Molecular modeling and simulation studies to prioritize sequence variants identified by whole-exome sequencing in a South African family with Parkinson's disease

By

Maryam Hassan (3576772)

Thesis submitted in the fulfillment for the degree of Master of Science (Bioinformatics) at the South African National Bioinformatics Institute, Faculty of Natural Sciences, University of the Western Cape.

Supervisor: Dr Ruben Cloete

South African National Bioinformatics Institute

Faculty of Natural Sciences

University of the Western Cape

Co-supervisor: Prof. Soraya Bardien

Division of Molecular Biology and Human Genetics

Faculty of Medicine and Health Sciences

Stellenbosch University

August 2021

Declaration

I, **Maryam Hassan,** declare that the contents of this thesis represent my work and that the thesis has not been submitted before for any degree or academic examination in any other university. Furthermore, it expresses my own opinions and not necessarily those of the University of the Western Cape.

Signed: _M Hassan_   **Date:** 11th August 2021

Dedication

This one is for you Hooyo (Mum), Aabo (Dad), and all those whose lives have been affected by

Parkinson's disease in some way.

Acknowledgments

Firstly, a huge thank you to the Almighty for granting me the strength and ability to complete this research. Say: '...then when you have taken a decision, put your trust in God. God loves those who put their trust (in Him)'. Surat Al-Emran Ayah (159) 'Quran'.

To my supervisors Dr. Ruben Cloete and Prof. Soraya Bardien: thank you for all your guidance, advice, patience, and unwavering support throughout this research. Thank you for always taking the time out when necessary and going the extra mile.

To my parents, 'Abdulqadir Hassan and Hukumo Mohamed', no words can explain what you have done and still do for me. Thank you for being my pillars of strength and for always encouraging me to pursue my dreams and accomplish my goals. You have sacrificed so much for me and without your support, I would not have made it this far. Many thanks to my siblings and family for constantly reminding me that nothing is impossible, and that one can accomplish whatever one sets their mind on.

My sincere appreciation to the National Research Foundation (NRF) for providing the financial support required to conduct this study. Many thanks to all South African National Bioinformatics Institute members for their support, encouragement and positive attitude. Many thanks to the administrative team, Mrs. Ferial Mullins, Mrs. Fungiwe Mpithi and Mr. Duane Kellerman. I would also like to appreciate Miss. Katelyn Cuttler, Miss. Rumbidzai Chitongo, Mr. Darius Martin and Mr. Darren Isaacs; thank you for all the training, assistance and enrichment both on an academic and personal level. Many thanks to the Parkinson's research group at Stellenbosch University for all their support.

To my friends Tusekile Kangwa, Keletso Modise, Asmaa Mohamed and Asmaa Adem; thank you for always reminding me that there is a light at the end of the tunnel and most importantly, for making this MSc journey a memorable one.

Table of Contents

UNIVERSITY *of the*

WESTERN CAPE

List of tables

List of figures:

Scientific contributions

1. Oral presentation (virtual) at the Genetic Epidemiology of Parkinson's disease (GeoPD) consortium at Milan, Italy, October 2020.

   Title: Prioritizing a putative causal variant on the *NPSR1* gene identified in a South African Indian family with Parkinson's disease.

2. Flash presentation (virtual) at the Africans in STEM conference at Cambridge University, United Kingdom, April 2021.

   Title: Prioritizing a putative causal variant on the *NPSR1* gene identified in a South African Indian family with Parkinson's disease.

   Comment: Winner of best flash presentation.

3. Poster presentation (virtual) at the Africans in STEM conference at Cambridge University, United Kingdom, April 2021.

   Title: Prioritizing a putative causal variant on the *NPSR1* gene identified in a South African Indian family with Parkinson's disease.

4. Talk at the South African Medical Research Council (SAMRC), August 2021.

   Title: Molecular modeling and simulation studies to prioritize sequence variants identified by whole-exome sequencing in a South African family with Parkinson's disease.
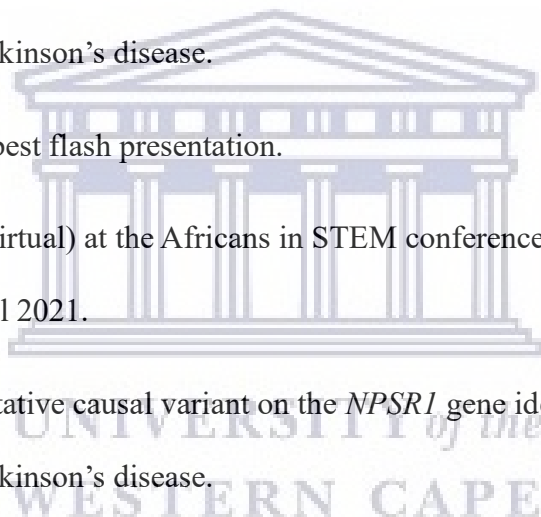
Abbreviation of terms

CPU   Central processing unit

GPU   Graphics processing unit

MD   Molecular dynamics

MT   variant

NMR   Nuclear magnetic resonance

PCA   Principal component analysis

PD   Parkinson's disease

Rg   Radius of gyration

RMSF   Root mean square fluctuation

RMSD   Root mean square deviation

SSA   Sub-Saharan Africa

UMD   Universal mutation databases

WES   Whole-exome sequencing

WT   wild-type

Abstract

Parkinson's disease (PD) is a neurodegenerative disorder that occurs due to a loss of dopaminergic neurons in the substantia nigra. It is one of the most common neurodegenerative disorders, ranking second only to Alzheimer's disease. Research on the genetic causes of PD over the past two decades has led to the discovery of several PD-associated genes. Currently, researchers have identified 23 genes that are linked to rare monogenic forms of PD with Mendelian inheritance. In sub-Saharan Africa (SSA), PD has received little attention due to factors such as underfunded healthcare infrastructure, the absence of epidemiological data, and a scarcity of neurologists. In the relatively few published studies, it has been shown that the known PD mutations play a minor role in disease etiology in SSA populations. In the current study, we follow up on previous work done in an MMed study investigating a South African family with several family members (mother and three sons) suffering from PD. The parents are both of Indian descent. Whole-exome sequencing (WES) had been performed on three affected members of this family, and 11 sequence variants had been prioritized for further study. The present study aimed to use *in-silico* methods to provide additional support for putative causal variants from WES data on a South African Indian family with PD.

In this study, the Universal Mutation Databases (UMD) predictor was used to confirm the positions of the 11 sequence variants. Next, pathway, protein-protein interactions, and expression analyses were performed to determine brain-specific gene function and mRNA expression profiles to prioritize the candidate genes/proteins. Three-dimensional (3D) protein structures for the genes that satisfied the prioritized criteria were then constructed to identify the location of variants with respect to the catalytic site that might affect protein-substrate or drug interactions. Afterward, the phenotypic effect of the variants on the 3D protein structures was determined using the DUET web server and validated using molecular dynamic (MD) simulations. Molecular dynamic simulations were run to compare the variant (MT) and wild-type (WT) systems. Finally, several trajectory parameters were evaluated using

GROMACS (2020) utilities to determine any change in the dynamics of the protein structure when the variant was introduced.

From the 11 sequence variants prioritized in the multiplex family, only three were selected for further analyses, namely, the p.T871M located in the leucine-rich repeat kinase 1 (*LRRK1),* the p.V138I located in the neuropeptide S receptor 1 (*NPSR1*), and p.R663C located in the erythrocyte membrane protein band 4.1 like 2 (*EPB41L2). LRRK1* could not be pursued for further studies due to a lack of a solved homologous template structure covering the variant position at the start of this study. Pathway and expression data analysis revealed that both *NPSR1* and *EPB41L2* are involved in signaling and cellular processes that play a critical role in PD. It was further shown that *NPSR1* is involved in the G protein-coupled receptor signaling pathway, neuropeptide signaling pathway, and the up-regulation of the release of sequestered calcium ions into the cytosol. *EPB41L2*, on the other hand, plays a role in the dopamine receptor-mediated signaling pathway as well as the nicotine pharmacodynamics pathway. Homology models of both *NPSR1*, using the 6J21 template, and *EPB41L2*, using the 3F31 template, were successfully constructed and validated. As detected by the Structure Analysis and Verification Server (SAVES), the structural quality assessment indicated high reliability of the 3D protein structures of both the *NPSR1* and *EPB41L2* models. MD simulation results suggest that both p.V138I and p.R663C variants on the *NPSR1* and *EPB41L2* proteins, respectively, had a destabilizing effect on the fold of the protein structures based on the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), hydrogen bond (H-bond) analysis, and principal component analysis (PCA) values.

Based on the findings of this study, we propose that both variants (p.V138I in *NPSR1* and p.R663C in *EPB41L2*) can potentially alter protein dynamics. The findings of this study contribute significantly to our current understanding of the genetic etiology of PD and support the notion that novel variants could be responsible for PD in SSA.

CHAPTER I: Background

1.1 Introduction

Parkinson's disease (PD) is a neurodegenerative condition caused by the death of dopaminergic neurons in a part of the brain known as the substantia nigra (Aarsland et al., 2005). It is one of the most common disorders affecting the nervous system, ranking second only to Alzheimer's disease, and hence imposes significant healthcare and economic burdens on society. Over the past two decades, several causative monogenic mutations linked to PD have been discovered (Blauwendraat et al., 2020). This sparked a surge in interest in the disease's genetic etiology. Although much has been elucidated from genetic research, the exact pathobiological mechanisms of PD are not yet fully understood. However, a few probable causes such as mitochondrial dysfunction, abnormal protein aggregation, neuro-inflammation, and impaired cerebral glucose metabolism are standard processes of insulin resistance, diabetes, and neurodegeneration and have been identified as possible fundamental mechanisms for PD progression (Jankovic, 2008; Xu et al., 2018).

Conversely, in sporadic cases, PD is thought to be induced by non-genetic factors that interact with genetic risk factors (Lesage & Brice, 2012). However, more insight is needed to clarify and further understand PD pathogenesis to ultimately develop effective therapeutic strategies (Lin et al., 2016).

It has been reported by the Parkinson's Foundation that there are more than 10 million people currently living with PD globally (https://www.parkinson.org/Understanding-Parkinsons/Statistics). It is estimated that approximately 60 000 Americans are currently diagnosed with PD each year (Marras et al., 2018). By 2030, it is estimated that 1.2 million individuals in the United States will be living with PD (https://www.parkinson.org/Understanding-Parkinsons/Statistics). Regarding the incidence of PD, studies show that it is more common in the elderly, particularly in those above the

age of 65 years (Akinyemi, 2012; Dorsey et al., 2018; Karimi-Moghadam et al., 2018). However, an estimated four percent of people with PD are diagnosed before the age of 50 years, and it has also been shown that men are 1.5 times more prone to having PD than women (Georgiev et al., 2017).

Presently, much of what we know about PD is based on data from studies carried out in North America, Asia, and Europe (Khalil et al., 2020). Therefore, information on the clinical, environmental, and genetic risk factors from other geographical regions and ethnicities worldwide is needed to understand better PD's etiology and pathophysiology (Khalil et al., 2020). The field of PD genetics is understudied in sub-Saharan Africa. This is a notable omission since populations of African ancestry have the most significant genomic diversity globally. Africa has been reported to be the birthplace of anatomically modern humans and the geographic origin of human expansion globally within the last 100 000 years (Nielsen et al., 2017; Shriner et al., 2014, 2016). Therefore, studying African genomes and understanding their complexity is critical to understanding human biology and disease's differential distribution by geography and ancestry. It is speculated that due to the rich genetic heterogeneity present in the region, there are unique genotypes, environmental factors, and cultural influences that may affect the disease onset, phenotype, and progression. The present study is a step towards improving our understanding of the genetic factors underlying PD in sub-Saharan African patients. Specifically, it is focused on a South African multiplex family living with PD and employed next-generation sequencing and bioinformatic approaches to prioritize variants to identify potentially a novel PD gene and or genes.

1.2 Review of literature

1.2.1 The history of PD

Parkinson's disease, initially referred to as 'the shaking palsy' was first described in Western medical literature by James Parkinson in the early nineteenth century as a "neurological syndrome that causes involuntary tremulous motion and lessened muscular strength in certain parts of the body as well as a stooped posture and abnormal gait." (Goetz, 2011; Parkinson, 2002). Previously, researchers such as Sylvius de la Boe and Sauvages wrote about resting tremor and festination, defined as the involuntary shortening of stride and quickening of gait in the late seventeenth and eighteenth-century, respectively (Pearce, 1989). Moreover, much earlier traditional Indian and ancient Chinese texts dating as early as 1000 BC also provided clinical descriptions that suggest PD (Goetz, 2011).

The term 'Parkinson's disease' was first suggested by Jean-Martin Charcot. He rejected the 'shaking palsy's' earlier designations as he recognized that patients diagnosed with the disease did not necessarily appear weak or present with tremors. He was also the first to note two forms: the tremorous and akinetic PD forms (Goetz, 2011). Many other researchers such as William Gowers (1880), Richer and Meige (1895), Babinski (1921), Brissaud (1925), Tretiakoff (1921), Foix and Nicolesco (1925), Greenfield and Bosanquet (1953) as well as Hoehn and Yahr (1967) also contributed to the more profound and precise understanding of the PD phenotype (Goetz, 2011).

1.2.2 Clinical and neuropathological features

1.2.2.1 Clinical features

PD falls under a broader category of parkinsonism, the latter defined as a combination of bradykinesia with either rest tremor, rigidity, or both (Postuma et al., 2015). Bradykinesia is the slowness of movement or difficulty moving the body voluntarily. Parkinsonism is characterized by a large number

of motor and non-motor features that can impact function. Included under the general term parkinsonism is a group of neurological disorders that display quite similar PD symptoms. Broadly, these can be divided into two categories, namely, primary parkinsonism and secondary parkinsonism. Primary parkinsonism, also known as idiopathic PD, occurs in about 80 – 85 % of patients diagnosed with PD, with the cause of disease currently unknown.

On the other hand, secondary parkinsonism, also referred to as atypical parkinsonism, includes drug-induced parkinsonism, vascular (Arteriosclerotic) parkinsonism, and dementia with Lewy bodies (DLB). It also includes essential tremor, normal pressure hydrocephalus (NSA) corticobasal degeneration (CBD), progressive supranuclear palsy (PSP), and multiple system atrophy (MSA) (Hughes et al., 1992; Jankovic, 2008). In these conditions, the disease's cause is generally known; however, they do not respond well to dopaminergic medications such as levodopa.

There are four cardinal manifestations of PD: tremor at rest, rigidity, akinesia/bradykinesia, and postural instability that all fall under the category of motor features (Jankovic, 2008). Additionally, freezing of gait (motor blocks) is also included in PD's classic symptoms. Non-motor symptoms of PD include, but are not restricted to, mood disorders such as depression, cognitive changes such as problems with focused attention, hallucinations and delusions, orthostatic hypotension, sleep disorders such as insomnia, constipation, and early signs of satiety, pain fatigue, excessive sweating as well as sexual problems (Chaudhuri et al., 2006).

Amongst the early signs of PD is the presence of a slight shaking or tremor in an individual's finger, thumb, hand, or chin, and their handwriting getting much smaller than it usually was (micrographia). Additionally, loss of sense of smell, difficulty in moving or walking, trouble sleeping, soft or low voice, masked face, frequent dizziness or fainting, and stooped posture is experienced (Chaudhuri et al., 2006).

1.2.2.2 Neuropathological features

Neuropathological hallmarks of PD are the progressive loss of dopaminergic neurons containing neuromelanin in the substantia nigra pars compacta (SNpc) and the presence of fibrillar aggregates referred to as Lewy bodies (LB) (Braak et al., 1999; Wakabayashi et al., 2013). There are four dopaminergic pathways: the mesolimbic pathway, mesocortical pathway, tuberoinfundibular pathway, and the nigrostriatal pathway. The nigrostriatal pathway projects from the dopamine-producing neurons in the substantia nigra (SN) to the basal ganglia or striatum and mediates motor movement (Livingstone & Wonnacott, 2009; Marsden, 2006). In PD patients, reduced levels of neuromelanin and dopamine result in depigmentation of the SNpc, as shown in Figure 1.1. This leads to the emergence of motor impairment. Other brain regions implicated in progressive neuronal loss are the hypothalamus, locus coeruleus (LC), the vagus's dorsal motor nucleus, basal nucleus of Meynert, and the cranial nerve motor nuclei cerebral cortex (Petrucelli & Dickson, 2008).

While the appearance of LB's in post-mortem PD brains is indicative of a definite PD diagnosis, not all PD patients present with these fibrillar aggregates in which the α-synuclein protein (encoded by the *SNCA* gene) is the main constituent (Wakabayashi et al., 2013). α-Synuclein is a presynaptic terminal protein that plays a direct role in PD pathogenesis. Studies report several mutations in the *SNCA* gene (including A53T, A30P, E46K) associated with PD development (Oczkowska et al., 2013). LB's consist of a heterogeneous mixture of more than 90 molecules, including mitochondria-related proteins, molecules implicated in the ubiquitin-proteasome system, autophagy, and aggresomal formation (Wakabayashi et al., 2013). Although LB's are recognized as the histopathological hallmark of PD, their role in promoting neurotoxicity remains poorly understood (Wakabayashi et al., 2013). Braak and colleagues (2017) presented a pathologic staging system for PD, where early pathologic manifestations of PD are detected in the dorsal motor nucleus of the vagus and anterior olfactory nucleus, with a consecutive spread of the pathology to LC, SN, and basal forebrain (Petrucelli & Dickson, 2008). In the final stages, pathology extends to the neocortex,

particularly limbic and multimodal association cortices of the frontal and temporal lobes. Although the Braak PD staging system has been proven in autopsy studies of patients with disorders associated with LB's and controls, not all PD patients exhibit LB pathology (Burke et al., 2008).
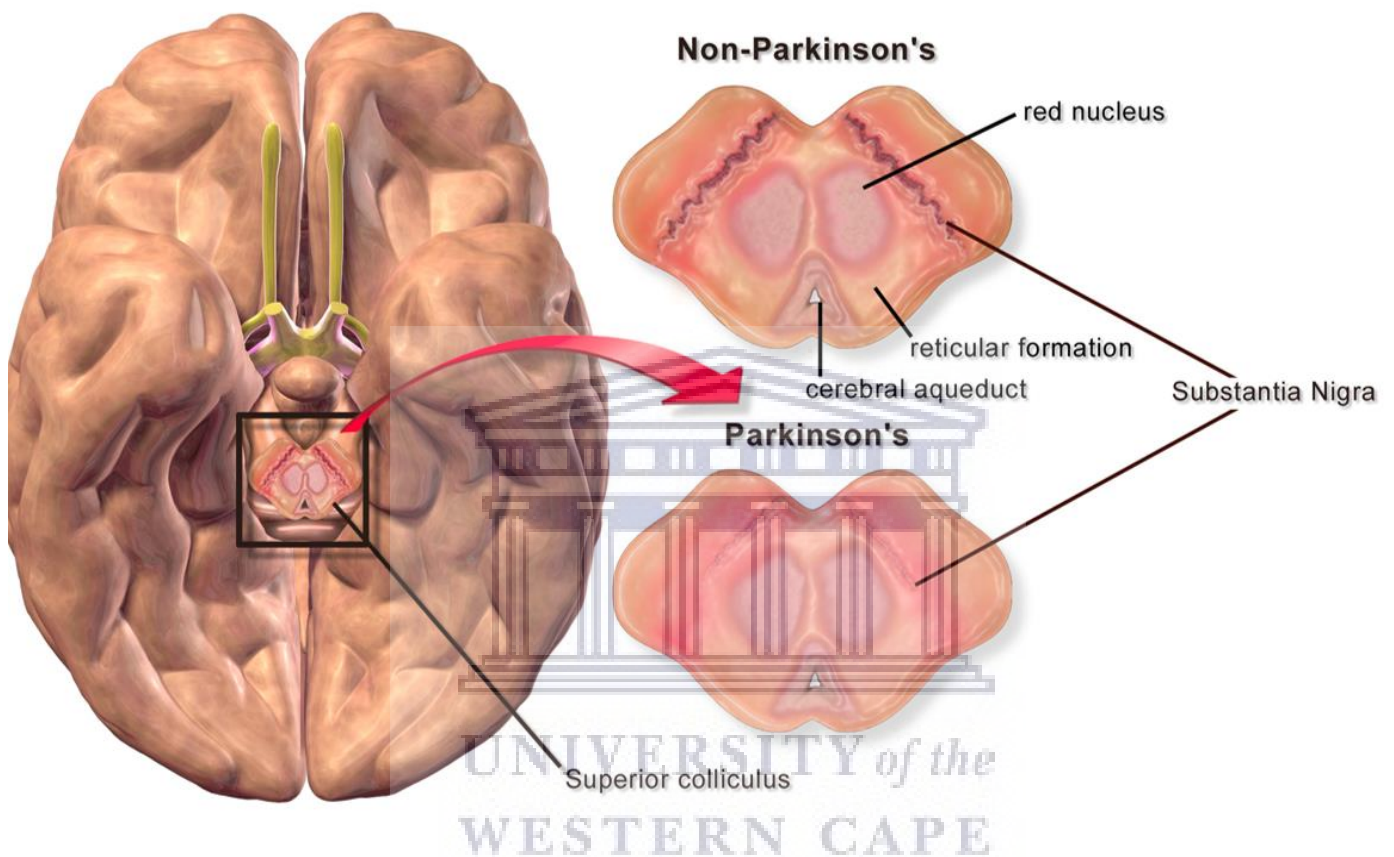
Figure 1.1: The brain region predominantly affected by PD (SNpc) and the difference in the appearance of the SN in PD patients and non-PD affected individuals. (Source: https://en.wikiversity.org/wiki/WikiJournal_of_Medicine/Medical_gallery_of_Blausen_Medical_2014#/media/File:Blausen_0704_ParkinsonsDisease.png). Accessed: 25/02/2021

1.2.3 Diagnosis

PD is diagnosed clinically by neurologists or physicians, basing it on the patient's medical history, reviewing their signs and symptoms, and performing a neurological and physical exam. This clinical diagnosis of PD is dependent on the Movement Disorder Society PD criteria (MDS-PD) (Postuma et al., 2015). In the MDS-PD criteria, the motor syndrome's centrality remains the core feature by which

clinical PD is defined. However, non-motor manifestations such as olfactory loss are present in most patients and often dominate the clinical presentation.

Several disease progression rating scales are used for the evaluation of motor impairment and disability to diagnose PD. The most frequently used rating scales are the Hoehn and Yahr and the Unified Parkinson's Disease Rating Scale (UPDRS) (Ramaker et al., 2002). Hoehn and Yahr first introduced a rating scale in the late 20th century. Clinicians used this rating scale to rate motor symptoms on a scale of 1 to 5, depending on the severity of the symptoms shown (Goetz, 2011). Another comprehensive tool developed to account for the non-motor symptoms is UPDRS, which also accounts for social interaction, moods, and mental functioning (Martinez-Martin et al., 2013). Despite the different rating scales' availability, PD diagnosis is still challenging as PD shares certain phenotypes with other Parkinsonian syndromes (Mhyre et al., 2012).

1.2.4 Incidence and prevalence of PD

Incidence is defined as the measure of new cases in a population over a given period and is typically measured as the number of people diagnosed annually. Conversely, prevalence is a measurement of individuals affected by the disease at a particular time. The incidence of PD is linked to certain risk factors such as age, industrial chemicals, and pollutants like pesticides and metals (Dorsey et al., 2018).

The Global Burden of Disease study (Dorsey et al., 2018) labels neurological disorders as the leading causes of disability and the second-highest death cause globally. Regarding the leading cause of disability-adjusted life years (DALs), neurological disorders accounted for 276 million cases (11.6 %) of the global DALYs than other diseases in 2016. Neurological disorders were also the underlying

causes of 9 million deaths (16.5 %) of the absolute global deaths recorded in 2016, second only to cardiovascular-related deaths (Dorsey et al., 2018).

PD is the second most common age-dependent neurodegenerative disease, currently affecting 1-2 percent of the population above 65 years (Johri & Beal, 2012; Karimi-Moghadam et al., 2018). The Parkinson's Prevalence Project was established to determine a more accurate representation of the number of people currently living with PD (Marras et al., 2018). It is estimated that in the year 2030, there will be about 1.2 million people in the United States only living with PD (https://www.parkinson.org/Understanding-Parkinsons/Statistics).

PD has been identified as the fastest-growing neurodegenerative disorder in prevalence, disability, and death (Dorsey et al., 2018). On a global scale, the prevalence of PD has been predicted to affect 41 per 100 000 individuals in the sixth decade of life (Karimi-Moghadam et al., 2018). In individuals aged 80 years and older, this figure is predicted to be greater than 1 900 per 100 000 (Cacabelos, 2017; Pringsheim et al., 2014). Furthermore, another study reports that PD is more predominant in males than females with a ratio of 3:1 respectively (Karimi-Moghadam et al., 2018).

A study carried out by Elbaz and colleagues (2018) reported that 6,1 million people worldwide had PD, of whom 47,5 % were women and the 52.5 % were men (Dorsey et al., 2018). When these statistics were further categorized, the study showed that 34,4 % of those individuals were from high Socio-demographic Index (SDI) countries such as North America, 50,8 % from high-middle or middle SDI countries such as North Africa, and 14,8 % from low-middle or low SDI countries such as West Africa (Dorsey et al., 2018). Compared to other regions, epidemiological studies in Sub-Saharan Africa (SSA) are either very old or have a minimal sample size (Williams et al., 2018). As seen in the data obtained from the study carried out by Elbaz and colleagues (2018), the prevalence

of PD in SSA is perceived to be lower than that reported in the more developed countries (high SDI countries).

1.2.5 Treatment

PD currently has no cure, but several available therapies and disease management strategies may be implemented to alleviate symptoms of this neurodegenerative disorder. Early PD treatments were based on empirical observations, and anticholinergic drugs were used as early as the nineteenth century. One such description can be found in the book *Ru Men Shi Qin* (Zhang et al., 2006) from traditional Chinese medicine that records a typical PD case where an "anti-tremor pill" was recommended containing a mixture of *gastrodia tuber* and scorpion. *Gastrodia tuber* contains gastrodin, which has been shown to increase dopamine levels in guinea pigs and inhibit monoamine oxidase type B in aging mice models (Zhan et al., 2016; Zhang et al., 2006). *Gastrodia tuber* has further been shown to possess potential neuroprotective effects.

In India, early descriptions of PD are recorded in the Ayurvedic treaties describing it as a category of disorders known as V*ata* and characterized by tremors (K*ampa*), stiffness (*stambha*), and depression (*vishada*) (Auddy et al., 2003; Zhang et al., 2006). Treatment recommendations for the tremors by Ayurvedic included the ingestion of the seeds of the *Mucuna pruriens* (cowhage) plant, which contains levodopa, a drug currently widely used to treat PD.

Figure 1.2: *Gastrodia tuber* **plant (A) and *Mucuna pruriens* seeds (B).** The *Gastrodia tuber* plant, part of the 'anti-tremor pill' recommended in ancient times by traditional Chinese medicine. Seeds of the *Mucuna pruriens* plant contained levodopa and were recommended in ancient Indian medicine to treat PD symptoms. (Sources: https://commons.wikimedia.org/wiki/File:Gastrodia_elata_1.JPG, https://commons.wikimedia.org/wiki/File:Mucuna_pruriens_flower.jpg). Accessed: 25/02/2021

The traditional approach to treating patients with PD is the administration of drugs to alleviate symptoms (Mhyre et al., 2012). Medications prescribed may help manage difficulties experienced with walking, movement, and tremor and may increase or substitute dopamine levels in the body. These include anticholinergic agents, Amantadine, levodopa, and synthetic dopamine agonists.

Levodopa is the cornerstone of symptomatic therapy. As dopamine does not readily cross the blood-brain barrier, levodopa, a dopamine precursor, is administered to treat PD symptoms (Koller & Rueda, 1998). It is decarboxylated to dopamine and thus, acts on all dopamine receptors (Calne, 1993). Levodopa is converted to dopamine by aromatic-L-amino-acid, raising the concentration of dopamine, which is beneficial within the striatum but deleterious outside the blood-brain barrier as such dopamine levels cause adverse reactions. Levodopa administration may be delayed as prolonged usage of this drug is associated with motor complications. These include uncontrolled involuntary

movement (dyskinesia), involuntary muscle contractions resulting in repetitive movements (dystonia), involuntary twitching and writhing affecting posture (choreoathetosis), and motor function fluctuations (Mhyre et al., 2012).

Neurosurgery, gene therapy, and stem cell transplantation are other therapies explored for motor symptoms to restore normalized motor function and optimally maintain daily living activities. Surgical procedures are advised during later stages of PD. An example is the deep brain stimulation (DBS) procedure, where surgeons implant electrodes into the thalamus, subthalamic nucleus, or globus pallidus (Volkmann, 2004). DBS has been shown to effectively control erratic and fluctuating responses to medication or control dyskinesia that do not improve with medication adjustments. Lifestyle changes such as aerobic exercises and diets high in fiber have also been recommended to assist with PD symptoms.

1.2.6 Non-genetic risk factors for PD

Apart from the primary risk factors associated with PD, such as sex, age, and genetic influences, there are other risk factors that researchers have identified that increase a person's risk of developing PD. Exposure to pesticides, rural residence, certain occupations such as farming, welding, carpentry, and workers chronically exposed to metals have been associated with increased risk to PD (Chade et al., 2006; Gorell et al., 2004). Other risk factors include exposure to toxins such as herbicides (paraquat), fungicides (maneb), insecticides, and pesticides (Lee & Gilbert, 2016). Scientists have also suggested prolonged exposure to metals such as copper, manganese, mercury, lead, and solvents such as trichloroethylene and polychlorinated biphenyls increase the risk of PD (De Miranda & Greenamyre, 2020).

1.2.7 Genetic causal and risk factors for PD

Research on the genetic causes of PD has led to discoveries of several PD-causing or associated genes over the past two decades (Bandres-Ciga et al., 2020). PD may either appear as sporadic or familial. Sporadic PD is when an individual presents with PD while there is no known history of other family members with the disease. On the other hand, familial PD refers to more than one disease occurrence in a particular family.

PD researchers have identified 23 genes linked to rare monogenic familial forms of PD with Mendelian inheritance that can be divided into either X-linked, autosomal recessive, or autosomal dominant (Bandres-Ciga et al., 2020). Autosomal recessive genes include *Parkin (PRKN)*, *PINK1*, *DJ-1*, *ATP13A2*, *PLA2G6*, *FBXO7*, *DNAJC6*, *SYNJ1*, *SLC6A3,* and *VPS13C*. *RAB39B* has been associated with an X-linked disorder displaying parkinsonism features. Examples of the autosomal dominant genes include *GBA*, *SNCA*, *LRRK2*, *GIGYF2*, *HTRA2*, *VPS35*, *EIF4G1*, *DNAJC13*, *CHCHD2*, *TMEM230*, *RIC3,* and *GCH1* (Karimi-Moghadam et al., 2018). Table S1 in the appendix shows the 23 genes linked to rare monogenic familial forms of PD. It highlights the disease features, locus and position, the pathways they are involved in, and the disease onset.

The clinical characteristics shared by patients who have mutations in the autosomal dominant PD genes are that of typical PD, presenting an excellent response to levodopa treatment during the early stages of disease progression and Lewy body formation (Dekker et al., 2003; Kouli et al., 2018). It is also seen that most of these genes result in late-onset PD (disease age at onset of > 50 years) (Hicks et al., 2002). Contrastingly, the attributes shared by the autosomal recessive PD genes are atypical parkinsonism features (Gasser, 2009; Karimi-Moghadam et al., 2018). However, patients with mutations in the *PRKN, PINK1,* and *DJ-1* genes display typical PD symptoms (Gasser, 2009). Juvenile disease onset is associated with the autosomal recessive PD genes, good response to

levodopa during the early stages of PD, and Lewy bodies' absence. Dementia has also been linked to autosomal recessive PD genes (Karimi-Moghadam et al., 2018).

Studies on the genetic causes of PD have been carried out over the past two decades and have shown that the genetic causes may vary in specific populations. Various PD-causing or PD-associated genes have been recognized, which have aided in drug discovery and the establishment of personalized treatment. This highlights the need to identify new PD-causing or PD-associated genes by utilizing new sequencing technologies and bioinformatics approaches.

1.2.8 Bioinformatic approaches to identify novel disease-causing genes

1.2.8.1 Overview

Bioinformatics is an interdisciplinary field that develops software tools and methods to interpret and understand biological data. It combines computer science, statistics, mathematics as well as engineering to analyze and interpret data. Analyses such as DNA sequencing result in exponential amounts of biological data, making it hard to analyze manually. With the advent of computer science, various computational techniques and algorithms are used to analyze large biological datasets more accurately and efficiently using automated processes.

As one of the most influential biomedical research tools, DNA sequencing has improved its productivity exponentially. It has evolved into a new layout of technological territories towards engineering and physical disciplines over the past four decades (Shuikan et al., 2019). DNA sequencing technology has played an essential role in the advancement of molecular biology ever since its invention. From the early manual sequencing operation that Frederick Sanger developed to the first-generation automated sequencer driven by Sanger chemistry, and right up to the most recent next-generation sequencing (NGS), a tremendous change in the field has been observed (Chow-White

et al., 2012). The sequencing speed has improved exponentially over the last few decades, fundamentally changing how we can examine the blueprint of all life and help us develop other branches of genomic studies such as comparative genomics, proteomics, and metabolomics (Shuikan et al., 2019). In the current study, bioinformatics approaches were used to query gene ontology information in several databases and protein-protein interaction analysis, to prioritize putative genes involved in PD. Additional structural methods employed in this study include molecular modeling, change in protein stability predictions, and molecular dynamic simulations to validate further the sequence variants' effect on the protein structure and behavior.

1.2.8.2 Gene Prioritization techniques

Genes contain the information required to produce proteins for daily metabolic activities to take place. A series of nucleotide bases in a gene may alter during mitosis or due to environmental exposure. The result is the acquisition of specific mutations in genes that alter protein function. These genetic mutations may result in either failure of protein production or the formation of faulty proteins with altered activity (Raj & Sreeja, 2018). As a result, genetic disorders may arise from the improper protein production of cells, which may be classified as Mendelian disorders (mutation of a single gene) or complex disorders that arise from mutations in several genes (Raj & Sreeja, 2018).

High-throughput methods such as genome-wide association studies (GWAS) and RNA interference screens have revolutionized how we can couple genetic variations to diseases. These techniques have become very efficient in connecting large chromosomal regions with genetic disorders, generating lists of hundreds of potential gene candidates (Guala et al., 2017). Techniques for prioritizing candidate disease genes are vital in maximizing the information from these high-throughput experiments to produce a list of potential genes that can be validated with experimental methods (Guala et al., 2017).

Gene prioritization is defined as the process of filtering potential candidate genes based on their association with a particular disease. Computational methods such as biological knowledge sources are often used to rank the genes to identify which of the prioritized genes is most likely the causative gene. These will then be studied further by performing a series of *in-silico* and *in-vitro* experiments. Biological knowledge sources can be classified according to the nature of data collection and the type of data (Raj & Sreeja, 2018). This can be further divided into primary, secondary, and composite databases. For this study, composite databases are utilized, which is a collaboration of multiple databases such as Online Mendelian Inheritance in Man (OMIM), National Centre for Biotechnology Information (NCBI), gene ontology, Kyoto encyclopedia of genes and genomes (KEGG), panther pathway analysis, Reactome, and the Allen brain atlas.

1.2.8.3 Molecular dynamic simulations to evaluate potential pathogenicity of variants

Biological macromolecules are large molecules built from smaller molecules divided into four classes: proteins, carbohydrates, lipids, and nucleic acids. Biological macromolecules are essentially dynamic systems, as insights into their internal motions can be demonstrated using calculations based on physical models (Karplus & McCammon, 2002). Dynamic simulations encompass molecular dynamic movements, refinement of structures based on nuclear magnetic resonance (NMR) and X-ray data, and the evaluation of free energy changes induced by mutations. Molecular dynamics (MD) simulations are vital in understanding these motions.

MD is a computer simulation method used to analyze the physical movements of molecules and atoms. MD simulations play a role in predicting molecular systems' properties, structure, and function (Geng et al., 2019). They enable researchers to calculate atomic movements in a molecular system and investigate the proteins' structural dynamics on timescales of nanoseconds up to microseconds (Ode et al., 2012). This is important for understanding the motion of molecular systems.

MD simulation studies have provided an increased understanding of protein structure-function relationships (Nair & Miners, 2014). The ability of MD simulations to provide individual particle movements as a function of time and measure the change in Gibbs free energy are a few characteristics that have played an essential role in the broad applicability of MD simulations (Karplus & McCammon, 2002). The Gibbs free energy is given by $G = H - TS$, where H is the enthalpy, T is temperature, and S is the entropy. Enthalpy is mathematically described as $H = U + pV$, where U is the internal energy, p is the pressure, and V is the volume. Additionally, free energy is a quantitative measure that describes the relationship between denatured and native molecules whose value depends on other factors such as the variation of enthalpy (the amount of potential energy: bonded and non-bonded interactions) and the variation of entropy (different protein conformations). This, in turn, describes deeper aspects of the protein such as bond angles, bond lengths, and hydrogen bonds. Furthermore, this describes energetic changes such as thermal fluctuations and rearrangements of chemical bonds. Simulating systems over nanosecond timescales can identify motions such as atomic fluctuations, conformational changes in amino acid side chains, and loop motions (Nair & Miners, 2014).
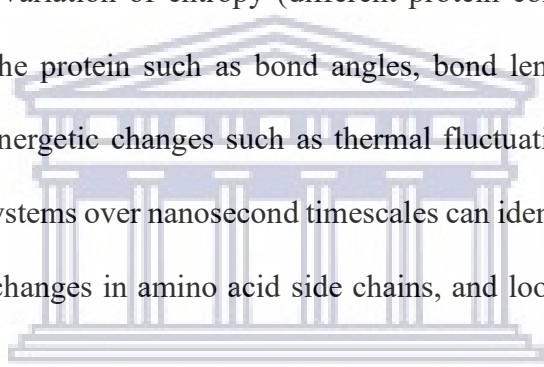
MD is a well-established method that provides insights into the workings of biomolecular systems at spatial (locations) and temporal (time) scales that are difficult to access experimentally (Klepeis et al., 2009). Experimentally, not all prioritized genes can be tested by wet-lab experiments due to complications that arise, such as protein aggregation, factors such as protein solubility, as well as the costs incurred during wet-lab experiments. Utilizing MD simulations is advantageous as it mimics what atoms do in an organism, assuming a given potential energy function. MD's application is made to study structure-function relationships, disease pathways, and drug design (Lemkul, 2019). MD can be used to explore conformational space, behavior, and movement of the residues in the protein structure as well as the folding and unfolding of the protein and is often the method of choice for biological macromolecules.

MD simulations provide powerful tools for exploring the conformational energy landscape accessible to molecules to determine the thermodynamic properties of the biological macromolecules (Karplus & McCammon, 2002). Recent advances in computer hardware and software have rapidly improved MD's performance and precision. Simulations can now be run on central processing units (CPUs) and graphics processing units (GPUs), providing significant internal parallelism with the potential for high combinational throughput at a relatively low cost (Klepeis et al., 2009). Combining increased computer power and improved potential energy functions for biological macromolecules has resulted in the researchers' ability to generate simulations that approach native experimental behavior. Furthermore, an increase in readily available computer power and the standardization of simulation protocols is now feasible for non-specialists to reproduce simulation analyses published by others.

1.2.8.4 Limitations of MD

Although MD simulations' capability has significantly increased due to improved high-performance computing systems, there are still limitations concerning the investigation of many critical biological systems (Kim et al., 2017). For instance, magnesium (Mg^{2+}) ions play an essential role in biological processes such as DNA repair mechanisms and the self-assembly of RNA nanoparticles. However, Mg^{2+} ion dynamics only occur on a microsecond timescale, which is currently beyond regular MD simulation timescales. Coarse-grained approaches can overcome this limitation on large-scale systems such as membranes, viruses, and chromatin fibers. In a coarse-grained approach, each nucleotide in an RNA molecule can be represented by a few beads so that the total number of atoms in the RNA nanostructure is markedly reduced (Kim et al., 2017; Paliy et al., 2010). Moreover, multiscale computational approaches using both quantum mechanical and molecular mechanical (QM/MM) methods, all-atom MD simulations, and coarse-grained techniques can provide vital information for large biomolecular systems.

Another principal challenge MD simulations face is that force fields require further refinement (Durrant & McCammon, 2011). Force fields are a collection of equations that describe the potential energy of a system. Here the potential energy is derived from bonded and non-bonded interactions observed between atoms. Force fields are inherently approximate as they are derived from classical mechanics and not quantum mechanics. However, there have been fast improvements made to match experimental data. While simulations can accurately predict critical molecular motions, simulations are poorly suited to systems where quantum mechanics are essential; for instance, when transitional metal atoms are involved in binding. Some researchers have introduced quantum mechanical calculations into classic MD force fields to overcome this challenge (Durrant & McCammon, 2011).

Another limitation faced by force fields is partly due to polarizability. The atoms' electrostatic properties in the classic MD simulations are represented by the atomic nuclei's partial charges (Kim et al., 2017). The magnitudes and positions of the partial charges are treated as constants throughout the MD simulations. However, the magnitudes and distributions of charges in a biomolecule continuously change in response to the environment. The constant partial charge approximation is insufficient to describe the polarizability of the biomolecules. To bridge the gap between the constant partial charge and polarizability, various theoretical models have been applied to major force field groups such as the Amber and CHARMM force field groups (Kim et al., 2017).

1.2.9 Computational studies in PD research

Computational methods involving comparative protein structure modeling, molecular docking, and MD simulations have provided fundamental support in PD research to unravel details of PD biology (Coskuner-Weber & Uversky, 2018). Computational studies complement wet-lab experiments and provide structural information at the atomic level, with MD simulations contributing insights into the structure-function relationships (Delemotte, 2020; Hollingsworth & Dror, 2018; Nair & Miners, 2014). For instance, the α-Synuclein protein with a genetic missense mutation (A53T) was initially

identified in Italian and Greek families with familial PD (Coskuner & Wise-Scira, 2013). With the aid of Replica Exchange Molecular Dynamic (REMD) simulations (40 ns), researchers established a direct relationship between the rapid conformational changes and free energy landscapes upon introducing the A53T mutation within the α-synuclein protein structure.

Conformationally, a decrease in the number of β-sheets formed in the non-amyloid-β component (NAC) and C-terminal region was seen upon introducing the A53T mutation. Conversely, an increase in β-sheet abundance around the mutation site was seen. Additionally, long-range interactions were lost between the NAC region and the N- and C-terminal regions due to the A53T mutation, indicating that the NAC region was more solvent-exposed, contributing to an increase in the aggregation rate of the protein. α-Synuclein aggregates then formed insoluble fibrils characterized by Lewy bodies. This study extensively explored the contribution of MD simulations in representing the molecular structure of α-synuclein and its contribution to the formation of Lewy bodies, enhancing our fundamental knowledge of PD (Coskuner & Wise-Scira, 2013; Coskuner-Weber & Uversky, 2018). Its results also agree with several experimental measurements previously conducted (Coskuner & Wise-Scira, 2013). However, it also highlights the lack of experts in both drug discovery and MD simulations, posing a significant limitation in this field of study.

Another study employed MD simulations to provide a mechanistic view underlying the WD40 dimerization and unveiled the structural basis by which interface-based mutations compromise the complex binding network along with the dimerization interface (Li et al., 2020). Other studies utilized MD simulations to understand the role of *LRRK2*, *DJ1*, and *VPS35* mutations in PD (Anderson & Daggett, 2008; Y.-B. Guo et al., 2016; Zimprich et al., 2011).

From these studies, the strengths of utilizing MD simulations are highlighted to understand the mechanical impact of mutations on protein structure and, in doing so, provide helpful insight into the

mechanism of PD development. Hence, in the present study, we apply a similar approach to prioritize the effect of variants on our target protein structure to provide support for the variant's role in PD and propose it for further experimental investigation.

1.2.10 Rationale of the present study

In sub-Saharan Africa (SSA), PD research has received little attention (Dekker et al., 2020). Consequently, a lack of knowledge of the condition leads to most people living with PD in SSA being misdiagnosed and untreated resulting in high mortality rates (Mokaya et al., 2017). The estimated number of PD patients in the SSA region is relatively low, which may be due to the widespread belief that neurological disorders are part of natural aging. Therefore, patients remain undiagnosed (Williams et al., 2018). Other factors contributing to the low numbers include the absence of epidemiological data, a scarcity of neurologists, and underfunded healthcare infrastructure (Bower & Zenebe, 2005; Lekoubou et al., 2014). Though chronic non-communicable diseases like PD are considered urgent health priorities in high-income countries, in SSA, the impact and burden of PD are unfortunately neglected and go unrecognized. Therefore, despite the imminent epidemic of PD, countries in SSA are ill-equipped to deal with this problem (Dotchin & Walker, 2012).

It has been shown that the known PD mutations play a minor role in disease etiology in SSA populations (Bardien et al., 2010; Blanckenberg et al., 2013; Wang et al., 2015). This illustrates a knowledge gap in SSA PD genetics, emphasizing the importance of utilizing new technologies to discover novel PD-causing genes or novel mutations in known PD genes. Ultimately, this insight might contribute significantly to our current understanding of the genetic etiology of PD, both locally and internationally.

In the current study, we follow up on previous work done in an MMed study by Dr. Roopnarain (unpublished data) investigating a South African family with several family members (mother and

three sons) suffering from PD. The parents are both of Indian descent. In the MMed study, whole-exome sequencing (WES) had been performed on three affected members of this family, and 11 sequence variants had been prioritized for further research. We used the data generated by the MMed study as a basis for the current study, and the study's aims and objectives are outlined below.

1.2.11 Aim and objectives:

The aim of this study was to use *in-silico* methods to prioritize a causal variant from WES data on a South African Indian family with PD. The project is divided into six specific objectives:

1   To annotate the 11 sequence variants using the Universal Mutation Databases (UMD) predictor v.2016.

2   To perform pathway, expression and STRING analysis on the 11 genes to determine brain-specific gene function, mRNA expression profiles and protein-protein interactions with known PD genes.

3   To construct accurate three-dimensional protein structures for the prioritized genes most likely to be associated with PD.

4   To calculate destabilizing effects of the variants on the respective protein structure upon introducing the variant using the DUET web server.

5   To simulate the effects of only the variants that destabilize the respective protein structure compared to the wild-type (WT) structure.

6   To analyze the simulation trajectory data and, in doing so, validate the effects of variants on the protein structure.

CHAPTER II: Methodology

2.1 Ethical considerations:

Ethical approval was obtained from the Health Research Ethics Committee of Stellenbosch University, Cape Town, South Africa (Protocol number: 2002C/059), with annual renewals.

2.2 Previous work:

2.2.1 Filtering and variant prioritization:

A previous MMed study (conducted by Dr. K Roopnarain) investigated a South African Indian family with PD, denoted as family ZA 398. DNA was collected from the three affected individuals and eight unaffected family members. WES was then carried out on the three affected family members (siblings) and yielded 166 variants shared by all the affected family members. Due to this large number of variants, they were further filtered down to 11 gene variants using strict criteria by Dr. Roopnarain. This criterion involved using a bioinformatics workflow (Appendix Figure S3) to find uncommon variants shared by the three affected siblings. Following this, variants were further prioritized using a population database (EXaC database) and *in-silico* pathogenicity prediction tools (Polyphen, SIFT, MutationTaster, Mutation Assessor, and CADD).

Additionally, the criteria included selecting only heterozygous variants shared between the three patients, filtering out all synonymous variants and only including non-synonymous variants. The Gtex portal was used to determine which of the variant genes were expressed in brain tissue. Any known PD genes reported in the AnnEx database (a database containing thousands of PD exomes) were excluded. Subsequently, common variants (Minor Allele Frequency ≥ 0.001) recorded in the Single Nucleotide Polymorphism database (dbSNP) and 1000 genome project were excluded. Other online

platforms utilized include the disease-related variant databases such as ClinVar, the Human Gene Mutation Data (HGMD), and the Online Mendelian Inheritance in Man (OMIM) database. This yielded a list of eleven candidate genes with novel variants pursued further in the current study (Table 2.1). Moreover, none of the currently known PD mutations was found in any of the sequenced family members. Additionally, Sanger sequencing was performed on the unaffected family members to determine the presence or absence of variants found in the affected family members.

Table 2.1: List of 11 non-synonymous variants in candidate genes (in chromosomal order).

Gene symbol	Gene name	Variant present or absent in unaffected members	Number of unaffected members
ELK4 (p.R48C)	ETS transcription factor 4	Present	4
CEP170 (p.T377M)	Centrosomal protein 170	Present	3
EEFSEC (p.V184M)	Eukaryotic elongation factor	Present	4
TRIM59 (p.K114R)	Tripartite Motif containing 59	Absent	0
HTT (p.R2063H)	Huntingtin	Absent	0
ELMOD2 (p.A215S)	ELMO domain containing 2	Absent	0
EPB41L2 (p.R663C)	Erythrocyte membrane protein band 4.1 like 2	Present	1
NUP153 (p.T1284I)	Nucleoporin 153	Present	2
NPSR1 (p.V138I)	Neuropeptide S receptor 1	Present	5
LRRK1 (p.T871M)	Leucine-rich repeat kinase 1	Present	6
MAN2B1 (p.V825A)	Mannosidase alpha class 2B member 1	Present	4
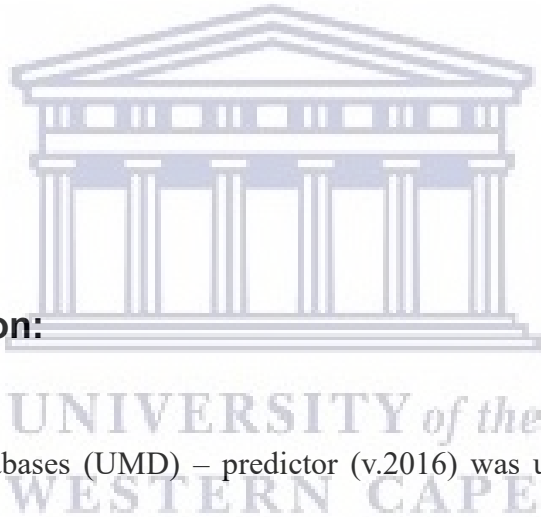
Abbreviations: *ELK4*= ETS transcription factor, *CEP170*= Centrosomal protein 170-KD, *EEFSEC*= Eukaryotic elongation factor selenocysteine-tRNA-specific, *TRIM59*= Tripartite motif-containing protein 59, *HTT*= Huntingtin, *ELMOD2*= ELMO domain containing 2, *EPB41L2*= Erythrocyte membrane protein 4.1 like 2, *NUP153*= Nucleoporin 153-KD, *NPSR1*= Neuropeptide S receptor 1, *LRRK1*= Leucine-rich repeat kinase 1, *MAN2B1*= Mannosidase alpha class 2B member 1, *SNCA*= Synuclein alpha, *PARK7*= Parkinson disease protein 7, ETS= E26 transformation-specific and ELMO= Engulfment and motility.

