COMPUTATIONAL STRUCTURE ACTIVITY RELATIONSHIP STUDIES ON THE CD1D/GLYCOLIPID/TCR COMPLEX USING AMBER AND AUTODOCK

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Janos Nadas

Graduate Program in Chemistry

The Ohio State University

2009

Dissertation Committee:

Peng George Wang, Advisor

Avraham Benatar

Christopher Hadad

Chenglong Li

ABSTRACT

The human immune system is very powerful and whose one function is to detect and eliminate foreign, pathogenic compounds that enter the body. The conventional major histocompatibility complexes (MHC) are able to bind peptides and present them to the T-cell lymphocytes thereby allowing the cell to communicate whether it is healthy or has been compromised. A different category of T-cell known as natural killer T (NKT) cells play an important role in bridging the innate and the adaptive immune systems, where NK cells and conventional T-cells exist, respectively. The innate immune response of these specific NKT cells has been associated with tumor rejection activities with the adaptive immune response being associated with protection against primarily bacterial infections, but also with viral and parasitic attacks.

These NKT cells are also unique in that their T-cell receptor (TCR) proteins recognize foreign glycolipid antigens, not peptides, presented by MHC-I like Cluster of Differentiation 1 (CD1) molecules. The marine-sponge derived glycolipid α-galactosylceramide (α-GalCer) has so far been the most potent iNKT stimulatory ligand when presented by the CD1d protein. It is more than unusual that an α-glycolipid derived from a marine sponge could yield such a massive immune response in humans, since it is doubtful that humans have evolved with a defense mechanism against a possible invasion of marine sponges. Currently, the glycosphingolipids from the alpha-proteobacteria *Sphingomonas* are considered the natural foreign ligands for the system since they were

found to activate iNKT cells but to a lesser degree than α-GalCer.

The massive immune response cascade that follows after CD1d presentation of a glycolipid to iNKT cells has yielded a search for a better ligand with either comparable activity as α-GalCer but with less of its pharmaceutical hindrances or a ligand that can control the immune response. To date, superficial structure-activity relationships have been defined wherein modifications to either the sphingosine chain or acyl chain of the lipid can lead to a bias in the immune response, and modifications to the galactose sugar have led to null activity.

Herein, this scientific project entailed the use of computational means to determine how the TCR protein of iNKT cells can differentiate so selectively between glycolipids presented by the CD1d protein in order to be able to design a better ligand for the system. Molecular dynamics simulations using AMBER found the crystallized CD1d/α-GalCer/TCR tertiary complex to be stable and relatively rigid in explicit solvent. A combination of high-level docking with AUTODOCK and simulation showed that modifications to the 2′- and 3′- positions of the galactose sugar are indeed not tolerated, whereas, modifications to the 4′- position were semi-tolerated. The simulations of CD1d/glycolipid binary complexes showed that glycolipids incapable of stimulating iNKT cells changed the direction of the CD1d residues that interact with TCR away from optimum orientation. Lastly, the evidence of a non-glycolipid ligand activating iNKT

cells led to an undertaking of a virtual screening program to find a replacement for the galactose sugar yielding a library of viable aromatic-based lipid ligands.

DEDICATION

Dedicated to the Nadas Family, A Group of Wonderful Intellectuals

"A picture is worth a thousand experiments"

ACKNOWLEDGEMENTS

The author would like to acknowledge first and foremost his advisor, Professor Peng George Wang, for all his guidance and encouragement. He would also like to acknowledge his fellow lab mates who have helped on this research project: Wenlan Alex Chen, Dr. Chengfeng Charles Xia, Wenpeng Zhang, and Yalong Zhang; and Dr. Chenglong Li for his guidance with the computational methodology. The Department of Chemistry and the Department of Biochemistry are acknowledged as well as The Ohio State University for providing such a wonderful learning and research environment. Lastly, the Ohio Supercomputer Center is acknowledged for without their resources and assistance this project could not have been accomplished.

VITA

EDUCATION

December 2009…….. Ph.D. Chemistry, The Ohio State University

May 2004……... B.S. Chemistry, Duke University

PUBLICATIONS

Nadas, J.; Li, C.; Wang, P.G. Computational structure activity relationship studies on the CD1d/glycolipid/TCR complex using AMBER and AUTODOCK. *J. Chem. Inf. Mod.* **2009**, 49, 410-423.

Chen, W.; Xia, C.; Wang, J.; Thapa, P.; Li, Y.; Talukdar, A.; Nadas, J.; Zhang, W.; Zhou, D.; Wang, P.G. Synthesis and structure-activity relationship study of isoglobotrihexosylceramide. *J. Org. Chem.* **2007**, 72, 9914-9923.

Dongning, L.; Nadas, J.; Zhang, G.; Johnson, W.; Zweier, J. L.; Cardounel, A. J.; Villamena, F.; Wang, P. G. 4-Aryl-1,3,2-oxathiazolylium-5-olates as pH-controlled NO-donors: the next generation of S-nitrosothiols. *J. Am. Chem. Soc.* **2007**, 129, 5503-5514.

Battisti, R. F.; Zhong, Y.; Fang, L.; Gibbs, S.; Shen, J.; Nadas, J.; Zhang, G.; Sun, D. Modifying the sugar moieties of daunorubicin overcomes P-gp-mediated drug resistance. *Mol. Pharmaceutics* **2007**, 4, 140-153.

Nadas, J.; Sun, D. Anthracyclines as effective anticancer drugs. *Expert. Opin. Drug. Discovery* **2006**, 1, 549-568.

Fang, L.; Zhang, G.; Chenglong, L.; Zheng, X.; Zhu, L.; Xiao, J. J.; Szakacs, G.; Nadas, J.; Chan, K. K.; Wang, P.G.; Sun, D. Discovery of a daunorubicin analogue that exhibits potent antitumor activity and overcomes P-gp-mediated drug resistance. *J. Med. Chem.* **2006**, 49, 932-941.

Pan, Y.; Ayani, T.; Nadas, J.; Wen, S.; Guo, Z. Accessibility of *N*-acyl-D-mannosamines to *N*-acetyl-D-neuraminic acid aldolase. *Carbohydr. Res.* **2004**, 339, 2091-2100.

FIELDS OF STUDY

Major Field: Chemistry

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

IMMUNOLOGY BACKGROUND INTRODUCTION

ANTIGEN PRESENTING PROTEINS

Conventional major histocompatibility complexes (MHC) are proteins whose function is to provide a means by which the immune system can survey cells. By binding peptides and presenting them to T-cell lymphocytes, the cell is able to communicate whether it is healthy with the presentation of self peptides or compromised by presenting peptides of a foreign origin such as those that would come from a microbial infection[1]. Specifically, class II and class I MHC proteins present peptides to the $CD4^+$ and $CD8^+$ T-cells of the immune system, respectively[2-4]. Upon recognition, $CD4^+$ cells secrete cytokines and regulate the responses of other cell types, whereas the $CD8^+$ cells even though they also secrete cytokines are in themselves potent killers of cells[5]. Unlike the recognition event between an antibody and antigen, the complexation between the peptide presenting MHC (pMHC) protein and the T-cell receptor (TCR) protein of a T-cell lymphocyte is substantially weaker existing usually in the low micromolar range[6].

A third category of T-cell has also been observed. These have been named natural killer T (NKT) cells for their co-expression of TCR proteins and other surface receptor proteins that are typical of natural killer (NK) cells[7]. The NKT cells play an important

role in bridging the innate and the adaptive immune systems, where NK cells and conventional T cells exist, respectively[5]. The unique nature of these NKT cells further lies in the fact that they recognize self and foreign glycolipid antigens presented by MHC-I like CD1 molecules and not peptides[8, 9]. These MHC-1 like presenting proteins were designated Cluster of Differentiation 1 (CD1) molecules based on their leukocyte staining characteristics[10]. Five classes of CD1 proteins have been identified in humans based on sequence homology, comprised of group I (CD1a, CD1b, CD1c), group II (CD1d), and an outlying group (CD1e)[11]. Only one or two types of all five CD1 proteins exist in other species, though in mice, there is only a single class of CD1d protein (mCD1d) which is homologous to the human isoform CD1d[12]. The variable presence of CD1 proteins between species has been attributed to the animals evolving only those CD1 isoforms within specific intracellular trafficking routes that satisfied a need to sample specific antigens within those routes[13].

The intracellular trafficking route of the CD1 proteins has been found to be similar to that of the MHC proteins (**Figure 1**)[9]. After synthesis, the CD1 proteins are translocated into the lumen of the endoplasmic reticulum after which they are diverted to the trans-Golgi network and secreted onto the plasma membrane. It is believed that self lipids are loaded onto the CD1 molecules within the ER and are bound throughout the trafficking process, however, the process is extremely thermodynamically unfavorable and more than likely requires assistance from mediators in the endocytic pathways[14]. To focus the route to the topic at hand, the CD1d protein, the protein in question, is then internalized in a clathrin-coated pit via interaction with the adaptor complex AP2 and is trafficked to the late endosomal and lysosomal compartments via interaction with an AP3 protein. The

AP3 protein interacts with tyrosine-based motifs in the cytoplasmic tail of CD1d whereupon if the tail is truncated then trafficking to the lysosomes does not occur and NKT cells are not stimulated[15]. Additionally in AP3 deficient mice, increased levels of CD1d protein were found on the cell surface and the antigen presentation ability of cells was dramatically decreased[14]

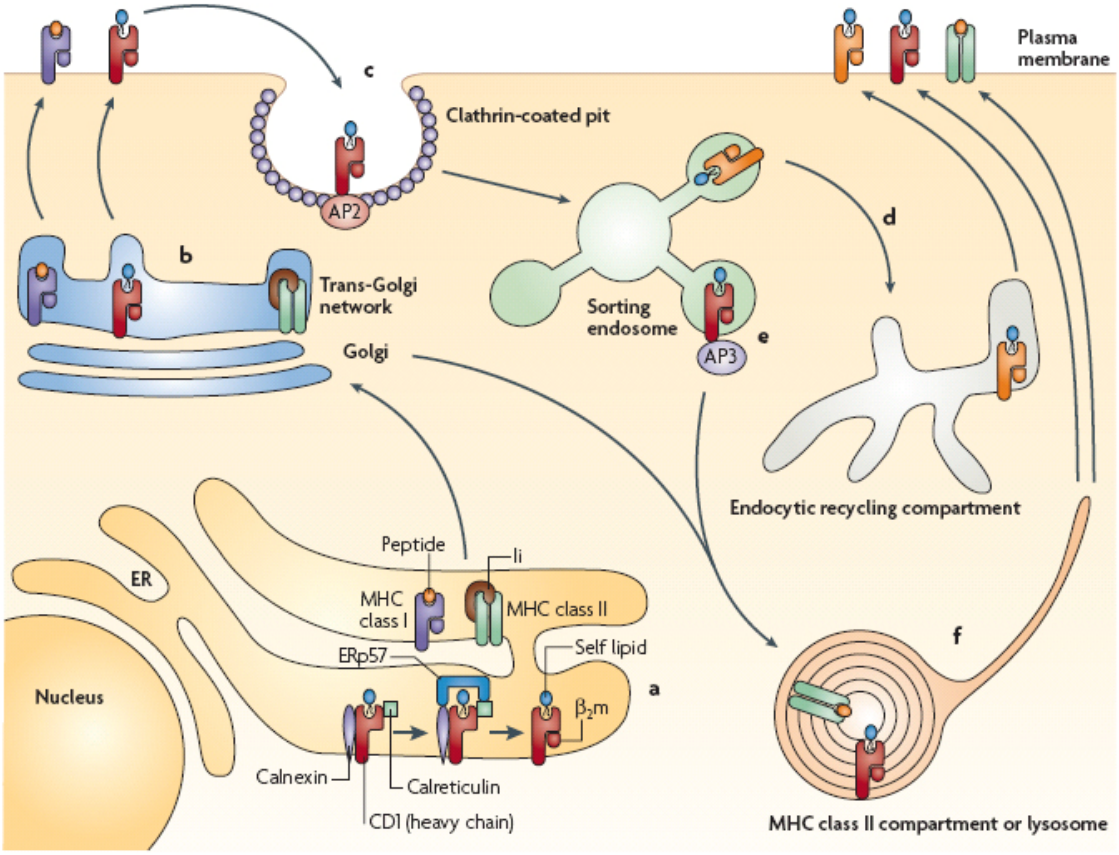


Figure 1. Intracellular trafficking of CD1 molecules within humans. (a) Assembly in ER. (b) Secreted through Golgi apparatus. (c) CD1 molecules internalized. (d) CD1a and CD1c can follow slow recycling path. (e) CD1b and CD1d traffic to late endosomal and lysosomal compartments via AP3. (f) Molecules recycled to plasma membrane. During the entire trafficking even, CD1 molecules are thought to be loaded with a self lipid molecule. *Reproduced from Barral,D.C.; et. al. Nature Rev. Immunol. 2007.*

CD1 proteins are able to survey the entire cell for foreign lipids through differential endosomal and lysosomal trafficking. Foreign lipids are continuously taken up by cells for their metabolic needs. Even though our understanding of cellular lipid trafficking is still very incomplete, it is known that lipid loading for CD1 proteins occurs within lysosomal compartments (**Figure 2**)[9, 15, 16]. Foreign lipids, usually of microbial origin, are loaded onto CD1 proteins via saposins which are small, non-enzymatic proteins that facilitate the hydrolysis of a variety of sphingolipids within the lysosome; specifically for CD1d, saposin B was shown to facilitate its antigen presentation[17].

The CD1d proteins present their foreign antigens specifically to CD1d-restricted invariant natural killer T (iNKT) cells' TCR proteins (**Figure 3**)[18]. If the foreign antigen is of harmful or pathogenic origin an immune response is generated. The innate immune

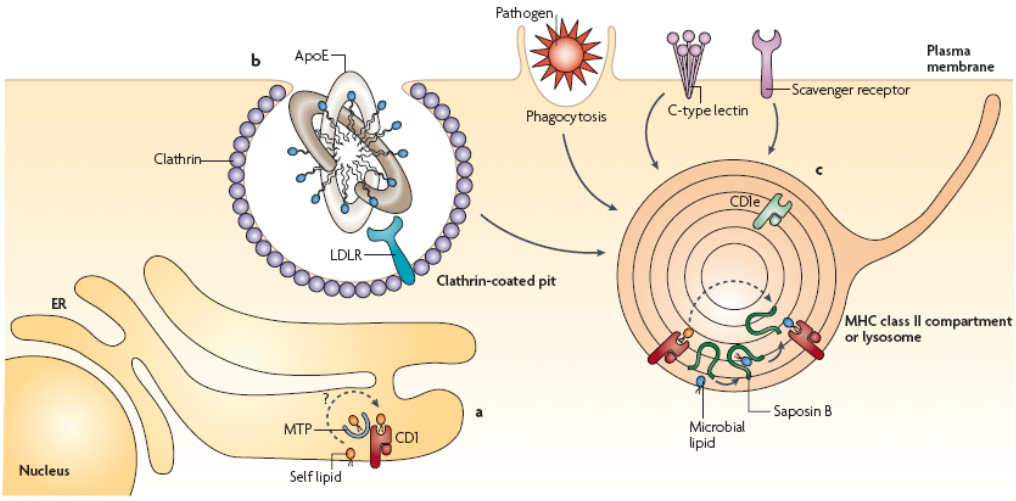


Figure 2. Loading of lipid antigens onto CD1 proteins. (a) Self lipids are loaded with help from MTP protein. (b) Four mechanisms through which foreign lipids enter the cell: (1) clathrin-dependent internalization of apoliprotein E-lipid complexes; (2) phagocytosis; (3) C-type lectin binding; and (4) scavenger receptor internalization. (c) The exchange of self-lipids with foreign lipids takes place in lysosomes. *Reproduced from Barral,D.C.; et. al. Nature Rev. Immunol. 2007.*

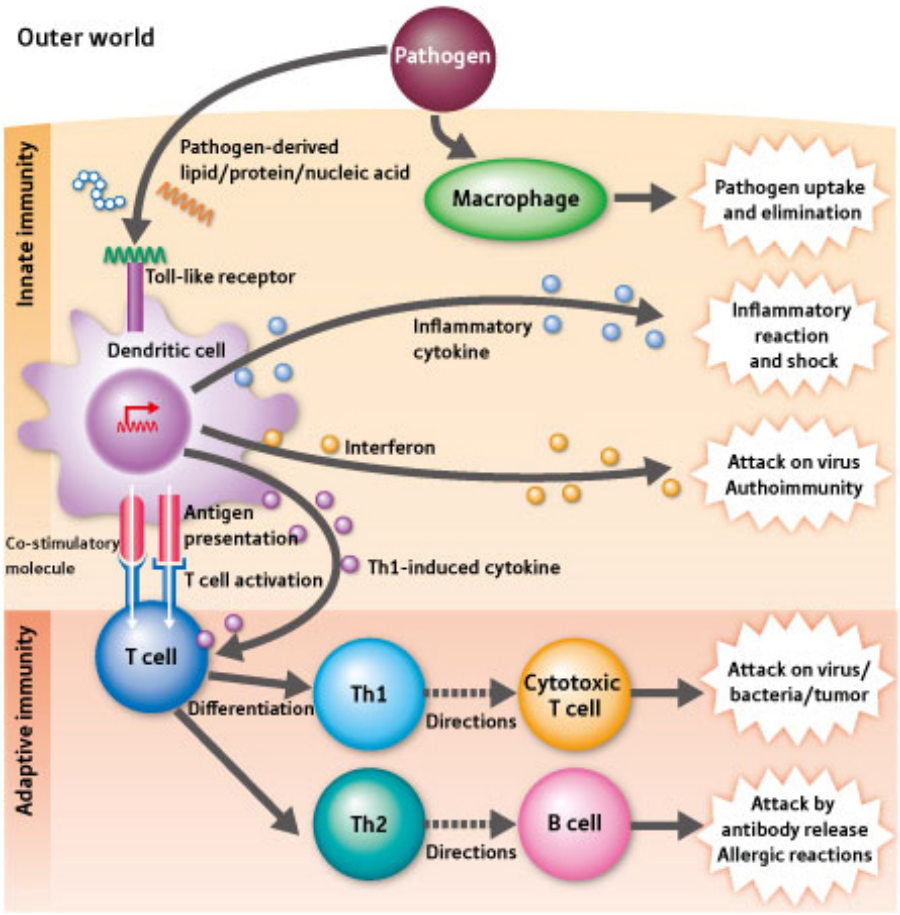


Figure 3. Bridge between innate and adaptive immunity. Upon CD1d antigen presentation to iNKT cells' TCR proteins a cascade of cytokines, interferons, and Th1 and Th2 cytokines are released leading to a very strong immune response to a foreign pathogen such as a virus, bacteria, or even tumor cells. The presentation is thought to be the bridge between the innate and adaptive immunity of the human immune system. *Adapted from Riken Research Website.*

response of iNKT cells is primarily associated with tumor-rejection activities and is relatively complex with reciprocal stimulation existing between iNKT cells and dendritic cells (DC). Upon TCR binding and recognition, iNKT cells produce such cytokines as interferon-γ (IFN-γ) and interleukin-4 (IL-4). The IFN-γ production accomplishes two things with the first being the activation of macrophages and the second being the stimulation of DC maturation, which causes increased IL-12 production which in turn

further activates the iNKT cells[19, 20]. This self-amplification cycle allows for a broad immune response capable of being generated by minimal antigen presentation. Furthermore, the complexity of the response increases wherein the chemokine profiles are variable depending on the type of iNKT cell stimulated. For example, $CD8^+$ cells more likely produce Th1-type cytokines whereas $CD4^+$ cells produce both Th1-type and Th2-type chemokines, and $CD4^-$ cells are those whose response is primarily associated with broad tumor rejection activities[21, 22].

The antigen presentation though does has a very negative side-effect where iNKT cells become unresponsive to continued antigen presentation. Typical adaptive immune responses cause a long-term memory to be associated with the specific peptide antigen presented, however, with iNKT cells this does not appear to happen. Upon stimulation, the iNKT cell population can rapidly proliferate while simultaneously down-regulating their TCR proteins which then results in a long-lasting depletion and cell apoptosis[23]. The process of rapid proliferation happens within 3 days of antigen presentation with cell apoptosis increasing for the next 7 days until a steady-state population is reached[24]. This effect of iNKT cell unresponsiveness correlating to their diminished proliferation, minimal production of cytokines, and inability to metastasize tumors resembles the result of when conventional T cells are activated by strong stimuli as from a "superantigen." It has been termed a state of anergy[25, 26].

The adaptive immune response of CD1 molecules is associated with protection against primarily bacterial infections, but also with viral and parasitic attacks[5]. For example, $CD1d^{-/-}$ mice infected with *Pseudomonas aeruginosa* yielding a pulmonary infection and then treated with anti-CD1d antibodies were markedly less able to clear bacteria from

their lungs[27]. Similar CD1d$^{-/-}$ mice infected with skin-based herpes simplex virus type 1 (HSV-1) and genital type 2 (HSV-2) were shown to be more susceptible to infection[28, 29]. In humans, it has been observed that these viruses along with HIV led to a decrease in CD1d expression albeit each by different mechanisms[30]. When *Jα18*$^{-/-}$ mice which are incapable of forming the necessary CD1d-restricted iNKT TCR proteins were infected intraperitoneally with *T. cruzi*, a type of parasite, a significant increase in the production of cytokines in the spleen cells was observed showing that the mice died from immunopathology as opposed to parasite infection[31]. This data added to the complexity of the role of iNKT cells whereby stimulation lead to an anti-inflammatory role rather than an inflammatory one.

STRUCTURE AND RECOGNITION OF pMHC and TCR PROTEINS

In order to understand the interaction between CD1d and TCR proteins which is the crucial point in this immune response pathway, it is necessary to first elucidate the mechanism of interaction between pMHC (peptide presenting) and TCR proteins. Structurally, MHC class I and II have similar arrangements where they are made up of three domains, one α-helix / β-sheet (α β) superdomain forms the peptide binding site and two immunoglobulin G (IgG)-like domains. The peptide antigen which is usually 8-10 residues long binds in the shallow groove created by the α1 and α2 helices of the heavy chain, with longer peptides extending out of the C-terminus or bulging vertically out of the groove (**Figure 4**)[32, 33]. The hydrogen bonding network between the α1 and α2 helices dictates peptide binding and specificity for the pMHC proteins with rare exceptions being possible[34].

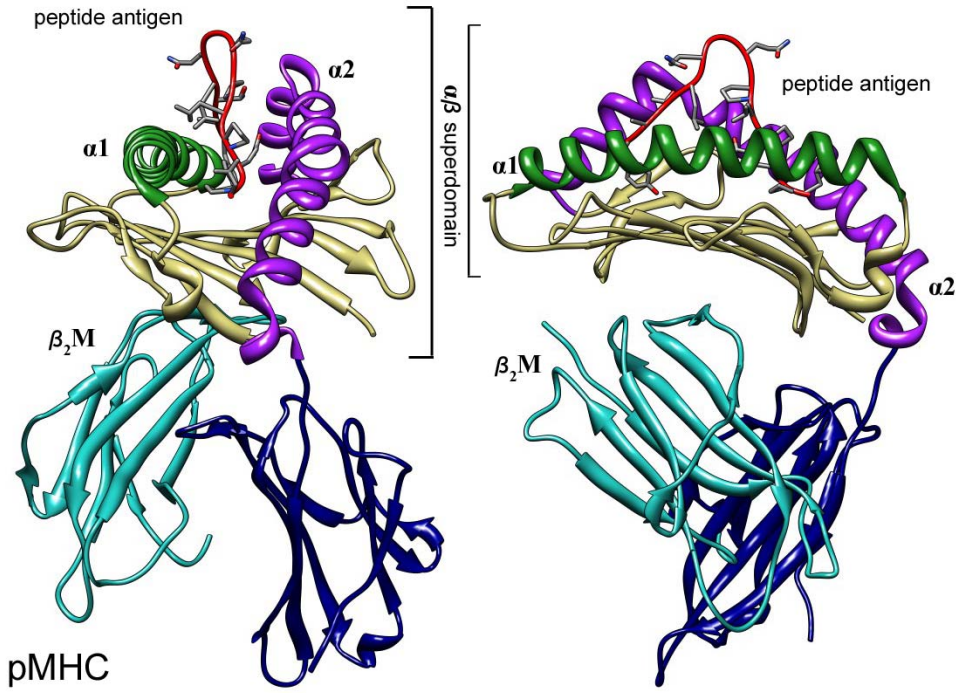


Figure 4. Crystal Structure of pMHC with peptide antigen. The pMHC protein is comprised of an one α-helix / β-sheet (α β) superdomain which makes up the peptide binding groove along. Two IgG-like β sheets comprise the rest of the transmembrane protein. The peptide antigen bound is shown to bulge out of the binding groove. (*PDB ID# 1ZHL*)

From the available crystal structures, it would appear that peptide antigen orientation adjustments are permissible upon their presentation (**Figure 5**). A peptide is capable of maintaining its presented orientation even if it extremely bulged out of the pMHC binding groove as was found in the crystal structure TCR-HLB-B*3508-LPEP complex[35]. If it is bulged, it is also capable of flattening out upon recognition as was observed in the crystal structure of LS4-HLA-B*3501-EPLP[36]. The most peculiar case that shows the flexibility of peptide presentation is in the Tax-5K-IBA/HLA-A2 complex when a peptide antigen was elongated and bulked up at its point of recognition and upon TCR binding this modified sidechain folded back over itself to accommodate the TCR protein[37].

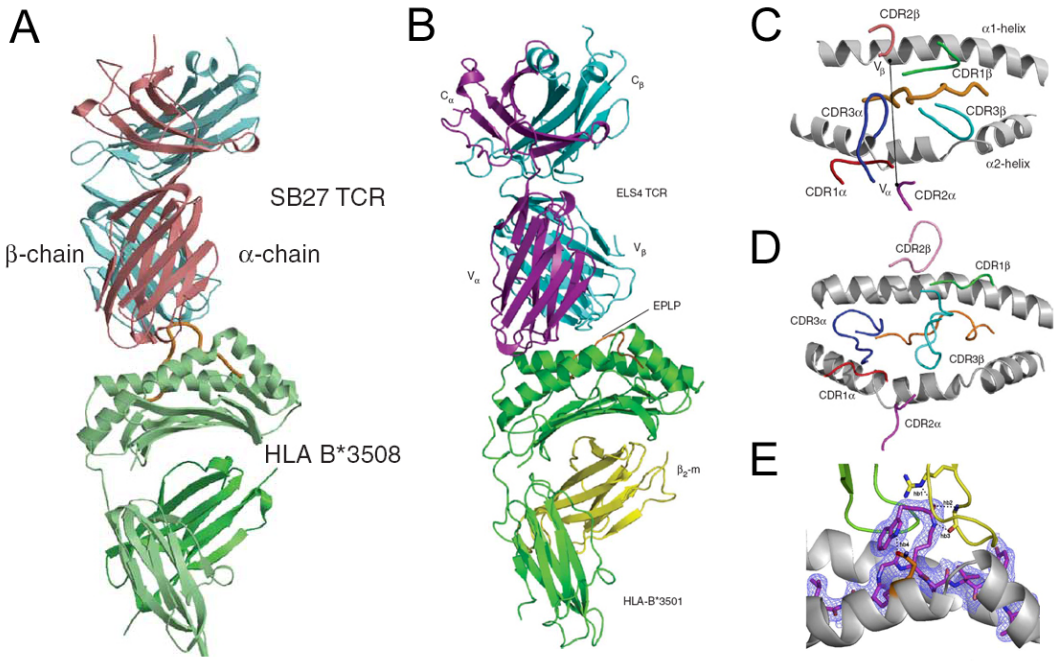


Figure 5. Assortment of pMHC/TCR binding footprints. (A) The TCR-HLB-B*3508-LPEP complex. (B) The LS4-HLA-B*3501-EPLP complex wherein the peptide bulge is flattened. (C) The binding footprint of TCR-HLB-B*3508-LPEP showing the CDR loop interactions with the α helices of pMHC. (D) The LS4-HLA-B*3501-EPLP binding footprint showing a different CDR loop interaction scheme. (E) An up-close view of the Tax-5K-IBA/HLA-A2 complex wherein the peptide folded back upon itself during TCR recognition. *The images are reproduced from the cited sources.*

The TCR protein structure make-up has evolved to anticipate any encounter with a variety of possible antigens as would become a protein which plays an inherent role in adaptive immune responses[38]. The structural portions of TCR are named after the gene segments from which they are randomly assembled from, variable (V), diversity (D), and joining (J) segments. TCRs within humans and mice consist of non-covalently associated α- and β- chains which are divided into the amino-terminal V region that is homologous to the V domains of immunoglobulins and the carboxy-terminal constant (C) region. The Vα and Vβ domains are responsible for pMHC/peptide antigen recognition through the non-covalent interactions of loops termed complementarity-determining regions (CDR)[39].

In 1999, I.A. Wilson proposed general principles for TCR recognition of pMHC proteins that still holds true[39]: (1) Most importantly, the diagonal binding footprint of TCR over pMHC is observed throughout all the crystal structures even if the actual contacts vary widely[40]; (2) The α1 and α2 helices are conserved for pMHC proteins allowing for ease of TCR sampling of different bound peptide antigens, approximately 10^4-10^6 different associations have been estimated[41]; (3) All bound peptide antigens do interact with TCR primarily at their centers; (4) The binding footprint shows extreme plasticity with TCR being able to adapt to different bound ligands leading to a variety of biological outcomes[37]; (5) The Vα chain primarily CD1α and CD2α loops appear to be the primary driving force that orients the TCR onto the pMHC protein; (6) All the CDR loops are capable of conformational changes thereby expanding TCR specificity[42]; (7) None of the crystal structures show large-scale conformational changes within the proteins upon recognition; (8) The recombinant building of TCR proteins by varied Vα and Vβ chain pairings provides an added level of TCR specificity; and lastly (9) All the complex pMHC/TCR protein structures are monomeric and stable even though clustering of the complexes does appear to be an integral part of the TCR signaling event.

An addendum to this list of principles is that binding kinetics, specifically the half-life, appears to play an important role concerning the pMHC and TCR recognition event where it has been shown to correlate to the variety of immune response profiles generated. For a long time, enthalpic or entropic contributions appeared to provide an explanation for the differences concerning the pMHC and TCR recognition, however, there has only been a narrow overall binding free energy amongst all the recorded complexes with no substantial correlation between structural features and TCR-binding

thermodynamics[43]. On the other hand, binding kinetics were correlated to T-cell activation where the activation depended on a pMHC/TCR half-life that is neither too short nor too prolonged[44, 45]. In regards to the short half-lives, T-cell activation occurred only when the half-life threshold was reached implying that a sufficiently long interaction is necessary to complete the intracellular signaling cascade[46]. Those interactions that do not reach the required half-life threshold do not cause activation such as those with bound self-peptides and not those of foreign origin. On the other hand an unusual occurrence was found to explain activation in regards to the too long half-lives. Since pMHC proteins were found to exist in relatively low density on the cell surface but only 1-50 are needed to activate T cells, it was proposed that upon TCR binding a single pMHC protein is internalized and another TCR, up to ~200, is then bound to the same pMHC protein[47, 48]. Therefore, if the half-life binding event was too prolonged then it would block the next TCR from interacting causing a blockade and the signaling cascade event would fade out[49].

STRUCTURE AND RECOGNITION OF CD1d and iNKT TCR PROTEINS

The overall structure of CD1d and iNKT cell TCR proteins differ very minimally from those of pMHC and T-Cell TCR proteins. CD1 proteins are closely related to the class I pMHC proteins wherein they also consist of an αβ superdomain that is associated with a β2 microglobulin (β_2M). The specific subclass of TCR proteins that are CD1d-restricted exist on invariant natural killer T (iNKT) cells. The term invariant refers to the Vα chain that comprises half of the TCR protein whereby only the Vα14-Jα18 chain paired with Vβ8.2 or Vβ7 chains for mice or the Vα24-Jα18 chain paired with a Vβ811 chains for

humans is able to recognize glycolipid presenting CD1d proteins[50-52]. As compared to the Vβ chain, the invariant Vα chain is primarily responsible for recognition of the CD1d presented glycolipid since $V\alpha^{-/-}$ or $J\alpha^{-/-}$ populations were unable to stimulate iNKT cells[53-55]. This reliance on the Vα chain has caused more extensive research to be performed on the Vβ chains where it was found to influence selectivity and avidity to the CD1d/glycolipid complexes but unable to abolish stimulation of iNKT cells[56, 57]. Studies have shown preference amongst the Vβ chains where in the mouse the Vβ8.2 conferred higher avidity for binding than the Vβ7 chain[58]. Even though the human and mouse iNKT TCR proteins are highly selective being restricted to CD1d, they happen to be cross-species reactive with Vα and Vβ chains being interchangeable and able to recognize either CD1d proteins from either species[12, 59].

The first crystal structure solved of CD1d at a resolution of 2.8 Å without a ligand was of murine origin (mCD1d) and it deviated from pMHC complexes in possessing a narrower, deeper, and extremely hydrophobic binding groove[60]. The depth of the binding groove was caused by the α2 helix being in closer proximity to the α1 helix causing the α1 helix to rise further from the β-sheet thereby accentuating the kink in the middle of it. The length of the groove is around ~30 Å which was slightly longer than for the pMHC proteins where they were typically around 25 Å long. The groove was narrower being only 10-15 Å wide and 7 Å deep with both ends being closed showing that the lipid loading must occur from the center (top-loading) of the binding groove rather than through its sides. The binding groove of CD1d was smaller than those of other CD1 proteins studied by possessing roughly 75% of the capacity, 1,650 $Å^3$ as compared to 2,200 $Å^3$ for human CD1b[61].

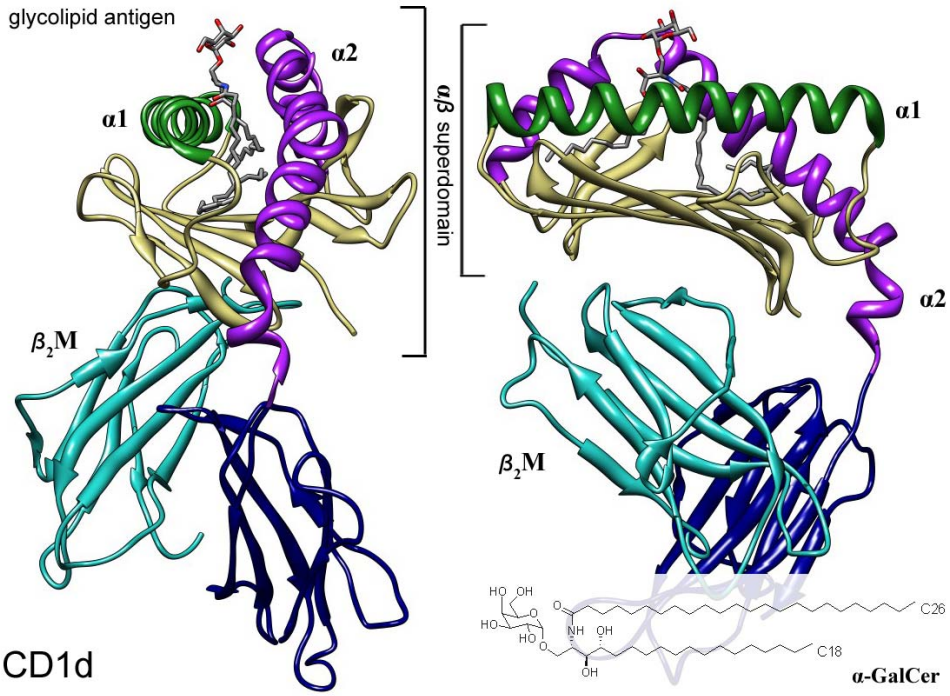


Figure 6. Crystal Structure of human CD1d with the α-galactosylceramide (α-GalCer) glycolipid antigen. The CD1d protein is similar in structure to the pMHC proteins except that the binding groove is much deeper and narrower due to the proximity of the α-helices to each other. The molecular structure of α-GalCer is also shown for clarity. (*PDB ID# 1ZT4*)

In 2005, human CD1d (hCD1d) was crystallized with and without the most potent, capable of generating the most robust immunological stimulation profile, antigen , α-galactosylceramide (α-GalCer)[62, 63], 12 years after the antigen was discovered and 8 years after the mCD1d crystal structure was solved (**Figure 6**)[64]. The total volume of the binding groove was 85% less than for the mCD1d being 1,400 $Å^3$, yet both mouse and human CD1d proteins are capable of presenting the glycolipid α-GalCer to iNKT TCR proteins of either species showing extreme conservation within mammalian evolution[12]. A more detailed structure-activity analysis of α-GalCer and other glycolipid antigens will be presented in the next section, but in order to fully describe CD1d antigen presentation it is important to look first at the non-covalent interactions which play a role in binding

the glycolipid.

The binding groove of hCD1d appeared to be built to fit α-GalCer almost perfectly. The ligand is specifically a glycosphingolipid (glycolipid for short) whose lipid is made up of a 18 carbon phytosphingosine chain and a 26 carbon acyl chain with the sphingosine chain inserting into the C′ pocket and the acyl chain inserting and rotating around counter-clockwise in the A′ pocket. Three important hydrogen bonds anchored and oriented the glycolipid within the groove: (1) the 2′-OH of the galactose ring, which is crucial for the antigenicity, was hydrogen-bonded with Asp151; (2) the 3-OH on the sphingosine chain formed a hydrogen bond with Asp80; and (3) the glycosidic linkage 1′-*O* formed the third hydrogen bond with Thr154 (**Figure 7A**).

Over the last few years, the only structural data available were the binary crystal structures of CD1d with the ligand antigens α-GalCer[64], α-galacturonosyl ceramide (GalA-GSL)[65], synthetic variant of α-GalCer wherein the acyl chain is truncated down to 8 carbons (PBS-25)[66], 3′-sulfogalactosyl ceramide (sulfatide)[67], the self-lipid phosphatidylcholine (PC)[68], and the complex mycobacterial phosphatidylinositol-dimannoside (PIM2)[69] (**Figure 7**). There are some important differences between the crystal structures, to begin with mCD1d lacks the tryptophan residue at position 153 and instead it has a glycine residue. The Trp153 in hCD1d provides a non-covalent platform upon which the nonpolar back of the galactose sugar sits whereas the mCD1d cannot provide this steric bulk behind the sugar. Due to the difference in amino acid, a change in the galactose sugars of PBS-25 (**Figure 7D**) and GalA-GSL (**Figure 7E**) can be seen where both were slightly angled back towards the A′ pocket, whereas with α-GalCer the

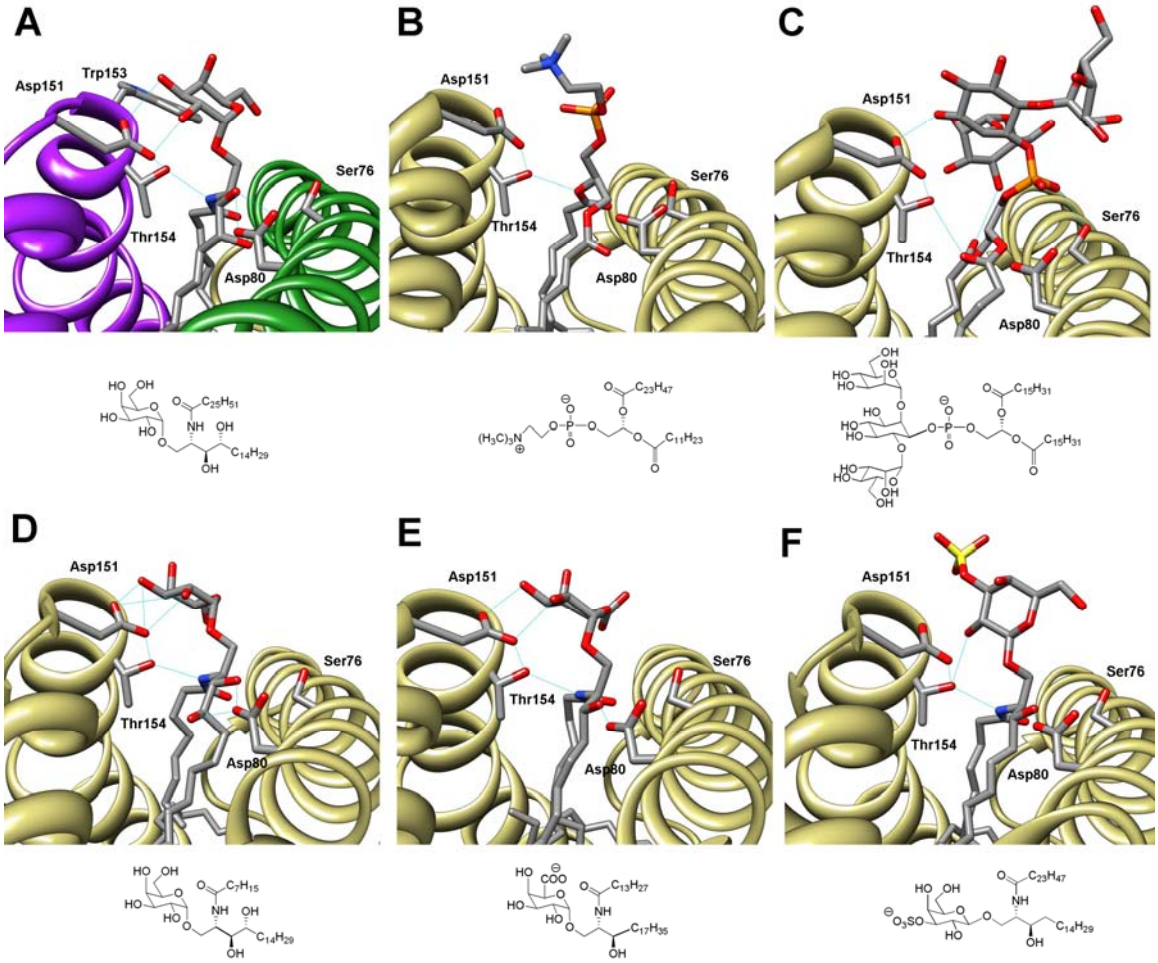


Figure 7. Crystal structures showing the hydrogen bonds between the α helices and a variety of antigens and a self-lipid. (A) The hCD1d/α-GalCer binary complex showing the important hydrogen bonds between the glycolipid and the residues on hCD1d. [*1ZT4*] (B) The mCD1d/phosphatidylcholine (PC) self-lipid binary complex is missing many of the important hydrogen bonds that play a role in anchoring the other antigens. [*1ZHN*] (C) The mCD1d/phosphatidylinositol-dimannoside (PIM2) binary complex wherein the ligand is plays an agonist role. [*2GAZ*] (D) The mCD1d/PBS-25 binary complex is capable of maintaining all the necessary hydrogen bonds even though it possesses a truncated acyl chain. [*1ZTL*] (E) The mCD1d/α-galacturonosyl ceramide (GalA-GSL) binary complex has similar hydrogen bonds as for the hCD1d/α-GalCer complex. [*2FIK*] (F) The mCD1d/3´-sulfogalactosyl ceramide (sulfatide) binary complex is able to maintain the necessary hydrogen bonds even though the glycosidic linkage is β rather than α. [*2AKR*] *PDB ID # are in brackets.*

sugar was angled towards the C´ pocket. This backward angling can also be attributed to the shortened acyl chains on both PBS-25 and GalA-GSL causing them to sit deeper in the C´ pocket. Otherwise, the hydrogen bonds are maintained in all the antigens with the exception that the Thr154 residue appeared in the images to make a hydrogen bond with

the amide nitrogen rather than the glycosidic bond. This discrepancy between image and report was further shown to be the case when α-GalCer and a few deoxy-sphingosine derivatives bound to hCD1d underwent molecular dynamics simulations and showed Thr154 hydrogen bonded to the amide nitrogen on the acyl chain[70].

The most striking difference structural difference between the bound molecules are those found when comparing the self and non-self ligands. The self-lipid phosphatidylcholine (PC) retained only the hydrogen bond to Thr154 by its ester oxygen and seemed to be held in place by the hydrophobic interactions between its lipid tail and the CD1d binding groove (**Figure 7B**). Since PC has been shown to be a self-lipid for mCD1d but incapable of stimulating iNKT cells, it is possible that its lessened hydrogen bonding allows for ease of unloading upon antigen loading[71, 72]. The more complex phosphatidylinositol-dimannoside (PIM2) ligand derived from the plasma membrane of mycobacteria was thought to be the natural foreign antigen for the system, but even though it was crystallized in mCD1d, it has not been found to stimulate iNKT cells (**Figure 7C**)[69, 73, 74]. Nonetheless, the fact that phospho-based lipids are capable of being bound by CD1d does show that it is capable of binding a diverse array of lipids.

The myelin derived 3′-sulfogalactosyl ceramide (sulfatide) was also capable of binding to and being crystallized with CD1d[67]. The β- anomeric linkage between the sugar and lipid portion caused the sugar to flip up whilst still being able to maintain its hydrogen bonds to all the amino acids that form contacts with α-GalCer with the exception of the 3′- sulfo substituent that no longer is hydrogen bonded to Asp151. Nevertheless, no β-linked monoglycosylceramide has been shown to be able to elicit any iNKT activity[55, 75]. A more in-depth description of a variety of other natural and synthetic lipids in regards to

their structure-activity relationships will be provided in the glycolipid section of this manuscript.

The binding footprint of iNKT TCR proteins to CD1d was widely believed to resemble the binding footprint of the related pMHC/ TCR footprint where the two proteins would adopt a diagonal orientation relative to each other. The crystallization of lone TCR proteins led two groups to create theoretical models of the CD1d/glycolipid/TCR ternary complex based on this plausible assumption[76, 77]. However, the successful crystallization of the ternary complex hCD1d/α-GalCer/TCR at 3.2 Å further elucidated the differences between CD1d and pMHC proteins that not only do they bind different antigens but also bind TCR proteins in different orientations[78]. In the crystal structure, TCR was found to be positioned over the C′ pocket of the CD1d-glycolipid binding cleft and not sitting diagonally over it like the pMHC/TCR binding footprint (**Figure 8**).

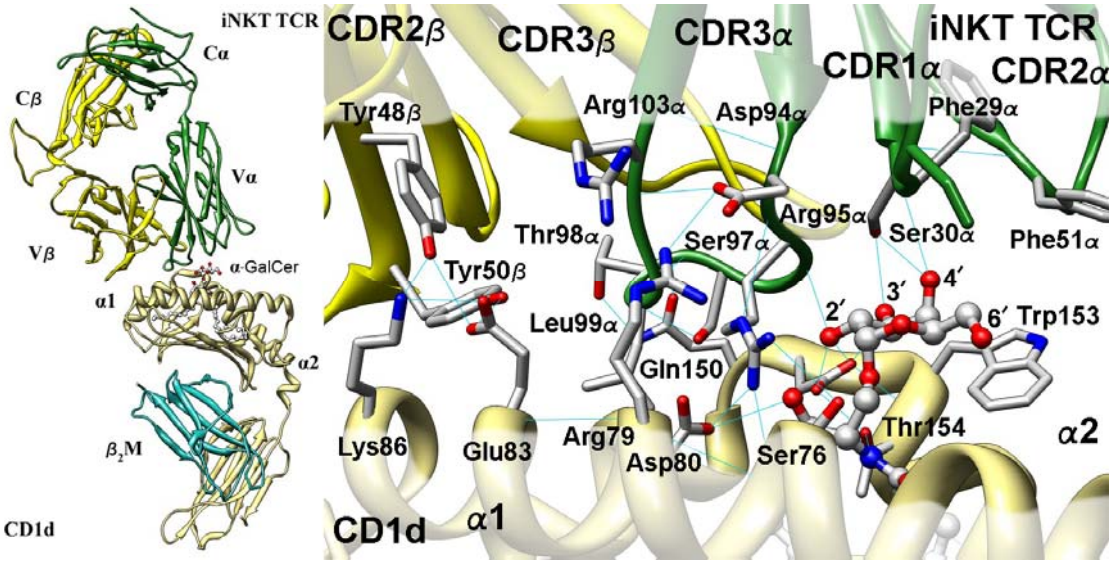


Figure 8. The crystal structure of the CD1d/α-GalCer/TCR complex. The novel binding orientation is shown with the hydrogen bonding network between the proteins and also for those with the glycolipid. *Vα-CDRα loops are green and Vβ-CDRβ are yellow. PDB ID 2PO6.*

The total buried surface area (BSA) between the proteins was roughly 910 $Å^2$ which was 40-75% less than the range of BSA values (1,200-2,400 $Å^2$) for the pMHC/TCR complexes[32]. The Vα chain contributed more to the BSA by 65% as compared to the Vβ chain which conferred only 35%. The Vα chain formed the primary interactions with the galactose sugar of the glycolipid with contributions from both the CDR1α and CDR3α loops. The overall BSA contribution was comprised more from the CDR3α loop with it comprising 52% of the BSA whereas the CDR1α loop only contributed 11%. Of the 35% BSA for the Vβ chain, the CDR2β loop contributed 28% with the non-covalent interactions primarily occurring at the edge of the α1 helix of CD1d. The roles of CDR2α, CDR1β, and CDR3β loops are therefore quite minimal in the binding footprint between CD1d and TCR. This orientation of TCR relative to CD1d along with the BSA percentages have provided strong structural evidence for the immunological and genetic studies performed on the Vα and Vβ chains wherein the invariant Vα chain was essential for iNKT stimulation but the Vβ chain was found to be semi-variable. Furthermore, this data also provided a probable reason for the invariance of the Vα chain throughout evolution since it played the primary role in antigen recognition.

A more in-depth look at the hydrogen bond network of the binding footprint has yielded a variety of interesting results. Most importantly, the overall structures of both TCR and CD1d have deviated almost negligibly if at all from their lone crystal structures implying a lock-and-key mechanism as compared to the plasticity exhibited in the interaction between pMHC and TCR[78]. Furthermore, the orientation of α-GalCer in CD1d did not change upon TCR binding compared to the binary CD1d/glycolipid structure and its hydrogen bonding network was also maintained where: the 2′- and 3′-

OH of the galactose with Asp151; the 3- and 4-OH of the sphingosine chain with Asp80; and both the glycosidic *O* or amide *N* with Thr154. Interestingly, only the galactose ring extended above the surface from CD1d lipid binding groove and formed the only direct interactions with TCR with the lipid not interacting at all with TCR. However, only the serine 30α residue of TCR was identified in the crystal structure as forming a direct hydrogen bond with both the 3′- and 4′-OH groups on the galactose ring. The other hydrogen bonding interactions between the galactose and TCR occurred with the backbone, specifically with that of Gly96α and Phe29α.

The hydrogen bond network existing between the interface of CD1d and TCR leading to the novel binding footprint surprisingly played a small energetic role with only a few of the residues being important for recognition, but it was a larger energetic contribution than the interactions between the glycolipid and TCR[79]. Five residues on the α1 helix and one residue on the α2 helix of CD1d formed hydrogen bonds with TCR (Ser76, Arg79, Asp80, Glu83, Lys86, and Gln150, respectively), and on TCR five residues formed hydrogen bonds with CD1d (Arg95α and Asp94α on the CDR3α loop, Tyr48β and Tyr50β on the CDR2β loop, and Thr98α also on the CDR3α loop). (*As a point of clarity, any residue with a greek symbol is located on the TCR with α placing it on the Vα chain and β on the Vβ chain, and a residue without a symbol is located on the CD1d protein*). The Glu83 residue on CD1d appears to be the focal point of part of the hydrogen bond network with it forming contacts with both tyrosines on the CDR2β loop along with the Lys86 residue. Another focal point would be the Arg95α on CDR3α loop where it made contacts with Asp80, Ser76, and the sphingosine chain hydroxyls of the glycolipid. The only residue on the α2 helix that formed a hydrogen bond to TCR was Gln150 to Thr98α,

but Ser97α also was reported to make a hydrogen bond contact to the backbone of the α2 helix.

The release of the crystal structure was soon followed by an extensive mutation study wherein the residues of CD1d and TCR that were found to form any non-covalent interaction between the proteins were mutated to alanine[79]. It was found that either the Glu83Ala or Arg95αAla mutation had a massive impact on the binding affinity as could be expected since they both formed multiple contacts within the binding footprint. The Arg79Ala mutation had no adverse effect on binding, but the Asp94αAla mutation to which it is hydrogen bonded did drastically affect the binding affinity of TCR to CD1d. It would appear that the role of Asp94α lied in its abilities of maintaining the shape of the CDR3α loop by hydrogen bonding to its backbone and thereby probably influencing the orientation of Arg95α.

The Tyr48βAla or the Tyr50βAla mutation were both equally critical to the binding affinity, which was unusual since it would be expected that one tyrosine should compensate for the loss of the other's hydrogen bond. The importance of having both tyrosine residues was further elucidated by the fact that the Vβ7 chain which was found to confer less avidity to the CD1d/glycolipid complex contained a lysine residue in place of the Tyr48β residue, so it was still capable maintaining the hydrogen bond to Glu83 but probably not as efficiently due to it being shorter than a tyrosine along with possessing different overall electrostatic properties.

The importance of CDR3α loop was further elucidated by the mutations of Ser97α and Leu99α both of which caused a decrease in TCR binding affinity. According to the visual inspection, it appeared that Ser97α formed a hydrogen bond within the loop's backbone

specifically at the Leu99α position, therefore, it may have a similar effect as Asp94α by playing a role in maintaining the correct orientation of the loop. Upon the Leu99αAla mutation, the ability for the CDR3α loop to fill the hydrophobic binding groove completely was removed. This may be due to the creation of unfavorable entropic contributions such as a water remaining in the gap as a singular example. Interestingly, the mutations of either residue Gln150 or Thr98α had only minimal effect on binding, which was unusual since they both participate in the only distinct hydrogen bond between the α2 helix and TCR.

GLYCOLIPIDS PRESENTED BY CD1d TO THE iNKT TCR

Over the last 15 years, the glycolipid antigen used as a standard for all immunological studies and synthetic manipulations concerning CD1d and iNKT cells has been the α-galactosylceramide (α-GalCer) molecule. This glycosphingolipid was originally discovered by the Pharmaceutical Division of the Kirin Brewery Company during a screen of agelasphins derived from the marine sponge *Agelas mauritianus* and was shown to prevent tumor metastases in mice[62, 63]. At the time of the discovery of α-GalCer, surface monoglycosylated galactosylceramides found in almost all cells were shown to be essential components of the neural receptor for the type 1 human immunodeficiency virus (HIV) surface glycoprotein gp120 making the finding even more significant[80]. During the synthetic structure activity relationship (SAR) studies on a variety of glycosphingolipids, an optimized ligand, KRN7000, was produced and shown to be a nonspecific immuno-stimulating agent. KRN7000 (from now on referred to as α-GalCer as it has become accepted in this area of immunology) was found to be optimal

for CD1d binding having an 18-carbon sphingosine chain and a 26-carbon acyl chain (**Figure 9**)[81-83]. The versatility of α-GalCer was that it could be readily loaded onto both mouse and human CD1d, and could be recognized by the TCR of all types of iNKT cells[84-89]. The CD1d/α-GalCer/TCR association is capable of triggering a rapid, transient, and massive response of iNKT cells that lead to a release of Th1 and Th2 cytokines and chemokines.

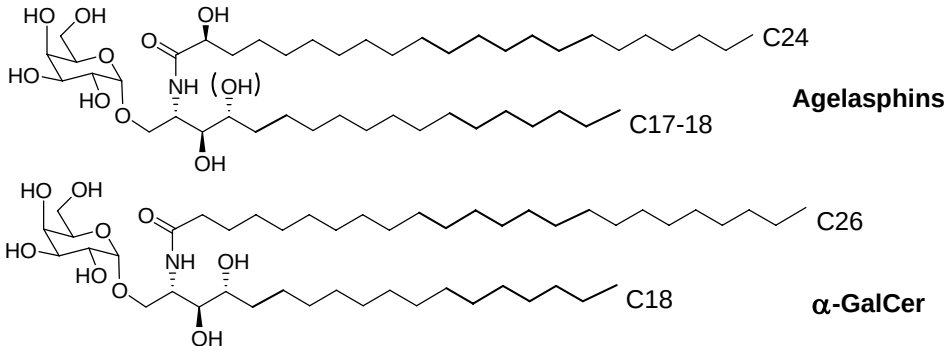


Figure 9. Structures of Agelasphins and α-GalCer (KRN7000). In regards to the agelasphins, the hydroxyl in parentheses exists only in some and a few even have branches in the alkyl chain.

Many biological and synthetic studies have been undertaken by a variety of research groups aimed at understanding of mechanism of α-GalCer recognition in regards to both CD1d and TCR proteins with the hope of finding novel analogs with improved biological activities[85, 86]. From a medicinal and pharmacological perspective, α-GalCer has severe limitations in that its immune response event is very broad and yields opposing results concerning the release of Th1 (IFN-γ) and Th2 (IL-4) cytokines. The difference between Th1 cells and Th2 cells is that Th1 cells participate in cell-mediated immunity and control against intracellular pathogens, whereas Th2 cells participate in antibody-

mediated immunity concerning extracellular pathogens[90]. Furthermore, the up-regulation of either pathway causes the down-regulation of the other making it a more difficult process to control[91]. Nevertheless, the ability of α-GalCer to influence both innate and adaptive immunities through iNKT cell stimulation makes it a very important target in immunology.

It is more than unusual that an α-glycolipid derived from a marine sponge can yield such a massive immune response in humans, since it is doubtful that humans have evolved with a defense mechanism against a possible invasion of marine sponges. Therefore, a search was undertaken to find both the self (endogenous) and non-self (foreign/exogenous) ligands that play a role in the CD1d-mediated iNKT cell activated immune response. As mentioned earlier, it is assumed that upon CD1d biosynthesis a self lipid is presented to it within the ER more than likely to ensure CD1d binding groove stability throughout its intracellular and extracellular trafficking cycle. It has also been shown that this natural self-lipid of CD1d is capable of eliciting a very weak response from iNKT cells even in the absence of any foreign lipid[92, 93]. In mammalian cells, there have been more than 60 different sphingoid lipid bases and more than 300 oligosaccharide chains identified making a very large glycosphingolipid library[94]. Typically, mammalian cells contain a β- linked ceramide instead of the α- linkage present in α-Galcer[55, 95]. So far no β- linked glycosylceramide or glycosphingolipid has been shown to elicit any type of response from iNKT cells thereby discarding the possibility of a natural mammalian glycosphingolipid from being the self-lipid for CD1d[55, 75].

There has been evidence showing that phospho-lipids could be the potential self-lipids for CD1d proteins (**Figure 10**). A glycosylphosphatidylinisotol (GPI) molecule was the first phospho-lipid shown to bind CD1d, and whose discovery was shortly followed by the more specific phosphatidylinositol (PI) being also capable of binding CD1d[72, 96, 97]. Phosphatidylcholine (PC) which is a lipid secreted from *Drosopholia* cells was then able to be crystallized bound to CD1d as was previously shown[68]. The more complex phosphatidylinositol dimannoside (PIM2) isolated from mycobacteria was also crystallized bound to CD1d[69]. Yet, none of these compounds were shown to be able to elicit the appropriate iNKT response expected of self-lipids[98]. Interestingly, the larger PIM4 containing an additional two mannoses was shown to modestly stimulate iNKT cells, but it was later determined that it was not by CD1d presentation but through an adjuvant role elsewhere in the trafficking cycle[99].

To date, the self-ligand that has caused the most controversy has been the isoglobotrihexosylceramide (iGb3) (**Figure 10**). As originally reported, β-hexosaminidase b deficient mice ($Hexb^{-/-}$) exhibited a severe reduction in the number of

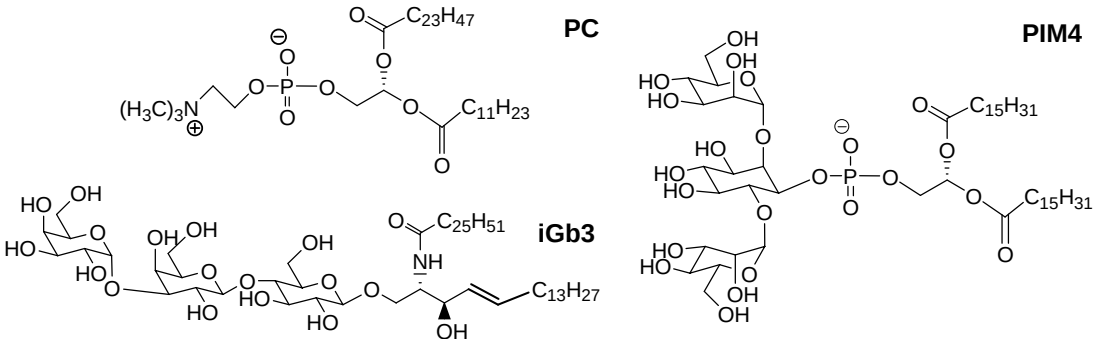


Figure 10. Structures of possible self-lipids for CD1d. Neither phosphatidylcholine (PC) nor phosphatidylinositol dimannoside (PIM2) was shown to elicit the necessary iNKT response expected of self-lipids. Isoglobotrihexosylceramide (iGb3) remains controversial but is no longer regarded as a self-lipid even though it does elicit an markedly diminished iNKT stimulation as compared to α-GalCer.

Vα14 iNKT cells by causing a specific defect in the generation of the lysosomal ligands since CD1d surface expression was unaltered[100]. Therefore, it was hypothesized that one or more of the natural products of Hexb-dependent enzymes must be the self-lipid presented to iNKT cells, and only the iGb3 lipid was found to elicit an iNKT response. The implications of iGb3 being the endogenous self-lipid presented by CD1d are far-reaching considering its total deviation from the structure of α-Galcer, having a β-anomeric linkage between the lipid and oligosaccharide portion and having a trisaccharide instead of a monosaccharide for the sugar head group. The crystal structure of iGb3 bound to CD1d was also solved with the exception of the outermost sugar whose motions were too varied to pinpoint in the x-ray diffraction, and its bound orientation was that expected of a β-linked glycolipid[101]. Since then, biochemical evidence has found no presence of iGb3 in mouse or human thymus or dendritic cells where they would be expected if they did indeed play a role in CD1d-mediated iNKT responses[102]. To further investigate the physiological role of iGb3 in regards to iNKT cell selection, iGb3 synthase enzyme deficient mice (iGb3S$^{-/-}$) developed normally with iNKT cells not diminishing in number as would be the case if the endogenous self-lipid were not present[103]. In the end, although iGb3 is able to elicit an iNKT response even though the structural reason behind remains to be determined, it is no longer considered the self-lipid for CD1d proteins.

The search for the foreign (exogenous) lipid presented by CD1d to iNKT cells however has been more successful. In 2005, nearly 12 years after the discovery of α-GalCer, two groups of collaborating scientists independently reported that glycosphingolipids from the Gram-negative, lipopolysaccharide (LPS) negative alpha-proteobacteria

Sphingomonas were able to activate iNKT cells through CD1d antigen presentation[104, 105]. Gram-negative bacteria derive their name from being unable to retain the crystal violet dye in the Gram staining protocol of bacteria[106]. Unlike Gram-positive bacteria, gram-negative bacteria have a more complex cell wall being composed of both inner and outer membranes with the outer membrane containing a LPS layer which in the case of *Sphingomonas* is comprised of glycosphingolipids instead of large polysaccharide units attached to lipids[107]. The cell wall of *Sphingomonas* have been extensively characterized and have been shown to contain a variety of glycosphingolipids (GSL) (**Figure 11**)[108].

Structurally, GSL-1 is differs from α-GalCer by having a shorter acyl chain and an

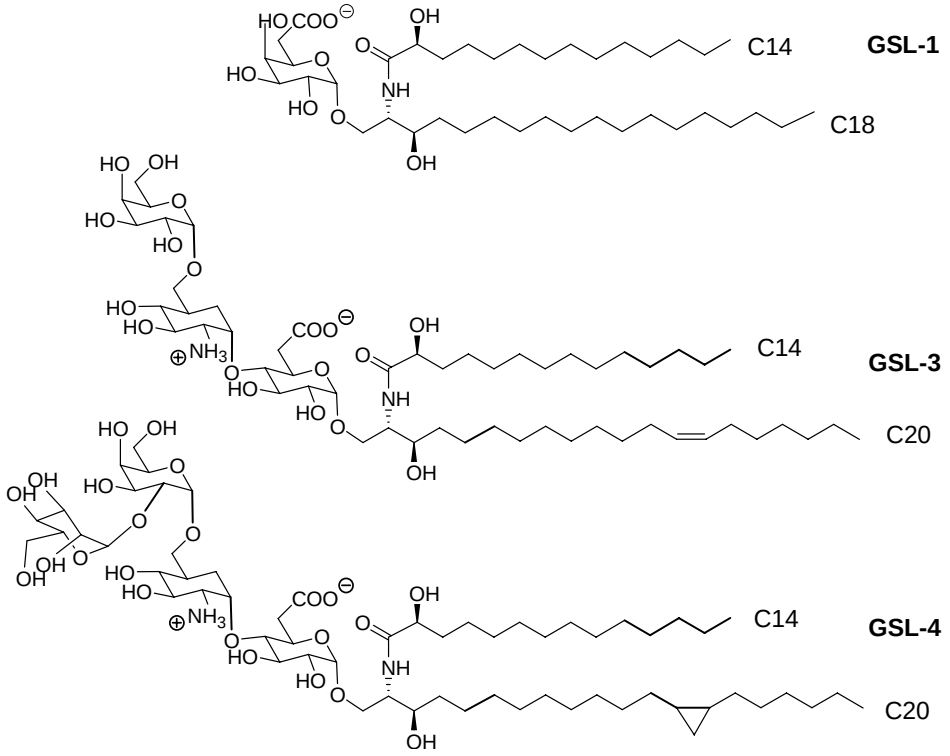


Figure 11. Glycosphingolipids characterized from *Sphingomonas* bacteria cell wall. Any of the lipid components can be paired with any oligosaccharide head group, even though for simplicity reasons they were drawn as such. GSL-1 (as drawn with lipid) was found to stimulate iNKT cells whereas the two polysaccharide lipids were unable.

hydroxyl group along with lacking the 4′-OH on the sphingosphine chain. Synthetic manipulations have yielded GSL-1′ that lacks the hydroxyl on the acyl chain along with GalGSL where the 6′-COO^- is replaced with the -CH_2OH group found on α-GalCer[90, 105]. However, the natural GSL along with the synthetic variants were all inferior to α-GalCer in their iNKT stimulatory capacity where the loss is attributed to the lack of the 4′-OH on the sphingosphine chain causing the glycolipid to sit lower in the binding groove[90]. Considering the size of the *Sphingomonas* genus wherein 20 species have been characterized each exhibiting simple and complex glycosphingolipids, a more extensive synthetic program was undertaken to synthesize a wider library of possible mono-, di-, tri-, and tetraglycosylceramides where in the end only the monosaccharides were found to be potent stimulators of iNKT cells[99, 109].

The physiological threat of *Sphingomonas* remains to be determined since most exist widely in the natural environment such as in water and soil and therefore appear not to threaten humans. Recently, the strain of *Spingomonas mucosissima* has been associated with a patient having sickle cell disease wherein the bacteria through residing in hospital equipment entered the patient causing her health to severely worsen[110]. From an evolutionary standpoint though, *Sphingomonas* are interesting considering that the majority of pathogenic Gram-negative bacteria have their outer membrane covered in LPS which is easily recognized and immediately destroyed by our innate immune system, so it is quite possible that *Sphingomonas* have evolved to evade our natural immune responses by removing their LPS while simultaneously our immune system co-evolved with the bacteria by designing a newer adaptive immune response that recognizes the newer membrane composed of glycosphingolipids[111].

The lucrative nature of α-GalCer, though, is in its ability to yield a wide cascade of immune responses. Therefore, much research has been focused on being able to control the cascade by attempting to bias the cytokine release one way or another through structural manipulation of both the galactose and the ceramide portions of α-GalCer. Results from modifications to the sphingosine or acyl chain have shown to affect not only the magnitude of the iNKT cell stimulation, but also the profile of the stimulation[112].

On the phytosphingosine chain, the 3-OH was found to be crucial for activity wherein the 3-deoxy analog with a 4-OH along with the 3,4-deoxy analog were shown not to be able to elicit an iNKT response[63]. Furthermore, the 3-OH containing glycosphingolipids and even iGb3 were able stimulate iNKT cells, albeit slightly diminished compared to α-GalCer since the lack of the 4-OH caused the glycolipid to sit deeper in the binding groove[90]. Molecular dynamics simulations carried out on the 3,4-deoxy analog bound to CD1d, however, showed that the sugar orientation was quite similar to that of α-GalCer implying that maybe their removal influenced some other part of the system[70]. The length of the sphingosine chain also played a role in biasing the release profile where the analog OCH, a shortened fully saturated lipid (9 carbon length), was found to induce a Th2-biased cytokine release that lowered the IFN-γ production by the iNKT cells but maintained the production of IL-4 cytokines (**Figure 12**)[113]. The simulation of OCH showed similar results until the 8 ns second mark in the trajectory where the galactose shifted forward causing the sphingosine chain to shift down projecting the sugar more into the solvent much like the orientation seen in the β- linked glycosylceramide crystal structures[70]. Since the truncation of the sphingosine chain was shown to lead to a less stable complex with CD1d in the simulation, it is possible that this then leads to a less

stable tertiary complex with TCR thereby lowering the half-life of interaction necessary for it to cause a full cascade of immune response[112, 114].

On the other half of the ceramide, the acyl chain length has also been shown to be able to bias the cytokine profile of iNKT cells in a similar way as the sphingosine chain analogs were found to do[112]. The C20:2 analog that contains a diunsaturated C20 fatty acid instead of the saturated acyl chain induced a Th2-biased cytokine profile with diminished IFN-γ profile along with increased IL-4 production (**Figure 12**)[86]. PBS-25, where the PBS stands for the investigator whose lab the compound was synthesized in (Paul B. Savage), also showed a similar cytokine profile as the C20:2 analog[115]. The complete shortening of the acyl chain down to 2 carbons, however, completely diminished activity much like the complete shortening of the sphingosine chain down to 9 carbons did[112, 115]. Furthermore, just as for the sphingosine chain where there appeared to be an optimal length so did it appear for the acyl chain length wherein chain lengths of

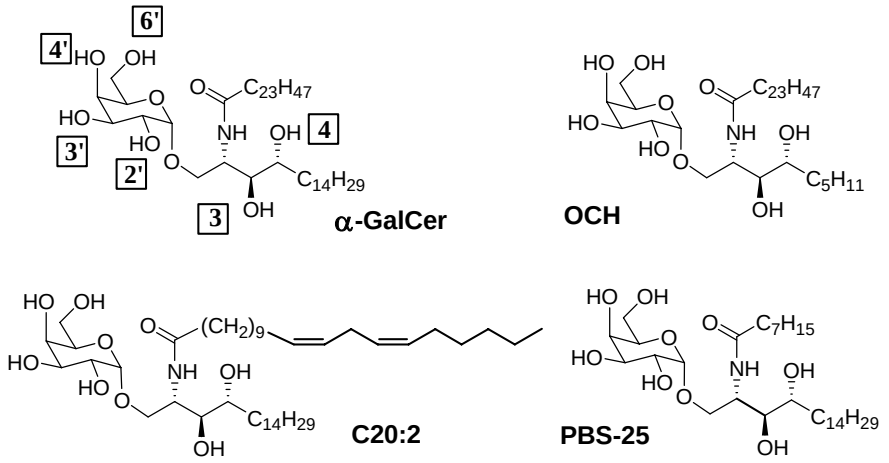


Figure 12. Synthetic glycolipids modified on the ceramide portion. Lipid modifications all caused a TH2-biased cytokine profile upon iNKT stimulation. OCH has a truncated sphingosine chain, whereas, PBS-25 has a truncated acyl chain. The addition of unsaturation into the acyl chain also caused a biased cytokine profile.

11 carbons such as in the C11:1 analog or 8 carbons in PBS-25 were both able to bind to CD1d and stimulate iNKT cells. Interestingly, upon the crystallization of PBS-25, the electron density showed a spacer lipid to exist in the binding groove of the A´ pocket directly below the acyl chain much like as was found in similar crystal structures of CD1b and pMHC compounds without peptides bound[61, 116]. The existence of the spacer lipid implicated the necessity of a self-lipid to exist within CD1d as a stabilizer for the hydrophobic binding pocket until a suitable foreign antigen is loaded[66]. Nevertheless, it can be concluded upon observing that the kinetic data comparing C20:2 to OCH where both possessed similar dissociation constants (K_d) when measured by surface plasma resonance experiments (SPR), that through structural perturbation of the ceramide lipid portion the kinetics of association are affected and the diminishment in affinity between CD1d and TCR caused the change in the stimulatory profile of iNKT cells[112]. This theory was further propositioned by other studies when a terminal aromatic group was placed on a shortened acyl chain showing decreased IFN-γ levels but unchanged IL-4 levels[84].

Much less investigation has been done on how the sugar moiety of the glycolipid participates in the stimulation of iNKT cells, and whatever has been studied has shown that seemingly any modification no matter how minute to the galactose was able to demolish the iNKT response. To begin with, the anomeric linkage as has been discussed has been found to be vital towards TCR recognition with β-GalCer being unable to stimulate iNKT cells (**Figure 13**)[55, 117]. The replacement of the glycosidic oxygen with a carbon yielding a C-glycoside analog was able to stimulate iNKT cells, and even more interestingly a Th1-bias was observed where IFN-γ and IL-12 cytokines were increased with a concurrent decrease in Th2 cytokine IL-2 production[118]. On the other hand, a

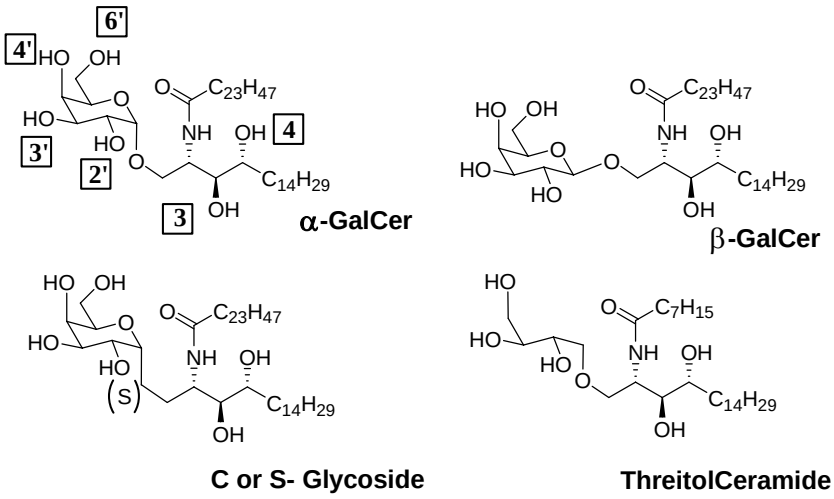


Figure 13. Glycolipid analogs with modifications to the sugar. The 2′- and 3′- OH have been shown to be critical for activity whereas the 4′- and 6′-OH of α-Galcer are more open to modification. Neither β-GalCer nor thioglycoside analog possess iNKT stimulatory abilities, whereas the C-glycoside and threitolceramide analogs are capable of iNKT activation.

sulfur glycosidic linkage yielding a thio-glycosylceramide was incapable of generating an iNKT response *in vivo* or *in vitro*[119].

The replacement of the full galactose sugar with glucose, where the 4′-OH was equatorial rather than axial, resulted in a weaker agonist than α-GalCer[55, 120]. Studies were also performed where certain positions of the galactose were substituted with a sugar and only on truncation by a glycosidase were they able to stimulate iNKT cells with the exception of a sugar on the 6′- position[55, 121]. The freedom at the 6′- position of the galactose was also shown when a fluorophore and a biotin group were introduced at it without causing a diminished iNKT response[122].

Unlike the 6′- position, the 2′-OH of the sugar was found to be very critical for CD1d binding[123]. When it was substituted by hydrogen, fluoro, acetylamide or modified to an axial orientation (mannose), the biological activities of the resulting analogs were

dramatically decreased[124]. The 3′-OH has shown similar reluctance to modifications with the exception of a 3′-sulfo-α-galactose analog[125]. Recently, our laboratory investigated the restrictions on the 4′-OH by synthetic modifications to the galactose sugar finding that unlike the 2′- and 3′-OH it is more open to a variety of modifications but to not as large of a degree as the 6′-OH position (*unpublished results*).

Even though glycosylceramide analogs containing modifications to the sugar appear to be not very well tolerated by the system, it appears that the sugar is not necessary for stimulation. The analog threitolceramide where the C5 and C6 carbons and associated hydroxyl were clipped from the galactose sugar was capable of stimulating iNKT cells along with possessing comparable affinity as α-Galcer to TCR[126]. Two other non-glycosidic analogs one where also the C4 position was removed in addition to the C5 and C6 positions and one where the C5 and C6 were retained but the ether linkage was removed showed much weaker binding affinities and therefore much weaker immune responses were generated. This confounding result has made understanding the structure-activity relationship of the glycolipid so much more difficult. Nevertheless, it can be generally concluded that modifications to the ceramide portion of α-Galcer are more tolerated and have the propensity to bias the cytokine profile whereas modifications to the sugar basically eradicate any iNKT stimulatory capability of the glycolipid.

CHAPTER 2:

MOLECULAR DYNAMICS SIMULATION – AMBER

MOLECULAR DYNAMICS SOLVATION ENVIRONMENTS

The solution of the crystal structure of the CD1d/α-GalCer/TCR tertiary complex created for us the possibility to begin analyzing what interactions play a role between the sugar of any glycolipid and TCR. Considering that the binding footprint of TCR on CD1d had such a minimal surface area with few non-covalent contacts as compared to the pMHC/TCR structures along with the crystal structure being solved to 3.2 Å, molecular dynamics (MD) simulations were chosen as the starting point in our computational investigation to assess the stability of complex. Furthermore, we were interested to determine whether or not the TCR Cα or Cβ and the CD1d β_2M and β sheet regions were important for analyzing the binding footprint and its interactions with the glycolipid. The CD1d protein without a bound glycolipid was also analyzed to determine the stability of the hydrophobic binding groove.

The starting coordinates for all the stimulations were taken from the crystal structure corresponding to the Protein Data Bank[127] entry 2PO6[78] which contained the human iNKT TCR protein that was also expressed and crystallized a few years earlier[77] along with the hCD1d protein. All of the CD1d/glycolipid crystal structures solved to date have

been the mCD1d structures with the exception of original the hCD1d/α-GalCer structure, but due to cross-species reactivity of these proteins, it is possible to extract structure-activity relationship results from either human or mouse CD1d which should then hold true for the system in general.

The AMBER[128] molecular dynamics simulation suite of programs allowed us to explore 4 possible solvation states for the full and truncated complexes: 1) a fully, explicitly solvated box using TIP3P[129] waters under periodic boundary conditions (PBC) in the NVE ensemble where moles (N), volume (V), and energy (E) are conserved; 2) an explicit shell solvation where a thin layer of water was placed around the entire system but the PBC implementation is lost; 3) a spherical cap of explicit waters was used in conjunction with a "pairwise" general Born implicit solvation[130]; and 4) only a "pairwise" generalized Born implicit solvation (**Figure 14**). Over the course of time that this project was ongoing, AMBER was upgraded from version 8 to 9 and is currently at version 10. In version 10 of AMBER, the Particle Mesh Ewald Molecular Dynamics (PMEMD) module was implemented in order to improve upon the former Simulated Annealing with NMR-Derived Energy Restraints (SANDER) module in relation to large solvated systems being simulated for an extended period of time[131]. When only a good average over the entire production was necessary and sampling energies every step was found to be very costly, the PMEMD module could be further improved by the use of the NVE ensemble over the NVT (moles, volume, temperature) ensemble or the NPT (moles, pressure, temperature) ensemble which was the least efficient.

