoire of limited diversity in the periphery.

Antigen recognition by T cell is a crucial feature for adaptive immunity and clearance of viruses. Since the TCR is the only T-cell surface molecule that senses antigens, a TCR is required to ligate a free pMHC displayed on the APC to initiate a cascade of T-cell signaling events that lead to activation. The TCR-pMHC interaction not only plays a pivotal role in

Peptide–MHC topology	Featureless	Featured	Bulged
Specific TCR repertoire	Restricted	Diverse	Restricted
Possible number of TCR clonotypes in naive T-cell repertoire capable of recognition	Few	Many	Few
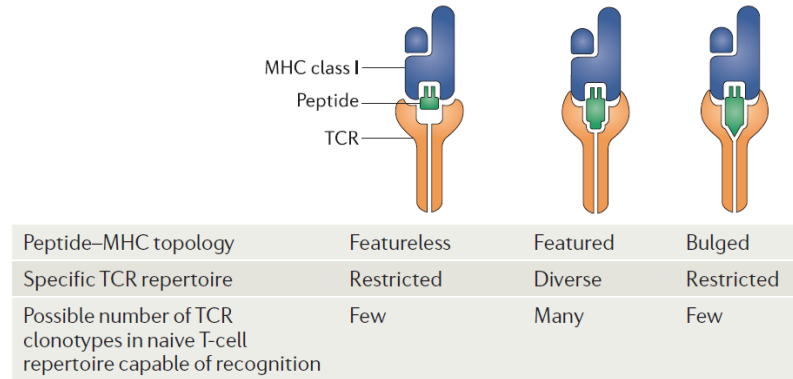
Figure 12: The topology of peptide-MHC class I complexes can influence TCR repertoire diversity. A schematic of pMHC class I complexes that are "featureless", "featured" or "bulged", and the types of specific TCR repertoire that are selected after infection. Featureless peptides are 8-10 residues but have few or no solvent-exposed amino acids with prominent side-chains; thus the peptides tend to look more similar to self and T cells that are capable of recognizing these peptides are deleted from the repertoire. Featured peptides are also 8-10 amino-acid long but with few prominent side-chains residues being exposed for recognition by specific TCRs. The selection process in the thymus results in a diverse array of T cells. Bulged peptides are longer peptides (>11 amino acids) and the peptides are bent in the center to allow binding to the MHC class I. This results in a prominent bulge that protrudes out of the MHC and limits TCR ligation. There are few T cells that can overcome both the structural constraints and this results in an immune T-cell repertoire of limited diversity. Image is taken from ref [86].

activation but also in T-cell selection, development, differentiation, fate, and function. The question of what governs the TCR interacting with different pMHC to trigger intracellular signals is still unanswered. It has long been hypothesized that the TCR-pMHC binding parameters determine the subsequent T cell response [90] and mechanical force plays a critical role in T-cell activation [92]. Experimentally, a force, applied across the T cell ligated to an APC, triggers intracellular calcium signaling–a known hallmark of early T-cell activation– only if a cognate pMHC is detected but not for irrelevant pMHCs. Force spectroscopy experiments such as a Biomembrane force probe (BFP) permit the manipulation of single molecules under physiological conditions. The BFP is capable of producing pico Newton forces necessary to dissociate the ligand from its receptor and sensitively measure its bond lifetime with a high precision. This produces a specific molecular signature for the interaction between a receptor and its ligand called the bond lifetime vs force profile. By

measuring the lifetimes under different forces, the profile determines the effects of mechanical forces on single receptor-ligand bonds. The force exerted on a single bond interaction is kept at constant until the bond dissociates. A bond is called "**slip**" if the average bond lifetime decreases as the force increases. A bond is called "**catch**" if the lifetime increases as the force increases under a certain regime of force. The catch bond would then transition to slip bond behavior as this could be due to forces overpowering the formed bond. Catch bonds are unique phenomena as they exhibit nonlinear responses that prolong the duration of the receptor-ligand interaction.

TABLE 1: *Affinity of RA14 TCR for NLV variants and functional impact on cytotoxic response of the RA14 T cell clone[a]*

	Sequence	K_D (μM)	EC_{50} (nM)
NLV	NLVPMVATV	27.7 ± 2.3	$5.10^{-2} \pm 0.01$
P4G	NLVGMVATV	69.2 ± 1.5	50 ± 3
P4A	NLVAMVATV	56.5 ± 4.0	$5.10^{-1} \pm 0.1$
M5S	NLVPSVATV	n.d.	$5.10^2 \pm 20$
M5V	NLVPVVATV	60.2 ± 2.6	$5.10^2 \pm 50$
M5T	NLVPTVATV	58.2 ± 1.7	$10^3 \pm 30$
M5Q	NLVPQVATV	57.5 ± 1.3	$10^2 \pm 20$
T8A	NLVPMVAAV	n.d.	$>10^4$
T8S	NLVPMVASV	42.7 ± 8.4	$10^{-1} \pm 0.05$
T8V	NLVPMVAVV	n.d.	$>10^4$

[a] Dissociation constants (K_D) were determined by steady-state SPR experiments for the interaction between RA14 TCR and HLA-A*0201 loaded with either the native viral NLV peptide or one of the nine different peptide variants (the mutated amino acids are underlined). The ability of a T cell clone bearing the RA14 TCR to recognize HLA-A*0201-positive target cells loaded with each peptide variant was evaluated by determining the peptide concentration necessary to achieve half maximal target-cell lysis (EC_{50} at an effector to target ratio of 10:1).
n.d., No binding detected.

Figure 13: The native crystal structure of TCR-pMHC [21] reveals that peptide residue Pro4^P, Met5^P, and Thr8^P (A, B, C) are interacting with the TCR. Residues Met5^P and Thr8^P are important since their side chains are longer, reactive, and flexible for adopting any changes. The structural analysis (B, C) reveals that Met5^P side chain is solvent-exposed while Thr8^P is barely exposed for interacting with the TCR; the functional study (Table I) also suggest that those two residues are hot spots.

Using the powerful BFP, it is demonstrated currently that the best predictor of T cell response is the bond lifetime of TCR-pMHC interactions [92]. The catch bond is clearly observed for potent pMHCs compared to the weak peptides. Indeed, the TCR-pMHC catch bonds are less pronounced for weaker ligands and are converted into slip-only bonds for antagonist pMHCs. Note this behavior is different from the shorter bond lifetimes measured in the absence of force (or so-called zero force) for the potent pMHCs compare to the bond lifetimes of weaker ligands. The idea of force regulation the TCR-pMHC interaction

is consistent with the fact that cells are constantly not only experiencing forces from the environment but also from the cellular components that made the membrane along with the intracellular components of the cytoskeleton. Thus, a role of mechanical force in TCR triggering has long been suspected and it is proposed that catch bonds may serve as a mechanism for the force to regulate TCR-pMHC interactions. This plethora of information was collected via a pMHC engaged to a TCR on a naive transgenic T cell. Here it is proposed to study persistent viral infections involving human TCRs; this is a more physiologically relevant system.

Persistent viral infection uses latent strategies to escape the immune response. The virus epitope is usually expressed in low-to-moderate levels when it is presented by the APC, and its pMHC landscape tends to look more similar to self-peptides ("featureless"), as observed in the crystal structures of (1) Epstein-Barr Virus (EBV) LC13 TCR bound to an HLA-B*08:01-FLR [142]; (2) Influenza JM22 TCR bound to an HLA-A*02:01-GIL [143]; (3) HIV TRAV17/TRBV7-3 TCR bound to an HLA-B51-TI8 [144]; (4) Human Cytomegalovirus (HCMV) R14, C7, and C25 TCRs bound to an HLA-A*02:01-NLV [21, 85]. Thus, the virus survives for long-term due to corresponding low levels of CTL activation. In addition, viruses such as HIV-1 and HCV adopt a more improved mutation strategy. This is searching evolutionary the best mutation that provides the "survival of the fittest" that leads to a decreased response of the immune system. Viruses do not have the luxury to modify their sequence as will since most mutations would render them dysfunctional; thus the best mutation is most of the time a single-residue replacement in the peptide sequence. This Aim employs human CMV to model persistent viral infection. CMV is an excellent human infection model since: 1) It is used to study "public" vs "private" TCRs due to their diverse repertoire; 2) It is not only a life-threatening in compromised individuals, but also it has been shown recently that CMV infection may impact and decreases host immune responses in healthy individuals over time; 3) CMV continues to have a tremendous impact in organ transplantation as it affects allograft function and

increase patient morbidity.

MD simulations are employed to understand how local conformational changes at the interface of TCR-pMHC enable T-cell activity. When interactions are measured in experiments, both TCR and pMHC are bound to the cell membrane or bead. Although the simulations only include the TCR-pMHC without the cytoplasmic tail and membrane, it is assumed that the ectodomains are attached to an imaginary cell membrane. α and β chain of TCR and α_3 domain of MHC are constrained at their C-terminal for this purpose. In the experiment, MHC is of a mutant type (MT) and it is not able to bind the CD8; hence the contribution of CD8 to the TCR-pMHC interaction is completely suppressed. Therefore, the molecular model only includes the TCR and its pMHC without any co-receptor to mimic the BFP experiments. CMV infection is employed to model persistent viral infection using wild-type and altered peptides. The native crystal structures of CMV specific-TCR bound to NLV-MHC [21, 85] reveal that the peptide residues Pro4^P, Met5^P, and Thr8^P are interacting with the TCR. Met5^P is important since its side chain is longer, reactive, and flexible for adopting different conformations. Thr8^P is smaller but it is able to form multiple hydrogen bonds with the TCR. The structural analysis and the functional studies (see Fig. 13 for details) suggest that those residues are indeed hot spots [21, 85].

The hypothesis is that peptides forming catch bonds are detected by the TCR, whereas mutant peptides that present slip-only bond behavior easily decrease recognition (binding affinity) from the TCR. In other words, to escape detection, epitope mutations have to convert their mechanical response from catch-slip to slip-only bond behavior. To test the hypothesis, the *in silico* mutants, M5I and T8V, are proposed and used together with the NLV peptide. Note that M5I is a naturally occurring infrequent mutation in human CMV-seropositive patients[145]. This chapter focuses on performing steered MD simulations. SMD simulations are used to apply an external force to proteins, and it closely represents the manipulation of proteins through Biomembrane force probe, atomic force microscopy

or optical tweezers. Note that the addition of external forces in MD simulations is a computational technique used to reduce the energy barriers and access different conformations and most importantly it mimics most force-spectroscopy experiments.

The equilibrium state is usually observed in the crystal structure, and the others are either rare events or steady states. In conventional SMD simulations, the interacting proteins are pulled apart by the external force following a specific direction and it can be applied to one or more atoms in the proteins. In free MD simulations, on the other hand, the system is simulated without any external force applied to the protein and here it is mostly used to equilibrate the system since MD does not effectively access or sample all possible configurations of a protein during nanoscale simulations. SMD is also a standard methodology to study the unfolding or dissociation pathway of proteins. The simulations are run with the external force being constant and variable over time, and this is accomplished by using constant-force or constant-velocity pulling, respectively. Using computer simulations provides a unique perspective to understand how viral peptides utilize mutation to modify TCR recognition and manipulate molecularly the T-cell functional outcome.

4.3   Results and Discussion

4.3.1   RA14 TCR forms an "*in silico*" catch bond with agonist NLV-HLA2.

This Aim describes the mechanical behavior of an NLV-specific TCR (RA14) interacting with wild type and altered pMHCs using a total of 8.7μs of simulations including all atoms and explicit solvents. Thus far it was shown experimentally that TCRs are capable of forming catch bonds when ligated to agonist pMHCs; however, it still remains elusive if this counter-intuitive phenomenon could be captured using "*in silico*" biology experiments and studied molecularly. For the first time, here it is reported that TCRs forming catch bonds can be predicted by using computer simulation and that a single peptide mutation can convert a catch into a slip bond as it happens in experiments.

Receptor-ligand interactions are highly dominated by hydrogen bonds (H-bonds), salt

bridges, and hydrophobic contacts across their binding site [146, 147] as these interactions densely pack and deeply bury the residues located at the binding hot spot and for the unbinding of a receptor from its ligand, they need to be broken and desolvated. The hot spot residues are energetically important since they tend to cluster into small regions with a low solvent-accessible surface area (SASA) in order to favor the binding [147, 148]. The buried SASA is commonly used as a predicting parameter of protein folding and misfolding and hot spot prediction.

The buried SASA is used here to measure indirectly the contribution of the hydrophobic effect since it buries non-polar surfaces in aqueous solution and provides a driving force for receptor-ligand compaction and strength. The total number of contacts also include the contribution of the H-bonds, the hydrophobic effect, and the salt bridge. It is shown here that the total number of H-bonds is an excellent predictor of experimental catch bonds of TCR-pMHC interactions and they play a seminal role in decreasing their dissociation time as applied tensile forces are increased in the bonds.

It is defined an *in silico* catch bond when **the total number** of H-bonds and van der Waals contacts at the binding interface of protein-protein interactions (PPI) increase as a function of force. The underlying assumption is that by creating new forced-induced H-bonds, the binding interface is strengthened, and this prolongs the bond lifetime. A tighter binding between the TCR and the pMHC also promotes effectively killing of infected cells and accelerates clonal expansion. Increasing the number of H-bonds at the interface and as a result prolonging the bond lifetime is consistent with studies performed in a large data set of PPI structures with documented binding affinities [149]. It is shown that there is a high positive correlation between affinities and the total number of H-bonds at the binding interface of PPI. It is also shown experimentally that only 2 H-bonds formed by two base-pairs survive for more than 0.1s even at high forces ($\approx$60 pN) [150]. Thus forming a few forced-induced H-bonds couple with an increase of their total count at the binding interface could prolong the bond lifetime.

It was shown recently that force produces structural rearrangement of a PPI to induce the formation of several *de novo* H-bonds [151] and it was assumed *a priori* that they are responsible for producing experimentally catch bonds. However, it is unclear from these computational experiments, whether the new H-bonds are formed at the expenses of breaking the existing ones. Or they replace only a few but more H-bonds are produced so that the total replacement is smaller compared to the created ones; thus the *de novo* H-bonds could be responsible for the catch bond behavior. In other words, it is unknown if the total number of H-bonds increases, decreases, or remains the same at the binding interface in the presence of force. For example, consider that a PPI has eight H-bonds in force-free conditions and formed 3 *de novo* H-bonds in the presence of force; however, their formation disrupts and breaks five existing H-bonds. At the end, the total number (six H-bonds) is smaller than the original state (eight H-bonds). In this case, if an experimental catch-bond is confirmed, it may not be only the result of the *de novo* H-bonds or H-bonds alone may not be a good predictor of catch bonds in this PPI.

The total number of H-bonds at the TCR-pMHC is used to predict catch bonds between TCRs and pMHCs and they are determined based on cutoffs for angle (30°) and distance (3.5Å) for donor-acceptor atoms. The angle is considered zero when the atoms are aligned and extended. "OH" and "NH" groups are regarded as donors, while "O" and "N" are acceptors. Fig. 14 shows the total (A) and the average (B) number of H-bonds at the binding interface between the TCR ligated with NLV-HLA2, T8V-HLA2, and M5I-HLA2 in free MD simulations; here the complexes are considered in force-free conditions (zero force). Experimentally NLV, T8V, M5I epitopes are considered agonist, very weak agonist, and antagonist, respectively. In the absence of force, the number of H-bonds does not correlate with the peptide potency. Apparently, weak agonist and antagonist form a similar number of H-bonds when compared to the agonist peptide. The histogram plots are constructed from averaging the MD simulations over time without including the minimization and equilibration phase.
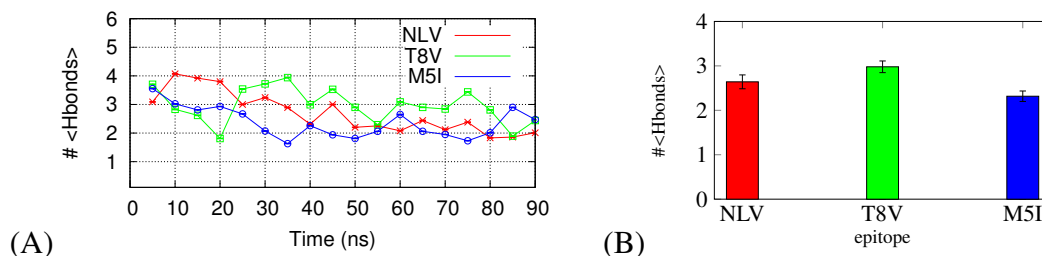
Figure 14: Free MD simulation of TCR-pMHC interactions. (A) Time evolution of the number of H-bonds between TCR and pMHC for wild type NLV, and mutants T8V and M5I in the absence of force. (B) The average number of H-bonds at the binding interface between the TCR and NLV-, T8V-, and M5I-HLA2 in free MD simulations.

Fig. 15 (A) illustrates the time evolution of the total number of H-bonds between the TCR and the pMHC for NLV (red) and mutants, T8V (green) and M5I (blue) peptides; the curves are produced by "block averaging" the number of H-bonds every 10 ns. This post-processing does not change the overall response as they are plotted together with the raw data in Fig. 15 for each peptide ligand, NLV (B), T8V (C), and M5I (D). The complexes are undergoing a "constant-velocity pulling" with a ramping rate of $kv = 7pN/ns$ and this is considered a normal ramping rate in the simulations. The constant-velocity pulling is depicted in Fig. 15 (E). For TCR-NLV-HLA2, the total number of H-bonds increases at the binding interface as force is applied at the protein's C-terminal, and it is simultaneously growing, building and propagating over time; thus, forming an *in silico* catch bond. The overall increase is a result of the formation of new forced-induced H-bonds since its total count is higher. The catch bond, then, transitions to slip bond since the applied force becomes big enough to overpower the new formed H-bonds at the binding interface. The transition also corresponds to the rupture of H-bonds and the decrease in their total number. Fig. 16 visually illustrates the time evolution of the number of H-bonds between the NLV-specific TCR (RA14) and the NLV-HLA2 using average structures at the different time point in the simulation. Clearly, there are multiple new H-bonds when the TCR is forming a catch bond (Fig. 16 D). The interacting residues are not labeled as they are analyzed in the next sections; H-bonds are drawn as the dashed cyan lines. RA14-T8V-HLA2

54
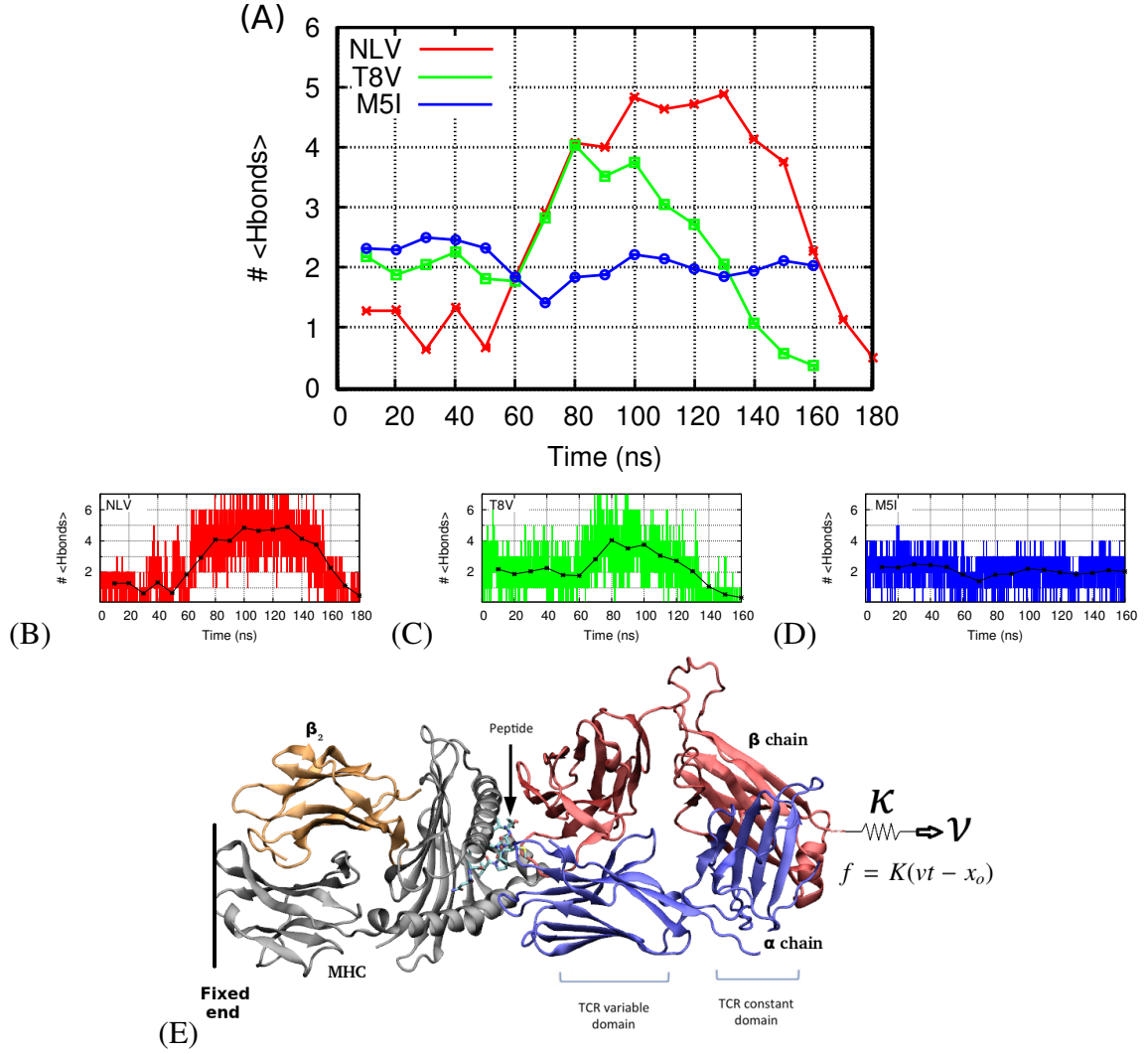
Figure 15: "*In silico*" catch bond of TCR-pMHC interactions. (A) Time evolution of the total number of H-bonds between TCR and peptide-MHC for wild type NLV, mutant T8V, and M5I, in the presence of force. (B), (C), and (D) show the raw data for each peptide ligand with its averaged curve. NLV shows clearly a catch-bond phase represented by an increased in the total number of the H-bonds; there is a small increased for T8V; M5I does not have any increase and apparently displays an ideal bond. The complexes are undergoing a constant-velocity pulling. The number of independent points (or the bin size) of the block averaging method are determined by using the autocorrelation time, τ_f, of the number of H-bonds in the simulation as $N_{sample} \sim t_{total}/\tau_f$. N_{sample} denotes the number of independent samples in the trajectory and if $N_{sample} \gg 1$ would suggest good sampling for any measured variable.

complex initially increases the number of H-bonds at the binding interface; however, its catch bond is weak and small. Thus it is unstable and unable to sustain the pulling force and starts dissociating from the RA14 TCR early. T8V peptide does not allow the force

to build and propagate over time. Initially, T8V and NLV peptides produce a similar catch bond pathway. TCR-M5I-HLA2 complex does not increase the number of H-bonds at the binding interface, its total count is independent of force, and it is considered an ideal bond.

The number of independent points or the bin size of the block averaging method are determined by using the autocorrelation time, τ_f, of the number of H-bonds in the simulation. τ_f measures the time required for the trajectory to lose memory of earlier values. τ_f is used for estimating the number of statistically independent values in a simulation of length t_{total} as $N_{sample} \sim t_{total}/\tau_f$. N_{sample} denotes the number of independent samples contained in the trajectory and if $N_{sample} \gg 1$ would suggest good sampling for any measured variable.
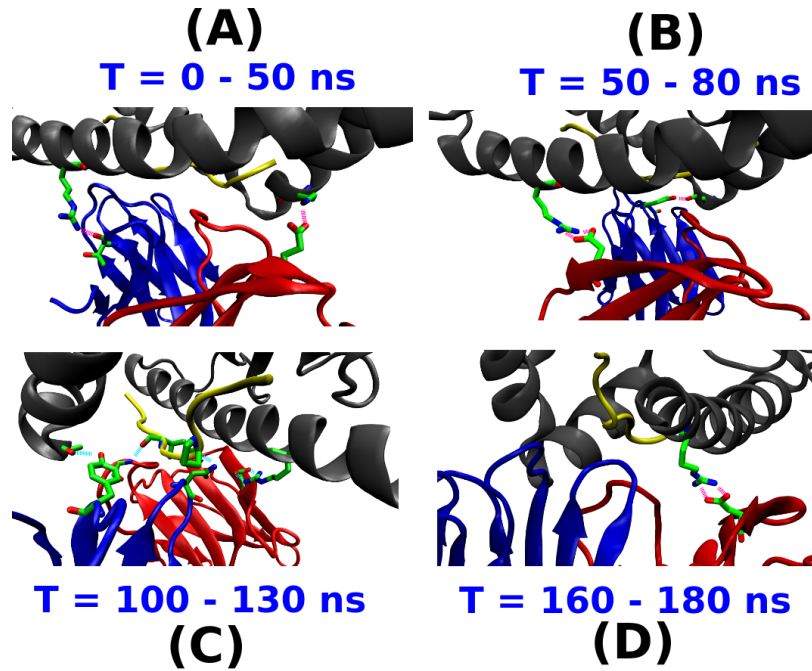
Figure 16: Illustration of the time evolution of the number of H-bonds between TCR and peptide for wild type in the presence of force. There is clearly an increased in the number of the H-bonds in the presence of force under the umbrella of catch bond, $t = 100$ to 160 ns. The H-bonds are colored in cyan at the catch phase.

Similar to experiments, the results may depend on several factors; for instance, in simulations differences could arise due to numerical computing errors defined by physical nature of the machine compiler (hardware); due to numerical methods (software), e.g., the integration step size, PME algorithm, etc; due to initial assignment of velocities (random

number seed) affecting the kinetic energy of the system; as it was completely explained in the previous chapter. Thus repeating the simulations is always performed to evaluate not only the convergence and reproducibility of the results and avoid random events, but also address concerns of whether or not the simulations are depended on any parameters (e.g. pulling velocity, spring constant, etc). The main constraint is that multiple simulations are always desired but with current computing power only a few are possible.
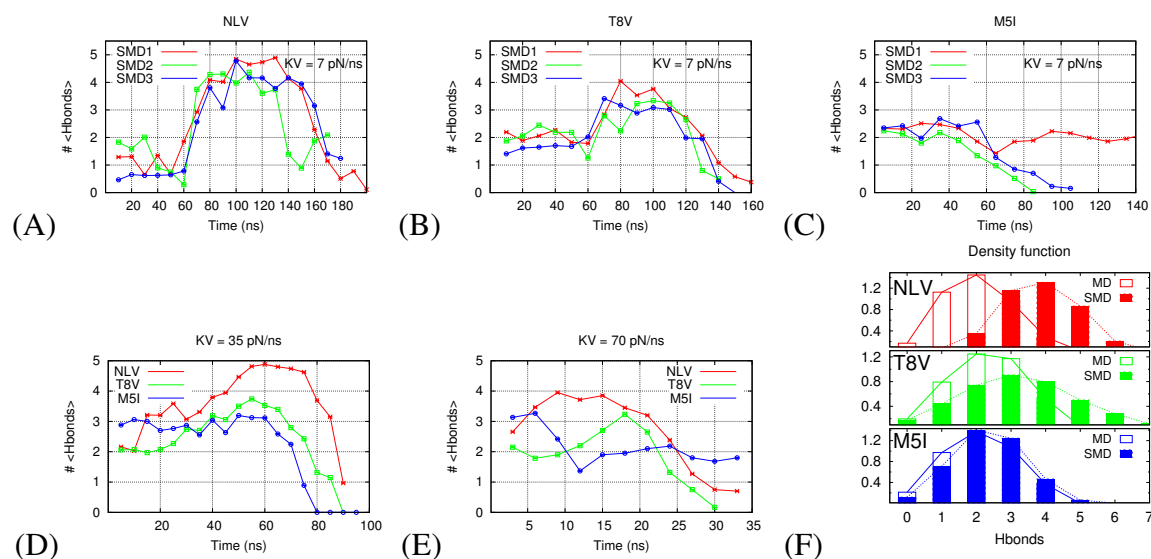
Figure 17: Steered MD simulation of RA14 interacting with different pMHCs.(A-C) The total number of hydrogen bonds of RA14 to NLV-HLA2 and altered peptide ligands(T8V- and M5I-MHC) using three different initial configurations at normal constant velocity. (D-E) Single simulation of RA14 interacting with NLV- and altered T8V- and M5I-HLA ligands at medium and high ramping rates. (F) Histogram of the number of hydrogen bond profile of free vs. steered MD simulation of RA14 to NLV and altered peptide ligands. The results show that: 1) Changing the IC has no significant difference and the H-bond curves are indistinguishable for each epitope. 2) The responses are also not dependent on the pulling parameters as the NLV complex always forms a clear catch bond phase and the T8V shows a very weak, small catch, while the M5I is unsuccessful forming any catch bond. Note that increasing the ramping rate impacts the quality of the catch bond (height and duration) and it is expected to eventually transition to a slip bond at a very high ramping rate.

To evaluate reproducibility, two additional simulations starting with different initial configuration (IC) are performed for NLV, T8V, M5I complexes using a normal constant-velocity pulling; in total 3 simulations per peptide ligand. The three complexes are also

pulled using two different ramping rates. Fig. 17 shows the time evolution of the number of H-bonds between the TCR and the peptide ligand for NLV and mutants, T8V and M5I, at different ramping rates and equilibrated ICs. First, steered MD (SMD) simulations are performed with the three equilibrated ICs only at a normal ramping rate of $kv = 7\text{pN/ns}$ for the NLV (A), T8V (B), and M5I (C). The results show that there is no significant difference, and the H-bond curves are indistinguishable for each peptide ligand. Second using another random IC, the pulling parameters are varied as shown in Fig. 17 (D) and (E) where they are increased to medium ($KV = 35\text{pN/ns}$) and high ramping rate ($kv = 70\text{pN/ns}$), respectively. The responses are also not strongly dependent on the pulling parameters as the NLV peptide always forms a clear catch bond phase, and the T8V shows a small catch, while the M5I is unsuccessful forming any catch bond. It should be noted that as the ramping rate is increased, the quality of the catch bond (height and duration) is decreased and at a very high ramping rate all curves are expected to eventually transition to a slip bond, consistent with the experiments. Fig. 17 (F) shows the probability density function (PDF) of the number of H-bonds for each peptide ligand where all performed SMD simulations are included in the calculation and compared to the free MD simulations. In free MD simulations, all TCR-pMHC complexes form in average two H-bonds; however, in the presence of force, only NLV and T8V complexes shift to five and three H-bonds in average, respectively. The TCR-pMHC complex has a narrow PDF for the NLV, while the complex for M5I has a wider one. M5I has a similar profile in the presence or absence of force. This shows that the overall peptide-ligand responses are not rare or random events and they do not dependent strongly on pulling parameter or initial configuration; thus the simulations are reproducible in general. Fig. 18 (A) shows that there is statistical significance in the number of H-bonds for the NLV and T8V complexes in the presence of force compared the absence of force. There is no statistical significance for the M5I complex in the absence or presence of force.

Last, SMD simulation is performed using a constant-force pulling only for the NLV complex. In this case, the C_{α} atom of the C-terminal residue of the MHC is kept fixed in

Figure 18: Examination of statistical significance using box and whisker plot of the number of hydrogen bond of the number of H-bonds, contacts and, buried SASAs for RA14 interacting with NLV, T8V, and M5I peptide ligands in free MD vs. steered MD simulation using constant-velocity pulling. A Welch two sample t-test is employed reporting p-value $< 10^{-16}$ with a 95% confidence interval for the sample with a statistical significance. No statistical significance is shown in the figure and it is indicated as n.s.

space and the C_α atoms of the TCR's C-terminus are subjected to a constant force in the direction perpendicular to the binding interface. Fig. 19 (A) shows the time evolution of the number of H-bonds between TCR and NLV-HLA2 at different constant-force pulling of 25, 50, and 100 pN. Dissociation is not observed for the NLV complex experiencing forces ≤ 100pN. Rising upward the force in the simulation increases the total number of H-bonds at the binding interface after the force propagates and the duration of forced-induce H-bonds decreases with an increasing force as shown in the Fig. 19 (A). Fig. 19 (B) shows the evolution of the average number of H-bonds between the RA14 and the NLV-HLA2 at different constant forces. The NLV complex forms an *in silico* catch bond (Fig. 19 B): The total number of H-bonds increases at the binding interface while the applied constant force is varied and increased in different independent simulations. There is a transition from a small number of H-bonds (short-lived state) at zero force (Free MD) to a larger number of H-bonds (long-lived state) at big force (around 50 to 100 pN). At the long-lived state, the catch bond reaches a peak in the total number of H-bonds since larger force will eventually break even the strongest H-bonds. For forces ≤ 100pN, each point represents the average of H-bonds from a 200 ns MD simulation of the NLV complex subject to varying constant forces. For forces ≥ 100pN, it is the average of H-bonds until the complex completely dissociates. Although this is a closer representation of a BFP experimental dataset where

the bond lifetime is measured at different constant forces, it is worth noticing that creating such curve (Fig. 19 B) for the mutants is computationally expensive and demanding. For instance using this protocol requires about 2000 ns of simulation for generating a single H-bond profile (# H-bonds vs forces) while pulling at constant velocity requires about 200 ns of simulation per curve. This limitation constrains the simulations to be performed using constant-velocity pulling. It is important to note that increasing the pulling velocity introduces nonlinear effects to the system. Nevertheless, both computational protocols provide consistent and comparable results. Furthermore, the critical force is around 50 to 100 pN for both protocols where the *"in silico"* catch-bond is increased the highest. Fig. 24 (A) shows that there is statistical significance in the number of H-bonds for the NLV complex in the presence of force and that in average it display a well-defined catch bond in the number of H-bonds. Using both protocols, the total number of H-bonds at the binding interface between TCR and pMHC is an excellent predictor for experimental catch bonds.
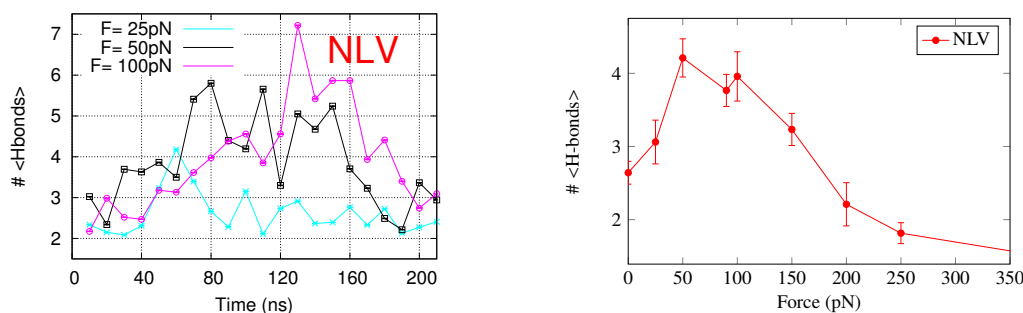
Figure 19: *"In silico"* catch bond using constant-force pulling. (A) H-bonds profiles pulling at a constant force of RA14 to NLV peptide over time for 25, 50, and 100 pN. (B) NLV complex shows clearly a catch-bond phase represented by an increased in the average number of the H-bonds at different constant forces.

4.3.2 The total number of contacts displays a catch-bond behavior

Since receptor-ligand interactions are mostly dominated by H-bonds, salt bridges, and hydrophobic contacts across their binding site, the number of contacts assess their total contribution. Fig. 20 shows that the total number of contacts increases in a similar manner as

the number of H-bonds.

Fig. 20 shows the time evolution of the number of contacts between the TCR and the peptide ligands for NLV, T8V, and M5I, at different IC and ramping rates. The results show that the NLV-HLA2 (A, D, G) increases the number of contacts and forms a catch bond phase, T8V-HLA2 (B, E, H) has a subtle increase in the number of contacts and displays a small catch, while the M5I (C, F, I) is unsuccessful increasing it and mostly display a constant profile. Similarly, increasing the ramping rate impact negatively the quality of the catch bond; however, only the RA14-NLV-HLA2 still displays a catch phase in the number of contacts at medium and high ramping rate, Fig. 17 (J and K). The other peptide ligands become very insensitive when the ramping rate is increased. In addition, the number of contacts is slightly higher in the absence of force. Nevertheless, when force is applied, it decreases and starts increasing when the complex is entering the catch-bond phase. Fig. 18 (B) shows that there is statistical significance in the number of contacts for all complexes in the presence of force compared the absence of force. In average, NLV decreases the number of contacts at the binding interface, while T8V and M5I increase it from the initial value. Overall this quantity is in good agreement with the number of H-bonds; however, it is not very sensitive to exhibit small changes in its value when it is calculated using constant-velocity pulling.

