Molecular simulations of potential agents and targets of Alzheimer's disease

A thesis submitted in partial fulfilment of the requirements for the degree
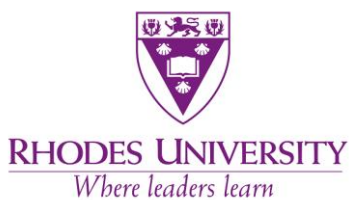
of

MASTER OF SCIENCE
IN BIOINFORMATICS

of

RHODES UNIVERSITY, SOUTH AFRICA

Department of Biochemistry and Microbiology
Faculty of Science

by

Tayna Carlisle

Abstract:

The World Alzheimer Report stated in 2016 that approximately 46.8 million people were living with dementia and this figure is expected to triple by 2050. Alzheimer's Disease was discovered to be a precursor to dementia in 1976 and since then efforts to understand Alzheimer's have been prioritized. To date, there are very few effective forms of treatment for Alzheimer's, many are known to offer only mild calming of the symptoms and have side effects such as diarrhea, nausea, loss of appetite and sleep disturbances. This has been due to lack of understanding on how Alzheimer's is caused. With the two main hallmarks of the disease now being more understood it has opened the doorway into the discovery of new treatments for this disease. This study focuses on the hallmark involving the aggregation of the β-amyloid protein to form plaques surrounding the neurons of the brain. Copper, Zinc and Iron have also been found in high concentrations in and surrounding these plaques. This study focused on the screening of the South African Natural Compound database (SANCDB) to discover hits that have potential destabilizing action against the Beta-amyloid aggregate. If one of these compounds could prove to have destabilizing action on the aggregate it could open the doorway to new potential forms of treatment. Over 700 SANCDB compounds were docked, and the top hits were taken to molecular dynamics to further study the interactions of the compounds and the aggregate. However, the hits identified had strong binding to the aggregate causing it to become stable instead of the desired effect of destabilizing the structure. This information, however, does not rule out the possibility of these compounds preventing the formation of the aggregates. Further, interactions of copper with β-amyloid and copper were determined by solubilizing the aggregate and introducing copper ions in a dynamics simulation. Possible interactions between copper and the methionine residues were visualised.

Table of Contents

Abbreviations

Apolipoprotein E ……………………………………………………..……… (apoE)

Alzheimer's disease ………………………………………………………..… (AD)

Amyloid precursor protein ……………………………………………….....(APP)

Amyloid precursor protein -like protein ……………………….……(APLP)

Beta-amyloid ……………………………………………………….……(Aβ)

Beta APP cleaving enzyme ………………………………………………...…(βACE)

Cerebrovasculature ……………………………………………..…(CAA)

Central nervous system ………………………………………………(CNS)

Chemistry at HARvard Macromolecular Mechanics …………………………..(CHARMM)

Chlorine …………………………………………………...…(Cl)

Circular dichroism ………………………………………………..……...(CD)

COOH-terminal derivative ……………………………………..……(C99)

COOH-terminal fragment ……………………………………...…….(CTFα).

Copper …………………………………………….……...……..(Cu)

Deoxyribonucleic acid ……………………………….…..……....(DNA)

Endoplasmic reticulum ……………………………………..…..(ER)

Early-onset Alzheimer's disease ………………………………………...(EOAD)

External magnetic field ………………………………………...……(B0)

Familial Alzheimer's disease ……………………………………….....(FAD)

Food and Drug Administration …………………………………...……(FDA)

Iron ………………………………………………….……...……...(Fe)

Kunitz-type serine protease inhibitory domain …………………..………(KPI)

Late-onset Alzheimer's disease …………………………….…………...(LOAD)

LDL-receptor-related protein …………………………………….……(LDL)

Logarithm of partition coefficient ………………………….………(logP)

Microtubule associated protein ………………………………….…(MAP)

Molecular Mechanics/Poissan-Boltzman Surface Area ………………..………...(MMPBSA)

Molecular dynamics …………………………………….…...………(MD)

Nanoseconds …………………………………….……………...(ns)

Neurofibrillary tangles ……………………………..……………...(NFT)

Nuclear magnetic resonance …………………………..……………(NMR)

Paired helical fragment …………………………………...………...(PHF)

Periodic boundary conditions ……………………………………………………….(PBC)

Picoseconds ……………………………………………………………...…………….(ps)

Presenilin ……………………………………………………………………… (PSEN)

Protein data bank …………………………………………………………………...(PDB)

Protein misfolding disorders ………………………………………….……...…(PMD)

Quantum mechanics ………………………………………………………………(QM)

Radio frequency ……………………………………………………………………(RF)

Radius of gyration ………………………………………………………………...(RoG)

Reactive oxygen species …………………………………………………...…(ROS)

Research unit in Bioinformatics …………………………………………………(RUBi)

Root mean squared deviation ………………………………………………….(RMSD)

Root mean squared fluctuation …………………………………...……………(RMSF)

Soluble secreted derivative ……………………………………………………(sAPPβ)

Solid-state nuclear magnetic resonance …………………………………..………(SS-NMR)

South African Natural Compound Database …………………………...……....(SANCDB)

Sporadic Alzheimer's disease ………………………………………………...…(SAD)

Topological polar surface area ……………………………………………..……...(TPSA)

Transmission electron microscopy …………………………………………………(TEM)

Zinc …………………………………………………………………………....(Zn)

Amino acid	3-letter symbol	1-letter symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

List of Figures

List of Tables

Acknowledgements:

Prof Özlem Tastan Bishop and RUBi

I cannot even imagine what it must be like to have created and to have successfully run such a fulfilling course. I thank you for accepting me and allowing me to enter this incredible field of study. To the other members of the RUBi, thank you for accepting me and helping me throughout the course. Never have I met such an incredibly talented group of people who are so willing to help each other. I wish you nothing but the best for the future.

My masters class

We have really grown together after this course. We were each other's shoulders to cry on and I would not have been able to do this without you. Special thanks to Jessica and Bertha for our informative study sessions and to Washington for being the best laboratory partner ever. You are all amazing and your futures are very bright.

My family

Special thanks must go to the people who kept me going when I thought I could not. You are my rock and I could not ask for a better group of people to support me no matter what. I love you all.

National Research Foundation

Chapter 1: Introduction

1.1 Alzheimer's Disease

According to the World Alzheimer Report 2016 it is estimated that there are approximately 46.8 million people living with dementia worldwide with 2.2 million of them being South Africans in 2011 (De Jager et al., 2017). It is also the sixth leading cause of all deaths and the fifth leading cause of death in persons aged ≥65 years (Reitz, 2012). With the amount of people living with dementia being expected to triple by 2050 it makes it a growing public health problem (Estrada and Soto, 2007).

Alzheimer's disease (AD) is a neurodegenerative disease that was first described in 1906 but only in 1976 was it recognized to be a cause of dementia. AD is a multi-factorial disease with genetic (70%) and environmental (30%) causes (Dorszewska *et al.*, 2016). It is characterized by loss of short-term memory disorientation, and impairment of judgment and reasoning. Dementia affects individuals' ability to perform everyday activities by minimising their ability to remember, communicate properly and problem-solve due to the damage of neurons in the brain (Estrada and Soto, 2007).

AD can be characterized into four types; Early-onset AD (EOAD), Late-onset (LOAD), Familial Alzheimer's disease (FAD) and sporadic AD (SAD).

Whether the disease is EOAD or LOAD depends on the patient's age when the first symptoms arise. EOAD categorizes patients affected before 65 years of age and is a very rare form that affects up to 5% of all people with AD. This form also appears to be linked with a defect in chromosome 14 of the patient's DNA. Myoclonus, a form of muscle twitching and spasm, is a symptom more common in EOAD. LOAD categorizes patients over 65 years of age and is the most common form of the disease (Dorszewska et al., 2016).

FAD is a form of AD that is linked to the genotype of the patient and accounts for less than 1% of all cases of AD. FAD is due to changes or alterations in specific genes that can be directly passed on from parent to child. In families that are affected, members of at least two generations have had the disease. SAD, however, has no specific familial link and is the more common form of AD. SAD is caused by a combination of the patients genes, environment and lifestyle and generally only affects patients over the age 60-65 (Bird, 1998; Dorszewska *et al*., 2016). The disease initiates with the individual's inability to remember new information that worsens as the patient ages. This occurs as a result of the damage of the neurons found in the brain regions involved in forming new memories. As

more neurons in different regions of the brain are destroyed more neurobehavioral symptoms form such as agitation, insomnia and often delusions (Gaugler *et al.*, 2016). The pace at which the disease progresses often varies between individuals and as progression continues cognitive and functional abilities decline. Advanced stages of this disease require the individuals to be completely cared for as they are unable to perform basic activities (Gaugler *et al.*, 2016).

Severe cases of dementia result in the individual being bedridden. When an individual is immobile contaminated mucus, produced in the lungs, pools in the lower part of the airway. In a mobile person this mucus is usually disposed of by the movement of it to the pharynx which allows the mucus to be swallowed. If the individual is dehydrated the pooled mucus becomes thick and this leads to the contraction of infections such as pneumonia. As a result of the patient's weakened immune system, they are not able to combat the infection and this proves to be fatal (Nigam *et al.*, 2009).

In 2014, 59% of people worldwide incorrectly believed that Alzheimer's disease is a typical part of aging (Estrada and Soto, 2007). It is the reason for much emotional and financial strain on the individuals affected and therefore a considerable amount of research has been aimed at the treatment of this disease, however as the pathology of AD is not fully understood it has made the discovery of new treatments difficult.

1.2 Causes of Alzheimer's

Two types of abnormal protein aggregates are associated with AD; the aggregation of the protein fragment beta-amyloid (Aβ) and tangles of the tau protein inside neurons (NFT) (Estrada and Soto, 2007).

Amyloid refers to fibrillar aggregates that are of a β-sheet conformation that share morphological characteristics. Some diseases involving the accumulation of amyloid deposits (protein misfolding disorders) include AD, Parkinson's disease, Huntington disease, transmissible spongiform encephalopathies, serpin deficiency disorders, secondary amyloidosis, diabetes type II, Amyotrophic Lateral Sclerosis and dialysis-related amyloidosis. Protein misfolding disorders (PMD) are characterized by the misfolding, aggregation and tissue deposition of an otherwise normal protein (Estrada and Soto, 2007). The misfolded proteins occur in various cellular compartments. These compartments include the cytoplasm, nucleus and endoplasmic reticulum (ER) (Ciechanover and

Kwon, 2015). These aggregates are generally found in the region of the hippocampus and the neocortex as well as in the cerebrovasculature (CAA) in the brain (Chen *et al.*, 2017).

The 'amyloid cascade hypothesis' suggests aggregation of the Aβ peptide in neural tissue to be the key of the AD (Fan *et al.*, 2015). The Aβ protein is derived from the amyloid precursor protein (APP) which is cleaved to form Aβ. In conjunction with APP three other genes have be linked to the disease; apolipoprotein E (apoE, chromosome 19), which is present in the earliest stages of plaque formation, presenilin 1 (PSEN1 chromosome 14) and presenilin 2 (PSEN2 chromosome 1). Many of these proteins are molecular 'chaperones' that act in enhancing Aβ aggregation (Armstrong, 2014).

The full function of Aβ is unknown however the mutation of it is known to lead to neuronal loss. The aggregation of the Aβ protein fragment can be reversible or irreversible and the aggregates formed can be soluble/insoluble, covalent/non-covalent and native/non-native (Berrill *et al.*, 2011). The main reason for protein aggregation is the decrease in free surface energy by the removal of hydrophobic residues from contact with the solvent. When the aggregates size is large enough to exceed their solubility they become insoluble (Berrill *et al.*, 2011). Misfolded oligomers can influence the misfolding of more Aβ molecules causing aggregation of these oligomers leading to the formation of insoluble plaques outside neurons in the brain. The plaque formation initiates a series of events which result in cognitive decline via synaptic dysfunction, neural loss and general neurodegeneration (Haass and Selkoe, 2007).

The tau protein is a major microtubule associated protein (MAP) found in in neurons in the central nervous system. MAP's main function is the interaction with tubulin and promotion of its transformation into microtubules, and the stabilization of these newly assembled microtubules. The microtubules play an important role as they are an internal support and transport system used to carry nutrients and other essential materials throughout the brain (Gaugler *et al.*, 2016). The tau protein's activity is regulated by its level of phosphorylation with the regular level being 2-3 moles of phosphate per mole of tau protein (Iqbal *et al.*, 2010). AD is caused by the hyperphosphorylation of this protein. The phosphorylation of tau proteins found in AD suffering individuals is 2-3 times higher than normal. A hyperphosphorylated tau protein is polymerized into paired helical fragments (PHF) and neurofibrillary tangles (NFTs). The PHF-tau is responsible for dissembling microtubules and other versions of tau such as MAP1, MAP2, and ubiquitin which are important in normal neuronal function (Iqbal *et al.*, 2010).

Even with both hallmarks of AD not yet being fully understood, the effect of them on the brain is evident. Both cause the damage and destruction of brain cells and the loss of these neuronal connections are proved to be fatal. However, the aggregation of the protein fragment Aβ is more prevalent and therefore this was decided to be the focus of this study.

1.3 Treatment

AD involves the loss of acetyl-choline-releasing neurons in brain areas that are related to memory. One of the available treatments involves cholinesterase inhibitors which prevent the hydrolysis of the critical neurotransmitter acetylcholine and boosts the levels of cell communication (Estrada and Soto, 2007). The results of this treatment are mild, but it can improve the symptoms of depression and agitation. However, it also has side effects such as diarrhea, nausea, loss of appetite and sleep disturbances. The cholinesterase inhibitors used include donepezil (Aricept), galantamine (Razadyne) and rivastigmine (Exelon) (Gaugler *et al.*, 2016).

Another form of treatment involves the use of memantine (Namenda) which slows the progression of symptoms. It is a non-competitive N-methyl-D-aspartate (NMDA) channel blocker that reduces the activity of the neurotransmitter glutamate, which plays a role in learning and memory by binding to the NMDA receptor (Chen *et al.*, 2017). This drug, however, also comes with side effects such as constipation, dizziness and headaches (Gaugler *et al.*, 2016).

The most recent drug approved by the Food and Drug Administration (FDA) in 2014 is Namzaric which is a combination of the two types of drugs that reduce the levels of both cholinesterase and glutamate (Chen *et al.*, 2017). Table 1 summarizes current treatments for Alzheimer's disease.

Table 1 : Summary of the available treatments for AD

Name	FDA approved	Targets
Donepezil (Aricept)	1996	Cholinesterase inhibitor
Rivastigmine (Exelon)	2000	Cholinesterase inhibitor
Galantamine (Razadyne)	2001	Cholinesterase inhibitor
Memantine (Namenda)	2003	NMDA receptor antagonist
Donepezil and Memantine (Namzaric)	2014	Cholinesterase inhibitor and NMDA receptor antagonist

Recently a novel drug candidate, Wgx-50, was discovered by Fan and colleagues (2015) which has desirable binding results to an aggregated set of Aβ proteins. Earlier known as gx-50, N-[2-(3,4-dimethoxyphenyl)ethyl]-3-phenyl-acrylamide was isolated from extracts of Sichuan pepper (*Zanthoxylum Bungeanum*) (Fan *et al.*, 2015).

Both *in vivo* and *in vitro* experiments have been performed to determine whether Wgx-50 has therapeutic effects on AD. The in vivo studies involved techniques such as pharmacokinetic assays, cognitive abilities tests, and immunohistochemical analyses of brain sections of transgenic mouse model (Hou *et al.*, 2017). Wgx-50 was shown to be able to pass through the blood brain barrier, improve the cognitive abilities of mice, and decrease the accumulation of Aβ oligomers in the cerebral cortex. The *in vitro* studies used atomic force microscopy of Aβ oligomers and cell apoptosis assays to provide evidence of the disassembling effect of Wgx-50 on Aβ oligomers. The *in vitro* studies also showed that Wgx-50 inhibits Aβ-induced neuronal apoptosis and has anti-inflammatory effects by counteracting Aβ-triggered microglial over activation (Hou, Gu and Wei, 2017). Microglial cells protect the central nervous system (CNS) from brain injury or immunological stimuli. However, over-activation of the microglia can enhance the inflammatory effects and mediate cellular degeneration which leads to the death of neurons (Peng *et al.*, 2015).

Based on the results of Fan and colleagues (2015) the stable binding lead to the destabilization of the aggregate. This potential drug is used as a reference in this research.
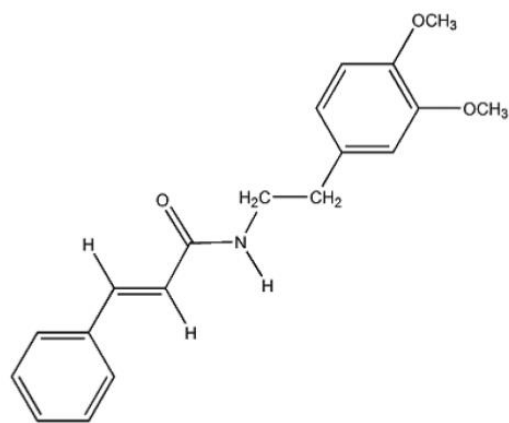
Figure 1: Diagram of the structure of Wgx-50 obtained from Fan and colleagues (2015).

1.4 Amyloid Beta

1.4.1 Amyloid precursor protein

As mentioned previously Aβ is encoded on the chromosome 21 as part of a larger protein named amyloid precursor protein (APP) which is a transmembrane glycoprotein (Estrada and Soto, 2007). APP is created in the endoplasmic reticulum (ER) and then transported to the Golgi, where it completes maturation and is transported to the plasma membrane (Chen *et al.*, 2017). APP can be found in many tissues including within the synapses of neurons. APP consists of a single membrane-spanning domain, a large extracellular glycosylated N-terminus and a shorter cytoplasmic C-terminus (Chen *et al.*, 2017). It is one of three members of a gene family in humans with the other two members being APP-like protein 1 (APLP1) and APP-like protein 2 (APLP2). It has been linked to synaptic formation regulation and repair, anterograde neuronal transport and iron export. APP can be produced with varying lengths ranging in size from 695 to 700 amino acids with APP695 mainly expressed in neurons and APP751 and APP770 mainly expressed on peripheral cells and platelets. APP695 is the isoform mostly found in the brain and the difference between this isoform and its counterparts is the lack of the Kunitz-type serine protease inhibitory domain (KPI) sequence in its ectodomain (Figure 2) (Chen *et al.*, 2017).

APP is cleaved by β-secretase and γ-secretase (proteolysis required for the amyloidogenic pathway discussed in a later section). As a result of γ-secretase providing non-specific cleavage of APP various Aβ proteins are formed with different residue lengths (39-43 residues) (Estrada and Soto, 2007). The Aβ (1–42) variation exhibits higher toxicity and tendency for aggregation than the more abundant 40-residue Aβ (1–40) variation. When compared, the increased production of Aβ (1–42) over Aβ (1–40) has been linked with early onset of AD due to numerous pathogenic mutants of γ-secretase (Xiao *et al.*, 2015).
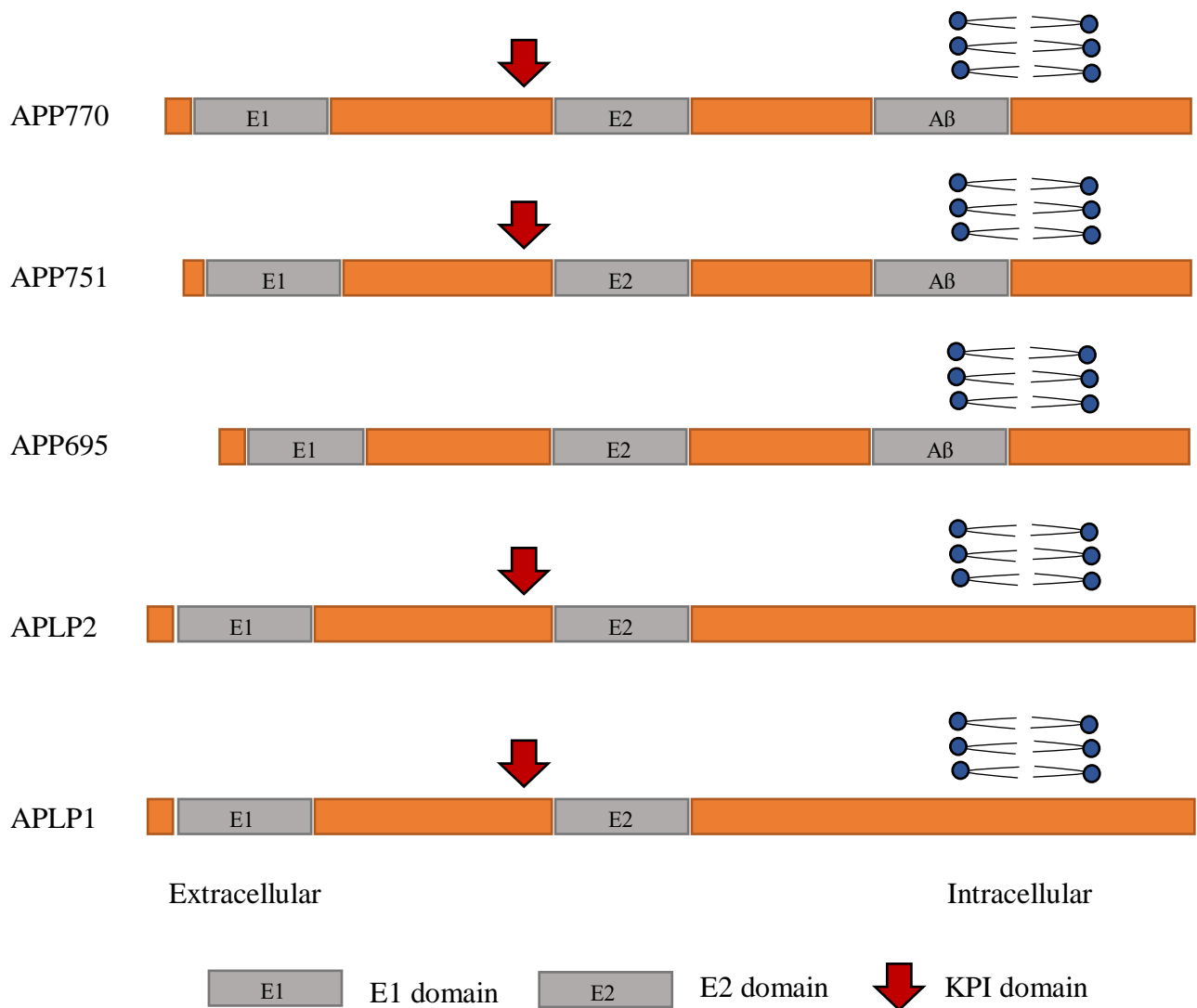
Figure 2: Genotype of the various forms of APP adapted from Chen and colleagues.

1.4.2 Structure

The Aβ (1–40) structure is characterized by a U-shaped β-sheet strand; (loop, β-sheet strand or "β-arch" motif). Two parallel β-sheets are connected by a short-curved loop region (between residues Asp23 and Gly29), often with stabilization by a salt-bridge between Asp23 and Lys28 side-chains (Estrada and Soto, 2007; Fan *et al*., 2015; Xiao *et al*., 2015). The N-terminal portion of Aβ is hydrophilic, whereas the C terminus amino acids 29–42 are rich in hydrophobic residues and originate from the transmembrane region of the amyloid precursor protein (Syme *et al.*, 2004).

However, the structural details for the more pathogenic Aβ(1–42) fibril are poorly defined despite exhaustive efforts. Aβ is commonly thought to be intrinsically unstructured and hence is unable to be

crystallized by common methods. Its high misfolding propensity and structural and morphological heterogeneity also limits analysis of this fibril (Xiao et al., 2015).

However, there are studies that have optimized conditions for stabilizing the aggregates for analysis and structures of the Aβ aggregate derived from both solid and liquid state nuclear magnetic resistance (NMR) techniques are available on the Protein Data Bank (PDB) (Berman *et al.*, 2000).

Many of the structures in the PDB are created using X-Ray crystallography which enables the atomic-level visualization of protein structures. X-ray crystallography utilizes very high energy X-rays ($\lambda \sim$ 0.1 nm) as the electromagnetic radiation to analyze the structure of the protein of interest (Henzler-Wildman and Kern, 2007). The molecular electron clouds and X-rays interact causing the deflection and scattering of the X-rays. The deflected beams can either interfere constructively or destructively and the effect of this interference creates a scattering distribution proportional to the scattering angle (diffraction). This process is described by Bragg's law which observes that the angle of diffraction is inversely proportional to the interplanar spacing of the crystal lattice. The diffraction pattern of the crystal is the Fourier transform of its structure. The diffraction pattern is reverse-Fourier transformed to obtain a structure without phase information. Phase information is obtained through either molecular or isomorphous replacement or anomalous dispersion techniques, which is then back-calculated through Fourier transforms into the electron density, giving form to a structural model of the protein (Lorieau and McDermott, 2006).

NMR focuses rather on the absorption of electromagnetic radiation in the radio-frequency (RF) range (10 MHz - 1 GHz) by atomic nuclei placed into a strong external magnetic field (B_0). The magnetic field introduces energy levels associated with possible alignments of the nuclei relative to the magnetic field, and the RF excites the nuclear spins to higher energy levels. After excitation of nuclei by a RF pulse signals (at the Larmor frequency of the nucleus observed, associated with the difference in energy level) are detected as the nuclei return to their lower energy state. This detected radio frequency signal is amplified by the NMR spectrometer and Fourier transformed to produce the NMR spectrum (Lorieau and McDermott, 2006).
.
A major difference between these two techniques is how the data is analysed in each. In X-Ray crystallography an electron density map is generated by indexing and analyzing the diffraction pattern using Fourier transforms. The protein structure can then be completed after refinement. X-Ray crystallography produces a single structure for the protein. By comparison, under NMR proteins are normally analysed in solution and this technique only provides structural restraints. The 20 to 40

lowest energy structures satisfying these restraints are provided. This solution state analysis allows for the native movement of the structures, and the resultant data collected is representative of a large population of a protein's many conformations (Lorieau and McDermott, 2006).

