Turun yliopisto
University of Turku

Immunological interactions of virus peptides at the antigen presenting MHC I proteins

Azam Rashidian

Master's Thesis

Master's Degree Programme in Digital Health and Biosciences

Department of Future Technologies

University of Turku

2018

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

UNIVERSITY OF TURKU
Department of Future Technologies/Faculty of Science and Engineering

RASHIDIAN, AZAM: Immunological interactions of virus peptides at the antigen presenting MHC I proteins

Master's Thesis, 71 p (+ 17 appendix pages)
Bioinformatics
November 2018

Regarding the recent outbreak of Zika virus (ZIKV) infection, there is an urgent need to develop a preventive or a therapeutic ZIKV vaccine. In this thesis, computational analysis was performed to predict suitable peptide candidates for vaccine design.

Computational approaches such as docking and molecular dynamics simulations (MD simulations) were employed to evaluate the binding energy and stability of candidate T-cell epitope peptides of ZIKV proteins at the antigen-presenting MHC class I molecules. For the docking step and the following MD simulations, MHC I alleles HLA-A*0101, HLA-A*0201, HLA-B*2705 and HLA-C*0801 were used as receptor structures and eight different peptides from ZIKV proteins (E, NS3, NS5) were docked to the MHC I molecules.

All predicted peptide-HLA complexes and experimental reference peptide-HLA complexes were submitted to a 10-ns MD simulation in explicit water for further refinement and to examine and compare their stability. Hydrogen bonding network, Root-Mean-Square Deviation (RMSD) for both the MHC I peptide-binding domain and the peptides, atomic fluctuation and solvent accessibility of the bound peptides, interaction energies and the dimensions of the peptide binding groove were analyzed to evaluate the stability and strength of the peptide-HLA complexes.

The computational analysis provided two T-cell epitopes from the ZIKV proteins (GLDFSDLYY, FSDLYYLTM) with a high affinity to the studied MHC I alleles. These could be introduced as putative candidates for vaccine development.

Keywords: Antigenic epitope, Binding energy, HLA, MHC I molecule, Molecular dynamics simulation, Peptide docking, T-cell, Zika infection

Abbreviations

3D	Three-dimentional
AMBER	Assisted Model Building with Energy Refinement
APC	Antigen Presenting Cell
BLAST	Basic Local Alignment Search Tool
CD	Cluster of Differentiation
CD4+ cell	T Helper Cell
CD8+ cell	Cytotoxic T-Cell
GROMACS	GROningen MAchine for Chemical Simulations
HLA	Human Leukocyte Antigen
MD	Molecular Dynamics
MHC I	Major Histocompatibility Complex Class I
MSA	Multiple Sequence Alignment
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
NAMD	Nanoscale Molecular Dynamics
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
Prime-MMGBSA	Prime-Molecular Mechanics Generalized Born Surface Area
RMSD	Root-Mean-Square Deviation
RMSF	Root-Mean-Square Fluctuation
RSA	Relative Solvent Accessibility
SARS	Severe Acute Respiratory Syndrome
SASA	Solvent Accessible Surface Area
TCR	T-cell receptor
ZIKV	Zika Virus

TABLE OF CONTENTS

1. Introduction 1
2. Literature review 2
2.1 MHC I molecules 2
2.2 Vaccine development 6
2.3 General Principles in Rational Drug Design 7
3. Aims 8
4. Materials and Methods 9
4.1 Ligands: Selection of potential antigenic peptides from ZIKV proteins 9
4.2 Targets: Selection of MHC I molecules 11
4.3 Amino Acid Sequence Retrieval 11
4.4 Multiple Sequence Alignment 13
4.5 Docking: Building the peptide-MHC I complexes 13
4.6 Molecular Dynamics Simulations 14
4.7 Analysis of the stability of the peptide-MHC I interactions 17
5. Results and Discussion 19
5.1 Multiple Sequence Alignment 20
5.2 Evaluation of the docking results in the light of the MD refinement 24
5.2.1 Peptides from the ZIKV E protein 24
5.2.2 Peptides from the ZIKV NS3 protein 28
5.2.3 Peptides from the ZIKV NS5 protein 31
5.3 Evaluation of the molecular dynamics simulation trajectories 34
5.3.1 HLA-A*0101 in complex with ZIKV E protein peptides 34
5.3.2 HLA-A*0201 in complex with ZIKV E protein peptides 37
5.3.3 HLA-B*2705 in complex with ZIKV NS3 protein peptides 52
5.3.4 HLA-C*0801 in complex with ZIKV NS5 protein peptides 55
6. Conclusion 61
7. References 64
Appendices 72

Acknowledgments

I would like to express my gratitude to my supervisors Professor Outi Salo-Ahen at Åbo Academi University who patiently and eagerly assisted me during this project and Dr. Martti Tolvanen at University of Turku for giving me the opportunity to explore the Bioinformatics world and providing me with guidance. I thank Dr. Jukka Lehtonen for his assistance to computational challenges I was facing during the project. I would also like to thank our programme coordinator Eija Nordlund, and Juho Heimonen Project researcher and Coordinator in Bioinformatics department for all their support.

I wish to acknowledge Professor Mark Johnson for the excellent computing facilities at the Åbo Akademi University and CSC – IT Center for Science, Finland, for computational resources. The use of Biocenter Finland infrastructure at Åbo Akademi (bioinformatics) is also acknowledged.

1. Introduction

Viruses are highly dependent upon their target host cell for replication. This means that, based on the type of virus, DNA or RNA, they assemble their genome in the host cell nucleus or cytoplasm, respectively.

Zika virus (ZIKV) belongs to the Flaviviridae family that has a single-stranded RNA genome. Its genome encodes a polyprotein comprised of seven non-structural, namely, NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5 and three structural proteins: capsid (C), premembrane/membrane protein (PrM/M) and envelope (E) protein (Lindenbach & Rice, 2003; Mirza et al., 2016). The virus was isolated in 1940 from a Ugandan forest where its name comes from (Dick et al., 1952). Monkeys were the primary vertebrate host for the virus (Fauci & Morens, 2016).

There are several ways of virus transmission such as Mosquito bites, sexual contacts, pregnancy and blood transfusion (Mansuy et al., 2016). During pregnancy, the virus can transmit transplacentally to the fetus and lead to several adverse outcomes such as congenital microcephaly, malformation and miscarriage (Brasil et al., 2016; Malkar et al., 2016), as well as ocular lesion in surviving infants (de Paula Freitas et al., 2016; Ventura et al., 2016).

There was an epidemic outbreak caused by ZIKV during 2007-2016 across the Pacific Ocean. The adverse complexities of the disease, especially the teratogenic lesions, demand urgent action to develop a vaccine against the infection (Oehler et al., 2014).

It has been shown that the identification of antigenic epitopes recognized by T-cells and B-cells is the first and most significant step to design a vaccine since a component of immune system, such as T-cell as a mediator, could be stimulated via such epitopes (Backert & Kohlbacher, 2015). Identifying these epitopes has been possible using immunoinformatics methods (Florea et al., 2003). It has also been suggested that the degree of elicited T-cell response is highly correlated to the strength of binding between the epitope and the MHC molecule (Lazarski et al., 2005; Liu et al., 2010). This study employed peptide docking and molecular dynamics (MD) simulation methods to select potential peptide epitopes as candidates for ZIKV vaccine development.

2. Literature review

2.1 MHC I molecules

Proteins are substantial macromolecules that play vital roles in living systems and many biological processes. They are polymers consisting of monomers called amino acids. Based on the characteristics of amino acids, a protein has a broad range of functional properties which allow it to fold and adapt different conformations from linear or primary structure to quaternary structure (National Research Council, 1989). Regarding the important role of proteins in living systems, it is of great importance to access the three-dimensional (3D) structure of proteins. There are two commonly used experimental methods that are capable in solving high-resolution 3D structures of proteins: X-ray crystallography and NMR spectroscopy.

The proteins that are studied in this project are called major histocompatibility complex (MHC) molecules. Many MHC structures have been determined with X-ray crystallography. Due to the major effects that these molecules have on the histocompatibility, they were called major histocompatibility molecules.

The MHC genes are divided into three different classes according to their locations, namely classes I, II, and III (Abbas et al., 2014). MHC class I (MHC I) genes are present on chromosome 6 in humans and they are known as Human Leukocyte Antigens (HLAs). They are coded at three different locations (loci) on the genome named A, B and C. MHC molecules are found on all the nucleated cells in the body, and they play an important role in the development of both humoral and cellular immunity. T-cells (cellular immunity) recognize antigens only as peptides that are bound to a MHC I molecule. The main function of MHC I molecules is to bring antigens to the cell surface for recognition by a T-cell. An antigenic peptide binds to the MHC molecule with high affinity to make a stable complex.

Of the three existing types of human MHC molecules, we focus on the MHC class I molecules. MHC I molecules are cell surface glycoproteins and key actors in the process of adaptive immunity. They are on the surface of antigen-presenting cells (APCs). They bind to peptides ranging from eight to ten residues and are recognized by CD8+ cytotoxic T lymphocytes. The antigenic peptides

presented by MHC I can be from pathogens such as viruses that are to be destructed by CD8+ T-cells. The MHC I pathway is shown in Figure 1.

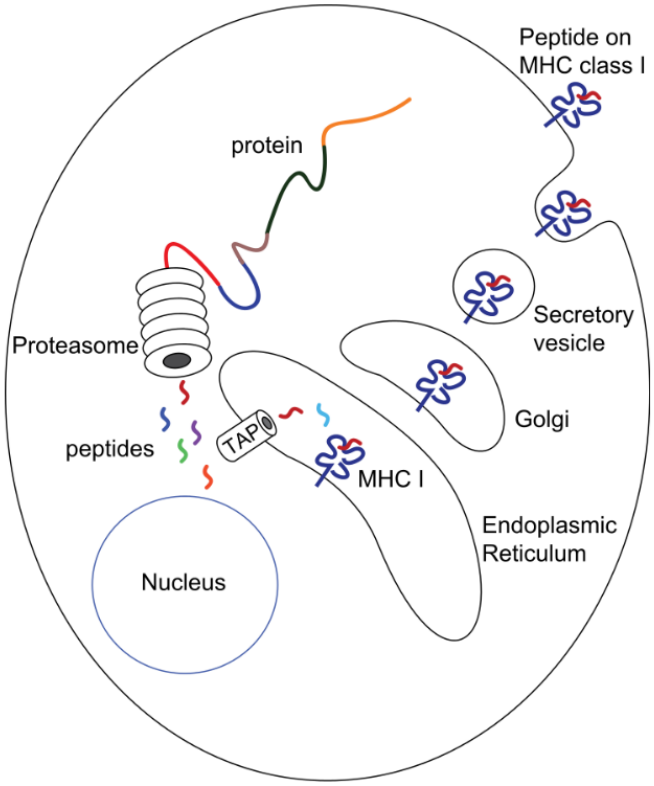


Figure 1. Cytosolic pathway of MHC I. Source: Scray, MHC Class I processing, CC BY-SA 3.0.

MHC I molecules are heterodimeric transmembrane proteins that consist of a single α chain that is coupled with the β2 microglobulin protein (Figure 2A). The α chain has been divided into three different parts, namely α1, α2, and α3. The α1 and α2 segments build the wall of the peptide-binding groove while eight antiparallel β-stranded sheets create the floor of the groove. This structure is known as α1α2 domain (Figure 2B). α3 domain is located in the membrane and it is the region where MHC I interacts with the CD8+ molecule (Toh et al., 2000). The α chain (270 amino acid residues) is polymorphic and encoded by HLA genes, whereas β2 microglobulin protein (100 amino acid residues) is not polymorphic and is encoded by the β2 microglobulin gene.

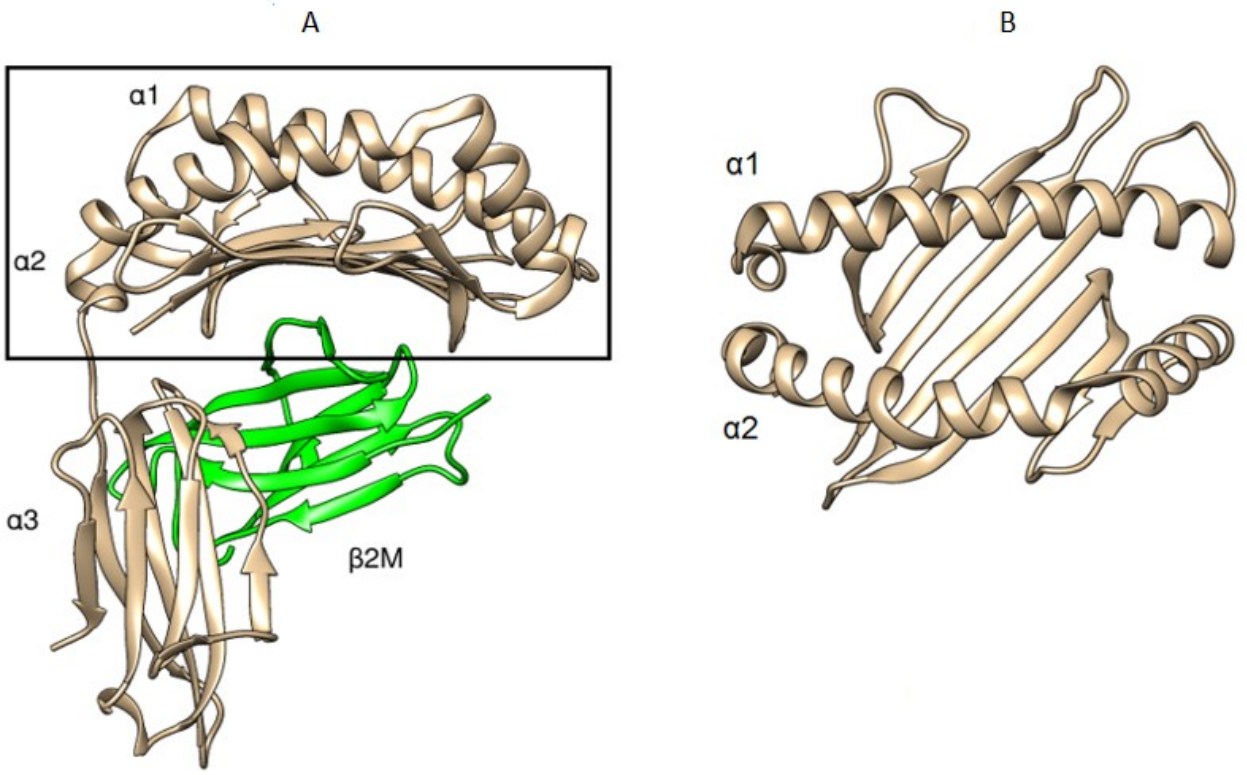


Figure 2. Cartoon representation of the MHC class I protein visualized with Chimera (Pettersen et al., 2004) (PDB ID: 5SWQ). A: The arrangement of α domains (tan color) and β2 microglobulin (β2M) domain (green color); black box denotes the peptide-binding groove. B: Top view of the peptide-binding groove.

As depicted in Figure 3, the peptide anchor residues to be buried in the peptide-binding groove of the MHC I molecule are at position 2 (P2) and 9 (P9). Positions 4 to 7 (P4-P7) are usually T-cell binders (Brown JH, 1993; Falk et al., 1991; Hunt et al., 1992; Stern et al., 1994). The geometry, charge distribution, and hydrophobicity of the binding groove determine the type of interaction between the peptide side chains and the MHC I binding groove (Wieczorek et al., 2017) (Figure 4).

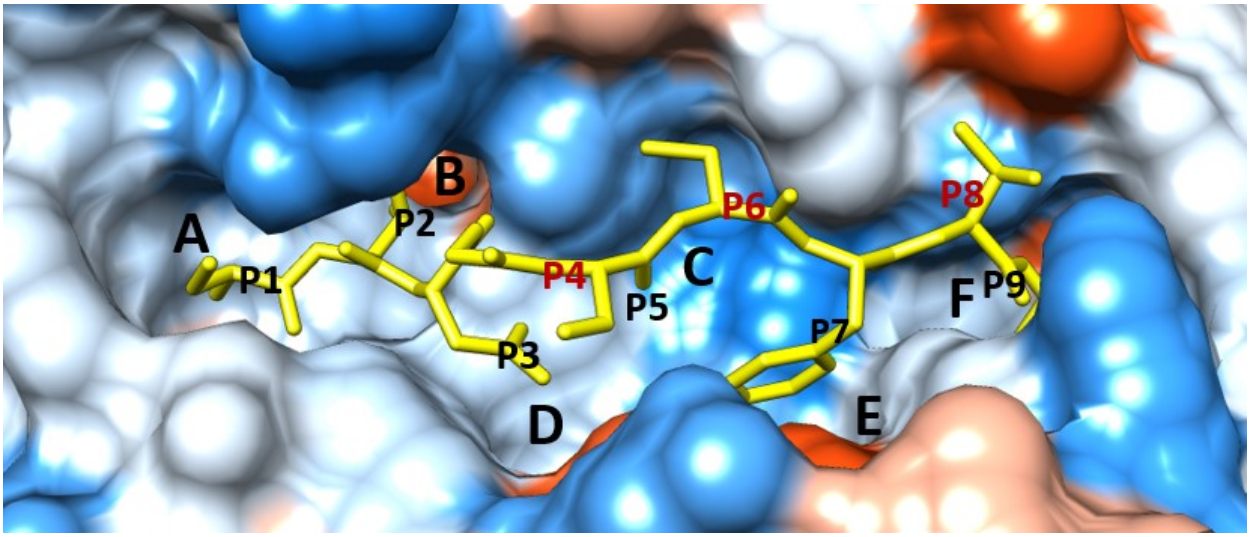


Figure 3. A stick model of a nonapeptide (yellow) binding to the MHC I molecule, HLA-A*0201; PDB ID: 5SWQ (surface representation). The subpockets are labeled from A to F. The main anchor residues (P2, P9) are deeply buried into the B and F subpockets, respectively. T-cell-receptor-binding residues (P4, P6 and P8) are labeled in red. P1, P5 and P7 side chains may bind both the MHC I groove and the T-cell receptor. Visualization with Chimera v. 1.13 (Pettersen et al., 2004) using the hydrophobicity scale of Kyte & Doolittle (1982): Hydrophobicity property from blue for the most hydrophilic, to white for intermediate hydrophobicity, to orange-red for the most hydrophobic.

A B C

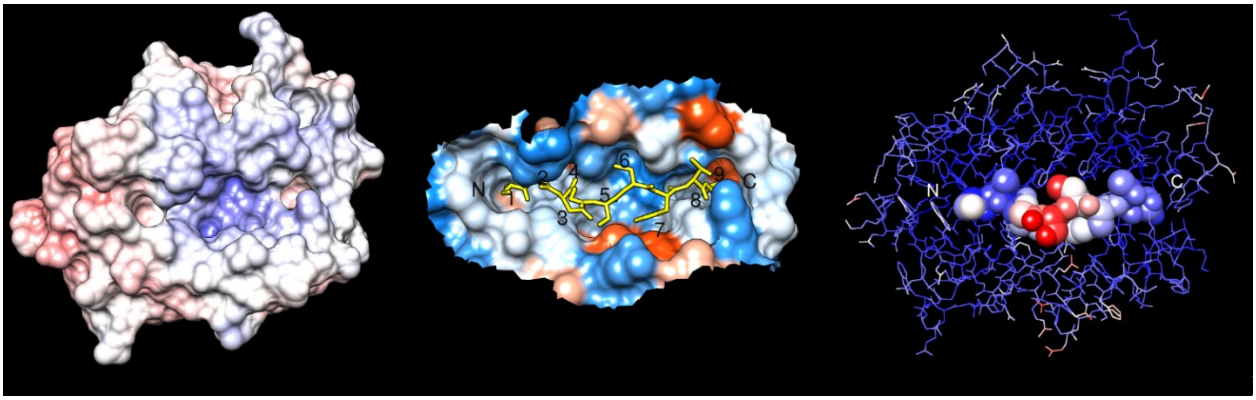


Figure 4. The MHC I peptide-binding groove properties determine the type of interactions formed with the peptide side chains. Top view; visualization with Chimera v. 1.13.1, PDB ID: 5SWQ. A: Molecular electrostatic potential (Coulombic surface coloring: blue for positive potential, white for neutral potential and red for negative potential). B: Hydrophobicity property from blue for the most hydrophilic, to white for intermediate hydrophobicity, to orange-red for the most hydrophobic (kdHydrophobicity scale) (Kyte, & Doolittle, 1982). C: B-factor distribution; the lowest B-factors (blue color) are in the MHC I groove core, peptide N-terminus (P1-P3) and C-terminus (P8 and P9) and the highest B factor (red color) is seen in the middle positions (P4-P7) of the peptide residues.

2.2 Vaccine development

Since viruses have multiple mechanisms to evade the host's immune response and exhibit diversity in their genes, a good approach for production of vaccines that stimulate T-cells is using highly conserved proteins of viruses as candidates (Rosendahl et al., 2014).

In recent years, bioinformatics has had a vital role in drug and vaccine design and development (Soria-Guerra et al., 2015). One goal of bioinformatics is to extract essential knowledge from genome, transcriptome and proteome to advance health issues (Brusic & Flower, 2004). Vaccine design by bioinformatics is more convenient, safer and less expensive than by earlier conventional approaches (Soria-Guerra et al., 2015).

The first demonstration of efficacy of a peptide vaccine containing a 15-mer peptide was published by Aichele et al. (1990). There is now a wide range of bioinformatics tools that are helpful for the prediction of novel T-cell epitopes based on the peptide anchor-residue binding positions in the MHC molecules (Adamczyk-Poplawska et al., 2011). Currently, it is well-established that T-cell epitopes interact with the MHC I binding groove in a linear mode (Aichele et al., 1990; Falk et al., 1991). This insightful concept inspired to develop a large number of algorithms to identify T-cell epitopes (DeLisi & Berzofsky, 1985; McMurry et al., 2005). Rosalia and co-workers (2013) showed that longer peptides result in enhanced CD8+ T-cell activation. CD8+ T-cells are the principal agents that clear the virus-infected cells (Rosalia et al., 2013).

There are antiviral drugs for the treatment of viral infectious diseases. However, viruses might escape such treatment by a mutation in their protein structure, which in turn leads to resistance against anti-viral drugs. To combat this challenge, activation and stimulation of host immune system can be an alternative CD4+ T helper cell and CD8+ cytotoxic T-cell response. The most notable advantage of peptide vaccines over protein and live attenuated vaccines is their flexibility to match escape variants. As mentioned earlier, mutation is a mechanism by which a virus can escape from the host immune system. Peptide vaccines are also safe and easy to produce. Nevertheless, choosing the right epitopes and avoiding the overstimulation of immune response is of great importance to be noted (Rosendahl et al., 2014).approach (Rosendahl et al., 2014). There are plenty of antibody-mediated vaccines which are, however, beneficial only in the

preventive stage of a viral infection since after the cells are infected only cellular immune response can eliminate the infection. This can be achieved by T-cell based vaccination. There are several types of T-cell-based vaccines, i.e. protein vaccines, live attenuated vaccines and peptide vaccines. Peptide vaccines are entirely synthetic, and they are composed of small protein fragments, peptides. Based on the included peptides, they may also be capable of inducing humoral response which stimulates both

2.3 General Principles in Rational Drug Design

There are generally two major requirements we need to achieve in order to be able to proceed in rational drug design. First, we need to understand the normal human physiology and pathology and also how the physiological/pathological pathways are regulated. This will help in mapping and identifying the target steps and molecules involved in the case of a disease. Second, we need knowledge about the 3D structure of the target molecule and/or its ligands. For structure-based drug design (direct design) there must exist knowledge about the target molecule and for ligand-based drug design (indirect design) (Tollenaere, 1996) there must exist knowledge about the molecules that bind to the target (Liljefors et al., 2002; Merz et al., 2010). Finally, after obtaining this knowledge and knowing about the disease at the molecular level, a drug can be designed to interact with the target to prevent or interrupt the disease. For this study, all requirements for rational design have been met (desired physiological effect: T-cell response against ZIKV virus; target protein(s): MHC I molecule(s); ligands: antigenic peptides from ZIKV proteins).

3. Aims

The overall aim of this study was to identify appropriate T-cell epitope candidates from Zika virus for the development of vaccine design using *in silico* techniques.

The specific aims were to:

- Evaluate sequence conservation among human MHC I molecules via sequence analysis.
- Evaluate the binding interactions inside the peptide-binding groove of the experimentally determined peptide-MHC I structures.
- Study the binding interactions in docked peptide-MHC I complexes in light of the binding interactions from known crystal complexes.
- Estimate the peptide binding affinity according to the peptide binding interactions.
- Rank the studied peptides based on their binding affinity as candidates for Zika vaccine development.

4. Materials and Methods

4.1 Ligands: Selection of potential antigenic peptides from ZIKV proteins

Mirza and co-workers (2016) predicted potential antigenic T-cell epitopes from three ZIKV proteins, namely E, NS3 and NS5 by employing the NetCTL-1.2 server (Larsen et al., 2007). In this study, eight epitopes from the total identified potential T-cell epitopes have been randomly selected for detailed binding interaction studies (Tables 1A-C).

Table 1A. Predicted T-cell epitopes from ZIKV E proteins (Mirza et al., 2016).

Residue number	Peptide sequence	Predicted MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport affinity	Prediction score*	MHC ligand
53	MAEVRSYCY	0.4403	1.8696	0.9649	3.021	2.1654	Yes
85	QSDTQYVCK	0.2076	0.8813	0.5921	0.443	0.9922	Yes
129	SIQPENLEY	0.4389	1.8634	0.9329	3.175	2.1621	Yes
159	ETDENRAKV	0.1722	0.7311	0.9755	−0.11	0.872	Yes
195	GLDFSDLYY	0.6604	2.8041	0.7805	2.599	3.0511	Yes
198	FSDLYYLTM	0.5872	2.4933	0.9307	−0.065	2.6297	Yes
205	TMNNKHWLV	0.2026	0.8603	0.9214	0.369	1.0169	Yes
308	CTAAFTFTK	0.1876	0.7964	0.6855	0.467	0.9226	Yes
324	GTVTVEVQY	0.1984	0.8423	0.9769	2.779	1.1278	Yes
368	STENSKMML	0.2251	0.9558	0.9741	0.675	1.1357	Yes

*NetCTL-1.2 prediction score threshold >0.75000. The epitopes GLDFSDLYY and TMNNKHWLV were selected to dock to MHC I HLA-A*0101 and the epitopes MAEVRSYCY and FSDLYYLTM were selected to dock to MHC I HLA-A*0201. Colors are in line with the resulting graphs in Chapter 5.

Table 1B. Predicted T-cell epitopes from ZIKV NS3 proteins (Mirza et al., 2016).

Residue number	Peptide sequence	Predicted MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport affinity	Prediction score*	MHC ligand
1	TTDGVYRVM	0.1514	0.643	0.9115	0.084	0.7839	Yes
47	RLDPYWGDV	0.1625	0.6897	0.9687	0.372	0.8536	Yes
75	HSEVQLLAV	0.277	1.176	0.2538	0.09	1.2186	Yes
105	DIGAVALDY	0.2072	0.8797	0.9491	2.544	1.1492	Yes
136	GVVIKNGSY	0.1489	0.6321	0.9462	3.103	0.9292	Yes

*NetCTL-1.2 prediction score threshold >0.75000. The colored epitopes were selected to dock to MHC I HLA-B*2705. Colors are in line with the resulting graphs in Chapter 5.

Table 1C. Predicted T-cell epitopes from ZIKV NS5 proteins (Mirza et al., 2016).

Residue number	Peptide sequence	Predicted MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport affinity	Prediction score*	MHC ligand
51	RTWAYHGSY	0.3114	1.3221	0.9789	3.267	1.6323	Yes
67	ASSLINGVV	0.1418	0.6019	0.8948	0.549	0.7636	Yes
91	IAMTDTTPY	0.1612	0.6844	0.8903	3.035	0.9697	Yes
93	MTDTTPYGQ	0.2728	1.1583	0.1576	−0.102	1.1769	Yes
124	MSMVSSWLW	0.2011	0.8539	0.8448	1.084	1.0349	Yes
233	FLEFEALGF	0.1492	0.6333	0.3368	2.447	0.8062	Yes
313	ALALAIIKY	0.216	0.9172	0.9296	3.237	1.2185	Yes
315	ALAIIKYTY	0.2376	1.0088	0.9516	3.153	1.3092	Yes
363	FTNLVVQLI	0.1698	0.7208	0.4951	0.334	0.8118	Yes
501	ETACLAKSY	0.3732	1.5844	0.3155	2.848	1.7742	Yes
509	YAQMWQLLY	0.6788	2.8819	0.8387	2.979	3.1567	Yes
555	MTTEDMLVV	0.1969	0.8361	0.2914	0.248	0.8922	Yes
581	VTKWTDIPY	0.2405	1.0211	0.8992	2.818	1.2969	Yes
584	WTDIPYLGK	0.3037	1.2893	0.9376	0.341	1.447	Yes
631	YMDYLSTQV	0.288	1.223	0.9517	0.341	1.3828	Yes

*NetCTL-1.2 prediction score threshold >0.75000. The colored epitopes were selected to dock to MHC I HLA-C*0801 (PDB ID: 4NT6). Colors are in line with the resulting graphs in Chapter 5.

4.2 Targets: Selection of MHC I molecules

As mentioned earlier, one of the key things in drug design is the selection of the target molecule and obtaining its 3D structure. The Protein Data Bank (PDB, http://www.rcsb.org/; Berman et al., 2000) is a freely accessible public database storing 3D structural data of biological macromolecules such as proteins and nucleic acids. PDB was used to retrieve the 3D structures of our target MHC I molecules that were selected following the work of Mirza et al (2016). The X-ray crystal structures of MHC I types HLA-A*0101, HLA-A*0201, HLA-B*2705 and HLA-C*0801 with the resolution of less than 2.5 Å were retrieved from PDB. Through superimposing all retrieved experimentally determined structures, eleven complexes following the common MHC I-peptide complex pattern were totally selected for this study: one structure for HLA-A*0101 (PDB ID: 4NQV) (N.B. this MHC I type was not studied in Mirza et al., 2016), eight structures for HLA-A*0201 (PDB IDs: 1OGA, 2GIT, 2GTW, 2GTZ, 3I6G, 3TO2, 4K7F and 5SWQ), one structure for HLA-B*2705 (PSB ID: 2BST) and one structure for HLA-C*0801 (PDB ID: 4NT6). The reason for studying eight different crystal structures of the MHC I HLA-A*0201 molecule was to map any possible consensus interactions that are always present when an antigenic peptide binds to that HLA type. This step was performed with Chimera v. 1.13.1, a molecular graphics tool developed by UCSF (http://www.rbvi.ucsf.edu/chimera) (Pettersen et al., 2004).

4.3 Amino Acid Sequence Retrieval

For the amino acid sequence analysis of the MHC-I molecules, the selected types of HLA sequences were obtained from the UniProt database (The UniProt Consortium, 2017, http://www.uniprot.org/; HLA-A*0101: P30443.1|HLA-A-1; HLA-A*0201: P01892.1|HLA-A-2; HLA-B*2705: P03989.2|HLA-B-27; HLA-C*0801: P30505.1|HLA-Cw-8) and stored in FASTA format. UniProt is a database containing information about the sequence and function of proteins (Chen et al., 2017). FASTA is a text-based format to represent nucleotide or peptide sequences in single-letter code (Lipman & Pearson, 1985). The sequence header starts with a greater-than (>) symbol.

Next, we looked for homologous HLA proteins for each selected MHC I protein in the

SwissProt database within UniProt. Accordingly, five homologs with an identity percent ranging from 70 to 95 percent for HLA-A*0101 and HLA-A*0201 types, six homologs for HLA-B*2705 and five homologs for HLA-C*0801 with the same range of identity percent were selected with the Basic Local Alignment Search Tool (BLAST, https://www.ncbi.nlm.nih.gov/BLAST/; Altschul et al. 1990). BLAST is one of the most used heuristic sequence searching algorithms in bioinformatics (it is heuristic in speeding up the computational process for a considerable amount of data) for finding high-scoring sequences among related sequences in a sequence database. Figure 5 presents the selected parameters for 'blasting'.

