Strategies for Identification of Small Molecule Inhibitors of Ad2 E3-19K/HLA-A2 Binding Interaction

BY

NIKITA RAYMOND DSOUZA

B. Pharm, Mumbai University, 2013

THESIS

Submitted as a partial fulfillment of the requirements for the degree of Master of Science in Bioinformatics in the Graduate College of the University of Illinois at Chicago, 2017

Chicago, Illinois

Defense Committee

Dr. Michael E. Johnson, Chair and Advisor

Dr. Marlene Bouvier, Microbiology and Immunology

Dr. Ao Ma

AKNOWLEDGEMENTS

I am extremely grateful to my advisor, Prof. Michael E. Johnson for giving me the chance to be a part of his group, providing strong support and excellent guidance throughout the research. I am very thankful for his continuous motivation and assistance throughout. It was great to learn and work under the guidance of a highly intellectual and serene advisor like him.

I would also like to express my gratitude towards Dr. Jinhong Ren, my mentor. It was her continuous vigilance and excellent guidance that helped me accomplish my research goals. Her broad background and practical experiences were very helpful.

I would also like to thank all the members in Dr. Johnson's lab, for their constant encouragement, support and friendship throughout the course of my project.

I would like to acknowledge and thank Prof. Marlene Bouvier and her laboratory, specially Dr. Lenong Li for determining and providing us with the crystal structures used throughout the project. I sincerely appreciate Dr. Hui Deng, supervised by Prof. Marlene Bouvier, and Dr. Hyun Lee for the time and effort they invested to carry out experimental assays for this project.

I also sincerely appreciate the advice and guidance of my committee members Dr. Michael E. Johnson, Dr. Marlene Bouvier, and Dr. Ao Ma.

Finally, I would like to thank my parents and my sister for their constant encouragement, love and being there always.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACE: Analytical continuum electrostatics;

Ad: Adenovirus;

Ad2: Adenovirus serotype 2;

Ad4: Adenovirus serotype 4;

ADME: Adsorption, Distribution, Metabolism and Excretion;

Ag: Antigen;

CTL: cytotoxic T-lymphocyte;

CS: Consensus Site;

CSM: Computational Solvent Mapping;

ER: Endoplasmic Reticulum;

FFT: fast Fourier transform;

GA: Genetic Algorithm;

GAFF: General AMBER force field;

HLA: Human Leukocyte Antigen;

HTS: High throughput Screening;

LBVS: Ligand-based virtual screening;

MD: Molecular Dynamics;

MHC: Major Histocompatibility Complex;

MM-GBSA: Molecular Mechanics Generalized Born Surface Area;

MM-PBSA: Molecular Mechanics Poisson–Boltzmann Surface Area;

MSCS: Multiple solvent crystal structure;

PAINS: Pan-assay interference compounds;

PDB: Protein Data Bank;

PLP: Piecewise Linear Potential;

PME: Particle mesh Ewald;

REOS: Rapid Elimination of Swill;

RMSD: Root-mean-square deviation;

SBDD: Structure-Based Drug Design;

SBVS: Structure Based Virtual Screening;

S_c: Shape-complementarity coefficient;

SPR: Surface Plasmon Resonance

TBSA: Total Buried Surface Area;

UCSF: University of California, San Francisco;

VMD: Visual Molecular Dynamics;

VDW: van der Waals;

SUMMARY

Adenovirus (Ad) infections are widespread in the human population, and causes infections linked to many gastrointestinal, respiratory and ocular illnesses. *In vitro* and *in vivo* studies have shown that the binding of the adenovirus protein E3-19K of the adenovirus with the Major Histocompatibility Complex (MHC) class I molecules plays a role in infection [1, 2]. "E3-19K binds to MHC class I molecules and retains these molecules in the endoplasmic reticulum (ER) of the Ad infected cells, thus inhibiting the lysis of the Ad infected cells by cytotoxic T-lymphocytes (CTLs) [3-5]."

The goal of this research study is to evaluate the Ad2 E3-19K – HLA-A2 complex as a target for its potential to bind small molecule inhibitors. The knowledge of the crystal structure [23], and the interaction surface of the Ad2 E3-19K – HLA-A2 complex provides us a foundation to conduct Structure-Based Drug Design (SBDD). We aim to disrupt the binding of E3-19K protein of adenovirus with the HLA-A2 molecule, which could then sensitize the adenovirus infected cells to lysis by the CTLs.

To achieve this, we used computational methods to determine potential binding sites for small molecules. Hence, we analyzed the crystal structure of the complex, including the binding surface of the proteins involved in the complex; we compared the Ad2 E3-19K – HLA-A2 and Ad4 E3-19K – HLA-A2 structures [25] to determine whether the binding site selected varies between strains. A structure-based virtual screening (SBVS) protocol was designed to probe the Ad2 E3-19K – HLA-A2 interface for its potential to be targeted by small molecules. Per-residue decomposition of the interface interactions was obtained from the analysis of Molecular Dynamic (MD) trajectories which was used to determine residues in peptides that could also be used to target the protein.

We shortlisted compounds to be tested experimentally, using the scoring functions of the molecular docking program; we found 7 compounds with binding affinity to HLA-A2 out of the 36 compounds tested. In addition, the peptide sequence obtained from per-residue decomposition has been used to synthesize peptides that can be tested for its binding potential through experiments.

CHAPTER 1

CRYSTAL STRUCTURE OF E3-19K – HLA-A2 COMPLEX

1.1   Introduction

Adenoviruses (Ad) are not usually of major concern in healthy adults, but they can be severe and even fatal in immunocompromised children and adults [6-10]. Generally, healthy adults can control the virus, although some may develop persistent infections that last for a lifetime. There is no specific treatment, but antiviral drugs like cidofovir, ribavirin, and ganciclovir reduce virus replication to some extent. However, the toxicity, and ineffectiveness of these drugs in some patients drives us to explore small molecules for their potential to inhibit the binding of the adenovirus protein E3-19K to MHC class I molecules. Ads have ≥ 70 known serotypes that are further classified into seven species from A through G, based on the studies conducted [11-13]. MHC class I antigen (Ag) mediates antigen presentation; human MHC molecules are known as Human Leukocyte Antigen (HLA). The HLA complex is a group of proteins derived from the information provided by the HLA gene. HLA is responsible for helping the body distinguish the body's own protein from a foreign protein.

"In the late 1980's the role of the E3-19K protein was brought to light; Ad virus protein E3-19K binds to MHC class I molecules and retains them in the endoplasmic reticulum of the Ad infected cell, blocking their escape to the surface of the cell [3-5, 42]." By subverting the function of MHC class I molecules (Ag presentation), the virus makes the infected cells less susceptible to lysis by cytotoxic T-lymphocytes (CTLs) [4, 14-16]. A study conducted on cotton rats supported the role of E3-19K in Ad infections [2]. The lungs of the cotton rat were infected with the wild-type Ad and a mutant Ad. The mutant contained a deletion of the gene encoding E3-19K. The immunopathological responses were less severe for the wild-type Ad, when compared to the

mutant Ad, when E3-19K was absent; supporting the evidence of the association of the E3-19K protein with MHC class I molecules. E3-19K (Type I transmembrane glycoprotein) is comprised of an N-terminal ER-lumenal domain that connects with the MHC class I molecule's ER-lumenal domain, whereas the dilysine motif in the C-terminal of the cytosolic tail of E3-19K is responsible for signals to localize the E3-19K – MHC I complex in the ER [14, 17-23].

1.2   Crystal structure of Ad2 E3-19K – HLA-A2 complex

The determination of the crystal structure by Li et al [23] gives us a basis to explore the complex for its potential to bind small molecules which would inhibit their interaction, and further make Ad infected cells susceptible to lysis by CTLs. The Ad2 E3-19K of species C bound to HLA-A2 complex was resolved at 1.95 Å by X-ray diffraction [23]. The authors analyzed the binding interface, in which a molecular interface of 1929 $Å^2$ is shielded from solvent. The shape-complementarity coefficient (S_c) is 0.68 (1 being the S_c for a perfect geometrical fit) [23]. The paper separates the binding interface into site 1, site 2, site 3, and site 4 (Fig. 1). Site 1 occupies 36% of the total buried surface area (TBSA), site 2 occupies 16%, site 3 contributes 17%, and site 4 represents 31% of the TBSA [23]. The crystal structure gives us insights into the binding interface, the interactions between the two proteins, and residues important for the interaction. Based on these insights we integrate computational methods into the structural analysis for potential binding sites. Site 1 has a salt bridge between Lys27 (E3-19K) and Glu53 (HLA-A2), and hydrogen bonds between Lys 16 and Tyr 48 of E3-19K with HLA-A2 residues Arg48 and Gln53, respectively [23]. Solvent-mediated hydrogen bonds as well as other hydrophobic contacts supplement the above interactions. Site 2 features a salt bridge between Lys42 (E3-19K) and Glu177 (HLA-A2), and is an important interaction as shown previously by mutation of the Glu[177] to Lys[177] [24]. This mutation was responsible for the failed interaction between the Ad2 E3-19K

in species B, C, D and E and HLA-A2. The interaction at site 3 involves water-mediated hydrogen bonds and hydrophobic contacts. However, site 3 lacks key interactions between conserved residues, as is observed in the other three sites. Site 4 includes strictly conserved residues Met89, Gln92 and Tyr93 of the E3-19K protein, forming hydrophobic interactions and hydrogen bonds with HLA-A2 residues His 13, Lys19, Ser20, Asn21, Phe22 and Glu69 [23].

a.

b.
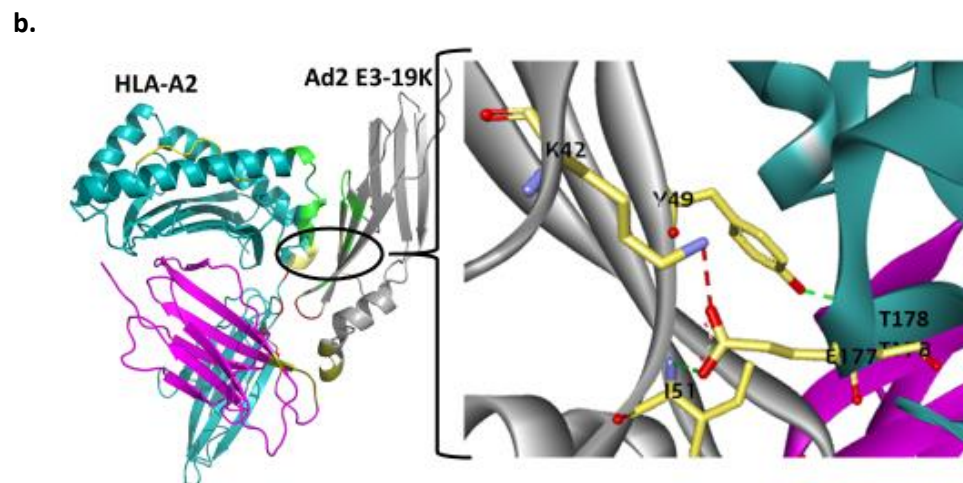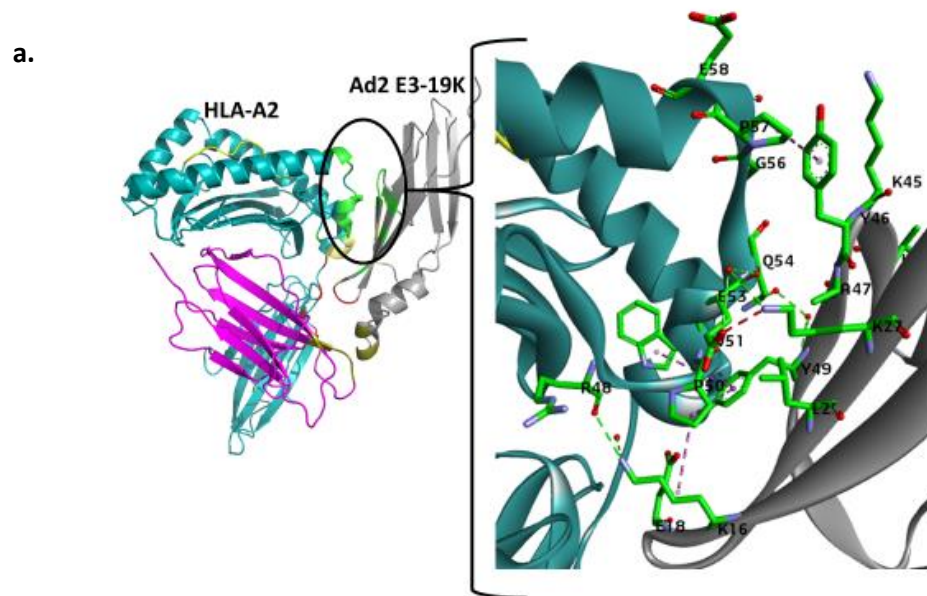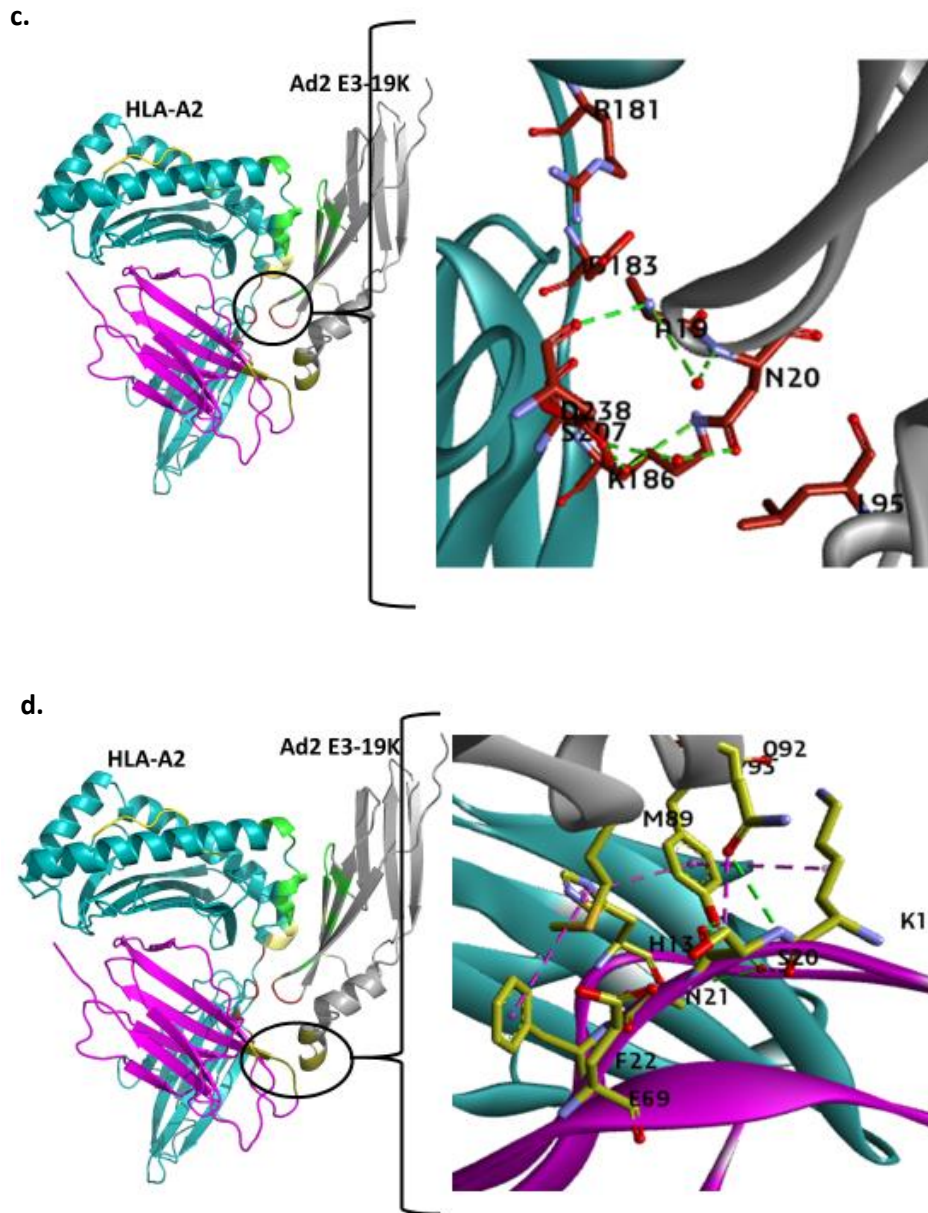
Figure 1 Crystal structure of Ad2 E3-19K – HLA-A2 and the interface interaction sites 1 (a), 2 (b), 3 (c), and 4 (d). Nitrogen, oxygen and Sulphur atoms are colored in blue, red and yellow respectively. Water molecules are shown as red spheres. Hydrogen bonds, charged interactions and hydrophobic contacts are indicated by green, red and purple dashed lines, respectively. Images were prepared using Discovery Studio 2016.

