

MOLECULAR SIMULATIONS OF POTENTIAL AGENTS AND TARGETS OF
ALZHEIMER'S DISEASE

A thesis submitted in fulfilment of the requirements for the degree

of

MASTER OF SCIENCE

in CHEMISTRY

of

RHODES UNIVERSITY, SOUTH AFRICA

Department of Chemistry

Faculty of Science

By

LUXOLO JOJI

December 2019



RHODES UNIVERSITY
Where leaders learn

To:

my supervisor Prof K.A Lobb

To:

The Joji family

“One of The Lessons That I Grew Up With Was To Always Stay True To Yourself And Never
Let What Somebody Else Says Distract You From Your Goals.” – Michelle Obama

ABSTRACT

Alzheimer's Disease (AD) is a neurodegenerative brain disorder that was first discovered in 1901 by Dr Aloïs Alzheimer and was later reported publicly in 1906. The German doctor had a 51-year-old woman patient called Auguste Deter, who was suffering from a rare brain disorder with early signs of memory loss and cognition. Alzheimer's Disease is the most common type of dementia that affects people with the age of 65 years and older. There is no single known cause of Alzheimer's disease however, amyloid β -peptide ($A\beta_{1-40/42}$) was found to be at the centre of AD pathogenesis and this connection was referred as "amyloid hypothesis". It is suspected that an accumulation of amyloid β -peptide is a major contributor to neuronal dysfunction and degeneration. Alzheimer's disease is complex and therefore, currently there is no medication available that treats the disease. However, there are approaches that focus on helping people maintain mental function, manage behavioral symptoms, and slow down the symptoms of disease. According to South Africa's 2011 census, there are approximately 2.2 million people in South Africa with some form of dementia and therefore there is a need to find a treatment for the disease. This study aims to find agents and targets of Alzheimer's Disease by using different computational techniques such as molecular modelling. The study will use compounds from the South African Compounds Database (SANCDB) and the following therapeutic targets α -, β - and γ -secretase, acetylcholinesterase, tau protein and neprilysin. A successful High-throughput Virtual Screening (HTVS) study to determine lead compounds was performed using a computational program called KNIME. Molecular docking was achieved with GLIDE as it allows for exhaustive ligand flexibility. The docking calculations were carried out using the high level of precision XP (extra precision) for enhanced docking accuracy. The binding affinities (docking scores) for the best bound ligands obtained from docking were in the order of -5 kcal/mol or less. The ligand SANC00370 was the best binding ligand against the protein 1J1C_B and had the best binding energy of -13.94 kcal/mol compared to others. The receptor-ligand complexes were analyzed using the interaction diagrams obtained from the Discovery Studio Visualizer and Maestro programs. Molecular Dynamics simulations were performed on the complexes obtained from docking to help in optimizing their interactions. The simulations were performed using the Desmond tool with the OPLS3 force field. 100 ns simulations were performed for six systems with the best docking score results representing each of the therapeutic targets and for the other complex systems, 50 ns simulations were performed. The Desmond simulations were analyzed using the Simulations Interaction Diagrams such as PL-RMSD, L-RMSF, P-RMSF, L-Torsions, P-SSE, LP-Contacts and L-Properties. Maestro was used to visualize the stability of the ligands in the active site during the simulation. All 13 Desmond simulations were successful however, there were 9 simulations which produced satisfactory results while the others were unsatisfactory. Based on the molecular docking and Molecular Dynamics results of this study, 9 potential targets and 6 potential agents were obtained successfully and can be studied further as therapeutics for Alzheimer's Disease.

ACKNOWLEDGEMENTS

To my supervisor Professor Kevin Alan Lobb, from the time I was still an undergraduate, I was astonished of how amazing a lecturer you are, no matter how difficult the concepts you taught us were at the time, you made them simple and for that I thank you. I would have never chosen another supervisor except you and from my experiences with you as my supervisor from Honours to Masters, I have no doubt that you're the best. I am grateful for all the computer skills you have taught me in all the different computational techniques I have used under your supervision.

I would also like to thank the National Research Fund (NRF) for providing funds to cover my MSc studies expenses for the past two years. I highly appreciate your help because it has provided me an opportunity to not only further my studies but also advance my skills as a researcher. Furthermore, I would like to appreciate the staff from the Rhodes Post-Graduate Funding Office particularly Mr John Gillam for providing me with the Rhodes University Council and Study Loan. I received the loan when I had no other financial assistance at the time and I was able to register for the degree because of it, thank you.

Gratitude's to Rhodes University, for providing the computational facilities particularly the Chemistry department for the Perkin server made available to me for this project. My research project was based on computational studies and therefore I am grateful for the computational time made available to our research group by the Center of High-Performance Computing in Cape Town (CHPC). I would like to mention Doctor Khrishna Govender, who has assisted me a lot during this study to be able to execute my Desmond Simulations on CHPC, I appreciate his help a lot.

Lastly, gratitude's to all my colleagues from my research group your inputs and advices have moulded my project to be what it is now and for that I thank you. Special mentions to Dr Arthur Sarron for his help with KNIME and Schrodinger Maestro, and to Dr Tendamudzimu Tshiwawa with helping me with Molecular Dynamics (Gromacs and Charmm).



National
Research
Foundation



DEDICATION

I dedicate this thesis to:

- my mentor and supervisor, Prof. K.A. Lobb and his family
- my parents, Vukile Gladstone Joji and Lindiwe Mavis Joji
- my siblings, Nokuphiwa Joji, Bukiwe Joji and Buntu Joji

They are my inspiration and they give me strength to continue towards achieving my goals.

TABLE OF CONTENTS

Contents

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	iv
DEDICATION.....	v
CHAPTER 1.....	1
Introduction.....	1
1.1 Dementia.....	1
1.2 Alzheimer’s disease.....	2
1.3 Alzheimer Disease hypotheses and associated targets.....	6
1.5 Ligands/Inhibitors.....	20
1.6 Aims and objectives of the study.....	23
1.7 Computational Methods.....	24
1.8 Previous studies from the Literature.....	30
CHAPTER 2.....	37
Computational Methods/Simulations.....	37
2.1 High-Throughput Virtual Screening.....	37
2.2 Molecular Dynamics.....	44
CHAPTER 3.....	47
Results and Discussion.....	47
3.1 Physicochemical Properties and Molecular Docking Results.....	47
3.2 Desmond: Molecular Dynamics Results (100 ns).....	79
3.3 Further protein-ligand complex dynamics.....	110
CHAPTER 4.....	112
Conclusion.....	112
CHAPTER 5.....	114
Bibliography.....	114
CHAPTER 6 Appendix.....	121
6.1 HTVS interaction diagrams.....	121
6.1.1 Docking Scores: Heatmap.....	121
6.2 Additional SANCDB HTVS Docking Detail.....	123
6.2.1 Protein-ligand complex interaction diagrams.....	123
6.2.2 H-Bonds.....	127

6.2.3 Hydrophobicity	128
6.2.4 Aromaticity	130
6.3 Molecular Dynamics Interaction Diagrams (50 ns)	134
6.3.1 Protein-Ligand RMSD	134
6.3.2 Protein RMSF	136
6.3.3 Protein SSE Histogram	138
6.3.4 Protein SSE timeline	139
6.3.5 Ligand RMSF.....	140
6.3.6 Ligand-Protein Contacts 2D summary.....	142
6.3.7 Protein-Ligand Contacts Histogram	144
6.3.8 Protein-Ligand Contacts Timeline	145
6.3.9 Ligand Torsions	147
6.3.10 Ligand Properties	150
6.4 PubChem target Molecular Dynamics	152
6.4.1 Protein-Ligand RMSD	152
6.4.2 Protein RMSF	153
6.4.3 Protein SSE Histogram	154
6.4.4 Protein SSE timeline	155
6.4.5 Ligand RMSF.....	156
6.4.6 Ligand-Protein Contacts 2D summary.....	157
6.4.7 Protein-Ligand Contacts Histogram	158
6.4.8 Protein-Ligand Contacts Timeline	159
6.4.9 Ligand Torsion.....	160
6.4.10 Ligand Properties	161
6.5 Molecular Dynamics 100 ns additional information	163
6.5.1 Ligand RMSF.....	163
6.6 Ligand Structures	165
6.6.1 Best Ligand Structures from SANCDB	165
6.6.2 Best Ligand Results from PubChem.....	167
6.6.3 Docking Scores Plots	169

CHAPTER 1

Introduction

In 1901, Dr Alois Alzheimer who was a psychiatrist and neuropathologist, observed a 51-year-old female patient at the Frankfurt Asylum in Germany. The patient showed bizarre behavioural symptoms and the loss of short-term memory, which was later named after him as Alzheimer's Disease (AD). He followed her case until she died in 1906, following which reported publicly on it. The autopsy of the patient's brain revealed plaques and tangles of neurofibrils, generated by insoluble fragments of the tau protein[1].

There are now several proposed causes of AD, and although there are current treatments of AD, none of these are successful in reversing or halting the progression of the disease[2]. Amyloid- β peptides ($A\beta_{1-40/42}$), the amyloid precursor protein (APP) cleavage products, have been identified as the main component of senile plaques and cerebrovascular inclusions, while tau protein has been identified as the major component of AD neurofibrillary lesions.[3]. Consequently, this has led to the development of numerous drug prospects based on the amyloid and tau hypotheses for Alzheimer's disease. [2].

1.1 Dementia

Dementia is a group of symptoms characterized by a loss in memory, vocabulary, problem solving and other cognitive abilities that affect the ability of an individual to perform daily activities. The deterioration is caused in parts of the brain involved in cognitive function by disruption or loss of neurons. This syndrome is known as a severe neurocognitive disorder as it interferes with both cognitive function and daily activities results. Cognitive function refers to the capacity of memory, voice, language, decision, reasoning, planning and other thinking abilities.

There are many cases where people suffer from mixed dementia which refers to anomalies in the brain associated with more than one cause of the syndrome. In some cases, people with dementia

symptoms do not have dementia, but have a disorder whose symptoms are close to those of dementia. Depression, delirium, side effects from treatment, thyroid disorders, vitamin deficiencies, and heavy use of liquor are common causes of dementia-like symptoms. These conditions can often be overturned with medication, apart from dementia[4]. Dementia is induced by an underlying disease of the brain and is not a typical result of aging. Some people experience a slight age-related memory loss that does not interfere with their everyday work and is not the same as dementia[5].

Dementia is an increasing concern for public health worldwide, with a total estimated cost of US\$604 billion per year in 2010. There is little information known about the prevalence of dementia or its impact on older people living in low- and middle-income countries (LMICs) in Africa, including South Africa. In addition, minimal work has been conducted on the aetiology and risk factors in LMICs. Studies to examine these variables are therefore very crucial in South Africa and the rest of the continent[6].

1.2 Alzheimer's disease

Alzheimer's disease is a neurodegenerative disease and the most common type of dementia. The damage and destruction of neurons in an AD patient eventually affects other parts of the brain, including those that enable a person to carry out basic bodily functions such as walking and swallowing.

As mentioned above, it was first identified more than 100 years ago, but 70 years passed before it was recognized as being the most common cause of dementia, as well as a major cause of death. There is information yet to be discovered about the precise biological changes that cause AD, why it progresses more quickly in some than in others, and how the disease can be prevented, slowed or stopped. Researchers believe that early detection of Alzheimer's will be key to preventing, slowing and stopping the disease[4].

There are two types of lesions present in the cerebral cortex: (1) amyloid plaques are found at the extracellular level and associated with A β -peptide; and (2) neurofibrillary tangles (NFTs), found

inside the neurons, which are associated with tau protein[7]. In addition to the formation of the senile plaques, studies have shown that $A\beta_{1-40/42}$ accumulation outside neurons could also lead to destructive changes to the brain, such as inflammation, oxidative stress, mitochondrial dysfunction, neurodegeneration, and clinical symptoms such as memory loss, mood changes, and difficulties in communication and reasoning[8].

1.2.1 Prevalence

The prevalence of Alzheimer's disease refers to the proportion of people in a population who have Alzheimer's at a given point in time[4]. There are very few surveys of AD or other dementias that have been conducted in developing countries but nonetheless by 2025, an estimated 68% of the world's elderly population will be living in these countries. China is the most studied developing country however, most of the work on dementia has been carried out in cities. Very few of the hundreds of studies of the prevalence of dementia have been carried out in Africa and much of the early work had been concerned with small hospitalized samples. No drug utilization or pharmaco-epidemiological studies are in the literature that specifically investigate the treatment of AD in South Africa since the introduction of memantine (a neuroreceptor antagonist, launched in March 2006)[9]. However, according to South Africa's 2011 census, there are approximately 2.2 million people in South Africa with some form of dementia[10].

1.2.2 Symptoms

The symptoms of AD differ among individuals but, the most common initial symptom is memory loss. This is due to the damage and destruction of the first neurons occurring in the brain regions involved in creating new memories. As the damage and destruction of neurons in different parts of the brain advances, individuals develop additional challenges.

The most common symptoms of Alzheimer's disease[11]:

- Memory loss.
- Problem solving difficulties.
- Challenges in executing duties at home or at the workplace.
- Confusion

- Trouble understanding visual images and spatial relationships.
- Speech problems
- Low or poor judgment.
- Retirement from work or social activity.
- Mood swings

1.2.3 Causes of Alzheimer's disease

There are two forms of AD: early and late. The early form occurs around 40 years or earlier and corresponds to only 2% of cases. It is produced through mutation of three different genes that encode presenilin's 1 and 2 (PS1 & PS2) and the amyloid protein precursor (APP). The late form occurs from around 60 years and above, it is the most common form of the disease[1]. Familial AD[12] (FAD) cases, caused by these autosomal dominant mutations in the genes for APP and the presenilin (PS1 and PS2), increase production of $A\beta_{42}$, relative to $A\beta_{40}$. The amyloid plaques are one of the neuropathological features of AD that result from the deposition of aggregated $A\beta_{1-40/42}$ peptides. Moreover, patients with APP gene duplications or individuals with Down's syndrome (trisomy 21), who have increased dosage of the APP gene (located on chromosome 21), develop early-onset AD and overproduce $A\beta_{42}$. These findings, along with a large body of evidence from other sources, strongly suggest that $A\beta_{1-40/42}$ plays a central, early role in AD pathogenesis. APP, β - and γ -secretases are the three principal players involved in $A\beta_{1-40/42}$ production[12].

There is no single known cause of Alzheimer's disease (AD). Research states that most cases, AD is most prevalent in people older than the age 65. A family history, apolipoprotein E-e4 (APOE-e4) and age are some of the risk factors for developing the disease. Age has been described as the greatest risk factor for the development of AD. While the mechanism by which APOE-e4 contributes to the disease is poorly understood, research has shown that Alzheimer's disease is not developed by everyone carrying the gene. As mentioned earlier, certain forms of AD are caused by mutations in very rare genes, and people who inherit these genes are almost certain of developing AD, often as early as their 30s and 40s[13],[14].

There are two suggested main causes of AD and they are in the forms of the amyloid and tau hypotheses.

The amyloid theory suggests that surplus amyloid β -peptide ($A\beta_{1-40/42}$) is a major contributor to the neuron dysfunction and degeneration that happens in Alzheimer's disease[15]. In addition to choline loss, multiple complex processes involving enzymes like β -secretase are involved in the first step of the amyloid precursor protein (APP) decomposition pathway leading to amyloid- β peptide synthesis and deposition[16].

In the tau hypothesis, it is proposed that the irregular accumulation of phosphorylated tau proteins along with the subsequent development of neurofibrillary tangles (NFTs) contribute to AD progression.[17],[18]. Tau phosphorylation controls the binding of tau to microtubules[19] and thus changes to it, like tau hyperphosphorylation, result in the loss of function by preventing its association with microtubules[20]. In this hypothesis, tau aggregation and the neurotoxicity associated with the hyperphosphorylated forms are suggested as key contributors to the phenomenon of neurodegeneration in AD[21].

It is believed that the progression of the multiple factors presented by both theories are the main reason why there is no effective cure currently available. Consequently, the molecules that can handle these numerous problems together could be potentially effective in treating AD[16].

1.2.4 Treatment of Alzheimer's disease

One of the big unmet needs of modern medicine is effective treatment of Alzheimer's disease[11]. Alzheimer's disease is complex and not well known, so there are currently no medications available to treat the disease effectively, halt or reverse its progression. There are, however, strategies that focus on helping people to maintain mental function, handle psychological symptoms, and delay disease progression. Such strategies include medications that can help to reduce certain symptoms and help control certain symptoms of behaviour. These medications are galantamine, rivastigmine, donepezil and memantine. The mechanism by which these drugs alleviate some symptoms of AD will be discussed later in the chapter.

The most common signs of AD are the amyloid plaques and neurofibrillary tangles. These cannot be used in the treatment of AD patients, since these structural defects can only be seen in autopsy and cannot be identified by existing brain imaging techniques[13]. The β -secretase (BACE1), a substrate for amyloid precursor protein (APP), is a promising molecular target for anti-Alzheimer drug development. BACE1 activates the production of A β -peptide, the main component of AD patients' amyloid plaques[2].

1.2.4.1 Pharmacologic treatment

None of today's pharmacological medications for AD prevents or avoids neuron damage and destruction which triggers the symptoms of Alzheimer's disease and makes it fatal. Through increasing the number of neurotransmitters in the brain, the six medications approved by the U.S. Food and Drug Administration (FDA) to treat AD temporarily boost symptoms. Several factors contribute to the challenge of creating successful Alzheimer's disease therapies. Such considerations include the uncertainty about the cause of AD, the high cost of drug discovery, the fairly long time needed to monitor whether an investigational treatment affects the progression of disease, and the brain structure covered by the blood-brain barrier through which only very advanced small-molecule drugs can cross.

1.2.4.2 Non-pharmacologic treatment

Non-pharmacological therapies are drug-free therapies and are often used to maintain or enhance cognitive function, the ability to perform daily living activities, or the overall quality of life. These can also be used to reduce behavioral symptoms such as depression apathy, wandering, sleep disturbances, anger and hostility [4].

1.3 Alzheimer Disease hypotheses and associated targets

As mentioned earlier, there are two distinctive structural irregularities that are found in Alzheimer's disease patients:

- *Senile plaques*—A β -peptide clumps. A $\beta_{1-40/42}$ is a result of cleavage of the APP; It is found in the membranes around the nerve cells as an aggregate; these A $\beta_{1-40/42}$ clumps collect outside the cells of the brain;
- Neurofibrillary tangles — twisted protein strands called "tau" which collect inside brain cells.

This has led to the two hypotheses about AD pathogenesis.

1.3.1 The Tau Hypothesis

Tau is one of the microtubule-associated proteins that regulate the stability of tubulin assemblies. It is thought that tau aggregates first develop in a limited number of brain cells, from which they spread to other regions, contributing to neurodegeneration and disease[22]. Research says that hyperphosphorylated tau, dissociates from microtubules and aggregates into paired helical filaments (PHFs) and NFTs. According to the hypothesis, tau phosphorylation and aggregation are the cause of neurodegeneration that occurs in AD[23].

1.3.1.1 Tau protein

A family of proteins called Microtubule Associated Proteins (MAP) stabilizes the formation of microtubules. A major member of the MAP family is the tau protein, which is involved in their association-dissociation by interacting with tubulin dimers that constitute the microtubules. Tau proteins are approximately separated into two parts: first, the extremity of the amino-terminus, which is the projection domain because it is not connected to the microtubules. In addition, the second part is the carboxy-terminus containing repeat domains that make up the Microtubules Binding Region (MTBR)[7]. Tau is a soluble protein in its normal state that encourages assembly and stabilization of microtubules[24].

In Alzheimer's disease, two types of kinases phosphorylate at least 30 serine / threonine residues: proline-directed kinases, such as GSK3; and non-proline-directed kinases, such as PKA[25]–[27]. Phosphorylation regulates tau binding to microtubules and GSK3 plays a major role in tau phosphorylation pathological control[19], [28]. Hyperphosphorylated tau protein accumulates as paired helical filaments that aggregate into masses within the neurofibrillary tangles associated with senile plaques[29].

The aggregated tau protein consists of neurofibrillary tangles (NFT). Tau is hyperphosphorylated during Alzheimer's disease, becomes removed from the microtubules, and accumulates in paired helical filaments and straight filaments in the somatodendritic compartment. The deposition of tangles occurs in a hierarchical fashion and only affects primary sensory areas in late stages of the disease. NFT deposition in human AD correlates with cognitive decline and neuronal loss[30].

Even though mutations in APP, but not MAP, cause inherited AD, neurofibrillary lesions are required for the development of dementia. People with abundant A β _{1-40/42} plaques, but no or only few neurofibrillary lesions, do not have AD[3]. A picture of a tau protein kinase is depicted on the Figure 1 below.

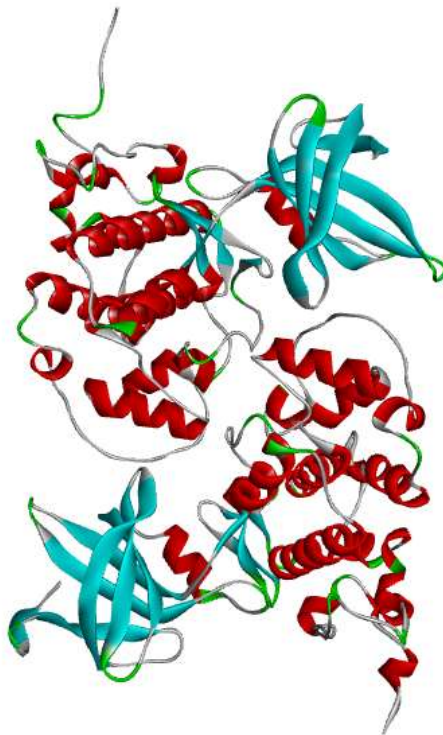


Figure 1: *Binary complex structure of human tau protein kinase I, PDB ID: 1J1C*[31]

1.3.2 The beta-amyloid hypothesis (amyloidogenic pathway)

The amyloid hypothesis is an assumption that a build-up of the amyloid- β peptide is the key cause of AD. The ‘amyloid hypothesis’ is that accumulation of the $A\beta_{1-40/42}$ peptide in the brain generates neurodegenerative processes that lead to the various symptoms detected in AD patients such as cognitive dysfunction[32]. The amyloid hypothesis also provides the reason for the pharmaceutical industry's current focus, namely the successful targeting and disruption of $A\beta_{1-40/42}$ aggregates[11].

1.3.2.1 Amyloid beta-peptide production

The formation of $A\beta_{1-40/42}$ in the amyloidogenic pathway proceeds as follows. APP is first cleaved by β -secretase (BACE1) at amino acid 1 (of $A\beta_{1-40/42}$). When the resulting C-terminal fragment, C99, undergoes γ -secretase cleavage, it releases the amyloidogenic $A\beta_{1-40/42}$ peptide. Key proteolytic components of γ -secretase are presenilin-1 and 2 (PS1, PS2). Amyloid beta-peptide has two key isoforms, soluble $A\beta_{40}$ and insoluble $A\beta_{42}$, the latter peptide depicts higher percentage concentration in AD patients and is more prone to aggregate[33].

1.3.2.2 Amyloid beta-peptide and Alzheimer's disease

Amyloid β -peptide ($A\beta_{1-40/42}$) is a 38 to 43 amino acid peptide that is derived from the β -amyloid precursor protein (APP) through sequential cleavages by β - and γ -secretase enzyme activities. The cleavage site for another APP processing enzyme, α -secretase, lies within the $A\beta_{1-40/42}$ sequence and thus precludes $A\beta_{1-40/42}$ formation. The amino terminal fragment generated through α - or β -secretase is called secreted APP (sAPP) α or β , respectively. The α -secretase activity is mediated by one or more enzymes from the family of disintegrin and metalloproteinase domain proteins (ADAM), with ADAM 9, 10, 17 and 19 being the most likely candidates.

The major beta-secretase in the brain is the β -site APP cleaving enzyme 1 (BACE1). On the other hand, a multi-protein complex constitutes γ -secretase activity. Four proteins are minimally

required for this complex: presenilin (PS) 1 or 2, nicastrin (Nct), presenilin enhancer 2 (Pen2) and anterior pharynx defective 1 (Aph-1). $A\beta_{1-40/42}$ is at the core of AD drug discovery studies and it is evident from this brief overview that at least pathways can be followed to minimize $A\beta_{1-40/42}$ production, i.e. to inhibit β - or π -secretase, or to increase α -secretase activity[15].

Recent findings indicate that amyloid deposition begins around 10-20 years before dementia starts, indicating that $A\beta$'s cerebral accumulation plays a crucial early role in AD pathogenesis. When given early enough during the disease, inhibition of $A\beta_{1-40/42}$ accumulation in the brain will help AD. Neurons are the major producers of $A\beta_{1-40/42}$ in the brain, although glia, astrocytes, can also contribute to $A\beta$ production, particularly during physiological stress which causes glial activation as in AD[34].

At the microscopic level, two characteristic brain lesions define AD: (1) amyloid plaques, extracellular deposits primarily composed of the 4-kDa, $A\beta_{40-42}$ [35], a product of amyloid precursor protein (APP) proteolysis; and (2) neurofibrillary tangles, intracellular aggregates of the microtubule associated protein tau[36]. The relationships between amyloid plaques, neurofibrillary tangles, and the pathophysiological mechanisms underlying AD have been controversial. However, evidence strongly suggests that $A\beta_{1-40/42}$ is critically involved at an early stage in the pathogenesis of AD[37]. Figure 2 below depicts the amyloid beta-peptide fragment $A\beta_{42}$.

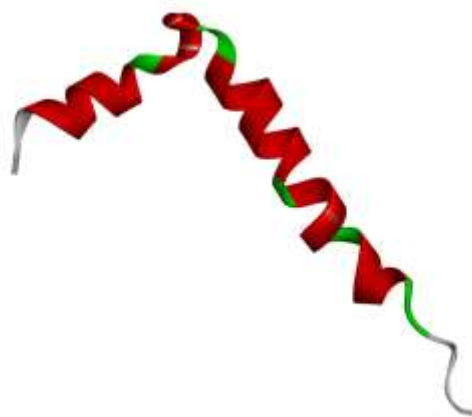


Figure 2: *Alzheimer's amyloid-beta peptide fragment 1-42 ($A\beta_{1-42}$) PDB ID: 1IYT[31].*

The prevailing theory, the ‘amyloid hypothesis’, suggests that accumulation of the amyloid- β peptide in protein aggregates triggers numerous pathophysiological changes that ultimately lead to cognitive dysfunction. The amyloid hypothesis also provides the rationale for the current focus of the pharmaceutical industry, namely, to successfully target the $A\beta_{1-40/42}$ aggregates.

This theory has contributed to the discovery of promising therapeutic strategies for AD therapy or prevention, such as the secretases/proteases involved in $A\beta_{1-40/42}$ production. Based on this it is now feasible to develop agents that interfere with the formation of $A\beta_{1-40/42}$. Aggregation also depends on concentration; reducing the concentration of $A\beta_{1-40/42}$ will therefore prevent aggregation and therefore such agents could manifest as a treatment for AD[11].

1.3.3 The Secretases

The fragment of the amino terminal produced by α -or β -secretase is called the α or β secret APP (sAPP). CTF83 and CTF99, respectively, are the carboxyterminal fragments (CTF) produced by α -and β -secretase. The CTF83 and CTF99 secretase cleavage will produce p3 and $A\beta_{1-40/42}$, respectively, as well as the intracellular amino-terminal APP (AICD) domain.[15]. There are three types of secretases that are crucially involved in the Alzheimer’s Disease and they will be briefly discussed below. They are α -, β - and γ -secretases; each has a different function towards $A\beta_{1-40/42}$ and the development of AD.

1.3.3.1 Beta-secretase

The β -secretase activity is from BACE1 (β -site APP cleaving enzyme 1) (Fig. 3), a type I transmembrane aspartyl protease, which has its active site in the extracellular space[3]. This results in APP releasing β -APP and generating CTF99. BACE1 can also break APP in a carboxy-terminal position to produce CTF89[15]. In the initial and rate-limiting step, the β -secretase characteristics give rise to the N-terminus of $A\beta_{1-40/42}$. It eliminates majority of the protein's extracellular portion, keeping only the C-terminal portion of APP.

BACE1 has all the molecular and cellular properties for β -secretase *in vitro* and *in vivo* predicted earlier. As required for β -secretase, the BACE active site is topologically oriented on the same side of the membrane as the β -secretase cleavage site in APP. In addition, BACE's activity has an optimal acid pH and the catalytic domain lies within acidic intracellular compartments, including endosomes and the trans-Golgi network. Moreover, BACE levels are greatest in neurons of the central nervous system, it has the right sequence specificity and if it is overexpressed in cells it cleaves APP and generates more amyloid β -peptide[38].

BACE1 is a key therapeutic target for reducing cerebral $A\beta_{1-40/42}$ levels. It effectively inhibits lower brain $A\beta_{1-40/42}$ levels, but two clinical trials with mild to moderate AD patients have failed due to unknown side effects and lack of efficacy. Since the β -amyloid deposition is a dynamic process in which initial plaque seed formation is followed by gradual growth, the potential therapeutic effects of BACE1 inhibition must be monitored, especially the $A\beta_{1-40/42}$ levels[39].

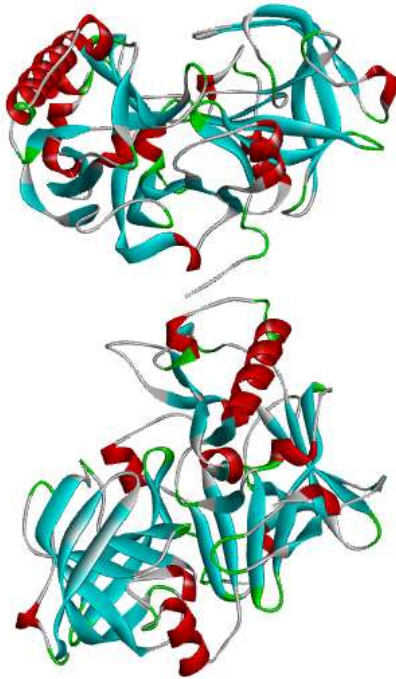


Figure 3: *The beta-secretase complex with PDB ID: 1FKN[31].*

1.3.3.2 Alpha-secretase

In the non-amyloidogenic pathway, APP is cleaved within the A $\beta_{1-40/42}$ domain by an α -secretase thus preventing deposition of intact A $\beta_{1-40/42}$. Research with a variety of different proteinase inhibitors found that α -secretase is a zinc metalloproteinase because hydroxamic acid-based compounds such as batimastat inhibit the expression of APP α without affecting full-length APP or β -secretase cleavage rates. There are suggestions that at least two different but closely related zinc metalloproteinases are involved in the constitutive and regulated α -cleavage of APP[12].

The activity of α -secretase is normally related to the protease family of the ADAM. It has been shown that ADAM 9, 10, 17 and 19 exercises α -secretase activity among this family. Which of these is the important protease in AD patients is presently unclear[15]. According to studies, ADAM10 is the major α -secretase in neurons. The bulk of APP is cleaved by the α - and γ -secretases, to form non-pathogenic fragments, nonetheless some is cleaved by β - and γ -secretases to produce amyloid β -peptide[3].

An improvement in α -secretase activity therefore is an attractive strategy for treatment of AD and can be accomplished by modulating selective signaling pathways[40]. However, the molecular identity of α -secretase is controversially discussed and remains to be fully established. ADAM9, 10 and 17 may all contribute to α -secretase activity and may still arbitrate APP α -secretase cleavage in the absence of one of them[41]. Figure 4 depicts the α -secretase ADAM10 extracellular domain protein used in this study.

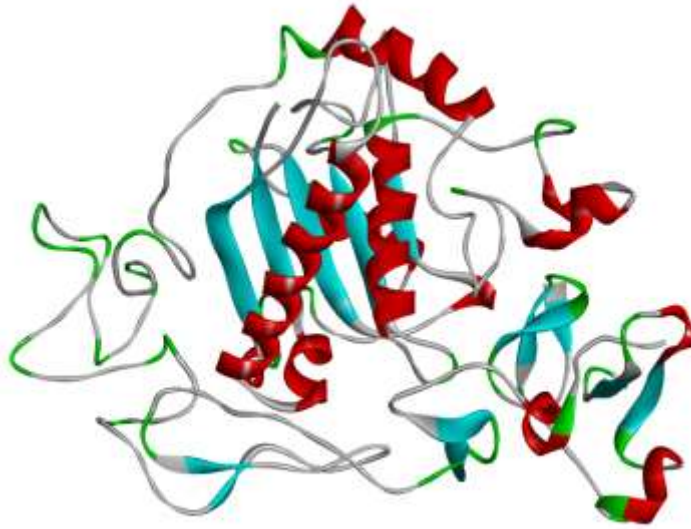


Figure 4: *The α -secretase ADAM10 extracellular domain from PDB ID: 6BDZ[31].*

1.3.3.3 Gamma-secretase

Gamma-secretase is a multi-subunit aspartyl protease that cleaves APP and several other type 1 transmembrane proteins within their transmembrane domains[11]. It cleaves APP in the intra-membrane area at the γ -cleavage site to generate $A\beta_{1-40/42}$. At least four proteins need to interact for γ -secretase activity to unfold: PSEN1 or PSEN2, Nct, Aph-1 and Pen2[15].

The presenilin 1 (PSEN1) and 2 (PSEN2) together with three accessory proteins, anterior pharynx defective 1 (APH1), presenilin enhancer-2 (PEN2) and nicastrin (NCT) function as the catalytic core of the intramembrane cleaving protease called γ -secretase (Fig. 5). This multi-subunit protease cleaves within the transmembrane domains (TMDs) of over 100 type 1 membrane proteins[42]. The three accessory proteins seem to be involved in the maturation and stability of the complex[11]. The γ -secretase was originally identified as the protease responsible for the generation of $A\beta_{1-40/42}$, and thus considered a prime therapeutic target in Alzheimer's disease[42].

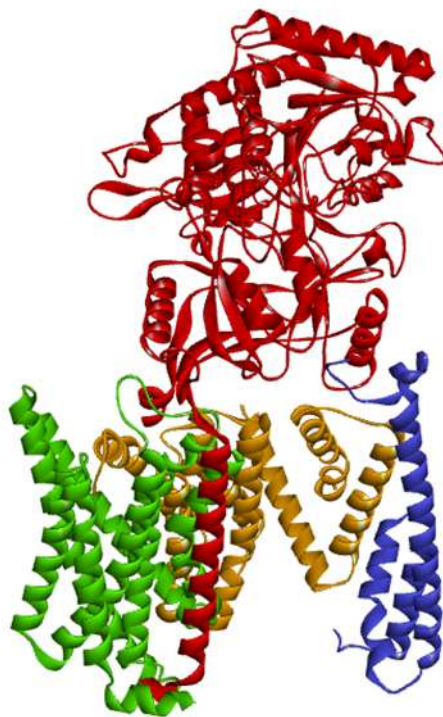


Figure 5: The γ -secretase complex, with nicastrin (red), presenilin-1 (orange), presenilin-2 (blue), and the gamma secretase subunits APH-1 (green). The structure was solved by cryo-electron microscopy. PDB ID: 5FN2[31].

1.3.4 Acetylcholinesterase

Nerve transmission can be improved by a reduction or an inhibition of the acetylcholinesterase enzyme's (AChE) activity. The AChE is accountable for lowering the levels of acetylcholine (ACh) by rapid hydrolysis in the synaptic cleft. A decline in AChE activity, retains upper acetylcholinesterase levels in the cleft, consequently intensifying the transmission of the electric impulses. In this way, anticholinesterase inhibition drugs will alter the progression of AD, enhance the patient's cognitive ability and improve the quality of life of the patient without, stopping the disorder from progressing.[1].

Reports suggested that AChE also bears responsibility for various non-catalytic acts like A β 's pro-aggregating activity. The AChE interacts with A β _{1-40/42} and facilitates the development of amyloid fibrils through an amino acid pool near the enzyme's peripheral anionic site (PAS). Molecules that

interact either exclusively with the PAS or with both catalytic and peripheral binding sites of AChE may thus prevent the pro-aggregating activity of AChE toward $A\beta_{1-40/42}$.

Moreover, studies have also found that many inhibitors of AChE promote cholinergic transmission and also impede with toxic $A\beta_{1-40/42}$ synthesis, deposition and aggregation. AChE inhibition has therefore been reported as a vital target for effective management of AD caused by an increase in the synaptic supply of ACh in brain regions and a reduction in $A\beta_{1-40/42}$ deposition. Consequently, molecules with dual AChE binding, *i.e.* binding to both catalytic and peripheral sites, are potential therapeutic agents for the effective management of Alzheimer's disease. [43].

Acetylcholinesterase (Fig. 6) has known different inhibitors such as tacrine, donepezil, rivastigmine and metrifonate; all of which will be discussed briefly. Tacrine, an amino-acridine, has several actions, but the most significant is as a reversible cholinesterase inhibitor that is centrally active.

In a first-pass metabolism, tacrine is quickly consumed and cleansed by the liver. Because of hepatic hydroxylation, when taken orally, tacrine has a very low bioavailability of 2-3% than that of an intravenous injection. Tacrine's short, dramatically varying half-life removal rises with higher doses and multiple doses from 1.4 to 3.6 h. It is easily consumed by the brain where the plasma concentration is tenfold.

Donepezil is a specifically designed piperidine derivative with reversible acetylcholinesterase inhibitor activity. It has a higher specificity for acetylcholinesterase inhibition relative to tacrine and its CNS selectivity is demonstrated by the absence of peripheral tissue involvement. In terms of pharmacokinetics, peak plasma concentration is reached in 4 hrs, but the elimination half-life is long in the range of 70-80 hrs.

Rivastigmine is a selective inhibitor of AChE carbamate in the brain. It is regarded as a 'pseudo-irreversible' inhibitor since it imitates ACh by attaching to a carbamylated complex formed by AChE. This prohibits additional enzyme-catalyzed hydrolysis of ACh for many hours after removal of the drug from the plasma. Therefore, rivastigmine has an action length of about 10 hours with a half-life of just 1 hour. Similarly, to donepezil, rivastigmine has marked the selectivity

of CNS, but it doesn't bind significantly to plasma proteins and is inactive during enzyme inhibition by cleavage. This avoids hepatic metabolism, and the drug is quickly excreted via the kidneys.

Metrifonate was considered in the late 1980s as a potential treatment for AD based on its anticholinesterase properties. It has a low plasma half-life (2h) but, thanks to its irreversible nature, it has a prolonged lifespan in the CNS. The active metabolite 2,2-dimethyldichlorovinylphosphate (DDVP) is hydrolyzed non-enzymatically. This compound binds stably to the enzyme's catalytic site and provides continuous inhibition of AChE in the CNS. The compound lacks in specificity for central inhibition but is quickly consumed and transmitted to the brain. In plasma, proteins remain largely unbound (< 15%) and the cytochrome P450 system is avoided[44].

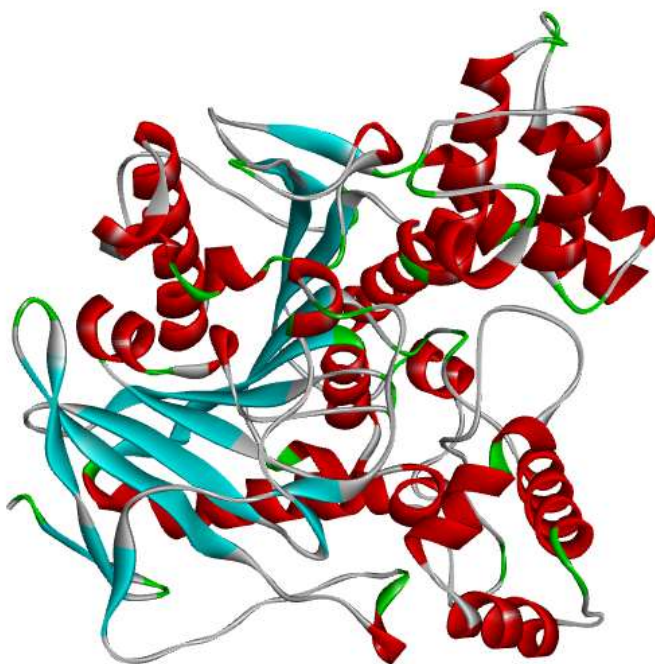


Figure 6: *The acetylcholinesterase protein with PDB ID:1GQR*[31].

1.3.5 Neprilysin

Increasing evidence has indicated that A β -degrading enzymes, including neprilysin (NEP), insulin-degrading enzyme (IDE), and endothelin converting enzyme (ECE), play a key role in A $\beta_{1-40/42}$ protein clearance[45]. Neprilysin, a plasma membrane glycoprotein of the neutral zinc metallo-

endopeptidase family, is a major A β -degrading enzyme capable of degrading not only monomeric but also oligomeric forms of A β_{1-40} . Furthermore, NEP is also capable of degradation of the oligomeric form of A β_{1-42} , a significant A β species in early pathogenesis. Nevertheless, NEP levels decline with age, resulting in increased A $\beta_{1-40/42}$ deposition in the brain. Reduced NEP levels and activity or polymorphisms in the NEP gene are associated with increased AD risk[8].

Neprilysin (Fig. 7) is considered one of the most important enzymes for the control of cerebral A $\beta_{1-40/42}$ levels in the brain. This enzyme works in the peripheral and CNS where small peptides have been shown to be degraded. The 50 amino acid catalytic core cleaves on the N-terminal side of hydrophobic residues. A study using radiolabeled A $\beta_{1-40/42}$, displayed that A β_{42} primarily experienced degradation by NEP in their *in vivo* assay[46]. In addition, the introduction of inhibitors to NEP in the rat brain resulted in significant endogenous amyloid β -peptide plaque deposition[47]. In addition, observations have revealed that NEP over expression imparts significant reductions in A $\beta_{1-40/42}$ plaque deposition in APP-transgenic mice and in some experiments, improved cognitive performance[47]. It has been demonstrated that NEP mRNA and protein expression levels are decreased in age-related or AD subjects, in concordance with its proposed effect of A $\beta_{1-40/42}$ clearance in the brain.

The spike in A $\beta_{1-40/42}$ levels after treatment with NEP inhibitors, leads also to the search for alternate A $\beta_{1-40/42}$ degrading enzymes. These NEP-like proteases are crucial due to their potential contribution to the removal of toxic plaques, and one such enzyme is neprilysin-2 (NEP2). NEP2 is also a zinc metallo-endopeptidase belonging to the same M13 family as NEP. Studies utilizing NEP2 as a therapeutic agent are still warranted because NEP2 was found to be even more selective for A $\beta_{1-40/42}$ [48].

Neprilysin levels is plentiful in the kidney, however, its content is very low in other organs, such as the brain. It is an ectoenzyme with the bulk of its structure, including the active site, facing the extracellular space. One feature of the NEP catalytic site is its restricted size which prevents access of large peptides and proteins but allows peptides containing up to 50 amino acid residues. This is consistent with A $\beta_{1-40/42}$ as a preferred substrate of NEP.

Importantly, NEP is the most effective A β -degrading enzyme in the brain and can deteriorate the monomeric forms of A $\beta_{1-40/42}$ and the more toxic oligomers. Research in vivo and in vitro now have a strong connection between NEP and AD pathogenesis and have rendered it a potential therapeutic target. There may be functional units that can be regulated by NEP modulators such as learning and memory, pain and inflammation, depression and anxiety, and further studies into the precise molecular mechanisms associated with tissue and cell-specific control of this peptidase could provide a valuable tool for enhancing human health and well-being[49].

The functional role of NEP in the brain is primarily determined by the physiological properties of its substrates and the roles they play in the nervous system. NEP is regarded as a major amyloid-degrading enzyme and although A β -peptide is poorly understood, evidence proposes that they can operate as modulators of neuronal function and synaptic plasticity. The role of NEP in controlling functional levels of A $\beta_{1-40/42}$ may be vital for healthy brain activity[49].

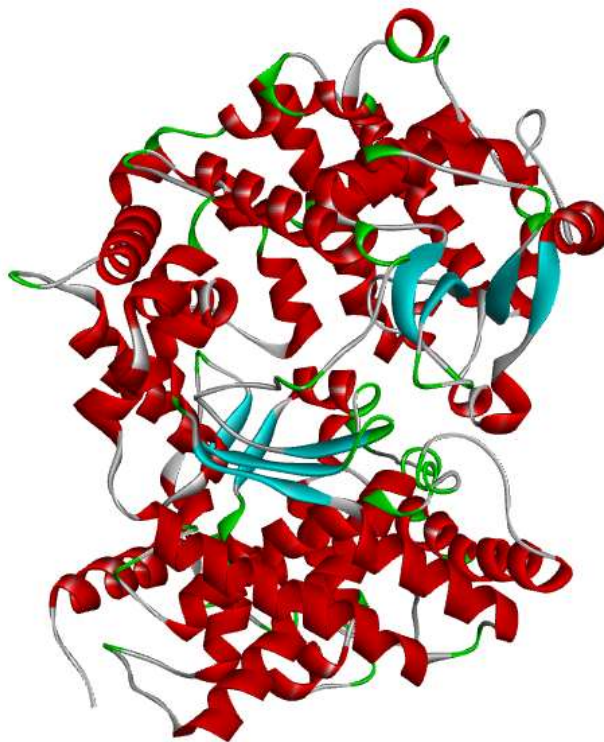


Figure 7: *The crystal structure of a human neprilysin (NEP) PDB ID:1R1H[31].*

1.5 Ligands/Inhibitors

1.5.1 *Virtual Chemical Databases*

The innovation of finding new lead compounds is one of the most important chemical issues in drug discovery[50]. It is important to find or design new compounds as there are many diseases yet to be cured, there is emergence of drug resistance in some diseases, and some of the current lead compounds are failing clinical trials or have no activity against the tested diseases. There are many virtual chemical databases that have been developed to aid researchers in finding new lead compounds and some of those will be briefly discussed below.

In the last 100 years, the synthetic chemistry has produced over 60 million compounds as collected by Chemical Abstracts Service. The combined corporate, academic, and commercial collections worldwide possibly total over 100 million different small molecules[51]. However, there are still difficulties in the development of new small molecule drugs, mainly due to lack of efficacy, side effects, and toxicity issues[52], [53].

On the other hand, researchers have been able to computationally produce virtual libraries of possible compounds for drug discovery. The GDB-17 database contains 166.4 billion organic compounds formed by an enumeration of all possible organic compounds up to 17 atoms of C, N, O, S, and halogens. There are millions of isomers of known drugs that can be found in GDB-17. This database contains molecules that are mostly nonaromatic heterocycles with several quaternary centres and stereoisomers[54].

Structure-based virtual screening has been successful over recent years and is a regularly used technique in early stage drug development at pharmaceutical companies and certain university groups. A virtual screening library could be generated from the 2-dimensional compound data supplied by several compound distributors. The ZINC database does this, and it contains purchasable molecules that are “drug-like” or “lead-like” and are available in many 3D formats that can be used by many popular docking programs[55].

DrugBank is a freely available web resource containing detailed drug, drug-target, drug action and drug interaction information about FDA-approved drugs along with experimental drugs going through the FDA approval process. The DrugBank is one of the world's most widely used reference drug resources due to its rich, high quality, primary-sourced content[56].

1.5.2 Natural product databases

Natural products (NPs) have been commonly used to cure virtually every known disease, with global reports stretching back as far as 4,600 years ago. NPs are the products of natural processes that provide them with structural complexity, chemical variety and biological specificities that are not observed in strictly synthetic compounds. NP scaffolds are crucial to the process of drug discovery and roughly 64% of approved drugs were obtained from or inspired by NPs between 1981–2010. Since 2008, there have been roughly 25 newly approved drugs extracted from NPs, with 31 extra drugs either in or past phase III clinical trials.

Several NP databases are generated to help with *in silico* drug discovery. In Africa, one such NP database is ConMedNP, which contains compounds isolated medicinal plants in Central Africa[57]. Similarly, to the ConMedNP, the AfroDb contains about one thousand NPs obtained from African flora but they were collected from various regions of the African continent. The AfroDb compounds display different biological activities and have been incorporated to the ZINC database[58]. This small compound library contains NPs that have recorded activities against different tropical diseases and diseases such as cancer and hypertension[59].

The Northern African Natural Products Database (NANPDB) contains compounds obtained largely from plants, with contributions from some endophytes, animals (e.g., corals), fungi, and bacteria. NANPDB represents the largest collection of annotated NPs derived from natural species found in Northern Africa. The database contains various known drugs and drug leads, however most of them have untested biological activity. Consequently, the medicinal potential is yet to be evaluated and this gives a chance for investigations in drug development[58].

NPs with biochemical specificity occupy a bigger chemical space than synthetic compounds, are favourable as lead compounds for drug development if they comply with drug-like or lead-like criteria. The natural product library (NPL) consists of pure structurally elucidated natural products and was intended to acquire compounds with improved drug-like/lead-like properties. This library is mainly based on compounds that have no violation to the Lipinski Rule of Five[60].

The Southern African region is rich with plant and marine life however, it doesn't have a NPs database. South Africa is rich in terrestrial biodiversity, and it is projected that 3,000 diverse plant species are used for medicinal purposes, largely as traditional medicine. Marine NPs are a new source of pharmacological agents but have diverse and distinctive chemistry. The studies on South African marine chemistry has provided compounds that depict potential as anticancer agents.

There is a lot of data available in literature about the isolation of molecules from organism's native to South Africa. However, these compounds are not easy to find, and it is difficult to distinguish whether which compounds from literature have been included in databases like ZINC and PubChem. In this study use is made of a database of compounds isolated specifically from South African plants and marine life, which is a useful resource for NP research.

The ligands used in this study were taken from the South African Nation Compounds Database (SANCDDB). The SANCDDB consist of compounds isolated from the plant and marine life in and around South Africa. Some of these compounds are known to be inhibitors of certain diseases while some are presently unknown.

SANCDDB is a MySQL database, which is integrated into a Django app. SANCDDB's primary objective was to connect all information to at least one referenced work, so this information was always recorded before uploading other specifics. Literature looked for journal articles, theses and textbooks that documented South African plant or marine life extraction of compounds. The set of theses from Rhodes University was used to find theses that met those criteria, while PubMed and ScienceDirect were used to find publications that matched those criteria.

Further literature was collected using Google and similar links to previously published papers and textbooks in the literature. This applied to compound data, source information and uses reported for each compound. The sources relate to the species from which a compound has been extracted and uses refer to any specific activity examined for a compound; e.g. antimalarial activity. Certain names are recorded as the name used in a particular publication to refer to a compound, and also names assigned by SciFinder to the compound if applicable[61].

This work will make use of different computational techniques such as molecular docking and molecular dynamics in achieving its aims and objectives. These techniques were designed for different purposes and they will be briefly discussed in Section 1.6.

1.6 Aims and objectives of the study

a) Aims

This study aims to focus on the following aspects of Alzheimer's Disease:

1. High-throughput screening to identify potential leads against secretases, a potential AD agonist.
2. Lead modification and scaffold hopping to improve lead compounds in the prevention of plaques.
3. Molecular dynamics in the evaluation of lead compounds.

b) Objectives

The goal of the study is to use molecular docking to determine the best binding ligands using binding energies and explore the receptor-ligand interactions of the various complexes. The most compatible ligands with the receptor will be modified using the PubChem[81] computational program. These ligands will be docked and compared to the SANCDB ligands to determine if there is any improvement in the binding scores and the protein-ligand interactions.

The results from the docking will then be evaluated using molecular dynamics to determine the behaviour of the receptor-ligand complex in a solvated and ionized system under the following conditions temperature (300 K) and pressure (1 bar).

1.7 Computational Methods

1.7.1 Molecular Modelling/Docking

About a century ago, Emil Fischer proposed a “lock and key” hypothesis that has become the basis for structure-based drug design. The "locks" are reported as the protein targets for active and selective drugs (the "keys") to fit in their designated active site. The achievement of this hypothesis in drug discovery is supported through decades of studies exploring “keys” by the structure/activity analysis of the experimental test compounds.

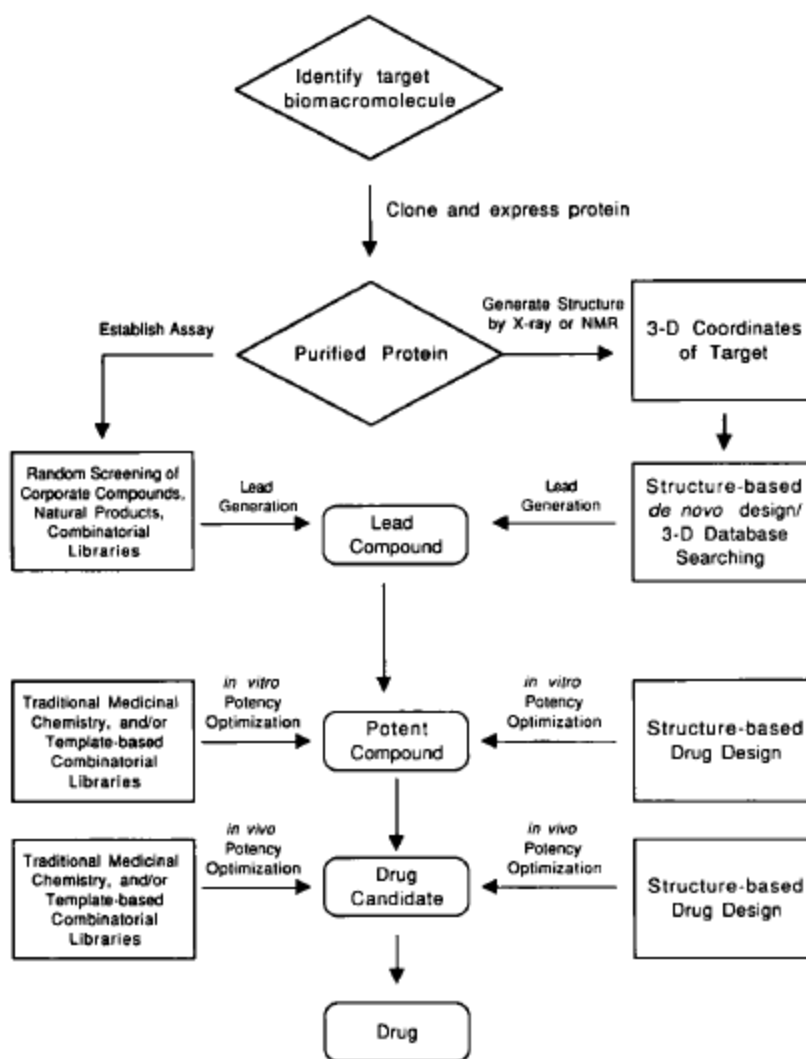


Figure 8: Drug discovery and design flow chart showing the different techniques used in drug development[62].

The ligand-based drug design approach plays a vital role in drug development. Figure 8 illustrates a drug design and discovery flow chart with techniques such as rational drug design (including structure-based methods), traditional screening and synthesis, and the recent combinatorial approaches. This scheme illustrates how structure-based drug design informs the process from lead identification to final optimization, in the discovery of novel agents. However, the hypothesis has some limitations such as the producing limited data regarding the “lock” and also often the ligand and receptor are treated as rigid counterparts which may preclude the discovery of certain binding modes[62]. In most cases, static receptor/ligand models do not match reality, where both

compounds and proteins have to encounter reciprocal changes in order to accommodate each other in a complex state. The ligand-protein binding mechanism has now been defined as a 'hand and glove,' suggesting that the best fit remains an important variable in a flexible setting. Noteworthy improvements have been made in the techniques used by docking software such as GLIDE to allow exhaustive ligand flexibility during docking[63].

Glide (energetic grid-based ligand docking) is a docking technique utilized in the Schrodinger software package used in this study. It has been designed to perform as close to an exhaustive search of the positional, orientational, and conformational space available to the ligand as is feasible while retaining enough computational speed to screen large libraries. Glide's present performance features are as follows:

- (i) The average docking times is less than 1 min for data sets with rotatable bonds between 0-10 on a Linux-operated AMD Athelon MP 1800 + processor.
- (ii) Predicting robustness in binding mode is qualitatively superior to what is stated in the existing literature for widespread use of docking methods.
- (iii) Binding affinity predictions are reasonable compared to experimental data for co-crystallized complexes (2.3 kcal/mol RMSD), however there is room for improvement.
- (iv) Library screening findings are very promising and the database enrichment factors obtained in Glide 2.5 have been improved[64].

1.7.2 Molecular Mechanics

Approximate Quantum Mechanics methods are used in isolated cases for studying the properties of isolated drug-like molecules and performing receptor-ligand complex simulations in the presence of water. However, the extremely reduced computational costs of Molecular Mechanics (MM) force fields, make Molecular Mechanics the method of choice for simulations of large systems [65]. Molecular Mechanics is a powerful method for studying molecular structure and related properties and calculates potential energy of systems using force fields[66]. It is based on the following assumptions:

- i) The nucleus is a perfect sphere.
- ii) Molecular bonds are treated as springs.

- iii) The potential function relies on experimental parameters such as force constants and equilibrium bond-length, bond-angle or dihedral angle values.
- iv) The potential energy function is the sum of individual functions for bond stretching, angle bending and torsional energies and non-bonded interactions

The force field is a model for the potential surface on which the atoms of the molecule are located[67]. Some widely used force field parameters include AMBER[68], CHARMM[69], GROMOS[70] and OPLS[71]. This study will focus on the OPLS3 force field which will be used in the Desmond molecular dynamics simulations. The functional form of the OPLS force field is given in equation 1, where E is the total energy.

$$\begin{aligned}
 E = \sum_{i < j} & \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) \right] f_{ij} \\
 & + \sum_{bonds} K_r (r - r_{eq})^2 \\
 & + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \\
 & + \sum_{dihedrals} \left[\frac{V_1}{2} (1 + \cos \varphi) + \frac{V_2}{2} (1 - \cos 2\varphi) + \frac{V_3}{2} (1 + \cos 3\varphi) \right. \\
 & \left. + \frac{V_4}{2} (1 - \cos 4\varphi) \right]
 \end{aligned} \tag{1}[71]$$

Bond stretch and angle bending energies are represented by harmonic terms where the subscript “eq” denotes the equilibrium bond length r and angle θ . The torsion term is represented by a truncated Fourier series that sums over all dihedral angles, where φ_i is the dihedral angle, V_1 , V_2 , and V_3 are the coefficients in the Fourier series. Non-bonded energies are represented by sums over Coulomb and Lennard-Jones terms covering pairwise interactions between available interacting sites (denoted i and j). For intramolecular atom pairs separated by 3 bonds or less the energy is scaled by f_{ij} . The Lennard-Jones σ and ϵ parameters for each ij pair are based on atom site-specific parameters coupled according to the geometric combining rule[72].

To obtain as close to full coverage of medicinal chemistry space for small molecules as is feasible, a very large number of parameters for equation 1 above are included in the OPLS3 force field. The general fitting protocol that has been used to populate these parameters is outlined as follows:

- (1) The stretching and bend terms have been fit to quantum chemical data. These terms are insensitive to the other terms in the force field and so were determined first.
- (2) The van der Waals terms were obtained from liquid state simulations, along with the core charge set used in the liquid simulations.
- (3) Bond charge corrections (BCC's) for the CM1A-BCC charge model were defined and tested.
- (4) Finally, in the last step the torsional parameters were fit to quantum chemical data.

The protocol defined above may also be applied to model compounds for which there is no experimental data, if the van der Waals parameters are well covered by the current force field. The OPLS3 force field delivers good performance in protein simulations and can predict protein–ligand binding affinities to high accuracy over a wide range of targets and ligands (less than 1 kcal/mol RMS error has been reported)[71].

1.7.3 Molecular Dynamics

Classical molecular dynamics (MD) simulations nowadays allow implementation of structure-based drug design (SBDD) strategies that fully account for structural flexibility of the overall drug–target model system. Classical MD is a technique used for studying the interaction and motion of atoms and molecules according to Newton's laws of motion with the use of a force field[73]. MD is a commonly used computational technique to study the equilibration structures and dynamic interactions of biological systems. It gives greater insights into time-dependent variations and configuration changes in biological system structures that may be connected to their functionality. The major benefit of the MD method is its ability to imitate the experimental conditions under which a typical biological issue is dealt with. In synthesis, experiments are performed by controlling different factors, like temperature, pressure, number of atoms, ionic concentration, and the type of solvent used to solvate the interacting molecules. All these factors can be easily modified and controlled in MD simulations within the context of statistical mechanics ensembles. These ensembles include the microcanonical ensemble (constant total energy),

canonical ensemble (constant volume and temperature), and isothermal-isobaric ensemble (constant temperature and pressure)[63].

1.7.2.2 Molecular Simulations

MD simulations can be performed on the receptor-ligand complexes obtained from docking in order to assist in the optimization of their interactions[63]. Nevertheless, the reliability of the effects of MD simulation depends heavily on the accuracy of the force field used[74]. MD simulations generate successive system configurations, providing trajectories that define particle positions and velocities over time. The MD trajectories produced during the simulation can be used to calculate free energy, kinetics measures, and other macroscopic quantities to compare with experimental results[73].

Biomolecular dynamics simulations can be applied in three different ways. Firstly, MD simulation is used to bring biomolecular structures to biological temperatures in simulation, giving a deep understanding into the natural dynamics on various timescales of biomolecules in solution. Secondly, MD simulation affords thermal averages of molecular properties. The ergodic hypothesis states that, one can simulate for a period a single molecule with its environment and obtain time-averaged molecular properties that approach the experimentally measurable averages of the ensemble. This is used to calculate the bulk properties of fluids and the free energy differences for chemical processes such as ligand binding. Thirdly, MD will investigate the thermally accessible conformations of a compound or complex. This method is used, for example, in ligand-docking applications to discover conformational space.

