De novo design of therapeutic peptides and their characterization

by

Ravindra Thakkar

B.Sc., Sardar Patel University, 2004
M.Sc., Sardar Patel University, 2006

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Anatomy and Physiology
College of Veterinary Medicine

KANSAS STATE UNIVERSITY
Manhattan, Kansas, USA

2022

Abstract

Peptides are low-cost, flexible, and biocompatible and can be designed to serve various functions in biotechnology and medicine. Peptides can be designed such that they fold spontaneously and adopt a specific conformation under specific conditions, including when in contact with the three-dimensional structure of a protein or the two-dimensional structure of graphene. They are promising for design of functional materials for biotechnology and medical applications. I have studied peptide design for biotechnology, including peptide self-assembly on a graphene surface, and for medical applications such as cancer immunotherapy and treatment of coronavirus disease 2019 caused by SARS-CoV-2.

In chapter 1, I describe my study of the self-assembly of a designed cyclic peptide on graphitic surfaces by molecular dynamics simulations. In experiments, it was found that hydrocarbon contaminants may interfere with this self-assembly, so we undertook a computational study of the behavior of these contaminants at the graphene–water interface and compared it to experimental data, as detailed in chapter 2.

Peptides are also promising in medicine, particularly for inhibiting protein-protein interactions in situations where conventional small-molecule drugs can be unsuitable. Many viruses important for public health including SARS-CoV-2 and influenza enter cells by means of binding between viral proteins and cell surface proteins. The blockade of these undesirable protein-protein interactions has definite clinical significance. Another medical application where blocking protein-protein interactions is essential is the immune checkpoint blockade used in cancer

immunotherapy. Immune checkpoint proteins most studied for cancer immunotherapy have flat and relatively hydrophobic interfaces that have impeded small-molecule drug development. Therefore, the application of peptide molecules that mimic the interacting surface of a natural binding protein is a promising alternative to small- molecule drugs.

Immunotherapy activates the patient's own immune system to treat cancer. When any foreign substance enters in the body, immune cells recognize it as a threat and neutralize it. But unfortunately, cancer cells often evolve to evade the immune system. Cytotoxic T-Lymphocyte Associated protein 4 (CTLA4) plays a crucial role in self-recognition and is an immune checkpoint protein that cancer cells may express to prevent attack from the immune system. Cancer cells frequently overexpress proteins of the B7 family, which allows them to evade the immune response by binding between these B7 proteins and CTLA4 on the surface of T cells. As presented in chapter 3, I have designed a 17-residue cyclic peptide targeting the CTLA4 protein that binds to it with a significant affinity. The binding activity was experimentally confirmed by the bio-layer interferometry (BLI) method. Studies performed by our collaborators showed an increase in CD8+ T cell-induced death of Lewis Lung Carcinoma (LLC) cells due to treatment with this peptide *in vitro*. *In vivo*, the designed peptide attenuated tumor growth in mouse models using orthotopic LLC cell allografts.

A disease caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), called as COVID-19, has threatened global public health and the global economy. The WHO has reported 434 confirmed million cases and 6 million deaths. Although effective vaccines have been developed against SARS-CoV-2, many regions in the world still have a low rate of vaccination

and even vaccinated individuals may experience reinfection. Deaths continue to be reported worldwide, exacerbated by continued mutation of the viral spike protein. SARS-CoV-2 enters the host cell through association of this spike protein, present on the envelope of the virus, and Angiotensin Converting Enzyme (ACE2), a protein expressed on the surface of host cells. As detailed in chapter 4, I have designed a 17-residue long peptide targeting the receptor-binding domain (RBD) of the spike protein to prevent COVID-19 infection. My designed peptide binds to the spike protein RBD with nanomolar affinity and blocks the binding site of ACE2. I have confirmed the binding activity using a microcantilever-based method and determined the dissociation constant using a BLI system. SARS-CoV-2 continues to mutate and produce variants. I have tested the binding activity of the designed peptide for the Delta variant, considered highly transmissible and declared as a variant of the concern (VOC) by the WHO. The BLI experiment revealed weaker binding of the designed peptide for the Delta variant spike protein compared to that for the original wild-type due to the mutations present in the receptor-binding domain of the spike protein.

De novo design of therapeutic peptides and their characterization

by

Ravindra Thakkar

B.Sc., Sardar Patel University, 2004
M.Sc., Sardar Patel University, 2006

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Anatomy and Physiology
College of Veterinary Medicine

KANSAS STATE UNIVERSITY
Manhattan, Kansas, USA

2022

Approved by:

Major Professor
Jeffrey R. Comer

Copyright

Abstract

Peptides are low-cost, flexible, and biocompatible and can be designed to serve various functions in biotechnology and medicine. Peptides can be designed such that they fold spontaneously and adopt a specific conformation under specific conditions, including when in contact with the three-dimensional structure of a protein or the two-dimensional structure of graphene. They are promising for design of functional materials for biotechnology and medical applications. I have studied peptide design for biotechnology, including peptide self-assembly on a graphene surface, and for medical applications such as cancer immunotherapy and treatment of coronavirus disease 2019 caused by SARS-CoV-2.

In chapter 1, I describe my study of the self-assembly of a designed cyclic peptide on graphitic surfaces by molecular dynamics simulations. In experiments, it was found that hydrocarbon contaminants may interfere with this self-assembly, so we undertook a computational study of the behavior of these contaminants at the graphene–water interface and compared it to experimental data, as detailed in chapter 2.

Peptides are also promising in medicine, particularly for inhibiting protein-protein interactions in situations where conventional small-molecule drugs can be unsuitable. Many viruses important for public health including SARS-CoV-2 and influenza enter cells by means of binding between viral proteins and cell surface proteins. The blockade of these undesirable protein-protein interactions has definite clinical significance. Another medical application where blocking protein-protein interactions is essential is the immune checkpoint blockade used in cancer

immunotherapy. Immune checkpoint proteins most studied for cancer immunotherapy have flat and relatively hydrophobic interfaces that have impeded small-molecule drug development. Therefore, the application of peptide molecules that mimic the interacting surface of a natural binding protein is a promising alternative to small- molecule drugs.

Immunotherapy activates the patient's own immune system to treat cancer. When any foreign substance enters in the body, immune cells recognize it as a threat and neutralize it. But unfortunately, cancer cells often evolve to evade the immune system. Cytotoxic T-Lymphocyte Associated protein 4 (CTLA4) plays a crucial role in self-recognition and is an immune checkpoint protein that cancer cells may express to prevent attack from the immune system. Cancer cells frequently overexpress proteins of the B7 family, which allows them to evade the immune response by binding between these B7 proteins and CTLA4 on the surface of T cells. As presented in chapter 3, I have designed a 17-residue cyclic peptide targeting the CTLA4 protein that binds to it with a significant affinity. The binding activity was experimentally confirmed by the bio-layer interferometry (BLI) method. Studies performed by our collaborators showed an increase in CD8+ T cell-induced death of Lewis Lung Carcinoma (LLC) cells due to treatment with this peptide *in vitro*. *In vivo*, the designed peptide attenuated tumor growth in mouse models using orthotopic LLC cell allografts.

A disease caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), called as COVID-19, has threatened global public health and the global economy. The WHO has reported 434 confirmed million cases and 6 million deaths. Although effective vaccines have been developed against SARS-CoV-2, many regions in the world still have a low rate of vaccination

and even vaccinated individuals may experience reinfection. Deaths continue to be reported worldwide, exacerbated by continued mutation of the viral spike protein. SARS-CoV-2 enters the host cell through association of this spike protein, present on the envelope of the virus, and Angiotensin Converting Enzyme (ACE2), a protein expressed on the surface of host cells. As detailed in chapter 4, I have designed a 17-residue long peptide targeting the receptor-binding domain (RBD) of the spike protein to prevent COVID-19 infection. My designed peptide binds to the spike protein RBD with nanomolar affinity and blocks the binding site of ACE2. I have confirmed the binding activity using a microcantilever-based method and determined the dissociation constant using a BLI system. SARS-CoV-2 continues to mutate and produce variants. I have tested the binding activity of the designed peptide for the Delta variant, considered highly transmissible and declared as a variant of the concern (VOC) by the WHO. The BLI experiment revealed weaker binding of the designed peptide for the Delta variant spike protein compared to that for the original wild-type due to the mutations present in the receptor-binding domain of the spike protein.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Back in 2008, I was analyzing negatively stained beta-amyloid proteins under the transmission electron microscope as my first project in my professional career in electron microscopy. I was astonished to see the beta sheet-like structure depicted in the textbooks. I have always been curious to see these beautiful and intricate protein molecules in action. However, the analysis of protein dynamics at the atomic level using electron microscopy is still challenging due to the harsh and unfavorable environment of the transmission electron microscope chamber for living cell activities.

In 2017, I attended a talk presented by Dr. Jeffrey Comer. He presented a study on the adsorption of small molecules and amino acids on graphene surfaces by molecular dynamics simulation. Here, I learned that the molecular dynamics simulations could be used as a "computational microscope" to lead the research where experiments cannot be implemented. His talk inspired me, and the idea of protein design for therapeutic application was shaped in my mind. I approached Dr. Comer and expressed my interest in learning molecular dynamics and joining his lab as a graduate student. With his acceptance, my journey as a graduate research scholar had started.

I want to thank all the people who helped me in my personal and professional journey. First of all, I would like to thank my Ph.D. committee, Dr. Jeffrey Comer, Dr. Masaaki Tamura, Dr. Ryan Rafferty, and Dr. Santosh Aryal. I especially want to thank my major advisor, Dr. Comer, for providing me an opportunity to research under his guidance and giving shape to my research ideas; teaching me computer coding, molecular modeling, molecular dynamics simulation, and free energy calculation techniques; nurturing my scholarly activities like presentation, critical thinking abilities; and being an amazing mentor. Without his continued and ingenious guidance, I

wouldn't have made it this far. I also thank Drs. Tamura, Rafferty, and Aryal provided me with their continued support and critical feedback and made me aware of real-world science. Especially I would like to thank Dr. Tamura for collaborating to check the efficacy of my CTLA-4 inhibitory peptide on cell culture and animal experiments and taking the lead to test the biological activities of the peptide pep39. Dr. Aryal, thank you for guiding me and suggesting some scientific inputs to address critical scientific questions and providing care as a guardian. Dr. Rafferty, thank you for giving your brainstorming input towards improving the stability of designed structures.

I would like to thank the rest of Dr. Comer's group, including Dr. Azigiya Singam, Sandun Gajaweera, and Dr. Lina María Preciado, for providing a supportive environment. Especially, Singam, I thank you for teaching the Linux commands and introducing me to cool molecular modeling techniques. I greatly appreciate your selfless help you provided despite your busy schedule at your new employer. Sandun, thanks for coordinating with the logistics in my absence. I also thank Dr. Tamura's group, including Deepa Upreti, Dr. Susumu Ishiguro, Mayme Loyd, Nicole Robben, Paige Cote, Molly Bassette, and Morgan Phillips, for providing wonderful presentations on experimental biology and explaining your research activities in the journal club meetings. Especially Deepa and Susumu, thanks for testing my designed peptide for its biological activities *in vitro* and *in vivo* conditions.

I am grateful to Dr. Hans Coetzee, head of the department, for approving my requisition to pursue a Ph.D. program along with my full-time employment. Also, I am thankful to Dr. Nancy A. Monteiro-Riviere, who provided me an opportunity to work at Kansas State University and introduced me to an amazing team of the center's NICKS and ICCM. I want to thank Drs. Peying Fong and T. G. Nagaraja for providing and updating information about the curriculum requirements for the Ph.D. program. I would like to thank Dr. Brian Giesbrecht for giving his time

to run a binding experiment for one of the projects. I thank Dr. Meena Kumari for encouraging me, providing her technical suggestions, and bringing my attention to assess possibilities of the incapability of the buffer solution for the selected binding experiments. Especially, I thank Dr. Raghavendra Amachawadi for suggesting courses to develop my scientific and professional growth. I also thank Gail Eyestone and Dan Galbraith for providing administrative support. I also thank Dr. Govind, Division of Biology, for granting me access to the FortéBio BLItz instrument for binding experiments.

I am lucky to have a skillful professional as my friend; they are a real treasure. Expressly, I am gratified to Dr. Dilip Agarwal, a post-doctoral fellow at Northwestern University who helped to perform the binding assay of my designed stapled peptide by the micro-cantilever-based method. His spectacular efforts have revived the protein design project. I am thankful to Dr. Chathurange Ranaweera for teaching me the basics of techniques to determine binding affinity and his altruistic help to carry out bio-layer interferometry experiments to determine the dissociation constant for my designed peptides. I am also thankful to Dr. Ashish Kumar for providing his thoughtful inputs and being a helpful classmate for some Ph.D. courses. Dr. Simaran Bawa, thank you for determining the secondary structure of my designed peptides. Dr. Anup Aggarwal, thank you for addressing my silly questions and providing insights into structural biology. Dr. Geetanjali Asampille, Dr. Seetharaman Jayaraman, Pradyumana Baviskar.

I express my gratitude to my framily (friends that become family). Special thanks to Dr. Pritam Sidhu for encouraging me to enroll for a Ph.D.; it was the first brick toward building my doctoral degree, and thanks for your motherly care. Amazing Sifat, that's what my phonebook says about you. Thanks for bringing joy, happiness, and laughter to my face; and giving me positive energy whenever needed. Dr. Ghashiya Kiyani and Dr. Mian Farrukh Raza, thanks for persistently

motivating me to join Ph.D. Nitesh Verma, thanks for introducing the living style of the USA and showing me around Manhattan, KS, and most of the USA, my favorite travel buddy. My brothers-in-arm, Jigesh Patel and Chintal Rawal, thank you for fueling my confidence and having a wonderful time. Leon Brown, Jenny Brown, John Ranum, Justine, Nathan Harstine, and David Stuckenschmidt for helping and making living comfortable in a beautiful town, Manhattan, KS, located 8000 miles away from home.

Last but not least, I would like to thank the entire team of Electron Microscopy shared resources, St. Jude Children's Research Hosptial, for supporting me. Especially, thanks to Dr. Camenzind Robinson for being considerate and accommodating with my Ph.D. research work. Randall Wakefield, Amanda Johnson, Amber Balkcom, and Nathan Kurtz, thanks for your cooperation and for showing your love.

The research reported in this thesis was supported by US National Science Foundation under grant no. DMR-1945589, Kansas State University (KSU) College of Veterinary Medicine Dean's funds [2015CVM-SMILE], Johnson Cancer Research Center 2018 JCRC-IRA, National Cancer Institute [1 R15 CA219919-01]. A majority of the computing for this project was performed on the Beocat Research Cluster at Kansas State University, funded in part by NSF grants CHE-1726332, CNS-1006860, EPS-1006860, and EPS-0919443. We also thank the Kansas Idea Network of Biomedical Research Excellence (K-INBRE) CORE Facility support award (grant number P20 GM103418) for providing access to the BLItz facility.

Dedication

I dedicate this Ph.D. work to the memory of my beloved elder brother. Harikrishna Thakkar, who passed away in 2009. He sacrificed his Ph.D. fellowship to take care of my high school and university education after my parent's demise. He was the visionary of my educational career. I wouldn't be the person I am today without his sacrifice and unconditional love. I am feeling happy to accomplish his dream to get a doctoral degree.

I also dedicate this thesis to my siblings Ashok, Jagdish, and Aarti for encouraging and providing cooperation. I dedicate my thesis to my nine-year-old nephew Samarth and a newborn baby girl, a bundle of joy, my niece Vidisha with hope to inspiring them for their bright academic future.

Preface

This thesis is submitted for the degree of Doctor of Philosophy in the Department of Anatomy and Physiology, College of Veterinary Medicine, Kansas State University. All the research work presented in this thesis was carried out under the able guidance of Dr. Jeffrey Comer between May 2018 to January 2022.

To the best of my knowledge, this work is original, except where references and acknowledgments are made. This dissertation, as well as any substantially similar one, has not been or will not be submitted for any other degree, diploma, or other qualification. Some portions of this thesis work have been published or submitted in the following peer-reviewed journals:

1. “Atomically resolved interfacial water structures on crystalline hydrophilic and hydrophobic surfaces” Manuel R. Uhlig, Simone Benaglia, Ravindra Thakkar, Jeffrey Comer, and Ricardo Garcia. *Nanoscale* 13 (10), 5275-5283 DOI: 10.1039/D1NR00351H

2. "Design of Peptides that Fold and Self-Assemble on Graphite" Legleiter, Justin; Thakkar, Ravindra; Miranda-Carvajal, Ingrid; Velásquez-Silva, Astrid; Whitaker, Susan; Tomich, John; and Comer, Jeffrey. *Journal of Chemical Information and Modeling* (Under review).

3. “Organic contaminants and atmospheric nitrogen at the graphene-water interface: A simulation study” prepared by Ravindra Thakkar and Jeffrey Comer. *Nanoscale Advances* (Accepted for publication).

Chapter 1 - Self-assembly of a designed cyclic peptide on a graphitic surface by molecular dynamics simulations.

1.1 Introduction

Peptides can form a wide array of architectures with diverse functionalities in ways that are well understood compared to other classes of biologically compatible molecules. With peptides, one can construct a novel molecular structure with the ability to mimic the complexes that exist in nature like virus-like protein assemblies[1], amyloid-like superstructures,[2] metalloporphyrin arrays,[3] antibacterial agents[4], and various catalytic scaffolds.[5] Because of their remarkable bioactive properties, peptide assemblies have gained substantial interest in biomedical and material research. Due to their extraordinary physical and electronic properties, 2D carbon nanomaterials have found increased applications over the past decade.[6–8] Assembly of peptides on 2D carbon nanomaterials not only improves the mechanical resilience of the resulting structures but promotes formation of ordered structures by reducing their entropic cost (through restricting the molecules to quasi-2D structures) and facilitates the visualization of the structure by microscopic techniques.[9] But the hydrophobicity and susceptibility to aggregation of carbon nanomaterials has made their application to biology and biotechnology challenging. However, chemical functionalization and the amphiphilic properties of peptides may enhance the water dispersability of carbon nanomaterials.[10,11]

In this project, I performed molecular dynamics simulations of cyclic peptide structures designed by Dr. Comer and studied their self-assembly on a graphene surface.

1.2 Methods

I was provided with the atomic structure of the peptide denoted CHP1404, which has the sequence cyc(SGTGGPGGGCGTGTGSGPGGTG). This peptide structure was designed by Dr. Comer using computational methods like sequence optimization (Rosetta and PyRosetta), temperature replica-exchange simulation (NAMD), clustering of conformations (GROMACS), and free energy estimation (GBSA and the adaptive biased force algorithm implemented in the Colvar module).[12–17]

All molecular dynamics simulations were run using NAMD.[18] The positions of the atoms were integrated with a time step of four femtoseconds, which was made stable by repartitioning hydrogen masses (on non-water hydrogen molecules) and rigidifying bonds to hydrogen. The particle-mesh Ewald electrostatics was used to provide computationally efficient calculation of long-range electrostatic forces.[19–22] Lennard-Jones forces were truncated at a cut-off distance of 12 Å.[23] The simulations were performed in the NPT ensemble . (constant atom number, pressure, temperature), the temperature and pressure were maintained by a Langevin thermostat and Langevin piston barostat, respectively.[24,25] The interatomic forces for peptides and graphene were defined by CHARMM36m and CHARMM General force fields.[26,27] The water was explicitly represented by the TIP3P model.[28] All structures were prepared and analyzed using the VMD software suite.[29]

1.2.1 Unfolding in free solution

The peptide was replicated, making 4 individual molecules, and randomly placed in a 93×93×52 $Å^3$ cube of solvent containing 9250 water molecules and 150 mmol/L NaCl. Energy minimization was performed and followed by the equilibrium simulation for 20 ns. The peptides rapidly unfolded in a free solution.

1.2.2 Folding of peptides on a graphene surface

The unfolded peptides from the above simulation were placed above two 88×82 $Å^2$ graphene layers at the 20 Å distance. Peptides and graphene layers were solvated in the 88×82×50 $Å^3$ cube of solvent containing 9603 water molecules and 150 mmol/L NaCl. Harmonic restraints were applied to backbone carbon atoms of the peptides to prevent unrealistic changes in their conformations during equilibration of the systems. The lower layer of the graphene was restrained to represent a large graphite flake as used in experiments. After running minimization, restraints applied to maintain the peptide conformation were removed; however, harmonic restraints were applied to prevent the peptides from crossing the periodic boundary along the *z* axis and adsorbing to the lower layer of graphene. Production simulations were run for more than 1.5 μs at a temperature of 370 K.

1.3 Results and discussion

1.3.1 Effects of graphene surface on a peptide folding

The hairpin structure of a peptide CHP1404 (**Fig. 1.1A**) in free solution is not stable; within 0.08 ns, all four molecules of a peptide lose their β-hairpin conformation (**Fig. 1.1B**). The peptide has a tendency to aggregate and form seemingly random conformations. This aggregated complex of peptide molecules was used as the starting structure to study peptide folding at the graphene–water interface, so that a biased conformation of the peptide can be avoided. The peptide aggregates were placed 2 nm above the graphene-water interface (**Fig. 1.2A and C**), and the simulation was run for more than 1.5 μs at a 370 K temperature. Peptides are mostly adsorbed on the graphene surface within 0.16 ns of simulation (**Fig. 1.2B**), and the aggregate of peptide molecules dispersed into individual molecules within 48.8 ns of simulation (**Fig.1.2D**). The peptides adsorbed on a graphene surface in different orientations; two molecules were found in upright orientation while

other two molecules were oriented in upside-down orientation (**Fig. 1.2D**). After 104.4 ns of the simulation, all the peptides were lying flat on the graphene surface **(Fig. 1.2E)**. However, after 745.3 ns of the simulation, the two upright-oriented peptides were able to achieve the folded hairpin conformation (**Fig. 1.2F**), while the two peptides oriented upside-down did not fold. After 854.8 ns, all peptides were arranged in a parallel formation on the graphene surface (**Fig. 1.2G**). However, the two upside-down peptides didn't fold even after 1.5 μs of simulation and exhibited variability in their structure for the entire length of the simulation, while the folded peptides maintained their hairpin conformation for rest of the simulation.

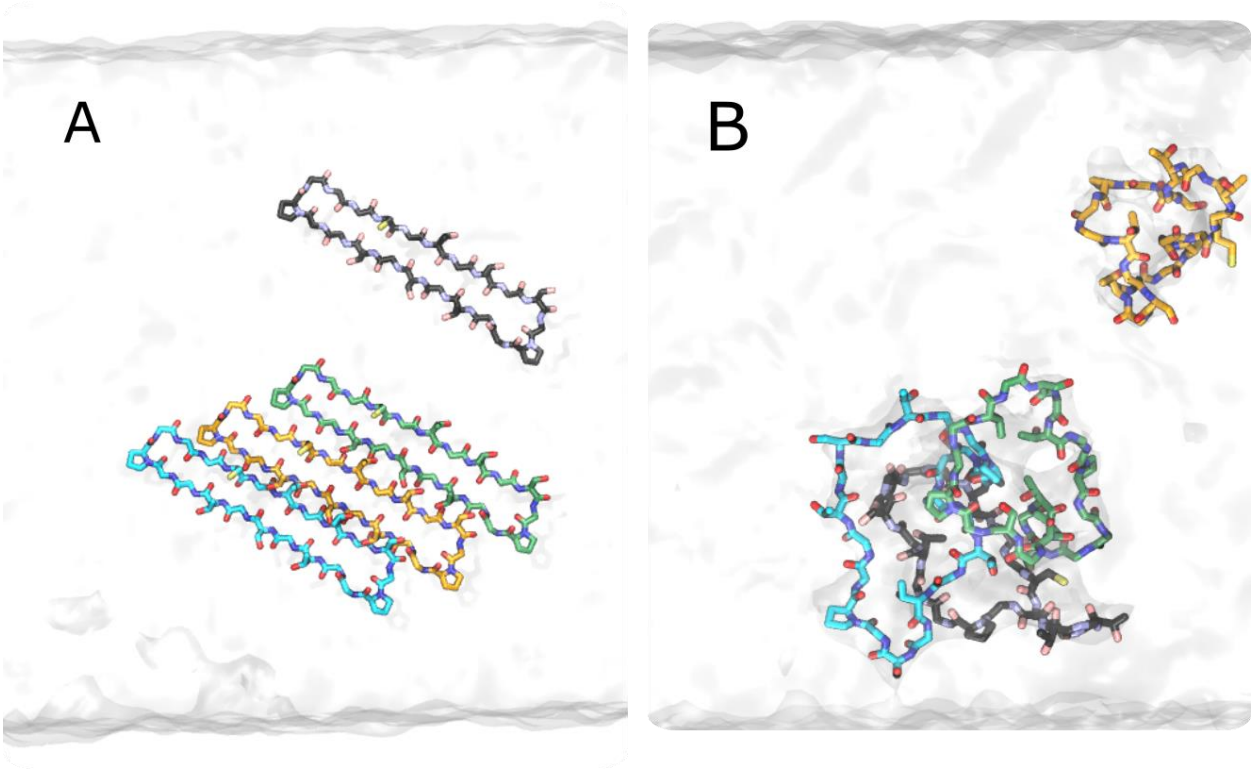

Figure 1.1 The hairpin structure of the peptide CHP1404.

(A) The starting structure of the cyclic hairpin peptide molecules. **(B)** Peptide molecules rapidly unfold in the free water simulation.

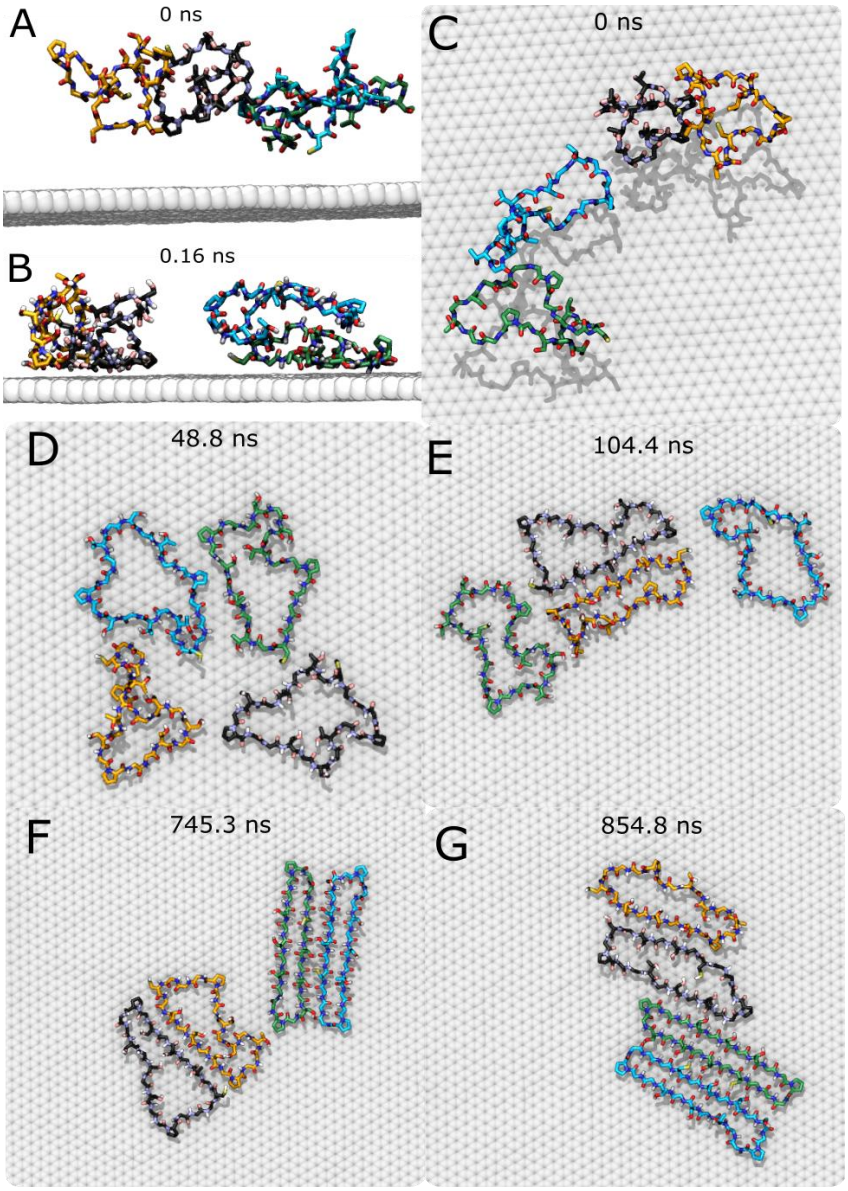

Figure 1.2 Spontaneous folding of the peptide CHP1404.

(A) The starting structure of peptides placed at a 20 Å distance from the graphene–water interface. **(B)** Peptides started adsorbing at 0.16 ns of the simulation. **(C)** The top view of a peptides on the surface of the graphene at 0 ns. **(D)** After 48.8 ns, two peptides were oriented upside-down and other two were upright. **(E)** All peptides were lying flat at 104.4 ns. **(F)** At 745.3 ns, two peptides were found in hairpin conformation. **(G)** All four peptides seen arranged in parallel formation after 854.8 ns of the simulation.