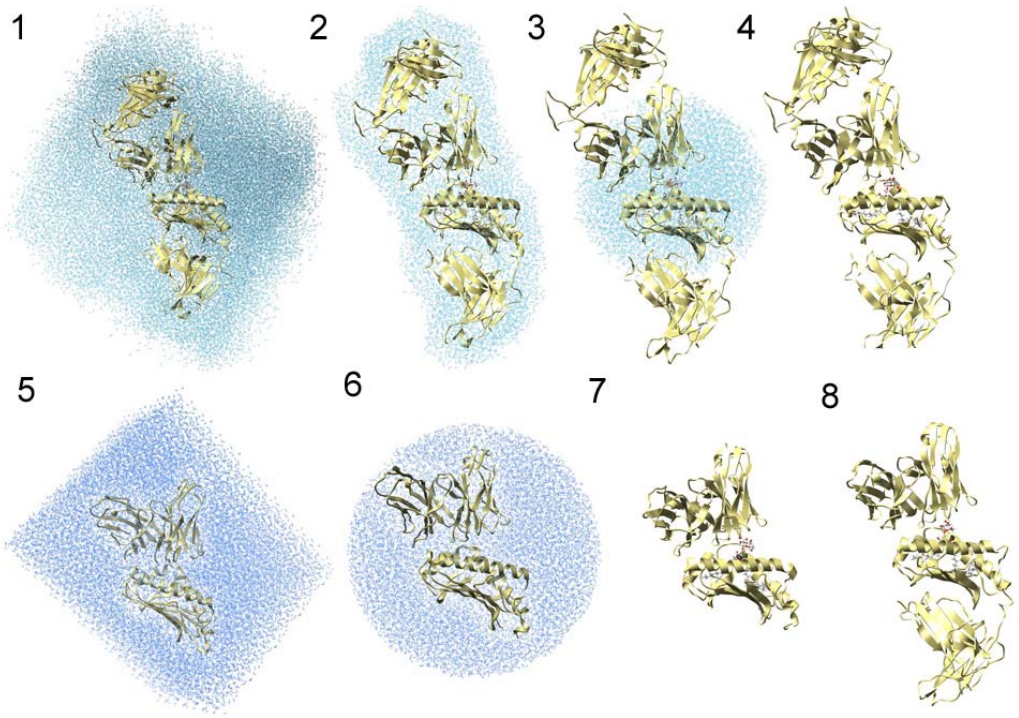


Figure 14. Solvation environments in AMBER. Different environments tested for the full and truncated complex: explicit water box solvation 1 and 5; shell and cap explicit solvation 2, 3, and 6; and implicit solvation 4, 7, and 8.

The full protein complex is roughly twice the size of the truncated protein complex where the TCR Cα & Cβ and the CD1d β_2M and transmembrane β sheet regions were removed. The addition of hydrogens to the full protein complex yielded a system comprised of 13,063 atoms whereas the truncated protein complex consisted of only 6,497 atoms. The full complex was solvated by a 118.22 x 57.71 x 117.44 $Å^3$ box totaling 40,846 TIP3P water molecules and the truncated complex was solvated by a 89.53 x 49.41 x 67.45 $Å^3$ box totaling 17,886 TIP3P water molecules. For the MD simulations, the proteins were defined by the ff03 force field[132] whereas the α-GalCer ligand was originally split into sugar and lipid portions where the sugar was defined as the 1LA residue available from the Glycam parameters[133] and the lipid portion was

defined using the Antechamber program[134] with the general amber force field (GAFF)[135]. The problem with this methodology was that the anomeric linkage was not defined upon splitting the parameters between sugar and lipid so problems upon simulation were observed. Therefore, the sugar parameters were then ignored and the entire glycolipid was defined by the GAFF force field which was found to be capable simulating sugars.

Following conventional molecular dynamics simulation protocols involving explicit solvation of large proteins, waters were minimized while the complex was held frozen using an initial 500 steps of the steepest descent (SD) algorithm with an additional 500 steps to ensure complete minimization. The entire system was then minimized with an initial 1000 steps of the SD algorithm and an additional 4000 steps of minimization to allow the complex to minimize in concert with the waters. With the complex fixed again, the solvent was then equilibrated to 300 K over the course of 25 ps and then the entire system was allowed to equilibrate at 300 K for an additional 25 ps. This was to avoid the system exploding upon heating the waters and protein simultaneously up to 300 K. A few other additional parameters that were used: time steps of 2 psec were taken, the SHAKE algorithm was applied, the pressure relaxation time (taup) was set to 5.0, Langevin dynamics with a collision frequency (γ) of 1.0 was used, and a nonbonded cutoff of 16.0 Å was employed.

The first half of the simulation (~5.0 ns) was performed using the SANDER module of AMBER v8.0 on 16 Intel Itanium 2 processors with the simulation being finished using the PMEMD module of AMBER v9.0 on 16 new IBM Cluster 1350 processors. Parallel jobs in AMBER can only scale to 16 processors, and it was indeed observed that the PMEMD module was slightly faster, so Table 1 contains the molecular dynamics

processing data based on the simulation times for the PMEMD module (**Table 1**). Simulation end points of 10.0 ns for the full complex (**Figure 14.1**), 16.0 ns for the truncated complex (**Figure 14.5**), and 10.0 ns for the implicitly solvated full complex (**Figure 14.4**) were performed since the 10.0 ns limit is generally accepted as the minimum necessary simulation time to determine system stability.

It was found that the best solvation environment in which to analyze the system was the full explicit water solvation under periodic boundary conditions in an NVE ensemble, which was good since it is the currently accepted method for accurate MD simulations of proteins. All of the simulations were run to the 3 ns mark before undergoing

Table 1. Molecular dynamics simulation averaged processing data. TIME corresponds to the simulation portion with REAL and CPU TIME showing how long it took to run that simulation period of time. TOTAL REAL TIME relates to how many actual days the simulation took with TOTAL CPU TIME showing computer processing hours. The truncated simulation is the least expensive of the three, taking about half the time as for the full complex.

	Full	**Truncated**	**Implicit**
Atoms	13,000	6,500	13,000
Residues	850	400	850
Waters	41,000	18,000	0
Total Atoms	136,000	60,500	13,000
Processors	16	16	16
Time	0.5 ns	0.5 ns	0.5 ns
Real Time	100 h	45 h	60 h
CPU Time	1,900 h	750 h	900 h
Total Real (10 ns)	83 days	37 days	50 days
Total CPU (10 ns)	32,400 h	15,000 h	18,000 h

visual analysis. The pressure constraints of maintaining the waters within the cap and shell explicit solvation environments forced the complex to not deviate at all from the crystal structure. Since the loops of TCR are assumed to be somewhat flexible, these solvation environments were believed to not provide an accurate picture of the complex's behavior in solution. Furthermore, the shell solvation parameters were more prone to crash with a random water molecule leaving the boundary of simulation. The implicitly solvated truncated complexes were found to be too flexible thereby causing the complete unraveling of the proteins' structures. Therefore, only simulations of the solvation environments 1, 4, and 5 (explicit NVE PBC full complex, explicit NVE PBC truncated complex, implicit full complex) were extended to the full 10.0 ns mark for complete analysis.

The PTRAJ (processing trajectory) module of AMBER was used to calculate the overall backbone (Cα atoms) root mean square deviations (RMSD) for each simulation (**Figure 15**). The truncated complex, which was computationally less expensive, was extended to 16 ns to determine whether or not the simulation stabilized past the 10 ns point where it appeared to still be fluctuating, and was found to have stabilized around the 3.0 Å mark. Both simulation trajectories of the full and truncated complexes showed similar trends where they gradually increased fluctuations starting at ~2.0 Å until about 5 ns at which point the system jumped to fluctuating about ~3.5 Å for the full complex and ~3.0 Å for the truncated. The implicit solvation trajectory did not follow the trend observed for the explicit solvation environments by rising very quickly to a RMSD of 4.0 Å. The fluctuations of the implicit system also spanned a range of ~ 2.0 Å in a higher frequency, whereas, the explicit solvation environments had fluctuations of about ~ 0.5

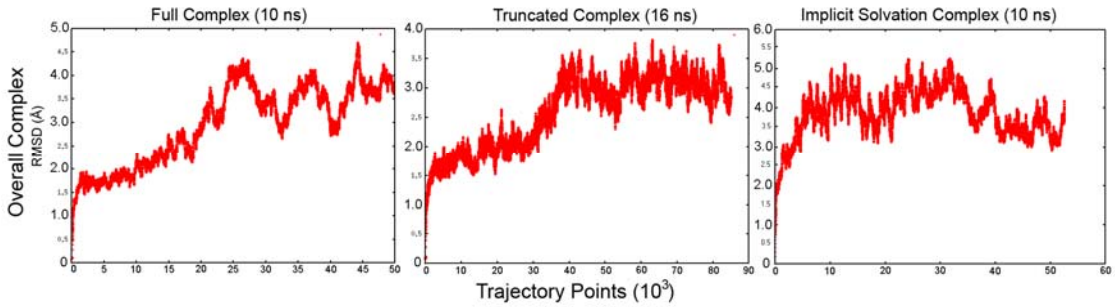


Figure 15. The overall RMSD of the simulation trajectories. The implicit solvation was found to be the least stable as compared to the full and truncated complexes, and that both the full and truncated follow similar RMSD deviation paths. The shift to the 3.0-3.5 Å of both was due to a slight relocation of TCR and CD1d relative to each other.

Å. These fluctuations with the implicit system caused the structural disruption of the binding cavity and the erratic behavior of α-GalCer.

A more extensive protein backbone RMSD analysis was performed on the individual parts comprising both protein where these portions were then found to be very relatively rigid. These sections were the CD1d full protein, binding groove, and β_2M sheet (**Figure 16**), and the TCR full protein, TCR Vα, Vβ, Cα, and Cβ chains (**Figure 17**). The full CD1d protein is very unstable in implicit solvation fluctuating at 4.0 Å with the β_2M sheet fluctuating very freely, whereas, the explicit solvation environments show the full CD1d protein fluctuating at around 2.25, and specifically it was observed in the full complex that both the binding groove and β_2M sheet each fluctuated at around 1.5 Å. The CD1d binding groove fluctuated at 1.75 Å for the explicit solvation systems, but around 3.0 Å for the implicit solvation where the two parallel α helices became destabilized. Unlike CD1d, the full TCR protein was shown to possess comparable stability in both the explicit and implicit solvation environments fluctuating at 2.5 Å for both. The removal of the TCR Cα/Cβ chains to yield the truncated complex caused it to become even more

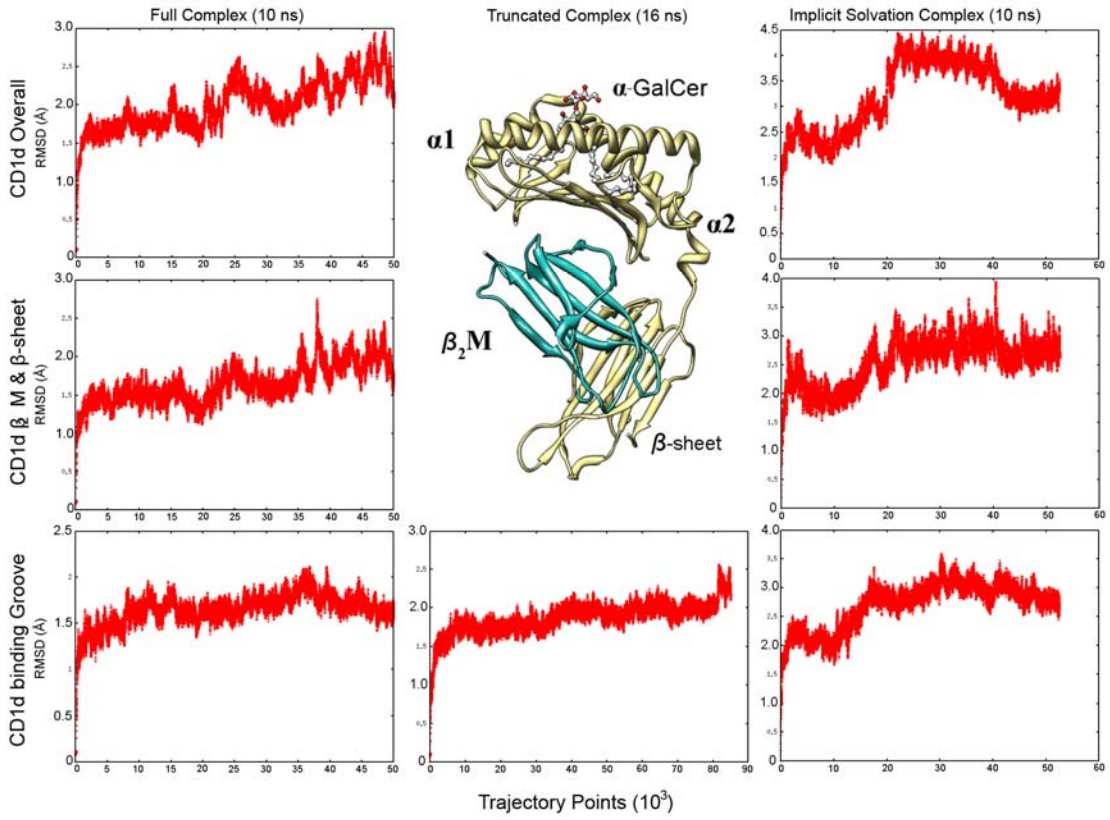


Figure 16. RMSD plots for the CD1d protein. The implicit solvation did not treat the CD1d protein well showing deviations about 1-2 Å higher than found for the explicit solvation systems. Both the individual portions of CD1d fluctuated < 2.0 Å indicative of rigidity implying that the overall RSMD is more in relation to global motions of the protein.

rigid where the remaining TCR Vα/Vβ chains fluctuated around 1.5 Å. Even though TCR is overall a rigid protein as was shown in the crystal structures of different TCR proteins[76, 77], the rigidity is provided specifically by the Vα/Vβ chains and any flexibility arises from the Cα/Cβ chains.

The overall RMSD evaluations for the explicit solvation systems did not show the instability of the TCR and CD1d proteins, but instead represented the act of TCR repositioning itself relative to CD1d as will be seen below in the analysis of the hydrogen bonding network. Furthermore, much of the flexibility in the proteins are located within

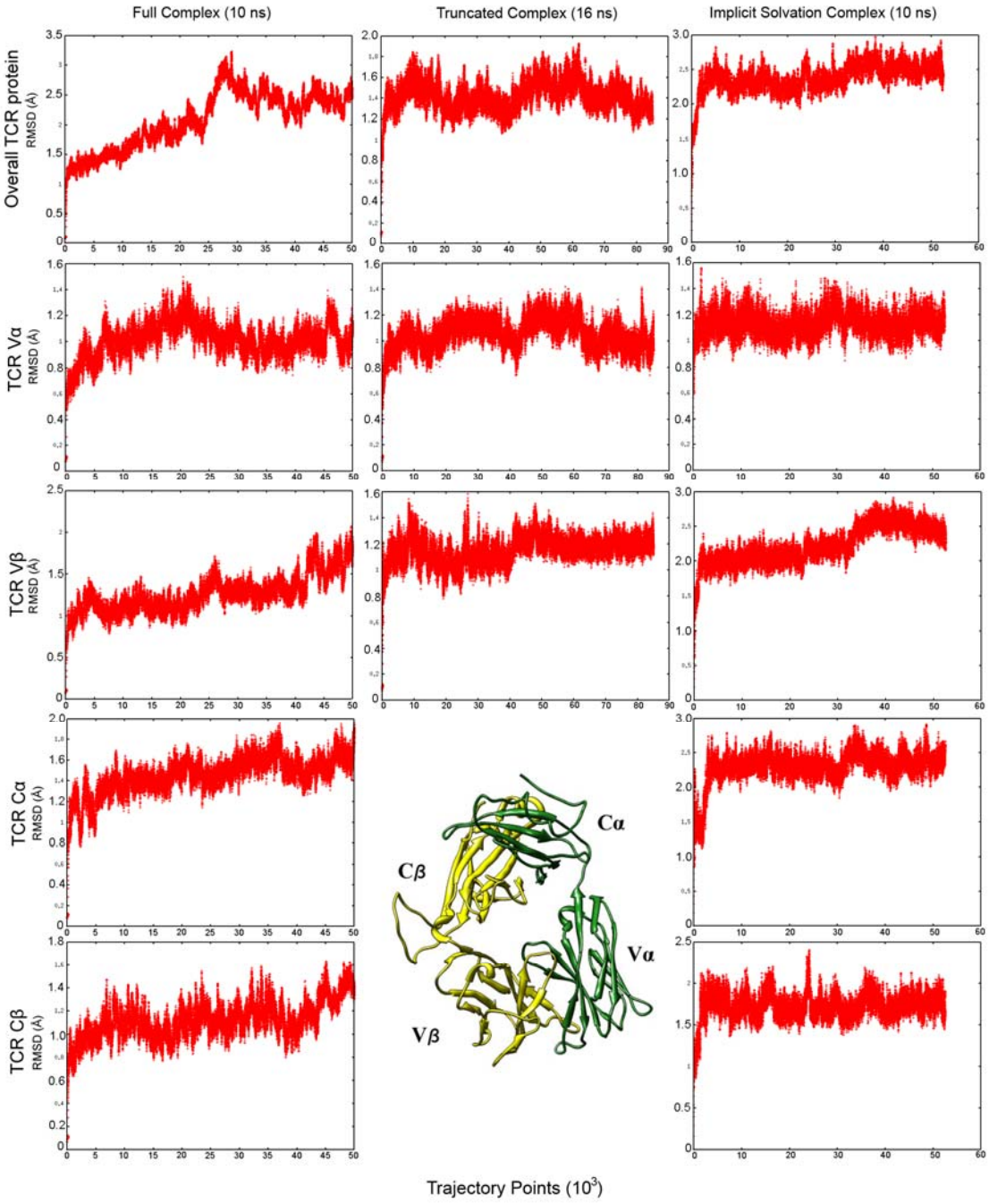


Figure 17. RMSD plots for the TCR protein. In this case, all the solvation environments showed TCR to be a rigid protein showing fluctuations of < 1.5 Å in the explicit environments and under < 2.5 Å for the implicit. The overall RSMD shift seen in the full complex is once again correlated to TCR shifting in its global motion, but more specifically in relation to its Cα an Cβ domains.

the TCR Cα/Cβ chains and the CD1d β_2M sheet as can be seen in the truncated complex having an overall lower RMSD fluctuation compared to the full complex. Considering that the protein sections of TCR and CD1d which make up the binding pocket for the glycolipid are inherently rigid having similar RMSDs in both explicit solvent simulations and the overall trajectories are similar up until 5 ns, it is reasonable to use the smaller and more efficient system to perform future simulations. These results also provide support for the lock-and-key interaction proposed upon the analysis of the binding footprint between the TCR and CD1d wherein the proteins were observed to have deviated minimally from their lone crystal structures upon binding[78].

A visual analysis of the fluctuations was also performed in order to observe the actual motions of the proteins and correlate them to the results of the RMSD calculations. Snapshots of the complexes were taken from every nanosecond up till 10 ns and were aligned to the 50 ps orientation (equilibrated crystal structure) using the Swiss-PDB viewer[136] (**Figure 18**). The explicitly solvated complexes were found to maintain their structural integrity with only a slight shift of TCR over the C′ pocket. On the other hand, the implicit solvation environment allowed too much flexibility in the system as was also shown in the RMSD calculations. The flexibility was localized primarily to CD1d causing the binding cavity to lose structural integrity thereby causing it to eject the glycolipid out of the pocket.

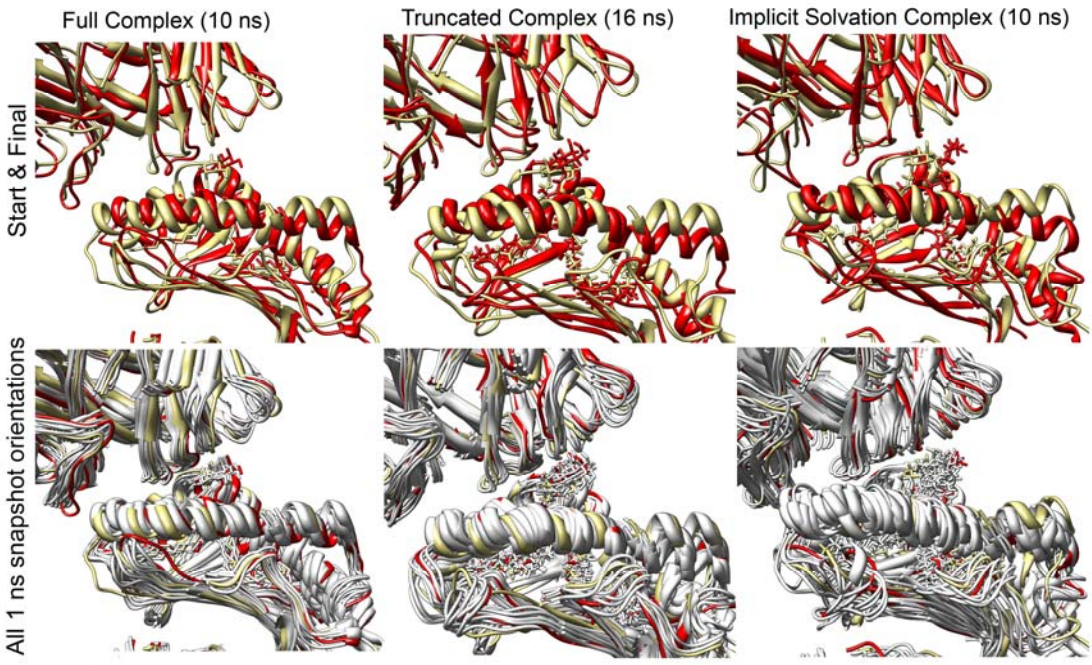


Figure 18. Snapshot alignments from the simulation trajectories. A view of the binding pockets for the stable explicitly solvated full complex, the truncated complex, and the extremely unstable implicitly solvated full complex. In both the explicit solvent environments a slight shift can be seen of TCR relative to CD1d. The ejection of the glycolipid can be seen in the implicit solvation simulation. The khaki color represents the 50 ps starting point and the red color the 10 ns ending point.

Upon observing the relative RMSD for each α-GalCer glycolipid, it can be seen that its overall binding orientation was maintained throughout the simulation with the exception for the implicit solvation system where it experienced a high degree of flexibility (**Figure 19**). The glycolipid at each nanosecond was removed from the complex and manually aligned by visual inspection for each system. It is remarkable how little the lipid portion fluctuated throughout the simulation, showing the extremely compact space the hydrophobic groove formed around it. However, there was some flexibility with the sugar portion, but even this is limited due to the extensive hydrogen bonds holding it in position. The α-GalCer in the implicit solvation was observed to be ejecting itself from

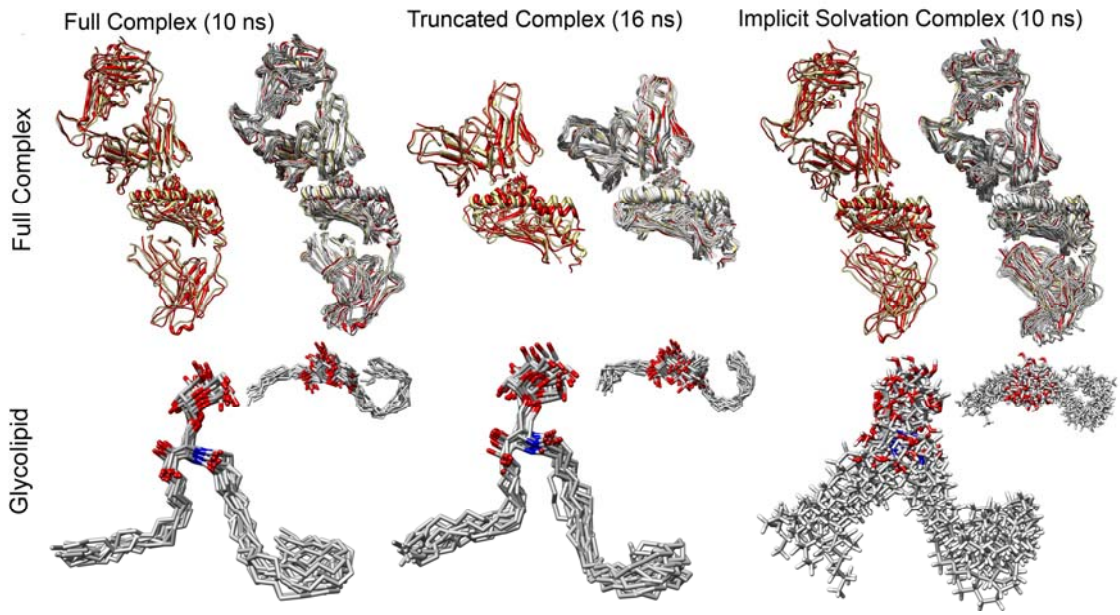


Figure 19. The alignment of the full proteins and the glycolipids. The proteins are shown to maintain their structural integrity and the overall shape of the glycolipid is maintained throughout the simulation with the exception of the implicit solvation where its shape becomes extremely distorted.

the binding pocket by shifting up and out of the pocket through having its lipid portions come together developing a U-shape versus the bound wave-like shape.

A few nanosecond long simulations on the truncated CD1d without α-GalCer and the truncated complex also without a ligand were performed to determine the stability of the binding pocket without a ligand. Interestingly, the simulation after only 2 ns provided a view of what the closed conformation of CD1d binding pocket would look like (**Figure 20**). Unlike in the open conformation crystallized of mCD1d (PDB 1ZT4) which appeared to just be the bound conformation of CD1d without a ligand, the simulation showed that without a ligand the A′ pocket closed up with the C′ remaining open with the hinge point being the Trp153 residue. This flexibility of the A′ pocket appeared to show that the ligand will be loaded into the CD1d binding groove by entering the C′ pocket causing the A′ pocket to slowly open and accommodate the ligand. Furthermore,

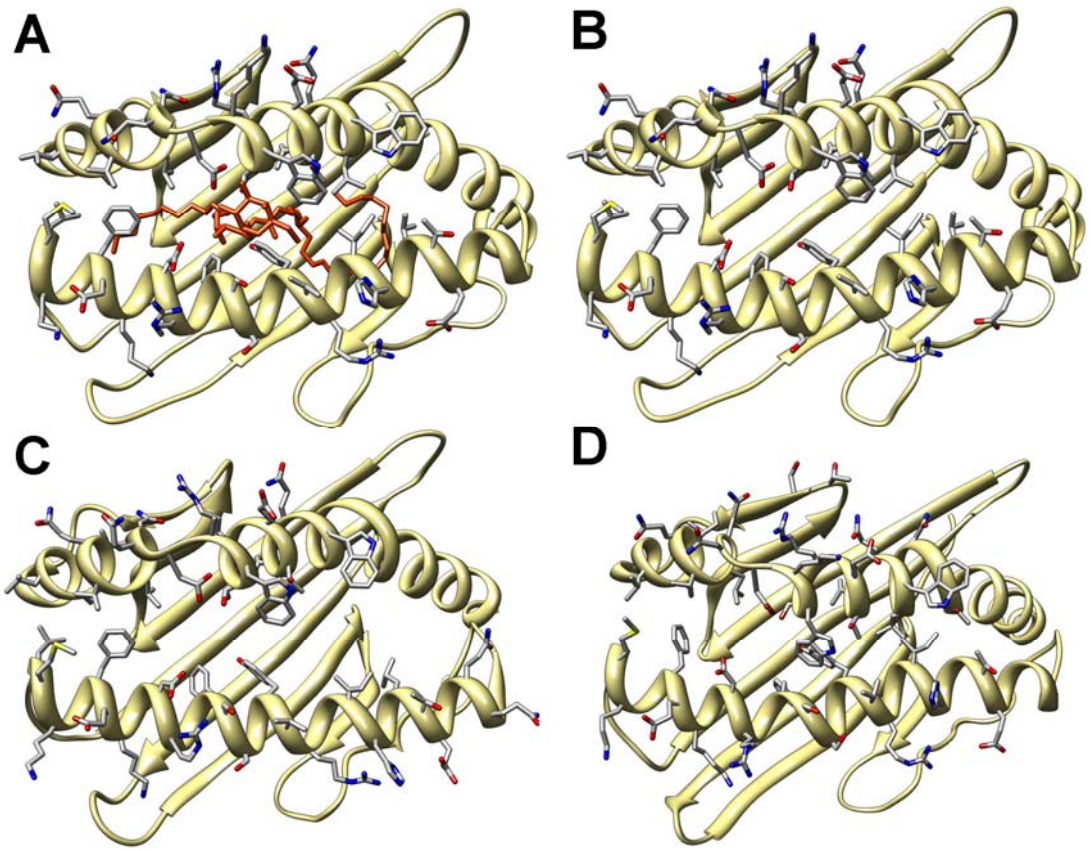


Figure 20. Unbound and bound states of CD1d binding groove. (**A**) Tertiary crystal structure CD1d binding groove with α-GalCer bound; (**B**) Tertiary crystal structure CD1d binding groove with α-GalCer removed; (**C**) Crystal structure of open conformation; and (**D**) closed conformation of CD1d binding groove after 2 ns of simulation.

this also showed the importance of having the appropriate length acyl chain or spacer molecule as was seen with GalA-GSL and PBS-25[65, 66]. And even though the A′ pocket does not form any contacts with TCR, the rotation of the helices inward will cause the C′ pocket residues to be displaced enough to not be in an optimum orientation to form the necessary contacts with TCR.

Upon ascertaining that the global motions of the protein complex along with the glycolipid appear to be stable, a closer look was taken at the hydrogen bond network of the binding footprint. Importantly, the molecular dynamics simulations did not show

TCR reverting to the binding footprint of the pMHC proteins nor did they show it disengaging itself from CD1d. This data provided some additional support that this is indeed the binding footprint between TCR and CD1d. Of course, to truly observe such a dramatic rotation would require much longer simulations, but any instability in the binding would have been observed even minutely after 10 ns of simulation. The lack of dissociation of TCR from CD1d can be primarily attributed to the strong forces of the hydrogen bond network. Considering that through the formation of non-covalent interactions, TCR can scan CD1d proteins presenting glycolipids and thereby recognize foreign from self lipids, the hydrogen bonds were closely investigated. The MD simulation results were compared to the network found in crystal structure along with the results of the mutation studies[79] performed on the key residues.

The PTRAJ module was used once again to analyze the full and truncated complex simulations to determine the permanence of the hydrogen bonds as was a visual inspection performed at varying time intervals (**Table 2**, **Figure 21**). The network between the α1 helix of CD1d and the TCRVα/Vβ chains provides for the unusual binding footprint of this complex. Although the TCR protein sits atop the whole C′ pocket of CD1d, there was only one hydrogen bond between the α2 helix of CD1d and TCR. Based on the crystal structure and as previously described, there are two hydrogen bond networks between the α1 helix of CD1d and TCR. The network existing at the far end of the C′ pocket has at its focus the Glu83 residue to which then Lys86, Tyr48β, and Tyr50β are bonded. The second hydrogen bond network occurs at the other end of the binding footprint with Arg95α on the CDR3β loop being at its focus and having Asp94α, Ser76α, Arg79, Asp80, and the glycolipid hydrogen bonded to it.

Table 2. The permanence of the hydrogen bonds. Key residues' atoms were analyzed throughout the 10 ns simulation where %Oc refers to the period of time of the simulation that hydrogen bond was occupied, i.e. maintained. (*Specific atoms are not shown for the residues below but see Appendix A for these and additional statistics*).

Full Complex			**Truncated Complex**		
Donor	*Acceptor*	*%Oc*	*Donor*	*Acceptor*	*%Oc*
Asp80	AGH_O4	96.79	Asp80	AGH_O4	97.84
Asp80	AGH_O3	92.27	Asp80	AGH_O3	93.63
Asp94α	Arg79	90.70	Asp151	1LA_O3	93.27
Asp151	Thr154	85.46	Asp151	Thr154	83.53
Asp80	Arg95α	84.35	Asp80	Arg95α	49.78
Asp94α	Arg79	83.50	Glu83	Arg103α	37.35
Asp151	1LA_O3	82.78	Asp80	Arg95α	35.43
Glu83	Arg103α	60.02	Asp94α	Arg79	33.27
Glu83	Tyr50β	57.09	Asp94α	Arg79	30.35
Glu83	Tyr48β	54.79	Glu83	Arg103α	20.38
Glu83	Arg103α	30.07	Glu83	Arg103α	19.90
Glu83	Lys86	29.63	AGH_O	1LA_O2	19.51
Asp151	1LA_O2	24.25	Asp94α	Arg79	19.06
Asp94α	Arg79	20.19	AGH_O	Thr154	13.72
Glu83	Lys86	15.98	1LA_O3	Ser30α	11.90
Glu83	Lys86	13.94	Glu83	Tyr48β	10.46
Asp151	1LA_O3	13.53	1LA_O4	Ser30α	10.07
Asp151	Thr154	12.66	Asp151	1LA_O2	10.05
Glu83	Tyr50β	12.02	Glu83	Tyr50β	9.16
Glu83	Arg103	8.39	Asp94α	Arg79	9.13
1LA_O3	Ser30α	6.79	Glu83	Lys86	8.10
AGH_O1	1LA_O2	6.70	Glu83	Arg103α	7.08
1LA_O4	Ser30α	5.87	Glu83	Lys86	5.80

** AGH refers to the lipid portion of α-GalCer and 1LA refers to its galactose headgroup. 1LA_O# refers to hydroxyls on the galactose sugar, AGH_O1 is the glycosidic bond linking the lipid and sugar portions, AGH_O is the amide carbonyl on the acyl chain, AGH_O# are the hydroxyls on the sphingosine chain.

The visual analysis showed that the hydrogen bond network was in flux, however, both the Glu83 and Arg95α residues remained hydrogen-bonded to their constituent residues. Observing the network involving Glu83, it was found to form the most permanent hydrogen bond with Arg103α in both complexes where this residue was not shown to be

hydrogen bonded to Glu83 in the crystal structure. The flexibility of the arginine residue allowed for it to swing over quite early on in the simulation. In the full complex, the hydrogen bond web involving Tyr48β, Tyr50β, and Lys86 for the Glu83 was maintained for ~50% of the simulation, however, this web was less prevalent in the truncated complex where it existed for less of the time. As the simulation progressed, this hydrogen bond web began strong and then faded away, just more quickly for the truncated complex. With the loss of these hydrogen bonds a slight shift of the TCR complex off the C′ pocket was observed. The shift was also likely due to the Arg103α swinging over to form hydrogen bonds to Glu83 thereby bringing the TCR with it.

Correlating these results to the mutation studies carried out, it was found that the Glu83Ala mutation had a massive impact on the binding affinity as could be expected since it is the focal point of one of the hydrogen bond networks and without it all would be lost. Interestingly, the Arg103αAla mutation had no effect on binding affinity implying that its nonexistence can be supplanted by the other residues of the web. The Tyr48βAla and Tyr50βAla mutations were also critical to the binding affinity. This was unusual because it would be expected that one tyrosine could compensate for the loss of the other's hydrogen bond capabilities as could the Arg103α. Considering that in the simulation the hydrogen bonds between the tyrosines and Glu83 were not well maintained, it is quite possible that their importance lies in providing other electrostatic stabilizing interactions with the α1 helix that outweigh the hydrogen bonding interactions.

The other focal point residue, Arg95α, of the hydrogen bond network at the other end of the binding footprint maintained its hydrogen bond to Asp94α for 80% of the full

complex simulation, and about 50% of the time for the truncated complex. This is probably the more important hydrogen bond formation that truly links TCR to CD1d and the glycolipid. Indeed, the Arg95αAla mutation did drastically affect the binding affinity of TCR to CD1d. The Arg79Ala mutation had no adverse effect on binding even though it was shown hydrogen bonding to Asp94α whose mutation did adversely affect binding. Since, Asp94α hydrogen bonds to Arg79, an unimportant residue, it can be assumed that the Asp94αAla mutation causes a displacement of the Arg95α making it unable to form the necessary hydrogen bond to Asp80. The Asp80 residue not only maintained a hydrogen bond with TCR but also exhibited the most permanent hydrogen bonding in both complexes with the hydroxyl groups of the sphingosine chain, existing >90% of the time with both. Therefore, Asp80 located on CD1d is the most important residue in the system by forming strong and lasting hydrogen bonds with both TCR and α-GalCer.

The other hydrogen bond network investigated was that existing between the galactose sugar of the glycolipid and both proteins. During the simulation, the Asp151 which has been thought to be strongly bonded to the 2′-OH of the galactose whose removal, 2′-deoxy, caused the glycolipid to lose all activity preferred instead to hydrogen bond to the 3′-OH by a difference of >50% permanence between the two hydrogen bonds. This preference was furthered by Asp151 hydrogen bonding to the backbone of Thr154 85% of the time, thereby drawing the Asp151 away from the 2′-OH and closer to the 3′-OH. The Ser30α residue appeared to form a hydrogen bond with the galactose sugar of α-GalCer in the crystal structure, however, the Ser30αAla mutation did not affect the binding affinity of TCR to CD1d, and it only remains hydrogen bonded to the 2′- and 3′-OH for <10% of the time. Overall, the galactose sugar remained within its binding pocket

between TCR and CD1d and was able to maintain its overall orientation and hydrogen bonds.

The Swiss-PDB Viewer was used to align the crystal structure, the full complex, and truncated complex at 3 ns, before the RMSD shift to a higher deviation and at the 10 ns mark, in order to provide a visual idea of how CD1d/α-GalCer/TCR interact during the simulation (**Figure 21**). It is important to point out that the RMSDs of the trajectories appear to make the situation more drastic than it truly was when comparing the crystal structure to the 3 and 10 ns final orientations. Interestingly, an arginine stack formed by the 3 ns mark and was maintained more or less throughout the simulation involving the Arg95α, Arg79, and Arg103α residues. The effect if any of this arginine stack remains to be determined, but considering that the Arg79Ala mutation had no effect on binding affinity it may just be a random occurrence of the system.

Three other residues that do not participate in hydrogen bonding but do provide van der Waals interactions with the sugar are also important to discuss. The Phe51α and Trp153 in the crystal structure are separated by 6.7 Å which appeared to be a large enough distance for the Phe51α to adopt an edge to face orientation rather than a offset or stacked orientation relative to the Trp153 creating a barrier behind the 4′-OH position. This closing creates a more compact electrostatic pocket for the galactose and may also hinder modifications to the 4′- position of the galactose that are relatively bulky. The Arg95α also behaves in a similar fashion by angling in more towards the 1′- position thereby further compacting the space within which the sugar sits. This propensity for the TCR cavity surrounding the sugar to compact even more around the sugar supports why modifications to it are not tolerated by the system.

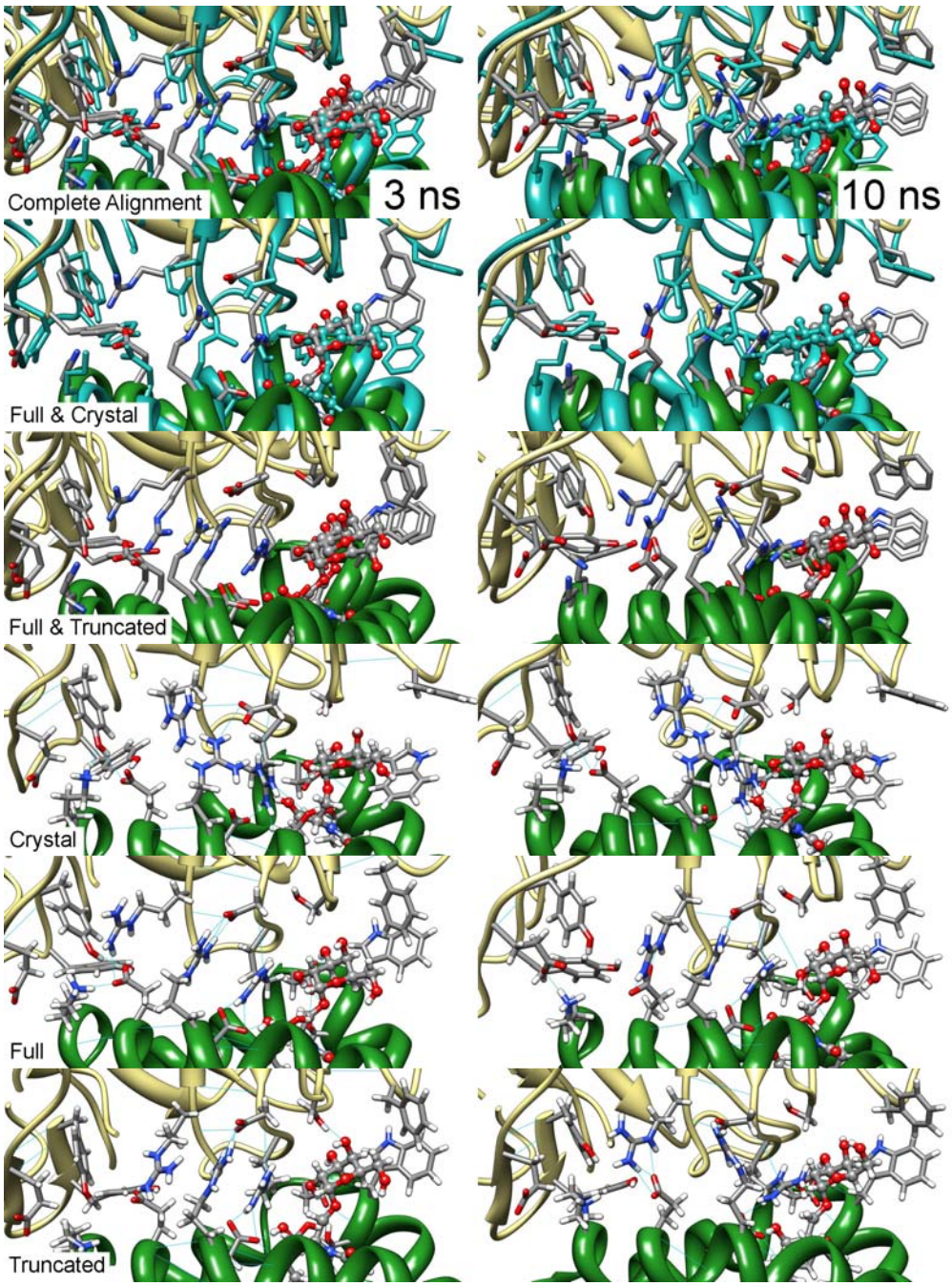


Figure 21. The alignment of the crystal structure, the full complex, and truncated complex. Alignments were done at the 3 (LH) and 10 ns (RH) marks in order to provide a visual idea of how CD1d/α-GalCer/TCR interaction evolves during the simulation. In the complete alignment and full & crystal images the crystal structure is colored cyan, and in each picture TCR is khaki and CD1d is green. The hydrogens were removed from the overlapping structures to provide a clearer image, whereas in the individual proteins' images they were maintained to show the hydrogen bonding network. Residue labels were also omitted for the sake of clarity.

In summary, it appeared that the hydrogen bonding interactions between the CD1d/α-GalCer/TCR evolved as the simulations progress. The loss of hydrogen bonds and the formation of new ones were correlated to the shifting of TCR slightly over the C′ pocket of CD1d. This repositioning cannot be considered an instability and hence unreliability in the simulations, but should be an expected motion for these two proteins. To begin with, the total buried surface area between TCR and CD1d was quite small at ~910Å^2 compared to buried surface area between TCR and MHC proteins, implying that there are less stabilizing interactions between TCR and CD1d and therefore more ability to move relative to each other. Furthermore, TCR and CD1d remained relatively rigid around the binding footprint exhibiting a lock-and-key type interaction where this type of protein binding interaction requires TCR and CD1d to come together in roughly the correct orientation, so a slight shift of TCR relative to CD1d just implies a larger keyhole. All in all, the crystal structure is just one stable, energy-minimum, frozen representation of this system that happened to be crystallizable. The structures resulting from simulation should be thought of as additional representations for how this system might behave in solution.

CHAPTER 3:

DOCKING AND SIMULATION

OF THE 2′-, 3′-, AND 4′-α-GALCER DERIVATIVES

DOCKING PARAMETERS AND RESULTS

In the introduction on the structure-activity relationships (SAR) between various natural and synthetic glycolipids the findings have yielded certain guidelines concerning binding affinity. To begin with, it appears that CD1d is capable of binding a diverse assortment of glycolipids as long as their lipid chains are able to occupy the hydrophobic binding groove along with the sphingosine chain maintaining its ability to hydrogen bond to Asp80 on the α1 helix. When it comes to TCR binding affinity, the picture is more complex with lipid modifications tending to be able to shift the cytokine profile release towards Th2-bias (IL-2). Sugar modifications on the other hand have not yet been fully clarified. The anomeric linkage has been shown to be able to negate binding (thio) or bias cytokine release towards Th1 cytokines (methylene). Furthermore, the 2′- and 3′-OHs seem quite adverse to any modifications even though they appear to only hydrogen bond to Asp151 with the 2′-OH being crucial due to it hydrogen bonding to the backbone of the CDR3β loop at the Gly96α residue. The 6′- position is open to any modification whereas the 4′ position has not really been investigated. Therefore, to begin

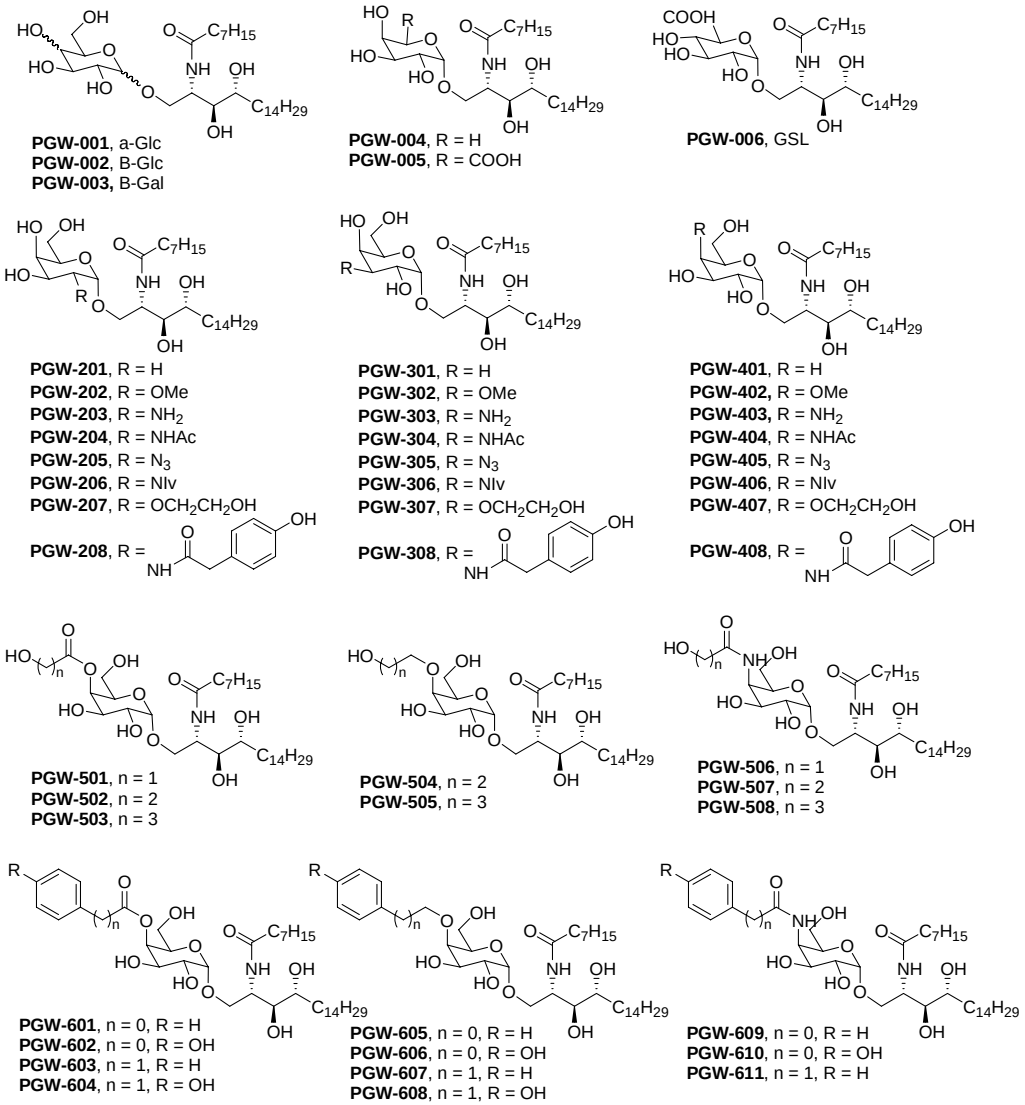


Figure 22. The list of 49 various sugar modified glycolipid analogs. Identical modifications were done on the 2′-, 3′-, and 4′- positions (PGW20#, PGW30#, and PGW40#, respectively). The tolerance of the 4′- position was further investigated by adding a variety of substituents that varied in size and types of aromatic substituents.

understanding the differences between the 2′-, 3′-, and 4′-OH positions a thorough computational investigation was undertaken.

A library of 49 sugar modified glycolipid analogs were built using the MacroModel[137] suite of programs based on α-GalCer in the crystal structure 2PO6 (**Figure 22**).

The first group of analogs (PGW001-006) were built to correlate their binding energies to currently available experimental evidence where it has been shown that CD1d can bind both α and β linked sugars, however, β linked sugars cannot be recognized by TCR, that the 4′-OH can be either equatorial or axial (Gal versus Glc), and that a carboxylic acid group at the 6′- position (GSL) remains capable of being recognized by TCR. The second group of analogs (PGW20X, PGW30X, and PGW40X) were created to analyze whether or not a preference could be found between the 2′-, 3′-, or 4′- positions when analyzing binding energies. The substitutions involved deoxy to eliminate the hydrogen bond capabilities of the position, methoxy to eliminate the hydrogen bond donor capabilities, a group of amine derivatives, and two bulkier substituents. Based on the crystal structure of the tertiary complex[78, 79], a library of 4′- analogs (PGW50X and PGW6XX) were built to analyze how bulky of a substitution would be tolerated at the position, whether the Phe51α and Trp153 could lend themselves to π-π interactions with the substituent, and if the flexibility of the substituent as determined by the sugar linkage (ester, ether, or amide) played a role in binding.

The glycolipid analogs were minimized using the PRCG method[138] with a maximum 500 steps in water defined by the OPLS2005 force field[139, 140] with a dielectric constant of 80.0 while the lipid portion was held frozen. Snapshots taken every 1 ns of both the full and truncated explicit box solvated complex trajectories were used as the receptors to which α-GalCer and the 49 glycolipids were docked. Additionally, TCR was removed to

determine the binding of the glycolipid to just the CD1d protein also at every nanosecond. The glycolipids were submitted to AUTODOCK v.3[141] where the AutoDockTools package was used to generate all of the necessary input files and the docking grids. The Lamarckian flexible ligand genetic algorithm search was employed. The torsions to be varied were kept under a total of ~15. One torsion in both the sphingosine and acyl chains were permitted to provide for some flexibility upon docking for the lipid. The 6′-OH was chosen to not have torsional freedom unless the substitution was on it such as for GSL. The anomeric linkage had 3 degrees of freedom along with the 2′-, 3′-, and 4′- positions being free to rotate. Each population consisted of 250 individuals which underwent a maximum number of 5.0 x 10^6 energy evaluations with a maximum number of 2.7 x 10^4 generations and a mutation rate of 0.02 with a crossover rate of 0.8.

The free energy of binding in solvent is calculated in Autodock by estimating the free energy change associated with the solvation of the protein and the ligand, then subtracting those from the free energy of binding in vacuum calculated by Autodock and from the estimated free energy change for the solvation of the complex:

$$\Delta G_{binding,solution} = \Delta G_{binding,vacuo} + \Delta G_{solvation(EI)} - \Delta G_{solvation(E+1)}$$

Upon analyzing the docking of α-GalCer back into the energy minimized crystal structure corresponding to the 50 ps time mark in the simulation, it appears that docking was capable of placing the glycolipid back into the binding pocket albeit with a slight perturbation (**Figure 23**). In the simulation, it appeared that the galactose sugar would fluctuate between the hydrogen bonding of the 2′- and 3′-OH to Asp151, and at the time of 50 ps, the 3′-OH was preferentially bonded to Asp151 and Ser30α simultaneously with

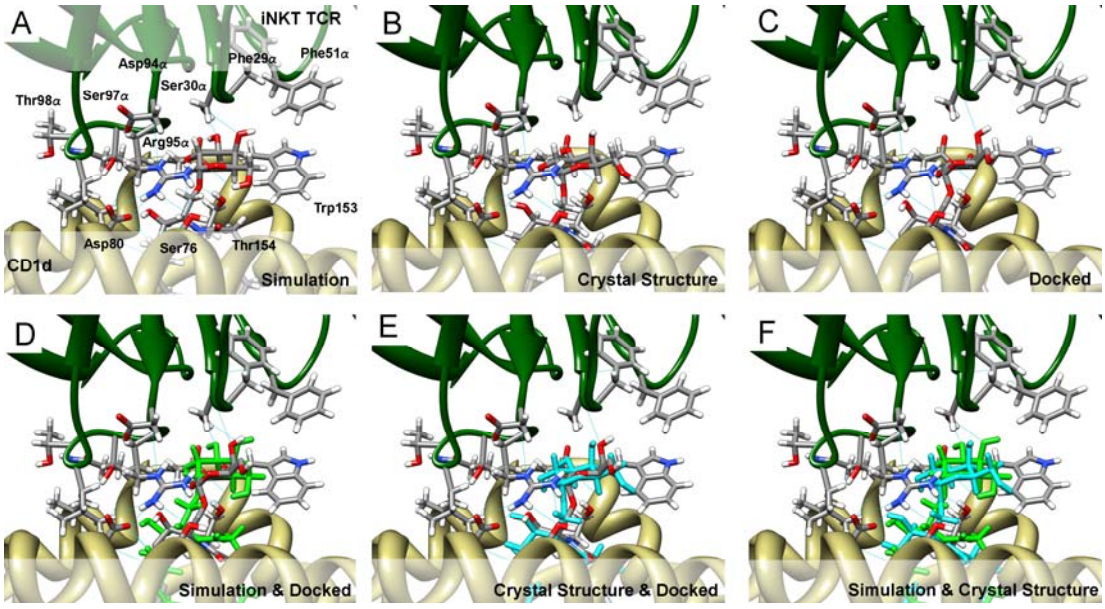
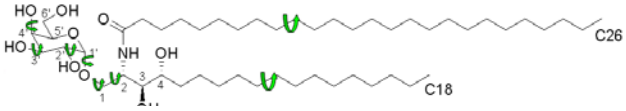


Figure 23. Structural correlations of the α-GalCer docking result. (A) Orientation of α-GalCer in simulation upon energy minimization (50 ps). (B) Orientation of α-GalCer as it existed in the original crystal structure. (C) Docked orientation of α-GalCer into energy minimized complex. (D) Overlay of docked orientation with simulation orientation (green) at time of snapshot. (E) Overlay of docked orientation with crystal structure (cyan) orientation. (F) Overlay of simulation (green) and crystal structure (cyan) orientations. Autodock was found capable of docking α-GalCer back into the binding pocket of the CD1d/TCR complex.

the 2′-OH interacting with Gly96α of the CDR3β loop. The docking result though favored the orientation wherein both 2′ and 3′-OH are hydrogen bonded to Asp151 which is more reminiscent of the original crystal structure orientation. Overall, the docked orientation was similar to both the simulation and crystal structure orientations. Since Autodock handles glycosidic linkages somewhat poorly from an energetic standpoint, this was the best result that could be expected.

The other fact taken into account upon determining Autodock's capabilities as the choice docking program was a question of efficiency. The parameters were chosen to allow for relatively quick docking of 50 glycolipid derivatives into 10 snapshots of the tertiary complex along with 10 snapshots of CD1d (**Table 3**). Additionally, a few

Table 3. Autodock parameters and processing data. Each glycolipid derivative was docked into 10 snapshots of the tertiary complex along with 10 snapshots of just CD1d yielding a total of 1,000 docked results.

	Data	Statistics	
Torsions	8-15	Tertiary Complex	10
Population	250	Binary Complex	10
Energy Evaluations	5.0×10^6	Compounds	50
Generations	2.7×10^4	Total Docked	1,000
Processors	1	Total CPU Time	4,000 hrs
CPU Time	4 hrs		
Torsions			

assumptions had to be made in regards to the docking results: (1) since all the glycolipids were built from the crystal structure orientation of α-GalCer their docked orientations will be slightly off from those found in the simulation of α-GalCer; (2) since the goal was to observe deviations in the sugar orientation due to the varied substitutions: (a) lipid binding should be similar considering that no changes were made to it, and (b) energetic differences should resemble the differences in substitutions; (3) since proteins are in constant flux, in order to simulate the idea of flexible residues, the glycolipid derivatives were docked into a variety of static snapshots from the simulation trajectory; and (4) any inconsistencies in docking arising from poor docked orientations or skewed energetic data should be evened out by averaging the results from all the docking.

As a starting point, all the 50 glycolipid analogs were docked into the 50 ps energy-minimized tertiary structure which is more or less the crystal structure to visualize if docking would yield the answers we were seeking. The experimental-based compounds,

α-GalCer, α-GlcCer (PGW001), 6-deoxy-α-GalCer (PGW004), and GSL-1 (PGW006), were found bind similarly to each other (**Figure 24**). This was expected since all of these compounds have been found to stimulate iNKT cells. As has been previously suggested, the shift of the 4′-OH from the axial to the equatorial position creating α-GlcCer caused it to lose the hydrogen bond to backbone of the CDR1α loop. Both the β-GalCer (PGW002) and β-GlcCer (PGW003) possessed docked orientations reminiscent of those found in the crystal structures of the β-linked glycolipids with both the 2′- and 3′-OH hydrogen bonded to Asp151. The β- anomeric linkage also caused the 4′-OH to be in closer proximity to the CDR1α loop whether or not it was axial or equatorial.

These seemingly acceptable docked orientations though bring to light the problem being investigated, wherein β-linked glycolipids should indeed be capable of being recognized by TCR and therefore should elicit an iNKT response. Structurally specific drugs usually are very susceptible to small changes in chemical structure wherein activity and potency can be quite altered.[142] This logic does appear to hold true for this ligand, for it appears that β-linked glycolipids are able to maintain all the necessary hydrogen bonds with CD1d, form better hydrogen bonds with TCR, and are spatially similar in their sugar head group, and yet, the change from α- to β- causes all activity to be lost. Therefore, it must be concluded that some other event is causing β- linked glycolipids from not being able to elicit an iNKT response.

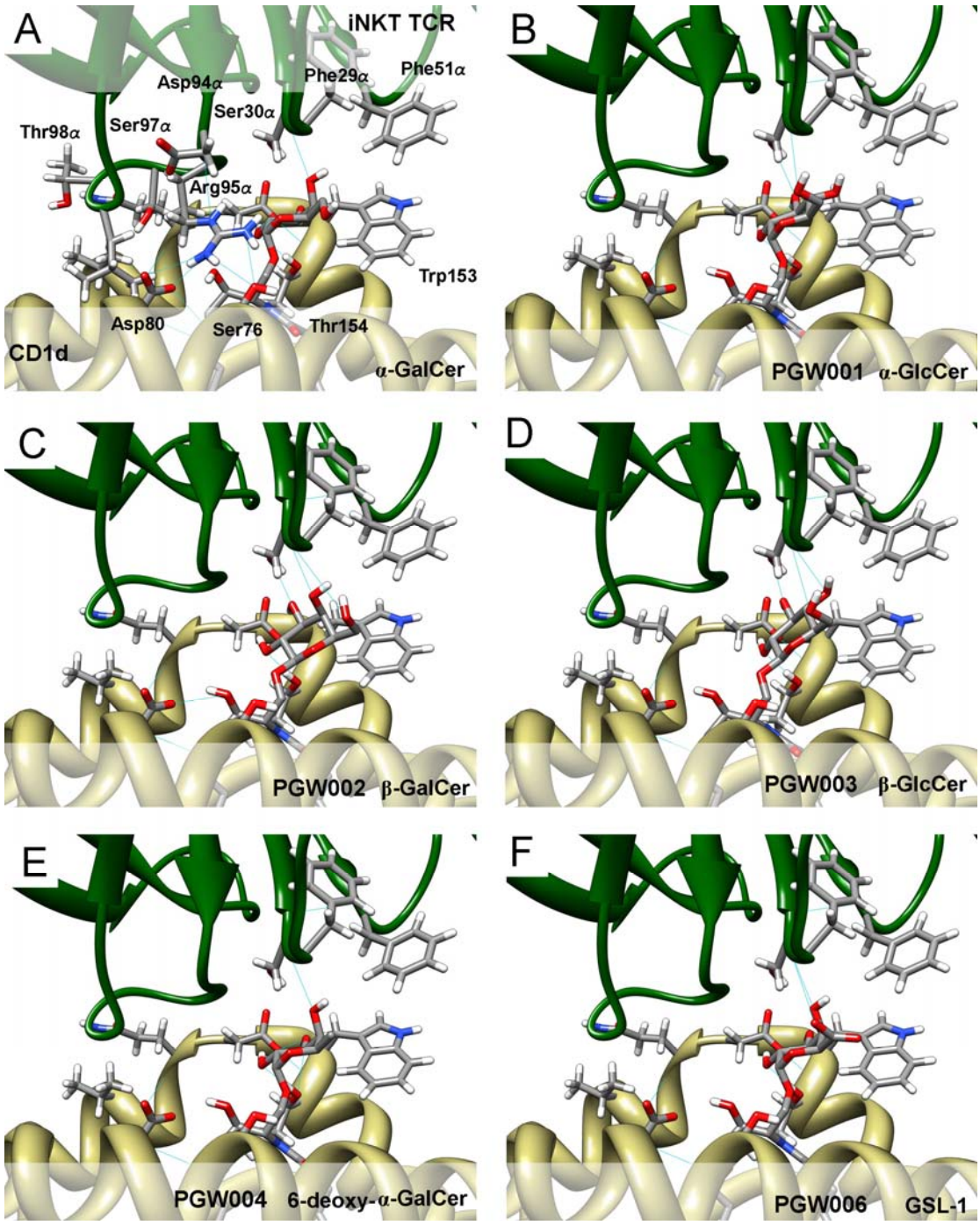


Figure 24. Docked results of experimentally tested glycolipids. (A) α-GalCer. (B) α-GlcCer [PGW001] with 4′-OH being axial. (C) β-GalCer [PGW002] with a β anomeric linkage. (D) β-GlcCer [PGW003] with both the 4′-OH being axial along with having a β anomeric linkage. (E) 6-deoxy-α-GalCer [PGW004] possessing no hydroxyl group at the 6′- position. and (F) GSL-1 [PGW006] wherein a carboxylic acid is at the 6′- position. The α- linked glycolipids all possessed similar docked orientations whereas the β- linked glycolipids developed orientations similar to that of the crystal structure orientations of β- linked glycolipids.

This issue became further evident upon analyzing the docked orientations of the 2′-, 3′-, and 4′- glycolipid derivatives wherein the substitution is very minimal such as with H-, NH_2-, OMe-, and the NHAc- substituents (**Figure 25**). The three deoxy derivatives appear to have maintained the same orientation as α-GalCer. The methoxy substitution showed a slight shift in the sugar head to compensate for the loss of hydrogen bond in the 2′-OMe, however, the 3′-OMe was able to maintain the hydrogen bond with Ser30α not permitting the 2′-OH to swing and compensate for the loss in hydrogen bonding to Asp151. The 4′-OMe analog remained in the same orientation as α-GalCer. Both the 2′- and 3′-NH_2 analogs were oriented same as α-GalCer which was to be expected considering they are both capable of providing hydrogen bond donors much like the hydroxyl group. The 4′-NH_2 was able to form a hydrogen bond with the CDR1α loop considering its substitution is bidirectional with the hydrogens pointing in opposite directions unlike α-GalCer where its hydroxyl hydrogen is aimed at the Trp153 residue. The bulky substitution of NHAc onto the galactose was the first to show deviations from the binding orientation of α-GalCer due to substitutions on the sugar. There appeared to be a small pocket that can accommodate the 2′-NHAc group allowing for the 3′-OH to maintain its hydrogen bond with Asp151, whereas the 3′-NHAc had to shift the sugar away from Asp151 it to fit into the same pocket that accommodated the 2′- substitution. Even the 4′-NHAc seemed to make a similar shift rather than maintaining both hydrogen bonds to Asp151, probably to avoid clashing with the Phe51α on the CDR2α loop.

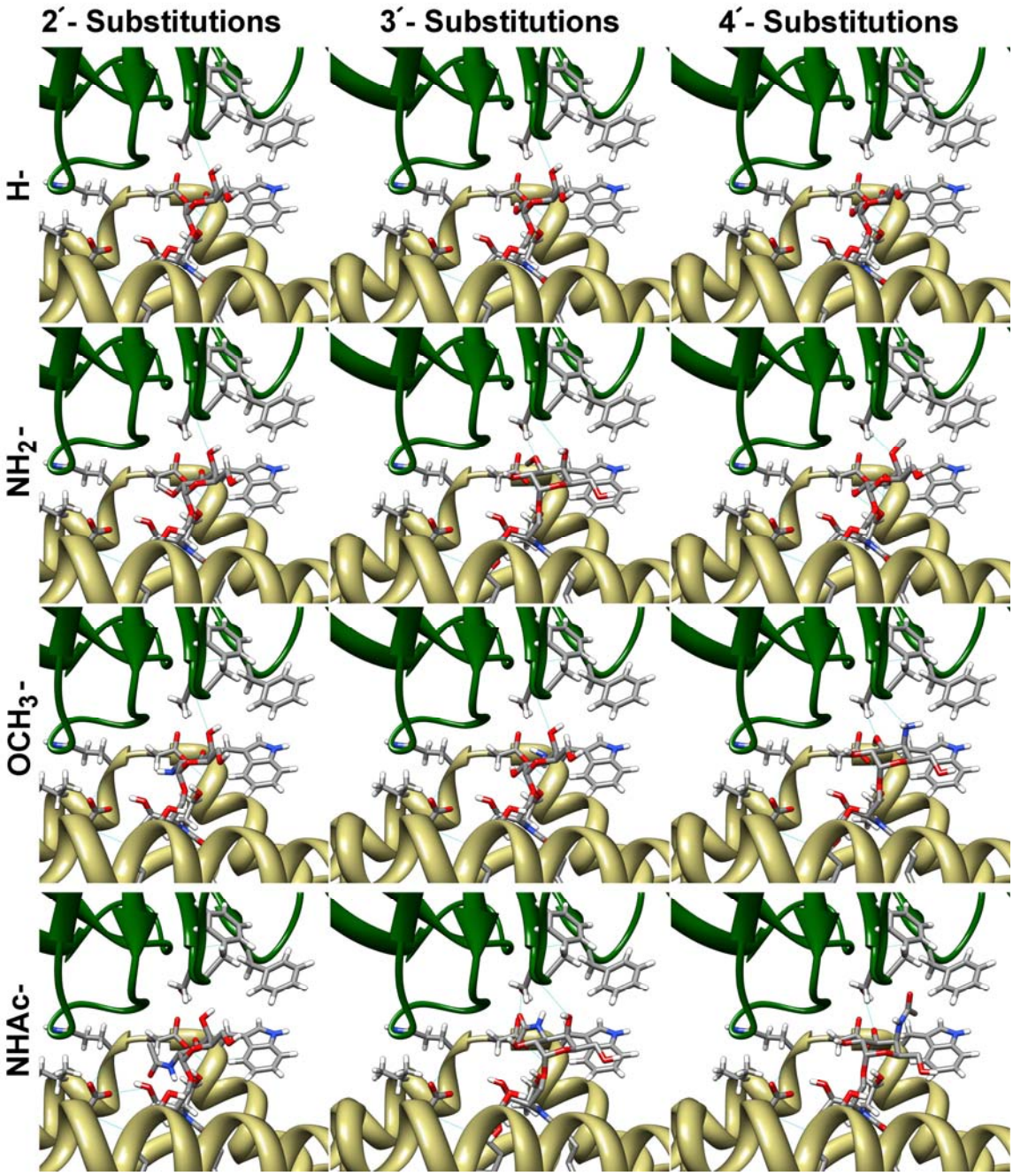


Figure 25. Docked orientations of comparing small substitutions at 2´-, 3´-, and 4´- positions on α-GalCer. The substitutions are H-, NH_2-, OMe-, and the NHAc-, respectively. Noticeable changes in sugar orientation are only visible with the bulkier substituent of NHAc with the small substitutions showing almost no deviation from the binding orientation of α-GalCer. Furthermore, the small substitutions also show minimal variation regardless of on what position the substituent was placed.

A more noticeable difference was observed when the α-GalCer analogs with bulky substituents were docked; these being N_3-, NHIv-, OCH_2CH_2OH-, and $NHCH_2PhOH$ (**Figure 26**). The 2′- substitutions were all accommodated within the small space beside Asp151 as was done for 2′-NHAc except for 2′-NHIv whose substitution did not allow for a reasonable docked orientation where the sugar was completely rotated with the 6′-OH was hydrogen bonding to Asp151. The 3′- substitutions also were found to cause a rotation of the sugar away from Asp151 to fit the bulky substituent into the same space as was occupied by the 2′- substituents. Only the 4′- modifications were able to maintain somewhat similar binding orientations as α-GalCer with both the 2′- and 3′-OH groups maintaining their hydrogen bonds to Asp151. From these preliminary results, it was concluded that the perturbation of the sugar must play a role in negating TCR recognition and thereby cancelling out iNKT activity. Yet, such a small change as replacing the OH with an H should not be the cause for such a drastic difference in iNKT stimulatory profiles. Even with the bulkier substituents, the 3′-OH appears to remain hydrogen bonded to Asp151 when there was a loss of hydrogen bonding at the 2′- position.

The 4′- analogs (PGW50# and PGW60#) were built based on the hypothesis that the 5-7 Å space between Phe51α and Trp153 should be able to accommodate larger substituents. However, the two aromatic residues were just close enough to force the substituents away towards the CDR2α loop with the exception of the less flexible amide linked substituents (**Figure 27**). The overlay of all the PGW50# compounds on each other along with the overlay of all the PGW60# compounds showed that regardless of their linkers, lengths, or bulkiness they all bind in a similar fashion. When the amide aromatic

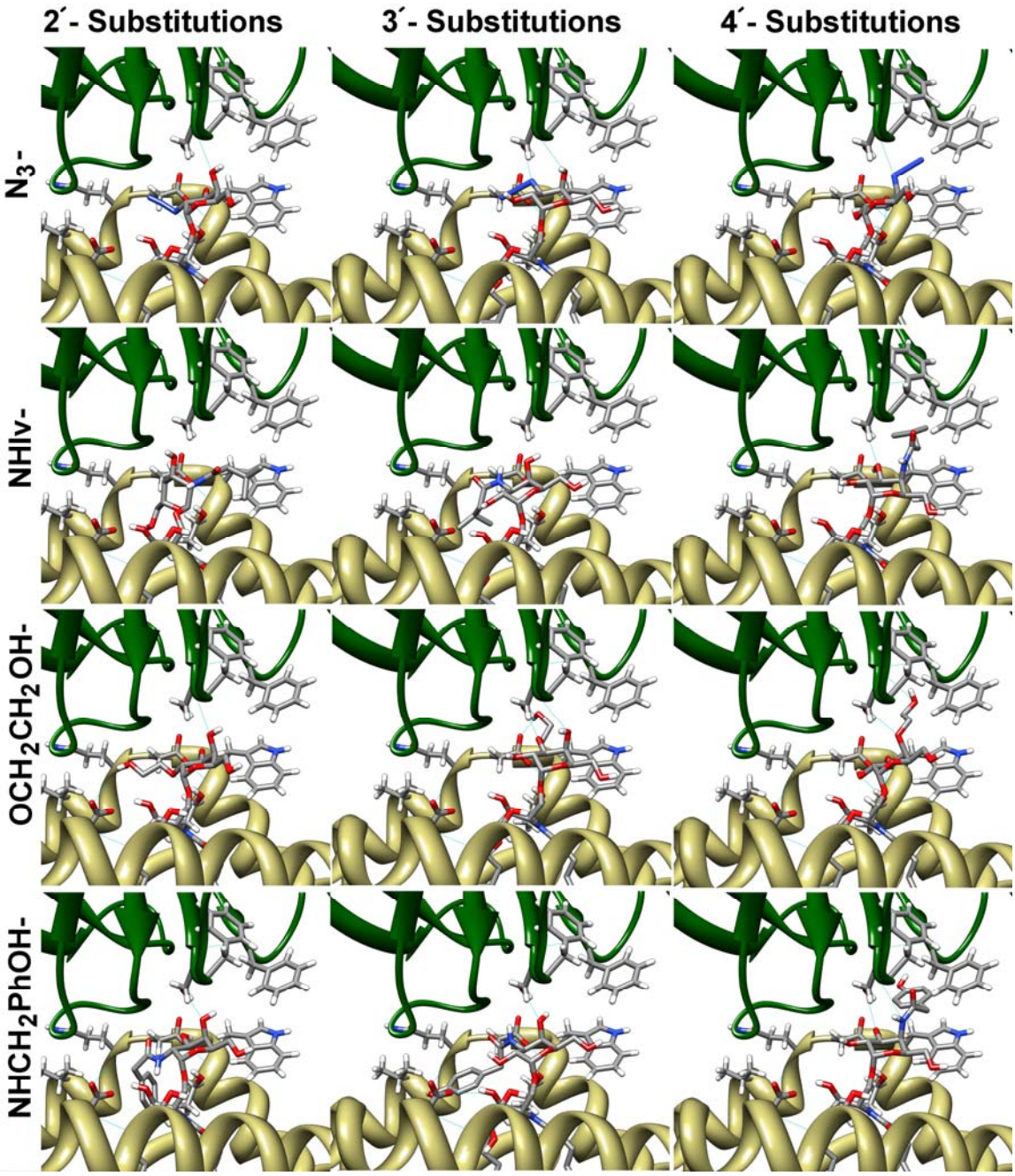


Figure 26. Docked orientations of comparing bulkier substitutions at 2´-, 3´-, and 4´- positions on α-GalCer. The substitutions are N_3-, NHIv-, OCH_2CH_2OH-, and $NHCH_2PhOH$, respectively. Unlike with the smaller substitutions, these compounds show noticeable changes upon changes in substituent position. Both the 2´- and 3´- substituted analogs shift to accommodate the bulky group in the small space beside Asp151, whereas the 4´- analogs are capable of maintaining their hydrogen bonds to Asp151.

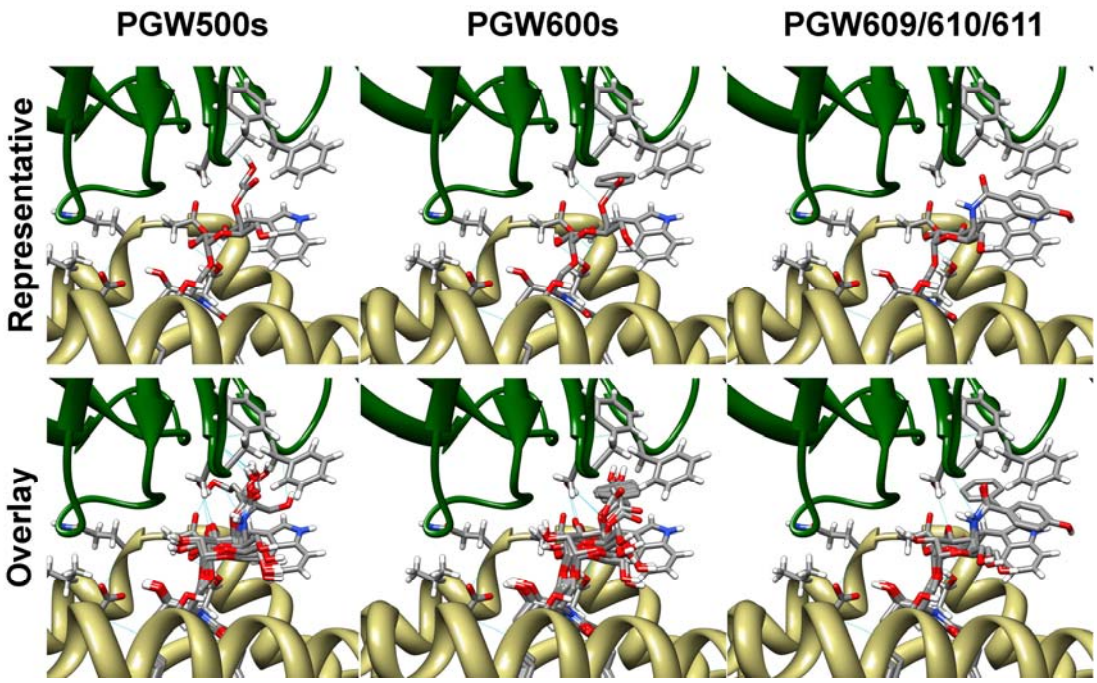


Figure 27. Docked orientations of the PGW50# and PGW60# series. Regardless of linker, length, and bulkiness, the analogs all bound in similar orientation wherein the group was accommodated and bound in the cavity below the CDR2α loop. The top row shows the representative compound of each series, PGW501, PGW601, and PGW609, respectively. Only the rigid, short PGW609 and PGW610 substitutions were found capable of intercalating between Phe51α and Trp153.

compounds were docked, only the longer substituent (PGW611) was able to bind in a similar fashion as the other PGW500s and PGW600s compounds. The short, rigid PGW609 and PGW610 were forced to intercalate between the aromatic residues (**Figure 27**). Since the 2′- and 3′- OH groups maintained their hydrogen bonds to Asp151, it is quite probable that 4′- substituents should be tolerated by the system as long as the interactions with the CDR1α and CDR2α loops are not detrimental to TCR binding.

The analysis of the docked orientations to the crystal structure were not conclusive enough to correlate iNKT stimulatory profiles to the structural changes of the glycolipid. Even though the 2′- and 3′- bulky analogs caused the sugar to rotate away from Asp151,

attributing the complete lack of iNKT stimulation to the loss of one or two hydrogen bonds to Asp151 seemed unjustifiable. Especially when considering that the deoxy-residues and the amine substitutions, both yielded similar binding orientations as α-GalCer and yet neither was capable of eliciting an iNKT response. If the loss of hydrogen bonding to Asp151 or the replacement with a more potent hydrogen bond donor to Asp151 causes the complete loss of iNKT stimulatory activity, something else must be occurring upon binding of these glycolipid analogs that cause TCR to not recognize their CD1d presentation.

After all of the docked orientations into the rest of the snapshots from the tertiary complex and CD1d were analyzed and found to be the same throughout, the binding energies were then analyzed to determine if there was a correlation between energy and loss of iNKT activity. It is accepted that the binding kinetics between TCR and the CD1d presenting glycolipid complex have been attributed to determining iNKT response wherein those CD1d/glycolipid complexes that do not reach the binding kinetic threshold do not cause stimulation to occur and those that bind well causing the binding affinity to increase will cause a rapid immune response cascade that immediately fades. Therefore, it was hypothesized that glycolipids with similar binding energies to α-GalCer should cause similar iNKT stimulation profiles with those that bind stronger being way too powerful agonists and those that bind weaker not being acceptable glycolipids.