There are still issues in development. For instance, in the recent past parameters describing alkanes were insufficient to replicate the properties of lipid bilayer membranes during lengthy simulations of big systems that are currently possible. This resulted in better parameters for alkanes being generated within the OPLS3, CHARMM and GROMOS96 force fields, respectively[75]. All these force fields share specific basic characteristics: they rely on an electrostatic fixed charge model, they use a standard van der Waals potential to model electronic repulsion and dispersive non-

bonded interactions, and they use harmonic stretching and bending conditions and torsional dihedral angle-based potentials to model the valence component of the energy[71].

1.8 Previous studies from the Literature

The complexity of Alzheimer's disease designates that there are various factors that might be involved in its pathogenesis[76]. Some of these include genetic factors, family history of the individual affected with the disease, cerebrovascular disease, traumatic brain injury, depression, hormonal disturbance, inflammation, hyperlipidemia and hyperglycemia. The disease is also distinguished by huge loss of cells, particularly in the basal nuclei of cholinergic neurons, resulting in permanent dementia [77].

There have been no new drug treatments approved by the FDA (Food and Drug Administration) for AD since 2003 and therefore several studies have been conducted in search for new treatments. There are five drugs approved for the treatment of AD with four being cholinesterase inhibitors and an *N*-methyl-D-aspartate (NMDA) receptor AD antagonist as mentioned previously in section 1.3.4. The shortcomings in AD drug discovery with both small compounds and immunotherapies have been largely due failures in showing drug difference or having high toxicity[78]. Below we will review some of the failures of the drugs that were discovered for the therapeutic targets that are used in this study to discover the causes and the approaches that other scientists have used in tackling Alzheimer's disease.

1.8.1 Declining the A β -peptide production

Reducing the production of A $\beta_{1-40/42}$ either by inhibiting β and α -secretase or by activating α -secretase is a desirable approach to the creation of AD therapeutics[79].

1.8.1.1 Inhibition of β -secretase

The most promising inhibitor of BACE1 is MK-8931 (Fig. 9), which in 2012 successfully passed a safety study involving 88 healthy volunteers. There were several other studies conducted in the

evaluation of the MK-8931 compound with different groups of volunteers and they confirmed its safety and efficacy. The compound is currently conducting phase II/III trials involving thousands of patients with mild-to-moderate Alzheimer's disease and prodromal AD.

The AZD 3293 compound has passed phase I clinical trials with the concentration of $A\beta_{1-40/42}$ in CSF being reduced by 75% and currently is undergoing phase II/III clinical trials. After this, the compound will undergo a different study of 2 years with a group of patients with mild AD type of dementia using the name AMARANTH.

Another β -secretase inhibitor, LY2886721 (Fig. 9), successfully passed a phase I trial involving healthy volunteers. A phase II study with patients showing mild AD was withdrawn in 2013 after the compound caused side effects in the liver while these test results were prior not observed in animal models. It was found that the symptoms experienced by some patients was related to the administration of the drug.

The high attrition rate among this group of compounds reduces the validity of the amyloid hypothesis, thereby challenging the desirability of finding effective therapy in this category of drugs. Given excellent results from immunization trials to minimize amyloid deposits, neurodegeneration progression was not anticipated[77].

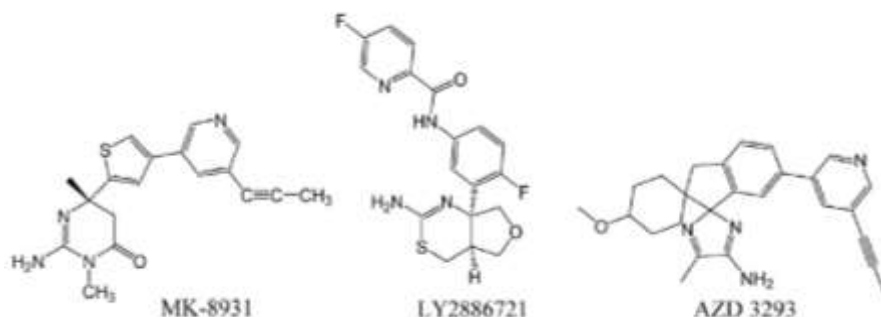


Figure 9: Chemical structures of β -secretase inhibitors[77].

1.8.1.2 Inhibition of γ secretase

The two γ -secretase inhibitors ELND006 and semagacestat have entered the clinical trials but unfortunately have both failed with the latter failing in the phase III trials. The compounds both caused significant side effects such as gastrointestinal ailments and high risk of skin cancer. The lack of predicted therapeutic effects; accelerated cognitive decline relative to the placebo control group for semagacestat were the reasons why these medications were omitted from clinical trials.

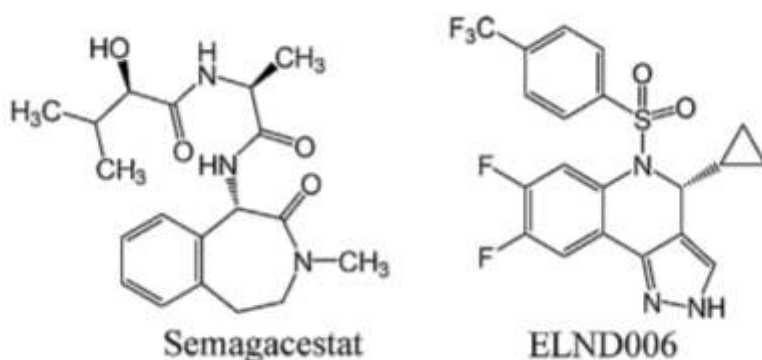


Figure 10: Chemical structures of the first generation of γ -secretase inhibitors[77].

The explanation for the failure of non-selective inhibitors was probably associated with presenilin 1's omnidirectional function, a catalytic subunit of γ -secretase. The Notch protein, which regulates cell proliferation, differentiation and development, is one of the most notable substrates of γ -secretase. A similar effect could be obtained by compounds detected as γ -secretase modulators that modify the enzyme's proteolytic activity, allowing it to interact with shorter forms of A β _{1-40/42} without affecting Notch proteolysis.

Nonetheless, the findings of modulator/secretase selective inhibitor clinical trials aren't very encouraging. The first γ -secretase modulator to undergo clinical trials was avagacestat (BMS-708163) (Fig. 11) – it has been tested in two large, long-term phase II trials (NCT00810147, NCT00890890). However, due to side effects, like stomach problems and skin reactions, one of these trials has been terminated. Furthermore, there was no change in cognitive function detected

in the patients. Only high doses of the drug led to high decreases in the level of $A\beta_{1-38}$, $A\beta_{1-40}$ and $A\beta_{1-42}$, but at the risk of serious adverse reactions and cognitive decline.

Another γ -secretase modulator, begacestat (GSI-953), was halted in phase I as the compound only influenced the plasma concentration of $A\beta_{1-40/42}$ (not its CSF concentration). Compound CHF-5074, with a formulation based on the structure of R-flurbiprofen, is also being tested as a γ -secretase modulator. So far the compound has exhibited an acceptable safety profile in phase I and is currently undergoing phase II clinical trials[77].

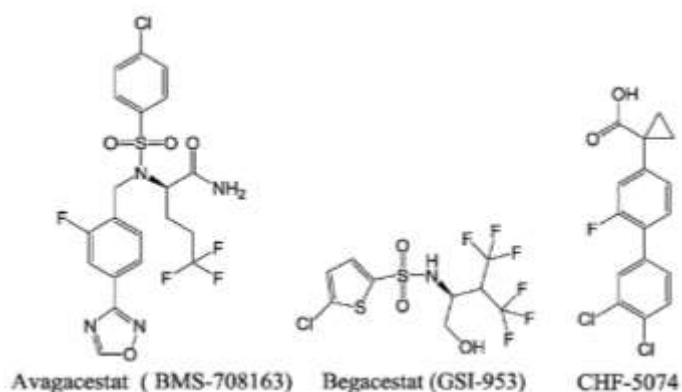


Figure 11: Chemical structures of the second generation of γ -secretase inhibitors[77].

1.8.1.3 Activation of α -secretase

An α -secretase cleavage activation may have two distinguishable benefits for AD. First, it is the lowering of the neurotoxic and pathogenic $A\beta_{1-40/42}$ peptide levels. Second, the volume of neuroprotective $APP\alpha$ is increased. Activation of APP α -secretase cleavage is therefore known to be a therapeutic approach to AD.

This activation has been further validated by the development of pharmacological α -secretase activators. For pre-clinical studies and patients with AD, a new generation of M1 muscarinic acetylcholine receptor agonists were tested. For α -secretase cleavage, the M1 receptor is needed as stated in a recent study using M1 receptor-deficient mice. The new M1 agonists are much more accurate than previous muscarinic agonists and are better tolerated. In the hippocampus and cortex

of an AD mouse model, one compound (AF267B, also known as NGX267) decreased $A\beta_{1-40/42}$ and tau protein pathology. A similar drug, AF102B, decreased $A\beta_{1-40/42}$ levels in AD patients' CSF and again pointed to α -secretase activation as a possible therapeutic strategy for lower $A\beta_{1-40/42}$ levels.

The etazolate (EHT 0202) compound which modulates the GABAA receptor is presently studied in phase II clinical trials. The compound unfortunately did not affect $A\beta_{1-40/42}$ levels and therefore cannot be classified as an $A\beta$ -lowering drug. Its positive effect can be mainly due to increased neuroprotective APPs α production. However, it may also contribute to increased secretion of other substrates of ADAM10/ α -secretase, which remains to be tested. Furthermore, the exact mechanism of activation of α -secretase by etazolate must be fully clarified.

Even though there has not yet been intensively tested a therapeutic activation of α -secretase, the above studies show the potential of this method. It is certain that further signaling pathways will be thoroughly tested, including the SIRT1/RAR pathways for transcriptional ADAM10 activation.[80].

1.8.1.4 Tau protein

The tau protein is also an essential biological target for innovative therapies. Potential tau-centered therapies include tau phosphorylation inhibition, microtubular stabilization, tau oligomerization prevention, and tau degradation enhancement, and also tau immunotherapy. Most anti-tau protein agents are still in preclinical testing, with only a handful of clinical trials.

LMTX is a promising compound which is currently in phase III of clinical trials. This compound is a derivative of a well-known Methylene Blue dye, i.e. methylthioninium chloride (MTC), whose molecule (leuco-methylthioninium) (Fig. 12) was improved for enhanced bioavailability and tolerability. LMTX acts as an inhibitor of tau hyperphosphorylation. It can stop tau interactions and facilitates clearance of tau from the brain; in addition, it also has anti- $A\beta$ aggregation activity.

Several phase I clinical trials with a microtubular stabilizing effect include epothilone D (BMS-241027) and TPI-287. Epothilone D (Fig. 12) is a small molecule, that can cross the blood-brain barrier and standardize tau binding at low doses. TPI-287 is a synthetic derivative of taxol with tubulin-binding and microtubule-stabilizing properties which can cross the BBB[77].

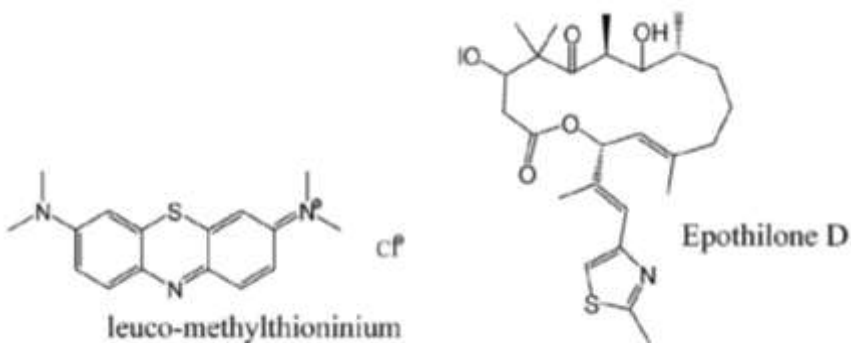


Figure 12: Chemical structure of compounds which focus on the tau protein[77].

1.8.1.5 Neprilysin

There are currently only few characterized endogenous tissue specific inhibitors of NEP. Bovine spinal cord isolated the first inhibitor, a heptapeptide spinorphin that also inhibited dipeptidyl peptidases and angiotensin-converting enzyme. Opiorphin was the first human saliva-isolated NEP inhibitor and had some pain-suppressive potency. Phosphoramidon and thiorphan are the most active and commonly used NEP inhibitors, and the 3D structure of the NEP extracellular domain in a phosphoramidon complex has been determined to allow a greater understanding of the enzyme's catalytic properties. The NEP catalytic site has a small size which prohibits large peptides and proteins from entry but enables peptides that contain up to 50 residues of amino acids. This is compatible with A β -peptide as a favoured neprilysin substrate. Another feature of NEP is its sensitivity to phosphoramidon and thiorphan inhibition at nanomolar levels. Even though phosphoramidon also inhibits a closely related NEP homologue endothelin-converting enzyme (ECE-1), it is only sensitive to the inhibitor's micromolar concentrations and is not affected by thiorphan[49].

1.8.1.6 Acetylcholinesterase

There are four acetylcholinesterase inhibitors currently available and these make up the 4 of 5 Alzheimer's Disease drug treatments currently available. The search for new acetylcholinesterase inhibitors is ongoing in hope to find those which have less side effects, improved half-life and possibly cure the disease amongst other things.

Chapter 2 is the experimental section and will explain the steps taken to achieve the aims and objectives above. It will also briefly explain how the steps will be taken and the programs used for the whole analysis.

CHAPTER 2

Computational Methods/Simulations

2.1 High-Throughput Virtual Screening

2.1.1 Docking

There are 716 compounds that were used for the high-throughput virtual screening, and they were taken from the South African Natural Compounds Database (SANCDB)[61]. The SANCDB contains compounds isolated from the plant and marine life in and around South Africa. Fourteen AD proteins were used in this study, and each protein was docked with the 716 compounds. Various computer programs were executed in carrying out the HTVS process which will be detailed in the following discussion.

The proteins were carefully examined prior to HTVS using Discovery Studio Visualizer[31]. At this time crystal structure water from the X-ray crystal structures was removed, ligands present were separated, and dimeric chains, where activity was not dependent on the dimeric interface, were separated into discrete structures for the screening process. This preparation was found to be essential, informed by previous preliminary work in which crystal structure ligands were found to interfere during some docking protocols in HTVS, thus giving poor binding energies compared to proteins that didn't have ligands.

HTVS was performed within a program environment called KNIME[82] and included in KNIME are different kinds of nodes that perform different functions; these nodes come from a variety of software vendors together with community, native KNIME and open source nodes. These nodes are then connected to provide a workflow that manages the flow of data and information for a project. An example of one of the workflows that were used for this project is shown in Figure 13; this is a simple workflow for the screening of the full set of SANCDB compounds against a single target, combined with the calculation of physicochemical properties for all compounds in the set. Schrödinger nodes are included with KNIME, and with these KNIME program connects seamlessly with Schrödinger Maestro[83] which handles the calculations required for the high-throughput virtual screening. In this workflow, a single protein (previously prepared in Discovery Studio) is read into the workflow via the PDB Reader node in KNIME. Similarly to how one works in the visual modeling software Maestro, the protein is then prepared using the Protein

Preparation Wizard[84] which allows you to take a protein from its raw state, (which may be missing hydrogen atoms and have incorrect bond order assignments, charge states, or orientations of various groups) to a state in which it is properly prepared for docking calculations using the Glide[64] docking software. After protein preparation, a Glide grid is then generated from the protein receptor using the Glide Grid Generation node. Within the setting of the KNIME Grid Generation node, xyz coordinates of a central point within the active site of the protein were entered manually and the size of the grid box was increased to 22 Å for both the ligand length and the ligand diameter midpoint box to allow the ligands to be able to fit within the active site during docking. Within KNIME the Glide Grid Writer node writes the Glide grid in a zip format so that it can be stored and re-used when performing other dockings with the same receptor. Although the grid may simply be passed onto a node after generation, the Glide Grid Reader may alternately be used to open the saved Glide grid from the Glide Grid Writer node. This portion of the workflow prepares the receptor for Glide docking. However, ligands also require preparation for docking. The ligands were downloaded from SANCDB and were combined into a sdf file format using openbabel. The ligands were brought into the workflow with the use of the SDF Reader node and then converted to a mae (maestro) format file using the Molecule-to-MAE node since this is a requirement for most of the Schrödinger nodes in KNIME. The ligands were prepared using the LigPrep[85] node which generates high quality, all-atom 3D conformations for large numbers of drug-like molecules. This program can take as input 1D, 2D or 3D structures in Maestro format; in this case 3D structures were treated with LigPrep. The simplest use of LigPrep produces a single, low-energy, 3D structure with correct chirality's for each successfully processed input structure. However, LigPrep can also produce several structures from each input structure with various ionization states, tautomers, stereochemistries, and ring conformations. The output from LigPrep across the full SANCDB dataset was used in the ligand input for docking using the Glide-Ligand Docking node. Ligand Docking uses Glide to dock ligands into existing grids. The docking calculations can be carried out in either Standard Precision (SP), Extra Precision (XP)[86], or High-throughput Virtual Screening (HTVS) mode. In this study the docking calculations were performed using the Extra Precision (XP) mode and the ligand sampling was set to flexible to allow the greater conformational sampling within the active site, providing a greater chance of finding the correct docking mode.

The dockings were performed only in one step; therefore, it was important to use the highest level of precision that is XP (extra precision) for enhanced docking accuracy. The key novel features characterizing XP Glide scoring are (a) the application of large desolvation penalties to both ligand and protein polar and charged groups in appropriate cases and (b) the identification of specific structural motifs that provide exceptionally large contributions to enhanced binding affinity. The XP allows the ligand to dock as close to the centre of the binding site as it can and imposes desolvation penalties for burial of protein or ligand polar and charged groups. The desolvation term plays a major role in reducing false positives in XP virtual screening[86].

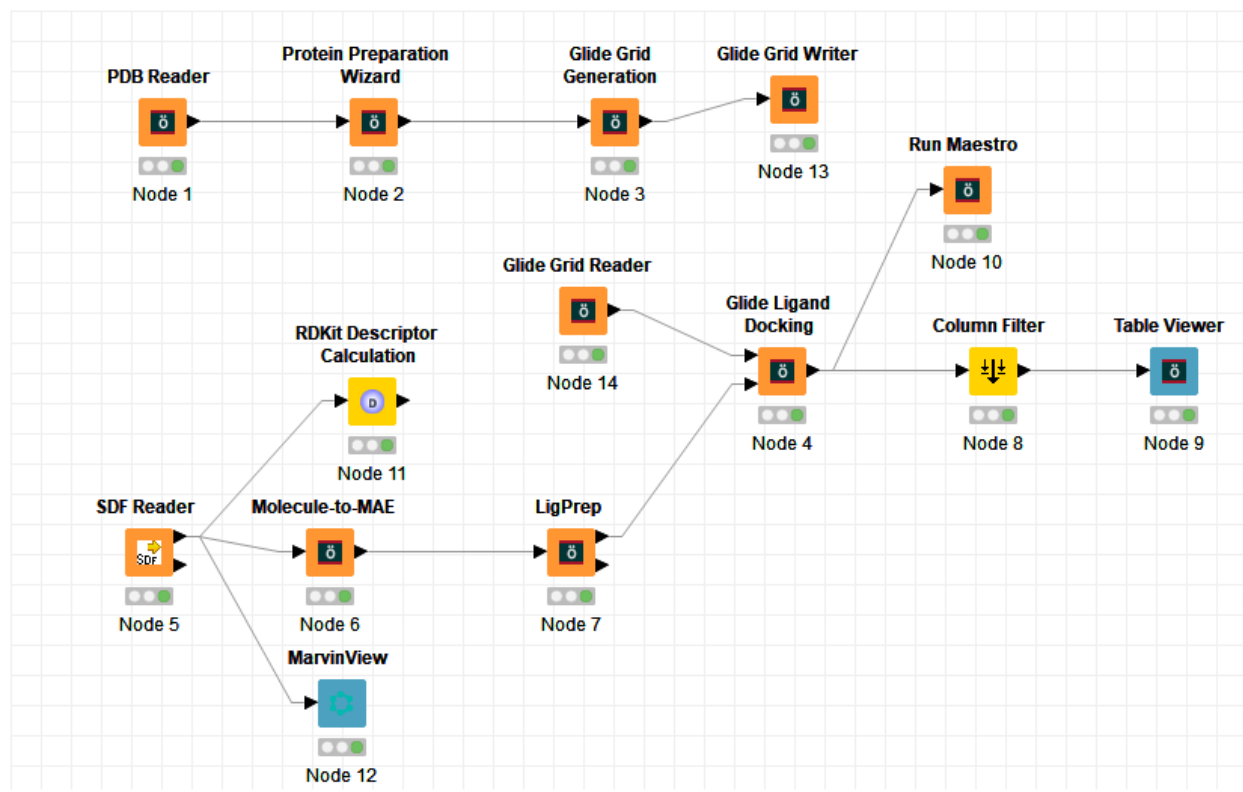


Figure 13: An example of a KNIME workflow used in this work for HTVS.

Although several Schrödinger nodes were used in KNIME, there are many other node sets from different software development groups, several of which focus on working with datasets of molecules in a chemoinformatics sense. Shown on Figure 13 above, are nodes that provide

information about the ligands, such as Marvin View and RDKit Descriptor Calculation. Marvin View provides the names or code of the ligands and the structure for easy visualization of molecules within KNIME. RDKit is a chemoinformatics library, and the RDKit Descriptor Calculation node may be used to calculate a plethora of different physicochemical properties of each ligand. The diagrams below depict the different physicochemical properties of all the 716 compounds used in this study.

2.1.2 Physicochemical properties analysis

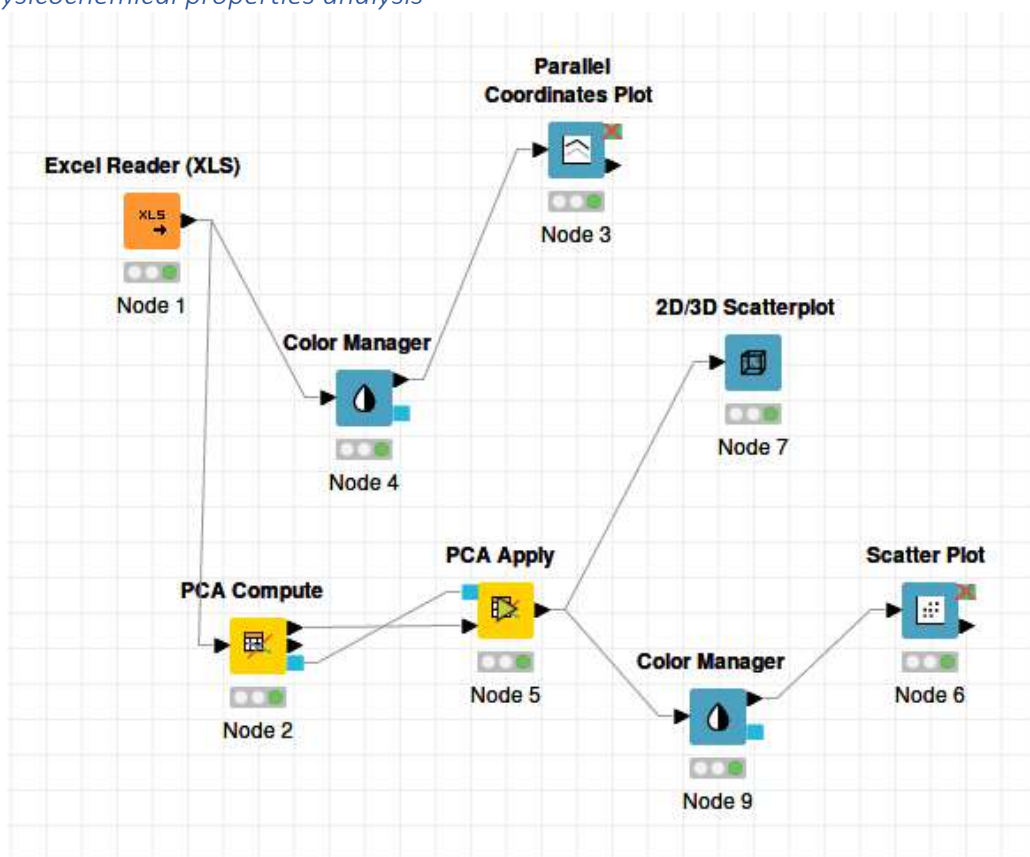


Figure 14: A workflow of the physicochemical properties' analysis of SANCDB compounds

The physicochemical properties of the SANCDB compounds were analyzed within KNIME using the workflow shown in Figure 14. The physicochemical properties of these compounds were previously saved in an Excel file format (using an Excel Writer node) and were uploaded into the Excel Reader node in KNIME. A Colour Manager node was used to assign the colours ranging between the minimum and maximum values of the SANCDB number. A parallel coordinates plot

displays multidimensional data as polylines over multiple parallel and evenly spaced axes. The data chosen for the plot was the Lipinski Rule of Five parameters (LogP, MW, NumHBD and NumHBA), TPSA and NumRings. These parameters were chosen because they are relatable to each other and are important in drug discovery to determine a drug-likeness of a compound. A principal component analysis (PCA) was performed using the PCA compute and PCA apply nodes. The PCA was performed on the 42 molecular quantum numbers (MQNs) that had been calculated using RDKit. The PCA compute node was used to perform a PCA on the 42-dimensional input data (across all SANCDB compounds), and the PCA apply node was used to apply a projection to three of the principal components on the given input data. The colour of the values were assigned in the Colour Manager node and a 2D Scatter Plot was performed. A 3D Scatter Plot was performed in the 2D/3D Scatter Plot node.

2.1.3 Docking results analysis

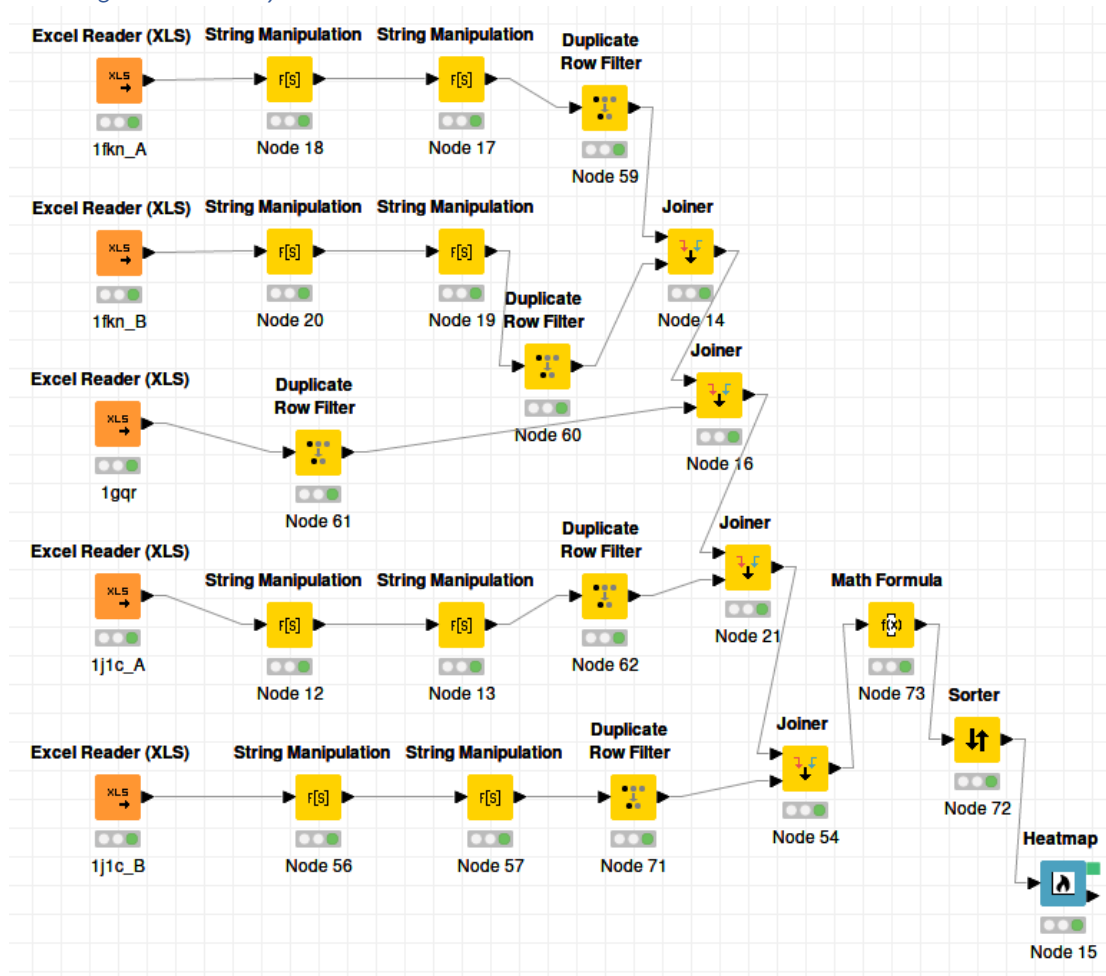


Figure 15: The analysis of the docking results KNIME workflow

Since the dockings per receptor were performed individually, the single-dimensional docking results were combined into a two-dimensional dataset relating each ligand docking to each target. The collation of the information was through the use of the workflow above in Fig. 15. The docking scores of each receptor were uploaded in the Excel Reader node. The String Manipulation node was used to fix the column headings in the docking scores table. Some ligands had docking scores of its different conformations, a Duplicate Row Filter node was used to remove all the duplicate docking scores from the same ligand and leave only the one with the lowest binding energy. Each table from all the dockings was combined into one table using the Joiner node. When all the tables are combined into one table, a Math Formula node which evaluates a mathematical expression based on the values in a row was used to express the docking scores according to their average.

The average docking scores were sorted in an ascending order using the Sorter node. Finally, when all docking results across all targets were collated (Figure 15 only shows collation of 5 of the 14 datasets), a heatmap of all the best ligand dockings performed in this study was produced using the heatmap node.

2.2 Molecular Dynamics

The dockings carried out on KNIME were executed using 14 structures against the 716 ligands. The structures for docking were prepared from 10 potential targets which were as follows 1FKN, 1GQR, 1J1C, 3LII, 4CTH, 4EY7, 4R12, 5E2W, 6BDZ and 6C2I (details of these targets are presented in the introduction of this thesis). As previously mentioned, the targets that were dimers were separated according to their chains A or B to make the total of 14 structures used. After the dockings were completed, the ligands from SANCDB identified as the best docked agents to these targets were downloaded for the simulations. The target and its corresponding agent were saved as a complex and prepared using protein preparation on Maestro using the default settings with an OPLS3 force field.

Molecular dynamics simulations were performed using Desmond[87], which is another program within the Schrödinger suite of programs, and which is accessible from within Maestro. A 100 ns simulation run was performed for six systems of the best docked ligands from KNIME; these six systems represent a range of targets, *viz*, secretases, acetylcholinesterase, tau protein and neprilysin. For systems that were duplicates of these enzymes, shortened simulations were performed for 50 ns of MD. The simulations all done on Desmond were prepared within Maestro but submitted to the CHPC (Centre for High Performance Computing) for faster computation of the trajectories.

Atomistic MD simulations help in better understanding of biological systems. They provide time dependent investigations of protein-ligand interactions and conformational dynamics of studied complex systems[88]. For the classical MD simulations of the predicted best binding compounds against Alzheimer's Disease targets using Desmond, the apo form of the target structure and selected compounds from docking simulations were incorporated into a combined complex as preparation for the MD simulations. All systems were solvated in an orthorhombic box with explicit TIP3P water molecules. Atomistic interactions were calculated with the OPLS3 force field. The ion and salt placement were excluded within 20.0 Å of the docked ligand. Sodium chloride was used for introduction of ions into the system with a concentration of 0.15 M.

A cut-off radius of 9.0 Å was applied for short-range van der Waals and Coulomb interactions. The systems were simulated under an isothermal–isobaric ensemble (NPT) with the temperature of 300 K and the pressure of 1 bar. A Nose–Hoover thermostat and Martyna–Tobias–Klein method were implemented to maintain the temperature and the pressure of the systems, respectively. A time step of 2.0 fs was used for the overall simulations. The systems were minimized and equilibrated with the default protocols of the Desmond. Finally, 100 ns non-constrained MD simulations were performed for each system, and the coordinates were saved every 9.6 ps. After performing the Desmond calculations remotely at the CHPC, the resulting files were transferred to the local computer and opened on Maestro for analysis. Analysis also included use within Maestro of the simulation interactions diagram wizard[88].

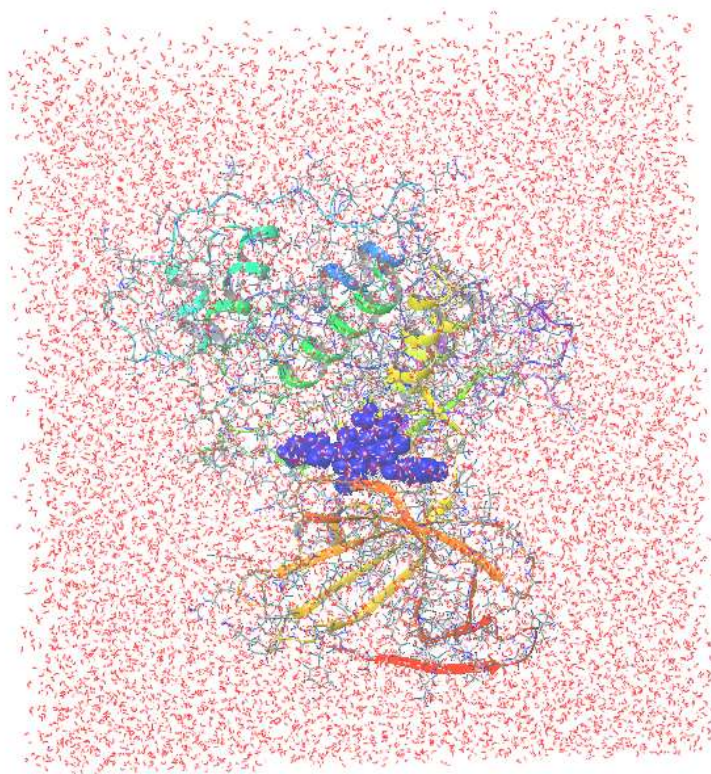


Figure 16: A snapshot of the constructed system of 1J1C_B protein and SANC00370 ligand complex MD simulation

The above snapshot (Fig. 16) is an example of a protein-ligand complex MD simulation performed by Desmond and depicts the solvent (water) in red, the protein (1J1C_B) in a chain and the ligand inside the binding pocket in blue. This snapshot was taken after the 100 ns MD simulation of this complex was performed. It serves as an example of how a constructed system of a protein-ligand complex MD simulation looks like.

Within the next Chapter 3, collating the results and discussion, explanations and observations of the results of these experiments are addressed. This chapter has two main sections dealing with the HTVS and the Molecular Dynamics.

CHAPTER 3

Results and Discussion

3.1 Physicochemical Properties and Molecular Docking Results

The *in-silico* docking simulations were performed for the following potential targets 1FKN_A, 1FKN_B, 1GQR, 1J1C_A, 1J1C_B, 3LII_A, 3LII_B, 4CTH, 4EY7_A, 4EY7_B, 4R12, 5E2W, 6BDZ and 6C2I (the subscripts _A and _B denote separate chains of homodimers where these have been separated). The 716 ligands from the SANCDB were docked against these targets as mentioned in chapter 2. The ligands were characterized in terms of chemical properties in addition to the docking experiments. The following details the results obtained from the *in silico* docking and evaluation of these as potential agents for the targets of Alzheimer's Disease.

3.1.1 Physicochemical properties of the ligands

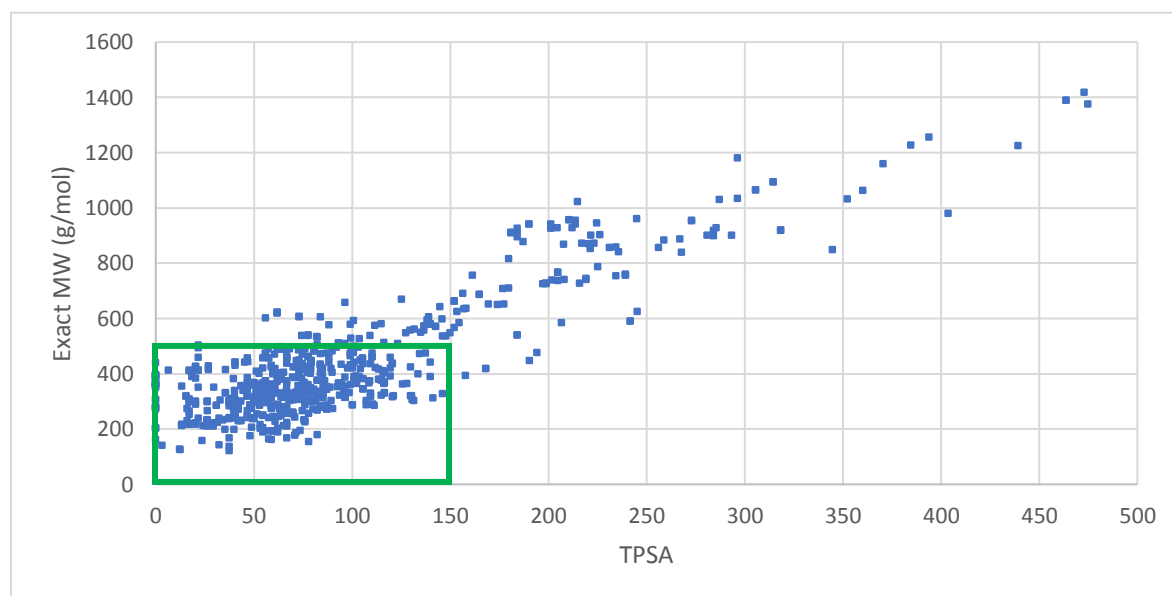


Figure 17: A graph showing the Molecular Weight (MW) vs Topological Surface Area (TPSA)

The above Fig. 17 shows a graph of molecular weight vs topological surface area (TPSA) (as calculated using RDKit nodes in KNIME) of all the compounds used from SANCDB. According to Lipinski Rule of 5, the molecular weight of a potential drug should be less or equal to 500 g/mol. The TPSA for a drug-likeness compound should be $\leq 140 \text{ \AA}^2$. The topological surface area

measures a compound's ability to permeate a cell membrane. The plot illustrates the distribution of compounds from SANCDB, and furthermore, most of these compounds are within the ranges of the Lipinski rules and therefore they have a potential to be used as drugs.

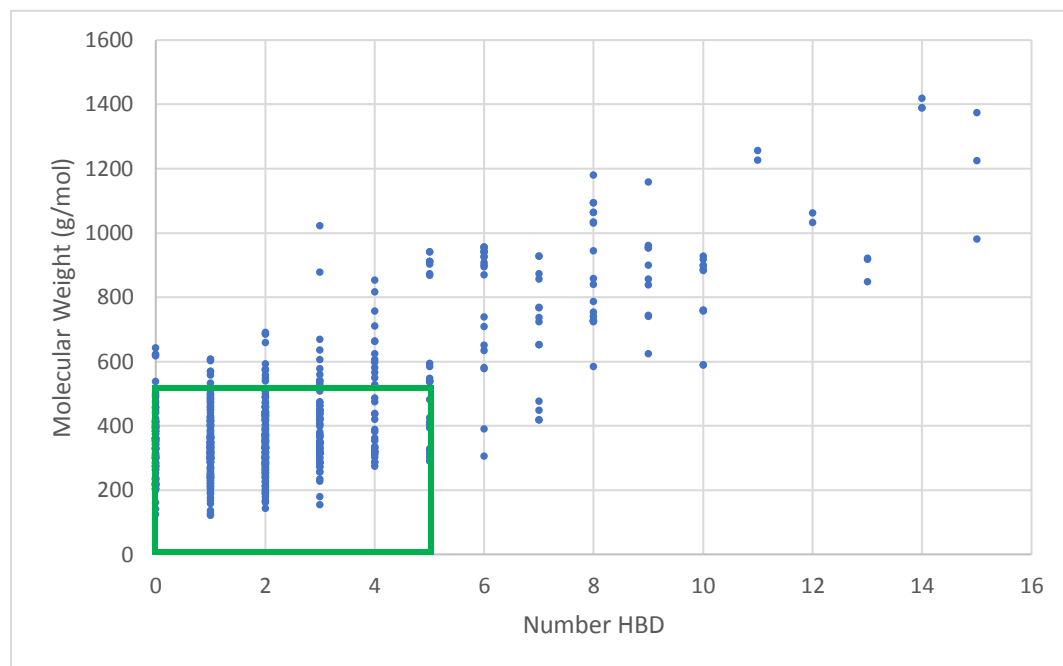


Figure 18: A graph showing the exact MW vs Number of HB

Figure 18 shows another way to explore the SANCDB dataset, with a plot of exact molecular weight against the number of hydrogen bond donor atoms within each compound in the dataset. Again, looking at Lipinski's rules, molecular weight should be ≤ 500 g/mol and the number of hydrogen bonds should be ≤ 5 . Similarly, to the previous graph, also most of these compounds follow the Lipinski rule of 5 according to this graph above.

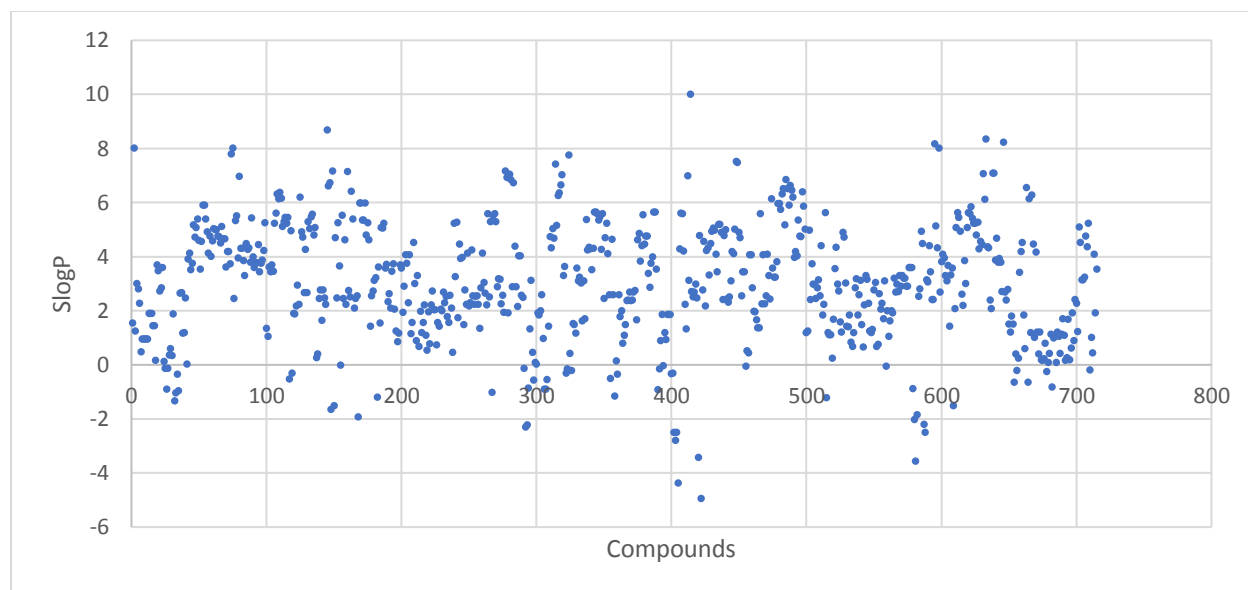


Figure 19: A graph showing the number of compounds vs SlogP

Without plotting properties against each other, it is also possible to explore the range of SlogP and SMR variation in the dataset.

Figure 19 simply plots SlogP against the molecule identity from SANCDB. According to Lipinski's rules, $\log P < 5$ and based on the above plot many of the 716 compounds used in HTVS fulfil another of Lipinski's rules. LogP is the octanol-water partition coefficient it really is a measure of the lipophilicity of a compound. From this simple plot a substantial number of compounds from SANCDB fulfil this requirement for drug-likeness. Although in this case there are many ligands that also do not fulfil this requirement, given that the Lipinski Rules of five is not absolute, ligands violating them can still be potential drugs.

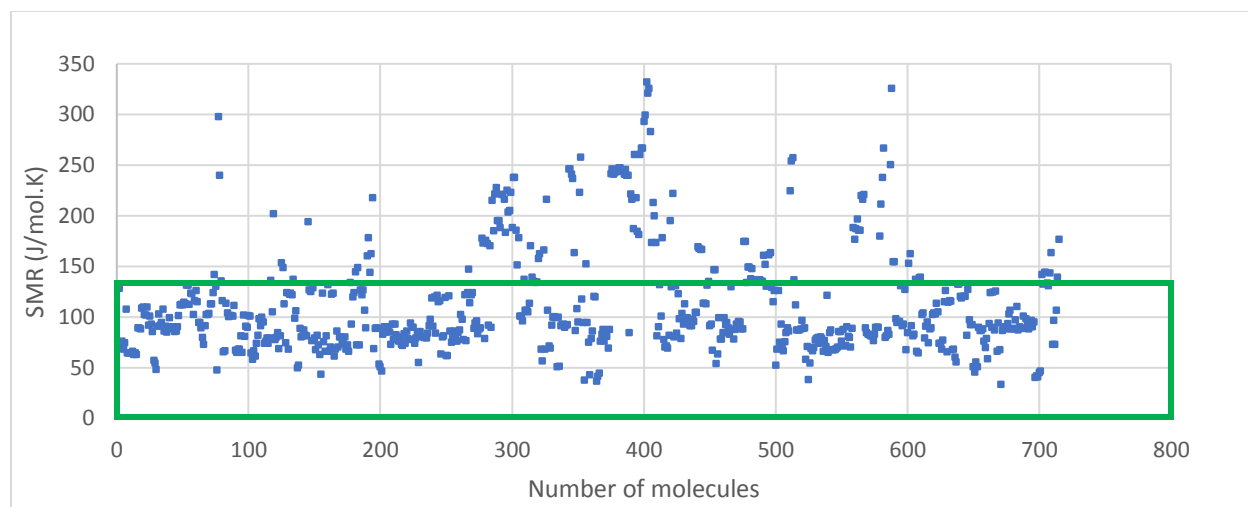


Figure 20: A graph showing the number of molecules vs molecular refractivity (SMR)

Molecular refractivity (SMR) is the measure of polarizability and depends on temperature and pressure. The SMR of a molecule should, for a drug-like molecule, be in the range of 40-130 J/mol.K. The molar refractivity is defined as:

$$A = \frac{4\pi}{3} N_A \alpha$$

Where $N_A = 6.022 \times 10^{23}$ is Avogadro's number and α is the mean polarizability of a molecule[89]. According to the plot in Figure 20, it can be observed that the majority of these SANCDB ligands are within the required range of SMR for a molecule and are polarizable.

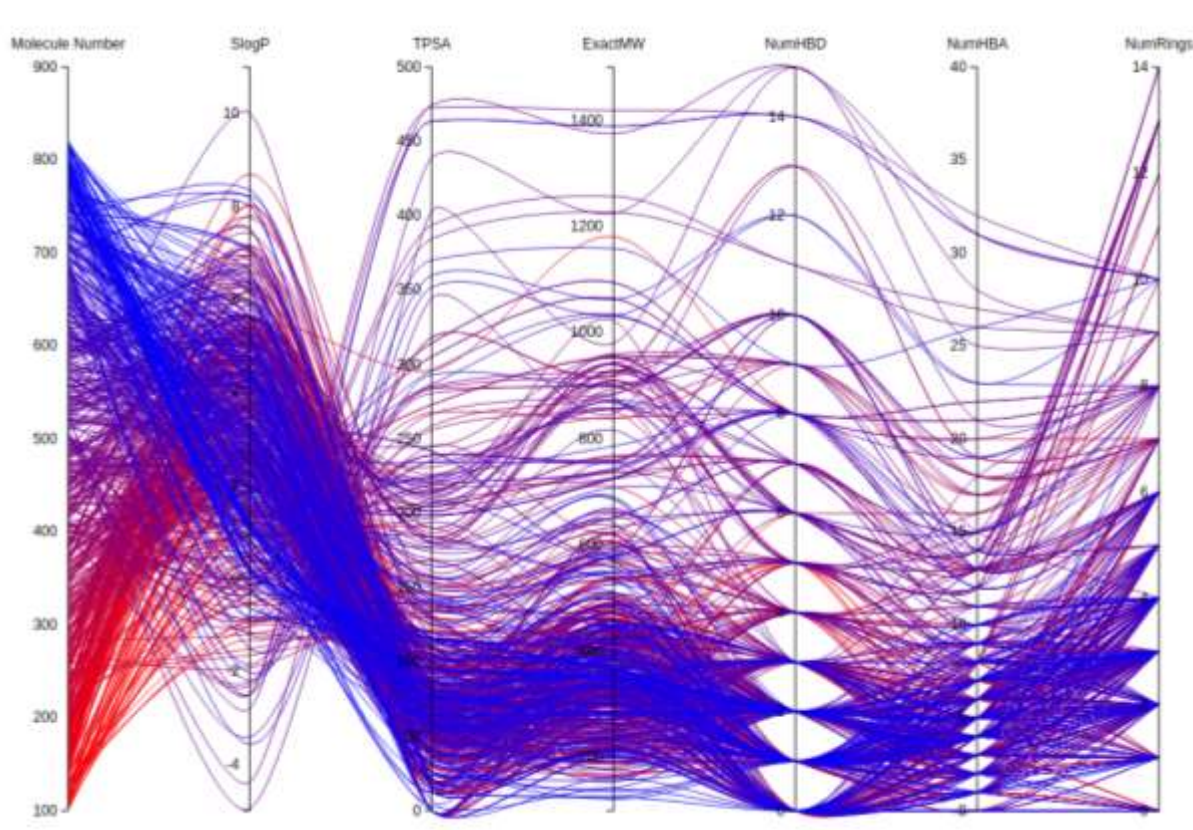


Figure 21: A Parallel Coordinates plot of the SANCDB compounds

The above diagram on Figure 21, depicts a parallel coordinates plot of the SANCDB ligands with axes for all Lipinski Rule of 5 parameters, the topological surface area (TPSA) and number of rings (NumRings), together with an axis for the compound SANCDB identity coding from SANC00101 to SANC00820. Although this is a static plot representing 7 dimensions of data, within KNIME this is dynamic and allows selection of compounds within specific ranges along each axis. Each line on the plot connects for a single compound the values of the different parameters that correspond to that SANCDB compound. The colouring in this case is according to the identity of the compound within SANCDB, with low ids coloured red and high ids coloured blue.

3.1.2 The MQN Plot

It is estimated that the number of possible small drug-like molecules reaches to about 10^{60} compounds and therefore visualization of large databases with so many compounds is a challenge.

However, the “chemical space” view of compound sets addresses this challenge by defining multidimensional property spaces in which dimensions are assigned to selected numerical descriptors of molecular structure, for instance Lipinski’s Rule of Five parameters. To describe the portion of chemical space in which a compound resides require a series of values or properties of that compound collected in a vector or fingerprint, each value defining a separate dimension[90]. Chemical space may also be represented as a property space whose dimensions measure a combination of structural parameters and predicted physicochemical properties[91]. Among the various approaches reported so far to represent chemical space Molecular Quantum Number (MQN)-space has the advantage of simplicity[90]. The MQN system can analyze large molecular databases and clusters compounds with similar structure, physicochemical properties and bioactivities[90], [91].

The MQN-system classifies organic molecules using 42 integer value descriptors of molecular structure including classical topological indexes such as atom and ring counts, and a few additional counts such as cyclic and acyclic unsaturation numbers, numbers of atoms and bonds in fused rings, and electrostatic charges predicted for neutral pH. The 42 MQNs form a 42-dimensional chemical space[91]. A principal component analysis (PCA) can be performed to project the multidimensional property space in a lower dimensionality space, typically a 2D- or 3D-space which can be visualized. This approach has been used to compare compound collections such as drugs, screening compounds, and natural products[92].

The following diagrams (Fig. 22-23) depict the MQN 2D and 3D scatter plots of the SANCDB compounds respectively. The similar compounds are grouped together according to the colour of the dots on the diagrams. According to the plots below, the points are scattered and therefore most of the SANCDB compounds are dissimilar. However, there are some similar compounds, and these are depicted by the red dots which are close to each other. It is interesting to note that while the SANCDB have a range of properties in terms of the physicochemical properties explored above, SANCDB covers only a small region of possible MQN space. This is attributed to the limited number of compounds (716) presently in the database, and that these compounds are all-natural products. The 3-D PCA plot is even more sparse in terms of chemical representation. However,

there is diversity within SANCDB that still makes it useful, although not exhaustive, in the search for lead compounds.

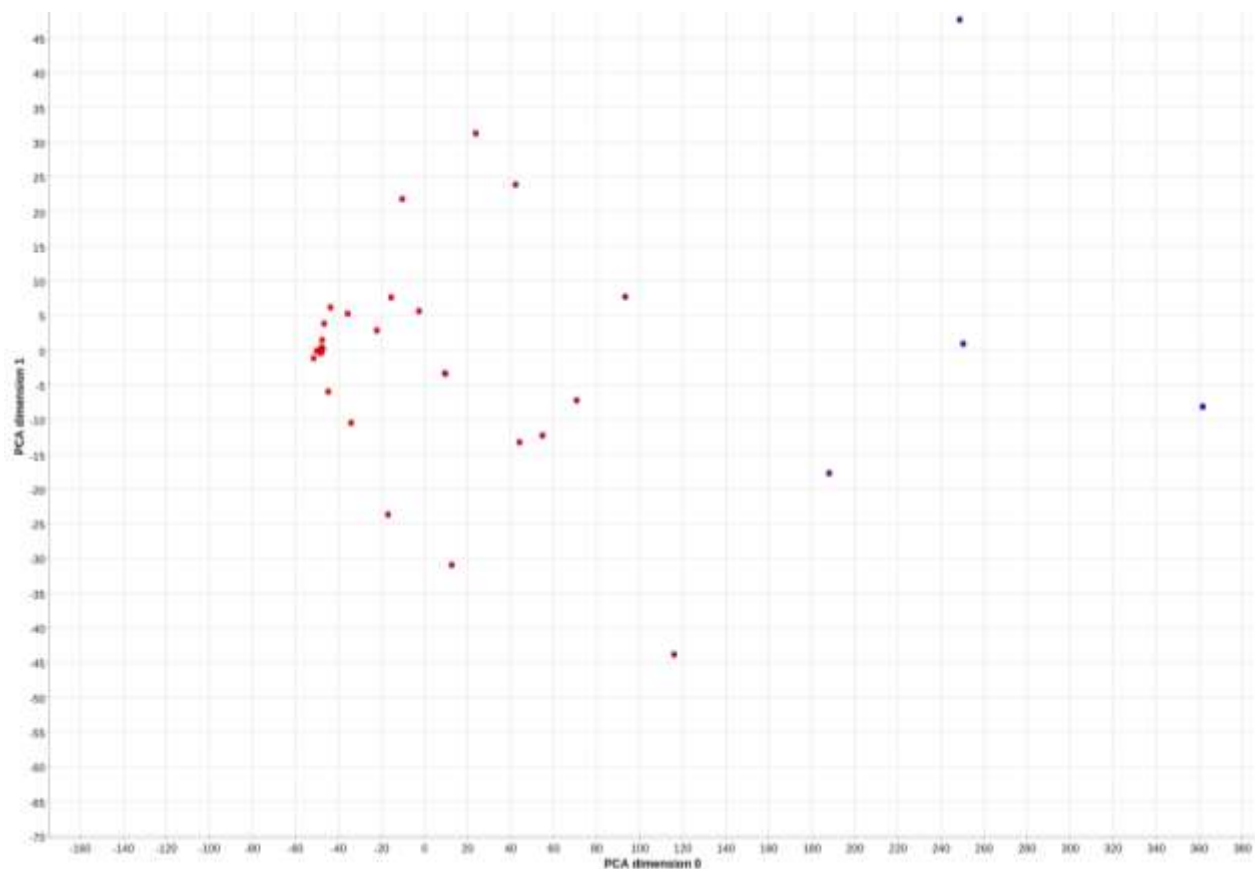


Figure 22: A 2D Scatter plot of the SANCDB chemical space: a projection of the 42D MQN properties of compounds in SANCDB using PCA.

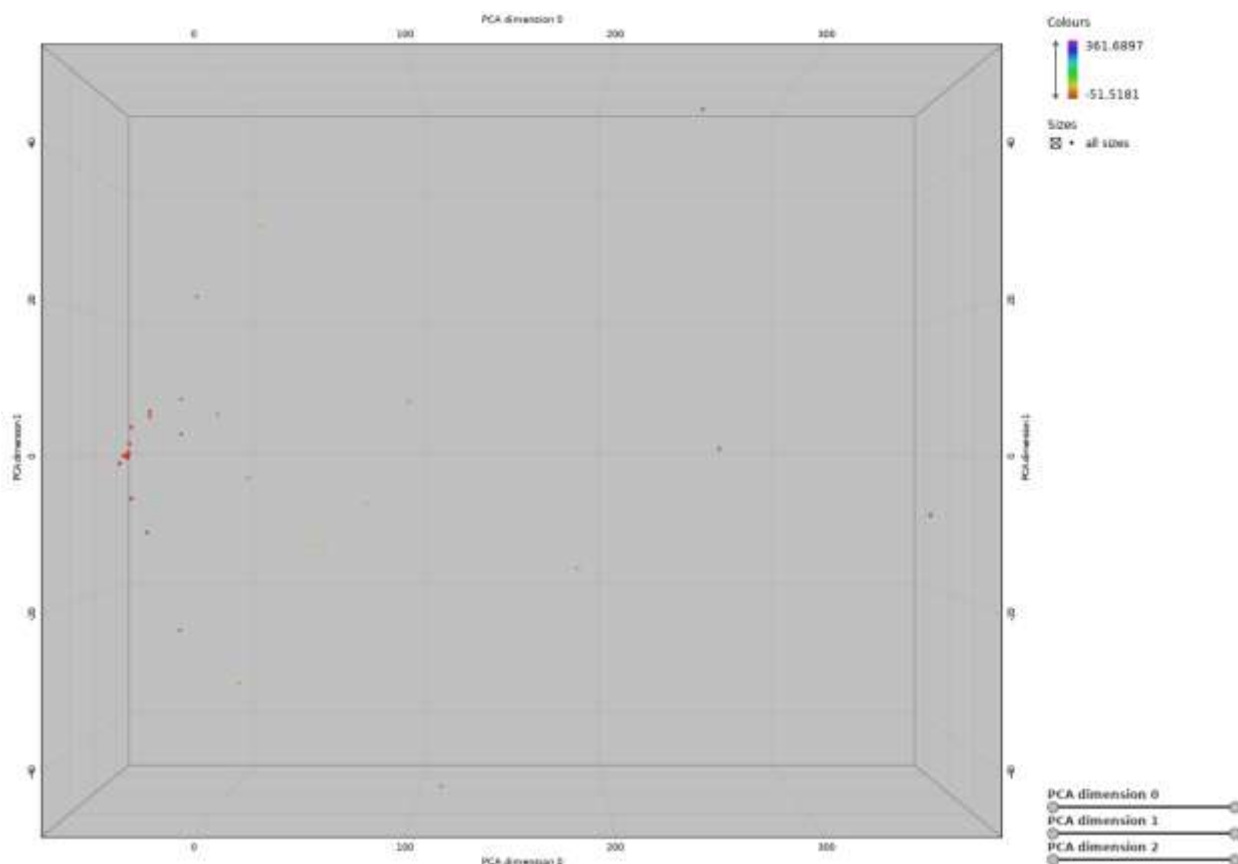


Figure 23: A 3D Scatter plot of the SANCDB chemical space: a projection of the 42D MQN properties of compounds in SANCDB using PCA.

3.1.3 Molecular Docking Results: docking scores

Receptor-ligand complexes provides atomic-level descriptions of molecular recognition between the ligand and the target. On the other hand, often it is difficult to make quantitative correlations between protein-ligand structures and their actual experimental binding affinities. Therefore, a wide variety of computational tools have become crucial components in drug discovery projects. Accurate prediction of ligand binding mode and affinity remains challenging mainly due to a protein's structural flexibility and inaccuracy of scoring functions[93]. In this project the assumption has been made that minimum calculated binding energy for a ligand to be considered with a target is ≤ -5 kcal/mol and anything > -5 kcal/mol is poor binding, although, in any case, only the systems with the most favourable binding energies are taken for further investigation.

The docking score used is a mathematical function used to approximately predict the binding affinity between the molecule and receptor after they have been docked[94]. **Table 1** below shows for each target the best observed binding ligands according to the calculated docking score. Most notable is that some ligands are not specific to the targets that they were docked against. These ligands are SANC00370 and SANC00103, and both ligands were the best binding ligands with three entirely different proteins. Based on the structures of these ligands (Figure 289 and 290 respectively), they contain many hydroxy groups and therefore interact with strong hydrogen bonding with many targets. SANC00103 was the best docked ligand to these proteins 4CTH, 4EY7_A and 1J1C_A (neprilysin, acetylcholinesterase and a tau protein kinase); although this ligand appears highly promiscuous, when comparing the binding scores protein 4EY7_A has the best binding energy at -10.56 kcal/mol (which is about 2 kcal/mol better than the next target). SANC00370 on the other hand was the best docked ligand to these proteins 1J1C_B, 6BDZ and 6C2I, but bound better by about 5 kcal/mol the protein 1J1C_B (a tau protein kinase) with a binding energy of -13.94 kcal/mol. However, these protein-ligand complexes will be further studied and will not be omitted from the study despite binding well to disparate proteins (the concern relates to issues of selectivity for this ligand).