The eleven candidate genes containing the novel variants were *ELK4* (R48C), *CEP170* (T377M), *EEFSEC* (V184M), *TRIM59* (K114R), *HTT* (R2063H), *ELMOD2* (A215S), *EPB41L2* (R663C), *NUP153* (T1242I), *NPSR1* (V138I), *LRRK1* (T871M) and *MAN2B1* (V825A). Most of the 11 prioritized variants found in the affected family members were also found in the eight unaffected

family members who were Sanger sequenced. Most frequent variants included: genes *ELK4* (four unaffected individuals), *CEP170* (three unaffected individuals), *NPSR1* (five unaffected individuals), *LRRK1* (six unaffected individuals), *MAN2B1* (four unaffected individuals), and *EEFSEC* (four unaffected individuals) (Table 2.1). However, the variants in genes *TRIM59*, *HTT*, and *ELMOD2* were not found in unaffected family members (Table 2.1). What is apparent from this is that all the unaffected family members are much younger than the age of onset observed for their parents. Therefore, the affection status of these individuals is currently unknown. However, now that their genotyping status for these 11 candidate variants is known, we will continue to monitor them for signs of PD. The 11 gene variants were further prioritized in this study using gene ontology, protein-protein interaction analysis, and pathway analysis to determine their possible role in PD development.

2.3 Current work:

2.3.1 Variant annotation:

The Universal Mutation Databases (UMD) – predictor (v.2016) was used for confirming variant positions at the start of this study. It is an online-based computational combinatory system used for next-generation sequencing (NGS) analysis that efficiently annotates complementary DNA substitutions of human transcripts for their potential pathogenicity (Frédéric et al., 2009; Salgado et al., 2016). Variant call format (VCF) files of the affected family members were uploaded to the UMD-predictor for variant annotation. The UMD-predictor investigates the substitutions' biochemical properties and their impact on splicing signals, the localization in protein domains, conservation through the BLOSUM62 global substitution matrix, and protein-specific conservation among 100 species (Salgado et al., 2016). The UMD-predictor tool has been shown to outperform the seven most cited prediction tools: SIFT version 5.1.1 (Sim et al., 2012), Polyphen version 2.2.2, Provean 1.1.3 (Choi et al., 2012), Mutation Assessor 2 (Reva et al., 2011), Condel, CADD (Kircher et al., 2014) and

MutationTaster 2 (Schwarz et al., 2014). This system showed a higher specificity of 95 % compared to SIFT having 79 %, Polyphen having 72 %, Provean having 73 %, Mutation Assessor having 75 %, Condel having 78 %, CADD having 76 %, and MutationTaster achieving 69 %. This trend was consistent in other parameters measured, such as the diagnostic odds ratio (DOR), accuracy, speed, and Matthew's correlation coefficient (MCC) (Salgado et al., 2016). The MCC is a model evaluation tool that measures the differences between actual values and predicted values, while the DOR is a statistic in epidemiology expressing the strength of association between exposure and disease; DOR can be applied to describe the strength of the association between test results and disease (Glas et al., 2003). The UMD-predictor also provided the shortest list of candidate mutations for WES data (Salgado et al., 2016). For these reasons, we utilized this prediction tool for our study.

2.3.2 Prioritizing the 11 candidate genes as potential PD-associated genes:

2.3.2.1. Functional gene annotation using data mining:

Data mining was performing using previously published data to determine each of the 11 genes' functions to investigate whether they have been reported previously as being involved in any PD-linked biological pathways associated with PD as well as other neurological disorders.

2.3.2.2 Pathway and expression analysis:

2.3.2.2.1 Allen Brain Atlas:

The Allen Brain Atlas (v.2016) (Jones et al., 2009) is an online resource available for researchers to query gene expression data in brain tissue under different diseased conditions. For this study, we used the microarray data resource. We selected 'Microarray gene search' under the human brain tool and searched each gene by its gene symbol as recorded on the UniProt resource (The UniProt Consortium, 2015). A minimum Z-score of 3, which indicated increased gene expression levels in the brain region

of interest, was set as the cut-off score when selecting the specific brain regions to confirm if the gene was expressed in that region. The purpose of searching each gene individually against the Allen Brain Atlas was to determine whether they were highly expressed in the basal ganglia, hypothalamus, and the substantia nigra as these regions have been associated with PD in previous studies. This also aided in narrowing down the number of genes to a manageable list for further studies.

2.3.2.2.2 KEGG pathway and PANTHER pathway analysis:

The KEGG (v.2020) and PANTHER (v.15.0) pathway databases were utilized to identify the specific pathways for each of the eleven genes (Du et al., 2014; Thomas et al., 2003). This was performed by searching the databases using each gene symbol as recorded on the UniProt database. The output was recorded as a diagram containing the biological pathways they were involved in. The purpose was to determine if any of the eleven genes interacted with any known PD genes and if they were involved in any PD-linked or dopamine-related pathways.

2.3.2.2.3 STRING Analysis:

The search tool for the retrieval of interacting genes (STRING) analysis (v.11.0b) is a database predicting protein-protein interactions. STRING utilizes experimental data and information from curated databases to determine protein-protein interactions based on interaction predictions. These interaction predictions are derived from systematic co-expression analyses, detection of shared selective signals across genomes, automated text-mining of the scientific literature, and computational transfer of interaction knowledge between organisms based on gene orthology (Szklarczyk et al., 2017). STRING analysis was performed by searching the database using each gene symbol as recorded on the UniProt database. For this analysis, all the different types of protein-protein interaction data were used. The output was recorded as a diagram containing the protein-protein interactions each gene was involved in. The purpose was to determine if any of the eleven genes interacted with any known PD genes.
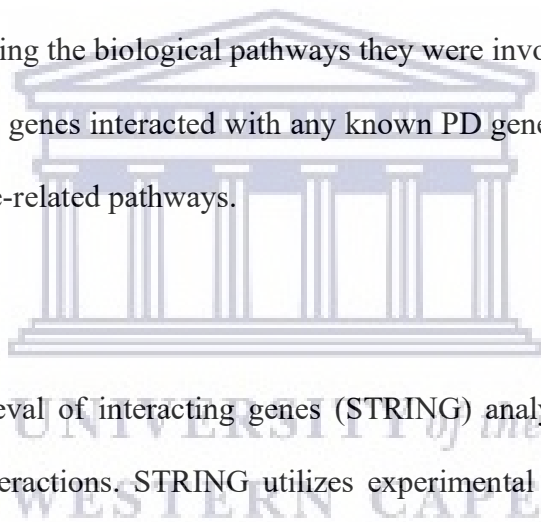
2.3.2.2.4 Human Protein Atlas:

The Human Protein Atlas (HPA) (v.20.1) is a web-based bioinformatics tool that maps the gene of interest against all tissues and organs present in the human body to determine the level of expression of the specified protein (Pontén et al., 2008). It is a publicly available tool and contains millions of high-resolution images that show the spatial distribution of proteins in healthy human tissue and various cancer types.

The HPA consists of six sub-atlases: the tissue atlas, the cell atlas, the pathology atlas, the brain atlas, the blood atlas, and the metabolic atlas. For this study, the tissue atlas tool was utilized to determine the gene expression level found in the various tissues. This specified which tissues the gene was highly expressed in and was done by searching each gene by its gene symbol as recorded on the UniProt database. The output was shown as graphs illustrating the expression level and was recorded. Using pathway (KEGG and PANTHER), STRING and expression analysis (HPA) allowed us to prioritize three novel genes as putative PD-associated genes, *LRRK1, NPSR1* and *EPB41L2*. Due to the unavailability of a template structure covering the variant position (p.T871M) of *LRRK1,* only genes *NPSR1* and *EPB41L2,* were further explored in structural studies to understand the influence of the variant on the protein structure and behavior. The structural methods are introduced in the following sections.

2.3.3 Protein structure homology modeling:

Homology modeling techniques were used to predict the target proteins' three-dimensional (3D) structures as there were no experimental structures available for the two proteins. For this study, we used the SWISS-MODEL (v.2020) web server (Guex et al., 2009), an online tool to perform homology modeling. There are several steps required to build a 3D protein model of a target protein: i) the identification of the amino acid (AA) sequence of the target protein, ii) selection of a

homologous structural template, iii) performing an accurate alignment of the target sequence and template structure, iv) 3D model-building and v) 3D model quality evaluation.

2.3.3.1 Template search and target-template alignment:

The WT canonical AA sequences for genes *NPSR1* and *EPB41L2* were retrieved from UniProt with accession numbers O43491 and Q6W5P4, respectively. The target protein sequences were aligned to a library of protein sequences extracted from the Protein Data Bank (PDB) (Cheng, 2008) to identify homologous protein sequences. SWISS-MODEL implemented the Basic Local Alignment Search Tool (BLAST) (Camacho et al., 2009) and HHblits (Steinegger et al., 2019) algorithms to search for related template structures. After a list of templates was generated, the homologous template structure with the highest sequence identity and sequence coverage to the target sequence that encompassed the variant site was selected for model construction.

2.3.3.2 3D Protein model construction:

Three-dimensional protein models for each target sequence were built using the target-template alignment file and the template PDB structure. The PDB file of the modeled protein 3D structure was downloaded from the SWISS-MODEL website and visualized in PyMol (Yuan et al., 2017). The predicted protein model was structurally aligned to the template structure to assess the structural fold. The template used for the *NPSR1* protein had two ligands bound: the 5-[[(2~{R},3~{S})-2-[(1~{R})-1-[3,5-bi(trifluoromethyl)phenyl] ethoxy]-3-(4-fluorophenyl) morpholin-4-yl] methyl]-1,2-dihydro-1,2,4-triazol-3-one (GBQ) and the (2R)-2,3-dihydroxypropyl (9Z)-octadec-9-enoate (OLC). The PyMol align and extract tools were simultaneously utilized to create the *NPSR1*-GBQ-OLC complex and the *NPSR1*-GBQ complex. The *NPSR1*-GBQ-OLC complex was needed to perform stability prediction analyses as well as calculate polar interactions, while the *NPSR1*-GBQ complex was used to perform MD simulations. This will be explained in detail in the next chapter. Additionally, the PyMol mutagenesis wizard tool was used to mutate the wild-type (WT) AA residues p.V138I on the

NPSR1 protein structure and the p.R663C on the *EPB41L2* protein system. The resulting variant structures were saved in PDB format.

2.3.3.3 3D Protein model quality evaluation:

To evaluate the accuracy of the predicted protein structures, both internal and external tools were used. Firstly, quality scores were calculated with SWISS-MODEL inbuilt algorithms, including the Global Model Quality Estimation (GMQE) and the Qualitative Model Energy Analysis (QMEAN). GMQE is a quality estimation that combines properties from the target-template alignment and the template structure. The final GMQE score is a numerical value between 0 and 1, representing the expected accuracy of the generated model. A score closer to 1 indicates higher reliability, and a score closer to 0 reflects a lower accuracy of the model (Biasini et al., 2014). QMEAN is a composite scoring function assessing the significant geometrical aspects of protein structures, such as their fold, hydrogen bonding, chirality, and solvation potential, providing both global (entire protein structure) and local (per residue) total quality estimates based on the protein model structure (Higueruelo et al., 2012; Persson et al., 2018; Studer et al., 2020). The QMEAN indicates the similarity of the generated protein model's quality score relative to experimental structures of a similar size (Benkert et al., 2011). A quality score close to 0 indicates a high-quality protein structure. In contrast, any score that falls within the range of -1 to -4 indicates low-quality models compared to high-resolution experimental structures of a similar size (Benkert et al., 2011; Studer et al., 2020).

For external evaluation, the Structure Analysis and Verification Server (SAVES) v6.0 web server was utilized (https://saves.mbi.ucla.edu/). SAVES is an online web-based tool that accepts files in PDB format and has five different quality parameter checks, namely, Verify 3D (Bowie et al., 1991; Lüthy et al., 1992), ERRAT (Colovos & Yeates, 1993), Prove (Pontius et al., 1996), PROCHECK (Laskowski et al., 1996; Morris et al., 1992) and WHATCHECK (Engh & Huber, 1991). The

predicted WT 3D models were uploaded to the SAVES web server to evaluate the overall quality of the models.

Verify 3D, the first of the five quality parameter checks, determines the compatibility of an atomic model (3D) with its amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar) and comparing the results to high-resolution structures (Eisenberg et al., 1997). A model should have at least 80 % of the residues in the sequence with a score greater than or equal to 0.2 in the 3D-1D profile for the protein model to be considered reliable. Any score below that indicates problematic regions within the model.

ERRAT is a quality evaluation tool that analyses the statistics of non-bonded interactions between different atom types and plots on a graph the value of the error function versus the position of a 9-residue sliding window (Colovos & Yeates, 1993). This is calculated by comparison with statistics from reliable high-resolution structures. Structures are considered of high accuracy when they have values of 50 % or above (Colovos & Yeates, 1993; Tran et al., 2015).

PROCHECK assesses the quality of a given protein structure by comparing its stereochemical parameters, phi and psi dihedral angle distributions of amino acid side chains, with those in high-resolution protein structures (Pontius et al., 1996). For a protein model to be considered reliable, at least 90 % of the residues must be in the most favorable regions of the Ramachandran plot and have the best combinations of phi and psi dihedral angle distributions (Morris et al., 1992). A score below that indicates problematic side chain orientations and loop regions. In summary, a protein model needs to satisfy all or some of these criteria to be deemed reliable. For this study, we utilized the Verify 3D, ERRAT, and PROCHECK quality evaluation tools.

2.3.4 Stability prediction and polar interaction calculations:

The DUET web server (v.14) was used to calculate the destabilizing or stabilizing effects of the variant on the respective protein structure (Pandurangan et al., 2017; Pires & Ascher, 2016). DUET implements both the Site-Directed Mutator (SDM) and the mutation Cut-off Scanning Matrix (mCSM) algorithms to calculate a combined score using the two software. Here, the WT structure of the respective protein was used to predict the stabilizing or destabilizing effect of the variant on the protein structure. The analysis was performed as a regression task consisting of predicting the numerical value and direction of change of the difference in Gibbs free energy (Pires et al., 2014). In brief, the web server took as input the WT homology model structure in PDB format. Next, information on the type of variant, its position, and the chain location was provided to the web server, and the job was submitted. The predicted change in Gibbs free energy score upon introducing the variant ($\Delta\Delta G$ in Kcal/mol) was calculated. DUET calculates either a negative score, indicating a destabilizing effect of the variant on the structure, or a positive score indicating a stabilizing effect of the variant on the protein structure.

PyMol was used to calculate changes in the number of polar contacts between neighboring residues of the mutated residue and the WT residue using the 'find polar contacts' option in PyMol. This was done to determine if the loss or gain of interactions occurred in the immediate amino acid environment of the protein residue, thereby possibly affecting the protein's fold. An increase or decrease in the number of polar interactions may further result in a conformational change of the protein (Chen et al., 2019). Furthermore, MD simulations were also performed to understand the effect of the variants on the dynamics of the protein structures, which are described below.

2.3.5 Molecular Dynamic (MD) simulations:

Molecular dynamics simulation (MD) is a computational method that analyses the physical movements of atoms and molecules. These simulations can provide details concerning individual particle motions as a function of time, enabling one to address specific questions about a model

system (Karplus & McCammon, 2002). MD's ability to simulate large biomolecular systems and sample more significant conformational changes over long-time scale periods makes it possible to determine functional implications that specific changes such as mutations may have on protein structures, thus playing an essential role in our understanding of biology (Karplus & McCammon, 2002).

MD simulations are widely employed in studies investigating structure-function relationships, disease pathways, and drug design, relying on the relationship between a given configuration of atoms and its energy to propagate dynamics (Nair & Miners, 2014). In this study, we used the GROMACS v.2020 suite to perform 200 ns MD simulations for each of the four systems: *NPSR1* WT and variant (p.V138I) and for *EPB41L2* WT and variant (p.R663C). The WT and variant *NPSR1* protein systems were simulated with the GBQ ligand-bound, forming the *NPSR1*-GBQ complex, which will be discussed further in the next chapter.

2.3.5.1 Preparation of MD simulation input files with CHARMM-GUI:

CHARMM-GUI (v.24), an online-based tool, was utilized to build the coordinates and topology of the WT and MT protein structures of *NPSR1* (p.V138I) and *EPB41L2* (p.R663C) (Brooks et al., 2009). For the *NPSR1*-GBQ complex, the topology of the GBQ ligand was prepared separately using the PARAMCHEM CHARMM (v.24) general force field. Topology is defined as a description of the properties and connectivity of all atoms in a system. CHARMM-GUI has several modules, and for this study, the CHARMM-GUI solution builder interface was employed, which accepts PDB, PDBx/mmCIf, and CHARMM formats as input files (Jo et al., 2008). The four systems' atomic coordinates were uploaded separately to the solution builder interface to solvate the systems with water molecules (Jo et al., 2008; Lee et al., 2016). The computed energy used was the CHARMM single-point energy, ensuring all coordinates were defined (Vanommeslaeghe et al., 2010). A rectangular TIP3 water-box with an edge distance of 10.0Å from the edges of the protein was defined.

The systems were neutralized by adding counter ions to each of the systems. These were 138 positive potassium (K) and 149 negative chloride (Cl) ions to a 0.15M concentration of the *NPSR1* WT and MT (p.V138I) systems. Additionally, 140 positive K and 140 negative Cl ions were also added to a 0.15M concentration of the *EPB41L2* WT and MT (p.R663C) systems. For all four protein systems, the Monte-Carlo ion placing method was used. An empirical force field, which is an accurate mathematical description of the relationship of conformation to energy in and between molecules, is vital for MD simulations, and in this study, the CHARMM36M was used as it is optimized for proteins and small molecules as well as nucleotides and lipid molecules (Huang & MacKerell, 2013). The NVT (constant number of particles, Volume, and Temperature) and NPT (constant number of particles, Pressure, and Temperature) ensembles were performed at a constant temperature of 303.15K.

2.3.5.2 Energy minimization:

CHARMM-GUI solution builder interface generated input files necessary for performing energy minimization of the various protein systems in an aqueous solvent environment (Lee et al., 2016). Each of the four systems underwent 50000 steps of steepest descent minimization integrator to energy minimize the solvated structures using the CHARMM36M force field, applying constraints to the hydrogen bonds, and employing the LINCS constraint algorithm as well as the Verlet cutoff-scheme (Vanommeslaeghe et al., 2010). This number of energy minimization steps was utilized to completely relax the systems' atomic coordinates to obtain a low energy state. Gromacs v.2020 was used to run all simulations in this study (Gajula et al., 2016).

2.3.5.3 System equilibration:

In the equilibration stage of MD simulations, atoms of the macromolecules and the surrounding solvent undergo a relaxation that usually lasts for tens or hundreds of picoseconds before the system reaches a stationary state (Stella & Melchionna, 1998). The simulation time required to equilibrate

protein systems depends on two significant factors, namely, the system size (the larger the system, the slower the equilibration) and the starting structure used (quality of the protein model used and how many solvent molecules is present in the system) (Kandt et al., 2007).

After completing the energy minimization step, all the systems underwent a two-step equilibration phase, namely NVT and NPT. In both the isothermal-isochoric (NVT) and the isothermal-isobaric (NPT) ensembles, the number of particles and temperature were conserved (Binder et al., 2004). The volume was conserved for the NVT ensemble, while pressure was conserved in the NPT ensemble. NVT was run for 125 picoseconds (ps) to stabilize the system's temperature, and a short position restraint NPT was run after that for 125 ps to stabilize the system's pressure by keeping the protein restrained. The Nose-Hoover temperature-coupling method (Ruhle, 2007) was used for the NVT ensemble, with a constant coupling of 1.0 ps at 303.15 K under a random sampling seed. For NPT, Parrinello-Rahman pressure coupling was turned on with a constant coupling of 2.0 ps at 303.15 K and compressibility of 4.5×10^{-5} bar^{-1} under conditions of position restraints (all-bonds) (Cherniavskyi et al., 2021; Martoňák et al., 2003).
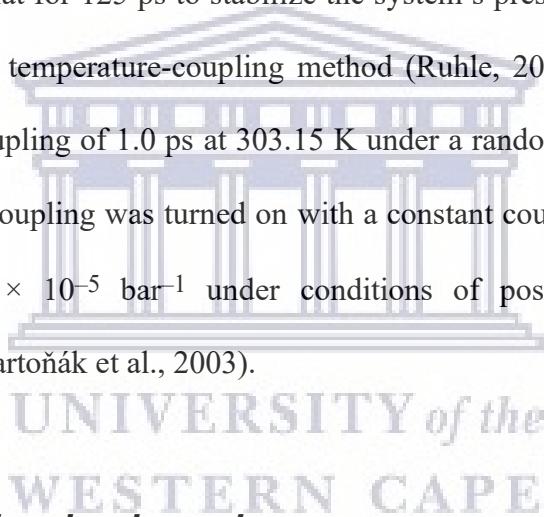
2.3.5.4 Production molecular dynamics:

After all the systems were equilibrated at the desired temperature and pressure, position restraints were removed from the protein, and the simulation proceeded in an unbiased manner. The simulation conditions included an integration time step of 0.002 ps, and the MD trajectories were recorded every 10 ps. The Centre for High-Performance Computing (CHPC) resources (Amolo, 2018) was utilized due to its ability to run simulations on multiple computer processors. An extensive production MD was initiated for a 200 nanosecond (ns) period to analyze structural changes in protein dynamics for all four systems at random seed numbers. Random seed number refers to random starting protein conformation (Lian, 2018). Each simulation was repeated to validate the reproducibility of results.

2.3.6 Analysis of MD simulations:

The analysis of the resulting MD trajectory files was done using GROMACS v.2020 utilities. The root mean square deviation (RMSD) was calculated for the backbone atoms of the protein using *gmx rms,* and the root mean square fluctuation (RMSF) per-residue analysis was done using *gmx rmsf*. RMSD reveals the average displacement of atoms compared to a reference structure during the simulation (Martínez, 2015). If the RMSD achieves a stable plateau, the simulation is considered to have attained a stable conformation and reached equilibration (Agrahari et al., 2019). The last 50 ns of the trajectory were analyzed, representing the most stable part of the simulation. RMSF measures the average fluctuation of a particular residue or atom over time (Martínez, 2015). It is used to obtain information on the local structure flexibility, thermal stability, and heterogeneity of macromolecules (Kuzmanic & Zagrovic, 2010). In this study, RMSF was utilized to inspect each residue's flexibility. Additionally, the mean and standard deviation (stdev) was calculated for the RMSD and RMSF values of both the WT and MT systems.

The radius of gyration (Rg) was calculated for the backbone using *gmx gyrate.* Rg, *which* provides information on the compactness and size of the protein molecules (Arnittali et al., 2019), was carried out to analyze the impact of the variant on the folding state of the protein system (Agrahari et al., 2019). The solvent-accessible surface area (SASA) of the protein atoms of the four protein systems was calculated using *gmx sas*. SASA is defined as the surface area of a biomolecule that is accessible to a solvent and the ability of its atoms to form contacts with the solvent (Ausaf Ali et al., 2014). SASA is a critical factor in studies inspecting protein folding and stability, as it plays a significant role in understanding the structure-function relation of proteins and their residues. The total intra-molecular hydrogen bonds (H-bonds) between the *NPSR1* protein and its ligand GBQ was calculated using *gmx hbond*. Additionally, the non-bonded pairwise interaction energy analysis was performed using *gmx energy*. In this calculation, the Lennard-Jones energy and the coulombic electrostatic energy were considered.

The Visual Molecular Dynamics (VMD) package was used to inspect the systems' motions along the complete 200 ns of the trajectory. VMD is a molecular visualization program for displaying, animating, and analyzing large biomolecule systems using 3D graphics and built-in scripting (Humphrey et al., 1996). As MD simulations are time-dependent and each frame of time carries vital information related to interactions, analysis and proper understanding of the simulation are very important. For this study, VMD was used to inspect the motions along the trajectory of the protein systems and reported as movies.

2.3.6.1 Principal component analysis:

Principal Component Analysis (PCA) is a statistical technique that reduces the complexity of data sets to extract biologically relevant movements of protein domains from irrelevant localized motions of atoms (David & Jacobs, 2014). PCA is a linear transform method that extracts basic movements from the data set using a covariance matrix constructed from atomic coordinates that describe the accessible degrees of freedom of the protein residues (David & Jacobs, 2014). This method is a means to transform the data to a new coordinate system, such that the most significant variance determined by the scalar projection of the data comes to lie on the first coordinate (first principal component) and the second most significant variance on the second coordinate (Ayaz, 2003).

For this analysis, firstly, we generated a covariance matrix using *gmx covar*. This was done to understand how the atom positions vary from the mean with respect to each other. Next, we computed the eigenvectors and eigenvalues of the covariance matrix to identify the principal components. Eigenvectors of the covariance matrix were used as they represent the directions of the axes where there is most variance; eigenvalues are the coefficients attached to eigenvectors, which give the percentage of variance carried in each principal component. Lastly, we created a 2D graph plotting variance distribution along the two major principal component axes. This was to understand

differences in large-scale movements of the protein systems throughout the simulation landscape.

Figure 2.1 illustrates a flowchart summarizing the methodology utilized to carry out this study.
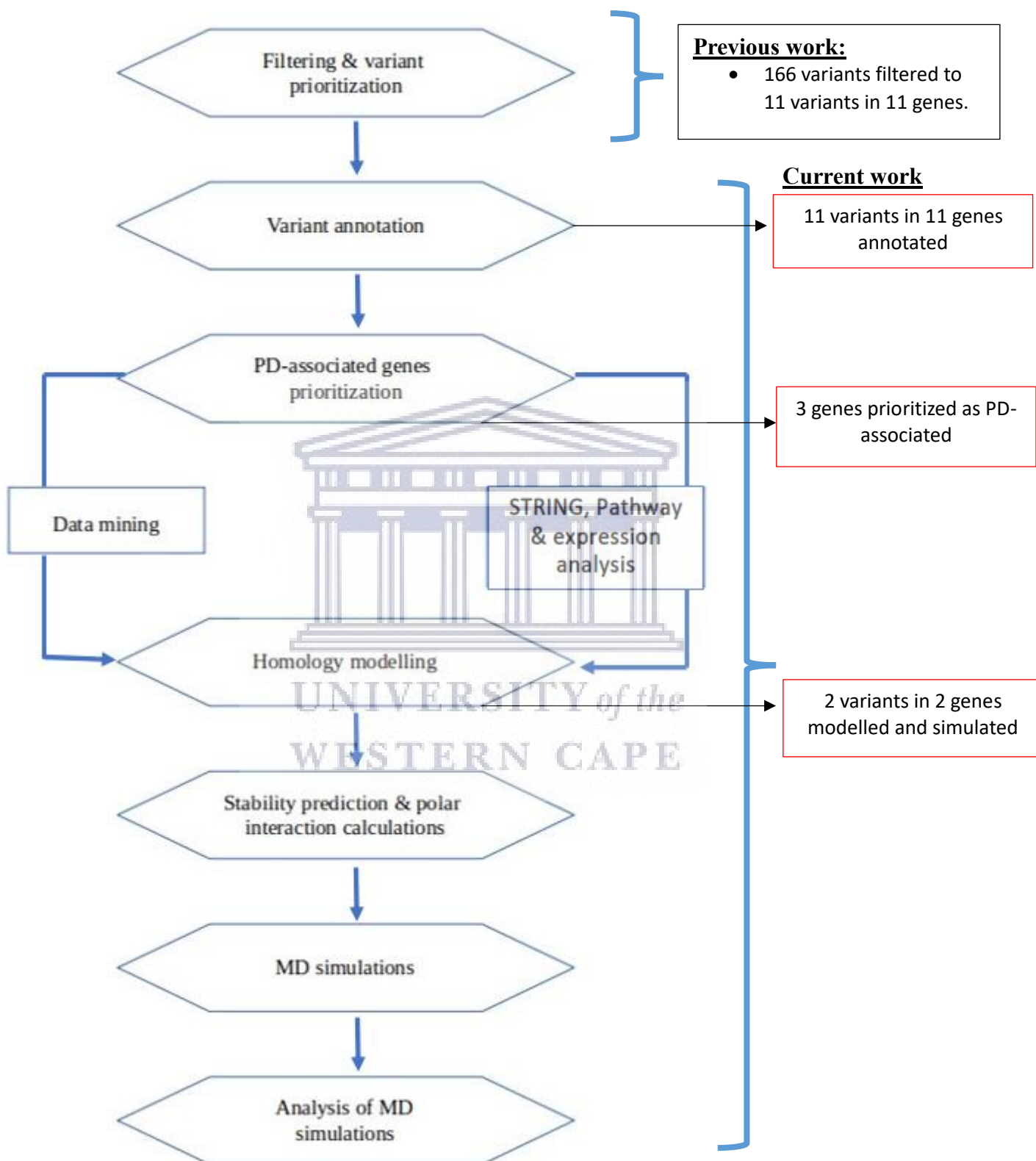
Figure 2.1: Flowchart representing the methodology of this study. PD: Parkinson's disease, MD: molecular dynamics.

CHAPTER III: Results

3.1 Description of the family:

A South African family of Indian ancestry with Parkinson's disease (PD) was identified by a neurologist, Dr. Jonathan Carr and clinically assessed by Dr. Roopnarain. A simplified version of the pedigree indicating the family members who took part in this study (four members affected with PD; a mother and her three sons) are shown in Figure 3.1 and denoted as family ZA 398.

Unfortunately, the mother passed away many years before starting this study, which did not allow for a clinical assessment of her symptoms by the neurologist or to collect a blood sample for DNA analysis. We did, however, obtain clinical information about the mother from her children by word of mouth.

As mentioned previously, all three affected individuals share the 11 variants that were prioritized in the previous study. Therefore, the focus of the current study was to further investigate the 11 variants and to prioritize putative pathogenic variants for future functional studies.
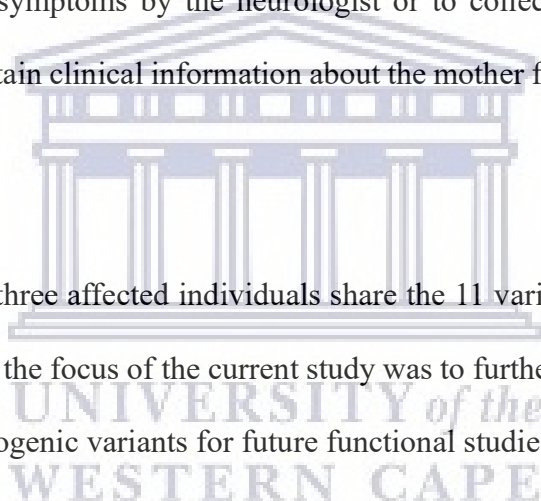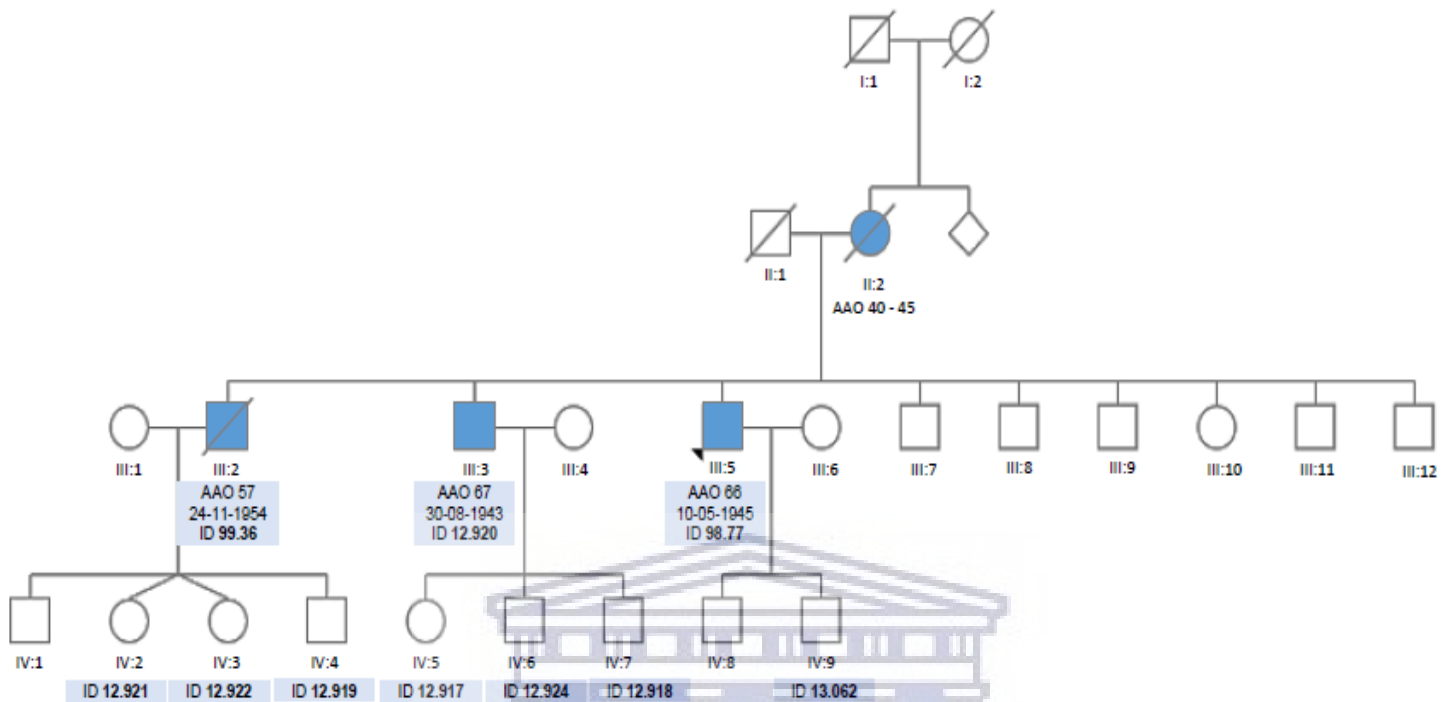
Figure 3.1: Pedigree of family ZA 398. Affected family members (PD patients) shown in blue. Squares represent males, while circles represent females. Dashed lines through the squares and circles indicate deceased individuals. Roman numerals represent generations of offspring. AAO = Age at onset of PD, ID = Identification number of individuals that took part in the study. Personal information removed to allow anonymity of individuals.

3.2 Variant annotation:

The UMD–predictor was used to verify the positions of variants on the respective genes. This web-based tool gives the chromosome and chromosome position on which the gene is found, gene name, the protein position, both the wild-type and the variant amino acid and indicates whether the variant is pathogenic or a benign polymorphism. We do not report here the pathogenicity of the variants, as this initial analysis was done only to confirm the position of the variants.

Table 3.1: Tabulated representation of the variant position annotation by the UMD-predictor.

Chromosome (Chr)	Chromosome position	Gene symbol	Amino acid position	Wild-type amino acid	Variant amino acid
Chr1	205592869	*ELK4*	48	R	C
Chr1	243349703	*CEP170*	377	T	M
Chr3	127980996	*EEFSEC*	184	V	M
Chr3	160156631	*TRIM59*	114	K	R
Chr4	3210535	*HTT*	2063	R	H
Chr4	1411464647	*ELMOD2*	215	A	S
Chr6	131199300	*EPB41L2*	663	R	C
Chr6	17626089	*NUP153*	1284	T	I
Chr7	34851409	*NPSR1*	138	V	I
Chr15	101567928	*LRRK1*	871	T	M
Chr19	12759179	*MAN2B1*	825	V	A

Abbreviations: *ELK4*= ETS transcription factor, *CEP170*= Centrosomal protein 170-KD, *EEFSEC*= Eukaryotic elongation factor selenocysteine-tRNA-specific, *TRIM59*= Tripartite motif-containing protein 59, *HTT*= Huntingtin, *ELMOD2*= ELMO domain containing 2, *EPB41L2*= Erythrocyte membrane protein 4.1 like 2, *NUP153*= Nucleoporin 153-KD, NPSR1= Neuropeptide S receptor 1, *LRRK1*= Leucine-rich repeat kinase 1, *MAN2B1*= Mannosidase alpha class 2B member 1, R= Arginine, C= Cysteine, T= Threonine, M= Methionine, V= Valine, K= Lysine, H= Histidine, A= Alanine, S= Serine, I= Isoleucine.

3.3 Functional gene annotation using data mining:

Presented below are the results obtained after data mining was performed on the 11 candidate genes. ETS-domain protein (*ELK4*) is a member of the E-twenty-six (ETS) family of transcription factors and the ternary complex factor (TCF) subfamily (Rickman et al., 2009). ETS-domain proteins play a role as either transcriptional repressors or activators, and their activities are often regulated by signal transduction pathways, including the MAP kinase pathway (Sharrocks et al., 1997). ETS-domain transcription factors regulate a diverse array of biological functions, including mammalian hematopoiesis and Drosophila eye development. Hematopoiesis refers to the production of the cellular components of blood and blood plasma. A study conducted by Peng *et al.* (2016) identified *ELK4* as a protein-binding partner with cyclin-dependent kinase 2 (CDK2), a catalytic subunit of the CDK complex whose activity is restricted to the G1-S of the cell cycle, where cells make proteins

necessary for mitosis and DNA replication (Peng et al., 2016). Mutations in the *ELK4* gene have been associated with prostate cancer and melanoma in previous studies (Makkonen et al., 2008; Peng et al., 2016).

The centrosomal protein of 170 kDa (*CEP170*) is a component of the centrosome, a non-membranous organelle that functions as the major microtubule-organizing center in humans and is highly expressed in the ovaries and testis (Ishikawa et al., 1997). During mitosis, *CEP170* localizes to centrosomes as well as spindle microtubules and promotes microtubule organization and microtubule assembly. (Bärenz et al., 2018) CEP170 interacts with the intraflagellar transport protein 81 (IFT81), the SH3-domain-containing protein PRAX-1 and is phosphorylated by CDK1 and polo-like kinase 1 (PLK1) (Guarguaglini et al., 2005; Welburn & Cheeseman, 2012).

The eukaryotic elongation factor, selenocysteine-TRNA specific (*EEFSEC*), is a protein-coding gene that promotes selenocysteine incorporation into selenoproteins (Dobosz-Bartoszek et al., 2016). SNPs identified in the *EEFSEC* locus were significantly associated with gestational length, preterm birth, increased risk of prostate cancer, and reduced risk of hypospadias (G. Zhang et al., 2017). Diseases associated with *EEFSEC* include Karyomegalic, Interstitial Nephritis, and Drug-Induced Lupus Erythematosus (Perico et al., 2018).

The tripartite motif-containing protein 59 (*TRIM59*) is an essential member of the TRIM family, regulating biological processes such as cell cycle, cell apoptosis, and natural immunity to viruses (P. Zhang et al., 2019). A study by Geng *et al.* (2019) identified *TRIM59* as a critical oncoprotein relating to lung cancer proliferation and metastasis (Geng et al., 2019). *TRIM59* has also been shown to promote retinoblastoma progression and tumor growth in hepatocellular carcinoma (Wu et al., 2020; Ying et al., 2020).

The huntingtin (*HTT*) gene is linked to Huntington's disease, a neurodegenerative disorder characterized by loss of striatal neurons (Sienes Bailo et al., 2020). This may be caused by an unstable, expanded trinucleotide repeat in the *HTT* gene, translating as a polyglutamine repeat in the protein product (Futter et al., 2009). *HTT* is widely expressed as two alternatively polyadenylated forms displaying different abundance in various fetal and adult tissues required for normal development.

ELMO domain-containing 2 (*ELMOD2*) gene encodes one of six engulfment and motility (ELMO) domain-containing proteins and may play a role in antiviral responses. Mutations in *ELMOD2* may be involved in the cause of familial idiopathic pulmonary fibrosis (Pulkkinen et al., 2010).

Erythrocyte membrane protein 4.1-like 2 (*EPB41L2*), also known as protein 4.1G, is a member of the protein 4.1 families, which function as adaptors linking transmembrane proteins to the cytoskeleton (Yang et al., 2011). This family includes four members: 4.1R, 4.1G, 4.1B, and 4.1N that share high sequence homology in three functional domains, namely, the N-terminal 4.1-ezrin-radixin-moesin (FERM) domain, the internal spectrin-actin-binding domain (SABD), and the C-terminal domain (CTD) (Yang et al., 2011). A study carried out using real-time PCR revealed that *EPB41L2* is predominantly expressed in the brain, spinal cord, and along the membrane of Sertoli cells in the testis (Pm et al., 2005). *EBP41L2* was found to be localized at two specific regions in Schwann cells forming myelin in the peripheral nervous system: Schmidt-Lanterman incisures (SLIs) and paranodes (Ohno et al., 2006; Saitoh et al., 2017).

Nucleoporin 153 (*NUP153*) protein belongs to the nucleoporin family, glycoproteins involved in regulating the transport of macromolecules between the nucleus and cytoplasm (McMorrow et al., 1994). *NUP153* plays a role in repairing double-strand breaks and activating DNA damage checkpoints (Lemaître et al., 2012). This protein is also necessary for HIV-1 replication (Buffone et al., 2018).

Neuropeptide S receptor 1 (*NPSR1*) gene encodes a membrane protein and is a member of the vasopressin/oxytocin subfamily of G protein-coupled receptors. The resulting protein acts as a receptor for neuropeptide S and affects various cellular processes through its signaling. Neuropeptide S is a neurotransmitter expressed in the brainstem, and its mRNA occurs in some regions of the fear circuit, such as the amygdala and hypothalamus (Gechter et al., 2019). Polymorphisms in the *NPSR1* gene have also been associated with asthma susceptibility, insomnia, panic disorders, inflammatory bowel disease, and rheumatoid arthritis (Ghazal, 2016; Hall et al., 2019; Ober & Yao, 2011).

Leucine-rich repeats kinase 1 (*LRRK1*) is a multi-domain protein belonging to the ROCO family of complex proteins (Korr et al., 2006). ROCO proteins constitute a novel subgroup of Ras-like GTPases that are characterized by unique domain architecture. All ROCO proteins contain a GTPase-like domain named *R*as *o*f *c*omplex proteins (Roc) followed by the *C*-terminal *o*f *R*oc (COR) domain, a stretch of 300–400 amino acids of unknown function showing no significant sequence homology to any functionally described protein domain (Bosgraaf & Van Haastert, 2003). Additionally, *LRRK1* has been shown to play a critical role in regulating bone mass in humans, and mutations in it have been identified in patients with osteosclerotic metaphyseal dysplasia (L. Guo et al., 2017; Iida et al., 2016).

Mannosidase alpha class 2B member 1 (*MAN2B1*) gene encodes an enzyme that hydrolyses terminal, non-reducing alpha-D-mannose residues in alpha-D-mannosidase. The activity of this enzyme is necessary for the catabolism of N-linked carbohydrates released during glycoprotein turnover, and it is a member of family 38 of glycosyl hydrolases. Mutations in *MAN2B1* have been linked to Alpha-mannosidosis, a rare autosomal recessive disorder caused by the deficiency of lysosomal alpha-mannosidase (Sbaragli et al., 2005).

3.4 Pathway, expression and STRING analysis:

To further prioritize the possible causal role of these variants and genes in PD, gene expression profiles and pathway databases were interrogated for each gene. The publicly available databases used in this study included the Allen Brain Atlas, KEGG's pathway, PANTHER pathway, Human Protein Atlas, and STRING database. Gene expression analysis was done to determine which genes were highly expressed in the brain, and pathway analysis was done to determine if any of the 11 genes were involved in PD-linked or dopamine-related biological pathways. Additionally, STRING analysis was performed to identify if any of the 11 genes interacted with PD-linked genes. Two known PD genes were also included as positive controls in these analyses to test the efficacy of this prioritization method namely, *SNCA* and *PARK7*. *SNCA* was chosen as it encodes a presynaptic protein (α-synuclein) that plays a significant role in PD-related pathways. *PARK7*, on the other hand, encodes the *DJ-1* protein and plays a key role in mitochondrial function, which has been implicated as an important biological process underlying the development of PD. *SNCA* is highly expressed in brain regions such as the thalamus, basal ganglia and midbrain and is involved in PD-linked pathways such as calcium signaling and dopamine metabolism. *PARK7* is expressed in the amygdala, hippocampus and cerebral cortex and is broadly involved in pathways previously linked to neurodegeneration in multiple diseases. The results of the positive controls were then used as a guide to prioritize putative candidate variants from the 11 sequence variants in the current study, for further analysis.

STRING analysis indicates that the *EPB41L2* protein interacts with several synaptic cell-adhesion molecules such as neuroligins and neurexins (Figure S5). Neuroligins and neurexins connect pre- and postsynaptic neurons at synapses, shape neural network properties, and mediate trans-synaptic signaling (Südhof, 2008). In humans, alterations in neuroligin or neurexin genes are implicated in autism and other cognitive diseases. Several significant protein-protein interactions between *EPB41L2-NLGN1* (0.907), *EPB41L2-NLGN2* (0.904), *EPB41L2-NRXN1* (0.920), *EPB41L2-NRXN2* (0.920), and *EPB41L2-NRXN3* (0.920) were recorded indicating the strength of the association

between the genes and amount of evidence available for the interactions. The closer the combined score is to 1, the more evidence available for interactions between the genes. Furthermore, it also revealed that *NPSR1* interacts with proteins involved in the positive regulation of the circadian sleep/wake cycle as well as in hormone-mediated apoptotic signaling pathways (Figure S4). Several significant protein-protein interactions between *NPSR1-PTh* (0.951), *NPSR1-CASR* (0.967), and *NPSR1-CALCA* (0.953) were recorded indicating the strength of the association between the genes and amount of evidence available for the interactions (Figure S4). Additionally, STRING analysis indicates that *LRRK1* interacts with four PD-associated genes, namely, *PARK2, PINK1, PARK7* and *VPS35* (Figure S6). *LRRK1* recorded a combined score of significant protein-protein interactions between LRRK1-PARK2 = 0.730; LRRK1-PARK7 = 0.666; LRRK1-PINK1 = 0.741; LRRK-VPS35 = 0.737. STRING analysis of the remaining eight genes showed that they did not interact with any PD-associated genes.

Table 3.2: Gene expression profiles and pathway data of the 11 candidate genes (in alphabetical order) under investigation compared to two known PD genes.

Gene	Allen Brain Atlas	Human Protein Atlas	KEGG Pathway Analysis	PANTHER Pathway Analysis
CEP170	Cerebral cortex, Grey matter	Testis, Caudate, Cerebellum, Cerebral cortex, Pons & medulla, Midbrain, Thalamus, Bone marrow, Spinal cord	Ciliogenesis, Centrosome formation	Centriole appendage assembly, Microtubule organization
EEFSEC	Grey matter, Pons, Cerebral cortex	Parathyroid gland, Skeletal muscle, Cardiac muscle, Cerebellum, Cerebral cortex, Hippocampal formation	Translation factors, Elongation factors	Translational elongation, Selenocysteine
ELK4	Basal ganglia, White matter	Prostate, Testis, Oesophagus,	MAPK signaling pathway,	Interleukin signaling pathway, PDGF

		Adipose tissue, Cerebellum	Transcriptional misregulation in cancer	signaling pathway, CCKR signaling map
ELMOD2	Cerebellar cortex, Grey matter, Hypothalamus, Substantia nigra, Pons, Cerebral cortex, Thalamus, Ventricles, Cerebellum	Thyroid gland, Adrenal gland, Bronchus, Duodenum, Testis, Pituitary gland	Membrane trafficking	Antiviral defence
EPB41L2	White matter, Grey matter	Hippocampus, Caudate, Spinal cord, Retina, Thalamus, Midbrain, Cerebral cortex, Testis, Basal ganglia	Signaling & cellular processes	Dopamine receptor-mediated signaling pathway, Nicotine pharmacodynamics pathway
HTT	Cerebellar cortex	Cerebral cortex, Parathyroid gland, Cerebellum, Amygdala, Lung	Huntingtin Disease	Huntingtin's disease
LRRK1	Not expressed in the brain	Detected in all tissue, Lymph node, Appendix, Tonsil, Spleen	Transferases	Canonical wnt signaling pathway
MAN2B1	Hypothalamus, White matter, Thalamus, Ventricle, Pons, Grey matter, Cerebellum	Lymph node, Spleen, Monocytes.	Glycan degradation, Lysosome process	Protein deglycosylation, neutrophil degranulation, cellular protein modification process.
NPSR1	Cerebral Cortex, Thalamus, Grey matter, Pons, Hypothalamus, Basal ganglia	Brain, Retina, Stomach, Cerebellum, Hypothalamus, Testis, Pons& medulla, Corpus callosum, Cerebral cortex	Signaling & cellular process	G-protein coupled receptor signaling pathway (Reactome), Neuropeptide signaling pathway, Up-regulation of the release of sequestered calcium ions into the cytosol (UniProtKB)
NUP153	Grey matter, Cerebral cortex	Bone Marrow, Testis, Cerebellum,	RNA transport	Nuclear pore assembly

		Thymus, Skeletal muscle		
PARK7	Hippocampal formation, Amygdala, Gray matter	Hippocampus, Cerebral cortex, Cerebellum, Amygdala	Parkinson's disease, Pathways of neurodegeneration – multiple diseases, Transcription.	Sumoylation of transcription co-factors (Reactome)
SNCA	Thalamus, Grey matter	Cerebral cortex, Hippocampal formation, Amygdala, Basal ganglia, Hypothalamus, Midbrain, Thalamus	Alzheimer's disease, Parkinson's disease, Pathways of neurodegeneration – multiple diseases, Calcium signaling, Dopamine metabolism	Parkinson's disease
TRIM59	White matter	Brain & lymphoid tissue, Corpus callosum, Thalamus, Midbrain, Thymus,	Ubiquitin system	I-kappaB kinase/ NF-kappaB signaling

Abbreviations: *ELK4*= ETS transcription factor, *CEP170*= Centrosomal protein 170-KD, *EEFSEC*= Eukaryotic elongation factor selenocysteine-tRNA-specific, *TRIM59*= Tripartite motif-containing protein 59, *HTT*= Huntingtin, *ELMOD2*= ELMO domain containing 2, *EPB41L2*= Erythrocyte membrane protein 4.1 like 2, *NUP153*= Nucleoporin 153-KD, *NPSR1*= Neuropeptide S receptor 1, *LRRK1*= Leucine-rich repeat kinase 1, *MAN2B1*= Mannosidase alpha class 2B member 1, *SNCA*= Synuclein alpha, *PARK7*= Parkinson disease protein 7, MAPK: Mitogen-activated protein kinase, PDGF: Platelet-derived growth factor, CCKR: Cholecystokinin receptor.

