Computational Examination of Biomolecular Systems Related to Alzheimer's and Parkinson's Diseases

Loizos Savva

A thesis submitted to
Cardiff University
in accordance with the requirements for the degree of
Doctor of Philosophy

School of Chemistry
Cardiff University
United Kingdom
July 2023

Acknowledgements

Publications

(*1*) Savva, L., and Platts, J. A. (2023) Exploring the impact of mutation and post-translational modification on α-Synuclein: Insights from molecular dynamics simulations with and without copper, *Journal of Inorganic Biochemistry 249*, 112395.

(*2*) Savva, L., and Platts, J. A. (2023) How Cu(II) binding affects structure and dynamics of α-synuclein revealed by molecular dynamics simulations, *Journal of Inorganic Biochemistry 239*, 112068.

(*3*) Savva, L., and Platts, J. A. (2022). Evaluation of implicit solvent models in molecular dynamics simulation of α-Synuclein. *Journal of Biomolecular Structure and Dynamics*, 1-16.

(*4*) Al-Shammari, N., Savva, L., Kennedy-Britten, O., and Platts, J. A. (2021). Forcefield evaluation and accelerated molecular dynamics simulation of Zn(II) binding to N-terminus of amyloid-β. *Computational Biology and Chemistry 93*, 107540.

Abstract

The aggregation of proteins has long been implicated in the pathogenesis of neurodegenerative disorders, such as Alzheimer's and Parkinson's diseases, through their deposition in amyloid plaques and Lewy bodies. The interaction of metal ions with these proteins has attracted significant attention due to their potential role in accelerating protein aggregation and neurotoxicity. In this thesis, Amyloid-β (Aβ) and α-Synuclein (αS) were studied using molecular dynamics (MD), to investigate the effect of metal ions on their structure and folding. Given the wide array of force fields available, the first part of this thesis focused on the evaluation of force fields and solvent models in simulating the average structure of Aβ_{16} in complexation with Zn(II), derived from an NMR study. The parameterisation of the metal ion and coordinating atoms was performed using quantum mechanic (QM) calculations on the metal-binding site (His6, His13, His14, Glu11), and incorporated into the force field to allow for the description of the metal ion and coordinating residues. The conformational landscape explored during the MD was expanded using accelerated MD (aMD), through the introduction of an energy bias to permit the crossing of energy barriers. The simulations revealed the ff14SB force field with the GBSA implicit solvent model to be the most accurate in reproducing the experimental structure.

The parameterisation described above was thus applied to a more disordered system, looking at the coordination of Cu(II) to αS. The simulations revealed that the force field was less ideal in reproducing the experimental characteristics of the protein, with better representation instead coming from ff03ws with the OBC continuum model. The aMD simulations revealed that the Cu(II) coordination to αS increased the stability of β-hairpins, while decreasing the N-terminal helical content, which has the potential to increase the rate of secondary nucleation. The Cu(I) coordination to αS was also investigated, due to the copper ions' interconversion during the catalytic release of reactive oxygen species. The system's average structure was suggestive of an intermediary state between the Cu(II) and apo forms. Following that, a different way of simulating the metal ion was implemented, through the use of cationic dummy atom models, eliminating the need for pre-defined bonded interactions with the coordinating atoms. This allowed the calculation of relative binding affinities to the metal ion. The model was also applied to study the αS-dimer in the presence and absence of Cu(II). The simulations on these systems, suggests the metal ion is a stabilising factor in the aggregation of αS, facilitating the formation of β-strand interlinkages between the chains.

The last part of this thesis, looked at two of the modifications often described in PD patients, in particular the phosphorylation at S129 (pS129) and the A53T mutation. The former systems suggested a protective effect to the aggregation of the protein, while the A53T mutation, especially in the case of the Cu(II)-bound system, presented longer-lasting β-characteristics, which could be indicative of a more stable aggregation with other peptides. Taken together, the results provide an understanding of the structural changes elicited by the association of these metal ions with the proteins, along with their influence on the aggregation process.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

The primary scope of this thesis was to investigate how metal ions affect the aggregation capacities of two peptides involved in two of the most common neurodegenerative diseases: amyloid-β (Aβ) in Alzheimer's Disease (AD) and α-Synuclein (αS) in Parkinson's Disease (PD). Initially, an assessment was performed on the efficacy of computational methods to provide results comparable to an experimentally-defined structure of the N-terminal Aβ (residues 1-16), bound to Zn(II).[1] Here, ff14SB with GBSA implicit solvent provided results most comparable to experiments, from the force field – solvent combinations tested. Following that, αS was simulated in its unbound and Cu(II)-bound forms, after evaluating different force field-solvent combinations against structural characteristics observed in experimental studies.[2] The Cu(II)-bound form of the peptide, was found to increase the stability of β-hairpins, believed to increase the formation and stability of multi-chain αS assemblies. The wild-type (WT)-αS was then modified, to examine the A53T-mutated and phosphorylated-S129 systems, again both in the unbound and copper-bound systems. The findings suggested increased aggregation of the A53T-mutated form, especially when Cu(II) was coordinated with αS, while the pS129 systems did not provide any significant differences to the structures observed in the WT-form. Acknowledging the susceptibility of the peptide to Cu-catalysed oxidation, resulting in the release of reactive oxygen species (ROS), the Cu(I)-αS was also investigated. The aforementioned assessment of the structural alterations from the coordination of the metal ions was performed at experimentally defined sites, through the implementation of harmonic M-L bonds. This limits the energetic information that can be recovered on the metal ion and the ligating atoms. For the evaluation of the binding affinities of the metal ions in each of the α-Synuclein systems, steered molecular dynamics (SMD) simulations using a cationic dummy atom model (CaDAM) of Cu(II) allowed the recovery of the free energy landscape and the interaction kinetics of those systems. Lastly, the dimeric αS system, both in its unbound and Cu(II)-bound forms, was examined using conventional all-atom MD simulations. Here, we find an increased stability of dimers bridged with Cu(II), suggesting a contributing effect to the association and conservation of interactions between αS chains. Overall, the results presented in this thesis, expand on the work performed on these systems, by offering an understanding of the folding mechanisms that contribute to this increase in aggragation we observe from experimental evidence. The effects of metal binding with these two proteins, point towards a stabilisation of the folding characteristics, aiding in the association of the monomers with other chains. Taken together, the results suggest that the metal-free proteins moderate their aggregation into deposits, as opposed to the metal-bound monomers, which are believed to accelerate motor deficit. In this chapter, a

detailed breakdown of the literature background on the aforementioned simulations is given.

1.2   Proteins

Proteins are macromolecular substances comprised of amino acids, which constitute an essential building block of life. Shorter sequences of amino acids, may be referred to as peptides. The coupling of amino acids occurs through peptide bonds, which are amide linkages formed by the condensation of the carboxylic acid group of one amino acid with the amino group of another. Given the plethora of biomolecular assemblies that are possible from the combination of these amino acids, and their varying activity within living organisms, it comes as no surprise that the way these proteins may be affected, both from biological and synthetic substances, is of great importance to the scientific community, driving studies on ways to aid their regulation and predict their behaviour, or response to extraneous factors. In this work, Amyloid-β (Aβ) and α-Synuclein (αS) are studied – the former derived from the amyloid precursor protein (APP), after it has been cleaved by β- and γ-secretase,[3] while αS is encoded by the SNCA gene.

There are twenty primary amino acids that make up protein sequences; their structure and key characteristics are given in Figure 1.1. The N- and C-termini of these amino acids allow the formation of peptide bonds between them. The variation in the sidechain functional group they carry, dictates some of their properties such as polarity and charge.
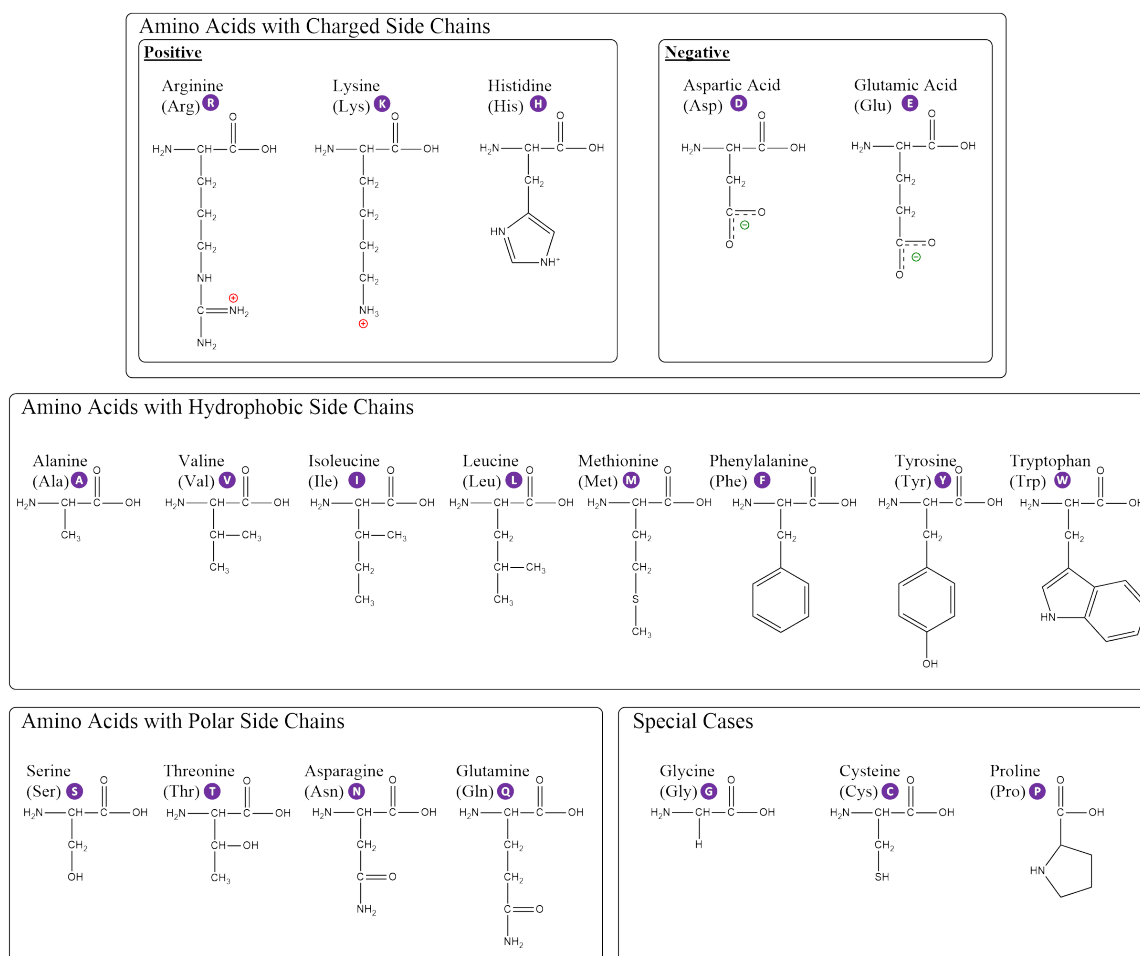
Figure 1.1: The 20 primary amino acids, with their 3- and 1-letter codes.

1.2.1   Secondary Structure

Proteins tend to fold into various structures, which often dictate their functions. These secondary folding elements can be identified by considering the backbone dihedral angles, stabilised by the presence of hydrogen bonds.[4] The principal dihedral angles, presented in Figure 1.2 and redefined here, are: ψ (N-C$_\alpha$-C-N), φ (C-N-C$_\alpha$-C), and ω (C$_\alpha$-C-N-C$_\alpha$). The angle where the plane created by the first three atoms is traversed by the plane formed by the last three atoms, is the dihedral angle.
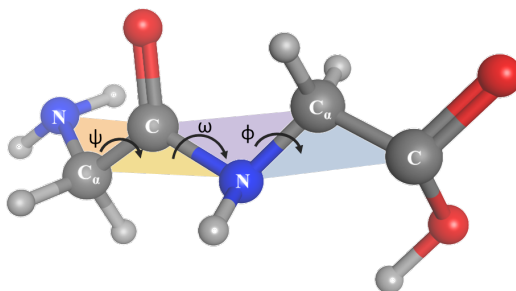
Figure 1.2: Dihedral angle representation, illustrated on Glycylglycine.

The analysis of the secondary structure of biomolecules, yields information about the α-helices and β-sheets, which predominantly dictate a protein's folding. These, along with the less structured characteristics, are summarised in Table 1.1:

Table 1.1: Summary of the principal secondary structures found in biomolecules.

α-helices[5]	α-helix 3_{10}-helix π-helix	Helical structures are separated into three main categories. These are stabilised by intrachain hydrogen-bonds, cross-linking carbonyl and amine groups. Specifically, the carbonyl group of residue *i* with the amine group of residue *i+3* (3_{10}-helix), *i+4* (α-helix) or *i+5* (π-helix). A representation of the interactions, which form the helices is shown in Figure 1.3 and Table 1.3.
β-sheets[6]	Parallel	β-strands hydrogen-bond with each other in a parallel direction and are often arranged in triads around the helical axis, Figure 1.5.
	Antiparallel	β-strands alternate directions successively, resulting in the N-terminus of one strand being adjacent to the C-terminus of the next, Figure 1.5.
Other	Turn	Chain reverses its overall direction.
	Bend	Sole non-hydrogen-bond conformation.
	Coil	Conformation that doesn't fit in any of the other categories.

(A)

$i,i+4$ H-bond

| 0 CONH | 1 CONH | 2 CONH | 3 CONH | 4 CONH | 5 CONH | 6 CONH | 7 CONH | 8 CONH |

C_α 1 C_α 2 C_α 3 C_α 4 C_α 5 C_α 6 C_α 7 C_α 8

(B)

$i,i+3$ H-bond

| 0 CONH | 1 CONH | 2 CONH | 3 CONH | 4 CONH | 5 CONH | 6 CONH | 7 CONH | 8 CONH |

C_α 1 C_α 2 C_α 3 C_α 4 C_α 5 C_α 6 C_α 7 C_α 8

(C)

$i,i+5$ H-bond

| 0 CONH | 1 CONH | 2 CONH | 3 CONH | 4 CONH | 5 CONH | 6 CONH | 7 CONH | 8 CONH |

C_α 1 C_α 2 C_α 3 C_α 4 C_α 5 C_α 6 C_α 7 C_α 8
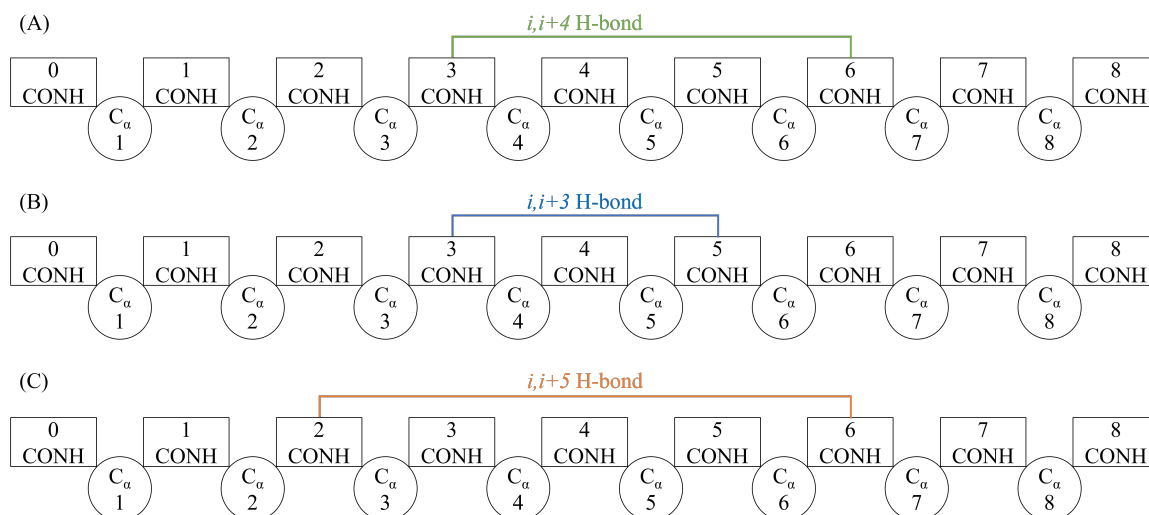
Figure 1.3: Representations of (A) α-, (B) 3_{10}- and (C) π-helices. Circles and rectangles are used to represent backbone-C ($C\alpha$) and peptide groups. Schematic adapted from original source.[5]

(A)   (B)   (C)
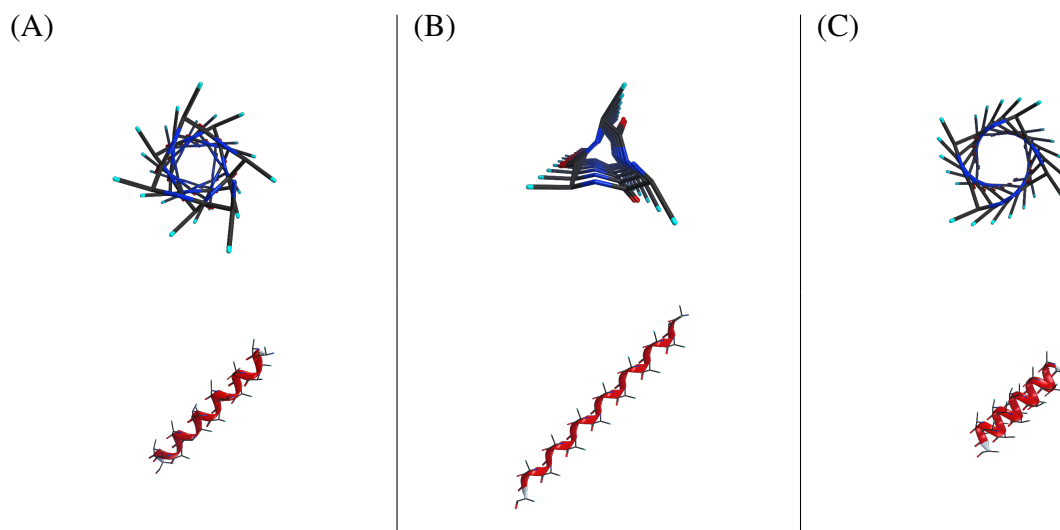
Figure 1.4: 3D-illustrations of (A) α-, (B) 3_{10}- and (C) π-helices, top and side views.
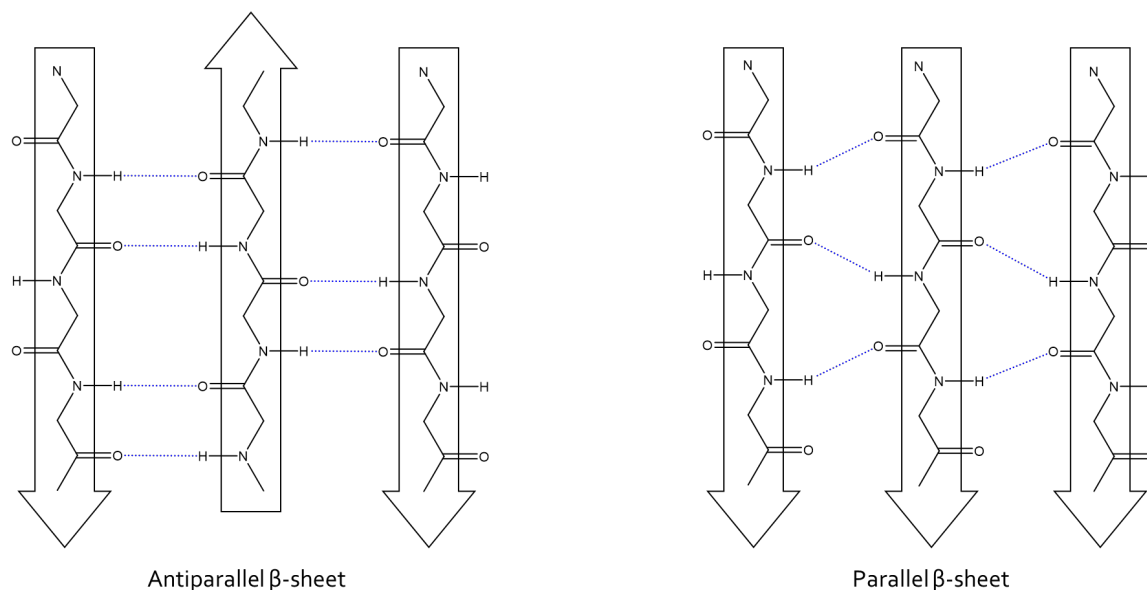
Figure 1.5: Antiparallel and parallel β-sheet arrangement, with the hydrogen bonds (blue dotted line) between each strand, characteristic of the two geometries.