All the ligands in Table 1 have good binding energies less than -5 kcal/mol and can be further studied to determine their potential as drug agents. The protein-ligand complex with the lowest binding energy is the binding to 1J1C_B of SANC00370, with a docking score of -13.94 kcal/mol. Many ligand complexes illustrated very good binding energies (in terms of our own criteria) and eight complexes were observed to have docking scores of less than -8 kcal/mol. The ligand SANC00269 which was docked with the target 3LII_A, although being the best ligand available from SANCDB to dock to the acetylcholinesterase 3LII_A, registered the worst docking score within this set of -5.06 kcal/mol.

a) SANCDB ligands

PDB ID	PROTEIN NAME	LIGAND NAME	BINDING ENERGY (kcal/mol)
1J1C_B	Tau protein	SANC00370	-13.94
4EY7_B	Acetylcholinesterase	SANC00374	-12.62
4EY7_A	Acetylcholinesterase	SANC00103	-10.56
1FKN_A	β -secretase	SANC00664	-10.30
1GQR	γ -secretase	SANC00111	-9.25
1J1C_A	Tau protein	SANC00103	-9.11
1FKN_B	β -secretase	SANC00562	-8.89
6C2I	β -secretase	SANC00370	-8.12
4CTH	Neprilysin	SANC00103	-7.63
5E2W	Tau protein	SANC00349	-6.55
3LII_B	Acetylcholinesterase	SANC00559	-6.49
6BDZ	α -secretase	SANC00370	-6.33
4R12	γ -secretase	SANC00256	-5.95
3LII_A	Acetylcholinesterase	SANC00269	-5.06

Table 1: *The binding energies of the best docked agent with their respective targets.*

Figure 24 displays a portion of the heatmap of the docking scores obtained from the high-throughput virtual screening (HTVS) performed in this study (and visualized within KNIME). This heatmap contains all the docking scores from all the receptors used for the study and they are shown according to the average binding energies in ascending order. The colour ranges from red (high/poor binding scores) to green (low/good binding scores), the black colour depicts the ligands for which no acceptable docking was obtained with the set parameters (due to ligands not fitting within the active site space specified). As previously mentioned, the KNIME workflow that collated these results intentionally was set to extract only the best docked energy for a ligand/target complex, so the diagram in Figure 24 represents only the top binding energies. The unshown portion of the heatmap (from Figure 24) are attached in Chapter 6 (Appendix A).

Figure 25 shows an example of the docking scores of all SANCDB compounds that were docked against the protein 1FKN_A. Each docking score is marked by a point, and the docking results are ordered from strongest to lowest binding. This graph serves as an example; all the other docking scores for other proteins (these can be found in Appendix 6.6.3 from Figure 305 -317) follow the similar trend as the graph in Figure 25. The lowest docking score for against this protein was -10.301 kcal/mol by ligand SANC00664 while the highest was above 5 kcal/mol which was the poorest docking with this protein.

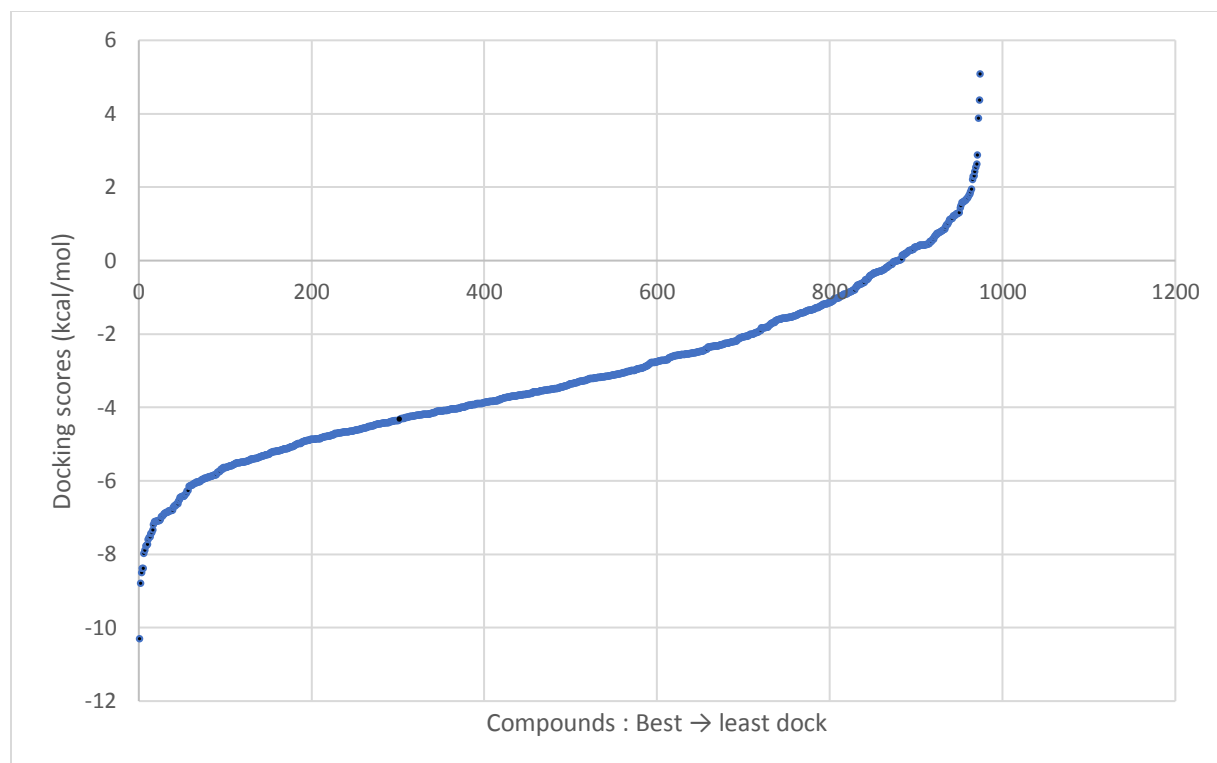


Figure 25: A graph showing all the docking scores of the compounds docked against IFKN_A

b) PubChem substructures' results

The best docked ligands obtained from the HTVS were used to provide search criteria for PubChem[81] in the search for similar structures for further docking. Some of the best docked ligands from SANCDB did not produce hits of similar structures from PubChem, however, those that produced structures are shown in **Table 2**, together with the number of PUBCHEM compounds produced. SANC00103 produced thousands of similar structures and therefore were filtered according to physicochemical properties to reduce the number to 50; these structures were first filtered according to physicochemical properties and the first 50 most drug-like compounds were used. The structures were accessed from PubChem in August 2018, due to updates within PubChem, a similar search at the present time might increase the number of structures retrieved. **Table 3** depicts the best docked compounds (informed by SANCDB) obtained from the PubChem structures corresponding to the targets with their docking scores. These best binding PubChem structures are named according to their PubChem compound CID. Only one of these ligands had a poor binding energy of -4.741 kcal/mol while the rest had good docking scores with the lowest binding observed to be -11.671 kcal/mol. All these structures except for the one docked with acetylcholinesterase (3LII_A) exhibited an improvement in the docking scores when compared to the SANCDB ligands docked against the same receptor. The structure with the biggest improved docking score was the one docked with neprilysin (4CTH) which was about 1.7 kcal/mol better than the SANCDB ligand SANC00103.

PDB ID	Best Ligands	No. of similar compounds
1FKN_A	SANC00664	30
1GQR	SANC00111	41
1J1C_A	SANC00103	50
3LII_A	SANC00269	2
4CTH	SANC00103	50
4EY7_A	SANC00103	50

Table 2: A table showing the number of substructure compounds of the best docked ligands obtained from HTVS with their respective proteins.

PDB ID	PROTEIN	PubChem CID	BINDING ENERGY (kcal/mol)	BEST SANCDB ENERGY (kcal/mol)
4EY7_A	Acetylcholinesterase	Molecule: 68103282	-11.67	-10.56
1FKN_A	β -secretase	Molecule: 101608678	-10.63	-10.30
1GQR	γ -secretase	Molecule: 11777124	-10.06	-9.25
1J1C_A	Tau protein	Molecule: 122542902	-9.30	-9.11
4CTH	Neprilysin	Molecule: 59370901	-9.40	-7.63
3LII_A	Acetylcholinesterase	Molecule: 11209156	-4.74	-5.06

Table 3: The binding energies of the best docked substructure agents with their respective targets. Bold indicates an improvement from the SANCDB best binder.

c) Lipinski Rule of Five

In the drug discovery setting, the Lipinski Rule of 5 predicts that poor absorption or permeation is more likely when there are more than 5 H-bond donors, 10 H-bond acceptors, the molecular weight is greater than 500, and the calculated Log P (Clog P) is greater than 5[95]. These physicochemical parameters are associated with acceptable aqueous solubility and intestinal permeability and comprise the first steps in oral bioavailability. If a compound fails the RO5, there is a high probability that oral activity problems will be encountered. However, passing the RO5 is no guarantee that a compound is drug-like[96].

More than 60% of compounds in the database violate none of Lipinski's rules and these are included in the approximately 80% of compounds that violate, at most, one of the rules. This is a promising indication that compounds in the database have potential to be investigated as prospective drug molecules[61]. Table 4 below depicts the Lipinski Rule of Five of the best docked ligands from SANCDB (according to receptor) and their violations of the rule. Out of the 14 receptor-ligand complexes, 5 ligands violate the Lipinski Rule of Five however, SANC00370 is the lead compound for 3 different proteins so there are only 3 violations. The ligands that violate the rule are SANC00664, SANC00370 and SANC00559.

PDB ID	LIGAND NAME	Lipinski's Rule of Five				Violations
		Molecular weight (≤ 500 Da)	HBD (≤ 5)	HBA (≤ 10)	logP (≤ 5)	
1FKN_A	SANC00664	758.45	10	14	-0.0505	3
1FKN_B	SANC00562	390.13	6	8	0.4469	1
1GQR	SANC00111	239.15	1	4	0.9509	0
1J1C_A	SANC00103	306.07	6	7	1.2517	1
1J1C_B	SANC00370	624.21	9	15	-1.0159	3
3LII_A	SANC00269	312.10	5	9	-1.9163	0
3LII_B	SANC00559	590.12	10	13	3.4369	3
4CTH	SANC00103	306.07	7	6	1.2517	1
4EY7_A	SANC00103	306.07	7	6	1.2517	1
4EY7_B	SANC00374	325.13	1	5	2.8816	0
5E2W	SANC00349	248.11	1	4	1.4936	0
6BDZ	SANC00370	624.21	9	15	-1.0159	3
6C2I	SANC00370	624.21	9	15	-1.0159	3
4R12	SANC00256	180.05	3	5	-0.0012	0

Table 4: *The Lipinski Rule of Five of the best docked ligands*

3.1.3 The HTVS Protein-Ligand Interactions

Due to the plethora of data, not all the protein-ligand complexes were examined from the point of view of evaluating the ligands as potential leads. Therefore, the six complexes with the lowest binding energies with each receptor (representing the different therapeutic targets investigated as potential Alzheimer's Disease agonists) will be briefly discussed. The targets of these protein-ligand complexes are as follows 1FKN_A, 1J1C_B, 4EY7_B, 4R12, 4CTH, and 6BDZ. The

following protein-ligand complex interactions are represented by the target's name. Most of these interactions were analyzed within Discovery Studio Visualizer and Schrödinger Maestro.

3.1.3.1 Aromaticity

Although π - π stacking, defined as an attractive, non-covalent interaction between aromatic rings, is not as emphasized as hydrogen bonds and hydrophobic contacts, it still is important in biological recognition and biomolecular structural organization. Dispersive interactions and electrostatics are the predominant components of π - π interactions, and as such many molecular force fields and scoring functions simply rely on the van der Waals and Coulomb potentials to account for these without introducing an explicit π - π interaction term. There is therefore heavy dependence on this approximation, and this has been the subject of careful evaluation relating experiment to π - π stacking prediction from computer models [97].

Aromatic systems geometrically and interactively have a large influence on protein-ligand recognition and therefore on drug design. π -systems of aromatic rings give rise to three types of interactions: (i) π - π , (ii) cation- π and (iii) X-H- π . These interactions give rise to three different types of geometries, namely, (A) Edge-to-face (or T-shaped), (B) Face-to-face and (C) Parallel displaced (or offset stacked) interactions as shown in Figure 26[98].

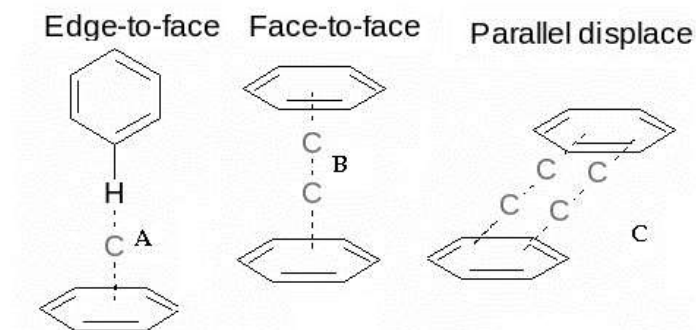


Figure 26: Schematic representation of the three types of geometries found in aromatic-aromatic interactions[98]

Figures 27-32 explore these types of aromatic interactions between receptor and ligand, coloured by the observed type of interaction. Ligand (SANC00664) docked with target 1FKN_A (Figure 27) does not have any aromatic ring, and this is evident in Figure 27 with the low level of colouring

for aromatic interactions (edge to face, blue and face to face, orange). Similarly, these types of interactions are not predominant for targets 1J1C_B and 6BDZ (Figures 30 and 31) where these aromatic interactions are at best weak with their docked ligand (in both cases this is SANC00370). However, unlike SANC00664, the ligand SANC00370 includes two benzene rings at separate ends of the molecule's chain. In both complexes, the interaction between these benzene rings on the ligand and protein atoms depict a weak edge-to-face geometry.

Figure 28 depicts the complex of 4EY7_B and SANC00374, where there are a wide variety of dominant aromatic interactions, varying between edge-to-face and face-to-face. The ligand is a flat molecule with interconnected aromatic rings which therefore explains the prevalence of these observed interactions. The 4R12 and 4CTH (Figure 29 and 32) complexes also feature a mixture of dominant edge-to-face and face-to-face geometries, however, the edge-to-face geometry is the most notable.

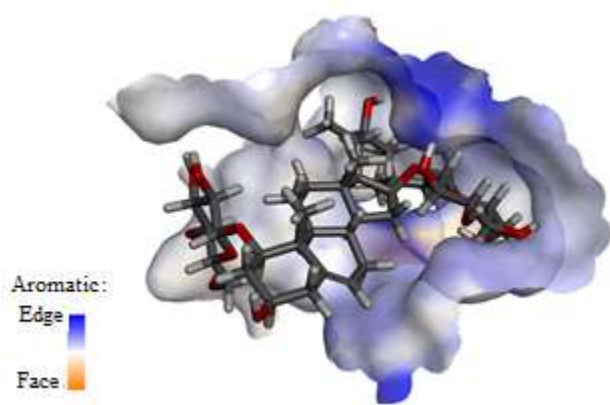


Figure 27: 1FKN_A-SANC00664

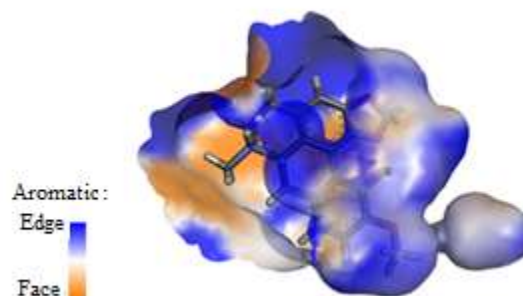


Figure 28: 4EY7_B-SANC00374

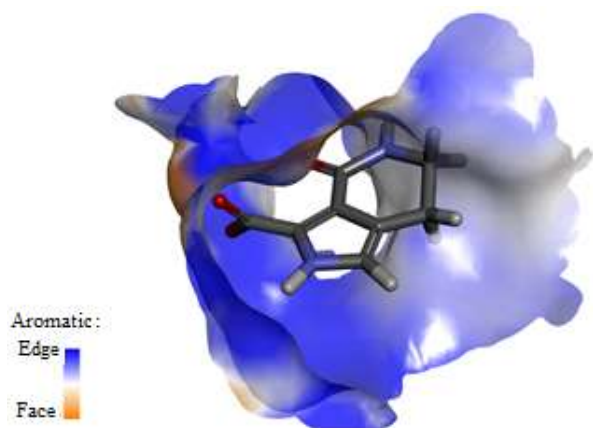


Figure 29: 4R12-SANC00256

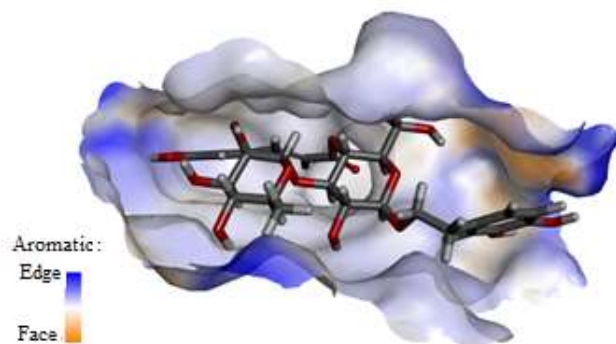


Figure 30: 1J1C_B-SANC00370

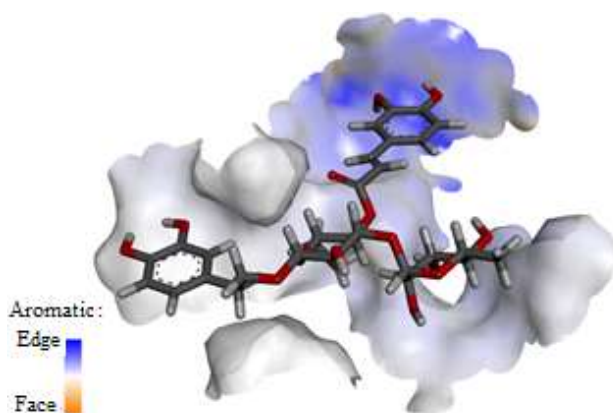


Figure 31: 6BDZ-SANC00370

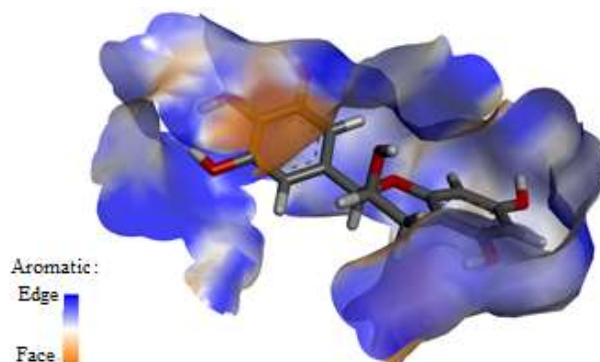


Figure 32: 4CTH-SANC00103

3.1.3.2 Solvent Accessible Surface Area (SASA)

The Solvent Accessible Surface Area (SASA) is the surface area of a molecule accessible by water molecules. In Figures 33-38 the least accessible surface is marked at 10.0 \AA^2 (green) while the highest is 25.0 \AA^2 (blue). The SASA of the binding surface of the protein-ligand complexes includes protein surface having a high solvent accessible surface area of above 20.0 \AA^2 together with pockets having a low surface area less than 12.5 \AA^2 . Therefore, it is likely that these ligands are replacing significant water from the active site upon binding even though there are areas of low accessibility of water for some pockets of the protein. Figure 33 depicts a ligand that extends

well into areas that are the most accessible by water (blue region). The ligand contains a few O or O-H bonds in this region, and it is most likely that hydrogen bonding takes place between them and the surrounding water molecules in the system.

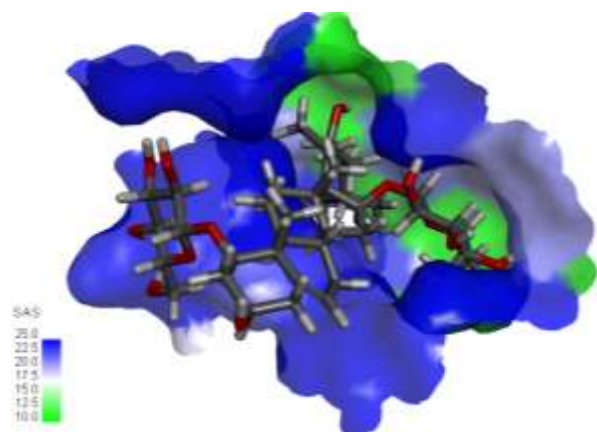


Figure 33: 1FKN_A- SANC00664

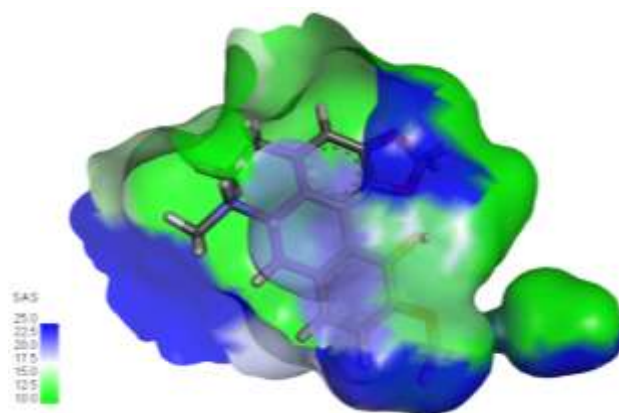


Figure 34: 4EY7_B-SANC00374

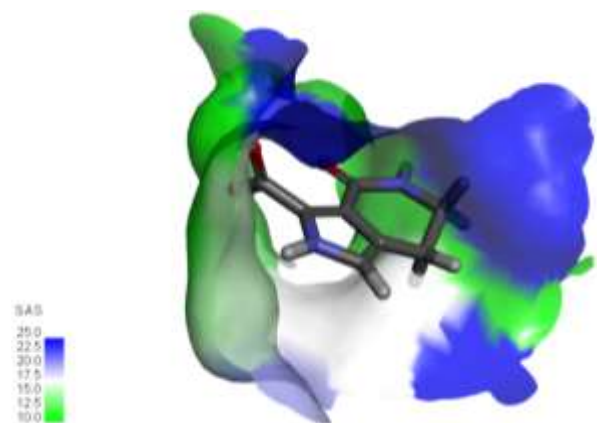


Figure 35: 4R12-SANC00256

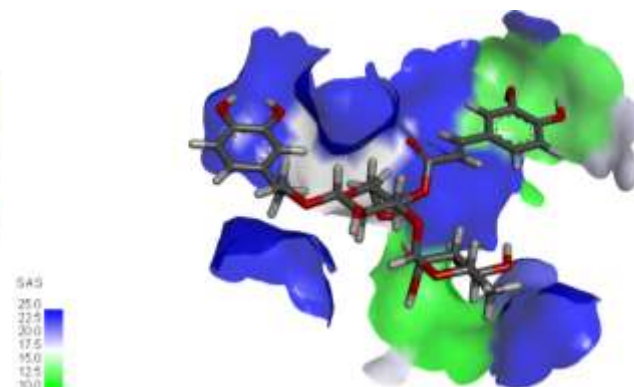


Figure 36: 6BDZ-SANC00370

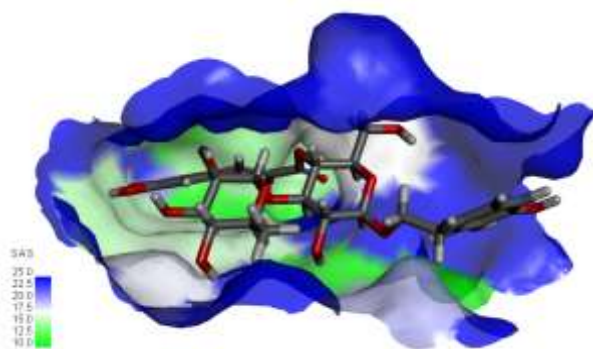


Figure 37: 1J1C_B-SANC00370

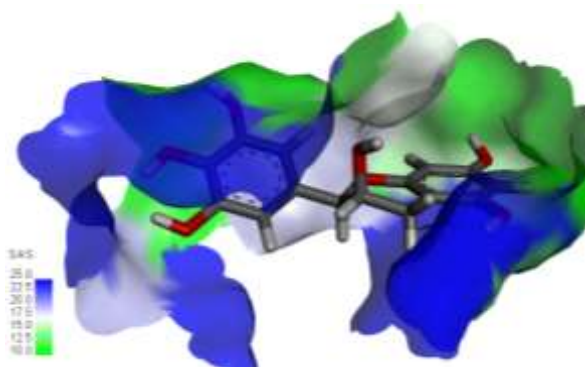


Figure 38: 4CTH-SANC00103

3.1.3.3 Protein-Ligand complex interaction diagrams

A third metric to follow in the evaluation of potential inhibitors involves inspection of interaction diagrams of the docked ligand with the respective protein. These interaction diagrams were performed from two different computer programs, Discovery Studio Visualizer and Maestro. Figure 39-44 depicts 3D diagrams taken from Discovery Studio while Figure 45-50 shows 2D diagrams from Maestro.

a) *Three-dimensional interaction diagrams*

The 3D diagrams in Figure 39-44 detail the different interactions between the protein and our best binding SANCDB ligands. Also identified are the protein residues and the nature of the interaction with the ligand. SANC00256 is a small molecule, suitable for fitting into the active site of 4R12, but as such has the least number of interactions compared with the other receptor-ligand complexes (Figure 42). A large number of binding interactions will increase the ability of the ligand to maintain its pose in the binding pocket during dynamics. It is interesting to note the number of hydrogen-bonding interactions that are highlighted for 1FKN_A-SANC00664 and 6BDZ-SANC00370 complexes (Figures 39 and 44). Also interesting is for instance in the ligand bonded to 1J1C_B, where although there are two aromatic groups, only one shows pi interactions (Figure 40).

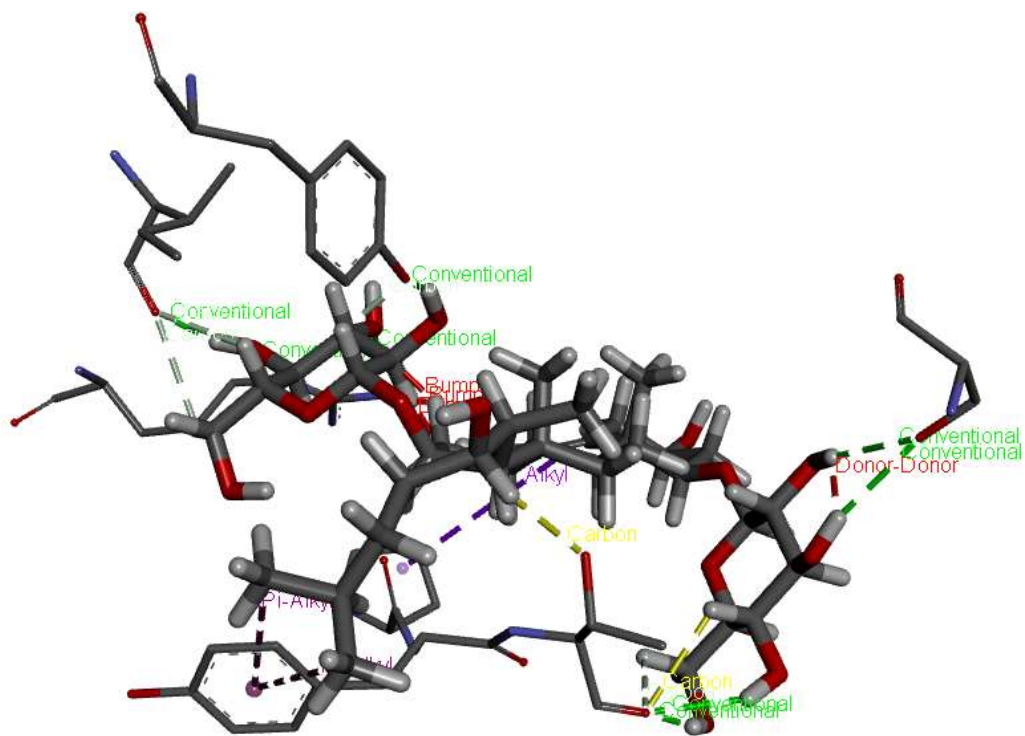


Figure 39: 1FKN_A- SANC00664

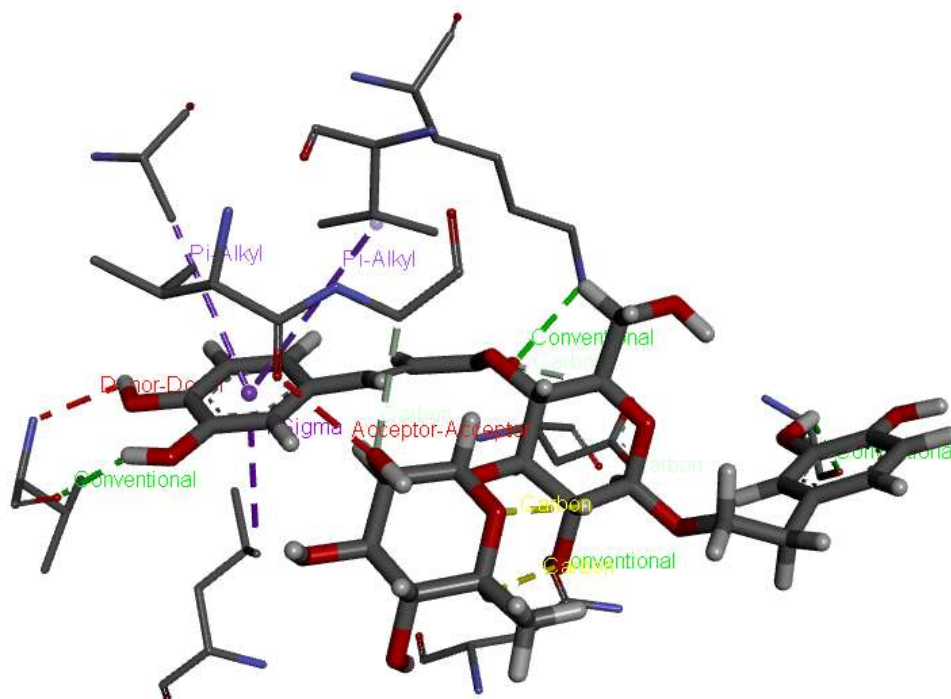


Figure 40: 1J1C_B-SANC00370

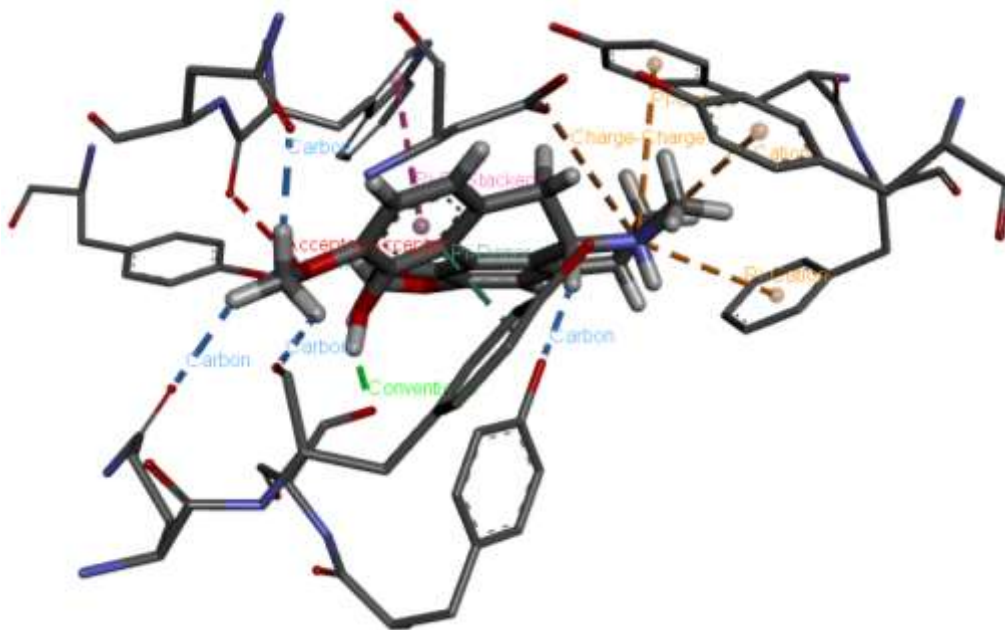


Figure 41: 4EY7_B-SANC00374

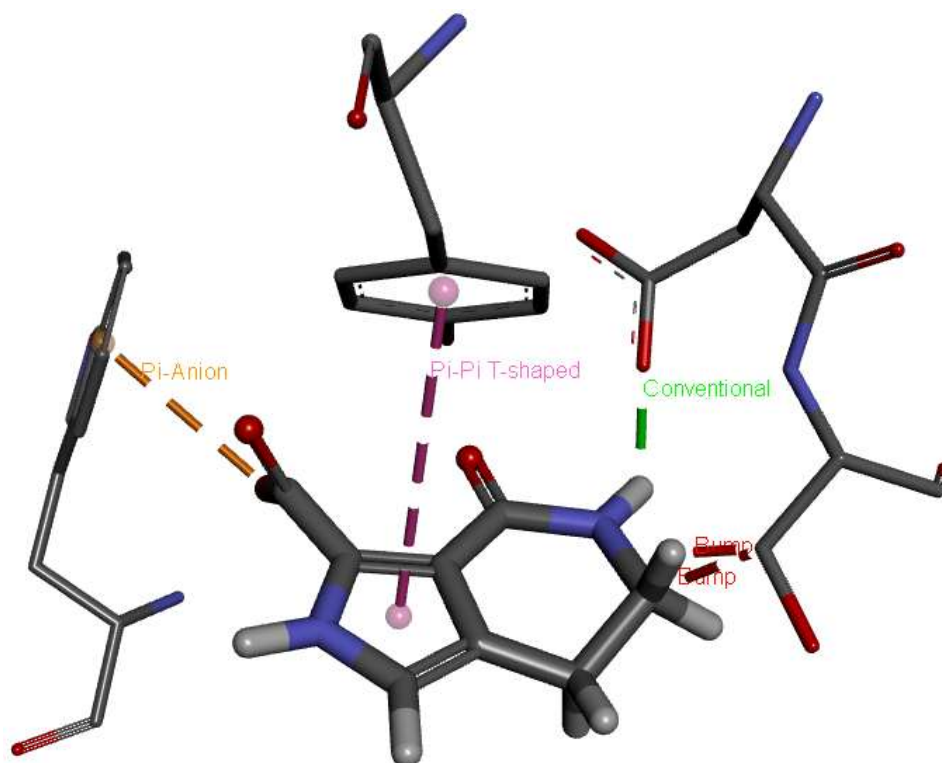


Figure 42: 4R12-SANC00256

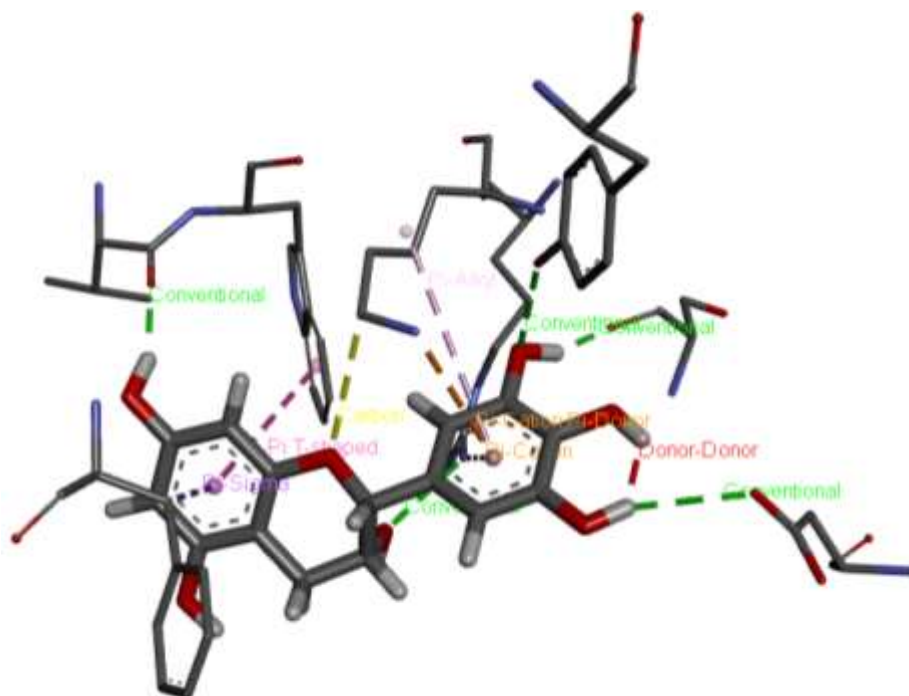


Figure 43: 4CTH-SANC00103

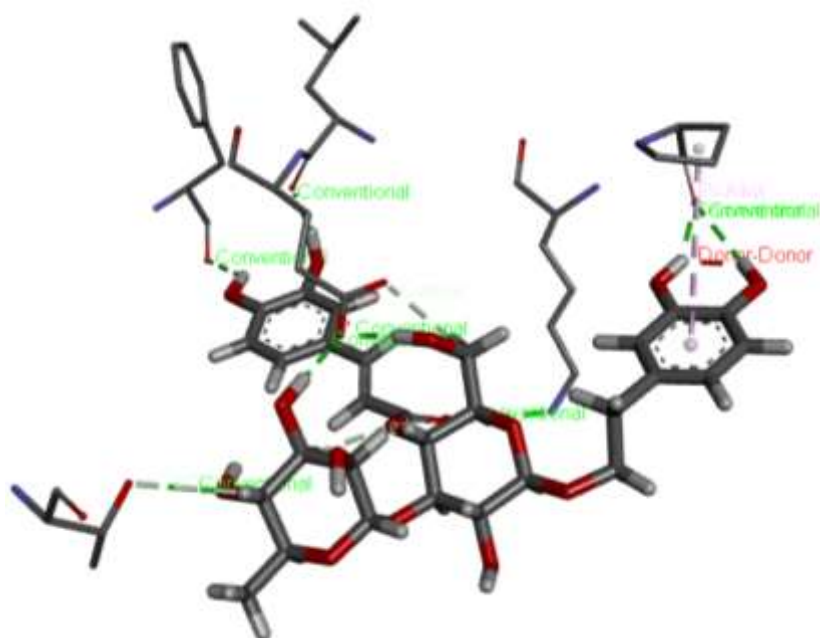


Figure 44: 6BDZ-SANC00370

b) *Two-dimension representations of protein-ligand complexes*

In Figures 45-50 are detailed schematic diagrams of ligand atom interactions to the protein residues. According to this analysis, most of the interactions identified from the complexes are H-bonding and π -cation interactions. During the docking process, ligand preparation (ligprep) explored and provided different protonation states for ligands, and this is observed with the best docked pose for SANC00256 in Figure 48, where the carboxylate is observed rather than the carboxylic acid. The richness of hydrogen bonding interactions with the 1FKN_A-SANC00664 and 6BDZ-SANC00370 are again observed (Figures 45 and 49).

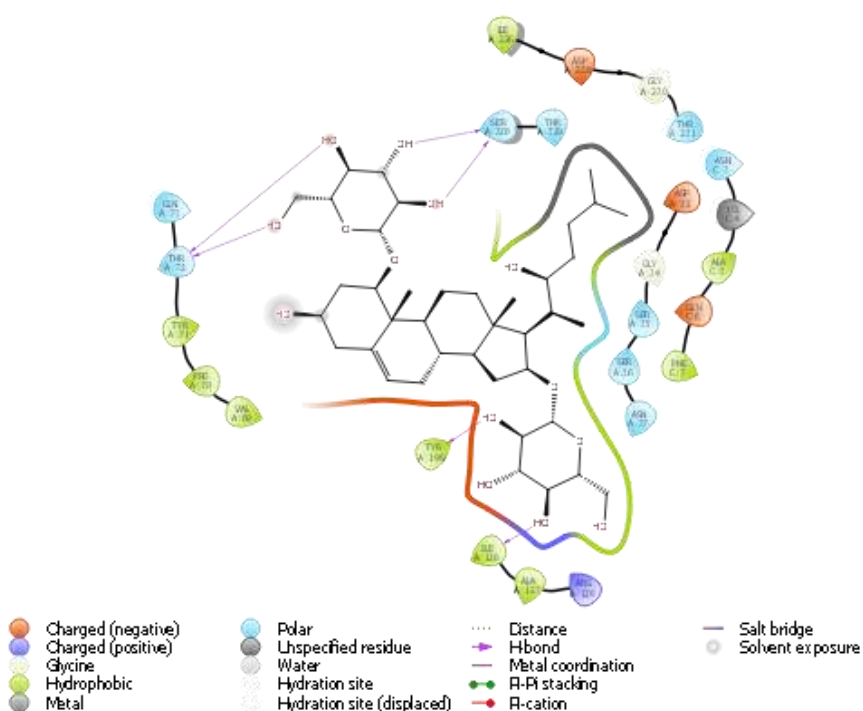


Figure 45: 1FKN_A-SANC00664

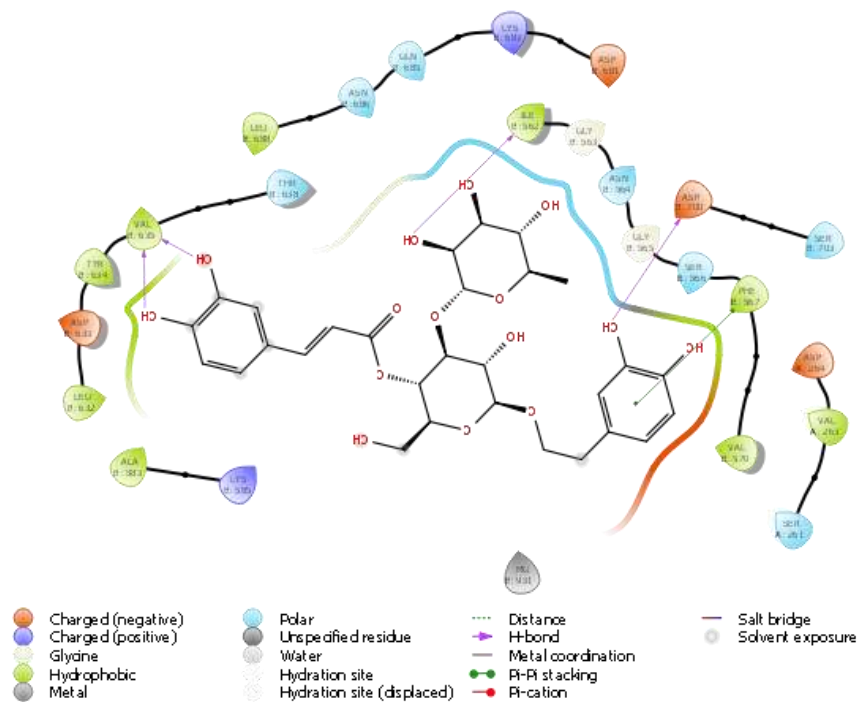


Figure 46:1J1C_B-SANC00370

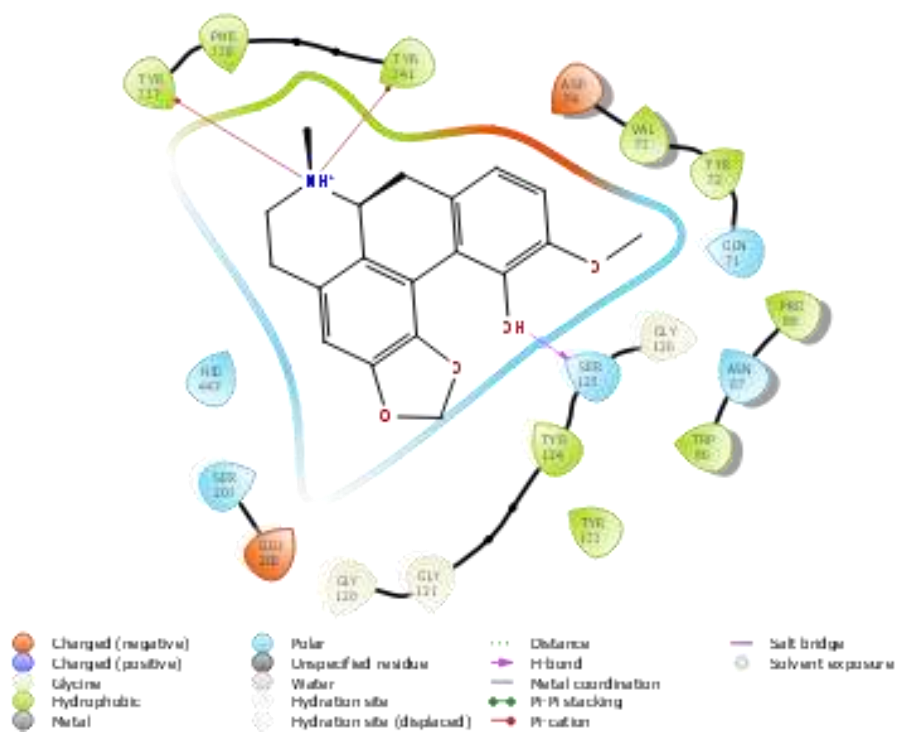


Figure 47: 4EY7_B-SANC00374

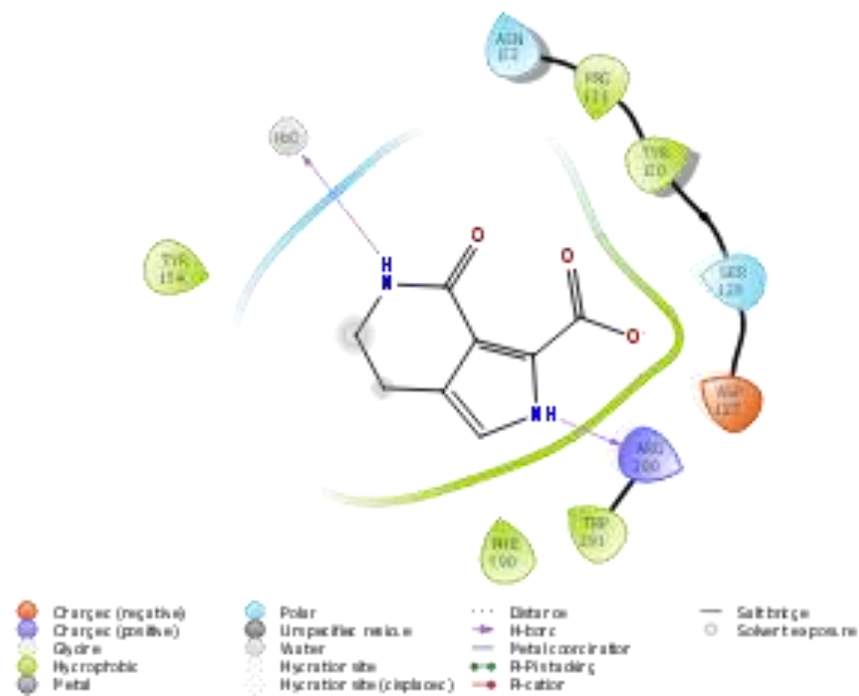


Figure 48: 4R12-SANC00256

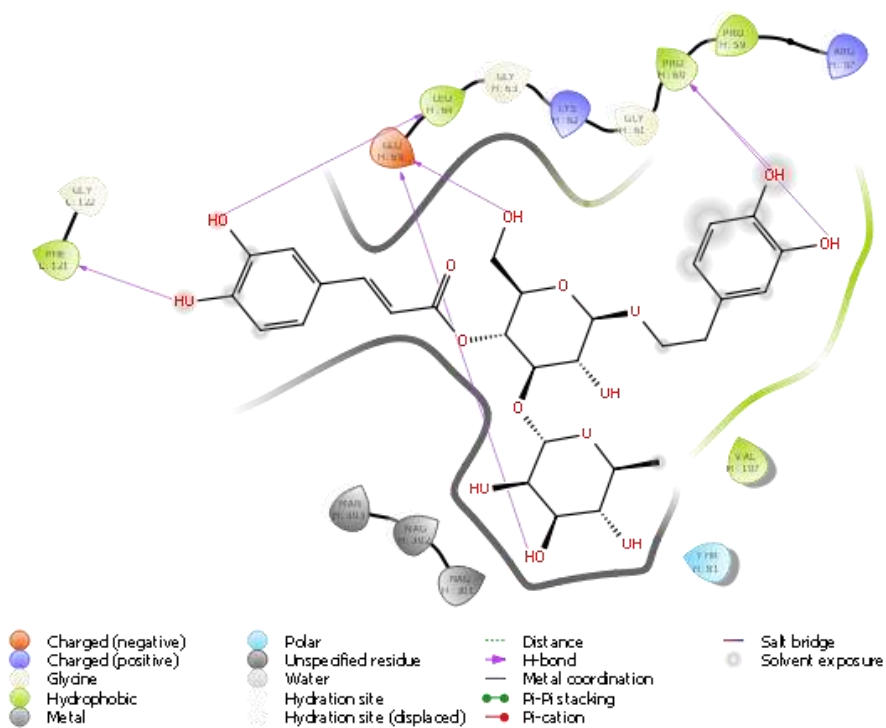


Figure 49: 6BDZ-SANC00370

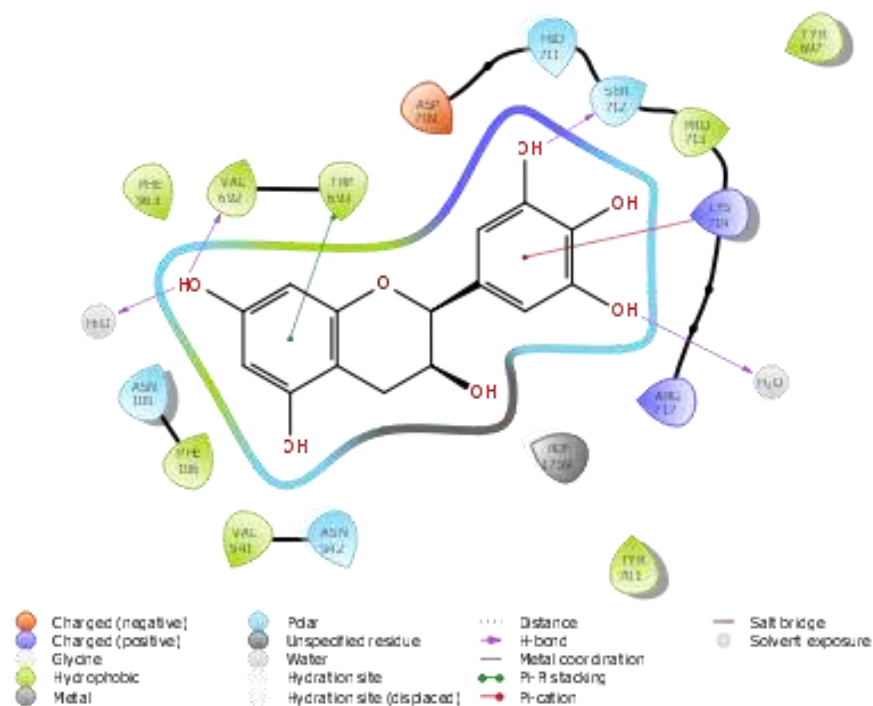


Figure 50: 4CTH-SANC00103

3.1.3.4 H-Bond

Hydrogen bonding has already been explored in previous interaction diagrams. The hydrogen bond is a highly attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and a separate electronegative atom in the same or in a different molecule [99]. The greater the electronegativity of the hydrogen bond acceptor, the stronger the hydrogen bond is likely to be. The hydrogen bond is one of the strongest intermolecular attractions, but weaker than a covalent or an ionic bond. Further information on hydrogen bond interactions will also be discussed in the next section on molecular dynamics.

The HTVS protein-ligand complex hydrogen bonding is depicted in the diagrams from Figure 51-56. In this analysis hydrogen bonds are differentiated by donor/acceptors. Where the protein is the H-bond acceptor, the surface is represented with the colour green on the diagrams below, and this is the most common with all the protein-ligand complexes and is due to the OH groups contained in the ligands (which are the corresponding H-bond donors in H-bond formation). Again, with exploration of 1FKN-SANC00664 and 6BDZ-SANC00370 complexes (Figures 51

and 54), the extent of hydrogen bonding is seen; however, the differentiated sites for hydrogen bonding are now evident within the active site surface.

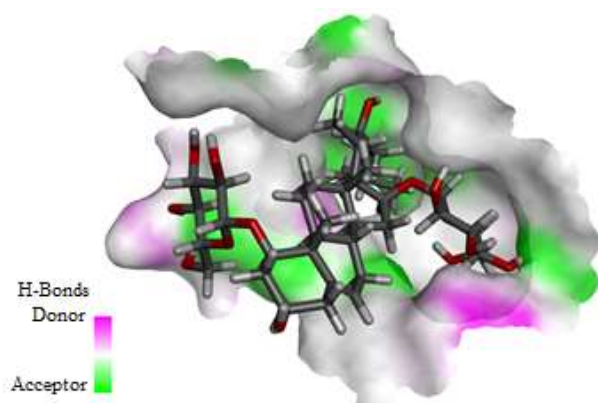


Figure 51: 1FKN_A-SANC00664

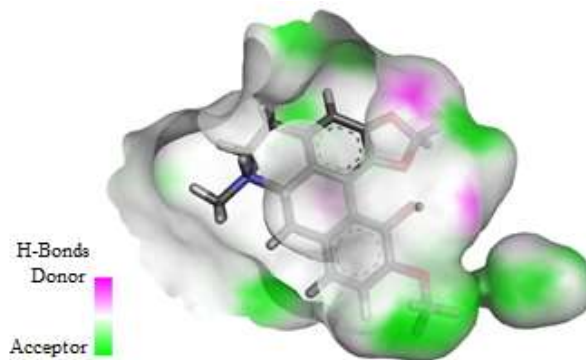


Figure 52: 4EY7-SANC00374

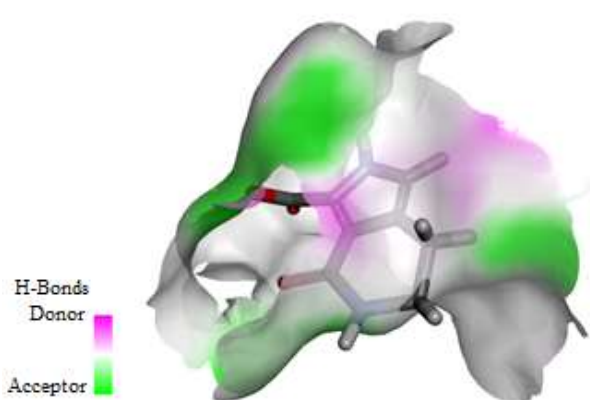


Figure 53: 4R12-SANC00256

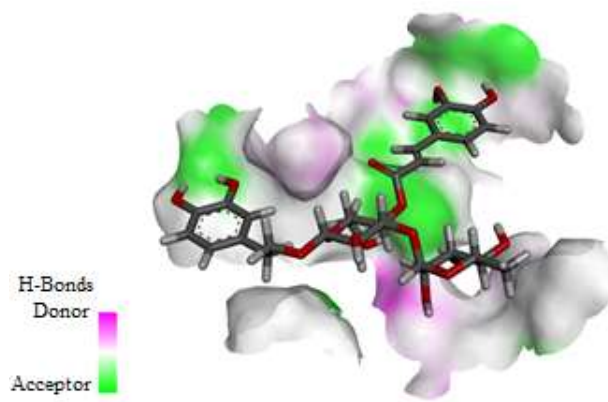


Figure 54: 6BDZ-SANC00370

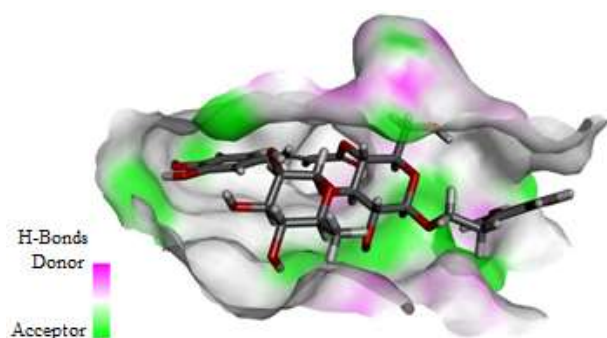


Figure 55: 1J1C_B-SANC00370

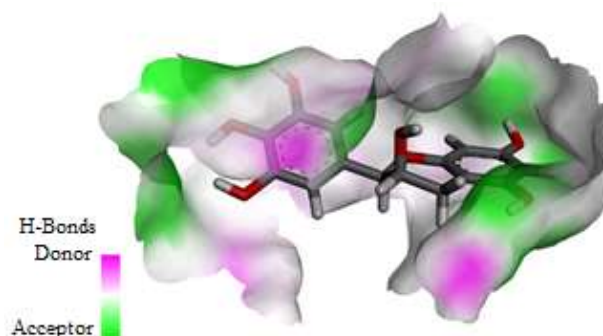


Figure 56: 4CTH-SANC00103

3.1.3.5 Hydrophobicity

A final detail in ligand binding, not unrelated to the aromatic interactions is hydrophobicity. Hydrophobicity (or lipophilicity) is the tendency of non-polar molecules to form aggregates in order to reduce their surface of contact with polar molecules such as water. Hydrophobic interactions also drive a variety of physical and biological phenomena[100]. Most of these protein-ligand complexes involve regions of the protein that have low hydrophobicity. Figure 61 depicts a small region of stronger hydrophobic interaction with one of the benzene rings within the ligand making a face edge interaction with the protein (compare with figures 44 and 49). Similarly, Figure 58 and 62 have a face-edge interaction between respective protein-ligand complexes but the intensity of the interaction is lower.

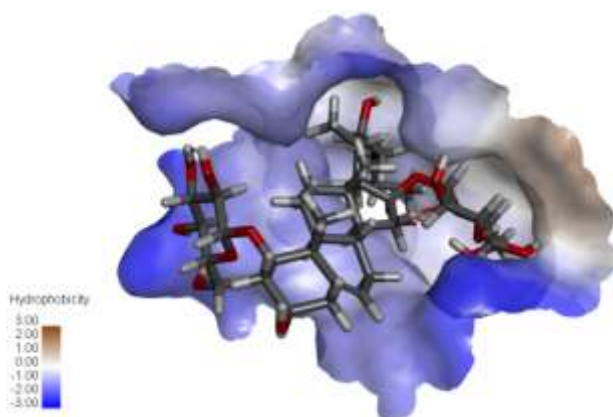


Figure 57: 1FKN_A- SANC00664

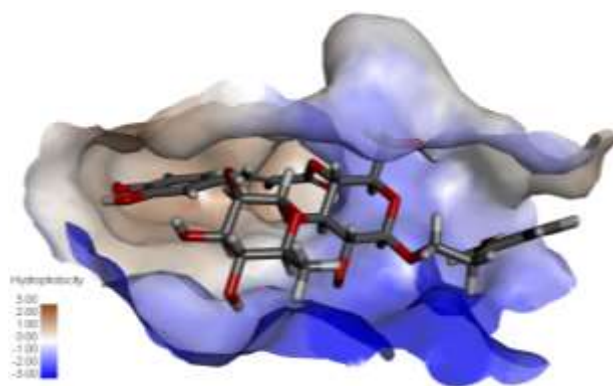


Figure 58: 1J1C_B-SANC00370

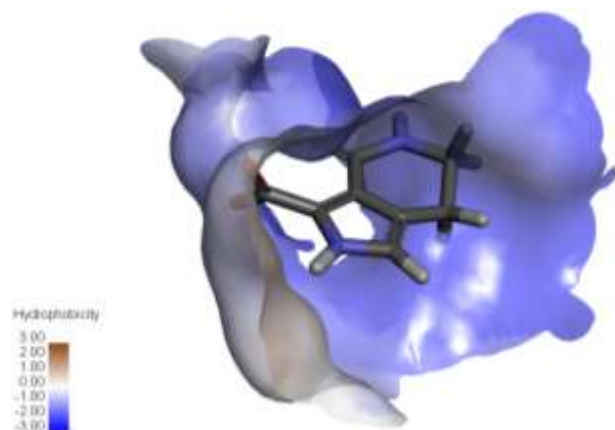


Figure 59: 4R12-SANC00256

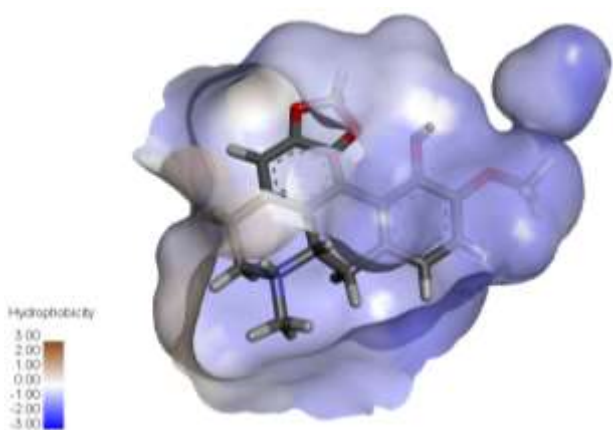


Figure 60: 4EY7_B-SANC00374

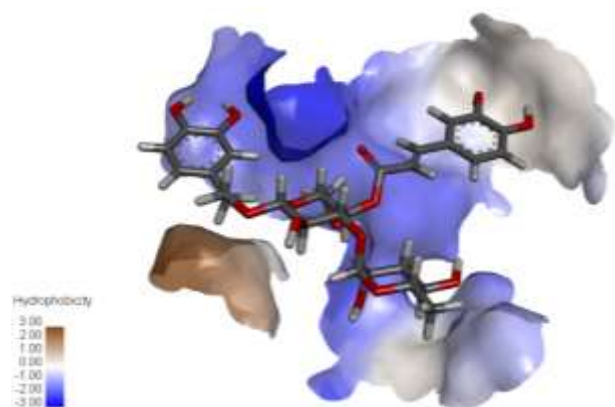


Figure 61: 6BDZ-SANC00370

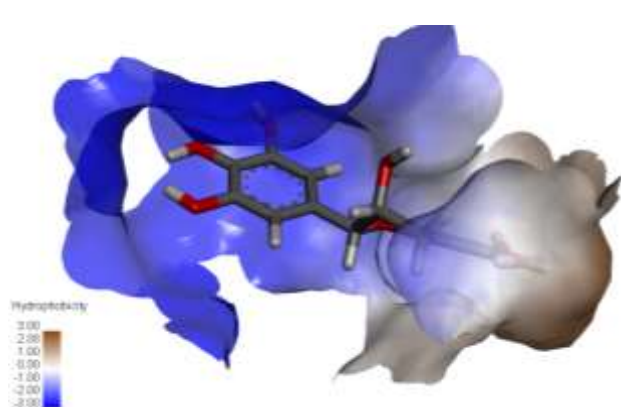


Figure 62: 4CTH-SANC00103

3.1.3.6 Conclusion

SANC00256 is the smallest of all of the ligands explored; however, it does make two hydrogen bonding interactions to 4R12 (Maestro analysis) and there are individual pi-pi (T-shaped) and pi cation interactions (DSV analysis). These interactions may be considered dense when the size of the ligand is taken into consideration (interactions per total ligand atoms or interactions per ligand mass). This ligand is small enough to inhabit a very small pocket which has largely a very low

SASA. Given the aromatic nature of the 4R12 pocket this appears to be an ideal ligand for this particular binding site.