The results tabulated show that *NPSR1* and *EPB41L2* were noted by both the Allen Brain Atlas (https://celltypes.brain-map.org/) and Human Protein Atlas (HPA) (https://www.proteinatlas.org/) to be highly expressed in the midbrain, thalamus and basal ganglia, which are vital regions known to be involved with PD pathogenesis (Table 3.2, Table S2 and Figures S1 and S2). Pathway analysis (http://pantherdb.org/; https://www.genome.jp/kegg/) has shown that both *NPSR1* and *EPB41L2* are involved in signaling and cellular processes. *NPSR1* is involved in the G protein-coupled receptor signaling pathway, neuropeptide signaling pathway and the up-regulation of the release of sequestered

calcium ions into the cytosol. *EPB41L2*, on the other hand, plays a role in the dopamine receptor-mediated signaling pathway as well as the nicotine pharmacodynamics pathway. This agrees with the results recorded from the known PD genes, *SNCA* and *PARK7* genes, which were used as positive controls as explained above. *ELMOD2* is expressed in the substantia nigra, hypothalamus and thalamus but only involved in membrane trafficking and antiviral defense pathways. *MAN2B1* and *ELK4* are also expressed in the hypothalamus, and basal ganglia, respectively but are not involved in any PD related pathways. *NUP153*, *CEP170*, *LRRK1*, *EEFSEC*, *TRIM59* and *HTT* are not involved in PD related pathways, as shown in Table 3.2. Interestingly, *LRRK1* is involved in protein-protein interactions with three autosomal recessive PD genes (*PARK2, PINK1* and *PARK7*) and one autosomal dominant PD gene (*VPS35*) as shown in Figure S6.

From the functional gene annotation performed using data mining, STRING analysis, as well as the data recorded in Table 3.2, it is evident that *LRRK1, EPB41L2* and *NPSR1* fit the criteria of this study. This is because they are either involved in protein-protein interactions with PD-associated genes (Figure S6) or highly expressed (as shown in Table S2) in the basal ganglia, midbrain and thalamus and involved in PD related pathways. Examples of the PD related pathways include the dopamine receptor-mediated signaling pathway, G protein-coupled receptor signaling pathway, up-regulation of the release of calcium ions and the neuropeptide signaling pathway to mention a few. Regulation of the dopamine receptor-mediated signaling pathway is crucial, as a loss of dopamine neurons is the main feature of PD (Saikia, 2018). Moreover, maintaining calcium homeostasis is critical in the mitochondria of healthy individuals and a dysregulation in the levels of calcium ions present in the mitochondria act as a major pathological hallmark of PD (Ludtmann & Abramov, 2018). Additionally, components in the G protein-coupled receptor-signaling pathway are utilized as therapeutic targets in reducing psychiatric symptoms displayed by PD patients (Galet et al., 2021). Specific proteins in these pathways have been implicated in PD development and could act as significant risk factors as both dopamine regulation and maintenance of calcium homeostasis in the human body are critical factors (Chan et al., 2009). For these reasons, genes; *LRRK1, NPSR1* and *EPB41L2* were selected to

carry out further *in-silico* structural analyses. Unfortunately, due to a lack of a solved homologous template structure covering the variant position for *LRRK1* at the start of this study, we only performed *in-silico* structural analyses on *NPSR1* and *EPB41L2*. The remaining eight genes are considered to be of lower importance as they do not fit all the criteria mentioned above. However, they are all still plausible candidates and therefore should not be excluded from future studies.

3.5 Homology modeling:

The two prioritized proteins, *NPSR1* and *EPB41L2*, containing amino acid substitutions at positions p.V138I and p.R663C, respectively, were further pursued in homology modeling studies to assess the effects of the variants on the protein structures. For the *NPSR1* protein, the homologous x-ray diffraction crystal structure of the human NK1 substance P (SP) receptor with Protein Data Bank ID: 6J21 was selected to construct the 3D protein structure of *NPSR1*. SP receptor, a distantly related homolog to *NPSR1*, is involved in the G-protein coupled receptor protein-signaling pathway. The template 6J21 had the highest sequence coverage (0.79) and low sequence identity (21.77 %) to *NPSR1* and was therefore selected for homology modeling.

Furthermore, the 6J21 template protein structure had two ligands bound to it; the 5-[[(2~{R},3~{S})-2-[(1~{R})-1-[3,5-bi(trifluoromethyl)phenyl]ethoxy]-3-(4-fluorophenyl)morpholin-4-yl]methyl]-1,2-dihydro-1,2,4-triazol-3-one (GBQ) and the (2R)-2,3-dihydroxypropyl (9Z)-octadec-9-enoate (OLC), attached to the protein structure. GBQ, a neurokinin 1 (NK1) receptor antagonist, is a drug used in chemotherapy and general anaesthesia. In previous studies, SP and its tachykinin NK1 receptor have been identified as novel neuroprotective targets for dopaminergic degeneration (Thornton & Vink, 2015). OLC, a molecule from lipid monoolein, is an amphiphilic featuring a polar head group and a non-polar hydrocarbon chain (Kulkarni et al., 2011). It is an endogenous ligand and was included to solve the 6j21 template structure (Chen et al., 2019; Cherniavskyi et al., 2021). With

the aid of PyMol, both ligands were extracted to *NPSR1* to generate a complex (the *NPSR1*-GBQ-OLC complex). This was done by firstly aligning the 6j21 template and the solved *NPSR1* protein structure. The PyMol extract tool was then used to extract both GBQ and OLC ligands from the 6j21 template and onto the *NPSR1* protein structure. The resulting *NPSR1*-GBQ-OLC complex was then saved in PDB format. The two ligands were included to determine whether the binding of ligands to *NPSR1* was affected by introducing the p.V138I variant. Figure 3.2 illustrates the 3D cartoon representation of the *NPSR1* membrane protein homology model, consisting of 10 alpha-helices, 13 loops and two beta-sheets. The C-terminal (shown to the right) of the *NPSR1* membrane protein projects into the lumen, while the N-terminal (shown to the left) is in the cytosol.
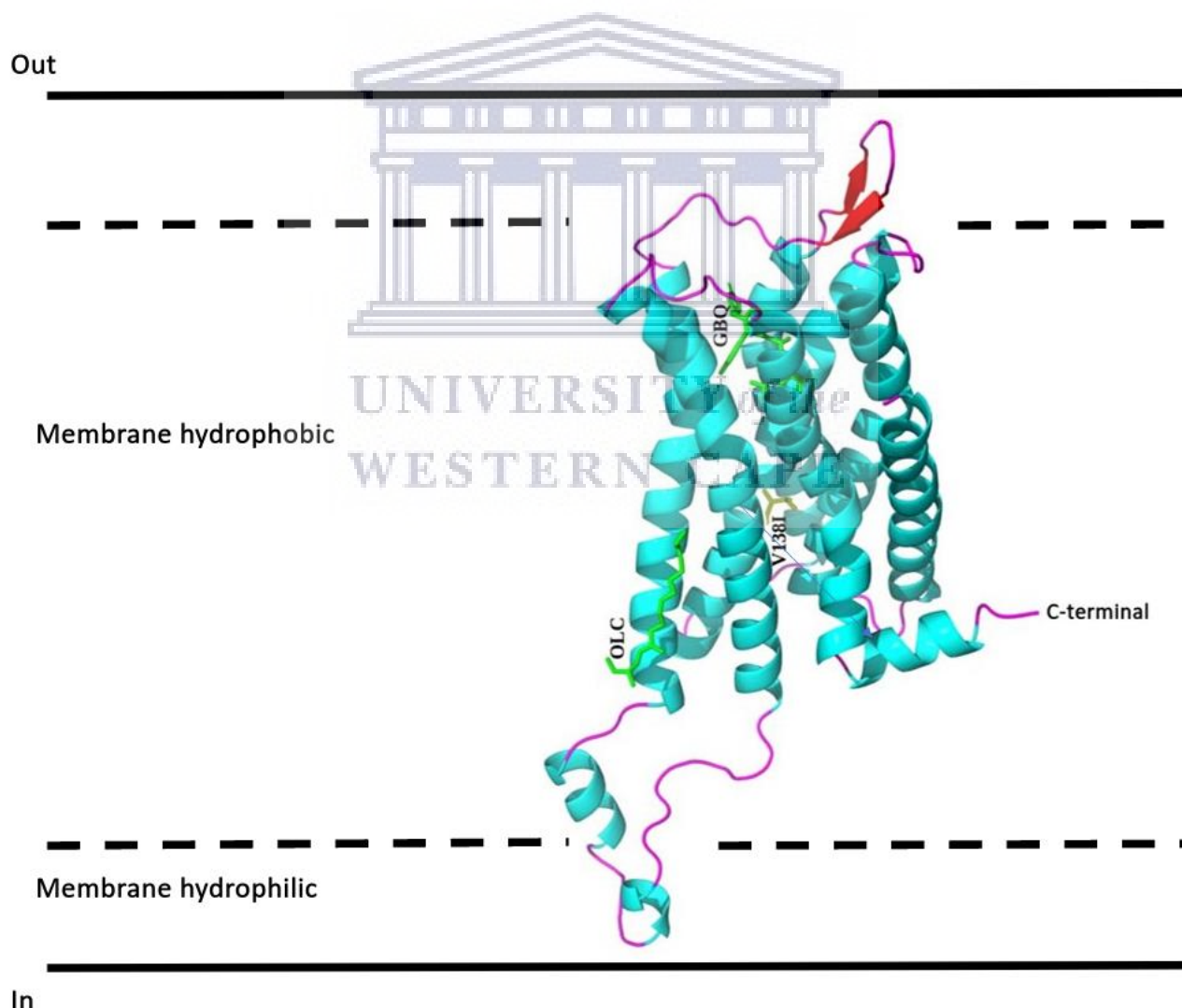
Figure 3.2: 3D cartoon representation of the *NPSR1* homology structure predicted with SWISS-Model. The protein A-chain is colored by its secondary structure, with alpha helices shown in cyan, beta sheets shown in red, and loops shown in magenta. The mutated residue Isoleucine 138 is labeled and shown as yellow sticks, and the GBQ and OLC ligands are labeled and shown as green sticks.

In addition, for the *EPB41L2* protein, the X-ray diffraction crystal protein structure of the AlphaII-spectrin tetramerization domain with PDB ID: 3F31, which is from homo sapiens, was selected to construct the 3D structure of the spectrin-actin-binding domain of this protein. The criteria used for selecting a homologous template were based on high sequence identity and sequence coverage, and coverage of the variant site. For the *EPB41L2* gene, the 3F31 template covering the AA region 634 – 770, as shown in the sequence alignment (Figure S18), presented the highest sequence coverage (0.73) and low sequence identity (27.42 %) relative to the other templates. The GMQE score was 0.21, indicating low confidence in the homology model's quality. Nevertheless, this is expected because of the low sequence identity between the target *EPB41L2* sequence and the homologous template 3F31 and because the predicted protein model still requires additional energy refinement. The predicted protein structure for *EPB41L2* consists of 9 alpha helices and 10 loop regions with no beta-sheets indicative of a transmembrane protein structure (Figure 3.3).
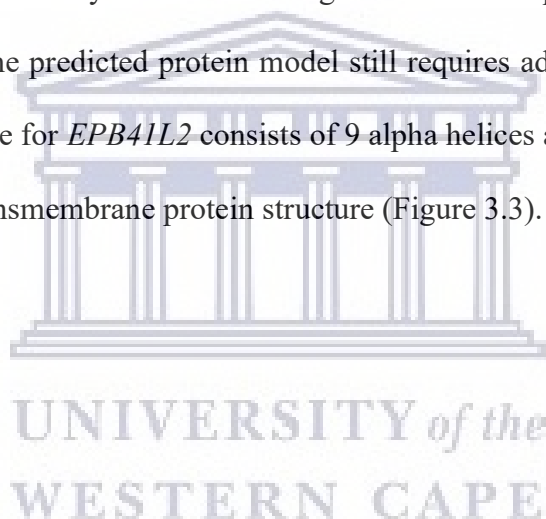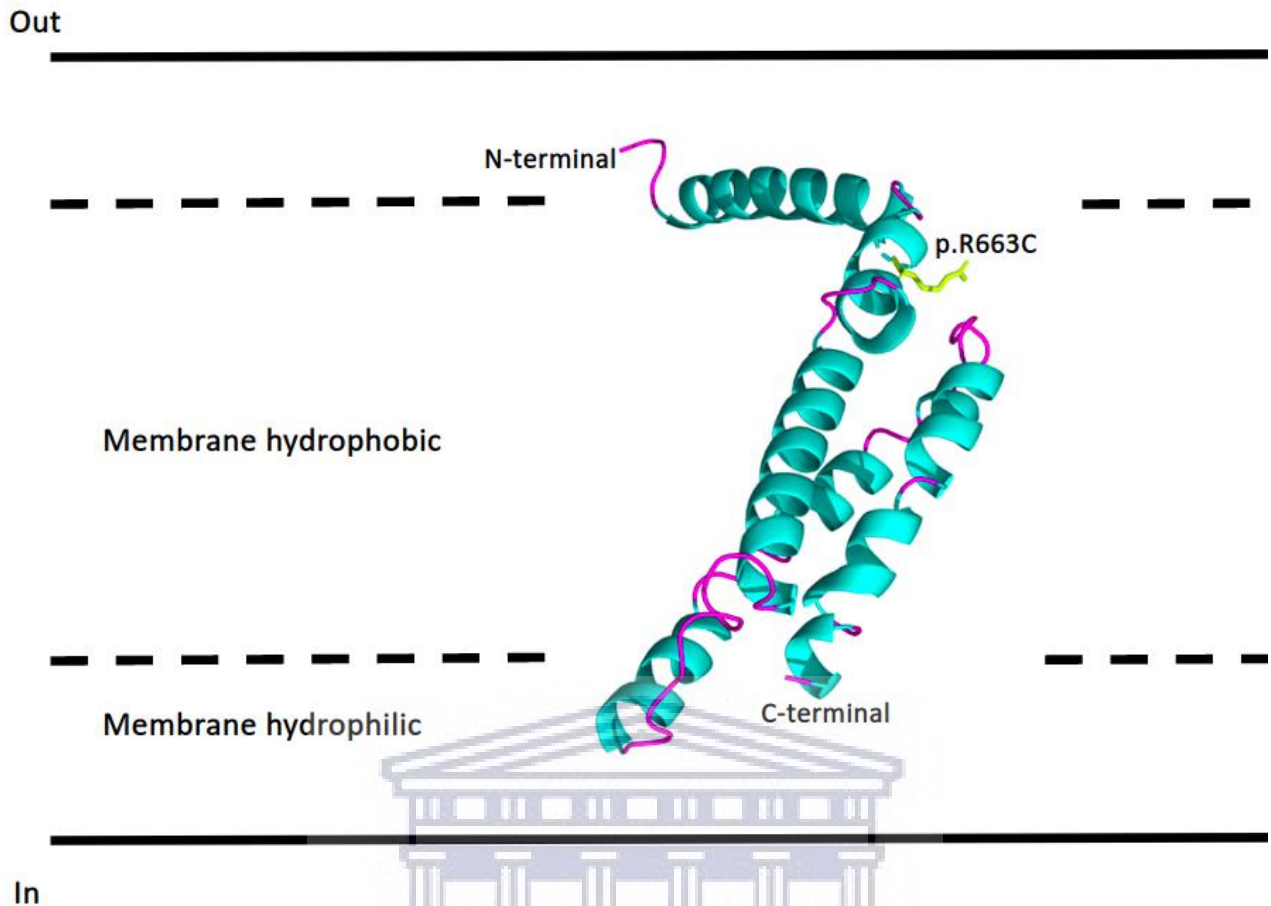
Figure 3.3: 3D cartoon representation of the *EPB41L2* homology structure predicted with SWISS-Model. The protein B-chain, a monomer, is colored by its secondary structure, with alpha helices shown in cyan and loops shown in magenta. The mutated residue Cys 663 is labeled and shown as yellow sticks.

3.6 3D protein model quality assessment:

The QMEAN score calculated for the *NPSR1* WT 3D protein complex was -5.00, demonstrating moderate reliability of the predicted *NPSR1*-GBQ-OLC protein complex compared to experimental structures of a similar size (Benkert et al., 2011; Studer et al., 2020). Moreover, the protein complex had a GMQE score of 0.54, which was closer to 1.00, indicating high confidence in the homology model's quality. Additionally, a low root mean square deviation (RMSD) score (0.121 Å) was calculated using PyMol align command between the *NPSR1* WT homology model and the 6J21 template. This indicated minimal deviation of backbone carbon atoms between target and template, suggesting high homology and structural similarity between the two structures. Quality assessment of the *NPSR1* WT and 6J21 template 3D protein structures was performed using the SAVES v6.0 web

server. Interestingly, PROCHECK analysis of the *NPSR1* WT homology model indicated satisfaction of stereo-chemical restraints. The *NPSR1* WT model contained 93 % of its residues in the most favored regions in the Ramachandran plot. The overall quality factor score, as evaluated by ERRAT, was high (93.33 %), indicating that all residues in the A-chain had a low error rate for the *NPSR1* WT homology model. *NPSR1* WT 3D structure failed the VERIFY3D quality check with only 45.24 % of the residues averaging a 3D-1D score >= 0.2. However, the 6J21 template scores also failed the VERIFY3D quality analysis and had a lower (37.94 %) score than our predicted model.

For the *EPB41L2* protein (WT), a quality assessment of the spectrin-actin-binding region and the 3F31 template was performed using the SAVES v6.0 web server. The WT homology model had a QMEAN score of -3.74, indicating that the generated model was of moderate quality compared to other experimental structures. The RMSD analysis calculated between the *EPB41L2* WT homology model and the 3F31 template using the PyMol align command revealed minimal deviation (0.626 Å) of the backbone carbon atoms. This suggests high homology and structural similarity between both structures as the RMSD (2.0 Å) was less than the RMSD cut-off value of 3.5 Å. PROCHECK and ERRAT analyses of the *EPB41L2* WT homology protein structure was satisfactory, with 84.9 % of the residues found in the most favored regions of the Ramachandran's plot and an overall quality factor score of 86.8 %. However, the WT homology model failed the VERIFY3D quality checks with only 47.45 % of residues averaging a 3D-1D score >= 0.2. Upon inspecting the VERIFY3D score of the 3F31 template, it was seen that only 51.62 % of the residues averaged a 3D-1D score >= 0.2. This indicates that the template structure 3F31 with a resolution of 2.3 Å and an R-factor of 0.28, used to generate the *EPB41L2* WT homology model, is of moderate quality. Subsequent modeling studies should consider using a template with higher resolution and higher sequence identity to the target sequence of *EPB41L2* to obtain a more accurate 3D structure of the target protein. In summary, based on the satisfaction of several qualitative factors, the protein models used here were considered reliable and were subsequently selected for MD simulation studies.

3.7 Protein stability and polar interaction calculations:

The most frequent cause of monogenic disorders is a single base DNA variant that results in an amino acid (AA) substitution that affects the enzyme activity (Yue et al., 2005). As shown in Table 3.3, both the p.V138I and p.R663C variants have slightly destabilizing effects on their respective protein structures, as predicted by the DUET web server. Additionally, no change in the number of polar contacts formed between the *NPSR1* protein and ligands (GBQ and OLC) is seen after introducing the variant. However, a slight change is seen in the distance between the H-bond formed between the oxygen atom in the GBQ ligand and the hydrogen atom of the ARG125 residue after the introduction of the p.V138I variant (Figure 3.4 A and 3.4 B). No change is seen in the polar interactions between OLC and *NPSR1* after the p.V138I variant is introduced. For this reason, when running the MD simulations, we only explored the *NPSR1*-GBQ complex further.

Table 3.3: Stability prediction of the variants on the energy minimized protein systems and analysis of their effect on the polar contacts

Protein	Predicted ΔG (Kcal/mol)		Polar contacts	
	DUET	Stability	Before variant	After variant
NPSR1 (p.V138I)	-0.324	Destabilizing	GBQ: Arg125 OLC: 0	GBQ: Arg125 OLC: 0
EPB41L2 (p.R663C)	-0.102	Destabilizing	1 (GLU668)	1 (GLU668)

Abbreviations: *NPSR1*= Neuropeptide S receptor 1, *EPB41L2*= Erythrocyte membrane protein 4.1 like 2 Arg= Arginine (R), Glu= Glutamic acid/ Glutamate, V= Valine, I= Isoleucine C= Cysteine.
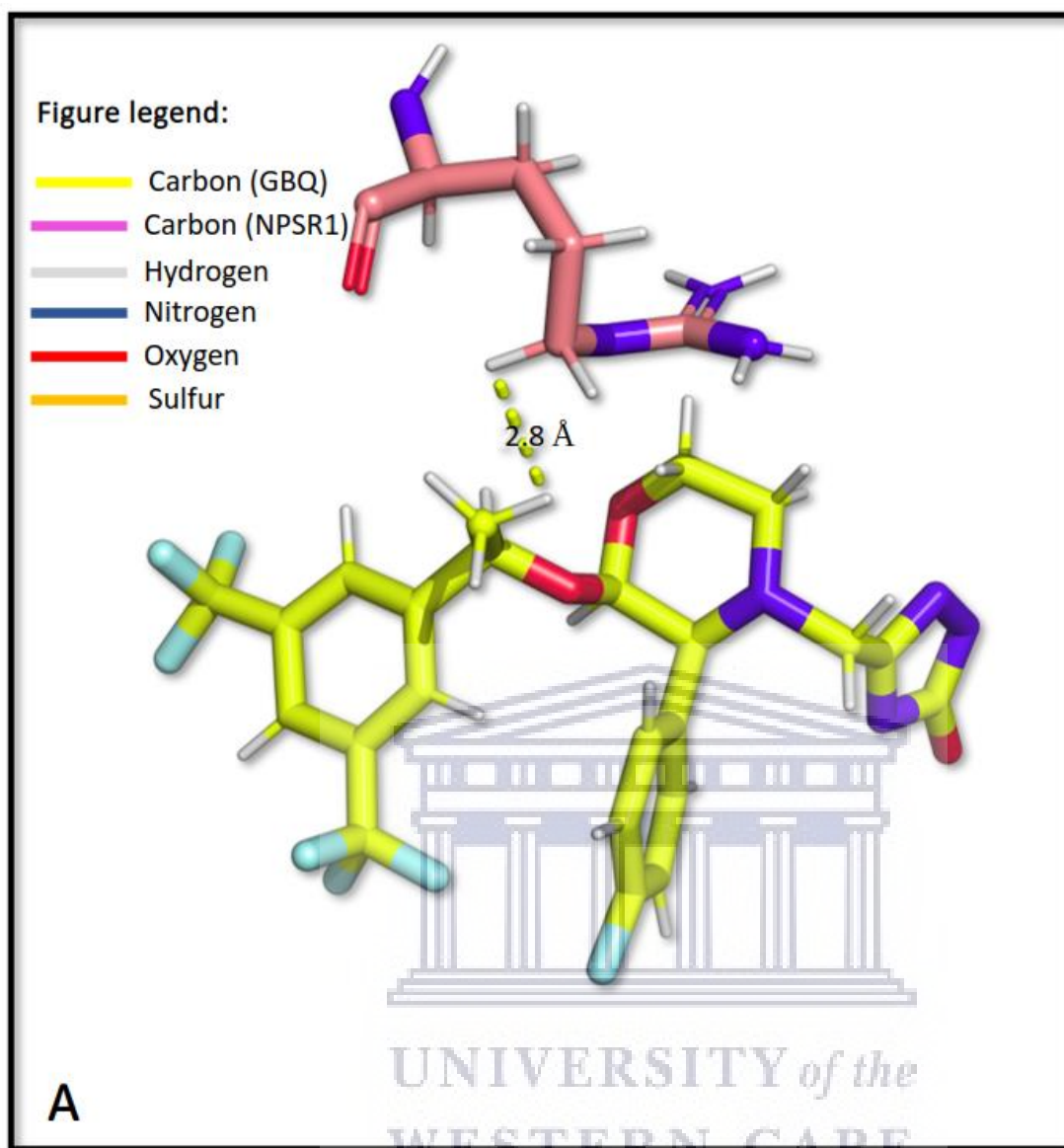
Figure 3.4: *NPSR1* **active site showing interaction with the GBQ ligand.**
A) Polar contacts formed between the GBQ drug, colored according to atom type (carbon: yellow, hydrogen: white, nitrogen: blue, oxygen: red, sulfur: orange). The residue of the *NPSR1* WT protein (ARG 125) forms a polar contact; labeled and colored according to atom type (carbon: magenta, hydrogen: white, nitrogen: blue, oxygen: red, sulfur: orange). Polar contacts shown as yellow dashes and labeled.
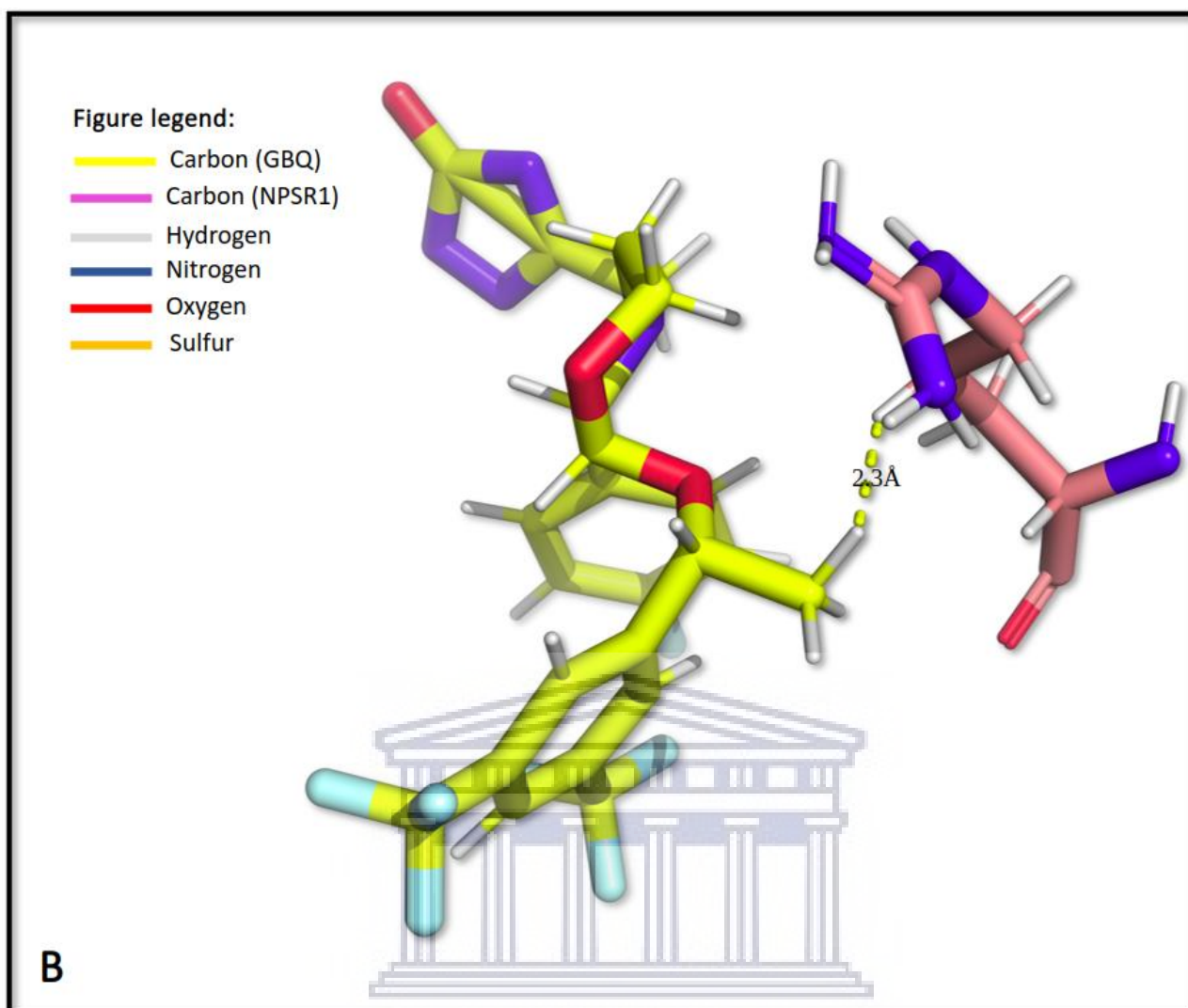
B) Polar contact formed between the GBQ drug, which is colored according to atom type (carbon: yellow, hydrogen: white, nitrogen: blue, oxygen: red, sulfur: orange). Residue that it forms a polar contact with (ARG 125) of the *NPSR1* MT (p.V138I) protein labeled and colored according to atom type (carbon: magenta, hydrogen: white, nitrogen: blue, oxygen: red, sulfur: orange). Polar contacts shown as yellow dashes and labeled.

3.8 Production MD simulations:

3.8.1 Energy minimization of *NPSR1* and *EPB41L2* protein structures:

Before running simulations, it was necessary to relax the solvated systems to a low-energy state. For the *NPSR1* WT and MT (p.V138I) protein structures, the potential energy (E_{pot}) was both negative values and fell within the $10^5 – 10^6$ range, suggesting that the energy minimization (EM) was successful as this is one of two critical factors used in determining convergence of the systems. The second factor, the maximum force (WT: Fmax = 9.0867 x10^2; MT: Fmax = 9.0909 x10^2), the target

for which was set in the EM script as 'emtol = 1000.0', which indicates a target Fmax of no greater than 1000 kJ mol^{-1} nm^{-1}. We observed that the Fmax of both the WT and MT was less than the set target indicating a successful EM.

A similar trend was seen in both the WT and MT (p.R663C) of the *EPB41L2* protein structures, with the E_{pot} values falling within the $10^5 - 10^6$ range and the Fmax values less than 1000 kJ mol^{-1} nm^{-1} as shown in Table 3.4, suggesting that all solvated systems were relaxed to a low-energy state. Energy minimization of all systems was repeated and showed similar results as observed for the first run (Table 3.4).

Table 3.4: Energy minimization results for the WT and MT protein structures of *NPSR1* and *EPB41L2*

		E_{pot} x E06 (kJ mol^{-1})	F_{max} x E02 (kJ mol^{-1} nm^{-1})	Repeat	E_{pot} x E06 (kJ mol^{-1})	F_{max} x E02 (kJ mol^{-1} nm^{-1})
NPSR1	WT	-2.349	9.0867	WT	-2.345	9.104
	MT (p.V138I)	-2.322	9.0909	V138I	-2.345	8.687
EPB41L2	WT	-2.126	8.7039	WT	-2.181	9.555
	MT (p.R663C)	-2.252	9.072	R663C	-2.256	9.765

Abbreviations: WT= wild-type, MT= variant, Epot= potential energy, Fmax= maximum force.

3.8.2 Equilibration of the *NPSR1* and *EPB41L2* protein structures:

Before performing simulations, equilibration was carried out to obtain constant temperature (T), pressure (P), and density for all the systems. We first employed the canonical (NVT) ensemble, where the number of particles (N) and volume (V) were constants. The results show that all systems, including the repeats, deviate at an average temperature range of 303.17 – 303.38 K. Subsequently, the isothermal-isobaric NPT ensemble was run to equilibrate the systems to constant pressure and density conditions. The average value of *NPSR1* systems pressure was at 2.2 Pa for the (WT) and -0.195 Pa for the (MT), while the *EPB41L2* system's average pressure was at 0.82 Pa for the (WT), -

0.166 Pa for the (MT). Overall, all four systems had an average density value of 1026 Kg/m^3, indicating a successful equilibration across all systems, including for each of the four repeats.

3.9 Simulation trajectory analyses:

To understand the conformational changes and dynamic stability of both *NPSR1* and *EPB41L2* upon introducing the variants, we initiated MDS runs for the *NPSR1* protein structures and the *EPB41L2* protein structures using Gromacs v.2020. The following sections will report the simulation results for both systems.

3.9.1 *NPSR1*-GBQ protein structure complex:

RMSD of both the WT and MT (p.V138I) of the *NPSR1* protein systems was calculated to inspect the trajectory's convergence (Figure 3.5). Initially, a sudden rise in the RMSD was seen in the first few nanoseconds, resulting from the initial kinetic shock experienced by both systems at the start of the simulation. In the RMSD plot below, the WT reached a stable conformation from 50 ns till the end of the simulation, while the MT had two stable conformations at 50 ns – 75 ns, 100 ns – 140 ns, and 150 ns – 200 ns, respectively. For the simulation duration, the mean and stdev of the WT and MT trajectories were at close range and followed a similar trend with the MT (p.V138I), having a slightly higher average RMSD value of 0.64 ± 0.07 nm compared to the WT with 0.63 ± 0.08 nm. A similar trend was observed for the repeat simulations, with the RMSD values for the WT and MT (p.V138I) structures being 0.60 ± 0.07 nm and 0.63 ± 0.08 nm, respectively (Figure S19). When the RMSD of both the WT and MT (p.V138I) were compared, it was observed that they both achieved a stable conformation after 50 ns suggesting a stable dynamic behavior for both systems; hence, the RMSD trajectories for the last 50 ns were utilized for further analysis. Additionally, the GBQ ligand adopted two stable conformations based on RMSD analysis in both the WT and MT (p.V138I) systems (Figure S20).
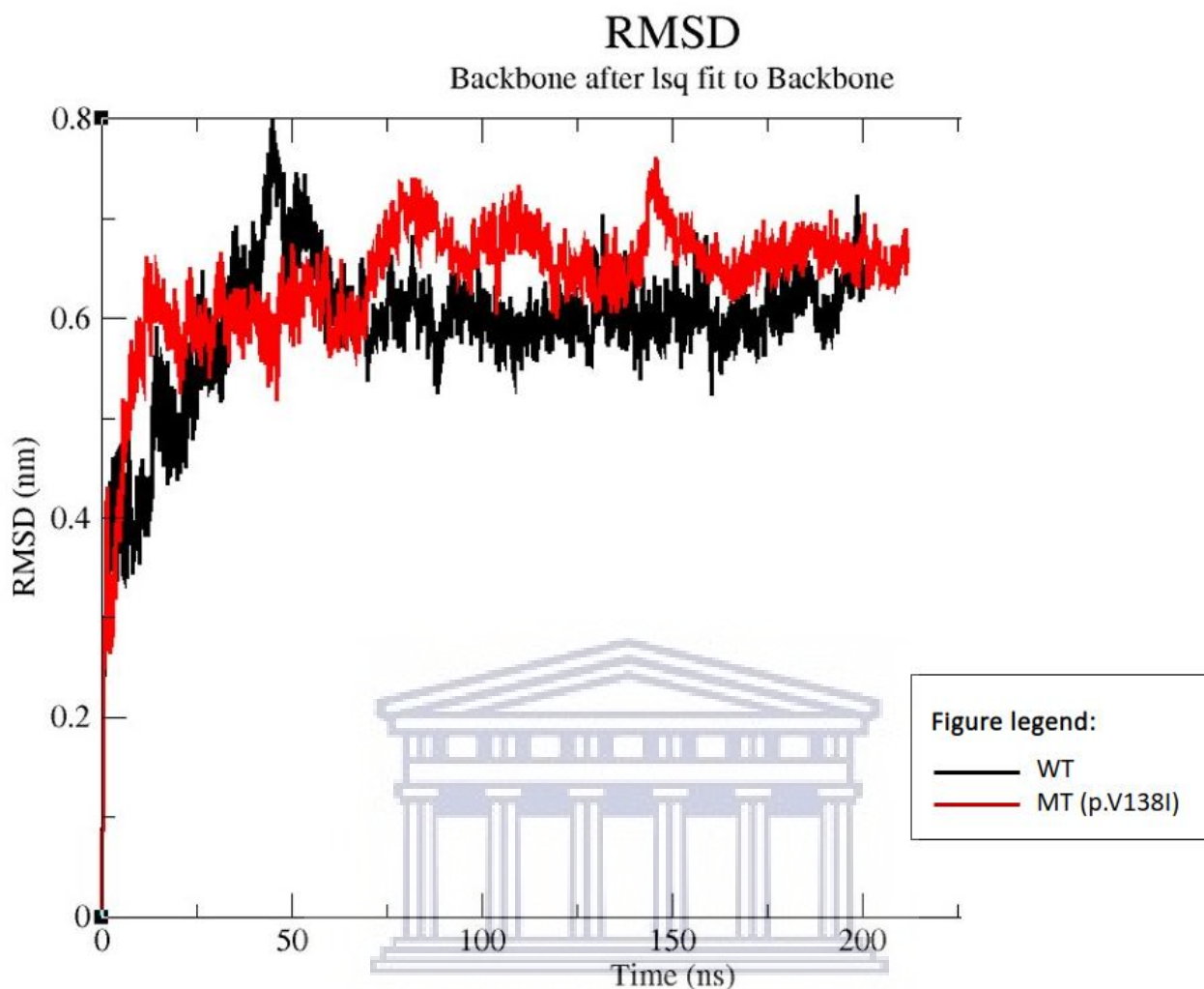
Figure 3.5: Backbone RMSD analysis of the WT *NPSR1* protein and p.V138I MT systems at 303 K over 200 ns of the simulation period.

The RMSF was calculated for the most stable part of the conformation (100 ns – 150 ns) and plotted to assess the AA flexibility for both the WT and MT (p.V138I) systems. In the RMSF plot below (Figure 3.6), the overall fluctuations of the p.V138I MT were found to be much higher in comparison to the WT. The mean and stdev RMSF values of the WT residues fluctuated between 0.17 ± 0.13 nm and the MT (p.V138I) system with an average of 0.2 ± 0.13 nm. RMSF analysis of the fluctuations showed that the MT (p.V138I) system had more flexible regions than the WT. Conformational flexibility was observed in the regions spanning residues 100-116, 138, 234-240, 251-260, and 296-306. Furthermore, the flexible regions span the active site residues ASP 105, TRP 108, ARG 109, PHE 109, and ASN 298 crucial for ligand binding, suggesting that the presence of the MT (p.V138I), which is found on the alpha-helices, impacts the overall dynamics and stability of the protein system.
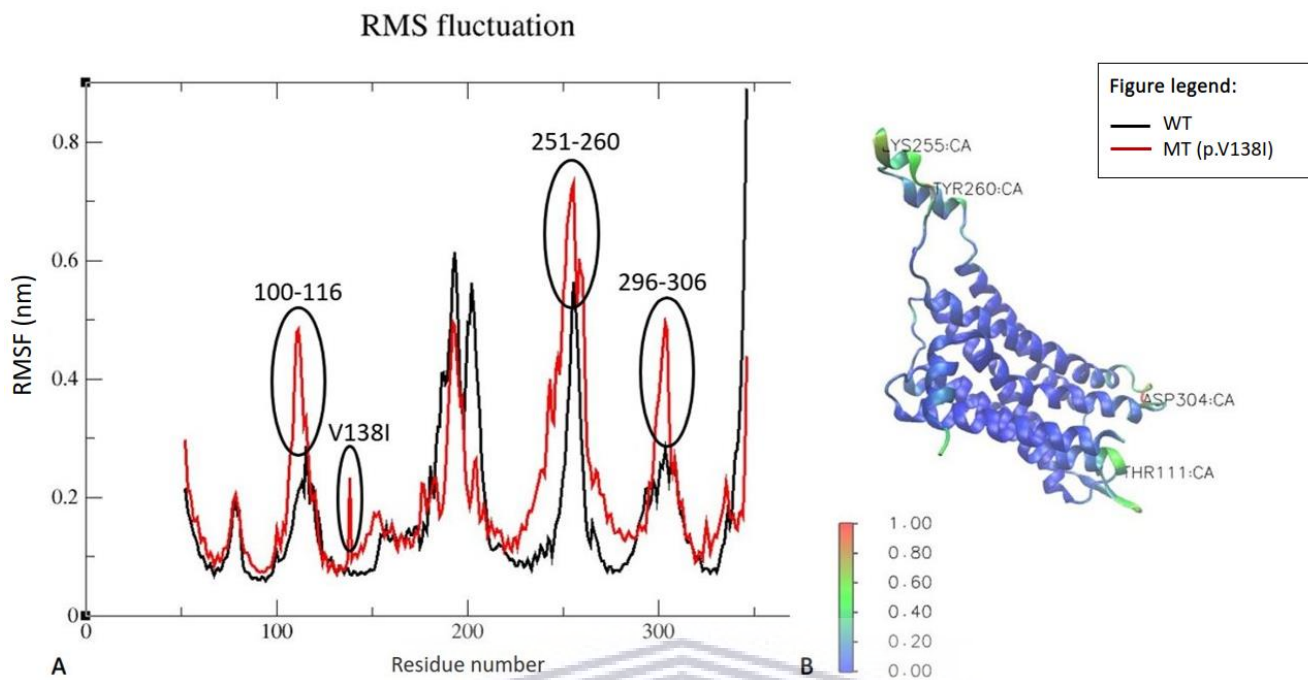
Figure 3.6: A) RMSF analysis of the protein residues. WT *NPSR1* system (shown in black) and the MT (p.V138I) system (shown in red) at 303 K. Black circles with residue numbers indicate regions of higher flexibility in the MT (p.V138I) protein system. B) Cartoon structure of the MT (p.V138I) protein structure colored according to RMSF with a value of 0 representing regions with low flexibility (blue) and 1 showing higher flexibility regions (green, yellow, and red). Labeled residues on cartoon structure correspond to regions of high flexibility.

To gain insight into the WT and MT (p.V138I) protein systems' stability and compactness, the radius of gyration (Rg) was computed, as shown in Figure 3.7. For the simulation duration, the MT (p.V138I) system had a higher mean and stdev of 2.27 ± 0.03 nm compared to the WT protein system, which had 2.23 ± 0.03 nm. The MT (p.V138I) protein system having a higher Rg value suggests a less compact protein structure than the WT. This agrees with the RMSF values, where the residues in the MT (p.V138I) protein system showed higher flexibility. Overall, these findings indicate that variant p.V138I impacts the stability and compactness of the *NPSR1* protein.

Radius of gyration (total and around axes)
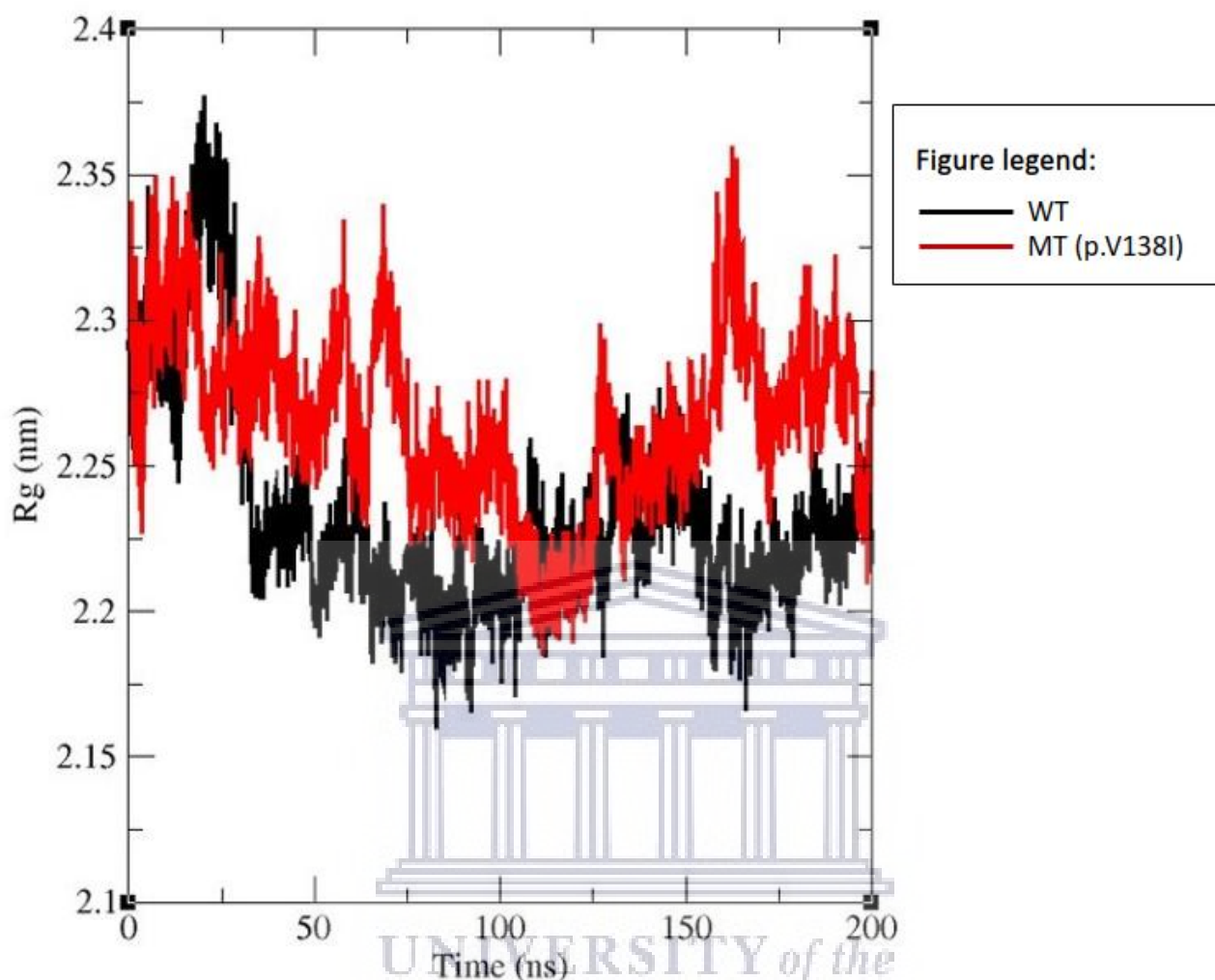
Figure 3.7: Rg for backbone atoms of the WT and MT (p.V138I) *NPSR1* **protein systems at 303 K are shown as a function of time.** The black line color represents the WT protein system, while the red line indicates the MT (p.V138I) protein system.

Hydrogen bonds (H-bonds) are essential interactions in protein-ligand complexes. They have many directional interactions underpinning protein folding, protein structure conformation and molecular recognition. H-bond analysis was performed to calculate the average number of H-bonds formed between the protein and GBQ ligand for both the WT and MT (p.V138I) protein systems. This was done for the most stable part of the simulation trajectory. The mean and stdev of the H-bonds formed between the *NPSR1* protein and the GBQ ligand were calculated to be 0.16 ± 0.365 nm and 0.03 ± 0.2 nm for the WT and MT (p.V138I) structures, respectively. In Figure 3.8 A and B, the H-bond bar

diagrams formed between the MT structure and GBQ ligand are sparse compared to the WT complex, which formed a higher number of H-bonds within the structure between 80 ns and 140 ns.

To understand hydrogen bond interactions, conformational snapshots of the *NPSR1*-GBQ complex during the last 50 ns of the simulation were extracted using Gromacs v.2020. The overall protein structure of the WT varied from the MT (p.V138I) in the loop regions as is shown in Figure S23. Furthermore, the GBQ ligand formed fewer contacts with surrounding residues because of the p.V138I variant on the protein structure (Table 3.5 and 3.6). The higher number of interactions seen for the WT NPSR1-GBQ agree with higher average number of H-bonds observed for the WT system compared to the MT (p.V138I) (Figures 3.8A and B and S21 and S22). Common interacting residues that could be exploited for further drug development include I181, P202, Y290, F299, and I315.

Table 3.5: Polar contacts formed between the GBQ ligand and surrounding residues in the *NPSR1* WT protein complex during the last 50 ns of the simulation.

Last 50 ns of simulation	Polar contacts and residues involved in the interactions
160 ns	3.0 Å (M211), 3.0 Å (Y290), 3.9 Å (I315).
170 ns	3.6 Å (L200), 3.2 Å (W207), 2.4 Å (M211), 3.7 Å (I315).
180 ns	2.6 Å (P202), 3.3 Å (I315).
190 ns	3.2 Å (P202), 3.6 Å (W201), 3.9 Å (S312), 2.9 Å (Q316).
200 ns	2.9 Å (L296)

Abbreviations: W= tryptophan, L= leucine, M= methionine, I= isoleucine, P= proline, F= phenylalanine, V= valine, R= arginine, E= glutamic acid/ glutamate, S= serine, Q= glutamine, T= threonine, Y= tyrosine, C= cysteine, A= alanine.

