Theoretical Aspects of Designing New Vaccines for Breast Cancer: Docking Studies of Peptide/HLA-A2.1 Complexes

Tugba Kucukkal

Follow this and additional works at: https://dsc.duq.edu/etd

Recommended Citation
THEORETICAL ASPECTS OF DESIGNING NEW VACCINES FOR BREAST CANCER:
DOCKING STUDIES OF PEPTIDE/HLA-A2.1 COMPLEXES

A master’s thesis

Presented to the Bayer School of Natural and Environmental Sciences, Duquesne University

In partial fulfillment of the requirements for the degree of Master of Science

by

Tugba G. Kucukkal

July 2003
Name: TUGBA GUL KUCUKKAL

Thesis Title: THEORETICAL ASPECTS OF DESIGNING NEW VACCINES FOR BREAST CANCER: DOCKING STUDIES OF PEPTIDE/HLA-A2.1 COMPLEXES

Degree: MASTER OF SCIENCE

APPROVED

Jeffrey D. Evanseck
Associate Professor of Chemistry

APPROVED

Charles Dameron
Associate Professor of Chemistry

APPROVED

Jeffry D. Madura
Associate Professor of Chemistry
Chair, the Department of Chemistry and Biochemistry

APPROVED

David W. Seybert
Professor of Chemistry
Dean, the Bayer School of Natural and Environmental Sciences
HER2/neu is a transmembrane glycoprotein that is overexpressed in many tumors, including ovarian and breast cancers. The HER2/neu peptide IISAVGIL (GP2) is recognized by tumor-specific cytotoxic T lymphocytes in the context of human class I major histocompatibility complex (MHC) HLA-A2.1. One limiting-factor for using GP2 as a tumor vaccine is its poor affinity for HLA-A2.1, even though it has the correct peptide-binding motif. The research aims are to develop an accurate docking method for the binding of GP2 to HLA-A2.1, to understand the molecular forces that give rise to strong binding, and to predict mutations that lead to new tumor vaccines. AutoDock and GOLD have been used for docking calculations. The binding free energies from AutoDock correlate qualitatively with experiment. The docked structures for 14 ligands from Autodock3 are in good agreement with experiment. However, the ligands are not fully flexible. GOLD allowed full flexibility to reproduce experimental GP2 structure.
Acknowledgements

I would like to thank Professor Jeffrey D. Evanseck for his guidance and support during the past years. I am also very thankful to Professor Jeffry D. Madura for inspiring discussions on different computational methods, including the molecular docking and for his participation in my committee.

Special thanks to Professor Wilson S. Meng from Pharmacy Department of Duquesne University for his help with the immunological aspects of this project. Professor Charles Dameron helped me in the biochemistry aspects and I want to thank him also for his participation in my committee.

I wish to express my thanks to Dr. Emilio Exavier Esposito and Dr. Pranav Dalal for their help in Autodock and CHARMM and to the members of the theory group at Duquesne University for creating a friendly environment for all of us.

I thankfully acknowledge the financial support from the Department of Chemistry and Biochemistry at the Duquesne University during my studies.

I want to thank my parents for guiding me, for encouraging me to have ambitions, and for teaching me integrity and honesty. I thank Mustafa for his love, and for his patience and support.
Contents

1. Introduction  
   1.1. Structures of class I MHC Proteins ................................. 2  
   1.2. MHC-Associated peptides ........................................... 4  
   1.3. T cell - MHC Interactions .......................................... 8  
   1.4. Vaccination Strategies against Cancer ............................ 9  

2. Goals of MS Research  
   2.1. Statement of Problem ................................................ 12  
   2.2. Previous Studies ..................................................... 14  
   2.3. Goals of This Research ............................................. 18  

3. Computational Strategy  
   3.1. Docking Studies ..................................................... 19  
   3.2. Genetic Algorithms .................................................. 21  

4. Methodology  
   4.1. AutoDock Version 3.0.5 ............................................. 24  
      4.1.1 Preparation of the Ligand and Receptor ....................... 24  
      4.1.2 The Autogrid Procedure ...................................... 25  
      4.1.3 The Genetic Algorithm Implementation ...................... 26  
      4.1.4 The Fitness Function and Free Energy Calculation .......... 27  
   4.2. GOLD Version 2.0 ................................................... 29
4.2.1 Initialization of the Protein and Ligand ........................................... 29
4.2.2 Genetic Algorithm Implementation ................................................. 31
4.2.3 The Fitness Function ................................................................. 33

5. Results and Discussion ................................................................. 36
   5.1. Why GP2? ................................................................. 36
   5.2. Docking Results from AutoDock .............................................. 37
       5.2.1 GP2-MHC Docking Study ............................................. 38
       5.2.2 FLU-MHC Docking Study ............................................. 39
   5.3. Docking Results from GOLD .................................................. 46

6. Conclusions ........................................................................... 49
7. Bibliography ........................................................................ 51
List of Figures

1.1 The X-ray structure of the MHC Complex: PDB code 1QR1 ......................... 3
1.2 Binding cleft of HLA-A2.1 complexed with GP2 peptide ................................. 5
1.3 Overlay of 14 different MHC-associated peptide X-ray structures .................. 7
1.4 The pathway of class I MHC molecules [1] ............................................. 9
2.1 Electron density map of GP2 peptide [6] ............................................... 14
3.1 A sample one-dimensional fitness function illustrating local and global minima ... 21
4.1 The size and location of docking box ..................................................... 25
4.2 Docking box and grid points [54] ............................................................. 26
4.3 The size and location of docking sphere ................................................ 31
4.4 Genetic algorithm implementation in GOLD [55] ..................................... 32
5.1 Overlay of experimental and docked GP2 ............................................... 39
5.2 Overlay of experimental and docked FLU .............................................. 40
5.3 Relative ΔG and # of H-bonds vs. exp. relative binding affinity plot complexes ... 43
5.4 PRO 7 and receptor contacts ................................................................. 45
5.5 Overlay of best docked GP2 and the X-ray structure of GP2 (RMSD: 1.31 Å) .... 47
List of Tables

5.1 Docking results for different peptide-MHC complexes .................................. 41
5.2 Docking results for experimentally modified GP2-analogs ......................... 42
5.3 Docking results for proposed modified GP2-analogs. ............................... 44
5.4 The interaction energy values and number of H-bonds for the nine docked structures in from GOLD ................................................................. 48
Chapter 1

Introduction

The study of the molecular interactions between the components of the immune system and specific molecules of interest is essential for our understanding of disease control as well as normal immune regulation. As is well known, immune system control involves a complex interaction between T cells, T cell receptors (TCR’s), antigens and antibodies. One essential component of immune control is the histocompatibility complex molecule (MHC). These molecules are glycoproteins expressed at the surface of all vertebrate cells. Their name arises from the fact that they are largely responsible for the determining the compatibility of tissues between genetically different individuals [1].

T cell receptors act as ‘recognition sites’ for specific antigens, i.e. MHC-associated peptides. During their differentiation, T-cells become tolerant to complexes of self-peptides and other self-molecules. Thus, if any novel molecule, i.e., from an infectious agent, is introduced, they are most usually recognized as ‘foreign’ by T-cells [2-4].
1.1 Structures of class I MHC Proteins

Two classes of MHC proteins are distinguished: class I, and class II. Class I molecules consist of three parts: a transmembrane protein, called the heavy chain, most of which is exposed at the cell surface. Two segments of alpha helix form a groove between them with two edges in the outermost domains. A small molecule (e.g., a short peptide) attached noncovalently in the groove between the two alpha helices. The third component of class I MHC proteins is a molecule of $\beta_2$-microglobulin ($\beta_2$m), which is also attached noncovalently. An example of crystallographic structure of class I MHC molecules is shown in Figure1.1 [5].