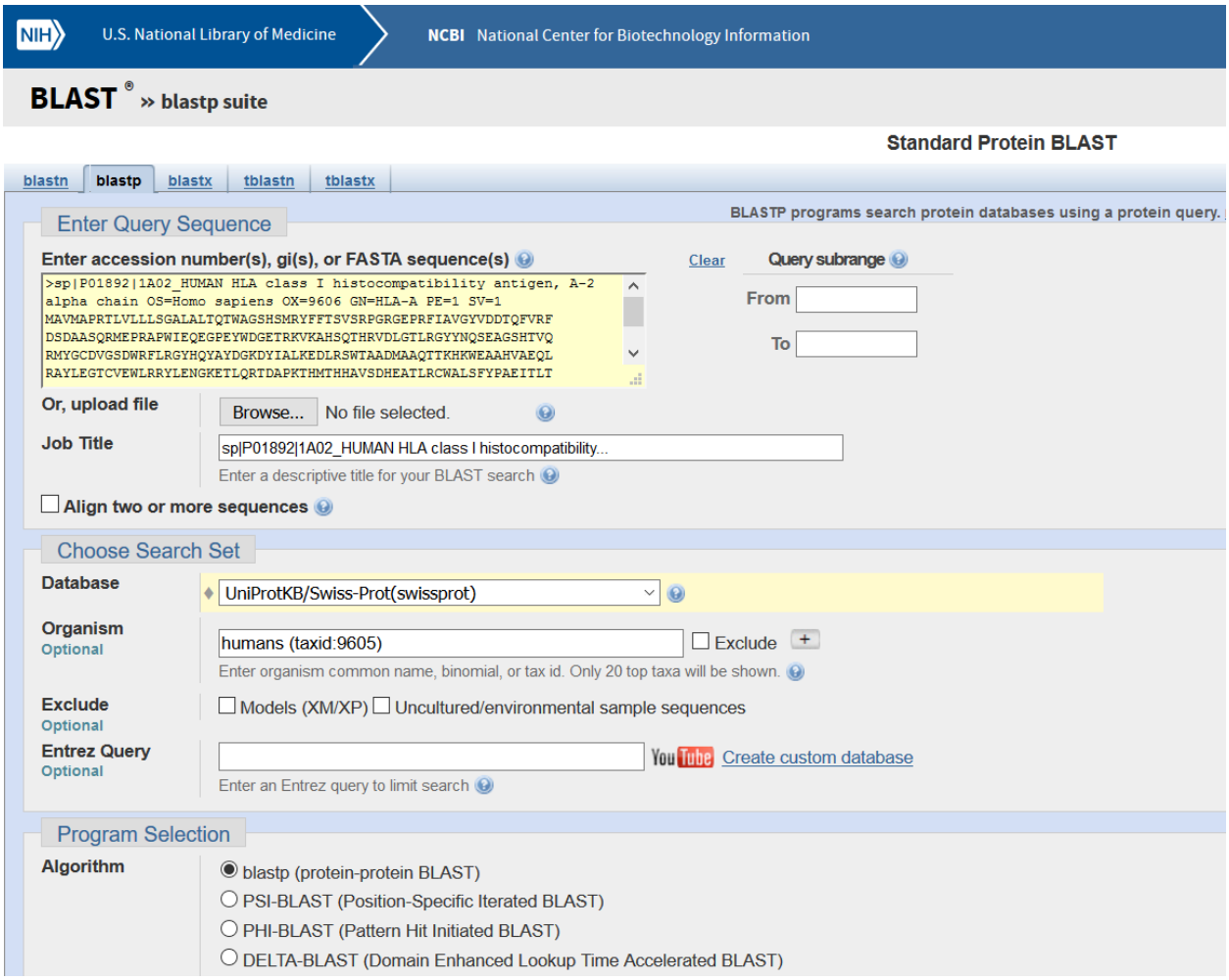


Figure 5. BLAST search for obtaining homologous MHC I sequences.

The next step was to perform a Multiple Sequence Alignment (MSA) of these twenty homologous sequences:

P30443.1|HLA-A-1	**P03989.2|HLA-B-27**	**P30505.1|HLA-Cw-8**
P01892.1|HLA-A-2	P01889.3|HLA-B-7	P30499.1|HLA-Cw-1
P04439.2|HLA-A-3	P30460.1|HLA-B-8	P30501.1|HLA-Cw-2
P13746.1|HLA-A-11	P30685.1|HLA-B-35	P04222.2|HLA-Cw-3
P18462.1|HLA-A-25	Q04826.1|HLA-B-40	P30504.1|HLA-Cw-4
P30512.2|HLA-A-29	P30481.1|HLA-B-44	Q9TNN7.1|HLA-Cw-5
P10314.2|HLA-A-32	P18465.1|HLA-B-57	

4.4 Multiple Sequence Alignment

MSA is the first step to analyse the sequence, function, and structure of biomolecules. The used aligner in this project was MUSCLE (multiple sequence comparison by log-expectation) (Edgar, 2004). MUSCLE is a progressive method for multiple sequence alignment. After generating the initial alignment, an optimal solution can be achieved by repeatedly modifying the suboptimal solution until no more improvement in the aligning score exists. The used parameters were as follows: the format for the generated multiple sequence alignment was Pearson/FASTA and the guide tree to output was set to zero. The resulting aligned FASTA file was imported to Clustal X (Larkin et al., 2007) to evaluate the redundancy and misfitting of the sequences. The MSA result was viewed with the GeneDoc tool (Nicholas & Nicholas, 1997) that can also be used for editing and annotating multiple sequence alignments.

4.5 Docking: Building the peptide-MHC I complexes

The crystal structures from PDB are not immediately ready for molecular modeling. The experimental structures need to be first prepared by removing metal ions, water molecules and cofactors, if any, and adding the hydrogen atoms. Moreover, multimeric structures may need to be reduced to a single unit and if the protein chains have missing residues, they should be assigned. For this purpose, the Protein Preparation Wizard of Maestro (Schrödinger Suite 2018-4) was used. In the preprocessing step, the following modifications were carried out: assign bond orders, add hydrogens, create disulfide bonds, filling missing side chains using Prime. Water

molecules beyond 5 Å from any non-protein groups were also removed. Further, the highest average occupancy position for any alternative amino acid side chain rotamers was selected. In the final step, the structure was refined with restrained minimization to remove steric clashes and relax side chains, allowing the heavy atoms to move at most 0.30 Å from their initial positions. After the protein structure preparation, Chimera v. 1.13.1 (Pettersen et al., 2004) was used to manually mutate the amino acids in the co-crystallized peptides of the PDB structures types, altogether eleven MHC I structures from PDB, (see section 4.2) to the corresponding residues of the ZIKV peptides (see Tables 1A-C). Sidechain rotamers were selected from the Dunbrak backbone-dependent rotamer library (Shapovalov & Dunbrack, 2011). Generally, a large number of steric clashes were found between the new peptide atoms and atoms of the interacting residues at the MHC I binding groove. To remove such clashes and refine the structures, the built MHC I-peptide complexes were submitted to the FlexPepDock tool (Londen et al., 2011; Raveh et al., 2010). The number of low-resolution and high-resolution structures were set to 200. Rosetta FlexPepDock aims at building models of protein-peptide complexes whilst trying to overcome the issue of conformational flexibility of the peptide. It also helps to evaluate peptide binding affinities and specificities based on the generated models. The protocol can produce high-resolution peptide-protein models by refining a coarse starting complex to a relatively native model. The protocol optimizes the peptide backbone and rigid body orientation by applying iteratively a Monte-Carlo approach and subsequent energy minimization. The hydrogen bonding interactions and steric clashes within resulting complexes were analyzed by Chimera. Donor-acceptor distance limit was set as < 3.3 Å.

4.6 Molecular Dynamics Simulations

All the refined docked complexes and, for comparison, the experimentally determined complexes were submitted to a molecular dynamics (MD) simulation for further refinement. MD simulations were first developed at the theoretical physics community during 1950s (Alder & Wainwright, 1957). The first MD simulation was performed using a crude molecular mechanics potential and only lasted for 9.2 ps (Artymiuk et al., 1979). MD presents the detailed physical motions and interactions of atoms of a molecule in a simulated environment (e.g. in vacuum or in water at a

certain temperature) as a function of time (Patodia et al., 2014). The regularly saved atomic coordinates over a course of time form the MD dynamics trajectory (different conformations) of a molecule. Multiple factors are substantial while performing a MD simulation; e.g. solvent, temperature and ions. The conformational flexibility of a protein in surface and loops is considerably related to the surronding environment. MD is an insightful approach that connects the structure and dynamics of a molecule by providing the means to study the conformational energy of proteins.

It can be said that the force field is the cornerstone of a MD simulation (Vitkup et al., 2000). MD simulations can be divided into two main categories; namely, classical molecular mechanics and quantum mechanics-based MD simulations. Since quantum mechanics-based MD simulation requires a huge amount of computational resources, currently the classical molecular mechanics-based simulation is the most common MD simulation method (Höltje et al., 2003). In molecular mechanics, a function that defines the energy of a molecule and describes the evolution of bonded (bond lengths, bond angles and torsions) and non-bonded (van der Waals and electrostatic interactions) interaction energies between atoms in a molecule over the course of time is called a force field. It is comprised of equations and parameters that determine the desired geometry of a particular molecule. The parameters for the function can be obtained either experimentally or theoretically. There are different force fields for different molecules based on the various atomic interactions within a molecule. In this study, the AMBER force field (Cornell, 1995; Elber et al., 1995) was applied.

There are several MD simulation packages such as NAMD (James & Phillips, 2005), Gromacs (Pronk et al., 2013) and Amber (Case et al., 2005). Amber16 (Assisted Model Building and Energy Refinement) simulation package was used in this study. The tleap module of AMBER (the basic tool to construct the force field files) was used to prepare the proteins and peptides (add hydrogens, build disulfide bonds) and the simulation systems employing the AMBER ff03 force field (Case et al, 2005). Each peptide-protein complex was dissolved in an octahedral TIP3P water (Jorgensen et al, 1983) box using Na+ as neutralizing counter ions. The minimum distance of the solute to the edge of the simulation box was set to 10.0 Å and periodic boundary conditions and particle-mesh Ewald electrostatics (Essman et al, 1995) with a cut-off of 9 Å for non-bonded

interactions were applied. SHAKE algorithm (Ryckaert et al., 1997) was applied to constrain the bonds to hydrogen atoms when using the 2-fs time step.

After preparing the simulation systems, the potential energy of the systems was minimized using the same stepwise protocol as described by Mirza et al. (2016). The purpose was not to find a global energy minimum but to find local minima on the potential energy surface and to obtain favorable bond lengths and bond angles minimizing the force pulling or pushing atoms together, removing the possible steric clashes and adjusting the distribution of solvent molecules (Dalke et al., 1997).

Before the actual production simulations, all the systems were equilibrated. The aim of this step is to equilibrate the energy of the system, relax the structure and the solvent.

Equilibration was performed in five steps, using a time step of 1 fs for the Langevin dynamics steps 1-4 and 2 fs for the last step of conventional dynamics. The protocol followed that of Mirza et al. (2016). Shortly:

Step 1: Heating the system for 10 ps from 10 K to 300 K at constant volume using Langevin dynamics [gamma_ln (collision frequency) = 1.0 ps^{-1}], with a fixed solute (restraint force constant 5 kcal/mol$Å^2$).

Step 2: A new heating of the system for 20 ps from 10 K to 300 K using Langevin dynamics [gamma_ln = 1.0 ps^{-1}], no positional restraints.

Step 3: Equilibration of the system for 20 ps at constant temperature (300 K) and constant volume using Langevin dynamics [gamma_ln = 0.5 ps^{-1}], no positional restraints.

Step 4: Equilibration of the system for 50 ps at constant temperature (300 K) and constant pressure of 1.0 bar using Langevin dynamics [gamma_ln = 0.5 ps^{-1}] to adjust the density of the system, no positional restraints.

Step 5: Equilibration of the system for 400 ps at constant temperature (300 K) and constant pressure of 1.0 bar (coupling constant for temperature = 5.0 ps and for pressure = 2.0 ps), using conventional MD, no positional restraints.

The actual production simulation was run at constant temperature of 300 K and pressure of 1 bar and the length of the production run was 10 ns. The coupling constants for temperature and pressure were 5.0 and 2.0 ps, respectively. In this step, velocities and coordinates of system

are saved as trajectories to be analyzed.

4.7 Analysis of the stability of the peptide-MHC I interactions

The cpptraj module of AmberTools16 was used to analyze the trajectories from the MD simulations (Roe & Cheathman III, 2013). Root Mean Square Deviation (RMSD) calculation is the most common way to measure how stable a structure remains during the simulation. RMSD is the distance between the corresponding positions of two atoms in the two structures (e.g. the starting structure before MD simulation and the final structure after MD simulation). It is a global measurement and calculated with respect to one reference structure (Devadoss & Raj, 2014). We used the starting structure as the reference and compared its backbone atoms with the backbone atoms of all the saved structure conformations along the trajectory. We also calculated RMSD of Cα atoms between the initial docked peptide conformation and the minimized conformation after the MD simulations. Root Mean Square Fluctuations (RMSF) of atoms per residue were also calculated. The RMSF graphs show which residues fluctuate the most or least and provide thus information on for example which of the peptide residues are engaged in stable/tight interactions (anchor residues) with the receptor and which ones are free to move more.

Prime-MMGBSA (Molecular Mechanics Generalized Born Surface Area) module of Maestro was used to estimate the free energy of binding of the peptides both in the initial (crystal or docked) complexes as well as the final minimized complexes from the MD simulations (Schrödinger, 2016). Prime-MMGBSA binding energy for the final complexes was calculated using a rigid peptide-MHC I complex (Jacobson et al., 2002; Jacobson et al., 2004; Schrödinger Release 2018-1**:** Maestro/Prime, Schrödinger, LLC, New York, NY, 2018)

Residues that bind to the MHC I can be identified by their high degree of burial (Höltje et al., 2003). For this purpose, the solvent accessible surface area (SASA) was calculated for the peptide residues with **NACCESS** v. 2.1.1 (Hubbard & Thornton, 1993) to evaluate their binding to the MHC I pockets or availability to interact with a TCR. The program uses the Lee & Richard's (1971) method. The SASA value depends on the tightness, strength and the number of interactions of the peptide residues with the MHC I pocket and is a measure of a residue's depth of binding to the pocket (Höltje et al., 2003). The measurement unit is square Ångströms ($Å^2$) and

the result is given as a relative accessibility (%) value for all atoms of a particular residue.

Change in the size of the MHC I subpocket F was also measured before and after the MD simulation. It has been suggested that a tightly binding peptide closes the MHC I binding groove and a poorly binding peptide widens the groove (Fleischmann et al., 2015) and the size of the flexible F pocket can give some insight into the groove size (Abualrous et al., 2015). For this purpose the distances between Cα atom of residue 74 in α1 helix and residue 149 in α2 helix (d1) and between Cα atom of residue 85 in α1 helix and residue 138 in α2 helix (d2) were measured.

5. Results and Discussion

In this chapter we will first analyze the results from the Multiple Sequence Alignment for different MHC I HLA types, namely HLA-A, HLA-B, and HLA-C to find the degree of conservation/variability particularly in the peptide-binding domain. Knowing about the variability in different alleles of MHC I HLA molecules will assist to explain the broad specificity of MHC I molecules to a wide range of antigens.

Next, we will proceed to the analysis of the docked peptide-MHC I complexes compared to the experimentally determined peptide-MHC I (crystal) complexes before the MD simulation. This will give insight into the goodness of the docked poses. For example, are there big steric clashes between the peptide and the MHC I binding pocket? Do the predicted peptides have similar interactions with the MHC I binding groove compared to an experimentally verified peptide? Moreover, analyzing the complexes before and after the MD simulation helps us to evaluate the significance of MD in refining the peptide binding poses in the MHC I groove.
Eventually, we will analyze the resulting trajectories from the MD simulations. In this section, we will discover how the different ZIKV peptide-MHC I complexes behave during the simulations with the aim to find the most stable complex, i.e. the peptide that has the best binding affinity to the MHC I binding groove. Such a peptide or peptides could be suitable candidates for the design of a vaccine against the ZIKV infection. Here, the analyzed complexes include both experimentally determined (crystal) and docked peptide-MHC I complexes. The crystal complexes act as a reference to compare the binding of the docked peptides with the experimentally verified binding of antigenic peptides (co-crystalized peptides).

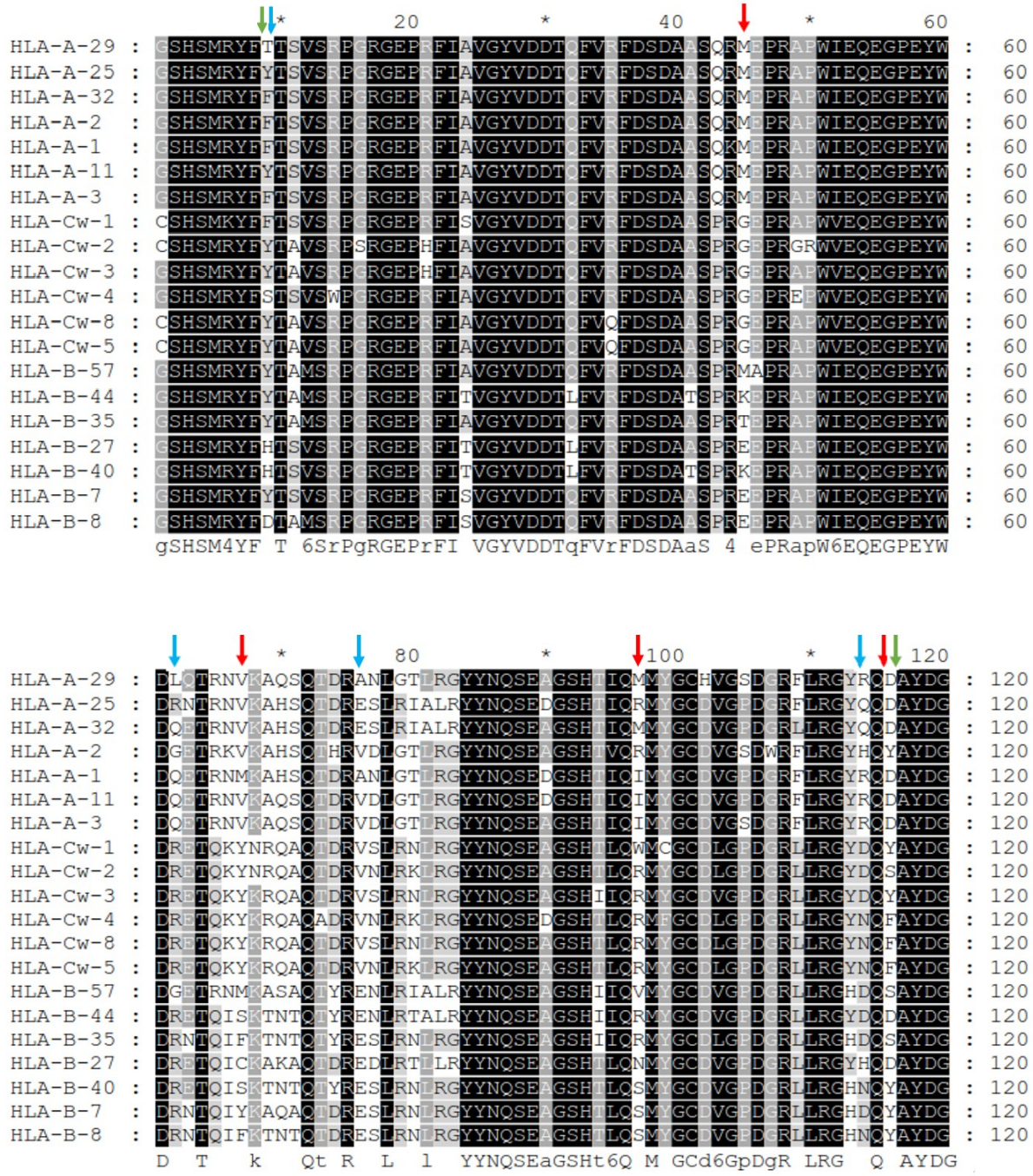
As mentioned before, we had eleven experimental peptide-MHC I structures from PDB plus two docked ZIKV peptide complexes for each experimental MHC I structure; altogether, we analyzed 33 complexes. We analyzed the predicted antigenic peptides derived from the ZIKV E protein in complex with the following MHC I HLA structures: HLA-A*0101 and HLA-A*0201 and the peptides from the ZIKV NS3 protein in complex with HLA-B*2705 and the peptides from ZIKV NS5 protein in complex with HLA-C*0801. The reason for using several different crystal structures of the HLA-A*0201 molecule was to identify the key residues interacting with different antigenic peptides. We wanted to investigate if there was any consensus binding pattern for all the binding

peptides or significant variation in binding interactions depending on the antigenic peptide bound.

5.1 Multiple Sequence Alignment

To investigate the degree of conservation among the MHC I HLA molecules, especially in the peptide-binding groove (active site), sequences were aligned with the MUSCLE algorithm. As it can be seen from the alignment (Figure 6.1), the gap-free areas are mostly belonging to the more rigid parts corresponding to the core areas of the protein while the gap-rich areas belong to the more flexible parts such as loop regions. In addition to the gaps, physicochemical characteristics of amino acids are another criterion when investigating MSAs. A change in the type of an amino acid in a sequence can, in turn, lead to a change in the protein structure and characteristics. Generally, the charged amino acids, namely lysine, arginine, glutamic acid and aspartic acid are common residues involved in ligand binding.

HLA genes are extremely polymorphic. It means most HLA genes consist of a huge number of allelic variants. Among the MHC I molecules, HLA-B has the greatest number of allelic variants (Reche & Reinherz, 2003). Although variations in the sequence are distributed through the whole sequence of the MHC I molecules, these polymorphic sites are more notable in the α1α2 domain (residues 1 to 180) which form the peptide-binding cleft (Clements et al., 2005; Reche & Reinherz, 2003). It should also be noted that the number of variable amino acids is higher in the region between residues 62 to 84 than the remaining part of the domain (Figure 6.1).


20
40
60
HLA-A-29 : GSHSMRYFTTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEPRAPWIEQEGPEYW : 60
HLA-A-25 : GSHSMRYFYTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEPRAPWIEQEGPEYW : 60
HLA-A-32 : GSHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEPRAPWIEQEGPEYW : 60
HLA-A-2 : GSHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEPRAPWIEQEGPEYW : 60
HLA-A-1 : GSHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQKMEPRAPWIEQEGPEYW : 60
HLA-A-11 : GSHSMRYFYTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEPRAPWIEQEGPEYW : 60
HLA-A-3 : GSHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEPRAPWIEQEGPEYW : 60
HLA-Cw-1 : CSHSMKYFFTSVSRPGRGEPRFISVGYVDDTQFVRFDSDAASPRGEPRAPWVEQEGPEYW : 60
HLA-Cw-2 : CSHSMRYFYTAVSRPSRGEPHFIAVGYVDDTQFVRFDSDAASPRGEPRGRWVEQEGPEYW : 60
HLA-Cw-3 : GSHSMRYFYTAVSRPGRGEPHFIAVGYVDDTQFVRFDSDAASPRGEPRAPWVEQEGPEYW : 60
HLA-Cw-4 : GSHSMRYFSTSVSWPGRGEPRFIAVGYVDDTQFVRFDSDAASPRGEPREPWVEQEGPEYW : 60
HLA-Cw-8 : CSHSMRYFYTAVSRPGRGEPRFIAVGYVDDTQFVQFDSDAASPRGEPRAPWVEQEGPEYW : 60
HLA-Cw-5 : CSHSMRYFYTAVSRPGRGEPRFIAVGYVDDTQFVQFDSDAASPRGEPRAPWVEQEGPEYW : 60
HLA-B-57 : GSHSMRYFYTAMSRPGRGEPRFIAVGYVDDTQFVRFDSDAASPRMAPRAPWIEQEGPEYW : 60
HLA-B-44 : GSHSMRYFYTAMSRPGRGEPRFITVGYVDDTLFVRFDSDATSPRKEPRAPWIEQEGPEYW : 60
HLA-B-35 : GSHSMRYFYTAMSRPGRGEPRFIAVGYVDDTQFVRFDSDAASPRTEPRAPWIEQEGPEYW : 60
HLA-B-27 : GSHSMRYFHTSVSRPGRGEPRFITVGYVDDTLFVRFDSDAASPREEPRAPWIEQEGPEYW : 60
HLA-B-40 : GSHSMRYFHTSVSRPGRGEPRFITVGYVDDTLFVRFDSDATSPRKEPRAPWIEQEGPEYW : 60
HLA-B-7 : GSHSMRYFYTSVSRPGRGEPRFISVGYVDDTQFVRFDSDAASPREEPRAPWIEQEGPEYW : 60
HLA-B-8 : GSHSMRYFDTAMSRPGRGEPRFISVGYVDDTQFVRFDSDAASPREEPRAPWIEQEGPEYW : 60
gSHSM4YF T 6SrPgRGEPrFI VGYVDDTqFVrFDSDAaS 4 ePRapW6EQEGPEYW
80
100
120
HLA-A-29 : DLQTRNVKAQSQTDRANLGTLRGYYNQSEAGSHTIQMMYGCHVGSDGRFLRGYRQDAYDG : 120
HLA-A-25 : DRNTRNVKAHSQTDRESLRIALRYYNQSEDGSHTIQRMYGCDVGPDGRFLRGYQQDAYDG : 120
HLA-A-32 : DQETRNVKAHSQTDRESLRIALRYYNQSEAGSHTIQMMYGCDVGPDGRLLRGYQQDAYDG : 120
HLA-A-2 : DGETRKVKAHSQTHRVDLGTLRGYYNQSEAGSHTVQRMYGCDVGSDWRFLRGYHQYAYDG : 120
HLA-A-1 : DQETRNMKAHSQTDRANLGTLRGYYNQSEDGSHTIQIMYGCDVGPDGRFLRGYRQDAYDG : 120
HLA-A-11 : DQETRNVKAQSQTDRVDLGTLRGYYNQSEDGSHTIQIMYGCDVGPDGRFLRGYRQDAYDG : 120
HLA-A-3 : DQETRNVKAQSQTDRVDLGTLRGYYNQSEAGSHTIQIMYGCDVGSDGRFLRGYRQDAYDG : 120
HLA-Cw-1 : DRETQKYNRQAQTDRVSLRNLRGYYNQSEAGSHTLQWMCGCDLGPDGRLLRGYDQYAYDG : 120
HLA-Cw-2 : DRETQKYNRQAQTDRVNLRKLRGYYNQSEAGSHTLQRMYGCDLGPDGRLLRGYDQSAYDG : 120
HLA-Cw-3 : DRETQKYKRQAQTDRVSLRNLRGYYNQSEAGSHIIQRMYGCDVGPDGRLLRGYDQYAYDG : 120
HLA-Cw-4 : DRETQKYKRQAQADRVNLRKLRGYYNQSEDGSHTLQRMFGCDLGPDGRLLRGYNQFAYDG : 120
HLA-Cw-8 : DRETQKYKRQAQTDRVSLRNLRGYYNQSEAGSHTLQRMYGCDLGPDGRLLRGYNQFAYDG : 120
HLA-Cw-5 : DRETQKYKRQAQTDRVNLRKLRGYYNQSEAGSHTLQRMYGCDLGPDGRLLRGYNQFAYDG : 120
HLA-B-57 : DGETRNMKASAQTYRENLRIALRYYNQSEAGSHIIQVMYGCDVGPDGRLLRGHDQSAYDG : 120
HLA-B-44 : DRETQISKTNTQTYRENLRTALRYYNQSEAGSHIIQRMYGCDVGPDGRLLRGYDQDAYDG : 120
HLA-B-35 : DRNTQIFKTNTQTYRESLRNLRGYYNQSEAGSHIIQRMYGCDLGPDGRLLRGHDQSAYDG : 120
HLA-B-27 : DRETQICKAKAQTDREDLRTLLRYYNQSEAGSHTLQNMYGCDVGPDGRLLRGYHQDAYDG : 120
HLA-B-40 : DRETQISKTNTQTYRESLRNLRGYYNQSEAGSHTLQSMYGCDVGPDGRLLRGHNQYAYDG : 120
HLA-B-7 : DRNTQIYKAQAQTDRESLRNLRGYYNQSEAGSHTLQSMYGCDVGPDGRLLRGHDQYAYDG : 120
HLA-B-8 : DRNTQIFKTNTQTDRESLRNLRGYYNQSEAGSHTLQSMYGCDVGPDGRLLRGHNQYAYDG : 120
D T k Qt R L l YYNQSEaGSHt6Q M GCd6GpDgR LRG Q AYDG

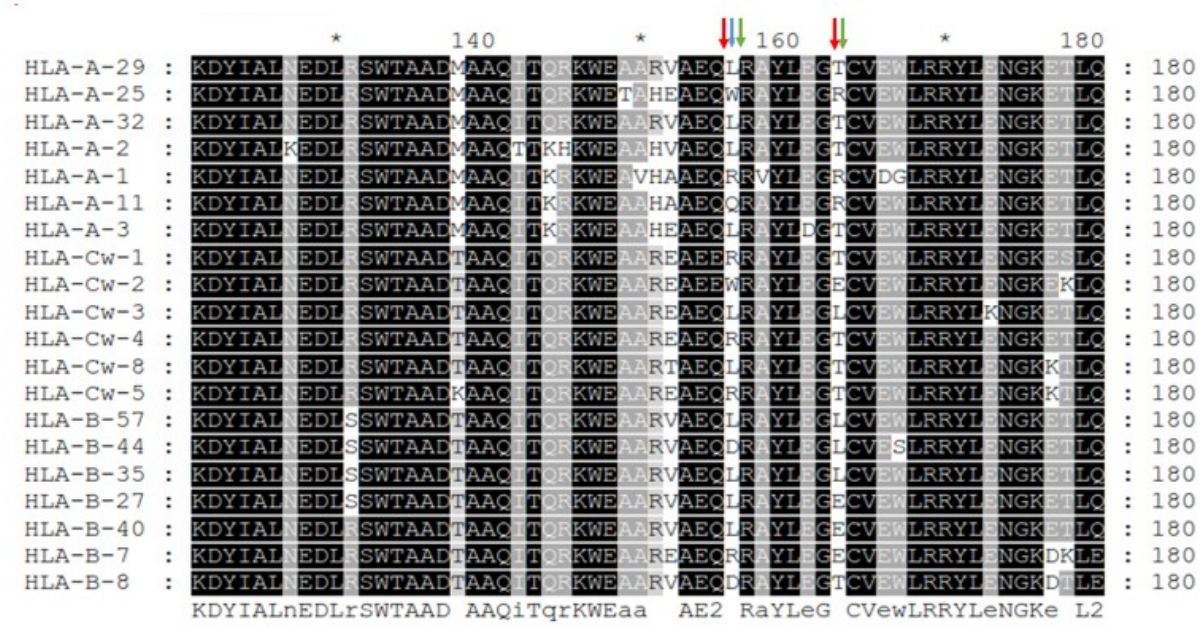


Figure 6.1. Multiple Sequence Alignment of α1 and α2 domain (residues 1-180) from the HLA-A, HLA-B and HLA-C alleles. The twenty sequences were obtained from the UniProt database and aligned by the MUSCLE algorithm. Blue, red and green arrows demonstrate the polymorphic peptide contact sites for HLA-A, HLA-B and HLA-C molecules, respectively. The alignment was visualized with the GeneDoc tool.

Black background shading with white text indicates that a single residue is conserved across the entire alignment. The second level of shading is a dark gray background with white text. It identifies the columns in the alignment where a single residue is conserved in all of the sequences and the third level of shading with black text and gray background color identifies the columns in the alignment where a single physicochemical property group is conserved in all of the sequences (Nicholas & Nicholas, 1997).

Reche and Reinherz' (2003) study reinforces this notion. They calculated sequence variability from relevant multiple sequence alignments of HLA types A, B and C and defined a variability metric (*V*) that was plotted for each site in the alignment. Their result showed that all peptide-binding residues obtained *V*>1 (polymorphic site). They also showed a variability profile for each HLA molecule which is consistent with our alignment (Figure 6.1): residues 9, 62, 76, 114 and 156 in the HLA-A, residues 45, 67, 97, 116, 156 and 163 in the HLA-B and residues 9, 116, 156 and 163 in the HLA-C are the most variable sites in each MHC I HLA group. These profiles demonstrate the specificity of peptide-binding in different MHC I allels. When this variability profile is mapped onto the 3D structure of the MHC I molecules, some residues, such as 9 and 114 in HLA-A reside in the β-sheet that forms the floor of the peptide- binding groove, while residues 67 in HLA-B and

156 in HLA-C locate on the α-helices (Figure 6.2). It can be concluded that the greatest diversity is exhibited by the binding groove (α1α2 domain), in consistence with the observed polymorphism. In sum, each HLA molecule has its variability profile and this high level of allelic diversity is attributed to the immune protection against a wide range of pathogens (Reche & Reinherz, 2003).