Table 3.6: Polar contacts formed between the GBQ ligand and surrounding residues in the *NPSR1* **MT (p.V138I) protein complex during the last 50 ns of the simulation.**

Last 50 ns of simulation	Polar contacts and residues involved in the interactions
160 ns	3.6 Å (T181), 4.0 Å (Y210), 3.3 Å (I315).
170 ns	3.9 Å (T181), 3.0 Å (V214).
180 ns	3.2 Å (Y210), 3.5 Å (P299).
190 ns	3.4 Å (Y129), 2.7 Å (L132), 3.0 Å (Y290), 2.7 Å (F299).
200 ns	3.0 Å (F299)

Abbreviations: W= tryptophan, L= leucine, M= methionine, I= isoleucine, P= proline, F= phenylalanine, V= valine, R= arginine, E= glutamic acid/ glutamate, S= serine, Q= glutamine, T= threonine, Y= tyrosine, C= cysteine, A= alanine.
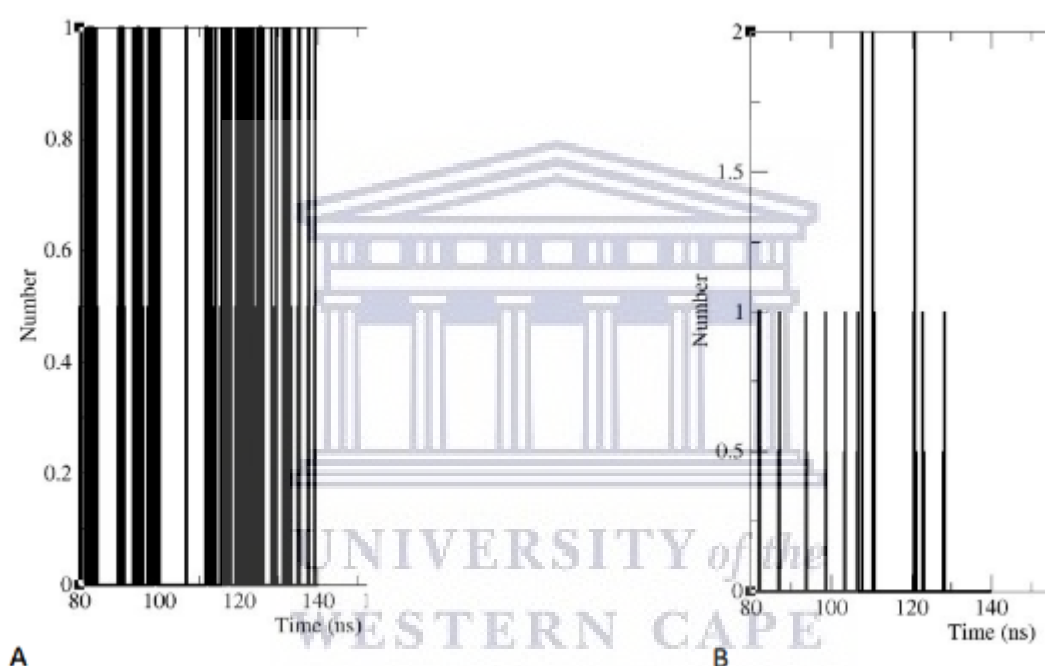
Figure 3.8: Average intramolecular protein-drug HBs among the WT, MT *NPSR1* **protein complex, and the GBQ ligand at 303 K for the most stable part of the simulations (80 ns-140 ns).** A) Average number of HBs formed between WT and GBQ. B) Average number of HBs formed between MT and GBQ.

The non-bonded interaction energy between the *NPSR1* protein and the GBQ ligand was calculated to assess the strength of the interaction. For the WT and MT (p.V138I) *NPSR1*-GBQ complex systems, the total pairwise non-bonded interaction energy between the *NPSR1* protein and GBQ ligand was computed for the simulation period 100 ns - 150 ns. Our study included the short-range Coulombic electrostatic interaction energy term and the short-range Lennard-Jones energy term, yielding the total non-bonded pairwise interaction energy as the sum of the two energy components

evaluated. Entropy change is not considered in the calculation. The overall non-bonded pairwise interaction energies between the *NPSR1* protein and GBQ ligand were greater for the MT (-146.4198 kJ/mol) than for the WT structure with it having a non-bonded pairwise interaction energy of -123.3012 kJ/mol.

Table 3.7: Non-bonded interaction energy between the WT and MT (p.V138I) NPSR1-GBQ complexes.

Energy (kJ/Mol)	WT	MT (p.V138I)
Coul-SR	-6.48424	-27.2028
LJ-SR	-116.817	-119.217
Total interaction energy	-123.3012	-146.4198

Abbreviations: Coul-SR: short-range Coulombic electrostatic interaction energy, LJ-SR: short-range Lennard-Jones interaction energy.

SASA was calculated to quantify solvent exposure for the hydrophobic and hydrophilic residues. In Figure 3.9, the p.V138I variant protein structure has a higher overall total SASA value than the WT protein structure, implying more buried residues are becoming solvent exposed. For the simulation duration, the MT (p.V138I) system had a slightly higher mean and stdev of 165.89 ± 2.08 nm^2 compared to the WT protein system, which had 163.39 ± 1.71 nm^2. This exposed the buried residues in the *NPSR1* protein structure to the solvent, raising the p.V138I variant's solvent accessible surface area.
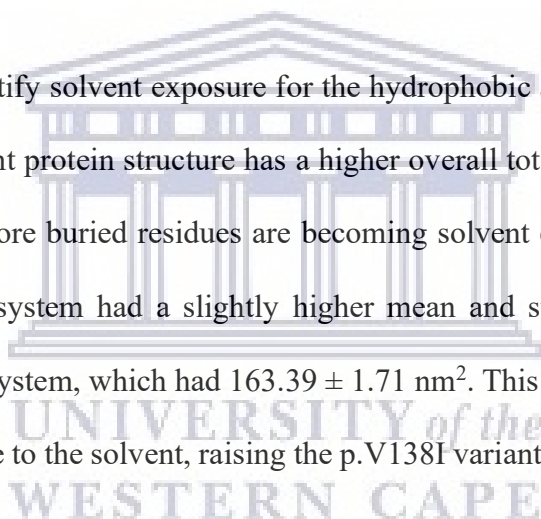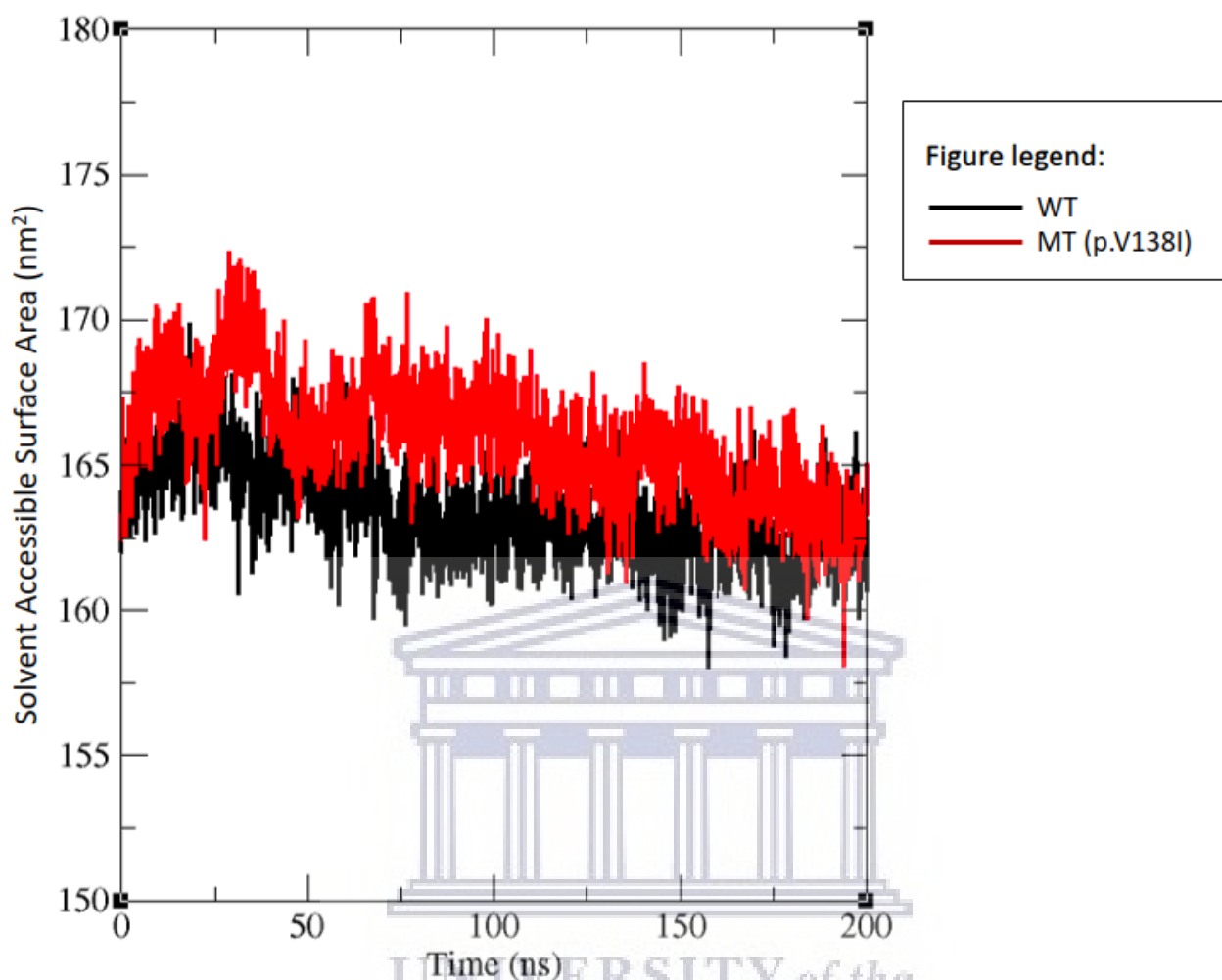
Figure 3.9: SASA analysis for the backbone atoms of the WT and MT (p.V138I) *NPSR1* protein systems at 303 K are shown as a function of time. The black line color represents the WT protein system, while the red line indicates the MT (p.V138I) protein system.

Calculation of the contribution of each of the top ten principal components (PCs) indicated that the first two PCs (PC1 and PC2) contributed significantly to the movement of the protein. For the WT system, PC1 contributed 52 %, while PC2 contributed 26 %. In the MT (p.V138I) system, PC1 contributed 51 % while PC2 contributed 20 %. Therefore, 2D projections of PC1 and PC2 for both systems were plotted and shown in Figure 3.10. The scatterplot observed that the MT (p.V138I) had more randomized motions and occupied a larger phase space than the WT, resulting in a less stable variant structure.
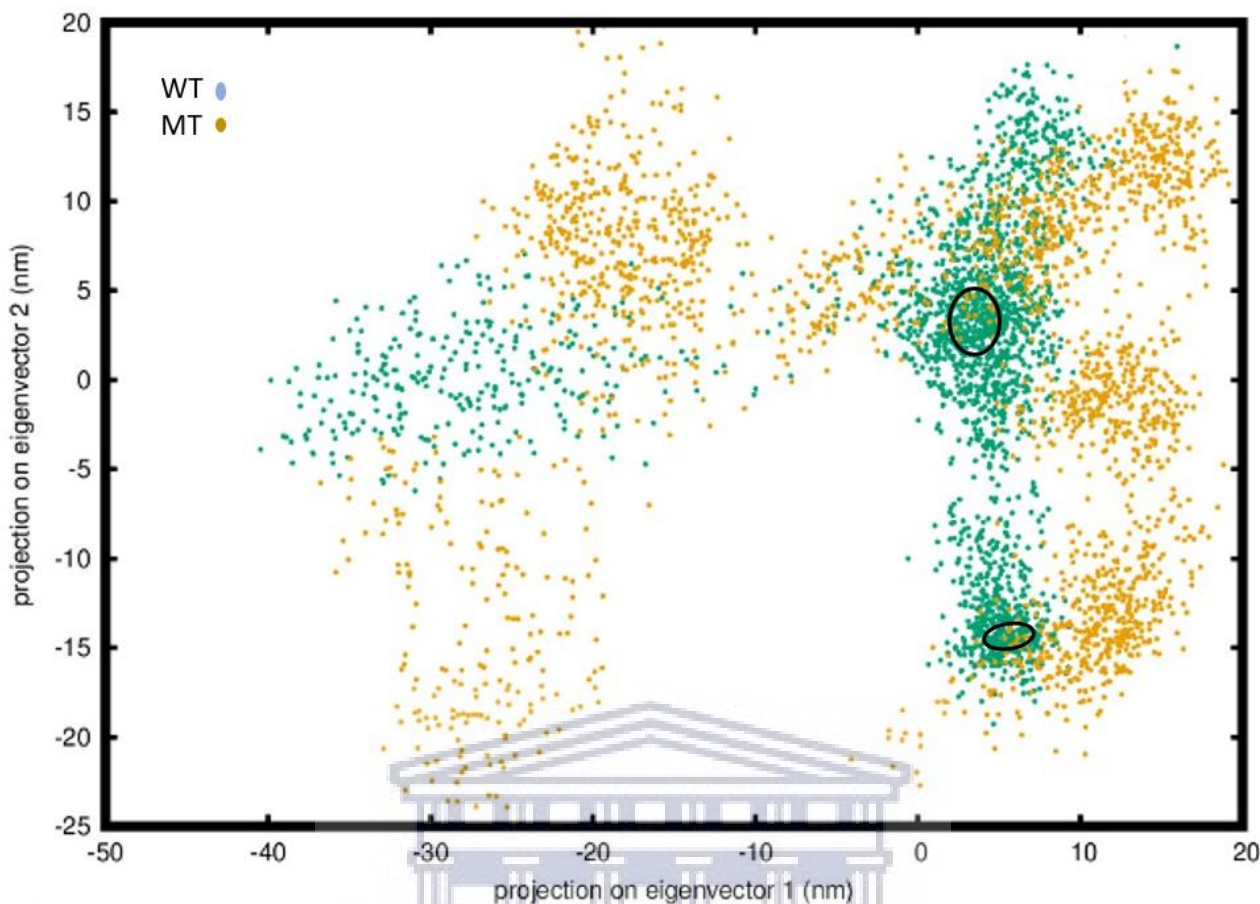
Figure 3.10: 2D Projections of the first two PCs during the stable 100-150 ns simulation period for both the WT and MT (p.V138I) protein systems. WT represented in turquoise, and the MT (p.V138I) shown in orange. Distinct cluster observed circled in black. Circles represent stable clusters for the WT.

3.9.2 *EPB41L2* protein structure:

The RMSD of both the WT and MT (p.R663C) *EPB41L2* protein systems were calculated to analyze trajectory convergence and equilibrium (Figure 3.11). Initially, a sharp increase in the RMSD was observed in the first few nanoseconds, which could be attributed to both systems' initial kinetic shock during the simulation phase. The WT and MT (p.R663C) protein systems reached equilibrium after 75 ns based on the RMSD plot shown below (Figure 3.11). The mean and standard RMSD deviation values for the MT (p.R663) system were higher at 1.74 ± 0.18 nm than that of the WT system at 1.63 ± 0.24 nm. When the RMSD of the WT and MT (p.R663C) trajectories were compared, it was found that both simulation trajectories achieved equilibrium, implying a stable dynamic behavior of both systems; thus, the last 100 ns of the two trajectories were used for further analysis.
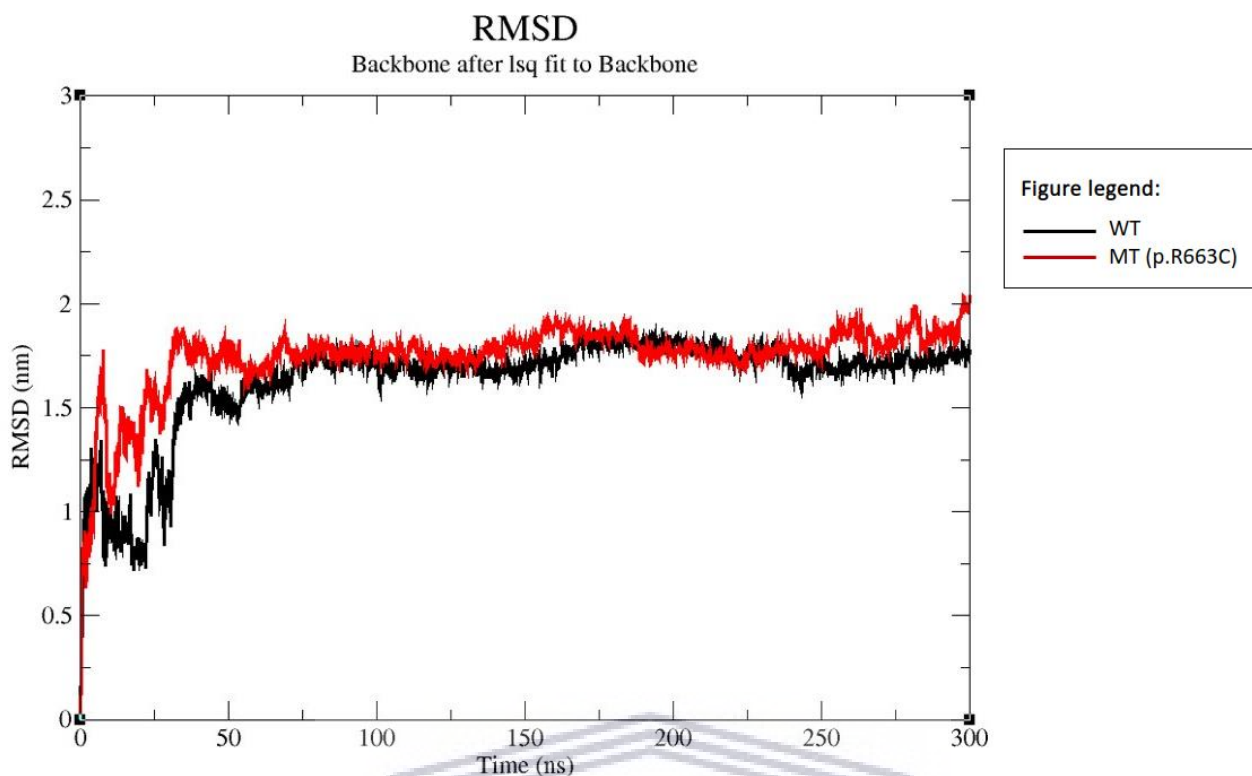
Figure 3.11: Backbone RMSD analysis of the WT *EPB41L2* protein and p.R663C MT systems at 303 K over 300 ns of the simulation period.

Protein dynamics, structure, and function are all intrinsically linked. The complex nature of the *EPB41L2* protein and the impact of the p.R663C variant on the protein structure can be revealed by studying the internal atomistic motions of protein systems. The RMSF for the most stable part of the conformation (100 ns – 300 ns) was calculated and plotted to assess the AA flexibility for both the WT and MT (p.R663C) systems. In the RMSF plot below (Figure 3.12), the overall fluctuation of the p.R663C MT was found to be much higher in comparison to the WT. The mean and stdev RMSF values of the WT residues fluctuated between 0.21 ± 0.11 nm and the MT (p.R663C) system with an average of 0.28 ± 0.15 nm. RMSF analysis of the fluctuations showed that the MT (p.R663C) system had higher flexibility than the WT. Conformational flexibility was observed in the region spanning residues 699-717 and corresponded to a loop region in the MT (p.R663C). In the WT protein system, residues 660-673 (Figure S21) which correspond to alpha-helices, showed increased flexibility compared to the MT. Moreover, the region spanning the variant p.R663C showed lower flexibility than the WT. The reduced flexibility of the 660-673 region is important for regulating access to the catalytic site and controlling the spectrin-actin-binding domain of the *EPB41L2* membrane protein.
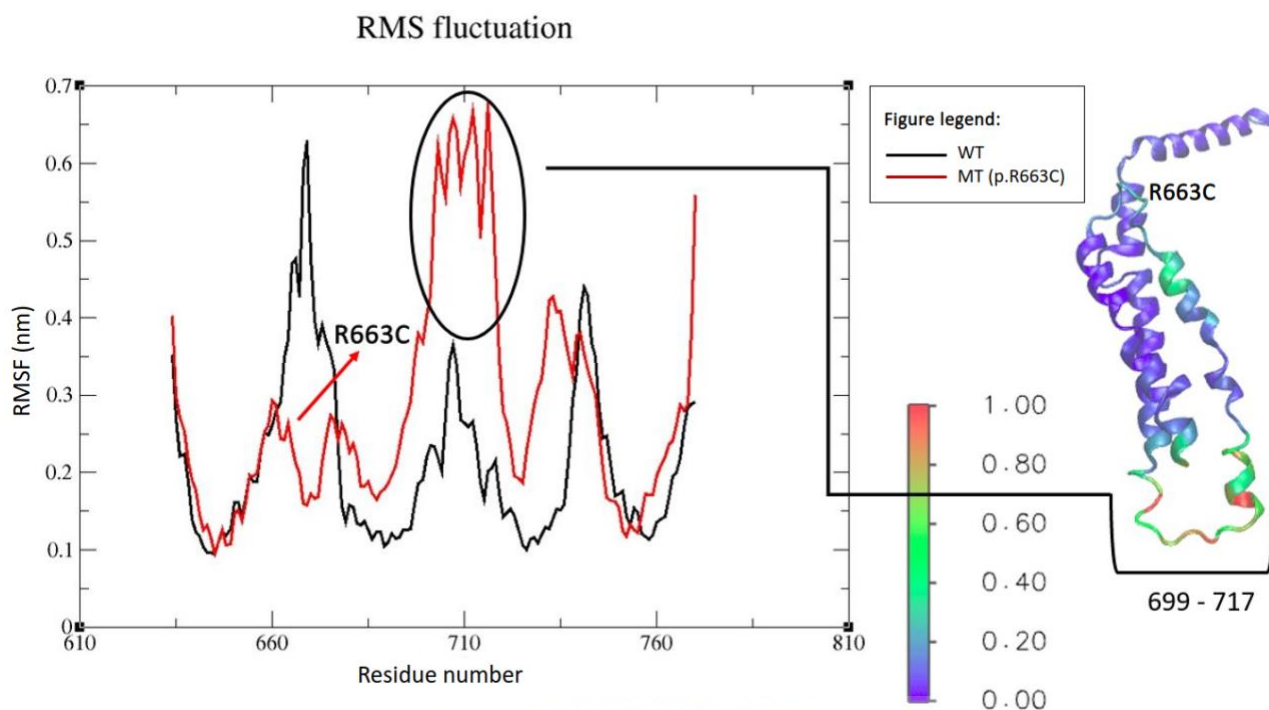
Figure 3.12A: RMSF analysis of the protein residues of WT *EPB41L2* protein (shown in black) and the MT (p.R663C)) system (shown in red) at 303 K. Black circles with residue numbers indicate regions of higher flexibility in the MT (p.R663C) protein system. Cartoon structure of the p.R663C MT protein structure colored according to RMSF flexibility with a value of 0 representing regions with low flexibility (blue) and 1 showing regions of high flexibility (green and red). The red arrow indicates the position of the p.R663C variant.
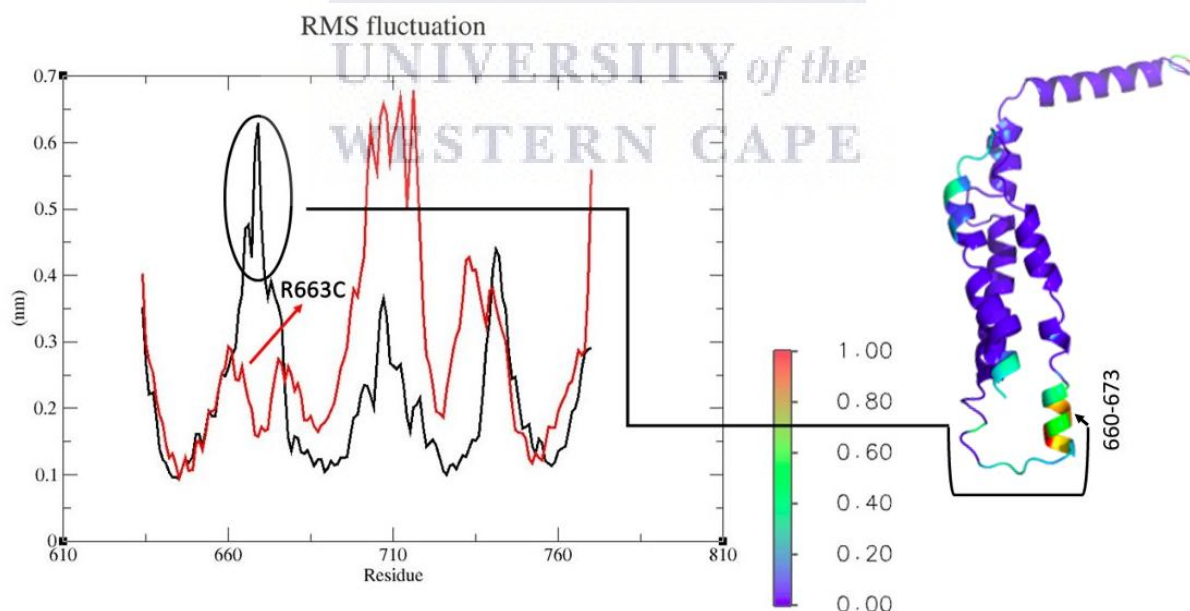
Figure 3.12B: RMSF analysis of the protein residues of WT *EPB41L2* protein (shown in black) and the MT (p.R663C)) system (shown in red) at 303 K. Black circles with residue numbers indicate regions of higher flexibility in the WT protein system. Cartoon structure of the WT protein structure colored according to RMSF flexibility with a value of 0 representing regions with low flexibility (blue) and 1 showing regions of high flexibility (green, yellow and red). The red arrow indicates the position of the p.R663C variant.

Next, the Rg plot (Figure 3.13) measures the change in the center of mass of the spectrin-actin-binding region of the *EPB41L2* protein structure before and after introducing the p.R663C variant. For the duration of the simulation (300 ns), the Rg value of the p.R663C variant was higher than the WT system. Significant fluctuations were observed in both protein systems in the first 100 ns. The MT (p.R663C) had a mean and stdev Rg values of 1.93 ± 0.11 nm, while the WT protein system had 1.81 ± 0.16 nm. The MT (p.R663C) protein system's Rg value is higher than the WT, indicating that the MT protein structure is less compact than the WT protein structure.
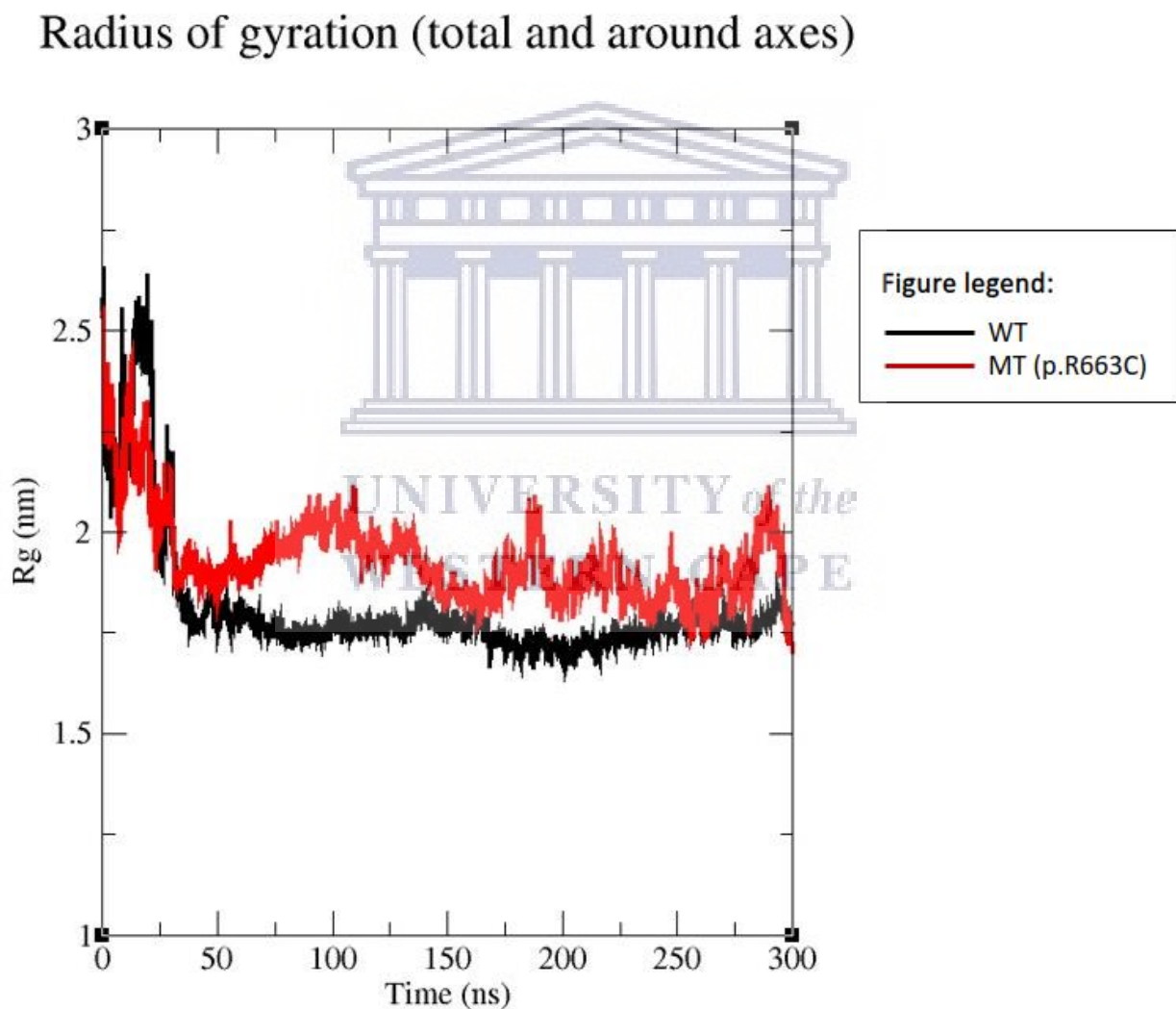
Figure 3.13: Rg for backbone atoms of the WT and MT (p.R663C) *EPB41L2* protein systems at 303 K are shown as a function of time. The black line color represents the WT protein system, while the red line indicates the MT (p.R663C) protein system.

69

The total SASA measures the conformational change in solvent exposure for hydrophobic and hydrophilic residues. The wild-type protein structure displayed lower overall SASA values as compared to the p.R663C variant structure, each having mean and stdev values of 77.9 ± 1.5 nm^2 and 82.08 ± 1.47 nm^2, respectively. This suggests that more buried residues became solvent exposed in the spectrin-actin binding domain for the p.R663C variant compared to the WT (Figure 3.14).
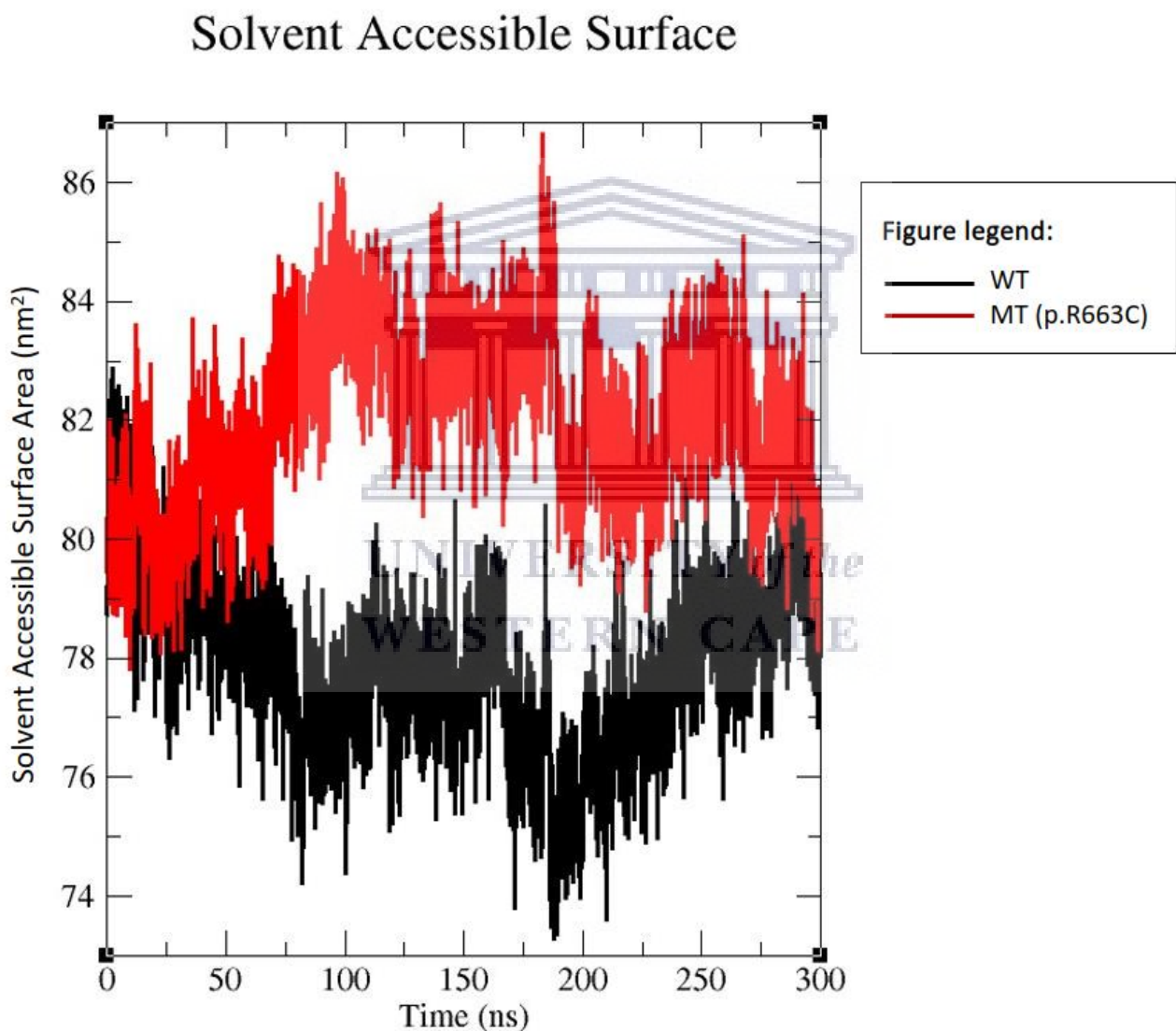
Figure 3.14: SASA analysis for the backbone atoms of the WT and MT (p.R663C) *EPB41L2* **protein systems at 303 K are shown as a function of time.** The black line color represents the WT protein system, while the red line indicates the MT (p.R663C) protein system.

The conformational motion of both systems was projected along the first two principal components (PC1 and PC2) directions to see if any differences were observed in protein motion between the WT

and MT (p.R663C) protein systems. PC1 contributed 56 %, while PC2 contributed 23 % for the WT protein system. In the MT (p.R663C) protein system, PC1 contributed 61 % while PC2 contributed 16 %. From the scatterplot, it is observed that the MT (p.R663C) clusters were more randomized and occupied a larger region of phase space as compared to the WT. However, we observed one distinct cluster circled in black for the MT and three for the WT system (Figure 3.15). These clusters represent the specific conformation adopted by the WT protein system when the protein is most stable over the simulation duration.
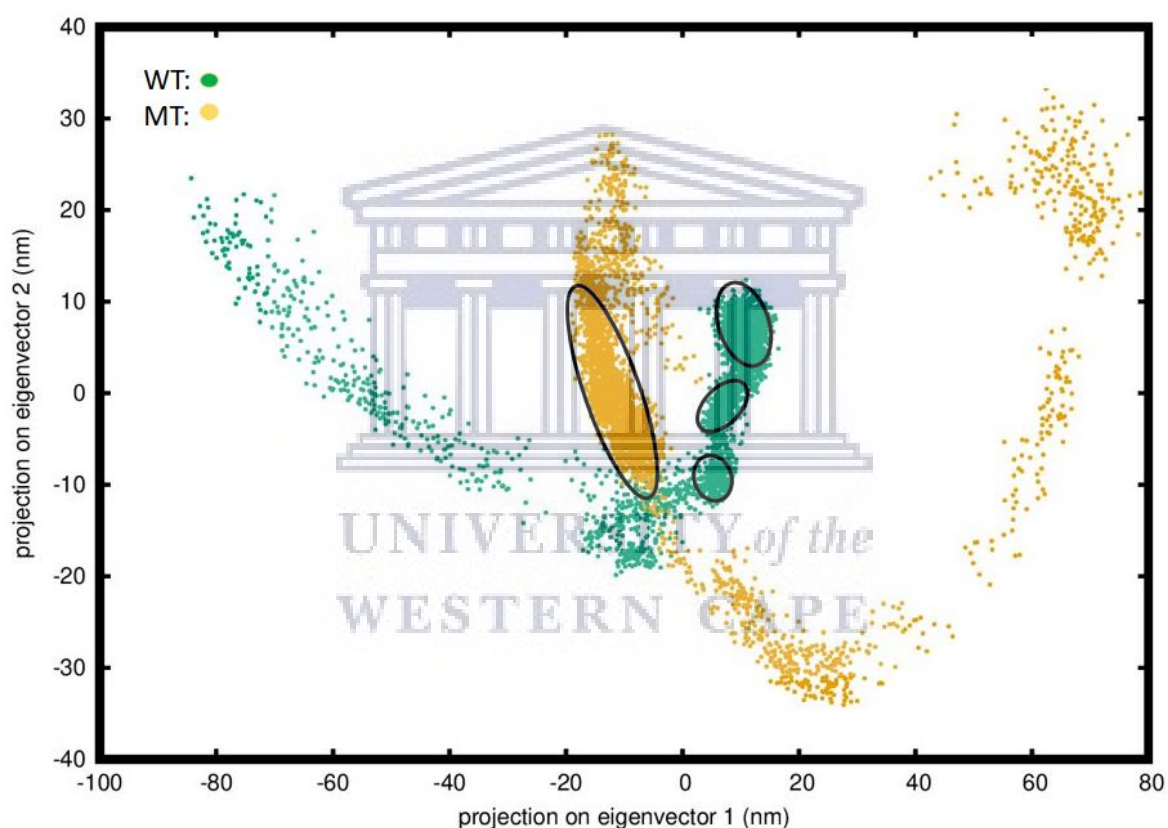
Figure 3.15: 2D Projection of Eigenvector 1 and 2 during the 300 ns simulation period for both the WT and MT (p.R663C) *EPB41L2* protein systems. The WT represented in turquoise, and the MT (p.R663C) shown in orange along with the first two principal components (PC1 and PC2). Circles represent stable clusters for the WT (3) and MT (1).

In summary, our findings revealed that both *NPSR1* (p.V138I) and *EPB41L2* (p.R663C) were attractive PD candidate genes based on several prioritization criteria's and were therefore, selected for further structural analyses. Moreover, pathway and expression data analysis revealed that both *NPSR1* and *EPB41L2* were involved in signaling and cellular processes. Furthermore, the structural

quality assessment indicated high reliability of the 3D protein structures, specifically conserved domains with the *NPSR1* protein model and the *EPB41L2* protein model each satisfying QMEAN, GMQE, ERRAT, Procheck and ERRAT quality analyses. Additionally, MD simulation results showed that both p.V138I and p.R663C variants had destabilizing effects on the protein structures of *NPSR1* and *EPB41L2* based on the stability predictions, trajectory parameter values that included RMSD, RMSF, Rg, SASA, and PCA analysis. Similarly, H-bond analysis confirmed fewer hydrogen bonds being formed between NPSR1 and ligand GBQ due to the presence of mutation p.V138I. The results visualized in the simulation movies are in agreement with the RMSD, RMSF, Rg, SASA, and PCA analyses of all four protein systems (https://drive.google.com/drive/folders/15h3iv0yRGtb04kdaZxl4xbrqV1ztqVJE?usp=sharing). In the *NPSR1*-GBQ protein systems, the MT (p.V138I) had more flexible loop regions compared to the WT. Overall, the protein structure adopted a less compact fold upon introduction of the p.V138I variant. In the actin-spectrin-binding domain of the *EPB41L2* protein system, the beta sheet connector and alpha helix region is very flexible spanning residues (699-717) going from an extended to flexed conformation in the p.R663C variant system. Future drug development should target this flexible region of the variant protein p.R663C spanning residues (699-717) to restore stability to the protein structure.
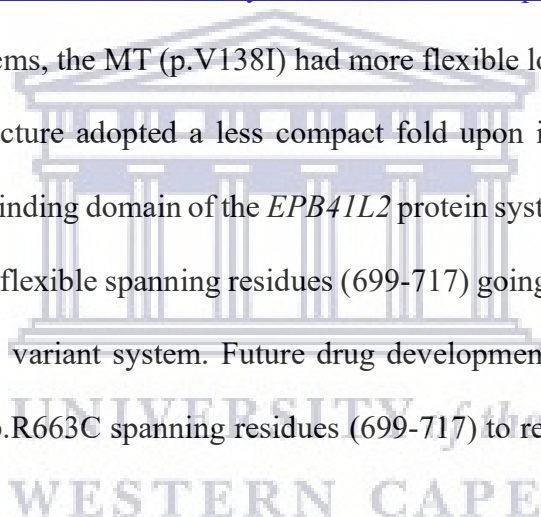
CHAPTER IV: Discussion and Conclusion

4.1 Discussion

In this study, the genetic etiology of PD in a SA family of Indian ancestry was investigated using bioinformatic approaches that included data mining, pathway, expression and STRING analysis. Thereafter, *in-silico* structural modeling methods were used to interrogate the effect of variants on the protein structure of the most likely PD associated genes. Interestingly, the 11 variants shared by the three affected family members reported by Dr. Roopnarain (unpublished data) are not reported in literature. Further inspection of the 11 variants identified in the multiplex family revealed that they are novel variants.

As estimated in 2016, India is home to nearly half a million people living with PD, with an expected significant increase in prevalence in the near future (Rajan et al., 2020). Mutations in genes such as *DJ-1*, *SNCA*, *VPS35*, and *PINK-1* have been shown to not significantly contribute to PD amongst individuals of Indian ancestry (Biswas et al., 2006; Nagar et al., 2001; Punia et al., 2006; Sudhaman et al., 2013). This strengthens the speculation that the genetic causes of PD in this population may differ from other populations, necessitating further research in this area.

As mentioned in the literature review, studies inspecting the genetic causes of PD have shown that the genetic elements contributing to PD may vary in specific populations (Rajan et al., 2020). The absence of previously identified PD mutations in this multiplex family led to the conclusion that other novel mutations could be responsible for the disease phenotype. From the 11 variants identified in 11 genes (*ELK4*, *CEP170*, *EEFSEC*, *TRIM59*, *HTT*, *ELMOD2*, *EPB41L2*, *NUP153*, *NPSR1*, *LRRK1*, and *MAN2B1*), candidate genes that fulfilled the prioritization criteria were selected for further studies. Factors such as genes involved in protein-protein interactions with PD-associated genes, genes highly expressed in brain tissue and involved in PD-related pathways were used as the criteria

for selecting the candidates. The *LRRK1* p.T871M, *NPSR1* p.V138I and *EPB41L2* p.R663C variants met the above filtration criteria and were chosen as top candidates to investigate the effect of the variants on the protein structure. Unfortunately, due to the lack of a solved homologous template structure that covered the p.T871M variant position located in *LRRK1,* structural modelling was not pursued. However, recently, a new homologous template (7LI3) (https://www.rcsb.org/structure/7LI3) was solved which can be used in future studies to predict the 3D structure of LRRK1 that encompasses the p.T871M region.

The 3D structures predicted for *NPSR1* and the spectrin-actin-binding domain of *EPB41L2* were found to be reliable approximations of the native protein structure, based on the satisfaction of several quality parameters such as PROCHECK and ERRAT. However, we recognize the predicted models' medium accuracy due to low sequence identity between the target sequence and the template and propose that these protein models be interpreted with caution. However, we are confident in the modeling of specific domains within the protein structures due to sequence conservation and similar structural fold to the homologous template structures.

Variants are significant in evolution as they introduce diversity into genomes. Tang et al. (2016) reported that non-synonymous variants, which are also referred to as missense mutations, may have deleterious or advantageous effects on an organism by affecting the protein's stability and function. Therefore, structural-based methods are essential as they predict the impact of amino acid substitution on protein stability (Tang & Thomas, 2016). Examples of tools that can predict a change in phenotype include SIFT, PolyPhen, and the stability prediction tool DUET. SIFT and Polyphen consider the composition of amino acids and predict whether a substitution is deleterious. However, both these tools rely on sequence data and usually provide contradictory results (Pereira et al., 2020; Sebate et al., 2021). Therefore, we utilized structure-based methods only to calculate the protein's free energy change using the DUET web server. DUET inspects the change in Gibbs free-energy differences

($\Delta\Delta G$) between folded and unfolded protein states upon introduction of the variant. DUET results showed that the variants introduced into the protein structure resulted in slightly destabilizing effects of the p.V138I (ΔG = -0.324 Kcal/mol) and p.R663C (ΔG = -0.102 Kcal/mol) variants on the *NPSR1* and *EPB41L2* protein structures, respectively. Regardless of the magnitude of the destabilizing effects being relatively small (<0.5 Kcal/mol), we speculated that the variants can still affect the stability and folding of the protein, as shown in other studies (Anderson & Daggett, 2008; Sebate et al., 2021). However, stability predictions do not take protein dynamics into account after the introduction of a variant. Therefore, more in-depth analyses such as the change in protein dynamics and per residue fluctuation are essential for understanding protein structure change after introducing variants.

To understand the impact of the variants on the protein structure and dynamics, MD simulations of the *NPSR1* and *EPB41L2* protein structures were performed. MD analyses effectively detect subtle changes in protein structure (Nair & Miners, 2014). Trajectory analyses of the *NPSR1* and *EPB41L2* protein structures verified the DUET stability prediction results. Variants p.V138I and p.R663C affected the stability of the *NPSR1* and *EPB41L2* protein structures based on RMSD, RMSF, Rg, SASA values, H-bond, and PCA analysis (Figures 3.5 – 3.15). From the analyses, the current study found that the p.V138I variant impacted the flexibility and compactness of the *NPSR1* protein structure. Overall fluctuations of the p.V138I variant were higher in various regions, including the regions spanning residues 100-116, 138, 234-240, 251-260, and 296-306. Apart from residue 138, representing the variant position, all other regions represent loop regions and important catalytic binding sites. The SASA analyses indicated that more residues in the NPSR*1* p.V138I MT system became solvent-exposed, suggesting that the MT is less stable than the WT structure. Furthermore, fewer inter-molecular H-bonds between the MT (p.V138I) protein structure and its ligand, GBQ was formed. The lower number of H-bonds indicate weaker interaction between the MT (p.V138I) and GBQ.

Similarly, the PCA result was in agreement with RMSF, Rg, SASA, and H-bond analysis. However, the non-bonded interaction analysis showed contradictory results to the H-bond analysis with the MT showing stronger interaction energy than the WT. Correspondingly, the change from valine to isoleucine resulted in perturbations in the hydrophobic core, inducing changes in the *NPSR1* proteins' overall 3D structure. This observation has not been reported in previous MD studies investigating mutations in *NPSR1* protein. A study conducted by Neufang *et al.* (2015) which looked at a polymorphism occurring at residue 107 (p.N107I) located in the *NPSR1* protein, reported that this substitution caused a 10-fold increase in *NPSR1* expression and NPS efficacy at the receptor. It also resulted in increased activation of the prefrontal cortex and locus coeruleus, the latter of which has been reported to be involved in the generation of tremors in PD (Isaias et al., 2012). These findings are in line with previous studies that report that variants in the NPS/*NPSR1* protein have been associated with psychiatric illnesses such as anxiety, panic disorders, and disorders related to alcohol misuse (Ghazal, 2016; Neufang et al., 2015).

As mentioned in the literature review, *NPSR1* is a G protein-coupled receptor (GPCR) family member and encodes a membrane protein that acts as a receptor for a neuropeptide known as Neuropeptide S (NPS) (Pietras et al., 2011). Neuropeptides are small protein-like molecules produced and secreted primarily by the peripheral and central nervous system neurons. The majority of neuropeptides exert their action by binding to GPCRs. GPCRs are the most prominent family of cell surface receptors in the human genome and are used as therapeutic targets for PD (Lemos et al., 2018). Signals from peptide-GPCRs are associated with specific brain functions such as stress and anxiety, memory and learning, glucose metabolism, sleep and wakefulness, and neuroprotection (Zheng et al., 2021). Some neuropeptides have played neuroprotective roles in PD in recent studies, including the *NPY*, ghrelin, substance P, and neurotensin, as shown in *in vitro* and *in vivo* studies (Bayliss et al., 2016; Li et al., 2019; Maasz et al., 2017; Shi et al., 2017; Wang et al., 2015).

Strikingly, *NPSR1* is involved in the up regulation of the release of sequestered Ca^{2+} ions into the cytosol and affects various cellular processes through its signaling. When activated by its cognate ligand NPS, *NPSR1* causes intracellular calcium ions to mobilize and stimulate cyclic adenosine monophosphate (cAMP) synthesis, which results in increased cellular excitability (Reinscheid et al., 2005; Xu et al., 2004). Calcium (Ca^{2+}) channels are found throughout the brain and regulate neurotransmitter release, primarily at presynaptic terminals (Zaichick et al., 2017). Since Ca^{2+} signaling has such a broad impact on neural biology, Ca^{2+} levels must constantly be regulated to prevent uncontrolled responses that could lead to diseases such as PD. We speculate that the novel variant in *NPSR1* could have deleterious effects on the regulation of Ca^{2+} levels resulting in PD symptoms.