An important characteristic of MHC proteins is their polymorphicity. The heavy chain in the MHC molecule is extremely polymorphic in nature [3,6]. Indeed, a characteristic of this family is the presence of hundreds of MHC alleles in a species. When the influence of recombination is also taken into consideration with the high degree of polymorphism, there is a heightened possibility of the expression of numerous combinations of MHC alleles, and also a high degree of heterozygosity. Under such genetic influences, it would be rare to find two unrelated individuals with an identical combination of HLA genes [3]. HLA-A, HLA-B, and HLA-C are the ‘classical’ class I molecules of humans. HLA stands for human leukocyte antigen primarily due to the fact that the molecules were first studied on leukocytes [1]. These differ only in their heavy chain, with all sharing the same type of $\beta_2$m. All HLA molecules are encoded on chromosome 6, with the exception of $\beta_2$m, which is located on chromosome 15 [2]. A human MHC molecule allele HLA-A2.1, one of the most common class I alleles [7,8], is
dealt with in this thesis. In many cancers, HLA-A2.1 plays a significant role in antigen presentation of both viral and tumor antigens [7,8]. The X-ray crystallographic structures of HLA-A2.1 proteins complexed with a limited number of peptides have been resolved [5, 9-15]. Figure 1 shows the X-ray structure of HLA-A2.1 protein complexed with a molecule of β2m and a short peptide, IISAVGIL (GP2) [5].

![Figure 1.1: The X-ray structure of the MHC complex 1QR1, including the HLA-A2.1 protein, GP2 peptide, and a molecule of β2m. Backbones of the HLA-A2.1 and β2m proteins shown in red and yellow colors. The peptide GP2 is shown in green color as a ball-and-stick. Stretches of beta conformation are represented by the broad yellow arrows (pointing N → C terminal). Regions of α helix are shown as red helical ribbons. The]
pairs of violet spheres represent the disulfide bridges. The molecule of \( \beta_{2m} \) is bound to the junction of the \( \alpha_1 \) and \( \alpha_2 \) domains (at the blue line) and to the \( \alpha_3 \) domain (blue circle) by noncovalent interactions only. A small peptide, shown in green color, is attached noncovalently in the groove between the \( \alpha \) helices of the \( \alpha_1 \) and \( \alpha_2 \) domains.

HLA-A2.1 and \( \beta_{2m} \) expressed in *Escherichia coli* are produced as inclusion bodies, purified and folded [5]. Crystals were grown by hanging drop vapor pressure [5]. The structure is a high-resolution X-ray structure that the resolution is 2.4 Å [5]. It is believed that glycosylation sites are far from the binding groove that the binding is not affected from it.

### 1.2 MHC-Associated Peptides

MHC proteins have a broad specificity providing different peptides to T cells, whose activity is limited to these peptides presented by the MHC proteins [6]. There are several hundred different MHC proteins within the human species, each having different specificity [2]. Thus, MHC proteins bind many peptides with diverse sequences.

Class I molecules, both heavy and light chains, are synthesized in the endoplasmic reticulum (ER) [3]. Cytosolic proteins are the source of peptides that are loaded on the class I molecules. There is a common opinion about endopeptidase activity of the proteasome that it partially degrades these cytosolic proteins. The proteasome has several activities and it is found in the cytosol [3]. The TAP molecule (transporter associated with processing) carries these peptides into the ER [3]. The fate of these peptides inside the ER lumen is not exactly known. According to one hypothesis class I
MHC proteins is immediately loaded with these peptides after release from the transporter [2]. Another suggest that peptides first binds to a chaperone, gp96, and then are carried to class I MHC proteins perhaps with some trimming of the peptides underway [3]. Regardless of these pathways, these peptides reach the class I MHC molecule and binds into the groove shown in Figure 1.2, perhaps after a final trimming step while already in touch with MHC [3].

![Figure 1.2: Binding cleft of HLA-A2.1 complexed with GP2 peptide. GP2 binds in the groove of HLA-A2.1 in an extended conformation.](image)

Class I MHC proteins require three important factors for binding of peptides to them. First, MHC-associated peptides have allele-specific lengths. HLA-A2.1 protein requires 9±1 residue peptides [2-4]. Second, MHC-associated peptides have anchor residues [2,16-19]. A subset of peptide amino acids in N and C termini, termed as anchor residues are important as their side-chains are critical for interacting with polymorphic complementary pockets in MHC peptide binding grooves [2,16-21]. The combination of
amino acids that may bind at the anchor positions is known as the “peptide-binding motif” [2, 5, 6, 19]. Third factor is called occupancy, i.e. binding within “specificity pockets” that are primarily formed by the polymorphic residues within the binding cleft of the MHC molecule [2, 20, 21]. Kuhns et. al. [4] stated that peptides that bind with high affinity to a given allotype are typically found to have one of a few preferred amino acids at each anchor position. The corresponding hypothesis is, peptides that do not have those preferred amino acids at the anchor positions will not bind well [5]. Previous studies explaining the role of anchor residues in binding of peptides to MHC proteins are in agreement with the fact that anchor residues are required by MHC proteins [2,5,16-19]. However, a number of studies show that anchor residues are not sufficient for strong binding, and there are factors other than peptide size and the presence of anchor residues playing a decisive role in determining binding to HLA-A2.1 [17, 22-25]. The center residues of MHC-associated peptides are shown to be important for the recognition by TCRs as anchor residues in N and C termini are important for binding to class I MHC proteins [17-26]. However, Buus et. al. [27] stated that other residues within the peptide besides the anchors may also be used to generate increased binding affinity. Meng et. al. [29] also support this concept and they reported that nonanchor residues are more tolerant to mutations than the anchor residues.

As it can be seen from Figure 1.2, MHC-associated peptides have extended conformation in the binding cleft. Also, all class I MHC associated peptides identified so far have similar extended backbone conformations. Interestingly, the anchor positions, N and C termini of the peptide, show further similarity in backbone conformation as shown in Figure 1.3.
Figure 1.3: Overlay of 14 different MHC-associated peptide X-ray structures. PDB codes are 1QR1 [5], 1HHI [9], 1HHJ [9], 1HHK [9], 1HHH [9], 1JF1 [10], 1JHT [10], 2CLR [11], 1I4F [12], 1DUY [13], 1DUZ [13], 1B0G [14] and 1QSE [15]. Backbones of the aminoacids in positions 1, 2, 8, and 9 do not vary with the sequence.

To summarize, peptides that bind to class I MHC proteins are usually restricted in length and often contain anchor residues at specific locations on the peptide. The anchor residues are required for binding to class I MHC proteins but may not be sufficient to generate a strong binding. The center residues play an important role in allele-specific recognition of antigenic peptides.

The identification and characterization of principles that govern peptide binding to specific structures on MHC proteins is of critical importance in the context of vaccine development and therapeutics, and ultimately to also enhance our knowledge of the basic principles of protein-protein recognition.
1.3 T cell / MHC Interaction

The surface of each T cell displays thousands of identical T cell receptors (TCRs). The TCR binds a bimolecular complex of a small peptide, i.e. a fragment of an antigen, and MHC molecule displayed at the surface of the antigen-presenting cell (APC). Most of the T cells in the body belong to one of two subsets. There are two types of T cells in the body. They differ from the glycoproteins that they have on their surfaces: CD4 and CD8. Which of these molecules is present determines what types of cells the T cell can bind. CD8+ T cells bind epitopes that are part of class I MHC proteins. Class I MHC proteins are expressed by almost all the cells of the body [1].

CD8+ T cells are specific to antigens presented by class I molecules. Most CD8+ T cells are cytotoxic T cells (CTLs). They might destroy cells whose class I epitope they recognize. Viral or tumor-associated cells might be destroyed after the recognition by the T cells specific to that antigen [2-4, 27-31]. To summarize, T cells function as protection mechanism in the body against viruses or tumors. The pathway of the class I MHC proteins are shown in Figure 1.4.
The pathway of class I MHC proteins [1]. The proteins synthesized within the cell are chopped up into small peptides. Then, these peptides bind to MHC proteins. The peptide-bound MHC proteins are transported to the cell surface, where T cell receptors will bind to them.

1.4 Vaccination Strategies against Cancer

Current strategies of tumor vaccination involve utilizing whole tumor cells, or targeting defined antigens [32]. The latter, i.e. antigen-specific vaccination strategies can be divided into two subgroups; antigens not normally expressed somatic cells, and antigens that are normally expressed on somatic cells [32]. The latter class of tumor antigens includes tissue-specific antigens, such as the melanoma/melanocyte antigens, and antigens that are widely expressed in normal tissue albeit at very low levels (e.g. p53 and Her-2/neu) [32]. To exploit these antigens as targets for immunotherapy one has to take into account two major factors: the existence of immunological tolerance, and the
possibility of inducing autoimmune reactions to normal tissues [32]. Work done in animal models and results from a number of clinical studies indicated that efficient immunity against autoantigens can be obtained as long as expression of the antigen concerned is not excessively high and widespread [32].

