

Computational Study of α -Synuclein Structure and Druggability

by

Mark Anthony Healey

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

Doctor of Philosophy

Department of Physics

University of Alberta

©Mark Anthony Healey, 2016

ABSTRACT

α -Synuclein has been implicated in the progression of Parkinson's disease, a neurodegenerative disorder that affects millions of people worldwide. This work reports on the structural propensity and druggability of this protein using a combination of Monte Carlo, Molecular Dynamics, and Virtual Screening simulation methods. In performing Monte Carlo simulations, we analysed the extended secondary structure formation for Wild-Type and Mutant α -synuclein above and below the apparent phase transition temperature. We also used the low energy phase of our simulations to approximate the potential compact structures that the protein may exhibit in aggregates. From these results, we were able to determine regions of the protein which exhibit variable extended beta sheet structure, suggesting a possible role in the aggregation of α -synuclein, as well as a number of drugs which hold the potential to bind and inhibit this aggregation.

PREFACE

The research conducted in this thesis is the original work of Mark Healey.

Chapter 4.1 of this thesis was published as MA Healey, M Woodside, and JA Tuszynski, “Phase Transitions and Structure Analysis in Wild-Type, A30P, E46K, and A53T Mutants of α -Synuclein”, *European Journal of Biophysics*, Dec 2015. I was responsible for simulations, data collection, analysis, and manuscript composition. M Woodside and JA Tuszynski contributed to manuscript edits, and JA Tuszynski was the supervisory author.

ACKNOWLEDGEMENTS

I would first and foremost like to thank my PhD supervisor, Dr. Jack Tuszynski, for his patient and thoughtful guidance throughout my studies.

I would also like to thank my committee members, Dr. Nils Petersen, Dr. Michael Woodside, and Dr. Mariusz Klobukowski. Their encouragement and advice throughout my degree were invaluable. I would also like to thank Dr. Donald Weaver and Dr. Alex Brown for being a part of my examination committee.

Finally, I would like to thank my family for their life-long encouragement and my fiancée for her enduring patience and support.

CONTENTS

1 MOTIVATION AND OUTLINE	1
1.1 MOTIVATION.....	1
1.2 OUTLINE	2
2 INTRODUCTION.....	4
2.1 PARKINSON'S DISEASE.....	4
2.1.1 <i>History</i>	4
2.1.2 <i>Clinical</i>	5
2.2 α -SYNUCLEIN	12
2.2.1 <i>Intrinsically Disordered Proteins</i>	12
2.2.2 <i>Properties of α-Synuclein</i>	15
2.2.3 <i>Proposed Functions</i>	17
3 METHODS	18
3.1 INTRODUCTION TO COMPUTATIONAL METHODS	18
3.2 PROTEIN MONTE CARLO METHODS	18
3.2.1 <i>Introduction</i>	18
3.2.2 <i>Canonical Protein Monte Carlo</i>	22
3.2.3 <i>Replica Exchange Monte Carlo</i>	25
3.3 MOLECULAR DYNAMICS	26
3.3.1 <i>Introduction</i>	26
3.3.2 <i>Equations of Motion</i>	28
3.3.3 <i>Temperature Control</i>	32
3.3.4 <i>Pressure Control</i>	34
3.4 DRUGGABILITY	37
3.4.1 <i>Introduction</i>	37
3.4.2 <i>Model and Probes</i>	38
3.4.3 <i>Analysis</i>	39
3.5 VIRTUAL SCREENING AND DOCKING.....	39
4 α-SYNUCLEIN MONOMERS AND DIMERS	42
4.1 α -SYNUCLEIN MONOMERS AND MUTANTS	42
4.1.1 <i>Outline</i>	42

4.1.2 <i>Methods</i>	43
4.1.3 <i>Analysis</i>	43
4.1.4 <i>Results</i>	44
4.2 DIMERS	54
4.2.1 <i>Outline</i>	54
4.2.2 <i>Methods</i>	54
4.2.3 <i>Analysis</i>	54
4.2.4 <i>Results</i>	55
4.3 DISCUSSION.....	59
5 MONTE CARLO PULLING SIMULATIONS	63
5.1 OUTLINE	63
5.2 METHODS.....	63
5.3 ANALYSIS.....	65
5.4 RESULTS	66
5.4.1 <i>Monomers</i>	66
5.4.2 <i>Dimers</i>	69
5.5 DISCUSSION.....	72
6 FRAGMENT PROBE ANALYSIS AND DRUG SCREENING	74
6.1 OUTLINE	74
6.2 REPLICA EXCHANGE MONTE CARLO	74
6.3 BINDING SITES	76
6.3.1 <i>Clustering</i>	76
6.3.2 <i>Fragment Probe Analysis</i>	78
6.4 VIRTUAL SCREENING AND DOCKING.....	83
6.4.1 <i>NCI-Diversity Set</i>	84
6.5 DISCUSSION.....	87
7 SUMMARY AND FUTURE WORK.....	89
7.1 SUMMARY	89
7.2 DIRECTIONS FOR FUTURE WORK.....	90

LIST OF TABLES

TABLE 2-1: PROPOSED STAGING OF PD PATHOGENESIS, TAKEN FROM <i>BRAAK ET AL</i> [73].	11
TABLE 5-1: VARIATION OF THE NUMBER OF RUPTURE EVENTS WITH CHANGES IN SPRING CONSTANT (k). WHERE ν IS IN FM/STEP, AND k IS IN pN/NM.	67
TABLE 5-2: COMPARISON OF NUMBER OF TRANSITION EVENTS FOR EACH DIMER PEAK.	69
TABLE 6-1: RESULTS OF DRUGGABILITY ASSESSMENT USING FRAGMENT PROBE ANALYSIS ON CLUSTERS OF ALPHA-SYNUCLEIN.	80
TABLE 6-2: TOP 3 DRUGS BY LOWEST MEAN BINDING ENERGY FROM BLIND DOCKING OF THE NCI-DIVERSITY SET AGAINST AS CLUSTER REPRESENTATIVES.	86

LIST OF FIGURES

FIGURE 2.1: ALPHA-SYNUCLEIN. (PURPLE) NON-POLAR RESIDUES, (BLUE) BASIC RESIDUES, (RED) ACIDIC RESIDUES, (GREEN) POLAR RESIDUES.	15
FIGURE 2.2: GRAPHICAL REPRESENTATION OF THE REGIONAL DIVISION OF AS, FROM THE AMPHIPATHIC N-TERMINUS, NAC, AND ACIDIC C-TERMINUS REGIONS.	16
FIGURE 3.1: SCHEMATIC REPRESENTATION OF THE USE OF MONTE CARLO TYPE SIMULATIONS TO DETERMINE THE VALUE OF Π	19
FIGURE 3.2: SCHEMATIC REPRESENTATION OF POSSIBLE MC MOVES, INCLUDING SIDE-CHAIN ROTATION, AS WELL AS ROTATION ABOUT REGIONS OF PEPTIDE BONDS.	23
FIGURE 3.3: ATOMIC REPRESENTATION OF A TIP3P WATER BOX.	26
FIGURE 3.4: GRAPHICAL DEPICTION OF MD ALGORITHM.	27
FIGURE 3.5: GRAPHICAL DEPICTION OF FORCES CALCULATED IN MD FORCE FIELDS.	30
FIGURE 3.6: SCHEMATIC OF PRESSURE CORRECTION BY VOLUME CORRECTION.	37
FIGURE 4.1: DISTRIBUTION OF CONFORMATIONAL ENERGIES OF WILD-TYPE AND MUTANT A-SYNUCLEIN AS A FUNCTION OF TEMPERATURE. (A) 340 K. (B) 330 K. (C) 320 K. (D) 310 K. (E) 300 K.	45
FIGURE 4.2: ENERGY DISTRIBUTIONS OF TWO SETS OF 16 INDEPENDENT SIMULATIONS RUN AT T=320K, FOR 12 MILLION MC SWEEPS. SHADED REGIONS REPRESENT JACKKNIFE 1σ ERRORS AND EXHIBIT SIGNIFICANT DEGREES OF OVERLAP BETWEEN THE TWO SIMULATIONS, INDICATING CONVERGENCE IN ENERGY.	46
FIGURE 4.3: $1-P_{ij}$ VALUES FOR RESIDUES IN OUR SIMULATIONS CORRESPONDING TO WT I=19 (RED), 90 (BLUE), AND 132 (BLACK), IN OUR HIGH ENERGY ENSEMBLES PRODUCED AT T=320K. THESE MEASUREMENTS HAVE BEEN USED AS AN ESTIMATION OF SPIN LABELLED PRE RESULTS.	47

FIGURE 4.4: PER-RESIDUE PROBABILITY OF DIHEDRAL ANGLES CONSISTENT WITH B-STRAND SECONDARY STRUCTURE AS A FUNCTION OF TEMPERATURE. (A) 340K. (B) 330K. (C) 320K. (D) 310K. BLACK: WILD-TYPE A-SYNUCLEIN, BLUE: E46K, GREEN: A53T, RED: A30P.....	49
FIGURE 4.5: PER-RESIDUE PROBABILITY OF DIHEDRAL ANGLES CONSISTENT WITH A-HELIX SECONDARY STRUCTURE AS A FUNCTION OF TEMPERATURE. (A) 340K. (B) 330K. (C) 320K. (D) 310K. BLACK: WILD-TYPE A-SYNUCLEIN, BLUE: E46K, GREEN: A53T, RED: A30P.....	50
FIGURE 4.6: JOINT PROBABILITY PLOTS USED TO DETERMINE COMMON SIZES OF EXTENDED BETA SHEET STRUCTURE FORMATION AT 330K.	51
FIGURE 4.7: PROBABILITY OF A GIVEN RESIDUE FORMING PART OF A B-STRAND OF LENGTH 5-12 RESIDUES. (A) 340K. (B) 330K. (C) 320K. (D) 310K. BLACK: WILD-TYPE A-SYNUCLEIN, BLUE: E46K, GREEN: A53T, RED: A30P. INSET OF FIGURE A SHOWS THE SEGMENT FROM RESIDUE 50 TO 60.....	52
FIGURE 4.8: PROBABILITY OF A GIVEN RESIDUE FORMING PART OF A HELIX OF LENGTH 11–20 RESIDUES. (A) 340K. (B) 330K. (C) 320K. (D) 310K. BLACK: WILD-TYPE A-SYNUCLEIN, BLUE: E46K, GREEN: A53T, RED: A30P.....	53
FIGURE 4.9: PLOT OF ENERGY DISTRIBUTIONS FOR EACH SIMULATION TEMPERATURE PRODUCE FOR DIMERIC α -SYNUCLEIN.....	55
FIGURE 4.10: PLOT OF DISTRIBUTIONS OF RADIUS OF GYRATION MEASUREMENTS FOR DIMERIC α -SYNUCLEIN	56
FIGURE 4.11: FREE ENERGY CONTOUR PLOT FOR COVALENTLY LINKED AS DIMERS. FREE ENERGY IS CALCULATED AS $F = -\text{LN}(R_g, \text{SASA})$	57
FIGURE 4.12: RESIDUE-RESIDUE CORRELATION FOR COVALENTLY BOUND DIMERS OF AS FOR EACH OF 4 TEMPERATURES A) 300K B) 310K C) 320K D) 330K.....	58
FIGURE 4.13: PROPORTION OF COVALENTLY LINKED DIMER EXHIBITING SPECIFIC SECONDARY STRUCTURE. (RED) BETA SHEET (BLUE) HELIX (GREEN) NO STRUCTURE.....	59
FIGURE 5.1: PLOT OF EXT V. FORCE FOR CYS-TERMINATED AS MONOMERS. PLOT IN RED REPRESENTS RAW DATA, WHEREAS BLUE IS THE RESULT OF SUBSEQUENT SMOOTHING OF THE RAW NOISY DATA.	66

FIGURE 5.2: PLOT OF HISTOGRAMS OF DISTRIBUTION OF DELTA LC VALUES FOR AS MONOMERS, AS WELL AS DISTRIBUTION OF FORCE RELATIVE TO LC CHANGE.	68
FIGURE 5.3: DISTRIBUTION OF <i>LC</i> , AND FORCES FOR THE LOWEST ENERGY PEAK (PEAK 1) IN THE DIMERIC ENSEMBLE.....	70
FIGURE 5.4: DISTRIBUTION OF <i>LC</i> AND FORCES FOR PEAK 2 OF THE AS DIMER ENSEMBLES.....	71
FIGURE 5.5: DISTRIBUTION OF <i>LC</i> AND FORCES FOR THE PEAK 3 ENSEMBLE OF DIMERIC AS.....	72
FIGURE 6.1: DISTRIBUTION OF ENERGYIES FOR WT REMC SIMULATIONS OF AS.....	75
FIGURE 6.2: PLOT OF CONTACT MAP FOR RESIDUES WITHIN 6 NM OF ONE ANOTHER FOR CLUSTERS 1-9 IN LIGHT BLUE. THE CONTACT MAP PRODUCED BY <i>VILAR ET AL.</i> IS SHOWN IN YELLOW, WITH DIRECT OVERLAP SHOWN IN RED.	77
FIGURE 6.3: REPRESENTATIVE STRUCTURES FOR CLUSTER 1. A. SECONDARY STRUCTURE REPRESENTATION. YELLOW ARROWS REPRESENT BETA-SHEET STRUCTURES. B. ELECTRONIC SURFACE AREA REPRESENTATION.	78
FIGURE 6.4: BINDING SITES FOR AS CLUSTER 1 REPRESENTATIVE. A. BINDING SITE 1 B. BINDING SITES 2 AND 3.....	81
FIGURE 6.5: FREQUENCY OF RESIDUES IN BINDING SITES BY DRUGGABILITY ANALYSIS.....	82
FIGURE 6.6: DISTRIBUTION OF THE MEAN MAXIMUM BINDING ENERGY'S OF DRUGS IN THE NCI-DIVERSITY SET.....	85

LIST OF ABBREVIATIONS AND ACRONYMS

PD – Parkinson’s Disease

AD – Alzheimer’s Disease

AS – α -Synuclein

LBD – Lewy Body Dementia

MC – Monte Carlo

REMC – Replica Exchange Monte Carlo

ANS – Autonomic Nervous System

OH – Orthostatic Hypotension

UD – Urogenital Dysfunction

UPDRS – Unified Parkinson’s Disease Rating Scale

MDS – Movement Disorder Society

IDP – Intrinsically disordered proteins

IDPR – Intrinsically disordered protein regions

LN – Lewy Neurites

LB – Lewy Body

NAC – Non Amyloid Containing

CBMC – Configurational-Bias Monte Carlo

RS – Rosenbluth Sampling

BGS – Biased Gaussian Steps

MD – Molecular Dynamics

NMR – Nuclear Magnetic Resonance

FF – Force Field

LJ – Lennard-Jones

IPAM – isopropylamine
LGA – Lamarckian Genetic Algorithm
WT – Wild Type
VMD – Visual Molecular Dynamics
PRE – Paramagnetic relaxation experiment
RMSD – root mean squared deviation
 R_g – Radius of gyration
 R_h – Radius of hydration
ASA – Accessible Surface Area
BS – Beta Strand
AH – Alpha Helix
AFM – Atomic Force Microscopy
EWLC – Extensible Worm-Like Chain
 L_c – Contour Length
 L_p – Persistence Length
DBSCAN – Density Based Clustering Algorithm
NCI – National Cancer Institute

LIST OF APPENDICES

APPENDIX 1 UNIFIED PARKINSON'S DISEASE RATING SCALE	110
APPENDIX 2: FULL TABLE OF BINDING REGIONS FOR DRUGGABILITY ANALYSIS	112
APPENDIX 3: TOP BINDING ENERGIES FROM BLIND DOCKING	121

1 MOTIVATION AND OUTLINE

1.1 Motivation

Parkinson's disease (PD) is 2nd only to Alzheimer's Disease (AD) in worldwide prevalence of neurodegenerative diseases, and has been found to affect approximately 1-2% of the population over 60 [1]. The disease exists in several forms, with the 'sporadic' form being the most prevalent, which has been linked to such potentially causative (or at least correlative) factors as head trauma, and presence of environmental toxins [2]. The remaining population of those diagnosed with the disease include those with familial, or genetic, predisposition to the disorder which involves the presence of specific mutations involved in the progression of the disease [2].

As the main portion of this thesis, I have used several computational methodologies to investigate the structural properties of α -Synuclein (AS), in collaboration with Dr. Michael Woodside's and Dr. Nils Petersen's research groups, in order to further understand this elusive protein. AS has been implicated as the toxic species in PD, and Lewy

Body Dementia (LBD), and is found as fibrillar aggregates (Lewy Bodies) in the post-mortem brains of those afflicted with PD.

AS appears to have several functions, including transport of vesicles [3], dopamine regulation (release/transport/metabolism) [4, 5], lipid metabolism [6]. It has been found to exist in a natively unstructured form in solution, and is classed with over 400 other proteins as an Intrinsically Disordered Protein (IDP). Oligomerisation and aggregation of toxic AS appear to form over several hours/days (depending on experimental setup), and any method that may interfere with the interaction, or the ability of the protein to form necessary secondary structures, may reduce or eliminate the progression of the disease in patients.

Little work has been done in understanding how extended structure formation may occur within such a (nearly) random coil protein, as well as production of models of fibrillated monomers. The purpose of this thesis is to elucidate the underlying structures that are formed by AS and how they might aid in the formation of oligomers and aggregate structures, as well as determining potential drug-binding regions and strategies in drug discovery for AS.

1.2 Outline

Chapter 2 will introduce the topics of Parkinson's Disease and α -Synuclein, focusing on current knowledge concerning both. Chapter 3 will then focus on relevant computational methods used in the process of studying AS in this work. This includes the underlying principles of MC, including the Metropolis-Hastings Criteria, force field used in this work, and the concepts of replica exchange. We will also focus on the principles that are key to molecular dynamics simulations, as well as fragment/druggability analysis.

Chapter 4 will focus on the results, and quickly overview the analytical methods, used in the study of monomeric AS and its mutants in

order to further understand the potential structural reasons for variation in aggregation of these monomers.

Chapter 5 focuses on the simulation of AS dimers produced through covalent linkage of AS monomers by canonical MC, REMC, and MC pulling simulations to determine the presence of potential interacting structures between the two monomers

In chapter 6, we have applied methods of Replica Exchange Monte Carlo to produce models of the fibril core region (approximately residue 35 to 95) [7, 8]. These results were then clustered using Density Based Clustering to produce a set of representative structures, which were then compared to experimental data to determine the possible validity of the models. The representative structures were analysed using fragment based druggability simulations to determine possible drug-binding regions. As part of this search for possible regions of interest, we also simulated full length AS under restraint to maintain a portion of the N-terminus in a helical state using Replica Exchange Molecular Dynamics, to determine the potential effects of a drug binding to stabilize this conformation on the formation of secondary structure in the fibril core and the potential implications on drug discovery.

Conclusions and summarizations of this work are presented in chapter 7.

2 INTRODUCTION

2.1 Parkinson's Disease

2.1.1 History

In the 19th century, James Parkinson's provided the first description of paralysis agitans, later to be termed Parkinsonism and then Parkinson's Disease (PD) [9]. In his essay, he discusses both the relevant clinical features that he found as well as 6 case studies. He describes the primary symptoms as being tremor, gait disturbance, and weakness, especially in the feet and hands. He also refers to the presence of disturbance in the patient's sleep, as well as constipation and speech abnormalities.

After the presentation of the disorder by Parkinson, several other researchers presented further studies and discussions of diagnosis and management over the remaining 19th century. The next major presentation of the disease was presented by Jean-Martin Charcot. One of Charcot's most important contributions to the discussion of tremor disorders was the separation of tremor types and disorders [10]. This involved the differentiation of rest tremor from action tremor, and the symptoms that accompany both. Upon establishment of the major symptomology of PD separate from other similar disorders, Charcot was able to classify both typical and atypical forms of PD, termed Parkinson Plus.

The 20th century presented itself with further developments in the understanding and clinical features of PD. Pathologic anatomical findings in PD were described first by Frederic Lewy in 1912, noting the formation of eosinophilic (readily stained by eosin) inclusions in PD affected patients. The effect of dopamine and L-dopa on PD patients was first described in the 1950's by Carlsson and Hornykiwicz [11].

In the late 20th century, several genetic markers which increase the risk of PD were found. These include the discovery of the first mutation associated with PD, that being of PARK1 which codes for AS, as well as the identification of AS as major portion of Lewy Body's [12, 13]. Another seminal study discovered the role of misfolded AS as transmission vectors capable of producing Parkinsons-Like symptoms [14].

2.1.2 Clinical

2.1.2.1 Symptoms

PD typically presents initially with some grouping of what are considered its cardinal symptoms of motor dysfunction. These symptoms include resting tremor, bradykinesia/hypokinesia (slowness of movement), rigidity and postural instability, also known as postural reflex impairment [15, 16]. Tremor is the most common complaint amongst sufferers of Parkinson's disease, with over 70% of those diagnosed exhibiting it as a primary symptom [17]. This tremor is most concentrated at the region of the limb furthest from midline, and is typically inversely proportionate to the degree of rest experienced by the limb. A common expression of this symptom is the "pill-rolling" motion, in which the patients thumb and forefinger produce an involuntary back and forth motion [15, 17]. Periods of heightened stress, such as anxiety, depression or fatigue, can increase the degree of tremor, whereas the tremor virtually disappears during sleep and under voluntary movement [18]. The head, neck, and trunk are rarely affected except when the degree of motion in the limbs is of such a degree as to disturb the resting position of these regions [19].

Hypokinesia/Bradykinesia typically presents itself amongst fine motor tasks such as writing, and is described as potentially the most disabling symptom [20]. This slowness of voluntary movement may also present itself as akinesia, or the near absence of voluntary movement [21, 22]. Bradykinesia can also manifest itself as loss of spontaneous motion, including drooling, facial expression and blinking [23]. Hypophonia, or soft speech, has also been described as a likely presentation of bradykinesia that affects those muscles involved in speech such as the vocal cord and respiratory muscles [24]. This can also be exhibited as a development of an apparent stutter or dysarthria, which is the inability of a person to properly articulate speech [25]. Assessment of bradykinesia is accomplished in slightly different manners depending on what areas are being measured. In the case of limb bradykinesia, the patient may be asked to repeatedly tap the thumb and index finger together, alternately rotate the forearm (pronate/supinate), open and close the hand or tap the foot rhythmically [26]. The assessment of fine motor skills may simply be assessed by examining variation in the patients writing ability over time [27]. Like resting tremor, bradykinesia is typically asymmetric and affected significantly by the emotional or physical state of the patient.

Muscular rigidity is another cardinal symptom of PD, and is due to the increase in muscle tone, and/or abnormal long-latency reflexes experienced by those with Parkinson's disease [28-32]. This increase in tone is due to increased activation of muscular contractile elements [33]. Rigidity is assessed throughout the passive range of motion of the potentially affected limb, in flexion, extension or rotation about that joint [26]. This rigidity is frequently accompanied by pain and mistaken as simple bursitis or arthritis [34]. Rigidity is believed to be due to multiple factors

The fourth most common motor sign of PD is postural instability, also referred to as postural reflex impairment, which can also affect the gait of the patient [35-39]. This typically occurs in 3.5-18% of diagnosed patients, and typically occurs 5 years after initial diagnosis, with a larger portion of late-onset PD patients exhibiting this symptom than those with

early-onset PD [36, 40]. As the disease progresses, initiation of gait may become more difficult as part of the gait disturbance or bradykinesia [36].

Other non-motor symptoms may also be apparent in patients with PD, including sensory symptoms, autonomic dysfunction, sleep disorders and disruptions, as well as neuropsychiatric disorders [41, 42].

Sensory symptoms of PD may include pain symptoms that may accompany neuro-motor issues such as muscular or joint/rheumatologic pain which may be related to stiffness or muscular dystonia, or back pain related to postural instability or gait disturbances [43]. Other sensory symptoms may include neurologic pain classified as paraesthesia or paraesthesia-like, which produce a numb or tingling sensation in affected limbs, and can also cause a feeling of coldness or even a burning pain [43]. Sensory symptoms have often been found to precede motor symptoms in PD patients [43, 44]. These symptoms can also be found in a majority of PD diagnosed patients responsive to levodopa (67%), which is significantly higher than the age-matched general population [43]. This may have significant impacts in the quality of life experienced by PD patients. Studies have reported variable responsiveness of these sensory symptoms to L-dopa therapy, depending on both type and location of pain [45, 46].

Autonomic dysfunction (or dysautonomia) is another near universal class of symptoms affecting PD patients, and are often potential symptoms prior to onset of motor dysfunction [47, 48]. The autonomic nervous system (ANS) contains two subsystems; the parasympathetic and sympathetic nervous systems [49]. Most organs exhibit innervation by both of these subsystems. The ANS is responsible for the regulation of many bodily functions including body temperature, balance, blood pressure etc [50]. Dysautonomia is exhibited in many variable ways as malfunctions of various bodily functions. In PD patients, this is typically revealed as orthostatic hypotension (OH), urogenital dysfunction (UD), and constipation [51, 52]. OH is a form of low blood pressure that occurs to changes in orientation of the body, for example when quickly standing from a sitting posture [53]. Chart reviews have shown symptomatic OH in

30% of cases, with tilt-table tests of PD patients showing 58% exhibiting significant (>20mmHG) drops in systolic blood pressure [54]. Other studies have shown clinically significant impacts of OH in 50% of patients with PD [55]. UD may include several issues including erectile dysfunction or ejaculatory failure, urinary incontinence/frequency/urgency, incomplete bladder emptying, double micturition, and urging continence. UD is typically a later stage symptom of PD.

Sleep disorders are another nonmotor symptom commonly reported by patients with PD [56-58]. Typical sleep related complaints of PD patients include restless-leg syndrome, nighttime incontinence, hallucinosis, and difficulty falling asleep [59, 60]. Due to degeneration of neuronal structures in the PD patient, they may also suffer from issues related to sleep fragmentation, disruption of REM sleep, and reduced slow-wave sleep [59, 60]. Sleep fragmentation refers to frequent sleep disruptions which may be due to multiple reasons, with 76% of PD patients in the UK reporting [56, 57]. This fragmentation may be due to autonomic disturbance including frequent nocturnal urination or incontinence, or motor issues such as tremor or muscle cramps. Other potential causes include reduction in the effect of medications taken during the day to reduce symptoms, resulting in increased PD motor and sensory symptoms.

Other typical nonmotor symptoms related to PD include those disorders classified as neuropsychiatric dysfunction [61]. This may include more common disorders such as depression and anxiety, and to those less common disorders such as dementia, or disorders affecting higher level executive function such as impulse control disorders [62, 63].

2.1.2.2 Assessment

Clinical assessment of Parkinsonism is typically performed through a checklist of symptoms and their severity in order to assess the patients

degree of motor symptoms and/or disability [26]. Several scales have been proposed or placed in use by clinicians including;

- Webster, Columbia University Rating Scale, and Parkinson's disease Impairment Scale, which assesses primarily motor symptoms [64].
- Schwab and England and Northwestern University Disability Scale, which assesses patient disability [64].
- Unified Parkinson Disease Rating Scale (UPDRS), which provides assessment of both [26].
- MDS-UPDRS, which is a rating scale developed by the Movement Disorder Society (MDS-UPDRS) to replace the UPDRS due to perceived weakness in the UPDRS attention to non-motor symptoms [65, 66]

In this section, I will focus primarily on the two forms of the UPDRS, as they are generally perceived to be the most complete, as well as the most commonly used scales in the assessment of Parkinsonism.

The UPDRS scale is the most commonly used by clinicians to assess PD patients due to its general utility as well as its completeness [64]. Prior to the advent of the UPDRS, multiple scales were used across clinics to assess PD, resulting in difficulty in comparisons [65]. The UPDRS attempts to incorporate many aspects of these scales into a cohesive framework. This assessment is divided into 4 parts, each consisting of questions related to symptoms developed by PD patients. Each focused on different aspects of the disorder, and two of which include other scales as subsections to the UPDRS [64]. The UPDRS is divided into 4 parts, which measure different aspects of the disease and its progression. These include;

1. Part I: Mentation, Behaviour and Mood.
2. Part II: Activities of Daily Living.
3. Part III: Motor Disturbances.
4. Part IV: Complications

Each part is divided into sets of questions related to the section of the assessment, with each question response being given on a scale of 0-4, where 0 is no symptom and 4 is severe, the full checklist is located in Appendix I [26]. The Movement Disorder Society's (MDS) Task Force was set to analyse the UPDRS for potential deficiencies and recommendations on how to improve upon it. The major criticisms of the original UPDRS presented by the Movement Disorder Society include [65];

1. Limited usefulness in early stages of disease.
2. Inadequate measure for depression, dementia, or psychosis.
3. Some motor portions have poor inter-rater reliability.
4. Part II exhibits potential cultural bias.
5. Lack of separation between patient perceived effects of PD on daily living, and those symptoms directly associated with PD.
6. Inability to distinguish symptoms from potential comorbidities.

These issues were subsequently resolved in the MDS-UPDRS. This form of the UPDRS maintains many of the salient features of the original, including a total score assessment, as well as similar division of questions into 4 parts that are conceptually similar to the original UPDRS [65]. A method of score conversion between the UPDRS and the MDS-UPDRS has also been proposed [67]. The result of this study showed the ability to convert Parts II and III. Part I and IV were not able to be converted due to inherent differences [67]

2.1.2.3 Progression

PD is typically diagnosed in patients in the range of 50-65 years of age, although in rare circumstances, early onset PD is found in patients younger than 35 [68]. The progression of PD can be subdivided into typical clinical (upon evaluation by a medical professional) and pre-clinical (prior to evaluation). This can also be described as whether or not clinically significant neurophysiological deficits or dysfunctions have presented in the patient. Many studies have noted that upon the onset of

the traditional symptomology of PD, significant pathological changes of the patient’s brain have already occurred. This include significant loss of dopaminergic neurons in the striatum and substantia nigra [69-72]. More subtle changes that may indicate PD pathology may go unnoticed until later periods, and are only readily apparent in hindsight, or retrospective analysis.

In 2003, *Braak et al* proposed a staging paradigm for PD [73].

Stage	Pathological Changes
Stage 1; Medulla Oblongata	Lesions in the dorsal IX/X motor nucleus and/or intermediate reticular zone.
Stage 2: Medulla oblongata and pontine tegmentum	Pathology of stage 1 plus lesions in the caudal raphe nuclei, gigantocellular reticular nucleus, and coeruleus-subcoeruleus complex.
Stage 3; Midbrain	Pathology of stage 2 plus midbrain lesions, in particular in the pars compacta of the substantia nigra.
Stage 4; Basal prosencephalon and mesocortex	Pathology of stage 3 plus prosencephalic lesions. Cortical involvement is confined to the temporal mesocortex (transentorhinal region) and allocortex (CA2-plexus). The neocortex is unaffected.
Stage 5; neocortex	Pathology of stage 4 plus lesions in the high order sensory association areas of the neocortex and prefrontal neocortex.
Stage 6; neocortex	Pathology of stage 5 plus lesions in first order sensory association areas of the neocortex and premotor areas, occasionally mild changes in primary sensory areas and the primary motor field.

Table 2-1: Proposed staging of PD pathogenesis, taken from *Braak et al* [73].

Their proposed stages are shown Table 2-1. Staging of PD depends on the underlying belief that pathological changes in the PD brain

are not random, and different sections of the brain are more sensitive to the development of pathological changes. Braak's staging has garnered wider support since it was proposed. Braak and colleagues have shown particular susceptibility of long, thin projection neurons which show little to no myelination. In contrast, those projection neurons with heavily myelinated axons appear to be resistant [74, 75].

2.2 α -Synuclein

2.2.1 Intrinsically Disordered Proteins

Intrinsically disordered proteins (IDPs) represent a unique class of protein that lack an organised 3D structure [76-80]. There also exists a subset of proteins which exhibit intrinsically disordered protein regions (IDPRs or IDRs) [80]. IDPs and IDRs are quite common in nature, and a database (DisProt) of known proteins and regions has been set up to catalogue disordered proteins, and currently contains over 600 proteins and 1500 regions [81].

These discoveries have changed the way in which we think about how a protein functions and the way in which they may express this function. For much of the 20th century and late 19th century, the prevailing model concerning the structure/activity relationship of proteins was described using the "Lock and Key" model of enzyme-substrate interactions described by Fisher. In this model, a substrate (whether another protein, chemical, etc) acts as a "key" capable of initiating an enzyme to action by fitting to the enzymes active site, or "lock". This requires each enzyme to be static, and only able to bind a single substrate, hence the other name for the model, the one protein-one substrate model [82]. Even to this day, this idea is frequently used successfully to discover potential binding sites on proteins which exhibit more static structures. Such methods may include volumetric or geometric methods which attempt to determine binding sites through determining where a binding pocket (or "lock") may exist on the protein [83]. Further work in to understanding the behaviour and binding of proteins and substrates lead to

the work of Koshland in developing his “induced fit” model [84]. This model highlights the potential importance of flexibility in binding substrates by suggesting that the interaction of the protein with the substrate alters the proteins tertiary structure in order to better accommodate the substrate.