There are many different structural versions of the Aβ peptide available on the PDB website. These structures range in residue length, and observation of the available structures provides an idea of how the different residue lengths affect the overall structure. When looking at models consisting of Aβ(1-28) isomers, their structures present as α-helical structures with a β-sheet conversion in membrane-like solution (Figure 3A). This structure is the main component of amyloid deposits in AD. The solution structure created of Aβ (1-40) suggests an α-helical structure at the C-terminus (residues 15-36) with a connecting structure at residues 25-27 while the residues 1-14 are unstructured and are likely solvated by water due to their polarity (Chen *et al.*, 2017). Deprotonation of two acidic amino acids leads to the creation of a helix-to-coil formation that precedes the aggregation of Aβ(1-40) (Figure 3B). Solid-state NMR models of the Aβ peptide (10-35) show that in some systems the peptide changes its conformation to a series of loops, strands and turns without α-helical or β-sheet structure (Figure 3C). Van der Waals and electrostatic forces seem to maintain this conformational stabilization. The surface of the structure is partly uninterruptedly hydrophobic, and the compact coil is meta-stable, which may lead to conformational change. Formation of an intermolecular β-sheet secondary structure could be associated with fibrilization. The 3D NMR structures of Aβ(8-25) and Aβ(28-38) show two helical regions joined by a regular type I β-turn (Figure 3D). Aβ(25-35) is a highly toxic version of the Aβ peptide (Figure 3F). The peptide behaves as a transmembrane helix in a lipidic environment and forms fibrillar aggregates which suggests a direct mechanism of neurotoxicity (Chen *et al.*, 2017).
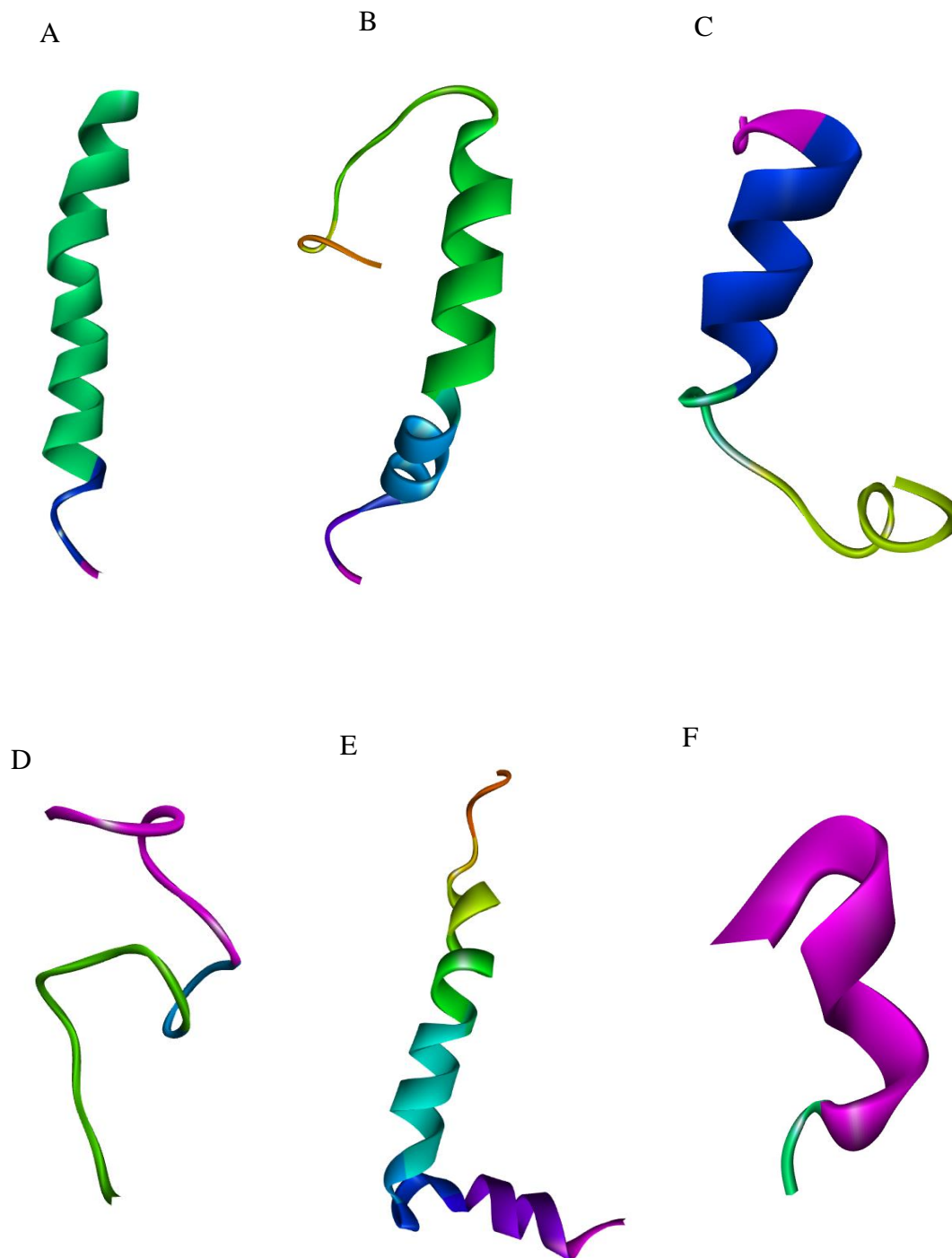
Figure 3: Various monomeric structures of the Aβ peptide generated by Discovery Studio. A) Aβ(1-28) generated by NMR (PDB ID: 1AMC). B) Aβ(1-40) with Met(O) generated by NMR (PDB ID: 1BA4). C) Aβ(10-35) generated by NMR (PDB ID: 1HZ3). D) Aβ(1-42) generated by NMR (PDB ID: 1IYT). E) Mutant of Aβ(1-28) generated by NMR that could stabilize the helix and promote fibrilization and aggregation (PDB ID: 1BJB). F) Aβ(25-35) generated by NMR (PDB ID: 1QWP).

Aβ40 and Aβ42 both are found in the plasma and cerebrospinal fluid at a concentration ratio of 10:1 respectively, however, Aβ42 is deposited first during the development of AD and is more neurotoxic than Aβ40. An increase in the Aβ42/Aβ40 ratio is associated with early onset familial AD (Bitan *et al.*, 2003). NMR Simulations, which include NMR constraints of Aβ(1-40) and Aβ(1-42), suggest that the two have different conformational states. Originally, studies proposed a "cross β" patterned structure for the amyloid polypeptide involving the folding of adjacent chain segments in an anti-parallel manner with fiber lattice. Later it was revealed that the peptide chains of the β-strand segments run perpendicular to the fibril and the intermolecular hydrogen bonds of the β-strands run parallel to the axis (Chen et al., 2017). The 3D structure of residues 15-42 of Aβ42 adopts a cross-β-sheet formation with buried hydrophobic side chains in which residues 1-14 are partially ordered. These residues 1-14 are in a β-strand conformation in Aβ42 which aggregates much faster and dominates in plaque in Alzheimer's disease patients. Studies show that the C-terminus of Aβ42 has less flexibility due to the β-hairpin formation of residues 31-34 and 38-41. This may be a reason for the more fibrillogenic nature of Aβ42 and its ability to form amyloids (Chen *et al.*, 2017).

Two structures were identified from the PDB as ideal structures for this study. One is a mature amyloid fibril (PDB ID: 2MXU) derived from solid-state nuclear magnetic resonance (SS-NMR) (Xiao et al., 2015) and the other is a 3D Structure of Alzheimer's Aβ (1-42) fibrils (PDB ID; 2BEG) constructed using solution NMR (Luhrs *et al.*, 2005). The primary amino acid sequence of Aβ was initially discovered from extracellular deposits and amyloid plaques in 1984 with the sequence being DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA (Chen *et al.*, 2017). The amino acid sequence for 2MXU is EVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA and the amino acid sequence of 2BEG is LVFFAEDVGSNKGAIIGLMVGGVVIA.. The reason for the missing 1-16 residues in the 2BEG model is due to the residues of each monomer being disordered (Luhrs *et al.*, 2005).

The 2MXU structural model presented by Xiao and colleagues (2015) was obtained after incubating an Aβ(1–42) solution for 24 hours with the addition of seeded amyloid fibrils. The morphology was then observed using transmission electron microscopy (TEM). The morphology shows a unique triple-β motif, which is made of three β-sheets using residues 12–18 (β1), 24–33 (β2), and 36–40 (β3) that differ from the Aβ (1–40) variation (Figure 4). This structure also contains a salt bridge between Lys28 side chain and Ala42 carboxyl terminus (Xiao *et al.*, 2015). On the other hand, the 2BEG structure adopts a strand-loop-strand motif, consisting of two β-sheets whose side chains join each other in an antiparallel way. The direction of the backbone hydrogen bonds in 2BEG are parallel to the fibril axis, with the β-strands perpendicular to this. Each U-shape peptide consists of an N-terminal

β-strand (β1) encompassing residues V18−S26, a C-terminal β-strand (β2) including residues I31−A42, and a loop (residues N27−A30) connecting them (Figure 4). The loop region of residues 27–30 is connected to sheet β1 by the salt bridge at Asp23–Lys28, which also forms bonds to residues Ile32 and Leu34 of sheet β2 (Luhrs *et al.*, 2005).
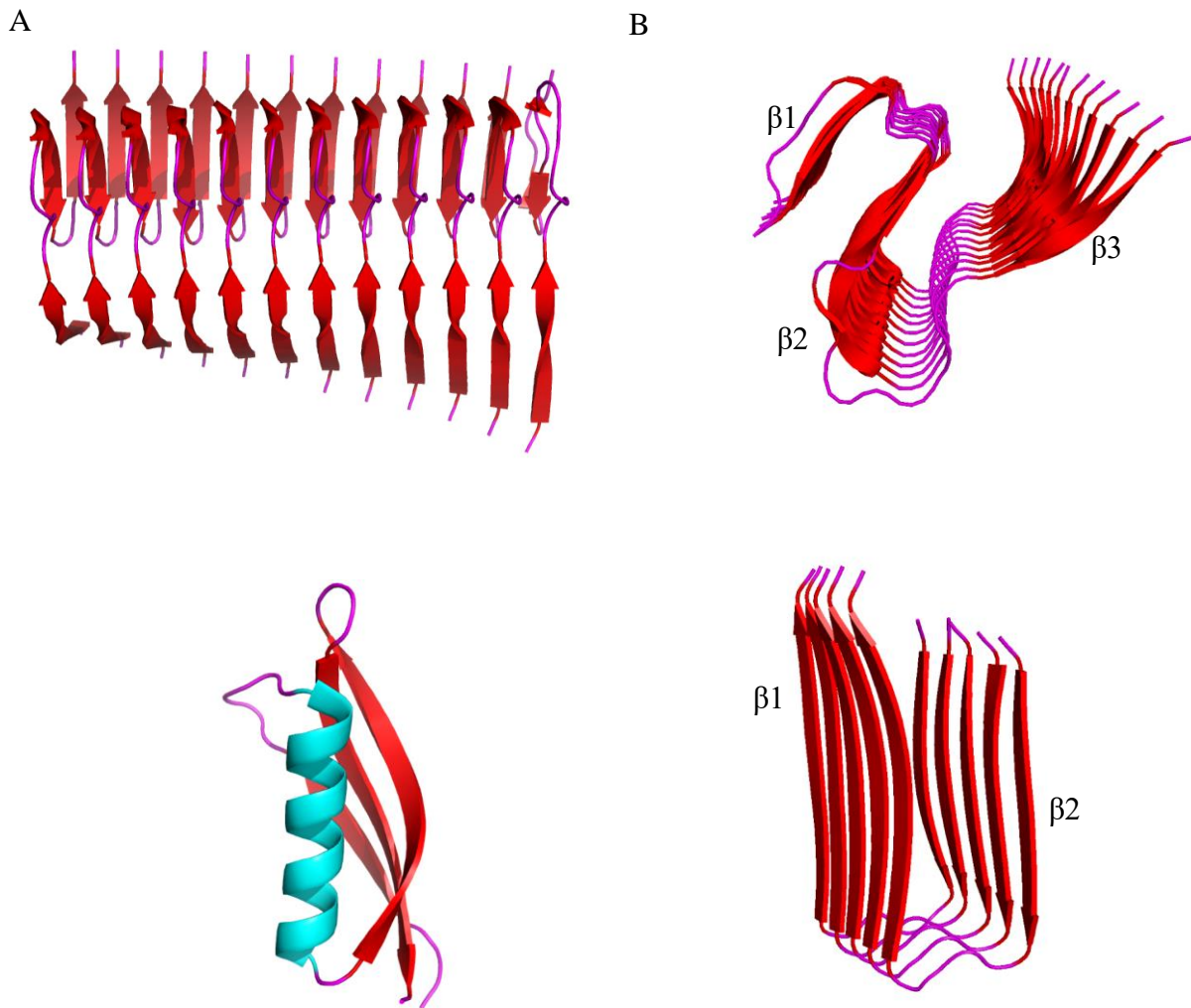

Figure 4: Images of various structures obtained on PyMol. A) A mature Beta Amyloid Fibril (PDB ID: 2MXU). B) Side view of 2MXU. C) Alzheimer's Amyloid Precursor Protein (APP) (PDB ID: 2FK2). D) 3D Structure of Alzheimer's Aβ (1-42) fibrils (PDB ID; 2BEG).

1.4.3 Proteolysis and the amyloidogenic pathway

Studies have shown the correlation between the mutations of APP and AD. There are two main pathways which APP is involved in; the amyloidogenic (producing Aβ) and the nonamyloidogenic (does not produce Aβ) pathway (Estrada and Soto, 2007; Armstrong, 2014) (Figure 5).

The nonamyloidogenic pathway involves the enzyme α-secretase, a metalloprotease enzyme. APP is cleaved by α-secretase between positions 16 and 17 of Aβ to produce αAPPs and an 83-residue COOH-terminal fragment (CTFα). The γ-secretase cleavage of CTFα produces the more benign p3 fragment instead of Aβ (Murphy and Levine, 2010).

On the other hand, the amyloidogenic pathway, involves the enzymes β-secretase and γ–secretase. β-Secretase is a membrane-bound aspartyl protease; however, it cleaves APP outside the lipid bilayer. βACE1 (Beta APP cleaving enzyme 1) and βACE2 (Beta APP cleaving enzyme 2) are the two major forms of the enzyme. βACE1, the form more associated with Aβ production, is highly expressed in brain. The second form, βACE2, is low in the brain but is present in most peripheral tissues. βACE1 cleaves APP on the amino side of Aβ releasing large, more soluble secreted derivative (sAPPβ). What remains is a 99-residue membrane associated COOH-terminal derivative (C99) (Estrada and Soto, 2007; Murphy and Iii, 2010; Armstrong, 2014).

The C99 product is sequentially cleaved by γ-secretase. γ-Secretase is a multisubunit enzyme composed of the proteins APH1, PEN2, nicastrin, and presenilin (PS1 or PS2). The enzyme complex is responsible for the cleavage of APP as well as other membrane proteins. γ-Secretase cleaves within the lipid bilayer and can only process substrates that are first cleaved by another protease. The cleavage by γ-secretase is often imprecise, therefore, many different Aβ species exist all with different residue lengths. The main two forms of Aβ formed by γ-secretase cleavage are; the more soluble Aβ40 or the more hydrophobic and fibrillogenic Aβ42 found largely in discrete Aβ deposits (Estrada and Soto, 2007; Murphy and Iii, 2010; Armstrong, 2014). The newly generated Aβ is either is released to the extracellular space or remains associated with the plasma membrane and lipid raft structures.

The binding of Aβ to ganglioside GM1 present in the lipid rafts favors the aggregation of the peptide. The binding of ApoE to Aβ taken up by the cells through receptor-mediated endocytosis mediated by LDL-receptor-related protein (LDL) and LDLR regulates aggregation but also the cellular uptake of Aβ. Once endocytosed, Aβ has access to other subcellular compartments through the vesicular transport system (Chen *et al.*, 2017).

Mutations of genes PSEN1 and PSEN2 have been linked to early-onset familial AD. The PSEN protein is composed of nine trans-membrane domains located in the membrane of the endoplasmic reticulum (ER). The endoproteolytic cleavage and assembly of PSEN into γ-secretase is said to potentially affect APP processing. The mutant PSEN1 could enhance 42-specific-γ-secretase cleavage of normal APP resulting in increased accumulation of aggregate forming species (Armstrong, 2014).
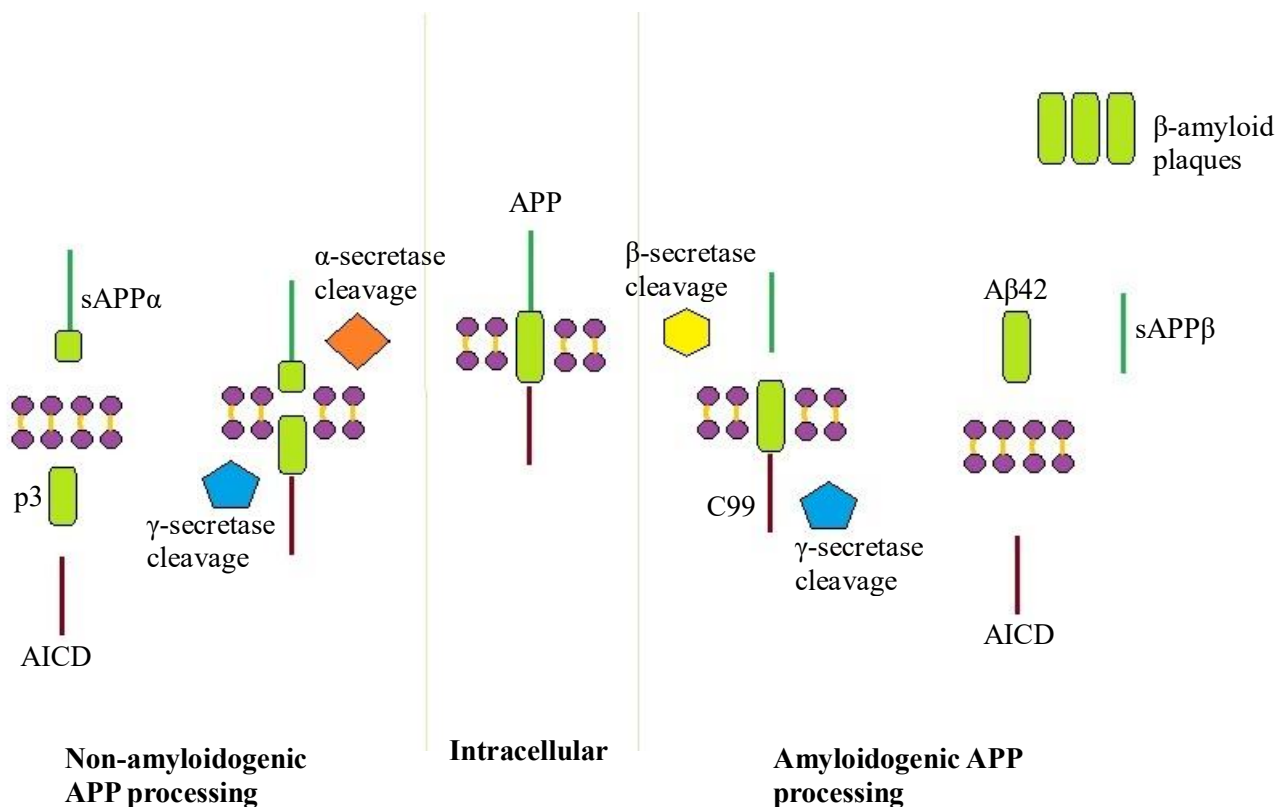
Figure 5: Diagram showing the non-amyloidogenic and amyloidogenic pathways (Adapted from Biolegend.com, 2018)

1.4.4 Aggregation of Aβ into fibrils

Earlier studies pointed to Aβ fibrils as the neurotoxic agent leading to cellular death, memory loss, and other AD characteristics. Over the last two decades, further investigation has suggested that oligomeric or prefibrillar species of the Aβ peptide are the most damaging to neuronal cells (Chen *et al.*, 2017).

Three intermediates in the process of Aβ aggregation have been determined; oligomers, paranuclei and protofibrils. The small oligomers are soluble structures formed by the collection of misfolded Aβ monomers via non-covalent interactions. Once the oligomers have reached a critical concentration protofibrils are formed. Protofibrils are aggregates of a more fibrillar nature with a 4–10 nm diameter and a length of 200 nm. Protofibrils are larger and more insoluble and culminate in mature fibrils (Murphy and Levine, 2010; Chen *et al.*, 2017).

Aβ monomers can form various types of assemblies which can be low molecular weight oligomers such as dimers, trimers, tetramers and pentamers and range to higher molecular weight oligomers such as hexamers, nonamers and dodecamers to protofibrils and fibrils (Figure 6). Recent studies of structure-activity relationships among fibril assembly intermediates have revealed that many intermediates are neurotoxic, including dimers, trimers, and protofibrils (Bitan *et al.*, 2003). According to the study performed by Ono and colleagues (2009) they stated that dimers were ~3-fold more toxic than monomers and tetramers were ~13-fold more toxic. This study showed a correlation between the structure of the intermediate and its toxicity (Ono *et al.,* 2009).

Information on the amyloid oligomers is limited unlike the fibril structure. The different structural versions of the oligomers share a common structure and mechanism of toxicity. The oligomers initially appear as spherical aggregates then elongate in a bead-like formation preceding the formation of protofibrils which undergo maturation to form fibrils (Chen *et al.*, 2017). Aβ40 and Aβ42 have been shown to have different formations of oligomers with Aβ42 having the ability to form fibrils substantially faster than Aβ40. The early assembly of Aβ42 involves formation of pentamer/hexamer units called paranuclei (Figure 6C). The paranuclei then self-associate into larger oligomers, which give rise to protofibrils. In contrast early Aβ40 assembly produces a mixture of monomer, dimer, trimer, and tetramer assemblies. These differences are likely to highlight the distinct characteristics of the two peptides (Bitan *et al.*, 2003).

The preparation of oligomers is complicated as a result of their states being more transient than fibrils. They can be stabilized by detergents and when prepared in the presence of these detergents the oligomers adopt a β-sheet conformation with mixed parallel and antiparallel features. In 2010, it was discovered that low temperatures and the presence of salt in solution made it possible to isolate pentameric disc shaped oligomers. Circular dichroism (CD) and infrared spectroscopy showed that Aβ oligomers are extended coil or β-sheet structures. Further analysis indicated a stable core with 40% of the total backbone hydrogens being resistant to exchange in oligomeric conformation. This contrasts with 50% of the backbone hydrogens being resistant to exchange in the mature amyloid

fibril. This 10% difference shows that a small increase could lead to the transformation to a more fibrillar conformation (Chen *et al.*, 2017).

The structures of the oligomers and the fibrils seem to have some similarities as they both are extended or β-sheet structures and both display a similar percentage of main chain hydrogen bonding that is resistant to exchange. Oligomers are an intermediate appearing at early stages of the devolvement's of fibrils and have been shown to be toxic to neuronal cells at nanomolar levels *in vitro* (Bitan *et al.*, 2003).
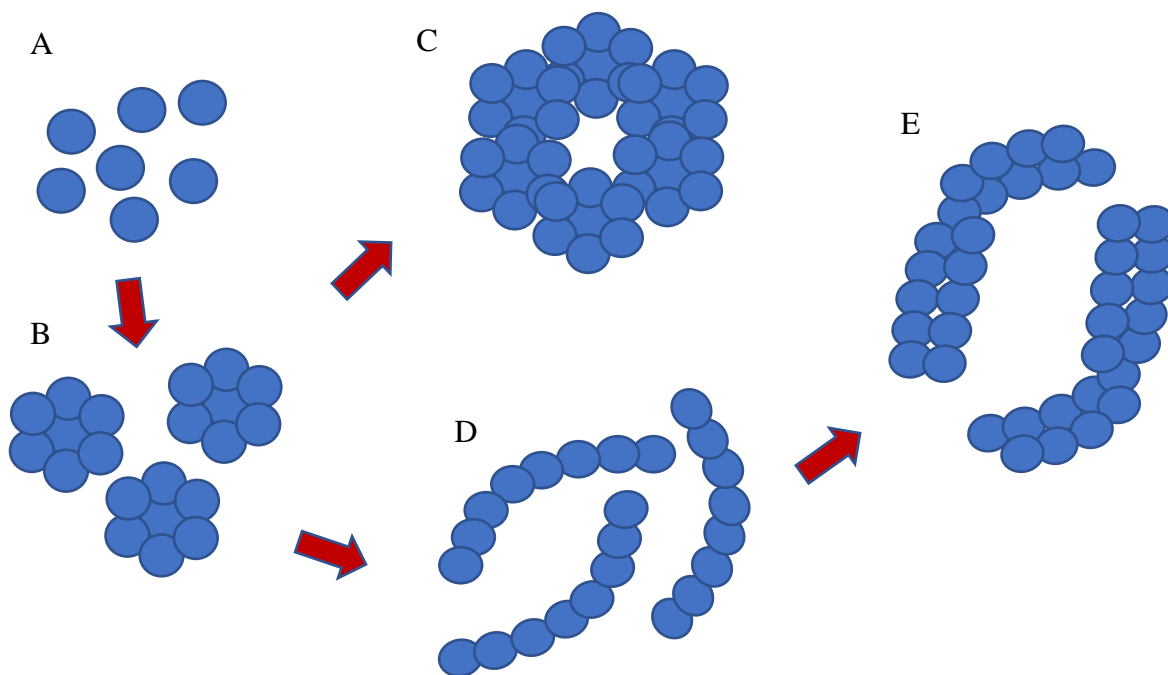
Figure 6: Various structures of the Aβ peptide. A) Misfolded Aβ monomers. B) Oligomers. C) Paranucleus. D) Protofibrils. E) Mature Fibrils.

1.5 Prevention of Aβ aggregation as a therapeutic treatment of AD

Soluble Aβ is a protein that is produced by many cell types and is a normal constituent of biological fluids. This suggests that APP processing is not required for amyloid formation. Aβ in its native form is soluble and harmless, therefore, the issue arises with Aβ misfolding and the formation of aggregates (Murphy and Iii, 2010).

The drugs currently used to treat AD have limited therapeutic value. New, potentially disease-modifying, therapeutic approaches are aimed at targeting Aβ. Targeting the inhibition, reversion and elimination of Aβ aggregation could prevent the disruption of neuronal cells and could ultimately prevent AD (Reitz, 2012). The screening of large libraries of chemical or natural compounds could identify several diverse small molecules have with ability to prevent Aβ fibrillogenesis. Finding a compound that can bind to the Aβ region necessary for the peptide self-assembly could prevent the aggregation of the protein and could lead to the discovery of a novel treatment for AD (Estrada and Soto, 2007).