The docked energies were used to rank and compare the glycolipids but these energies could not be correlated directly to experimental binding results. Autodock is capable of providing for each docked ligand the docked energy which is the sum of the intermolecular energy and the ligand's internal energy, the binding energy which is the

sum of the intermolecular energy and the torsional free-energy penalty, and the K_i constant calculated by

$$\Delta G_{obs} = -RT \ln K_i$$

where *R* is the gas constant, 1.987 cal K^{-1} mol^{-1}, and *T* is the absolute temperature, 298.15 K. However, only the docked energy was usable as a value for comparison in the case of the glycolipid analogs due to the placed restriction on the torsional degrees of freedom. Both the binding energy and the K_i are dependent on this value which is derived from the number of active torsions multiplied by 0.3113 which is the forcefield torsional free energy parameter. Since the torsional degrees were limited, neither the binding energy nor the inhibition constant are representative of the real binding energy and inhibition constants for the system. This was further shown to be the case when the calculated binding free energies from the experimental dissociation constants (K_D) for α-GalCer to mouse CD1d (mCD1d) were compared to the binding free energies calculated by Autodock for α-GalCer binding to hCD1d where the big differences in average binding energy were -7 and -12 kcal/mol, respectively[143].

The docked energies ended up showing that α-GalCer was neither the best nor the worst ligand when it came to binding, however, the results are more complicated than would be expected upon comparing simple singular modifications on the glycolipid. In addition to the docked energies calculated for the crystal structure, an additional two sets of energies were obtained from averaging the docked energies to the tertiary complex and the CD1d protein (**Table 4**). Also, a hypothetical TCR-glycolipid binding energy was calculated by subtracting the tertiary complex binding energy from the CD1d binding energy. This is a different value than the binding of the CD1d/α-GalCer binary complex

to TCR, it instead showed the favorability of the glycolipid interaction with TCR and therefore cannot be correlated to the binding kinetic or thermodynamic data available. Currently, protein-protein docking would be the only recourse to calculate the CD1d/α-GalCer binary complex to TCR binding energy, however, this is still a very underdeveloped area.

The averaged energies show a very diverse spread of the glycolipids with seemingly no trend able to be discerned especially with α-GalCer being spread throughout the other 49 analogs. If the binding energy of α-GalCer was either the lowest or the highest, it would be easy to conclude that glycolipids with binding energies close to α-GalCer should behave similarly, however, this is not the case with α-GalCer possessing lower average binding energies than most of the glycolipids in both the tertiary complex and the CD1d protein. It does not appear that even the cluster argument can be applied to the docked energies where the theory is that those glycolipids close to α-GalCer should behave like it.

The docked energies are also quite close together with sometimes only a difference of 0.01 kcal/mol existing between the ranked analogs. The overall spread of the tertiary complex energies is ~7 kcal/mol with the CD1d docked energies being spread over ~4 kcal/mol. This minimal range for CD1d docking was expected considering that CD1d has been shown to be able to bind a diverse spectrum of glycolipids, however, the miminal spread for the tertiary complex implies that more than likely the energetic window for TCR recognition of the CD1d presented antigen will be quite narrow. The TCR glycolipid interacting energies further this idea by showing that only minimal docked energy coincides with the sugar head portion interacting with TCR. With the amount of

Table 4. Ranked docked energies of the glycolipid analogs bound to the tertiary complex. The docked energies (kcal/mol) include glycolipids docked into the crystal structure along with the averages of the glycolipids docked into groups of snapshots: 3 ns (50 ps, 1 ns, 2ns, and 3 ns – orange), 5 ns (green), and 10 ns (pink). The Analysis section contains observations from the docked results where YES showed that the 1st docked conformation was the same as the predicted correct docked orientation for the sugar, whereas NO showed that they differed which is also why an Alternative Energy (AltE) but correct docking orientation was included in the calculations. The NULL was when the glycolipid was not able to be docked by Autodock and OVERALL was an assessment where G-good and B-bad were based on having more YESes than NOs and no NULLs. (*Continued*)

ANALYSIS								Average (10 ns)				Average (5 ns)				Average (3 ns)				Crystal Structure			
YES		NO		NULL		OVERALL		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)	
303	10	303	0	0	0	0	G	607	-28.05	607	-27.43	607	-29.23	607	-28.72	607	-29.40	606	-29.16	208	-32.98	208	-32.98
403	9	4	1	2	0	1	N	601	-27.76	508	-27.42	608	-28.77	606	-28.67	505	-29.33	607	-28.99	603	-31.71	606	-31.02
4	9	402	1	3	0	2	G	508	-27.68	608	-27.22	606	-28.73	207	-28.36	606	-29.26	505	-28.89	602	-31.31	602	-30.96
402	9	403	2	407	0	3	G	604	-27.66	605	-27.12	601	-28.56	605	-28.27	601	-29.08	207	-28.69	606	-31.12	604	-30.88
205	8	205	2	506	0	4	N	608	-27.58	207	-27.06	508	-28.52	508	-28.22	608	-28.96	605	-28.54	601	-31.10	607	-30.83
207	8	207	2	404	0	5	G	606	-27.53	206	-26.92	502	-28.52	608	-28.12	502	-28.90	502	-28.52	607	-31.02	603	-30.77
6	8	6	2	403	0	6	G	603	-27.46	604	-26.90	605	-28.42	611	-28.05	208	-28.88	508	-28.41	605	-30.93	605	-30.61
203	8	203	2	303	0	201	N	502	-27.44	606	-26.84	207	-28.36	502	-28.01	207	-28.69	205	-28.34	604	-30.88	601	-30.50
2	7	204	3	401	0	202	N	505	-27.31	503	-26.81	611	-28.34	604	-27.96	605	-28.62	608	-28.25	608	-30.59	206	-30.28
3	7	208	3	201	0	203	G	605	-27.30	603	-26.71	604	-28.23	308	-27.79	508	-28.57	611	-28.24	611	-30.57	406	-30.24
204	7	503	3	4	0	204	G	206	-27.28	505	-26.70	408	-28.04	206	-27.76	611	-28.51	303	-28.21	408	-30.56	408	-30.11
208	7	0	4	305	0	205	G	408	-27.13	601	-26.64	505	-27.96	408	-27.73	507	-28.39	405	-28.16	505	-30.39	508	-30.07
503	7	2	4	405	0	206	N	207	-27.06	502	-26.62	206	-27.93	205	-27.72	408	-28.37	307	-28.08	503	-30.33	505	-30.03
508	7	3	4	402	0	207	N	503	-27.02	501	-26.61	603	-27.87	405	-27.50	205	-28.34	0	-27.95	206	-30.28	611	-29.98
0	6	401	4	202	0	208	G	308	-26.92	407	-26.48	205	-27.85	303	-27.46	405	-28.25	6	-27.88	502	-30.07	507	-29.83
407	6	305	4	205	0	301	G	501	-26.84	303	-26.46	503	-27.84	503	-27.46	307	-28.24	408	-27.85	508	-30.07	503	-29.70
506	6	202	4	204	0	302	G	611	-26.80	308	-26.44	308	-27.84	203	-27.45	407	-28.22	507	-27.84	406	-29.98	504	-29.65
401	6	206	4	208	0	303	N	407	-26.80	610	-26.30	407	-27.82	407	-27.44	303	-28.21	203	-27.84	504	-29.97	502	-29.43
305	6	508	4	307	0	304	N	507	-26.71	602	-26.28	204	-27.72	2	-27.43	204	-28.12	2	-27.83	507	-29.83	407	-29.19
405	6	605	4	5	0	305	N	504	-26.67	205	-26.27	405	-27.66	204	-27.38	501	-28.10	407	-27.83	407	-29.19	306	-29.13
202	6	602	4	6	0	306	G	609	-26.67	611	-26.18	507	-27.66	603	-27.31	2	-28.07	503	-27.79	306	-29.13	207	-29.00
206	6	608	4	203	0	307	G	506	-26.65	507	-26.16	2	-27.60	406	-27.24	206	-28.04	604	-27.79	501	-29.09	307	-28.99
605	6	1	5	502	0	308	G	610	-26.61	405	-26.14	609	-27.53	1	-27.23	503	-28.02	206	-27.78	207	-29.00	205	-28.81
609	6	501	5	503	0	401	G	602	-26.52	2	-26.12	3	-27.51	0	-27.21	3	-27.96	406	-27.69	307	-28.99	501	-28.78
602	6	407	5	601	0	402	N	306	-26.51	203	-26.11	208	-27.50	609	-27.20	0	-27.95	501	-27.66	205	-28.81	308	-28.64
608	6	506	5	603	0	403	N	406	-26.48	204	-26.10	0	-27.47	307	-27.19	6	-27.88	5	-27.65	308	-28.64	506	-28.53
1	5	304	5	605	0	404	G	2	-26.48	3	-26.09	303	-27.46	505	-27.16	406	-27.88	204	-27.61	506	-28.53	405	-28.31
501	5	301	5	608	0	405	N	303	-26.46	506	-26.06	203	-27.46	6	-27.14	604	-27.86	308	-27.60	405	-28.31	608	-28.27
404	5	405	5	1	1	406	N	405	-26.37	408	-26.05	306	-27.43	403	-27.13	203	-27.84	1	-27.51	204	-28.30	204	-28.24

Table 4 continued

608	6	**506**	5	**603**	0	**403**	N	**406**	-26.48	**204**	-26.10	**0**	-27.47	**307**	-27.19	**6**	-27.88	**5**	-27.65	**308**	-28.64	**506**	-28.53
1	5	**304**	5	**605**	0	**404**	G	**2**	-26.48	**3**	-26.09	**303**	-27.46	**505**	-27.16	**406**	-27.88	**204**	-27.61	**506**	-28.53	**405**	-28.31
501	5	**301**	5	**608**	0	**405**	N	**303**	-26.46	**506**	-26.06	**203**	-27.46	**6**	-27.14	**604**	-27.86	**308**	-27.60	**405**	-28.31	**608**	-28.27
404	5	**405**	5	**1**	1	**406**	N	**405**	-26.37	**408**	-26.05	**306**	-27.43	**403**	-27.13	**203**	-27.84	**1**	-27.51	**204**	-28.30	**204**	-28.24
304	5	**308**	5	**501**	1	**407**	G	**205**	-26.37	**504**	-26.05	**406**	-27.41	**3**	-27.13	**5**	-27.70	**202**	-27.50	**2**	-28.20	**404**	-28.05
301	5	**5**	5	**304**	1	**408**	G	**3**	-26.36	**406**	-26.04	**6**	-27.38	**5**	-27.00	**308**	-27.68	**403**	-27.50	**404**	-28.05	**304**	-27.97
201	5	**504**	5	**301**	1	**501**	G	**204**	-26.30	**301**	-25.99	**307**	-27.38	**507**	-26.95	**403**	-27.61	**208**	-27.41	**3**	-28.00	**303**	-27.89
302	5	**609**	5	**306**	1	**502**	N	**208**	-26.24	**609**	-25.94	**1**	-27.28	**202**	-26.94	**1**	-27.58	**3**	-27.38	**304**	-27.97	**305**	-27.87
308	5	**604**	5	**406**	1	**503**	G	**203**	-26.24	**6**	-25.93	**403**	-27.20	**602**	-26.93	**202**	-27.51	**4**	-27.28	**303**	-27.89	**610**	-27.81
5	5	**404**	6	**302**	1	**504**	N	**307**	-26.19	**0**	-25.79	**504**	-27.20	**610**	-26.86	**304**	-27.39	**609**	-27.27	**305**	-27.87	**609**	-27.80
504	5	**201**	6	**206**	1	**505**	N	**6**	-26.15	**402**	-25.75	**602**	-27.19	**4**	-26.85	**603**	-27.32	**402**	-27.08	**610**	-27.81	**202**	-27.79
507	5	**406**	6	**504**	1	**506**	N	**0**	-26.15	**307**	-25.74	**5**	-27.18	**501**	-26.75	**302**	-27.30	**301**	-27.07	**609**	-27.80	**6**	-27.72
601	5	**302**	6	**505**	1	**507**	N	**301**	-26.12	**208**	-25.69	**610**	-27.17	**301**	-26.71	**4**	-27.28	**401**	-26.98	**1**	-27.79	**302**	-27.71
603	5	**505**	6	**507**	1	**508**	N	**1**	-25.88	**403**	-25.68	**506**	-27.11	**402**	-26.70	**306**	-27.27	**603**	-26.90	**202**	-27.79	**5**	-27.64
604	5	**507**	6	**508**	1	**601**	N	**202**	-25.85	**306**	-25.68	**501**	-27.05	**601**	-26.68	**609**	-27.27	**302**	-26.82	**6**	-27.72	**1**	-27.53
606	5	**601**	6	**607**	1	**602**	G	**5**	-25.83	**202**	-25.67	**304**	-26.99	**401**	-26.61	**402**	-27.22	**305**	-26.72	**302**	-27.71	**402**	-27.51
610	5	**603**	6	**611**	1	**603**	N	**404**	-25.82	**1**	-25.65	**302**	-26.96	**302**	-26.58	**301**	-27.17	**201**	-26.68	**5**	-27.64	**0**	-27.45
406	4	**606**	6	**602**	1	**604**	N	**304**	-25.81	**4**	-25.64	**202**	-26.95	**305**	-26.55	**305**	-27.15	**404**	-26.47	**402**	-27.51	**403**	-27.38
505	4	**610**	6	**604**	1	**605**	N	**402**	-25.80	**5**	-25.53	**4**	-26.85	**208**	-26.53	**201**	-27.07	**504**	-26.47	**0**	-27.45	**2**	-27.23
607	4	**306**	7	**606**	1	**606**	N	**403**	-25.75	**401**	-25.44	**305**	-26.84	**306**	-26.37	**504**	-26.99	**601**	-26.46	**403**	-27.38	**3**	-27.14
306	3	**607**	7	**207**	2	**607**	N	**305**	-25.74	**305**	-25.23	**402**	-26.79	**201**	-26.36	**401**	-26.98	**306**	-26.42	**203**	-27.14	**203**	-27.14
307	3	**408**	7	**308**	2	**608**	N	**401**	-25.72	**201**	-25.09	**301**	-26.78	**506**	-26.31	**506**	-26.89	**602**	-26.37	**201**	-27.10	**4**	-27.08
611	3	**307**	8	**609**	2	**609**	G	**4**	-25.65	**302**	-24.69	**201**	-26.70	**404**	-26.24	**404**	-26.84	**610**	-26.15	**4**	-27.08	**401**	-26.87
408	3	**502**	8	**610**	2	**610**	N	**201**	-25.44	**304**	-24.36	**401**	-26.66	**504**	-26.16	**610**	-26.64	**506**	-26.11	**301**	-26.92	**301**	-26.53
502	2	**611**	8	**408**	2	**611**	N	**302**	-25.18	**404**	-20.80	**404**	-26.61	**304**	-25.07	**602**	-26.49	**304**	-24.97	**401**	-26.87	**201**	-26.45

Table 5. Ranked docked energies of glycolipid analogs bound to the CD1d protein. The docked energies (kcal/mol) include glycolipids docked into the crystal structure along with the averages of the glycolipids docked into groups of snapshots: 3 ns (50 ps, 1 ns, 2ns, and 3 ns – orange), 5 ns (green), and 10 ns (pink). The Analysis section contains observations from the docked results where YES showed that the 1st docked conformation was the same as the predicted correct docked orientation for the sugar, whereas NO showed that they differed which is also why an Alternative Energy (AltE) but correct docking orientation was included in the calculations. The NULL was when the glycolipid was not able to be docked by Autodock and OVERALL was an assessment where G-good and B-bad were based on having more YESes than NOs and no NULLs. (*Continued*)

ANALYSIS								Average (10 ns)				Average (5 ns)				Average (3 ns)				Crystal Structure			
YES		NO		NULL		OVERALL		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)	
303	10	**303**	0	**0**		**0**	G	**208**	-28.04	**604**	-27.69	**608**	-27.93	**604**	-27.79	**608**	-27.87	**608**	-27.87	**604**	-29.95	**604**	-29.95
203	9	**203**	1	**1**		**1**	N	**611**	-27.92	**208**	-27.67	**604**	-27.85	**608**	-27.65	**607**	-27.76	**607**	-27.67	**608**	-29.92	**608**	-29.92
407	8	**506**	2	**2**		**2**	N	**608**	-27.86	**608**	-27.66	**611**	-27.83	**208**	-27.33	**208**	-27.74	**604**	-27.44	**208**	-29.56	**208**	-29.56
506	8	**202**	2	**3**		**3**	N	**604**	-27.83	**611**	-27.28	**208**	-27.64	**508**	-27.20	**611**	-27.69	**603**	-27.43	**603**	-29.30	**603**	-29.30
202	8	**504**	2	**501**		**4**	G	**408**	-27.78	**603**	-27.27	**603**	-27.64	**603**	-27.17	**603**	-27.61	**505**	-27.29	**607**	-29.30	**607**	-29.30
504	8	**605**	2	**407**		**5**	G	**308**	-27.64	**607**	-27.25	**408**	-27.52	**601**	-27.04	**604**	-27.53	**208**	-27.28	**505**	-29.23	**505**	-29.23
605	8	**501**	3	**506**		**6**	G	**603**	-27.62	**601**	-27.13	**206**	-27.47	**611**	-27.00	**505**	-27.43	**611**	-27.23	**503**	-29.22	**503**	-29.22
604	8	**407**	3	**404**		**201**	G	**206**	-27.58	**605**	-27.11	**607**	-27.45	**607**	-26.97	**308**	-27.41	**503**	-27.09	**602**	-28.87	**602**	-28.87
608	8	**401**	3	**304**		**202**	N	**607**	-27.54	**508**	-27.09	**505**	-27.38	**605**	-26.95	**408**	-27.31	**508**	-26.97	**601**	-28.77	**601**	-28.77
501	7	**402**	3	**403**		**203**	G	**610**	-27.50	**505**	-27.00	**508**	-27.37	**505**	-26.88	**508**	-27.23	**203**	-26.88	**606**	-28.72	**606**	-28.72
403	7	**607**	3	**303**		**204**	G	**508**	-27.44	**303**	-26.99	**308**	-27.37	**504**	-26.79	**206**	-27.19	**602**	-26.88	**508**	-28.52	**508**	-28.52
401	7	**604**	3	**401**		**205**	G	**505**	-27.40	**203**	-26.88	**610**	-27.21	**609**	-26.76	**503**	-27.09	**601**	-26.86	**605**	-28.41	**605**	-28.41
402	7	**606**	3	**301**		**206**	N	**609**	-27.38	**506**	-26.81	**503**	-27.18	**303**	-26.73	**601**	-27.04	**408**	-26.85	**611**	-28.34	**611**	-28.34
503	7	**608**	3	**201**		**207**	G	**601**	-27.29	**503**	-26.79	**601**	-27.17	**602**	-26.69	**605**	-26.95	**605**	-26.84	**408**	-28.27	**408**	-28.27
607	7	**404**	4	**4**		**208**	N	**306**	-27.29	**408**	-26.77	**609**	-27.11	**503**	-26.69	**610**	-26.93	**606**	-26.83	**504**	-28.24	**504**	-28.24
602	7	**403**	4	**305**		**301**	N	**503**	-27.24	**606**	-26.72	**605**	-27.02	**506**	-26.66	**606**	-26.92	**507**	-26.80	**303**	-28.14	**303**	-28.14
606	7	**502**	4	**405**		**302**	G	**605**	-27.21	**504**	-26.66	**306**	-27.00	**203**	-26.58	**203**	-26.88	**504**	-26.71	**502**	-28.14	**502**	-28.14
0	6	**503**	4	**306**		**303**	N	**507**	-27.12	**609**	-26.63	**602**	-26.95	**606**	-26.55	**602**	-26.88	**303**	-26.69	**308**	-28.05	**609**	-27.99
404	6	**601**	4	**406**		**304**	N	**406**	-27.12	**602**	-26.61	**406**	-26.92	**405**	-26.50	**609**	-26.86	**502**	-26.68	**609**	-27.99	**507**	-27.93
201	6	**603**	4	**302**		**305**	N	**203**	-27.01	**507**	-26.55	**507**	-26.90	**407**	-26.40	**306**	-26.85	**206**	-26.63	**507**	-27.93	**206**	-27.90
405	6	**602**	4	**402**		**306**	G	**303**	-26.99	**406**	-26.50	**606**	-26.85	**408**	-26.40	**507**	-26.81	**506**	-26.48	**206**	-27.90	**207**	-27.72
204	6	**0**	5	**202**		**307**	G	**606**	-26.99	**502**	-26.43	**203**	-26.83	**406**	-26.35	**504**	-26.71	**609**	-26.33	**306**	-27.86	**406**	-27.68
208	6	**201**	5	**205**		**308**	N	**602**	-26.97	**407**	-26.32	**504**	-26.79	**502**	-26.28	**303**	-26.69	**407**	-26.31	**207**	-27.72	**610**	-27.68
502	6	**405**	5	**204**		**401**	G	**307**	-26.91	**206**	-26.28	**303**	-26.73	**507**	-26.25	**502**	-26.68	**207**	-26.31	**406**	-27.68	**407**	-27.66
507	6	**204**	5	**206**		**402**	N	**504**	-26.86	**404**	-26.28	**506**	-26.70	**402**	-26.12	**406**	-26.62	**406**	-26.29	**610**	-27.68	**506**	-27.65
601	6	**208**	5	**207**		**403**	N	**506**	-26.86	**402**	-26.23	**5**	-26.69	**207**	-26.07	**506**	-26.55	**404**	-26.28	**407**	-27.66	**203**	-27.63
603	6	**505**	5	**208**		**404**	G	**502**	-26.86	**207**	-26.23	**502**	-26.69	**403**	-26.03	**204**	-26.49	**405**	-26.21	**506**	-27.65	**501**	-27.56
406	5	**507**	5	**307**		**405**	N	**204**	-26.71	**501**	-26.17	**6**	-26.66	**404**	-25.99	**1**	-26.48	**204**	-26.07	**203**	-27.63	**204**	-27.35
207	5	**301**	6	**308**		**406**	N	**207**	-26.69	**403**	-26.13	**207**	-26.63	**401**	-25.96	**207**	-26.43	**501**	-26.05	**501**	-27.56	**405**	-27.32

Table 5 continued

505	5	**4**	6	**5**		**407**	N	**304**	-26.67	**401**	-26.13	**307**	-26.60	**5**	-25.95	**307**	-26.41	**403**	-26.02	**204**	-27.35	**404**	-27.28
508	5	**406**	6	**6**		**408**	N	**405**	-26.58	**205**	-26.12	**405**	-26.56	**4**	-25.93	**6**	-26.36	**402**	-25.96	**405**	-27.32	**403**	-27.15
609	5	**205**	6	**203**		**501**	G	**404**	-26.55	**204**	-26.05	**204**	-26.53	**0**	-25.90	**305**	-26.33	**205**	-25.92	**404**	-27.28	**205**	-27.14
301	4	**207**	6	**502**		**502**	G	**6**	-26.54	**4**	-26.02	**1**	-26.47	**308**	-25.89	**5**	-26.33	**308**	-25.91	**403**	-27.15	**402**	-26.99
4	4	**5**	6	**503**		**503**	G	**5**	-26.53	**202**	-25.95	**205**	-26.45	**501**	-25.82	**407**	-26.31	**4**	-25.86	**205**	-27.14	**0**	-26.85
305	4	**6**	6	**504**		**504**	G	**205**	-26.52	**610**	-25.95	**407**	-26.40	**202**	-25.82	**405**	-26.31	**6**	-25.82	**402**	-26.99	**4**	-26.85
205	4	**508**	6	**505**		**505**	G	**1**	-26.52	**5**	-25.91	**404**	-26.40	**610**	-25.80	**2**	-26.28	**307**	-25.81	**0**	-26.85	**401**	-26.76
5	4	**609**	6	**507**		**506**	G	**501**	-26.48	**6**	-25.85	**304**	-26.40	**205**	-25.75	**404**	-26.28	**401**	-25.81	**4**	-26.85	**6**	-26.64
6	4	**305**	7	**508**		**507**	N	**305**	-26.48	**0**	-25.77	**2**	-26.34	**307**	-25.73	**304**	-26.24	**610**	-25.72	**305**	-26.78	**1**	-26.62
408	4	**408**	7	**601**		**508**	G	**407**	-26.45	**201**	-25.76	**305**	-26.33	**206**	-25.66	**205**	-26.22	**202**	-25.68	**401**	-26.76	**202**	-26.59
611	3	**1**	8	**603**		**601**	G	**2**	-26.35	**3**	-25.74	**501**	-26.30	**304**	-25.62	**4**	-26.12	**0**	-25.65	**6**	-26.64	**2**	-26.44
1	2	**611**	8	**605**		**602**	G	**403**	-26.29	**306**	-25.72	**4**	-26.19	**3**	-25.57	**501**	-26.11	**5**	-25.47	**1**	-26.62	**307**	-26.28
2	2	**2**	9	**607**		**603**	G	**402**	-26.27	**308**	-25.71	**402**	-26.16	**6**	-25.56	**3**	-26.04	**2**	-25.46	**202**	-26.59	**301**	-26.27
206	2	**3**	9	**609**		**604**	N	**4**	-26.22	**307**	-25.69	**3**	-26.16	**201**	-25.50	**403**	-26.03	**3**	-25.43	**3**	-26.51	**5**	-26.24
307	2	**206**	9	**611**		**605**	N	**3**	-26.17	**2**	-25.63	**403**	-26.10	**301**	-25.49	**402**	-26.03	**1**	-25.32	**2**	-26.44	**201**	-26.22
610	2	**307**	9	**602**		**606**	G	**401**	-26.16	**301**	-25.63	**201**	-26.05	**2**	-25.45	**0**	-25.87	**301**	-25.23	**307**	-26.41	**3**	-26.02
3	1	**610**	9	**604**		**607**	G	**201**	-26.13	**304**	-25.61	**0**	-26.04	**302**	-25.44	**401**	-25.83	**201**	-25.18	**301**	-26.27	**306**	-26.02
304	1	**304**	10	**606**		**608**	G	**0**	-26.10	**1**	-25.59	**401**	-25.99	**204**	-25.38	**201**	-25.82	**304**	-25.07	**5**	-26.24	**308**	-25.58
302	1	**302**	10	**608**		**609**	G	**302**	-26.00	**302**	-25.49	**202**	-25.87	**1**	-25.27	**202**	-25.75	**302**	-25.07	**201**	-26.22	**305**	-25.34
308	1	**308**	10	**610**		**610**	N	**202**	-26.00	**305**	-25.18	**302**	-25.86	**306**	-25.19	**302**	-25.69	**306**	-25.02	**304**	-26.05	**302**	-25.28
306	0	**306**	11	**408**		**611**	N	**301**	-25.83	**405**	-21.73	**301**	-25.76	**305**	-24.62	**301**	-25.52	**305**	-23.77	**302**	-25.66	**304**	-25.10

Table 6. Ranked energies of the glycolipid analogs interacting with TCR. The docked energies (kcal/mol) include glycolipids docked into the crystal structure along with the averages of the glycolipids docked into groups of snapshots: 3 ns (50 ps, 1 ns, 2ns, and 3 ns – orange), 5 ns (green), and 10 ns (pink). The Analysis section contains observations from the docked results where YES showed that the 1st docked conformation was the same as the predicted correct docked orientation for the sugar, whereas NO showed that they differed which is also why an Alternative Energy (AltE) but correct docking orientation was included in the calculations. The NULL was when the glycolipid was not able to be docked by Autodock and OVERALL was an assessment where G-good and B-bad were based on having more YESes than NOs and no NULLs. (*Continued*)

ANALYSIS								Average (10 ns)				Average (5 ns)				Average (3 ns)				Crystal Structure			
YES		NO		NULL		OVERALL		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)	
303	10	303	0	0		0	N	502	-0.58	405	-4.41	207	-1.99	206	-2.10	207	-2.27	305	-2.96	208	-3.42	208	-3.42
402	7	402	3	1		1	N	207	-0.51	308	-0.82	502	-1.84	207	-2.08	502	-2.21	205	-2.42	307	-2.58	306	-3.11
203	7	203	3	2		2	N	601	-0.46	206	-0.80	302	-1.49	204	-2.00	205	-2.12	207	-2.38	605	-2.52	308	-3.06
403	6	403	5	3		3	N	501	-0.46	207	-0.73	0	-1.43	1	-2.00	0	-2.08	2	-2.38	602	-2.44	304	-2.87
401	5	401	5	501		4	N	301	-0.36	501	-0.52	407	-1.42	2	-1.98	601	-2.04	0	-2.30	603	-2.41	307	-2.71
202	5	202	5	407		5	N	407	-0.35	2	-0.49	205	-1.40	205	-1.97	501	-1.99	307	-2.28	606	-2.40	406	-2.56
207	5	4	6	506		6	N	3	-0.19	301	-0.44	605	-1.40	305	-1.93	405	-1.94	1	-2.20	206	-2.38	305	-2.53
608	5	205	6	404		201	N	2	-0.13	3	-0.35	601	-1.39	502	-1.73	3	-1.92	5	-2.18	601	-2.33	302	-2.43
0	4	207	6	304		202	N	605	-0.09	508	-0.34	3	-1.36	302	-1.61	407	-1.91	6	-2.06	406	-2.30	206	-2.38
407	4	6	6	403		203	G	0	-0.04	610	-0.30	2	-1.26	6	-1.58	505	-1.91	3	-1.96	408	-2.29	606	-2.30
506	4	504	6	303		204	G	508	-0.01	502	-0.18	204	-1.18	3	-1.56	307	-1.83	405	-1.95	611	-2.23	605	-2.20
201	4	605	6	401		205	G	505	0.08	407	-0.16	403	-1.10	308	-1.53	2	-1.79	301	-1.84	302	-2.05	602	-2.09
4	4	608	6	301		206	N	202	0.14	205	-0.14	405	-1.10	307	-1.47	202	-1.76	502	-1.83	502	-1.93	507	-1.90
205	4	0	7	201		207	N	604	0.15	6	-0.08	202	-1.09	605	-1.32	605	-1.67	202	-1.83	304	-1.92	408	-1.84
204	4	501	7	4		208	N	205	0.15	204	-0.05	1	-1.08	0	-1.32	301	-1.65	302	-1.76	507	-1.90	601	-1.73
208	4	407	7	305		301	N	603	0.16	305	-0.05	301	-1.02	301	-1.23	204	-1.63	605	-1.70	2	-1.76	205	-1.67
6	4	506	7	405		302	N	504	0.18	307	-0.04	608	-0.85	202	-1.12	302	-1.62	501	-1.61	504	-1.73	611	-1.64
503	4	201	7	306		303	N	206	0.20	1	-0.04	508	-0.82	403	-1.10	403	-1.58	505	-1.60	607	-1.72	508	-1.55
504	4	204	7	406		304	N	405	0.21	0	-0.02	307	-0.78	408	-1.08	6	-1.53	204	-1.54	205	-1.67	407	-1.53
601	4	208	7	302		305	N	506	0.22	503	-0.02	501	-0.75	5	-1.06	303	-1.52	303	-1.52	508	-1.55	607	-1.53
603	4	503	7	402		306	G	503	0.22	605	-0.01	303	-0.73	407	-1.04	5	-1.37	407	-1.52	501	-1.53	603	-1.47
605	4	601	7	202		307	G	608	0.28	202	0.28	6	-0.73	508	-1.02	201	-1.24	201	-1.51	407	-1.53	504	-1.41
607	4	603	7	205		308	N	6	0.39	306	0.30	401	-0.67	405	-1.00	402	-1.19	403	-1.48	3	-1.49	5	-1.40
609	4	607	7	204		401	N	204	0.42	505	0.31	503	-0.67	501	-0.93	4	-1.16	508	-1.44	5	-1.40	502	-1.29
501	3	609	7	206		402	N	401	0.44	602	0.32	4	-0.66	4	-0.93	401	-1.15	4	-1.42	207	-1.28	207	-1.28
404	3	404	8	207		403	N	402	0.47	5	0.38	201	-0.65	203	-0.87	304	-1.15	401	-1.18	306	-1.27	501	-1.22
405	3	301	8	208		404	N	602	0.49	4	0.38	402	-0.63	201	-0.86	208	-1.13	206	-1.16	202	-1.20	202	-1.20
508	3	405	8	307		405	N	1	0.51	608	0.43	203	-0.63	611	-0.86	1	-1.10	402	-1.12	1	-1.17	3	-1.12
602	3	5	8	308		406	N	303	0.53	403	0.45	304	-0.59	306	-0.78	608	-1.09	308	-1.06	505	-1.16	6	-1.08

Table 6 continued

604	3	**502**	8	**5**		**407**	N	**403**	0.54	**601**	0.48	**505**	-0.58	**503**	-0.77	**507**	-1.03	**203**	-0.96	**503**	-1.11	**405**	-0.99
606	3	**505**	8	**6**		**408**	N	**4**	0.57	**402**	0.48	**305**	-0.50	**303**	-0.73	**203**	-0.96	**406**	-0.91	**305**	-1.09	**604**	-0.93
2	2	**508**	8	**203**		**501**	G	**507**	0.60	**303**	0.53	**5**	-0.49	**401**	-0.65	**503**	-0.93	**306**	-0.79	**6**	-1.08	**1**	-0.91
301	2	**602**	8	**502**		**502**	N	**302**	0.64	**507**	0.53	**206**	-0.47	**406**	-0.59	**206**	-0.85	**503**	-0.70	**405**	-0.99	**204**	-0.89
305	2	**604**	8	**503**		**503**	N	**201**	0.68	**603**	0.56	**507**	-0.41	**402**	-0.58	**508**	-0.85	**611**	-0.61	**204**	-0.95	**506**	-0.88
206	2	**606**	8	**504**		**504**	N	**5**	0.70	**302**	0.57	**504**	-0.41	**610**	-0.56	**305**	-0.82	**507**	-0.49	**604**	-0.93	**505**	-0.80
5	2	**2**	9	**505**		**505**	N	**307**	0.72	**406**	0.63	**506**	-0.41	**507**	-0.49	**406**	-0.71	**408**	-0.45	**506**	-0.88	**2**	-0.79
502	2	**305**	9	**507**		**506**	N	**404**	0.73	**201**	0.66	**604**	-0.39	**608**	-0.47	**404**	-0.57	**608**	-0.38	**201**	-0.88	**404**	-0.77
505	2	**206**	9	**508**		**507**	N	**305**	0.74	**401**	0.70	**308**	-0.31	**505**	-0.29	**408**	-0.35	**604**	-0.35	**404**	-0.77	**0**	-0.60
507	2	**507**	9	**601**		**508**	N	**308**	0.76	**504**	0.70	**603**	-0.23	**404**	-0.25	**506**	-0.34	**404**	-0.20	**608**	-0.67	**402**	-0.52
3	1	**1**	10	**603**		**601**	N	**203**	0.77	**604**	0.75	**404**	-0.21	**602**	-0.21	**604**	-0.33	**208**	-0.13	**301**	-0.65	**503**	-0.48
304	1	**3**	10	**605**		**602**	N	**304**	0.82	**506**	0.76	**602**	-0.17	**604**	-0.18	**504**	-0.28	**304**	0.10	**0**	-0.60	**301**	-0.26
406	1	**304**	10	**607**		**603**	N	**406**	0.85	**203**	0.76	**306**	-0.16	**603**	-0.14	**611**	-0.27	**609**	0.14	**308**	-0.59	**403**	-0.23
302	1	**406**	10	**609**		**604**	N	**408**	0.85	**408**	0.88	**611**	-0.15	**609**	0.29	**306**	0.00	**504**	0.24	**402**	-0.52	**201**	-0.23
308	1	**302**	10	**611**		**605**	N	**609**	0.93	**609**	1.03	**406**	-0.10	**506**	0.34	**308**	0.02	**506**	0.38	**403**	-0.23	**4**	-0.23
611	1	**308**	10	**602**		**606**	N	**306**	0.94	**304**	1.18	**408**	-0.05	**601**	0.37	**609**	0.22	**610**	0.38	**4**	-0.23	**610**	-0.13
610	1	**611**	10	**604**		**607**	N	**610**	1.06	**611**	1.22	**609**	0.00	**304**	0.55	**603**	0.29	**601**	0.40	**610**	-0.13	**401**	-0.11
408	1	**610**	10	**606**		**608**	N	**611**	1.30	**208**	1.98	**208**	0.14	**504**	0.63	**602**	0.49	**603**	0.53	**401**	-0.11	**609**	0.19
1	0	**408**	10	**608**		**609**	N	**208**	1.80	**606**	2.31	**610**	0.42	**208**	0.81	**610**	0.82	**602**	0.60	**609**	0.19	**303**	0.25
306	0	**306**	11	**610**		**610**	N	**606**	1.96	**607**	2.31	**606**	2.91	**606**	2.65	**606**	4.97	**606**	4.96	**303**	0.25	**203**	0.49
307	0	**307**	11	**408**		**611**	N	**607**	2.04	**404**	5.48	**607**	3.08	**607**	3.03	**607**	5.71	**607**	5.92	**203**	0.49	**608**	1.65

data and the close proximity of the compounds in energy, only general ideas and trends could be derived from the data.

Autodock was capable of finding the correct docking orientation of the glycolipid analog most of the time as the first docked conformation, however, there were instances where only in the second cluster was the correct binding mode located which was only sometimes in the top 10 docked energies. The assumption was made that the correct docked mode would have the 2´- and 3´- OH pointing at the Asp151 residue. It could be possible that the flexibility inherent in the glycosidic linkage between the sugar and lipid is flexible enough to permit rotation thereby allowing the 6´- OH to bind to Asp151 which was usually the case for the incorrect docked orientations. Why experimentally that orientation is not preferred remains to be seen. This issue was taken into consideration and analyzed wherein those glycolipids that did not yield correct docked modes as the 1st conformation were counted by YESes (1st conformation is acceptable) and NOs (an alternative docked conformation was needed). There were even instances wherein the glycolipids were incapable of being docked to the tertiary complex not necessarily due to bulkiness and these were given a NULL count. As a hypothetical analysis, it was assumed that good glycolipids for the system will have more YESes than NOs also without NULLs which was tallied as an OVERALL impression of the glycolipid.

Since it appears that only general trends can be deduced from the amount of docked data, the ability of Autodock to dock glycolipids correctly was used as a starting point. On average bulkier substituents were found to have more difficulty in binding to the tertiary structure as can be seen with the amount of NOs for the PGW50# and PGW60#

series. There was no trend concerning the PGW20#, PGW30#, and PGW40# where in the case of the 2′- and 4′- substitutions, the bulky groups were favored whereas with the small 3′- substitutions were preferred. Furthermore, when it came to observing the experimentally-based glycolipids, α-GlcCer (PGW001) and the 6′-deoxy-α-GalCer (PGW004) were both considered as NOs whereas they have been found to elicit an iNKT response. When looking at the OVERALL analysis for CD1d binding it appears that most glycolipids were able to be docked correctly into the CD1d binding pocket. The analysis for the TCR interacting energies did of course show mostly NOs but that is because it combines the energies from the tertiary complex binding and CD1d binding making it more prone to have differences. Therefore, it was concluded that even though this was a nice analytical tool for attempting to distinguish the glycolipids it was only capable of showing that docking to CD1d was easier rather than the tertiary complex.

The next step was to analyze actual docked energies and draw trends out of their rankings, however, since it was found to be quite random the only point to be drawn has to do with the tendency to solely dock to crystal structures. It was found in all three docked energy tables that upon calculating the average docked energy associated with docking to static snapshots taken of the protein to mimic flexible residues the overall docked energy changes. The ramification of this is that it is no longer reasonable to solely dock ligands to a rigid crystal structure since the flexible nature of the residues play a role in recognizing the ligands and contribute to their binding.

The docked average energy of α-GalCer in the tertiary complex went from very poor to moderate and then back down to poor following the trend from crystal structure to 3ns, to 5 ns, and ending with all 10 ns (marked as PGW000 and in green). The CD1d docked

energy was on the other hand found to be relatively similar throughout for α-GalCer with it being a relatively poor binder to CD1d. However, the TCR interaction energy of α-GalCer was found to be amongst the highest throughout the simulation. Therefore, if an energetic correlation had to be associated with α-GalCer from this seemingly incomprehensible data, it would be that its substituents and orientation create for the most optimum interaction with TCR. However, applying this theory to the other glycolipids as a technique to predict iNKT stimulation by correlating docked energies was not as forthright.

If the docked energies of α-GalCer are excluded and only the overall trends of the other 49 glycolipids are analyzed then some interesting trends are able to be ascertained. For example, the 2′- , 3′-, and 4′-deoxy analogs were found to be poor binders to the tertiary complex along with CD1d protein. This same poor binding was also observed for the experimentally tested glycolipids (PGW001-006) in the tertiary complex, but they were found to be moderate binders to CD1d. Further generalizing, the poor binders to the tertiary complex appeared to be the PGW20#, PGW30#, and PGW40# series with the PGW50# and PGW60# being the better binders; this also held true for binding to CD1d. However, the TCR interacting energy was not as clean to generalize where the best glycolipids were a smattering of 2′- , 3′-, and 4′- bulky substitutions along with the β-linked glycolipids.

It was quite disappointing to find out in the end that the docked energies are not a reliable to use in order to predict glycolipid antigen potential. Furthermore, the docked orientations to the crystal structure along with all the docked orientations to each snapshot of the tertiary complex and CD1d also provided no clear support for the

differences in iNKT stimulatory abilities. However, the fact that the substitutions were able to perturb the sugar orientation in the binding pocket provided a basis to continue using computational tools to analyze the system.

MD SIMULATIONS ON CD1D/GLYCOLIPID/TCR COMPLEXES

Docking was found not to be a reliable predictor for glycolipid selectivity and iNKT stimulatory capability. Furthermore, the small 2′-, 3′-, and 4′- substituted analogs (H-, OMe-, NH_2-) were all docked with near identical orientations even though experimentally any of these modifications to the 2′- and 3′- positions led to complete loss of iNKT activity. Therefore, no structural explanation could be found to explain this drastic change. The docked results of the bulkier substituents did show a slight difference in the orientation of the analogs. The 3′- substitutions caused the sugar to rotate to accommodate the bulky substituent thereby losing all hydrogen bonding to the Asp151 residue. On the other hand, the 2′- substitutions were able to be accommodated in the small pocket beside Asp151 with the 3′-OH maintaining hydrogen bonding to Asp151. Lastly, the 4′- substitutions were capable of maintaining their hydrogen bonds to Asp151 which was to be expected, and were found to accommodate their substituents in the space underneath the CDR1α and CDR2α loops.

All of this evidence did seem to support the idea that hydrogen bonding to Asp151 was important, but if it were not for the small substituents, the glycolipid selectivity could be resigned to this simple observation. The fact that substitutions were able to perturb the sugar orientation forced us to question whether the substitutions would also be capable of perturbing the overall CD1d/TCR complex. Based on the lines of evidence that analogs

with small substitutions were able to bind with similar orientation as α-GalCer, it was hypothesized that it is not the orientation nor the binding energy of the glycolipid that causes TCR not to recognize the presented glycolipid, but that something else is disrupted within the overall complex. The only tool available that is capable of observing such a change is that of molecular dynamics simulations.

A few chosen docked CD1d/glycolipid/TCR tertiary complexes were therefore submitted to MD simulations where the proteins, which were held rigid in the docking runs, were allowed to relax and equilibrate to the presence of the new glycolipid. It was hypothesized that in the MD simulation the sugar modifications will cause the orientation of the glycolipid, TCR, CD1d, or the entire complex to change, therefore, the following analogs were chosen to perform 3 ns simulations on: β-GlcCer (PGW002) to determine the effect of the glycosidic linkage on the orientation of the complex: 2′-, 3′-, and 4′-NHAc-α-GalCer (PGW402-404); 2′-, 3′-, and 4′-OMe-α-GalCer (PGW302-304); and 2′-, 3′-, and 4′-HE-α-GalCer (PGW307, PGW308, PGW505) to determine the effect of substitutions on the galactose sugar; and PGW306, PGW508, and PGW408 to determine the effect of bulky aromatic substituents at the 2′- and 4′- position. The truncated CD1d/glycolipid/TCR complex in explicit solvent was chosen as it required less intensive computational resources and time, hence permitting more simulations. The docked orientations of the glycolipids into the crystal structure were used as the starting point for all the simulations with the 3 ns mark being chosen as an endpoint since both the full and truncated complexes appeared to behave similarly until this point.

It was hypothesized that any modification of the sugar will be immediately felt by the complex, and its structural motions will be directed at accommodating the glycolipid.

This was supported upon the analysis of the simulations' RMSD calculations where the 4′- analog complexes showed stable trajectories that fluctuated between 2.5-3.0 Å similar to the truncated TCR/α-GalCer/CD1d simulation, whereas the 2′- and 3′- analogs' trajectories were slightly more unstable and fluctuating between 3.0-3.5 Å. Also, the RMSDs of the 2′- and 3′- analogs appeared to be increasing versus stabilizing as did the 4′- trajectories, however, 3 ns is not a conclusive enough length of simulation to determine overall complex stability. Nevertheless, it may be extrapolated that the 2′- and 3′- analogs deny a stable tertiary structure to be formed upon TCR recognition of the CD1d binary complex, whereas, the 4′- analogs allow the creation of a seemingly stable tertiary structure.

As in the crystal structures of the β linked glycolipids, the simulated β-GlcCer adopted a similar perpendicular orientation projecting out of the cavity rather than the flatter orientation of α-GalCer (**Figure 28**). There were no differences in the hydrogen bond network of the binding footprint for either α-GalCer or β-GlcCer. The orientation of the β-glucose does, however, cause the hydrogen bond to the Gly96α to be lost, while simultaneously allowing the 6′-OH to form a new hydrogen bond with Ser76 on the α1 helix of CD1d. Furthermore, the CDR2α loop containing Phe51α shifted drastically away during the simulation of β-GlcCer from the Trp153 and the 4′- position, thereby opening up a space between it and the CDR1α loop that measures roughly ~8 Å in width from Ser30α to Phe51α. Many currently tested diglycosylceramides are derivatives of the α-GalCer and not β-GlcCer, and since there is no space for the second sugar at any position on the α- linked sugar, these are only tolerated upon being truncated by a glycosidase back to the α-GalCer. However, the opening at the 4′- position of β-GlcCer provides

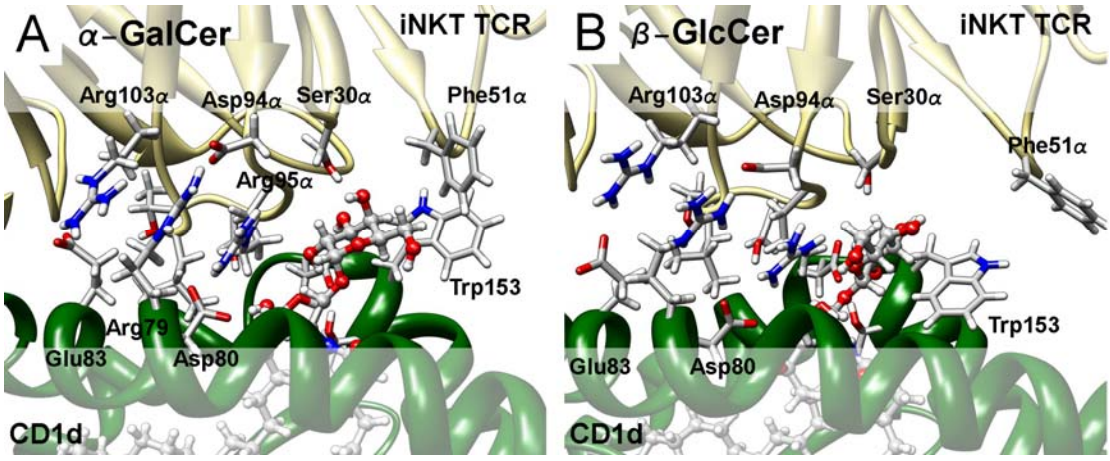


Figure 28. The orientations of (**A**) α-galactose and (**B**) β-glucose glycolipids after 3 ns of simulation. α-GalCer was found to maintain its orientation as it exists in the crystal structure, whereas, β-GlcCer was found to adopt a more perpendicular orientation that does cause the CDR2α loop to shift away from CD1d and Trp153.

some evidence for why iGb3, a trisaccharide, has been found to be a viable antigen for the system. Yet, the disaccharide Galβ1-4Glcβ1-1′Cer was found to be incapable of stimulating $V_{\alpha}14$ NKT cells showing that the disaccharide even though it could fit into this opening probably does not form necessary electrostatic interactions to ensure TCR stimulation, which is why the third sugar may be necessary.

The 3 ns MD simulations of the 2′-, 3′-, and 4′- analogs' tertiary complexes showed that many events occurring simultaneously more than likely contribute to the lack of iNKT cell stimulation for the 2′- and 3′- substituted glycolipids (**Figure 29**). The most obvious difference between the substitutions was the orientation of the sugar and its hindered ability to form the necessary hydrogen bonds within its binding pocket. Only the 4′- analogs maintained all of the necessary hydrogen bonds that exist between the 2′- and 3′-OH and TCR, thereby allowing it to also maintain a similar orientation as α-GalCer inside the binding pocket. This evidence coincides with what was found during

the docking of the glycolipids. Unlike in the docking results though where the 2′-OH was found not reorient itself to hydrogen bond with Asp151 upon loss at the 3′- position, the simulations did show the 2′- and 3′- OME and HE analogs' sugars rotating slightly allowing the 2′-OH to compensate for the lack of hydrogen bond donating of the 3′-substituent to Asp151, and, of course, vice versa.

In the 2′- and 3′-NHAc analogs, the sugar orientation was found to be more drastically affected, where the 2′-NHAc analog adopted an orientation reminiscent of β glycolipids. The 2′-NHAc substituent was unable to be accommodated into the small pocket beside the Asp151whereas the 2′-HE substituent was accommodated. This caused the sugar to rotate back along with the CDR3α loop to shift up and out of the hydrophobic pocket of CD1d, but overall the sugar remained in the pocket. The 3′-NHAc analog was not as tolerated whereupon none of the necessary hydrogen bonds to Gly96α, Asp151, nor Thr154 were formed because the sugar was being ejected from the binding pocket. The sugar was found not to rotate to displace its 3′- substitution into the small pocket beside the Asp151 as happened with the docked compounds. Therefore, whatever space thesmall pocket provided, it was not preferred as a viable place for any of the 3′- substitutions with all of them seeming to prefer the 2′-OH hydrogen bond to Asp151 and shifting the 3′- substitution up and over α2 helix of CD1d. The drastic behaviors of the 2′- and 3′-NHAc were in stark contrast to the 4′-NHAc, which behaved like the other 4′- analogs, maintaining a similar orientation as well as the hydrogen bonding of α-GalCer.

In conjunction with the perturbation of the sugars' orientations, the more important consequence of binding the analogs was that the hydrogen bonding network of the binding footprint was disrupted. Specifically, the hydrogen bonds between Glu83, Lys86,

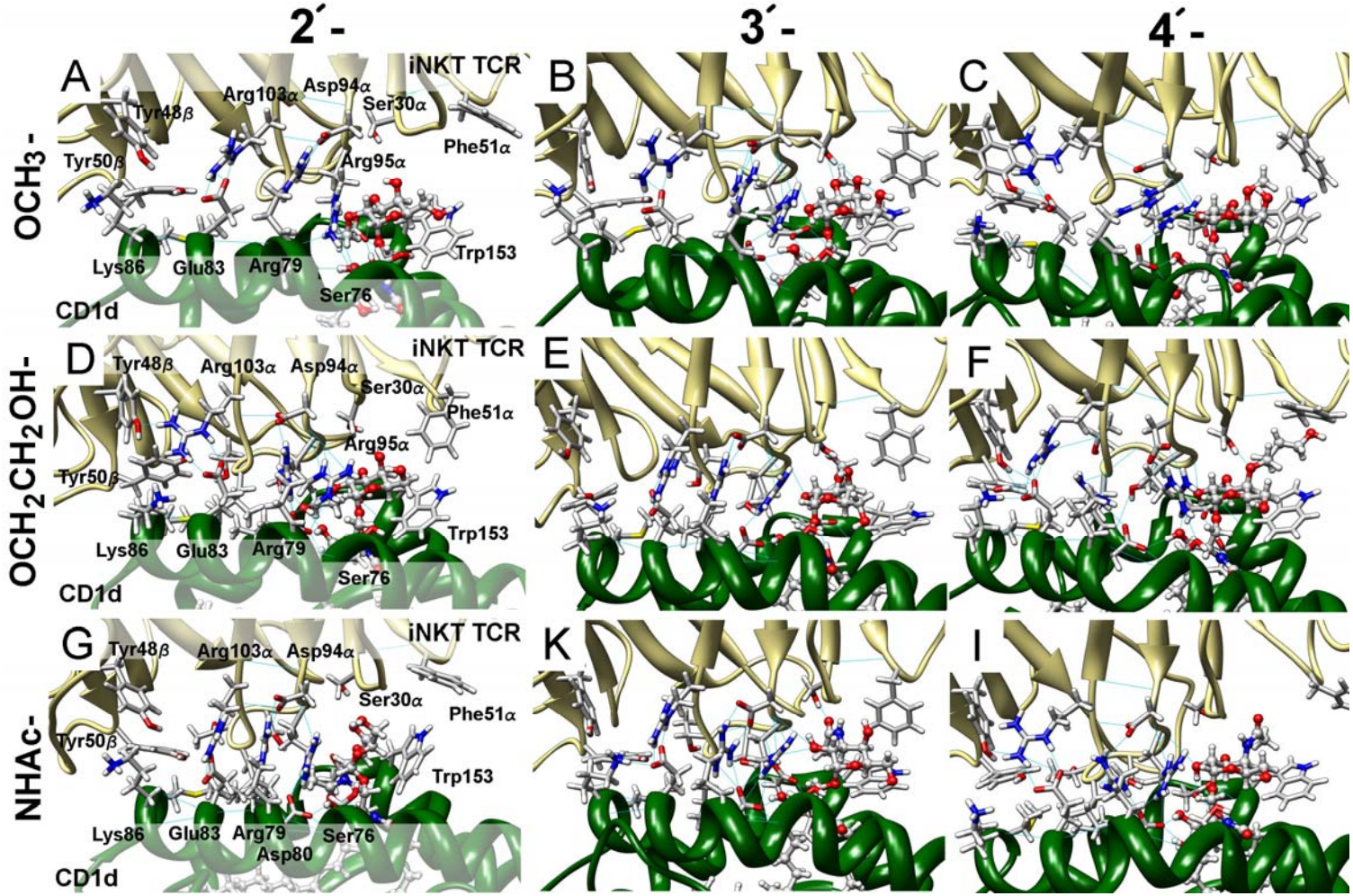


Figure 29. Comparison of the 2´-, 3´-, and 4´- analogs post MD. (**A-C**) 2´-, 3´-, and 4´-OMe-α-GalCer; (**D-F**) 2´-, 3´-, and 4´-HE-α-GalCer; and (**G-I**) 2´-, 3´-, and 4´-NHAc-α-GalCer. The sugar orientations were relatively unaffected by substitutions except with the NHAc- substitutions. The disruption of the tyrosine hydrogen bonds by the 2´- and 3´- substitutions appears to differentiate them from the 4´- analogs implying that the disruption of the binding footprint is the cause for glycolipid selectivity.

Tyr48β, and Tyr50β were disrupted for all the 2′- and 3′- analogs. The arginine stack that was observed in the original simulations also seems to exist in each of the simulations on the glycolipid analogs with Arg103α swinging over and binding to Glu83, however, only in the case of 4′- analogs were the tyrosines and lysine able to maintain their hydrogen bonds to Glu83. As mentioned earlier, when either tyrosine was mutated into an alanine the binding affinity between TCR and CD1d dropped significantly, so that if either tyrosine cannot form hydrogen bonds to Glu83 then TCR cannot bind to CD1d. In all the simulations, the other hydrogen bonds between Asp94α, Arg95α, Arg79, Asp80, and the hydroxyl groups on the sphingosine chain were maintained. Even if all the other non-covalent interactions were maintained within the binding footprint, the loss of both tyrosine hydrogen bonds will have a negative consequence on the TCR affinity to the CD1d presented glycolipid.

The reason for the disruption of the tyrosine binding is speculated to be that the CDR3α loop was being not buried far enough into the hydrophobic binding cavity of CD1d. This was more apparent in the 2′- analogs rather than the 3′- analogs as was previously mentioned. As it turned out with the increase in bulkiness of the 3′- substituent, the propensity for the analog to block the CDR3α loop from being buried in the hydrophobic groove was possible as was the case with the 3′-NIv-α-GalCer (PGW306) (**Figure 30**). Interestingly, the 4′- aliphatic amide substituent, PGW508, maintained the tyrosine hydrogen bonding network, however, its bulkier counterpart the 4′- aromatic amide substituent, PGW408, did not implying that probably this glycolipid would not be an ideal candidate for iNKT stimulation.

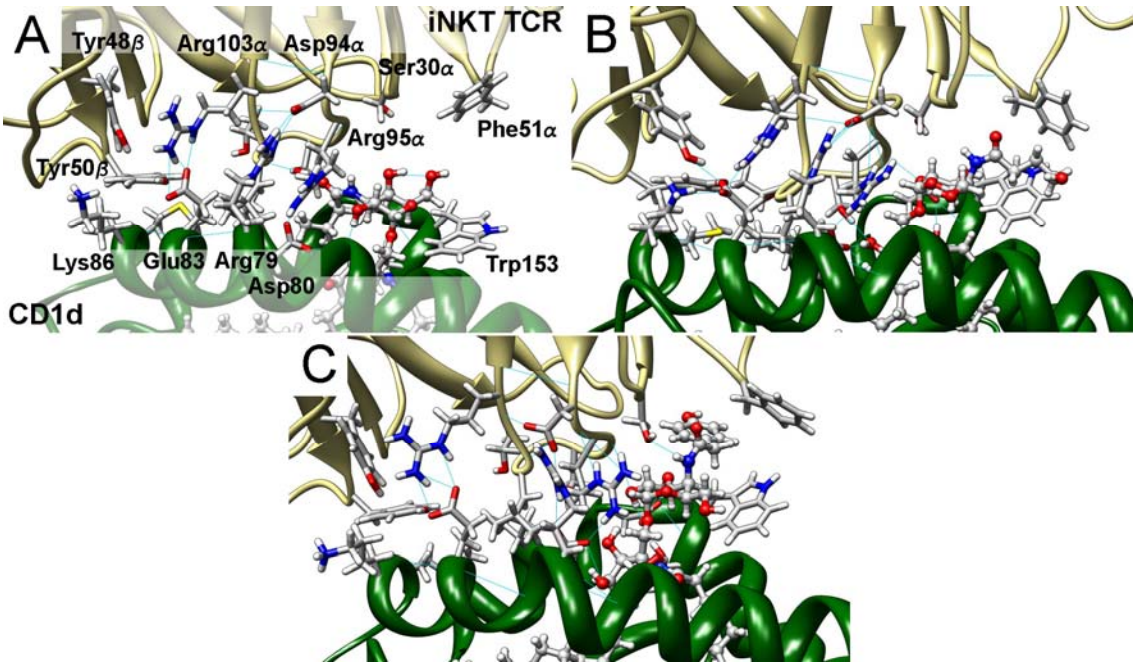


Figure 30. Comparison of the 3′- and 4′- bulky substituents after MD. (**A**) The extremely bulky 3′-NIv (PGW306) substituent disrupted the tyrosine hydrogen bonding by forcing the CDR3α loop out of the binding pocket along with totally rotating the sugar away from Asp151. (**B**) PGW508 was able to maintain the binding footprint between CD1d and TCR, (**C**) whereas, a bulkier substitution, PGW408, was not able to maintain it causing the tyrosines to lose bonding to Glu83.

The MD simulations were able to expound upon the structural observations made from the docking results by truly showing what is the difference between good and poor glycolipids for the system. The simulations confirmed that depending upon the substituents placed at the 2′-, 3′-, and 4′- positions on the galactose sugar, it would cause the rotation of the sugar away from the Asp151. Until now, there has been no evidence to delineate why a lack of hydrogen bond such as with the 2′-deoxy or the addition of another hydrogen bond donor sucha s 2′-NH_2 were able to totally destroy iNKT stimulation activity. If the motion to or away from Asp151 by the OH position compensating for where the substitution was made causes a shift in the overall sugar orientation, then the sugar orientation will affect the binding footprint by blocking the

CDR3α loop from being able to sit within the hydrophobic binding pocket of CD1d thereby causing the tyrosines from being able to form hydrogen bonds with Glu83. In medicinal chemistry, modifications to a ligand usually cause direct problems between the ligand and its binding protein structure, however, in this case the structure-activity relationship showed that bad modifications would cause a disruption away from the location of the modification. For the CD1d/TCR tertiary complex, glycolipid modifications will directly disrupt the binding footprint between the two proteins, which will in turn lower the binding kinetics between them not allowing for TCR to reach the binding threshold necessary to elicit an iNKT response.

CHAPTER 4:

SIMULATION OF THE CD1d/GLYCOLIPID BINARY COMPLEXES

HYPOTHESIS OF CD1d SELECTIVITY IN GLYCOLIPIDS

The search for understanding the extreme glycolipid selectivity of the CD1d/TCR tertiary complex within the immune system by computational means has been able to provide us with a reason that experimental data has not been capable of doing in the last 15 years. Nevertheless, experiment has shown us that the CD1d and iNKT TCR are relatively rigid proteins even upon complexation and possess a very minimal binding footprint wherein only a half of each available surface interacts with the other's. Furthermore, it has been accepted that the binding kinetics of this interaction drive iNKT response, that modifications within the glycolipid can create a biased cytokine release, and that the sugar is extremely intolerant to modification. The CD1d/TCR interaction differs extremely from that of the pMHC/TCR interaction wherein the latter is capable of an induced fit interaction showing the plasticity of the binding footprint along with the flexibility within the peptide ligand presented. The CD1d/TCR interaction does not possess these qualities of this relative system wherein the interaction is considered a lock-and-key type. The computational data have added to this picture provided by experiment wherein substitutions on the sugar are capable of changing the orientation of

the sugar which in turn displaces the CDR3α loop causing the tyrosines, Tyr48β and Tyr50β, to lose hydrogen bonding with Glu83 at the far end of the binding footprint away from the sugar. The ability of minimal glycolipid changes to disrupt the binding footprint of CD1d to TCR adds additional validity to the lock-and-key interaction hypothesis where only a certain orientation of the glycolipid causes the right orientation of CD1d to TCR to be achieved.

Assuming that this immune system process follows the path where a foreign lipid displaces a self-lipid bound to CD1d and CD1d then presents the foreign lipid to the TCR proteins of iNKT cells, it must be the case that the glycolipid is capable of disturbing the CD1d protein before the TCR interacts with it. In the immune system, iNKT cells scan the surface of dendritic cells in search of CD1d presented foreign antigens which is the sign of an infected cell. Concurrently, though, healthy cells also possess CD1d proteins presenting self-lipids which also interact with TCR proteins but cause a very diminished iNKT response. If TCR needs to bind to CD1d presenting a ligand to determine whether it is a foreign or self-lipid and either binding event causes a vast or minimal iNKT response, then those ligands that are incapable of causing any iNKT response, synthetically modified or biological-derived ligands, must bar even the slightest of TCR binding to the CD1d protein.

When considering the mutation studies performed on the key hydrogen bonding residues in the binding footprint between CD1d and TCR, if a single mutation of these residues to alanine caused a drastic loss in binding affinity and thereby extreme loss in the iNKT response then a combination of mutations should in essence completely negate binding and in turn cause complete loss of the iNKT response. This sensitivity to singular

mutations on either the TCR or the CD1d shows that simple modifications on the glycolipid should also have negative consequences on the interaction. The removing or even adding a hydrogen bond donor on a sugar wherein two positions are bonded to the same residue, Asp151, was able to cause a complete loss of binding provides evidence that this system is truly structurally very specific. This specificity is due to simple glycolipid modifications being capable of disrupting the binding footprint and in more than one location. Therefore, since it is hypothesized that non-ligands for the system bar TCR binding, then they must be able to affect the CD1d proteins residues that interact with TCR, specifically with the CDR3α and CDR2β loop.

MOLECULAR DYNAMICS PARAMETERS

The starting point for determining whether or not glycolipids are capable of changing the CD1d binding surface was performing an analysis of the currently existing crystal structures of CD1d binding a variety of glycolipids. However, it appeared that the fitting parameters for these crystal structures which were all solved between 2-3 Å were quite similar making all of the CD1d residues possess the same relative orientation (**Figure 31**).

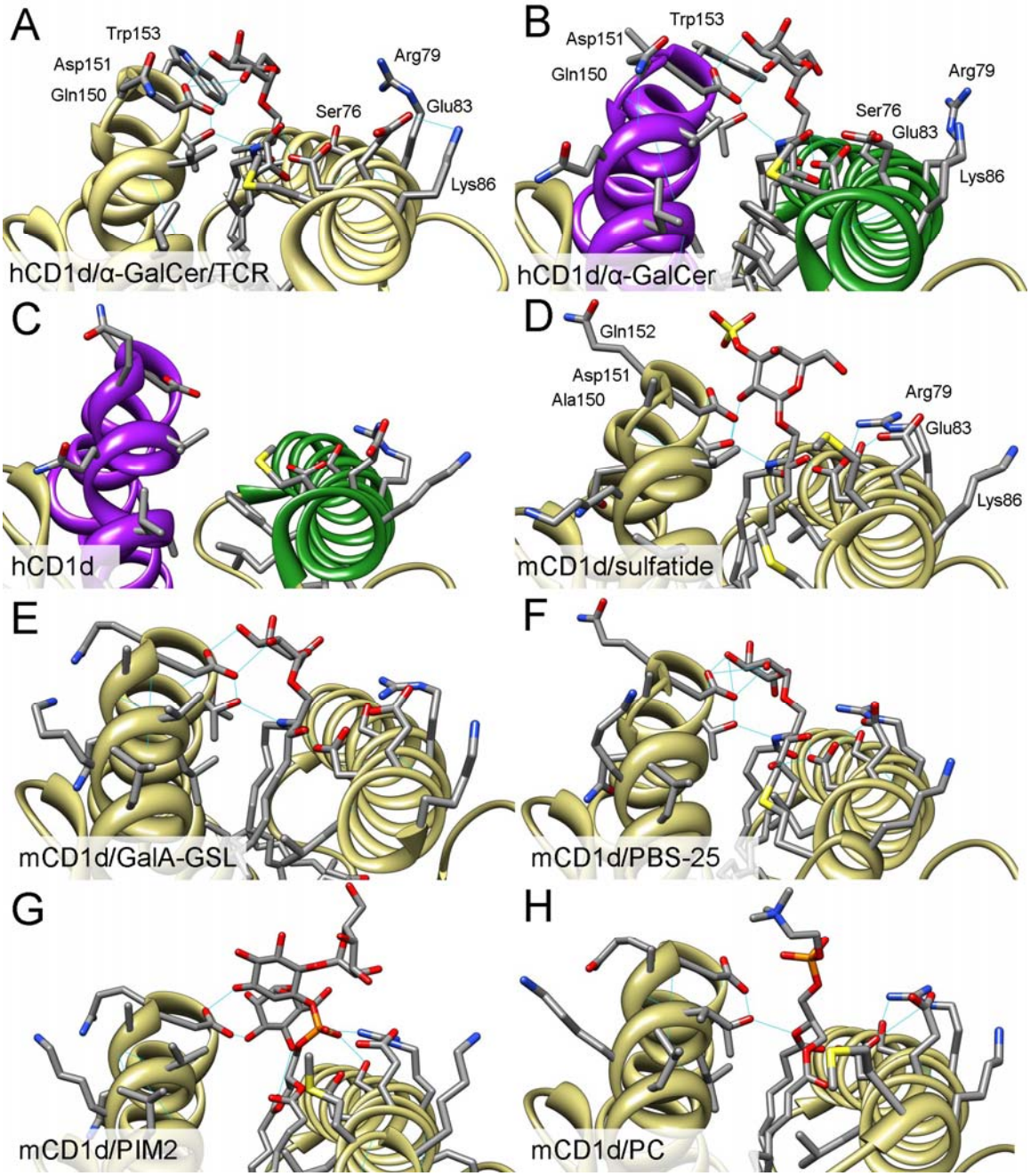


Figure 31. View down hydrophobic binding groove of CD1d/glycolipid crystal structures. The important TCR interacting residues are labeled for both mouse and human CD1d proteins. (A) The hCD1d/α-GalCer binary complex with TCR removed. [*2PO6*] (B) The original crystallized hCD1d/α-GalCer binary complex. [*1ZT4*] (C) The crystallized hCD1d protein without ligand. [*1ZT4*] (D) The mCD1d/sulfatide binary complex. [*2AKR*] (E) The mCD1d/ GalA-GSL binary complex. [*2FIK*] (F) The mCD1d/PBS-25 binary complex. [*1ZTL*] (G) The mCD1d/ PIM2 binary complex. [*2GAZ*] (H) The mCD1d/PC self-lipid binary complex. [*1ZHN*] Both residues for hCD1d proteins are in similar orientation as are all the mCD1d in relation to each other. The only protein showing any displacement in residues is hCD1d without bound ligand. *PDB ID # are in brackets.*

Neither the sulfatide nor PIM2 elicit an iNKT response making them non-ligands and PC is a self-lipid, therefore, it did not bode well that their protein structures as well as the bound ligand were oriented much like α-Galcer, GalA-GSL, and PBS-25. The hCD1d without a bound ligand showed two residues Gln150 and Ly86 being slightly displaced with Gln150 pointing vertically up and Lys86 point laterally out of CD1d. It remains to be seen what information this crystal structure can provide considering that CD1d always has a bound lipid. The alignment of mCD1d to hCD1d showed that only the residues interacting with TCR on the α1 helix were conserved between the species (**Figure 32**). Even though both Asp151 and Thr154 which are known to form direct hydrogen bonds to the glycolipid were conserved between the two species none of the surrounding residues on the α2 helix kink were found to be. The Gln150 residue which upon mutation to alanine showed a decreased iNKT response is located at position 152 for mCD1d whereas the mCD1d possesses an alanine at its 150 location. If both mouse and human CD1d proteins can present α-GalCer to iNKT cells and elicit a strong response, the Gln150/Ala mutation on hCD1d should have had no effect.

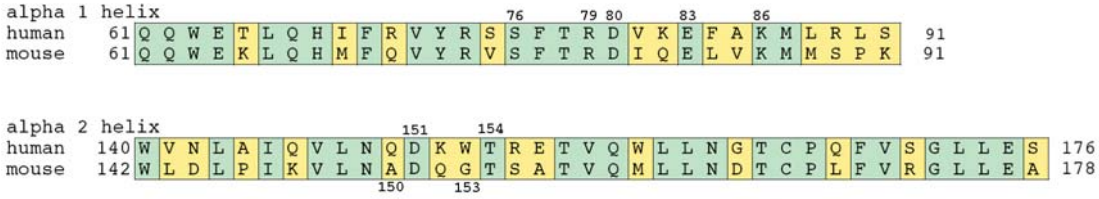


Figure 32. Alignment of mouse and human CD1d alpha helices. The conserved residues are shown in green whereas yellow colored are not conserved. The important TCR contacting residues have been specifically numbered along with the positions 150 and 153.

Molecular dynamics simulations were turned to once again as the tool to be used to gauge whether or not the crystal structure orientations of CD1d were a valid means in qualifying ligand presentation. The simulation protocol as was used for the truncated complex simulations was applied with an explicit water solvation environment applying the NVE ensemble and PBC conditions.

The same question for the sake of efficiency was investigated that concerned the truncated complexes. In order to determine whether or not the β_2M sheet along with the adjoining transmembrane β sheet were necessary to observe the glycolipid behavior in the hydrophobic binding pocket created by the two α helices of CD1d, the full CD1d protein was submitted to a full 10 ns long simulation. The truncation of the CD1d protein occurred at the end of the α2 helix where a long loop connected the αβ superdomain to the β sheet. Two things needed to be true in order to use the truncated CD1d in the rest of the binary complex simulations: (1) the full CD1d protein had to be quite stable for the entire duration of the 10 ns simulation, and (2) the truncated protein had to behave similarly to the CD1d protein as far as the 3 ns mark which was found to be a good length for the truncated complex simulations and it could then be extrapolated it would behave much like the full CD1d if it were simulated out to the 10 ns mark.

The full CD1d protein was found to be quite stable throughout the 10 ns simulation. Snapshots taken at the 0.5 ns, 3.0 ns, and 10.0 ns mark were aligned and the RMSD deviations were calculated to both visually and statistically observe the fluctuations of the

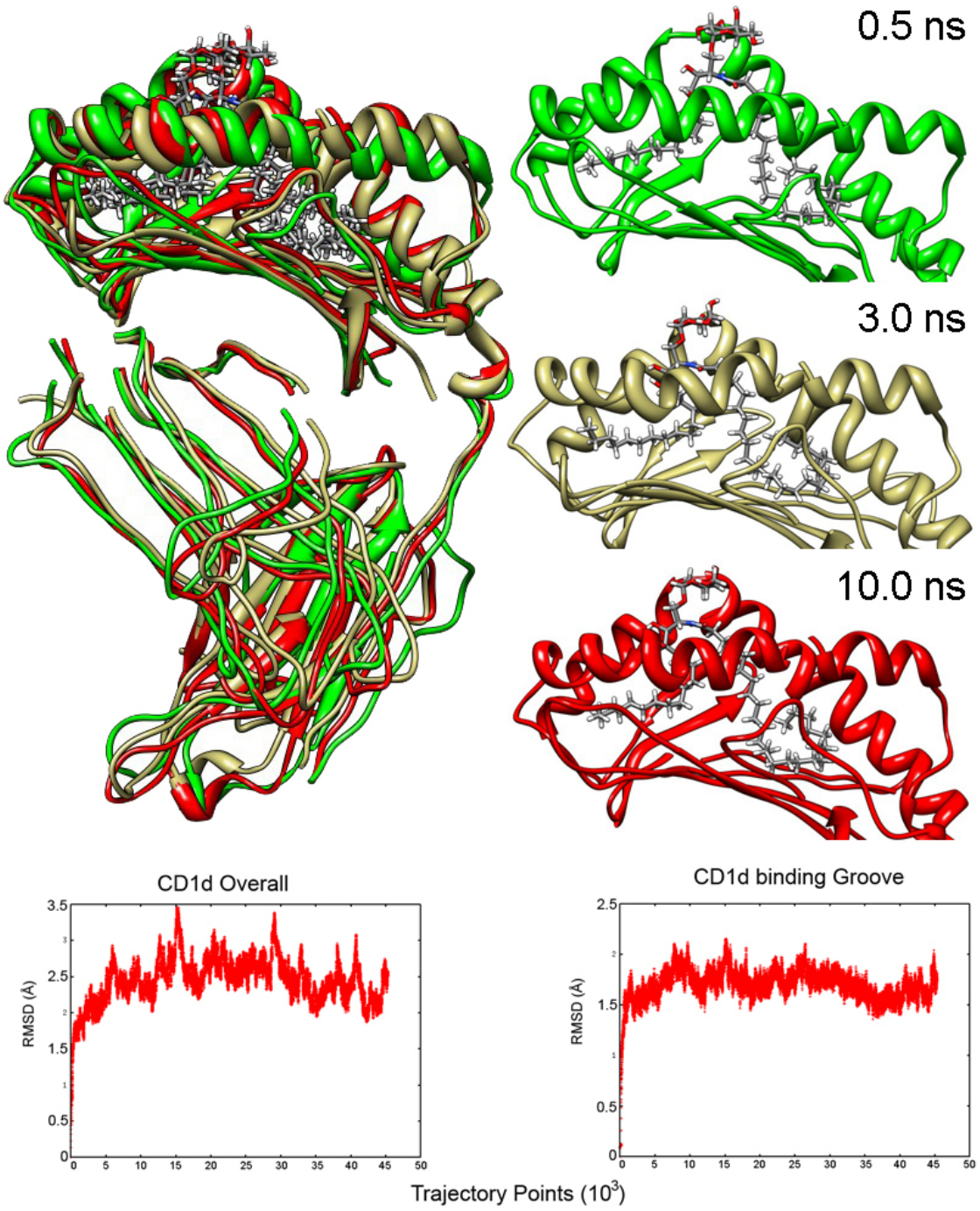


Figure 33. Alignment and RMSD plots of the full CD1d protein simulation. The protein fluctuates very minimally and has a similar RMSD profile in both the overall fluctuation as well as the binding groove to the tertiary complex simulation. The snapshots that were aligned were at the 0.5 ns mark (green), 3.0 ns (khaki), and the 10.0 ns mark (red).

protein (**Figure 33**). Neither α helix was found to fluctuate during the simulation and α-GalCer was able to maintain all its hydrogen bonds along with its overall orientation throughout the simulation. However, in this simulation the galactose appeared to right itself without the TCR present to adopt a more head on orientation towards Asp151 with both the 2′- and 3′- OH groups hydrogen bonding to it. Since the hydrogen bonds between the 3 and 4-OH on the sphingosine chain to Asp80 are the most durable, this shift of the sugar was compensated for by having the bottom of the sphingosine chain curl more towards the acyl chain rather than disrupting the hydrogen bonds to Asp80.

The RMSD deviation for the overall complex fluctuated around 2.5 Å whereas the binding groove fluctuated around 1.75 Å. These resemble almost exactly the RMSD plots for CD1d in the tertiary complex where it had an overall fluctuation around 2.5 Å and 1.75 Å for its binding groove. The fact that the CD1d protein fluctuates quite similarly regardless of whether or not TCR is bound to it confirms the lock-and-key theory of the interaction along with the inherent rigidity of the CD1d protein. Furthermore, since the majority of fluctuation arises in the β_2M sheet along with the adjoining β sheet, it is quite reasonable to ignore them and focus primarily on the binding groove of CD1d for the future simulations.