Although IDP/IDPRs have been discovered independently over the past century, it has only recently become a significant area of research, with the number of publications regarding these proteins increasing exponentially over the past 25 years [76, 77, 79]. This explosion in interest can be attributed to several factors, including the advent of genetic sequence analysis and bioinformatics methods [76, 85]. For example, traditionally, a protein mixture would be assayed according to activity, with the resultant homogenate being fractionated, assayed, and analysed for structural properties. This may result in the degradation of any IDPs during the isolation stage, as they are more prone to destruction by proteases released during the homogenization process [77]. Modern methods of analysing activity relationships using genetic methods may involve the mapping of a particular function to a region of an organism’s genetic sequence. Once this mapping has been complete, the protein described by this region is transcribed and further analysed for structure and activity, thus resulting in less protein loss and more robust descriptions of their structure and function [77, 85].

IDP/IDPRs exhibit several unique structural properties. The amino acid distribution in the proteins and regions differs quite remarkably from ordered regions [86-89]. They are found to contain few amino acids containing bulky hydrophobic groups such as Valine, Leucine, Isoleucine, Methionine, Phenylalanine, Tryptophan, and Tyrosine. This would typically produce the portion of a protein that is internal to the protein-water interface due to entropic/London dispersion forces [90-92]. In the absence of these bulky hydrophobic groups, IDP/IDPRs exhibit a higher propensity for amino acids with highly charged side chains or structure breaking side chains. The relation of hydrophobicity and net charge is best expressed in an Uversky plot (where mean scaled hydropathy is plotted

against mean net charge), where a separation of folded and unfolded/disordered proteins is evident.

IDP/IDPRs exhibit a broad range of expressed functions. Even a single protein can exhibit several functions and abilities, owing to its degree of flexibility. Examples of functions exhibited by IDPs/IDPRs include entropic springs, cellular signalling and regulation, chaperones, protein assemblers, and scavengers [77, 87, 93-103].

Entropic chains provide their function directly due to their flexibility, such as linking protein regions [104]. These linkers are essential in maintaining interactions of some protein domains, or protein-protein interaction [105, 106]. An example of flexible protein linking domain is the interaction of the SH2 and SH3 domains with the catalytic domains of Src kinases [105]. These chains may also act as spacers, separating two proteins or protein regions. An example being the region projecting from the microtubule associated protein 2, which acts as a dendritic spacer affected by phosphorylation [107]

The role of IDP/IDPRs in cellular signalling and regulation has been extensively reviewed in two recent publications by Wright and Dyson and Pompa [78, 108]. IDP/IDPRs function in this respect by a number of methods, including the presence of multiple interaction motifs, in which the interaction of the target and substrate cause the process of coupled folding to occur [109-111]. Other IDP/IDPRs maintain their disordered structure upon binding, or maintain their previous level of structural dynamicism, while still exhibiting a static interface with its target [112-116]. This flexibility also allows IDP/IDPRs to potentially interact with a number of targets, which allows them to act as protein interaction hubs [115, 117-119]. IDP/IPDRs are also prone to post-translational modifications, such as phosphorylation, which alters the proteins conformation, conformational flexibility, and activity [120]. The flexibility, and ability to bind to multiple targets also allows IDP/IDPRs to act as chaperones, protein assemblers, and scavengers [121].

The disorder and flexibility that provide IDP/IDPRs with their ability to operate in multiple ways, also makes them inherently difficult to study. This flexibility gives the protein a rough energy landscape, instead of the more funnel shaped landscape of structured proteins [122]. This makes experimental evaluation of the structural landscape ensembles to be multiply degenerate, and finding an average structure virtually impossible. Simulations of proteins are traditionally performed based on the evaluation of a potential energy function and the resulting trajectories produced. These functions are typically parameterized empirically or from *ab Initio* calculations, and oftentimes use folded proteins for their parametrization [123-125]. This has resulted in the ensembles produced in simulations with different force fields having very different structural characteristics [126]. This issue can be addressed through the addition of experimental restraints to the system in question, the appropriate choice of force fields, and/or appropriately choosing the best method for your particular goal.

2.2.2 Properties of α -Synuclein

α -Synuclein (AS) is a 140 residue IDP that, as previously mentioned in Section 2.1, is implicated in the pathogenesis of PD. An example of AS structure is shown in Figure 2.1. In those patients with PD, it is found as a major component of the structure of eosinophilic inclusions called Lewy Bodies, as well as Lewy Neurites (LN) [127].

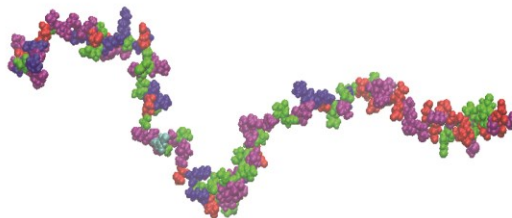


Figure 2.1: Alpha-synuclein. (Purple) Non-polar residues, (Blue) Basic residues, (Red) Acidic residues, (Green) polar residues.

AS is typically described as consisting of three separate regions. These include an amphipathic N-terminal region, from residue 1 to 60, which contains an 11 residue imperfect repeat consisting of the consensus

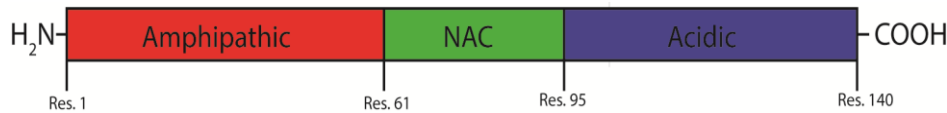


Figure 2.2: Graphical representation of the regional division of AS, from the Amphipathic N-Terminus, NAC, and Acidic C-Terminus regions.

sequence KTKEGV [128]. This region of the protein has been shown to be necessary for the binding of AS to lipid membranes and vesicles. It also contains a hydrophobic core, or the Non Amyloid Containing (NAC) region, from residue 61-95, and an acidic C-terminal region [129, 130].

Typically, AS can be found within the cell in a primarily disordered state, or containing significant degrees of helical structure, such as when it is bound to membranes [131, 132]. In PD, AS can be found as β -sheet rich, misfolded structures, found in small proto-fibrillar aggregates, or as longer fibrillar aggregates. The rate of production of these aggregates *in vitro* has been found to be increased by a number of agents. These include the transition metals Cu, Zn, Fe, and Mn, as well as Cu and Zn-superoxide dismutase [133].

AS has been implicated in the pathogenesis of PD in a number of ways. Several point mutations of AS, as well as duplication and triplication of gene locus repeats, have been noted to increase a person's susceptibility to developing PD [12, 134-137]. Also, animal models in which AS is overexpressed resulted in formation of LN and LB as well as neurodegeneration [138]. One of the more notable discoveries in recent years has been the observation of cell-to-cell transmission of AS in animal models as a mechanism of pathology [14]. LB's and LN's are composed of several structures, most notably that of AS fibrils. [13]. AS monomers within these fibrils are believed to form a stacked, apparently ordered arrangement consisting of a high proportion of beta sheets within the core

region [8]. Several studies have shown the rate of aggregation of AS to be affected by a number of sources, including the presence of transition metals (Zn, Cu, Fe, Mn) and Cu- and Zn-superoxide dismutase [133]. The NAC region of the core of AS has also been shown to be necessary for the aggregation of the protein, with the segment from residue 68 to 78 being of notable importance [130, 139-143]. Recently, experimental evidence has shown that these monomers actually form a Greek Key type topology in its core [144]. However, this may be due to the significant degree of flexibility of AS, which could result in its ability to form multiple types of internal arrangements.

2.2.3 Proposed Functions

The function of AS is not well defined, however there is evidence of its many potential roles. AS is primarily located near, or associated with, the neurons presynaptic terminal [145-147]. This is suggestive of its role in neurotransmitter release, as well as synaptic function and plasticity. Experiments have also been performed on mice in which *Snca* (the gene responsible for production of AS) has been knocked out. These mice have hippocampal synapses showing a reduced response to high frequency stimulation, as well as reduced replenishment of neurotransmitter pools [148, 149]. Experiments on mice which exhibited overexpression of *Snca* have also been performed. These mice exhibit multiple impairments appearing consistent with pre-clinical signs of PD, including gastrointestinal, motor activity, and olfactory abnormalities [150-152]. Overexpression of AS has also shown to be responsible for inhibition of vesicle exocytosis in cultured neurons, and neurotransmitter release [153, 154].

3 METHODS

3.1 Introduction to Computational Methods

In this chapter, we will be discussing various methods used in the study of the physical and chemical properties determined in this work. These include forms of protein Monte Carlo simulations, molecular dynamics calculations, as well as druggability analysis and virtual screening/docking. Each of these methods has its own specific benefits, applicability, and underlying algorithms which will be discussed.

3.2 Protein Monte Carlo Methods

3.2.1 Introduction

The Monte Carlo (MC) computational method takes its name from the city in Monaco known for its gambling, and games of chance.

Random sampling, the simplest form of MC method, is frequently used in the calculation of integrals instead of the use of numerical quadrature methods. This can provide a significant speed increase in calculation, as well as calculating integrals which may not lend themselves well to linearization. For a function $f(x)$ over the interval $[a,b]$, we can calculate the integral by randomly choosing a number in the defined

interval, evaluating the function for the chosen number, performing this again for some predefined number of points, and summing the results. This can be better explained through a common example of calculating π . For this example, we might inscribe the upper right quadrant of a circle of radius $R=1$, inside a square with a side of length 1. We know that the area of the square is 1, and the area of this quadrant of the circle to be $\pi/4$. The ratio of the two areas is thus $\pi/4$. If we randomly choose points between 0 and 1, the ratio of the number of points that will fall within the circle quadrant M to the number of points selected N is proportionate to the ratio of the areas, as expressed in Figure 3.1. This then gives us an expression for π ,

$$\pi = 4 \frac{M}{N}, \quad 3.1$$

where the accuracy increases with the number of points selected.

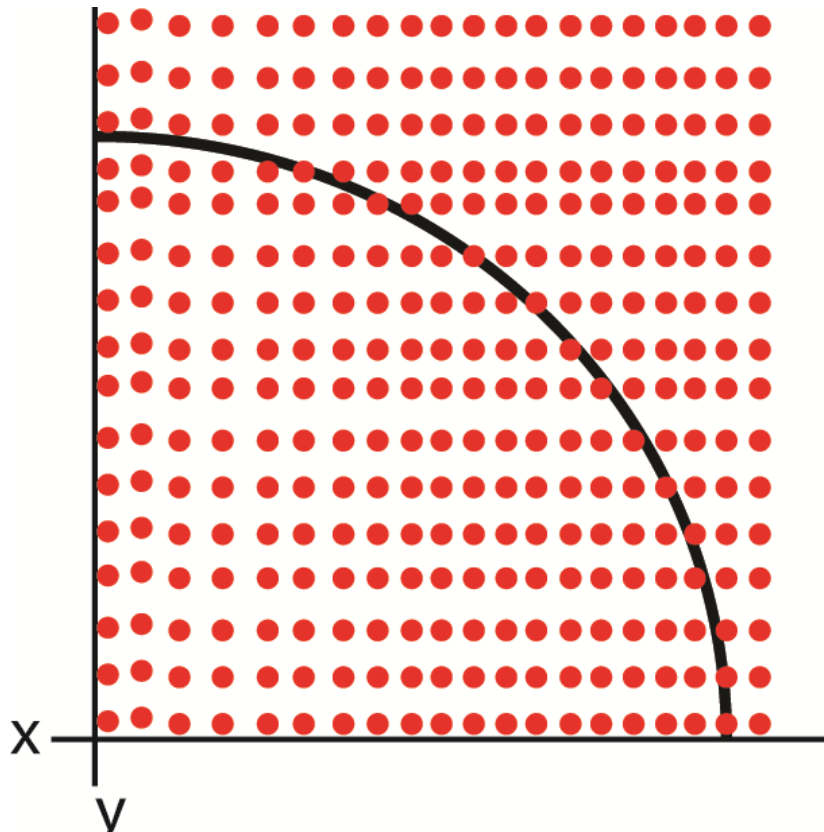


Figure 3.1: Schematic representation of the use of Monte Carlo type simulations to determine the value of π .

The essence of MC simulations is to effectively sample the space of possible results according to some predetermined distribution. If, for example, we define our partition function as,

$$Z \equiv \int dr^N \exp[-\beta U(r^N)] \quad 3.2$$

where β is $1/k_{\beta}T$, r^N represents the N-dimensional coordinate systems describing the particles in the system, and $U(r^N)$ is the potential energy of the particles[155]. This can then be used in the calculation of the mean properties of the system expressed by,

$$\langle A \rangle = \frac{\int dr^N \exp[-\beta U(r^N)] A(r^N)}{Z} \quad 3.3$$

where A is the property in question. We can also note that the probability density of a system existing in some state r^N can be represented using our partition function as well,

$$N(r^N) = \frac{\exp[-\beta U(r^N)]}{Z} \quad 3.4$$

If this distribution is known prior to sampling our configurational space, allowing us to then perform moves with probability $N(r^N)$, we can then approximate equation 3.3 as a summation,

$$\langle A \rangle \approx \frac{1}{L} \sum_{i=1}^L n_i A(r_i^N) \quad 3.5$$

where n_i is the number of particles in configuration r_i^N and L is the total number of particles in the system. In fact, equation 3.5 is representative of the Boltzmann's distribution.

The Monte Carlo Method as we know it was first elucidated by *Metropolis et al.* [156, 157]. Many of the algorithms used in the construction of MC simulations are based on the construction of Markov Chain random walkers in order to sample an *a priori* probability distribution [158]. Markov Chains are constructed via a random walker that is capable of occupying discrete stages, with each stage dependent on the previous state of the system. A Markov Process is one for which the

current state of the system is dependent only on the state that came immediately before it [158]. The conditional probability of system X_t being in state S_i in a Markov Process is therefore,

$$P(X_t = S_i | X_{t-1} = S_{i-1}). \quad 3.6$$

The sequence of states produced by such a process is termed the Markov Chain, and these conditional probabilities can be termed the transition probabilities. Transition probabilities are commonly expressed as a transition matrix from one set of states \mathbf{x}' to a new set of states \mathbf{x} ,

$$T(\mathbf{x}' | \mathbf{x}) = P(\mathbf{x}' \rightarrow \mathbf{x}) \quad 3.7$$

where T is our transition matrix. Each transition matrix must satisfy the following conditions,

$$T(\mathbf{x}' | \mathbf{x}) \geq 0, \quad 3.8$$

$$\sum_{\mathbf{x}'} T(\mathbf{x}' | \mathbf{x}) = 1 \quad 3.9$$

Simulation of such a system was first derived by *Metropolis et al.* and was termed the Metropolis Scheme, which deals with the canonical (NVT) ensemble [157]. The more generalized form was further developed by WK Hastings [159]. When a Markov Chain MC simulation is constructed, we must ensure that the principle of detailed balance is maintained. That is, at equilibrium the simulation must ensure that the average probability of accepting a trial move from one state to the next must be equal to the probability of the reverse, implying that

$$N(\mathbf{x}')T(\mathbf{x}' | \mathbf{x}) = N(\mathbf{x})T(\mathbf{x} | \mathbf{x}') \quad 3.10$$

We can calculate the transition probabilities by noting that each transition involves the performance of a trial move, which is accepted with some probability and can be expressed as

$$T(\mathbf{x}' | \mathbf{x}) = \alpha(\mathbf{x}' | \mathbf{x})P(\mathbf{x}' | \mathbf{x}), \quad 3.11$$

where $\alpha(\mathbf{x}' | \mathbf{x})$ is the probability of performing a trial move and $P(\mathbf{x}' | \mathbf{x})$ is the probability of accepting this trial move. This provides the necessary condition proposed by Metropolis et al.,

$$P(x'|x) = \begin{cases} \frac{N(x')}{N(x)} = \exp\{-\beta[U(x') - U(x)]\} & \text{if } N(x') < N(x) \\ 1 & \text{if } N(x') > N(x) \end{cases} \quad 3.12$$

3.2.2 Canonical Protein Monte Carlo

In canonical MC simulations, we consider simulations with a constant number of atoms (N), constant volume (V), and constant temperature (T). This is also frequently termed the NVT ensemble. Thus the NVT ensemble describes the system that is in thermal equilibrium with a heat bath. The NVT ensemble has several properties that are well described in any number of statistical mechanics texts.

The NVT ensemble is easily simulated through the Metropolis-Hastings scheme. The algorithm for such a simulation, as proposed from by *Metropolis et al.*, is,

1. A particle r is selected at random.
2. The system energy is calculated.
3. A random displacement of the chosen particle is performed to r' , and the energy of this new configuration is calculated.
4. The move from r to r' is accepted with probability,

$$P(r'|r) = \min(1, \exp\{-\beta[U(r') - U(r)]\}) \quad 3.13$$

The displacement induced by the trial move of step 1 can be of several different types. The most basic type is a translational motion. When performing a translational trial move, one might add a random number to the position of the centres of mass of some particle. A simple expression may involve the calling of a vector of random numbers (**Ran**) in the range of 0 and 1, and then applying it to the vector of the particles coordinates \mathbf{r} , which might result in a move in the range of $-A/2$ to $A/2$. This displacement can be simply expressed as

$$r' = r + \frac{\Delta}{2} \left(Ran - \frac{1}{2} \right) \quad 3.14$$

The choice of value for Δ is dependent on the necessary speed of the simulation, as well as the properties of the system to be simulated.

The other class of trial moves include those that change the relative orientation of one molecule to another. This is particularly relevant to simulations of proteins and polymer chains, or other irregularly shaped molecules. In moving a rigid molecule or chain, the trial move may be performed in much the same way as a translation. Non-rigid molecular motion is slightly more complex.

Non-rigid molecules presents additional difficulties in the calculation and performance of trial moves. In this case, we resolve to treat cases of internal rearrangement of atoms in the molecule, as well as large scale rotations about backbone bonds. The simpler of these two cases involves the rotation about a single backbone bond without fixed-ends, as is shown in Figure 3.2.

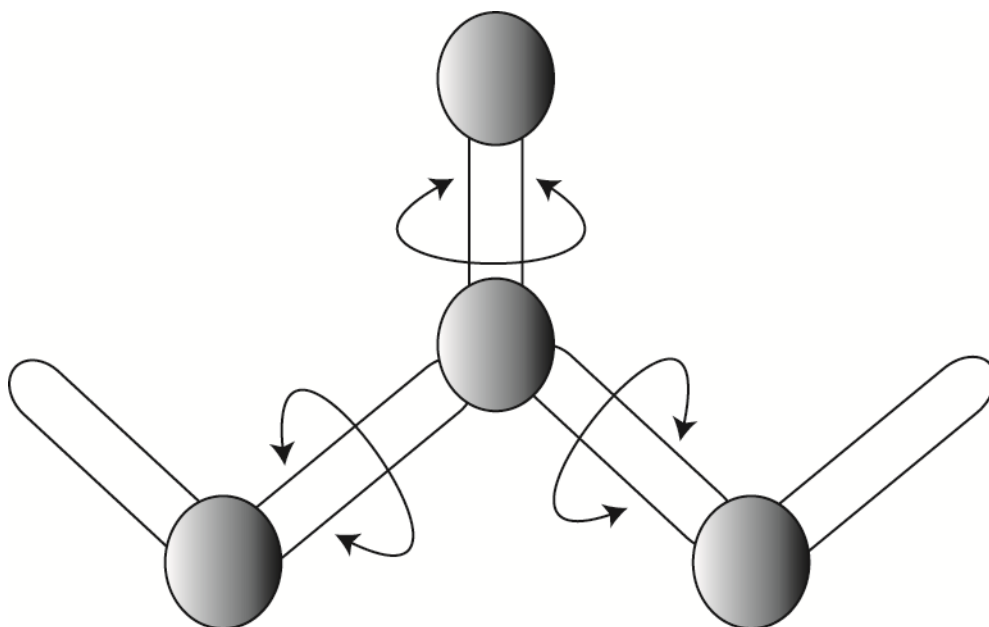


Figure 3.2: Schematic representation of possible MC moves, including side-chain rotation, as well as rotation about regions of peptide bonds.

Simple transformation of the Eulerian angles as presented for the rigid case is frequently possible. This can be accomplished simply by rotating the center of mass of one side of the bond in relation to that of the other. The case of fixed-end rotations produces a more complex system of sampling trial moves. This is accomplished through Configurational-Bias MC (CBMC) techniques. The first CBMC method was proposed by *Seipmann et al.* [160]. Their bias technique was built upon the work of *Rosenbluth et al.*, who produced a technique of sampling extended chain molecules via a self-avoiding random walk algorithm, referred to as Rosenbluth Sampling (RS) from here on [161]. The primary drawback of RS was its inability to produce a proper Boltzmann distribution in its results. The basic implementation of the CBMC algorithm involves the introduction of a biasing term to the calculation of the acceptance probability. The derivation presented in *Frenkel et al.* for the generalized case will be repeated here for brevity and completeness [155]. The derivation begins by first assuming our ability to perform a MC move according to some biasing function,

$$\begin{aligned}\alpha(x'|x) &= f[U(x')], \\ \alpha(x|x') &= f[U(x)].\end{aligned}\tag{3.15}$$

If we are then sampling configurations according to a Boltzmann distribution, this results in detailed balance being satisfied by

$$\frac{P(x'|x)}{P(x|x')} = \frac{f[U(x)]}{f[U(x')]} \exp\{-\beta[U(x') - U(x)]\}.\tag{3.17}$$

This subsequently gives us the acceptance rule,

$$P(x'|x) = \min\left(1, \frac{f[U(x)]}{f[U(x')]} \exp\{-\beta[U(x') - U(x)]\}\right).\tag{3.18}$$

In this work, sampling of internal configurations of the protein is performed using an approach termed Biased Gaussian Steps (BGS) [162]. The method used for performing MC simulations in this work relies on static bond lengths, necessitating biasing methods in torsional space. In

this method, n torsional angles along the backbone of the protein are considered. In our application of the BGS method, we have $n=8$. The algorithm for this method is dependent on the conformation dependent matrix \mathbf{G} , where $\delta\phi^T\mathbf{G}\delta\phi \approx 0$ [162]. This matrix provides a method to draw moves from a Gaussian distribution, hence the name Biased Gaussian steps. This bias potential is of the form,

$$P(\delta\bar{\phi}) \propto \exp\left[-\frac{a}{2}\delta\bar{\phi}_r(\mathbf{1} + b\mathbf{G})\delta\bar{\phi}\right], \quad 3.19$$

where a and b are parameters capable of being tuned to set the acceptance rate and bias. After these trial moves are attempted, the Metropolis-Hastings accept/reject step (eqn 3.13) is applied [162].

3.2.3 Replica Exchange Monte Carlo

Replica Exchange MC (REMC), also called Parallel Tempering, is a method of improving the overall sampling of the proteins conformational space [163-165]. This is required as a protein may get stuck in a local energetic minimum instead of traversing the conformational space. Exchanging simulation temperatures allow for the protein to access higher energy structures. This is accomplished by using multiple independent replicates M , over a temperature range $[T_0, T_{M-1}]$, where $T_0 < T_1 < \dots < T_{M-1} < T_M$. This method allows for the sampling of higher energy states at higher temperatures, while allowing for the investigation of potential minimum energy basins at the lower temperatures. In theory, this allows any low energy conformations to escape free energy troughs of the rough energy landscape of most biomolecular systems [164]. This inherently allows us to more effectively sampled the disordered conformational space of IDP's. This style of simulation involves an attempt at swapping temperatures between neighbouring replicas after a predefined number of steps, with an acceptance probability of

$$P(T_j | T_i) = \min\left\{1, \exp\left[(\beta_j - \beta_i)(U(\mathbf{r}_i^N) - U(\mathbf{r}_j^N))\right]\right\} \quad 3.20$$

where $\beta_{ij}=1/kT_{ij}$, and $U(\mathbf{r}_{ij}^N)$ are the energies of systems i and j respectively [163, 164].

3.3 Molecular Dynamics

3.3.1 Introduction

Molecular Dynamics (MD) simulations provide a means of numerically simulating complex N-body systems, and were first introduced by *Alder et al.* in 1955 [166]. Since these first MD simulations, significant improvements have been made in the implementation of various algorithms used to implement the simulation, as well as improvements in hardware speed and capability. This has led to MD becoming a widely-used tool in many areas of physics, chemistry, and biology. An example of a box setup for simulating water is shown in Figure 3.3.

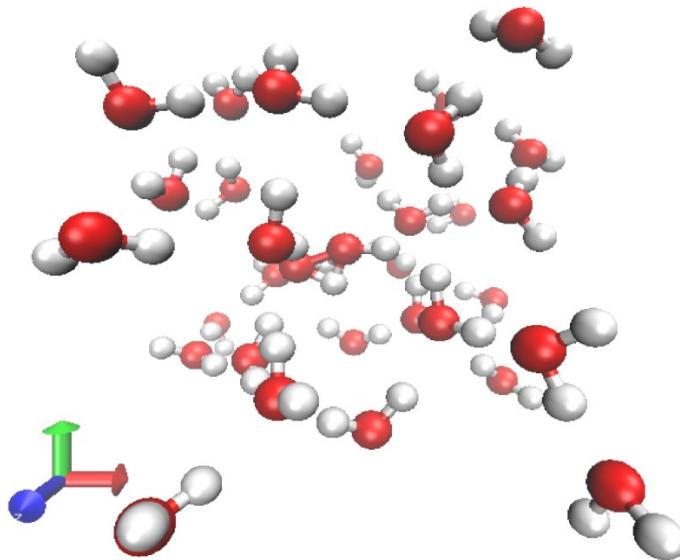


Figure 3.3: Atomic representation of a TIP3P water box.

The first uses of MD in the simulation of biomolecules occurred in the 1970's. These first simulations concerned the dynamics and folding of

bovine pancreatic trypsin inhibitor (BPTI), owing to its small size, stability, and accurate X-ray structure [167, 168]. Traditionally, statistical mechanics provides us a gamut of ways to calculate system properties. However, many of these solutions require assumptions which may not be correct at all times, or unfeasible in a laboratory setting. These may include assumptions of infinite size, or the requirement of extremely low or high temperatures. When these methods are inadequate, or a system is of sufficient complexity as to make such solutions unavailable, one can rely instead on computational simulations such as MD. At its heart, MD simulations provide the means of creating a virtual laboratory developed from these essential elements. The computation of these experiments is built around a simple algorithm, shown graphically in Figure 3.4. Once a system has been initialized (definition of initial velocities, positions, temperatures, etc),

1. Compute the forces acting on each atom.
2. Solve the equations of motion.
3. Update atomic velocities and positions etc.
4. Repeat.

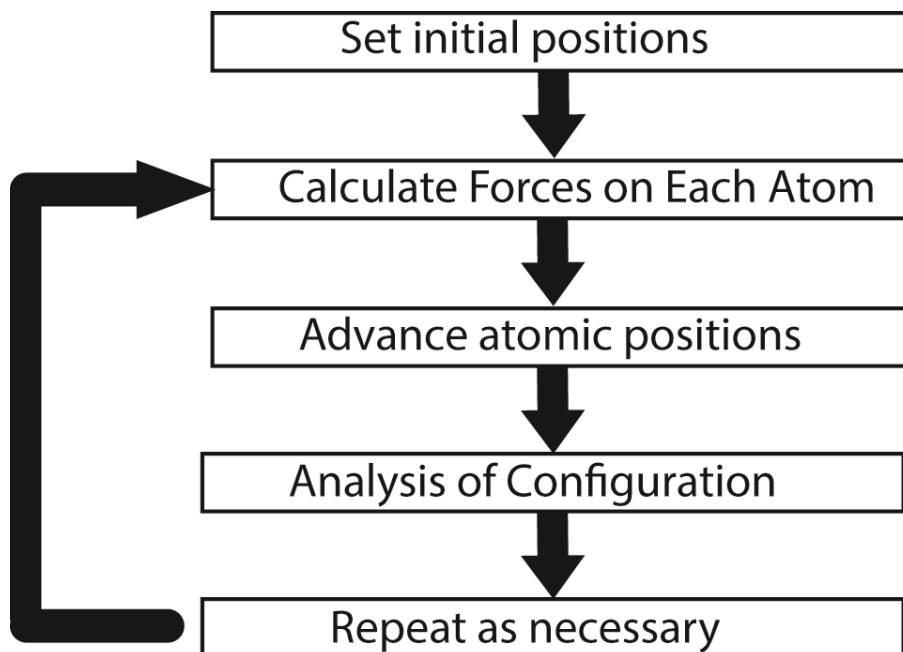


Figure 3.4: Graphical depiction of MD algorithm.

Thus, the evolution of the system is dependent/correlated with the initial conditions. Therefore, high quality initial structures increase the accuracy of the simulations. These structures are usually determined through experimental methods such as X-ray diffraction, or Nuclear Magnetic Resonance (NMR) [169, 170].

When performing a MD simulation, there is a typical progression of methods. Initially, the energy of the structure provided as input to the MD simulation package is calculated, and then the structure is changed step-wise in order to reduce or minimize the energy of the initial structure. Once the energy has been minimized, the system to simulate is then slowly heated to our predefined simulation temperature. After the system has finally reached this temperature, the system temperature and pressure are equilibrated to 300K and 1 bar respectively. The production simulation is then run for a prescribed length.

3.3.2 Equations of Motion

In order to calculate forces on each atom in our MD simulations, we first require the determination of the energy of the system on each particle. This may be accomplished through the calculation of the particles Hamiltonian,

$$H = H_0 + H_1, \quad 3.21$$

$$H = \sum_{i=1}^N \frac{p_i^2}{2m_i} + \sum_{i<j}^N u(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i<j}^N u^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots + H_1 \quad 3.22$$

where H_0 is the intra-protein term, dependent on the interactions of the atoms of our system with one another, and H_1 is the term describing the external forces or energy applied to the system. The potential energy term of H_0 is referred to as the force field (FF) describing the system. The FF is typically an empirical, or semi-empirical, potential of varying forms and parameters. These force-fields have been developed by a number of groups, with some examples being the CHARMM FF, and the AMBER FF [171, 172].

Common terms included in the atomic FF for proteins include bond-length, bond angles, torsional potentials, and atomic pair potentials. This can be expressed through the following potential energy function,

$$\begin{aligned}
 U(\vec{r}_1, \dots, \vec{r}_N) = & \sum_{\text{bonds}} \frac{a_i}{2} (l_i - l_0)^2 + \sum_{\text{angles}} \frac{b_i}{2} (\theta_i - \theta_{i0})^2 + \dots \\
 & + \sum_{\text{torsions}} c_i (1 + \cos(n(\theta_i - \theta_{i0}))) + \sum_{\text{atom pairs}} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \dots \\
 & + k_c \frac{q_i q_j}{r_{ij}}.
 \end{aligned} \tag{3.23}$$

Where l is the bond length, θ is the bond angle, σ is the minimum interparticle distance, q is the charge, ε is the potential well depth, r is the interparticle distance, and k_c is Coulomb's constant. This expression of the potential energy encloses descriptions of bond deformation (stretching and bending) in the first two terms, and torsional (dihedral) energies in the third term. These terms are described via harmonic parameters in the bending and stretching potentials (b_i and a_i respectively). The torsional term describes the rotational energy about a chemical bonds, which are periodic in nature, where n describes the periodicity of the interaction, and c_i describes the height of the rotational energy barrier. These forces are shown in Figure 3.5.

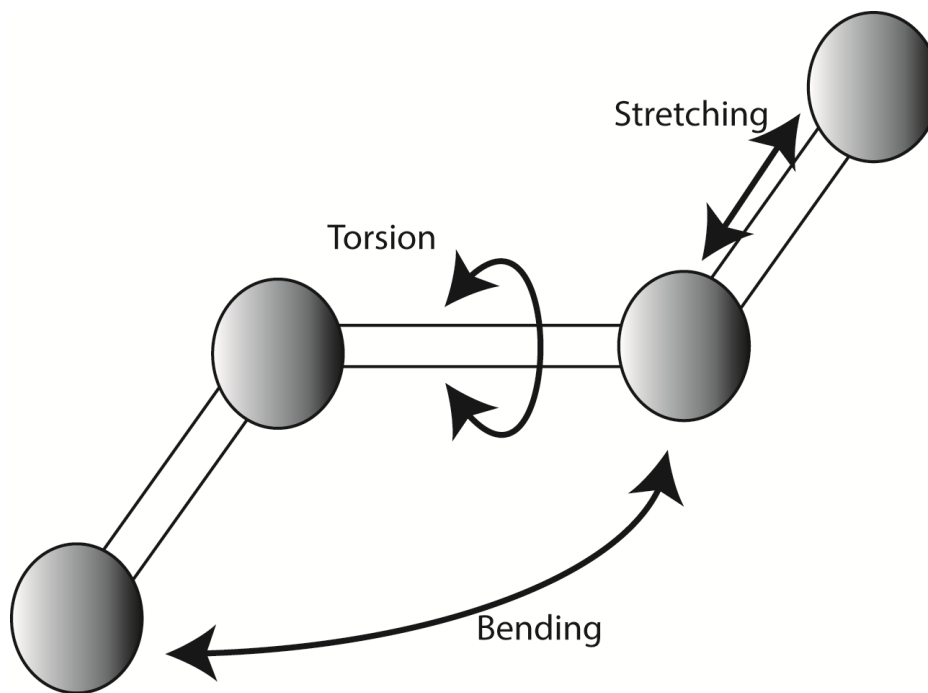


Figure 3.5: Graphical depiction of forces calculated in MD force fields.