1.6 Copper and Aβ

It has been proposed that metal ions play a role in the pathogenesis of AD. Copper (Cu), Zinc (Zn) and Iron (Fe) have been found in high concentrations in and surrounding AD plaques in the brain. The interactions of Aβ with transition metals have revealed potential pathogenic interactions and structural consequences. Oligomers that may normally be embedded in the membrane bind to transition metals such as Cu, Zn and Fe. Constitutively, metal-bound plaques play a role in accelerating the aggregation of amyloid beta peptide (Chen *et al.*, 2017). Aβ has selective high and low affinity Cu^{2+}- and Zn^{2+}- binding sites that mediate its aggregation with Aβ42 having the greatest affinity for Cu^{2+} (Maynard *et al.*, 2005).

Electron paramagnetic resonance and nuclear magnetic resonance studies proposed a model of monomer Aβ binding to a Cu ion via three histidines and a tyrosine. At neutral pH, Zn^{2+} binds to Aβ to form insoluble aggregates, while Cu^{2+} binding induces a soluble conformation due to the binding being more competitive (Maynard *et al.*, 2005). In a more aged patient, the levels of Cu increase and the pH becomes more acidic, this is also as a response to inflammation. The elevated copper levels and the more acidic pH induces Cu^{2+} to lead to the conversion of Aβ from a functional peptide to a self-aggregating neurotoxin. This abnormal interaction of Aβ and Cu^{2+} leads to Aβ reducing Cu^{2+} to Cu in a catalytic reaction cycle that uses O_2 and biological reducing agents as substrates to generate neurotoxic H_2O_2 (Bush, Masters and Tanzi, 2003). This highlights the importance of the balance of Zn and Cu concentrations, as well as the maintenance of physiological pH in the prevention of Aβ aggregation and amyloid formation (Syme *et al.*, 2004; Maynard *et al.*, 2005).

One of the most significant studies that link copper with AD has shown that metal chelators specific to Cu^{2+} reversed the aggregation of normally insoluble amyloid deposits and solubilized amyloid in

postmortem human brain specimens. Studies have also shown the link between the use of Cu^{2+} chelators and inhibition of amyloid accumulation in AD transgenic mice (Maynard *et al.*, 2005; Syme *et al.*, 2004).

However, despite an increasing body of evidence to link Cu^{2+} with AD, the precise coordination geometry and the residues involved in Cu^{2+} ligation are yet to be established, therefore it would be beneficial for this study to further investigate the role of Cu^{2+} and Aβ destabilization. In order to do this a computational approach such as molecular dynamics could be adopted to understand the interactions between Cu^{2+} and Aβ.

However, the issue arises in molecular dynamics as a result of the force field for Cu^{2+}, specifically in the context of the system we are investigating, not being available. Force fields are available for a single chain; however, the system is that of an aggregate therefore the binding will differ. This poses as an issue, however there are techniques such as an ab initio or quantum mechanics (QM) approaches or using van der Waals forces and electrostatic interactions to simulate binding of the metal to the protein.

Techniques such as an *ab initio* or quantum mechanics (QM) approaches or using van der Waals forces and electrostatic interactions can be used to simulate binding of the metal to the protein. The van der Waals forces and electrostatic interactions may not be strong enough to keep the ion bound and although this may interfere with correct coordination of the ion, it may in turn give a good idea of the possible biding of the ion to the protein (Moses, Tastan Bishop and Lobb, 2017). Using QM approaches in conjunction with molecular dynamics (MD) can provide information into the binding of Cu^{2+} to the Aβ peptide therefore allowing the relationship between the two to be understood.

1.7 The South African Natural Compound Database (SANCDB) and small compounds in drug discovery

Natural products have been discovered to be an incredible source of therapeutic agents. This is based on their structures having the characteristics of high chemical diversity, biochemical specificity and other molecular properties (Siddiqui *et al.*, 2014). The heteroatoms and chiral centers of natural products make them beneficial for drug discovery as they have specifically evolved to interact with biological macromolecules (Newman and Cragg, 2013). Small molecules have good drug-like

properties as a result of their size and this enables them to be taken orally by patients which maximizes the amount of drug that gets taken to the brain.

The South African Natural Compounds Database (SANCDB) was developed by the Rhodes University Research Unit in Bioinformatics (RUBi) group. The SANCDB is a database containing over 700 fully curated and referenced natural compounds extracted directly from journal articles, book chapters and theses. The database allows entries from researchers, through a submission pipeline, which assists in the growth of the database. As the only web-based natural product database in Africa, the SANCDB aims to provide a useful resource for the *in silico* screening of South African natural products used in drug discovery (Hatherley *et al.*, 2015).

1.8 Problem Statement:

Currently, approximately 50 million people are suffering from dementia worldwide. With AD being one of the main causes of dementia the desire for a treatment is great. However, the only available forms of treatments offer mild calming of the symptoms in conjunction with many side effects. With the disease affecting so many people worldwide finding a treatment that is effective and not toxic to the individual has become imperative.

As a result of the uncertainty related to the causes of AD progress towards drug discovery has been slow. Recently two main hallmarks of AD, the aggregation of beta-amyloid protein and the hyperphosphorylation of the tau protein, have become better understood. These processes have now been targeted by researches to prevent AD.

Recent advances in *in silico* studies and availability of many small compounds in various databases now make it possible to discover new drugs with ability to dissemble the aggregated beta-amyloid protein in a cost-effective manner through high throughput virtual screening. By docking small compounds, obtained from the SANCDB, to the beta-amyloid fibril the aggregation of the beta-amyloid protein could be prevented. Using techniques such as molecular dynamics the possible dissembling of the protein can be determined.

1.9 Hypothesis:

High throughput screening and molecular dynamics of compounds against the target (PDB ID: 2MXU) will result in hits able to destabilize the Aβ aggregate.

1.10 Aims:

The aim of this research is to use *in silico* methods such as; molecular docking and molecular dynamics to screen for compounds that display destabilizing action against the Aβ aggregate. Additionally, the determination of the interactions between copper and the Aβ aggregate will be analysed using available techniques.

1.11 Goals/Objectives:

To successfully implement the proposed research, the following objectives will be carried out:

1. High throughput virtual screening of SANCDB compounds against the Aβ oligomer (PDB ID: 2MXU) using Autodock Vina

2. Docking of the potential drug candidate Wgx-50 to 2MXU using Autodock Vina

3. Molecular dynamics to analyse the results of the dockings using GROMACS

4. Simulation of copper binding to the Aβ oligomer

Chapter 2: High throughput virtual screening

2.1 Introduction

2.1.1 Docking and High-Throughput Virtual Screening

The first step in drug discovery is to identify compounds with potential inhibitory action against the potential drug target. High-throughput virtual screening (HTVS) is a technique widely used to discover new lead compounds for drug design. HTVS is based off high-throughput screening (HTS) which identifies lead molecules by performing individual biochemical assays with over millions of compounds. HTS, however comes with huge cost and is time consuming (Subramaniam, Mehrotra and Gupta, 2008). Nowadays virtual HTS is performed *in silico* allowing one to screen through databases containing thousands of compounds in order to identify hits that could have inhibitory activity against the drug target. HTVS methods have discovered novel molecules that bind to drug targets, and given the efficiency of these methods in producing results, they are under constant development (Bajorah, 2002; McGovern *et al.*, 2002).

HTVS in this context involves a technique referred to as molecular docking. Molecular docking is the process of computationally placing a molecular structure (ligand) into a binding site of a macromolecule (drug target) and scoring the complex based on how the ligand and drug target complement each other (Bleicher *et al.*, 2003). These ligands are small potential drug-like molecules which can prevent the function of the target proteins, and good performing ligands are further optimized to act as a therapeutic drug against a targeted disease (Subramaniam *et al.*, 2008). Molecular docking attempts to predict noncovalent binding of these macromolecules and ligands, starting with their unbound structures obtained from MD simulations, or homology modeling, etc. The aim of molecular docking is to predict where the ligand would bind to the macromolecule, its pose within the macromolecule and determination of the binding affinity. Molecular docking is important because it identifies small compounds with potential inhibitory activity, and screening virtual libraries of drug-like molecules using this technique will result in lead compounds for further drug development (Trott and Olson, 2009).

The docking program used in this study was AutoDock Vina. AutoDock Vina is a commercial, open-source program used to perform molecular docking. AutoDock Vina was developed as the successor of AutoDock 4. AutoDock Vina significantly improves the average accuracy of the binding mode predictions as compared to AutoDock 4. AutoDock Vina was tested against a virtual screening

benchmark called the "Directory of Useful Decoys" created by the Watowich group and was found to be a strong competitor against the other programs (Trott and Olson, 2009).

Docking programs generally use a scoring function, which attempts to approximate the standard chemical potentials of the system. The van der Waals interactions and Coulomb energies are used in the scoring function and are empirically weighted to account for the difference between energies and free energies (Subramaniam *et al.*, 2008). Autodock Vina uses a hybrid scoring function (empirical and knowledge-based) based on the X-Score function with some different parameters which are not currently published (Trott and Olson, 2009).

AutoDock Vina is used in this study to identify ligands with good binding affinities to the aggregate. The hits (compounds exhibiting good binding) will then be taken further to analyse if they have destabilising action against the aggregate.

Druglikeness is a combination of various molecular properties that is used as qualitative concept in drug design to describe how a substance will perform in the human body. This is determined with respect to factors such as hydrophobicity, electronic distribution, hydrogen bonding characteristics, molecule size and flexibility (Nair *et al.*, 2016). Drug-likeness scores help to optimise pharmacokinetic and pharmaceutical properties such as solubility, chemical stability, bioavailability and distribution profile. Drug-likeness is estimated from the molecular structure before the substance is synthesised and tested.

2.1.2. Physico-chemical Properties

A drug, once taken by a patient, needs to be able to; pass through the intestinal lining, be carried by the blood and penetrate the lipid-based cell membrane to reach the inside of a cell. As the drug must pass through the blood (an aqueous media) it must be water-soluble. The logarithm of partition coefficient (logP) of a particle determines its solubility as well as number of hydrogen bond donors and alkyl sidechains in the molecule. A low solubility score means the drug will absorb slowly and therefore it will not be as effective in a short span of time. If there are too many hydrogen bond donors this leads to low fat solubility and means the drug will not be able to penetrate the cell membrane. With regards to molecular weight, the smaller compound is beneficial. This is because size and diffusion are directly related (Vistoli, Pedretti and Testa, 2008).

Lipinski's rule of 5 is used to evaluate druglikeness or to determine if a chemical compound has properties that would make it a likely orally active drug in humans. The five rules are as follows; the compound must have no more than 5 hydrogen bond donors, no more than 10 hydrogen bond

acceptors, the molecular mass must be less than 500 g/mol and its octanol-partition coefficient (logP) must not be greater than 5 (Lipinski, 2004). Three different online sites were used to determine the chemical properties of each of the SANCDB compounds. These sites were Molinspiration, SwissADME and Chemicalize.

2.2 Methods:

2.2.1 Ligand preparation

Seven hundred and twenty eight natural compounds were obtained, already minimized from the SANCDB (Hatherley *et al.*, 2015). The ligands were then prepared for docking using a script (*prepare_ligand4.py -l filename*) from the Autodock Tools suite of software. This script assigns the ligands with the correct atom types and charges necessary for the AutoDock program, defines the relevant torsions for conformational searching, and the ligand files are saved in pdbqt file format. The drug candidate Wgx-50 was drawn using Schrodinger Suites 2018 program Maestro based on the structure (provided by Fan *et al.*). The structure was prepared and minimized within Maestro and was saved as a sdf file. The sdf file then was prepared for docking, first by conversion to pdb format, then further prepared using the same script as for the SANCDB compounds.

2.2.2 Receptor preparation

The focus is the interaction between ligands and the Aβ aggregate. The 42-Residue Beta Amyloid Fibril (PDB ID: 2MXU), obtained through solid-state nuclear magnetic resonance (NMR) (Xiao *et al.*, 2015) was retrieved from the Protein Data Bank (RCSB.org, 2017) as well as the 3D Structure of Alzheimer's Aβ (1-42) fibrils (PDB ID; 2BEG) obtained through solution NMR (Luhrs *et al.*, 2005). The receptors were prepared using a script (*prepare_receptor4.py -r filename*) also from the Autodock Tools suite.

2.2.3 Molecular docking

The docking simulations were performed using the SANCDB compounds and the described recently discovered structure of the aggregate 2MXU. Vina scripts were created for each compound to dock to this receptor (the vina scripts were created from a controlling python script) and these Vina scripts were fed as input to AutoDock Vina (Trott and Olson, 2009) on a Linux cluster. The docking was performed in duplicate for each of three different docking sites. The initial docking (Dock 1) was a targeted docking focusing on the bottom of the U-shape of the protein. Dock 1 (centre) was performed

with a grid box 20 angstrom (Å) in size, centred with coordinates x=0, y=0, z=0 (given the pdb coordinates of the aggregate are centred at the origin). The second docking was performed to focus on the final (L) chain of the aggregate (Dock 2). Dock 2 (targeted to L chain) was performed with a grid box 40 Å in size, centred with coordinates x=22.493, y=-21.342, z=-1.702. The third docking (dock 3) that was performed was a blind docking. Dock 3 (blind) was performed with a grid box 100 Å in size, centred with coordinates x=0, y=0, z=0. All the dockings were performed across 4 CPU cores utilizing an exhaustiveness of 128 (see scripts in appendix). In order to validate the docking results all three of the dockings were performed in duplicate.

2.2.4 Physico-chemical properties

Molinspiration is an online web service that offers free online services for calculation of molecular properties such as logP, polar surface area, number of hydrogen bond donors and acceptors and others, as well as prediction of bioactivity score for the most important drug targets (GPCR ligands, kinase inhibitors, ion channel modulators, nuclear receptors). This site also offers a broad range of cheminformatics software tools such as SMILES and SDfile conversion, normalization of molecules, generation of tautomers, molecule fragmentation, calculation of various molecular properties needed in QSAR, molecular modelling and drug design, high quality molecule depiction, molecular database tools supporting substructure and similarity searches. They also offer fragment-based virtual screening, bioactivity prediction and data visualization (Mabkot *et al.*, 2016).

The Swiss institute of bioinformatics is an academic non-profit organization which offers an online site that provides tools that allows one to compute physicochemical descriptors as well as to predict ADME parameters, pharmacokinetic properties, druglike nature and medicinal chemistry friendliness of one or multiple small molecules to support drug discovery (Daina *et al.*, 2017)

Chemicalize is a free online platform owned by ChemAxon which offers various cheminformatics tools such as chemical property predictions (elemental analysis, names and identifiers (IUPAC name, SMILES, InChI, pKa, logP/logD, and solubility), structure-based and text-based search of chemicals, chemical text processing, and checking compounds with respect to national regulations of different countries (Swain, 2012).

Molecules were submitted to each of these platforms to determine their drug-likeness and physico-chemical properties.

2.2.5 Wgx-50

Wgx-50 was docked to the aggregate (2MXU) following the same procedures as the natural compounds. This drug reference was also docked to the other Aβ fibril available on the PDB site (PDB ID: 2BEG) as a comparison in preparation for MD as a comparison against the results obtained by Fan and colleagues (2015).

2.2.6 Analysis

Seven of the best hits for each docking were identified and molecular graphic images were created of them and the receptor using PyMOL (Delano, 2002). The ligand interactions were determined using DiscoveryStudio version 4.1(BIOvIA, 2015) and LigPlot+ version 1.4.5 (Laskowski and Swindells, 2011). The best compounds SMILES were obtained from the SANCDB and their molecular properties were obtained from three online sites Molinspiration, SwissADME and Chemicalize. The molecular properties of Wgx-50 were also obtained by drawing the structure on the molinspiration site based on the structure published by Fan and colleagues (2015).

2.3 Results and Discussion:

The first step of the drug discovery process is to identify a potential drug target. The Aβ fibril aggregates to form a plaque that prevents the communication of neurons in the brain, leading to neuronal loss. The destabilisation of this aggregate has now been identified to combat AD. In order to study this, a structure of the aggregate needs to be identified. There are many structures available on the PDB however, 2MXU was decided on for this molecular docking study as it is the most recently discovered structure and contains the three β-sheet motif the aggregates are discovered to possess (Xiao *et al.*, 2015). There is no catalytic domain within the aggregate due to it being a collection of monomeric Aβ proteins. As a result of this, three dockings were performed to encompass all possible binding sites the aggregate may possess (Figure 10). The first docking was performed in order to target the area that was the most central to the aggregate. The initial docking showed the interaction of SANCDB compounds with the residues found in β1 (Val12–Phe20), β2 (Asn27–Ile32) and the loop connecting them (Ala21–Ser26). The second docking was performed to target the last and most variable chain of the protein. The docking of the compounds to the last chain could help discover if the compounds would bind to the variable chain in such a way to prevent the binding of another monomer to the aggregate. The third docking was a blind docking that was performed to identify the preferential site of binding for the SANCDB compounds.

Once the dockings were performed using AutoDock Vina (Trott and Olson, 2010), seven of the top SANCDB hits for each of the three docking runs were identified, extracted, and their binding was visualised using PyMol. Dassault Systemes' Discovery Studio program was used to identify the protein-ligand interactions for these SANCDB compounds to the aggregate (BIOvIA, 2015). The interactions shown are those mediated by conventional and carbon hydrogen bonds, Van der Waals forces and alkyl interactions (Table 3-5).
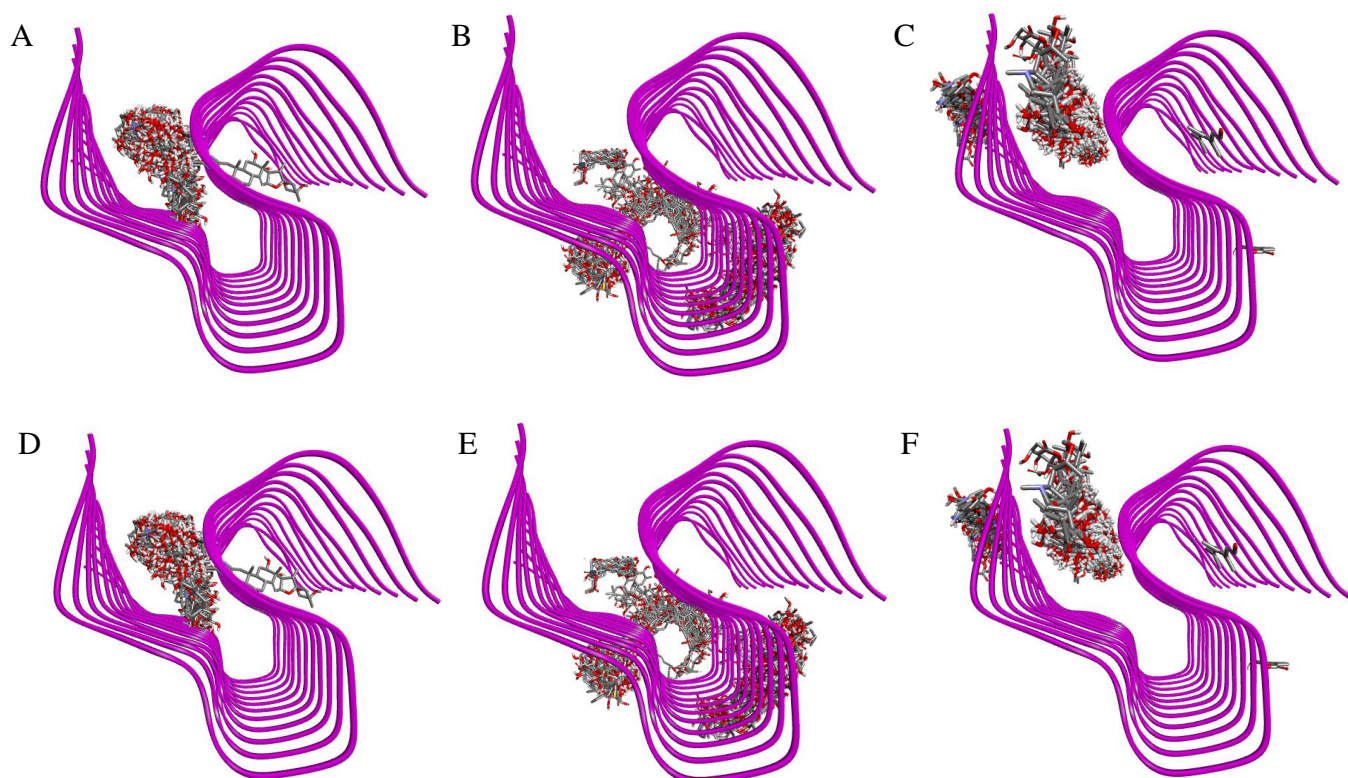
Figure 7: Images of the dockings performed in this study using AutoDock Vina visualised on Discovery Studio. A) Dock 1 (centre). B) Dock 2 (targeted to L chain). C) Dock 3 (blind). D) Duplicate of Dock 1 (centre). E) Duplicate of Dock 2 (targeted to L chain). F) Duplicate of Dock 3 (blind).

The results in Figure 7 show the docking results that were formed. Seven hundred and twenty-eight natural compounds were docked against 2MXU and were performed in duplicate for three different docking sites. The duplicated results were identical to the original dockings and the compounds generally all were bound to similar regions near the targeted sites.

Table 2: Table summarising the docking scores between the receptor and the top hits from all the dockings.

SANCDB compound	Docking score	Dock number
SANC00175	-9,0	1
SANC00290	-9,4	1
SANC00347	-9,1	1
SANC00348	-9,0	1
SANC00518	-9,1	1
SANC00553	-9,2	1
SANC00700	-9,2	1
SANC00178	-9.8	2
SANC00447	-9.5	2
SANC00478	-9.5	2
SANC00480	-9.4	2
SANC00481	-9.9	2
SANC00482	-9.2	2
SANC00486	-9.5	2
SANC00220	-9.8	3
SANC00342	-9.8	3
SANC00384	-9.9	3
SANC00686	-9.9	3
SANC00693	-9.7	3
SANC00700	-9.8	3
SANC00715	-9.8	3

The docking scores are predicted using a scoring function in AutoDock Vina (Table 2). All the docking scores ranged from -9.0 to -9.9 (for the seven best scoring ligands in each case) with SANC00384, SANC00481 and SANC00686 having the lowest binding energies. Dock 3 (blind) contained the lowest set of docking scores and Dock 1 (centre) had the highest, indicating the docking site for Dock 3 (blind) is a more preferred site for strong binding.
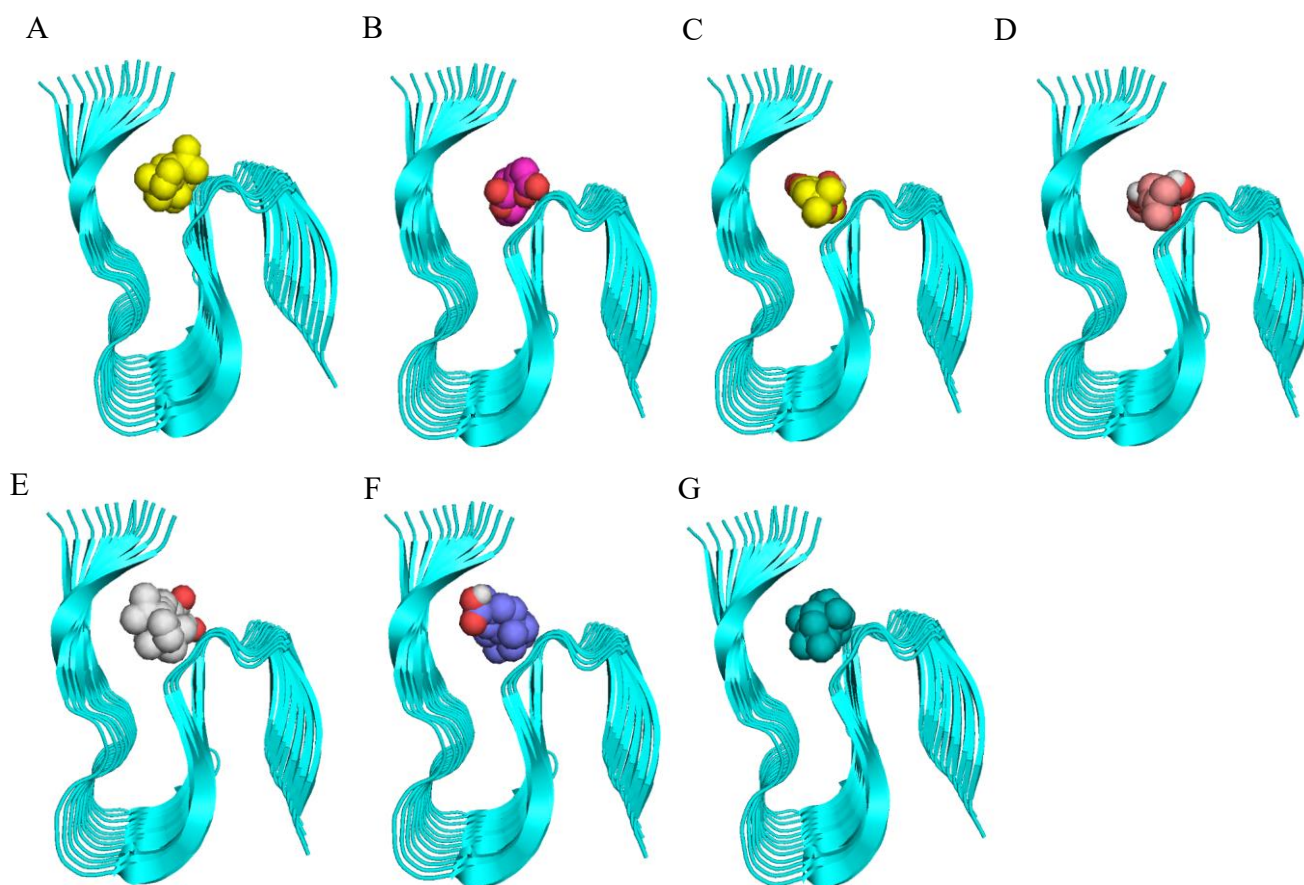
Figure 8: Visual representations of the results of the SANCDB compounds and the receptor 2MXU after Dock 1 (centre) created using PyMol. A) SANC00175. B) SANC00290. C) SANC00347. D) 348. E) SANC00518. F) SANC00553. G) SANC00700.

Table 3: Table summarising the docking score and interactions between the receptor and the top hits from Dock 1 (centre) obtained from Discovery Studio. *Vdw – Van der Waals, HB – hydrogen bonds, () – indicate chain where interaction is present.