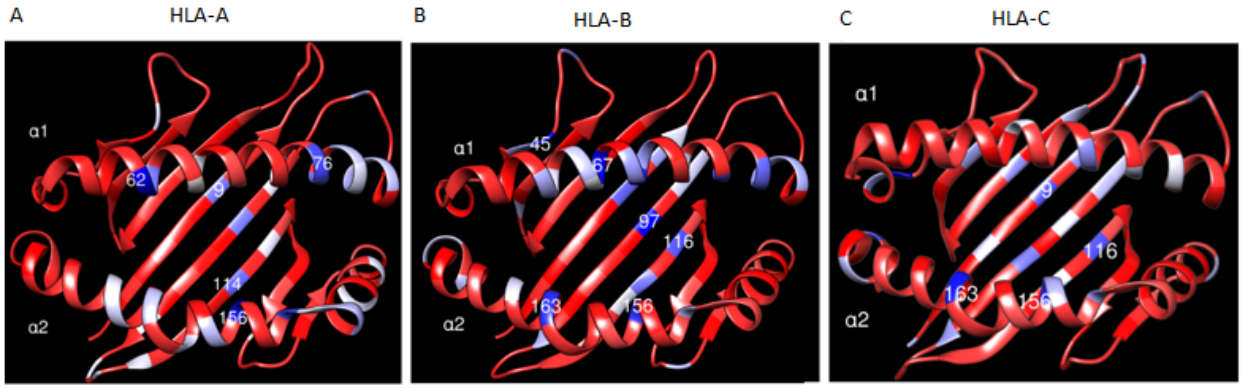


Figure 6.2. Sequence variability in the peptide-binding groove (α1α2 domain) of the MHC I HLA molecules. (A1-C1) The peptide-binding groove depicted using the ribbon style denoting the most variable residue sites in HLA-A, HLA-B, HLA-C. The following PDB structures were used for rendering the conservation attributes: HLA-A: 5SWQ, HLA-B: 2BST, HLA-C: 4NT6. Red color denotes the non-variable areas; blue denotes variable regions and white represents the intermediately variable regions. AL2CO was selected as Conservation style (Pei & Grishin, 2001). The images were created with Chimera v. 1.13.1.

5.2 Evaluation of the docking results in the light of the MD refinement

5.2.1 Peptides from the ZIKV E protein

Structural analysis of the co-crystalized CTELKLSDY peptide bound to the **HLA-A*0101** molecule (PDB ID: 4NQV, resolution of 2.4 Å) indicated that residues at peptide positions P2 (Thr) and P9 (Tyr) are the anchor residues while the residue at P3 (Glu) is buried in the D pocket forming a salt bridge with Arg156. Residues at P4 (Leu), P5 (Lys) and P8 (Asp) are exposed to the solvent (Table 2A). However, the P8 backbone atom (O) interacts with Trp147. An extended peptide conformation stabilizes the HLA-A*0101 molecule and these interactions enhance the stability of the complex, leading to the peptide presentation to TCR and immunogenicity. Evaluation of the docking results of the predicted antigenic peptides TMNNKHWLV and GLDFSDLYY from ZIKV E protein complexed with HLA-A*0101 indicated that the two peptides retained some of their initial hydrogen bonds during the MD refinement. However, this was more remarkable for GLDFSDLYY (Table 2A). Tyr159 (atom OH) interacting with the peptide backbone (P1 oxygen) is the only residue retaining its interaction in both initially docked complexes and after MD simulation complexes. Glu63 (atoms OE1 and OE2), Asn77 (atom OD1) and Lys146 (atom NZ) seem to be the most crucial residues in this HLA molecule since they participate in hydrogen bond interactions in the experimental complex both before and after the MD simulation. Among the peptide residues, residue 9 (Tyr) from GLDFSDLYY has the largest number of H-bond interactions in the pocket F (Table 2A).

The two docked complexes exhibit an increased number of H-bond interactions in the peptide-binding groove after the MD simulation (Table 2A). No steric clashes were found in the initial docked complexes or after the MD simulation. Besides, the GLDFSDLYY-HLA complex has a lower (better) value of hydrogen bonding energy than the TMNNKHWLV-HLA complex (Table 3).

Table 2A. Hydrogen bonding interactions in HLA-A*0101 (PDB ID: 4NQV) complexes.

Peptide Interacting atom	HLA-A*0101 Interacting atom	CTELKLSDY (native peptide)				TMNNKHWLV[a]				GLDFSDLYY[a]			
		D^b_{init}		D^c_{MD}		D^b_{init}		D^c_{MD}		D^b_{init}		D^c_{MD}	
P1(N)	GLU 63 OE1			*	3.448							*	2.693
	GLU 63 OE2			*	2.751								
	TYR 159 OH	*	2.671	*	2.637								
P1(O)	TYR 159 OH					*	2.757	*	3.142	*	2.743	*	2.664
P1(OG1)	GLU 63 OE1							*	2.588				
P2(N)	GLU 63 OE1			*	2.742			*	3.264				
	GLU 63 OE2	*	2.755			*	3.223	*	2.868	*	3.327		
P2(O)	ARG 163 NH2	*	2.609			*	2.861			*	2.835		
	ARG 163 NH1					*	3.411			*	3.297		
P2(OG1)	GLU 63 OE2			*	2.597							*	2.869
P3(OD2)	ARG 156 NH1			*	2.939					*	2.895	*	2.865
P3(OE1)	ARG 156 NH1	*	2.872										
P6(OD1)	ARG 114 NH2									*	2.918	*	2.869
P7(O)	ASN 77 ND2	*	3.268	*	2.840	*	3.336			*	3.248	*	3.411
	ARG 114 NH1			*	2.961								
P7(N)	THR 73 OG1							*	3.060				
P8(O)	TRP 147 NE1	*	2.782	*	2.812	*	3.174			*	2.881	*	2.902
P9(N)	ASN 77 OD1	*	2.946			*	3.060	*	3.301	*	3.127	*	2.980
P9(O)	TYR 84 OH			*	2.868			*	2.683				
	THR 143 OG			*	2.733								
	LYS 146 NZ			*	3.043			*	3.170				
P9(OH)	ASP 116 OD2	*	2.460	*	2.693					*	2.712	*	2.672
P9(OXT)	LYS 146 NZ			*	2.801			*	2.806			*	2.859
	THR 80 OG1											*	2.604
Total no. of interactions		8		14		7		9		10		11	
Side chain interactions		2		3		-		2		3		4	

[a]Docked ZIKV E peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera v. 1.13.1 (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after molecular dynamics (MD) simulation. Stars denote the existence of an H-bond.

The co-crystallized GLMWLSYFV peptide (from SARS coronavirus membrane protein) is deeply buried via residues at P2 (Leu) and P9 (Val) in the B and F pockets of the **HLA-A*0201** (PDB ID: 3I6G) peptide-binding groove, respectively. The other two anchor positions, known as secondary anchor residues are located at P3 (Met) and P6 (Ser). Met3 resides in the subpocket D and Ser6 is buried in the subpocket C (Liu et al, 2010). These interactions enhance the stability of the complex. The peptide FSDLYYLTM from ZIKV E protein docked to the 3I6G structure forms hydrogen bonds via P1 and P2 backbone atoms (O and N) in the N-terminus of the peptide in the same way as the co-crystalized peptide does, whereas Met1 of MAEVRSYCY is not involved in binding to the HLA-A*0201 binding groove in this particular HLA-A*0201 structure (Appendix 2E; PDB ID: 3I6G). Moreover, the C-terminal segment (positions P8 and P9) of both MAEVRSYCY and FSDLYYLTM are involved in binding to the MHC I cleft. The difference in the binding pattern for the two predicted peptides is observed for the central residues (P4 to P7). Residues at P4 and P6 of MAEVRSYCY are involved in the hydrogen bonding network with the binding groove of the HLA molecule similarly to the co-crystalized peptide while those residue positions in FSDLYYLTM might rather interact with the TCR. These differences contribute to differential T-cell recognition which leads to diverse immunogenicity. Trp147, Tyr159, and Asp77 are the most crucial residues in this MHC I HLA molecule to interact with the co-crystallised peptide as well as the docked peptides.

Table 3 presents the FlexPepDock hydrogen bonding energy for the two docked complexes, that of FSDLYYLTM being slightly better (-23.087 kg/mol). No steric clashes were observed in the complexes after manual docking and FlexPepDock refinement nor after the MD simulation for either of the docked peptides.

The most important residue for interacting with the peptides in the HLA-A*0201 structure PDB ID: 3TO2 is Trp147. It interacts with the docked peptides (MAEVRSYCY and FSDLYYLTM) as well as the co-crystallized peptide (LACFVLAAV from SARS coronavirus membrane glycoprotein) (Appendix 2F). Peptide positions P1, P2, P6, P8 and P9 are responsible for forming hydrogen bond interactions in the crystal complex and the docked complex with MAEVRSYCY , whereas P6 does not participate in H-bond interactions in the complex with FSDLYYLTM before MD. On the other hand, P3 (Glu) of MAEVRSYCY interacts with the binding groove only via the backbone oxygen whereas no interaction was formed between P3 and HLA crystal structure binding groove. H-

bonding energies and the observed initial clashes between the docked peptides and the MHC I molecules are presented in Table 3.

Post-docking analysis of the ZIKV E peptides docked at the HLA-A*0201 structure PDB ID: 5SWQ revealed that the FSDLYYLTM peptide retained most of the initial hydrogen bonds within the binding cleft, whereas MAEVRSYCY could keep only two hydrogen bonds out of seven initial H-bonding interactions (Appendix 2H). Significant hydrogen bonds were formed between Cys8 and Thr9 from both ZIKV E peptides and Trp147, Asp77 from the HLA protein. Hydrogen bonding energy values resulting from FlexPepDock were -23.288 kcal/mol and -24.398 kcal/mol for MAEVRSYCY and FSDLYYLTM, respectively (Table 3). Both peptides showed some initial clashes with the MHC I peptide-binding groove before the MD refinement but the clashes were removed after the MD simulation (Table 3).

In addition, the post-docking results of MAEVRSYCY and FSDLYYLTM in complex with the other crystal structures of HLA-A*0201 (PDB IDs: 1OGA, 2GIT, 2GTZ, 2GTW and 4K7F) are presented in Table 3 and Appendices 2A-D and 2G, respectively. Overall, the post-docking analysis of the HLA-A*0201 structures in complex with the different predicted antigenic peptides revealed that all the analyzed HLA-A*0201 complexes share some identical amino acid residues that are involved in binding with the peptides. For instance, Glu63, Tyr159, Trp147 and Asp77 interacted with the bound peptide almost in all the HLA-A*0201 crystal structures. In sum, the binding interactions depend on the peptides but interactions with certain groove residues are especially favorable for the peptide binding.

As it was mentioned earlier, choosing eight different crystal structures for the MHC I HLA-A*0201 type was used for investigating the crucial interacting residues of the HLA-A*0201 molecule to find a possible consensus binding pattern for the peptides. Looking at the interacting residues in the binding-groove of the HLA-A*0201 crystal complexes revealed that despite the identical interacting residues from HLA-A*0201 (Glu63, Tyr159, Trp147 and Asp77) with P2 and P9 (the main anchor residues of the peptides), there are also some variable interacting residues binding with the partial anchor residues (at P1, P3, P6 and P8) or the central region residues (at P4-P7). These impact the orientation adopted by the main anchor residues, which then affects the adopted pose of the peptide inside the MHC I binding-groove, which in turn affects the

stability and, finally, the immunogenicity of the presented peptide. The crystal complex PDB ID: 5SWQ is a good example on this. The aromatic residue Phe at P7 of the CVNGSCFTV peptide (an NA231 influenza epitope) is buried inside the groove and interacts with Trp147 of the HLA molecule and is not solvent-exposed (Appendix 2H). This affects the conformation of Cys6 in such a way that the general SASA of the peptide is reduced (Grant et al., 2016).

In conclusion, structural analysis of the experimentally determined HLA-A*0201 complexes demonstrated that the peptides LLFGKPVYV (PDB ID: 2GIT), ALGIGILTV (PDB ID: 2GTZ) and GLMWLSYFV (PDB ID: 3I6G) form more stable complexes than the other peptides in the HLA-A*0201 crystal complexes. This is due to the different interacting residues engaged in such complexes beside the identical interacting residues. For example, LLFGKPVYV engaged Tyr171 and Tyr 84 with its residues at P1 (Leu) and P9 (Val), respectively. ALGIGILTV engaged Tyr 59 with its P1 (Ala) and Arg97 was interacting with P6 (Ser) of GLMWLSYFV. Interestingly, Tyr171 was also involved in interacting with P1 of the two docked peptides (MAEVRSYCY and FSDLYYLTM) and Tyr84 was involved in interacting with FSDLYYLTM. However, neither Tyr59 nor Arg97 formed any interactions with the docked peptides.

5.2.2 Peptides from the ZIKV NS3 protein

Previous studies have revealed that the arginine at P2 in the co-crystalized peptide SRYWAIRTR (derived from Influenza A virus H1N1) is an anchor for the peptide to accommodate in the MHC I HLA-B*2705 (PDB ID: 2BST) pocket B. It was suggested that the acidic Asp116 from pocket F plays a crucial role in the intaraction with the basic arginine at P9 (Jardetzky, 1991; Madden et al., 1991; Madden et al., 1992). The P9 arginine contacts both Asp77 and Aps116 in the base of the F pocket through a salt bridge. There is also a salt bridge between Glu63, Glu45 and the arginine at P2. The central region of the peptide (P4-P7) is located out of the binding groove.

Evaluation of the docking results for the ZIKV NS3 peptides (DIGAVALDY, HSEVQLLAV) docked at the HLA-B*2705 structure demonstrated a more strong hydrogen bonding energy for DIGAVALDY (-21.251 kcal/mol) than HSEVQLLAV (-16.589 kcal/mol) (Table 3). Six hydrogen bonds formed between the DIGAVALDY peptide and the HLA-B*2705 molecule (Table 2B). The strongest hydrogen bond formed between the backbone oxygen of Asp8 with the side chain nitrogen (NE1)

of Trp147 while the weakest H-bond formed between Asp1 (N) and Tyr59 (OH) which is part of the subpocket A. Pocket B demonstrates a quite strong H-bond interaction between Ile2 (N) and Glu63 (OE1). Tyr9 could form two H-bonds in pocket F with Asp77 (OD1) and Lys146 (NZ).

On the other hand, another predicted epitope HSEVQLLAV from the ZIKV NS3 protein formed eight H-bonds with HLA-B*2705 (Table 2B). Five H-bonds were formed between His1 from the peptide and Glu63, Tyr59, Arg62 and Tyr159 from the subpocket A in the MHC I molecule. One H-bond interaction is between Glu63 from the HLA-B*2705 molecule and Ser2 from the peptide and the other two H-bonds are between Trp147 and Asp77 from the MHC I, and Ala8 and Val9 from the peptide, respectively. Peptide residues 4 to7 did not form any interaction with the MHC I molecule, the rule that is followed by the co-crystalized peptide as well. It should be mentioned that there existed an initial steric clash between Tyr9 (HH atom) from DIGAVALDY and Arg96 (HE atom) from HLA-B*2705, whereas no clash was found in the HSEVQLLAV-HLA-B*2705 complex. No steric clashes were observed in the complexes after the MD simulation.

Table 2B. Hydrogen bonding interactions in HLA-A*2705 (PDB ID: 2BST) complexes.

Peptide Interacting atom	HLA-A*2705 Interacting atom	SRYWAIRTR (native peptide)				DIGAVALDY[a]				HSEVQLLAV[a]			
		D^{b}_{init}		D^{c}_{MD}		D^{b}_{init}		D^{c}_{MD}		D^{b}_{init}		D^{c}_{MD}	
P1(N)	GLU 63 OE1							*	2.727			*	2.706
	TYR 7 OH									*	2.852		
	GLU 63 OE2			*	3.184							*	3.321
	GLU 45 OE2											*	2.781
	TYR 59 OH					*	3.509			*	3.042		
P1(ND1)	ARG 62 NE									*	3.154		
	ARG 62 NH2									*	3.080		
P1(O)	TYR 159 OH	*	2.560	*	2.743	*	2.855			*	2.789		
P2(N)	GLU 63 OE1			*	2.864	*	2.984						
	GLU 63 OE2	*	2.928							*	2.888		
P2(NE)	GLU 45 OE2	*	2.609	*	3.169								
P2(NH1)	THR 24 OG1	*	2.836	*	2.865								
P2(NH2)	THR 24 OG1	*	2.957										
	GLU 45 OE2	*	2.715	*	2.748								
P2(O)	LYS 65 NZ	*	3.187										
	TYR 159 OH											*	2.766
P3(OE1)	LYS 70 NZ											*	3.482
P3(OE2)	LYS 70 NZ											*	2.760
P5(OE1)	GLN 155 NE2											*	2.800
P6(O)	LYS 70 NZ			*	2.747								
P8(O)	TRP 147 NE1	*	2.903							*	2.783		
P8(OD1)	LYS 146 NZ							*	2.878				
P9(N)	ASP 77 OD1	*	2.790	*	2.967	*	2.864			*	2.892	*	3.462
	ASP 116 OD1			*	3.462								
	ASP 116 OD2			*	2.723								
P9(O)	LYS 146 NZ	*	2.487			*	2.889						
	TRP 147 NE1					*	2.762						
P9(NE)	ASP 116 OD1	*	2.652	*	2.862								
	ASP 116 OD2			*	3.135								
P9(OXT)	LYS 146 NZ			*	2.862								
	TYR 84 OH	*	2.537										
Total no of interactions		12		13		6		2		8		8	
Side chain interactions		4		5		_		1		2		4	

[a]Docked ZIKV NS3 peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera v. 1.13.1 (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after molecular dynamics (MD) simulation. Stars denote the existence of an H-bond.

5.2.3 Peptides from the ZIKV NS5 protein

The co-crystallized GILGFVFTL peptide (derived from Influenza A virus) is held in the B and F pockets of the MHC I **HLA-C*0801** peptide-binding groove via the anchor residues at P2 (Ile) and P9 (Leu), respectively (PDB ID: 4NT6). This particular peptide shows very similar interactions in complex with HLA-A*0201 (PDB ID:1OGA). Ser77, Asn80, Lys146 and Phe116 in pocket F of HLA-C*0801 interact with the antigenic peptide while in HLA-A*0201 the corresponding residues are Asp77, Thr80, Lys146 and Tyr116. GILGFVFTL is accommodated in the F pocket through hydrogen bonding with residues 77, 80, 146 and hydrophobic interaction with residue 116 in both HLA structures. Similarly, the B pocket also engages the antigenic peptide in hydrogen bonding via Glu63 and Lys66 and hydrophobic interaction via Met45 in HLA-C*0801 and Tyr7 in HLA-A*0201. It should be mentioned that apart from Tyr7 and Tyr9, residues of the binding-groove floor do not form any notable bonds with the peptides (Choo et al., 2014).

Docking results for the ZIKV NS5 peptides (MTTEDMLVV, FTNLVVQLI) docked into the HLA-C*0801 molecule suggest that Ser77, Asn80 and Lys146 are the significant residues from the MHC I molecule that hold the ZIKV NS5 peptides inside the F pocket (Table 2C). The hydrogen bond energy values resulting from FlexPepDock are -24.163 kcal/mol and -24.269 kcal/mol for MTTEDMLVV and FTNLVVQLI peptides, respectively. One steric clash was found between Thr2 (HG1 atom) of the peptide and Tyr9 (HH atom) of the MHC I molecule (Table 3). The greater number of interactions observed in the F pocket compared to the B pocket suggests that the F pocket plays a more significant role in providing stability to the complex.

Table 2C. Hydrogen bonding interactions in HLA-C*0801 (PDB ID: 4NT6) complexes.

Peptide Interacting atom	HLA-C*0801 Interacting atom	GILGFVFTL (native peptide)				MTTEDMLVV[a]				FTNLVVQLI[a]			
		D^b init		D^c MD		D^b init		D^c MD		D^b init		D^c MD	
P1(N)	GLU 63 OE1			*	2.776							*	3.088
	GLU 63 OE2			*	2.852					*	3.278	*	2.754
	TYR 9 OH	*	2.802			*	2.431						
	TYR 7 OH	*	2.608							*	2.822		
	TYR 171 OH					*	2.851						
	TYR 59 OH			*	2.851								
P1(O)	TYR 159 OH	*	2.658	*	2.713	*	3.009			*	3.294		
	GLU 63 OE1							*	3.129	*	2.882		
	GLU 63 OE2							*	2.722				
P2(O)	LYS 66 NZ	*	2.831										
P2(N)	GLU 63 OE2	*	2.853	*	2.855								
P2(OG1)	GLU 63 O											*	2.896
	TYR 9 OH					*	2.431						
	GLU 63 OE2							*	2.608				
P3(O)	GLN 70 NE2					*	3.380	*	2.946				
P3(N)	TYR 99 OH	*	3.091										
P3(OG1)	TYR 159 OH							*	3.146				
P4(N)	GLN 70 OE1											*	2.890
P5(O)	GLN 70 NE2											*	2.930
P5(OD1)	GLN 155 NE2							*	3.312				
P5(OD2)	GLN 155 NE2							*	2.863				
	GLN 155 NE2							*	2.931				
P7(OE1)	ARG 97 NH1											*	2.995
	ARG 97 NH2											*	2.810
	GLN 70 NE2											*	3.409
P8(O)	TRP147 NE1	*	2.966	*	2.918	*	2.965	*	2.931	*	2.812	*	2.868
	LYS 146 NZ	*	3.091										
P8(OG1)	LYS 146 NZ	*	2.613	*	2.939								
P9(N)	SER 77 OG	*	3.006	*	3.037	*	3.047			*	3.114		
P9(O)	ASN 80 ND2	*	2.937	*	2.958	*	2.943	*	3.127	*	2.820		
	LYS 146 NZ	*	2.863			*	2.845					*	3.116
P9(OXT)	LYS146 NZ	*	2.963					*	2.804			*	3.008
	THR 143 OG1			*	2.662							*	2.654
TOTAL no of interactions		13		10		9		11		7		12	
Side chain interactions		1		1		1		5		-		4	

[a]Docked ZIKV NS5 peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera v. 1.13.1 (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after molecular dynamics (MD) simulation. Stars denote the existence of an H-bond.

Table 3. FlexPepDock energy for the peptide-MHC I complexes after the refinement of the built complexes and the measured steric clashes

	Steric Clash[a]			FlexPepDock		Steric Clash[a]			FlexPepDock	
	Atom1 Atom2	Overlap Å[b]	Distance Å	H-bond energy[c]	Total score[d]	Atom1 Atom2	Overlap Å[b]	Distance Å	H-bond energy[c]	Total score[d]
*1OGA (HLA-A*0201)*	ZIKV E protein peptide: **MAEVRSYCY**					ZIKV E protein peptide: **FSDLYYLTM**				
	GLU 5 H TYR 99 HH	0.994	1.006	-23.01	-531.123	ASP 3 H TYR 99 HH	0.807	1.193	-24.090	-536.234
	GLU 5 N TYR 99 HH	0.659	1.966			-	-	-		
	TYR 9 HH TYR 116 HH	0.637	1.363			-	-	-		
	No steric clashes were found after MD simulation									
*2GTZ (HLA-A*0201)*	-	-	-	-20.623	-538.840	-	-	-	-21.697	-541.334
	No steric clashes were found after MD simulation									
*2GIT (HLA-A*0201)*	-	-	-	-25.101	-524.232	-	-	-	-27.022	-539.109
	No steric clashes were found after MD simulation									
*2GTW (HLA-A*0201)*	More than 20 clashes			-21.090	-559.436				-20.111	-536.430
*3I6G (HLA-A*0201)*	GLU 5 H TYR 99 HH	0.878	1.122	-21.535	-474.067	PHE 1 H2 TYR 7 HH	0.847	1.153	-23.087	-476.302
	GLU 5 N TYR 99 HH	0.667	1.122			PHE 1 N TYR 7 HH	0.806	1.819		
	No steric clashes were found after MD simulation									
*3TO2 (HLA-A*0201)*	TYR 9 HH ARG 97 HH	0.991	1.009	-23.682	-485.236	-	-	-	-21.756	-489.814
	TYR 9 HH ARG 97 NH2	0.680	1.945			-	-	-		
	MET 1 H1 TYR 7 HH	0.641	1.359			-	-	-		
	No steric clashes were found after MD simulation									
*4K7F (HLA-A*0201)*	TYR 9 HH ARG 96 HE	0.747	1.253	-19.069	-524.893	PHE 1 N TYR 6 HH	0.777	1.848	-20.432	-527.281
	No steric clashes were found after MD simulation									
*5SWQ (HLA-A*0201)*	TYR 9 HH ARG 97 HH	0.841	1.253	-23.288	-487.568	PHE 1 H1 TYR 7 HH	0.714	1.286	-24.398	-490.135
	GLU 3 H TYR 99 HH	0.673	1.327			-	-	-		
	MET 1 H1 TYR 7 HH	0.626	1.374			-	-	-		
	TYR 9 HH ARG 97 NH2	0.617	2.008			-	-	-		
	No steric clashes were found after MD simulation									
*2BST (HLA-B*2705)*	ZIKV NS3 protein peptide: **DIGAVALDY**					ZIKV NS3 protein peptide: **HSEVQLLAV**				
	TYR 9 HH ARG 96 HE	0.747	1.253	-21.251	-515.097	-	-	-	-16.589	-521.751
	No steric clashes were after MD simulation									
*4NT6 (HLA-C*0801)*	ZIKV NS5 protein peptide : **MTTEDMLVV**					ZIKV NS5 protein peptide: **FTNLVVQLI**				
	THR 2 HG1 TYR 9 HH	0.704	1.296	-24.163	-536.303	-	-	-	-24.269	-535.095
	No steric clashes were found after MD simulation									
*4NQV (HLA-A*0101)*	ZIKV E protein peptide: **TMNNKHWLV**					ZIKV E protein peptide: **GLDFSDLYY**				
	-	-	-	-26.705	536.634	-	-	-	-27.953	541.860
	No steric clashes were found after MD simulation									

[a]Steric clashes were measured with Chimera; [b]Allowed overlap is 0.6 Å; H-bond overlap reduction is 0.4 Å; [c]H-bond energy is for the side chain-side chain interaction (kcal/mol); [d]Total score is Rosetta energy.

5.3 Evaluation of the molecular dynamics simulation trajectories

MD simulations of the peptide-MHC I complexes were performed in explicit water at 300 K for a period of 10 ns to investigate the stability of the docked peptide-MHC I protein complexes and comparing the result with the behavior of experimentally verified antigenic peptides (reference peptides). In general, the potential energy of all simulation systems remained stable during MD simulation. Several parameters were measured before, during and after the MD simulation to analyze the quality and strength of ZIKV peptide binding to the studied MHC I proteins. These parameters included hydrogen bonding network (see Tables 2A-C, Appendices 2A-H), RMSD of the MHC I peptide-binding domain as well as the peptide molecule with respect to the starting structure, atomic fluctuations and solvent accessibility of the peptide residues, peptide interaction energies (free energy of binding) and the size of the F pocket both in the initial complexes as well as the final structures from the MD simulations. Here the aim was to find the most stable peptide-MHC I complex in order to predict the ZIKV peptide or peptides with the highest binding affinity to the MHC I binding groove. Such high-affinity peptides could be suitable candidates for design of a vaccine against the ZIKV infection.

5.3.1 HLA-A*0101 in complex with ZIKV E protein peptides

In this section, we will review the analysis of the following MHC I complexes with the predicted antigenic peptides from the ZIKV E protein and the experimental peptide from the respective MHC I crystal structure:

HLA-A*0101 (PDB ID: 4NQV) in complex with ZIKV peptides TMNNKHWLV and GLDFSDLYY and the co-crystalized peptide CTELKLSDY.

Inspection of the hydrogen bonding network of the predicted ZIKV E epitopes TMNNKHWLV and GLDFSDLYY complexed with HLA-A*0101 demonstrated that peptide GLDFSDLYY forms a larger number of H-bond interactions in the binding groove of the MHC I molecule than peptide TMNNKHWLV. However, the number of H-bond interactions with TMNNKHWLV increased during

the simulation (Table 2A). Free energy of binding, calculated by the Prime-MMGBSA approach, decreased (improved) for the GLDFSDLYY-HLA-A*0101 complex as well as the crystal complex CTELKLSDY-HLA-A*0101 after the MD simulation compared to the initial values while this value increased (worsened) for the TMNNKHWLV- HLA-A*0101 complex. Also, the F pocket enlarged in both docked peptide-MHC I complexes (Table 4, Appendix 3: Figure 11.1).

RMSD of the peptide-binding groove and the bound peptides was used as another criterion to inspect the stability of the peptide-MHC I complexes. RMSD of the backbone atoms of the HLA-A*0101 binding groove was about 1.6-1.7 Å in both docked peptide-MHC I complexes (Figure 7.1). RMSD of the backbone atoms of the GLDFSDLYY peptide bound to HLA-A*0101 was about 1.2 Å, whereas peptide TMNNKHWLV had a higher value of about 2 Å (Figure 8.1).

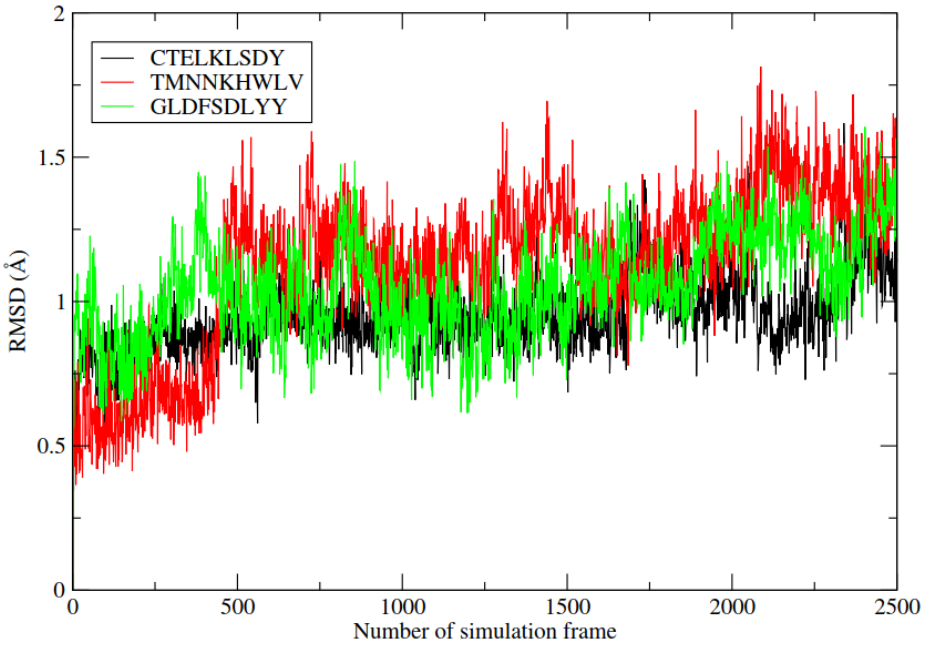


Figure 7.1. RMSD of the backbone atoms of the MHC I binding groove (residues 1-180 in PDB ID: 4NQV) during the 10-ns MD simulation of the ZIKV E peptide–HLA-A*0101 complexes (green and red lines) and the reference complex (black line).

Atomic fluctuations are a significant indicator of the strength of a ligand's binding that depends on hydrophobic and electrostatic interactions. High atomic fluctuations indicate loose binding and vice versa (Höltje et al., 2003). Therefore, in this study atomic fluctuations were computed for each docked complex to be compared with the fluctuations of the experimental complexes. While TMNNKHWLV demonstrates less atomic fluctuations inside the binding groove

in the N-terminal segment, GLDFSDLYY shows less fluctuation in its C-terminus (Figure 9.1).

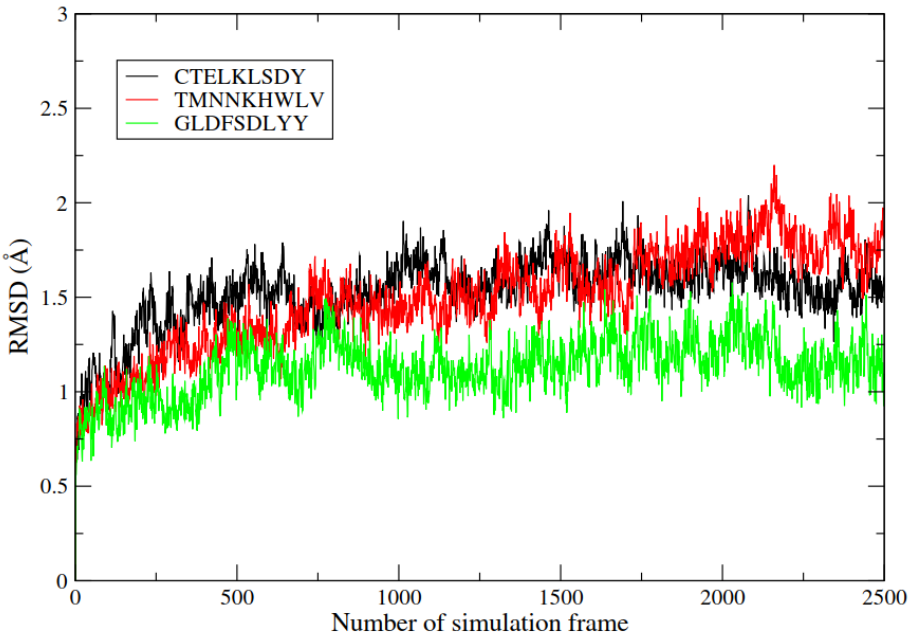


Figure 8.1. RMSD of the backbone atoms of the ZIKV E peptides (green and red lines) and the reference peptide (black line) bound to HLA-A*0101 (PDB ID: 4NQV) during the 10-ns MD simulation.