1.1.1 Peptide-peptide interaction on a graphene surface

The folded peptide molecules coordinate in pairs to form extended β-sheet-like structures. As you see in (**Fig. 1.3A**), those peptide molecules that were folded in the hairpin conformation after 745 ns are structurally similar to the original design of the CHP1404 peptide (**Fig. 1.1A**) and remain stable in a folded state for the rest of the simulation. However, the remaining two peptide molecules were structurally unstable, and very long simulations are likely required to achieve the hairpin conformation.

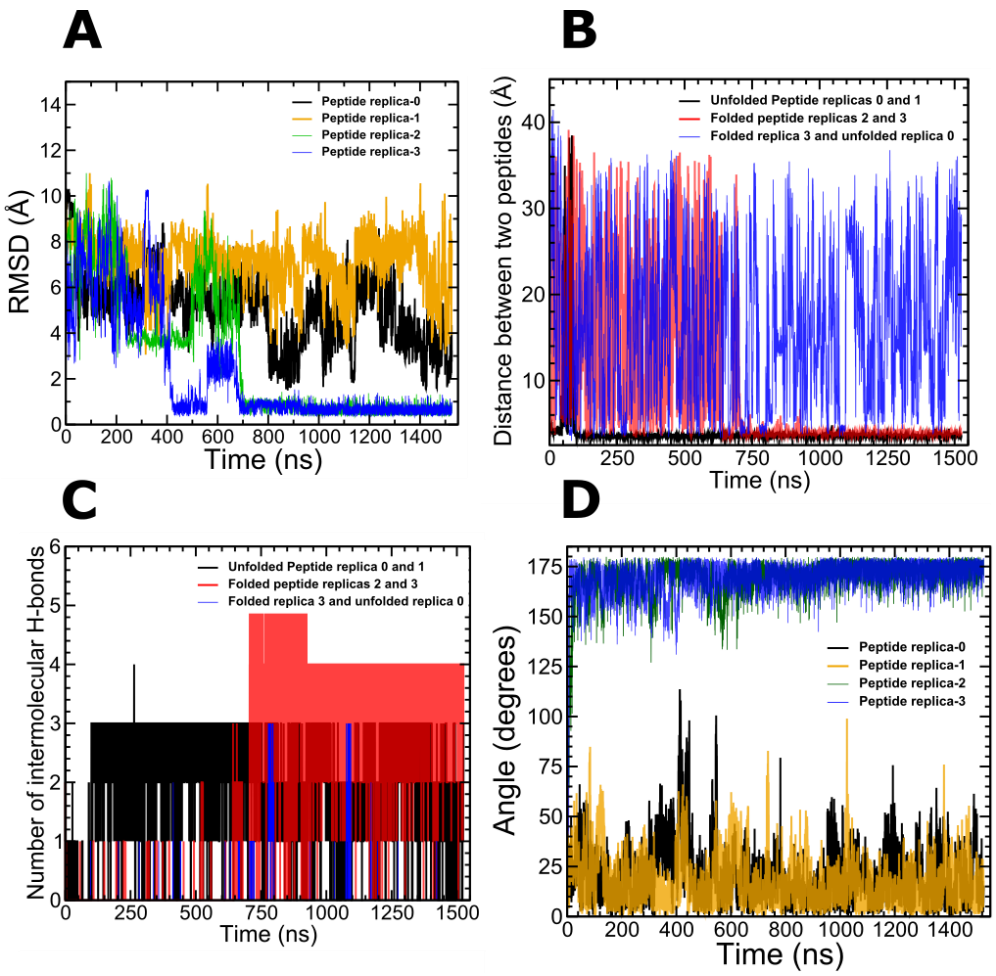

Figure 1.3 Analysis of peptide self-assembly.

(A) Root mean-square distance (RMSD) between the α carbons of the peptides in the simulation and those of a folded reference structure as a function of time. This plot indicates that two of the four molecules folded and attained structural similarity with the original design of the CHP1404 peptide. **(B)** Distance between the pair of folded peptides (black curve), unfolded peptides (red

curve) and mixture of folded and unfolded peptides (blue curve). This distance is the shortest atom–atom distance between any part of each molecule. The folded peptides remain for a stable pair. **(C)** Number of intramolecular hydrogen bonds between the pair of folded peptides, unfolded peptides and a mixture of folded and unfolded peptides. **(D)** The orientation angle of the peptides with respect to the graphene surface as function of time shows that folded peptides maintained the angle of orientation more than 135° degrees, which corresponds to the upright orientation, while the unfolded peptides showed an orientation angle of less than 45° degrees, which is associated with upside-down orientation.

In regard to the peptide-peptide interaction, the most commonly observed phenomenon was formation of intermolecular hydrogen bonds similar to a β-sheet. The distance between the pair of folded peptides varied for half of the simulation length, but after 745 ns the distance between them stayed constant, indicating that they formed a pair. The phenomenon of pairing of the peptides was corroborated with the intermolecular hydrogen bond analysis (**Fig. 1.3C**). It is observed that the pair of folded peptides have a greater number of intermolecular hydrogen bonds after their pairing at 745 ns, which is consistent with the RMSD analysis of those peptide molecules and the analysis of the distance between them.

When initially adsorbed from solution, the peptide is presumably equally likely to be in the orientation that we refer to as "upright," which is associated with folded form, or the "upside-down" orientation. I measured the orientation of a peptide by calculating the vector product of the vector between consecutive residues as per (**Eq. 1.1**).

$$\mathbf{O} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{r}_i^{\mathrm{CA}} - \mathbf{r}_{i-1}^{\mathrm{CA}}\right) \times \left(\mathbf{r}_{i+1}^{\mathrm{CA}} - \mathbf{r}_i^{\mathrm{CA}}\right),$$

$$\theta = \cos^{-1}\left(\hat{\mathbf{z}} \cdot \frac{\mathbf{O}}{|\mathbf{O}|}\right)$$

where $n = 22 =$ number of residues,

$\mathbf{r}_i^{\mathrm{CA}} =$ position of the C_α carbon of the i^{th} residue,

$\hat{\mathbf{z}} =$ direction orthogonal to the graphene plane,

$\theta =$ angle between the orientation vector $\mathbf{O}$ and $\hat{\mathbf{z}}$.

Equation 1.1 vector product of the vector between consecutive residues.

The upright-oriented peptide has an orientation vector **O** that faces towards the graphene (θ > 135°), while an upside-down-oriented peptide has an **O** that faces away from the graphene (θ < 45°). As shown in (**Fig. 1.3D**), folded and unfolded peptides occupy upright and upside-down orientations, respectively, because it is difficult for the upside-down oriented peptides to fold.

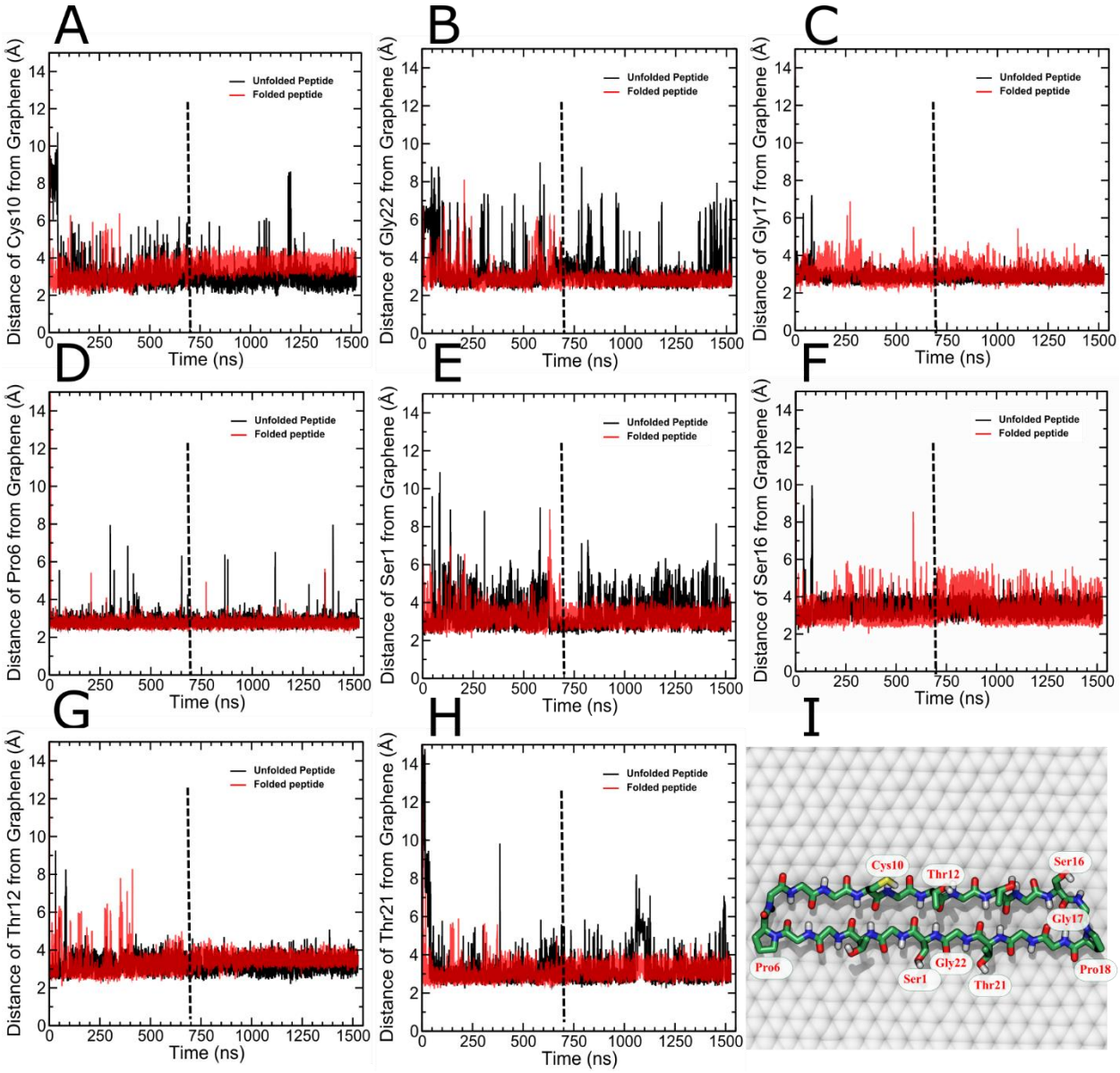

Figure 1.4 Interaction between peptide and graphene.

(A) Distance between the peptide's cysteine side chain and the graphene surface. This distance was more consistent for the folded peptides than for the unfolded conformations. For the curves labeled "Folded peptides," folding occurred by 745 ns (indicated by the dashed line); before 745 ns both peptides were unfolded. **(B)** Likewise, the distance between the glycine residue at the middle of the peptide and the graphene surface is more stable for the folded peptide. **(C)** The distance between glycine present at loop region and graphene surface fluctuates in both conformations of peptide. **(D)** Similarly, contact with the graphene surface is stable for the proline residues at the loop region. **(E)** The distance between the serine residue present at the linear region of the peptide stays consistent in the folded state **(F)** whereas the position of the serine at the loop oscillates in folded state. **(G and H)** The interaction of threonines with graphene surface was stable once the peptide folds for both threonines considered (one in the loop region and one in the linear region). **(I)** Residues glycine, serine and threonine adopt a planar backbone conformation at the graphene-water interface in the folded state.

1.3.2 The role of particular amino acids in peptide-graphene interaction

A previous study reported that the small peptide Ac-Ala-NHMe changes its conformational preferences on a graphitic surface from typical α-helical or β-sheet conformations to distinct planar versions, which are stabilized at the graphene surface due to amide-π stacking.[30] Aromatic amino acids have demonstrated a high affinity for a graphene surface, whereas aliphatic and carboxylic amino acids have demonstrated a weak interaction with a graphene surface.[31–33] Dr. Comer has selected electrically neutral residues like glycine, proline, threonine, cysteine, and serine to design the CHP1404 peptide that is highly likely to form the β planar conformation at the graphene–water interface. Inspired from the natural structure of silk proteins containing GA repeats, Dr. Comer designed peptides with GX repeats, where X are residues other than glycine. Cyclization of the peptide facilitates the planar β conformation. Previous studies suggest that the cyclization of a peptide by disulfide bonds can stabilize the β-sheet conformation in solution; however, the geometry imposed by disulfide bonds is not quite optimal for the β-sheet conformation.[34,35]On the other hand, we found that head-to-tail cyclization can better stabilize the β-sheet conformation, and non-glycine residues were added at specific locations to force the β turn to occur.

When the peptide adopts a stable folded conformation and the separation between parts of the peptide and the graphene layer remains consistent. On the hand, this separation can fluctuate considerably when the peptide is the unfolded state as it visits a wide variety of conformations. Hence to determine the conformational stability of peptide, I calculated the distance between individual amino acids and the upper graphene layer, using the convention that the reported distance is the shortest atom–atom distance between any part of the amino acid and any part of the graphene. The hydrophobic interaction between cysteine and the graphene surface was found to be more consistent when the peptide was in the folded state than when it was in the unfolded state (**Fig. 1.4A**). As shown in (**Fig. 1.4B**), the distance between the graphene layer and glycine positioned in the linear region (Gly22 in **Fig. 1.4I**) of the peptide fluctuates in the unfolded state and the folding of peptide stabilizes the distance between this glycine and the graphene layer. In contrast, irrespective of the peptide conformation, inconsistency in the distance was found when glycine present in the loop region (**Fig. 1.4C**). Except for a few occurrences, proline exhibited consistent distance from graphene surface in irrespective of its position in the peptide and conformational state. Both prolines are in loop regions and force the turn to occur. Presumably, the pyrrolidine group of proline consistently makes hydrophobic interaction with graphene surface for the entire length of the simulation (**Fig. 1.4D and E**). The serine residue at the linear region of the peptide stays at constant distance from graphene layer after folding of the peptide, whereas the distance of serine from graphene remains inconsistent even after peptide folding for the serine residue present at the loop. Interestingly, in the unfolded peptide, the serine of loop region remains stable compared with the serine located in the linear region of the peptide (**Fig. 1.4F and G**). The distance profile for the threonine, whether it was located near the loop or at the center of linear region, did not show a

difference in the folded state. However, in the unfolded state, the threonine located near loop region fluctuates more (**Fig. 1.4H and I**). Hence, it is depicted that due to amide–π stacking between planar amide and graphene, residues like glycine, serine and threonine present in the linear region of the peptide stabilize the distance with graphene's surface, while glycine located at the loop region did not form a planner geometry even in the folded state. A broken line in the figure–4 B, E and G shows that the peptides were fully folded after 745 nanoseconds of the simulation.

1.4 Conclusion

In this work, I investigated, using computational modeling and molecular dynamics simulations, how a graphene surface affects folding of a peptide and influences a peptide-peptide interaction. I observed spontaneous folding of the peptide into a stable β-sheet-like structure and spontaneous pairing of two folded peptide molecules. As a result of my research, I have provided insights into designing peptides for self-assembly on graphitic materials, which can be used to develop biosensors and drug delivery devices.

However, surface contamination on graphitic surfaces might interfere with the self-assembly of peptides. Hence, identification and removal of surface contamination are crucial in biotechnology and medical applications of graphitic surfaces. I will discuss the identification of the organic contaminants on a graphene surface in the next chapter.

1.5 References

1. Grigoryan, G. *et al.* Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* **332**, 1071–1076 (2011).

2. Shaham-Niv, S., Adler-Abramovich, L., Schnaider, L. & Gazit, E. Extension of the generic amyloid hypothesis to nonproteinaceous metabolite assemblies. *Sci Adv* **1**, e1500137 (2015).

3. Yu, T., Lee, O.-S. & Schatz, G. C. Molecular Dynamics Simulations and Electronic Excited State Properties of a Self-Assembled Peptide Amphiphile Nanofiber with Metalloporphyrin Arrays. *J. Phys. Chem. A* **118**, 8553–8562 (2014).

4. Rodrigues de Almeida, N. *et al.* Design, Synthesis, and Nanostructure-Dependent Antibacterial Activity of Cationic Peptide Amphiphiles. *ACS Appl Mater Interfaces* **11**, 2790–2801 (2019).

5. Jang, H.-S. *et al.* Tyrosine-mediated two-dimensional peptide assembly and its role as a bio-inspired catalytic scaffold. *Nat Commun* **5**, 3665 (2014).

6. Brown, C. L., Aksay, I. A., Saville, D. A. & Hecht, M. H. Template-Directed Assembly of a de Novo Designed Protein. *J. Am. Chem. Soc.* **124**, 6846–6848 (2002).

7. Li, P., Chen, X. & Yang, W. Graphene-induced self-assembly of peptides into macroscopic-scale organized nanowire arrays for electrochemical NADH sensing. *Langmuir* **29**, 8629–8635 (2013).

8. Li, Q. *et al.* Modulating aβ33-42 peptide assembly by graphene oxide. *Chemistry* **20**, 7236–7240 (2014).

9. Iglesias, D., Bosi, S., Melchionna, M., Da Ros, T. & Marchesan, S. The Glitter of Carbon Nanostructures in Hybrid/Composite Hydrogels for Medicinal Use. *Curr Top Med Chem* **16**, 1976–1989 (2016).

10. EL-Mahdy, A. F. M. & Kuo, S.-W. Diphenylpyrenylamine-functionalized polypeptides: secondary structures, aggregation-induced emission, and carbon nanotube dispersibility. *RSC Adv.* **8**, 15266–15281 (2018).

11. Iglesias, D. *et al.* Oxidized Nanocarbons-Tripeptide Supramolecular Hydrogels: Shape Matters! *ACS Nano* **12**, 5530–5538 (2018).

12. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).

13. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput* **13**, 3031–3048 (2017).

14. BEKKER, H. *et al.* GROMACS - A PARALLEL COMPUTER FOR MOLECULAR-DYNAMICS SIMULATIONS: 4th International Conference on Computational Physics (PC 92). *PHYSICS COMPUTING '92* 252–256 (1993).

15. Tanner, D. E., Chan, K.-Y., Phillips, J. C. & Schulten, K. Parallel Generalized Born Implicit Solvent Calculations with NAMD. *J. Chem. Theory Comput.* **7**, 3635–3642 (2011).

16. The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask | The Journal of Physical Chemistry B. https://pubs.acs.org/doi/10.1021/jp506633n.

17. Fiorin, G., Klein, M. L. & Hénin, J. Using collective variables to drive molecular dynamics simulations. *Molecular Physics* **111**, 3345–3362 (2013).

18. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802 (2005).

19. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).

20. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952–962 (1992).

21. Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics* **52**, 24–34 (1983).

22. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

23. Daintith, J. D. Lennard-Jones potential. in *A Dictionary of Chemistry* (ed. Daintith, J.) (Oxford University Press, 2008).

24. Davidchack, R. L., Handel, R. & Tretyakov, M. V. Langevin thermostat for rigid body dynamics. *J. Chem. Phys.* **130**, 234101 (2009).

25. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).

26. Huang, J. *et al.* CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **14**, 71–73 (2017).

27. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields - Vanommeslaeghe - 2010 - Journal of

Computational Chemistry - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.21367.

28. Mark, P. & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **105**, 9954–9960 (2001).

29. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).

30. Poblete, H., Miranda-Carvajal, I. & Comer, J. Determinants of Alanine Dipeptide Conformational Equilibria on Graphene and Hydroxylated Derivatives. *J. Phys. Chem. B* **121**, 3895–3907 (2017).

31. Hughes, Z. E. & Walsh, T. R. What makes a good graphene-binding peptide? Adsorption of amino acids and peptides at aqueous graphene interfaces. *J. Mater. Chem. B* **3**, 3211–3221 (2015).

32. Imai, Y. N., Inoue, Y., Nakanishi, I. & Kitaura, K. Amide-pi interactions between formamide and benzene. *J Comput Chem* **30**, 2267–2276 (2009).

33. Dasetty, S., Barrows, J. K. & Sarupria, S. Adsorption of amino acids on graphene: assessment of current force fields. *Soft Matter* **15**, 2359–2372 (2019).

34. Cochran, A. G. *et al.* A Minimal Peptide Scaffold for β-Turn Display: Optimizing a Strand Position in Disulfide-Cyclized β-Hairpins. *J. Am. Chem. Soc.* **123**, 625–632 (2001).

35. Russell, S. J., Blandl, T., Skelton, N. J. & Cochran, A. G. Stability of Cyclic β-Hairpins: Asymmetric Contributions from Side Chains of a Hydrogen-Bonded Cross-Strand Residue Pair. *J. Am. Chem. Soc.* **125**, 388–395 (2003).

Chapter 2 - Identification of organic contaminants at the graphene-water interface.

Adopted in parts from the manuscripts "Organic contaminants and atmospheric nitrogen at the graphene-water interface: A simulation study" prepared by Ravindra Thakkar and Jeffrey Comer. Submitted to Nanoscale Advances – The Royal Society of Chemistry. (Under review) and "Atomically resolved interfacial water structures on crystalline hydrophilic and hydrophobic surfaces" Manuel R. Uhlig, Simone Benaglia, Ravindra Thakkar, Jeffrey Comer and Ricardo Garcia. Nanoscale 13 (10), 5275-5283 DOI: 10.1039/D1NR00351H

2.1 Introduction

Aqueous solutions interact with solid surfaces in many biochemical processes, such as heterogeneous catalysis and biomineralization. As a promising solid surface material for nanotechnology, graphene can modulate peptide assembly when present on substrates. However, the technical application of graphene, graphite, and carbon nanotubes depends upon the interaction between the graphitic surface and the media. The presence of contaminants at the interface of graphitic surface and water is found by physical experimental methods like atomic force microscopy, infrared spectroscopy, contact angle measurements, and capacitance measurements.[1–5] The presence of contaminants was reported even on freshly cleaved graphitic surfaces, and the type of contaminants depended upon the environment to which graphite was exposed. *Kozbial et al.* reported a change in water contact angle after exposing freshly cleaved graphite to air and measured infrared spectrum peaks that suggested the presence of linear alkanes as contaminant.[2] In another study, the decrease in capacitance of the freshly cleaved graphite in water due to hydrocarbon contamination was detected by *Hurst et al.*[6] From these studies, it is speculated that these contam-

inants are made up of hydrocarbon mixture. Ambient air and purified water have a trace of hydrocarbon contaminants between ranges of 5 to 10 ppb. Adsorption of hydrocarbon contaminants affects the order of hydration on the graphitic surface.[7] Although hydrocarbons only make up a small portion of ambient air; their high affinity makes them likely to adsorb on graphene surfaces.[8,9] The laboratory environment and apparatus also may also serve as sources of hydrocarbon contamination. As Seibert et al. noted, syringes with plastic tips were able to cause stripes on the graphite surface, but such stripes were not observed when glass syringes were used.[10] In the recent study, *Garcia et al.* observed that the force profile obtained by AFM for the graphene-water interface is surprisingly similar to that obtained for the graphene-hexane interface. From the results of computational simulations, it is speculated that the water molecules were expelled from the interface and replaced by alkane-like hydrocarbon molecules, forming a 15–20 Å thick layer between water and hydrophobic graphitic surface.[3]

Molecular dynamics simulations can be a useful tool to screen the wide range of hydrocarbon materials and explore possibilities of their presence as contaminants at the graphene-water interface. Molecular dynamics simulations have not been used extensively to study hydrocarbon contaminants at the graphene-water interface. In the present study, a few volatile hydrocarbons like pentane, octane, toluene and dodecane commonly found in indoor air and in laboratories were selected and studied by molecular dynamics simulations. The results of these simulations were compared with the reported experimental evidence of contaminants at graphene-water interface.

2.2 Methods

2.2.1 Molecular modeling and system construction

The organic molecules pentane, octane, toluene, and dodecane, and the carbon atoms of graphene were represented by the CHARMM General Force Field.[11] Two rectangular patches of graphene with dimensions of 2.9 nm × 3.0 nm containing 672 carbon atoms were built using the program VMD.[12] The PackMol package was used to generate random configurations of 48, 34, 30 and 20 molecules of pentane, toluene, octane, and dodecane, respectively, and added to the top of the graphene layers.[13] A model of a fragment of an AFM tip was created from a diamondoid carbon framework using the Avogadro and the TopoTools module of the VMD program and parameterized with the CGenFF webserver.[12,14–16] The relatively small forces measured in the experiments suggest that contact between the AFM tip and surface occurs at an asperity only a few atoms in size. To model such an asperity, the central carbon was attached a methyl group that made direct contact with the graphitic surface; remaining carbon atoms were capped with either an OH group or two hydrogen atoms. Each system was solvated in a 6 nm^3 cubic box. Each water molecule was explicitly represented by the TIP3P model and kept rigid using the SETTLE algorithm.[17,18] An exemplary model of the system is shown in (**Fig. 2.1**).

2.2.2 Molecular dynamics simulations

All molecular dynamics simulations were performed using NAMD version 2.13.[19] To improve the computational efficiency, non-water hydrogen mass was repartitioned, and a time step of 4 femtoseconds was applied.[20] Lennard-Jones interactions were smoothly truncated with a cut-off distance of 10 – 12 Å, and particle-mesh Ewald electrostatics with a grid spacing less than 1.2 Å was used to calculate long-ranged electrostatic interactions.[21,22] All simulations were performed at 295 K

and 1.01325 bar pressure by using a Langevin thermostat and Langevin piston algorithm, respectively.[23,24] Periodic boundary conditions were imposed to make a pair of graphene sheets infinite in the 2D dimension. The lower layer of graphene was restrained to its initial position to mimic a mounted graphite flake as used in experiments. The solvent and solutes were allowed to interact with the unrestrained upper layer of the graphene. The orientation of the AFM tip model was maintained by applying restraints as implemented in the Colvars module of NAMD.[25] Minimization for each system was performed for 2000 steps prior to running the production simulation.

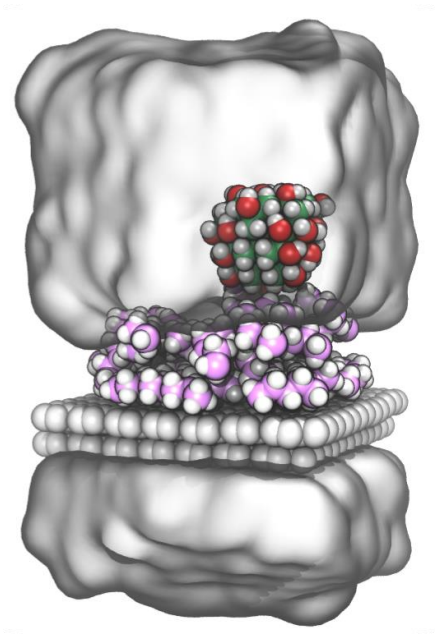

Figure 2.1 A molecular model.

A molecular system containing a model of a fragment of an AFM tip and pentane molecules at the water and graphite interface. The water is represented as gray transparent surface and carbon atoms of the AFM tip, graphene layer and pentane are shown as a green, gray and pink spheres, respectively.

2.2.3 Free energy calculation

The adaptive biasing force (ABF) method was used to calculate the mean force on the AFM tip model as a function of the distance between this model and the graphene surface.[26,27] ABF enhances sampling so that the accurate estimates of the mean force can be obtained over the entire domain (0.25–1.40 nm). Using the Colvars module, the transition coordinate was defined as the distance from the center of mass of the upper graphene sheet to the center of mass of the AFM tip model projected along the *z*-axis (the axis perpendicular to the graphene planes). The grid size for transition coordinate was set to 0.005 nm, and all ABF simulations were run for around 1.5 μs.

2.3 Results and discussion

2.3.1 Presence of hydrocarbon at a graphene-water interface

As described in our previous paper, the simulation-derived force profile for the graphene–hexane interface agreed well with corresponding AFM-derived data. In contrast, the simulation-derived force profile for the graphene–water interface was wildly different than the AFM results for what is nominally a graphene–water system. (**Fig. 2.2A**).[3,28] This observation suggests the presence of some contaminants at the graphene-water interface that alter the structure of the interface. To test this hypothesis, we prepared four different molecular systems, each containing a graphene surface, the AFM tip model, and alkane molecules (pentane, toluene, octane, or dodecane) solvated in water. Using the adaptive biasing force method, the average force on the AFM tip model was calculated as a function of distance from the graphene-water interface for each system.[26] The resultant force profile for pentane, toluene, octane, and dodecane (**Fig. 2.2B**) showed high similarity with the force profile obtained for the graphene-hexane interface from the previous study (**Fig. 2.2A**).