Compound	Vdw interactions	Alkyl interactions	Conventional HB	Carbon HB
SANC00175	Leu17(H); Gly33(E,F,H); Leu34(H)	His14(E); Leu17(F, G); Ile(F,G,H); Leu34(E,F,G)		Gly33(G)
SANC00290	Leu17(F,G,H); Ile32(H); Gly33(E,G,H,I); Leu34(I)	Ile32(F,G); Leu34(G,H)	Leu34(F)	Gly33(F); Leu34(E)
SANC00347	His14(F,G); Leu17(F,G); Gly33(D,G,H); Leu34(F)	Leu17(D,E, H); Ile32(D,E,F,G,H)		Gly33(E,F)
SANC00348	Leu17(E,F,G); Ile32(F); Gly33(D,E,F,G,H); Leu34(D,E,F,G)	Leu17(H); Ile(D,E,G,H)		
SANC00518	Leu17(I); Ile32(G,H,I); Gly33(F,G,H); Leu34(F,G)	Val12(E); His14(E,F); Leu17(E,F,G); Ile32(E,F)	Leu34(I)	Leu34(H); Gly33(I)
SANC00553	Val12(D); Leu17(E,F,H); Gly33(E,F,G); Leu34(E,F)	Val12(E); His14(E); Leu17(G); Ile32(F,G); Leu34(G)		
SANC00700	Leu17(G); Ile32(D,F,G); Gly33(E,F,G); Leu34(D,G)	Leu17(E,F); Ile32(E); Leu34(E,F)		

Table 3 shows that majority of the interactions from Dock 1 (centre) are present from chains D-I, the more central chains of the aggregate (Figure 8). Conventional hydrogen bonds were formed between SANC00290 and SANC00518 and 2MXU, however on dissimilar chains. Carbon hydrogen bonds were also formed for SANC00290 and SANC00518 as well as SANC00175 and SANC00347. For Dock 1 (centre) the interactions are formed by residues such as Leu, Val, Gly, His and Ile.
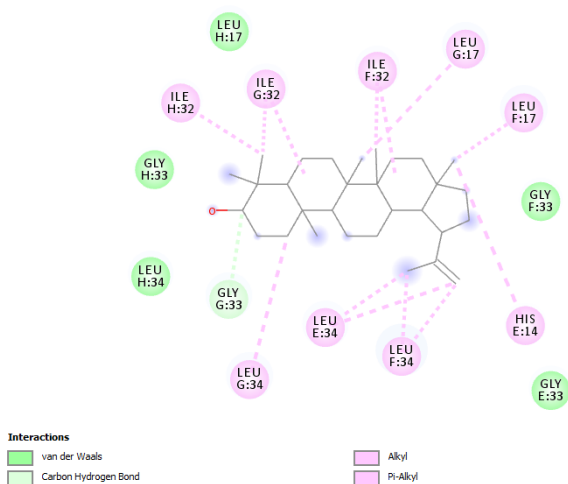
Figure 9: Image obtained as an example of the outputs from Dock 1 (centre) obtained from discovery studio of the interactions between the natural compound SANC00175 and 2MXU.

Discovery studio provides a tool that allows the user to visualise the interactions between the compound and the aggregate. This figure is shown as 2-Dimensional figure in order to fully grasp the various interactions and their placement (Figure 9, 11 and 13). The difference types of interactions present are alkyl, Van der Waals, carbon hydrogen bond and conventional hydrogen bond. Alkyl interactions are non-covalent interactions that form between the electrons of the compounds and receptors (Ribas *et al.*, 2002). Van der Waals forces are a different set of interactions that are deemed to be weak and electrostatic. The Van der Waals forces are distant dependant and are not a result of a chemical bond (Petrucci, 1997). Carbon hydrogen bonds (C-H) are formations of bonds between a carbon and a hydrogen. The carbon shares its outer valence electrons with the hydrogen making it a covalent and relatively stable bond. The strongest of all the bonds, however, is the conventional hydrogen bond. The conventional hydrogen bond (NH-O, OH-O, OH-N, and NH-N) is one of the main stabilising forces in molecular structures. Oxygen and Nitrogen contain higher electronegativity which allows them to form stronger bonds with hydrogens (Scheiner, Kar and Pattanayak, 2002).

Figure 9 shows how the interactions are placed and with there being more alkyl interactions than Van der Waals. There is no presence of any conventional hydrogen bonds and with only one carbon hydrogen bond it does not seem to be a tightly bonded compound to 2MXU.

Figure 10: Visual representations of the results of the SANCDB compounds and the receptor 2MXU after Dock 2 (targeted to L chain) created using PyMol. A) SANC00178. B) SANC00447. C) SANC00478. D) SANC00480. E) SANC00481. F) SANC00482. G) SANC00486.

Table 4: Table summarising the docking score and interactions between the receptor and the top hits from Dock 2 (targeted to L chain) obtained from Discovery Studio. *Vdw – Van der Waals, HB – hydrogen bonds, () – indicate chain where interaction is present.

Compound	Vdw interactions	Alkyl interactions	Conventional HB	Carbon HB
SANC00178	Ser26(G,J,K,L); Asn27(H,J,K); Lys28(G,H)	Lys28(I,J,K)	Ser26(H); Asn27(I)	Ser26(I)
SANC00447	Ser26(G,J,K,L); Asn27(H,I,J,K); Lys28(G,H)	Lys28(I,J,L)	Ser26(H)	Ser26(I); Lys28(K)
SANC00478	Ser26(H,I,JK,L); Asn27(L)	Lys28(I,J,K,L)	Asn27(I); Lys28(H)	
SANC00480	Gly25(L); Ser26(J,K); Asn27(I,L)	Lys28(H,I,J,K,L)	Ser26(H)	Ser26(I,L)
SANC00481	Ser26(I,J,K,L); Asn27(H,I,J,K); Lys28(L)	Lys28(G,H,I,J,K)	Ser26(H)	
SANC00482	Ser26(H,I,K); Asn27(H,I,K,L); Lys28(G)	Lys28(H,I,J,K,L)	Ser26(L)	
SANC00486	Ser26(H,I,J,K,L); Asn27(H,I,J,K); Lys28(L)	Lys28(G,H,I,J,K)		Ser26(L)
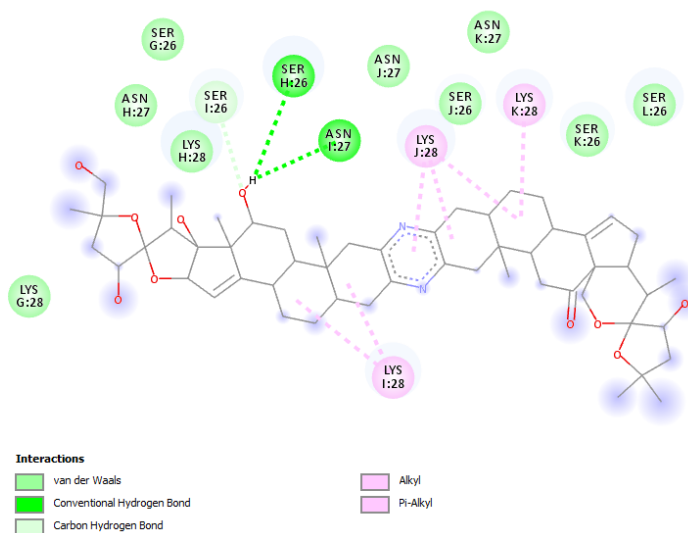
Figure 11: Image obtained as an example of the outputs obtained from discovery studio of the interactions between the natural compound SANC00178 and 2MXU.

Dock 2 (targeted to L chain), which was performed to target the L-chain of the aggregate showed that the compounds were interacting with completely new residues, these being Ser26, Asn27 and Lys28 which are the residues found on the loop region connecting β1 and β2 as well as the residues on β2 (Table 4 and Figure 10). There are more conventional bonds present in this set of compounds for Dock 2 (targeted to L chain) compared to Dock 1 (centre), generally showing tighter bonding. Ser and Asn are both small, polar molecules so this could show there is a higher probability hydrogen bond will form with these types of residues. The other residues that formed hydrogen bonding contacts were glycine, leucine and histidine with histidine also sharing the polar characteristic. This correlates with the slightly lower docking scores seen for Dock 2 (targeted to L chain) when compared to Dock 1 (centre). Xiao and colleagues identified a salt bridge between the residues Lys28 and Ala42 which is said to stabilise the system (Xiao *et al.*, 2015). The interactions of the SANCDB compounds with Lys28 could therefore possibly interfere with the formation of the salt bridge. Ideally interaction with Ala2 would be preferential as it was identified as a stabilizing salt bridge contact (Xiao *et al.*, 2015).
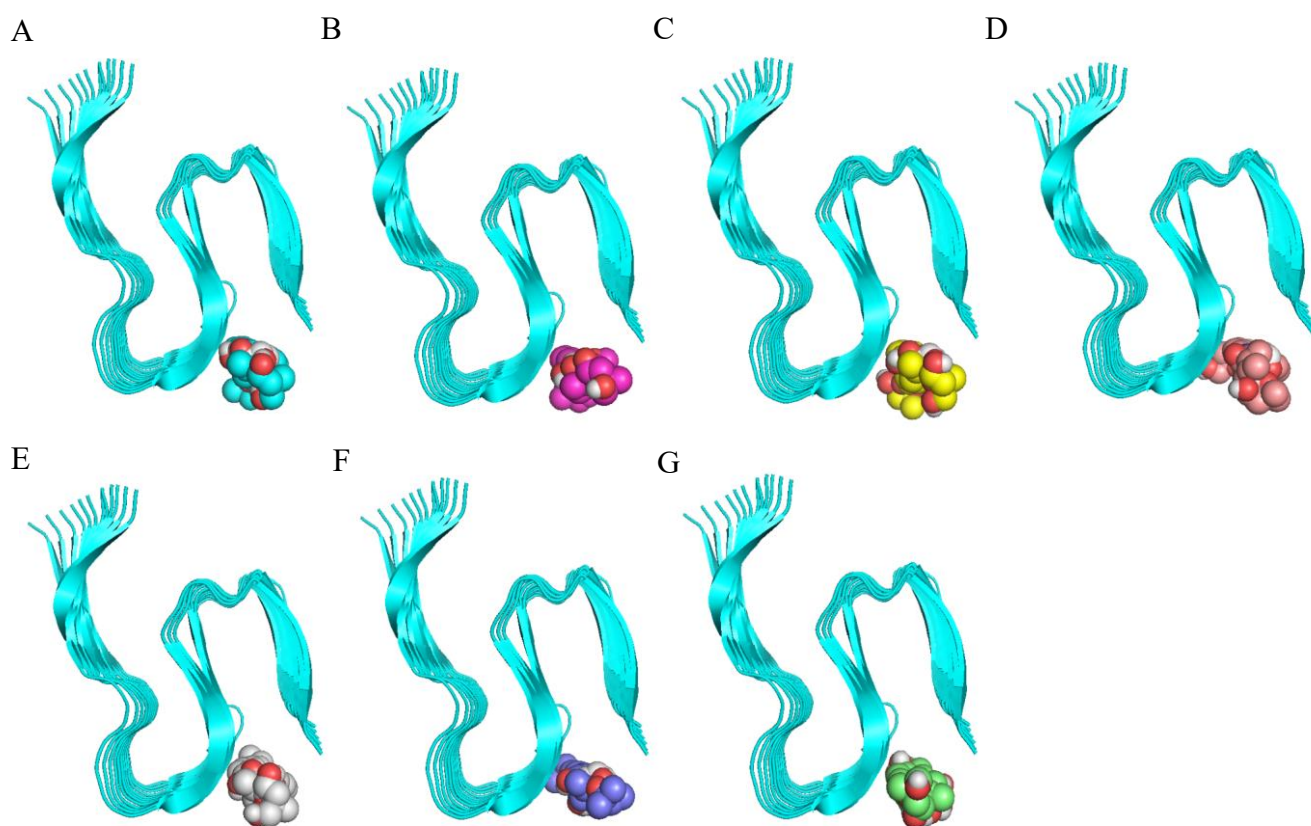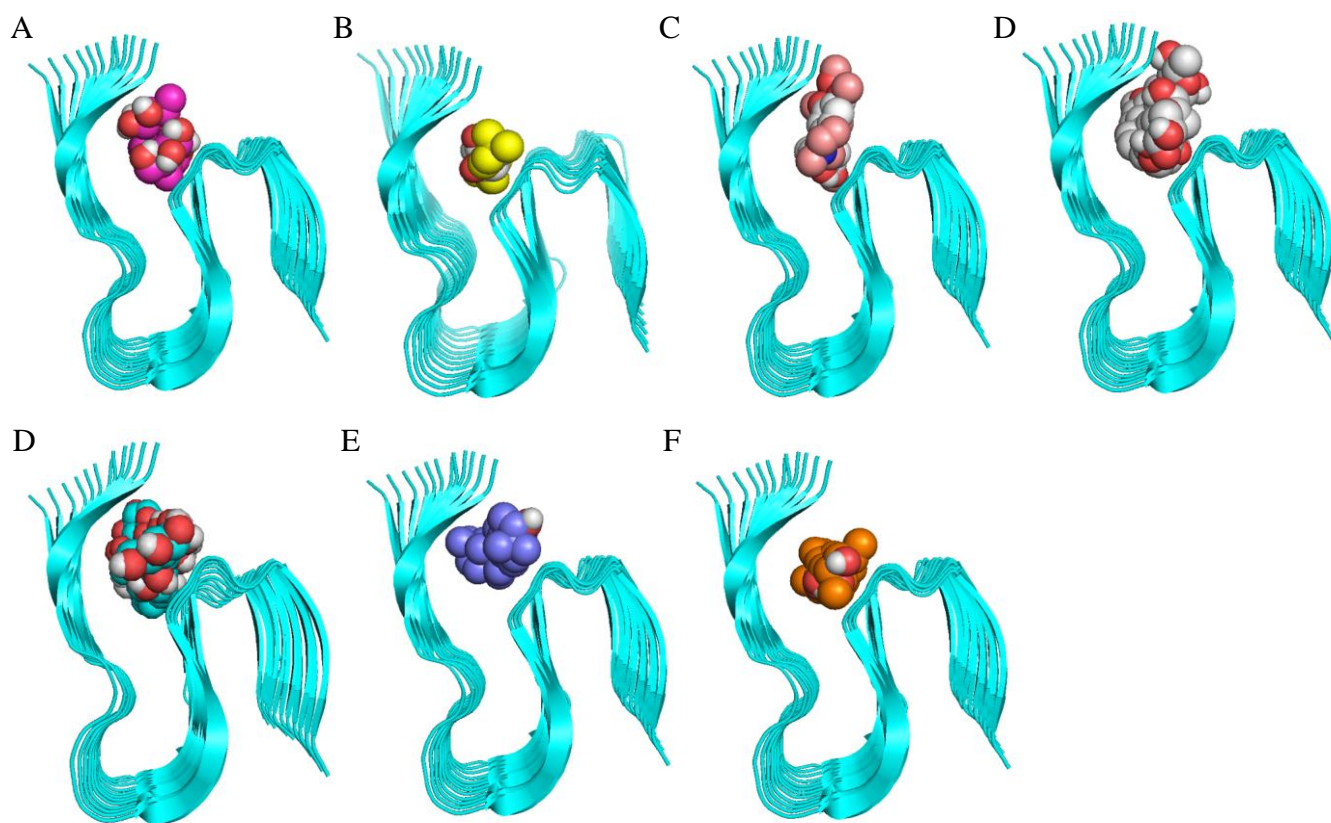
Figure 12: Visual representations of the results of the SANCDB compounds and the receptor 2MXU after Docking 3 created using PyMol. A) SANC00220. B) SANC00342. C) SANC00384. D) SANC00686. E) SANC00693. F) SANC00700. G) SANC00712.

Table 5: Table summarising the docking score and interactions between the receptor and the top hits from Docking 3 obtained from Discovery Studio. *Vdw – Van der Waals, HB – hydrogen bonds, () – indicate chain where interaction is present.

Compound	Vdw interactions	Alkyl interactions	Conventional HB	Carbon HB
SANC00220	Val12(E,C); Leu17(E,D,H); Ile32(C,F,H); Gly33(C,F,G)	Val12(D); His14(F); Leu17(F,G); Ile32(G,E,D); Leu34(D)		Gly33(D,H)
SANC00342	Val12(C); Leu17(C); Ile32(C); Gly33(B,D,E)	Val12(A,B); Leu17(E,F); Ile32(D,E); Leu34(A)		His14(B,C); Leu34(B,C)
SANC00384	Val12(B,E); His14(B); Leu17(E,F); Ile32(C,D,E); Gly33(C,D,E,F)	Val12(C,D); His14(C); Leu34(C,D,E)		Ile32(F)
SANC00686	Glu11(C); Val12(A,B); His14(D,E,F); Leu17(C,D,E,G); Ile32(B,C,D,G); Gly33(C,D,E); Leu34(B,C,D,F)	Val12(C,D,E,F); Leu17(F); Ile32(E,F)		Gly33(F)
SANC00693	Val12(A,B,F,G,H); His14(A,C,F); Leu17(A,C,H,I); Ile32(A,B,C,D); Gly33(D,E,F,H); Leu34(A,C,D,E,F)	His14(H); Leu17(E,F,G); Ile32(E,F,G)	His14(B); Gly33(B,C)	Gly33(A)
SANC00700	His14(D); Leu17(E); Gly33(C,D)	Val12(B,C,D,E); His14(B,C,E); Leu34(B,C,D)		
SANC00715	Val12(B,D); His14(B); Leu17(C); Ile32(B,F); Gly33(B,E,F); Leu34(E)	His14(D); Leu17(D, E,F); Ile32(D,E); Leu34(B,C)		His14(C); Ile32(C); Gly33(C)

Docking 3 was performed to target the more variable chain of the aggregate. Targeting of this chain could prevent further monomers from binding to the structure. Dock 3 (blind) allowed the natural compounds to bind to residues the 20Å grid box did not encompass in the initial docking. Therefore, interactions with residues such as Glu11, Val12 and His13 appeared in the Docking 3 (blind docking).

Dock 3 (blind) contained the lowest set of docking scores. However, the reason for this is not evident from the interactions (Table 5). There are no conventional hydrogen bonds present in any of the dockings in this set and there is a larger number of Van der Waals interactions (probably due to the larger sizes of these compounds). The SANCDB compounds interacted at similar sites in Dock 1 (centre) and Dock 3 (blind), interacting only with the residues on the β1 and β2 (Val12–Phe20 and Val36–Ile41) sheets, and the loop region connecting them (Ala21–Ser26). SANC00700 was a common compound in Dock 1 (centre) and 3 and interacted with the same residues in both apart from the interactions with Val12 and His14 due to the smaller box size in Dock 1 (centre). The appearance of the interactions of Val12 and His14 to SANC00700 decreased the binding energy from -9.2 in Dock 1 (centre) to -9.8 in Docking 3.

Figure 13: Image obtained as an example of the outputs obtained from discovery studio of the interactions between the natural compound SANC00220 and 2MXU.

Figure 14: Images of the three docking sites used in the molecular docking study visualised on PyMol. A) Dock 1 (centre). B) Dock 2 (targeted to L chain). C) Dock 3 (blind). *orange – Glu11, salmon – Val12, dark green – His13, red – His14, green - Leu17, sky blue – Ser26, purple – Asn27, grey – Lys28, blue – Ile32, yellow - Gly33, magenta - Leu34.

Table 6: Table summarising the main compounds obtained from the three dockings done against the Aβ aggregate 2MXU, structures retrieved from the SANCDB

Compound and Dock number	Name	Structure
SANC00175 – Dock 1 (center)	Lupeol	
SANC00178 – Dock 2 (targeted to L chain)	Cephalostatin 1	

SANC00220 - Dock 3 (blind)	Urginin	
SANC00290 – Dock 1 (center)	Clionamine D	
SANC00342 - Dock 3 (blind)	Kraussianone 1	
SANC00347 – Dock 1 (center)	Kraussianone 4	
SANC00348 – Dock 1 (center)	Kraussianone 5	
SANC00384 - Dock 3 (blind)	12-O-Methylcurine	
SANC00447 – Dock 2 (targeted to L chain)	Cephalostatin 14	

SANC00478 – Dock 2 (targeted to L chain)	Cephalostatin 2	
SANC00480 – Dock 2 (targeted to L chain)	Cephalostatin 4	
SANC00481 – Dock 2 (targeted to L chain)	Cephalostatin 7	
SANC00482 – Dock 2 (targeted to L chain)	Cephalostatin 8	
SANC00486 – Dock 2 (targeted to L chain)	Cephalostatin 12	
SANC00518 – Dock 1 (center)	20(29)-Lupene-3β-isoferulate	
SANC00553 – Dock 1 (center)	Octandronic acid	
SANC00686 - Dock 3 (blind)	Ornithosaponin C	

SANC00693 - Dock 3 (blind)	3β-[(O-β-D-Glucopyranosyl-(1->4)-O-β-D-Glucopyranosyl-(1->6)- β-D-Glucopyranosyl)Oxy)]-17α-Hydroxy-16β-[(O-(2-O-3,4-Dimethoxybenzoyl-β-D-Xylopyranosyl)-(1->3)-2-O-Acetyl-α-L-Arabinopyranosyl)Oxy]Cholest-5-En-22-One	
SANC00700 – Dock 1 (center) and Dock 3 (blind)	α-Glutinol	
SANC00715 - Dock 3 (blind)	Acovenoside A	
Wgx-50 – All dockings	N-[2-(3,4-dimethoxyphenyl)ethyl]-3-phenyl-acrylamide	

The compounds (Table 6) were all derived from various plant and animal species found in South Africa. Many of these structures are linear polycyclic systems. Some of these compounds have already been used as therapeutic agents in other diseases. SANC00178, SANC00447, SANC00478, SANC00480, SANC00481, SANC00482 and SANC00486 were obtained from *Cephalodiscus gilchristi* and all have been shown to have anti-cancer abilities. *Cephalodiscus gilchristi* (Figure 15) is a genus of Hemichordates and are a type of marine worms. These worms are built up of branched tubes attached to a disk-shaped head (Ridewood, 1908). Cephalostatin itself has proven to have potent growth inhibitory effects in the NCI 60 cancer cell line screen (Pettit et al., 1988).

Figure 15: Image of *Cephalodiscus gilchristi* discovered by Ridewood in 1908.

Currently the design of drug candidate targets that can cause the destabilization of the Aβ fibril is in its early stages. However, comparisons between structures of effective compounds can provide information for structure-based drug design. It has been noticed that structures with hydrophobic aromatic groups are closely associated to many Aβ-binding compounds with the hydrophobic interaction being the reason for the binding of these compounds to the fibril. The size of the compounds is also an important feature with the smaller molecules being able to penetrate the interior of the cross-β subunit and deform the protofibril. It is hypothesized that larger compounds are only able to bind to the surface/edge of the fibril, causing minimal to no damage. Polar compounds assist in the destabilization of the salt bridge at Asp23-Lys28, however this could also be positive and may help stabilize the fibril (Fan, Xu and Wei, 2017). Table 7 summarizes the results of the calculation of physico-chemical properties of the HTVS best performing ligands.

Table 7: Summary of the druglikeness results for each SANCDB compound and Wgx-50 from the three online sites; SwissADME (S), Molinspiration (M) and Chemicalize (C). * logP- partition coefficient. TPSA- the polar surface area prediction.

Compounds	Mass (g/mol)			Hydrogen Donors			Hydrogen Acceptors			LogP			Rule violations		
	S	M	C	S	M	C	S	M	C	S	M	C	S	M	C
Rule		<500			<5			<10			≤5			≤1	
SANC00175	426,72	426,73	426,73	1,00	1,00	1,00	1,00	1,00	1,00	4,76	8,29	7,45	0	1	1
SANC00178	911,17	911,19	911,19	5,00	5,00	5,00	12,00	12,00	12,00	5,71	6,88	3,74	4	4	3
SANC00220	838,93	838,94	838,94	9,00	9,00	9,00	17,00	17,00	16,00	4,12	-0,12	-0,19	3	3	3
SANC00290	401,54	401,55	401,55	1,00	2,00	1,00	5,00	5,00	3,00	2,81	0,82	3,01	0	0	0
SANC00342	418,44	418,44	418,45	2,00	2,00	2,00	6,00	6,00	6,00	3,99	5,34	5,18	0	1	1
SANC00347	434,44	434,44	434,44	2,00	2,00	2,00	7,00	7,00	7,00	3,83	4,26	4,54	0	0	0
SANC00348	436,45	436,46	436,46	3,00	3,00	3,00	7,00	7,00	7,00	3,54	4,83	5,07	0	0	1
SANC00384	608,72	608,74	608,74	1,00	1,00	1,00	8,00	8,00	6,00	4,98	6,27	6,06	1	2	2
SANC00447	941,20	941,22	941,22	5,00	5,00	5,00	13,00	13,00	13,00	6,54	6,64	3,67	4	4	3
SANC00478	927,19	927,19	927,19	6,00	6,00	6,00	13,00	13,00	13,00	5,93	5,93	2,82	4	4	3
SANC00480	943,17	943,19	943,19	6,00	6,00	6,00	14,00	14,00	14,00	5,49	5,36	2,19	4	4	3
SANC00481	929,21	929,21	929,21	7,00	7,00	7,00	13,00	13,00	13,00	6,28	6,28	2,52	4	4	3
SANC00482	927,22	927,23	927,23	6,00	6,00	6,00	12,00	12,00	12,00	6,33	7,54	3,31	4	4	3
SANC00486	945,19	945,20	945,20	8,00	8,00	8,00	14,00	14,00	14,00	5,46	5,37	1,60	4	4	3
SANC00518	602,89	602,90	602,90	1,00	1,00	1,00	4,00	4,00	3,00	6,23	9,26	10,02	2	2	2
SANC00553	456,70	456,71	456,71	1,00	1,00	1,00	3,00	3,00	3,00	3,63	6,56	7,31	0	1	1
SANC00686	1033,11	1 033,12	1033,12	12,00	12,00	12,00	23,00	23,00	23,00	4,87	-1,87	-2,39	3	3	3
SANC00693	1 389,48	1 389,49	1389,50	14,00	14,00	14,00	31,00	31,00	29,00	3,54	-0,91	-0,86	3	3	3
SANC00700	426,72	426,73	426,73	1,00	1,00	1,00	1,00	1,00	1,00	4,80	8,02	7,40	0	1	1
SANC00715	550,68	550,69	550,69	4.00	4,00	4,00	9,00	9,00	8,00	3,88	1,26	1,68	1	1	1
Wgx-50	311,37	311,38	311,39	1,00	1,00	1,00	3,00	4,00	3,00	3,24	3,22	3,25	0	0	0

From Table 7 only a few of these best-performing compounds, across all dockings, abide by Lipinski's rule of 5. These compounds are SANC00175, SANC00290, SANC00347, SANC00348, SANC00553, SANC00700, SANC00715 and Wgx-50. Other compounds do abide by the rules as determined by a single website but fail according to other websites. The websites generate generally the same values for each compound except for the logP score. These scores differ greatly for each website.