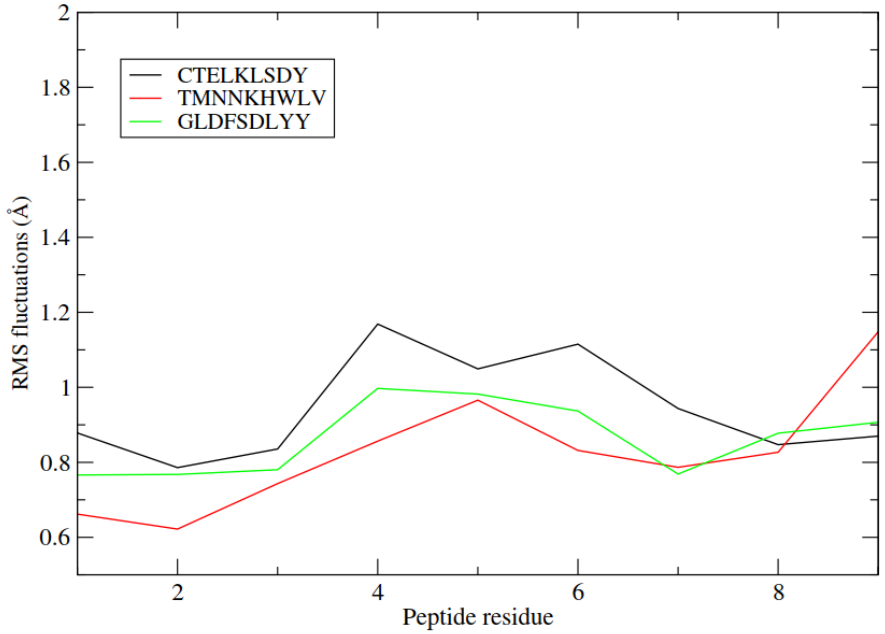


Figure 9.1. Per-residue RMS fluctuations of the ZIKV E protein peptides (green and red lines) and the reference peptide (black line) in the binding groove of MHC I HLA-A*0101 (PDB ID: 4NQV) during the 10-ns MD simulations.

SASA has also been known as an excellent measurement to evaluate the affinity of a peptide and it may be used as the decisive criteria to select the good binders from various peptides (Höltje et al., 2003). As it can be seen from the plot in Figure 10.1, GLDFSDLYY mimics the behavior of the co-crystallized peptide especially at positions P2 and P9 (anchor residues).

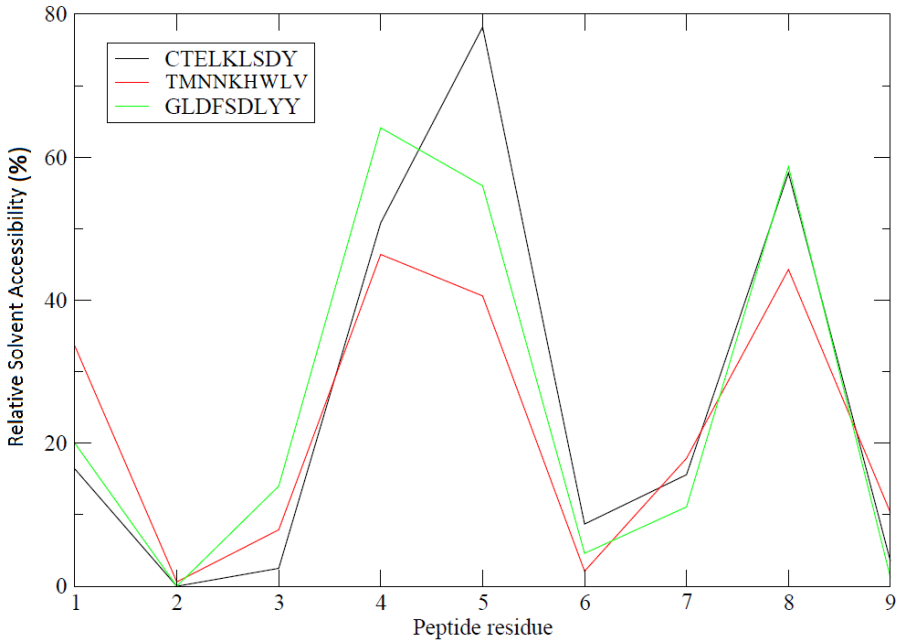


Figure 10.1. Per-residue relative solvent accessibility of the ZIKV E protein peptides (green and red lines) and the reference peptide (black line) in the binding groove of MHC I HLA-A*0101 (PDB ID: 4NQV) after the MD simulations.

Based on the above analysis, it can be concluded that GLDFSDLYY seems to form a more stable complex with the MHC I HLA-A*0101 than TMNNKHWLV. This suggests that it would have a better binding affinity to the MCH I molecule, thus being able to stimulate a strong immune response.

5.3.2 HLA-A*0201 in complex with ZIKV E protein peptides

In this section, we will review the analysis of the following MHC I complexes with the predicted antigenic peptides from the ZIKV E protein and the experimental peptides from the respective MHC I crystal structures:

HLA-A*0201 (PDB IDs: 1OGA, 2GIT, 2GTW, 2GTZ, 3I6G, 3TO2, 4K7F, 5SWQ) in complex with the ZIKV peptides MAEVRSYCY and FSDLYYLTM as well as the co-crystalized peptides GILGFVFTL, LLFGKPVYV, LAGIGILTV, ALGIGILTV, GLMWLSYFV, LACFVLAAV, VCWGELMNL, CVNGSCFTV, respectively.

PDB ID: 1OGA. MD analysis of the HLA-A*0201 molecule (PDB ID: 1OGA) in complex with the ZIKV E peptides MAEVRSYCY and FSDLYYLTM demonstrated that RMSD of the backbone atoms of the binding groove (residues 1-180) was about 1 Å in both the MAEVRSYCY-MHC I complex and the crystal complex (GILGFVFTL-MHC I) (Figure 7.2A). However, RMSD of the peptide-binding groove in the FSDLYYLTM-MHC I complex was somewhat larger (about 1.5 Å). On the other hand, in this crystal structure the co-crystalized peptide GILGFVFTL and the docked peptide FSDLYYLTM were more stable with a RMSD value of about 1.5 Å compared to the docked peptide MAEVRSYCY with a RMSD of about 2 Å (Figure 8.2A). The F pocket size was also enlarged in all the peptide-HLA-A*0201 complexes, including the crystal complex (Table 4 and Appendix 3, Figure 11.2). The Prime-MMGBSA energy was lower (i.e. better) for the FSDLYYLTM-HLA-A*0201 as well as the crystal complex GILGFVFTL- HLA-A*0201 after the MD simulation compared to the initial values while this value rose (worsened) from -91.53 to -68.74 kcal/mol for the docked peptide MAEVRSYCY (Table 4). Besides, the RMS fluctuations of residues 2 and 9 of the both docked peptides in the binding groove were about 1 Å, which is comparable with the corresponding fluctuations of the co-crystalized peptide GILGFVFTL (about 0.7 Å) (Figure 9.2A). Moreover, the relative solvent accessibility (RSA) of residues 2 (P2) and 9 (P9) was about 20 percent for the two docked peptides showing that the residues in N-terminus (P2) and C-terminus (P9) are deeply buried in the binding groove (B and F subpockets) of the HLA molecule. The same trend was also observed for the co-crystalized reference peptide (Figure 10.2A). Inspecting the atomic clashes in the modeled complexes showed that three clashes existed between the atoms of peptide MAEVRSYCY and the MHC I binding groove, whereas no clash was found between the atoms of peptide FSDLYYLTM and the MHC I binding groove (Table 3). It is probable that the existence of such clashes affects the flexibility of MAEVRSYCY peptide inside the MHC cleft I, which then resulted in the low value of RMSD in the binding groove. Moreover, the evaluation of hydrogen

bonds before and after MD simulation revealed that MAEVRSYCY has retained three of the initial H-bonds in the binding groove after the MD simulation within 2.7 Å to 3.3 Å. One persistent interaction was between Arg97 and the glutamic acid at P3 of the peptide while methionine at P1 lost some of its initial H-bonds (Appedix 2A). No hydrogen bonds were formed at P2 (anchor residue) but Tyr9 at P9 formed two H-bonds via its backbone with Asp77 (OD2) and Lys146 (NZ). Peptide FSDLYYLTM exhibited more H-bonds inside the MHC I groove than MAEVRSYCY after the MD simulation. Additionally, as the binding of the peptide in the MHC I binding groove depends on the characteristics of the groove residues (Höltje et al., 2003), the larger number of aromatic residues in the FSDLYYLTM peptide than in MAEVRSYCY could improve the interaction between the peptide and the binding groove.

PDB ID: 2GIT. Assessing the MD simulation of the HLA-A*0201 (PDB ID: 2GIT) complexed with the ZIKV E peptides MAEVRSYCY and FSDLYYLTM showed that RMSD of the backbone atoms of the binding groove was about the same for the two docked complexes: near 1.2 Å (Figure 7.2B). RMSD of the crystal complex (LLFGKPVYV-HLA-A*0201) was slightly higher (over 1.5 Å) for the MHC I peptide-binding domain. However, among all the three peptides the co-crystalized peptide LLFGKPVYV was the most stable in the binding groove with the RMSD value of only 0.5 Å. The second most stable peptide was the docked peptide FSDLYYLTM with the RMSD value of 1 Å. The docked peptide MAEVRSYCY had the highest RMSD value (about 2 Å), meaning that this peptide had the least stable pose in the binding groove of the studied MHC I structure (Figure 8.2B). Prime-MMGBSA free energies of binding were lower (better) than initially for all the complexes after the MD simulations (Table 4). The F pocket that binds the C-terminal area of the nonapeptides enlarged in size during the MD simulations in all the complexes (Table 4 and Appendix 3, Figure 11.3). Moreover, RMSF of the residue 2 (P2) and 9 (P9) of the docked peptide FSDLYYLTM in the binding groove is also comparable with the co-crystalized peptide (LLFGKPVYV), about 0.7 Å, which is significantly less than that of peptide MAEVRSYCY (about 1.7 Å) (Figure 9.2B). This suggests that peptide FSDLYYLTM is able to bind more tightly to the binding groove of also this HLA-A*0201 structure. In addition, RSA of both the anchor residues (P2 and P9) in all complexes is below 20 percent and even less for FSDLYYLTM and the co-crystalized peptide (Figure

10.2B).

PDB ID: 2GTW. MD analysis of the HLA-A*0201 (PDB ID: 2GTW) in complex with the ZIKV E peptides MAEVRSYCY and FSDLYYLTM showed again (see the 2GIT structure above) that RMSD of the backbone atoms of the binding groove in both the docked peptides adopted a smaller value of about 1.2-1.5 Å than the crystal complex (LAGIGILTV-HLA-A*0201) that adopted a value of about 1.7 Å (Figure 7.2C). Moreover, the co-crystalized peptide LAGIGILTV and the docked peptide FSDLYYLTM had a higher RMSD value of about 1 Å compared to the docked peptide MAEVRSYCY with a RMSD of about 0.5 Å (Figure 8.2C). Although RMSD for the backbone atoms of MAEVRSYCY is the smallest in this MHC I structure among all the HLA-A*0201 complexes (Figure 8.2), there was a huge increase (worsening) in Prime-MMGBSA energy of this complex (Table 4). It can be assumed that a large number of clashes between the peptide side chain atoms and the MHC I binding groove atoms prevented the peptide from moving freely in the binding cleft to obtain stability (Table 3). Another reason could be attributed to the short time of the MD simulation in which the complex could not reach the desired stability. From the docking evaluation, it was observed that the FSDLYYLTM-HLA-A*0201 complex retained more H-bonds than the MAEVRSYCY-HLA-A*0201 (Appendix 2C). On the other hand, the Prime-MMGBSA free energy of binding for the co-crystalized peptide LAGIGILTV and FSDLYYLTM was improved during the MD simulation (Table 4). The F pocket was also enlarged in the two docked peptide-MHC I complexes and the crystal complex (Table 4 and Appendix 3, Figure 11.4). Besides, the atomic fluctuation of residues 2 (P2) and 9 (P9) of MAEVRSYCY in the binding groove is higher than that of FSDLYYLTM and the co-crystalized peptide (Figure 9.2C). Further, the relative accessibility of the peptide residues 2 and 9 to the solvent was below 10 percent for the co-crystalized and FSDLYYLTM peptides. Thus, these residues in N-terminus (P2) and C-terminus (P9) are deeply buried in the B and F subpockets of the HLA molecule, respectively, which suggests stable binding to the MHC I molecule (Figure 10.2C).

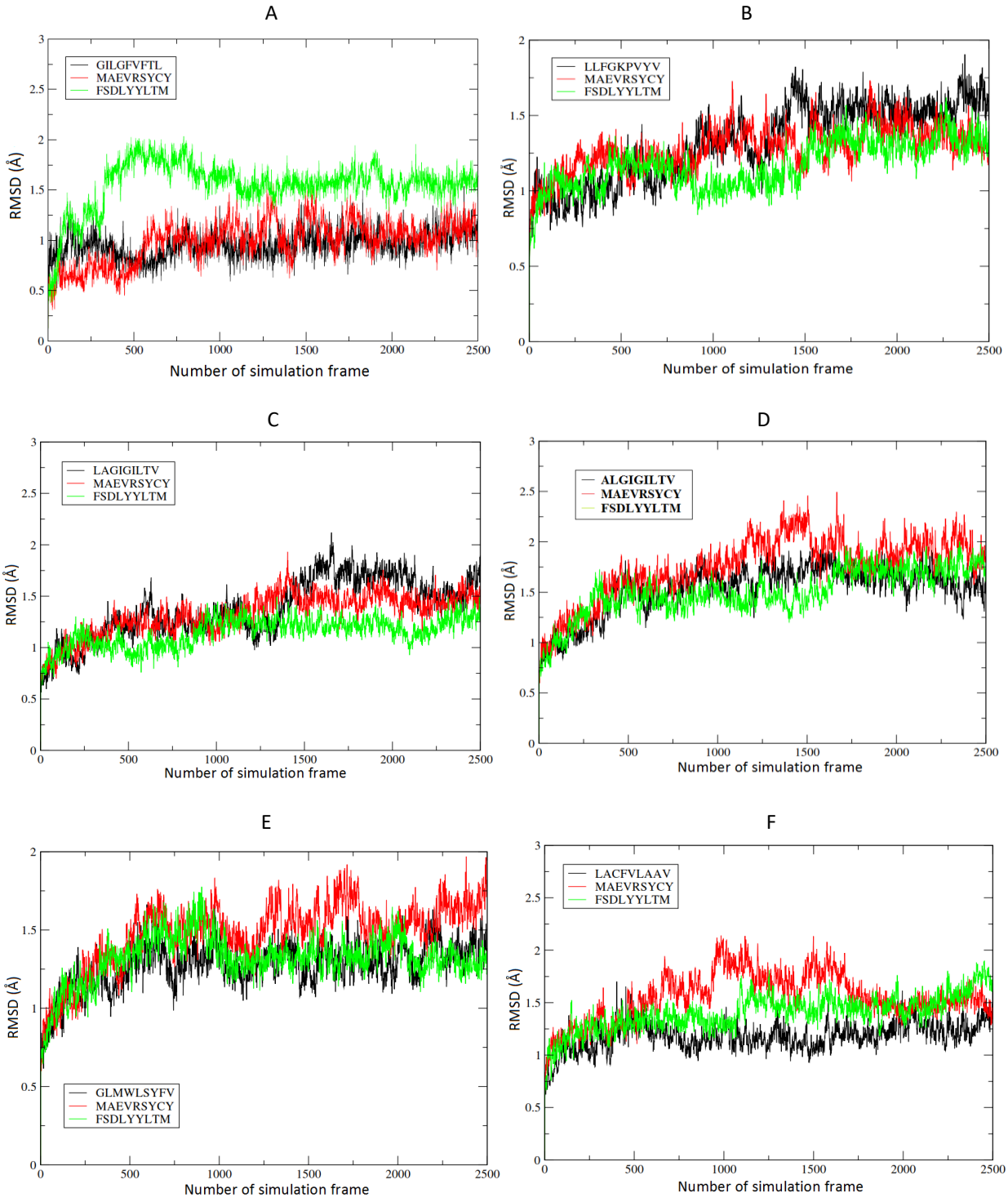
PDB ID: 2GTZ. MD analysis of the HLA-A*0201 (PDB ID: 2GTZ) complexed with the ZIKV E peptides MAEVRSYCY and FSDLYYLTM revealed that RMSD of the backbone atoms of the binding groove is

about the same for the two docked complexes (near 1.7 Å) (Figure 7.2D). However, RMSD for the MHC I peptide-binding domain of the crystal complex (ALGIGILTV-HLA-A*0201) was lower (below 1.5 Å). In this structure, the co-crystalized peptide ALGIGILTV and the docked peptide MAEVRSYCY were more stable with a RMSD value of about 1.5 Å compared to the docked peptide FSDLYYLTM. Similar to the crystal complex, Prime-MMGBSA free energy of binding was lower (better) in both the docked complexes (Table 4) and the F pocket was enlarged after the MD simulations (Appendix 3, Figure 11.5). Besides, RMSF of residues 2 (Ser) and 9 (Met) of the docked peptide FSDLYYLTM in the binding groove is also comparable with the co-crystalized peptide (ALGIGILTV) and less than that of peptide MAEVRSYCY (Figure 9.2D). Moreover, RSA of residue 9 in FSDLYYLTM and the crystal peptide was also less than that of Tyr9 in MAEVRSYCY. However, P2 (Ser) is more solvent accessible in both docked peptides, suggesting a lower affinity of the peptide N-terminus than the C-terminus (Figure 10.2D).

PDB ID: 3I6G. When it comes to the the HLA-A*0201 structure in complex with GLMWLSYFV peptide (human SARS coronavirus, PDB ID: 3I6G), the number of hydrogen bond interactions between the docked peptide and the MHC I binding groove in the two peptide-HLA complexes (MAEVRSYCY-HLA-A*0201 and, FSDLYYLTM-HLA-A*0201) increased during the MD simulations (Appendix 2E). The largest number of interactions were formed between Glu3 from MAEVRSYCY peptide and Arg97 and His114 from the HLA molecule. Trp147 and Asp77 of the MHC I seem to be important interaction partners since they retained their hydrogen bonds in the docked complexes as well as the experimental complex throughout the MD simulations. RMSD of the backbone atoms of the HLA-A*0201 binding groove for the complex with the docked peptide FSDLYYLTM had approximately the same value as the complex with the co-crystalised peptide GLMWLSYFV (about 1.5 Å), whereas it was larger (about 2 Å) for the complex with MAEVRSYCY (Figure 7.2E). On the other hand, RMSD of the backbone atoms for the two predicted peptides was approximately the same (1.5 Å) as it was for the co-crystallised peptide (GLMWLSYFV) (Figure 8.2E). Prime-MMGBSA free energy of binding worsened for the MAEVRSYCY-HLA-A*0201 complex from -96.3 to -77.7 kcal/mol during the simulation. By contrast, Prime-MMGBSA free energy of binding for the co-crystalized peptide GLMWLSYFV and FSDLYYLTM improved during

the MD simulation (Table 4). The F pocket size of all the complexes enlarged during the MD simulations (Table 4 and Appendix 3, Figure 11.6). Besides, RMS fluctuation of residues 2 and 9 of MAEVRSYCY in the binding groove is comparable with the co-crystalized peptide and less than that of FSDLYYLTM. The atomic fluctuations of FSDLYYLTM are at lowest at P3 and P8 (Figure 9.2E). On the other hand, residues 2 and 9 of FSDLYYLTM have less accessibility to the solvent than the same residues of peptide MAEVRSYCY and follow the same trend as co-crystalized peptide (Figure 10.2E). Consistent with the above analysis, peptide FSDLYYLTM from the ZIKV E protein seems to be a stronger binder to the HLA-A*0201 molecule than MAEVRSYCY.

PDB ID: 3TO2. RMSD of the binding groove and the backbone atoms of peptide FSDLYYLTM in the HLA-A*0201 structure (PDB ID: 3TO2) were somewhat higher when compared to the MAEVRSYCY-HLA-A*0201 complex (Figures 7.2F and 8.2F). The two predicted peptides improved their Prime-MMGBSA free energy of binding during the MD simulations (Table 4). On the other hand, low RSA and RMSF of the N-terminal residues (P2-P3) of MAEVRSYCY suggest that only the N-terminus of the peptide binds strongly to this MHC I structure (Figures 9.2F and 10.2F and Appendix 3, Figure 11.7). FSDLYYLTM has a somewhat higher RMSF at P2 but the C-terminal P9 is fluctuating less than that of MAEVRSYCY; the anchor residues P2 and P9 of FSDLYYLTM are also well buried. However, P3 of FSDLYYLTM is not so buried as the corresponding residues in the co-crystalized peptide LACFVLAAV and MAEVRSYCY. As discussed earlier, another measured parameter to evaluate the binding affinity of the peptide to the binding cleft is the size of the F pocket that accommodates the C-terminus of the nonapeptide. In most complexes, including the experimentally determined structures, the size of the F pocket increased somewhat. The MAEVRSYCY-HLA-A*0201 complex (PDB ID: 3TO2) was the only one that showed a decrease in the F pocket size (Appendix 3, Figure 11.7A) and this seems to result from the fact that the C-terminus of the peptide was not properly bound in the pocket.


A
GILGFVFTL
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame
B
LLFGKPVYV
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame
C
LAGIGILTV
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame
D
ALGIGILTV
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame
E
GLMWLSYFV
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame
F
LACFVLAAV
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame

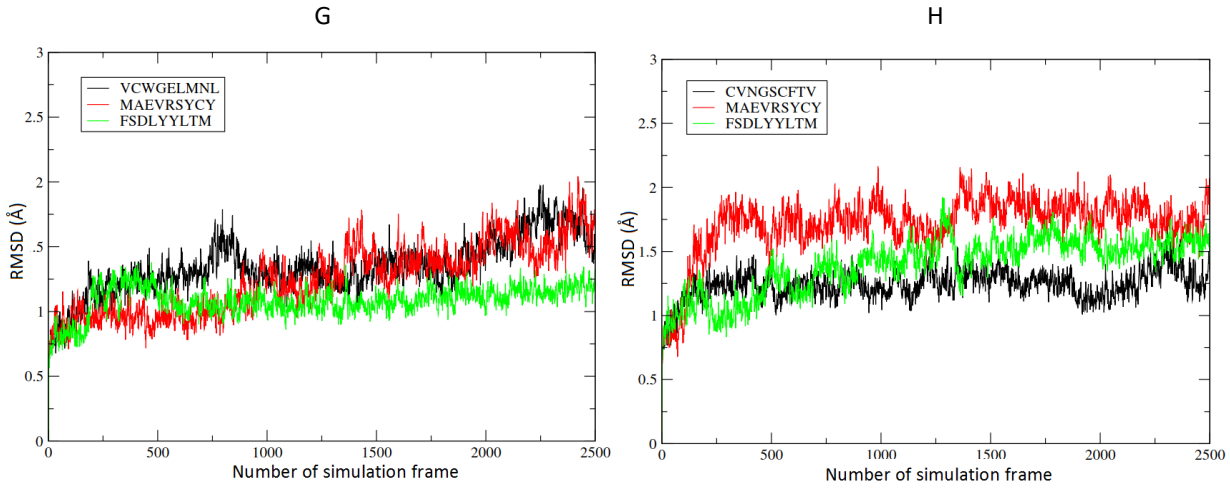
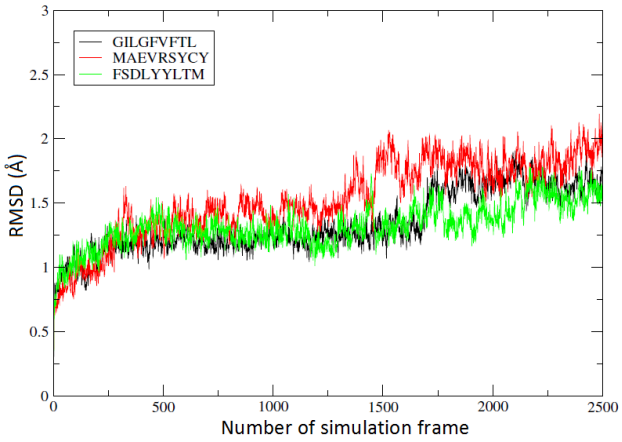


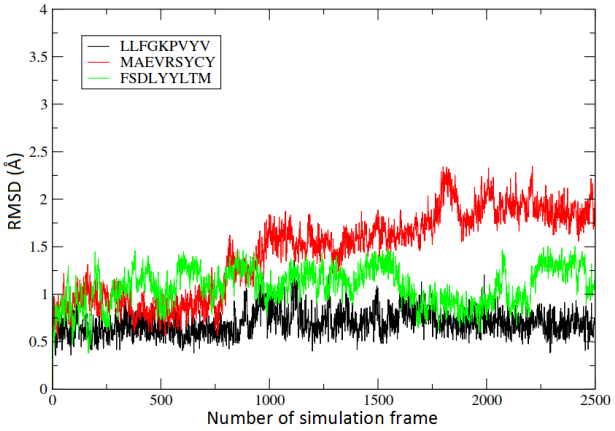
Figure 7.2. RMSD of the backbone atoms of the MHC I binding groove (residues 1-180) during the 10-ns MD simulations of the HLA-A*0201–ZIKV E protein peptide complexes. The ZIKV peptides were docked to the crystal structures: A. 1OGA, B. 2GIT, C. 2GTW, D. 2GTZ, E. 3I6G, F. 3TO2, G. 4K7F, and H. 5SWQ. The black line denotes the reference peptides bound in the respective crystal structures.

PDB ID: 4K7F. From Figure 7.2G, it can be seen that the HLA-A*0201 (PDB ID: 4K7F) in complex with ZIKV E peptide FSDLYYLTM adopted a RMSD value of about 1.2 Å for the backbone atoms of the binding groove. The same domain of the crystal complex (VCWGELMNL-HLA-A*0201) and the complex with the docked peptide MAEVRSYCY showed a RMSD of about 1.5 Å. On the other hand, it can be observed from Figure 8.2G that the lowest peptide RMSD belongs to the co-crystalized peptide VCWGELMNL (about 1 Å). This experimental peptide is thus considered as the most stable peptide in the binding groove of this HLA-A*0201 structure, which is in line with the low Prime-MMGBSA energy of the peptide after the MD simulation. On the contrary, the Prime-MMGBSA energy for the predicted peptide MAEVRSYCY rose (worsened) from -106.5 to -96.04 kcal/mol during the simulation while it again decreased (improved) for peptide FSDLYYLTM (Table 4). The F pocket size also enlarged in all the complexes (Table 4; Appendix 3, Figure 11.8). Besides, the atomic fluctuations of residues 2 and 9 of FSDLYYLTM and MAEVRSYCY are comparable with the co-crystalized peptide. However the fluctuation of residue 9 is higher than that of the co-crystalized peptide (Figure 9.2G). The relative accessibility of residues 2 and 9 to the solvent was below 10 percent for the co-crystalized peptide and FSDLYYLTM. This means that the N-terminus (P2) and C-terminus (P9) are deeply buried in the B and F subpockets of the HLA molecule, suggesting high binding affinity to the MHC I groove (Figure 10.2G).