1.3   Crystal Structure of Ad4 E3-19K – HLA-A2 Complex

The crystal structure of the Ad4 E3-19K – HLA-A2 complex identifies species-specific

features in recognizing HLA-A2, and common binding characteristics between the two crystal

structures. Li et al reported the x-ray crystal structure of the E3-19K protein of the Adenovirus serotype 4 belonging to species E bound to HLA-A2 at 2.64 Å resolution [25]. The binding interface shielded a molecular interface of 2112 $Å^2$ and S_c was determined to be 0.66. The binding interface was again divided into four sites like the Ad2 E3-19K/HLA-A2 binding interface [25].

1.4   Comparison of Ad2 E3-19K/HLA-A2 and Ad4 E3-19K/HLA-A2

The specific mechanisms to interfere with MHC I pathway, and suppression of Ag have been evolved by viruses. Both Ad4 and Ad2 E3-19K protein avoid contacts with the bound antigenic peptide present in the groove on HLA-A2, thus eluding peptide specificity [25]. At sites 1 through 3 compared to Ad2 E3-19K, the E3-19K protein of Ad4 forms no salt bridges and fewer hydrogen bonds, instead relying mostly on hydrophobic bonds [25]. The sites 1 and 2 of the Ad2 and Ad4 proteins mostly have conserved residues, except for a few; site 3 comprises mostly of polymorphic (different) residues. However, site 4 has strictly conserved residues in the Ad2 and Ad4 E3-19K protein contacting the same residues on HLA-A2 [25]. Only 36% of sequence identity is shared by Ad2 and Ad4 proteins, but they have similar complex structures because of the conservation of key structural residues and determinants, and MHC I binding function [25]. Furthermore, the Gln54 and Glu177 residues of both E3-19K proteins make considerable contacts with HLA-A2 residues. The Q54G mutant of HLA-A2 interacted weakly with Ad2 and Ad4 E2-19K protein. The interaction of HLA-A2 (E177K) with Ad2 and Ad4 E3-19K was abolished. The above results were determined by using native gel band-shift assay [25]. The results provided evidence for the importance of the two residues, highlighting the importance of the Glu177 residue present at the interface on HLA-A2 [25].

CHAPTER 2

ANALYSIS OF THE BINDING INTERFACE

2.1   Introduction

The crystal structures of the two proteins highlighted the importance of the interaction of Glu177 (HLA-A2) and Lys42 (E3-19K) for binding, and consequently subverting HLA-A2 antigen presentation function. We decided to look at the energy contribution of each of the residues at the interface. The paper also mentions about the chemical properties of the binding surface on both the proteins of the complex [23]. We used UCSF Chimera to visualize the surface of the E3-19K and HLA-A2 protein, also visually analyzing for the presence of potential binding pockets on the binding surface of each protein. Computational Solvent Mapping was also conducted to identify potentially favorable binding sites. The details of the methods used will be described in the individual sections below. The Ad2 E3-19K/ HLA-A2 crystal structure was downloaded from the Protein Data Bank (PDB) code: 4E5X [23].

2.2   Per-Residue Decomposition to calculate the binding free energy

Assisted **M**odel **B**uilding with **E**nergy **R**efinement – the AMBER14 suite of programs [26] was used to perform MD simulations for the Ad2 E3-19K/HLA-A2 complex for 6ns. The ff99SB forcefield parameters were used for the protein complex, i.e., the Ad2 E3-19K – HLA-A2 complex. The protein-protein complex was solvated in a 20 Å octahedral box of TIP3P water molecules surrounding the protein on all sides. We used the SHAKE algorithm, which constrains all the bonds, including the hydrogen bonds. For long range electrostatics, the particle mesh Ewald (PME) is used by the AMBER MD package. Hence, to calculate the interactions, the long range electrostatic cut-off distance was kept at 10 Å. 10,000 steps of conjugate gradient minimization was used to minimize the system. The minimization was followed by MD equilibration for 100

picoseconds (ps) over constant volume with weak restraints on the solvated system. The system was gradually heated at approximately 20 ps from 0K to 300K. Now that our system has been heated over constant volume, we run a 100 ps equilibration over constant pressure to relax the density of the water. The system was already heated to 300K, which was maintained in the second equilibration. We need to verify whether the system has been equilibrated, which was done by looking at the plots for total energy, temperature, density and backbone root mean square deviation (RMSD). For the density graph, the data is recorded only for the second equilibration, because the first heating run was performed under constant volume conditions.
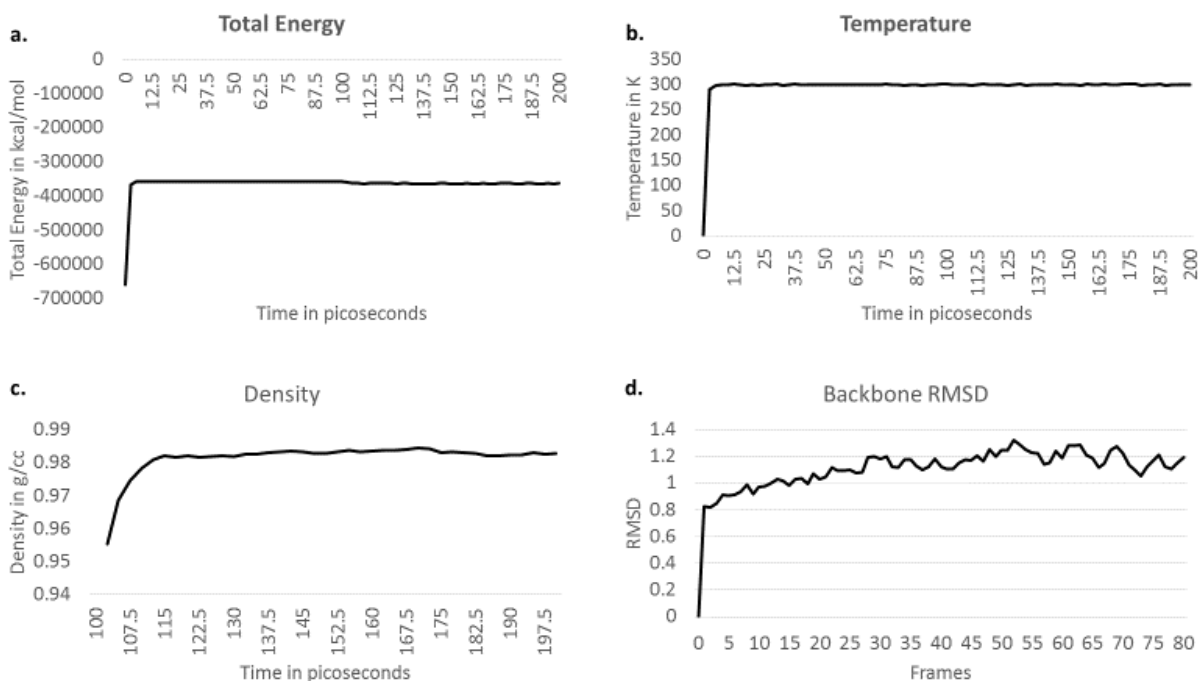
Figure 2 Plots of Total energy (a), Temperature (b), Density (c) and Backbone RMSD (d).

Fig. 2 displays the plots for total energy, temperature, density and backbone RMSD. The total energy, temperature, density plots have converged, and the RMSD of the backbone atoms seems to have leveled.

A 6 nanosecond production runs was then performed, recording coordinated for every 2.5 picoseconds. A total of 2400 frames or snapshots were extracted. We look at the RMSD of the complex backbone atoms relative to the starting complex structure to assess the stability. In Fig. 3 we can see that the RMSD seems to have converged for the complex. We also used Visual Molecular Dynamics (VMD) http://www.ks.uiuc.edu/Research/vmd/ [41] to visualize the MD trajectories, to see the interaction of the two proteins surrounded by a solvent box.
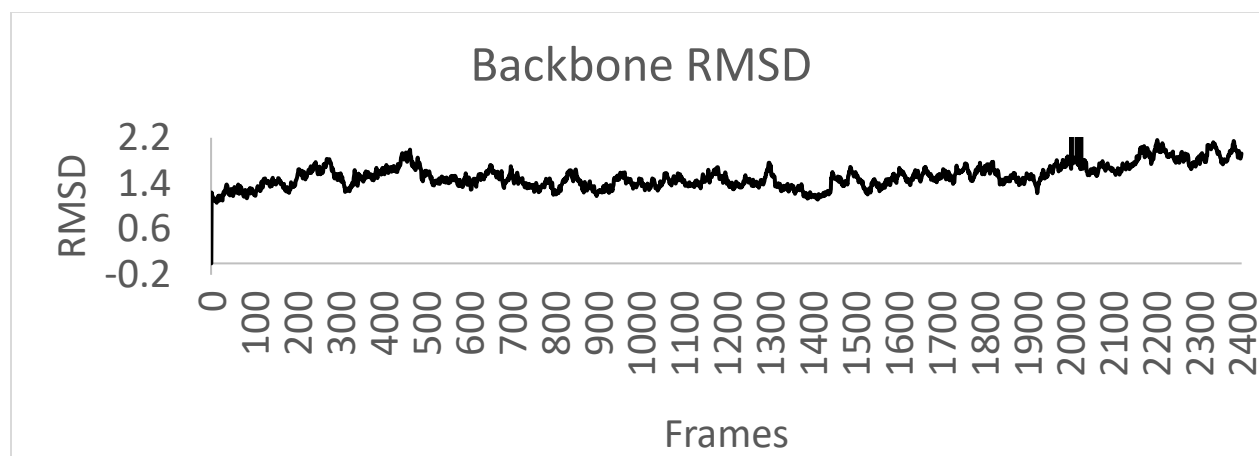
Figure 3 Backbone RMSD plot during the 6 nanosecond production run.

For the per-residue energy computation we used the MM-GBSA (Molecular Mechanics Generalized Born Surface Area) and MM-PBSA (Molecular Mechanics Poisson–Boltzmann Surface Area) method in Ambertools [26]. We selected the option which provides for van der Waals and electrostatic energy contributions separately, which was calculated from the Molecular Mechanics. We get an output file containing the interaction energy of each atom in a residue with other atoms in the system.

The per-residue decomposition calculates the energy by adding up the energy contributed by a single residue's interactions with the other residues. The binding interface residue energies of

both Ad2 E3-19K and HLA-A2 were calculated using the MD trajectories. The results obtained included binding energies of both Ad2 E3-19K and HLA-A2 protein interface residues, and was correlated with the crystal structure [23].

2.2.1 HLA-A2 Residues

The table below gives the energy decomposition, summing up the individual contributions to give a total binding free energy of each residue. We picked the set of residues present on the binding interface to analyze. Most of the interface residues noted in the crystal structure had a negative total binding free energy value. Given below are the residues identified in the crystal structure study. Site 1 highlighted in green in Table 1 below consists of residues Pro50, Trp51, Glu53, Gln54, Gly56 and Pro57, and has a significant binding free energy calculated with the interaction of these residue with other residues. As noted above Site 1 features a salt bridge involving the residue Glu53, the energy contribution of which can be seen in the table under average electrostatic energy, but this is compensated by the average polar solvation energy. Glu177 at Site 2 also has a significant energy, supporting the role Glu177 plays in the E3-19K/HLA-A2 interaction [25]. The free energy by the electrostatic interaction of Glu177 residue is compensated by the polar solvation energy, thus supporting the presence of salt bridge at Site 2. Site 3 residues, Ser207 and Asp238 have a total binding free energy contributed by all hydrophobic interactions and solvent-mediated hydrogen bonds. Site 4 being an apolar site as mentioned in the crystal structure paper by our collaborator Dr. Marlene Bouvier and her group, has hydrophobic interactions that contribute significantly to the total binding free energy of the conserved residues involved in the interaction. The per-residue decomposition of the HLA-A2 residues shows that residues at sites 1, 2 and 4 each have binding free energies substantially greater than the residues present on the binding interface of HLA-A2 at site 3.