Drug-likeness is a good indicator of how effective a drug may be; however, it has its limitations. Drug-likeness can be estimated for any molecular and does not evaluate the specific effect that the drug achieves. Several poisons have a good drug-likeness score and many best-selling drugs have characteristics that cause them to have a low drug-likeness score. This technique is also not relevant for proteins since they are digested if eaten and therefore need to be injected (Vistoli *et al.*, 2008).

As a result of these limitations variations of Lipinski's rule of 5 were created to improve the predictions. One of them is the Ghose filter. This filter further states that the logP value should be in the range of −0.4 to +5.6, the molar refractivity should be from 40-130, the molecular weight from 180-480 g/mol and the number of atoms from 20-70 (Ghose *et al*, 1999). Veber's rule disagrees with the 500 g/mol molecular weight cutoff and states that the polar surface area and number of rotatable bonds better determines which compounds will be able to be administered orally. The criteria for Veber's rule are 10 or fewer rotatable bonds and a polar surface area no larger than 140 $Å^2$ to have good oral bioavailability (Veber *et al.*, 2002).

The topological polar surface area (TPSA) of a compound is the surface sum over all polar atoms, also including their attached hydrogen atoms. It is used in medicinal chemistry for the optimization of a drug's ability to permeate cells. If the TPSA is above 140 $Å^2$ the compound will tend to be poor at permeating the cell membranes. A TPSA of below 90 $Å^2$ is required to penetrate the blood-brain-barrier (Prasanna and Doerksen, 2009).

Table 8 shows the calculated TPSA properties of the identified binding molecules.

Table 8: Summary of the TPSA results for each SANCDB compound and Wgx-50 from the three online sites; SwissADME (S), Molinspiration (M) and Chemicalize (C).

Compounds	TPSA (Å²)		
	S	M	C
SANC00175	20,23	20,23	20,23
SANC00178	180,92	180,93	180,92
SANC00220	267,66	267,67	263,75
SANC00290	78,62	78,63	78,62
SANC00342	89,13	89,14	85,22
SANC00347	94,45	94,46	94,45
SANC00348	105,45	105,46	105,45
SANC00384	72,86	72,87	72,86
SANC00447	190,15	190,16	190,15
SANC00478	201.16	201,16	201,15
SANC00480	213,68	213,69	213,68
SANC00481	204.32	204,32	204,31
SANC00482	184,08	184,09	184,08
SANC00486	224,54	224,54	224,54
SANC00518	55,76	55,77	55,76
SANC00553	54,37	54,37	54,37
SANC00686	352,13	352,15	352,13
SANC00693	463,65	463,68	463,65
SANC00700	20,23	20,23	20,23
SANC00715	134,91	134,92	134,91
Wgx-50	47,56	47,57	47,56

The results in table 8 show that only few of the compounds, these being SANC00175, SANC00290, SANC00342, SANC00347, SANC00348, SANC00384, SANC00518, SANC00553, SANC00700, SANC00715 and Wgx-50 would be able to permeate the cell membranes. Compounds SANC00175, SANC00290, SANC00342, SANC00384, SANC00518, SANC00553, SANC00700 and Wgx-50 would be able to pass through the blood-brain barrier.

By comparison the docking of the known inhibitor Wgx-50, is presented in the three docking scenarios to 2MXU in Figure 15. Table 9 lists the observed interactions.
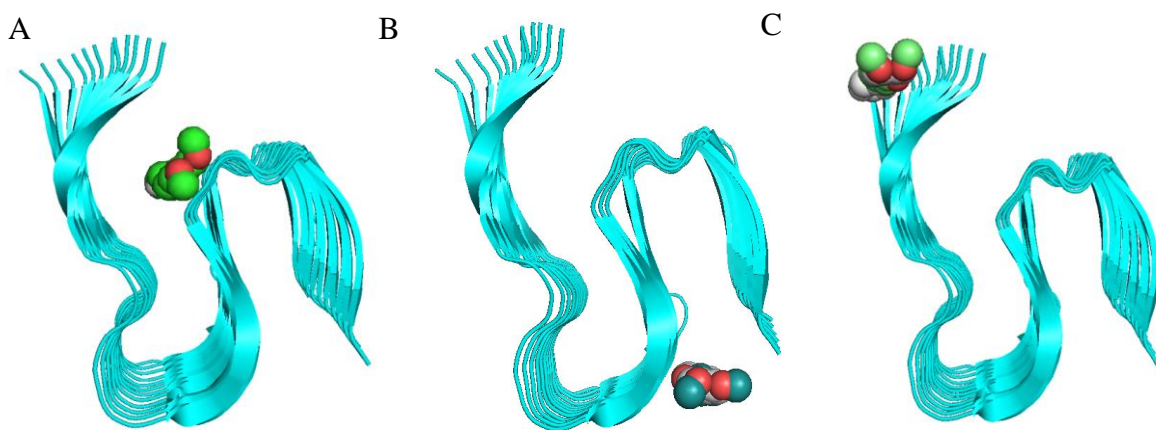
Figure 16: Images of the three different dockings of the Wgx-50 drug candidate against the 2MXU receptor generated on PyMol. A) Dock 1 (centre). B) Dock 2 (targeted to L chain). C) Dock 3 (blind).

Table 9: Table summarising the docking score and interactions between the receptor and the Wgx-50. *Vdw – Van der Waals, HB – hydrogen bonds, () – indicate chain where interaction is present.

Wgx-50	Docking score	Vdw interactions	Alkyl interactions	Conventional HB	Carbon HB
Dock 1	-6.4	Leu17(F,G,H,I); Ile32(E.G); Gly33(F,G,H,I); Leu34(E)	Ile32(H,I)	Leu34(F)	Ile32(F); Gly33(E)
Dock 2	-6.4	Ser26(H,I,J,K); Asn27(H,I,J,K); Lys28(G,I)	Lys28(H,J,K)		Ser26(G)
Dock 3	-6.9	Glu11(E, F, G, H, I); Val12(F, G, I); His13(E, F, G, I)	His13(H)		

Figure 16 shows the various types of binding between Wgx-50 and 2MXU for each of the dockings. All the dockings have different results and no docking shows Wgx-50 reacting with similar residues. Dock 1 (centre)'s results show the drug candidate interacting with central residues. Dock 2 (targeted to L chain)'s interactions are like the other Dock 2 (targeted to L chain) results as in they react with the same residues (Ser26, Asn27 and Lys28). Dock 3 (blind) shows Wgx-50 reacting with the first few residues of the chains. Dock 3 (blind) has the lowest docking score compared to the other dockings however, these docking scores if compared to the SANCDB dockings have higher docking

scores. This suggests the binding of Wgx-50 and 2MXU is not as strong as other interactions shown previously.

As it can be seen in Table 9 Wgx-50 does not form any hydrogen bonds with the receptor. This does not indicate strong binding and neither does the docking scores for the dockings which are lower than the SANCDB compounds docking scores. However, this does not indicate that there will be no disruption when analysed using molecular dynamics.

Chapter 3: Molecular Dynamics

3.1 Introduction:

AD is directly correlated to the formation of amyloid fibrils. These fibrils are also associated with other diseases such as the neurodegenerative Parkinson's and Huntington's, prion diseases and type 2 diabetes. These fibrils, formed by an aggregation of β-amyloid peptides, is important to understand in order to combat AD and the other mentioned diseases. A major exploratory path for treatment is to combat the aggregation of the peptide or to destabilize the fibril itself. Information about the molecular structures were previously limited to their basic structure of β-sheets arranged in a cross-β orientation indicated by X-ray fiber diffraction data. Recently, due to the use of solid-state NMR, more information is available on the fibril structure such as measured interatomic distances that constrain the supramolecular organization, dihedral angles and estimated distances that constrain the molecular conformation (Buchete, Tycko and Hummer, 2005). Studies such as molecular dynamics can delve into the native movements of the aggregate, and in simulations combining both aggregate and ligands, determine the effect of these ligands on the aggregate.

Only few decades ago proteins were thought to be rigid body structures with very limited conformational flexibility. However, it is now clear that proteins are highly dynamic structures, the internal organization of which is the key to their 3D spatial arrangement and hence biological function (Vlachakis *et al.*, 2014). Historically the study of protein dynamics in the wet laboratory has been a very complicated, expensive, and time-consuming process. Therefore, much effort has been placed in the use of computers and the *in silico* study of protein structure using molecular dynamics (MD) (Vlachakis *et al.*, 2014).

MD is a powerful computing technique that provides the user with information of macromolecules such as the patterns of motion, structural strength, and properties of protein behavior at biological temperatures. In addition to this, MD may reveal drug–receptor interactions, the solvation of molecules and the conformational changes that a protein or molecule may undergo under various conditions (Vlachakis *et al.*, 2014). Biological macromolecules in MD simulations are subjected to forces, and their motions at various temperatures and pressures are computed at timesteps in computer simulations, capturing in full atomic detail key biochemical processes such as protein folding, drug binding, membrane transport, and the conformational changes critical to protein function (Berhanu and Masunov, 2014). MD simulations treat the molecule as a collection of interacting classical particles and integrate the Newton's laws of motion to simulate the motion of macromolecules and ligands (Berhanu and Masunov, 2014). The energy terms used and calculated

in a MD procedure allows the atoms to move and even collide in a physically realistic manner into neighbouring atoms (Pitman and Menz, 2006)

The result of the simulation is a trajectory of the system over a certain simulated period, usually tens to hundreds of nanoseconds. Various structural and dynamic properties of the system can then be calculated from the trajectory to gain a kinetic and thermodynamic understanding of the system (Berhanu and Masunov, 2014). Simulations are performed using empirically parameterized force fields that typically include explicit solvent. Examples of some of the most commonly used force fields are AMBER, CHARMM, GROMOS and OPLS (Berhanu and Masunov, 2014).

There have been several theoretical attempts to study the interactions between current inhibitors and amyloid aggregates at the atomic level. However, MD of the top hits from the molecular docking of the SANCDB compounds could provide evidence of new potential treatments. This is where the compound interacts with the fibril in a way that it interrupts bonds between the monomers and thus destabilizes the aggregate.

3.2 Methods:

3.2.1 Molecular dynamics

All simulations were run using the GROMACS software package, version 4.6, using the GROMOS96 43a1 force field. Initially the two protein models 2MXU (apo structure 1) and 2BEG (apo structure 2) were analysed using MD without any ligands present. The protein structures were solvated in a cubic box 1 nm away from the model border. The solvent water molecules were explicitly represented by the TIP3P model. Positive sodium ions (Na+) and negative chlorine ions (Cl-) replaced water molecules to neutralize the system (with an excess of +12 with 2MXU and an excess of +5 with 2BEG) to neutralize the system. The systems were energy-minimized using the steepest-descent method. The NVT ensemble (constant number of particles, volume and temperature) was applied to the system for 100 picoseconds (ps) using the Berendsen weak coupling method to maintain the temperature at 300 K. An isothermal-isobaric (NPT) ensemble was performed at 100 ps to maintain temperature (300 K) and pressure (1 bar) using a Nose´-Hoover thermostat and Parrinello-Rahman barostat.

The protein models were then simulated using MD at 200 nanoseconds (ns). The first models of all the top hits of the SANCDB docking as well as Wgx-50 were submitted to PRODRG in order to obtain the GROMACS topology and the polar/aromatic hydrogens GROMACS files needed for MD

simulations. The 2MXU receptor and ligand Gromacs files were combined and were solvated in a cubic box 1 nm away from the model border. The MD simulations were performed identically to the apo protein simulations after the combination of the topology files. A MD simulation was also performed with 2BEG and Wgx-50 as described before.

3.2.2 RMSD, RMSF and radius of gyrations

Three tools used for the analysis of the MD results are available from GROMACS; root mean squared deviation (RMSD), root mean squared fluctuation (RMSF) and radius of gyration.

RMSD compares two structures by computing the root mean square deviation with each structure in the trajectory compared with a reference in the structure file. The starting structure of the protein is used as the reference ($\mathbf{r}^{ref}$) and is compared to the structures that evolve over time in the simulation. The RMSD of certain atoms in a molecule is calculated where $\sum_{i=1}^{N} m_i$ and $r_i(t)$ is the position of atom i at time t after least square fitting the structure to the r^{ref} (Devadoss and Raj, 2014; Mutt and Sowdhamini, 2016).

$$\text{RMSD}(t) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i |\mathbf{r}_i(t) - \mathbf{r}_i^{ref}|^2 \right]^{1/2}$$

The RMSF is a measure of the deviation between the position of particle i and some reference position: where T is the time over which one wants to average and $\mathbf{r}_i^{ref}$ the reference position of particle i. It is the most common measure of the spatial extent of random motion. Typically this reference position will be the time-averaged position of the same particle i (Devadoss and Raj, 2014; Mutt and Sowdhamini, 2016).

$$\text{RMSF}_i = \left[\frac{1}{T} \sum_{t_j=1}^{T} |\mathbf{r}_i(t_j) - r_i^{ref}|^2 \right]^{1/2}$$

The difference between RMSD and RMSF is that the latter is averaged over time, giving a value for each particle i. For the RMSD the average is taken over the particles, giving time specific values.

The radius of gyration is the measure of compactness of a structure where m_i is the mass of atom i and $\mathbf{r}_i$ the position of atom i with respect to the centre of mass of the molecule (Devadoss and Raj, 2014; Mutt and Sowdhamini, 2016).

$$R_g = \left(\frac{\sum_i |\mathbf{r}_i|^2 m_i}{\sum_i m_i} \right)^2$$

RMSD, RMSF and Radius of gyration plots were generated using GROMACS for each of the results. RMSD and RMSF were generated with respect to the backbone atoms and radius of gyration was generated with respect to the entire system.

3.2.3 MMPBSA

MMPBSA analysis was performed between 100 ns and 110 ns on all the simulated complexes using GROMACS. This region was chosen because it was the most stable in all the RMSD results.

Molecular Mechanics/Poissan-Boltzman Surface Area (MMPBSA). The MMPBSA method calculates a binding free energy by the free energies of solvation for the complex ($\Delta G_{complex}$), protein ($\Delta G_{protein}$) and ligand (ΔG_{ligand}):

$$\Delta G_{bind} = \Delta G_{complex} - \Delta G_{protein} - \Delta G_{ligand}$$

Each of these values are calculated by determining the enthalpic energy of the solute using molecular mechanics (E_{MM}), the polar solvation free energy (ΔG_{solv}), the nonpolar solvation free energy (ΔG_{pp}) and the entropic contribution (ΔS):

$$\Delta G = (E_{MM}) + (\Delta G_{solv}) + (\Delta G_{pp}) - T(\Delta S)$$

The solute is computed as the average enthalpic term over the molecular mechanics force field. The solute vibrational entropy is estimated using either normal mode analysis or quasiharmonic approximation. The polar solvation free energy is solved using the Poisson-Boltzman (PB) equation. The nonpolar term is solved using the Generalized Born (GB) method and is assumed to be proportional to the SASA:

$$\Delta G_{np} = \gamma SASA + \beta$$

Where γ is the surface tension, set to 0.0072 kcal/Å^2. β is an offset value used to correct for the nonpolar contribution to the solvation free energy term and is dependent on the GB model used (Hou *et al.*, 2011).

The MMPBSA results show the van der Waal energy contribution from MM, Electrostatic energy calculated by the MM force field, SASA energy and the final estimated binding free energy calculated from the terms above for the 2MXU complex for each docking.

3.3 Results and Discussion:

Molecular dynamics is an important computing tool used to understand the native movement of a protein or a system. Using MD on the aggregate in complex with compounds was performed in order to deepen understanding of how it interacts with various compounds added to it. Good binding SANCDB compounds docked in 2MXU were taken to MD.

The MD simulations were conducted with the aggregate alone, with three sets of SANCDB docked compounds (according to dock1 – docking at the center of the aggregate, dock2 – docking focussed on chain L and dock3 – blind docking), and with docked wgx40. The *apo* 2MXU dynamics analysis is included with all sets of MD.

For Dock 1 (center) the MD simulations were performed on the complexes with SANC00175, SANC00290, SANC00347, SANC00348, SANC00518, SANC00553 and SANC00700. The MD analysis for Dock 2 (targeted to L chain) was conducted with compounds SANC00178, SANC00447, SANC00478, SANC00480, SANC00481, SANC00482 and SANC00486. The MD analysis for Dock 3 (blind) was conducted with compounds SANC00220, SANC00342, SANC00384, SANC00686, SANC00693, SANC00700 and SANC00715.

The RMSD of the backbone atoms of 2MXU alone in an MD simulation increased from 0.15 to 0.59 (top of Figures 17, 20 or 23). The RMSD of the backbone atoms of Dock 1 (centre) of 2MXU and SANCDB compounds complexes did not exceed 0.4 (Figure 17). These RMSD scores of the complexes are lower than the *apo* structure 1. The RMSD of the Dock 2 (targeted to L chain)'s complexes also did not exceed 0.4 except for SANC00480 (Figure 20) which was slightly above 0.4 but did still not reach the value of the *apo* structure's RMSD. The same was seen for the Dock 3 (blind)'s complexes however SANC00693 did match the *apo* structure the best (Figure 23). The RMSD results (Figure 17, 20 and 23) do not indicate a high level of structural change.

The RMSF's of the complexes show distinctive fluctuations every ~470 atoms associated with the termini of the respective chains in all systems whether for the *apo* system or for complexes. The fluctuation of the *apo* structure's first two chains is relatively low however the complexes show increased fluctuation in this area except for SANC00347, SANC00518 and SANC00700. Only SANC00175, SANC00347 and SANC00518 contain peaks that are higher than the highest peak of the *apo* structure 1 (Figure 18). However, for SANC00347, it is interesting to see that there is destabilization towards the center of each chain (i.e. at the surface of the complex furthest away from the termini). Compare this to the known inhibitor of aggregation (Wgx-50, Figure 26, first docked

RMSF), where the "doubling" of peaks on the RMSF plot is also evident. It is possible that the good binder SANC00347 is affecting the aggregate in a similar manner to that of Wgx-50.

Dock 2 (targeted to L chain)'s set of complexes show more instances of this general increase in the number of peaks of the RMSF (Figure 21) especially for SANC00178 and SANC00447, this indicating a high number of atom fluctuations, particularly at what would be presumed to be the strongest bound region of the aggregate, furthest from the termini. There is increased fluctuation just after atom 2000 (this is at the boundary between chain 7 and chain 8) in all these complexes, however the fluctuation is higher for SANC00482 (Figure 21). Of this set, SANC00178, SANC00447, SANC00480, SANC00481 hold promise in terms of following the same destabilizing patterns as Wgx-50. The destabilization by SANC00481 is different in terms of its concentrated effect on the first chains of the aggregate.

The RMSF results for Dock 3 (blind) complexes (Figure 24) showed the lowest fluctuation compared to the other sets of complexes with SANC00693 exhibiting the lowest atom fluctuation. SANC00700 showed high fluctuation in the first 1000 atoms compared to the apo structure. SANC00715 showed an unusually high initial peak with a height of 0.6nm showing a large distance of fluctuation. SANC00220 matched the apo results the best of the complexes (Figure 24). SANC00686 and SANC00715 were the most promising in terms of the "doubling" of peaks in the RMSF, minimizing the Wgx-50 effect.

The radius of gyration for the *apo* structure 1 decreases from 2.2 – 2.055 over time (top of Figure 19, 22 and 25). The only complexes that provide a higher radius of gyration are SANC00175, SANC00347, SANC00700 (Figure 19), SANC00447, SANC00486 (Figure 22), SANC00220 and SANC000693 (Figure 25). The second set of complexes showed little change in radius of gyration. Even though these compounds have a higher radius of gyration when in complex with 2MXU, suggesting a less tightly packed structure, the values are not high enough to suggest a large difference between the packing of the complex and the apo structure. As such, most evidence for destabilization of the aggregate comes from the RMSF plots. The timescale of 200ns for dynamics may not be sufficient for the radius of gyration to show significant deviation.

Figure 17: RMSD plots of the results of MD for Dock 1 (centre) performed using GROMACS.

Figure 18: RMSF plots of the results of MD for Dock 1 (center) performed using GROMACS

Figure 19: Radius of gyration plots of the results of MD for Dock 1 (center) performed using GROMACS

Figure 20: RMSD plots of the results of MD for the Dock 2 (targeted to L chain) performed using GROMACS.

Figure 21: RMSF plots of the results of MD for Dock 2 (targeted to L chain) performed using GROMACS.

Figure 22: Radius of gyration plots of the results of MD for Dock 2 (targeted to L chain) performed using GROMACS.
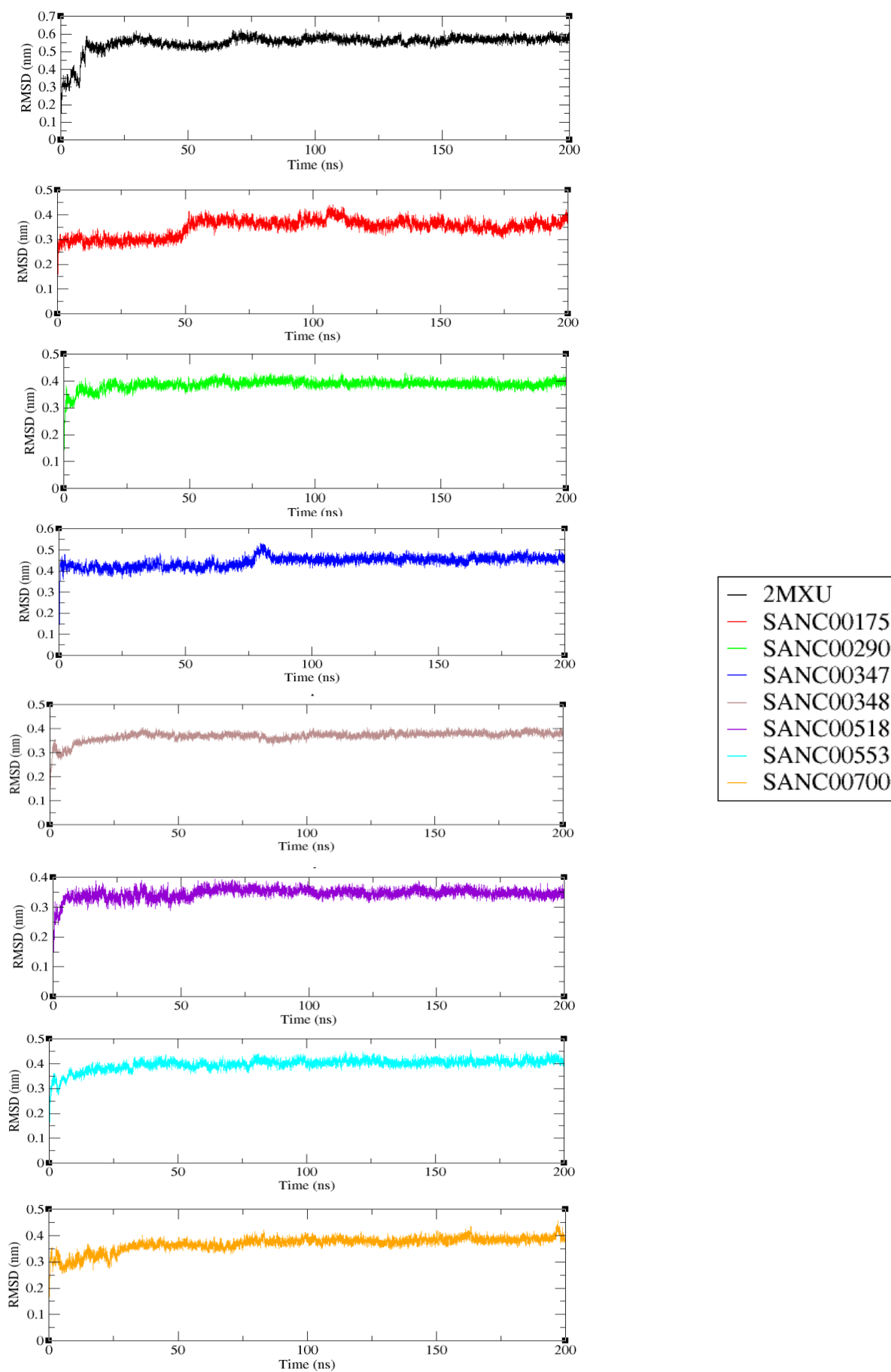
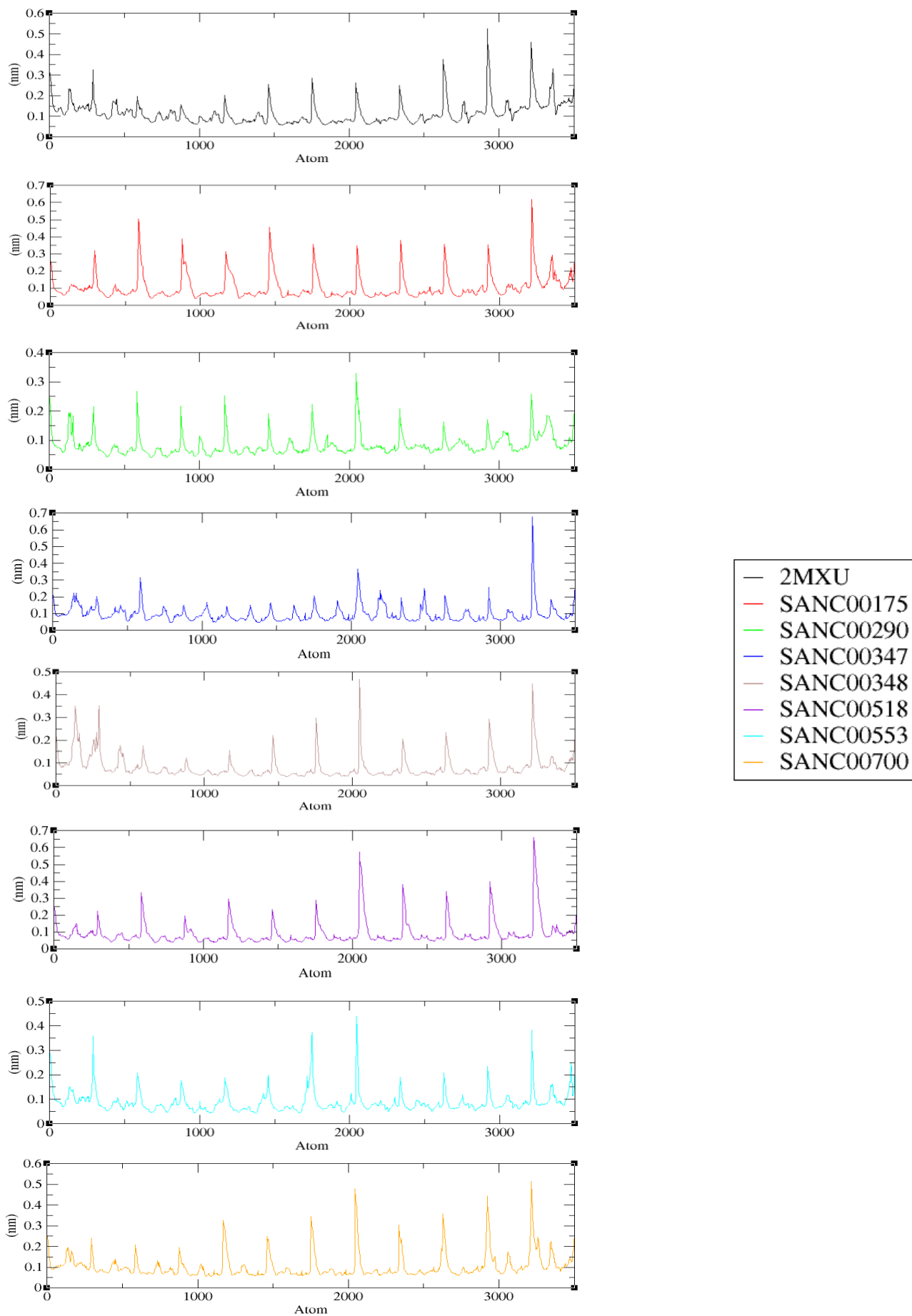**Figure 23:** RMSD plots of the results of MD for Dock 3 (blind) performed using GROMACS.