1.2.2   Energy Landscape Theory

The protein folding mechanisms have long attracted the attention of scientists studying biological systems.[7] This interest arises from the proteins' activity related to their composition, but also in their ability to spontaneously fold into their native structure.[8] The energy landscape theory aims to investigate the mechanics behind the folding process of biomolecular substances, by considering the kinetic and thermodynamic control in the composition of these structures. The folding funnel hypothesis has been used to describe how a protein assumes its native structure, by reaching an energy minimum.[9] This tendency is closely related to Anfinsen's dogma,[10] which postulates that a protein's native structure is governed solely by the amino acid sequence, granted standard conditions are present. Also known as the thermodynamic hypothesis, this was developed from research on the folding of ribonuclease A, which showed that the protein's native state was reached spontaneously, by the system collapsing down to its energy minimum, without any intervention from other biological systems.[11] This proposal has, understandably, received criticism for its exclusivity in later years, with the discovery of chaperone-mediated protein folding,[12] and the highly unpredictable structures of intrinsically disordered proteins.

Levinthal's paradox,[13] perfectly illustrates the infeasibility of a protein folding into its native state by randomly sampling conformations in the phase space, if the predicament that a protein's native folding state is governed by the amino acid sequence is held true. Levinthal showed the dimensionality of this problem, even if only the dihedrals are considered, would require an unrealistically long time to settle on the true native state; thus there must be an energy bias towards the native state that explains the ability of proteins to fold within the very short time-scales seen in nature.[14] Levinthal essentially proposed that protein folding is determined by kinetics, rather than thermodynamics, meaning a protein's native state is not necessarily the global lowest energy state (i.e. most stable), but rather the one that is most likely to be reached by the system.[15] Computational experiments disagree with that approach, where thermodynamic control is favoured and the native state is regarded as the lowest energy one. This is in

line with the folding funnel models that have more recently been developed. As the protein collapses to its native state, during the folding process, a reduction in entropy is observed. The term "rugged funnel," has been used to describe the tendency of proteins to adopt their native structure, owing to the competing interactions between the biomolecule and surrounding solvent.[16, 17] Granted, these are often quite irregular owing to the presence of non-native local minima (kinetic trap in Figure 1.6), where the system may spend time. Enhanced sampling techniques have been implemented as a way to circumvent those, and push the system out of those wells, so that more of the conformational space can be sampled, increasing the possibilities of accessing lower energy minima.
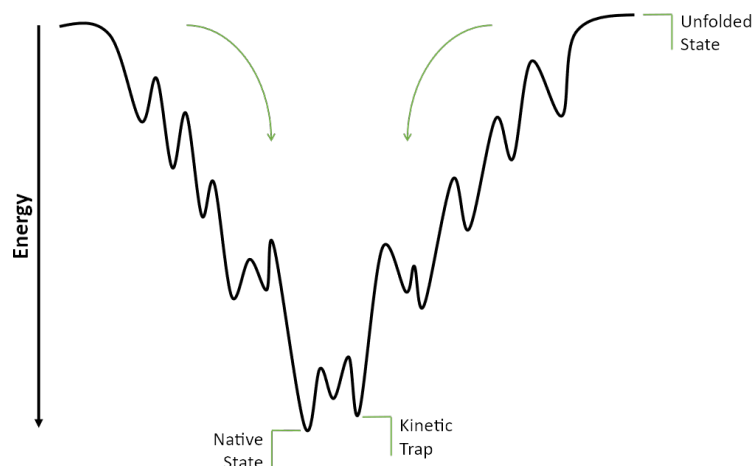
Figure 1.6: Schematic of a funneled energy landscape, with the native state being the global minimum. Kinetic traps are also present along the way, where the system may spend time in, as these may be separated by high energy barriers.

1.3   Protein Folding

Proteins display a diverse range of functions depending on their assembly and folding. Arguably, the most daunting task in molecular biology is the attempt to reverse-engineer the folding process, so as to understand the mechanisms that govern the formation of the native state of a protein (i.e. the state that is most stable and functional).[18] There have been several ventures into hypothesizing a solution to the 'protein folding problem,' with the most recent attempts being the development of machine learning algorithms.[19] The seemingly stochastic nature of the folding process, along with the highly complex folding landscape, described above, makes the problem of protein structure a very challenging one. More recently, AlphaFold has received a great deal of attention, on its potential to predict the native structure of proteins. Although, considering the highly dynamic nature of some peptide systems, the predictions obtained from a machine learning model, must be treated with care. Considering models such as this, have been trained on pre-existing database structures, that may not accurately describe the conformation of the protein in different environments, some predictions, such as in the case of α-Synuclein, fail to accurately predict the structure of the protein in solution. Nevertheless, estimates such as these can still provide a good starting point for MD simulations.[20]

The physiological and/or pathological functions of a protein are often determined by its folding state. Several neurodegenerative diseases have been linked to the misfolding of proteins

in the brain, with the most common being Alzheimer's and Parkinson's diseases, studied in this thesis. This misfolding can be caused by several factors, such as mutations in the primary sequence, environmental factors like temperature and pH, but also dysregulation of metal ion concentration in the brain.

Here, the folding process of the proteins, is examined using MD simulations, which despite limitation, pertaining to the time-scales that can be accessed, have been instrumental in the study of biomolecular systems.[21] Considering the years it takes for the development of neurodegenerative diseases, related to these systems, the nanoseconds of simulations performed here can only offer an insight into the early stages of the aggregation process. The overall manifestation of the diseases is a complex process, very much dependent on the auxiliary action of multiple proteins, along with how well the biological defences function, against the development, and towards the expulsion of the abberant forms of these proteins, which again depend on the genetic makeup of the individual.

1.3.1   Effect of Metal Ions

The involvement of metals in the folding capacity of proteins is well established within the community, with increasing evidence substantiating the importance of metal homeostasis in preserving the physiological functions of proteins and restricting the manifestation of pathologies associated with their aggregation into deposits.[22, 23] The accumulation of metal ions in the brain, has been shown to proceed through several mechanisms, including the impairement of ATPase transporter proteins, responsible for the transport and sequestration of copper ions in the brain.[24] The development of Wilson's disease, a genetic disorder that results in the accumulation of copper in the liver and brain, is a prime example of the importance of metal homeostasis in the brain, with an increased risk of developing a neurodegenerative disease being associated with the disease.[25] The metal hypothesis has thus been developed as an attempt to explain this phenomenon of metal accumulation and offer an understanding of the mechanisms in play, that result in an increase in the aggregation propensity of proteins, and ultimately seed the development of neurodegenerative diseases.[26] Initially, the hypothesis was centred around Alzheimer's disease and Amyloid-β, and the potential therapeutic opportunities that may emerge by targetting the dyshomeostasis of metal ions. The hypothesis has since been extended to other systems where metal deposition has been found to influence the misfolding of proteins associated with pathologies, such as Parkinson's disease and α-Synuclein.[27, 28] Other diseases not explored in the present work, but still associated with the metal hypothesis, include the amyotrophic lateral sclerosis (ALS), Huntington's and Prion diseases.[29]

A recent review into the encoded proteins exhibiting a Zn(II)-binding site has shown the prevalence of this metal ion in the human proteome.[30] In the majority of cases, this metal ion was found to coordinate with Cys, His, and Glu/Asp residues. Its coordination with Cys and His, in particular, has often been implicated in the formation of zinc finger motifs, which play a key role in fold stabilisation.[31] These motifs are not just structural elements, but also serve as interaction domains for other proteins, resulting in the formation of inter-peptide interactions. It's also worth noting that the specific geometry and coordination environment can vary depending on the protein and the biological context, further adding to the complexity of metal-protein interactions.

1.3.2 Reactive Oxygen Species

Besides the effect on the protein folding, described above, metal ions have also been implicated with the catalytic release of reactive oxygen species (ROS), triggering oxidative stress in tissues. This imbalance has been closely related to the release of neurotoxins, accelerating neurodegeneration by contributing to dopaminergic neuronal degradation.[32–34] These ROS are usually derivatives of oxygen metabolism, with hydroxyl radicals being the main initiators of oxidative damage. It is worth noting that the breakdown of other small molecules, can also result in the formation of these moieties. In the case of α-Synuclein, for instance, an increase in the uptake of dopamine (DA) has been associated with an accelerated production of ROS.[35] This, along with the reported interaction of DA with the C-terminal of α-Synuclein, Figure 1.11, has the potential to exacerbate neurodegeneration, as a result of the interaction between the two.

The elevated production of ROS, related to the presence of metal ions, is attributed to Fenton-like reactions through a redox mechanism, as the one shown in Figure 1.7 for Cu(II)/Cu(I),[36, 37] with hydrogen peroxide produced as a by-product from biological oxidation reactions.[38]

$$Cu^+ \qquad Cu^{+2}$$
$$H_2O_2 \qquad \overset{\bullet}{O}H + OH^-$$
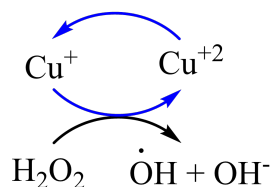
Figure 1.7: Catalytic release of ROS from hydrogen peroxide in a Fenton-like reaction.

1.4 α-Synuclein and Parkinson's Disease

Second to Alzheimer's disease (AD), Parkinson's disease (PD) is a significantly common neurodegenerative disease, with an increasing incidence with age.[39] Although PD is rare in individuals below the age of 50, early onset mutations can result in its development in younger populations.[40] The disease accounts for 15% of people diagnosed with dementia, exhibiting physical as well as neuropsychiatric symptoms.[41]

1.4.1 History of Parkinson's Disease

The main symptoms accompanying the disease, relating to the motor function of patients, were first described in 1817 by James Parkinson in his report "An Essay on the Shaking Palsy".[42] The anatomical substrate of the disease was not realised until over a century later. Edouard Brissaud, influenced by the publication of Blocq and Marinesco's case report of a person admitted with tuberculosis who displayed signs of unilateral Parkinsonism,[43] speculated a causal link between PD and the presence of ischemic lesions in the *substantia nigra*.[44] Brissaud based his hypothesis on the presence of a noisette (enucleated tuberculoma), discovered during the autopsy of that person, which was found to have selectively damaged the *substantia nigra*. Evidence supporting his theory came in 1919, when neuropathologist Konstantin Tretiakoff in his doctoral dissertation looked at the *substantia nigra* of 54 brains, of which 9 had PD and 3 post-encephalitic parkinsonism. In these cases, lesioning was identified in the *pars compacta* (SNpc) domain of the *substantia nigra*. In 6 of the PD cases he also identified the presence

of "corps de Lewy" (Lewy bodies), which are abnormal deposits of the protein α-Synuclein, named as an homage to Fritz Heinrich Lewy, who had discovered the presence of these spherical neuronal inclusions 7 years earlier.

The underlying mechanisms of the development of Lewy bodies, and the reason these result in motor behaviour deficits remained unknown until much later – starting with the discovery of dopamine (DA) in the late 1950s by Arvid Carlsson.[45] Not long after, in 1960, Oleh Hornykiewicz looked into the concentration of DA and norepinephrine in post-mortem brains of people with and without PD or post-encephalitic parkinsonism, and noticed selective DA depletion in the *putamen* and *causate nucleus*, in both forms of the disease.[46] He thus proposed that DA depletion was a result of loss of neurons in the midbrain. His prediction was confirmed a year later, when he rightly suggested the administration of L-Dopa (DA's precursor) as a possible treatment of PD symptoms.[47] Hornykiewicz's contribution in the understanding of the disease led to the development of new drugs for the treatment of PD, targeting an improved channelling of DA. The underlying genetic connection to PD was eventually discovered in 1997, in chromosome 4q22 – gene Synuclein Alpha (SNCA), with a G209A point mutation resulting in a A53T substitution in the encoded αS.[48] In the same year, the connection between Lewy bodies (LB) and αS was made,[49] where the latter was found to be the main, insoluble component of LBs. The A53T mutation was then confirmed to promote aggregation of αS, a year later.[50]
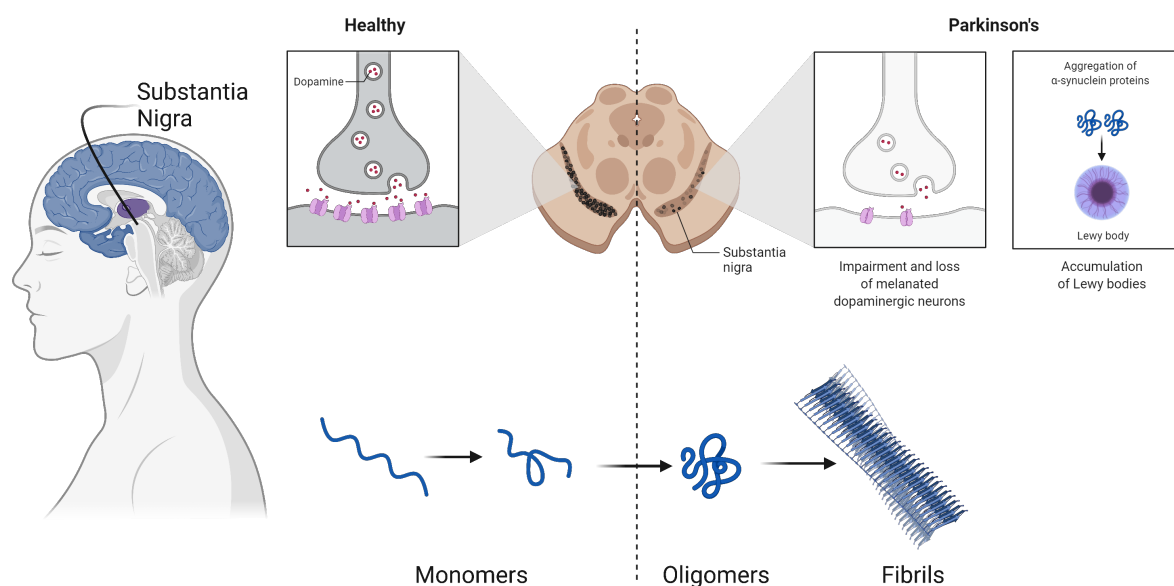
Figure 1.8: Schematic of the region in the brain affected by α-Synuclein aggregates. Created with BioRender.com.

1.4.2 Structure of α-Synuclein

α-Synuclein was first identified in 1988[51] as a protein that was present in the synaptic vesicles of the brain.[a] The Synuclein Alpha (SNCA) gene is found in region 4q22.1 (band 22 on the

---

[a]Note that in the present work, *protein* and *peptide* are used interchangeably to describe αS.

long arm of the 4th chromosome), and is responsible for encoding the 140-residue-αS protein. The system is generally partitioned into three main regions: the N-terminal (residues 1-60), the central non-Amyloid-β component (NAC) (residues 61-95), and the C-terminal (residues 96-140), Figure 1.9. The N-terminal region is characterised by a high content of amphipathic helices, as a result of residue repeats that extend into the NAC region, where they are responsible for the formation of β-strands. The NAC region is largely populated by hydrophobic residues, making up *ca.* 66% of the total in that region (8 Val, 7 Ala, 6 Gly, 1 Phe and 1 Ile), with the remaining consisting of 9 polar (5 Thr, 2 Gln, 1 Asn and 1 Ser) and 3 charged (2 Glu and 1 Lys) residues. The C-terminal contains the majority of the acidic residues in the peptide accounting for 64% of the total, Figure 1.10, and displays a generally unstructured character.[52]

The N-terminal is distinguished by the four 11-mer [EQS]-K-T-K-[EQ]-[GQ]-V-X_4 repeats, which constitute the region where the peptide assumes the most stable order of secondary characteristics – considering its inherently disordered nature, any region that is responsible for the formation of even transient well-defined structural elements, is significant. Within these repeats, β-hairpins and α-helices have been found to form, either in solution or as a response to lipid membrane interactions.[53–59] The NAC central region of the peptide, is arguably the most studied, considering its involvement, not only in inter-peptide interactions in PD, but also the co-localization of the NAC region with Aβ, accumulating in the extracellular senile plaques, described in the succeeding sections.[54, 60] The fibrillation propensity expressed here, follows the hydrophobicity of the residues composing the region.[61]

1		60	95	140
	N-terminal		NAC	C-terminal

^{1}MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKEGVVHGVATVAEKTK^{60}
^{61}EQVTNVGGAVVTGVTAVAQKTVEGAGSIAAATGFV^{95}
^{96}KKDQLGKNEEGAPQEGILEDMPVDPDNEAYEMPSEEGYQDYEPEA^{140}
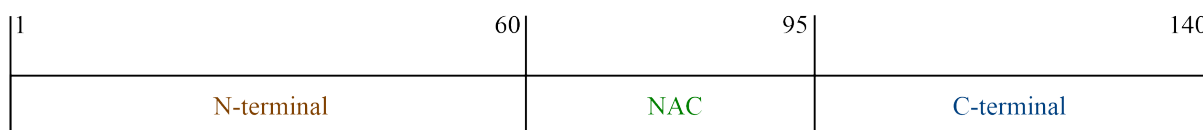
Figure 1.9: The three regions comprising αS.
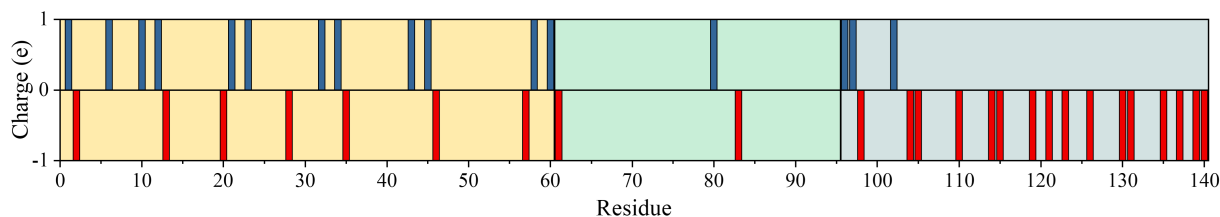
Figure 1.10: The charged residues present in the WT-αS system.

As alluded to earlier, the widely-accepted subcellular factor that underlies Parkinson's neuropathology is the presence of Lewy bodies with inclusions of aggregated αS.[62] Lewy body formation, involves "cross-seeding" of αS, from interaction with other peptide systems, such as the tau protein and amyloid-β. These intra-cytoplasmic build-ups mainly consist of αS, which is why this peptide has been the focus of studies looking to expose the settings in which

the disease thrives. Fibril inclusions of αS, have been earmarked as the primary form of the peptide occupying Lewy bodies. The fibrillation of this peptide has been associated with the presence of β-sheets, through which inter-peptide interactions take place.[63, 64] Figure 1.11 shows, some of the key locations assigned to specific functions of αS are presented – each of which will be addressed in the sections that follow.
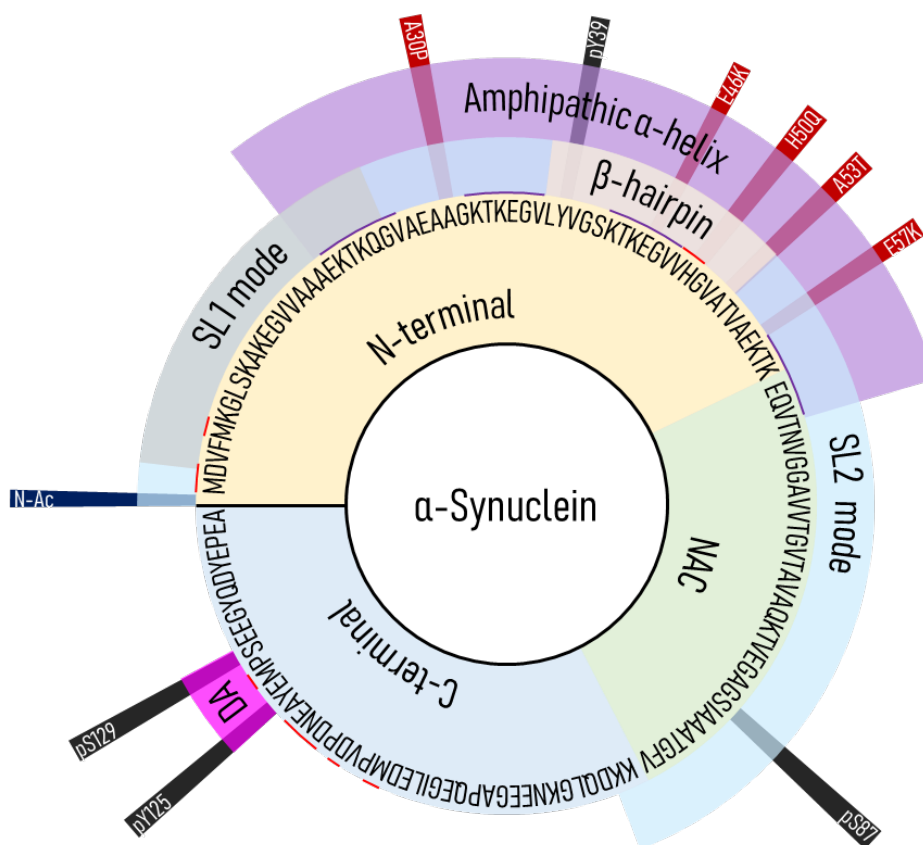
Figure 1.11: Schematic of some of the key localisation characteristics of αS. The lines above some of the amino acids represent the Cu(II) and Cu(I) binding sites (red) and the key repeating motifs (purple). DA = dopamine-binding region, NAC = non-amyloid-β component, SL1 = lipid-binding region 1, SL2 = lipid-binding region 2.