SANC00370 may have issues with selectivity, and although it may be felt that this non-selective binding may be driven by many hydrogen bond donor and acceptor groups present in the ligand (there are 8 hydrogen bonds to 6BDZ), the binding to 1J1C_B is also driven by pi-alkyl interactions (with only 4 hydrogen bonds), and in this case one portion of the ligand is occupying a hydrophobic pocket. In terms of physicochemical characteristics, SANC00370 has too many hydrogen bond acceptors; this implies that SANC00370 requires modification of these properties to be acceptable as a drug like molecule, but also therefore that binding to 1J1C has a greater potential since this is based more on the conformation and shape of SANC00370, and less on the formation of many hydrogen bonds.

SANC00103 also was the top binding ligand to more than one target, but as a ligand it has fewer hydrogen bond donors (6) and acceptors (1) than SANC00370. In its binding to 4CTH there are four hydrogen bonds (again this is a large number for the relatively small size of SANC00103). This ligand has two aromatic systems, and this appears to be highly relevant to its binding to 4CTH. There is a complex combination of hydrophobic and hydrogen bonding interactions between this ligand and 4CTH.

SANC00374 is a relatively rigid planar molecule, and aromatic interactions are not essential in its interaction with 4EY7_B. There are few hydrogen bonds in this interaction and hydrophobicity appears to be relevant. It is interesting that this ligand fits within a narrow site where much of the receptor surface is relatively solvent inaccessible, and it appears that the shape of the ligand is very well matched for its interaction with 4EY7_B.

Finally, SANC00664 is a large flexible system and there appear to be portions of this ligand that lie outside of the active site pocket binding to the surface of 1FKN_A. SANC00664 has no aromatic rings at all, although it does get involved in some hydrophobic pi-alkyl interactions to histidine and tyrosine residues. DSV does identify some unfavorable van der Waals interactions.

3.2 Desmond: Molecular Dynamics Results (100 ns)

The simulations for all these complexes was performed for 100 ns. For SANC00370, molecular dynamics was only pursued in the case of 1J1C_B (and not for 6BDZ). As such, presented below are the results of molecular dynamics simulations of five receptor-ligand complexes, namely 1FKN_A (SANC00664), 1J1C_B (SANC00370), 4EY7_B (SANC00374), 4CTH (SANC00103) and 4R12 (SANC00256). All of the molecular dynamics simulations and analysis (and diagrams presented) were produced using the Schrödinger Desmond program.

3.2.1 RMSD calculations during dynamics

The Root Mean Square Deviation (RMSD) metric measures the average displacement of a selection of atoms for a particular frame of molecular dynamics with respect to a reference frame. It is calculated for all frames in the trajectory. The RMSD for a single frame x is calculated using the following formula,

$$\text{RMSD}_x = \sqrt{\frac{1}{N} \sum_{i=1}^N \left((r'_i(t_x)) - r_i(t_{ref}) \right)^2}$$

where N is the number of atoms in the simulation under consideration, t_{ref} is the reference time, (typically the first frame is used as the reference and it is regarded as being at time $t=0$) and r' is the position of the selected atoms in frame x after superimposing on the reference frame, where frame x is recorded at time t_x . The procedure is repeated to calculate an RMSD value for every frame in the simulation trajectory.

In terms of protein analysis RMSD, Protein RMSD plots show the RMSD evolution of a protein (in Figures 63-67 in the next section, the Protein RMSD is indicated on the left Y-axis according to time on the X-axis). For this calculation, all protein frames are first aligned on the reference frame backbone, and then the RMSD is calculated based on the protein atom selection (in this case C α atoms). Monitoring the RMSD of the protein can provide insights into its structural conformation throughout the simulation. In the first case, RMSD analysis can indicate if the simulation has equilibrated — if it has equilibrated its fluctuations towards the end of the

simulation will be around some thermal average structure. It is important that simulations are observed to converge as observed by monitoring this fluctuation. Increasing or decreasing trends in RMSD indicate that equilibration has not been achieved, and longer dynamics may be required. Changes in RMSD of the order of 1-3 Å are regarded as acceptable for small, globular proteins. RMSD fluctuations greater than 3 Å, however, indicate that the protein is undergoing a large conformational change during the simulation.

In terms of the long-term stability of the ligand within the context of the protein during dynamics, the Ligand RMSD provides a suitable metric for analysis (in the next section, Figures 63-67 show the evolution through time of Ligand RMSD using right Y-axis). In the Ligand RMSD plots used for our analysis, 'Lig fit Prot' indicates that the reference for the RMSD calculation is based on alignment of the protein-ligand complex to the protein backbone of the reference. After this alignment the RMSD of the ligand heavy atoms is measured. As a result of this, if the Ligand RMSD values observed are significantly larger than the RMSD of the protein, then it is likely that the ligand has diffused away from its initial binding site.

3.2.1.1 Protein and Ligand RMSD

The protein RMSDs in Figures 63-67 have fluctuations within the required 1-3 Å range which is acceptable for these proteins and confirms that they are stable during the simulation. At the start of the simulations there are large fluctuations in RMSD associated with equilibration, however, in all cases toward the end of the simulations the RMSD values stabilize around a fixed value, indicating that all systems are fully equilibrated.

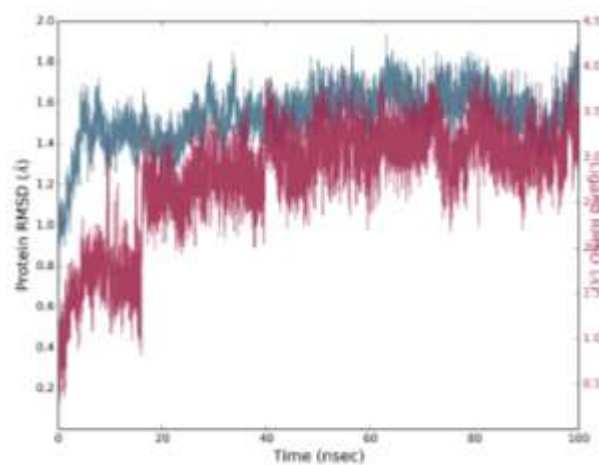


Figure 63: 4CTH-SANC00103

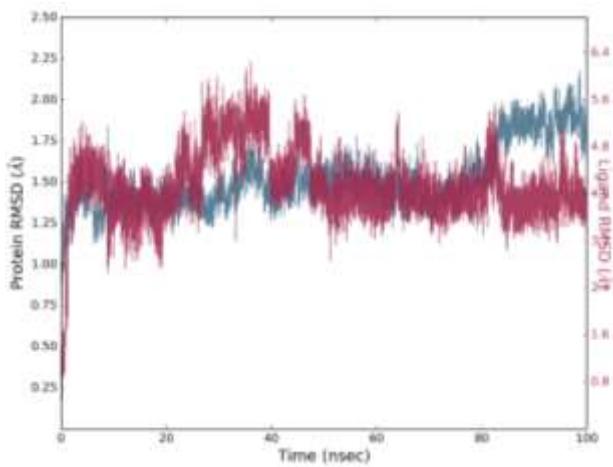


Figure 64: 4EY7_B-SANC00374

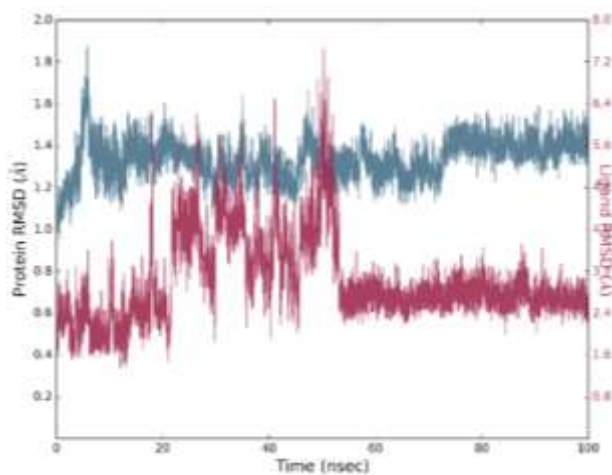


Figure 65: 1FKN_A-SANC00664

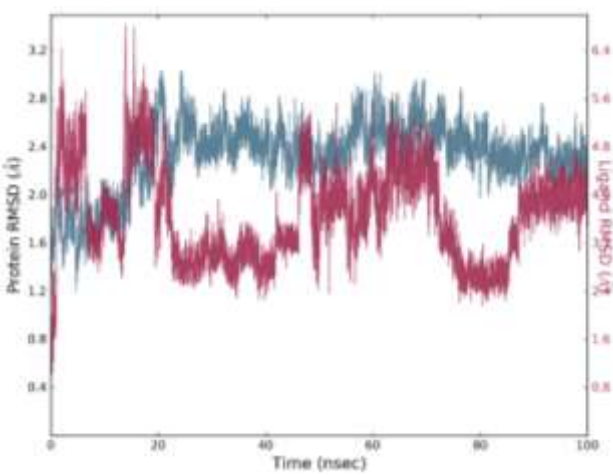


Figure 66: 1J1C_B-SANC00370

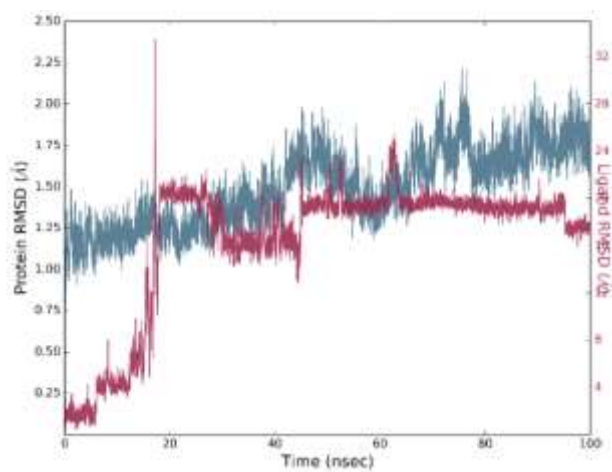


Figure 67:4R12-SANC00256

On the other hand, the ligand RMSDs displayed in Figures 63-67 vary in terms of their fluctuations compared to the protein RMSD in the respective complexes. For SANC00103 (Figure 63) and SANC00374 (Figure 64), these are simulations where the ligand largely maintains its position (is stable) within the active site, although from careful observation of Figure 64 it is apparent that SANC00374 does shift somewhat prior to stabilization within the binding site. Visualization of SANC00374 during the simulation depicts that at no point does it leave the binding pocket, there is not leeway in terms of space for it to move around as it is deep inside the binding site of 4EY7. In the other receptor-ligand complexes, the ligand RMSD at some point makes fluctuations indicating reorganization of the ligand during dynamics.

In Figures 65 and 66 SANC00664 and SANC00370 encounter a very large ligand RMSD early in the simulation, but the RMSD then stabilizes. Visualizations show that the ligands do not leave the binding site during the simulations; however, the ligands go through conformational changes due to rotation of some bonds within the respective ligands. It is this rotation of bonds during the simulations that causes the observed increase in ligand RMSD. On the other hand, in Figure 67 for SANC00256 depicts similar trends of the ligand RMSD; however, the visualizations of the trajectory show this is for a very different reason. At about 17 ns the ligand leaves the binding site, returning 1 ns later, thereafter, staying in the binding site until the simulation ends.

3.2.2 RMSF calculations during dynamics

The Root Mean Square Fluctuation (RMSF) is useful for characterizing local changes along the protein chain. The RMSF for a particular residue i is obtained by superposition of the chain on the reference, and then applying the following formula.

$$\text{RMSF}_i = \sqrt{\frac{1}{T} \sum_{t=1}^T \langle \left(r_i'(t) - r_i(t_{ref}) \right)^2 \rangle}$$

where T is the trajectory time over which the RMSF is calculated, t_{ref} is the reference time, r_i is the position of residue i and r' is the position of atoms in residue i . The average of the square of this distance is taken over the selection of atoms in the residue.

The RMSF plot (for proteins) draws attention to regions of the protein that are more variable in position during the simulation. A common observation is with respect to the termini (N- and C-termini) which are naturally expected to be less fixed than the rest of the protein, where the RMSF is generally large. Regions of defined secondary structure (including α -helices or β -strands) are expected to be rigid with a low RMSF, while unstructured regions (flexible loops) are expected to have a large RMSF. In the RMSF graphs in the next section, ligand contacts are included in this plot and they show the protein residues that interact with the ligand as being marked with green-coloured vertical bars.

The Ligand Root Mean Square Fluctuation (L-RMSF) is useful for characterizing changes in the ligand atom positions and is calculated in the same manner. The protein-ligand complex is first aligned to the reference frame protein backbone, following which the ligand RMSF is calculated for the ligand heavy atoms.

The Ligand RMSF plot shows the ligand's fluctuations broken down by atom. Since there are fewer atoms in a ligand than in the protein chain, analysis is by atom rather than by residue. In the next section, the ligand atom numbers for the RMSF plot are provided in a 2D structure preceding the RMSF plot. This ligand RMSF provides detail of what parts of the ligand are associated with reorganization within the active site, where this occurs, and also may give you insights on how different chemical moieties from the ligand interact within the active site.

3.2.2.1 Protein RMSF

The plot on Figure 68 for 4CTH interestingly presents three regions in the middle of the protein that during simulation display distinct fluctuations of above 2 Å (three small flexible loops centered on PHE431, LEU457 and LYS528 that all connect α -helix regions). 4CTH also interestingly has low levels of fluctuations in the N- and C-terminal region during this simulation. Bear in mind that the ligand contacts to the protein's residues are not far spatially from the N- and

C-terminals. Figure 69 for 4EY7 shows the expected high RMSF for the N- and C-terminals during this simulation, particularly the RMSF for the C-terminal ALA542 = 5.29Å. There are two other residues with very high RMSF towards the C-terminal of 4EY7, associated with flexible loop regions and they are THR262 = 5.21Å and LYS496 = 3.98Å. None of these fluctuations are associated with the ligand binding residues.

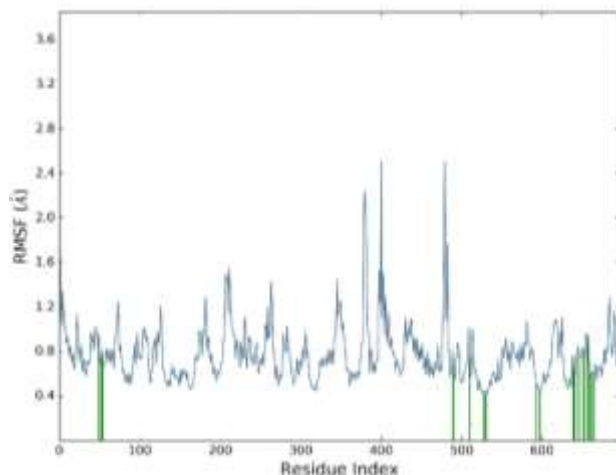


Figure 68: 4CTH (SANC00103)

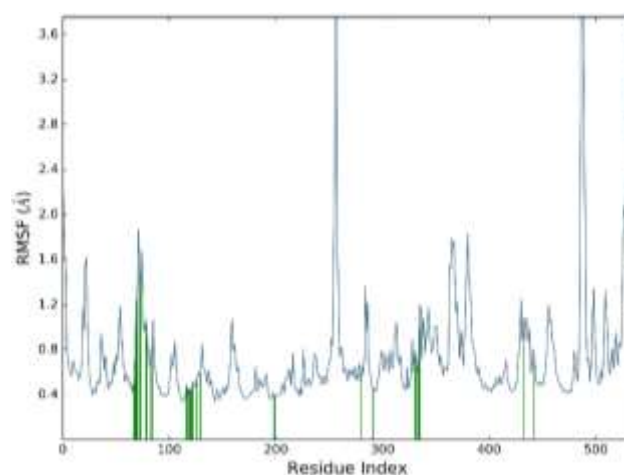


Figure 69: 4EY7_B (SANC00374)

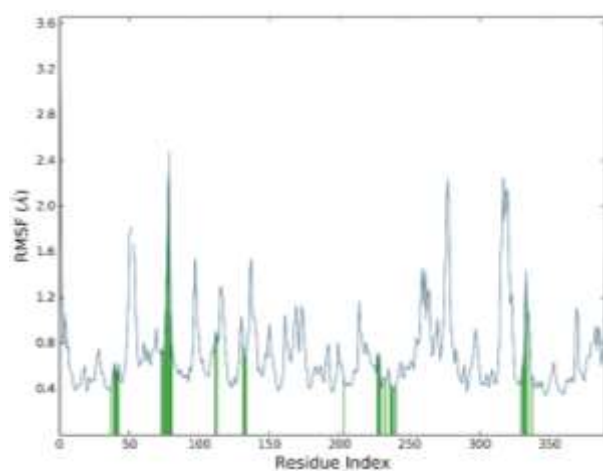


Figure 70: 1FKN_A (SANC00664)

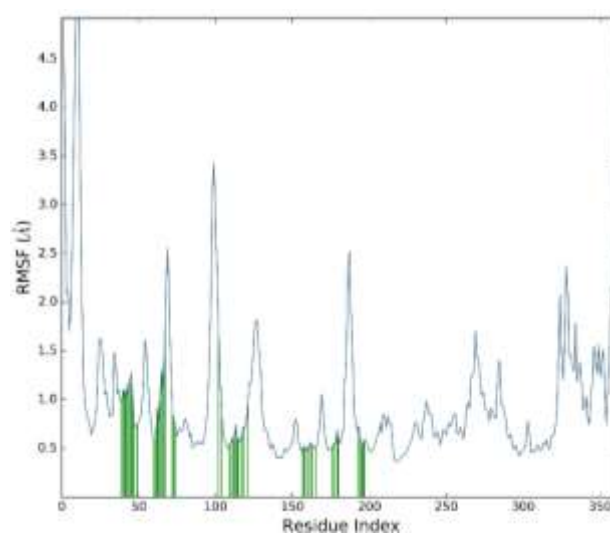


Figure 71: 1J1C_B (SANC00370)

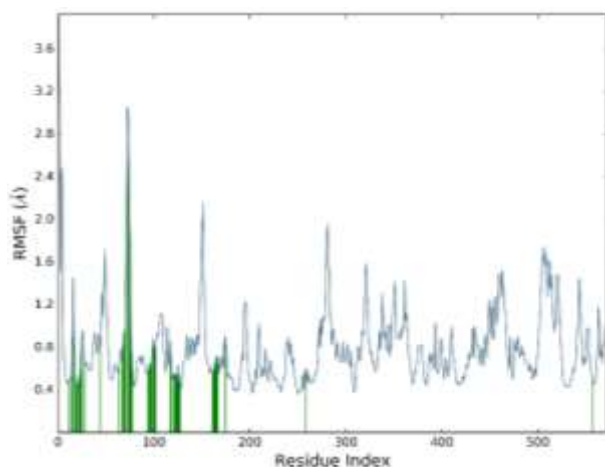


Figure 72:4R12 (SANC00256)

During the molecular dynamics simulation with 1FKN_A (Figure 70) there is a broad region of mild fluctuation towards the middle of the protein. LEU119 (indexed 75) is part of a β -sheet which is influenced indirectly by the SANC00664 ligand and displays an RMSF of 2.5\AA . In Figure 71 (1J1C_B) again there are large fluctuations at both the N- and C-termini. Of note is the large fluctuation of VAL135 in 1J1C, which is in a loop region in close proximity to the active site. By contrast, in Figure 72 the ligand contacts HIS104 which presents the greatest RMSF variation of all residues. This HIS104 is part of the binding groove for the ligand but is on a flexible loop of the protein. The fluctuations for the remainder of the protein are mostly within 2.0\AA except for this residue with an RMSF value of about 3.0\AA .

3.2.2.1 Ligand RMSF

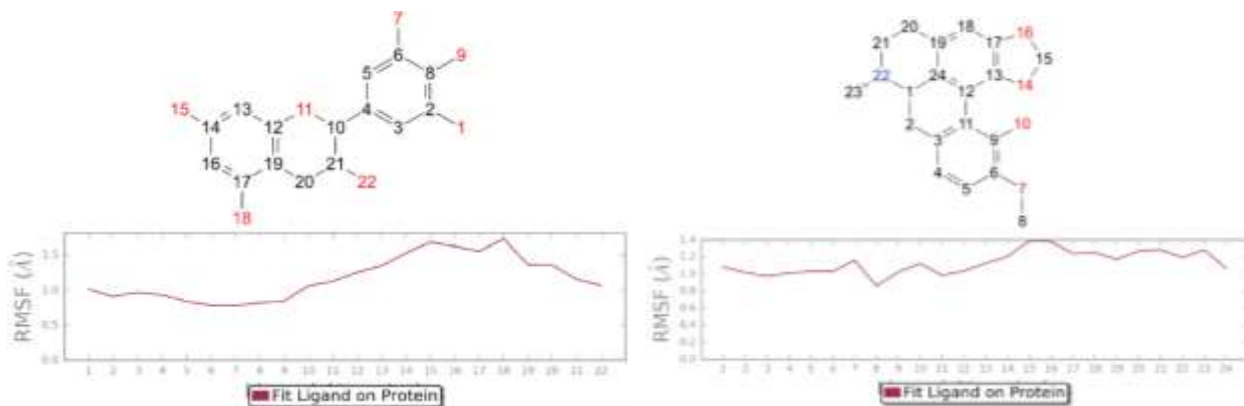


Figure 73: 4CTH (SANC00103)

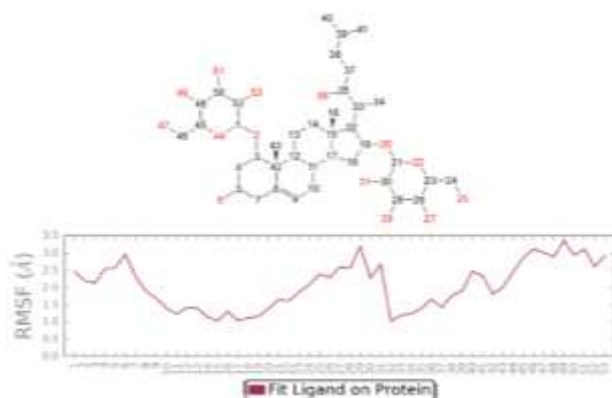


Figure 74: 4EY7_B (SANC00374)

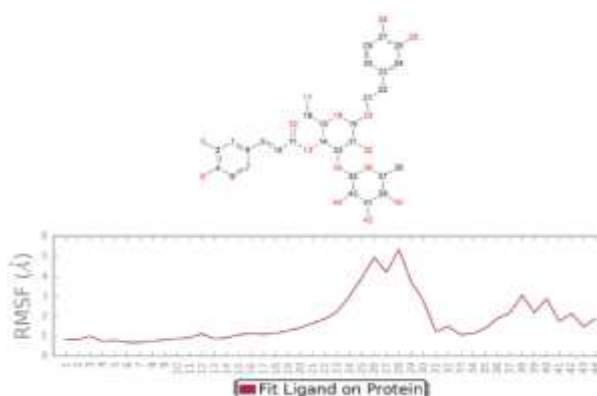


Figure 75: 1FKN_A (SANC00664)

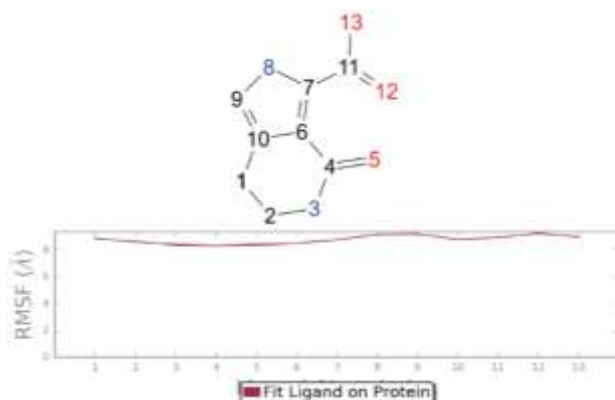


Figure 76: 1J1C_B (SANC00370)

Figure 77: 4R12 (SANC00256)

In Figure 73, for SANC00103, the atoms numbered 1-10 remain stably deep within the 4CTH binding pocket during the simulation, hence the low fluctuations. On the other hand, atoms in the region of 15, 17 and 18 are somewhat solvent exposed and can rotate and interact with various waters in the system. The ligand in Figure 74, SANC00374, is generally solvent exposed with a more consistent RMSF across the molecule, with few significant differences in fluctuation per atom. The ligand is planar and has few rotatable bonds, and even atoms such as 8 and 10 which are rotatable do not exhibit an increased relative RMSF. However, the edge defined by atoms 14, 15 and 16 shows a relatively increased RMSF indicating a rocking of this whole rigid molecule,

where the pivot for this rocking is far from these atoms. Figure 75 illustrates the flexibility of SANC00664. For this ligand several large RMSF fluctuations are observed as the ligand has space to rotate (as it extends beyond the binding site and interacts with water in the system). It is interesting that the “core” of SANC00664 maintains a lower RMSF while the substituent sugars present the largest RMSF values. In Figure 76 SANC00370 atoms 22-44 have a wider range of motion, while atoms 1-21 are maintained deep inside the binding pocket with low observed values of RMSF during simulation. SANC00256 is a small rigid ligand, there is only one bond that can rotate which is between atom 7 and 11 (the rotation is for a carboxylate group). In Figure 77 SANC00256 shows that RMSF values are maintained for the whole ligand (showing the ligand rigidity) but the high overall RMSF (9 Å) correlates with the ligand RMSD (Figure 67), where the ligand leaves the active site at the start of dynamics before settling back into the protein.

3.2.3 Protein SSE during dynamics

Protein secondary structure elements (SSE) like alpha-helices (red) and beta-strands (blue) were monitored throughout these simulations (Figures 78-87). Secondary structure elements are particularly relevant in the Alzheimer’s disease context. For example, although the β -Amyloid ($A\beta_{1-40/42}$) protein is not explored in this study, its precursor the Amyloid Precursor Protein (APP) contains helical structures, but the $A\beta_{1-40/42}$ protein stacks as β -sheet-rich oligomeric structures linked with pathological states. The $A\beta_{1-40/42}$ protein’s oligomeric β -sheet-rich form is implicated as a cause of Alzheimer’s Disease. [101].

3.2.3.1 Protein SSE Histograms

Figures 78-82 report SSE distribution by residue index throughout the simulation as a histogram. Figure 78, for instance illustrates that 4CTH comprises mostly alpha helices (red). Since the 4CTH protein has more α -helices than β -strands, we may infer some expectations about the protein during dynamics, given that α -helices are more stable than β -strands due to hydrogen bonding formed by the amide NH groups and the carbonyl oxygen atoms C=O. By contrast, Figure 80 shows that 1FKN contains largely beta-strands during simulation. The SSE of the other proteins in Figures 79, 81 and 82 is distributed between alpha-helices and beta-strands – the distribution in 1J1C

(Figure 81) is such that the first 100 residues comprise mainly β -strands while the remainder of the residues predominantly form α -helices.

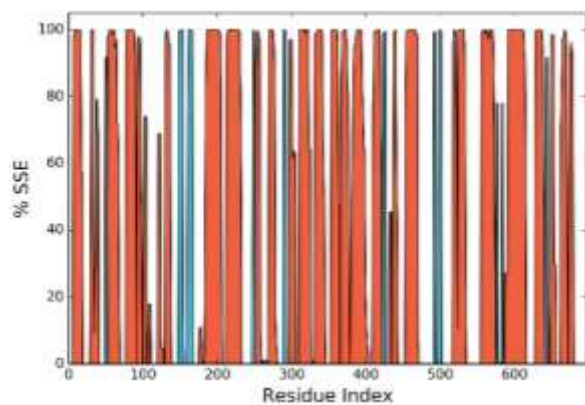


Figure 78: 4CTH -SANC00103

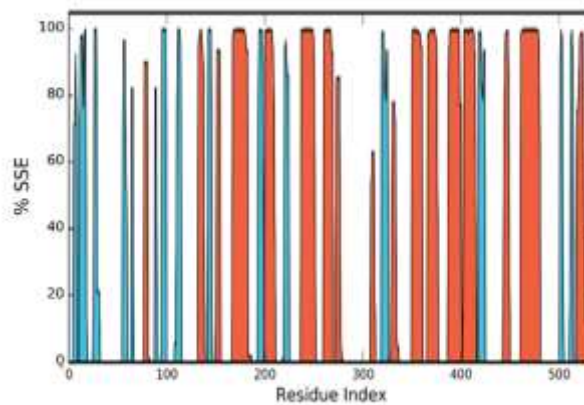


Figure 79: 4EY7_B-SANC00374

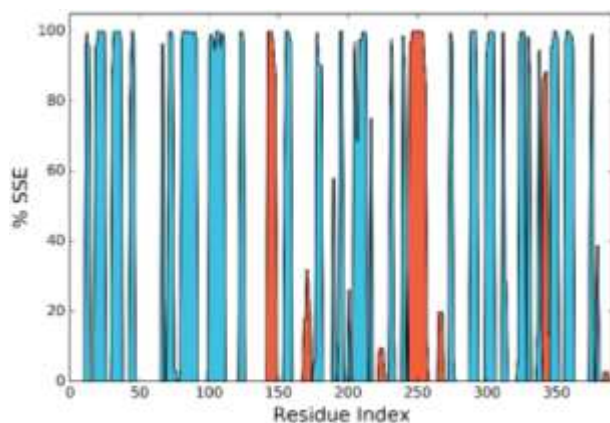


Figure 80: 1FKN_A-SANC00664

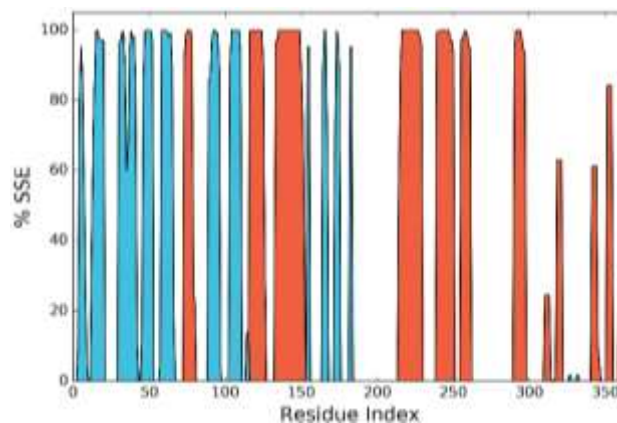


Figure 81: 1J1C_B-SANC00370

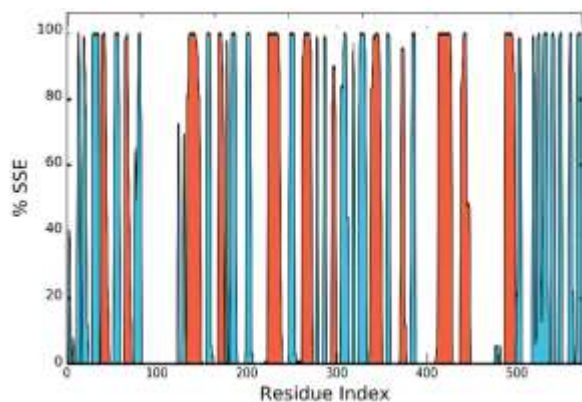


Figure 82: 4R12-SANC00256

Most of the residues of the proteins have an SSE distribution during simulation of up to 100% showing that secondary structure is maintained during molecular dynamics in the presence of ligands.

3.2.3.2 Protein SSE Timeline

Although the plots in Figures 78-82 are promising with respect to secondary structure, they do not detail points of variation during dynamics where secondary structure may vary. The plots in Figures 83-87 summarize the SSE composition for each trajectory frame during the course of the simulation, in terms of both total % SSE per time step (top), and in terms of SSE monitoring and assignment (on a per residue basis) over time.

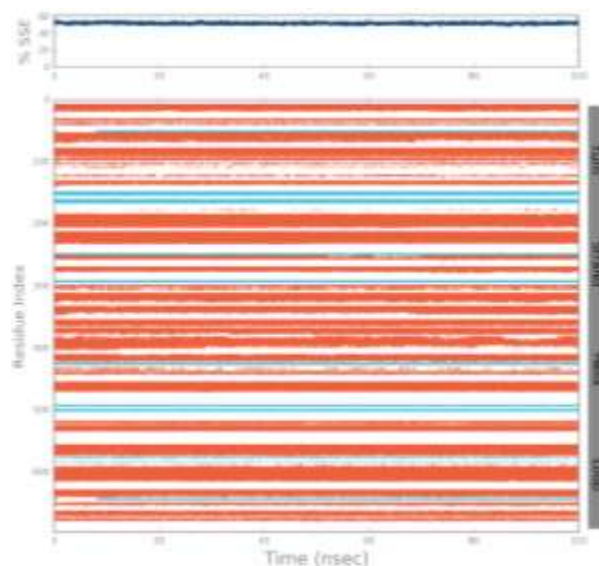


Figure 83: 4CTH-SANC00103

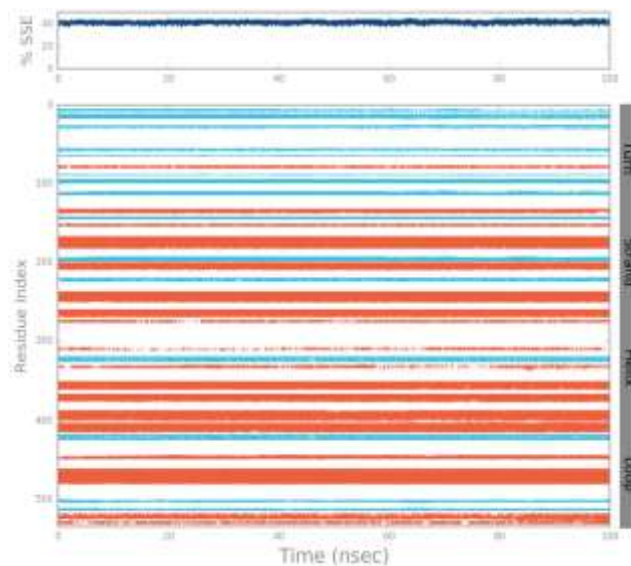


Figure 84: 4EY7_B-SANC00374

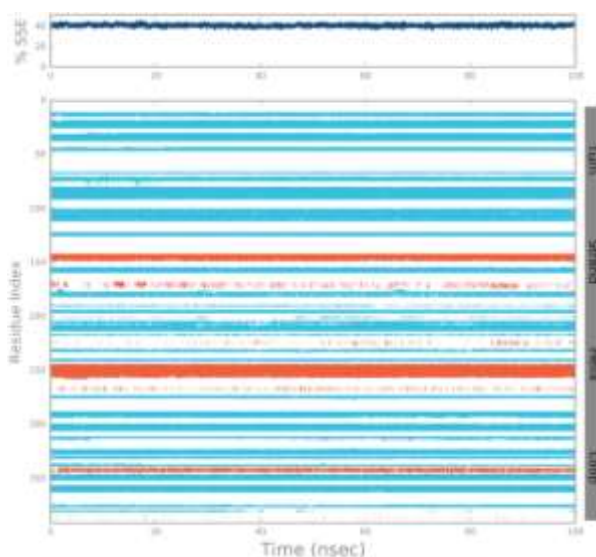


Figure 85: 1FKN_A-SANC00664

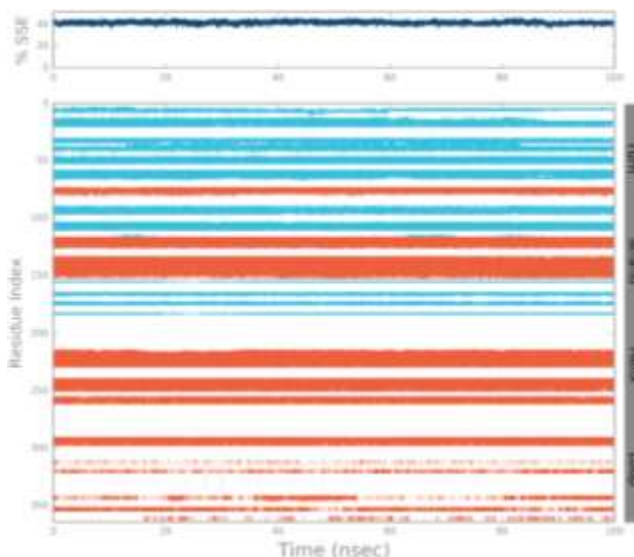


Figure 86: 1J1C_B-SANC00370

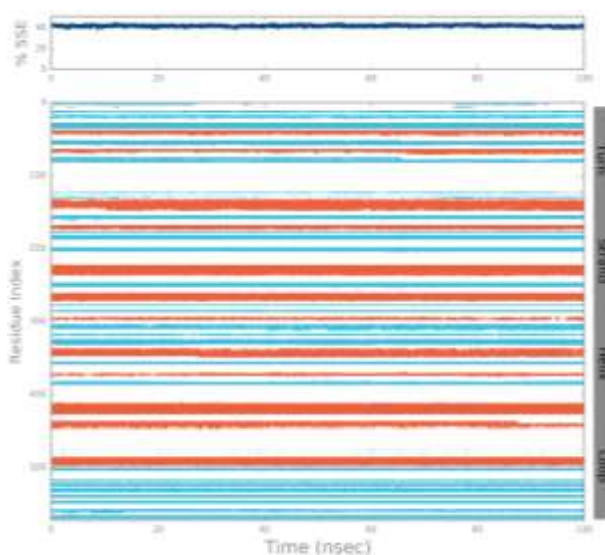


Figure 87: 4R12-SANC00256

The 4CTH protein maintains 55% SSE throughout the simulation, all other proteins maintain 40% SSE composition through the 100 ns Desmond simulations. In terms of detail, all proteins secondary structure elements for each of the proteins appear constant during the entire MD simulations. In terms of the SANCDB ligands properties, this suggests that the ligands chosen simply block and bind the active site, as is their intended purpose, during dynamics.

3.2.4 Protein-Ligand Contacts

It was of interest to monitor specific interactions of the ligands during the simulation. Available in analysis were four categories of ligand-protein interactions, namely Hydrogen Bonding, Hydrophobic, Ionic and Water Bridge interactions. Within the analysis subtypes of interactions were available for more detail on the behavior of the ligand with respect to the protein. All interactions were normalized over the full trajectory (such that maintaining an interaction for all the trajectory would result in a value of 1.0, interaction for half the trajectory would be 0.5 and so on). It was possible for ligands to interact greater than the normalized 1.0, particularly where ligands interacted through many contacts with a single residue.

3.2.4.1 Protein-Ligand Contacts Bar Graphs

In order to generate the protein-ligand contacts bar graphs, strict criteria were used in the evaluation of interactions.

For hydrogen bonds, the distance cut-off used was 2.5 Å between the donor H and the acceptor atom A (in the hydrogen bond D—H···A). Two angles requirements were also to be met for the identification of a hydrogen bond; firstly, the angle donor-hydrogen-acceptor (D—H···A) was required to be $\geq 120^\circ$, and secondly the angle to the acceptor anchor X on the molecule (H···A—X) was required to be $\geq 90^\circ$.

In terms of analysis of hydrophobic contacts, separate criteria were used for the three subtypes of hydrophobic interactions. For π -Cation interactions, the cutoff for the distance between the charged group and the aromatic system was 4.5 Å and π - π interactions the focus was on face to face and edge to face geometries. All other hydrophobic interactions were classed into a third category as non-specific interactions, where the criterion was a hydrophobic side chain of the protein being within 3.6 Å of aromatic or aliphatic carbon atoms of a ligand

For ionic interactions, the cut-off was 3.7 Å for the distance between two atoms of opposite charge.

The identification of hydrogen bonding mediated by Water Bridges followed less strict criteria than for direct protein ligand hydrogen bonding. For a Water Bridge, the criteria were a cutoff of 2.7 Å between H and A in the hydrogen bond D—H···A, with the associated angle being $\geq 110^\circ$, and a requirement for the H···A—X angle (as described above) to be $\geq 80^\circ$.

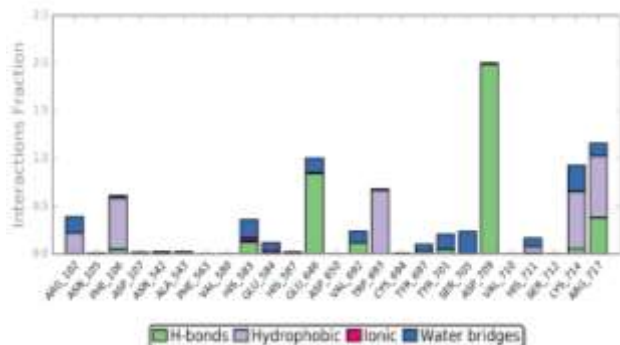


Figure 88: 4CTH-SANC00103

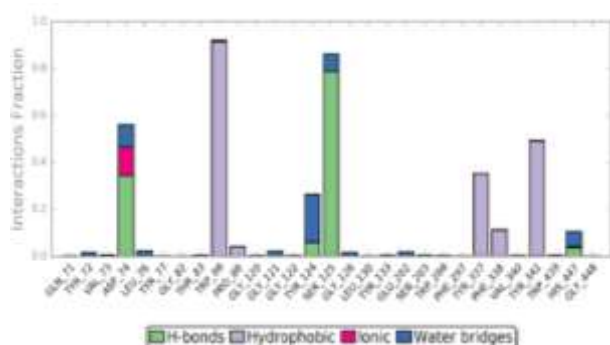
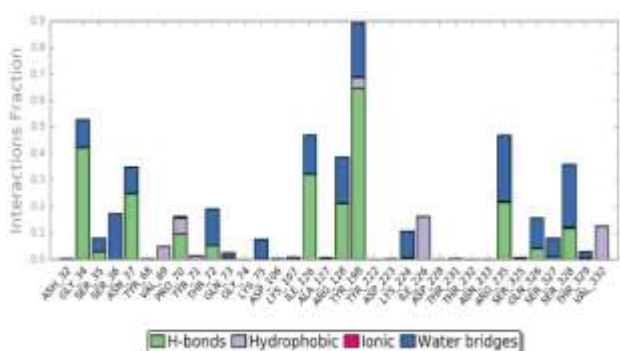


Figure 89: 4EY7_B-SANC00374



Figures 88-92 show normalized contact bar graphs for these interactions for each of the protein ligand contacts for the duration of the molecular dynamics simulation.

Figure 88 depicts that ASP_709 of 4CTH continuously interacts with SANC00103 during the course of the 100 ns simulation. The value of 2.0 (200%) for this interaction indicates that there are two hydrogen bond contacts continuously for the duration of the molecular dynamics (since the value per interaction is normalized). The reason for this is that two ligand hydroxyl groups hydrogen bond separately but continuously to both oxygen atoms of the ASP_709 carboxylate group. GLU_646 of 4CTH also provides a predominant interaction with the ligand SANC00103 as is observed by the 0.9 or 90% value for the hydrogen bonding contact. There are four residues providing a value of 0.5 for the hydrophobic contact – meaning that each of these interactions are present for half of the simulation. There are also several water bridging interactions (water mediated interaction between residues on the protein and the ligand) but it is likely that these are short-lived on the timespan of the dynamics.

Protein 4EY7_B (Fig. 89) interacts with the ligand SANC00374 mainly through hydrophobic interactions, with the residue making the most contact with SANC00374 being TRP_86 with an almost constant interaction (0.9) during simulation. SER_125 of 4EY7 provides a dominant hydrogen bonding interaction that is maintained for a 0.8 fraction (or 80%) of the entire simulation. The protein-ligand complex has ionic interactions with the carboxylate anion of residue ASP_74, due to the presence of the positively charged tertiary ammonium center in ligand SANC00374.

The ligands from the receptor-ligand complexes in Figures 90 (1FKN - SANC00664) and 91 (1J1C - SANC00370) are large with many hydroxyl groups, and so, as expected several residues show hydrogen bonding and water bridging interactions to the respective ligands during dynamics.

For the 4R12 - SANC00256 complex (Figure 92), there are three residues dominate the protein ligand interactions (these interactions are observed for 30% or more of the total course of simulation). The LYS_44 of 4R12 interacts with SANC00256 for 64% of the simulation through a hydrogen bonding interaction.

3.2.4.2 Protein-Ligand Contacts Timeline

In the analysis of the protein ligand interactions through the Protein-Ligand Contact Bar Graph, where contacts are not continuous, it is not clear as to whether interactions occur during reorganization of the ligands (where this occurs), or after the ligand has reached an equilibrium position. To this end, timelines of the interactions/contacts (**H-bonds**, **Hydrophobic**, **Ionic**, **Water bridges**) during the dynamics were generated (Figures 93-97). In Figures 93-97, two panels are visible using the same time variable horizontal coordinate. The upper panel simply counts the total number of protein-ligand contacts at each frame of dynamics, while the lower panel details the specific residues that is interacting with the ligand for each frame of the trajectory. Some residues in a single frame interact through multiple interactions, and this is represented by a darker shade of orange, according to the scale to the right of each plot.

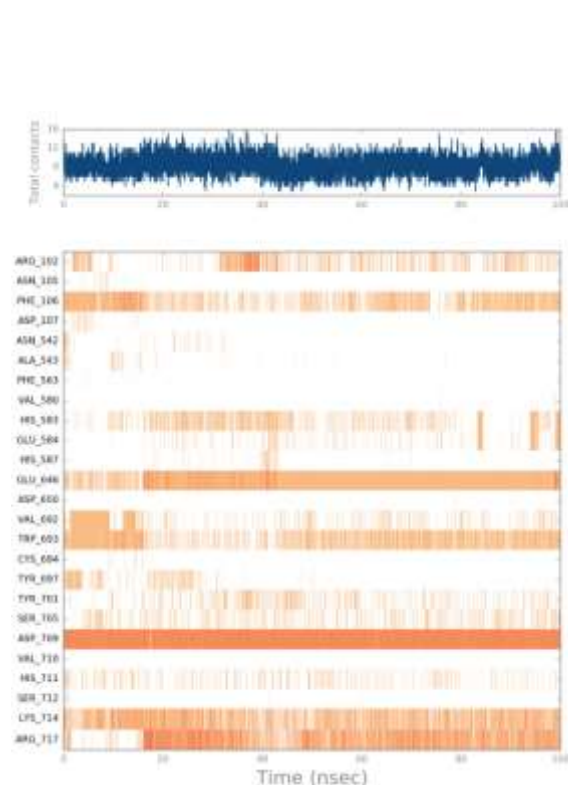


Figure 93: 4CTH-SANC00103

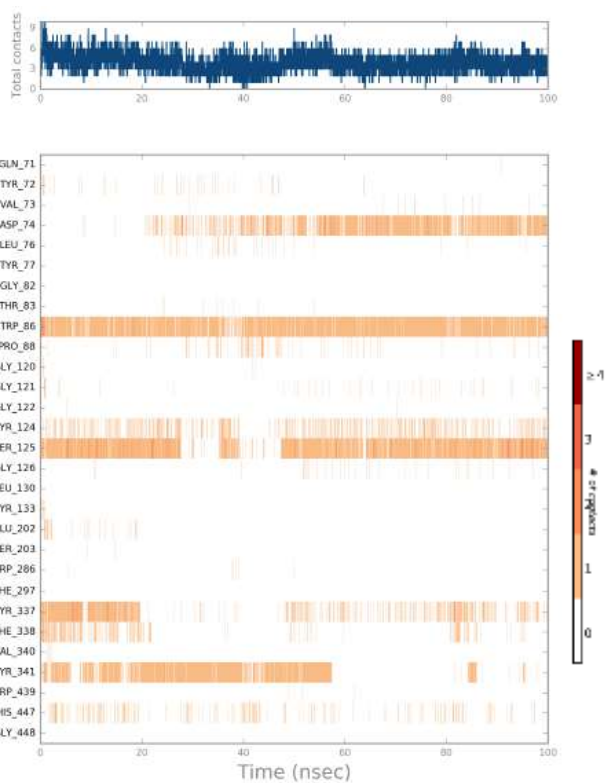


Figure 94: 4EY7_B-SANC00374

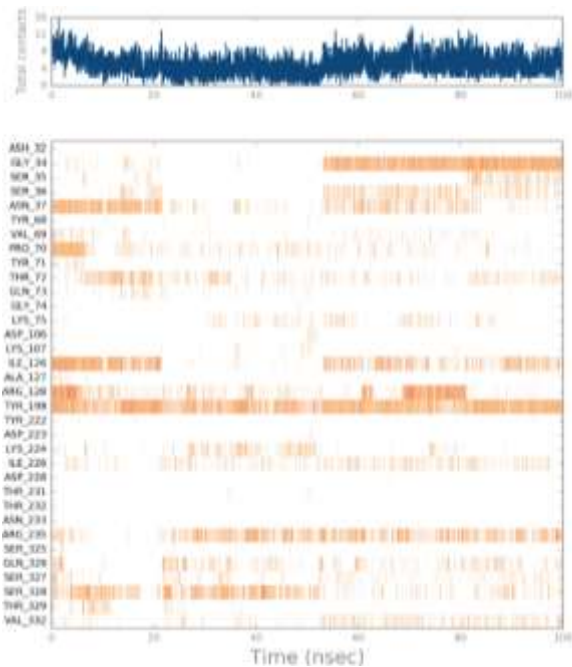


Figure 95: 1FKN_A-SANC00664

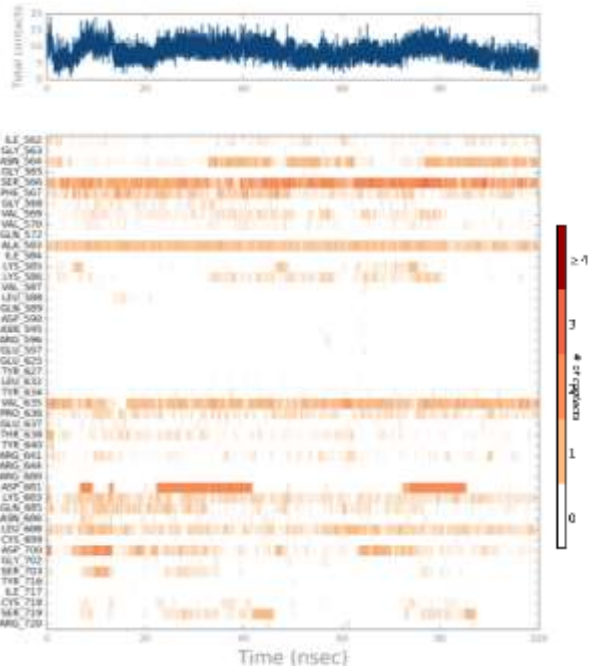


Figure 96: 1J1C_B-SANC00370

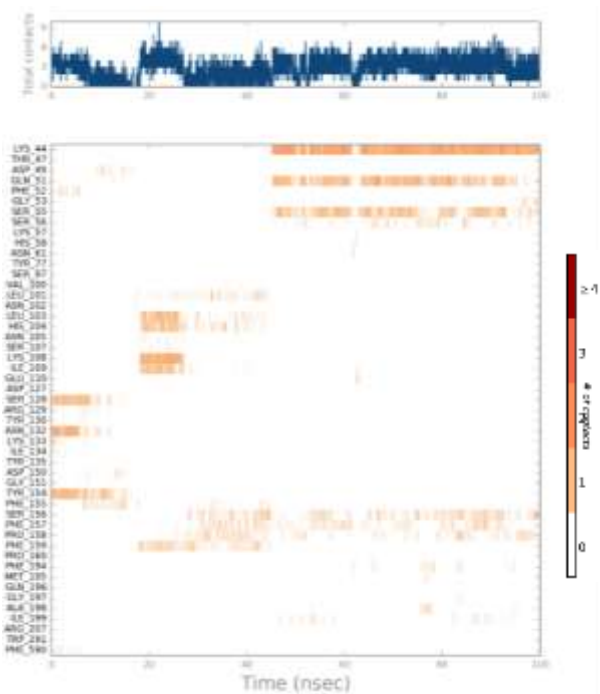


Figure 97: 4R12-SANC00256

In Figure 93 ASP_709 of 4CTH makes the most contacts with SANC00103 throughout the simulation (in agreement with Figure 88 which shows the contact number of 2.0) Following this is GLU_646; now it is apparent that this hydrogen bond has not formed during the first 18 ns of the simulation, however, after this point the bond is formed and is stable for the remainder of the dynamics. However, as also suggested in Figure 88 GLU_646 only makes a single contact relative to the two hydrogen bonds from the ligand to ASP_709, and this is reflected in the darkness of the orange colour in Figure 93. LYS_714 and ARG_717 also both contribute significant contacts with the ligand but at a lower intensity; these contacts made are mostly hydrophobic bonds. During the 100 ns MD simulation, there are many identified contacts between 4CTH and SANC00103; at several frames this number reaches 16 identified contacts.

Figure 94 shows that TRP_86 of 4EY7_B is a key residue in terms of interaction with the SANC00374 ligand for the entire simulation; this interaction is entirely hydrophobic (compare Figure 89). Residue SER_125 also makes significant contacts with the ligand but during the simulation, between the 30 and 50 ns marks, this hydrogen bonding interaction is broken; later in the simulation this interaction reappears. It is interesting that water bridging is also observed between SER_125 and SANC00374. This would suggest that when the hydrogen bond breaks, the ligand still has opportunity to interact with SER_125 through the formation of water bridges. There appear to be fewer residues in this case from the 4EY7_B protein making significant contacts with the ligand SANC00374 (compared to the previous 4CTH SANC00103). However, there appears to be a firm progression during simulation away from the hydrophobic interactions with tyrosine 337 and 341 (at the start of simulation) towards the very highly favoured ionic/hydrogen bonding between ASP_74 and the tertiary ammonium center of SANC00374.

In Figure 95 1FKN-SANC00664 it is now evident that the ARG_128 hydrogen bonding/water contact interaction is very strong, but periodic through the simulation. Further to the ASN_37 and ILE_126 are strong hydrogen bonding contacts at the start of simulation that have been lost or significantly weakened by the end of simulation. The only constant interaction with the ligand is through TYR_198 which maintains a hydrogen bond (or for a small part of the simulation a water bridge) with the only aliphatic alcohol of SANC00664. This complex makes up to 16 simultaneous

contacts during simulation, but towards the end of simulation this fluctuates around 5 simultaneous contacts.

With the 1J1C-SANC00370 complex (Figure 96) again we see a strong periodic interaction between the ligand and ASP_681 (similar to the ARG interaction in Figure 95). There are three residues consistently interacting with SANC00370, namely SER_566, ALA_583 and VAL_635. The interaction to SER_566 is particularly rich due to the interaction of the serine OH with two oxygen atoms of the SANC00370 ligand throughout the 100 ns simulation. Up to 20 simultaneous contacts are made during the simulation, but this is early in the simulation.

As expected with the small ligand in the 4R12-SANC00256 complex (Figure 97) few contacts are made at any one time to the ligand during the 100 ns simulation. However, it is clear how reorganization of the ligand is occurring. From initial contacts with SER_128, ASN_132 and TYR_154 the ligand settles to interact instead with LYS_44, GLN_51 and SER_55 after 40 ns of the simulation. The residue which dominates the interaction between protein and receptor is LYS_44.

3.2.4.3 Protein-Ligand Contacts Detail

Supporting the analysis above was generation of 2-D contact maps detailing contacts occurring during dynamics. In these maps, interactions that occurred for more than 30.0% of the simulation time in the selected trajectory (0.00 - 100.00 ns) were included (Figures 98-102).

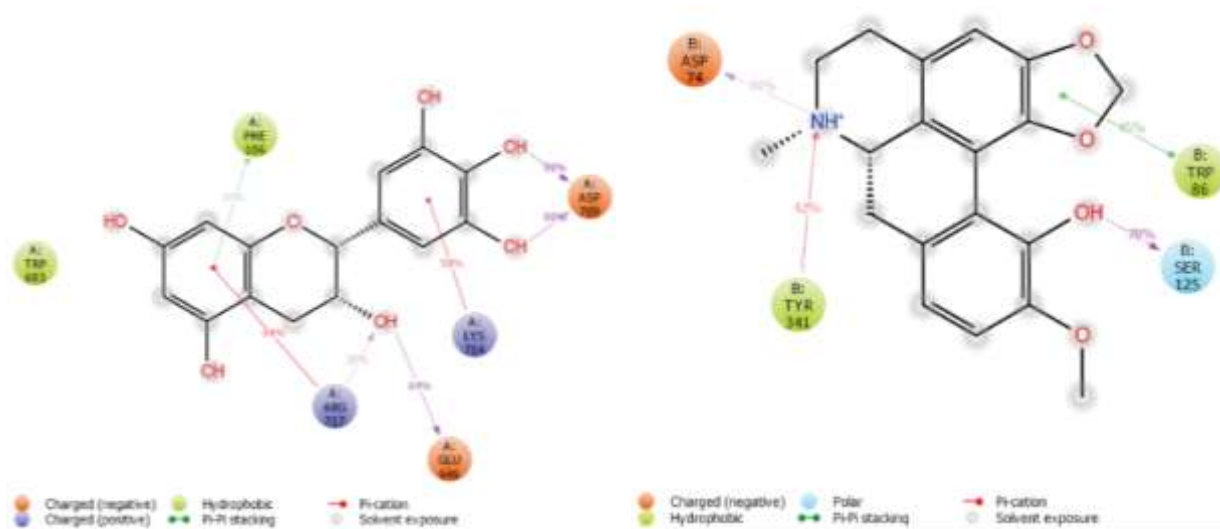


Figure 98: 4CTH-SANC00103

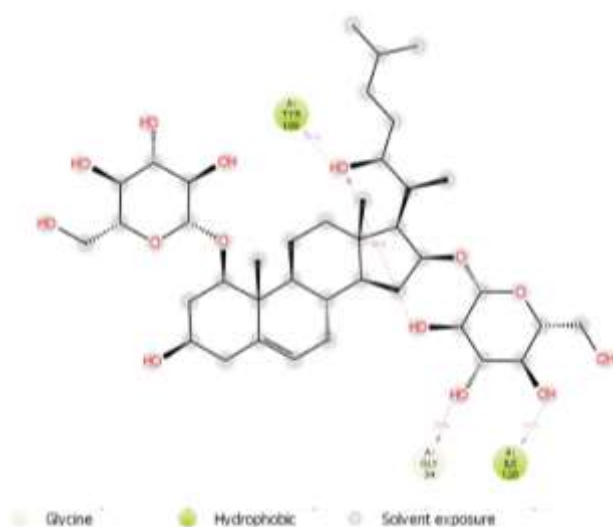


Figure 99: 4EY7_B-SANC00374

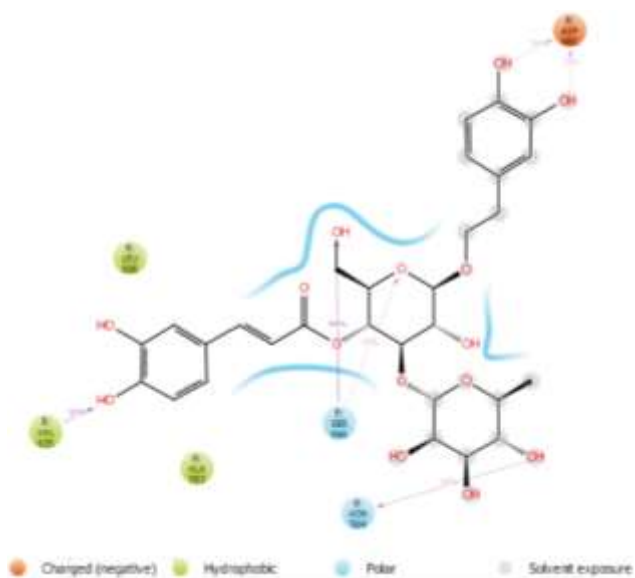


Figure 100: 1FKN_A-SANC00664

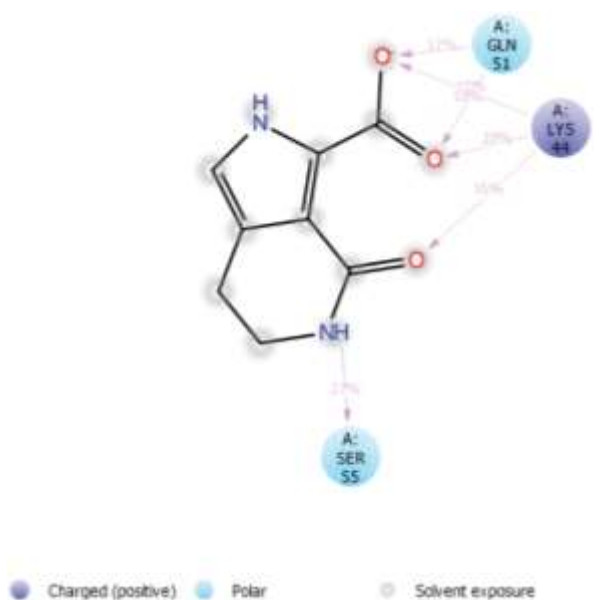


Figure 101: 1J1C_B-SANC00370

Figure 102: 4R12-SANC00256

A schematic of detailed ligand atom interactions with the protein residues. Interactions that occur more than 30.0% of the simulation time in the selected trajectory (0.00 - 100.00 ns), are shown.