Figure 24: RMSF plots of the results of MD for Dock 3 (blind) performed using GROMACS.
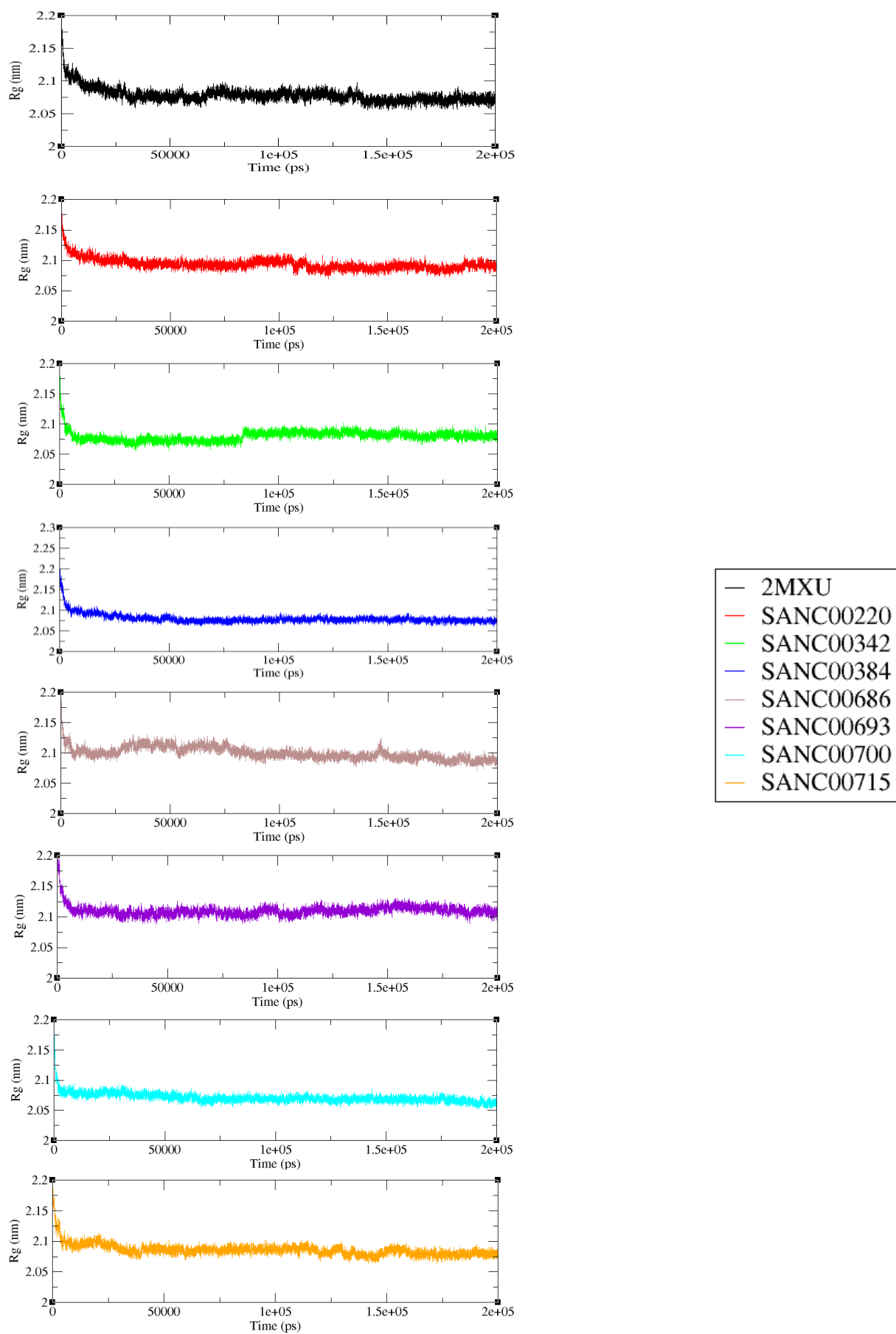
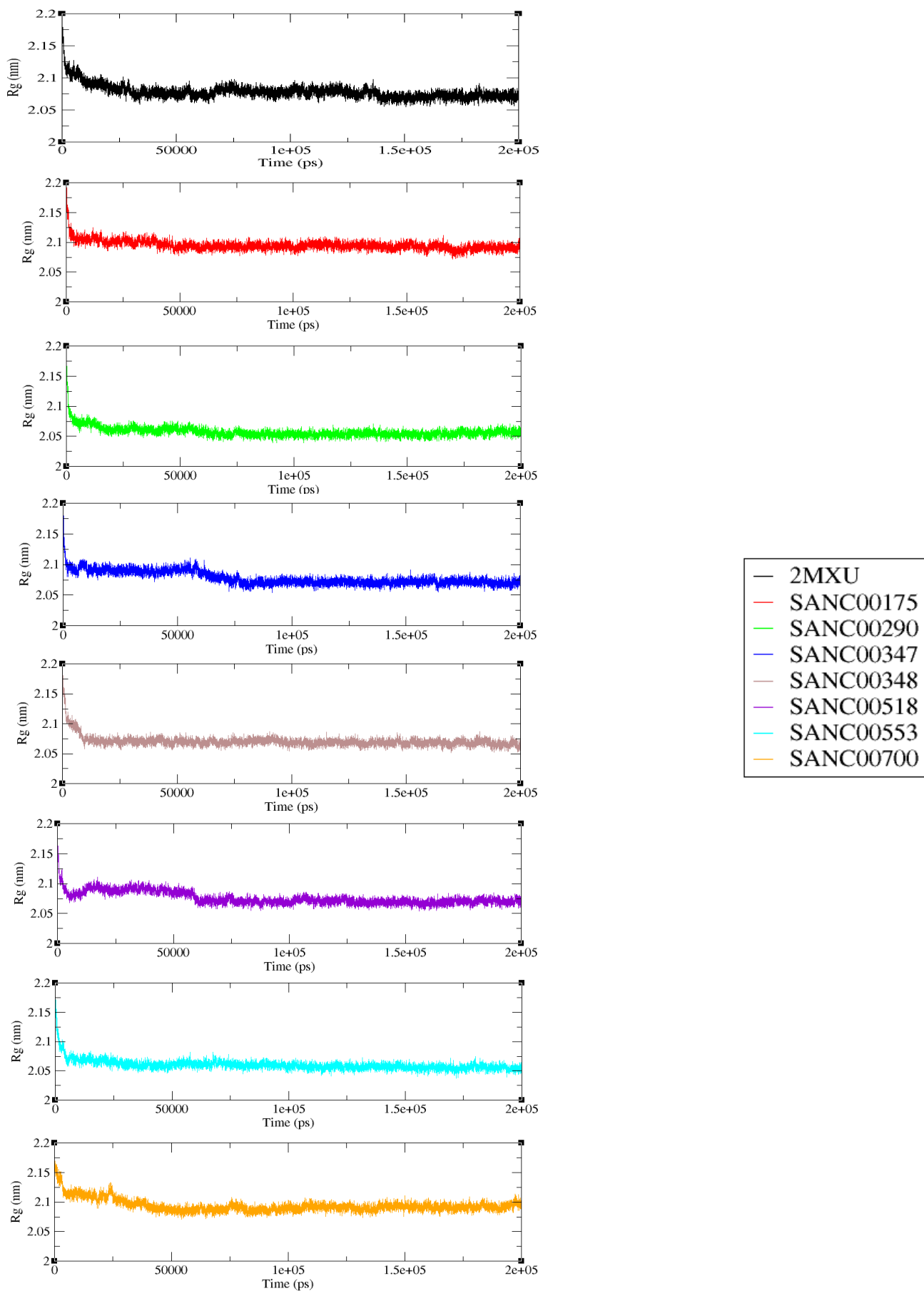**Figure 25:** Radius of gyration plots of the results of MD for Dock 3 (blind) performed using GROMACS.
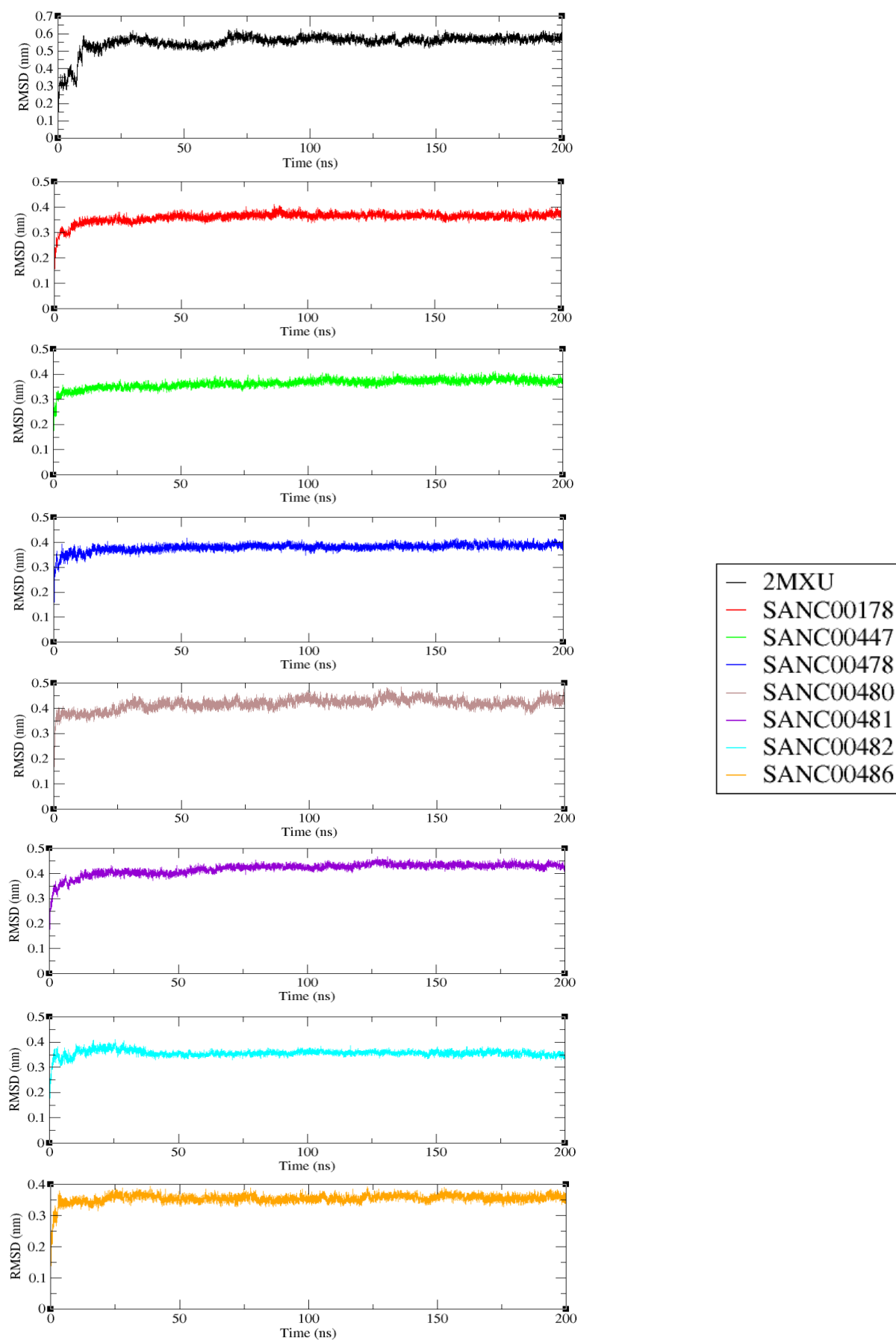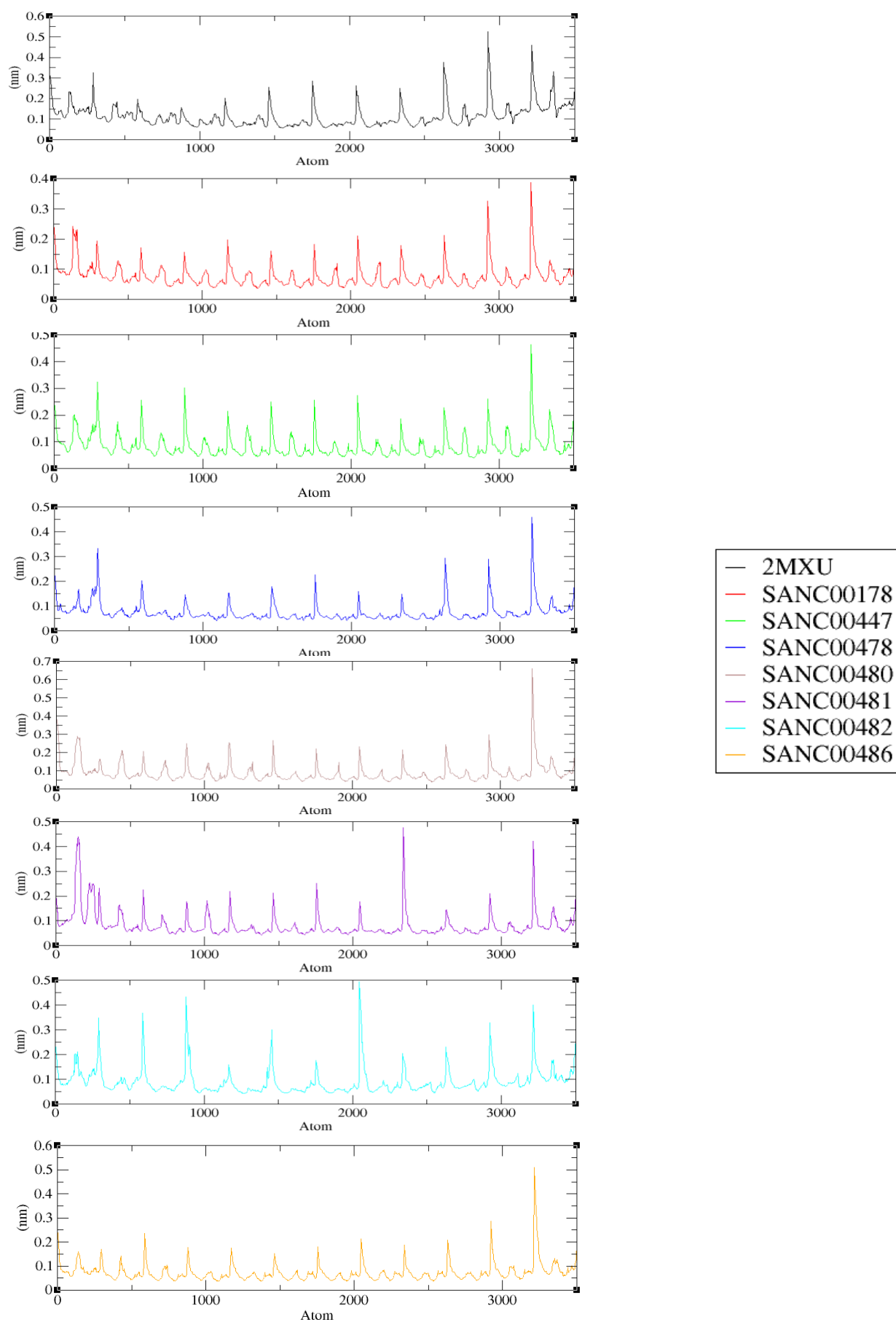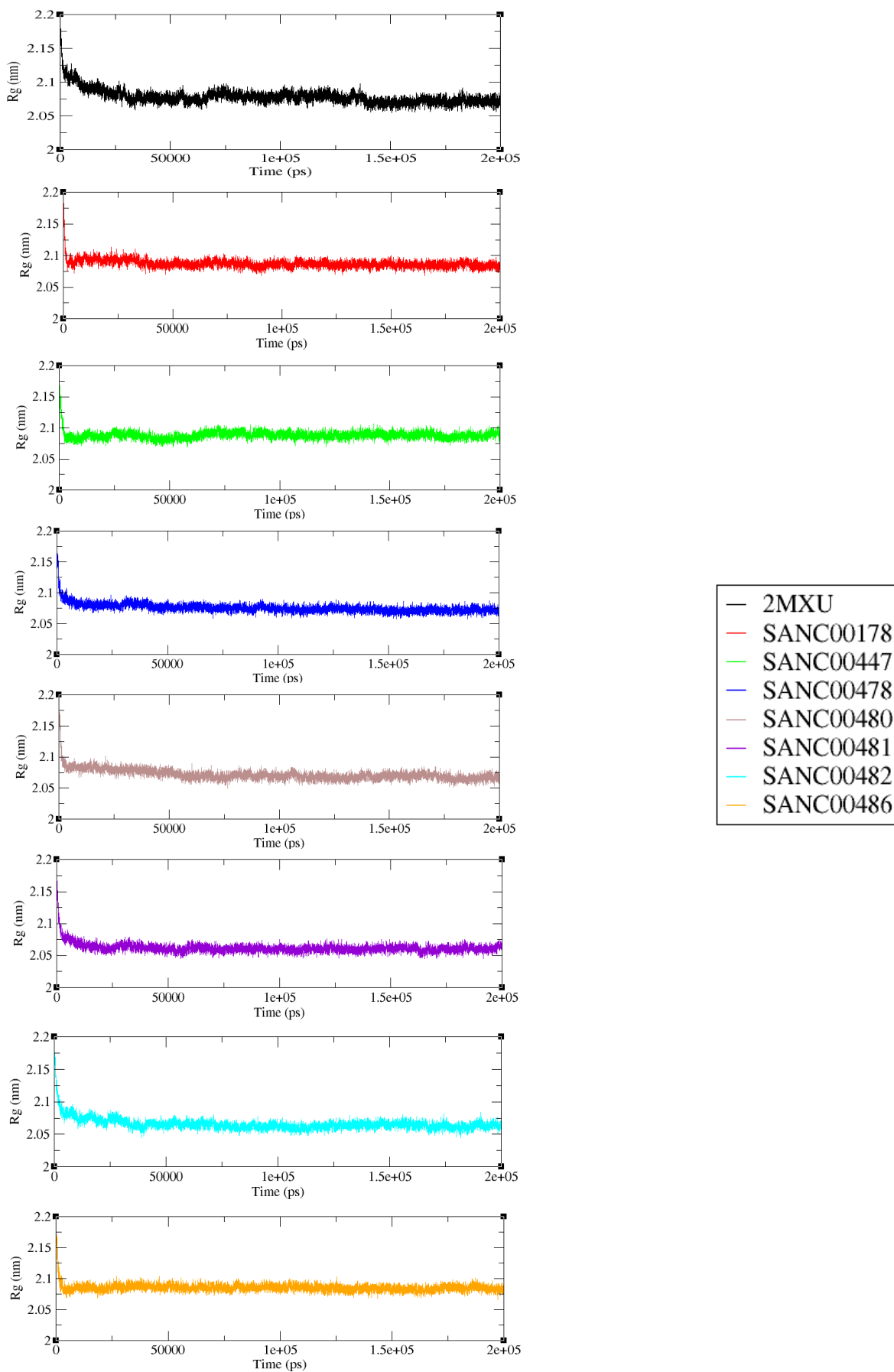
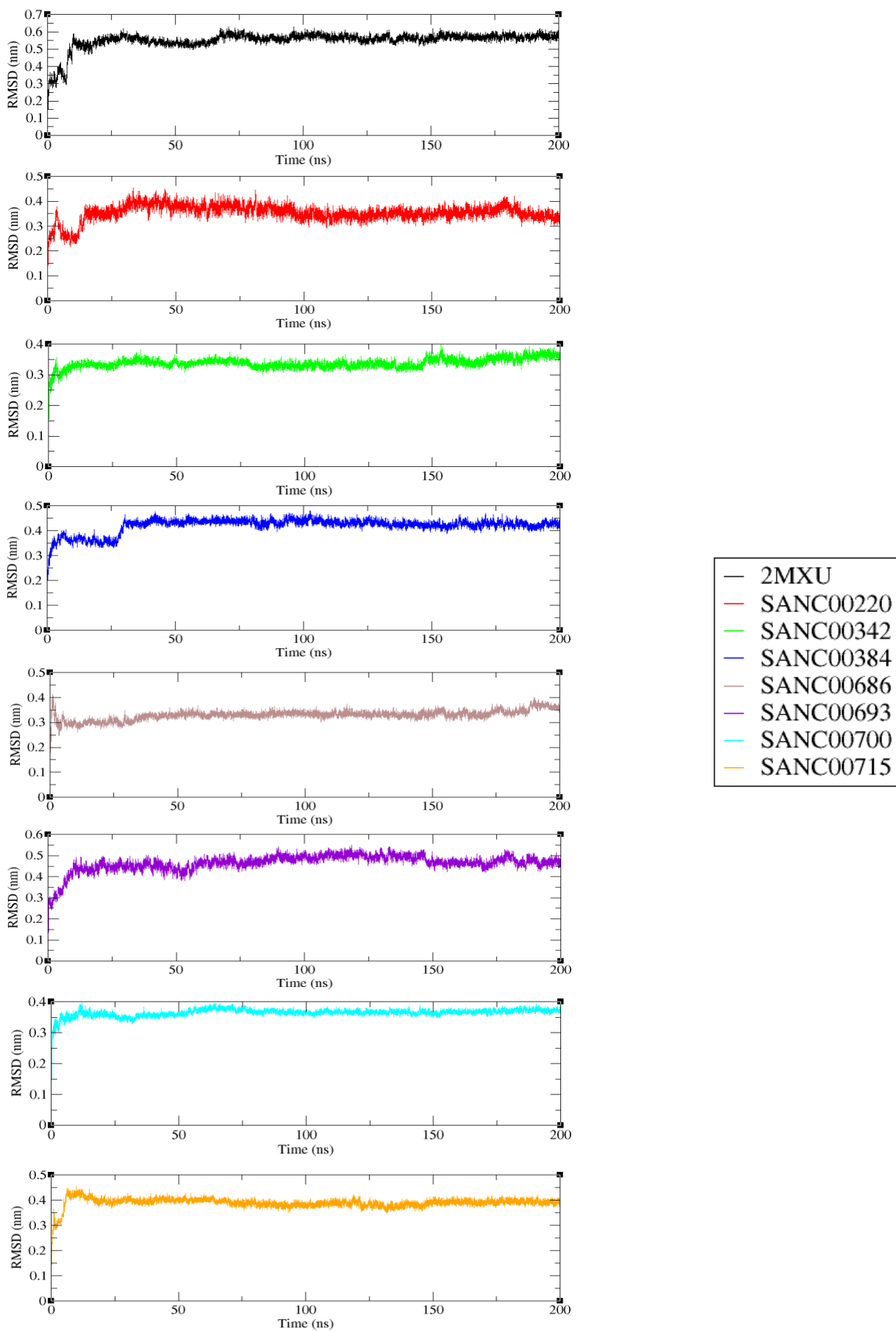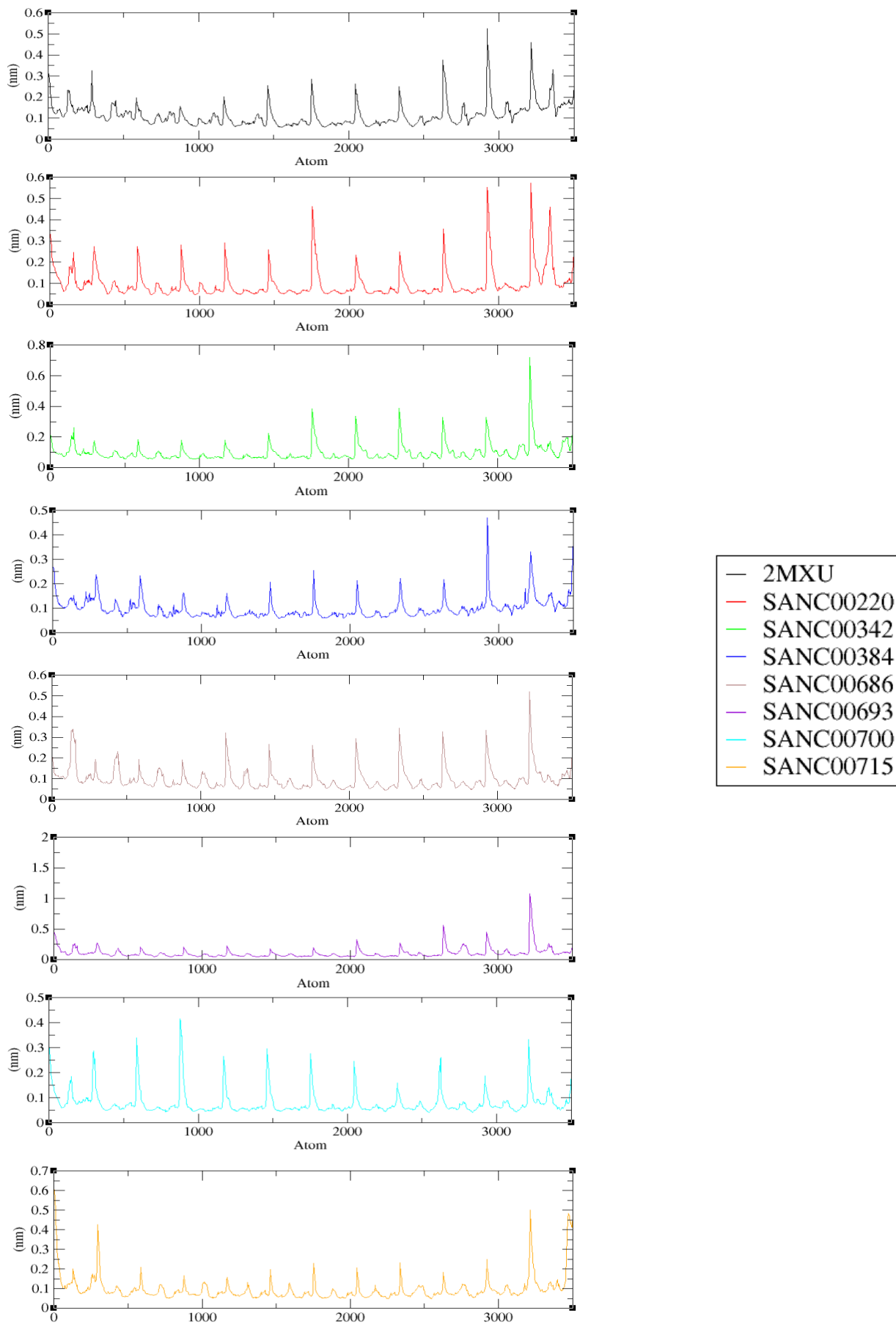**Figure 26:** RMSD and RMSF plots of the results of MD for the three dockings of Wgx-50 that was docked against 2MXU performed using GROMACS.
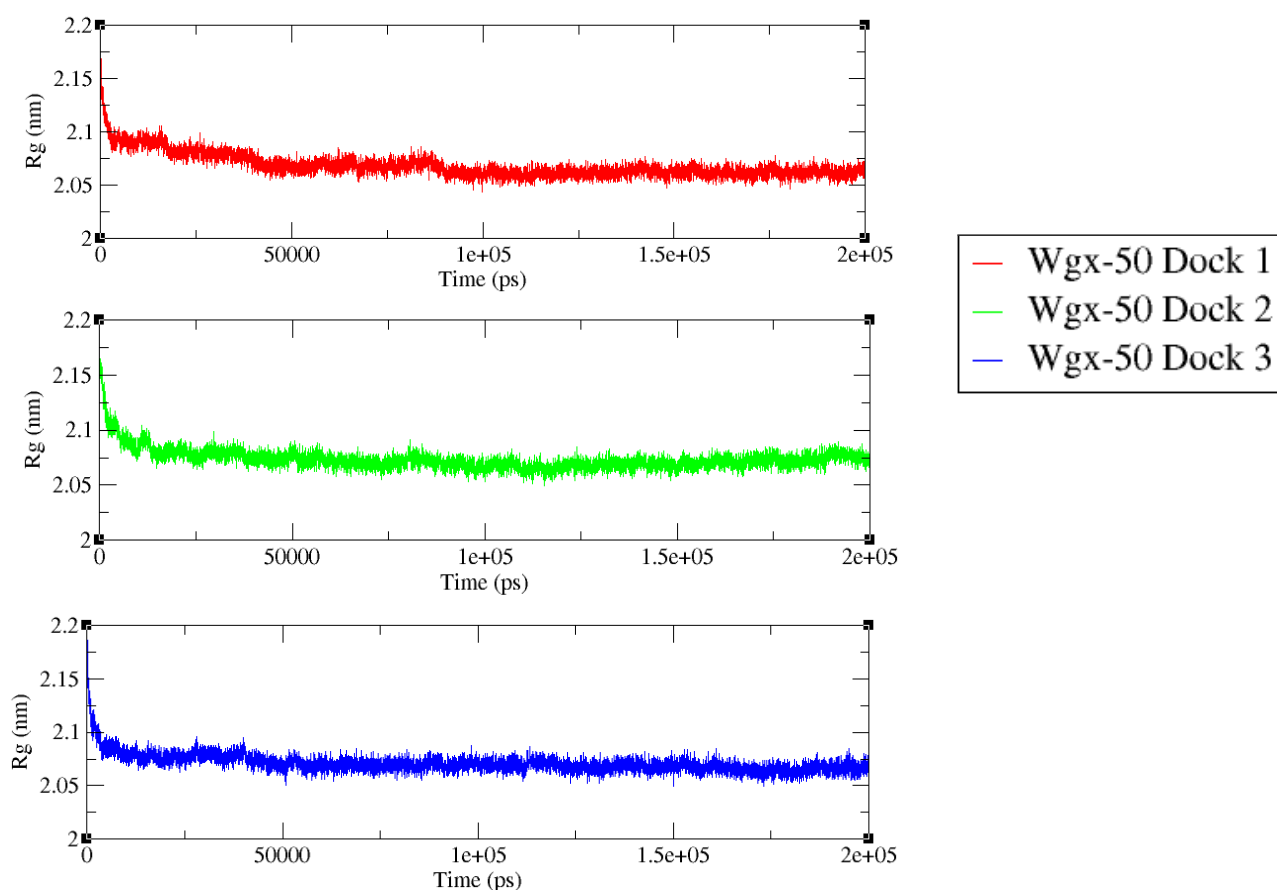
Figure 27: Radius of gyration plots of the results of MD for the three dockings of Wgx-50 that was docked against 2MXU performed using GROMACS