Fig. 21 shows the time evolution of the number of H-bonds between TCR and NLV-HLA2 when the NLV complex is pulled at different constant-force pulling of 25, 50, 90, 100, 150, and 200 pN and their average are plotted in Fig. 22. Increasing the force increases the total number of contacts at the binding interface. The NLV complex forms an *in silico* catch bond, Fig. 22: The total number of contacts increases at the binding interface while the applied constant force is increased in different independent simulations. Fig. 24 (B) shows that there is statistical significance in the number of contacts for NLV complexes at different constant forces. In average, NLV increases the number of contacts from the initial value (zero force) when force is increased. Overall this quantity is in good agreement with
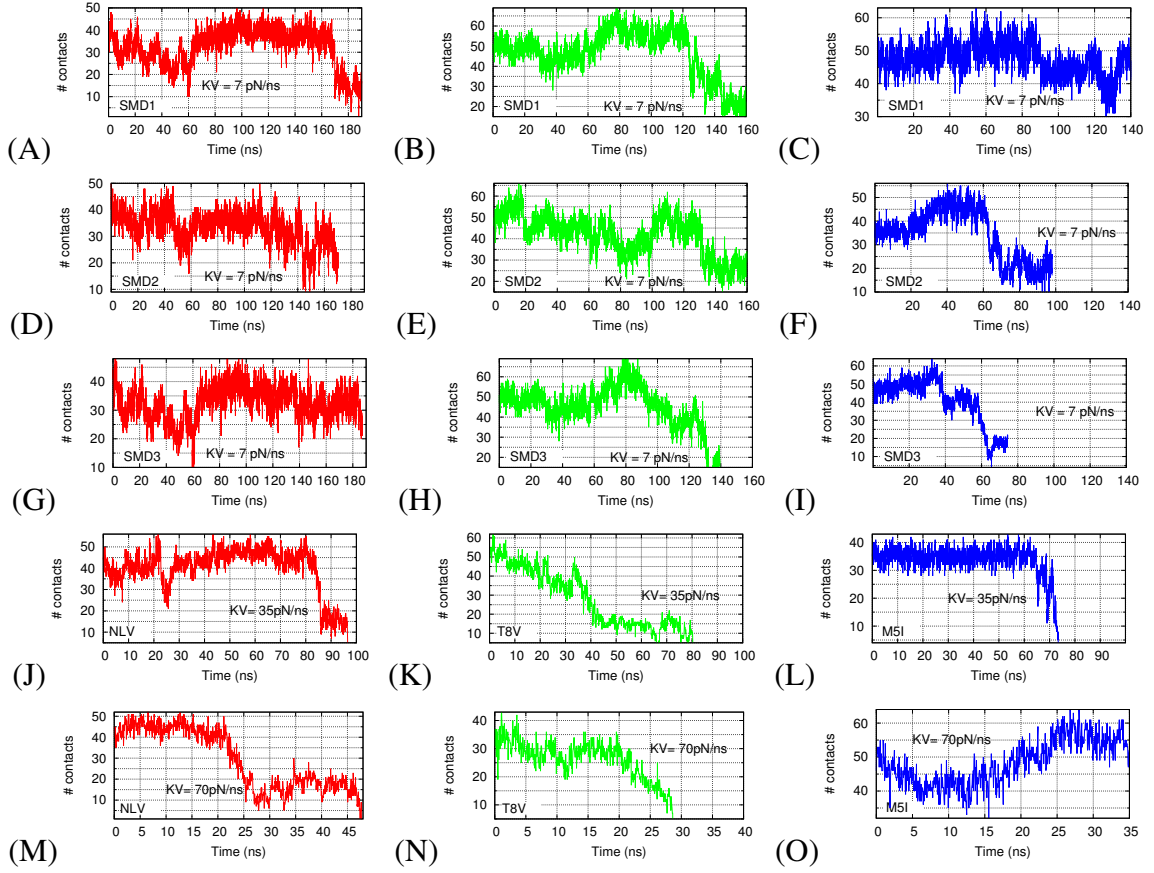
Figure 20: The time evolution of the total number of contacts between the TCR and the peptide ligands for NLV, T8V, and M5I, at different ICs and ramping rates. At the catch phase, NLV and T8V display catch-bond profiles are, with T8V having a small one. M5I mostly display a slip profile.

the number of H-bonds when it is calculated using constant-force pulling. In general, this

quantity is in good agreement with the number of H-bonds.
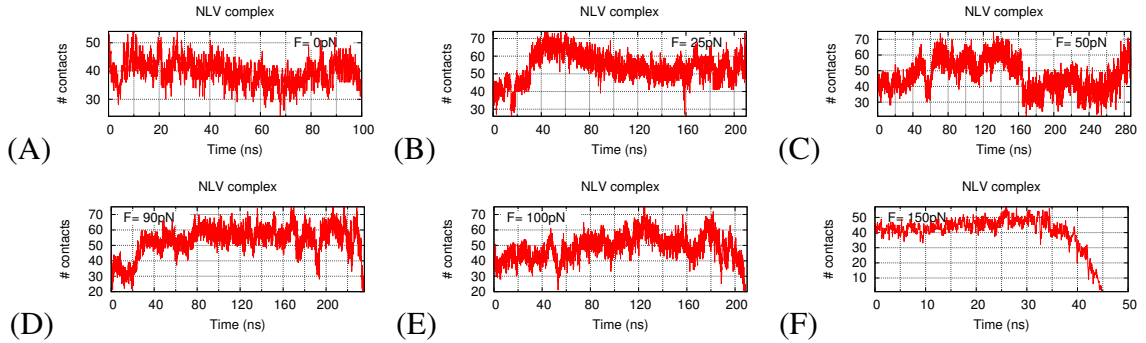
Figure 21: The total number of vdW contacts over time for the TCR-NLV-HLA2 at varying constant forces of 0, 25, 50, 90, 100, and 150 pN.
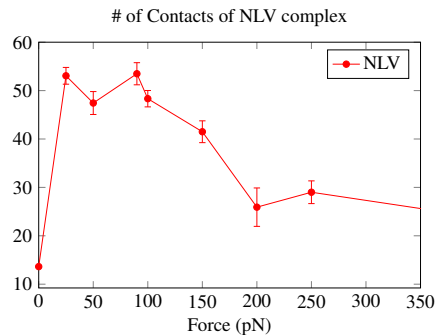
Figure 22: "*In silico*" catch bond in the number of contacts for the NLV complex using constant-force pulling. NLV complex shows clearly a catch-bond phase represented by an increased in the total number of contacts at different constant forces.
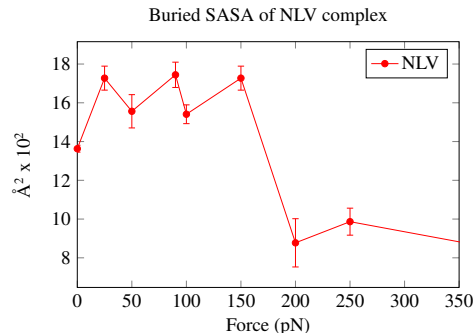
Figure 23: Buried SASA of RA14 interacting with NLV peptide ligand using constant-force pulling. NLV complex shows a subtle increase in the area; however, they are basically constant for $\leq$ 200 pN. The buried SASA is not a good predictor of catch bonds.

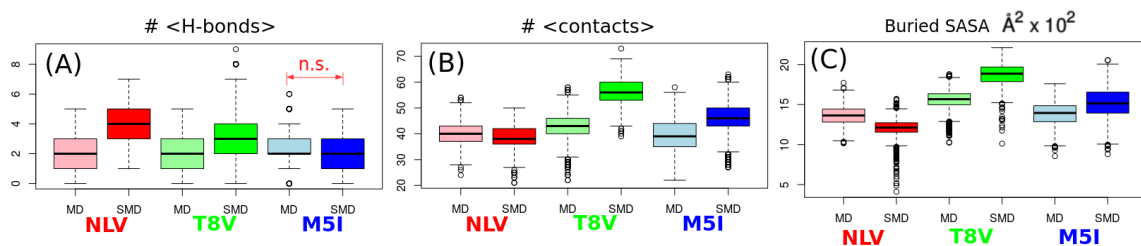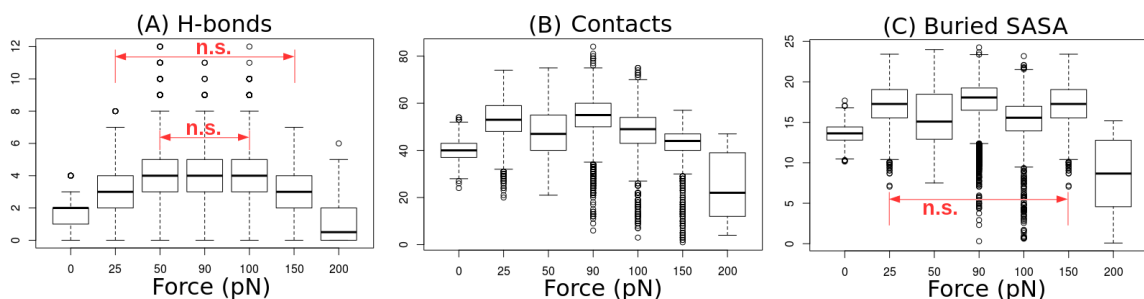**Figure 24:** Examination of statistical significance using box and whisker plot of the number of H-bonds, contacts and, buried SASA for RA14 interacting with NLV peptide ligand in the presence of force using constant-force pulling. A Welch two sample t-test is employed reporting p-value $< 10^{-16}$ with a 95% confidence interval for the sample with a statistical significance. No statistical significance is shown in the figure and it is indicated as n.s.

4.3.3 The buried SASA is not a good predictor of catch bonds when using constant velocity pulling, it becomes constant at the catch phase

Fig. 25 shows the time evolution of the buried SASA between the TCR and the peptide ligands for NLV, T8V, and M5I, at different IC and ramping rates. The results show that the buried SASA of NLV-HLA2 (A, D, G) at the catch phases is roughly constant, T8V-HLA2 (B, E, H) has a subtle increase in the buried SASA and roughly stay constant at catch phase, while the M5I (C, F, I) mostly display a slip profile. Increasing the ramping

rate does not impact the buried SASA profile for the peptide ligands. The buried SASA is very insensitive to changes in the ramping rate. The initial buried SASA is slightly higher in the absence of force. However, when force is applied, it decreases before reaching the catch phase and then it becomes constant until dissociation occurs. Fig. 18 (C) shows that there is statistical significance in the buried SASA for all complexes in the presence of force compared the absence of force. In average, NLV decreases the buried SASA, while T8V and M5I increase it from the initial value. In summary, the buried SASA is not a parameter for predicting catch bonds when it is calculated using constant-velocity pulling.
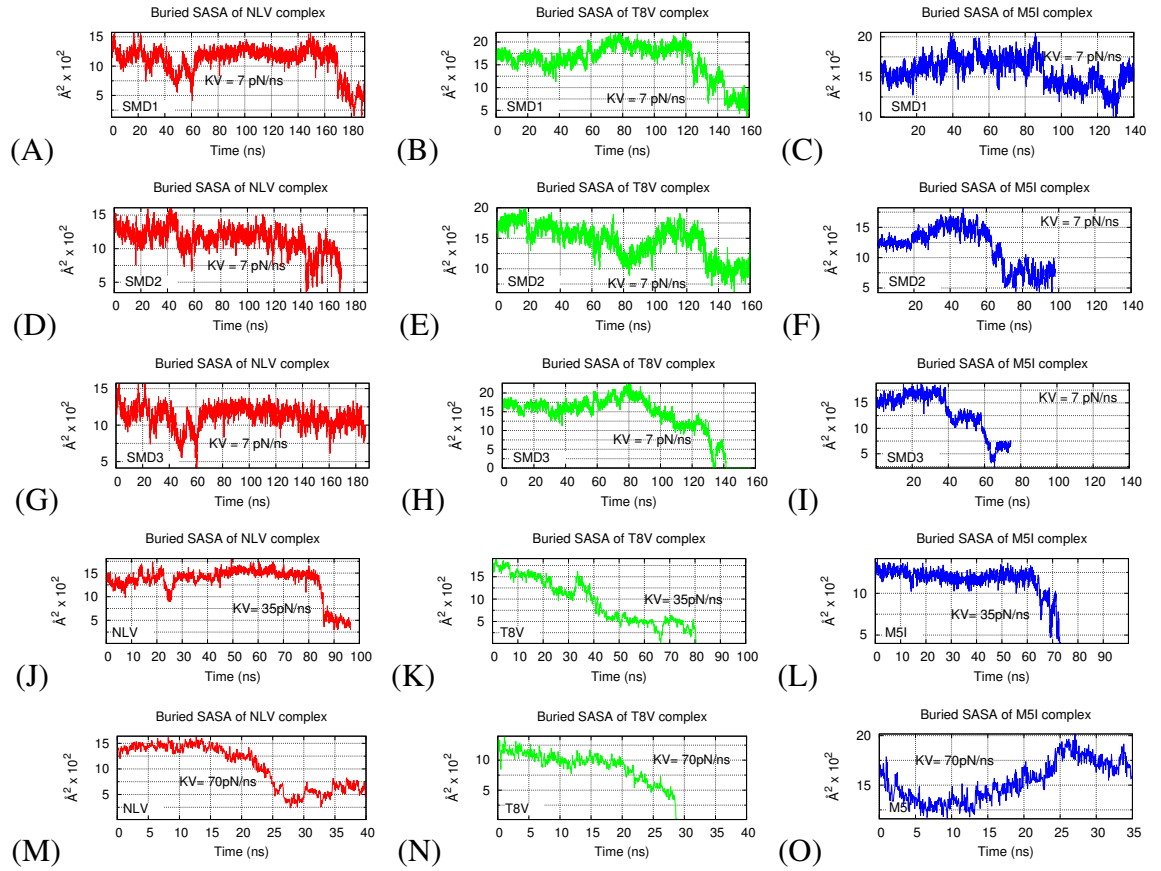
Figure 25: Buried solvent accessible surface area (SASA). The time evolution of the buried SASA between the TCR and the peptide ligands for NLV, T8V, and M5I, at different IC and ramping rates. At the catch phase, NLV and T8V are roughly constant, with T8V having a subtle increase in the buried SASA. M5I mostly display a slip profile.

Fig. 21 show the buried SASA for different constant-force pulling of 25, 50, 90, 100, 150, and 200 pN and its average is plotted in Fig. 23. Although, it is believed that the

hydrophobic effect is a major driving force for binding since it buries nonpolar surfaces from aqueous solution, here it is shown that this parameter is not very sensitive to the subtle changes in the binding interface for the TCR interacting with peptide ligands. For forces ≤200, the buried SASA is roughly constant. This is because the binding interface is small and its changes in the contacting interface are not very sensitive. Buried SASA is not capable of distinguishing slightly buried residues, of which most are not hot spots; the deeply buried ones are usually inside and are hot spots. Overall the buried SASA is not a good parameter to predict catch bonds. Fig. 24 (C) shows that there is statistical significance in the buried SASA for NLV complexes at different constant forces. In average, it increases from the initial value (zero force) when force is increased. Overall this quantity is in good agreement with the number of H-bonds when it is calculated using constant-force pulling.
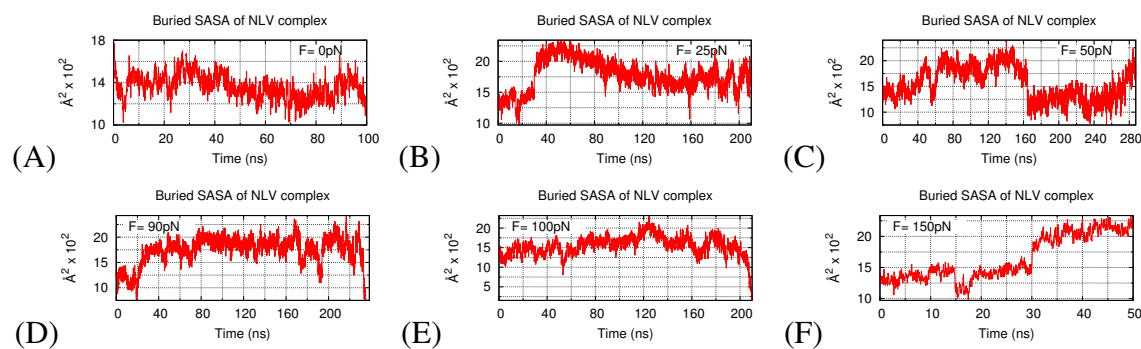
Figure 26: The dependence of buried SASA over time for the TCR-NLV-HLA2 at varying constant forces of 0, 25, 50, 90, 100, and 150 pN.

In Summary, the total number of H-bonds is a good predictor of catch bonds when the complex undergoes either constant-velocity or constant-force pulling. However, the total number of contacts and the buried SASA display only catch-bond behaviors when the complex undergoes constant-force pulling. This is because in a single simulation at constant-velocity pulling, the complex visits all forced-induce states (starting from zero to dissociation) and in average it provides a value that is smaller than the initial one (Fig. 18).

4.3.4 The RA14 changes its binding orientation on the pMHC to form a catch bond.

One important parameter for characterizing TCR recognition of antigenic peptides is the TCR binding (or docking) orientation on the pMHC and it is called the crossing angle. The TCR-pMHC crossing angle is usually calculated between the MHC-bound peptide axis and a line that connects the center of mass of the CDR loops of the TCR α and β chain. As a general rule, TCR CDR1 and CDR2 recognize the MHC α-helices, whereas the hypervariable CDR3 regions interact mainly with the presented and diverse peptide as seen in most TCR-pMHC crystal structures. This topological rearrangement generally produced in most TCRs a diagonal docking angle. The binding orientation is always measured in force-free conditions in experiments and it is believed that when a specific TCR binds to a cognate pMHC, the complex obtains a defined crossing angle and its structural rearrangement does not change after the initial binding as the docking orientation is unique and it provides one specific TCR footprint.

The protein crystal structure shows that NLV-specific RA14 bounds to NLV-HLA2 in diagonal docking mode with a crossing angle of about 35°. The association and dissociation between a TCR and a pMHC is a dynamic and stochastic process where their engagement is subjected to several pico Newtons of mechanical forces during immune surveillance possible through cell movement and by cytoskeletal rearrangements at the immunological synapse after cell migration stops. However, this important component has been omitted in most structural studies of TCRs when characterizing recognition of antigens.

The Fig. 27 shows the docking (or crossing) angle in free MD simulations for all complex. After reaching thermodynamic equilibrium ($\approx$ 20ns), all complex oscillate around an average angle that probably defines a particular footprint on the pMHC and it is constant over simulation time after reaching equilibrium. The angle for the NLV complex slightly decreases from its experimental value and becomes around 25°. A caveat is that T8V and M5I peptide ligands are single *in silico* mutations performed on the wild type NLV complex and they are expected not to change the docking topology significantly. T8V crossing

66

angle changes considerably; however, it is assumed that this is the natural response of the T8V complex and not a computational artifact. The angle of RA14-NLV-HLA2 initially decreases from its experimental value of 35° to 20°; however, later it oscillates between 20° to 30°. In the absence of force, thus, the angle and interface are stable and this orientation and molecular structure are similar to the one observed in the crystal structure. The *in silico* mutants, T8V and M5I, show a different trend. RA14-T8V-HLA2 increases to 55° and oscillate around it, while the RA14-M5I-HLA2 does not change its original angle and oscillates around 30°. For T8V complex, the crossing angle diverges from the NLV trajectory as it initially increases and moves to a different value (55°) in free dynamics.

Fig. 28 shows the dependence of docking angles over time in the presence of force for RA14 interacting with NLV, T8V, and M5I peptide ligands. In the presence of force, the angle starts to linearly increase at 40ns to reach its experimental value (35°) at 60 ns and then it suddenly jumps to 55° in a stepwise manner forming a new state. This second state is very stable and the complex remains in this orientation until dissociation occurs. Fig. 29 (A) and (B) show visually the "force-free" and "forced-induced" orientation, respectively. The complex is aligned according to the MHC to highlight the difference in docking angles for RA14. The unbinding of RA14 from NLV-HLA2 requires that the orientation of the complex returns to its initial force-free state (35°). The transitioning to the second orientation is fast. Surprisingly the second state is very stable even though RA14-NLV-HLA2 is experiencing a huge load. Note that the initial angle is formed in a force-free condition whereas the second is a forced-induced state.

Interestingly, the footprint of RA14 interacting with T8V peptide ligand in free dynamics is similar in angle (50°) to the forced-induced state of NLV complex (55°). However, in the presence of force, the angle decreases to 20° and forms an orientation similar to NLV complex when it is in the force-free state. T8V complex remains here until 100ns. The angle for T8V is reversed compared to the NLV complex. After it reaches 110 ns, the complex moves again to the second state just before the α_3-MHC starts unfolding, and as a

consequence, the angle suddenly increases. This implies that the interface between RA14 and pMHC is unstable and weak for T8V peptide.

For M5I complex, the crossing angle does not change significantly over time in the absence or presence of force and the RA14 footprint is in a similar to the force-free state of NLV complex and it stays until dissociation occurs. After it reaches 120 ns, the α_3-MHC also unfolds. The unfolding happens only with the *"in silico"* mutants. This could potentially be an artifact arising from the fact that those structures are produced by performing *"in silico"* mutations in the wild type NLV complex and they are not structures obtained by protein crystallography. Importantly, the unfoldings are also consistent with two experimental studies describing viral immune escape strategies where mutant epitopes were observed to affects T cell recognition by reducing pMHC-complex stability [152] and allow peptides get out of the groove [153].

A striking result is that not only complexes transitioning to a forced-induced orientation are able to form a catch-bond, but also their discrete shifting in angle coincides with the rise of a catch bond. Although all measured angles in the simulation oscillate around a central value over time, they are considered constant since their amplitude is relatively very small.
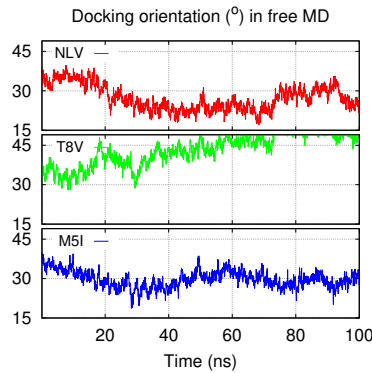
Figure 27: The docking orientation between the RA14 interacting with NLV, T8V, and M5I in the absence of force.
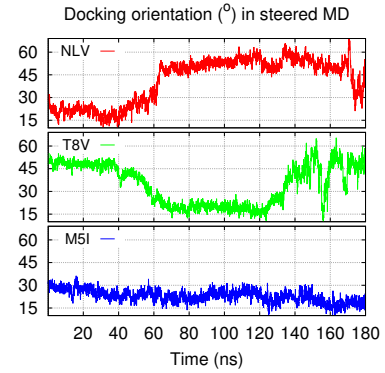
Figure 28: The docking orientation between the RA14 interacting with NLV, T8V, and M5I in the presence of force.

Fig. 30 shows the docking angle over time for NLV, T8V, and M5I complexes at two different IC and two ramping rates. The two additional ICs are pulled at a normal ramping
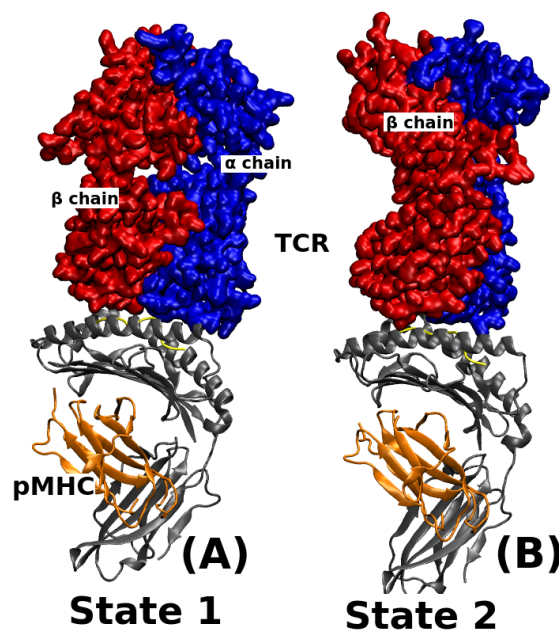
Figure 29: General view of the change of RA14 TCR binding orientation on the pMHC to form a catch bond. (A) Initial state similar to the crystal structure. (B) The force-induced, the TCR has rotated over the binding site. The complex is aligned according to the MHC to highlight the difference in docking angles for the TCR. The unbinding of the TCR from the pMHC requires that the orientation of the complex returns to its initial force-free state (35°). The transitioning to the second orientation is short and fast. The second state is very stable even though the complex is experiencing a huge load. Note that the initial angle is formed in a force-free condition whereas the second is a forced-induced state.

rate. Fig. 30 shows the angle for NLV (A, D), T8V (B, E), and M5I (C, F) complexes. Fig. 30 shows simulations pulled at medium (G, H, I) and high (J, K, L) ramping rate for the three peptide ligands. The color coding is red, green, and blue for RA14 ligated with NLV-, T8V-, and M5I-HLA2, respectively. As the ramping rate is increased for the NLV complex, the docking angle is affected and starts decreasing; this is consistent with the experimental observation that high ramping rates transform catch to slip bonds. A high rate does not allow the complex to move to the second state; the protein dynamics becomes very fast and it is mostly dominated by the rate that introduces nonlinear effects to the system. A closer look to the atomic structure of T8V complex reveals that residue $V8^P$ (mutated from Threonine) is unable to form any stable H-bond with the RA14 and thus it loses this anchor point, originally seen in the wild type. RA14 is initially anchored to this

residue and if removed, RA14 is able to rotate around the binding axis and it can either increase or decreases its docking angle. Note that VAL (T8V) is unable to form H-bonds so it is not anchor point. M5I complex does not change its docking and it is invariant to the presence of force. The results are in general consistent and there is no significant difference between different runs for the same complex. The NLV complex always shows a stable forced-induced angle and the T8V shows an unstable second phase, while the M5I is unsuccessful changing its initial state.
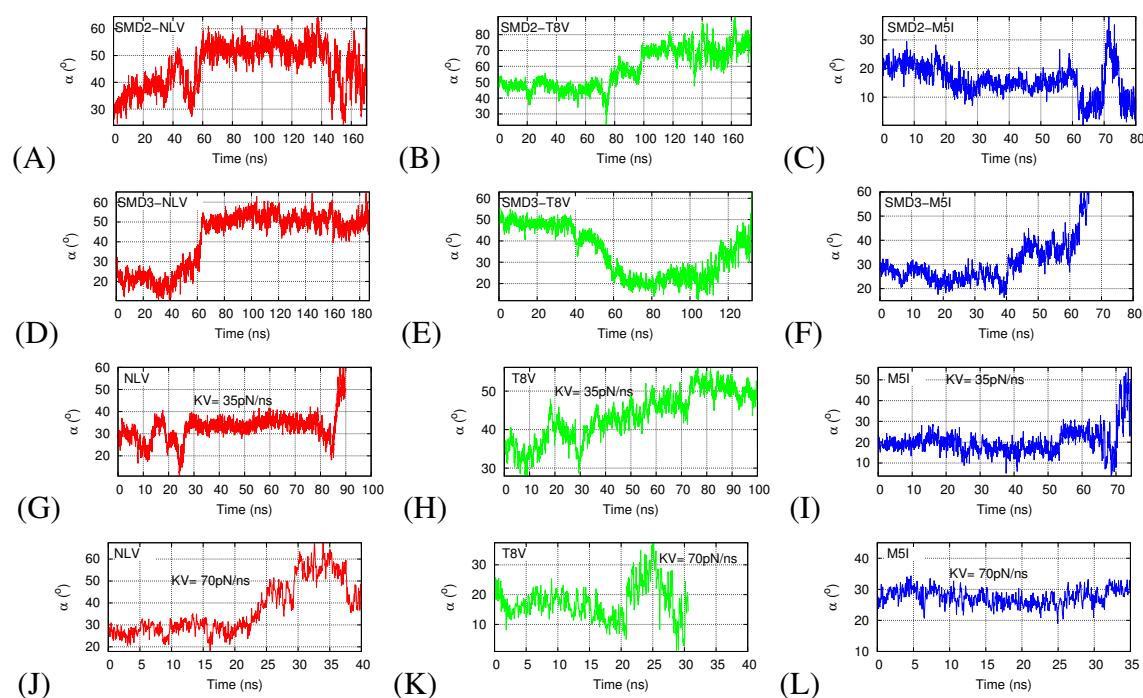
Figure 30: The docking orientation over time for NLV, T8V, and M5I complexes at two different IC and two ramping rates in the presence of force. The color coding is red, green, and blue for TCR ligated with NLV-, T8V-, and M5I-HLA2, respectively. As the ramping rate is increased for the NLV complex, the docking angle decreases; this is consistent with the experimental observation that a high ramping rate will transform a catch into slip bond. A high rate also does not allow the complex to move to the second stable state as the protein dynamics becomes very fast and be dominated by the high rate. In T8V, residue $V8^P$ is unable to form any stable H-bond with the TCR and thus it loses this anchor point. NLV is anchored at least in two residues, Met5^P and Thr8^P and Without them, it is free to rotate around the binding. M5I complex does not change its docking. The results show there is no significant difference between different simulations for the same epitopes. The NLV peptide always shows a stable second angle and the T8V shows an unstable second phase, while the M5I is unsuccessful forming any stable state.

70

Fig. 31 shows the dependence of docking angles over time for the RA14-NLV-HLA2 pulled at constant forces of 25, 50, 90, 100, 150, and 200 pN. At lower ($\leq$ 25 pN) or higher forces (>100), NLV complex has only one state and displays a relatively constant angle before dissociation (Fig. 31 A, E, F). This angle is close and similar to the one measured in the crystal structure. In the catch phase, (50 pN to 100 pN, Fig. 19), the complex shifts to a second state by changing the docking angles. 50 pN is considered the peak or critical force since it forms more H-bonds compared to other forces (Fig. 19). At the critical force (Fig. 31 B), the complex initially has a sudden increased shown as a prominent peak (at 45 ns) and then set the docking angle to 40°; this is the second state and it remains at this orientation for at least 60 ns. Importantly the complex not only transitions to the second orientation discretely but also is able to increase in average the total number of H-bonds simultaneously (Fig. 19). After this point (120 ns), the angle starts slowly decreasing; however, the complex resists and opposes the change by returning to the forced-induced state. This is clearly seen around 160 to 180ns where the angle basically moves back to 40°. Passing this region the forces become large and start overpowering the H-bonds. This level of force does not allow the complex to reach the optimal docking orientation to form catch bonds. At 90 and 100 pN, the complex's angle initially moves to 15° and then increases and maintain a relatively constant value (30°) below the optimal orientation; however the complex always remains moving in an upward direction trying exhaustively to increase it as seen in Fig. 31 (C, D). Although the angles fluctuate, it is considered constant since their amplitude is relatively small.

Important to note that measuring the angle is very sensitive to small variation in the binding site and the problem with the docking orientation oscillating widely is character-istic of the manner is defined initially by scientists in the field. Using the peptide's end terminus or defining a line going through the MHC α_2 chain only works if the complex is static and rigid structures. In MD simulations, they are flexible and fluctuate. A better methodology is to use the axis parallel to the MHC binding groove defined by calculating

the principal axes of molecules. This is a well-defined direction and does not fluctuate.
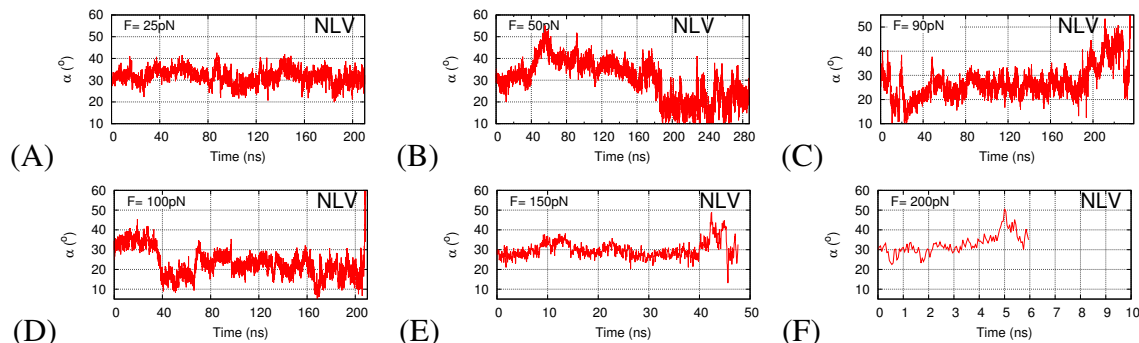
Figure 31: The dependence of docking orientation over time for the RA14-NLV-HLA2 at varying constant forces of 25, 50, 90, 100, 150, and 200 pN. At lower forces ($\leq$ 25 pN) or higher (>100) forces, the complex has only one state and displays a relatively constant angle before dissociation. In the catch region, (50 to 100 pN), the complex shifts to a second state by changing its docking angles. At this orientation, the complex is able to increase in average the total number of H-bonds. Passing this region the forces become large and start overpowering easily the H-bonds and does not allow the formation of an optimal docking orientation. At 90 and 100 pN, the complex's angle initially moves to 15° and then increases and maintain a relatively constant value (30°) below the optimal orientation; however, the complex always remains moving in an upward direction to increase the angle.

4.3.5   NLV peptide has a second conformation associated with the catch bond

The problem of describing structurally catch bonds is tangled since the peptide is experiencing *positional, orientational,* and *conformational* changes at the TCR-MHC interface; however these small changes are inconsistent with the enormous variation in functional outcomes. Having multiple configurations may also modulate the affinity and the duration of the bond lifetime. There is a question if the peptide configuration inside the MHC binding groove is modulated by force when the complex moves to the second state and forms a catch bonds. Or the epitope conformation does not change or it is moving in a randomly. The dynamic nature of the peptide, when is inside the MHC and it is interacting with the TCR, has been never addressed adequately in experiments or simulation.

The root-mean-square deviation (RMSD) measures an average displacement of atoms and it is routinely used to examine if the protein has equilibrated; it can also provide information if (parts of) proteins have experienced any conformational change. Using the

RMSD, the protein behavior over time is compared to a starting point of the simulation. The initial reference and the time-varying target are usually aligned before calculating the RMSD to remove any translational and rotational diffusion (deviation) and explore mostly local conformational changes of the protein.

Fig. 32 shows the RMSD of each peptide in the absence of force (Free MD). In free dynamics, the RMSD of NLV, T8V, and M5I peptides are initially expected to increase since crystal structures are produced by artificially packing the protein inside a very dense crystal; thus, proteins need to relax in order to achieve their native conformation as they are not rigid entities. An increase of 0.5Å is reasonable and expected for native states; however, if larger changes are observed in the RMSD, it implies that a conformational change has occurred. The RMSD reaches a plateau value for the NLV and T8V peptides meaning that they are progressing towards their equilibrium state, where all properties become time independent apart from small fluctuations; this is the thermodynamic equilibrium state. This plateau is not observed immediately for the M5I epitope. M5I displays larger fluctuations and changes in its RSMD from beginning to middle of the MD simulations. Basically, the RMSD measures distance and changing its value means that the atoms are displaced from its original position and they are potentially forming a different configuration. At the new position, the RMSD for M5I epitope eventually displays a constant value (after 50 ns) indicating that its structure has also converged. Finally, the NLV and *"in silico"* mutants, T8V and M5I, are moving towards their equilibrium state in the absence of force.

Fig. 33 shows the RMSD of each peptide in the presence of force (SMD) using normal constant-velocity pulling. The NLV epitope transitions fast in a stepwise fashion to a stable second state defined when the RMSD value reaches 2Å and it remains in this phase inasmuch as the complex is inside the catch-bond region (Fig. 15). This forced-induced conformation is optimal for producing new H-bonds at the binding interface and prolonging the bond lifetime. The T8V epitope initially moves to the second state; however, it is unstable and unable to sustain this conformation and returns back to its original state.

73

The M5I epitope does not have any well-defined state and its conformation is independent of force in the sense that it is always increasing and displaying large fluctuations. The peptides, NLV and T8V, discretely transition to a forced-induced state when the docking orientation changes and it also coincides with the rise of a catch bond, Fig. 15 and 28.
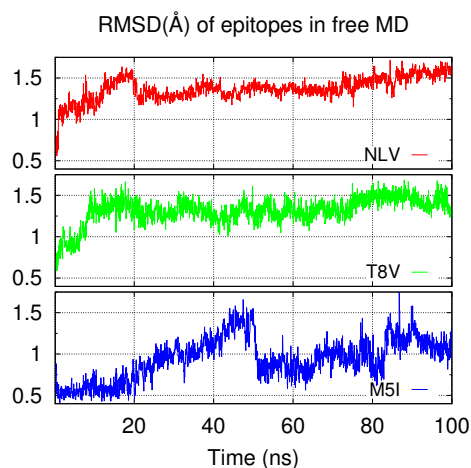
Figure 32: The RMSD of each peptide for NLV, T8V, and M5I peptides in the absence of force.
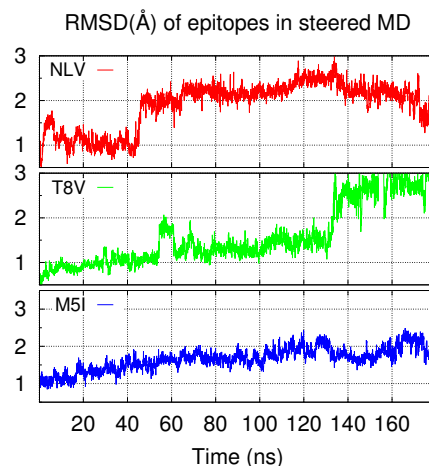
Figure 33: The RMSD of each peptide for NLV, T8V, and M5I peptides in the presence of force.

Fig. 34 shows the RMSD of NLV, T8V, and M5I complexes at different IC and ramping rate. The two independent ICs are pulled using constant-velocity pulling at a normal ramping rate and are shown for the NLV (A, D), T8V (B, E), and M5I (C, F). Fig. 34 shows the simulations pulled at medium (G, H, I) and high (J, K, L) ramp rate for the three epitopes. normal, medium and high ramping rate have been previously defined. The color coding is red, green, and blue for NLV, T8V, and M5I peptides, respectively. As the ramping rate is increased, their peptides are losing the ability to transition to the new state; nevertheless, all NLV simulations display a well-defined, forced-induce state as the RMSD changes to 2Å, excluding only the high rate pulling. Only one simulation for the T8V epitope is able to access the second state. Although one simulation for the M5I epitope reaches a stable state, it either dissociates faster or force unfolds the MHC α_3 domain in the other runs, being is

consistent with experimental studies [152, 153]. The second state formed by the NLV epitope is longer-lived, while the mutants have their second state shorter-lived. Overall there is no significant difference between runs for the epitopes. The NLV peptide always shows a longer-lived, stable second state and the T8V shows an unstable second phase, while the M5I is unsuccessful forming any stable state.
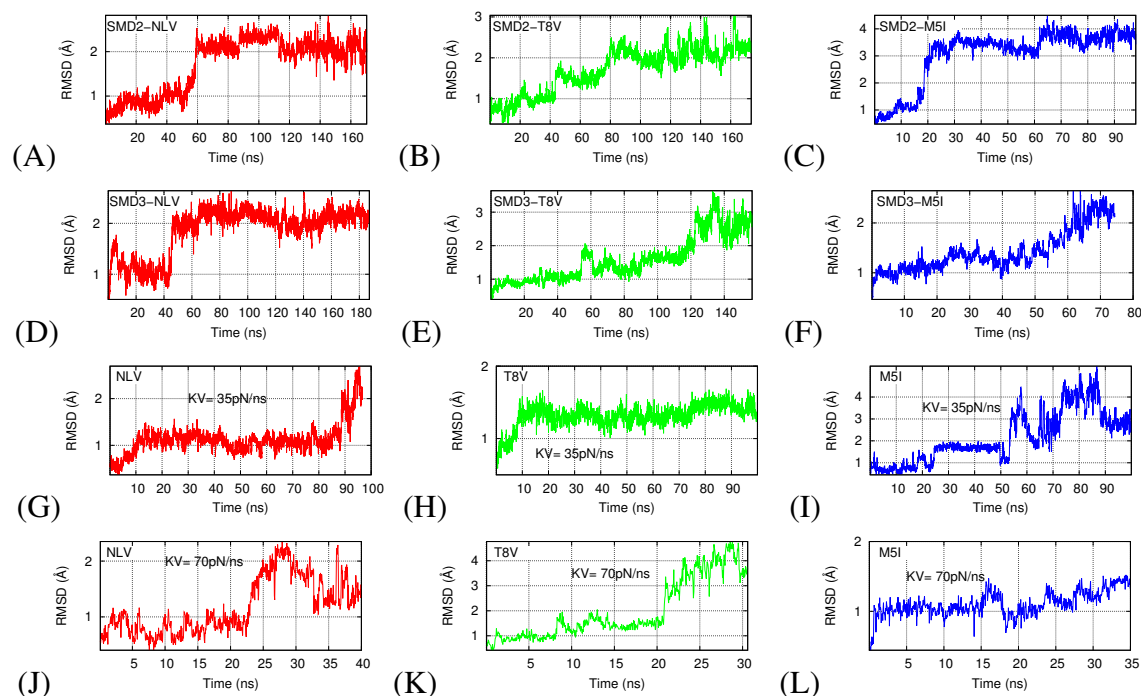
Figure 34: The RSMD for NLV, T8V, and M5I complexes at two different IC and two ramping rates. The color coding is red, green, and blue for NLV, T8V, and M5I peptides, respectively. As the ramping rate is increased, the peptides are losing the ability to transition to the new state. All NLV simulations display a forced-induce state as the RMSD changes to 2Å, excluding the high rate pulling. Only one simulation for the T8V epitope is able to access the second state. Although one simulation for the M5I epitope reaches a stable state, it either dissociates faster or unfolds the MHC α_3 domain in the other runs. The second state formed by the NLV epitope is longer-lived, while the mutants have their second state shorter-lived.