9

Location	van der Waals (VDW) (kcal/mol)		Electrostatic (kcal/mol)		Polar Solvation (kcal/mol)		Non-Polar Solvation (kcal/mol)		TOTAL (kcal/mol)	
	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD
PRO50	**-3.74**	**0.49**	**-3.47**	**0.49**	**5.07**	**0.6**	**-0.71**	**0.04**	**-2.87**	**0.63**
TRP51	**-2.67**	**0.3**	**-4.93**	**0.58**	**4.41**	**0.53**	**-0.17**	**0.02**	**-3.36**	**0.35**
GLU53	**-1.95**	**0.77**	**-63.8**	**3.07**	**64.91**	**2.58**	**-0.55**	**0.03**	**-1.43**	**0.85**
GLN54	**-4.61**	**0.44**	**-6.08**	**0.94**	**9.62**	**0.81**	**-0.68**	**0.03**	**-1.75**	**0.56**
GLU55	-1.08	0.22	-39.7	1.71	40.3	1.79	-0.03	0.02	-0.49	0.21
GLY56	**-1**	**0.29**	**1.06**	**0.32**	**-0.75**	**0.3**	**-0.21**	**0.03**	**-0.89**	**0.3**
PRO57	**-0.92**	**0.15**	**1.1**	**0.23**	**-1.12**	**0.21**	**-0.19**	**0.03**	**-1.13**	**0.2**
GLY175	-0.73	0.15	1.5	0.31	-1	0.42	-0.04	0.01	-0.26	0.2
GLU177	-1.33	0.77	-37.9	3.21	36.95	3	-0.52	0.09	-2.76	1.03
SER207	**-0.84**	**0.14**	**0.93**	**0.67**	**-0.61**	**0.5**	**-0.13**	**0.03**	**-0.65**	**0.72**
PHE208	-0.11	0.01	-0.59	0.13	0.48	0.14	0	0	-0.21	0.06
TYR209	-0.18	0.02	0.6	0.09	-0.77	0.12	0	0	-0.35	0.08
ASP238	**-2.37**	**0.44**	**-11**	**2.29**	**12.57**	**2.36**	**-0.52**	**0.07**	**-1.35**	**0.68**
GLY239	-0.62	0.13	0.57	0.13	-0.74	0.2	-0.03	0.02	-0.82	0.23
THR240	-0.89	0.13	-0.28	0.12	0.63	0.18	-0.1	0.02	-0.65	0.18
HIS13	**-4.24**	**0.66**	**-5.08**	**0.72**	**5.99**	**0.52**	**-0.68**	**0.04**	**-4**	**0.86**
PRO14	-0.38	0.72	-4.05	0.7	2.04	0.43	-0.13	0.03	-2.52	0.55
LYS19	**-2.87**	**0.61**	**0.97**	**4.23**	**-0.15**	**3.98**	**-0.51**	**0.07**	**-2.56**	**0.79**
SER20	**-2.37**	**0.52**	**-1.68**	**1.02**	**3.23**	**0.94**	**-0.52**	**0.05**	**-1.35**	**0.7**
ASN21	**-2.08**	**0.22**	**-0.62**	**0.49**	**1.18**	**0.37**	**-0.07**	**0.02**	**-1.6**	**0.35**
PHE22	**-1.82**	**0.4**	**-0.04**	**0.28**	**0.38**	**0.32**	**-0.29**	**0.07**	**-1.78**	**0.44**
PHE70	-0.18	0.02	0.08	0.14	-0.16	0.16	0	0	-0.26	0.07

Table 1. Per-residue decomposition of HLA-A2 residues
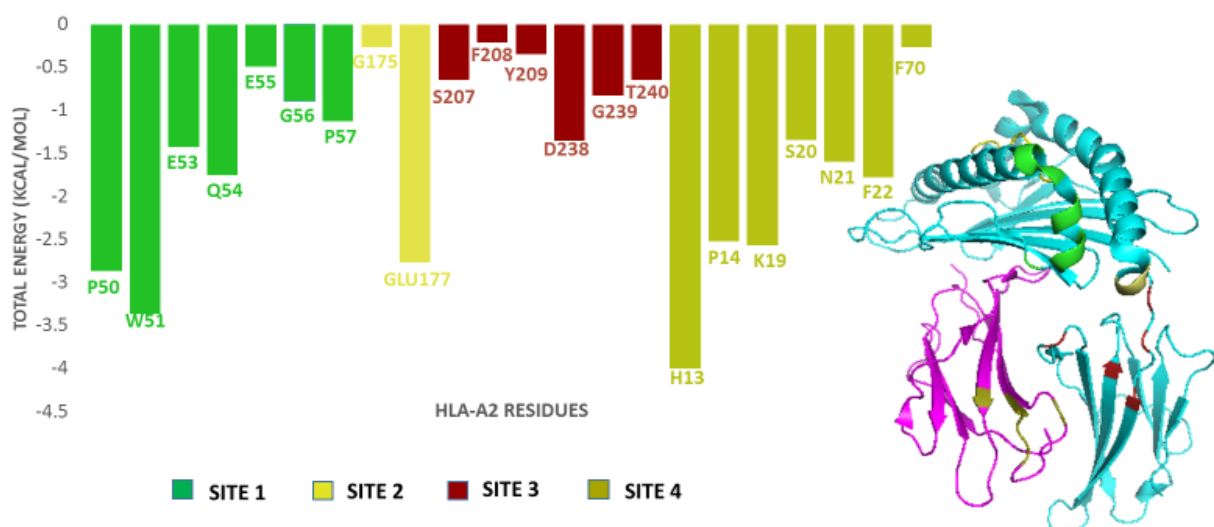
Per residue decomposition HLA-A2

Figure 4 Per-residue decomposition for the HLA-A2 residues highlighted based on the sites.

2.2.2 Ad2 E3-19K Residues

The Table 2 below displays the separate energy contributions by different interactions towards the total binding free energy in Ad2 E3-19K protein residues. We will look at the per residue decomposition results in the order of the sites at the binding interface. The Leu25, Lys27, Tyr46, Tyr49 are residues highlighted in the crystal structure. Lys27 forms a salt bridge with the Glu53 residue of the HLA-A2 protein, which can be seen in the electrostatic energy sections, although the polar solvation energy compensates the energy contributed by the electrostatic energy. Tyr49 not only interacts with HLA-A2 residues at site 1, but also interacts with HLA-A2 residues at site 2, explaining the significant total binding free of this residue. Lys42 is a highly-conserved residue present on site 2 that is involved in a salt bridge interaction with Glu177 residue (HLA-A2). The electrostatic energy contributed by the breaking of the salt bridge is substantial -

11

88.48 kcal/mol. Ala19 and Leu95 at site 3, like HLA-A2 residues at site 3, have significant total

energy because of hydrophobic interactions and solvent-mediated hydrogen bonds. Site 4 residues

Met89 and Tyr93, strictly conserved in Adenoviruses also have extensive total binding free energy.

Gln92 also has a noticeable binding free energy level. For the E3-19K protein of the Ad type 2,

the residues of sites 2 and 4 are more significant than residues at the other two sites.

Location	van der Waals (VDW) (kcal/mol)		Electrostatic (kcal/mol)		Polar Solvation (kcal/mol)		Non-Polar Solvation (kcal/mol)		TOTAL (kcal/mol)	
	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD
ALA 19	-2.74	0.41	-3.07	0.67	3.63	0.53	-0.66	0.04	-2.83	0.47
LEU 25	-2.45	0.42	-0.91	0.33	1.71	0.35	-0.37	0.04	-2.03	0.48
LYS 27	-0.37	0.74	-86.2	2.85	85	2.35	-0.26	0.03	-1.78	0.84
LYS 42	-0.4	0.68	-88.48	3.41	88.8	3.03	-0.44	0.04	-0.53	1.13
TYR 46	-3.82	0.41	-2.81	0.66	4.31	0.55	-0.64	0.04	-2.96	0.46
VAL 48	-1.13	0.21	-2.53	0.52	2.26	0.43	-0.04	0.02	-1.43	0.41
TYR 49	-5.76	0.8	-1.12	0.95	0.48	0.62	-0.86	0.04	-7.26	0.84
ILE 51	-1.3	0.31	-0.47	0.27	0.61	0.27	-0.29	0.04	-1.45	0.29
TYR 88	-0.94	0.93	-8.62	2.99	9.06	2.08	-0.53	0.07	-1.03	0.72
MET 89	-5.46	0.59	-2.19	0.54	4.14	0.67	-0.83	0.05	-4.33	0.78
GLN 92	-2.48	0.73	-7.07	3.43	8.58	2.08	-0.58	0.04	-1.56	1.08
TYR 93	-6.16	0.92	-7.07	1	8.64	0.78	-0.96	0.04	-5.56	0.83
LEU 95	-1.5	0.39	0.53	0.29	0.37	0.33	-0.39	0.04	-0.99	0.45

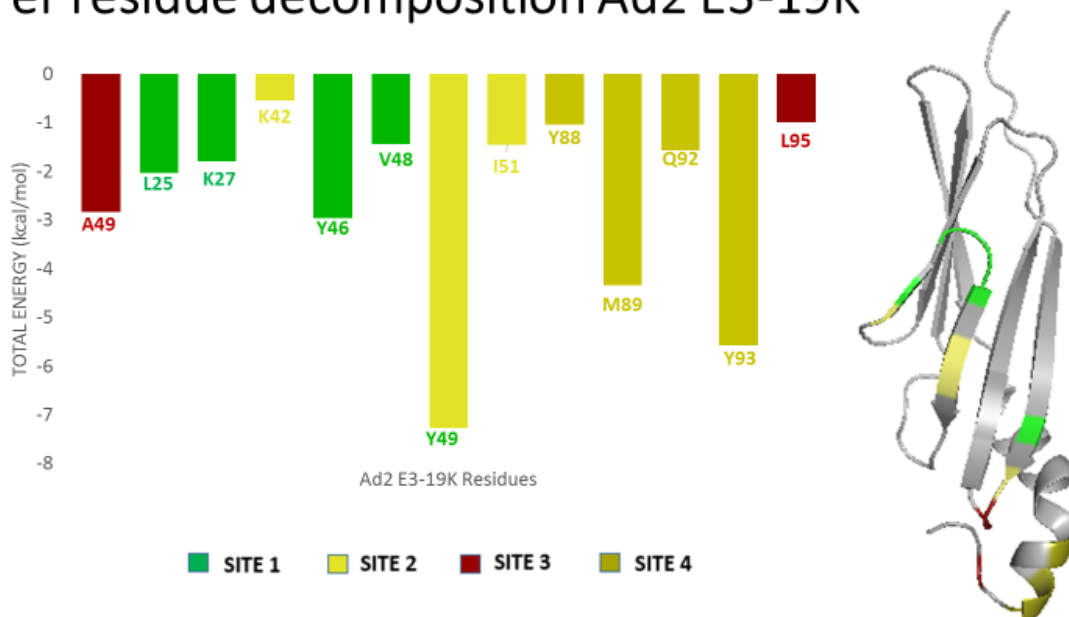Table 2. Per-residue decomposition of Ad2 E3-19K residues

Figure 5 Per-residue decomposition of Ad2 E3-19K residues highlighted based on the sites.

2.3 Binding Surface Analysis

Next, looking at the binding surface to determine the presence of a pocket through visual inspection would help us to select a potential binding site for small molecules. The surface is colored based on coulombic surface coloring in UCSF's Chimera [27]. "As described in the website https://www.cgl.ucsf.edu/chimera/docs/ContributedSoftware/coulombic/coulombic.html coulombic surface coloring calculates electrostatic potential per Coulomb's law:
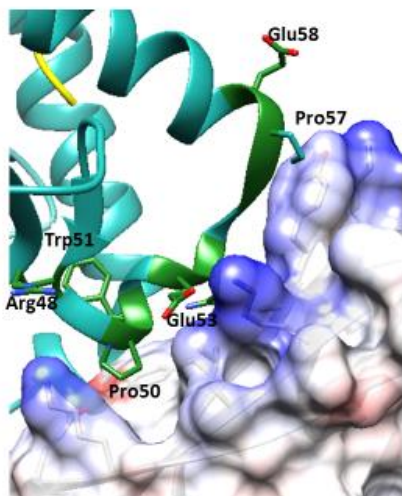
$\varphi = \Sigma \ [q_i \ / \ (\varepsilon d_i)]$

In the above equation, φ is the potential, q are the atomic partial charges, ε is the dielectric (representing screening by the medium or solvent), and d is the distance from the atoms [27]." We chose to use the Coulombic surface coloring over Electrostatic surface coloring, because the latter reads a pre-existing grid; we needed to see if the results obtained by computational methods

13

confirm the results obtained from the study of the crystal structure. The binding surface is colored by potential values calculated. We highlight the residues belonging to each site, and visually inspect the surface for the presence of a pocket on either Ad2 E3-19K or HLA-A2. The coloring of the surface also gives us an idea of the kind of region at the binding interface.

2.3.1 Surface Analysis at SITE 1

Since the binding of the proteins is tight, the pockets are shallow. Through visual inspection we found that Ad2 E3-19K protein had a shallow pocket at site 1. As suggested by the crystal structure study done by Dr. Bouvier and the team [23], we see in Fig. 6 that the electropositive (blue) region of E3-19K targets the electronegative (red) region of HLA-A2. Thus, for docking we require the presence of a pocket to dock small molecules which would fit in the position.
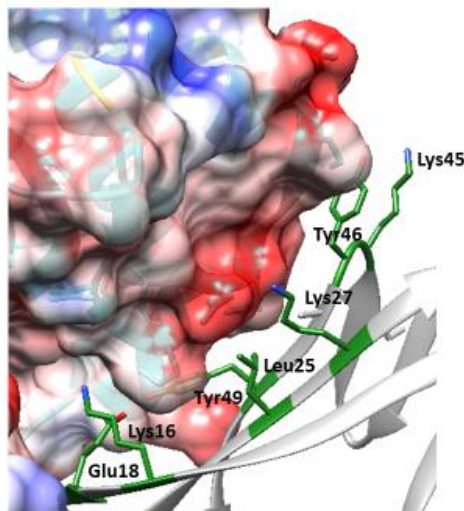

Figure 6 Binding surface analysis for site 1 (residues highlighted in green). The surface is colored using coulombic surface coloring in UCSF's Chimera. Images were prepared using UCSF's Chimera.