1.4.3   Function of α-Synuclein

Maroteaux et al., who had first described αS, speculated the possible function of the protein, which they thought to be related with the regulation of gene expression, and/or the control of synaptic events.[51] Shortly after the initial publication, the group provided evidence for the peptide's pre-synaptic function, as well as discovering the other two isoforms of synuclein, β- and γ-synuclein.[65] Now, some 25 years later we have yet to fully understand the function of

αS, but we do know that it is largely found in the pre-synaptic regions of the brain, where it has been implicated with synapse plasticity and neurotransmitter release.[66–69] The role of αS in neurotransmitter release has been argued in the scientific community, with some studies reporting an increased snap receptor protein (SNARE) formation, as a result of αS's association with v-SNARE VAMP2,[70, 71] and by extension an enhanced neurotransmitter release. Other studies, propose the impairment of such function as a consequence of the over-accumulation of the peptide.[72, 73]

1.4.4   Mutations and Post-Translational Modifications in α-Synuclein

The spatiotemporal heterogeneity of intrinsically disordered proteins (IDPs), such as αS, has been reported to be influenced by environmental factors, mutations and post translational modifications (PTMs), with function-related conformations with varied retention times depending on the peptide and interactions with binding partners.[74, 75] Promotion of fibrillation of αS has been proposed as a result of some of these factors, such as low pH, increased temperature and mutations, owing to changes in the structure resulting in effects such as a decrease in the diffusivity of the aggregates.[76, 77] In the present thesis, the changes to the wild-type (WT) form, which have been prominently associated with the onset of PD are examined. These include the first point mutation flagged in the early discovery of the peptide's association with PD, A53T,[48] and the post-translational modification reported to be present in 90% of diseased brains, pS129.[78–80] N-terminal acetylation (N-Ac) has also been associated with reduced fibrillation capacities of αS,[81] although the system has not been examined here.

Experimental studies on the nature of αS in patients suffering from synucleinopathies, have identified the presence of an alternate form of αS, phosphorylated at S129. Despite its 90% expression in patients with PD, as opposed to the 4% occurrence in normal brains,[78–80] studies on the effect of this form of αS in its ability to aggregate still have not settled on a definite conclusion. In spite of a report suggesting the promotion of fibrillation upon phosphorylation at Ser129,[78, 82] other studies argue the phosphorylated form may be an outcome of the disease itself, as a result of proteolysis impairment acting as a signal for degradation,[83, 84] with no correlation to its ability to form fibrils, or in certain cases inhibiting their expression altogether.[83–85] Notwithstanding the disagreement of studies on the impact of pS129 on the aggregation capacities of αS, it has been shown that oxidative stress induces phosphorylation at S129.[86] Recognizing the effect of copper coordination to αS, catalysing the production of reactive oxidative species (ROS),[87] as well as experimental evidence suggesting an increase in the binding affinity of divalent metal ions upon phosphorylation of αS,[80, 88] the Cu(II)-bound pS129-αS was also simulated. Experimental studies on the copper-bound pS129-αS, also report higher binding affinity in the C-terminal binding site.[89–91]

Further phosphorylation sites exist in αS, covering all three regions of the peptide, Figure 1.11. In the N-terminal, Tyr39 has recently been the focus of a multimethod research on the importance of the residue's aromaticity in the aggregation effects of membrane-bound αS.[92] Phosphorylation of that residue has been proposed to decrease the aggregation capacities of the peptide when in solution, while promoting helical membrane-bound forms,[93] as well as resulting in the formation of fibril polymorphs.[94] In the NAC region, Ser87 has been detected to phosphorylate *in vivo*, with an increased occurrence in synucleinopathies; a feature also observed in pS129.[95] The pS87 modification is believed to inhibit fibrillation of αS, although not enough evidence exists to arrive to a categorical conclusion. Lastly, in the C-

terminal, second to arguably the most studied pS129, Tyr125 has been the focus of several studies, owing to its close proximity to pS129, and the known small molecule binding features of that region.[96] The phosphorylation of this residue was found to inhibit aggregation,[97] while it has also been suggested to occur alongside pS129, and aid in the aggregation process of the peptide.[98] Epitope mapping was more recently used to evaluate the incidence of this PTM in diseased brains, where it was found to not constitute a significant enough percentage of the forms of αS present in diseased brains.[99] Some additional phosphorylation sites still exist, such as Y133 and Y136, but not addressed here. From the evidence presented here, it is apparent that the phosphorylation of αS is a complex process, with many sites available, each with the potential to alter the functions of the peptide. Following the literature survey of the available studies, it has become quite evident that the characterisation of these effects is very much in its infancy, with many studies offering speculations of the mechanisms at play, from inconclusive, or conflicting results.

Thus far, eight natural point mutations have been identified in the SNCA gene,[100] with the aggregation process shown to accelerate through at least five of those mutations. The A30P mutation has been found to decrease the kinetic conversion of oligomers to fibrils, due to a rapid formation of amorphous aggregates and their increased retention times.[101–103] One of the key point mutations exhibited in the SNCA gene is A53T, which has been shown to increase the formation of protofibrilar intermediates.[104, 105] This point mutation was first found to exist in familial-PD patients of Greek and Italian descent;[48, 106] while later the mutation was also characterised in patients from Sweden[107] and Korea.[108] The H50Q mutation has also been found to correlate with an increased aggregation of the peptide *in vitro*, and an increased rate of β-sheet-rich oligomer formation, compared to WT.[109, 110] Along with the A53T and H50Q mutations, E46K has also been characterised by an increased rate of fibrillation, *in vitro* and *in vivo*, along with an increased stability of the resulting fibrils,[111–114] as a result of destabilisation of α-helices in the N-terminal and NAC.[115] The G51D-mutated protein has been shown to aggregate slower than the WT form, while displaying a decreased membrane affinity *in vitro*,[116] and an increased helical content of the oligomers, compared to the ones in the WT form.[117, 118] The same impaired binding affinity has also been found in A30P forms of αS,[118, 119] possibly owing to the increased formation of amorphous aggregates, discussed above. In addition to the naturally-occurring mutations in αS, several artificial mutations have been introduced to the peptide, to aid in the understanding of the role of each residue in the aggregation process. The E57K mutation is an example of this, which was found to increase membrane interference *in vivo*, resulting in an increased toxicity and decreased fibrillation rate.[120, 121]

The function of N-terminal acetylation of αS has been unclear thus far, partly due to its expression in both normal and diseased brains. The high occurrence of this form of synuclein, is possibly a result of the apparent lack of N-terminal deacetylases, making it a potentially irreversible modification.[122, 123] The general consensus among studies of N-Ac-αS, is the increased stabilisation of the helical characteristics in the N-terminal region of the peptide,[124, 125] which has been shown to be a key region for membrane interactions,[126] but also may result in an increased homogeneity of fibrils.[127] N-terminal acetylation has been proposed to increase the formation of a fibrillation-inhibiting, membrane-binding promoting helix between residues 1-12, as opposed to fibrillation-promoting helix between residues 14-31.[125, 128, 129] Evidence suggests an increased affinity to lipid membranes,[126, 130] this is not to say a definite conclusion, as arguments exist that the increased affinity is seen from the facilitation

of interactions with binding partners, and not as a direct a result of N-Ac.[127] Despite the evidence for an increased binding affinity to membranes upon acetylation, the peptide has been shown to display decreased aggregation rates in spectroscopic studies.[81, 124, 128, 131] This brings into perspective the possible interconnection of N-terminal helical formation, membrane binding and aggregation propensity.

Considering the evidence that exists for the increase of oligomer formation upon membrane binding, an additional link is established on the influence of membrane-binding and oligomer-formation on the aggregation rate.[132, 133] The mutation studies, described above, provide many examples of structural and functional evidence that can be used to study the aggregation mechanism. These include: (1) the decrease in the conversion of oligomers to fibrils, as a result of the increased formation of amorphous aggregates, seen in the A30P mutation,[102, 103] along with the decreased membrane affinity,[119] (2) the increased aggregation rate of the H50Q form, along with the accelerated production of β-sheet-rich oligomers,[109, 110] (3) the decreased aggregation rate as a result of the increased helical content of the oligomers, seen in the G51D mutation,[117] and (4) the decreased rate of fibrillation, as well as the increased affinity to lipid membranes, in the E57K mutation.[120] The increase in the rate of fibrillation, as a result of membrane binding regulation, has also been shown in the past through experimental studies on the WT-αS.[134–136]

1.4.5   Aggregation of α-Synuclein

Under physiological conditions, αS normally exists in an equilibrium between the monomeric and membrane-bound forms,[137] or in a tetrameric form which resists the formation of the aberrant aggregates.[138] Concerning the aggregation of biomolecular systems, one usually looks for a nucleation-elongation process,[139] whereby the polymerisation of the monomeric chains is commenced through the formation of nuclei in the *lag phase*, containing unfolded and partially unfolded chains, in a process referred to as primary nucleation.[140] This initial nucleation step is followed by a growth phase, *elongation*, through the addition of further monomers to the growing chains. A recent look at the aggregation mechanism, revealed a further step in the process, where the newly formed fibril surfaces catalyse the formation of new nuclei from pre-existing filaments.[141] This process is referred to as secondary nucleation, and is thought to be the chief accelerator for the formation of fibrils. The mechanism believed to drive this secondary nucleation is a filament fragmentation process, thus relating the rate of fragmentation to the formation of aggregates.[142] This aggregation mechanism is illustrated in Figure 1.12. In the previous section, the off-pathway formation of amorphous aggregates, was found to decrease the formation of fibrils, but at the same time increase neurotoxicity.[143, 144] This is in line with the mechanism seen in other systems, such as Amyloid-β, where similar aggregates have also been suggested to modulate fibril formation, although for this system a reduced toxicity was observed as a result of these structures.[145]

Of note is that the mechanism described here, has recently also been shown to undergo a liquid-liquid phase separation (LLPS), mediating the initial nucleation.[146–148] LLPS involves the formation of liquid-droplet inclusions of protein and/or RNA, resulting in the formation of membrane-less organelles (MLOs), which have been associated with various biological functions, such as regulation of gene expression and response to stress stimuli.[149] These condensates have a high content of intrinsically disordered proteins/regions (IDPs/IDRs) and proteins with stretches of low complexity regions (LCRs).[150, 151] The pre-nucleation MLOs have

been shown to decrease the nucleation barrier, resulting in the ready formation of heterogeneous nuclei, as opposed to their more enthalpically demanding homogeneous counterparts in the classical mechanism.[152]
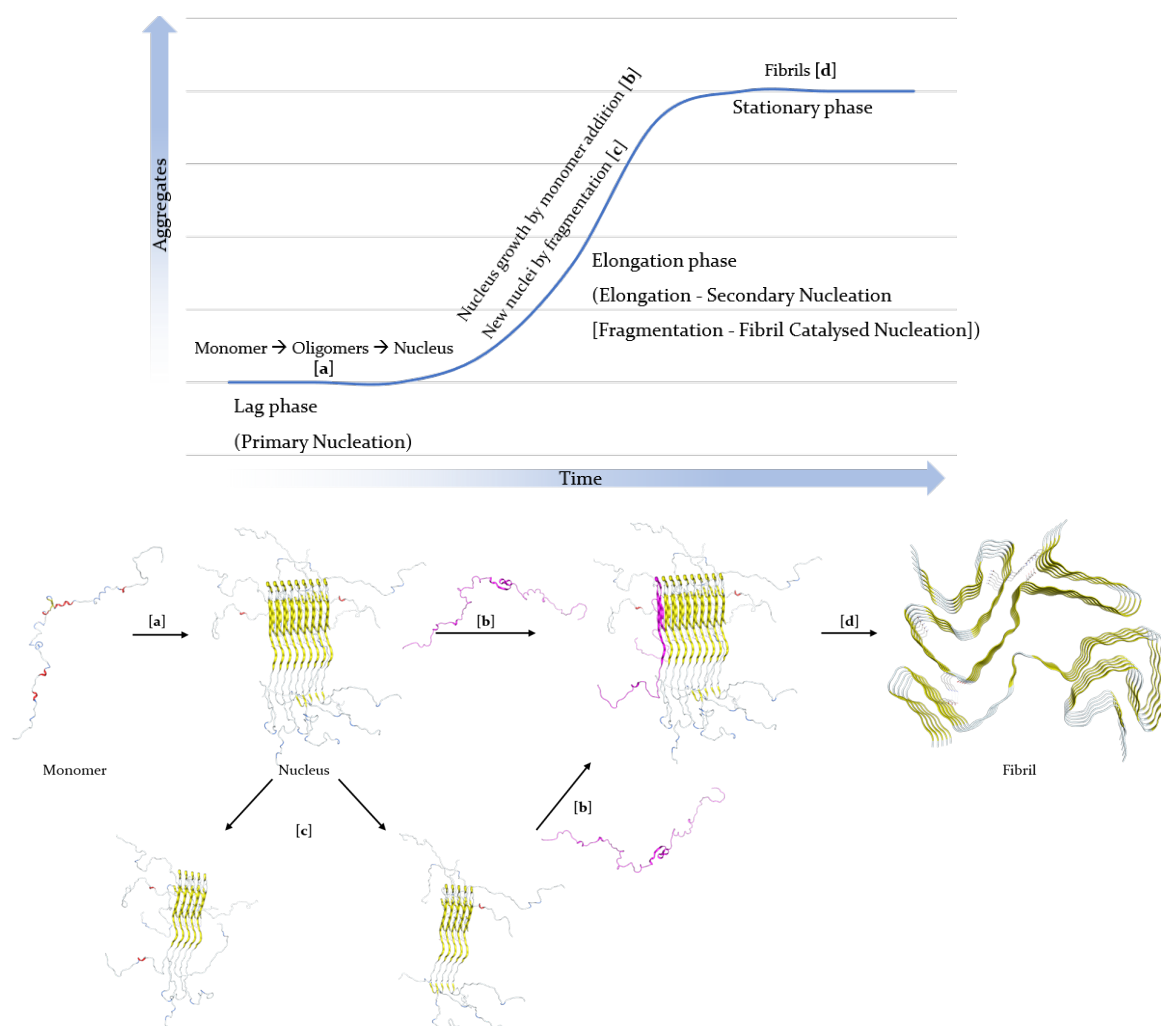
Figure 1.12: Plot of the stages in the classical aggregation mechanism, with a pictorial representation of them. Step [a] involves the primary nucleation of monomers to nuclei, followed by elongation of the nuclei in step [b], through secondary nucleation (i.e. addition of monomers to the growing nuclei). Step [c] still may occur during the elongation phase, whereby nuclei may fragment, resulting in the addition of monomers to multiple nuclei, accelerating the elongation process. Step [d] involves the formation of fibrils from the elongated nuclei, and the association of fibrils together. The structures of nucleus and fibril were obtained from the PDB database (PDB IDs: 2N0A[58] and 6XYO[153]), and just serve to illustrate the stages – not necessarily representing the actual structures.

There are several Cryo-EM structures of αS fibrils in the literature, such as:

1. PDB ID: 6A6B – fibril from residues 37-99, where the chains associate through the formation of left-handed α-helices;[154, 155]

2. PDB ID: 6CU8 – full-length αS fibrillar polymorph, where β-arches were found in the two interacting protofilaments, with inter-fibril interfaces between residues 47-56 and 68-78;[156, 157]

3. PDB ID: 7UAK – A53E αS fibrils, where lower stability was found compared to the WT, A53T and H50Q fibrils.[*158*, *159*]

1.4.6   Membrane Binding

Amphipathic helices, particularly in the N-terminal of αS, have long been thought to mediate the coordination of the peptide to lipid membranes.[*160*] Two key domains have been implicated with this facilitation, labelled in Figure 1.11 as SL1 and SL2, encompassing residues starting in the N-terminal and extending up to Gly25 and Lys97, respectively.[*133*, *161*] As mentioned in the previous section, the protein naturally exists in an equilibrium between the monomeric and membrane-bound forms;[*137*] thinking about the potential implications both in the pathological and physiological functions of the protein, one is compelled to consider how structural changes to the WT-form impact this equilibrium. Assessing the effects of membrane-coordination is certainly not an easy feat, considering how chemical properties of different lipids may affect this interaction,[*162*] as well as the subsequent protein folding.[*118*] Mutation and PTM studies offer a way to assess the effect of membrane interactions on the aggregation of αS. An example of this comes from a study examining the membrane affinity and aggregation capacities of several different mutations of αS, including A30P and G51D, discussed above, where a lower binding affinity to lipid membranes was reported, compared to WT, along with a reduced helical formation upon interaction and increased aggregation propensity.[*118*]

1.4.7 α-Synuclein and Transition Metal Ions

Transition metal ions have been observed to affect fibril formation, in a similar way to their coordination with Aβ.[*76*, *89*, *163–165*] Research on the influence of metals on the onset and progression of PD have gained traction over the last 20 years, Figure 1.13, with the majority of research, evidently conducted on the role of iron. The sudden increase in interest around 2016, could potentially be a result of ongoing clinical trials conducted on iron-chelation therapy as a potential treatment for PD,[*166*] with the aim of regulating iron overload. αS has been suggested to aggregate intracellularly as a response to divalent metal ions (Fe(II), Mn(II), Co(II), Ni(II) and in particular Cu(II)), bound to the two termini, with the removal of either coordinated metal resulting in disruption of aggregation.[*167*, *168*]