Similarly, *EPB41L2* is a highly expressed gene in the thalamus, midbrain, and basal ganglia and is also involved in signaling and cellular processes (Table 3.2). This gene plays a role in the dopamine receptor-mediated signaling pathway and the nicotine pharmacodynamics pathway. The dopamine (DA) receptor-mediated pathway is essential as alterations in DA receptors and DA signaling pathways are associated with PD (Juárez Olguín et al., 2015; Tozzi et al., 2018). Since DA is a signaling molecule in the DA receptor-mediated signaling pathway, maintaining its homeostasis is crucial as DA plays a vital role in motor coordination and memory (Saikia, 2018). Any changes occurring in the proteins/molecules involved in this pathway could cause an imbalance in the DA homeostasis of an individual's body. The substitution occurring at residue 663 from arginine to cysteine (p.R663C) in the *EPB41L2* protein affected the protein structure and dynamics. The substitution from arginine, a polar, positively charged amino acid, to cysteine, an uncharged polar amino acid, results in a change in charge that could affect protein stability. A study conducted by Chakkalakal (2018) reports that point mutations involving a change from arginine to cysteine result in a broad spectrum of unusual phenotypes depending on the protein backbone on which the change occurs (Chakkalakal et al., 2018).

From the MD results, the variant p.R663C located on the spectrin-actin-binding domain of the *EPB41L2* protein structure recorded higher fluctuations and decreased compactness compared to the WT structure. Fluctuations in protein residues emerge as a critical factor in deciding biological activity, indicating that protein functional sites are uniquely coupled with structural fluctuations (Agrahari et al., 2019). RMSF analysis demonstrated overall higher fluctuations of the *EPB41L2* p.R663C variant structure with maximum fluctuations observed at the loop region. This region regulates the morphological and functional dynamics of the protein (Nestor et al., 2011). Furthermore, the p.R663C variant structure exhibited higher Rg values than the WT, illustrating a less tightly packed and less stable protein structure. SASA results agreed with the RMSF and Rg analysis of all parameters suggesting that the *EPB41L2* p.R663C MT affects the stability of the protein structure, making it more flexible and less compact, leading to a thermodynamically unstable system. From the PCA analysis, it is observed that the p.R663C variant structure covers a larger subspace with less stable cluster conformations in comparison to the WT, suggesting it has a less stable conformation. It is speculated that the resulting altered MT (p.R663C) structure may lead to the increased flexibility of the loop region that regulates calcium transportation. This may implicate that the dopamine receptor-mediated signaling pathway is involved in the development of PD symptoms.
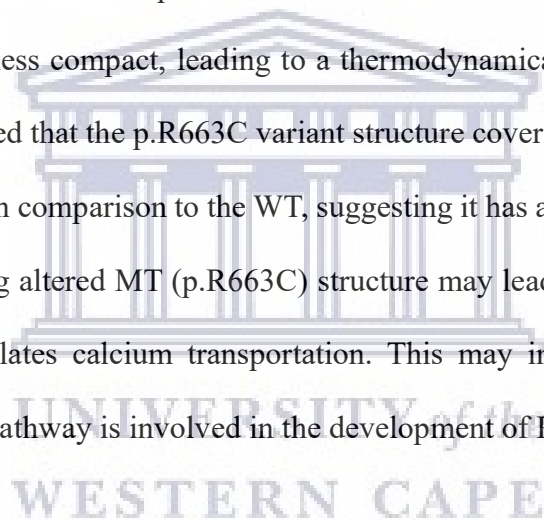
In summary, a schematic illustration of the findings of this study and their implications in PD are shown in Figure 4.1 below.
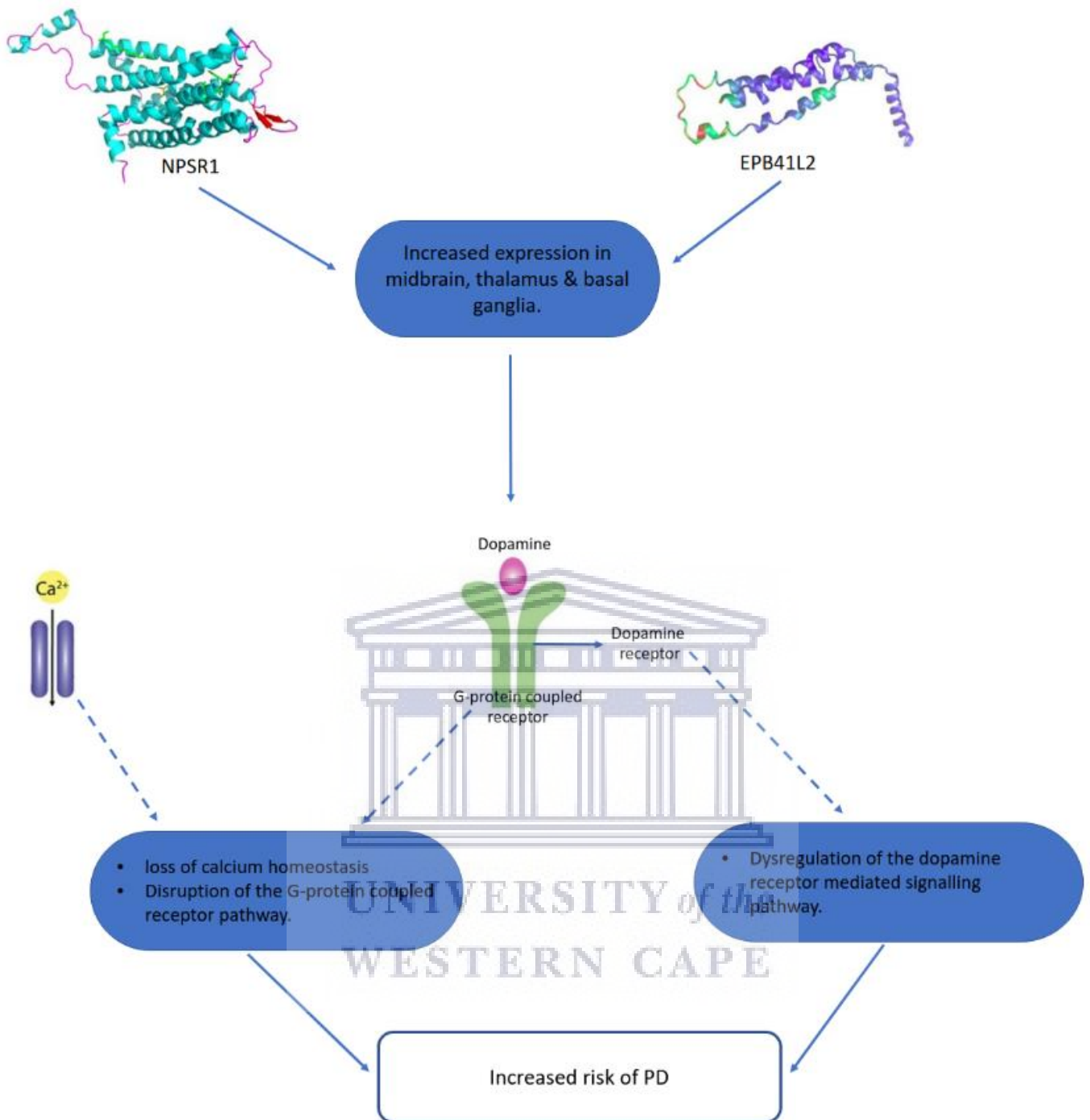
Figure 4.1: Schematic representation of the findings of this study. Dashed arrows indicate a disruption of the biological process of interest.

79

4.2 Limitations of study and future work

The initial study only whole-exome sequenced three family members, a small sample size. However, the study was not dependent upon a large sample size as this was a familial study. Future studies should screen more unrelated families with PD to determine if these variants are unique within this South African family. The genetic diversity of PD continues to complicate our understanding of PD. This is an evolving epidemic (Dorsey et al., 2018); it is vital to keep identifying many causal variants to be one step closer to achieving personalized treatments for PD patients. Another limitation of this study was the use of protein models of moderate accuracy. To overcome this limitation, subsequent modeling studies should consider using templates with higher resolution and higher sequence identity to the target sequence when they become available.

Additionally, short MD simulations were ran for this study (200 ns for *NPSR1* and 300 ns for *EPB41L2*). Lengthy simulations may be required to capture large and subtle changes of the protein systems correctly. Another limitation of this study was the exclusion of Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) free energy of binding calculations between the protein *NPSR1* and GBQ ligand to obtain absolute measures of protein binding by taking entropy into account. Lastly, protein-membrane simulations were not pursued for *EPB41L2* and *NPSR1*, which are crucial to account for long-range electrostatic interactions. Future experimental work should include protein expression assays that look at the cell viability and apoptotic effects of the putative causal variants (p.V138I and p.R663C) on the *NPSR1* and *EPB41L2* proteins, respectively.

4.3 Conclusion

This study successfully filtered three candidate genes: *LRKK1, NPSR1* and *EPB41L2* as potential PD causative genes. The filtering method applied in this study also validated two known genes, *SNCA*

and *PARK7*, providing a solid basis for using these methods to prioritize other candidate genes for future investigation. Our findings in this study propose that the p.V138I and p.R663C variants on the *NPSR1* and *EPB41L2* protein structures, respectively, can alter protein dynamics. Furthermore, it is interesting to note that both variants (p.V138I and p.R663C) caused increased fluctuations, protein fold changes, and instability in the protein structures compared to the wild-type. However, these findings need to be validated using wet-lab experiments. While it is well-recognized that neurodegenerative diseases are characterized by aberrant protein misfolding and aggregate formation, the mechanisms that initiate or promote proteinopathy in PD remain poorly understood. If *NPSR1* and *EPB41L2* are indeed found to be the causal genes of PD in this multiplex family after experimental validation, this could reveal a new mechanism for the pathobiology of Parkinson's disease.

References

Aarsland, D., Zaccai, J., & Brayne, C. (2005). A systematic review of prevalence studies of dementia in Parkinson's disease. *Movement Disorders*, *20*(10), 1255–1263. https://doi.org/10.1002/mds.20527

Agrahari, A. K., Doss, G. P. C., Siva, R., Magesh, R., & Zayed, H. (2019). Molecular insights of the G2019S substitution in LRRK2 kinase domain associated with Parkinson's disease: A molecular dynamics simulation approach. *Journal of Theoretical Biology*, *469*, 163–171. https://doi.org/10.1016/j.jtbi.2019.03.003

Akinyemi, R. O. (2012). Epidemiology of Parkinsonism and Parkinson's disease in Sub-Saharan Africa: Nigerian profile. *Journal of Neurosciences in Rural Practice*, *03*(03), 233–234. https://doi.org/10.4103/0976-3147.102586

Amod, F. H., & Bhigjee, A. I. (2019). Clinical series of Parkinson's disease in KwaZulu-Natal, South Africa: Retrospective chart review. *Journal of the Neurological Sciences*, *401*, 62–65. https://doi.org/10.1016/j.jns.2019.03.023

Amolo, G. O. (2018). The Growth of High-Performance Computing in Africa. *Computing in Science Engineering*, *20*(3), 21–24. https://doi.org/10.1109/MCSE.2018.03221926

Anderson, P. C., & Daggett, V. (2008). Molecular Basis for the Structural Instability of Human DJ-1 Induced by the L166P Mutation Associated with Parkinson's Disease. *Biochemistry*, *47*(36), 9380–9393. https://doi.org/10.1021/bi800677k

Appel-Cresswell, S., Rajput, A. H., Sossi, V., Thompson, C., Silva, V., McKenzie, J., Dinelle, K., McCormick, S. E., Vilariño-Güell, C., Stoessl, A. J., Dickson, D. W., Robinson, C. A., Farrer, M. J., & Rajput, A. (2014). Clinical, positron emission tomography, and pathological studies of DNAJC13 p.N855S Parkinsonism. *Movement Disorders*, *29*(13), 1684–1687. https://doi.org/10.1002/mds.26019

Aras, S., Bai, M., Lee, I., Springett, R., Hüttemann, M., & Grossman, L. I. (2015). MNRR1 (formerly CHCHD2) is a bi-organellar regulator of mitochondrial metabolism. *Mitochondrion*, *20*, 43–51. https://doi.org/10.1016/j.mito.2014.10.003

Arnittali, M., Rissanou, A. N., & Harmandaris, V. (2019). Structure Of Biomolecules Through Molecular Dynamics Simulations. *Procedia Computer Science*, *156*, 69–78. https://doi.org/10.1016/j.procs.2019.08.181

Auddy, B., Ferreira, M., Blasina, F., Lafon, L., Arredondo, F., Dajas, F., Tripathi, P. C., Seal, T., & Mukherjee, B. (2003). Screening of antioxidant activity of three Indian medicinal plants, traditionally used for the management of neurodegenerative diseases. *Journal of Ethnopharmacology*, *84*(2), 131–138. https://doi.org/10.1016/S0378-8741(02)00322-7

Ausaf Ali, S., Imtaiyaz Hassan, Md., Islam, A., & Ahmad, F. (2014). A Review of Methods Available to Estimate Solvent-Accessible Surface Areas of Soluble Proteins in the Folded and Unfolded States. *Current Protein and Peptide Science*, *15*(5), 456–476.

Ayaz, F. (2003). On the two-dimensional differential transform method. *Applied Mathematics and Computation*, *143*(2), 361–374. https://doi.org/10.1016/S0096-3003(02)00368-5

Balsinde, J., & Balboa, M. A. (2005). Cellular regulation and proposed biological functions of group VIA calcium-independent phospholipase A2 in activated cells. *Cellular Signaling*, *17*(9), 1052–1062. https://doi.org/10.1016/j.cellsig.2005.03.002

Bandres-Ciga, S., Diez-Fairen, M., Kim, J. J., & Singleton, A. B. (2020). Genetics of Parkinson's disease: An introspection of its journey towards precision medicine. *Neurobiology of Disease*, *137*, 104782. https://doi.org/10.1016/j.nbd.2020.104782

Bardien, S., Keyser, R., Lombard, D., Plessis, M. D., Human, H., & Carr, J. (2010). Novel non-sense GCH1 mutation in a South African family diagnosed with dopa-responsive dystonia. *European Journal of Neurology*, *17*(3), 510–512. https://doi.org/10.1111/j.1468-1331.2009.02725.x

Bärenz, F., Kschonsak, Y. T., Meyer, A., Jafarpour, A., Lorenz, H., & Hoffmann, I. (2018). Ccdc61 controls centrosomal localization of Cep170 and is required for spindle assembly and symmetry. *Molecular Biology of the Cell*, *29*(26), 3105–3118. https://doi.org/10.1091/mbc.E18-02-0115

Bayliss, J. A., Lemus, M., Santos, V. V., Deo, M., Elsworth, J. D., & Andrews, Z. B. (2016). Acylated but not des-acyl ghrelin is neuroprotective in an MPTP mouse model of Parkinson's disease. *Journal of Neurochemistry*, *137*(3), 460–471. https://doi.org/10.1111/jnc.13576

Ben-David, Y., Kagan, S., Cohen Ben-Ami, H., Rostami, J., Mizrahi, T., Kulkarni, A. R., Thakur, G. A., Vaknin-Dembinsky, A., Healy, L. M., Brenner, T., & Treinin, M. (2020). RIC3, the cholinergic anti-inflammatory pathway, and neuroinflammation. *International Immunopharmacology*, *83*, 106381. https://doi.org/10.1016/j.intimp.2020.106381

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, *27*(3), 343–350. https://doi.org/10.1093/bioinformatics/btq662

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., & Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*(Web Server issue), W252–W258. https://doi.org/10.1093/nar/gku340

Binder, K., Horbach, J., Kob, W., Paul, W., & Varnik, F. (2004). Molecular dynamics simulations. *Journal of Physics: Condensed Matter*, *16*(5), S429–S453. https://doi.org/10.1088/0953-8984/16/5/006

Biswas, A., Gupta, A., Naiya, T., Das, G., Neogi, R., Datta, S., Mukherjee, S., Das, S. K., Ray, K., & Ray, J. (2006). Molecular pathogenesis of Parkinson's disease: Identification of mutations in the Parkin gene in Indian patients. *Parkinsonism & Related Disorders*, *12*(7), 420–426. https://doi.org/10.1016/j.parkreldis.2006.04.005

Blanckenberg, J., Bardien, S., Glanzmann, B., Okubadejo, N. U., & Carr, J. A. (2013). The prevalence and genetics of Parkinson's disease in sub-Saharan Africans. *Journal of the Neurological Sciences*, *335*(1), 22–25. https://doi.org/10.1016/j.jns.2013.09.010

Blauwendraat, C., Nalls, M. A., & Singleton, A. B. (2020). The genetic architecture of Parkinson's disease. *The Lancet Neurology*, *19*(2), 170–178. https://doi.org/10.1016/S1474-4422(19)30287-X

Bonifati, V., Rizzu, P., Baren, M. J. van, Schaap, O., Breedveld, G. J., Krieger, E., Dekker, M. C. J., Squitieri, F., Ibanez, P., Joosse, M., Dongen, J. W. van, Vanacore, N., Swieten, J. C. van, Brice, A., Meco, G., Duijn, C. M. van, Oostra, B. A., & Heutink, P. (2003). Mutations in the DJ-1 Gene Associated with Autosomal Recessive Early-Onset Parkinsonism. *Science*, *299*(5604), 256–259. https://doi.org/10.1126/science.1077209

Bosgraaf, L., & Van Haastert, P. J. M. (2003). Roc, a Ras/GTPase domain in complex proteins. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1643*(1), 5–10. https://doi.org/10.1016/j.bbamcr.2003.08.008

Bower, J. H., & Zenebe, G. (2005). Neurologic services in the nations of Africa. *Neurology*, *64*(3), 412–415. https://doi.org/10.1212/01.WNL.0000150894.53961.E2

Bowie, J. U., Lüthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science (New York, N.Y.)*, *253*(5016), 164–170. https://doi.org/10.1126/science.1853201

Braak, H., & Del Tredici, K. (2017). Neuropathological Staging of Brain Pathology in Sporadic Parkinson's disease: Separating the Wheat from the Chaff. *Journal of Parkinson's Disease*, *7*(s1), S71–S85. https://doi.org/10.3233/JPD-179001

Braak, H., Sandmann-Keil, D., Gai, W., & Braak, E. (1999). Extensive axonal Lewy neurites in Parkinson's disease: A novel pathological feature revealed by α-synuclein immunocytochemistry. *Neuroscience Letters*, *265*(1), 67–69. https://doi.org/10.1016/S0304-3940(99)00208-6

Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., … Karplus, M. (2009). CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, *30*(10), 1545–1614. https://doi.org/10.1002/jcc.21287

Buffone, C., Martinez-Lopez, A., Fricke, T., Opp, S., Severgnini, M., Cifola, I., Petiti, L., Frabetti, S., Skorupka, K., Zadrozny, K. K., Ganser-Pornillos, B. K., Pornillos, O., Di Nunzio, F., & Diaz-Griffero, F. (2018). Nup153 Unlocks the Nuclear Pore Complex for HIV-1 Nuclear Translocation in Nondividing Cells. *Journal of Virology*, *92*(19). https://doi.org/10.1128/JVI.00648-18

Burke, R. E., Dauer, W. T., & Vonsattel, J. P. G. (2008). A Critical Evaluation of The Braak Staging Scheme for Parkinson's Disease. *Annals of Neurology*, *64*(5), 485–491. https://doi.org/10.1002/ana.21541

Cacabelos, R. (2017). Parkinson's Disease: From Pathogenesis to Pharmacogenomics. *International Journal of Molecular Sciences*, *18*(3). https://doi.org/10.3390/ijms18030551

Calne, D. B. (1993). Treatment of Parkinson's Disease. *New England Journal of Medicine*, *329*(14), 1021–1027. https://doi.org/10.1056/NEJM199309303291408

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Chade, A. R., Kasten, M., & Tanner, C. M. (2006). Nongenetic causes of Parkinson's disease. In P. Riederer, H. Reichmann, M. B. H. Youdim, & M. Gerlach (Eds.), *Parkinson's Disease and Related Disorders* (pp. 147–151). Springer. https://doi.org/10.1007/978-3-211-45295-0_23

Chakkalakal, S. A., Heilig, J., Baumann, U., Paulsson, M., & Zaucke, F. (2018). Impact of Arginine to Cysteine Mutations in Collagen II on Protein Secretion and Cell Survival. *International Journal of Molecular Sciences*, *19*(2). https://doi.org/10.3390/ijms19020541

Chan, C. S., Gertler, T. S., & Surmeier, D. J. (2009). Calcium homeostasis, selective vulnerability and Parkinson's disease. *Trends in Neurosciences*, *32*(5), 249–256. https://doi.org/10.1016/j.tins.2009.01.006

Chartier-Harlin, M.-C., Dachsel, J. C., Vilariño-Güell, C., Lincoln, S. J., Leprêtre, F., Hulihan, M. M., Kachergus, J., Milnerwood, A. J., Tapia, L., Song, M.-S., Le Rhun, E., Mutez, E., Larvor, L., Duflot, A., Vanbesien-Mailliot, C., Kreisler, A., Ross, O. A., Nishioka, K., Soto-Ortolaza, A. I., … Farrer, M. J. (2011). Translation Initiator EIF4G1 Mutations in Familial Parkinson Disease. *The American Journal of Human Genetics*, *89*(3), 398–406. https://doi.org/10.1016/j.ajhg.2011.08.009

Chaudhuri, K. R., Healy, D. G., & Schapira, A. H. (2006). Non-motor symptoms of Parkinson's disease: Diagnosis and management. *The Lancet Neurology*, *5*(3), 235–245. https://doi.org/10.1016/S1474-4422(06)70373-8

Chen, S., Lu, M., Liu, D., Yang, L., Yi, C., Ma, L., Zhang, H., Liu, Q., Frimurer, T. M., Wang, M.-W., Schwartz, T. W., Stevens, R. C., Wu, B., Wüthrich, K., & Zhao, Q. (2019). Human substance P receptor binding mode of the antagonist drug aprepitant by NMR and crystallography. *Nature Communications*, *10*. https://doi.org/10.1038/s41467-019-08568-5

Cheng, J. (2008). A multi-template combination algorithm for protein comparative modeling. *BMC Structural Biology*, *8*(1), 18. https://doi.org/10.1186/1472-6807-8-18

Cherniavskyi, Y. K., Oliva, R., Stellato, M., Vecchio, P. D., Galdiero, S., Falanga, A., Dames, S. A., & Tieleman, D. P. (2021). Structural characterization of the antimicrobial peptides myxinidin and WMR in bacterial membrane mimetic micelles and bicelles. *BioRxiv*, 2021.03.30.437760. https://doi.org/10.1101/2021.03.30.437760

Chitongo, R., Obasa, A. E., Mikasi, S. G., Jacobs, G. B., & Cloete, R. (2020). Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding. *PLOS ONE*, *15*(5), e0223464. https://doi.org/10.1371/journal.pone.0223464

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, *7*(10), e46688. https://doi.org/10.1371/journal.pone.0046688

Chow-White, P.A., & García-Sancho, M. (2012). Bidirectional shaping and spaces of convergence: Interactions between biology and computing from the first DNA sequencers to global genome databases. *Science, Technology, & Human Values, 37*(1), 124-164.

Cloete, R., Kapp, E., Joubert, J., Christoffels, A., & Malan, S. F. (2018). Molecular modelling and simulation studies of the Mycobacterium tuberculosis multidrug efflux pump protein Rv1258c. *PLOS ONE*, *13*(11), e0207605. https://doi.org/10.1371/journal.pone.0207605

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science: A Publication of the Protein Society*, *2*(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

Coskuner, O., & Wise-Scira, O. (2013). Structures and free energy landscapes of the A53T mutant-type α-synuclein protein and impact of A53T mutation on the structures of the wild-type α-synuclein protein with dynamics. *ACS Chemical Neuroscience*, *4*(7), 1101–1113. https://doi.org/10.1021/cn400041j

Coskuner-Weber, O., & Uversky, V. N. (2018). Insights into the Molecular Mechanisms of Alzheimer's and Parkinson's Diseases with Molecular Simulations: Understanding the Roles of Artificial and Pathological Missense Mutations in Intrinsically Disordered Proteins Related to Pathology. *International Journal of Molecular Sciences*, *19*(2). https://doi.org/10.3390/ijms19020336

Cosnett, J. E., & Bill, P. L. (1988). Parkinson's disease in blacks. Observations on epidemiology in Natal. *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde*, *73*(5), 281–283.

Crosby, N. J., Deane, K., & Clarke, C. E. (2003). Amantadine in Parkinson's disease. *Cochrane Database of Systematic Reviews*, *1*. https://doi.org/10.1002/14651858.CD003468

Cubo, E., Doumbe, J., Martinez-Martin, P., Rodriguez-Blazquez, C., Kuate, C., Mariscal, N., Lopez, I., Noubissi, G., Mapoure, Y. N., Jon, J. L., Mbahe, S., Tchaleu, B., & Catalan, M.-J. (2014). Comparison of the clinical profile of Parkinson's disease between Spanish and Cameroonian Cohorts. *Journal of the Neurological Sciences*, *336*(1), 122–126. https://doi.org/10.1016/j.jns.2013.10.021

David, C. C., & Jacobs, D. J. (2014). Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. In D. R. Livesay (Ed.), *Protein Dynamics: Methods and Protocols* (pp. 193–226). Humana Press. https://doi.org/10.1007/978-1-62703-658-0_11

De Miranda, B. R., & Greenamyre, J. T. (2020). Trichloroethylene, a ubiquitous environmental contaminant in the risk for Parkinson's disease. *Environmental Science. Processes & Impacts*, *22*(3), 543–554. https://doi.org/10.1039/c9em00578a

Dehay, B., Ramirez, A., Martinez-Vicente, M., Perier, C., Canron, M.-H., Doudnikoff, E., Vital, A., Vila, M., Klein, C., & Bezard, E. (2012). Loss of P-type ATPase ATP13A2/PARK9 function induces general lysosomal deficiency and leads to Parkinson disease neurodegeneration. *Proceedings of the National Academy of Sciences*, *109*(24), 9611–9616. https://doi.org/10.1073/pnas.1112368109

Dekker, M. C. J., Bonifati, V., & van Duijn, C. M. (2003). Parkinson's disease: Piecing together a genetic jigsaw. *Brain*, *126*(8), 1722–1733. https://doi.org/10.1093/brain/awg172

Dekker, M. C. J., Coulibaly, T., Bardien, S., Ross, O. A., Carr, J., & Komolafe, M. (2020). Parkinson's Disease Research on the African Continent: Obstacles and Opportunities. *Frontiers in Neurology*, *11*, 512. https://doi.org/10.3389/fneur.2020.00512

Delemotte, L. (2020). Chapter 7: Bridging the Gap Between Atomistic Molecular Dynamics Simulations and Wet-lab Experimental Techniques: Applications to Membrane Proteins. In *Computational Techniques for Analytical Chemistry and Bioanalysis* (pp. 247–286). https://doi.org/10.1039/9781788015882-00247

Dobosz-Bartoszek, M., Pinkerton, M. H., Otwinowski, Z., Chakravarthy, S., Söll, D., Copeland, P. R., & Simonović, M. (2016). Crystal structures of the human elongation factor eEFSec suggest a non-canonical mechanism for selenocysteine incorporation. *Nature Communications*, *7*, 12941. https://doi.org/10.1038/ncomms12941

Dorsey, E. R., Elbaz, A., Nichols, E., Abd-Allah, F., Abdelalim, A., Adsuar, J. C., Ansha, M. G., Brayne, C., Choi, J.-Y. J., Collado-Mateo, D., Dahodwala, N., Do, H. P., Edessa, D., Endres, M., Fereshtehnejad, S.-M., Foreman, K. J., Gankpe, F. G., Gupta, R., Hankey, G. J., … Murray, C. J. L. (2018). Global, regional, and national burden of Parkinson's disease, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, *17*(11), 939–953. https://doi.org/10.1016/S1474-4422(18)30295-3

Dorsey, E. R., Sherer, T., Okun, M. S., & Bloem, B. R. (2018). The Emerging Evidence of the Parkinson Pandemic. *Journal of Parkinson's Disease*, *8*(s1), S3–S8. https://doi.org/10.3233/JPD-181474

Dotchin, C., Msuya, O., Kissima, J., Massawe, J., Mhina, A., Moshy, A., Aris, E., Jusabani, A., Whiting, D., Masuki, G., & Walker, R. (2008). The prevalence of Parkinson's disease in rural Tanzania. *Movement Disorders*, *23*(11), 1567–1672. https://doi.org/10.1002/mds.21898

Dotchin, C., & Walker, R. (2012). The management of Parkinson's disease in sub-Saharan Africa. *Expert Review of Neurotherapeutics*, *12*(6), 661–666. https://doi.org/10.1586/ern.12.52

Du, J., Yuan, Z., Ma, Z., Song, J., Xie, X., & Chen, Y. (2014). KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Molecular BioSystems*, *10*(9), 2441–2447. https://doi.org/10.1039/C4MB00287C

Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, *9*(1), 71. https://doi.org/10.1186/1741-7007-9-71

Edvardson, S., Cinnamon, Y., Ta-Shma, A., Shaag, A., Yim, Y.-I., Zenvirt, S., Jalas, C., Lesage, S., Brice, A., Taraboulos, A., Kaestner, K. H., Greene, L. E., & Elpeleg, O. (2012). A Deleterious Mutation in DNAJC6 Encoding the Neuronal-Specific Clathrin-Uncoating Co-Chaperone

Auxilin, Is Associated with Juvenile Parkinsonism. *PLOS ONE*, *7*(5), e36458. https://doi.org/10.1371/journal.pone.0036458

Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. In *Methods in Enzymology* (Vol. 277, pp. 396–404). Academic Press. https://doi.org/10.1016/S0076-6879(97)77022-8

Engh, R. A., & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica Section A*, *47*(4), 392–400. https://doi.org/10.1107/S0108767391001071

Femi, O. L., Ibrahim, A., & Aliyu, S. (2012). Clinical profile of parkinsonian disorders in the tropics: Experience at Kano, northwestern Nigeria. *Journal of Neurosciences in Rural Practice*, *03*(03), 237–241. https://doi.org/10.4103/0976-3147.102589

Fonzo, A. D., Dekker, M. C. J., Montagna, P., Baruzzi, A., Yonova, E. H., Guedes, L. C., Szczerbinska, A., Zhao, T., Dubbel-Hulsman, L. O. M., Wouters, C. H., Graaff, E. de, Oyen, W. J. G., Simons, E. J., Breedveld, G. J., Oostra, B. A., Horstink, M. W., & Bonifati, V. (2009). FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. *Neurology*, *72*(3), 240–245. https://doi.org/10.1212/01.wnl.0000338144.10967.2b

Frédéric, M. Y., Lalande, M., Boileau, C., Hamroun, D., Claustres, M., Béroud, C., & Collod-Béroud, G. (2009). UMD-predictor, a new prediction tool for nucleotide substitution pathogenicity— application to four genes: FBN1, FBN2, TGFBR1, and TGFBR2. *Human Mutation*, *30*(6), 952–959. https://doi.org/10.1002/humu.20970

Futter, M., Diekmann, H., Schoenmakers, E., Sadiq, O., Chatterjee, K., & Rubinsztein, D. C. (2009). Wild-type but not mutant huntingtin modulates the transcriptional activity of liver X receptors. *Journal of Medical Genetics*, *46*(7), 438–446. https://doi.org/10.1136/jmg.2009.066399

Gajula, M. P., Kumar, A., & Ijaq, J. (2016). Protocol for Molecular Dynamics Simulations of Proteins. *Bio-Protocol*, *6*(23), e2051–e2051.

Galet, B., Ingallinesi, M., Pegon, J., Thi, A. D., Ravassard, P., Biguet, N. F., & Meloni, R. (2021). G-protein coupled receptor 88 knockdown in the associative striatum reduces psychiatric symptoms in a translational male rat model of Parkinson disease. *Journal of Psychiatry & Neuroscience: JPN*, *46*(1), E44–E55. https://doi.org/10.1503/jpn.190171

Gasser, T. (2009). Molecular pathogenesis of Parkinson disease: Insights from genetic studies. *Expert Reviews in Molecular Medicine*, *11*, e22. https://doi.org/10.1017/S1462399409001148

Gechter, J., Liebscher, C., Geiger, M. J., Wittmann, A., Schlagenhauf, F., Lueken, U., Wittchen, H.-U., Pfleiderer, B., Arolt, V., Kircher, T., Straube, B., Deckert, J., Weber, H., Herrmann, M. J., Reif, A., Domschke, K., & Ströhle, A. (2019). Association of NPSR1 gene variation and neural activity in patients with panic disorder and agoraphobia and healthy controls. *NeuroImage: Clinical*, *24*, 102029. https://doi.org/10.1016/j.nicl.2019.102029

Geng, H., Chen, F., Ye, J., & Jiang, F. (2019). Applications of Molecular Dynamics Simulation in Structure Prediction of Peptides and Proteins. *Computational and Structural Biotechnology Journal*, *17*, 1162–1170. https://doi.org/10.1016/j.csbj.2019.07.010

Georgiev, D., Hamberg, K., Hariz, M., Forsgren, L., & Hariz, G.-M. (2017). Gender differences in Parkinson's disease: A clinical perspective. *Acta Neurologica Scandinavica*, *136*(6), 570–584. https://doi.org/10.1111/ane.12796

Ghazal, P. (2016). The Physio-Pharmacological Role of the NPS/NPSR System in Psychiatric Disorders: A Translational Overview. *Current Protein and Peptide Science*, *17*(4), 380–397.

Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. M. (2003). The diagnostic odds ratio: A single indicator of test performance. *Journal of Clinical Epidemiology*, *56*(11), 1129–1135. https://doi.org/10.1016/S0895-4356(03)00177-X

Goetz, C. G. (2011). The History of Parkinson's Disease: Early Clinical Descriptions and Neurological Therapies. *Cold Spring Harbor Perspectives in Medicine*, *1*(1), a008862–a008862. https://doi.org/10.1101/cshperspect.a008862

Gorell, J. M., Peterson, E. L., Rybicki, B. A., & Johnson, C. C. (2004). Multiple risk factors for Parkinson's disease. *Journal of the Neurological Sciences*, *217*(2), 169–174. https://doi.org/10.1016/j.jns.2003.09.014

Guala, D., & Sonnhammer, E. L. L. (2017). A large-scale benchmark of gene prioritization methods. *Scientific Reports*, *7*(1), 46598. https://doi.org/10.1038/srep46598

Guarguaglini, G., Duncan, P. I., Stierhof, Y. D., Holmström, T., Duensing, S., & Nigg, E. A. (2005). The forkhead-associated domain protein Cep170 interacts with Polo-like kinase 1 and serves as a marker for mature centrioles. *Molecular Biology of the Cell*, *16*(3), 1095–1107. https://doi.org/10.1091/mbc.e04-10-0939

Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *ELECTROPHORESIS*, *30*(S1), S162–S173. https://doi.org/10.1002/elps.200900140

Guo, L., Girisha, K. M., Iida, A., Hebbar, M., Shukla, A., Shah, H., Nishimura, G., Matsumoto, N., Nismath, S., Miyake, N., & Ikegawa, S. (2017). Identification of a novel LRRK1 mutation in a family with osteosclerotic metaphyseal dysplasia. *Journal of Human Genetics*, *62*(3), 437–441. https://doi.org/10.1038/jhg.2016.136

Guo, Y.-B., Chen, J., Zhang, X.-D., Xu, S.-B., & Liu, H.-Y. (2016). Molecular dynamics simulations to understand LRKK2 mutations in Parkinson. *Molecular Simulation*, *42*(1), 64–70. https://doi.org/10.1080/08927022.2015.1008471

Gusdon, A. M., Zhu, J., Van Houten, B., & Chu, C. T. (2012). ATP13A2 regulates mitochondrial bioenergetics through macroautophagy. *Neurobiology of Disease*, *45*(3), 962–972. https://doi.org/10.1016/j.nbd.2011.12.015

Gustavsson, E. K., Trinh, J., Guella, I., Vilariño-Güell, C., Appel-Cresswell, S., Stoessl, A. J., Tsui, J. K., McKeown, M., Rajput, A., Rajput, A. H., Aasly, J. O., & Farrer, M. J. (2015). DNAJC13 genetic variants in parkinsonism. *Movement Disorders*, *30*(2), 273–278. https://doi.org/10.1002/mds.26064

Hall, R., Hall, I. P., & Sayers, I. (2019). Genetic risk factors for the development of pulmonary disease identified by genome-wide association. *Respirology*, *24*(3), 204–214. https://doi.org/10.1111/resp.13436

Healy, D. G., Falchi, M., O'Sullivan, S. S., Bonifati, V., Durr, A., Bressman, S., Brice, A., Aasly, J., Zabetian, C. P., Goldwurm, S., Ferreira, J. J., Tolosa, E., Kay, D. M., Klein, C., Williams, D. R., Marras, C., Lang, A. E., Wszolek, Z. K., Berciano, J., … Wood, N. W. (2008). Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: A case-control study. *The Lancet Neurology*, *7*(7), 583–590. https://doi.org/10.1016/S1474-4422(08)70117-0

Hicks, A. A., Pétursson, H., Jónsson, T., Stefánsson, H., Jóhannsdóttir, H. S., Sainz, J., Frigge, M. L., Kong, A., Gulcher, J. R., Stefánsson, K., & Sveinbjörnsdóttir, S. (2002). A susceptibility gene for late-onset idiopathic Parkinson's disease. *Annals of Neurology*, *52*(5), 549–555. https://doi.org/10.1002/ana.10324

Higueruelo, A. P., Schreyer, A., Bickerton, G. R. J., Blundell, T. L., & Pitt, W. R. (2012). What Can We Learn from the Evolution of Protein-Ligand Interactions to Aid the Design of New Therapeutics? *PLOS ONE*, *7*(12), e51742. https://doi.org/10.1371/journal.pone.0051742

Hollingsworth, S., & Dror, R. (2018). Molecular Dynamics Simulation for All. *Neuron*, *99*, 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

Holloway, R., I, S., S, F., K, K., A, L., K, M., M, M., J, S., W, W., B, M., C, K., M, W., A, S., R, P., L, B., J, H., P, L., J, M., O, S., … undefined. (2004). Pramipexole vs levodopa as initial treatment for Parkinson disease: A 4-year randomized controlled trial. *Archives of Neurology*, *61*(7), 1044–1053. https://doi.org/10.1001/archneur.61.7.1044

Huang, J., & MacKerell, A. D. (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry*, *34*(25), 2135–2145. https://doi.org/10.1002/jcc.23354

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery & Psychiatry*, *55*(3), 181–184. https://doi.org/10.1136/jnnp.55.3.181

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Iida, A., Xing, W., Docx, M. K. F., Nakashima, T., Wang, Z., Kimizuka, M., Van Hul, W., Rating, D., Spranger, J., Ohashi, H., Miyake, N., Matsumoto, N., Mohan, S., Nishimura, G., Mortier, G., & Ikegawa, S. (2016). Identification of biallelic LRRK1 mutations in osteosclerotic metaphyseal dysplasia and evidence for locus heterogeneity. *Journal of Medical Genetics*, *53*(8), 568–574. https://doi.org/10.1136/jmedgenet-2016-103756

Isaias, I. U., Marzegan, A., Pezzoli, G., Marotta, G., Canesi, M., Biella, G. E. M., Volkmann, J., & Cavallari, P. (2012). A role for locus coeruleus in Parkinson tremor. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00179

Ishikawa, K., Nagase, T., Nakajima, D., Seki, N., Ohira, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., & Ohara, O. (1997). Prediction of the Coding Sequences of Unidentified Human Genes. VIII. 78 New cDNA Clones from Brain Which Code for Large Proteins in vitro. *DNA Research*, *4*(5), 307–313. https://doi.org/10.1093/dnares/4.5.307

Iyer, M., Venkatesan, D., Subramaniam, M., & Vellingiri, B. (2020). Mutational analysis of SLC6A3 gene in Parkinson's disease (PD) patients in Coimbatore Population, India. *Parkinsonism & Related Disorders*, *79*, e32. https://doi.org/10.1016/j.parkreldis.2020.06.136

Jankovic, J. (2008). Parkinson's disease: Clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, *79*(4), 368–376. https://doi.org/10.1136/jnnp.2007.131045

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, *29*(11), 1859–1865. https://doi.org/10.1002/jcc.20945

Joch, M., Ase, A. R., Chen, C. X.-Q., MacDonald, P. A., Kontogiannea, M., Corera, A. T., Brice, A., Séguéla, P., & Fon, E. A. (2007). Parkin-mediated Monoubiquitination of the PDZ Protein PICK1 Regulates the Activity of Acid-sensing Ion Channels. *Molecular Biology of the Cell*, *18*(8), 3105–3118. https://doi.org/10.1091/mbc.e05-11-1027

Johri, A., & Beal, M. F. (2012). Mitochondrial dysfunction in neurodegenerative diseases. *The Journal of Pharmacology and Experimental Therapeutics*, *342*(3), 619–630. https://doi.org/10.1124/jpet.112.192138

Jones, A. R., Overly, C. C., & Sunkin, S. M. (2009). The Allen Brain Atlas: 5 years and beyond. *Nature Reviews Neuroscience*, *10*(11), 821–828. https://doi.org/10.1038/nrn2722

Kandt, C., Ash, W. L., & Peter Tieleman, D. (2007). Setting up and running molecular dynamics simulations of membrane proteins. *Methods*, *41*(4), 475–488. https://doi.org/10.1016/j.ymeth.2006.08.006

Karimi-Moghadam, A., Charsouei, S., Bell, B., & Jabalameli, M. R. (2018). Parkinson Disease from Mendelian Forms to Genetic Susceptibility: New Molecular Insights into the Neurodegeneration Process. *Cellular and Molecular Neurobiology*, *38*(6), 1153–1178. https://doi.org/10.1007/s10571-018-0587-4

Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, *9*(9), 646–652. https://doi.org/10.1038/nsb0902-646

Khalil, H., Chahine, L., Siddiqui, J., Aldaajani, Z., & Bajwa, J. A. (2020). Parkinson's disease in the MENASA countries. *The Lancet Neurology*, *19*(4), 293–294. Scopus. https://doi.org/10.1016/S1474-4422(20)30026-0

Kim, T., Kasprzak, W. K., & Shapiro, B. A. (2017). Protocols for Molecular Dynamics Simulations of RNA Nanostructures. In E. Bindewald & B. A. Shapiro (Eds.), *RNA Nanostructures: Methods and Protocols* (pp. 33–64). Springer. https://doi.org/10.1007/978-1-4939-7138-1_3

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315. https://doi.org/10.1038/ng.2892

Kisoli, A., Gray, W. K., Dotchin, C. L., Orega, G., Dewhurst, F., Paddick, S.-M., Longdon, A., Chaote, P., Dewhurst, M., & Walker, R. W. (2015). Levels of functional disability in elderly people in Tanzania with dementia, stroke and Parkinson's disease. *Acta Neuropsychiatrica*, *27*(4), 206–212. https://doi.org/10.1017/neu.2015.9

Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., & Shaw, D. E. (2009). Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology*, *19*(2), 120–127. https://doi.org/10.1016/j.sbi.2009.03.004

Koller, W. C., & Rueda, M. G. (1998). Mechanism of action of dopaminergic agents in Parkinson's disease. *Neurology*, *50*(6 Suppl 6), S11. https://doi.org/10.1212/WNL.50.6_Suppl_6.S11

Kononenko, N. L., & Haucke, V. (2015). Molecular Mechanisms of Presynaptic Membrane Retrieval and Synaptic Vesicle Reformation. *Neuron*, *85*(3), 484–496. https://doi.org/10.1016/j.neuron.2014.12.016

Köroğlu, Ç., Baysal, L., Cetinkaya, M., Karasoy, H., & Tolun, A. (2013). DNAJC6 is responsible for juvenile parkinsonism with phenotypic variability. *Parkinsonism & Related Disorders*, *19*(3), 320–324. https://doi.org/10.1016/j.parkreldis.2012.11.006

Korr, D., Toschi, L., Donner, P., Pohlenz, H.-D., Kreft, B., & Weiss, B. (2006). LRRK1 protein kinase activity is stimulated upon binding of GTP to its Roc domain. *Cellular Signalling*, *18*(6), 910–920. https://doi.org/10.1016/j.cellsig.2005.08.015

Kouli, A., Torsney, K. M., & Kuan, W.-L. (2018). Parkinson's Disease: Etiology, Neuropathology, and Pathogenesis. In T. B. Stoker & J. C. Greenland (Eds.), *Parkinson's Disease: Pathogenesis and Clinical Aspects*. Codon Publications. http://www.ncbi.nlm.nih.gov/books/NBK536722/

Krebs, C. E., Karkheiran, S., Powell, J. C., Cao, M., Makarov, V., Darvish, H., Paolo, G. D., Walker, R. H., Shahidi, G. A., Buxbaum, J. D., Camilli, P. D., Yue, Z., & Paisán-Ruiz, C. (2013). The Sac1 Domain of SYNJ1 Identified Mutated in a Family with Early-Onset Progressive Parkinsonism with Generalized Seizures. *Human Mutation*, *34*(9), 1200–1207. https://doi.org/10.1002/humu.22372

Kuzmanic, A., & Zagrovic, B. (2010). Determination of Ensemble-Average Pairwise Root Mean-Square Deviation from Experimental B-Factors. *Biophysical Journal*, *98*(5), 861–871. https://doi.org/10.1016/j.bpj.2009.11.011

Langlais, P., Dong, L. Q., Ramos, F. J., Hu, D., Li, Y., Quon, M. J., & Liu, F. (2004). Negative Regulation of Insulin-Stimulated Mitogen-Activated Protein Kinase Signaling By Grb10. *Molecular Endocrinology*, *18*(2), 350–358. https://doi.org/10.1210/me.2003-0117

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR*, *8*(4), 477–486. https://doi.org/10.1007/BF00228148

Lee, A., & Gilbert, R. M. (2016). Epidemiology of Parkinson Disease. *Neurologic Clinics*, *34*(4), 955–965. https://doi.org/10.1016/j.ncl.2016.06.012

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., MacKerell, A. D., Klauda, J. B., & Im, W. (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, *12*(1), 405–413. https://doi.org/10.1021/acs.jctc.5b00935

Lekoubou, A., Echouffo-Tcheugui, J. B., & Kengne, A. P. (2014). Epidemiology of neurodegenerative diseases in sub-Saharan Africa: A systematic review. *BMC Public Health*, *14*(1), 653. https://doi.org/10.1186/1471-2458-14-653

Lemaître, C., Fischer, B., Kalousi, A., Hoffbeck, A.-S., Guirouilh-Barbat, J., Shahar, O. D., Genet, D., Goldberg, M., Betrand, P., Lopez, B., Brino, L., & Soutoglou, E. (2012). The nucleoporin 153, a novel factor in double-strand break repair and DNA damage response. *Oncogene*, *31*(45), 4803–4809. https://doi.org/10.1038/onc.2011.638

Lemkul, J. (2019). From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. *Living Journal of Computational Molecular Science*, *1*(1). https://doi.org/10.33011/livecoms.1.1.5068

Lemos, A., Melo, R., Preto, A. J., Almeida, J. G., Moreira, I. S., & Cordeiro, M. N. D. S. (2018). In Silico Studies Targeting G-protein Coupled Receptors for Drug Research Against Parkinson's Disease. *Current Neuropharmacology*, *16*(6), 786–848. https://doi.org/10.2174/1570159X16666180308161642

Lesage, S., & Brice, A. (2012). Role of Mendelian genes in "sporadic" Parkinson's disease. *Parkinsonism & Related Disorders*, *18*, S66–S70. https://doi.org/10.1016/S1353-8020(11)70022-0

Lesage, S., Drouet, V., Majounie, E., Deramecourt, V., Jacoupy, M., Nicolas, A., Cormier-Dequaire, F., Hassoun, S. M., Pujol, C., Ciura, S., Erpapazoglou, Z., Usenko, T., Maurage, C.-A., Sahbatou, M., Liebau, S., Ding, J., Bilgic, B., Emre, M., Erginel-Unaltuna, N., … Brice, A. (2016). Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. *The American Journal of Human Genetics*, *98*(3), 500–513. https://doi.org/10.1016/j.ajhg.2016.01.014

Li, X., Ye, M., Wang, Y., Qiu, M., Fu, T., Zhang, J., Zhou, B., & Lu, S. (2020). How Parkinson's disease-related mutations disrupt the dimerization of WD40 domain in LRRK2: A comparative molecular dynamics simulation study. *Physical Chemistry Chemical Physics*, *22*(36), 20421–20433. https://doi.org/10.1039/D0CP03171B

Lian, P. (2018). *How can I run a repetition of MD simulation using NAMD*. https://www.researchgate.net/post/How_can_I_run_a_repetition_of_MD_simulation_using_ NAMD/5a902ae493553b666569f062/citation/download.

Lin, L., Göke, J., Cukuroglu, E., Dranias, M. R., VanDongen, A. M. J., & Stanton, L. W. (2016). Molecular Features Underlying Neurodegeneration Identified through In Vitro Modeling of Genetically Diverse Parkinson's Disease Patients. *Cell Reports*, *15*(11), 2411–2426. https://doi.org/10.1016/j.celrep.2016.05.022

Livingstone, P. D., & Wonnacott, S. (2009). Nicotinic acetylcholine receptors and the ascending dopamine pathways. *Biochemical Pharmacology*, *78*(7), 744–755. https://doi.org/10.1016/j.bcp.2009.06.004

Lombard, A. J., & Gelfand, M. D. (1978). Parkinson's disease in the African. *The Central African Journal of Medicine*.