Fig. 35 shows the RMSD for the NLV epitope pulled at constant forces. When force is ≤ 25 pN, the epitope is flexible and does not adopt a clear state and the changes observed in the RMSD correspond to the free movement of its side chains. In the catch region, (50, 90, and 100 pN), the RMSDs increase roughly to a constant value of 2Å and it defines the second conformer. At the critical force, 50 pN (Fig. 35 B), the epitope has a sudden

increased and forms the second state; it remains at this conformation until dissociation and it mostly happens a the end of the simulation. Importantly the NLV not only transitions to the second conformation discretely but also is promoting the increase of the total number of H-bonds at the same time (Fig. 27). At 90 and 100 pN, the peptide also changes its conformation reaching 2Å at the end of the simulation Fig. 35 (C, D). For higher forces (>100), the peptide does not change its RMSD and has only one state until dissociation. This state is similar to the crystal structure, Fig. 35 (E, F).

Fig. 37 are probability density function (PDF) and show pattern distribution of the RMSD values for NLV, T8V and M5I epitopes in free and steered MD simulations, respectively. All simulation are included in the PDF calculation. NLV epitope has only one state (population located at 1Å) in the absence of force (Fig. 37 A); however when force is applied it obtained a second state at 2.5Å (Fig. 37 D). Both states are easily differentiated since they lie within a closely packed interval. The second state has greater height and it means this conformation is the most dominant. There is only one state for the T8V epitope in free dynamics (Fig. 37 B); however, a second conformation appears in the presence of force (Fig. 37 E). This conformer is not the most populated; the dominant is the force-free state for the T8V epitope. Fig 37 (C) and (F) show that the M5I epitope displays only one state in the absence or presence of force, respectively. The distribution of this state is wider implying that the epitope is mostly moving randomly.

4.3.6 Mutants reduce the conformational states of the peptide

In order to understand how local conformational states and internal motion of the epitope change in the absence and presence of force, the per-residue root mean square fluctuation (RMSF) averaged over simulation time is calculated. It quantitatively compares the flexibility and total displacement among peptides per residue. RMSF is also calculated in the standard manner after rotational and translational alignment to the initial structure. All simulation are included in the RMSF calculation. Fig. 38 shows the mobility of the
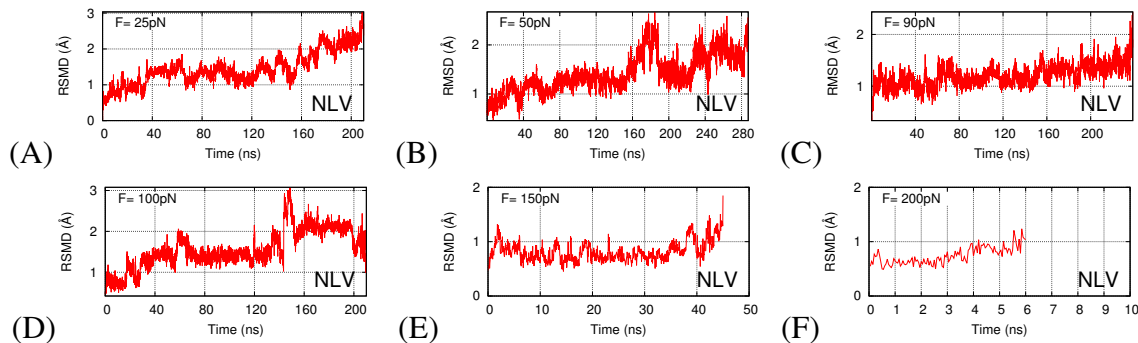
Figure 35: The RMSD for the NLV epitope pulled at constant forces of 25, 50, 90, 100, 150, and 200 pN. When force $\leq$ 25 pN, the epitope is flexible and does not adopt a clear state. In the catch region, (50, 90, and 100 pN), the RMSD increases. At the critical force, 50 pN, the epitope has a sudden increased forming a second state and remains at this conformation until dissociation. At 90 and 100 pN, the epitope also changes its conformation reaching 2Å at the end of the simulation. For higher forces (>100), the peptide does not change its RMSD and has only one state before it dissociates.
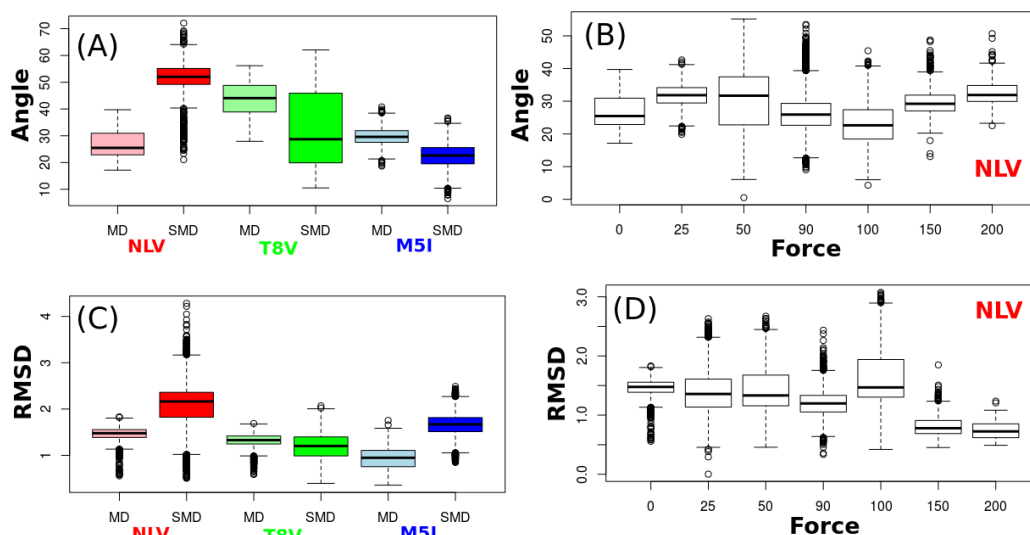
Figure 36: Examination of statistical significance using box and whisker plot of the docking angle and RMSD of the peptide for RA14 interacting with NLV, T8V, and M5I peptide ligands in the absence and presence of force using constant-velocity pulling. A Welch two sample t-test is employed reporting p-value $< 10^{-9}$ with a 95% confidence interval for the sample with a statistical significance.

peptides in the absence and presence of force. NLV epitope (Fig. 38 A) has a higher displacement compared to mutant T8V and M5I in the absence of force. The NLV epitope is more flexible than the mutants and it implies that a flexible epitope could potentially have
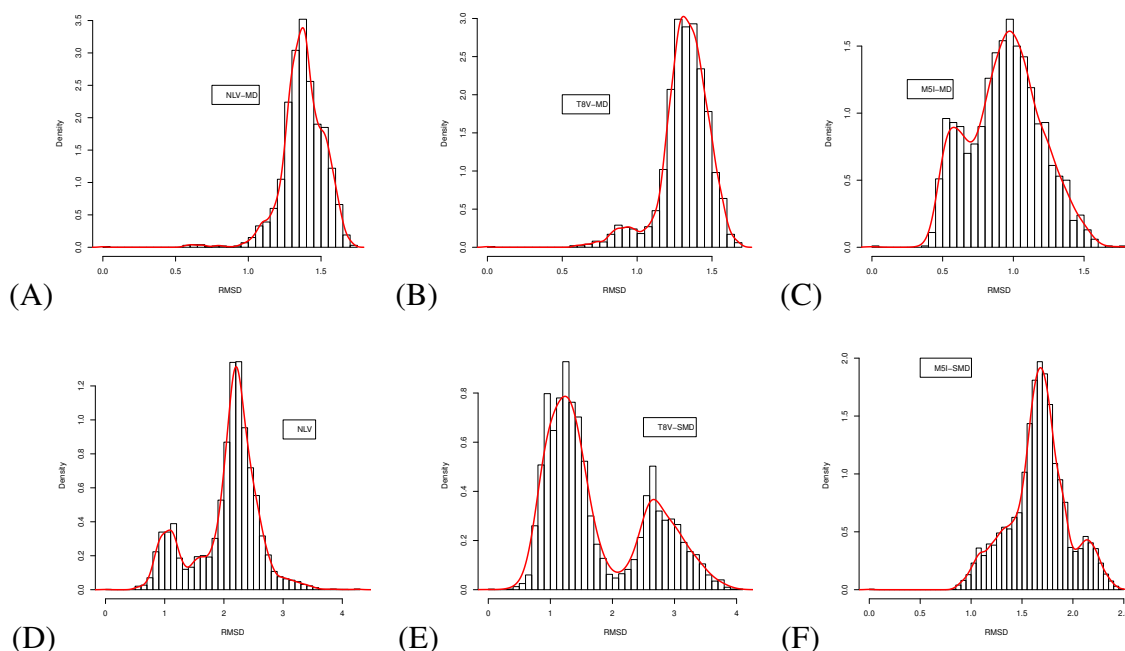
Figure 37: Histograms comparing the peptide conformation in the absence and presence of force. All simulations are included in their calculation. NLV epitope has only one state in the absence of force; however when force is applied it obtained a second state. Both states are easily differentiated. The second conformation is the most dominant as shown by the height of the histogram. There is only one state for the T8V epitope in free dynamics; however, a second conformation appears in the presence of force but it is not the dominant. M5I epitope displays only one state in the absence or presence of force.

more conformational states. However, NLV epitope (Fig. 38 B) experienced a large displacement and is rigidified inside the binding groove in the presence of force. M5I and T8V also experience a small displacement. Force promotes the formation of new H-bonds that constrains the spatial motion and bias the epitope to a specific conformational state. Point mutations disrupt the epitope flexibility, mobility, and dynamic footprint by abolishing the formation of those interactions.

Perhaps a great mobility is observed in NLV because the epitope is switching conformations across multiple states in the absence of force and allowing TCR scanning. Then, if the interacting regions are highly mobile within the binding interface in the absence of force; the probability of complementary matching the opposing interfaces can be favorably enhanced. And once the match is found, force enhances their stability of what would

be otherwise a weak TCR-pMHC interaction. Force modulates and induces a bias in the energy landscape; for instance, at zero force (Free MD), peptide moves across multiple conformations; at critical force, the peptide is mostly restricted to one state (longer bond-lifetime).

(A) (B)
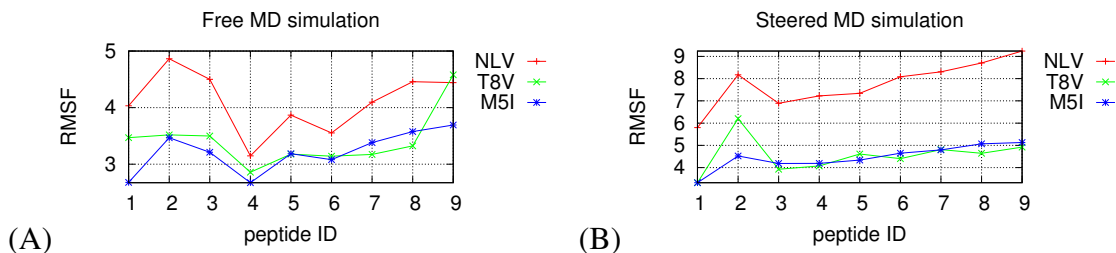
Figure 38: Local mobility of the NLV peptide, T8V, and M5I mutants measured by the RMSF of their C_α atoms in the absence (MD) and the presence (SMD) of force. NLV has comparable displacement to T8V and M5I in the absence of force. However, NLV displaced to a new state and becomes rigid in the presence of force.

4.3.7 The principal component analysis isolates the two conformations

The dihedral angle principal component analysis (dPCA) are used to separate different conformations of a protein occurring through an MD trajectory. The dPCA differentiates conformational changes by using orthogonal eigenvectors (or principal components) that describe the axes of a maximal variance in the distribution of structures in the simulation. The distribution of structures are projected into the largest eigenvectors which provide a reduced dimensional representation of the data set and their corresponding eigenvalues represents the percentage of total variance of atom fluctuations obtained in each dimension. This reduces the dimensional space while still retaining most of the dynamics of the original protein.

Fig. 39 shows the dPCA plots with a k-means clustering for different conformers of NLV, T8V, and M5I peptide ligands, respectively. The distribution of conformations along the principal component plane is represented by dots. Dots with the same color indicates that they belong to the same groups as they have a similar conformer throughout the MD trajectory simulation. Fig 39 (A) shows that NLV complex has two well-defined population

of conformer, while Fig 39 (B) shows that T8V epitope has also two conformers being the second less populated. The transition between distinct structural clusters corresponds to significant dynamical changes in conformation of the peptide sidechain. On the contrary, M5I epitopes (Fig 39) shows a densely pack plot meaning that it has only one conformer. Interestingly only 3 principal components are capable to capture more than 50% of the total dynamics in the simulation. Fig 40 shows the force-free (A) and forced-induced states (B) and it highlights the important interactions between residues.
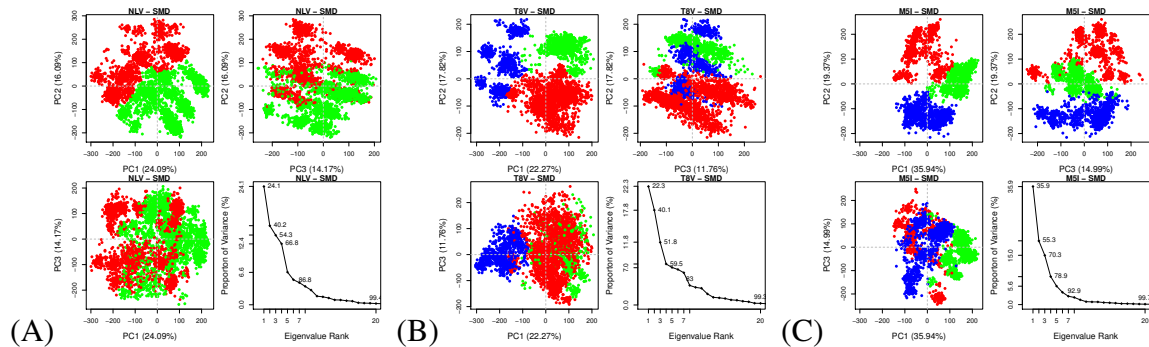
Figure 39: dPCA and k-means clustering for NLV, T8V, and M5I epitopes, respectively. The distribution of conformations along the principal component plane is represented with dots. Dots with the same color indicates they belong to the same groups and they have a similar conformer throughout the trajectory of the simulation.

4.3.8 The molecular basis of *"in silico"* catch bonds in RA14 interacting with the NLV peptide ligand

The RA14– $\alpha\beta$ TCR –is genetically encoded by rearranging TRAV24 and TRAJ49 gene segments for the α-chain and TRBV6-5, TRBD1, and TRBJ1-2 gene segments for the β-chain and these genes mandate the structural arrangement and interaction between the RA14 and the peptide ligand (NLV-HLA2). The crystal structure of RA14-NLV-HLA2 [21] shows special TCR germline-encoded residues interacting with HLA2. Residues Asn29$^\alpha$ and Tyr31$^\alpha$ located at CDR1α are exclusively expressed in TRAV24 and TRAV21 genes and they interact with the NLV-HLA2. TRBV6-5 gene features a hydrophobic sequence core located at CDR2$^\beta$ and they form multiple van der Waals contacts with the ligand. Residues Tyr48$^\beta$ and Asp56$^\beta$ located at the ends of CDR2β are highly conserved residues
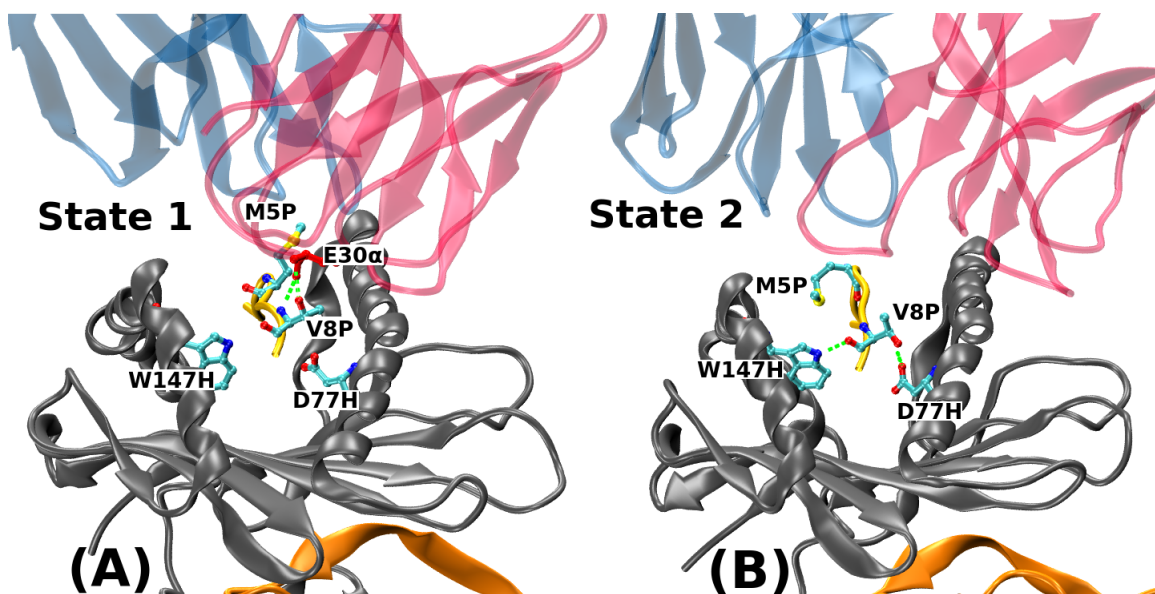
Figure 40: Catch-phase exhibit a second state of RA14-NLV-HLA2 interaction. (A-B) Molecular structure for the two states. The transition from force-free to forced-induced state involves the changing of the docking orientation and the peptide flexibility. In the presence of force, the TCR not only forms new interaction with the peptide-MHC; but also promotes the formation of new H-bonds between the peptide and the MHC. NLV epitope forms two additional anchor points with residues located inside the MHC binding groove, $Asp77^H$ and $Trp147^H$. These residues anchored the NLV epitope in the new conformer. Strikingly these residues are highly conserved across HLAs. This suggests that MHCs have evolutionarily conserved these two residues to lock and modulate the peptide conformations when the TCR is scanning.

across TCR β chains and form H-bonds with the HLA2. $Pro4^P$, $Met5^P$ and $Thr8^P$ from NLV peptide are hot spots since mutating them rendered the epitope ineffective in activating the T cells.

In protein structure is observed that the RA14 and NLV-HLA2 engage in a conventional diagonal docking orientation which allows all CDRs to participate in HLA2 binding while mostly the CDR1s and CDR3s loops participate in peptide recognition [21]. This orientation specifically allows CDR1α and CDR3α loops to be position closer to the N-terminal half of the peptide and the α_1-MHC helix, while CDR1β and CDR3β loops are close to the C-terminal end of the peptide and the α_2-MHC helix. In addition only CDR3α forms

H-bonds with the NLV-HLA2, whereas CDR1β and CDR3β forms H-bonds with the NLV-HLA2. Residue Thr8^P forms one and three H-bonds with Glu30$^{CDR1\beta}$ and Thr97$^{CDR3\beta}$, respectively, while residue Met5^P forms an H-bond with Asn96$^{CDR3\beta}$. Residue Gln72^H forms one H-bonds with Asn96$^{CDR3\alpha}$ and the conserved residues Tyr48$^\beta$ and Asp56$^\beta$. Residue Lys146^H and ALA149^H forms H-bonds with Glu30$^{CDR1\beta}$ and Tyr101$^{CDR3\beta}$, respectively.

Fig 41 and 42 show the stability of those H-bonds reported in the crystal structure [21] for the RA14-HLA2 and RA14-NLV, respectively. During free MD simulations, only 2 H-bonds are long-lived stable interactions, while two others are unstable and reforming intermittently. The others are extremely weak and completely unstable. This is consistent with the lower 3D binding affinity ($K_D = 27.7 \pm 2.3\,\mu M$) reported for the RA14-NLV-HLA2 and not detected for the mutant complexes by using surface plasmon resonance (SPR)[21].
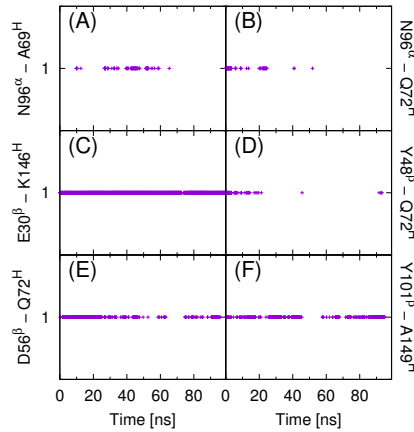
Figure 41: Stability of H-bonds between TCR and MHC from the crystal structure.
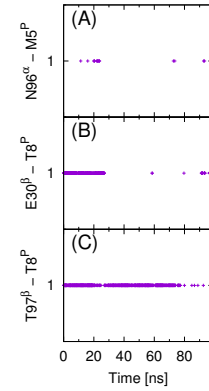
Figure 42: Stability of H-bonds between TCR and peptide from the crystal structure.

4.3.9 The forced-induced H-bonds formed at the catch-bond phase

In protein crystallography, the structures only provide a quantity (# of H-bonds) and they are unable to assess their quality: strong vs weak or stable vs unstable. Thus, simulation is required to provide complementary dynamic information. A similar analogy is made in most biophysical experiments, for instance, micropipette adhesion frequency experiment provides only the quantity of single bonds as 2D binding affinities, where Biomembrane

force probe assess their quality by reporting their bond lifetime. In structural immunology, the underlying assumption is that the docking orientation formed between the TCR and the pMHC is fixed for immune recognition and their footprint does not change after binding. This inherently implies that the binding contacts such as H-bonds are mostly stable and not changing over time. However in free dynamic, these interactions are continuously changing and some are strong while others are weak and not completely stable since they do not form a continuous interaction, but sometimes intermittent. An H-bond is considered intermittent if it breaks and then it subsequently reforms after a period of time. This is different from the stable and strong H-bond that continuously remains attached over a long period of time.

In the previous sections, it is shown how, in the presence of force, the TCR initially interacts and then varies its molecular footprint on the pMHC to form a stable forced-induce state that is functional producing catch bonds. Fig 43 and Fig 44 show the stability of the forced-induced H-bonds created when the RA14 rotates and forms a catch bond with the NLV peptide ligand, for the TCR-MHC and peptide-MHC, respectively. Importantly these interactions are not observed in the crystal structure or free MD simulations. Importantly the H-bonds are generated when the complex changes its docking angle and simultaneously the epitope is switched to the new second state; thus a catch bond is formed. A total of 5 long-lived and stable H-bonds are created during pulling simulations and they are continuous and strong under the umbrella of the catch phase. The others are unstable and reforming intermittently. The H-bonds are visually displayed in Fig 45.

4.3.10 The RA14-NLV-HLA2 exhibits a second state at catch bond phase

In the presence of force, the TCR not only forms new interaction with the peptide ligand; but also promotes the formation of two H-bonds between the peptide and the MHC. Fig 44 shows that the NLV epitope forms two additional H-bond anchor points with residues, $Asp77^H$ and $Trp147^H$, located inside the MHC binding groove. They are visually displayed in Fig 40 (B). Note $Trp147^H$ contact is a weak and unstable interaction present in
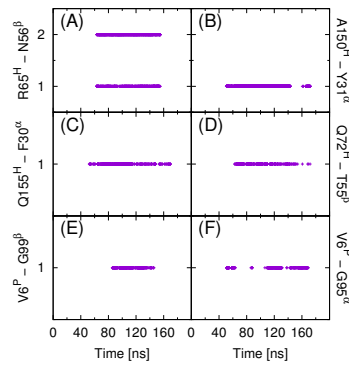
Figure 43: Stability of forced induced H-bonds between TCR and MHC at the second conformation.
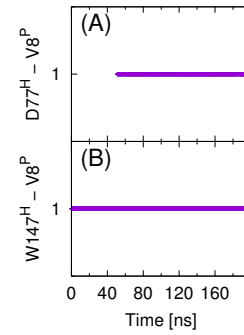
Figure 44: Stability of forced induced H-bonds between MHC and the epitope at the second conformation.

the crystal; however, in SMD it becomes very stable. These residues anchored the NLV epitope in the new conformer by reducing its mobility and decreasing its flexibility right after peptide has switched to the second state ("on"). Strikingly these residues are not only essential for anchoring but they are also highly conserved across HLA2s. There are about nine conserved residues located at each end of the peptide binding groove of MHCs and they are Tyr7, Tyr59, Asp77, Tyr84, Thr143, Lys146, Trp147, Tyr159, and Tyr171 [154]. This suggests that MHCs have remained unchanged and evolutionarily conserved these two residues to lock and modulate the peptide conformations when the TCR is scanning the binding cleft and thus maximizing the peptide read out. The simulation is assigning an important functional role for these conserved anchor points.

4.3.11 The RA14 uses a molecular to lever read out a featureless viral peptide ligand

Initially, residue $Met5^P$ is surrounded by a small molecular cage formed at the interface of TCR α and β chains as depicted in Fig. 40. The interaction of $Met5^P$ with TCR is mostly of a hydrophobic nature since it does not form any stable H-bond in free nor steered MD simulations. This is opposed as it is reported forming an H-bond with $Asn96^{CDR3\beta}$. In Free MD simulations, the $Met5^P$ side chain is position inside the molecular cage; however, in the presence of force, $Met5^P$ is dislodged from the molecular cage and it is forced to

84
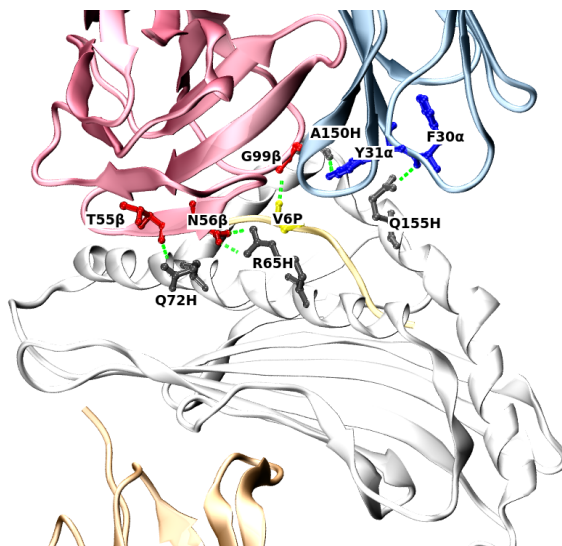
Figure 45: The *"in silico"* catch bond exhibits a second state of RA14-NLV-HLA2 interaction. The two conformations are shown in full molecular detail, the new forced-induced H-bonds are omitted in the second state for easy visualization. The first state is similar to the one observed in the crystal structure. The forced-induced H-bonds between the TCR and MHC participate indirectly in the mechanism of RA14 using a molecular lever. They provide essential support to maintain together the complex when the TCR rotates and changes the docking orientation when the pulling force is growing. They hold together the RA14 and the pMHC otherwise they will dissociate.

relocate its hydrophobic side chain inside the MHC binding groove. This is because the TCR rotating around the binding axis kicks out Met5^P side chain outside the molecular cage and creates a lever. The molecular lever is formed by the TCR pushing down (from one end) the tip of the Met5^P side chain (C$_\epsilon$) inside the binding groove as it is rotating. The "out-of-cage" movement of Met5^P not only pushed down and buried its side chain but also: (1) lifts the C$_\alpha$ of Met5^P (the load) from the other end; (2) levers up the peptide hydrophobic core (Pro4^P, Met5^P, Val6^P); (3) and exposes this core to the molecular cage. This movement also dynamically locked the epitope in this forced-induced conformation by creating simultaneously another anchor point, Thr8^P. Inserting the side chain of Met5^P also brings together Thr8^P to MHC α_1 domains where the conserved residue, ASP77^H, is located. This promotes the formation of an H-bond between Thr8^P and ASP77^H. This defined the transition to the forced-induced conformation since the peptide is no longer

allowed to move and it is locked at this state by losing its mobility and flexibility.

Thr8^P (the fulcrum) has two important functions: one is to interact with the MHC to form an anchor point for the lever and stand the pressure extorted by RA14 when it is rotating and lifting the hydrophobic core out of the binding groove; second reduce the flexibility of the epitope by locking it in this second conformation. This is also opposed as it is reported that Thr8^P forms multiple H-bonds with RA14 and not with MHC. Note that pushing down Met5^P side chain facilitates the insertion of its side chain into the binding groove since Met is hydrophobic in nature.

The sequential events of the molecular lever are quantified in the Fig. 47. For this propose, three distances are defined as: d_1 is the distance from the tip of the side chain of Met5^P (C_ϵ) to the plane formed by the MHC binding groove; d_2 represents the distance from the C_α of Met5^P to the C_α of a hydrophobic residue (Phe98$^\alpha$) located inside the molecular cage; d_3 is the distance between the anchored point (Thre8^P) and the conserved residue Asp77^H. The distances are visually depicted as purple arrows and the interacting residues are highlighted in Fig. 46. The curves in Fig. 47 are composed of free (0 to 100 ns) and steered MD simulation (100 to 280 ns). This is done to study the behavior of the molecular lever in the presence and absence of force. The black arrow depicts the moment forces is applied to the complex and the blue vertical lines enclose the catch-bond region. Initially, there is a delay of 60 ns to start forming the catch bond as shown Fig. 15; this is because constant-velocity pulling is used and thus the force is proportional to the time as $f = K(vt - x_o)$. Force starts from zero and requires about 60 ns to reach about $\approx$ 50 pN.

In the absence of force, the Met5^P side chain is always inside the molecular cage and its tip (C_ϵ) is far away from the plane of the binding MHC; however, as force is applied RA14 rotates and places Met5^P side chain outside the molecular cage. Met5^P is buried inside the binding groove and distance, d_1, decreases about 6Å in the catch region, Fig. 47 (A). The tip is then closer to the groove plane. As the catch phase is finished the Met5^P side chain returns to its original position since RA14 is not pushing it down and RA14 is dissociating

from the pMHC.

Simultaneously distance, d_2 (Fig. 47 B), decreases once the tip is dislodged and buried by the TCR. RA14 uses the molecular lever to lift the peptide hydrophobic core (P4 and P6) out of the binding groove; thus the core is rapidly sucked into the molecular cage to avoid contact with water molecules. This also increases the probability of the epitope to be exposed and maximizes the read out by RA14. This is shown as the distance from the Met5^P C_α to inside the molecular cage decreases only in the presence of force.

Finally, the TCR pushing the lever positions together Thre8^P and Asp77^H and creates an anchor point (H-bond) so that the pMHC is stable and does not release the peptide. This interaction provides stability for the optimal engagement between RA14 and pMHC as the peptide is being pulled by TCR. This is shown as distance, d_3, (Fig. 47 C), sequentially decreases and reaches an average of 2.6Å in the catch phase; such distance classifies the H-bond as "strong". The H-bond is formed between side chain atoms, Thr8^P-OG1 and Asp77^H-OD2. Fig. 47 (D) shows that the H-bond is strong and stable only in the catch region.

The forced-induced H-bonds formed between RA14 and NLV peptide ligand plotted in Fig. 43 and Fig. 44 are visually shown in Fig. 45. Those H-bonds participate indirectly in forming the mechanism of the molecular lever since they provide essential support to maintain together the complex when the TCR is being pulled. They are stickum points and without them, the TCR will dissociate when rotating.

All simulations performed using constant-velocity pulling at normal and medium ramping rate show the transition to the forced-induced state for the NLV complex. This state is longer-lived inside the catch bond. Higher ramping rates render the complex unable to transition to the second state; thus its dissociation from the pMHC with a footprint similar to the crystal structure. For simulations performed using constant-force pulling, the complex shifts to the new state only for forces inside the catch bond region. Outside the catch bond phase, they present only one state that is similar to the crystal structure.
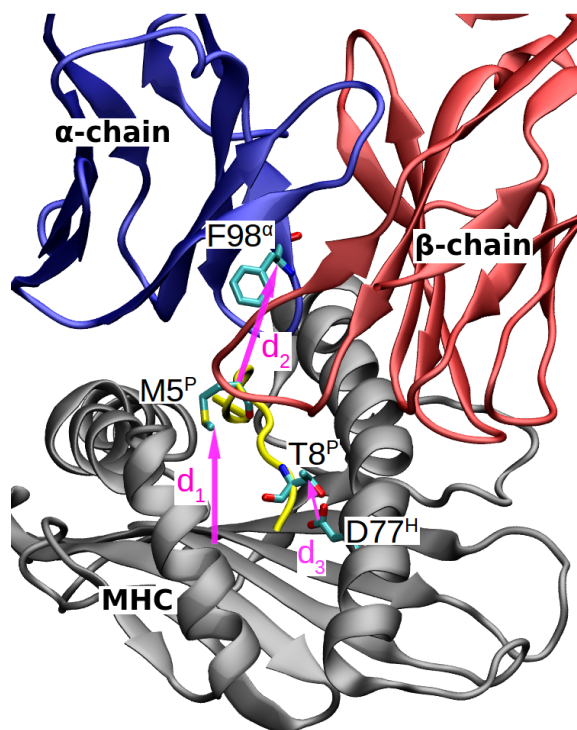
Figure 46: Illustration of the TCR using a molecular lever to form an *"in silico"* catch bond. d_1 represents the distance from the tip of Met5^P side chain to the plane formed by the MHC binding groove; d_2 is the distance from the backbone of Met5^P to the C_α of hydrophobic residue located inside the molecular cage; d_3 is the distance between the anchored point (Thre8^P) and Asp77^H. The distances are depicted as purple arrows and the interacting residues are displayed.

4.3.12 Understanding the evolution of viral escape mutants

Mutating Met5^P to Ile5^P is unlikely to produce any significant effect since their side chains have similar chemical properties; however, this is not entirely true. Functional studies report that NLV peptide ligand is detected, while mutant M5I experiences a different faith. M5I^P peptide ligand is an escaped mutation that is not easily recognized by NLV-specific TCRs and it also has a decreased binding affinity. T8V peptide ligand is not detected in SPR experiments and its 2D binding affinity is higher than M5I but still lower than NLV epitope.

If Thr8^P is substituted as in the mutant T8V, the epitope will be either easily removed from the MHC binding groove from the unanchored end when pulling force grows or will
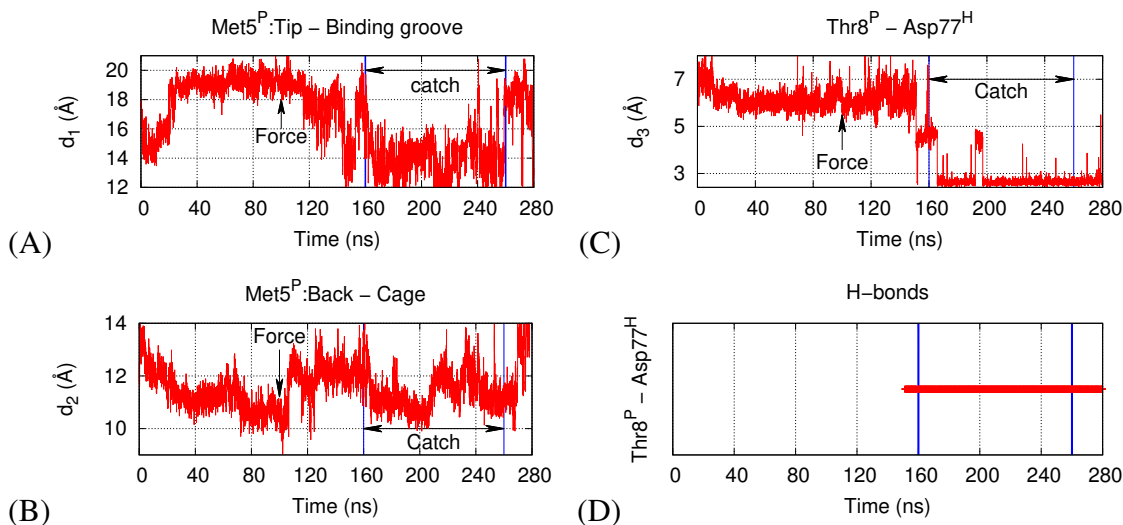
Figure 47: Quantification of the sequential events of the molecular lever. d_1 is the distance from the tip of the side chain of Met5^P to the plane formed by the MHC binding groove; d_2 is the distance from the backbone of Met5^P to the C_α of hydrophobic residue located in the molecular cage; d_3 is the distance between the anchored point (Thre8^P) and Asp77^H. Both Free and steered MD simulations are shown together to study their time-changing behaviors in the absence and the presence of force. The black arrow indicates the moment force is applied starting from 0 pN to the complex. There is a delay of 60 ns before the force reaches about 50 pN and becomes significant to propagate through the complex and induce changes in the docking angle and RMSD of the epitope. The catch phase is enclosed and label. Importantly the mutants, T8V complex fails to secure the second state, while M5I complex has the same state in the presence of force.

not lock the epitope in the second conformation for a longer period of time. This is because Valine is unable to form H-bonds with its side chain. For smaller forces, the pMHC would be able to hold the epitope with the another conserved residue Trp147^H and an H-bond formed with the backbone of Thr8. This interaction is present in both NLV and T8V peptide ligands in force-free conditions and it becomes strong in the presence of force. At higher forces beyond the critical force, both contacts are required, Asp77^H and Trp147^H, to resist the increasing load.

If M5^P is replaced as in the mutant M5I, Thr8^P could still form an H-bond with MHC; however, this replacement disrupts the out-of-pocket movement since the ILE is always located inside the molecular cages whether or not the TCR rotates. Thus the rotating lever is lost and not anchor point is formed. As the pulling force grows, it will be easy to remove

the epitope from the binding groove. This would render eventually the pMHC unstable and unfolding is possible. In our simulations, the MHC unfolds mostly for the mutant peptide ligands and this suggests that it may not be an artifact from the modeling but actually a functional escape pathway for the virus. M5I is a naturally occurring infrequent mutation in human CMV seropositive patients [145]. Apparently, this mutation does not disrupt the MHC class I presentation but it impairs its stability in the presence of force. This is an intelligent mutation strategy from the virus since in the absence of force, the complex, pMHC, is probably long-lived and stable; however, it becomes very unstable and unfolds only when it is scanned by the TCR as it applies force. Thus M5I peptide ligand avoids down regulating the MHC class I presentation and being targeted by the Natural Killer Cell.

The structural and molecular basis is supported by BFP experiments where NLV epitopes form a strong catch bond, while T8V forms a very weak catch bond. M5I does not form catch bonds and its force profile is similar to an ideal bond. This is consistent with the mutation impacting mostly peptide-MHC Binding stabilization but not pMHC presentation [152, 153]. As soon as the complex is pulled, it dissociates the peptide from the groove and unfolds the MHC; thus the lifetime reported in the BFP experiment is mostly the time required to unload the epitope from the MHC binding groove; this is independent of force since the mutant severally impacts the stability of the complex.