2.3.2   Surface Analysis at SITE 2

Site 2 looks like a potential site that could be targeted, based on the mutation results, and the interactions at the site. HLA-A2 has a pocket on site 2, as can be seen by the visual inspection of the surface of HLA-A2in the Fig. 7. Once again, as the study suggests, the electronegative region is present on Ad2 E3-19K, with the corresponding HLA-A2 region being electropositive [23]. It can be seen in Fig. 4 below that it correlates with the data obtained in the study.
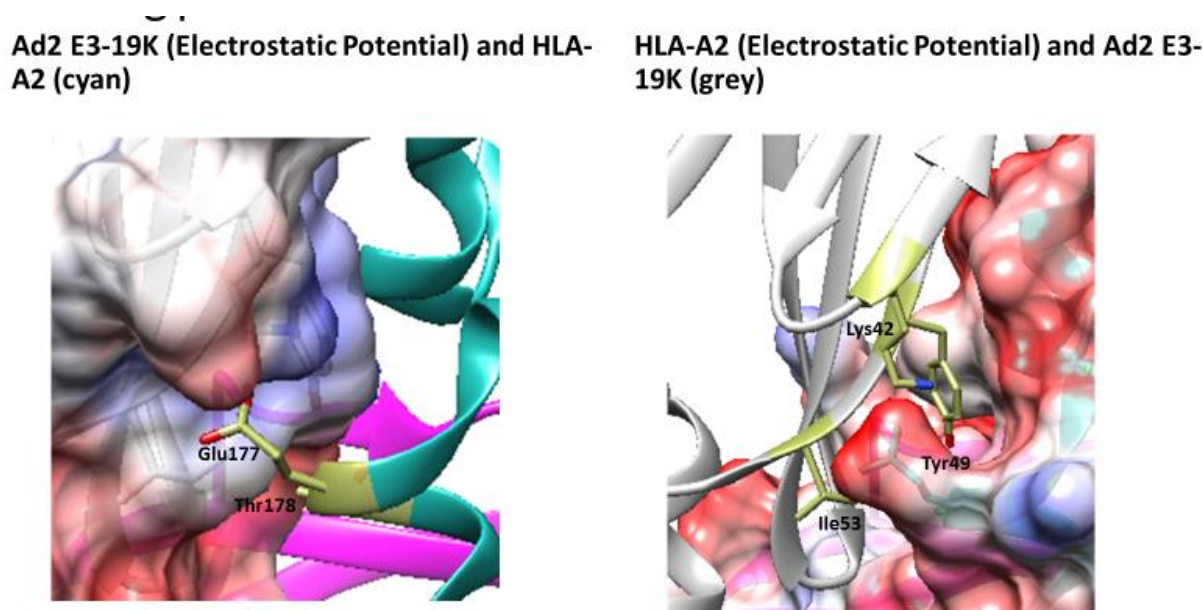

Figure 7 Binding surface analysis for site 2 (residues highlighted in pale yellow). The surface is colored using coulombic surface coloring in UCSF's Chimera. Images were prepared using UCSF's Chimera.

2.3.3   Surface Analysis at SITE 3

Site 3 comprises of polymorphic residues, varying in different species of adenoviruses; targeting this site would not be suitable because the residues may differ for differing serotypes. Visual inspection as in Fig. 8 helps us to see the pocket for site 3 is present on HLA-A2. The site

3 region consists of both electropositive and electronegative regions present on both the proteins, with the areas being complementary to each other on the respective protein.
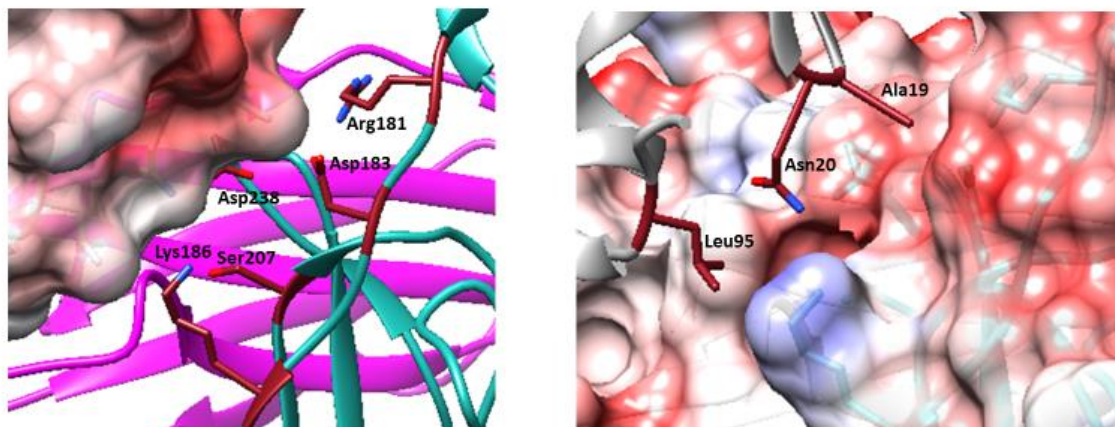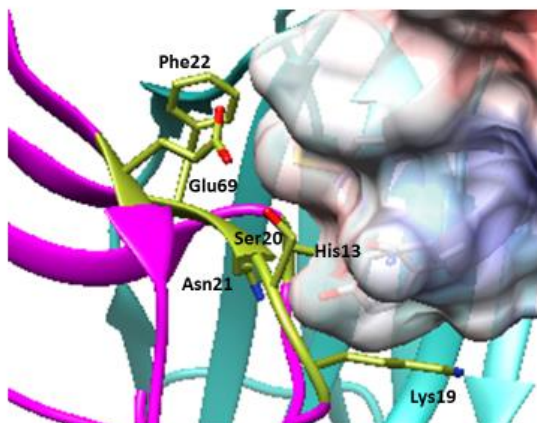

Figure 8 Binding surface analysis for site 3 (residues highlighted in red). The surface is colored using coulombic surface coloring in UCSF's Chimera. Images were prepared using UCSF's Chimera.

2.3.4 Surface Analysis at SITE 4

Site 4 consists mostly of apolar regions as can be seen in Fig. 9. The pocket is present on HLA-A2, which can be seen clearly when the surface of the Ad2 E3-19K protein is displayed. This site consists of highly conserved residues, shown in the comparison of Ad2 and Ad4 E3-19K proteins. This site also has a high shape complementarity coefficient of 0.801 [23].
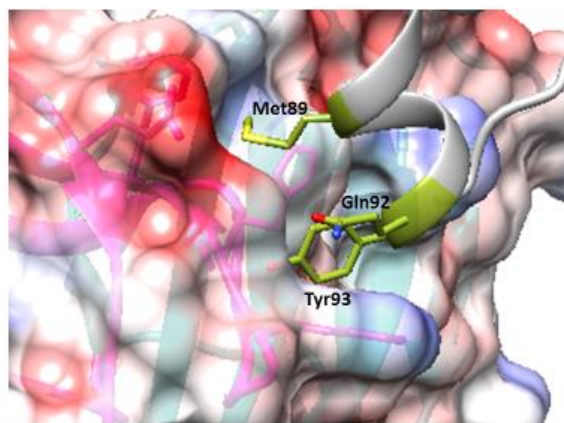
Figure 9 Binding surface analysis for site 3 (residues highlighted in red). The surface is colored using coulombic surface coloring in UCSF's Chimera. Images were prepared using UCSF's Chimera.

2.4   Computational Solvent Mapping (CSM)

A computational solvent mapping (CSM) study was carried out on the complex, as well as individual proteins. P. J. Goodford first developed a computational procedure to determine binding sites that were favorable energetically on biologically important molecules in 1985 [28]. Further, experimental tests have been conducted in 1990's using 6-8 organic solvents, multiple solvent crystal structure (MSCS) [29, 30]. X-ray crystallography was used to resolve the structures of proteins soaked in aqueous solutions of probes. Regions binding multiple probes were identified when the structures were superimposed [29, 30]. CSM, an analogous computational strategy is useful in identifying hotspots within potential binding sites. CSM moves small organic probe molecules around the protein surface, finding favorable binding positions on the protein surface, clustering the probes, and then, based on the average energy, ranks the clusters.

2.4.1   FTMap server

We used the FTMap server [31] to carry out CSM on the Ad2 E3-19K – HLA-A2 complex, and the individual HLA-A2 and Ad2 E3-19K proteins. FTMap is a close computational equivalent for X-ray crystallography or NMR-based screening experiments carried out to determine binding hot spots. The experimental methods are time-consuming in comparison to computational methods. 16 small organic molecules were used as probes. Ethane, ethanol, isopropanol, *tert*-butanol, acetonitrile, methanamine, *N,N*-dimethylformamide, dimethyl ether, benzaldehyde, benzene, cyclohexane, phenol, acetamide, acetone, acetaldehyde and urea were the 16 probes used. The PDB file was processed, and the probe molecules were docked on the protein surface after finding the most favorable position. The FTMap algorithm [31] produces bound positions using rigid body docking; in this step billions of probe positions are sampled on both translational and rotational grids using fast Fourier transform (FFT) correlation approach. The FFT algorithm speeds up calculations [31]. The positions are then scored based on Poisson-Boltzmann calculations (including attractive and repulsive VDW terms, electrostatic interaction energy, structure-based pairwise interaction potential and a cavity term that represents effects of non-polar enclosures [31]). Based on the calculation, 2,000 lowest energy poses of each of the 16 probes are retained, after which the CHARMM potential is used to minimize the energy [40]. The energy minimization is done using Analytical continuum electrostatics (ACE) model in association with CHARMM potential to explain the electrostatics and solvation [37]. ACE potential solves the Poisson equation to give an approximate potential with less computational efforts. The conformation consisting of minimized probes were then clustered with a 4 Å radius, with the lowest-energy structure first. Six clusters with lowest energies were retained for each probe, based on the ranking conducted using their Boltzmann averaged energies [31]. After this, the next step was to form consensus clusters

after clustering the clusters of different probes. A center of mass was specified for each of the different probe cluster in the step above, so the cluster with the maximum cluster centers within a radius of 4 Å was ranked as the highest consensus cluster [31]. The site comprising the cluster was called the Consensus Site (CS) [31].

The results are available after all the above steps are complete. It consisted of a PDB file containing the input structure, and the coordinates for the individual consensus cluster. The result also contained a PyMOL [38] session consisting of the protein structure, and the clusters colored individually, probes summary, nonbonded interactions and Hydrogen bond interactions of the protein residues with the small organic probe molecules. Nonbonded interactions defined the main and secondary hot spots, determined by the residue with the highest percentage of nonbonded interactions.

$$\% \; contact \; frequency \; of \; aa_i = \frac{Number \; of \; nonbonded \; contacts \; for \; aa_i}{Sum \; of \; nonbonded \; contacts \; for \; all \; aa} \times 100\%$$

The above formula is directly adapted from the paper, and was used to calculate the percent of contact frequency for individual amino acids (aa_i) where i = 1 .. n [31]. The strength (S) of the consensus clusters obtained from the FTMap algorithm, defined as the number of probe clusters within the consensus clusters should be greater than 16, for the site to be considered a hotspot. When the strength of the consensus cluster is greater than 16, and if there exists at least one more hot spot within 8 Å from the strong hotspot, the binding site is considered a potentially druggable site. The distance is measured from the from the center of mass for the two consensus clusters [13]. The site detected could be a potential binding site for small molecules. We use Ad2 E3-19K/HLA-A2 complex, and the HLA-A2 and Ad2 E3-19K individual proteins as input PDB's using the FTMap server, which performs the calculations.

2.4.2   Computational solvent mapping of the Ad2 E3-19K/HLA-A2 complex

The complex structure was obtained from the PDB ID 4e5x. We performed the computational solvent mapping for the complex to determine possible hotspot sites present on the binding interface. The FTMap server was used to perform CSM using the complex. We considered looking at both the CS and nonbonded interaction. The binding interface of Ad2 E3-19K and HLA-A2 have shallow pockets. Most of the CSs are present in the cavity on HLA-A2 protein as seen in Fig. 10.
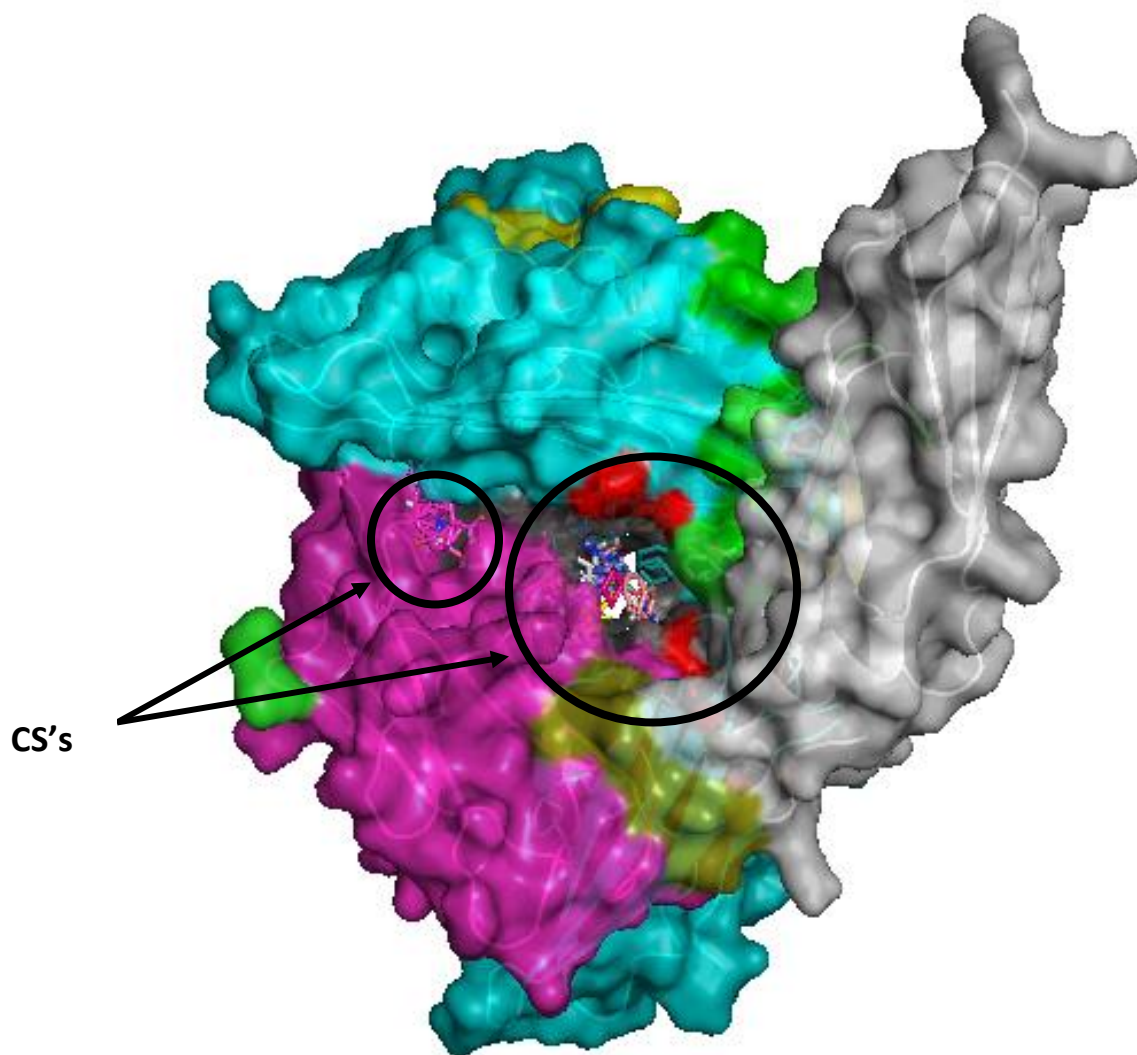
Figure 10 Consensus site clusters with the surface of the Ad2 E3-19K/HLA-A2 complex displayed