Lücking, C. B., Dürr, A., Bonifati, V., Vaughan, J., De Michele, G., Gasser, T., Harhangi, B. S., Meco, G., Denèfle, P., Wood, N. W., Agid, Y., Nicholl, D., Breteler, M. M. B., Oostra, B. A., De Mari, M., Marconi, R., Filla, A., Bonnet, A.-M., Broussolle, E., … Brice, A. (2000). Association between Early-Onset Parkinson's Disease and Mutations in the Parkin Gene. *New England Journal of Medicine*, *342*(21), 1560–1567. https://doi.org/10.1056/NEJM200005253422103

Ludtmann, M. H. R., & Abramov, A. Y. (2018). Mitochondrial calcium imbalance in Parkinson's disease. *Neuroscience Letters*, *663*, 86–90. https://doi.org/10.1016/j.neulet.2017.08.044

Lüthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, *356*(6364), 83–85. https://doi.org/10.1038/356083a0

Maasz, G., Zrinyi, Z., Reglodi, D., Petrovics, D., Rivnyak, A., Kiss, T., Jungling, A., Tamas, A., & Pirger, Z. (2017). Pituitary adenylate cyclase-activating polypeptide (PACAP) has a neuroprotective function in dopamine-based neurodegeneration in rat and snail parkinsonian models. *Disease Models & Mechanisms*, *10*(2), 127–139. https://doi.org/10.1242/dmm.027185

Makkonen, H., Jääskeläinen, T., Pitkänen-Arsiola, T., Rytinki, M., Waltering, K. K., Mättö, M., Visakorpi, T., & Palvimo, J. J. (2008). Identification of ETS-like transcription factor 4 as a novel androgen receptor target in prostate cancer cells. *Oncogene*, *27*(36), 4865–4876. https://doi.org/10.1038/onc.2008.125

Marras, C., Beck, J. C., Bower, J. H., Roberts, E., Ritz, B., Ross, G. W., Abbott, R. D., Savica, R., Van Den Eeden, S. K., Willis, A. W., & Tanner, C. M. (2018). Prevalence of Parkinson's disease across North America. *Npj Parkinson's Disease*, *4*(1), 1–7. https://doi.org/10.1038/s41531-018-0058-0

Marsden, C. A. (2006). Dopamine: The rewarding years. *British Journal of Pharmacology*, *147*(S1), S136–S144. https://doi.org/10.1038/sj.bjp.0706473

Martínez, L. (2015). Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. *PLOS ONE*, *10*(3), e0119264. https://doi.org/10.1371/journal.pone.0119264

Martinez-Martin, P., Rodriguez-Blazquez, C., Alvarez-Sanchez, M., Arakaki, T., Bergareche-Yarza, A., Chade, A., Garretto, N., Gershanik, O., Kurtis, M. M., Martinez-Castrillo, J. C., Mendoza-Rodriguez, A., Moore, H. P., Rodriguez-Violante, M., Singer, C., Tilley, B. C., Huang, J., Stebbins, G. T., & Goetz, C. G. (2013). Expanded and independent validation of the Movement Disorder Society–Unified Parkinson's Disease Rating Scale (MDS-UPDRS). *Journal of Neurology*, *260*(1), 228–236. https://doi.org/10.1007/s00415-012-6624-1

Martoňák, R., Laio, A., & Parrinello, M. (2003). Predicting Crystal Structures: The Parrinello-Rahman Method Revisited. *Physical Review Letters*, *90*(7), 075503. https://doi.org/10.1103/PhysRevLett.90.075503

McMorrow, I., Bastos, R., Horton, H., & Burke, B. (1994). Sequence analysis of a cDNA encoding a human nuclear pore complex protein, hnup153. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1217*(2), 219–223. https://doi.org/10.1016/0167-4781(94)90040-X

Mhyre, T. R., Boyd, J. T., Hamill, R. W., & Maguire-Zeiss, K. A. (2012). Parkinson's Disease. *Sub-Cellular Biochemistry*, *65*, 389–455. https://doi.org/10.1007/978-94-007-5416-4_16

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins*, *12*(4), 345–364. https://doi.org/10.1002/prot.340120407

Nagar, S., Juyal, R. C., Chaudhary, S., Behari, M., Gupta, M., Rao, S. N., & Thelma, B. K. (2001). Mutations in the α-synuclein gene in Parkinson's disease among Indians. *Acta Neurologica Scandinavica*, *103*(2), 120–122. https://doi.org/10.1034/j.1600-0404.2001.103002120.x

Nair, P. C., & Miners, J. O. (2014). Molecular dynamics simulations: From structure function relationships to drug discovery. *In Silico Pharmacology*, *2*(1), 4. https://doi.org/10.1186/s40203-014-0004-8

Nalls, M. A., Pankratz, N., Lill, C. M., Do, C. B., Hernandez, D. G., Saad, M., DeStefano, A. L., Kara, E., Bras, J., Sharma, M., Schulte, C., Keller, M. F., Arepalli, S., Letson, C., Edsall, C., Stefansson, H., Liu, X., Pliner, H., Lee, J. H., … Singleton, A. B. (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature Genetics*, *46*(9), 989–993. https://doi.org/10.1038/ng.3043

Nelson, D. E., Randle, S. J., & Laman, H. (n.d.). Beyond ubiquitination: The atypical functions of Fbxo7 and other F-box proteins. *Open Biology*, *3*(10), 130131. https://doi.org/10.1098/rsob.130131

Nestor, M. W., Cai, X., Stone, M. R., Bloch, R. J., & Thompson, S. M. (2011). The Actin Binding Domain of βI-Spectrin Regulates the Morphological and Functional Dynamics of Dendritic Spines. *PLOS ONE*, *6*(1), e16197. https://doi.org/10.1371/journal.pone.0016197

Neufang, S., Geiger, M. J., Homola, G. A., Mahr, M., Akhrif, A., Nowak, J., Reif, A., Romanos, M., Deckert, J., Solymosi, L., & Domschke, K. (2015). Modulation of prefrontal functioning in attention systems by NPSR1 gene variation. *NeuroImage*, *114*, 199–206. https://doi.org/10.1016/j.neuroimage.2015.03.064

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, *541*(7637), 302–310. https://doi.org/10.1038/nature21347

Ober, C., & Yao, T.-C. (2011). The genetics of asthma and allergic disease: A 21st century perspective. *Immunological Reviews*, *242*(1), 10–30. https://doi.org/10.1111/j.1600-065X.2011.01029.x

Oczkowska, A., Kozubski, W., Lianeri, M., & Dorszewska, J. (2013). Mutations in PRKN and SNCA Genes Important for the Progress of Parkinson's Disease. *Current Genomics*, *14*(8), 502–517.

Ode, H., Nakashima, M., Kitamura, S., Sugiura, W., & Sato, H. (2012). Molecular dynamics simulation in virus research. *Frontiers in Microbiology*, *3*. https://doi.org/10.3389/fmicb.2012.00258

Ohno, N., Terada, N., Yamakawa, H., Komada, M., Ohara, O., Trapp, B. D., & Ohno, S. (2006). Expression of protein 4.1G in Schwann cells of the peripheral nervous system. *Journal of Neuroscience Research*, *84*(3), 568–577. https://doi.org/10.1002/jnr.20949

Olgiati, S., De Rosa, A., Quadri, M., Criscuolo, C., Breedveld, G. J., Picillo, M., Pappatà, S., Quarantelli, M., Barone, P., De Michele, G., & Bonifati, V. (2014). PARK20 caused by SYNJ1 homozygous Arg258Gln mutation in a new Italian family. *Neurogenetics*, *15*(3), 183–188. https://doi.org/10.1007/s10048-014-0406-0

Osuntokun, B. O., Adeuja, A. O., Schoenberg, B. S., Bademosi, O., Nottidge, V. A., Olumide, A. O., Ige, O., Yaria, F., & Bolis, C. L. (1987). Neurological disorders in Nigerian Africans: A community-based study. *Acta Neurologica Scandinavica*, *75*(1), 13–21. https://doi.org/10.1111/j.1600-0404.1987.tb07883.x

Ottolini, D., Calì, T., Negro, A., & Brini, M. (2013). The Parkinson disease-related protein DJ-1 counteracts mitochondrial impairment induced by the tumour suppressor protein p53 by enhancing endoplasmic reticulum–mitochondria tethering. *Human Molecular Genetics*, *22*(11), 2152–2168. https://doi.org/10.1093/hmg/ddt068

Paliy, M., Melnik, R., & Shapiro, B. A. (2010). Coarse-graining RNA nanostructures for molecular dynamics simulations. *Physical Biology*, *7*(3), 036001. https://doi.org/10.1088/1478-3975/7/3/036001

Pan, H., Zhao, Y., Mei, J., Fang, Z., Wang, Y., Zhou, X., Zhou, Y., Zhang, R., Zhang, K., Jiang, L., Zeng, Q., He, Y., Wang, Z., Liu, Z., Xu, Q., Sun, Q., Yang, Y., Hu, Y., Chen, Y., … Guo, J. (2020). GCH1 variants contribute to the risk and earlier age-at-onset of Parkinson's disease: A two-cohort case-control study. *Translational Neurodegeneration*, *9*(1), 31. https://doi.org/10.1186/s40035-020-00212-3

Pandurangan, A. P., Ochoa-Montaño, B., Ascher, D. B., & Blundell, T. L. (2017). SDM: A server for predicting effects of mutations on protein stability. *Nucleic Acids Research*, *45*(W1), W229–W235. https://doi.org/10.1093/nar/gkx439

Parkinson, J. (2002). An Essay on the Shaking Palsy. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *14*(2), 223–236. https://doi.org/10.1176/jnp.14.2.223

Peng, C., Zeng, W., Su, J., Kuang, Y., He, Y., Zhao, S., Zhang, J., Ma, W., Bode, A. M., Dong, Z., & Chen, X. (2016). Cyclin dependent kinase 2 (CDK2) is a key mediator for EGF-induced cell transformation mediated through the ELK4/c-Fos signaling pathway. *Oncogene*, *35*(9), 1170–1179. https://doi.org/10.1038/onc.2015.175

Pereira, R., Oliveira, J., & Sousa, M. (2020). Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *Journal of Clinical Medicine*, *9*(1), 132. https://doi.org/10.3390/jcm9010132

Perico, N., Askenazi, D., Cortinovis, M., & Remuzzi, G. (2018). Maternal and environmental risk factors for neonatal AKI and its long-term consequences. *Nature Reviews Nephrology*, *14*(11), 688–703. https://doi.org/10.1038/s41581-018-0054-y

Persson, F., Söderhjelm, P., & Halle, B. (2018). The geometry of protein hydration. *The Journal of Chemical Physics*, *148*(21), 215101. https://doi.org/10.1063/1.5026744

Petrucelli, L., & Dickson, D. W. (2008). Chapter 3—Neuropathology of Parkinson's Disease11Supported by NIH grants P50-NS40256, P50-AG25711, P50-AG16574, P01-AG17216, and P01-AG03949. In R. Nass & S. Przedborski (Eds.), *Parkinson's Disease* (pp. 35–48). Academic Press. https://doi.org/10.1016/B978-0-12-374028-1.00003-8

Pietras, C. O., Vendelin, J., Anedda, F., Bruce, S., Adner, M., Sundman, L., Pulkkinen, V., Alenius, H., D'Amato, M., Söderhäll, C., & Kere, J. (2011). The asthma candidate gene NPSR1 mediates isoform specific downstream signalling. *BMC Pulmonary Medicine*, *11*(1), 39. https://doi.org/10.1186/1471-2466-11-39

Pires, D. E. V., & Ascher, D. B. (2016). mCSM-AB: A web server for predicting antibody–antigen affinity changes upon mutation with graph-based signatures. *Nucleic Acids Research*, *44*(W1), W469–W473. https://doi.org/10.1093/nar/gkw458

Pires, D. E. V., Ascher, D. B., & Blundell, T. L. (2014). mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, *30*(3), 335–342. https://doi.org/10.1093/bioinformatics/btt691

Pm, T.-H., Le, F., Ej, B., Rc, F., Mh, Y., Aj, B., Pj, B., & Jc, P. (2005). Expression of human membrane skeleton protein genes for protein 4.1 and betaIISigma2-spectrin assayed by real-time RT-PCR. *Cellular & Molecular Biology Letters*, *10*(1), 135–149.

Polymeropoulos, M. H., Higgins, J. J., Golbe, L. I., Johnson, W. G., Ide, S. E., Iorio, G. D., Sanges, G., Stenroos, E. S., Pho, L. T., Schaffer, A. A., Lazzarini, A. M., Nussbaum, R. L., & Duvoisin, R. C. (1996). Mapping of a Gene for Parkinson's Disease to Chromosome 4q21-q23. *Science*, *274*(5290), 1197–1199. https://doi.org/10.1126/science.274.5290.1197

Pontén, F., Jirström, K., & Uhlen, M. (2008). The Human Protein Atlas—A tool for pathology. *The Journal of Pathology*, *216*(4), 387–393. https://doi.org/10.1002/path.2440

Pontius, J., Richelle, J., & Wodak, S. J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. *Journal of Molecular Biology*, *264*(1), 121–136. https://doi.org/10.1006/jmbi.1996.0628

Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A. E., Halliday, G., Goetz, C. G., Gasser, T., Dubois, B., Chan, P., Bloem, B. R., Adler, C. H., & Deuschl, G. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders*, *30*(12), 1591–1601. https://doi.org/10.1002/mds.26424

Pringsheim, T., Jette, N., Frolkis, A., & Steeves, T. D. L. (2014). The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Movement Disorders: Official Journal of the Movement Disorder Society*, *29*(13), 1583–1590. https://doi.org/10.1002/mds.25945

Pulkkinen, V., Bruce, S., Rintahaka, J., Hodgson, U., Laitinen, T., Alenius, H., Kinnula, V. L., Myllärniemi, M., Matikainen, S., & Kere, J. (2010). ELMOD2, a candidate gene for idiopathic pulmonary fibrosis, regulates antiviral responses. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *24*(4), 1167–1177. https://doi.org/10.1096/fj.09-138545

Punia, S., Behari, M., Govindappa, S. T., Swaminath, P. V., Jayaram, S., Goyal, V., Muthane, U. B., Juyal, R. C., & Thelma, B. K. (2006). Absence/rarity of commonly reported LRRK2 mutations in Indian Parkinson's disease patients. *Neuroscience Letters*, *409*(2), 83–88. https://doi.org/10.1016/j.neulet.2006.04.052

Quadri, M., Fang, M., Picillo, M., Olgiati, S., Breedveld, G. J., Graafland, J., Wu, B., Xu, F., Erro, R., Amboni, M., Pappatà, S., Quarantelli, M., Annesi, G., Quattrone, A., Chien, H. F., Barbosa, E. R., Oostra, B. A., Barone, P., Wang, J., & Bonifati, V. (2013). Mutation in the SYNJ1 Gene Associated with Autosomal Recessive, Early-Onset Parkinsonism. *Human Mutation*, *34*(9), 1208–1215. https://doi.org/10.1002/humu.22373

Quan, L., Lv, Q., & Zhang, Y. (2016). STRUM: Structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics*, *32*(19), 2936–2946. https://doi.org/10.1093/bioinformatics/btw361

Raj, M. R., & Sreeja, A. (2018). Analysis of Computational Gene Prioritization Approaches. *Procedia Computer Science*, *143*, 395–410. https://doi.org/10.1016/j.procs.2018.10.411

Rajan, R., Divya, K. P., Kandadai, R. M., Yadav, R., Satagopam, V. P., Madhusoodanan, U. K., Agarwal, P., Kumar, N., Ferreira, T., Kumar, H., Sreeram Prasad, A. V., Shetty, K., Mehta, S., Desai, S., Kumar, S., Prashanth, L. K., Bhatt, M., Wadia, P., Ramalingam, S., … Sharma, M. (2020). Genetic Architecture of Parkinson's Disease in the Indian Population: Harnessing Genetic Diversity to Address Critical Gaps in Parkinson's Disease Research. *Frontiers in Neurology*, *11*. https://doi.org/10.3389/fneur.2020.00524

Ramaker, C., Marinus, J., Stiggelbout, A. M., & Hilten, B. J. van. (2002). Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. *Movement Disorders*, *17*(5), 867–876. https://doi.org/10.1002/mds.10248

Rascol, O., Brooks, D. J., Korczyn, A. D., De Deyn, P. P., Clarke, C. E., & Lang, A. E. (2000). A Five-Year Study of the Incidence of Dyskinesia in Patients with Early Parkinson's Disease Who Were Treated with Ropinirole or Levodopa. *New England Journal of Medicine*, *342*(20), 1484–1491. https://doi.org/10.1056/NEJM200005183422004

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Research*, *39*(17), e118–e118. https://doi.org/10.1093/nar/gkr407

Rickman, D. S., Pflueger, D., Moss, B., VanDoren, V. E., Chen, C. X., de la Taille, A., Kuefer, R., Tewari, A. K., Setlur, S. R., Demichelis, F., & Rubin, M. A. (2009). SLC45A3-ELK4 is a Novel and Frequent ETS Fusion Transcript in Prostate Cancer. *Cancer Research*, *69*(7), 2734–2738. https://doi.org/10.1158/0008-5472.CAN-08-4926

Ross, J. P., Dupre, N., Dauvilliers, Y., Strong, S., Ambalavanan, A., Spiegelman, D., Dionne-Laporte, A., Pourcher, E., Langlois, M., Boivin, M., Leblond, C. S., Dion, P. A., Rouleau, G. A., & Gan-Or, Z. (2016). Analysis of DNAJC13 mutations in French-Canadian/French cohort of Parkinson's disease. *Neurobiology of Aging*, *45*, 212.e13-212.e17. https://doi.org/10.1016/j.neurobiolaging.2016.04.023

Ruhle, V. (2007). *Berendsen and Nose-Hoover thermostats*. https://scholar.googleusercontent.com/scholar?q=cache:GA3iQjyLcJEJ:scholar.google.com/&hl=en&as_sdt=0,5&scioq=Nose-Hoover+temperature-coupling+method

Saikia, A., Bhattacharya, P., & Paul, S. (2018). Importance of dopamine in Parkinson's disease. *Adv. Tissue Eng. Regen. Med. Open Access*, *4*.

Saitoh, Y., Ohno, N., Yamauchi, J., Sakamoto, T., & Terada, N. (2017). Deficiency of a membrane skeletal protein, 4.1G, results in myelin abnormalities in the peripheral nervous system. *Histochemistry and Cell Biology*, *148*(6), 597–606. https://doi.org/10.1007/s00418-017-1600-6

Salgado, D., Desvignes, J.-P., Rai, G., Blanchard, A., Miltgen, M., Pinard, A., Lévy, N., Collod-Béroud, G., & Béroud, C. (2016). UMD-Predictor: A High-Throughput Sequencing Compliant System for Pathogenicity Prediction of any Human cDNA Substitution. *Human Mutation*, *37*(5), 439–446. https://doi.org/10.1002/humu.22965

Sbaragli, M., Bibi, L., Pittis, M. G., Balducci, C., Heikinheimo, P., Ricci, R., Antuzzi, D., Parini, R., Spaccini, L., Bembi, B., & Beccari, T. (2005). Identification and characterization of five novel MAN2B1 mutations in Italian patients with alpha-mannosidosis. *Human Mutation*, *25*(3), 320–320. https://doi.org/10.1002/humu.9310

Schapira, A. H. (2008). Mitochondria in the aetiology and pathogenesis of Parkinson's disease. *The Lancet Neurology*, *7*(1), 97–109. https://doi.org/10.1016/S1474-4422(07)70327-7

Schoenberg, B. S., Osuntokun, B. O., Adeuja, A. O., Bademosi, O., Nottidge, V., Anderson, D. W., & Haerer, A. F. (1988). Comparison of the prevalence of Parkinson's disease in black populations in the rural United States and in rural Nigeria: Door-to-door community studies. *Neurology*, *38*(4), 645–646. https://doi.org/10.1212/wnl.38.4.645

Schreglmann, S. R., & Houlden, H. (2016). VPS13C-Another Hint at Mitochondrial Dysfunction in Familial Parkinson's Disease. *Movement Disorders*, *31*(9), 1340. https://doi.org/10.1002/mds.26682

Schwarz, J. M., Cooper, D. N., Schuelke, M., & Seelow, D. (2014). MutationTaster2: Mutation prediction for the deep-sequencing age. *Nature Methods*, *11*(4), 361–362. https://doi.org/10.1038/nmeth.2890

Sebate, B., Cuttler, K., Cloete, R., Britz, M., Christoffels, A., Williams, M., Carr, J., & Bardien, S. (2021). Prioritization of candidate genes for a South African family with Parkinson's disease using in-silico tools. *PLOS ONE*, *16*(3), e0249324. https://doi.org/10.1371/journal.pone.0249324

Shamekhi Amiri, F. (2019). Intracellular organelles in health and kidney disease. *Néphrologie & Thérapeutique*, *15*(1), 9–21. https://doi.org/10.1016/j.nephro.2018.04.002

Sharrocks, A. D., Brown, A. L., Ling, Y., & Yates, P. R. (1997). The ETS-domain transcription factor family. *The International Journal of Biochemistry & Cell Biology*, *29*(12), 1371–1387. https://doi.org/10.1016/s1357-2725(97)00086-1

Shi, L., Du, X., Jiang, H., & Xie, J. (2017). Ghrelin and Neurodegenerative Disorders—A Review. *Molecular Neurobiology*, *54*(2), 1144–1155. https://doi.org/10.1007/s12035-016-9729-1

Shinzawa, K., Sumi, H., Ikawa, M., Matsuoka, Y., Okabe, M., Sakoda, S., & Tsujimoto, Y. (2008). Neuroaxonal Dystrophy Caused by Group VIA Phospholipase A2 Deficiency in Mice: A Model of Human Neurodegenerative Disease. *Journal of Neuroscience*, *28*(9), 2212–2220. https://doi.org/10.1523/JNEUROSCI.4354-07.2008

Shriner, D., Tekola-Ayele, F., Adeyemo, A., & Rotimi, C. N. (2014). Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. *Scientific Reports*, *4*, 6055. https://doi.org/10.1038/srep06055

Shriner, D., Tekola-Ayele, F., Adeyemo, A., & Rotimi, C. N. (2016). Ancient Human Migration after Out-of-Africa. *Scientific Reports*, *6*(1), 26565. https://doi.org/10.1038/srep26565

Shuikan, A., Alharbi, S. A., Alkhalifah, D. H. M., & Hozzein, W. N. (2019). High-throughput sequencing and metagenomic data analysis. In *Metagenomics-Basics, Methods and Applications*. IntechOpen.

Siekevitz, P. (1957). Powerhouse of the Cell. *Scientific American*, *197*(1), 131–144.

Sienes Bailo, P., Lahoz, R., Sánchez Marín, J. P., & Izquierdo Álvarez, S. (2020). Incidence of Huntington disease in a northeastern Spanish region: A 13-year retrospective study at tertiary care centre. *BMC Medical Genetics*, *21*. https://doi.org/10.1186/s12881-020-01174-z

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, *40*(W1), W452–W457. https://doi.org/10.1093/nar/gks539

Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J., & Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*, *20*(1), 473. https://doi.org/10.1186/s12859-019-3019-7

Stella, L., & Melchionna, S. (1998). Equilibration and sampling in molecular dynamics simulations of biomolecules. *The Journal of Chemical Physics*, *109*(23), 10115–10117. https://doi.org/10.1063/1.477703

Strauss, K. M., Martins, L. M., Plun-Favreau, H., Marx, F. P., Kautzmann, S., Berg, D., Gasser, T., Wszolek, Z., Müller, T., Bornemann, A., Wolburg, H., Downward, J., Riess, O., Schulz, J. B., & Krüger, R. (2005). Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease. *Human Molecular Genetics*, *14*(15), 2099–2111. https://doi.org/10.1093/hmg/ddi215

Studer, G., Rempfer, C., Waterhouse, A. M., Gumienny, R., Haas, J., & Schwede, T. (2020). QMEANDisCo—Distance constraints applied on model quality estimation. *Bioinformatics*, *36*(6), 1765–1771. https://doi.org/10.1093/bioinformatics/btz828

Sudhaman, S., Behari, M., Govindappa, S. T., Muthane, U. B., Juyal, R. C., & Thelma, B. K. (2013). VPS35 and EIF4G1 mutations are rare in Parkinson's disease among Indians. *Neurobiology of Aging*, *34*(10), 2442.e1-2442.e3. https://doi.org/10.1016/j.neurobiolaging.2013.04.025

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P., Jensen, L. J., & von Mering, C. (2017). The STRING database in

2017: Quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research*, *45*(Database issue), D362–D368. https://doi.org/10.1093/nar/gkw937

Tanaka, K., Suzuki, T., Hattori, N., & Mizuno, Y. (2004). Ubiquitin, proteasome and parkin. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1695*(1), 235–247. https://doi.org/10.1016/j.bbamcr.2004.09.026

Tang, H., & Thomas, P. D. (2016). Tools for Predicting the Functional Impact of Nonsynonymous Genetic Variation. *Genetics*, *203*(2), 635–647. https://doi.org/10.1534/genetics.116.190033

The UniProt Consortium. (2015). UniProt: A hub for protein information. *Nucleic Acids Research*, *43*(D1), D204–D212. https://doi.org/10.1093/nar/gku989

Thomas, P. D., Kejariwal, A., Campbell, M. J., Mi, H., Diemer, K., Guo, N., Ladunga, I., Ulitsky-Lazareva, B., Muruganujan, A., Rabkin, S., Vandergriff, J. A., & Doremieux, O. (2003). PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Research*, *31*(1), 334–341.

Thornton, E., & Vink, R. (2015). Substance P and its tachykinin NK1 receptor: A novel neuroprotective target for Parkinson's disease. *Neural Regeneration Research*, *10*(9), 1403–1405. https://doi.org/10.4103/1673-5374.165505

Tofaris, G. K. (2012). Lysosome-dependent pathways as a unifying theme in Parkinson's disease. *Movement Disorders*, *27*(11), 1364–1369. https://doi.org/10.1002/mds.25136

Tozzi, A., Tantucci, M., Marchi, S., Mazzocchetti, P., Morari, M., Pinton, P., Mancini, A., & Calabresi, P. (2018). Dopamine D2 receptor-mediated neuroprotection in a G2019S Lrrk2 genetic model of Parkinson's disease. *Cell Death & Disease*, *9*(2), 1–13. https://doi.org/10.1038/s41419-017-0221-2

Tran, N. T., Jakovlić, I., & Wang, W.-M. (2015). *In silico characterisation, homology modelling and structure-based functional annotation of blunt snout bream (Megalobrama amblycephala)*

Hsp70 and Hsc70 proteins. Journal of Animal Science and Technology; Korean Society of Animal Science and Technology. https://doi.org/10.1186/s40781-015-0077-x

Valente, E. M., Abou-Sleiman, P. M., Caputo, V., Muqit, M. M. K., Harvey, K., Gispert, S., Ali, Z., Turco, D. D., Bentivoglio, A. R., Healy, D. G., Albanese, A., Nussbaum, R., González-Maldonado, R., Deller, T., Salvi, S., Cortelli, P., Gilks, W. P., Latchman, D. S., Harvey, R. J., … Wood, N. W. (2004). Hereditary Early-Onset Parkinson's Disease Caused by Mutations in PINK1. *Science*, *304*(5674), 1158–1160. https://doi.org/10.1126/science.1096284

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., & Mackerell, A. D. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, *31*(4), 671–690. https://doi.org/10.1002/jcc.21367

Vecchione, A., Marchese, A., Henry, P., Rotin, D., & Morrione, A. (2003). The Grb10/Nedd4 Complex Regulates Ligand-Induced Ubiquitination and Stability of the Insulin-Like Growth Factor I Receptor. *Molecular and Cellular Biology*, *23*(9), 3363–3372. https://doi.org/10.1128/MCB.23.9.3363-3372.2003

Vilariño-Güell, C., Wider, C., Ross, O. A., Dachsel, J. C., Kachergus, J. M., Lincoln, S. J., Soto-Ortolaza, A. I., Cobb, S. A., Wilhoite, G. J., Bacon, J. A., Behrouz, B., Melrose, H. L., Hentati, E., Puschmann, A., Evans, D. M., Conibear, E., Wasserman, W. W., Aasly, J. O., Burkhard, P. R., … Farrer, M. J. (2011). VPS35 Mutations in Parkinson Disease. *The American Journal of Human Genetics*, *89*(1), 162–167. https://doi.org/10.1016/j.ajhg.2011.06.001

Volkmann, J. (2004). Deep Brain Stimulation for the Treatment of Parkinson's Disease. *Journal of Clinical Neurophysiology*, *21*(1), 6–17.

Wakabayashi, K., Tanji, K., Odagiri, S., Miki, Y., Mori, F., & Takahashi, H. (2013). The Lewy Body in Parkinson's Disease and Related Neurodegenerative Disorders. *Molecular Neurobiology*, *47*(2), 495–508. https://doi.org/10.1007/s12035-012-8280-y

Wang, Q., Oyarzabal, E., Wilson, B., Qian, L., & Hong, J.-S. (2015). Substance P enhances microglial density in the substantia nigra through neurokinin-1 receptor/NADPH oxidase-mediated chemotaxis in mice. *Clinical Science*, *129*(8), 757–767. https://doi.org/10.1042/CS20150008

Wang, X., Guo, G., Zhang, J., Aebez, N., Liu, Z., Liu, C.-F., Ross, C. A., & Smith, W. W. (2021). Mutant-TMEM230-induced neurodegeneration and impaired axonal mitochondrial transport. *Human Molecular Genetics*, *ddab128*. https://doi.org/10.1093/hmg/ddab128

Welburn, J. P. I., & Cheeseman, I. M. (2012). The microtubule-binding protein Cep170 promotes the targeting of the kinesin-13 depolymerase Kif2b to the mitotic spindle. *Molecular Biology of the Cell*, *23*(24), 4786–4795. https://doi.org/10.1091/mbc.E12-03-0214

Williams, D. R., Hadeed, A., al-Din, A. S. N., Wreikat, A.-L., & Lees, A. J. (2005). Kufor Rakeb Disease: Autosomal recessive, levodopa-responsive parkinsonism with pyramidal degeneration, supranuclear gaze palsy, and dementia. *Movement Disorders*, *20*(10), 1264–1271. https://doi.org/10.1002/mds.20511

Williams, U., Bandmann, O., & Walker, R. (2018). Parkinson's Disease in Sub-Saharan Africa: A Review of Epidemiology, Genetics and Access to Care. *Journal of Movement Disorders*, *11*(2), 53–64. https://doi.org/10.14802/jmd.17028

Wilson, G. R., Sim, J. C. H., McLean, C., Giannandrea, M., Galea, C. A., Riseley, J. R., Stephenson, S. E. M., Fitzpatrick, E., Haas, S. A., Pope, K., Hogan, K. J., Gregg, R. G., Bromhead, C. J., Wargowski, D. S., Lawrence, C. H., James, P. A., Churchyard, A., Gao, Y., Phelan, D. G., … Lockhart, P. J. (2014). Mutations in RAB39B Cause X-Linked Intellectual Disability and Early-Onset Parkinson Disease with α-Synuclein Pathology. *The American Journal of Human Genetics*, *95*(6), 729–735. https://doi.org/10.1016/j.ajhg.2014.10.015

Wu, C., Shang, X.-Q., You, Z.-P., Jin, Q.-F., Zhang, Y.-L., Zhou, Y., Zhang, Y.-Z., & Shi, K. (2020). TRIM59 Promotes Retinoblastoma Progression by Activating the p38-MAPK Signaling Pathway. *Investigative Ophthalmology & Visual Science*, *61*(10), 2. https://doi.org/10.1167/iovs.61.10.2

Xu, C., Chen, J., Xu, X., Zhang, Y., & Li, J. (2018). *Potential Therapeutic Drugs for Parkinson's Disease Based on Data Mining and Bioinformatics Analysis* [Research Article]. Parkinson’s Disease; Hindawi. https://doi.org/10.1155/2018/3464578

Yang, S., Weng, H., Chen, L., Guo, X., Parra, M., Conboy, J., Debnath, G., Lambert, A. J., Peters, L. L., Baines, A. J., Mohandas, N., & An, X. (2011). Lack of Protein 4.1G Causes Altered Expression and Localization of the Cell Adhesion Molecule Nectin-Like 4 in Testis and Can Cause Male Infertility. *Molecular and Cellular Biology*, *31*(11), 2276–2286. https://doi.org/10.1128/MCB.01105-10

Ying, H., Ji, L., Xu, Z., Fan, X., Tong, Y., Liu, H., Zhao, J., & Cai, X. (2020). TRIM59 promotes tumor growth in hepatocellular carcinoma and regulates the cell cycle by degradation of protein phosphatase 1B. *Cancer Letters*, *473*, 13–24. https://doi.org/10.1016/j.canlet.2019.12.030

Yuan, S., Chan, H. C. S., & Hu, Z. (2017). Using PyMOL as a platform for computational drug design. *WIREs Computational Molecular Science*, *7*(2), e1298. https://doi.org/10.1002/wcms.1298

Yue, P., Li, Z., & Moult, J. (2005). Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease. *Journal of Molecular Biology*, *353*(2), 459–473. https://doi.org/10.1016/j.jmb.2005.08.020

Zaichick, S. V., McGrath, K. M., & Caraveo, G. (2017). The role of Ca2+ signaling in Parkinson's disease. *Disease Models & Mechanisms*, *10*(5), 519–535. https://doi.org/10.1242/dmm.028738

Zhan, H.-D., Zhou, H.-Y., Sui, Y.-P., Du, X.-L., Wang, W., Dai, L., Sui, F., Huo, H.-R., & Jiang, T.-L. (2016). The rhizome of Gastrodia elata Blume – An ethnopharmacological review. *Journal of Ethnopharmacology*, *189*, 361–385. https://doi.org/10.1016/j.jep.2016.06.057

Zhang, G., Feenstra, B., Bacelis, J., Liu, X., Muglia, L. M., Juodakis, J., Miller, D. E., Litterman, N., Jiang, P.-P., Russell, L., Hinds, D. A., Hu, Y., Weirauch, M. T., Chen, X., Chavan, A. R., Wagner, G. P., Pavličev, M., Nnamani, M. C., Maziarz, J., … Muglia, L. J. (2017). Genetic

Associations with Gestational Length and Spontaneous Preterm Birth. *The New England Journal of Medicine*, *377*(12), 1156–1167. https://doi.org/10.1056/NEJMoa1612665

Zhang, P., Zhang, H., Wang, Y., Zhang, P., & Qi, Y. (2019). Tripartite Motif-Containing Protein 59 (TRIM59) Promotes Epithelial Ovarian Cancer Progression via the Focal Adhesion Kinase(FAK)/AKT/Matrix Metalloproteinase (MMP) Pathway. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, *25*, 3366–3373. https://doi.org/10.12659/MSM.916299

Zhang, Z. X., & Román, G. C. (1993). Worldwide occurrence of Parkinson's disease: An updated review. *Neuroepidemiology*, *12*(4), 195–208. https://doi.org/10.1159/000110318

Zhang, Z.-X., Dong, Z.-H., & Román, G. C. (2006). Early Descriptions of Parkinson Disease in Ancient China. *Archives of Neurology*, *63*(5), 782–784. https://doi.org/10.1001/archneur.63.5.782

Zheng, Y., Zhang, L., Xie, J., & Shi, L. (2021). The Emerging Role of Neuropeptides in Parkinson's Disease. *Frontiers in Aging Neuroscience*, *13*. https://doi.org/10.3389/fnagi.2021.646726

Ziegler, D. K. (1990). Headache: Public Health Problem. *Neurologic Clinics*, *8*(4), 781–791. https://doi.org/10.1016/S0733-8619(18)30316-5

Zimprich, A., Benet-Pagès, A., Struhal, W., Graf, E., Eck, S. H., Offman, M. N., Haubenberger, D., Spielberger, S., Schulte, E. C., Lichtner, P., Rossle, S. C., Klopp, N., Wolf, E., Seppi, K., Pirker, W., Presslauer, S., Mollenhauer, B., Katzenschlager, R., Foki, T., … Strom, T. M. (2011). A Mutation in VPS35, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease. *The American Journal of Human Genetics*, *89*(1), 168–175. https://doi.org/10.1016/j.ajhg.2011.06.008

Appendix

Link to simulation movies for protein systems; *NPSR1* (WT), *NPSR1* (p.V138I), *EPB41L2* (WT), and *EPB41L2* (p.R663C):

https://drive.google.com/drive/folders/15h3iv0yRGtb04kdaZxl4xbrqV1ztqVJE?usp=sharing

NPSR1 (WT): Protein has highly flexible beta sheets.

NPSR1 (p.V138I): Protein has more prominent loop regions that are highly flexible. Beta sheets are less flexible compared to the WT.

EPB41L2 (WT): Protein structure made up of alpha helices and loop regions.

EPB41L2 (p.R663C): Beta sheets present in the protein structure. More loop regions present, and fewer alpha helices are seen in the protein structure due to the introduction of the p.R663C variant.

Table S1: Genes linked to monogenic familial forms of PD.

Gene	Features/Traits	Locus & Position	Pathways Involved in	Disease Onset	References
AUTOSOMAL DOMINANT PD GENES					
SNCA	Psychiatric symptoms Moderate levodopa response Dementia Lewy body formation	PARK1,4 4q21	Chaperone-mediated autophagy pathway	Variable onset	Polymeropoulos et al., 1996 Karimi-Moghadam et al., 2018
LRRK2	Good response to levodopa Dementia	PARK 8 12q12	Mitochondrial fission/fusion	Late onset	Healy et al., 2008 Karimi-Moghadam et al., 2018
VPS35	Good response to levodopa Rare cognitive decline	PARK 17 16q11.22	Lysosomal-mediated proper protein quality control pathway	Late onset	Karimi-Moghadam et al., 2018 Vilarino-Guell et al., 2011
GCH1	Dopa-responsive dystonia	14q22.2	Folate biosynthesis, metabolic pathways, biosynthesis of cofactors	Juvenile onset	Pan et al., 2020
HTRA2	Mitochondrial dysfunctiion	PARK 13 2p13	Proteolysis of misfolded and damaged proteins	Late onset	Karimi-Moghadam et al., 2018 Strauss et al., 2005
GIGYF2	Not available (N/A)	PARK 11	IGF-1/Insulin receptor	Late	Langlais et al., 2004

Gene	Clinical features	PARK / Locus	Pathway/Function	Onset	References
		2q36	signalling pathway	onset	Vecctione et al., 2003 Karimi-Moghadam et al., 2018
DNAJC13	Lewy body formation Levodopa effective in early stages	PARK 21 3q22.1	Clathrin coating of vesicles	40 – 83 years	Vilarino-Guell et al., 2014 Appel-Cresswel et al., 2014 Gustavsson et al., 2015 Ross et al., 2016
TMEM230	N/A	20p13-p12.3	N/A	N/A	Wang et al., 2021
EIF4G1	Preserved cognition Good response to levodopa	PARK 18 3q27.1	MRNA translation initation	Late onset	Chartier—Harlin et al., 2011 Karimi-Moghadam et al., 2018
CHCHD2	Good response to levodopa	PARK 22 7p17.2	Mitochondrial metabolism regulation	Late onset	Karimi-Moghadam et al., 2018 Aras et al., 2015
RIC3	N/A	11p15.4	N/A	N/A	Ben-David et al., 2020
GBA	N/A	1q21	N/A	Unclear	Karimi-Moghadam et al., 2018
AUTOSOMAL RECESSIVE PD GENES					
PRKN	Slow progression Lewy bodies absent Dementia	PARK 2 6q25-27	PARKIN-mediated ubiquitylation pathway	Early onset (juvenile)	Lucking et al., 2000 Tanaka et al., 2004
PINK1	Slow progression Lewy bodies absent Dementia	PARK 6 1p35-p36	Mitochondrial pathway Mitophagy	Early onset (juvenile)	Valente et al., 2004 Ibanez et al., 2006
DJ-1	Psychotic symptoms Good response to levodopa	PARK 7 1p36	Transcriptional regulation Antioxidative stress reaction	Early onset	Ottolini et al., 2009 Bonifati et al., 2003
ATP13A2	Dementia Cognitive impairment Adequate response to levodopa	PARK 9 1p36	Lysosomal-mediated proper protein and mitochondrial quantity and quality control pathway	Severe early onset	Dehay et al., 2012 Gusdon et al., 2012 Tofaris 2012
PLA2G6	Dystonia-Parkinsonism Visual disturbance Dementia Initial good response to levodopa	PARK 14 22q12 -q13	Mitochondrial inner membrane remodelling processes	Early onset	Balsinde and Balboa 2005 Shinzawa et al., 2008

	Levodopa-induced dyskinesia				
FBOX7	Saccadic eye movement Appreciable response to levodopa	PARK 15 22q12 - q13	Ubiquitin-proteasome degradation pathway	Early onset	Nelson et al., 2013 Di Fonzo et al., 2009
DNAJC6	Wheelchair bound state Poor response to levodopa Mental retardation Seizures	PARK 19 1p31.3	Presynaptic endocytosisof clathrin-coated vesicles	Early onset	Edvardson et al., 2012 Koroglu et al., 2013 Kononenko and Hauke 2015
SYNJ1	Apraxia of eyelid opening (ALO) Cognitive decline	PARK 20 21q22.11	Endocytic recyclcing pathway	Early onset (juvenile)	Quadri et al., 2013 Krebs et al., 2013 Olgiati et al., 2014
VPS13C	Cognitive decline Autonomic dysfunction Good levodopa response within early stages	PARK 23 15q22.2	Co-operates with *PRKN/PINK1* pathway	Early onset	Lesage et al., 2016 Nalls et al., 2014 Schreglmann and Houlden 2016
SLC6A3	N/A	5p15.33	Synaptic vesicle cycle, dopaminergic synapse, Parkinson disease	N/A	Iyer et al., 2020
X-LINKED DISORDER WITH PARKINSONIM FEATURES					
RAB39B	N/A	Xq28	Pathways of neurodegeneration - multiple diseases	Early onset	Wilson et al., 2014

Abbreviations: *SNCA* = Synuclein alpha, *LRRK2* = Leucine-rich repeat kinase 2, *VPS35* = Vacuolar protein sorting ortholog 35, *GCH1* = GTP Cyclohydrolase 1, *HTRA2* = HtrA Serine Peptidase 2, *GIGYF2* = GRB10 Interacting GYF Protein 2, *DNAJC13* = DnaJ Heat Shock Protein Family Member C13, *TMEM230* = Transmembrane protein 230, *EIF4G1* = Eukaryotic translation initiation factor 4 gamma 1, *CHCHD2* = Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 2, *RIC3* = Resistance to inhibitors of cholinesterase 3, *GBA* = Glucocerebrosidase, *PRKN* = Parkin RBR E3 Ubiquitin Protein Ligase, *PINK1* = PTEN-induced kinase 1, *DJ-1* = Protein deglycase, *ATP13A2* = ATPase Cation Transporting 13A2, *PLA2G6* = Phospholipase A2 Group VI, *FBOX7* = F-box protein 7, *DNAJC6* = DnaJ Heat Shock Protein Family Member C6, *SYNJ1* = Synaptojanin 1, *VPS13C* = Vacuolar Protein Sorting 13 Homolog C.
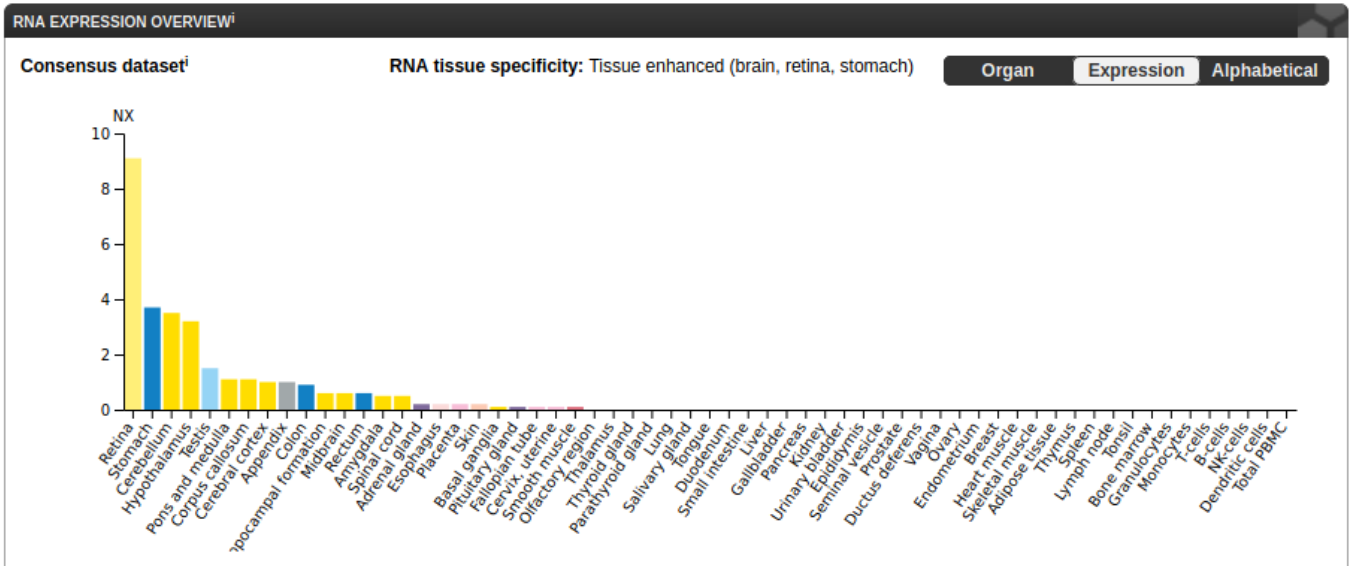
Figure S1: RNA tissue specificity expression data for the _NPSR1_ gene. Ranked according to expression levels in descending order (Source:https://www.proteinatlas.org/ENSG00000187258-NPSR1/tissue). Accessed: 28/07/2021
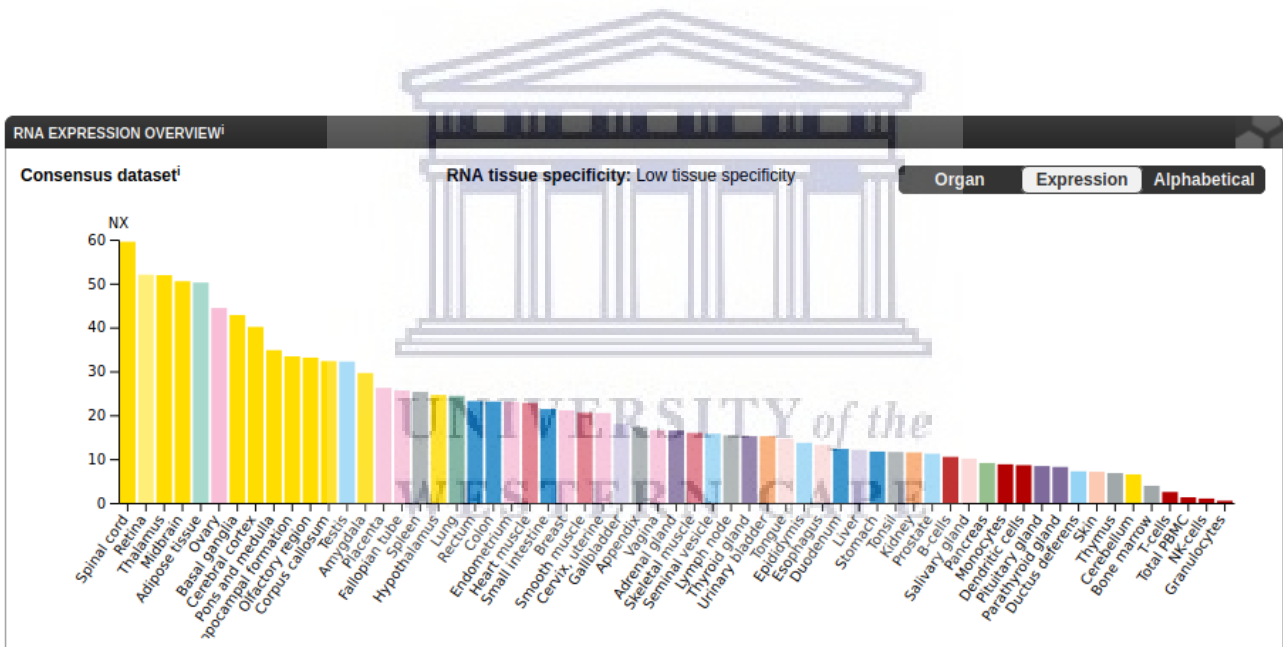
Figure S2: RNA tissue specificity expression data for the _EPB41L2_ gene. Ranked according to expression levels in descending order (Source: https://www.proteinatlas.org/ENSG00000079819-EPB41L2/tissue). Accessed: 28/07/2021

Table S2: Microarray data for gene expression (Allen Brain Atlas)

Gene Symbol	Structure	Probe name	Expression: Z-score	Log2 Intensity	Donor Information
ELK4	globus pallidus, external segment (GPe)	A_23_P161098	3.2543	7.79445	H0351.1009, 57 yrs, M, White or Caucasian
ELK4	corpus callosum (cc)	CUST_14695_PI416261804	3.70433	9.3546	H0351.1015, 49 yrs, F, Hispanic
ELK4	cingulum bundle (cgb)	CUST_14695_	3.17808	9.72976	H0351.2001, 24 yrs, M, Black

		PI416261804			or African American
CEP170	cingulum bundle (cgb)	A_23_P23151	3.41149	7.49535	H0351.2001, 24 yrs, M, Black or African American
CEP170	CA4 field (CA4)	A_23_P23151	3.57055	8.10207	H0351.1009, 57 yrs, M, White or Caucasian
CEP170	ventral thalamus (VT)	A_23_P23151	3.09337	7.75621	H0351.1016, 55 yrs, M, White or Caucasian
EEFSEC	lateral group of nuclei, dorsal division (DTLd)	A_23_P144202	4.649	5.67189	H0351.2001, 24 yrs, M, Black or African American
EEFSEC	nucleus subceruleus (SubC)	CUST_11498_PI416261804	5.5024	7.18942	H0351.1012, 31 yrs, M, White or Caucasian
EEFSEC	precentral gyrus (PrG)	CUST_11498_PI416261804	3.93401	5.74492	H0351.1015, 49 yrs, F, Hispanic
EEFSEC	locus ceruleus (LC)	CUST_11498_PI416261804	3.72774	5.60927	H0351.1015, 49 yrs, F, Hispanic
TRIM59	cingulum bundle (cgb)	A_23_P407718	3.19333	8.06338	H0351.2001, 24 yrs, M, Black or African American
HTT	IX (Cb-IX)	A_23_P212749	3.04422	10.7383	H0351.2002, 39 yrs, M, Black or African American
ELMOD2	I-II (Ve-I-II)	A_23_P305692	3.29953	4.20595	H0351.2001, 24 yrs, M, Black or African American
ELMOD2	VIIB (Ve-VIIB)	A_24_P385185	3.71756	3.9222	H0351.2001, 24 yrs, M, Black or African American
ELMOD2	V (Cb-V)	A_24_P385185	3.44992	3.80494	H0351.2001, 24 yrs, M, Black or African American
ELMOD2	inferior olivary complex (IO)	A_24_P385185	3.5154	3.83363	H0351.2001, 24 yrs, M, Black or African American
ELMOD2	paraventricular nuclei of thalamus (Pa)	A_23_P305692	3.59334	4.40736	H0351.2002, 39 yrs, M, Black or African American
ELMOD2	lateral hypothalamic area, anterior region (LHA)	A_24_P385185	3.21239	3.42527	H0351.2002, 39 yrs, M, Black or African American
ELMOD2	supraoptic nucleus (SO)	A_23_P305692	3.23039	4.22893	H0351.2002, 39 yrs, M, Black or African American
ELMOD2	substantia nigra (SN)	A_24_P385185	2.98152	3.32655	H0351.2002, 39 yrs, M, Black or African American
ELMOD2	trochlear nucleus (4)	A_24_P385185	3.82777	3.68837	H0351.2002, 39 yrs, M, Black or African American