The HER-2/neu oncogenic protein is a well defined tumor antigen [33]. It is shared among multiple tumor types, including breast and ovarian cancers [33]. Patients with HER-2/neu protein over-expressing breast, ovarian, non-small cell lung, and prostate cancers have been shown to have a pre-existent T cell immunity to HER-2/neu [33-35]. In general, endogenous immunity to HER-2/neu detected in cancer patients demonstrates two prominent characteristics [33]. First, HER-2/neu specific immune responses are found in only a minority of patients whose tumors over-express HER-2/neu. Secondly, immunity, if detectable, is of low magnitude. These observations have led to the development of vaccine strategies design to boost HER-2/neu-specific T cell response in patients [33-35]. HER-2/neu is a non-mutated self-protein, therefore vaccines must be developed based on immunological principles focused on circumventing tolerance [33-35]. Early results in clinical trials demonstrate that significant levels of HER2/neu can be generated with active immunization [33]. Importantly, it was learned that the antigen-specific T cell immunity remains detectable after the completion of vaccinates [33].

In conclusion, with the recent progress in the study of natural and experimentally induced immunity, it is appears that immunotherapy, particularly in relation to cancer treatment, is entering a new era. The continuation of such progress may depend on the progressive development, refinement and use of antigen-specific cancer vaccines in
patients displaying less-progressed stages of disease. More powerful and accurate methodologies may also arise with closer assimilation between, for instance immunologists/oncologists specializing in cell-specific aspects of vaccine development and those researchers in other less ‘classical’ fields of chemistry. For instance, it is hoped that the computational modeling of protein binding may help to catalyze biological research by assessing the specificity, force and predictability of certain protein binding characteristics. In the present thesis, it is this latter field that will be used examined in the context of immunotherapeutics, in an attempt to assess, and hopefully enhance, the predictability of vaccine design.
Chapter 2

Goals of MS Research

2.1 Statement of Problem

As previously stated, the immune system has the capability of recognizing the tumor-associated peptides in the context of class I MHC proteins and ultimately destroying the tumor-associated cell. However, not all tumors are recognized efficiently. One reason hypothesized for poor T cell recognition of tumor-associated peptides is poor binding of potentially immunologically reactive peptides to class I MHC proteins [5]. If the peptides dissociate from class I MHC proteins too quickly, the cell presenting the peptides do not have a sufficient concentration of the specific peptide bound MHC at the surface of the cell to be recognized by circulating T cells [5].

It is known that peptide vaccines have the potential to induce immune responses in vivo that are specific for epitopes on the tumor cells and therefore could have therapeutic potential [33-36]. Several tumor-associated peptides have been identified in breast and ovarian cancer patients [32-36]. The proto-oncogene product HER-2/neu is overexpressed in approximately 30% of patients with breast and ovarian cancer, and such
tumors are associated with a worse prognosis than HER-2/neu negative tumors [32-36]. Many peptides derived from the HER-2/neu have been shown to be recognized by cytotoxic T cells derived from HLA-A2 patients with breast cancer and other carcinomas [26,32-36]. In particular, IISAVGIL (GP2) peptide (HER-2/neu 654-662) is recognized by T cell lines developed from patients [5,26]. Thus, it is one of the potential candidates for tumor vaccines. [5,26]. GP2 binds very poorly even though it is predicted to bind well based upon the presence of the correct HLA-A2.1 peptide binding motif [4,27]. GP2 (IISAVGIL) has the anchor residues ILE at position 2 and LEU at position 9 [5,28]. The crystallographic structure of the GP2 bound to HLA-A2.1 [5] provided an explanation for the poor binding: The central region of the peptide (positions 5, 6, and 7) is flexible, i.e. has multiple conformations, and does not make stabilizing contacts with residues in the MHC binding cleft. It is reported that the electron density map of GP2, which is shown in Figure 2.1, shows uninterpretable electron density within the center of the peptide (the residues VVG at positions 5, 6, and 7, respectively) [5]. Substitutions at the 2 and 9 anchor positions resulted in little or no improvement in binding affinity, but altered the conformation of the putative T cell receptor contact region in the center [37]. Therefore, the next logical step is to explore substitutions at positions other than the primary anchors by focusing on the flexible P5-P7 region. Some of the previous studies addressing the recognition problem of HLA-A2.1-restricted peptides are reviewed briefly in the next subsection of this thesis.
2.2 Previous Studies

Since the binding of peptides to MHC protein is a prerequisite for the recognition by TCR, a significant number of effort has gone to investigate the binding of small peptides to MHC proteins and to improve the binding by performing mutations on the peptides as a strategy [see ref.28 as a review] [5,24-26,37-57]. To investigate the poor binding, several GP2 mutations have been made experimentally as high affinity ligands for cancer immunotherapy [5,26]. Kuhns et. al. [5] measured the relative binding constants of GP2 and several GP2 analogs as well as a high affinity ligand, called MelA, and its analogs to MHC. The GP2 and MelA analogs were obtained by modifying the anchor residues since it is thought that center residues were important for T cell response. The results showed that although anchor substitutions of MelA improved binding significantly to MHC, anchor substitutions did not significantly increase the binding affinity of GP2 to MHC. This is because these anchor substitutions did not address the fundamental problem that the peptide has in the center. That is the center residues do not

Figure 2.1: Electron density map of GP2 peptide [6]. There is an interpretable electron density in the region of positions 5, 6, and 7 [6].
make stabilizing contacts with the binding cleft of the class I MHC molecule. In their subsequent article [37], it is shown that the anchor residue mutations alters the conformation of the T cell receptor contacts, i.e., the center of the peptide. Tanaka et. al. [26] performed modifications on anchor residues together with the position 1 and studied the binding affinities as well as the CTL response. A high affinity peptide FLU [PDP code 1HH1, 9] was used as a positive control. Only the residues at positions 1, 2, and 9 were modified since center residues were thought to be important for T cell responses. However, as it is stated by Sharma et. al. [37] that, altering one residue at positions 2 and 9 changes the peptide structure in the center. In addition, Meng and Butterfield [28] reported that non-anchor residues can also be used for mutations besides the anchor residues. By introducing the third modification at position 1, Tanaka et. al. were able to improve both the binding affinity to HLA-A2 and the recognition by GP2-specific CTL.

There are a number of theoretical studies to investigate the peptide/HLA-A2.1 binding properties, though none of them examines GP2 specifically [24,25,38-57]. Meng et. al. [48] studied influenza matrix protein peptide FLU by using molecular dynamics simulations. They concluded that a dynamic, groove based water network contributes to the formation of the FLU/HLA-A2 complex. A flexible network of water molecules in the binding groove might allow the protein side chains to adjust their conformation in response to different peptides. Also, they proposed that the presence of groove-based water might play a role in determining the affinity of specific peptide-MHC interactions. Froloff et. al. [42] described a methodology to calculate the binding free energy of a protein-ligand complex using a continuum model of the solvent. A formal
thermodynamic cycle is used to decompose the binding free energy into electrostatic and non-electrostatic contributions. In this cycle, the reactants are discharged in water, associated as purely nonpolar entities, and the final complex is then recharged. The total electrostatic free energies of the protein-ligand complex in water are calculated with the finite difference Poisson-Boltzmann method. The nonpolar (hydrophobic) binding energy is calculated using a free energy-surface area relationship, with a single alkane/water surface tension coefficient. The loss in backbone and side-chain configurational entropy upon binding is estimated and added to the electrostatic and the nonpolar components of the binding free energy. Rognan et. al. [45] developed a free energy scoring function, Fresno, to predict the binding free energy of peptides to class I MHC proteins. Fresno has the form:

\[
\Delta G_{\text{binding}} = K + \alpha(\text{HB}) + \beta(\text{LIPO}) + \gamma(\text{ROT}) + \delta(\text{BP}) + \varepsilon(\text{DESOLV})
\]