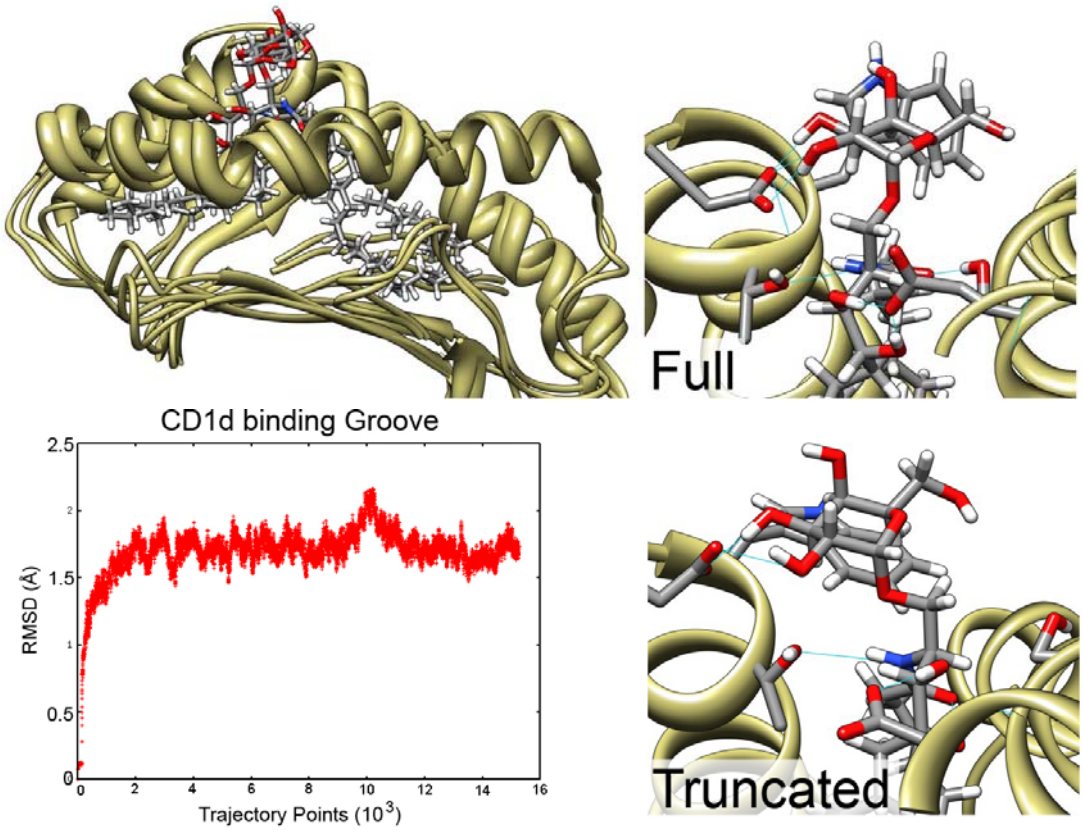


Figure 34. The alignment of the full and truncated CD1d proteins. The overall binding pocket remained similar with the RMSD plot showing exactly the same fluctuation as in the full complex. The glycosidic bond did though develop a different staggered conformation, but glycosidic bonds are known to be quite flexible. Regardless, all the hydrogen bonds were maintained by the glycolipid in the

The binding groove truncated CD1d simulation ran for 3 ns and its end point was aligned to the 3 ns mark for the full CD1d protein and found to be quite similar (**Figure 34**). The primary difference between the two concerned the glycosidic linkage where in the truncated complex it adopted a different staggered conformation as that existing in the full complex. Regardless, the α-GalCer in the truncated complex was able to maintain all the important hydrogen bonds to CD1d. Even more importantly, the RMSD plot for the binding groove showed exactly the same fluctuation around 1.75 Å as did the binding groove in the full CD1d protein. Since multiple glycolipids bound to CD1d would be

simulated, the fact that only the binding groove was necessary for observing glycolipid behavior provided a very fast, efficient, and computationally non-intensive method to be used.

MD SIMULATIONS OF CD1D/GLYCOLIPID COMPLEXES

The starting set of glycolipid analogs used to gage what protein motions occur upon binding to CD1d were those that underwent experimental testing along with a few that possessed crystal structures. The analogs contained simple structural differences compared to α-GalCer: (1) the 4′- position was equatorial [α-GlcCer]; (2) the glycosidic linkage was β- [β-GalCer and β-GlcCer]; (3) the 6′- position was a carboxylic acid subsituent [GSL]; (4) the glyolipid was of bacterial origin [GSL-1 and GalA-GSL]; (4) the sphingosine chain had its hydroxyl groups removed [3- or 4- deoxy and 3,4-deoxy]; (5) the acyl chain length was short [PGW-25]; (6) the sphingosine chain length was short (OCH); (7) the acyl chain contained unsaturation [C20:2]; (8) the glycosidic linkage was a methylene group [C-glycoside]; and (9) the sugar was replaced by a hydroxylated alkyl chain [threitolceramide] (**Figure 35**). According to experiment, only the following compounds from the list were incapable of eliciting an iNKT response: β-GalCer, β-GlcCer, 3-deoxy-α-GalCer, and 3,4-deoxy-α-GalCer. Biased cytokine profiles were achieved by PBS-25, OCH, and the C-glycoside with lessoned iNKT responses achieved by the GSL analogs, 4-deoxy-α-GalCer, C20:2, and threitolceramide.

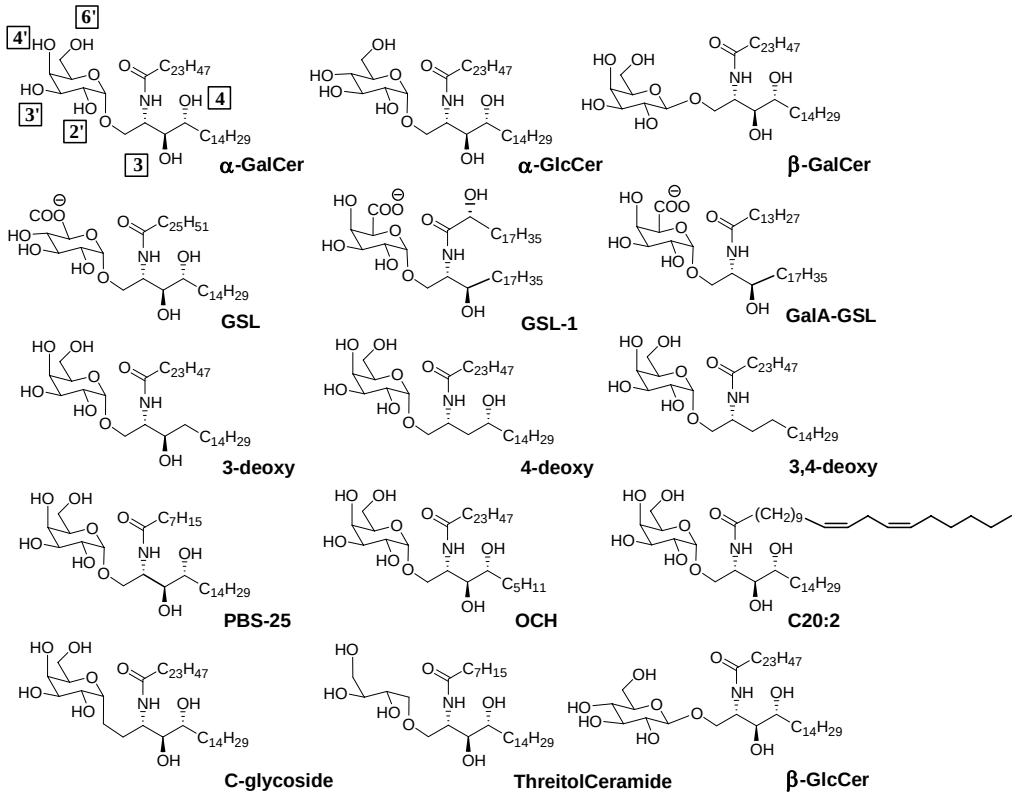


Figure 35. List of glycolipids that underwent MD simulation in CD1d. The list represented simple differences in experimentally tested compounds: the glycosidic linkage, 4´- or the 6´- position, the lipid chain lengths, or not having a sugar at all.

The glycosidic linkage was the first to be analyzed since humans rarely if ever produce α-linked glycolipids and the β-linked glycolipids are not capable of eliciting any iNKT activity (**Figure 36**). The residues that underwent observation were those that played a role in TCR or glycolipid binding: Gln150 and Asp151 on the α2 helix; and Ser76, Arg79, Asp80, Glu83, and Lys86 on the α1 helix. Both the full and truncated CD1d/α-GalCer binary complexes maintained all of the necessary hydrogen bonding between the glycolipid and CD1d. The TCR interacting residues also remained in an orientation similar to those found in tertiary crystal structure and the binary crystal structure with the Gln150 remaining in a horizontal position to the α2 helix, and Arg79, Glu83, and Lys86

all pointing vertically up from the α1 helix. The Met87 residue does not play a role in the hydrogen bonding of the binding footprint, however, its mutation to alanine was found to have a negative effect on TCR binding affinity. Since the hydrophobic binding cavity needs to be in a certain orientation and of a certain size in keeping with the lock-and-key theory, it is assumed that the Met87/Ala mutation caused the cavity to open at that end and thereby causing its disruption.

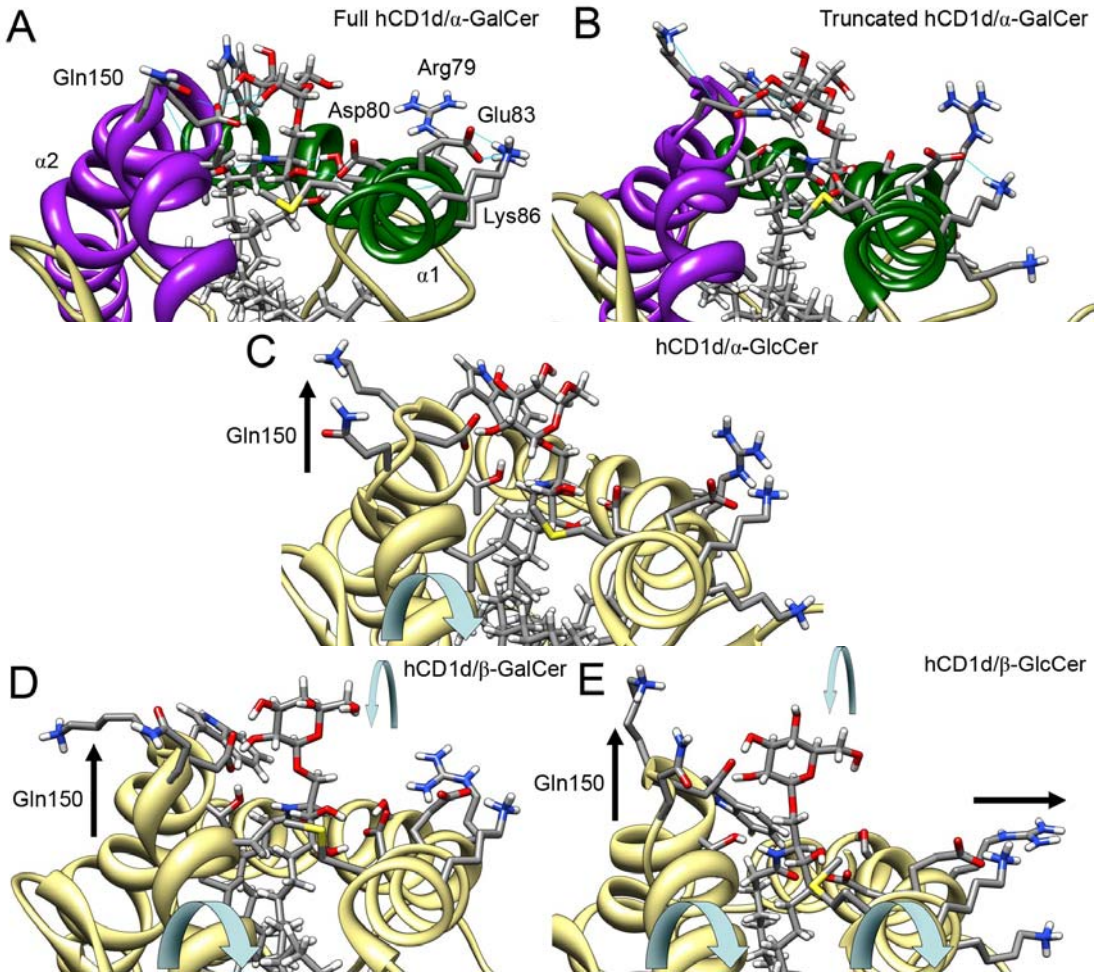


Figure 36. MD results of CD1d bound (**A** & **B**) α-GalCer, (**C**) α-GlcCer, (**D**) β-GalCer, and (**E**) β-GlcCer. The α-GalCer complexes maintained a similar orientation as their starting structures. The α-GlcCer complex only showed minimal variation but both β-linked glycolipids varied widely in structure after 3 ns of simulation.

The α-GlcCer analog is also able to stimulate iNKT cells albeit a little less than α-GalCer, and luckily there were a few structural deviations found in its binary complex. To begin with, it did a very unusual shift of its glucose making it resemble a β- linked glycolipid. It is quite possible that the equatorial orientation of the 4′-OH makes that end of the sugar charged and thereby causing unfavorable interactions with the Trp153 residue where tryptophan residues have been known to interact with the less polar underside of sugars as can be seen by its orientation in the truncated CD1d/α-GalCer complex. The Gln150 residue shifted away from its horizontal position to a more vertical position on the α2 helix. Since the Gln150/Ala mutation caused a marked loss of iNKT stimulation it is possible that this vertical orientation is capable of achieving two things simultaneously where it loses its hydrogen bond with the Thr98α on the CDR3α loop and then provides a physical barrier for the CDR3β loop when TCR is trying to bind to CD1d. On the α1 helix, the TCR interacting residues remain in an upright position.

The β-linked glycolipids are found to not stimulate iNKT cells at all and their CD1d protein structures were found to be the most changed. Both sugars were able to adopt the β-linked orientation with the sugar jutting vertically out of the CD1d binding pocket. Since TCR has to sit flat onto CD1d, it is possible that this extra structural hump blocks the CDR1α and CDR2α loops from coming down and forming the necessary interactions with the sugar. This new vertical position was way more drastic than for the α-GlcCer which still appeared to sit within the binding cavity. Even though the β-linked glycolipids were now vertical, they were still able to maintain both their 2′- and 3′-OH hydrogen bonds to Asp151. However, this caused the Asp151 to pull up and towards the 3′-OH position bringing with it the α2 helix and thereby pulling the Gln150 with it. This new

Gln150 orientation is also much more drastic than found with α-GlcCer where its backbone was in a same orientation as α-GalCer but now with the β-linked glycolipids the whole residue including backbone was vertically projected upward. The α1 helix between the two β-linked glycolipids were different wherein the β-GalCer was able to maintain the optimal vertical orientation for TCR binding, however, the residues for β-GlcCer were projected out laterally from the CD1d binding groove. It was observed that the new orientation of the 6′-OH was capable of occasionally forming a hydrogen bond with Ser76 thereby pulling it upward. This shift then screws the α1 helix which pulls the Arg79 residue outward along with but to a lesser degree the Asp80 and Lys86 residues. The orientation of the 6′-OH on β-GalCer was instead found to form an intramolecular hydrogen bond to the 4′-OH, however, it is quite possible that if the simulation is lengthened a similar result will occur as for β-GlcCer.

The results of the α-,β-Gal/Glc ceramides show a possible three things occurring within the CD1d protein upon binding a glycolipid. The orientation of the sugar is capable of changing and if it happens to extrude too far out of the groove, it will provide a steric hindrance to the CDR1α and CDR2α loops potentially also causing the necessary hydrogen bonds to not form between the sugar and TCR. In conjunction with its motion, the sugar is capable of bringing with it the Asp151 residue to which its hydrogen bonded. The motion of pulling Asp151 with the sugar is able to cause the α2 helix kink to shift thereby also changing the orientation of the Gln150 which is the only residue on the helix to form a direct hydrogen bond with the CDR3α loop. If the Gln150 residue shifts into a vertical orientation rather than lying horizontally across the helix then it no longer is in proximity to Thr98α and can potentially clash sterically with the CDR3β loop or even

with the CDR3α loop upon TCR binding. The residues on the α1 helix are also capable of changing their orientation upon glycolipid binding. Since the majority of the binding footprint occurs between this helix and TCR, any deviations here are potentially disastrous. If the lock-and-key hypothesis must hold true, then all of the residues on the α1 helix must be oriented in a vertical fashion to interact with their partner hydrogen bond residues on TCR. Overall, it would appear that if Gln150 is projecting vertically along with the sugar of the glycolipid in conjunction with the α1 helix residues projecting laterally out of the binding cavity than the CD1d protein is no longer in the conformation necessary for TCR to even be capable of binding with it. It remains to be seen whether or not all three have to occur or if one of these events is quite capable of destroying TCR recognition.

All three of the GSL analogs have been found to elicit iNKT activity albeit to a much lower degree than α-GalCer. The differences between GSL and GSL-1 are in their lipid portions wherein GSL has the α-GalCer lipid and GSL-1 is of the bacterial origin possessing only 3′-OH on its sphingosine chain and an additional hydroxyl on the acyl chain. The GalA-GSL glycolipid contains a shorter acyl chain than the α-GalCer lipid by 10 carbons. After the 3 ns of simulation, both GSL and GalA-GSL maintained similar CD1d protein conformations along with similar overall glycolipid orientations as was found for α-GalCer (**Figure 37**). However, the bacterial GSL-1 glycolipid was found to change its position within the binding pocket of CD1d. The entire glycolipid shifted downward into the groove while still maintaining its hydrogen bonds to Asp151 and forming a new hydrogen bond between its 6′- position and Ser76. The downward shift did cause the Asp80 hydrogen bond to the sphingosine chain to be lost, but the Thr154

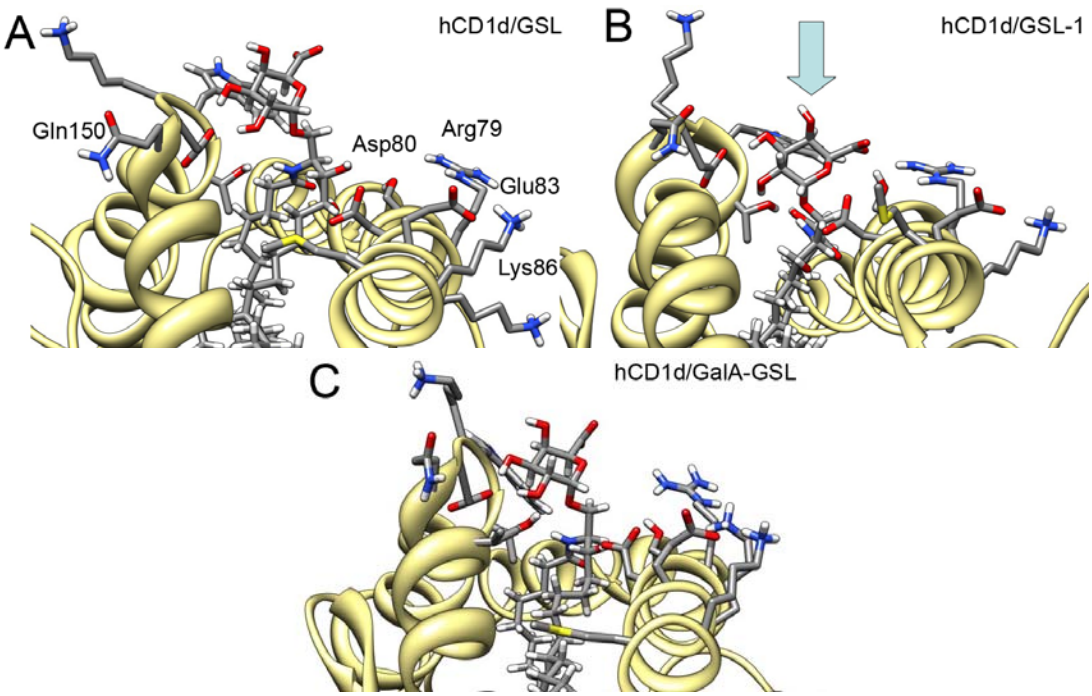


Figure 37. MD Results of the CD1d bound GSL analogs. (**A**) The hCD1d/GSL and (**C**) hCD1d/GalA-GSL complexes were both found to maintain very similar overall orientations as found in α-GalCer. (**B**) hCD1d/GSL-1 complex result showed that CD1d maintained similar residue orientations as well, however, the glycolipid shifted further downward into the binding groove making that the probable cause for its diminished iNKT response.

was able to form a new hydrogen bond with the hydroxyl on the acyl chain while maintaining its hydrogen bond to the nitrogen. This global downward motion of the glycolipid did not change the orientation of the CD1d residues which provides evidence for why TCR is capable of still binding the complex. However, the fact that the glycolipid is now further buried in the binding groove making potential contacts between the TCR and the sugar much farther away and therefore harder to make, probably the reason for the diminished iNKT response caused by GSL-1. So now the picture for the sugar becomes slightly more complicated in that either being jutted out too far (vertical orientation) causes loss of activity or being buried too far in the pocket will too.

When the hydroxyl groups on the sphingosine chain where removed individually (3-deoxy-α-GalCer or 4-deoxy-α-GalCer) or together (3,4-deoxy-α-GalCer), only the 4-deoxy-α-GalCer glycolipid was found capable of eliciting an iNKT response. However, the conclusion that the 3-OH is critical for stimulation was premature because the 3-deoxy analog has never been tested. So far in all the studies, the conclusion that the 3-OH was critical resulted from the observation that the 4-deoxy glycolipid was capable of eliciting a response but the 3,4-deoxy glycolipid was not. However, it should be possible that upon removing the 3-OH the 4-OH should be able to compensate with hydrogen bonding to the Asp80.

The MD results are a little more complex for the individual deoxy compounds but the 3,4-deoxy glycolipid definitely predict that TCR should not bind to its CD1d complex (**Figure 38**). The relative position of the sugar in the 3-deoxy glycolipid as compared to the 4-deoxy glycolipid is slightly higher. This is probably due to the 4-OH being shifted upward to form a better hydrogen bond with Asp80. Even though the starting orientation of the glycolipid is from the crystal structure of α-GalCer, it is quite possible that upon glycolipid insertion into the binding groove that it would remain jutted out since the 4-OH would be the only one forming a hydrogen bond with Asp80 and the ligand need not further insert to have the 3- position be in closer proximity. In the case of the 4-deoxy analog, this appears to hold true with the glycolipid being slightly more buried in the pocket as can be seen with the Asp151 pointing downward. Both of these events were to be expected since the deoxy analog need no longer to share hydrogen bonding with Asp80 but can rearrange to form a more direct hydrogen bond with it. In both deoxy analogs the Gln150 remained horizontal but rather than being in parallel to the α2 helix it

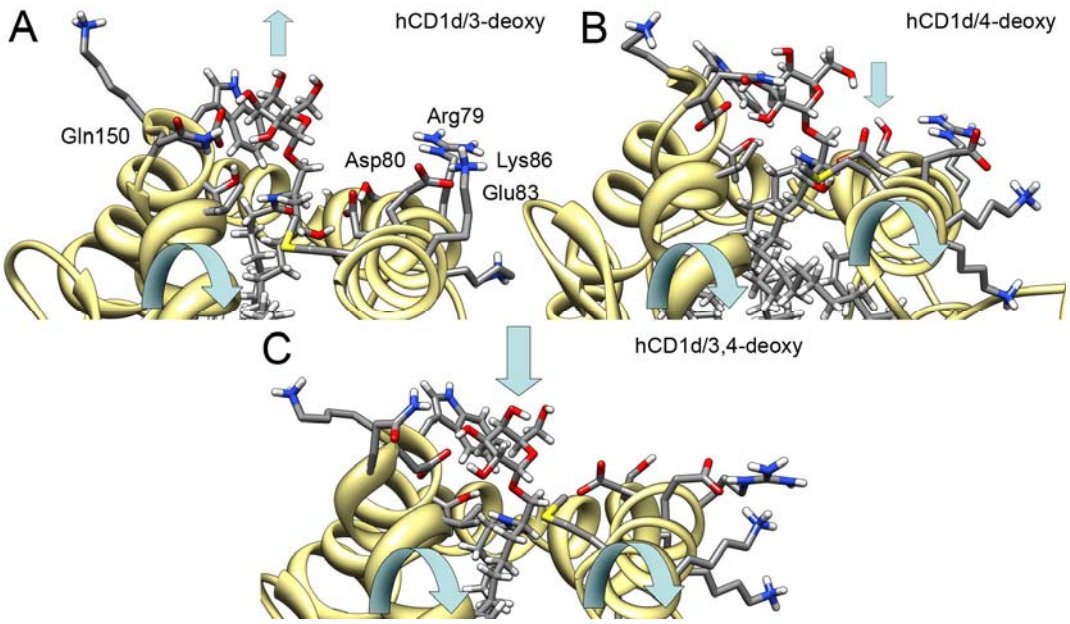


Figure 38. MD results of the CD1d bound deoxy sphingosine analogs. The 3-deoxy and 4-deoxy analog are very negligibly different with a slight shift occurring of the glycolipid due to the reorientation of the hydrogen bonds to Asp80. However, the 3,4-deoxy analog definitely caused the CD1d to adopt an undesirable orientation that is not conducive for TCR binding.

appeared to be oriented toward the sugar of the glycolipid thereby probably barring the CDR3α loop from being able to insert into the binding groove. Only in the case of the 4-deoxy analog were the residues on the α1 helix shifted slightly laterally. This shift was caused once again by the Ser76 being in close enough proximity to the 6′-OH of the sugar that it could form a hydrogen bond pulling it upward and with that screwing the α1 helix. Even though, there was a slight shift in the sugar head group and Gln150 was horizontal 90 degrees to where it should be oriented, it would appear that overall both CD1d proteins and glycolipids had similar orientations and therefore would be quite hard to differentiate. Therefore, to return to the experimental problem caused by the lack of experimental data on the 3-deoxy analog, the idea should become that either the 3-OH or the 4-OH on the sphingosine chain are critical for activity but not both. However, it has

been known that the 3,4-deoxy analog is not capable of eliciting iNKT activity. The MD results would predict that this would indeed be the expected outcome since the glycolipid was extremely buried in the binding groove, the positioning of the Gln150 was vertical and outward, and there was a drastic rotation laterally of the α1 helix. This overall CD1d orientation most certainly could not be recognized or even bound to by TCR.

When the more drastic lipid modifications underwent the 3 ns MD simulations then a huge difference in glycolipid freedom was observed between the mCD1d and hCD1d proteins. The PBS-25 crystal structure was chosen to be run in simulation to determine whether or not it would retain its original orientation like the hCD1d/α-GalCer. However, it turned out that the differences at the α2 helix kink where ^{150}Ala-Asp-Gln-Gly-Thr^{154} in the mCD1d versus the ^{150}Gln-Asp-Lys-Trp-Thr^{154} in the hCD1d created a more flexible space for the sugar (**Figure 39**). The galactose on PBS-25 was observed to rotate around forming a hydrogen bond between the 6′-OH and Asp151 while the 3′- and 4′-OH were still oriented upwards towards the TCR binding region. However, it appeared that the short acyl chain was slightly ejected from the binding groove causing the entire glycolipid to shift upward from the pocket. This caused the 3- and 4-OHs on the sphingosine chain to pull the Asp80 residue with them, but no lateral displacement of the residues on the α1 helix occurred.

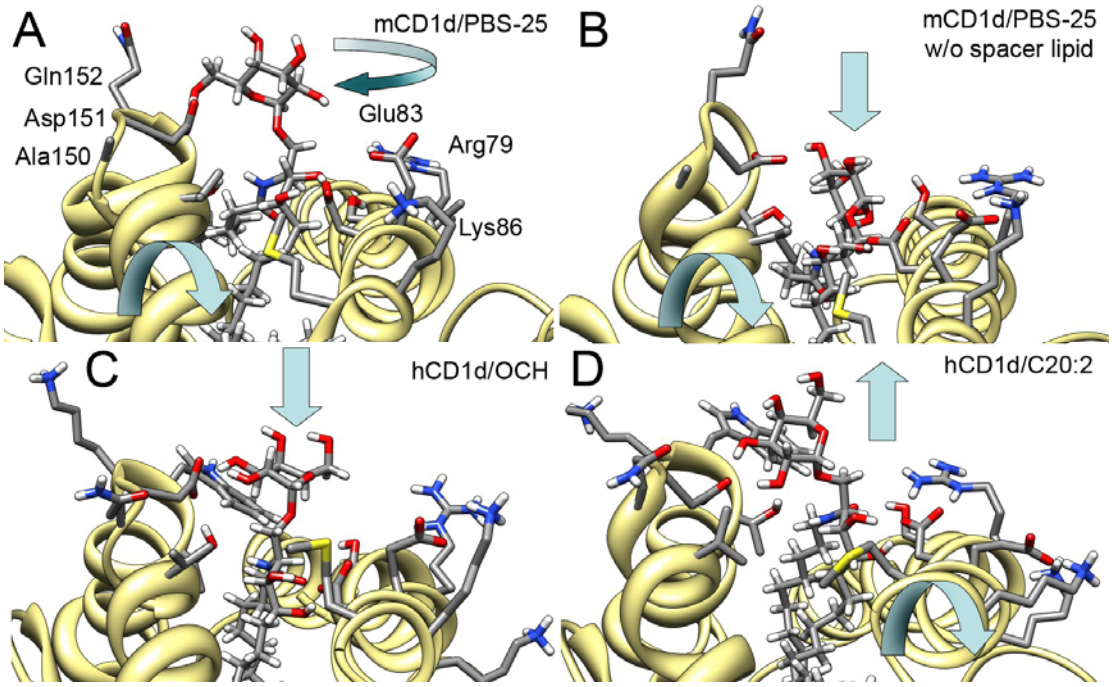


Figure 39. MD results of CD1d bound glycolipids with lipid modifications. (**A**) The mCD1d/PBS-25 crystal structure permitted enough freedom that the galactose sugar was capable of rotating around. (**B**) The mCD1d/PBS-25 structure without the spacer lipid in the A´ pocket was shown to be completely buried in the binding groove implying that the groove must be filled. (**C**) The hCD1d/OCH complex only showed the glycolipid sitting somewhat lower in the groove. (**D**) The hCD1d/C20:2 complex showed that saturation provides some rigidity into the acyl chain causing it to kick up the sugar in the binding groove.

There are two possibilities in regards to what should be conceived as an untolerable sugar orientation: either there is something inherently wrong in the forcefield used to parameterize the glycolipid linkage whereby it permitted too much flexibility or the simulation is valid and glycolipid linkages are indeed flexible and the lack of a blocking residue such as a tryptophan permits such a rotation to occur. It is true that the forcefield is not calibrated for sugars let alone glycosidic linkages between sugars and lipids, however, until this point all of the simulations appeared to handle the glycolipids relatively well. On the other hand, glycosidic linkages are indeed flexible, however, there

discrete preferred conformations within which the sugar resides with the potential of exchanging between them. Therefore, it is more than likely that the sugar rotation event is a combination of both effects.

The cross-reactivity of both species has been validated, but if there is more flexibility in the mCD1d protein then the current dogma must be reevaluated. If it is indeed the case that in the mCD1d the sugar is permitted free rotation with the 6′-OH capable of forming a hydrogen bond, then no longer are the 2′- nor the 3′-OH critical positions. Considering that the residues on the α1 helix are still oriented in the correct fashion to interact with TCR and there is no Gln150 on mCD1d that would sterically clash upon binding, it may be possible that the rotated sugar orientation could be a viable possibility.

The simulation of mCD1d/PBS-25 without the spacer lipid with which it was crystallized showed that the filling of the A′ pocket of CD1d was necessary for a stable binding of the glycolipid. Without the spacer lipid the glycolipid completely disappeared into the binding cavity but since it maintained its hydrogen bonds with Asp151 it pulled the α2 helix kink way upward which would create a severe steric clash with TCR. Furthermore, the fact that the sugar is no longer visible out of the groove would also cause TCR to be unable to bind to the complex. On the other hand, the OCH glycolipid wherein the sphingosine chain is shortened bound quite well and seemingly adopted a similar overall orientation as was found for CD1d bound α-GalCer. Even though, OCH is capable of eliciting a biased cytokine response, since all the residues of CD1d appear to be able to form the necessary TCR contacts some other structural activity must be going on. For example, it is quite possible that the C′ pocket closes just enough because of the lacking sphingosine chain that the CDR3α loop sits in the groove only briefly causing

fast binding and therefore a full iNKT response is not possible.

The hCD1d/C20:2 complex MD simulation predicted that C20:2 should not be able to elicit an iNKT response. The di-unsaturated acyl chain caused it to become quite rigid thereby pushing the sugar upward and pulling in the sphingosine chain. This almost circular shifting caused the sugar to push down on Asp151 which kept the Gln150 residue in a horizontal orientation. The pulling back of the sphingosine chain caused the 3- and 4-OHs to shift up in the binding pocket pulling the Asp80 residue with them which then caused the Glu83 and Lys86 to shift laterally away from CD1d. Since all the glycolipids were built on the original binding orientation of α-GalCer, it is quite possible that this erroneous simulation result could be due to methodology used to build the system. In that it is possible that the di-unsaturated acyl chain does not bind in the same conformation as the saturated acyl chain of α-GalCer, and therefore, the result is not consistent with what would be found in nature from the beginning.

The two most interesting antigens that have been synthetically created for this system have been the C-glycoside and Threitolceramide (TCER) because neither exhibit a true sugar as their head group while still being able to stimulate iNKT cells. The C-glycoside was found capable of maintaining the right sugar orientation along with its hydrogen bonds to Asp151 even though its anomeric linkage was a methylene group (**Figure 40**). Its Gln150 was slight pointing vertically but it is not because the kink in the α2 helix was disrupted by the Asp151 being displaced and so was probably due to normal rotations of its torsions. The TCER was also capable of maintaining the hydrogen bonds to the sugar albeit a little differently as was hypothesized in its report. It was assumed that the flexible hydroxylated alkyl chain would bind in the same fashion as the galactose sugar and

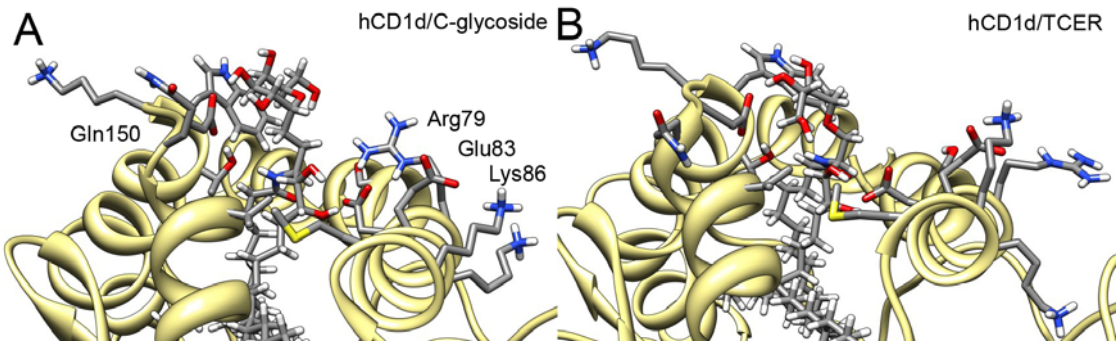


Figure 40. MD results of CD1d bound C-glycoside and Threitolceramide simulations. Both ligands were able to maintain all the necessary hydrogen bonds with CD1d along with the overall orientation of the important TCR binding residues.

therefore the starting coordinates were built as such. However, the 4′-OH was observed to flop forward into the binding groove and bind to the Asp151 residue which then shifted all the other groups. It was as if they reversed order and the 4′-OH played the role of the 2′-OH. Nevertheless, the TCR binding residues with the exception of Arg79 were all found to be in the correct orientation with the Gln150 horizontal and the α1 helix residues being vertically displayed out. Since the Arg79/Ala mutation was found to have no effect on iNKT activity it is not as important for it to be oriented vertically.

In all of the 3 ns simulations performed, the CD1d protein fluctuated very minimal as would be expected of a rigid protein. However, the residues on the α helices were capable of being displaced if only a few angstroms. It was found that three potential things can occur that would bar TCR from binding to the CD1d presenting glycolipid complex: (1) the glycolipid or sugar can be shifted upward or downward in the pocket; (2) the repositioning of the glycolipid causes hydrogen bonded residues to shift as well thereby screwing the rest of the α helix; and (3) the slight shift of the α helices will displace the TCR binding residues from the optimum binding orientation. When the Asp151 was

pulled up or down to maintain its hydrogen bonds with the ligand, it displaced the Gln150 residue disallowing it to form a hydrogen bond with Thr98α and making it a steric block to TCR sitting down onto the groove of CD1d. The intermittent hydrogen bonds with Ser76 would also cause the entire α1 helix to screw and displace the Arg79, Glu83, and Ly86 residues out laterally from CD1d rather than being in the optimum vertical position for TCR binding. Even though all these motions are relatively small, if the lock-and-key theory of CD1d/TCR binding is the accepted reality for the system then any motion of CD1d away from that found in the CD1d/α-GalCer complex would disrupt binding. And in the case of the MD simulations those ligands that are not supposed to elicit any iNKT activity were able to accomplish that, whereas, those with iNKT stimulatory activity were capable of maintaining the correct CD1d residue orientation.

CHAPTER 5:

A VIRTUAL SCREEN TO FIND A NOVEL ANTIGEN

HYPOTHESIS OF A NON-GLYCOLIPID ANTIGEN AND METHODOLOGY

There are a few lines of evidence that a glycolipid with structure of α-GalCer, sugar and lipid, is not needed to elicit an iNKT response. The finding of the self-lipid, phosphatidylcholine (PC), was the first sign of such a ligand. Other examples are the C-glycoside which does not possess a true anomeric linkage rather it has a methylene group and threitolceramide (TCER) which consists only of a hydroxylated alkyl chain. The final and best example is the phenyl 2,2,4,6,7-pentamethylbenzofuran-5-sulfonate which was found to be recognized by iNKT cells when presented by CD1d[144, 145].

These four compounds along with our computational studies led us to the belief that only the right guest molecule needed to be bound by the host CD1d. This ligand must be able to maintain hydrogen bonding with Asp151 and Asp80 along with being at the right height in the groove and not capable of displacing the residues on the CD1d α helices. Normally a vast biological screen of available compounds to find starter molecules from which to build a synthetic program, however, our laboratory as are most academic labs is not equipped for such a process. Instead, we chose to apply a virtual screen but limited to only finding a substitute for the galactose sugar. Our reasoning stemmed from the fact

that the majority of lipids were capable of being bound to CD1d so novel ligands without a sugar but possessing the α-GalCer ceramide backbone should be bound by CD1d without difficulty and therefore a lack of iNKT response can be directly attributed to the poor choice in the sugar-replacing group.

Our virtual screen program was a bit more complicated than normal since after finding choice molecules to replace the sugar they still needed to be docked back into the tertiary complex and then undergo simulation to verify that they do not disrupt the binding of TCR to CD1d (**Figure 41**). As a choice starting point, the library of 2,000,000 lead-like compounds were downloaded from the Zinc database[146] rather than the smaller fragment-based library since more complex groups were sought to replace the sugar. The benefit of the Zinc database is that it provides all the molecules pre-built in 3D for parallel docking runs whereas most databases only possess 2D information. The library was split into smaller 150,000 groups which were used in our high-throughput docking (HTD) runs.

DOCK v6 was the choice program to perform the HTD docking since it was built for it being scalable for parallel processors[147]. The tertiary complex was used as the acceptor compound to which docking was performed, however, three different sizes of the α-GalCer glycolipid were used to define the energetics of the cavity: the full glycolipid, truncated glycolipid (where the truncation occurred at the point where both chains bent), and the lone sugar. As it turned out, if the full glycolipid is used all the compounds are docked deep into the binding groove of CD1d and if the lone sugar was used then most compounds would be docked somewhere above CD1d but below TCR, i.e. none of the necessary hydrogen bonds such as to Asp151 were even considered by the docking program. Therefore, the best definition for the binding site was the truncated glycolipid

which contained both a shortened hydrophobic lipid tail along with the hydrophilic sugar portion. After the grid was defined for this space, the default virtual screening parameters were applied where 2500 max orientations were performed with a minimum of 40 anchors with each anchor undergoing 500 max iterations and growth iterations.

The top 1,500 docked compounds from the 15,000 scored of the 150,000 underwent visual analysis to determine binding and sugar replacement viability yielding roughly 100

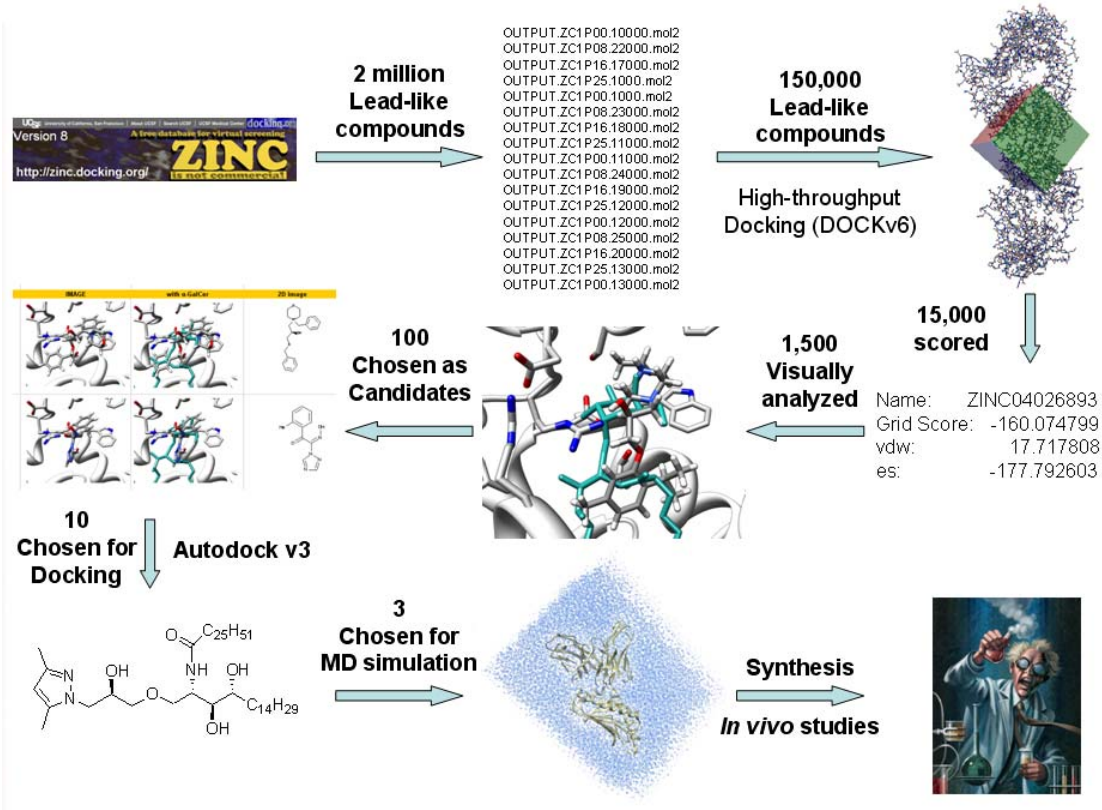


Figure 41. Virtual screening program to search for a non-glycolipid antigen. The library of 2,000,000 lead-like compounds was downloaded and split into smaller files of 150,000 compounds from the ZINC database. These portions were docked into the sugar binding cavity of the tertiary complex using DOCK v6. The top 1,500 docked compounds from the 15,000 scored of the 150,000 underwent visual analysis to determine binding and sugar replacement viability yielding roughly 100 compounds. These 100 were then analyzed and 10-15 were chosen in regards to availability or synthetic ease. The chosen compounds were then built into potential ligands by attaching them to the lipid portion of α-GalCer. These were then submitted to Autodock following the previous methodology. The top 3 best binding ligands then underwent a 3 ns simulation following the MD protocol used for the modified glycolipids.

compounds (*See Appendix B*). These 100 were then analyzed and 10-15 were chosen in regards to commercial availability or synthetic ease. The chosen compounds were then built into potential ligands by attaching them to the lipid portion of α-GalCer, and were then submitted to Autodock to be docked back into CD1d following the previous methodology used for the 2′-, 3′-, and 4′- substituted glycolipid. The top 3 best binding ligands then underwent a 3 ns simulation within the tertiary complex following the MD protocol used for in the 2′-, 3′-, and 4′- substituted glycolipids. To date, 600,000 of the 2,000,000 lead-like compounds have been docked and visually analyzed for sugar replacement potential.

From the first 150,000 compounds that were docked, a library of 10 head groups were chosen, built onto the ceramide, and docked into the tertiary complex (**Figure 42**). The majority of lead-like compounds docked into the tertiary complex contained aromatic moieties. Only those compounds were chosen that appeared to possess hydrogen bonding capabilities along with the potential of linking to the lipid portion. Compounds Z1-12 were all derivatives of the lead-like compound hits with the exception of Z3 and Z4 which were just permutations chosen based on the Z2 compound. The compounds Z13 and Z14 were not hits from the HTD docking but rather were manually built in order to determine whether or not lone aromatic groups could satisfy replacing the sugar. The beauty of theoretical docking can be appreciated in compounds Z15 and Z16 which were created after it was determined that compound Z7 would be synthetically difficult. The original ketone linker between the two aromatic rings was replaced with an amide bond which is an easier linker to manipulate, and the meta position on the aromatic ring linked to the glycolipid was replaced with a methoxy substituent. The compounds Z17 and Z18

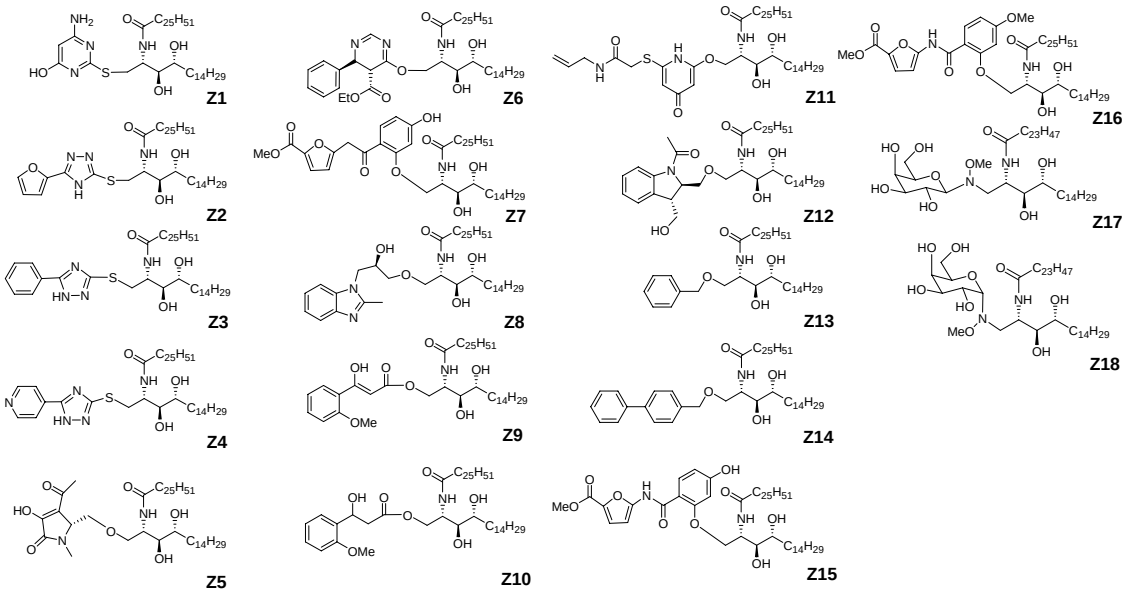


Figure 42. Original non-glycolipid compounds submitted to Autodock. These 12 compounds were created from the first 150,000 high-throughput docked lead-like compounds. The Z13-Z18 compounds were manually created to either satisfy scientific curiosity (Z13 and Z14), to solve synthetic issues with previous compounds (Z15 and Z16), or answer questions from previous synthetic studies (Z17 and Z18).

were remnants of an earlier investigation where it was thought that a nitrogen could be a potential linker between the sugar and lipid much like the thiol or methylene group were.

The compounds were docked using Autodock following the same methodology and using the same structures that were used in the 2′-, 3′-, and 4′-glycolipid investigations. Much like then, α-GalCer was found to be neither the best nor the worst ligand, but existed energetically about half-way with the exception of the TCR binding energies where it was the top ligand. This caused a similar concern as previously wherein the docked energies could not provide a clear answer to which compound would be an ideal ligand for the system. Therefore, the hypothesis was made that compounds that exhibited high binding energies to the tertiary complex, to CD1d, and to the TCR would probably be ideal ligands. The tertiary complex was found to bind Z7, Z8, Z11 and Z12 the best,

and these four compounds were also the top binders to CD1d (**Tables 7-9**). With the exception of Z7, the other compounds were found to only partially contribute to the TCR binding energy. Interestingly, its derivative, Z15, was found to also be a good binder of TCR, and was capable of binding to the tertiary complex and CD1d as well.

These original 150,000 compounds yielded quite interesting results, but the docking of the next 150,000 lead-like compounds were able to validate our methodology and hypothesis. Compounds containing phosphorous groups and inositol-like compounds were returned back as hits. Since the PC self-lipid contained a phosphorous group and α-GalCer was comprised of a sugar, this was a good sign for the strength and accuracy of the docking program. No phosphorous containing compounds were chosen for further docking runs, but some additional aromatic compounds compounds were chosen for continued analysis (Z19-Z22) (**Figure 43**). The two inositol-like compounds were also built into glycolipids (Z23-24) and a few heterocyclic nitrogen rings were built as well (Z25-Z26). The compounds Z27-Z29 were built based on the HTD docking orientation of the inositol-like compounds where they mimicked the orientation that would be found if an L-sugar glycolipid existed, therefore, L-talose, L-ribose, and L-tagastose were attached to the ceramide. The Z30 compound was also docked to determine whether an additional nitrogen in the aromatic ring would have any affect as compared to the original Z11 compound. The final group of compounds, Z31-Z34, were those that were finally synthesized based on the original investigation into the 2′-, 3′-, and 4′-glycolipid modifications, and as such they were included in the docking methodology. As it turned out, the addition of the amide linker was still found to be difficult, so it was thought that flipping the bond would make it synthetically easier and hence the Z35 compound.

Table 7. Ranked docked energies of the Z1-Z18 compounds bound to the tertiary complex. The compounds Z7, Z8, Z11, and Z12 were found to bind well to the tertiary complex as well as Z15 and Z16 which were derivatives of compound Z7.

TERTIARY COMPLEX RESULT (kcal/mol)															
Average (10 ns)				Average (5 ns)				Average (3 ns)				Crystal Structure			
(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)	
Z11	-28.15	**Z10**	-25.77	**Z11**	-28.68	**Z11**	-28.68	**Z7**	-28.65	**Z7**	-28.65	**Z16**	-30.67	**Z16**	-30.67
Z12	-27.66	**Z9**	-25.29	**Z7**	-28.48	**Z7**	-28.48	**Z11**	-28.53	**Z11**	-28.53	**Z15**	-30.52	**Z15**	-30.52
Z7	-27.60	**Z3**	-25.09	**Z12**	-28.38	**Z12**	-28.38	**Z15**	-28.44	**Z15**	-28.44	**Z7**	-30.45	**Z7**	-30.45
Z8	-27.49	**Z6**	-26.15	**Z15**	-28.11	**Z15**	-28.11	**Z16**	-28.29	**Z16**	-28.29	**Z8**	-29.86	**Z8**	-29.86
Z5	-26.82	**Z2**	-25.82	**Z8**	-28.03	**Z8**	-28.03	**Z8**	-28.24	**Z8**	-28.24	**Z12**	-29.63	**Z12**	-29.63
Z15	-26.82	**Z5**	-26.56	**Z5**	-27.74	**Z5**	-27.74	**Z12**	-28.22	**Z12**	-28.22	**Z11**	-29.61	**Z11**	-29.61
Z14	-26.81	**Z1**	-26.26	**Z14**	-27.58	**Z14**	-27.58	**Z5**	-28.19	**Z5**	-28.19	**Z14**	-28.57	**Z14**	-28.57
Z6	-26.40	**Z14**	-27.39	**Z0**	-27.47	**Z6**	-27.23	**Z14**	-27.99	**Z14**	-27.99	**Z6**	-28.51	**Z6**	-28.51
Z3	-26.32	**Z7**	-27.39	**Z6**	-27.23	**Z0**	-27.21	**Z0**	-27.95	**Z0**	-27.95	**Z5**	-28.32	**Z5**	-28.32
Z0	-26.15	**Z16**	-26.02	**Z4**	-27.06	**Z4**	-27.06	**Z6**	-27.44	**Z6**	-27.44	**Z9**	-27.64	**Z9**	-27.64
Z4	-26.04	**Z15**	-25.61	**Z3**	-26.97	**Z3**	-26.97	**Z3**	-27.16	**Z3**	-27.16	**Z0**	-27.45	**Z0**	-27.45
Z16	-26.02	**Z11**	-28.21	**Z2**	-26.51	**Z2**	-26.51	**Z4**	-27.13	**Z4**	-27.13	**Z18**	-27.32	**Z18**	-27.32
Z9	-25.62	**Z0**	-27.70	**Z9**	-26.47	**Z9**	-26.47	**Z2**	-26.72	**Z2**	-26.72	**Z3**	-27.24	**Z3**	-27.24
Z10	-25.51	**Z13**	-25.01	**Z16**	-26.33	**Z16**	-26.33	**Z9**	-26.61	**Z9**	-26.61	**Z4**	-27.21	**Z4**	-27.21
Z2	-25.34	**Z4**	-26.52	**Z10**	-26.19	**Z10**	-26.19	**Z18**	-26.43	**Z18**	-26.43	**Z10**	-27.14	**Z10**	-27.14
Z1	-25.34	**Z12**	-26.63	**Z1**	-26.15	**Z1**	-26.08	**Z10**	-26.27	**Z10**	-26.27	**Z2**	-26.75	**Z2**	-26.75
Z13	-25.20	**Z8**	-25.55	**Z18**	-26.00	**Z18**	-26.00	**Z13**	-26.25	**Z13**	-26.25	**Z13**	-26.11	**Z13**	-26.11
Z17	-24.80	**Z17**	-24.67	**Z13**	-25.94	**Z13**	-25.94	**Z1**	-26.21	**Z1**	-26.21	**Z17**	-25.94	**Z17**	-25.94
Z18	-23.71	**Z18**	-23.01	**Z17**	-24.08	**Z17**	-24.08	**Z17**	-25.91	**Z17**	-25.91	**Z1**	-25.58	**Z1**	-25.58

Table 8. Ranked docked energies of the Z1-Z18 compounds bound to the CD1d protein. The compounds same Z7, Z8, Z11, and Z12 were found to bind just as well to the CD1d protein.

CD1d RESULT (kcal/mol)															
Average (10 ns)				Average (5 ns)				Average (3 ns)				Crystal Structure			
(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)	
Z8	-28.29	**Z8**	-27.78	**Z8**	-27.98	**Z8**	-27.98	**Z8**	-27.78	**Z8**	-27.73	**Z11**	-28.71	**Z8**	-27.88
Z11	-28.21	**Z12**	-27.43	**Z11**	-27.66	**Z11**	-27.66	**Z11**	-27.78	**Z11**	-27.24	**Z8**	-28.09	**Z7**	-27.66
Z12	-27.78	**Z11**	-27.40	**Z12**	-27.60	**Z12**	-27.60	**Z12**	-27.52	**Z12**	-27.12	**Z7**	-27.66	**Z12**	-27.61
Z7	-27.42	**Z16**	-27.08	**Z7**	-27.12	**Z7**	-27.12	**Z7**	-27.15	**Z14**	-26.75	**Z12**	-27.61	**Z11**	-27.46
Z14	-27.40	**Z14**	-27.00	**Z14**	-27.10	**Z14**	-27.10	**Z14**	-27.13	**Z16**	-26.73	**Z16**	-27.45	**Z16**	-27.45
Z15	-27.14	**Z7**	-26.79	**Z5**	-26.89	**Z5**	-26.89	**Z16**	-26.73	**Z7**	-26.44	**Z15**	-27.40	**Z15**	-27.40
Z16	-27.14	**Z15**	-26.72	**Z16**	-26.79	**Z16**	-26.79	**Z5**	-26.61	**Z15**	-26.41	**Z17**	-27.18	**Z17**	-27.18
Z5	-26.93	**Z5**	-26.46	**Z15**	-26.76	**Z15**	-26.76	**Z15**	-26.60	**Z5**	-26.08	**Z14**	-27.01	**Z14**	-27.01
Z4	-26.65	**Z6**	-26.22	**Z3**	-26.48	**Z3**	-26.48	**Z3**	-26.39	**Z3**	-25.72	**Z0**	-26.85	**Z0**	-26.85
Z3	-26.57	**Z3**	-26.17	**Z4**	-26.47	**Z4**	-26.47	**Z4**	-26.38	**Z0**	-25.65	**Z5**	-26.70	**Z5**	-26.70
Z9	-26.39	**Z13**	-25.94	**Z2**	-26.22	**Z2**	-26.22	**Z13**	-26.04	**Z13**	-25.60	**Z3**	-26.55	**Z3**	-26.55
Z6	-26.36	**Z2**	-25.89	**Z13**	-26.08	**Z13**	-26.08	**Z9**	-25.97	**Z6**	-25.51	**Z4**	-26.52	**Z4**	-26.52
Z13	-26.32	**Z0**	-25.80	**Z0**	-26.04	**Z0**	-26.04	**Z2**	-25.94	**Z2**	-25.46	**Z9**	-26.35	**Z9**	-26.35
Z2	-26.24	**Z9**	-25.70	**Z9**	-26.04	**Z9**	-26.04	**Z0**	-25.87	**Z9**	-25.45	**Z18**	-26.28	**Z18**	-26.28
Z0	-26.08	**Z4**	-25.62	**Z6**	-25.91	**Z6**	-25.91	**Z6**	-25.64	**Z10**	-25.43	**Z13**	-26.23	**Z13**	-26.23
Z10	-25.99	**Z10**	-25.45	**Z10**	-25.75	**Z10**	-25.75	**Z10**	-25.63	**Z1**	-24.85	**Z2**	-26.13	**Z2**	-26.13
Z1	-25.42	**Z1**	-25.29	**Z1**	-25.21	**Z1**	-25.21	**Z1**	-24.98	**Z17**	-24.54	**Z6**	-26.04	**Z6**	-26.04
Z17	-24.90	**Z17**	-24.81	**Z17**	-24.96	**Z17**	-24.96	**Z18**	-24.81	**Z18**	-24.45	**Z10**	-25.82	**Z10**	-25.82
Z18	-24.62	**Z18**	-24.41	**Z18**	-24.83	**Z18**	-24.83	**Z17**	-24.77	**Z4**	-24.05	**Z1**	-25.27	**Z1**	-25.27

Table 9. Ranked docked energies of the Z1-Z18 compounds interacting with TCR. Compounds Z7 and Z15 appeared to be the only compounds continuing to possess high binding energies.

TCR BINDING RESULT (kcal/mol)															
Average (10 ns)				Average (5 ns)				Average (3 ns)				Crystal Structure			
(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)	
Z7	-0.13	**Z7**	-0.73	**Z0**	-1.43	**Z4**	-2.26	**Z0**	-2.08	**Z4**	-3.07	**Z16**	-3.22	**Z16**	-3.22
Z0	-0.04	**Z11**	-0.72	**Z7**	-1.37	**Z7**	-2.08	**Z15**	-1.84	**Z0**	-2.30	**Z15**	-3.12	**Z15**	-3.12
Z6	0.03	**Z5**	-0.34	**Z15**	-1.35	**Z15**	-1.48	**Z6**	-1.80	**Z7**	-2.21	**Z7**	-2.79	**Z7**	-2.79
Z11	0.10	**Z4**	-0.34	**Z6**	-1.32	**Z6**	-1.46	**Z18**	-1.62	**Z5**	-2.11	**Z6**	-2.47	**Z6**	-2.47
Z1	0.12	**Z12**	-0.24	**Z18**	-1.17	**Z18**	-1.41	**Z5**	-1.58	**Z15**	-2.03	**Z12**	-2.02	**Z11**	-2.15
Z12	0.15	**Z6**	-0.09	**Z11**	-1.02	**Z11**	-1.38	**Z7**	-1.50	**Z18**	-1.98	**Z8**	-1.77	**Z12**	-2.02
Z5	0.19	**Z3**	-0.09	**Z1**	-0.94	**Z5**	-1.32	**Z1**	-1.24	**Z6**	-1.93	**Z5**	-1.62	**Z8**	-1.98
Z17	0.19	**Z10**	-0.05	**Z5**	-0.85	**Z0**	-1.32	**Z17**	-1.14	**Z3**	-1.44	**Z14**	-1.56	**Z5**	-1.62
Z3	0.34	**Z15**	-0.05	**Z12**	-0.79	**Z12**	-1.16	**Z16**	-1.05	**Z1**	-1.36	**Z10**	-1.32	**Z14**	-1.56
Z15	0.40	**Z0**	-0.02	**Z4**	-0.59	**Z3**	-1.13	**Z14**	-0.86	**Z17**	-1.36	**Z9**	-1.29	**Z10**	-1.32
Z10	0.55	**Z17**	-0.02	**Z3**	-0.49	**Z1**	-1.04	**Z2**	-0.78	**Z11**	-1.29	**Z18**	-1.04	**Z9**	-1.29
Z14	0.64	**Z1**	0.04	**Z14**	-0.48	**Z2**	-0.87	**Z3**	-0.77	**Z2**	-1.26	**Z11**	-0.90	**Z18**	-1.04
Z4	0.68	**Z14**	0.12	**Z10**	-0.45	**Z9**	-0.86	**Z11**	-0.75	**Z14**	-1.25	**Z4**	-0.69	**Z4**	-0.69
Z9	0.83	**Z9**	0.15	**Z9**	-0.43	**Z10**	-0.76	**Z4**	-0.75	**Z9**	-1.16	**Z3**	-0.69	**Z3**	-0.69
Z8	0.87	**Z8**	0.31	**Z2**	-0.30	**Z14**	-0.74	**Z12**	-0.70	**Z12**	-1.11	**Z2**	-0.62	**Z2**	-0.62
Z18	0.94	**Z2**	0.60	**Z8**	-0.05	**Z8**	-0.36	**Z9**	-0.64	**Z16**	-1.05	**Z0**	-0.60	**Z0**	-0.60
Z2	0.96	**Z13**	0.69	**Z13**	0.14	**Z13**	-0.27	**Z10**	-0.63	**Z10**	-0.84	**Z1**	-0.31	**Z1**	-0.31
Z13	1.17	**Z18**	0.74	**Z16**	0.46	**Z16**	0.46	**Z8**	-0.46	**Z13**	-0.64	**Z13**	0.12	**Z13**	0.12
Z16	1.32	**Z16**	1.25	**Z17**	0.88	**Z17**	0.73	**Z13**	-0.21	**Z8**	-0.51	**Z17**	1.24	**Z17**	1.24

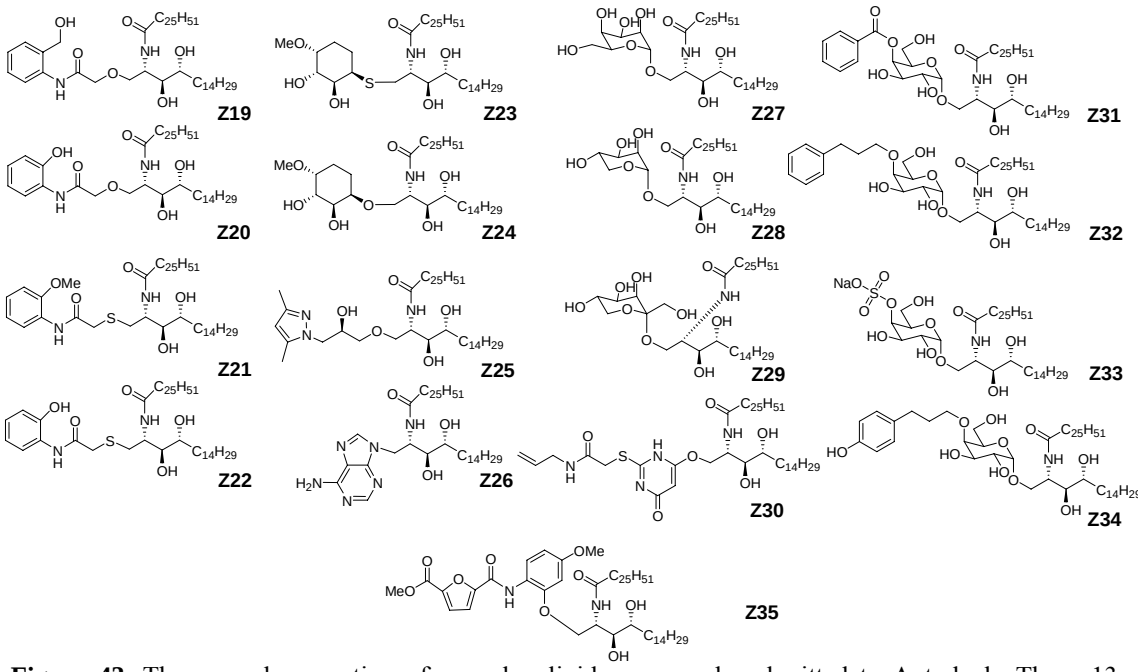


Figure 43. The second generation of non-glycolipid compounds submitted to Autodock. These 13 compounds were created from the second 150,000 high-throughput docked lead-like compounds. The Z27-Z29 sugar compounds were manually created based on the inositol-like compounds. The Z30 and Z36 were created to solve synthetic issues with previous compounds docked. The Z31-Z34 list of compounds were to test these synthesized compounds from the previous 4´- investigations.

The docking energy results showed that all four of the 4´-α-GalCer analogs bound well to both the tertiary complex and CD1d protein (**Tables 10-12**). Only in the crystal structure binding did they show good contribution to TCR binding with Z32 and Z35 being able to remain in the top. From the lead-like compounds, compound Z19 appeared to be the next best bound ligand. Interestingly, the heterocyclic aromatic compound Z26 which had almost no hydrogen bond donors or acceptors and limited flexibility was shown to be the worst binder followed very closely by Z29 which had many hydrogen bond donors and lots of flexible substituents. However, Z29 was found to be able to provide the most energetic contacts with TCR.

Table 10. Ranked docked energies of the Z19-Z35 compounds bound to the tertiary complex. The compounds Z7, Z8, Z11, and Z12 were found to bind well to the tertiary complex as well as Z15 and Z16 which were derivatives of compound Z7.

TERTIARY COMPLEX RESULT (kcal/mol)															
Average (10 ns)				Average (5 ns)				Average (3 ns)				Crystal Structure			
(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)	
Z32	-28.23	**Z25**	-27.89	**Z32**	-29.43	**Z32**	-29.43	**Z32**	-30.38	**Z32**	-30.38	**Z32**	-32.16	**Z32**	-32.16
Z25	-27.89	**Z32**	-27.87	**Z34**	-29.19	**Z34**	-29.19	**Z34**	-30.09	**Z34**	-30.09	**Z31**	-32.12	**Z31**	-32.12
Z30	-27.77	**Z30**	-27.56	**Z25**	-28.93	**Z25**	-28.93	**Z25**	-29.57	**Z25**	-29.57	**Z34**	-31.75	**Z34**	-31.75
Z34	-27.65	**Z34**	-27.54	**Z30**	-28.74	**Z30**	-28.65	**Z30**	-29.13	**Z30**	-29.13	**Z35**	-30.90	**Z35**	-30.90
Z35	-26.97	**Z35**	-26.88	**Z35**	-28.56	**Z35**	-28.46	**Z35**	-28.88	**Z35**	-28.73	**Z25**	-30.02	**Z25**	-30.02
Z19	-26.67	**Z19**	-26.67	**Z0**	-27.47	**Z19**	-27.33	**Z0**	-27.95	**Z0**	-27.95	**Z30**	-29.59	**Z30**	-29.59
Z31	-26.39	**Z20**	-26.29	**Z31**	-27.35	**Z27**	-27.31	**Z27**	-27.81	**Z27**	-27.81	**Z33**	-28.46	**Z33**	-28.46
Z20	-26.30	**Z22**	-26.13	**Z19**	-27.33	**Z0**	-27.21	**Z19**	-27.63	**Z19**	-27.63	**Z19**	-28.18	**Z19**	-28.18
Z0	-26.15	**Z31**	-25.92	**Z27**	-27.31	**Z20**	-27.19	**Z20**	-27.40	**Z20**	-27.40	**Z27**	-28.15	**Z27**	-28.15
Z22	-26.13	**Z21**	-25.91	**Z20**	-27.21	**Z31**	-26.96	**Z31**	-27.37	**Z31**	-27.37	**Z21**	-28.05	**Z21**	-28.05
Z27	-26.00	**Z0**	-25.79	**Z22**	-26.85	**Z22**	-26.85	**Z24**	-27.19	**Z24**	-27.19	**Z20**	-27.93	**Z20**	-27.93
Z21	-25.91	**Z27**	-25.72	**Z21**	-26.75	**Z21**	-26.75	**Z22**	-26.95	**Z22**	-26.95	**Z0**	-27.45	**Z0**	-27.45
Z24	-25.48	**Z24**	-25.43	**Z24**	-26.66	**Z24**	-26.66	**Z21**	-26.92	**Z21**	-26.92	**Z24**	-27.13	**Z24**	-27.13
Z28	-25.22	**Z28**	-25.10	**Z29**	-26.38	**Z29**	-26.38	**Z28**	-26.74	**Z28**	-26.65	**Z22**	-27.04	**Z22**	-27.04
Z33	-25.15	**Z33**	-25.07	**Z28**	-26.29	**Z28**	-26.23	**Z29**	-26.61	**Z29**	-26.61	**Z23**	-27.00	**Z23**	-27.00
Z23	-25.00	**Z29**	-24.97	**Z23**	-26.01	**Z23**	-25.96	**Z23**	-26.57	**Z23**	-26.54	**Z28**	-26.68	**Z28**	-26.68
Z29	-24.97	**Z23**	-24.76	**Z33**	-25.89	**Z33**	-25.89	**Z33**	-26.22	**Z33**	-26.22	**Z29**	-26.26	**Z29**	-26.26
Z26	-23.37	**Z26**	-23.30	**Z26**	-24.46	**Z26**	-24.33	**Z26**	-24.49	**Z26**	-24.49	**Z26**	-25.83	**Z26**	-25.83

Table 11. Ranked docked energies of the Z19-Z35 compounds bound to the tertiary complex. The compounds Z7, Z8, Z11, and Z12 were found to bind well to the tertiary complex as well as Z15 and Z16 which were derivatives of compound Z7.

CD1d RESULT (kcal/mol)															
Average (10 ns)				Average (5 ns)				Average (3 ns)				Crystal Structure			
(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)	
Z30	-28.25	**Z25**	-27.62	**Z34**	-27.93	**Z30**	-27.33	**Z30**	-27.84	**Z34**	-27.71	**Z34**	-29.20	**Z34**	-29.20
Z34	-28.13	**Z30**	-27.54	**Z25**	-27.87	**Z25**	-27.31	**Z34**	-27.81	**Z25**	-27.60	**Z32**	-28.96	**Z32**	-28.96
Z25	-28.10	**Z35**	-27.26	**Z30**	-27.84	**Z34**	-27.15	**Z25**	-27.73	**Z32**	-27.49	**Z30**	-28.61	**Z31**	-28.58
Z32	-27.80	**Z34**	-26.90	**Z32**	-27.65	**Z32**	-26.91	**Z32**	-27.49	**Z30**	-27.44	**Z31**	-28.58	**Z25**	-28.38
Z19	-27.43	**Z32**	-26.76	**Z19**	-27.16	**Z35**	-26.85	**Z35**	-27.24	**Z35**	-26.97	**Z25**	-28.38	**Z30**	-27.68
Z35	-27.42	**Z19**	-26.57	**Z35**	-27.08	**Z20**	-26.28	**Z19**	-27.10	**Z31**	-26.91	**Z35**	-27.63	**Z35**	-27.38
Z31	-27.19	**Z20**	-26.45	**Z31**	-27.02	**Z31**	-26.20	**Z31**	-27.01	**Z20**	-26.50	**Z0**	-26.85	**Z0**	-26.85
Z20	-27.18	**Z31**	-26.23	**Z20**	-26.82	**Z19**	-26.09	**Z20**	-26.79	**Z19**	-26.23	**Z19**	-26.80	**Z19**	-26.78
Z22	-26.96	**Z21**	-26.01	**Z22**	-26.74	**Z27**	-25.93	**Z22**	-26.66	**Z24**	-25.88	**Z22**	-26.75	**Z27**	-26.69
Z21	-26.74	**Z27**	-25.88	**Z21**	-26.46	**Z0**	-25.90	**Z21**	-26.39	**Z27**	-25.86	**Z27**	-26.69	**Z24**	-26.67
Z27	-26.45	**Z0**	-25.80	**Z27**	-26.39	**Z21**	-25.84	**Z27**	-26.17	**Z21**	-25.78	**Z24**	-26.67	**Z20**	-26.50
Z28	-26.32	**Z22**	-25.74	**Z28**	-26.26	**Z22**	-25.38	**Z28**	-26.06	**Z33**	-25.68	**Z21**	-26.64	**Z28**	-26.39
Z24	-26.29	**Z28**	-25.34	**Z24**	-26.13	**Z23**	-25.24	**Z24**	-26.03	**Z0**	-25.65	**Z20**	-26.57	**Z23**	-26.32
Z0	-26.08	**Z24**	-25.08	**Z0**	-26.04	**Z33**	-24.96	**Z0**	-25.87	**Z28**	-25.43	**Z28**	-26.39	**Z21**	-26.26
Z23	-26.06	**Z23**	-25.03	**Z33**	-26.01	**Z28**	-24.93	**Z23**	-25.87	**Z22**	-25.31	**Z23**	-26.32	**Z26**	-26.17
Z33	-26.05	**Z33**	-25.01	**Z23**	-25.98	**Z29**	-24.75	**Z33**	-25.85	**Z23**	-25.10	**Z26**	-26.17	**Z33**	-26.12
Z29	-25.07	**Z29**	-24.97	**Z26**	-25.15	**Z24**	-24.74	**Z26**	-25.13	**Z26**	-24.90	**Z33**	-26.12	**Z22**	-25.92
Z26	-24.92	**Z26**	-24.40	**Z29**	-24.87	**Z26**	-24.44	**Z29**	-24.70	**Z29**	-24.51	**Z29**	-25.59	**Z29**	-25.59

Table 12. Ranked docked energies of the Z19-Z35 compounds bound to the tertiary complex. The compounds Z7, Z8, Z11, and Z12 were found to bind well to the tertiary complex as well as Z15 and Z16 which were derivatives of compound Z7.

TCR BINDING RESULT (kcal/mol)															
Average (10 ns)				Average (5 ns)				Average (3 ns)				Crystal Structure			
(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)		(1ST CONF.)		(AltE)	
Z32	-0.14	**Z32**	-0.85	**Z29**	-1.51	**Z32**	-2.23	**Z32**	-2.20	**Z0**	-2.30	**Z31**	-3.54	**Z31**	-3.54
Z0	-0.04	**Z34**	-0.59	**Z35**	-1.48	**Z24**	-1.92	**Z0**	-2.08	**Z32**	-2.20	**Z35**	-3.27	**Z35**	-3.52
Z29	0.13	**Z22**	-0.31	**Z0**	-1.43	**Z34**	-1.72	**Z29**	-1.91	**Z29**	-2.10	**Z32**	-3.20	**Z32**	-3.20
Z25	0.26	**Z25**	-0.21	**Z32**	-1.34	**Z29**	-1.63	**Z25**	-1.83	**Z25**	-1.97	**Z34**	-2.55	**Z34**	-2.55
Z27	0.48	**Z24**	-0.19	**Z25**	-1.06	**Z25**	-1.62	**Z27**	-1.64	**Z27**	-1.95	**Z33**	-2.34	**Z33**	-2.34
Z30	0.49	**Z19**	-0.10	**Z27**	-0.92	**Z35**	-1.62	**Z35**	-1.64	**Z35**	-1.76	**Z25**	-1.64	**Z30**	-1.91
Z35	0.50	**Z33**	-0.07	**Z30**	-0.90	**Z22**	-1.48	**Z34**	-1.53	**Z30**	-1.69	**Z27**	-1.46	**Z21**	-1.79
Z34	0.80	**Z0**	-0.02	**Z34**	-0.78	**Z27**	-1.39	**Z30**	-1.29	**Z34**	-1.66	**Z21**	-1.41	**Z25**	-1.64
Z31	0.81	**Z30**	-0.02	**Z31**	-0.54	**Z30**	-1.32	**Z24**	-1.16	**Z22**	-1.64	**Z19**	-1.38	**Z27**	-1.46
Z19	0.83	**Z29**	0.03	**Z24**	-0.53	**Z0**	-1.32	**Z23**	-0.70	**Z23**	-1.45	**Z20**	-1.36	**Z20**	-1.43
Z24	0.86	**Z20**	0.05	**Z20**	-0.39	**Z28**	-1.30	**Z28**	-0.68	**Z19**	-1.40	**Z30**	-0.98	**Z19**	-1.40
Z21	0.93	**Z27**	0.15	**Z21**	-0.29	**Z19**	-1.24	**Z20**	-0.61	**Z24**	-1.31	**Z23**	-0.68	**Z22**	-1.12
Z22	0.94	**Z21**	0.27	**Z19**	-0.16	**Z33**	-0.93	**Z19**	-0.53	**Z28**	-1.23	**Z29**	-0.67	**Z23**	-0.68
Z33	0.94	**Z28**	0.36	**Z22**	-0.12	**Z20**	-0.91	**Z21**	-0.53	**Z21**	-1.14	**Z0**	-0.60	**Z29**	-0.67
Z20	1.00	**Z35**	0.45	**Z28**	-0.03	**Z21**	-0.91	**Z33**	-0.37	**Z20**	-0.90	**Z24**	-0.46	**Z0**	-0.60
Z28	1.13	**Z23**	0.45	**Z23**	-0.03	**Z23**	-0.72	**Z31**	-0.36	**Z33**	-0.54	**Z22**	-0.29	**Z24**	-0.46
Z23	1.14	**Z31**	0.66	**Z33**	0.12	**Z31**	-0.23	**Z22**	-0.29	**Z31**	-0.47	**Z28**	-0.29	**Z28**	-0.29
Z26	1.62	**Z26**	1.13	**Z26**	0.69	**Z26**	0.11	**Z26**	0.64	**Z26**	0.41	**Z26**	0.34	**Z26**	0.34

MOLECULAR DYNAMICS SIMULATION

Nine of the thirty compounds that showed the best overall binding energies were chosen to undergo molecular dynamics simulations to determine whether they are acceptable ligands for the system. If any of these novel ligands would perturb the binding footprint of the CD1d/TCR complex, they can then be assumed to not be recognized by TCR, and therefore would be unacceptable ligands. So in keeping with the methodology of the 2′-, 3′-, and 4′- glycolipids, the crystal structure docked orientations were chosen as starting points for the 3 ns tertiary complex simulations.

The nine compounds chosen were Z2, Z11, Z15, Z19, Z25, Z29, Z31, Z32, and Z33. They contained small and bulky aromatic groups, hydrogen bond donors and acceptors, one was an L-sugar, and the last three were 4′-α-GalCer derivatives (**Figure 44**). The MD simulations showed that only two of the nine would be viable ligands for this system (**Figure 45**).

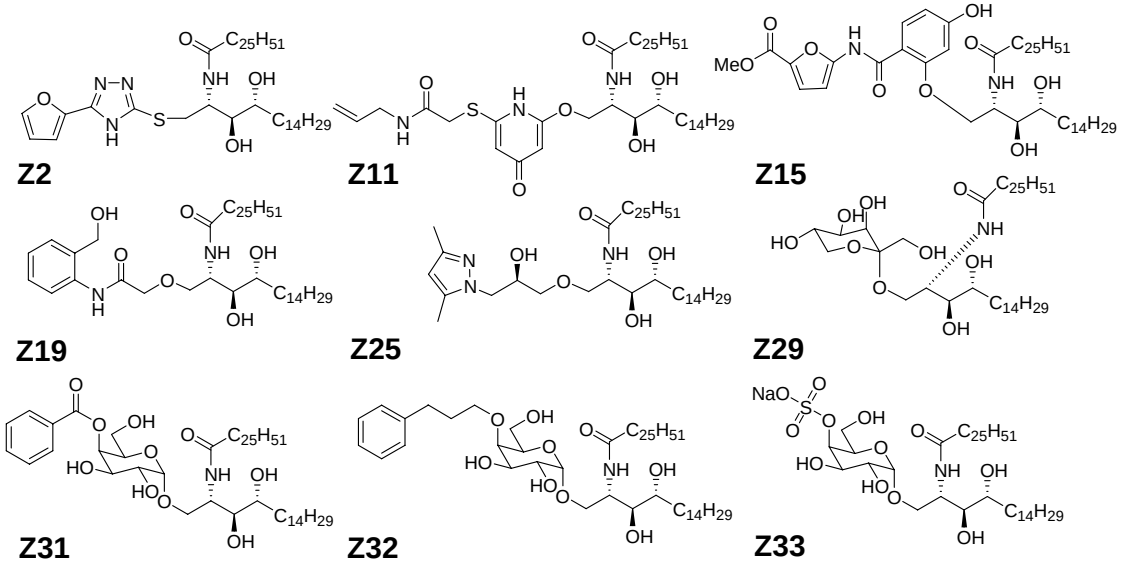


Figure 44. The virtual screen ligands that underwent MD simulations.