ELMOD2	facial motor nucleus (7)	A_24_P38518 5	3.51889	3.55631	H0351.2002, 39 yrs, M, Black or African American
ELMOD2	CA4 field (CA4)	A_23_P30569 2	4.0832	4.78219	H0351.1009, 57 yrs, M, White or Caucasian
ELMOD2	supramarginal gyrus (SMG)	A_24_P38518 5	4.67612	4.45775	H0351.1009, 57 yrs, M, White or Caucasian
ELMOD2	claustrum (Cl)	A_24_P38518 5	3.34446	3.77329	H0351.1009, 57 yrs, M, White or Caucasian
ELMOD2	tuberomammillary nucleus (TM)	A_24_P38518 5	6.4376	5.36314	H0351.1009, 57 yrs, M, White or Caucasian
ELMOD2	posterior group of nuclei (DTP)	A_24_P38518 5	5.70108	4.4238	H0351.1012, 31 yrs, M, White or Caucasian
ELMOD2	cochlear nuclei (8Co)	A_24_P38518 5	3.21613	3.30532	H0351.1012, 31 yrs, M, White or Caucasian
ELMOD2	choroid plexus of the lateral ventricle (CPLV)	A_23_P30569 2	3.97306	4.74337	H0351.1012, 31 yrs, M, White or Caucasian
ELMOD2	parahippocampal gyrus (PHG)	A_24_P38518 5	3.86654	3.72909	H0351.1015, 49 yrs, F, Hispanic
ELMOD2	arcuate nucleus of the hypothalamus (ARH)	A_23_P30569 2	4.08247	4.96858	H0351.1015, 49 yrs, F, Hispanic
ELMOD2	superior colliculus (SC)	A_24_P38518 5	4.01029	3.80254	H0351.1015, 49 yrs, F, Hispanic
ELMOD2	globose nucleus (Glo)	A_24_P38518 5	3.11386	4.66866	H0351.1016, 55 yrs, M, White or Caucasian
EPB41L2	cingulum bundle (cgb)	CUST_10588_ PI416261804	3.07842	10.3154	H0351.2001, 24 yrs, M, Black or African American
EPB41L2	central glial substance (CGS)	CUST_10587_ PI416261804	3.43252	9.79024	H0351.1012, 31 yrs, M, White or Caucasian
NUP153	central glial substance (CGS)	A_23_P12225 4	3.23847	9.01955	H0351.2001, 24 yrs, M, Black or African American
NUP153	dentate gyrus (DG)	CUST_10891_ PI416261804	3.2699	5.92154	H0351.1016, 55 yrs, M, White or Caucasian
NPSR1	planum polare (PLP)	A_23_P19886	5.78725	4.42121	H0351.2001, 24 yrs, M, Black or African American
NPSR1	lateral group of nuclei,	CUST_476_PI	3.7146	4.38354	H0351.2001, 24 yrs, M, Black

121

	ventral division (DTLv)	416408490			or African American
NPSR1	medial group of nuclei (DTM)	CUST_476_PI 416408490	3.81673	4.44403	H0351.2001, 24 yrs, M, Black or African American
NPSR1	superior colliculus (SC)	A_23_P19886	4.9865	4.10303	H0351.2001, 24 yrs, M, Black or African American
NPSR1	red nucleus (RN)	CUST_16724_ PI416261804	4.03132	4.01214	H0351.2001, 24 yrs, M, Black or African American
NPSR1	pretectal region (PTec)	CUST_16724_ PI416261804	3.03403	3.63143	H0351.2002, 39 yrs, M, Black or African American
NPSR1	medial geniculate complex (MG)	A_23_P19886	5.10127	3.92499	H0351.2002, 39 yrs, M, Black or African American
NPSR1	trigeminal nuclei (5)	A_23_P19886	3.39098	3.23388	H0351.2002, 39 yrs, M, Black or African American
NPSR1	cuneate nucleus (Cu)	A_23_P19886	3.1588	3.14006	H0351.2002, 39 yrs, M, Black or African American
NPSR1	lateral hypothalamic area, anterior region (LHA)	CUST_476_PI 416408490	3.68097	3.73325	H0351.1009, 57 yrs, M, White or Caucasian
NPSR1	mammillary body (MB)	CUST_476_PI 416408490	4.28009	4.02847	H0351.1009, 57 yrs, M, White or Caucasian
NPSR1	caudal group of intralaminar nuclei (ILc)	CUST_16724_ PI416261804	3.05815	4.15417	H0351.1009, 57 yrs, M, White or Caucasian
NPSR1	lateral parabrachial nucleus (LPB)	CUST_16724_ PI416261804	3.76174	4.59551	H0351.1009, 57 yrs, M, White or Caucasian
NPSR1	pineal gland (PIN)	A_23_P19886	3.67069	3.01265	H0351.1012, 31 yrs, M, White or Caucasian
NPSR1	Lateral tuberal nucleus (LTu)	A_23_P19886	3.13643	2.80366	H0351.1012, 31 yrs, M, White or Caucasian
NPSR1	lateral hypothalamic area, mammillary region (LHM)	A_23_P19886	7.62803	4.56071	H0351.1012, 31 yrs, M, White or Caucasian
NPSR1	mammillary body (MB)	CUST_16724_ PI416261804	3.37439	4.03607	H0351.1012, 31 yrs, M, White or Caucasian
NPSR1	tuberomammillary nucleus (TM)	CUST_476_PI 416408490	3.17282	4.65152	H0351.1012, 31 yrs, M, White or Caucasian
NPSR1	preoptic region (PrOR)	CUST_476_PI 416408490	3.48437	4.92602	H0351.1012, 31 yrs, M, White or Caucasian
NPSR1	body of the caudate nucleus (BCd)	A_23_P19886	3.19862	3.13934	H0351.1015, 49 yrs, F, Hispanic
NPSR1	posterior hypothalamic	A_23_P19886	4.95779	4.08826	H0351.1015, 49 yrs, F,

	area (PHA)				Hispanic
NPSR1	parolfactory gyri (PaOG)	CUST_16724_PI416261804	6.12278	4.81824	H0351.1016, 55 yrs, M, White or Caucasian
NPSR1	subiculum (S)	A_23_P19886	3.52953	3.23112	H0351.1016, 55 yrs, M, White or Caucasian
NPSR1	nucleus subceruleus (SubC)	CUST_16724_PI416261804	3.46878	3.3031	H0351.1016, 55 yrs, M, White or Caucasian
NPSR1	Paramedian pontine reticular formation (PPRF)	A_23_P19886	5.08295	4.08545	H0351.1016, 55 yrs, M, White or Caucasian
LRRK1	Not available (N/A)	N/A	N/A	N/A	N/A
MAN2B1	lateral hypothalamic area, mammillary region (LHM)	CUST_151_PI417507815	5.32486	7.72601	H0351.2001, 24 yrs, M, Black or African American
MAN2B1	cingulum bundle (cgb)	A_24_P370156	3.50073	5.85948	H0351.2001, 24 yrs, M, Black or African American
MAN2B1	lateral group of nuclei, dorsal division (DTLd)	CUST_151_PI417507815	3.02976	5.56675	H0351.2002, 39 yrs, M, Black or African American
MAN2B1	corpus callosum (cc)	A_23_P27613	3.67153	11.6327	H0351.2002, 39 yrs, M, Black or African American
MAN2B1	choroid plexus of the lateral ventricle (CPLV)	A_23_P27613	5.12188	13.936	H0351.1009, 57 yrs, M, White or Caucasian
MAN2B1	trigeminal nuclei (5)	CUST_151_PI417507815	3.20617	5.42835	H0351.1012, 31 yrs, M, White or Caucasian
MAN2B1	central glial substance (CGS)	A_23_P27613	3.29846	11.7684	H0351.1012, 31 yrs, M, White or Caucasian
MAN2B1	globus pallidus, external segment (GPe)	A_24_P370156	3.57386	6.64669	H0351.1016, 55 yrs, M, White or Caucasian

Abbreviations: *ELK4*= ETS transcription factor, *CEP170*= Centrosomal protein 170-KD, *EEFSEC*= Eukaryotic elongation factor selenocysteine-tRNA-specific, *TRIM59*= Tripartite motif-containing protein 59, *HTT*= Huntingtin, *ELMOD2*= ELMO domain containing 2, *EPB41L2*= Erythrocyte membrane protein 4.1 like 2, *NUP153*= Nucleoporin 153-KD, NPSR1= Neuropeptide S receptor 1, *LRRK1*= Leucine-rich repeat kinase 1, *MAN2B1*= Mannosidase alpha class 2B member 1, R= Arginine, C= Cysteine, T= Threonine, M= Methionine, V= Valine, K= Lysine, H= Histidine, A= Alanine, S= Serine, I= Isoleucine.

Previous work:
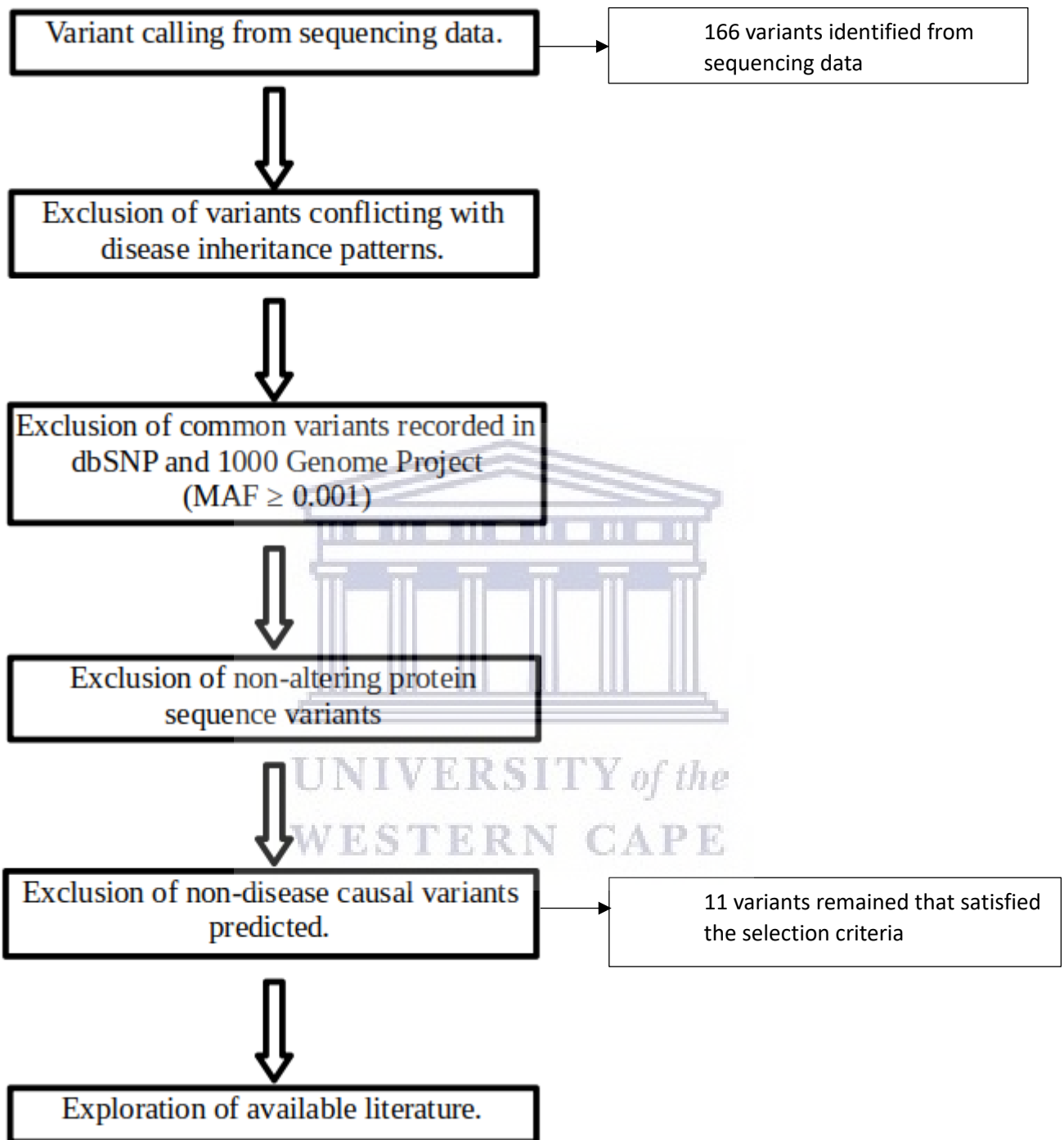
Figure S3: Bioinformatics pipeline. Steps followed by the MMED student to prioritize the 11 sequence variants.

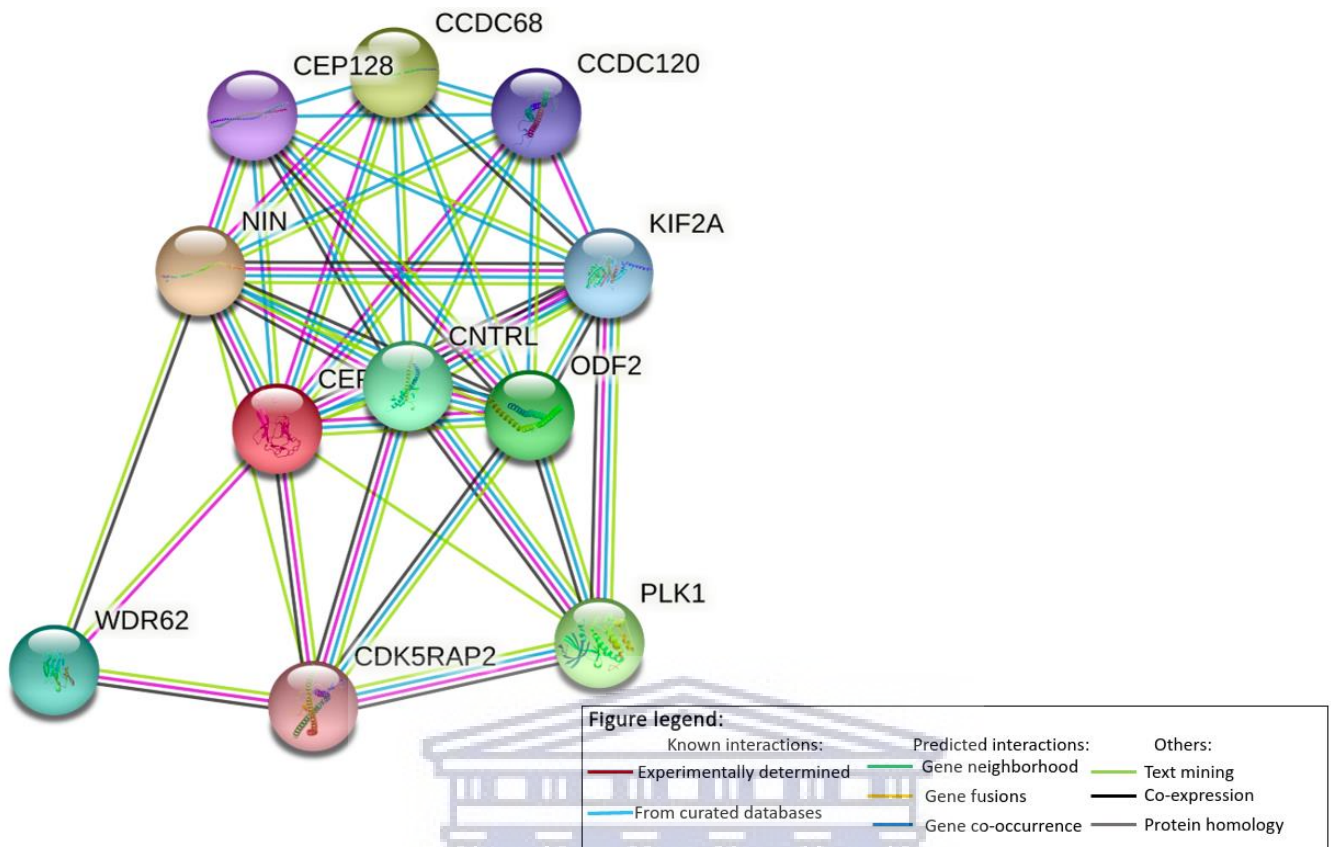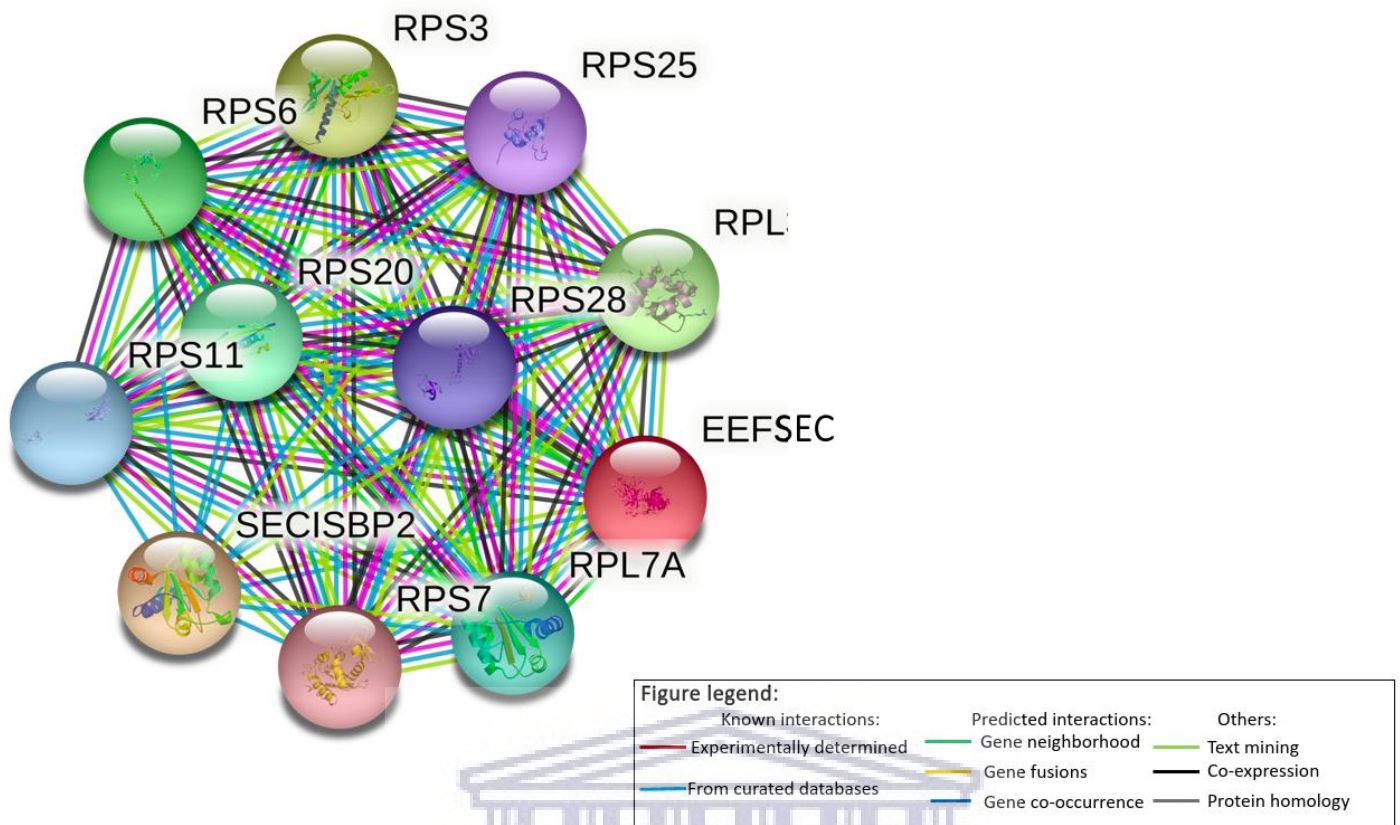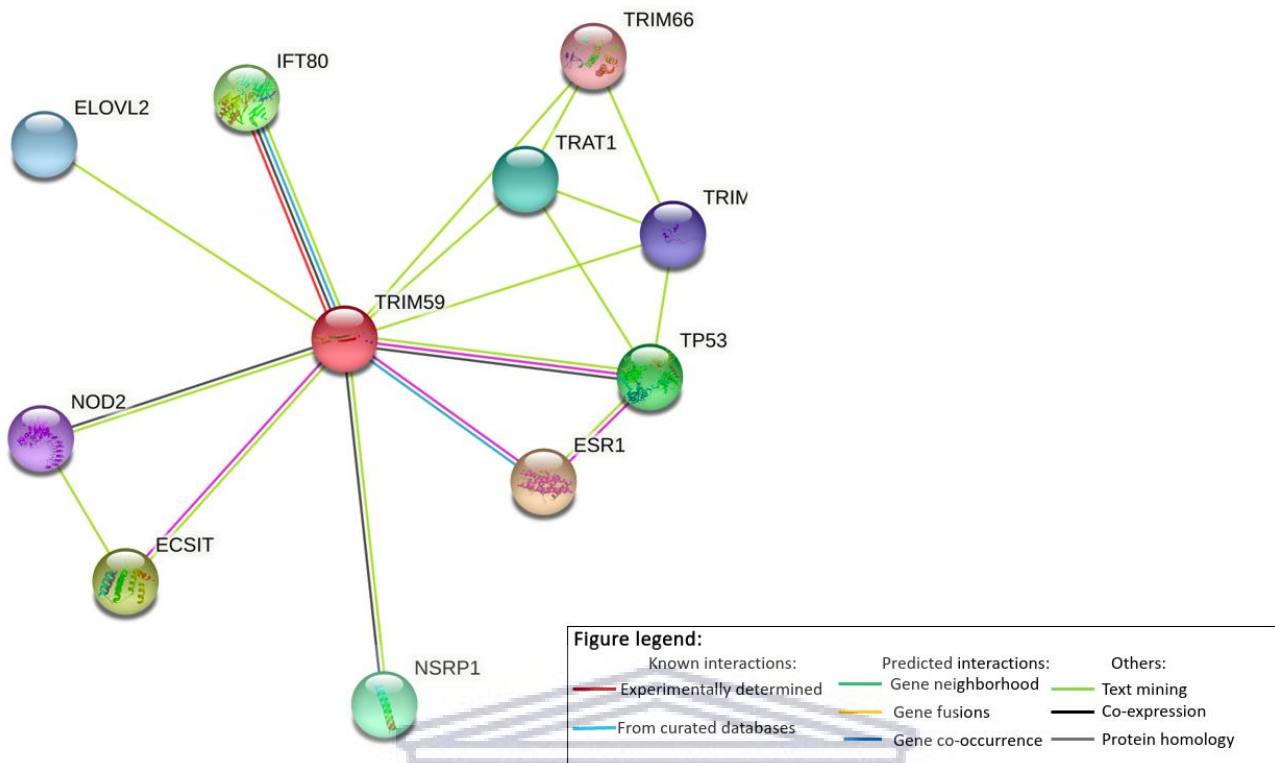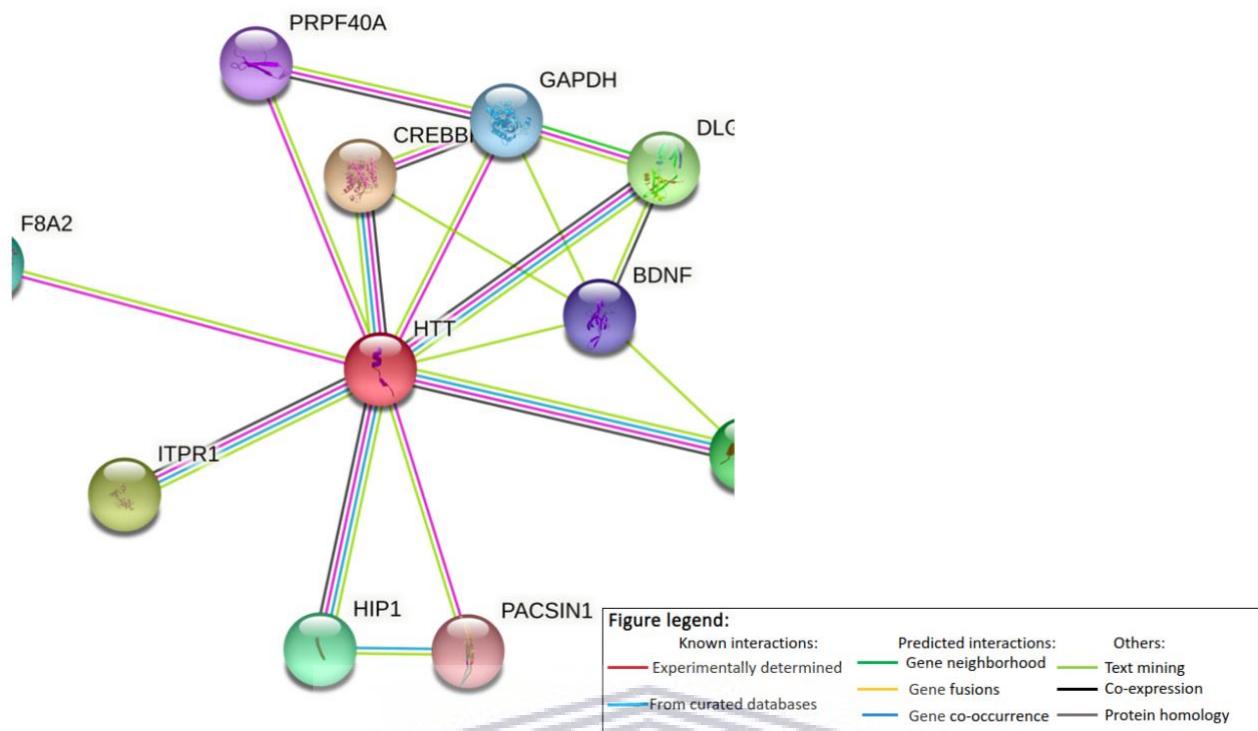**Current work:**
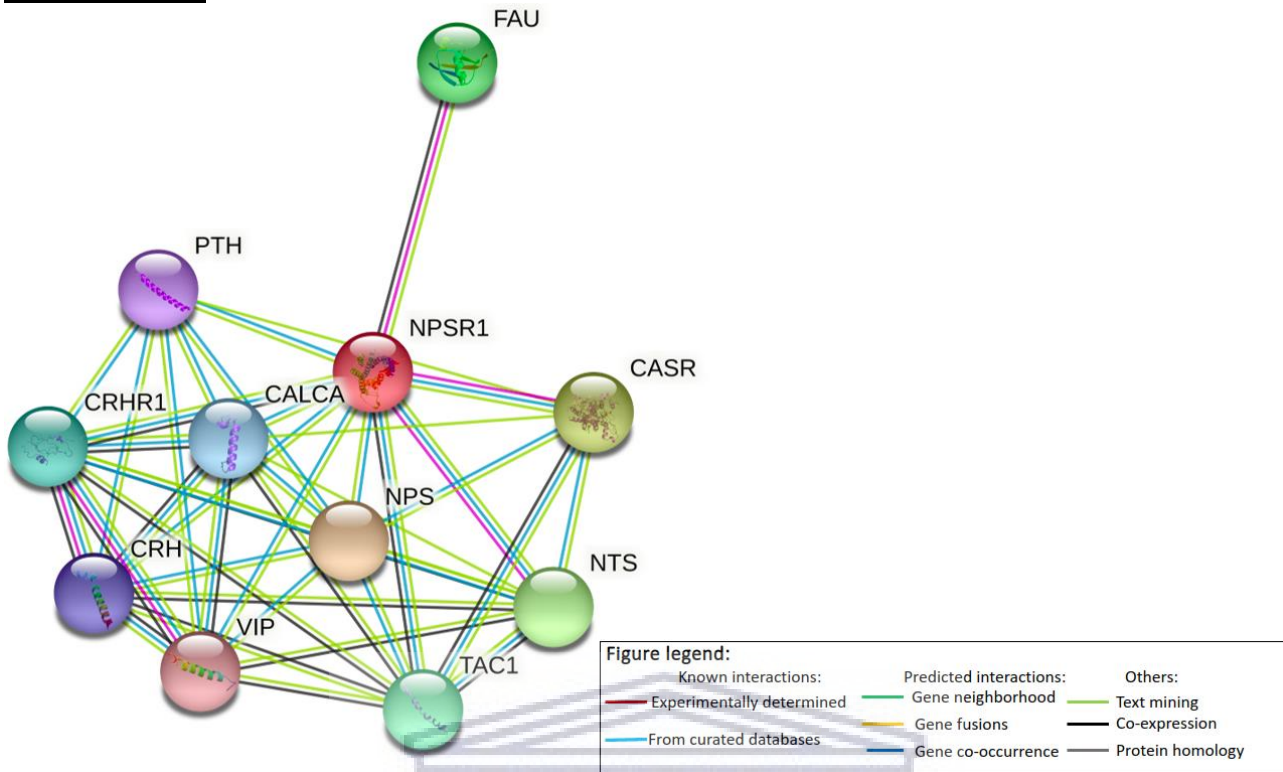
Figure S4: STRING analysis of *NPSR1* protein-protein interactions. Colored according to the type of interactions present; red/magenta represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=bqL9ERYTRRDC&sessionId=bbHQfln5VLM5). Accessed:28/07/2021
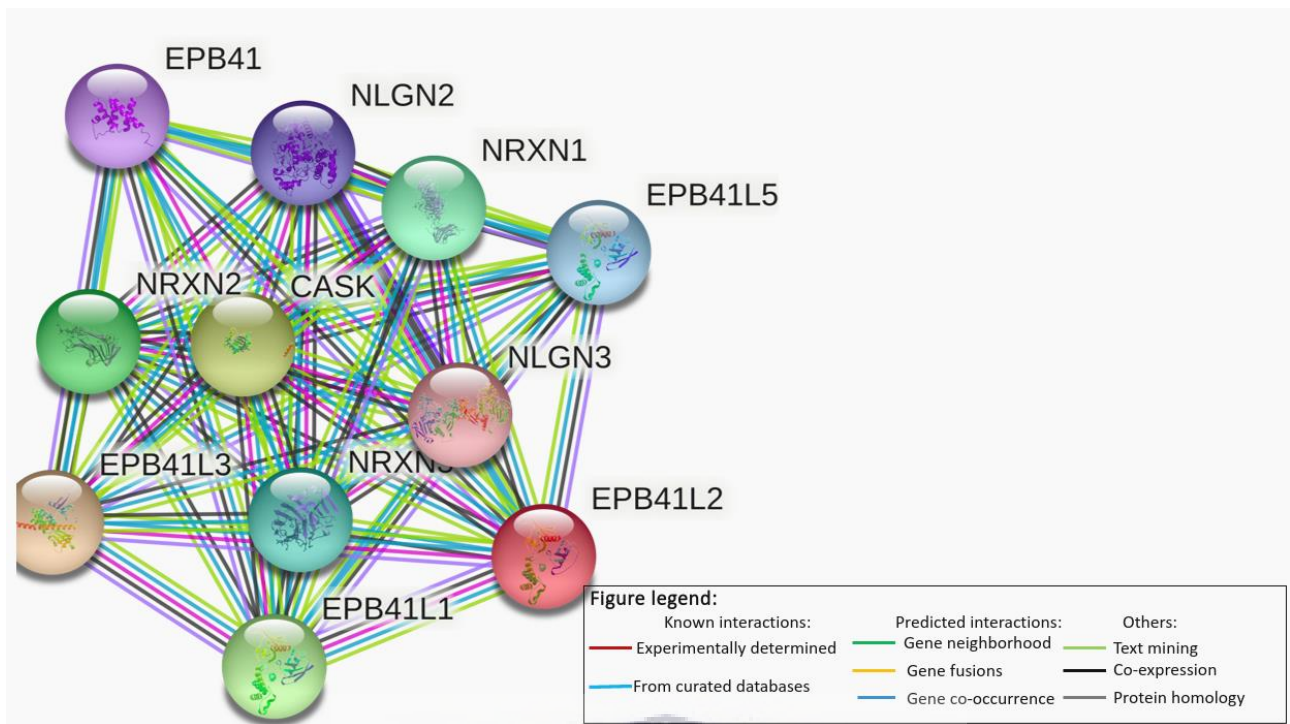
Figure S5: STRING analysis of *EPB41L2* protein-protein interactions. Colored according to the type of interactions present; red/magenta represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=bvbm1paWXg2g&sessionId=bbHQfln5VLM5). Accessed: 28/07/2021
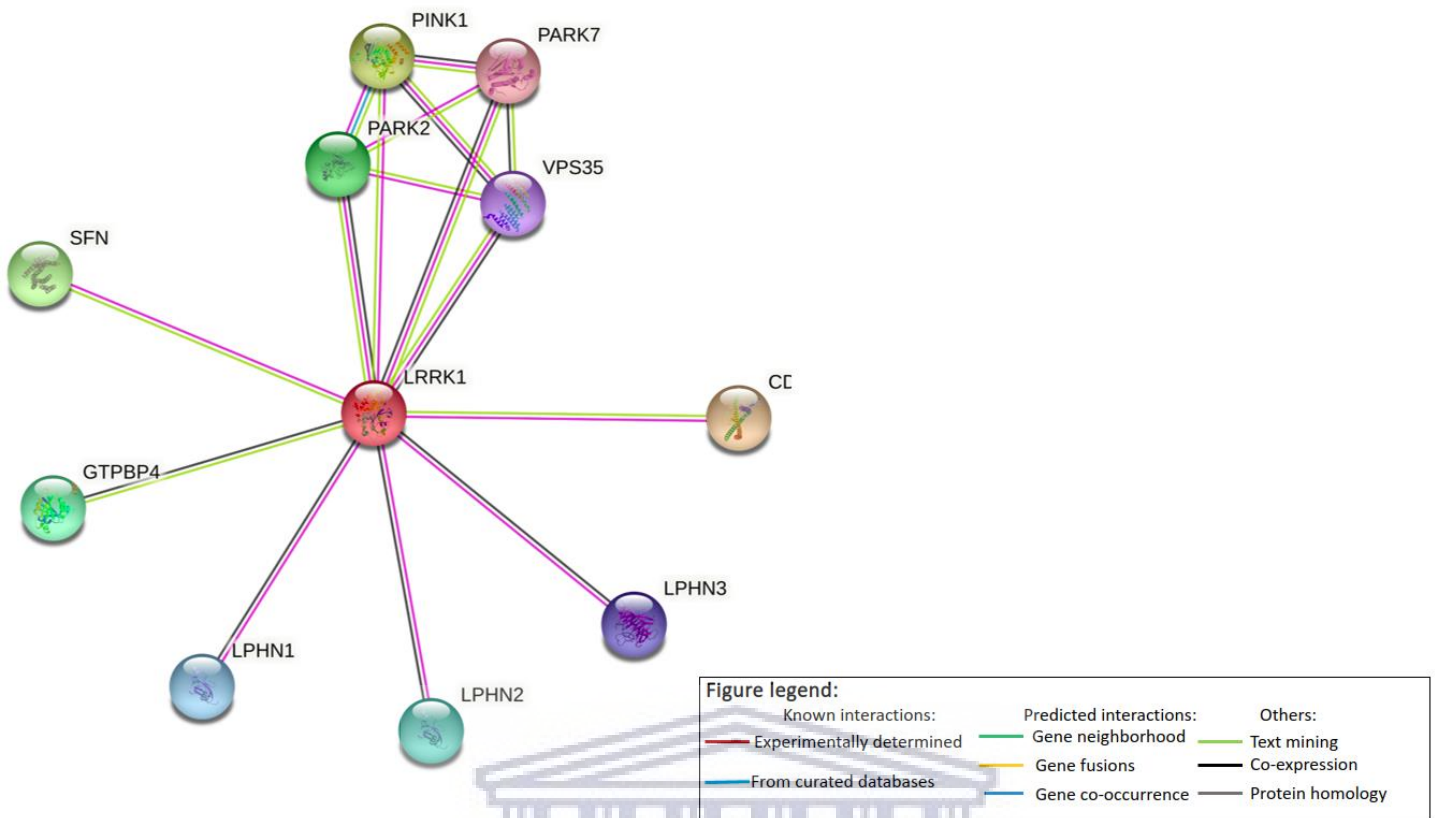
Figure S6: STRING analysis of *LRRK1* protein-protein interactions. LRRK1 interactions with PD-associated genes are illustrated, namely, VPS35, PARK7, PINK1, and PARK2. Colored according to the type of interactions present; red/magenta represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=b2AHy5Oeuwph&sessionId=bbHQfln5VLM5). Accessed: 28/07/2021
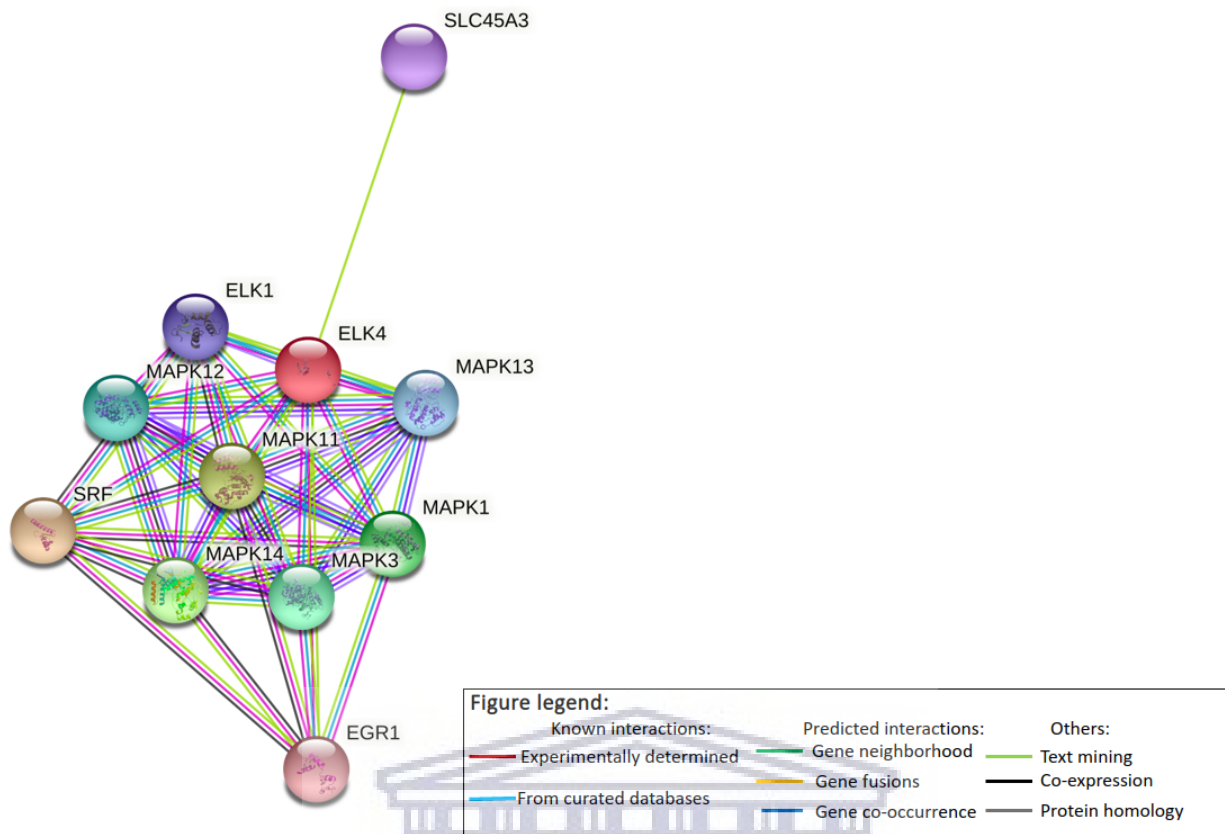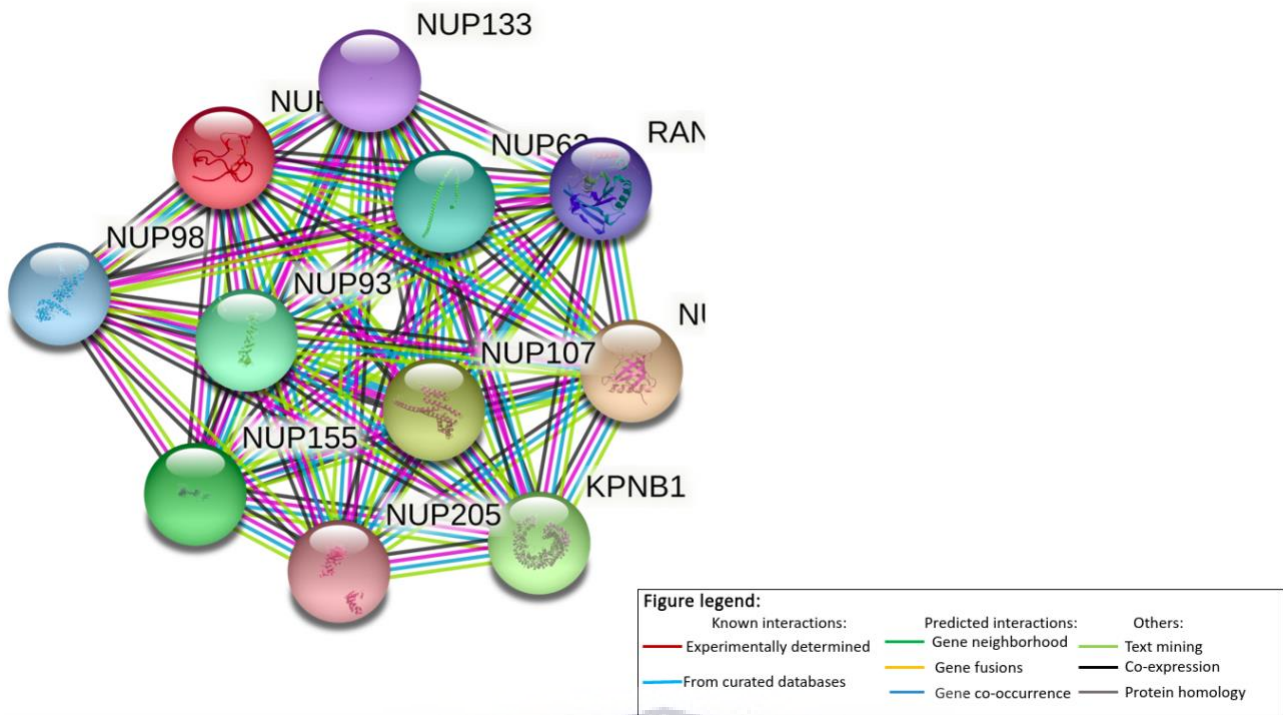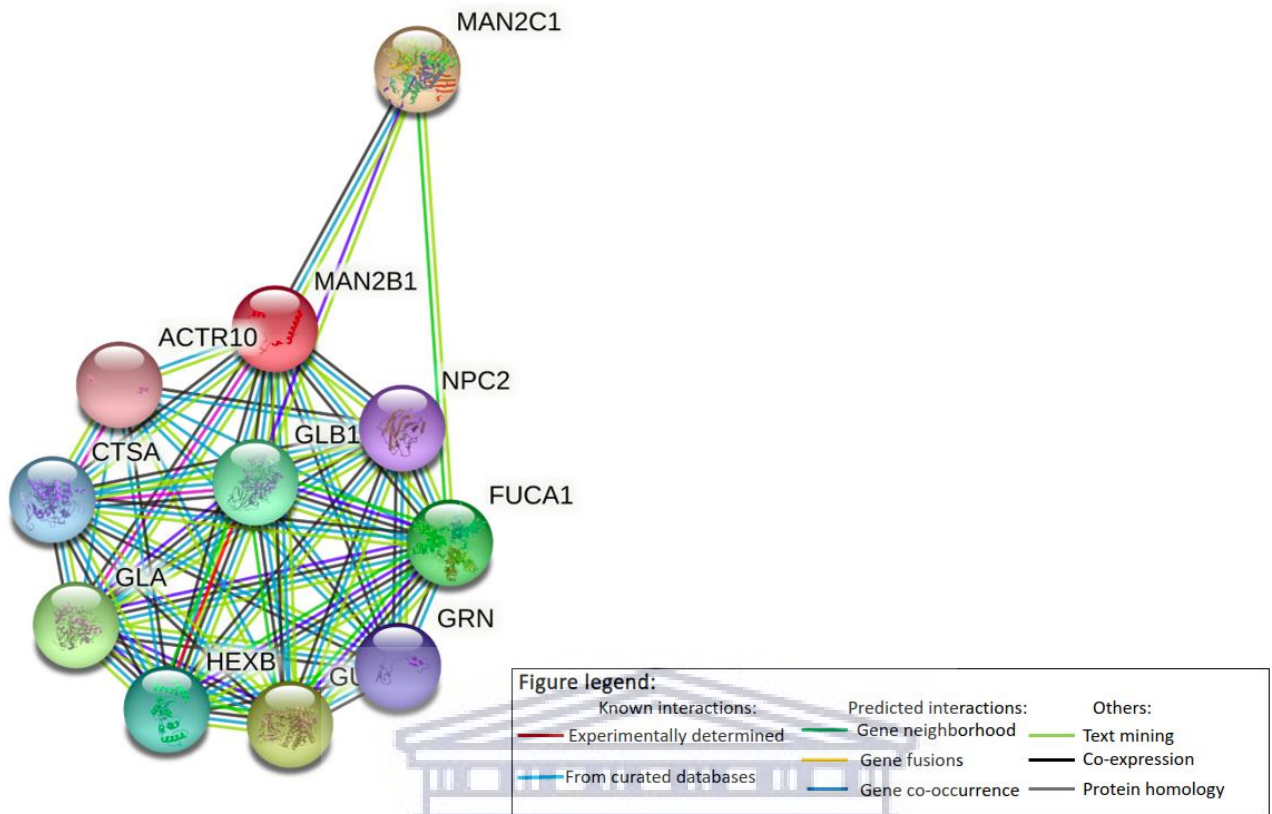
Figure S7: STRING analysis of _ELK4_ protein-protein interactions. Colored according to the type of interactions present; red represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=bZz5i2MqU3BP&sessionId=bbHQfln5VLM5). Accessed: 28/07/2021

Figure S8: STRING analysis of *CEP170* protein-protein interactions. Colored according to the type of interactions present; red represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=bD9zYB0zpKAU&sessionId=bbHQfln5VLM5). Accessed: 28/07/2021

Figure S9: STRING analysis of *EEFSEC* protein-protein interactions. Colored according to the type of interactions present; red represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=bAgt3M6YFLH3&sessionId=bbHQfln5VLM5). Accessed: 28/07/2021

Figure S10: STRING analysis of *TRIM59* protein-protein interactions. Colored according to the type of interactions present; red represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=b6dYF3030PjC&sessionId=bbHQfln5VLM5). Accessed: 28/07/2021

Figure S11: STRING analysis of *HTT* protein-protein interactions. Colored according to the type of interactions present; red represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=bNXc8On5lAoC&sessionId=bbHQfln5VLM5). Accessed: 28/07/2021
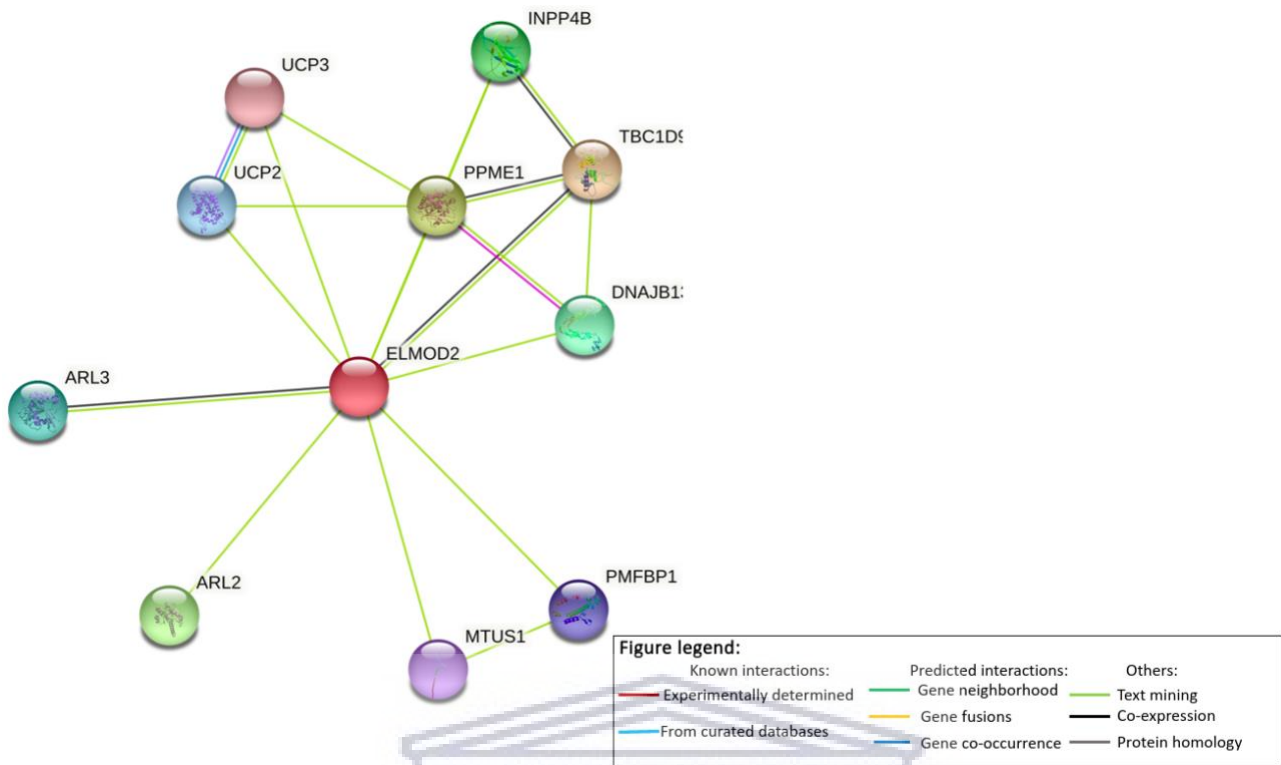
Figure S12: STRING analysis of *ELMOD2* protein-protein interactions. Colored according to the type of interactions present; red represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=bJhI0SVr3eQh&sessionId=bbHQfln5VLM5). Accessed: 28/07/2021