2D diagram plots are presented and they provide detail information about the interaction among TCR, peptide, MHC in the absence and presence of force. Fig 48 and Fig 49 show the interactions mediated by hydrogen bonds and by hydrophobic contacts for the force-free and forced-induced states, respectively. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.
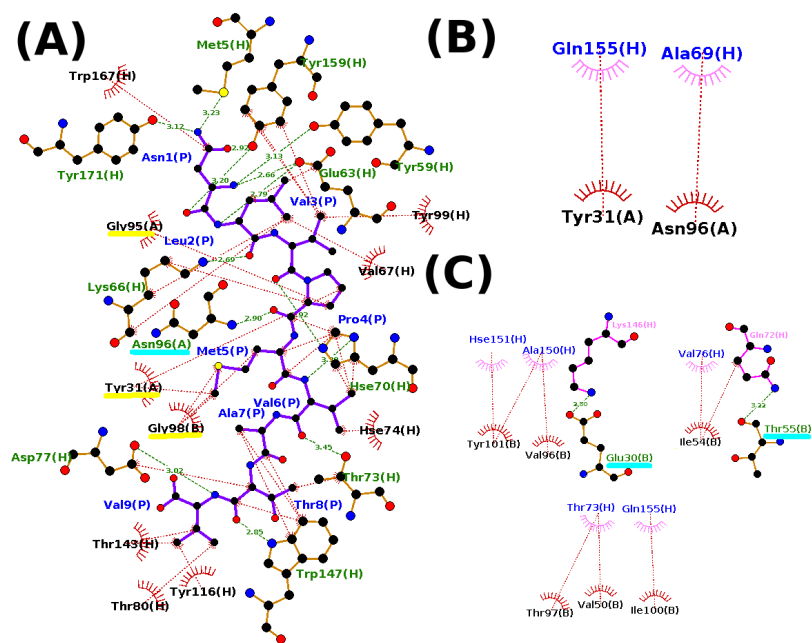
Figure 48: 2D Diagram plot of the first state: The interaction between the RA14 and NLV-HLA2 in force-free conditions. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.
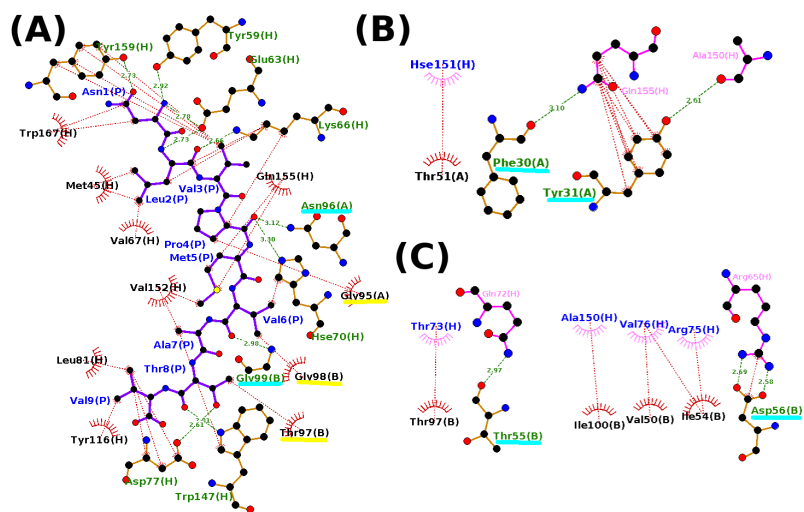
Figure 49: 2D Diagram plot of the second state: The interaction between the RA14 and NLV-HLA2 at the catch-bond phase. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.
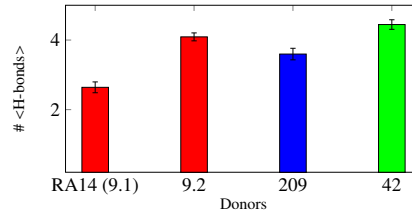
Figure 50: Free MD simulation: The average number of H-bonds at the binding interface between the TCR and NLV-HLA2 for different donors.

	TRAV	TRAJ	CDR3a	TRBV	TRBJ	TRBD	CDR3b
Donor		24*01	49*01	CARNTGNQFYF	6-5*01	1-2*01	1*01	CASSPQTGAIYGYTF
9	3*01	26*01	CAVYYGQNFVF	28*01	1-1*01	1*01	CASSFQGYTEAFF
				12-4*01	2-1*01	2*01	CASSSVNEQFF
42	24*01	31*01	CASLLYARLMF	6-5*01	1-5*01	1*01	CASTTGTGGQPQHF
	3*01	31*01	CAVRDTNARLMF	12-4*02	2-1*01	2*01	CASSVVNEQFF
209	24*01	49*01	CARNTGNQFYF	6-5*01	1-2*01	1*01	CASSPVSGTGHYGYTF
	26-2*01	40*01	CILLGAYKYIF	7-8*01	1-1*01	1*01	CASSPLYGTDKGTEAFF
				6-5*01	1-4*01	1*01	CASSYSNVGAVTEKLFF

Figure 51: TCR genes shown in the table according to IMGT nomenclature. The data shown are from three T cell lines generated from three individuals.
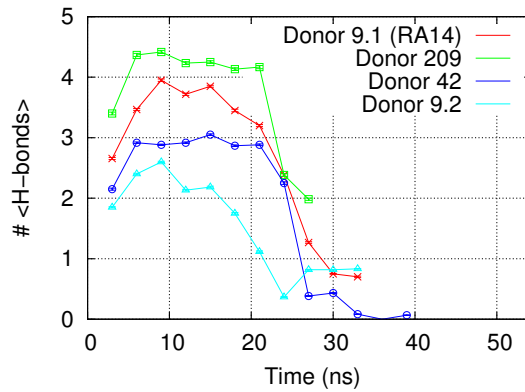
Figure 52: "*In silico*" catch bond for different donors. Time evolution of the number of H-bonds between TCR and NLV-pMHC in the presence of force. Catch-bond is represented by an increased in the number of the H-bonds. It is used constant-velocity pulling. Donor 42 displays a small catch bond while donor 209 has basically a similar profile with RA14 (donor 9.1). Donor 9.2 displays also a small catch bond.

4.3.13 Structural differences of NLV-specific TCRs from different donors forming "*in silico*" catch bonds when ligated to NLV-MHC

Finally, the simulations are initially compared with experiments conducted in Donor 9. Two new donors are included to address the possibility of donor-to-donor variations. The two donors are 42 and 209 and they are collected and provided by the Grakoui lab at Emory
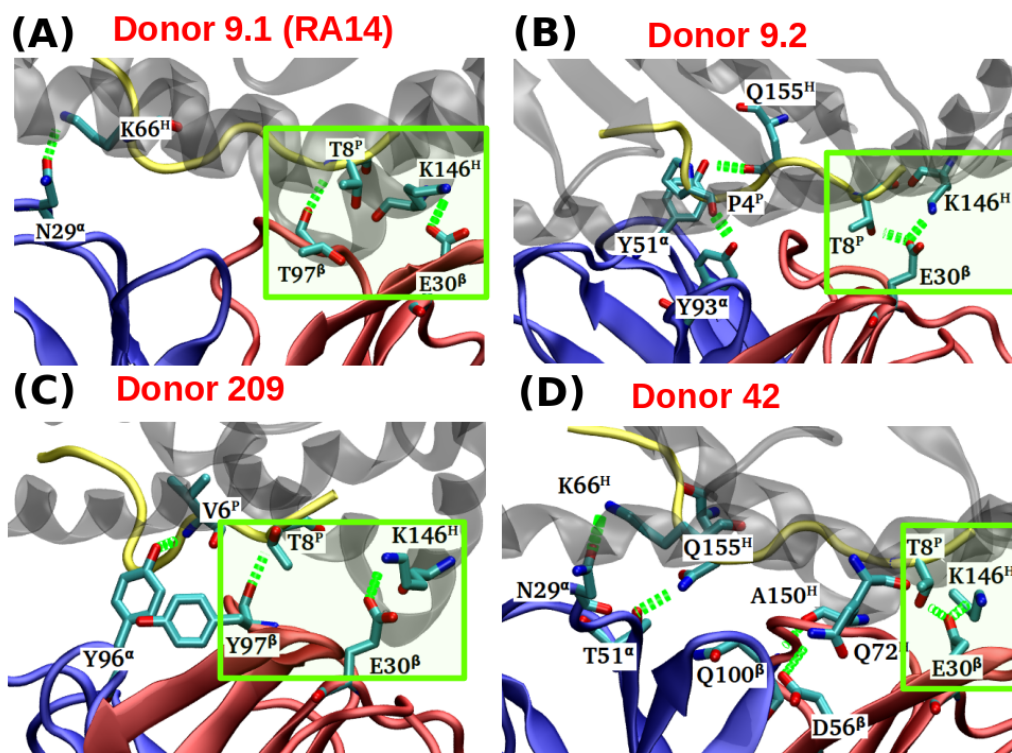
Figure 53: Structural differences of TCRs from different donors at resting phase. Interaction of immunodominant CMV-TCR to NLV peptide in the absence of force for different donors.

University, School of Medicine. The three donors are CMV seropositive and their NLV-specific TCRs have indistinguishable 2D affinities as they are in the same log order of magnitude. They are sequenced in the lab of Dr. David Prize at Cardiff University School of Medicine and it is shown that Donor 9 has two T-cell populations, one expressing RA14 (called donor 9.1) and a variation (called donor 9.2). Donor 42 and 209 mostly present one immunodominant T-cell population. Interesting the new donors possess TCRs with similar CD3α and CD3β; however they present variation in key residues. The sequence and populations are shown in Fig 51. Since their crystal structures are not currently available, they are generated performing *"in silico"* mutations on the RA14-NLV-HLA2. The new *"in silico"* complexes and their responses are compared with their respective experimental profiles. The structures of the new donors are minimized and equilibrated for about 200 ns and it is observed that their RMSD reaches a stable value (data not shown). For the SMD
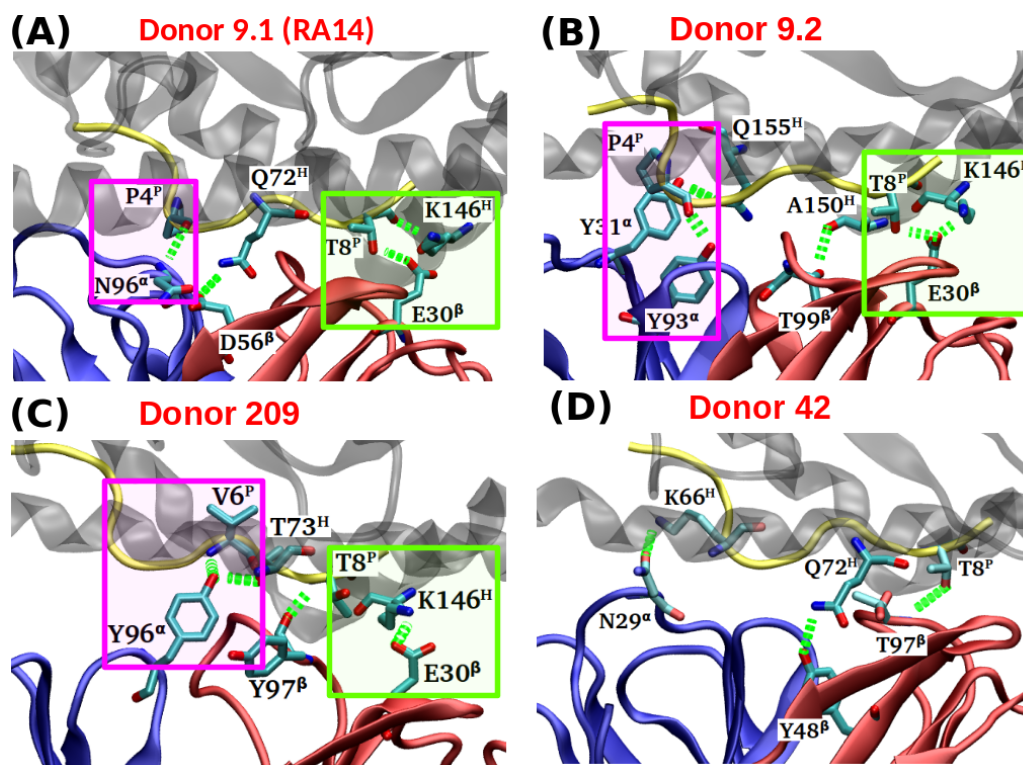
Figure 54: Structural differences of TCRs from different donors at catch phase. Molecular interaction of immunodominant CMV-TCR to NLV peptide in the presence of force for different donors.

simulations, all donor were pulled at high force for 5 to 10 ns to further drive them into low energy states and reequilibrated for 50 ns. Those configurations are pulled at high constant velocity and Their responses are consistent with the experiments: Donor 42 displays a small catch bond while donor 209 has basically a similar profile with RA14 (Donor 9.1). Donor 9.2 displays also a small catch bond. The results are shown in Fig 52 and their molecular details at the catch phase are shown in Fig 54.

CHAPTER V

THE PRINCIPLES FOR MOLECULAR RECOGNITION OF ANTIGENS AND ACTIVATION OF CD3 SIGNALING DOMAINS BY THE T-CELL RECEPTOR

5.1 Abstract

The overall goal of this chapter is understanding the principles for recognition of antigens and then the activation of CD3 signaling domains by the TCR. The first part is focused to investigate how the information encoded in the peptide is decoded mechanically by the TCR and what is the structural role of catch bonds. It is established the general principles for TCR molecular recognition of antigens by studying the response of five biologically different TCRs in the presence of force. It is demonstrated that 1) Catch bonds are required to effectively recognize epitopes; 2)TCR musts rotates around the binding axes to form *"in silico"* catch bonds and induce functional conformational states of the peptide. The second part of this chapter is to study how TCRs uses mechanical forces to determine whether or not the presented antigen is a threat. The focus is to identify how information read from the TCR distal-membrane binding site is propagated to the CD3 signaling domains. It is proposed the peptide-decoding process and intracellular signaling are connected by conformational changes traveling across the TCR and it is described the unique features of this mechanism. The model shows that the TCRs are deformable proteins that can experience large conformational changes and they use mechanical forces to modulate their conformations. The ability of TCRs to deform without releasing the peptide ligands is the key to understand this complex mechanism. Finally, it is unveiled that TCRs uses their C_β FG loop as a molecular mechanism to propagate information and connect the CD3 signaling domains. This is accomplished by changing the conformational states of C_β FG loop.

5.2 Background

There are continually strong demands in understanding the principles for molecular recognition of antigens and activation of CD3 signaling domains by the T-Cell receptor because of their broad implications in immunology and the benefits in cancer immunotherapy. Owing to the progress of nanobiotechnology, scientists have investigated the functional and mechanical properties of the TCR at the level of single molecules and a plethora of information was collected; however, the studies appear to be disconnected or even contradicting one from another.

The first part is focused on establishing the principles for molecular recognition of antigens by the T-Cell Receptor. Previously it is shown that RA14 must rotate around the binding site in order to produce an *"in silico"* catch bond and this is associated with the TCR functionally switching to a "forced-induced state". Interesting, RA14 ligated with NLV-HLA2 also forms experimentally a catch bond. This poses the following question for antigen recognition of TCRs interacting with peptide ligands: 1) Are rotating and changing conformations (for the peptide and the complex) intrinsic properties of TCRs? Or is it a particular property for RA14?; 2) Are catch bonds required to effectively recognize epitopes? To address this question, the response of five different TCRs are investigated in the presence of force using similar pulling protocols as the previous chapter. Importantly, the chosen TCRs are completely different biologically from one another and their functionality has been recorded experimentally. This part systematically investigates and tests the rules for molecular recognition.

Antigen processing and presentation by MHC molecules was extensively studied; however, how TCR recognizes peptides and then activates the signaling machinery is not understood. This question has puzzled immunologist for decades. Understanding if there exists a preconceived traveling pathway, that is genetically encoded in the TCR genes, to propagate information read from the TCR distal-membrane binding site to the CD3 signaling domains is considered the holy grail of molecular immunology. Multiple competing models

have been proposed with no clear answer; many of them contradict each other, despite all models being supported by experimental evidence. Is it possible for a molecule to have apparently multiple mechanisms to accomplish the same task? The second part describes the molecular mechanism of how this process is accomplished. It is proposed that the decoding process and intracellular signaling are connected by conformational changes in the TCR, and a unifying model is presented and its unique features are described to solve this puzzle. The model is consistent with all experimental observations up to date. The TCR interactions with either or both of the pMHC and CD3 are hardwired on their structural dynamics and not in their static structure. The TCRs are not rigid entities, as currently thought, but they are deformable proteins and mechanical forces modulate their local and global conformations. Yes, TCRs could potentially experience large conformational changes that are functional. The ability of TCRs to deform without releasing the pMHCs and to transform a linear force into functional conformational changes are the key to understand this complex mechanism. Finally, it is unveiled the molecular mechanism used by the TCR to open and close its C_β FG loop. This functional mechanism highlights the important role of the C_β FG loop in activating the CD3 signaling machinery.

It is suggested that TCRs mediate sequentially recognition and activation by connecting their dimeric $\alpha\beta$ chains with CD3 $\epsilon\gamma$, CD3 $\epsilon\delta$, and CD3 $\zeta\zeta$ subunits using an "unknown mechanism". The C_β FG loop in TCRs is regarded as very important in this process since its removal severely affects T cell activation and development. For instance, C_β FG-loop deleted T cells manifest reduced proliferation and cytokine production abilities upon T cell stimulation and dramatically impaired negative selection. Therefore, it has been long asked how recognition of peptide ligands by TCRs evoke activity on T cells since TCR antigen-specific binding site is located at least 40 Å away from the C_β FG loop and the overhead of CD3 subunits. There must exist a structural mechanism that propagates information, presented at the TCR binding site, to the CD3 complex. This traveling pathway, if exists, is still a mystery.

In general, scientists have attempted to understand the TCR interactions by characterizing their thermodynamic properties, affinity measurements, protein crystallization, and single-molecule force spectroscopy, with no much success. The problem with not capturing the deformable states of TCRs on those experiments, especially in crystallography, is that determining their 3D structure is a delicate and complex task where multiple techniques are required to produce proteins with increased crystallization propensity such as molecular cloning, recombinant protein expression, and mutagenesis. In addition, the structure of proteins dynamically switches between different functional states and it is altered/modulated by their ligand under very specific conditions. Those conditions are not easily achieved in protein crystallization and prevent scientists to capture them in their physiological condition. Thus they obtain structures that do not usually represent the functional state of the protein but rather an ensemble of many states. On the other hand, MD simulation is a powerful technique to explore and observe conformational changes in proteins. MD simulations can access rare states by employing sampling methods to cross the barrier and increase the probability for a protein to transition to a different state.

It is also believed that the specific binding orientation and interaction of the TCR complementary-determining regions (CDRs) with the peptide loaded onto MHC molecule are the basis of immune recognition. It is still debatable whether the CDR amino acid contacts have co-evolved and thus are germline encoded to govern the TCR-MHC restriction or the MHC bias is externally imposed on the TCR during thymic selection by co-receptors. Structural analysis of TCR-pMHC crystals supports a model of physical specificity between the germline CDR regions and MHC. Other suggest that co-receptors mediating the thymic selection are responsible for the MHC bias since TCRs can robustly engage non-MHC ligands when given the chance; thus TCR repertoire is not innately specific for MHC but instead is subject to externally imposed influences. This proposed a different role for co-receptors in which they directly influence TCR recognition of pMHC. Nevertheless, what governs TCR specificity and sensitivity are not understood and they may be also related

98

to the ability of TCRs to form catch bonds. For instance, it is difficult to assess if a given epitope is able to activate the T cells by reading the affinities. This is the case especially when its binding affinities are not at the end of the spectrum for T cells, very high (1μM) or very low (100μM). A better predictor to measure the strength of TCR-pMHC interaction is the bond lifetime measured under forces. If a TCR forms a catch bond, it will activate the T cell. Conversely, if a T cell forms a slip bond it may produce a reduced activity. In this case, catch bonds can be useful for assessing functionality. However, the mechanisms of how TCRs employ catch bonds for molecular recognition of antigens and activation of CD3 signaling domains is not established. Thus this chapter dissects the role of catch bonds in recognition of antigen and activation of signaling machinery by the TCR.

In molecular immunology, there are still many unanswered questions concerning the structural dynamics of TCRs such: Does a naive TCR have a similar structure to a memory TCR? How does the structural landscape of the pMHC binding site define the TCR repertoire and subsequently its effector function? How a public TCR respond differentiate from a private TCR? In addition, it is still under debate what quantity best defines a functional binding between TCR and pMHC and what role it has in the T-cell response. The TCR-pMHC are also likely to form parallel bonds instead of serial at the binding interface; however, the effect of bonds in parallel on the T-cell responses are still not fully studied; How close do bonds have to be before they can be lumped together or treated separately?

There is no doubt that structure defines function and influences the TCR biology as it was found that the structural accessibility of a pMHC binding surface dramatically influences not only the antigen-specific T-cell response but also biases the expressing profiles of TCR that are selected from a diverse, naive repertoire during infection, transplantation, autoimmunity, or hypersensitivity. These support the view that TCR functional response is mechanically-regulated for immune recognition and activation. Here simulations are used to molecularly manipulate the TCR conformational changes with position and force resolutions on the order of Å and pN, respectively and understand its functional conformational

Table 1: Summary of TCRs used in this chapter with their respective recorded experimental data. The TCRs are completely different biologically. Simulation time refers to all combined simulation performed per TCR in this study. Most simulations are carried out using RA14 and 1E6. C7, C25, and 2C are used for comparison. M5I* forms an ideal bond experimentally. Ligands** can be considered very weak agonist or antagonist. † represents unpublished experiments conducted in the Zhu lab. Not determined or not detected are abbreviated as n.d.

TCR	RA14	1E6	C7	C25	2C
Species	Homo Sapiens	Homo Sapiens	Homo Sapiens	Homo Sapiens	Transgenic mice
Type	Viral	Autoimmune	Viral	Viral	Alloreactive
Specific for/ Causing	CMV	Type 1 Diabetes	CMV	CMV	PIFS
Agonist epitope	NLV	YQF, RQW, RQFA, RQFI	NLV	NLV	SIYR
Crystal structure	Yes	Yes	Yes	Yes	Yes
Affinity	Medium	High	High	High	High
Bond type	Catch†	Catch†	n.d.	n.d.	Catch
Weak agonist epitope	T8V, M5I*	ALW, AQW MVW**, YLG**	n.d.	n.d.	Many
Crystal structure	No	Yes	n.d.	n.d.	Yes
Affinity	n.d.	Very low	n.d.	n.d.	Medium
Bond type	Catch and slip†	Slip†	n.d.	n.d.	Slip
Reference	[21]	[97, 98]	[85]	[85]	[155, 156]
Simulation Time (μs)	12.0	20.0	1.0	1.0	1.0

changes. This methodology is novel because it is a bottom-up approach that builds from a molecular perspective and it provides insight into the dynamics of TCR and how its conformations modify its specificity and activity. These ideas have not been entirely explored in experiments due to an inability to characterize high molecular vibrations.

5.3 Results and Discussion

In this section, the general rules for molecular recognition of antigens and activation of CD3 signaling domains by the TCR are outlined. The observations obtained in the previous chapter are verified if they are generally applicable to most TCRs forming "catch bonds". In this chapter, it is used MD, steered MD, targeted MD simulations (proposed as an artificial-ligand protocol), *"in silico"* mutations, and computationally deleting TCR

loops to systematically understand the dynamics and conformational change in TCRs.

For a few decades of intensive study of the TCR biology, the factors that enable productive signaling upon a TCR binding to a pMHC still remain elusive and to fully understand this critical process requires the systematic study of TCR conformational changes as a mechanosensor and how they interact (bind) with CD3 signaling domains. To solve this puzzle, it is important to understand that under physiological conditions, protein structures are not **static** but they do change dynamically and access an ensemble of conformations that have significantly narrower distribution compared to random colloids. This is different from the traditional view that proteins have a native state or unique structure genetically encoded by their amino acid sequence; but rather the native state– observed in the crystal– is an ensemble of microstates. The microstates share the same overall folding and 3D structure but differ in their molecular details such as different lengths in bond and angles that could significantly impact their local loop conformations, domain packing, or even entire domain orientations. The ability of proteins to deform their native state is important to properly and biologically function. The "jigglings and wigglings" of atoms in a protein enable the access of structural conformers that potentially could have different functional states. The "jigglings and wigglings" are not only related locally but they are long-range vibrations that involve interdependence and correlated motion of atoms throughout the entire protein. Disrupting the long-range vibration and breaking the interdependence may render the protein functionally obsolete or provide a pathway for "remote" control to either enhance or inhibit the physiological function.

It is shown that the RA14– an NLV-specific TCR– must rotate around its binding axes in order to form new H-bonds and produce an *"in silico"* catch bond; the rotation is also associated with the epitope changing to a "forced-induced state". These are observed either the complex was pulled at "constant velocity" or "constant-force". This suggests that, for molecular recognition and activation, rotating around the binding axis and generate new states for the entire complex is an intrinsic property for all TCR ligated to cognate

pMHC. To address this question, the response of RA14 [21] and 4 new different TCRs: C25 [85], C7 [85], 2C [155], and 1E6 [97, 98], are individually investigated in the presence of force using similar pulling protocols. Similarly to RA14, C7 and C25 are human TCRs specific for HCMV viral infections and they recognize the NLV epitope. RA14 and C25 are public TCRs used to study clonal diversity. NLV-specific TCRs are notorious for being very diverse and immunodominant after infection, shaping the TCR repertoire and producing public T cell responses. 2C is an alloreactive TCR in transgenic mice capable of binding self and foreign peptide. It recognizes the superagonist peptide, SIYR, loaded into H-2K^b and it is responsible for causing a peptide-induced fatal syndrome (PIFS) mediated by CD8$^+$ T cells in naïve 2C transgenic mice. 1E6 is an autoimmune TCR specific for human type I diabetes that recognizes the preproinsulin peptide (ALW) and it is responsible for for the killing of human beta cells. 1E6 TCR is also able to recognize more than a million different peptides and here it is included to study multi-ligand discrimination by a single TCR. Importantly these TCRs are completely different from each other biologically and their functionality has been documented experimentally and they are summarized in table 1.

5.3.1 C7, C25, 2C, and 1E6 TCRs are capable of forming "*in silico*" catch bonds

In this section, the biophysical property of C7, C25, 2C, and 1E6 TCRs are studied. These TCRs have been verified experimentally to either form catch bonds and inducing early T cell activation (RA14, 1E6, 2C), or to have higher 3D affinities (C7, C25) compared to most TCRs and are closed to the higher end of the affinity spectrum. 1E6 TCR also is capable of forming experimentally slip and catch bonds with weak and strong ligands, respectively. See table 1 for a complete summary. Fig. 55 illustrates the time evolution of the total number of H-bonds between the TCR interacting with a pMHC for C25, C7, and 2C. The TCR-pMHC complexes are undergoing a "constant-velocity pulling". In human C25 and C7 TCRs and mouse 2C TCR, the total number of H-bonds increases at the TCR-pMHC binding interface as an applied force at the proteins' C-terminal, is simultaneously

growing; and thus forming an *in silico* catch bond. Then it transitions to slip bond since the applied force becomes big enough to overpower the H-bonds at the binding interface. The figure. 55 also shows the average number of H-bonds in free MD simulations for the C25, C7, 2C and 1E6 TCRs; This information is used later to compared with SMD simulations.

A longstanding question in TCR signaling is how structurally similar ligands that have similar (2D or 3D) affinities, can have different biological functionality and activity. The crystal structures of 1E6 TCR bound to a panel of ligands ranging from antagonist to strong agonist are used to study how the rules for molecular recognition and activation are affected by the ligand potency. 1E6 TCR is interacting with HLA-A*0201 loaded with a preproinsulin peptide (ALW) and 7 altered peptides–MVW, YLG, AQW, RQF(A), RQF(I), RQW, and YQF. The epitopes differ only by a single amino-acid substitution. The experimental benchmark is that recognition of ALW peptide by 1E6 autoreactive TCR induces the killing of human islet beta cells in type 1 diabetes. MVW and YLG have a lower 2D affinity compared to ALW and require substantially higher peptide concentration than physiological conditions to induce T cell killing. RQF(A), RQF(I), RQW, and YQF epitopes have a higher 2D affinity and induce T cell activation and killing from low to normal peptide concentration.

Fig. 56 shows the total number of H-bonds vs time at the TCR-pMHC binding interface for the peptide ligands. Note that the applied force is proportional to the time, $f = k(vt - x_o)$. ALW, MVW, and YLG epitopes display an *in silico* slip bond where the average number of H-bond decreases as the force increases with the peaks at their H-bond profile decreasing in that order, respectively. RQF(A), RQF(I), RQW, and YQF have a catch bond and their prominence (height and duration) is proportional to the experimental peptide potency. AQW is roughly showing a very small catch bond or ideal bond– where the number of H-bonds is independent of force. AQW is considered here the phase-transition epitope between switching the complex from slip to catch bonds or vice versa. The phase transition also separates functionality, epitopes forming strong catch bonds have higher

affinities and are effective inducing the killing of cells, while the ones having slip bonds may have a lower affinity and promote little activity in T cells.
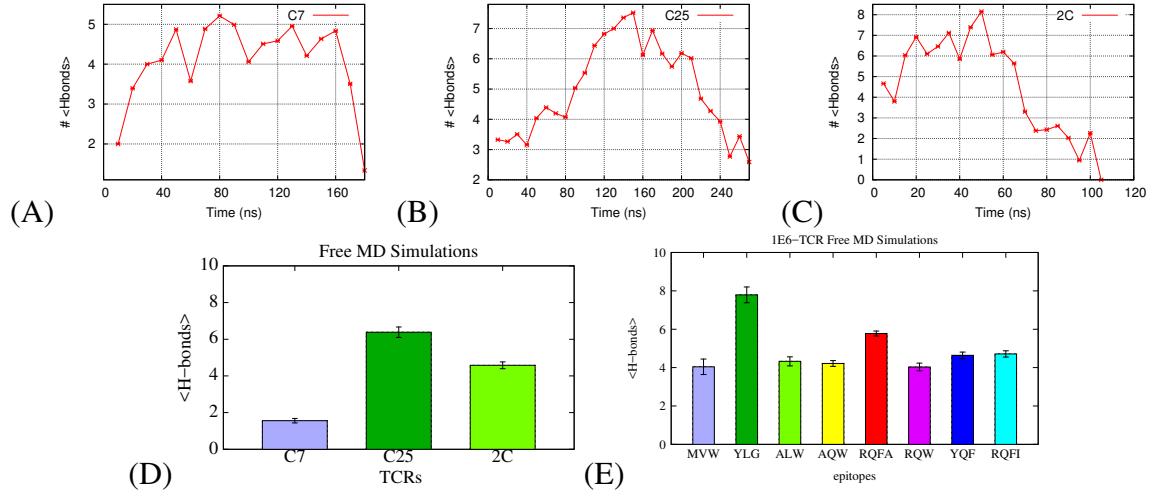
Figure 55: "*In silico*" catch bonds of C7, C25, 2C interacting with cognate peptide ligands. The number of H-bonds between TCR and pMHC is plotted over time for C7 (A), C25 (B), 2C (C). The TCRs are interacting with their respective agonist epitopes. The "*in silico*" catch bond is represented by a total increase in the number of the H-bonds at the binding interface. (D) and (E) represent the total number of H-bonds at the binding interface in the absence of force (free MD simulation), for the C7, C25, and 2C (D) and 1E6 interacting with a panel of peptide ligands ranging from antagonist (MVW, YLG), weak agonist (ALW, AQW) and strong agonist (RQW, RQFI, RQFA, YQF).

In summary, before entering the catch-bond phase, at zero and small forces, the TCR-pMHC complex stochastically fluctuates around the initial native structure. In the catch-bond phase, the TCR-pMHC moves to other intermediate states (with higher energy) and it starts diffusing slowly along the axis not parallel to the pulling direction. This result in high mechanical stability that allows the complex to cross any high energy barrier. Increasing the ramping rates makes the stochastic fluctuations of the TCR-pMHC along the pulling axis to be irrelevant and the dissociation pathway becomes deterministic. This is because pulling very fast does not allow the complex to explore its surrounding energy landscape.
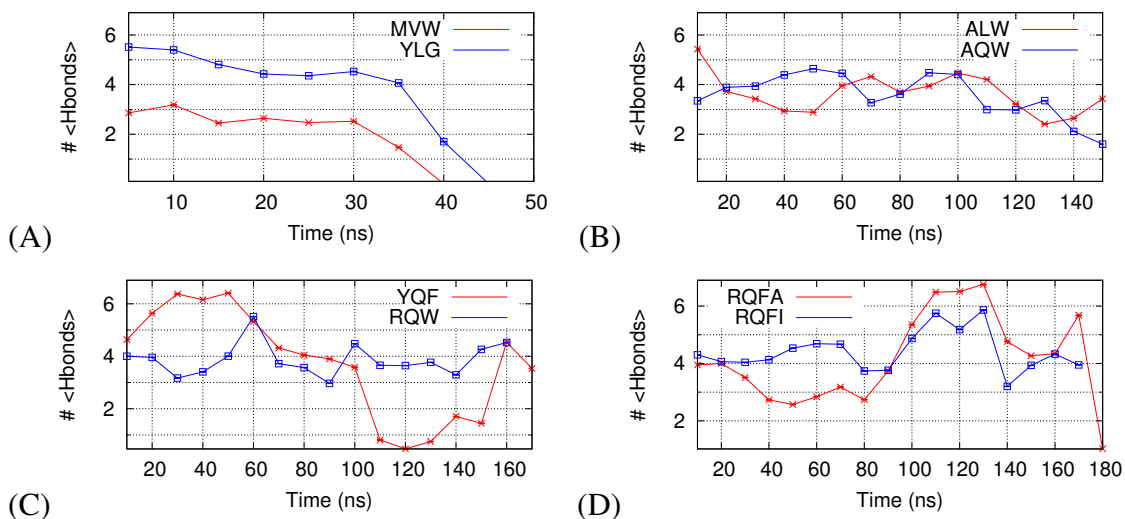
Figure 56: Time evolution of the number of H-bonds between 1E6 TCR interacting with a panel of peptide ligands ranging from antagonist (MVW, YLG), weak agonist (ALW, AQW) and strong agonist (RQW, RQFI, RQFA, YQF). Clearly the strong agonist form an "*In silico*" catch bonds. The weak agonists display a slip or ideal bond. Ideal bond is when the lifetime is independent of force. The catch bond potency, height, and duration are proportional to the experimental peptide potency.

5.3.2 Inside the catch-bond phase, TCRs start rotating with a polarity around the binding axes defined by the right-hand rule.

Previously it was shown that RA14 TCR changes its docking orientation by gyrating around the binding axes to form a catch bond. C7 and C25 TCRs dynamically gyrate around the binding axes to form a *in silico* catch bond and thus experience a transitional change in the docking orientation. For instance, C25 increases while C7 decreases their docking angle. There is a subtle swing in the angle of C25 without breaking any H-bond. It is because the C25 interaction with the pMHC is centrally focused where most of the contacts are located around Met5^P. This provides a small freedom of rotation while still forming H-bonds. Nevertheless, the C25 TCR is always moving to and staying in the active conformation. For both TCRs, the angles change about 20°. 2C TCR also has this common feature of rotating and increasing the number of H-bonds at the binding interface. Although 2C TCR displays a relatively small change in the docking angle ($\approx 7^o$), it is statistically significant based on their normal distribution of both states and the p-value $< 10^{-16}$ reported by a Welch two

sample t-test (data not shown). This is consistent with 2C having a higher affinity to the super agonist peptide. Unfortunately only 2C TCR forms a stable docking angle; however, C7, C25, and 2C are dynamically trying to gyrate counter-clockwise around the binding axes, see Fig. 57.

For the 1E6 TCR, MVW and YLG epitopes display an *in silico* slip bond. They do not attain a well-defined docking angle; they are changing over time; and they dissociate fast before the complex reaches any stable state; thus, a docking angle cannot be clearly identified. Dissociation occurs first probably because they have a very low affinity, see Fig. 58(A, B).

ALW has a weak affinity but it is high compared to MVW and YLG. It initially decreases its angle in a stepwise manner and reaches a stable state and stays in this state for most of the simulation. Finally, it then attempts to return to its initial value before dissociation occurs. AQW has only one stable state in the whole simulation; however, at the end, it abruptly increases its docking angle and then shortly deattaches from the complex. AQW is the phase transition peptide, see Fig. 58(C, D).

RQF(A), RQF(I), RQW, YQF show clearly well-defined transitional states for the docking angle where the change is bigger ($\approx 30^o$) and significant compared to the other peptides. RQW apparently is dynamically trying to increase its docking orientation to a higher value, displaying cyclic changes in its angle. The states are also longer-lived. The complex gyrate counter-clockwise around the binding axes around the catch bond phase, see Fig. 58(E, F, G, H).

Interesting, the large transition in the angle corresponds to changes in the # of H-bonds at the binding interface for the peptide ligands. all TCR-pMHC complexes are free to rotate in any direction around the binding axis in free MD. However, when the complex is inside the catch bond phase for agonist, or around dissociation for weak ligands, all TCRs apparently start rotating around the binding axis with a defined "polarity". The direction is defined by positioning the thumb of the right-hand along the TCR's largest principal

axis and pointing to its membrane-distal binding site; then rotation is defined by curling the other fingers. This is a counter-clockwise rotation around the TCR main axis. This is apparently an intrinsic property of the TCRs since all simulations display this behavior. Changing the docking orientation in step and attaining stable angles is not observed on those TCRs with the exception of 2C. 2C behaves in a similar fashion as described for RA14 interacting with agonist NLV-HLA2.

(A)   (B)   (C)
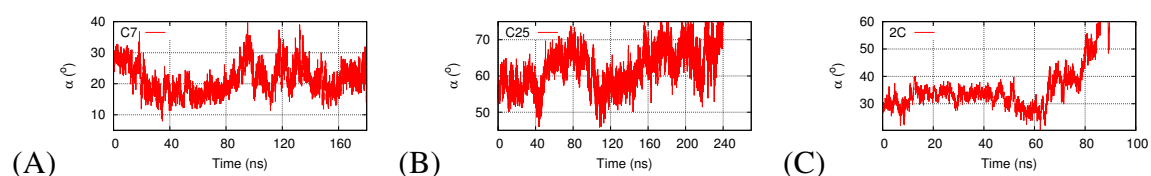
Figure 57: The docking orientation around the binding site for the C7, C25, and 2C TCRs interacting with their respective agonist pMHC in the presence of force. Only 2C forms a stable docking angle; C7 and C25 are dynamically trying to gyrate counter-clockwise around the binding axes.

(A)   (B)   (C)
(D)   (E)   (F)
(G)   (H)
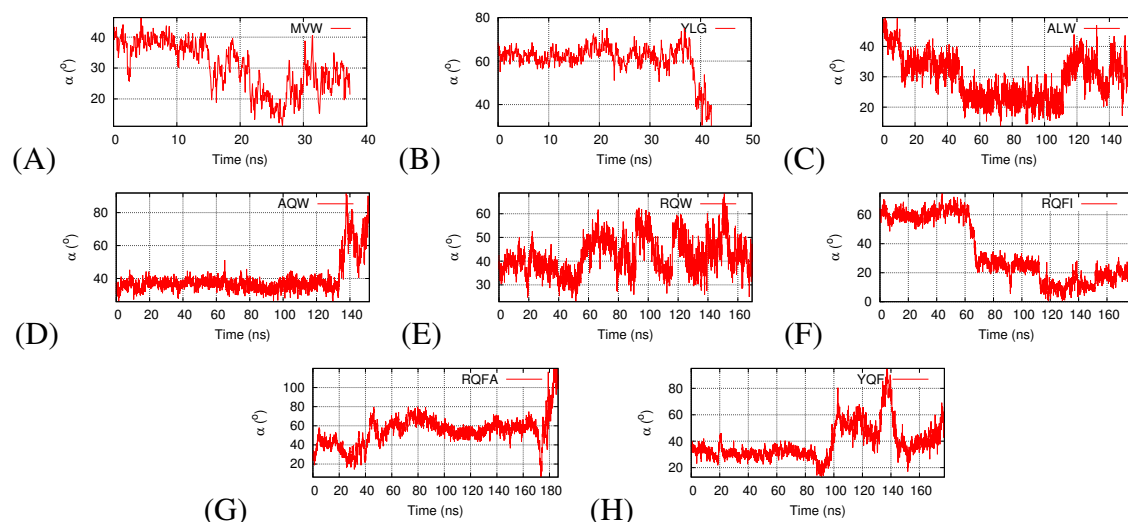
Figure 58: The docking orientation around the binding site for 1E6 TCR interacting with a panel of ligands classified as antagonists (MVW, YLG), weak agonists (ALW, AQW) and strong agonists (RQW, RQFI, RQFA, YQF) in the presence of force. MVW and YLG complexes do not attain a well-defined docking angle; the agonist complex displays multiple longer-lived stays. Apparently weak- and strong-agonist complexes gyrate counter-clockwise around the binding axes either around or after passing the catch bond phase or critical force.