The consensus clusters can be seen in the cavity when the surface of the complex protein is displayed as shown above in Fig. 11. Two residues, Arg48 (HLA-A2 site 1) and Asp238 (HLA-A2 site 3) showed a significant amount of nonbonded interactions with the probe molecules. A percentage of the nonbonded interaction between protein residues and the probe molecule was calculated using the formula for percent contact frequency for the amino acid residue. Ranking the amino acids in the order of the highest percentage of nonbonded contact frequencies, we found two residues mentioned above.
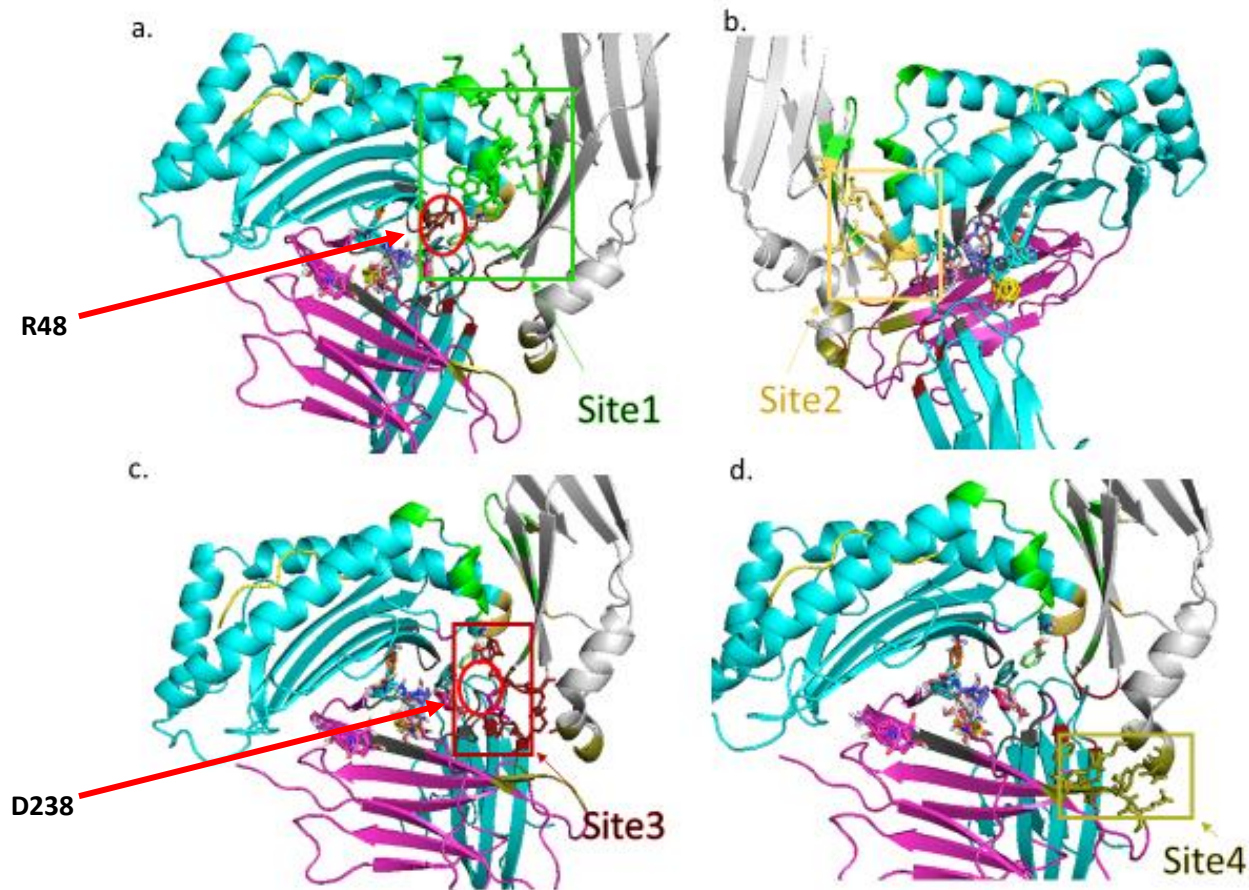
Figure 11 Displaying the side chains of the HLA-A2 (ribbon representation) residues present on the binding interface. Images were prepared using PyMOL.

As seen in Fig. 11 (a), the Arg48 residue of site 1 is present close to the probe clusters, allowing it to interact with them. In Fig. 8 (c) the Asp238 residue of site 2 is close to the probe clusters, thus associating with them through nonbonded interactions. Sites 2 and 4 are not close to the probe clusters.

2.4.3 Computational solvent mapping of HLA-A2

The computational solvent method again detected the cavity to be a potential binding site for small molecules. However, the site is more than 8 Å away from the binding interface, and our focus is to probe the binding interface with small molecules to determine its potential to bind compounds. Both the complex and HLA-A2 have a potential binding site in the cavity (Fig. 12), so we map the Ad2 E3-19K protein.
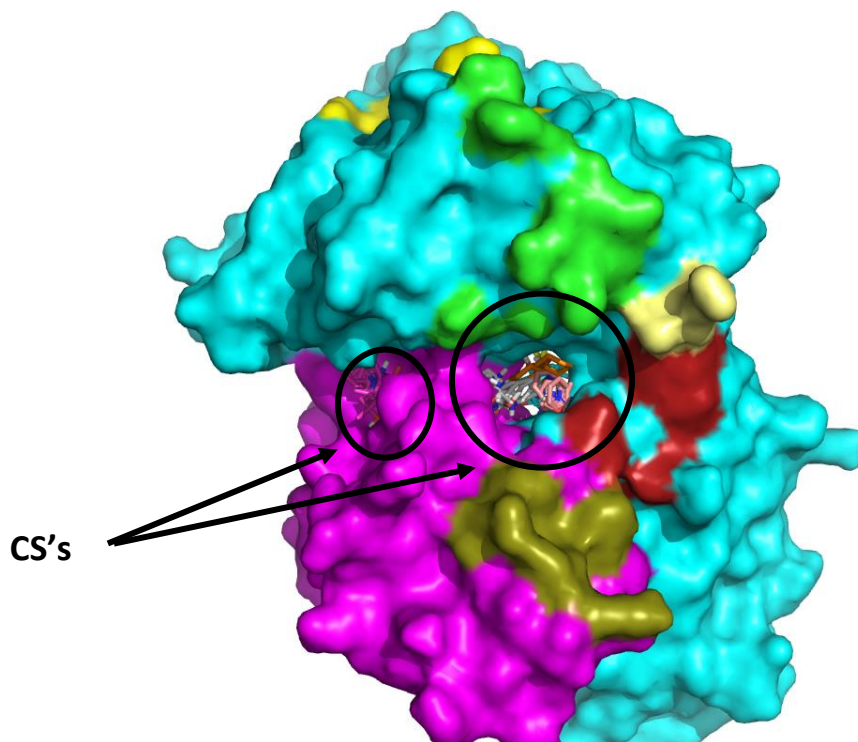

Figure 12 Consensus site clusters with the surface of the HLA-A2 displayed

2.4.4 Computational solvent mapping of Ad2 E3-19K

In Fig. 13, residues of sites 1 and 2 surround Consensus Site 6 (CS6), and site 1 residues surround Consensus Site 5 (CS5). CS5 includes of probe molecules acetonitrile, benzaldehyde, cyclohexane, acetamide, acetone, acetaldehyde, benzene, *N,N*-dimethylformamide, dimethyl ether, ethane, phenol, isopropanol, urea and butanol. They form hydrogen bonds, electrostatic interactions and hydrophobic interactions with the site 1 residues on Ad2 E3-19K protein. Lys45, Tyr 46 and Ala47. CS6 includes of acetaldehyde, benzaldehyde, *N,N*-dimethylformamide, urea, acetamide, methanamine, benzene, ethanol, and phenol. The probes formed hydrogen bonds, electrostatic and hydrophobic interactions with residues Lys42, Ile43, Gly44, Lys45, Tyr46, Ala47, Val48 and Tyr49 present at the binding interface. The percent of nonbonded interactions for these residues is limited, but it is still obvious. Fig. 14 shows a graph depicting the residues in order of the percentage of nonbonded interaction based on the ranking.
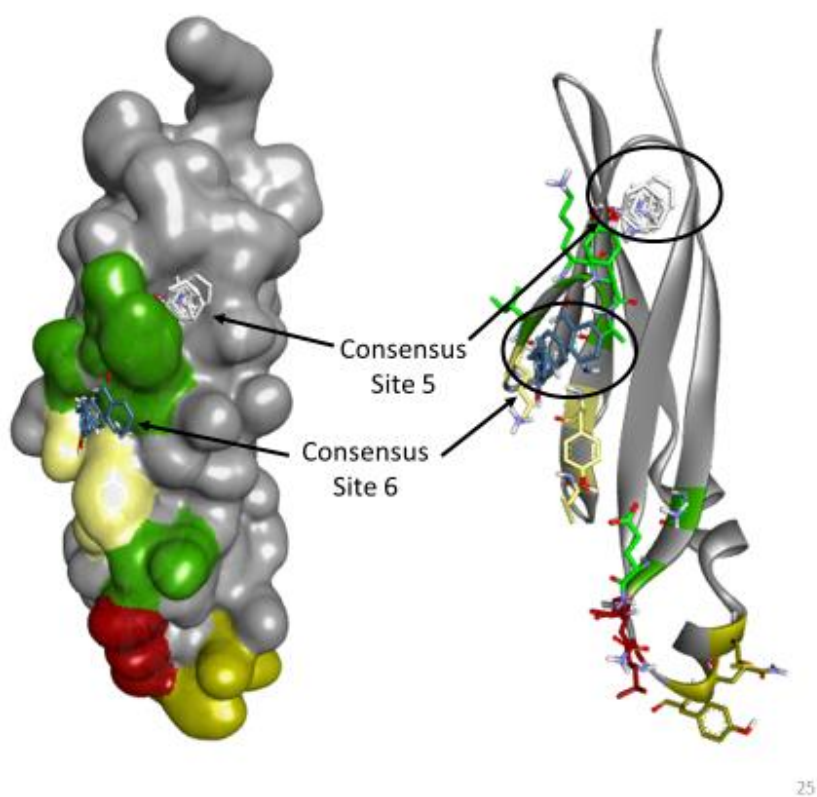
Figure 13 The surface and ribbon representation of Ad2 E3-19K protein is shown highlighting the consensus clusters CS5 and CS6
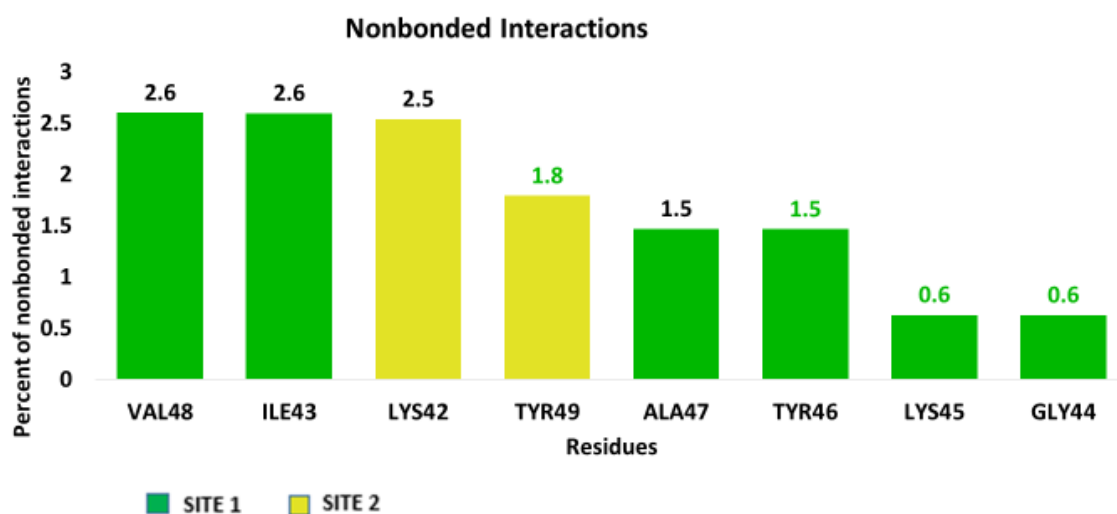

Figure 14 Graph of the Percent of nonbonded interactions vs Ad2 E3-19K residues

2.5   Summary of the Binding Surface Analysis

Viruses are continuously changing as they undergo mutation; they infect human cells; for example, Ads binds to the HLA-A2, blocking their egress to the surface of the cell. In an article written by Bjorn Carey on a way to combat viruses by targeting host cell proteins, Shirit Einav suggests that targeting the host cell protein instead of targeting the virus directly, since viruses live within our cells, while counting on the host cell machineries to replicate [39]. The aim of our research is to target the HLA-A2 (host) region containing sites 1 and 2 with chemical libraries. Based on the study conducted by Dr. Marlene Bouvier and her group [23] [25], Glu177 (HLA-A2) was found to be a hotspot residue. The per-residue decomposition confirms this, the binding free energy of Glu177 (HLA-A2) and Lys42 (Ad2 E3-19K) was from the electrostatic interaction. Computational solvent mapping for the complex and HLA-A2 identified the cavity present on the HLA-A2 molecule to be a favorable site, the CSM for Ad2 E3-19K also identified sites 1 and 2 as potential binding sites. Since our aim was to target the HLA-A2, targeting the counterpart of sites 1 and 2 on Ad2 E3-19K protein sounds reasonable. Thus, considering the results obtained from our structure and surface analysis we use HLA-A2 as a receptor to dock large chemical libraries on the region including both these sites. The next step would be conducting a high-throughput virtual screening on HLA-A2. While targeting the host cell, we need to keep in mind that the drug inhibiting the protein is safe, allowing the functions carried out by HLA molecules to be as close to the usual function. Also, it should not cause any toxicity to the HLA molecules.

CHAPTER 3
STRUCTURE-BASED VIRTUAL SCREENING

3.1 Introduction

Computational methods in drug discovery have emerged as a reliable and indispensable tool. It expedites the otherwise long process of drug discovery, making it a fast and more cost-effective tool. Virtual Screening is one of the key areas of computer-aided drug discovery. Virtual screening is the *in silico* equivalent to high-throughput screening (HTS), screening a large compound database using computational methods to prioritize compounds for experimental testing. Structure-based and ligand-based virtual screening are the two methods used in Virtual Screening. The *structure-based virtual screening (SBVS)* approach uses the knowledge of the three-dimensional structure of the target protein. Molecular docking is a SBVS approach, where the ligand and protein alter their conformations to find the 'best-fit', referred to as the 'induced-fit' effect. *Ligand-based virtual screening (LBVS)* utilizes known-compounds that bind to the target of interest as a query to search databases for identifying potential compounds. LBVS methods like substructure or similarity search, pharmacophore search and electrostatic or shape search have been developed and utilized for hit identification.