The interaction energies between non-bonded atoms in the last two terms encompasses the charge-charge interaction of the last term, as well as the Lennard-Jones (LJ) interaction potential of the second to last term. In these terms, $r_{ij}=|\mathbf{r}_i-\mathbf{r}_j|$ is the absolute distance between atomic centers, q_i is the partial charge the atoms, ϵ_{ij} is the depth of the LJ well, and σ_{ij} is the minimum distance between atomic centers. For some potentials, such as that of the CHARMM FF, a term specific to out of plane bending is also included [172].

The speed of calculation of these potential functions is typically increased through specific methods for considering long-range interactions. These are classified as truncation methods or approximation methods to calculating long-range interactions [173]. Truncation methods can be classified as abrupt, in which particle interactions of particle-pairs whose distance is beyond some defined cut-off (r_{cut}) are not considered in calculating energies. Another choice of truncation may be in the use of a shifting function, in which the form of the potential energy function is

modified for distances between the switch distance and cut off, with pairs with distances greater than r_{cut} are not considered. Another possible option for calculating long distance interactions include mean-field techniques, such as the Reaction Field technique, which calculates interaction energies beyond a cut-off distance, using a mean-field method [173-175]. Where beyond r_{cut} , the interaction is expressed as

$$U(r_i, r_j) = q_i q_j \left[\frac{1}{r_{ij}} + \frac{(\varepsilon - 1)r_{ij}^2}{(2\varepsilon + 1)r_{cut}^2} \right], \quad 3.24$$

where ε is the dielectric constant of the medium that the simulation is taking place in.

The most common methods of calculating long-range interactions, such as the Particle Mesh Ewald and Particle-Particle/Particle Mesh Ewald, use the Ewald summation to calculate interaction energies beyond r_{cut} . This summation is expressed as

$$\begin{aligned} E_{Ewald} = & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left(\sum_{|\mathbf{n}|=0}^{\infty} q_i q_j \frac{\text{erfc}(\kappa |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} \right) + \\ & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{1}{\pi L^3} \sum_{\mathbf{k} \neq 0} q_i q_j \left(\frac{4\pi^2}{k^2} \right) \exp\left(\frac{-k^2}{4\kappa^2}\right) \cos(\mathbf{k} \cdot \mathbf{r}_{ij}) \right) \\ & - \left(\frac{\kappa}{\pi^{1/2}} \right) \sum_{i=1}^N q_i^2 + \left(\frac{2\pi}{3L^3} \right) \left| \sum_{i=1}^N q_i \mathbf{r}_i \right|^2 \end{aligned} \quad 3.25$$

where k is the reciprocal lattice vector, $q_{i/j}$ are the charges of particles i and j , κ is a parameter to set prior to simulation, and L is the length of one side of our simulation box.

Upon calculation of the energy of a particle, we can then move to calculating and solving the necessary equations of motion describing the force on the system. Newton's equation of motion for particles of position $\mathbf{r}_i(t)$ and velocity $\mathbf{v}_i(t)$ is

$$F_i = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = m_i \mathbf{a}_i = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_i). \quad 3.26$$

Solving this equation for each particle provides the complete information of the system. However, due to the many-bodies involved in typical MD simulations, the system of equations must be solved numerically. The most common methods for solving such a system of time-discrete equations are the Leap-Frog and Verlet algorithms, which are types of Finite-Difference methods [175, 176]. The Verlet Method solves for position and velocity pairs $(\mathbf{r}_i(t+\Delta t), v_i(t+\Delta t))$ at each subsequent time step based on the current position and velocity, assuming the acceleration of particle i is known. The expression for this integrator is thus,

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + (\Delta t)^2 a_i(t), \quad 3.27$$

$$v_i(t + \Delta t) = v_i(t) + \frac{a_i(t) + a_i(t + \Delta t)}{2} \Delta t. \quad 3.28$$

The algorithm for solving such a set of equations typically takes the form of the half-kick implementation (or velocity-Verlet), where the velocity at half-time step is first calculated, which is implement in the NAMD package for instance [175, 176]. This can be described by the following step-wise implementation;

1. First calculate the half-time step velocity:

$$v_i(t + (1/2)\Delta t) = v_i(t) + (1/2)a_i(t).$$
2. Then calculate the full time-step position:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + v_i(t + (1/2)\Delta t)\Delta t$$
3. Calculate $a_i(t+\Delta t)$ from the energy function at position $\mathbf{r}_i(t+\Delta t)$.
4. Calculate full time-step velocity:

$$v_i(t + \Delta t) = v_i(t + (1/2)\Delta t) + (1/2)a_i(t + \Delta t)\Delta t$$

3.3.3 Temperature Control

Temperature in a MD simulation is calculated as an extension of the equipartition theorem for kinetic energy, which states

$$\left\langle \frac{m_i v_{i,\alpha}^2}{2} \right\rangle = \frac{1}{2} k_B T, \quad 3.29$$

where T is the temperature of the system, k_B is Boltzmann's constant, m_i is the mass of particle i , and $v_{i,\alpha}$ is the component velocity $\alpha=(x,y,z)$ of particle i . Note that $\langle \dots \rangle$ represents the ensemble average of the system. This then gives us a calculable, system dependent expression for temperature,

$$T = \frac{1}{N_f} \sum_{i=1}^N m_i v_i^2, \quad 3.30$$

where N_f is the number of degrees of freedom of the system to be simulated.

Several methods to control or constrain temperature in MD simulations have been devised and can generally be classified as Stochastic dynamics, Strong-coupling, Weak-coupling, or as Extended system dynamics. These methods are used in the simulation of the canonical ensemble (NVT), which conserves the particle number, volume and temperature of the system, as well as the isobaric-isothermal ensemble (NPT), which conserves the number of particles, pressure, and temperature.

Temperature control in this work is controlled via a Langevin Thermostat, a type of stochastic method of constraining temperature, in NAMD [176]. This method is based on solving the modified Newton's equation,

$$m_i \ddot{\mathbf{r}}_i(t) = -\nabla U(\mathbf{r}_i(t)) - \gamma \dot{\mathbf{r}}_i(t) + \left(\sqrt{\frac{2\gamma k_B T}{m_i}} \right) R(t), \quad 3.31$$

where T is the temperature at which the simulation is to run, γ is the collision frequency, and $R(t)$ is a Gaussian Distributed Random Process. This can also be described through the particle momenta

$$\dot{\mathbf{r}}_i(t) = \frac{p_i}{m_i} \quad 3.32$$

$$\dot{p}_i(t) = F_i(t) - \gamma p_i(t) + f_i, \quad 3.33$$

where f_i is a random force. The expression for the random force in Langevin dynamics is that of a Gaussian Process, with dispersion/standard deviation of $\sigma^2 = 2\gamma m_i k_B T$. The numerical solution of the Langevin process for position is accomplished through the Brunger-Brooks-Karplus method [176, 177]. The original expression for this integrator, from *Brunger et al.*, is

$$\begin{aligned} \mathbf{r}_i(t + \Delta t) = & \frac{1}{\beta} \left\{ 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{[F(t) + R(t)](\Delta t)^2}{m_i} \right\} \\ & + \frac{1}{\beta} + \frac{\mathbf{r}_i(t - \Delta t)\gamma \Delta t}{2} \end{aligned} \quad 3.34$$

where $\beta = (1 + (1/2)\gamma \Delta t)$, and $R(t)$ is our previously stated Gaussian Random Variable, with the properties

$$\langle R(t) \rangle = 0, \quad 3.35$$

$$\langle R(t_i)R(t_j) \rangle = \delta(t_i - t_j).$$

This expression and method assumes larger particles are moving through a sea of smaller ones, which provide friction to the larger particles, resulting in velocity correction of the larger particles. This makes consideration of the collision frequency γ important for the results of the simulation.

3.3.4 Pressure Control

The NPT ensemble requires the maintenance of an approximately constant pressure, temperature and particle number. The temperature of the system can be calculated as mentioned in Section 3.3.3. The pressure however, is usually calculated through the ensemble average of the microscopic pressure [175, 178]. The expression for pressure is derivable through the ensemble average of the Clausius Virial Function $W(\mathbf{r}_1 \dots \mathbf{r}_N)$,

$$W(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i^{Tot}, \text{ and} \quad 3.36$$

$$\langle W \rangle = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau \sum_{i=1}^N \mathbf{r}_i(\tau) \cdot m_i \ddot{\mathbf{r}}_i(\tau). \quad 3.37$$

Through integrating Eq 3.37 by parts, and invoking the equipartition law,

$$\langle W \rangle = -\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau \sum_{i=1}^N m_i |\dot{\mathbf{r}}_i(\tau)|^2, \quad 3.38$$

$$\langle W \rangle = -DNk_B T, \quad 3.39$$

where D is the dimensionality of the system, N is the number of particles, T is the temperature of the system, and k_B is Boltzmann's Constant. If one were to separate the total force in to interatomic forces, and external applied forces, so that $\mathbf{F}_i^{Tot} = \mathbf{F}_i + \mathbf{F}_i^{Ext}$, and particles are enclosed within a container of volume V , we can express the external portion of the ensemble averaged Virial Function,

$$\langle W \rangle = \langle W^{Int} \rangle + \langle W^{Ext} \rangle, \quad 3.40$$

$$-DNk_B T = \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle - DPV, \quad 3.41$$

$$P = \frac{Nk_B T}{V} + \frac{1}{DV} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle. \quad 3.42$$

In this work, the NPT ensemble simulated in NAMD is based on an extension of the Nose-Hoover Langevin Piston Barostat [175, 176, 179, 180]. This method has been derived independently by two groups, resulting in virtually identical methods [176, 181]. This method relies on the introduction of a dynamic container volume V , a fictitious piston of mass W , and parameter e analogous to the strain introduced by *Quigley et al.* [181, 182]. The resulting equations to be solved are

$$\begin{aligned}
\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} + \dot{e}\mathbf{r}_i, \\
\dot{\mathbf{p}}_i &= \mathbf{F}_i - (\dot{e} + \gamma)\mathbf{p}_i + R, \\
\dot{V} &= 3V\dot{e}, \\
\ddot{e} &= \frac{3V(P - P_0)}{W} - \gamma_e\dot{e} + \frac{R_e}{W}, \\
W &= 3N\tau^2 k_B T, \\
\langle R^2 \rangle &= \frac{2m\gamma k_B T}{h}, \\
\langle R_e^2 \rangle &= \frac{2W\gamma_e k_B T}{h}, \\
\tau &= \text{The oscillation period},
\end{aligned}
\tag{3.43}$$

where R and R_e are the stochastic noise on the particle and piston, and γ and γ_e are the friction coefficients for the particle and parameter e respectively. This method requires a more complex numerical integration method, an example of which is shown in *Quigley et al* [181]. Within this method, the volume of the box is altered in order to maintain the preassigned pressure, as in Figure 3.6.

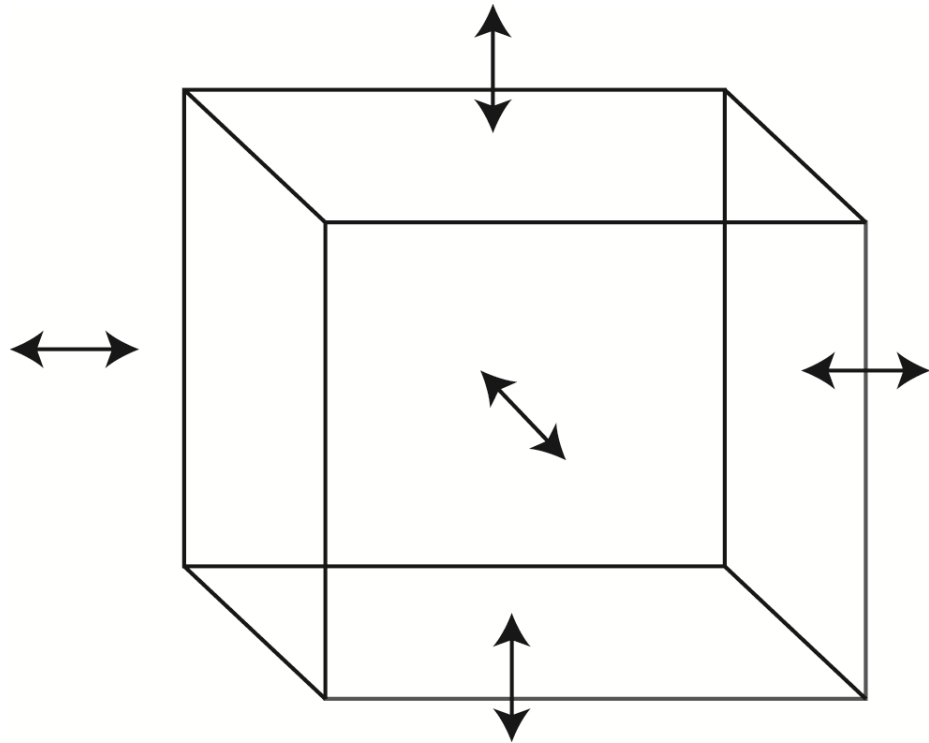


Figure 3.6: Schematic of pressure correction by volume correction.

3.4 Druggability

3.4.1 Introduction

A potential drug target is said to be druggable if it's able to bind small molecules at an appropriate binding affinity. Assessment of druggability through experiment or simulation is becoming a more dominant method of determining drug targets and potential therapeutics. This is due in part to the high failure rate, and exorbitant costs, in more traditional methods of therapeutic identification [183]. Druggability typically refers to binding pockets of a target, and not simply the target itself. A pocket or region of the target is considered druggable if it is able to bind drug-like molecules, which are defined by their physical properties[184].

The first step in assessing druggability is the identification of potential targets for therapeutics. The work of Hopkins and Groom estimates that approximately 10% of the human genome is potentially druggable [185]. However, simply being druggable is not enough to support a target's potential utility in drug design, it must also be therapeutically relevant. Upon successfully identifying the target protein, identification of potential binding regions in the protein becomes the next step. This can be accomplished through the use of various databases of ligand binding sites or experimental identification of druggable sites in concert with identification algorithms, or computational modelling of the protein, as is attempted in this work [186-189].

Several forms of computational assessment of protein druggability exist, including [183],

1. Cavity Detection Algorithms - These algorithms look for regions of a protein which exhibit inward curvature as the basis of locating regions which may act as binding pockets/domains. They are typically either geometric or energetic in nature [83].
2. Interaction Potentials – These methods are based around the physico-chemical properties of the underlying protein structures and probe molecules. These can include docking of fragments, or probes, to static targets. Other methods, which use more complete configurational space sampling via MD simulations, are capable of sampling the configurational space with probe molecules.

3.4.2 Model and Probes

In this work, we use the MD simulation based method of assessing druggability provided by the software package DruGUI [190]. DruGUI uses a mixture of probes in water to explore potential sites of interaction on the surface of the target. The main probes in the analysis would also need to be small enough in order to maximize the efficacy and speed of the simulation. The set of primary probes was decided through the occurrence of their functional groups in FDA approved drugs. The probes suggested

by DruGUI are isopropanol, isopropylamine (IPAM), acetate, and acetamide.

Each simulation consists of the target, and a mixture of probes to water of 20/80% ratio by volume. Initially, the system is minimized for a predetermined amount of time. The simulation then constrains the target to the initial position, and heats the system to 600K in order to maximize mixing of the probe/water mixture about the protein. The temperature is then lowered to 300K, the constraints are released, and the system is equilibrated in the NPT ensemble. The simulation is then run for a production time of 32-40ns.

3.4.3 Analysis

Analysis of the simulation takes place as outlined by *Bakan et al.* [190]. Each frame of the simulation is divided into a grid, and the density of probe molecules in each voxel is determined relative to reference. This allows for calculation of the binding free energy of each voxel as

$$\Delta G_{probe\ binding} = -RT \ln\left(\frac{n_i}{n_0}\right), \quad 3.44$$

where R is the ideal gas constant, T is the temperature in Kelvin, and n_i/n_0 is the density of probe molecules relative to a reference ensemble. This grid is further refined to identify regions near the protein called interaction spots. Interaction spots within 5.5 to 6.5 Å are then merged. When 7 or 8 interaction spots are merged, if the resulting maximal binding affinity is at least 10 μM, with a binding free energy of -6.86 kcal/mol or stronger).

3.5 Virtual Screening and Docking

Virtual screening refers to the use of computational methods to screen a library of ligands against some known structure or binding site to determine the ability of the ligand to bind. In this work, we employed the method of blind docking to screen a library of ligands for the ability to bind to our proposed structures. In blind docking, one does not assign a specific binding site to the structure, and instead attempts are made to bind

the ligand to the structure as a whole. This method was employed through the use of the AUTODOCK and AUTODOCK Tools program suite [191].

In this work, we employed the use of the Lamarckian Genetic Algorithm (LGA) to perform docking simulations. This method can be described as analogous to Lamarck's discredited ideas of transference of phenotypic characteristics as heritable traits. LGA uses many of the concepts of genetics in its description, including genes, genotype, phenotype, chromosome, fitness, and mutation. State variables for a ligand (orientation, conformation, etc) are referred to as a ligand's genes, and the coordinates of atoms in the ligand are referred to as its phenotype. The genotype of the ligand is described as a single chromosome. The fitness of the ligand in this context is the total interaction energy of the ligand and protein, evaluated using the AUTODOCK energy function [191]. Pairs of ligands may be mated through crossover, with selection of offspring based on fitness. In this process, LGA allows for the "death" of unsuitable ligands. Some ligands instead undergo mutation, at a rate defined by the user. These mutations alter the ligand's genes by a random amount, which then may alter a particular ligand's fitness, affecting its survival and ability to crossover.

Initially, a population of ligands is created within a box containing the protein/binding site defined by the user. Each ligand is given a random value for each of its genes, with torsion angles between $+180^\circ$ and -180° , and translation positions are randomly distributed through the grid. After the creation of the population, the search is looped over a number of generations. Between each generation, local searches about each member of the ligand population are performed to improve the interaction and intramolecular energies. Each generation consists of mapping and evaluating the ligand's fitness, selection, crossover, mutation, and elitist selection [191, 192]. Mapping simply takes the state variables of each genotype and translates them to an appropriate phenotype. This is followed by determining the fitness of the individual ligands by calculating their free energy:

$$\Delta G = \Delta G_{vdm} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) - \Delta G_{hbond} \sum_{ij} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^6} + E_{hbond} \right) + \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + \Delta G_{tor} N_{tor} + \Delta G_{sol} \sum_{i,j} S_{ij} V_j \exp\left(\frac{-r_{ij}^2}{2\sigma^2}\right), \quad 3.45$$

where the coefficients are determined by linear regression, q_i are the atomic charges, E_{hbond} is the mean energy of hydrogen bonding of water with a polar atom, N_{tor} is the number of torsional angles, and r_{ij} is the distance between atoms i and j [191, 192].

Upon determining each ligand's fitness, pairs of individuals are then proportionally selected to determine reproduction, with their reproductive success determined by fitness. The number of offspring is determined by

$$n_i = \frac{f_w - f_i}{f_w - \langle f \rangle}, \quad f_w \neq \langle f \rangle, \quad 3.46$$

where n_i is the number of offspring for ligand i , f_i is the fitness of ligand i , f_w is the fitness of the ligand with the highest energy, and $\langle f \rangle$ is the mean fitness of the entire population.

After reproduction, crossover and mutation are next performed. A random member of the ligand population is chosen, and two point crossover is performed. Crossover exchanges two non-contiguous gene segments. For example, for a pair of ligands, a chromosome that is broken into three pieces (ABC, abc) could result in the formation of AbC and aBc after crossover is performed, where AbC and aBc would replace the original parent ligands. Mutation is then performed on the resulting ligands. Mutation involves adding a random Cauchy-distributed number to one of the state variables representing the ligand's genes [191, 192].

4 α -SYNUCLEIN

MONOMERS AND

DIMERS

A version of this chapter was published in the European Biophysical Journal, 2016 May; 45(4):355-364.

4.1 α -Synuclein Monomers and Mutants

4.1.1 Outline

In section 4.1, we use canonical Monte Carlo protein simulations to effectively sample the phase space of monomeric wild type (WT) cysteine-terminated α -synuclein, as well as the well-known disease associated mutants A30P, A53T, and E46K. We aim to determine potential important variations in structural propensity that may help to elucidate regions of the protein which may be important for aggregation.

4.1.2 Methods

Simulations of WT α -synuclein and its mutants associated with familial disease were performed via canonical protein MC using the ProFASI software package [193]. 16 independent simulations of 12 million sweeps were run at temperatures between 290K and 340K. ProFASI simulations operate in the typical MC fashion outlined in section 3.1. The software calculates the energy of each conformation according to equation 4.1,

$$E = E_{ev} + E_{loc} + E_{hb} + E_{hp} \quad \mathbf{4.1}$$

Where E_{ev} is the energy due to the excluded volume effect, E_{loc} is the local electrostatic energy, E_{hb} and E_{hp} are the energies of hydrogen bonding and hydrophilic effect respectively.

4.1.3 Analysis

Individual trajectories produced by ProFASI for the α -synuclein simulations were transformed to dcd format via the catdcd plugin available with the Visual Molecular Dynamics (VMD) program [194]. The Amber Tools trajectory analysis program, ptraj, was used to determine the secondary structure propensities for each residue, as well as the correlation matrices for the structures in each trajectory [195]. In the case of the bimodal WT ensemble, a second set of simulations was run to determine if the initial simulation had been run long enough to converge. We analyzed the lengths of secondary structures formed, the frequency of each length, and the specific residues involved in these various lengths using custom scripts in the Octave package [196, 197]. Comparison of ensembles to the available NMR results was performed using the SHIFTX2 package, which calculates theoretical chemical shifts for each structure, and generates ensemble averages [198]. These theoretical values were then compared to NMR structures available in the Biological Magnetic Resonance Bank [199]. Further comparison to experiment was performed by calculating $J(H^N, H^\alpha)$ coupling values via the Karplus equation [200], as well as calculating R_h values using the program HYDROPRO [201].

4.1.4 Results

We first examined the distribution of energies for the conformations of wild-type and mutant α -synuclein at each temperature value, as a sign of possible structural differences between the protein variants (Figure 4.1). At the highest simulation temperature, 340 K, the energy distributions for all four versions of the protein (WT and mutant) converged to a single peak (Figure 4.1A). As the temperature was lowered, a bimodal distribution indicative of a phase transition was observed for each protein (Figure 4.1,B–D), being most evident at 320 K (Figure 4.1 C). Variations in the distributions between the WT and mutant proteins also became apparent, implying differences in the temperature for the structural phase transition in each protein variant. Wild-type α -synuclein was found to have the lowest average energy compared to the three mutants, in each case. Whereas the differences between the α -synuclein variants were in most cases not very large, the A30P mutant was notably different from the other variants. To ensure convergence, a second set of 16 independent simulations was run at 320K for the WT, and found their energy distributions to have converged (to within error) (Figure 4.2). With this mixed phase simulation showing convergence, the setup for our simulations would likely also converge at higher temperatures, given the increased sampling that occurs at higher temperature. At the lowest temperature (Figure 4.1F), the distributions appeared to converge again towards a single peak, although small differences between the four variants remained.

To ensure that our results were consistent with previous work, we compared the NMR chemical shifts that would be expected theoretically from the ensemble of simulated structures (using SHIFTX2) to the values measured experimentally by NMR. The bimodal peaks were decomposed to form distinct high and low energy peaks of approximately the same size. The sides of each peak were fitted to a Gaussian distribution in order to produce these distinct peaks. Because experimental results exhibit properties consistent with mixed phases at low temperature, or mainly high-energy phase at higher temperatures, we focused on the high-energy

peak in the simulated distribution. The high-energy peak of the distribution at 320 K led to predicted chemical shifts that agreed well with NMR data collected at 285.5 K and a pH 6.5 [202]; the observed root-mean-square deviations (RMSDs) of 0.48 and 1.81 ppm for ^{13}C and ^{15}N shifts, respectively, compared to expected RMS errors for SHIFTX of 0.9 ppm (^{13}C), and 2.4 ppm (^{15}N) [198, 203]. Differences in simulation and experimental temperature are due to the fact that PROFASI's temperature scale is not physical, and instead calibrated using a smaller protein, Trp-Cage. We note that our results are also consistent with previous MC simulations of α -synuclein [204].

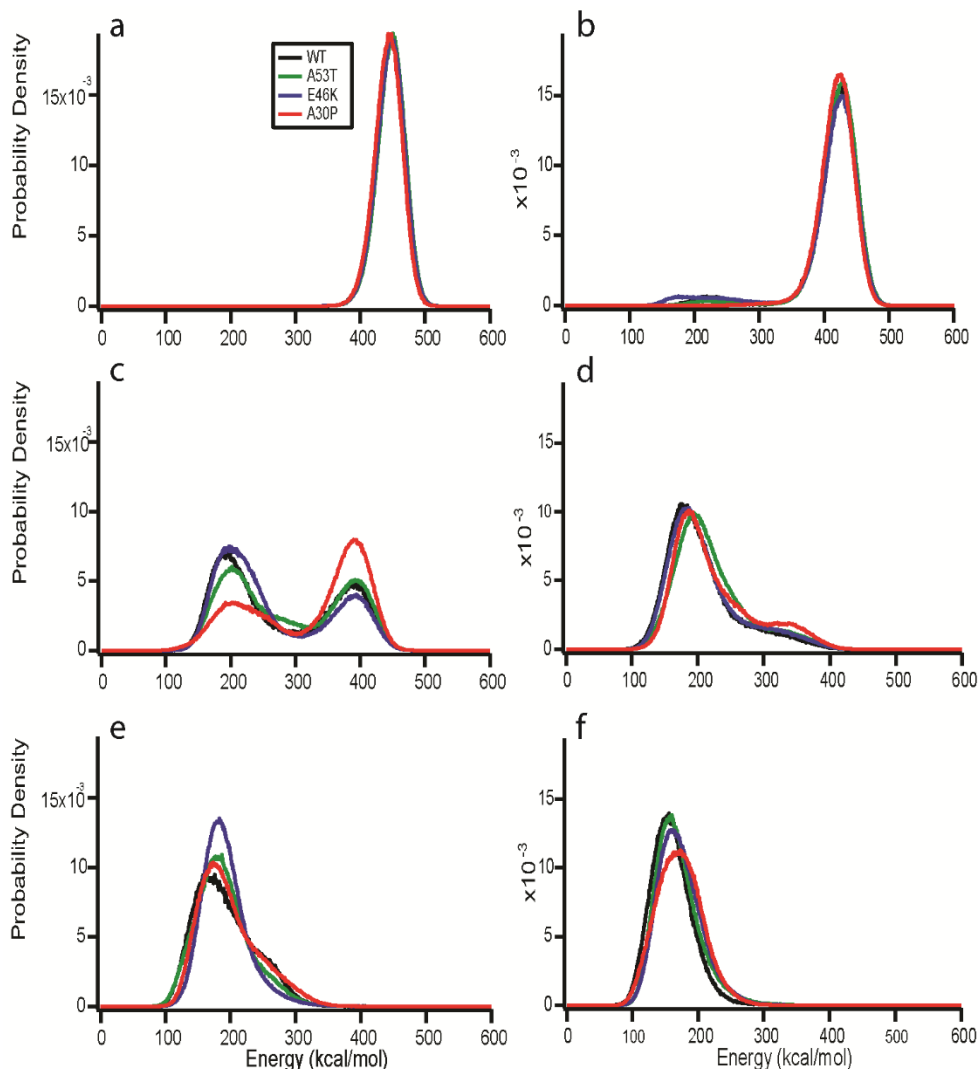


Figure 4.1: Distribution of conformational energies of wild-type and mutant α -synuclein as a function of temperature. (a) 340 K. (b) 330 K. (c) 320 K. (d) 310 K. (e) 300 K.

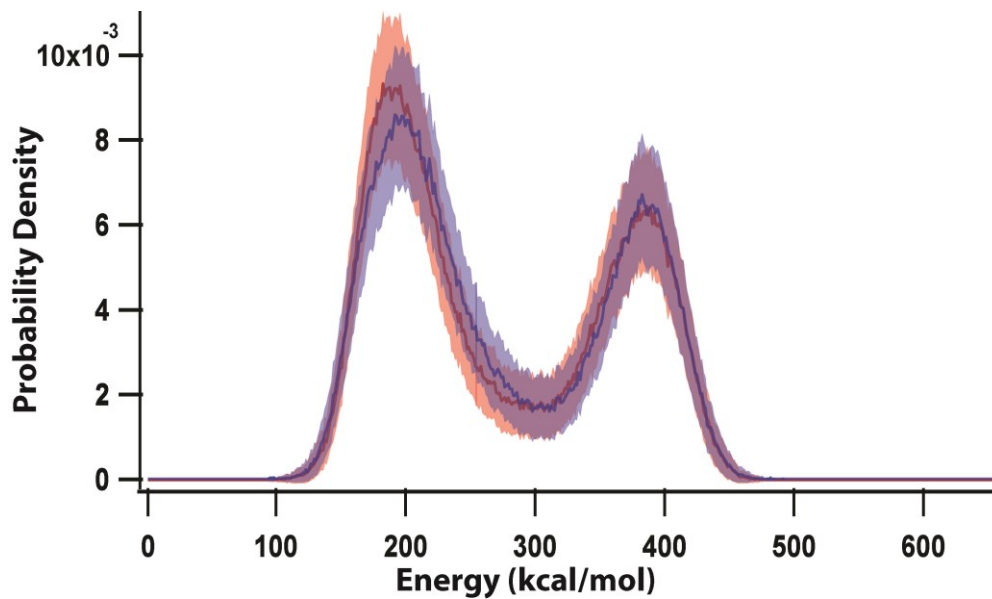


Figure 4.2: Energy distributions of two sets of 16 independent simulations run at $T=320\text{K}$, for 12 million MC sweeps. Shaded regions represent Jackknife 1σ errors and exhibit significant degrees of overlap between the two simulations, indicating convergence in energy.

Theoretically predicted paramagnetic relaxation enhancement (PRE) intensity ratios were estimated through the use of the measure $I-p_{ij}$, where p_{ij} is the contact probability, and were compared to experiment. We found that our results are comparable to the measurements performed by *Wu et al.* as well as those computed by *Jónsson et al.* [204, 205]. This includes an apparent extended region of long-range contacts about residue 90 (residue 91 in Cys-terminate α -synuclein), and the presence of a contact “bump” between residues 80-100 for PRE-labelled residue 132 (133 in Cys-terminated α -synuclein). Extended contact regions also exist about residue 19 and 132 (residue 20 and 133 in Cys-terminated α -synuclein) also exhibit extended regions similar to those found by *Wu et al.* and *Jónsson et al.* [204, 205]. This is shown in Figure 4.3.

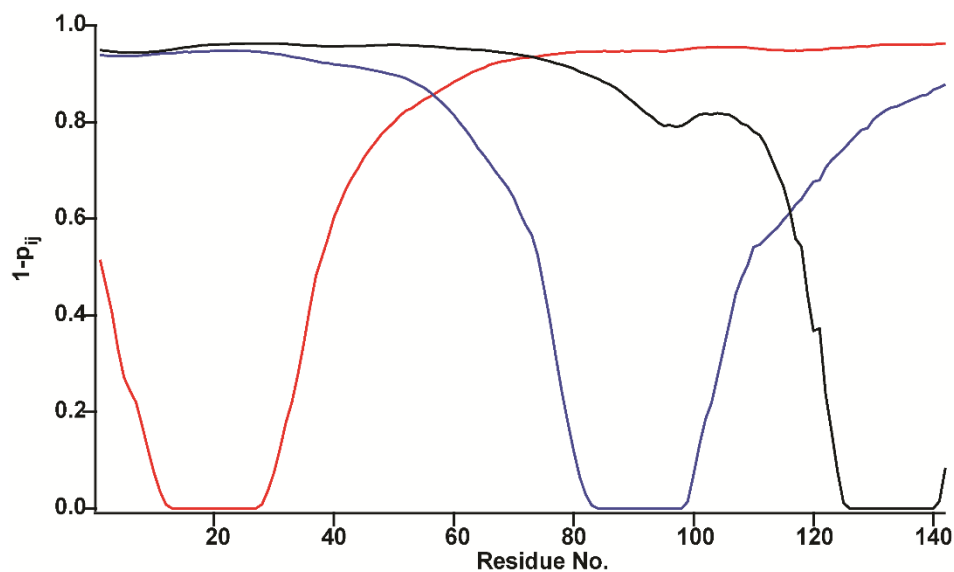


Figure 4.3: $1-p_{ij}$ values for residues in our simulations corresponding to WT $i=19$ (red), 90 (blue), and 132 (black), in our high energy ensembles produced at $T=320\text{K}$. These measurements have been used as an estimation of spin labelled PRE results.