For example, with the 4CTH-SANC00103 complex, the ASP_709 interaction that was reaching a value of 2.0 in Figure 88 is now detailed in Figure 98 with two OH groups of SANC00103 interacting through hydrogen bonds to ASP_702 each for 99% of the simulation. Clearer in this diagram is that ARG_717 contacts one of the SANC00103 benzene rings through a π -cation interaction for 64% of the simulation, while PHE_106 makes a π - π stacking contact for 30% of the simulation to this same aromatic system of SANC00103.

For the 4EY7-SANC00374 complex (Figure 99), a phenolic OH of SANC00374 makes hydrogen bond interactions with SER_125 for 78% of the simulation (compare Figures 89 and 94). As previously mentioned, ASP_74 forms both ionic and hydrogen bonds to the tertiary ammonium NH^+ ion while this center on the ligand also interacts with TYR_341 through hydrophobic interactions. These maps also pick up on another aspect, that of solvent exposure of the ligand. In Figure 99, during simulation, many of the atoms of ligand SANC00374 are also exposed to the solvent throughout the simulation.

The maps for the complexes 1FKN-SANC00664 and 1J1C-SANC00370 (Figures 100 and 101) detail the hydrogen bonding interactions that persisted through the respective simulations (compare Figures 95 and 96).

Finally, the limited, mainly charge-based interactions that are present in the complex 4R12-SANC00256 are detailed in Figure 102.

3.2.5 Ligand behavior during dynamics

In order to assess the stability of the ligand within the active site, it is also of use to explore geometrically the conformations available to the ligand during dynamics, together with physical properties of the ligand. Figures 103-107 show the ligand torsion profiles during the course of dynamics, while Figures 108-112 detail physical properties of the ligand within the complex over the 100ns simulation.

3.2.5.1 Ligand Torsion Profiles

The ligand torsion plots (Figures 103-107) summarize conformational change through the 100ns trajectory by detailing all rotatable bonds. The upper panel in these figures colour-codes the rotatable bonds within each ligand. Below this are the dial plot, which describes the torsion throughout the course of the simulation. For these earlier frames in the trajectory result in a point towards the center of the dial plot, later torsions are plotted radially outwards. The bar plots are a summary of the dial plots, by showing the probability density of the torsion.

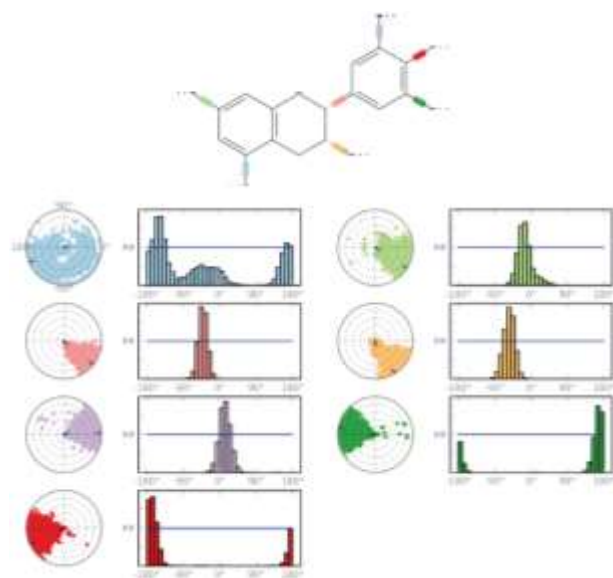


Figure 103: 4CTH-SANC00103

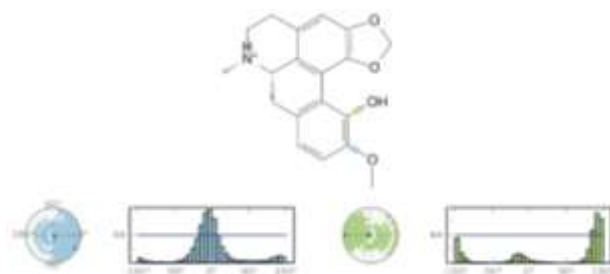


Figure 104: 4EY7_B-SANC00374

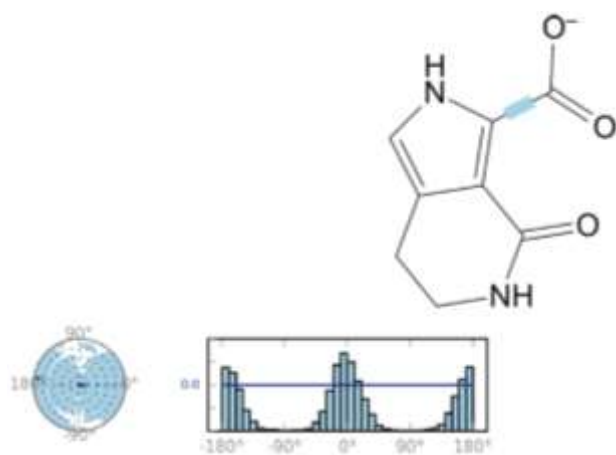
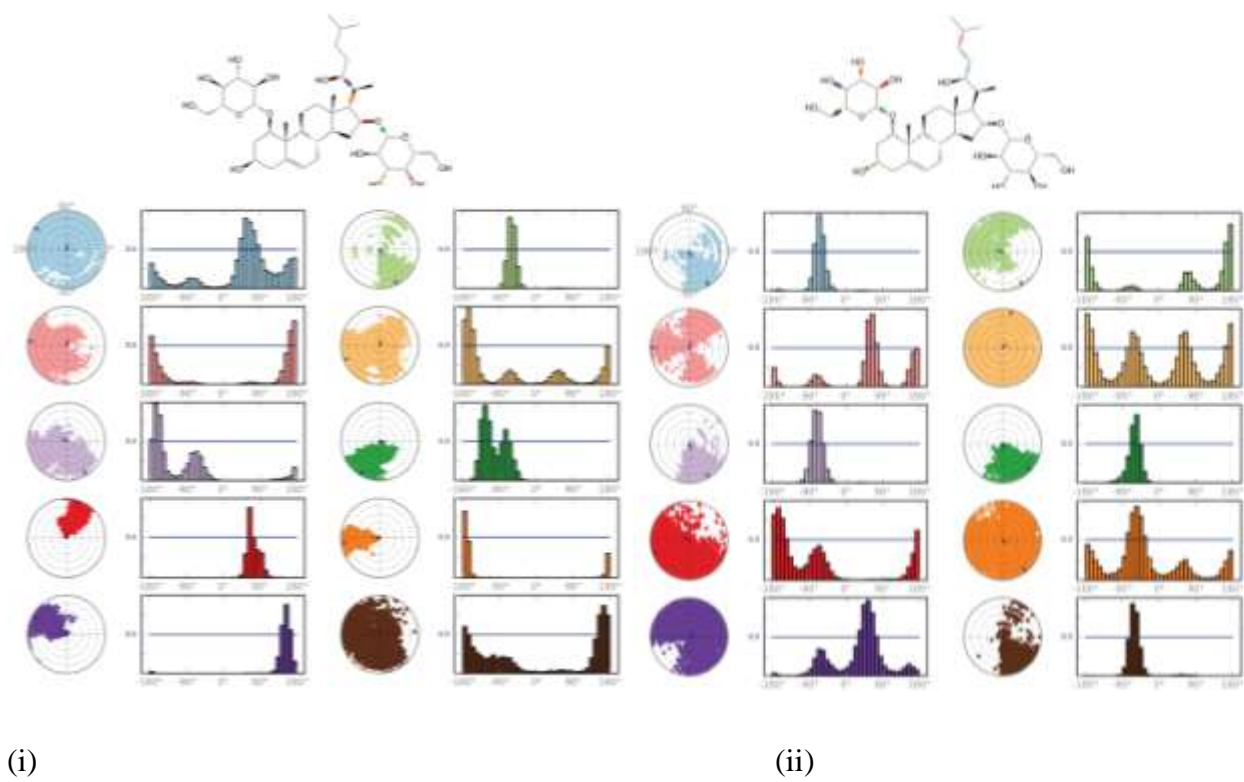
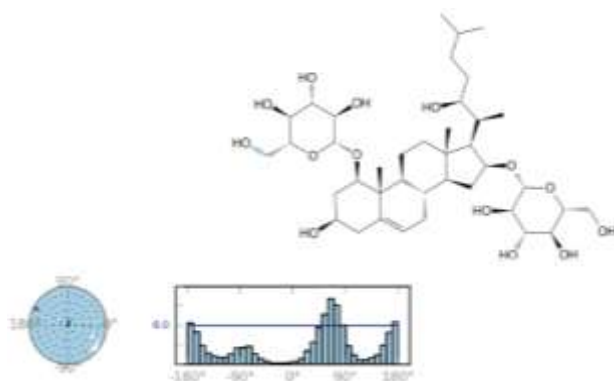


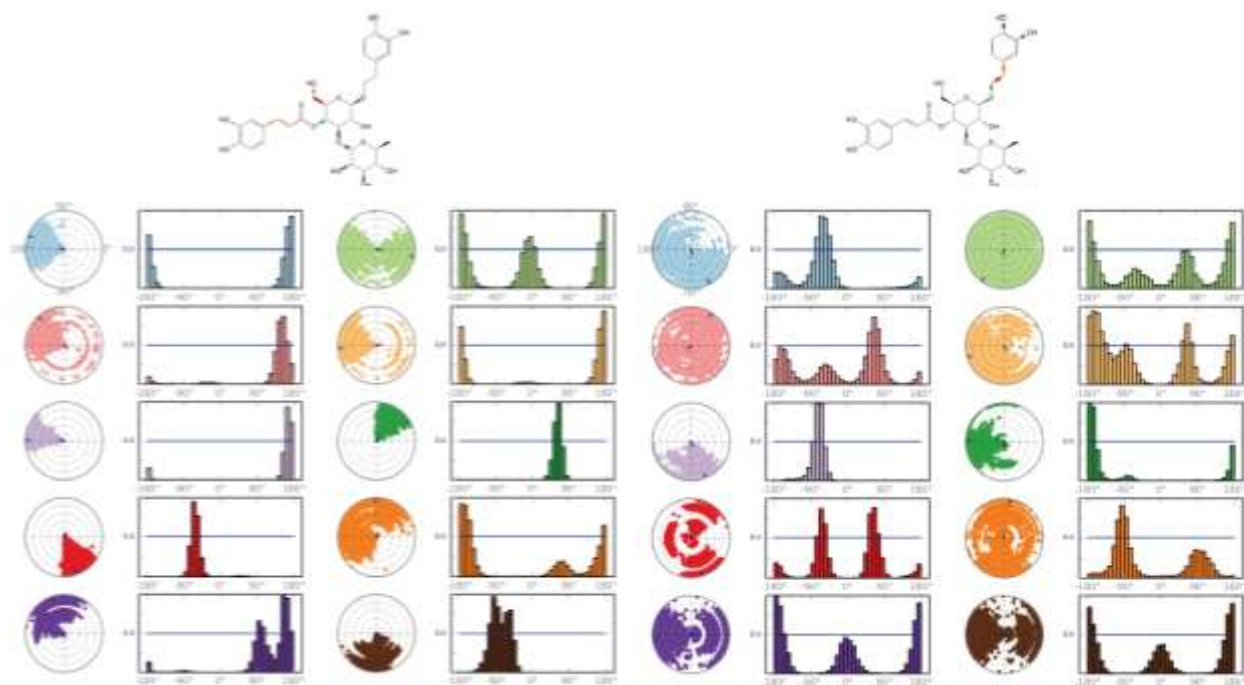
Figure 105: 4R12-SANC00256





(iii)

Figure 106: 1FKN_A-SANC00664



(i)

(ii)

Figure 107: 1J1C_B-SANC00370

Ligand SANC00103 in Figure 103 has seven rotatable bonds. Structurally, the two ring systems in the molecule are separated by the pink rotatable C-C bond, and it is interesting to note the stability of the molecule conformation during dynamics, as this bond has a torsion angle close to

-45° during the course of simulation. The blue rotatable bond is that of a hydroxyl substituent (one that is not involved in any bonding during simulation) which has the full range of torsions covered during simulation.

For SANC00374 (Figure 104), this is a planar molecule due to the interconnected rings. The only rotatable bonds are for two substituents (methoxy and hydroxy substituents). The methoxy group appears to fluctuate around 0° (blue torsion), while the hydroxy bonds spend most of the simulation at 170° but a small amount of the simulation at 0°. The 170° orientation is associated with the hydrogen bonding interaction to SER_125 of 4EY7, and this correlates to an interaction for 75% of the simulation (compare Figures 84, 99 and 94).

Figure 105 depicts SANC00256 with one rotatable bond (carboxyl group), and due to symmetry 0° will have identical geometry to 180°. Still it is interesting that this group is not fixed during simulation, given also that this carboxylate interacts with LYS_44 and GLN_51 for most of the trajectory (Figures 92, 97 and 102).

The complex 1FKN-SANC00664 under simulation leads to the torsional profile for SANC00664 shown in Figure 106. Although there are many torsion angles (some of which may be critical for hydrogen bond formation) structurally four torsions describe the orientation of the sugar moieties relative to the central steroid. Figure 106 (i) shows two of these (red and green) while Figure 106 (ii) shows the other two (green and lilac), which show the ligand is retaining its basic conformation during simulation. Several torsions describe the conformation of the aliphatic side chain, and examination of these shows that the full conformational space of this side chain is available and is explored during simulation.

Finally, for the 1J1C-SANC00370 complex (Figure 107), examining the torsional variation of the three ring systems surrounding the central sugar reveals extensive conformational change in this molecule during simulation. This information was not clear in Figures 91, 96 and 101, and it may be that although interactions are rich and maintained with certain residues throughout simulation, the ligand is able to retain flexibility in spite of being tied down at certain points to the receptor.

3.2.5.2 Ligand Structural Properties

6 properties of the ligands were also followed through each trajectory. Calculated were ligand RMSD (with respect to the ligand alone), ligand Radius of Gyration (rGyr), measuring the compactness of the ligand [102], [103], intramolecular hydrogen bonding (intraHB), molecular surface area (MolSA) (this correlates to the van der Waals surface and was calculated using a 1.4 Å probe radius) and the Solvent Accessible Surface Area (SASA), to see how much of the ligand was accessible at each point in the trajectory by water [104], [105]. Finally, the polar surface area (PSA), relating to the solvent accessible surface area, but only considering heteroatoms such as nitrogen and oxygen, was also calculated throughout the trajectory. Figures 108-112 plot these values as a function of trajectory time step.

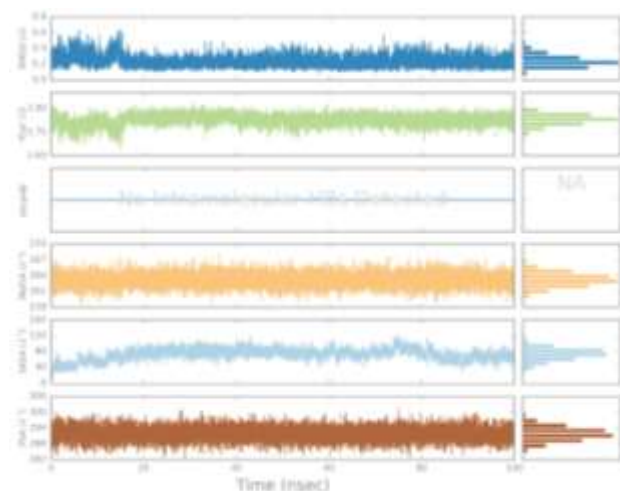


Figure 108: 4CTH-SANC00103

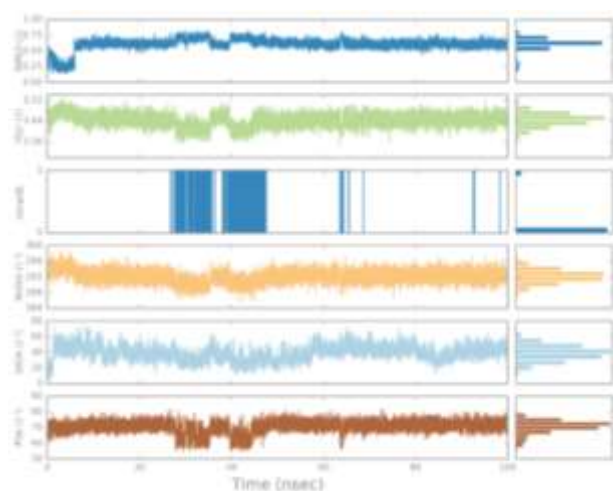


Figure 109: 4EY7_B-SANC00374

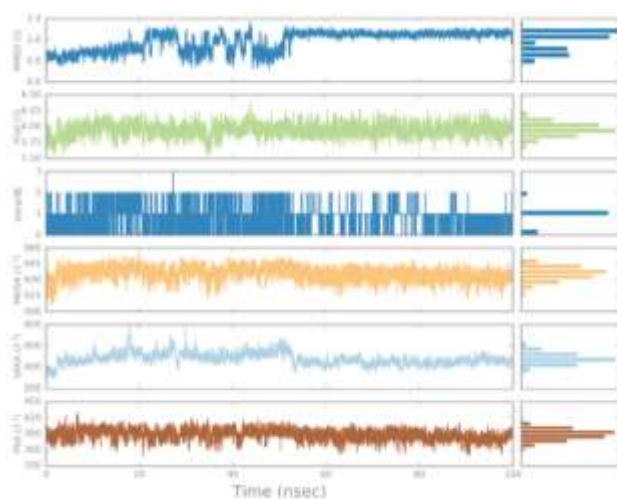


Figure 110: 1FKN_A-SANC00664

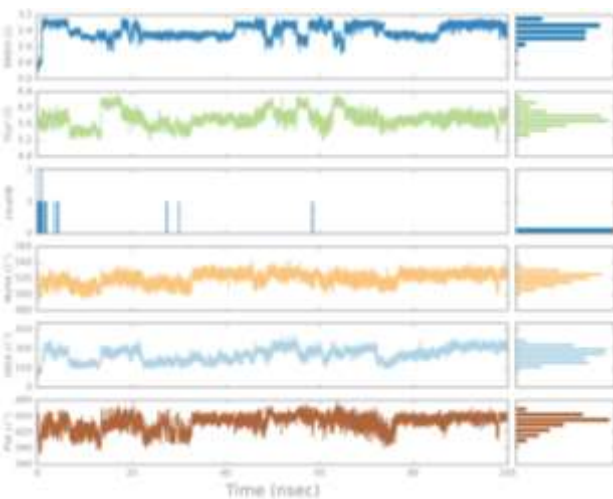


Figure 111: 1J1C_B-SANC00370

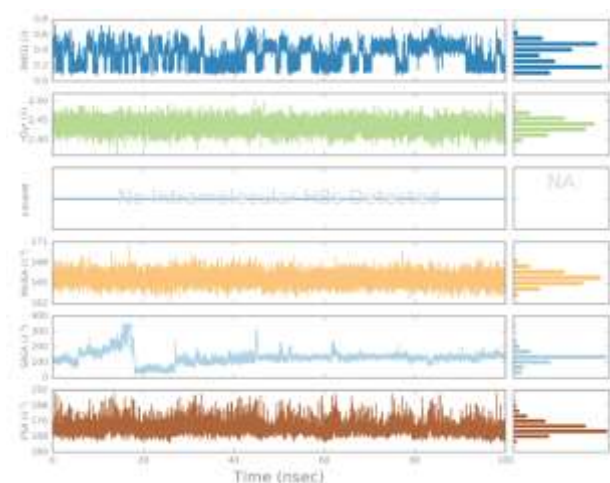


Figure 112: 4R12-SANC00256

From these plots' certain events (that have been discussed) may be followed. For example, when SANC00256 leaves the protein there is an increase in SASA (Figure 112), or the conformational flexibility of SANC00370 is easy to follow in terms of the polar surface area (Figure 111). Table 4, for convenience, summarizes the ligand properties depicted from Figure 108-112:

Protein	L-RMSD (Å)	rGyr(Å)	intraHB	molSA (Å ²)	SASA (Å ²)	PSA (Å ²)
1FKN_A (SANC00664)	2.0	6.00	1	635	450	390
1J1C_B (SANC00370)	2.6	5.60	0	522	300	448
4CTH (SANC00103)	0.2	3.80	-	263	80	290
4EY7_B (SANC00374)	0.6	3.46	0	293	42	72
4R12 (SANC00256)	0.2	2.44	-	166	120	170

Table 4: A table of the ligand properties obtained from the Desmond simulations study.

The ligand RMSD for SANC00103 (Figure 108) seems to fluctuate a lot in the first 20 ns of the simulation up to 0.6 Å but then fluctuates less for the rest of the simulation with a mean ligand RMSD of 0.2 Å. This settling-down in terms of RMSD correlates very well with the loosening of the hydrophobic PHE_106 and TRP_693 interactions, and the strengthening of the GLU_646 and ARG_717 hydrogen bonding interactions (Figure 93). These values provide information on the stability of the ligand conformation due to the ligand RMSD of less than 1.5 Å (this may be compared to Figure 65 in which the orientation or positioning of the ligand within the protein settles much later, at about 50ns). This also links up well with the torsion angle plot for SANC00103 (Figure 103), with the single torsion defining the basic (non-substituent) conformation of this ligand. The two hydrogen bonds to ASP_709 anchor one end of the ligand during the entire simulation (Figure 93),

For SANC00374 a similar picture emerges. The RMSD (Figure 109) decreases from 0.5 Å to about 0.2 Å in the first 5 ns of the simulation (correlating well with loss of a TYR_337 interaction, Figure 94) and increases significantly after 5 ns to about 0.75 Å. Towards the end of the simulation the fluctuations stabilize and the mean RMSD of this ligand is 0.60 Å. However, again comparison with Figure 66 shows that although conformation is stable, there is still some change in position within the protein later in the simulation. Bear in mind that SANC00374 only has two torsions, a

methoxy and hydroxyl group torsion available for conformation change (Figure 104). Variations in the molecular surface area, polar surface area and solvent accessible surface for SANC00374 match well the formation or loss of hydrogen bonding to SER_125 of 4EY7. Polar surface area (PSA) is a commonly used medicinal chemistry metric for the optimization of a drug's ability to permeate cells. Molecules with a PSA of $\geq 140 \text{ \AA}^2$ tend to be poor at permeating cell membranes[106]. A PSA $\leq 90 \text{ \AA}^2$ is typically needed for molecules to penetrate the blood-brain barrier[107]. According to these values, SANC00374 which is docked to the acetylcholinesterase 4EY7_B can penetrate the blood-brain barrier because it has a PSA less than 90 \AA^2 . This is in comparison to the other four SANCDB compounds with PSA values greater than 140 \AA^2 , which will therefore make them poor at permeating membranes. This ligand has the lowest SASA mean value of 42 \AA^2 , because the ligand is buried inside the protein binding site and has lower opportunity for interacting with the solvent in the system.

For SANC00664 (Figure 110), although there is large fluctuation for the first half of the simulation, the ligand is very stable conformationally for the remainder of the simulation. This matches the emergence of an interaction of the ligand with GLY_34 (Figure 95), and it is likely that it is in the early stages of the simulation that the indirect effect it has on the β -sheet around LEU_119 is in force (Figure 70, protein RMSF). SANC00664 is very flexible, and although the two substituent two rings stabilize in terms of conformation, the aliphatic substituent remains in motion. Further, there are not extensive interactions with 1FKN, with total contacts fluctuating around 5 during the latter portion of the simulation (Figure 95). Given this, the remarkable stability of the ligand within the active site in the latter portion of the simulation is worth noting. Structurally the steroid core of SANC00664 is not deep within the active site (see for example Figure 27) yet is held stably to the surface of the protein on the edge of the active site through hydrophobic interactions. Given the size of the ligand, and its fixed conformation with rings spread out, it is not surprising that it has the largest rGyr value compared to all other ligands studied (Table 4). One contributor to the consistency of conformation for SANC00664 may be the 1-2 intramolecular hydrogen bonds observed, locking the conformation independently of protein interactions.

Concerns were noted earlier on the non-selectivity of SANC00370 during docking. There is a correlation between ligand RMSD (Figure 111) and the aligned ligand RMSD of Figure 66. This

is attributed to large changes in conformation taking place during the simulation affecting the binding pose, in accordance with the observations regarding Figure 107 in which there is a large torsional variation with respect to substituents on the central ring. Although there are three interactions maintained during simulation (to SER_566, ALA_583 and VAL_635) this is not enough to lock conformation of this system. Although the docking values were good, the ligand does not maintain a specific pose during dynamics. The rGyr is quite high for SANC00370 (although not as high as for SANC00664), but the comparison between them (Figures 110 and 111) show how the compactness varies greatly for SANC00370 during simulation as its conformation continually changes, while that for SANC00664, the rGyr is constant as SANC00664 maintains a relatively fixed conformation in the active site. This ligand SANC00370 together with previously discussed SANC00664 have high ligand SASA fluctuations with mean SASA values $\geq 300 \text{ \AA}^2$; and this indicates that both structures have significant solvent exposure during the simulation. Further, both molecules have high molSA values compared to others as shown in Table 4 due to their bulky sizes.

Finally, the “bimodal” fluctuation in the RMSD of SANC00256 during simulation is interesting (Figure 112). This ligand was observed to leave the active site and return during simulation. The solvent-accessible surface area in Figure 112 is a good indication of the position of SANC00265 in relation to the protein as it leaves and returns to the active site. Even after settling in the active site, this fluctuation is maintained. For SANC00256 the only rotatable bond is that of a carboxylate side chain; the symmetry of the side chain means that both conformations of the carboxylate are identical. However, in RMSD calculation this will not be considered, hence the variation. Thus the “switches” in RMSD are a good indication of the rate of rotation of the carboxylate group. As expected, this small ligand has the lowest rGyr (Table 4), and the lowest molSA compared to other ligands.

Given the mobility of SANC00370 in the active site of 1J1C_B, and that SANC00256 leaves 4R12 during dynamics, these two ligands will not be taken for further study as lead compounds. However, of the three remaining complexes, 1FKN-SANC00664, 4CTH-SANC00103 and 4EY7-SANC00374, all will be taken further, but the conformational consistency of SANC00664 will make 1FKN-SANC00664 an interesting study in lead optimization. Further, given that PubChem

101608678 is very similar to SANC00664, there is a richness to be followed in the study of these substituted steroids as β -secretase inhibitors. 4CTH and 4EY7 are neprilysin and acetylcholinesterase.

3.3 Further protein-ligand complex dynamics

Twelve further molecular dynamics simulations using Desmond were performed in total (on further targets and using the SANCDB-informed PubChem structures). Of these 9 simulations produced promising results with the RMSDs showing that the ligand is stable and remains within the active site for the full course of simulation. The extent of these ‘good’ results will not be discussed; however, all the diagrams are provided in Chapter 3.2 and Appendix B.

Table 5 lists the additional simulations that have been performed and for which analyses are included in the supplementary information. These include γ -secretase, tau protein and α -secretase targets to evaluate leads for these systems.

Targets	Agents
1FKN_B	(SANC00562)
1GQR	(SANC00111)
1J1C_A	(SANC00103)
3LII_A	(SANC00269)
3LII_B	(SANC00559)
4EY7_A	(SANC00103)
5E2W	(SANC00349)
6C2I	(SANC00370)

1FKN_A	(101608678)
1GQR	(11777124)
1J1C_A	(122542902)
4CTH	(59370901)

Table 5: *A table showing pairs of potential agents and targets also taken further to molecular dynamics.*

CHAPTER 4

Conclusion

Alzheimer's Disease is a disease that does not have a cure, and therefore there is a need to identify new leads and inhibitors for all known potential AD targets. Although it has been more than a century since AD was discovered, the causes of the disease are still poorly understood, and this has been a major obstacle in drug design. Nonetheless, amyloid β -peptide has been found as a potential centre of AD pathogenesis, and this has led to the identification of associated potential therapeutic targets, which we have used in this study. These targets we have used have been extensively studied, with one of them having produced an FDA approved drugs against the disease. The aim of this study was to identify potential agents of AD from SANCDB through the use of molecular docking, where binding energy and receptor-ligand complex interactions would help in the determination and evaluation of potential compounds. These potential lead compounds were further evaluated using molecular dynamics simulations to study the receptor-ligand complex interactions. Receptor, ligand and receptor-ligand properties such as RMSD, RMSF, Radius of gyration, contacts, torsions etc., informed the analysis of which agents or complexes will be taken for further study.

In silico molecular docking was performed with 715 SANCDB compounds docked against 14 protein targets using a high-throughput virtual screening (HTVS) technique. Where ligands were appropriate for the active site, a successful docking result was obtained; all such ligands docked within the binding site, and the best binding ligands exhibited good binding affinities of less than -5 kcal/mol. The complex comprising ligand SANC00370 docked against protein 1J1C_B exhibited the greatest binding affinity of -13.94 kcal/mol compared to the other ligands, and from the docking had clear strong receptor-ligand interactions.

19 protein ligand complexes were taken to molecular dynamics, of which 5 were analyzed in detail. Of these five complexes it was interesting to observe that although the SANC00370-1J1C complex showed promising docking results, the variability of the ligand pose during dynamics, together with what appears to be lower selectivity shown by SANC00370 (binding well to very different proteins) rules SANC00370 out for further study.

One other ligand was observed to leave the target during dynamics and return to the active site. This ligand (SANC00256) has also been determined to be not suitable for further study.

By contrast, three simulations provided promising results based on detailed study of RMSD and RMSF of both ligand and target during dynamics, the interactions and their nature and timeframe during dynamics, and information with respect to ligand conformation in the active site. The physicochemical properties of all three ligands are satisfactory. These three receptor-ligand complexes are 1FKN-SANC00664, 4CTH-SANC00103 and 4EY7-SANC00374.

The molecular docking of the PubChem analogues of the SANCDB best ligands provided for 6 protein-ligand complexes with improved binding affinity compared to the SANCDB ligands. 4 of these complexes have been taken forward to molecular dynamics simulation, and the results for 3 of these are promising, as determined by the stability of the ligand within the active site. These will also be taken for further study.

Further studies will involve optimization of compounds SANC00664, SANC00103 and SANC00374 in the context of their respective targets, together with further analysis of the additional dynamics trajectories. In the context of AD, it would be interesting to follow a similar exploration of other potential targets such as the A β _{1-40/42} peptide and prevention of A β -mediated oxidative stress, and inhibition of associated proteins surrounding tau hyperphosphorylation, [108],[109],[110],[111].

CHAPTER 5

Bibliography

- [1] A. L. Pascoini, L. B. Federico, A. L. F. Arêas, B. A. Verde, P. G. Freitas, and I. Camps, "In silico development of new acetylcholinesterase inhibitors," *J. Biomol. Struct. Dyn.*, vol. 37, no. 4, pp. 1007–1021, 2019.
- [2] Y. . Hamada and Y. Kiso, "Discovery of BACE1 Inhibitors for the Treatment of Alzheimer's Disease," *Quant. Struct. Relatsh.*, p. 27, 2017.
- [3] M. Goedert, "Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled A β , tau, and α -synuclein," *Science* (80-.), vol. 349, no. 6248, pp. 61–69, 2015.
- [4] J. Gaugler, B. James, T. Johnson, K. Scholz, and J. Weuve, "2016 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, vol. 12, no. 4, pp. 459–509, 2016.
- [5] G. M. A. Nowotny Petra, Kwon M Jennifer, "Alzheimer Disease," *Nat. Publ. Gr.*, 2001.
- [6] C. A. de Jager, J. A. Joska, M. Hoffman, K. E. Borochowitz, and M. I. Combrinck, "Dementia in rural South Africa: A pressing need for epidemiological studies," *South African Med. J.*, vol. 105, no. 3, pp. 189–190, 2015.
- [7] M. Jouanne, S. Rault, and A. S. Voisin-Chiret, "Tau protein aggregation in Alzheimer's disease: An attractive target for the development of novel therapeutic agents," *Eur. J. Med. Chem.*, vol. 139, pp. 153–167, 2017.
- [8] Y. Li, J. Wang, S. Zhang, and Z. Liu, "Neprilysin gene transfer: A promising therapeutic approach for Alzheimer's disease," *J. Neurosci. Res.*, vol. 93, no. 9, pp. 1325–1329, 2015.
- [9] I. Truter, "Prescribing of drugs for Alzheimer's disease: A South African database analysis," *Int. Psychogeriatrics*, vol. 22, no. 2, pp. 264–269, 2010.
- [10] M. Marketing, "Alzheimer's SA NPC," *Dementia Care Facility - Livewell Villages*, 1985. [Online]. Available: <https://alzheimers.org.za/prevalence-of-dementia/>.
- [11] B. De Strooper, R. Vassar, and T. Golde, "The secretases: Enzymes with therapeutic potential in Alzheimer disease," *Nat. Rev. Neurol.*, vol. 6, no. 2, pp. 99–107, 2010.
- [12] R. Epis, E. Marcello, F. Gardoni, and M. Di Luca, "Alpha, beta- and gamma-secretases in Alzheimer's Disease," *Front. Biosci.*, vol. 42, no. 2, pp. 1126–1150, 2012.
- [13] G. M. Pasinetti and S. Hiller-Sturmhöfel, "Systems biology in the study of neurological disorders: focus on Alzheimer's disease," *Alcohol Res. Health*, vol. 31, no. 1, pp. 60–5, 2008.
- [14] N. Zhao, C. C. Liu, W. Qiao, and G. Bu, "Apolipoprotein E, Receptors, and Modulation of Alzheimer's Disease," *Biol. Psychiatry*, vol. 83, no. 4, pp. 347–357, 2018.
- [15] G. M. Chow, M. Vivian, Mattson P. Mark, Wong C. Philip, "An Overview of APP Processing Enzymes and Products," *Neuromolecular Med.*, vol. 12, no. 1, pp. 1–12, 2010.
- [16] A. Hiremathad, K. Chand, and R. S. Keri, "Development of coumarin–benzofuran hybrids as versatile multitargeted compounds for the treatment of Alzheimer's Disease," *Chem. Biol. Drug Des.*, vol. 92, no. 2, pp. 1497–1503, 2018.
- [17] P. J. Dolan and G. V. W. Johnson, "The role of tau kinases in Alzheimer's disease," *Curr. Opin. Drug Discov. Dev.*, vol. 13, no. 5, pp. 595–603, 2010.
- [18] B. Nizynski, W. Dzwolak, and K. Nieznanski, "Amyloidogenesis of Tau protein," *Protein Sci.*, vol. 26, no. 11, pp. 2126–2150, 2017.
- [19] J. Avila, "Tau phosphorylation and aggregation in Alzheimer's disease pathology," *FEBS*

- Lett.*, vol. 580, no. 12, pp. 2922–2927, 2006.
- [20] A. M. Pooler, W. Noble, and D. P. Hanger, “A role for tau at the synapse in Alzheimer’s disease pathogenesis,” *Neuropharmacology*, vol. 76, no. PART A, pp. 1–8, 2014.
 - [21] R. B. Maccioni, G. Farias, I. Morales, and L. Navarrete, “The Revitalized Tau Hypothesis on Alzheimer’s Disease,” *Arch. Med. Res.*, vol. 41, no. 3, pp. 226–231, 2010.
 - [22] F. Kametani and M. Hasegawa, “Reconsideration of amyloid hypothesis and tau hypothesis in Alzheimer’s disease,” *Front. Neurosci.*, vol. 12, no. JAN, 2018.
 - [23] P. Thomson, “Amyloid Hypothesis vs Tau Hypothesis,” 2018. [Online]. Available: <https://www.stressmarq.com/amyloid-hypothesis-vs-tau-hypothesis/?v=68caa8201064>. [Accessed: 24-Oct-2019].
 - [24] F. M. LaFerla and S. Oddo, “Alzheimer’s disease: A β , tau and synaptic dysfunction,” *Trends in Molecular Medicine*. 2005.
 - [25] C. X. Gong, F. Liu, I. Grundke-Iqbal, and K. Iqbal, “Post-translational modifications of tau protein in Alzheimer’s disease,” *J. Neural Transm.*, vol. 112, no. 6, pp. 813–838, 2005.
 - [26] M. Morishima-Kawashima; *et al.*, “Proline-directed and Non-proline-directed Phosphorylation of PHF-tau,” *J. Biol. Chem.*, vol. 270, no. 2, pp. 823–829, 1995.
 - [27] C. W. Scott *et al.*, “Phosphorylation of Recombinant Tau by Protein Kinase,” *Biochemistry*, vol. 268, no. 2, pp. 1166–1173, 1993.
 - [28] J. Busciglio, A. Lorenzo, J. Yeh, and B. A. Yankner, “ β -Amyloid fibrils induce tau phosphorylation and loss of microtubule binding,” *Neuron*, vol. 14, no. 4, pp. 879–888, 1995.
 - [29] R. A. Goedert, Michel; Klug, Aaron; Crowther, “Tau protein, the paired helical filament and Alzheimer’s disease,” *J. Alzheimer’s Dis.*, vol. 9, no. 3, pp. 195–207, 2006.
 - [30] T. L. Spires-Jones and B. T. Hyman, “The Intersection of Amyloid Beta and Tau at Synapses in Alzheimer’s Disease,” *Neuron*, vol. 82, no. 4, pp. 756–771, 2014.
 - [31] D. S. BIOVIA, “Discovery Studio Visualizer.” Dassault Systèmes, San Diego, 2016.
 - [32] S. Makin, “Alzheimer’s on trial,” *Nature*, vol. 559, 2018.
 - [33] X. Sun, W. D. Chen, and Y. D. Wang, “ β -Amyloid: The key peptide in the pathogenesis of Alzheimer’s disease,” *Front. Pharmacol.*, vol. 6, no. SEP, pp. 1–9, 2015.
 - [34] R. Vassar, “BACE1 inhibitor drugs in clinical trials for Alzheimer’s disease,” *Alzheimer’s Res. Ther.*, vol. 6, no. 9, pp. 1–14, 2014.
 - [35] G. G. Glenner and C. W. Wong, “Alzheimer’s disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein,” *Biochem. Biophys. Res. Commun.*, vol. 120, no. 3, pp. 885–890, 1984.
 - [36] J. Lee, VM, Balin, BJ, Otvos Jr, L, Trojanowski, “A68. A major subunit of paired helical filaments and derivatized forms of normal tau,” *Am. Assoc. Adv. Sci.*, vol. 251, no. 4994, pp. 675–678, 1991.
 - [37] R. Vassar, “The β -Secretase Enzyme in Alzheimer’s Disease.pdf,” vol. 23, pp. 105–113, 2004.
 - [38] R. Vassar, “BACE1 Inhibition as a Therapeutic Strategy for Alzheimer’s Disease,” *J. Sport Heal. Sci.*, vol. 5, no. 4, pp. 388–390, 2016.
 - [39] F. Peters *et al.*, “BACE1 inhibition more effectively suppresses initiation than progression of β -amyloid pathology,” *Acta Neuropathol.*, vol. 135, no. 5, pp. 695–710, 2018.
 - [40] F. Fahrenholz, “Alpha-Secretase As a Therapeutic Target,” *Curr. Alzheimer Res.*, vol. 4,

- no. 4, pp. 412–417, 2007.
- [41] P. H. Kuhn *et al.*, “ADAM10 is the physiologically relevant, constitutive α -secretase of the amyloid precursor protein in primary neurons,” *EMBO J.*, vol. 29, no. 17, pp. 3020–3032, 2010.
 - [42] T. E. Golde, E. H. Koo, K. M. Felsenstein, B. A. Osborne, and L. Miele, “ γ -Secretase inhibitors and modulators,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1828, no. 12, pp. 2898–2907, 2013.
 - [43] P. Anand, B. Singh, and N. Singh, “A review on coumarins as acetylcholinesterase inhibitors for Alzheimer’s disease,” *Bioorganic Med. Chem.*, vol. 20, no. 3, pp. 1175–1180, 2012.
 - [44] A. . McGleenon, B.M, Dynan, K.B, Passmore, “Acetylcholinesterase Inhibitors in Alzheimer’s Disease,” *Br. J. Clin. Pharmacol.*, vol. 48, no. 4, pp. 471–480, 1999.
 - [45] H. Zhang, D. Liu, Y. Wang, H. Huang, Y. Zhao, and H. Zhou, “Meta-analysis of expression and function of neprilysin in Alzheimer’s disease,” *Neurosci. Lett.*, vol. 657, pp. 69–76, 2017.
 - [46] K. Shirotani *et al.*, “Neprilysin Degrades Both Amyloid β Peptides 1-40 and 1-42 Most Rapidly and Efficiently among Thiorphan- and Phosphoramidon-sensitive Endopeptidases,” *J. Biol. Chem.*, vol. 276, no. 24, pp. 21895–21901, 2001.
 - [47] S. S. El-Amouri, H. Zhu, J. Yu, R. Marr, I. M. Verma, and M. S. Kindy, “Neprilysin: An enzyme candidate to slow the progression of Alzheimer’s disease,” *Am. J. Pathol.*, vol. 172, no. 5, pp. 1342–1354, 2008.
 - [48] R. Marr and D. Hafez, “Amyloid beta and Alzheimer’s Disease: The role of neprilysin-2 in amyloid-beta clearance,” *Front. Aging Neurosci.*, vol. 6, no. JUL, pp. 1–7, 2014.
 - [49] N. N. Nalivaeva, N. D. Belyaev, I. A. Zhuravin, and A. J. Turner, “The Alzheimers amyloid-degrading peptidase, neprilysin: Can we control it?,” *Int. J. Alzheimers. Dis.*, vol. 2012, 2012.
 - [50] L. C. Blum and J. L. Reymond, “970 Million druglike small molecules for virtual screening in the chemical universe database GDB-13,” *J. Am. Chem. Soc.*, vol. 131, no. 25, pp. 8732–8733, 2009.
 - [51] S. . Renner *et al.*, “Recent trends and observations in the design of high-quality screening collections,” *Future Med. Chem.*, vol. 3, no. 6, pp. 751–766, 2011.
 - [52] I. . Kola and J. Landis, “Can the pharmaceutical industry reduce attrition rates?,” *Nat. Rev. Drug Discov.*, vol. 3, no. 8, p. 711, 2004.
 - [53] M. M. Hann, “Molecular obesity, potency and other addictions in drug discovery,” *Medchemcomm*, vol. 2, no. 5, pp. 349–355, 2011.
 - [54] L. Ruddigkeit, R. Van Deursen, L. C. Blum, and J. L. Reymond, “Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17,” *J. Chem. Inf. Model.*, vol. 52, no. 11, pp. 2864–2875, 2012.
 - [55] J. J. Irwin and B. K. Shoichet, “ZINC - A free database of commercially available compounds for virtual screening,” *J. Chem. Inf. Model.*, vol. 45, no. 1, pp. 177–182, 2005.
 - [56] D. S. Wishart *et al.*, “DrugBank 5.0: A major update to the DrugBank database for 2018,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D1074–D1082, 2018.
 - [57] F. Ntie-Kang *et al.*, “ConMedNP: A natural product library from Central African medicinal plants for drug discovery,” *RSC Adv.*, vol. 4, no. 1, pp. 409–419, 2014.
 - [58] F. Ntie-Kang *et al.*, “NANPDB: A Resource for Natural Products from Northern African

- Sources,” *J. Nat. Prod.*, vol. 80, no. 7, pp. 2067–2076, 2017.
- [59] F. Ntie-Kang *et al.*, “AfroDb: a select highly potent and diverse natural product library from African medicinal plants,” *PLoS One*, vol. 8, no. 10, pp. 1–15, 2013.
 - [60] R. J. Quinn *et al.*, “Developing a drug-like natural product library (Journal of Natural Products (2008) 71, (468)),” *J. Nat. Prod.*, vol. 71, no. 10, p. 1792, 2008.
 - [61] R. Hatherley *et al.*, “SANCDB: A South African natural compound database,” *J. Cheminform.*, vol. 7, no. 1, pp. 1–9, 2015.
 - [62] R. S. Bohacek, C. McMartin, and W. C. Guida, “The art and practice of structure-based drug design: A molecular modeling perspective,” *Med. Res. Rev.*, vol. 16, no. 1, pp. 3–50, 1996.
 - [63] A. Ganesan, M. L. Coote, and K. Barakat, “Molecular dynamics-driven drug discovery: leaping forward with confidence,” *Drug Discov. Today*, vol. 22, no. 2, pp. 249–269, 2017.
 - [64] R. A. Friesner *et al.*, “Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy,” *J. Med. Chem.*, vol. 47, no. 7, pp. 1739–1749, 2004.
 - [65] D. A. Vanommeslaeghe, Kenno; Guvench, Olgun; MacKerell Jr., “Macromolecular mechanics,” *Curr. Pharm. Des.*, vol. 20, no. 20, pp. 3281–3292, 2014.
 - [66] N. L. Allinger, Y. H. Yuh, and J. H. Lii, “Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1,” *J. Am. Chem. Soc.*, vol. 111, no. 23, pp. 8551–8566, 1989.
 - [67] G. R. Brubaker and D. W. Johnson, “Molecular mechanics calculations in coordination chemistry,” *Coord. Chem. Rev.*, vol. 53, no. C, pp. 1–36, 1984.
 - [68] D. A. Pearlman *et al.*, “AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules,” *Comput. Phys. Commun.*, vol. 91, no. 1–3, pp. 1–41, 1995.
 - [69] B. R. Brooks *et al.*, “CHARMM: Molecular dynamics simulation package,” *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, 2009.
 - [70] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, “GROMACS: A message-passing parallel molecular dynamics implementation,” *Comput. Phys. Commun.*, vol. 91, no. 1–3, pp. 43–56, 1995.
 - [71] E. Harder *et al.*, “OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins,” *J. Chem. Theory Comput.*, vol. 12, no. 1, pp. 281–296, 2016.
 - [72] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, “Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids,” *J. Am. Chem. Soc.*, vol. 118, no. 45, pp. 11225–11236, 1996.
 - [73] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, “Role of Molecular Dynamics and Related Methods in Drug Discovery,” *J. Med. Chem.*, vol. 59, no. 9, pp. 4035–4061, 2016.
 - [74] P. Robustelli, S. Piana, and D. E. Shaw, “Developing a molecular dynamics force field for both folded and disordered protein states,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 21, pp. E4758–E4766, 2018.
 - [75] T. Hansson, C. Oostenbrink, and W. F. Van Gunsteren, “Molecular dynamics simulations Hansson, Oostenbrink and van Gunsteren 191,” pp. 190–196, 2002.
 - [76] R. J. Castellani and G. Perry, “The complexities of the pathology-pathogenesis relationship in Alzheimer disease,” *Biochem. Pharmacol.*, vol. 88, no. 4, pp. 671–676, 2014.

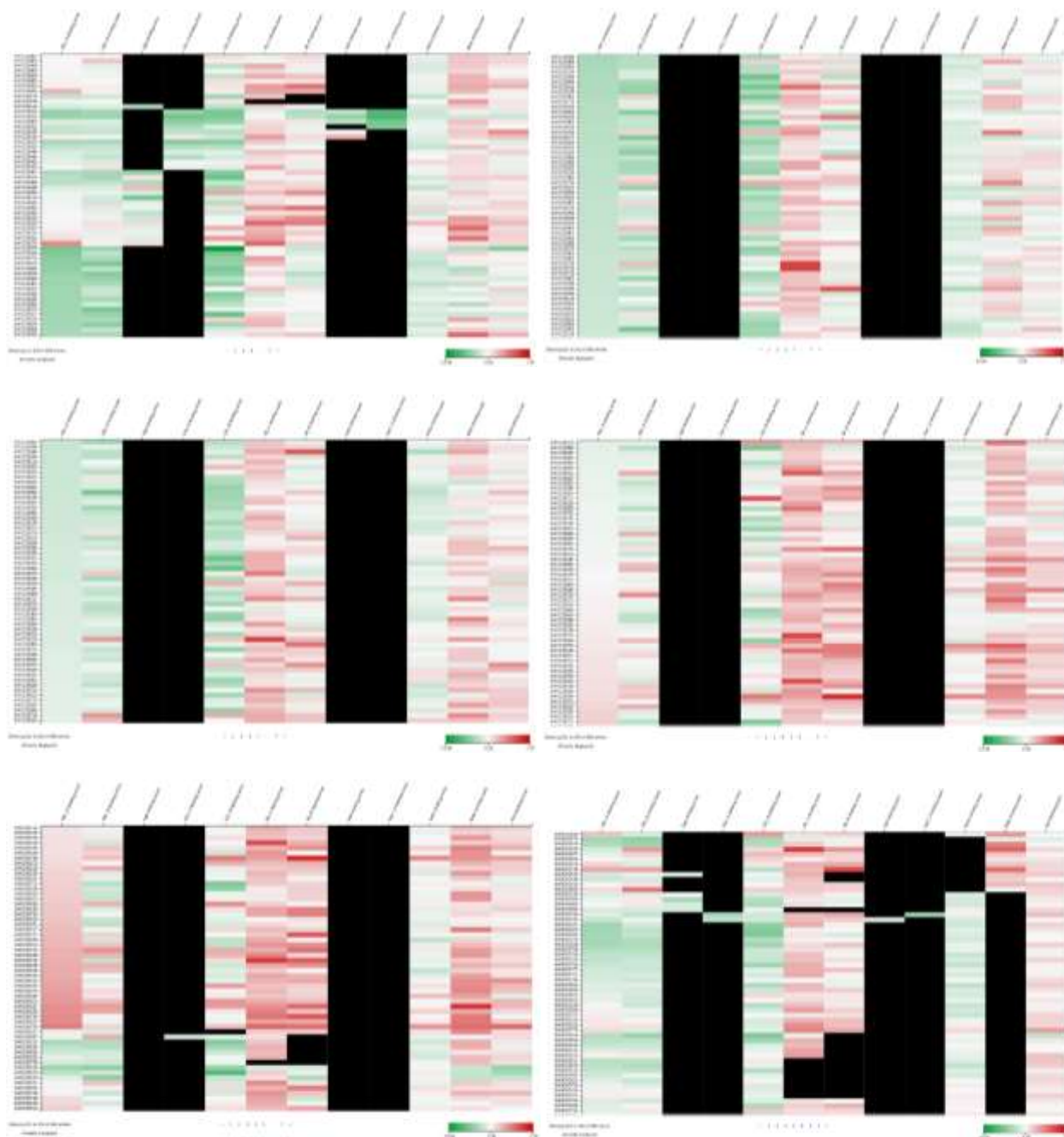
- [77] J. Godyń, J. Jończyk, D. Panek, and B. Malawska, "Therapeutic strategies for Alzheimer's disease in clinical trials," *Pharmacol. Reports*, vol. 68, no. 1, pp. 127–138, 2016.
- [78] J. L. Cummings, T. Morstorf, and K. Zhong, "Alzheimer's disease drug-development pipeline: Few candidates, frequent failures," *Alzheimer's Res. Ther.*, vol. 6, no. 4, pp. 1–7, 2014.
- [79] S. H. Barage and K. D. Sonawane, "Amyloid cascade hypothesis: Pathogenesis and therapeutic strategies in Alzheimer's disease," *Neuropeptides*, vol. 52, pp. 1–18, 2015.
- [80] S. F. Lichtenthaler, "Alpha-secretase in Alzheimer's disease: Molecular identity, regulation and therapeutic potential," *J. Neurochem.*, vol. 116, no. 1, pp. 10–21, 2011.
- [81] S. Kim *et al.*, "PubChem 2019 update: Improved access to chemical data," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D1102–D1109, 2019.
- [82] R. M. Berthold *et al.*, "KNIME: The Konstanz Information Miner. In: Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization.," *Springer, Berlin, Heidelb.*, 2008.
- [83] Schrodinger, "Maestro." LLC, New York, 2019.
- [84] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, "Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments," *J. Comput. Aided. Mol. Des.*, vol. 27, no. 3, pp. 221–234, 2013.
- [85] Schrodinger, "LigPrep." LLC, New York, 2019.
- [86] R. A. Friesner *et al.*, "Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes," *J. Med. Chem.*, vol. 49, no. 21, pp. 6177–6196, 2006.
- [87] D. D. Version and D. D. Version, "Desmond Tutorial," *Schroedinger*, no. April, 2011.
- [88] S. B. Mirza, R. E. Salmas, M. Q. Fatmi, and S. Durdagi, "Virtual screening of eighteen million compounds against dengue virus: Combined molecular docking and molecular dynamics simulations study," *J. Mol. Graph. Model.*, vol. 66, pp. 99–107, 2016.
- [89] E. Born, Max; Wolf, "Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light (7th ed.)," 7th ed., Cambridge: Cambridge University Press, 1999, pp. 92–94.
- [90] R. Van Deursen, L. C. Blum, and J. L. Reymond, "Visualisation of the chemical space of fragments, lead-like and drug-like molecules in PubChem," *J. Comput. Aided. Mol. Des.*, vol. 25, no. 7, pp. 649–662, 2011.
- [91] K. T. Nguyen, L. C. Blum, R. Van Deursen, and J. L. Reymond, "Classification of organic molecules by molecular quantum numbers," *ChemMedChem*, vol. 4, no. 11, pp. 1803–1805, 2009.
- [92] M. Awale, R. Van Deursen, and J. L. Reymond, "MQN-mapplet: Visualization of chemical space with interactive maps of DrugBank, ChEMBL, PubChem, GDB-11, and GDB-13," *J. Chem. Inf. Model.*, vol. 53, no. 2, pp. 509–518, 2013.
- [93] E. Stjernschantz and C. Oostenbrink, "Improved ligand-protein binding affinity predictions using multiple binding modes," *Biophys. J.*, vol. 98, no. 11, pp. 2682–2691, 2010.
- [94] A. Jain, "Scoring Functions for Protein-Ligand Docking," *Curr. Protein Pept. Sci.*, vol. 7, no. 5, p. 14, 2006.
- [95] L. Z. Benet, C. M. Hosey, O. Ursu, and T. I. Oprea, "BDDCS, the Rule of 5 and

- Drugability Graphical abstract HHS Public Access,” *Adv Drug Deliv Rev*, vol. 101, pp. 89–98, 2016.
- [96] C. A. Lipinski, “Lead- and drug-like compounds: The rule-of-five revolution,” *Drug Discov. Today Technol.*, vol. 1, no. 4, pp. 337–341, 2004.
- [97] M. Brylinski, “Aromatic interactions at the ligand–protein interface: Implications for the development of docking scoring functions,” *Chem. Biol. Drug Des.*, vol. 91, no. 2, pp. 380–390, 2018.
- [98] R. Anjana *et al.*, “Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance,” *Bioinformation*, vol. 8, no. 24, pp. 1220–1224, 2012.
- [99] E. Arunan *et al.*, “Definition of the hydrogen bond (IUPAC Recommendations 2011),” *Pure Appl. Chem.*, vol. 83, no. 8, pp. 1637–1641, 2011.
- [100] A. Sarkar and G. Kellogg, “Hydrophobicity - Shake Flasks, Protein Folding and Drug Discovery,” *Curr. Top. Med. Chem.*, vol. 10, no. 1, pp. 67–83, 2010.
- [101] R. Nelson *et al.*, “Structure of the cross- β spine of amyloid-like fibrils,” *Nature*, vol. 435, no. 7043, pp. 773–778, 2005.
- [102] M. Y. Lobanov, N. S. Bogatyreva, and O. V. Galzitskaya, “Radius of gyration as an indicator of protein structure compactness,” *Mol. Biol.*, vol. 42, no. 4, pp. 623–628, 2008.
- [103] F. N. Cele, H. Kumalo, and M. E. S. Soliman, “Mechanism of Inhibition of Hsp90 Dimerization by Gyrase B Inhibitor Coumermycin A1 (C–A1) Revealed by Molecular Dynamics Simulations and Thermodynamic Calculations,” *Cell Biochem. Biophys.*, vol. 74, no. 3, pp. 353–363, 2016.
- [104] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, “Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions,” *J. Phys. Chem. B*, vol. 113, no. 18, pp. 6378–6396, 2009.
- [105] M. Feig and C. L. Brooks, “Recent advances in the development and application of implicit solvent models in biomolecule simulations,” *Curr. Opin. Struct. Biol.*, vol. 14, no. 2, pp. 217–224, 2004.
- [106] H. Pajouhesh and G. R. Lenz, “Medicinal chemical properties of successful central nervous system drugs,” *NeuroRx*, vol. 2, no. 4, pp. 541–553, 2005.
- [107] S. A. Hitchcock and L. D. Pennington, “Structure-brain exposure relationships,” *J. Med. Chem.*, vol. 49, no. 26, pp. 7559–7583, 2006.
- [108] Z. Chen and C. Zhong, “Oxidative stress in Alzheimer’s disease,” *Neurosci. Bull.*, vol. 30, no. 2, pp. 271–281, 2014.
- [109] C. Cheignon, M. Tomas, D. Bonnefont-Rousselot, P. Faller, C. Hureau, and F. Collin, “Oxidative stress and the amyloid beta peptide in Alzheimer’s disease,” *Redox Biol.*, vol. 14, no. October 2017, pp. 450–464, 2018.
- [110] X. Wang, W. Wang, L. Li, G. Perry, H. gon Lee, and X. Zhu, “Oxidative stress and mitochondrial dysfunction in Alzheimer’s disease,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1842, no. 8, pp. 1240–1247, 2014.
- [111] A. Tramutola, C. Lanzillotta, M. Perluigi, and D. A. Butterfield, “Oxidative stress, protein modification and Alzheimer disease,” *Brain Res. Bull.*, vol. 133, pp. 88–96, 2017.