The knowledge of the crystal structure of the Ad2 E3-19K/HLA-A2 complex provides a solid foundation for structure-based virtual screening. Since there is no known-compound that inhibits the complex, we moved forward with the SBVS. We first screened the Prestwick chemical library, then moved on to larger chemical libraries. The primary objective is to screen the chemical libraries to identify novel scaffolds that inhibit the complex.

3.2 Molecular Docking using Prestwick Chemical Library

The Prestwick chemical library consists of 1200 compounds, used for virtual screening to reduce the number of compounds to a manageable number to be tested experimentally. We used the crystal structure obtained from the PDB code 4E5X [23]. The crystal structure is a dimer; we used a monomer for carrying out the docking. The 'Protein Preparation Wizard' available in the Schrödinger Suite 2011 [32] was used to prepare the protein structure (HLA-A2) for SBVS. Water molecules were removed. Hydrogens were added to the receptor protein; the OPLS 2005 forcefield was used to add charges. We used restrained minimization to refine the hydrogens and the entire protein. Next, the 'LigPrep' of the Schrödinger Suite 2011 [32] was used to prepare the Prestwick chemical libraries. The EPIK software in Schrödinger was used to generate multiple ionization states and tautomeric forms at pH 7.4 ± 0 (Human body pH), We retained the stereoisomers for compounds containing chiral centers.

We utilized GOLD v5.2.2[33] to conduct molecular docking, GOLD being one of the most cited programs. GOLD uses genetic algorithm (GA), accounting for the flexibility of the ligand for it to fit into the binding site. The binding site sphere was defined within 10 Å, thus allowing us to include sites 1 and 2 to be used as a binding pocket to dock the compounds. ChemPLP was used as a scoring function; it utilizes ChemScore's hydrogen bonding term along with the multiple linear potential used to model repulsive terms and VDW's. ChemScore takes account of the hydrophobic interactions contact area, hydrogen bonding, metal interaction and ligand flexibility. 10 GA runs were conducted for each compound. Default settings were used for the fitness and search options.

The results obtained from the docking were ranked in descending order based on the scoring function. The compound selection was done based on the scoring function, starting with

the docking pose analysis of the highest scoring function and the interaction of the compound with the receptor. Glu177 being an important residue, we prioritized the ligand to interact with it by forming a hydrogen bond with it. A set of 44 compounds fulfilling the criteria was shortlisted for experimental tests. Fig. 15 shows the docking poses of Raloxifene (ChemPLP = 57.3) and Trazodone (ChemPLP = 47.2), which showed binding affinity by the experiments conducted by Dr. Hui Deng. As seen in Fig. 15 they can form electrostatic interactions with Glu177 residue, and form hydrogen bonds and hydrophobic interactions with residues of sites 1 and 2. The residues are highlighted in green for site 1 and pale yellow for site 2. The ones not on the binding interface just have their side chains displayed.
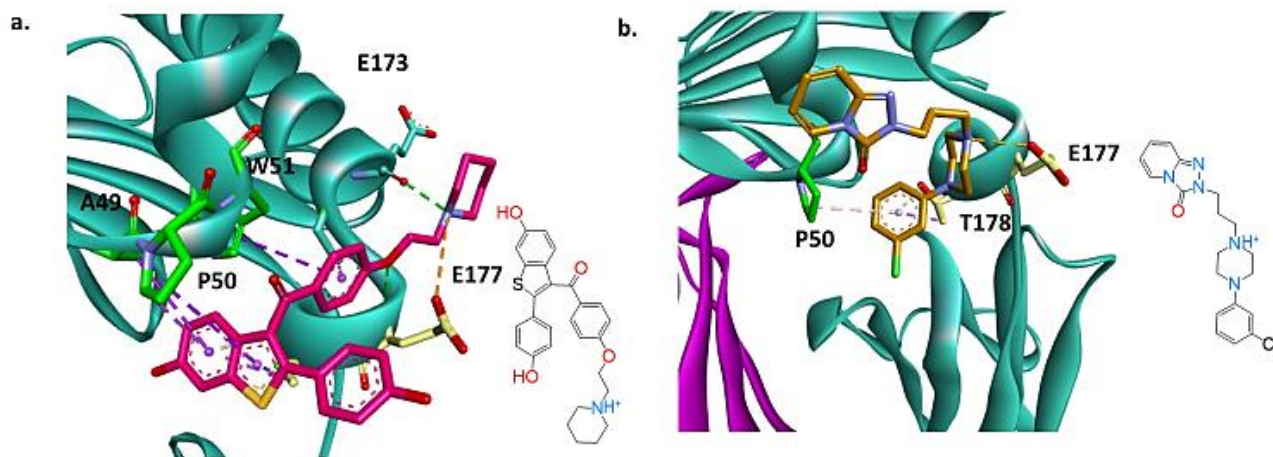
Figure 15 Docking poses of Raloxifene (a) and Trazodone (b), with their respective structures. Nitrogen, oxygen and Sulphur atoms are colored in blue, red and yellow respectively. Water molecules are shown as red spheres. Hydrogen bonds, charged interactions and hydrophobic contacts are indicated by green, red and purple dashed lines, respectively. Images were prepared using Discovery Studio 2016.

3.3   Consensus Docking with Chembridge, LifeChemicals and Specs

The libraries Chembridge, LifeChemicals and Specs in total consisted of 1,330,288 compounds to be screened. First, we needed to identify and eliminate compounds that could

potentially be problematic. We used computational filtering methods to do so, thus removing compounds that could have led to false positives in the assays to be conducted. The inbuilt filters of Canvas [32] were used to eliminate the compounds. Second, the potential binding site being targeted i.e. sites 1 and 2 are electronegative in nature. The compounds were clustered based on the binary fingerprint of the structure, and then the compounds were selected which would complement the electronegative nature of the site. Third, after narrowing down the compounds to a reasonable number, we would carry-out docking using Gold, Autodock and Surflex-Dock. Fourth, the results obtained were compared and analyzed to pick out compounds to be subjected to Molecular Dynamic (MD) simulations to visualize the interaction of the receptor (HLA-A2) and small molecule complex in a solvent mediated environment.

3.3.1 Filtering chemical libraries

Canvas (Schrödinger, LLC) was used to filter the screening libraries using the inbuilt Qikprop descriptors, Rapid Elimination of Swill (REOS) and pan-assay interference compounds (PAINS) to remove the compounds, that could later be problematic. Qikprop predicts ADME properties by generating descriptors relevant to the compounds. #stars are the overall score representing the ADME compliance as well as the parameter of drug likeness. The QikProp descriptor #stars denote the number of property descriptors falling outside the optimum range of values for 95% of the known drugs. The inbuilt REOS filter in Canvas flags compounds containing functional groups that are reactive, toxic or other problem causing structures. PAINS are compounds showing activity across a range of assays for different proteins, resulting in false positives. The PAINS filter in Canvas is used to eliminate the compounds during the computational filtering. Overall the role of all three filters used is to eliminate compounds that would cause any

hindrance when conducting an assay to test the binding activity of the compound. 177,545 compounds were eliminated in this step. The rest were further clustered using K-means clustering.

3.3.2   Ligand Clustering and Selection

Binary fingerprints of the compounds were generated using MOLPRINT 2D, a molecular fingerprint method. K-means clustering was used to form a thousand clusters. K-means clustering is a partitioning method, that can cluster for a larger data set. K-means clustering partitions the data into k (1000 in this study) mutually exclusive clusters based on the information obtained from the binary fingerprints of the compounds, and the compounds with nearest binary fingerprints are placed in the same cluster. The compounds in each cluster varied depending on the binary fingerprint properties; we picked around 10,000 compounds for conducting VS. The 10,000 selection was done based on the presence of an amine group, the reason being the presence of an electronegative region on the receptor surface, and presence of the Glu177 hotspot residue.

3.3.3   Preparation and Docking

The small molecule preparation and protein preparation were done using 'LigPrep' and 'Protein Preparation Wizard' in Schrödinger's software, respectively. The same parameters were used as the preparation as described in section 3.2. Docking was carried out using GOLD, Autodock [34] and Surflex-Dock [35]. Autodock is also one of the most cited docking software program. Autodock uses a Lamarckian genetic algorithm and an empirical free energy scoring function, per the manual. Surflex-Dock uses empirical scoring function as well. Autodock and Surflex-Dock use energy scoring functions, thus we used the MMGBSA in Prime [32] to calculate the scoring function of the receptor (HLA-A2)/small molecule complex after it was docked using GOLD.

The binding site sphere was defined within 10 Å, to include the sites 1 and 2 to be used as a binding pocket to dock the 10,000 compounds. 10 GA runs were run for each ligand in GOLD, the default settings were used to conduct the molecular docking. The best docking poses were extracted for the HLA-A2 and small molecule complex, the MMGBSA in Prime was used to measure the energy of the complex; the gold docking poses are selected; default MMGBSA calculation settings were used. Autodock v4.2.6 was used to dock the 10,000 compounds. The default parameters were used, we specified the sites to conduct the docking (Sites 1 and 2); 75,000 energy evaluations; 5 GA runs per ligand. The results are sorted by the lowest energy. For Surflex-Dock the binding site was defined by generating the residue based protomol, including residues at sites 1 and 2. The docking was run using default settings, and the Surflex-Score function was used to rank the docked poses.

3.3.4   Results of the Consensus Docking

The results obtained from Autodock, Surflex-Dock and MMGBSA were sorted by the lowest energy. The top 1000 docking results were extracted and compared to find common compounds. As seen in Fig. 16 we selected a total of 406 compounds to study the best docking pose, and further shortlist it to around 50. Out of the 406 compounds, 13 were common for all three. A total of 54 compounds were selected based on best docking poses, including 13 common compounds. MD simulations were then carried out using the 54 compounds.
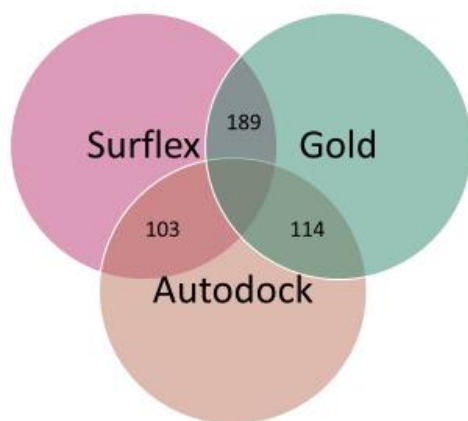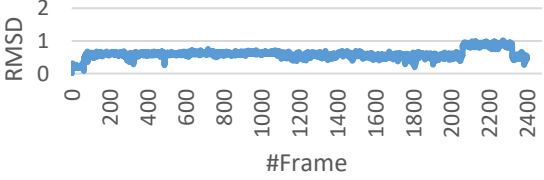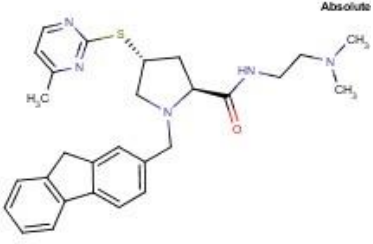
Figure 16 Venn Diagram of the docking results common between each docking software
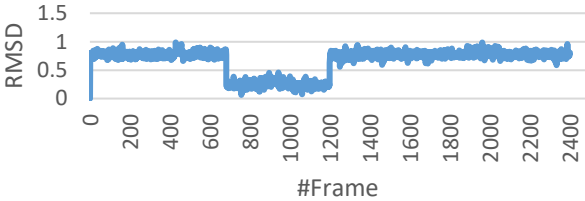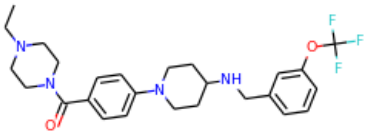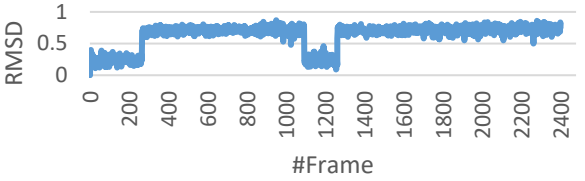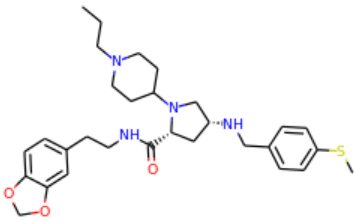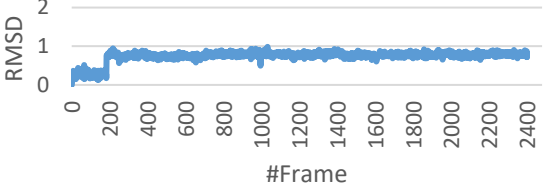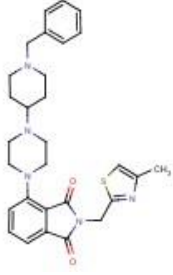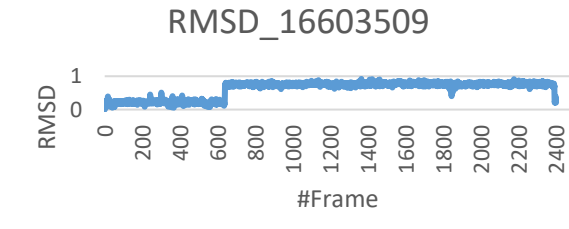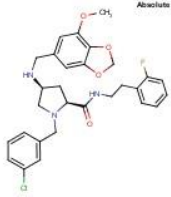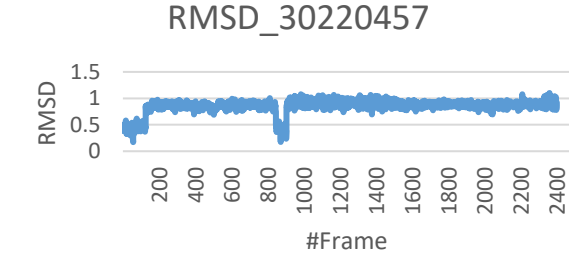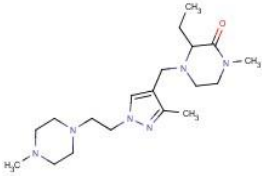
3.3.5 Molecular Dynamic Simulations
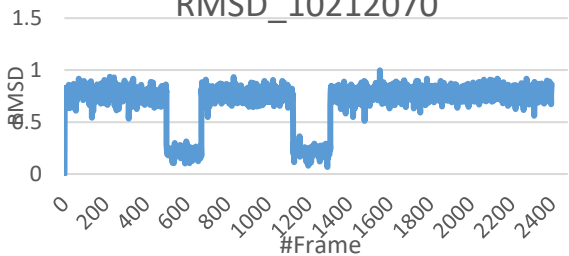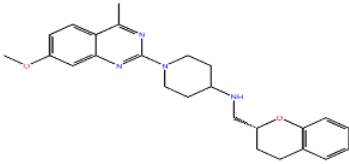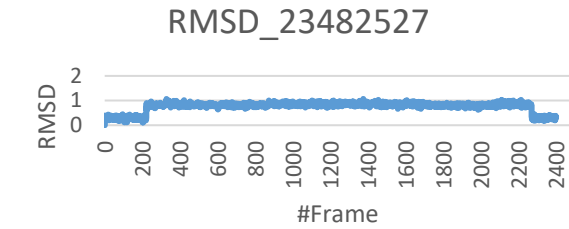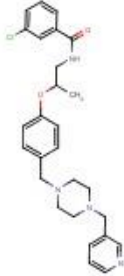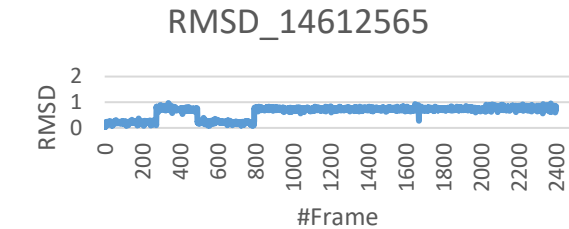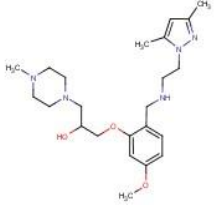
The molecular dynamics (MD) simulations were performed by using the AMBER14 suite of programs [26]. The ff99SB forcefield parameters were used for the protein. The system was neutralized by the addition of chloride ions. We used antechamber from the AmberTools [26], using the 'general AMBER force field' (GAFF) designed for pharmaceutical molecules. The traditional AMBER force fields and GAFF are compatible in a way that both can be mixed during the simulation. Each of the protein-ligand (ligand being the small molecule) complexes was solvated using a 10 Å cubic octahedral box of TIP3P water molecules. We used the SHAKE algorithm to constrain all the bonds, including the hydrogen bonds. For long range electrostatics particle mesh Ewald (PME) was used by the AMBER MD package. To calculate the interactions, the long range electrostatic cut-off distance was kept at 10 Å. Once again, 10,000 steps of conjugate gradient minimizations were used to minimize the system. The minimization was followed by MD equilibration for 100 picoseconds (ps) over constant volume with weak restraints on the solvated system. The system was gradually heated over approximately 20 ps from 0K to 300K. Now that our system has been heated over constant volume, we ran a 100 ps equilibration
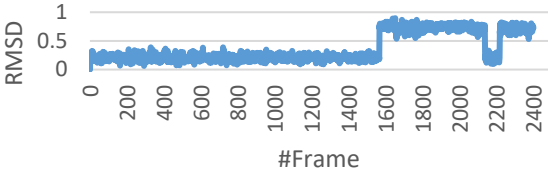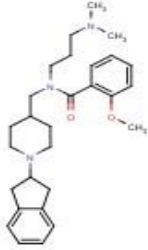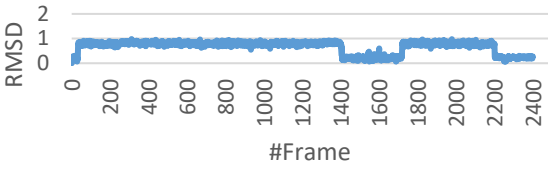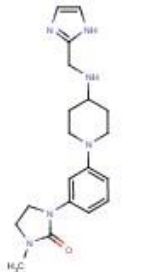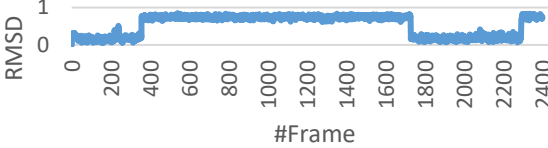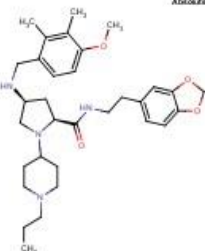
over constant pressure to relax the density of the water. The system was already heated to 300K, this will be maintained in the second equilibration. A 6 nanosecond production runs was then performed, recording coordinates every 2.5 picoseconds. A total of 2400 frames or snapshots were extracted. In Table 3 the RMSD of the complex backbone atoms relative to the starting complex structure was used to assess the stability of each complex. We also used Visual Molecular Dynamics (VMD) [39] to visualize the MD trajectories, to see the interaction of the protein-ligand surrounded by a solvent box.
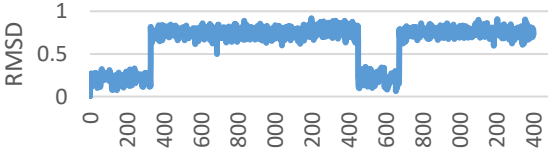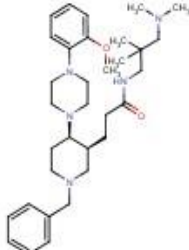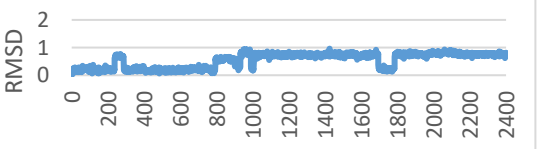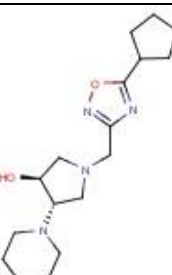
3.3.6   Results of the Molecular Dynamics Simulations