Figure S13: STRING analysis of *NUP153* protein-protein interactions. Colored according to the type of interactions present; red represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=bsx1ZDfMzFZc&sessionId=bbHQfln5VLM5). Accessed: 28/07/2021

Figure S14: STRING analysis of *MAN2B1* protein-protein interactions. Colored according to the type of interactions present; red/magenta represents interactions determined experimentally; turquoise represents interactions determined from curated databases. Predicted interactions are shown as green (gene neighborhood), yellow (gene fusions), and blue (gene co- occurrence). Other interactions determined by text mining (lime), co-expression (black), and protein homology (grey) are also shown. (Source: https://string-db.org/cgi/network?taskId=bWTwAe3g22NJ&sessionId=bbHQfln5VLM5). Accessed: 28/07/2021
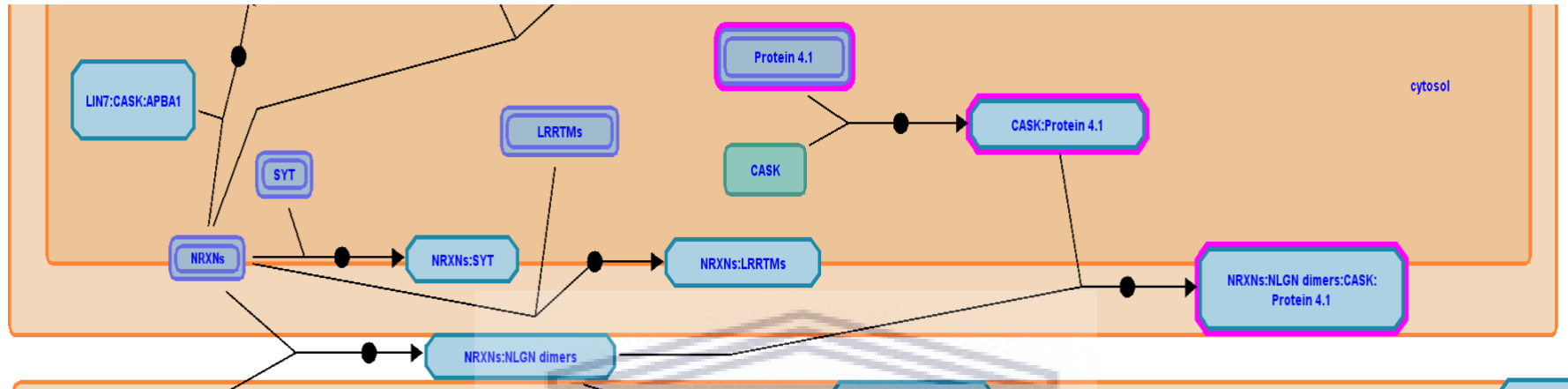
Figure S15: Pathway interaction data of the *EPB41L2* (Protein 4.1) gene. Interactions that *EPB41L2* (Protein 4.1) is involved in highlighted with pink border. Pathway information obtained from Reactome (Source: https://reactome.org/). Accessed: 28/07/2021
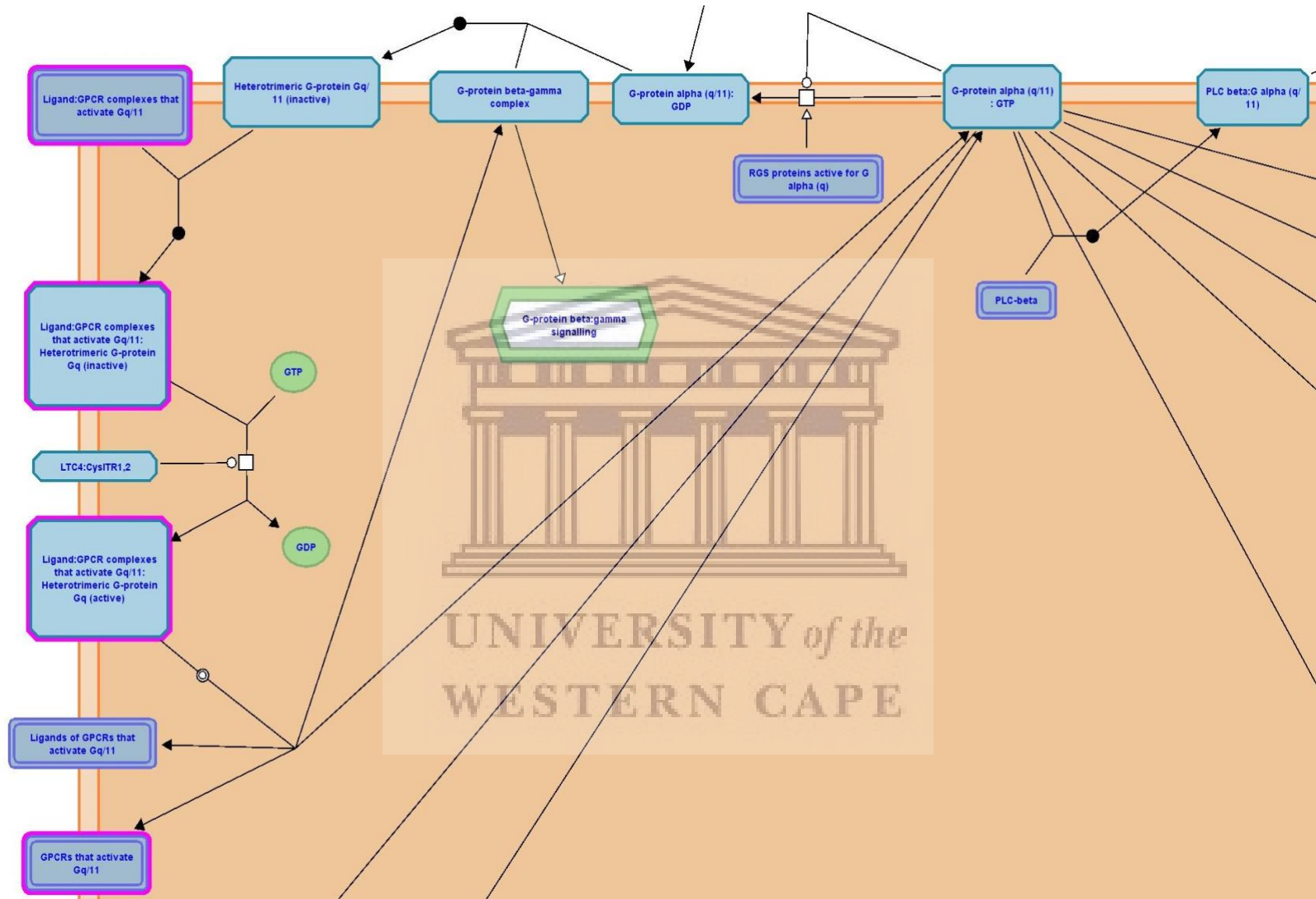
Figure S16: Pathway interaction data of the *NPSR1* (G-proteins) gene. Interactions that *NPSR1* (G-protein) is involved in are highlighted. Pathway information obtained from Reactome (Source: https://reactome.org/). Accessed: 28/07/2021

Figure S17: Sequence alignment of the *NPSR1* protein (Model_01) and its template (6j21.1.A). Colored according to hydrophobicity ranging from the least hydrophobic (darker regions) to the most hydrophobic (lighter regions). Mismatch residues are faded while conserved residues are shown in bold andgaps are shown as dashes.

Figure S18: Sequence alignment of the spectrin-actin-binding domain of the *EPB41L2* protein and its template (3F31). Colored according to hydrophobicity ranging from the least hydrophobic (darker regions) to the most hydrophobic (lighter regions). Mismatch residues are faded while conserved residues are shown in bold andaps are shown as dashes.

Table S3: NVT/NPT of the *NPSR1* wild-type system after EM and equilibration

	Average	Err. Est	RMSD	Tot-Drift
Potential	-2.34903×10^6	8200	17059.3	-57499.1 (kJ/mol)
Temperature	303.174	0.052	5.47481	-0.316022 (k)
Pressure	2.20459	1.1	68.0104	3.36789 (bar)
Density	1019.16	0.1	2.14414	0.768824 (kg/m^3)

Table S4: NVT/NPT of the *NPSR1* (p.V138I) variant system after EM and equilibration

	Average	Err. Est	RMSD	Tot-Drift
Potential	-2.32227×10^6	4600	9903.33	-32755.7 (kJ/mol)
Temperature	303.202	0.042	5.13256	-0.301053 (k)
Pressure	-0.194827	0.98	65.6804	-3.38891 (bar)
Density	971.096	0.13	1.48729	0.407359 (kg/m^3)

Table S5: NVT/NPT of the *EPB41L2* wild-type system after EM and equilibration.

	Average	Err. Est	RMSD	Tot-Drift
Potential	-1.35658×10^6	0.00	204.729	-499.125 (kJ/mol)
Temperature	303.224	0.04	5.68616	-0.266789 (k)
Pressure	1.97473	1	101.086	-3.71275 (bar)
Density	1025.52	0.16	3.38796	0.434789 (kg/m^3)

Table S6: NVT/NPT of the *EPB41L2* (p.R663C) variant system after EM and equilibration.

|	Average	Err. Est	RMSD	Tot-Drift
Potential	-1.35439×10^6	0.00	267.614	-650.875 (kJ/mol)
Temperature	303.178	0.026	4.6431	-0.182804 (k)
Pressure	1.43574	1.1	101.943	3.22227 (bar)
Density	1034.64	0.14	3.80982	0.893166 (kg/m^3)
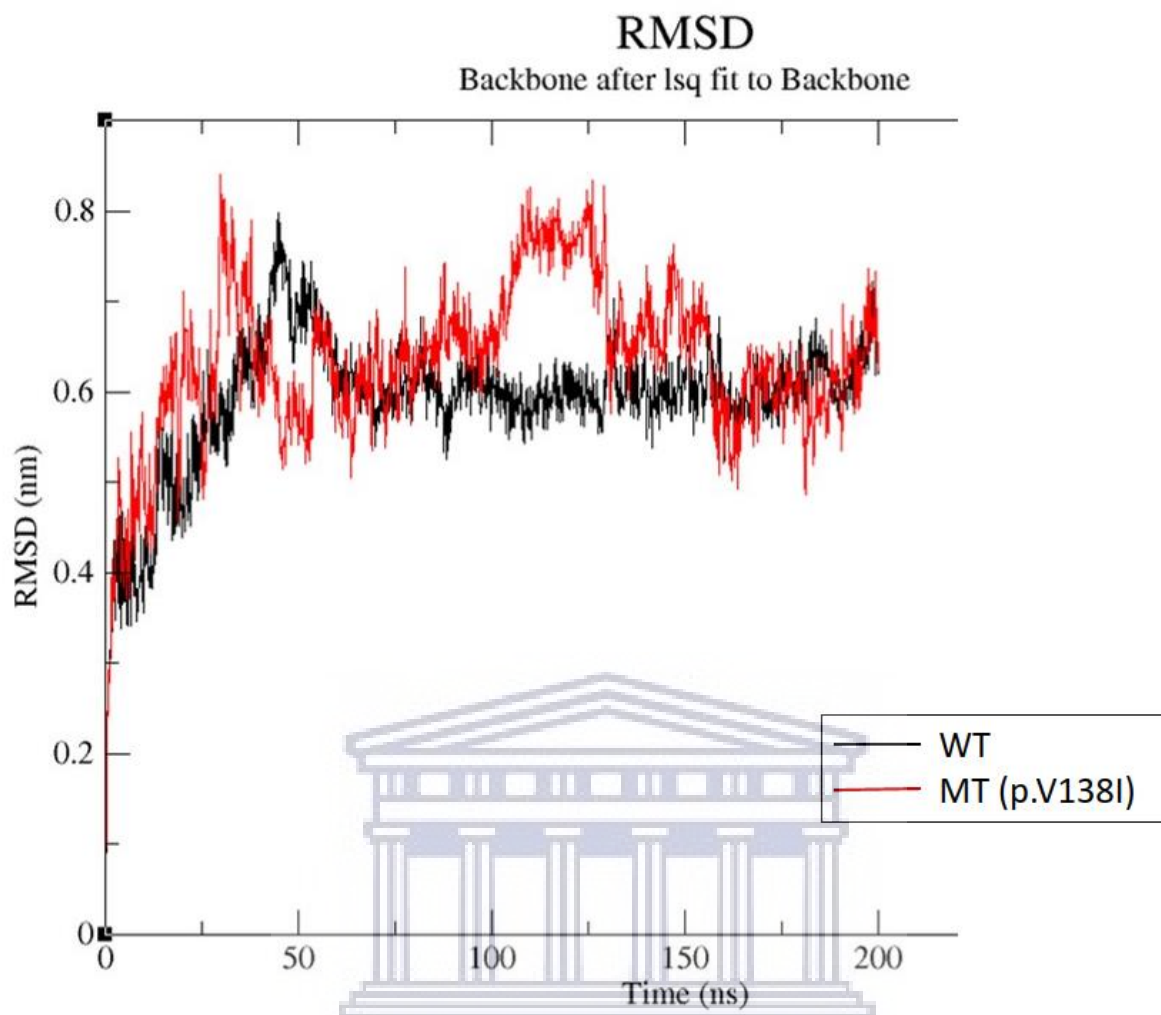
Figure S19: Backbone RMSD analysis of the WT *NPSR1* **protein and p.V138I MT systems (repeats) at 303 K over 200 ns of the simulation period.**
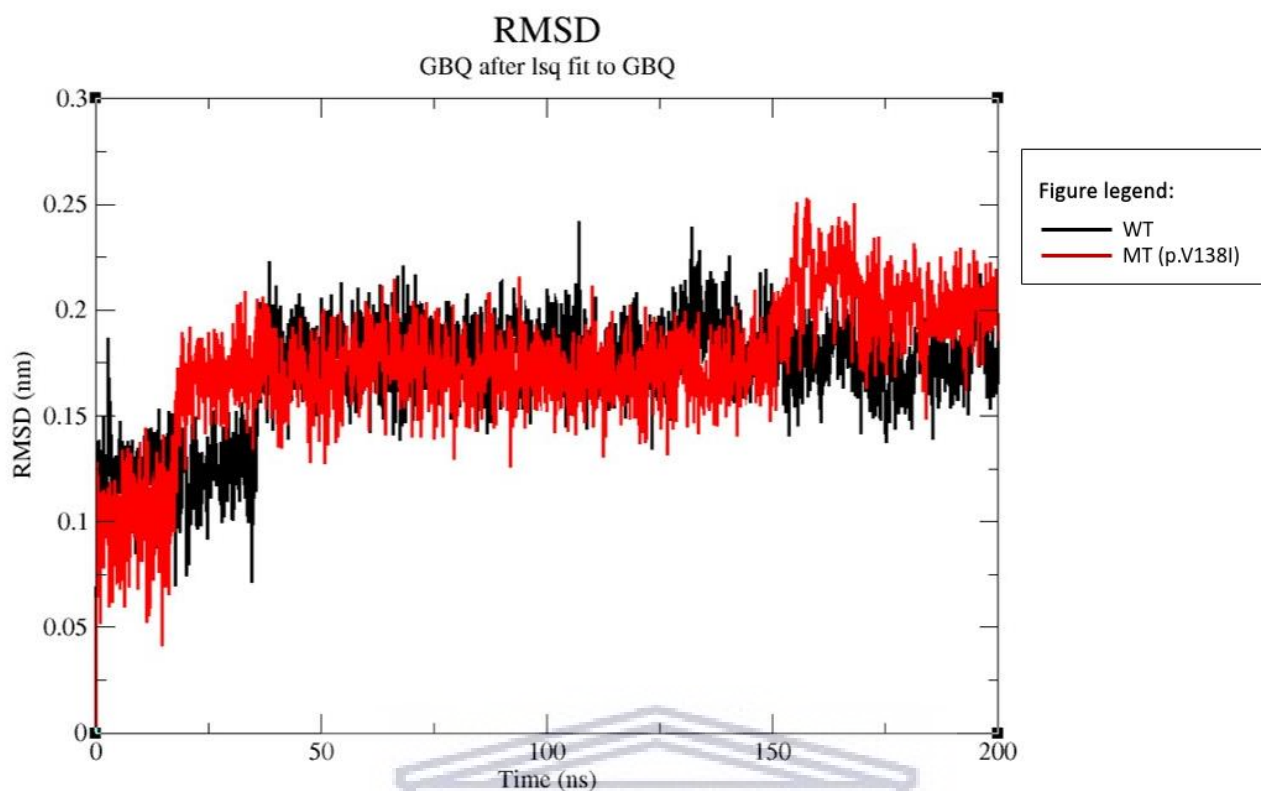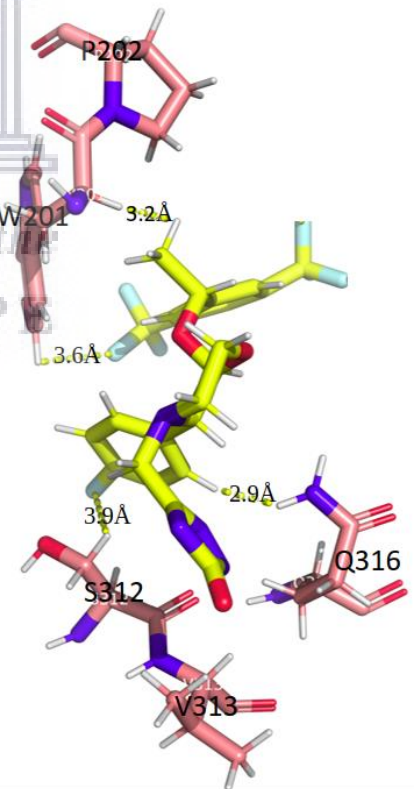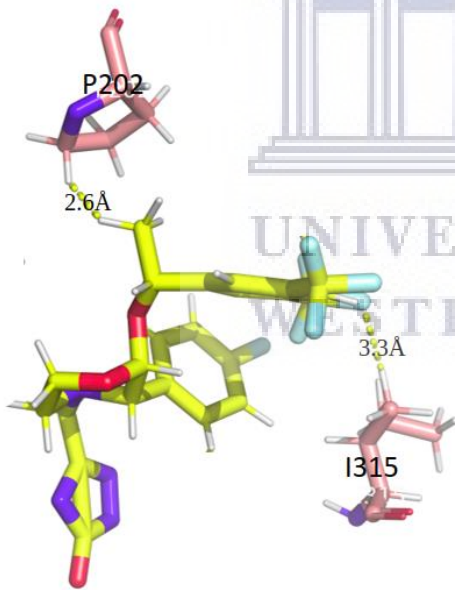
Figure S20: RMSD analysis of the GBQ ligand in the WT *NPSR1* protein and p.V138I MT systems at 303 K over 200 ns of the simulation period.
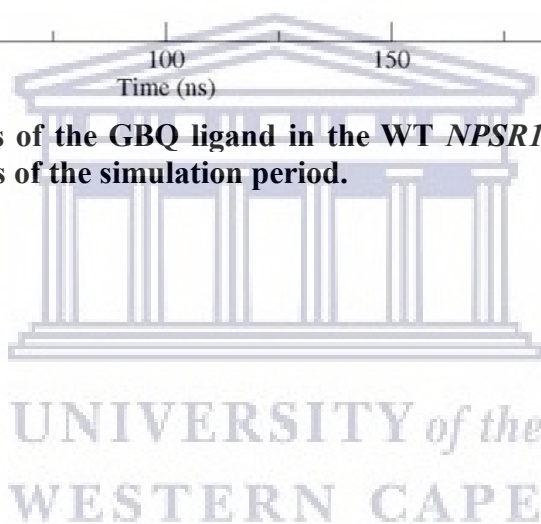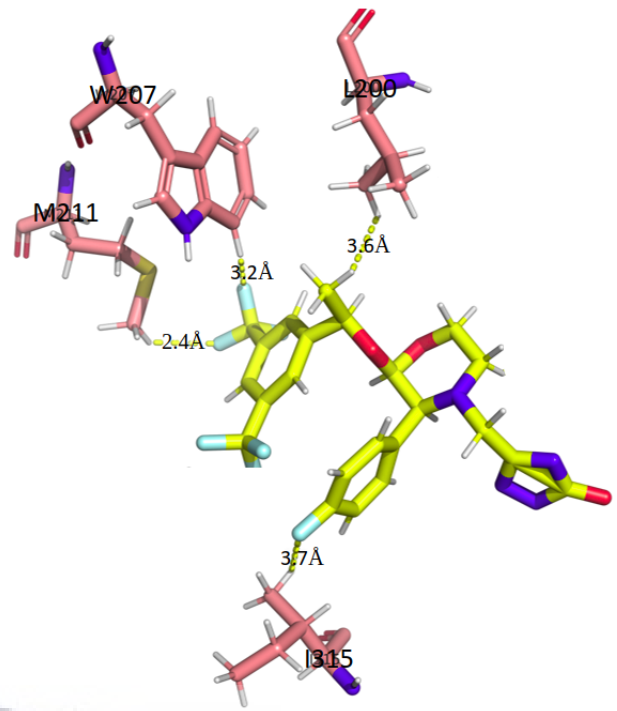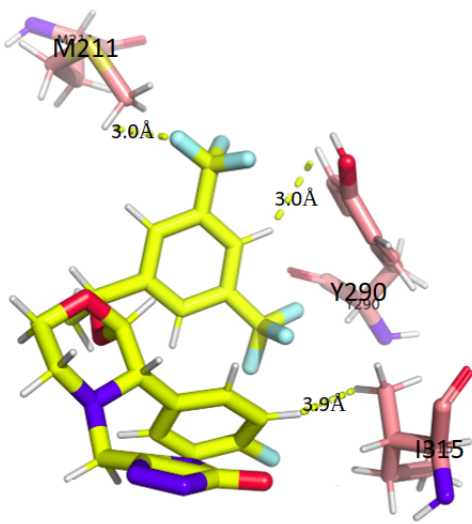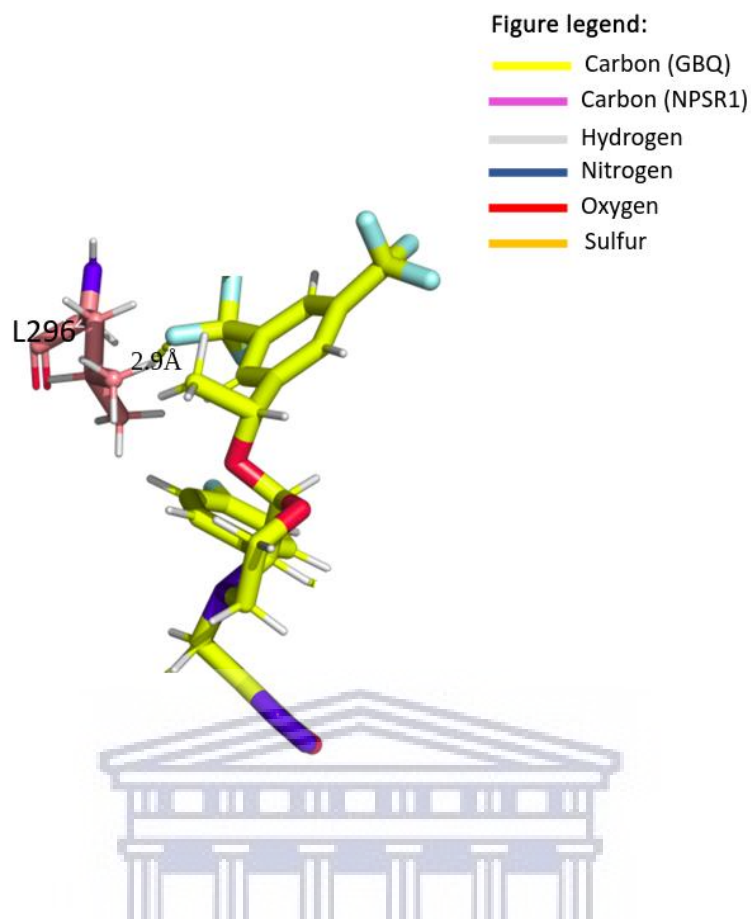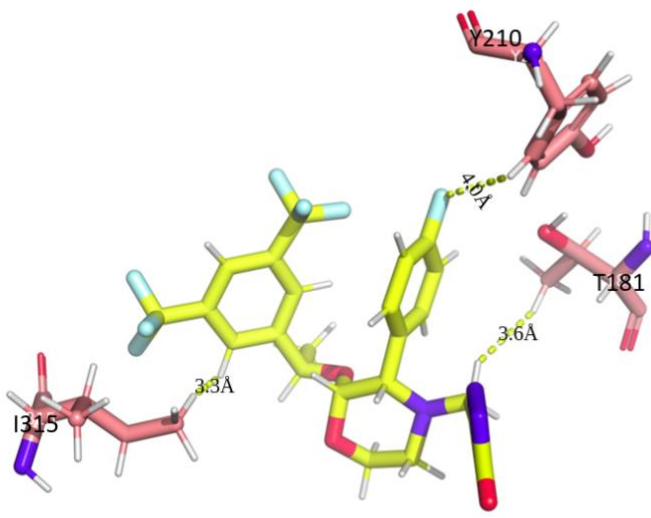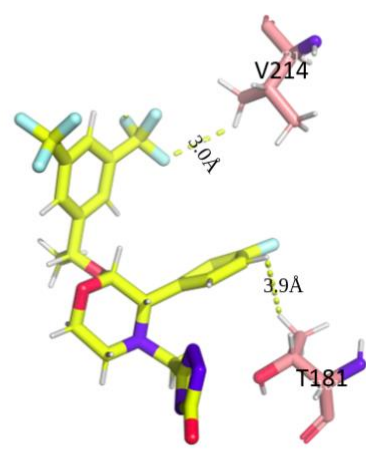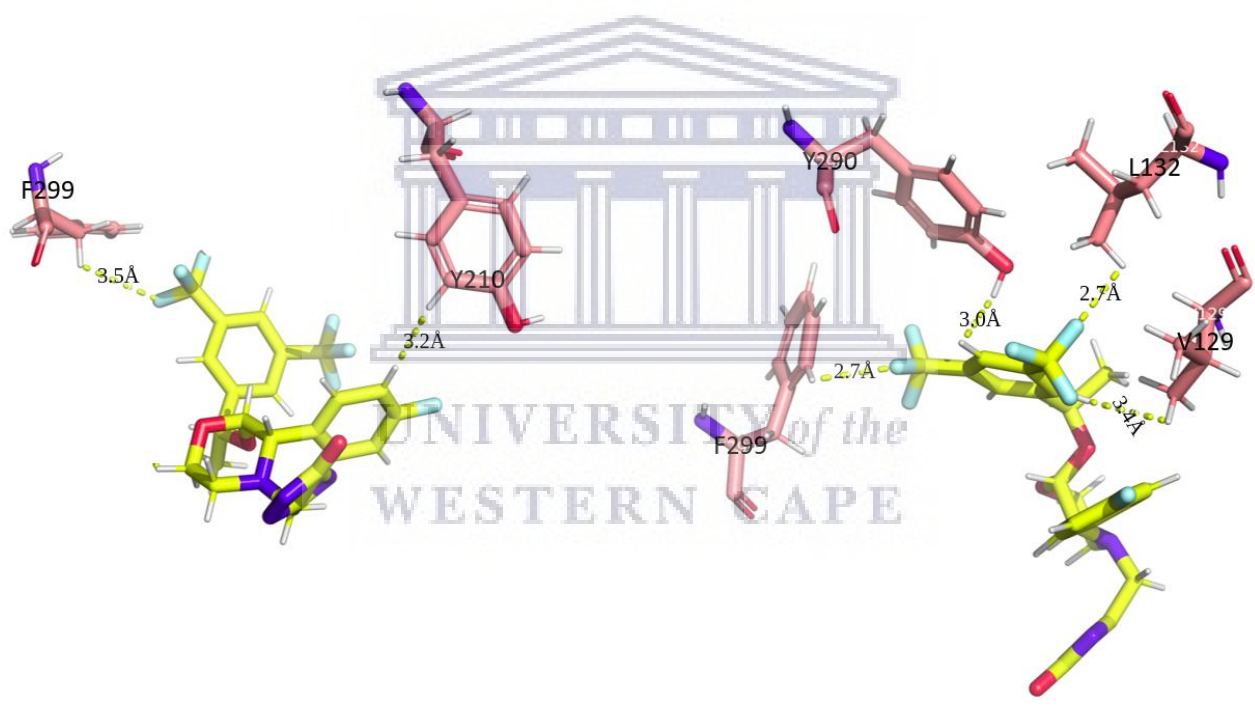
A

B

C

D

Figure S21: Snapshots of the GBQ ligand and the residues it interacts with as well as the polar contacts formed in the *NPSR1* WT protein system during the last 50 ns of the simulation. (A) 160 ns (B) 170 ns (C) 180 ns (D) 190 ns (E) 200 ns. Polar contact formed between the GBQ drug, which is colored according to atom type (carbon: yellow, hydrogen: white, nitrogen: blue, oxygen: red, sulfur: orange). Residues that it forms a polar contact with of the *NPSR1* WT protein labeled and colored according to atom type (carbon: magenta, hydrogen: white, nitrogen: blue, oxygen: red, sulfur: orange). Polar contacts shown as yellow dashes and labeled.
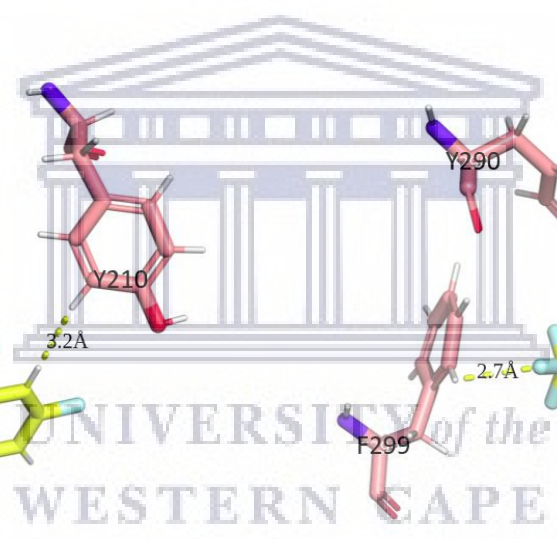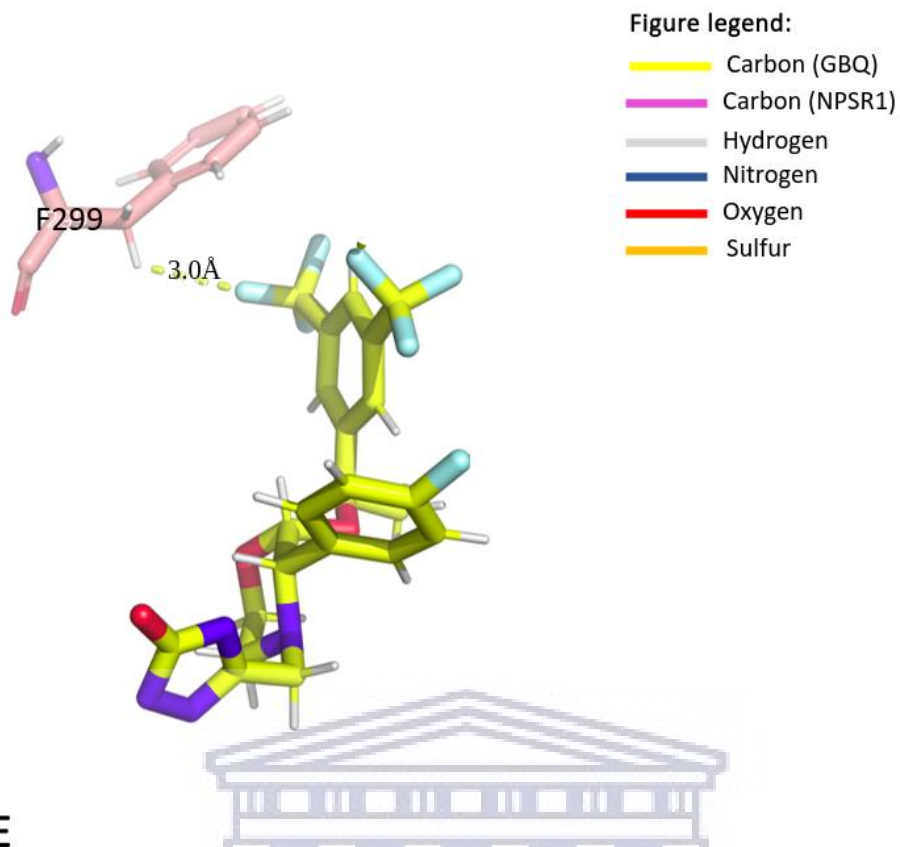
144

A

B

C

D

Figure S22: Snapshots of the GBQ ligand and the residues it interacts with as well as the polar contacts formed in the *NPSR1* MT (p.V138I) protein system during the last 50 ns of the simulation. (A) 160 ns (B) 170 ns (C) 180 ns (D) 190 ns (E) 200 ns. Polar contact formed between the GBQ drug, which is colored according to atom type (carbon: yellow, hydrogen: white, nitrogen: blue, oxygen: red, sulfur: orange). Residues that it forms a polar contact with of the *NPSR1* MT protein labeled and colored according to atom type (carbon: magenta, hydrogen: white, nitrogen: blue, oxygen: red, sulfur: orange). Polar contacts shown as yellow dashes and labeled.
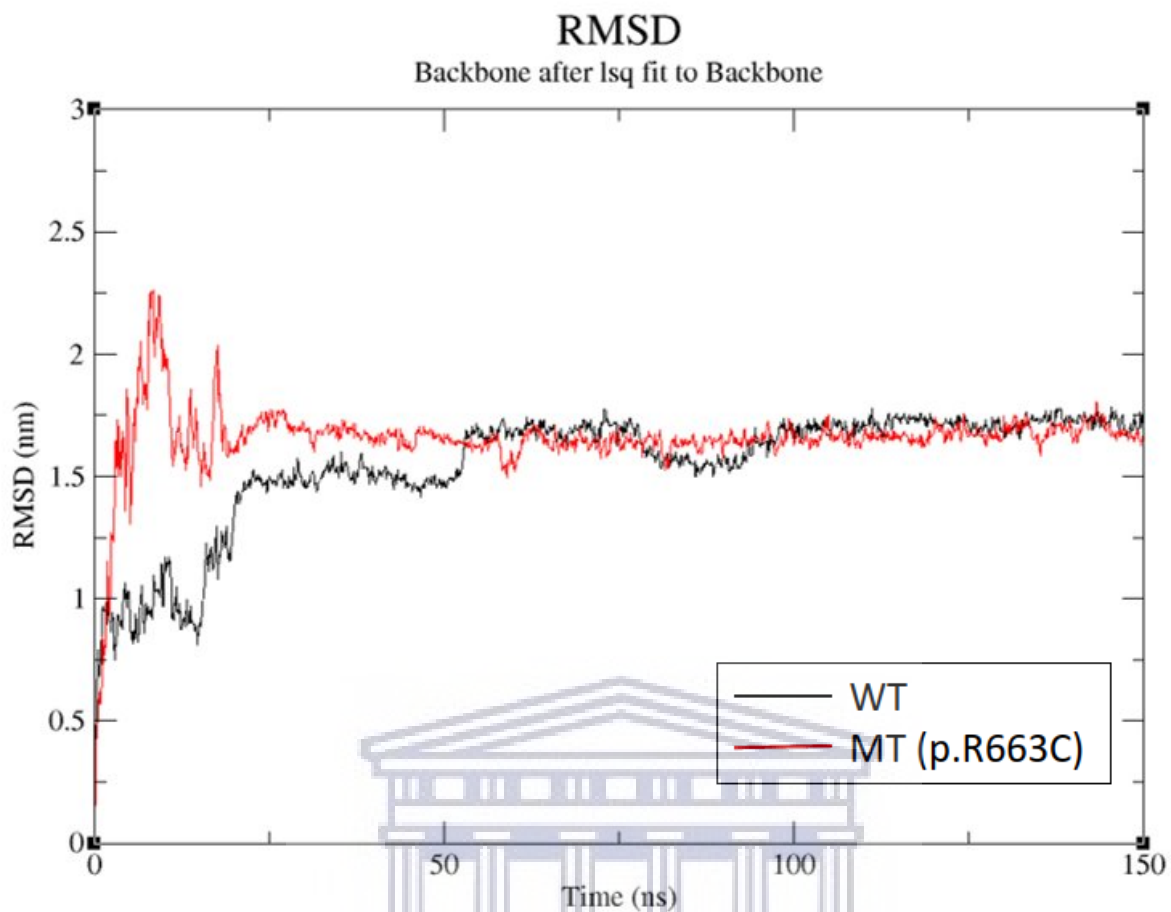
Figure legend:
— NPSR1 WT
— NPSR1 MT (p.V138I)
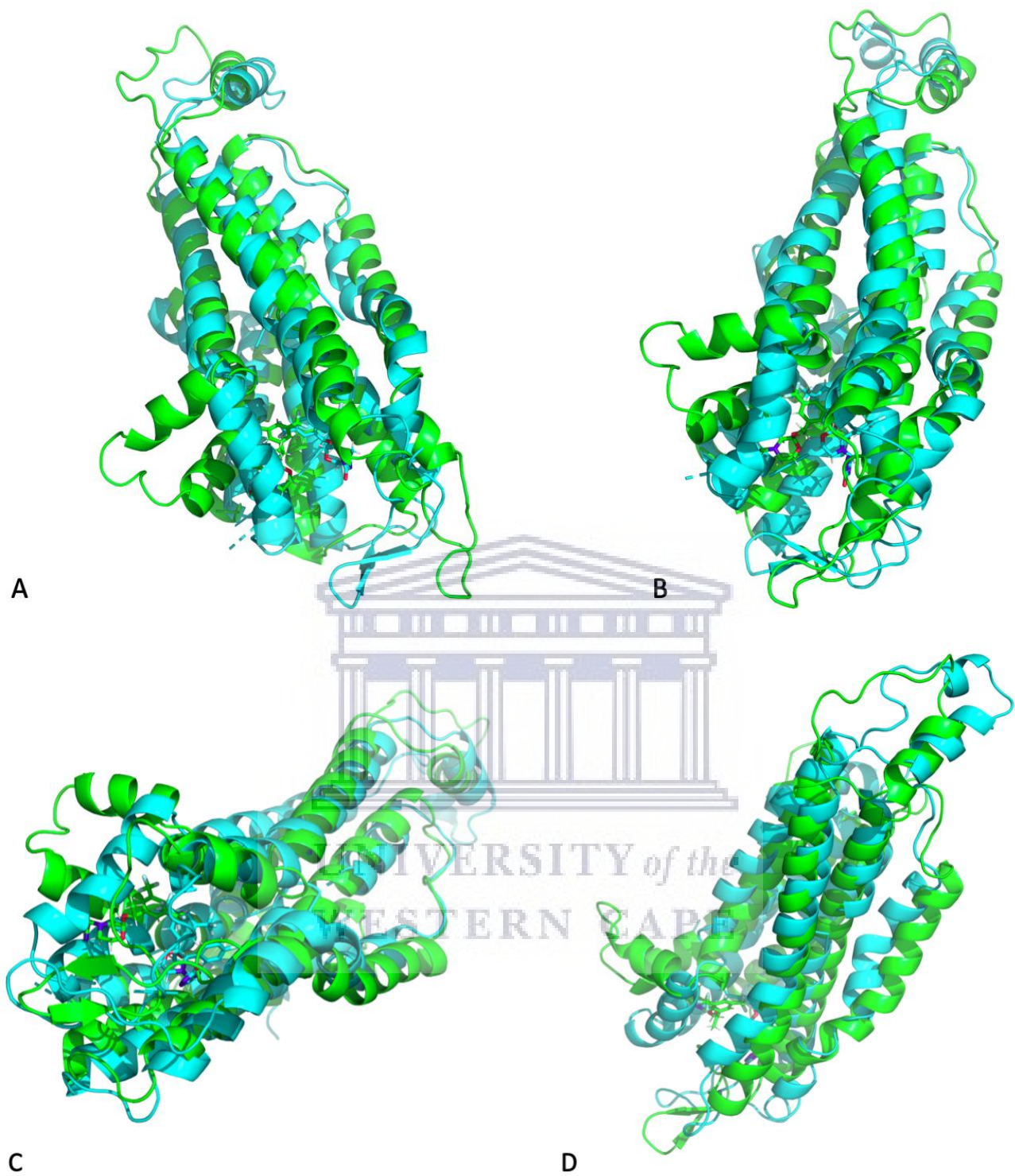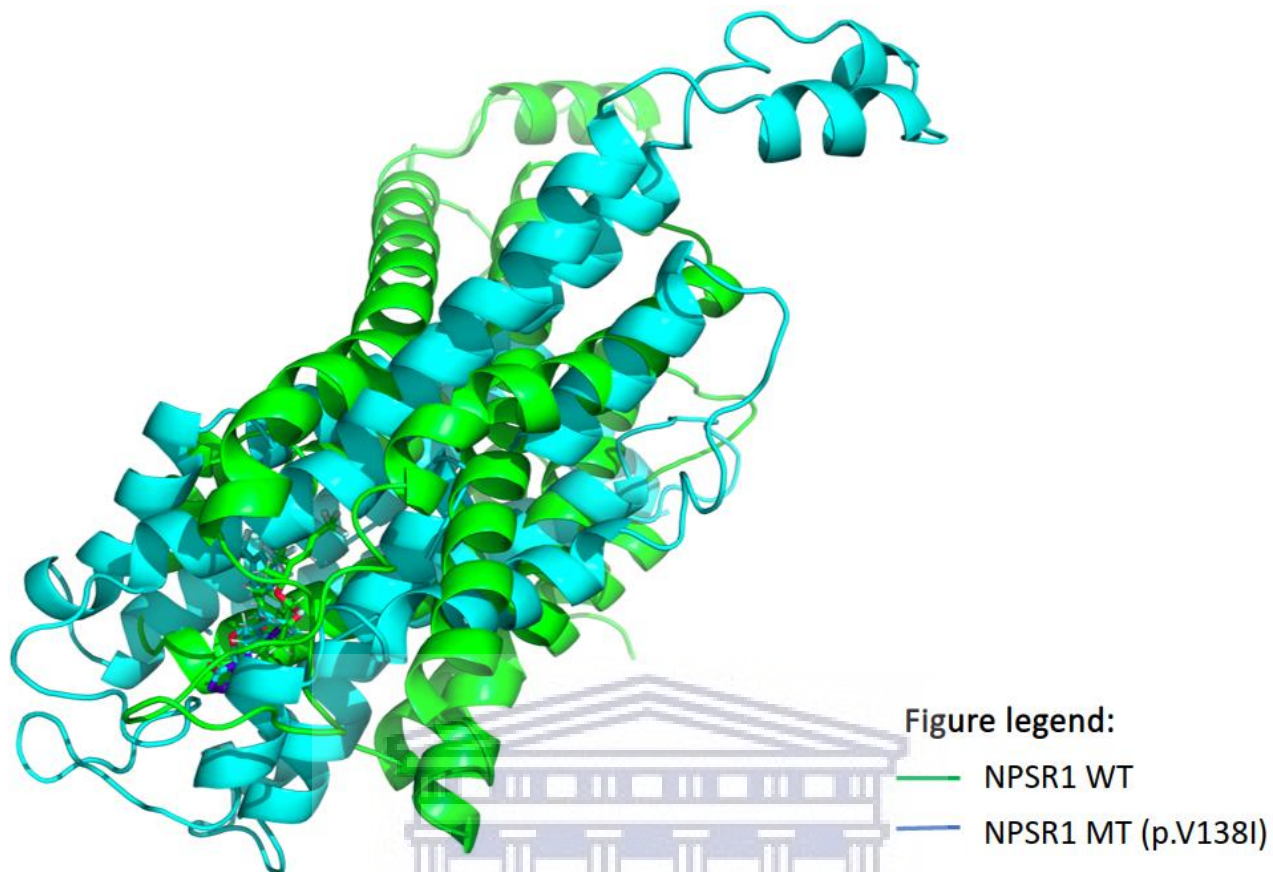
E

Figure S23: Snapshots of the *NPSR1* (WT) and *NPSR1* MT (p.V138I) protein systems during the last 50 ns of the simulation. (A) 160 ns (B) 170 ns (C) 180 ns (D) 190 ns (E) 200 ns. WT shown in green while the MT is shown in turquoise.

Figure S24: Backbone RMSD analysis of the WT *EPB41L2* protein and p.R663C MT systems (repeats) at 303 K over 150 ns of the simulation period.

Approval Letter
Progress Report

20/03/2019

Project ID: 7506

Ethics Reference #: 2002C/059

Title: Genetic analysis of inherited Parkinson's Disease and other related movement disorders

Dear Prof Jonathan Carr,

Your request for extension/annual renewal of ethics approval dated 08/03/2019 11:35 refers.

The Health Research Ethics Committee (HREC) reviewed and approved the annual progress report you submitted at a convened meeting of HREC2 on 20 March 2019.

The approval of this project is extended for a further year.

Approval date: 20 March 2019

Expiry date: 19 March 2020

Kindly be reminded to submit progress reports two (2) months before expiry date.

Where to submit any documentation

Kindly note that the HREC uses an electronic ethics review management system, *Infonetica*, to manage ethics applications and ethics review process. To submit any documentation to HREC, please click on the following link: https://applyethics.sun.ac.za.

Please remember to use your **Project ID [7506]** and Ethics Reference Number **[2002/C059]** on any documents or correspondence with the HREC concerning your research protocol.

Yours sincerely,

Mr. Francis Masiye,

HREC Coordinator,

Health Research Ethics Committee 2 (HREC2).

National Health Research Ethics Council (NHREC) Registration Number:
REC-130408-012 (HREC1)•REC-230208-010 (HREC2)

Federal Wide Assurance Number: 00001372
Office of Human Research Protections (OHRP) Institutional Review Board (IRB) Number:
IRB0005240 (HREC1)•IRB0005239 (HREC2)

The Health Research Ethics Committee (HREC) complies with the SA National Health Act No. 61 of 2003 as it pertains to health research. The HREC abides by the ethical norms and principles for research, established by the *World Medical Association (2013). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*; the South African *Department of Health (2006).Guidelines for Good Practice in the Conduct of Clinical Trials with Human Participants in South Africa (2nd edition)*; as well as the Department of Health (2015). Ethics in Health Research: Principles, Processes and Structures (2nd edition).

The Health Research Ethics Committee reviews research involving human subjects conducted or supported by the Department of Health and Human Services, or other federal departments or agencies that apply the Federal Policy for the Protection of Human Subjects to such research (United States Code of Federal Regulations Title 45 Part 46); and/or clinical investigations regulated by the Food and Drug Administration (FDA) of the Department of Health and Human Services.

Approval Letter
Progress Report

09/06/2020

Project ID: 7506

Ethics Reference No: 2002C/059

Project Title: Genetic analysis of inherited Parkinson's Disease and other related movement disorders

Dear Prof Jonathan Carr

We refer to your request for an extension/annual renewal of ethics approval dated 13/04/2020 15:12.

The Health Research Ethics Committee reviewed and approved the annual progress report through an expedited review process.

The approval of this project is extended for a further year.

Approval date: 14 May 2020

Expiry date: 13 May 2021

Kindly be reminded to submit progress reports two (2) months before expiry date.

Where to submit any documentation

Kindly note that the HREC uses an electronic ethics review management system, *Infonetica*, to manage ethics applications and ethics review process. To submit any documentation to HREC, please click on the following link: https://applyethics.sun.ac.za.

Please remember to use your Project Id 7506 and ethics reference number 2002C/059 on any documents or correspondence with the HREC concerning your research protocol.

Yours sincerely,

Mrs. Ashleen Fortuin
Health Research Ethics Committee 2 (HREC2)

National Health Research Ethics Council (NHREC) Registration Number:
REC-130408-012 (HREC1)•REC-230208-010 (HREC2)

Federal Wide Assurance Number: 00001372
Office of Human Research Protections (OHRP) Institutional Review Board (IRB) Number:
IRB0005240 (HREC1)•IRB0005239 (HREC2)

The Health Research Ethics Committee (HREC) complies with the SA National Health Act No. 61 of 2003 as it pertains to health research. The HREC abides by the ethical norms and principles for research, established by the World Medical Association (2013). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects; the South African Department of Health (2006).Guidelines for Good Practice in the Conduct of Clinical Trials with Human Participants in South Africa (2nd edition); as well as the Department of Health (2015). Ethics in Health Research: Principles, Processes and Structures (2nd edition).

The Health Research Ethics Committee reviews research involving human subjects conducted or supported by the Department of Health and Human Services, or other federal departments or agencies that apply the Federal Policy for the Protection of Human Subjects to such research (United States Code of Federal Regulations Title 45 Part 46); and/or clinical investigations regulated by the Food and Drug Administration (FDA) of the Department of Health and Human Services.