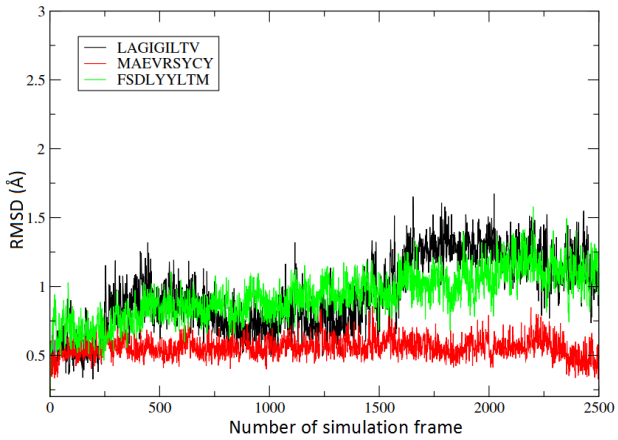
A


GILGFVFTL
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame

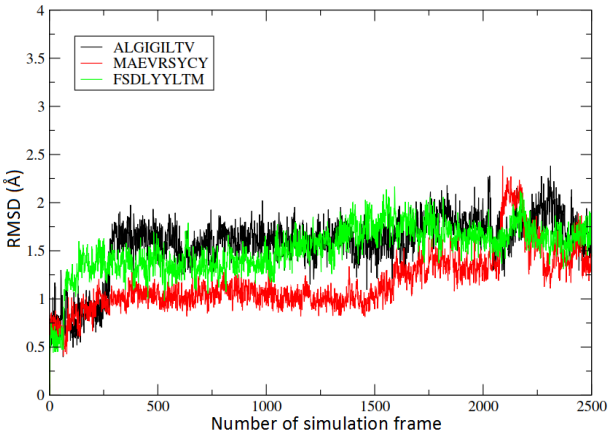
B


LLFGKPVYV
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame

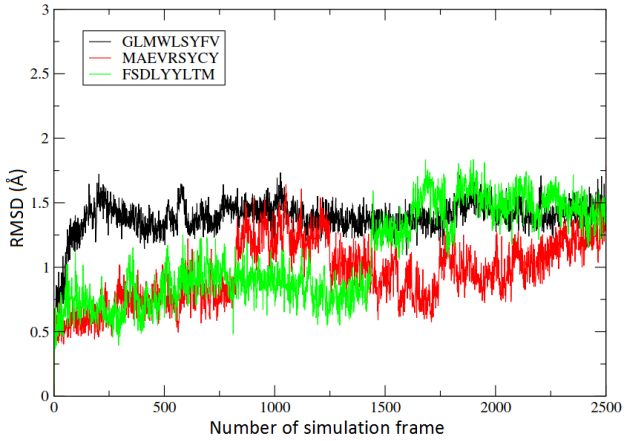
C


LAGIGILTV
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame

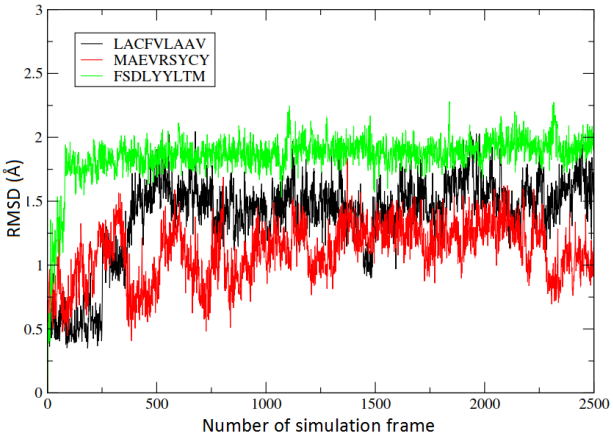
D


ALGIGILTV
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame

E


GLMWLSYFV
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame

F


LACFVLAAV
MAEVRSYCY
FSDLYYLTM
RMSD (Å)
Number of simulation frame

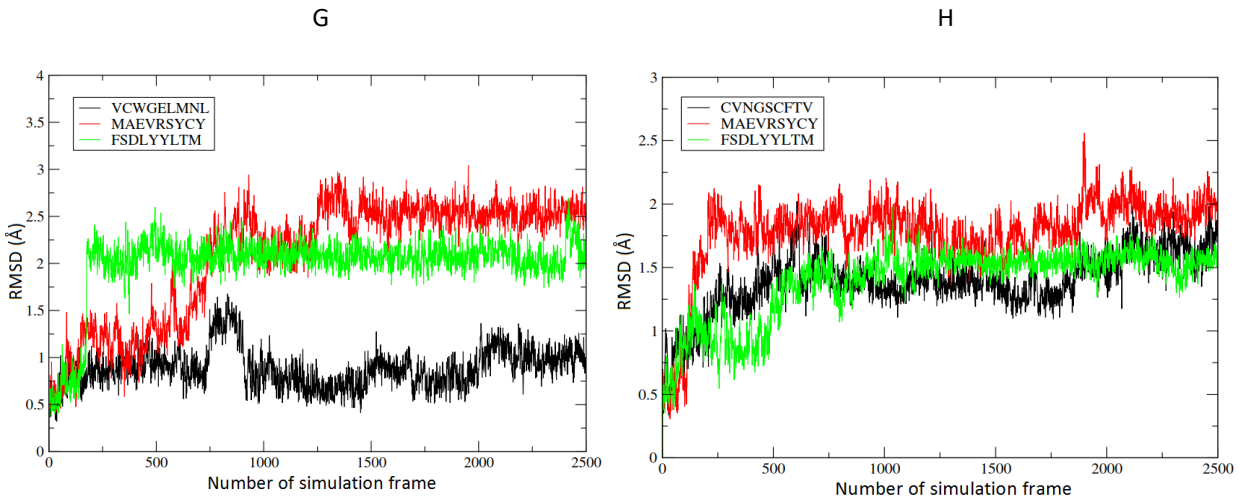
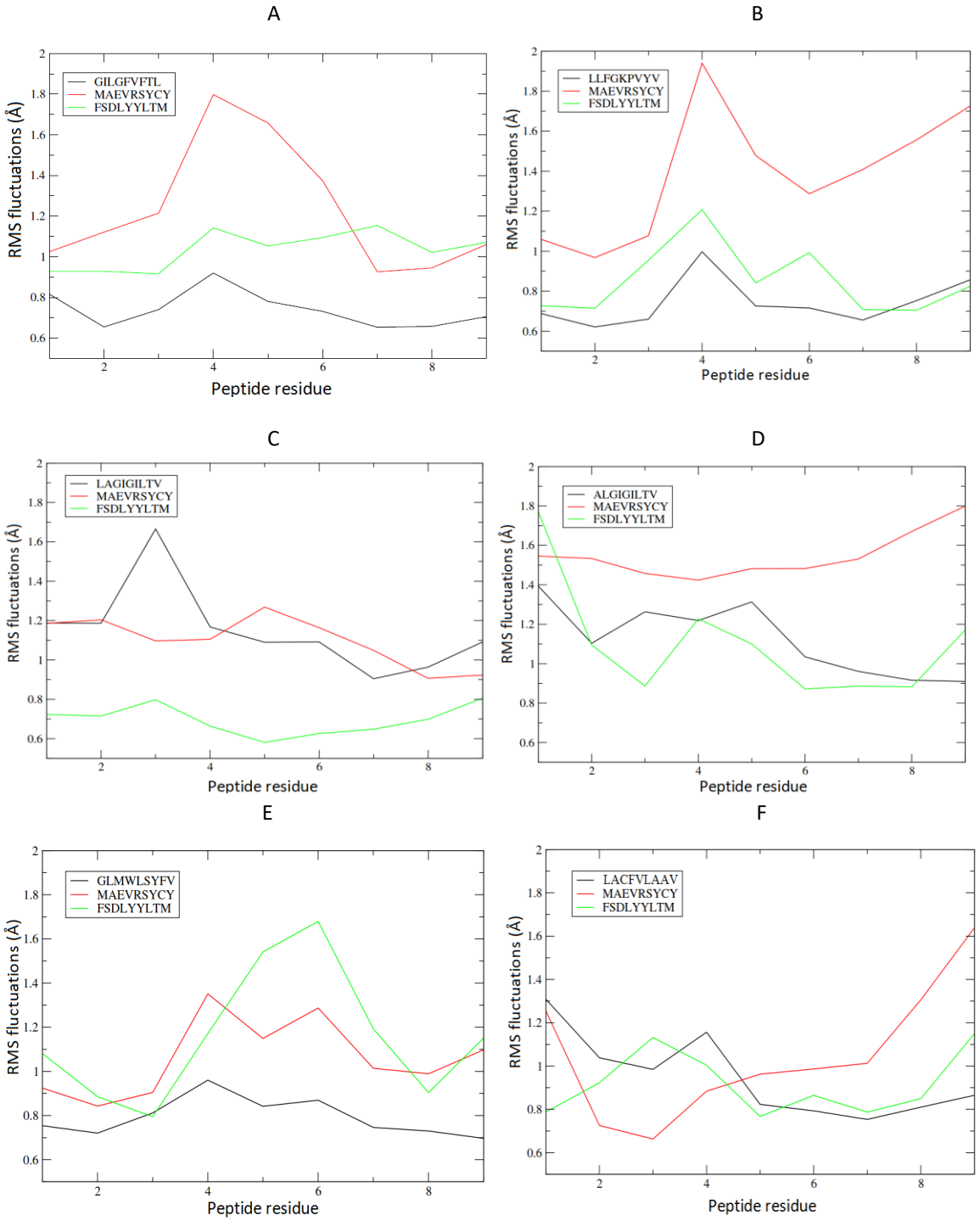


Figure 8.2A. RMSD of the backbone atoms of the peptides bound to HLA-A*0201 during the 10-ns MD simulations of the HLA-A*0201–ZIKV E protein peptide complexes The ZIKV peptides were docked to the crystal structures: A. 1OGA, B. 2GIT, C. 2GTW, D. 2GTZ, E. 3I6G, F. 3TO2, G. 4K7F, and H. 5SWQ. The black line denotes the reference peptides bound in the respective crystal structures.

PDB ID: 5SWQ. MD analysis of the HLA-A*0201 molecule (PDB ID: 5SWQ) complexed with the docked peptides MAEVRSYCY and FSDLYYLTM revealed that there is an increase in the number of H-bond interactions in the binding groove after the MD simulation (Appendix 2H). The extra H-bonds occurred between Cys1 (N), Asn3 (N) and Val9 (N) of the peptide with Glu63 (OE1, OE2), Tyr99 (OH) and Asp77 (OD) of the MHC I molecule. All the newly formed H-bonds are between the peptide backbone and the MHC I protein. Tracing the H-bonds for both the docked peptides revealed that FSDLYYLTM forms additional H-bonds in the N- and C-terminus at P2 and P9 while MAEVRSYCY forms additional H-bonds with residues at P5, P6 and P7. Based on the fact that N- and C-terminus are the anchor locations of the peptides when they interact with the MHC I binding groove, FSDLYYLTM seems to act as a stronger binder than MAEVRSYCY when regarding only the H-bond interactions. RMSD of the backbone atoms of the MHC I binding groove for the experimental complex was below 1.5 Å, for FSDLYYLTM about 1.5 Å and for MAEVRSYCY about 2.0 Å (Figure 7.2H). In addition, RMSD of the peptides was about 1.5 Å for FSDLYYLTM and the crystal peptide CVNGSCFTV and 1.7 Å for MAEVRSYCY (Figure 8.2H). The seeming stability of the FSDLYYLTM-MHC I complex is also reflected in the value of Prime-MMGBSA energy that decreased from -61.87 kcal/mol for the initial docked complex to -101 kcal/mol for the conformation after

the MD simulation (Table 4). RMSF of the residues at P9 of FSDLYYLTM and MAEVRSYCY was 1.1 Å and 1.7 Å, respectively (Figure 9.2H). However, at P2 both peptides showed a value similar to that of the co-crystalized peptide, about 1.0 Å. Thus, FSDLYYLTM has less fluctuation at the anchor positions P2 and P9, which results in a higher affinity to the B and F pockets, respectively.



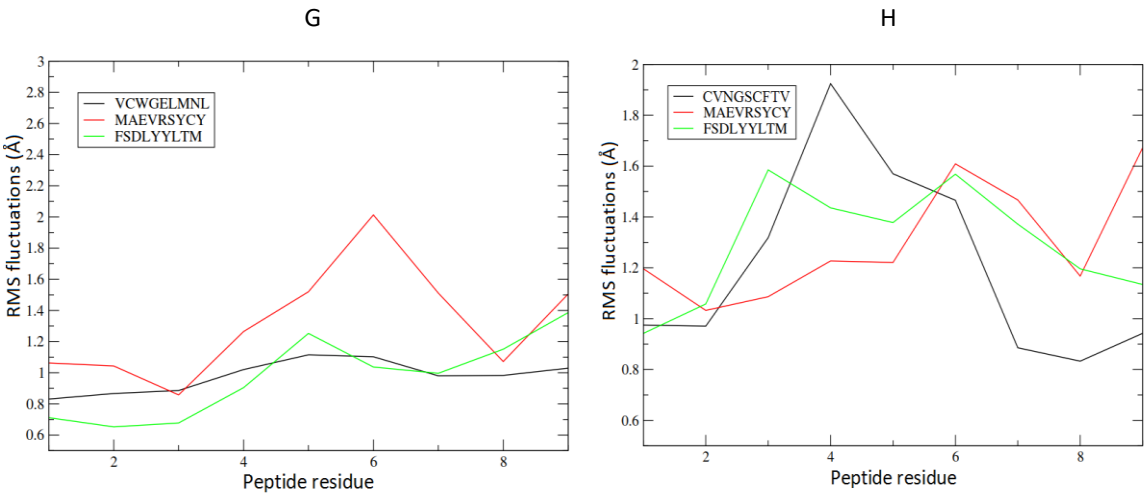
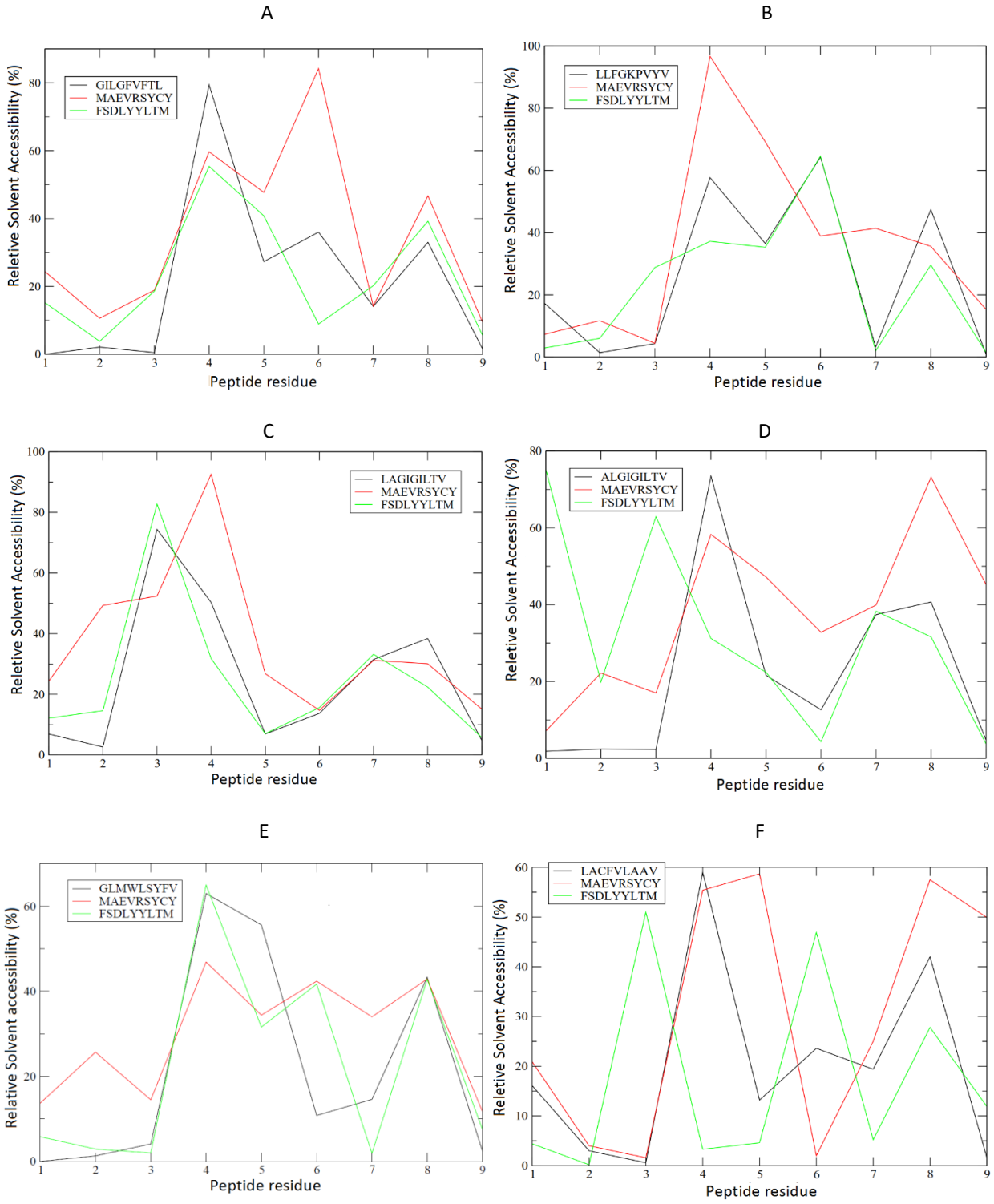


Figure 9.2. Per-residue RMS fluctuations of the ZIKV E protein peptides in the binding groove of MHC I HLA-A*0201 during the 10-ns MD simulations. The black line denotes the reference peptide. PDB IDs: A. 1OGA, B. 2GIT, C. 2GTW, D. 2GTZ. E. 3I6G, F. 3TO2, G. 4K7F, H. 5SWQ.

As mentioned earlier, residues at P2 and P9 are the most important sites for binding to the MHC I molecule. In contrast, the other residues (at P4 to P7) are generally more exposed to the solvent and interpreted as inactive and low-affinity binders to the MHC I molecule but high-affinity and active binders to the T-cell receptors (Höltje et al., 2003). It should be noted that P5 and P7 (Chapter 2, Figure 3) could bind to both MHC I and TCR. Referring to the peptides FSDLYYLTM and MAEVRSYCY in complex with this particular HLA-A*0201 structure, it can be observed that the residues at P2, P7 and P9 were completely buried inside the MHC I binding groove (less than 20 percent of solvent accessibility area) (Figure 10.2H), thus being important for the binding interaction. By contrast, the TCR binding residues show a high percentage of solvent accessibility (more than 50 percent) for their side chains at P3, P6 and P8. All in all, the above analysis of the different HLA-A*0201 crystal structures suggests that FSDLYYLTM would be a stronger binding peptide to MHC I HLA-A*0201 than MAEVRSYCY.


A
GILGFVFTL
MAEVRSYCY
FSDLYYLTM
Reletive Solvent Accessibility (%)
Peptide residue
B
LLFGKPVYV
MAEVRSYCY
FSDLYYLTM
Reletive Solvent Accessibility (%)
Peptide residue
C
LAGIGILTV
MAEVRSYCY
FSDLYYLTM
Reletive Solvent Accessibility (%)
Peptide residue
D
ALGIGILTV
MAEVRSYCY
FSDLYYLTM
Reletive Solvent Accessibility (%)
Peptide residue
E
GLMWLSYFV
MAEVRSYCY
FSDLYYLTM
Relative Solvent accessibility (%)
Peptide residue
F
LACFVLAAV
MAEVRSYCY
FSDLYYLTM
Reletive Solvent Accessibility (%)
Peptide residue

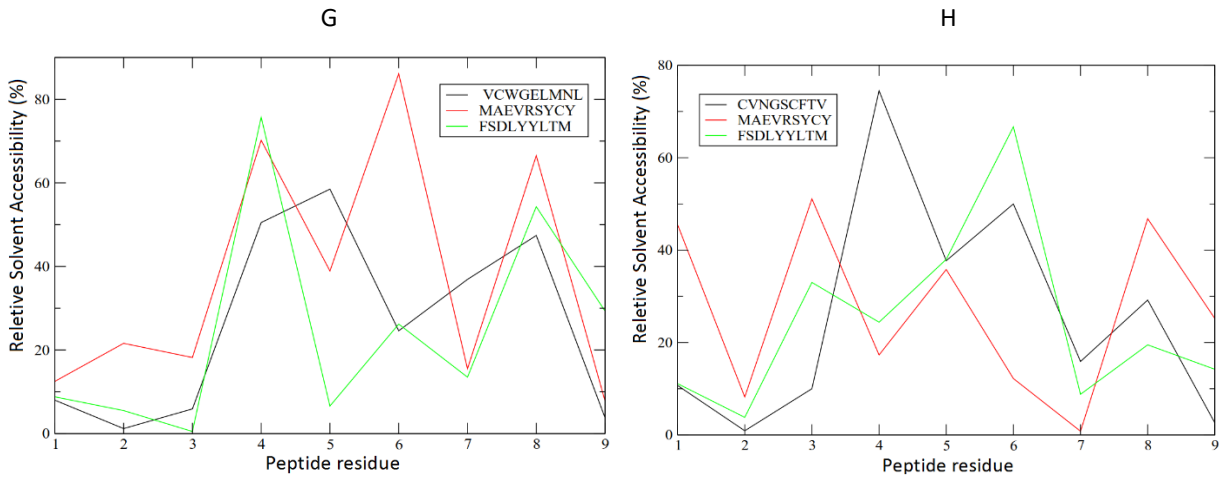


Figure 10.2. Per-residue relative solvent accessibility of the ZIKV E protein peptides in the binding groove of MHC I HLA-A*0201 after the MD simulations. The black line denotes the co-crystalized peptides. PDB IDs: A. 1OGA, B. 2GIT, C. 2GTW, D. 2GTZ, E.3I6G, F. 3TO2, G. 4K7F, H. 5SWQ.

Overall, analysing the ZIKV E peptide-HLA-A*0201 complexes and comparing the results with the experimentally determined structures of HLA-A*0201, demonstrated that RMSD of the peptide-binding domain was generally about 1.5 Å in the experimentally determined complexes while this value was about 2 Å in the docked peptide-MHC I complexes (Figures 7.2A-H). Yet, RMSD of the binding groove in the HLA-A*0201 structures complexed with FSDLYYLTM was usually smaller than in the structures complexed with MAEVRSYCY. There were only two exceptions: MAEVRSYCY complexes with PDB ID: 1OGA and 3TO2 (Figures 7.2A and F). This indicates that the HLA-A*0201 complexes with the docked peptide FSDLYYLTM are more stable. On the other hand, RMSD of the backbone atoms of the peptide bound to HLA-A*0201 during the 10-ns MD simulation was variable in the different structures. In the structures PDB ID: 1OGA and 3TO2 (Figures 7.2A and F) FSDLYYLTM had the greatest RMSD, whereas in the structures PDB ID: 2GIT, 4K7F and 5SWQ, MAEVRSYCY adopted the greatest RMSD value (Figures 7.2B, G, H).

Generally, the Prime-MMGBSA energy for the peptide-HLA-A*0201 complexes was improved during the MD simulations. However, for example for the MAEVRSYCY-HLA-A*0201 (PDB ID: 4K7F) complex the energy increased 10 kcal/mol (from -106.59 to -96.06 kcal/mol) which is compatible with the higher value of RMS deviation in the complex. However, despite the higher RMSD, there is a decrease in the Prime-MMGBSA energy in the MAEVRSYCY-HLA-A*0201 complexes PDB ID: 2GIT and 5SWQ. Consequently, the RMSD may not always indicate loose

binding but only the adaptation of the docked peptide to the binding pocket. Meanwhile, Prime-MMGBSA energy for FSDLYYLTM complexes was improved (decreased) in all cases. For example, FSDLYYLTM complexed with PDB IDs: 1OGA and 4K7F was well-buried via P1, P2 and P3 (Figures 10.2A, G) and in those cases the binding energy decreased (improved) about 50 kcal/mol. Although all MHC I HLA-A*0201 structures complexed with the predicted peptide FSDLYYLTM retained some of their initial hydrogen bonds after the MD simulations, this is the most notable with four H-bonds in the complexes with PDB ID: 2GTW and PDB ID: 3I6G (Appendices 2C and E). All the studied HLA-A*0201 complexes formed more H-bond interactions with FSDLYYLTM compared to MAEVRSYCY.

As mentioned earlier, in only one of the HLA-A*0201 complexes the F pocket size reduced during the MD simulation (MAEVRSYCY with PDB ID: 3TO2, Table 4), whereas in all other predicted as well as the experimentally determined peptide-MHC I complexes the groove widened somewhat at the pocket F region (Table 4; Appendix 3, Figures 11.1-11). Also, Prime-MMGBSA binding energy was generally reduced with some exceptions (Table 4). The reason for the exceptions could be for example a loss of some particular hydrogen bonding interaction in the docked peptide-MHC I complexes. However, one should bear in mind that the value of Prime-MMGBSA free energy of binding is comprised of many components such as Coulomb energy, van der Waals energy, lipophilic energy, pi-pi packing energy, Generalized Born electrostatic solvation energy as well as hydrogen-bonding energy. The Prime-MMGBSA ΔG(bind), the binding free energy is calculated as follows (Eq. 1) (Li et al., 2011):

ΔG(bind) = E_complex[minimized] - (E_ligand[minimized] + E_receptor[minimized]) Eq.1

It is based on the calculation of five fundamental energies, namely:

- Optimized free receptor
- Optimized free ligand
- Optimized complex
- Receptor from minimized/optimized complex
- Ligand from minimized/optimized complex

The improved Prime-MMGBSA energy indicates that most of the docked peptides adopted a better pose during the simulation (exceptions: MAEVRSYCY complexed with PDB ID: 2GTW, 3I6G,

and 4K7F). All but one (3TO2) of the experimentally determined HLA-A*0201 complexes show at least a slight improvement in their Prime-MMGBSA energy after the MD simulations.

Fleishman et al. (2015) suggested that low-affinity peptides widen the peptide-MHC I binding groove and that the groove will be tightened by high-affinity peptides (Fleishman et al., 2015). The strong disulfide bonds between Cys101 from the α2 helix and Cys164 from the β-sheet reinforce the stability of the B pocket while the F pocket lacks such bonds, leading to increased flexibility of this region (Zacharias & Springer, 2004). Thus, it was suggested that the F pocket size could help recognize the tight binding peptides (Abualrous et al., 2015). This study shows that only one of the predicted peptides with high Prime-MMGBSA energy reduced the F pocket size, whereas most predicted peptides and the crystal peptides with the high Prime energy widen the F pocket. Thus, we can draw a conclusion that the F pocket size may not be a decisive criterion to determine the degree of binding affinity of the peptide to the binding groove.

5.3.3 HLA-B*2705 in complex with ZIKV NS3 protein peptides

In this section, we will review the MD analysis of the MHC I HLA-B*2705 (PDB ID: 2BST) complexes with the predicted antigenic peptides from the ZIKV NS3 protein (DIGAVALDY and HSEVQLLAV) and the experimental peptide SRYWAIRTR (Influenza A virus peptide). Results from the MD analysis are presented in Tables 2B and 4, Figures 7.3, 8.3, 9.3, 10.3 and Appendix 3: Figure 11.10).

H-bond interaction analysis revealed that Glu45, Glu63 and Asp77 are the most important residues to sustain the hydrogen bonding network in the HLA-B*2705 complexes before and after the MD simulations (see Table 2B). While the MD simulation resulted in fewer H-bonds in the DIGAVALDY-HLA-B*2705 complex, the number of H-bonds increased in the HSEVQLLAV-HLA-B*2705 complex. The greatest number of H-bonds was formed by His1 in the HSEVQLLAV-HLA-B*2705 complex. RMSD of the backbone atoms of the MHC I binding groove was approximately 1.5 Å and 1.7 Å for the HSEVQLLAV-HLA-B*2705 and DIGAVALDY-HLA-B*2705 complexes, respectively (Figure 7.3). RMSD of the backbone atoms of the peptide bound to HLA-B*2705 rises up to about 2.0 Å for the docked peptides while the experimentally verified antigenic peptide (SRYWAIRTR) had a RMSD value of only 1.2 Å, being the most stable peptide in the groove during the simulations (Figure 8.3).

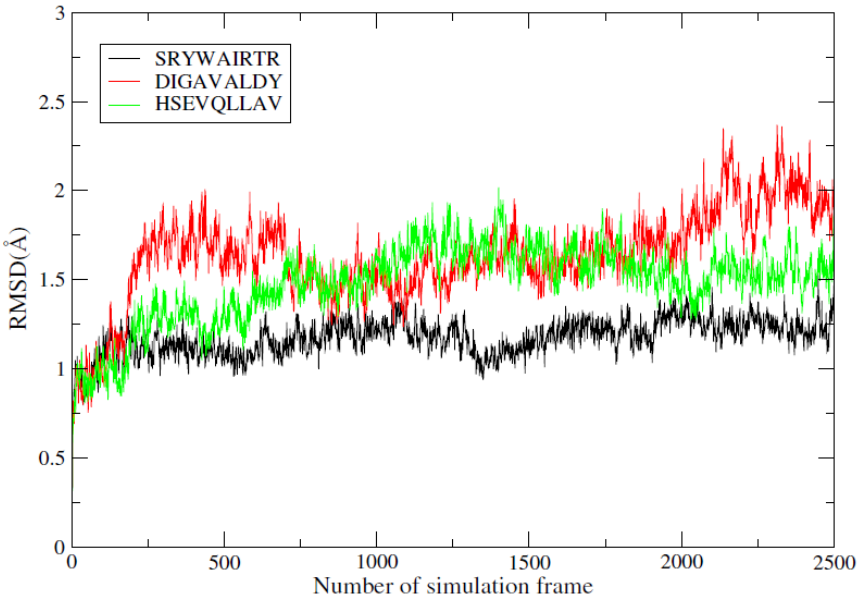


Figure 7.3. RMSD of the backbone atoms of the MHC I binding groove (residues 1-180 in PDB ID: 2BST) during the 10-ns MD simulation of the ZIKV NS3 peptide-HLA-B*2705 complexes. The black line denotes the reference peptide.

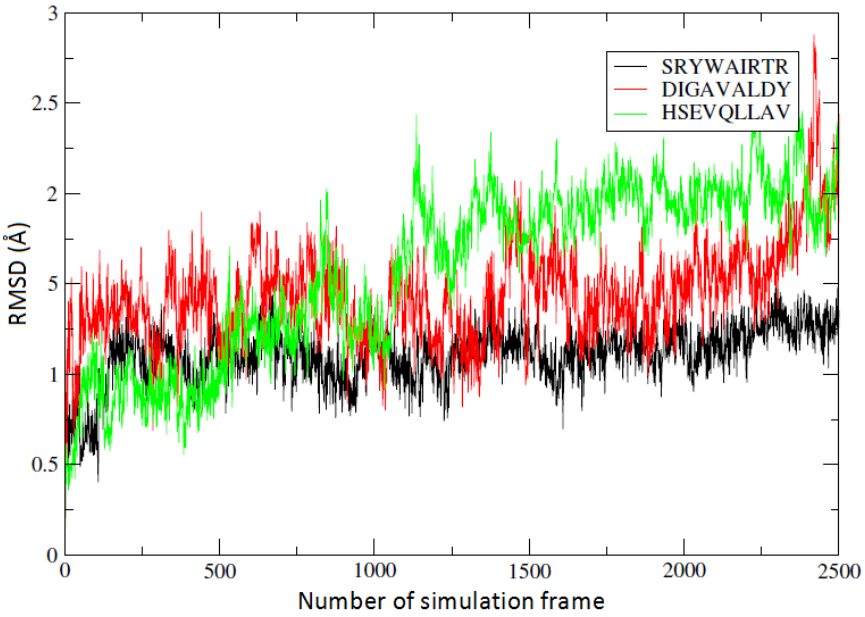


Figure 8.3. RMSD of the backbone atoms of the ZIKV NS3 peptides (green and red lines) and the reference peptide (black line) bound to HLA-B*2705 (PDB ID: 2BST) during the 10-ns MD simulation.

RMSF of the peptide residues and per-residue RSA were also measured for the crystal complex as well as the docked peptide-HLA-B*2705 complexes (Figures 9.3 and 10.3). RMSF of the residues at P2 of HSEVQLLAV and DIGAVALDY rises almost over 1.2 Å, which indicates relatively large flexibility at the binding site (few interactions from the peptide to the pocket). RMSF is

dramatically higher for the residue 9 in both peptides, which indicates that large flexibility and movement of the C-terminus of the peptide does not allow for a tight binding to the groove. The low fluctuations of the co-crystalized peptide SRYWAIRTR emphasize the tight and stable binding of the anchor residues to the binding groove. The relative accessibility of the DIGAVALDY residues to the solvent follows the RSA pattern of the co-crystalized peptide but is higher at the anchor positions P2 and P9 (30-40 % vs 1-7 %). The RSA at the anchor positions in HSEVQLLAV is comparable with DIGAVALDY (31-36 %). On the other hand, P7 and P8 from peptide HSEVQLLAV have lower accessibility to the TCR, which might affect the ability of the peptide to stimulate the immune system (Figure 10.3).

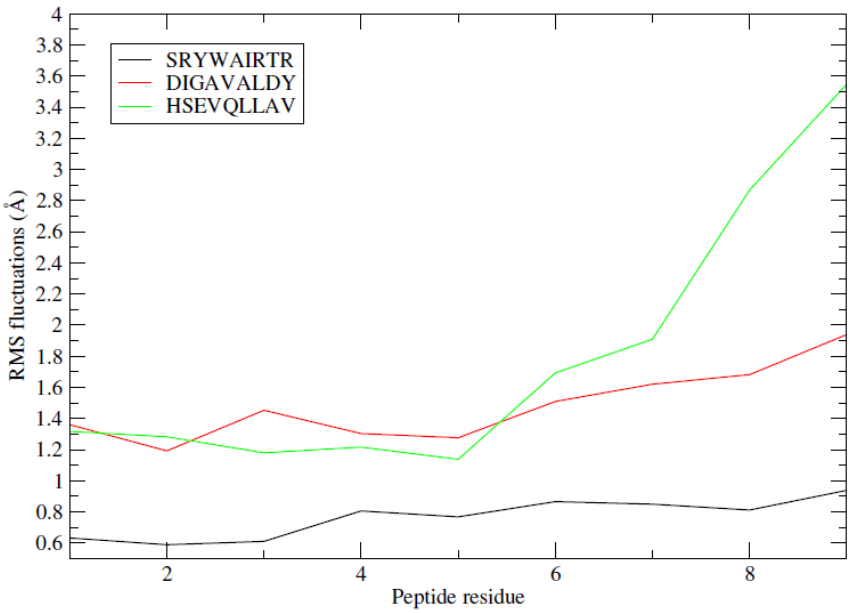


Figure 9.3. Per-residue RMS fluctuations of the ZIKV NS3 protein peptides the binding groove of MHC I HLA-B*2705 (PDB ID: 2BST; the black line denotes the reference peptide) during the 10-ns MD simulations.

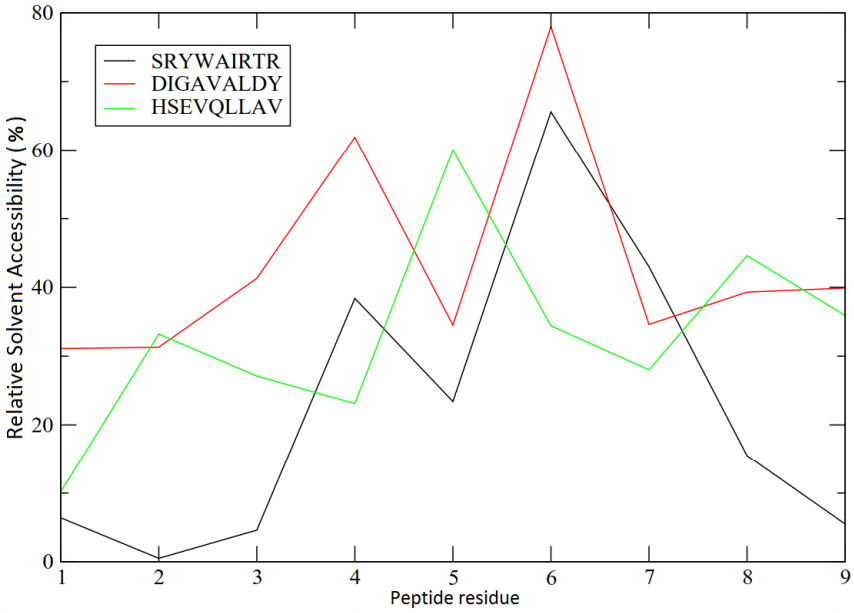


Figure 10.3. Per-residue relative solvent accessibility of the ZIKV NS3 protein peptides in the binding groove of MHC I HLA-B*2705 (PDB ID: 2BST; the black line denotes the reference peptide) after the MD simulations.

The F pocket size in both ZIKV NS3 peptide-MHC I complexes was smaller before the MD simulations than after MD as defined by the distances (d) between the alpha carbon atoms of residues 85 and 138 (d1) and residues 74 and 149 (d2) (see Table 4, Figure 11.1.D). The Prime-MMGBSA free energy of binding for the peptides was poor(-43.90 kcal/mol and -31.32 kcal/mol for DIGAVALDY and HSEVQLLAV, respectively) and did not improve much during the simulations (Table 4). The above analysis suggests that neither of the predicted peptides is a strong binder at HLA-A*B2705. However, DIGAVALDY seems to be somewhat more stable particularly within the F pocket, probably due to the rigidity of Tyr9 compared to Val9 of HSEVQLLAV. It also has a more solvent-exposed stretch in the central area to be recognized by TCR.