As discussed in previous sections, iron has closely been linked with the stimulated release of ROS, and the subsequent oxidative damage to cells. Despite the lack of a causal link between the accumulation of iron in the brain and the onset or progression of PD, it is quite well established by now that diseased brains display a higher iron content than healthy ones.[*169*, *170*] Figure 1.14, shows the differences in iron-related signal in the putamen of a healthy and diseased brain, acquired by magnetic resonance imaging (MRI).[*169*] The majority of metal ions found to bind to αS trigger the aberrant form of the protein – this is not the case with magnesium, which has been found to exhibit a reduced presence in diseased brains.[*171*, *172*] Regardless, the focus of the present work is on the role of Cu(II) in the aggregation of αS, and the potential for this metal to accelerate the fibrillation process, as it is the metal ion that has been found to exhibit the greatest affinity to the protein.[*173*]
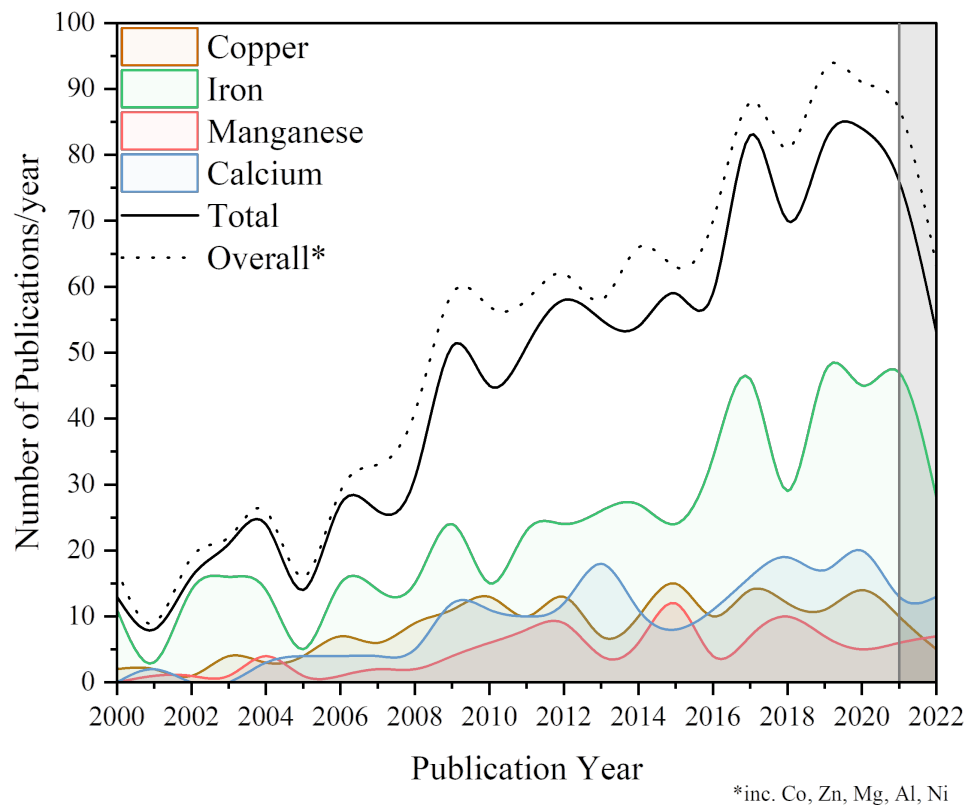
Figure 1.13: Number of publications per year with titles relating to Parkinson's disease or αS, and copper, iron, manganese and calcium ions (source: Web of Science; accessed: 25-01-2023).
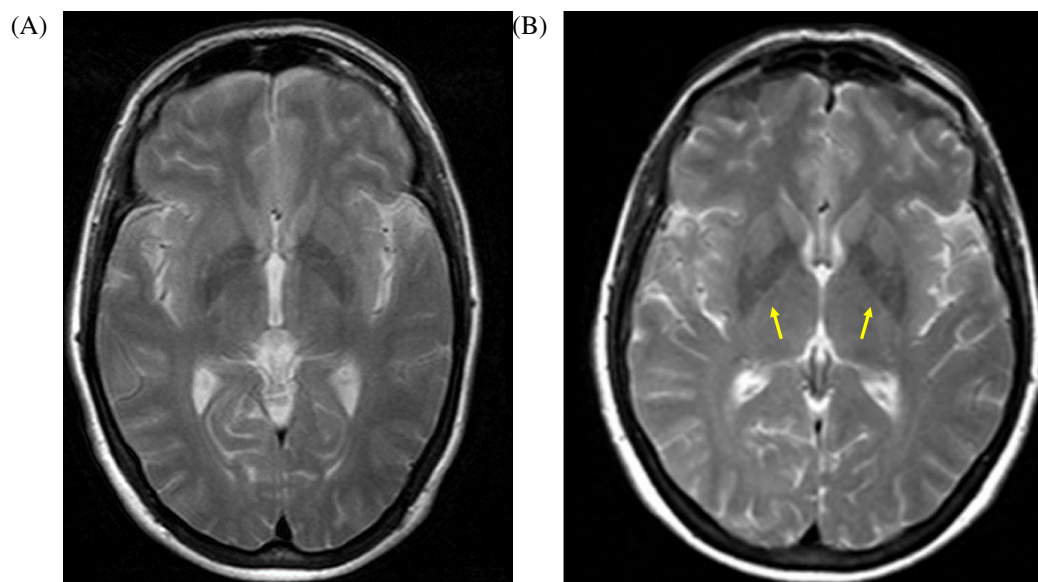
Figure 1.14: T_2-weighted scans of the iron-related signal loss in (A) a healthy brain and (B) the brain of a person with PD. The arrows point to the putamen of the diseased brain. Figure adapted from source.[169]

Some reports have been published on the role of metal ions in the promotion of free-radical mediated oxidative processes,[167, 174, 175] which have been associated with the overproduction of peptides, such as in the case of Aβ, through activation of β- and γ-secretases (*vide infra*, Section 1.5.2).[176] Most reports on α-Synuclein – metal ion interaction, have focused either on the specific regions interacting with the ions,[27, 89, 163, 165, 177–183] or the possible self-oligomerisation and aggregation mechanisms involving the metal ions.[164, 167, 184, 185] Studies employing small-angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR), circular dichroism (CD) and electron paramagnetic resonance (EPR) spectroscopy, on metal ion-αS interactions, indicate coordination of metal ions with the regions comprising of [1]MDVFMKGLS[9], [48]VAHGV[52] and [119]DPDNEA[124].[89, 163, 178, 186]

Computational studies on the binding of Cu(II) to αS have been lacking in the community, with most of the published ones focusing on modelling the free peptide.[27, 183, 187, 188] Two QM/MM studies examined the coordination of Cu(II) in the N-terminus of αS, one of them looking at the M1-D2-H_2O binding site,[189] believed to result in the more stable Cu(II)-complex, owing to the formation of a (5,6)-joined chelate ring from (NH_2, N-, β-COO-);[190] while the other focusing on the V48-H50 region,[191] involved in the Cu(II)-coordination of the N-terminally acetylated αS. One of the computational studies researching the copper-bound peptide involved a fragment of αS, simulating the first 12 residues using temperature replica-exchange molecular dynamics (T-REMD), coordinating the copper ion on the first two amino acids and a water molecule.[182] The study found an increased retention of β-hairpins upon metal ion coordination – a feature believed to increase the fibrillation of the protein, by offering a stable secondary structure region where inter-peptide interactions can take place. Other computational studies on the full peptide have used coarse-grained molecular dynamics (CG-MD), through scaling seen in the ff03ws force field applied in SIRAH,[181, 183] and *ab initio*[27, 180] and MD simulations using the CHARMM27 in explicit solvent.[27] These studies have highlighted the high affinity of Cu(II) coordination to Asp121 and His50, a feature also seen through the *ab initio* calculation of force constants, as detailed in Chapter 4, and binding free energy simulations, given in Chapter 5. These increased affinities to the two residues, have also been expressed in experimental studies, through spectroscopic experiments, suggesting a contributing effect to the coordination with the metal centre.[163, 192]
In view of the involvement of copper ions in the oxidative stress within proteins, through the redox-release and accumulation of reactive oxygen species (ROS),[167, 193, 194] this study would not be complete had I not considered the effect of monovalent copper ions on αS. This, understandably, changes the docking site for the metal ion, now coordinating with the sulfur atoms in Methionine (site 1: Met1 and Met5; site 2: Met116 and Met127).[195, 196] The effect of this reduction of Cu(II) to Cu(I) has driven research into the effect of this transition on the release of said ROS, although the influence on the protein's structure has been less explored. Experimental studies looking at this system have identified a decrease in the binding affinity to lipid membranes,[197] as a result of the oxidation at the Met residues, hypothesizing an increased accumulation of cytosolic αS that can nucleate.[198]

1.4.7.1   Metal Coordination

Further research into the specific binding modes of Cu(II) and the aforementioned sites, in the WT-form, proposed one of those to consist of macro-chelation (i.e. interaction of the metal ion with distanced residues) between residues M1, D2 and H50, and a second site encompassing D119, D121, N122 and E123.[163, 165] Of note is that Cu(II)-coordination at the N-terminus

has also been proposed to occur through either coordination with M1, D2 and H_2O, or V49, H50 and H_2O.[*199, 200*] The former has also been thought to exist in the membrane-bound form of αS, while the latter has been linked to the acetylated form of αS.[*201, 202*] In its monomeric WT form, however, these near-range binding modes have been put into dispute from electron spin-echo envelope modulation (ESEEM) spectroscopy studies,[*192, 203, 204*] thus, here, even though all the aforementioned N-terminal binding modes are considered, the majority of the focus is on the more established macro-chelated coordination mode (Chapter 4).[*87, 178, 188, 192, 203, 204*] This coordination mode has also been proposed to happen in an interpeptide fashion,[*205*] studied here on the dimeric system, Chapter 5. The coordination of the first site occurs in a 3N1O fashion, while the C-terminal binding involves a 4O coordination mode.[*27, 163, 165, 206*] One of the anchoring residues for metal ions, especially Cu(II), binding on the peptide chain is H50. In this context, it is notable that a H50A mutant results in quite different aggregation profile on coordination with Cu(II) compared to the wild type.[*178, 191, 207*]

1.4.8   Targeting α-Synuclein

Several approaches to addressing the accumulation of α-Synuclein in the brain have been proposed, with a significant amount of those undergoing clinical trials, at the time of writing. Categorising some of those therapeutic approaches based on the mechanism of action, aids towards a better understanding of the current state of the field.

Starting off with the most conventional approach – antibodies to target αS – two methods are currently being explored: passive and active immunization. The difference, between the two, is that the former involves the administration of antibodies to reduce the accumulation of the protein, while the latter involves the administration of antigens to stimulate the immune system to produce antibodies against it. Thus far, no successful candidate has been identified, although several are currently undergoing clinical trials.[*208–211*] Perhaps a more promising approach to target αS is through the use of small molecules, that either work to promote its clearance from the brain, or to inhibit its aggregation. One such a molecule is Anle138b,[*212*] which is currently undergoing its second phase 1 clinical trial.[*213*] The supposed neuroprotective effects of Anle138b have been attributed to its ability to modulate misfolding and the development of toxic oligomers in the brain, not only of αS, but also of Tau and Amyloid-β proteins.[*212, 214*] The specific mechanism of action and binding site of Anle138b is still a matter of discussion, with studies suggesting its association with the fibril form of the proteins.[*215*]

Recent research into mammalian aging, proposed a novel approach to targetting degenerative disorders, by resetting the epigenetic information in cells, restoring them to a 'youthful' state.[*216, 217*] This approach, known as rejuvenation biotechnology, appears to be a promising pathway towards the treatment, or even cure, of age-related disorders – whether a breakthrough can be achieved within realistic timeframes, to benefit present generations, remains to be seen.[*218*]

Lastly, and perhaps the resolution most relevant to the current work, is the regulation of metal ion concentration in the brain. One approach to achieving this, is through the use of metal chelators, which are compounds that bind to metal ions and promote their excretion from the body.[*166*] This is by no means a novel approach; deferiprone is an example of metal chelator used in the treatment of thalassemia, as a way to regulate the concentration of iron in the

body.[*219*] In fact, this particular chelator has recently undergone a phase 2 clinical trial, for the treatment for Parkinson's disease, although it was found to be ineffective, even showing signs of clinical worsening.[*220*]

1.5   Amyloid-β and Alzheimer's Disease

Alzheimer's disease (AD) has long been associated with the cognitive decline that often comes with aging, currently accounting for 2/3^{rds} of the overall dementia diagnoses.[*221*] Aside from the short-term effects that the disease entails, such as memory loss and cognitive impairment, cognitive degradation comes as a result of the chronic development of the disease. Characterised by the neural atrophy and loss of synaptic affinity in the cerebral cortex and the hippocampus, AD is mainly involved with memory and cognition, while it may further develop in loss of bodily functions and eventual death.

1.5.1   History and Onset of Alzheimer's Disease

The disease has been named after Alois Alzheimer who, having performed an autopsy on a woman who suffered from severe dementia, noted the presence of senile plaques and neurofibrillary tangles in the cerebral cortex.[*222*] The connection involving neurofibrillary tangles had also been noted before Alzheimer's case study, while correlation between senile plaques and dementia had been mentioned as far back as 1887.[*223*] These plaques have been found to mostly consist of fibrils of Aβ, and exist extracellularly, as opposed to the intra-neuronal presence of neurofibrillary tangles, which comprise of twisted fibers of the tau protein.[*224, 225*] Later, it was discovered that inclusions of αS, also accumulate in these amyloid plaques, through coordination of the non-Amyloid-β component (NAC), which is situated in the central region of αS.[*54*]

AD is caused by degeneration and death of nerve cells, due to the presence of insoluble plaques and fibrils in the affected individuals.[*226–229*] It has been reported that 5% of AD diagnoses come from early-onset familial AD (EOFAD),[*230*] with the majority of cases arising in patients with no family history of AD. The disease progression is characterised by the development of these neurofibrillary tangles and amyloid plaques, prompting neuronal cell death by disrupting the communication between the nerve cells. This disruption eventually advances to shrinkage of brain tissue, contributing to the cognitive decline that is characteristic of AD. Figure 1.15 shows a cross-section of the brain of a patient with AD, where this shrinkage is particularly striking, considering the right section was taken just one year after the left section. The amyloid,[*231–234*] metal ion,[*26, 228*] and oxidative stress[*235*] hypotheses lay the prevailing theoretical framework in determining the preconditions which result in the onset of AD. These theories are by no means mutually exclusive, in fact the work presented here, looks at the proposals in each of these theories and aims to unify them towards a better understanding of the possible mechanisms involved in the development of AD. The metal ion and oxidative stress hypotheses have already been approached in previous sections, thus the focus hereafter is on the amyloid-cascade hypothesis, with mentions on the metal ion coordination that accelerates the aggregation of the peptide.
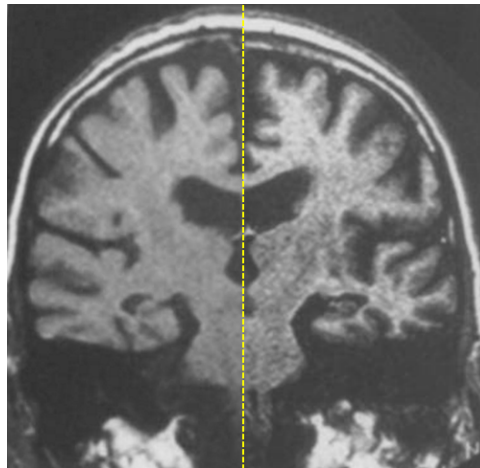
Figure 1.15: Coronal T_1-weighted MRI of the same patient, with Alzheimer's disease – the right-hand side taken one year after the left-hand side. Figure adapted from source.[*236*]

1.5.2   Involvement of Amyloid-β in Alzheimer's Disease

Amyloid-β has long been considered as the primary seed for neuronal degeneration in AD, through the deposition of the peptide in aggregates, resulting in the formation of amyloid plaques. There are several mechanisms by which Aβ contributes to the development of AD, amongst which is the toxicity of the aggregate form to neurons and the disruption of their physiological function, such as synaptic transmission.[*237*] The contribution of the fibrillar form of Aβ to the development of tau-deposits in neurofibrillar tangles, has also been supported by studies, suggesting that it acts as a trigger for the phosphorylation of tau proteins.[*238, 239*]

The amyloid precursor protein (APP) is a large transmembrane protein, cotranslationally translocated into the endoplasmic reticulum, spanning from the extracellular region (where the majority of the residues are found), through the transmembrane region (where Aβ is partly found) and finally in the cytoplasm through the short, C-terminal, tail.[*240*] APP is cleaved through two pathways: the non-amyloidogenic pathway, producing a fragment of Aβ involving residues 17-40/42 (p3), which has not, as-of-yet, been linked to the pathology of AD;[*241*] and the amyloidogenic pathway, where Aβ is produced, Figure 1.16. In the former case, APP is initially cleaved by α-secretase at the region between Aβ residues 16 and 17, resulting in the release of two fragments, a C-terminal 83-residue peptide (C83), and a soluble extracellular peptide (sAPPα). The C83 fragment is further cleaved by γ-secretase, into the APP intracellular fragment (AICD) and peptide p3. In the amyloidogenic pathway, APP is first cleaved by β-secretase, releasing again a soluble extracellular peptide (sAPPβ), and now a C-terminal 99-residue fragment, which is further cleaved by γ-secretase, releasing the AICD fragment in the cytoplasm and $Aβ_{1-40/42}$ extracellularly, where it aggregates into plaques. Following that, a cascade of events takes place, causing neuronal degeneration and cognitive decline. It has been shown, through various studies, that mutations in the APP may induce aggregation of Amyloid-β (Aβ) peptides, resulting in senile plaque formation and neurofibrillary tangles in the brain.[*242*] This effect has been described in the amyloid cascade hypothesis, discussed below. The protein has been considered as the chief molecule in AD pathogenesis, since its discovery in 1984;[*243*] making it the causative agent of the disease, trigerring the onset of dementia, along with cellular aggregation and degradation, as a direct result of the amyloid's deposition in the brain.[*229*]
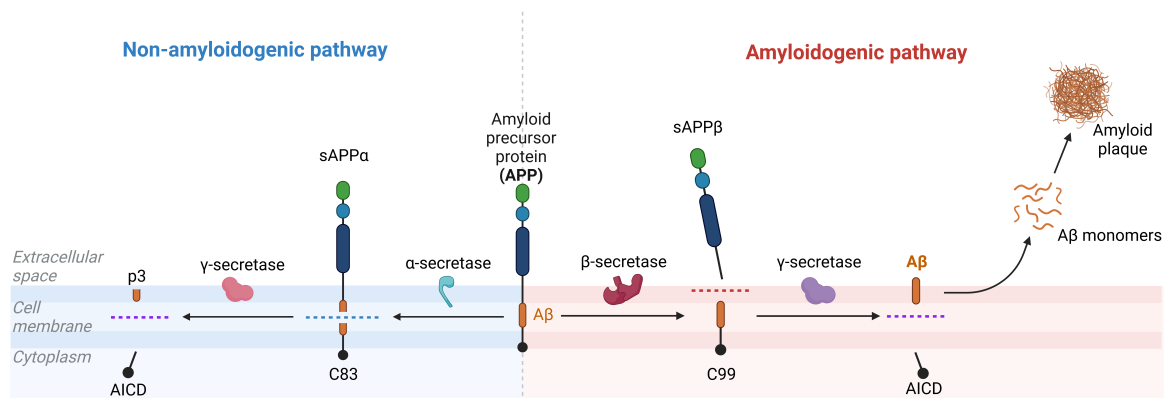
Figure 1.16: APP processing through the amyloidogenic and non-amyloidogenic pathways, from cleavage by secretases. sAPP[β/α]: soluble amyloid precursor protein [β/α]; C[99/83]: [99/83] amino acid C-terminal fragment of APP; AICD: APP intracellular fragment; p3: $A\beta_{17-40/42}$. Created with BioRender.com.

Amyloid fibrils have been shown to exhibit a high degree of polymorphism, self-assembling into various morphologies.[244] This process is believed to primarily occur in the early stages of the aggregation mechanism, during the lag phase, Figure 1.12,[245] where the peptide is still largely in its monomeric form. The fibrils are thus formed, mainly through antiparallel β-sheet packing.[246] As a result of the documented implications of this peptide in the development of the disease, recent advancements towards the treatment of AD, have focused on targetting Aβ. Aducanumab is an example of such medication, where monoclonal antibodies are used to target the Aβ plaques, with the aim of reducing their accumulation in the brain.[247] The drug has recently been approved by the FDA, although it has been met with controversy, due to the lack of evidence supporting its efficacy.[248, 249] Nevertheless, more monoclonal antibody treatments targeting Aβ have since been developed, with several of them currently undergoing clinical trials.[250]

1.5.2.1 Amyloid Cascade Hypothesis

The amyloid cascade hypothesis has become the dominant theory in trying to understand the involvement of Aβ in AD.[251] The main stipulation being that the peptide is accumulated in plaques in the form of toxic oligomers, resulting in the decline of brain function. One of the key structural characteristics of Aβ, that influences the aggregation of the peptide, is the presence of β-sheet structures. These are believed to play a catalytic role in the association of monomers into fibrils. The subsequent accumulation of those fibrils into plaques, increase the toxicity to neurons. The validity of this hypothesis has been put into question, with regards to the consideration of a linear chain of steps; rather than a multifaceted network of events, with studies supporting the role of the intermediate oligomeric structures, as being responsible for its toxicity. Evidence for this can be seen from the absence of a correlation between the plaque burden and the degree of neuronal degeneration and memory impairment.[252, 253] The hypothesis was thus modified to consider the oligomers as the pathogenic agents.[254–256]

1.5.2.2 Metal-ions in Alzheimer's Disease

Many studies report that metal ions act as strong inducers of Aβ aggregation,[257–259] due to their anchoring effect, which provides structural restraint and a seed for aggregation. The coor-

dination chemistry of several metal ions with Aβ protein and its aggregates have been studied extensively through the years.[*228*] Naturally occurring ions, such as Zn(II), Fe(II) and Cu(II), bind onto biological ligands, granting them high affinity to bind onto nucleophilic residues. Throughout the literature, the most notable coordination residues for Zn(II) are His6, His13 and His14, with the latter two being the more favourable as equatorial ligands.[*260*] Zn has been observed, through NMR and other spectroscopic studies,[*261–268*] to coordinate through a tetrahedral to six-ligand fashion[*269, 270*]; with the three residues already mentioned and a fourth one, whose identity is still debatable, but Asp1, Glu3 and Glu11 have been strong candidates.[*260, 271*] Experimental results from a Raman spectroscopic study, also proposed a His-Zn-His bridge coordination, through an inter-peptide aggregation mode.[*272*] The metal-Aβ complexes mostly form a 1:1 ratio, though a secondary, lower affinity binding mode has also been observed.[*228, 270, 273–275*]

Transition metal ions have been known to promote aggregation of Aβ, as described above.[*262, 276*] Contamination of the peptide with trace amount of metals has been determined necessary in the formation of fibrils.[*259, 273, 277*] Many studies have also concluded that soluble oligomeric precursor species of the peptide constitute the most toxic form of the protein, resulting in high neurotoxicity,[*278–280*] rather than the more aggregated forms, like fibrils and protofibrils.[*240, 251, 281, 282*] The importance of metal ions in plaque formation, can be seen from their abundance in diseased brain tissue. Zinc, in particular, has been found to be the most abundant at a concentration of up to 1 mM.[*283*] Zn is essentially released into brain synapses and its distribution in a healthy brain echo the areas most prone to amyloid deposition.[*284*] The Zn binding sites are located within the first 16 amino acids of the amyloid-β peptide,[*285–288*] which is why many studies have opted to focus on that particular region, but also the more extended $Aβ_{1-28}$ peptide.[*262, 264, 268, 289, 290*]

MD studies on the full-length peptide, suggest a significant contribution to the aggregation of Aβ, to come from the central hydrophobic region, through increasing β-sheet population of the peptide.[*291*] This characteristic has been found to be more prevalent in the zinc and iron bound systems, which presented similar development of the stabilising characteristics, between the two metal ions. Furthermore, another study implementing MD and REMD simulations, studying the Zn(II)-bound full-length peptide, offered evidence into the development of oligomer polymorphism, as a result of metal-binding and the way in which the metal bridges regions within the peptide.[*292*] These evidence suggest the binding of metals, particularly zinc, can lead to a variety of oligomeric structures, further complicating the aggregation process, while highlighting the complex nature of Aβ aggregation and the significant role that metal ions, play in this process.