Calculation of scalar $^3J(\text{H}^N, \text{H}^\alpha)$ couplings was performed for both peaks via the Self Consistent Karplus Equation [200]. The results were then compared against experiment. *Kim et al.* found a mean coupling value of 7.44 Hz, whereas we found a mean of 8.072 Hz for the low-energy phase and 6.31 Hz for the high-energy phase, which are very close to the values of 8.09 Hz and 6.43 Hz, respectively, found by *Jónsson et al.* [204, 206]. The low energy peak exhibited J-coupling values in good agreement with *Cho et al.*'s low pH system, with RMSD of 1.006 Hz and mean offset of 0.2 Hz [207]. However, when compared to the results of *Lendel et al.*, the low energy-peak exhibited an RMSD of 1.50 Hz, and a mean offset of 1.2002 Hz [208]. The high-energy phase exhibited greater variation, with an RMSD of 1.40 Hz, and a mean offset of 1.03 Hz, when compared to *Cho et al.* [207], whereas, the high energy phase showed an RMSD of 0.58 Hz, with a mean offset of 0.05 Hz when compared to *Lendel et al.* [208]. These results are also consistent with the observations of *Jónsson et al.*, using the Wang-Landau Monte Carlo method [204].

We also calculated the radius of hydration (R_h) values for both the low- and high-energy distributions. 1000 structures were randomly selected from each peak and the values of R_h were determined using HYDROPRO [201] by randomly extracting 1000 structures from each distribution, and we found the WT peaks to be at $25.9 \pm 1.5 \text{ \AA}$ and $30.8 \pm 2.4 \text{ \AA}$, respectively. The R_h value for the high-energy distribution aligns quite well with previous experimentally measured values, as well as previous simulations, which put it at approximately 32 \AA [204, 205, 207, 209, 210]. The measurement of the low energy phase R_h is also consistent with *Jónsson et al.*'s simulations where they measure a mean value of 26.0 \AA , and *Wu et al.*'s low-temperature measurements of 28.1 \AA [204, 205]. The consistency of our results with that of previous simulations and experiments, as well as consistency in comparison to a second set of simulations at 320K, allow us to justify our simulation set-up, and temperature-comparisons. Peaks produced by AS mutants were also decomposed in order to determine potential variations in structural propensities for the high and low energy peaks.

We next analysed the propensity of each temperature ensemble to form secondary structure for each protein variant, as a function of temperature, concentrating on the range 310–340 K around the phase transition. Our focus in this case is primarily on those distributions formed at our phase transition temperature and those formed at temperatures greater than the apparent phase transition temperature. Distributions below this point exhibited properties that were unverifiable, or outside of the known properties of low temperature experiments (eg. R_h values significantly lower than known values), and were hence excluded. The frequency with which each residue took on dihedral angles consistent with β -strands (Figure 4.4) and α -helices (Figure 4.5) was measured from the simulations. For both types of secondary structure, the per-residue frequency at high temperature (340 K) was quite similar across the entire protein for wt α -synuclein and the A53T and E46K mutants (Figure 4.4A, Figure 4.105A), but significant local differences were seen for A30P near the mutation site, including a reduction in helical propensity. At 330 K, the

results for helices were similar to the results at 340 K (Figure 4.105B), but differences were observed in the β -strand propensity, with E46K being more prone to β -strands than wt α -synuclein whereas the other two mutants were generally less prone (Fig. 3B). There was also an overall increase in the per-residue probabilities for both structures. At 320 K, the trends in β -strand propensity continued (Figure 4.94C), but A30P became more liable than WT α -synuclein to form helices, whereas E46K became less so (Figure 4.94C); again, the overall per-residue probabilities increased in all cases. Finally, at lower temperature the four variants begin to converge again in their β -strand propensity (Figure 4.4D), albeit with a few notable deviations, but A30P remains somewhat different from the others in its helical tendencies, especially near the N terminus (Figure 4.5D).

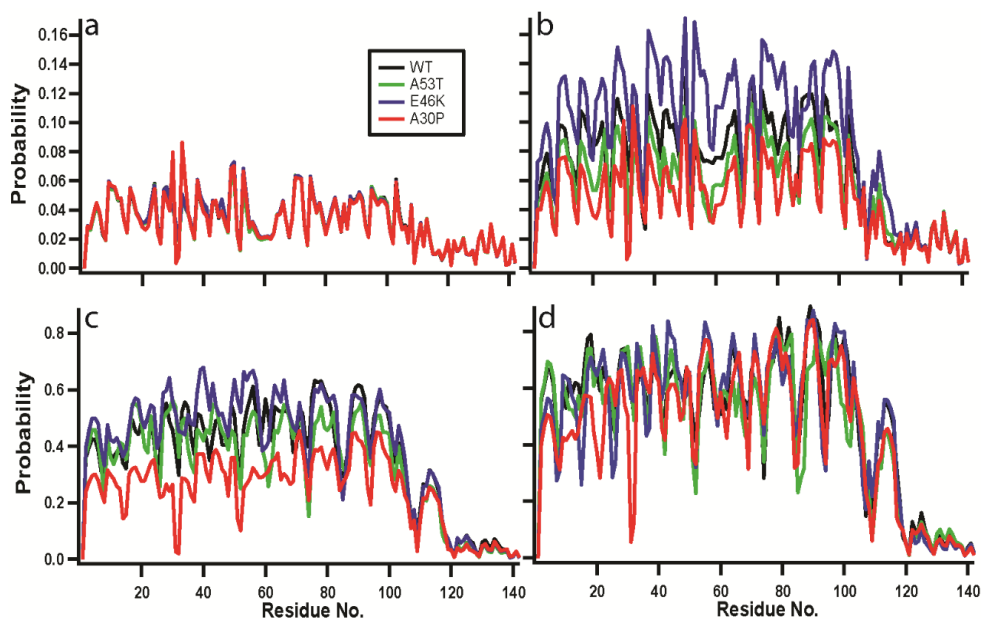


Figure 4.4: Per-residue probability of dihedral angles consistent with β -strand secondary structure as a function of temperature. (a) 340K. (b) 330K. (c) 320K. (d) 310K. Black: wild-type α -synuclein, blue: E46K, green: A53T, red: A30P.

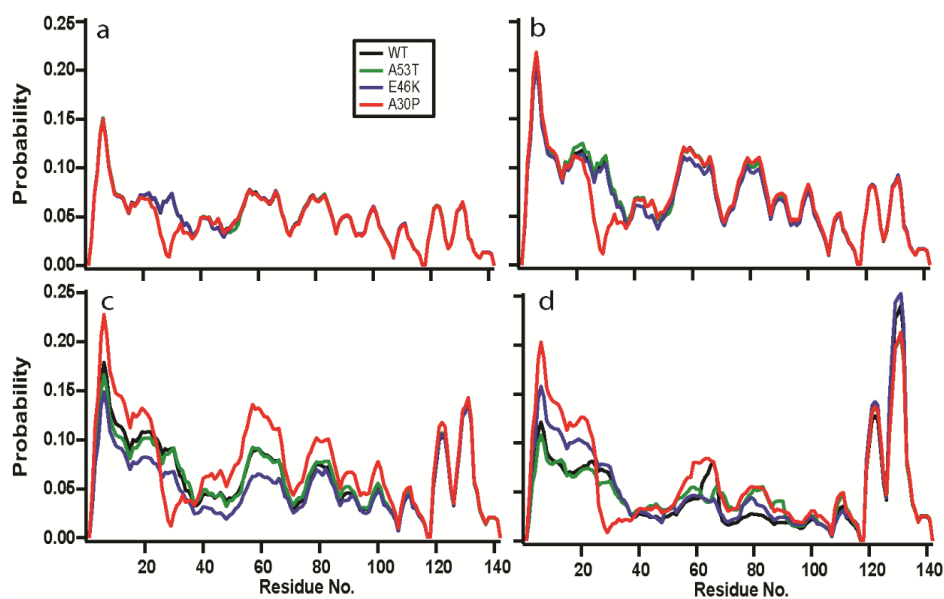


Figure 4.5: Per-residue probability of dihedral angles consistent with α -helix secondary structure as a function of temperature. (a) 340K. (b) 330K. (c) 320K. (d) 310K. Black: wild-type α -synuclein, blue: E46K, green: A53T, red: A30P.

Per-residue conformational frequency analysis is a useful tool, but it does not capture correlations between neighbouring residues and thereby reveal the frequency with which amino acid residues might form extended structures. The latter is more indicative of actual secondary structure formation, rather than transient fluctuations that are merely consistent with the ability to form structure. We therefore also investigated the frequency with which extended segments of the peptide chain formed dihedral angles consistent with β -strands or α -helices, reflecting extended structure.

Considering first the β -strands, we divided structures into “short” strands of fewer than 5 residues, and “long” strands of 5-12 residues; strands longer than 12 residues were ignored as they occurred with a frequency of less than 0.001%. These lengths were chosen as they are commonly-reported strand lengths in NMR experiments on aggregates [8, 211], and they are also the most commonly-formed extended structures as determined via joint probability plots (Figure 4.6).

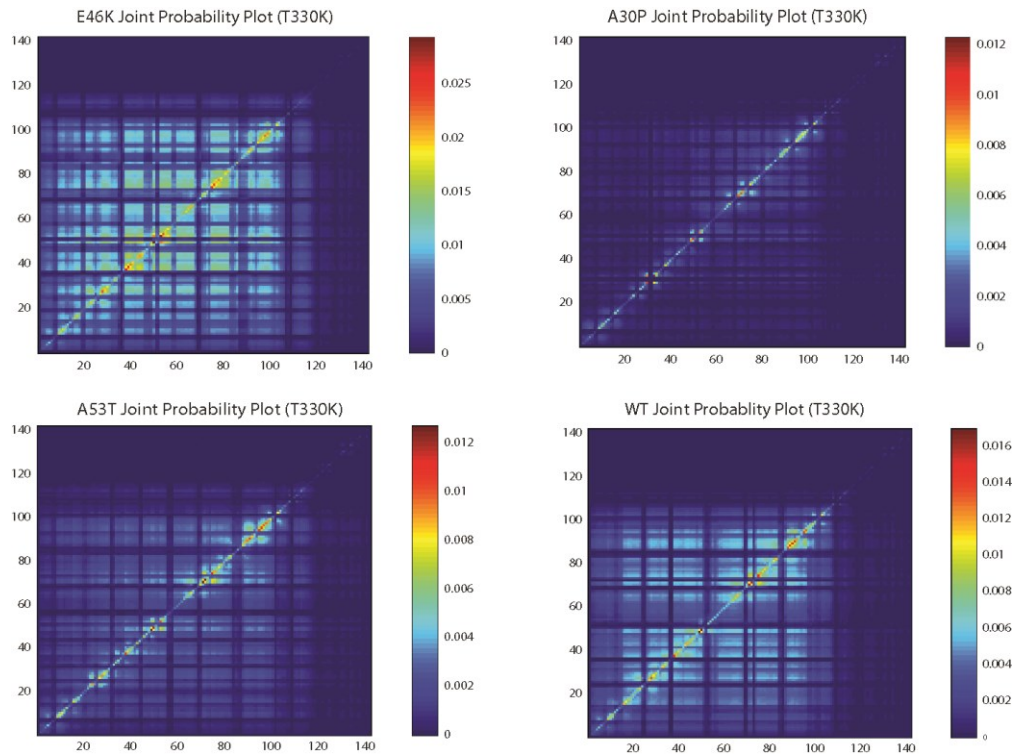


Figure 4.6: Joint probability plots used to determine common sizes of extended beta sheet structure formation at 330K.

Short β -strand segments were dominated by “strands” of only 2 residues; perhaps not surprisingly, the short strands occurred at a rate that was very similar to the per-residue frequency of β -strand angles (Figure 4.6).

The probability that a given residue formed part of a longer strand (Figure 4.7) displayed some of the same trends observed in the per-residue probabilities, including a decreasing probability of structure at higher temperature and quasi-periodic regions with increased propensity for strand formation. At 340 K, all four variants had similarly low strand probability, although A30P was notably lower in the region around residue 30. At 330 K, the variants were more differentiated: E46K was generally more prone to form strands than the other forms of α -synuclein, especially near the N terminus, between residues 2-18, 21-31, 38-45, 50-57, 60-69, 73, 75-82, 85-88, 91-92, 97-102, 110-122, 127, 130-136, and residue 139. A53T was more prone than WT to form strands at residues 24-25, 36-55, 59-64, 66, 69-73, 86, 110-119, 123-127, 129-138, and 140. A30P was less likely to form strands over most of the sequence, but showed increased propensity at residues 36-37, 50-56, 70-73, 86, and residues 140 and 141.

At 320 K, A30P remained much less likely to form strands than the other variants, which were otherwise quite similar over the C-terminal half of the protein; A53T and especially E46K were more prone than wt to form strands in the N-terminal half. Finally, at 310 K, the probability for strand formation in A30P was only slightly lower than for WT (except at residue 30), and in some regions (e.g. 50–100) A53T was lowest.

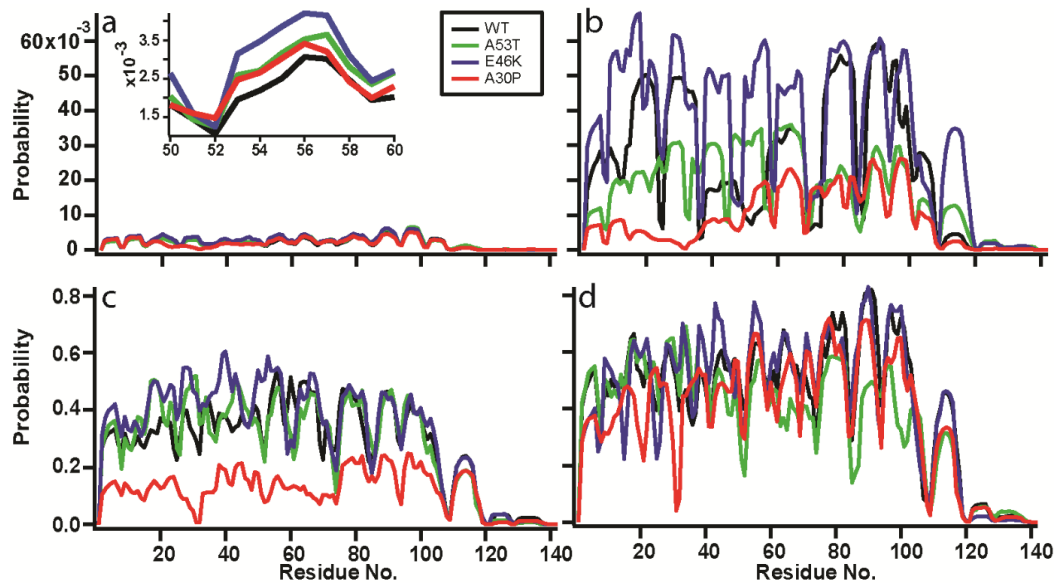


Figure 4.7: Probability of a given residue forming part of a β -strand of length 5-12 residues. (a) 340K. (b) 330K. (c) 320K. (d) 310K. Black: wild-type α -synuclein, blue: E46K, green: A53T, red: A30P. Inset of figure a shows the segment from residue 50 to 60.

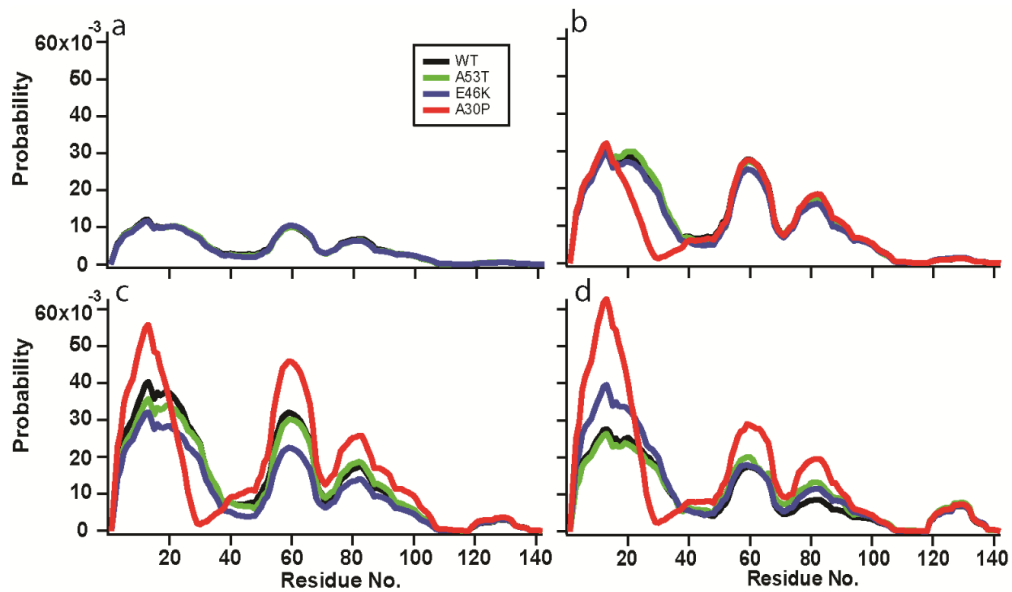


Figure 4.8: Probability of a given residue forming part of a helix of length 11–20 residues. (a) 340K. (b) 330K. (c) 320K. (d) 310K. Black: wild-type α -synuclein, blue: E46K, green: A53T, red: A30P.

The same kind of analysis was applied for α -helical structures, again divided into short helices (2–10 residues) and long helices, Figure 4.8, (11–20 residues; helices longer than 20 residues occurred in less than 0.005% of extended structures). At 330–340 K, both long and short helical segments formed with similar frequency in E46K and A53T mutants as in WT α -synuclein, but helix formation was depressed in residues 17–39 for A30P. In contrast, at 320 and 310 K (below the phase transition), A30P formed long helices more frequently than the other variants for residues 2–19 and from 40–100, but less frequently for residues 22–38 around the mutation site. The A53T mutant behaved very similarly to WT α -synuclein at both temperatures, with a slightly higher helix propensity for residues 60–100, but whereas E46K had generally lower helix propensity than WT at 320 K, it was generally a bit higher at 310 K, especially at the N terminus.

4.2 Dimers

4.2.1 Outline

In this section, we will be discussing temperature varying canonical MC simulations performed on α -synuclein monomers that have been covalently linked by a Gly-Ser-Gly tag to produce a contiguous dimer. We will then analyse the resulting distributions temperature dependent behaviour and compare to experiment where possible. We will also analyse the the distributions for the presence of multiple phases. These phases will subsequently be used in the production of discrete peaks to be used in the pulling simulations performed in the following chapters, in order to determine variation in formation of stable structures.

4.2.2 Methods

Simulations were performed using the standard protein MC protocol outlined in Chapter 3. These simulations were performed over the MC temperature range of 300-360K in 10K increments. At each temperature, 16 independent MC runs were performed for 5.5 million MC sweeps each in order to maximize the sampling of the proteins conformational space. Due to the size of the protein, these simulations took in excess of 12 weeks to perform, making substantially longer simulations unfeasible. The subsequent trajectories were then converted to dcd format through the use of the catdcd plugin, in order to take advantage of AMBER's extensive analysis program, cpptraj [194, 195].

4.2.3 Analysis

Trajectories produced by these simulations were grouped according to temperature, and their chemico-physical properties were then analysed using cpptraj, and the open source math package Octave [195, 196]. Analysis was undertaken to determine temperature dependence of energy, radius of gyration, secondary structure, and free energy landscape.

Comparison to the experiments of *Pivato et al.* were performed wherever possible [212].

4.2.4 Results

In first analysing our results, the simulations for each temperature were combined in order to maximize the apparent sampling for each of the potential distributions. The energy distribution at each temperature was then plotted and analysed to determine the potential existence of multiple phases, the resulting distributions are shown in Figure 4.9.

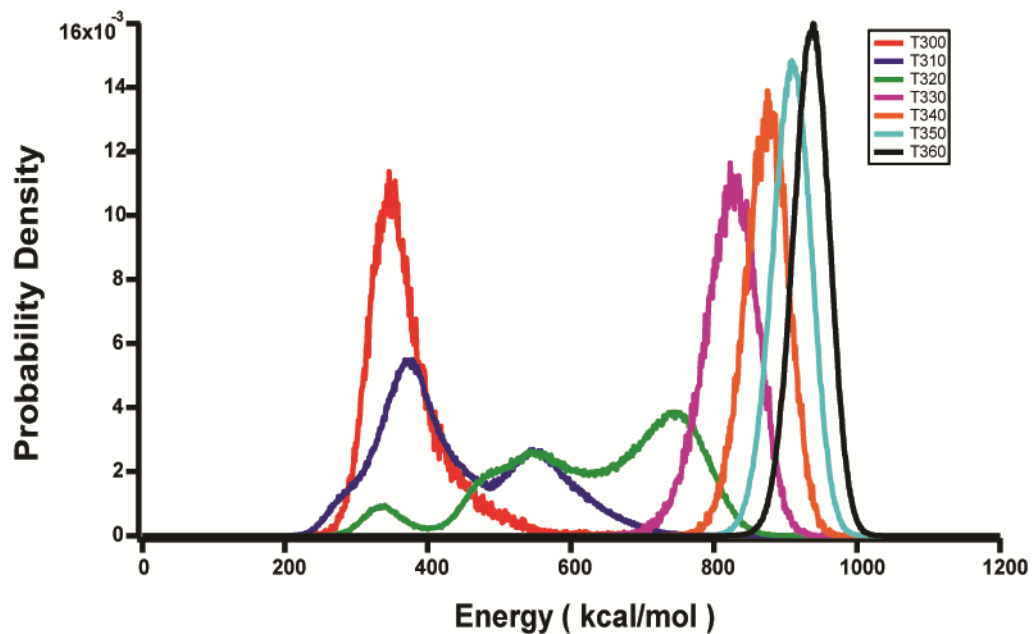


Figure 4.9: Plot of energy distributions for each simulation temperature produce for dimeric α -synuclein.

We can see that with increasing temperature, the energy of the distributions also increases as expected. However, at simulation temperatures 310K and 320K, we see the potential existence of three distinct phases. This is further supported by the appearance of multiple peaks in the distribution of R_g values, as shown in Figure 4.10. In this figure, we can see there exists two clearly defined peaks at $T=300$ K and 310K. One sharp peak between 20 and 25 Å, with a second peak between 25 and 35 Å. There also appears to be a transition to a broader peak above 35 Å, which suggests a potential third R_g peak. This suggests a more

complex free energy landscape than the two phase peaks of monomeric α -synuclein.

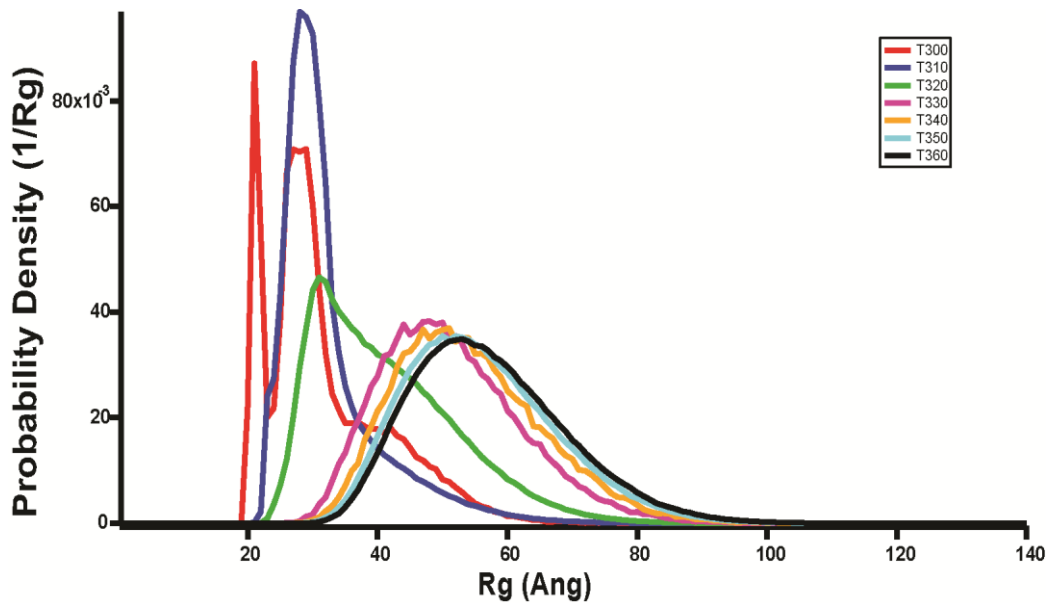


Figure 4.10: Plot of distributions of Radius of Gyration measurements for dimeric α -synuclein

We then grouped the results for simulations at $T = 310\text{K}$ and 320K , as they exhibit the apparent three energy states, resulting in the energy distribution shown in Figure 4.9. The resulting free energy landscape of the dimer can then be determined through the equation,

$$F = -\ln(P(R_g, ASA)) \quad 4.2$$

where F is the free energy, R_g is the radius of gyration, ASA is the accessible free surface area, and $P(R_g, ASA)$ is the probability of occurrence of both measurements. This results in a multi-phased free energy shown in Figure 4.11.

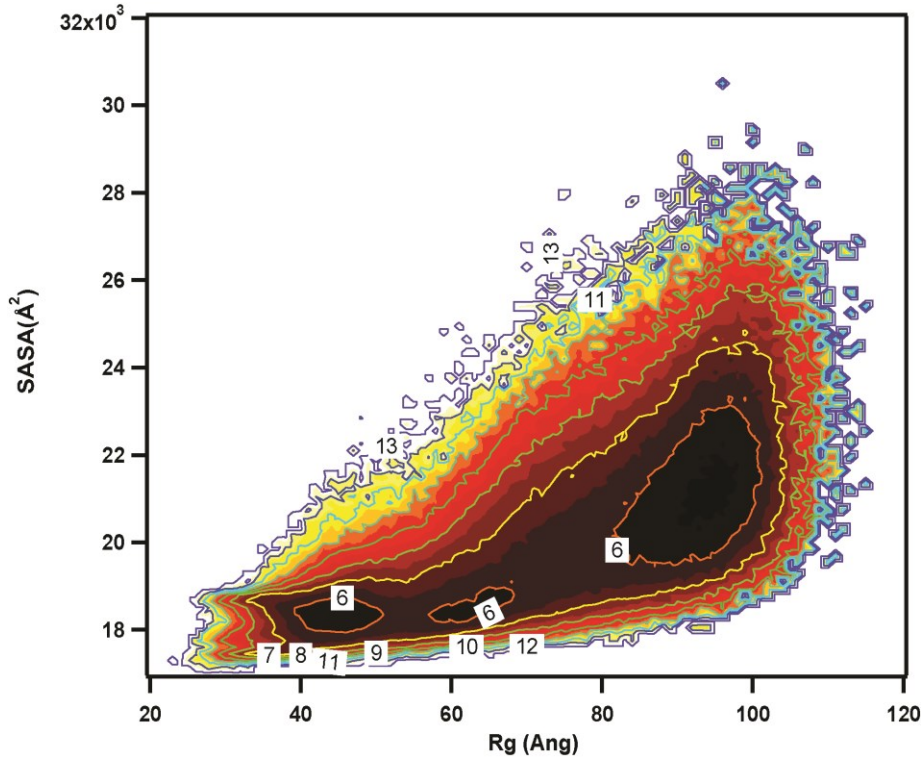


Figure 4.11: Free energy contour plot for covalently linked AS dimers. Free energy is calculated as $F = -\ln(Rg, SASA)$.

Measurements of the correlation between the centres of mass of residue pairs were also analysed for each temperature. The correlation is calculated as

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}}, \quad 4.3$$

where \bar{x} and \bar{y} are the mean positions of residue x and y . Plots of correlations for simulation temperatures of 300K, 310K, 320K, and 330K are shown in Figure 4.12.

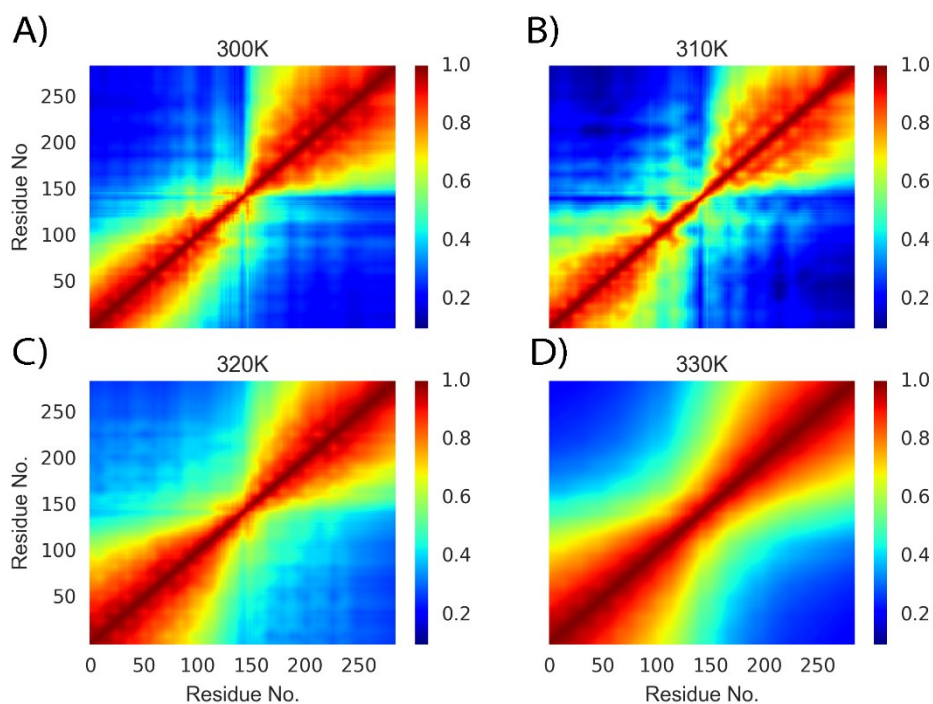


Figure 4.12: Residue-Residue correlation for covalently bound dimers of AS for each of 4 temperatures A) 300K B) 310K C) 320K D) 330K.

As can be seen from these plots, the two linked monomer domains exhibit little to no correlation with one another, which has been suggested by FTIR measurements performed by *Pivato et al.* [212]. Although *Pivato et al.* used a different linker between monomers, it is likely that the comparison is still valid due to the intrinsic disorder of AS, and comparable linker length. Further comparison to experiment was performed by comparing the degree of secondary structure formation over the simulation temperature range, including proportion of the structure in BS, Helix, and Random structure. This is shown in Figure 4.13.

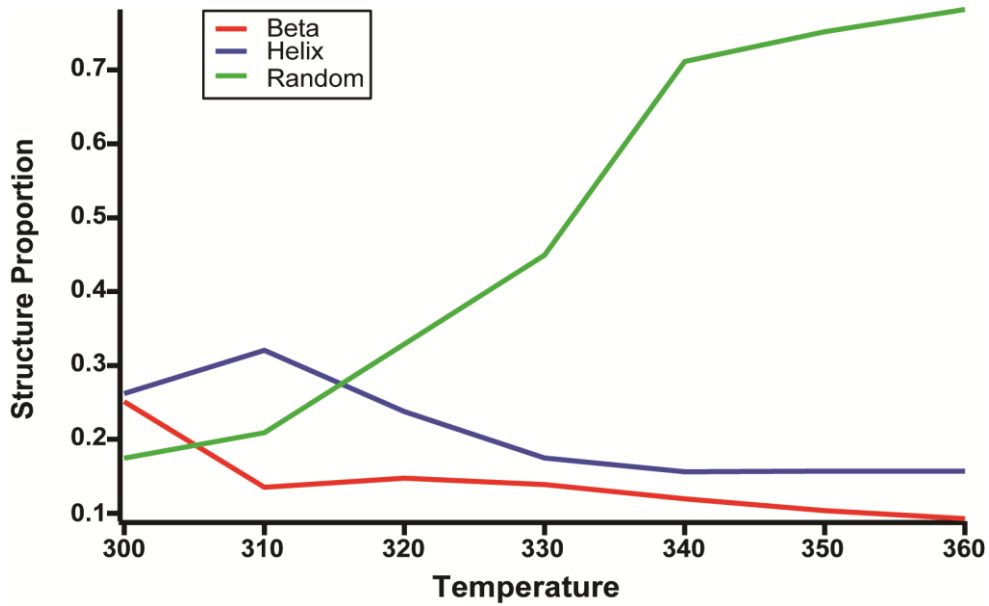


Figure 4.13: Proportion of covalently linked dimer exhibiting specific secondary structure. (Red) Beta Sheet (Blue) Helix (Green) No structure.