5.3.4 HLA-C*0801 in complex with ZIKV NS5 protein peptides

The results of the MD analysis of the HLA-C*0801 molecule (PDB ID: 4NT6) complexed with ZIKV NS5 peptides MTTEDMLVV and FTNLVVQLI are reviewed in this section. Both docked peptides kept some of their initial H-bonds in the binding groove and the number of H-bonds for both the predicted and the crystal peptide-MHC I complexes rose during the simulations (Table 2C). This

rise is more observable for FTNLVVQLI than MTTEDMLVV. Among the residues of the MHC I HLA-C*0801 binding groove, Trp147 (NE1 atom) plays a crucial role as it forms a hydrogen bond interaction to the peptide C-terminus in both the docked complexes and the crystal before and after the MD simulations. In addition, one clash was found between Thr2 of MTTEDMLVV and Tyr9 of the MHC I binding groove before MD (Table 3), suggesting that FTNLVVQLI would be a better fitting peptide to the MHC I binding groove compared to MTTEDMLVV. However, both docked peptides showed equal improvement in the Prime-MMGBSA energy (Table 4). MTTEDMLVV had a lower RMSD value (1.7 Å) for the binding groove compared to that of FTNLVVQLI (2.0 Å; Figure 7.4). Besides, RMSD of the backbone atoms of the peptides during the 10-ns MD simulation rose up to about 1.5 Å for the docked peptides, whereas it was only about 0.5 Å for the co-crystalised peptide GILGFVFTL (Figure 8.4).

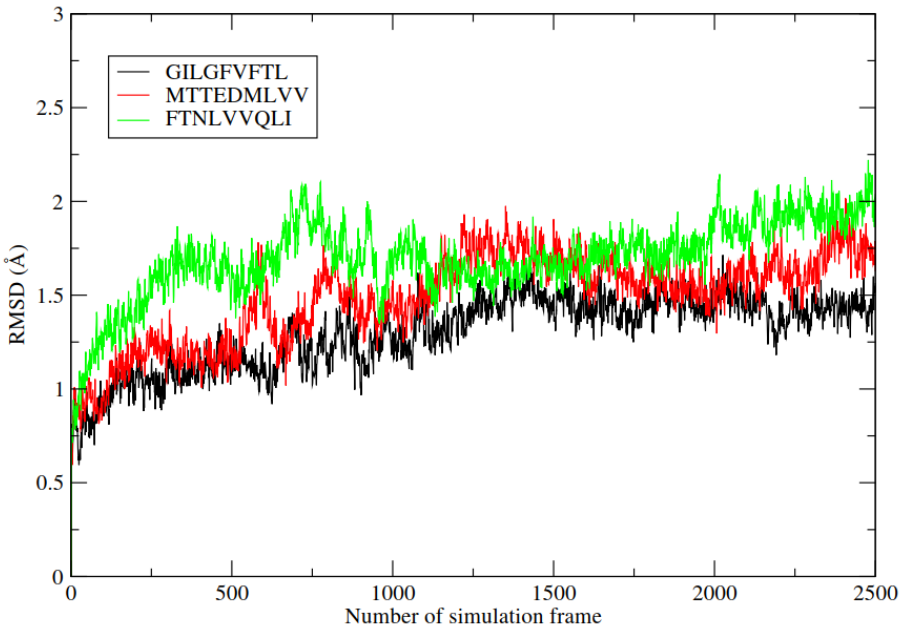


Figure 7.4. RMSD of the backbone atoms of the MHC I binding groove (residues 1-180 in PDB ID: 4NT6) during the 10-ns MD simulations of the ZIKV NS5 peptide-HLA-C*0801 complexes. The black line denotes the reference peptide.

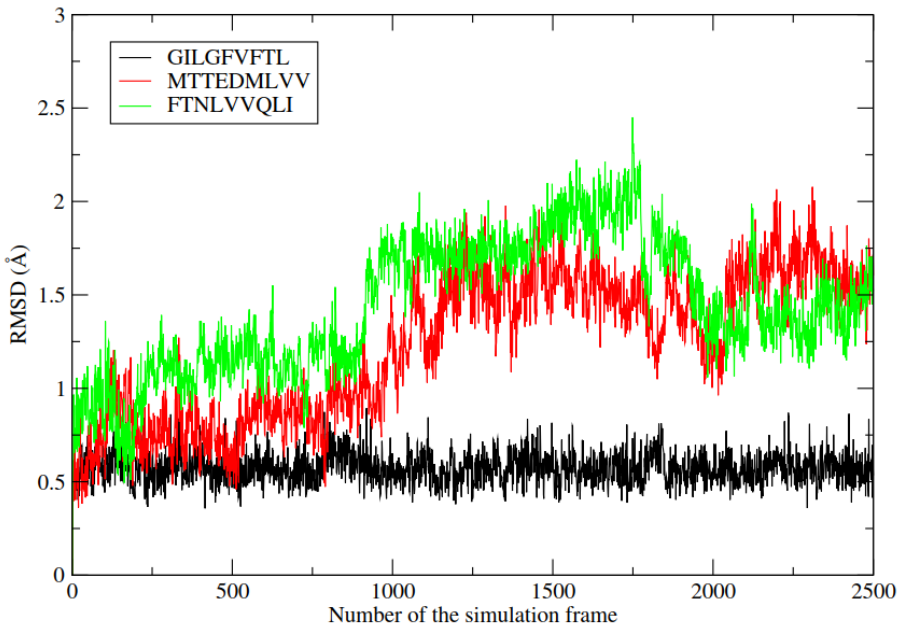


Figure 8.4. RMSD of the backbone atoms of the ZIKV NS5 peptides (green and red lines) and the reference peptide (black line) bound to HLA-C*0801 (PDB ID: 4NT6) during the 10-ns MD simulations.

Per-residue RMSF of the ZIKV NS5 peptides in the binding groove of the MHC I HLA-C*0801 demonstrates less fluctuation in the N-terminus of FTNLVVQLI and the C-terminus of MTTEDMLVV (Figure 9.4). The fluctuations at P2 rise up to about 1.2 Å for FTNLVVQLI and about 1.8 Å for MTTEDMLVV. This means that the flexibility of especially MTTEDMLVV in the peptide-binding domain is high, which results in a less stable peptide-MHC I complex (few interactions from the peptide to the pocket). Therefore, it is likely that MTTEDMLVV is not a strong binder to HLA-A*0801 (and not a suitable vaccine candidate if it cannot bind to any other HLA molecules either). It should be also mentioned that the F pocket enlarged in the crystal complex as well as the docked peptide-HLA-C*0801 complexes.

Measured per-residue RSA profile for the docked peptide-HLA-C*0801 complexes shows that MTTEDMLVV follows the RSA profile of the crystal peptide, suggesting that the anchor residues at P2 and P9 have a stronger binding affinity to the MHC I and residues at P4-P6 to the TCR. However, both the docked peptides have overall much greater per-residue RSA values than the crystal peptide, suggesting that they cannot bind strongly to the MHC I HLA-C*0801 molecule (Figure 10.4).

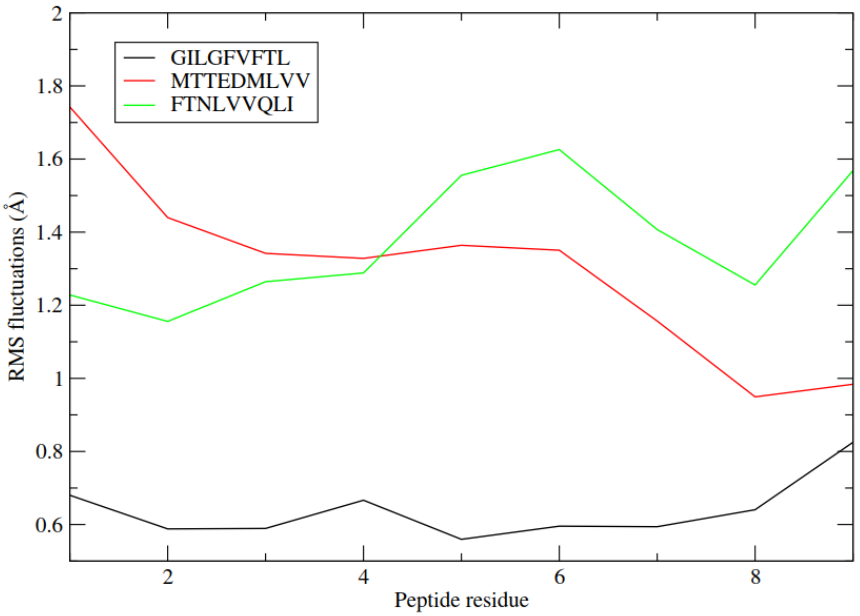


Figure 9.4. Per-residue RMS fluctuations of the ZIKV NS5 protein peptides in the binding groove of MHC I HLA-C*0801 (PDB ID: 4NT6; the black line denotes the reference peptide) during the 10-ns MD simulations.

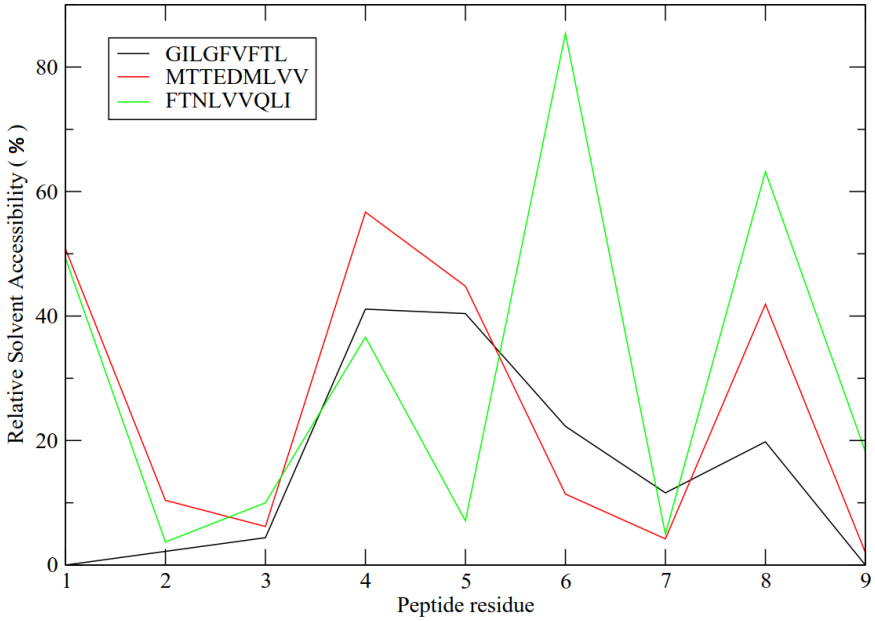


Figure 10.4. Per-residue relative solvent accessibility of the ZIKV NS5 protein peptides in the binding groove of MHC I HLA-C*0801 (PDB ID: 4NT6, the black line denotes the reference peptide) after the MD simulations.

Table 4. Dynamics and energetics of the ZIKV peptide-MHC I complexes (experimental complexes are denoted in italics)

PDB structure	ZIKV peptide sequence	RMSD peptide (Å)[a]	Change in the MHC I binding groove (F pocket) size		MM-GBSA dG (kcal/mol) before MD	MM-GBSA dG (kcal/mol) after MD
			d1/d2[b] (Å) (initial)	d1/d2 (Å) (after MD)		
4NQV	***Antigenic epitopes from ZIKV E protein bound to MHC I HLA-A*0101***					
	TMNNKHWLV	0.889	9.44/20.76	9.18/22.8	-136.607	-108.447
	GLDFSDLYY	0.93	9.44/20.70	12.44/20.98	-133.661	-152.772
	CTELKLSDY	*1.24*	*9.74/20.73*	*10.15/21.76*	*-107.604*	*-133.745*
1OGA	***Antigenic epitopes from ZIKV E protein bound to MHC I HLA-A*0201***					
	MAEVRSYCY	0.882	9.6/20.03	10.43/21.02	-91.53	-68.74
	FSDLYYLTM	1.06	9.6/20.3	9.11/21.45	-61.36	-117.156
	GILGFVFTL	*0.85*	*9.6/20.3*	*12.34/20.52*	*-104.87*	*-123.38*
2GIT	**MAEVRSYCY**	0.995	9.43/20.02	12.1/22.8	-79.307	-88.785
	FSDLYYLTM	1.55	9.7/20.37	10.6/20.26	-85.650	-97.445
	LLFGKPVYV	*0.607*	*9.32/20.21*	*11.53/19.94*	*-106.223*	*-118.193*
2GTW	**MAEVRSYCY**	1.32	9.55/19.73	12.44/21.32	-44.874	-11.917
	FSDLYYLTM	1.07	9.53/19.72	11.24/19.72	-35.543	-125.471
	LAGIGILTV	*0.750*	*9.55/19.72*	*12.33/21.26*	*-69.1*	*-90.817*
2GTZ	**MAEVRSYCY**	0.992	9.42/20.03	11.73/23.3	-33.27	-78.04
	FSDLYYLTM	1.35	9.43/19.95	13.8/22.53	-59.34	-86.76
	ALGIGILTV	1.167	9.57/19.58	10.35/20.55	-80.90	-86.016
3I6G	**MAEVRSYCY**	0.992	9.87/19.72	11.96/21.15	-96.39	-74.76
	FSDLYYLTM	0.913	9.87/19.72	12.08/21.19	-86.03	-131.405
	GLMWLSYFV	*0.925*	*9.87/19.72*	*11.36/20.95*	*-97.66*	*-131.89*
3TO2	**MAEVRSYCY**	1.2	10.16/19.9	**9.46/19.76[c]**	-69.93	-105.27
	FSDLYYLTM	1.15	10.16/19.9	12.05/21.27	-56.63	-117.49
	LACFVLAAV	*0.926*	*9.87/19.72*	*10.91/21.08*	*-92.09*	*-90.72*
4K7F	**MAEVRSYCY**	0.954	10.51/20.44	12.09/21.79	-106.59	-96.04
	FSDLYYLTM	1.32	10.51/20.44	11.34/22.11	-66.78	-112.79
	VCWGELMNL	*1.04*	*10.51/20.44*	*12.57/21.07*	*-92.42*	*-126.61*

5SWQ	**MAEVRSYCY**	0.981	9.88/20.19	12.5/21.44	-70.95	-90.20
	FSDLYYLTM	0.676	9.88/20.19	11.42/21.66	-61.87	-101.26
	CVNGSCFTV	*0.757*	*9.878/20.19*	*10.28/19.3*	*-82.27*	*-110.170*
2BST	***Antigenic epitopes from ZIKV NS3 protein bound to MHC I HLA-B*2705***					
	DIGAVALDY	1.135	9.5/20	12.36/20.08	-43.90	-66.83
	HSEVQLLAV	1.1	9.5/20	13.52/18.96	-31.32	-44.06
	SRYWAIRTR	*1.02*	*9.5/20*	*10.41/20.54*	*-118.37*	*-169.418*
4NT6	***Antigenic epitopes from ZIKV NS5 protein bound to MHC I HLA-C*0801***					
	MTTEDMLVV	1.17	9.19/19.63	10.24/20.86	-86.06	-101.112
	FTNLVVQLI	0.679	9.19/19.63	14.55/19.8	-62.06	-86.14
	GILGFVFTL	*0.373*	*9.19/19.63*	*11.71/20.44*	*-127.25*	*-129.59*

[a] RMSD of the Cα atoms between the initial docked peptide conformation and the minimized conformation from the final MD frame; [b]d1=distance between the Cα atom of Tyr85 in α1 helix and Met138 in α2 helix (Thr138 in HLA-B*2705 and HLA-C*0801); d2=distance between the Cα atom of Asp74 in α1 helix and Ala149 in α2 helix (Tyr74 in HLA-B*2705, His74 in HLA-A*0201); [c]F pocket size has been reduced during the MD simulations.

6. Conclusion

Regarding the recent outbreak of ZIKV infection, urgent action is needed to develop a preventive or a therapeutic ZIKV vaccine. This project employed computational methods such as docking and MD simulations to evaluate the binding affinity of candidate T-cell epitope peptides of ZIKV proteins to the MHC class I molecules.

MHC I alleles HLA-A*0101, HLA-A*0201, HLA-B*2705 and HLA-C*0801 were used as receptor structures and eight different peptides from ZIKV proteins (E, NS3, NS5) were docked at these MHC I molecules (see Chapter 4). The initial PDB structures were prepared with the Protein Preparation Wizard tool implemented in Maestro. Residues of the experimental peptide were mutated to the residues in the desired ZIKV peptide with Chimera. Next, mutated peptide-MHC I complexes were submitted to the FlexPepDock tool to refine the geometry of the complex structures and remove the atomic clashes that resulted from the manual docking. All the docked peptide-MHC I complexes from FlexPepDock were submitted to MD simulation for further refinement along with the experimentally determined peptide-MHC I complexes to examine and compare their stability. The MD simulations were performed in an explicit TIP3P water model at 300 K and at 1 bar pressure for 10 ns. Na+ ions were used for neutralizing the simulation system.

Multiple parameters were measured to analyze the binding strength of the peptides to the MHC I molecules such as: hydrogen bonding network, RMSD for both the backbone atoms of the MHC I peptide-binding domain and the peptides, atomic fluctuations of the peptide residues, per-residue SASA of the peptides, Prime-MMGBSA free energies of binding and the size of the F pocket before and after the MD simulations. Among all, RMSF (atomic fluctuation) and atomic solvent accessibility play crucial roles in the evaluation of the peptide affinity to the MHC I. It has been argued that the calculation of interaction energies should not have a decisive role in predicting the peptide affinity since the potentials used for the calculations are generally approximations and for example the hydrophobic effect has been neglected in many cases (Höltje et al., 2003). However, Prime-MMGBSA free energy calculation may be a more accurate method as it includes the lipophilic energy and solvation effects and thus, it can be useful in at least comparing the different peptides and providing an estimate on any likely favorable interactions.

To sum up, the hydrogen bonding network of crystal structures could be partially

regenerated for the docked structures as well. Although there were some differences in the hydrogen bonding patterns, low-affinity binders could be generally recognized by the loss of H-bond interactions compared to the better binders and especially the co-crystalized peptides. There was a considerable change in the hydrogen bonding network regenerated for some crystal complexes and almost all docked complexes after the MD simulation. Generally, it can be said that the binding affinity of the peptide highly correlates to the number of hydrogen bonds between the peptide (particularly its N- and C-terminus) and the MHC I molecule. In addition, all crystal structures retained some of the initial H-bonds after the MD simulation. Besides, despite some steric clashes in the docked complexes before the MD simulation, these conflicts were removed via forming new favorable hydrogen bond interactions during the MD simulations (Table 3). Overall, the complexes get stabilized during the MD simulation.

Since the fluctuation of the peptide directly depends on its binding interactions, measuring the atomic fluctuations of the peptide residues provides a good insight into the ligand activity-related properties on (Höltje et al., 2003). In this regard, calculating the fluctuations of the docked peptides and comparing them to the fluctuations of the co-crystalized peptides gives possibilities to evaluate and understand the degree of flexibility of the docked peptide. One should bear in mind that not only the degree of burial of the peptide residues in the corresponding MHC I pockets is crucial, but the docking pose of the peptide is of importance as well. This means that sometimes the docked peptides can adopt bulged or even helical conformation rather than an extended pose, which is not the proper geometry to stimulate the TCR. Although some peptides seemed to interact with the MHC I molecule via the residues between P4 to P7, most of the interactions formed between the peptide N- and C-terminus, which supports the well-known pattern of the peptide-MHC I complex to elicit a cellular immune response.

Our study concluded that the changes in the size of the F pocket as a decisive criterion to determine the binding affinity of the peptide to the MHC molecule is arguable. In this regard, we refer to the finding by Zakharias and Springer (2004) who suggest that the binding of the C-terminal segment of the peptide to the F pocket may require stabilization by a chaperon protein during the peptide loading (Zacharias & Springer, 2004).

Considering the sequence variability analysis of the MHC I molecules and also the analysis of the degree of affinity of the peptide to the binding groove of the MHC I molecule, this study reveals that each allelic polymorphism in the MHC I molecules (particularly in the peptide binding site) has a substantial effect on the binding affinity of the peptide to the MHC I peptide-binding domain. Moreover, in a research conducted by Reche and Reinherz (2003) where they investigated the sequence variability and peptide/TCR contacts in the peptide-binding domain of the MHC I (α1 and α2 domains), the result indicated that almost all residues that are in contact with TCR are in the proximity of the peptide (< 5 Å). In addition, the diversity of the residues in the α3 domain (CD8+ binding domain), especially in position 225-232 belongs to the motifs that are in contact with the CD8+ co-receptor. This co-receptor assists with the stability of the overall complex when peptide-MHC I is connected to TCR (Margulies, 1997). Accordingly, it can be concluded that the affinity of the peptides to the MHC I (the scope of this study) is not the only contributing factor for the immunogenesis. The affinity of MHC I molecules to CD8+ in α3 domain, as well as the affinity of the peptide-MHC I complex to the TCR in α1 and α2 domains should also be taken into account when measuring the immunogenicity of the peptide-MHC I complexes.

The final results demonstrated that the two predicted peptides, GLDFSDLYY and FSDLYYLTM, exhibited the highest affinity to the studied MHC I alleles (HLA-A*0101 and HLA-A*0201). However, the rest of the studied ZIKV peptides seem not to bind strongly to the respective MHC I molecules used in the study.

Finally, the study showed that *in silico* approach can be used as an effective and fast method to predict candidate peptides for vaccine design and at the same time the approach reduces the number of in vitro experiments. As a suggestion for future study, other peptides could also be studied in a similar way for their binding affinity. Meanwhile, the peptides could also be studied using different docking tools to investigate their binding affinity for the same, or other, MHC I HLA molecules.

7. References

Abbas, A. K., Lichtman, A. H., & Pillai, S. (2014). *Basic immunology: functions and disorders of the immune system*. Elsevier Health Sciences.

Abualrous, E. T., Saini, S. K., Ramnarayan, V. R., Ilca, F. T., Zacharias, M., & Springer, S. (2015). The carboxy terminus of the ligand peptide determines the stability of the MHC class I molecule H-2Kb: a combined molecular dynamics and experimental study. *PloS one*, *10*(8), e0135421.

Adamczyk-Poplawska, M., Markowicz, S., & Jagusztyn-Krynicka, E. K. (2011). Proteomics for development of vaccine. *Journal of proteomics*, 74(12), 2596-2616.

Aichele, P. E. T. E. R., Hengartner, H. A. N. S., Zinkernagel, R. M., & Schulz, M. A. N. F. R. E. D. (1990). Antiviral cytotoxic T cell response induced by in vivo priming with a free synthetic peptide. *Journal of Experimental Medicine*, 171(5), 1815-1820.

Alder, B. J., & Wainwright, T. (1957). Phase transition for a hard sphere system. *The Journal of Chemical Physics*, *27*(5), 1208-1209.

Altschul, S. F., Gish, W., Miller, W., Meyers, E. W., & Lipman, D.J., (1990). Basic local alignment search tool. *Journal of Molecular Biology,* 215(3), 403-410.

Artymiuk, P. J., Blake, C. C. F., Grace, D. E. P., Oatley, S. J., Phillips, D. C., & Sternberg, M. J. E. (1979). Crystallographic studies of the dynamic properties of lysozyme. *Nature*, *280*(5723), 563.

Backert, L., & Kohlbacher, O. (2015). Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome medicine*, *7*(1), 119.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The protein data bank. *Nucleic acids research*, *28*(1), 235-242.

Brasil, P., Pereira Jr, J. P., Moreira, M. E., Ribeiro Nogueira, R. M., Damasceno, L., Wakimoto, M., ... & Zin, A. A. (2016). Zika virus infection in pregnant women in Rio de Janeiro. *New England Journal of Medicine*, *375*(24), 2321-2334.

Brown JH, J. T. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* , 364:33–9.10.1038/364033a0.

Brusic, V., & Flower, D. R. (2004). Bioinformatics tools for identifying T-cell epitopes. *Drug Discovery Today: Biosilico*, *2*(1), 18-23.

Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M., & Fernández-Viña, M. A. (2001). Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the

United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Human immunology*, *62*(9), 1009-1030.

Case, D. A., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden III, T. A., Duke, R. E., ... & Izadi, S. AMBER 2016 (University of California, 2016). *Google Scholar*.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., ... & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of computational chemistry*, *26*(16), 1668-1688.

Chen, C., Huang, H., & Wu, C. H. (2017). Protein bioinformatics databases and resources. In *Protein Bioinformatics* (pp. 3-39). Humana Press, New York, NY.

Choo, J. A., Liu, J., Toh, X., Grotenbreg, G. M., & Ren, E. C. (2014). The immunodominant Influenza A virus M158-66 CTL epitope exhibits degenerate class I MHC restriction in humans. *Journal of Virology*, JVI-00855.

Clements, C. S., Kjer-Nielsen, L., Kostenko, L., Hoare, H. L., Dunstone, M. A., Moses, E., ... & McCluskey, J. (2005). Crystal structure of HLA-G: a nonclassical MHC class I molecule expressed at the fetal–maternal interface. *Proceedings of the National Academy of Sciences*, *102*(9), 3360-3365.

Cole, D. K., Rizkallah, P. J., Gao, F., Watson, N. I., Boulter, J. M., Bell, J. I., ... & Jakobsen, B. K. (2006). Crystal structure of HLA-A* 2402 complexed with a telomerase peptide. *European Journal of Immunology*, *36*(1), 170-179.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., ... & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, *117*(19), 5179-5197.

Dalke, A., Humphrey, W., & Ulrich, J. (1997). Theoretical Biophysics Group, University of Illinois and Beckman Institute. *Urbana, USA*.

Delisi, C., & Berzofsky, J. A. (1985). T-cell antigenic sites tend to be amphipathic structures. *Proceedings of the National Academy.*

de Paula Freitas, B., de Oliveira Dias, J. R., Prazeres, J., Sacramento, G. A., Ko, A. I., Maia, M., & Belfort, R. (2016). Ocular findings in infants with microcephaly associated with presumed Zika virus congenital infection in Salvador, Brazil. *JAMA Ophthalmology*, *134*(5), 529-535.

Devadoss, F. R., & Raj, V. P. (2014). Analysis and visual summarization of molecular dynamics simulation. *Journal of cheminformatics*, *6*(1), O16.

Dick, G. W. A., Kitchen, S. F., & Haddow, A. J. (1952). Zika virus (I). Isolations and serological

specificity. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, *46*(5), 509-520.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797.

Elber, R., Roitberg, A., Simmerling, C., Goldstein, R., Li, H., Verkhivker, G., ... & Ulitsky, A. (1995). MOIL: A program for simulations of macromolecules. *Computer Physics Communications*, *91*(1-3), 159-189.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of chemical physics*, *103*(19), 8577-8593.

Falk, K., Rötzschke, O., Stevanović, S., Jung, G., & Rammensee, H. G. (1991). Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature*, *351*(6324), 290.

Fauci, A. S., & Morens, D. M. (2016). Zika virus in the Americas—yet another arbovirus threat. *New England Journal of Medicine*, *374*(7), 601-604.

Fleischmann, G., Fisette, O., Thomas, C., Wieneke, R., Tumulka, F., Schneeweiss, C., ... & Tampé, R. (2015). Mechanistic basis for epitope proofreading in the peptide-loading complex. *The Journal of Immunology*, 1501515.

Florea, L., Halldorsson, B., Kohlbacher, O., Schwartz, R., Hoffman, S., & Istrail, S. (2003). Epitope prediction algorithms for peptide-based vaccine design. In *Bioinformatics Conference, 2003. CSB 2003. Proceedings of the 2003 IEEE* (pp. 17-26).

Grant, E. J., Josephs, T. M., Valkenburg, S. A., Wooldridge, L., Hellard, M., Rossjohn, J., ... & Gras, S. (2016). Lack of heterologous cross-reactivity towards HLA-A* 02: 01 restricted viral epitopes is underpinned by distinct αβT cell receptor signatures. *Journal of Biological Chemistry*, jbc-M116.

Hubbard, S. J., & Thornton, J. M. (1993). NACCESS: Department of Biochemistry and Molecular Biology, University College London. *Software available at http://www. bioinf. manchester. ac. uk/naccess/nacdownload. html*.

Hunt, D. F., Henderson, R. A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., ... & Engelhard, V. H. (1992). Characterization of peptides bound to the class I MHC molecule HLA-A2. 1 by mass spectrometry. *Science*, *255*(5049), 1261-1263.

Höltje, H. D., Sippl, W., Rognan, D., & Folkers, G. (2003). Molecular Modeling, Basic Principles and Applications. 2nd Edition. Wiley-VCH, Weinheim, Germany.

Jacobson, M. P., Friesner, R., Xiang, Z., & Honig, B. (2002). On the Role of Crystal Packing Forces in Determining Protein Sidechain Conformations. *J. Mol. Biol*, 320, 597-608.

Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E., & Friesner, R. A. (2004). A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins: Structure, Function and Bioinformatics*, 55, 351-367.

James C. Phillips, R. B. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26:1781-1802.

Jardetzky, T. S. (1991.). Identification of self peptides bound to purified HLA-B27. *Nature , 353:*, 326–329.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, *79*(2), 926-935.

Kolaskar, A. S., & Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS letters*, *276*(1-2), 172-174.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, *157*(1), 105-132.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., ...& Thompson, J. D. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947-2948.

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*, *8*(1), 424.

Lazarski, C. A., Chaves, F. A., Jenks, S. A., Wu, S., Richards, K. A., Weaver, J. M., & Sant, A. J. (2005). The kinetic stability of MHC class II: peptide complexes is a key parameter that dictates immunodominance. *Immunity*, *23*(1), 29-40.

Lee, B., & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *Journal of molecular biology*, *55*(3), 379-IN4.

Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., & Friesner, R. A. (2011). The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins: Structure, Function, and Bioinformatics*, *79*(10), 2794-2812.

Liang, T. J. (2013). Current progress in development of hepatitis C virus vaccines. *Nature Medicine*, *19*(7), 869.

Liljefors, T., Krogsgaard-Larsen, P., & Madsen, U. (Eds.). (2002). *Textbook of Drug Design and Discovery*. CRC Press.

Lindenbach, B. D., & Rice, C. M. (2003). Molecular biology of flaviviruses. *Advances in Virus Research*, *59*, 23-62.

Lipman, D. J., & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science*, *227*(4693), 1435-1441.

Liu, J., Sun, Y., Qi, J., Chu, F., Wu, H., Gao, F., ... & Gao, G. F. (2010). The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. *The Journal of Infectious Diseases*, *202*(8), 1171-1180.

London, N., Raveh, B., Cohen, E., Fathi, G., & Schueler-Furman, O. (2011). Rosetta FlexPepDock web server—high resolution modeling of peptide–protein interactions. *Nucleic acids research*, *39*(suppl_2), W249-W253.

Madden, D. R., Gorga, J. C., Strominger, J. L., & Wiley, D. C. (1991). The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature*, *353*(6342), 321.

Madden, D. R., Gorga, J. C., Strominger, J. L., & Wiley, D. C. (1992). 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.

Mansuy, J. M., Dutertre, M., Mengelle, C., Fourcade, C., Marchou, B., Delobel, P., ... & Martin-Blondel, G. (2016). Zika virus: high infectious viral load in semen, a new sexually transmitted pathogen. *Lancet Infect Dis*, *16*(4), 405.

Margulies, D. H. (1997). Interactions of TCRs with MHC-peptide complexes: a quantitative basis for mechanistic models. *Current Opinion in Immunology*, *9*(3), 390-395.

McMurry, J., Sbai, H., Gennaro, M. L., Carter, E. J., Martin, W., & De Groot, A. S. (2005). Analyzing Mycobacterium tuberculosis proteomes for candidate vaccine epitopes. *Tuberculosis*, *85*(1-2), 95-105.

Merz Jr, K. M., Ringe, D., & Reynolds, C. H. (Eds.). (2010). *Drug design: structure-and ligand-based approaches*. Cambridge University Press.

MHC Sequencing Consortium. (1999). Complete sequence and gene map of a human major histocompatibility complex. *Nature*, *401*, 921-923.

Mirza, M. U., Rafique, S., Ali, A., Munir, M., Ikram, N., Manan, A., ... & Idrees, M. (2016). Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Scientific Reports*, *6*, 37313.

Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., ... & Vizjak, A. (2016). Zika

virus associated with microcephaly N Engl J Med.

National Research council (1989). Opportunities in Biology. Washington, DC: The national Academic Press. https://doi.org/10.17226/742.

Nicholas, K. B., Nicholas, H. B. J. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments Distributed by the author.