Bibliography

(*1*) Al-Shammari, N., Savva, L., Kennedy-Britten, O., and Platts, J. A. (2021). Force-field evaluation and accelerated molecular dynamics simulation of Zn(II) binding to N-terminus of amyloid-β. *Computational Biology and Chemistry 93*, 107540.

(*2*) Savva, L., and Platts, J. A. (2022). Evaluation of implicit solvent models in molecular dynamics simulation of α-Synuclein. *Journal of Biomolecular Structure and Dynamics 41*, 1–16.

(*3*) Chow, V. W., Mattson, M. P., Wong, P. C., and Gleichmann, M. (2010). An overview of APP processing enzymes and products. *Neuromolecular medicine 12*, 1–12.

(*4*) Sneha, P., and Priya Doss, C. G., *Molecular Dynamics: New Frontier in Personalized Medicine*, 1st ed.; Elsevier Inc.: 2016; Vol. 102, pp 181–224.

(*5*) Rohl, C. a., and Doig, A. J. (1996). Models for the 3 10 -helix/coil, π-helix/coil, and α-helix/3 10 -helix/coil transitions in isolated peptides. *Protein Science 5*, 1687–1696.

(*6*) Cheng, P. N., Pham, J. D., and Nowick, J. S. (2013). The supramolecular chemistry of β-sheets. *Journal of the American Chemical Society 135*, 5477–5492.

(*7*) Frauenfelder, H., Parak, F., and Young, R. D. (1988). Conformational substates in proteins. *Annual review of biophysics and biophysical chemistry 17*, 451–479.

(*8*) Nelson Onuchic, J., Nymeyer, H., García, A. E., Chahine, J., and Socci, N. D. In *Proceedings of the European Conference on Synthetic Aperture Radar, EUSAR*, 2000; Vol. 1-4, pp 87–152.

(*9*) Dill, K. A. (1985). Theory for the Folding and Stability of Globular Proteins. *Biochemistry 24*, 1501–1509.

(*10*) Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science 181*, 223–230.

(*11*) Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. (1961). The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proceedings of the National Academy of Sciences 47*, 1309–1314.

(*12*) Fink, A. L. (1999). Chaperone-Mediated Protein Folding. *Physiological Reviews 79*, 425–449.

(*13*) Levinthal, C. (1969). How to fold graciously. *Mössbauer Spectroscopy in Biological Systems Proceedings 24*, 22–24.

(*14*) Martínez, L. (2014). Introducing the Levinthal's protein folding paradox and its solution. *Journal of Chemical Education 91*, 1918–1923.

(*15*) Finkelstein, A. V. et al. (2017). There and back again: Two views on the protein folding puzzle. *Physics of Life Reviews 21*, 56–71.

(*16*) Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997). Theory of Protein Folding: The Energy Landscape Perspective. *Annual Review of Physical Chemistry 48*, 545–600.

(*17*) Hansmann, U. H., Okamoto, Y., and Onuchic, J. N. (1999). The folding funnel landscape for the peptide Met-enkephalin. *Proteins: Structure, Function and Genetics 34*, 472–483.

(*18*) Cieplak, A. S., *Protein folding, misfolding and aggregation: The importance of two-electron stabilizing interactions*; 9, 2017; Vol. 12, pp 1–71.

(*19*) Noé, F., De Fabritiis, G., and Clementi, C. (2020). Machine learning for protein folding and dynamics. *Current Opinion in Structural Biology 60*, 77–84.

(*20*) Jumper, J. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 596*, 583–589.

(*21*) Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology 9*, 646–652.

(*22*) Breydo, L., and Uversky, V. N. (2011). Role of metal ions in aggregation of intrinsically disordered proteins in neurodegenerative diseases. *Metallomics 3*, 1163–1180.

(*23*) Kim, A. C., Lim, S., and Kim, Y. K. (2018). Metal ion effects on Aβ and tau aggregation. *International Journal of Molecular Sciences 19*, 1–15.

(*24*) Schipper, H. M. (2012). Neurodegeneration with brain iron accumulation - Clinical syndromes and neuroimaging. *Biochimica et Biophysica Acta - Molecular Basis of Disease 1822*, 350–360.

(*25*) Litwin, T. et al. (2013). Brain metal accumulation in Wilson's disease. *Journal of the Neurological Sciences 329*, 55–58.

(*26*) Bush, A. I., and Tanzi, R. E. (2008). Therapeutics for Alzheimer's disease based on the metal hypothesis. *Neurotherapeutics 5*, 421–432.

(*27*) Rose, F., Hodak, M., and Bernholc, J. (2011). Mechanism of copper(II)-induced misfolding of Parkinson's disease protein. *Scientific Reports 1*, 1–5.

(*28*) Belaidi, A. A., and Bush, A. I. (2016). Iron neurochemistry in Alzheimer's disease and Parkinson's disease: targets for therapeutics. *Journal of Neurochemistry 139*, 179–197.

(*29*) Haywood, S. (2019). Brain–Barrier Regulation, Metal (Cu, Fe) Dyshomeostasis, and Neurodegenerative Disorders in Man and Animals. *Inorganics 7*, 108.

(*30*) Ireland, S. M., and Martin, A. C. (2019). ZincBind - The database of zinc binding sites. *Database 2019*, 1–8.

(*31*) Cassandri, M. et al. (2017). Zinc-finger proteins in health and disease. *Cell Death Discovery 3*, 17071.

(*32*) Xu, J. et al. (2002). Dopamine-dependent neurotoxicity of α-synuclein: A mechanism for selective neurodegeneration in Parkinson disease. *Nature Medicine 8*, 600–606.

(*33*) Meiser, J., Weindl, D., and Hiller, K. (2013). Complexity of dopamine metabolism. *Cell Communication and Signaling 11*, 1–18.

(*34*) Boland, B. et al. (2018). Promoting the clearance of neurotoxic proteins in neurodegenerative disorders of ageing. *Nature Reviews Drug Discovery 17*, 660–688.

(*35*)   Junn, E., and Mouradian, M. M. (2002). Human α-Synuclein over-expression increases intracellular reactive oxygen species levels and susceptibility to dopamine. *Neuroscience Letters 320*, 146–150.

(*36*)   Wolozin, B., and Behl, C. (2000). Mechanisms of Neurodegenerative Disorders. *Archives of Neurology 57*, 801.

(*37*)   Ostrerova-Golts, N. et al. (2000). The A53T α-synuclein mutation increases iron-dependent aggregation and toxicity. *Journal of Neuroscience 20*, 6048–6054.

(*38*)   Kulagina, N. V., and Michael, A. C. (2003). Monitoring hydrogen peroxide in the extracellular space of the brain with amperometric microsensors. *Analytical Chemistry 75*, 4875–4881.

(*39*)   Pringsheim, T., Jette, N., Frolkis, A., and Steeves, T. D. (2014). The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Movement Disorders 29*, 1583–1590.

(*40*)   Valente, E. M. et al. (2004). Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science 304*, 1158–1160.

(*41*)   Cookson, M. R. (2009). α-Synuclein and neuronal cell death. *Molecular Neurodegeneration 4*, 9.

(*42*)   Parkinson, J. (2002). An Essay on the Shaking Palsy. *The Journal of Neuropsychiatry and Clinical Neurosciences 14*, 223–236.

(*43*)   Blocq, P., and Marinesco, G. (1893). Sur un cas de tremblement parkinsonien hémiplégique symptomatique d'une tumeur du pédoncule cerebral. *Comptes Rendus Séances Société Biol. 5*, 105–111.

(*44*)   Brissaud, E., *Leçons sur les Maladies Nerveuses*, 1895.

(*45*)   Carlsson, A., Lindqvist, M., Magnusson, T., and Waldeck, B. (1958). On the Presence of 3-Hydroxytyramine in Brain. *Science 127*, 471–471.

(*46*)   Ehringer, H., and Hornykiewicz, O. (1960). [Distribution of noradrenaline and dopamine (3-hydroxytyramine) in the human brain and their behavior in diseases of the extrapyramidal system]. *Klinische Wochenschrift 38*, 1236–9.

(*47*)   Birkmayer, W., and Hornykiewicz, O. (1961). [The L-3,4-dioxyphenylalanine (DOPA)-effect in Parkinson-akinesia]. *Wiener klinische Wochenschrift 73*, 787–8.

(*48*)   Polymeropoulos, M. H. et al. (1997). Mutation in the α-synuclein gene identified in families with Parkinson's disease. *Science 276*, 2045–2047.

(*49*)   Spillantini, M. G. et al. (1997). α-Synuclein in Lewy bodies. *Nature 388*, 839–840.

(*50*)   Conway, K. A., Harper, J. D., and Lansbury, P. T. (1998). Accelerated in vitro fibril formation by a mutant α-synuclein linked to early-onset Parkinson disease. *Nature Medicine 4*, 1318–1320.

(*51*)   Maroteaux, L., Campanelli, J. T., and Scheller, R. H. (1988). Synuclein: A neuron-specific protein localized to the nucleus and presynaptic nerve terminal. *Journal of Neuroscience 8*, 2804–2815.

(*52*)   Stefanis, L. (2012). α-Synuclein in Parkinson's disease. *Cold Spring Harbor Perspectives in Medicine 2*, 1–23.

(*53*)   Segrest, J. P., De Loof, H., Dohlman, J. G., Brouillette, C. G., and Anantharamaiah, G. M. (1990). Amphipathic helix motif: Classes and properties. *Proteins: Structure, Function, and Bioinformatics 8*, 103–117.

(*54*)   Uéda, K. et al. (1993). Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc. Natl. Acad. Sci. USA 90*, 11282–11286.

(*55*)   Bussell, R., and Eliezer, D. (2003). A structural and functional role for 11-mer repeats in α-synuclein and other exchangeable lipid binding proteins. *Journal of Molecular Biology 329*, 763–778.

(*56*)   Jao, C. C., Hegde, B. G., Chenb, J., Haworth, I. S., and Langen, R. (2008). Structure of membrane-bound α-synuclein from site-directed spin labeling and computational refinement. *Proceedings of the National Academy of Sciences of the United States of America 105*, 19666–19671.

(*57*)   Xu, L., Nussinov, R., and Ma, B. (2016). Coupling of the non-amyloid-component (NAC) domain and the KTK(E/Q)GV repeats stabilize the α-synuclein fibrils. *European Journal of Medicinal Chemistry 121*, 841–850.

(*58*)   Tuttle, M. D. et al. (2016). Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nature Structural & Molecular Biology 23*, 409–415.

(*59*)   Mor, D. E., Ugras, S. E., Daniels, M. J., and Ischiropoulos, H. (2016). Dynamic structural flexibility of α-synuclein. *Neurobiology of Disease 88*, 66–74.

(*60*)   Hashimoto, M. et al. (2000). The Role of NAC in Amyloidogenesis in Alzheimer's Disease. *The American Journal of Pathology 156*, 734–735.

(*61*)   Ulmer, T. S., Bax, A., Cole, N. B., and Nussbaum, R. L. (2005). Structure and dynamics of micelle-bound human α-synuclein. *Journal of Biological Chemistry 280*, 9595–9603.

(*62*)   Goedert, M., Jakes, R., and Spillantini, M. G. (2017). The Synucleinopathies: Twenty Years on. *Journal of Parkinson's Disease 7*, S53–S71.

(*63*)   Serpell, L. C., Berriman, J., Jakes, R., Goedert, M., and Crowther, R. A. (2000). Fiber diffraction of synthetic α-synuclein filaments shows amyloid-like cross-β conformation. *Proceedings of the National Academy of Sciences of the United States of America 97*, 4897–4902.

(*64*)   Vilar, M. et al. (2008). The fold of α-synuclein fibrils. *Proceedings of the National Academy of Sciences of the United States of America 105*, 8637–8642.

(*65*)   Maroteaux, L., and Scheller, R. H. (1991). The rat brain synucleins; family of proteins transiently associated with neuronal membrane. *Molecular Brain Research 11*, 335–343.

(*66*)   Gómez-Benito, M. et al. (2020). Modeling Parkinson's Disease With the Alpha-Synuclein Protein. *Frontiers in Pharmacology 11*, 1–15.

(*67*)   Burré, J. (2015). The synaptic function of α-synuclein. *Journal of Parkinson's Disease 5*, 699–713.

(*68*)   Cheng, F., Vivacqua, G., and Yu, S. (2011). The role of alpha-synuclein in neurotransmission and synaptic plasticity. *Journal of Chemical Neuroanatomy 42*, 242–248.

(69) Venda, L. L., Cragg, S. J., Buchman, V. L., and Wade-Martins, R. (2010). α-Synuclein and dopamine at the crossroads of Parkinson's disease. *Trends in Neurosciences 33*, 559–568.

(70) Burré, J. et al. (2010). α-Synuclein promotes SNARE-complex assembly in vivo and in vitro. *Science 329*, 1663–1667.

(71) Lou, X., Kim, J., Hawk, B. J., and Shin, Y. K. (2017). α-Synuclein may cross-bridge v-SNARE and acidic phospholipids to facilitate SNARE-dependent vesicle docking. *Biochemical Journal 474*, 2039–2049.

(72) Lundblad, M., Decressac, M., Mattsson, B., and Björklund, A. (2012). Impaired neurotransmission caused by overexpression of α-synuclein in nigral dopamine neurons. *Proceedings of the National Academy of Sciences of the United States of America 109*, 3213–3219.

(73) Phan, J. A. et al. (2017). Early synaptic dysfunction induced by α-synuclein in a rat model of Parkinson's disease. *Scientific Reports 7*, 1–17.

(74) Uversky, V. N. (2013). Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta - Proteins and Proteomics 1834*, 932–951.

(75) Oldfield, C. J., and Dunker, A. K. (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual Review of Biochemistry 83*, 553–584.

(76) Uversky, V. N., Li, J., and Fink, A. L. (2001). Metal-triggered structural transformations, aggregation, and fibrillation of human α-synuclein: A possible molecular link between parkinson's disease and heavy metal exposure. *Journal of Biological Chemistry 276*, 44284–44296.

(77) Ahmad, B., Chen, Y., and Lapidus, L. J. (2012). Aggregation of α-synuclein is kinetically controlled by intramolecular diffusion. *Proceedings of the National Academy of Sciences of the United States of America 109*, 2336–2341.

(78) Fujiwara, H. et al. (2002). A-Synuclein Is Phosphorylated in Synucleinopathy Lesions. *Nature Cell Biology 4*, 160–164.

(79) Anderson, J. P. et al. (2006). Phosphorylation of Ser-129 is the dominant pathological modification of α-synuclein in familial and sporadic lewy body disease. *Journal of Biological Chemistry 281*, 29739–29752.

(80) Oueslati, A. (2016). Implication of Alpha-Synuclein Phosphorylation at S129 in Synucleinopathies: What Have We Learned in the Last Decade? *Journal of Parkinson's Disease 6*, 39–51.

(81) Bell, R. et al. (2022). N-Terminal Acetylation of α-Synuclein Slows down Its Aggregation Process and Alters the Morphology of the Resulting Aggregates. *Biochemistry 61*, 1743–1756.

(82) Uversky, V. N. (2007). Neuropathology, biochemistry, and biophysics of α-synuclein aggregation. *Journal of Neurochemistry 103*, 17–37.

(83) Azeredo da Silveira, S. et al. (2008). Phosphorylation Does Not Prompt, Nor Prevent, the Formation of -synuclein Toxic Species in a Rat Model of Parkinson's Disease. *Human Molecular Genetics 18*, 872–887.

(84) Bell, R., and Vendruscolo, M. (2021). Modulation of the Interactions Between α-Synuclein and Lipid Membranes by Post-translational Modifications. *Frontiers in Neurology 12*, 1–22.

(85)   Ghanem, S. S. et al. (2022). α-Synuclein phosphorylation at serine 129 occurs after initial protein deposition and inhibits seeded fibril formation and toxicity. *Proceedings of the National Academy of Sciences of the United States of America 119*, e2109617119.

(86)   Smith, W. W. et al. (2005). α-synuclein phosphorylation enhances eosinophilic cytoplasmic inclusion formation in SH-SY5Y cells. *Journal of Neuroscience 25*, 5544–5552.

(87)   De Ricco, R. et al. (2015). Remote His50 Acts as a Coordination Switch in the High-Affinity N-Terminal Centered Copper(II) Site of α-Synuclein. *Inorganic Chemistry 54*, 4744–4751.

(88)   Lu, Y., Prudent, M., Fauvet, B., Lashuel, H. A., and Girault, H. H. (2011). Phosphorylation of α-synuclein at Y125 and S129 alters its metal binding properties: Implications for understanding the role of α-synuclein in the pathogenesis of Parkinson's disease and related disorders. *ACS Chemical Neuroscience 2*, 667–675.

(89)   Binolfi, A. et al. (2006). 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 128*, 9893–9901.

(90)   Castillo-Gonzalez, J. A. et al. (2017). Phosphorylated α-Synuclein-Copper Complex Formation in the Pathogenesis of Parkinson's Disease. *Parkinson's Disease 2017*, 1–9.

(91)   González, N. et al. (2019). Effects of alpha-synuclein post-translational modifications on metal binding. *Journal of Neurochemistry 150*, 507–521.

(92)   Buratti, F. A. et al. (2022). Aromaticity at position 39 in α-synuclein: A modulator of amyloid fibril assembly and membrane-bound conformations. *Protein Science 31*, e4360.

(93)   Dikiy, I. et al. (2016). Semisynthetic and in Vitro Phosphorylation of Alpha-Synuclein at Y39 Promotes Functional Partly Helical Membrane-Bound States Resembling Those Induced by PD Mutations. *ACS Chemical Biology 11*, 2428–2437.

(94)   Zhao, K. et al. (2020). Parkinson's disease-related phosphorylation at Tyr39 rearranges α-synuclein amyloid fibril structure revealed by cryo-EM. *Proceedings of the National Academy of Sciences of the United States of America 117*, 20305–20315.

(95)   Paleologou, K. E. et al. (2010). Phosphorylation at S87 is enhanced in synucleinopathies, inhibits α-synuclein oligomerization, and influences synuclein-membrane interactions. *Journal of Neuroscience 30*, 3184–3198.

(96)   Herrera, F. E. et al. (2008). Inhibition of α-synuclein fibrillization by dopamine is mediated by interactions with five C-terminal residues and with E83 in the NAC region. *PLoS ONE 3*, DOI: `10.1371/journal.pone.0003394`.

(97)   Hejjaoui, M. et al. (2012). Elucidating the role of C-terminal post-translational modifications using protein semisynthesis strategies: α-synuclein phosphorylation at tyrosine 125. *Journal of the American Chemical Society 134*, 5196–5210.

(98)   Kosten, J. et al. (2014). Efficient modification of alpha-synuclein serine 129 by protein kinase CK1 requires phosphorylation of tyrosine 125 as a priming event. *ACS Chemical Neuroscience 5*, 1203–1208.