where, HB stands for H-bond, LIPO for lipophilic, ROT for rotational, BP for buried-polar terms, and last term for desolvation scores. First, ChemScore scoring function is used to calculate the binding energies of the training set, i.e. peptide/HLA-A2.1 complexes of PDB codes 1HHG, 1HHH, 1HHI, 1HHJ, and 1HHK. Then, the Fresno is obtained by recalibrating the computed values fitting to the experimental data. It differs from the other scoring functions [58-62] developed before mainly by the explicit treatment of ligand desolvation and of unfavorable protein-ligand contacts. This method requires the three-dimensional structures of protein-ligand complexes. Predictions were more accurate for HLA-A2 binding peptides as the training set had been built from
experimentally determined structures (average error: 3.1 kJ/mol). For the homology model-derived equation, i.e. the computed data is fitted to the homology modeled data since the experimental data were not available, the average error in predicting binding free energy of peptides to the MHC molecule of mouse \((K^k)\) was higher (5.4 kJ/mol) but still acceptable. Then, in their next paper [53], the Fresno has been compared to six universal scoring functions (ChemScore, Dock, FlexX, GoldScore, Pmf, Score). Fresno performed better than other six scoring functions. Another computational study of Rognan \textit{et. al.} [50] investigated the mouse histocompatibility molecule \((K^k)\) complexed with an Influenza Hemagglutinin peptide, \(\text{Ha}_{255-262}\), and the TCR by using the tools of homology modeling as well as molecular dynamics. First, \(K^k\) was built by homology modeling and subsequently refined by simulated annealing and restrained molecular dynamics. TCR models are then docked into the \(\text{Ha}_{255-262}/K^k\) model. The experimental mutations have been examined systematically and the computational results are in good agreement with the experiment. It is believed that such models may guide the future rational experimentation. In another significant molecular modeling study, Michielin \textit{et. al.} [49] presented a methodology for generating a homology model of a TCR-MHC-peptide complex. It is shown that by using computational tools together with the experimental mutation data, the interactions of complex biological systems can be better understood. This model investigates the effects of mutations on the binding of TCR to its ligand and moreover suggests new mutations for TCR. The mutational data obtained is qualitative and it is noted that for quantitative analyses, more computer-intensive methods need to be used.

2.3 Goals of this research
The long-term goal of this computational study is to develop a protocol so that better binding GP2 analogs for cancer immunotherapy can be designed. To do this the following short-term goals have been established. First of all, it is planned to develop a practical and accurate docking method to investigate the binding of GP2 to HLA-A2.1. Then, gain insight into the molecular forces that give rise to strong binding by performing the same type of calculations on different peptide/MHC complexes including strong, moderate, and weak binders. Finally, propose mutations that could lead to new tumor vaccines.
Chapter 3

Computational Strategy

The docking calculations for peptide/HLA-A2.1 complexes have been planned as a strategy to obtain the best ligand confirmation inside the binding groove. Two main steps have been followed. First, the experimental data needed to be reproduced by performing docking calculations on ligand/HLA-A2.1 complexes. Second, the GP2 analogs, i.e. the mutated GP2 ligands have been docked by using the same docking method to obtain a better binding ligand. AutoDock version 3.0.5 [55] was used for both purposes. Then the GP2 peptide have been investigated further by using GOLD program [55]. The genetic algorithm has been used as implemented in AutoDock and GOLD programs. In the following parts of this thesis, simple introductory information on docking studies and genetic algorithms is provided.

3.1 Docking studies

Molecular docking predicts the orientation of the ligands bound to receptors by assuming that the receptor conformation is known [66]. Molecular docking is a fast method to explore substrate/receptor complexes in the field of drug discovery as well as in understanding biochemical processes [66]. The major techniques used for molecular
docking are: genetic algorithms, simulated annealing, molecular dynamics, Monte Carlo methods, distance geometry methods, point complementary methods, fragment-based methods, tabu searches, and systematic searches [66].

Docking procedures are composed of two components: a search algorithm and a scoring function. The search algorithm finds different conformations for the ligand by using one of the methods listed above. Systematic searches explore all possible binding modes between the ligand and receptor. However, this takes a huge amount of computational time especially for large flexible ligands. The amount of conformational space explored and the computational time required for the search must be balanced. Scoring functions are used to rank the different conformations obtained by the search algorithm. A good scoring function distinguishes the experimentally obtained conformation from all other conformations explored through the search algorithm. The major scoring methods are empirical free energy scoring functions, molecular mechanics force fields, and knowledge-based functions [66].

Some of the docking programs currently in use are the DOCK [67,68], GOLD [64], AutoDock [63], Surflex [69], MOE-Dock [70,71], FlexX [72-75], FTDOCK [76-79]. The differences between them are derived from the different search algorithms or different implementation of the same algorithms, and different scoring functions. Most of these programs hold the receptor rigid and allow a certain degree of flexibility to the ligands. The docking methodology in AutoDock and GOLD is explained in the methodology section in detail.

3.2 **Genetic algorithms**

The Genetic algorithms [80] are considered to be stochastic global optimization
methods. Since they do not use gradient information, they can be applied various different types of optimization problems. The genetic algorithm searches the parameters forming the conformational space simultaneously. Judson [80] simply explains the genetic algorithms by using Figure 6 as an example. Three local minima labeled as I, II (the global minimum), and III are shown in the f(x) function. Then, the genetic algorithm terms are introduced; fitness, populations, and chromosomes. The function f(x) is the fitness. The populations are set of individuals sampling the conformational space.

![Figure 3.1: A sample one-dimensional fitness function illustrating local minima labeled as I, II, III, and the global minimum labeled as II [80].](image)

The x in the example denotes for one of the chromosomes, which are the parameters forming each individual. Other genetic algorithm terms are mutations, selection, crossovers, and migrations [81]. The mutation operator obtains the individuals by randomly changing the chromosomes. The best individuals based on the fitness function are then selected for crossover. The crossover allows an exchange of a set of
chromosomes from one parent to another. The migration moves individual chromosomes from one sub-population to another. The mutation, crossover, and migration procedures continue until some stopping criteria are met. The following section includes the information on how genetic algorithms were implemented in AutoDock and GOLD programs.
Chapter 4

Methodology

This section examines AutoDock and GOLD methodologies, including information on how those programs work and the details of docking studies performed in this study.

The crystal structures of 14 peptide/HLA-A2.1 complexes were obtained from the Protein Data Bank (PDB) [82] having the PDB codes of 1QR1 [5], 1HHI [9], 1HHJ [9], 1HHG [9], 1HHK [9], 1HHH [9], 1JF1 [10], 1JHT [10], 2CLR [11], 1I4F [12], 1DUY [13], 1DUZ [13], 1B0G [14], and 1QSE [15]. Docking studies were performed on these systems by utilizing AutoDock program [63]. The 1QR1 complex, i.e. the GP2/HLA-A2.1 system has been investigated in GOLD program [64] since in AutoDock, the backbone of the GP2 ligand had to be fixed. GOLD allows full flexibility for the ligands. The energy relaxations in MOE program [71] were performed before all docking calculations to obtain the structures that are free of steric clashes. The details of these energy minimizations are given in section 5.2.

4.1 AutoDock Version 3.0.5

The program AutoDock was developed to provide an automated procedure for predicting the interaction of ligands with biomacromolecular targets. The docking simulation is carried out using one of a number of possible search methods. The
Lamarckian genetic algorithm and the Monte Carlo simulated annealing are available in AutoDock. The receptor is considered as rigid, and the allowed flexibility for the ligand is 32 torsion angles. The docking methodology in AutoDock is examined in the following four sub-sections: Preparation of the ligand and receptor, the Autotors and Autogrid procedure, the genetic algorithm implementation, and the fitness function and evaluation of the free energy.

4.1.1 Preparation of the Ligand and Receptor

First the docking box has been placed into the active side of the receptor. For GP2, a docking box with a grid consisting of 92 X 70 X 70 points and 0.375 Å grid spacing were employed. The box was oriented so that the long side was along the direction of the center of binding site to the center of the entrance of the binding site. In this orientation, the box included the entire binding site and some area just outside of the binding pocket entrance. The docking boxes of sizes between 82 X 60 X 60 and 110 X 80 X 80 with 4 grid point increments have been used in the calculations. The size 92 X 70 X 70 has been found optimum so that the ligand is free to have different conformations in the binding cleft and also it is small enough to save the computational time. The size and location of the docking box used in this study is shown in Figure 4.1.
The degree of flexibility to be assigned to the ligand can be imported by using the Autotors utility. Autotors uses a “Rigid Root” as a starting point for the “torsion angle tree” [reference 90 has detailed information on torsion tree algorithms].