Additionally, at the graphene-water interface, pentane, toluene, octane, and dodecane have a distance of ≈ 0.45 nm between the first two minima (denoted as the distance between two broken lines in Figures 2A and B). This results corroborate with the previous AFM study that suggested 0.5 nm spacing distance between first two maxima for the graphite-water and graphite-hexane interface.[3]

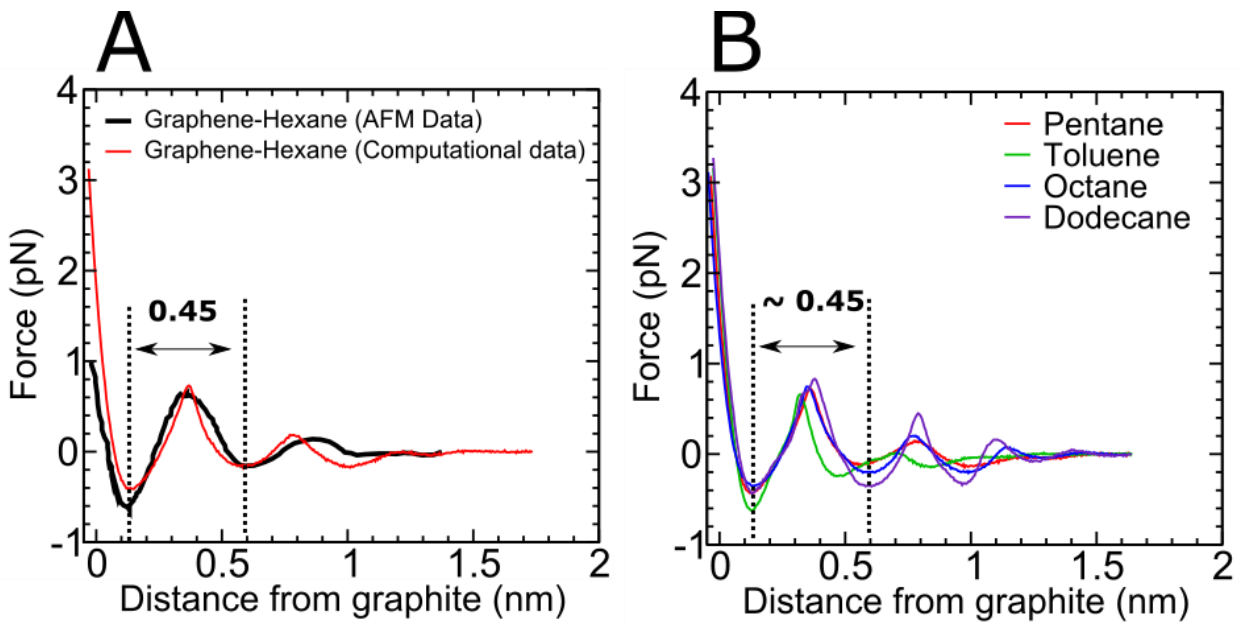

Figure 2.2 The force profile a graphene-hexane and graphene-water interface.

(A) The force profile a graphene-hexane interface obtained by AFM and computational simulation; **(B)** The force profile for graphene-water interface populated by alkane molecules (pentane, toluene, octane, or dodecane) as predicted by computational simulations.

2.3.2 The mass density of hydrocarbons at a graphene-water interface

As I discussed in the previous section, the distance between first and second minima in the force profile of octane, pentane, toluene, and dodecane from my study shows similarity with the distance between first and second minima in force profile of the graphene-hexane interface obtained by AFM experiment by the previous study.[3] However, the density of the hydrocarbon varies

depending on the molecular weight of the molecule. As shown in (**Fig. 2.3**), the density of dodecane, a hydrocarbon containing higher molecular weight, was found higher at the first maxima compared with the hydrocarbons containing lower molecular weight like pentane, octane, and toluene. However, the maximum of mass density for the water molecules are closer to the graphene layer as compared with the organic compounds. Interestingly, the distance between the first two density peaks for the toluene was found shorter than the first two maxima reported for the graphene-water interface by Uhlig et al.[3] This observation suggests that the aromatic hydrocarbons are unlikely to be the dominant contaminants present on a graphene surface in experiments.

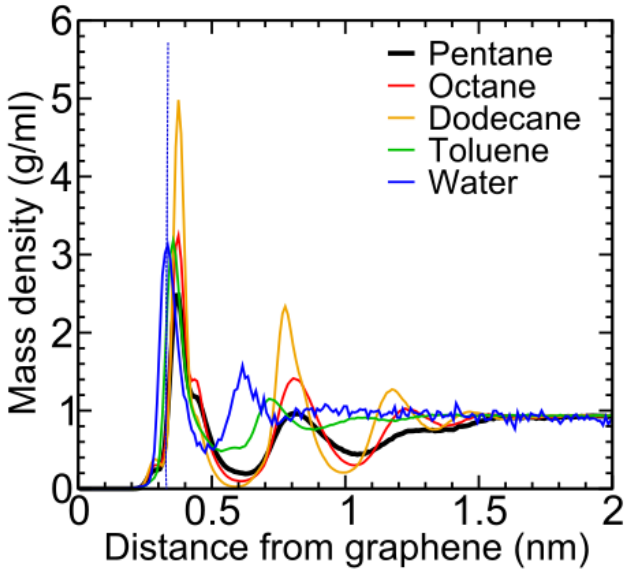

Figure 2.3 The mass density profile for hydrocarbons on the graphene-water interface. The distance of maximum of the water mass density is indicated by a dotted blue line.

2.3.3 An ordered arrangement of hydrocarbon on the graphene-water interface

At the beginning of the simulation, dodecane molecules were arbitrarily dispersed in the water phase, but within 7 ns, dodecane molecules started adsorbing on the graphene surface. From the observation of the simulation trajectories for the system containing hydrocarbons, specifically, dodecane on the graphene-water interface, it is speculated that the hydrocarbons containing long-

chain alkane form an ordered arrangement on the graphene-water interface. The dodecane molecules have adopted a straight conformation to maximize the contact area by lying flat against the surface (**Fig. 2.4**). This arrangement of hydrocarbon molecules might be responsible for the stripes observed on the graphene-water interface by the AFM analysis in many previous studies.[3,10,28–32] These stripes possibly show the presence of long-chain hydrocarbon adsorbed on the graphene surface.[10]

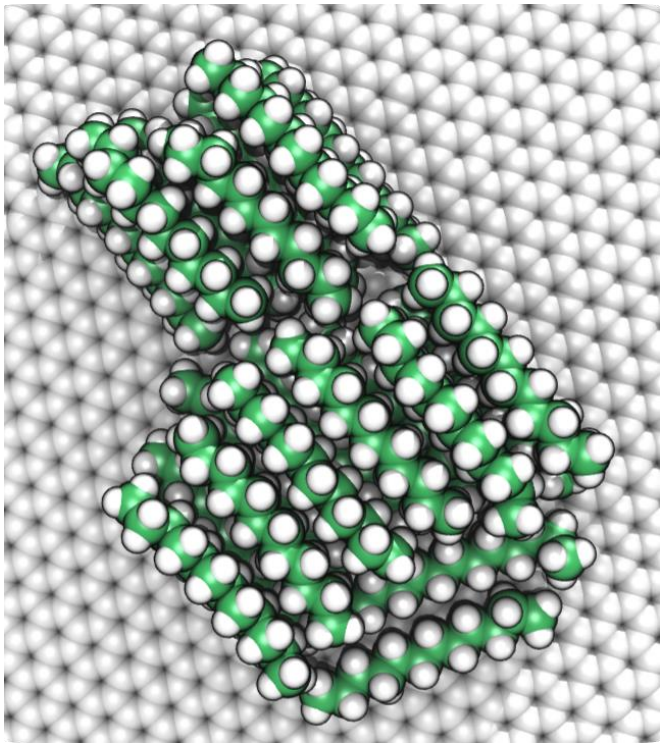

Figure 2.4 Ordered arrangement of dodecane on the graphene-water surface.

2.4 Conclusion

In conclusion, the mass density and force profiles at the graphene-water interface in the presence of hydrocarbons predicted by simulations using an AFM tip model agree with the previous experimental report. The presence of stripes on graphene surfaces observed by AFM is likely due to an ordered arrangement of linear hydrocarbons originating in the ambient medium that collect at the surface. Even the presence of trace amounts of hydrocarbons in purified water, ambient air, or laboratory apparatus can adsorb onto the graphene surface and able to alter its properties.

2.5 References

1. Schrader, M. E. Ultrahigh vacuum techniques in the measurement of contact angles. IV. Water on graphite (0001). *J. Phys. Chem.* **79**, 2508–2515 (1975).

2. Kozbial, A., Zhou, F., Li, Z., Liu, H. & Li, L. Are Graphitic Surfaces Hydrophobic? *Acc. Chem. Res.* **49**, 2765–2773 (2016).

3. R. Uhlig, M., Benaglia, S., Thakkar, R., Comer, J. & Garcia, R. Atomically resolved interfacial water structures on crystalline hydrophilic and hydrophobic surfaces. *Nanoscale* **13**, 5275–5283 (2021).

4. Haghighian, N. *et al.* Ripple morphology of graphitic surfaces: a comparison between few-layer graphene and HOPG. 11.

5. Teshima, H., Nakamura, N., Li, Q.-Y., Takata, Y. & Takahashi, K. Zigzag gas phases on holey adsorbed layers. *RSC Advances* **10**, 44854–44859 (2020).

6. Hurst, J. M. Adventitious Hydrocarbon Adsorption and Its Effect on the Solid-Liquid Interface. 177.

7. Li, Z. *et al.* Water Protects Graphitic Surface from Airborne Hydrocarbon Contamination. *ACS Nano* **10**, 349–359 (2016).

8. Xia, X. R. *et al.* Mapping the Surface Adsorption Forces of Nanomaterials in Biological Systems. *ACS Nano* **5**, 9074–9081 (2011).

9. Comer, J., Chen, R., Poblete, H., Vergara-Jaque, A. & Riviere, J. E. Predicting Adsorption Affinities of Small Molecules on Carbon Nanotubes Using Molecular Dynamics Simulation. *ACS Nano* **9**, 11761–11774 (2015).

10. Seibert, S., Klassen, S., Latus, A., Bechstein, R. & Kühnle, A. Origin of Ubiquitous Stripes at the Graphite–Water Interface. *Langmuir* **36**, 7789–7794 (2020).

11. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields - Vanommeslaeghe - 2010 - Journal of Computational Chemistry - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.21367.

12. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).

13. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J Comput Chem* **30**, 2157–2164 (2009).

14. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform | Journal of Cheminformatics | Full Text. https://jcheminf.biomedcentral.com/articles/10.1186/1758-2946-4-17.

15. Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **52**, 3144–3154 (2012).

16. Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **52**, 3155–3168 (2012).

17. Mark, P. & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **105**, 9954–9960 (2001).

18. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952–962 (1992).

19. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802 (2005).

20. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).

21. Daintith, J. D. Lennard-Jones potential. in *A Dictionary of Chemistry* (ed. Daintith, J.) (Oxford University Press, 2008).

22. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

23. Davidchack, R. L., Handel, R. & Tretyakov, M. V. Langevin thermostat for rigid body dynamics. *J. Chem. Phys.* **130**, 234101 (2009).

24. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).

25. Fiorin, G., Klein, M. L. & Hénin, J. Using collective variables to drive molecular dynamics simulations. *Molecular Physics* **111**, 3345–3362 (2013).

26. The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask | The Journal of Physical Chemistry B. https://pubs.acs.org/doi/10.1021/jp506633n.

27. Darve, E. & Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **115**, 9169–9183 (2001).

28. Uhlig, M. R., Martin-Jimenez, D. & Garcia, R. Atomic-scale mapping of hydrophobic layers on graphene and few-layer MoS2 and WSe2 in water. *Nat Commun* **10**, 2606 (2019).

29. Wastl, D. S. *et al.* Observation of 4 nm Pitch Stripe Domains Formed by Exposing Graphene to Ambient Air. *ACS Nano* **7**, 10032–10037 (2013).

30. Yaghoubi, H. & Foroutan, M. Wettability of striped patterned mono-and multilayer graphene supported on platinum. *Applied Surface Science* **500**, 144002 (2020).

31. Ko, H.-C., Hsu, W.-H., Yang, C.-W., Fang, C.-K. & Lu, Y.-H. Preferential Gas Adsorption at the Graphene-Water interface. 19.

32. Gallagher, P. *et al.* Optical Imaging and Spectroscopic Characterization of Self-Assembled Environmental Adsorbates on Graphene. *Nano Lett.* **18**, 2603–2608 (2018).

Chapter 3 - Computational design of cyclic peptide for cancer immunotherapy and its characterization.

3.1 Introduction

Immune cells typically infiltrate the tumor microenvironment and have the ability to detect cancer-specific antigens; however, cancer cells often evolve the ability to hijack self-tolerance mechanisms, inhibiting the immune response and allowing them to evade destruction.[1,2] These self-tolerance mechanisms are referred as immune checkpoints and involve inhibitory receptors such as CTLA4, PD1, TIM3, LAG3, and BTLA, which are expressed on immune cells. Immuno-therapies based on antibodies that bind these receptors (CTLA4, PD1) or their natural ligands (PDL1 for the PD1/PDL1 pathway) have dramatically improved clinical results for some cancers. The first immune checkpoint inhibitor approved by the US Food and Drug Administration was ipilimumab,[3] a monoclonal antibody that binds to CTLA4, preventing the binding of its natural ligands B7-1 and B7-2 (also known as CD80 and CD86).[4,5] At the current time, approved immune checkpoint inhibitors that target PD1 and PDL1 also find clinical use. Targeting both the CTLA4 and PD1 pathways appears promising and combined therapies can yield improved clinical out-comes for some cancers.[6]

Thus far, all successful immune checkpoint inhibitors have been monoclonal antibodies. However, such antibodies are difficult and expensive to produce and have short shelf lives,[7] leading to unfavorable cost effectiveness in some cases.[8] Small-molecule immune checkpoint inhibitors could provide simpler synthesis and lower costs; however, developing potent small molecules has proven difficult due to the relatively at hydrophobic interfaces of immune checkpoint receptors and many of their ligands.[7,9] Most proteins involved in inhibitory immune checkpoint signaling, including CTLA4, B7-1, B7-2, PD1, PDL1, PDL2, TIM3, LAG3, BTLA, B7-H4 (VTCN1), and

B7-H3 (CD276), have immunoglobulin-like V (IgV) domains and some also possess similar immunoglobulin-like C1 or C2 domains.[10] These domains have a β-sandwich structure,[11] consisting of two β-sheets lying face-to-face, which are tethered together by one or more disulfide bridges. The binding interfaces of these domains are typically found on the relatively flat surfaces of these β-sheets, which include no obviously druggable pockets. For instance, the human CTLA4:B7-2 complex,[12] has a binding interface involving the β-sheet faces of CTLA4 and B7-2, as well as the conserved MYPPPY (residues 99-104) loop of CTLA4 **(Figure 3.4)**. Although this paper focuses on designing peptides that bind to this interface of CTLA4, we expect that the design principles might be used to design inhibitors for similar proteins involved immune checkpoint signaling.

While protein-protein interfaces can be difficult targets for conventional small molecule drugs, peptides, by their nature, should be able to mimic natural protein-protein interactions.[13] Moreover, the facile synthesis and modification of peptides make them a promising alternative to antibodies and other drugs.[14] Various studies have reported the therapeutic applications of synthetic peptides in drug delivery, cell membrane penetration, and specific cell targeting and activation of immune response.[15,16] Additionally, some clinical studies suggests that the application of peptides can be site-specific and safe for the drug delivery purpose.[17] Tens of therapeutic peptides are already approved to treat cancer, diabetes, and cardio-vascular diseases.[18] One major challenge to the use of therapeutic peptides is rapid degradation by proteolytic enzymes present in the physiological environment.[13] However, these limitations can be overcome by taking inspiration from natural protease-resistant peptides, which are often cyclized to improve the proteolytic stability. Like natural peptides, synthetic peptides can be cyclized by one or more disulfide bridges or by peptide linkages involving the N-terminus, C-terminus, or sidechains (usually Lys, Asp, or

Glu).[19,20] Furthermore, crosslinks of various chemistries (referred to as “staples") can be used to cyclize peptides in ways never found in nature.[13]

Computational protein design tools have matured over the last decade, enabling rational design of peptides with desired structure and function can be possible.[21,22] Combining protein modeling programs like Rosetta,[22–24] Modeller,[25,26] and associated tools[27] with molecular dynamics simulation and free-energy calculation techniques, [28,29] results in a powerful platform for designing therapeutic peptides.[30,31] In this paper, we apply such an approach to designing a CTLA4-binding peptide and verify its affinity experimentally by using the Bio-Layer Interferometry (BLI) method.[32]

3.2 Material and methods

3.2.1 Materials

We contracted a commercial service (GenScript USA Inc., Pscataway, NJ, USA) to synthesize the cyclic peptide and obtained human recombinant Avi-tagged CTLA4 and CD86 (as positive control) from commercial sources (R&D Systems, Inc., Minneapolis, MN USA). High precision streptavidin (SAX) BLItz biosensor tips were purchased from FortéBio; Freemont, CA, USA).

Cell lines of the mouse immature dendritic cells JAWSII JAWSII (CRL-11904) and the mouse Lewis Lung Carcinoma cells LLC (CRL-1642) were obtained from American Type Culture Collection (ATCC, Manassas, VA). Cell culture media RPMI 1640, Fetal bovine serum (FBS), 2-mercaptoethanol and Penicillin-streptomycin solution were purchased from Mediatech, Inc. (Manassas, VA), Biowest (Riverside, MO), Sigma-Aldrich (St. Louis, MO) and Lonza Rockland, Inc. (Allendale, NJ), respectively. The sodium pyruvate (200 mM), L-glutamine (200 mM), 50x

Gibco® brand antibiotic-antimycotic, 100x MEM non-essential and essential amino acids, were obtained from Thermo Fisher Scientific (Waltham, MA). The Ultra-LEAF™ Purified anti-mouse CD274 (B7-H1, PD-L1) antibody (10F.9G2), the LIVE/DEAD™ Fixable Violet Dead Cell Stain Kit (Invitrogen™) and Fluorescent conjugated antibodies targeting FoxP3 (R16-715) were purchased from BioLegend (San Diego, CA), Thermo Fisher Scientific and BD Bioscience (Franklin Lakes, NJ), respectively. Fluorescent conjugated antibodies targeting mouse CD4 (H129.19), CD8b (YTS156.7.7), IFNγ (XMG1.2), CTLA4 (UC10-4B9), PD-1 (RMP1-30) and isotype controls were obtained from BioLegend.

C57BL/6 mice were purchased from Charles River Laboratories International, Inc and were acclimated for a week in a clean facility. Kansas State University Institutional Animal Care and Use Committee (Protocol # 4393) and Institutional Biosafety Committee (Protocol # 1317) procedures were followed for all animal experiments.

3.2.2 Peptide modeling

An x-ray structure of the CTLA4:B7-2 protein complex was obtained from the Protein Data Bank (PDB ID: 1I85).[12] The CTLA4 structure was missing some residues (residues 27-30 and 42-44, inclusive), which we inserted using Modeller 9.13, a homology modeling tool.[25,26] A single monomer unit of CTLA4 and B7-2 protein were extracted using VMD[33] (a molecular modeling and visualization software) from the crystal structure (**Fig. 3.1A**). The hotspot residues at the binding interface between CTLA4 and B7-2 protein predicted using KFC web server.[34,35] The fragment of B7-2 protein comprising residue numbers 85 to 101, which was found to be in close contact with CTLA4, was selected as a template for the peptide design. A CTLA4 protein and residue number 85 to 101 from B7-2 protein as shown in (**Fig. 3.1B**) was used for further processing.

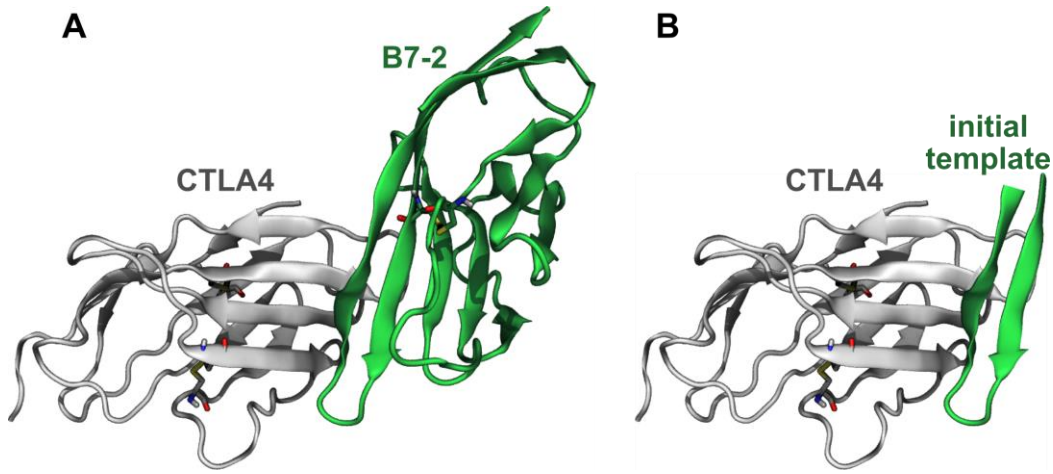

Figure 3.1 Generating the template selection.

(A) CTLA4 (gray) bound to its ligand B7-2 (green) from a published x-ray structure (PDB ID: 1I85). Disulfide bridges are shown explicitly by bonds representation. **(B)** A template for the designed peptides created by extracting a fragment (residues 85-101) from the x-ray structure.

The probable poses of template peptide on the receptor protein CTLA4 were predicted using FlexPepDock ab-initio, a module of the Rosetta, a protein modeling suite.[27,36] VMD was used to modify the structure le for the template peptide and receptor to make its format compatible with the Rosetta (a molecular modeling suite). The secondary structure of the template peptide sequence was predicted using the protein structure prediction web-server Phyre2.[37] and used as input for FlexPepDock.

Hundreds of different poses of the template peptide were generated. Few best poses with the lowest energies and RMSD values relative to the native structure were selected. Selected structures were processed with the series of algorithms (PyRosetta, a Python-based interface for Rosetta)[38] to optimize side-chain and amino acid sequence.[24,39]

3.2.3 Molecular dynamics and free energy calculation using the MM-GBSA method

Molecular dynamics simulations for each designed peptides bound to the CTLA4 receptor protein were carried out using NAMD 2.13, a scalable molecular dynamics simulation program.[40] The CHARMM36m force field[41] was used to define forces between the atoms. The two terminal residues of resulting peptides were patched by a *trans head-to-tail* (N-terminal to C-terminal) peptide bond. The structures of the complexes were solvated using the TIP3P[42] water model, which was kept rigid using the SETTLE algorithm.[43] Salt ions were added to give Na^+ and Cl^- concentrations of approximately 150 mmol/L. Additional Na^+ ions where included to neutralize the system. An exemplary simulation system is shown in (**Fig. 3.2**).

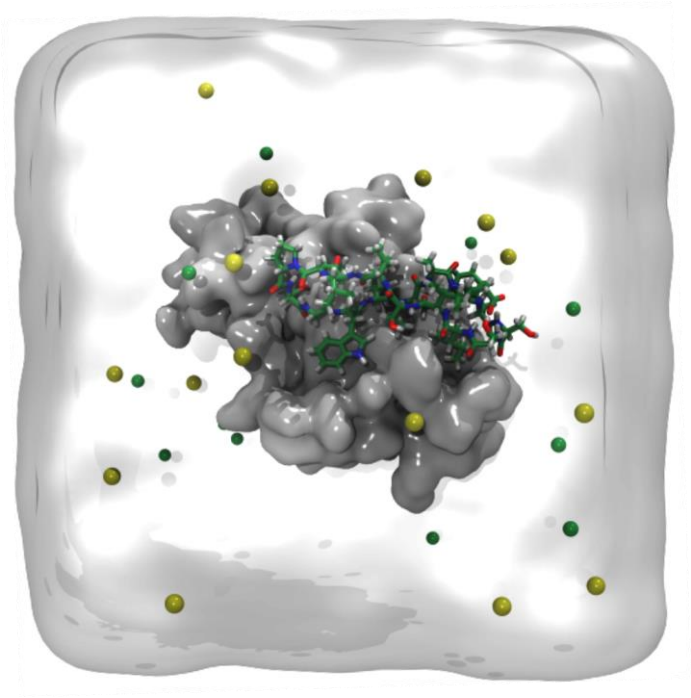

Figure 3.2 Explicit solvent model for the biomolecular system.

The CTLA4 receptor protein represented as a gray surface and the designed peptide R10 (H, white; C, green; N, blue; O, red). Na^+ and Cl^- ions are shown as yellow and greens spheres, respectively. For clarity, the explicit water molecules are shown only as a transparent surface.

The mass of solute hydrogen atoms was repartitioned to allow a 4 femtosecond time step to improve computational ecffiency.[44] The Lennard-Jones interactions between pair of atoms was

calculated using cutoff distance of 12 Å, smoothly truncated beginning at 10 Å. A temperature of 300 K and a pressure of 1.01325 bar were maintained using the Langevin thermostat[45] and Langevin piston method[46], respectively. The electrostatic interactions were calculated by particle mesh Ewald (PME) method using grid spacing of 1.2 Å.[47] Energy minimization for each system were performed.[48] After minimization, 0.1 ns of molecular dynamics were performed with restraints applied to non-hydrogen atoms of the protein, followed by 1.0 ns with restraints applied to the only the Cα carbon atoms. The production simulation was performed without using any restraints and set to stop when the Cα atoms of the candidate peptide showed a root mean square displacement (RMSD) of more than 25 Å from their initial positions. The Colvars module of NAMD was used to implement the stopping criterion.[49] Every 200 ps, a frame of the trajectory was written for further analysis. For each of these frames, the binding free energy of the candidate peptide to the CTLA4 was estimated using the molecular mechanics generalized Born surface area (MM-GBSA) method.[50–52] To implement the GBSA calculation, the structures of the receptor protein, candidate peptide, and protein-peptide complex were extracted from the output trajectories and snapshots were created. For each snapshot of the receptor protein, candidate peptide, and protein-peptide complex MM-GBSA calculated as per (**Eq. 3.1**).

$$\Delta G_{binding} = \Delta G_{protein:peptide} - \Delta G_{protein} - \Delta G_{peptide}$$

Equation 3.1 Binding free energy calculation by MM-GBSA method.

The ∆G values were calculated using the NAMD implementation of generalized Born implicit solvent with a dielectric constant of 78.5 and a surface tension 0.00542 kcal/(mol Å^2). The conformational entropy was not calculated because it does not necessarily improve the accuracy of the relative MM-GBSA energies and a rigorous free-energy calculation method, which included this contribution, was applied in the promising cases.

3.2.4 Explicit Solvent Free Energy Calculations

The MM-GBSA free energy of candidate peptide: CTLA4 complexes were compared with the CTLA4:B7-2 reference protein complex MM-GBSA results. The designed cyclic peptides with higher mean MM-GBSA binding affinity for the CTLA4 receptor protein were shortlisted for the rigorous calculation of ab-solute binding free energy using the geometric route.[29,53] The simulation frame having the minimum MM-GBSA energy was used as the starting structure for the complex and used as input for the Binding Free Energy Estimator (BFEE), a plugin of VMD.[54] To make the free energy calculation feasible, the calculation is partitioned into separate stages **(Table–3.1)**. The basic idea is to calculate the free energy for turning on a cumulative series of restraints (stages 1-6). The free energy for extracting the restrained ligand from the receptor is then calculated in stage 7, which is typically the most difficult and time-consuming step. The contribution associated with turning off the conformational restraints for the free ligand is calculated in stage 8. Finally, the contribution of the remaining geometrical restraints is computed analytically (which can be called stage 9). Because the initial and final states in these stages represent an unrestrained complex and an unrestrained free ligand/protein, the sum of the free energy over all stages is the binding free energy **(Eq. 3.2)**. Stages 2-5 were computed with the adaptive biasing force (ABF) method,[55,56] consistent with the BFEE plugin:

$$\Delta G^{0}_{\text{binding}} = k_{\text{B}}T \ln C^{0}$$

$$\mathrm{K_{eq}} = \exp\{-\beta\Delta \mathrm{G_c^{site}} + \Delta \mathrm{G_0^{site}} + \Delta \mathrm{G_\alpha^{site}} - \mathrm{k_B T \ln (S * I * C^0)} + \Delta \mathrm{G_0^{bulk}} + \Delta \mathrm{G_c^{bulk}}\}$$

where $C^{0} = \frac{1}{1660.539}\,\text{Å}^{-3}$ is the standard 1 mol/L concentration.

Equation 3.2 Binding free energy calculation by explicit solvent geometric route.

Table 3.1 Free energy values and simulation time for each stage of the BFEE method.

Free energy values and simulation time for each stage of BFEE, a rigorous absolute biding free energy calculation method. All stages were computed with ABF, except stage 9, which is computed analytically and requires no simulation.

Stage	System	Free-Energy Term	Free Energy (kcal/mol)	Simulation time (ns)
1	protein-ligand	$\Delta G_{conform}$	–11.03±1.59	91
2	protein-ligand	ΔG_{θ}	–0.23±0.02	19
3	protein-ligand	ΔG_{φ}	–0.29±0.17	17
4	protein-ligand	ΔG_{ψ}	–0.20±0.02	19
5	protein-ligand	ΔG_{θ}	–0.05±0.01	20
6	protein-ligand	ΔG_{ϕ}	–0.49±0.11	17
7	protein-ligand	$-k_B T \ln (S * I * C^0)$	–12.92±0.79	6954
8	ligand only	$\Delta G_{conform}^{unbound}$	+11.97±2.96	389
9	ligand only	$\Delta G_{\Theta\phi\Psi}^{unbound}$	+6.61±0.00	0
Total	-	ΔG^0	–6.63±3.45	2380

3.2.5 Peptide design workflow

Fig. 3.3 illustrates the workflow for the cyclic peptide design. **Fig. 3.3A** shows the CTLA4/B7-2 complex structure. **Fig. 3.3B** shows a fragment from the B7-2 protein taken as the primary template peptide. Different poses of template peptide on CTLA4 were predicted **(Fig. 3.3C)**. The 4-5 structures with the most favorable poses from the above step were selected for sidechain and sequence optimization. **Fig. 3.3D** shows the template peptide in black carbon back-bone and a one of the selected poses from the various favorable poses of the template peptide in green carbon backbone. All optimized peptides were cyclized by a head-to-tail peptide bond (one of the optimized peptides is shown in **Fig. 3.3E**). MD simulation and MM-GBSA calculations were carried out to estimate the binding affinity of all the designed cyclic peptides for CTLA4 receptor protein **(Fig. 3.3F)**. Peptides with noteworthy MM-GBSA scores and longer simulation times without dissociating were selected for rigorous binding free energy calculation **(Fig. 3.3G)**.