(99)   Fayyad, M. et al. (2020). Investigating the presence of doubly phosphorylated α-synuclein at tyrosine 125 and serine 129 in idiopathic Lewy body diseases. *Brain Pathology 30*, 831–843.

(*100*) Gadhe, L. et al. (2022). Intermediates of α-synuclein aggregation: Implications in Parkinson's disease pathogenesis. *Biophysical Chemistry 281*, 106736.

(*101*) Li, J., Uversky, V. N., and Fink, A. L. (2001). Effect of familial Parkinson's disease point mutations A30P and A53T on the structural properties, aggregation, and fibrillation of human α-synuclein. *Biochemistry 40*, 11604–11613.

(*102*) Li, J., Uversky, V. N., and Fink, A. L. (2002). Conformational Behavior of Human α-Synuclein is Modulated by Familial Parkinson's Disease Point Mutations A30P and A53T. *NeuroToxicology 23*, 553–567.

(*103*) Lemkau, L. R. et al. (2012). Mutant protein A30P α-synuclein adopts wild-type fibril structure, despite slower fibrillation kinetics. *Journal of Biological Chemistry 287*, 11526–11532.

(*104*) Lashuel, H. A. et al. (2002). A-Synuclein, Especially the Parkinson'S Disease-Associated Mutants, Forms Pore-Like Annular and Tubular Protofibrils. *Journal of Molecular Biology 322*, 1089–1102.

(*105*) Bengoa-Vergniory, N., Roberts, R. F., Wade-Martins, R., and Alegre-Abarrategui, J. (2017). Alpha-synuclein oligomers: a new hope. *Acta Neuropathologica 134*, 819–838.

(*106*) Athanassiadou, A. et al. (1999). Genetic analysis of families with Parkinson disease that carry the Ala53Thr mutation in the gene encoding α-synudein [1]. *American Journal of Human Genetics 65*, 555–558.

(*107*) Puschmann, A. et al. (2009). A Swedish family with de novo α-synuclein A53T mutation: Evidence for early cortical dysfunction. *Parkinsonism and Related Disorders 15*, 627–632.

(*108*) Ki, C.-S. et al. (2007). The Ala53Thr mutation in the α-synuclein gene in a Korean family with Parkinson disease. *Clinical Genetics 71*, 471–473.

(*109*) Ghosh, D. et al. (2013). The parkinson's disease-associated H50Q mutation accelerates α-synuclein aggregation in vitro. *Biochemistry 52*, 6925–6927.

(*110*) Khalaf, O. et al. (2014). The H50Q mutation enhances $\alpha\alpha$-synuclein aggregation, secretion, and toxicity. *Journal of Biological Chemistry 289*, 21856–21876.

(*111*) Zarranz, J. J. et al. (2004). The new mutation, E46K, of α-synuclein causes parkinson and Lewy body dementia. *Annals of Neurology 55*, 164–173.

(*112*) CONWAY, K. A. et al. (2006). Accelerated Oligomerization by Parkinson's Disease Linked α-Synuclein Mutants. *Annals of the New York Academy of Sciences 920*, 42–45.

(*113*) Boyer, D. R. et al. (2020). The α-synuclein hereditary mutation E46K unlocks a more stable, pathogenic fibril structure. *Proceedings of the National Academy of Sciences of the United States of America 117*, 3592–3602.

(*114*) Zhao, K. et al. (2020). Parkinson's disease associated mutation E46K of α-synuclein triggers the formation of a distinct fibril structure. *Nature Communications 11*, DOI: `10.1038/s41467-020-16386-3`.

(*115*) Guzzo, A. et al. (2022). Wild-Type α-Synuclein and Variants Occur in Different Disordered Dimers and Pre-Fibrillar Conformations in Early Stage of Aggregation. *Frontiers in Molecular Biosciences 9*, 1–20.

(*116*) Fares, M. B. et al. (2014). The novel Parkinson's disease linked mutation G51D attenuates in vitro aggregation and membrane binding of α-synuclein, and enhances its secretion and nuclear localization in cells. *Human Molecular Genetics 23*, 4491–4509.

(*117*) Xu, C. K. et al. (2022). The Pathological G51D Mutation in Alpha-Synuclein Oligomers Confers Distinct Structural Attributes and Cellular Toxicity. *Molecules 27*, DOI: `10.3390/molecules27041293`.

(*118*) Ysselstein, D. et al. (2015). Effects of impaired membrane interactions on α-synuclein aggregation and neurotoxicity. *Neurobiology of Disease 79*, 150–163.

(*119*) Jensen, P. H., Nielsen, M. S., Jakes, R., Dotti, C. G., and Goedert, M. (1998). Binding of α-synuclein to brain vesicles is abolished by familial Parkinson's disease mutation. *Journal of Biological Chemistry 273*, 26292–26294.

(*120*) Winner, B. et al. (2011). In vivo demonstration that α-synuclein oligomers are toxic. *Proceedings of the National Academy of Sciences of the United States of America 108*, 4194–4199.

(*121*) Reynolds, N. P. et al. (2011). Mechanism of membrane interaction and disruption by α-synuclein. *Journal of the American Chemical Society 133*, 19366–19375.

(*122*) Narita, T., Weinert, B. T., and Choudhary, C. (2019). Functions and mechanisms of non-histone protein acetylation. *Nature Reviews Molecular Cell Biology 20*, 156–174.

(*123*) Aksnes, H., Ree, R., and Arnesen, T. (2019). Co-translational, Post-translational, and Non-catalytic Roles of N-Terminal Acetyltransferases. *Molecular Cell 73*, 1097–1114.

(*124*) Kang, L. et al. (2012). N-terminal acetylation of α-synuclein induces increased transient helical propensity and decreased aggregation rates in the intrinsically disordered monomer. *Protein Science 21*, 911–917.

(*125*) Maltsev, A. S., Ying, J., and Bax, A. (2012). Impact of N-terminal acetylation of α-synuclein on its random coil and lipid binding properties. *Biochemistry 51*, 5004–5013.

(*126*) Runfola, M., De Simone, A., Vendruscolo, M., Dobson, C. M., and Fusco, G. (2020). The N-terminal Acetylation of α-Synuclein Changes the Affinity for Lipid Membranes but not the Structural Properties of the Bound State. *Scientific Reports 10*, 1–10.

(*127*) Iyer, A. et al. (2016). The impact of N-terminal acetylation of α-synuclein on phospholipid membrane binding and fibril structure. *Journal of Biological Chemistry 291*, 21110–21122.

(*128*) Kang, L., Janowska, M. K., Moriarty, G. M., and Baum, J. (2013). Mechanistic Insight into the Relationship between N-Terminal Acetylation of α-Synuclein and Fibril Formation Rates by NMR and Fluorescence. *PLoS ONE 8*, 1–10.

(*129*) Rossetti, G. et al. (2016). Conformational ensemble of human α-synuclein physiological form predicted by molecular simulations. *Physical Chemistry Chemical Physics 18*, 5702–5706.

(*130*) Dikiy, I., and Eliezer, D. (2014). N-terminal Acetylation stabilizes N-terminal Helicity in Lipid- and Micelle-bound α-Synuclein and increases its affinity for Physiological Membranes. *Journal of Biological Chemistry 289*, 3652–3665.

(*131*) Watson, M. D., and Lee, J. C. (2019). N-Terminal Acetylation Affects α-Synuclein Fibril Polymorphism. *Biochemistry 58*, 3630–3633.

(*132*)   Fantini, J., and Yahi, N. (2010). Molecular insights into amyloid regulation by membrane cholesterol and sphingolipids: Common mechanisms in neurodegenerative diseases. *Expert Reviews in Molecular Medicine 12*, 1–22.

(*133*)   Pfefferkorn, C. M., Jiang, Z., and Lee, J. C. (2012). Biophysics of α-synuclein membrane interactions. *Biochimica et Biophysica Acta - Biomembranes 1818*, 162–171.

(*134*)   Uversky, V. N., Li, J., and Fink, A. L. (2001). Trimethylamine-N-oxide-induced folding of α-synuclein. *FEBS Letters 509*, 31–35.

(*135*)   Zhu, M., and Fink, A. L. (2003). Lipid binding inhibits α-synuclein fibril formation. *Journal of Biological Chemistry 278*, 16873–16877.

(*136*)   Pineda, A., and Burre, J. (2017). Modulating membrane binding of α-synuclein as a therapeutic strategy. *Proceedings of the National Academy of Sciences of the United States of America 114*, 1223–1225.

(*137*)   Bendor, J. T., Logan, T. P., and Edwards, R. H. (2013). The function of α-synuclein. *Neuron 79*, 1044–1066.

(*138*)   Bartels, T., Choi, J. G., and Selkoe, D. J. (2011). α-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature 477*, 107–111.

(*139*)   Frieden, C. (2007). Protein aggregation processes: In search of the mechanism. *Protein Science 16*, 2334–2344.

(*140*)   Cao, Y., Tang, X., Yuan, M., and Han, W., *Computational studies of protein aggregation mediated by amyloid: Fibril elongation and secondary nucleation*, 1st ed.; Elsevier Inc.: 2020; Vol. 170, pp 461–504.

(*141*)   Knowles, T. P. et al. (2009). An analytical solution to the kinetics of breakable filament assembly. *Science 326*, 1533–1537.

(*142*)   Tanaka, M., Collins, S. R., Toyama, B. H., and Weissman, J. S. (2006). The physical basis of how prion conformations determine strain phenotypes. *Nature 442*, 585–589.

(*143*)   Breydo, L., Wu, J. W., and Uversky, V. N. (2012). α-Synuclein misfolding and Parkinson's disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease 1822*, 261–285.

(*144*)   Verma, G., and Bhat, R. (2023). The Anthocyanidin Peonidin Interferes with an Early Step in the Fibrillation Pathway of α-Synuclein and Modulates It toward Amorphous Aggregates. *ACS Chemical Neuroscience 14*, 1424–1438.

(*145*)   Wu, J. et al. (2022). Identifying the role of co-aggregation of Alzheimer's amyloid-β with amorphous protein aggregates of non-amyloid proteins. *Cell Reports Physical Science 3*, DOI: `10.1016/j.xcrp.2022.101028`.

(*146*)   Ray, S. et al. (2020). α-Synuclein aggregation nucleates through liquid–liquid phase separation. *Nature Chemistry 12*, 705–716.

(*147*)   Sawner, A. S. et al. (2021). Modulating α-Synuclein Liquid-Liquid Phase Separation. *Biochemistry 60*, 3676–3696.

(*148*)   Mukherjee, S. et al. (2022). Liquid-liquid Phase Separation of α-Synuclein: A New Mechanistic Insight for α-Synuclein Aggregation Associated with Parkinson's Disease Pathogenesis. *Journal of Molecular Biology*, 167713.

(*149*) Uversky, V. N. (2017). Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current Opinion in Structural Biology 44*, 18–30.

(*150*) Shin, Y., and Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. *Science 357*, DOI: `10.1126/science.aaf4382`.

(*151*) Babinchak, W. M., and Surewicz, W. K. (2020). Liquid–Liquid Phase Separation and Its Mechanistic Role in Pathological Protein Aggregation. *Journal of Molecular Biology 432*, 1910–1925.

(*152*) Yuan, C. et al. (2019). Nucleation and Growth of Amino Acid and Peptide Supramolecular Polymers through Liquid–Liquid Phase Separation. *Angewandte Chemie 131*, 18284–18291.

(*153*) Schweighauser, M. et al. (2020). Structures of α-synuclein filaments from multiple system atrophy. *Nature 585*, 464–469.

(*154*) Li, Y. et al. (2018). Amyloid fibril structure of α-synuclein determined by cryo-electron microscopy. *Cell Research 28*, 897–903.

(*155*) Li, Y. et al. cryo-em structure of alpha-synuclein fiber (PDB:6A6B), 2018.

(*156*) Li, B. et al. Alpha Synuclein fibril formed by full length protein - Twister Polymorph (PDB: 6CU8), 2018.

(*157*) Li, B. et al. (2018). Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nature Communications 9*, 1–10.

(*158*) Sun, C. et al. (2023). Cryo-EM structure of amyloid fibril formed by α-synuclein hereditary A53E mutation reveals a distinct protofilament interface. *Journal of Biological Chemistry 299*, 104566.

(*159*) Zhou, K., and Zhou, H. Structure of recombinantly assembled A53E alpha-synuclein fibrils (PDB: 7UAK), 2023.

(*160*) Davidson, W. S., Jonas, A., Clayton, D. F., and George, J. M. (1998). Stabilization of α-Synuclein Secondary Structure upon Binding to Synthetic Membranes. *Journal of Biological Chemistry 273*, 9443–9449.

(*161*) Bodner, C. R., Maltsev, A. S., Dobson, C. M., and Bax, A. (2010). Differential Phospholipid Binding of α-Synuclein Variants Implicated in Parkinson's Disease Revealed by Solution NMR Spectroscopy. *Biochemistry 49*, 862–871.

(*162*) Galvagnion, C. et al. (2016). Chemical properties of lipids strongly affect the kinetics of the membrane-induced aggregation of α-synuclein. *Proceedings of the National Academy of Sciences of the United States of America 113*, 7065–7070.

(*163*) Rasia, R. M. et al. (2005). Structural characterization of copper(II) binding to α-synuclein: Insights into the bioinorganic chemistry of Parkinson's disease. *Proceedings of the National Academy of Sciences of the United States of America 102*, 4294–4299.

(*164*) Paik, S. R., Shin, H. J., Lee, J. H., Chang, C. S., and Kim, J. (1999). Copper(II)-induced self-oligomerization of α-synuclein. *Biochemical Journal 340*, 821–828.

(*165*) Binolfi, A., Quintanar, L., Bertoncini, C. W., Griesinger, C., and Fernández, C. O. (2012). Bioinorganic chemistry of copper coordination to alpha-synuclein: Relevance to Parkinson's disease. *Coordination Chemistry Reviews 256*, 2188–2201.

(*166*) Devos, D. et al. (2014). Targeting chelatable iron as a therapeutic modality in Parkinson's disease. *Antioxidants and Redox Signaling 21*, 195–210.

(*167*) Wang, C., Liu, L., Zhang, L., Peng, Y., and Zhou, F. (2010). Redox reactions of the α-Synuclein-Cu2+ complex and their effects on neuronal cell viability. *Biochemistry 49*, 8134–8142.

(*168*) Brown, D. R. (2010). Oligomeric alpha-synuclein and its role in neuronal death. *IUBMB Life 62*, 334–339.

(*169*) Wallis, L. I. et al. (2008). MRI assessment of basal ganglia iron deposition in Parkinson's disease. *Journal of Magnetic Resonance Imaging 28*, 1061–1067.

(*170*) Gröger, A., and Berg, D. (2012). Does structural neuroimaging reveal a disturbance of iron metabolism in Parkinson's disease? Implications from MRI and TCS studies. *Journal of Neural Transmission 119*, 1523–1528.

(*171*) Golts, N. et al. (2002). Magnesium inhibits spontaneous and iron-induced aggregation of α-synuclein. *Journal of Biological Chemistry 277*, 16116–16123.

(*172*) Santner, A., and Uversky, V. N. (2010). Metalloproteomics and metal toxicology of α-synuclein. *Metallomics 2*, 378–392.

(*173*) Moons, R. et al. (2020). Metal ions shape α-synuclein. *Scientific Reports 10*, 1–13.

(*174*) Paik, S. R., Shin, H. J., and Lee, J. H. (2000). Metal-catalyzed oxidation of α-synuclein in the presence of copper(II) and hydrogen peroxide. *Archives of Biochemistry and Biophysics 378*, 269–277.

(*175*) Wright, J. A., and Brown, D. R. (2008). Alpha-synuclein and its role in metal binding: Relevance to Parkinson's disease. *Journal of Neuroscience Research 86*, 496–503.

(*176*) Wang, L. et al. (2020). Current understanding of metal ions in the pathogenesis of Alzheimer's disease. *Translational Neurodegeneration 9*, 1–13.

(*177*) Sung, Y. H., Rospigliosi, C., and Eliezer, D. (2006). NMR mapping of copper binding sites in alpha-synuclein. *Biochimica et Biophysica Acta - Proteins and Proteomics 1764*, 5–12.

(*178*) Drew, S. C. et al. (2008). Cu2+ binding modes of recombinant α-synuclein - Insights from EPR spectroscopy. *Journal of the American Chemical Society 130*, 7766–7773.

(*179*) Brown, D. R. (2009). Metal binding to alpha-synuclein peptides and its contribution to toxicity. *Biochemical and Biophysical Research Communications 380*, 377–381.

(*180*) Ramis, R., Ortega-Castro, J., Vilanova, B., Adrover, M., and Frau, J. (2017). Copper(II) Binding Sites in N-Terminally Acetylated α-Synuclein: A Theoretical Rationalization. *Journal of Physical Chemistry A 121*, 5711–5719.

(*181*) Ramis, R. et al. (2019). A Coarse-Grained Molecular Dynamics Approach to the Study of the Intrinsically Disordered Protein α-Synuclein. *Journal of Chemical Information and Modeling 59*, 1458–1471.

(*182*) Cao, Z., Liu, L., Zhao, L., Li, H., and Wang, J. (2013). Comparison of the structural characteristics of Cu2+-bound and unbound α-syn12 peptide obtained in simulations using different force fields. *Journal of Molecular Modeling 19*, 1237–1250.

(*183*) Ramis, R., Ortega-Castro, J., Vilanova, B., Adrover, M., and Frau, J. (2021). Cu2+, Ca2+, and methionine oxidation expose the hydrophobic α-synuclein NAC domain. *International Journal of Biological Macromolecules 169*, 251–263.

(*184*)   Wright, J. A., Wang, X., and Brown, D. R. (2009). Unique copper-induced oligomers mediate alpha-synuclein toxicity. *The FASEB Journal 23*, 2384–2393.

(*185*)   Brown, D. R. (2007). Interactions between metals and α-synuclein - Function or artefact? *FEBS Journal 274*, 3766–3774.

(*186*)   Camponeschi, F. et al. (2013). Copper(I)-α-Synuclein Interaction: Structural Description of Two Independent and Competing Metal Binding Sites. *Inorganic Chemistry 52*, 1358–1367.

(*187*)   Mandaci, S. Y., Caliskan, M., Sariaslan, M. F., Uversky, V. N., and Coskuner-Weber, O. (2020). Epitope region identification challenges of intrinsically disordered proteins in neurodegenerative diseases: Secondary structure dependence of α-synuclein on simulation techniques and force field parameters. *Chemical Biology and Drug Design 96*, 659–667.

(*188*)   Choi, T. S. et al. (2018). Supramolecular Modulation of Structural Polymorphism in Pathogenic α-Synuclein Fibrils Using Copper(II) Coordination. *Angewandte Chemie - International Edition 57*, 3099–3103.

(*189*)   Binolfi, A. et al. (2010). Bioinorganic chemistry of Parkinson's disease: Structural determinants for the copper-mediated amyloid formation of alpha-synuclein. *Inorganic Chemistry 49*, 10668–10679.

(*190*)   Kállay, C., Várnagy, K., Micera, G., Sanna, D., and Sóvágó, I. (2005). Copper(II) complexes of oligopeptides containing aspartyl and glutamyl residues. Potentiometric and spectroscopic studies. *Journal of Inorganic Biochemistry 99*, 1514–1525.

(*191*)   Villar-Piqué, A. et al. (2016). Environmental and genetic factors support the dissociation between α-synuclein aggregation and toxicity. *Proceedings of the National Academy of Sciences of the United States of America 113*, E6506–E6515.

(*192*)   Dudzik, C. G., Walter, E. D., and Millhauser, G. L. (2011). Coordination features and affinity of the Cu2+ site in the α-synuclein protein of Parkinson's disease. *Biochemistry 50*, 1771–1777.

(*193*)   Solomon, E. I., Chen, P., Metz, M., Lee, S. K., and Palmer, A. E. (2001). Oxygen binding, activation, and reduction to water by copper proteins. *Angewandte Chemie - International Edition 40*, 4570–4590.

(*194*)   Ha, Y. et al. (2014). Dopamine and Cu +/2+ can induce oligomerization of α-synuclein in the absence of oxygen: Two types of oligomerization mechanisms for α-synuclein and related cell toxicity studies. *Journal of Neuroscience Research 92*, 359–368.

(*195*)   Gentile, I. et al. (2018). Interaction of Cu(i) with the Met-X3-Met motif of alpha-synuclein: binding ligands, affinity and structural features. *Metallomics 10*, 1383–1389.

(*196*)   Miotto, M. C. et al. (2014). Site-specific copper-catalyzed oxidation of α-synuclein: Tightening the link between metal binding and protein oxidative damage in Parkinson's disease. *Inorganic Chemistry 53*, 4350–4358.