4.1.2 The Autogrid Procedure

AutoDock requires pre-calculated grid maps, one for each atom type present in the ligand being docked. This helps to make the docking calculations extremely fast. These maps are calculated by AutoGrid. A grid map consists of a three dimensional lattice of regularly spaced points, surrounding (either entirely or partly) and centered on some region of interest of the macromolecule under study. Each point within the grid map stores the potential energy of a ‘probe atom’ or functional group that is due to all the atoms in the macromolecule. Figure 4.2 illustrates the main features of a grid map [63]:

**Figure 4.1:** The size and location of docking box
Figure 4.2: Docking box and grid points [63]. In the figure, the whole protein is in the docking box which is defined by the grid points with a user defined grid spacing.

The energetics of a particular substrate configuration is found by tri-linear interpolation of affinity values of the eight grid points surrounding each of the atoms in the substrates. The time to perform an energy calculation using the grids is proportional only to the number of atoms in the substrate, and is independent of atoms in the protein.

4.1.3 The Genetic Algorithm Implementation

The particular arrangement of a ligand and a protein is defined by state variables, which include a set of variables describing the translation, orientation, and conformation of the ligand with respect to protein. Each state variable corresponds to a gene, and the ligand’s state corresponds to a genotype, whereas its atomic coordinates correspond to the phenotype. In AutoDock implementation [63], the chromosome is composed of real valued genes: three Cartesian coordinates for the ligand translation; four variables
defining a quaternion specifying the ligand orientation; and one real-value for each ligand
torsion angle, in that order. The genetic algorithm begins by creating a random
population of individuals, where the number of individuals in the population is user
defined. For each random individual in the population, random values are assigned for
the genes. A loop over generations then takes place, repeating until the maximum number
of generations or the maximum number of energy evaluations is reached, whichever
comes first. A generation consists of five stages: mapping and fitness evaluation,
selection, crossover, mutation, and elitist selection, in that order. Mapping translates from
each individual’s genotype to its corresponding phenotype. This allows each individual’s
fitness to be evaluated. The fitness function and the energy evaluation are explained in
the next sub-section. Every time an individual’s energy is calculated. This is followed by
proportional selection to decide which individuals will reproduce. Crossover and
mutation are performed on random members of the population according to user-defined
rates of crossover and mutation. First, crossover is performed. The new members are
replaced the parents in the population, keeping the population size constant. Crossover is
followed by mutation. Optional user-defined integer parameter elitism determines how
many of the top individuals also automatically survive into the next generation. The
genetic algorithm iterates over generations until one of the termination criteria is met.

4.1.4 The Fitness Function and Free Energy Calculation

The fitness is the sum of the intermolecular interaction energy between the ligand
and the protein, and the intramolecular energy of the ligand. At the end of each docking,
AutoDock reports the fitness (the docked energy), the state variables, the coordinates of
the docked conformation, and the estimated free energy of binding (ΔG).
\[ \Delta G = \Delta G_{vdw} \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \Delta G_{hbond} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \]

\[ + \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij})(r_{ij})} + \Delta G_{tor} N_{tor} + \Delta G_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{\left(\frac{r_{ij}^2}{2\sigma^2}\right)} \]

(2)

The Lennard-Jones 12-6 dispersion-repulsion term is first, the second is a directional 12-10 H-bonding term, where E(t) is a directional weight based on the angle, t, between the probe and the target atom, the third term is a screened Coulombic electrostatic potential, the fourth term is a measure of the unfavorable entropy of ligand binding due to the restriction of conformational degrees of freedom, which is proportional to the number of of \(sp^3\) bonds in the ligand, \(N_{tor}\), and finally the last term accounts for the desolvation effect. For desolvation, the pairwise, volume-based method of Stouten et. al. is used [63]. This method has the advantage that it is consistent with the pre-calculated affinity grid formulation used by AutoDock [63]. For each atom in the ligand, fragmental volumes of surrounding protein are weighted by an exponential function and then summed, evaluating the percentage of volume around the ligand atom that is occupied by protein atoms [63]. This percentage is then weighted by the atomic solvation parameter of the ligand atom to give the desolvation energy [63]. The full method may be broken into four separate components: burial of apolar atoms in the ligand, burial of apolar protein atoms, burial of polar and charged atoms in the ligand,
and burial of polar and charged protein atoms [63]. Great success has also been reported in using simply the amount of hydrophobic surface area buried upon complexation as a measure of the ‘hydrophobic effect’, so they tested several formulations that included only the volume lost around ligand carbon atoms [63]. The burial of polar atoms caused particular problems. Apart from the volume-based method, a simpler formulation for the solvent transfer of polar atoms has been tested, where a constant term corresponding to the favorable free energy of interaction of a polar atom with solvent is estimated, and this is subtracted from the binding free energy [63].

4.2 GOLD Version 2.0

GOLD [64] (Genetic Optimization for Ligand Docking) is an automated docking program that uses genetic algorithm to search the conformational space and allows full flexibility for acyclic ligands and partial protein flexibility in the neighborhood of the protein binding cleft [64]. The docking methodology of GOLD will be examined in three parts: initialization of the protein and ligand, genetic algorithm implementation, and the fitness function.

4.2.1 Initialization of the Protein and Ligand

GOLD needs a user defined docking sphere that is placed within the binding cleft of the protein with a user defined center and a radius. The center can be a point, or an atom. Since there are X-ray crystal structures for ligands bound to HLA-A2.1 receptors, the binding site of the receptor is already known. The center of the binding cleft has been placed by using CHARMM program [86]. The receptor atoms within the 10 Å distances of ligand atoms were selected and the center of the selected volume obtained by using
‘stats’ option. The coordinates of the center is then found (4.00, 16.1, -6.70) in x, y, z directions, respectively. As a radius going along with this center; 20, 25, and 30 Å have been used. The 20 Å radius was found optimum so that the computational time and the accuracy can be balanced. Figure 4.3 shows the docking sphere used in the calculations.

The protein is considered as rigid except OH groups of SER, THR and TYR, and NH₄⁺ group of LYS in the active site neighborhood. The ligand can be prepared as fully flexible. The simplest constraints to apply to the ligand are keeping the ring corners, amide bonds, planar nitrogens, and/or internal hydrogen bonds constant. The other constraints available are the covalent constraints, distance constraints, H-bond constraints, structure-based constraints, and similarity constraints. The default value for the number of runs, which varies between 1 and 50, is 10. The protein and ligand input files must be in pdb or tripos mol2 format. The latter is suggested. If the pdb format is used, the program will assign the partial charges using a modified Tripos force field. Since mol2 files contain the partial charge information, different force fields can be used to prepare the mol2 files. In this study, the input files were prepared by using MOE program [83] and mol2 files are used.
Figure 4.3: The size and location of docking sphere. The center coordinates are (4.00, 16.1, -6.70) in x, y, z directions, respectively with 20 Å radius has been used in docking calculations in GOLD.

4.2.2 Genetic Algorithm Implementation

GOLD uses a steady-state operator-based genetic algorithm to sample the conformational space and ligand binding modes [64]. This genetic algorithm is illustrated in Figure 4.4 [64].
1. A set of reproduction operators (crossover, mutation, etc) is chosen. Each operator is assigned a weight.
2. An initial population is randomly created and the fitnesses of its members determined.
3. An operator is chosen using roulette wheel selection based on operator weights.
4. The parents required by the operator are chosen using roulette wheel selection based on scaled fitness.
5. The operator is applied and child chromosomes produced. Their fitness is evaluated.
6. If not already present in the population, the children replace the least fit members of the population.
7. If 100000 operators have been applied stop otherwise go to 3.

**Figure 4.4:** Genetic algorithm implementation in GOLD [64].

There are three operators used: mutation, crossover, and migration. The mutation operator creates the individuals in each population by randomly changing the rotatable bonds in the protein and ligand. Torsion angle values vary between -180° and 180° in step-sizes of 1.4°. The default values for the number of individuals in one population and number of populations are 100 and 5, respectively. The crossover operator provides the exchange of chromosomes between the individuals. The migration operator copies an individual from one population to another population. Operators were chosen using roulette-wheel selection based on operator weights. These weights were chosen so that crossover and mutation were applied with equal probability and migration was applied 5% of the time. After reaching the default value of number of operators, 100000, the algorithm terminates.