The molecular dynamics results for the dockings of Wgx-50 to 2MXU also showed no major difference between the complex and the apo (Figure 25 and 26), except in the RMSF plots (bottom of Figure 25). The RMSD plots had lower deviations than the *apo*. The RMSF plot for the third binding showed destabilization of the aggregate focused on the first chains (compare with SANC00481), while all showed some increase in the RMSF at the center of the chains relative to the *apo* as discussed earlier. Xu and colleagues stated that in their findings Wgx-50 was able to destabilize the aggregate structure due to the binding of the compound to the hydrophobic/aromatic side chains of the Val18-Phe20 groove and the Ile31-Met35 groove (Fan, Xu and Wei, 2017). Since Wgx-50 is hydrophobic this area is favorable for binding. Studies have shown that the hydrophobic/aromatic interactions are stabilizing forces for the binding of several ligands. Although evidence is not seen for effects in these areas, the RMSF plots do show an increase in fluctuation towards residue Glu22, and therefore the effect may be levered remote to the binding. The RMSF plots provide the most evidence of destabilization of the aggregate.

In Fan and colleagues' study they used the 2BEG structure. Therefore, for comparison MD was performed with Wgx-50 and the 2BEG structure (Figure 28). In this study the aromatic ring of Wgx-50 was packed against the side chains of Ile32 and Leu34 on β2, partially disrupted the salt bridges of Asp23-Lys28 which are crucial to the stabilization of the loop region. In our study no destabilization is evident. The RMSD of the complex is significantly lower, the RMSF shows less fluctuation of the complex (certainly no destabilization of the center of chains) and the gyration shows the structure is slightly less tightly bound. The effects on 2MXU have been more successfully identified in this study, and we are unable to identify effects on 2BEG unlike Fan and colleagues' results.
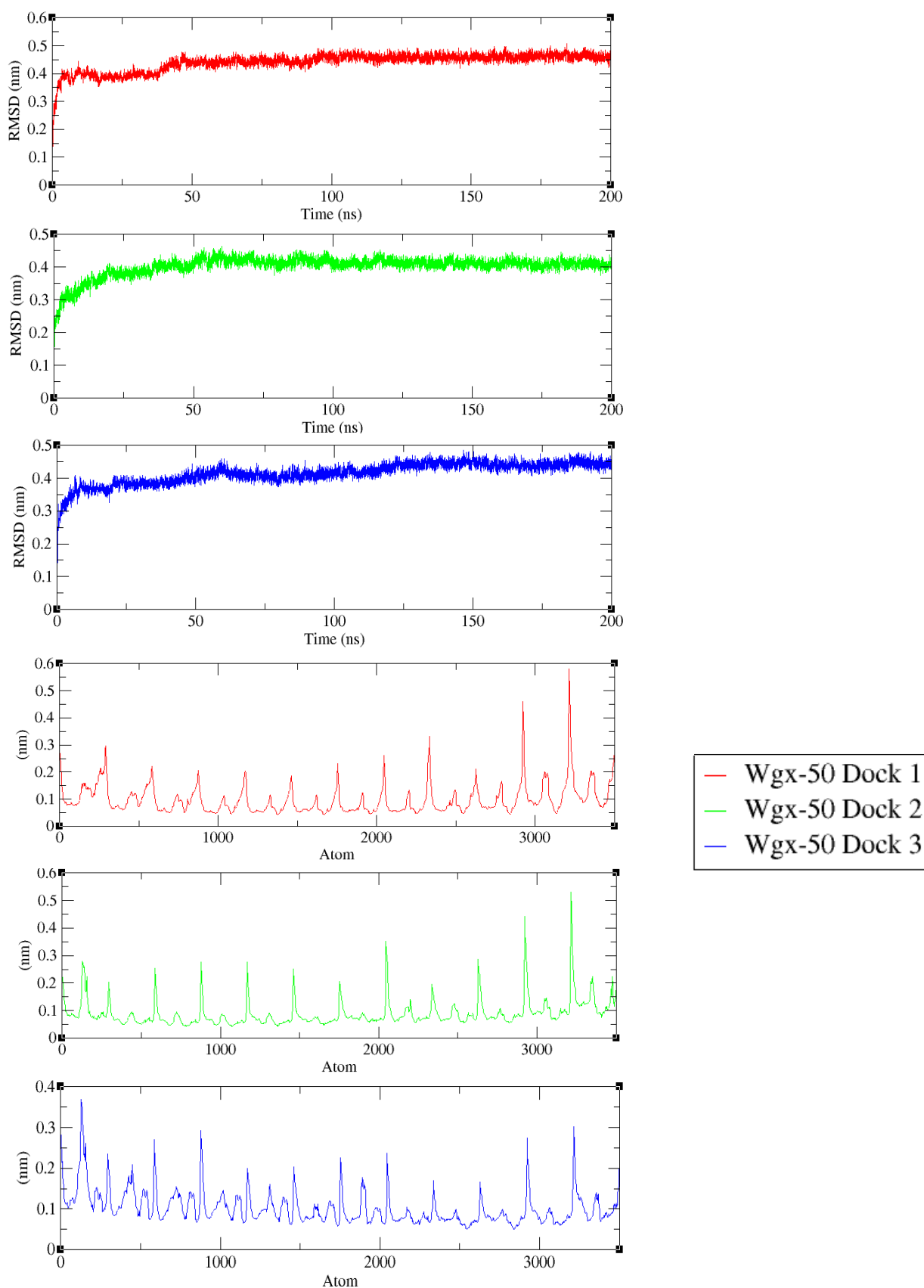

Figure 28: RMSD and RMSF plots of the results of MD for the three dockings of Wgx-50 that was docked against 2BEG performed using GROMACS.
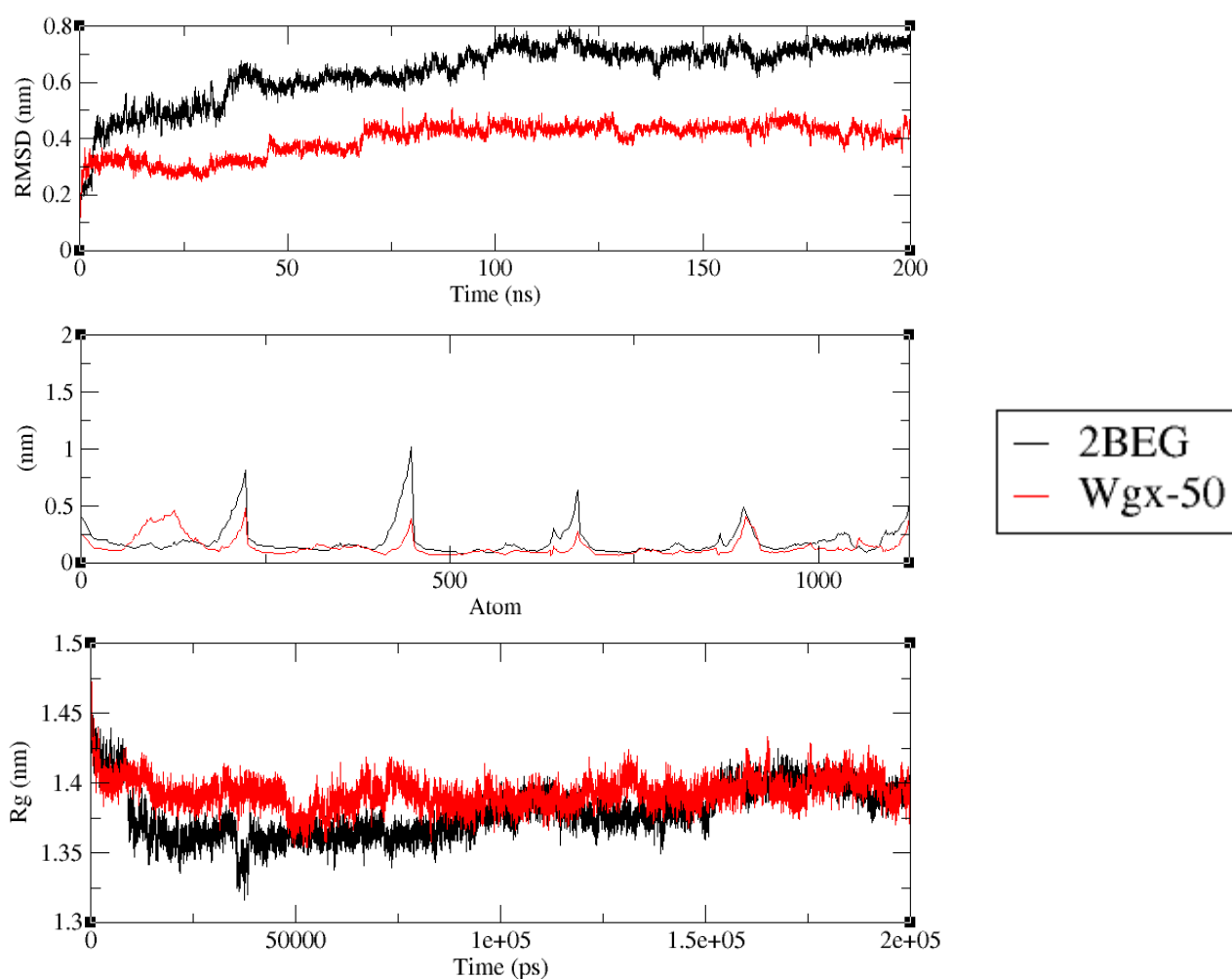
The binding free energy scores ranged from -100 to -700 with SANC00686 and SANC00693 having the lowest scores. These binding scores show satisfactory binding of the compounds and 2MXU. The reason for the low binding scores of SANC00686 and SANC00693 could be due to their large mass, allowing the compound to make more connections with the residues of the aggregate (Table 11). Wgx-50 had the highest score for the second (Table 10) and third docking (Table 11) with the third dockings score being the highest of all the results. This could be explained by its smaller size and its inability to form as many hydrogen bonds as the other compounds.

Table 10: Table summarising the MMPBSA results obtained using GROMACS for Dock 1 (centre)

Compound	Van der Waal energy(kJ/mol)	Electrostatic energy(kJ/mol)	SASA energy (kJ/mol)	Binding energy (kJ/mol)
SANC00175	-209.361 +/- 10.332	-1.292 +/- 0.973	-18.728 +/- 1.114	-201.573 +/- 10.663
SANC00290	-260.606 +/- 9.821	-42.918 +/- 5.769	-19.371 +/- 0.911	-227.366 +/- 9.730
SANC00347	-334.965 +/- 10.477	-2.422 +/- 1.440	-21.597 +/- 0.808	-315.826 +/- 10.368
SANC00348	-332.869 +/- 11.216	-3.631 +/- 2.803	-22.868 +/- 0.894	-302.085 +/- 11.735
SANC00518	-340.454 +/- 14.184	-5.776 +/- 2.837	-28.731 +/- 1.281	-326.764 +/- 13.712
SANC00553	-243.437 +/- 10.605	-19.423 +/- 5.053	-20.030 +/- 0.994	-209.495 +/- 11.281
SANC00700	-173.786 +/- 16.384	1.934 +/- 0.912	-15.502 +/- 1.129	-151.387 +/- 13.516
Wgx-50	-207.802 +/- 11.094	-25.793 +/- 6.961	-14.213 +/- 0.756	-205.132 +/- 12.548

Table 11: Table summarising the MMPBSA results obtained using GROMACS for Dock 2 (targeted to L chain)

Compound	Van der Waal energy(kJ/mol)	Electrostatic energy(kJ/mol)	SASA energy (kJ/mol)	Binding energy (kJ/mol)
SANC00178	-306.568 +/- 11.391	-19.746 +/- 6.748	-21.944 +/- 1.718	-235.752 +/- 17.437
SANC00447	-298.831 +/- 13.868	-23.391 +/- 7.783	-23.586 +/- 2.701	-232.536 +/- 36.831
SANC00478	-285.376 +/- 13.940	-27.751 +/- 6.441	-24.206 +/- 2.407	-204.394 +/- 25.334
SANC00480	-315.900 +/- 14.053	-11.546 +/- 7.005	-22.820 +/- 2.853	-236.008 +/- 59.328
SANC00481	-195.153 +/- 9.333	7.890 +/- 5.056	-14.847 +/- 2.071	-178.344 +/- 17.468
SANC00482	-344.311 +/- 14.173	-8.490 +/- 5.295	-25.152 +/- 1.789	-244.681 +/- 22.197
SANC00486	-308.968 +/- 13.331	90.932 +/- 13.163	-23.719 +/- 1.278	-242.222 +/- 16.032
Wgx-50	-146.312+/- 9.336	4.039 +/- 14.427	-12.143 +/- 0.987	-111.577 +/- 15.362

Table 12: Table summarising the MMPBSA results obtained using GROMACS for Dock 3 (blind)

Compound	Van der Waal energy(kJ/mol)	Electrostatic energy(kJ/mol)	SASA energy (kJ/mol)	Binding energy (kJ/mol)
SANC00220	-404.493 +/- 15.623	-13.459 +/- 4.384	-34.096 +/- 1.371	-379.859 +/- 16.236
SANC00342	-324.321 +/- 10.326	-1.207 +/- 1.679	-21.361 +/- 0.885	-312.716 +/- 10.522
SANC00384	-292.417 +/- 19.256	-2.106 +/- 2.181	-20.155 +/- 1.224	-250.871 +/- 19.538
SANC00686	-508.957 +/- 16.343	-8.541 +/- 4.652	-39.810 +/- 1.340	-449.590 +/- 14.986
SANC00693	-712.318 +/- 21.512	-13.387 +/- 5.919	-55.702 +/- 1.742	-645.738 +/- 21.179
SANC00700	-248.033 +/- 12.185	-0.971 +/- 1.567	-18.967 +/- 1.183	-211.262 +/- 12.575
SANC00715	-240.239 +/- 11.823	-28.176 +/- 7.423	-21.003 +/- 0.940	-224.341 +/- 12.456
Wgx-50	-172.817 +/- 10.631	-18.722 +/- 6.426	-13.602 +/- 0.844	-99.088 +/- 15.224

Chapter 4: CHARMM simulations with copper

4.1 Introduction

The main cause for the amyloid cascade hypothesis of AD is the extracellular deposition in the brain of the Aβ peptides. *In vivo* it is still unknown why Aβ forms deposits but there is much speculation on the relationship of metal ions and the conformational changes which lead to the aggregation. The binding of Cu^{2+} to the peptide prevents the peptide from adopting its usual β-sheet conformation resulting in aggregation of the peptide (Mold *et al.*, 2013). As mentioned earlier; Cu, Zn and Fe have been found in high concentrations in and surrounding AD plaques (Maynard *et al.*, 2005). The redox-active nature of Cu and Fe and defective regulation of these metals can lead to reaction with O_2 and the production of reactive oxygen species (ROS), resulting in cellular toxicity. The AD brain exhibits marked oxidative damage of proteins, lipids and nucleic acids with oxidative damage being highly concentrated in and around amyloid plaques (Maynard *et al.*, 2005). Studies are still divided over the effect of copper therefore it is important to understand the true interactions between copper and Aβ.

In order to study these interactions molecular dynamics can be used. However, these simulations require a correct set of potential energy functions commonly referred to as a force field. The force field information is generally lacking for metal ions as the pairwise-additive force fields cannot be applied to them due to their inability to handle polarization and ligand metal charge transfer effects. In order to combat this limitation quantum mechanics (QM) can be used to account for the electron structure of the atoms of the system. This technique cannot be used for systems larger than 100

atoms due to the computational expense. Therefore the use of QM in conjunction with molecular mechanics combats this issue (Musyoka *et al.*, 2018). Quantum mechanics (QM) has enhanced the understanding of the structure and reactivity of small molecular systems. Structural information, chemical reactions and accurate interation energies for hydrogen-bonded or dispersive systems are a few of the incredible outcomes of QM based methods (Merz, 2014).

Chemistry at HARvard Macromolecular Mechanics (CHARMM) is a molecular simulation program that supports multi-scale techniques such as QM/MM, MM/CG and a range of implicit solvent models. CHARMM is used to analyze biological systems including peptides, proteins, prosthetic groups, small molecule ligands etc. CHARRM provides a large suite of computational tools that include conformational and path sampling methods, free energy estimators, molecular minimization, dynamics, and analysis techniques, and model-building capabilities. Calculations with CHARMM can be performed using a number of different energy functions and models, from mixed quantum mechanical-molecular mechanical force fields, to all-atom classical potential energy functions with explicit solvent and various boundary conditions, to implicit solvent and membrane models (Brooks *et al.*, 2009)br.

4.2 Methods

4.2.1 Structure preparation

The 2MXU structure was reduced to the first five chains, chains A-E for efficiency and to match the five chains of the 2BEG structure. The various models of the two structures and by visual inspection model nine of the multimodel 2MXU structure was chosen along with model 1 of multimodel 2BEG. The periodic boundary conditions (PBC) for structures 2MXU model 9 and 2BEG model 1 were computed using a perl script. The protonation states for these structures were determined using H++ (Anandakrishnan, Aguilar and Onufriev, 2012).

4.2.2 CHARMM

The initial script (Setup.inp) required each chain to be input separately, therefore the selected models from 2MXU and 2BEG were split into their separate chains for input. After successful input the next step was vacuum minimization. The structures were then solvated and neutralized with a high molar content (1.5M) of Cu and Chlorine (Cl). The systems underwent their final minimizations, were heated in a way that ensured correct equilibration and underwent molecular

dynamics. The initial dynamics simulations were too short, so they were run again for a longer time period (200 ns).

4.2.3 Analysis

The structures were loaded onto Discovery Studio and the trajectories of the copper were analysed. Using a perl script the distances of the coppers and the closest residues to them were discovered.

4.3 Results and discussion

In order to submit the structures to CHARMM set periodic boundary conditions (PBC) need to be formed for 2MXU and 2BEG. PBCs are a set of boundary conditions, commonly used in computer simulations, which are used to approximate large systems using a portion of the structure referred to as a unit cell. The large systems approximated by PBCs consist of an infinite number of unit cells (Makov and Payne, 1994). Figure 29 shows the final structures of 2MXU and 2BEG , set as aggregates under PBC with an infinite number of chains.

A

B
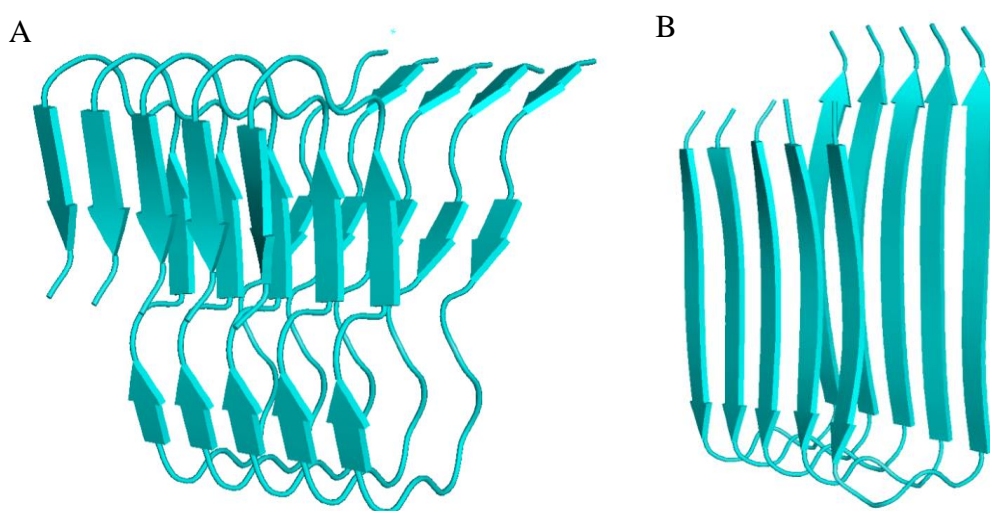
Figure 29: The resultant structures of 2MXU (A) and 2BEG (B) after their PBC were created.

The CHARMM results allowed for investigation of the movement of the Cu ions throughout the simulation. Figure 26 shows an overlay of the copper atoms in all frames for the trajectories, showing the localization of the copper during the molecular dynamics. The distance script illustrated that the residue the coppers interact most with (this interaction in the simulation is only from the set van der

Waal's and electrostatic terms) is Met35 which visually may also be seen in Figure 26B. Studies have shown a relationship between Met35 and Cu ions and their contribution to the neurotoxic activity of Aβ.

Methionine residues can be spontaneously oxidized to methionine sulfoxide (Met(O)35) by oxidants such as H_2O_2 and molecular oxygen. This oxidation pathway does not result in the formation of reactive oxygen species or free radicals. Methionine oxidation to sulfoxide is shown to inhibit fibril formation and this observation has also lead to various speculations about the essential role of the oxidation of Met35 in AD (Friedemann *et al.*, 2015).