Oehler, E., Watrin, L., Larre, P., Leparc-Goffart, I., Lastere, S., Valour, F., ... & Ghawche, F. (2014). Zika virus infection complicated by Guillain-Barre syndrome–case report, French Polynesia, December 2013. *Eurosurveillance*, *19*(9), 20720.

Patodia, S., Bagaria, A., & Chopra, D. (2014). Molecular Dynamics Simulation of Proteins: A Brief Overview. *Journal of Physical Chemistry & Biophysics*, *4*(6), 1. DOI: 10.4172/2161-0398.1000166.

Pei, J., & Grishin, N. V. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, *17*(8), 700-712.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605-1612. http://www.cgl.ucsf.edu/chimera.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., ... & Hess, B. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, *29*(7), 845-854.

Pundir, S., Martin, M. J., & O'Donovan, C. (2017). UniProt protein knowledgebase. In *Protein Bioinformatics* (pp. 41-55). Humana Press, New York, NY.

Raveh, B., London, N., & Schueler-Furman, O. (2010). Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins: Structure, Function, and Bioinformatics*, *78*(9), 2029-2040.

Reche, P. A., & Reinherz, E. L. (2003). Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *Journal of Molecular Biology*, *331*(3), 623-641.

Roe, D. R., & Cheatham III, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation*, *9*(7), 3084-3095.

Rosalia, R. A., Quakkelaar, E. D., Redeker, A., Khan, S., Camps, M., Drijfhout, J. W., ... & Janssen, G. (2013). Dendritic cells process synthetic long peptides better than whole protein, improving antigen presentation and T-cell activation. *European Journal of Immunology*, *43*(10), 2554-2565.

Rosendahl Huber, S., van Beek, J., de Jonge, J., Luytjes, W., & van Baarle, D. (2014). T cell responses to viral infections–opportunities for peptide vaccination. *Frontiers in Immunology*, *5*, 171.

Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, *23*(3), 327-341.

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design*, *27*(3), 221-234.

Schrödinger, R. (2016). 2: Maestro, Schrödinger, LLC, New York, NY, 2017. *Received: February*, *21*, 2018.

Sette, A., Vitiello, A., Reherman, B., Fowler, P., Nayersina, R., Kast, W. M., ... & Sidney, J. (1994). The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *The Journal of Immunology*, *153*(12), 5586-5592.

Shapovalov, M. V., & Dunbrack Jr, R. L. (2011). A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*, *19*(6), 844-858.

Soria-Guerra, R. E., Nieto-Gomez, R., Govea-Alonso, D. O., & Rosales-Mendoza, S. (2015). An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *Journal of Biomedical Informatics*, *53*, 405-414.

Stern, L. J., Brown, J. H., Jardetzky, T. S., Gorga, J. C., Urban, R. G., Strominger, J. L., & Wiley, D. C. (1994). Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature*, *368*(6468), 215.

Stewart-Jones, G. B., di Gleria, K., Kollnberger, S., McMichael, A. J., Jones, E. Y., & Bowness, P. (2005). Crystal structures and KIR3DL1 recognition of three immunodominant viral peptides complexed to HLA-B* 2705. *European Journal of Immunology*, *35*(2), 341-351.

The UniProt Consortium, (2017). UniProt: the universal protein knowledgebase. *Nucleic acids research*, 45(D1), D158-D159. URL: https://www.uniprot.org.

Toh, H., Savoie, C. J., Kamikawaji, N., Muta, S., Sasazuki, T., & Kuhara, S. (2000). Changes at the floor of the peptide-binding groove induce a strong preference for Proline at position 3 of the bound peptide: Molecular dynamics simulations of HLA-A* 0217. *Biopolymers: Original Research on Biomolecules*, *54*(5), 318-327.

Tollenaere, J. P. (1996). The role of structure-based ligand design and molecular modelling in drug

discovery. *Pharmacy World and Science*, *18*(2), 56-62.

Toussaint, N. C., Maman, Y., Kohlbacher, O., & Louzoun, Y. (2011). Universal peptide vaccines–optimal peptide vaccine design based on viral sequence conservation. *Vaccine*, *29*(47), 8745-8753.

Ventura, C. V., Maia, M., Bravo-Filho, V., Góis, A. L., & Belfort, R. (2016). Zika virus in Brazil and macular atrophy in a child with microcephaly. *The Lancet*, *387*(10015), 228.

Vitkup, D., Ringe, D., Petsko, G. A., & Karplus, M. (2000). Solvent mobility and the protein'glass' transition. *Nature Structural and Molecular Biology*, *7*(1), 34.

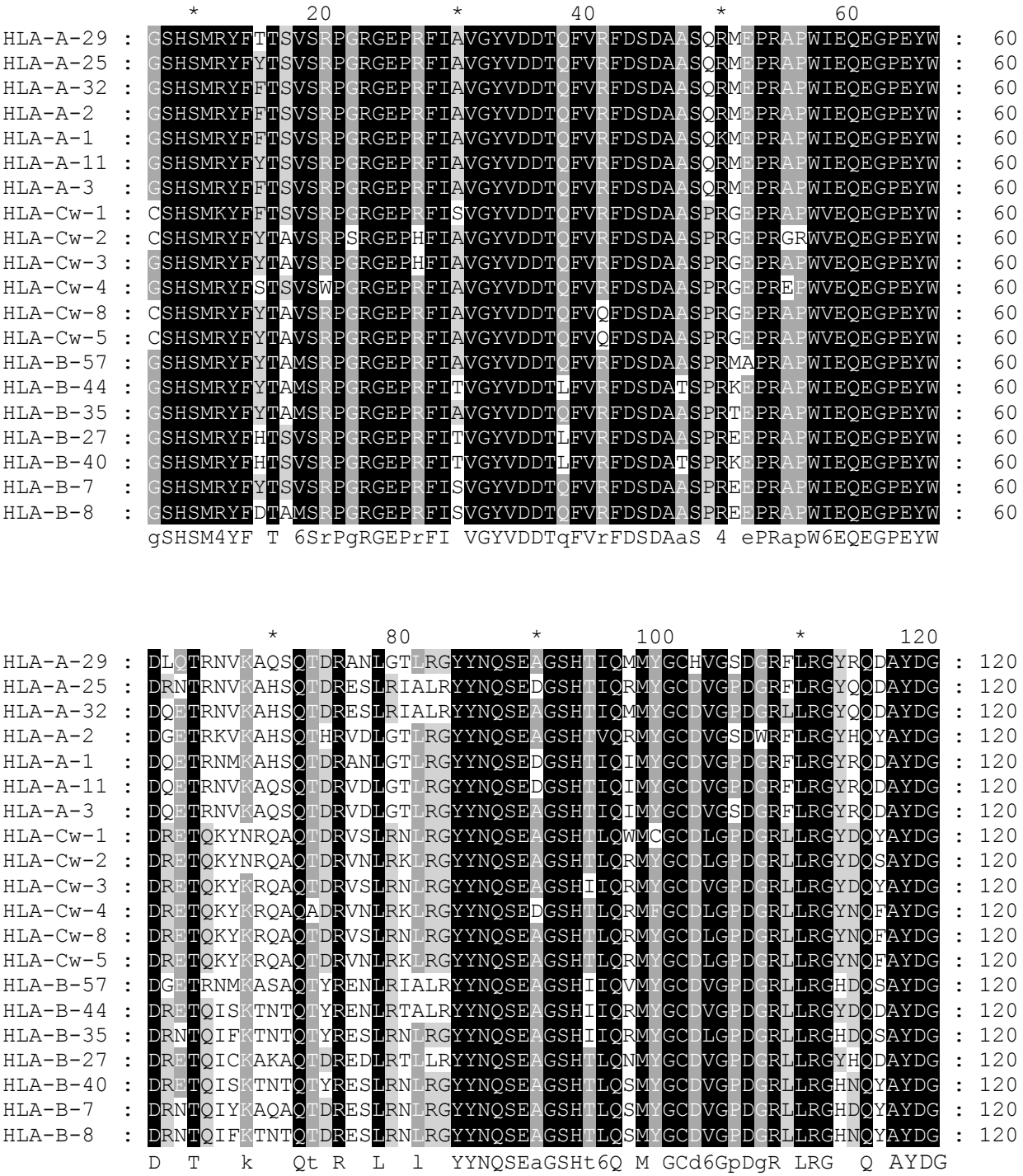
Wieczorek, M., Abualrous, E. T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., & Freund, C. (2017). Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. *Frontiers in immunology*, *8*, 292.

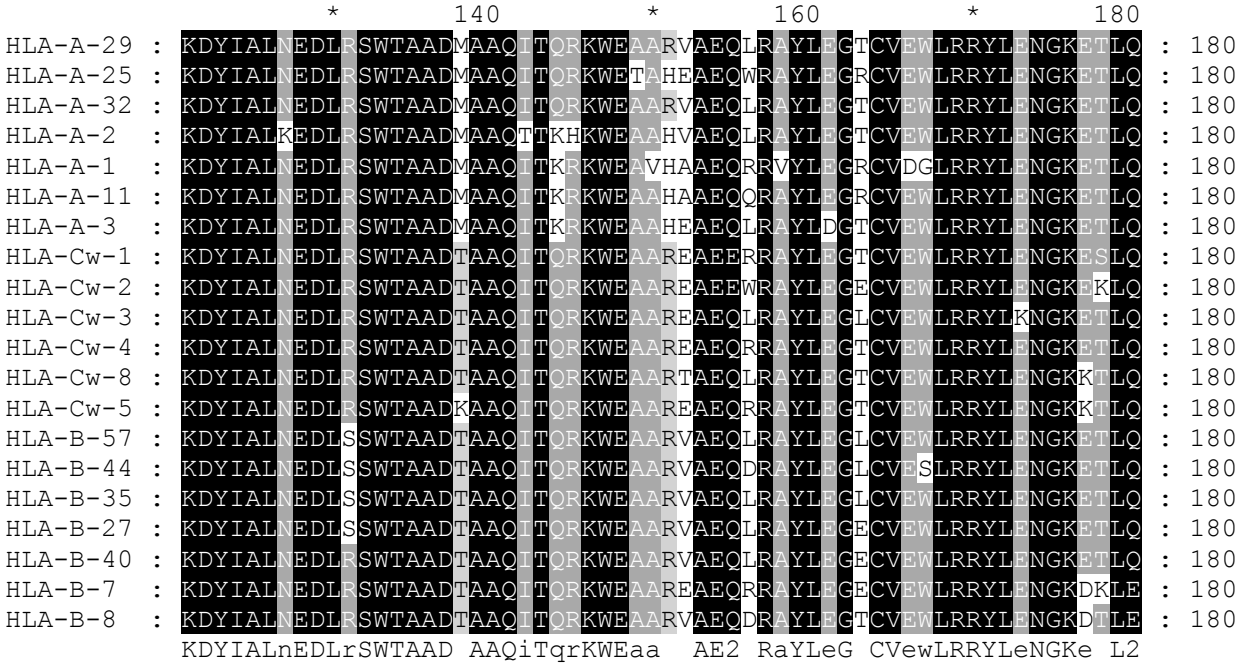
Zacharias, M., & Springer, S. (2004). Conformational flexibility of the MHC class I α1-α2 domain in peptide bound and free states: a molecular dynamics simulation study. *Biophysical Journal*, *87*(4), 2203-2214.

Zinkernagel, R. M., & Doherty, P. C. (1997). The discovery of MHC restriction. *Immunology Today*, *18*(1), 14-17.

Appendices

Appendix 1. Multiple Sequence Alignment of HLA-MHC I molecules




* 140 * 160 * 180
HLA-A-29 : KDYIALNEDLRSWTAADMAAQITQRKWEAARVAEQLRAYLEGTCVEWLRRYLENGKETLQ : 180
HLA-A-25 : KDYIALNEDLRSWTAADMAAQITQRKWETAHEAEQWRAYLEGRCVEWLRRYLENGKETLQ : 180
HLA-A-32 : KDYIALNEDLRSWTAADMAAQITQRKWEAARVAEQLRAYLEGTCVEWLRRYLENGKETLQ : 180
HLA-A-2 : KDYIALKEDLRSWTAADMAAQTTKHKWEAAHVAEQLRAYLEGTCVEWLRRYLENGKETLQ : 180
HLA-A-1 : KDYIALNEDLRSWTAADMAAQITKRKWEAVHAAEQRRVYLEGRCVDGLRRYLENGKETLQ : 180
HLA-A-11 : KDYIALNEDLRSWTAADMAAQITKRKWEAAHAAEQQRAYLEGRCVEWLRRYLENGKETLQ : 180
HLA-A-3 : KDYIALNEDLRSWTAADMAAQITKRKWEAAHEAEQLRAYLDGTCVEWLRRYLENGKETLQ : 180
HLA-Cw-1 : KDYIALNEDLRSWTAADTAAQITQRKWEAAREAEERRAYLEGTCVEWLRRYLENGKESLQ : 180
HLA-Cw-2 : KDYIALNEDLRSWTAADTAAQITQRKWEAAREAEEWRAYLEGECVEWLRRYLENGKEKLQ : 180
HLA-Cw-3 : KDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQLRAYLEGLCVEWLRRYLKNGKETLQ : 180
HLA-Cw-4 : KDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQRRAYLEGTCVEWLRRYLENGKETLQ : 180
HLA-Cw-8 : KDYIALNEDLRSWTAADTAAQITQRKWEAARTAEQLRAYLEGTCVEWLRRYLENGKKTLQ : 180
HLA-Cw-5 : KDYIALNEDLRSWTAADKAAQITQRKWEAAREAEQRRAYLEGTCVEWLRRYLENGKKTLQ : 180
HLA-B-57 : KDYIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRAYLEGLCVEWLRRYLENGKETLQ : 180
HLA-B-44 : KDYIALNEDLSSWTAADTAAQITQRKWEAARVAEQDRAYLEGLCVESLRRYLENGKETLQ : 180
HLA-B-35 : KDYIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRAYLEGLCVEWLRRYLENGKETLQ : 180
HLA-B-27 : KDYIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRAYLEGECVEWLRRYLENGKETLQ : 180
HLA-B-40 : KDYIALNEDLRSWTAADTAAQITQRKWEAARVAEQLRAYLEGECVEWLRRYLENGKETLQ : 180
HLA-B-7 : KDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQRRAYLEGECVEWLRRYLENGKDKLE : 180
HLA-B-8 : KDYIALNEDLRSWTAADTAAQITQRKWEAARVAEQDRAYLEGTCVEWLRRYLENGKDTLE : 180
KDYIALnEDLrSWTAAD AAQiTqrKWEaa AE2 RaYLeG CVewLRRYLeNGKe L2


* 260 * 280 * 300
HLA-A-29 : FQKWASVVVPSGQEQRYTCHVQHEGLPKPLTLRWEPSSQPTIPIVGIIAGLVLFGAV-FA : 299
HLA-A-25 : FQKWASVVVPSGQEQRYTCHVQHEGLPKPLTLRWEPSSQPTIPIVGIIAGLVLFGAV-IA : 299
HLA-A-32 : FQKWASVVVPSGQEQRYTCHVQHEGLPKPLTLRWEPSSQPTIPIVGIIAGLVLFGAM-FA : 299
HLA-A-2 : FQKWAAVVVPSGQEQRYTCHVQHEGLPKPLTLRWEPSSQPTIPIVGIIAGLVLFGAV-IT : 299
HLA-A-1 : FQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWELSSQPTIPIVGIIAGLVLLGAV-IT : 299
HLA-A-11 : FQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWELSSQPTIPIVGIIAGLVLLGAV-IT : 299
HLA-A-3 : FQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWELSSQPTIPIVGIIAGLVLLGAV-IT : 299
HLA-Cw-1 : FQKWAAVMVPSGEEQRYTCHVQHEGLPEPLTLRWEPSSQPTIPIVGIVAGLAVLAVLAVL : 300
HLA-Cw-2 : FQKWAAVVVPSGEEQRYTCHVQHEGLPEPLTLRWEPSSQPTIPIVGIVAGLAVLAVLAVL : 300
HLA-Cw-3 : FQKWAAVVVPSGEEQRYTCHVQHEGLPEPLTLRWEPSSQPTIPIVGIVAGLAVLAVLAVL : 300
HLA-Cw-4 : FQKWAAVVVPSGEEQRYTCHVQHEGLPEPLTLRWKPSSQPTIPIVGIVAGLAVLAVLAVL : 300
HLA-Cw-8 : FQKWAAVVVPSGEEQRYTCHVQHEGLPEPLTLRWGPSSQPTIPIVGIVAGLAVLAVLAVL : 300
HLA-Cw-5 : FQKWAAVVVPSGEEQRYTCHVQHEGLPEPLTLRWGPSSQPTIPIVGIVAGLAVLAVLAVL : 300
HLA-B-57 : FQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSSQSTVPIVGIVAGLAVLAVV-VI : 299
HLA-B-44 : FQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSSQSTVPIVGIVAGLAVLAVV-VI : 299
HLA-B-35 : FQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSSQSTIPIVGIVAGLAVLAVV-VI : 299
HLA-B-27 : FQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSSQSTVPIVGIVAGLAVLAVV-VI : 299
HLA-B-40 : FQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSSQSTVPIVGIVAGLAVLAVV-VI : 299
HLA-B-7 : FQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSSQSTVPIVGIVAGLAVLAVV-VI : 299
HLA-B-8 : FQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSSQSTVPIVGIVAGLAVLAVV-VI : 299
FQKWAaV6VPSG2EQRYTCHVQHEGLP PLTLRWepSSQ T6PIVGI6AGL 6l 6

```
                        *       320         *       340
HLA-A-29 : GAVVAAVRWRRKSSDRKGGSYSQAASSDSAQGSDMSLTACKV : 341
HLA-A-25 : GAVVAAVMWRRKSSDRKGGSYSQAASSDSAQGSDMSLTACKV : 341
HLA-A-32 : GAVVAAVRWRRKSSDRKGGSYSQAASSDSAQGSDMSLTACKV : 341
HLA-A-2  : GAVVAAVMWRRKSSDRKGGSYSQAASSDSAQGSDVSLTACKV : 341
HLA-A-1  : GAVVAAVMWRRKSSDRKGGSYTQAASSDSAQGSDVSLTACKV : 341
HLA-A-11 : GAVVAAVMWRRKSSDRKGGSYTQAASSDSAQGSDVSLTACKV : 341
HLA-A-3  : GAVVAAVMWRRKSSDRKGGSYTQAASSDSAQGSDVSLTACKV : 341
HLA-Cw-1 : GAVVAVVMCRRKSSGGKGGSCSQAASSNSAQGSDESLIASKA : 342
HLA-Cw-2 : GAVVAVVMCRRKSSGGKGGSCSQAASSNSAQGSDESLIASKA : 342
HLA-Cw-3 : GAVVAVVMCRRKSSGGKGGSCSQAASSNSAQGSDESLIACKA : 342
HLA-Cw-4 : GAMVAVVMCRRKSSGGKGGSCSQAASSNSAQGSDESLIACKA : 342
HLA-Cw-8 : GAVMAVVMCRRKSSGGKGGSCSQAASSNSAQGSDESLIACKA : 342
HLA-Cw-5 : GAVMAVVMCRRKSSGGKGGSCSQAASSNSAQGSDESLIACKA : 342
HLA-B-57 : GAVVAAVMCRRKSSGGKGGSYSQAACSDSAQGSDVSLTA--- : 338
HLA-B-44 : GAVVAAVMCRRKSSGGKGGSYSQAACSDSAQGSDVSLTA--- : 338
HLA-B-35 : GAVVATVMCRRKSSGGKGGSYSQAASSDSAQGSDVSLTA--- : 338
HLA-B-27 : GAVVAAVMCRRKSSGGKGGSYSQAACSDSAQGSDVSLTA--- : 338
HLA-B-40 : GAVVAAVMCRRKSSGGKGGSYSQAACSDSAQGSDVSLTA--- : 338
HLA-B-7  : GAVVAAVMCRRKSSGGKGGSYSQAACSDSAQGSDVSLTA--- : 338
HLA-B-8  : GAVVAAVMCRRKSSGGKGGSYSQAACSDSAQGSDVSLTA--- : 338
           GA66A Vm RRKSS  KGGS 3QAA S1SAQGSD SL A
```

Appendix 2A. Hydrogen bonding interactions in HLA-A*0201 (PDB ID: 1OGA) complexes.

Peptide Interacting atom	HLA-A*0201 Interacting atom	GILGFVFTL (native peptide)				MAEVRSYCY[a]				FSDLYYLTM[a]			
		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}	
P1(N)	GLU 63 OE1			*	2.704								
	GLU 63 OE2											*	2.685
	TYR 7 OH					*	2.890						
P1(O)	TYR1 59 OH	*	2.766	*	2.649	*	3.011						
	TYR 7 OH											*	2.774
P2(N)	GLU 63 OE1	*	2.934	*	2.805					*	2.934		
P2(O)	LYS 66 NZ	*	2.799	*	2.999					*	2.799		
	TYR 99 OH											*	2.777
P2(OG)	HIS 70 NE2											*	2.913
	GLU 63 OE1											*	2.521
P3(N)	TYR 99 OH			*	3.300								
P3(OE1)	ARG 97 NH1							*	3.368				
P3(OD1)	ARG 97 NH1											*	3.169
P3(OD2)	ARG 97 NH2											*	2.911
P3(OE2)	ARG 97 NH1							*	2.802				
P3(O)	HIE 70 NE2			*	2.799	*	2.900						
P4(O)	LYS 66 NZ					*	2.909						
P6(OG)	THR 73 OG1					*	2.842						
P6(OH)	THR 73 OG1											*	2.953
	ASP 77 OD2											*	2.658
P7(O)	THR 73 OG1							*	2.986				
P7(OH)	TYR 116 OH					*	2.701						
P8(O)	TRP 147 NE1	*	2.874	*	2.991	*	2.871					*	2.837
	LYS 146 NZ	*	3.322	*	2.786								
P9(N)	ASP 77 OD1	*	2.880	*	2.773	*	2.941						
	ASP 77 OD2											*	3.005
P9(O)	THR 143 OG1			*	2.736								
	LYS 146.A NZ											*	2.808
P9(OXT)	LYS 146 NZ	*	3.009										
	THR 80 OG1											*	2.719
Total no of interactions		7		10		8		3		2		13	
Side chain interactions		-		_		2		2		_		6	

[a]Docked ZIKV E peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after molecular dynamics simulation (MD). Stars denote the existence of an H-bond.

Appendix 2B. Hydrogen bonding interactions in HLA-A*0201 (PDB ID: 2GIT) complexes.

Peptide Interacting atom	HLA-A*0201 Interacting atom	LLFGKPVYV (native peptide)				MAEVRSYCY[a]				FSDLYYLTM[a]			
		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}	
P1(N)	GLU 63 OE1			*	2.938								
	TYR 171 OH	*	2.951			*	2.955			*	2.914		
	GLU 63 OE2			*	2.779			*	2.751			*	2.713
	TYR 7 OH									*	2.915		
P1(O)	TYR159 OH	*	2.758	*	2.782								
P2 (N)	GLU 63 OE1	*	3.528										
	GLU 63 OE2	*	3.003	*	2.908	*	2.916			*	3.018		
P2(O)	LYS66 NZ					*	2.887			*	2.823		
P2(OG)	GLU 63 OE2											*	2.675
P3(N)	TYR 99 OH	*	3.028	*	3.394	*	3.016			*	3.024		
	HID 70 NE2							*	3.172				
P3(OE2)	HID 114 ND1					*	3.178						
P5(NH1)	GLN 155 O							*	3.025				
P6(O)	THR 73 OG1					*	3.219						
P6(OG)	ARG 97 NH2							*	2.926				
P7(O)	ARG 97 NH1					*	3.286						
P7(OH)	GLN 155 OE1					*	3.145						
P8(O)	TRP147 NE1	*	2.818			*	2.923	*	2.880	*	2.892	*	2.926
	LYS 146 NZ							*	2.951				
P8(OG1)	ASP 77 OD1											*	2.650
P9(N)	ASP 77 OD1	*	3.046	*	2.833	*	3.397			*	2.915	*	2.965
P9(O)	TYR 84 OH	*	2.846	*	2.631					*	2.801		
	THR 143 OG	*	2.750					*	2.706	*	2.760		
	LYS 146 NZ											*	2.887
	THR 143 OG1	*	3.502	*	2.855							*	2.635
P9(OH)	ASP 77 OD2							*	2.721				
P9(OXT)	LYS146 NZ	*	2.884	*	2.906			*	2.905	*	2.847		
	TYR 84 OH					*	2.798						
	THR 143 OG1					*	2.763						
Total no of interactions		11		9		12		9		10		7	
Side chain interactions		_		_		2		2		_		2	

[a]Docked ZIKV E peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after molecular dynamics simulation (MD). Stars denote the existence of an H-bond.

Appendix 2C. Hydrogen bonding interactions in HLA-A*0201 (PDB ID: 2GTW) complexes.

Peptide Interacting atom	HLA-A*0201 Interacting atom	LAGIGILTV (native peptide)				MAEVRSYCY[a]				FSDLYYLTM[a]			
		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}	
P1(N)	GLU 63 OE1			*	3.536			*	2.750				
	GLU 63 OE2	*	3.100	*	2.699	*	3.100			*	3.100	*	2.696
P1(O)	LYS 66 NZ	*	2.951			*	2.951			*	2.950	*	2.832
P2(O)	HIS 70 NE2							*	2.866			*	2.773
P2(OG)	TYR 99 OH											*	2.910
P5 (O)	HIS 114 NE2			*	2.994								
P6(O)	THR 73 OG1							*	2.614			*	2.640
P7(O)	ARG 97 NH1	*	3.216	*	3.078	*	3.216			*	3.215		
P7(OH)	GLN 155 NE2					*	2.462						
P8(O)	TRP147 NE1	*	2.887	*	2.882	*	2.887			*	2.887	*	2.880
P9(N)	ASP 77 OD1	*	2.834	*	2.877	*	2.834	*	2.811	*	2.835S	*	2.809
P9(O)	LYS 146 NZ			*	3.060			*	2.832			*	3.075
P9(OXT)	LYS146 NZ			*	2.791							*	3.127
	THR 143 OG1			*	2.617							*	2.625
Total no of interactions		5		9		6		5		5		10	
Side chain interactions		_		_		1		_		_		1	

[a]Docked ZIKV E peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after molecular dynamics simulation (MD). Stars denote the existence of an H-bond.

Appendix 2D. Hydrogen bonding interactions in HLA-A*0201 (PDB ID: 2GTZ) complexes.

Peptide Interacting atom	HLA-A*0201 Interacting atom	ALGIGILTV (native crystal)				MAEVRSYCY[a]				FSDLYYLTM[a]			
		d^b_{init}		d^c_{MD}		d^b_{init}		d^c_{MD}		d^b_{init}		d^c_{MD}	
P1(N)	GLU 63 OE1			*	2.746								
	GLU 63 OE2			*	2.940			*	2.705			*	2.661
	TYR 59 OH			*	3.129								
P1(O)	TYR 159 OH	*	2.673			*	3.365			*	2.963	*	2.736
P2(N)	GLU 63 OE2	*	2.900	*	2.818	*	2.918			*	2.873		
P2(O)	LYS 66 NZ	*	2.940	*	2.893	*	2.885S			*	2.813		
P2(OG)	GLU 63 OE1											*	2.601
	GLU 63 OE2											*	3.073
P3(OD1)	LYS 66 NZ											*	2.775
P3(OE1)	HIE 70 NE2							*	2.892				
P3(OD2)	TYR 99 OH											*	2.723
P5(O)	GLN 155 NE2							*	2.804				
P5(NH1)	THR 163 OG1							*	3.059				
P6(O)	THR 73 OG1					*	3.336						
	HIE 114 NE2											*	2.978
	ARG 97 NH2											*	2.810
P6(OG)	HIE 114 NE2							*	3.037				
P7(O)	ARG 97 NH1					*	3.274			*	3.196		
	ARG 97 NH2											*	2.797
P8(O)	TRP147 NE1	*	2.885			*	2.905			*	2.962	*	3.214
	LYS 146 NZ			*	2.994			*	2.879			*	2.882
P8(OG1)	LYS 146 NZ			*	2.872								
	ASP 77 OD1											*	2.771
P9(N)	ASP 77 OD1	*	2.889	*	2.905	*	3.366			*	3.215	*	2.921
P9(O)	LYS 146 NZ			*	2.848							*	2.761
P9(OXT)	THR 143 OG1			*	2.695								
Total no of interactions		5		10		7		6		6		14	
Side chain interactions		-		1				3		-		4	

[a]Docked ZIKV E peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after Molecular dynamics simulation (MD). Stars denote the existence of an H-bond.

Appendix 2E. Hydrogen bonding interactions in HLA-A*0201 (PDB ID: 3I6G) complexes.

Peptide Interacting atom	HLA-A*0201 Interacting atom	GLMWLSYFV (native crystal)				MAEVRSYCY[a]				FSDLYYLTM[a]			
		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}	
P1(N)	GLU 63 OE1			*	2.974			*	2.670			*	3.439
	GLU 63 OE2			*	2.806			*	3.092			*	2.695
P1(O)	TYR 159 OH	*	2.698	*	2.705	*	2.799			*	2.739	*	2.819
P2(N)	GLU 63 OE1	*	2.698	*	2.799					*	2.842	*	2.754
P2(O)	LYS 66 NZ	*	2.759	*	3.030								
P2(OG)	GLU 63 OE1											*	2.606
P3(N)	TYR 99 OH			*	3.251								
P3(O)	HIS 70 NE2			*	2.828							*	2.889
P3(OE1)	HIS 114 NE2							*	2.876				
	ARG 97 NH1							*	2.854				
P3(OE2)	ARG 97 NH1							*	3.363				
	ARG 97 NH2							*	2.812				
P4(O)	LYS 66 NZ					*	2.842						
P5(NH1)	ALA 150 O							*	2.783				
	GLN 155 OE1							*	2.845				
P6(OG)	ARG 97 NH2	*	2.698										
	THR 73 OG1					*	2.801						
P7(O)	ARG 97 NH2											*	2.938
P8(O)	TRP147 NE1	*	2.698	*	2.828	*	2.891	*	2.927	*	3.019	*	2.866
	LYS 146 NZ											*	3.341
	ASP 77 OD1											*	2.670
P9(N)	ASP 77 OD1	*	2.698	*	3.003	*	2.904	*	3.080	*	2.885	*	2.821
	TYR 384 N							*	3.488				
P9(O)	TYR 84 OH			*	3.311								
	LYS 146 NZ			*	2.985	*	2.900						
P9(OXT)	LYS 146 NZ			*	2.828							*	2.780
	TYR 84 OH			*	2.740								
	THR 143 OG1			*	2.760			*	2.649				
Total no of interactions		6		14		6		12		4		12	
Side chain interactions		1		_		1		6		_		1	

[a]Docked ZIKV E peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after Molecular dynamics simulation (MD). Stars denote the existence of an H-bond.

Appendix 2F. Hydrogen bonding interactions in HLA-A*0201 (PDB ID: 3TO2) complexes.

Peptide Interacting atom	HLA-A*0201 Interacting atom	LACFVLAAV (native crystal)				MAEVRSYCY[a]				FSDLYYLTM[a]			
		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}	
P1(N)	GLU 63 OE1			*	2.691							*	2.689
	TYR 7 OH	*	2.902.							*	2.746		
	TYR 159 OH	*	2.169							*	3.374		
P1(O)	TYR 7 OH					*	2.826					*	2.781
P2(N)	GLU 63 OE1	*	2.794			*	2.848			*	2.895		
	GLU 63 OE2	*	2.725					*	2.704				
P2(O)	LYS 66 NZ												
P2(OG)	GLU 63 OE2											*	2.592
P3(N)	TYR 99 OH	*	3.038										
P3(O)	HIE 70 NE2					*	2.915						
	TYR 159 OH											*	2.697
P3(OD1)	LYS 66 NZ											*	2.944
	LYS 66 NZ											*	2.943
P3(OE1)	HIE 114 NE2							*	2.897				
	ARG 97 NH1							*	2.957				
P3(OE2)	HIS 74 NE2							*	2.856				
	HIS 70 NE2							*	2.763				
	ARG 97 NH1							*	2.984				
P6(O)	THR 73 OG1					*	3.379						
	HIE 114 NE2			*	2.993								
	ARG 97 NH1											*	2.964
	ARG 97 NH2											*	2.914
	ARG 97 NE			*	2.910								
P6(OG)	ARG 97 NH1					*	3.074						
	THR 73 OG1					*	2.969						
P7(O)	ARG 97 NH2			*	2.810								
P8(O)	TRP147 NE1	*	2.728	*	2.788	*	2.798			*	2.988	*	2.914
P8(N)	ASP 77 OD1			*	2.993								
P8(OG1)	ASP 77 OD1	*	2.955										
P9(N)	ASP 77 OD1									*	2.830		
P9(O)	LYS 146 NZ	*	2.915									*	2.832
P9(OH)	ASP 77 OD1							*	2.753				
	ARG 97 NH2					*	2.884						
P9(OXT)	LYS146 NZ	*	3.287	*	2.899							*	2.876
	TYR 84 OH	*	2.589	*	2.968								
	THR 143 OG1	*	2.773	*	2.680								
Total no of interactions		11		9		8		7		5		11	
Side chain interactions		1		_		3		6		_		3	

[a]Docked ZIKV E peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after Molecular dynamics simulation (MD). Stars denote the existence of an H-bond.