Given below in Table 3 are the 15 compounds with the RMSD graph, chemical structure and the MMGBSA energy Δ in kcal/mol, ranked in order of their alchemical free energy (lowest to highest). Though the RMSD for the 15 complexes fluctuates, the RMSD for most of the complexes seems to converge. Fig. 17 shows the docking poses for the top two compounds ranked based on their alchemical free energy calculations. As seen in Fig. 17 they form electrostatic interactions with the Glu177 residue, and form hydrogen bonds and hydrophobic interactions with residues of sites 1 and 2. The residues are highlighted in green for site 1 and pale yellow for site 2. The ones not on the binding interface just have their side chains displayed.

RMSD	Chemical Structure	MMGBSA Energy Calculation Delta Total (Avg ± SD)
RMSD_1154916		-35.43 ± 3.01
RMSD_44362439		-35.19 ± 2.41
RMSD_48514691		-34.58 ± 2.70
RMSD_15212815		-31.57 ± 3.15

RMSD_10212070		-27.36 ± 2.07
RMSD_23482527		-27.09 ± 3.08
RMSD_14612565		-26.97 ± 2.16
RMSD_16603509		-26.02 ± 2.18
RMSD_30220457		-25.52 ± 2.16

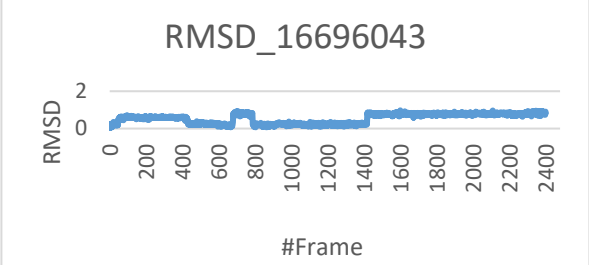|		
RMSD_31626394		-17.43 ± 1.66
RMSD_88331399		-16.21 ± 2.26
RMSD_43799825		-15.94 ± 2.17
RMSD_20841963		-14.85 ± 3.04
RMSD_1776159		-14.08 ± 2.17
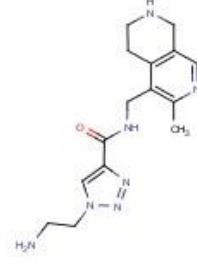
RMSD_16696043	
	-2.08 ± 1.81

Table 3. RMSD graphs and the binding free energy of the complex
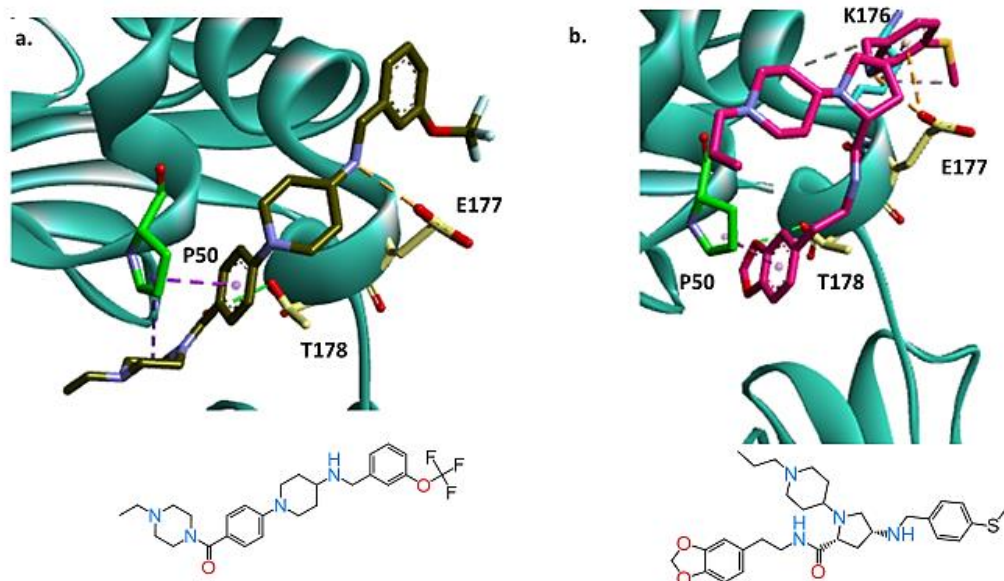
Figure 17 Docking poses of the compounds with alchemical free energy of 35.43 kcal/mol (a) and 35.19 kcal/mol (b). Nitrogen, oxygen and Sulphur atoms are colored in blue, red and yellow respectively. Water molecules are shown as red spheres. Hydrogen bonds, charged interactions and hydrophobic contacts are indicated by green, red and purple dashed lines, respectively. Images were prepared using Discovery Studio 2016.

3.4 Surface Plasmon Resonance Experiment conducted by Dr. Hui Deng and Dr. Hyun Lee

36 Prestwick compounds were tested for their binding affinity, with 7 showing significant binding activity. The Q56G HLA-A2 mutant was used in the Ad2 E3-19K/HLA-A2 complex, attached to the sensor chip of the SPR machine. The sample containing the second interaction

partner (Prestwick Chemical compounds) was passed over the surface of the chip. The changes in angle of the reflected light is measured in real time. K_D the binding affinity is also obtained from this experiment. The graph in Fig. 18 displays the binding affinity of the 7 compounds. We have also shown the docking pose and interactions for two compounds, Raloxifene and Trazodone in Fig. 16.
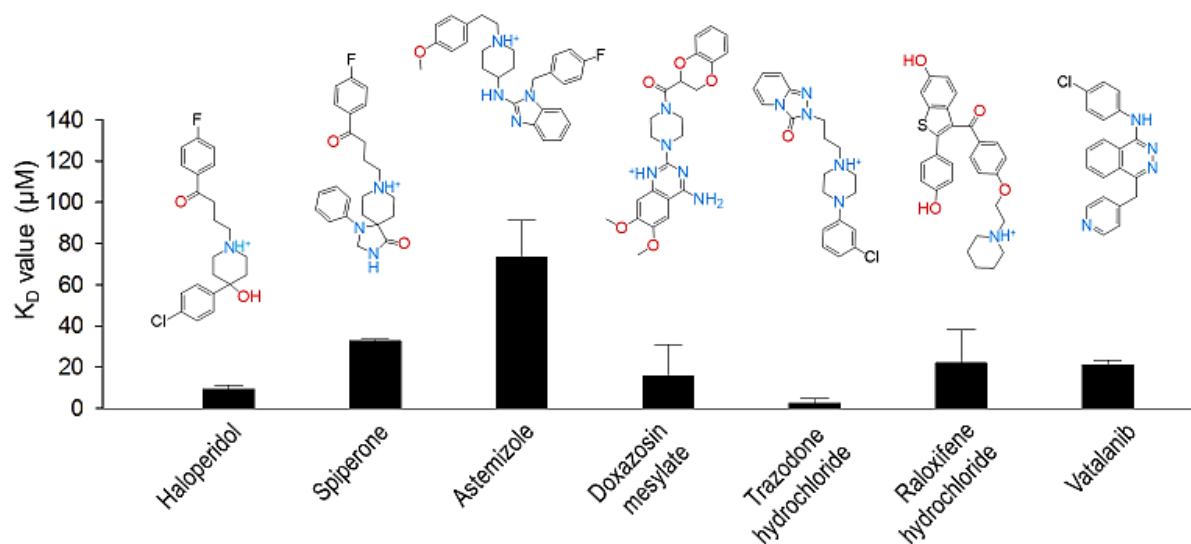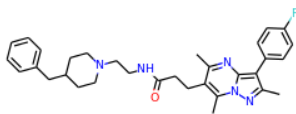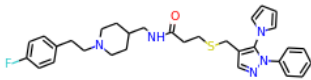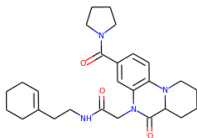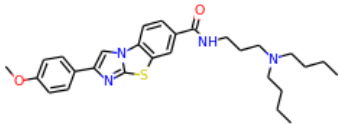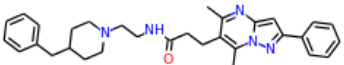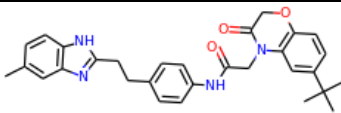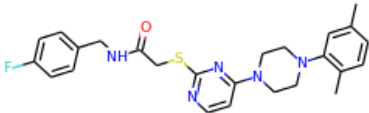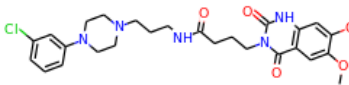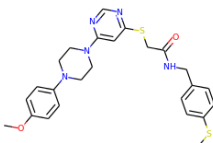
Figure 18 Graph of the binding affinity (K_D) vs Prestwick compounds

3.5   Protein-Protein Interaction (PPI) library

The PPI library consisted of 10,000 compounds. With the predicted Prestwick chemical library compounds showing significant binding activity, we decided to use Gold for docking and scoring the PPI library. The method described in section 3.2 above, was used for the PPI library. As can be seen in the top two docking poses of the PPI library compounds are larger, and hence a part of the compound binds in the cavity (Fig. 10 and Fig. 11) on HLA-A2 as shown in the Fig. 16. The top 100 compounds are selected to analyze the binding activity using SPR. Shown below

(Fig. 19) are the docking poses of the top two compounds with the highest ChemPLP score of

92.5 and 90.5 respectively.