Potentially the TCR dynamically rotating could be biologically functional and a possible explanation is that potent epitopes amalgamate the TCR with pMHC and allow the propagation of force by converting the linear load to angular torque, twisting the TCR in a specific polarity and producing conformational changes. This is supported by the fact that the TCR's structure is very asymmetric with respect to a fragment antigen-binding (Fab); the TCR C_α and C_β domains are not attached similarly as the ones in the Fab. They are connected asymmetrically exposing residues in the interface of the β-sheet of the C_β that are buried in the Fab. The asymmetry also impacts the buried surface area between V_β and C_β domains which are higher than the Fabs and probably they are stabilized by the C_β FG loop. Note that Fab has almost identical 3D structure compared to the TCR.

5.3.3 TCRs rotating around the binding axis induce functional conformational states in the peptide

In the previous chapter, it is observed that in order to form a catch bond, not only RA14 needs to rotate and change its docking angle but also the peptide obtains a stable conformational state. This is clearly observed in the RMSD of the epitopes interacting with human C25 and C7, 1E6 TCRs, and mouse 2C TCR. Fig. 59 and 60 shows that the peptides forming catch bonds have at least two states and the transition from the initial inactive (off) to active (on) states occurs in a stepwise fashion. The forced-induced conformation is long-lived, well-defined, and stable; it is apparently optimal for forming new H-bonds at the binding interface; and the peptide remains in the "on" state inasmuch as the complex is forming a catch bond. Importantly the number of long-lived states increases with the potency of the epitope. This suggests that: First, multiple-state epitopes permit an easy scanning by the TCR and enhance their engagement and these epitopes move in the antigen-binding groove and increase the probability of interacting with the TCR CDRs. Second, not only mobility is necessary for a successful engagement, but also the TCR CDR3 loops have to be mobile and vibrate at the same frequency of the peptide. Third, after a successful

engagement, force modulate the flexibility of the epitope, docking the TCR-pMHC complex into a stable angle, and the epitope becomes rigid at catch bond phase or critical force. This increases the adhesion of the complex.

Conversely epitopes forming slip bonds (MVW, YLG, ALW, T8V, M5I) have either one state or they initially try to achieve a second state; however, they are unstable and unable to sustain this conformation for a longer period of time. Thus, it unsuccessfully returns back to its original state and continuously moves randomly until it dissociates from the TCR. The phase transition peptide, AQW, has multiple states that initially display smaller fluctuations and its conformation finally reaches a long-lived state. It is possible to consider those states as a single one since they have a small deviation in their RMSD; this suggests that their structural difference is probably insignificant (also p-value > 0.05).

As a summary potent peptides are mobile and malleable; thus their conformation can be modulated when the TCR is reading out the peptide ligands. Mobility and malleability are high at the initial binding; however, they are reduced when the epitope is progressing to the active state and it is docked into a rigid conformation. This would only occur if it forms dynamically a complementary match with the TCR.
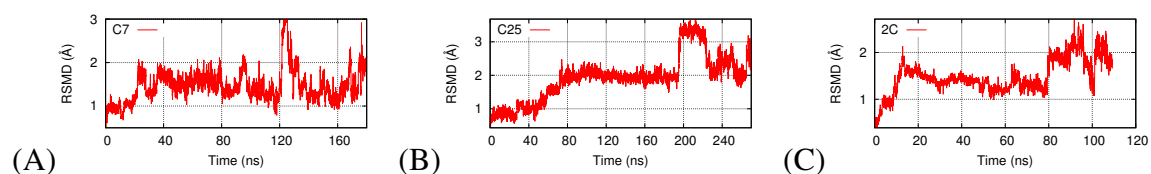
Figure 59: The RMSD of each peptide for C7, C25, and 2C TCR interacting with their agonist pMHC in the presence of force. The TCRs modulate the peptide conformation; it is shown by multiple states in the RMSD.

5.3.4 TCRs forming catch bonds experience the opening and closing of their C_β FG loop

Pulling the TCR ligated with pMHC from the C-terminus "allosterically" regulate the conformations of the C_β FG loop while the pMHC C-terminal is constrained in this process. This happens only for complexes capable of forming *"in silico"* catch bonds. The C_β FG
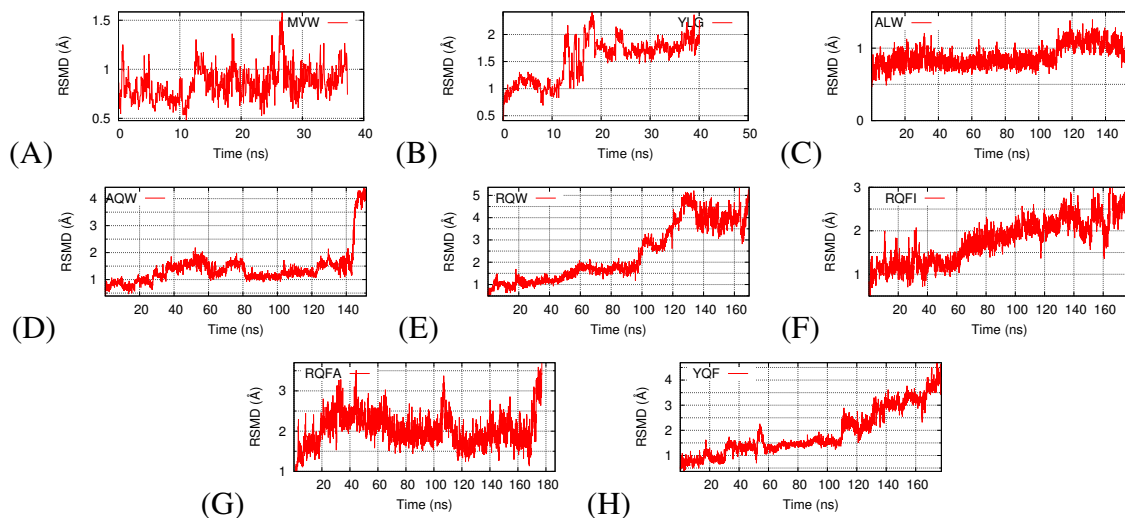
Figure 60: The RMSD of 1E6 interacting with a panel of peptides in the presence of force. Epitopes forming catch bonds have at least two states and the transition from the initial inactive (off) to active (on) states occurs in a stepwise fashion. For potent epitopes Mobility and malleability are high at the initial binding; however, they are reduced when the epitope is docked into a rigid state if it formed dynamically a good match for the TCR.

loop is a string comprised from Leu219$^\beta$ to Pro232$^\beta$ that is extending from the body of the β chain. The C$_\beta$ FG loop is a solvent-exposed, highly-charged 14-residue insertion at the β-chain constant domain and situated alongside the chain. In humans, the C$_\beta$ FG loop has a distinctive charge distribution: 2 glutamate, 2 aspartate, 1 arginine, and 1 lysine residues, while mice, it has: 3 glutamate, 1 aspartate, and 2 lysine residues. In both cases, the FG loops are negatively charged and protruding from the elbow region of the TCR β chain. Residues Leu219, Asp223, Trp225, Lys231, and Pro232 are highly conserved in human and mouse while the reminding ones are relatively shared among multiple species. Fig. 4 shows the overall location of the C$_\beta$ FG loop (A), its atomic details(B) and the common residues shared among species. What is intriguing from the structure of human TCRs, is that the C$_\beta$ FG loop is well-structured having its charged residues spatially arranged so that all negative are facing away while positively charged are facing inward the center of the loop; they are also separated by charged and cluster at each FG loop end. Lys229$^\beta$ interacts with the backbone of Asp223$^\beta$, while Trp225$^\beta$ and Pro232$^\beta$ form strong hydrophobic interactions with β chain. This forms and provides a rigid frame for the C$_\beta$ FG loop as observed

in the crystal structure and here it is termed as the "**neutral**" conformer (Fig. 71 A). The neutral state is also very stable and rigid in free MD simulations (data not shown).

5.3.5   The forced-induced states of the C_β FG loop at the catch phase

In the presence of force, the C_β FG loop undergoes conformational changes indicated by first deforming and becoming flexible, then emigrating away from the V_β-C_β interface. The C_β FG loop then comes to a "**down**" or "**open**" conformation (Fig. 61 B) after moving downward from the elbow region at this point. Subsequently, it returns back to the "neutral" state to further advance and become very close to the V_β domain. This is termed the "**up**" or "**closed**" conformer (Fig. 61 C). All conformational states of the C_β FG loop are displayed on Fig. 61. Close and open states are quantified by measuring the RMSD using the neutral conformation as a reference. RMSD measures the average displacement of the atoms of the C_β FG loop after optimal superposition. An RMSD of ≈ 3Å and ≈ 5Å correspond to a closed and open C_β FG loop, respectively. Importantly the presence of force fosters the transition between conformations of the C_β FG loop.

Fig. 62 and Fig. 62 show correlation between the opening of the C_β FG Loop and the occurrence of the catch-bond at constant-velocity and constant-force pulling, respectively, for the RA14 ligated with agonist NLV-HLA2. The C_β FG loop always remains in the closed conformation in Free MD simulations or when the pulling force is small. There exists a positive correlation between the opening of the C_β FG Loop and the rise of a computational catch-bond– represented by an increase in the number of H-bonds between the TCR and the pMHC. In both simulations are observed that the C_β FG Loop transitions only to the open state; it also experiences three cyclic transitions between "neutral" to "open" conformation inside the catch-bond phase; finally the complex dissociates before the C_β FG Loop reaches the closed state. Apparently, the opening is a reversible process since the C_β FG Loop returns to the neutral conformer once the TCR is detached from the peptide ligand.
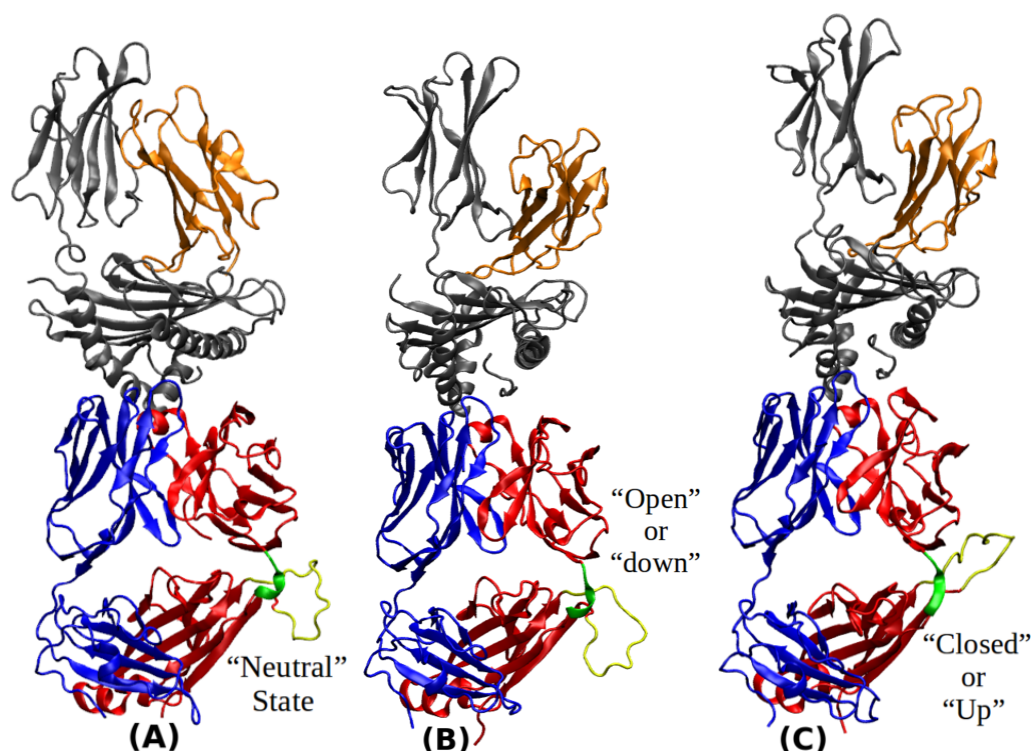
Figure 61: Illustration of the C_β FG loop states in the presence of force. The C_β FG loop is colored in green. The MHC is loaded with an agonist RQFI peptide. (A) shows the neutral state, this is common at the beginning of SMD simulation where force is little. (B) show an "open" or "down" FG loop, the simulation is at the "*in silico*" catch-bond phase, where force is moderate and the number of H-bonds is at the peak. The FG loop is "closed" or "up"around or right before dissociation from pMHC only for potent epitopes as shown in (C). This state is only achieved by strong agonist peptides.

5.3.6 Quantification of the states of the C_β FG loop of C7, C25, 2C, and 1E6 TCRs forming "*in silico*" catch bonds

All complexes in Fig. 64 undergo constant-velocity pulling and force starts from 0 pN. C7 and C25 TCR initially opens the C_β FG loop as indicated by an increase in the RSMD, Fig. 64 (A, B). C25 shows cyclic transitions between open and neutral conformation. There are a total of five cycles, three is long-lived while two are short-lived. C7 has only two long-lived open states. Although C7 experiences smaller deformation at the end of the simulation, both TCRs converge to the closed conformer and settle to this state; C25 experiences the largest deformation as compared to C7. Nevertheless, before dissociation and exiting the catch-bond phase, both TCRs achieved the close conformation at the end of the
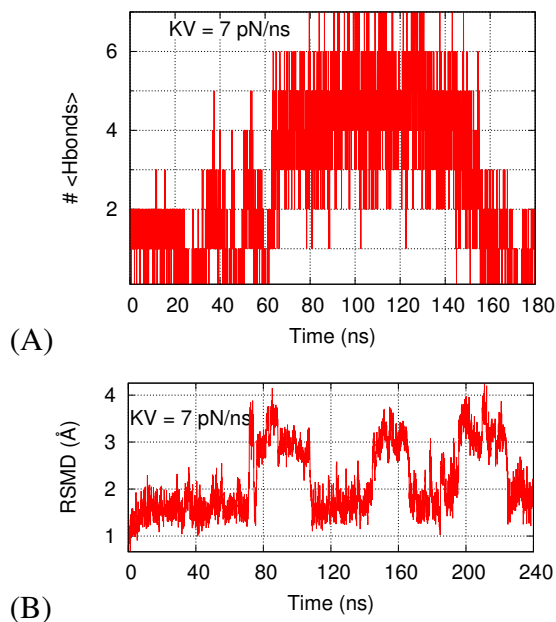
Figure 62: Correlation between the opening of the C_β FG Loop and the occurrence of the catch-bond. (A) shows the total number of H-bonds between the TCR and pMHC in the presence of force. (B) shows the RSMD of the C_β FG loop. The increase in the RMSD from 1.8 to 3.5Å correspond to the "opening" of the C_β FG loop. The complex underwent a constant-velocity pulling.
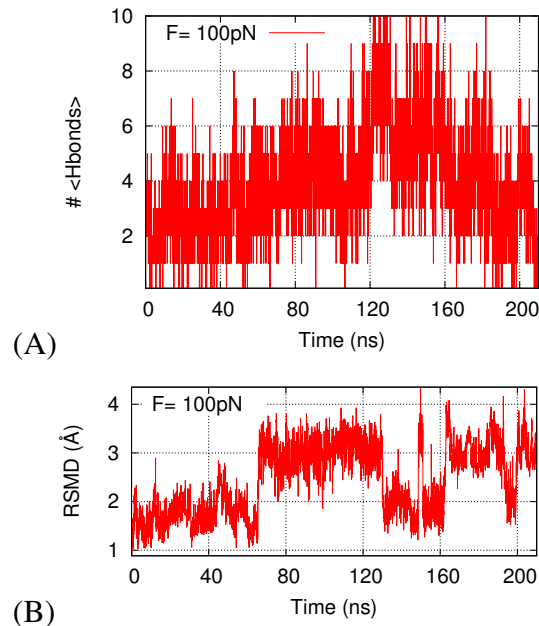
Figure 63: Correlation between the opening of the C_β FG Loop and the occurrence of the catch-bond. (A) shows the total number of H-bonds between the TCR and pMHC in the presence of force. (B) shows the RSMD of the C_β FG loop. The increase in the RMSD from 1.8 to 3.5Å correspond to the "opening" of the C_β FG loop. The complex underwent a constant-force pulling.

simulation.

2C TCR also experiences deformation in the C_β FG loop and initially opens it; however, after this initial transition, it starts moving to the closed conformation as shown by the monotonically increase in the RMSD (Fig. 64 C). In the simulation, the C_β FG loop only experiences larger deformations that lead to its closing; however, it is not clear what prevented its complete opening and its cyclic transitions. It is observed that Lys11$^\beta$ residue initially forms a strong H-bond with Asp223$^\beta$ and at the closed conformation, the C_β FG Loop forms multiple H-bonds (≈ 3 to 4) with residues located in the first strand at the V_β domain, before 2C dissociates from the complex. The H-bonds form strong interactions as they are arranged into a β sheet in the closed state. See Fig. 65 and read the caption

for details. Note that the 2C is a transgenic mouse TCR that shares only key residues with human TCRs; thus they are not easily compared.

In Fig. 64(D, E, F), the 1E6 TCR is interacting with a panel of peptide ligands, and it is observed that all ligands from weak (D) to strong (E and F) agonists are able to induce the opening of the C_β FG Loop as indicated by the increase in their RMSD (≈ 3Å). Weak agonists do not display cyclic transitions between states. They only switch one time from neutral to open; however, it is a long-lived state, (Fig. 64 D). All strong agonists wave the TCR C_β FG Loop (neutral to open) from two to five times (Fig. 64 E, F). In addition, at the end of the simulation, all strong agonist switched to closed conformer before exiting the catch-bond phase and dissociating from the peptide ligand. This is quantified by the large increase in the RMSD. The strongest agonist (YQF and RQFA) undergo larger deformation than others, RQW and RQFI complexes. Apparently, they are proportional to the peptide potency. Antagonist YLG is also able to open the C_β FG Loop; however, it is a very short-lived state since it occurs when it is dissociating from the pMHC. Antagonist MVW does not induce any substantial deformation of the loop.

In summary, for potent peptide ligands, there is a positive correlation between the deformation of the C_β FG Loop and the appearance of a catch-bond. Force fosters the transition between states of the C_β FG loop only when the number of H-bonds increases at the binding interface of TCR-pMHC. The transition between neutral to open only occurs inside the catch-bond region. First, the open state happens and it is then followed by the closed state. The closed conformer usually appears at the end of the simulation, after passing the critical force and when the complex is exiting the catch phase. Potent peptide ligands also generate more cyclic transitions to finally achieve the closing of the FG loop, and it experiences large deformation proportional to their potency.
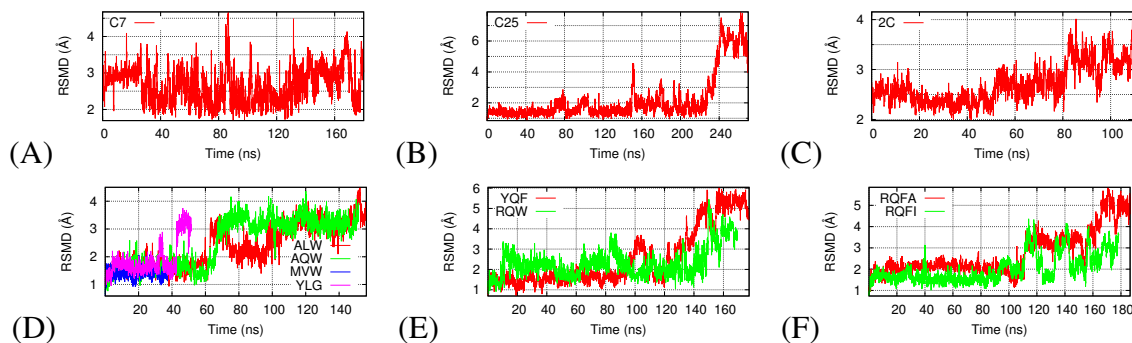
Figure 64: Time evolution of the RMSD of the C_β FG loop of C7(A), C25(B), 2C(C) TCRs interacting with their agonist epitopes in the presence of force. (D), (E), and (F) display the RMSD of the C_β FG loop of 1E6 TCR interacting with a panel of ligands classified as antagonists (MVW, YLG), weak agonists (ALW, AQW) and strong agonists (RQW, RQFI, RQFA, YQF). All epitopes with the exception of MVW and YLG are capable inducing the opening of the C_β FG loop, represented by changing their RMSD to ≈ 3Å. The deformation and duration of the state depended on the epitope potency. Only TCR interacting with potent epitopes transition the state of their C_β FG loop to the closed conformer; 1E6 complex with RQW, RQFI, RQFA, YQF and C25 and 2C complexed with their respective agonist epitope display the closed state before dissociating form the complex.
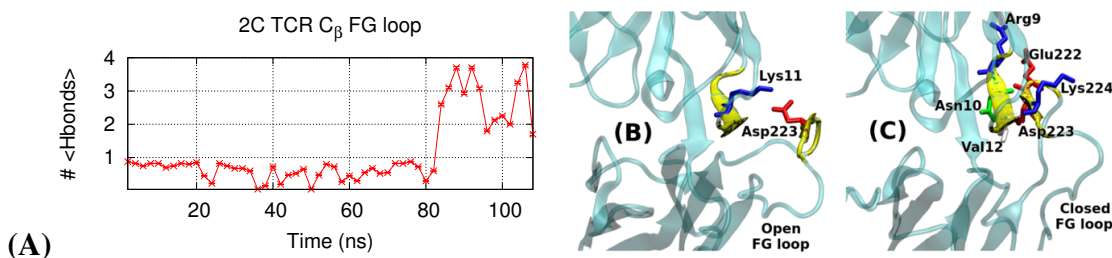
Figure 65: Forced-induced H-bonds between the C_β FG loop and the V_β domain for the 2C TCR. They initially form a strong H-bond between Lys11$^\beta$ and Asp223$^\beta$, not present in the crystal structure. When the FG loop switches to the closed conformation, the H-bond is replaced by 4 other H-bonds: Lys 224 forms 2 H-bonds with Asn10 and Val12; Asn10 forms one with Asp223; and Arg9 forms one with Glu222. (A) Show the total number of H-bonds between the C_β FG loop and the V_β domain over simulation time. (B) and (C) depict those interactions in open and closed conformation.

5.3.7 Quantification of the states of RA14 C_β FG loop forming an *"in silico"* catch bond

In the chapter 4, it is studied the force response of RA14 complexed with, antagonist (M5I), weak agonist(T8V) and agonist (NLV) peptide ligands under different conditions. The initial configuration (IC) is varied using different ramping rate with constant-velocity pulling

on those ligands. The ramping rates are defined as normal, medium, and high. A total of 8.9 μs of simulation is generated. The focus here is to evaluate the conformational changes induced by the presence of force in the C_β FG loop in these simulations. Fig. 66 shows the SMD simulations using constant-velocity pulling for RA14 interacting with its peptide ligands, and they are ordered in columns and colored as NLV (red), T8V (green), and M5I (blue) complexes. The varying parameter is indicated on the label of the Figures, for instance: SMD1 and SMD3 represent the simulation using the IC1 and IC2; KV represents the simulation changing the ramping rate from normal to high.

In Fig. 66, NLV complexes (red curves) show that the C_β FG loop undergoes deformation and cyclic transitioning from neutral to open for all IC pulled independently at the normal ramping rate (Fig. 66 A, D, G). They oscillate between neutral and open from three to five times and these states are long-lived. For T8V complexes (green curves) pulled at normal ramping rate (Fig. 66 B, D, H) are observed that although the C_β FG loop experiences deformations, it does not completely transition to the open conformer; whenever it is increasing the RMSD, it returns to the base level. The deformation is also periodically occurring. For M5I complexes (blue curves) show that the C_β FG loop is opened once (Fig. 66 C, F, I); however, this happens at the expenses of unfolding the MHC α_3 domain. At the same time, it was also unable to switch to open and also unfolding happens in the simulations. Nevertheless, deformations are also observed for the M5I complexes.

In summary, the dependence of the state of the C_β FG loop on the ramping rate is displayed in Fig. 66 (J, K, L, M, N, O). It is shown that increasing the ramping rate attenuates the catch-bond by decreases its total number of H-bond and its duration. This is translated into a rapid dissociation of the complex and less force is being propagated to induce considerable deformation of the C_β FG loop. For instance, the NLV complex is still experiencing small cyclic deformations at the medium ramping rate, Fig. 66 (J); this is drastically reduced at higher pulling as seen in Fig. 66 (M). Similarly for T8V and M5I complexes shown in Fig. 66 (K, N, L, O), respectively; increasing the ramping rate renders the peptide

ligands unable to induce any detectable deformation in the C_β FG loop. It also appears that all peptide ligands ranging from weak to strong agonist may initially induce the deformation of the FG loop and open it to a certain degree when pulled at a normal ramping rate; however, antagonist or weak agonist is not capable of inducing the closed state. It is clear that the high ramping rate renders the peptide ligands unable to produce conformational changes in the FG loop. This is because force starts building up rapidly at the interface and it does not allow the complex to rearrange and sustain the force, then any transition (or state) is diminished and only the neutral state is most dominant. This similar to making mutations or reducing the potency of the peptide.
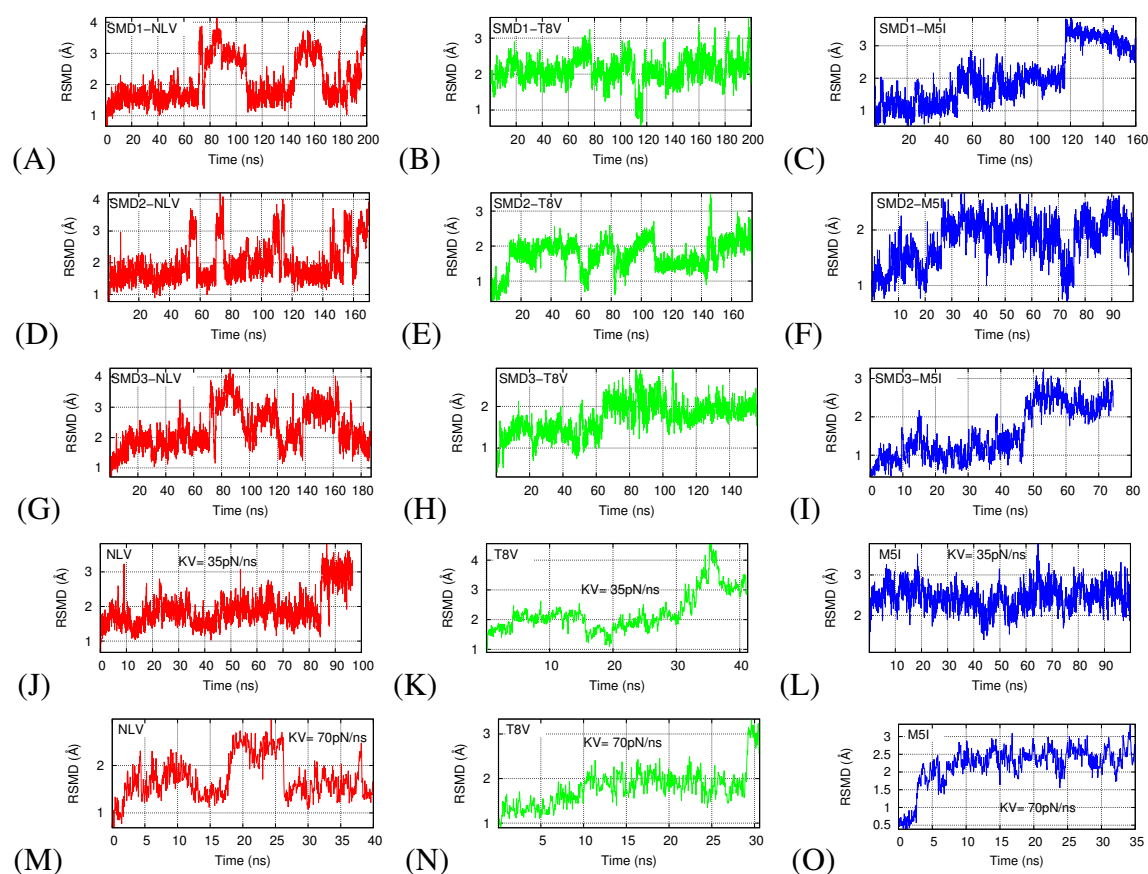
Figure 66: Time evolution of the RMSD of the C_β FG loop of RA14 TCR interacting with peptide agonist NLV (A), mutant weak agonist T8V (B), and antagonist M5I (C) in the presence of force using different simulating parameters. Three different IC and three varying ramping rates (normal, medium and high) are employed. The peptide ligands are ordered in columns and have a color coding of red for NLV, green for T8V, and blue for M5I. The changing parameter is indicated on the label of the Figures.

It is also varied the SMD protocol by pulling the RA14-NLV-HLA2 using constant-force. Of note, this part is conducted only in the NLV complex because simulating the mutants is computationally demanding. Six constant forces are employed varying from 0 (free MD) to 200 pN. At zero force (free MD, data not shown), deformation is not observed in the C_β FG loop at the nanosecond scale. As force is increased to 25 pN, three short-lived open states are observed; importantly they appear to be unstable and the states return back to the neutral configuration at this level of force, Fig. 67 (A). As force is increased to 50 pN (Fig. 67 B), at the beginning of the simulation, there starts appearing two distinguishable short-lived deformations of the C_β FG loop and it finally settles to the open conformation for about 40 ns. Subsequently, it undergoes substantial deformations (at 160 ns) that lead to a rapid transitioning from open to the closed state; this final state is long-lived and stable and last for about 40 ns. At this point, the RA14 attempts unsuccessfully to open C_β FG loop (at 200 ns) and is forced to experience large irreversible deformations to finally close the C_β FG loop. Dissociation is not observed at this level of force and time scale. As force is increased to 90 pN (Fig. 67 C), the transitions between states become cyclic as the C_β FG loop oscillates between neutral and open conformer four times; the cycles are also long-lived and stable. Finally, at this force, it achieves large deformations that induces the closing of the C_β FG loop and the TCR dissociation from the peptide ligand is observed. The closed conformer appears to be stable. At a force of 100 pN (Fig. 67 D), the complex only displays three long-lived and stable cyclic transition from open to neutral. At this force, the complex is unable to reach the closed conformation before TCR dissociation occurs. The open conformation is a reversible process; after the TCR dissociates from the pMHC, the C_β FG loop returns to neutral conformation as seen in Fig. 68. At 150 pN (Fig. 67 E), the complex displays only one transition from neutral to open and afterward it dissociates. For forces $\geq$ 200 pN, (Fig. 67 F), there is no noticeable deformation of the C_β FG loop; dissociation is achieved rapidly in these cases. Note that 50 pN is the critical force in the "*in silico*" catch bond (Fig. 19 B) and 90 and 100 pN are close and considered

optimal forces in terms of increasing the total number of H-bonds at the binding interface.

In summary, force modulates the states of the C_β FG loop and it is observed that: 1) It decreases the complex dissociation time. 2)It increases the number of cyclic transition between reversible states. 3) After the threshold force is passed, the number of cycles is reduced. 3) The closed conformation is observed around to the critical force and it is reached at the end of the simulation; it appears to be a one-time transition. 4) Transitioning to the open conformer is reversible while closing the C_β FG loop is an irreversible process. Note that deformation at low or high force is minimal and undetectable.
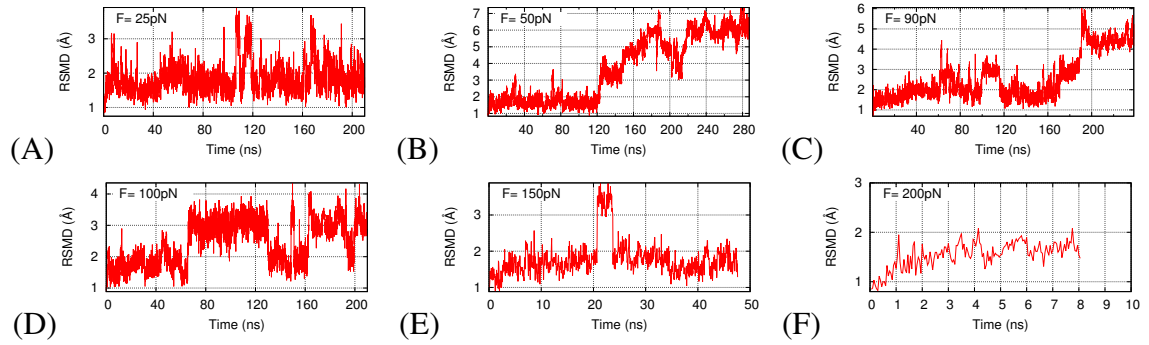
Figure 67: Time evolution of the RMSD of the C_β FG loop of RA14 TCR interacting with agonist NLV peptide ligand using six constant forces of 25, 50, 90, 100, 150, 200 pN.

It is observed that the state of the C_β FG loop is easily modulated by many parameters such as changing the ramping rate, increasing the force, mutating the epitope, replacing by potent epitope, or even using a different species TCR C_β FG loop. However, it generates new scientific question: 1)Why does opening appear first and closing usually occurs later when the complex is exiting the catch-bond phase? 2) Why does the open-to-neutral transition present periodic cycles? and is the closing a one-time event? 3) Why are the transitions (from open to neutral) reversible processes while closing is irreversible for potent ligands? Do forced-induced states of the C_β FG loop relates to the specificity or sensitivity? Apparently, the human C_β FG loop is more mobile, and thus more sensitive than mouse TCRs. This results in the fact that the humans have more T cells (one order-of-magnitude or so). The next section is focused on addressing these questions.
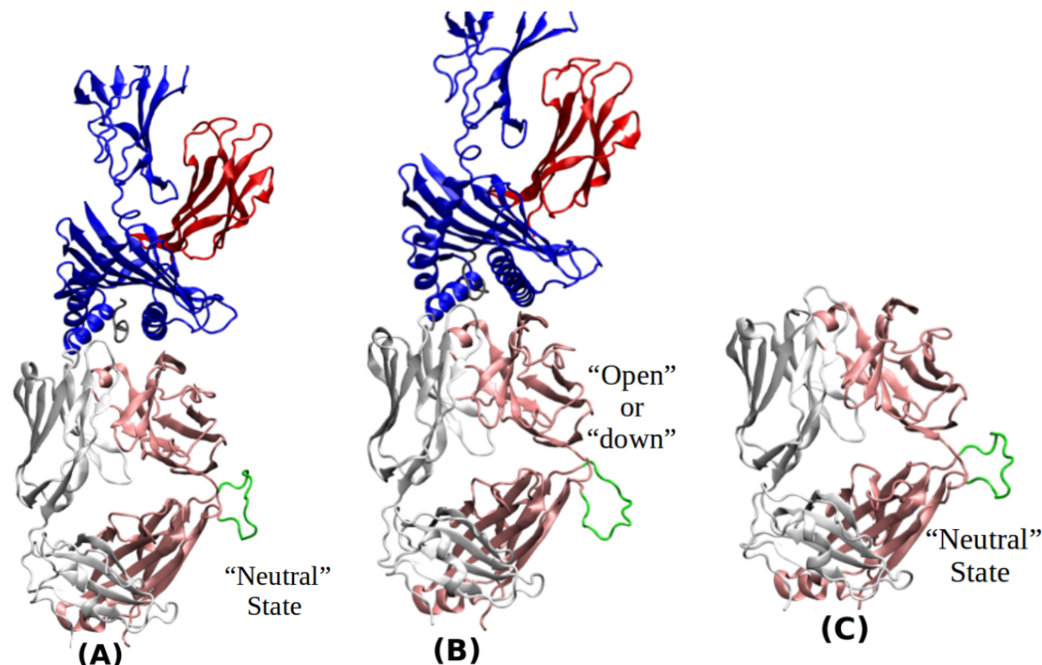
Figure 68: Representation of the reversible states of the C_β FG loop at different simulation times in the presence of force. The FG loop is colored in green. The MHC is loaded with a RQFI peptide. (A) shows a FG loop in the neutral state when the simulation is beginning and force is little. (B) show an open FG loop; the simulation is at the middle and force is already built-up; the number of H-bonds is also at the peak. The FG loop returns back to the neutral state after the TCR dissociates from pMHC as shown in (C).

5.3.8 Rotating around the binding axis and waving the FG loop is an intrinsic property of TCRs

Normal mode analysis (NMA) is used to observed large-scale motions in the TCR-pMHC complex, based solely on the structure without including molecular details of the residues that form the complex. There is a question of whether the observed conformational changes are typical functional modes of TCRs interacting with pMHCs (intrinsic property). HM-MER is used for searching the databases for sequence homologs (Fig. 66 A, B). Based on the HMMER search, BLAST is used to identify and collect protein species down to a pairwise sequence identity of 70% (Fig. 66 C). There are about 275 structures and the list of unique species is displayed in Table 2. Interesting RA14 share its sequence identity broadly with multiple species while 1E6 is more restricted. RA14 is a public CMV-specific

Table 2: List of unique species identified using the searching database HMMER for sequence homologs. RA14 share its sequence identity broadly with multiple species while 1E6 is more restricted. RA14 is a public CMV-specific TCR while 1E6 is an autoimmune specific TCR for type 1 diabetes.

RA14	1E6	Identity(%)
[1] Homo sapiens [2] Mus musculus [3] Rattus norvegicus [4] Oryctolagus cuniculus [5] Bos taurus	[1] Homo sapiens [2] Mus musculus [3] Manduca sexta [4] Escherichia coli	$\geq 70\%$
[6] Macaca mulatta [7] Gallus gallus [8] Cricetulus migratorius [9] Lama Glama		<70%
1167	400	Total hits

TCR while 1E6 is an autoimmune TCR. The protein query is performed based on TCR β chain sequences. Blast is used to search the PDB database and to identify related structures. The query is constrained to mammal proteins. NMA is performed in the identified structures under study and it shows that these complexes have a remarkable similarity in their profile's fluctuations. The analysis reveals that all structures have as significant angular displacements around the binding axis and linear displacements for the C$_\beta$ FG loop. For instance, modes 7 and 8 of the NMA correspond to rotating around the binding axis while mode 9 to 14 correspond to large displacements in the C$_\beta$ FG loop. Apparently, these 2 modes are essential and functionally important for the TCR-pMHC complexes since all studied structures acquire those modes. Remarkably the TCR α chain does not display any conformational change or large displacement while the TCR β chain is flexible in most of its loops. Notably MHC α_1, α_2, and the β_2 microglobulin are very stable having little-to-no deformation; however, α_3 domain becomes flexible and its modes (12 to higher) resemble a potential unfolding of this domain from its C-terminal.
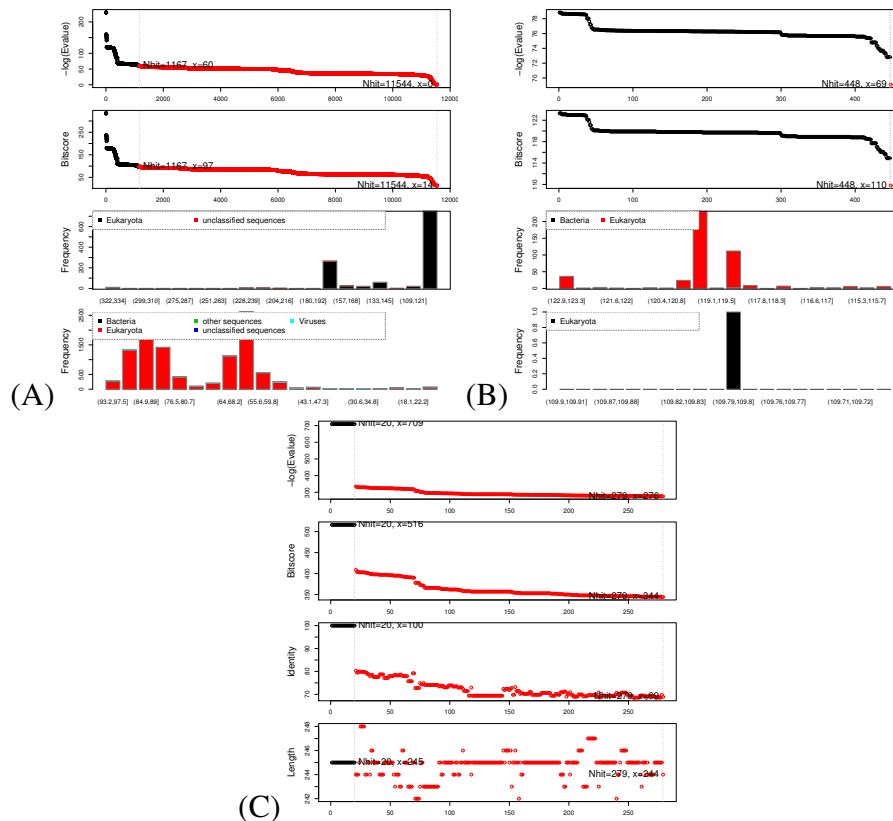
Figure 69: Summary of protein query produced by the HMMER and BLAST search: (A) and (B) show the number of hits obtained when searching the database with HMMER. The two upper figures in (A) and (B) show the normalized score and bitscore, respectively. The remaining two lower figures show the distribution of significant hits according to the bitscores and taxonomic kingdom, respectively. The first bar plot corresponds to the hits above the cutoff, while the second bar plot corresponds to the hits below this cutoff. The majority of the top scoring hits is from Eukaryote (red) while the others correspond to other species that are considered insignificant. (C) Blast is used to search the database for top scoring structures. The NMA is employed only with the top scoring hits (black).