Cyclic peptides with higher binding affinity were selected as subsequent templates for further optimization cycles till we achieved a significant affinity for binding between the cyclic peptide and CTLA4 receptor protein **(Fig. 3.3H)**.

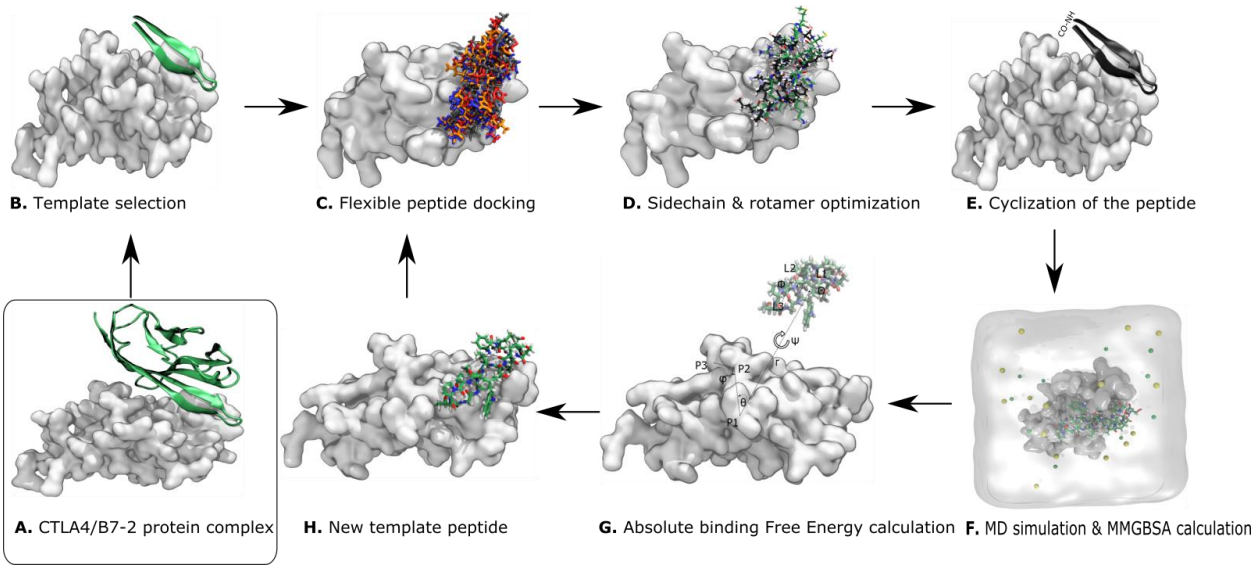

Figure 3.3 Graphical representation of the work-flow for cyclic peptide design.

3.2.6 Binding assay by bio-layer interferometry

The peptide with the best predicted binding affinity (denoted R10) and ordered proteins were dissolved in PBS buffer. The binding affinity of best performing peptide for the CTLA4 protein was evaluated using a BLItz bio-layer interferometer (FortéBio; Freemont, CA, USA) at room temperature. High precision streptavidin (SAX) BLItz biosensor tips (FortéBio; Freemont, CA, USA) were used to immobilize the CTLA4 protein and all the tips were hydrated for 15-30 minutes in PBS buffer before each experiment. A plateau of the signal during the washing step (after loading the CTLA4 on biosensor tips) indicated immobilization of the CTLA4 protein on the high precision streptavidin (SAX) biosensor tip. The PBS buffer without the designed cyclic peptide or any protein was used to record the baseline. A solution of 400 nM of the CD86 protein,

a natural binding partner of CTLA4, was used as positive control and different molar concentrations (150, 175 and 200 μM) of the designed peptide R10 were tested. The values the on and off rates and the dissociation constant were obtained using the BLItz Pro 1.2 software.

3.2.7 Effects of designed peptide R10 on T cell-induced death of LCC cells

The LLC lysate was subcutaneously injected into mice on Day 0 (0.5×106 cells/mouse in 200 μl). A 1:1 ratio of X-ray (100 Gy) irradiated LLC cells and JAWSII immature dendritic cells were cocultured for 48 hours with LPS (1 g/ml). JAWS-irrLLC (above mentioned) cells were collected and intravenously injected into mice on Day 7 (0.5×104 cells per mouse in 200 μl). The splenocytes were harvested on Day 21 and labeled with a MojoSort™ Mouse CD8+ T Cell Isolation Kit (Biolegend, San Diego, CA) and isolated with a MACS® Column (Miltenyi Biotec, Bergisch Gladbach, Germany). Using purified CD8+ T cells, LLC cells were primed with antigens to produce antigen-primed CD8+ T cells (AP-CD8+ T cells).

LLC cells expressing GFP were seeded into a 12-well plate (1x104 cells per well) and treated with murine interferon-gamma (mIFNγ) at 25 ng/ml for 48 hours. AP-CD8+ T cells were added to each well at a 1:16 ratio (LLC cells: AP-CD8+ T cells) and treated with 10 μM peptide R10, 0.5 or 1.0 μg/ml mouse anti-PD-L1 antibody (αPD-L1). At 18 and 36 hours, the cytotoxicity of AP-CD8+ T cells for LLC cells was measured using LIVE/DEAD™ Fixable Violet Dead Cell Stain Kit and evaluated by BD LSRFortessa X-20 flow cytometer (BD Biosciences, San Jose, CA, USA). Using BD FACSDiva software (BD Bioscience), GFP+ LIVE/DEAD+ gating based specific death of GFP-LLC cells was identified.

3.2.8 Anti-tumor effects of the designed peptide R10 for murine lung carcinoma

A 1.5×106 per 200µl LLC cells were inoculated in the C57BL/6 mice. A 0.5×106 cells/200µl JAWS-irrLLC were injected intravenously into all mice after 5 days. After 48 hours, mice were separated into negative control (PBS), positive control (anti-PD-L1; 10 mg/kg/day and IP every other day 4 times) and treatment group (peptide R10; 10 mg/kg/day and IP every other two days 3 times). On 23rd day of LLC inoculation, all mice were sacrificed, and blood was collected by cardiac puncture for cytometry analysis. Measured weights of lung and spleen and followed by histological analysis.

3.2.9 Effects of the designed peptide R10 on the T cell population and an expression of immune check points

Leukocytes were isolated and immune-stained with antibodies anti-CD4, anti-FoxP3, anti-CD8b, anti-INFγ to detect the population of helper, regulatory, cytotoxic and activated cytotoxic T cells, respectively by flow cytometry. Immune checkpoints CTLA4 and PD-1 expression was evaluated using antibodies anti-CTLA4 and anti-PD-1. An isotype control was used to verify non-specific reaction of antibodies.

3.2.10 Statistical analysis:

Statistical significance of effects of the peptide R10 for *in vivo* and *in vitro* experiments were evaluated by ANOVA or unpaired t-test. All values were obtained from the multiple samples and were expressed as the mean ± standard deviation of the mean. Statistical significance was accepted with minimum 5% probability.

3.3 Results

3.3.2 Template selection

The prediction of hotspot residues at the CTLA4:B7-2 complex (PDB ID: 1I85)[12] interface shows that the residues Ile86, His88 and Met95 of B7-2 make contact with the conserved MYPPPY loop of CTLA4,[57] as illustrated in the **(Fig. 3.4)**. We chose residues 85 to 101 of B7-2, (some of the residues are part of the β-hairpin) to serve as the template for designing the peptide.

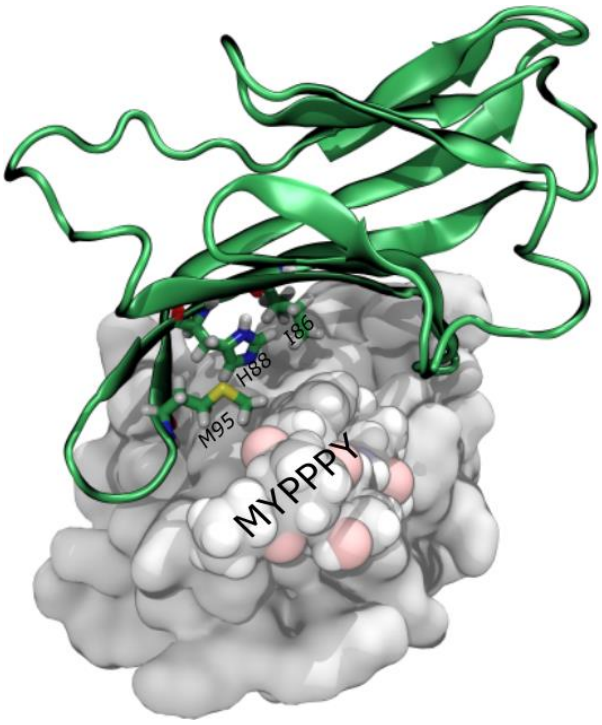

Figure 3.4 Interface of the CTLA4:B7-2 protein complex.

CTLA4 receptor protein (gray) is represented as a surface with conserved motif MYPPPY in solid spheres. The B7-2 protein (green) is represented as by its secondary structure along with key residues represented as sticks.

3.3.3 Flexible peptide docking

We performed the flexible docking of the template peptide with the CTLA4 protein at the MYPPPY surface and hundreds of different poses of the template peptide were generated **(Fig. 3.5A)**. Out of these, a few poses with low docking scores and low RMSD values from the initial

template were selected using a plot of docking score versus RMSD **(Fig. 3.5B)**. The resulting poses were applied for fixed-backbone sequence optimization by Rosetta. The selected structures had RMSD values of 4.6, 3.3, 2.4 and 3.9 Å with the template and docking scores of 167.2, 173.6, 172.4 and 170.8. Because the FlexPepDock module of Rosetta does not currently support cyclic peptides, the linear version of the subsequent candidate peptides was used for the docking procedure. The effect of cyclization on the backbone conformation was later tested in molecular dynamics simulation.

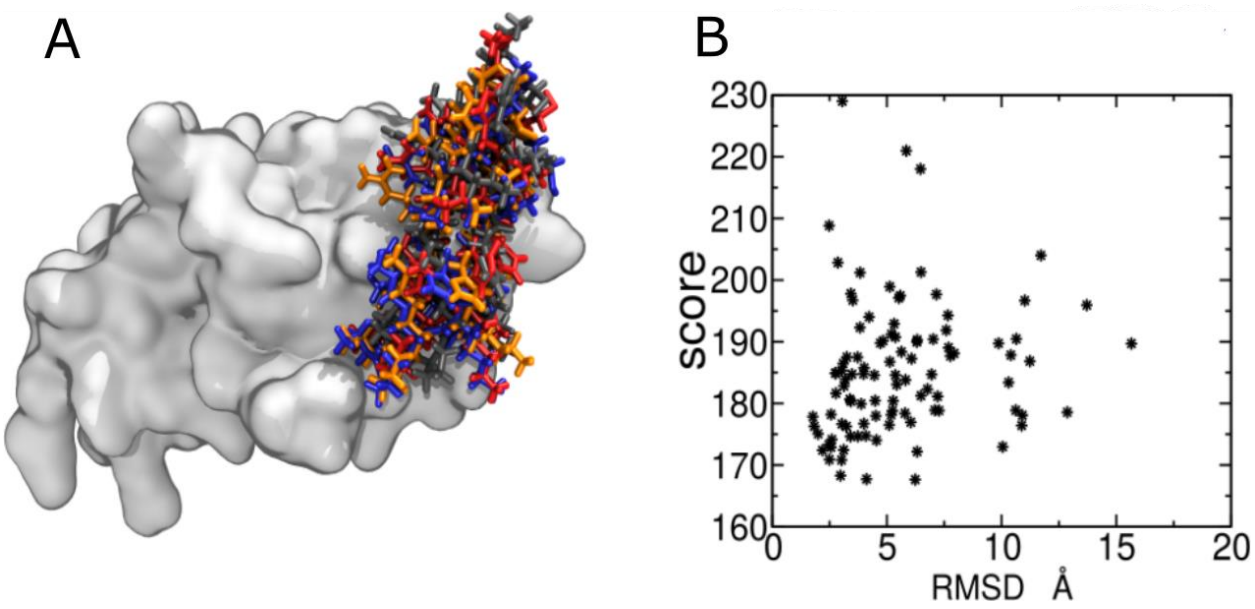

Figure 3.5 Flexible peptide docking and selection of favorable poses.

(A) Selected docking poses for the template peptide (represented in different color stick models) The CTLA4 receptor protein (gray) is represented as a surface. **(B)** A docking score versus RMSD plot for the different poses of the template peptide on the CTLA4 surface generated by flexible peptide docking.

3.3.4 Sidechain and sequence optimization

This is the crucial step in the peptide design. The purpose of this step is to make the connection between peptide and protein stronger by manipulating the sidechain and amino acid sequence of the peptide. For each selected pose, sequence and side chain optimization was carried

out. As this algorithm was based on Rosetta modeling software, it doesn't include the information about cyclic nature of the peptide.

The sequence logo in the **(Fig. 3.6)** represents the probability of the amino acids found at each position in the candidate peptides.[58] It indicates that Pro8 and Gly10, which are part of the hairpin loop, can't be replaced by other residues while maintaining the same backbone structure.

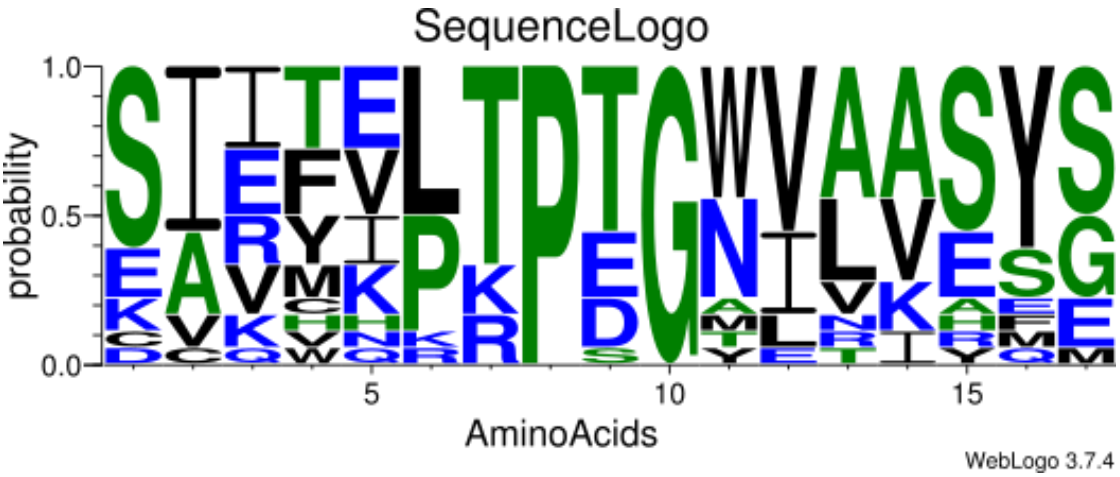

Figure 3.6 The plot of probabilities for occurrence of amino acids at specific positions in the designed peptides.

3.3.5 MD simulation and binding free energy calculation

The MD simulation techniques allow us to explore more conformational dynamics taking the place between protein and ligand during the binding phenomena.[59] We cyclized the optimized peptides by a head-to-tail peptide bond using in-house VMD scripts before running the simulations. In the MD simulations the candidate peptides stayed bound with CTLA4 for more than a microsecond, which is an indication of considerable binding affinity. We performed the free energy calculations using the MM-GBSA method to short list the optimized peptides and performed a rigorous free energy calculation using the BFEE plugin of VMD (by the geometric rather than

alchemical route). We shortlisted 6 optimized peptides that had MM-GBSA scores less than –20 kcal/mol (**see Appendix A**). Upon applying the rigorous calculation, the two most promising peptides exhibited binding affinities of –10.21 ± 2.41 and –12.58 ± 3.76 kcal/mol, which may be sufficient to block the MYPPPY motif on the surface of the CTLA4 receptor for B7 family proteins. As illustrated in (**see Appendix B**), we attempted cyclize the peptide by adding a disulfide bond between cysteine residues at the termini, but the results were not promising. Furthermore, head-to-tail cyclization may have advantages in promoting β-sheet structure.

3.3.6 Binding assay by bio-layer interferometry

A bio-layer interferometry-based binding assay was performed to determine the binding affinity of the designed peptide (R10) for the CTLA4 protein. First, we confirmed that the biotinylated recombinant human CTLA4 Fc-chimera Avi-tagged protein was non-covalently immobilized on the high precision streptavidin (SAX) biosensor tip. As shown in **Fig. 3.7B**, the CTLA4 protein was immobilized on the high precision streptavidin biosensor tips (red curve), whereas we did not get any binding response while applying only PBS at the loading step (black curve). Immobilization of CTLA4 on the high precision streptavidin biosensor tip was further confirmed when CD-86 was used as an analyte at the association step **(Fig. 3.7A)** (red curve). Our binding kinetics results revealed a dissociation constant of 60 ± 3 nM (K_d) for the binding of h-CTLA4 to CD-86, which is in agreement with the previous reports.[60] Thus, the high precision streptavidin (SAX) biosensor tip and CTLA4 protein could be used to evaluate the binding affinity of the designed peptide R10. The binding kinetics results for the R10 peptide are shown in **(Fig. 3.7A)**. The K_d value for the designed peptide R10 was found to be 30 ± 6 μM, which overlaps with the range predicted by the BFEE calculation using MD simulations (34 to 4400 μM).

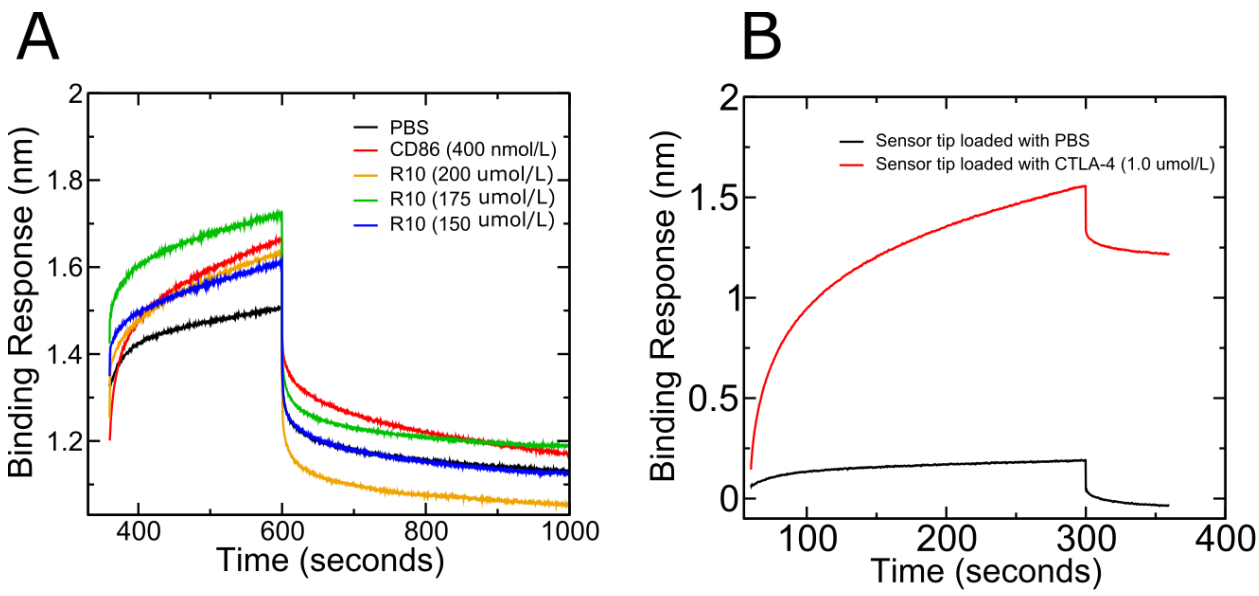

Figure 3.7 Time-based binding kinetics study of the R10 peptide binding to CTLA4 by bio-layer interferometry.

(A) Different molar concentrations of the R10 peptide were used. The CD-86 protein was used as a positive control and PBS solution alone served as a negative control. The experiments were performed using the biotinylated recombinant human CTLA4 Fc-chimera Avi-tagged protein loaded on a biosensor tip. The binding response was set zero at the beginning of the association, and the dissociation phase was initiated at 600 seconds. **(B)** Bio-layer interferometry experiment confirming loading of CTLA4 (human CTLA4 conjugated to Fc Avi tag) to the high precision streptavidin biosensor tip.

3.3.7 Effects of designed peptide R10 on T cell-induced death of LCC cells

Effects of the peptide R10 on T cell-induced death of LLC were assessed by *in vitro* co-culturing with LLC cell antigen primed CD8+ T cells which was isolated from the spleen of the immature dendritic cell line (JAWSII) cells stimulated mouse with irradiated LLC cells. The death of LLC cells was measured after 18 and 36 hours by flow cytometry. As per **(Fig. 3.8),** results showed that after 18 hours the LLC-specific cell death was 43.8 ± 3.0% in the 10 μM peptide R10 treated cells, 66.6 ± 1.5% in the 1.0 μg/ml anti-PD-L1 treated (positive control) group and 12.7 ± 0.3% in the PBS treated group. However, 1.0 g/ml anti-PD-L1 was the most effective in promoting LLC cell death induced by AP-CD8+ T cells compared with other groups, but after 36 hours the

efficacy of anti-PD-L1was decreased to 27.5 ± 12.9% while the efficacy of the peptide R10 (68.2 ± 12.3%) was sustained significantly. According to these results, designed peptide R10 can effectively inhibit the interaction between CTLA4 on AP-CD8+ T cells and CD80/CD86 on LLC cancer cells, thus facilitating the cytotoxic effect of AP-CD8+ T cells.

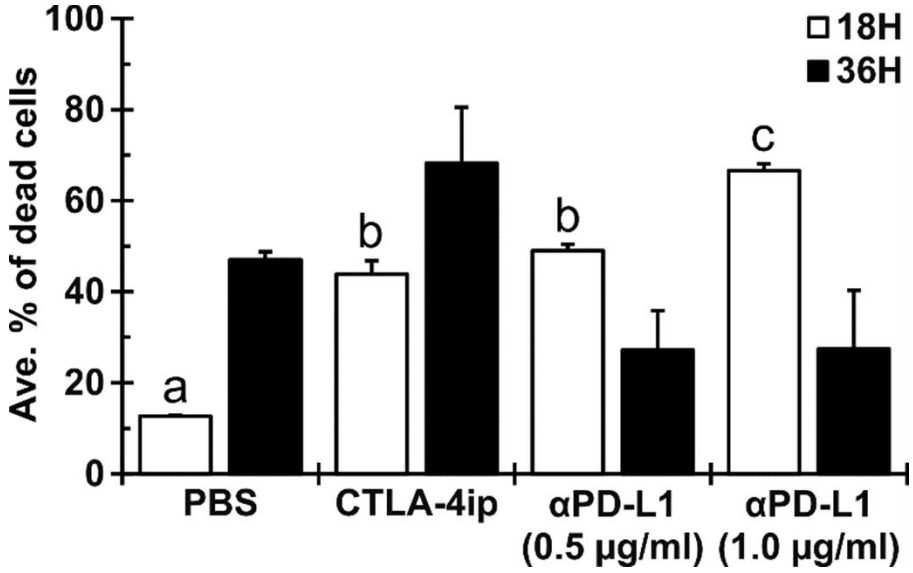

Figure 3.8 The designed peptide R10 promotes antigen primed CD8+ T cell induced LLC cell death.

Cocultured LLC cells with antigen primed CD8+ T cells treated with the peptide 10 μM R10 (CTLA4-ip), anti-PD-L1 antibody (αPD-L1; 0.5 or 1.0 μg/ml) and Flow cytometry was used to identify dead cells after 18 and 36 hours of coculture.

3.3.8 Anti-tumor effects of the designed peptide R10 for murine lung carcinoma

LLC mouse allografts were used to test the effectiveness of peptide R10 against lung tumor growth. Irradiated LLC were coculture with the mouse dendritic cell line, JAWSII cells for 24 hours and injected into all mice. After 48 hours, mice were treated with the 10mg/kg/day peptide R10 (CTLA4-ip), 10 mg/kg/day anti-PD-L1 as positive control or PBS as negative control and tumor growth was analyzed by macroscopic study. It was demonstrated that fewer tumor nodules (4 out of 5, Ave. 1.60 ± 0.80) were present in the mice treated with the designed peptide R10

(CTLA4-ip) compared with (5 out of 6, Ave. 3.50 ± 2.57) the PBS treated, a negative control group whereas anti-PD-L1 treated, positive control mice showed similar effects (4 out of 6, Ave. 1.17 ± 1.07) as negative control group **(Fig. 3.9A and B)**. Based on results, it appears that R10, a novel designed peptide, triggers anticancer immunity by inhibiting tumor growth.

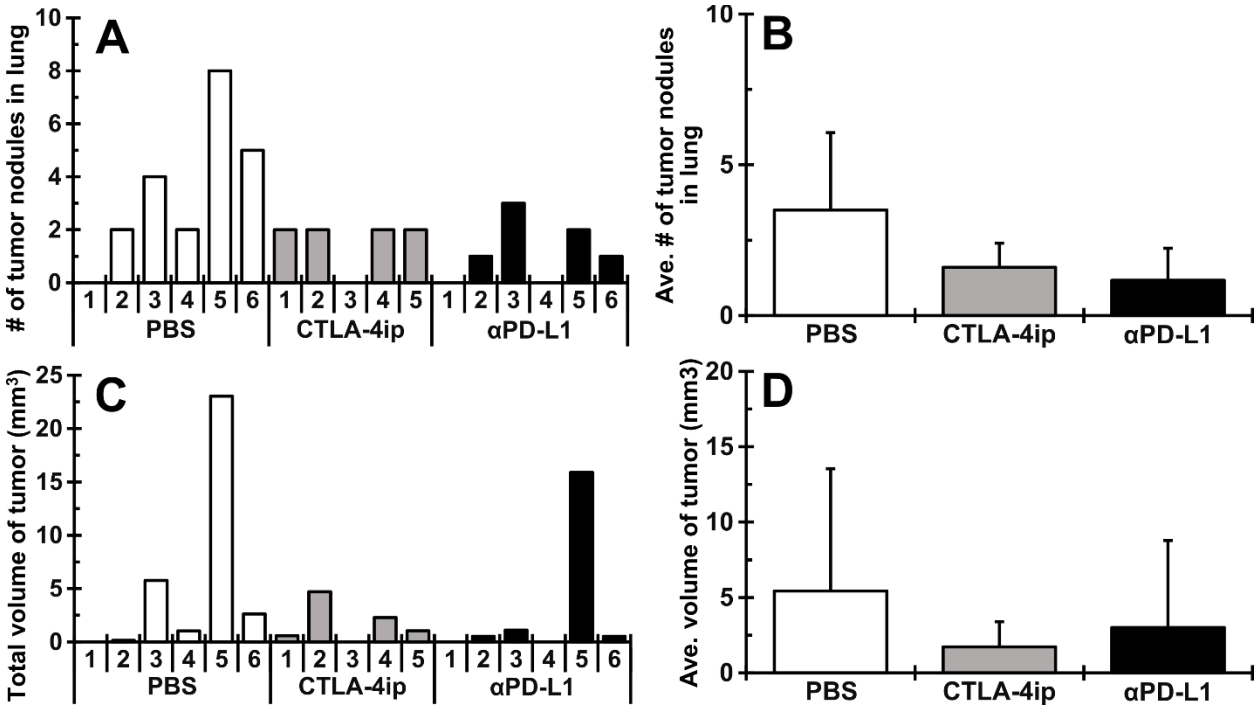

Figure 3.9 The designed peptide R10 attenuates the tumor growth in mouse lung.

Mice were treated with PBS (JAWS-irrLLC alone as negative control), the designed peptide R10 (CTLA4-ip, 10mg/kg/day) or anti-PD-L1 antibody (10 mg/kg/day, as positive control) and the number of tumor nodules and its size were noted after 5 days of treatment. **(A)** The number of tumor nodules in each mouse and **(B)** the average number of tumor nodules in each group. **(C)** The volume of tumor in each mouse and **(D)** the average volume of tumor in each group.

3.3.9 Effects of the designed peptide R10 on the T cell population and an expression of immune check points

Leukocytes were isolated from the peripheral blood of each mouse and cell population of CD4+ and CD8+ T cells was enumerated to evaluate the effects of the designed peptide R10 on the T-cell population and expression of immune checkpoint proteins using flow cytometry.