From this figure, we see that increasing temperature is correlated with increasing proportion of the protein exhibiting no secondary structure. However, it appears that the amount of helical structure appears to level off at approximately 15.7% of the total structure, whereas the degree beta sheet structure appears to continue to reduce with increasing temperature.

4.3 Discussion

Our work extends the previous study of Jónsson and colleagues who analyzed bimodal distributions of WT AS via multi-canonical Wang-Landau simulations, determining the ensemble-averaged secondary and tertiary structural properties of each mode of the distribution [204]. Here we focus instead on the simulation of temperature-dependent properties of AS and its familial point mutations (A30P, E46K, and A53T), as well as covalent dimers of AS. Reassuringly, our simulations exhibited comparable peak energy levels to those of Jónsson *et al.* for WT AS, as seen in the 320K simulation of Figure 4.1 [204], although there was a roughly 9 K difference in the apparent transition temperature, likely due to differences in the move frequency used in the simulations. Our simulation

of WT AS at a simulation temperature of 340K also agreed well with solution NMR experiments at 303K, providing a useful point of comparison to experimental results. Note that variation between the experimental and simulated temperatures are due to Protein Monte Carlo being weakly coupled to temperature.

Based on our simulations, we found significant structural variation between WT and mutant AS monomers. The E46K and A53T mutants both exhibited an increased propensity to form extended sheet structure when compared with either the WT or A30P mutant, most notably at the simulation temperature of 330K. This variation was found primarily within the portion of the protein that makes up the core structure of α -synuclein fibrils, from approximately residues 30-110 [8, 211, 213-216]. To test whether any of these structural variations might be correlated with changes in the aggregation propensities of the AS variants *in vitro*, we looked for regions in the primary sequence where the structural trends matched the trend in the aggregation rates, with E46K aggregation fastest, then A53T, then A30P and WT AS (where A30P and WT are taken to be similar, given the uncertainty in the experimental aggregation results for these variants) [211, 216-222].

Looking first at the propensity to form extended sheet structure, at 340K the results are generally similar for all variants, but at residues 24-25, 36-43, 46-51, 53-55, 59-60, 69, and 131-136 the propensity to form extended sheets does indeed match the order of aggregation propensity (E46K > A53T > A30P ~ WT). This pattern of extended-sheet propensities is even more prominent at 330K occurring over a wider range of residues, from 24-25, 38-43, 50-55, 60-64, 66, 69, 73, 110-119, 127, and 130-136. From these results, we see that this propensity is exhibited in both simulations at residues 24-25, 38-43, 50, 51, 53-55, 60-69, and 131-136. We note that previous work showed that the aggregation rates of α -synuclein mutants were linearly related to the propensity of the structure to form extended sheets, with residues 50-54 showing a marked increase in aggregation rate with increased sheet content [223], and residues 38-40, 50-54, 65-75, and 87-92 have been predicted to be some of the most

aggregation-prone regions of the AS sequence [224]. Our results reinforce the idea that the increase in extended β -sheet propensity is the underlying cause of increasing aggregation propensity. Regions showing consistent relative variation in β -sheet content are most likely due to the presence of mutations between monomers, whereas those that are not consistent may be more likely due to random variations. This suggests that the effect of the mutations is to increase the structural propensity of aggregation prone regions to form β -sheet structures, and that this may be the underlying driving force causing increased fibril formation. Below the phase transition temperature, the monomers exhibit less consistent variations in extended sheet content, with primary variations exhibited about the mutation sites and neighbours. However, E46K and A53T do exhibit increased local and long range correlations when compared to A30P and WT, which is indicative of structure formation. Propensity for short β -strands appears to converge quickly above the transition temperature; this is likely due to increased likelihood of random short chain fluctuations dominating the ensemble.

Turning next to the propensity to form helices, we found only minor variations in α -helical propensity, whether for extended or short helices, with the exception of the A30P mutant. The variations in helix content between mutants did not consistently match the effects of the mutations on aggregation rates in any region of the sequence, suggesting that the effects of the mutations on helix content are not significant for aggregation. The effects of the A30P mutation were most important in the region 17–39 at temperatures above the transition, consistent with observations linking depressed α -helical formation in residues 6-25 with reduced membrane binding [225] and the reduction in membrane/vesicle binding observed in A30P [226-228].

We next performed a similar set of simulations on the larger AS dimer formed through the covalent linkage of two monomers in a head-to-tail fashion via a GSG linker. The resulting simulations also showed an apparent multi-modal distribution of energies between simulation temperatures of 300K and 330K. We then approached comparing our

results to those of experiments in literature. We find that based on this comparison, our simulations agree very well with the limited amount of experiments performed on such a covalent arrangement. At each temperature, the attached monomers behave virtually independent of one another. This is consistent with the experimental evidence which showed little difference between monomers, but a reduction in NMR signal nearer the linkage region. This appears to be consistent over the entire length of the protein, with secondary structural propensities for each monomer being virtually identical to the monomer simulations. We also found that the total content of secondary structure within our distributions was consistent with that reported in literature for covalently linked dimers, resulting in linked unstructured monomers. However, due to the scarcity of information concerning these dimers, little more comparison able to be made.

In conclusion, our simulations support the notion that the ability of the monomers to form extended β -sheet structures in the fibril core region may be intimately related to the rates of aggregation. This might suggest a mechanism where aggregates form via the interaction of long chains, so that the ability of the protein to form these structures upon interaction is important in fibril growth. Our work also lends support to a drug design strategy targeting those regions of the protein exhibiting increased β -strand content proportional to aggregation tendency. Targeting these regions for small-molecule binding and aggregation-inhibition could reduce the propensity of these regions to form β -strand secondary structures as a likely method of reducing the effects of AS aggregation in disease. Through our dimer simulations, we were also able to produce ensembles which agree with the current state of literature, and from this we will be using these structures in MC pulling simulations reported in Chapter 5.

5 MONTE CARLO PULLING SIMULATIONS

5.1 Outline

In this chapter, we extend the discussion on structure formation in α -synuclein monomers and dimers presented in Chapter 4. This is accomplished through the performance of Monte Carlo pulling simulations on a sub-sample of each of the multimodal peaks described. These results were then compared against experimental optical tweezers and AFM pulling experiments to determine the validity of the results. Suggestions for improvements, and corroboration with experiments are then provided where possible.

5.2 Methods

The MC pulling simulations are performed as an extension of the ProFASI force field,

$$E_{pulling} = E_{ProF} - \frac{k_{Spring}}{2} (L_0 + vt - L(x))^2 \quad 5.1$$

where $E_{Pulling}$ is the energy of the simulation, E_{ProF} is the base forcefield for the ProFASI package, k_{spring} is the spring constant for the applied force, L_0 is the end-to-end distance of the protein, $L(x)$ is the position of the probes, v is the pulling velocity, and t is the monte carlo time. Initially, the probe position is set at the position of the terminal residues, and then proceeds to move away from this position at the state velocity v . The move/update probability for these simulations are 50% BGS, and 50% pivots, eliminating side chain rotations. Each simulation is run using the Metropolis Hastings criterion, with the probability of acceptance being

$$P = \min\left(1, \exp\left(\frac{-(E_2 - E_1)}{k_B T}\right)\right) \quad 5.2$$

where E_2 and E_1 are the new and old energy's respectively. The temperature of the simulations was set to the approximate temperature at which each phase of our multimodal peaks should be present. This temperature for our monomers was set at 329K, which was determined via Wang Landau simulations performed by *Jónsson et al.* [204]. The temperature for the simulation of the covalently linked dimers, which was estimated from the peak distributions of the Chapter 4, was set at 315K. 200 structures were then randomly extracted from each phase expressed by our previous MC simulations, for a total of 400 monomer structures and 600 dimer structures.

Simulations on monomeric AS were first performed using parameter values determined by *Jónsson et al.* [229]. In their simulations, the velocity corresponding best to results obtained via AFM pulling experiments [230] was 0.05 fm/move, and spring constant set at 37 pN/nm. Once completed, we then turned our attention to performing simulations with values expected to be consistent with optical tweezer measurements performed by *Neupane et al.* [231]. The optical tweezer measurements were performed using a similar velocity, but a 100 fold smaller spring constant, thus requiring a reduction in our spring constant to 0.37 pN/nm. This resulted in the simulations requiring a threefold longer simulation time. Upon analysing the subsequent results, we then turned our attention

to the simulation of dimeric AS, where our new simulation parameter values resulted in simulation times of approximately 3 months, making performance of larger covalently linked monomeric structures presently unfeasible.

5.3 Analysis

In analysing our simulations, we first determined the secondary structure content of the samples extracted from each phase, including presence of turns, and beta-sheet content of each conformer. This was determined through the use of the AMBER analysis package, cpptraj [195].

The output of our simulations exhibited a significant degree of apparent noise, so the results were smoothed over a time averaged window of 1000 MC sweeps, representing 620 000 MC steps, using the Octave Math Package [196]. This allowed for the easy visual identification of rupture events. These rupture events are recognizable through a sharp drop in applied force, with a concomitant increase in end-to-end distance.

The conformational length at which these rupture events occurred was determined through a fitting script written in house for Octave [196]. The extensible worm-like chain model (EWLC) was used to fit the resulting data, and parameters described therein. The EWLC model, as interpolated by Marko and Siggia is described by the following equation,

$$F = \frac{k_{\beta}T}{L_p} \left(\frac{1}{4} \left(1 - \frac{x}{L_c} + \frac{F}{K} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c} - \frac{F}{K} \right), \quad 5.3$$

where F is the applied force, L_c and L_p are the contour length and persistence length respectively, x is the extension length of the protein, and K is the proteins stretch modulus [232]. Initially, three-parameters were used in the fitting of each curve, $k_{\beta}T/L_p$, K , and L_c for the unfolded proteins. This fitting allowed us to get an approximate value for the ratio of $k_{\beta}T/L_p$, in order to reduce the number of fitting parameters, that being ~ 6.25 pN. L_c and K were then allowed to vary for each curve.

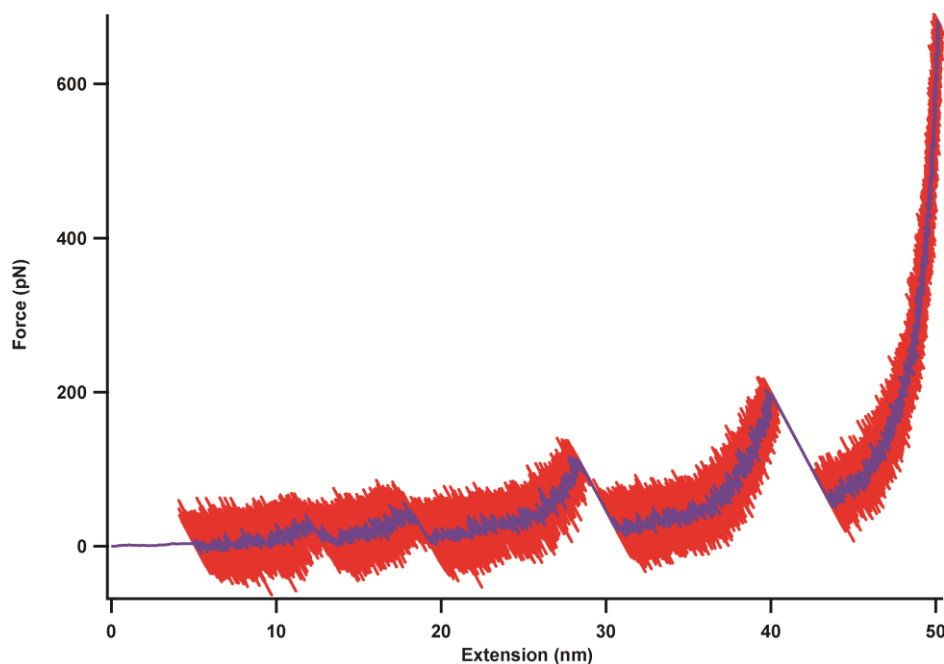


Figure 5.1: Plot of Ext v. Force for Cys-Terminated AS Monomers. Plot in red represents raw data, whereas blue is the result of subsequent smoothing of the raw noisy data.

5.4 Results

5.4.1 Monomers

Monomer simulations were performed at the previously reported MC phase transition temperature of 329K [204]. Our primary concern in this analysis, is the comparison of contour length change between simulation parameter sets, as well as comparison to experiment. Due to difficulties in comparing MC temperatures to physical/experimental temperatures, necessity of higher speeds/force constants due to time constraints, the forces are unlikely to be accurate. For all simulation parameters, the high energy peak exhibited no apparent structure. This is different from the simulations performed by *Jónsson et al.*, and is likely due to the lower number of monomers simulated. The lower energy peak, however, exhibited a substantial degree of secondary and tertiary structure. For this reason, we focused our analysis on the lower energy distribution.

Peak	v	k	Events
1	0.05	37	635
2	0.05	37	0

Table 5-1: Variation of the number of rupture events with changes in spring constant (k). Where v is in fm/step, and k is in pN/nm.

In analysing the results of each set of pulling simulations, we determined the number of apparent structural transitions, contour length changes, and distribution of forces. From these results, shown in Figure 5.2, we see that a most predominant contour length changes occur about 10 nm, with an apparently slowly decreasing tail segment, with slight increases at just below 20 nm. This is consistent with previous experimental measurements by *Neupane et al.* [231]. However, our simulations exhibit a substantial variation in measured transition force. Whereas *Neupane et al.* measured typical forces in the 5-15 pN range, our simulations exhibited forces greater than 500 pN. This is likely due to several issues. First, the values for pulling speed were empirically chosen by *Jonsson et al.* to match AFM data, which it does adequately. Also, the MC temperature at which these simulations were performed represents the approximate transition temperature of AS of -15°C , which may also play a role.

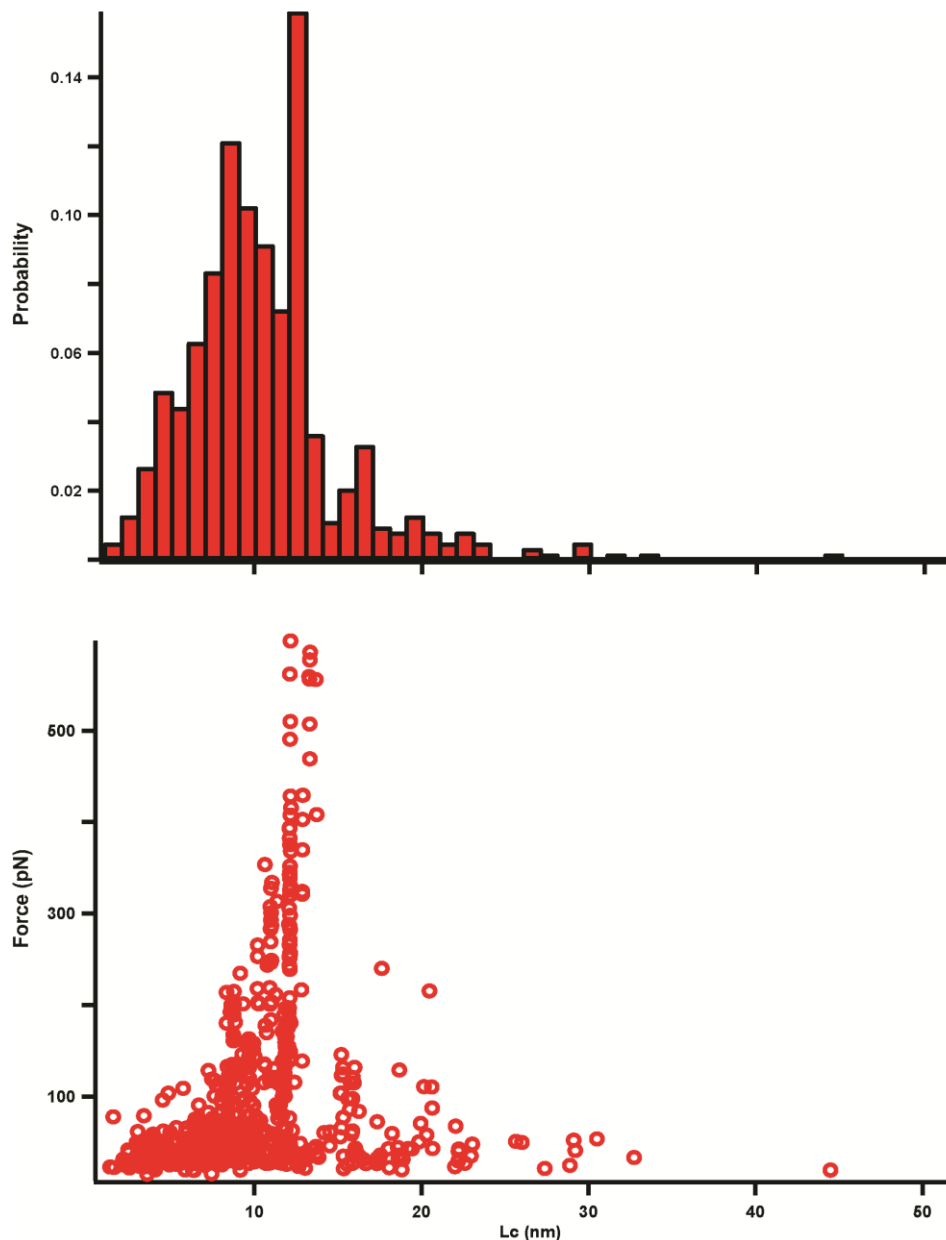


Figure 5.2: Plot of histograms of distribution of delta Lc values for AS monomers, as well as distribution of force relative to Lc change.

Given that the previous simulation parameters were empirically chosen to represent AFM data, a second set of runs using the same initial structures was chosen at values that were hoped to correspond with *Neupane et al.*'s optical tweezer experiments. In these experiments, typical spring constants were on the order of 100 fold smaller than their AFM counterparts, with similar order of magnitude pulling speeds. This led us to simply reduce the spring constant to 0.37 pN/nm, and maintain

the previous pulling speed. However, this greatly extended the simulation time to unpractical computationally times for larger proteins.

5.4.2 Dimers

In performing pulling simulations on covalently linked dimers of AS, we again extracted 200 structures from each of the three apparent peaks in our temperature varying simulations from Chapter 4. Simulation temperature was performed at the estimated three phase temperature of 315K. Simulation speed was maintained at 0.05 fm/step, and the spring constant at 37 pN/nm.

Each peak exhibits an increasing mean energy with increasing peak number, and reduction in the number of apparent rips from the sample of structures, with the number of events shown in Table 5-2.

Peak	No. of Events
1	1077
2	518
3	97

Table 5-2: Comparison of number of transition events for each dimer peak.

Peak 1 exhibits the highest number of transition events as might be expected due to its more compact, low energy arrangement. From

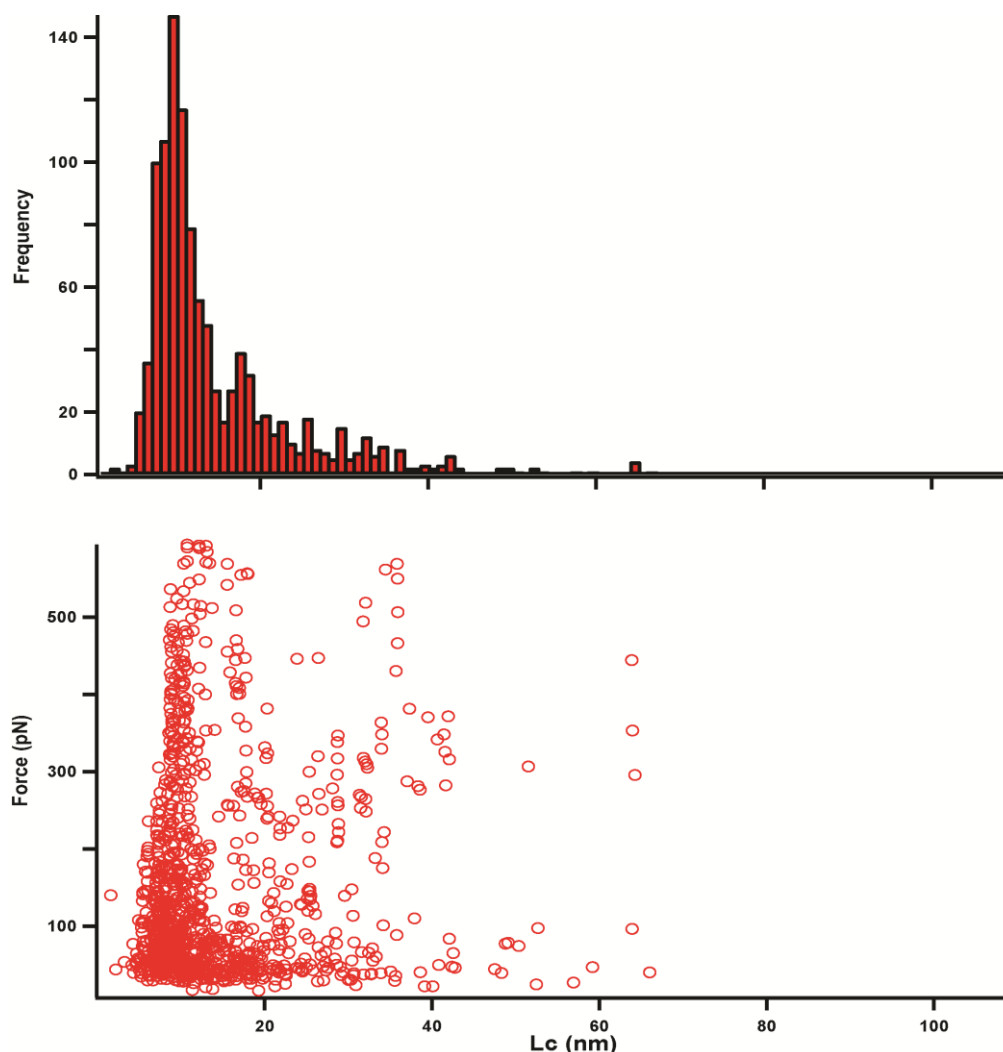


Figure 5.3: Distribution of L_c , and Forces for the lowest energy peak (Peak 1) in the dimeric ensemble.

analysing the changes in L_c , we see that this distribution exhibits a substantial peak about 10 nm, with a short tail decreasing with increasing ΔL_c as shown in Figure 5.3. We can also see that the high force regimes exists over much of the range of ΔL_c values. The Peak 2 distribution however exhibits a much longer tail, nearing the total length of the dimer, as shown in Figure 5.4. This distribution exhibits a long flat tail of L_c changes, consistent with the results of *Krishna et al* [231], and suggesting more extensive monomer-monomer interactions..

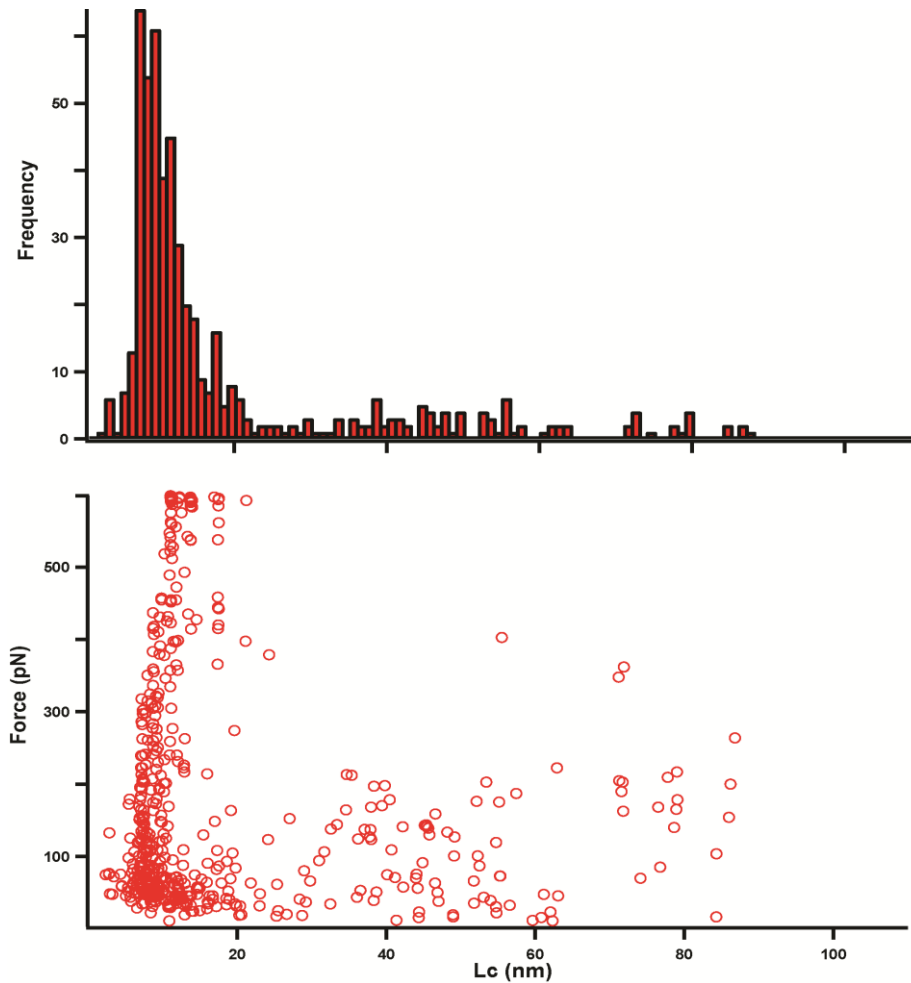


Figure 5.4: Distribution of L_c and Forces for Peak 2 of the AS Dimer ensembles.

Analysis was also performed on the Peak-3 distribution, which exhibited further extension of the tail region, and flattening of the large peak about 10 nm.

Much of the result of the analysis of structural transitions in dimeric AS appear to be due to intrachain interaction within each monomer, instead of interaction between monomers.

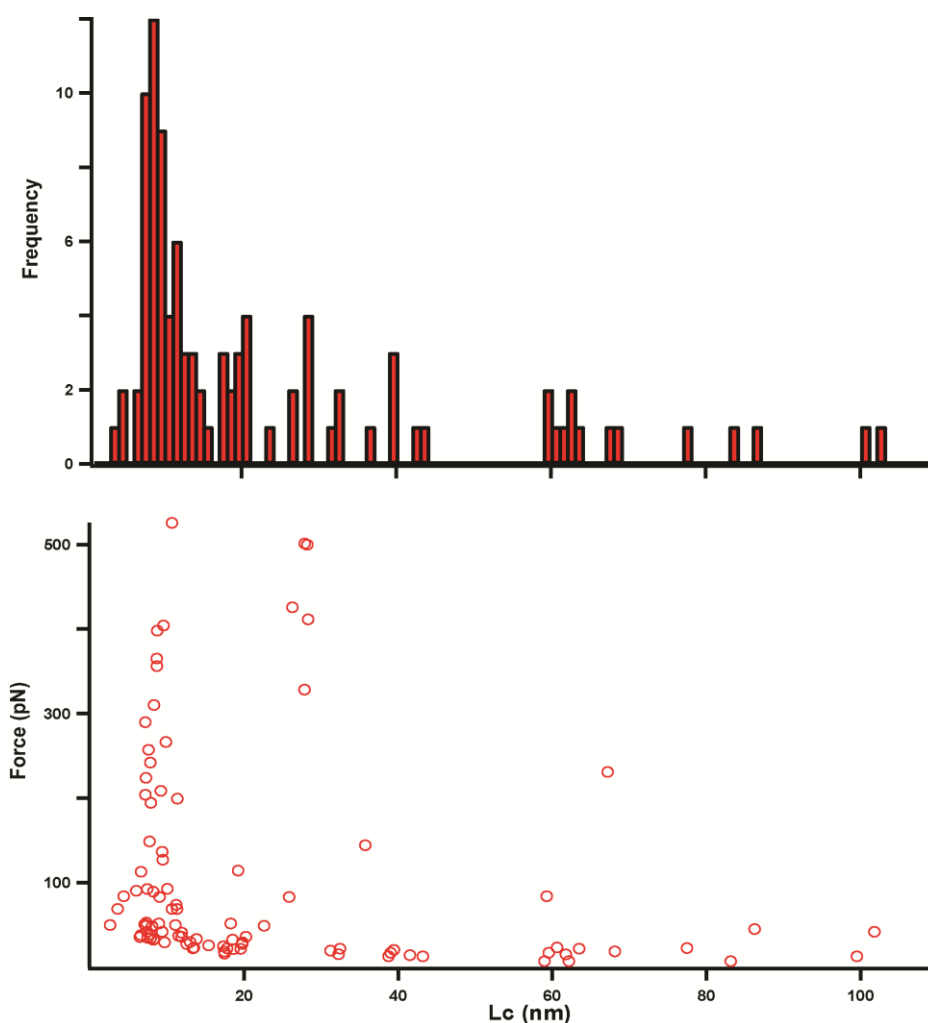


Figure 5.5: Distribution of L_c and Forces for the Peak 3 ensemble of Dimeric AS.

5.5 Discussion

In this chapter, we have presented methodology and results of pulling simulations using a modified MC force-field. The resulting monomer simulations were capable of approximating much of the results of *Neupane et al.*'s pulling simulations, and were thus implemented on dimeric AS. The resulting ΔL_c distributions for each peak showed a significant range of, from less than 5 nm to approximately 50 nm. These results exhibit significant peaks around ΔL_c values of 9nm, 13nm, 17nm, 19nm, and 30

nm. When we view the results of pulling experiments of Cys-terminated AS performed *Neupane et al*, we see that there appear to be no peaks below 10 nm. However, we do view peaks at values slightly greater than 10nm, and peaks at approximately 15nm, 18nm, 21nm, and 31nm. Given that the separation of apparent peaks appears to be consistent, it is likely that in this case variation of ΔL_c between simulation and experiment is due to assumptions made in the model. This includes the absence of DNA handles, as well as the bond lengths in the model being static.

Our pulling simulations of dimeric AS exhibit a further increase in the range of measured values, from approximately 2nm to greater than 100nm. Many of the resulting peaks appear to correspond to intrachain (or within monomer) interactions. However, we do see some values greater than the absolute length of AS monomers which would indicate interaction between monomers. When we view the results of the dimer pulling experiments, we see a number of apparent peaks at approximately 15nm, 24nm, 35nm, 50nm, 55nm, 60nm, 74nm, and 80nm. In our simulations, a number of these peaks are exhibited in our simulations as well, however this is also mixed with a number of intrachain monomer interactions.

Currently, attempts are being undertaken to more quantitatively compare the results of experiment with those found in simulation. Given the probabilistic nature of the simulation methods, a possible solution is to perform simulations consisting of multiple replicates for each structure. This will reduce the role of less stable intermediate structures and alter the distribution landscape. A second potential solution is to develop filtering methods to reduce our data to a sample that better approximates experiment, and subsequently analysing the resulting structures.

6 FRAGMENT PROBE ANALYSIS AND DRUG SCREENING

6.1 Outline

In this chapter, we look to investigate the possibility of using chemical-protein docking tools to investigate potential drug binding to α -synuclein. The goal is to discover such drugs that may provide therapeutic benefit to the patients suffering from disorders of Lewy Body (LB) formation such as PD or Lewy Body Dementia (LBD).

6.2 Replica Exchange Monte Carlo

Replica Exchange MC (REMC) was the chosen method to appropriately sample the space of extended and collapsed formations of α -synuclein. As discussed in Section 3.2.3, REMC (or parallel tempering) provides a means of enhancing the sampling of the proteins conformational space by performing the simulations at multiple

temperatures, above and below any potential phase transition. In this case, temperature bounds were chosen based on the formation of singular peaks of high and low energy ensembles of WT monomeric α -synuclein in Section 4.1. Thus, we set up 32 parallel REMC runs in the simulation temperature range of 300K to 340K, with temperature exchange between simulations attempted after each MC move. A significant overlap between energies at neighbouring temperatures would be expected at such small temperature separations. Once completed, a histogram of the energies of the simulation was compiled, which exhibited two peaks of high and low energy as expected, as can be seen in Figure 6.1.

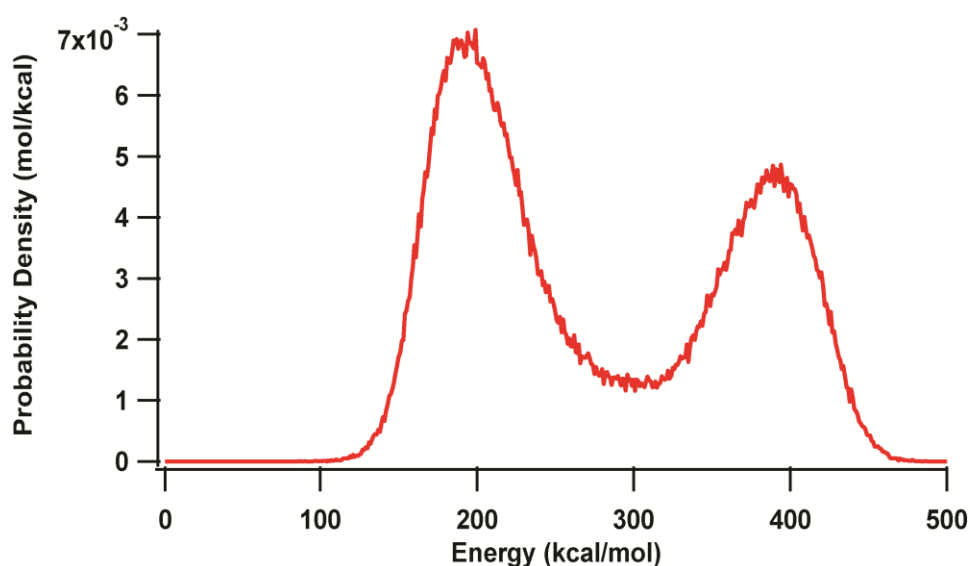


Figure 6.1: Distribution of energies for WT REMC simulations of AS.