5.3.8.1 *The partial least-squares functional mode analysis*

The partial least-squares (PLS) functional mode analysis [157] is used to identify and determine systematically hidden relation between coordinates and functional conformational changes. The algorithm not only searches for the best relation in the trajectory simulation between the atom coordinates that correlates best with the functional order parameter, α^o, but also, find the one that contributes the most to the fluctuation in α^o. The algorithm can be also employed to established relations among functional parameters in well-defined

conformational states.

The number of H-bonds, the docking angle, and the RMSD of FG loop are examined as the functional order parameter. All atoms of the protein are used. The 80% of all trajectories are used as a model training and the remaining 20% for cross-validation. Remarkably all simulations display that there is a strong relationship between rotating around the binding axis and inducing conformational changes in the FG loop when these are chosen as order parameter; however, using the RMSD of FG loop provides the most notorious conformations. Figure 70 shows visually the two dependent states for RA14 and their respective order parameter with the training and cross-validation sets.

(A)   (B)
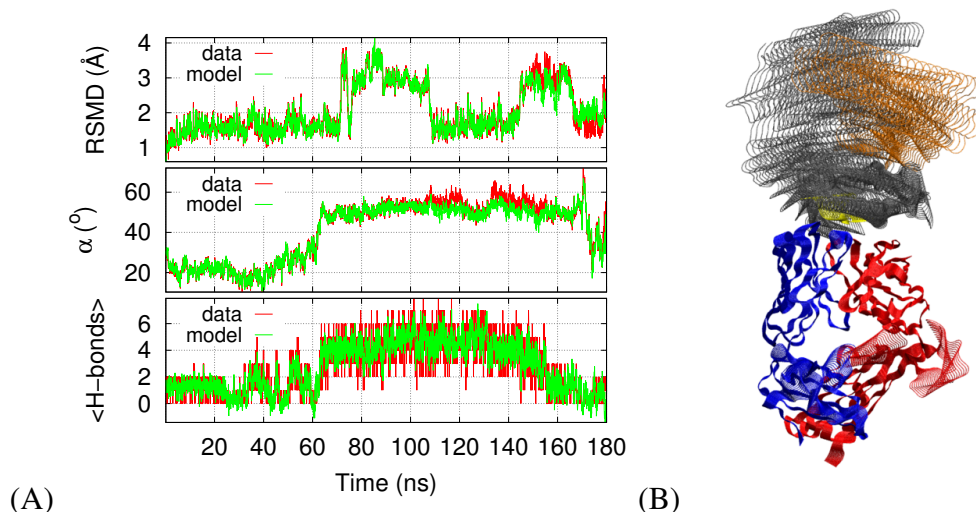
Figure 70: PLS-based functional mode analysis for RA14-NLV-HLA2 forming a "*in silico*" catch-bond. (A) The functional order parameter is calculated from a 200 ns steered MD. The training and cross-validation sets agree very well and show that rotation of the TCR induces opening of the C_βFG loop. (B) Image of the conformational changes of the TCR-pMHC complex, aligned with respect to the TCR.

5.3.9 The deformation of the FG loop is proportional to the peptide potency.

The FG loop is internally well-structured and closely packed to the β chain by a mini-hydrophobic core. This apparently integrates the Vβ and Cβ domains and forms a rigid structural entity. Thus it has no substantial deformation in the absence of force. However, in the presence of force, the FG loop becomes disorder and emigrates away from TCR β

chain. This is measured by the increase in its RSMD as shown in previous sections. The area under the curve delineated by the RMSD of the FG loop is calculated for all epitopes in the absence and presence of force and their ratio is plotted in Fig. 71. High constant-velocity pulling is used for the calculation. This represents the percentage increase in the area in the presence of force when compared to the force-free simulation. The FG loop becomes disorder and experiences larger conformational changes (deformation) when presented with agonist epitope (NLV), while its change is smaller when a weak agonist (T8V) and antagonist(M5I) are presented in the presence of force. This is consistent with our functional studies; agonist NLV produces significant interferon γ (IFNγ) while mutants T8V and M5I produce lower levels of IFNγ even at higher peptide concentrations. A similar analysis is performed for the panel of ligands (from weak to strong agonist) complexed with the 1E6 TCR, and the results are shown in Fig. 71 (B). Strong ligands, (RQW, RQFI, RQFA, YQF), induces the deformation of the FG loop compared to the antagonist (MVW) or weak agonist (YLG, ALW, AQW) and their deformation is proportional to the epitope potency. This correlation is very important since the FG loop is believed to play a critical role in activating T cells. This is also consistent with recent experiments showing that T cells develop adhesion forces with pMHC in a manner that correlates with the peptide potency to induce the production of T cell calcium and interleukin-2 [158].

(A)   (B)
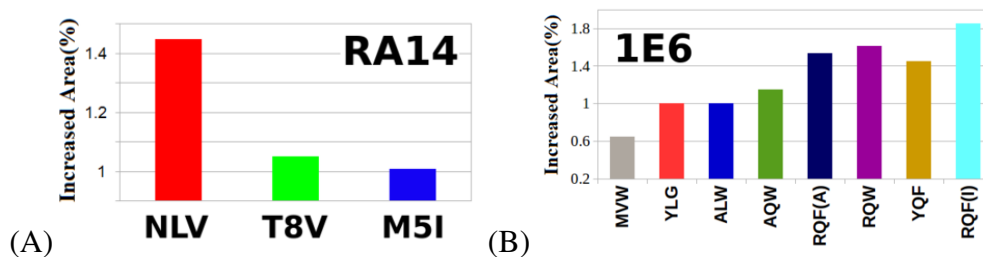
Figure 71: The percentage increase in the area of the C_β FG loop of RA14 and 1E6 interacting with a panel of ligands in the presence of force as compared to force-free simulation. (A) The NLV complex shows an increased of 45% in the area. There is no such increase in mutant T8V and M5I. (B) All strong agonist (RQW, RQFI, RQFA, YQF) induce large deformation of the FG loop.

124

5.3.10 The mechanism of opening and closing of the C_β FG loop

It is shown so far that the Cβ FG loop is deformable in the presence of force and it is reported here that it displays two new forced-induced conformers when it is compared to all available crystal structures up to date. This is the first time that such conformation is observed and it has not been reported in experiments or simulations. The two forced-induced states are defined in the previous section as the **"open"** or **"down"** and the **"closed"** or **"up"** conformations; they are visually depicted in Fig. 68.

In free MD simulations, the Cβ FG loop does not experience deformations; thus, it is in a **neutral** state and this is basically similar to the one observed in the crystal structures. Importantly, the states are not randomly appearing or transitioning from one to another when force is applied to the complex, but their transition (between states) occurs sequentially. In the presence of force The TCR initially opens its Cβ FG loop and stay in this state and cyclically transition from open to neutral and later it completely moves from neutral to the closed state.

5.3.10.1 The dependence of the states of the C_β FG loop on the peptide potency

Similarly, all epitopes (from weak to agonist) initially induce the deformation and open of the Cβ FG loop in the presence of force to a certain degree as it is shown in the previous section. The difference among peptides is that weak ones produce less deformation than stronger ones. However, peptides that are capable to stand some level of force without dissociating the complex, **opens** the FG loop. Interestingly only agonist epitopes are able to transition the Cβ FG loop to a closed state. This is observed in: RA14 interacting with the NLV-HLA2 but not for T8V- or M5I-HLA2; 1E6 interacting with RQW-, YQF-, RQFA-, and RQFI-HLA2 but not with YLG and ALW peptide ligands. MVW and YLG peptide ligands are not able to open the Cβ FG loop; they slightly deform it and then dissociate from the complex.

5.3.10.2 The dependence of the states of the C_β FG loop on the ramping rate

Interesting varying the ramping rate also impacts the states of the C_β FG loop. Normal pulling rates sequentially open and then close the C_β FG loop only for potent ligands. Weaker peptide ligands may induce some deformation and at most the opening. Pulling a low ramping rate also abolish the closing state for all ligands and decrease the deformation of the C_β FG loop. At this rate, only potent ligands induce the opening of the C_β FG loop. There is no significant difference among weaker ligands. This is observed in all ligands complexed with 1E6 and RA14 TCR, where dissociation occurs first and the opening is not seen for weaker ligands. This is shown in Fig. 72, where the 1E6 complexed with a panel of peptide ligands from antagonist to the strong agonist are pulled at slow constant velocity. No difference is observed in the H-bonds profiles of the ligands; indicating that it also diminished the transition between states. At the other end of the pulling spectrum, at high-ramping rates, it also abolishes the open and closed state for the ligands; potent ligands display more deformation of the C_β FG loop and in some cases, an opening can be observed. Finally, another important feature when changing the pulling is that a normal ramping rate increases the dissociation time when compared to low or higher pulling rate. For instance at low or high pulling rate all complex dissociate before or around 50ns, respectively; however, at normal pulling, they dissociate around 200 ns and some potent ligand even go further. In addition either at low or high ramping rate, the *"in silico"* catch bonds are transformed into either an ideal or slip bond for all ligands. No significant difference is established using a Welch two sample t-test; it reports p-values $< 10^{-9}$ with a 95% confidence interval for the samples.

Pulling at constant force also impact the state of the C_β FG loop. RA14-NLV-HLA2 complex is used and its NLV peptide ligand is considered agonist. At smaller forces, there is no significant deformation of the C_β FG loop. As force increases, the opening of the $C\beta$ FG loop starts appearing. At the forces around the catch bond, both open and neutral states are observed multiple times. As forces pass this region, mostly the closed state start

dominating. It is observed that the transition from neutral to open states and vice versa are a reversible process; however, once the closed state is achieved the TCR experience permanent and irreversible deformations in the C_β FG loop. Closed state is only observed for potent ligands. No significant difference is established using a Welch two sample t-test; it reports p-values $< 10^{-7}$ with a 95% confidence interval for the samples.

(A)
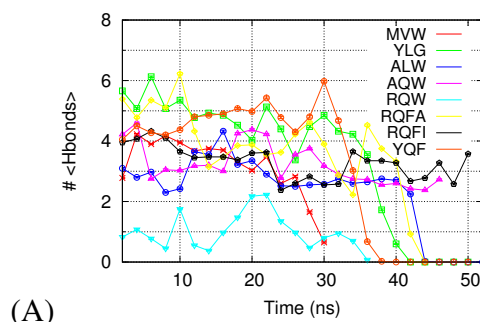
Figure 72: Time evolution of the number of H-bonds between 1E6 TCR interacting with a panel of peptide ligands ranging from antagonist (MVW, YLG), weak agonist (ALW, AQW) and strong agonist (RQW, RQFI, RQFA, YQF) at slow constant-velocity pulling. There is no clear difference in the H-bond profiles between the strong agonist that previously formed an "*in silico*" catch bonds at the normal ramping rate. All epitopes apparently decrease the number of total of H-bonds at the binding interface as force is increased and they appear to form a "slip" or "ideal" bond. Nevertheless, in the simulations, the potent epitopes induce more deformation of the C_β FG loop.

5.3.11 The moment of inertia of the TCR shows that linear force is translated into rotational torque that twists the TCR domains

To this point, it is clear that the FG loop has three states, "open", "neutral", and "closed" and force modulates its state and transitions as explained in the previous section. It is also involved the TCR rotating around the binding axis and reducing the flexibility of epitope. The moment of inertia (MI) is calculated to understand how linear forces are propagated to induce rotation and twisting of TCR, that could potentially induce deformations including the opening and closing of the C_β FG loop. MI dictates the required torque for changing the angular moment about an axis and it depends on the mass distribution. For TCRs, it is visually inspected that its largest principal axis is perpendicular and penetrates the surface of MHC binding groove as seen in Fig. 73 (A). It also goes through the CDRs loops;
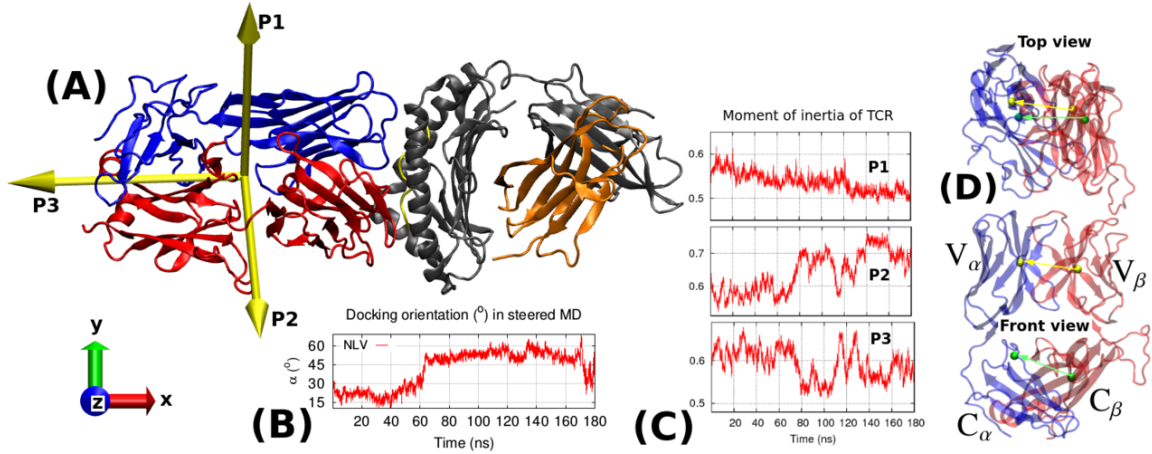
Figure 73: The moment of inertia of the RA14 TCR. (A) P_1, P_2, P_3 are the principal axes of rotation of the TCR. When the TCR rotates around this axis, its angular momentum vector **L** becomes parallel to the angular velocity vector ω. It is important for molecular recognition that P_3 is perpendicular to the binding interface, intersects the peptide ligand, and passes parallel to the interface formed between the TCR $\alpha\beta$ chain. (B) The docking orientation between the RA14 interacting with NLV in the presence of force. (C) The moment of inertia (MI) of the TCR around the principal axis P_1, P_2, P_3; they are normalized assuming that the TCR is a rectangular box. The TCR size in each dimension is approximately, d_{P_1}=5.1 nm, d_{P_2}=5.8 nm, and d_{P_3}=7.7 nm; it has a total mass of, m_{TCR} = 49.3 kDa. It is observed that P2 increases while P3 decreases its value inside the catch-bond phase. P2 corresponds to the axis where the FG loop opens and closes, while P3 corresponds to the axis that stretches or compresses the TCR. P1 slowly decreases and reaches an asymptotic value. Dramatic changes in the MI are indications that the overall cross sectional area along the major axis is impacted in the presence of force. The largest cross section is situated in the cross-sectional plane between the V_α-V_β and C_α-C_β domains; and they behave as an inter-domain rotational hinge. This hinge converts linear force into angular torque that easily twists the TCR from one end while is held from its membrane-distal binding site by the peptide ligand. (D) Vectors depicted in yellow and green are defined to measure the angular rotation of the C_α-C_β respect to V_α-V_β complex. The vectors are situated and passed the center of mass of each Ig-like domains. The TCR is shown in front and top view.

thus a role of changing the MI dynamically while the TCR rotates around to the binding interface is expected in the presence of force. The MI of the TCR in the presence of force is calculated and shown unitless in Fig. 73 (C). The values, P_1, P_2, P_3, are normalized assuming the TCR is a rectangular box with dimension in the axis of d_{P_1}=5.1 nm, d_{P_2}=5.8 nm, and d_{P_1}=7.7 nm, and a total mass of, m_{TCR} = 49.3 kDa. Then the TCR theoretical

128

moment of inertia would be in each axis as follow

$$MI_{P_1} = m_{TCR}\left(\frac{d_{P_2}^2 + d_{P_3}^2}{12}\right) \tag{16}$$

$$MI_{P_2} = m_{TCR}\left(\frac{d_{P_1}^2 + d_{P_3}^2}{12}\right) \tag{17}$$

$$MI_{P_3} = m_{TCR}\left(\frac{d_{P_1}^2 + d_{P_2}^2}{12}\right) \tag{18}$$

P_2 increases while P_3 simultaneously decreases its value inside the catch-bond phase. P_2 correspond to the axis where the FG loop opens and closes while P_3 correspond to the axis that stretches the TCR. Interesting, changes in P_2 is compensated with variations in P_3. P_1 slowly decreases to reach an asymptotic value. Such dramatic changes in the MI is an indication that the overall cross sectional area along the major axis decreases (or increases) while it is compensated in the other axis in the presence of force. The largest cross-section is situated parallel to the P_1-P_2 plane and it splits the TCR at the middle into halves. This structurally divides the V_α-V_β and C_α-C_β in two independent domains and they behave as an inter-domain rotational hinge. In other words, a linear force is converted into angular torque by separating these domains and, thus rotating the hinge at the middle section; the TCR is held from one end while it is pulled by its membrane-distal binding. This is shown in Fig. 73 (D) and Fig. 74.

5.3.12 Stretching the TCR rotates the inter-domain hinge and induces conformational changes of the C_β FG loop

In this section, controlled computational experiments are performed to understand how rotations of the inter-hinge domains impact or induce the opening and closing mechanism of the FG loop. Fig. 73 (D) shows the vectors defined to measure the angular rotation between the C_α-C_β respect V_α-V_β complex in front and top view, and they are situated in the center of mass of each TCR Ig-like domains. The rotation is visually depicted in Fig. 74. It is assumed that C_α-C_β and V_α-V_β complex independently move or rotate as a
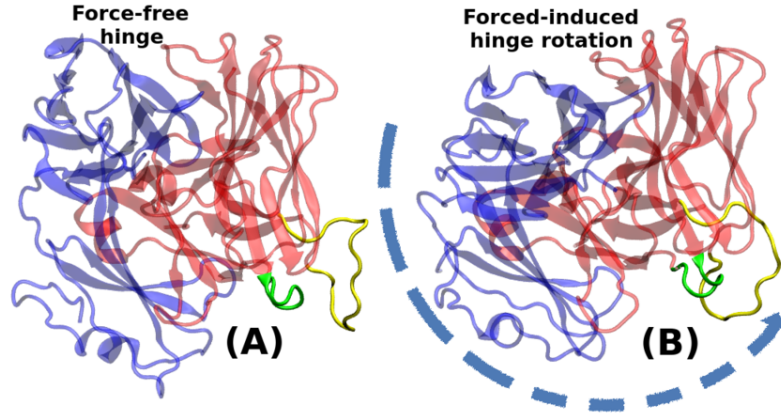
Figure 74: The TCR inter-domain rotational hinge. The V_α-V_β and C_α-C_β are structurally divided into two independent domains and they behave as an inter-domain rotational hinge. Pulling linearly the inter-hinge domains is translated into creating a torque that twists the TCR around the binding axis in a counter-clockwise manner. Linear forces also separate the TCR domains at the middle section. (A) The force-free state shows that the 3_{10} helix (green) is spatially away from the C_β FG loop (yellow). (B) The 3_{10} helix is above and close to the C_β FG loop when force induces the rotation of the inter-domain hinge. The cyan arrow depicts the direction of rotation that brings together the 3_{10} helix and the C_β FG loop.

whole unit. This is not a bad assumption since most structures of Ig-like domains are well-defined and stable. However, the exception is the C_α. This is an unusual domain that is not totally rigid. Here atoms strategically located in a β-strand of C_α are chosen. Those atoms are rigid even under load. See section 2.1 for properties of Ig-like domains.

5.3.12.1 The artificial-ligand protocol

The observation that potent ligands produce clearly the two forced-induced conformers further simplified our model. For the following simulation, an artificially potent ligand is proposed that does not allow dissociation. Before the complex was pulled from its ends and force was propagated to the binding site; this induces conformational changes in the epitope and glues the TCR with the MHC by producing and an "*in silico*" catch bond. The focus here is to study the TCR as a nanomachine; thus the artificial ligand will produce an infinite number of H-bonds and amalgamate the TCR to the binding interface, like a catch bond does. The process is accomplished by removing the pMHC and replacing it by two

dummy atoms connected to the center of mass of the TCR V_β and V_α domains. The dummy atoms function as a potent peptide ligand and their role is to attach the complex tightly and prolong the dissociation time and induce conformational changes in the presence of force as observed in previous simulations. This accomplishes three goals: 1) It reduces the system size and requires less computer power; 2) It removes the noise from the TCR interacting with the peptide ligand; 3) It move the complex to the catch-bond phase in a fast and controlled manner and the TCR stays at this state as long as the simulation runs. Note that this protocol is a variation of targeted molecular dynamics (TMD). TMD is used to induce conformational changes from an initial structure to an already known target conformation.

5.3.13 The role of the 3_{10} helix in changing the states of the C_β FG loop

Another important feature to consider is that all TCRs have in the β chain a small and unique 3_{10} helix that connects the V_β with C_β domains; this connector is also positioned near the vicinity of the C_β FG loop. The 3_{10} helix is very stable as it is held together by salt bridges formed between the Lys119$^\beta$ NH group and the Asp116$^\beta$ and Glu115$^\beta$ CO group. Fig. 75 (A) shows the molecular details of the 3_{10} helix and it is colored in green. Importantly the 3_{10} is observed to simultaneously unfolds only when the FG loop transition to the closed state in all potent ligands in the simulations. The unfolding process is visually depicted in Fig. 61 (C) and Fig. 75 (C). The unfolding occurs simultaneously when the FG loop is closed. In the neutral configuration, this connector is folded and formed a secondary structure of a 3_{10} helix (Fig. 61 A and Fig. 75 A), and the helix is commonly shared among structures of all TCRs known up to date. In the open conformation, this connector is displaying some level of linear stretching, Figure 61(B) Fig. 75 (B), while it is completely unfolded in the closed conformation, Figure 61(C) and Fig. 75 (C).
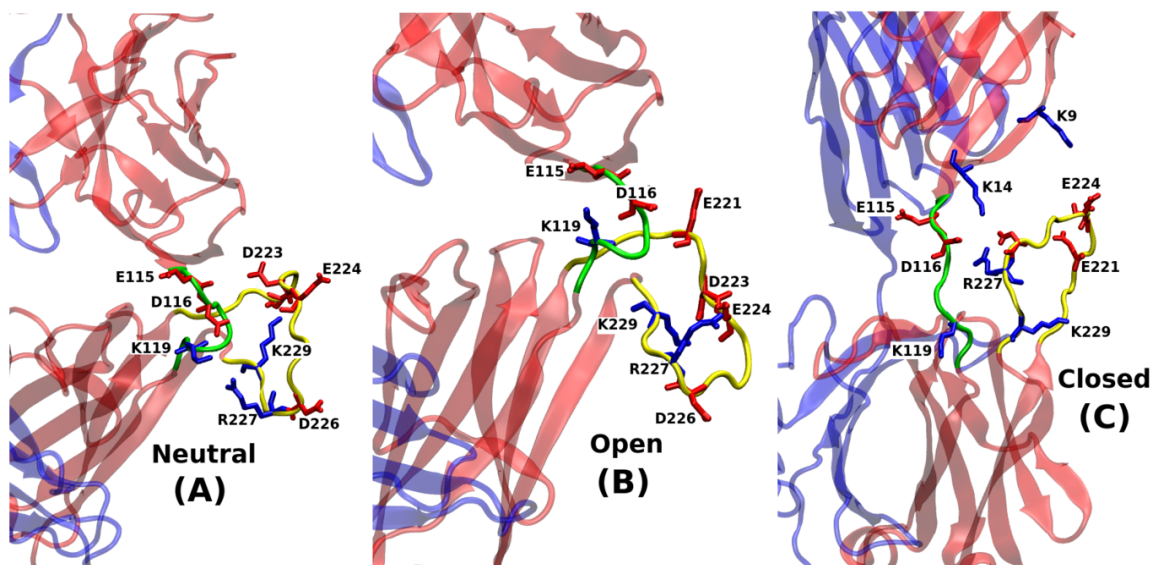
Figure 75: Molecular details of the 3_{10} helix and C_β FG loop. The 3_{10} helix and C_β FG loop are colored in green and yellow, respectively. (A) shows an FG loop in the neutral state; (B) show an open FG loop where force is built-up; (C) shows a closed FG loop. The interacting residues are drawn and labeled.

5.3.14 Hinge rotation and the open state occur simultaneously when linear force is translated into rotational torque that twists the TCR domains

Simulations are performed to validate if the artificial-ligand (AP) pulling produces the forced-induced conformers of the C_β FG loop. RA14 and 1E6 TCRs pulled at normal ramping rate using the AL pulling, 3 simulations are performed per TCR. It is observed that AL induces sequentially the open and closed states and the results are shown in Fig. 76. RA14 and 1E6 interacting with the AL peptide ligand are pulled using constant-velocity with a normal ramping rate. The pulling force is applied at 50 ns and starts growing from 0 pN (Shown in Figures). Before 50 ns all complexes display the neutral conformer. Open and closed states of the C_β FG loop are labeled and shown in black arrows. After a delay of 50 and 130 ns for RA14 and 1E6 TCRs, respectively, both complexes start experiencing conformational changes (deformations) as depicted: in the increase of the RMSD of their C_β FG loop (Fig. 76 A, D) and the 3_{10}-helix connector (Fig. 76 B, E) and the rotation of the TCR inter-domain hinge (Fig. 76 C, F).

5.3.15 The unfolding of the 3_{10}-helix connector allows the closing of the C_β FG loop

As force is applied, RA14 TCR starts slowly experiencing a linear increase in the RMSD corresponding to the deformation of its C_β FG loop and eventually leads to an open state, while it simultaneously rotates its inter-domain hinge angle. Pulling linearly the inter-hinge domains is translated into creating a torque that twists the TCR around the binding axis in a counter-clockwise manner (Fig. 74). The RMSD of their C_β FG loop experience four discrete deformation steps during the opening state; then it transitions to the neutral conformer as indicated by a suddenly decreased in the RMSD. Finally, it moves abruptly to the closed state and experiences irreversible deformations as the 3_{10}-helix connector completely unfolds. The transition between neutral to open states is mark by a linear increase in the inter-domain hinge.

As force is applied, RA14 TCR starts slowly experiencing a linear increase in the RMSD (Fig. 76 A), corresponding to the deformation of its C_β FG loop that eventually leads to its opening, while it simultaneously rotates the inter-domain hinge angle (Fig. 76 B). Pulling linearly the inter-hinge domains is translated into creating a torque that twists the TCR around the binding axis in a counter-clockwise manner. The RMSD of their C_β FG loop experience four small discrete deformation steps during the opening state; then it transitions to the neutral conformer as indicated by a sudden decreased in the RMSD. Finally, it moves abruptly to the closed state and experiences irreversible deformations and at the same time the 3_{10}-helix connector completely unfolds (Fig. 76 C). The transition between "neutral" to "open states" is marked by a linear increase in the inter-domain hinge while the transition from "open" to "close" is identified by an abruptly jump in the hinge angle (Fig. 76 B). The 3_{10}-helix only experiences major changes when the C_β FG loop moves from open to closed since it is abruptly unfolding at this state.

For 1E6 TCR, the C_β FG loop starts experiencing deformations later in the simulation that put it into a closed state in an stay there in a very short time ($\sim$20ns in Fig. 76 D). This is also depicted by the unfolding of the 3_{10}-helix connector (Fig. 76 E). After being closed, it
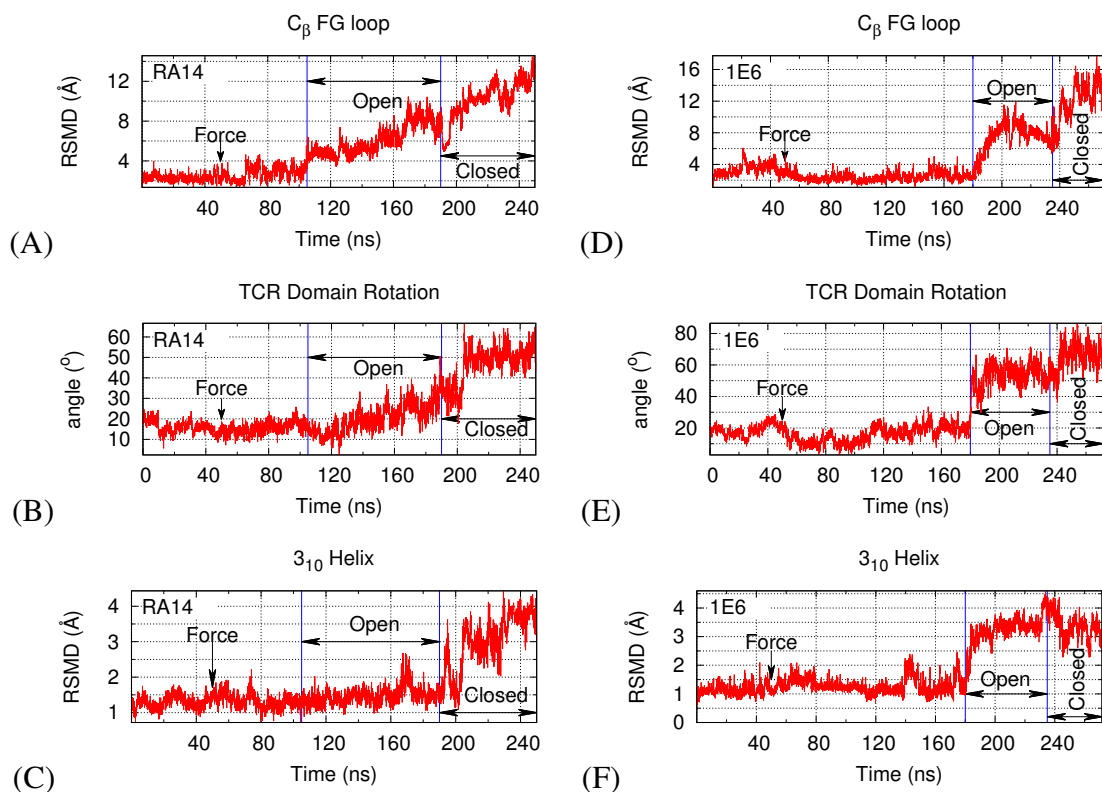
Figure 76: Characterization of the Opening and Closing states of the C_β FG loop in terms of its deformation (A, D), extension of the 3_{10}-helix connector (B, E), and changes of the inter-domain rotational hinge (C, F) for RA14 and 1E6 TCRs, respectively, using constant-velocity with a normal ramping rate. The pulling force is applied from 0 pN at 50 ns. Before applying the force, RA14 and 1E6 TCRs display the neutral conformer. Open and closed states of the C_β FG loop are labeled and shown in black arrows.

rapidly transitions to the open state by showing a decrease in the RMSD. Note that moving from closed to open state, it has to pass first through the neutral state and this is shown in the shortly-lived steepest valley at 200ns (and 190ns for RA14, Fig. 76 A). Finally, the load is huge that the C_β FG loop is forced to closed. Interesting the transitions between states are produced by the rotations of the inter-domain hinge angle. The unfolding process consists of two steps as shown in the RMSD of the 3_{10}-helix connector; however, the hinge angle is roughly constant in each unfolding step. It is shown that reaching the unfolding of the 3_{10}-helix connector is an irreversible process and after this step, only closed conformation is observed.

5.3.16 The hinge rotation brings together the 3_{10} helix and C_β FG loop and creates a huge electrostatic repulsion

Here the question is what opens and closes the C_β FG loop? The FG loop is highly charged having 4 negatively and 2 positively charged residues and it is able to form at most 8 salt-bridges. Salt-bridges are considered strong interaction since they are generated when two ionized residues form a hydrogen bond and an electrostatic interaction. For RA14 and 1E6 TCRs, their salt bridges become coordinated only when the C_β FG loop is undertaking an open or closed conformation as it is seen in Fig. 76 and 77. The Lys229 strongly interact with the negatively charged cluster to open the C_β FG loop, Fig. 77 (A and C). The electro-static force coordinating the strong slat-bridge is mainly between Lys229 and Glu219 and Asp 221. Importantly all species have the negatively charged cluster in a similar location and those three residues are highly conserved among them. The rotation of the inter-domain hinge brings together those residues in order to open the C_β FG loop and simultaneously the angle increases while their distance decreases, Fig. 76 and 77. Arg227 has similar be-havior; however it mostly interact with either Asp221 or Glu222, Fig. 77 (B and D). Note if the Arg227 is missing as in other species, it will impact in the duration of the open state; nevertheless, a shorter and less-pronounced opening should be still observed. In the closed conformation, the electrostatic interactions are pulled apart, Fig. 77. The molecular details of the interacting residues of the C_β FG loop and their location are shown in Fig. 75.

In summary, the rotation of the hinge causes the opening and closing of the C_β FG loop. In both TCRs, the transition from neutral to open is observed to be gradual and continuous while moving from open to close occurs abruptly and there is a jump in the value of the hinge angle and the RMSD of 3_{10}-helix connector and the C_β FG loop. Once the second unfolding occurs, it appears that the closed state becomes a punctuated equilibrium. The open state is a steady state at any specific level of force; if removed, it returns to the neutral state, or, if increased, it closes the C_β FG loop. Changing from open to close is presented in the RMSD by an abruptly drop and increase in the RMSD of the C_β FG loop. The neutral
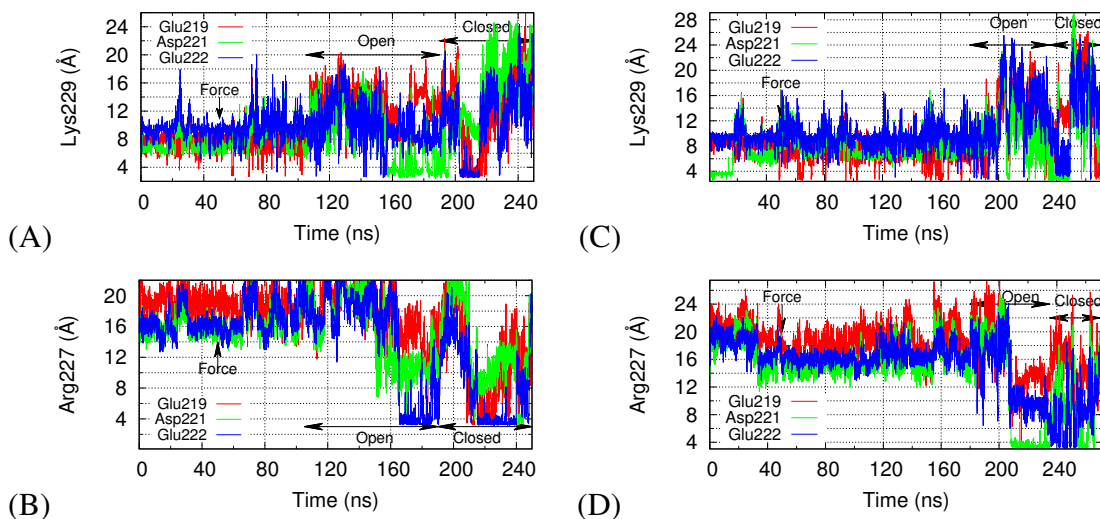
Figure 77: Salt-bridges formed by charged residues located in the C_β FG loop. The C_β FG loop has 4 negatively and 2 positively charged residues. The salt bridge becomes co-ordinated only when the C_β FG loop is undertaking an open or closed conformation. The Lys229 strongly interact with the negatively charged cluster to open the C_β FG loop. Open and closed states of the C_β FG loop are labeled and shown in black arrows.

conformer is unstable and short-lived in the presence of force. The inter-domain hinge rotation is responsible for the formation of open and closed states. Note that the jump in the RMSD of 3_{10}-helix connector (unfolding) corresponds to the subtle separation of the V_α-V_β from C_α-C_β interface. The transition from open to closed is fast and appears to be like releasing a compressed spring. Once the compressing force is removed the spring extends.

5.3.17 Performing single or double mutations on the C_β FG loop abolishes the open state

The next step is to test the model and determine if the opening or closing of the C_β FG loop can be manipulated by enhancing one phase while abolishing the other. Based on the previous understanding, it is predicted that pulling a high ramping rate the complex or performing a single or double *in silico* mutations on the C_β FG loop would abolish the open conformation. In addition, a closer inspection in the RA14 and 1E6 hinge angles, Fig. 76(D and E), reveal that when force is applied initially and it is in the neutral state,

the angles experience subtle swings (ups and downs) until enough force is built to start migrating continuously to the open, or drastically to the close conformation. Thus, computationally deleting the C_β FG loop would render the inter-domain hinge rotation smooth and monotonically increasing with little to no fluctuation.
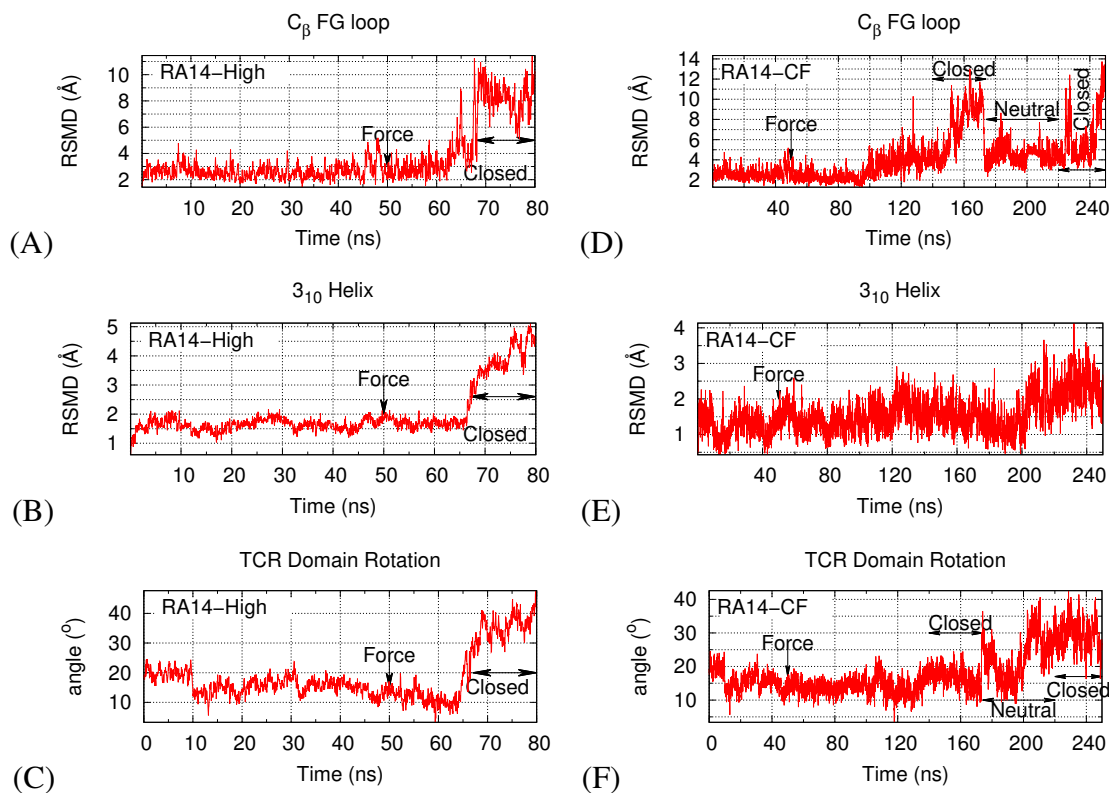
Figure 78: States of the C_β FG loop at high constant-velocity and constant-force pulling for RA14 TCR. The open state is diminished and becomes unstable. High pulling moves the C_β FG loop from the neutral to the closed conformer. The 3_{10}-helix connector is unfolded in this process and the hinge inter-domain angle displays an abruptly increased when the loop is closed.