CHAPTER 6 Appendix

6.1 HTVS interaction diagrams

6.1.1 Docking Scores: Heatmap



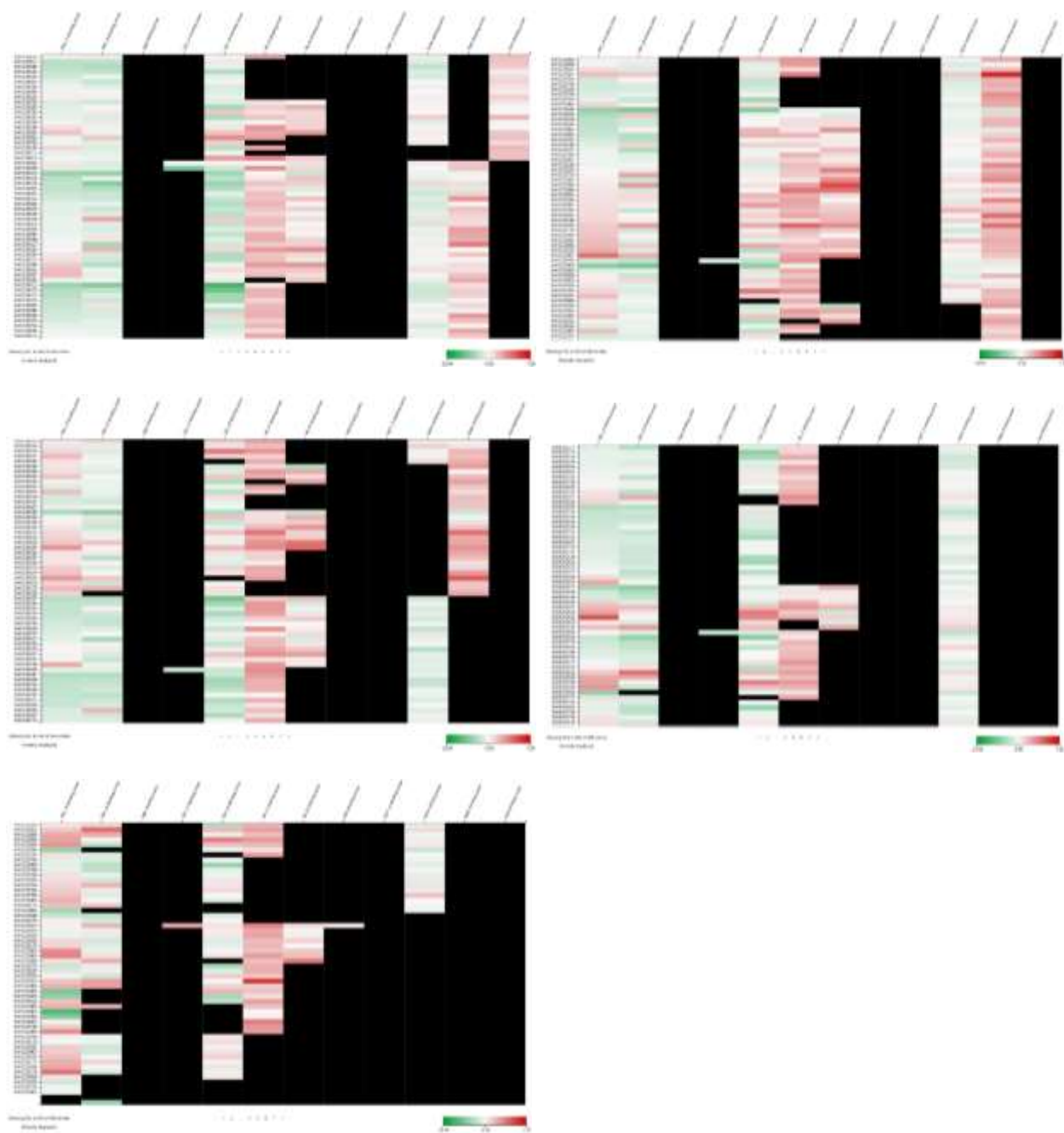


Figure 114: Heatmap

6.2 Additional SANCDB HTVS Docking Detail

6.2.1 Protein-ligand complex interaction diagrams

6.2.1.1 *Discovery Studio*

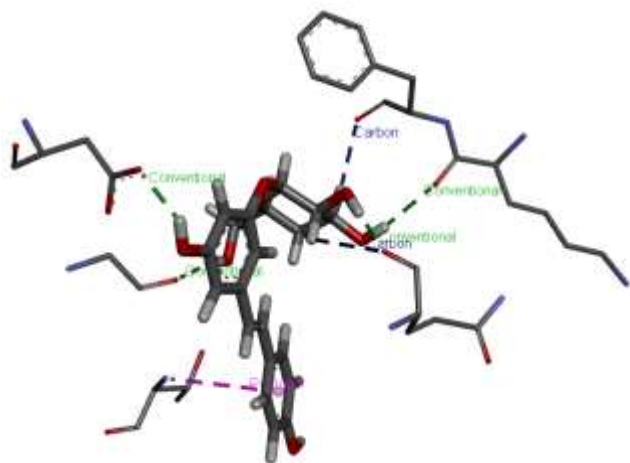


Figure 115: 1FKN_B (SANC00562)

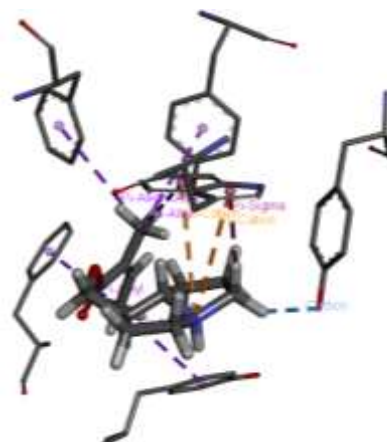


Figure 116: 1GQR (SANC00111)

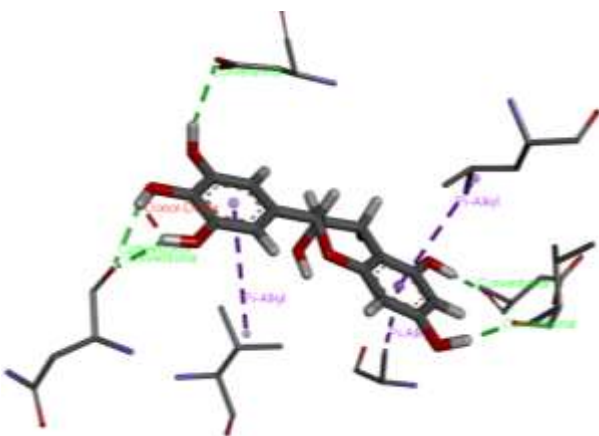


Figure 117: 1J1C_A (SANC00103)

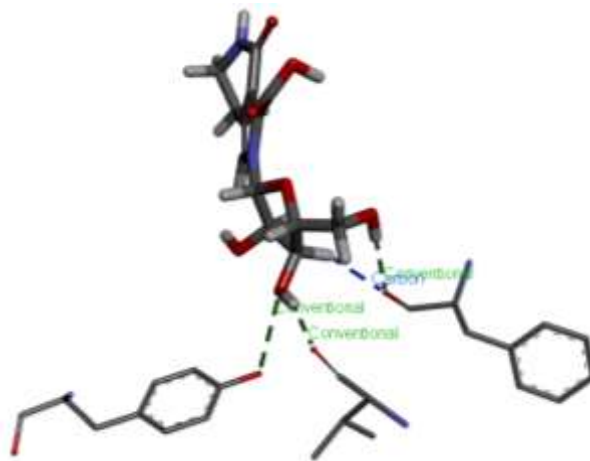


Figure 118: 3LII_A (SANC00269)

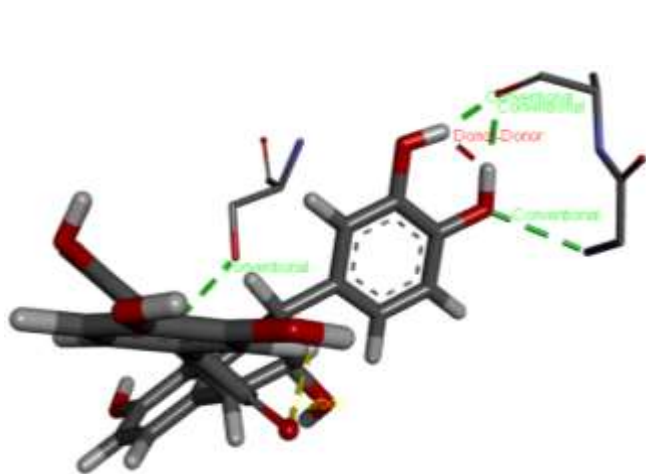


Figure 119: 3LII_B (SANC00559)

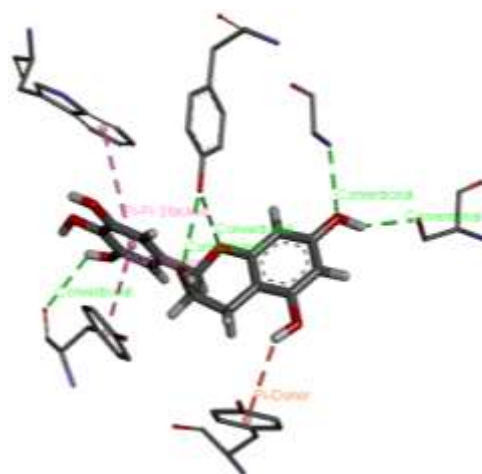


Figure 120: 4EY7_A (SANC00103)

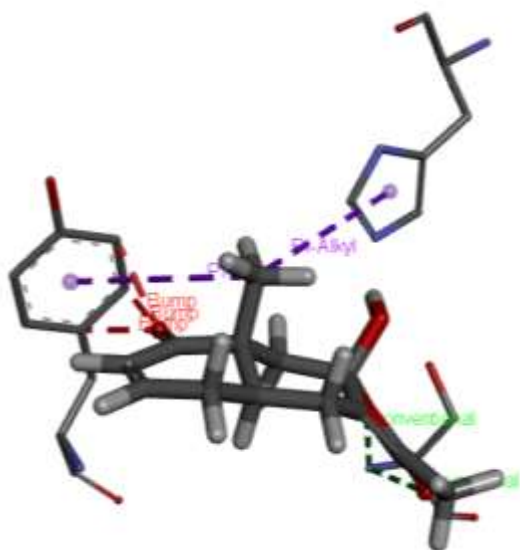


Figure 121: 5E2W (SANC00349)

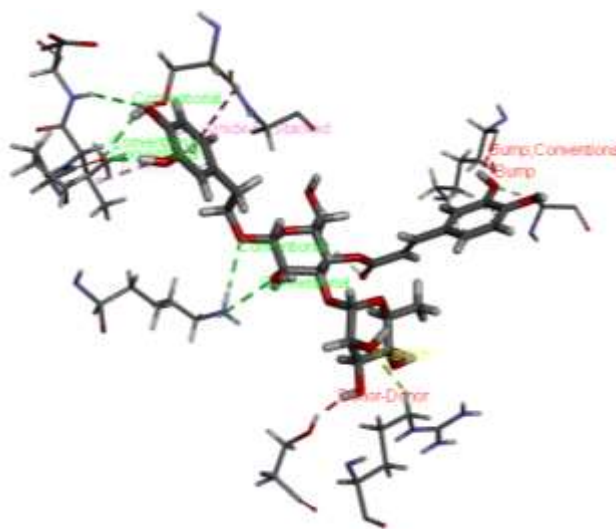


Figure 122: 6C2I (SANC00370)

6.2.1.2 Maestro

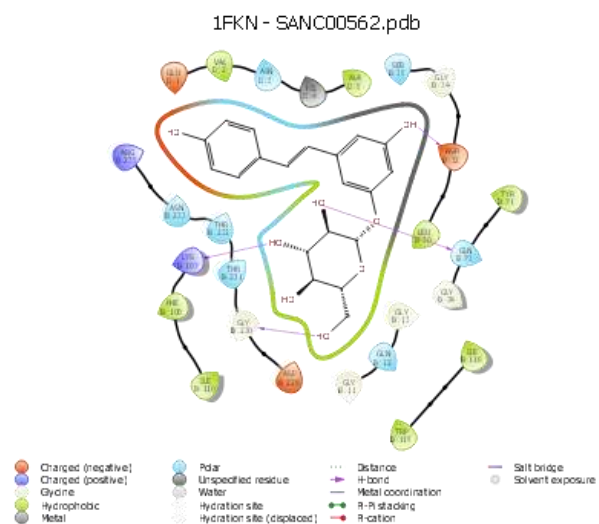


Figure 123: 1FKN_B (SANC00562)

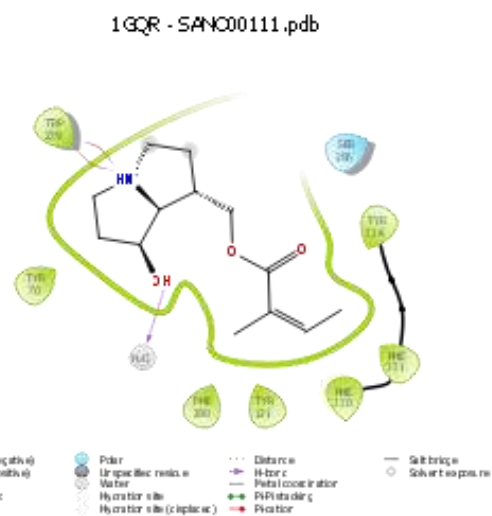


Figure 124: 1GQR (SANC00111)

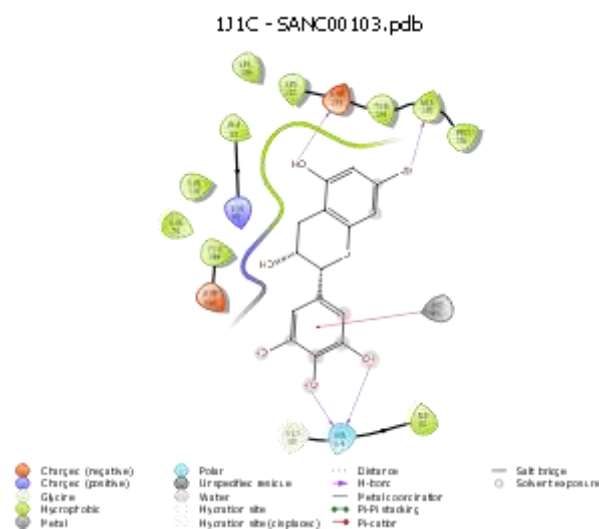


Figure 125: 1J1C_A (SANC00103)

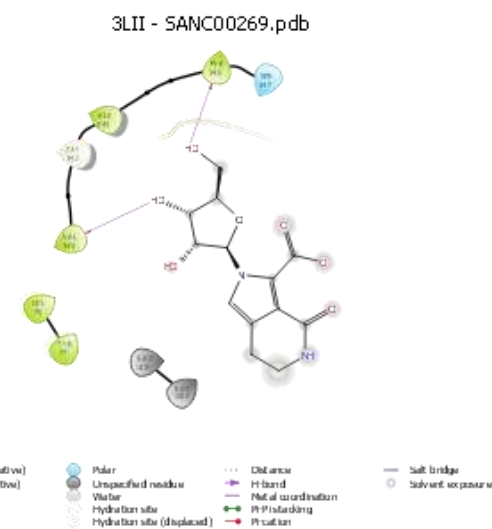


Figure 126: 3LII_A (SANC00269)

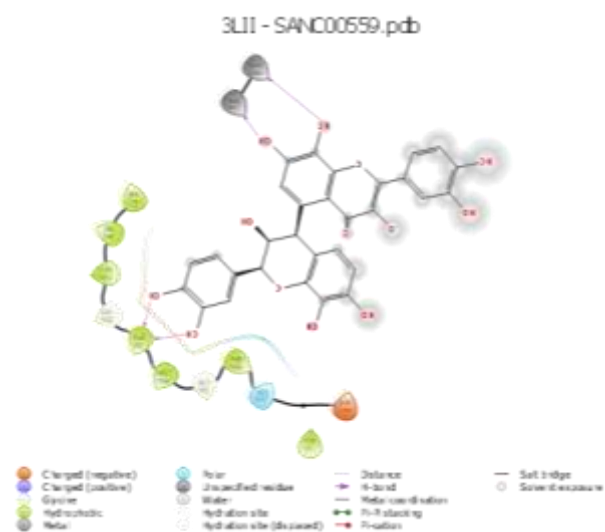


Figure 127: 3LII_B (SANC00559)

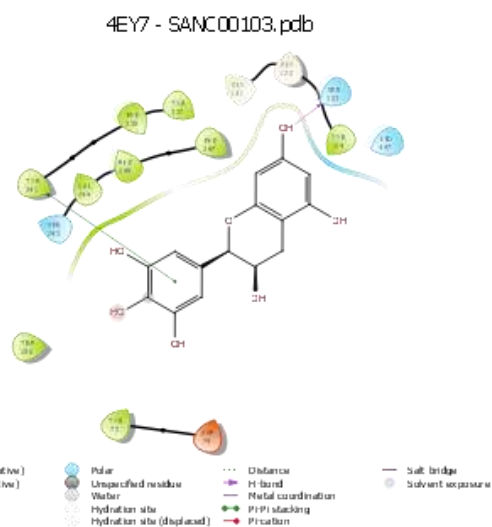


Figure 128: 4EY7_A (SANC00103)

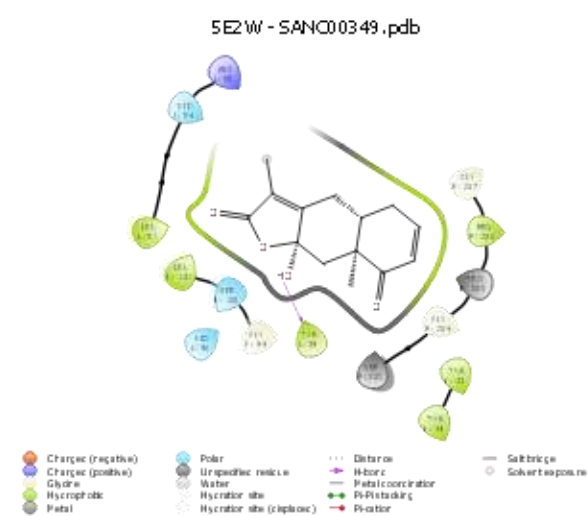


Figure 129: 5E2W (SANC00349)

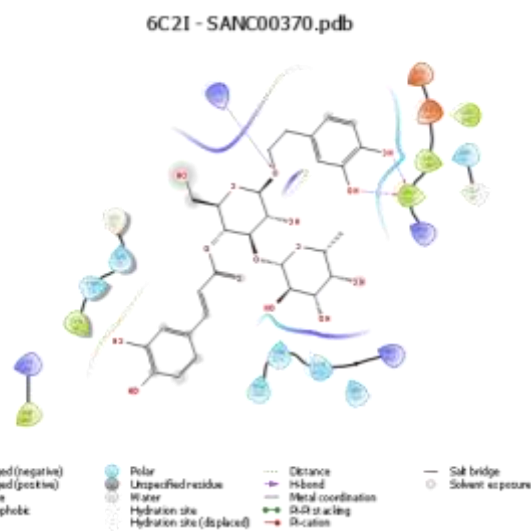


Figure 130: 6C2I (SANC00370)

6.2.2 H-Bonds

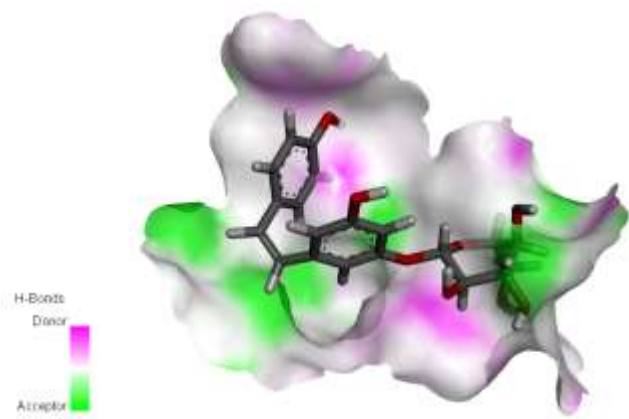


Figure 131: 1FKN_B (SANC00562)

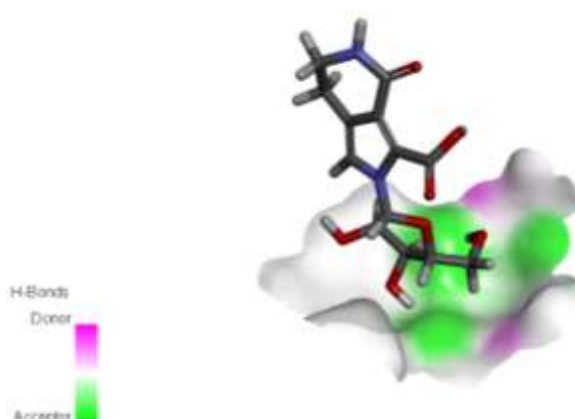


Figure 132: 3LII_A (SANC00269)

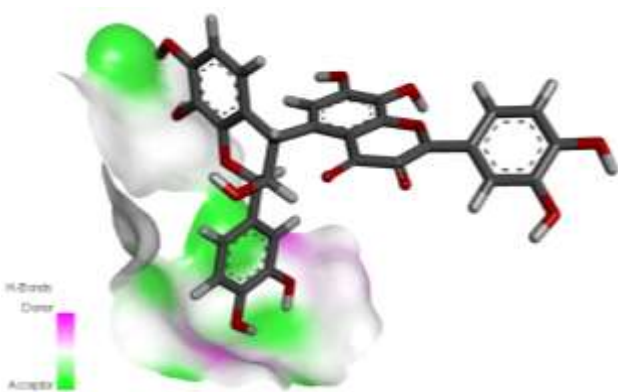


Figure 133: 3LII_B (SANC00559)

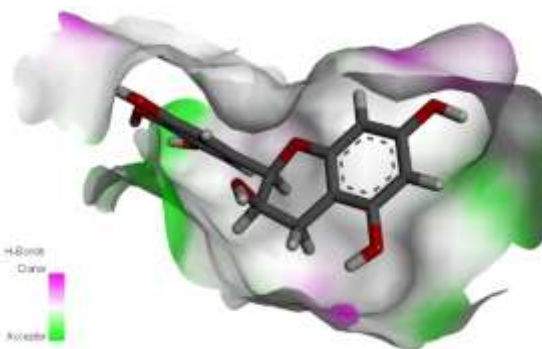


Figure 134: 4EY7_A (SANC00103)

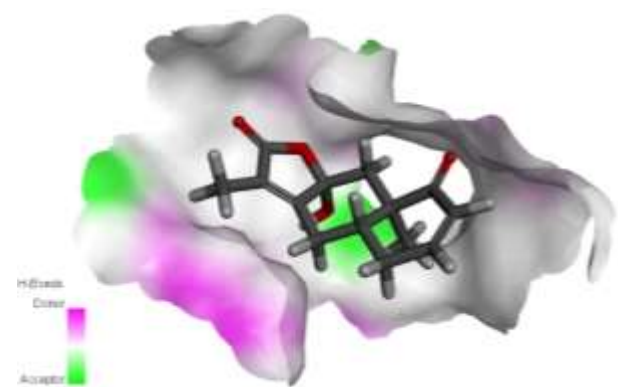


Figure 135: 5E2W (SANC00349)

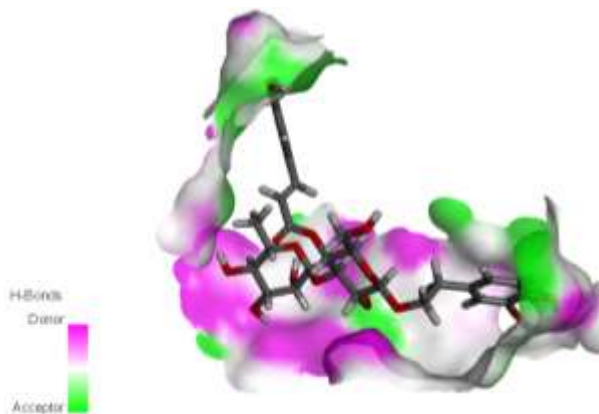


Figure 136: 6C2I (SANC00370)

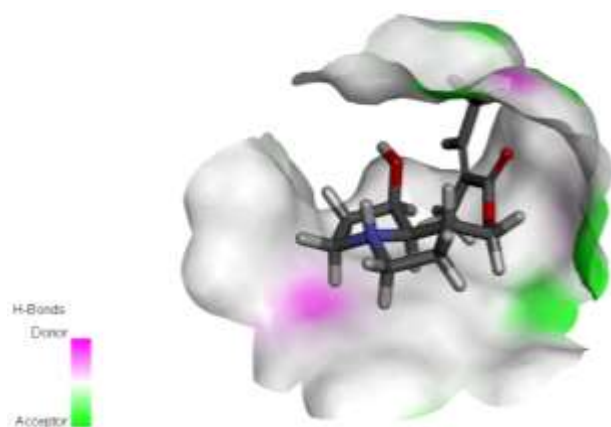


Figure 137: 1GQR (SANC00111)

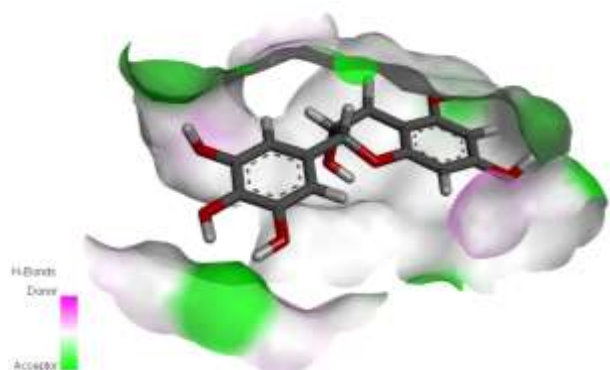


Figure 138: 1J1C_A (SANC00103)

6.2.3 Hydrophobicity

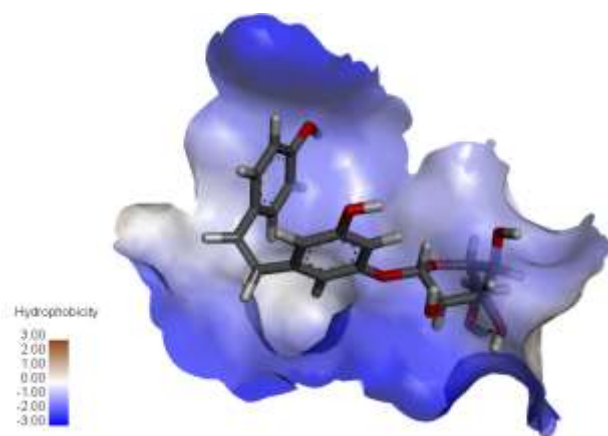


Figure 139: 1FKN_B (SANC00562)

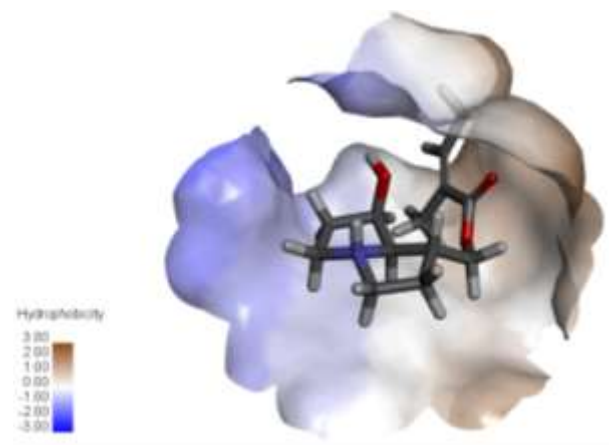


Figure 140: 1GQR (SANC00111)

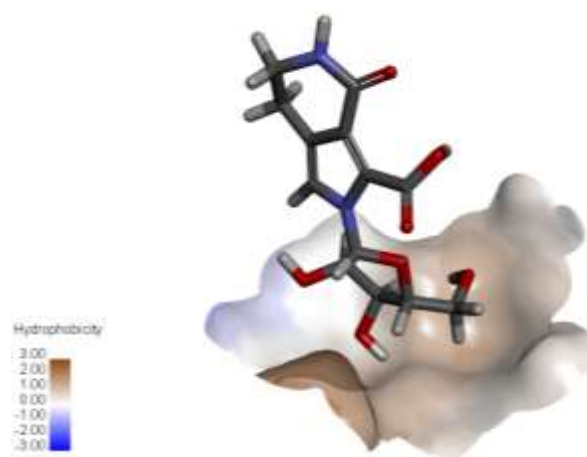


Figure 141: 3LII_A (SANC00269)

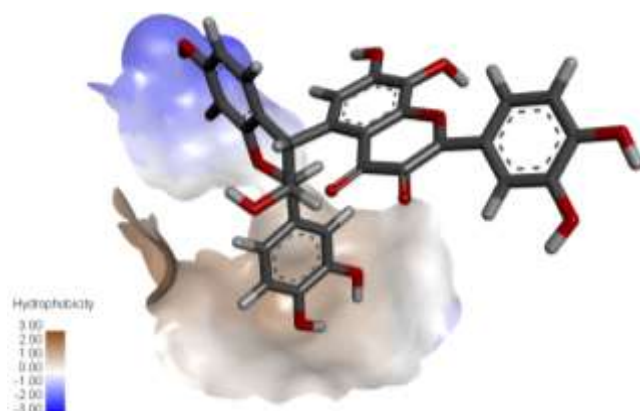


Figure 142: 3LII_B (SANC00559)

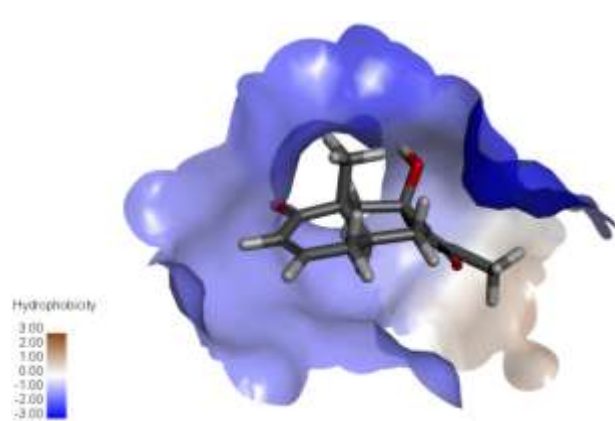


Figure 143: 5E2W (SANC00349)

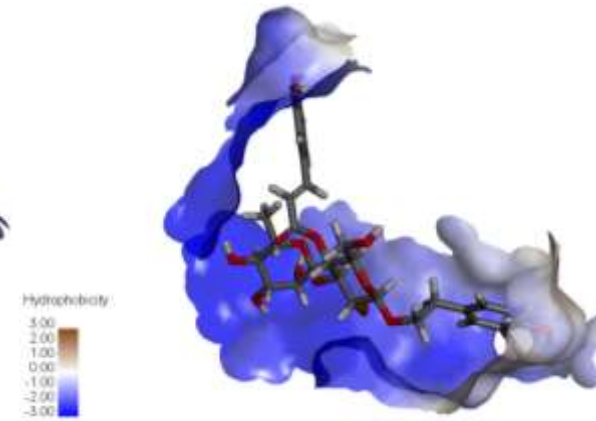


Figure 144: 6C2I (SANC00370)

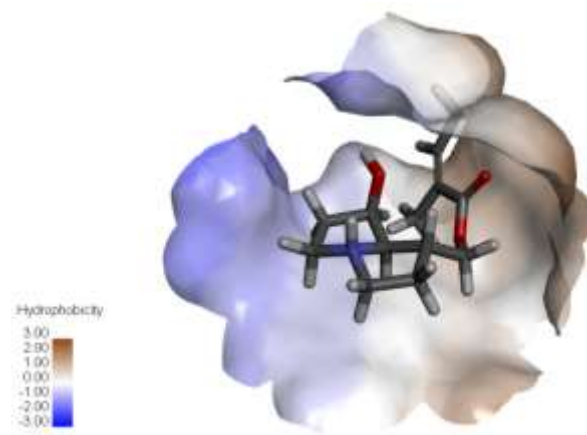


Figure 145: 1GQR

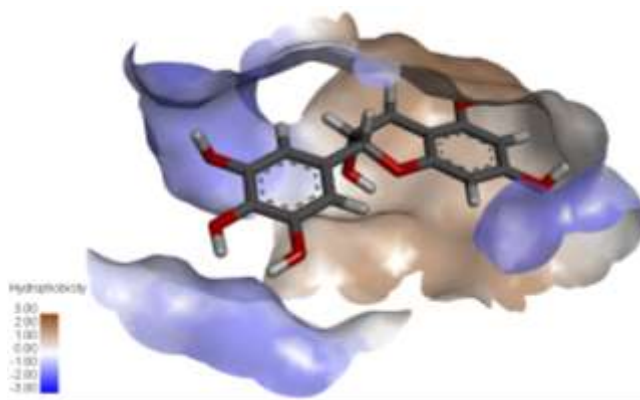


Figure 146: 1J1C_A (SANC00103)

6.2.4 Aromaticity

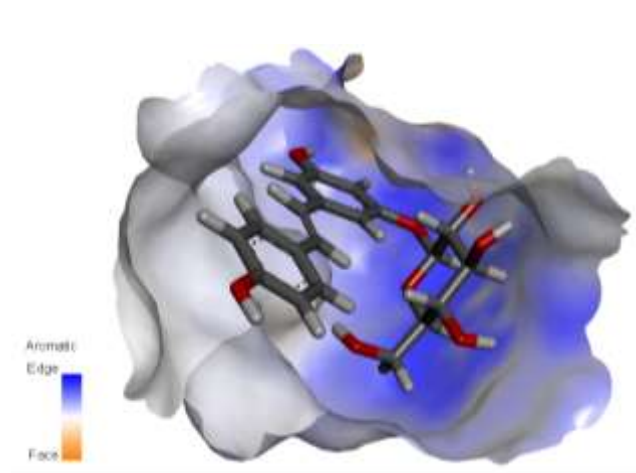


Figure 147: 1FKN_B (SANC00562)

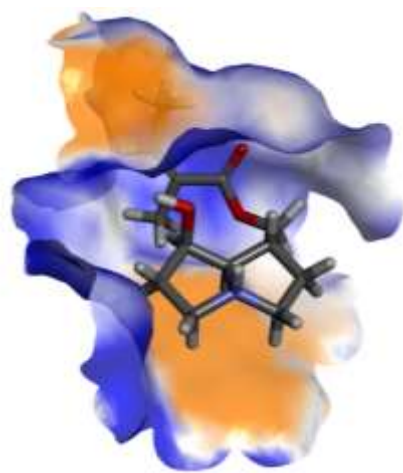


Figure 148: 1GQR (SANC00111)

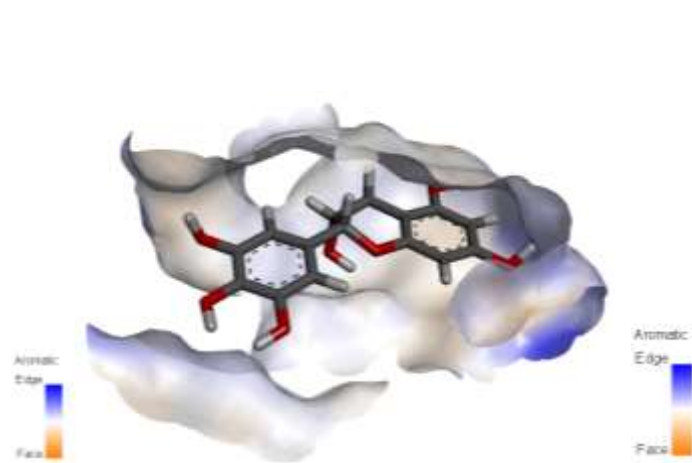


Figure 149: 1J1C_A (SANC00103)

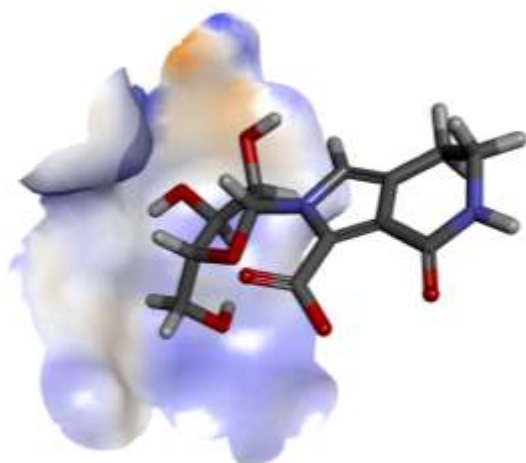


Figure 150: 3LII_A (SANC00269)

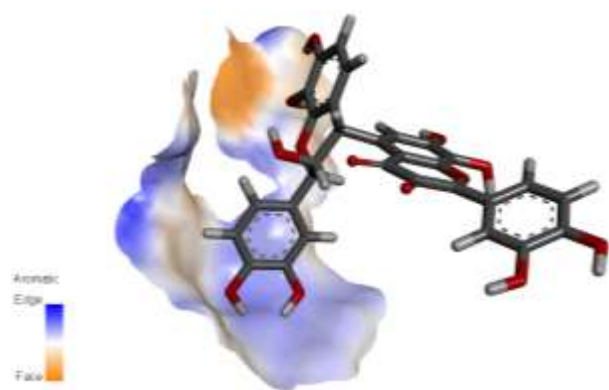


Figure 151: 3LII_B (SANC00559)

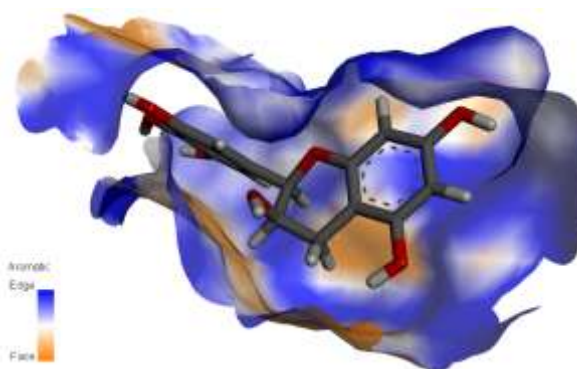


Figure 152: 4EY7_A (SANC00103)

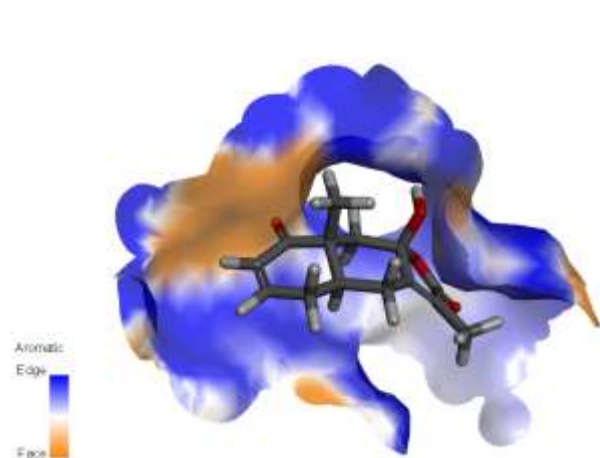


Figure 153: 5E2W (SANC00349)

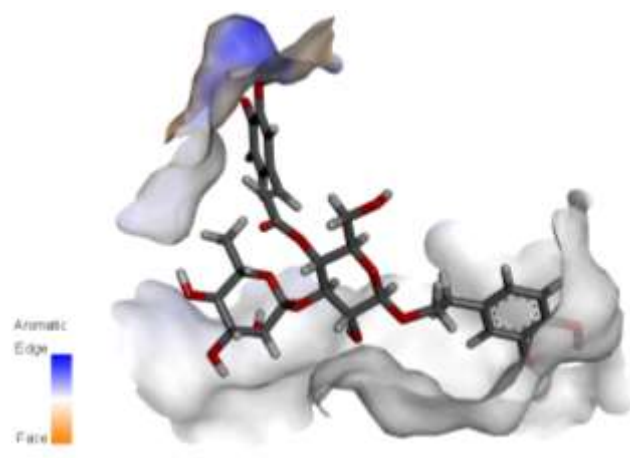


Figure 154: 6C2I (SANC00370)

6.2.5 SASA

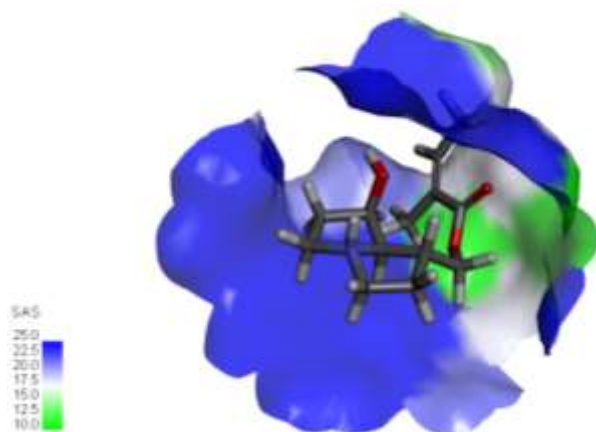


Figure 155: 1GQR (SANC00111)

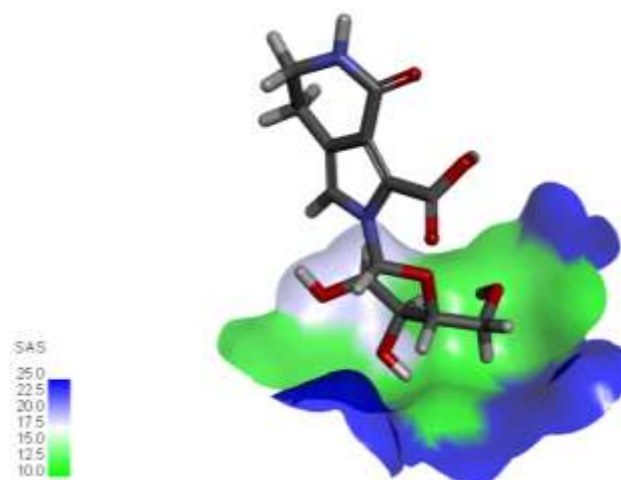


Figure 156: 3LII_A (SANC00269)

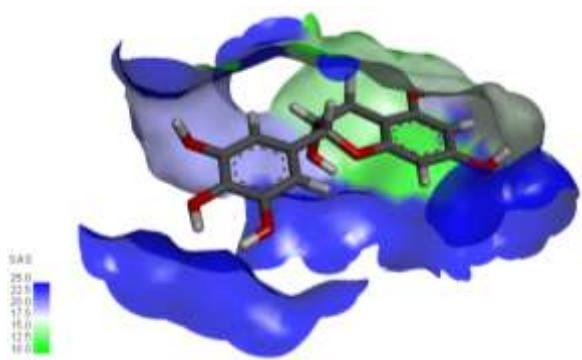


Figure 157: 1J1C_A (SANC00103)

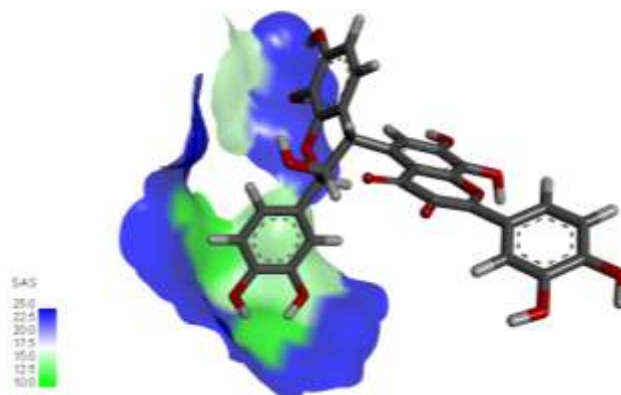


Figure 158: 3LII_B (SANC00559)

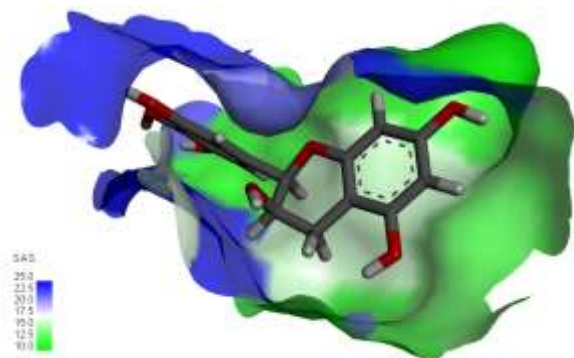


Figure 159: 4EY7_A (SANC00103)

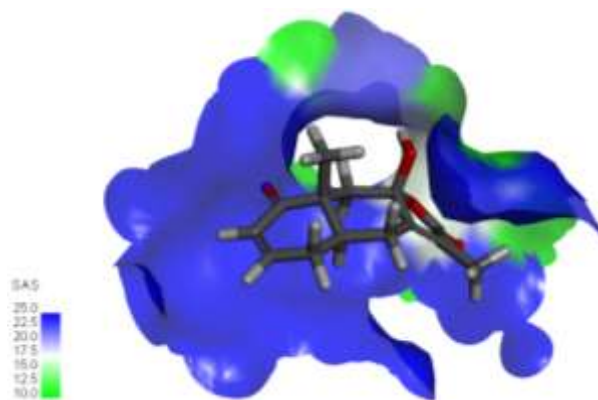


Figure 160: 5E2W (SANC00349)

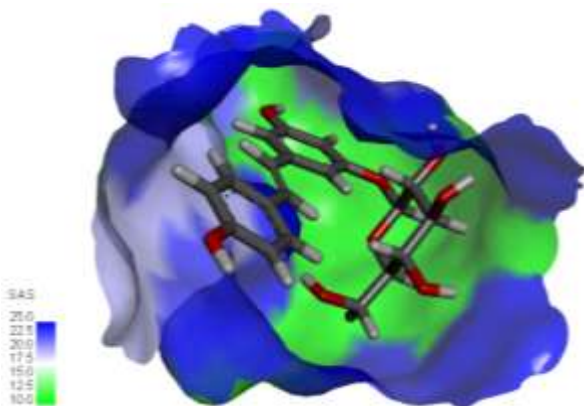


Figure 161: 1FKN_B (SANC00562)

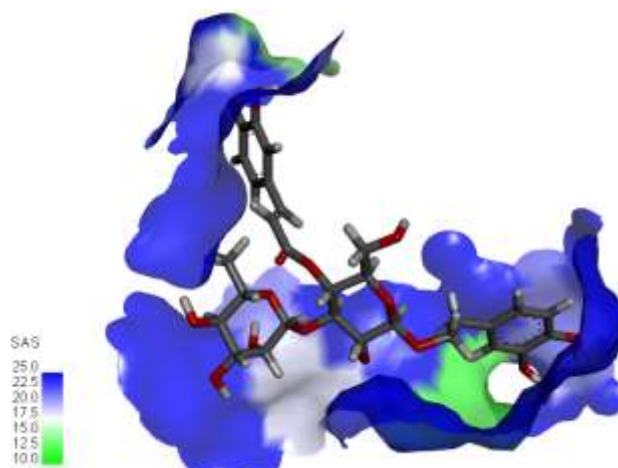


Figure 162: 6C2I (SANC00370)

6.3 Molecular Dynamics Interaction Diagrams (50 ns)

6.3.1 Protein-Ligand RMSD

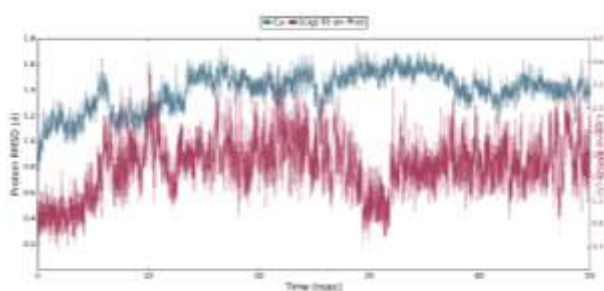


Figure 163: 1FKN_B (SANC00562)

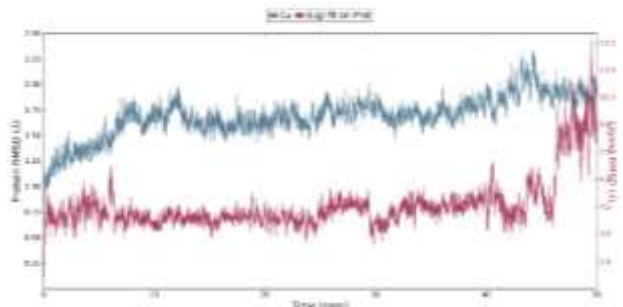


Figure 164: 3LII_A (SANC00269)

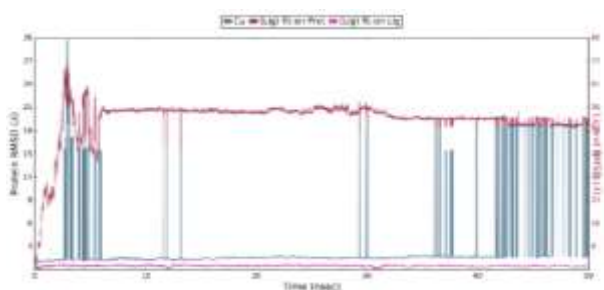


Figure 165: 3LII_B (SANC00559)

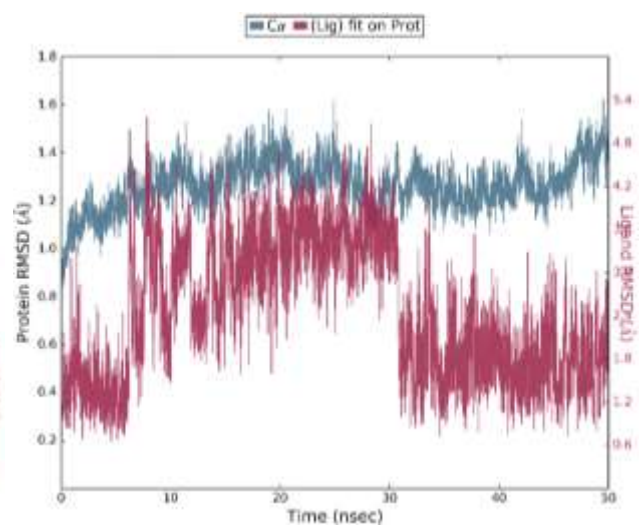


Figure 166: 1GQR (SANC00111)

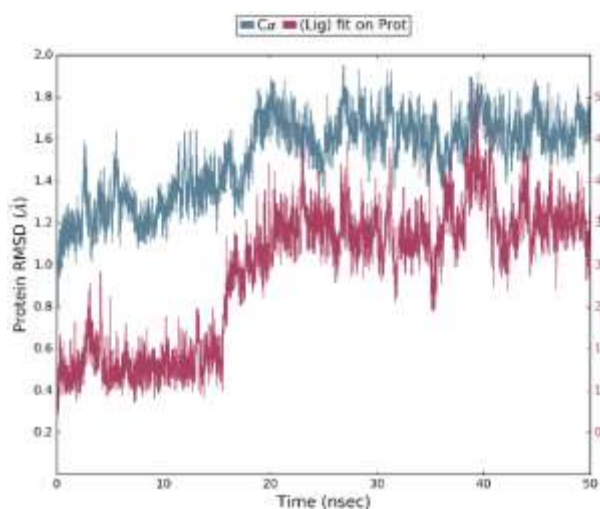


Figure 167: 6C2I (SANC00370)

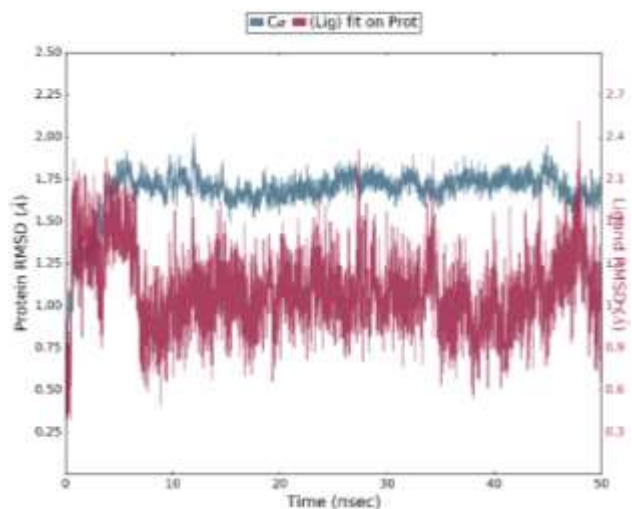


Figure 168: 4EY7_A (SANC00103)

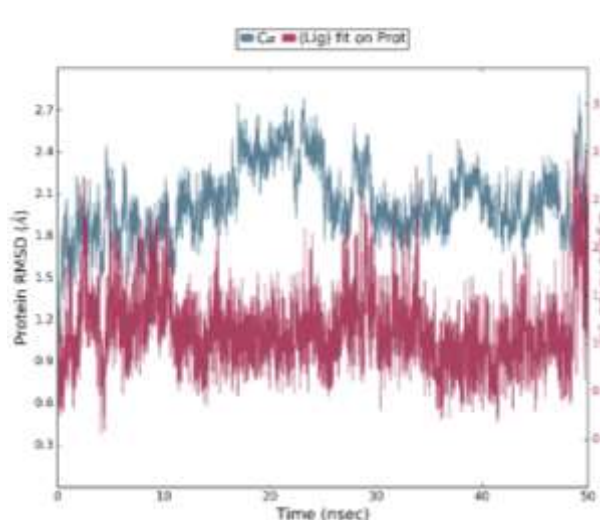


Figure 169: 1J1C_A (SANC00103)

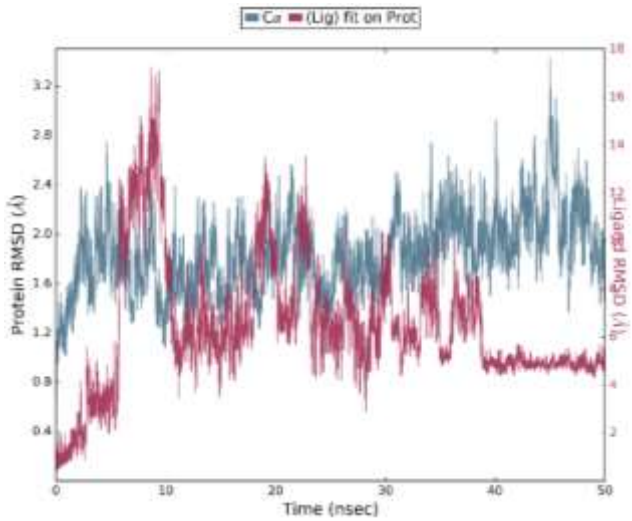


Figure 170: 5E2W (SANC00349)

6.3.2 Protein RMSF

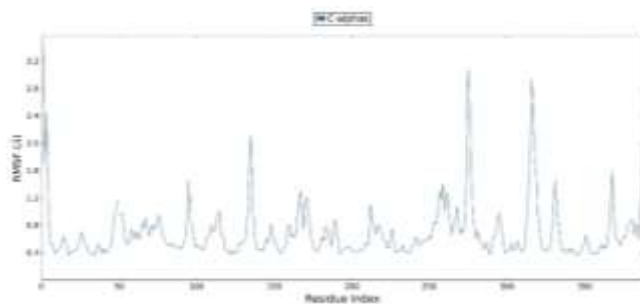


Figure 171: 1FKN_B (SANC00562)

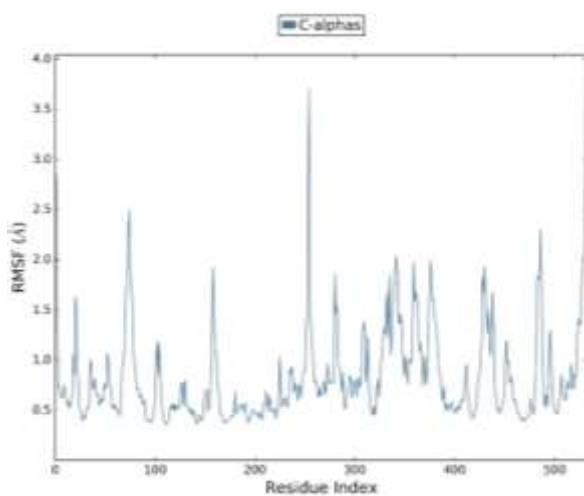


Figure 172: 3LII_A (SANC00269)

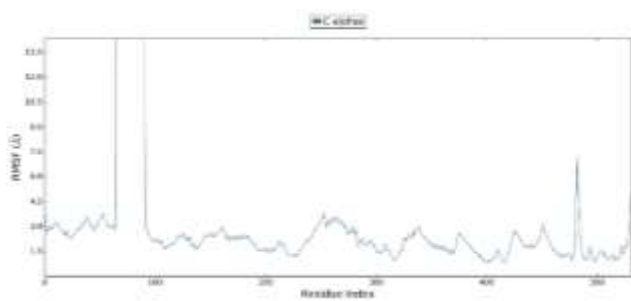


Figure 173: 3LII_B (SANC00559)

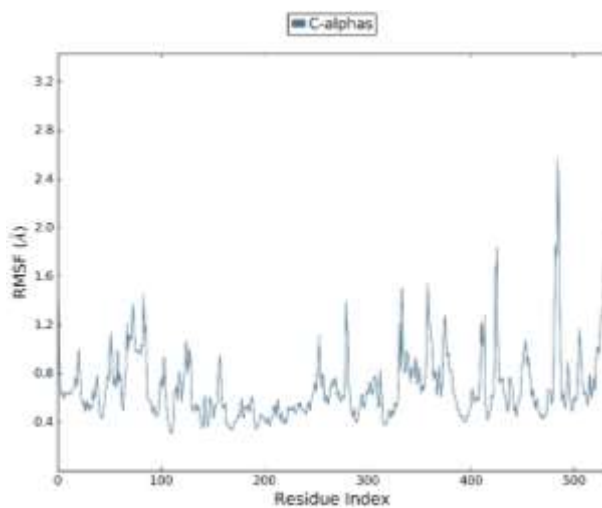


Figure 174: 1GQR (SANC00111)

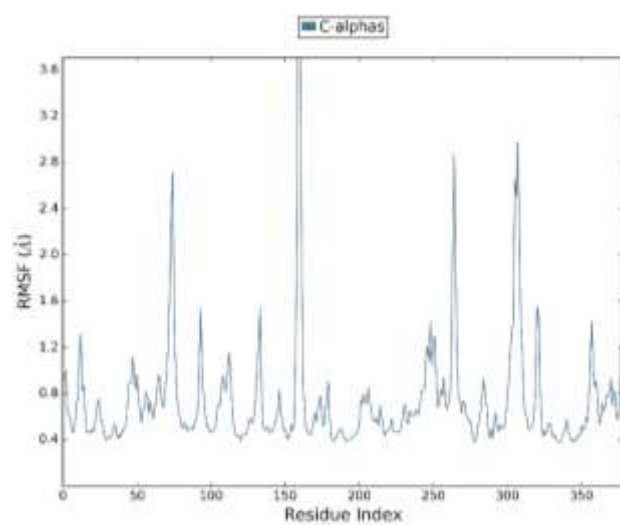


Figure 175: 6C2I (SANC00370)

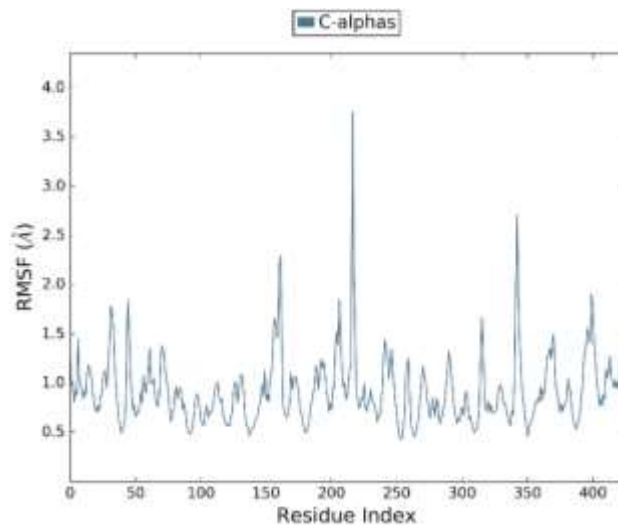


Figure 176: 5E2W (SANC00349)

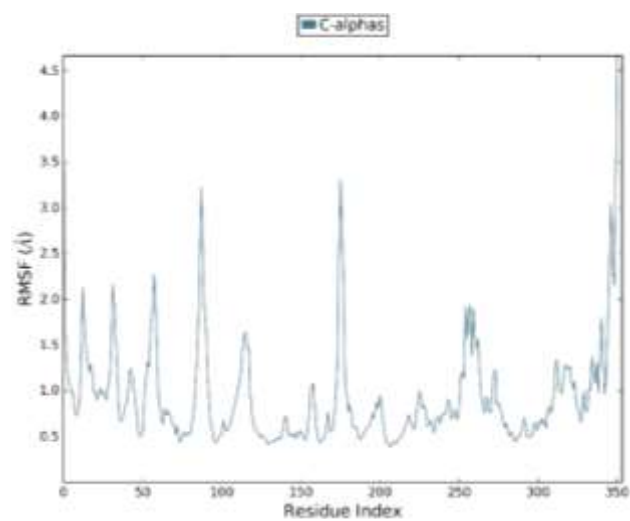


Figure 177: 1J1C_A (SANC00103)

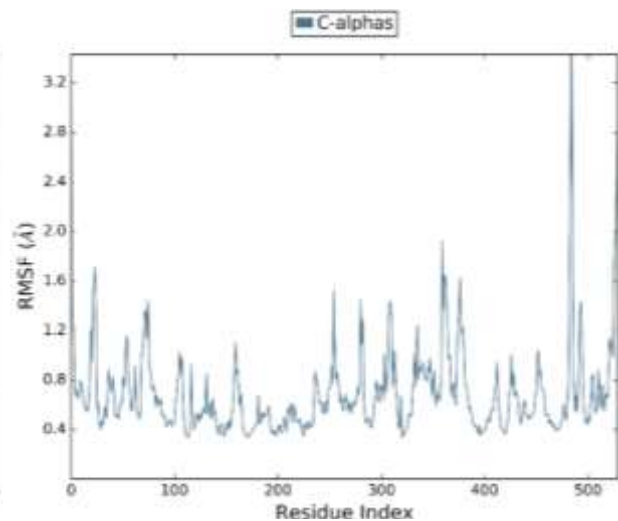


Figure 178: 4EY7_A (SANC00103)

6.3.3 Protein SSE Histogram

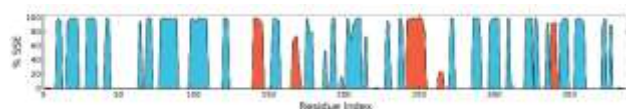


Figure 179: 1FKN_B (SANC00562)

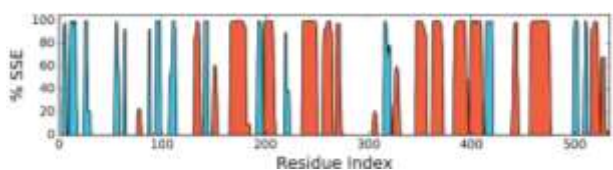


Figure 180: 3LII_A (SANC00269)

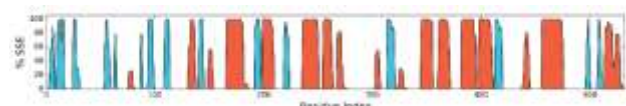


Figure 181: 3LII_B (SANC00559)

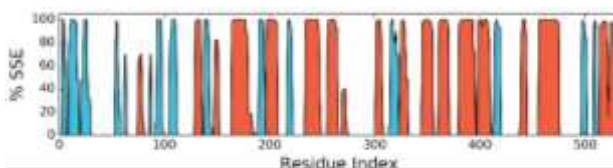


Figure 182: 1GQR (SANC00111)

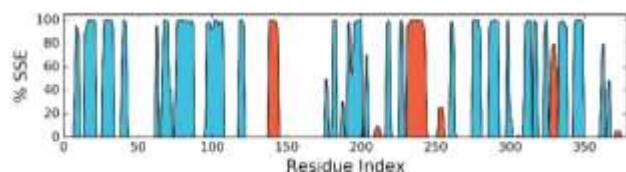


Figure 183: 6C2I (SANC00370)