As per **Fig. 3.10A**, in comparison to non-treated JAWS-irrLLC alone (1.05 ± 1.06%) and anti-PD-L1 mouse antibodies 0.83 ± 1.26%), designed peptide R10, CTLA4-ip (0.50 ± 0.74%) decreased T-regulator cell population. As per**,** the population of cytotoxic T-lymphocyte was found to be reduced when mice were treated with anti-PD-L1 antibody (3.01 ± 2.45%), while no significant difference in the cytotoxic T-lymphocyte population was detected between JAWS-irrLLC alone group (5.10 ± 4.26%) and designed peptide R10 group (5.33 ± 1.85%). In all three groups, there was no clear difference in IFNγ and PD-1 expression levels **(Fig. 3.10B)**. As shown in **Fig. 3.10C,** CD8+ cytotoxic T cell population did not affected by either of the treatment, but peptide R10 treatment tended to increase CTLA4 expression in T-regulatory cells (MFI: 863.02 ± 269.34) compared to JAWS-irrLLC alone (MFI: 624.98 ± 324.71) and anti-PD-L1 antibody treated groups (MFI:570.84 ± 292.13). In contrast, treatment with the designed peptide R10 reduces CTLA4 expression in CD8+ IFNγ+ activated cytotoxic T cells (MFI: 9.71 ± 10.79) compared to the JAWS-irrLLC alone (MFI: 18.52 ± 27.59) and αPD-L1 group (MFI: 41.52 ± 79.83) **(Fig. 3.10D)**. Based on these results, it is possible that the immune-suppressive effect of T-regulatory cell on CD8+ cytotoxic T cells was effectively inhibited by the peptide R10 treatment, and that an anticancer immune response led to reduced tumor growth.

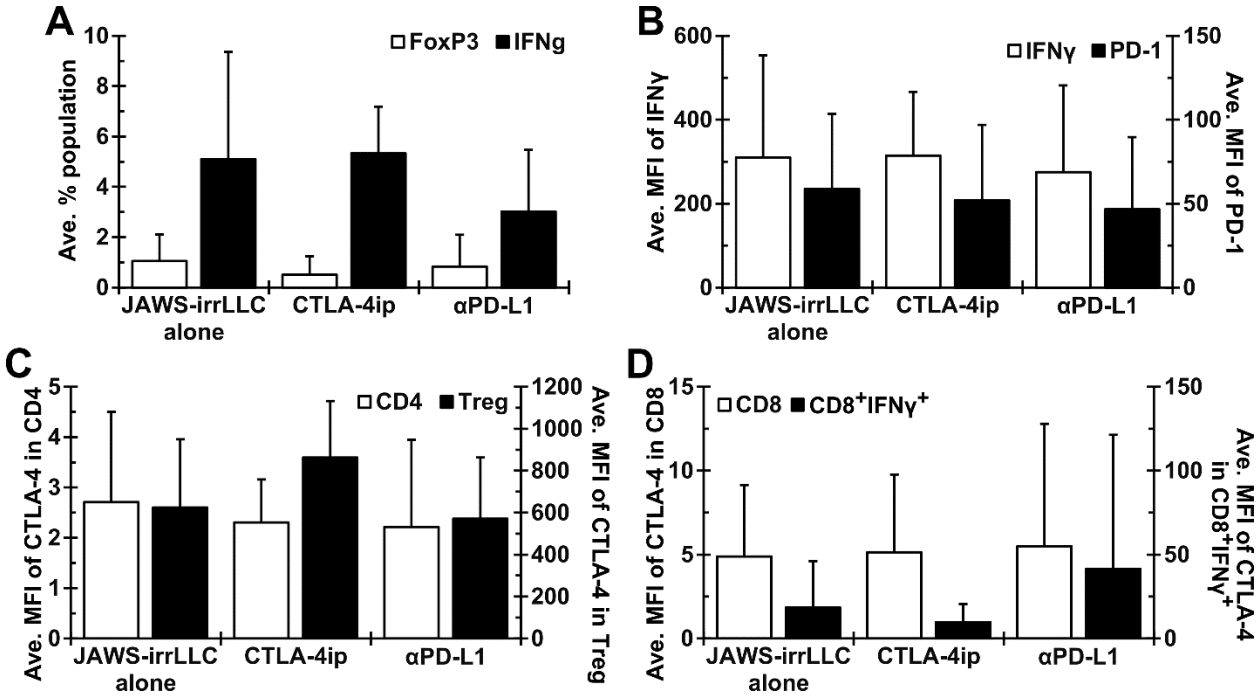

Figure 3.10 Population of CD4+, FoxP3+ regulatory T cell and CD8+, IFNγ+ cytotoxic T cell, and an expression level of immune check point proteins from the mouse blood.

(A) Population of CD4+ FoxP3+ regulatory T cell (Treg) and CD8+ IFNγ+ cytotoxic T cell (IFNγ) (an average percentage). **(B)** Mean fluorescence intensity (MFI) of IFNγ and PD-1 in CD8+ cytotoxic T cells. **(C)** Mean fluorescence intensity (MFI) of CTLA4 in CD4+ T cells (open bar) and Treg (filled bar). **(D)** Mean fluorescence intensity (MFI) of CTLA4 in CD8+ cytotoxic T cell (open bar) and CD8+ IFNγ+ activated cytotoxic T cell (filled bar).

3.4 Discussion

As we mentioned in the above section, only two peptides (R7) and (R10), have shown significant binding activity with CTLA4 protein. The peptide R7 was too hydrophobic and could not be synthesized. Hence, we focused our study on the interaction of peptide R10 and CTLA4.

3.4.2 Interaction of the R10 peptide with CTLA4

A two-microsecond-long MD simulation of the designed peptide (R10) at the binding surface of the CTLA4 where B-7 family protein binds gives sufficient conformational sampling to allow us to identify stable interactions. As we can see from the **(Fig. 3.11)** three hydrogen bonds,

one hydrophobic interaction and one π–π stacking interaction can be identified between the R10 peptide and CTLA4 protein. **Fig. 3.12** shows some of these interactions in a 3D view. As shown in **Fig. 3.13A** shows the distance between the alcohol hydrogen atom of Thr53 of CTLA4 and the carbonyl oxygen of Ala13 of the peptide. **Fig. 3.13B** shows the minimum distance between any hydrogen atom of the guanidinium group of Arg35 of CTLA4 and the carbonyl oxygen atom of Arg15 of the peptide. **Fig. 3.13C** shows the minimum distance between the oxygen atom of the carboxylate group of Glu48 of CTLA4 and any hydrogen atom of the guanidinium group of Arg14 of the peptide. All of these contacts are stable and stay bound for more than 1 microsecond. Furthermore, the aromatic ring of Tyr16 from the R10 peptide occupies the space between the aromatic rings of Tyr100 and Try104 of conserved motif (MYPPPY loop) of CTLA4 protein and makes a stable π–π stacking with Tyr104 of CTLA4 **(Fig. 3.13E). Fig. 3.13D** indicates a hydrophobic contact between the aliphatic sidechain of Pro8 of the peptide and the aliphatic sidechain of Glu59 of CTLA4.

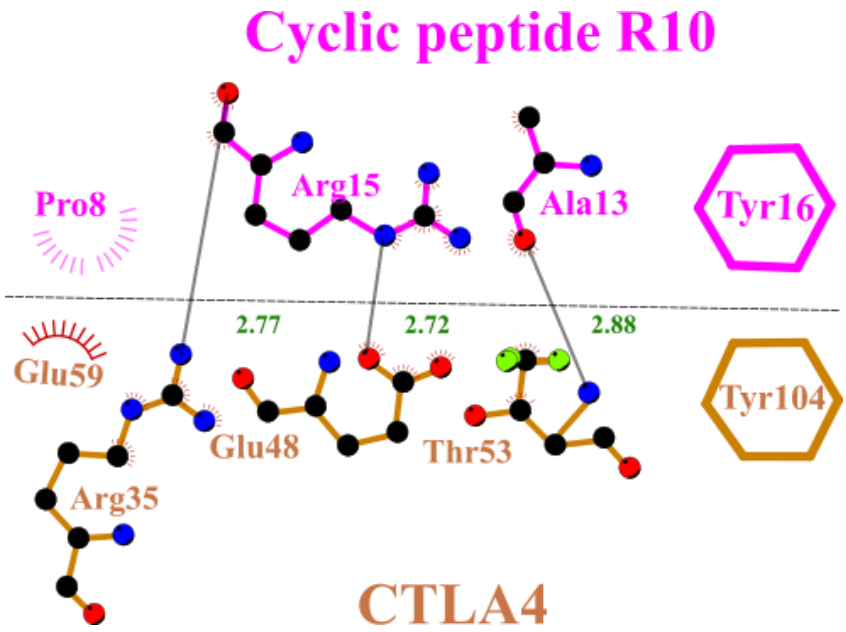

Figure 3.11 An atomic interaction between designed cyclic peptide R10 and CTLA4 in 2D.

A 2D diagram to illustrate the interaction between peptide R10 and CTLA4 in the frame of the MD simulation corresponding to the lowest MM-GBSA free energy. The residues (in a magenta) above the black dashed line belong to peptide R10, while those below (in orange) belong to CTLA4. The broken lines indicate hydrogen bonds between the R10 and CTLA4, while spoked arcs represent a hydrophobic contact.

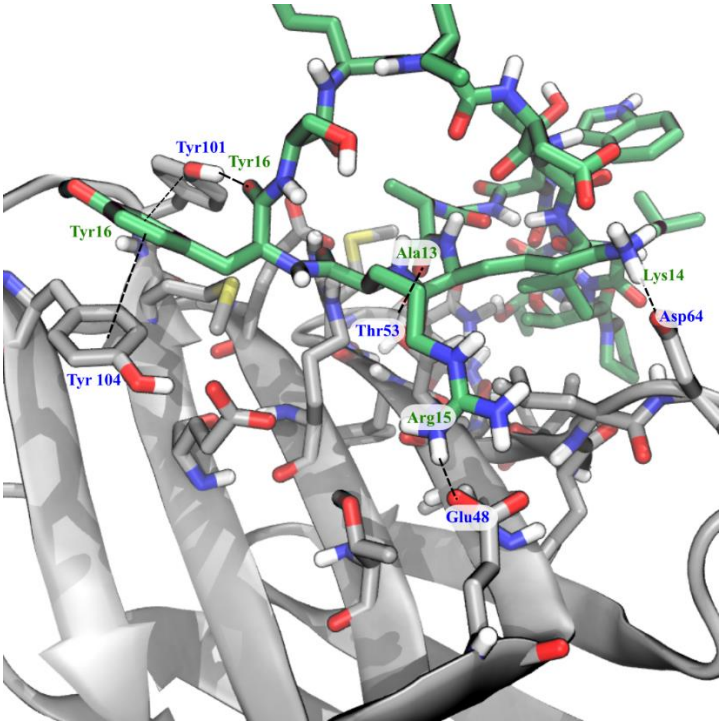

Figure 3.12 An atomic interaction between designed cyclic peptide R10 and CTLA4 in 3D.

An atomic interaction between the CTLA4 protein (grey cartoon) and the R10 peptide (green stick model) in 3D. Hydrogen bonds and π-π interaction is represented by dotted black lines.

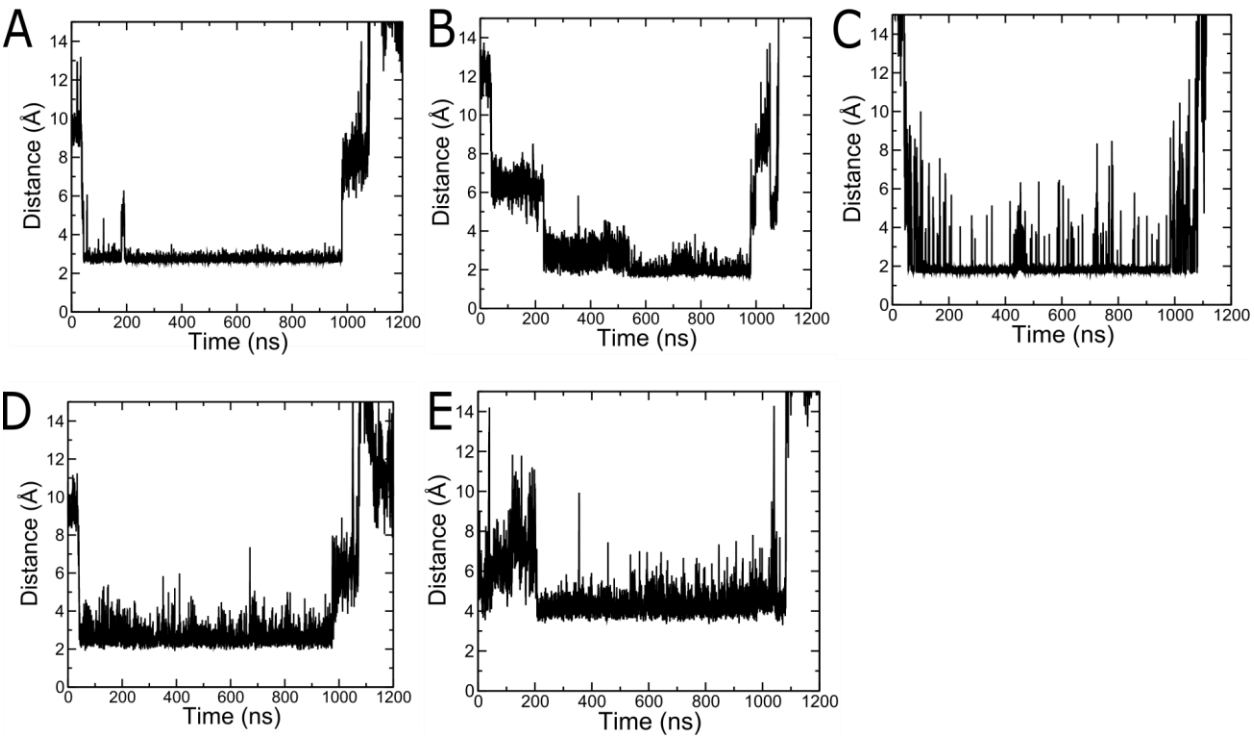

Figure 3.13 An atomic interaction between R10 peptide and CTLA4.

Hydrogen bonds between **(A)** the alcohol hydrogen atom of Thr53 of CTLA4 and the carbonyl oxygen of Ala13 of the peptide, **(B)** a hydrogen atom of the guanidinium group of Arg35 of CTLA4 and the carbonyl oxygen atom of Arg15 of the peptide, and **(C)** the oxygen atom of the carboxylate group of Glu48 of CTLA4 and a hydrogen atom of the guanidinium group of Arg14 of peptide. **(D)** Hydrophobic interaction the aliphatic sidechain of Pro8 of the peptide and the aliphatic sidechain of the Glu59 of CTLA4 (distance plotted is the minimum hydrogen-hydrogen distance). **(E)** Aromatic ring of Tyr16 making a π–π interaction with the aromatic ring of Tyr104 of CTLA4.

3.4.3 Conformational stability of the R10 peptide

We have analyzed the MD simulation of the peptide R10 in the absence of CTLA4 to study whether the peptide R10 adopts the conformation that is favorable to bind with CTLA4, or whether this conformation is induced upon binding. We extracted structure of peptide R10 from the lowest energy frame on the surface of CTLA4 as reference structure. An RMSD analysis on the trajectories of the R10 peptide shown that the conformation of R10 in the absence of CTLA4 fluctuates more than when R10 is bound **(Fig. 3.14)**. It can be seen that the R10 peptide adopts a stable conformation when it binds with its target protein CTLA4. In the presence of CTLA4, the R10

peptide remains stable on the surface of CTLA4 in the MD simulation, specifically between 500 to 1000 ns. Furthermore, the intramolecular hydrogen bonds between the backbone peptide groups of Val5-Val12, Thr4-Val12, and Thr7-Gly10 remain stable in the bound form, maintaining the β-sheet structure **(Fig. 3.15)**. These results suggest that the bound conformation of the peptide is not stable in solution and that the β-sheet structure is induced by the surface of the CTLA4 protein.

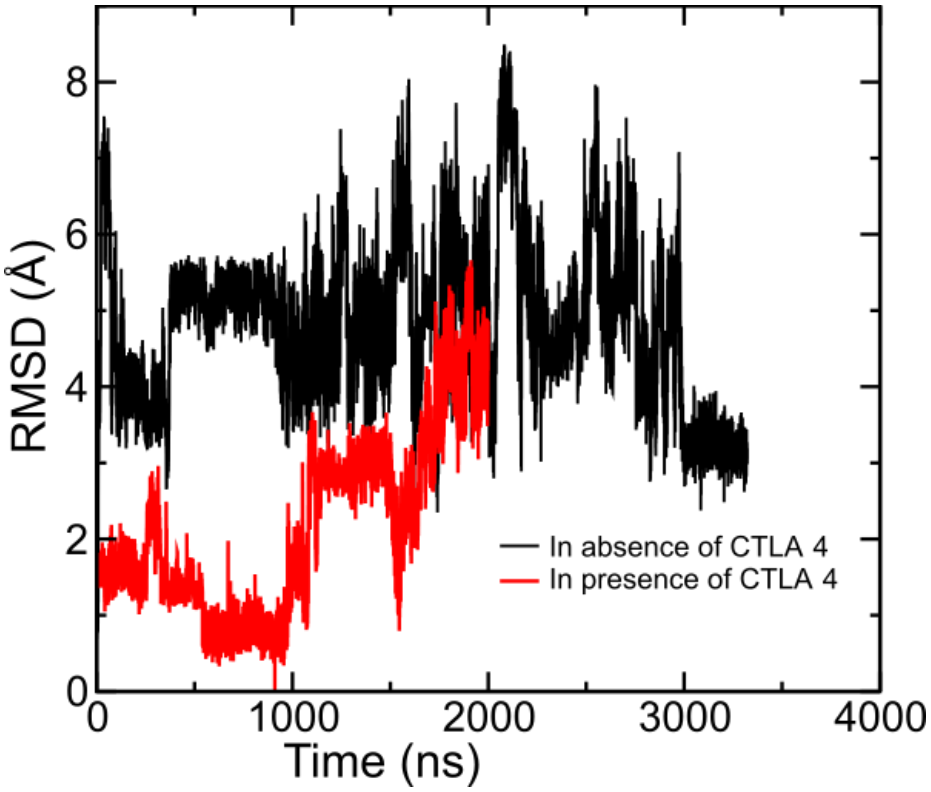

Figure 3.14 A conformational stability of peptide R10.

RMSD analysis of R10 peptide in the presence (red curve) and absence (black curve) of CLTA-4 protein comparing with lowest energy configuration of R10 peptide on the CTLA4 surface.

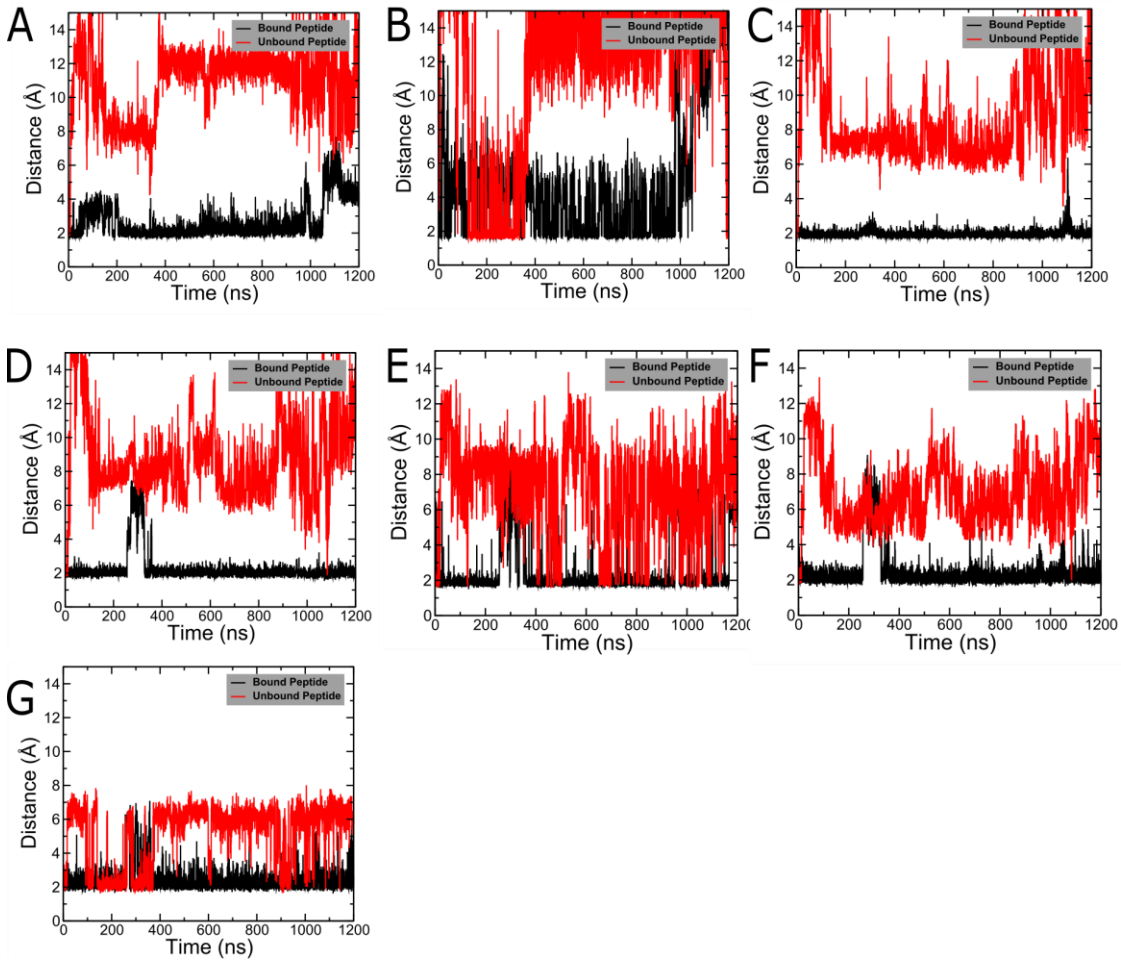

Figure 3.15 Intramolecular interactions of the R10 peptide when bound to CTLA4 and when in solution.

Intramolecular hydrogen bonds between **(A)** the backbone carbonyl oxygen of Asp3 and the NH group of Lys14, **(B)** a carboxylate oxygen from the sidechain of Asp3 and a hydrogen atom from the amino group of Lys14, **(C)** the NH group of Val5 and the carbonyl oxygen of Val12, **(D)** the NH group of Val12 and the carbonyl oxygen of Val5, **(E)** the alcohol hydrogen atom of Thr7 and the carbonyl oxygen of Gly10, **(F)** the NH group of Thr7 and the carbonyl oxygen of Gly10, and **(G)** the carbonyl oxygen of Thr7 and the NH group. Distances are shown for the CTLA4-bound and unbound states of the R10 peptide.

3.4.4 Bio-layer Interferometry analysis for the binding activity of R10 peptide for Template based protein

Although Surface Plasmon Resonance is more sensitive, we preferred the BLI method due to its flexibility in sample preparation. BLI can be used with a range of buffer solutions as this

method is less affected by a change in the refractive index.[61] An *in vitro* binding kinetics for designed peptide R10 are shown in **Fig. 3.7A** PBS buffer showed low nonspecific binding, but this response was selected as a baseline. Association and dissociation signals generated by CD-86 considered as a positive control and comparable with the responses obtained from different concentration of R10 peptide. Various concentrations (150, 175 and 200 μM) of R10 peptide were used to fit results globally and to determine dissociation constant (K_d). To improve the signal-to-noise ratio concentrations of the peptide R10 were used above its predicted (K_d) value, so that a sufficient number of binding events can be obtain. BLI based binding kinetics results indicate that our *de novo* protein design method can design the inhibitory peptide for the Template based protein with significant binding affinities.

3.4.10 Correlation between tumor growth, immune cell population, and an expression of immune checkpoint proteins.

The mouse with the high tumor growth showed high populations of T-regulatory and cytotoxic for each group of mice (**Fig. 3.16A**). As per shown in the **Fig. 3.16B,** IFNγ and PD-1 were found to be highly expressed in CD8+ cytotoxic T cells. As a result of induction of exhaustion in CD8+ cytotoxic T cells, tumor growth in the mouse appears to be promoted. A mice group treated with anti-PD-L1 antibody reduces the expression of IFNγ and PD-1 and reduces the tumor growth. On other hand, a high level of CTLA4 expression was noted in CD4+ cells (**Fig. 3.16C**), CD8+ cytotoxic T cells (**Fig. 3.16E**), and CD8+ IFNγ+ activated cytotoxic T cells (**Fig. 3.16F**). In spite of anti-PD-L1 antibody treatment, a tumor growth may be accelerated by the inhibition of CD8+ cytotoxic T cells by CTLA4 pathway. As can be seen in **Fig. 3.10** and **Fig. 3.16D**, the designed peptide R10 treated group had a higher level of CTLA4 expression than the other two groups.

According to these results, the treatment of the designed peptide R10 inhibits the immune suppressive effects of T-regulatory cells for CD8+ cytotoxic T cell and prevents the growth of tumor. An anti-cancer activity of the designed peptide R10 was supported by the analysis of tumor growth and leukocyte count.

.

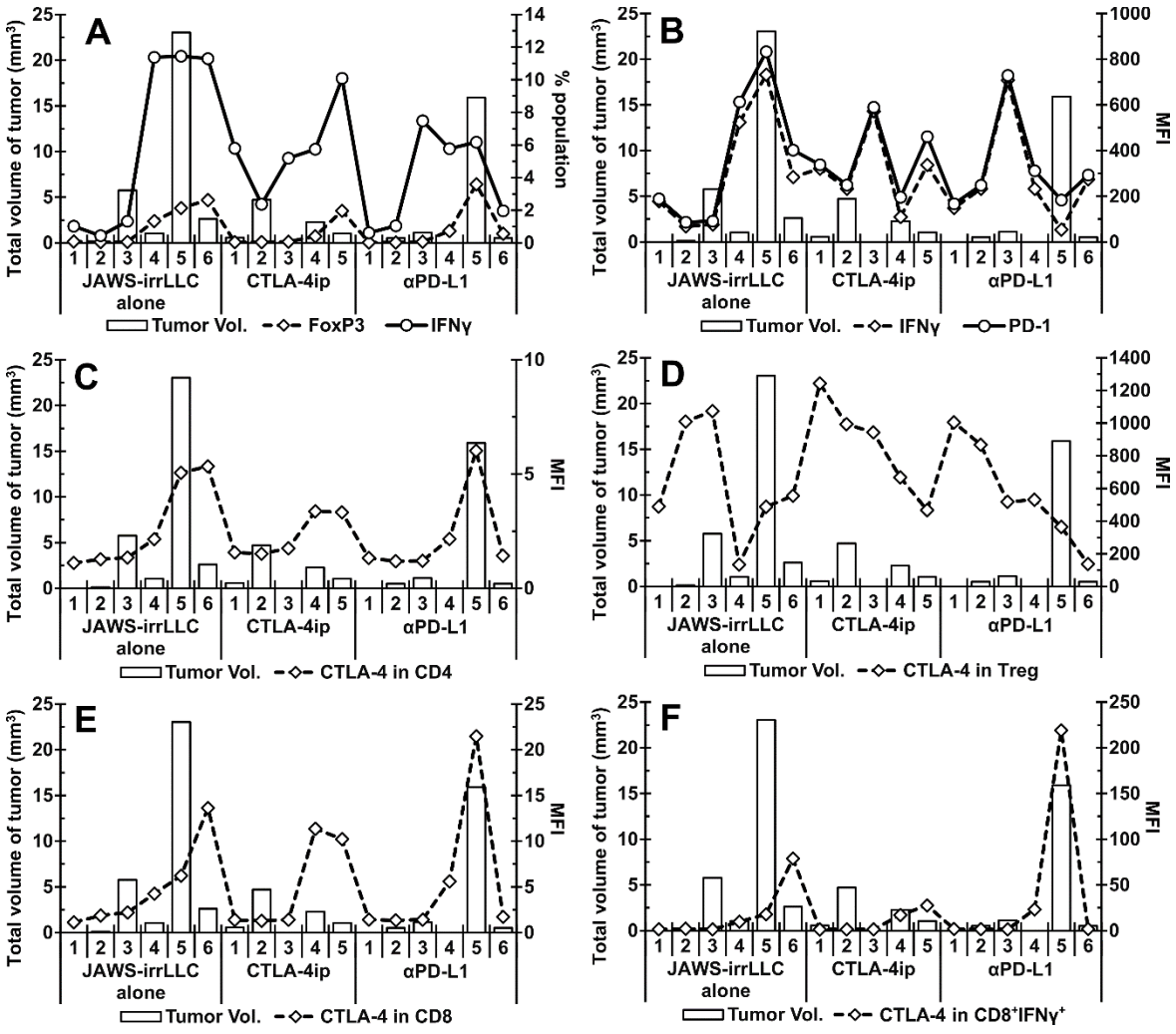

Figure 3.16 Correlation between tumor growth, immune cell population, and an expression of immune checkpoint proteins.

The tumor volume of each mouse were correlated with the **(A)** population of CD4+ FoxP3+ regulatory T cell (Treg) and CD8+ IFNγ+ cytotoxic T cell (IFNγ), **(B)** mean fluorescence intensity (MFI) of IFNγ and PD-1 in CD8+ T cell, **(C)** MFI of CTLA4 in CD4+ T cell, **(D)** MFI of

CTLA4 in Treg, **(E)** MFI of CTLA4 in CD8+ cytotoxic T cell, and **(F)** MFI of CTLA4 in CD8+ IFNγ+ activated T cell.

3.5 Conclusion

The biolayer interferometry results indicate that the computationally predicted binding affinity value is reliable and that our *de novo* peptide design algorithm can be helpful in designing inhibitory peptides for disease-related proteins that can be used as therapeutics.