(*197*)   Maltsev, A. S., Chen, J., Levine, R. L., and Bax, A. (2013). Site-specific interaction between α-synuclein and membranes probed by NMR-observed methionine oxidation rates. *Journal of the American Chemical Society 135*, 2943–2946.

(*198*)   Schildknecht, S. et al. (2013). Oxidative and nitrative alpha-synuclein modifications and proteostatic stress: implications for disease mechanisms and interventions in synucleinopathies. *Journal of Neurochemistry 125*, 491–511.

(*199*) Valensin, D. et al. (2011). The role of His-50 of α-synuclein in binding Cu(ii): PH dependence, speciation, thermodynamics and structure. *Metallomics 3*, 292–302.

(*200*) Villar-Piqué, A. et al. (2017). Copper(II) and the pathological H50Q α-synuclein mutant: Environment meets genetics. *Communicative and Integrative Biology 10*, 1–4.

(*201*) Dudzik, C. G., Walter, E. D., Abrams, B. S., Jurica, M. S., and Millhauser, G. L. (2013). Coordination of copper to the membrane-bound form of α-synuclein. *Biochemistry 52*, 53–60.

(*202*) Lucas, H. R., and Lee, J. C. (2011). Copper(ii) enhances membrane-bound α-synuclein helix formation. *Metallomics 3*, 280.

(*203*) Ahmad, A., Burns, C. S., Fink, A. L., and Uversky, V. N. (2012). Peculiarities of copper binding to α-synuclein. *Journal of Biomolecular Structure and Dynamics 29*, 825–842.

(*204*) Bortolus, M. et al. (2010). Structural characterization of a high affinity mononuclear site in the copper(II)-α-synuclein complex. *Journal of the American Chemical Society 132*, 18057–18066.

(*205*) Drew, S. C. (2015). The N Terminus of α-Synuclein Forms Cu II -Bridged Oligomers. *Chemistry - A European Journal 21*, 7111–7118.

(*206*) Abeyawardhane, D. L., Heitger, D. R., Fernández, R. D., Forney, A. K., and Lucas, H. R. (2019). C-Terminal Cu II Coordination to α-Synuclein Enhances Aggregation. *ACS Chemical Neuroscience 10*, 1402–1410.

(*207*) Wang, X., Moualla, D., Wright, J. A., and Brown, D. R. (2010). Copper binding regulates intracellular alpha-synuclein localisation, aggregation and toxicity. *Journal of Neurochemistry 113*, 704–714.

(*208*) McFarthing, K., and Simuni, T. (2019). Clinical trial highlights: Targetting alpha-synuclein. *Journal of Parkinson's Disease 9*, 5–16.

(*209*) Volc, D. et al. (2020). Safety and immunogenicity of the α-synuclein active immunotherapeutic PD01A in patients with Parkinson's disease: a randomised, single-blinded, phase 1 trial. *The Lancet Neurology 19*, 591–600.

(*210*) Lang, A. E. et al. (2022). Trial of Cinpanemab in Early Parkinson's Disease. *New England Journal of Medicine 387*, 408–420.

(*211*) Pagano, G. et al. (2022). Trial of Prasinezumab in Early-Stage Parkinson's Disease. *New England Journal of Medicine 387*, 421–432.

(*212*) Wagner, J. et al. (2013). Anle138b: A novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases such as prion and Parkinson's disease. *Acta Neuropathologica 125*, 795–813.

(*213*) Levin, J. et al. (2022). Safety, tolerability and pharmacokinetics of the oligomer modulator anle138b with exposure levels sufficient for therapeutic efficacy in a murine Parkinson model: A randomised, double-blind, placebo-controlled phase 1a trial. *eBioMedicine 80*, 1–12.

(*214*) Matthes, D., Gapsys, V., Griesinger, C., and De Groot, B. L. (2017). Resolving the Atomistic Modes of Anle138b Inhibitory Action on Peptide Oligomer Formation. *ACS Chemical Neuroscience 8*, 2791–2808.

(215) Antonschmidt, L. et al. (2022). The clinical drug candidate anle138b binds in a cavity of lipidic α-synuclein fibrils. *Nature Communications 13*, DOI: `10.1038/s41467-022-32797-w`.

(216) Gjoneska, E. et al. (2015). Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature 518*, 365–369.

(217) Yang, J.-H. et al. (2023). Loss of epigenetic information as a cause of mammalian aging. *Cell*, 305–326.

(218) de Magalhães, J. P., and Ocampo, A. (2022). Cellular reprogramming and the rise of rejuvenation biotech. *Trends in Biotechnology 40*, 639–642.

(219) Hider, R. C., and Hoffbrand, A. V. (2018). The Role of Deferiprone in Iron Chelation. *New England Journal of Medicine 379*, 2140–2150.

(220) Devos, D. et al. (2022). Trial of Deferiprone in Parkinson's Disease. *New England Journal of Medicine 387*, 2045–2055.

(221) Alzheimer's Association (2020). 2020 Alzheimer's disease facts and figures. *Alzheimer's & Dementia 16*, 391–460.

(222) Alzheimer, A. (1907). Über eine eigenartige Erkrankung der Hirnrinde. *Allgemeine Zeitschrift fur Psychiatrie und Psychisch-gerichtliche Medizin 64*, 146–148.

(223) Ramirez-Bermudez, J. (2012). Alzheimer's Disease: Critical Notes on the History of a Medical Concept. *Archives of Medical Research 43*, 595–599.

(224) Kosik, K. et al. (1988). Epitopes that span the tau molecule are shared with paired helical filaments. *Neuron 1*, 817–825.

(225) Goedert, M., Spillantini, M. G., Jakes, R., Rutherford, D., and Crowther, R. A. (1989). Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron 3*, 519–526.

(226) Kosik, K. S., Joachim, C. L., and Selkoe, D. J. (1986). Microtubule-associated protein tau (τ) is a major antigenic component of paired helical filaments in Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America 83*, 4044–4048.

(227) Karantzoulis, S., and Galvin, J. E. (2011). Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Review of Neurotherapeutics 11*, 1579–1591.

(228) Kepp, K. P. (2017). Alzheimer's disease: How metal ions define β-amyloid function. *Coordination Chemistry Reviews 351*, 127–159.

(229) Hardy, J. A., and Higgins, G. A. (1992). Alzheimer's Disease: The Amyloid Alzheimer's disease. *Science 256*, 184–185.

(230) Ballard, C. et al. (2011). Alzheimer's disease. *The Lancet 377*, 1019–1031.

(231) Selkoe, D. J. (1991). The molecular pathology of Alzheimer's disease. *Neuron 6*, 487–498.

(232) Walsh, D. M. et al. (2002). Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation in vivo. *Nature 416*, 535–539.

(233) Cleary, J. P. et al. (2005). Natural oligomers of the amyloid-β protein specifically disrupt cognitive function. *Nature Neuroscience 8*, 79–84.

(234) Haass, C., and De Strooper, B. (1999). The presenilins in Alzheimer's disease - Prote-olysis holds the key. *Science 286*, 916–919.

(235) Cheignon, C. et al. (2018). Oxidative stress and the amyloid beta peptide in Alzheimer's disease. *Redox Biology 14*, 450–464.

(236) Scheltens, P., Launer, L. J., Barkhof, F., Weinstein, H. C., and van Gool, W. A. (1995). Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: Interobserver reliability. *Journal of Neurology 242*, 557–560.

(237) Bloom, G. S. (2014). Amyloid-β and tau: The trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurology 71*, 505–508.

(238) Busciglio, J., Lorenzo, A., Yeh, J., and Yankner, B. A. (1995). β-Amyloid fibrils induce tau phosphorylation and loss of microtubule binding. *Neuron 14*, 879–888.

(239) Rapoport, M., Dawson, H. N., Binder, L. I., Vitek, M. P., and Ferreira, A. (2002). Tau is essential to β-amyloid-induced neurotoxicity. *Proceedings of the National Academy of Sciences of the United States of America 99*, 6364–6369.

(240) Selkoe, D. J. (2001). Alzheimer's Disease: Genes, Proteins, and Therapy. *Physiological Reviews 81*, 741–766.

(241) Kuhn, A. J., and Raskatov, J. (2020). Is the p3 Peptide (Aβ17-40, Aβ17-42) Relevant to the Pathology of Alzheimer's Disease?1. *Journal of Alzheimer's Disease 74*, ed. by Barron, A., 43–53.

(242) Grasso, G., and Danani, A. (2020). Molecular simulations of amyloid beta assemblies. *Advances in Physics: X 5*, 1770627.

(243) Glenner, G. G., and Wong, C. W. (1984). Alzheimer's disease: Initial report of the pu-rification and characterization of a novel cerebrovascular amyloid protein. *Biochemical and Biophysical Research Communications 120*, 885–890.

(244) Pedersen, J. S. et al. (2006). The changing face of glucagon fibrillation: Structural poly-morphism and conformational imprinting. *Journal of Molecular Biology 355*, 501–523.

(245) Kodali, R., and Wetzel, R. (2007). Polymorphism in the intermediates and products of amyloid assembly. *Current Opinion in Structural Biology 17*, 48–57.

(246) Härd, T. (2014). Amyloid Fibrils: Formation, Polymorphism, and Inhibition. *The Jour-nal of Physical Chemistry Letters 5*, 607–614.

(247) Sevigny, J. et al. (2016). The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature 537*, 50–56.

(248) Avgerinos, K. I., Ferrucci, L., and Kapogiannis, D. (2021). Effects of monoclonal anti-bodies against amyloid-β on clinical and biomarker outcomes and adverse event risks: A systematic review and meta-analysis of phase III RCTs in Alzheimer's disease. *Age-ing Research Reviews 68*, 101339.

(249) Walsh, S., Merrick, R., Milne, R., and Brayne, C. (2021). Aducanumab for Alzheimer's disease? *The BMJ 374*, 10–11.

(250) Ramanan, V. K., and Day, G. S. (2023). Anti-amyloid therapies for Alzheimer disease: finally, good news for patients. *Molecular Neurodegeneration 18*, 1–3.

(251) Hardy, J., and Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science 297*, 353–356.

(252)   Kayed, R. et al. (2003). Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science 300*, 486–489.

(253)   Hardy, J. (2009). The amyloid hypothesis for Alzheimer's disease: A critical reappraisal. *Journal of Neurochemistry 110*, 1129–1134.

(254)   Kirkitadze, M. D., Bitan, G., and Teplow, D. B. (2002). Paradigm shifts in Alzheimer's disease and other neurodegenerative disorders: The emerging role of oligomeric assemblies. *Journal of Neuroscience Research 69*, 567–577.

(255)   Caughey, B., and Lansbury, P. T. (2003). Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annual Review of Neuroscience 26*, 267–298.

(256)   Glabe, C. G. (2006). Common mechanisms of amyloid oligomer pathogenesis in degenerative disease. *Neurobiology of Aging 27*, 570–575.

(257)   Esler, W. P. et al. (1996). Zinc-induced aggregation of human and rat β-amyloid peptides in vitro. *Journal of Neurochemistry 66*, 723–732.

(258)   Bush, A. I. et al. (1994). Rapid induction of Alzheimer Aβ amyloid formation by zinc. *Science 265*, 1464–1467.

(259)   Huang, X. et al. (2004). Trace metal contamination initiates the apparent auto-aggregation, amyloidosis, and oligomerization of Alzheimer's Aβ peptides. *Journal of Biological Inorganic Chemistry 9*, 954–960.

(260)   Silva, K. I., Michael, B. C., Geib, S. J., and Saxena, S. (2014). ESEEM Analysis of Multi-Histidine Cu(II)-Coordination in Model Complexes, Peptides, and Amyloid-β. *The Journal of Physical Chemistry B 118*, 8935–8944.

(261)   Curtain, C. C. et al. (2001). Alzheimer's Disease Amyloid-β Binds Copper and Zinc to Generate an Allosterically Ordered Membrane-penetrating Structure Containing Superoxide Dismutase-like Subunits. *Journal of Biological Chemistry 276*, 20466–20473.

(262)   Faller, P., and Hureau, C. (2009). Bioinorganic chemistry of copper and zinc ions coordinated to amyloid-β peptide. *Dalton Transactions*, 1080–1094.

(263)   Faller, P. (2009). Copper and zinc binding to amyloid-β: Coordination, dynamics, aggregation, reactivity and metal-ion transfer. *ChemBioChem 10*, 2837–2845.

(264)   Alies, B. et al. (2011). PH-dependent Cu(II) coordination to amyloid-β peptide: Impact of sequence alterations, including the H6R and D7N familial mutations. *Inorganic Chemistry 50*, 11192–11201.

(265)   Alies, B. et al. (2012). Copper coordination to native N-terminally modified versus full-length amyloid-β: Second-sphere effects determine the species present at physiological pH. *Inorganic Chemistry 51*, 12988–13000.

(266)   Alies, B. et al. (2016). Zinc(II) Binding Site to the Amyloid-β Peptide: Insights from Spectroscopic Studies with a Wide Series of Modified Peptides. *Inorganic Chemistry 55*, 10499–10509.

(267)   Bousejra-Elgarah, F., Bijani, C., Coppel, Y., Faller, P., and Hureau, C. (2011). Iron(II) binding to amyloid-β, the Alzheimer's peptide. *Inorganic Chemistry 50*, 9024–9030.

(268)   Eury, H., Bijani, C., Faller, P., and Hureau, C. (2011). Copper(II) coordination to amyloid β: Murine versus human peptide. *Angewandte Chemie - International Edition 50*, 901–905.

(*269*)   Miller, Y., Ma, B., and Nussinov, R. (2012). Metal binding sites in amyloid oligomers: Complexes and mechanisms. *Coordination Chemistry Reviews 256*, 2245–2252.

(*270*)   Tõugu, V., and Palumaa, P. (2012). Coordination of zinc ions to the key proteins of neurodegenerative diseases: Aβ, APP, α-synuclein and PrP. *Coordination Chemistry Reviews 256*, 2219–2224.

(*271*)   Zirah, S. et al. (2006). Structural Changes of Region 1-16 of the Alzheimer Disease Amyloid β-Peptide upon Zinc Binding and in Vitro Aging. *Journal of Biological Chemistry 281*, 2151–2161.

(*272*)   Miura, T., Suzuki, K., Kohata, N., and Takeuchi, H. (2000). Metal binding modes of Alzheimer's amyloid β-peptide in insoluble aggregates and soluble complexes. *Biochemistry 39*, 7024–7031.

(*273*)   Atwood, C. S. et al. (2008). Characterization of Copper Interactions with Alzheimer Amyloid β Peptides. *Journal of Neurochemistry 75*, 1219–1233.

(*274*)   Tõugu, V., Karafin, A., and Palumaa, P. (2008). Binding of zinc(II) and copper(II) to the full-length Alzheimer's amyloid-β peptide. *Journal of Neurochemistry 104*, 1249–1259.

(*275*)   Kozin, S. A., Zirah, S., Rebuffat, S., Hui Bon Hoa, G., and Debey, P. (2001). Zinc binding to Alzheimer's Aβ(1-16) peptide results in stable soluble complex. *Biochemical and Biophysical Research Communications 285*, 959–964.

(*276*)   Furlan, S., and La Penna, G. (2009). Modeling of the Zn2+ binding in the 1-16 region of the amyloid β peptide involved in Alzheimer's disease. *Physical Chemistry Chemical Physics 11*, 6468–6481.

(*277*)   Sengupta, P. et al. (2003). The amyloid β peptide (Aβ1-40) is thermodynamically soluble at physiological concentrations. *Biochemistry 42*, 10506–10513.

(*278*)   Eckman, C. B., and Eckman, E. A. An Update on the Amyloid Hypothesis, 2007.

(*279*)   Drew, S. C., Noble, C. J., Masters, C. L., Hanson, G. R., and Barnham, K. J. Pleomorphic copper coordination by alzheimer's disease amyloid-β peptide, 2009.

(*280*)   Rana, M., and Sharma, A. K. (2019). Cu and Zn interactions with Aβ peptides: consequence of coordination on aggregation and formation of neurotoxic soluble Aβ oligomers. *Metallomics 11*, 64–84.

(*281*)   Haass, C., and Selkoe, D. J. (2007). Soluble protein oligomers in neurodegeneration: Lessons from the Alzheimer's amyloid β-peptide. *Nature Reviews Molecular Cell Biology 8*, 101–112.

(*282*)   Klein, W. L., Stine, W. B., and Teplow, D. B. (2004). Small assemblies of unmodified amyloid β-protein are the proximate neurotoxin in Alzheimer's disease. *Neurobiology of Aging 25*, 569–580.

(*283*)   Lovell, M., Robertson, J., Teesdale, W., Campbell, J., and Markesbery, W. (1998). Copper, iron and zinc in Alzheimer's disease senile plaques. *Journal of the Neurological Sciences 158*, 47–52.

(*284*)   Bush, A. I. (2002). Metal complexing agents as therapies for Alzheimer's disease. *Neurobiology of Aging 23*, 1031–1038.

(*285*)   Curtain, C. C. et al. (2003). Metal ions, pH, and cholesterol regulate the interactions of Alzheimer's disease amyloid-β peptide with membrane lipid. *Journal of Biological Chemistry 278*, 2977–2982.

(*286*)   Karr, J. W., Akintoye, H., Kaupp, L. J., and Szalai, V. A. (2005). N-terminal deletions modify the Cu2+ binding site in amyloid-β. *Biochemistry 44*, 5478–5487.

(*287*)   Guilloreau, L. et al. (2006). Structural and thermodynamical properties of CuII amyloid-β16/28 complexes associated with Alzheimer's disease. *Journal of Biological Inorganic Chemistry 11*, 1024–1038.

(*288*)   Kowalik-Jankowska, T., Ruta, M., Wiśniewska, K., and Łankiewicz, L. (2003). Coordination abilities of the 1-16 and 1-28 fragments of β-amyloid peptide towards copper(II) ions: A combined potentiometric and spectroscopic study. *Journal of Inorganic Biochemistry 95*, 270–282.

(*289*)   Karr, J. W., Kaupp, L. J., and Szalai, V. A. (2004). Amyloid-β binds Cu2+ in a mononuclear metal ion binding site. *Journal of the American Chemical Society 126*, 13534–13538.

(*290*)   Minicozzi, V. et al. (2008). Identifying the minimal copper- and zinc-binding site sequence in amyloid-β peptides. *Journal of Biological Chemistry 283*, 10784–10792.

(*291*)   Kennedy-Britten, O. D., Alshammari, N., and Platts, J. A. (2021). Accelerated Molecular Dynamics to Explore the Binding of Transition Metals to Amyloid-β. *ACS Chemical Neuroscience 12*, 4065–4075.

(*292*)   Miller, Y., Ma, B., and Nussinov, R. (2010). Zinc ions promote alzheimer aβ aggregation via population shift of polymorphic states. *Proceedings of the National Academy of Sciences of the United States of America 107*, 9490–9495.

Chapter 2

Theory

This chapter focuses on describing the theory behind the methods used throughout this thesis. Included are also theoretical concepts important to the general field, but not necessarily used to a large extent in the present work.

2.1 Quantum Mechanics

The probabilistic quantum mechanical concepts described below have been primarily derived from computational chemistry textbooks[1–3] and original published works on the theories.[4–11]

2.1.1 The Schrödinger Equation

The quantum non-relativistic equation, famously devised by Schrödinger, aims to describe the behaviour of small-mass particles, primarily electrons. Solving the Schrödinger's equation yields the wave function (ψ) as a function of time. Hence, it is used to define the behaviour of atomic and subatomic particles, as shown in the time-dependent equation – Eq. 2.1. The square of the wave function gives the probability of observing a particle at position r and time t, as per Eq. 2.2.

$$i\hbar \frac{\partial}{\partial t} \Psi(r, t) = \widehat{H}\Psi(r, t) = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(r, t)\right]\Psi(r, t) \tag{2.1}$$

$$P(r, t) = \Psi^2(r, t) \tag{2.2}$$

Where, $i\hbar\frac{\partial}{\partial t}$ = energy operator ($i = \sqrt{-1}$), $\widehat{H}$ = Hamiltonian, $\hbar$ = reduced Planck constant, m = mass of particles, ∇^2 = Laplace operator, V = potential, r = space and t = time.