After a conformation of the ligand and protein binding cleft generated, the ligand is placed within the active site using a least squares fitting procedure. Then, its fitness
score, which is explained in the next sub-section, is evaluated.

4.2.3 The Fitness Function

Two different scoring functions are available in GOLD: ChemScore and GoldScore [64]. Chemscore was derived empirically from a set of 82 protein-ligand complexes for which measured binding affinities were available. The ChemScore function was trained by regression against measured affinity data. The calculation of free energy of binding ($\Delta G_{\text{binding}}$) is shown in Equation 3.

$$\Delta G_{\text{binding}} = \Delta G_0 + \Delta G_{\text{hbond}} + \Delta G_{\text{metal}} + \Delta G_{\text{lip}o} + \Delta G_{\text{rot}}$$

(3)

Each component of this equation is the product of a term dependent on the magnitude of a particular physical contribution to free energy and a scale factor determined by regression. The final ChemScore value is then obtained by adding a clash penalty and internal torsion terms, which prevents close contacts in docking and poor internal conformations. Covalent and constraint scores may also be included.

$$\text{ChemScore} = \Delta G_{\text{binding}} + P_{\text{clash}} + c_{\text{internal}}P_{\text{internal}} + (c_{\text{covalent}}P_{\text{covalent}} + P_{\text{constraint}})$$

(4)

The GoldScore fitness function is composed of four components: protein-ligand hydrogen bond energy (external H-bond), protein-ligand van der Waals (vdw) energy (external vdw), ligand internal vdw energy (internal vdw), and ligand torsional strain energy (internal torsion). Optionally, a fifth component, ligand intramolecular hydrogen bond energy (internal H-bond), may be added. Empirical parameters used in the fitness function (hydrogen bond energies, atom radii and polarizabilities, torsion potentials,
hydrogen bond directionalities, etc.) are taken from GOLD parameter file, which can be edited by the user. The external vdw score is multiplied by a factor of 1.375 when total fitness score is computed. This is an empirical correction to encourage protein-ligand hydrophobic contact. The final GoldScore value is stated in the following expression.

\[
\text{GoldScore} = - (\text{H\_Bond\_Energy} + \text{Internal\_Energy} + \text{Complex\_Energy})
\]  

(5)

The first term denotes for hydrogen bonding energy, which is determined by taking the combinations of all donor and acceptor atoms whether they form a hydrogen bond. The internal energy of the ligand, which is the sum of the ligand steric and torsional energies, is calculated by using molecular mechanics expressions. The steric energy was calculated using a 6-12 potential of the form:

\[
E_{ij} = \frac{C}{d_{ij}^{12}} - \frac{D}{d_{ij}^{6}}
\]

(6)

The torsional energy \(E_{ijkl}\) is calculated by using the Tripos forcefield of the form:

\[
E_{ijkl} = \frac{1}{2} V_{ijkl} \left[ 1 + \frac{n_{ijkl}}{n_{ijkl}} \cos(n_{ijkl} \cdot \omega_{ijkl}) \right]
\]

(7)

The last term is a pairwise energy obtained for the steric energy of interaction between the protein and the ligand. A 4-8 potential of the following form with linear cut-off was used to calculate the interaction energy between the ligand and the protein.
The cut-off distance used was 1.5 times the sum of the van der Waals radii of the atoms.

The 4-8 potential was parametrized to reproduce the minimum of the more usual 6-12 potential.

\[ E_{ij} = \frac{A}{d_{ij}^8} - \frac{B}{d_{ij}^4} \]  

(8)
Chapter 5

Results and Discussion

5.1 Why GP2?

There are two primary reasons to study GP2 bound MHC proteins in detail. The first is to understand how class I MHC binds peptides. There is a great deal known about how class I MHC binds many peptides. Unfortunately, there is very little information known about how the protein binds any particular peptide. There are many examples of crystal structures of high affinity peptides bound to class I MHC. If a self-peptide binds to class I MHC with high affinity, there is a larger concentration of peptide bound MHC at the cell surface, and thus a greater chance that T cells would be able to recognize the complex well. However, GP2 bound complex has a low affinity, a poor thermal stability, and a very short cell surface half-life. Thus, GP2 is a perfect example of a poor binding peptide and as such offers the first opportunity to understand poor binding.

The second reason to examine GP2 is that poor affinity peptides are potentially better targets for immunotherapy [5]. It is known that the patients in breast cancer have
a significant amount of GP2-specific T-cells. Peptides with poor affinities like GP2 have been believed for cancer immunotherapy, because it is more likely that the immune system has not become tolerant to them. GP2 is therefore a good candidate for study.

5.2 Docking results from AutoDock

Docking studies have been performed to address the binding affinity of GP2 to HLA-A2 using the X-ray crystallographic structure of GP2/HLA-A2.1 complex [5]. Partial energy minimizations have been performed on all systems using the Amber 94 force field [83] and the MOE program [71] to avoid the steric clashes. The default sequence of minimization methods in MOE program has been used for partial energy minimizations.

For weaker binding peptides, such as GP2, where all degrees of freedom were free except for those associated with the amide bonds, the reported X-ray structure could not be reproduced (1QR1). With the backbone fixed, the docking studies returned a similar structure to experiment. This led us to use a different docking program, GOLD, for GP2/MHC complex. GOLD results for GP2 are discussed in section 5.3. However, interestingly, for other ligand-MHC complexes (1HII, 1JHT, and 1B0G), the experimental data have been reproduced with much better RMSD only by keeping amide bonds fixed. The results from the docking studies on 1QR1 and 1HII are presented in detail in Figures 4 and 5.

The ligand has been treated two different ways for docking calculations. First by keeping the backbone conformation fixed, then by keeping only amide bonds fixed. The receptor has been taken as rigid in all docking calculations.
molecules have been retained and implicit solvation effect has been added.

The systems used for docking studies in Autodock are GP2-MHC molecule:
1QR1, other ligand-MHC proteins: 1HHI, 1HHJ, 1HHG, 1JF1, 2CLR, 1HHK, 1HHH, 114F, 1JHT, 1DUY, 1DUZ, 1B0G, 1QSE and MHC proteins bound to GP2-analogs (experimental proposed modifications). The docking studies were carried out in different manners, as described below.

5.2.1 GP2/MHC Docking Study

Starting with the X-ray crystallographic structure [5], GP2/MHC complex has been partially energy minimized and investigated by docking. The receptor and the backbone conformation of GP2 have been kept fixed in docking calculations. The method and details of docking are identical to those described above. The MOE program has been used to calculate the root mean square deviation (RMSD) between the experimental and docked structures. In this procedure, a set of 3D protein structures are superposed in such a way as to maximize their spatial overlap. Such a superposition is useful because it highlights both the regions of conserved structure and the areas of divergence or modification in the protein set. A weighted non-linear optimization is used to determine the solid-body transformations required to maximize the superposition of the protein atomic coordinates. Then, the RMSD between protein atomic coordinates is taken to be:

\[
\sum_{k=1}^{M} \sum_{l>k} \sum_{i=1}^{N_{kl}} w_i \left| R_k(x_{ik} - t_k) - R_l(x_{il} - t_l) \right|^2
\]

(9)

where, \(M\) is the number of proteins, \(N_{kl}\) is the number of corresponding atoms, \(w\) is the
protein-to-protein importance weights, \( w_i \) is the importance weight of each atom correspondence, \( R_k \) and \( R_l \) are the rotation matrices, \( x_{ik} \) are the coordinates of the \( i^{th} \) corresponding atom in the \( k^{th} \) protein, \( x_{il} \) are the coordinates of the \( i^{th} \) corresponding atom in the \( l^{th} \) protein, and \( t_k \) and \( t_l \) are the translation vectors [34].

All atoms has been taken into account for calculating the superposition except hydrogen atoms. The comparison of experimental and docked structure of GP2 is shown in Figure 5.1.