3.6 References

1. He, X. & Xu, C. Immune checkpoint signaling and cancer immunotherapy. *Cell Res* **30**, 660–669 (2020).

2. Wilky, B. A. Immune checkpoint inhibitors: The linchpins of modern immunotherapy. *Immunol Rev* **290**, 6–23 (2019).

3. Margolin, K. *et al.* Ipilimumab in patients with melanoma and brain metastases: an open-label, phase 2 trial. *Lancet Oncol* **13**, 459–465 (2012).

4. Leach, D. R., Krummel, M. F. & Allison, J. P. Enhancement of Antitumor Immunity by CTLA-4 Blockade. *Science* **271**, 1734–1736 (1996).

5. Dyck, L. & Mills, K. H. G. Immune checkpoints and their inhibition in cancer and infectious diseases. *Eur J Immunol* **47**, 765–779 (2017).

6. Rotte, A. Combination of CTLA-4 and PD-1 blockers for treatment of cancer. *Journal of Experimental & Clinical Cancer Research* **38**, 255 (2019).

7. Li, K. & Tian, H. Development of small-molecule immune checkpoint inhibitors of PD-1/PD-L1 as a new therapeutic strategy for tumour immunotherapy. *J Drug Target* **27**, 244–256 (2019).

8. Verma, V. *et al.* A systematic review of the cost and cost-effectiveness studies of immune checkpoint inhibitors. *J Immunother Cancer* **6**, 128 (2018).

9. Skalniak, L. *et al.* Small-molecule inhibitors of PD-1/PD-L1 immune checkpoint alleviate the PD-L1-induced exhaustion of T-cells. *Oncotarget* **8**, 72167–72181 (2017).

10. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489 (2021).

11. Chothia, C., Hubbard, T., Brenner, S., Barns, H. & Murzin, A. Protein folds in the all-beta and all-alpha classes. *Annu Rev Biophys Biomol Struct* **26**, 597–627 (1997).

12. Schwartz, J.-C. D., Zhang, X., Fedorov, A. A., Nathenson, S. G. & Almo, S. C. Structural basis for co-stimulation by the human CTLA-4/B7-2 complex. *Nature* **410**, 604–608 (2001).

13. Nevola, L. & Giralt, E. Modulating protein-protein interactions: the potential of peptides. *Chem Commun (Camb)* **51**, 3302–3315 (2015).

14. AlDeghaither, D., Smaglo, B. G. & Weiner, L. M. Beyond Peptides and mAbs - Current Status and Future Perspectives for Biotherapeutics with Novel Constructs. *J Clin Pharmacol* **55**, S4–S20 (2015).

15. Komin, A., Russell, L. M., Hristova, K. A. & Searson, P. C. Peptide-based strategies for enhanced cell uptake, transcellular transport, and circulation: Mechanisms and challenges. *Adv Drug Deliv Rev* **110–111**, 52–64 (2017).

16. Song, H. *et al.* Injectable polypeptide hydrogel-based co-delivery of vaccine and immune checkpoint inhibitors improves tumor immunotherapy. *Theranostics* **9**, 2299–2314 (2019).

17. Zhang, L. *et al.* Peptide-based materials for cancer immunotherapy. *Theranostics* **9**, 7807–7825 (2019).

18. Thundimadathil, J. Cancer treatment using peptides: current therapies and future prospects. *J Amino Acids* **2012**, 967347 (2012).

19. Góngora-Benítez, M., Tulla-Puche, J. & Albericio, F. Multifaceted Roles of Disulfide Bonds. Peptides as Therapeutics. *Chem. Rev.* **114**, 901–926 (2014).

20. Gao, M., Cheng, K. & Yin, H. Targeting protein-protein interfaces using macrocyclic peptides. *Biopolymers* **104**, 310–316 (2015).

21. Khoury, G. A., Smadbeck, J., Kieslich, C. A. & Floudas, C. A. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol* **32**, 99–109 (2014).

22. Leman, J. K. *et al.* Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods* **17**, 665–680 (2020).

23. Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annu Rev Biochem* **77**, 363–382 (2008).

24. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).

25. Fiser, A. & Šali, A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. in *Methods in Enzymology* vol. 374 461–491 (Academic Press, 2003).

26. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit-5.6 (2006).

27. Raveh, B., London, N., Zimmerman, L. & Schueler-Furman, O. Rosetta FlexPepDock ab-initio: Simultaneous Folding, Docking and Refinement of Peptides onto Their Receptors. *PLOS ONE* **6**, e18934 (2011).

28. Woo, H.-J. & Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6825–6830 (2005).

29. Fu, H., Cai, W., Hénin, J., Roux, B. & Chipot, C. New Coarse Variables for the Accurate Determination of Standard Binding Free Energies. *J. Chem. Theory Comput.* **13**, 5173–5178 (2017).

30. Renfrew, P. D., Choi, E. J., Bonneau, R. & Kuhlman, B. Incorporation of Noncanonical Amino Acids into Rosetta and Use in Computational Protein-Peptide Interface Design. *PLOS ONE* **7**, e32637 (2012).

31. Garton, M. *et al.* Rapid and accurate structure-based therapeutic peptide design using GPU accelerated thermodynamic integration. *Proteins* **87**, 236–244 (2019).

32. Concepcion, J. *et al.* Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. *Comb Chem High Throughput Screen* **12**, 791–800 (2009).

33. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).

34. Darnell, S. J., Page, D. & Mitchell, J. C. An automated decision-tree approach to predicting protein interaction hot spots. *Proteins* **68**, 813–823 (2007).

35. Zhu, X. & Mitchell, J. C. KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins* **79**, 2671–2683 (2011).

36. Kaufmann, K. W., Lemmon, G. H., DeLuca, S. L., Sheehan, J. H. & Meiler, J. Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You. *Biochemistry* **49**, 2987–2998 (2010).

37. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845–858 (2015).

38. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).

39. Kuhlman, B. Designing protein structures and complexes with the molecular modeling program Rosetta. *J Biol Chem* **294**, 19436–19443 (2019).

40. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802 (2005).

41. Huang, J. *et al.* CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **14**, 71–73 (2017).

42. Comparison of simple potential functions for simulating liquid water: The Journal of Chemical Physics: Vol 79, No 2. https://aip.scitation.org/doi/abs/10.1063/1.445869.

43. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952–962 (1992).

44. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).

45. Brünger, A., Brooks, C. L. & Karplus, M. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters* **105**, 495–500 (1984).

46. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).

47. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

48. Hestenes, M. R. & Stiefel, E. Methods of conjugate gradients for solving linear systems. 28.

49. Fiorin, G., Klein, M. L. & Hénin, J. Using collective variables to drive molecular dynamics simulations. *Molecular Physics* **111**, 3345–3362 (2013).

50. Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **55**, 383–394 (2004).

51. Rastelli, G., Del Rio, A., Degliesposti, G. & Sgobba, M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J Comput Chem* **31**, 797–810 (2010).

52. Vergara-Jaque, A., Comer, J., Monsalve, L., González-Nilo, F. D. & Sandoval, C. Computationally Efficient Methodology for Atomic-Level Characterization of Dendrimer–Drug Complexes: A Comparison of Amine- and Acetyl-Terminated PAMAM. *J. Phys. Chem. B* **117**, 6801–6813 (2013).

53. Gumbart, J. C., Roux, B. & Chipot, C. Efficient Determination of Protein–Protein Standard Binding Free Energies from First Principles. *J. Chem. Theory Comput.* **9**, 3789–3798 (2013).

54. Fu, H. *et al.* BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **58**, 556–560 (2018).

55. Darve, E. & Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **115**, 9169–9183 (2001).

56. The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask | The Journal of Physical Chemistry B. https://pubs.acs.org/doi/10.1021/jp506633n.

57. Metzler, W. J. *et al.* Solution structure of human CTLA-4 and delineation of a CD80/CD86 binding site conserved in CD28. *Nat Struct Mol Biol* **4**, 527–531 (1997).

58. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188–1190 (2004).

59. Salmaso, V. & Moro, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front Pharmacol* **9**, (2018).

60. van der Merwe, P. A., Bodian, D. L., Daenke, S., Linsley, P. & Davis, S. J. CD80 (B7-1) binds both CD28 and CTLA-4 with a low affinity and very fast kinetics. *J Exp Med* **185**, 393–403 (1997).

61. Shah, N. B. & Duncan, T. M. Bio-layer Interferometry for Measuring Kinetics of Protein-protein Interactions and Allosteric Ligand Effects. *JoVE (Journal of Visualized Experiments)* e51383 (2014) doi:10.3791/51383.

4 *De novo* design of a stapled peptide targeting SARS-CoV-2 spike protein receptor-binding domain.

Ravindra Thakkar[1], Dilip Agarwal[2], Chathuranga B Ranaweera[3], Masaaki Tamura[1] and Jeff Comer[1].

[1], Department of Anatomy & Physiology, Kansas State University College of Veterinary Medicine, Manhattan, KS.

[2], Department of Material Science and Engineering and NUANCE Center, Northwestern University, Evanston, IL

[3], Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, KS.

Corresponding Author: Jeffrey Comer

Associate Professor

Department of Anatomy & Physiology

College of Veterinary Medicine

Kansas State University

Manhattan, KS 66506

Phone: (785) 532-6311

E-mail: jeffcomer@vet.k-state.edu

4.4 Abstract

Although effective vaccines have been developed against SARS-CoV-2, many regions in the world still have low rates of vaccination and new variants with mutations in the viral spike protein have reduced the effectiveness of most available vaccines and treatments. There is an urgent need for a drug to cure this disease and prevents infection. The SARS-CoV-2 virus enters the host cell through protein-protein interaction between the virus's spike protein and the host's Angiotensin Converting Enzyme (ACE-2). We have designed a 17-residue peptide, using protein design software and molecular dynamics simulations, that binds with the spike protein receptor-binding domain (RBD) and blocks interaction of spike protein with ACE-2. We have confirmed the binding activity of the designed peptide for the original spike protein and the Delta variant spike protein using micro-cantilever and bio-layer interferometry (BLI) based methods.

4.5 Introduction

A new infectious respiratory disease was reported in Wuhan, China, in December 2019.[1] The virus causing this disease was identified as a novel coronavirus and termed as COVID-19 by the WHO. This virus has become a threat to global public health and the economy. As of today, more than 453 million people are infected, and 6 million deaths are reported due to this worldwide pandemic.[2] This disease continues to cause disability and death across the globe, even after effective vaccines have been developed against SARS-CoV-2. Additionally, vaccination is less effective for immuno-compromised patients.[3] Cases of reinfection in patients who had fully recovered from COVID-19 have been reported.[4,5] Antibodies against SARS-CoV-2 seem to have a short life span, and the titer decreased in few months after the onset of symptoms.[6] An inconsistency in the

ability to produce effective antibodies against the spike protein has been observed in various patients.[7,8] An effective antiviral drug could complement vaccination and be a powerful tool in the continuing fight against COVID-19.[9]

The virus causing COVID-19 exhibits 80% sequence similarity with SARS-CoV, a virus that emerged in 2002–2003, which is why the virus has been dubbed SARS-CoV-2.[10] These coronaviruses enter the host cells by binding to the Angiotensin Converting Enzyme (ACE2), a receptor protein on surface of human cells. The receptor binding domain (RBD) of the prominent viral spike protein is responsible for this binding.[11] The spike protein of SARS-CoV-2 has evolved to bind ACE2 with high affinity and is an important factor in its high contagiousness.[12] X-ray crystallography of the complex between the SARS-CoV-2 spike protein and ACE2 has revealed some of the key amino acid residues where the ACE2 protein binds and provides valuable information needed to design therapeutic drugs that can block the spike protein surface, preventing entry of the of coronavirus into the host cell.[13]

Several small drug molecules have been predicted to target the spike protein RBD by computational studies. [14,15] However, small molecules are, by their nature, too small to occupy the entire portion of the RBD surface that forms the interface with ACE2 **(Fig. 4.1A)**. Synthesis of custom peptides has become routine and commercialized, while synthesis of novel conventional small-molecule drugs remains specialized.[16,17] Previous studies have reported that linear therapeutic peptides based on the human ACE2 alpha-1 helix have been shown to have an inhibitory effect on the interaction between ACE2 and the spike protein RBD.[18,19] However, linear peptides have poor conformational and proteolytic stability. Linear peptides are typically quite flexible and,

therefore, the entropic cost for adopting a more restricted conformation when bound to the target is high.[20] The introduction of chemical crosslinks in the peptide, referred to as staples, helps to maintain the bound conformation and reduce the entropic cost for binding. Some stapled peptides that bind to the SARS-CoV-2 spike protein have been developed, but they did not prevent virus internalization into host cells.[21] None of the work has discussed the effect of the designed peptides on the variants of the spike protein. These studies also did not address the Delta and Omicron variants of SARS-CoV-2, which have significantly mutated spike proteins that are less sensitive to host antibodies from recovered or vaccinated individuals.[17,18]

This study presents the *de novo* computational design of a stapled peptide and evaluation of its binding activity with the spike protein RBD and those of variants by molecular simulation and label-free binding techniques like micro-cantilever and bio-layer interferometry.

4.6 Materials and methods

4.6.10 Peptide modeling

We analyzed the x-ray structure of the complex between the receptor-binding domain of the spike protein and the ACE2 protein (PDB ID: 6LZG). To make peptide template, a fragment containing residues 26 to 42 were extracted from the ACE2 protein using VMD version 1.9.4 **(Fig. 4.1B)**.[22]

4.6.11 Flexible docking of template peptide with Spike protein RBD and sequence optimization

Flexible peptide docking was performed using the FlexPepDock module of the Rosetta molecular modeling suite.[23] Low energy conformations of the template peptide on the surface of

spike protein RBD were generated. Based on energy score and root mean square deviation (RMSD) from the initial conformation, out of 500, around 3–5 peptide poses were selected for sidechain and sequence optimization. The optimization algorithm iterates between a conformational optimization phase, where it attempts to find the lowest energy conformation of the sidechains, and the design phase, where the algorithm applies the substitution of user-defined residues that attempt to lower the energy of optimized conformation. The conformational optimization phase includes rotational and translational movement of the entire peptide while keeping the carbon backbone fixed. [24–26]

4.6.12 Bio-molecular system preparation

All the receptor protein models were built from the x-ray crystal structure of the spike protein RBD and human ACE2 complex (PDB ID:6LZG). All structures were parameterized using the CHARMM36m force field[27] and input generator module of the CHARMM-GUI webserver.[28] Glycosylation was performed, and disulfide bonds between residues 379–432; 488–480; 391–525; 336–361 were added. Models of spike protein RBD variants Alpha, Beta, Gamma, and Delta were generated by making the mutations N501Y; K417N, E484K, N501; K417T, E484K, N501 and E484Q, L452R respectively. The peptide structures with the optimized sequence were added to the appropriate binding site on spike protein RBD using ZDOCK, a protein-peptide docking algorithm.[29] The each protein-peptide complex was solvated using ≈9000 molecules of water (TIP3P water model), and 150 mM of sodium chloride ions were added, with additional ions to neutralize the system[30]. The overall volume of the system was (80 Å)^{3,} and the number of atoms was ≈50,000.

4.6.13 Molecular dynamics simulation and MMGBSA calculation to estimate binding free energy

All molecular dynamics simulations were performed in the NPT ensemble using the program NAMD version 2.13, where a Langevin thermostat was applied to maintain a temperature of 310 K and the Langevin piston barostat algorithm was used to maintain a pressure of 1 standard atmosphere.[31–33] Interatomic forces were defined by CHARMM36m force field. The Lennard–Jones interaction between pair of atoms calculated using a smooth 10–12 Å cut off distance and electrostatic interaction measured using Particle Mesh Ewald (PME) using 1.2 Å grid space.[30,31] Energy minimization for each system was performed for 1 ns and followed by a production simulation for 2 μs without applying any restraints on the atoms. The NAMD Colvars module was applied to terminate the MD simulation if the conformation of the peptide changed more than cut off value (RMSD > 15 Å) from the bound state.[36] At every 200 ps, the configuration of the biomolecular system was collected for further analysis.

The binding free energy for frame of the MD simulation for each system was estimated using molecular mechanics generalized Born surface area (MMGBSA)[37,38] method as described in **Eq. 4.1**. The implementation of this method involves an implicit solvent with a dielectric constant of 78.5 and a surface tension of 0.00542 kcal/(mol Å^2) to calculating the solvation free energy of the extracted protein, peptide, and protein-peptide complex for each frame of the MD trajectories.

$$\Delta G_{\text{binding}}^{\text{GBSA}} = \Delta G_{\text{protein:peptide}}^{\text{GBSA}} - \Delta G_{\text{protein}}^{\text{GBSA}} - \Delta G_{\text{peptide}}^{\text{GBSA}}$$

Equation 4.1 Binding free energy.

4.6.14 Stapling of better performing peptides to maintain the favorable binding conformation

Peptides were shortlisted based on the MMGBSA score. The most favorable configuration of the peptides was extracted. To maintain this most favorable conformation, sidechains that are away from the binding interface were selected for the i + 4 stapling process. Selected sidechains of aliphatic amino acids were linked using propane. Side chains of charged residues were connected using extra peptide bond. The stapled peptide was parameterized using CHARMM36m force field.[39] MD simulations and MMGBSA calculations were performed for each system of stapled peptide and spike protein RBD.

4.6.15 Absolute binding free energy calculation by geometric route

The configuration corresponding to the lowest MMGBSA energy was extracted and used as an input for the Binding Free Energy Estimator (BFEE) plugin of VMD version 1.9.4.[40] BFEE subdivides the binding free energy calculation into different sub-processes (**Table–4.1**). The key idea is that calculating the free energy for unbinding the peptide is much more efficient if artificial restraints are applied to the conformation and orientation of the peptide relative to the receptor.[52] In fact, the calculation is not feasible without these restraints. However, the restraints bias the result, so their effect must be calculated and removed to obtain the unbiased binding free energy. First, the free energy cost of adding these restraints to the bound peptide is determined by calculating a potential of mean force along each restrained coordinate using the extended adaptive biasing force (eABF) method as implemented in the Colvars module.[41,42] This comprises sub-processes 1–6 in **Table 4.2**. Next the free energy of unbinding the restrained peptide is calculated (sub-process 7). The free energy gained by releasing the restraints for the unbound peptide in solution are then calculated (sub-processes 8 and 9). Fewer MD simulations are needed to calculate

the effect of releasing the restraints because some can be computed analytically owing to the isotropy of the unbound peptide in solution. The absolute binding free energy is calculated using **Eq. 4.2**.

$$\Delta G^{0}_{\text{binding}} = k_{\text{B}}T \ln C^{0}$$

$$K_{\text{eq}} = \exp\{-\beta\left(\Delta G_{c}^{\text{site}} + \Delta G_{0}^{\text{site}} + \Delta G_{\alpha}^{\text{site}} - k_{\text{B}}T \ln (S^{*}I^{*}C^{0}) + \Delta G_{0}^{\text{bulk}} + \Delta G_{c}^{\text{bulk}}\right)\}$$

where $C^{0} = \frac{1}{1660.539}\text{Å}^{-3}$ is the standard 1 mol/L concentration.

Equation 4.2 Binding free energy calculation by the explicit-solvent geometric route.

Table 4.1 Sub-processes to calculate an entropic contribution involved in the protein-peptide association.

Stage	System	Free-Energy Term	Degree of freedom
1	protein-ligand	$\Delta G_{\text{conform}}$	RMSD of peptide in bound form.
2	protein-ligand	ΔG_{Θ}	Movement of peptide in Euler angle (Θ) with respect to protein
3	protein-ligand	ΔG_{φ}	Movement of peptide in Euler angle (φ) with respect to protein
4	protein-ligand	ΔG_{ψ}	Movement of peptide in Euler angle (ψ) with respect to protein
5	protein-ligand	ΔG_{θ}	Movement of peptide in Polar angle (θ) with respect to protein
6	protein-ligand	ΔG_{ϕ}	Movement of peptide in Polar angle (ϕ) with respect to protein
7	protein-ligand	$-k_{\text{B}}T \ln (\text{S}^{*}\text{I}^{*}\text{C}^{0})$	Separation of the peptide from protein
8	ligand only	$\Delta G_{\text{conform}}^{\text{unbound}}$	RMSD of the peptide in unbound form.
9	ligand only	$\Delta G_{\Theta\phi\Psi}^{\text{unbound}}$	Angular movement of unbound peptide. (calculated analytically)

4.6.16 Binding confirmation of the designed stapled peptide by the micro-cantilever method

We purchased silicon cantilevers from Nanoworld Inc. and the SARS-CoV-2 spike protein RBD, the antibody for the spike protein RBD and influenza H1N1 hemagglutinin protein were procured from Sino Biological Inc; and 1-Ethyl-3-(3-dimethyl aminopropyl) carbodiimide and sulfo-NHS from ThermoFisher Scientific. The designed stapled peptide was commercially synthesized from LifeTein LLC. The cantilever tips were plasma cleaned before immobilization. The microcantilevers were covalently immobilized by EDC-NHS chemistry using 100 μM of the designed staple peptide (referred to as Pep39) prepared in PBS and 0.05 % BSA (pH = 7.4) solution.[43] The immobilized micro-cantilevers were brought into a microfluidic chamber containing the spike protein RBD. Antibodies to the spike RBD protein and the influenza H1N1 hemagglutinin protein (1 μg/ml each) were used as positive and negative controls, respectively. All experiments were conducted on a Bruker Bioscope Resolve liquid imaging system at a constant temperature, and cantilever deflection was measured using an in-built optical detector.

4.6.17 Binding assay by the Bio-Layer Interferometry

The biotinylated recombinant SARS-CoV-2 spike protein RBD with His-tag, recombinant SARS-CoV-2 spike B.1.617.2 with His-tag and recombinant human ACE-2 protein were purchased from R&D Systems, Inc. High precision streptavidin (SAX) and anti-penta-His (high precision streptavidin (SAX)) biosensors obtained from the Sartorius Corporation. A solution of the spike protein at a 1 μg/ml concentration was loaded onto the corresponding hydrated biosensors. Each labeled biosensor was placed in different molar concentrations (120, 48, 24, 4.8, and 0.48 μM) of Pep39, and association was measured for 120 seconds and followed by dissociation with

PBS for 120 seconds. PBS buffer alone and 44 μg/ml human ACE-2 protein were used as a reference and a positive control, respectively. Experimental association and dissociation constants from all experiments were globally fitted using a 1:1 binding model to measure the dissociation constant K_d using the built-in software BLItzPro version 1.1. All binding assays were performed on the FortéBio BLItz instrument.

4.7 Results and discussion

4.7.10 Selection of the starting template for the peptide design

The experimentally determined x-ray crystal structure of the spike protein RBD and ACE2 protein complex (PDB ID: 6LZG) provided a high-resolution atomic model (2.50 Å). The α1–helix of the hACE2 protein occupies a long and flat interface of the spike protein RBD, which is reported to have an area of around 225 $Å^2$.[45] Residues Thr27, Phe28, Lys31, His34, Try41 and Lys353 of ACE2 protein were identified as key residues using KFC2, the knowledge-based protein-protein interface prediction webserver web-server **(Fig. 4.1A)**.[13,44] Hence, residues 26 to 42 from the α1–helix of the hACE-2 can provide an appropriate starting template. **(Fig. 4.1B)**.

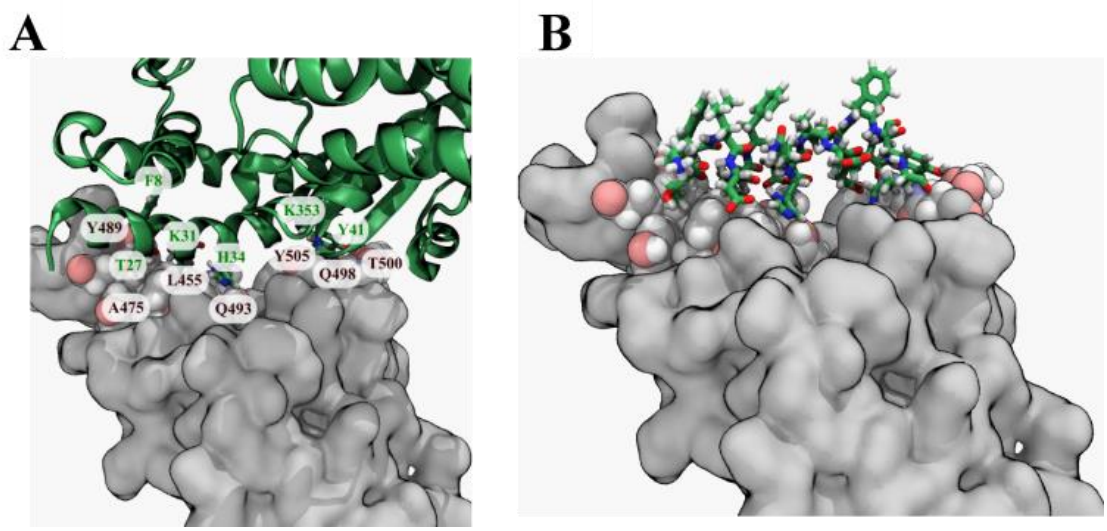

Figure 4.1 Template selection.

(A) The crystal structure of the spike protein RBD and ACE2 protein complex (PDB ID 6lzg): The spike protein RBD (grey surface) and its key residues (brown labels) interacting with amino acids (green labels) of the ACE2 protein (green ribbon). **(B)** A fragment of the ACE2 protein containing amino acid numbers 26–42 was selected as a template for the peptide design.

4.7.11 Docking and sequence modification to optimize peptide

We applied the structure based FlexPepDock protocol to design a peptide that binds the spike protein RBD. FlexPepDock performs a large-scale search of the backbone conformational space. We selected poses that were both low energies according to Rosetta and had structural similarity with the template peptide to create a peptide that binds to the RBD in the same location as the α1–helix of hACE2 protein. The poses were selected by plotting the docking score vs. RMSD **(Fig. 4.9)**. To increase the affinity of selected poses for the spike protein RBD, we applied a protein design protocol of Rosetta that performs sidechain and rotamer optimization to make the estimated binding energy more favorable. The selection of amino acids for substitution was unbiased (all 20 canonical proteinogenic amino acids were available). Although multiple rounds of optimization gave a variety of sequences, certain positions on the peptide favored particular residues. For example, the initial backbone structure was consistent with only proline and glycine at the 8th and 10th positions, respectively. The 2nd, 12th, 13^{th,} and 14th positions were dominated by hydrophobic residues. **Fig. 4.2** shows the occurrence of residues at each position in the peptide. [46]

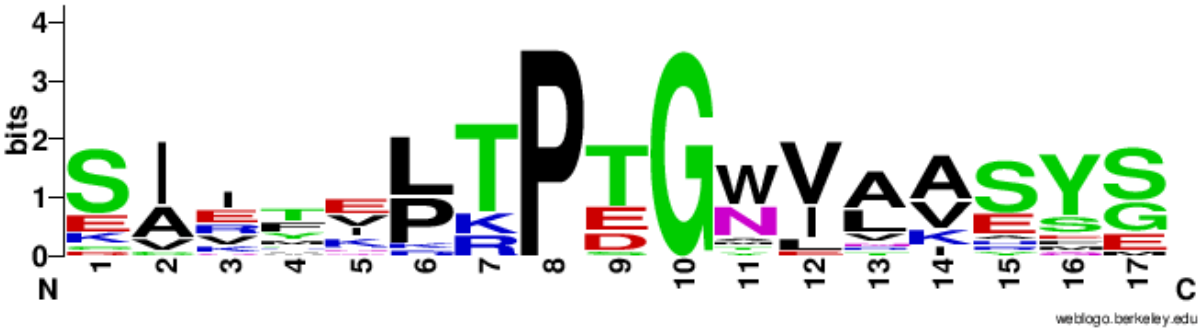

Figure 4.2 The plot of probabilities for occurrence of amino acids at specific positions in the designed peptides

The graphical representation of the frequency of residues at specific position on the peptide. The height of symbols represents the relative frequency for the occurrence of that residue at the corresponding position.

4.7.12 Estimation of the binding free energy of optimized peptides

We performed all-atom explicit solvent molecular dynamics simulations for 41 optimized peptide structures with the spike protein RBD. We set up the simulations such that a simulation would terminate if the peptide structure deviated (in RMSD of Cα atoms) more than 15 Å from the starting configuration. We then estimated binding free energy calculation for all simulations by the MMGBSA method. The simulated time for which the peptide remained bound (RMSD < 15 Å) was more correlated with the MMGBSA estimation of the binding free energy than the score assigned by the Rosetta. **(see Appendix C)**. The Rosetta scoring system uses empirical methods, and entropy components have not incorporated. These results in the undesirable folding of optimized peptides and the binding affinity may change.[25,47–50]

4.7.13 Effect of stapling on the backbone conformation of the peptide

Based on the MMGBSA score we shortlisted nine out of 41 optimized peptides. For all cases, the most favored conformation (according to the GBSA free energy function) of the peptide was completely changed from the starting conformation **(Fig. 4.8)**. To maintain the α-helical structure of the peptide and presumably reduce the entropic cost of binding, we added a propene staple between Val13 and Ser17, and covalently linked the carboxylic acid group of Asp14 (after making a E14D mutation) to the primary amine of Lys10 creating an amide linkage **(Fig. 4.3)**. We then performed an MD simulation followed by an MMGBSA calculation. Similar stapling strategies were applied to the other shortlisted peptides. Almost all the stapled peptides exhibited greater

conformational stability and stayed longer in bound state. In almost all cases, the MMGBSA binding free energy was more favorable for the stapled version than for the unstapled version. **(see Appendix D)**.