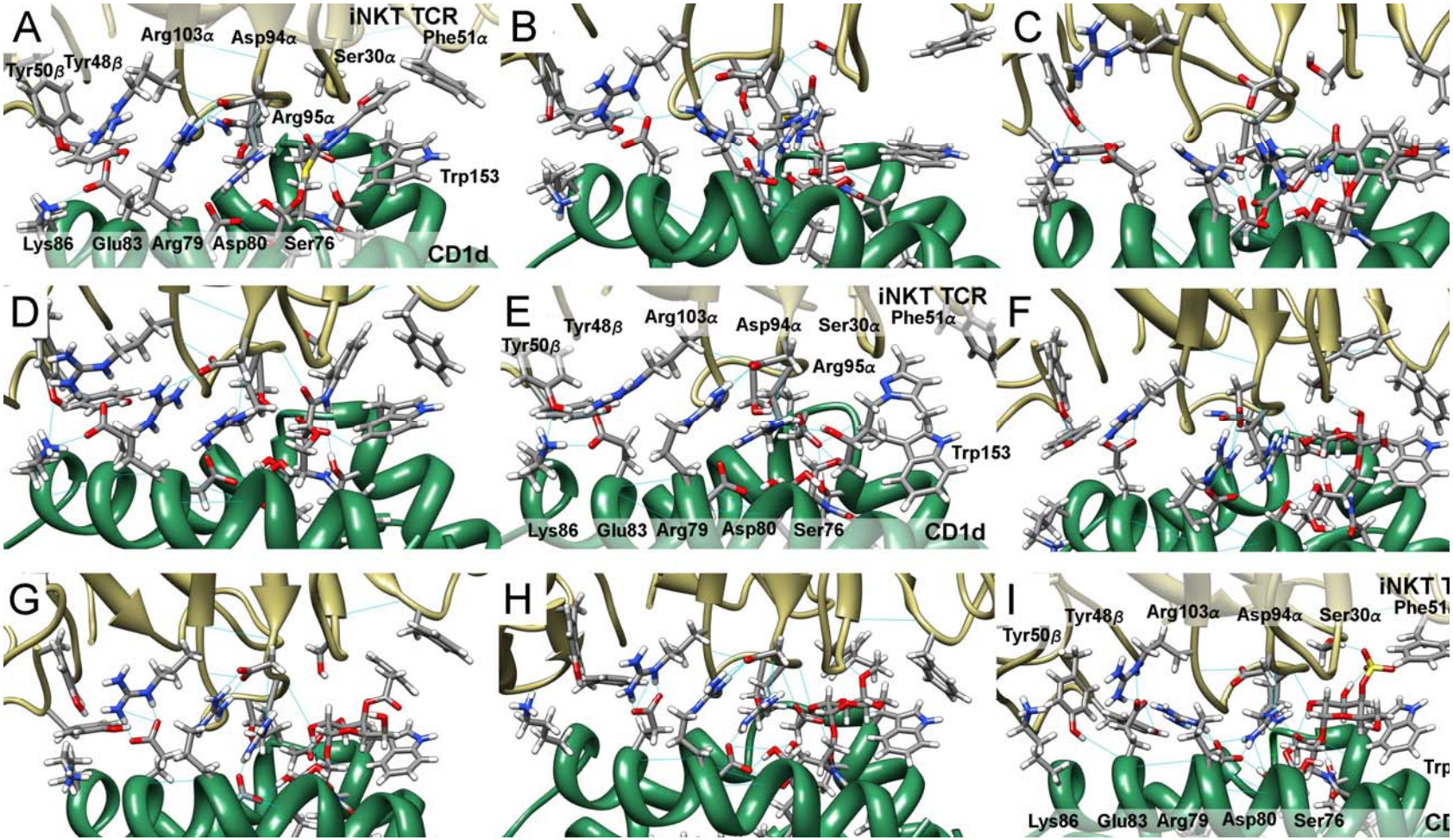


Figure 45. The MD results of the virtual screen ligands. **A**-Z2, **B**-Z11, **C**-Z15, **D**-Z19, **E**-Z25, **F**-Z29, **G**-Z31, **H**-Z32, and **I**-Z33. The compounds **B**-Z11, **F**-Z29, **G**-Z31, and **H**-Z32 appear to disrupt the tyrosine hydrogen bonds. Compound **A**-Z2 shows loss of one of tyrosine hydrogen bonds as do **D**-Z19 and **I**-Z33. Therefore, only **C**-Z15 and **E**-Z25 appear to be able to maintain the hydrogen bonding network of the CD1d/TCR binding footprint.

Only compound Z15 of the compounds Z2, Z11, and Z15 which were derived from the first 150,000 lead-like library docked that contained primarily aromatic type compounds was predicted to be a viable ligand for the system (**Figure 45 A-C**). It would appear that Z2 could be a viable ligand since it was able to maintain the majority of hydrogen bonding of the binding footprint and even hydrogen bond with Asp151. However, the Tyr50β residue appeared to be losing its hydrogen bond with Arg103α swinging over and beginning to monopolize the hydrogen bonding to Glu83. This was in stark contrast to the Z11 compound which completely disrupted the binding footprint between CD1d and TCR. The bulky sidechain was docked into the opening near the Asp151, but during the simulation it forced the CDR3α loop to eject from the hydrophobic binding groove of CD1d. This ejection caused both tyrosines to lose hydrogen bonding with Glu83. On the other hand, the equally bulky Z15 did not cause such an event to occur. Even though there was a slight displacement of the CDR3α loop out of the binding groove, the presence of the ester at the end of Z15 allowed a hydrogen bond to form between it and Ser97α. Furthermore, the tyrosine network was very well maintained, but the hydrogen bond between Arg79 and Asp94α was not. However, this is of no consequence since Arg79 is not found to be an important residue in the binding of TCR nor was Arg103α which rotated drastically away from the CD1d α1 helix. The important hydrogen bonding network between Asp94α, Arg95α, Asp80, and the ligand was maintained. Therefore, overall the Z15 compound even though it has caused the most difficulty from a synthetic point, appears to be the most viable ligand for the CD1d/TCR system.

The second set of compounds (Z19, Z25, and Z29) from the second library of 150,000 docked lead-like compounds also had similar results wherein only Z25 was found to be a

viable ligand for the CD1d/TCR system (**Figure 45D-F**). The Z19 compound much like its simple counterpart Z2 showed a slight disruption in the tyrosine hydrogen bonding network with Try50β being bonded to Ly86 rather than Glu83. However, unlike in Z2, the amide carbonyl group of Z19 was capable of forming a hydrogen bond with the Gly96α backbone on the CDR3α loop which is was an important hydrogen bond with the 2′-OH of α-GalCer. The Z19 compound was also able to maintain hydrogen bonding with Asp151, however, the bulky aromatic group appeared to shift the CDR3α loop out of the binding groove thereby allowing for a new hydrogen bond to form between the Ser97α and Asp151. Therefore, Z15 along with Z2 yielded interesting results that were only semi-conclusive.

The Z25 compound was found to maintain the hydrogen bonding network of the binding footprint. Other events though will probably contribute to its unacceptability such as the bulkiness and flexibility in the headgroup. The hydroxyl was not able to form a hydrogen bond with Asp151, but instead formed a novel hydrogen bond to Arg103α on the CDR3α loop. The bulky nitrogen ring also displaced the Phe51α residue which may or may not cause a significant displacement of the rest of the TCR protein. Nevertheless, all the necessary hydrogen bonds were maintained in the system with the exception of a direct one to Asp151, making this ligand appear to be a viable one for the system. The L-tagatose glycolipid derivate though was found to not maintain the hydrogen bonding of the tyrosine residues. It was also found that the L-sugar rotated around causing 2′- and 3′- OH groups to hydrogen bond to Asp151. The docking of the inositol-like compounds from which the L-glycolipids were derived showed that the axial hydroxyl should have maintained a hydrogen bond with the CDR3α loop to which it was originally also docked

to with Autodock. Regardless, the rotation of the sugar caused the lipid to be displaced from Arg103α and Asp80 preferring to hydrogen bond with Ser76 on the α1 helix of CD1d. Considering the importance of the glycolipid being anchored by Asp80 and sharing hydrogen bonds with Arg103α, it must be concluded that L-sugars would not be viable ligands for the CD1d/TCR system.

The two bulky aromatic 4′-α-GalCer derivatives were found to disrupt the hydrogen bonding of the tyrosines to Glu83 (**Figure 45G-I**). This was the same result found for PGW408 (**Figure 30C**) when it underwent 3 ns of MD simulation. On the other hand, the 4′-sulfo-α-GalCer though was able to maintain the hydrogen bonding of the binding footprint with a few perturbations. For example, Tyr50β was found to hydrogen bond to the backbone of the α1 helix of CD1d rather than Glu83. The Arg79 swung over to form hydrogen bonds with Glu83 much like the Arg103α residue. It would appear that the cause for these events was the new hydrogen bond between Ser30α and the sulfate on the sugar where it seemed to have allowed the CDR3α loop to be buried it deeper into the hydrophobic binding groove. Much like Z2 and Z19, the simulation results of Z33 are not conclusive enough to determine whether it would be a viable ligand or not. However, Z31 and Z32 due to their bulky 4′- substituents should not be ligands for the CD1d/TCR system.

At this time, two viable ligands were developed from the virtual screening program undertaken to find a non-glycolipid ligand for the CD1d/TCR system. These two ligands, Z15 and Z25, are both very different from each other but are both capable of maintaining the necessary binding footprint between CD1d and TCR. If the host-guest hypothesis is a more accurate representation for the system then these novel ligands should be able to

elicit some iNKT activity upon TCR recognition. However, to date, only 300,000 of the compounds have made it to the final simulation stage. At this stage, the next 300,000 are undergoing Autodocking of the built lipid ligands with the other 1.4 million undergoing HTD docking. After synthesis and animal testing, it will be seen whether or not the power of theoretical prediction was capable of providing a novel ligand for the CD1d/TCR system.

CHAPTER 6:

CONCLUSION

The scientific project presented herein entailed the use of computational means to determine how the TCR protein of iNKT cells can differentiate so selectively between glycolipids presented by the CD1d protein in order to be able to design a better ligand for the system. The massive immune response cascade that follows after CD1d presentation of a glycolipid to iNKT cells has yielded a want of a better ligand with either comparable activity as α-GalCer but with less of its pharmaceutical hindrances or a ligand that can control the immune response. To date, superficial structure-activity relationships have been defined wherein modifications to either the sphingosine chain or acyl chain of the lipid can lead to a bias in the immune response, and modifications to the galactose sugar have led to null activity.

This computational endeavor has provided a deeper understanding of the CD1d/TCR binding recognition event where it can be concluded that the lock-and-key and host-guest theories are what define the selectivity for this system. Molecular dynamics simulations using AMBER found the crystallized CD1d/α-GalCer/TCR tertiary complex to be stable and relatively rigid in explicit solvent. The 10 ns simulation of the full tertiary complex in explicit solvent defined by periodic boundary conditions in a NVE ensemble showed

little fluctuation in the binding footprint between the two proteins. This led to the testing of a truncated complex where the Cα and Cβ regions of TCR along with the β_2M associated protein and the β sheet of CD1d were eliminated, and this truncated complex was shown in simulation to be an adequate and a more efficient system as compared to the tertiary complex. The generalized Born implicit solvation parameters, albeit more efficient than the full tertiary complex simulations were found to provide too much flexibility in the proteins and ligand, along with being considerably less efficient than the truncated complex. Therefore, the truncated CD1d/TCR complex was used henceforward in all simulation studies as means to gage glycolipid interaction with both proteins.

A combination of high-level docking with AUTODOCK and simulation showed that modifications to the 2′- and 3′- positions of the galactose sugar are indeed not tolerated, whereas, modifications to the 4′- position were semi-tolerated. A library of 50 glycolipids were docked into the CD1d/TCR truncated complex to yield a long list of binding energies and orientations. However, the binding energies appeared to not be correlated to expected iNKT stimulatory ability of the glycolipid. Therefore, a small library of 2′-, 3′-, and 4′-α-GalCer derivative glycolipids docked to the crystal structure orientation were submitted to 3 ns long simulations in the truncated complex. The 3 ns mark was used as it was assumed that the proteins would immediately fluctuate to accommodate the novel modification on the glycolipid and that both tertiary and truncated proteins showed similar overall RMSD fluctuations until this point. It was observed that modifications to the 2′- and 3′- positions of the galactose sugar caused a disruption of the binding footprint between CD1d and TCR specifically between the tyrosine residues of TCR and the glutamine residue on CD1d. The disruption was found

to be exasperated on increasing the bulkiness of the substituent. The 4´-substituents were found to not disrupt the binding footprint and therefore are tolerable modifications to the galactose until the bulkiness of the substituent increased enough to shift TCR off the binding site.

The ability of glycolipids to perturb the binding footprint between CD1d and TCR led to an investigation of experimentally tested glycolipids bound solely to CD1d/glycolipid whose results showed that those glycolipids that have been found to be incapable of stimulating iNKT cells changed the direction of the CD1d residues that interact with TCR away from optimum orientation. The 10 ns simulations showed both the TCR and CD1d proteins to be extremely rigid at their binding footprint, and this evidence in conjunction with their tertiary crystal structure deviating little from the individual crystal structures provided the necessary evidence to support the lock-and-key interaction hypothesis. It was found that multiple events occurred when a non-antigen was bound by CD1d: the sugar could be displaced either out of or deeper into the binding groove; the displacement caused the Asp151 residue to shift displacing the α2 helix and thereby causing the Gln150 residue to adopt an orientation that would sterically bar TCR from sitting down on CD1d; and the ability to hydrogen bond with Ser76 would cause the α1 helix to screw such that the residues interacting with TCR would splay out laterally away from the vertical orientation necessary to form hydrogen bonds with TCR. Therefore, the accumulation of these changes within the CD1d protein causing residues to shift away from their optimum orientation would eliminate recognition of the CD1d presented glycolipid by TCR and hence cause no iNKT activity.

Lastly, the evidence of a few non-glycolipid ligands, the self-ligand PC, the C-

glycoside and threitolceramide, being able to activate iNKT cells led to an undertaking of a virtual screening program to find a replacement for the galactose sugar yielding a library of viable aromatic-based lipid ligands. The Zinc database 2,000,000 large library of lead-like compounds were high-throughput docked into the binding cavity of the galactose sugar of α-GalCer. The choice compounds were built onto lipids, re-docked using AUTODOCK, and then a select few underwent MD simulations in the truncated tertiary complexes to determine whether they would be viable ligands for the system or not. To date, two very different analogs have been predicted to be viable antigens, however, until many more are tested, synthesized, and submitted for animal studies, the hypothesis that a non-glycolipid may cause an iNKT response remains to be validated.

In the end, only by the analysis of the interactions between the CD1d and TCR proteins upon glycolipid presentation at the atomic level has a structure-activity relationship been defined for the system. Through intensive molecular dynamics simulations and thousands of docking runs was an understanding achieved of how TCR could selectively differentiate between glycolipids presented by CD1d. Even though until now, hydrogen bonding has been an important factor in explaining glycolipid presentation to TCR, it was found that the effect of the glycolipid on the overall residue orientations of CD1d that truly aid in TCR recognition was the cause behind null iNKT activity. As iNKT cells' TCR proteins scan the surface of cells with CD1d proteins presenting glycolipids, only those lipids that are capable of maintaining the correct CD1d residue orientation are recognized by TCR.

REFERENCES

1. Gumperz, J. E. The ins and outs of CD1 molecules: bringing lipids under immunological surveillance. *Traffic* **2006**, *7*, 2-13.
2. Krummel, M. F.; Davis, M. M. Dynamics of the immunological synapse: finding, establishing and solidifying a connection. *Curr. Opin. Immunol.* **2002**, *14*, 66-74.
3. Bromley, S. K.; Burack, W. R.; Johnson, K. G.; Somersalo, K.; Sims, T. N.; Sumen, C.; Davis, M. M.; Shaw, A. S.; Allen, P. M.; Dustin, M. L. The immunological synapse. *Annu. Rev. Immunol.* **2001**, *19*, 375-396.
4. Gao, G. F.; Jakobsen, B. K. Molecular interactions of coreceptor CD8 and MHC class I: the molecular basis for functional coordination with the T-cell receptor. *Immunol. Today* **2000**, *21*, 630-636.
5. Tupin, E.; Kinjo, Y.; Kronenberg, M. The unique role of natural killer T cells in the response to microorganisms. *Nat. Rev. Microbiol* **2007**, *5*, 405-417.
6. Matsui, K.; Boniface, J. J.; Reay, P. A.; Schild, H.; Fazekas de St. Groth, B.; Davis, M. M. Low affinity interaction of peptide-MHC complexes with T cell receptors. *Science* **1991**, *254*, 1788-1791.
7. Godfrey, D. I.; MacDonald, H. R.; Kronenberg, M.; Smyth, M. J.; Van Kaer, L. NKT cells: what's in a name? *Nat. Rev. Immunol.* **2004**, *4*, 231-237.
8. Van Kaer, L. a-Galactosylceramide therapy for autoimmune diseases: prospects and obstacles. *Nat. Rev. Immunol.* **2005**, *5*, 31-42.
9. Barral, D. C.; Brenner, M. B. CD1 antigen presentation: how it works. *Nat. Rev. Immunol.* **2007**, *7*, 929-941.
10. McMichael, A. J.; Pilch, J. R.; Galfre, G.; Mason, D. Y.; Fabre, J. W.; Milstein, C. A human thymocyte antigen defined by a hybrid myeloma monoclonal antibody. *Eur. J. Immunol.* **1979**, *9*, 205-210.
11. Porcelli, S. A.; Modlin, R. L. The CD1 system: Antigen-presenting molecules for T cell recognition of lipids and glycolipids. *Annu. Rev. Immunol.* **1999**, *17*, 297-329.
12. Brossay, L.; Chioda, M.; Burdin, N.; Koezuka, Y.; Casorati, G.; Dellabona, P.; Kronenberg, M. CD1d-mediated recognition of an a-galactosylceramide by natural killer T cells is highly conserved through mammalian evolution. *J. Exp. Med.* **1998**, *188*, 1521-1528.
13. Dascher, C. C.; Hiromatsu, K.; Xiong, X.; Sugita, M.; Buhlmann, J. E.; Dodge, I. L.; Lee, S. Y.; Roura-Mir, C.; Watts, G. F.; Roy, C. J.; Behar, S. M.; Clemens, D. L.; Porcelli, S. A.; Brenner, M. B. Conservation of CD1 intracellular trafficking patterns between mammalian species. *J. Immunol.* **2002**, *169*, 6951-6958.
14. Lawton, A. P.; Kronenberg, M. The third way: Progress on pathways of antigen processing and presentation by CD1. *Immunol. Cell Biol.* **2004**, *82*, 295-306.
15. Chiu, Y.-H.; Park, S.-H.; Benlagha, K.; Forestier, C.; Jayawardena-Wolf, J.; Savage, P. B.; Teyton, L.; Bendelac, A. Multiple defects in antigen presentation and T

cell development by mice expressing cytoplasmic tail-truncated CD1d. *Nat. Immunol.* **2002**, *3*, 55-60.
16. Chiu, Y.-H.; Jayawardena, J.; Weiss, A.; Lee, D.; Park, S.-H.; Dautry-Varsat, A.; Bendelac, A. Distinct subsets of CD1d-restricted T cells recognize self-antigens loaded in different cellular compartments. *J. Exp. Med.* **1999**, *189*, 103-110.
17. Yuan, W.; Qi, X.; Tsang, P.; Kang, S.-J.; Illaniorov, P. A.; Besra, G. S.; Gumperz, J.; Cresswell, P. Saposin B is the dominant saposin that facilitates lipid binding to human CD1d molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 5551-5556.
18. Riken Research. http://www.rikenresearch.riken.jp/frontline/303/ (accessed June 21, 2009),
19. Tomura, M.; Yu, W.-G.; Ahn, H.-J.; Yamashita, M.; Yang, Y.-F.; Ono, S.; Hamaoka, T.; Kawano, T.; Taniguchi, M.; Koezuka, Y.; Fujiwara, H. A novel function of Valpha 14+CD4+NKT cells: stimulation of IL-12 production by antigen-presenting cells in the innate immune system. *J. Immunol.* **1999**, *163*, 93-101.
20. Vincent, M. S.; Leslie, D. S.; Gumperz, J. E.; Xiong, X.; Grant, E. P.; Brenner, M. B. CD1-dependent dendritic cell instruction. *Nat. Immunol.* **2002**, *3*, 1163-1168.
21. Kim, C. H.; Johnston, B.; Butcher, E. C. Trafficking machinery of NKT cells: shared and differential chemokine receptor expression among Valpha 24+Vbeta 11+ NKT cell subsets with distinct cytokine-producing capacity. *Blood* **2002**, *100*, 11-16.
22. Thomas, S. Y.; Hou, R.; Boyson, J. E.; Means, T. K.; Hess, C.; Olson, D. P.; Strominger, J. L.; Brenner, M. B.; Gumperz, J. E.; Wilson, S. B.; Luster, A. D. CD1d-Restricted NKT Cells Express a Chemokine Receptor Profile Indicative of Th1-Type Inflammatory Homing Cells. *J. Immunol.* **2003**, *171*, 2571-2580.
23. Eberl, G.; MacDonald, H. R. Rapid death and regeneration of NKT cells in anti-CD3e- or IL-12-treated mice: a major role for bone marrow in NKT cell homeostasis. *Immunity* **1998**, *9*, 345-353.
24. Wilson, M. T.; Johansson, C.; Olivares-Villagomez, D.; Singh, A. K.; Stanic, A. K.; Wang, C.-R.; Joyce, S.; Wick, M. J.; Van Kaer, L. The response of natural killer T cells to glycolipid antigens is characterized by surface receptor down-modulation and expansion. *Proc. Nat. Acad. Sci. U.S.A.* **2003**, *100*, 10913-10918.
25. Parekh, V. V.; Wilson, M. T.; Olivares-Villagomez, D.; Singh, A. K.; Wu, L.; Wang, C.-R.; Joyce, S.; Van Kaer, L. Glycolipid antigen induces long-term natural killer T cell anergy in mice. *J. Clin. Invest.* **2005**, *115*, 2572-2583.
26. Sullivan, B. A.; Kronenberg, M. Activation or anergy: NKT cells are stunned by alpha -galactosylceramide. *J. Clin. Invest.* **2005**, *115*, 2328-2329.
27. Nieuwenhuis, E. E. S.; Matsumoto, T.; Exley, M.; Schleipman, R. A.; Glickman, J.; Bailey, D. T.; Corazza, N.; Colgan, S. P.; Onderdonk, A. B.; Blumberg, R. S. CD1d-dependent macrophage-mediated clearance of Pseudomonas aeruginosa from lung. *Nat. Med.* **2002**, *8*, 588-593.
28. Ashkar, A. A.; Rosenthal, K. L. Interleukin-15 and natural killer and NKT cells play a critical role in innate protection against genital herpes simplex virus type 2 infection. *J. Virol.* **2003**, *77*, 10168-10171.
29. Grubor-Bauk, B.; Simmons, A.; Mayrhofer, G.; Speck, P. G. Impaired clearance of herpes simplex virus type 1 from mice lacking CD1d or NKT cells expressing the semivariant Valpha 14-Jalpha 281 TCR. *J. Immunol.* **2003**, *170*, 1430-1434.

30. Hage, C. A.; Kohli, L. L.; Cho, S.; Brutkiewicz, R. R.; Twigg, H. L.; Knox, K. S. Human immunodeficiency virus gp120 downregulates CD1d cell surface expression. *Immunol. Lett.* **2005**, *98*, 131-135.
31. Duthie, M. S.; Kahn, M.; White, M.; Kapur, R. P.; Kahn, S. J. Critical proinflammatory and anti-inflammatory functions of different subsets of CD1d-restricted natural killer T cells during Trypanosoma cruzi infection. *Infect. Immun.* **2005**, *73*, 181-192.
32. Rudolph, M. G.; Stanfield, R. L.; Wilson, I. A. How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* **2006**, *24*, 419-466.
33. Tynan, F. E.; Borg, N. A.; Miles, J. J.; Beddoe, T.; El-Hassen, D.; Silins, S. L.; van Zuylen, W. J. M.; Purcell, A. W.; Kjer-Nielsen, L.; McCluskey, J.; Burrows, S. R.; Rossjohn, J. High Resolution Structures of Highly Bulged Viral Epitopes Bound to Major Histocompatibility Complex Class I: implications for T-cell receptor engagement and T-cell immunodominance. *J. Biol. Chem.* **2005**, *280*, 23900-23909.
34. Zhang, W.; Young, A. C. M.; Imarai, M.; Nathenson, S. G.; Sacchettini, J. C. Crystal structure of the major histocompatibility complex class I H-2Kb molecule containing a single viral peptide: Implications for peptide binding and T-cell receptor recognition. *Proc. Nat. Acad. Sci. U.S.A.* **1992**, *89*, 8403-8407.
35. Tynan, F. E.; Burrows, S. R.; Buckle, A. M.; Clements, C. S.; Borg, N. A.; Miles, J. J.; Beddoe, T.; Whisstock, J. C.; Wilce, M. C.; Silins, S. L.; Burrows, J. M.; Kjer-Nielsen, L.; Kostenko, L.; Purcell, A. W.; McCluskey, J.; Rossjohn, J. T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. *Nat. Immunol.* **2005**, *6*, 1114-1122.
36. Tynan, F. E.; Reid, H. H.; Kjer-Nielsen, L.; Miles, J. J.; Wilce, M. C. J.; Kostenko, L.; Borg, N. A.; Williamson, N. A.; Beddoe, T.; Purcell, A. W.; Burrows, S. R.; McCluskey, J.; Rossjohn, J. A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. *Nat. Immunol.* **2007**, *8*, 268-276.
37. Gagnon, S. J.; Borbulevych, O. Y.; Davis-Harrison, R. L.; Turner, R. V.; Damirjian, M.; Wojnarowicz, A.; Biddison, W. E.; Baker, B. M. T Cell Receptor Recognition via Cooperative Conformational Plasticity. *J. Mol. Biol.* **2006**, *363*, 228-243.
38. Mazza, C.; Malissen, B. What guides MHC-restricted TCR recognition? *Semin. Immunol.* **2007**, *19*, 225-235.
39. Garcia, K. C.; Teyton, L.; Wilson, I. A. Structural basis of T cell recognition. *Annu. Rev. Immunol.* **1999**, *17*, 369-397.
40. Sami, M.; Rizkallah, P. J.; Dunn, S.; Molloy, P.; Moysey, R.; Vuidepot, A.; Baston, E.; Todorov, P.; Li, Y.; Gao, F.; Boulter, J. M.; Jakobsen, B. K. Crystal structures of high affinity human T-cell receptors bound to peptide major histocompatibility complex reveal native diagonal binding geometry. *Protein Eng., Des. Sel.* **2007**, *20*, 397-403.
41. Mason, D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today* **1998**, *19*, 395-404.
42. Rudolph, M. G.; Wilson, I. A. The specificity of TCR/pMHC interaction. *Curr. Opin. Immunol.* **2002**, *14*, 52-65.

43. Armstrong, K. M.; Insaidoo, F. K.; Baker, B. M. Thermodynamics of T-cell receptor-peptide/MHC interactions: progress and opportunities. *J. Mol. Recognit.* **2008**, *21*, 275-287.
44. Carreno, L. J.; Gonzalez, P. A.; Kalergis, A. M. Modulation of T cell function by TCR/pMHC binding kinetics. *Immunobiology* **2006**, *211*, 47-64.
45. Cole, D. K.; Pumphrey, N. J.; Boulter, J. M.; Sami, M.; Bell, J. I.; Gostick, E.; Price, D. A.; Gao, G. F.; Sewell, A. K.; Jakobsen, B. K. Human TCR-Binding Affinity is Governed by MHC Class Restriction. *J. Immunol.* **2007**, *178*, 5727-5734.
46. McKeithan, T. W. Kinetic proofreading in T-cell receptor signal transduction. *Proc. Nat. Acad. Sci. U.S.A.* **1995**, *92*, 5042-5046.
47. Alcover, A.; Alarcon, B. Internalization and intracellular fate of TCR-CD3 complexes. *Crit. Rev. Immunol.* **2000**, *20*, 325-346.
48. Valitutti, S.; Mueller, S.; Cella, M.; Padovan, E.; Lanzavecchia, A. Serial triggering of many T-cell receptors by a few peptide-MHC complexes. *Nature* **1995**, *375*, 148-150.
49. Valitutti, S.; Dessing, M.; Aktories, K.; Gallati, H.; Lanzavecchia, A. Sustained signaling leading to T cell activation results from prolonged T cell receptor occupancy. Role of T cell actin cytoskeleton. *J. Exp. Med.* **1995**, *181*, 577-584.
50. Kronenberg, M.; Gapin, L. The unconventional lifestyle of NKT Cells. *Nat. Rev. Immunol.* **2002**, *2*, 557-568.
51. Parekh, V. V.; Wilson, M. T.; Van Kaer, L. INKT-cell responses to glycolipids. *Crit. Rev. Immunol.* **2005**, *25*, 183-213.
52. Lantz, O.; Bendelac, A. An invariant T cell receptor alpha chain is used by a unique subset of major histocompatibility complex class I-specific CD4+ and CD4-8- T cells in mice and humans. *J. Exp. Med.* **1994**, *180*, 1097-1106.
53. Cui, J.; Shin, T.; Kawano, T.; Sato, H.; Kondo, E.; Toura, I.; Kaneko, Y.; Koseki, H.; Kanno, M.; Taniguchi, M. Requirement for Valpha 14 NKT cells in IL-12-mediated rejection of tumors. *Science* **1997**, *278*, 1623-1626.
54. Gadola, S. D.; Dulphy, N.; Salio, M.; Cerundolo, V. Va24-JaQ-independent, CD1d-restricted recognition of a-galactosylceramide by human CD4+ and CD8ab+ T lymphocytes. *J. Immunol.* **2002**, *168*, 5514-5520.
55. Kawano, T.; Cui, J.; Koezuka, Y.; Toura, I.; Kaneko, Y.; Motoki, K.; Ueno, H.; Nakagawa, R.; Sato, H.; Kondo, E.; Koseki, H.; Taniguchi, M. CD1d-restricted and TCR-mediated activation of Va14 NKT cells by glycosylceramides. *Science* **1997**, *278*, 1626-1629.
56. Gui, M.; Li, J.; Wen, L.-J.; Hardy, R. R.; Hayakawa, K. TCRbeta chain influences but does not solely control autoreactivity of Valpha 14J281T cells. *J. Immunol.* **2001**, *167*, 6239-6246.
57. Sim, B.-C.; Holmberg, K.; Sidobre, S.; Naidenko, O.; Niederberger, N.; Marine, S. D.; Kronenberg, M.; Gascoigne, N. R. J. Surprisingly minor influence of TRAV11 (Valpha 14) polymorphism on NK T-receptor mCD1/alpha -galactosylceramide binding kinetics. *Immunogenetics* **2003**, *54*, 874-883.
58. Schumann, J.; Voyle Roger, B.; Wei, B.-Y.; MacDonald, H. R. Cutting edge: influence of the TCR V beta domain on the avidity of CD1d:alpha-galactosylceramide binding by invariant V alpha 14 NKT cells. *J. Immunol.* **2003**, *170*, 5815-5819.

59. Benlagha, K.; Weiss, A.; Beavis, A.; Teyton, L.; Bendelac, A. In vivo identification of glycolipid antigen-specific T cells using fluorescent CD1d tetramers. *J. Exp. Med.* **2000**, *191*, 1895-1903.
60. Zeng, Z. H.; Castano, A. R.; Segelke, B. W.; Stura, E. A.; Peterson, P. A.; Wilson, I. A. Crystal structure of mouse CD1: an MHC-like fold with a large hydrophobic binding groove. *Science* **1997**, *277*, 339-345.
61. Gadola, S. D.; Zaccai, N. R.; Harlos, K.; Shepherd, D.; Castro-Palomino, J. C.; Ritter, G.; Schmidt, R. R.; Jones, E. Y.; Cerundolo, V. Structure of human CD1b with bound ligands at 2.3.ANG., a maze for alkyl chains. *Nat. Immunol.* **2002**, *3*, 721-726.
62. Kobayashi, E.; Motoki, K.; Uchida, T.; Fukushima, H.; Koezuka, Y. KRN7000, a novel immunomodulator, and its antitumor activities. *Oncology Research* **1995**, *7*, 529-534.
63. Morita, M.; Motoki, K.; Akimoto, K.; Natori, T.; Sakai, T.; Sawa, E.; Yamaji, K.; Koezuka, Y.; Kobayashi, E.; Fukushima, H. Structure-Activity Relationship of a-Galactosylceramides against B16-Bearing Mice. *J. Med. Chem.* **1995**, *38*, 2176-2187.
64. Koch, M.; Stronge, V. S.; Shepherd, D.; Gadola, S. D.; Mathew, B.; Ritter, G.; Fersht, A. R.; Besra, G. S.; Schmidt, R. R.; Jones, E. Y.; Cerundolo, V. The crystal structure of human CD1d with and without a-galactosylceramide. *Nat. Immunol.* **2005**, *6*, 819-826.
65. Wu, D.; Zajonc, D. M.; Fujio, M.; Sullivan, B. A.; Kinjo, Y.; Kronenberg, M.; Wilson, I. A.; Wong, C. H. Design of natural killer T cell activators: structure and function of a microbial glycosphingolipid bound to mouse CD1d. *Proc. Nat. Acad. Sci. U.S.A.* **2006**, *103*, 3972-3977.
66. Zajonc, D. M.; Cantu, C.; Mattner, J.; Zhou, D.; Savage, P. B.; Bendelac, A.; Wilson, I. A.; Teyton, L. Structure and function of a potent agonist for the semi-invariant natural killer T cell receptor. *Nat. Immunol.* **2005**, *6*, 810-818.
67. Zajonc, D. M.; Maricic, I.; Wu, D.; Halder, R.; Roy, K.; Wong, C.-H.; Kumar, V.; Wilson, I. A. Structural basis for CD1d presentation of a sulfatide derived from myelin and its implications for autoimmunity. *J. Exp. Med.* **2005**, *202*, 1517-1526.
68. Giabbai, B.; Sidobre, S.; Crispin, M. D. M.; Sanchez-Ruiz, Y.; Bachi, A.; Kronenberg, M.; Wilson, I. A.; Degano, M. Crystal Structure of Mouse CD1d Bound to the Self Ligand Phosphatidylcholine: A Molecular Basis for NKT Cell Activation. *J. Immunol.* **2005**, *175*, 977-984.
69. Zajonc, D. M.; Ainge, G. D.; Painter, G. F.; Severn, W. B.; Wilson, I. A. Structural Characterization of Mycobacterial Phosphatidylinositol Mannoside Binding to Mouse CD1d. *J. Immunol.* **2006**, *177*, 4577-4583.
70. Henon, E.; Dauchez, M.; Haudrechy, A.; Banchet, A. Molecular dynamics simulation study on the interaction of KRN 7000 and three analogues with human CD1d. *Tetrahedron* **2008**, *64*, 9480-9489.
71. Gumperz, J. E.; Roy, C.; Makowska, A.; Lum, D.; Sugita, M.; Podrebarac, T.; Koezuka, Y.; Porcelli, S. A.; Cardell, S.; Brenner, M. B.; Behar, S. M. Murine CD1d-restricted T cell recognition of cellular lipids. *Immunity* **2000**, *12*, 211-221.
72. Joyce, S.; Woods, A. S.; Yewdell, J. W.; Bennink, J. R.; De Silva, A. D.; Boesteanu, A.; Balk, S. P.; Cotter, R. J.; Brutkiewicz, R. R. Natural ligand of mouse CD1d1: cellular glycosylphosphatidylinositol. *Science* **1998**, *279*, 1541-1544.

73. Fischer, K.; Scotet, E.; Niemeyer, M.; Koebernick, H.; Zerrahn, J.; Maillet, S.; Hurwitz, R.; Kursar, M.; Bonneville, M.; Kaufmann, S. H. E.; Schaible, U. E. Mycobacterial phosphatidylinositol mannoside is a natural antigen for CD1d-restricted T cells. *Proc. Nat. Acad. Sci. U.S.A.* **2004**, *101*, 10685-10690.
74. Speak, A. O.; Cerundolo, V.; Platt, F. M. CD1d presentation of glycolipids. *Immunol. Cell Biol.* **2008**, *86*, 588-597.
75. Brutkiewicz, R. R. CD1d Ligands: The Good, the Bad, and the Ugly. *J. Immunol.* **2006**, *177*, 769-775.
76. Gadola, S. D.; Koch, M.; Marles-Wright, J.; Lissin, N. M.; Shepherd, D.; Matulis, G.; Harlos, K.; Villiger, P. M.; Stuart, D. I.; Jakobsen, B. K.; Cerundolo, V.; Jones, E. Y. Structure and binding kinetics of three different human CD1d-a-galactosylceramide-specific T cell receptors. *J. Exp. Med.* **2006**, *203*, 699-710.
77. Kjer-Nielsen, L.; Borg, N. A.; Pellicci, D. G.; Beddoe, T.; Kostenko, L.; Clements, C. S.; Williamson, N. A.; Smyth, M. J.; Besra, G. S.; Reid, H. H.; Bharadwaj, M.; Godfrey, D. I.; Rossjohn, J.; McCluskey, J. A structural basis for selection and cross-species reactivity of the semi-invariant NKT cell receptor in CD1d/glycolipid recognition. *J. Exp. Med.* **2006**, *203*, 661-673.
78. Borg, N. A.; Wun, K. S.; Kjer-Nielsen, L.; Wilce, M. C. J.; Pellicci, D. G.; Koh, R.; Besra, G. S.; Bharadwaj, M.; Godfrey, D. I.; McCluskey, J.; Rossjohn, J. CD1d-lipid-antigen recognition by the semi-invariant NKT T-cell receptor. *Nature* **2007**, *448*, 44-49.
79. Wun, K. S.; Borg, N. A.; Kjer-Nielsen, L.; Beddoe, T.; Koh, R.; Richardson, S. K.; Thakur, M.; Howell, A. R.; Scott-Browne, J. P.; Gapin, L.; Godfrey, D. I.; McCluskey, J.; Rossjohn, J. A minimal binding footprint on CD1d-glycolipid is a basis for selection of the unique human NKT TCR. *J. Exp. Med.* **2008**, *205*, 939-949.
80. Bhat, S.; Spitalnik, S. L.; Gonzalez-Scarano, F.; Silberberg, D. H. Galactosyl ceramide or a derivative is an essential component of the neural receptor for human immunodeficiency virus type 1 envelope glycoprotein gp120. *Proc. Nat. Acad. Sci. U.S.A.* **1991**, *88*, 7131-7134.
81. Shimosaka, A. Role of NKT cells and alpha-galactosyl ceramide. *Int. J. Hematol.* **2002**, *76 Suppl 1*, 277-279.
82. Naumov, Y. N.; Bahjat, K. S.; Gausling, R.; Abraham, R.; Exley, M. A.; Koezuka, Y.; Balk, S. B.; Strominger, J. L.; Clare-Salzer, M.; Wilson, S. B. Activation of CD1d-restricted T cells protects NOD mice from developing diabetes by regulating dendritic cell subsets. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 13838-13843.
83. Taniguchi, M.; Seino, K.-i.; Nakayama, T. The NKT cell system: bridging innate and acquired immunity. *Nat. Immunol.* **2003**, *4*, 1164-1165.
84. Fujio, M.; Wu, D.; Garcia-Navarro, R.; Ho, D. D.; Tsuji, M.; Wong, C.-H. Structure-Based Discovery of Glycolipids for CD1d-Mediated NKT Cell Activation: Tuning the Adjuvant versus Immunosuppression Activity. *J. Am. Chem. Soc.* **2006**, *128*, 9022-9023.
85. Miyamoto, K.; Miyake, S.; Yamamura, T. A synthetic glycolipid prevents autoimmune encephalomyelitis by inducing TH2 bias of natural killer T cells. *Nature* **2001**, *413*, 531-534.
86. Yu, K. O. A.; Im, J. S.; Molano, A.; Dutronc, Y.; Illarionov, P. A.; Forestier, C.; Fujiwara, N.; Arias, I.; Miyake, S.; Yamamura, T.; Chang, Y.-T.; Besra, G. S.; Porcelli,

S. A. Modulation of CD1d-restricted NKT cell responses by using N-acyl variants of a-galactosylceramides. *Proc. Nat. Acad. Sci. U.S.A.* **2005**, *102*, 3383-3388.
87. Corver, J.; Moesby, L.; Erukulla, R. K.; Reddy, K. C.; Bittman, R.; Wilschut, J. Sphingolipid-dependent fusion of Semliki Forest virus with cholesterol-containing liposomes requires both the 3-hydroxyl group and the double bond of the sphingolipid backbone. *J. Virol.* **1995**, *69*, 3220-3223.
88. Lang, G. A.; Maltsev, S. D.; Besra, G. S.; Lang, M. L. Presentation of a-galactosylceramide by murine CD1d to natural killer T cells is facilitated by plasma membrane glycolipid rafts. *Immunology* **2004**, *112*, 386-396.
89. Rauch, J.; Gumperz, J.; Robinson, C.; Skoeld, M.; Roy, C.; Young, D. C.; Lafleur, M.; Moody, D. B.; Brenner, M. B.; Costello, C. E.; Behar, S. M. Structural Features of the Acyl Chain Determine Self-phospholipid Antigen Recognition by a CD1d-restricted Invariant NKT (iNKT) Cell. *J. Biol. Chem.* **2003**, *278*, 47508-47515.
90. Wu, D.; Fujio, M.; Wong, C.-H. Glycolipids as immunostimulating agents. *Bioorg. Med. Chem.* **2008**, *16*, 1073-1083.
91. Wilson, S. B.; Delovitch, T. L. Janus-like role of regulatory iNKT cells in autoimmune disease and tumour immunity. *Nat. Rev. Immunol.* **2003**, *3*, 211-222.
92. Park, S.-H.; Benlagha, K.; Lee, D.; Balish, E.; Bendelac, A. Unaltered phenotype, tissue distribution, and function of Valpha 14+ NKT cells in germ-free mice. *Eur. J. Immunol.* **2000**, *30*, 620-625.
93. van Der Vliet, H. J.; Nishi, N.; de Gruijl, T. D.; von Blomberg, B. M.; van den Eertwegh, A. J.; Pinedo, H. M.; Giaccone, G.; Scheper, R. J. Human natural killer T cells acquire a memory-activated phenotype before birth. *Blood* **2000**, *95*, 2440-2442.
94. Degroote, S.; Wolthoorn, J.; Van Meer, G. The cell biology of glycosphingolipids. *Semin. Cell Dev. Biol.* **2004**, *15*, 375-387.
95. Hayakawa, Y.; Godfrey, D. I.; Smyth, M. J. alpha -galactosylceramide: Potential immunomodulatory activity and future application. *Curr. Med. Chem.* **2004**, *11*, 241-252.
96. De Silva, A. D.; Park, J. J.; Matsuki, N.; Stanic Aleksandar, K.; Brutkiewicz Randy, R.; Medof, M. E.; Joyce, S. Lipid protein interactions: the assembly of CD1d1 with cellular phospholipids occurs in the endoplasmic reticulum. *J. Immunol.* **2002**, *168*, 723-733.
97. Park, J.-J.; Kang, S.-J.; De Silva, A. D.; Stanic, A. K.; Casorati, G.; Hachey, D. L.; Cresswell, P.; Joyce, S. Lipid-protein interactions: Biosynthetic assembly of CD1 with lipids in the endoplasmic reticulum is evolutionarily conserved. *Proc. Nat. Acad. Sci. U.S.A.* **2004**, *101*, 1022-1026.
98. Tsuji, M. Glycolipids and phospholipids as natural CD1d-binding NKT cell ligands. *Cell. Mol. Life Sci.* **2006**, *63*, 1889-1898.
99. Kinjo, Y.; Pei, B.; Bufali, S.; Raju, R.; Richardson, S. K.; Imamura, M.; Fujio, M.; Wu, D.; Khurana, A.; Kawahara, K.; Wong, C.-H.; Howell, A. R.; Seeberger, P. H.; Kronenberg, M. Natural Sphingomonas Glycolipids Vary Greatly in Their Ability to Activate Natural Killer T Cells. *Chem. Biol.* **2008**, *15*, 654-664.
100. Zhou, D.; Mattner, J.; Cantu, C., 3rd; Schrantz, N.; Yin, N.; Gao, Y.; Sagiv, Y.; Hudspeth, K.; Wu, Y.-P.; Yamashita, T.; Teneberg, S.; Wang, D.; Proia Richard, L.; Levery Steven, B.; Savage Paul, B.; Teyton, L.; Bendelac, A. Lysosomal glycosphingolipid recognition by NKT cells. *Science* **2004**, *306*, 1786-1789.

101. Zajonc, D. M.; Savage, P. B.; Bendelac, A.; Wilson, I. A.; Teyton, L. Crystal Structures of Mouse CD1d-iGb3 Complex and its Cognate Va14 T Cell Receptor Suggest a Model for Dual Recognition of Foreign and Self Glycolipids. *J. Mol. Biol.* **2008**, *377*, 1104-1116.
102. Speak Anneliese, O.; Salio, M.; Neville David, C. A.; Fontaine, J.; Priestman David, A.; Platt, N.; Heare, T.; Butters Terry, D.; Dwek Raymond, A.; Trottein, F.; Exley Mark, A.; Cerundolo, V.; Platt Frances, M. Implications for invariant natural killer T cell ligands due to the restricted presence of isoglobotrihexosylceramide in mammals. *Proc. Nat. Acad. Sci. U.S.A.* **2007**, *104*, 5971-5976.
103. Porubsky, S.; Speak, A. O.; Luckow, B.; Cerundolo, V.; Platt, F. M.; Groene, H.-J. Normal development and function of invariant natural killer T cells in mice with isoglobotrihexosylceramide (iGb3) deficiency. *Proc. Nat. Acad. Sci. U.S.A.* **2007**, *104*, 5977-5982.
104. Kinjo, Y.; Wu, D.; Kim, G.; Xing, G.-W.; Poles, M. A.; Ho, D. D.; Tsuji, M.; Kawahara, K.; Wong, C.-H.; Kronenberg, M. Recognition of bacterial glycosphingolipids by natural killer T cells. *Nature* **2005**, *434*, 520-525.
105. Mattner, J.; DeBord, K. L.; Ismail, N.; Goff, R. D.; Cantu, C.; Zhou, D.; Saint-Mezard, P.; Wang, V.; Gao, Y.; Yin, N.; Hoebe, K.; Schneewind, O.; Walker, D.; Beutler, B.; Teyton, L.; Savage, P. B.; Bendelac, A. Exogenous and endogenous glycolipid antigens activate NKT cells during microbial infections. *Nature* **2005**, *434*, 525-529.
106. Kohlerschmidt, D. J.; Musser, K. A.; Dumas, N. B., Identification of Gram-negative bacteria. In *Practical Handbook of Microbiology (2nd Edition)*, Goldman, E.; Green, L. H., Eds. CRC Press: Boca Raton, FL: 2009; pp 67-79.
107. Suvorov, M.; Fisher, J. F.; Mobashery, S., Bacterial cell wall morphology and biochemistry. In *Practical Handbook of Microbiology (2nd Edition)*, Goldman, E.; Green, L. H., Eds. CRC Press: Boca Raton, FL: 2009; pp 159-189.
108. Kawahara, K.; Moll, H.; Knirel, Y. A.; Seydel, U.; Zahringer, U. Structural analysis of two glycosphingolipids from the lipopolysaccharide-lacking bacterium Sphingomonas capsulata. *Eur. J. Biochem.* **2000**, *267*, 1837-1846.
109. Long, X.; Deng, S.; Mattner, J.; Zang, Z.; Zhou, D.; McNary, N.; Goff, R. D.; Teyton, L.; Bendelac, A.; Savage, P. B. Synthesis and evaluation of stimulatory properties of Sphingomonadaceae glycolipids. *Nat. Chem. Biol.* **2007**, *3*, 559-564.
110. Angelakis, E.; Roux, V.; Raoult, D. Sphingomonas mucosissima Bacteremia in patient with sickle cell disease. *Emerg. Infect. Dis.* **2009**, *15*, 133-134.
111. Arora, P.; Porcelli, S. A. A glycan shield for bacterial sphingolipids. *Chem. Biol.* **2008**, *15*, 642-644.
112. McCarthy, C.; Shepherd, D.; Fleire, S.; Stronge, V. S.; Koch, M.; Illarionov, P. A.; Bossi, G.; Salio, M.; Denkberg, G.; Reddington, F.; Tarlton, A.; Reddy, B. G.; Schmidt, R. R.; Reiter, Y.; Griffiths, G. M.; van der Merwe, P. A.; Besra, G. S.; Jones, E. Y.; Batista, F. D.; Cerundolo, V. The length of lipids bound to human CD1d molecules modulates the affinity of NKT cell TCR and the threshold of NKT cell activation. *J. Exp. Med.* **2007**, *204*, 1131-1144.
113. Miyamoto, K.; Miyake, S.; Yamamura, T. A synthetic glycolipid prevents autoimmune encephalomyelitis by inducing TH2 bias of natural killer T cells. *Nature* **2001**, *413*, 531-534.

114. Oki, S.; Chiba, A.; Yamamura, T.; Miyake, S. The clinical implication and molecular mechanism of preferential IL-4 production by modified glycolipid-stimulated NKT cells. *J. Clin. Invest.* **2004**, *113*, 1631-1640.
115. Goff, R. D.; Gao, Y.; Mattner, J.; Zhou, D.; Yin, N.; Cantu, C., III; Teyton, L.; Bendelac, A.; Savage, P. B. Effects of Lipid Chain Lengths in alpha -Galactosylceramides on Cytokine Release by Natural Killer T Cells. *J. Am. Chem. Soc.* **2004**, *126*, 13602-13603.
116. Apostolopoulos, V.; Yu, M.; Corper, A. L.; Li, W.; McKenzie, I. F. C.; Teyton, L.; Wilson, I. A. Crystal structure of a non-canonical high affinity peptide complexed with MHC class I: a novel use of alternative anchors. *J. Mol. Biol.* **2002**, *318*, 1307-1316.
117. Sidobre, S.; Naidenko, O. V.; Sim, B.-C.; Gascoigne, N. R. J.; Garcia, K. C.; Kronenberg, M. The Valpha 14 NKT cell TCR exhibits high-affinity binding to a glycolipid/CD1d complex. *J. Immunol.* **2002**, *169*, 1340-1348.
118. Schmieg, J.; Yang, G.; Franck, R. W.; Tsuji, M. Superior protection against malaria and melanoma metastases by a C-glycoside analogue of the natural killer T cell ligand alpha -galactosylceramide. *J. Exp. Med.* **2003**, *198*, 1631-1641.
119. Blauvelt, M. L.; Khalili, M.; Jaung, W.; Paulsen, J.; Anderson, A. C.; Wilson, S. B.; Howell, A. R. alpha -S-GalCer: Synthesis and evaluation for iNKT cell stimulation. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6374-6376.
120. Iijima, H.; Kimura, K.; Sakai, T.; Uchimura, A.; Shimizu, T.; Ueno, H.; Natori, T.; Koezuka, Y. Structure-activity relationship and conformational analysis of monoglycosylceramides on the syngeneic mixed leukocyte reaction. *Bioorg. Med. Chem.* **1998**, *6*, 1905-1910.
121. Zhang, W.; Zheng, X.; Xia, C.; Perali, R. S.; Yao, Q.; Liu, Y.; Zheng, P.; Wang, P. G. alpha -Lactosylceramide as a novel "sugar-capped" CD1d ligand for natural killer T cells: biased cytokine profile and therapeutic activities. *ChemBioChem* **2008**, *9*, 1423-1430.
122. Zhou, X.-T.; Forestier, C.; Goff, R. D.; Li, C.; Teyton, L.; Bendelac, A.; Savage, P. B. Synthesis and NKT Cell Stimulating Properties of Fluorophore- and Biotin-Appended 6"-Amino-6"-deoxy-galactosylceramides. *Org. Lett.* **2002**, *4*, 1267-1270.
123. Wu, D.; Xing, G.-W.; Poles, M. A.; Horowitz, A.; Kinjo, Y.; Sullivan, B.; Bodmer-Narkevitch, V.; Plettenburg, O.; Kronenberg, M.; Tsuji, M.; Ho, D. D.; Wong, C.-H. Bacterial glycolipids and analogs as antigens for CD1d-restricted NKT cells. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 1351-1356.
124. Savage, P. B.; Teyton, L.; Bendelac, A. Glycolipids for natural killer T cells. *Chem. Soc. Rev.* **2006**, *35*, 771-779.
125. Xing, G.-W.; Wu, D.; Poles, M. A.; Horowitz, A.; Tsuji, M.; Ho, D. D.; Wong, C.-H. Synthesis and human NKT cell stimulating properties of 3-O-sulfo-alpha /beta -galactosylceramides. *Bioorg. Med. Chem.* **2005**, *13*, 2907-2916.
126. Silk, J. D.; Salio, M.; Reddy, B. G.; Shepherd, D.; Gileadi, U.; Brown, J.; Masri, S. H.; Polzella, P.; Ritter, G.; Besra, G. S.; Jones, E. Y.; Schmidt, R. R.; Cerundolo, V. Cutting Edge: Nonglycosidic CD1d Lipid Ligands Activate Human and Murine Invariant NKT Cells. *J. Immunol.* **2008**, *180*, 6452-6456.
127. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a

computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535-542.
128. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. \"AMBER\", a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to stimulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1-42.
129. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.
130. Pellegrini, E.; Field, M. J. A Generalized-Born Solvation Model for Macromolecular Hybrid-Potential Calculations. *J. Phys. Chem. A* **2002**, *106*, 1316-1326.
131. Case, D. A.; Darden, T. A.; Cheatham, I., T.E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A., In *AMBER 10*. Eds.; University of California, San Francisco: San Francisco, CA, 2008; pp 1-306.
132. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999-2012.
133. Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. Molecular Mechanical and Molecular Dynamic Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM_93 Parameter Development. *J. Phys. Chem.* **1995**, *99*, 3832-3846.
134. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247-260.
135. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
136. Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer. An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714-2723.
137. *MacroModel*, version 8.5; Schrödinger Inc.: Portland, OR, 2003.
138. Polak, E.; Ribiere, G. Note sur la Convergence de Methodes de Directions Conjuguees. *Revue Francaise Inf. Rech. Oper., Serie Rouge* **1969**, *16*, 35-43.
139. Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657-1666.
140. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.
141. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639-1662.

142. Silverman, R. B., Drug Discovery, Design, and Development. In *The Organic Chemistry of Drug Design and Drug Action*. 2nd ed.; Hayhurst, J. Eds.; Elsevier: Amsterdam, The Netherlands, 2004; pp 8-120.
143. Cantu, C., III; Benlagha, K.; Savage, P. B.; Bendelac, A.; Teyton, L. The Paradox of Immune Molecular Recognition of a-Galactosylceramide: Low Affinity, Low Specificity for CD1d, High Affinity for ab TCRs. *J. Immunol.* **2003**, *170*, 4673-4682.
144. van Rhijn, I.; Young, D. C.; Jin, S. I.; Levery, S. B.; Illarionov, P. A.; Besra, G. S.; Porcelli, S. A.; Gumperz, J.; Cheng, T.-Y.; Moody, D. B. CD1d-restricted T cell activation by nonlipidic small molecules. *Proc. Nat. Acad. Sci. U.S.A.* **2004**, *101*, 13578-13583.
145. Young, D. C.; Moody, D. B. T-cell recognition of glycolipids presented by CD1 proteins. *Glycobiology* **2006**, *16*, 103R-112R.
146. Irwin John, J.; Shoichet Brian, K. ZINC--a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177-182.
147. Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 601-619.

Appendix A: The List of Hydrogen Bonding Data for the Tertiary Complexes

Permanence of hydrogen bonds between key residues' atoms throughout the 10 ns simulation where % occupy refers to what portion of the simulation that hydrogen bond existed. Where [blank space] 0-5% [.] 5-20% [-] 20-40% [o] 40-60% [x] 60-80% [*] 80-95% and [@] 95-100%.

Table 13. Full Complex Simulation Hydrogen Bonds.

DONOR		ACCEPTOR			STASTICAL DATA							
:res@atom		:res@Hatom	:res@atom		%occupied	distance	std	angle	std	lifetime	std	maxocc
:75@OD2	**Asp80**	:822@HO4	:822@O4	**AGH_O4**	96.79	2.69	0.11	16.19	8.24	56.7	67.2	612
:75@OD2	**Asp80**	:822@HO3	:822@O3	**AGH_O3**	92.27	2.751	0.11	16.31	9.07	14.2	19.4	240
:467@OD1	**Asp94**	:74@HH12	:74@NH1	**Arg79**	90.7	2.798	0.09	20.54	9.79	13.9	16.6	180
:146@OD1	**Asp151**	:149@HG1	:149@OG1	**Thr154**	85.46	2.677	0.11	15.25	8.38	73.8	98.9	999
:75@OD1	**Asp80**	:468@HH11	:468@NH1	**Arg95**	84.35	2.822	0.09	22.15	-11.31	7.1	7.3	79
:467@OD2	**Asp94**	:74@HH22	:74@NH2	**Arg79**	83.5	2.828	0.09	20.19	-10.27	7.2	9.2	111
:146@OD2	**Asp151**	:823@H3O	:823@O3	**1LA_O3**	82.78	2.724	0.12	17.11	-10.1	14.8	24.3	377
:78@OE2	**Glu83**	:474@HH21	:474@NH2	**Arg103**	60.02	2.82	0.09	21.4	-11.15	5.5	7.6	111
:78@OE1	**Glu83**	:626@HH	:626@OH	**Tyr50**	57.09	2.737	0.12	15.93	8.92	8.8	13.3	181
:78@OE1	**Glu83**	:624@HH	:624@OH	**Tyr48**	54.79	2.79	0.11	23.9	-11.82	4.6	5.9	101
:78@OE1	**Glu83**	:474@HE	:474@NE	**Arg103**	30.07	2.863	0.08	19.64	9.12	2.7	2.9	45
:78@OE2	**Glu83**	:81@HZ1	:81@NZ	**Lys86**	29.63	2.782	0.09	22.13	-11.28	14.9	21.3	211
:146@OD1	**Asp151**	:823@H2O	:823@O2	**1LA_O2**	24.25	2.854	0.1	37.17	-12.8	1.5	1	13
:467@OD1	**Asp94**	:74@HH22	:74@NH2	**Arg79**	20.19	2.881	0.08	35.85	9.59	1.7	1.4	31
:78@OE2	**Glu83**	:81@HZ2	:81@NZ	**Lys86**	15.98	2.792	0.09	21.59	-11.38	10.5	14.3	147
:78@OE2	**Glu83**	:81@HZ3	:81@NZ	**Lys86**	13.94	2.789	0.09	21.47	-11.34	11.8	14.8	151
:146@OD1	**Asp151**	:823@H3O	:823@O3	**1LA_O3**	13.53	2.761	0.12	25.87	-12.36	5.8	9.9	119
:146@OD2	**Asp151**	:149@HG1	:149@OG1	**Thr154**	12.66	2.673	0.11	17.97	8.45	66.7	79.3	352
:78@OE2	**Glu83**	:624@HH	:624@OH	**Tyr50**	12.02	2.808	0.12	40.02	-13.84	2.2	3.3	80
:78@OE1	**Glu83**	:474@HH21	:474@NH2	**Arg103**	8.39	2.882	0.08	30.18	-11.14	1.6	1.3	17
:823@O3	**1LA_O3**	:404@HG	:404@OG	**Ser30**	6.79	2.801	0.11	24.66	-11.92	5.9	8.2	83
:822@O1A	**AGH_O1A**	:823@H2O	:823@O2	**1LA_O2**	6.7	2.758	0.15	54.96	4.76	1.4	1.8	41
:823@O4	**1LA_O4**	:404@HG	:404@OG	**Ser30**	5.87	2.794	0.11	24.13	-12.33	4.3	7.3	72

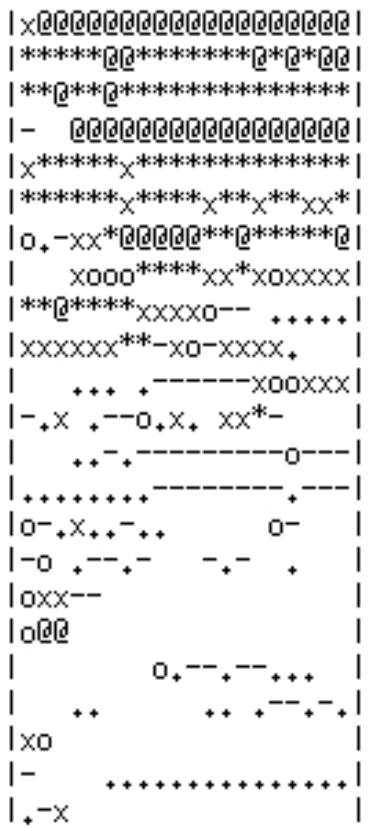


Table 14. Truncated Complex Simulation Hydrogen Bonds.

DONOR		ACCEPTOR			STASTICAL DATA							
:res@atom		**:res@Hatom**	**:res@atom**		**%occupied**	**distance**	**std**	**angle**	**std**	**lifetime**	**std**	**maxocc**
:75@OD2	**Asp80**	:404@HO4	:404@O4	**AGH_O4**	97.84	2.692	0.11	15.41	8.4	48.6	83.3	948
:75@OD2	**Asp80**	:404@HO3	:404@O3	**AGH_O3**	93.63	2.737	0.11	14.81	8.43	17	23.9	306
:146@OD2	**Asp151**	:405@H3O	:405@O3	**1LA_H3**	93.27	2.716	0.12	15.94	8.8	18.9	29.1	328
:146@OD1	**Asp151**	:149@HG1	:149@OG1	**Thr154**	83.53	2.694	0.11	16.36	8.9	36.7	47.3	700
:404@OAA	**AGH_OAA**	:71@HG	:71@OG	**Ser76**	71.84	2.725	0.12	16.46	9.24	11.9	19.9	403
:75@OD1	**Asp80**	:271@HH11	:271@NH1	**Arg95**	49.78	2.837	0.09	24.28	-11.91	4.1	5.6	94
:78@OE1	**Glu83**	:277@HH21	:277@NH2	**Arg103**	37.35	2.809	0.09	23.43	-11.69	6.5	11.6	208
:75@OD2	**Asp80**	:271@HH11	:271@NH1	**Arg95**	35.43	2.855	0.09	24.71	-11.73	3	3.8	56
:270@OD2	**Asp94**	:74@HH22	:74@NH2	**Arg79**	33.27	2.78	0.09	23.88	-10.41	14	27.3	306
:270@OD1	**Asp94**	:74@HH12	:74@NH1	**Arg79**	30.35	2.847	0.09	22.44	-11.47	3.3	4.5	82
:78@OE2	**Glu83**	:277@HH21	:277@NH2	**Arg103**	20.38	2.827	0.1	26.53	-12.2	3.9	7.3	183
:78@OE2	**Glu83**	:277@HE	:277@NE	**Arg103**	19.9	2.846	0.09	23.54	-10.67	3.8	5	95
:404@O1A	**AGH_OAA**	:405@H2O	:405@O2	**1LA_H2**	19.51	2.646	0.16	52.18	6.12	2.5	4.3	85
:270@OD2	**Asp94**	:74@HH12	:74@NH1	**Arg79**	19.06	2.849	0.09	30.68	-12.08	2.7	4	69
:404@O1A	**AGH_OAA**	:149@HG1	:149@OG1	**Thr154**	13.72	2.647	0.1	16.38	8.28	182.3	275	1313
:405@O3	**1LA_O3**	:207@HG	:207@OG	**Ser30**	11.9	2.805	0.11	27.92	-12.97	5.4	8	140
:78@OE1	**Glu83**	:336@HH	:336@OH	**Tyr48**	10.46	2.779	0.11	25.21	-12.04	6	10.1	176
:405@O4	**1LA_O4**	:207@HG	:207@OG	**Ser30**	10.07	2.797	0.11	25.79	-12.1	4.5	7.2	77
:146@OD1	**Asp151**	:405@H2O	:405@O2	**1LA_H2**	10.05	2.871	0.09	36.9	-12.67	1.3	0.7	13
:78@OE1	**Glu83**	:338@HH	:338@OH	**Tyr50**	9.16	2.725	0.12	19.79	-11.22	12.6	16.2	138
:270@OD1	**Asp94**	:74@HH22	:74@NH2	**Arg79**	9.13	2.811	0.1	28.02	-10.57	4.2	12.2	220
:78@OE2	**Glu83**	:81@HZ3	:81@NZ	**Lys86**	8.1	2.811	0.09	24.35	-12.76	5.7	7.3	68
:78@OE1	**Glu83**	:277@HE	:277@NE	**Arg103**	7.08	2.858	0.09	28.86	-10.8	2.6	3.5	48
:78@OE1	**Glu83**	:81@HZ2	:81@NZ	**Lys86**	5.8	2.811	0.09	24.82	-12.1	6.1	7.4	70
:78@OE1	**Glu83**	:81@HZ3	:81@NZ	**Lys86**	5.59	2.811	0.09	24.61	-12.66	5.5	7.8	69
:78@OE2	**Glu83**	:81@HZ2	:81@NZ	**Lys86**	5.54	2.808	0.09	25.5	-12.61	6.3	9.2	127
:78@OE2	**Glu83**	:81@HZ1	:81@NZ	**Lys86**	5.37	2.811	0.09	26.09	-13.07	5.7	8.8	132
:78@OE1	**Glu83**	:81@HZ1	:81@NZ	**Lys86**	5.06	2.808	0.1	24.06	-12.15	6.1	8.3	82

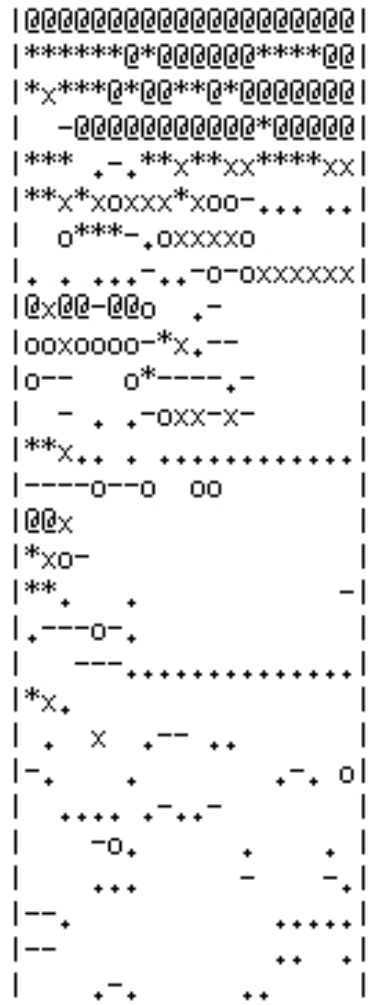


Table 15. Implicit Solvated Complex Simulation Hydrogen Bonds.

DONOR		ACCEPTOR			STASTICAL DATA							
:res@atom		:res@Hatom	:res@atom		%occupied	distance	std	angle	std	lifetime	std	maxocc
:822@O1A	**AGH_O1A**	:823@H2O	:823@O2	**1LA_O2**	80.95	2.544	0.12	45.32	8.69	6.4	9.7	210
:822@O1A	**AGH_O1A**	:468@HH22	:468@NH2	**Arg95**	54.8	2.816	0.09	29.11	-10.92	5.7	8.1	167
:822@O1A	**AGH_O1A**	:468@HH12	:468@NH1	**Arg95**	52.45	2.795	0.09	27.51	8.79	8	13.7	174
:467@OD2	**Asp94**	:474@HH21	:474@NH2	**Arg103**	38.17	2.816	0.09	23.4	-10.94	6.9	9.5	139
:149@OG1	**Thr154**	:468@HH11	:468@NH1	**Arg95**	35.52	2.882	0.08	25.11	-12.21	2.7	2.4	25
:146@OD2	**Asp151**	:149@HG1	:149@OG1	**Thr154**	33.54	2.783	0.12	18.58	-10.29	3.3	3.8	79
:78@OE2	**Glu83**	:474@HH11	:474@NH1	**Arg103**	33.38	2.814	0.09	20.29	9.67	6.8	6.8	54
:75@OD2	**Asp80**	:474@HH22	:474@NH2	**Arg103**	31.07	2.835	0.09	28.07	-11.05	3.5	4.3	62
:78@OE1	**Glu83**	:474@HH11	:474@NH1	**Arg103**	30.31	2.812	0.09	20.35	9.64	6.9	7.4	84
:75@OD1	**Asp80**	:474@HH12	:474@NH1	**Arg103**	28.49	2.834	0.09	26.8	-10.5	3.9	4.1	47
:75@OD1	**Asp80**	:474@HH22	:474@NH2	**Arg103**	27.53	2.836	0.09	28.43	-10.67	3.2	3.3	39
:146@OD1	**Asp151**	:149@HG1	:149@OG1	**Thr154**	24.02	2.767	0.12	18.92	-10.76	4.1	5.4	119
:75@OD2	**Asp80**	:474@HH12	:474@NH1	**Arg103**	22.79	2.835	0.09	28.33	-10.04	3.6	3.6	45
:75@OD1	**Asp80**	:74@HH21	:74@NH2	**Arg79**	22.05	2.849	0.09	25.43	-12.21	2.7	2.9	37
:467@OD1	**Asp94**	:474@HH21	:474@NH2	**Arg103**	21.81	2.81	0.09	25.11	-11.08	7	10.2	101
:467@OD1	**Asp94**	:468@HH21	:468@NH2	**Arg95**	20.99	2.815	0.09	24.44	-10.23	6.7	8.7	97
:75@OD2	**Asp80**	:74@HE	:74@NE	**Arg79**	20.53	2.851	0.09	23.9	-10.96	3	3.7	47
:823@O5	**1LA_O5**	:468@HH22	:468@NH2	**Arg95**	20.1	2.878	0.08	33.78	-11.49	2.2	2	21
:823@O5	**1LA_O5**	:468@HH12	:468@NH1	**Arg95**	18.05	2.848	0.09	33.98	-12.5	3.5	3.9	43
:467@OD2	**Asp94**	:74@HH22	:74@NH2	**Arg79**	17.06	2.837	0.09	23.61	-11.49	3.4	3.7	33
:75@OD2	**Asp80**	:74@HH21	:74@NH2	**Arg79**	16.05	2.857	0.09	30.62	-11.54	2.2	2.2	28
:75@OD2	**Asp80**	:468@HH22	:468@NH2	**Arg95**	13.6	2.831	0.09	32.41	-12.54	3.3	4	53
:467@OD1	**Asp94**	:474@HE	:474@NE	**Arg103**	13.22	2.875	0.08	26	-10.75	1.7	1.2	16
:822@O3	**AGH_O3**	:468@HH11	:468@NH1	**Arg95**	10.61	2.883	0.08	30.83	-13.32	2	1.5	15
:75@OD1	**Asp80**	:474@HH21	:474@NH2	**Arg103**	10.51	2.832	0.09	27.08	-10.71	4.3	4.4	46
:467@OD2	**Asp94**	:74@HH12	:74@NH1	**Arg79**	9.46	2.844	0.09	24.16	-10.58	3	3.2	33
:467@OD2	**Asp94**	:468@HH21	:468@NH2	**Arg95**	9.41	2.853	0.09	25.2	-10.45	2.5	3.1	45

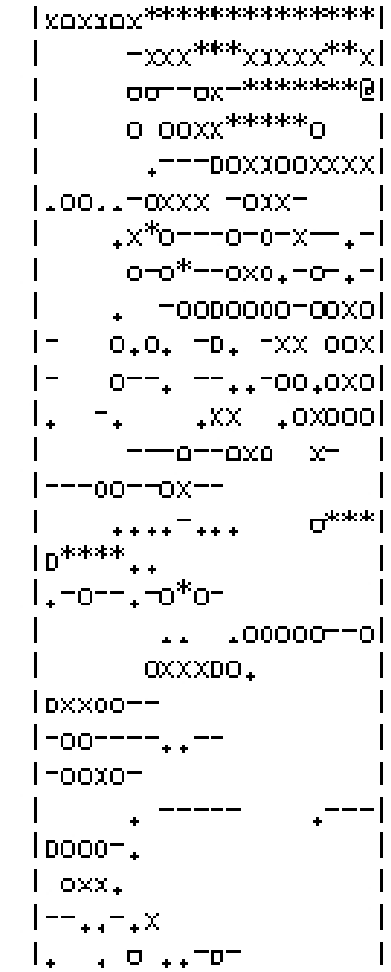
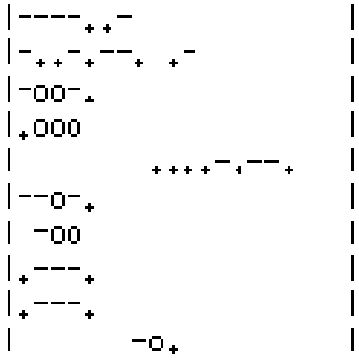


Table 15 continued

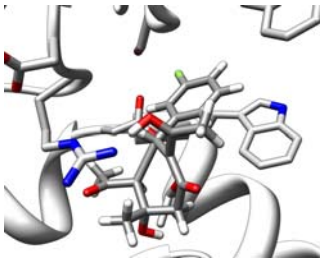
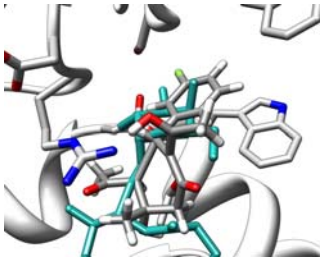
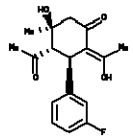
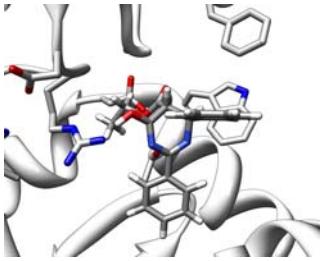
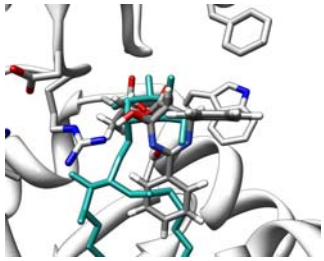
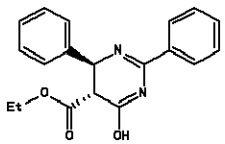
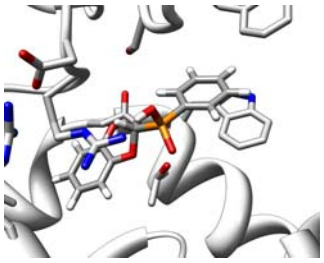
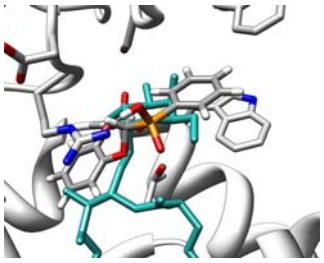
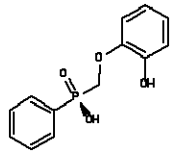
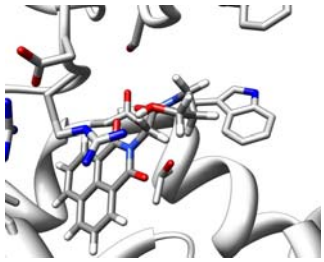
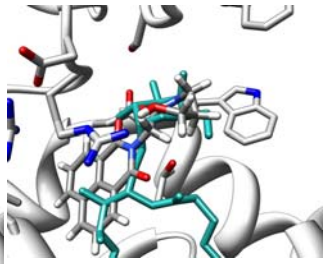
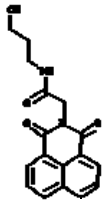
:467@OD1	**Asp94**	:74@HH22	:74@NH2	**Arg79**	9.02	2.851	0.09	29.8	-14.09	2.2	2.2	39
:75@OD1	**Asp80**	:74@HE	:74@NE	**Arg79**	8.98	2.861	0.09	27.75	-10.36	2.2	2.1	20
:467@OD1	**Asp94**	:468@HE	:468@NE	**Arg95**	8.5	2.875	0.08	30.78	8.62	1.9	1.5	13
:78@OE2	**Glu83**	:474@HH22	:474@NH2	**Arg103**	8.38	2.819	0.09	24.9	-10.4	4.2	4.2	33
:467@OD2	**Asp94**	:474@HE	:474@NE	**Arg103**	8.22	2.876	0.08	30.54	9.81	1.5	0.9	9
:78@OE2	**Glu83**	:474@HH12	:474@NH1	**Arg103**	6.88	2.831	0.09	26.58	-10.59	2.9	3.2	36
:75@OD1	**Asp80**	:474@HE	:474@NE	**Arg103**	6.09	2.862	0.09	28.74	9.6	2.3	1.9	18
:78@OE1	**Glu83**	:474@HH22	:474@NH2	**Arg103**	5.88	2.825	0.09	25.24	-10.51	3.7	3.3	23
:78@OE1	**Glu83**	:474@HH12	:474@NH1	**Arg103**	5.3	2.832	0.09	25.85	-10.55	3	3.5	58
:822@O1A	**AGH_O1A**	:74@HH22	:74@NH2	**Arg79**	5.3	2.857	0.09	17.58	9.24	3.2	3.5	49