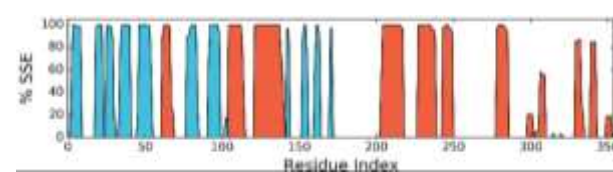


Figure 184: 1J1C_A (SANC00103)

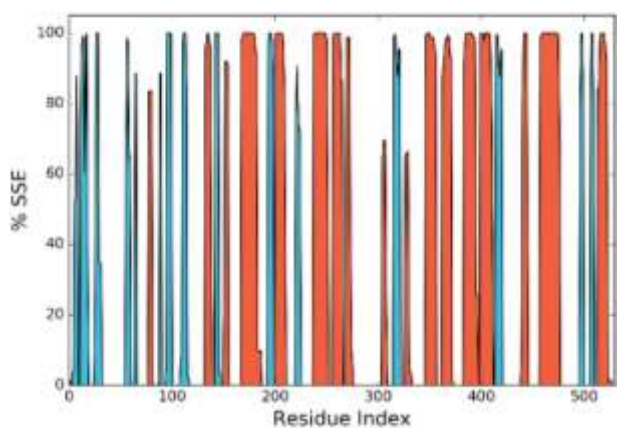


Figure 185: 4EY7_A (SANC00103)

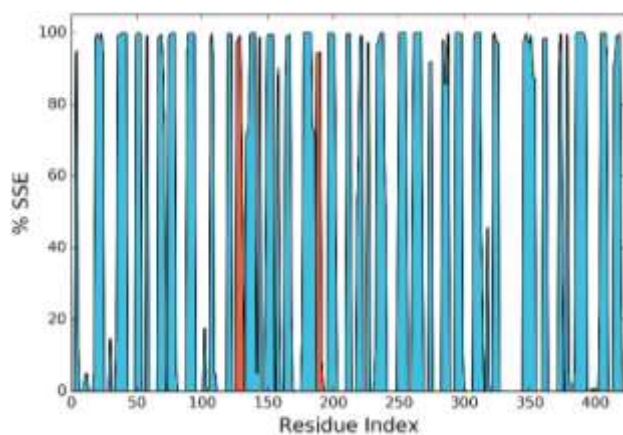


Figure 186: 5E2W (SANC00349)

6.3.4 Protein SSE timeline

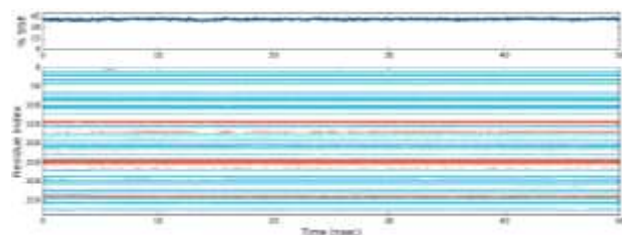


Figure 187: 1FKN_B (SANC00562)

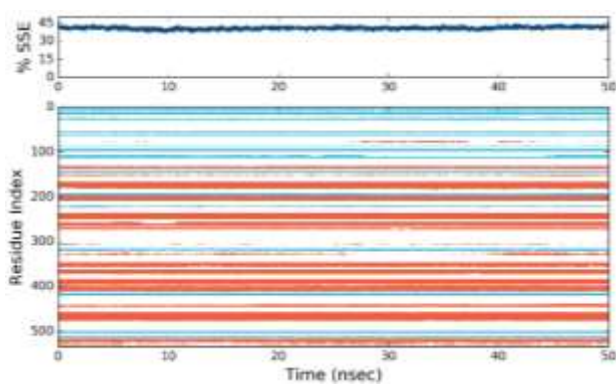


Figure 188: 3LII_A (SANC00269)

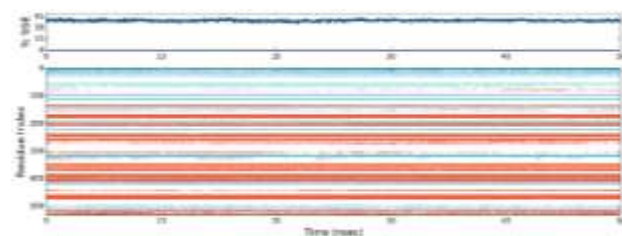


Figure 189: 3LII_B (SANC00559)

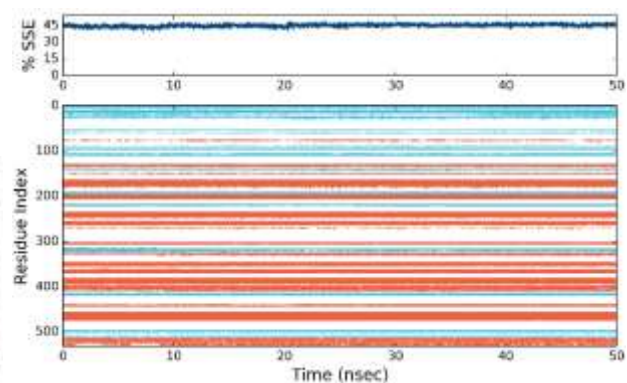


Figure 190: 1GQR (SANC00111)

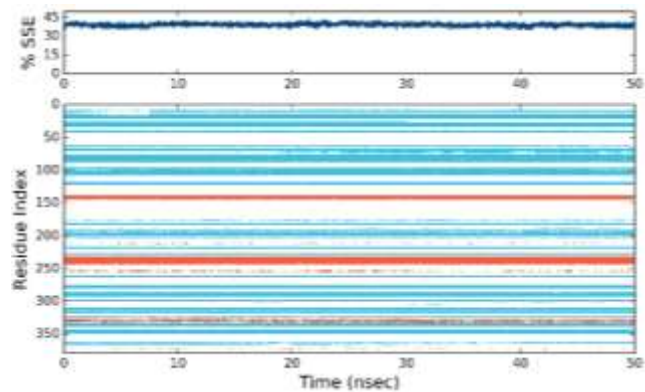


Figure 191: 6C2I (SANC00370)

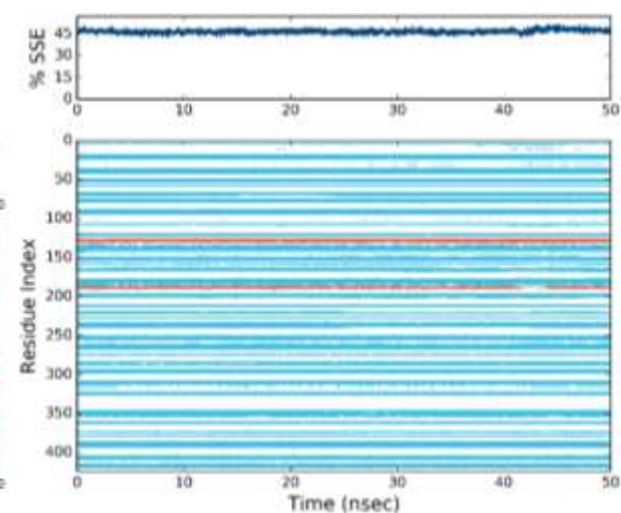


Figure 192: 5E2W (SANC00349)

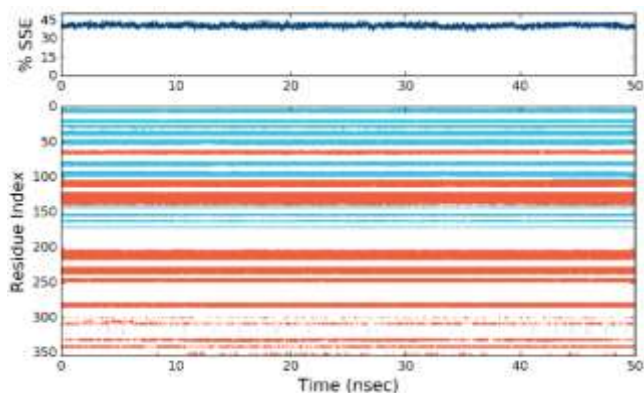


Figure 193: 1J1C_A (SANC00103)

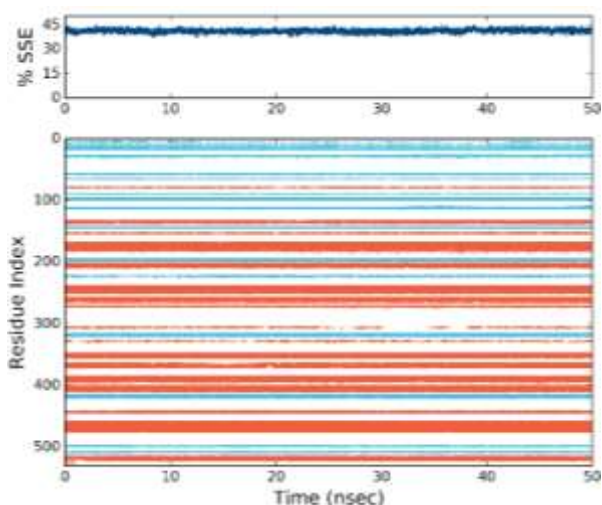


Figure 194: 4EY7_A (SANC00103)

6.3.5 Ligand RMSF

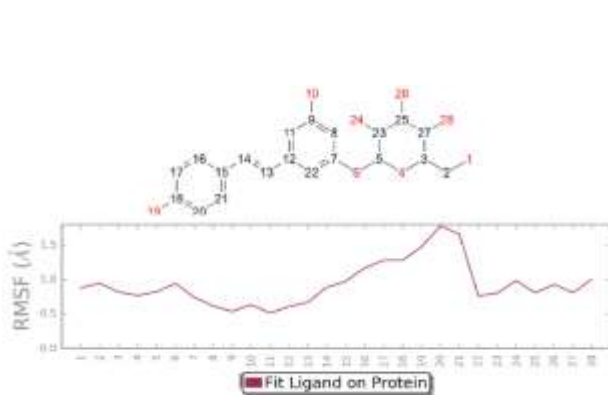


Figure 195: 1FKN_B (SANC00562)

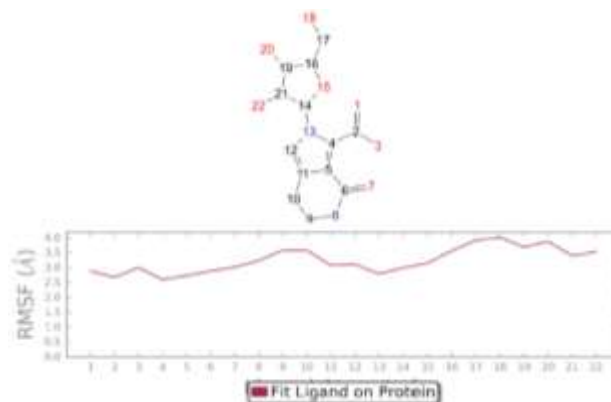


Figure 196: 3LII_A (SANC00269)



Figure 197: 3LII_B (SANC00559)

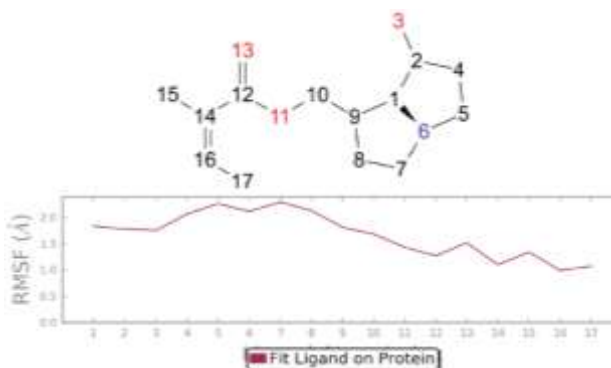


Figure 198: 1GQR (SANC00111)

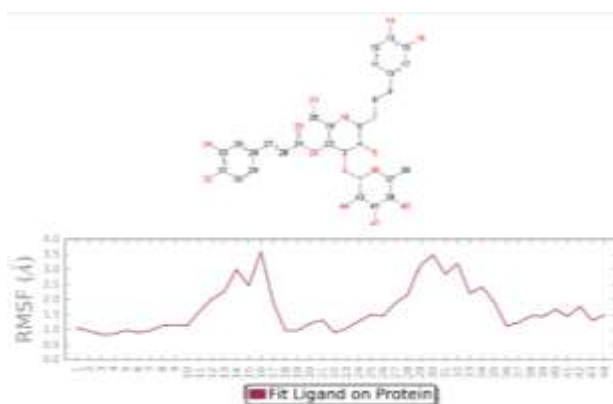


Figure 199: 6C2I (SANC00370)



Figure 200: 4EY7_A (SANC00103)

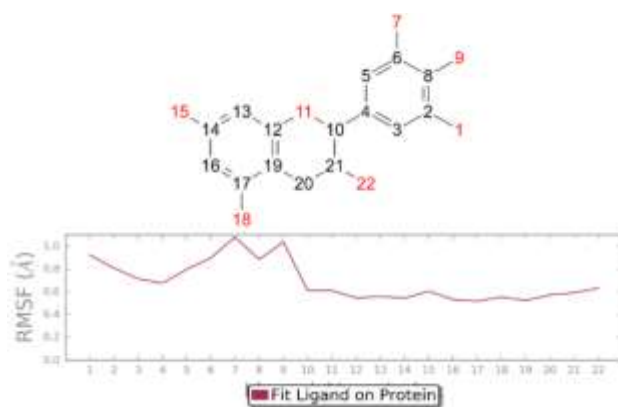


Figure 201: 1J1C_A (SANC00103)

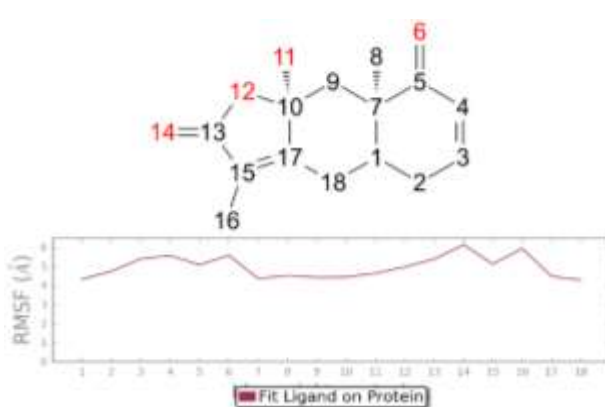


Figure 202: 5E2W (SANC00349)

6.3.6 Ligand-Protein Contacts 2D summary

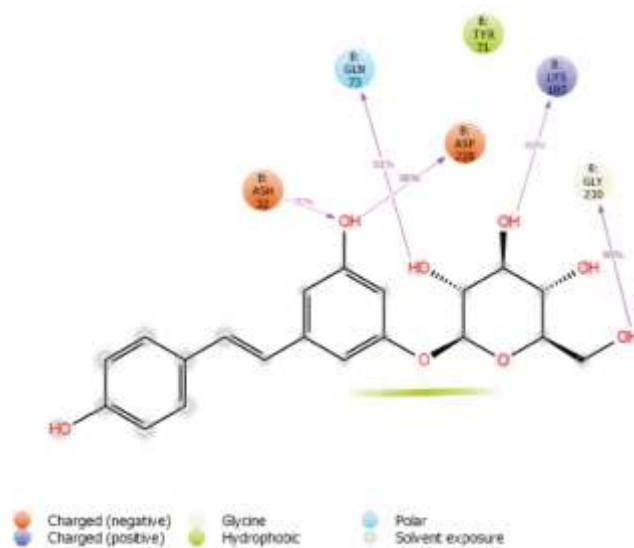


Figure 203: 1FKN_B (SANC00562)

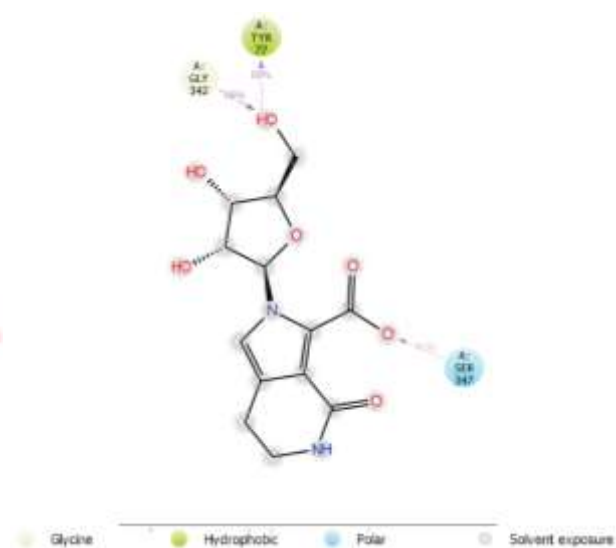


Figure 204: 3LII_A (SANC00269)

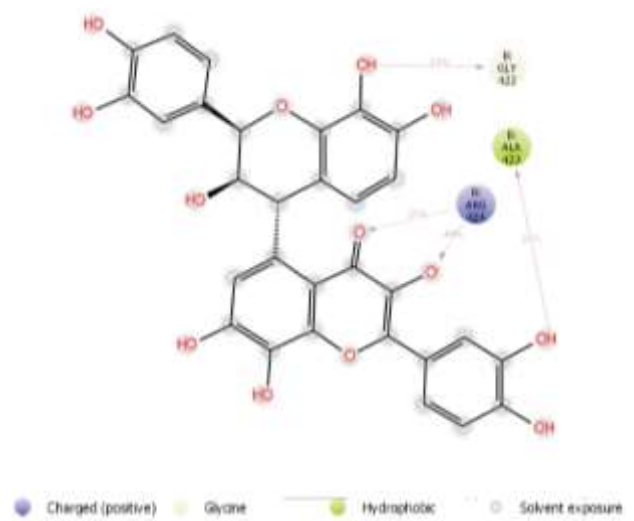


Figure 205: 3LII_B (SANC00559)

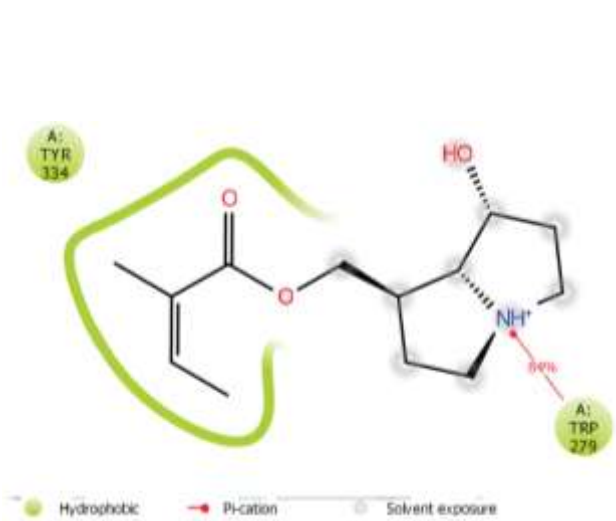


Figure 206: 1GQR (SANC00111)

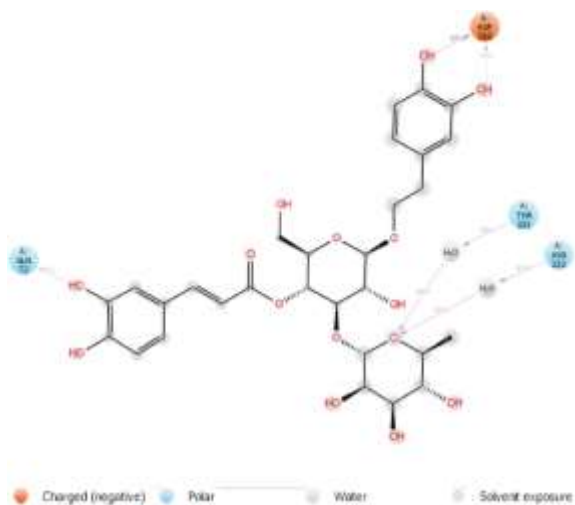


Figure 207: 6C2I (SANC00370)

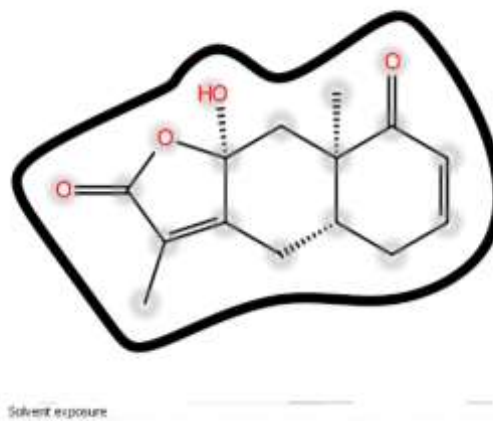


Figure 208: 5E2W (SANC00349)

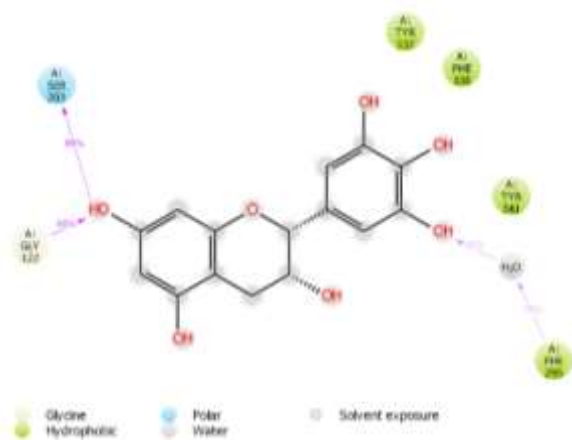


Figure 209: 4EY7_A (SANC00103)

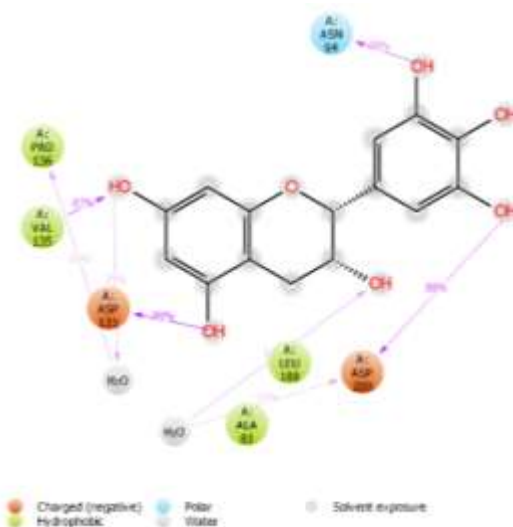


Figure 210: 1J1C_A (SANC00103)

Stacked bar chart showing the fraction of interactions for various residues. The y-axis is 'Interactions Fraction' (0.0 to 1.2). The x-axis lists residues: Ser, Asp, Glu, Lys, Thr, Ala, Val, Phe, Trp, Gly, Asn, Gln, Arg, Leu, Ile, Met, Pro, His, Tyr, and Cys. The legend indicates: H-bonds (green), Hydrophobic (purple), Ionic (pink), and Water bridges (blue).

Residue	H-bonds	Hydrophobic	Ionic	Water bridges
Ser	0.02	0.00	0.00	0.00
Asp	0.02	0.00	0.00	0.00
Glu	0.02	0.00	0.00	0.00
Lys	0.02	0.00	0.00	0.00
Thr	0.02	0.00	0.00	0.00
Ala	0.02	0.00	0.00	0.00
Val	0.02	0.00	0.00	0.00
Phe	0.02	0.00	0.00	0.00
Trp	0.02	0.00	0.00	0.00
Gly	0.02	0.00	0.00	0.00
Asn	0.02	0.00	0.00	0.00
Gln	0.02	0.00	0.00	0.00
Arg	0.02	0.00	0.00	0.00
Leu	0.02	0.00	0.00	0.00
Ile	0.02	0.00	0.00	0.00
Met	0.02	0.00	0.00	0.00
Pro	0.02	0.00	0.00	0.00
His	0.02	0.00	0.00	0.00
Tyr	0.02	0.00	0.00	0.00
Cys	0.02	0.00	0.00	0.00

[illegible]

Stacked bar chart showing the interaction fraction for various residues. The y-axis is 'Interactions Fraction' (0.0 to 1.4). The x-axis lists residues: Arg, Asn, Asp, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val, Ala, Cys, Gln, and Pro. The legend indicates: H-bonds (green), Hydrophobic (purple), Ionic (red), and Water Bridges (blue).

A bar chart titled "Fraction of interactions" on the y-axis, which ranges from 0.0 to 1.0. The x-axis lists residues: Trp_98, Asp_11, Glu_17, Glu_16, Asn_89, Trp_121, Trp_174, Asp_205, Phe_188, Asn_209, Met_240, Phe_240, Met_315, Trp_316, Glu_316, and Trp_410. A legend at the bottom identifies four interaction types: H-bonds (green), Hydrophobic (purple), Ionic (red), and Water bridges (blue). The bars are stacked: Water bridges (top, blue), Hydrophobic (middle-top, purple), H-bonds (middle-bottom, green), and Ionic (bottom, red). Notable values include Trp_98 having significant fractions in all categories, Trp_174 being almost entirely hydrophobic (~0.95), and Trp_316 being almost entirely hydrophobic (~1.0).

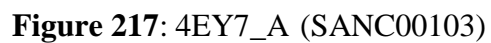
Residue	H-bonds	Hydrophobic	Ionic	Water bridges
Trp_98	0.15	0.15	0.05	0.25
Asp_11	0.25	0.05	0.05	0.05
Glu_17	0.20	0.05	0.05	0.05
Glu_16	0.15	0.05	0.05	0.05
Asn_89	0.00	0.00	0.00	0.00
Trp_121	0.00	0.10	0.00	0.00
Trp_174	0.00	0.95	0.00	0.00
Asp_205	0.00	0.00	0.00	0.00
Phe_188	0.00	0.00	0.00	0.00
Asn_209	0.00	0.00	0.00	0.00
Met_240	0.00	0.00	0.00	0.00
Phe_240	0.00	0.00	0.00	0.00
Met_315	0.00	0.25	0.00	0.00
Trp_316	0.00	1.00	0.00	0.00
Glu_316	0.00	0.00	0.00	0.00
Trp_410	0.00	0.00	0.00	0.00

[illegible]

A bar chart titled "Interactions Fraction" on the y-axis, which ranges from 0.0 to 1.8. The x-axis lists 16 amino acid pairs: EE-EQ, DD-DE, GG-GG, PP-PG, SS-SG, LL-LP, RR-RS, AA-AA, TT-TT, KK-KK, NN-NN, QQ-QQ, YY-YY, CC-CC, FF-FF, and WW-WW. Each pair has up to four bars representing different interaction types: H bonds (green), Hydrophobic (purple), Ionic (pink), and Water bridges (blue). The legend at the bottom identifies these colors.

Residue Pair	H bonds	Hydrophobic	Ionic	Water bridges
EE-EQ	0.0	0.0	0.0	0.0
DD-DE	0.0	0.0	0.0	0.0
GG-GG	0.75	0.0	0.0	0.05
PP-PG	0.0	0.0	0.0	0.1
SS-SG	0.0	0.0	0.0	0.0
LL-LP	0.0	0.0	0.0	0.0
RR-RS	0.0	0.15	0.0	0.0
AA-AA	0.8	0.0	0.0	0.0
TT-TT	0.0	0.0	0.0	0.0
KK-KK	1.0	0.0	0.0	0.0
NN-NN	0.9	0.0	0.0	0.3
QQ-QQ	0.0	0.0	0.0	0.0
YY-YY	0.0	0.0	0.0	0.05
CC-CC	0.0	0.0	0.0	0.0
FF-FF	0.0	0.0	0.0	0.0
WW-WW	0.0	0.55	0.0	0.0
DD-DD	0.0	0.0	0.0	0.0

144



6.3.8 Protein-Ligand Contacts Timeline

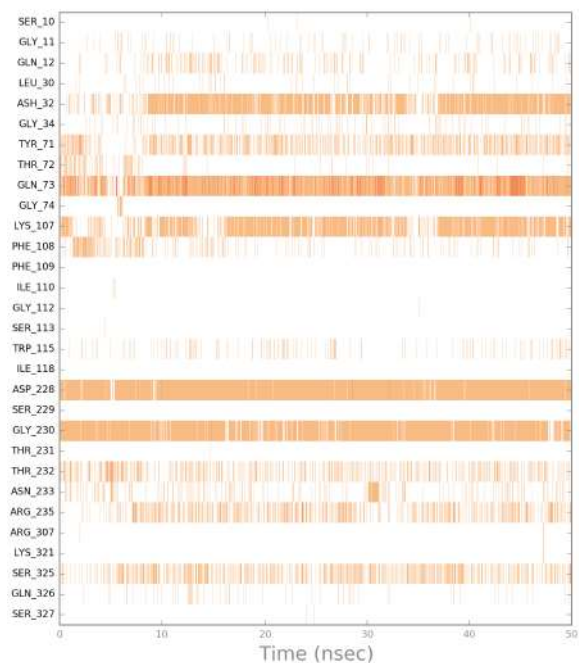


Figure 219: 1FKN_B (SANC00562)





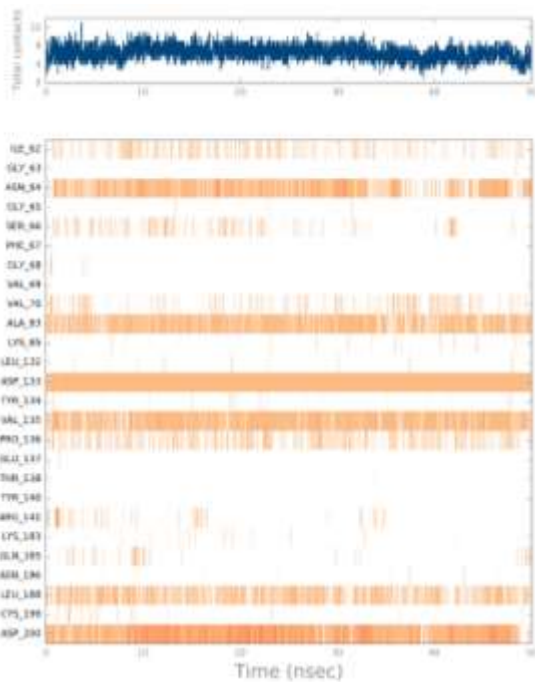


Figure 225: 1J1C_A (SANC00103)

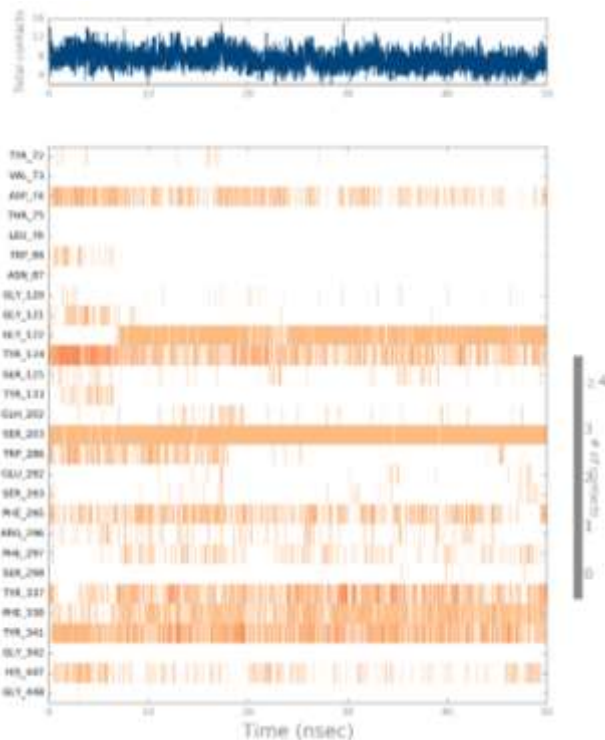


Figure 226: 4EY7_A (SANC00103)

6.3.9 Ligand Torsions

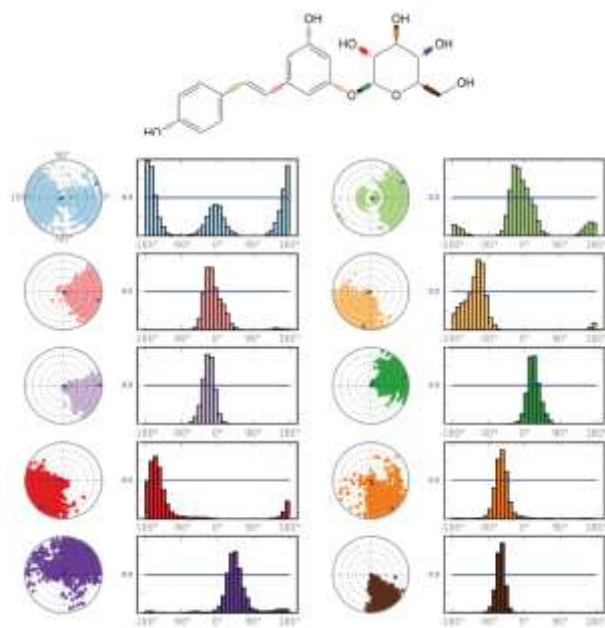


Figure 227: 1FKN_B (SANC00562)

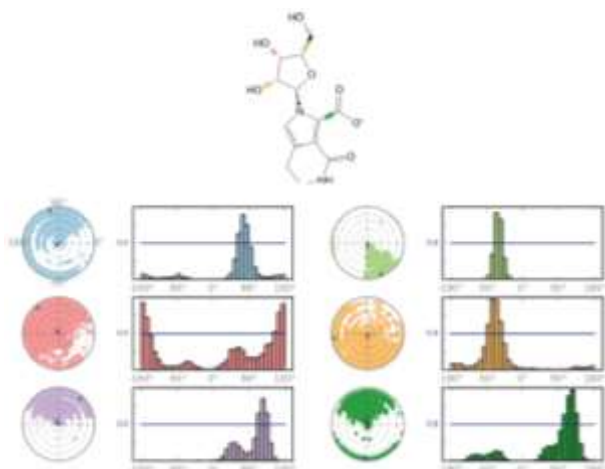


Figure 228: 3LII_A (SANC00269)

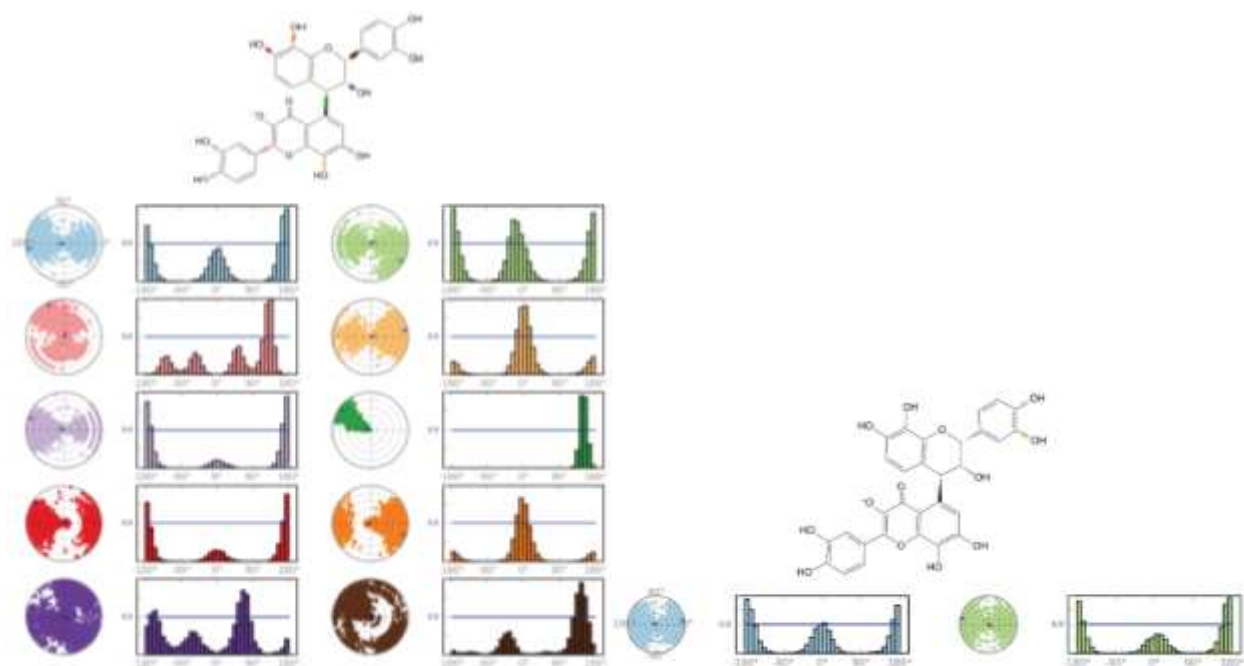


Figure 229: 3LII_B (SANC00559) (i)

(ii)

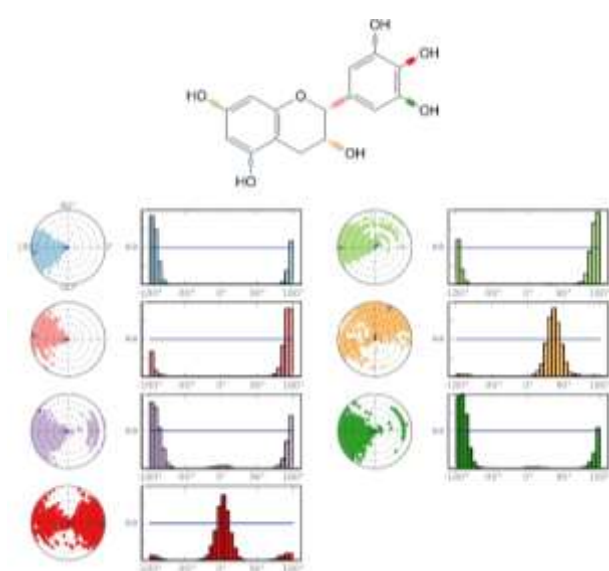


Figure 230: 1J1C_A (SANC00103)

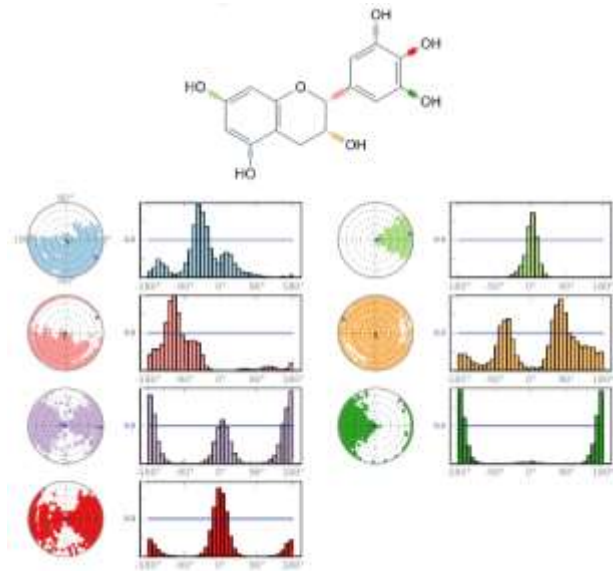


Figure 231: 4EY7_A (SANC00103)

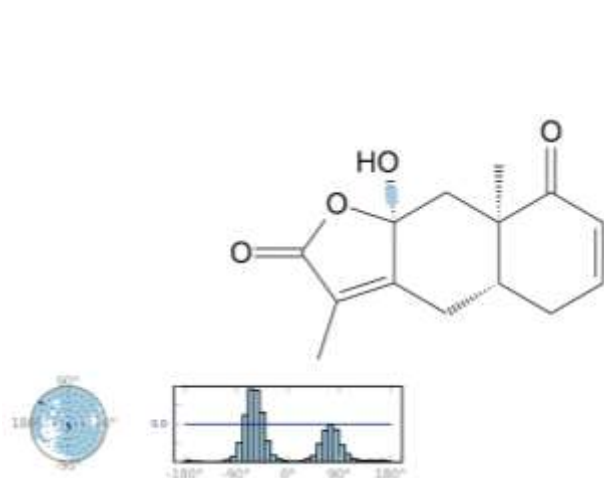


Figure 232: 5E2W (SANC00349)

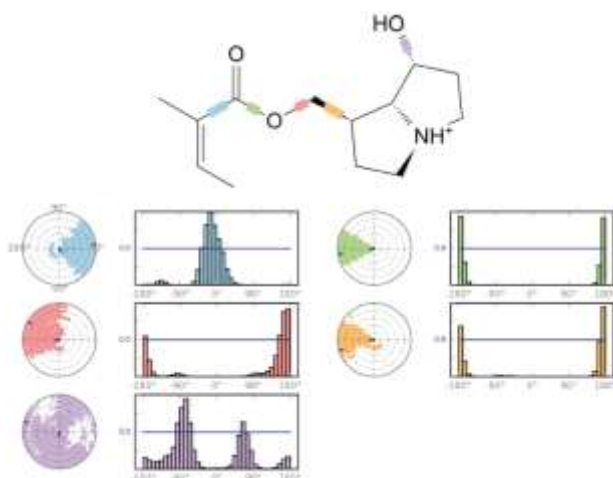


Figure 234: 1GQR (SANC00111)

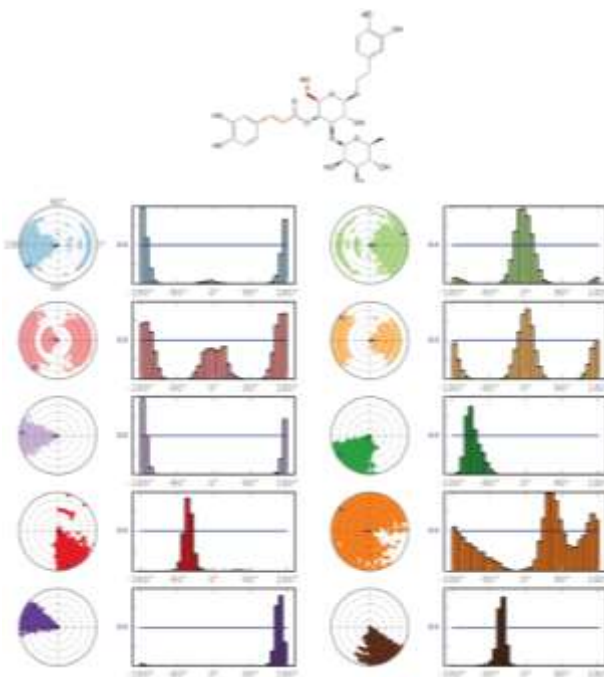
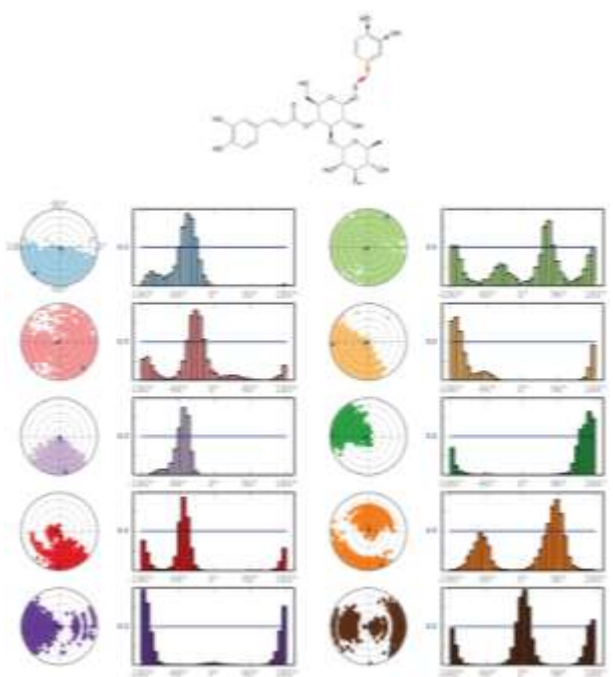


Figure 235: 6C2I (SANC00370) (i)



(ii)

6.3.10 Ligand Properties

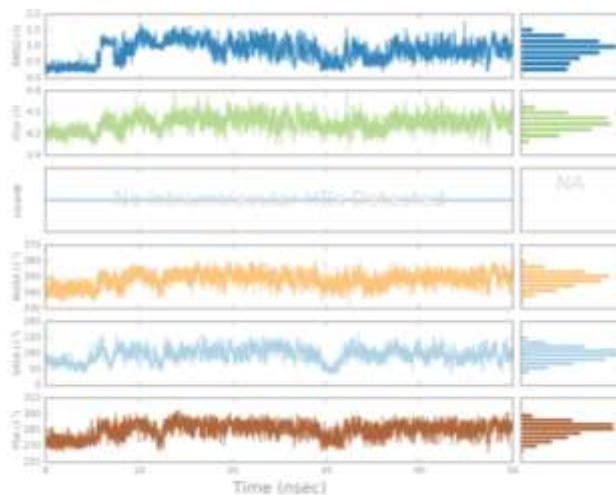


Figure 236: 1FKN_B (SANC00562)

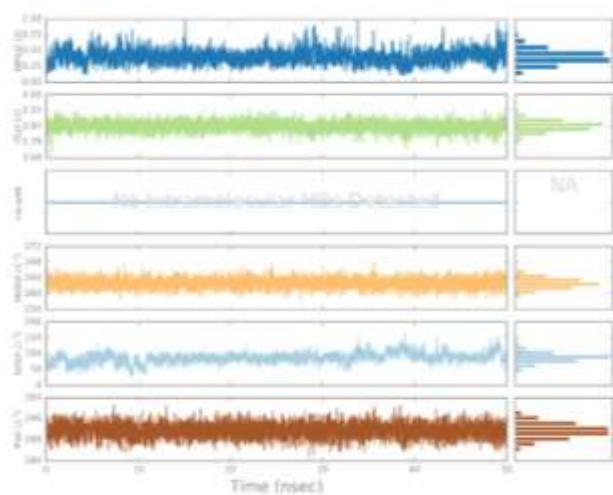


Figure 237: 1J1C_A (SANC00103)

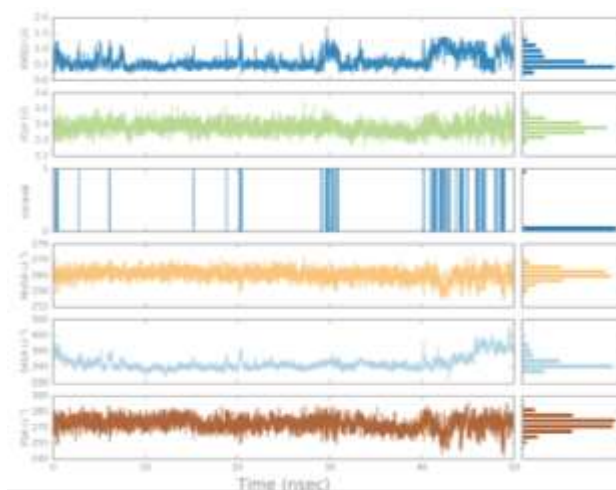


Figure 238: 3LII_A (SANC00269)

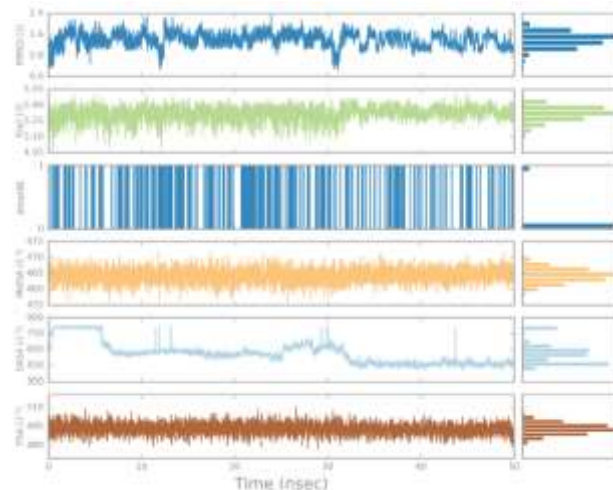


Figure 239: 3LII_B (SANC00559)

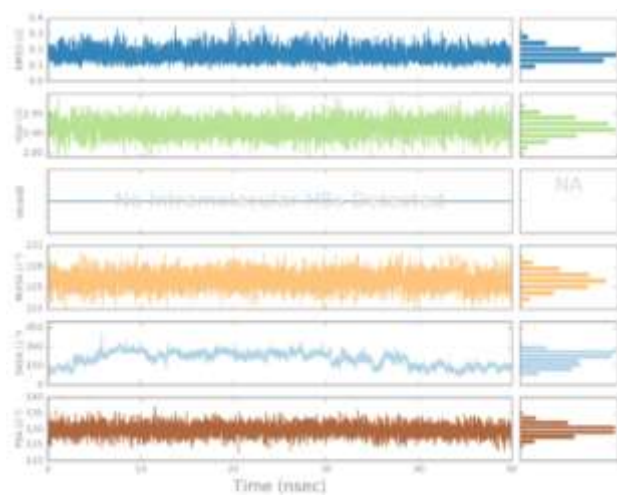


Figure 240: 5E2W (SANC00349)

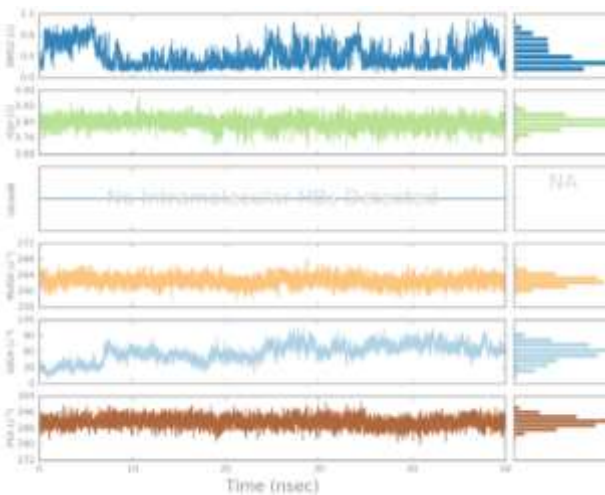


Figure 241: 4EY7_A (SANC00103)

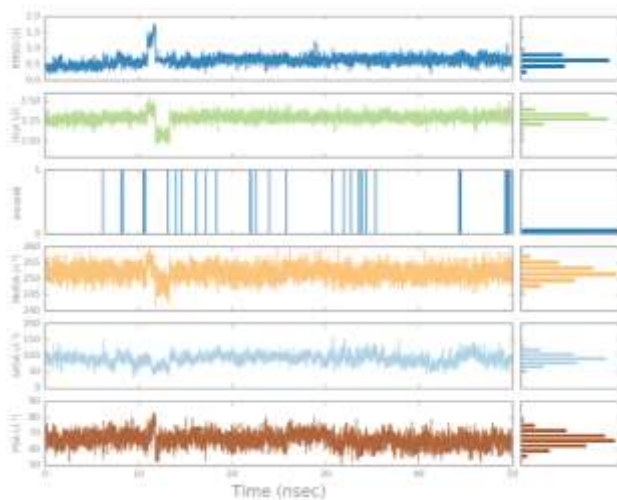


Figure 242: 1GQR (SANC00111)

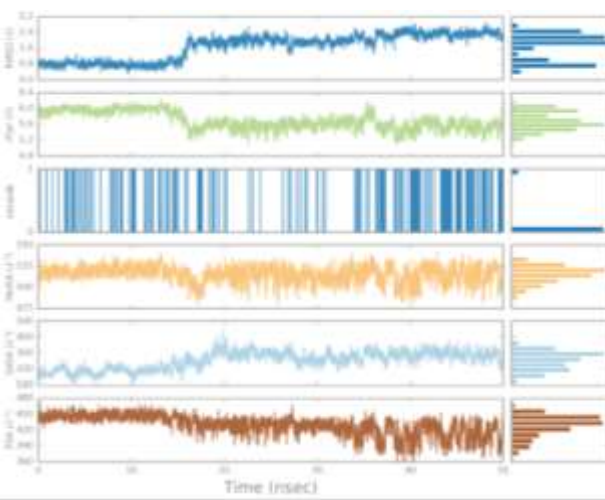


Figure 243: 6C2I (SANC00370)