Chemical ID	ChemPLP	Structure
C561-0793	92.45	
C798-1126	90.05	
G673-0590	89.78	
C797-0939	89.2	
C561-0670	88.17	
F255-0610	87.2	
M459-5188	86.18	
C191-0124	86.06	
M179-5512	85.89	

C741-0401	85.23	

Table 4 Top 10 docking scores for the PPI Library
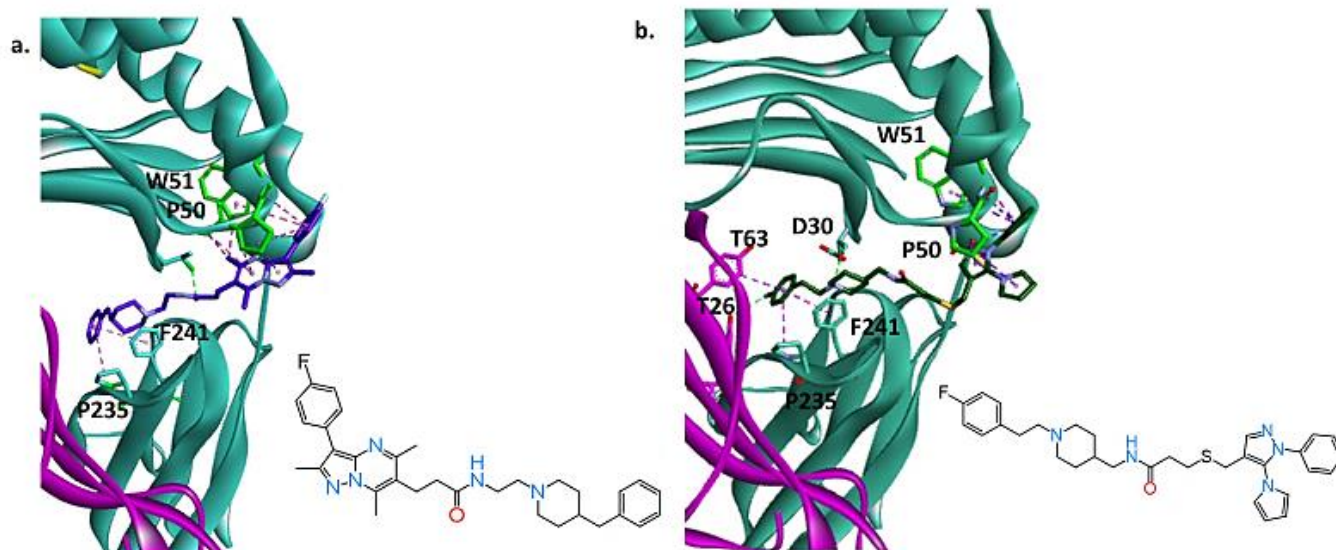
Figure 19 Top two docking poses based on the ChemPLP for the PPI library. ChemPLP 92.5 1(a) and 90.5 1(b), respectively. Nitrogen, oxygen and Sulphur atoms are colored in blue, red and yellow respectively. Water molecules are shown as red spheres. Hydrogen bonds, charged interactions and hydrophobic contacts are indicated by green, red and purple dashed lines, respectively. Images were prepared using Discovery Studio 2016.

CHAPTER 4

SELECTION OF RESIDUES FOR A PEPTIDE TO BE USED AS A POTENTIAL INHIBITOR

4.1   Peptide residue selection

The per-residue decomposition result was used to determine the sequence of a peptide, that could be used as a potential inhibitor of the Ad2 E3-19K – HLA-A2 interaction. The peptide selection is done by looking at the per-residue decomposition of the binding interface of both proteins, keeping in mind to select limited residues for further synthesis. From Fig. 4 and Fig. 5 we see that the residues on sites 1, 2 and 4 are significant for HLA-A2, but residues at sites 2 and 4 are significant for the Ad2 E3-19K protein. Selecting the residues in sequence is to be done for both the proteins. So, we decided to use the residues at sites 2 and 4 for this analysis. We selected residues which are in a sequence in the HLA-A2 molecule and the E3-19K protein of Ad2, inclusive of the residues with significant binding free energy. The peptides were determined such that they would have a counterpart i.e. HLA-A2 peptide to be used as a potential inhibitor against the E3-19K protein of Ad2 and vice-versa. Given below are the residues proposed for synthesizing the peptides:

1. ENGKET (173-178 of HLA-A2) residues: This includes the Glu177 residue (hotspot) and the residues on either side of Glu177 at site 2.

2. KIGKYAVYAI (42-51 of AD2) residues: We selected residues with significant binding free energy; the residues in this peptide includes site 1 residues as well, since the residues of site 2 Lys42 and Tyr49 are separated by the residues in site 1.

3. HPAENGKSNF (13-22 of HLA) residues: The binding free energy for the residues at site 4 were prominent, hence we considered these residues for forming a peptide.

4. MYMSKQYKL (88-95 of AD2) residues: The binding free energy of Met89 and Tyr93 were evident, both these residues and Gln92 are highly conserved in the Adenovirus types 2 and 4. They interact with the same residues on the HLA-A2 receptor. Thus, the sequence had to include the Met89, Gln92 and Tyr93 in the peptide.

CONCLUSION

Adenoviruses as such do not have a specific therapeutic treatment. Using computational methods for drug discovery methods to find novel scaffolds was the aim of this study. We decided to target the HLA-A2 molecule, since viruses keep mutating. We studied the binding interface of the Ad2 E3-19K and HLA-A2 molecules to determine a binding site to conduct the docking. This complex does not have a reference ligand, and so we opted for a SBVS method to find potential inhibitors of the complex using chemical libraries. The Glu177 hotspot residue from the crystal structure study correlated with the per-residue decomposition conducted by computational methods. The Ad2 E3-19K and HLA-A2 binding interface is divided into 4 sites, out of which we conducted a SBVS covering sites 1 and 2 of the HLA-A2 molecule. Glu177 is present on site 2 of the HLA-A2 molecule. Through SPR, 7 Prestwick chemical library compounds out of 36 showed significant binding affinity to the Q56G HLA-A2 mutant bound to Ad2 E3-19K protein. Further experimental tests are to be performed on the larger libraries screened, rescored using MD simulations; and the protein-protein interaction libraries to measure the binding affinity of these compounds to the HLA-A2 receptor.

CITED LITERATURE

1. Fu, J., Li, L. & Bouvier, M. Adenovirus E3–19K proteins of different serotypes and subgroups have similar, yet distinct, immunomodulatory functions towards major histocompatibility class I molecules. *J. Biol. Chem.* **286**, 17631–17639 (2011).

2. Ginsberg, H.S. *et al.* Role of early region 3 (E3) in pathogenesis of Adenovirus disease. *Proc. Natl. Acad. Sci. USA* **86**, 3823–3827 (1989).

3. Andersson, M., Paabo, S., Nilsson, T. & Peterson, P.A. Impaired intracellular transport of class I MHC antigens as a possible means for adenoviruses to evade immune surveillance. *Cell* **43**, 215–222 (1985).

4. Andersson, M., McMichael, A. & Peterson, P.A. Reduced allorecognition of adenovirus-2 infected cells. *J. Immunol.* **138**, 3960–3966 (1987).

5. Burgert, H.G. & Kvist, S. An adenovirus type 2 glycoprotein blocks cell surface expression of human histocompatibility class I antigens. *Cell* **41**, 987–997 (1985).

6. Hierholzer, J. C. 1992. Adenoviruses in the immunocompromised host. Clin. Microbiol. Rev. 5: 262–274.

7. Flomenberg, P., J. Babbitt, W. R. Drobyski, R. C. Ash, D. R. Carrigan, G. V. Sedmak, T. McAuliffe, B. Camitta, M. M. Horowitz, N. Bunin, and J. T. Casper. 1994. Increasing incidence of adenovirus disease in bone marrow transplant recipients. J. Infect. Dis. 169: 775–781.

8. Zahradnik, J. M., M. J. Spencer, and D. D. Porter. 1980. Adenovirus infection in the immunocompromised patient. Am. J. Med. 68: 725–732.

9. Krilov, L. R., L. G. Rubin, M. Frogel, E. Gloster, K. Ni, M. Kaplan, and S. M. Lipson. 1990. Disseminated adenovirus infection with hepatic necrosis in patients with human immunodeficiency virus infection and other immunodeficiency states. Rev. Infect. Dis. 12: 303–307.

10. Kosulin, K., E. Geiger, A. Vecsei, W. D. Huber, M. Rauch, E. Brenner, F. Wrba, K. Hammer, A. Innerhofer, U. Potschger, et al. 2016. Persistence and reactivation of human adenoviruses in the gastrointestinal tract. Clin. Microbiol. Infect. 22 (4): 381.e1–8. doi: 10.1016/j.cmi.2015.12.013.

11. Wadell, G. 1984. Molecular epidemiology of human adenoviruses. Curr. Top. Microbiol. Immunol. 110: 191–220.

12. Berk, A. J. 2007. Adenoviridae: the viruses and their replication. In Fields Virology, 5th Ed. D. M. Knipe, and P. M. Howley, eds. Lippincott Williams & Wilkins, Philadelphia, p. 2355–2394.

13. Hage, E., U. Gerd Liebert, S. Bergs, T. Ganzenmueller, and A. Heim. 2015. Human mastadenovirus type 70: a novel, multiple recombinant species D mastadenovirus isolated from diarrhoeal faeces of a haematopoietic stem cell transplantation recipient. J. Gen. Virol. 96: 2734–2742.

14. Burgert, H.-G. & Kvist, S. The E3/19K protein of adenovirus type 2 binds to the domains of histocompatibility antigens required for CTL recognition. *EMBO J.* **6**, 2019–2026 (1987).

15. Flomenberg, P., Piaskowski, V., Truitt, R.L. & Casper, J.T. Human adenovirus-specific CD8+ T-cell responses are not inhibited by E3–19K in the presence of gamma interferon. *J. Virol.* **70**, 6314–6322 (1996).

16. Rawle, F.C., Tollefson, A.E., Wold, W.S.M. & Gooding, L.R. Mouse anti-adenovirus cytoxic T lymphocytes. *J. Immunol.* **143**, 2031–2037 (1989).

17. Feuerbach, D. *et al.* Identification of amino acids within the MHC molecule important for the interaction with the adenovirus protein E3/19K. *J. Immunol.* **153**, 1626–1636 (1994).

18. Flomenberg, P., Gutierrez, E. & Hogan, K.T. Identification of class I MHC regions which bind to the adenovirus E3–19K protein. *Mol. Immunol.* **31**, 1277–1284 (1994).

19. Gabathuler, R., Levy, F. & Kvist, S. Requirement for the association of adenovirus type 2 E3–19K wild-type and mutant proteins with HLA antigens. *J. Virol.* **64**, 3679–3685 (1990).

20. Hermiston, T.W., Tripp, R.A., Sparer, R.A., Gooding, L.R. & Wold, W.S.M. Deletion mutation analysis of the adenovirus type 2 E3-gp19K protein: identification of sequences within the endoplasmic reticulum lumenal domain that are required for class I antigen binding and protection from adenovirus-specific cytotoxic T lymphocytes. *J. Virol.* **67**, 5289–5298 (1993).

21. Cox, J.H., Bennink, J.R. & Yewdell, J.W. Retention of adenovirus E19 glycoprotein in the endoplasmic reticulum is essential to its ability to block antigen presentation. *J. Exp. Med.* **174**, 1629–1637 (1991).

22. Pääbo, S., Bhat, B.M., Wold, W.S.M. & Peterson, P.A. A short sequence in the COOH-terminus makes an adenovirus membrane glycoprotein a resident of the endoplasmic reticulum. *Cell* **50**, 311–317 (1987).

23. Li, L., Y. Muzahim, and M. Bouvier. 2012. Crystal structure of adenovirus E3-19K bound to HLA-A2 reveals mechanism for immunomodulation. Nat. Struct. Mol. Biol. 19: 1176–1181.

24. Fu, J., Li, L. & Bouvier, M. Adenovirus E3–19K proteins of different serotypes and subgroups have similar, yet distinct, immunomodulatory functions towards major histocompatibility class I molecules. *J. Biol. Chem.* **286**, 17631–17639 (2011).

25. Li, L., Santarsiero, B. D., & Bouvier, M. (2016). Structure of the Adenovirus Type 4 (Species E) E3-19K/HLA-A2 Complex Reveals Species-Specific Features in MHC Class I Recognition. *The Journal of Immunology,197*(4), 1399-1407.

26. D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu and P.A. Kollman (2014), AMBER 14, University of California, San Francisco.

27. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis *J Comput Chem.* 2004 Oct;25(13):1605-12.

28. Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of medicinal chemistry*, *28*(7), 849-857.

29. Mattos, C., & Ringe, D. (1996). Locating and characterizing binding sites on proteins. *Nature biotechnology*, *14*(5), 595-599.

30. Allen, K. N., Bellamacina, C. R., Ding, X., Jeffery, C. J., Mattos, C., Petsko, G. A., & Ringe, D. (1996). An experimental approach to mapping the binding surfaces of crystalline proteins. *The Journal of Physical Chemistry*, *100*(7), 2605-2611.

31. Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., ... & Vajda, S. (2015). The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nature protocols*, *10*(5), 733-755.

32. Maestro, P. P. W. (2012). Schrödinger LLC. *New York*.

33. Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, *267*(3), 727-748.

34. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *J. Computational Chemistry* 2009, **16**: 2785-91.

35. Jain, A. N. (2003). Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of medicinal chemistry*, *46*(4), 499-511.

36. Wei, B.Q., L.H. Weaver, A.M. Ferrari, B.W. Matthews, And B.K. Shoichet, *Testing A flexible--- receptor docking algorithm in a model binding site.* J Mol Biol, 2004. **337**(5): pp. 1161---82.

37. Schaefer, M., & Karplus, M. (1996). A comprehensive analytical treatment of continuum electrostatics. *The Journal of Physical Chemistry*, *100*(5), 1578-1599.

38. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

39. Carey, B. (1970, October 07). A way to combat viruses by targeting host-cell proteins. Retrieved May 20, 2017, from http://med.stanford.edu/news/all-news/2013/10/a-way-to-combat-viruses-by-targeting-host-cell-proteins.html

40. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, *4*(2), 187-217.

41. Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", *J. Molec. Graphics*, 1996, vol. 14, pp. 33-38.

42. Cox, J. H., J. R. Bennink, and J. W. Yewdell. 1991. Retention of adenovirus E19 glycoprotein in the endoplasmic reticulum is essential to its ability to block antigen presentation. J. Exp. Med. 174: 1629–1637.

VITA

NAME: Nikita Raymond Dsouza

EDUCATION: Bachelor's in Pharmacy, Mumbai University, Mumbai, India 2013.

M.S., Bioinformatics, University of Illinois at Chicago, Chicago, Illinois, 2017 (Present).

RESEARCH EXPERIENCE: Research Assistant at the Center for Biomolecular Sciences, University of Illinois at Chicago, Chicago, Illinois, 2015 - present.

Advisor: Prof. Michael E. Johnson.