These two peaks were then separated into high and low energy phases. These two peaks were decomposed by separating the structures with energy greater than and less than the minimum value experienced in the dip between the two peaks. The R_g values, as well as the per residue secondary structure values were determined through the use of the AMBER analysis package, cpptraj [195]. The result of these portions of analysis are consistent with the values determined by both myself in Chapter 4, and *Jónsson et al.* [204].

6.3 Binding Sites

6.3.1 Clustering

The structure of the potential drug binding region is necessary in docking calculations/simulations. To determine candidate structures for the remaining portion of our results, we performed density clustering of the low energy phase structures from our REMC simulations.

In order to perform clustering of our simulations, we opted to choose the density based clustering algorithm (DBSCAN) [233]. DBSCAN is a clustering method that produces approximate clusters by allocating those points that are near in spatial distance to a single cluster. Points that are outside these regions, in low-density areas, are marked as outliers. DBSCAN also benefits in the context of MC simulations due to its ability to discriminate rare events from any cluster, thereby potentially eliminating spurious structures produced during the course of the simulation.

Due to the inherent lack of order exhibited by AS for much of its length, clustering was performed on the region associated with the fibril core of AS. This is approximately from residues 35 to 95 [8]. This resulted in greater than 60 total clusters, with the top 20 clusters representing over 70% of the total population of structures. The core region was then compared to the short range contact map of AS fibrils produced by *Vilar et al.* [8]. For convenience, only the top 9 clusters are shown in Figure 6.2.

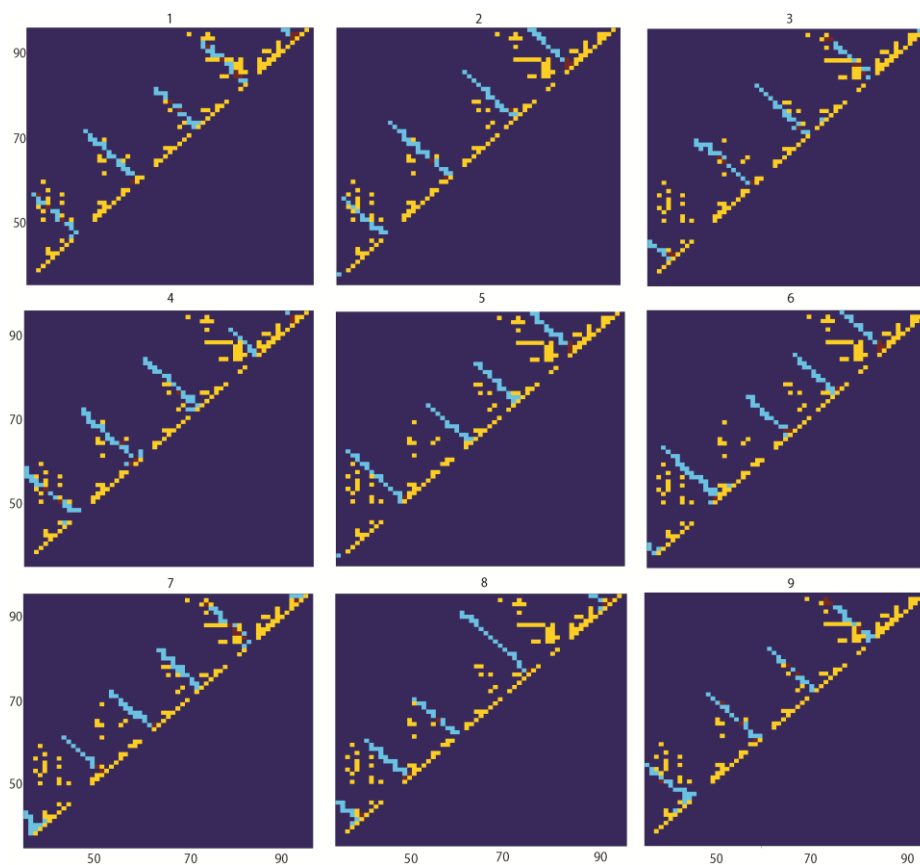


Figure 6.2: Plot of contact map for residues within 6 nm of one another for clusters 1-9 in light blue. The contact map produced by *Vilar et al.* is shown in yellow, with direct overlap shown in red.

From this plot we can see that the work of *Vilar et al.* exhibits more apparent flexibility, judging by the spread of contacting residues. One must remember that this is the result of an ensemble measurement when comparing the two results, where the structures produced computationally are static. From the resulting plots, we also see apparent significant overlap of the two distance plots. This is especially true for

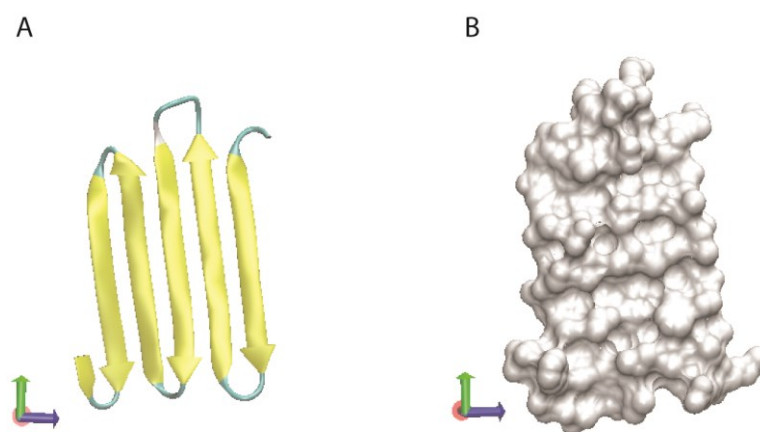


Figure 6.3: Representative structures for cluster 1. A. Secondary structure representation. Yellow arrows represent beta-sheet structures. B. Electronic surface area representation.

clusters 1, 2, 3, and 9. The most striking variation is the existence of an apparent turn at approximately residue 69 (GLY) in the work of *Vilar et al.*, whereas clustering results exhibit a turn at residue 72 (GLY). Each of these corresponds to a flexible glycine residue, which are capable of producing such a turn. This may also explain the apparent spread about the turn as well. This representative core region is shown in Figure 6.3.

6.3.2 Fragment Probe Analysis

In this section, we focus on fragment probe analysis, which was performed on the top twenty cluster representatives from Section 6.3.1. Each simulation was set up using a 10 Å border, with periodic conditions, using TIP3P waters. 1 in 20 water molecules were replaced by probes of

isopropanol (70%), acetamide (10%), acetate (10%), and isopropylamine (10%). Each simulation was first minimized for 2000 steps. The process of equilibration was begun by first heating the system to 300 K over 2 ps, and running at 300 K for 80 ps, keeping the protein constrained. The second stage involved heating the system to 600 K over 2 ps and holding it the temperature steady at 600 K for 600 ps. The temperature was then lowered to 300 K over another 2 ps with the position of the protein held stationary. This step ensures mixing of probe molecules and water molecules about the protein. The constraints are then released, and the protein is allowed to equilibrate for 600 ps. The production simulations were then run for a total of 40 ns each. Each simulation was then analysed to determine regions about the protein in which druggable sites may exist, the results of the first cluster simulations are shown in Table 6-1, with the complete table in Appendix 2.

Cluster	Site	Site Res.	Binding Aff. (nM)
1	1	21,28,29,3 0,40,41,42, 43,44,50,5 1,52,53,54, 66,67,68,6 9	16.93 - 22.86
	2	29,30,37,3 8,39,41,42, 43,51,55,5 7	488 - 630
	3	63,65,82,1 37	1.41 - 1.50

Table 6-1: Results of druggability assessment using fragment probe analysis on clusters of alpha-synuclein.

Binding for cluster 1 core region is shown in Figure 6.4.

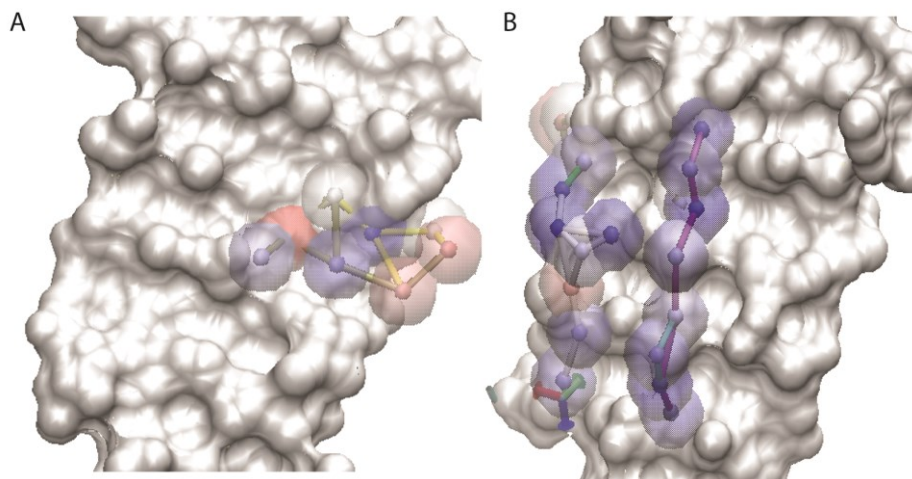


Figure 6.4: Binding sites for AS cluster 1 representative. A. Binding site 1 B. Binding sites 2 and 3.

From this figure, we can see that site 1 appears to have a much more condensed grouping, whereas sites 2 and 3 appear to exist in a chain-like fashion, which may also provide more information to the potential structures of ligands.

In analysing potential binding sites/regions, we also looked to determine which residues occurred with the highest frequencies. This was performed by grouping all residues in binding sites, and plotting a histogram of their occurrence, as shown in Figure 6.5.

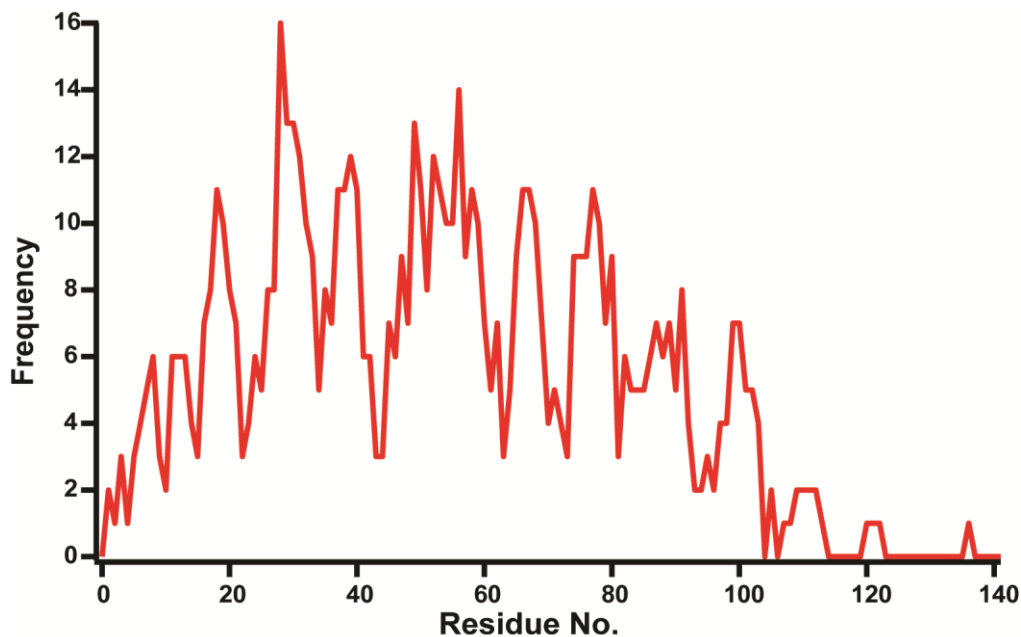


Figure 6.5: Frequency of residues in binding sites by druggability analysis.

From this, we see several regions that occur with higher relative frequency than others. The most frequently occurring residue is residue 28. However, the most frequently occurring region appears to be residues 49 to 59 with a mean per-residue frequency of 10.81 ± 1.72 . Other regions exhibiting increased occurrence (frequency greater than 7, or 50% of the max frequency or greater), are residues 26-41, 65-68, and 74-78. These regions exhibit a per-residue mean frequency of 10.27 ± 4.47 , 10.25 ± 0.95 , and 9.6 ± 0.89 respectively. This suggests drugs or peptides that target residues 49-59 may have the best chance of interrupting protein aggregation.

The range of binding energies determined via this in silico assay is 7 pM to 1700 μ M, representing a very broad range of potential affinities and energies. The maximal drug-like binding affinities are surprisingly strong, and likely have to be taken in context of the protein in question. In traditional, more stable structures, the majority of the structure would likely be stable. α -Synuclein, being a very dynamic protein, may lose some of this apparent structure in much longer simulations, which could

reduce the apparent binding affinities. In assessing druggability, we are looking for segments which exhibit increased binding frequency as potential targets for drugs. Secondary to these binding affinities, we can also determine the relative stability of each structure, as well as the stability of the core region of each.

6.4 Virtual Screening and Docking

From the previous section, we may note that there exists a variety of potential sites in the fibril core region for the binding of ligands. Optimally, if one were to design a pharmacophore to bind this area of the protein, it might be expected that the designer would focus primarily on those regions with high likelihood of binding. However, in the case of our screening, we have chosen to adopt a blind docking approach to assess docking in this region. This is in part due to the variety of potential binding solutions that may exist over the core region, and the apparent variety of structures that may exist.

In this section, we have chosen to adopt a multi-modal approach to determine potential ligands likely capable of reducing AS aggregation. This method is based on analysis performed by several researchers. Blind docking approaches have successfully identified possible aggregation inhibitors based on solid state 3D NMR models of amyloid- β ($A\beta$) fibrils [234]. In this work, the authors used a blind docking approach to bind beta-sheet breaker peptides to fibrillar core representatives, and used the resulting scores to determine the best selection of sheet breakers to test experimentally. Other work on identifying possible anti-aggregation drugs used PRE distance constrained MD simulations of AS in solution [235]. The resulting structures were then clustered, and cluster representatives were used to determine binding regions, and have ligand binding activity. This previous work was unable to determine aggregation inhibitors, but did find drugs capable of binding AS and interfering with membrane interaction.

Here we have employed the underlying methods of both these works in order to determine ligands that may have the potential to retard or halt aggregation of AS. We have taken the clusters of the previous section as possible representatives of fibril core of AS, and performed blind docking on these regions to determine the ligand binding.

Blind docking/Virtual screening was performed using the Autodock software package, with AutoDockTools [191]. Traditionally, drugs chosen for docking and testing satisfy the requirements proposed by *Lipinsky et al.*, the so called Rule of 5, in order to determine the potential viability of a drug [236]. However, recent research into the blood brain barrier (BBB) has shown the ability to increase its permeability through disruption via focused ultrasound or through novel BBB targets [237-239]. Other research has shown that neuroinflammation associated with degenerative diseases may alter the permeability of the BBB to drugs. For these reasons, we have chosen to not reduce the space of drugs by molecular descriptors, but to employ a diversity search.

6.4.1 NCI-Diversity Set

The ligands chosen for this work were taken from the NCI (National Cancer Institute) Diversity Set, which is a structurally diverse set of 1880 pharmacophores that are representative of the NCI's library of 140,000 compounds. This set of ligands was downloaded from the ZINC depository [240]. Simulation grids were maximized around the cluster representatives. Initial simulations were set to run for 2,500,000 evaluations, with a population of 250 ligands, and 40 simulation runs. The results of simulations for each structure were then ranked according to lowest energy structure in the largest cluster of ligands. This approach allows us to find the most likely binding energy for the ligand. Several methods for determining which ligands may exhibit the highest chance of binding the fibril core region were attempted. These include determining the lowest mean binding energy over the set of possible ligands, as well as determining those structures which bind with binding energies below a specified energy cut-off.

Following the analysis of the binding energies, we find that there exists two distinct binding energy distributions for drugs within the NCI Diversity Set, as shown in Figure 6.6.

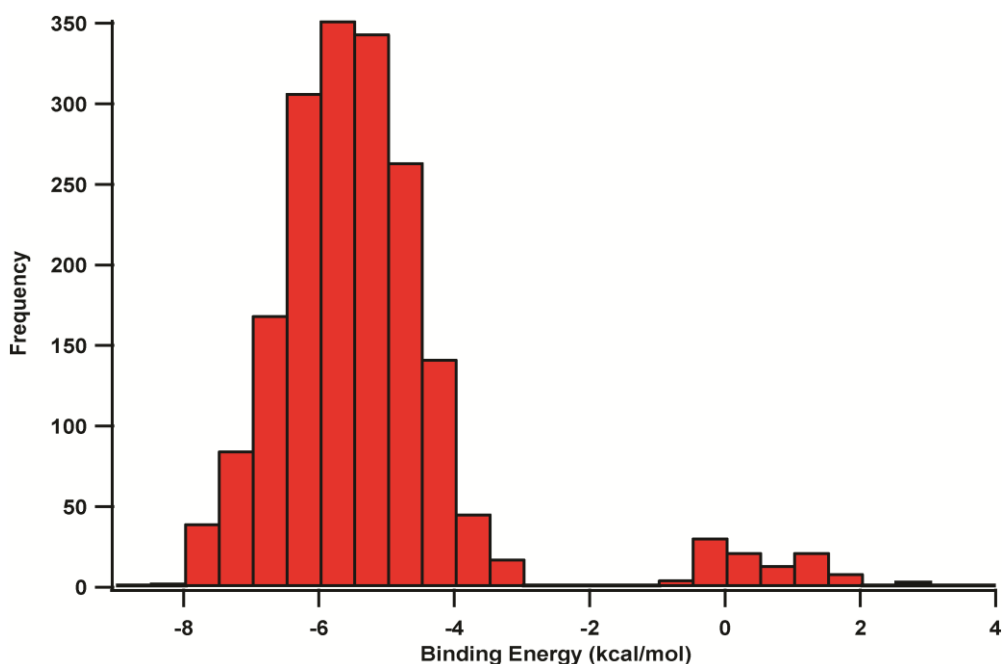


Figure 6.6: Distribution of the mean maximum binding energy's of drugs in the NCI-Diversity Set.

From this, we can see that a large segment of the population of drugs likely bind the condensed beta-sheet structures present in our simulations. However, the vast majority of these drugs would likely not interact to a high enough degree to be a relevant candidate. However, we do see a number of drugs which bind, to within error, of less than -8.0 kcal/mol. This represents a binding affinity of less than 1.35 μM . The top 3 drugs with the lowest binding energy are shown in Table 6-2, with the top 21 shown in Appendix 3.

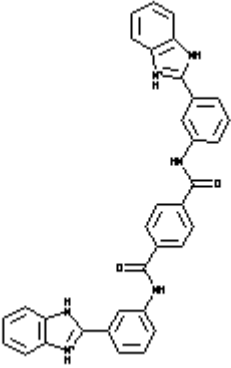
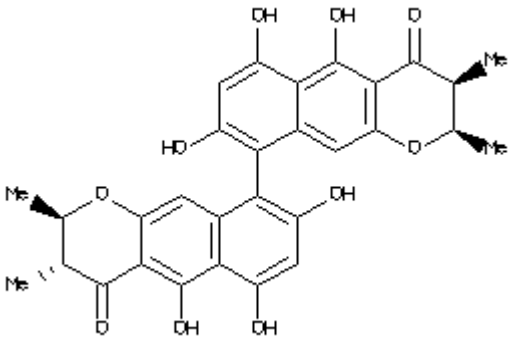
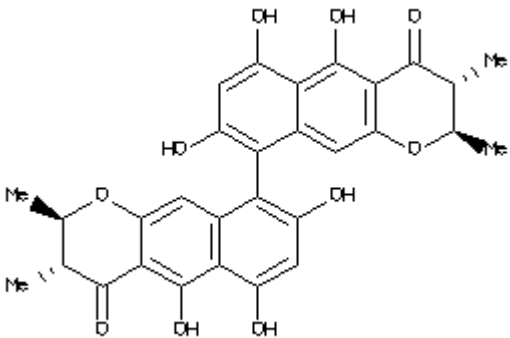
Structure (ZINC ID/ Name)	Binding Energy (kcal/mol)
 <p>ZINC ID: 1690699</p>	-8.6 ± 0.8
 <p>ZINC ID: 17465979</p>	-8.5 ± 0.8
 <p>ZINC ID: 17465983</p>	-8.5 ± 0.8

Table 6-2: Top 3 drugs by lowest mean binding energy from blind docking of the NCI-Diversity set against AS cluster representatives.

6.5 Discussion

In using REMD, we have produced a multimodal distribution of AS structures similar to those produced using Wang-Landau simulations. Our primary concern in these simulations was to approximate the stacked beta sheet structure believed to exist in the AS fibril core. The work of *Vilar et al* has provided us with experimental measures to compare the results of our simulations against [8]. From clustering the resulting low energy peak, we find a number of the core AS region to be in good agreement with their results. Most notably, the representative for cluster 1 matches very well. Variation between cluster representatives and experimental results is likely due to the fact that cluster representatives are static, whereas the experimental results are the ensemble mean of a dynamic structure. Recently, a fibril structure of AS was produced using restraints derived from NMR in various experimental setups. This solved structure exhibited a Greek key topology, versus the previously held belief of the core being a stacked anti-parallel beta sheet arrangement. Given the highly dynamic nature of AS, it's likely that this is simply one of the potential structural arrangements AS may exhibit.

In order to combat this inherent variability, the top 20 cluster representatives are used to determine regions of potential druggability. The results of these druggability simulations produce several regions with highly likelihood of being druggable. These include residues 49 to 59, 65-68 and 74-79. The results of these simulations provide us with regions of contiguous druggability which can be used in the production of drugs via rational drug design strategies. This may include the production of small-peptide inhibitors, such as those being researched in the treatment of A-Beta aggregation in Alzheimer's disease [241].

From the results of the previous druggability simulations and clustering, we subsequently applied blind-docking to the top 20 cluster representatives. This blind docking was undertaken using the NCI-Diversity set, which represents the structural diversity present in the NCI database of potential drug candidates. From these simulations, we

determine which members of the NCI-Diversity set exhibit the lowest mean binding energies over the entire set of simulations. Within error, a number of set of drugs exhibit binding energies of -8.0 kcal/mol or less. This binding energy represents binding constants on the order of 10^{-6} mol.

Our results concerning the application of a blind-docking diversity search to each of the cluster representatives provides us with a significant breadth of information to reduce the search space of potential binding candidates. From this point, application of similarity searches using the top binding candidates will likely produced far improved candidate species.

Currently, work is being performed determine these potential candidate species, as well as determining the best candidates for experimental assessment of the affects these drugs may have on the behaviour of AS in solution.

7 SUMMARY AND FUTURE WORK

7.1 Summary

In this work, we have employed a number of computational strategies to study the physical traits of AS that likely affect its ability to aggregate into higher order structures, as well potential methods of determining drugs that may inhibit such aggregation.

In Chapter 4, we employed the use of protein Monte Carlo simulations to effectively sample the conformational space of AS (and its familial mutants) over a broad range of temperatures. From these simulations, we noted that each of the monomers exhibit bimodal behaviour at the lower range of simulation temperatures. This suggested varying transition temperatures for each of the types of AS. In this work, we were also able to determine regions of the protein which exhibit increased propensity to form β -sheet structures which is proportional to the species propensity for aggregation, suggesting a potential role in the aggregation process. In this chapter, we also investigated the ability of Monte Carlo simulations to produce structural ensembles of covalently

linked monomers of AS. These simulations exhibit properties consistent with available experimental evidence, including propensity for secondary structural formation, as well as independence of each monomers in the linked structure.

In Chapter 5, we investigated those distributions of structures found in Chapter 4 through the use of a modified Monte Carlo force-field. This modified force-field approximated the experimental apparatus of pulling simulations through the use of the additional spring-term. These simulations were performed on both monomeric WT-AS, as well as dimeric WT-AS. From these simulations, we found some agreement of our findings with that of experimental results, with qualitative agreement of distributions of ΔL_c values.

Finally, in Chapter 6, we performed cluster analysis of the lower energy conformations of WT-AS from Chapter 4 and found that the core region agreed well with existing experimental data on AS fibrils. From this, we performed druggability simulations on the core region of these cluster representatives to determine its ability to bind potential drug candidates. We found a number of segments in this region that exhibited propensity for drug binding and thus initiated blind docking of the NCI-Diversity set against the fibril core region. Binding of drugs to this core region may allow blocking of potential aggregation prone regions from interaction with other monomers, or possible destabilize intra-chain contacts. We found a number of potential candidate drugs to be used in the future experimental or computational efforts of developing drug treatments.

7.2 Directions for Future Work

This work leads to a number of potential projects in the future. First and foremost, improvements in ΔL_c distributions produced by Monte Carlo pulling simulations may allow for the more accurate description of interacting regions of AS. Given the probabilistic nature of Monte Carlo

simulations, those structures which are less stable may be over-represented in our single replicate simulations. This may be corrected through the use of multiple replicates of each pulling structure in simulations instead of using a single simulation for each structure. Upon completion of these simulations, any further or necessary corrections may be accomplished through filtering of data to better approximate experiment.

As a possible continuation of Chapter 6, we may employ structural similarity search of our top ligands against the NCI drug database. Subsequently, we can perform virtual screening on the resulting ligand dataset to determine drugs which may exhibit improved binding affinity, or improved likelihood of bioavailability. Another approach may be to rationally design pharmacophores capable of binding those protein regions which appear to play a role in AS aggregation as well as exhibit higher frequency of druggability.

8 REFERENCES

1. Tanner CM, Aston DA, *Epidemiology of Parkinson's disease and akinetic syndromes*. Current Opinion in Neurology, 2000. **13**: p. 427-430.
2. de Lau LML, Breteler MMB, *Epidemiology of Parkinson's disease*. The Lancet Neurology, 2006. **5**(6): p. 525-535.
3. Cooper AA, et al., *Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models*. Science, 2006. **313**(5785): p. 324-328.
4. Wersinger C, Sidhu A, *Attenuation of dopamine transporter activity by alpha-synuclein*. Neuroscience Letters, 2003. **340**(3): p. 189-192.
5. Lee FJ, Liu F, Pristupa ZB, Niznik HB, *Direct binding and functional coupling of alpha-synuclein to the dopamine transporters accelerate dopamine-induced apoptosis*. FASEB Journal, 2001. **15**(6): p. 916-926.
6. Golovko MY, Barceló-Coblijn G, Castagnet PI, Austin S, Combs CK, Murphy EJ, *The role of alpha-synuclein in brain lipid metabolism: a downstream impact on brain inflammatory response*. Molecular and Cellular Biochemistry, 2009. **326**(1-2): p. 55-66.
7. Der-Sarkissian A, Jao CC, Chen J, Langen R, *Structural Organization of α -Synuclein Fibrils Studied by Site-directed Spin Labeling*. The Journal of Biological Chemistry, 2003. **278**: p. 37530-37535.
8. Vilar M H-TC, Luhrs T, Maji SK, Riek-Loher D, Verel R, Manning G, Stahlberg H, Riek R, *The fold of α -synuclein fibrils*. Proceeding of the National Academy of Sciences of the United States of America, 2008. **105**(25): p. 8637-8642.
9. Parkinson J, *An Essay on the Shaking Palsy*. 1817, London: Whittingham and Rowland.

10. Charcot J-M, *De la paralysie agitante. In Oeuvres Complètes (t 1) Leçons sur les maladies du système nerveux*, in *Lectures on diseases of the nervous system delivered at the Salpêtrière*. 1877, A Delahaye: Paris. p. 155-188.
11. Fahn S, *The history of dopamine and levodopa in the treatment of Parkinson's disease*. *Movement Disorders*, 2008. **23**(Supplement S3): p. S497-S508.
12. Polymeropoulos MH, et al., *Mutation in the alpha-synuclein gene identified in families with Parkinson's disease*. *Science*, 1997. **276**(5321): p. 2045-2047.
13. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M, *Alpha-synuclein in Lewy bodies*. *Nature*, 1997. **388**(6645): p. 839-840.
14. Luk KC, Kehm V, Carroll J, Zhang B, O'Brien P, Trojanowski JQ, Lee VM-Y, *Pathological α -Synuclein Transmission Initiates Parkinson-like Neurodegeneration in Nontransgenic Mice*. *Science*, 2012. **16**(6109): p. 949-953.
15. Jankovic J, *Parkinson's disease: clinical features and diagnosis*. *Journal of Neurology, Neurosurgery, and Psychiatry*, 2008. **79**(4): p. 368-374.
16. Marsden CD, *Movement disorders*, in *Oxford textbook of medicine*, L.J. Weatherall DJ, Warrell DA, Editor. 1996, Oxford University Press Inc.: New York. p. 3998-4022.
17. Calne D, Stoessl AJ, *Early parkinsonism*. *Clinical Neuropharmacology*, 1986. **9 Suppl 2**(S3-S8).
18. Chou KL, *Diagnosis and Management of the Patient with Tremor*. *Medicine and Health Rhode Island*, 2004. **87**(5): p. 135-138.
19. Ondo WG, Young R, *Gait and Movement Disorders*, in *Family Practice Curriculum in Neurology*, B.D. Martin RA, Moquist DC, Editor. 2013, American Academy of Neurology.
20. Samii A, Nutt JG, Ransom BE, *Parkinson's Disease*. *Lancet*, 2004. **363**(9423): p. 1783-1793.
21. Hallet M, *Clinical neurophysiology of akinesia*. *Reviews Neurology*, 1990. **146**(1299-1306).
22. Pascuel-Leone A, Valls-Sole J, Brasil-Neto JP, Cohen LG, Hallet M, *Akinesia in Parkinson's disease. I. Shortening of simple reaction time with focal, single-pulse transcranial magnetic stimulation*. *Neurology*, 1994. **44**: p. 884-891.
23. Berardelli A, Rothwell JC, Thompson PD, Hallet M, *Pathophysiology of bradykinesia in Parkinson's disease*. *Brain*, 2001. **124**: p. 2131-2146.
24. Baker KK, Ramig LO, Luschei ES, Smith ME, *Thyroarytenoid muscle activity associate with hypophonia in Parkinson disease and aging*. *Neurology*, 1998. **51**(6): p. 1592-1598.
25. *Parkinson's Disease and Movement Disorders: Diagnosis and Treatment Guidelines for the Practicing Physician*. Totowa, New Jersey: Humana Press.

26. Perlmutter JS, *Assessment of Parkinson Disease Manifestations*. Current Protocols in Neuroscience, 2010. **49**(10.1): p. 10.1.1-10.1.14.
27. Hong M, Perlmutter JS, Earhart GM, *Recommendations for Bradykinesia Assessment in Parkinson Disease*. Journal of Neurologic Physical Therapy, 2006. **30**(4): p. 205.
28. Berardelli A, Sabra AF, Hallett M, *Physiological mechanisms of rigidity in Parkinson's disease*. Journal of Neurology, Neurosurgery, and Psychiatry, 1983. **46**: p. 45-53.
29. Mortimer JA, Webster DD. *Relationships between quantitative measures of rigidity and tremor and the electromyography response to load perturbations in unselected normal subjects and Parkinson patients*. in *Progress in Neurophysiology*. 1978. Basel: Karger.
30. Lee R, Tatton WG. *Long loop reflexes in man. Clinical applications*. in *Progress in Clinical Neurophysiology*. 1978. Basel: Karger.
31. Mortimer JA, Webster DD, *Evidence for a quantitative association between EMG stretch responses and Parkinsonism rigidity*. Brain Research, 1979. **162**: p. 169-173.
32. Chan EWY, Kearney RE, Melvill Jones G, *Tibialis anterior responses to sudden angle displacements in normal and Parkinsonism subjects*. Brain Research, 1979. **173**: p. 303-314.
33. Melnick ME, *Basal Ganglia Disorders*, in *Neurological Rehabilitation*, L.R. Umphred DA, Roller M, Burton G, Editor. 2013, Elsevier: St. Louis, Missouri.
34. Duvoisin RC, Sage J, *Parkinson's Disease: A Guide for Patient and Family*. Vol. 5. Lipincott Williams and Wilkins.
35. Kim SD, Allen NE, Ganning CG, Fung VSC, *Postural Instability in Patients with Parkinson's Disease*. CNS Drugs, 2013. **27**: p. 97-112.
36. Gibb WR, Lees AJ, *A comparison of clinical and pathological features of young- and old-onset Parkinson's disease*. Neurology, 1988. **38**(9): p. 1402-1406.
37. Adkin AL, Frank JS, Jog MS, *Fear of falling and postural control in Parkinson's disease*. Movement Disorders, 2003. **18**(5): p. 496-502.
38. Bloem BR, Hausdorff JM, Visser JE, Giladi N, *Falls and freezing of gait in Parkinson's disease: A review of two interconnected, episodic phenomena*. Movement Disorders, 2004. **19**(8): p. 871-884.
39. Nallegowda M, et al., *Role of sensory input and muscle strength in maintenance of balance, gait, and posture in Parkinson's disease: a pilot study*. American Journal of Physical Medicine and Rehabilitation, 2004. **83**(12): p. 898-908.
40. Martila RJ, Rinne UK, *Disability and progression in Parkinson's disease*. Neurology, 1977. **56**(2): p. 159-169.
41. Riley DE, Lang AE, *The spectrum of levodopa-related fluctuations in Parkinson's disease*. Neurology, 1993. **43**: p. 1459-1464.