6.4 PubChem target Molecular Dynamics

6.4.1 Protein-Ligand RMSD

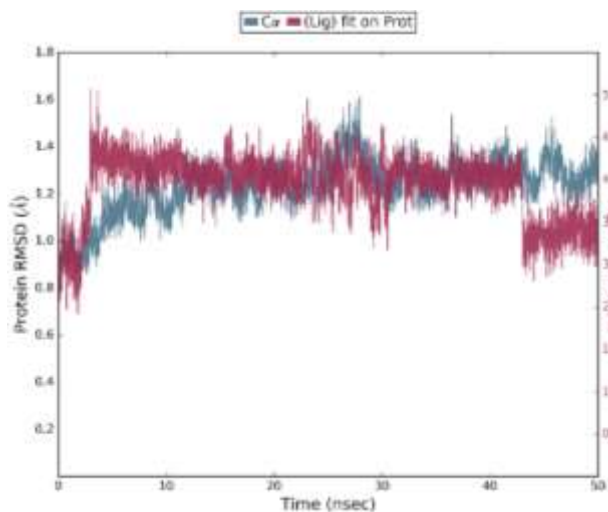


Figure 244: 1GQR_11777124

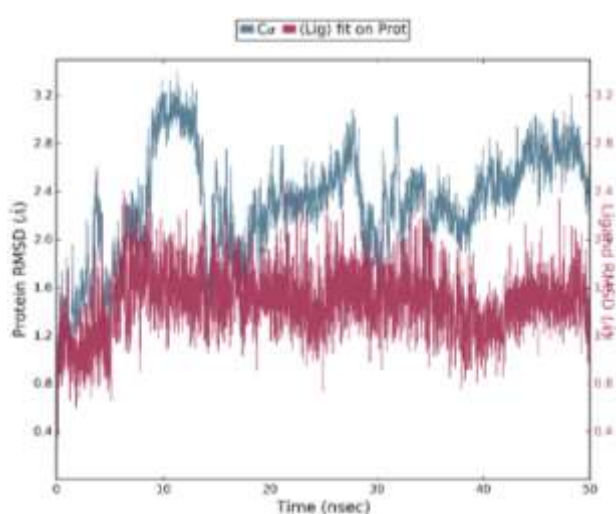


Figure 245: 1J1C_A_122542902

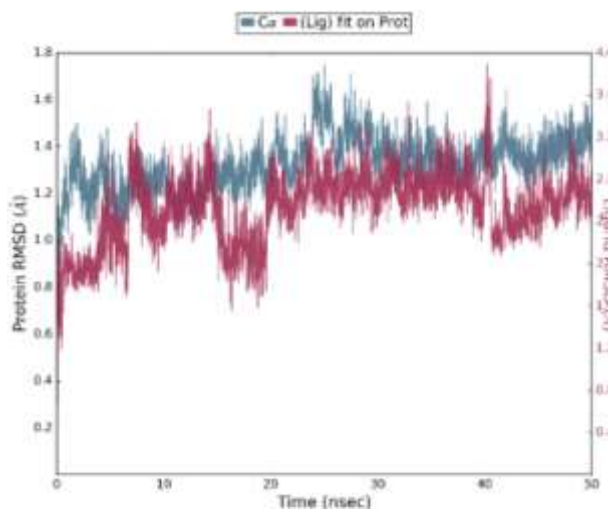


Figure 246: 1FKN_A_11777124

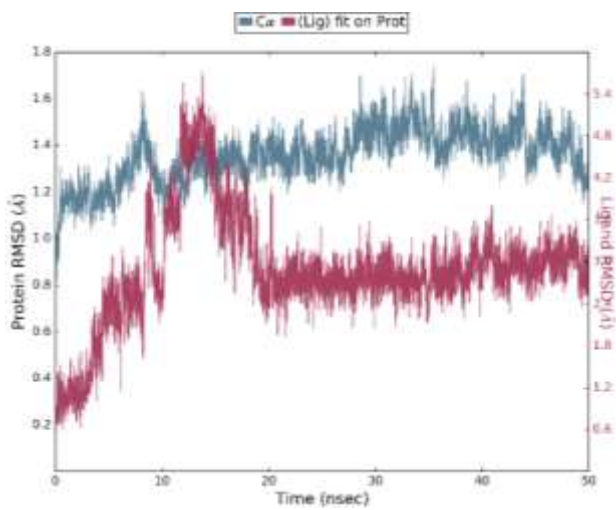


Figure 247: 4CTH_59370901

6.4.2 Protein RMSF

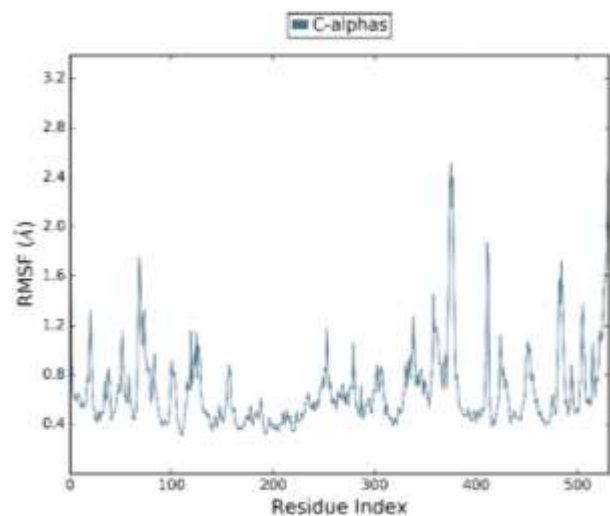


Figure 247: 1GQR_11777124

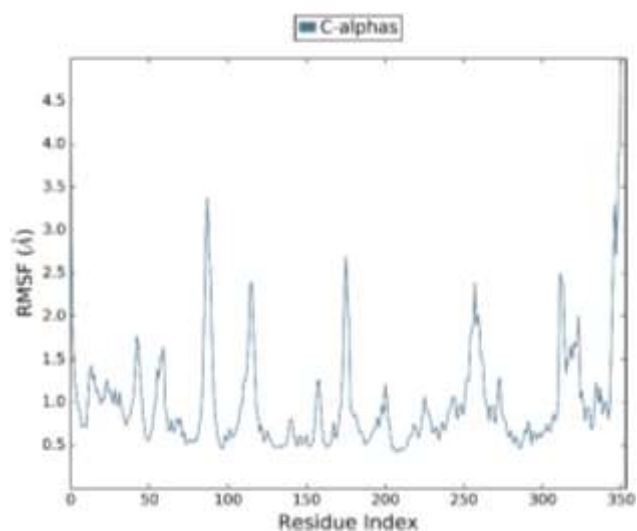


Figure 248: 1J1C_A_122542902

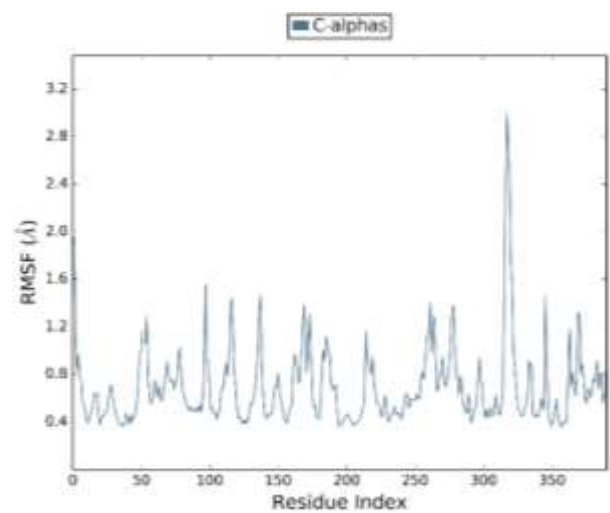


Figure 249: 1FKN_A_101608678

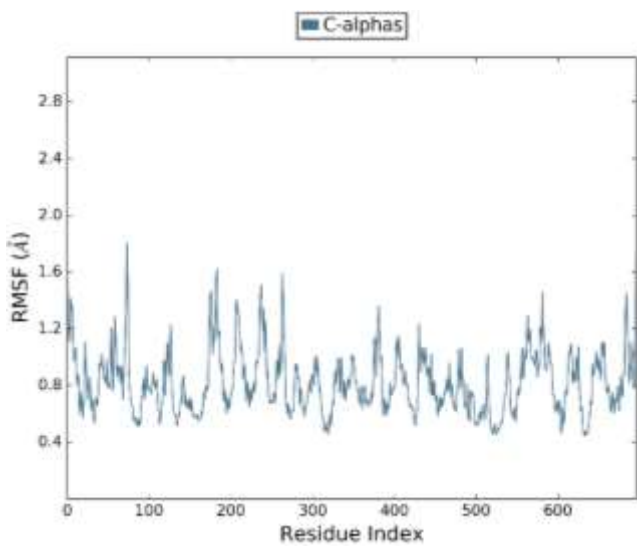


Figure 250: 4CTH_59370901

6.4.3 Protein SSE Histogram

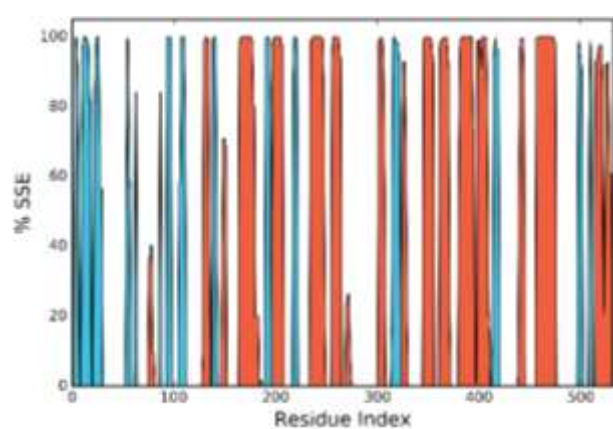


Figure 251: 1GQR_11777124

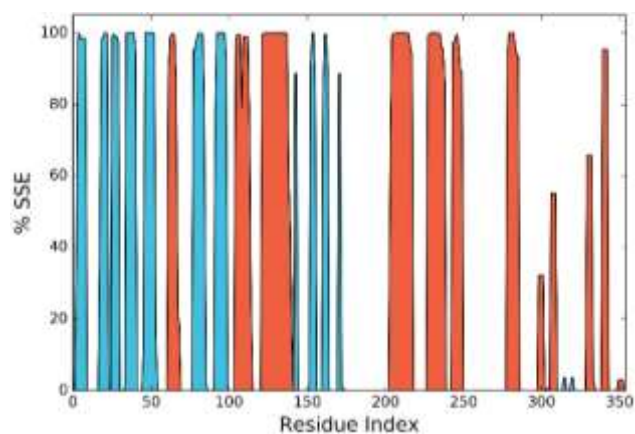


Figure 252: 1J1C_A_122542902

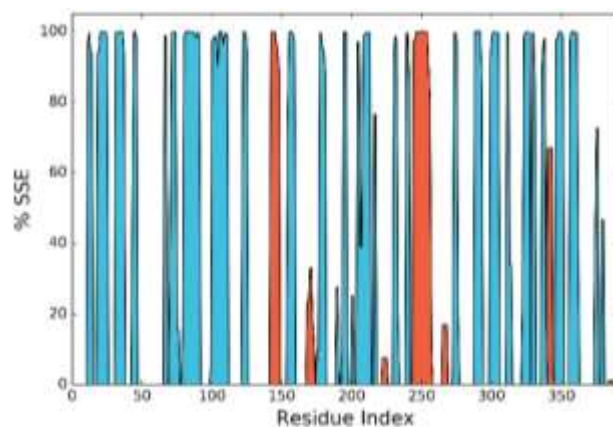


Figure 253: 1FKN_A_101608678

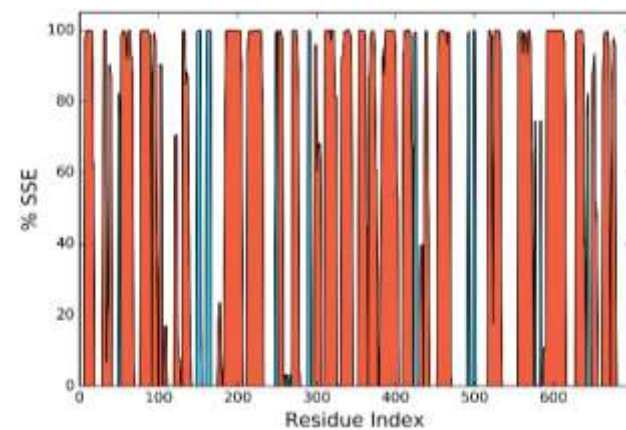


Figure 254: 4CTH_59370901

6.4.4 Protein SSE timeline

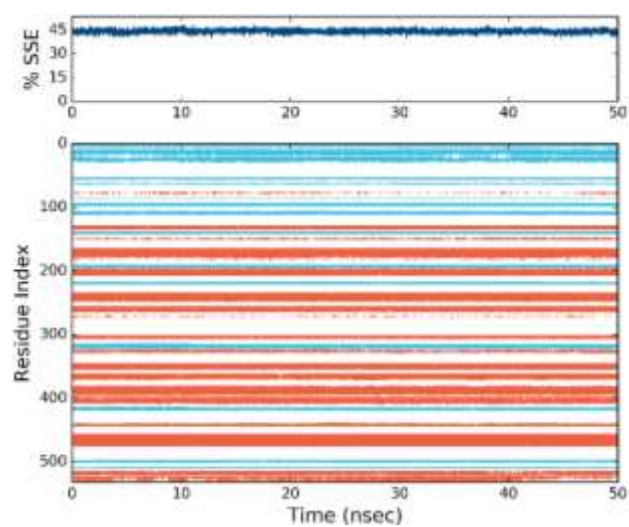


Figure 255: 1GQR_11777124

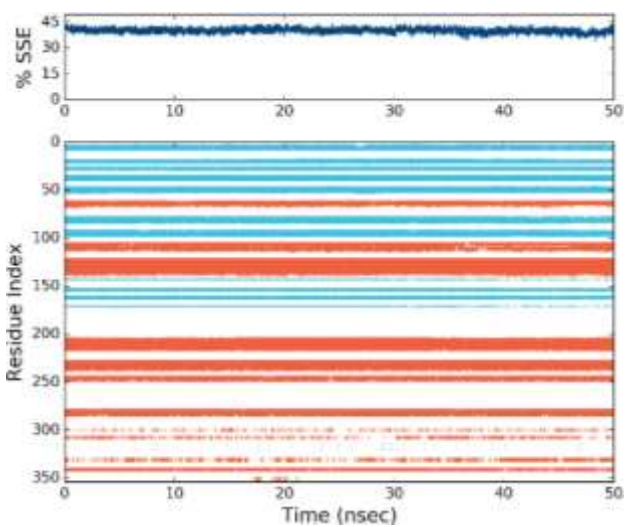


Figure 256: 1J1C_A_122542902

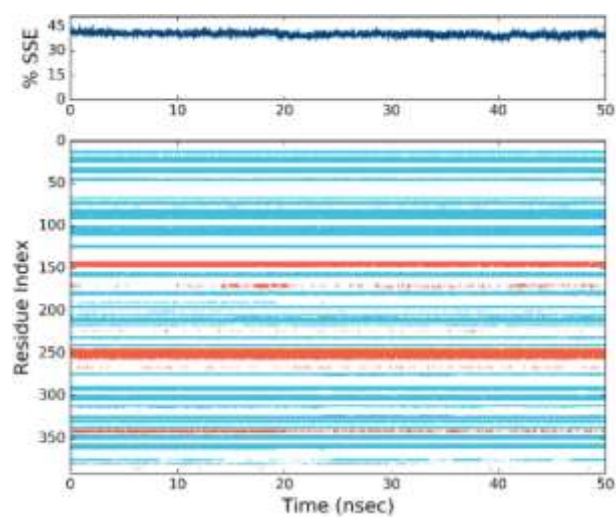


Figure 257: 1FKN_A_101608678

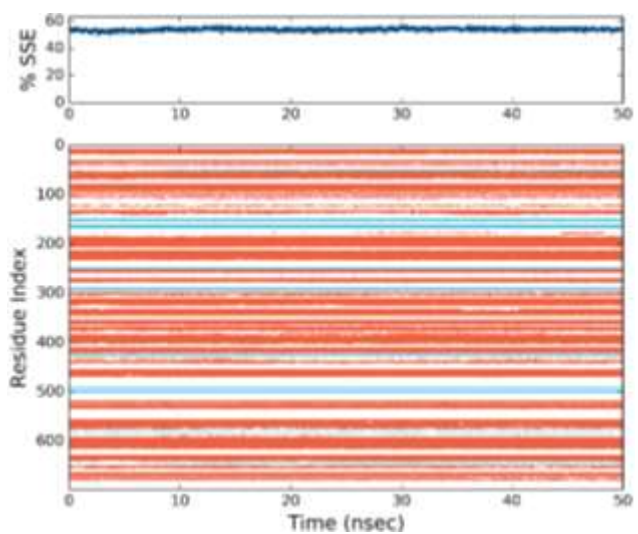


Figure 258: 4CTH_59370901

6.4.5 Ligand RMSF



Figure 259: 1GQR_11777124

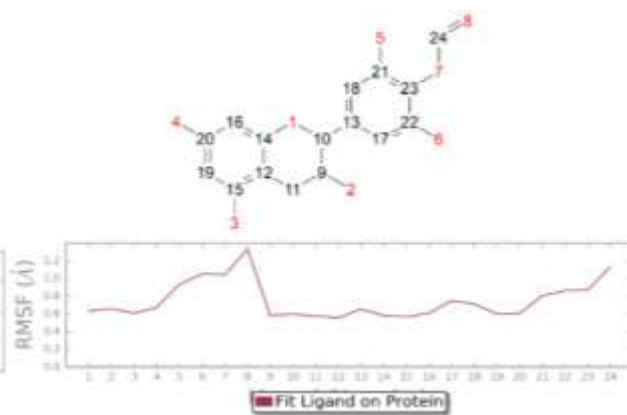


Figure 260: 1J1C_A_122542902

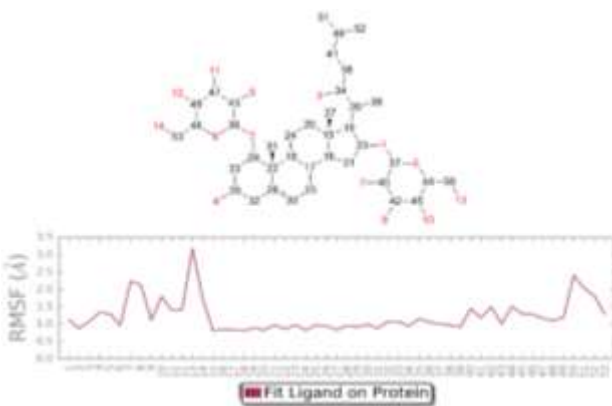


Figure 261: 1FKN_A_101608678

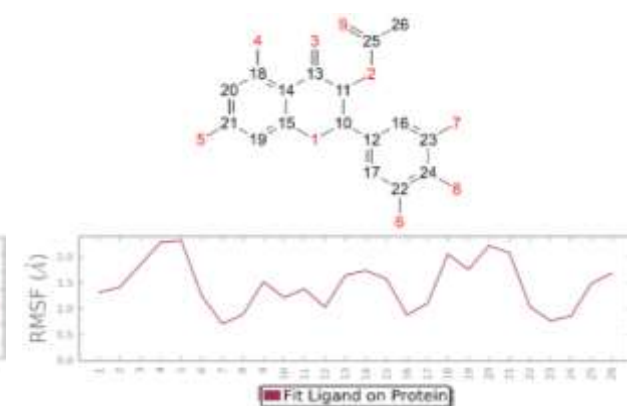


Figure 262: 4CTH_59370901

6.4.6 Ligand-Protein Contacts 2D summary

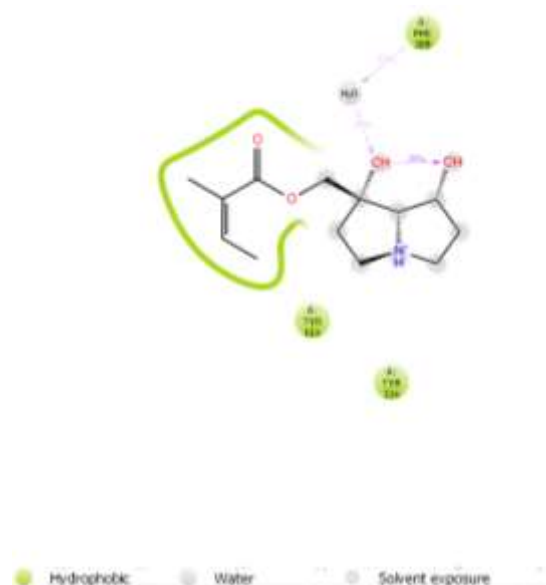


Figure 263: 1GQR_11777124

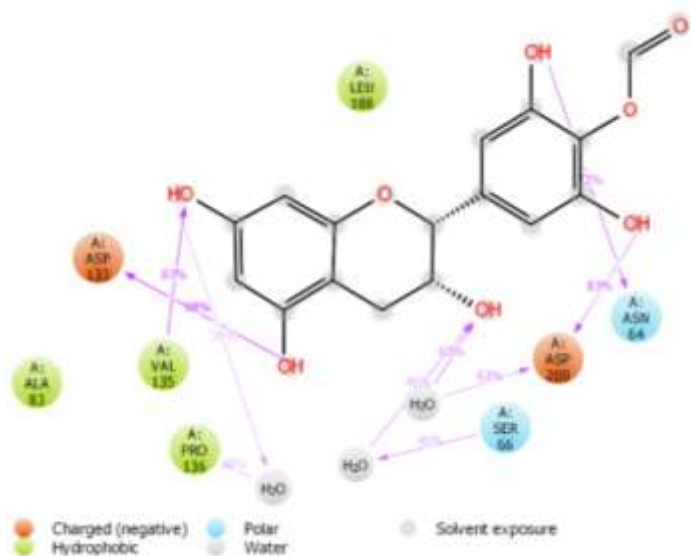


Figure 264: 1J1C_A_122542902

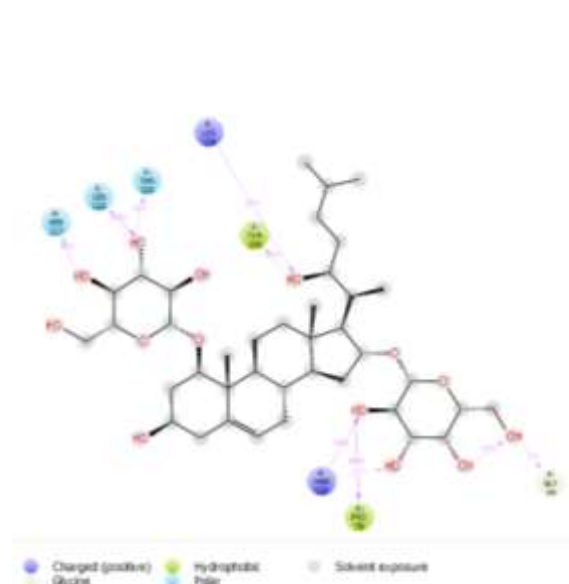


Figure 265: 1FKN_A_101608678

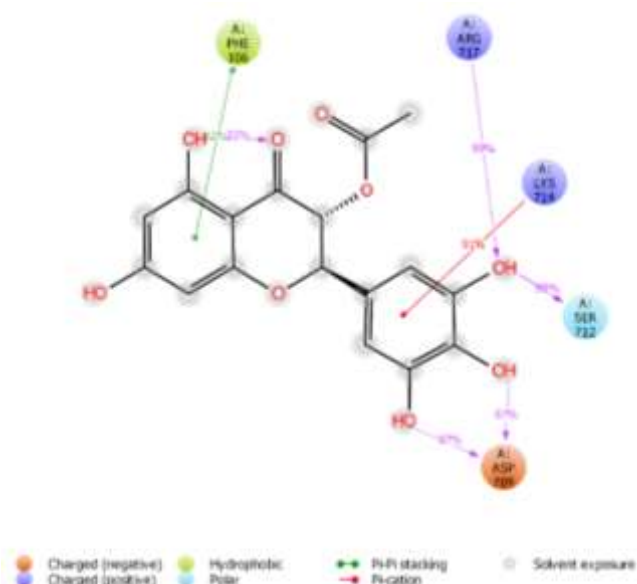


Figure 266: 4CTH_59370901

6.4.7 Protein-Ligand Contacts Histogram

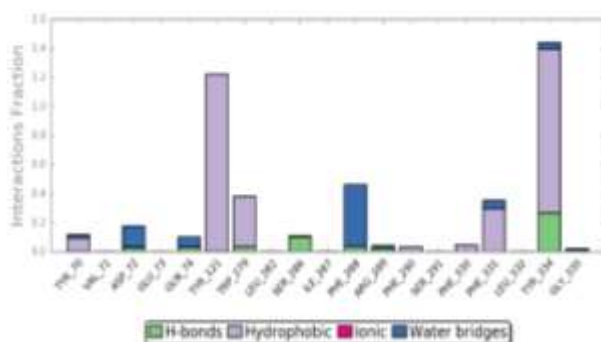


Figure 267: 1GQR_11777124

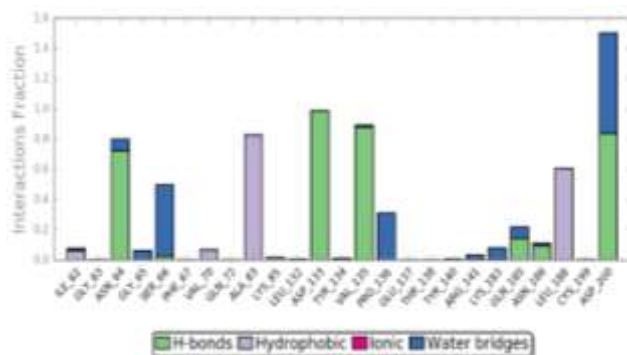


Figure 268: 1J1C_A_122542902

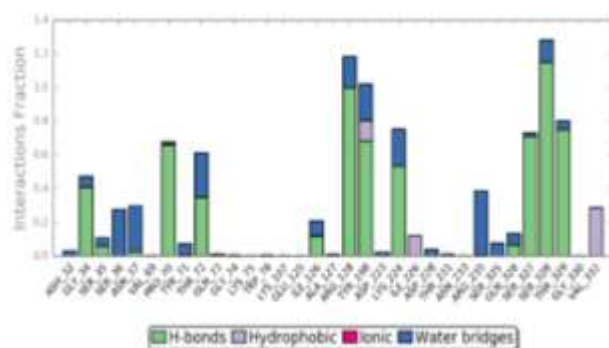


Figure 269: 1FKN_A_101608678

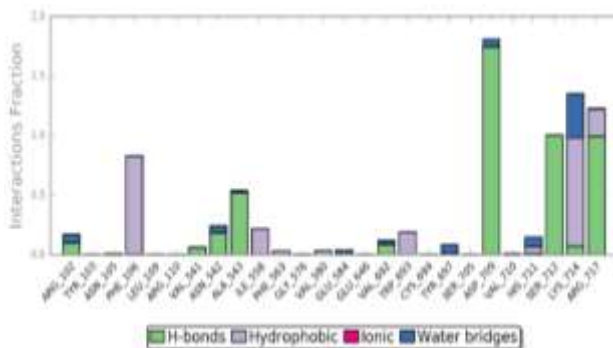


Figure 270: 4CTH_59370901

6.4.8 Protein-Ligand Contacts Timeline

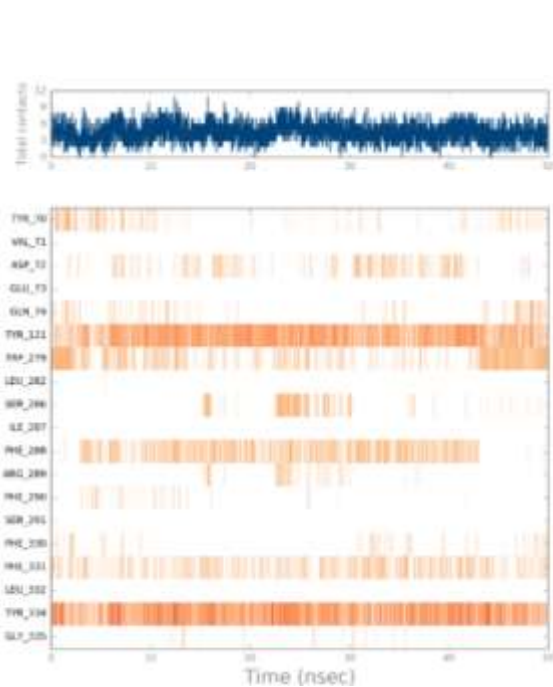


Figure 271: 1GQR_11777124

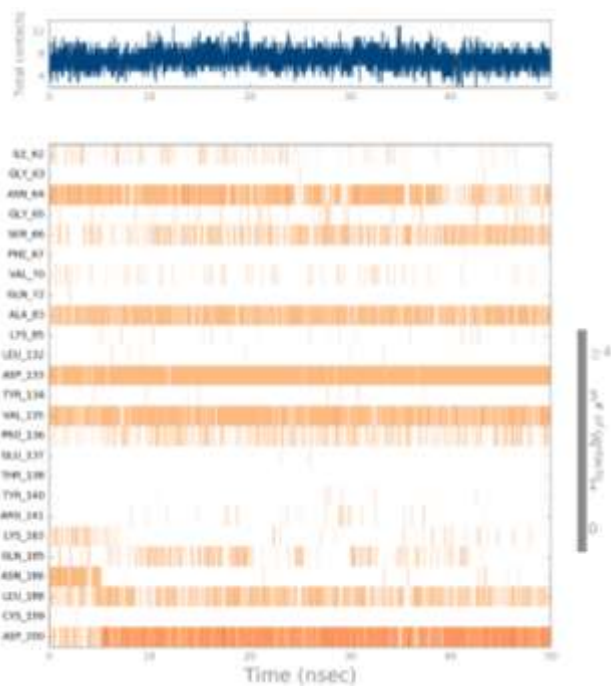


Figure 272: 1J1C_A_122542902

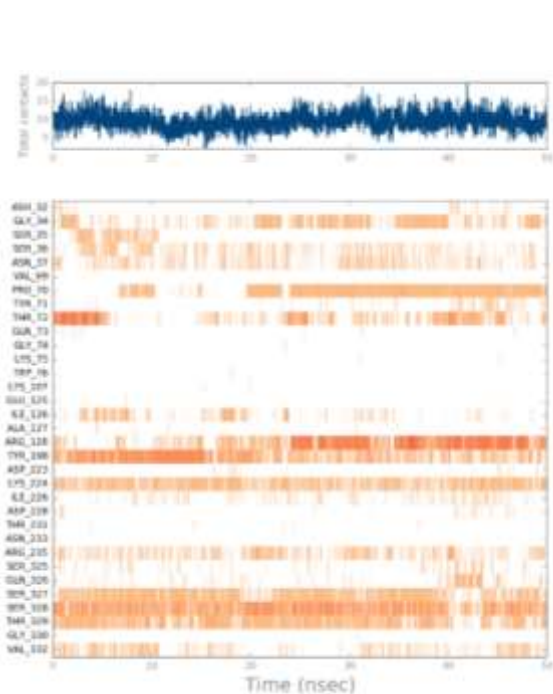


Figure 273: 1FKN_A_101608678

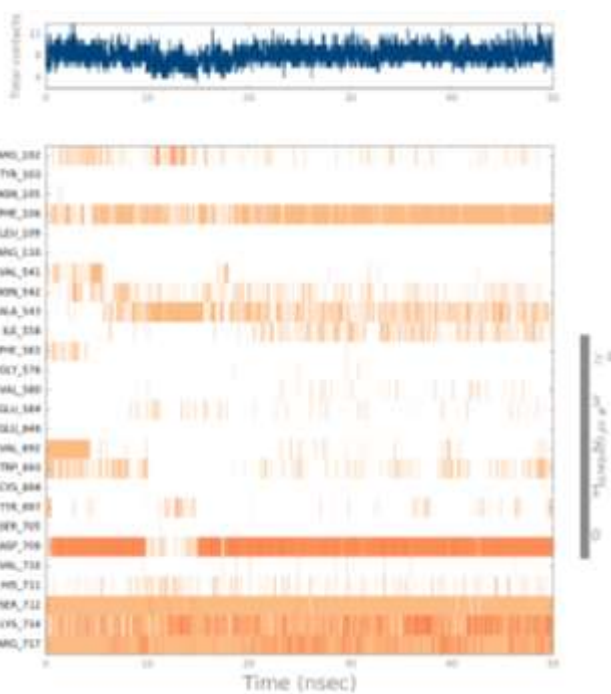


Figure 274: 4CTH_59370901

6.4.9 Ligand Torsion

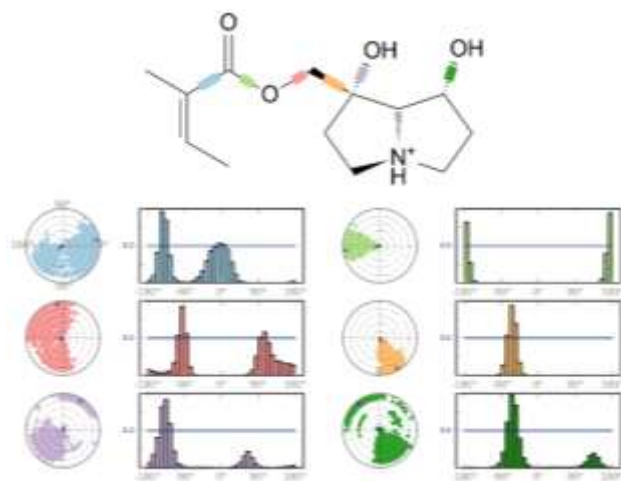


Figure 275: 1GQR_11777124

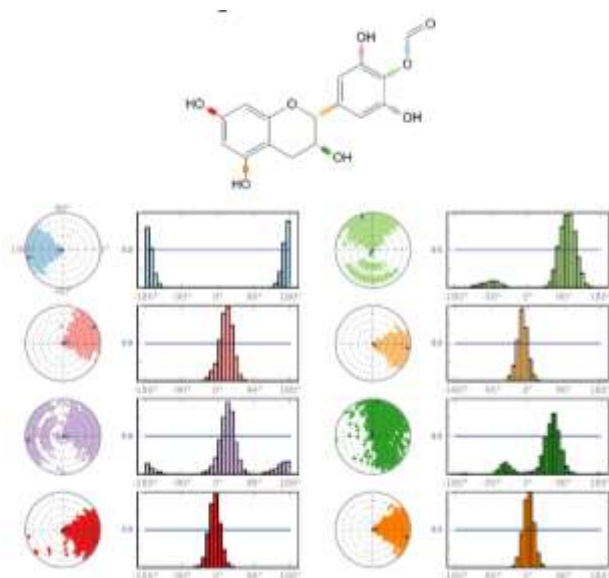


Figure 276: 1J1C_A_122542902

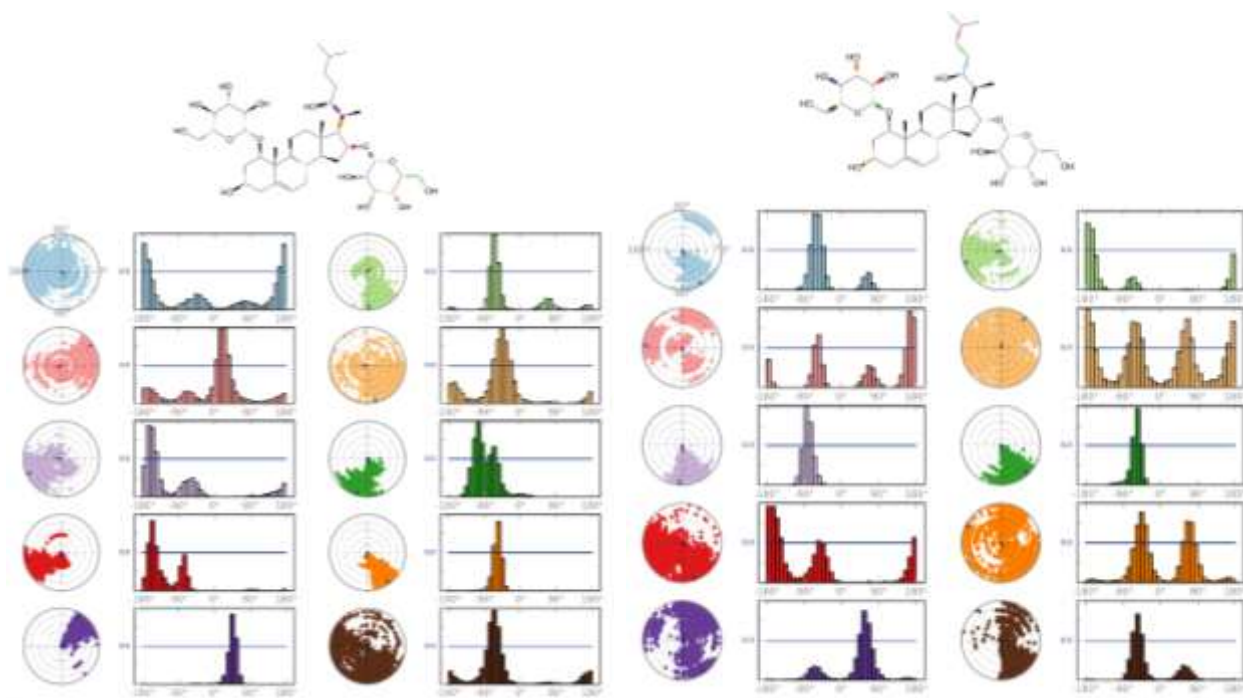


Figure 278: 1J1C_A_122542902 (i)

(ii)

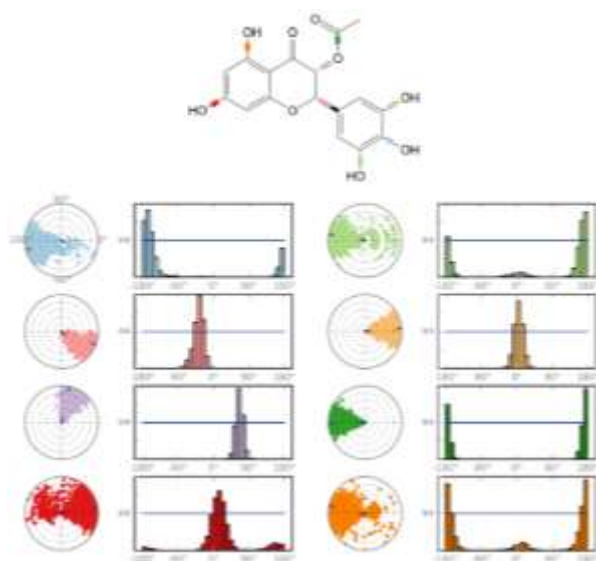


Figure 279: 4CTH_59370901

6.4.10 Ligand Properties

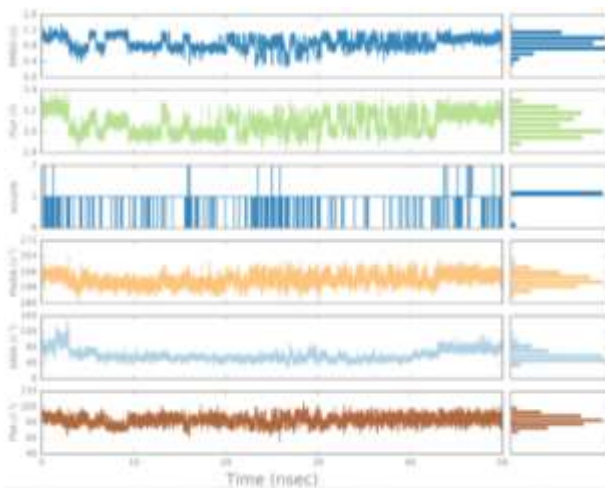


Figure 280: 1GQR_11777124

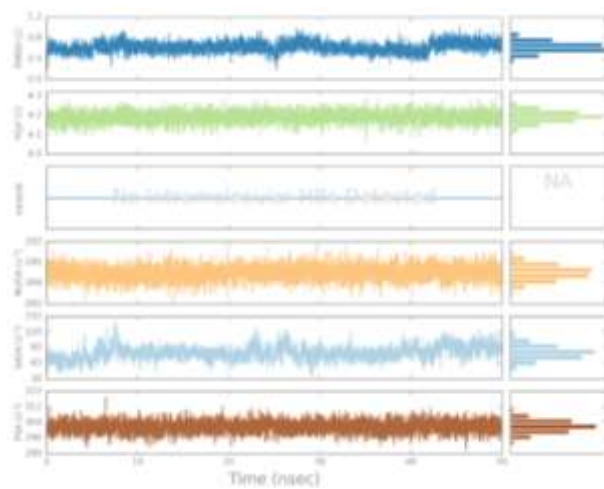


Figure 281: 1J1C_A_122542902

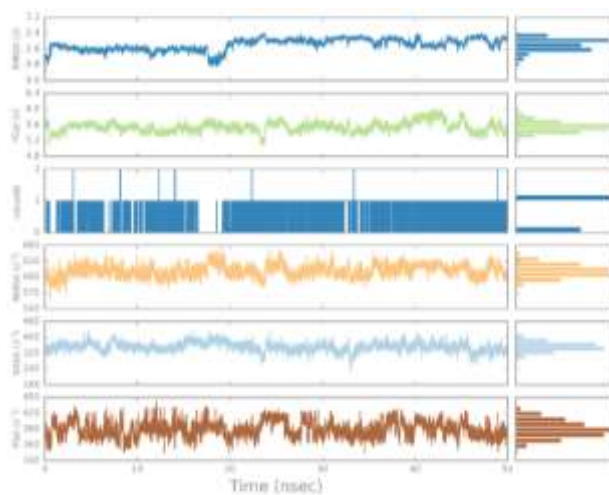


Figure 282: 1FKN_A_101608678

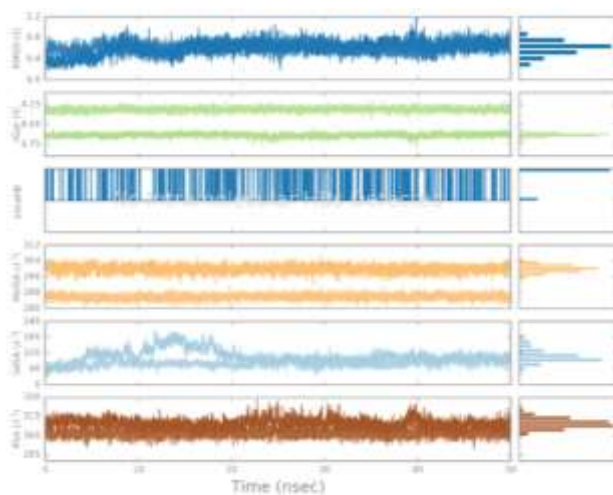


Figure 283: 4CTH_59370901

6.5 Molecular Dynamics 100 ns additional information

6.5.1 Ligand RMSF

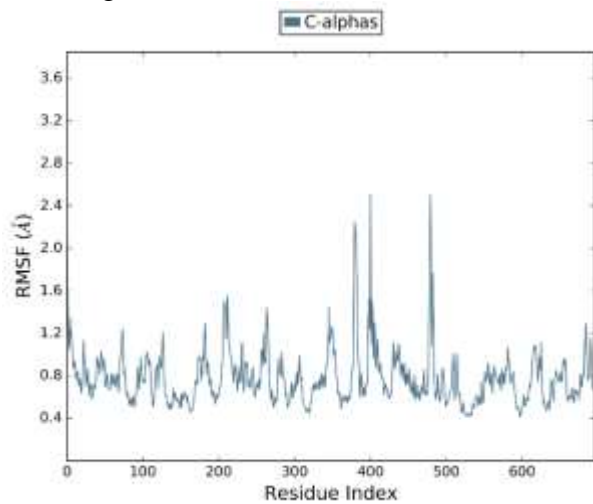


Figure 284: 4CTH (SANC00103)

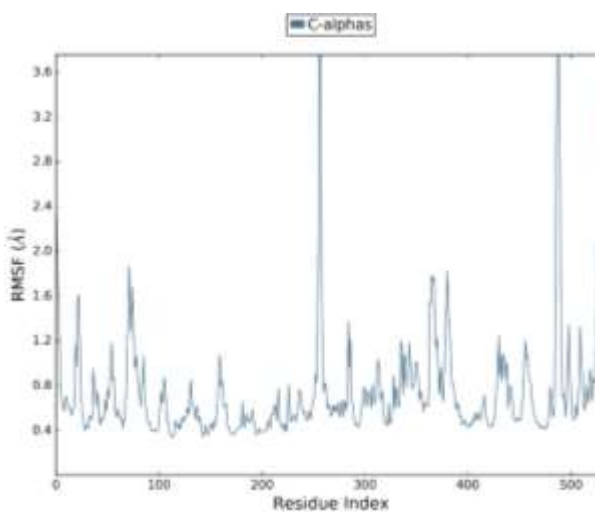


Figure 285: 4EY7_B (SANC00374)

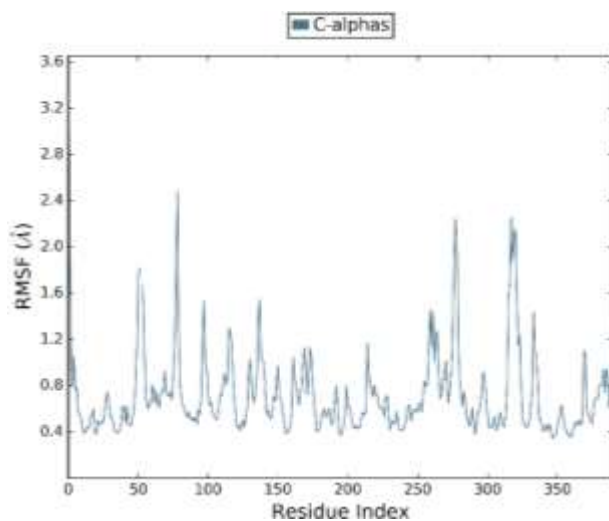


Figure 286: 1FKN_A (SANC00664)

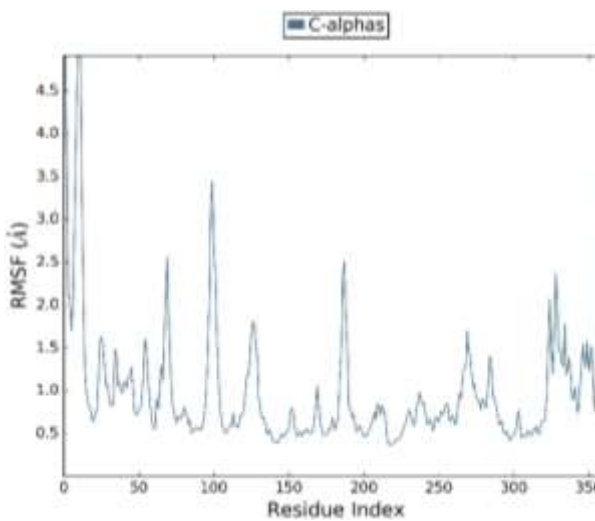


Figure 287: 1J1C_B (SANC00370)

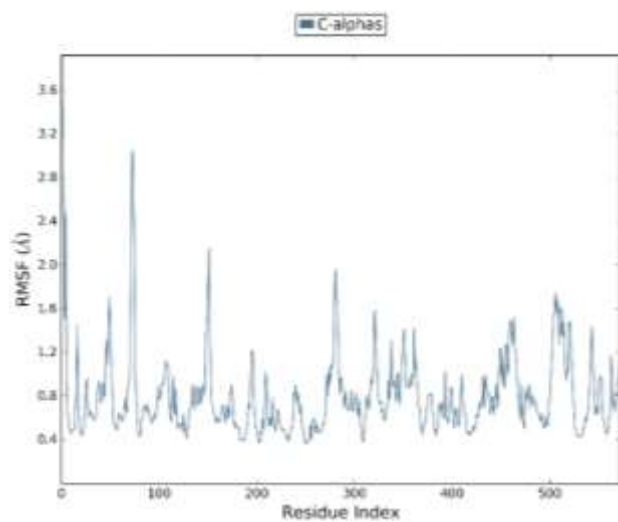


Figure 288:4R12 (SANC00256)

6.6 Ligand Structures

6.6.1 Best Ligand Structures from SANCDB

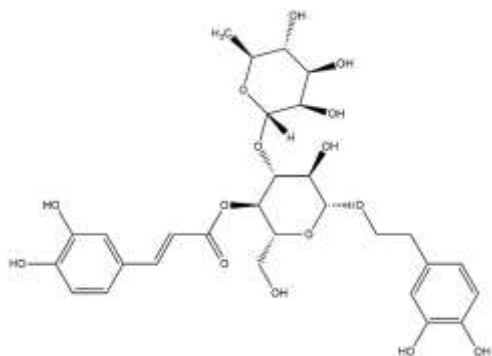


Figure 289: SANC00370

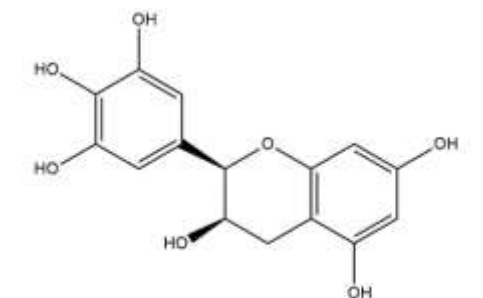


Figure 290: SANC00103

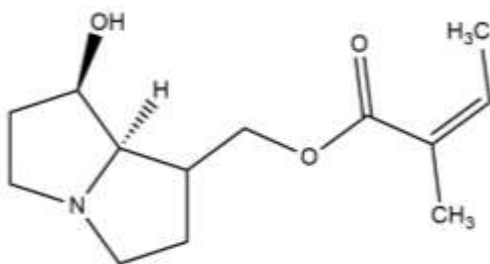


Figure 291: SANC00111

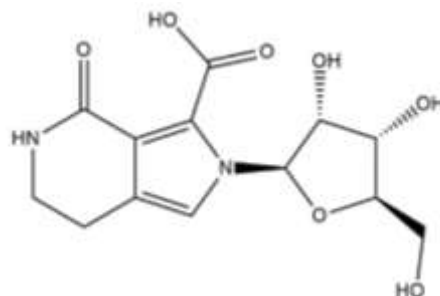


Figure 292: SANC00269

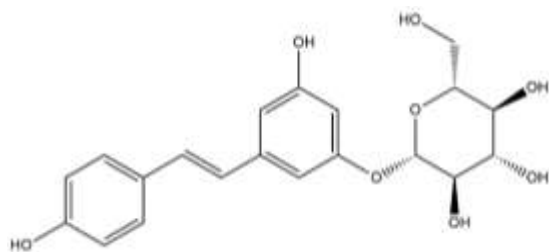


Figure 293: SANC00562

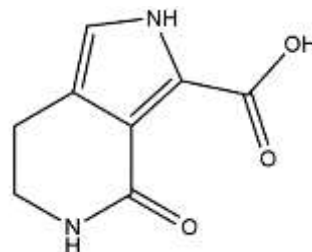


Figure 294: SANC00256

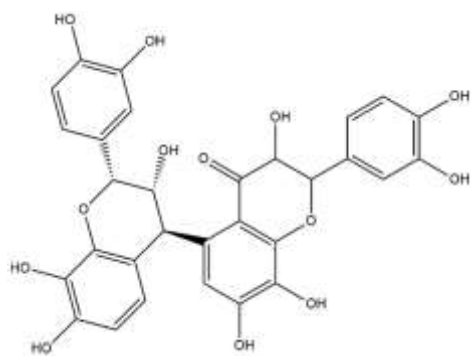


Figure 295: SANC00559

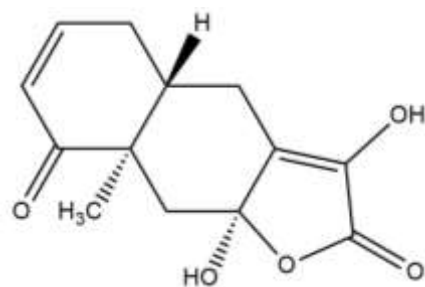


Figure 296: SANC00349

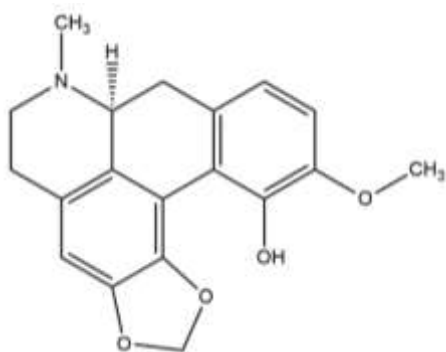


Figure 297: SANC00274

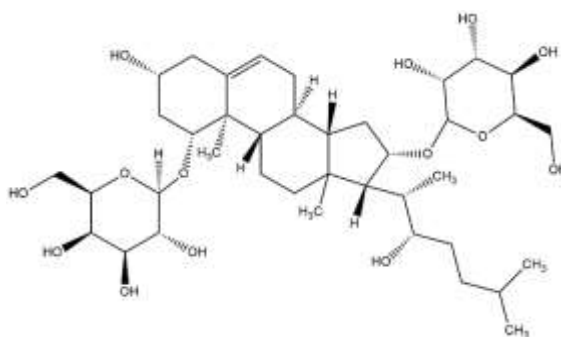


Figure 298: SANC00664

6.6.2 Best Ligand Results from PubChem.

The following are the best performing ligands from PubChem, using the best SANCDB binder as a search input into the respective database, and high throughput virtual screening on the results.

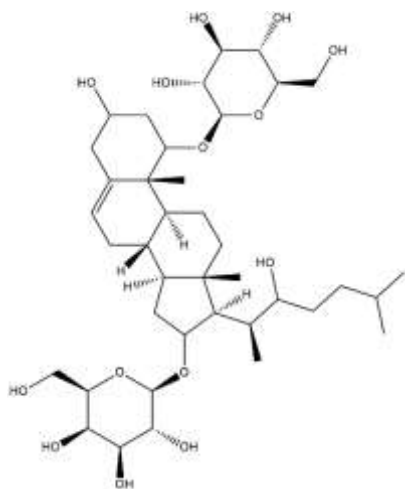


Figure 299: 1FKN_A-101608678

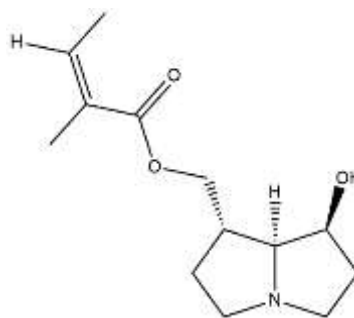


Figure 300: 1GQR-11777124

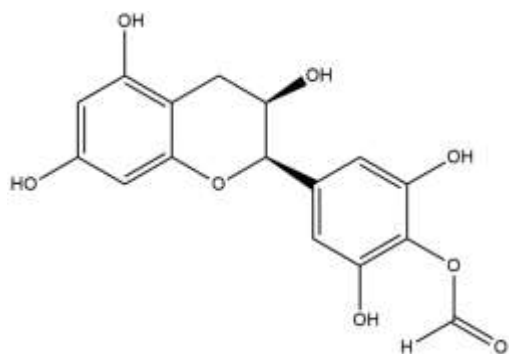


Figure 301: 1J1C_A-122542902

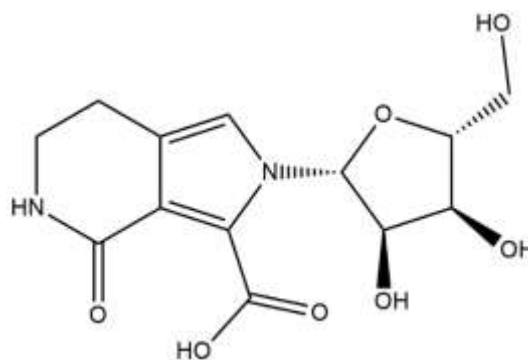


Figure 302: 3LII_A-11209156

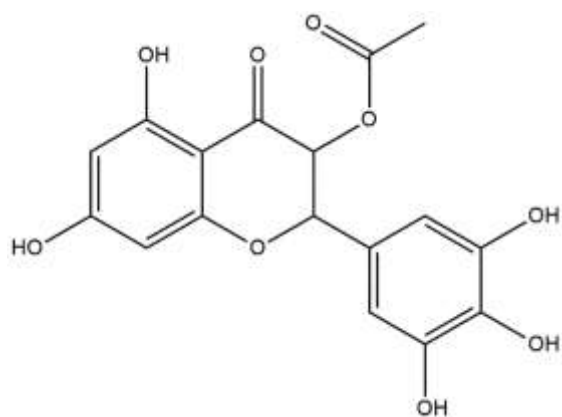


Figure 303: 4CTH-59370901

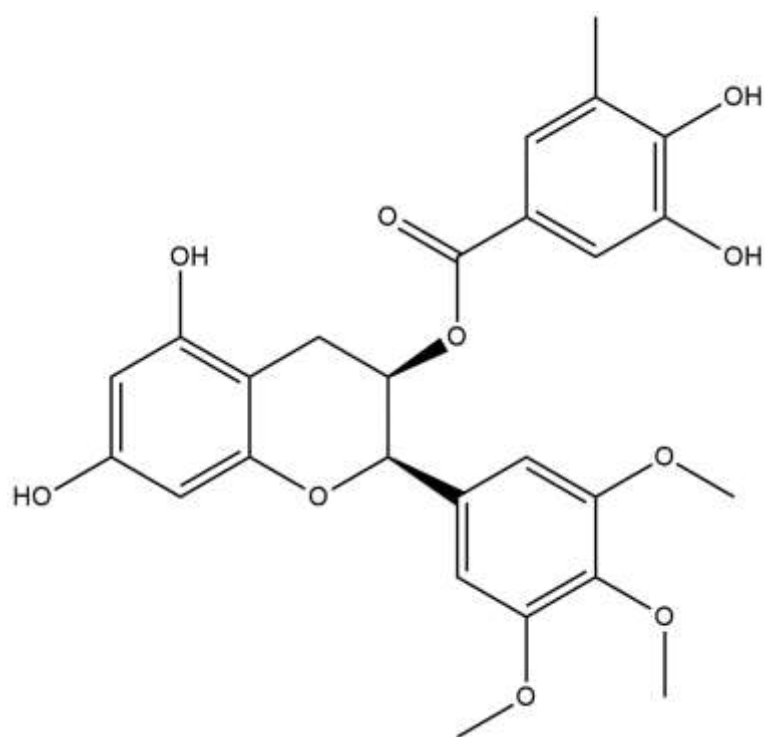


Figure 304: 4EY7_A-68103282

6.6.3 Docking Scores Plots

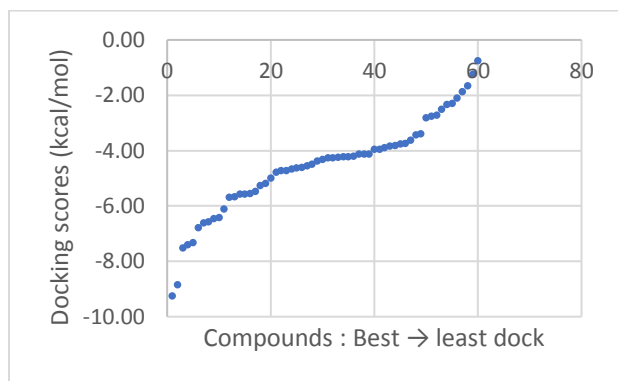


Figure 305: 1GQR

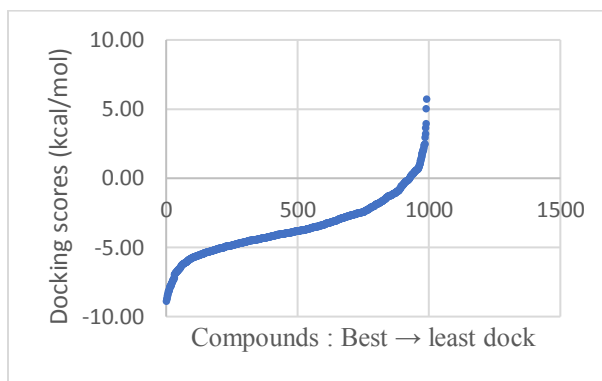


Figure 306: 1FKN_B

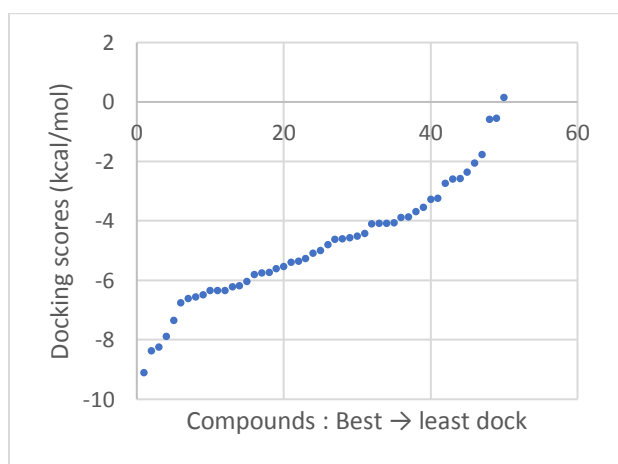


Figure 307: 1J1C_A

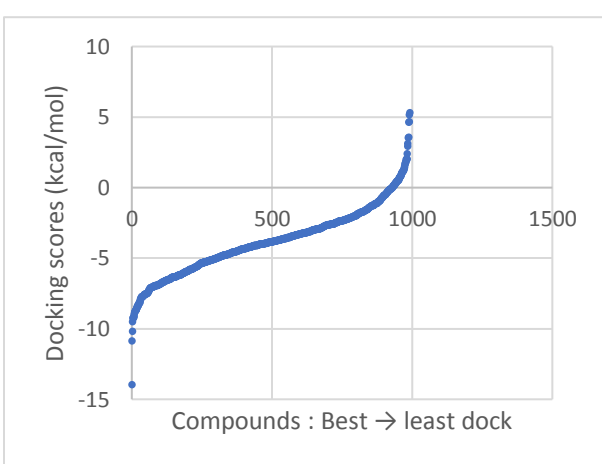


Figure 308: 1J1C_B

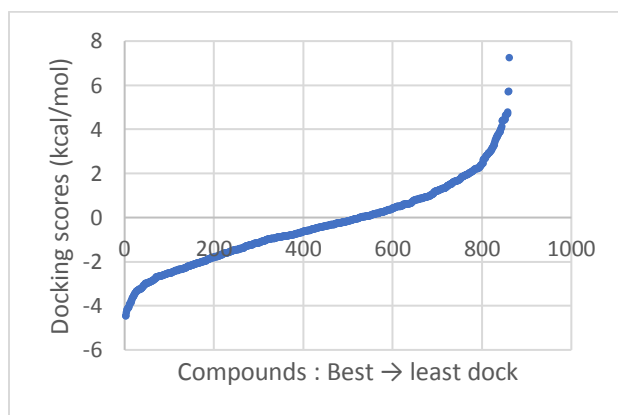


Figure 309: 3LII_A

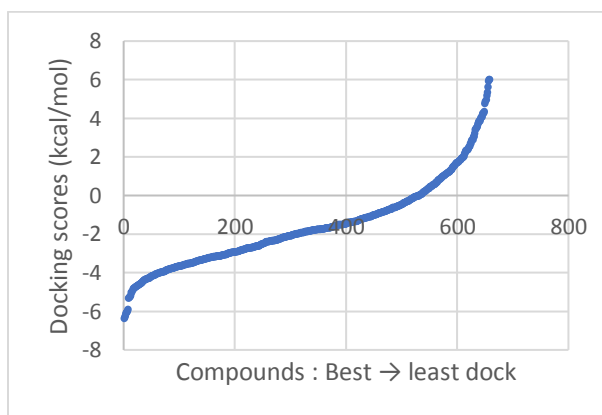


Figure 310: 3LII_B

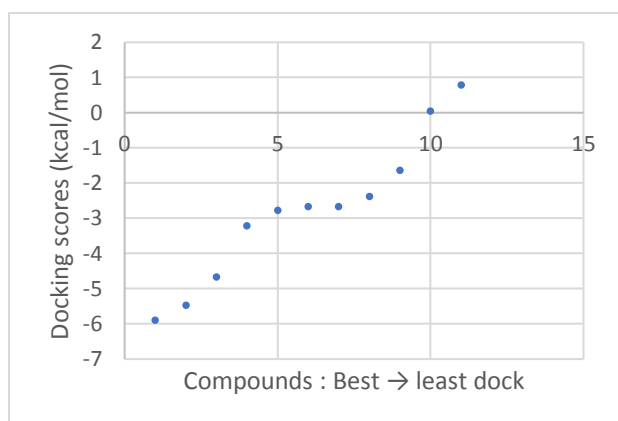


Figure 311: 4CTH

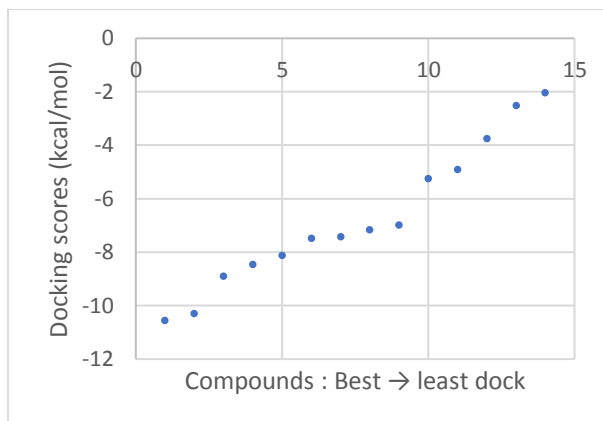


Figure 312: 4EY7_A

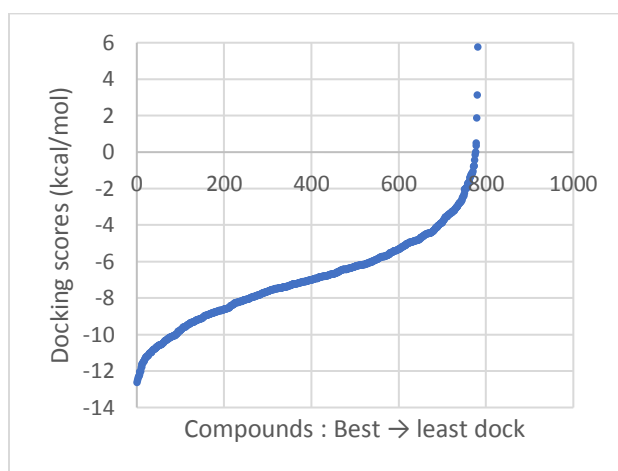


Figure 313: 4EY7_B

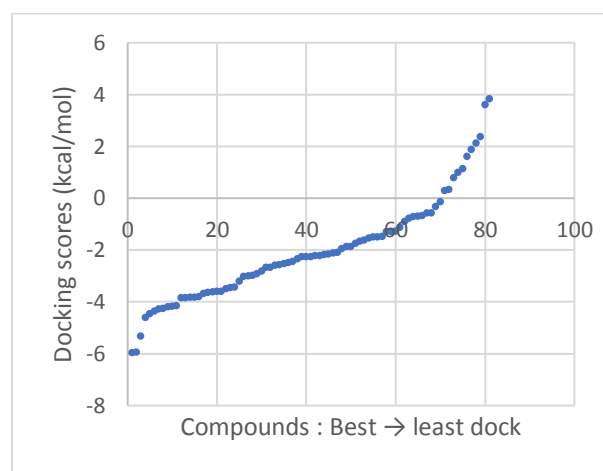


Figure 314: 4R12

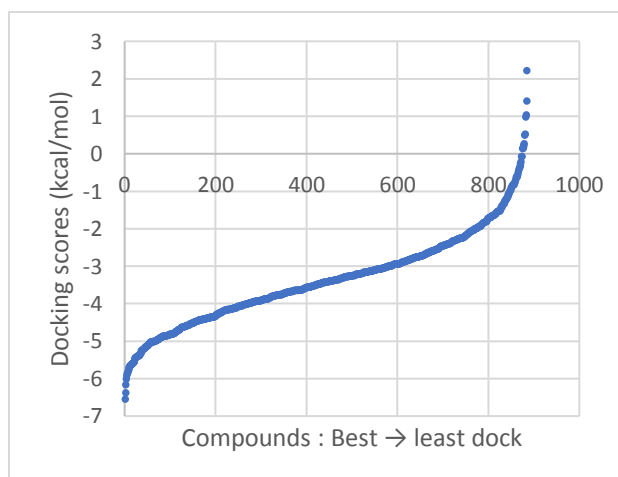


Figure 315: 5e2w

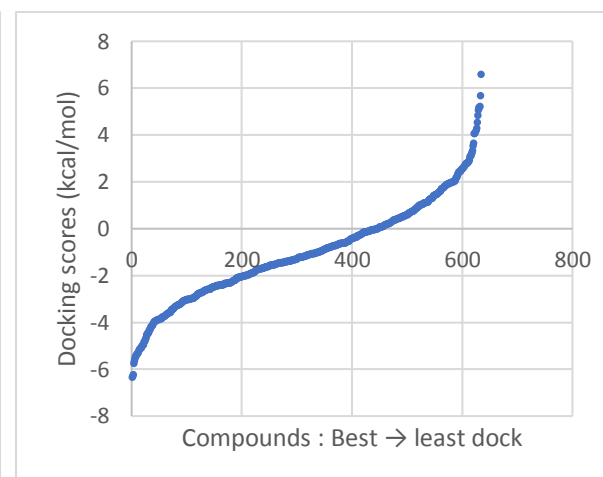


Figure 316: 6BDZ

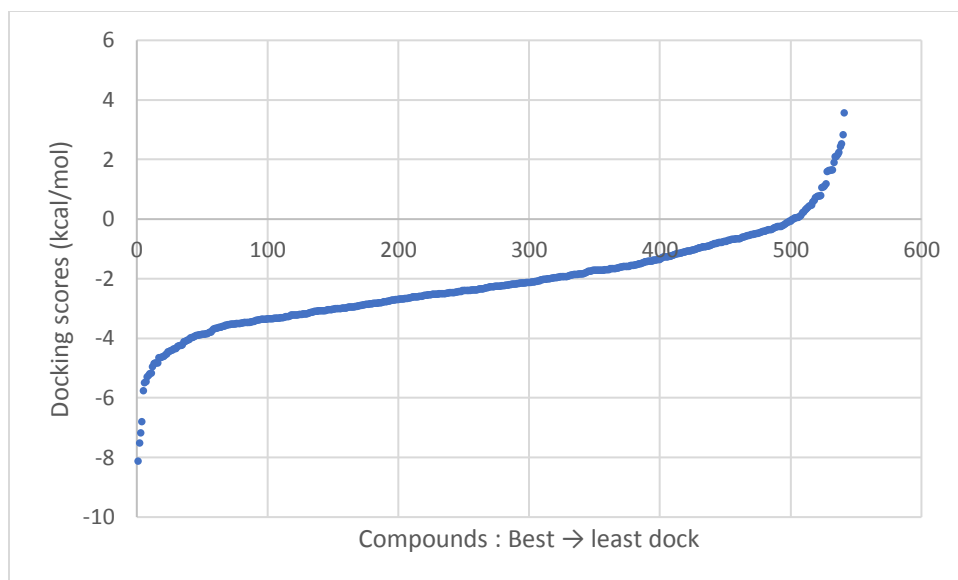


Figure 317: 6C2I