Appendix 2G. Hydrogen bonding interactions in HLA-A*0201 (PDB ID: 4K7F) complexes.

Peptide Interacting atom	HLA-A*0201 Interacting atom	VSWGQLMNL (native crystal)				MAEVRSYCY[a]				FSDLYYLTM[a]			
		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}	
P1(N)	GLU 62 OE1			*	2.739								
	GLU 62 OE2			*	2.857							*	2.702
P1(O)	TYR 158 OH	*	2.671	*	2.715	*	2.823			*	2.806	*	2.712
P2(N)	GLU 62 OE1					*	2.864			*	2.883	*	2.761
	GLU 62 OE2			*	2.779							*	3.447
P2(OG)	GLU 62 OE1											*	2.540
P2(O)	LYS 65 NZ	*	2.582										
	TYR 158 OH							*	2.975				
P2(N)	GLU 62 OE2	*	3.067										
P2(SG)	GLU 62 OE2			*	3.223								
P3(N)	TYR 98 OH			*	2.978								
P3(O)	HIE 69 NE2	*	2.956							*	3.089	*	3.059
	TYR 98 OH			*	2.955								
P3(OD1)	ARG 96 NH2											*	3.540
	ARG 96 NH1											*	2.812
P3(OD2)	ARG 96 NH1											*	3.337
	ARG 96 NH2											*	2.756
P3(OE1)	HIE 113 NE2							*	2.813				
P3(OE2)	ARG 96 NH1							*	2.781				
	HIS 69 NE2							*	2.783				
P5(O)	GLN 154 NE2			*	3.050								
P5(NE)	GLN 154 O							*	2.908				
P5(NH2)	GLN 154 O							*	3.551				
P6(O)	THR 72 OG1			*	3.232								
	ARG 96 NH1					*	2.925						
	THR 72 OG1					*	3.230						
P7(O)	TRP 146 NE1											*	2.970
	ARG 96 NH2					*	2.925						
P8(O)	TRP146 NE1	*	3.028	*	2.862	*	2.849	*	2.864	*	2.840		
P8(N)	ASP 76 OD1	*	2.758										
P8(ND2)	ASP 76 OD1			*	2.865								
P8(OG1)	ASP 76 OD1			*	2.869								
P9(N)	ASP 76 OD1					*	2.834			*	2.892	*	2.859
P9(O)	LYS 145 NZ							*	2.832				
P9(OH)	ARG 96 NE					*	2.917						
	ARG 96 NH2					*	3.122						
P9(OXT)	TYR 83 OH											*	2.886
	THR 142 OG1			*	2.675			*	2.740				
Total no of interactions		6		13		9		9		5		13	
Side chain interactions		1		3		2		5				5	

[a]Docked ZIKV E peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after Molecular dynamics simulation (MD). Stars denote the existence of an H-bond.

Appendix 2H. Hydrogen bonding interactions in HLA-A*0201 (PDB ID: 5SWQ) complexes.

Peptide Interacting atom	HLA-A*0201 Interacting atom	CVNGSCFTV (native crystal)				MAEVRSYCY[a]				FSDLYYLTM[a]			
		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}		d^{b}_{init}		d^{c}_{MD}	
P1(N)	GLU 63 OE1			*	2.741							*	2.911
	GLU 63 OE2			*	2.911			*	2.685			*	2.843
P1(O)	TYR 159 OH	*	2.480	*	2.708	*	2.809			*	2.840		
	TYR 7 OH											*	2.799
P2(N)	GLU 63 OE2			*	2.742								
P2(N)	GLU 63 OE1	*	3.153							*	2.895		
P2(OG)	GLU 63 OE1											*	2.701
P2(O)	LYS 66 NZ	*	2.687										
	TYR 99 OH											*	2.772
P2(N)	GLU 63 OE1					*	2.941						
P3(N)	TYR 99 OH			*	3.044								
P3(OD1)	HIS 70 NE2									*	2.704		
P4(O)	LYS 66 NZ									*	2.920		
P5(NH1)	ALA 150 O							*	2.858				
	GLN 155 OE1							*	2.799				
P6(OG)	HIE 70 NE2							*	3.112				
P7(O)	ARG 97 NH1							*	3.378			*	2.891
	ARG 97 NH2											*	2.864
P8(O)	TRP147 NE1	*	2.846	*	2.916	*	2.674	*	3.015	*	2.719	*	2.989
	LYS 146 NZ	*	2.866										
P8(N)	ASP 77 OD1	*	2.889										
P9(N)	ASP 77 OD1			*	2.851			*	2.931	*	3.015		
P9(O)	LYS 146 NZ			*	2.965								
	THR 143 OG1											*	2.599
P9(OH)	ARG 97 NE					*	3.149						
	ARG 97 NH2					*	2.961						
P9(OXT)	THR 142 OG1			*	2.675								
Total no of interactions		6		9		5		7		6		9	
Side chain interactions		_		_		_		3		1		1	

[a]Docked ZIKV E peptides, [b]Initial distance (after FlexPepDock) between the H-bond donor and acceptor; measured with the FindHBond tool in Chimera (H-bond constraints were relaxed by 0.4 and 20.0 Å degrees), [c]distance between the H-bond donor and acceptor after Molecular dynamics simulation (MD). Stars denote the existence of an H-bond.

Appendix 3. Superimposition of the initial and simulated complexes: the F pocket size.

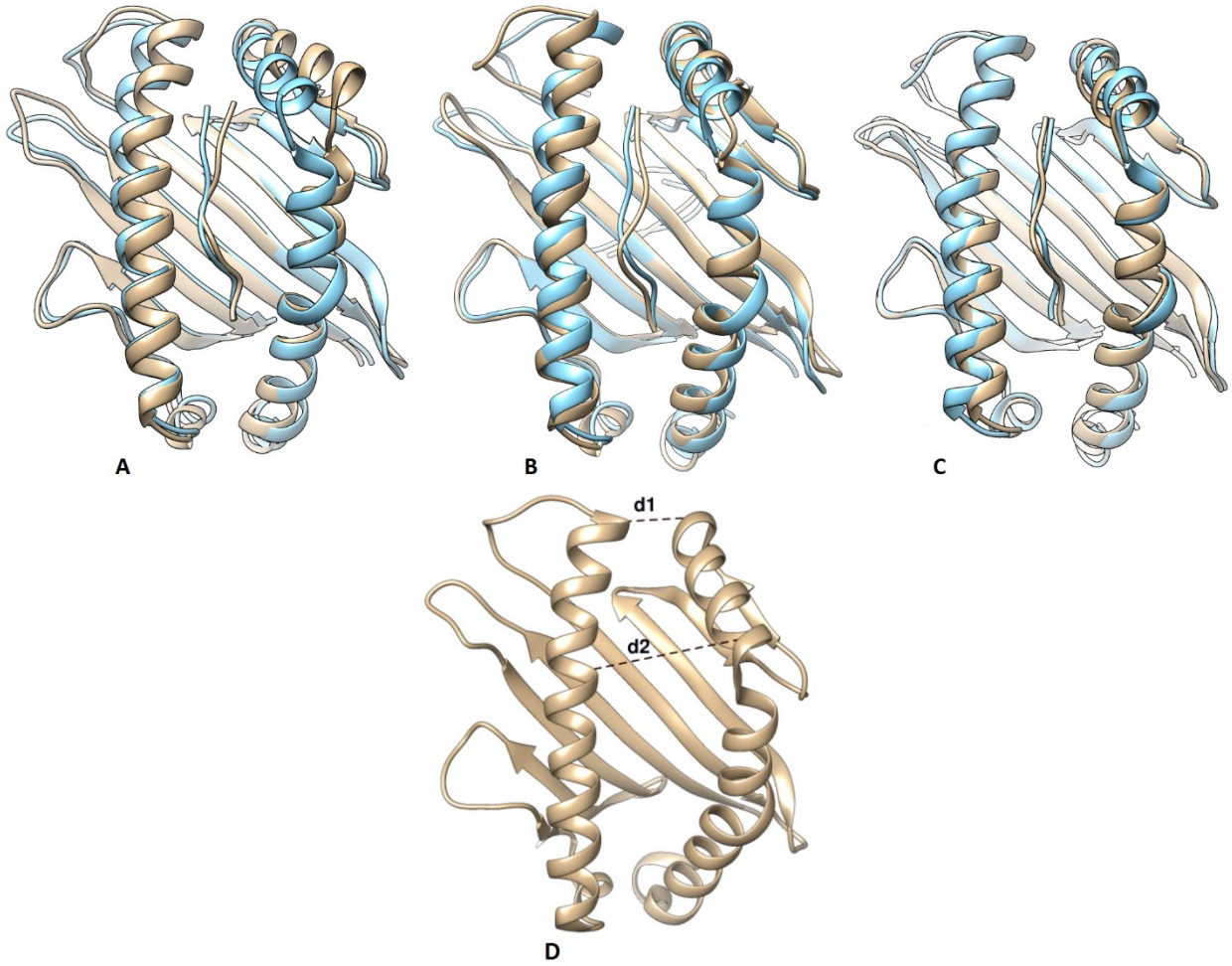


Figure 11.1. (A-B) E protein peptides from ZIKV in the binding groove of MHC I HLA-A*0101; (C) Crystal complex of HLA-A*0101 with Influenza A virus (H1N1) peptide (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) TMNNKHWLV; (B) GLDFSDLYY; (C) CTELKLSDY (PDB ID: 4NQV). D: MHC I binding groove (F pocket) (cartoon representation colored in tan, model before minimization); distance 1 (d1) and distance 2(d2) are shown as black dashed lines. Visualization with Chimera v. 1.13 (Pettersen et al., 2004).

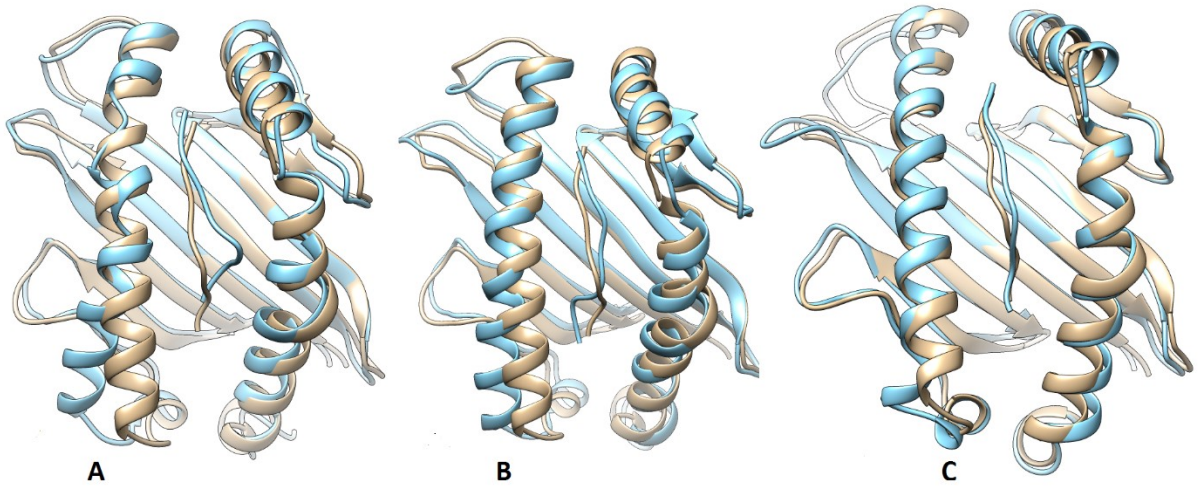


Figure 11.2. (A-B) E protein peptides from ZIKV in the binding groove of MHC I HLA-A*0201; (C) Crystal complex of HLA-A*0201 with Influenza A virus peptide (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) MAEVRSYCY; (B) FSDLYYLTM; (C) GILGFVFTL (PDB ID: 1OGA). Visualization with Chimera v. 1.13 (Pettersen et al., 2004).

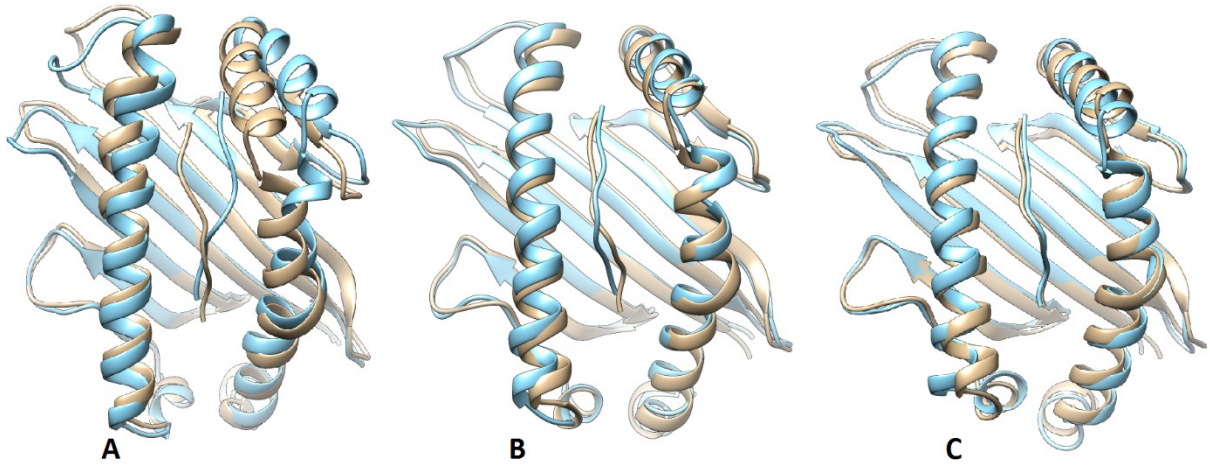


Figure 11.3. (A-B) E protein peptides from ZIKV in the binding groove of MHC I HLA-A*0201; (C) Crystal complex HLA-A*0201 with transcriptional activator TAX peptide (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) MAEVRSYCY; (B) FSDLYYLTM; (C) LLFGKPVYV (PDB ID: 2GIT). Visualization with Chimera v. 1.13 (Pettersen et al., 2004).

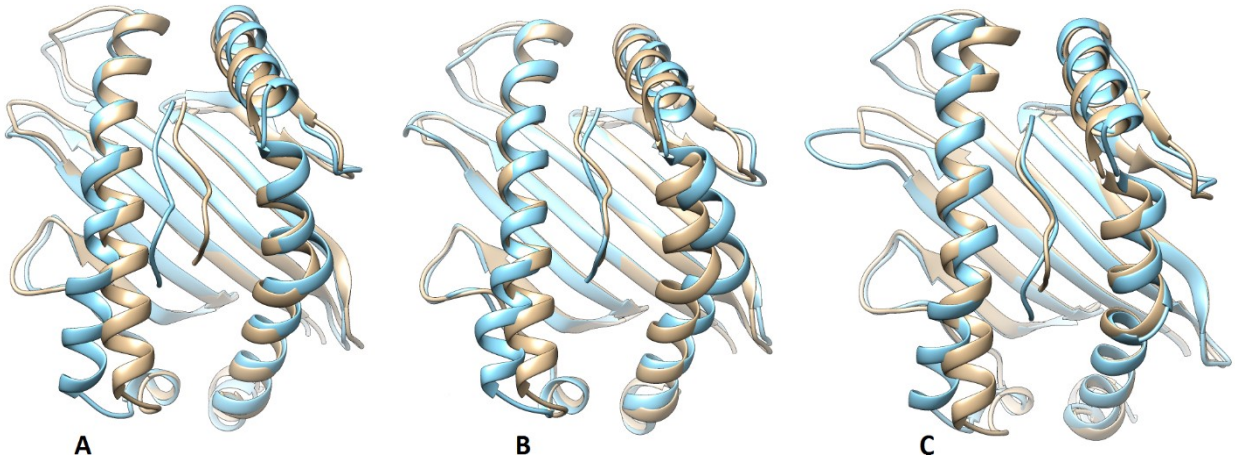


Figure 11.4. (A-B) E protein peptides from ZIKV in the binding groove of MHC I HLA-A*0201; (C) Crystal complex of HLA-A*0201 with melanoma peptide (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) MAEVRSYCY; (B) FSDLYYLTM; (C) LAGIGILTV (PDB ID: 2GTW). Visualization with Chimera v. 1.13 (Pettersen et al., 2004).

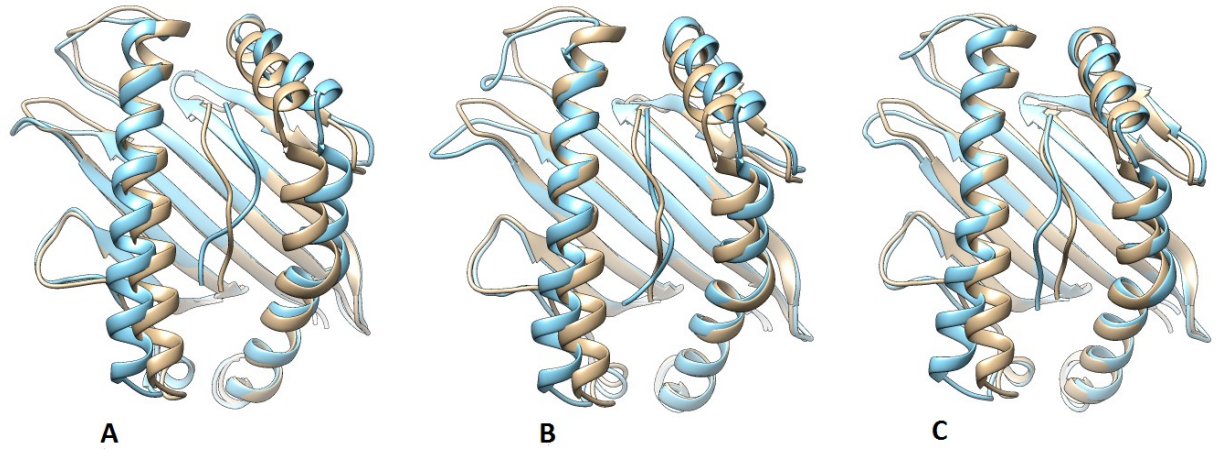


Figure 11.5. (A-B) E protein peptides from ZIKV in the binding groove of MHC I HLA-A*0201; (C) Crystal complex of HLA-A*0201 with melanoma peptide (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) MAEVRSYCY; (B) FSDLYYLTM; (C) ALGIGILTV (PDB ID: 2GTZ). Visualization with Chimera v. 1.13 (Pettersen et al., 2004).

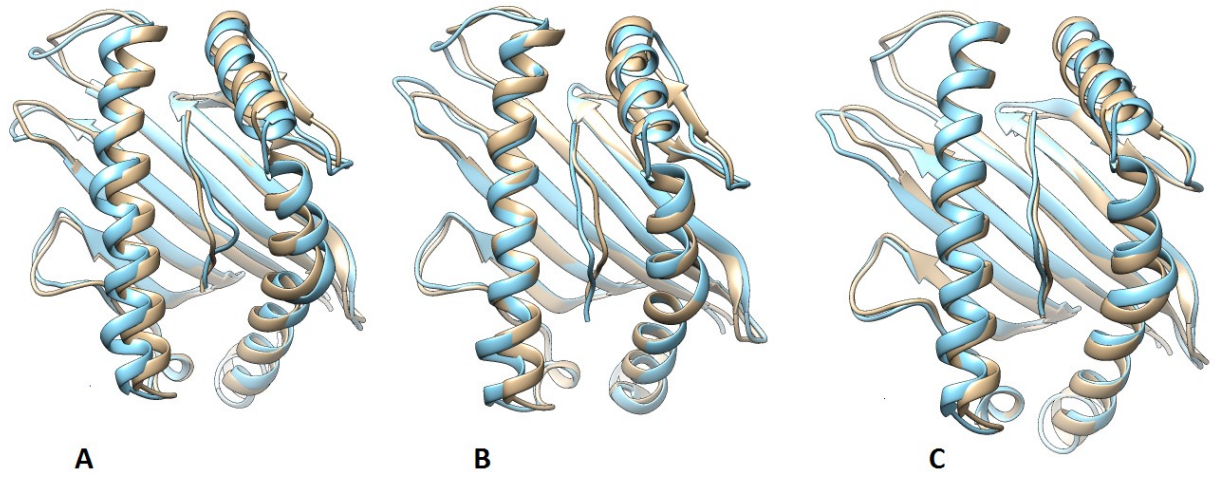


Figure 11.6. (A-B) E protein peptides from ZIKV in the binding groove of MHC I HLA-A*0201; (C) Crystal complex of HLA-A*0201 with a peptide from membrane protein of SARS virus (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) MAEVRSYCY; (B) FSDLYYLTM; (C) GLMWLSYFV (PDB ID: 3I6G). Visualization with Chimera v. 1.13 (Pettersen et al., 2004).

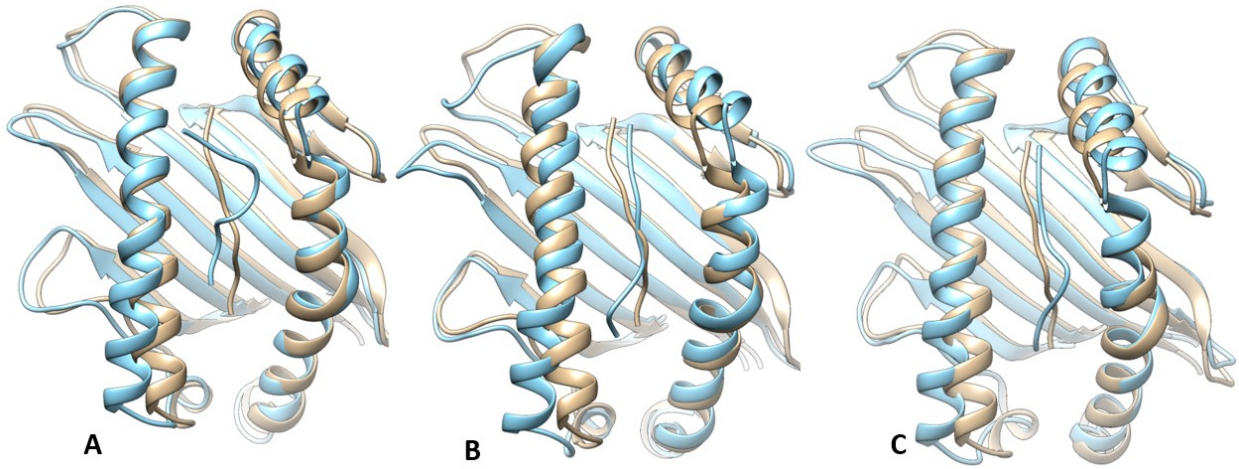


Figure 11.7. (A-B) E protein peptides from ZIKV in the binding groove of MHC I HLA-A*0201; (C) Crystal complex of HLA-A*0201 with membrane glycoprotein peptide (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) MAEVRSYCY; (B) FSDLYYLTM; (C) LACFVLAAV (PDB ID: 3TO2). Visualization with Chimera v. 1.13 (Pettersen et al., 2004).

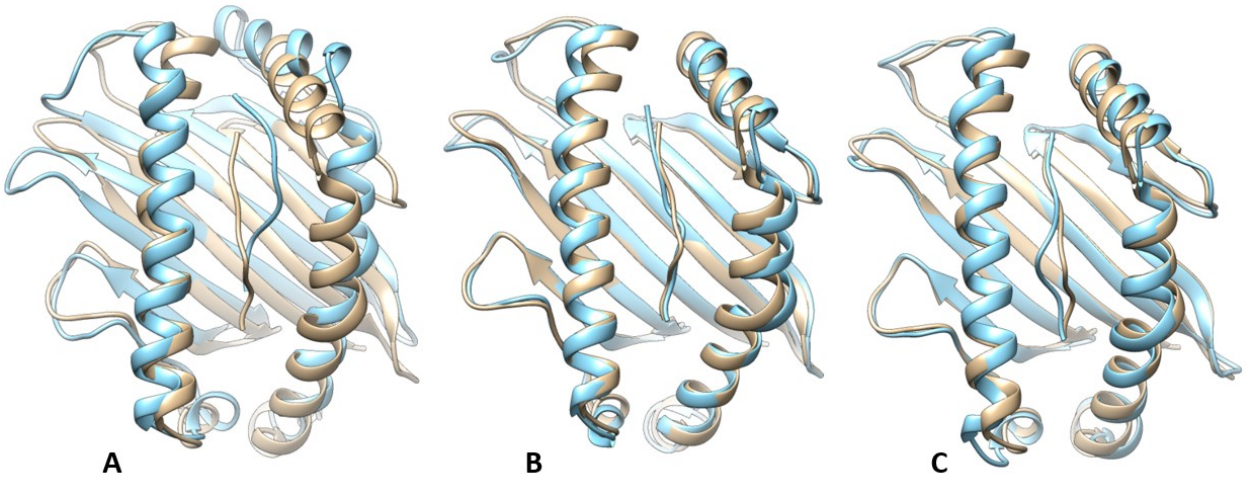


Figure 11.8. (A-B) E protein peptides from ZIKV in the binding groove of MHC I HLA-A*0201; (C) Crystal complex of HLA-A*0201 with Hepatitis B virus peptide (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) MAEVRSYCY; (B) FSDLYYLTM; (C) VCWGELMNL (PDB ID: 4K7F). Visualization with Chimera v. 1.13 (Pettersen et al., 2004).

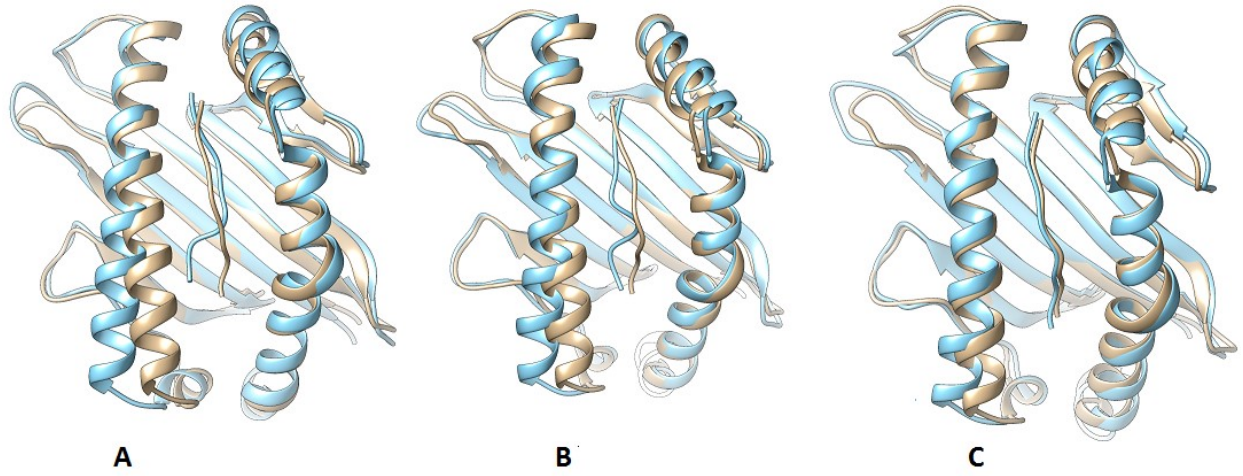


Figure 11.9. (A-B) E protein peptides from ZIKV in the binding groove of MHC I HLA-A*0201; (C) Crystal complex of HLA-A*0201 with NA231 influenza epitope (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) MAEVRSYCY; (B) FSDLYYLTM; (C) CVNGSCFTV (PDB ID: 5SWQ). Visualization with Chimera v. 1.13 (Pettersen et al., 2004).

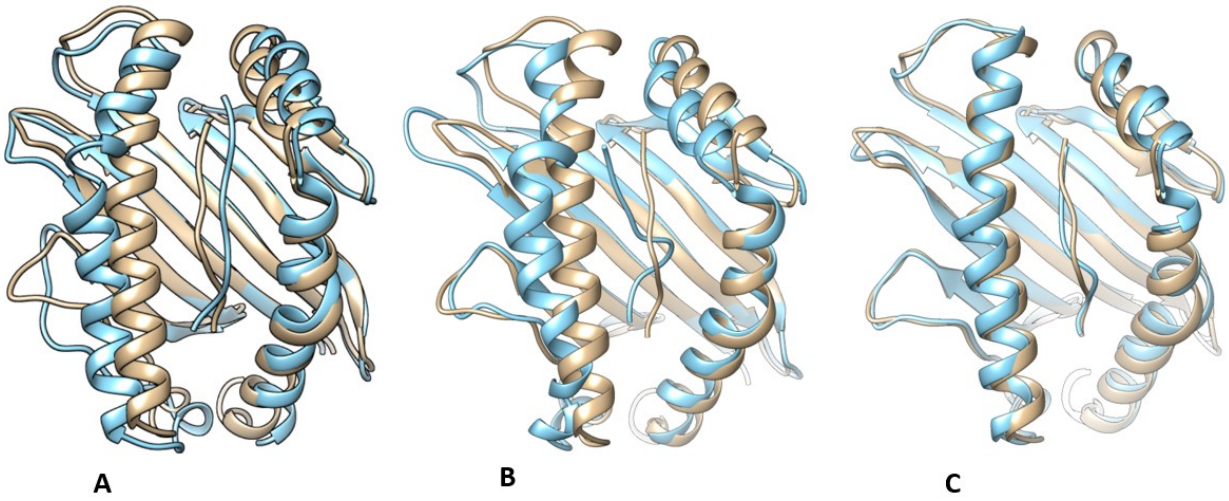


Figure 11.10. (A-B) NS3 protein peptides from ZIKV in the binding groove of MHC I HLA-B*2705; (C) Crystal complex of HLA-B*2705 with Influenza A virus peptide (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) DIGAVALDY; (B) HSEVQLLAV; (C) SRYWAIRTR (PDB ID: 2BST). Visualization with Chimera v. 1.13 (Pettersen et al., 2004).

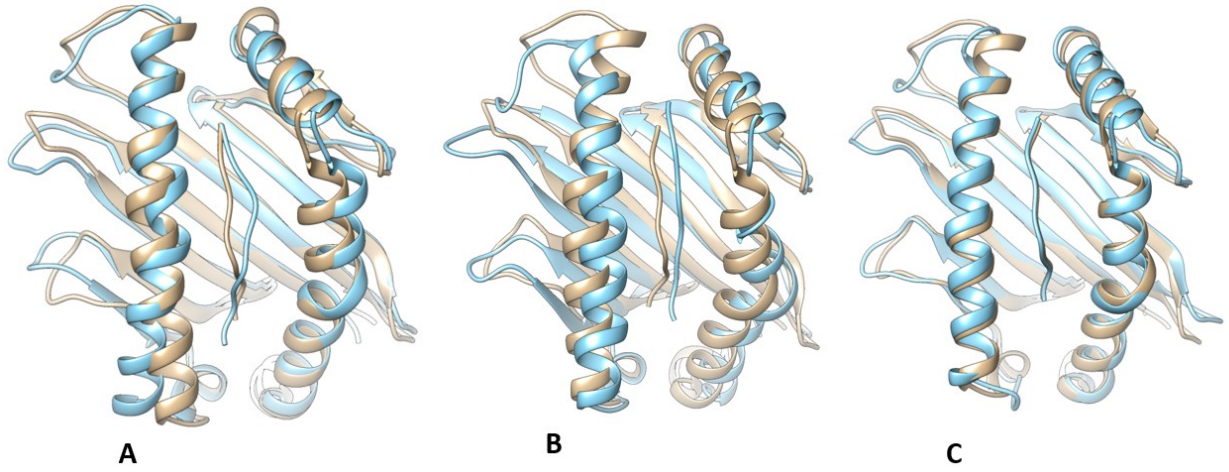


Figure 11.11. (A-B) NS5 protein peptides from ZIKV in the binding groove of MHC I HLA-C*0801; (C) Crystal complex of HLA-C*0801 with Influenza A virus peptide (cartoon representation); the complex obtained from docking or Protein Data Bank (in tan) is superimposed with the complex obtained after the 10-ns MD simulation and subsequent energy minimization (in cyan): (A) MTTEDMLVV; (B) FTNLVVQLI; (C) GILGFVFTL (PDB ID: 4NT6). Visualization with Chimera v. 1.13 (Pettersen et al., 2004).