42. Witjas T, Kaphan E, Azulay JP, *Non-motor fluctuations in Parkinson's disease*. Reviews Neurology, 2007. **163**(8-9): p. 846-850.
43. Giuffrida R, Vingerhoets FJ, Bogousslavsky J, Ghika J, *Pain in Parkinson's disease*. Reviews Neurology, 2015. **161**(4): p. 407-418.
44. Schrag A, Jahanshahi M, Quinn N, *How does Parkinson's disease affect quality of life? A comparison with quality of life in the general population*. Movement Disorders, 2000. **15**: p. 1112-1118.
45. Karlsen KH, Tandberg E, Arslan D, Larsen J, *Health related quality of life in Parkinson's disease: a prospective longitudinal study*. Journal of Neurology and Neurosurgical Psychiatry, 2000. **69**(5): p. 584-589.
46. Snider SR, Fahn S, Isgreen WP, Cote LJ, *Primary sensory symptoms in parkinsonism*. Neurology, 1976. **26**(5): p. 423-429.
47. Goldstein DS, *Orthostatic hypotension as an early finding in Parkinson's disease*. Clinical Autonomic Research, 2006. **16**: p. 46-54.
48. Magerkurth C, Schnitzer R, Braune S, *Symptoms of autonomic failure in Parkinson's disease: prevalence and impact on daily life*. Clinical Autonomic Research, 2005. **15**: p. 76-82.
49. Hamill RW, Shapiro RE, Vizzard MA, *Peripheral Autonomic Nervous System*, in *Primer on the Autonomic Nervous System*, B.I. Robertson David, Burnstock G, Low PA, Paton JFR, Editor. 2012. p. 17-26.
50. McCorry LK, *Physiology of the Autonomic Nervous System*. American Journal of Pharmaceutical Education, 2007. **71**(4): p. 78.
51. Korchounov A, Kessler KR, Yakhno NN, Damulin IV, Schipper HI, *Determinants of autonomic dysfunction in idiopathic Parkinson's disease*. Journal of Neurology, 2005. **252**(1530-1536).
52. Margerkurth C, Schnitzer R, Braune S, *Symptoms of autonomic failure in Parkinson's disease: prevalence and impact of daily life*. Clinical Autonomic Research, 2005. **15**: p. 76-82.
53. Gupta V, Lipsitz LA, *Orthostatic Hypotension in the Elderly: Diagnosis and Treatment*. The American Journal of Medicine, 2007. **120**(10): p. 841-847.
54. Senard JM, Brefel-Courbon C, Rascol O, Matastruc JL, *Orthostatic hypotension in patients with Parkinson's disease: pathophysiology and management*. Drugs and Aging, 2001. **2001**(18): p. 7.
55. Peralta C, Werner P, Holl B, Kiechl S, Willeit J, Seppi K, et al, *Parkinsonism following striatal infarcts: incidence in a prospective stroke unit cohort*. Journal of Neural Transmission, 2004. **111**(1473-1483).
56. Lees AJ, Blackburn NA, Campbell VL, *The nighttime problems of Parkinson's disease*. Clinical Neuropharmacology, 1988. **11**: p. 512-519.
57. Factor SA, McAlarny T, Sanchez-Ramos JR, Weiner WJ, *Sleep disorders and sleep effect in Parkinson's disease*. Movement Disorders, 1990. **5**: p. 280-285.

58. Oerlemans WG, De Weerd AW, *The prevalence of sleep disorders in patients with Parkinson's disease. A self-reported, community-based survey.* Sleep Medicine, 2002. **3**: p. 147-149.
59. Ondo WG, Dat Vuon K, Khan H, Atassi F, Kwak C, Jankovic J, *Daytime sleepiness and other sleep disorders in Parkinson's disease.* Neurology, 2001. **57**(8): p. 1392-1396.
60. Comella CL, *Sleep disorders in Parkinson's disease: An overview.* Movement Disorders, 2007. **22**(S17): p. S367-S373.
61. Poewe W, *Non-motor symptoms in Parkinson's disease.* European Journal of Neurology, 2008. **15**(Suppl 1): p. 14-20.
62. Rascol O, Brooks DJ, Korczyn AD, De Deyn PP, Clarke CE, Lang AE, *A five-year study of the incidence of dyskinesia in patients with early Parkinson's disease who were treated with ropinirole or levodopa. 056 Study Group.* New England Journal of Medicine, 2000. **342**: p. 1484-1491.
63. Fenelon G, Mahieux F, Huon R, Ziegler M, *Hallucinations in Parkinson's disease: prevalence, phenomenology and risk factors.* Brain, 2000. **123**(Pt 4): p. 733-745.
64. Ramaker C, Marinus J, Stiggelbout AM, Van Hilten BJ, *Systematic evaluation of rating scales for impairment and disability in Parkinson's disease.* Movement Disorders, 2002. **17**(5): p. 867-876.
65. Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease, *The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations.* Movement Disorders, 2003. **18**(7): p. 738-750.
66. Goetz CG, et al., *Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS0UPDRS): Process, format, and clinimetric testing plan.* Movement Disorders, 2007. **22**(1): p. 41-47.
67. Goetz CG, Stebbins GT, Tilley BC, *Calibration of Unified Parkinson's Disease Rating Scale Scores of Movement Disorder Society-Unified Parkinson's Disease Rating Scale Scores.* Movement Disorders, 2012. **27**(10): p. 1239-1242.
68. Giovanni P, et al., *Early-onset Parkinson's disease.* Movement Disorders, 1991. **6**(1): p. 36-42.
69. Marsden CD, *Parkinson's Disease.* Lancet, 1990. **335**(948-952).
70. Fearnley JM, Lees AJ, *Ageing and Parkinson's disease: Substantia nigra regional selectivity* Brain, 1992. **114**: p. 2283-2301.
71. Scherman D, Desnos C, Darchen F, Pollak P, Javoy-Agid F, Agid Y, *Striatal dopamine deficiency in Parkinson's disease: role of aging.* Annals of Neurology, 1989. **26**(4): p. 551-557.
72. Lee CS, et al., *Clinical observations on the rate of progression of idiopathic parkinsonism.* Brain, 1994. **117**: p. 501-507.

73. Braak H, Del Tredici K, Rüb U, de Vos RAI, Jansen Steur ENH, Braak E, *Staging of brain pathology related to sporadic Parkinson's disease*. *Neurobiology of Aging*, 2003. **24**: p. 197-211.
74. Braak H, Rub U, Gai WP, Del Tredici K, *Idiopathic Parkinson's disease: possible routes by which vulnerable neuronal types may be subject to neuroinvasion by an unknown pathogen*. *Journal of Neural Transmission*, 2003. **110**: p. 517-536.
75. Braak H, Braak E, *Pathoanatomy of Parkinson's disease*. *Journal of Neurology*, 2000. **247**: p. 3-10.
76. Uversky VN, *Introduction to Intrinsically Disordered Proteins (IDPs)*. *Chemical Reviews*, 2014. **114**(13): p. 6557-6560.
77. Dyson HJ, Wright PE, *Intrinsically unstructured proteins and their functions*. *Nature Reviews Molecular Cell Biology*, 2005. **6**: p. 197-208.
78. Wright PE, Dyson HJ, *Intrinsically disordered proteins in cellular signalling and regulation*. *Nature Reviews Molecular Cell Biology*, 2015. **16**: p. 18-29.
79. Babu MM, van der Lee R, Sanchez de Groot N, Gsponer J, *Intrinsically disordered proteins: regulation and disease*. *Current Opinion in Structural Biology*, 2011. **21**: p. 1-9.
80. Oldfield CJ, Dunker AK, *Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions*. *Annual Reviews of Biochemistry*, 2014. **83**: p. 553-584.
81. Sickmeier M, et al., *DisProt: the Database of Disordered Proteins*. *Nucleic Acids Research*, 2006. **35**: p. D786-793.
82. Fischer E, *Einfluss der Configuration auf die Wirkung der Enzyme (Influence of configuration on the action of enzymes)*. *Berichte der Deutschen chemischen Gessellschaft zu Berlin (in German)*, 1894. **27**(3): p. 2985-2993.
83. Henrich S, Salo-Ahen OMH, Huang B, Rippman FF, Cruciani G, Wade RC, *Computational approaches to identifying and characterizing protein binding sites for ligand design*. *Journals of Molecular Recognition*, 2010. **23**: p. 209-219.
84. Koshland DE, *Application of a Theory of Enzyme Specificity to Protein Synthesis*. *Proceeding of the National Academy of Sciences of the United States of America*, 1958. **44**(2): p. 98-104.
85. Habchi J, Tompa P, Longhi S, Uversky VN, *Introducing Protein Intrinsic Disorder*. *Chemical Reviews*, 2014. **114**(13): p. 6561-6588.
86. Dunker AK, et al., *Intrinsically disordered protein*. *Journal of Molecular Graphics and Modelling*, 2001. **16**: p. 26.
87. Tompa P, *Intrinsically unstructured proteins*. *Trend in Biochemical Science*, 2002. **27**(10): p. 527-533.
88. Vucetic S, Brown CJ, Dunker AK, Obradovic Z, *Flavors of Protein Disorder*. *Proteins*, 2003. **52**: p. 573-584.

89. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK, *Sequence Complexity of Disordered Proteins*. Proteins, 2001. **42**: p. 38-48.
90. Sengupta D, Kundu S, *Role of long- and short-range hydrophobic, hydrophilic and charged residues contact network in protein's structural organization*. BMC Bioinformatics, 2012. **13**: p. 142.
91. White SH, Jacobs RE, *Statistical distribution of hydrophobic residues along the length of protein chains. Implications for protein folding and evolution*. Biophysical Journal, 1990. **57**(4): p. 911-921.
92. Jayaraj V, Suhanya R, Vijayasarathy M, Anandagopu P, Rajasekaran E, *Role of large hydrophobic residues in proteins*. Bioinformation, 2009. **3**(9): p. 409-412.
93. Uversky VN, *Natively unfolded proteins: A point where biology waits for physics*. Protein Science, 2002. **11**: p. 739-756.
94. Wright PE, Dyson HJ, *Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm*. Journal of Molecular Biology, 1999. **293**: p. 321-331.
95. Namba K, *Roles of partly unfolded conformations in macromolecular self assembly*. Genes to Cells, 2001. **6**(1): p. 1-12.
96. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z, *Intrinsic Disorder and Protein Function*. Biochemistry, 2002. **41**(21): p. 6573-6582.
97. Ebert M-O, Bae S-H, Dyson HJ, Wright PE, *NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR*. Biochemistry, 2008. **47**: p. 1299-12308.
98. Uversky VN, Narizhneva NV, Ivanova TV, Kirkitadze MD, Tomashevski AY, *Ligand-free form of human α -fetoprotein: evidence for the molten globule state*. FEBS Letters, 1997. **410**: p. 280-284.
99. Hoh JH, *Functional protein domains from the thermally driven motion of polypeptide chains*. Proteins, 1998. **32**: p. 223-228.
100. Reichmann D, Jakob U, *The roles of conditional disorder in redox proteins*. Current Opinion in Structural Biology, 2013. **23**: p. 436-442.
101. Rauscher S, Pomes R, *Structural disorder and protein elasticity*, in *Fuzziness*, M. Fuxreiter and P. Tompa, Editors. 2012, Sperlag: New York. p. 159-183.
102. Trombitás K, Greaser M, Labeit S, Jin JP, Kellermayer M, Helmes M, Granzier H, *Titin extensibility in situ: entropic elasticity of permanently folded and permanently unfolded molecular segments*. Journal of Cellular Biology, 1998. **140**(4): p. 853-859.
103. Qian YQ, Otting G, Furukubo-Tokunaga K, Affolter M, Gehring WJ, K W, *NMR structure determination reveals that the homeodomain is connected through a flexible linker to the main body in the Drosophila antennapedia protein*. Proceeding of the National Academy of Sciences of the United States of America, 1992. **89**: p. 10738-10742.

104. van der Lee R, et al., *Classification of Intrinsically Disordered Regions and Proteins*. Chemical Reviews, 2014. **114**: p. 6589-6631.
105. George RA, Heringa J, *An analysis of protein domain linkers: their classification and role in protein folding*. Protein Engineering, 2003. **15**(11): p. 871-879.
106. Gokhale RS, Khosla C, *Role of linkers in communication between protein modules*. Current Opinion in Chemical Biology, 2000. **4**(1): p. 22-27.
107. Friedrich P, Aszodi A, *MAP2: a sensitive cross-linker and adjustable spacer in dendritic architecture*. Federation of European Biochemical Societies, 1991. **295**: p. 5-9.
108. Tompa P, *Multiteric Regulation by Structural Disorder in Modular Signalling Proteins: An Extension of the Concept of Allostery*. Chemical Reviews, 2013. **114**: p. 6715-6732.
109. Sugase K, Dyson HJ, Wright PE, *Mechanism of coupled folding and binding of an intrinsically disordered protein*. Nature, 2007. **447**: p. 1021-1025.
110. Spolar RS, Record MT, *Coupling of local folding to site-specific binding of proteins to DNA*. Science, 1994. **263**: p. 777-784.
111. Dyson HJ, Wright PE, *Coupling of folding and binding for unstructured proteins*. Current Opinion in Structural Biology, 2002. **12**: p. 54-60.
112. Tompa P, Fuxreiter M, *Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions*. Science, 2008. **18**: p. 2-8.
113. Mittag T, March J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, *Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit*. Science, 2010. **18**(4): p. 494-506.
114. Baker JM, Hudson RP, Kanelis V, Choy WY, Thibodeau PH, Thomas PJ, Forman-Kay JD, *CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices*. Nature Structural Molecular Biology, 2007. **14**(8): p. 738-745.
115. Wang Y, et al., *Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor p21*. Nature Chemical Biology, 2011. **7**(4): p. 214-221.
116. Bertagna A, Totygin D, Brand L, Barrick D, *The effects of conformational heterogeneity on the binding of the Notch intracellular domain to effector proteins: a case of biologically tune disorder*. Biochemical Society Transactions, 2008. **36**: p. 157-166.
117. Kim PM, Sboner A, Xia Y, Gerstein M, *The role of disorder in interaction networks: a structural analysis*. Molecular Systems Biology, 2008. **4**: p. 179.
118. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN, *Flexible nets. The roles of intrinsic disorder in protein interaction networks*. FEBS Journal, 2005. **272**: p. 5129-5148.

119. Hsu WL, et al., *Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding*. Protein Science, 2013. **22**(3): p. 258-273.
120. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK, *The importance of intrinsic disorder for protein phosphorylation*. Nucleic Acids Research, 2004. **32**(3): p. 1037-1049.
121. Tompa P, *Structure and Function of Intrinsically Disordered Proteins*. 2010, Boca Raton, FL: Taylor and Francis.
122. Dobson CM, *The nature and significance of protein folding.*, in *Mechanisms of Protein Folding*, R. Pain, Editor. 2000, Oxford University Press.
123. Tuckerman ME, *Statistical Mechanics: Theory and Molecular Simulation*. 2010, Oxford, UK & New York, NY, USA: Oxford University Press.
124. Karplus M, J.A. M, *Molecular dynamics simulations of biomolecules*. Nature Structural and Molecular Biology, 2002. **9**: p. 646-652.
125. Leach A, *Molecular Modelling: Principles and Applications*. Vol. 2. 2001, Harlow, UK & New York, NY, USA: Prentice Hall.
126. Rauscher S, Gapsys V, Gajda M, Zwecksetter M, de Groot B, Grubmüller H, *Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment*. Journal of Chemical Theory and Computation, 2015. **11**(11): p. 5513-5524.
127. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M, *α -Synuclein in Lewy Bodies*. Nature, 1998. **388**(6645): p. 839-840.
128. Clayton DF, George JM, *The synucleins: a family of proteins involved in synaptic function, plasticity, neurodegeneration and disease*. Trends in Neuroscience, 1998. **21**(6): p. 249-254.
129. Hoyer W, Cherny D, Subramaniam V, Jovin TM, *Impact of the Acidic C-Terminal Region Comprising Amino Acids 109–140 on α -Synuclein Aggregation in Vitro*. Biochemistry, 2004. **43**(51): p. 16233-16242.
130. Ueda K, et al., *Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer's disease*. Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(23): p. 11282-11286.
131. Georgieva ER, Ramlall TF, Borbat PP, Freed JH, Eliezer D, *Membrane-Bound α -Synuclein Forms an Extended Helix: Long-Distance Pulsed ESR Measurements Using Vesicles, Bicelles, and Rod-Like Micelles*. Journal of the American Chemical Society, 2008. **130**(39): p. 12856-12857.
132. Jao CC, Der-Sarkissian A, Chen J, Langen R, *Structure of membrane-bound α -synuclein studied by site-directed spin-labelling*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(22): p. 8331-8336.
133. Carboni E, Lingor P, *Insights on the interaction of α -synuclein and metals in the pathophysiology of Parkinson's disease*. Metallomics, 2015. **7**(3): p. 395-404.

134. Lesage S, et al., *G51D α -synuclein mutation causes a novel parkinsonian-pyramidal syndrome*. *Annals of Neurology*, 2013. **73**(4): p. 459-471.
135. Appel-Cresswell S, et al., *Alpha-synuclein p.H50Q, a novel pathogenic mutation for Parkinson's disease*. *Movement Disorders*, 2013. **28**(6): p. 811-813.
136. Zarranz JJ, et al., *The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia*. *Annals of Neurology*, 2004. **55**(2): p. 164-173.
137. Krüger R, et al., *Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease*. *Nature Genetics*, 1998. **18**(2): p. 106-108.
138. Jackson-Lewis V, Blesa J, Przedborski S, *Animal models of Parkinson's disease*. *Parkinsonism and Related Disorders*, 2012. **18**(Suppl 1): p. S183-S185.
139. Biere A, et al., *Parkinson's disease-associated alpha-synuclein is more fibrillogenic than bet- and gamma-synuclein and cannot cross-see its homologs*. *Journal of Biological Chemistry*, 2000. **275**(44): p. 34574-34579.
140. Du H, Tang L, Luo X, Li H, Hu J, Zhou J, Hu J, *A peptide motif consisting of glycine, alanine, and valine is required for the fibrillization and cytotoxicity of human alpha-synuclein*. *Biochemistry*, 2003. **42**(29): p. 8870-8878.
141. Giasson B, Murray I, Trojanowski J, Lee V, *A hydrophobic stretch of 12 amino acid residues in the middle of alpha-synuclein is essential for filament assembly*. *Journal of Biological Chemistry*, 2001. **276**(4): p. 2380-2386.
142. Periquet M, Fulga T, Myllykangas L, Schlossmacher M, Feany M, *Aggregated alpha-synuclein mediates dopaminergic neurotoxicity in vivo*. *Journal of Neuroscience*, 2007. **27**(12): p. 3336-3346.
143. Rodriguez J, et al., *Structure of the toxic core of α -synuclein from invisible crystals*. *Nature*, 2015. **525**(7570): p. 486-490.
144. Tuttle MD, et al., *Solid-state NMR structure of a pathogenic fibril of full-length human α -synuclein*. *Nature Structural and Molecular Biology*, 2016. **23**: p. 409-415.
145. Maroteaux L, Campanelli JT, Scheller RH, *Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal*. *Journal of Neuroscience*, 1988. **8**: p. 2804-2815.
146. George JM, Jin H, Woods WS, Clayton DF, *Characterization of a novel protein regulated during the critical period for song learning in the zebra finch*. *Neuron*, 1995. **15**: p. 361-372.
147. Iwai A, et al., *The precursor protein of non-A beta component of Alzheimer's disease amyloid is a presynaptic protein of the central nervous system*. *Neuron*, 1995. **14**: p. 467-475.
148. Cabin DE, et al., *Synaptic vesicle depletion correlates with attenuated synaptic responses to prolonged repetitive stimulation in mice lacking alpha-synuclein*. *Journal of Neuroscience*, 2002. **22**(20): p. 8797-8807.

149. Abeliovich A, et al., *Mice lacking alpha-synuclein display functional deficits in the nigrostriatal dopamine system*. *Neuron*, 2000. **25**(1): p. 239-252.
150. Fleming SM, Tetreault NA, Mulligan CK, Huston CB, Masliah E, Cesselet MF, *Olfactory deficits in mice overexpressing human wild-type alpha-synuclein*. *European Journal of Neuroscience*, 2008. **28**: p. 247-256.
151. Kuo YM, et al., *Extensive enteric nervous system abnormalities in mice transgenic for artificial chromosomes containing Parkinson disease-associated alpha-synuclein gene mutations precede central nervous system changes*. *Human Molecular Genetics*, 2010. **19**: p. 1633-1650.
152. Noorian AR, Rha J, Annerino DM, Bernhard D, Taylor GM, Greene JG, *Alpha-synuclein transgenic mice display age-related slowing of gastrointestinal motility associated with transgene expression in the vagal system*. *Neurobiology of Disease*, 2012. **48**: p. 9-19.
153. Scott DA, Tabarean I, Tang Y, Cartier A, Masliah E, Roy S, *A pathologic cascade leading to synaptic dysfunction in alpha-synuclein-induced neurodegeneration*. *Journal of Neuroscience*, 2010. **30**(24): p. 8083-8095.
154. Nemani VM, et al., *Increased expression of alpha-synuclein reduces neurotransmitter release by inhibiting synaptic vesicle recluster after endocytosis*. *Neuron*, 2010. **65**(1): p. 66-79.
155. Frenkel D, Smit B, *Understanding Molecular Simulation: From Algorithms to Applications*. 2002, San Diego, San Francisco, New York, Boston, London, Sydney, Tokyo: ACADEMIC PRESS: A division of Harcourt, Inc.
156. Metropolis N, Ulam S, *The Monte Carlo Method*. *Journal of the American Statistical Association*, 1949. **44**(247): p. 335-341.
157. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, *Equation of State Calculations by Fast Computing Machines*. *The Journal of Chemical Physics*, 1953. **21**(6): p. 1087-1092.
158. Landau DP, Binder K, *A Guide to Monte Carlo Simulations in Statistical Physics*. 2005, Cambridge, New York, Melbourne, Cape Town, Singapore, Sao Paulo: Cambridge University Press.
159. Hastings WK, *Monte Carlo Sampling Methods Using Markov Chains and their Applications*. *Biometrika*, 1970. **57**(1): p. 97-109.
160. Siepmann JI, Frenkel D, *Configurational bias Monte Carlo: a new sampling scheme for flexible chains*. *Molecular Physics*, 1992. **75**(1): p. 59-70.
161. Rosenbluth MN, Rosenbluth AW, *Monte Carlo simulations of the average extension of molecular chains*. *Journal of Chemical Physics*, 1955. **23**: p. 356-359.
162. Favrin G, Irbäck A, Sjunnesson F, *Monte Carlo update for chain molecules: Biased Gaussian steps in torsional space*. *Journal of Chemical Physics*, 2001. **114**(18).

163. Hansmann UHE, *Parallel tempering algorithm for conformational studies of biological molecules*. Chemical Physics Letters, 1997. **281**: p. 140-150.
164. Earl DJ, Deem MW, *Parallel Tempering: Theory, applications, and new perspectives*. Physical Chemistry Chemical Physics, 2005. **7**: p. 3910.
165. Swendsen RH, Wang J-S, *Replica Monte Carlo Simulation of Spin-Glasses*. Physical Review Letters, 1986. **57**(21): p. 2607-2609.
166. Alder BJ, Wainwright TE, *Studies in Molecular Dynamics, I. General Method*. The Journal of Chemical Physics, 1959. **31**(2): p. 459-466.
167. Levitt M, Warshel A, *Computer simulation of protein folding*. Nature, 1975. **253**: p. 694-698.
168. McCammon JA, Gelin BR, Karplus M, *Dynamics of folded proteins*. Nature 1977. **267**: p. 585-590.
169. Zolotoyabko E, *Basic Concepts of X-Ray Diffraction*. 2014, Weinheim, Germany: Wiley-VCH.
170. Hore PJ, *Nuclear Magnetic Resonance*. 2015, Oxford, United Kingdom: Oxford University Press.
171. Cornell WED, et al., *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules*. Journal of the American Chemical Society, 1995. **117**: p. 5179-5197.
172. Mackerell AD, Wiorkiewicz-Kuczera J, Karplus M, *An all-atom empirical energy function for the simulation of nucleic acids*. Journal of the American Chemical Society, 1995. **117**: p. 11946-11975.
173. Patra M, Hyvönen MT, Falck E, Sabouri-Ghomi M, Vattulainen I, Karttunen M, *Long-range interactions and parallel scalability in molecular simulations*. Computer Physics Communications, 2007. **176**(1): p. 14-22.
174. Onsager L, *Electric moments of molecules in liquids*. Journal of the American Chemical Society, 1936. **58**: p. 1486-1493.
175. Allen MP, Tildesley DJ, *Computer simulation of liquids*. 1997, Oxford: Clarendon Press.
176. Phillips JC, et al., *Scalable Molecular Dynamics with NAMD*. Journal of Computational Chemistry, 2005. **26**: p. 1781-1802.
177. Brunger A, Brooks III C, Karplus M, *Stochastic Boundary Conditions for Molecular Dynamics Simulations of ST2 Water*. Chemical Physics Letters, 1984. **105**(5): p. 495-500.
178. de Miguel E, Jackson G, *The nature of the calculation of the pressure in molecular simulations of continuous models from volume perturbations*. The Journal of Chemical Physics, 2006. **125**: p. 164109.

179. Feller SE, Zhang Y, Pastor RW, Brooks BR, *Constant pressure molecular dynamics simulation: The Langevin piston method*. The Journal of Chemical Physics, 1995. **103**: p. 4613-4621.
180. Martyna GJ, Tobias DJ, Klein ML, *Constant pressure molecular dynamics algorithms*. The Journal of Chemical Physics, 1994. **101**(5): p. 4177-4189.
181. Quigley D, Probert MIJ, *Langevin dynamics in constant pressure extended systems*. The Journal of Chemical Physics, 2004. **120**(24): p. 11432-11441.
182. Bhandarkar M, et al. *NAMD users's guide*.
183. Schmidtke P, Alvarez-Garcia D, Seco J, Barril X, *Expanding the Target Space: Druggability Assessments*, in *Physico-Chemical and Computational Approaches to Drug Discovery*, B.X. Luque FJ, Editor. 2012, The Royal Society of Chemistry: Cambridge. p. 302-315.
184. Schneider G, Fechner U, *Computer-based de novo design of drug-like molecules*. Nature Reviews Drug Discovery, 2005. **4**: p. 649-663.
185. Hopkins AL, Groom CR, *The druggable genome*. Nature Reviews: Drug Discovery, 2002. **1**: p. 727-730.
186. Keller TH, Pichota A, Yin Z, *A practical view of "druggability"*. Current Opinion in Chemical Biology, 2006. **10**: p. 357-361.
187. An J, Totrov M, Abagyan R, *Comprehensive Identification of "Druggable" Protein Ligand Binding Sites*. Genome Informatics, 2004. **15**(2): p. 31-41.
188. Kellenberger E, Muller P, Schalon C, Bret G, Foata N, Rognan D, *sc-PDB: an Annotated Database of Druggable Binding Sites from the Protein Data Bank*. Journal of Chemical Information and Modelling, 2006. **46**: p. 717-727.
189. Hajduk PJ, Huth JR, Tse C, *Predicting protein druggability*. Drug Discovery Today, 2005. **10**(1675-1682).
190. Bakan A, Neysa N, Lakdawala AS, Bahar I, *Druggability Assessment of Allosteric Proteins by Dynamics Simulations in the Presence of Probe Molecules*. Journal of Chemical Theory and Computation, 2012. **8**(7): p. 2435-2447.
191. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ, *Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility*. Journal of Computational Chemistry, 2009. **16**: p. 2785-2791.
192. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ, *Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function*. Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.
193. Irbäck A, Mohanty S, *PROFASI: A Monte Carlo simulation package for protein folding and aggregation*. Journal of Computational Chemistry, 2006. **27**: p. 1548-1555.
194. Humphrey WA, Dalke A, Schulten K, *VMD - Visual Molecular Dynamics*. Journal of Molecular Graphics, 1996. **14**: p. 33-38.

195. Case DA, et al., *AMBER 12*. University of California, San Francisco, 2012.
196. Eaton JW, Bateman D, Hauberg S, Wehbring R, *GNU Octave 3.8.1 manual: a high-level interactive language for numerical computation*. 2014, CreateSpace Independent Publishing Platform.
197. Eaton JW, Bateman D, Hauberg S, (*GNU Octave*) version 3.0.1 manual: a high-level interactive language for numerical computations. 2009: CreateSpace Independent Publishing Platform.
198. Han B, Liu Y, Ginzinger SW, Wishart DS, *SHIFTX2: significantly improved protein chemical shift prediction*. Journal of Biomolecular NMR, 2011(50): p. 43-57.
199. Ulrich EL, et al., *BioMagResBank*. Nucleic Acids Research, 2008. **36**: p. D402-D408.
200. Jürgen MS, Blümel M, Löhr F, Rüterjans H, *Self-consistent 3J coupling analysis for the joint calibration of Karplus coefficients and evaluation of torsion angles*. Journal of Biomolecular NMR, 1999(14): p. 1-12.
201. Ortega A, Amorós D, Garcia de la Torre J, *Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models*. Biophysical Journal, 2011. **101**: p. 892-898.
202. Bermel W, et al., *Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins*. Journal of the American Chemical Society, 2006. **128**: p. 3918-3919.
203. Neal S, Nip AM, Zhang H, Wishart DS, *Rapid and accurate calculation of protein 1H, 13C, and 15N chemical shifts*. Journal of Biomolecular NMR, 2003. **26**(3): p. 215-240.
204. Jónsson SA, Mohanty S, Irbäck A, *Distinct phase of free α -synuclein--a Monte Carlo study*. Proteins, 2012. **80**(9): p. 2169-2177.
205. Wu K-P, Weinstock DS, Narayanan C, Levy RM, Baum J, *Structural reorganization of α -Synuclein at low pH observed by NMR and REMD simulations*. Journal of Molecular Biology, 2009. **391**: p. 784-796.
206. Kim H-Y, Heise H, Fernandez CO, Baldus M, Zweckstetter M, *Correlation of amyloid fibril B-structure with the unfolded state of α -Synuclein*. ChemBioChem, 2007. **8**: p. 1671-1674.
207. Cho M-K, et al., *Structural characterization of α -synuclein in an aggregation prone state*. Protein Science, 2009. **18**: p. 1840-1846.
208. Lendel C, Damberg P, *3D J-resolved NMR spectroscopy for unstructured polypeptides: fast measurement of $^3J_{\text{HNHa}}$ coupling constants with outstanding spectral resolution*. Journal of Biomolecular NMR, 2009. **44**: p. 35-42.
209. Binolfi A, et al., *Interaction of α -synuclein with divalent metal ions reveals key differences: a link between structure, binding specificity and fibrillation*

- enhancement*. Journal of the American Chemical Society, 2006. **128**: p. 9893-9901.
210. Allison JR, P V, Dobson CM, Vendruscolo M, *Determination of the Free Energy Landscape of α -Synuclein Using Spin Label Nuclear Magnetic Resonance Measurements*. Journal of the American Chemical Society, 2009. **131**(51): p. 18314-18326.
 211. Lemkau LR, Comellas G, Kloepper KD, Woods WS, George JM, Rienstra CM, *Mutant Protein A30P α -Synuclein Adopts Wild-type Fibril Structure, Despite Slower Fibrillation Kinetic*. Journal of Biological Chemistry, 2012. **287**(14): p. 11526-11532.
 212. Pivato M, et al., *Covalent α -Synuclein Dimers: Chemico-Physical and Aggregation Properties*. PLoS One, 2012. **7**(12): p. e50027.
 213. Chen M, Margittai M, Jeannie C, Langen R, *Investigation of α -Synuclein Fibril Structure by Site-directed Spin Labelling*. The Journal of Biological Chemistry, 2007. **282**: p. 24970-24979.
 214. Heise H, Celej MS, Becker S, Riedel D, Pelah AK, A, Jovin TM, Baldus M, *Solid-State NMR Reveals Structural Differences between Fibrils of Wild-Type and Disease-Related A53T Mutant α -Synuclein*. Journal of Molecular Biology, 2008. **380**(3): p. 444-450.
 215. Heise H, Hoyer W, Becker S, Andronesi OC, Riedel D, Baldus M, *Molecular-level secondary structure, polymorphism, and dynamics of full-length α -synuclein fibrils studied by solid-state NMR*. Proceeding of the National Academy of Sciences of the United States of America, 2005. **102**(44): p. 15871-15876.
 216. Lemkau LR, Comellas G, Lee SW, Rikardsen LK, Woods WS, George JM, Rienstra CM, *Site-Specific Perturbations of Alpha-Synuclein Fibril Structure by the Parkinson's Disease Associated Mutations A53T and E46K*. PLoS One, 2013. **8**(3): p. e49750.
 217. Choi W, Zibae S, Jakes R, Serpell LC, Daveletov B, Crowther RA, Goedert M, *Mutation E46K increases phospholipid binding and assembly into filaments of human alpha-synuclein*. FEBS Letters, 2004. **576**(3): p. 363-8.
 218. Fredenburg RA, Rospigliosi C, Meray RK, Kessler JC, Lashuel HA, Eliezer D, Lansbury P, Jr., *The impact of the E46K mutation on the properties of alpha-synuclein in its monomeric and oligomeric states*. Biochemistry, 2007. **46**(24): p. 7107-7118.
 219. Greenbaum EA, et al., *The E46K Mutation in α -Synuclein Increases Amyloid Fibril Formation*. The Journal of Biological Chemistry, 2005. **280**(9): p. 7800-7807.
 220. Li J, Uversky VN, Fink AL, *Effect of Familial Parkinson's Disease Point Mutations A30P and A53T on the Structural Properties, Aggregation, and Fibrillation of Human α -Synuclein*. Biochemistry, 2001. **40**: p. 11604-11613.
 221. Narhi L, et al., *Both familial Parkinson's disease mutations accelerate α -synuclein aggregation*. Journal of Biological Chemistry, 1999. **274**(14): p. 9843-9846.