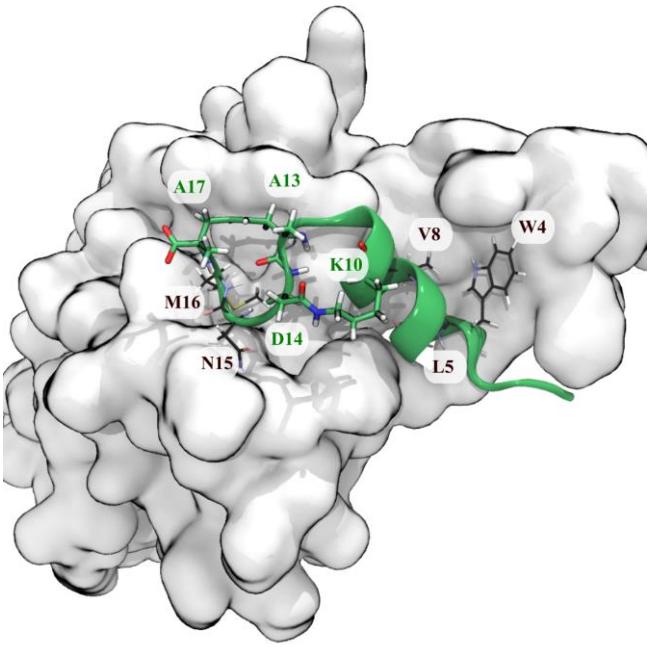

Figure 4.3 Favorable conformation of the optimized peptide on the spike protein RBD.
The pose of an optimized peptide (green cartoon) having the most favorable GBSA free energy on the spike protein RBD (gray surface). Some residues whose sidechains did not make contact with the protein were stapled (green labels), so as to improve the conformational stability of the peptide while not interfering with its interactions with residues of the spike protein RBD (black labels).

4.7.14 Conformational flexibility of the receptor-bound peptide in the stapled and unstapled form

We have analyzed the MD simulation trajectories for the stapled and unstapled versions of peptides in the presence of spike protein RBD and measured the deviation in atomic positions between the stapled and unstapled peptides relative to the structure of the unstapled peptide with the most favorable $\Delta G_{\text{binding}}^{\text{GBSA}}$. The stapled peptides in the bound state exhibit higher rigidity and lower mobility compared to their unstapled versions. A comparison of RMSD for the stapled peptide pep39 and its corresponding unstapled version is shown in **Fig. 4.4**.

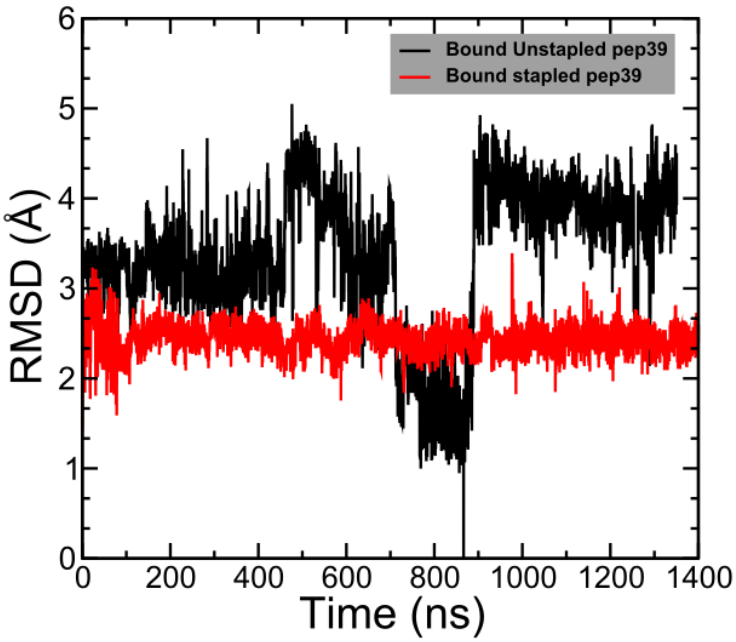

Figure 4.4 Conformational stability of the designed peptide.

Deviation in the atomic positions of the stapled and unstapled versions of the peptide pep39 on the surface of spike protein RBD. The RMSD is taken with respect to the stapled structure with the most favorable GBSA binding free energy ($t \approx 870$ ns).

4.7.15 Absolute binding free energy calculation for the designed stapled peptide

The MMGBSA calculation gives an estimate for the binding affinity of the designed stapled peptide for the spike protein RBD. However, we did not include the contribution from the conformational entropy and uses a continuum approximation for the hydration free energy, rather than explicit water molecules.[51] Hence, to validate the binding affinity of the stapled peptide, an absolute binding free energy calculation by the geometric route was carried out using the BFEE plugin of VMD[84]. The BFEE method explicitly includes the effects individual water molecules and is more rigorously derived from statistical mechanics than the MMGBSA method. It has also been demonstrated to yield better agreement with experiment.[52] The BFEE method includes entropy contributions, including conformational entropy of the peptide and configurational entropy of hydrating water molecules; however, for this reason, it requires extensive sampling of these

degrees of freedom to yield correct results. This makes BFEE calculations computationally expensive. Based on their MMGBSA score, we shortlisted 4 stapled peptides out of 15 and ran BFEE calculations. An eABF calculation for each sub-process was run, and the overall simulation length was more than 3.0 μs for all stapled peptides. The results obtained from the potential of mean force (PMF) calculations are given in **see Appendix D**. Here, we will discuss the best performing stapled peptide that is (pep39) in further detail.

The contribution of each sub-process for absolute binding free energy is given in **Table–4.1**. The unbound stapled peptide does not always maintain the α-helical structure when unbound, while this structure is stable when the bound to the spike protein RBD. These results are corroborated by the RMSD analysis (performed for the conventional MD simulation) of the configuration of the stapled peptide in bound and unbound form **(Fig. 4.10)**. An analysis of the PMF values computed from independent portions of the simulations for sub-process 2 to 7 show that they are well converged. Although the eABF calculations for sub-processes RMSD bound and RSMD unbound were ran for 1 and 3 μs respectively, PMF values did not appear to converge well **(Figure–4.12)**. This suggests that the estimation of conformational free energy of flexible macromolecules like peptides and proteins by computation methods remains a challenging task. This demands further improvement in the computational techniques to measure the contribution of the RMSD of bound and unbound peptide ligand in the free energy calculations.[53–55]

Table 4.2 The contribution of sub-processes in the absolute binding free energy calculation.

Sub-process	Free-Energy Term	PMF (kcal/mol)	Simulation time (ns)
1	$\Delta G_{\text{conform}}$	−17.94	1100
2	ΔG_{θ}	−0.25	18
3	ΔG_{φ}	−0.22	18
4	ΔG_{ψ}	−0.21	17
5	ΔG_{θ}	−0.09	19

6	ΔG_{ϕ}	−0.11	16
7	$-k_{\mathrm{B}}T \ln (S^{*}I^{*}C^{0})$	−29.03	1181
8	$\Delta \mathrm{G}^{\mathrm{unbound}}_{\mathrm{conform}}$	+14.75	2963
9	$\Delta \mathrm{G}^{\mathrm{unbound}}_{\Theta\phi\Psi}$	+6.80	–
	Total $\Delta \mathrm{G}^{0}_{\mathrm{binding}}$	−26.32	5332

4.7.16 Interaction of the designed stapled peptide pep39 with the spike protein RBD

We designed peptides to target key residues of the binding site on the RBD like Lys417, Leu455, Phe490, Gln493, and Tyr505 on RBD.[56] An analysis of unbiased MD simulation trajectories for pep39 bound to the spike protein RBD shows that the pep39 occupies the region where the host protein ACE2 binds. **Fig. 4.11A** shows the distance of pep39 from the binding site (consisting of residues Trp353, Arg403, Lys417, Asn439, Val445, Leu455 Phe456, Gln493, Asn501, and Tyr505) is less than 3 Å for almost 1700 ns of MD simulation. Some important interactions between pep39 and RBD are shown in the 3-dimensional view in **Fig. 4.11B**. These include a hydrogen bond between the indole NH hydrogen of Trp9 and the sidechain amide oxygen of Gln493 of the spike protein RBD (**Fig. 4.5A**). A π–π interaction between the aromatic rings of the Phe12 of pep39 and Phe490 of the spike protein (**Fig. 4.5B**) is also present for more than 500 ns out of 2000 ns of the simulation. The hydrophobic interaction between the aliphatic side chain of Val8 and aromatic ring of Phe490 of the spike protein was maintained for the almost entire simulation (**Fig. 4.5C**). Hence Phe490, one of the key residues in the binding site, was engaged with pep39 throughout the simulation. Overall, pep39 blocks the spike protein RBD surface where ACE2 protein binds.

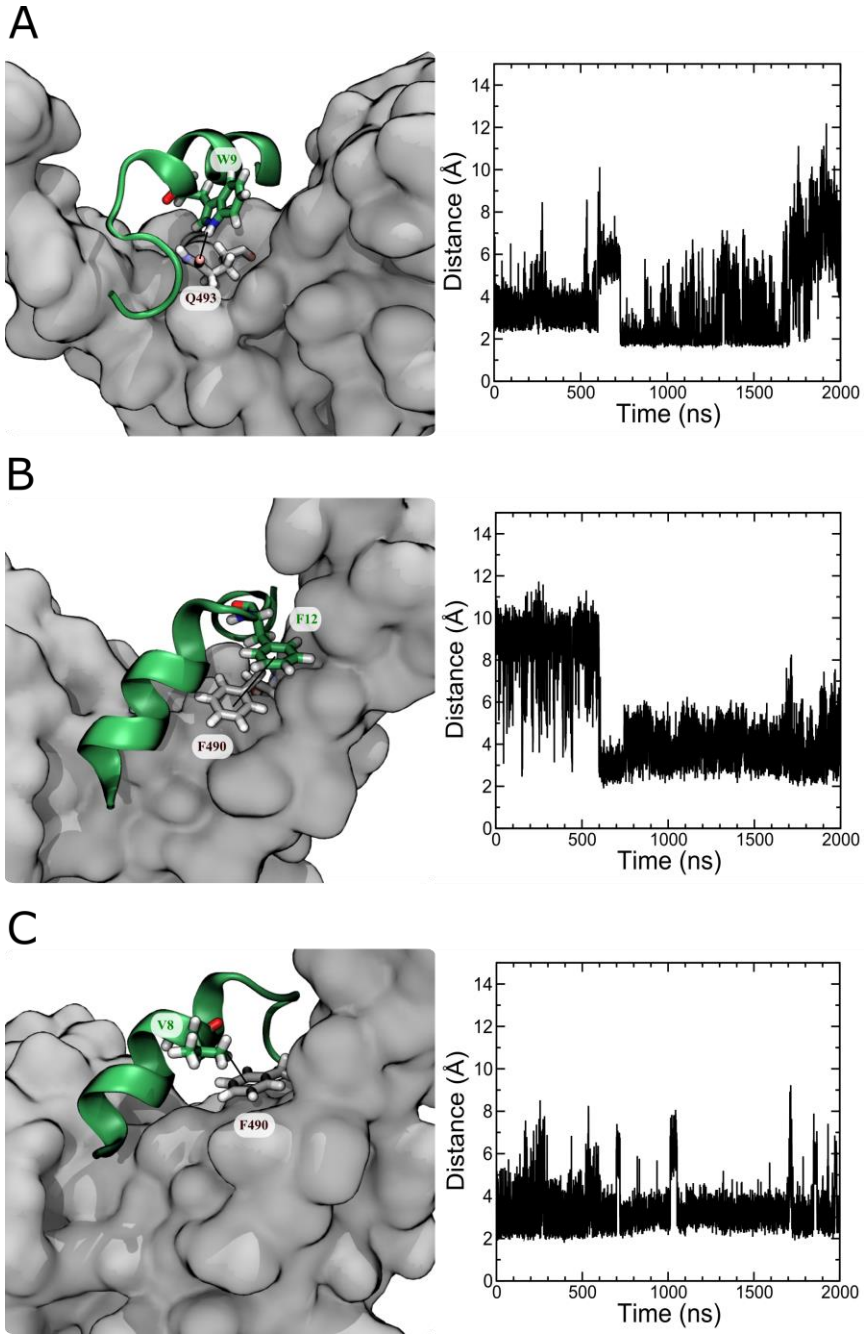

Figure 4.5 An atomic interaction between designed stapled peptide and spike protein RBD.

(A) H-bond involving the indole nitrogen of Trp9 and the sidechain amide oxygen of Gln493. **(B)** A π–π stacking interaction between Phe12 of pep39 and Phe490 of spike protein RBD. **(C)** The hydrophobic interaction between the aliphatic sidechain of Val8 and the aromatic ring of Phe490 of the spike protein RBD.

4.7.17 Experimental confirmation of the binding activity of the designed stapled peptide pep39

4.7.17.1 Micro-cantilever-based method

On the basis of the deflection of the micro-cantilever, we were able to confirm the binding of pep39 on the surface of the spike protein RBD. The method involves the real-time monitoring of microcantilever bending (deflection) resulting from surface stress induced by a specific protein-protein interaction on the cantilever surface.[57,58] The maximum deflection (signal) achieved was 33.27 nm for a 1 μg/ml target concentration, whereas the minimum signal of 8.46 nm was observed for the lowest concentration of the target analyte (1 ng/ml). For the positive control (anti-spike protein antibody), a target concentration of 1 μg/ml was used, and we observed a maximum deflection of 81.88 nm. For the negative control, as expected, we measured the lowest deflection, 4.71 nm, which is not significant and can be considered as noise (**Figure–4.6A**).

4.7.17.2 Bio-layer interferometry method

BLI analyses the interference pattern of white light reflected from a biosensor tip with conjugated with protein molecules and an internal reference surface.[59,60] The binding kinetics between the protein and peptide molecules are measured in real-time by detecting the shift in the interference pattern of the white light caused by binding and unbinding events at the surface of the protein immobilized biosensor tip. Human recombinant ACE2 was used as a positive control to validate the biosensor tip. In agreement with the previous reports, our BLI analysis has demonstrated a dissociation constant of 9 nM for binding between the spike protein RBD and ACE2 .[61,62] Concentrations of 120, 48, 24, 4.8, and 0.48 μM for pep39 were used to study binding activity with biotinylated recombinant SARS-CoV-2 spike protein RBD immobilized on high precision strep-

tavidin (SAX) biosensor tips (**Figure–4.6B**). The results were globally fit to determine the dissociation constant. The binding assay demonstrated that the pep39 binds to spike protein RBD with K_D value of 570 ± 50 nM using BLItz Pro version1.1 software.

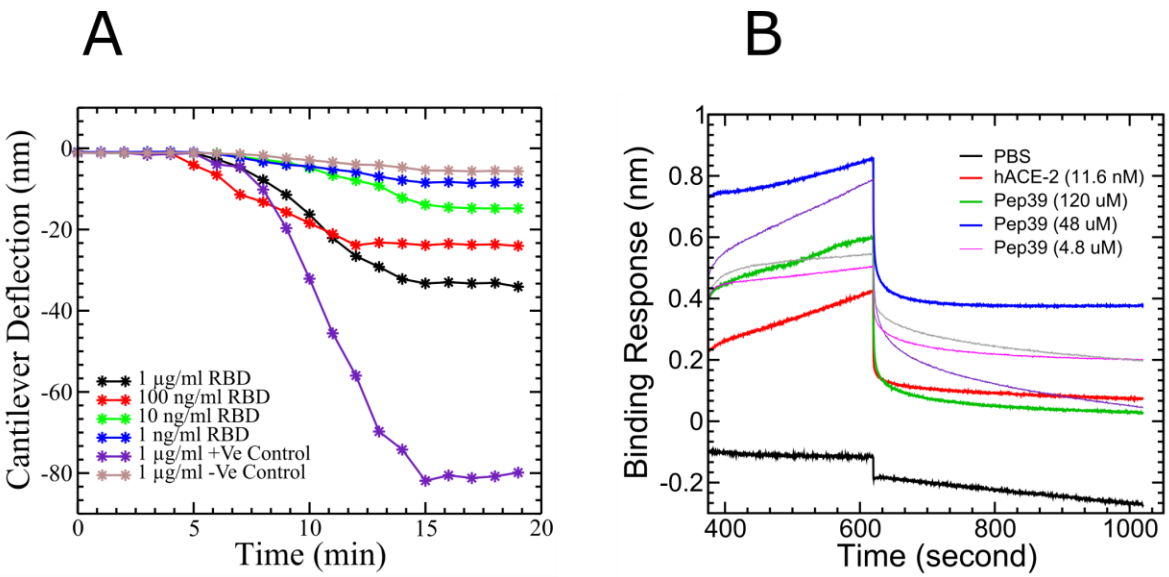

Figure 4.6 Binding kinetics study of the designed stapled peptide pep39 and the spike protein RBD.

(A) The deflection curve for the peptide pep39 immobilized on the cantilever in response to different concentrations of SARS-CoV-2 spike protein. Anti-S1 antibody served as a positive control and H1N1 protein, as a negative control. **(B)** The association and dissociation between immobilized spike protein RBD and different concentrations of pep39. Recombinant human ACE2 protein used as a positive control and PBS buffer as a reference blank.

4.7.18 Effect of pep39 on variants of the spike protein

Like other RNA viruses, new variants of SARS-CoV-2 are emerging due to mutations. Variants with the mutations in the spike protein are major health concerns and alarming because they are more transmissible and capable of evading the immune response.[63] Some variants like B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), and B.1.617.2 (Delta), which were first found in the United Kingdom, South Africa, Brazil, and India, respectively, were designated as variants of concern (VOC) by WHO.[64–66] To evaluate the effect of pep39 on the VOCs, we performed almost

2 μs long MD simulations and MMGBSA calculation. Our designed stapled peptide pep39 stays bound with all VOCs for the entire length of the simulation with significant binding affinity (**Table–4.2**). Among all VOCs, the Delta variant caused a major wave of COVID-19 pandemic.[67–71] *Rui Wang et al.* demonstrated that the Delta variant would be a vaccine breakthrough variant due to its ability to disrupt the antibody-RBD complexes using a computational study.[71] We performed BLI analysis to confirm the binding activity of pep39 with spike protein Delta variant. Because the Delta variant contains mutations in the N-terminal domain (NTD) and the receptor-binding domain (RBD), we immobilized the complete spike protein Delta variant B.1.617.2 which has both subunits. [72] The BLI assay reported that the ACE2 protein binds to spike protein delta variant B.1.617.2 with K_D 120 pM. As expected, the dissociation constant for the interaction of ACE2 and Delta variant is more favorable for binding than the wild-type spike protein. Our results with the positive control agreed well with previous studies.[73] To evaluate binding activity of pep39 with the spike protein Delta variant, we used a broad range of pep39 concentrations (375, 187, 93, 46, 25, 5 and 500 μM), and the data were globally fit to determine the dissociation constant. The binding assay demonstrated that the pep39 binds to spike protein Delta variant with an affinity of 4.1 ± 1.4 μM using BLItz Pro version 1.1 (**Fig. 4.7**).

Table 4.3 MD simulation of pep39 with spike protein RBD variants and resultant MMGBSA binding free energy.

Sl. No	Spike protein RBD Variants	MD Simulation Run time (ns)	Rough Estimate of Binding Energy by MMGBSA (kcal/mol)
1	Alpha	1866	-41.798 ± 0.076
2	Beta	1845	-37.215 ± 0.095
3	Gamma	1759	-37.700 ± 0.085
4	Delta	2000	-41.295 ± 0.128

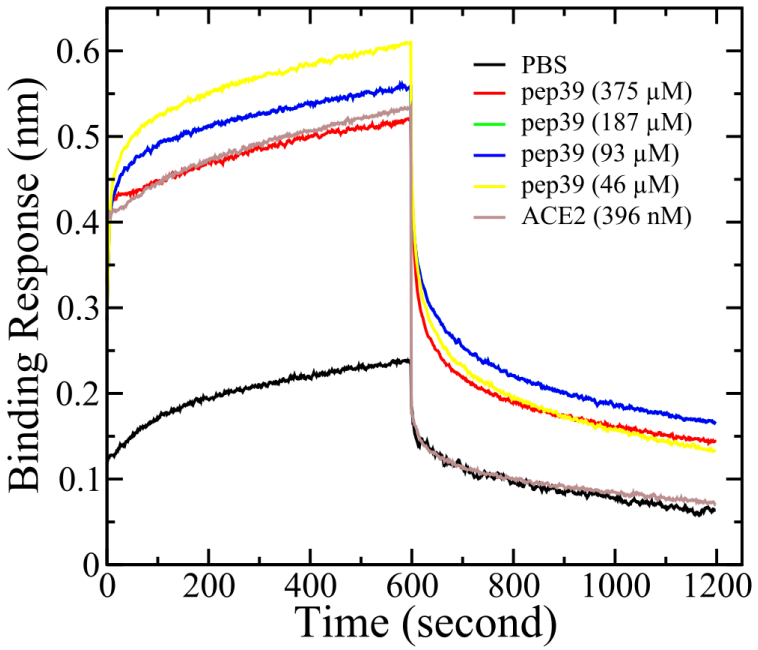

Figure 4.7 Binding kinetics study of the designed stapled peptide pep39 and Delta variant spike protein.

The association and dissociation between immobilized delta variant spike protein and various concentrations of the designed stapled peptide pep39. Recombinant human ACE–2 protein used as a positive control and PBS buffer as a reference blank.

In previous studies, computational models of some anti-microbial peptides, cell-penetrating peptides conjugated with FDA-approved drugs, *de novo* design of peptides, and lipopeptides were described to prevent the entry of SARS-CoV-2, but experimental validation remains to be reported.[74–79] *Danielle et al.* have reported binding activity of stapled peptides based on the ACE2 α1 helix, but designed peptides didn't prevent virus internalization.[21] *Curreli et al.* also used a similar strategy to design double stapled peptides; their peptides showed activity in the range of IC50 1.9 – 4.1 μM, but effects on the spike protein variant have not been reported yet.[80] A biochemically modified ACE2-targeting peptide derived from the spike protein showed significant inhibitory activity against ACE2 and spike protein association.[81] Protein decoys, synthetic antibodies, and nanobodies have been explored to target the epitopes on the spike protein and achieved desirable binding affinity and neutralization effects.[82,83] Recent reports claims picomolar binding

affinity of a synthetic antibody with the spike protein and its variants. Prophylactic and therapeutic effects were reported in the laboratory animals.[73]

These reports suggest, to inhibit the interaction between the spike protein and ACE2, a macromolecule is required as a drug. However, the production cost is the major limitation for antibodies and other macromolecular therapeutics and affordable options need to be explored. Our designed peptide exhibited a weaker binding affinity (500 nM for original spike protein RBD and 4.1 μM for Delta variant) compared with some reported macromolecules, but still, this affinity is considerable. There is scope for further improvement in the peptide design algorithm.

4.8 Conclusion

Due to rapid mutation of SARS-CoV-2 and limitations of vaccination for immunocompetent patients, there is an urgent need for anti-SARS-CoV-2 therapeutics. The availability of an affordable drug is required to treat patient that belong to the lower income categories. The presented computational calculations and binding experiments confirm that our designed stapled peptide pep39 binds to the original spike protein and its Delta variant with considerable binding affinity and could be considered as a potential candidate for the COVID-19 therapeutics to end this pandemic.

4.9 Acknowledgements

This research was supported by US National Science Foundation under grant no. DMR-1945589, Kansas State University (KSU) College of Veterinary Medicine Dean's funds [2015CVM-SMILE], Johnson Cancer Research Center 2018 JCRC-IRA, National Cancer Institute [1 R15 CA219919-01]. A majority of the computing for this project was performed on the Beocat

Research Cluster at Kansas State University, which is funded in part by NSF grants CHE-1726332, CNS-1006860, EPS-1006860, and EPS-0919443. We also thank Dr. Michal Zolkiewksi, Dr. Govind Vediyappan of Kansas State University and the Kansas Idea Network of Biomedical Research Excellence (K-INBRE) CORE Facility support award (grant number P20 GM103418) for providing access to BLItz facility.

4.10 References

1. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

2. COVID-19 Map. *Johns Hopkins Coronavirus Resource Center* https://coronavirus.jhu.edu/map.html.

3. Ljungman, P. Vaccination of immunocompromised patients. *Clinical Microbiology and Infection* **18**, 93–99 (2012).

4. Babiker, A., Marvil, C. E., Waggoner, J. J., Collins, M. H. & Piantadosi, A. The Importance and Challenges of Identifying SARS-CoV-2 Reinfections. *Journal of Clinical Microbiology* **59**, e02769-20.

5. Tillett, R. L. *et al.* Genomic evidence for reinfection with SARS-CoV-2: a case study. *Lancet Infect Dis* **21**, 52–58 (2021).

6. Crawford, K. H. D. *et al.* Dynamics of Neutralizing Antibody Titers in the Months After Severe Acute Respiratory Syndrome Coronavirus 2 Infection. *The Journal of Infectious Diseases* **223**, 197–205 (2021).

7. Long, Q.-X. *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med* **26**, 1200–1204 (2020).

8. Convergent antibody responses to SARS-CoV-2 in convalescent individuals | Nature. https://www.nature.com/articles/s41586-020-2456-9.

9. Kalil, A. C. Treating COVID-19—Off-Label Drug Use, Compassionate Use, and Randomized Clinical Trials During Pandemics. *JAMA* **323**, 1897 (2020).

10. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).

11. Brann, D. H. *et al.* Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia. *Science Advances* **6**, eabc5801 (2020).

12. Zhu, Y., Yu, D., Yan, H., Chong, H. & He, Y. Design of Potent Membrane Fusion Inhibitors against SARS-CoV-2, an Emerging Coronavirus with High Fusogenic Activity. *Journal of Virology* **94**, (2020).

13. Wang, Q. *et al.* Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **181**, 894-904.e9 (2020).

14. Carino, A. *et al.* Hijacking SARS-CoV-2/ACE2 Receptor Interaction by Natural and Semi-synthetic Steroidal Agents Acting on Functional Pockets on the Receptor Binding Domain. *Frontiers in Chemistry* **8**, 846 (2020).

15. Ahidjo, B. A., Loe, M. W. C., Ng, Y. L., Mok, C. K. & Chu, J. J. H. Current Perspective of Antiviral Strategies against COVID-19. *ACS Infect. Dis.* **6**, 1624–1634 (2020).

16. Josephson, K., Ricardo, A. & Szostak, J. W. mRNA display: from basic principles to macrocycle drug discovery. *Drug Discovery Today* **19**, 388–399 (2014).

17. Wu, C. *et al.* Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm Sin B* **10**, 766–788 (2020).

18. Zhang, G., Pomplun, S., Loftis, A. R., Loas, A. & Pentelute, B. L. *The first-in-class peptide binder to the SARS-CoV-2 spike protein*. 2020.03.19.999318 https://www.biorxiv.org/content/10.1101/2020.03.19.999318v1 (2020) doi:10.1101/2020.03.19.999318.

19. Karoyan, P. *et al. Human ACE2 peptide mimics block SARS-CoV-2 Pulmonary Cells Infection*. 2020.08.24.264077 https://www.biorxiv.org/content/10.1101/2020.08.24.264077v2 (2020) doi:10.1101/2020.08.24.264077.

20. Tan, Y. S., Lane, D. P. & Verma, C. S. Stapled peptide design: principles and roles of computation. *Drug Discovery Today* **21**, 1642–1653 (2016).

21. Morgan, D. C. *et al.* Stapled ACE2 peptidomimetics designed to target the SARS-CoV-2 spike protein do not prevent virus internalization. *Peptide Science* **113**, e24217 (2021).

22. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).

23. Raveh, B., London, N., Zimmerman, L. & Schueler-Furman, O. Rosetta FlexPepDock ab-initio: Simultaneous Folding, Docking and Refinement of Peptides onto Their Receptors. *PLOS ONE* **6**, e18934 (2011).

24. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).

25. Alford, R. F. *et al.* The Rosetta all-atom energy function for macromolecular modeling and design. *J Chem Theory Comput* **13**, 3031–3048 (2017).

26. Leman, J. K. *et al.* Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods* **17**, 665–680 (2020).

27. Huang, J. *et al.* CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **14**, 71–73 (2017).

28. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859–1865 (2008).

29. ZDOCK: An initial-stage protein-docking algorithm - Chen - 2003 - Proteins: Structure, Function, and Bioinformatics - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/prot.10389.

30. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

31. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802 (2005).

32. Davidchack, R. L., Handel, R. & Tretyakov, M. V. Langevin thermostat for rigid body dynamics. *J. Chem. Phys.* **130**, 234101 (2009).

33. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).

34. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

35. Daintith, J. D. Lennard-Jones potential. in *A Dictionary of Chemistry* (ed. Daintith, J.) (Oxford University Press, 2008).

36. Fiorin, G., Klein, M. L. & Hénin, J. Using collective variables to drive molecular dynamics simulations. *Molecular Physics* **111**, 3345–3362 (2013).

37. Azhagiya Singam, E. R., Rajapandian, V. & Subramanian, V. Molecular dynamics simulation study on the interaction of collagen-like peptides with gelatinase-A (MMP-2). *Biopolymers* **101**, 779–794 (2014).