Fig. 78 show indeed that pulling either at high constant-velocity or constant-force completely abolished the open state as it becomes unstable and short-lived. High ramping rates transition the loop almost directly from the neutral state to closed conformer; a similar response is observed at high constant-force pulling. Fig. 78 (A, D) show mostly loner-lived closed conformers, while the open ones become shorter-lived. At medium-to-high constant forces, the C_β FG loop mostly transitions from neutral to closed and vice versa before the

3_{10}-helix connector is unfolded, Fig. 78 (E). The hinge inter-domain angle also displays the abruptly increased when there is a transition from neutral to closed.

Fig. 79 show that performing a single and double *in silico* mutation on 1E6 TCR decreases and abolishes the open conformer. The single mutation disrupts the salt bridge and breaks the symmetry; thus the electrostatic interactions between the charges lose its coordination and become attractively weak, and the deformation is minimal in both the C_β FG loop and 3_{10}-helix connector, Fig. 79 (A, B). Instead of opening, there is a transition phase where the TCR attempts to deform the C_β FG loop; however, it fails to open it. The double mutation completely destroys the ability to move from neutral to the open conformer, Fig. 79 (D, E). Importantly the inter-domain rotational hinge is not affected by either mutation; instead, it continuously increases as it is allowed to move without any resistance nor repulsive electrostatic force. This is observed even before there is any sign of deforming in the loop, Fig. 79 (C, F). Unfolding of 3_{10}-helix is only observed once the C_β FG loop moves to the closed conformer, Fig. 79 (B, E).

5.3.18 Deleting the C_β FG loop allows the hinge angle to rotate and unfolds the 3_{10} helix

Fig. 80 (A, D) show that deleting the C_β FG loop computationally would permit the inter-domain hinge angle freedom to rotate around the binding axis. In addition 3_{10}-helix is easily extended in the presence of normal to huge load. The unfolding process of the 3_{10}-helix, previously involves its abrupt extension, but now it is replaced by a linear increase with little to no resistance, Fig. 80 (B, E). Note that before it was required 150 ns and 20 ns to induce any conformational changes in the 3_{10}-helix at medium and high constant-velocity pulling; however it now requires 60 ns and 5 ns at medium and high, respectively. The waiting time has dramatically reduced with the removal of the C_β FG loop. This means that the removal allows an easy transition and separation of the V_α-V_β from the C_α-C_β and their interface becomes weakly attached. Apparently, C_β FG loop provides a complementary structural support to maintain the 3D arrangement of V_β and C_β domains. In Fig. 80 (C) is
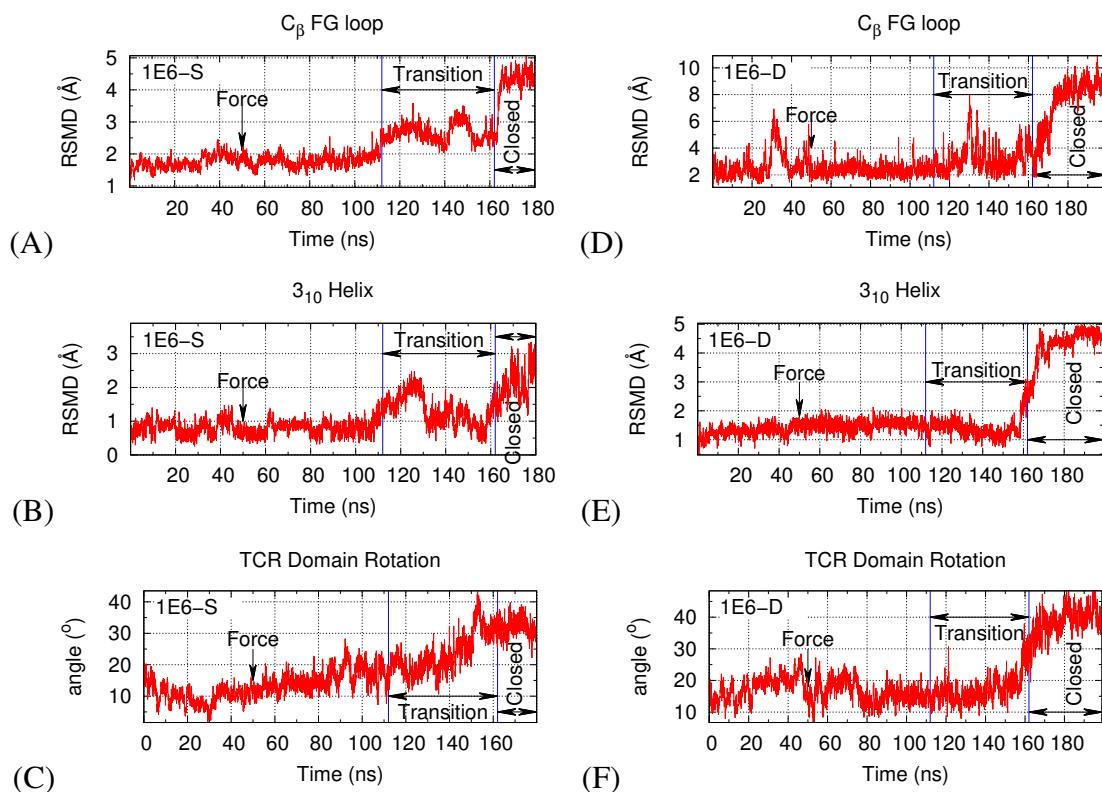
Figure 79: Single and double *in silico* mutations on 1E6 TCR. The single mutation disrupts the salt bridge and the electrostatic interactions are lost and their attractions become weak; the deformation is minimal in both the C_β FG loop and 3_{10}-helix connector. A double mutation abolishes completely the transition from neutral to open conformer. The inter-domain hinge is barely affected by either mutation; it increases and moves without any constraint or resistance. Unfolding of the 3_{10}-helix is only observed when the C_β FG loop is closed.

plotted the distance between V_β and C_β for a wild type and FG-loop deleted TCR. RA14 is used for comparing the two cases and both complexes are pulled at high ramping rates. It is observed that after 80 ns the FG-loop deleted RA14 TCR is already separated 15Åwhile the wild type is only displaced 5Å. This clearly shows the removal impacts the stability of the complex under load. A computationally FG-loop deleted TCR is depicted in the Fig. 80 (F); it shows the separation between domains.
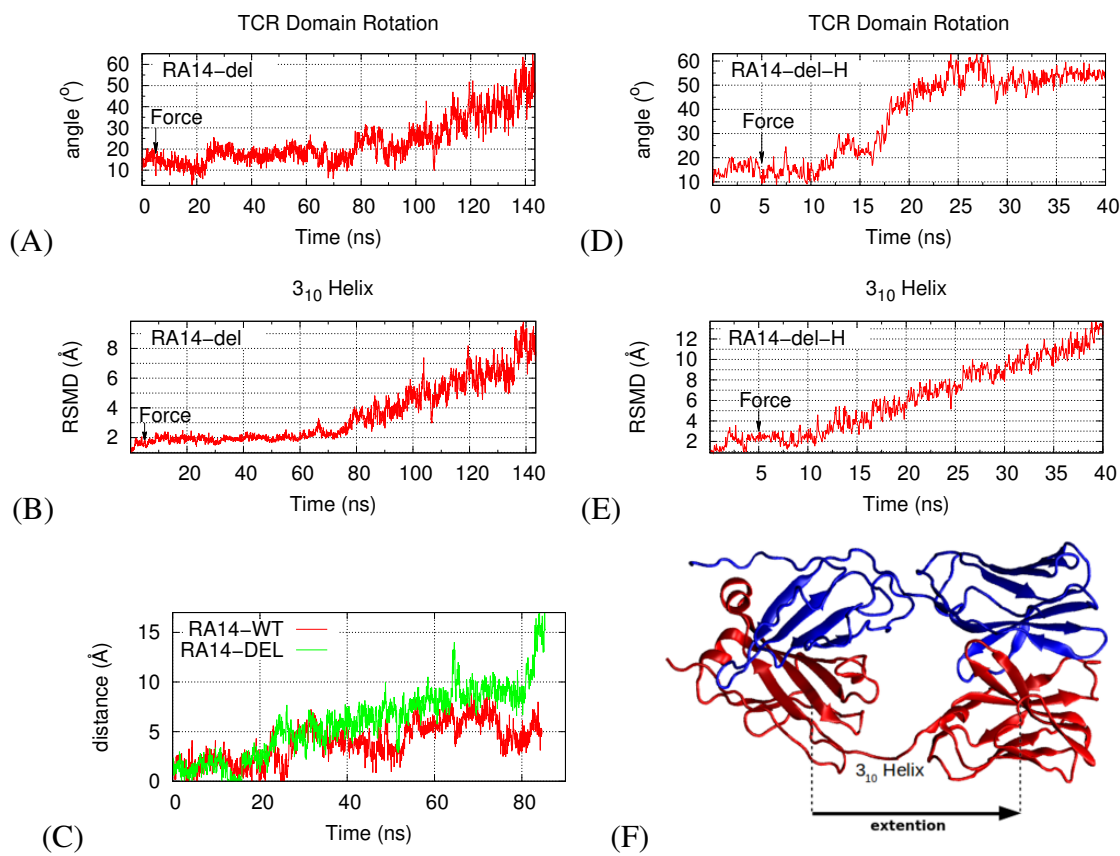
Figure 80: Characterization of conformational changes of C_β FG-loop deleted TCR for the RA14. FG-loop deleted TCRs allow the inter-domain hinge angle to rotate around the binding axis. 3_{10}-helix is easily extended once a small level of force is reached. The unfolding process of the 3_{10}-helix shows a linear increase in the RMSD. (F) shows an easily extended FG-loop deleted TCR.

5.3.19 Deformations that eventually open the C_β FG loop can be observed in long-time MD simulations.

It is also predicted that long-time MD simulations may capture deformations of the C_β FG loop that eventually cause the open of the loop. It is observed in the simulation that the domains, V_α-V_β and C_α-C_β, are not completely constrained, they are rotating respect to one another, and the rotation is mainly occurring the binding axis (the angular rotation is in the range of $\pm 15^o$. When the domains rotate in opposing directions as to increases the rotational angle, the change in the hinge angle is expected to induce deformations that potentially open of the C_β FG loop if the domains are moving at high velocity. In this

scenario, C_β FG loop should transition from neutral to open conformer and the duration should be proportional to the angular velocity. On the other hand, the closing of the C_β FG loop would require forces since it apparently has a bigger energy barrier since it also requires the unfolding of the 3_{10}-helix connector.
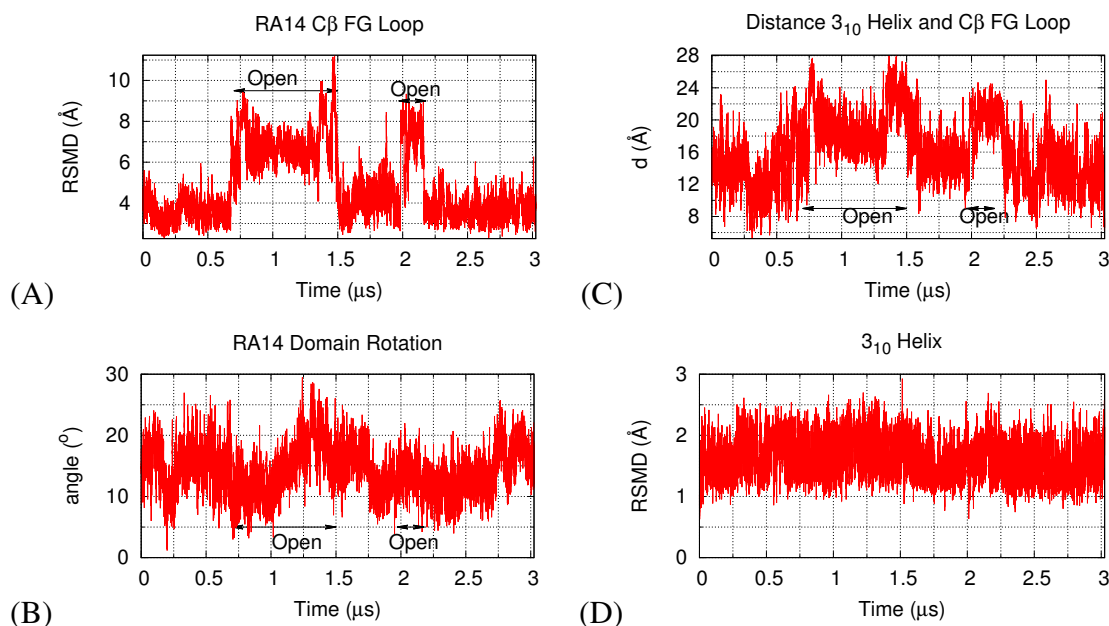
Figure 81: Characterization of the opening and closing states of the C_β FG loop in terms of its deformation (A), changes of the inter-domain rotational hinge (B), and the extension of the 3_{10}-helix connector (C, D) for RA14 , using microsecond MD simulations. The open states of the C_β FG loop are labeled and shown in black arrows.

Note that those would be rare events in free MD simulations. In order to access those rare conformers and cross the energy barrier between the states, an external force is applied to complex to reduce barriers and move the TCR across those configurations using steered MD simulation. The SMD simulations force the transition from an equilibrium conformation (neutral) to a rare event (steady state, open or closed). For instance, if the system is provided with enough time, the closed conformer or even the folding and unfolding process should be observed without the application of external forces.

Fig. 82 shows the potential conformational changes of TCRs when external forces are applied to the complex. The conformations occur sequentially and are characterized as the
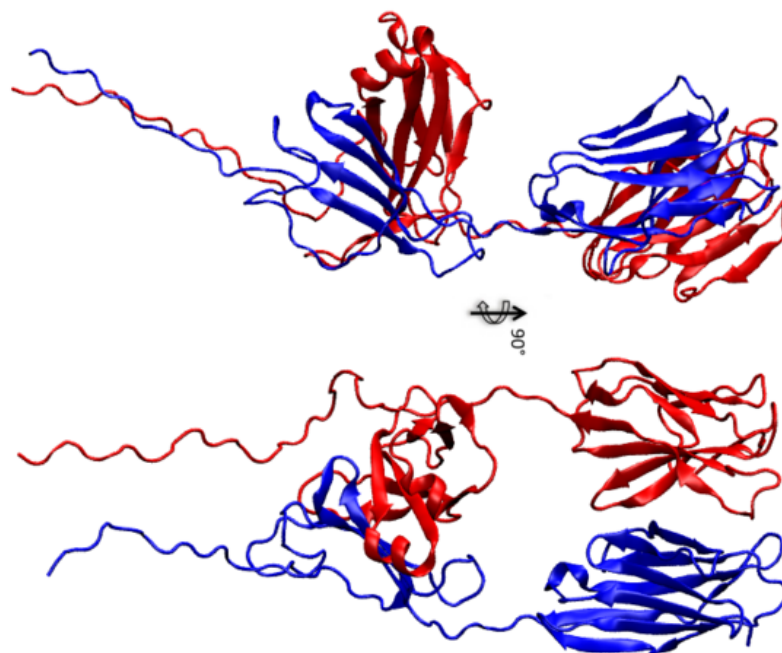
Figure 82: Conformational changes of TCRs upon recognition of peptide ligands. They fall into the following categories: the rotation of the inter-domain hinge, the opening and closing of the C_β FG loop, the unfolding of the 3_{10}-helix connector, the separation of TCR domains, and the unfolding of the TCR membrane-proximal stalk. The conformations impact the TCR structures and remodel the CD3 binding interface for activating the TCR and recognition of peptide ligands.

rotation of the inter-domain hinge, the opening and closing of the C_β FG loop, the unfolding of the 3_{10}-helix connector, the separation of TCR domains (V_α-V_β from C_α-C_β), and the unfolding of the TCR membrane-proximal stalk. The conformations not only impact the TCR structures but also remodel the CD3 binding interface for activating the TCR and recognition of peptide ligands.

5.3.20 The mechanism of how the opening and closing of the C_β FG loop induce T-cell activation by connecting with the CD3 signaling domains

Finally, the model is tested and verified using the super powerful **Anton2** machine–a special-purpose, massively-parallel supercomputer for performing microsecond MD simulations of proteins. The results support the model: 1) the inter-domain hinge is allowed to rotate and swing around 10° with an angular velocity of $\approx \pi/9 \, \mu s^{-1}$. 2) The deformation of

the C_β FG loop is observed and it is similar to the forced-induced one. 3) There is a symmetric up and down phase in the measured angle whenever hinge rotates. 3) The closed conformation is unstable in the absence of force. 4) Once the closed state is achieved, the complex experience globally irreversible deformation. 5) It is found that conserved residues in the TCR $_\alpha$ chain stabilize the closed conformation. 6) The TCR could experience large conformational changes that are potentially functional. In this part, the RA14 alone and complexed with pMHC is used, and the 1E6 (alone) in the different conformation of the C_β FG loop (open vs closed), are employed to run on Anton2 and the results support the proposed model.
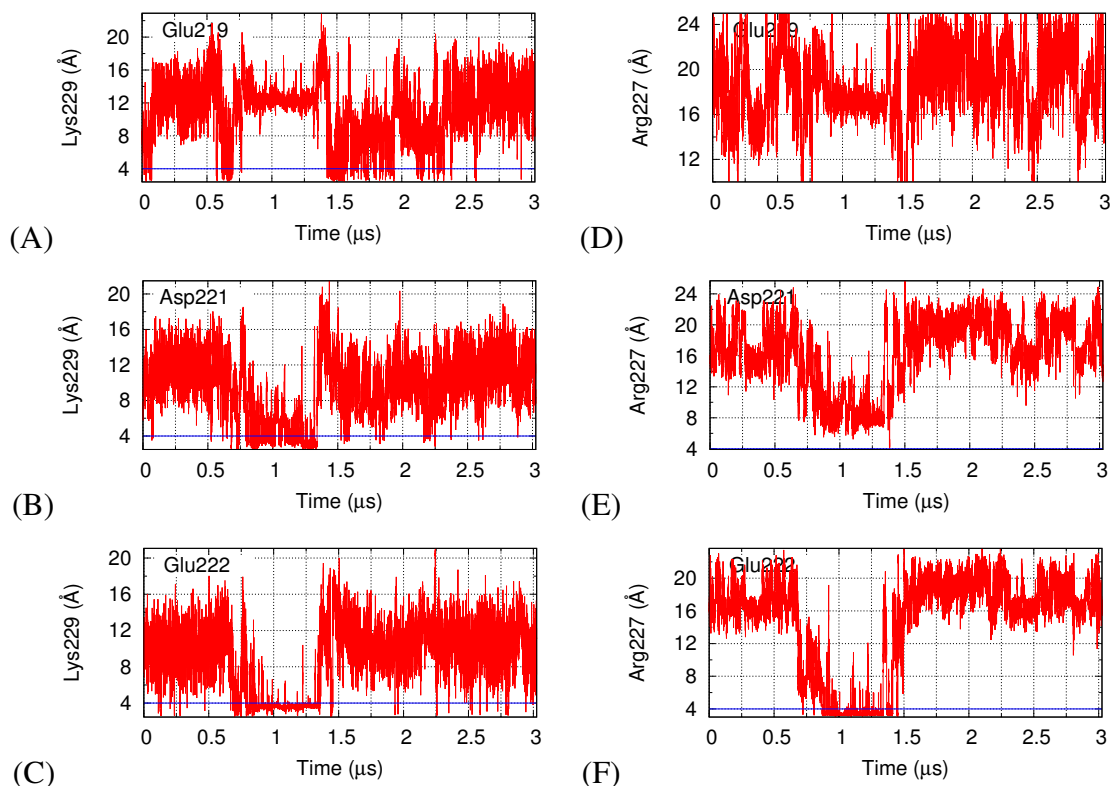
Figure 83: Salt-bridges formed by charged residues located in the C_β FG loop at the microsecond scale. The C_β FG loop has 4 negatively and 2 positively charged residues. The salt bridge becomes coordinated only when the C_β FG loop is undertaking an open conformation. The Lys229 strongly interact with the negatively charged cluster to open the C_β FG loop. Open states of the C_β FG loop is labeled and shown in black arrows.

143

5.3.21 What are the resting and activated states of the TCR?

The TCR-CD3 complex recognizes antigens and transmits signals into the T cell to initiate the adaptive immune response. By using computer simulations, in this section, a structural model is provided to unify conformational changes reported experimentally and it is proposed a connecting pathway between TCR and the CD3 subunits; and the molecular basis of TCR-CD3 interactions is examined.
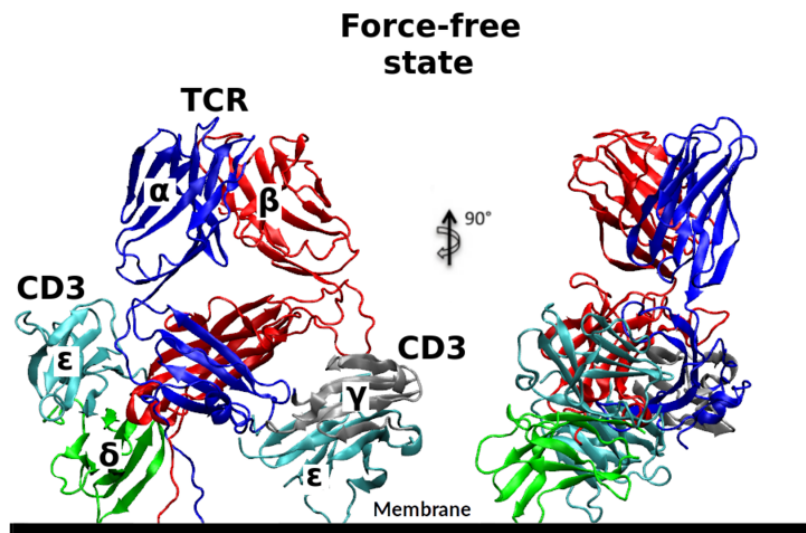
Figure 84: Molecular assembly of the TCR interacting with the CD3 signaling subunits at the resting state. CD3$\epsilon\delta$ and CD3$\epsilon\gamma$ could be located "below" TCRα and TCRβ, respectively, or they could be adjacent to the constant domain of the TCR. The model is created using the crystal structure of RA14 (PDB code 3GSN), CD3$\epsilon\gamma$ is made (PDB code 1SY6), and CD3$\epsilon\delta$ (PDB code 1XIW). The resting state is initially constructed by using the contacting residues between the TCR and the CD3 signaling subunits from reference [51].

The main conformational changes in this model for activating the TCR are: the rotation of the inter-domain hinge; the opening and closing of the C_β FG loop; the unfolding of the 3_{10}-helix connector; the separation of TCR domains (V_α-V_β from C_α-C_β); and the unfolding of the TCR membrane-proximal stalk, see Fig. 82 and Fig. 85. This model is consistent with all experimental observations reported up to now and it is explained in the following paragraph.

In the resting state of this model, the TCR is not required to be bound to the CD3 signaling subunits at the resting state. The CD3 subunits could be diffusing in the plasma membrane or they could be around the TCR when it encounters the peptide ligand. CD3$\epsilon\delta$ and CD3$\epsilon\gamma$ could be located "below" TCRα and TCRβ, respectively, or they could be adjacent to the constant domain of the TCR; thus, they do not have a specific binding site at the resting state (Fig. 84).

Upon the TCR ligation to the peptide ligand, a mechanical force is generated that pulls the TCR-pMHC complex. The pulling force induces the rotation of the TCR around the binding site and allows an easy scanning and attachment of the CD3 subunits to the TCR constant region. The rotation allows the interaction of the C$_\alpha$ AB and DE loop with the CD3$\epsilon\delta$ and the C$_\beta$ CC' loop with both CD3$\epsilon\gamma$ and CD3$\epsilon\delta$ and this movement firmly accommodates and places the CD3ϵ domains below the small cavity formed by the TCR constant domains.

The transition between the neutral and open state of the C$_\beta$ FG loop facilitates the contacting of the upper surface of the CD3$\epsilon\gamma$. At this point, the TCR behave as a **"bulldozer"** and the C$_\beta$ FG loop is the **"dozer blade"** that search for the CD3 subunits that are placed below. It is important that the C$_\beta$ FG is waved multiple times when the TCR is rotating; if not already in close contact with the signaling subunits; it would increase the probability of encountering and attaching to the CD3 subunits. Importantly the hinge rotation and the transition from neutral to open state occur simultaneously. This proposes that the CD3 subunits are reorientated and docked dynamically to the TCR upon antigen binding.

Once they are in good contact, and if force is propagating, it would eventually pull the CD3 subunits out of the plasma membrane. This is accomplished by unfolding the TCR connecting peptides and transmit a mechanical force across the membrane that finally dislodges the CD3 cytoplasmic tails from the plasma membrane and exposes their ITAM motifs for phosphorylation and activation. When the connecting peptides are extended, the C$_\beta$ helix-3 and helix 4-F strand and C$_\alpha$ F and C strands are also exposed and moved

aways from the plasma membrane and they are allowed to firmly interacts with the CD3$\epsilon\gamma$ and CD3$\epsilon\delta$, respectively. Then, the closing of the C$_\beta$ FG loop occurs when the 3$_{10}$-helix connector unfolds. At this step, V$_\alpha$-V$_\beta$ is slightly separated from C$_\alpha$-C$_\beta$ and the TCR is switched to the active state and the buried residues of the TCR α/β-interface are observed to undergo a systematic reorganization that produces significant conformational changes in these residues (Fig. 85).

Note that the hinge rotation brings together 3$_{10}$-helix connector and C$_\beta$ FG loop and it creates a huge electrostatic repulsion that opens the FG loop. Unfolding releases the electrostatic repulsion and put into contact opposed charged residues that pull up the C$_\beta$ FG loop and brings it to a closed state.
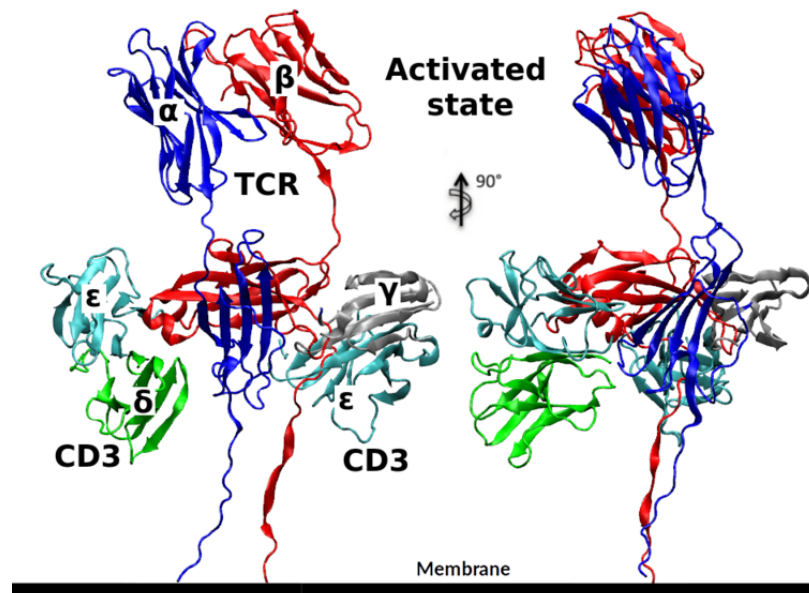
Figure 85: Molecular assembly of the TCR interacting with the CD3 signaling subunits at the activated state. The TCR behave as a "bulldozer" and the C$_\beta$ FG loop is the "dozer blade" that levels the CD3 subunits up. The C$_\beta$ FG is waved multiple times when the TCR is rotating so it attaches the CD3 subunits. Once they are in contact, and if force is continuing propagating, it would pull the CD3 subunits out of the plasma membrane. The model is created using the crystal structure of RA14 (PDB code 3GSN), CD3$\epsilon\gamma$ is made (PDB code 1SY6), and CD3$\epsilon\delta$ (PDB code 1XIW).

CHAPTER VI

CONCLUSIONS AND FUTURE DIRECTIONS

This thesis addresses various aspects of TCR mechanobiology by integrating computer simulation, biology, and engineering principles and the results are compared with experiments performed in the Zhu lab. The main goal of this thesis is directed to understand the molecular mechanisms of the TCR-pMHC-CD3 interactions and their underlying biological processes. The proposed methodology is studying this complex system by using the principles of statistical physics, which ensures that the models are theoretical sounded and it provides testable predictions that can be verified in experiments. The key benefit of this approach is that it enables the manipulation of complex systems at the molecular level and in a more logical way. This is the first step moving forward to the ultimate goal of immuno-engineering –a rational drug design, as it provides insights into the mechanisms of TCR function and generates new ideas for fighting infections, cancer, and autoimmune diseases using the body own's immune system.

This thesis proposes a mechanism to connect structurally the ligand binding, the catch bond, and the triggering of TCRs interacting with their peptide ligands and the CD3 signaling subunits. For instance, it describes a general mechanism used by TCR to recognize pMHC in the context of viral infections and how this information is propagated to the CD3 signaling domains. For understanding how T cells accomplish their function of transferring information from outside to inside the cells is very important to study the dynamics of their receptors and how they connect and interact with their ligands. Proteins are not static structures under physiological conditions as they dynamically change across multiple conformations. This is different from the current view of immune receptors that are believed to have unique docking interactions.

Aim 1 describes the mechanism how featureless viral peptide ligands are read by TCRs and explores many parameters that can be used to compare directly simulations with experiments. It is demonstrated that catch bonds formed experimentally between a TCR and a pMHC can be predicted by using simulations for the first time. The best predictor of catch bonds is the total number of H-bonds of TCR-pMHC interactions. This supports the idea that tensile force is a principal axis in regulating TCR function and induces the formation of *"in silico"* catch bonds. It is described the molecular characteristics and requirements for a TCR to form catch bonds in the context of antigen recognition and how mutant peptides convert it to slip bonds. It is then explained the mechanism of how the TCR changing its initial docking orientation modulates the peptide conformation inside the MHC binding groove. This highlights that the TCR docking angle is fundamental for immune recognition of antigenic peptides. At the catch-bond phase, TCRs start rotating with a defined polarity around the binding axis and the peptide has an on and off state for interacting with the TCR and forming catch bonds. It is proposed a molecular lever mechanism for recognition of viral epitopes. This aim could be also used to understand the mechanism of generating public vs private TCRs since most immunodominant viral epitopes resemble the NLV peptide (in terms of sidechain exposure) and elicit highly diverse T cell responses. A visual summary is presented in Fig 86 (A).

Aim 2 describes molecular mechanism of information is read and propagated to the CD3 signaling domains by the TCR. The first part is focused to investigate the structural role of catch bonds and how the information encoded in the peptide is decoded mechanically by the TCR. It is demonstrated that catch bonds are required to recognize epitopes and they are formed by the TCR rotating around the binding axes and inducing conformational changes in the peptide. The second part is focused to study how TCRs uses mechanical forces to determine whether or not the presented antigen is a threat. It is demonstrated that the peptide-decoding process and intracellular signaling are connected by conformational

148

changes traveling across the TCR. The molecular model shows that mechanical forces modulate the TCR states and that the TCRs are deformable proteins that can experience large conformational changes. It is shown that the TCR behaves as a "bulldozer" and the C_β FG loop is the "dozer blade" that levels the CD3 subunits up. The C_β FG is waved multiple times when the TCR is rotating so it attaches the CD3 subunits. Once they are in contact, and if force is continuing propagating, it would pull the CD3 subunits out of the plasma membrane as the TCR is undergoing large deformations. This subjects that proposed the ability of TCRs to deform without releasing the pMHC is fundamental to switch the TCR from the resting to activated states. A visual summary is presented in Fig 86 (B).

Overall, this thesis has laid the groundwork using computational methodologies to study how TCRs read viral and autoimmune peptide ligands and how they activate the signaling domains. However, this project can be expanded to study other TCRs as a unique framework is outlined to directly compare simulations with experiments. One intriguing and physiological relevant field is HIV-specific TCRs (infections) or cross-reactive TCRs for organ transplants. These would provide a better understanding into the molecular basis of the immune response leading to the loss of recognition of viral proteins or the strong attack of healthy cells, respectively. In addition, in Aim 2, the proposed mechanism of the opening and closing of the FG loop and the activated state of the TCR can be verified experimentally by doing site-directed mutagenesis since the molecular details are completely provided. This can be a project to determine if it is a functional mechanism or just an evolutionary relic.

Many problems in biology, in general, are just beginning to be thought in this framework, where molecular models with statistical-mechanical approaches are used to understand how proteins work. The major downside of this approach is that it is very computationally expensive and requires HPC supercomputer. It is time to simulate the TCR-pMHC complexed with co-receptors, the CD3 signaling domains including the cytoplasmic tails, and the plasma membrane. This new system would represent more physiologically the

149

TCR environment and it can be used to verify the proposed pathway of signal transduction between the ligand binding to the cytoplasmic tails.

The difficulty is that TCR simulations are still contained to the time scale and it would require at least hundreds of microseconds to observe conformational changes. A coarse-grain model can be built and used to track large propagations of structural changes through multiple domains. Importantly this should be performed in parallel with all-atom simulations to identify important interactions and assess their relevance in conformational changes and in the binding. The combination of these two types of simulations is a rising technique in MD simulations to study the structural changes in proteins.
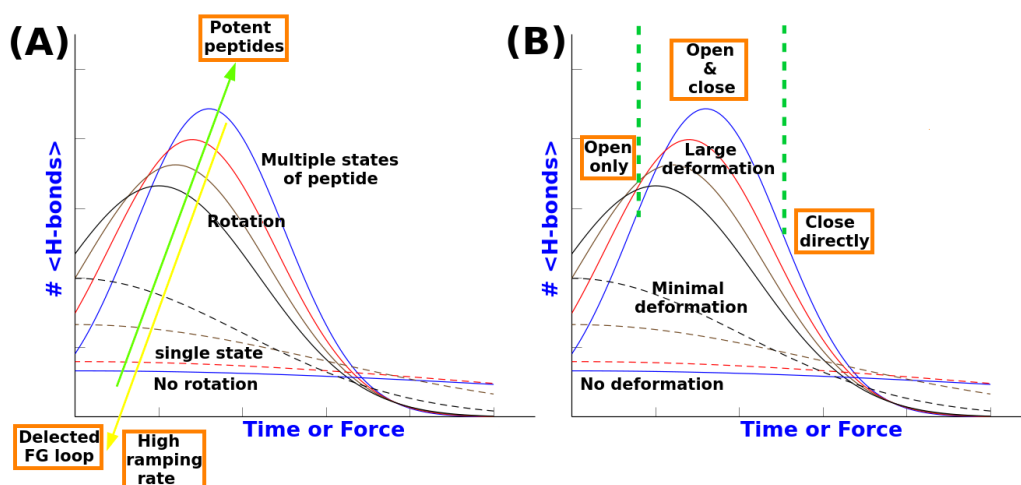
Figure 86: A visual summary of the results from Aim 1 and Aim 2. (A) The peptide potency increases the quality of the *"in silico"* catch bond while high ramping rates or pulling forces decrease it. The quality is defined as the duration and height of the catch bond. (B) Strong ligands open and close the C_β FG loop and induce large deformations in the TCR while weak ligands open the C_β FG loop and deform minimally the TCR. Very weak ligands do not deform the C_β FG loop and the TCR. Potent ligands produce catch bonds while weak ones convert them to slip or ideal bonds.

CHAPTER VII

LIST OF EXPECTED PUBLICATIONS

7.1   Aim 1:

Describe the molecular mechanism of how TCRs recognize featureless viral pMHC landscapes and how epitopes mutations decrease immune recognition in the context of human infection.

- *"Forced reorientation of CMV-specific TCR induces catch-bond formation with pMHC via a molecular lever mechanism"*; Paul Cardenas*, Prithi Jothikumar*, et. al. (Manuscript in preparation) *First co-author

7.2   Aim 2:

Describe a previously unknown mechanism of how recognition of peptides by TCRs propagates to the CD3 signaling domains to induce T-cell activation.