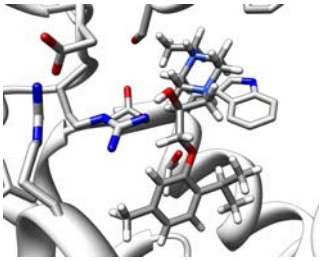
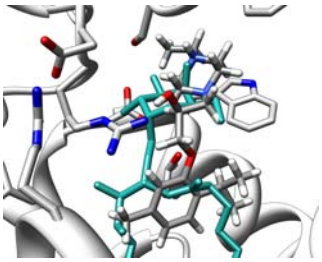
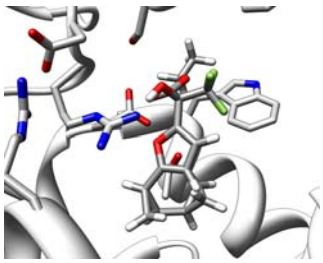
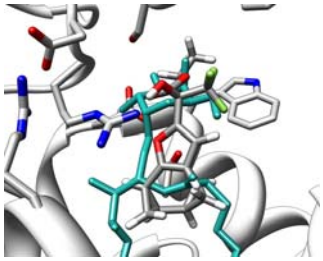
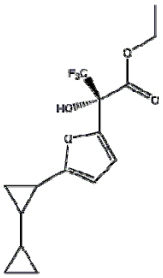
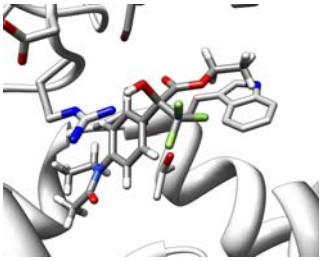
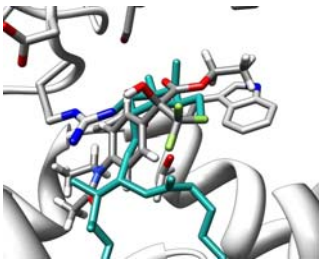
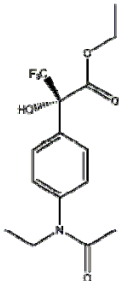
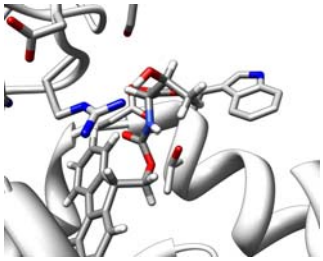
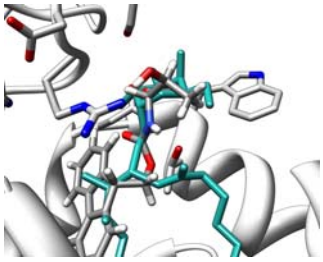
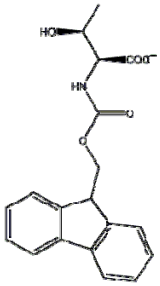
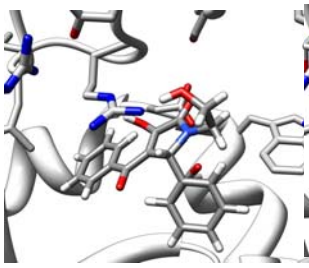
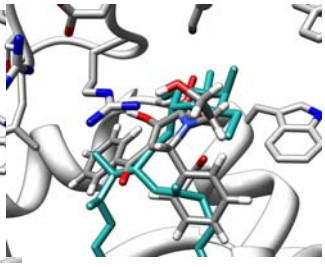
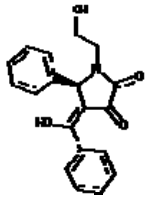


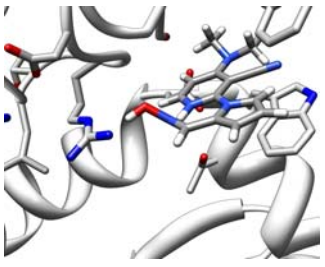
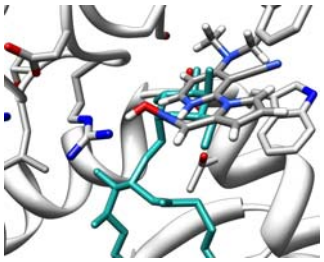
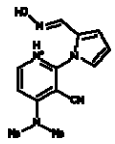
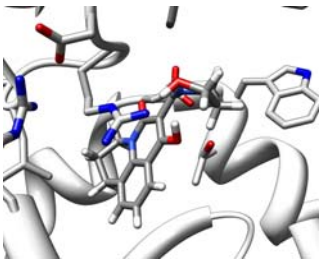
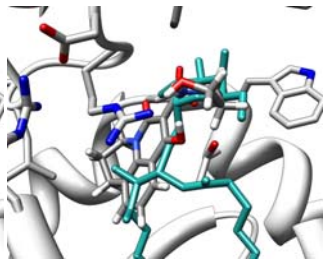
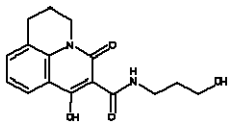
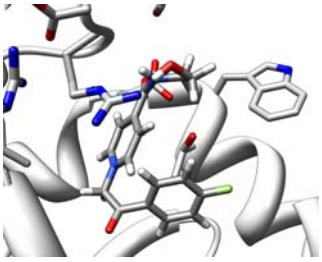
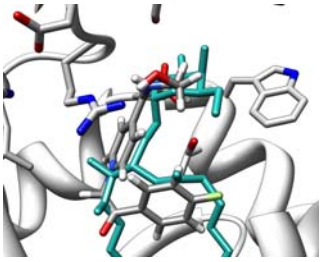
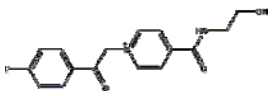
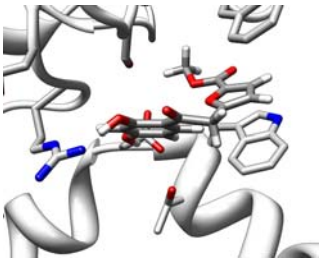
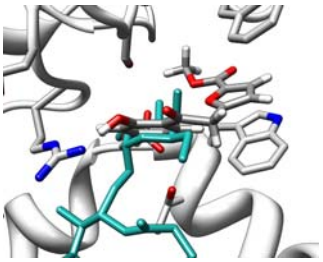
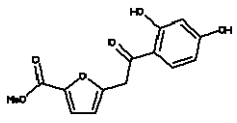
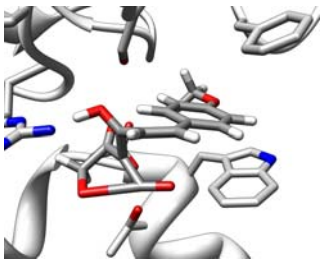
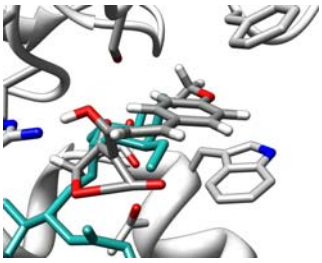
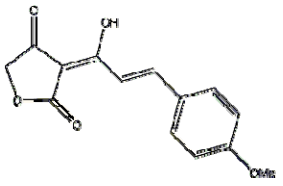
Appendix B: The List of Zinc Compounds from the HTD Virtual Screen

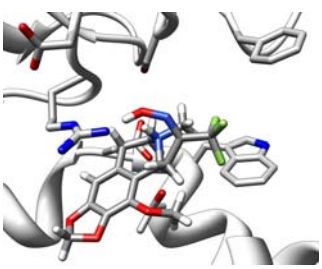
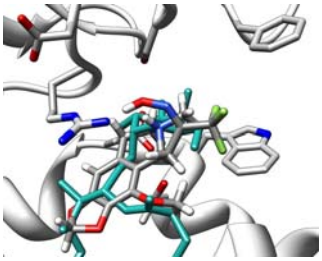
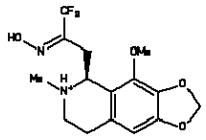
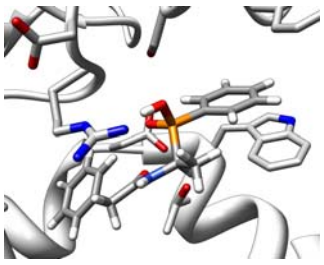
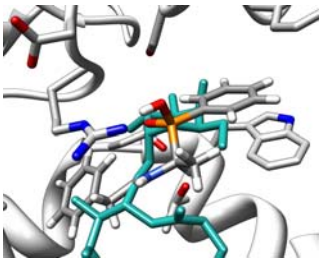
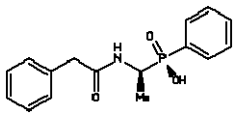
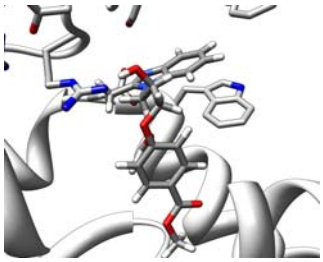
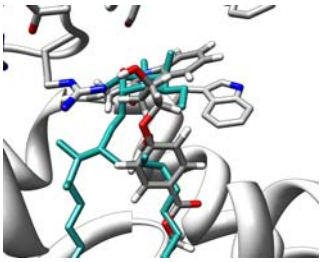
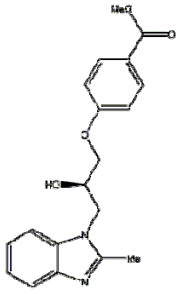
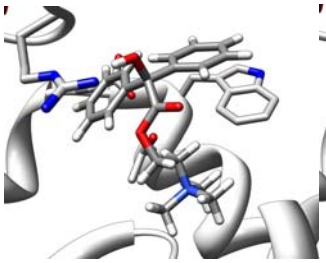
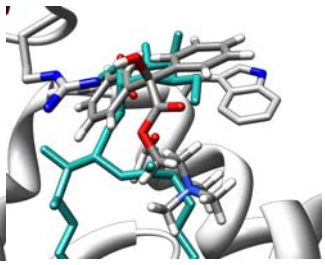
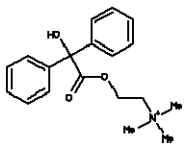
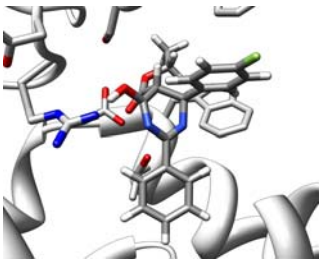
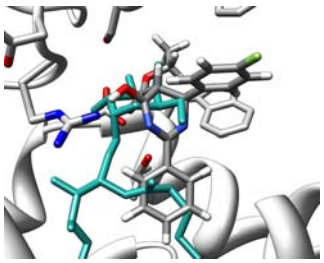
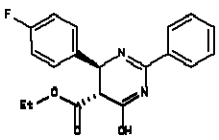
Appendix B is a list of all the best bound lead-like compounds from the Zinc database divided into sections where ZINC1P## represents ~30,000 compounds which are divided into 5 divisions of 5,000 of which the top 500 were scored and the top 25-50 were visually analyzed. The docked energies were not included since they are not necessary to determine whether or not the compound is a suitable replacement to the galactose sugar of α-GalCer. Only the Zinc database code, the 3D image of the compound in the binding site, the image overlaid upon α-GalCer, and the 2D structure are shown. *The red header implied nothing could be found in that series of compounds.*

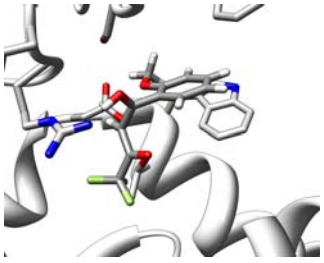
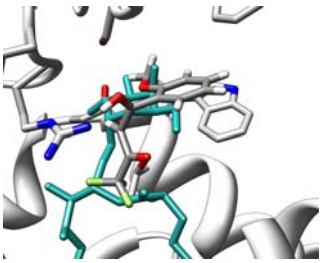
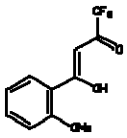
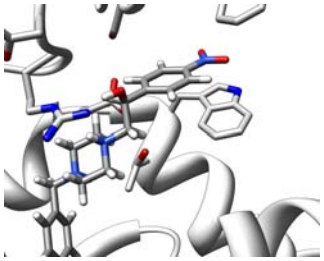
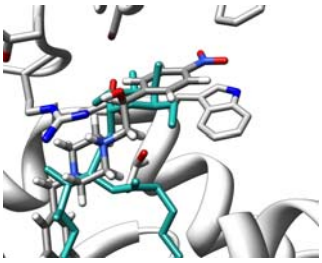
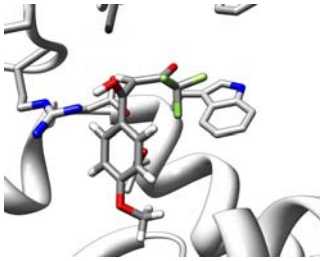
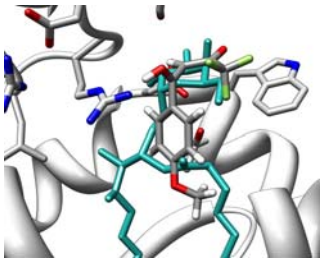
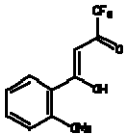
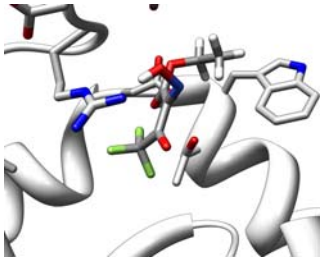
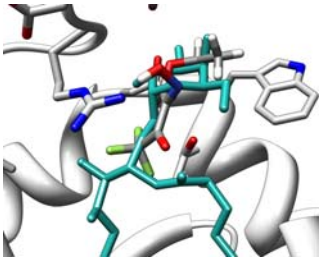
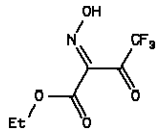
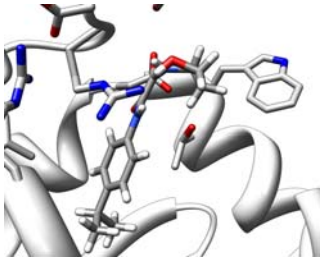
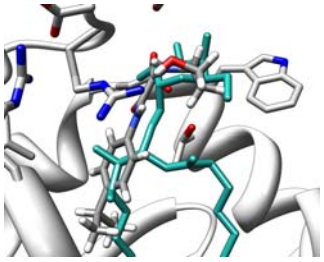
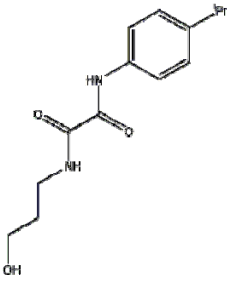
Table 16. List of all the virtual screen ligands chosen from visual analysis.

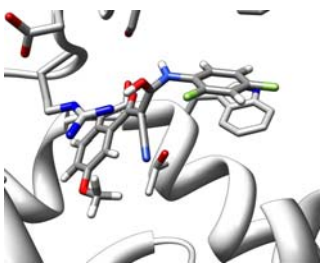
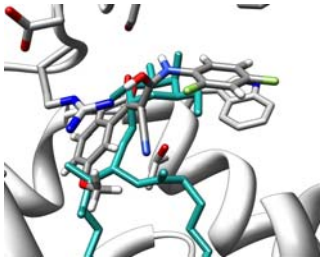
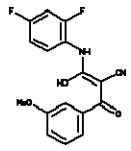
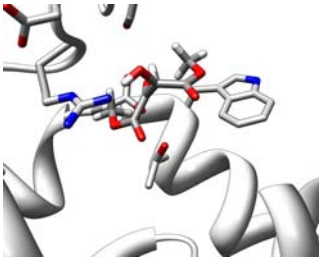
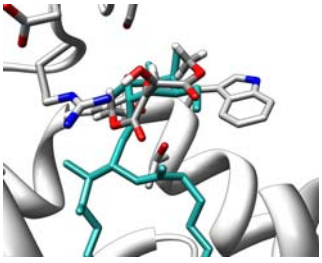
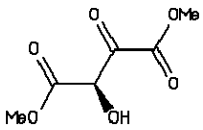
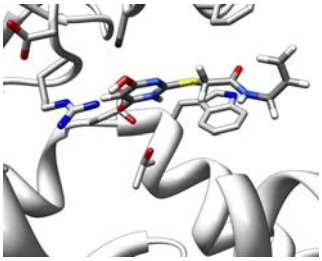
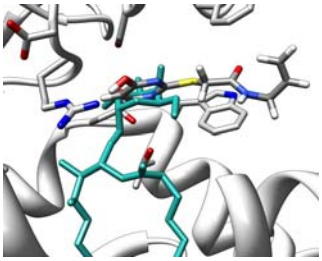
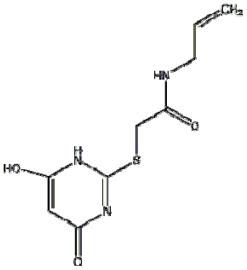
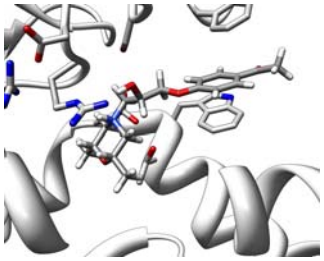
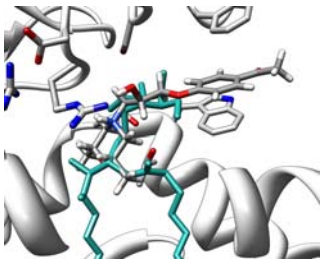
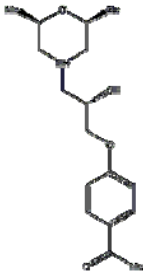
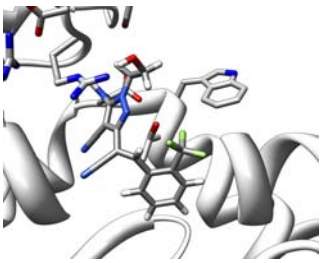
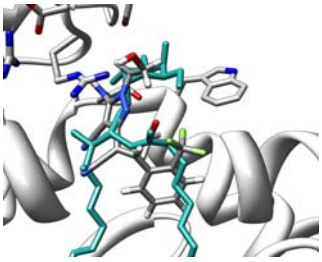
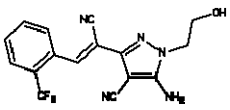
ZINC1P01			
5000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.01197040			
Z.01387480			
Z.00198190			
Z.04006225			

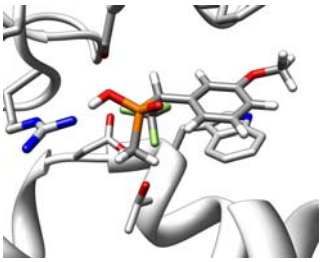
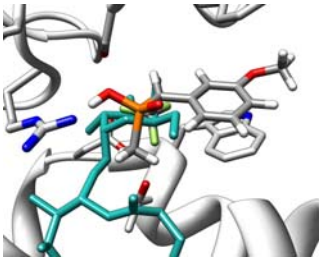
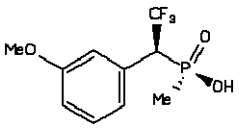
10000.TOP25	IMAGE	with α-GalCer	2D image
Z.00448030			
Z.04006742			
Z.02900767			
Z.01576232			
Z.00435667			

Z.01383318			
Z.02134030			
Z.03130952			
Z.00518876			
15000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.03985043			

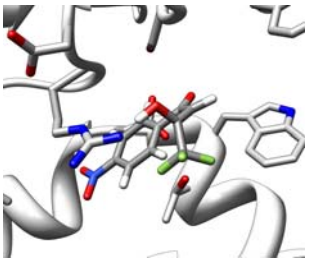
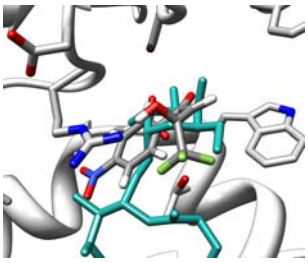
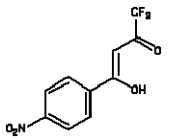
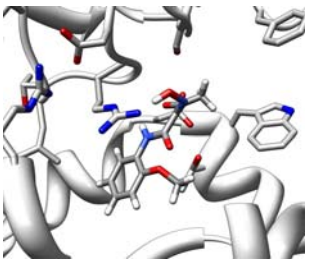
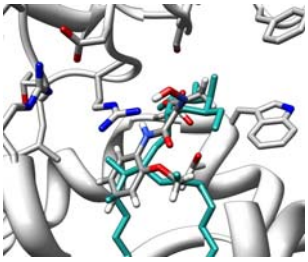
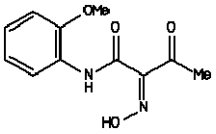
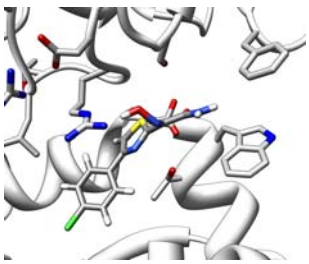
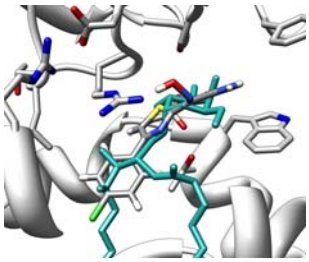
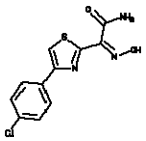
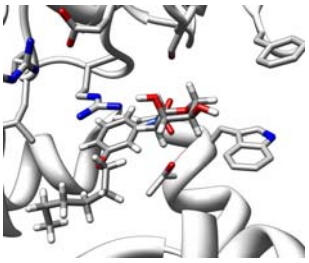
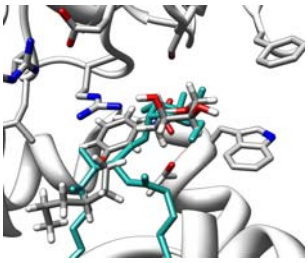
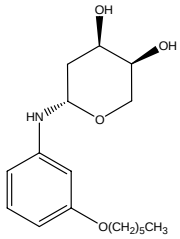
Z.00489920			
Z.03078597			
Z.00243147			
Z.01999358			
20000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.01404437			

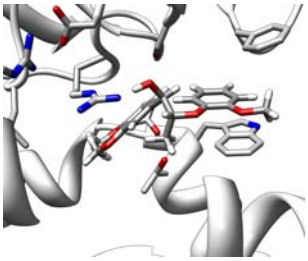
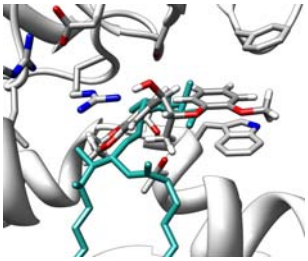
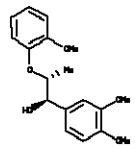
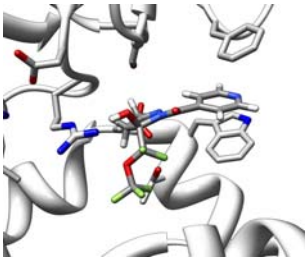
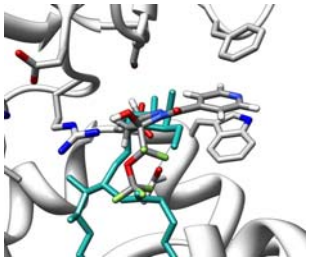
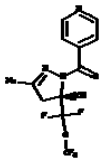
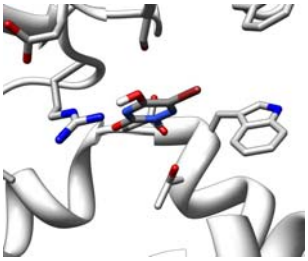
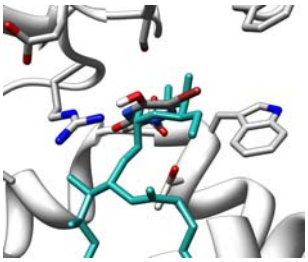
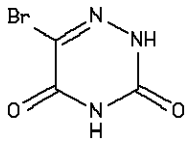
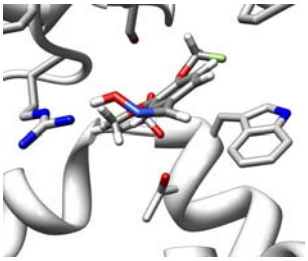
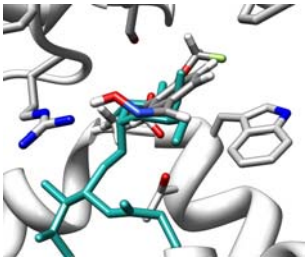
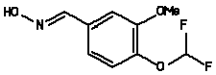
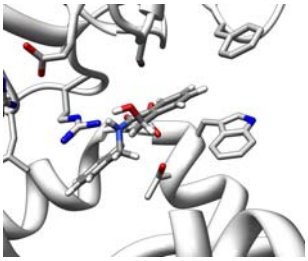
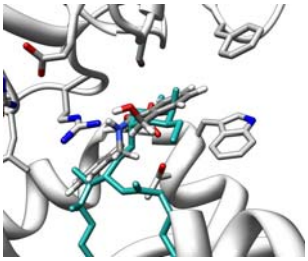
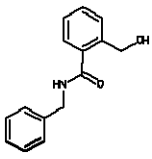
25000.TOP25	IMAGE	with α-GalCer	2D image
Z.04002942			
Z.00446529			
Z.04002941			
Z.02173245			
Z.02025846			

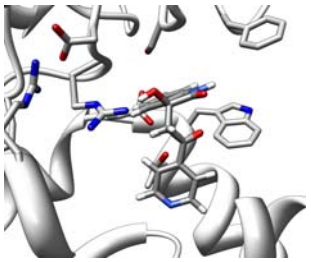
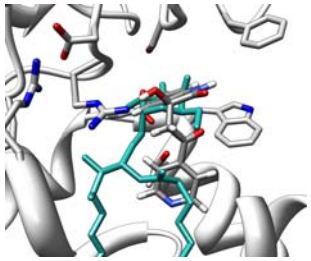
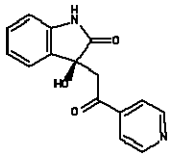
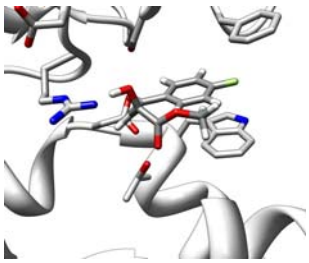
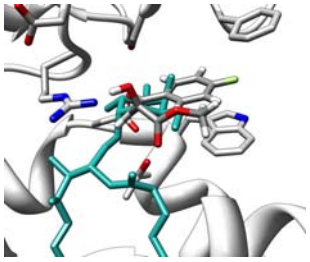
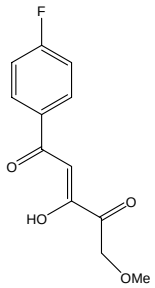
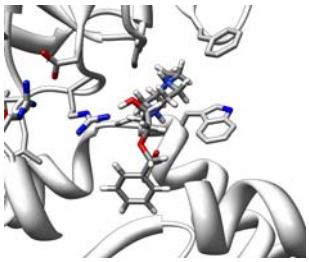
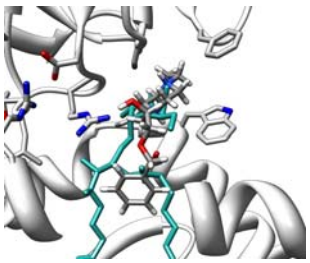
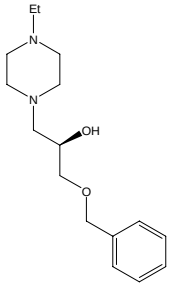
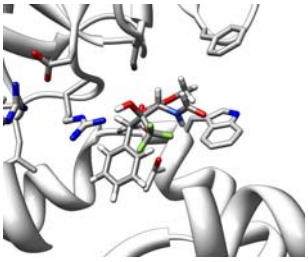
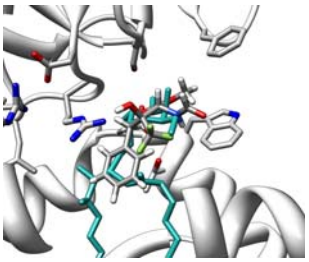
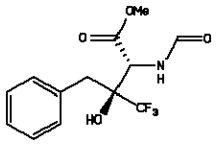
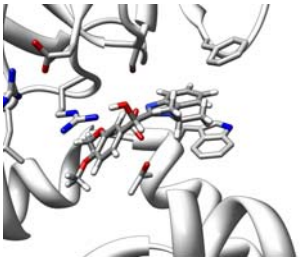
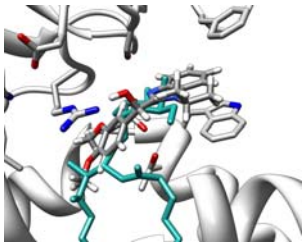
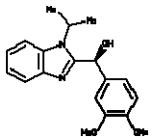
Z.04106687			
Z.04099934			
Z.02482576			
Z.04101501			
27319.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.04118906			

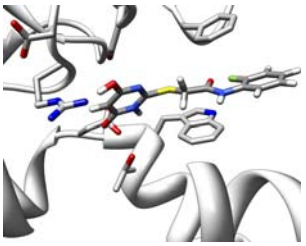
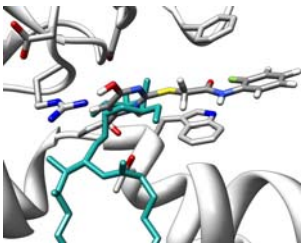
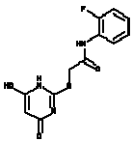
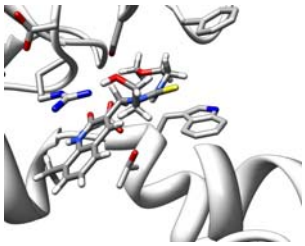
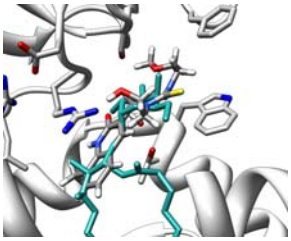
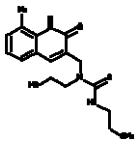
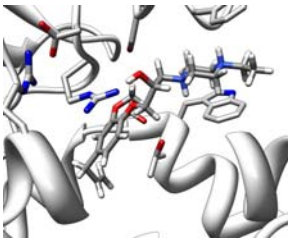
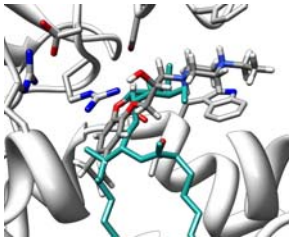
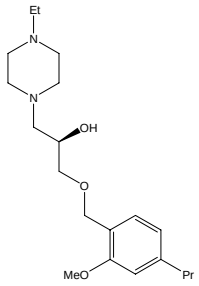
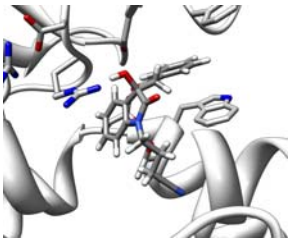
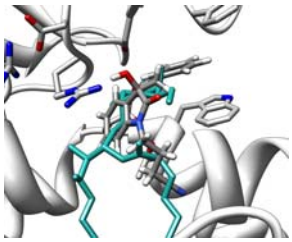
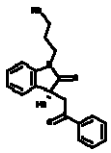
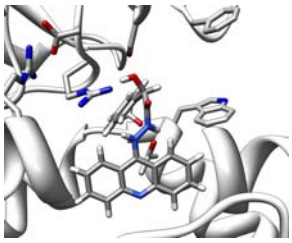
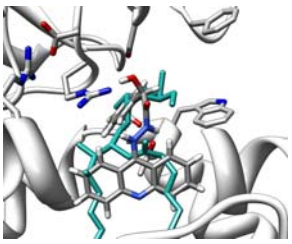
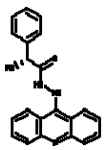
Z.00439979			

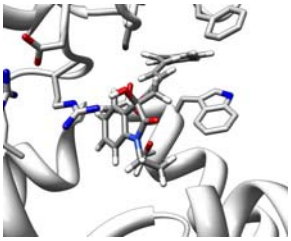
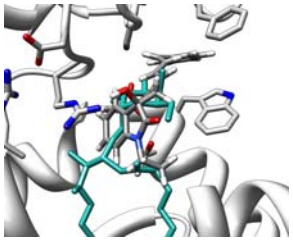
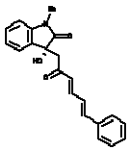
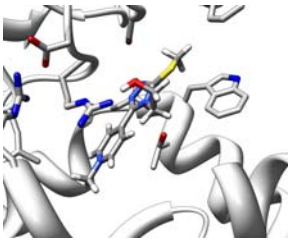
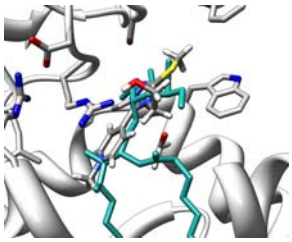
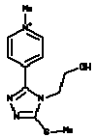
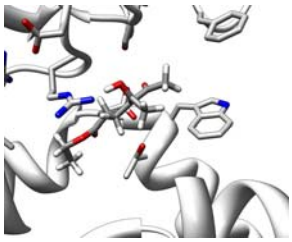
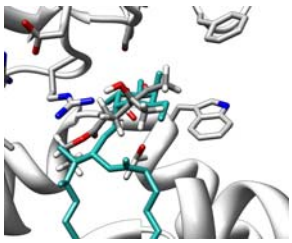
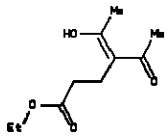
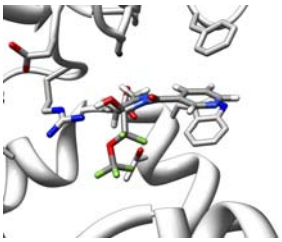
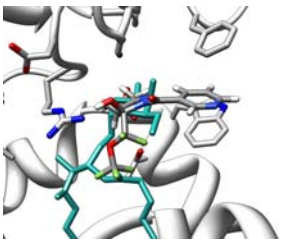
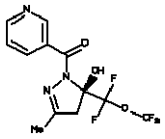
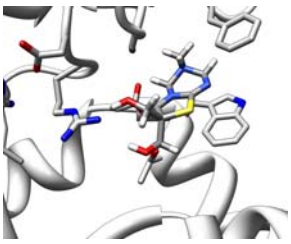
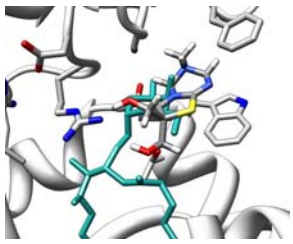
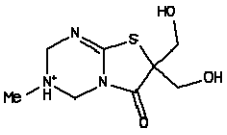
ZINC1P02

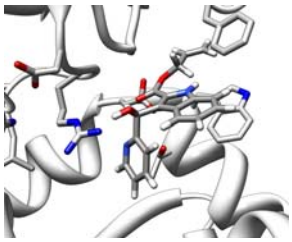
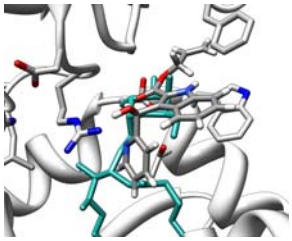
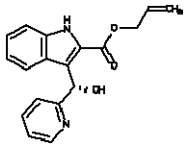
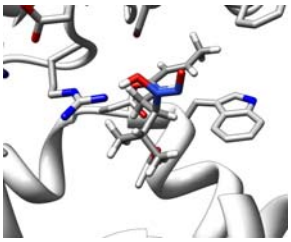
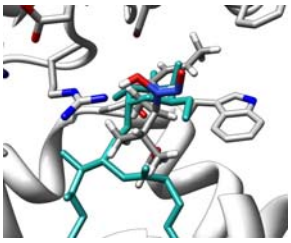
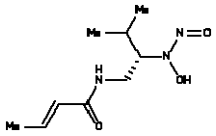
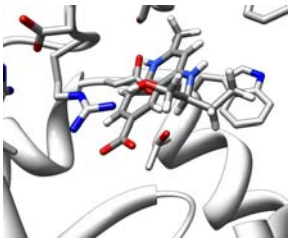
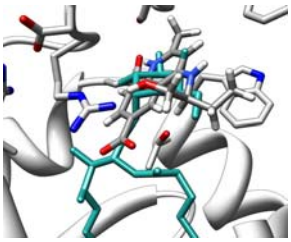
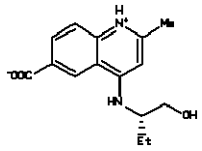
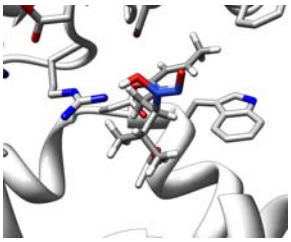
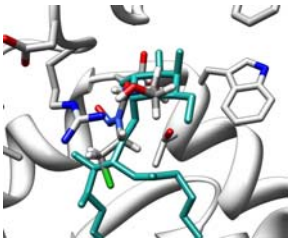
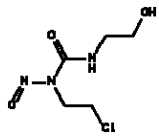
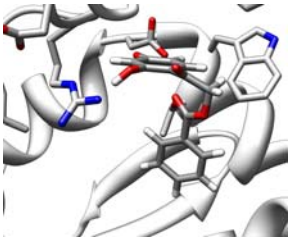
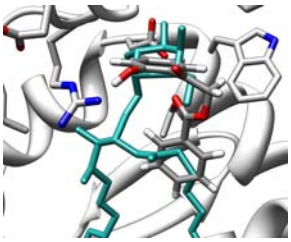
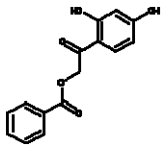
5000.TOP25	IMAGE	with α-GalCer	2D image
Z.02937193			
Z.03317406			
Z.00453325			
Z.04035935			

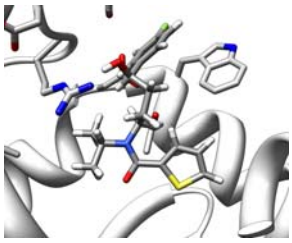
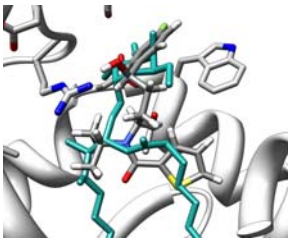
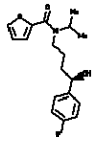
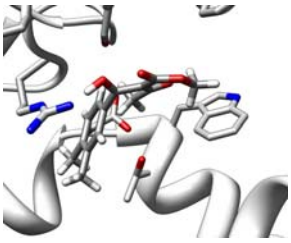
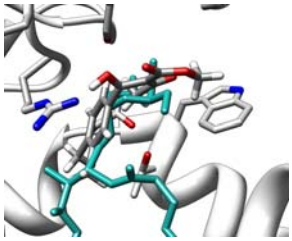
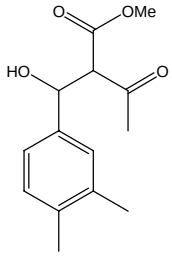
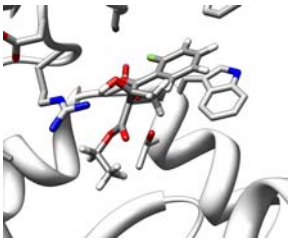
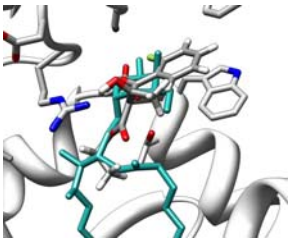
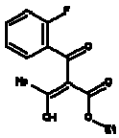
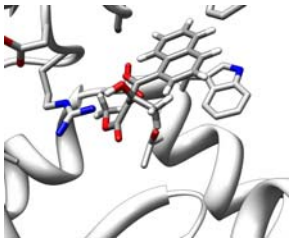
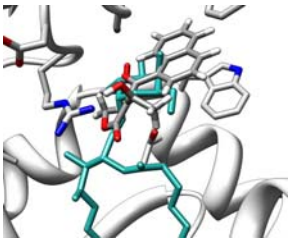
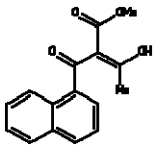
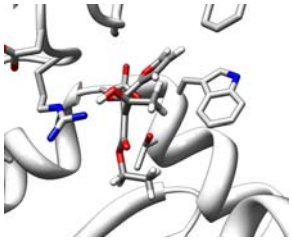
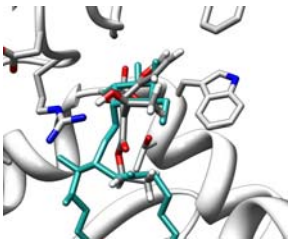
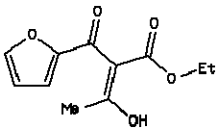
Z.00185130			
Z.00527768			
Z.04130780			
Z.03337366			
Z.00346887			

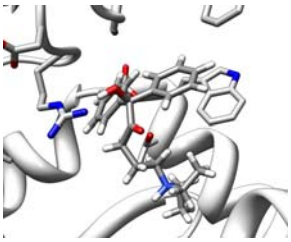
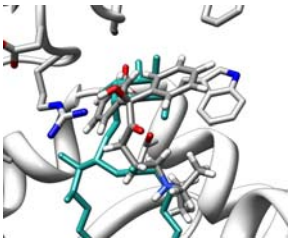
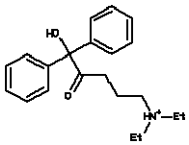
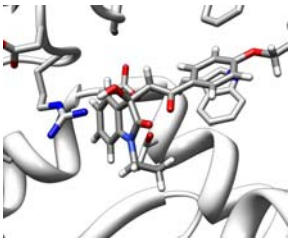
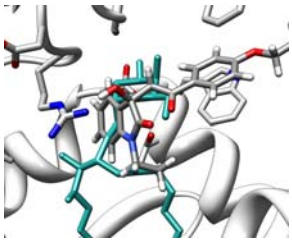
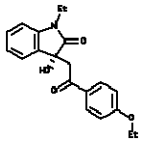
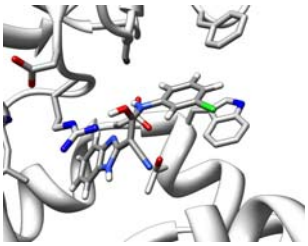
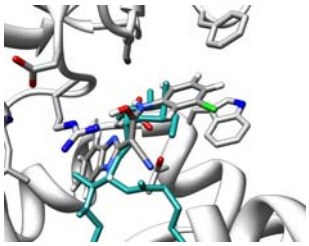
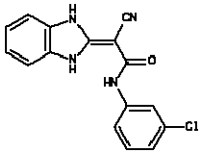
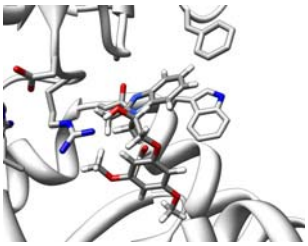
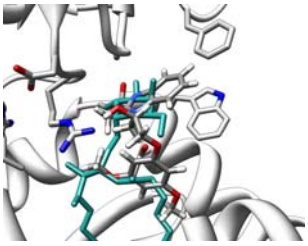
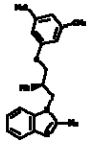
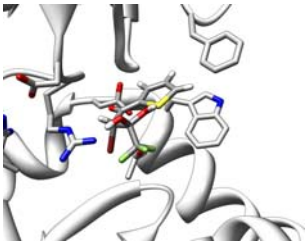
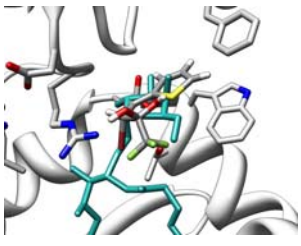
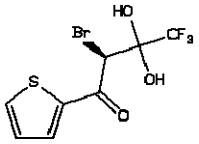
Z.00114593			
10000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.00033442			
Z.04162355			
Z.03102124			
Z.0079103			
15000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**

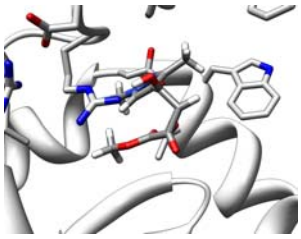
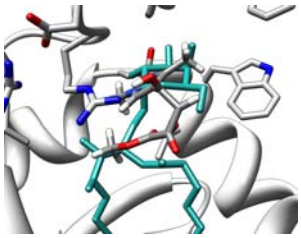
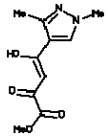
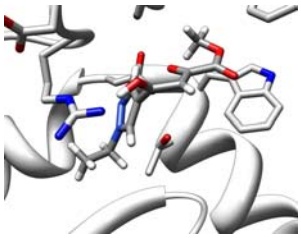
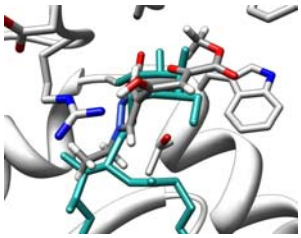
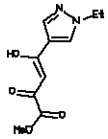
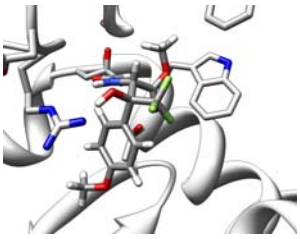
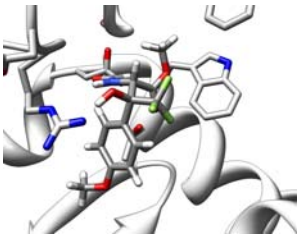
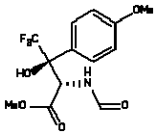
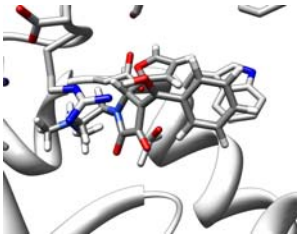
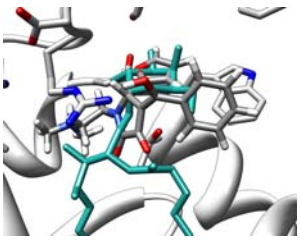
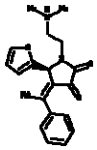
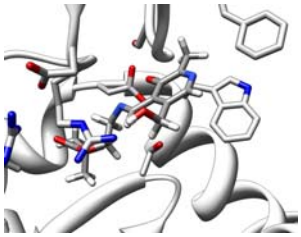
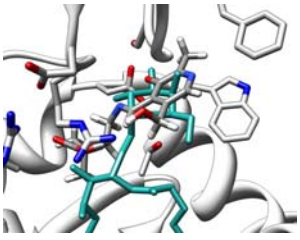
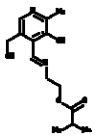
Z.00091155			
Z.04131782			
Z.00448073			
Z.04166007			
Z.04151273			
20000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**

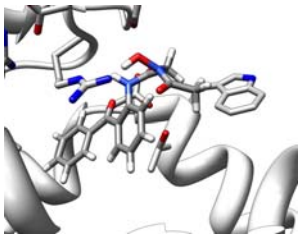
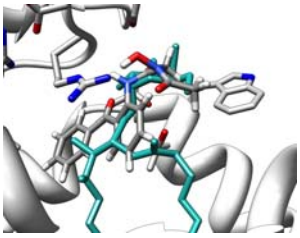
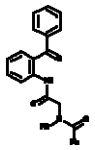
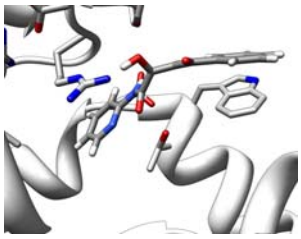
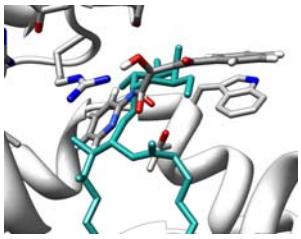
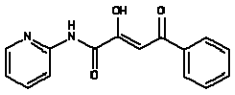
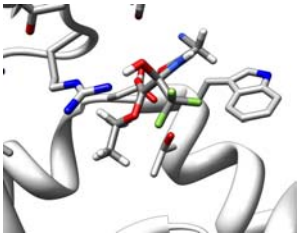
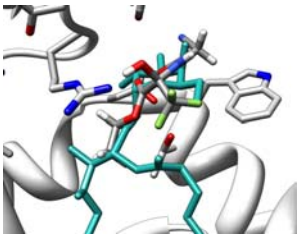
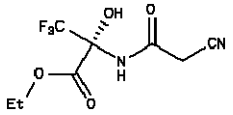
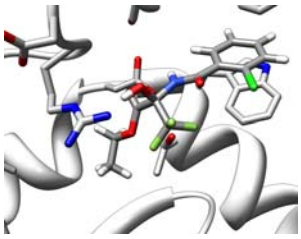
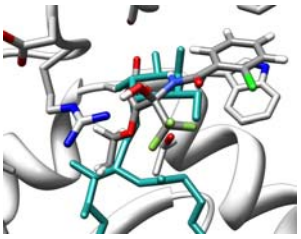
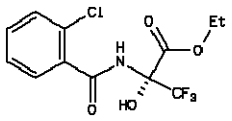
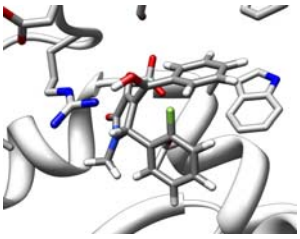
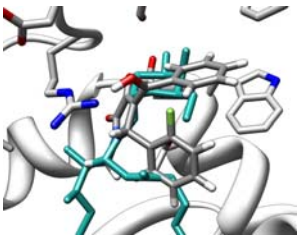
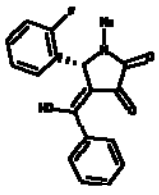
Z.02062685			
Z.00353156			
25000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.02534767			
Z.00531523			
Z.01016482			

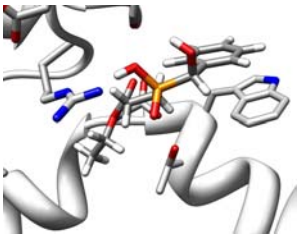
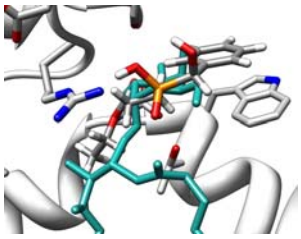
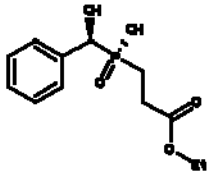
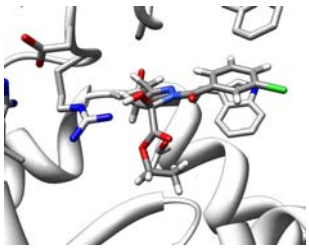
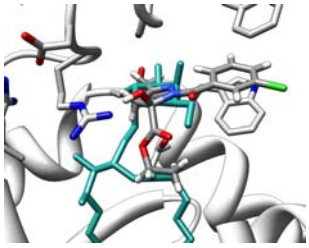
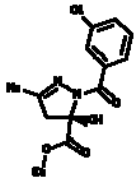
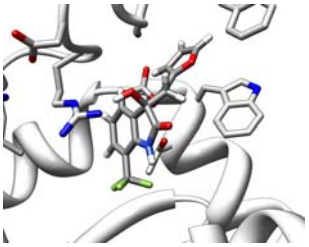
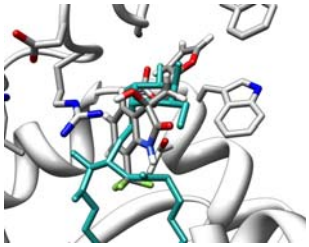
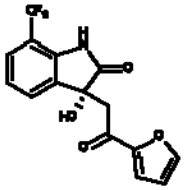
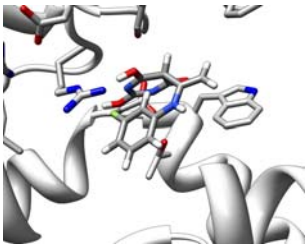
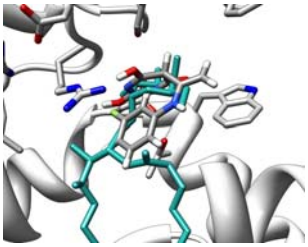
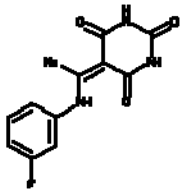
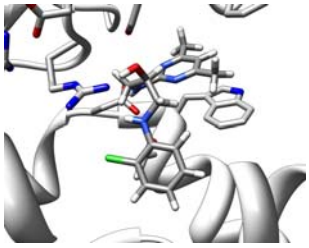
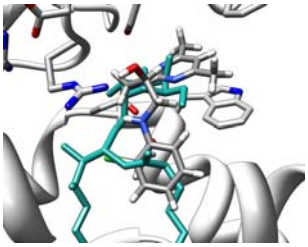
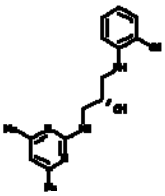
Z.02525396			
27274.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.04215723			
Z.04226485			
Z.04215768			
ZINC1P03			
5000.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
Z.02565998			

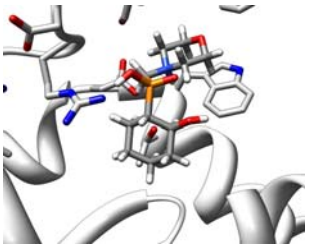
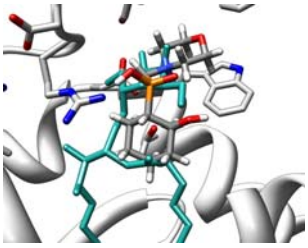
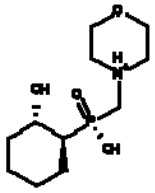
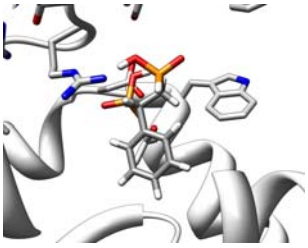
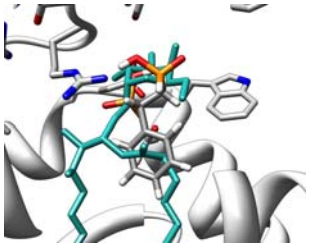
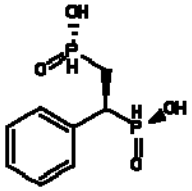
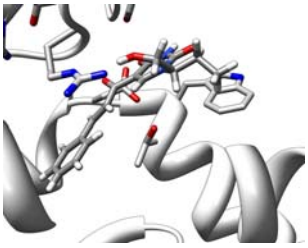
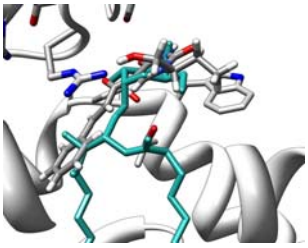
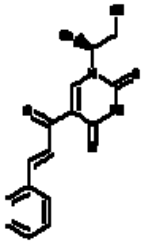
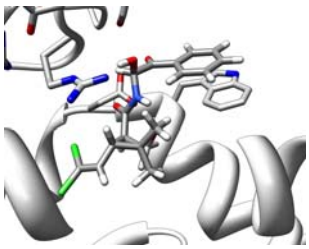
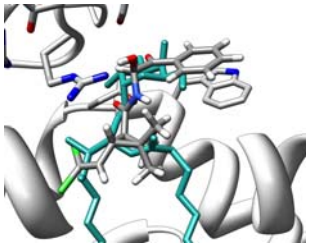
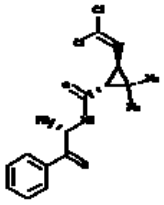
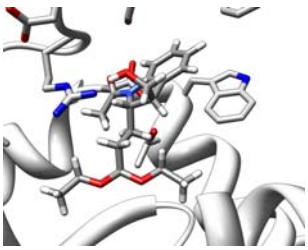
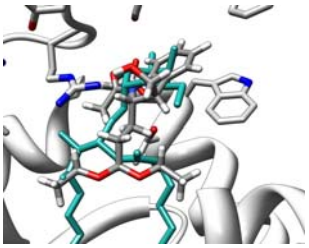
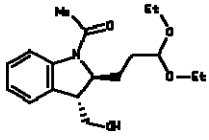
Z.01740832			
10000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.04243061			
Z.04243077			
Z.04243474			
Z.04243139			

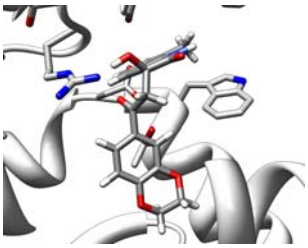
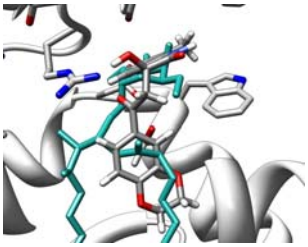
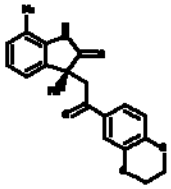
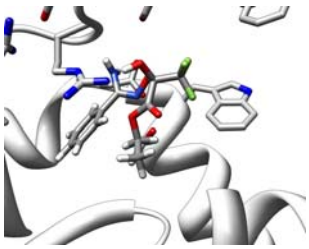
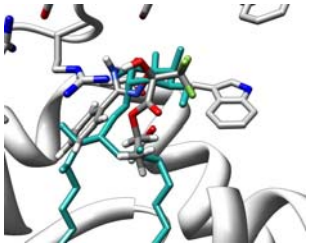
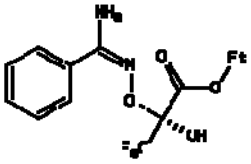
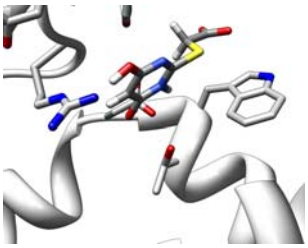
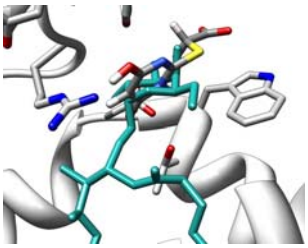
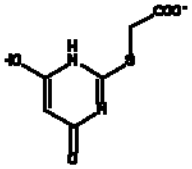
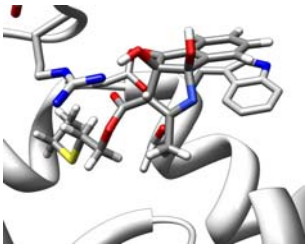
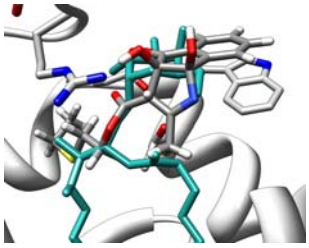
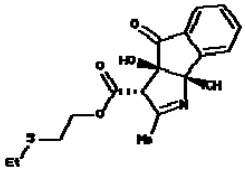
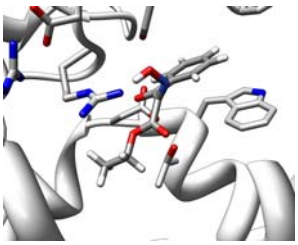
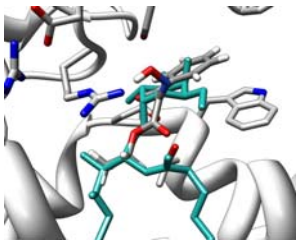
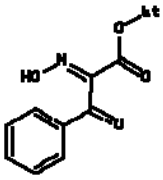
Z.04270091			
Z.00565232			
15000.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
Z.04037113			
Z.00810189			
Z.00159356			
20000.TOP50	**IMAGE**	**with α-GalCer**	**2D image**

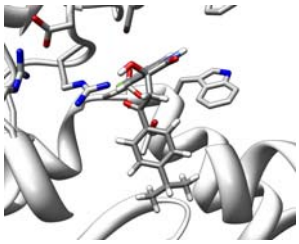
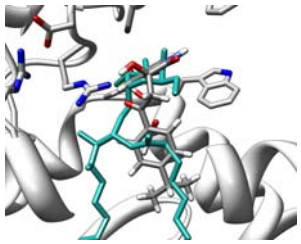
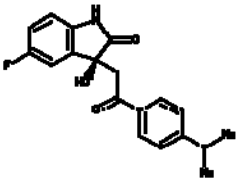
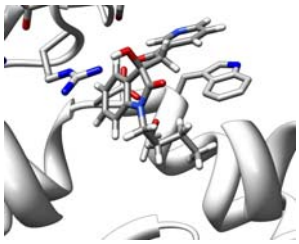
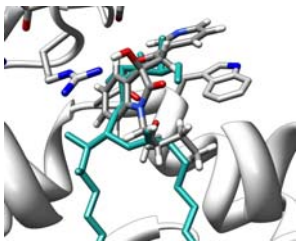
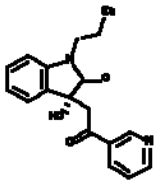
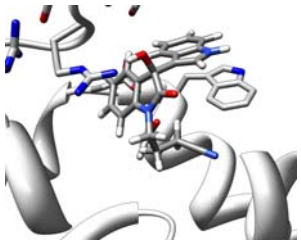
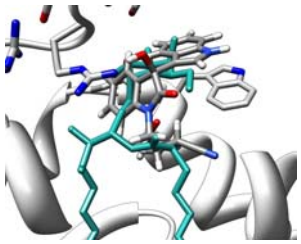
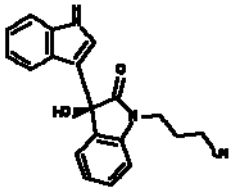
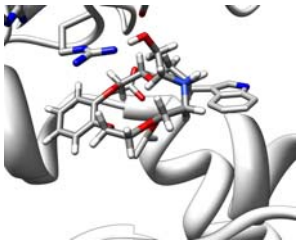
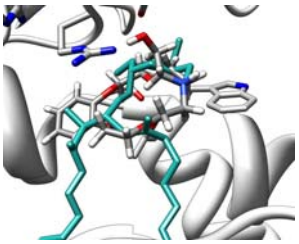
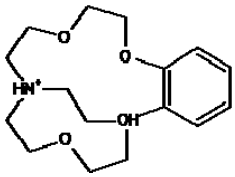
Z.04243786			
Z.04243767			
Z.00160722			
25000.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
Z.04310915			
Z.01778001			
28835.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
ZINC1P04			
5000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**

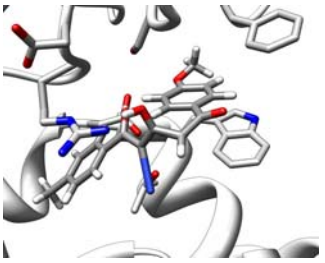
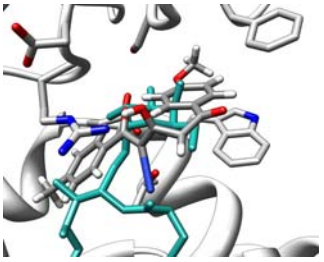
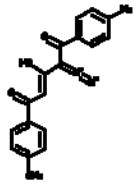
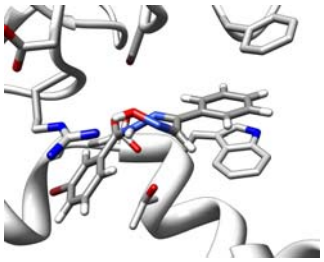
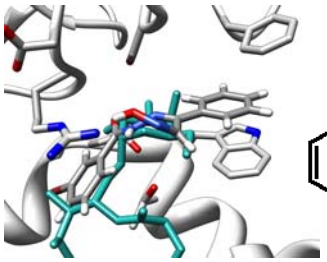
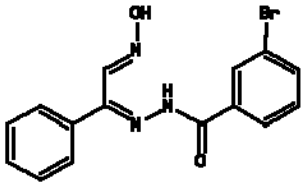
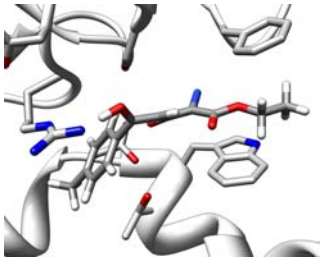
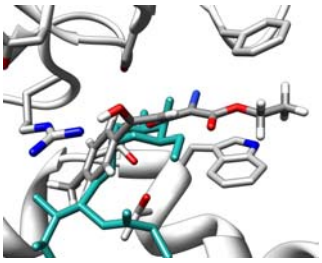
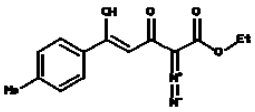
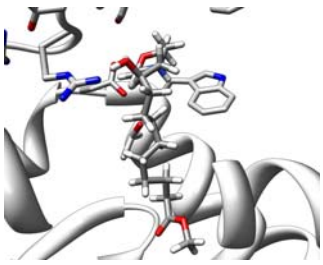
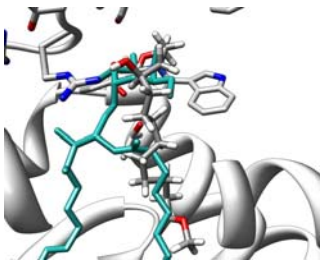
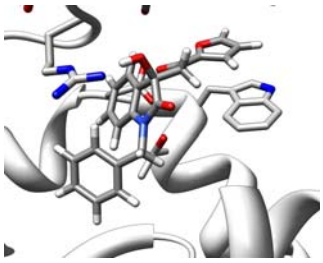
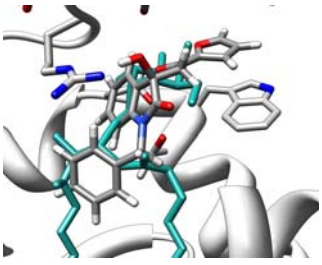
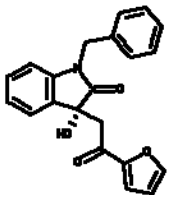
Z.00083443			
Z.02227601			
Z.03138609			
Z.04343704			
10000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.04312218			

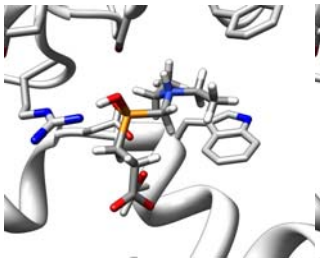
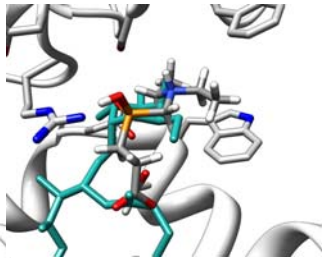
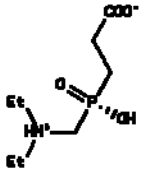
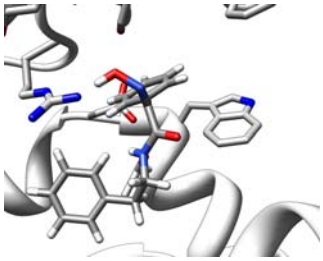
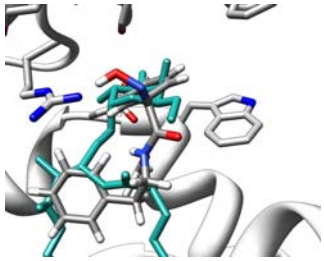
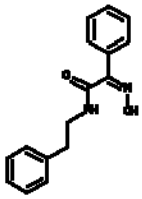
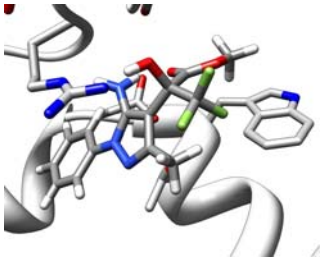
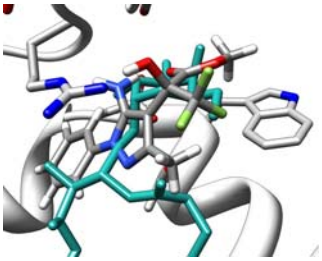
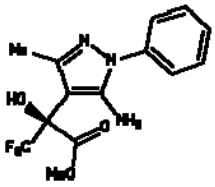
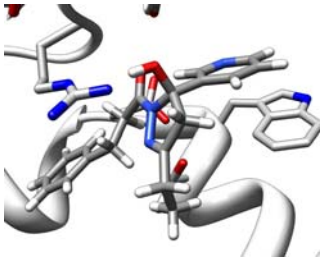
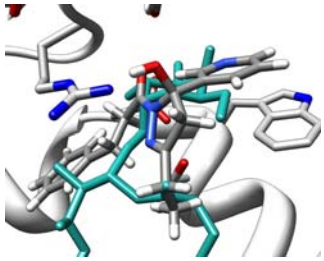
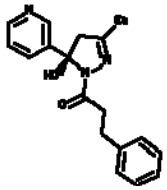
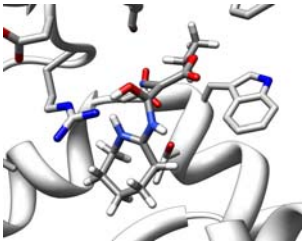
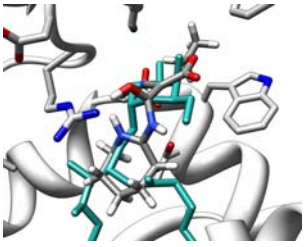
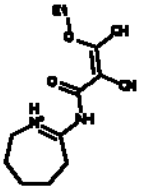
Z.01841416			
15000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.00033264			
Z.00614270			
20000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.01790470			
Z.03153788			

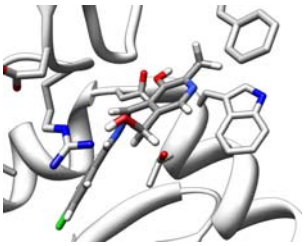
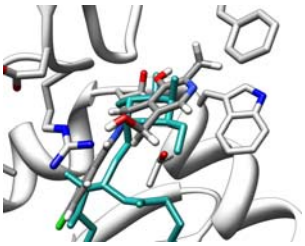
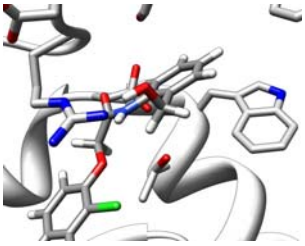
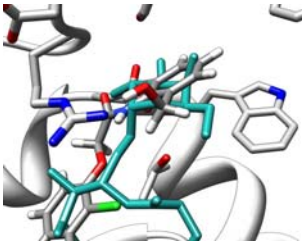
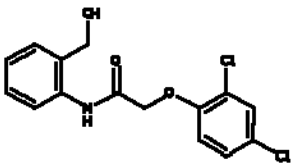
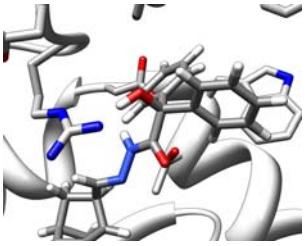
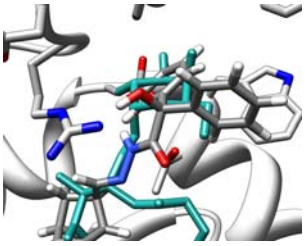
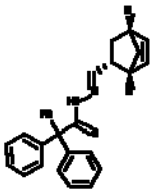
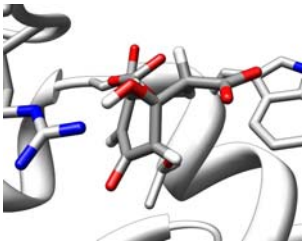
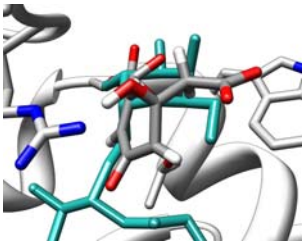
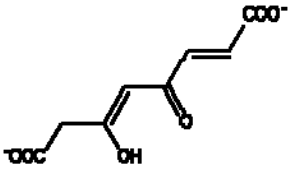
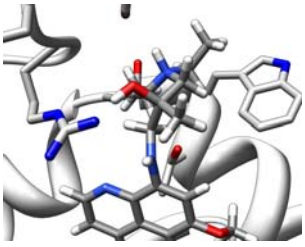
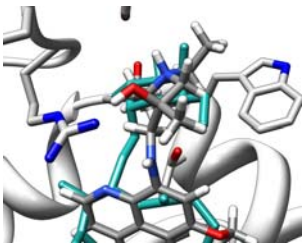
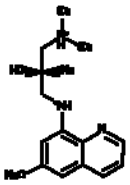
Z.00439300			
25000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.04425726			
Z.04419664			
28015.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.02317200			
ZINC1P05			
5000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.00481545			

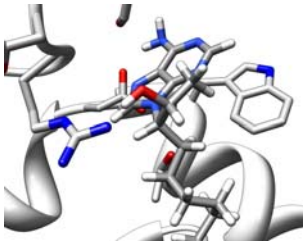
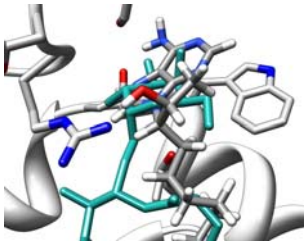
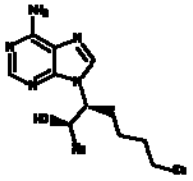
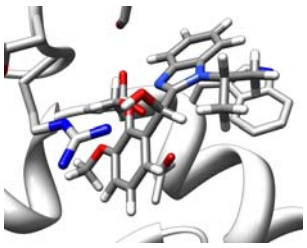
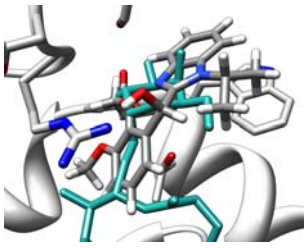
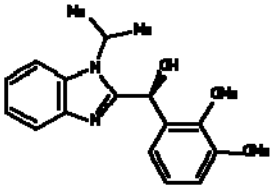
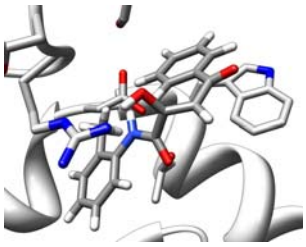
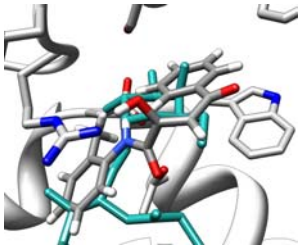
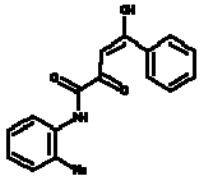
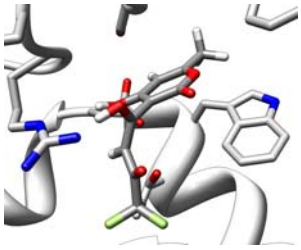
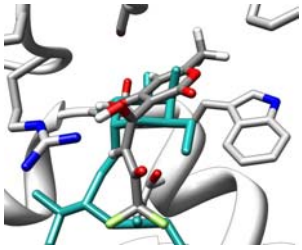
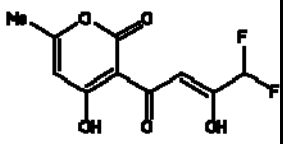
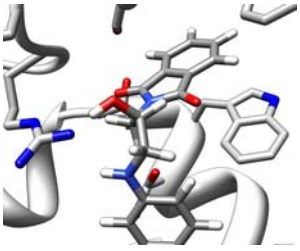
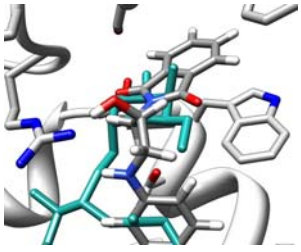
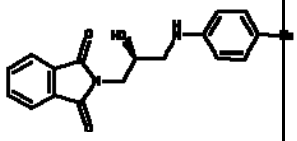
Z.00565397			
Z.04474601			
10000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.04396990			
Z.04503257			
15000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.04529800			
20000.TOP25	**IMAGE**	**with α-GalCer**	**2D image**

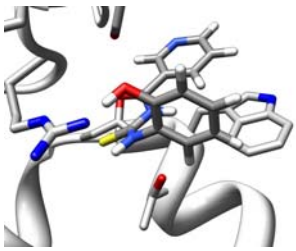
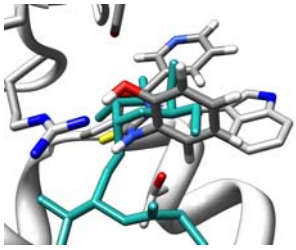
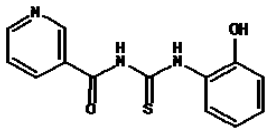
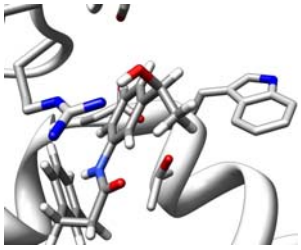
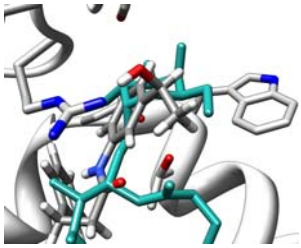
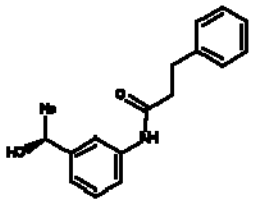
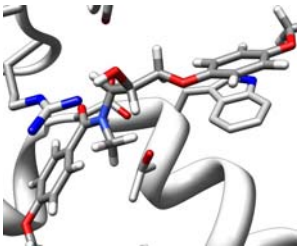
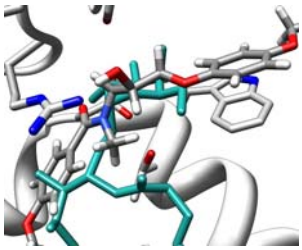
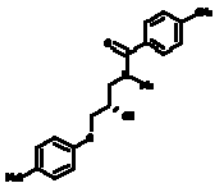
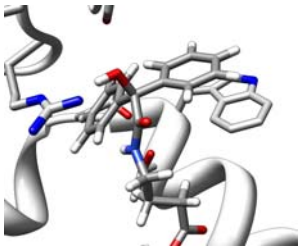
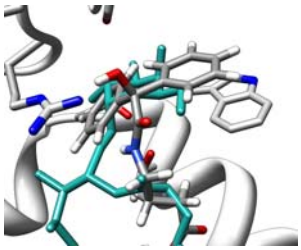
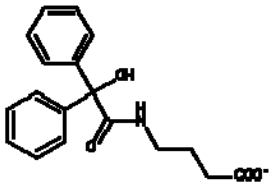
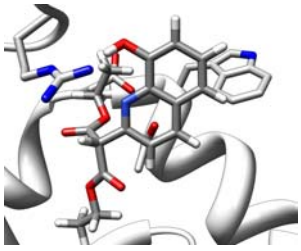
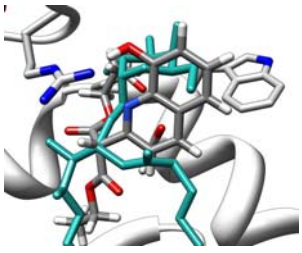
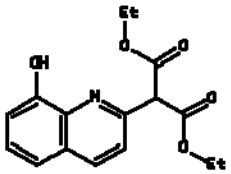
Z.01320528			
Z.04554904			
26945.TOP25	**IMAGE**	**with α-GalCer**	**2D image**
Z.02235661			
Z.04566328			

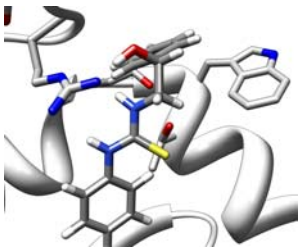
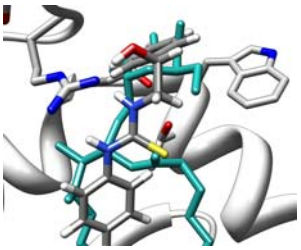
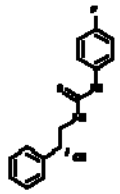
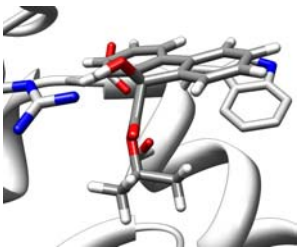
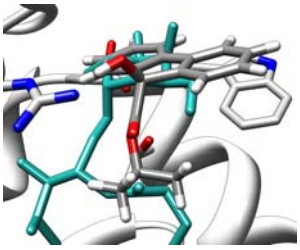
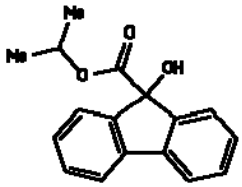
ZINC1P06			
5000.TOP35	IMAGE	with α-GalCer	2D image
Z.04607075			
Z.04603355			
Z.04576476			
Z.04611118			
Z.00565362			

Z.03173043			
Z.04616014			
Z.02432874			
Z.00371045			
10000.TOP65	**IMAGE**	**with α-GalCer**	**2D image**
Z.01714214			