38. Vergara-Jaque, A., Comer, J., Monsalve, L., González-Nilo, F. D. & Sandoval, C. Computationally Efficient Methodology for Atomic-Level Characterization of Dendrimer–Drug Complexes: A Comparison of Amine- and Acetyl-Terminated PAMAM. *J. Phys. Chem. B* **117**, 6801–6813 (2013).

39. Croitoru, A. *et al.* Additive CHARMM36 Force Field for Nonstandard Amino Acids. *J. Chem. Theory Comput.* **17**, 3554–3570 (2021).

40. Fu, H. *et al.* BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **58**, 556–560 (2018).

41. Extended Adaptive Biasing Force Algorithm. An On-the-Fly Implementation for Accurate Free-Energy Calculations | Journal of Chemical Theory and Computation. https://pubs.acs.org/doi/10.1021/acs.jctc.6b00447.

42. Lesage, A., Lelièvre, T., Stoltz, G. & Hénin, J. Smoothed Biasing Forces Yield Unbiased Free Energies with the Extended-System Adaptive Biasing Force Method. *J. Phys. Chem. B* **121**, 3676–3685 (2017).

43. Agarwal, D. K. *et al.* Highly Sensitive and Ultra-Rapid Antigen-based Detection of SARS-CoV-2 using Nanomechanical Sensor Platform. *Biosens Bioelectron* 113647 (2021) doi:10.1016/j.bios.2021.113647.

44. Zhu, X. & Mitchell, J. C. KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins* **79**, 2671–2683 (2011).

45. Chakraborty, S. Evolutionary and structural analysis elucidates mutations on SARS-CoV2 spike protein with altered human ACE2 binding affinity. *Biochemical and Biophysical Research Communications* **538**, 97–103 (2021).

46. Crooks, G. E. WebLogo: A Sequence Logo Generator. *Genome Research* **14**, 1188–1190 (2004).

47. Barros, E. P. *et al.* Improving the Efficiency of Ligand-Binding Protein Design with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **15**, 5703–5715 (2019).

48. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).

49. Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J Mol Biol* **332**, 449–460 (2003).

50. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein Structure Prediction Using Rosetta. in *Methods in Enzymology* vol. 383 66–93 (Academic Press, 2004).

51. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* **10**, 449–461 (2015).

52. Fu, H., Cai, W., Hénin, J., Roux, B. & Chipot, C. New Coarse Variables for the Accurate Determination of Standard Binding Free Energies. *J. Chem. Theory Comput.* **13**, 5173–5178 (2017).

53. Ekberg, V. & Ryde, U. On the Use of Interaction Entropy and Related Methods to Estimate Binding Entropies. *J. Chem. Theory Comput.* **17**, 5379–5391 (2021).

54. de Ruiter, A. & Oostenbrink, C. Advances in the calculation of binding free energies. *Current Opinion in Structural Biology* **61**, 207–212 (2020).

55. Gumbart, J. C., Roux, B. & Chipot, C. Efficient Determination of Protein–Protein Standard Binding Free Energies from First Principles. *J. Chem. Theory Comput.* **9**, 3789–3798 (2013).

56. Pavlova, A. *et al.* Machine Learning Reveals the Critical Interactions for SARS-CoV-2 Spike Protein Binding to ACE2. *J. Phys. Chem. Lett.* **12**, 5494–5502 (2021).

57. Translating Biomolecular Recognition into Nanomechanics. https://www.science.org/doi/abs/10.1126/science.288.5464.316.

58. Tamayo, J., Kosaka, P. M., Ruz, J. J., Paulo, Á. S. & Calleja, M. Biosensors based on nanomechanical systems. *Chem. Soc. Rev.* **42**, 1287–1311 (2013).

59. Concepcion, J. *et al.* Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. *Comb Chem High Throughput Screen* **12**, 791–800 (2009).

60. Shah, N. B. & Duncan, T. M. Bio-layer Interferometry for Measuring Kinetics of Protein-protein Interactions and Allosteric Ligand Effects. *JoVE (Journal of Visualized Experiments)* e51383 (2014) doi:10.3791/51383.

61. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).

62. Zhou, T. *et al.* Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* **28**, 867-879.e5 (2020).

63. McGill, A. R. *et al.* SARS-CoV-2 Immuno-Pathogenesis and Potential for Diverse Vaccines and Therapies: Opportunities and Challenges. *Infect Dis Rep* **13**, 102–125 (2021).

64. Wang, R. *et al.* Analysis of SARS-CoV-2 variant mutations reveals neutralization escape mechanisms and the ability to use ACE2 receptors from additional species. *Immunity* **54**, 1611-1621.e5 (2021).

65. Duerr, R., Crosse, K. M., Valero-Jimenez, A. M. & Dittmann, M. SARS-CoV-2 Portrayed against HIV: Contrary Viral Strategies in Similar Disguise. *Microorganisms* **9**, 1389 (2021).

66. Weekly epidemiological update - 2 February 2021. https://www.who.int/publications/m/item/weekly-epidemiological-update---2-february-2021.

67. Weekly epidemiological update on COVID-19 - 10 August 2021. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---10-august-2021.

68. Li, B. *et al. Viral infection and transmission in a large well-traced outbreak caused by the Delta SARS-CoV-2 variant*. 2021.07.07.21260122 https://www.medrxiv.org/content/10.1101/2021.07.07.21260122v1 (2021) doi:10.1101/2021.07.07.21260122.

69. Liu, Y. & Rocklöv, J. The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. *Journal of Travel Medicine* (2021) doi:10.1093/jtm/taab124.

70. Liu, Y. *et al. Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant*. 2021.08.12.456173 https://www.biorxiv.org/content/10.1101/2021.08.12.456173v1 (2021) doi:10.1101/2021.08.12.456173.

71. Wang, R., Chen, J., Hozumi, Y., Yin, C. & Wei, G.-W. Emerging vaccine-breakthrough SARS-CoV-2 variants. *ArXiv* arXiv:2109.04509v1 (2021).

72. Planas, D. *et al.* Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276–280 (2021).

73. Zhang, H. *et al. An ultrapotent neutralizing bispecific antibody with broad spectrum against SARS-CoV-2 variants*. 2021.08.10.455627 https://www.biorxiv.org/content/10.1101/2021.08.10.455627v1 (2021) doi:10.1101/2021.08.10.455627.

74. A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. https://www.science.org/doi/10.1126/sciadv.aav4580.

75. Ansari, M. A. *et al.* TAT-peptide conjugated repurposing drug against SARS-CoV-2 main protease (3CLpro): Potential therapeutic intervention to combat COVID-19. *Arabian Journal of Chemistry* **13**, 8069–8079 (2020).

76. Baig, M. S., Alagumuthu, M., Rajpoot, S. & Saqib, U. Identification of a Potential Peptide Inhibitor of SARS-CoV-2 Targeting its Entry into the Host Cells. *Drugs R D* **20**, 161–169 (2020).

77. Sekar, P. C. & Rajasekaran, R. Could Dermaseptin Analogue be a Competitive Inhibitor for ACE2 Towards Binding with Viral Spike Protein Causing COVID19?: Computational Investigation. *Int J Pept Res Ther* 1–14 (2021) doi:10.1007/s10989-020-10149-w.

78. Ling, R. *et al.* In silico design of antiviral peptides targeting the spike protein of SARS-CoV-2. *Peptides* **130**, 170328 (2020).

79. Huang, X., Pearce, R. & Zhang, Y. De novo design of protein peptides to block association of the SARS-CoV-2 spike protein with human ACE2. *Aging (Albany NY)* **12**, 11263–11276 (2020).

80. Curreli, F. *et al.* Stapled Peptides Based on Human Angiotensin-Converting Enzyme 2 (ACE2) Potently Inhibit SARS-CoV-2 Infection In Vitro. *mBio* **11**, e02451-20 (2020).

81. Gottschalk, G. *et al. ACIS, A Novel KepTide™, Binds to ACE-2 Receptor and Inhibits the Infection of SARS-CoV2 Virus in vitro in Primate Kidney Cells: Therapeutic Implications for COVID-19.* 2020.10.13.337584 https://www.biorxiv.org/content/10.1101/2020.10.13.337584v1 (2020) doi:10.1101/2020.10.13.337584.

82. Linsky, T. W. *et al.* De novo design of potent and resilient hACE2 decoys to neutralize SARS-CoV-2. *Science* **370**, 1208–1214 (2020).

83. Schoof, M. *et al.* An ultrapotent synthetic nanobody neutralizes SARS-CoV-2 by stabilizing inactive Spike. *Science* **370**, 1473–1479 (2020).

84. Gumbart, J. C., Roux, B. & Chipot, C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? J. Chem. Theory Comput. 9, 794–802 (2013).

4.11 Additional information

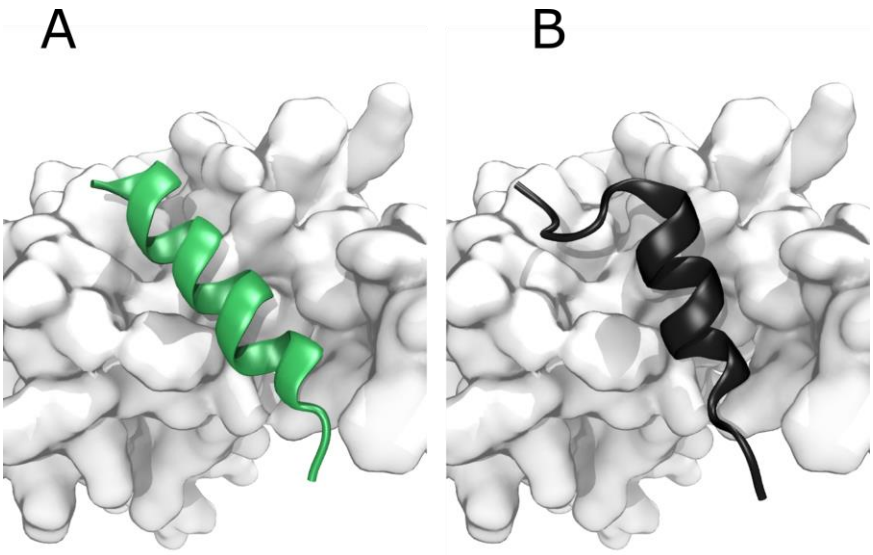

Figure 4.8 Starting configuration and the most favorable pose of the optimized peptide.

(A) The starting configuration of the best performing peptide with optimized sequence. **(B)** The same peptide in its the most binding favorable conformation (according to the GBSA free energy function).

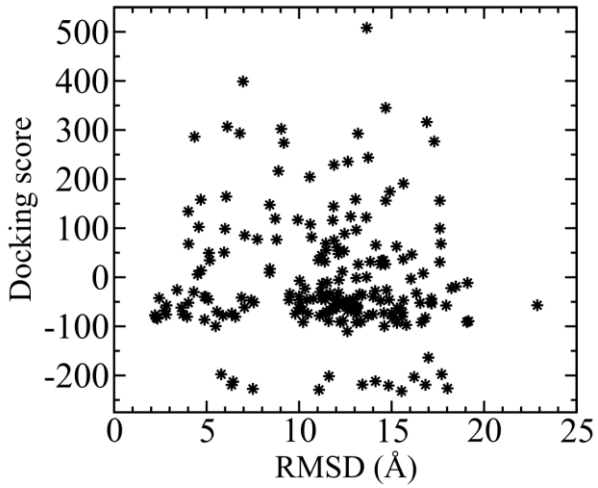

Figure 4.9 The docking score versus RMSD plot for the poses of template peptide on the spike protein RBD surface.

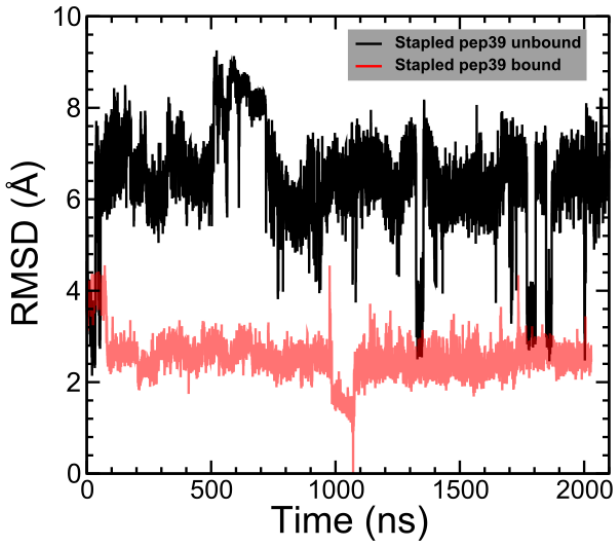

Figure 4.10 RMSD of the stapled peptide pep39 in bound and unbound form from a reference structure with the lowest GBSA free energy.

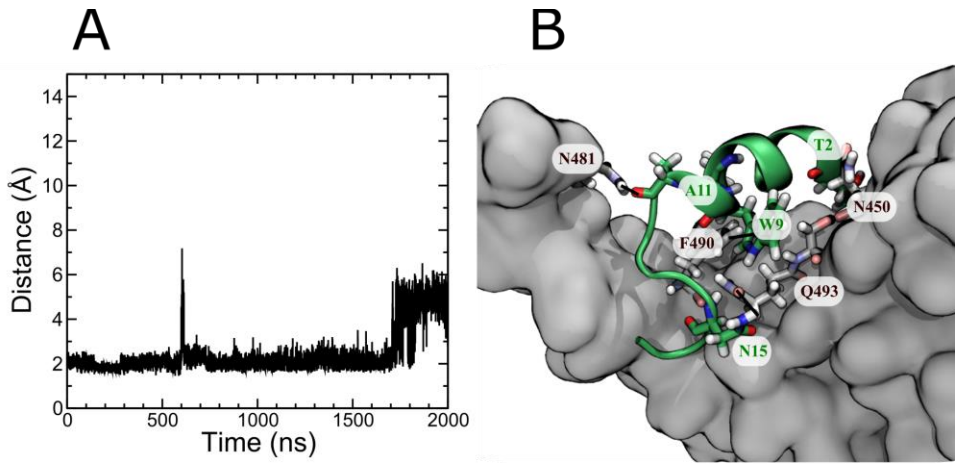

Figure 4.11 Designed cyclic peptide pep39, block ACE–2 binding site of spike protein RBD.

(A) Distance of peptide from the ACE-2 binding site on the spike protein RBD. **(B)** An interaction between designed peptide and spike protein RBD in 3D view.

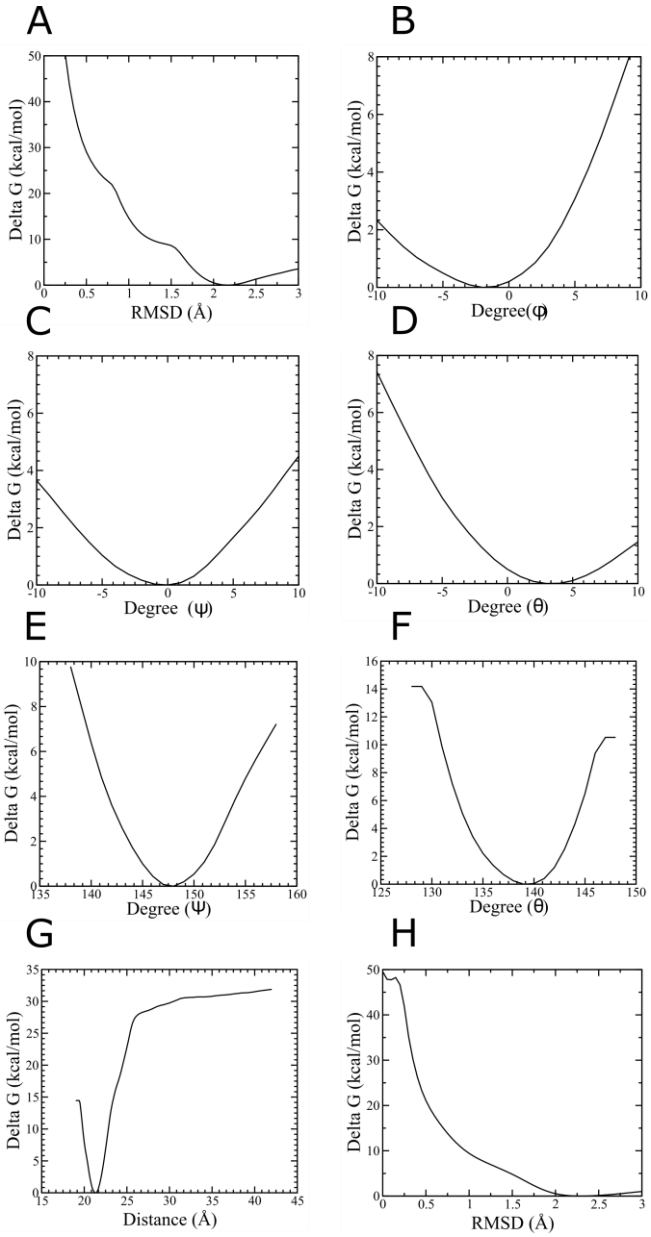

Figure 4.12 Estimated PMF values for each sub-process of absolute biding free energy calculation.

Sub-process is **(A)** RMSD of peptide in bound form, Movement of peptide in **(B)** Euler angle Θ, **(C)** Euler angle $\boldsymbol{\varphi}$, **(D)** Euler angle $\boldsymbol{\psi}$, **(E)** Polar angle $\boldsymbol{\theta}$, **(F)** Polar angle ϕ with respect to protein, **(G)** Separation of peptide from protein and **(H)** RMSD of peptide in unbound form.

Appendix A - **Manipulation in the sequence of the template peptide targeting CTLA4 protein.**

Sr. No.	Sequence	Optimization Score		Duration of MD Simulation (ns)	Free Energy by MMGBSA (kcal/mol)	Free Energy by BFEE (kcal/mol)
		Pre design	Post design			
1	Cyc(ECRYEPRPEGNILVSYS)	186.49	170.67	660.2	-22.67 ± 0.18	+2.61± 1.42
2	Cyc(SIVTKLTPTGWVAASYS)	185.57	175.15	899.4	-26.40 ± 0.14	+3.51 ± 13.24
3	Cyc(KVEFKRTPSGTITVSME)	176.95	165.86	12.8	-10.43 ± 0.64	N/A
4	Cyc(KVVYEPKPEGNIVVEYE)	211.67	194.39	48.4	-12.42 ± 0.34	N/A
5	Cyc(SAKFEPRPEGNIVVSYG)	202.6	200.6	134	-22.67 ± 0.24	N/A
6	Cyc(EARYQPRPDGNVLVSYG)	204.99	206.02	254.6	-14.20 ± 0.19	N/A
7	Cyc(SAKWNPKPEGAELIEEG)	226.5	222.82	16	-7.69 ± 0.67	N/A
8	Cyc(SAEFIPTPDGNLLKSSG)	216.4	214.02	13.8	-8.79 ± 0.72	N/A
9	Cyc(SIVVVLTPTGWVAASYS)	155.77	157.63	278	-26.94 ± 0.29	N/A
10	Cyc(EIITKLTPTGWVAASYS)	156.02	157.61	86	-19.29 ± 0.35	N/A
11	Cyc(SIEMELTPTGWVNKSSS)	159.12	157.96	44.2	-11.30 ± 0.33	N/A
12	Cyc(SIITVLTPTGWVAAEFS)	154.32	155.7	1110.6	-32.76 ± 0.14	-10.21 ± 2.41
13	Cyc(DIITILTPTGYVAAAYS)	152.66	154.02	395.6	-19.01 ± 0.18	N/A
14	Cyc(SIITVLTPTGWVAAYYS)	153.1	155.36	1463.2	-23.31 ± 0.09	+5.11 ± 21.31
15	Cyc(SIQCVLTPTGWVAARYS)	153.69	155.15	42.4	-23.59 ± 0.71	N/A
16	Cyc(EIDTVLTPTGWVAKRYS)	152	153.85	2000	-22.30 ± 0.11	-12.58 ± 3.76
17	Cyc(SIRMELTPTGWVAAEYE)	165.81	165.25	119.4	-18.47 ± 0.37	N/A

Manipulation in the template peptide sequence cyc(CIIHHKKPTGMIRIHQC) at its selected poses with post and pre optimization score, binding free energy by MM-GBSA and BFEE, a rigorous method.

Appendix B - **Cyclization of peptide by disulfide bond.**

Sr. No.	Sequence	Optimization Score		Duration of MD Simulation (ns)	Free Energy by MMGBSA (kcal/mol)	Free Energy by BFEE (kcal/mol)
		Pre design	Post design			
1	Cyc(CIVHENRPEGLVRVHLC)	180.51	167.9	23.6	-5.16 ± 0.30	N/A
2	Cyc(CVVFEPKPEGTEKVHEC)	218.12	211.6	74.2	-19.60 ± 0.53	N/A
3	Cyc(CIIWEDQPNGKVCVHSC)	174.096	165.25	89.6	-10.99 ± 0.28	N/A
4	Cyc(CVIEQYRPEGVVLIYEC)	182.95	172.16	578.8	-18.17 ± 0.14	+1.73 ± 4.84
5	Cyc(CTVAIPLPDGKICVKSC)	167.7	160.96	59.6	-12.97 ± 0.32	N/A
6	Cyc(CEVRKYSESGVIPIDSC)	177.21	176.28	106	-8.11 ± 0.27	N/A
7	Cyc(CLITAASESGVYTIYEC)	172.99	169.36	4.6	-3.95 ± 1.03	N/A
8	Cyc(CVLQQNAPEGIITIEEC)	159.66	161.89	28	-15.51 ± 0.94	N/A
9	Cyc(CQITVPLPEGVVIVETC)	172.57	164.97	13.8	-5.66 ± 0.65	N/A
10	Cyc(CLIVEYKPEGVEIIYEC)	191.79	191.78	37.6	-7.90 ± 0.34	N/A
11	Cyc(CSTKQNMPEGTVLIYSC)	180.703	181.63	5.8	-2.29 ± 0.85	N/A
12	Cyc(CVVSQDRPEGTVLLYTC)	162.55	164.24	402.4	-13.20 ± 0.15	N/A
13	Cyc(CVIAYAAPEGYILVTVC)	218.21	223.39	162.2	-19.03 ± 0.19	N/A

Manipulation in the template peptide sequence cyc(CIIHHKKPTGMIRIHQC) at its selected poses with pre and post-design, binding free energy by MM-GBSA and rigorous calculation.

Appendix C - **Sequence optimization for the template peptide targeting Spike protein RBD.**

ID	Sequence	Pre-design score	Post-design score	MD Sim Run time (ns)	Rough Estimate of Binding Energy by MMGBSA (kcal/mol)
1	NSLKELTEKEATDKYYK	-225.63	-232.95	66.6	+10.845 ± 0.586
2	SIEELIKKNKKTIEKYK	-77.33	-99.762	321	-17.274 ± 0.198
3	EDLKKWAKQLEELKYK	-228.43	-237.75	198	-24.442 ± 0.369
4	YESLLRFLTKLKDDKYK	-231.29	-240.42	67.4	-12.467 ± 0.396
5	ATEEETKKAKAEDKKYK	-226.55	-239.59	193.4	-14.902 ± 0.334
6	EEYEKMLESLKELKHYK	-81.938	-105.68	183.4	-29.737 ± 0.351
7	YTFKEKTDHEATDKYYK	-225.63	-230.47	13.6	-03.171 ± 0.669
8	YTFEVKEKHKRDIEKYK	-77.33	-91.879	164.6	-12.253 ± 0.322
9	YTFLKKWAHQAEELKYK	-228.43	-229.85	231.2	-36.851 ± 0.518
10	YTFLEKFLHKLKDKKYK	-231.29	-236.83	76.2	-12.328 ± 0.405
11	YTFEAKVKHEQEDKKYK	-226.55	-235.7	12.8	-05.877 ± 0.730
12	YTFEAKKKHEEELRHYK	-81.938	-92.738	1032.8	-28.023 ± 0.178
13	SEDLAKWAETLERLKYK	-231.87	-230.58	131.6	-21.556 ± 0.449
14	NVEKRTKEAKAEDSHYK	-234.67	-234.14	1572.2	-30.562 ± 0.164
15	SIEELIKRLKKTEEKYK	-219.99	-220.31	1148.6	-24.423 ± 0.131
16	YDDLEKWKKQLEELKYK	-232.34	-229.13	31.2	-10.648 ± 0.376
17	EEEEKTKKILTELAKYK	-215.51	-218.78	265.4	-24.564 ± 0.311
18	STEDEEKKAKAEDSLYK	-237.45	-238.36	165.8	-12.937 ± 0.340
19	STDQEYKKAVEHDRKYK	-232.84	-233.12	13.6	-07.521 ± 0.551
20	PDDLKIWAQILLEAFYK	-233.18	-232.02	2000	-34.147 ± 0.097
21	SIEDKIKKAKAVVAKYK	-228.34	-229.62	245.2	-18.212 ± 0.289
22	QTQDEWKKAVEEEKKYK	-233.95	-234.59	155.8	-29.358 ± 0.118
23	STDDEKKKALAEDQQYK	-233.02	-236.3	1039.4	-17.186 ± 0.152

24	SELEKALKALSSLSTYK	-225.2	-230.04	351.8	-22.292 ± 0.182
25	EEEEKKKTAKKELDTYK	-220.7	-226.02	274.6	-25.472 ± 0.336
26	YEELKKWAKTLEEAKYK	-230.58	-228.17	239.4	-16.526 ± 0.170
27	SLEDLIEEAKRKIEKYK	-220.1	-219.24	371.2	-16.070 ± 0.169
28	TFLEEWLKQKEELKLK	-168.15	-172.52	2000	-28.882 ± 0.092
29	TFLELWKRQAKDLRYK	-178.7	-181.95	435.4	-21.133 ± 0.240
30	TFLEKWLKQADDLKNK	-172.51	-175.92	111	-22.651 ± 0.477
31	DELEITADKLRQAMEK	-231.04	-233.77	365	-14.595 ± 0.178
32	DELEVLAEQLREALAK	-226.87	-230.03	83.8	-15.091 ± 0.310
33	DPWEIWAKLLKESRDK	-233.1	-237.51	1421	-29.714 ± 0.085
34	GTEWLYKVWKAFVENMS	NA	NA	1350.8	-29.688 ± 0.122
35	STEWLKKVHWKFHAKFS	NA	NA	29.6	-03.422 ± 0.468
36	GTEEIYKRWATVMKKKK	NA	NA	880.2	-11.344 ± 0.098
37	GTEWLKKVWEAFVKNMS	NA	NA	189.4	-28.698 ± 0.335
38	GTEEIYLWMVTVHKQKK	NA	NA	151	-17.313 ± 0.471
39	STEYLKRIFWMFHEYES	NA	NA	342	-29.252 ± 0.217
40	NVEKRTKEAKAEDSHYK	NA	NA	1253.8	-45.021 ± 0.153
41	QTQDEWKKAVEEEKKYK	NA	NA	1014.12	-23.381 ± 0.116

Optimized peptide sequences with pre and post-design score, length of the simulation and estimated binding free energy by MM-GBSA method.

Appendix D - **Stapling of optimized peptide sequences targeting Spike protein RBD.**

Sl. No	ID	Sequence	MD Sim Run Time (ns)		Rough Estimate of Binding Energy by MMGBSA (kcal/mol)		Absolute Binding Energy by BFEE (kcal/mol)
			Stapled	Unstapled	Stapled	Unstapled	
1	Ngu-123	GTEWLKKVWDAFVKNMS	2000	189.4	-24.958 ± 0.063	-24.958 ± 0.063	NA
2	Ngu-335	STDYLKRIFWMFHEYES	1547	342	-28.557 ± 0.106	-29.252 ± 0.217	NA
3	Ngu-129	PDDLKIWAQILDEAFKK	2000	2000.0	-29.000 ± 0.072	-34.147 ± 0.097	NA
4	Ngu-129a	PDDLKAWAQALDEAFKK	371		-18.303 ± 0.144		NA
5	Ngu-29*	NVEKRTKEAKAEDAHYK	2000	1253.8	-33.970 ± 0.119	-45.021 ± 0.153	-05.02 ± 15.37
6	Ngu-29a	NVEKRTDEAKAEDAHYK	162		-14.611 ± 0.901		NA
7	Ngu-29b*	NVEKRTKEAKAEDSHYK	2000		-44.746 ± 0.122		-13.31 ± 01.63
8	Ngu29c	NVDKRTKEAKAEDAHYK	77.6		-12.620 ± 0.455		NA
9	Ta-93*	YTFLKKWAHQAEEAKYK	1398.8	198.0	-33.945 ± 0.151	-36.851 ± 0.518	-17.41 ± 03.10
10	Ta-200	PDPWAIWAKLLKESRDK	18.2	1421.0	-13.137 ± 0.491	-29.714 ± 0.085	NA
11	Ta-196	YTFDAKKKHEDELKHYK	415.4	1032.80	-21.832 ± 0.213	-28.023 ± 0.178	NA
12	Ta-26	YAFLAEWLKQKDELKLK	189.8	2000	-29.407 ± 0.259	-28.882 ± 0.092	NA
13	Ta-39***	GTEWLYKVWKAFADNMA	2028.8	1350.8	-33.936 ± 0.071	-29.688 ± 0.122	-26.32 ± 08.64
14	Ta-39a	GTEWLYKVWKAFADNMA	2000		-28.884 ± 0.055		NA
15	Ta-39us	GTEWLYKVWKAFADNMA	1222		-27.037 ± 0.076		NA

Optimized peptide sequences with pre and post-design score, length of the simulation and estimated binding free energy by MM-GBSA method.