![Overlay of experimental and docked GP2.](image)

**Figure 5.1:** Overlay of experimental and docked GP2.

The RMSD has been calculated as 1.42 Å, while the resolution of the X-ray is 2.40 Å [4]. The predicted GP2 structure is in good agreement with the experimental findings since the RMSD value from the X-ray structure is much smaller. However, the whole backbone of GP2 have had been kept fixed. The GP2/MHC complex has been then further examined by using GOLD program.

### 5.2.2 FLU-MHC Docking Study

The partially minimized structure of FLU and MHC has been docked by using AutoDock. Only the amide bonds of the FLU has benn kept fixed and the receptor has
been taken as rigid. An implicit solvation effect has been used. The number of grid points and the grid spacing used for the docking box are 96 x 72 x 72 and 0.375 Å, respectively. The experimental and docked structures of FLU are shown in Figure 5.2. The resulting RMSD is 0.42 Å, which is good agreement between the docked and experimental structures.

**Figure 5.2:** Overlay of experimental and docked FLU. Although only amide bonds of FLU are fixed, the experimental structure has been reproduced with 0.42 Å values.

It is experimentally proven that while GP2 binds MHC poorly, FLU has a very high relative binding affinity for MHC [5]. One interpretation is that GP2 binding is less than FLU due to GP2’s increased flexibility. The results of docking studies on different MHC complexes and also experimentally modified GP2-analogs are shown in the Tables 5.1 and 5.2. Results have been presented in terms of ΔG (kcal/mol) and number of H-bonds between the ligand and the receptor. Relative binding estimates have been obtained by personal contact with Professor Wilson Meng from Pharmacy Department, Duquesne University. RMSD between the X-ray crystallographic and the docked structures (in Å) are given in Table 5.1. Two sets of numbers are reported for hydrogen bonds. The numbers on the left have been obtained according to the MOE criterion; bond distance of
HAB is equal to or less than 3.5 Å, for H-bonding. The numbers on the right have been obtained by using the following criteria; HAB distance is equal to or less than 3.2 Å and A - HAB angle is between 135 – 180° [84].

**Table 5.1:** Docking results for different peptide-MHC complexes

<table>
<thead>
<tr>
<th>PDB code</th>
<th>Sequences of Ligands</th>
<th>ΔG (calc. kcal/mol)</th>
<th>Bind. Est.</th>
<th>RMSD (Å)</th>
<th># of H-bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1QR1 (GP2)</td>
<td>I S A V V G I L</td>
<td>1</td>
<td>*</td>
<td>1.72</td>
<td>5</td>
</tr>
<tr>
<td>1HHI (FLU)</td>
<td>G I L G F V F T L</td>
<td>5.8</td>
<td>*******</td>
<td>0.40</td>
<td>10</td>
</tr>
<tr>
<td>1HHJ</td>
<td>I L K E P V H G V</td>
<td>5.8</td>
<td>***</td>
<td>0.59</td>
<td>14</td>
</tr>
<tr>
<td>1HHG</td>
<td>T L T S C N T S V</td>
<td>4.8</td>
<td>***</td>
<td>1.10</td>
<td>14</td>
</tr>
<tr>
<td>1JFI</td>
<td>E L A G I G I L T V</td>
<td>7.5</td>
<td>*******</td>
<td>0.66</td>
<td>20</td>
</tr>
<tr>
<td>2CLR</td>
<td>M L L S V P L L L G</td>
<td>5.3</td>
<td>*******</td>
<td>0.65</td>
<td>9</td>
</tr>
<tr>
<td>1HHK</td>
<td>L L F G Y P V Y V</td>
<td>5.5</td>
<td>*******</td>
<td>0.98</td>
<td>10</td>
</tr>
<tr>
<td>1HHH</td>
<td>F L P S D F F P S V</td>
<td>7.9</td>
<td>*******</td>
<td>0.79</td>
<td>10</td>
</tr>
<tr>
<td>114F</td>
<td>G V Y D G R E H T V</td>
<td>7.8</td>
<td>*******</td>
<td>1.38</td>
<td>28</td>
</tr>
<tr>
<td>1JHT</td>
<td>A L G I G I L T V</td>
<td>6.5</td>
<td>-</td>
<td>0.65</td>
<td>20</td>
</tr>
<tr>
<td>1DUY</td>
<td>L F G Y P V Y V</td>
<td>4.5</td>
<td>-</td>
<td>0.59</td>
<td>13</td>
</tr>
<tr>
<td>1DUZ</td>
<td>L L F G Y P V Y V</td>
<td>5.5</td>
<td>-</td>
<td>0.61</td>
<td>10</td>
</tr>
<tr>
<td>1B0G</td>
<td>A L W G F F P V L</td>
<td>4.3</td>
<td>-</td>
<td>1.88</td>
<td>8</td>
</tr>
<tr>
<td>1QSE</td>
<td>L L F G Y P R Y V</td>
<td>4.0</td>
<td>-</td>
<td>1.89</td>
<td>5</td>
</tr>
</tbody>
</table>

As can be seen from Table 5.1, the number of hydrogen bonds and the free energy
values correlate with the experimental binding strengths only qualitatively. For the complexes 2CLR, 1HHK, and 1HHH that are known as strong binders, the estimated number of H-bonds are 7, 7, and 6 respectively. Thus, hydrogen bonding is not the only factor affecting the ligand binding. Also, this might be due to the limitations of the method which are discussed later.

Finally, docking studies have been performed for experimentally modified GP2 analogs and the proposed GP2 analogs. The results are shown in Tables 1 and 2, respectively.

Table 5.2: Docking results for experimentally modified GP2-analogs. This is a qualitative correlation between the experimental binding affinities and the calculated binding free energies. Also, number of H-bonds correlates qualitatively with the experimental binding affinity data.

<table>
<thead>
<tr>
<th></th>
<th>ΔG (calc., relative)</th>
<th>Binding (exp.*, rel.)</th>
<th># of H-bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>IISAVVGIL</td>
<td>1.0</td>
<td>1.0</td>
<td>5</td>
</tr>
<tr>
<td>ILSAVVGIV</td>
<td>4.7</td>
<td>2.4</td>
<td>8</td>
</tr>
<tr>
<td>FLSAVVL</td>
<td>5.0</td>
<td>2.2</td>
<td>11</td>
</tr>
<tr>
<td>FISAVGIV</td>
<td>5.0</td>
<td>3.0</td>
<td>9</td>
</tr>
</tbody>
</table>

The correlation can be seen more clearly from the following plot, in Figure 5.3.
Figure 5.3: Relative ΔG and number of H-bond vs. experimental relative binding affinity plot.

The $R^2$ values for the correlations of calculated free energies and the number of hydrogen bonds with experimental binding strengths are 0.846 and 0.686, respectively. This shows that there is a poor correlation between the calculated and experimental findings.

Finally, in Table 5.3, the docking results are shown for proposed GP2-like ligands. According to docking results shown in Table 5.3, binding has been improved by selective mutations. Based upon proposed modifications, IISAVVPIIL has been studied experimentally and found that it increased the binding affinity 68%. IISAVVPIIL has been studied because PRO has been thought to decrease the backbone flexibility, and also experimentally stated that hydrophobic residues are preferred at position 3 [28]. The interaction between the PRO7 and the receptor potential energy surface is shown in
Figure 5.4.

Table 5.3: Docking results for proposed modified GP2-analogs

<table>
<thead>
<tr>
<th>Sequence</th>
<th>ΔG (calc., kcal/mol)</th>
<th>ΔG (calc., relative)</th>
<th># of H-bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>IISAVVGIL</td>
<td>-4</td>
<td>1.0</td>
<td>8</td>
</tr>
<tr>
<td>IISAVVPIL</td>
<td>-18</td>
<td>4.5</td>
<td>12</td>
</tr>
<tr>
<td>IISAVPGIL</td>
<td>-21</td>
<td>5.3</td>
<td>11</td>
</tr>
<tr>
<td>IISANVGIL</td>
<td>-16</td>
<td>4.0</td>
<td>11</td>
</tr>
<tr>
<td>IISAQVGIL</td>
<td>-21</td>
<td>5.3</td>
<td>7</td>
</tr>
<tr>
<td>IISAFVGIL</td>
<td>-21</td>
<td>5.3</td>
<td>9</td>
</tr>
<tr>
<td>IMSAVVGIL</td>
<td>-21</td>
<td>5.3</td>
<td>11</td>
</tr>
<tr>
<td>IISPVVPIl</td>
<td>-23</td>
<td>5.8</td>
<td>10</td>
</tr>
</tbody>
</table>
As can be seen from Figure 5.3, there is a hydrophobic interaction between PRO7 and the hydrophobic residues of the receptor. Another reason of the increased binding affinity can be the decreased flexibility of the backbone due to the PRO at position 7.