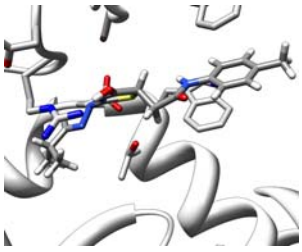
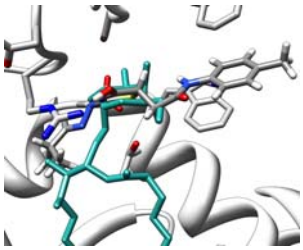
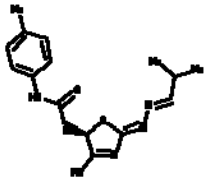
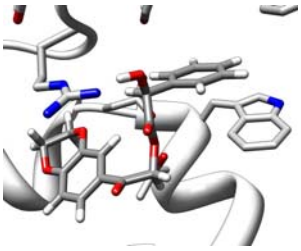
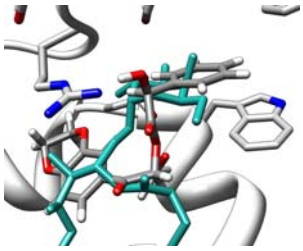
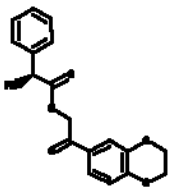
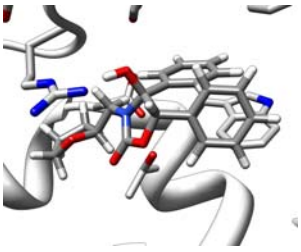
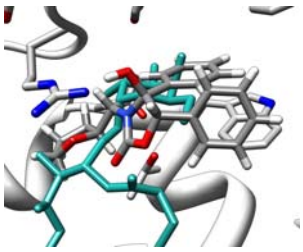
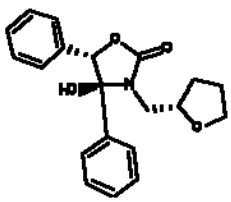
Z.03003132			
15000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.00455209			
Z.04652104			
Z.03869299			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.01656122			

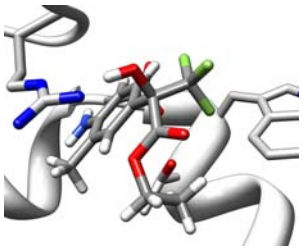
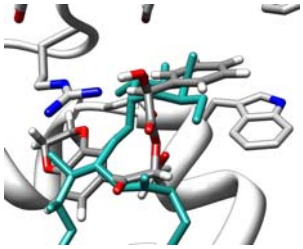
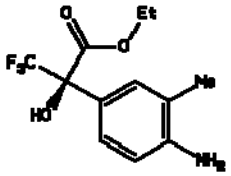
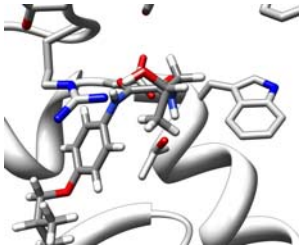
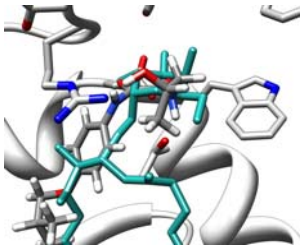
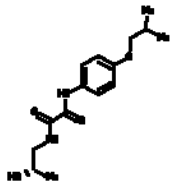
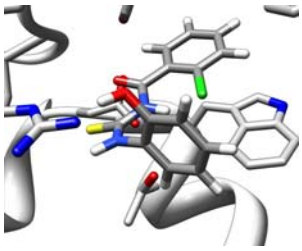
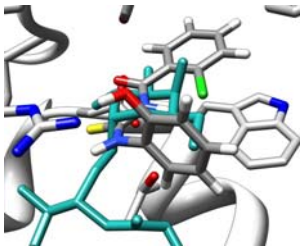
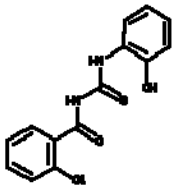
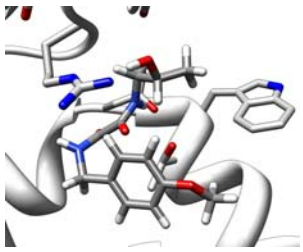
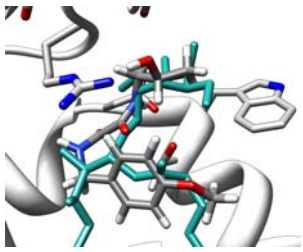
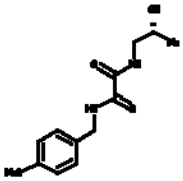
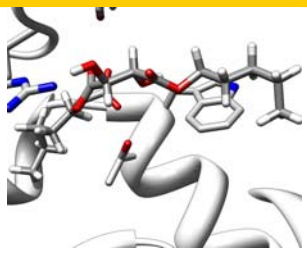
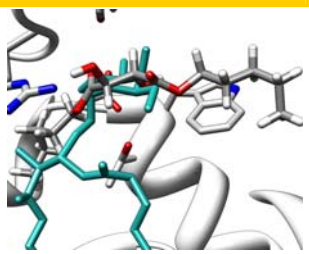
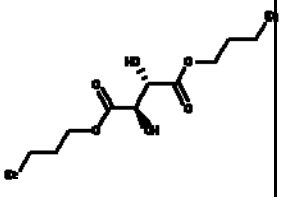
Z.01614354			
Z.00239203			
25000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.00273679			
Z.03771134			
Z.00384477			

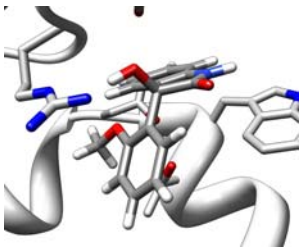
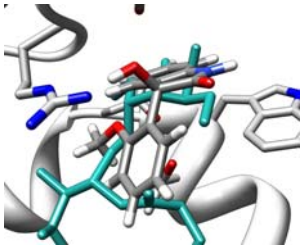
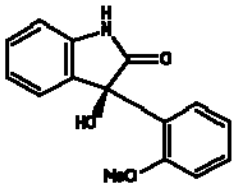
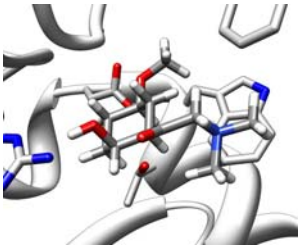
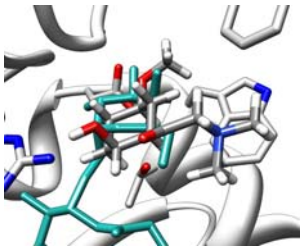
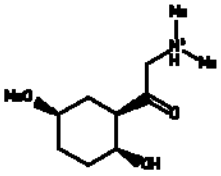
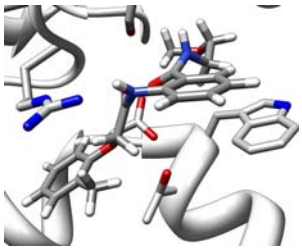
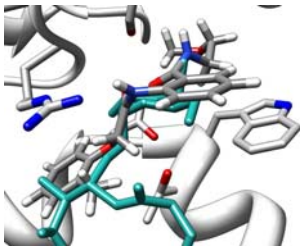
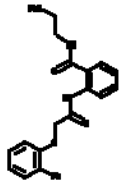
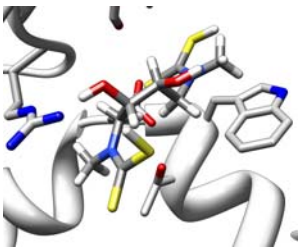
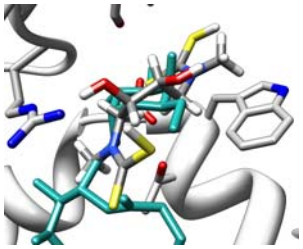
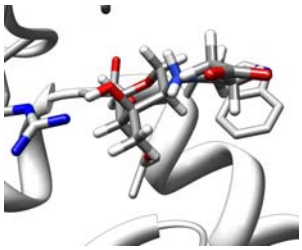
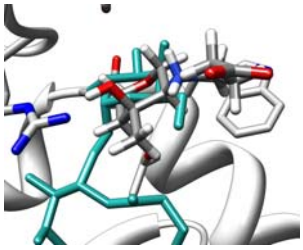
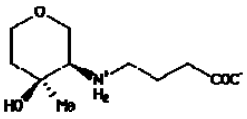
Z.00284350			
Z.01017190			
27016.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
Z.00465915			
Z.02891110			
Z.02026072			

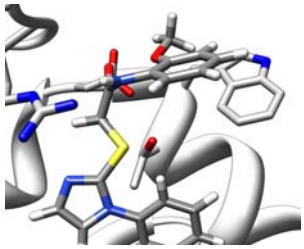
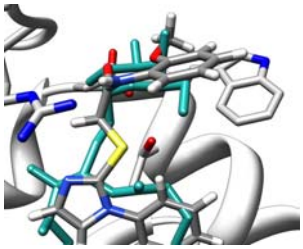
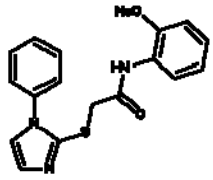
Z.00306222			
Z.02817280			

ZINC1P07

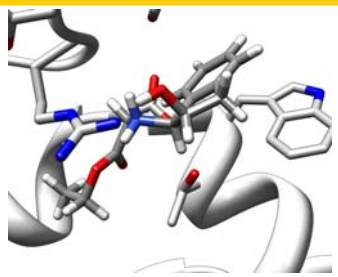
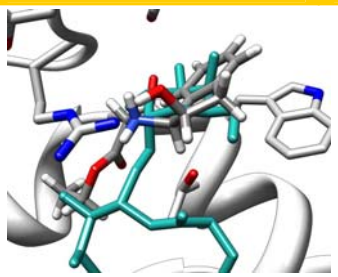
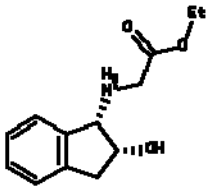
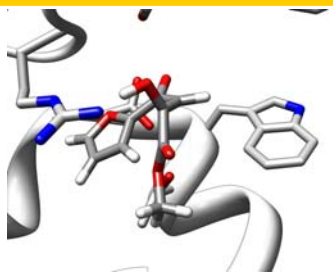
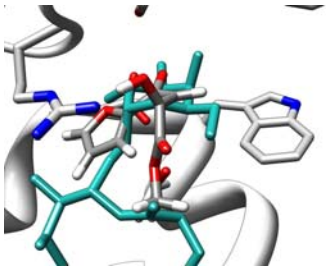
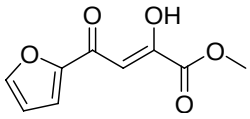
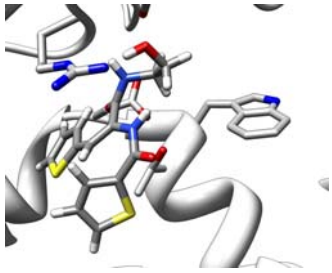
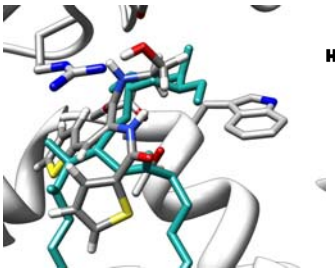
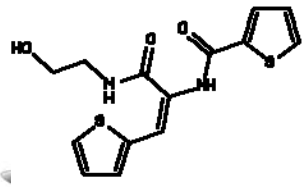
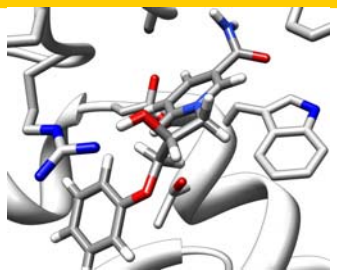
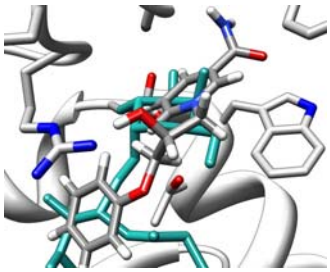
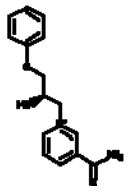
5000.TOP50	IMAGE	with α-GalCer	2D image
Z.04690145			
Z.02316160			
Z.04689274			

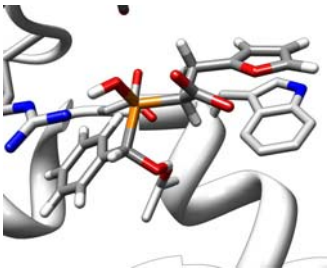
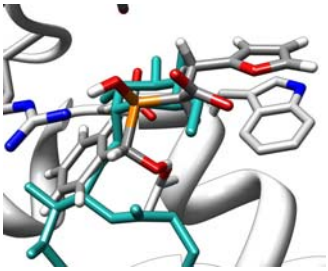
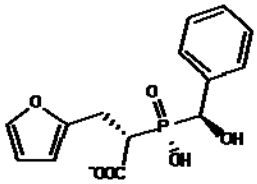
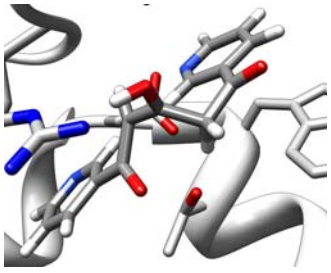
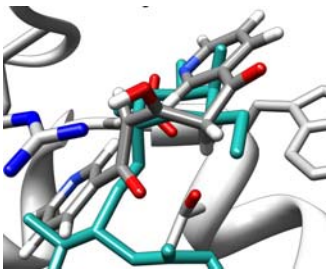
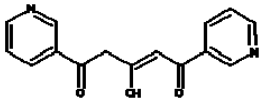
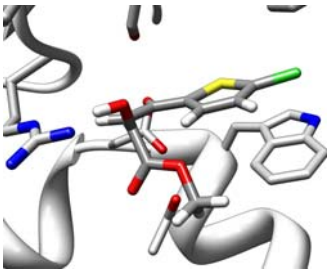
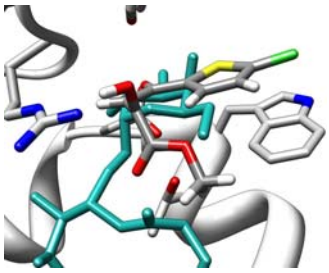
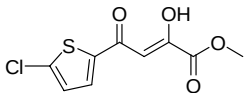
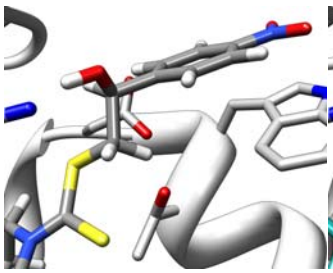
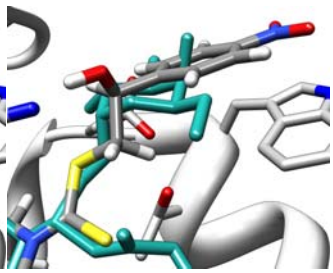
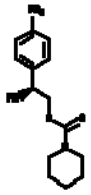
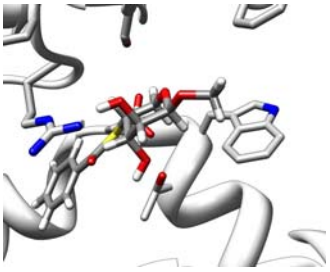
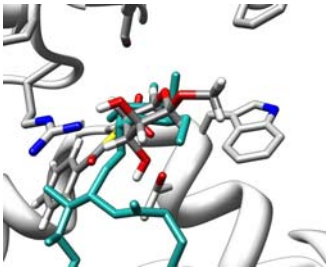
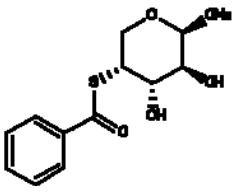
Z.02300362			
10000.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
Z.04703094			
Z.00456824			
Z.04702968			
15000.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
Z.04721900			

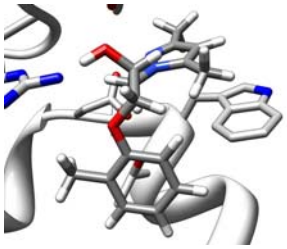
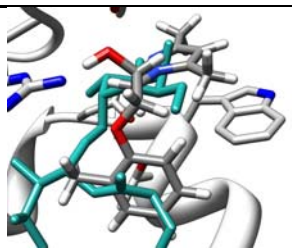
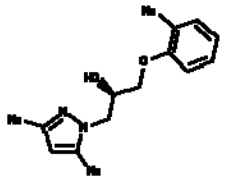
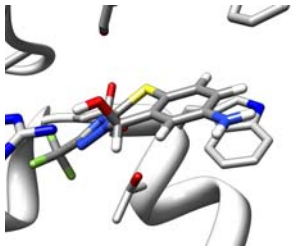
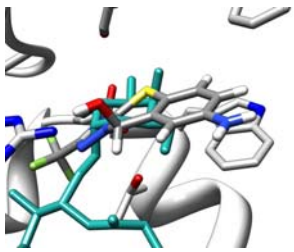
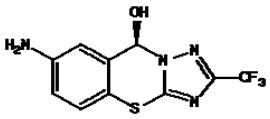
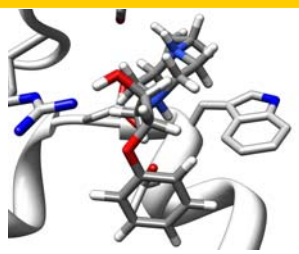
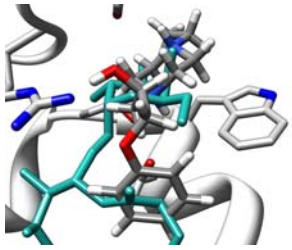
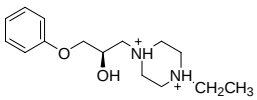
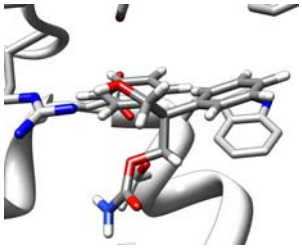
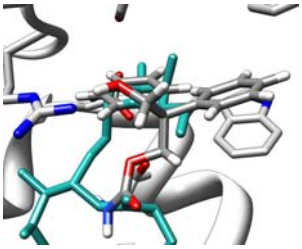
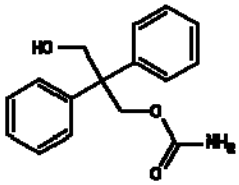
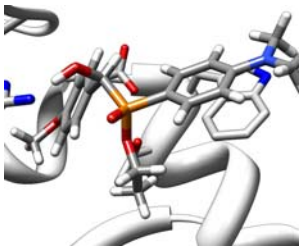
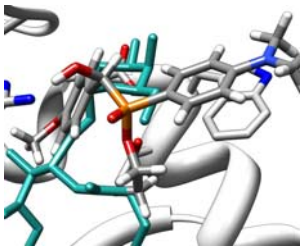
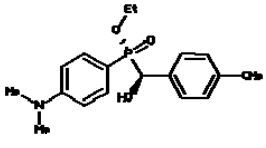
Z.04722637			
Z.04722606			
20000.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
ZINC04741079			
26565.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
Z.01661135			
Z.02227826			

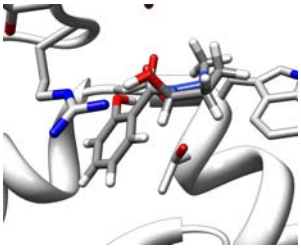
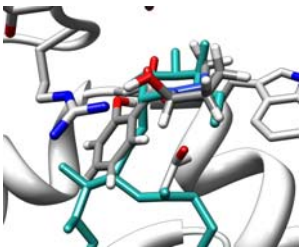
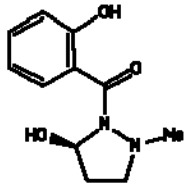
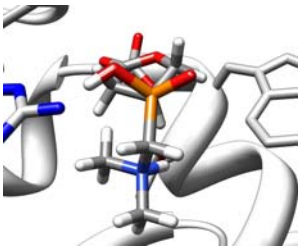
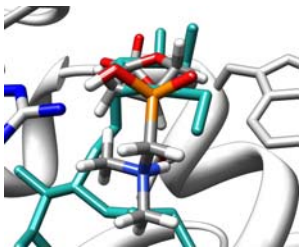
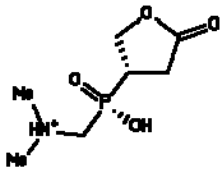
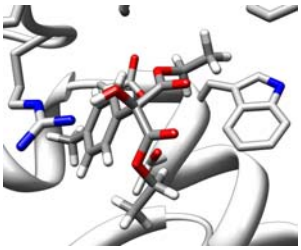
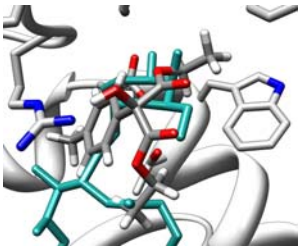
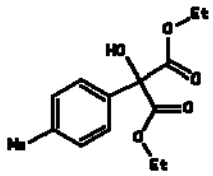
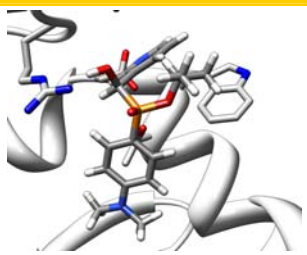
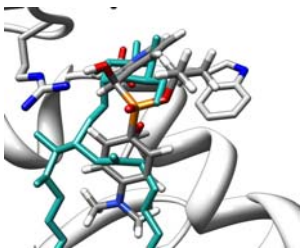
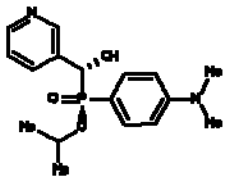
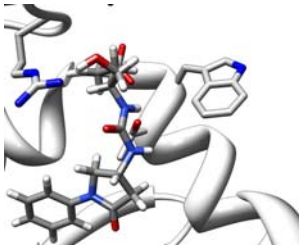
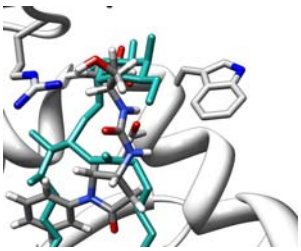
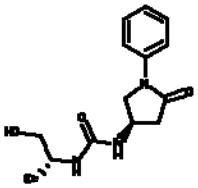
Z.00427355			

ZINC1P08

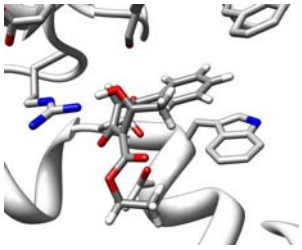
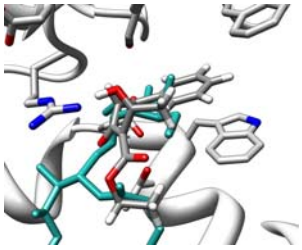
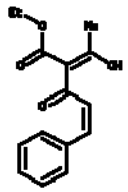
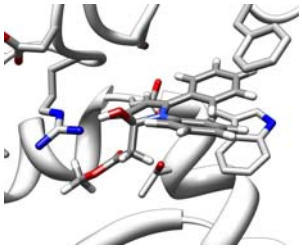
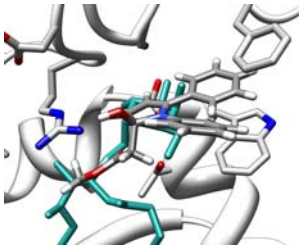
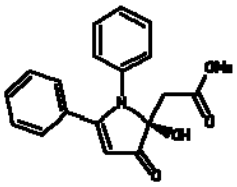
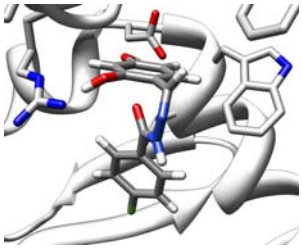
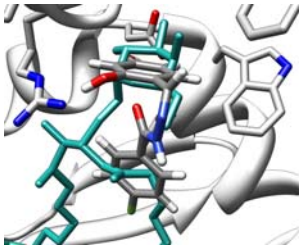
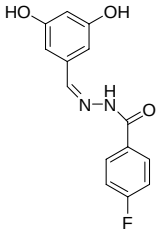
5000.TOP50	IMAGE	with α-GalCer	2D image
Z.01695674			
10000.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
Z.00535703			
Z.02297777			
15000.TOP50	**IMAGE**	**with α-GalCer**	**2D image**
Z.00297628			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**

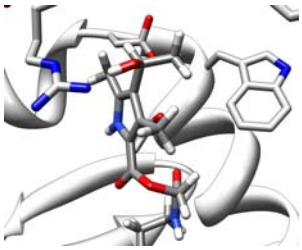
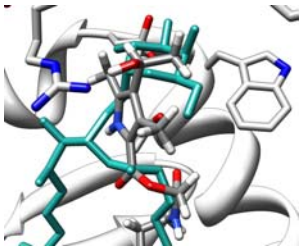
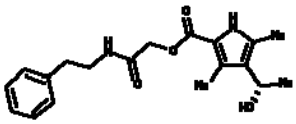
Z.04859764			
26415.TOP65	**IMAGE**	**with α-GalCer**	**2D image**
Z.04876933			
Z.03888466			
Z.00473233			
ZINC04896439			
ZINC1P09			
5000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
10000.TOP80	**IMAGE**	**with α-GalCer**	**2D image**

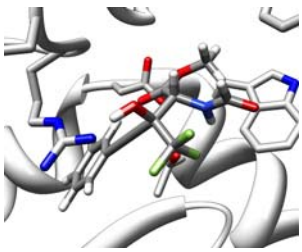
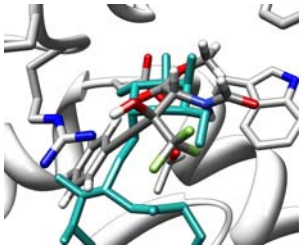
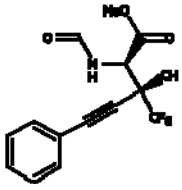
Z.00261364			
15000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.04245265			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.04991301			
Z.01700895			
26355.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
ZINC1P10			
10000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.00527946			

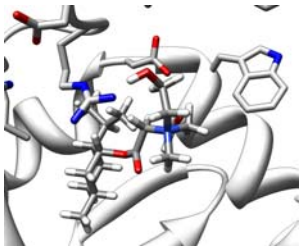
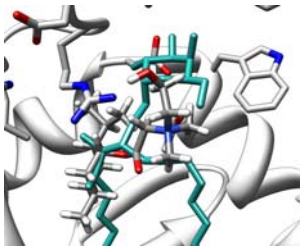
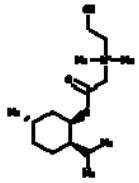
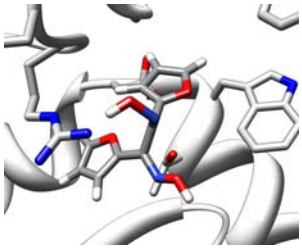
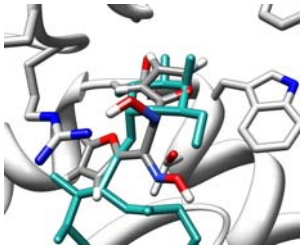
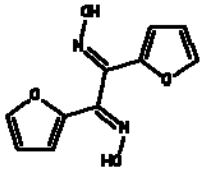
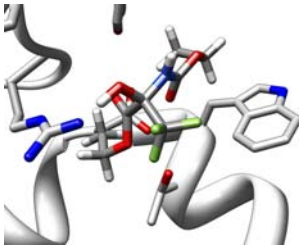
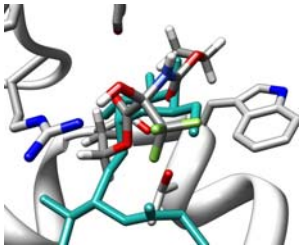
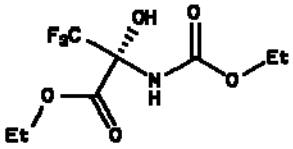
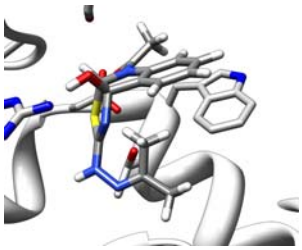
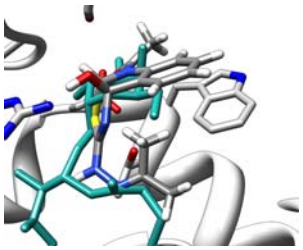
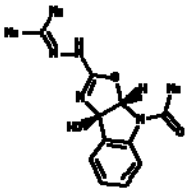
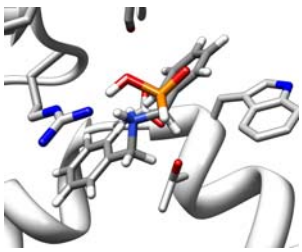
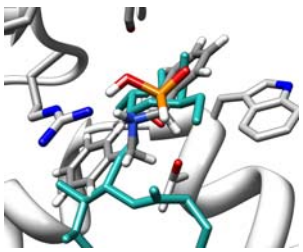
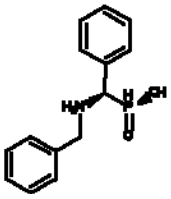
15000.TOP60	IMAGE	with α-GalCer	2D image
Z.02870289			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.05092045			
Z.01727318			
26080.TOP65	**IMAGE**	**with α-GalCer**	**2D image**
Z.00533416			
Z.05123042			

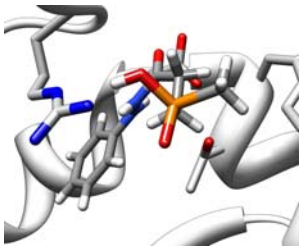
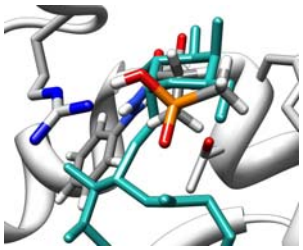
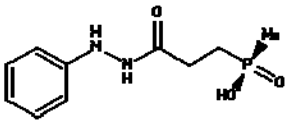
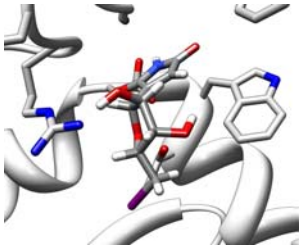
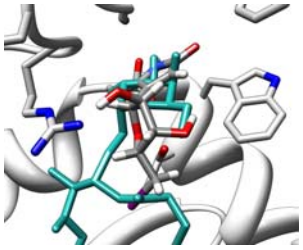
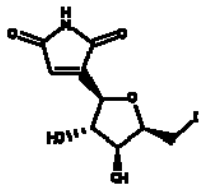
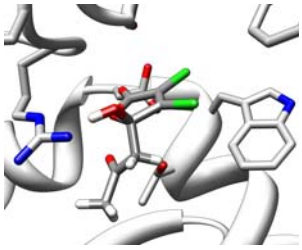
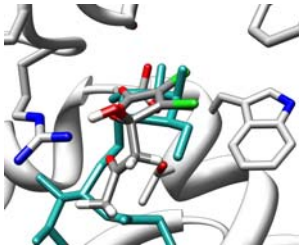
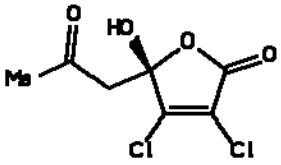
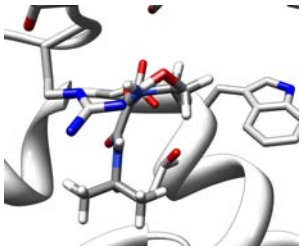
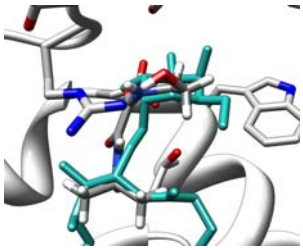
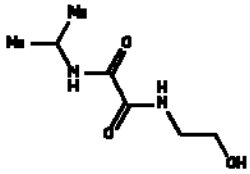
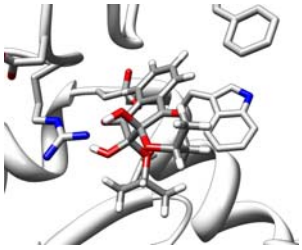
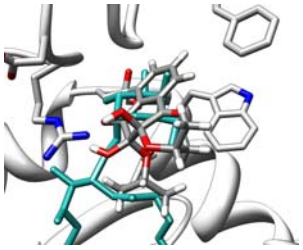
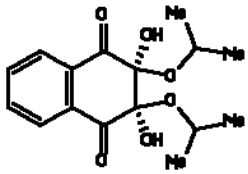
ZINC1P11

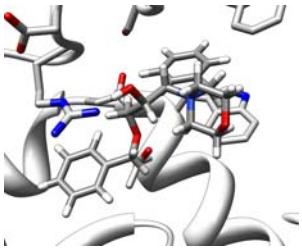
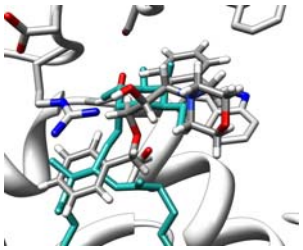
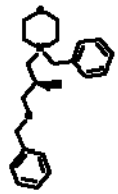
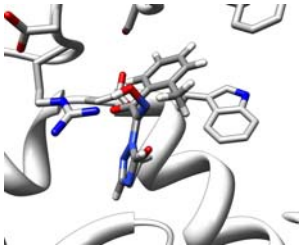
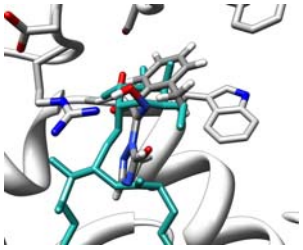
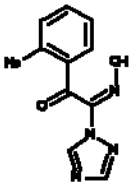
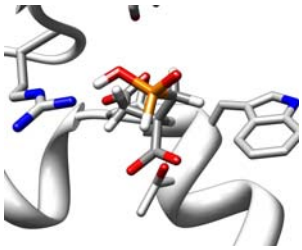
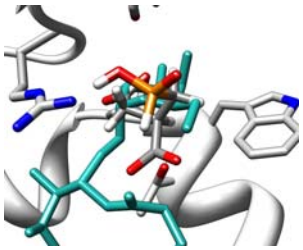
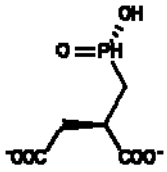
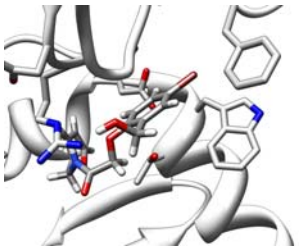
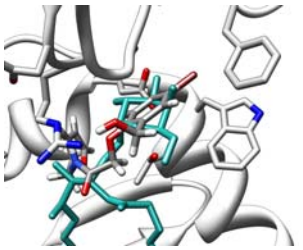
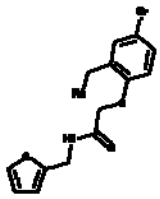
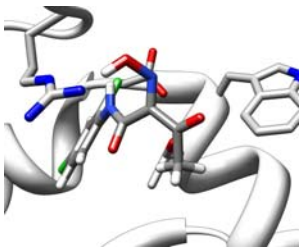
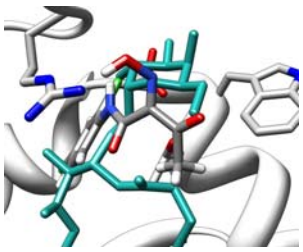
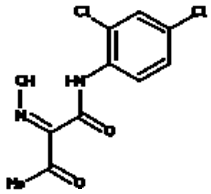
5000.TOP60	IMAGE	with α-GalCer	2D image
Z.05012694			
Z.00263691			
Z.00253972			

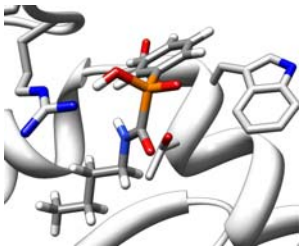
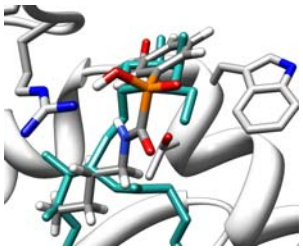
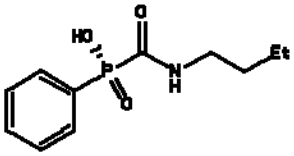
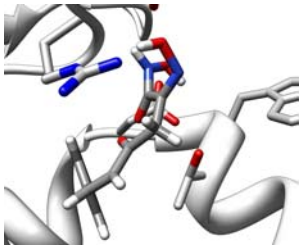
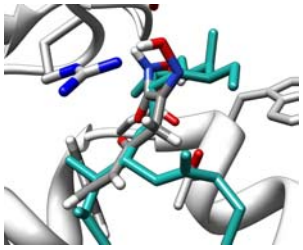
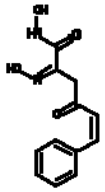
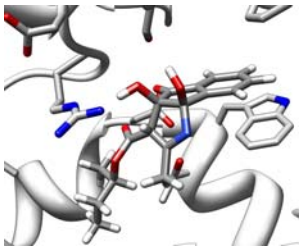
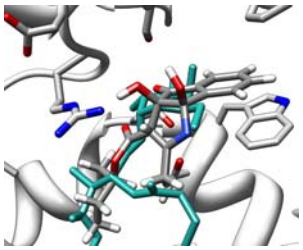
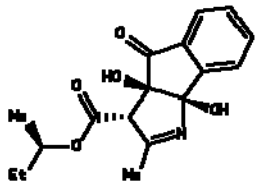
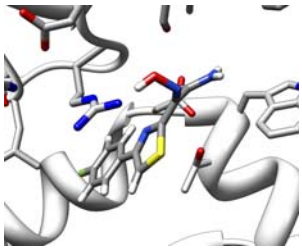
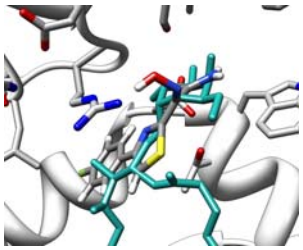
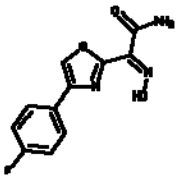
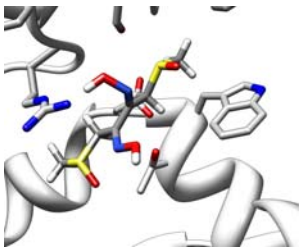
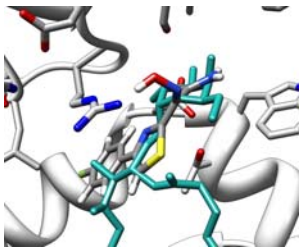
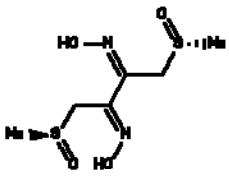
10000.TOP60	IMAGE	with α-GalCer	2D image
Z.03544226			

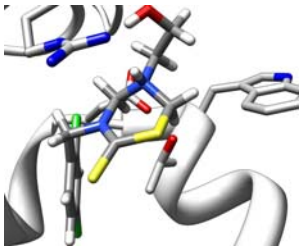
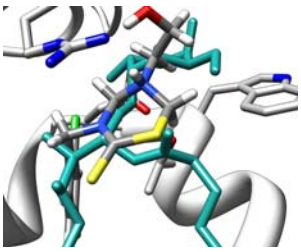
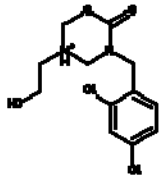
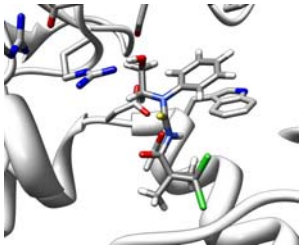
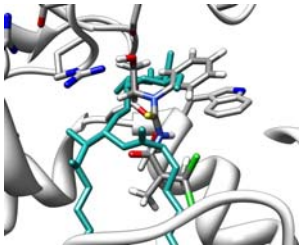
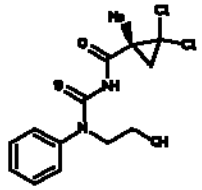
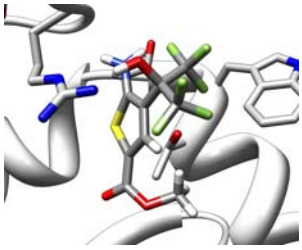
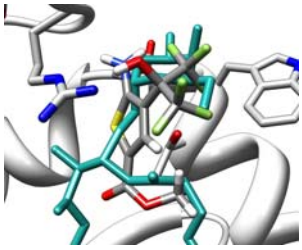
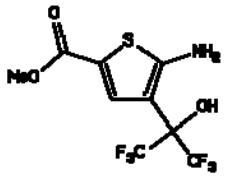
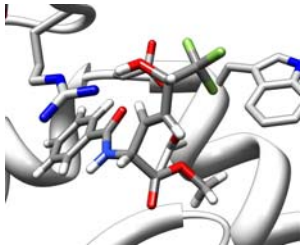
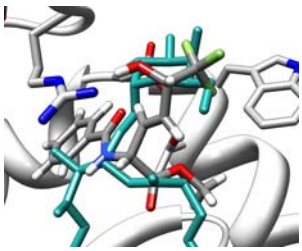
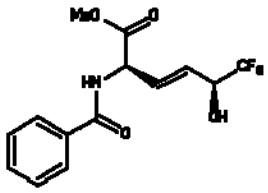
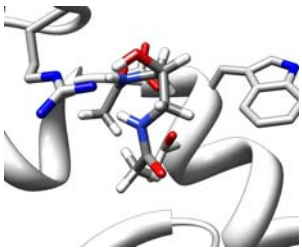
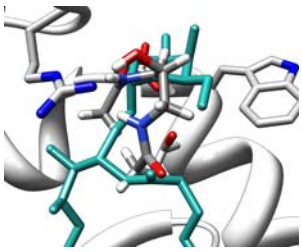
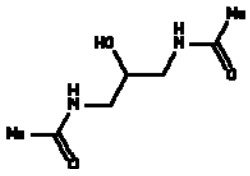
15000.TOP60	IMAGE	with α-GalCer	2D image
Z.03078170			

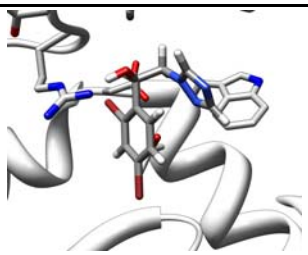
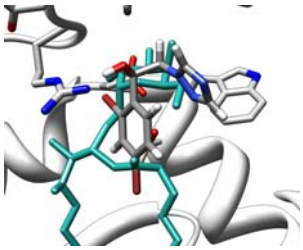
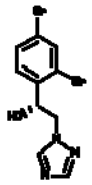
Z.02566571			
Z.05263535			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.02334030			
Z.00983686			
25000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.05345965			

Z.01277563			
Z.05369607			
30000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.05383020			
Z.05380985			
Z.00036944			
ZINC1P12			
5000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
10000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**

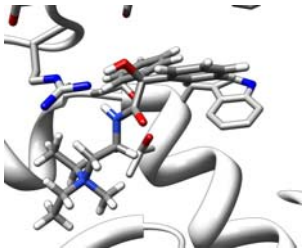
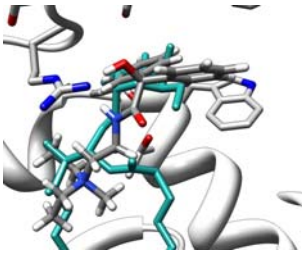
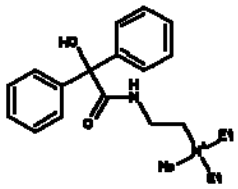
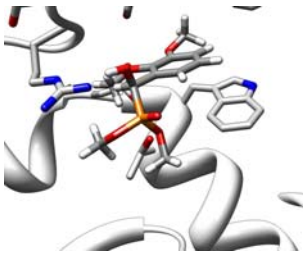
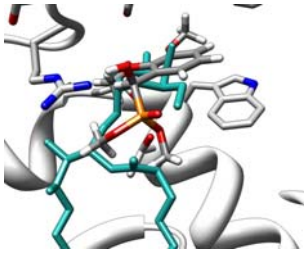
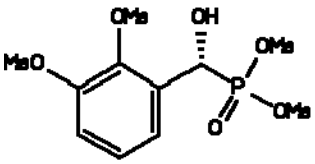
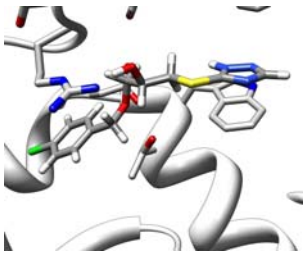
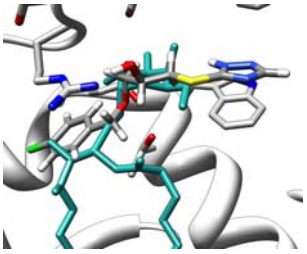
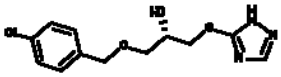
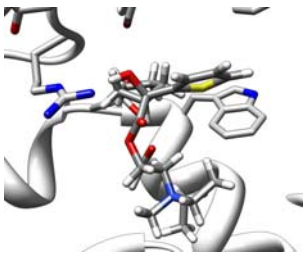
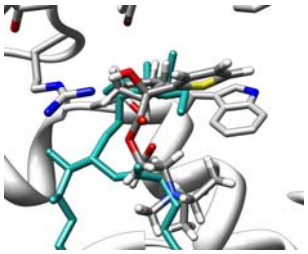
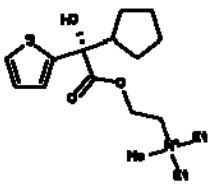
Z.02692841			
Z.05440411			
Z.02352688			
15000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.05468225			
Z.05495236			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**

Z.05552324			
Z.05485949			
Z.05550890			
25000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.05592186			
Z.05567214			

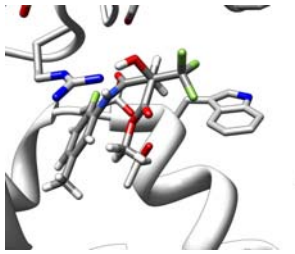
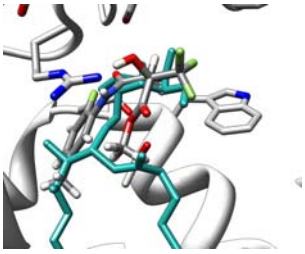
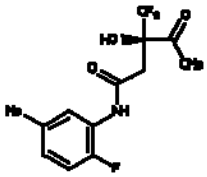
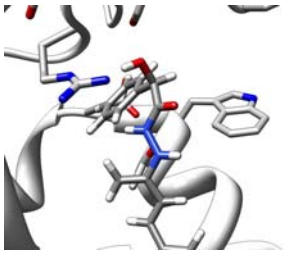
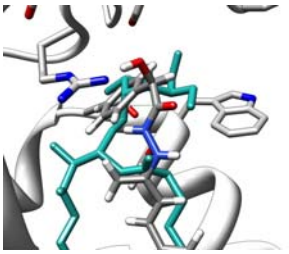
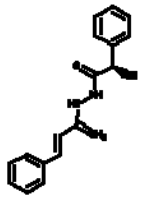
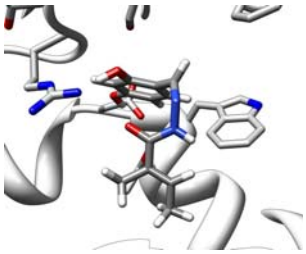
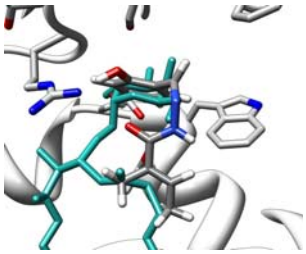
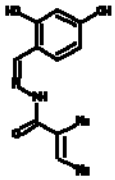
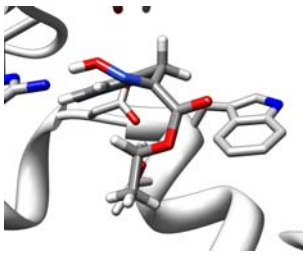
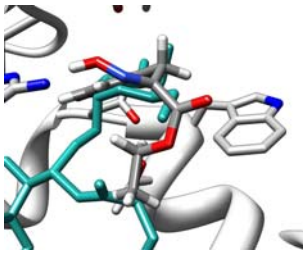
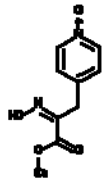
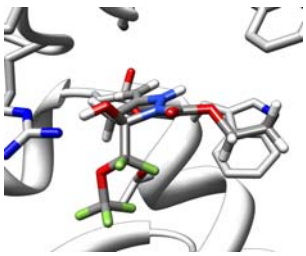
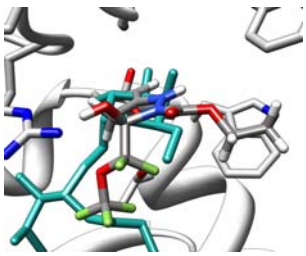
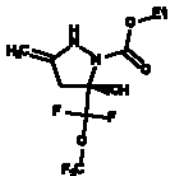
Z.01567475			
Z.03262894			
30000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.03200057			
Z.03202278			
Z.01574083			

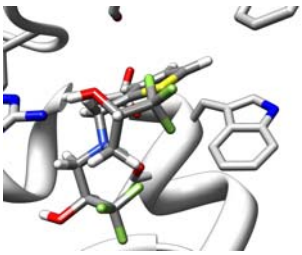
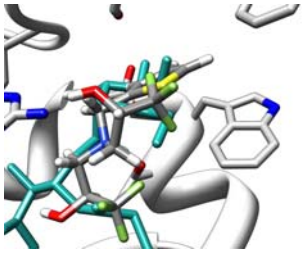
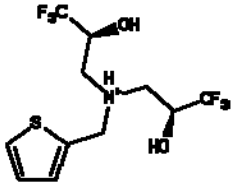
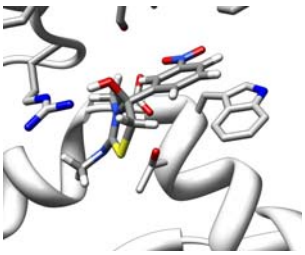
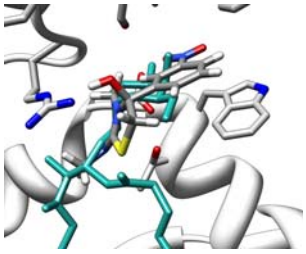
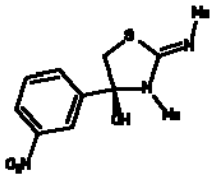
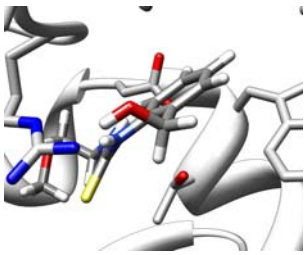
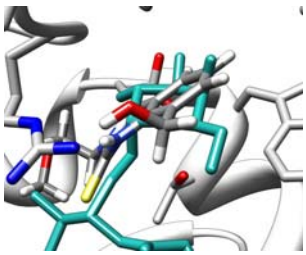
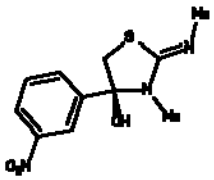
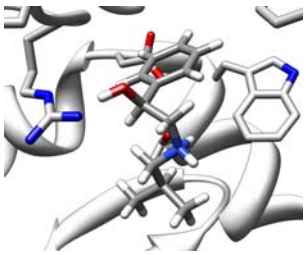
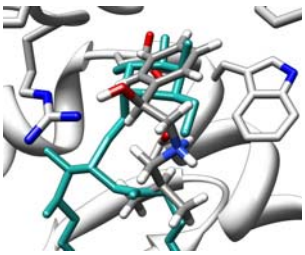
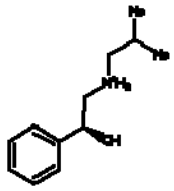
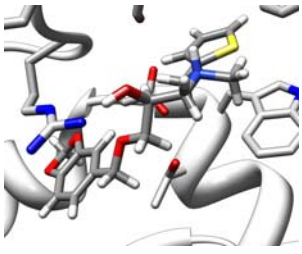
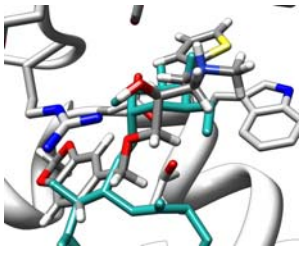
Z.00983319			

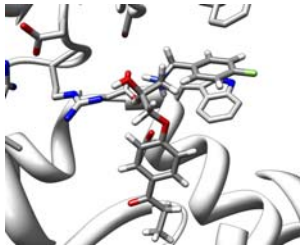
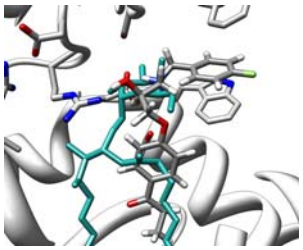
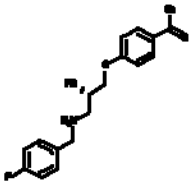
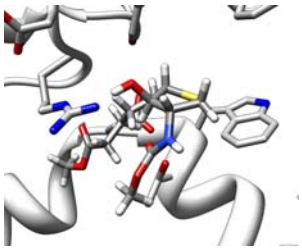
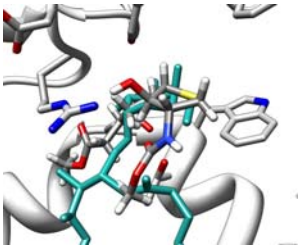
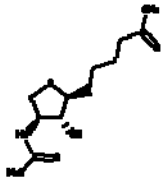
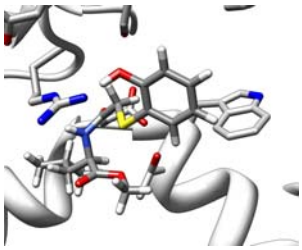
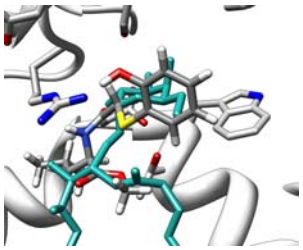
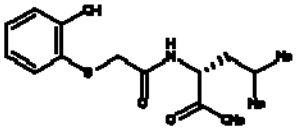
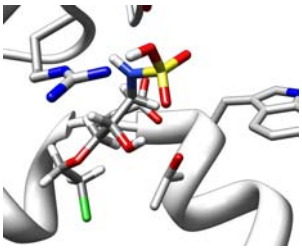
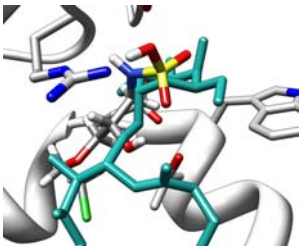
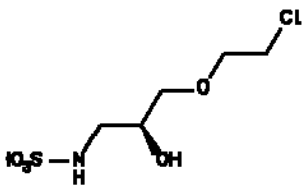
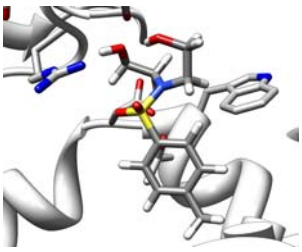
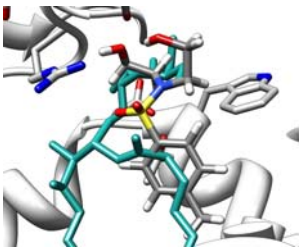
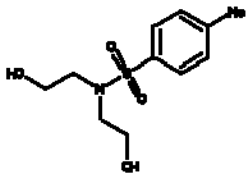
ZINC1P13

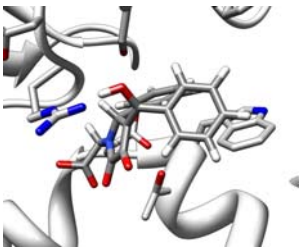
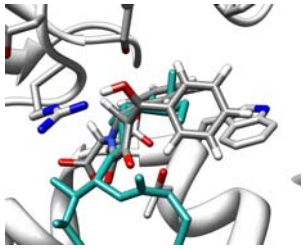
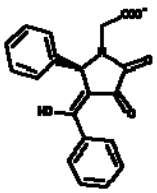
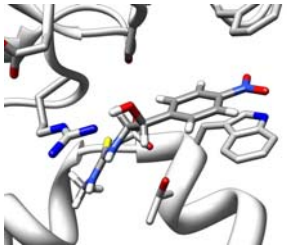
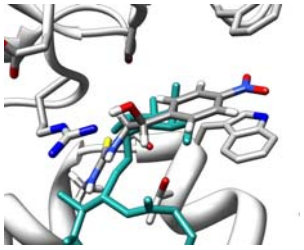
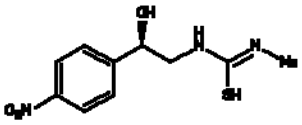
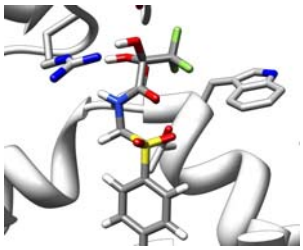
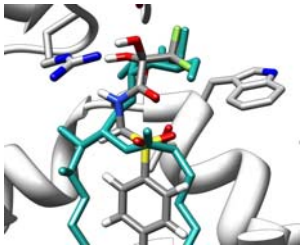
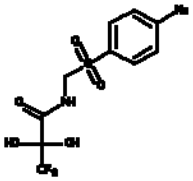
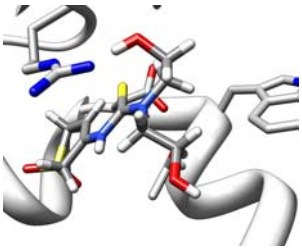
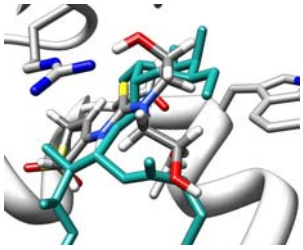
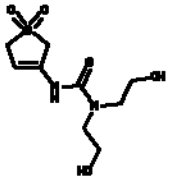
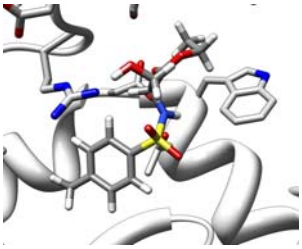
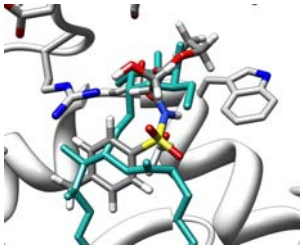
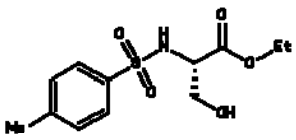
5000.TOP60	IMAGE	with α-GalCer	2D image
Z.01636925			
Z.02042329			
Z.01230594			
Z.02041344			

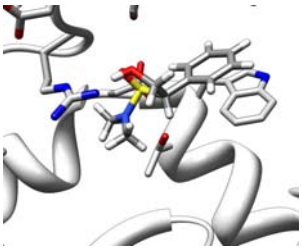
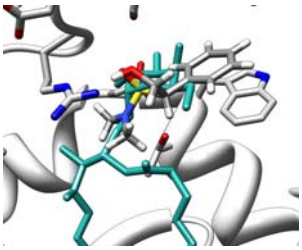
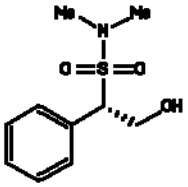
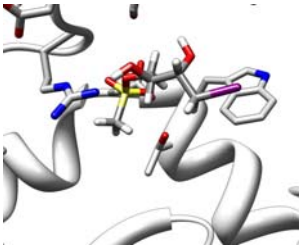
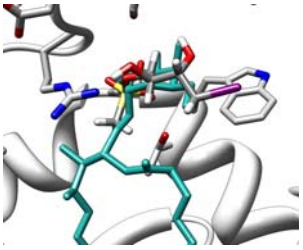
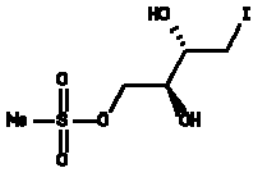
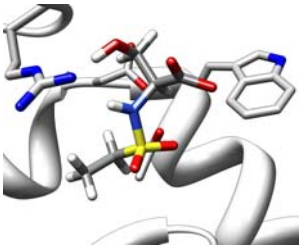
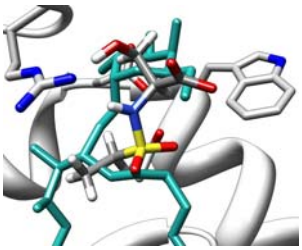
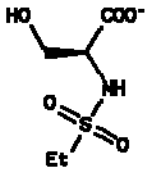
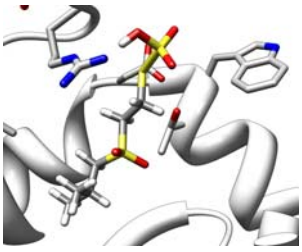
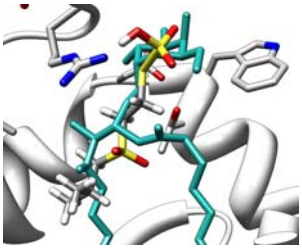
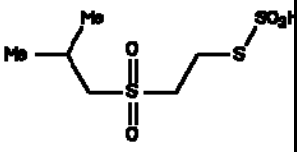
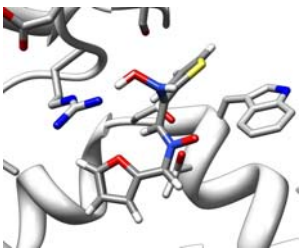
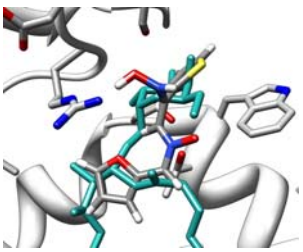
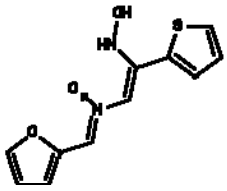
10000.TOP60	IMAGE	with α-GalCer	2D image

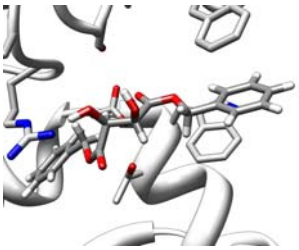
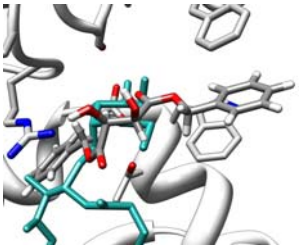
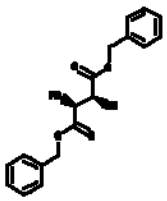
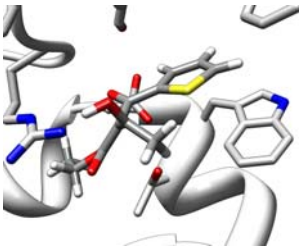
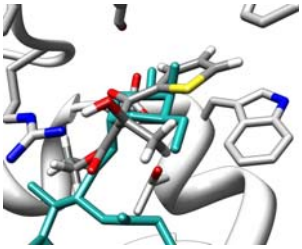
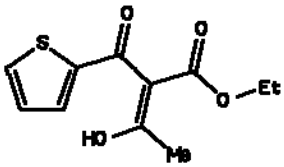
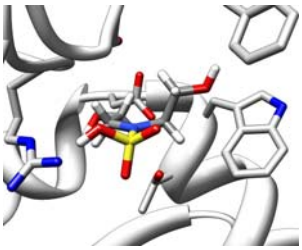
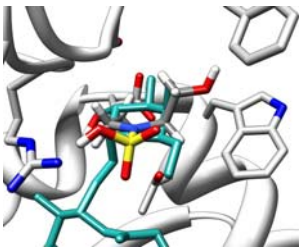
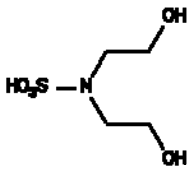
Z.02776150			
Z.02339989			
Z.05664708			
15000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.05734833			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.02355636			

Z.05770557			
25000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.02698096			
Z.03308706			
30000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.02655334			
ZINC1P14			
5000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.03495909			
10000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**

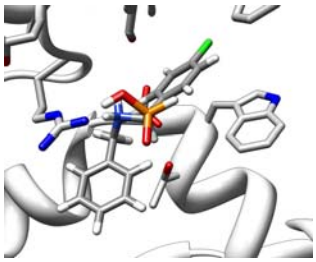
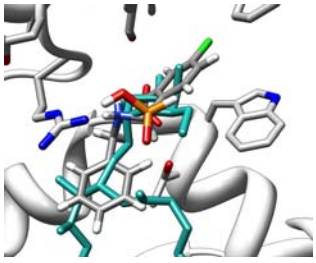
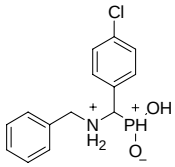
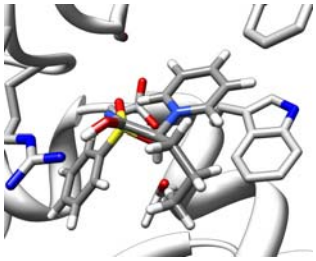
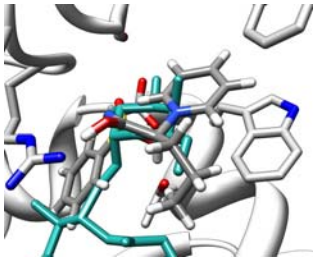
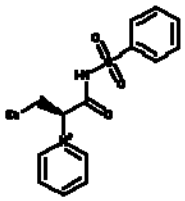
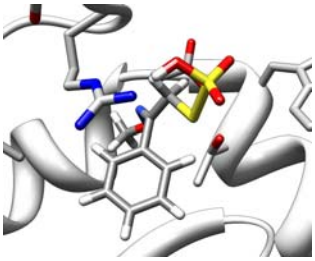
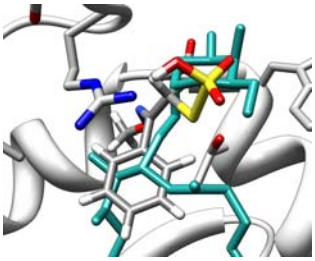
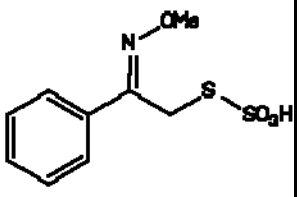
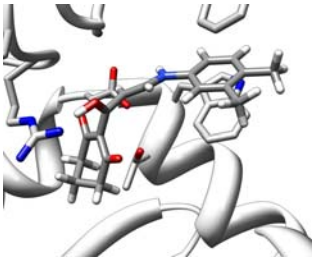
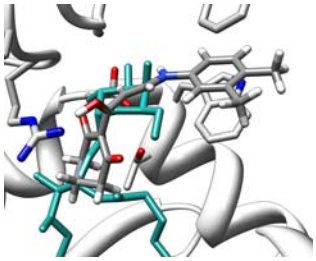
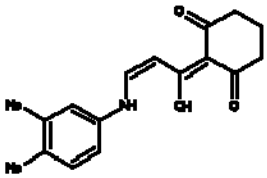
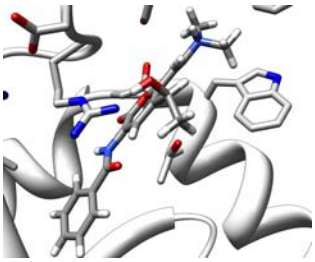
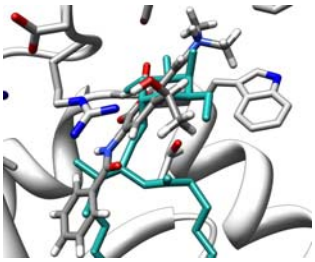
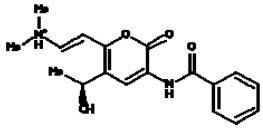
Z.03378271			
Z.02121425			
Z.05977016			
15000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.04697508			
Z.00056417			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**

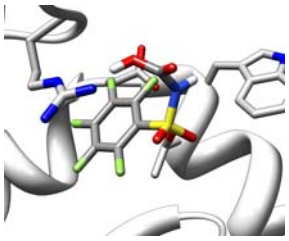
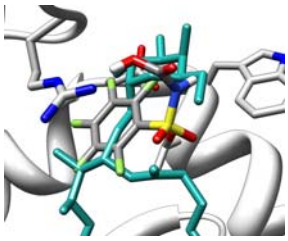
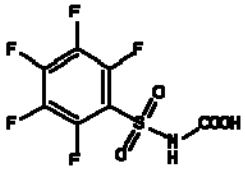
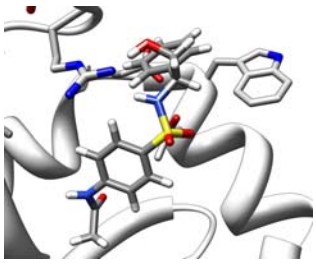
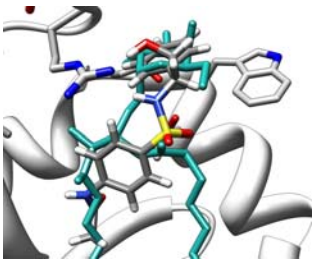
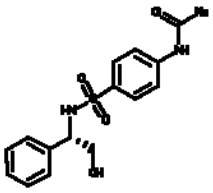
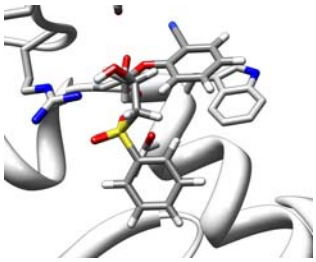
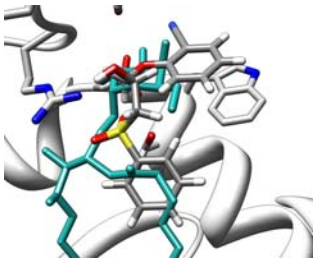
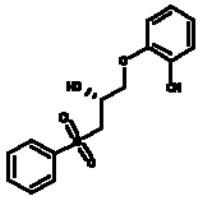
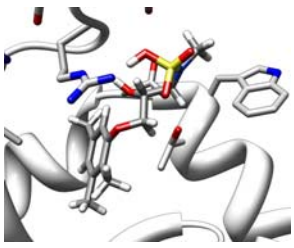
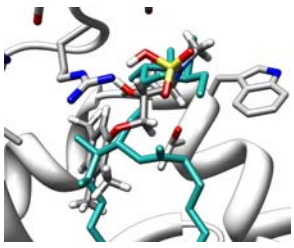
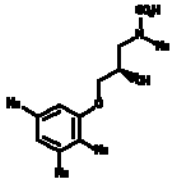
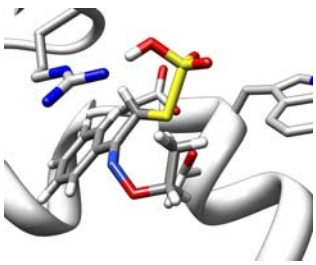
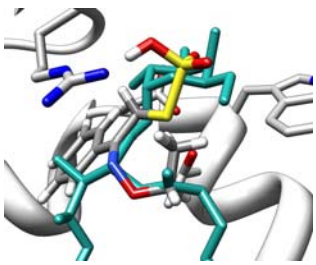
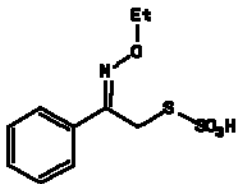
Z.05819255			
Z.02630412			
25000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.01648893			
Z.03866170			
30000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.01723849			

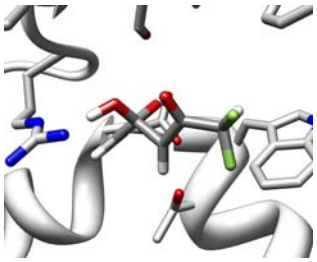
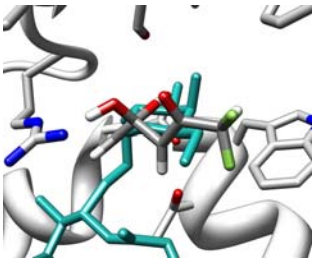
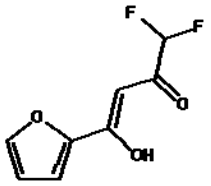
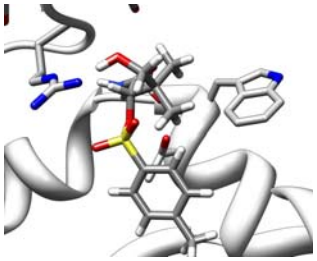
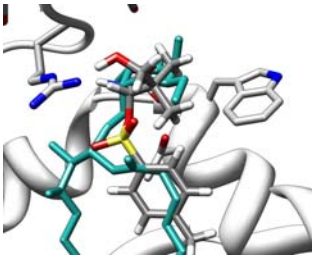
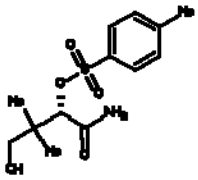
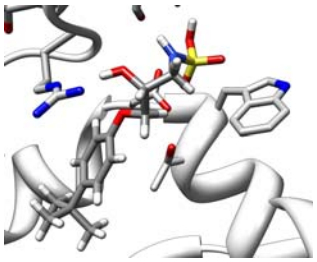
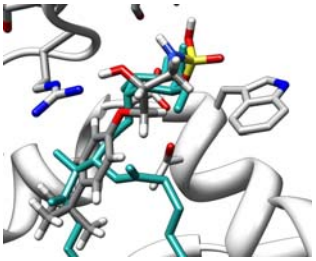
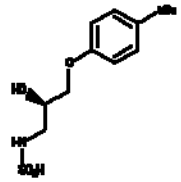
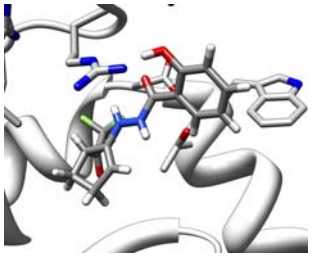
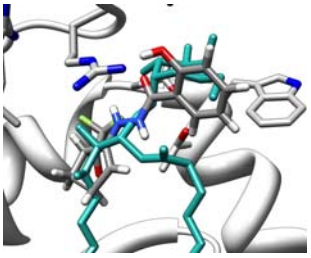
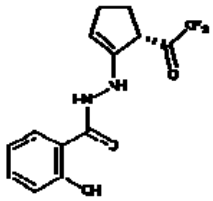
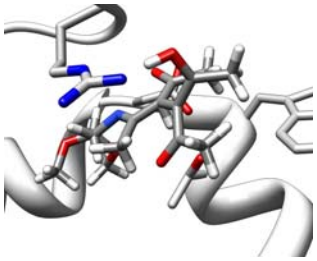
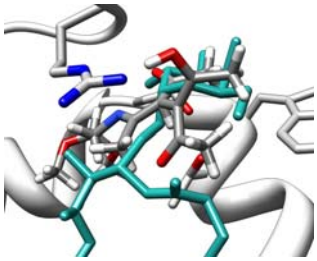
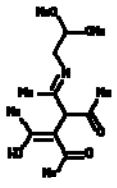
Z.01717691			
Z.01721554			
ZINC01691086			
ZINC1P15			
5000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
10000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.02283630			
15000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.01899744			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**

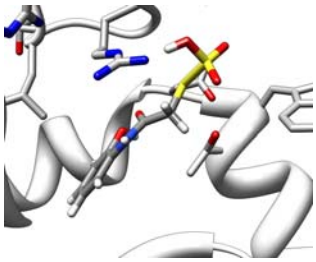
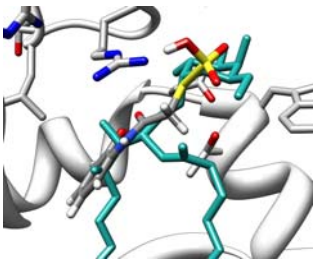
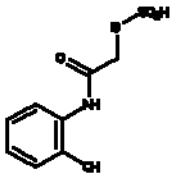
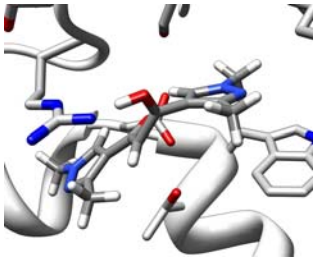
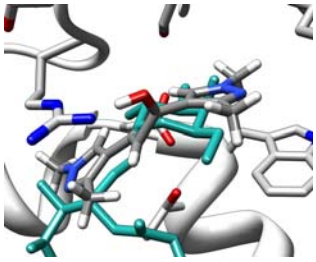
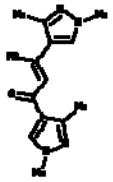
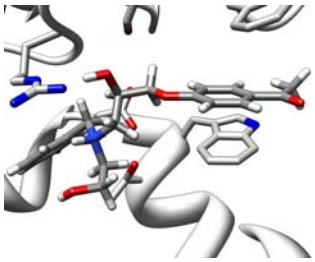
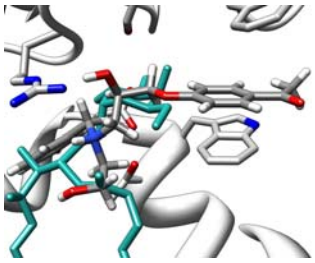
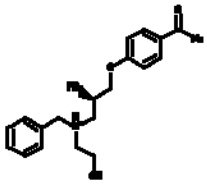
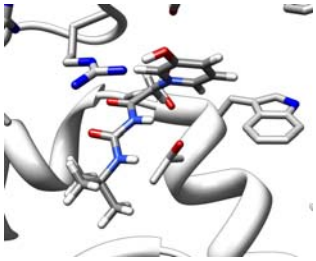
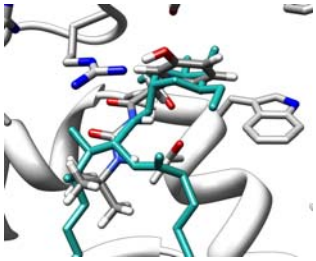
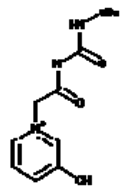
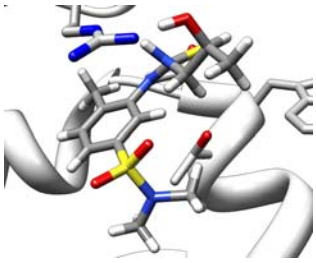
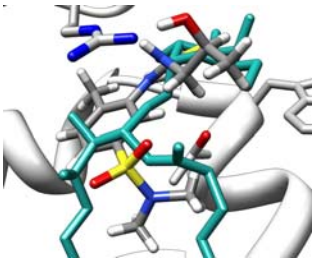
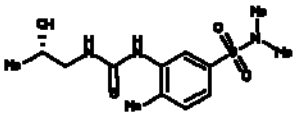
Z.02386959			
25000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.00166617			
30000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.04866021			