- *"Conformational changes in the TCR C_β FG loop: The mechanism of connecting ligand binding, catch bonds, and triggering of TCRs"*; Paul Cardenas and Cheng Zhu (Manuscript in preparation)

7.3   Other: Simulations of the PD-1 complexed with PD-1 ligands

- *"Structural mechanism of PD-1 catch-bonds and PD-Ligands"*; (Manuscript in preparation)

- *"The effect of stalk region in PD-1/PD-Ligands catch-bond and signalling"*; (Manuscript in preparation)

REFERENCES

[1] Abul Abbas. *Basic immunology : functions and disorders of the immune system.* Saunders/Elsevier, Philadelphia, Pa. London, 2011. 2

[2] Kenneth Murphy. *Janeway's immunobiology*. Garland Science, New York, 2012. 2, 2.1, 2

[3] Ed Palmer and Dieter Naeher. Affinity threshold for thymic selection through a t-cell receptor-co-receptor zipper. *Nature reviews. Immunology*, 9:207–213, March 2009. 2

[4] Alfred Singer, Stanley Adoro, and Jung-Hyun Park. Lineage fate and intense debate: myths, models and mechanisms of cd4- versus cd8-lineage choice. *Nature reviews. Immunology*, 8:788–801, October 2008. 2

[5] Mark M Davis, Michelle Krogsgaard, Morgan Huse, Johannes Huppa, Bjoern F Lillemeier, and Qi-jing Li. T cells as a self-referential, sensory organ. *Annu. Rev. Immunol.*, 25:681–695, 2007. 2

[6] Charles A Janeway Jr. The t cell receptor as a multicomponent signalling machine: Cd4/cd8 coreceptors and cd45 in t cell activation. *Annual review of immunology*, 10(1):645–674, 1992. 2, 2.10

[7] P Anton Van Der Merwe and Omer Dushek. Mechanisms for t cell receptor triggering. *Nature Reviews Immunology*, 11(1):47–55, 2011. 2, 2.3

[8] Abul Abbas. *Cellular and molecular immunology*. Elsevier/Saunders, Philadelphia, 2012. 2, 2.1

[9] Anne Davidson and Betty Diamond. Autoimmune diseases. *New England Journal of Medicine*, 345(5):340–350, 2001. 2

[10] Ian R. Mackay, Fred S. Rosen, Thomas Kamradt, and N. Avrion Mitchison. Tolerance and autoimmunity. *New England Journal of Medicine*, 344(9):655–664, mar 2001. 2

[11] Brian J Nankivell and Stephen I Alexander. Rejection of the kidney allograft. *The New England journal of medicine*, 363:1451–1462, October 2010. 2

[12] Michael S. Kuhns and Hemant B. Badgandi. Piecing together the family portrait of tcr-cd3 complexes. *Immunological Reviews*, 250(1):120–143, Oct 2012. 2.1

[13] Miguel Munoz-Ruiz, Julie C Ribot, Ana R Grosso, Natacha Goncalves-Sousa, Ana Pamplona, Daniel J Pennington, Jose R Regueiro, Edgar Fernandez-Malave, and Bruno Silva-Santos. Tcr signal strength controls thymic differentiation of discrete proinflammatory $\gamma\delta$ t cell subsets. *Nat Immunol*, 17(6):721–727, Apr 2016. 2.1

[14] Mark M. Davis and Pamela J. Bjorkman. T-cell antigen receptor genes and t-cell recognition. *Nature*, 334(6181):395–402, aug 1988. 2.1

[15] K. C. Garcia, M. Degano, R. L. Stanfield, A. Brunmark, M. R. Jackson, P. A. Peterson, L. Teyton, and I. A. Wilson. An alpha beta t cell receptor structure at 2.5 A and its orientation in the tcr-mhc complex. *Science*, 274(5285):209–219, Oct 1996. 2.1

[16] David N. Garboczi, Partho Ghosh, Ursula Utz, Qing R. Fan, William E. Biddison, and Don C. Wiley. Structure of the complex between human t-cell receptor, viral peptide and hla-a2. *Nature*, 384(6605):134–141, Nov 1996. 2.1

[17] J.-h. Wang. Atomic structure of an alpha beta t cell receptor (tcr) heterodimer in complex with an anti-tcr fab fragment derived from a mitogenic antibody. *The EMBO Journal*, 17(1):10–26, Jan 1998. 2.1

[18] A F Williams and A N Barclay. The immunoglobulin superfamily—domains for cell surface recognition. *Annual Review of Immunology*, 6(1):381–405, apr 1988. 2.1

[19] Lawrence J Stern and Don C Wiley. Antigenic peptide binding by class i and class ii histocompatibility proteins. *Structure*, 2(4):245–251, 1994. 2.2

[20] Markus G Rudolph, Robyn L Stanfield, and Ian A Wilson. How tcrs bind mhcs, peptides, and coreceptors. *Annu. Rev. Immunol.*, 24:419–466, 2006. 2.2

[21] Stéphanie Gras, Xavier Saulquin, Jean-Baptiste Reiser, Emilie Debeaupuis, Klara Echasserieau, Adrien Kissenpfennig, François Legoux, Anne Chouquet, Madalen Le Gorrec, Paul Machillot, et al. Structural bases for the affinity-driven selection of a public tcr against a dominant human cytomegalovirus epitope. *The Journal of Immunology*, 183(1):430–437, 2009. 1, 4, 2.8.1, 13, 4.2, 4.3.8, 1, 5.3

[22] Lawrence P Kane, Joseph Lin, and Arthur Weiss. Signal transduction by the tcr for antigen. *Current opinion in immunology*, 12(3):242–249, 2000. 2.3

[23] Arthur Weiss and Dan R Littman. Signal transduction by lymphocyte antigen receptors. *Cell*, 76(2):263–274, 1994. 2.3

[24] Matthew E Call and Kai W Wucherpfennig. Common themes in the assembly and architecture of activating immune receptors. *Nature Reviews Immunology*, 7(11):841–850, 2007. 2.3, 2

[25] Matthew E Call, Kai W Wucherpfennig, and James J Chou. The structural basis for intramembrane assembly of an activating immunoreceptor complex. *Nature immunology*, 11(11):1023–1029, 2010. 2.3

[26] Matthew E Call, Jason R Schnell, Chenqi Xu, Regina A Lutz, James J Chou, and Kai W Wucherpfennig. The structure of the $\zeta\zeta$ transmembrane dimer reveals features essential for its assembly with the t cell receptor. *Cell*, 127(2):355–368, 2006. 2.3

[27] Chenqi Xu, Etienne Gagnon, Matthew E Call, Jason R Schnell, Charles D Schwieters, Christopher V Carman, James J Chou, and Kai W Wucherpfennig. Regulation of t cell receptor activation by dynamic membrane binding of the cd3ϵ cytoplasmic tyrosine-based motif. *Cell*, 135(4):702–713, 2008. 2.3

[28] Clifford S Guy and Dario AA Vignali. Organization of proximal signal initiation at the tcr: Cd3 complex. *Immunological reviews*, 232(1):7–21, 2009. 2.3

[29] Paul E Love and Sandra M Hayes. Itam-mediated signaling by the t-cell antigen receptor. *Cold Spring Harbor perspectives in biology*, 2(6):a002485, 2010. 2.3

[30] Michael S Kuhns and Mark M Davis. The safety on the tcr trigger. *Cell*, 135(4):594–596, 2008. 2.3, 3

[31] Tetsuro Sasada, Maki Touma, Hsiu-Ching Chang, Linda K Clayton, Jia-huai Wang, and Ellis L Reinherz. Involvement of the tcr cβ fg loop in thymic selection and t cell function. *Journal of Experimental Medicine*, 195(11):1419–1431, 2002. 2.4

[32] Maki Touma, Hsiu-Ching Chang, Tetsuro Sasada, Maris Handley, Linda K Clayton, and Ellis L Reinherz. The tcr cβ fg loop regulates $\alpha\beta$ t cell development. *The Journal of Immunology*, 176(11):6812–6823, 2006. 2.4

[33] Jia-huai Wang and Ellis L Reinherz. The structural basis of $\alpha\beta$ t-lineage immune recognition: Tcr docking topologies, mechanotransduction, and co-receptor function. *Immunological reviews*, 250(1):102–119, 2012. 2.4, 2.6, 2.10

[34] Kristine N Brazin, Robert J Mallis, Dibyendu Kumar Das, Yinnian Feng, Wonmuk Hwang, Jia-huai Wang, Gerhard Wagner, Matthew J Lang, and Ellis L Reinherz. Structural features of the $\alpha\beta$tcr mechanotransduction apparatus that promote pmhc discrimination. *Frontiers in immunology*, 6, 2015. 2.4

[35] Masatsugu Oh-hora and Anjana Rao. Calcium signaling in lymphocytes. *Curr Opin Immunol.*, 20:250–258, 2008. 2.5

[36] Oreste Acuto, Vincenzo Di Bartolo, and Frederique Michel. Tailoring t-cell receptor signals by proximal negative feedback mechanisms. *Nature Reviews Immunology*, 8:699–712, 2008. 2.5

[37] Rolf Konig, Li-Yun Huang, and Ronald N. Germain. Mhc class ii interaction with cd4 mediated by a region analogous to the mhc class i binding site for cd8. *Nature*, 356:796–798, 1992. 2.5

[38] G Cammarota, A Scheirle, B Takacs, D M Doran, R Knorr, W Bannwarth, J Guardiola, and F Sinigaglia. Identification of a cd4 binding site on the beta 2 domain of hla-dr molecules. *Nature*, 356:799–801, April 1992. 2.5

[39] D. Gibbings and A. D. Befus. Cd4 and cd8: an inside-out coreceptor model for innate immune cells. *J Leukoc Biol.*, 86:251–259, 2009. 2.5

[40] Michel Mallaun, Dieter Naeher, Mark A Daniels, Pia P Yachi, Barbara Hausmann, Immanuel F Luescher, Nicholas RJ Gascoigne, and Ed Palmer. The t cell receptorâĂŹs α-chain connecting peptide motif promotes close approximation of the cd8 coreceptor allowing efficient signal initiation. *The Journal of Immunology*, 180(12):8211–8221, 2008. 2.5, 2.6

[41] Yiyuan Yin, Xin Xiang Wang, and Roy A Mariuzza. Crystal structure of a complete ternary complex of t-cell receptor, peptide-mhc, and cd4. *Proceedings of the National Academy of Sciences of the United States of America*, 109:5405–5410, April 2012. 5

[42] R. Wang, K. Natarajan, and D. H. Margulies. Structural basis of the cd8 mhc class i interaction: Focused recognition orients cd8 to a t cell proximal position. *Journal of Immunology*, 183:2554–2564, 2009. 5

[43] E. B. Day, C. Guillonneau, S. Gras, N. L. La Gruta, D. A. A. Vignali, P. C. Doherty, A. W. Purcell, J. Rossjohn, and S. J. Turner. Structural basis for enabling t-cell receptor diversity within biased virus-specific cd8+ t-cell responses. *PNAS*, 108:9536–9541, 2011. 5

[44] Sun Taek Kim, Koh Takeuchi, Zhen-Yu J Sun, Maki Touma, Carlos E Castro, Amr Fahmy, Matthew J Lang, Gerhard Wagner, and Ellis L Reinherz. The $\alpha\beta$ t cell receptor is an anisotropic mechanosensor. *Journal of Biological Chemistry*, 284(45):31028–31037, 2009. 2.6, 2.10

[45] Sun Taek Kim, Maki Touma, Koh Takeuchi, Zhen-Yu J Sun, Vibhuti P Dave, Dietmar J Kappes, Gerhard Wagner, and Ellis L Reinherz. Distinctive cd3 heterodimeric ectodomain topologies maximize antigen-triggered activation of $\alpha\beta$ t cell receptors. *The Journal of Immunology*, 185(5):2951–2959, 2010. 2.6

[46] Michael S Kuhns and Mark M Davis. Disruption of extracellular interactions impairs t cell receptor-cd3 complex stability and signaling. *Immunity*, 26(3):357–369, 2007. 2.6

[47] Yoseph Ghendler, Alex Smolyar, Hsiu-Ching Chang, and Ellis L Reinherz. One of the cd3ε subunits within a t cell receptor complex lies in close proximity to the cβ fg loop. *Journal of Experimental Medicine*, 187(9):1529–1536, 1998. 2.6

[48] Travis Beddoe, Zhenjun Chen, Craig S Clements, Lauren K Ely, Simon R Bushell, Julian P Vivian, Lars Kjer-Nielsen, Siew Siew Pang, Michelle A Dunstone, Yu Chih Liu, et al. Antigen ligation triggers a conformational change within the constant domain of the $\alpha\beta$ t cell receptor. *Immunity*, 30(6):777–788, 2009. 2.6

[49] Nuria Martínez-Martín, Ruth M Risueño, Antonio Morreale, Irene Zaldívar, Elena Fernández-Arenas, Fernando Herranz, Angel R Ortiz, and Balbino Alarcón. Cooperativity between t cell receptor complexes revealed by conformational mutants of cd3. *Sci. Signal*, 2:ra43, 2009. 2.6

[50] Yanan He, Sneha Rangarajan, Melissa Kerzic, Ming Luo, Yihong Chen, Qian Wang, Yiyuan Yin, Creg J Workman, Kate M Vignali, Dario AA Vignali, et al. Identification of the docking site for cd3 on the t cell receptor β chain by solution nmr. *Journal of Biological Chemistry*, 290(32):19796–19805, 2015. 2.6

[51] Aswin Natarajan, Vidushan Nadarajah, Klara Felsovalyi, Wenjuan Wang, Vivian R Jeyachandran, Riley A Wasson, Timothy Cardozo, Clay Bracken, and Michelle Krogsgaard. Structural model of the extracellular assembly of the tcr-cd3 complex. *Cell reports*, 14(12):2833–2845, 2016. 2.6, 84

[52] Michael E Birnbaum, Richard Berry, Yu-Shan Hsiao, Zhenjun Chen, Miguel A Shingu-Vazquez, Xiaoling Yu, Deepa Waghray, Suzanne Fischer, James McCluskey, Jamie Rossjohn, et al. Molecular architecture of the $\alpha\beta$ t cell receptor–cd3 complex. *Proceedings of the National Academy of Sciences*, 111(49):17576–17581, 2014. 2.6

[53] Marie-Agnès Doucey, Laurence Goffin, Dieter Naeher, Olivier Michielin, Petra Baumgärtner, Philippe Guillaume, Ed Palmer, and Immanuel F Luescher. Cd3δ establishes a functional link between the t cell receptor and cd8. *Journal of Biological Chemistry*, 278(5):3257–3264, 2003. 2.6

[54] Michael S Kuhns, Andrew T Girvin, Lawrence O Klein, Rebecca Chen, Kirk D C Jensen, Evan W Newell, Johannes B Huppa, BjÃűrn F Lillemeier, Morgan Huse, Yueh-Hsiu Chien, K Christopher Garcia, and Mark M Davis. Evidence for a functional sidedness to the alphabetatcr. *Proceedings of the National Academy of Sciences of the United States of America*, 107:5094–5099, March 2010. 2.6

[55] Michael S Kuhns and Hemant B Badgandi. Piecing together the family portrait of tcr-cd3 complexes. *Immunological reviews*, 250:120–143, November 2012. 2.6

[56] Ricardo A Fernandes, David A Shore, Mai T Vuong, Chao Yu, Xueyong Zhu, Selma Pereira-Lopes, Heather Brouwer, Janet A Fennelly, Claire M Jessup, Edward J Evans, Ian A Wilson, and Simon J Davis. T cell receptors are structures capable of initiating signaling in the absence of large conformational rearrangements. *The Journal of biological chemistry*, 287:13324–13335, April 2012. 2.6

[57] Zhen-Yu J Sun, Sun Taek Kim, Il Chul Kim, Amr Fahmy, Ellis L Reinherz, and Gerhard Wagner. Solution structure of the cd3epsilondelta ectodomain and comparison with cd3epsilongamma as a basis for modeling t cell receptor topology and signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 101:16867–16872, November 2004. 2.6

[58] Gijs I van Boxel, Samantha Holmes, Lars Fugger, and E Yvonne Jones. An alternative conformation of the t-cell receptor alpha constant region. *Journal of molecular biology*, 400:828–837, July 2010. 2.6

[59] William F Hawse, Matthew M Champion, Michelle V Joyce, Lance M Hellman, Moushumi Hossain, Veronica Ryan, Brian G Pierce, Zhiping Weng, and Brian M Baker. Cutting edge: Evidence for a dynamically driven t cell signaling mechanism. *Journal of immunology (Baltimore, Md. : 1950)*, 188:5819–5823, June 2012. 2.6

[60] Kannan Natarajan, Andrew McShan, Jiansheng Jiang, Vlad K Kumirov, Rui Wang, Huaying Zhao, Peter Schuck, Mulualem Tilahun, Lisa F Boyd, Jinfa Ying, Ad Bax, David Margulies, and Nikolaos Sgourakis. An allosteric site in the t-cell receptor cβ domain plays a critical signalling role. *Nature Communications*, 8(15260), 2017. 2.6

[61] Joseph A Trapani and Mark J Smyth. Functional significance of the perforin/granzyme cell death pathway. *Nature reviews. Immunology*, 2:735–747, October 2002. 2.7

[62] Alfons S K de Hooge, Dagmar Berghuis, Susy Justo Santos, Esther Mooiman, Salvatore Romeo, J Alain Kummer, R Maarten Egeler, Maarten J D van Tol, Cornelis J M Melief, Pancras C W Hogendoorn, and Arjan C Lankester. Expression of cellular flice inhibitory protein, caspase-8, and protease inhibitor-9 in ewing sarcoma and implications for susceptibility to cytotoxic pathways. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 13:206–214, January 2007. 2.7

[63] Pierre Tattevin, Yves Le Tulzo, Sophie Minjolle, Arnaud Person, Jean Marc Chapplain, Cedric Arvieux, Remi Thomas, and Christian Michelet. Increasing incidence of severe epstein-barr virus-related infectious mononucleosis: surveillance study. *Journal of clinical microbiology*, 44:1873–1874, May 2006. 2.7

[64] Sheetal Manicklal, Vincent C Emery, Tiziana Lazzarotto, Suresh B Boppana, and Ravindra K Gupta. The "silent" global burden of congenital cytomegalovirus. *Clinical microbiology reviews*, 26:86–102, January 2013. 2.7

[65] Babs E Verstrepen, Andre Boonstra, and Gerrit Koopman. Immune mechanisms of vaccine induced protection against chronic hepatitis c virus infection in chimpanzees. *World journal of hepatology*, 7:53–69, January 2015. 2.7

[66] Rebecca R Terilli and Andrea L Cox. Immunity and hepatitis c: a review. *Current HIV/AIDS reports*, 10:51–58, March 2013. 2.7

[67] Helene D Gayle and Gena L Hill. Global impact of human immunodeficiency virus and aids. *Clinical microbiology reviews*, 14(2):327–335, 2001. 2.7

[68] P Ghazal, JW Yewdell, and A Alcami. Viruses in control of the immune system workshop on molecular mechanisms of immune modulation: lessons from viruses. *EMBO reports-European Molecular Biology Organization*, 3(10):27–32, 2002. 2.7

[69] E John Wherry, Joseph N Blattman, Kaja Murali-Krishna, Robbert Van Der Most, and Rafi Ahmed. Viral persistence alters cd8 t-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. *Journal of virology*, 77(8):4911–4927, 2003. 2.7

[70] John S Yi, Maureen A Cox, and Allan J Zajac. T-cell exhaustion: characteristics, causes and conversion. *Immunology*, 129(4):474–481, 2010. 2.7

[71] Sine Reker Hadrup and Ton N Schumacher. Mhc-based detection of antigen-specific cd8+ t cell responses. *Cancer Immunology, Immunotherapy*, 59(9):1425–1433, 2010. 2.7

[72] A Plauzolles, M Lucas, and S Gaudieri. Hepatitis c virus adaptation to t-cell immune pressure. *The Scientific World Journal*, 2013, 2013. 2.7

[73] Kelly P Burke and Andrea L Cox. Hepatitis c virus evasion of adaptive immune responses: a model for viral persistence. *Immunologic research*, 47(1-3):216–227, 2010. 2.7

[74] EGM Berkhoff, MM Geelhoed-Mieras, EJ Verschuren, CA Van Baalen, RA Gruters, RAM Fouchier, ADME Osterhaus, and GF Rimmelzwaan. The loss of immunodominant epitopes affects interferon-γ production and lytic activity of the human influenza virus-specific cytotoxic t lymphocyte response in vitro. *Clinical & Experimental Immunology*, 148(2):296–306, 2007. 2.7

[75] Akiko Iwasaki. Antiviral immune responses in the genital tract: clues for vaccines. *Nature reviews. Immunology*, 10(10):699, 2010. 2.7

[76] X Zhou, S Ramachandran, M Mann, and DL Popkin. Role of lymphocytic choriomeningitis virus (lcmv) in understanding viral immunology: past, present and future. viruses 4: 2650–2669. *d x. doi. org/10*, 3390:v4112650, 2012. 2.7

[77] Robert F Kalejta. Tegument proteins of human cytomegalovirus. *Microbiology and molecular biology reviews*, 72(2):249–265, 2008. 2.8

[78] Craig R Roy. Immunology: professional secrets. *Nature*, 425(6956):351–352, 2003. 2.8, 6

[79] John Sinclair and Patrick Sissons. Latency and reactivation of human cytomegalovirus. *The Journal of general virology*, 87:1763–1779, July 2006. 2.8

[80] Beverly L. Davidson and Xandra O. Breakefield. Neurological diseases: Viral vectors for gene delivery to the nervous system. *Nature Reviews. Neuroscience*, 4:353–364, 2003. 2.8

[81] Maria Jamela Revilleza, Rui Wang, Janet Mans, Manqing Hong, Kannan Natarajan, and David H Margulies. How the virus outsmarts the host: function and structure of cytomegalovirus mhc-i-like molecules in the evasion of natural killer cell surveillance. *Journal of biomedicine & biotechnology*, 2011:724607, 2011. 2.8

[82] Thomas R Jones and Shi-Wu Lee. An acidic cluster of human cytomegalovirus ul99 tegument protein is required for trafficking and function. *Journal of virology*, 78:1488–1502, February 2004. 2.8

[83] John Paul Tomtishen. Human cytomegalovirus tegument proteins (pp65, pp71, pp150, pp28). *Virology journal*, 9:22, January 2012. 2.8

[84] G V Quinnan, W H Burns, N Kirmani, A H Rook, J Manischewitz, L Jackson, G W Santos, and R Saral. Hla-restricted cytotoxic t lymphocytes are an early immune response and important defense mechanism in cytomegalovirus infections. *Reviews of infectious diseases*, 6:156–163, 1984. 2.8

[85] Xinbo Yang, Mingming Gao, Guobing Chen, Brian G Pierce, Jinghua Lu, Nan-ping Weng, and Roy A Mariuzza. Structural basis for clonal diversity of the public t cell response to a dominant human cytomegalovirus epitope. *Journal of Biological Chemistry*, 290(48):29106–29119, 2015. 2.8.1, 4.2, 1, 5.3

[86] Stephen J Turner, Peter C Doherty, James McCluskey, and Jamie Rossjohn. Structural determinants of t-cell receptor bias in immunity. *Nature Reviews Immunology*, 6(12):883–894, 2006. 2.8.1, 4.2, 12

[87] Hanjie Li, Congting Ye, Guoli Ji, and Jiahuai Han. Determinants of public t cell responses. *Cell research*, 22(1):33, 2012. 2.8.1

[88] Klaus Dornmair, Norbert Goebels, Hans-Ulrich Weltzien, Hartmut Wekerle, and Reinhard Hohlfeld. T-cell-mediated autoimmunity: novel techniques to characterize autoreactive t-cell receptors. *The American journal of pathology*, 163:1215–1226, October 2003. 2.9

[89] Jun Huang, Veronika I Zarnitsyna, Baoyu Liu, Lindsay J Edwards, Ning Jiang, Brian D Evavold, and Cheng Zhu. The kinetics of two-dimensional tcr and pmhc interactions determine t-cell responsiveness. *Nature*, 464(7290):932–936, 2010. 2.9, 2.10

[90] Baoyu Liu, Shi Zhong, Karolina Malecek, Laura A Johnson, Steven A Rosenberg, Cheng Zhu, and Michelle Krogsgaard. 2d tcr-pmhc-cd8 kinetics determines t-cell responses in a self-antigen-specific tcr system. *European journal of immunology*, 44:239–250, January 2014. 2.9, 4.2

[91] Cheng Zhu, Ning Jiang, Jun Huang, Veronika I Zarnitsyna, and Brian D Evavold. Insights from in situ analysis of tcr-pmhc recognition: response of an interaction network. *Immunological reviews*, 251:49–64, January 2013. 2.9

[92] Baoyu Liu, Wei Chen, Brian D Evavold, and Cheng Zhu. Accumulation of dynamic catch bonds between tcr and agonist peptide-mhc triggers t cell signaling. *Cell*, 157:357–368, April 2014. 2.9, 4.2, 4.2

[93] Dhruv K Sethi, David A Schubert, Anne-Kathrin Anders, Annie Heroux, Daniel A Bonsor, Chantz P Thomas, Eric J Sundberg, Jason Pyrdol, and Kai W Wucherpfennig. A highly tilted binding mode by a self-reactive t cell receptor results in altered engagement of peptide and mhc. *The Journal of experimental medicine*, 208:91–102, January 2011. 2.9

[94] Yiyuan Yin, Yili Li, Melissa C Kerzic, Roland Martin, and Roy A Mariuzza. Structure of a tcr with high affinity for self-antigen reveals basis for escape from negative selection. *The EMBO journal*, 30:1137–1148, March 2011. 2.9

[95] Michael Hahn, Melissa J Nicholson, Jason Pyrdol, and Kai W Wucherpfennig. Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune t cell receptor. *Nature immunology*, 6:490–496, May 2005. 2.9

[96] Yili Li, Yuping Huang, Jessica Lue, Jacqueline A Quandt, Roland Martin, and Roy A Mariuzza. Structure of a human autoimmune tcr bound to a myelin basic protein self-peptide and a multiple sclerosis-associated mhc class ii molecule. *The EMBO journal*, 24:2968–2979, September 2005. 2.9

[97] Anna M Bulek, David K Cole, Ania Skowera, Garry Dolton, Stephanie Gras, Florian Madura, Anna Fuller, John J Miles, Emma Gostick, David A Price, Jan W Drijfhout, Robin R Knight, Guo C Huang, Nikolai Lissin, Peter E Molloy, Linda Wooldridge, Bent K Jakobsen, Jamie Rossjohn, Mark Peakman, Pierre J Rizkallah, and Andrew K Sewell. Structural basis for the killing of human beta cells by cd8(+) t cells in type 1 diabetes. *Nature immunology*, 13:283–289, January 2012. 2.9, 1, 5.3

[98] David K Cole, Anna M Bulek, Garry Dolton, Andrea J Schauenberg, Barbara Szomolay, William Rittase, Andrew Trimby, Prithiviraj Jothikumar, Anna Fuller, Ania Skowera, et al. Hotspot autoimmune t cell receptor binding underlies pathogen and insulin peptide cross-reactivity. *The Journal of clinical investigation*, 126(6):2191, 2016. 2.9, 1, 5.3

[99] Henk-Jan Aanstoot, Barbara J Anderson, Denis Daneman, Thomas Danne, Kim Donaghue, Francine Kaufman, Rosangela R RÃl'a, and Yasuko Uchigata. The global burden of youth diabetes: perspectives and potential. *Pediatric diabetes*, 8 Suppl 8:1–44, October 2007. 2.9

[100] Jeffrey A Bluestone, Kevan Herold, and George Eisenbarth. Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature*, 464:1293–1300, April 2010. 2.9

[101] Antonios Chatzigeorgiou, Vaggelis Harokopos, Christina Mylona-Karagianni, Emmanouil Tsouvalas, Vassilis Aidinis, and Elli F Kamper. The pattern of inflammatory/anti-inflammatory cytokines and chemokines in type 1 diabetic patients over time. *Annals of medicine*, 42:426–438, September 2010. 2.9

[102] Yuri Sykulev, Michael Joo, Irina Vturina, Theodore J Tsomides, and Herman N Eisen. Evidence that a single peptide–mhc complex on a target cell can elicit a cytolytic t cell response. *Immunity*, 4(6):565–571, 1996. 2.10

[103] Darrell J Irvine, Marco A Purbhoo, Michelle Krogsgaard, and Mark M Davis. Direct observation of ligand recognition by t cells. *Nature*, 419(6909):845–849, 2002. 2.10

[104] Marco A Purbhoo, Darrell J Irvine, Johannes B Huppa, and Mark M Davis. T cell killing does not require the formation of a stable mature immunological synapse. *Nature immunology*, 5(5):524–530, 2004. 2.10

[105] Timothy K Starr, Stephen C Jameson, and Kristin A Hogquist. Positive and negative selection of t cells. *Annual review of immunology*, 21(1):139–176, 2003. 2.10

[106] Michelle Krogsgaard, Qi-jing Li, Cenk Sumen, Johannes B Huppa, Morgan Huse, and Mark M Davis. Agonist/endogenous peptide–mhc heterodimers drive t cell activation and sensitivity. *Nature*, 434(7030):238–243, 2005. 2.10

[107] Michelle Krogsgaard, Jeremy Juang, and Mark M Davis. A role for "self" in t-cell activation. In *Seminars in immunology*, volume 19, pages 236–244. Elsevier, 2007. 2.10

[108] Nadia Anikeeva, Tatiana Lebedeva, Aaron R Clapp, Ellen R Goldman, Michael L Dustin, Hedi Mattoussi, and Yuri Sykulev. Quantum dot/peptide-mhc biosensors reveal strong cd8-dependent cooperation between self and viral antigens that augment the t cell response. *Proceedings of the National Academy of Sciences*, 103(45):16846–16851, 2006. 2.10

[109] Ya-Chen Li, Bing-Mae Chen, Pei-Chun Wu, Tian-Lu Cheng, Lung-Sen Kao, Mi-Hua Tao, Andre Lieber, and Steve R Roffler. Cutting edge: mechanical forces acting on t cells immobilized via the tcr complex can trigger tcr signaling. *The Journal of Immunology*, 184(11):5959–5963, 2010. 2.10

[110] Zhengyu Ma, Dennis E Discher, and Terri H Finkel. Mechanical force in t cell receptor signal initiation. *Frontiers in immunology*, 3:217, 2012. 2.10

[111] Salvatore Valitutti, Mark Dessing, Klaus Aktories, Harald Gallati, and Antonio Lanzavecchia. Sustained signaling leading to t cell activation results from prolonged t cell receptor occupancy. role of t cell actin cytoskeleton. *Journal of Experimental Medicine*, 181(2):577–584, 1995. 2.10

[112] Bryan T Marshall, Mian Long, James W Piper, Tadayuki Yago, et al. Direct observation of catch bonds involving cell-adhesion molecules. *Nature*, 423(6936):190, 2003. 2.11

[113] Sabyasachi Rakshit, Yunxiang Zhang, Kristine Manibog, Omer Shafraz, and Sanjeevi Sivasankar. Ideal, catch, and slip bonds in cadherin adhesion. *Proceedings of the National Academy of Sciences*, 109(46):18815–18820, 2012. 2.11

[114] Brian Bracegirdle. Microscopy and comprehension: the development of understanding of the nature of the cell. *Trends Biochem. Sci.*, 14:464–468, 1989. 3.1

[115] Peter F. Lindley. The use of synchrotron radiation in protein crystallography. *Radiat. Phys. Chem.*, 45:367–383, 1995. 3.1

[116] D E BRADLEY. A study of the negative staining process. *Journal of general microbiology*, 29:503–516, November 1962. 3.1

[117] Ilya Gertsman, Elizabeth A Komives, and John E Johnson. Hk97 maturation studied by crystallography and h/2h exchange reveals the structural basis for exothermic particle transitions. *Journal of molecular biology*, 397:560–574, March 2010. 3.1

[118] Qing Xie, Michael Spilman, Nancy L Meyer, Thomas F Lerch, Scott M Stagg, and Michael S Chapman. Electron microscopy analysis of a disaccharide analog complex reveals receptor interactions of adeno-associated virus. *Journal of structural biology*, 184:129–135, November 2013. 3.1

[119] Mourad Sadqi, David Fushman, and Victor Munoz. Atom-by-atom analysis of global downhill protein folding. *Nature*, 442:317–321, 2006. 3.1

[120] Ulam S. Fermi E., Pasta J. Studies of nonlinear problems, i - doe-osti. *Los Alamos report LA-1940 (1955)*, 1955. 3.2

[121] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, 31(2):459–466, 1959. 3.2

[122] A Rahman. Correlations in the motion of atoms in liquid argon. *Physical Review*, 136(2A):A405, 1964. 3.2

[123] John Maddox. Statistical mechanics by numbers. *Nature*, 334:561–561, 1988. 3.2

[124] Tamar Schlick. Pursuing laplace's vision on modern computers. In *Mathematical Approaches to Biomolecular Structure and Dynamics*, pages 219–247. Springer, 1996. 3.2

[125] Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical review. A, General physics*, 31:1695–1697, March 1985. 3.2, 3.2

[126] Hoover WG. *Computational Statistical Mechanics.* Elsevier Science, 1991. 3.2, 3.2

[127] Hoover. Constant-pressure equations of motion. *Physical review. A, General physics*, 34:2499–2500, September 1986. 3.2

[128] Scott E. Feller, Yuhong Zhang, Richard W. Pastor, and Bernard R. Brooks. Constant pressure molecular dynamics simulation: The langevin piston method. *J Chem Phys.*, 103:4613–4621, 1995. 3.2

[129] D Quigley and M I J Probert. Langevin dynamics in constant pressure extended systems. *The Journal of chemical physics*, 120:11432–11441, June 2004. 3.2

[130] R.W Hockney and J.W Eastwood. *Computer Simulation Using Particles.* CRC Press, 1989. 3.3.1.2

[131] Shneior Lifson and Arieh Warshel. Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules. *The Journal of Chemical Physics*, 49(11):5116–5129, 1968. 3.3.2.3

[132] Hui Lu, Barry Isralewitz, Andre Krammer, Viola Vogel, and Klaus Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical journal*, 75(2):662–671, 1998. 3.3.4

[133] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26:1781–1802, December 2005. 3.4

[134] David E Shaw, JP Grossman, Joseph A Bank, Brannon Batson, J Adam Butts, Jack C Chao, Martin M Deneroff, Ron O Dror, Amos Even, Christopher H Fenton, et al. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 41–53. IEEE Press, 2014. 3.6

[135] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996. 3.7

[136] John Stone. *An Efficient Library for Parallel Ray Tracing and Animation.* Master's thesis, Computer Science Department, University of Missouri-Rolla, April 1998. 3.7

[137] D. Frishman and P. Argos. Knowledge-based secondary structure assignment. *Proteins: structure, function and genetics*, 23:566–579, 1995. 3.7

[138] Felix Rico, Laura Gonzalez, Ignacio Casuso, Manel Puig-Vidal, and Simon Scheuring. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. *Science*, 342(6159):741–743, 2013. 3.8

[139] Janet L Maryanski, Jean-Laurent Casanova, Kirsten Falk, Hélène Gournier, Christian Jaulin, Philippe Kourilsky, François A Lemonnier, Roland Lüthy, Hans-Georg Rammensee, Olaf Rötzschke, et al. The diversity of antigen-specific tcr repertoires reflects the relative complexity of epitopes recognized. *Human immunology*, 54(2):117–128, 1997. 4.2

[140] Stephen J Turner, Katherine Kedzierska, Helen Komodromou, Nicole L La Gruta, Michelle A Dunstone, Andrew I Webb, Richard Webby, Helen Walden, Wiedong Xie, James McCluskey, et al. Lack of prominent peptide-major histocompatibility complex features limits repertoire diversity in virus-specific cd8+ t cell populations. *Nature immunology*, 6(4):382, 2005. 4.2

[141] Rob Meijers, Char-Chang Lai, Yuting Yang, Jin-huan Liu, Weimin Zhong, Jia-huai Wang, and Ellis L Reinherz. Crystal structures of murine mhc class i h-2 d b and k b molecules in complex with ctl epitopes from influenza a virus: implications for tcr repertoire selection and immunodominance. *Journal of molecular biology*, 345(5):1099–1110, 2005. 4.2

[142] Lars Kjer-Nielsen, Craig S Clements, Anthony W Purcell, Andrew G Brooks, James C Whisstock, Scott R Burrows, James McCluskey, and Jamie Rossjohn. A structural basis for the selection of dominant $\alpha\beta$ t cell receptors in antiviral immunity. *Immunity*, 18(1):53–64, 2003. 4.2

[143] Jeffrey Ishizuka, Guillaume BE Stewart-Jones, Anton van der Merwe, John I Bell, Andrew J McMichael, and E Yvonne Jones. The structural dynamics and energetics of an immunodominant t cell receptor are programmed by its vβ domain. *Immunity*, 28(2):171–182, 2008. 4.2

[144] Chihiro Motozono, Nozomi Kuse, Xiaoming Sun, Pierre J Rizkallah, Anna Fuller, Shinichi Oka, David K Cole, Andrew K Sewell, and Masafumi Takiguchi. Molecular basis of a dominant t cell response to an hiv reverse transcriptase 8-mer epitope presented by the protective allele hla-b* 51: 01. *The Journal of Immunology*, 192(7):3428–3434, 2014. 4.2

[145] John A Zaia, Ghislaine Gallez-Hawkins, Xiuli Li, Zhi-Qiang Yao, Norma Lomeli, Karen Molinder, Corinna La Rosa, and Don J Diamond. Infrequent occurrence of natural mutations in the pp65495–503 epitope sequence presented by the hla-a* 0201 allele among human cytomegalovirus isolates. *Journal of virology*, 75(5):2472–2474, 2001. 4.2, 4.3.12

[146] Ozlem Keskin, Attila Gursoy, Buyong Ma, and Ruth Nussinov. Principles of protein-protein interactions: What are the preferred ways for proteins to interact? *Chemical reviews*, 108(4):1225–1244, 2008. 4.3.1

[147] Zhenhua Li, Limsoon Wong, and Jinyan Li. Dbac: A simple prediction method for protein binding hot spots based on burial levels and deeply buried atomic contacts. *BMC systems biology*, 5(1):S5, 2011. 4.3.1

[148] Ozlem Keskin, Buyong Ma, and Ruth Nussinov. Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. *Journal of molecular biology*, 345(5):1281–1294, 2005. 4.3.1

[149] Ariel Erijman, Eran Rosenthal, and Julia M Shifman. How structure defines affinity in protein-protein interactions. *PLOS one*, 9(10):e110085, 2014. 4.3.1

[150] Alexander Fuhrmann, Sebastian Getfert, Qiang Fu, Peter Reimann, Stuart Lindsay, and Robert Ros. Long lifetime of hydrogen-bonded dna basepairs by force spectroscopy. *Biophysical journal*, 102(10):2381–2390, 2012. 4.3.1

[151] Kristine Manibog, Hui Li, Sabyasachi Rakshit, and Sanjeevi Sivasankar. Resolving the molecular mechanism of cadherin catch bond formation. *Nature communications*, 5:3941, 2014. 4.3.1

[152] Xiaoming Sun, Yi Shi, Tomohiro Akahoshi, Mamoru Fujiwara, Hiroyuki Gatanaga, Christian Schönbach, Nozomi Kuse, Victor Appay, George F Gao, Shinichi Oka, et al. Effects of a single escape mutation on t cell and hiv-1 co-adaptation. *Cell reports*, 15(10):2279–2291, 2016. 4.3.4, 4.3.5, 4.3.12

[153] Phillip Pymm, Patricia T Illing, Sri H Ramarathinam, Geraldine M O'Connor, Victoria A Hughes, Corinne Hitchen, David A Price, Bosco K Ho, Daniel W McVicar, Andrew G Brooks, et al. Mhc-i peptides get out of the groove and enable a novel mechanism of hiv-1 escape. *Nature Structural & Molecular Biology*, 24(4):387–394, 2017. 4.3.4, 4.3.5, 4.3.12

[154] Dean R Madden, Joan C Gorga, Jack L Strominger, and Don C Wiley. The three-dimensional structure of hla-b27 at 2.1 å resolution suggests a general mechanism for tight peptide binding to mhc. *Cell*, 70(6):1035–1048, 1992. 4.3.10

[155] Massimo Degano, K Christopher Garcia, Vasso Apostolopoulos, Markus G Rudolph, Luc Teyton, and Ian A Wilson. A functional hot spot for antigen recognition in a superagonist tcr/mhc complex. *Immunity*, 12(3):251–261, 2000. 1, 5.3

[156] Jeffrey A Speir, K Christopher Garcia, Anders Brunmark, Massimo Degano, Per A Peterson, Luc Teyton, and Ian A Wilson. Structural basis of 2c tcr allorecognition of h-2l d peptide complexes. *Immunity*, 8(5):553–562, 1998. 1

[157] Tatyana Krivobokova, Rodolfo Briones, Jochen S Hub, Axel Munk, and Bert L de Groot. Partial least-squares functional mode analysis: application to the membrane proteins aqp1, aqy1, and clc-ec1. *Biophysical journal*, 103(4):786–796, 2012. 5.3.8.1

[158] Tong Seng Lim, Alessandra Mortellaro, Chwee Teck Lim, Günter J Hämmerling, and Paola Ricciardi-Castagnoli. Mechanical interactions between dendritic cells and t cells correlate with t cell responsiveness. *The Journal of Immunology*, 187(1):258–265, 2011. 5.3.9