The limitations of docking studies performed in this study were two-fold: First, the receptor including crystallographic water molecules is fixed; also there are constraints on the ligands. For GP2 the backbone has been kept fixed. Second, implicit solvent is used. Third, only non-bonded terms have been taken into account for energy calculations. Therefore, it gives qualitative results for energy values. However, the main advantage of docking calculations in AutoDock is the reduced computational time. Therefore, these studies can be useful to determine the best set of ligands for molecular dynamics studies.
that takes much longer time than docking calculations.

5.3 Docking results from GOLD

Since the backbone of GP2 had to be fixed in AutoDock calculations, the 1QR1 complex (GP2/MHC) was further investigated by using GOLD program, which allows full flexibility for the ligand. The X-ray structure was reproduced without any constraints on the ligand GP2, and the receptor was treated as rigid except the OH groups of SER, THR, and TYR as well as the NH₄ group LYS amino acids in the binding cleft. Nine docked structures were identified as good in terms of RMSD values and the conformation having the RMSD values (in Å): 1.31, 1.49, 1.53, 1.57, 1.68, 1.79, 1.87, 1.98, and 2.03. These structures were obtained by efficiently sampling the conformational space and also by using different starting structures for the receptor. Also, the following parameters affecting docking studies have been investigated systematically: Different radius values for the docking sphere, different coordinates for the center of the binding cleft, different molecular mechanics force fields, genetic algorithm settings, and different degrees of flexibility for the ligand. The 20, 25 and 30 Å radius values were used and the 20 Å was found optimum, because the better and faster results have been obtained with it. Also, this size has been found big enough to provide enough space for the free conformational change of the ligand. Amber 94 [83], Triposs [85] and CHARMM [86] force fields were used and the latter was found to give better RMSD values. The receptor atoms 1521 and 1524 and also the coordinates (4.00, 16.1, -6.70) in x, y, z directions have been used. The latter was found optimum based on RMSD values from the experimental structure. In terms of genetic algorithm parameters; 10, 20, and 50 runs with default, doubled and
tripled population size and number of operations have been used, and 50 runs with doubled parameters have been found to give best and fastest results in terms of the RMSD values from the X-ray crystallographic structure. Thus, the nine good structures, having RMSD values smaller then 2.5 Å, have been obtained by using the optimum values but with different receptor starting structures. The comparison of the best docked-structure in terms of RMSD values out of these nine good structures and the X-ray structure is shown in Figure 5.5.

![Figure 5.5](image)

**Figure 5.5:** Overlay of best docked GP2 in GOLD and the X-ray structure of GP2 (RMSD: 1.31 Å). The docked structure and experimental structure are shown in blue and element colors, respectively as ball-and-stick.

The docked GP2 and the experimental data are in excellent agreement. Although GP2 is a fairly long and flexible ligand, the experimental data was reproduced efficiently. The criteria for being best docked-structure includes the binding energy values, H-bonds and hydrophobic contacts between the ligand and the receptor as well as the RMSD values. To obtain a better comparison, the interaction energy values were calculated by using the MOE program. Also, each H-bond between the ligand and the receptor is
examined carefully. The structures with lower RMSD values are expected to have lower interaction energies and/or stronger H-bonds relative to each other. Neither the interaction energy values nor the H-bond information correlates with the RMSD values. Table 5.4 shows the RMSD values of nine docked structures as well as the number of H-bonds using MOE criterion.

**Table 5.4:** The interaction energy values and number of H-bonds for the nine docked structures in from GOLD.

<table>
<thead>
<tr>
<th>GP2 / RMSD in Å</th>
<th>E_{int} (kcal)</th>
<th># of H-bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray</td>
<td>-103</td>
<td>8</td>
</tr>
<tr>
<td>1 / 1.31</td>
<td>-113</td>
<td>9</td>
</tr>
<tr>
<td>2 / 1.49</td>
<td>-90</td>
<td>10</td>
</tr>
<tr>
<td>3 / 1.53</td>
<td>-107</td>
<td>12</td>
</tr>
<tr>
<td>4 / 1.57</td>
<td>-81</td>
<td>6</td>
</tr>
<tr>
<td>5 / 1.68</td>
<td>-81</td>
<td>12</td>
</tr>
<tr>
<td>6 / 1.79</td>
<td>-32</td>
<td>7</td>
</tr>
<tr>
<td>7 / 1.87</td>
<td>-103</td>
<td>13</td>
</tr>
<tr>
<td>8 / 1.95</td>
<td>-105</td>
<td>12</td>
</tr>
<tr>
<td>9 / 2.03</td>
<td>-30</td>
<td>5</td>
</tr>
</tbody>
</table>

Some of the reasons for this lack of correlation might come from the following reasons. First, the method used in calculating the energies may not be sufficient. More thorough methods can be used for this purpose. For example, ZAP program [95] can be used to calculate the binding free energy after scaling the energy expression for the specific system used in this research, i.e., the peptide/MHC complexes. Second, the
docking method may not be accurate enough. Since the fitness calculation used in GOLD program is H-bonding oriented, it may not be able to locate the experimental data accurately for hydrophobic ligands. Also, solvation effect is not added in docking calculations in GOLD, and the receptor has a high degree of rigidity.
Chapter 6

Conclusions

The binding free energy values from docking studies correlate weakly with the experimental values and with the number of H-bonds between the ligand and MHC receptor. We find that hydrogen bonding is not the only factor affecting the binding. Other forces like hydrophobic interactions and charged-charged interactions are in operation.

Binding of GP2 to HLA-A2.1 has been improved by selective mutations. The center residues have been mutated so that the main problem of flexibility that GP2 has in the center can be addressed. The center residues of GP2 are bulged out of the binding cleft and do not make stabilizing contacts with the residues of the receptor in the binding groove. Based on these predictions the ligand IISAVVPIL has been studied experimentally and found that it increased the binding 68%.

Although the X-ray structure of GP2 was reproduced by flexible docking calculations in GOLD, the results need further work to calculate the binding free energies. For
example, parameters in the binding energy expression of ZAP program can be scaled to fit the known experimental values of peptide/MHC complexes, and then the binding energies of docked GP2 models can be calculated utilizing ZAP.

In conclusion, the experimental data was reproduced qualitatively by using fast docking calculations, and these findings can be useful for more thorough computational methods like molecular mechanics.
Chapter 7

Bibliography


[3] Rammensee, HG.; Friede, T.; Stevanovic, S. MHC ligands and peptide motifs: 
First listings. *Immunogenetics*, **1995**, 41, 178-228

Wiley, D.C. The foreign antigen binding site and T cell recognition regions of 

epitope (GP2) to HLA-A2.1 is due to a lack of interactions with the center of the 

project. Curr. Opin. in Immunol., 1999, 11, 209-213


[24] Altuvia, Y.; Schueler, O.; Margalit, H.; Ranking potential binding peptides to MHC proteins by a computational threading approach. J. Mol. Biol., 1995, 249, 244-250


21449

[38] Gopalakrishnan, B.; Roques, B.P. Do antigenic peptides have a unique sense of direction inside the MHC binding groove? A molecular modeling study. *Federation of Europ. Biochem. Societies, 1992, 303(2,3), 224-228*


[61] Murray, C. W.; Eldridge, M. D. Empirical scoring functions. II. The testing of an


[70] Kelly, K. MOE-Dock; 2002.03 ed.; Chemical Computing Group, Inc.: Montreal,
Canada H3A 2R7

[71] Molecular Orbital Environment; 2002.03 ed.; Chemical Computing Group, Inc.: Montreal, Canada H3A 2R7


[74] FlexX; Tripos, Inc.; Missouri, USA

[75] Tripos, Inc.; M issouri, USA


[79] 3D-Dock Suite: FTDock, RPSCore and MultiDock; Biomolecular Modeling Laboratory, Cancer Research Institute, London, UK

[80] Judson R. Genetic algorithms and their use in chemistry. In *Reviews in*