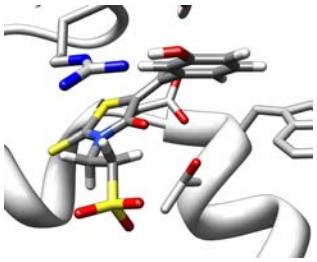
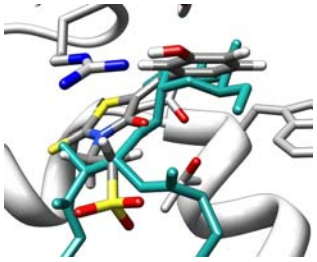
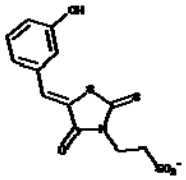
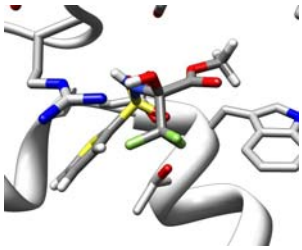
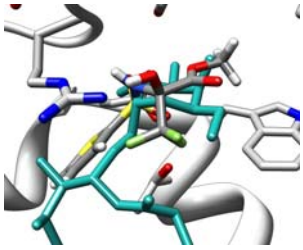
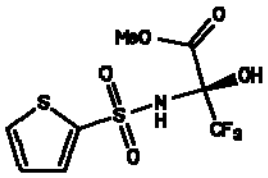
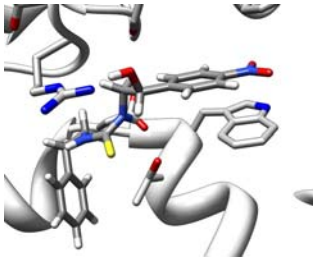
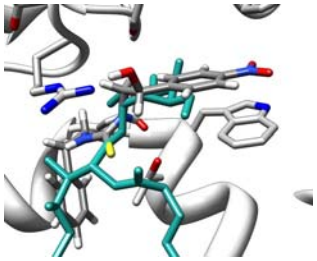
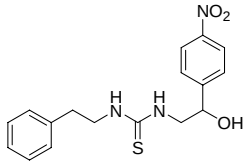
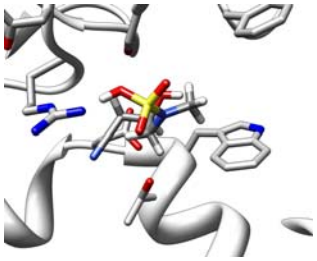
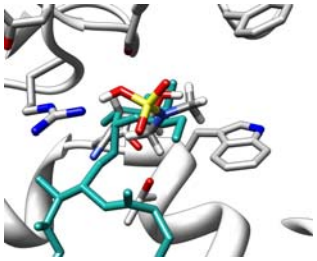
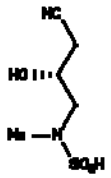
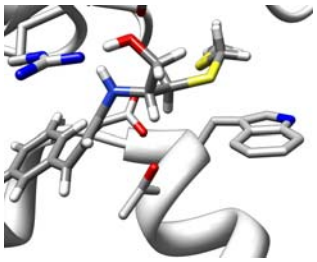
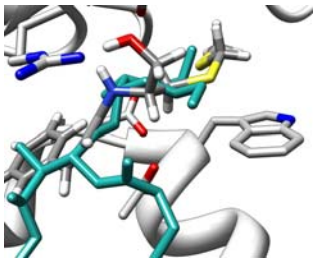
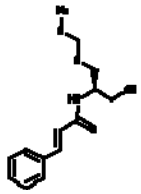
ZINC1P16

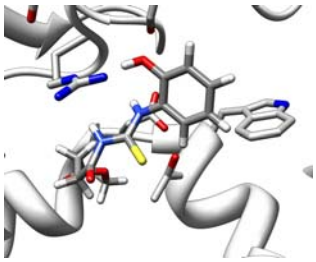
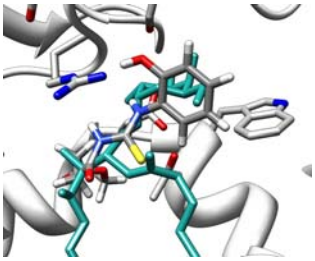
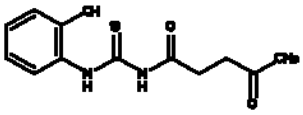
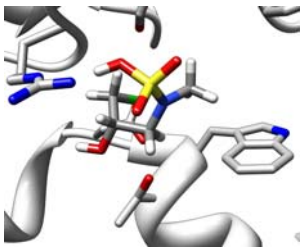
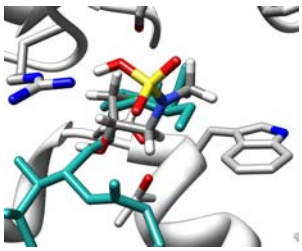
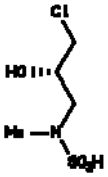
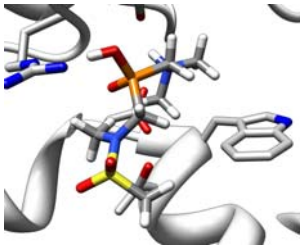
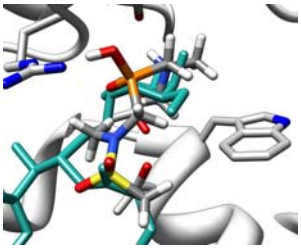
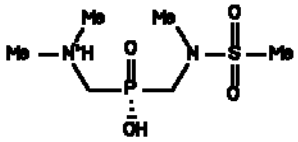
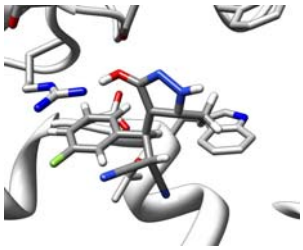
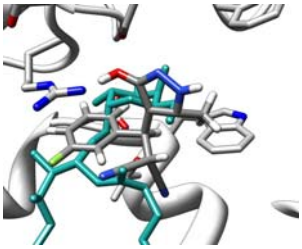
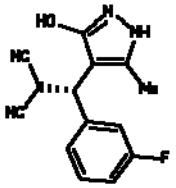
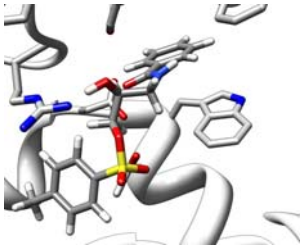
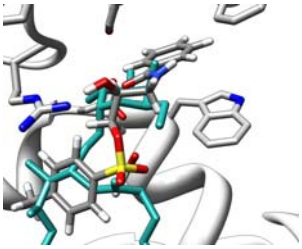
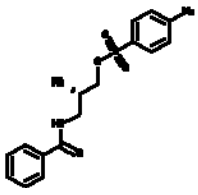
5000.TOP60	IMAGE	with α-GalCer	2D image
Z.02219133			
Z.02700445			
Z.05190634			
Z.02356827			
Z.06183050			

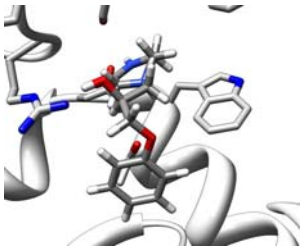
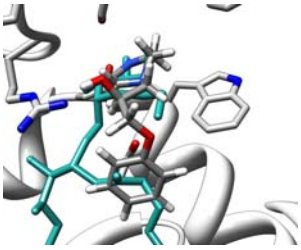
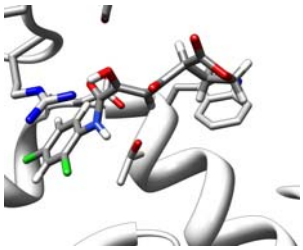
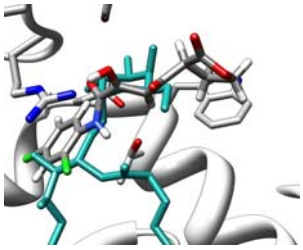
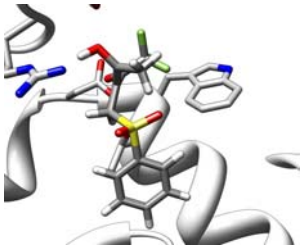
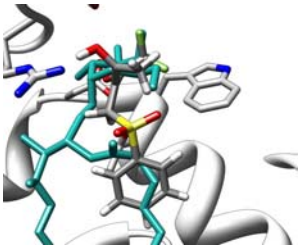
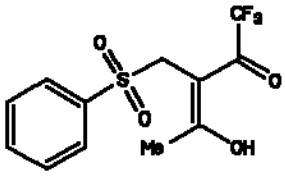
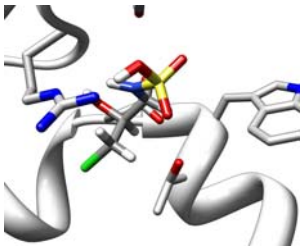
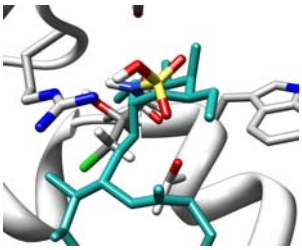
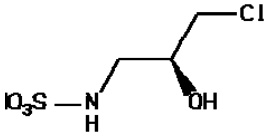
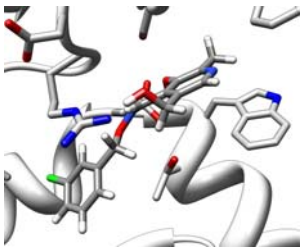
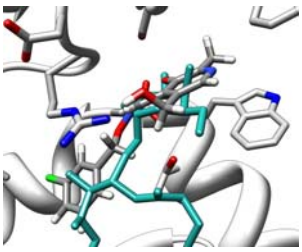
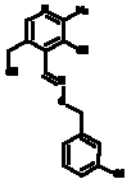
Z.05234587			
Z.01389759			
Z.01399235			
10000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.03156759			
Z.05048410			

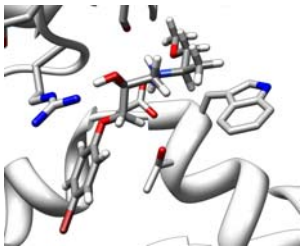
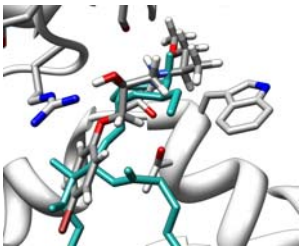
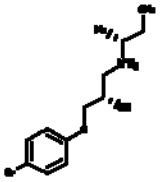
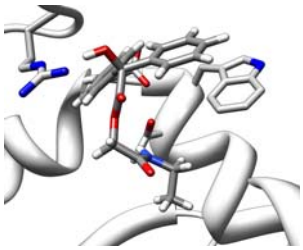
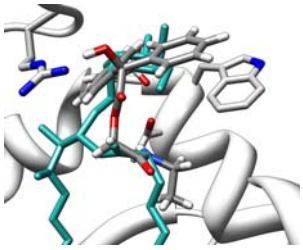
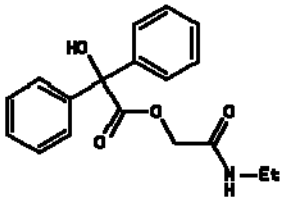
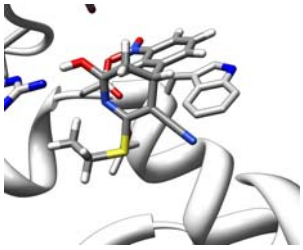
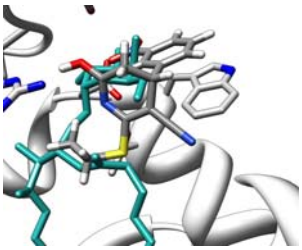
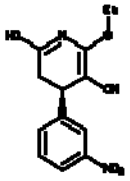
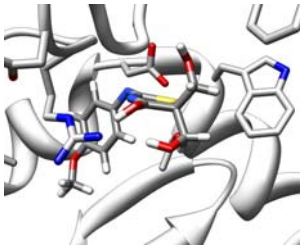
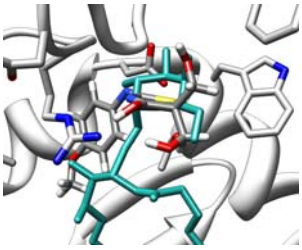
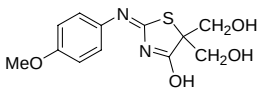
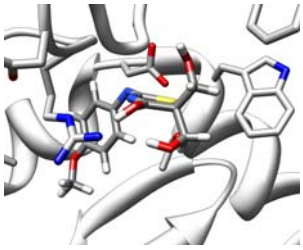
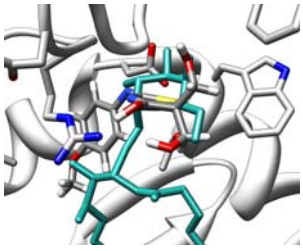
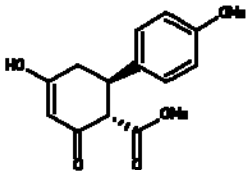
Z.05942173			
Z.03156074			
Z.04955814			
Z.02299332			
15000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.03984800			

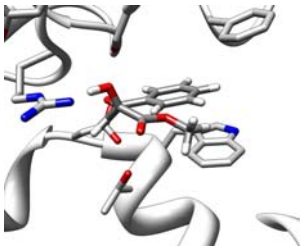
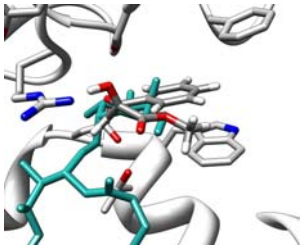
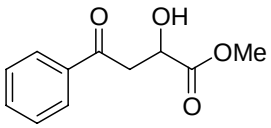
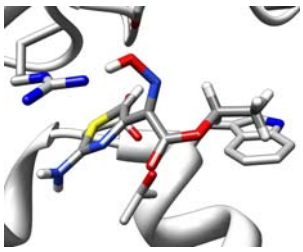
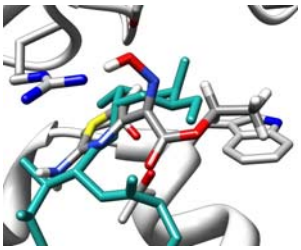
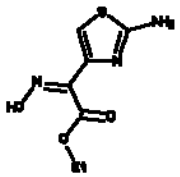
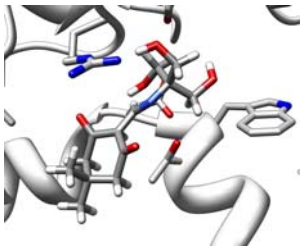
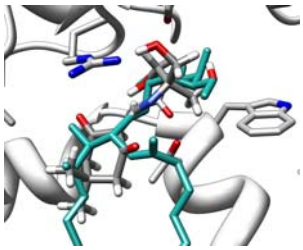
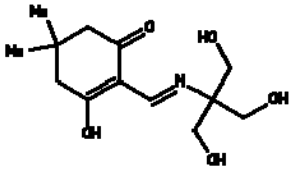
Z.03175591			
Z.05940996			
Z.02684805			
Z.06313417			
Z.05659888			

Z.05551794			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.03206142			
Z.06332182			
25000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.02522335			
Z.01731101			

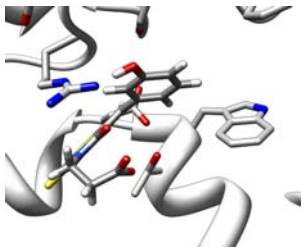
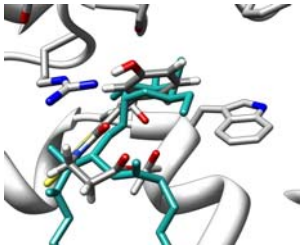
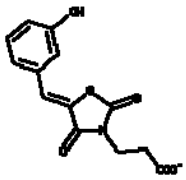
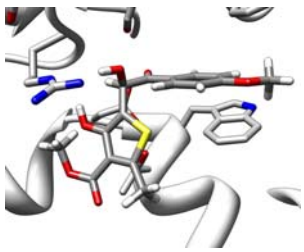
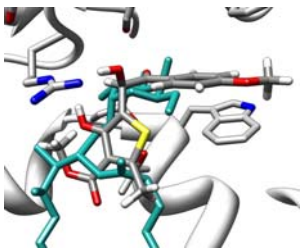
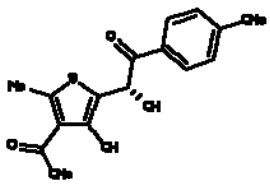
Z.06388665			
30000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
ZINC1P17			
5000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.02522333			
Z.04482338			
Z.00295860			
Z.04071748			

10000.TOP60	IMAGE	with α-GalCer	2D image
Z.03305607			
Z.05605947			
Z.04024524			
Z.04342764			
Z.01579045			

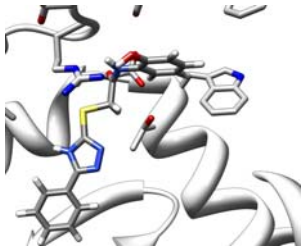
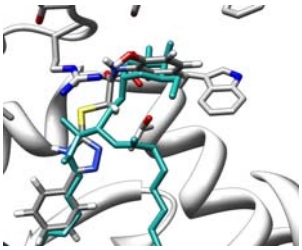
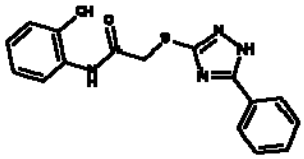
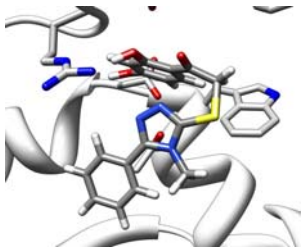
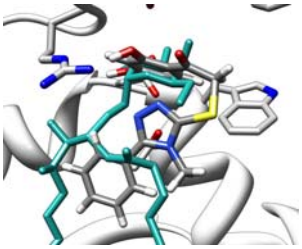
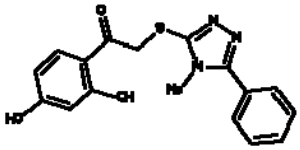
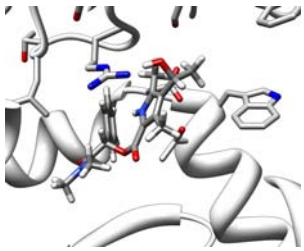
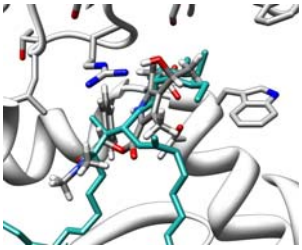
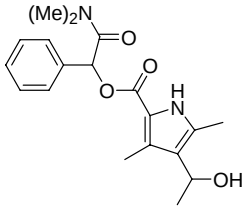
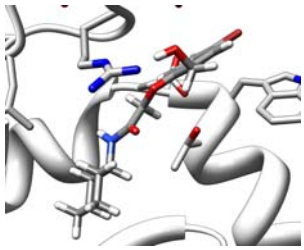
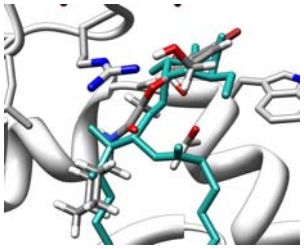
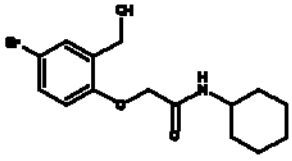
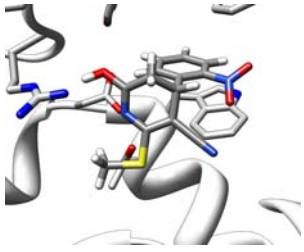
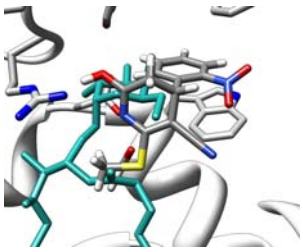
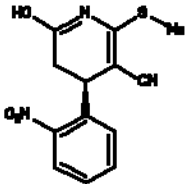
Z.03337956			
15000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.03555248			
Z.06265697			
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.01267266			
25000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.00249293			

Z.00085407			
Z.00056639			
Z.00027961			

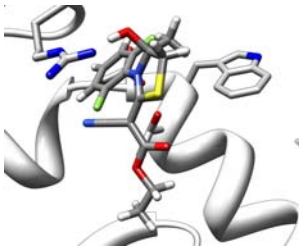
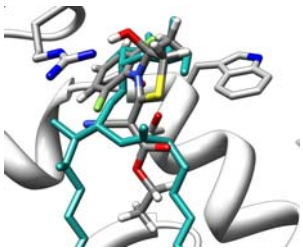
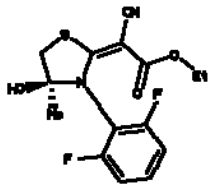
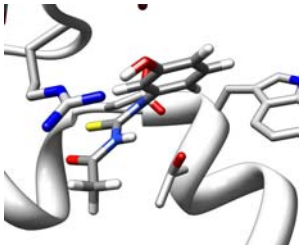
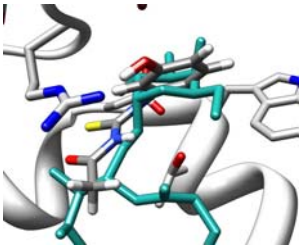
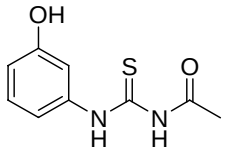
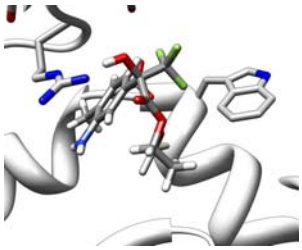
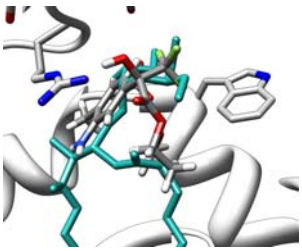
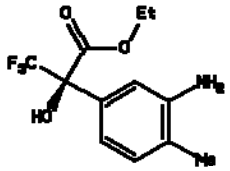
ZINC1P18

5000.TOP60	IMAGE	with α-GalCer	2D image
Z.01210154			
Z.06667795			

10000.TOP60	IMAGE	with α-GalCer	2D image

Z.06701224			
15000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
25000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.03300480			
Z.06846038			
Z.03458052			
Z.06411677			
30000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**

ZINC1P19

5000.TOP60	IMAGE	with α-GalCer	2D image
Z.00076585			
Z.06905313			
10000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
15000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
20000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
25000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**
Z.07078780			
30000.TOP60	**IMAGE**	**with α-GalCer**	**2D image**

Appendix C: Batch scripts for Efficient HTD Simulation Processing

Simple computing scripts were written by the author for this project over the course of its time in order to improve the efficiency of the high-throughput simulation and virtual screening studies performed. A typical simulation project entails simulating 1 or 2 protein systems making rewriting the necessary input files each time not a huge inconvenience. However, upon attempting to simulate upwards of 15 complexes with the overall trajectory split into 0.5 ns times intervals, or when a 2,000,000 large library where individual groups of 30,000 compounds that were split into smaller groups of 5,000 had to be docked simultaneously, the only logical recourse was to write programs that would be able to write and submit the jobs without human intervention. The methodology needed to have the first file manually written, but then the programs would copy and replace text within it for the other files along with permitting some user intervention. These batch scripts and processing scripts are included to share with others who are keen on making any of their computational studies more efficient. Furthermore, the Courier Font with smaller typeface was used to differentiate programs from the included header text.

MOLECULAR DYNAMICS SIMULATION FILES

INPUT FILES INCLUDING SIMULATION PARAMETERS

To determine what the shorthand notation for the various parameters please see the AMBER v10 manual.

The initial minimization of explicit water molecules with protein held rigid *parameter file*:

```
Initial minimisation of water in our complex
  &cntrl
    imin=1, maxcyc=1000, ncyc=500,
    cut=16, ntb=1, ntr=1,
    ntx=1,
 /
 Hold the Complex Fixed
 500.0
 RES 1 403
 END
```

The second minimization of explicit water molecules with protein allowed to fluctuate *parameter file*:

```
 Initial minimisation of water in our complex
  &cntrl
    imin=1, maxcyc=5000, ncyc=1000,
    cut=16, ntb=1, ntr=0,
    ntx=1,
 /
```

The initial molecular dynamics equilibration with protein held rigid *parameter file*:

```
Initial MD equilibration
  &cntrl
   imin = 0, ntb = 1, ntr =1, ntc = 2, ntf =2,
   igb = 0, ntpr = 100, ntwx = 100,
   ntt = 3, gamma_ln = 1.0,
   tempi = 0.0, temp0 = 300.0
   nstlim = 12500, dt = 0.002,
   cut = 16.0, nscm = 1000,
  /
 HOLD COMPLEX FIXED
 5.0
```

```
RES 1 403
END
```

The second molecular dynamics equilibration with protein allowed to fluctuate *parameter file*:

```
Initial MD equilibration
  &cntrl
   imin = 0, ntb = 2, ntc = 2, ntf =2,
   ntp = 2, taup = 5.0,
   igb = 0, ntpr = 100, ntwx = 100,
   ntt = 3, gamma_ln = 1.0,
   tempi = 300.0, temp0 = 300.0
   nstlim = 12500, dt = 0.002,
   cut = 16.0, nscm = 1000
  /
```

The first of the molecular dynamics performed at 300 K for 0.25 ns with each subsequent time interval having the time frame changed *parameter file*:

```
Initial MD equilibration
  &cntrl
   imin = 0, ntb = 2, ntc = 2, ntf =2,
   ntp = 2, taup = 5.0,
   igb = 0, ntpr = 100, ntwx = 100,
   ntt = 3, gamma_ln = 1.0,
   tempi = 300.0, temp0 = 300.0
   nstlim = 250000, dt = 0.002,
   cut = 16.0, nscm = 1000
  /
```

A sample batch script for molecular dynamics simulations where the INPCRD is set to the previous run's output:

```
#AMBER Example Batch Script
#
#PBS -N Z1_md1
#PBS -j oe
#PBS -m ae
#PBS -M jnadas@chemistry.ohio-state.edu
#PBS -l walltime=20:00:00
#PBS -l nodes=4:ppn=2
#PBS -S /bin/csh
set echo
setenv AMBERHOME /usr/local/amber/amber10
#
# The path below may need to be changed
set WORK=/nfs/proj03/PAS0248/ZINCMD/Z1
cd $TMPDIR
```

```
 #
 # The file names below may need to be changed
 set MDIN=md1.in
 set PRMTOP=complex.prmtop
 set INPCRD=complex_min_wat2.crd
 set MDOUT=complex_md1.out
 set RESTRT=complex_md1.rst
 set MDCRD=complex_md1.mdcrd
 set MDINFO=mdinfo
 set MDVEL=mdvel
 set MDEN=mden
 pbsdcp $WORK/$MDIN $TMPDIR
 pbsdcp $WORK/$PRMTOP $TMPDIR
 pbsdcp $WORK/$INPCRD $TMPDIR
 #
 # Periodic, explicit solvent simulations using the paricle-mesh Ewald
(PME)
 # method may perform better using the pmemd program; change sander to
pmemd.
 # Some applications may need the -O option for the sander/pmemd
command below:

 mpiexec -verbose $AMBERHOME/exe/pmemd -O -i $MDIN -o $MDOUT -inf
$MDINFO -p $PRMTOP -c $INPCRD -ref $INPCRD -x $MDCRD -v $MDVEL -e $MDEN
-r $RESTRT

 ls -al
 cp $MDOUT $MDINFO $MDCRD $MDVEL $MDEN $RESTRT $WORK
 exit
```

PROCESSING FILES IN THE SIMULATION STUDIES

The **simultaneous submission script** for MD simulations where values were asked from the user on where to start if for instance the number was different than 1 (startcount), the amount of simulations to process (quantity), how much to increment by which was usually 1 (incrementby), what file, batch script, was being processed (file), what was the associated input parameter file (mdin), was the original written file needed to be submitted (SUBMIT? Z1 Y1/N2), where are these files located (path RP/___/Z1), and do these files need to be submitted at a later time (WAIT Y1/N2):

```
 #!/bin/sh

 #Usage $0 <startcount> <quantity> <incrementby>
```

```
  NUMARGS=$#
  if [ ${NUMARGS} -lt 8 ]
  then
     echo "Usage $0 <startcount> <quantity> <incrementby> <file> <mdin>
<SUBMIT? Z1 Y1/N2> <path RP/___/Z1> <WAIT Y1/N2>"
     exit 1
  fi
  RP="/nfs/proj03/PAS0248"
  STARTCOUNT=${1}
  QUANTITY=${2}
  INCREMENTBY=${3}
  FILE=${4}
  MDIN=${5}
  SUBMIT=${6}
  PATH=${7}
  WAIT=${8}
  COUNTER=0
  RGN=${STARTCOUNT}
  QUERY="FALSE"

    if [ ${SUBMIT} -eq 1 ]
     then
        if [ ${WAIT} -eq 1 ]
           then
             echo "WAIT to Submit? yearmonthdaymilitarytime"
             read TIME
             /usr/local/torque-2.1.8/bin/qsub -a $TIME
${RP}/${PATH}/Z1/${FILE}
        else /usr/local/torque-2.1.8/bin/qsub ${RP}/${PATH}/Z1/${FILE}
       fi
    fi

  while [ ${COUNTER} -lt ${QUANTITY} ]
  do
  echo ' '
  echo ${COUNTER}'......'
  #   echo $QUERY
  #   echo ${RUNNINGNUMBER}

        cd ${RP}/${PATH}/Z1
        /bin/rm -r $RP/$PATH/Z$RGN/$FILE
        /bin/sed s/Z1/Z${RGN}/g ${FILE} > $RP/$PATH/Z${RGN}/${FILE}
        /bin/cp -r ${MDIN} $RP/$PATH/Z${RGN}/${MDIN}
        echo $RP/$PATH/Z$RGN/$FILE
        if [ ${WAIT} -eq 1 ]
           then
             echo "WAIT to Submit? yearmonthdaymilitarytime"
             read TIME2
             /usr/local/torque-2.1.8/bin/qsub -a $TIME2
${RP}/${PATH}/Z${RGN}/${FILE}
        else /usr/local/torque-2.1.8/bin/qsub
${RP}/${PATH}/Z${RGN}/${FILE}
        fi
```

```
        let RGN=${RGN}+${INCREMENTBY}
        let COUNTER=${COUNTER}+1
        let DONE=${QUANTITY}-1

  done
  exit 0
```

The simulation output generally contained gigabytes worth of material so it was easier to transfer files to the home workstation by **changing stop points (.rst files) into pdb files** in order to observe the progress of the simulation as well as the evolution of the proteins' motions over the course of the simulation. This also required the stripping of the waters which was necessary since they numbered in the tens of thousands. Much like before values were asked from the user on where to start if for instance the number was different than 1 (startcount), the amount of simulations to process (quantity), how much to increment by which was usually 1 (incrementby), since the simulation was divided into 0.5 ns time intervals the correlating integer to that was asked for (MD#?), what type of simulation was being changed into a pdb file (file=gly/ligand/complex), at the time the full complex was also being simulated in conjunction with the truncated complexes and since it had a different location the user was asked whether those files needed to be processed (MD_CD1D Y1/N2), and where are these files located (path RP/___/Z1):

```
  #!/bin/sh
  #Usage $0 <startcount> <quantity> <incrementby>


  NUMARGS=$#
  if [ ${NUMARGS} -lt 7 ]
  then
     echo "Usage $0 <startcount> <quantity> <incrementby> <MD#?>
<file=gly/ligand/complex> <MD_CD1D Y1/N2> <path RP/___/Z1> "
     exit 1
  fi

  RP="/nfs/proj03/PAS0248"
  STARTCOUNT=${1}
  QUANTITY=${2}
  INCREMENTBY=${3}
  MD=${4}
```

```
 FILE=${5}
 CD1D=${6}
 PATH=${7}
 COUNTER=0
 RGN=${STARTCOUNT}
 QUERY="FALSE"
 PLUSONE=1
 let PLUSONE=${MD}+1

 while [ ${COUNTER} -lt ${QUANTITY} ]
 do
 echo ' '
 echo ${COUNTER}'......'
 #   echo $QUERY
 #   echo ${RUNNINGNUMBER}

          cd ${RP}/${PATH}/Z${RGN}/
              echo "${RP}/${PATH}/Z${RGN}/"${FILE}"_md"${MD}".rst"
              /bin/rm -r ${FILE}_md${MD}.pdb
              /usr/local/amber/amber10/exe/ambpdb -p ${FILE}.prmtop
<${FILE}_md${MD}.rst> ${FILE}_md${MD}.pdb
              TIME=$(/bin/awk '{ if ( $1 == "begin" && $2 == "time" )
print $8 }' ${FILE}_md${PLUSONE}.out)
              /bin/rm -r  *_md${MD}_NOWAT_*
              /bin/awk '( $4 != "WAT" )' ${FILE}_md${MD}.pdb >>
${FILE}_md${MD}_NOWAT_${TIME}ps.pdb
              echo ${FILE}_md${MD}_NOWAT_${TIME}ps.pdb

       let RGN=${RGN}+${INCREMENTBY}
       let COUNTER=${COUNTER}+1
       let DONE=${QUANTITY}-1

 done

    if [ ${CD1D} -eq 1 ]
       then
          cd ${RP}/AGALCD1D/MD_CD1D/
              echo ' '
              echo 'END......'
              echo "AGALC1D/MD_CD1D/complex_md"${MD}".rst"
 # Make PDB FILE
              /bin/rm -r complex_md${MD}.pdb
              /usr/local/amber/amber10/exe/ambpdb -p complex.prmtop
<complex_md${MD}.rst> complex_md${MD}.pdb
 # Make Waterless PDB FILE
               TIME=$(/bin/awk '{ if ( $1 == "begin" && $2 == "time" )
print $8 }' complex_md${PLUSONE}.out)
               /bin/rm -r  *_md${MD}_NOWAT_*
               /bin/awk '( $4 != "WAT" )' complex_md${MD}.pdb >>
complex_md${MD}_NOWAT_${TIME}ps.pdb
               /bin/awk '( $4 != "WAT" && $4 != "AGH" )'
complex_md${MD}.pdb >> complex_md${MD}_NOWAT_NOAGH_${TIME}ps.pdb
              echo complex_md${MD}_NOWAT_${TIME}ps.pdb
              echo complex_md${MD}_NOWAT_NOAGH_${TIME}ps.pdb
          cd ${RP}/AGALCD1D/
```

```
    fi
exit 0
```

As mentioned above, the simulation output contained gigabytes worth of material, and this trajectory data was submitted to statistical processing using the PTRAJ functionality of AMBER to determine the RMSD and hydrogen bonding data of the system. The trajectory files with explicit water were extremely bulky and as **such the water molecules were stripped from them**. The input values from the user included where to start if for instance the number was different than 1 (startcount), the amount of simulations to process (quantity), the trajectory was analyzed at certain time intervals such as 1 ns, 3 ns, or 10 ns depending (start of MD? and amount of MD?), how much to increment by which was usually 1 (incrementby), what was the total nanoseconds achieved by the simulation (totalNS?#), what type of simulation was being changed into a pdb file (file=gly/ligand/complex), since the copying and writing of the parameter files occasionally caused issues a wait on the actual statistical analysis was allowed (ptraj Y1/N2), and where are these files located (path RP/___/Z1)):

```
#!/bin/sh
#Usage $0 <startcount> <quantity> <incrementby>


NUMARGS=$#
if [ ${NUMARGS} -lt 8 ]
then
    echo "Usage $0 <startcount> <quantity> <startofMD?> <amountofMD?>
<incrementby> <file=gly/ligand/complex> <PTRAJ Y1/N2> <path RP/___> "
    exit 1
fi

RP="/nfs/proj03/PAS0248"
STARTCOUNT=${1}
QUANTITY=${2}
CNT2=${3}
QT2=${4}
INCREMENTBY=${5}
FILE=${6}
```

```
 PTRAJ=${7}
 PATH=${8}
 COUNTER=0
 RGN=${STARTCOUNT}
 RGN2=${CNT2}
 RST=${CNT2}
 QUERY="FALSE"
 PLUSONE=1
 let QT3=${CNT2}+${QT2}

 echo ' '
 echo
".....................................BEGIN.........................
..........."

 while [ ${COUNTER} -lt ${QUANTITY} ]
 do
 echo ' '
 echo ${COUNTER}'......'
 #   echo $QUERY
 #   echo ${RUNNINGNUMBER}

          cd ${RP}/${PATH}/Z${RGN}/
          /bin/rm -r ptraj_WATSTRIP
     while [ ${CNT2} -lt ${QT3} ]
       do
              echo "${RP}/${PATH}/Z${RGN}/"${FILE}"_md"${RGN2}".mdcrd"
              /usr/bin/printf "trajin ${FILE}_md${RGN2}.mdcrd\ncenter
:1-183\nimage center familiar\nstrip :WAT\nrms first out
${FILE}_md${RGN2}_rms.out:3_183@CANOWAT\ntrajout
${FILE}_md${RGN2}_watless.mdcrd nobox\n" > ptraj_WATSTRIP
 #             /usr/bin/printf "\n" > ptraj_WATSTRIP
              /usr/bin/tail -n1000 ptraj_WATSTRIP
     if [ ${PTRAJ} -eq 1 ]
       then
         /usr/local/amber/amber10/exe/ptraj ${FILE}.prmtop
ptraj_WATSTRIP
       fi

       let CNT2=${CNT2}+1
       let RGN2=${RGN2}+1
     done


       CNT2=${RST}
       RGN2=${RST}
       let RGN=${RGN}+${INCREMENTBY}
       let COUNTER=${COUNTER}+1
       let DONE=${QUANTITY}-1

 done


 exit 0
```

After stripping the water molecules from the trajectory files, both the water and non-water trajectories were analyzed **to determine the RMSD fluctuations** of the protein along with **creating the necessary binpos file that allowed for hydrogen bond occupations to be calculated**. The input values from the user included where to start if for instance the number was different than 1 (startcount), the amount of simulations to process (quantity), the trajectory was analyzed at certain time intervals such as 1 ns, 3 ns, or 10 ns depending (start of MD? and amount of MD?), how much to increment by which was usually 1 (incrementby), what was the total nanoseconds achieved by the simulation (totalNS?#), what type of simulation was being changed into a pdb file (file=gly/ligand/complex), since the copying and writing of the parameter files occasionally caused issues a wait on the actual statistical analysis was allowed (ptraj Y1/N2), and where are these files located (path RP/___/Z1):

```
#!/bin/sh
#Usage $0 <startcount> <quantity> <incrementby>


NUMARGS=$#
if [ ${NUMARGS} -lt 9 ]
then
   echo "Usage $0 <startcount> <quantity> <startofMD?> <amountofMD?>
<incrementby> <totalNS? #> <file=gly/ligand/complex> <PTRAJ Y1/N2>
<path RP/___/Z1> "
   exit 1
fi

RP="/nfs/proj03/PAS0248"
STARTCOUNT=${1}
QUANTITY=${2}
CNT2=${3}
QT2=${4}
INCREMENTBY=${5}
NS=${6}
FILE=${7}
PTRAJ=${8}
PATH=${9}
COUNTER=0
RGN=${STARTCOUNT}
RGN2=${CNT2}
```

```
  RST=${CNT2}
  QUERY="FALSE"
  PLUSONE=1
  let QT3=${CNT2}+${QT2}
  let QT4=${QT3}-1

  echo ' '
  echo
".....................................BEGIN.........................
..........."

  while [ ${COUNTER} -lt ${QUANTITY} ]
  do
  echo ' '
  echo ${COUNTER}'......'
  #   echo $QUERY
  #   echo ${RUNNINGNUMBER}

           cd ${RP}/${PATH}/Z${RGN}/
           /bin/rm -r ptraj_ANALYZE
           /bin/rm -r ptraj_ANALYZE2
      while [ ${CNT2} -lt ${QT3} ]
        do
               echo "${RP}/${PATH}/MD${RGN}/"${FILE}"_md"${RGN2}".mdcrd"
                /usr/bin/printf "trajin
${FILE}_md${RGN2}_watless.mdcrd\n" >> ptraj_ANALYZE
                /usr/bin/printf "trajin ${FILE}_md${RGN2}.mdcrd\n" >>
ptraj_ANALYZE2
        let CNT2=${CNT2}+1
        let RGN2=${RGN2}+1
      done
  # HBOND & RMS analysis for Waterless System
      /usr/bin/printf "center :1-179\nimage center familiar\n" >>
ptraj_ANALYZE
      /usr/bin/printf "rms first out complex_rms_${NS}ns_ALL.out :3-
174@CA\n" >> ptraj_ANALYZE
      /usr/bin/printf "rms first out complex_rms_${NS}ns_A1.out :54-
83@CA\n" >> ptraj_ANALYZE
      /usr/bin/printf "rms first out complex_rms_${NS}ns_A1_R.out :54-
71@CA\n" >> ptraj_ANALYZE
      /usr/bin/printf "rms first out complex_rms_${NS}ns_A1_L.out :71-
83@CA\n" >> ptraj_ANALYZE
      /usr/bin/printf "rms first out complex_rms_${NS}ns_A2.out :135-
172@CA\n" >> ptraj_ANALYZE
      /usr/bin/printf "rms first out complex_rms_${NS}ns_A2_L.out :135-
146@CA\n" >> ptraj_ANALYZE
      /usr/bin/printf "rms first out complex_rms_${NS}ns_A2_R.out :146-
172@CA\n" >> ptraj_ANALYZE
      /usr/bin/printf "rms first out complex_rms_${NS}ns_A2_M.out :144-
149@CA\n" >> ptraj_ANALYZE
      /usr/bin/printf "rms first out complex_rms_${NS}ns_ALL.out :3-
373@CA\n" >> ptraj_ANALYZE
      /usr/bin/printf "rms first out complex_rms_${NS}ns_B2M.out :276-
373@CA\n" >> ptraj_ANALYZE
```

```
     /usr/bin/printf "rms first out complex_rms_${NS}ns_BOTTOM.out
:179-373@CA\n" >> ptraj_ANALYZE
     /usr/bin/printf "rms first out complex_rms_${NS}ns_GROOVE.out :3-
179@CA\n" >> ptraj_ANALYZE
     /usr/bin/printf "rms first out complex_rms39_TCR.out :375-
821@CA\n" >> ptraj_ANALYZE
     /usr/bin/printf "rms first out complex_rms39_TCRA.out :375-
486@CA\n" >> ptraj_ANALYZE
     /usr/bin/printf "rms first out complex_rms39_TCRB.out :578-
692@CA\n" >> ptraj_ANALYZE
     /usr/bin/printf "rms first out complex_rms39_TCRCA.out :487-
577@CA\n" >> ptraj_ANALYZE
     /usr/bin/printf "rms first out complex_rms39_TCRCB.out :693-
821@CA\n" >> ptraj_ANALYZE
     /usr/bin/printf "trajout complex_md${RST}_${QT4}_${NS}ns.binpos
binpos\n" >> ptraj_ANALYZE
  # HBOND & RMS analysis for WATER
     /usr/bin/printf "center :1-179\nimage center familiar\n" >>
ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_ALL.out :3-
174@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_A1.out :54-
83@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_A1_R.out :54-
71@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_A1_L.out :71-
83@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_A2.out :135-
172@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_A2_L.out :135-
146@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_A2_R.out :146-
172@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_A2_M.out :144-
149@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_ALL.out :3-
373@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_B2M.out :276-
373@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_BOTTOM.out
:179-373@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms_${NS}ns_GROOVE.out :3-
179@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms39_TCR.out :375-
821@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms39_TCRA.out :375-
486@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms39_TCRB.out :578-
692@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms39_TCRCA.out :487-
577@CA\n" >> ptraj_ANALYZE2
     /usr/bin/printf "rms first out complex_rms39_TCRCB.out :693-
821@CA\n" >> ptraj_ANALYZE2
```

```
     /usr/bin/printf "trajout
complex_md${RST}_${QT4}_${NS}ns_WATER.binpos binpos\n" >>
ptraj_ANALYZE2


  echo "..............ptraj ANALYZE FILE............."
     /usr/bin/tail -n1000 ptraj_ANALYZE
  echo
  echo "..............ptraj WATER ANALYZE FILE............."
     /usr/bin/tail -n1000 ptraj_ANALYZE2
  echo


    if [ ${PTRAJ} -eq 1 ]
        then
  #        echo "............ You are analyzing NONwater file
............"
  #        /usr/local/amber/amber10/exe/ptraj ${FILE}2.prmtop
ptraj_ANALYZE
           echo "............ You are analyzing WATER file ............"
          /usr/local/amber/amber10/exe/ptraj ${FILE}.prmtop
ptraj_ANALYZE2
        fi

        CNT2=${RST}
        RGN2=${RST}
        let RGN=${RGN}+${INCREMENTBY}
        let COUNTER=${COUNTER}+1
        let DONE=${QUANTITY}-1

  done

  exit 0
```

The last and most tedious script written was for the **analysis of the hydrogen bonds**. The parameter file was simple to create for the proteins since the same complex was used in all the simulations, however, the atom names of the different ligands had to be sometimes manually inputted. The input values from the user included where to start if for instance the number was different than 1 (startcount), the amount of simulations to process (quantity), how much to increment by which was usually 1 (incrementby), what was the total nanoseconds achieved by the simulation (totalNS?#), the name of the hydrogen bond data file (nameofbinposFILE), what type of simulation was being changed into a pdb file (file=gly/ligand/complex), since the copying and writing of the

parameter files occasionally caused issues a wait on the actual statistical analysis was allowed (PTRAJ Y1/N2), and where are these files located (path RP/___/Z1), and whether there were special hydrogen bond acceptor or donor cases (Special Cases? Y1/N2):

```
  #!/bin/sh
  #Usage $0 <startcount> <quantity> <incrementby>


  NUMARGS=$#
  if [ ${NUMARGS} -lt 9 ]
  then
     echo "Usage $0 <startcount> <quantity> <incrementby> <totalNS? #>
<nameofbinposFILE> <gly/ligand/complex> <PTRAJ Y1/N2> <path RP/___/Z1>
<Special Cases? Y1/N2>  "
     exit 1
  fi

  RP="/nfs/proj03/PAS0248"
  STARTCOUNT=${1}
  QUANTITY=${2}
  INCREMENTBY=${3}
  NS=${4}
  FILE=${5}
  FILE2=${6}
  PTRAJ=${7}
  PATH=${8}
  SC=${9}
  COUNTER=0
  RGN=${STARTCOUNT}
  RGN2=${CNT2}
  RST=${CNT2}
  QUERY="FALSE"
  PLUSONE=1

  echo ' '
  echo ".............................PTRAJ HBOND
ANALYSIS............................."
  echo
".....................................BEGIN.........................
..........."

  while [ ${COUNTER} -lt ${QUANTITY} ]
  do
  echo ' '
  echo ${COUNTER}'......'
  #   echo $QUERY
  #   echo ${RUNNINGNUMBER}
           cd $RP/$PATH/Z${RGN}
  #         cd ${RP}/${PATH}/MD${RGN}/
           /bin/rm -r ptraj_HBOND
```

```
          /bin/rm -r ptraj_HBOND2
          /bin/rm -r ptraj_HBOND3
              echo "${RP}/${PATH}/Z${RGN}/ ....... HBOND FILE CREATION"
 #DONORS
               /usr/bin/printf '#HBOND Analysis\n'  >> ptraj_HBOND
               /usr/bin/printf "\ntrajin $FILE.binpos\n\n" >>
ptraj_HBOND
               /usr/bin/printf '#-- Special SUGAR Amino-Acid Donors\n'
>> ptraj_HBOND
               /usr/bin/printf '\n# Asp80\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :75@OD1\n" >> ptraj_HBOND
               /usr/bin/printf "donor mask :75@OD2\n" >> ptraj_HBOND
               /usr/bin/printf '# Ser76\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :71@OG\n" >> ptraj_HBOND
               /usr/bin/printf '# Thr154\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :149@OG1\n" >> ptraj_HBOND
               /usr/bin/printf '# Asp151\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :146@OD1\n" >> ptraj_HBOND
               /usr/bin/printf "donor mask :146@OD2\n" >> ptraj_HBOND
               /usr/bin/printf '# Ser30\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :207@OG\n" >> ptraj_HBOND
               /usr/bin/printf '#-- Special TCR Amino-Acid Donors\n' >>
ptraj_HBOND
               /usr/bin/printf '# Tyr48\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :336@OH\n" >> ptraj_HBOND
               /usr/bin/printf '# Tyr50\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :338@OH\n" >> ptraj_HBOND
               /usr/bin/printf '# Glu55\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :344@OE1\n" >> ptraj_HBOND
               /usr/bin/printf "donor mask :344@OE2\n" >> ptraj_HBOND
               /usr/bin/printf '#-- Special CDR3a LOOP Amino-Acid
Donors\n' >> ptraj_HBOND
               /usr/bin/printf '# Asp94\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :270@OD1\n" >> ptraj_HBOND
               /usr/bin/printf "donor mask :270@OD2\n" >> ptraj_HBOND
               /usr/bin/printf "donor mask :270@O\n" >> ptraj_HBOND
               /usr/bin/printf '# Arg95\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :271@O\n" >> ptraj_HBOND
               /usr/bin/printf '# Gly96\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :272@O\n" >> ptraj_HBOND
               /usr/bin/printf '# Ser97\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :273@O\n" >> ptraj_HBOND
               /usr/bin/printf "donor mask :273@OG\n" >> ptraj_HBOND
               /usr/bin/printf '# Thr98\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :274@O\n" >> ptraj_HBOND
               /usr/bin/printf "donor mask :274@OG1\n" >> ptraj_HBOND
               /usr/bin/printf '# Leu99\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :275@O\n" >> ptraj_HBOND
               /usr/bin/printf '# Gly100\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :276@O\n" >> ptraj_HBOND
               /usr/bin/printf '# Arg101\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :277@O\n" >> ptraj_HBOND
               /usr/bin/printf '# Leu102\n' >> ptraj_HBOND
               /usr/bin/printf "donor mask :278@O\n" >> ptraj_HBOND
               /usr/bin/printf '# Tyr103\n' >> ptraj_HBOND
```

```
                /usr/bin/printf "donor mask :279@O\n" >> ptraj_HBOND
                /usr/bin/printf "donor mask :279@OH\n" >> ptraj_HBOND
                /usr/bin/printf '#-- Special CD1D Amino-Acid Donors\n'
>> ptraj_HBOND
                /usr/bin/printf '# Glu83\n' >> ptraj_HBOND
                /usr/bin/printf "donor mask :78@OE1\n" >> ptraj_HBOND
                /usr/bin/printf "donor mask :78@OE2\n" >> ptraj_HBOND
                /usr/bin/printf '# aGalCer AGH\n' >> ptraj_HBOND
  #ACCEPTORS
                /usr/bin/printf '\n\n#-- Special Amino-Acid Acceptors\n'
>> ptraj_HBOND2
                /usr/bin/printf '\n# Ser76\n'  >> ptraj_HBOND2
                /usr/bin/printf 'acceptor mask  :71@OG  :71@HG\n'  >>
ptraj_HBOND2
                /usr/bin/printf '# Thr154\n'  >> ptraj_HBOND2
                /usr/bin/printf 'acceptor mask  :149@OG1 :149@HG1\n'  >>
ptraj_HBOND2
                /usr/bin/printf '# Ser30\n' >> ptraj_HBOND
                /usr/bin/printf "acceptor mask :207@OG :207HG\n" >>
ptraj_HBOND
                /usr/bin/printf '#-- Special TCR Amino-Acid Acceptors\n'
>> ptraj_HBOND2
                /usr/bin/printf '# Tyr48\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :336@OH :336@HH\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Tyr50\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :338@OH :338@HH\n" >>
ptraj_HBOND2
                /usr/bin/printf '#-- Special CDR3a LOOP Amino-Acid
Acceptors\n' >> ptraj_HBOND2
                /usr/bin/printf '# Asp94\n' >> ptraj_HBOND
                /usr/bin/printf "acceptor mask :270@N :270@H\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Arg95\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :271@N :271@H\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :271@NE :271@HE\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :271@NH1 :271@1HH1\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :271@NH1 :271@2HH1\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :271@NH2 :271@1HH2\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :271@NH2 :271@2HH2\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Gly96\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :272@N :272@H\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Ser97\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :273@OG :273@HG\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :273@N :273@H\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Thr98\n' >> ptraj_HBOND2
```

```
                /usr/bin/printf "acceptor mask :274@OG1 :274@HG1\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :274@N :274@H\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Leu99\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :275@N :275@H\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Gly100\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :276@N :276@H\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Arg101\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :277@N :277@H\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :277@NE :277@HE\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :277@NH1 :277@1HH1\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :277@NH1 :277@2HH1\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :277@NH2 :277@1HH2\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :277@NH2 :277@2HH2\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Leu102\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :278@N :278@H\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Tyr103\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :279@OH :279@HH\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :279@N :279@H\n" >>
ptraj_HBOND2
                /usr/bin/printf '#-- Special CD1D Amino-Acid
Acceptors\n' >> ptraj_HBOND2
                /usr/bin/printf '# Lys86\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :81@NZ :81@HZ1\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :81@NZ :81@HZ2\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :81@NZ :81@HZ3\n" >>
ptraj_HBOND2
                /usr/bin/printf '# Arg79\n' >> ptraj_HBOND2
                /usr/bin/printf "acceptor mask :74@N :74@H\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :74@NE :74@HE\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :74@NH1 :74@1HH1\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :74@NH1 :74@2HH1\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :74@NH2 :74@1HH2\n" >>
ptraj_HBOND2
                /usr/bin/printf "acceptor mask :74@NH2 :74@2HH2\n" >>
ptraj_HBOND2
                /usr/bin/printf '# aGalCer AGH \n'  >> ptraj_HBOND2
```

```
  #DONOR/ACCEPTOR LIGAND Addition
              /bin/rm -r TEST
              /bin/awk '{ if ( $4 == "AGH" ) print $2 } '
complex_md1.pdb >> TEST
              CNT=$(/bin/awk '{ if ( NR == 1 ) print } ' TEST )
              END=$(/usr/bin/tail -n1 TEST)
              echo "AGH BEGINS @$CNT.......... and ends @$END"

      let END=${END}+1
         while [ ${CNT} -lt ${END} ]
            do
                ATOM=$(/bin/awk '{ if ( $2 == '$CNT' && $4 == "AGH" )
print $3 }' complex_md1.pdb)
              echo "$CNT.......$ATOM"
  #Donor Oxygen H's
                if [[ "$ATOM" == *O* && "$ATOM" != *H* ]]
                  then
                       echo $CNT"......."$ATOM"......DONOR Oxygen"
                       /usr/bin/printf "donor mask :AGH@$ATOM\n" >>
ptraj_HBOND
                fi
  #Donor Nitrogen H's
                if [[ "$ATOM" == *N* && "$ATOM" != *O* && "$ATOM" != *H*
&& "$ATOM" != *C* ]]
                then
                       echo $CNT"......."$ATOM"......DONOR Nitrogen"
                       /usr/bin/printf "donor mask :AGH@$ATOM\n" >>
ptraj_HBOND
                fi
  #Donor Sulfur H's
                if [[ "$ATOM" == *S* && "$ATOM" != *H* ]]
                  then
                       echo $CNT"......."$ATOM"......DONOR Sulfur"
                       /usr/bin/printf "donor mask :AGH@$ATOM\n" >>
ptraj_HBOND
                fi
  #Acceptor Oxygen H's
                if [[ "$ATOM" == *O* && "$ATOM" == *H* && "$ATOM" !=
"HOH" && "$ATOM" != "OH" ]]
                then
                       echo $CNT"......."$ATOM"......ACCEPTOR Oxygen H"
                       OXY=`echo $ATOM | /bin/sed 's/H//'`
                       /usr/bin/printf "acceptor mask :AGH@$OXY
:AGH@$ATOM\n" >> ptraj_HBOND2
                fi
  #Acceptor Nitrogen H's
               if [[ "$ATOM" == *N* && "$ATOM" == *H* && "$ATOM" != *O*
&& "$ATOM" != *C* ]]
                then
                       echo $CNT"......."$ATOM"......ACCEPTOR Nitrogen
H"
                       OXY=`echo $ATOM | /bin/sed 's/H//'`
                       /usr/bin/printf "acceptor mask :AGH@$OXY
:AGH@$ATOM\n" >> ptraj_HBOND2
                fi
```

```
  #Acceptor Sulfur H's
               if [[ "$ATOM" == *S* && "$ATOM" == *H* && "$ATOM" != *O*
&& "$ATOM" != *C* ]]
                  then
                       echo $CNT"......."$ATOM"......ACCEPTOR Sulfur H"
                       OXY=`echo $ATOM | /bin/sed 's/H//'`
                       /usr/bin/printf "acceptor mask :AGH@$OXY
:AGH@$ATOM\n" >> ptraj_HBOND2
                  fi

  #SPECIAL HOH - Oxygen
               if [[ "$ATOM" == "HOH" ]]
                  then
                       echo $CNT"......."$ATOM"......ACCEPTOR Oxygen
HOH"
                       /usr/bin/printf "donor mask :AGH@OH\n" >>
ptraj_HBOND
                       /usr/bin/printf "acceptor mask :AGH@OH
:AGH@HOH\n" >> ptraj_HBOND2
                  fi
  #SPECIAL aGalCer
                if [[ "$ATOM" == "HO5" || "$ATOM" == "HO2" ]]
                  then
                       echo $CNT"......."$ATOM"......ACCEPTOR aGalCer
HO5/HO2"
                       OXY=`echo $ATOM | /bin/sed 's/H//'`
                       /usr/bin/printf "acceptor mask :AGH@$OXY""A
:AGH@$ATOM\n" >> ptraj_HBOND2
                  fi

             let CNT=$CNT+1
           done

  #SPECIAL CASES?
           if [ ${SC} -eq 1 ]
               then
                  echo "HOW MANY DONORS?"
                  read NUM
                  CT=0
                    while [ ${CT} -lt $NUM ]
                    do
                     echo "DONOR ...."${CT}
                     read D${CT}
                     echo "donor mask :AGH@${DCT}"
                     /usr/bin/printf "donor mask :AGH@$DCT\n" >>
ptraj_HBOND
                  let CT=${CT}+1
                  done
                  CT=0
                  echo "HOW MANY ACCEPTORS?"
                  read NUM2
                  while [ ${CT} -lt $NUM2 ]
                    do
                     echo "ACCEPTOR .... (1) = ATOM [Hit ENTER] // (2)
Hydrogen [Hit Enter] ... (COUNTER) ... "${CT}
```

```
                    read D
                    read E
                    echo "acceptor mask :AGH@$D :AGH@$E"
                    /usr/bin/printf "acceptor mask :AGH@$D :AGH@$E\n"
>> ptraj_HBOND2

                let CT=${CT}+1
                done
          fi
#CREATION OF ptraj_HBOND file ANALYSIS of HBonds
              /bin/cat ptraj_HBOND ptraj_HBOND2 >> ptraj_HBOND3
              /usr/bin/printf '\n\n#-- HBOND ANALYSIS COMMAND\n' >>
ptraj_HBOND3
              /usr/bin/printf "\nhbond print .05 series hbt out
$FILE.dat\n" >> ptraj_HBOND3
              /usr/bin/printf "hbond distance 3.5 angle 120.0
solventneighbor 6 solventdonor WAT O series hbt out $FILE.WATER.dat\n"
>> ptraj_HBOND3

# HBOND analysis for WATER
#    /usr/bin/printf "trajout
complex_md${RST}_${QT4}_${NS}ns_WATER.binpos binpos\n" >>
ptraj_ANALYZE2
    echo
    echo "............................... FILE CREATED
................................."
    echo
    /usr/bin/tail -n1000 ptraj_HBOND3

#PAUSE TO INSPECT ptraj INPUT FILE
#read PAUSE


  if [ ${PTRAJ} -eq 1 ]
      then
#        /usr/local/amber/amber10/exe/ptraj ${FILE}2.prmtop
ptraj_ANALYZE
        /bin/rm -r $FILE.dat
        /bin/rm -r $FILE.WATER.dat
        /usr/local/amber/amber10/exe/ptraj ${FILE2}.prmtop ptraj_HBOND3
      fi

# MODIFY ATOM names >> RESIDUE NAMES
     /bin/rm -r FOO
     /bin/rm -r FOO.TMP
     /bin/rm -r $FILE.RESIDUES.dat
     /bin/rm -r $FILE.WATER.RESIDUES.dat

CNT3=0
QNT3=2

echo
echo "........................... MODIFY ATOMS -> RESIDUES
..........................."
echo
```

```
 while [ ${CNT3} -lt ${QNT3} ]
    do
     echo "......................... $CNT3"
     if [ ${CNT3} -eq 0 ]
       then
           echo "copying $FILE.dat to FOO"
           /bin/cp -r $FILE.dat FOO
       else
           echo "copying $FILE.WATER.dat to FOO"
           /bin/cp -r $FILE.WATER.dat FOO
     fi
# DONOR changes
     /bin/sed -e s/":75@OD1"/" Asp80 "/g FOO > FOO.TMP
     /bin/mv FOO.TMP FOO.TMP2
     /bin/sed -e s/":75@OD2"/" Asp80 "/g FOO.TMP2 > FOO.TMP
     /bin/mv FOO.TMP FOO.TMP2
     /bin/sed -e s/":71@OG"/" Ser76"/g FOO.TMP2 > FOO.TMP
     /bin/mv FOO.TMP FOO.TMP2
     /bin/sed -e s/":149@OG1"/" Thr154 "/g FOO.TMP2 > FOO.TMP
     /bin/mv FOO.TMP FOO.TMP2
     /bin/sed -e s/":146@OD1"/" Asp151 "/g FOO.TMP2 > FOO.TMP
     /bin/mv FOO.TMP FOO.TMP2
     /bin/sed -e s/":146@OD2"/" Asp151 "/g FOO.TMP2 > FOO.TMP
     /bin/mv FOO.TMP FOO.TMP2
# Acceptor changes
     /bin/sed -e s/":71@HG"/" Ser76"/g FOO.TMP2 > FOO.TMP
     /bin/mv FOO.TMP FOO.TMP2
     /bin/sed -e s/":149@HG1"/" Thr154 "/g FOO.TMP2 > FOO.TMP
     /bin/mv FOO.TMP FOO.TMP2
# AGH changes
     /bin/sed -e s/":178"/" LIG"/g FOO.TMP2 > FOO.TMP
     /bin/mv FOO.TMP FOO.TMP2
# FINAL MOVE
    if [ ${CNT3} -eq 0 ]
       then
           /bin/mv FOO.TMP2 $FILE.RESIDUES.dat
       else
           /bin/mv FOO.TMP2 $FILE.WATER.RESIDUES.dat
     fi
   let CNT3=$CNT3+1
done

# Delete TEMPORARY files
     /bin/rm -r ptraj_HBOND
     /bin/rm -r ptraj_HBOND2
     /bin/rm -r TEST
#     /bin/rm -r FOO
      CNT2=${RST}
      RGN2=${RST}
      let RGN=${RGN}+${INCREMENTBY}
      let COUNTER=${COUNTER}+1
      let DONE=${QUANTITY}-1

done
exit 0
```

AUTODOCK V3 PARAMETER FILES

GRID PARAMETER FILE

```
receptor complex_md2.pdbqs           # macromolecule
gridfld complex_md2.maps.fld         # grid_data_file
npts 92 76 90                        # num.grid points in xyz
spacing 0.375                        # spacing(A)
gridcenter 61.0 35.0 37.0            # xyz-coordinates or auto
types CANOSH                         # atom type names
smooth 0.5                           # store minimum energy w/in rad(A)
map complex_md2.C.map                # atom-specific affinity map
nbp_r_eps  4.00 0.0222750 12  6   # C-C lj
nbp_r_eps  3.75 0.0230026 12  6   # C-N lj
nbp_r_eps  3.60 0.0257202 12  6   # C-O lj
nbp_r_eps  4.00 0.0257202 12  6   # C-S lj
nbp_r_eps  3.00 0.0081378 12  6   # C-H lj
nbp_r_eps  3.00 0.0081378 12  6   # C-H lj
nbp_r_eps  3.00 0.0081378 12  6   # C-H lj
sol_par  12.77 0.6844    # C atomic fragmental volume, solvation
parameters
constant  0.000      # C grid map constant energy
map complex_md2.A.map                # atom-specific affinity map
nbp_r_eps  4.00 0.0222750 12  6   # A-C lj
nbp_r_eps  3.75 0.0230026 12  6   # A-N lj
nbp_r_eps  3.60 0.0257202 12  6   # A-O lj
nbp_r_eps  4.00 0.0257202 12  6   # A-S lj
nbp_r_eps  3.00 0.0081378 12  6   # A-H lj
nbp_r_eps  3.00 0.0081378 12  6   # A-H lj
nbp_r_eps  3.00 0.0081378 12  6   # A-H lj
sol_par  10.80 0.1027    # A atomic fragmental volume, solvation
parameters
constant  0.000      # A grid map constant energy
map complex_md2.N.map                # atom-specific affinity map
nbp_r_eps  3.75 0.0230026 12  6   # N-C lj
nbp_r_eps  3.50 0.0237600 12  6   # N-N lj
nbp_r_eps  3.35 0.0265667 12  6   # N-O lj
nbp_r_eps  3.75 0.0265667 12  6   # N-S lj
nbp_r_eps  1.90 0.3280000 12 10   # N-H hb
nbp_r_eps  1.90 0.3280000 12 10   # N-H hb
nbp_r_eps  1.90 0.3280000 12 10   # N-H hb
sol_par   0.00 0.0000    # N atomic fragmental volume, solvation
parameters
constant  0.000      # N grid map constant energy
map complex_md2.O.map                # atom-specific affinity map
nbp_r_eps  3.60 0.0257202 12  6   # O-C lj
nbp_r_eps  3.35 0.0265667 12  6   # O-N lj
nbp_r_eps  3.20 0.0297000 12  6   # O-O lj
nbp_r_eps  3.60 0.0297000 12  6   # O-S lj
nbp_r_eps  1.90 0.3280000 12 10   # O-H hb
```

```
nbp_r_eps  1.90 0.3280000 12 10   # O-H hb
nbp_r_eps  1.90 0.3280000 12 10   # O-H hb
sol_par   0.00 0.0000    # O atomic fragmental volume, solvation
parameters
constant  0.236      # O grid map constant energy
map complex_md2.S.map               # atom-specific affinity map
nbp_r_eps  4.00 0.0257202 12  6   # S-C lj
nbp_r_eps  3.75 0.0265667 12  6   # S-N lj
nbp_r_eps  3.60 0.0297000 12  6   # S-O lj
nbp_r_eps  4.00 0.0297000 12  6   # S-S lj
nbp_r_eps  2.50 0.0656000 12 10   # S-H hb
nbp_r_eps  2.50 0.0656000 12 10   # S-H hb
nbp_r_eps  2.50 0.0656000 12 10   # S-H hb
sol_par  0.000 0.000    #S atomic fragmental volume, solvation
parameters
constant  0.000          #S grid map constant energy
map complex_md2.H.map               # atom-specific affinity map
nbp_r_eps  3.00 0.0081378 12  6   # H-C lj
nbp_r_eps  1.90 0.3280000 12 10   # H-N hb
nbp_r_eps  1.90 0.3280000 12 10   # H-O hb
nbp_r_eps  2.50 0.0656000 12 10   # H-S hb
nbp_r_eps  2.00 0.0029700 12  6   # H-H lj
nbp_r_eps  2.00 0.0029700 12  6   # H-H lj
nbp_r_eps  2.00 0.0029700 12  6   # H-H lj
sol_par   0.00 0.0000    # H atomic fragmental volume, solvation
parameters
constant  0.118      # H grid map constant energy
elecmap complex_md2.e.map           # electrostatic potential map
dielectric -0.1146                  # <0, distance-dep.diel;>0,
constant
```

DOCK PARAMETER FILE

```
outlev 1                            # diagnostic output level
seed pid time                       # seeds for random generator
types CANOSH                        # atom type names
fld complex_md2.maps.fld            # grid_data_file
map complex_md2.C.map               # atom-specific affinity map
map complex_md2.A.map               # atom-specific affinity map
map complex_md2.N.map               # atom-specific affinity map
map complex_md2.O.map               # atom-specific affinity map
map complex_md2.S.map               # atom-specific affinity map
map complex_md2.H.map               # atom-specific affinity map
map complex_md2.e.map               # electrostatics map
move PGW.ZINC.D1.pdbq               # small molecule
about 61.5018 0.7042 13.6168        # small molecule center
tran0 random                        # initial coordinates/A or random
quat0 random                        # initial quaternion
ndihe 7                             # number of active torsions
dihe0 random                        # initial dihedrals (relative) or
random
tstep 2.0                           # translation step/A
qstep 50.0                          # quaternion step/deg
dstep 50.0                          # torsion step/deg
```

```
torsdof 44 0.3113                       # torsional degrees of freedom and
coeffiecent
intnbp_r_eps  4.00 0.0222750 12 6    # C-C lj
intnbp_r_eps  4.00 0.0222750 12 6    # C-A lj
intnbp_r_eps  3.75 0.0230026 12 6    # C-N lj
intnbp_r_eps  3.60 0.0257202 12 6    # C-O lj
intnbp_r_eps  4.00 0.0257202 12 6    # C-S lj
intnbp_r_eps  3.00 0.0081378 12 6    # C-H lj
intnbp_r_eps  4.00 0.0222750 12 6    # A-A lj
intnbp_r_eps  3.75 0.0230026 12 6    # A-N lj
intnbp_r_eps  3.60 0.0257202 12 6    # A-O lj
intnbp_r_eps  4.00 0.0257202 12 6    # A-S lj
intnbp_r_eps  3.00 0.0081378 12 6    # A-H lj
intnbp_r_eps  3.50 0.0237600 12 6    # N-N lj
intnbp_r_eps  3.35 0.0265667 12 6    # N-O lj
intnbp_r_eps  3.75 0.0265667 12 6    # N-S lj
intnbp_r_eps  2.75 0.0084051 12 6    # N-H lj
intnbp_r_eps  3.20 0.0297000 12 6    # O-O lj
intnbp_r_eps  3.60 0.0297000 12 6    # O-S lj
intnbp_r_eps  2.60 0.0093852 12 6    # O-H lj
intnbp_r_eps  4.00 0.0297000 12 6    # S-S lj
intnbp_r_eps  3.00 0.0093852 12 6    # S-H lj
intnbp_r_eps  2.00 0.0029700 12 6    # H-H lj
#
rmstol 2.0                              # cluster_tolerance/A
extnrg 1000.0                           # external grid energy
e0max 0.0 10000                         # max initial energy; max number
of retries
ga_pop_size 250                         # number of individuals in
population
ga_num_evals 5000000                    # maximum number of energy
evaluations
ga_num_generations 27000                # maximum number of generations
ga_elitism 1                            # number of top individuals to
survive to next generation
ga_mutation_rate 0.02                   # rate of gene mutation
ga_crossover_rate 0.8                   # rate of crossover
ga_window_size 10                       #
ga_cauchy_alpha 0.0                     # Alpha parameter of Cauchy
distribution
ga_cauchy_beta 1.0                      # Beta parameter Cauchy
distribution
set_ga                                  # set the above parameters for GA
or LGA
sw_max_its 300                          # iterations of Solis & Wets local
search
sw_max_succ 4                           # consecutive successes before
changing rho
sw_max_fail 4                           # consecutive failures before
changing rho
sw_rho 1.0                              # size of local search space to
sample
sw_lb_rho 0.01                          # lower bound on rho
ls_search_freq 0.06                     # probability of performing local
search on individual
```

```
set_sw1                         # set the above Solis & Wets
parameters
ga_run 25                       # do this many hybrid GA-LS runs
analysis                        # perform a ranked cluster
analysis
```

DOCK PARAMETER FILE

```
ligand_atom_file            ../1_struct/OUTPUT.ZC1P10.5000.DOG.mol2
limit_max_ligands                                           no
skip_molecule                                               no
read_mol_solvation                                          no
calculate_rmsd                                              no
orient_ligand                                               yes
automated_matching                                          yes
receptor_site_file              ../2_site/CM7CCAnHc_SHOULDER.sph
max_orientations                                            2500
critical_points                                             no
chemical_matching                                           no
use_ligand_spheres                                          no
flexible_ligand                                             yes
min_anchor_size                                             40
pruning_use_clustering                                      yes
pruning_max_orients                                         100
pruning_clustering_cutoff                                   100
use_internal_energy                                         yes
internal_energy_att_exp                                     6
internal_energy_rep_exp                                     12
internal_energy_dielectric                                  4.0
use_clash_overlap                                           no
bump_filter                                                 no
score_molecules                                             yes
contact_score_primary                                       no
contact_score_secondary                                     no
grid_score_primary                                          yes
grid_score_secondary                                        no
grid_score_rep_rad_scale                                    1
grid_score_vdw_scale                                        1
grid_score_es_scale                                         1
grid_score_grid_prefix          ../3_grid/CM7CCAnHc_SHOULDER_grid
dock3.5_score_secondary                                     no
continuous_score_secondary                                  no
gbsa_zou_score_secondary                                    no
gbsa_hawkins_score_secondary                                no
amber_score_secondary                                       no
minimize_ligand                                             yes
minimize_anchor                                             yes
minimize_flexible_growth                                    yes
use_advanced_simplex_parameters                             no
simplex_max_cycles                                          1
simplex_score_converge                                      0.1
simplex_cycle_converge                                      1.0
simplex_trans_step                                          1.0
simplex_rot_step                                            0.1
```

```
simplex_tors_step                                                10.0
simplex_anchor_max_iterations                                    500
simplex_grow_max_iterations                                      500
simplex_final_min                                                no
simplex_random_seed                                              0
atom_model                                                       all
vdw_defn_file                              ../vdw_AMBER_parm99.defn
flex_defn_file                                        ../flex.defn
flex_drive_file                                       ../flex_drive.tbl
ligand_outfile_prefix                          ZC1P10.5000.SHOULDER.DOG
write_orientations                                               no
num_scored_conformers                                            1
rank_ligands                                                     yes
```

DOCK Batch Script Submit File

```
#AMBER Example Batch Script for oscbw
#
#PBS -N ZC1P10mpi1
#PBS -j oe
#PBS -m ae
#PBS -M jnadas@chemistry.ohio-state.edu
#PBS -l walltime=75:00:00
#PBS -l nodes=4:ppn=4:pvfs
#
echo "Pre-HOME"
# The path below may need to be changed
WORK=$HOME/DKSBMPI/tutorials/mpi_demo/
OUTPUT=$HOME/DOCKTEST/ZC1P10/
DOCK_COMMAND_FOR_MPI="$HOME/DKSBMPI/bin/dock6.mpi -i mpi1.in -o
mpi1.out"
# The file names below may need to be changed
echo "Pre-COPY"
cd $WORK
cp -r flex* $PFSDIR
cp -r vdw_* $PFSDIR
cp -r 2_site/ $PFSDIR
cp -r 3_grid/ $PFSDIR
cp -r 4_dock/ $PFSDIR
cd $PFSDIR
mkdir 1_struct/
cd $WORK/1_struct/
cp -r OUTPUT.ZC1P10* $PFSDIR/1_struct
cd $PFSDIR
cd 4_dock/
#
echo "Pre-SCRIPT"
# Periodic, explicit solvent simulations using the paricle-mesh Ewald
(PME)
# method may perform better using the pmemd program; change sander to
pmemd.
#Some applications may need the -O option for the sander/pmemd command
below:
mpiexec $DOCK_COMMAND_FOR_MPI
ls -al
```

```
cp -rf mpi1.out *mol2 $OUTPUT
exit
```

ZINC DATABASE DIVISION SCRIPT

The mol2 file downloadable off the Zinc Database website contains all the compounds for that library. In the case of the lead-like compounds, this mol2 file contained 2,000,000 compounds. The program was written so that the split size could be user determined, in this case multiples of 1,000 were chosen as most ideal with 5,000 being a reasonable split size.

```
#!/bin/sh

#Usage $0 <startcount> <quantity> <incrementby>

NUMARGS=$#
if [ ${NUMARGS} -lt 7 ]
then
   echo "Usage $0 <startcount> <quantity> <incrementby> <splitsize>
<file> <output> <lastfew>"
   exit 1
fi

STARTCOUNT=${1}
QUANTITY=${2}
INCREMENTBY=${3}
SPLIT=${4}
FILE=${5}
OUTPUT=${6}
LAST=${7}

COUNTER=0
RUNNINGNUMBER=${STARTCOUNT}
QUERY="FALSE"
while [ ${COUNTER} -lt ${QUANTITY} ]
do
#   echo $QUERY
#   echo ${RUNNINGNUMBER}

#CREATION OF THE FIRST FILE

   if [ $QUERY = "FALSE" ]
    then
#       echo "FIRST IF"
       sed '1 s/@<TRIPOS>MOLECULE/BLEH/g' ${FILE} > MASTER
       csplit -f ${OUTPUT} MASTER '/@<TRIPOS>MOLECULE/'
       sed '1 s/BLEH/@<TRIPOS>MOLECULE/g' ${OUTPUT}00 >>
${OUTPUT}0.mol2
```

```
#       sed '1 s/MOLECULE/&${OUTPUT}${RUNNINGNUMBER}/g' ${OUTPUT}.mol2
>> ${OUTPUT}0.mol2
       QUERY="TRUE"
   fi
#   echo $QUERY

#CREATION OF ALL INDIVIDUAL FILES

   if [ $QUERY = "TRUE" ]
     then
#        echo "SECOND IF"
        sed '1 s/@<TRIPOS>MOLECULE/BLEH/g' ${OUTPUT}01 >> MASTER2
        csplit -f ${OUTPUT}A MASTER2 '/@<TRIPOS>MOLECULE/'
        sed '1 s/BLEH/@<TRIPOS>MOLECULE/g' ${OUTPUT}A00 >>
${OUTPUT}${RUNNINGNUMBER}.mol2
#        sed '1 s/MOLECULE/&${OUTPUT}${RUNNINGNUMBER}/g' ${OUTPUT}.mol2
>> ${OUTPUT}${RUNNINGNUMBER}.mol2
        mv ${OUTPUT}A01 ${OUTPUT}01
        rm -r MASTER2
        QUERY="TRUE"
   fi

#CONCATENATION OF INDIVIDUAL FILES INTO MULTIPLES OF ##

#  echo "MATH"
  let DIVISION=${COUNTER}
  INDEX=$(expr ${DIVISION} % ${SPLIT})
#  echo "DIVISION" ${DIVISION}
#  echo "INDEX" ${INDEX}

  if [ ${COUNTER} -lt 5 ]
    then
    BEGIN=0
  fi

  if [ ${INDEX} -eq 0 ]
    then
     echo "CONCATENATE" ${COUNTER}
     while [ ${BEGIN} -lt ${COUNTER} ]
     do
#     echo "CONCATENATION... " ${BEGIN}
     cat OUTPUT.${OUTPUT}.${COUNTER}.mol2 ${OUTPUT}${BEGIN}.mol2 >>
CC${COUNTER}.mol2
     mv CC${COUNTER}.mol2 OUTPUT.${OUTPUT}.${COUNTER}.mol2
     let BEGIN=${BEGIN}+1
     done
  fi

      let RUNNINGNUMBER=${RUNNINGNUMBER}+${INCREMENTBY}
      let COUNTER=${COUNTER}+1
      let DONE=${QUANTITY}-1

#LAST FEW TO CONCATENATE OTHERWISE THEY DISAPPEAR...

   let NEAREND=${QUANTITY}-${LAST}
```

```
   if [ ${COUNTER} = ${DONE} ]
     then
     echo "THE LAST FEW"
     while [ ${NEAREND} -lt ${QUANTITY} ]
     do
     echo "LAST DO LOOP" ${NEAREND}
     cat OUTPUT.${OUTPUT}.${QUANTITY}.mol2 ${OUTPUT}${NEAREND}.mol2 >>
CC${QUANTITY}.mol2
     mv CC${QUANTITY}.mol2 OUTPUT.${OUTPUT}.${QUANTITY}.mol2
     let NEAREND=${NEAREND}+1
     done
  fi

#ADDITION OF VERY LAST LIGAND TO LAST FEW CONCATENATED FILES and
REMOVAL of TEMP

  if [ ${COUNTER} = ${DONE} ]
     then
     cat OUTPUT.${OUTPUT}.${QUANTITY}.mol2 ${OUTPUT}01 >> ZLAST
     mv ZLAST OUTPUT.${OUTPUT}.${QUANTITY}.mol2
     rm -r ${OUTPUT}*
     rm -r MASTER
     exit
   fi

done

exit 0
```