The oxidation of Met35 to (Met(O)35) has been linked to Aβ-induced oxidative damage. Met35 is one of the most intriguing amino acid residues in the peptide since it is the most easily oxidized side chain in the peptide and it is partially oxidized in post mortem amyloid plaques (Friedemann *et al.*, 2015). Solution NMR studies of reduced and oxidized Aβ(1-40) and Aβ(1-42) have shown little conformational difference among the peptides, suggesting the effect of Met35 on the biological activity of the peptide is visible in the assembled state rather than the monomeric state (Syme *et al.*, 2004).
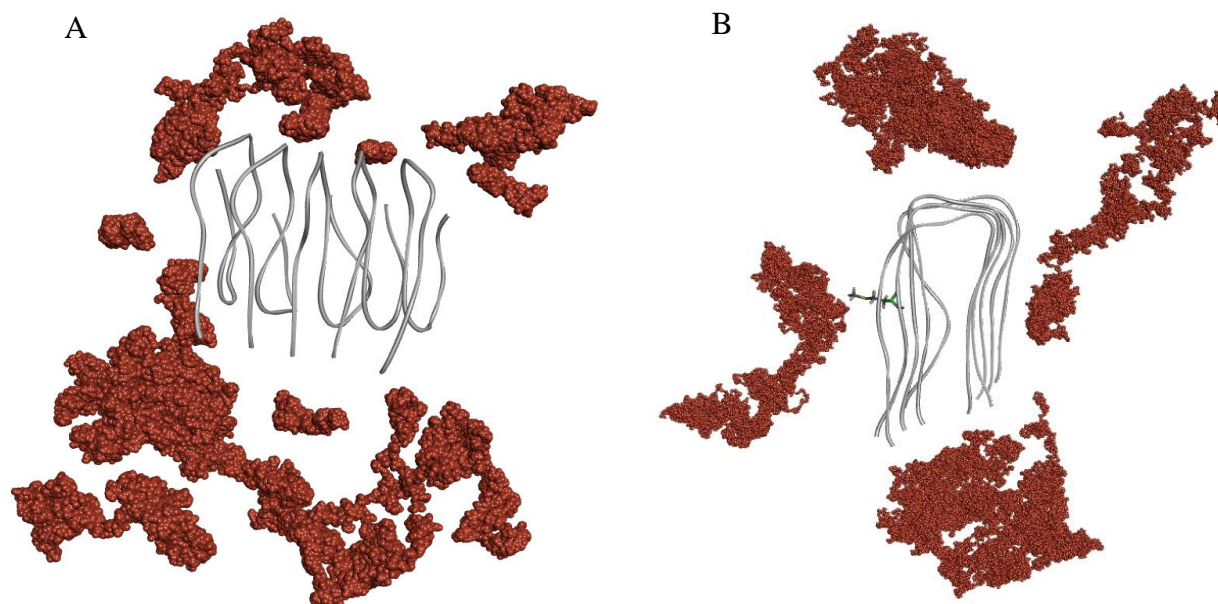

Figure 30: Images of the structures after CHARMM and the movement of the copper ions during the simulation. A) 2MXU. B) 2BEG.

Taking these results further could offer more insight into the interaction between Met35 and Cu and the effect that relationship has on aggregation and fibril formation.

5. Conclusion

Three separate dockings, each targeting separate sites of the 2MXU aggregate were performed and resulted in seven top hits for each. These top hits were analysed based on structure and were taken to molecular dynamics simulations. These simulations were analysed based on RMSD, RMSF and radius of gyration plots. The RMSF analyses showed promising results in terms of compounds destabilizing parts of the aggregate in a similar way to that of wgx-50. Further analysis was performed on the aggregates by solvating the structures including a high concentration of copper and performing molecular dynamics within CHARMM. The CHARMM results showed the residue Met35 being the residue the Cu ions were closes to and studies have shown a link between the relationship between copper and Met35, and AD. However, due to time constraints these results were not taken further to QM/MM molecular dynamics studies, but they do provide evidence that Cu is an important ion to study with regards to Aβ and its aggregation into mature fibrils.

6. Future work

In future, high throughput molecular dynamics could be done with compounds from other databases such as the ZINC database in order to broaden the scope of compounds and their effects on the aggregates. HTVS on a monomeric peptide could also be performed to analyse the interactions and to identify if any of them will prevent the aggregation of the fibrillar structure. Further QM/MM work with copper and within CHARMM could be done in order to explore the oxidation of Met25 and the chemical transformations taking place that lead to this oxidation.

7. References

1. A. Armstrong, R. (2014) 'A critical analysis of the "amyloid cascade hypothesis"', *Folia Neuropathologica*, 3(3), pp. 211–225. doi: 10.5114/fn.2014.45562.

2. Anandakrishnan, R., Aguilar, B. and Onufriev, A. V (2012) 'H + + 3 . 0 : automating p K prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations', 40(May), pp. 537–541. doi: 10.1093/nar/gks375.

3. Bajorah, J. (2002) 'Integration of virtual and high-throughput screening', *Nature Reviews Drug Discovery*, 1, pp. 882–894.

4. Berman, H. M. *et al.* (2000) 'The Protein Data Bank', 28(1), pp. 235–242.

5. Berrill, A., Biddlecombe, J., Bracewell, D. (2011) 'Chapter 13 - Product Quality During Manufacture and Supply', Peptide and Protein Delivery. *Academic Press*. pp. 313-339.

6. BIOvIA (2015) 'Discovery studio modeling environment', *Dassault Systemes*.

7. Bird, T. D. (1998) *Alzeihmer Disease Overview*.

8. Bitan, G. *et al.* (2003) 'A Molecular Switch in Amyloid Assembly: Met35 and Amyloid β-Protein Oligomerization', *Journal of the American Chemical Society*, 125(50). doi: 10.1021/ja0349296.

9. Bleicher, K. H. *et al.* (2003) 'Hit and lead generation: beyond high-throughput screening', *Nature Reviews Drug Discovery*, 2, pp. 369–378.

10. Brooks, B. *et al.* (2009) 'CHARMM: The Biomolecular Simulation Program', *NIH*, 30(10), pp. 1545–1614. doi: 10.1002/jcc.21287.CHARMM.

11. Buchete, N. V., Tycko, R. and Hummer, G. (2005) 'Molecular dynamics simulations of Alzheimer's β-amyloid protofilaments', *Journal of Molecular Biology*, 353(4), pp. 804–821. doi: 10.1016/j.jmb.2005.08.066.

12. Bush, A. I., Masters, C. L. and Tanzi, R. E. (2003) 'Copper, beta amyloid and AD: Tapping a sensitive connection', *Pnas*, 100(20), pp. 11193–11194.

13. C Lipinski (2004) 'Lead- and drug-like compounds: the rule-of-five revolution', *Elsevier*, 1(4), pp. 337–341.

14. Chen, G. F. *et al.* (2017) 'Amyloid beta: Structure, biology and structure-based therapeutic development', *Acta Pharmacologica Sinica*. Nature Publishing Group, 38(9), pp. 1205–1235. doi: 10.1038/aps.2017.28.

15. Ciechanover, A & Kwon, Y.T. (2015) 'Degradation of misfolded proteins in neurodegenerative diseases: therapeutic targets and strategies', *Experimental & Molecular Medicine*. 47, pp. e147. doi: 10.1038/emm.2014.117

16. Daina, A., Michielin, O. and Zoete, V. (2017) 'SwissADME : a free web tool to evaluate pharmacokinetics , drug- likeness and medicinal chemistry friendliness of small molecules', *Nature Publishing Group*. Nature Publishing Group, (January), pp. 1–13. doi: 10.1038/srep42717.

17. Delano, W. L. (2002) 'Pymol: An open-source molecular graphics tool.', *CCP4 Newsletter On Protein Crystallography*, 40, pp. 82–92.

18. Devadoss, F. R. and Raj, V. P. (2014) *Analysis and Visual Summarization of Molecular Dynamics Simulation*.

19. Dorszewska, J. *et al.* (2016) 'Molecular Basis of Familial and Sporadic Alzheimer's Disease', *Current Alzheimer Research*, 13(9), pp. 952–963. doi: 10.2174/1567205013666160314150501.

20. Estrada, L. and Soto, C. (2007) 'Disrupting β-Amyloid Aggregation for Alzheimer Disease Treatment', *Current Topics in Medicinal Chemistry*, 7(1), pp. 115–126. doi: 10.2174/156802607779318262.

21. Fan, H. M. *et al.* (2015) 'Destabilization of Alzheimer's Aβ42 Protofibrils with a Novel Drug Candidate wgx-50 by Molecular Dynamics Simulations', *Journal of Physical Chemistry B*, 119(34), pp. 11196–11202. doi: 10.1021/acs.jpcb.5b03116.

22. Fan, H., Xu, Q. and Wei, D. (2017) 'Recent studies on mechanisms of new drug candidates for Alzheimer's disease interacting with amyloid protofibrils using molecular dynamics simulations', (April). doi: 10.1007/978-94-024-1045-7.

23. Friedemann, M. *et al.* (2015) 'Effect of methionine-35 oxidation on the aggregation of amyloid- β peptide', *Biochemistry and Biophysics Reports*. Elsevier, 3, pp. 94–99. doi: 10.1016/j.bbrep.2015.07.017.

24. Gaugler, J. *et al.* (2016) '2016 Alzheimer's disease facts and figures', *Alzheimer's and Dementia*. Elsevier Inc., 12(4), pp. 459–509. doi: 10.1016/j.jalz.2016.03.001.

25. Ghose, A., Viswanadhan, V. . and Wendoloski, J. . (1999) 'A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases', *J. Comb .Chem*, 1(1), pp. 55–68.

26. Haass, C. and Selkoe, D. (2007) 'Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide', *Nature reviews*, 8, pp. 101–112.

27. Hatherley, R. *et al.* (2015) 'SANCDB: A South African natural compound database', *Journal of Cheminformatics*. Springer International Publishing, 7(1), pp. 1–9. doi: 10.1186/s13321-015-0080-8.

28. Henzler-Wildman, K. and Kern, D. (2007) 'Dynamic personalities of proteins', *Nature*,

450(7172), pp. 964–972. doi: 10.1038/nature06522.

29. Hou, S., Gu, R. X. and Wei, D. Q. (2017) 'Inhibition of β-Amyloid Channels with a Drug Candidate wgx-50 Revealed by Molecular Dynamics Simulations', *Journal of Chemical Information and Modeling*, 57(11), pp. 2811–2821. doi: 10.1021/acs.jcim.7b00452.

30. Hou, T. *et al.* (2011) 'Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations', *JCIM*, 51(1), pp. 69–82.

31. Iqbal, K. *et al.* (2010) 'Tau in Alzheimer Disease and Related Tauopathies', *National institute of health*, 7(8), pp. 656–664. doi: 10.2174/156720510793611592.

32. De Jager, C. A. *et al.* (2017) 'Dementia Prevalence in a Rural Region of South Africa: A Cross-Sectional Community Study', *Journal of Alzheimer's Disease*, 60(3), pp. 1087–1096. doi: 10.3233/JAD-170325.

33. Laskowski, R. A. and Swindells, M. B. (2011) 'LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery', *J. Chem. Inf. Model*, 51(10), pp. 2778–2786.

34. Lorieau, J. L. and McDermott, A. E. (2006) 'Conformational flexibility of a microcrystalline globular protein: Order parameters by solid-state NMR spectroscopy', *Journal of the American Chemical Society*, 128(35), pp. 11506–11512. doi: 10.1021/ja062443u.

35. Luhrs, T. *et al.* (2005) '3D structure of Alzheimer's amyloid-β(1–42) fibrils', *PNAS*, pp. 17342–17347. doi: 10.1073/pnas.95.11.6448.

36. Mabkot, Y. *et al.* (2016) 'Antimicrobial Activity of Some Novel Armed Thiophene Derivatives and Petra/Osiris/Molinspiration (POM) Analyses', *Molecules*, 27(222), pp. 1–16. doi: 10.3390/molecules21020222.

37. Makov, G. and Payne, M. (1994) 'Periodic boundary conditions in ab initio calculations', *Physical review*, 51(7).

38. Maynard, C. J. *et al.* (2005) 'Metals and amyloid-β in Alzheimer's disease', *International Journal of Experimental Pathology*, 86(3), pp. 147–159. doi: 10.1111/j.0959-9673.2005.00434.x.

39. McGovern, S. L. *et al.* (2002) 'A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening', *Journal of Medicinal Chemistry*, 45(8), pp. 1712–1722. doi: 10.1021/jm010533y.

40. Merz, K. M. (2014) 'Using Quantum Mechanical Approaches to Study Biological Systems'. doi: 10.1021/ar5001023.

41. Mold, M. *et al.* (2013) 'Copper prevents amyloid-b1–42 from forming amyloid fibrils under near-physiological conditions in vitro', *Scientific reports*, pp. 1–6. doi: 10.1038/srep01256.

42. Moses, V., Tastan Bishop, Ö. and Lobb, K. A. (2017) 'The evaluation and validation of

copper (II) force field parameters of the Auxiliary Activity family 9 enzymes', *Chemical Physics Letters*. Elsevier B.V., 678, pp. 91–97. doi: 10.1016/j.cplett.2017.04.022.

43. Murphy, M. P. and Levine, H. (2010) 'Alzheimer's Disease and the β-Amyloid Peptide', *Journal of Alzheimer's Disease*, 19(1), pp. 1–17. doi: 10.3233/JAD-2010-1221.Alzheimer.

44. Musyoka, T. *et al.* (2018) 'The determination of CHARMM force field parameters for the Mg2+containing HIV-1 integrase', *Chemical Physics Letters*. Elsevier, 711(September), pp. 1–7. doi: 10.1016/j.cplett.2018.09.019.

45. Mutt, E. and Sowdhamini, R. (2016) 'Molecular Dynamics Simulations and Structural Analysis to Decipher Functional Impact of a Twenty Residue Insert in the Ternary Complex of Mus musculus TdT Isoform', *PLoS ONE*, pp. 1–24. doi: 10.1371/journal.pone.0157286.

46. Nair, N. T., Nisthar, S. and Sunil, B. A. (2016) 'Identification of novel drug candidate against Mycobacterium Tuberculosis InhA protein through Computer aided drug discovery', 50(046). doi: 10.5530/ijper.50.4.9.

47. Nigam, Y., Knight, J. and Jones, A. (2009) 'Effects of bedrest 3: musculoskeletal and immune systems, skin and self-perception', *Nursing times*.

48. Ono, K., Condron, M.M., Teplow, D.B. (2009) 'Structure-neurotoxicity relationships of amyloid β-protein oligomers', *Proceedings of the National Academy of Sciences*. 106(35), pp. 14745-14750.

49. Peng, J. *et al.* (2015) 'Effects of cordycepin on the microglia-overactivation-induced impairments of growth and development of hippocampal cultured neurons', *PLoS ONE*, 10(5), pp. 1–18. doi: 10.1371/journal.pone.0125902.

50. Petrucci, R. H. (1997) *General chemistry : principles and modern applications /.*

51. Prasanna, S. and Doerksen, R. . (2009) 'Topological polar surface area: a useful descriptor in 2D-QSAR.', *J. Pub. Med*, 16(1), pp. 21–41.

52. Reitz, C. (2012) 'Alzheimer's disease and the amyloid cascade hypothesis: A critical review', *International Journal of Alzheimer's Disease*, 2012. doi: 10.1155/2012/369808.

53. Ribas, J. *et al.* (2002) 'Theoretical Study of Alkyl- π and Aryl- π Interactions . Reconciling Theory and Experiment', (30), pp. 7057–7065. doi: 10.1021/jo0201225.

54. Ridewood, W. (1908) 'A new species of Cephalodiscus (C. gilchristi) from the Cape Seas', 4, pp. 173–192.

55. Scheiner, S., Kar, T. and Pattanayak, J. (2002) 'Comparison of Various Types of Hydrogen Bonds Involving Aromatic Amino Acids', (11), pp. 13257–13264. doi: 10.1021/ja027200q.

56. Subramaniam, S., Mehrotra, M. and Gupta, D. (2008) 'Virtual high throughput screening (vHTS) – A perspective', *Bioinformation*, 3(1), pp. 14–17. doi: 10.6026/97320630003014.

57. Swain, M. (2012) 'chemicalize.org', *J. Chem. Inf. Model*, 52(2), pp. 613–615.

58. Syme, C. D. *et al.* (2004) 'Copper binding to the amyloid-β (aβ) peptide associated with Alzheimer's disease: Folding, coordination geometry, pH dependence, stoichiometry, and affinity of Aβ-(1-28): Insights from a range of complementary spectroscopic techniques', *Journal of Biological Chemistry*, 279(18), pp. 18169–18177. doi: 10.1074/jbc.M313572200.

59. Trott, O. and Olson, A. J. (2009) 'Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading', *Wiley InterScience*. doi: 10.1002/jcc.

60. Trott, O. and Olson, A. J. (2010) 'AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading', *J Comput Chem*, 31(2), pp. 455–461. doi: 10.1002/jcc.21334.AutoDock.

61. Veber, D. . *et al.* (2002) 'Molecular Properties That Influence the Oral Bioavailability of Drug Candidates', *J. Med. Chem*, 45(12), pp. 2615–2623.

62. Vistoli, G., Pedretti, A. and Testa, B. (2008) 'Assessing drug-likeness – what are we missing?', *Elsevier*, 13(7–8), pp. 285–294.

63. Xiao, Y. *et al.* (2015) 'Aβ(1–42) Fibril Structure Illuminates Self-recognition and Replication of Amyloid in Alzheimer's', *Nat Struct Mol Biol.*, 22(6), pp. 499–505. doi: 10.1038/nsmb.2991.A.

8. Appendix:

8.1. Vina script example dock 1

```
receptor = receptor/2mxu_model1.pdbqt
ligand = SANC00175_minRM1.pdbqt
out = results/SANC00175_minRM1.all.pdbqt
log = SANC00175_minRM1.log
center_x = 0
center_y = 0
center_z = 0

size_x = 20
size_y = 20
size_z = 20

exhaustiveness = 128
cpu = 4
```

8.2 Vina script example dock 2

```
receptor = receptor/2mxu_model1.pdbqt
ligand = SANC00365_minRM1.pdbqt
out = results/SANC00365_minRM1.all.pdbqt
log = SANC00365_minRM1.log
center_x = 0
center_y = 0
center_z = 0

size_x = 100
size_y = 100
size_z = 100

exhaustiveness = 128
cpu = 4
```

8.3 Vina script example dock 3

```
receptor = receptor/2mxu_model1.pdbqt
ligand = SANC00365_minRM1.pdbqt
out = results/SANC00365_minRM1.all.pdbqt
log = SANC00365_minRM1.log
center_x = 22.493
center_y = -21.342
center_z = -1.702

size_x = 40
size_y = 40
size_z = 40

exhaustiveness = 128
cpu = 4
```

8.4. Example PBC scripts (2BEG)

```perl
#!/usr/bin/perl


use Math::Trig;
use POSIX;

...

my $reference1="C    LEU A  17";
my $reference2="C    LEU E  17";

my @coords1;
my @coords2;

open(PDB,"< 2beg_model1.pdb");
while(my $line=<PDB>)
{
    if($line=~m/ATOM/)
    {
        chomp $line;
#        print "$line \n";
        my @segments = split /\s+/, $line;
#ATOM      4 H13                 -45.543  -8.510 -14.158  0.00  0.00          H
#ATOM   1856  HA  ALA E  42      -21.474   6.990 -16.560  1.00  0.00          H
        $atommatrix[$numberoffiles][$atomnumber][0]=$segments[6];
        $atommatrix[$numberoffiles][$atomnumber][1]=$segments[7];
        $atommatrix[$numberoffiles][$atomnumber][2]=$segments[8];
        $atommatrix[$numberoffiles][$atomnumber][3]=$segments[12];
        #print "x  $atommatrix[$numberoffiles][$atomnumber][0] y  $atomma-
trix[$numberoffiles][$atomnumber][1]  $atommatrix[$numberoffiles][$atom-
number][2]\n";
        if($line=~ m/$reference1/){@coords1=($segments[6],$segments[7],$seg-
ments[8]);print "success\n";}
        if($line=~ m/$reference2/){@coords2=($segments[6],$segments[7],$seg-
ments[8]);print "success\n";}
        $atomnumber++;

    }
}
close(PDB);
#my @atommatrix;
#for(my $i=1;$i<$atomnumber;$i++){print " $atommatrix[$numberoffiles][$i][0]
$atommatrix[$numberoffiles][$i][1] $atommatrix[$numberoffiles][$i][2] $atomma-
trix[$numberoffiles][$i][3]       \n"}
#$numberoffiles++;

print "REFERENCES:\n";
print @coords1;
print "\n";
print @coords2;
print "\n";
my @bondvector=($coords1[0]-$coords2[0],$coords1[1]-$coords2[1],$coords1[2]-
$coords2[2]);
print "bond vector\n";
print @bondvector;
print "\n";
normalize (\@bondvector);
```

```perl
printf "bondvector: %6.4f %6.4f %6.4f\n",$bondvector[0],$bondvector[1],$bondvec-
tor[2];
print "\n";
my @xvector=(1,0,0);

my @rotationvector=cross(\@bondvector,\@xvector);
#my @rotationvector=cross(\@xvector,\@bondvector);

my $theta=angle(\@coords1,\@coords2,\@xvector);


print "angle $theta rotate about $rotationvector[0] $rotationvector[1] $rota-
tionvector[2] \n";

$theta=ceil($theta)*10;
$theta=930;

my @newcoords1;
my @newcoords2;

my @max;
$max[0]=-100000;
$max[1]=-100000;
$max[2]=-100000;

my @min;

$min[0]=10000;
$min[1]=10000;
$min[2]=10000;

my @extent=(-3.56999010722755,-12,-41);

open(PDB,"< 2beg_model1.pdb");
while(my $line=<PDB>)
{
    if($line=~m/ATOM/)
    {
        chomp $line;
        #print "$line \n";
        my @segments = split /\s+/, $line;
#ATOM      4 H13                 -45.543  -8.510 -14.158  0.00  0.00           H

#01234567890123456789012345678901234567890123456789012345678901234567890123456789012345678
90
#ATOM   1856  HA   ALA E  42     -21.474   6.990 -16.560  1.00  0.00           H
        my $first=substr $line,0,27;
        my $last=substr $line,54,1000;

        my @atomcoords=($segments[6],$segments[7],$segments[8]);


        my @newcoords=translate(\@coords2,\@atomcoords);
        my @origin= (0,0,0);
        #print "**coords $atomcoords[0] $atomcoords[1] $atomcoords[2]  \n";
        my @rotated=translateandrotate(\@origin,\@rotationvector,\@new-
coords,$theta,\@cosinematrix,\@sinematrix);
        if($line=~ m/$reference1/){@newcoords1=($rotated[0],$rotated[1],$ro-
tated[2]);print "success\n";}
        if($line=~ m/$reference2/){@newcoords2=($rotated[0],$rotated[1],$ro-
tated[2]);print "success\n";}

        my @nrotated=translate(\@extent,\@rotated);
```

```perl
        if($nrotated[0]>$max[0]){$max[0]=$nrotated[0]}
        if($nrotated[1]>$max[1]){$max[1]=$nrotated[1]}
        if($nrotated[2]>$max[2]){$max[2]=$nrotated[2]}
        if($nrotated[0]<$min[0]){$min[0]=$nrotated[0]}
        if($nrotated[1]<$min[1]){$min[1]=$nrotated[1]}
        if($nrotated[2]<$min[2]){$min[2]=$nrotated[2]}


        printf "$first    "."%8.3f"."%8.3f"."%8.3f"."$last\n", $nrotated[0],$nro-
tated[1],$nrotated[2];
#       print "** $first $newcoords[0] $newcoords[1] $newcoords[2] $last \n";



    }
}
close(PDB);



my @newbondvector=($newcoords1[0]-$newcoords2[0],$newcoords1[1]-$new-
coords2[1],$newcoords1[2]-$newcoords2[2]);
print "new bond vector\n";
print @newbondvector;
print "\n";
normalize (\@newbondvector);
printf "newbondvector: %6.4f %6.4f %6.4f\n",$newbondvector[0],$newbondvec-
tor[1],$newbondvector[2];
print "\n";

my $theta1=angle(\@newcoords1,\@newcoords2,\@xvector);

my @zvector=(0,0,1);
my @origin=(0,0,0);
my @randomvector=(1,1,0);
my $theta1=angle(\@newcoords1,\@newcoords2,\@xvector);
#my $theta1=angle(\@xvector,\@origin,\@randomvector);
$theta1=ceil($theta1);

print "Final angle is $theta1\n";
print "max $max[0] $max[1] $max[2] min $min[0] $min[1] $min[2]\n";
```

8.5 Distance for CHARMM results scripts


```perl
#!/usr/bin/perl
#


#ATOM   1851 HG21 ILE E 129    -10.154   5.504  20.240  0.00  0.00       E

my $frame=0;
my $atomnumber=0;
my @atommatrix;
open(PDB,"< trajectory2mxu.pdb");

while(my $line=<PDB>)
{
  if(not ($line =~ m/ATOM/))
  {
    ################??????????????????
    for(my $i=0;$i<$atomnumber;$i++)
```

```perl
      {
         my $res1=$atommatrix[$i][4];
         my $rno1=$atommatrix[$i][6];
         my $ch1=$atommatrix[$i][5];
         my $num1=$atommatrix[$i][3];
         my $x1=$atommatrix[$i][0];
         my $y1=$atommatrix[$i][1];
         my $z1=$atommatrix[$i][2];
         if($res1 =~ m/CU/)
         {
             for(my $j=0;$j<$atomnumber;$j++)
             {
                 my $res2=$atommatrix[$j][4];
                 my $rno2=$atommatrix[$j][6];
                 my $ch2=$atommatrix[$j][5];
                 my $num2=$atommatrix[$j][3];
                 my $x2=$atommatrix[$j][0];
                 my $y2=$atommatrix[$j][1];
                 my $z2=$atommatrix[$j][2];
                 my $id=$atommatrix[$j][7];
                 if(not(($res2 =~ m/CU/)or($res2 =~ m/CLA/)))
                 {
                     my $distance=sqrt(($x1-$x2)*($x1-$x2)+($y1-$y2)*($y1-
$y2)+($z1-$z2)*($z1-$z2));
                     if($distance<3.5)
                     {
                         print "$frame : $res1 $rno1 $ch1 $num1 --- $res2 $rno2
$ch2 $num2 $id : $distance\n";
                     }
                 }
             }
         }
      }
      $atomnumber=0;
      $frame++;
   }
   else
   {
       my($atomcard,$number,$id,$resi-
due,$chain,$resno,$x,$y,$z,$temp1,$temp2,$temp3)=split /\s+/, $line;
       $atommatrix[$atomnumber][0]=$x;
       $atommatrix[$atomnumber][1]=$y;
       $atommatrix[$atomnumber][2]=$z;
       $atommatrix[$atomnumber][3]=$number;
       $atommatrix[$atomnumber][4]=$residue;
       $atommatrix[$atomnumber][5]=$chain;
       $atommatrix[$atomnumber][6]=$resno;
       $atommatrix[$atomnumber][7]=$id;
       $atomnumber++;
       #print "$x $y $z $number $residue $chain $resno\n";
   }

}
close PDB;
```