222. Serpell LC, Berriman J, Jakes R, Goedert M, Crowther RA, *Fiber diffraction of synthetic α -synuclein filaments shows amyloid-like cross- β conformation*. *Proceeding of the National Academy of Sciences of the United States of America*, 2000. **97**(9): p. 4897-4902.
223. Camilloni C, Vendruscolo M, *A Relationship between the Aggregation Rates of α -Synuclein Variants and the β -Sheet Populations in Their Monomeric Forms*. *The Journal of Physical Chemistry B*, 2013. **117**: p. 10737-10741.
224. Pawar AP, DuBay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CM, *Prediction of "Aggregation-prone" and "Aggregation-susceptible" Regions in Proteins Associated with Neurodegenerative Diseases*. *Journal of Molecular Biology*, 2005. **350**: p. 379-392.
225. Fusco G, De Simone A, Gopinath T, Vostrikov V, Vendruscolo M, Dobson CM, Veglia G, *Direct observation of the three regions in α -synuclein that determine its membrane-bound behaviour*. *Nature Communications*, 2014. **5**.
226. Diao J, et al., *Native α -synuclein induces clustering of synaptic-vesicle mimics via binding to phospholipids and synaptobrevin-2/VAMP2*. *eLife*, 2013. **2**: p. e000592.
227. Jensen PH, Nielsen MS, Jakes R, Dotti CG, Goedert M, *Binding of α -Synuclein to Brain Vesicles is Abolished by Familial Parkinson's Disease Mutation*. *The Journal of Biological Chemistry*, 1998. **273**: p. 26292 - 26294.
228. Jo E, Fuller N, Rand RP, St George-Hyslop P, Fraser PE, *Defective membrane interactions of familial Parkinson's disease mutant A30P α -synuclein*. *Journal of Molecular Biology*, 2002. **315**(4): p. 799-807.
229. Jónsson SÆ, Mitternacht S, Irbäck A, *Mechanical resistance in unstructured proteins*. *Biophysical Journal*, 2013. **104**(12): p. 2725-2732.
230. Hervás R, et al., *Common features at the start of the neurodegeneration cascade*. *PLoS Biol*, 2012. **10**: p. e1001335.
231. Neupane K, Solanki A, Sosova I, Belov M, Woodside MT, *Diverse Metastable Structures Formed by Small Oligomers of α -Synuclein Probed by Force Spectroscopy*. *PLoS: One*, 2014. **9**(1): p. e86495.
232. Marko JF, Siggia ED, *Stretching DNA*. *Macromolecules*, 1995. **28**(26): p. 8759-8770.
233. Ester M, Kriegel HP, Sander J, Xu X. *A density-based algorithm for discovering clusters in large spatial databases with noise*. in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. 1996. AAAI Press.
234. Chem D, Martin ZS, Soto C, Schein CH, *Computational Selection of Inhibitors of A-beta Aggregation and Neuronal Toxicity*. *Bioorganic and Medicinal Chemistry*, 2009. **17**(14): p. 5189-5197.
235. Toth G, et al., *Targetting the Intrinsically Disordered Structural Ensemble of α -Synuclein by Small Molecules as Potential Therapeutic Strategy for Parkinson's Disease*. *PLoS One*, 2014. **9**(2): p. e87133.

236. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ, *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. *Advanced Drug Delivery Reviews*, 2001. **46**(1): p. 3-26.
237. Kobus T, Vykhodtseva N, Pilatou M, Zhang Y, McDannold N, *Safety Validation of Repeated Blood-Brain Barrier Disruption Using Focused Ultrasound*. *Ultrasound in medicine and biology*, 2016. **42**(2): p. 481-492.
238. Zuchero YJY, et al., *Discovery of Novel Blood-Brain Barrier Targets to Enhance Brain Uptake of Therapeutic Antibodies*. *Neuron*, 2016. **89**(1): p. 70-82.
239. Gregori M, et al., *Novel Antitransferrin Receptor Antibodies Improve the Blood-Brain Barrier Crossing Efficacy of Immunoliposomes*. *Journal of Pharmaceutical Sciences*, 2016. **105**(1): p. 276-283.
240. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG, *ZINC: A Free Tool to Discover Chemistry for Biology*. *Journal of Chemical Information and Modelling*, 2012. **52**(7): p. 1757-1768.
241. Minicozzi V, et al., *Computational and Experimental Studies on β -sheet Breakers Targeting A β 1-40 Fibrils* *The Journal of Biological Chemistry*, 20154. **289**(16): p. 11242-11252.

9 APPENDICES

APPENDIX 1 UNIFIED PARKINSON’S DISEASE RATING SCALE	110
APPENDIX 2: FULL TABLE OF BINDING REGIONS FOR DRUGGABILITY ANALYSIS	112
APPENDIX 3: TOP BINDING ENERGIES FROM BLIND DOCKING	121

APPENDIX 1 UNIFIED PARKINSON'S DISEASE RATING SCALE

- I. Mentation, Behaviour, and Mood
 1. Intellectual Impairment
 2. Thought Disorder
 3. Depression
 4. Motivation/Initiative
- II. Activities of Daily Living
 5. Speech
 6. Salivation
 7. Swallowing
 8. Handwriting
 9. Cutting Food and Handling Utensils
 10. Dressing
 11. Hygiene
 12. Turning in Bed and Adjusting Clothes
 13. Falling (Unrelated to Freezing)
 14. Freezing when Walking
 15. Walking
 16. Tremor (Symptomatic complaint of tremor in any part of body.)
 17. Sensory Complaints Related to Parkinsonism
- III. Motor Examination
 18. Speech
 19. Facial Expression
 20. Tremor at Rest
 21. Action or Postural Tremor of Hands
 22. Rigidity
 23. Finger Taps
 24. Hand Movements
 25. Rapid Alternating Movements of Hands
 26. Leg Agility
 27. Arising from Chair

- 28. Posture
- 29. Gait
- 30. Postural Stability
- 31. Body Bradykinesia and Hypokinesia

IV. Complications

A. Dyskinesias

- 1. Duration
 - i. 0
- 2. Disability: How disabling are the dyskinesias?
- 3. Painful Dyskinesias: How painful are the dyskinesias?
- 4. Presence of Early Morning Dystonia

B. Clinical Fluctuations

- 1. Are “off” periods predictable?
- 2. Are “off” periods unpredictable?
- 3. Do “off” periods come on suddenly, within a the few seconds?
- 4. What proportion of the waking day is the patient “off” on average?

C. Other Complications

- 1. Does the patient have anorexia, nausea, or vomiting?
- 2. Any sleep disturbances, such as insomnia or hypersomnolence?
- 3. Does the patient have symptomatic orthostasis?

**APPENDIX 2: FULL TABLE OF BINDING REGIONS
FOR DRUGGABILITY ANALYSIS**

Rep.	Site	Site Residues	Maximal Binding Affinity (nM)
1	1	21,28,29,30,40,41,42,43,44,50,51,52,53,54,66,67,68,69	16.93 → 22.86
	2	29,30,37,38,39,41,42,43,51,55,57	488 → 630
	3	63,65,82,137	1.41 → 1.50
2	1	31,33,37,38,40,54,56,58,64,65,66,67,78,79,79,80,81,91,92	0.131 → 0.449
	2	29,30,39,40,41,52,53,54,68,69,78	0.321 → 0.367
3	1	4,6,15,16,	0.052 →

Chapter 9: Appendices

		17,25,26,2 7,28,50,51 ,52,53,54, 61,63,79,8 1,81	0.152
	2	12,13,14,2 5,27,29,30 ,31,32,48, 51	0.881 → 1.352
4	1	7,8,9,20,2 1,22,27,29 ,40,41,42, 43,44,50,5 1,52,66,68	0.215 → 0.312
	2	41,53,65,6 7,68,77,78 ,79,80,81, 90,90,91,9 2,95	1.224 → 1.398
	3	17,19,31,3 2,32,33,36 ,37,38,39, 55,56,57	5.02 → 6.526
5	1	40,61,63,6 6,67,68,69 ,70,79,80, 81,92,93,9 4	1.221 → 1.336
	2	17,30,31,3 2,34,39,41 ,42,43,58,	9.628 → 16.14

		59,60,62,7 0,71,78	
	3	60,65,66,6 7,68,69,82 ,83,84	0.851
6	1	No Sites	NA
7	1	41,57,59,6 8,69,70,77 ,79,81,87, 89,100	0.088 → 0.098
	2	7,8,9,20,2 1,22,29,30 ,31,32,33, 46,47,48,4 9,50,56,57 ,58	0.576 → 0.939
	3	70,72,75,7 7,91,92,93 ,96,97,98	1.927 → 87.93
8	1	50,51,52,5 3,54,55,69 ,70,71,83, 84,85,86	5.30 → 6.72
	2	29,34,46,5 5,57	117
	3	12,13,14,1 5,26,27,28 ,32,33,34, 35,47,48,4 9,50,54	267

Chapter 9: Appendices

9	1	2,3,4,11,1 3,35,36,37 ,54,55,56, 67,76	0.461 → 0.539
	2	10,12,36,3 8,40,53,55 ,66,67,68, 69,70,76,7 7,78	4.91→ 6.85
	3	18,20,23,2 9,30,31,34 ,57,59,62, 63,64,81,8 2,83	1558
10	1	No Sites	N/A
11	1	73,75,77,7 8,83,84,85 ,86,87,89, 100,101,1 02,103,10 4,110,111, 112,113,1 14	0.792 → 2.611
	2	19,28,29,3 0,41,47,48 ,49,50,59, 60,61,62,7 2,73,74,75 ,76,87,88, 99	1.029 → 2.62

	3	13,14,15,1 6,17,18,19 ,30,31,32, 33,34,35,3 6,37,38,39 ,50,51,52, 53,54,55,5 6,57,58,76	1.117 → 1.494
12	1	14,24,25,2 6,27,28,45 ,46,47,48, 49,59,60,6 1,62,63,76 ,78,80	0.327 → 0.779
	2	12,25,27,4 6,60,62,79 ,80,81,88, 89,90,100, 101,102	14.23 → 26.95
13	1	58,60,75,7 7,79,88,90 ,92,99,100 ,101,103,1 03	0.749
	2	6,18,19,20 ,29,30,31, 32,33,34,4 6,47,48,49 ,50,57,59, 60,61	0.781 → 1.113

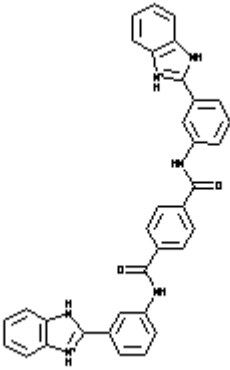
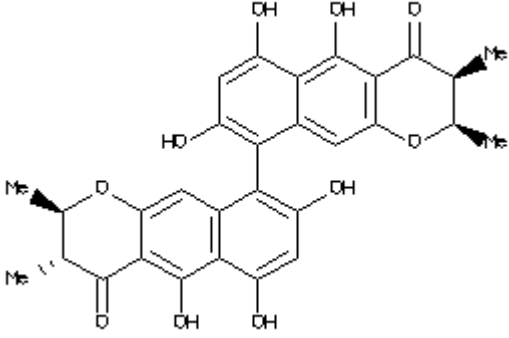
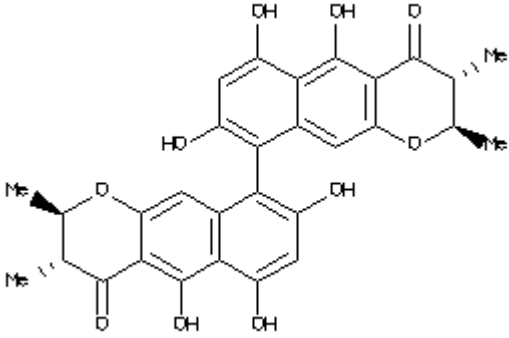
	3	20,22,29,3 0,31,32,33 ,46,48,49, 50,57,59	26.56
	4	78,80,85,8 6,87,88,89 ,100,101,1 02,103,10 4,108,109, 110,111,1 12,113,12 1,122,123	27.23 → 29.33
14	1	17,19,26,2 7,28,29,35 ,36,37,38, 39,53,54,6 3,78	0.00705 → 0.0076
	2	8,9,10,18, 19,20,21,2 4,25,26,29 ,36,40,53	0.36 → 0.641
	3	67,68,69,7 2,73,74,75 ,91,92,93, 94,95,96,9 7,98	202
15	1	69,71,75,7 6,77,96,98 ,99,101,10 4,106,106	14.77 → 17.66

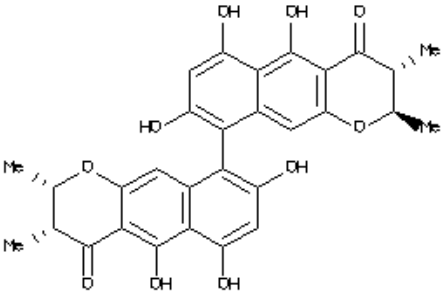
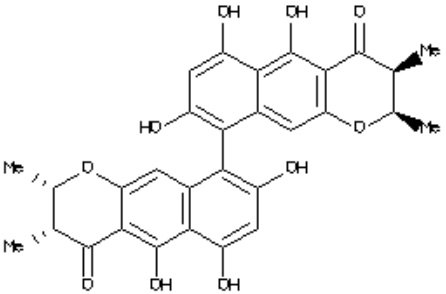
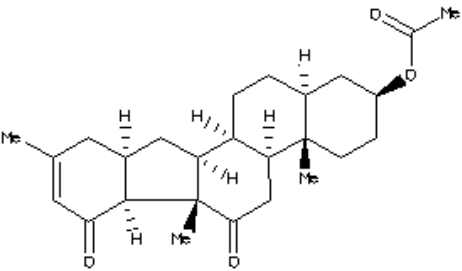
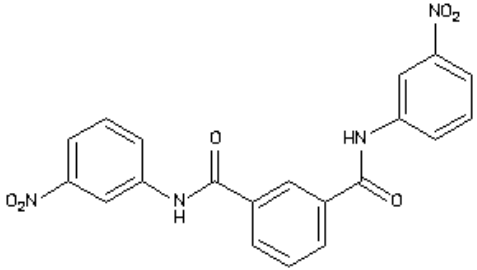
	2	5,6,7,8,9,1 2,13,14,15 ,16,,29,30, 31,32,33,3 4,38,39,40 ,41	14.78 → 31.24
16	1	2,13,18,19 ,20,31,33, 38,39,40,5 1,52,53	0.236 → 0.788
	2	50,51,59,6 0,61,72,73 ,74,75,76, 83,84,85,8 6,87,88,10 0,101,102, 104	0.617 → 0.839
	3	12,14,17,1 8,19,34,39	→ 8.933
17	1	28,29,30,3 1,32,41,42 ,43,54,55, 56,57,58,6 4,66	3.89 → 4.68
	2	18,19,20,2 1,22,29,31 ,32,33,38, 40	9.75 16. 07
	3	4,22,23,24 ,27,28,43, 44,45,51	396 → 420

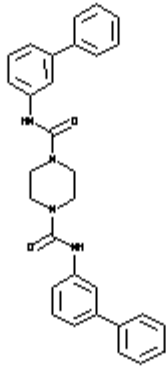
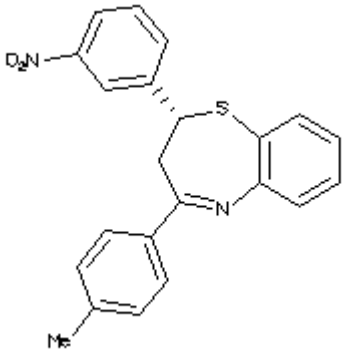
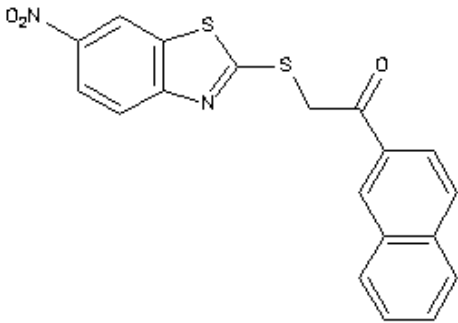
	4	38,39,40,5 3,55,56,57 ,58,59,60, 61,63	1589
	5	50,51,67,6 8,69,70,71 ,77,78,79, 80,94,95	1689
18	1	67,75,76,7 7,79,89,90 ,91,92,93, 98,99,100, 101,102	1.707 → 3.124
	2	34,36,45,4 6,56,57,58 ,66,67,68, 78,79,80,8 8,103	3.447
	3	52,53,54,5 5,56,68,69 ,70,71,72, 75,76,77,7 8,90,92	11.87 → 19.12
	4	7,8,9,20,2 1,22,23,24 ,25,35,36, 37,38,39,4 0,41,42	56.69 → 91.67
19	1	None	N/A
20	1	9,10,11,17 ,19,21,48,	11.67 → 17.50

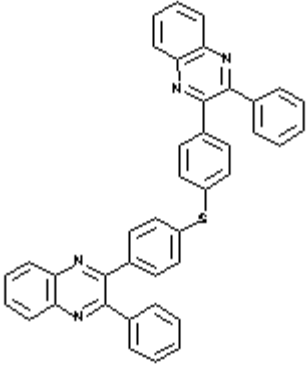
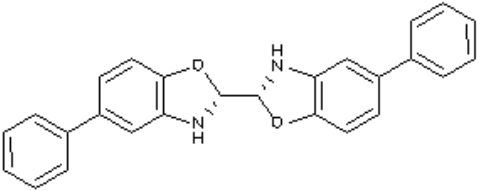
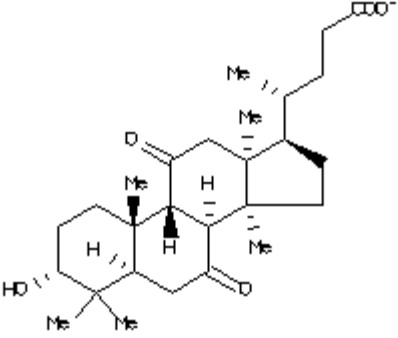
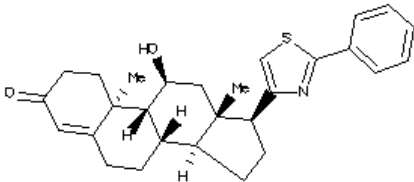
		50,57,59,	
	2	18,19,20,2 1,22,47,48 ,49,50,51, 56,57,58,5 9,60,65,66	40.73 → 78.31
	3	67,81,83,8 4,85,86,87 ,88,89,90	44.56 →62.91

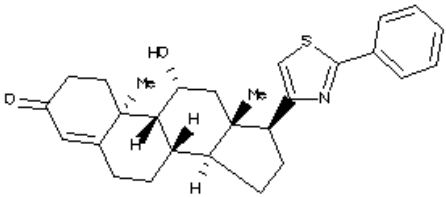
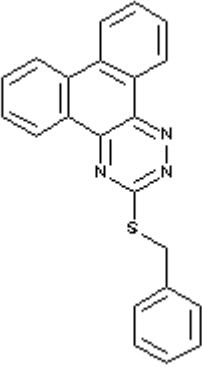
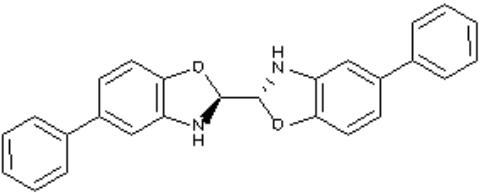
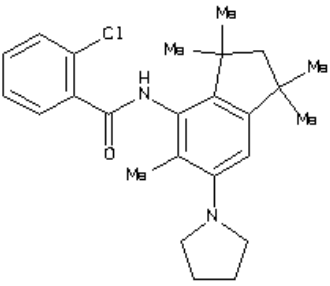
APPENDIX 3: TOP BINDING ENERGIES FROM BLIND DOCKING

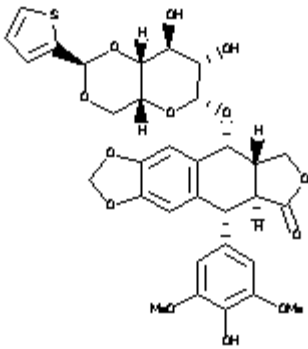
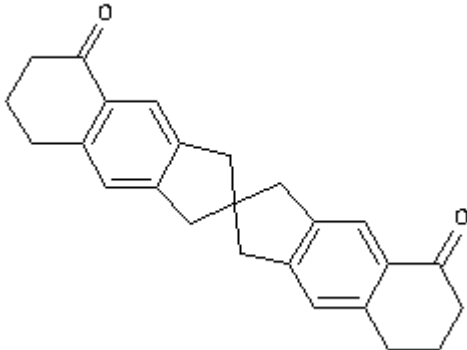
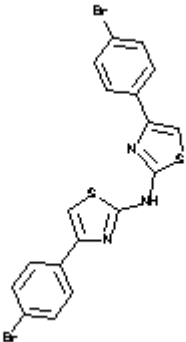
Structure + ZINC ID	Binding Energy (kcal/mol)
 <p>ZINC ID: 1690699</p>	-8.6 ± 0.8
 <p>ZINC ID: 17465979</p>	-8.5 ± 0.8
 <p>ZINC ID: 17465983</p>	-8.4 ± 0.8

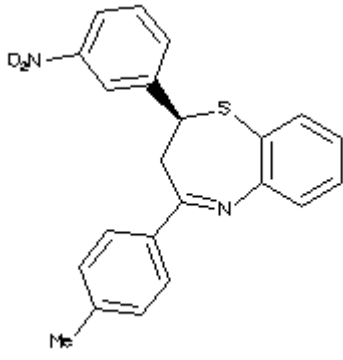
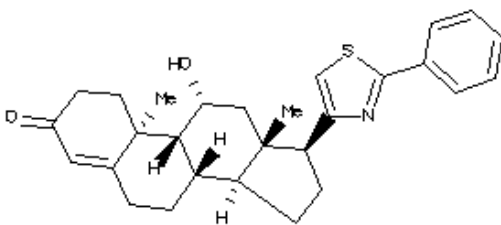
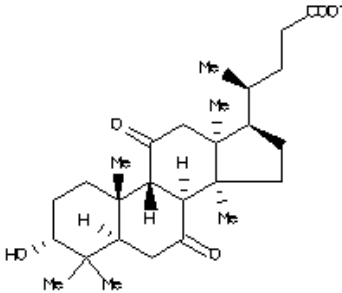
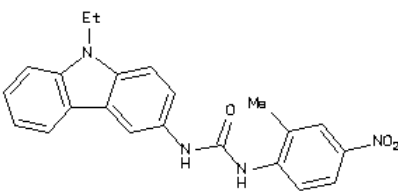
 <p>ZINC ID: 5462670</p>	<p>-8.4 ± 0.8</p>
 <p>ZINC ID: 5462670</p>	<p>-8.3 ± 0.8</p>
 <p>ZINC ID: 12671893</p>	<p>-8.0 ± 0.7</p>
 <p>ZINC ID: 1738764</p>	<p>-8.0 ± 0.9</p>

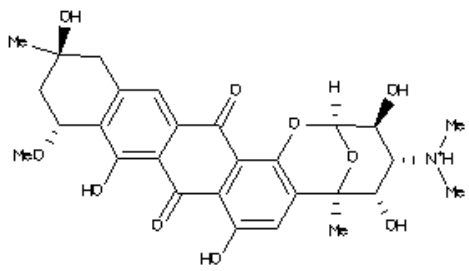
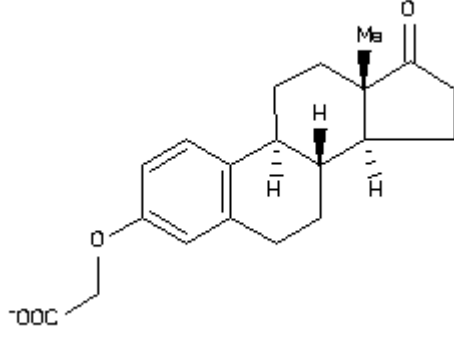
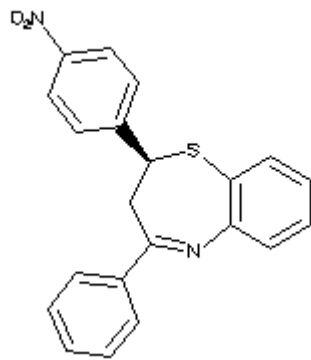
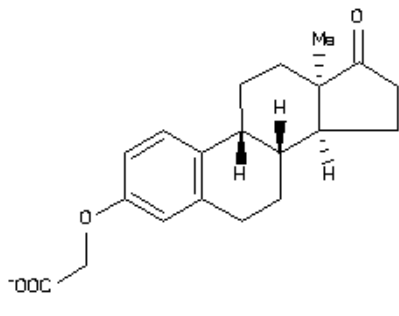
 <p>ZINC ID: 4783229</p>	<p>-7.9 ± 0.6</p>
 <p>ZINC ID: 3954520</p>	<p>-7.9 ± 0.5</p>
 <p>ZINC ID: 1163259</p>	<p>-7.9 ± 0.8</p>

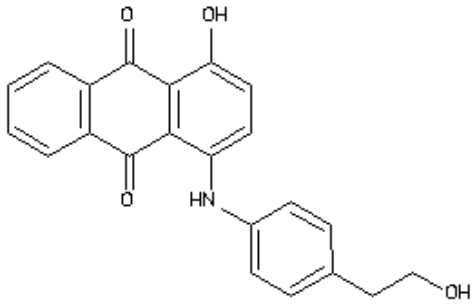
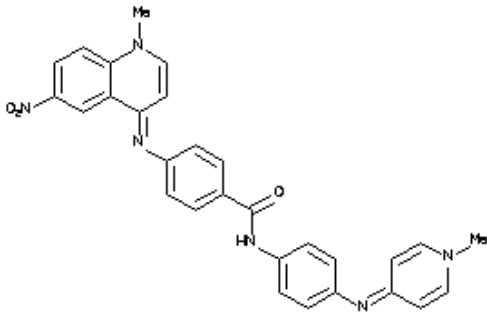
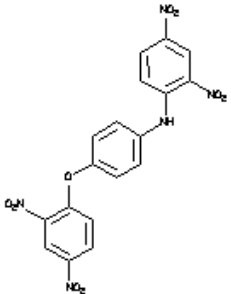
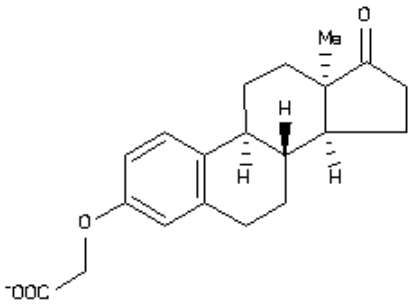
 <p>ZINC ID: 5492794</p>	<p>-7.8 ± 0.5</p>
 <p>ZINC ID: 1736228</p>	<p>-7.7 ± 0.5</p>
 <p>ZINC ID: 12670920</p>	<p>-7.7 ± 0.7</p>
 <p>ZINC ID: 13099051</p>	<p>-7.7 ± 0.6</p>

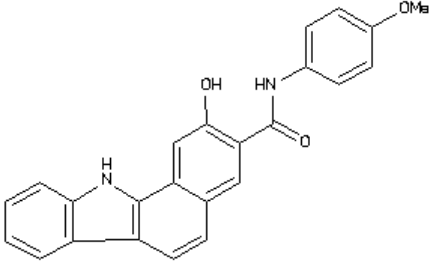
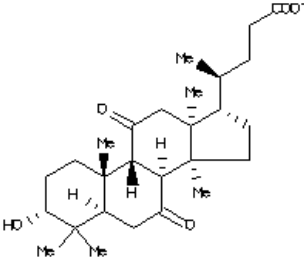
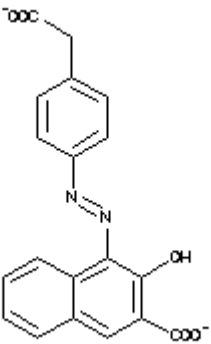
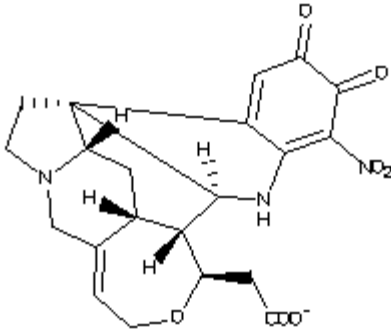
 <p>ZINC ID: 13099048</p>	-7.7 ± 0.6
 <p>ZINC ID: 1573829</p>	-7.7 ± 0.5
 <p>ZINC ID: 13152284</p>	-7.7 ± 0.5
 <p>ZINC ID: 1572767</p>	-7.7 ± 0.6

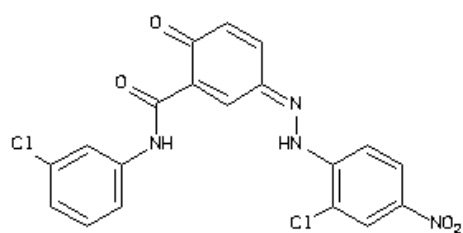
 <p>ZINC ID: 11616857</p>	<p>-7.7 ± 0.6</p>
 <p>ZINC ID: 1855333</p>	<p>-7.7 ± 0.5</p>
 <p>ZINC ID: 13154298</p>	<p>-7.6 ± 0.6</p>

 <p>ZINC ID: 4376856</p>	-7.6 ± 0.6
 <p>ZINC ID: 13099048</p>	-7.6 ± 0.6
 <p>ZINC ID: 12670914</p>	-7.6 ± 0.6
 <p>ZINC ID: 5390471</p>	-7.6 ± 0.6

 <p>ZINC ID: 11681161</p>	<p>-7.5 ± 0.6</p>
 <p>ZINC ID: 4428843</p>	<p>-7.5 ± 0.8</p>
 <p>ZINC ID: 1045090</p>	<p>-7.5 ± 0.6</p>
 <p>ZINC ID: 13099024</p>	<p>-7.4 ± 0.8</p>

 <p>ZINC ID: 5124960</p>	<p>-7.4 ± 0.6</p>
 <p>ZINC ID: 18057104</p>	<p>-7.4 ± 0.6</p>
 <p>ZINC ID: 17465965</p>	<p>-7.4 ± 1.3</p>
 <p>ZINC ID: 13099027</p>	<p>-7.4 ± 0.7</p>

 <p>ZINC ID: 4720972</p>	<p>-7.4 ± 0.7</p>
 <p>ZINC ID: 12670903</p>	<p>-7.3 ± 0.9</p>
 <p>ZINC ID: 5124957</p>	<p>-7.2 ± 0.9</p>
 <p>ZINC ID: 26730911</p>	<p>-7.2 ± 0.9</p>



ZINC ID: 17995347

 -7.7 ± 0.9