The Hamiltonian operator is applied to account for the kinetic (T) and potential (V) energy terms:

$$\widehat{H} = T + V \tag{2.3}$$

$$T = -\frac{\hbar}{2m}\nabla^2 \tag{2.4}$$

In a bound system, where the Hamiltonian operator is independent of time, the equation can be separated into spatial, $\varphi(r)$, and time-dependent, $\psi(t)$, parts. The equations for the wave function, are hence defined:

$$\Psi(r, t) = \varphi(r) \psi(t) \tag{2.5}$$

$$i \frac{\partial \psi(t)}{\partial t} = E \psi(t) \tag{2.6}$$

$$\widehat{H} \varphi(r) = E \varphi(r) \tag{2.7}$$

The Born-Oppenheimer approximation, assumes a separable wave function, owing to the electronic relaxation being much faster than nuclear motion. An electronic system is thus handled as a wave function at a fixed external potential of the nuclei. The ground state of such a system acts as the potential function for the nuclei. The interactions within the system are expressed using a Hamiltonian, as shown in Eq. 2.8, and simplified using the Born-Oppenheimer approximation, Eq. 2.9. This can be done by disregarding the negligible nuclear kinetic energy, when compared to that of the electrons. Hence, allowing for the wave function in a time-independent Schrödinger equation (TISE), to be separated, thus permitting the derivation of its electronic component.

$$\widehat{H} = -\frac{\hbar^2}{2m_i} \sum_i^N \nabla_i^2 - \frac{\hbar^2}{2m_A} \sum_A^M \nabla_A^2 - \sum_i^N \sum_A^M \frac{Z_A}{r_{iA}} + \sum_i^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{iA}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \tag{2.8}$$

$$\widehat{H}_{\text{elec}} = -\frac{\hbar^2}{2m_i} \sum_i^N \nabla_i^2 - \sum_i^N \sum_A^M \frac{Z_A}{r_{iA}} + \sum_i^N \sum_{j>i}^N \frac{1}{r_{ij}} \tag{2.9}$$

Where, M = number of nuclei in the system, m_A = mass of nuclei, Z = charge of nuclei, R = position of nuclei, A and B = nuclei, N = number of electrons in the system, m_i = mass of electrons, r = position of electron, i and j = electrons. R_{AB} = distance between nuclei, $r_{iA} = |r_i - R_A|$ = distance between electron i and nucleus A, and r_{ij} = distance between electrons.

2.1.2 Hartree-Fock

The electronic Schrödinger equation, given in Eq. 2.9, has been employed by the HF method, to simplify individual electrons' motion in a many-body quantum system. The equation, really only solves exactly for the H_2^+ molecule and other single-electron systems. Hence, for a broader application of the equation, approximations must be made, based on the assumption that electrons do not share interactions in space, thus providing independent probability distributions. These are approximated by a single Slater determinant.[1] The solution to this method is defined by the Hartree-product (HP). In the case of a two-electron system, where spin orbitals are defined as χ_α and χ_β, the Hartree-product can be represented as:

$$\psi_{ij}^{\text{HP}}(x_i, x_j) = \chi_\alpha(x_i) \chi_\beta(x_j) \tag{2.10}$$

Having neglected relativistic effects, an *ad hoc* quantum effect is introduced, concerning the spin quantum number of each electron. In the presence of a magnetic field, the spins of an electron pair can interchange states, either aligning along or opposite the field. This results to the antisymmetry principle not being satisfied, as: $\psi_{ij}^{\text{HP}}(x_i, x_j) \neq -\psi_{ji}^{\text{HP}}(x_j, x_i)$.

To approximate the energy of the wave function, the expectation value of the $\widehat{H}$, is divided by the norm of the wave function:

$$E = \frac{\langle \psi | \widehat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \tag{2.11}$$

A normalised wave function results in the denominator being equal to 1, hence $E = \langle \psi | \widehat{H} | \psi \rangle$. The matrix element can thus be expanded to $\int \psi^* \widehat{H} \psi dr$, where the bra, $<\psi|$, and ket, $|\psi>$, are the complex conjugate of the wave function, and the wave function, respectively. This can further be expanded, by considering the kinetic and potential contributions to the Hamiltonian, Eq. 2.3 and Eq. 2.4:

$$E = \int \psi^* \left(-\frac{\hbar}{2m} \nabla^2 \right) \psi dr + \int \psi^* V \psi dr \tag{2.12}$$

For the wave function to satisfy both the antisymmetry and Pauli principles, a linear combination of the Hartree-products must be used, by employing a single Slater determinant. For a two-electron system, the following equation can hence be used:

$$\psi_{ij}^{HP} \left(x_i, x_j \right) = \frac{1}{\sqrt{2}} \chi_\alpha \left(x_i \right) \chi_\beta \left(x_j \right) - \chi_\alpha \left(x_i \right) \chi_\beta \left(x_j \right) = \frac{1}{\sqrt{2}} \begin{vmatrix} \chi_\alpha \left(x_i \right) & \chi_\beta \left(x_i \right) \\ \chi_\alpha \left(x_j \right) & \chi_\beta \left(x_j \right) \end{vmatrix} \tag{2.13}$$

For an N-electron system with M-spin orbitals, Eq. 2.13 can be rewritten as:

$$\psi^{HP} \left(x_i, x_j, ..., x_N \right) = \frac{1}{\sqrt{(N!)}} \begin{vmatrix} \chi_\alpha \left(x_i \right) & \chi_\beta \left(x_i \right) & \cdots & \chi_M \left(x_i \right) \\ \chi_\alpha \left(x_j \right) & \chi_\beta \left(x_j \right) & \cdots & \chi_M \left(x_j \right) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_\alpha \left(x_N \right) & \chi_\beta \left(x_N \right) & \cdots & \chi_M \left(x_N \right) \end{vmatrix} \tag{2.14}$$

The use of single Slater determinants neglects any potential electron correlation. The rows in the determinant represent the electron coordinates and the columns the single-electron wave functions.

2.1.2.1   Basis Set Approximation

The integro-differential equations encountered in HF methods, may be solved through various basis functions, with regards to reaching the HF limit for atoms or diatomic molecules. In the case of molecular systems, however, it becomes less likely for the solutions to reach the HF limit, with increasing complexity. Hence, the linear combination of atomic orbitals (LCAO) method is used to expand the canonical molecular orbitals (ψ), in terms of basis functions χ, where M_{basis} represents the set of the different basis functions employed for the nuclei, defined in Eq. 2.15 with expansion coefficient $c_{\alpha i}$.

$$\psi = \sum_{\alpha}^{M_{basis}} c_{\alpha i} \chi_\alpha \tag{2.15}$$

The Roothaan-Hall equations,[12, 13] can be used to solve for the molecular orbitals within the basis set framework:

$$E \sum_{\alpha}^{M_{\text{basis}}} c_{\alpha i} \chi_\alpha = \varepsilon_i \sum_{\alpha}^{M_{\text{basis}}} c_{\alpha i} \chi_\alpha \tag{2.16}$$

All the M_{basis} equations may be written in terms of matrix, with F representing the Fock matrix, S matrix denoting the overlapping elements between basis functions, C for the coefficients' matrix and ε the matrix of transformed Lagrange multipliers, with the orbital energies of electrons on the diagonal:

$$FC = SC\varepsilon \tag{2.17}$$

This results in a self-consistent field (SCF) approximation, which is implemented in *ab initio* calculations, where the potential energy is computed from the solution to the electronic Schrödinger equation. It is an iterative method, that selects an approximate Hamiltonian, which is then implemented to solve the Schrödinger equation. This permits the acquirement of orbitals with greater accuracy, which are then used to solve the Schrödinger equation, again, until convergence is achieved.

2.1.3   Density Functional Theory

The foundation of Density Functional Theory (DFT), lies on the use of electron density to solve for the ground state electronic energy, using the three main theories described below. The DFT methods are considered more efficient in determining the properties of a many-body system, than HF, due to the use of electron densities, instead of wave functions. This is because the complexity of wave functions increases exponentially with the number of electrons in the system, whereas electron densities use the same number of variables regardless of the system size – *vide infra*, Section 2.1.3.2. DFT calculations are used to model functionals, with the aim of relating the electron density to the ground state energy of the system. Different DFT methods exist, based on the choice of functionals for the exchange-correlation energy. To account for the electron correlation in a system, the *Local Density Approximation* (LDA) or *Generalised Gradient Approximation* (GGA) methods (detailed in Sections 2.1.3.5 – 2.1.3.7) are implemented.

2.1.3.1   Thomas-Fermi

The Thomas-Fermi[4, 7] model may be considered as an early concept of DFT-like methods, being introduced shortly after the Schrödinger equation, as a way to use electron densities, rather than wave functions to model the energy of a system. The energy functional suggested here, makes use of the electron density, *n(r)*, to simulate the total kinetic energy of the electrons. One of the limitations that came with the proposal, is the assumption of an infinite nuclear charge. An expression for the calculation of the kinetic energy is given below:

$$T_{\text{TF}} = \frac{3}{5} \left(\frac{\hbar^2}{2m} \right) \left(3\pi^2 \right)^{2/3} \int n\left(r\right)^{5/3} d^3 r \tag{2.18}$$

2.1.3.2   Hohenberg-Kohn

Some of the most consequential theorems for DFT methods, came from Hohenberg and Kohn in 1964.[5] The theorems compute the ground-state electron density, based on the assumption that it is exact and non-degenerate. The major reason for the efficiency of this method is that, while the wave mechanics approach uses 4N variables to describe individual electrons in an N-electron system – 3 spatial and one spin coordinate – when describing a system using electron density, it can be represented using only the 3 spatial coordinates. Hence, one can see how the complexity of using wave functions increases exponentially with the number of electrons, while the electron density approach uses the same number of variables, regardless of the system size.

The existence theorem, is used to describe how the electron density is uniquely responsible for the determination of the Hamiltonian operator, and therefore all the properties of the system. Assuming the system consists of two external ground state potentials, $V_{ext}(r)$ and $V'_{ext}(r)$, then these are defined by two Hamiltonians, $\widehat{H}$ and $\widehat{H}'$, and thus two wave functions, ψ and ψ', and ground state energies, E_0 and E'_0. Since the external potential is a function of electron density and it may be used to fix the Hamiltonian, then the electron density can be used as a function for the many-particle ground state. This, in turn, allows for the Hamiltonian to be solved using the electron density. Taking into consideration the variational principle, which states that the Hamiltonian for a wave function is equal to or greater than the ground state energy; then the expected values for these energies, calculated for each external potential, may be defined as depicted in Eq. 2.22 and Eq. 2.23. The equations, hence, prove that the external potential is a unique functional of the electron density. The bra-ket notation has been used to denote the expectation value for the Hamiltonian on the wave function and prove the theorem by *reductio ad absurdum*:

$$E_0 = \left\langle \psi \left| \widehat{H} \right| \psi \right\rangle < \left\langle \psi' \left| \widehat{H} \right| \psi' \right\rangle \tag{2.19}$$

$$< \left\langle \psi' \left| \widehat{H} - \widehat{H}' + \widehat{H}' \right| \psi' \right\rangle \tag{2.20}$$

$$< \left\langle \psi' \left| \widehat{H} - \widehat{H}' \right| \psi' \right\rangle + \left\langle \psi' \left| \widehat{H}' \right| \psi' \right\rangle \tag{2.21}$$

Retaining the assumption that the ground-state electron density is non-degenerate, the inequality strictly holds.

$$E_0 < \left\langle \psi' \left| \widehat{H} \right| \psi' \right\rangle = \left\langle \psi' \left| \widehat{H}' \right| \psi' \right\rangle + \left\langle \psi' \left| \widehat{H} - \widehat{H}' \right| \psi' \right\rangle$$
$$= E'_0 + \int n(r) \left[V_{ext}(r) - V'_{ext}(r) \right] dr \tag{2.22}$$

Interchanging the primed and unprimed quantities:

$$E'_0 < \left\langle \psi \left| \widehat{H}' \right| \psi \right\rangle = \left\langle \psi \left| \widehat{H} \right| \psi \right\rangle + \left\langle \psi \left| \widehat{H}' - \widehat{H} \right| \psi \right\rangle$$
$$= E_0 - \int n(r) \left[V_{ext}(r) - V'_{ext}(r) \right] dr \tag{2.23}$$

Adding Eq. 2.22 and Eq. 2.23, results in the contradiction which proves the theorem:

$$E_0 + E'_0 < E_0 + E'_0 \tag{2.24}$$

A summation of the two ground state energies, show that each external potential results in discrete electron density values. The time-independent Schrödinger equation may now be rewritten in terms of electron density, instead of wave function, using the Hohenberg-Kohn functional, $F_{HK}[n(r)]$, which gives a description of the inter-electronic interactions and electronic kinetic energy:

$$\widehat{H} = \widehat{H}_{int} + V_{ext} \tag{2.25}$$

$$E = \left\langle \psi \left| \widehat{H} \right| \psi \right\rangle = F_{HK}[n(r)] + \int V_{ext}(r)\, n(r)\, \mathrm{dr} \tag{2.26}$$

The variational theorem, seeks to ensure that the calculated density is the ground state density, by stating that $F_{HK}[n]$ solves for the ground state energy only if the input density is the true ground state density. The idea is that a trial electron density, associated to an external potential, results in an energy which represents the upper bound to the ground state energy, $(E_0 < E)$. A shortcoming of this method, however, is the fact that the form of the functional F_{HK} is unknown and cannot be easily approximated. The Kohn-Sham approach aims to resolve this by defining the exchange-correlation functional.

2.1.3.3 Kohn-Sham

The Kohn-Sham[6] model works by considering the electrons found in a system as non-interacting. This allows for the derivation of equations, described by single Slater determinant wave function, for the electron density, n(r):

$$n(r) = \sum_{i=1}^{N_{elec}} |\psi_i(r)|^2 \tag{2.27}$$

2.1.3.4 Exchange-Correlation Functionals

The reinstatement of orbitals in KS, increases the complexity of the calculations from 3 to 3N. This permits the calculation of the kinetic energy, by implementing a non-interacting electron assumption, similar to HF. In theory, the exact kinetic energy is calculated as shown in Eq. 2.28, via a Slater determinant composed of orbitals $\psi_i(r)$:

$$T_s = \sum_{i=1}^{N_{elec}} \left\langle \psi_i(r) \left| -\frac{1}{2}\nabla^2 \right| \psi_i(r) \right\rangle \tag{2.28}$$

However, in reality electrons are interacting and the above equation does not provide the total kinetic energy, granted the difference between the two may be negligible. A general formula for the energy expression of DFT is given below, with the exchange correlation term assuming the remaining kinetic energy. Here, the TF theory is implemented in obtaining the kinetic energy of the system and in combination with the KS theory, result in the overall DFT energy:

$$E_{DFT}[n(r)] = T_s[n(r)] + V_{ne}[n(r)] + V_H[n(r)] + E_{xc}[n(r)] \tag{2.29}$$

Where, E_{xc} = exchange-correlation term, V_{ne} = attraction of electrons to nuclei, and V_H = Hartree potential;

$$V_{\text{ne}}\left[n\left(r\right)\right] = -\sum_{i}^{N_{\text{nuclei}}} \int \frac{Z_i n\left(r\right)}{\left|R_i - r\right|} \text{d}r \tag{2.30}$$

$$V_H\left[n\left(r\right)\right] = \frac{1}{2} \iint \frac{n\left(r\right)n\left(r'\right)}{\left|r - r'\right|} \text{d}r\text{d}r' \tag{2.31}$$

$$E_{\text{xc}}\left[n\left(r\right)\right] = \left(T_\Sigma\left[n\left(r\right)\right] - T_\text{s}\left[n\left(r\right)\right]\right) + \left(V_{\text{ee}}\left[n\left(r\right)\right] - V_H\left[n\left(r\right)\right]\right) \tag{2.32}$$

Where V_{ee} = electron-electron repulsion, R_i = atomic positions, and Z_i = atomic number.

$$T_\Sigma\left[n\left(r\right)\right] = T_{\text{ex}}\left[n\left(r\right)\right] + T_2\left[n\left(r\right)\right] + T_4\left[n\left(r\right)\right] + \ldots \tag{2.33}$$

$$T_2\left[n\left(r\right)\right] = \frac{1}{9} \int \frac{\left|\nabla n\left(r\right)\right|^2}{8n\left(r\right)} dr \tag{2.34}$$

$$T_4\left[n\left(r\right)\right] = \left(540\left(3\pi\right)^{\frac{2}{3}}\right)^{-1} \int n^{\frac{1}{3}}\left(r\right) \left\{ \left(\frac{\left|\nabla^2 n\left(r\right)\right|}{n\left(r\right)}\right) - \frac{9}{8}\left(\frac{\left|\nabla^2 n\left(r\right)\right|}{n\left(r\right)}\right)\left(\frac{\left|\nabla n\left(r\right)\right|}{n\left(r\right)}\right)^2 + \frac{1}{3}\left(\frac{\left|\nabla n\left(r\right)\right|}{n\left(r\right)}\right)^4 \right\} \text{d}r \tag{2.35}$$

Where, T_2 includes the $\frac{\left|\nabla n(r)\right|^2}{8n(r)}$ *von Weizsäcker* correction on the TF equation for the kinetic energy of a system.[14]

The exchange-correlation functional may be used to replace the difficult-approximated $F_{\text{HK}}\left[n(r)\right]$ from HK theory, with simple approximations, such as the local density approximation (LDA), described below.

2.1.3.5   Local Density Approximation

The Local Density Approximation (LDA) offers a way to accurately approximate E_{xc}, invoking how it is equivalent to the integral of the space with an exchange-correlation energy density at a given point assumed to be the same as in a homogeneous electron gas with identical density (ε_{xc}),[15, 16] such that:

$$E_{\text{xc}}\left[n\left(r\right)\right] = \int n\left(r\right)\varepsilon_{\text{xc}}\left[n\left(r\right)\right]\text{d}r \tag{2.36}$$

An issue faced when implementing the LDA, is the disregard of corrections to the exchange-correlation energy because of nearby inhomogeneities in the electron density.[17] Modifications have been made to the method, for the inclusion of spin-polarised systems and nonhomogeneous electron gases, using the Local Spin Density Approximation (LSDA) and Generalised Gradient Approximation (GGA).

2.1.3.6   Generalised Gradient Approximation

The Generalised Gradient Approximation (GGA) improves upon the uniform electron gas approximations, through the introduction of the dependency of the exchange-correlation energy on the gradient of the electron density, through the inclusion of the density's derivatives as variables. One of the earliest and most widely implemented GGA exchange functional is the B88 (or B), which was introduced as a correction to LSDA:[18]

$$\varepsilon_{\text{xc}}^{B88} = \varepsilon_{\text{xc}}^{\text{LDA}} + \Delta\varepsilon_{\text{xc}}^{B88} \tag{2.37}$$

The most extensively used GGA correlation functional, is LYP.[19] The equation used to describe this functional is rather extensive, hence not included here – the four empirical parameters used in the equation are derived from fitting to data for the He-atom. This functional does not include parallel spin correlation, when they are all aligned, and it is often combined with other GGA-functionals, such as B88.

The use of meta-GGA functionals permits the expansion of GGA functionals, by allowing the dependency of the exchange and correlation functionals on higher-order derivatives of the electron density. This is achieved through the inclusion of the Laplace operator of the electron density or the local kinetic energy density, $\nabla^2[n(r)]$.

2.1.3.7 Hybrid-GGA

The issue that comes with the aforementioned DFT functionals is the poor description of the exchange functional. Considering the exact estimation of this using HF, an obvious approach to overcome the electronic self-interactions, that result in the inaccuracies, is through the combination of DFT and HF functionals, such that: $E_{\text{xc}} = E_x^{\text{HF}} + E_c^{\text{KS}}$. However, the combination of the correlation part from DFT and exchange part from HF result in even less accurate outcomes for molecular systems. To understand the reason behind this, we must first consider the Fermi, Coulomb and total holes. These concepts refer to the quantum phenomena, which are a result of the Pauli exclusion principle, ensuring that two electrons with the same half-integer spin occupy a different position in space.[20] The Fermi holes describe the probability of finding an electron of a certain spin at a distance away from another electron of the same spin. In the case of Coulomb holes, the probability of finding an electron of one spin at a distance away from an electron of the opposite spin, is considered.[21]

The contribution of the delocalised exchange and correlation components to the total hole, has been beautifully illustrated in the past using H_2,[22, 23] thus I will be utilising the same example to conceptualise what is happening. Considering the Fermi hole here, removing the self-interaction leaves half the total density delocalised over the whole molecule. This, in turn, can be represented by the removal of half an electron from the vicinity of each nucleus. This becomes an issue at large distances, where the same contribution is removed. Figure 2.1, shows the Fermi, Coulomb and total holes for H_2. At large distances, $R_{\text{H-H}}$, even though both components are delocalised, the total hole is localised at the reference electron, which still experiences partial screening, ending up with a diffuse orbital.
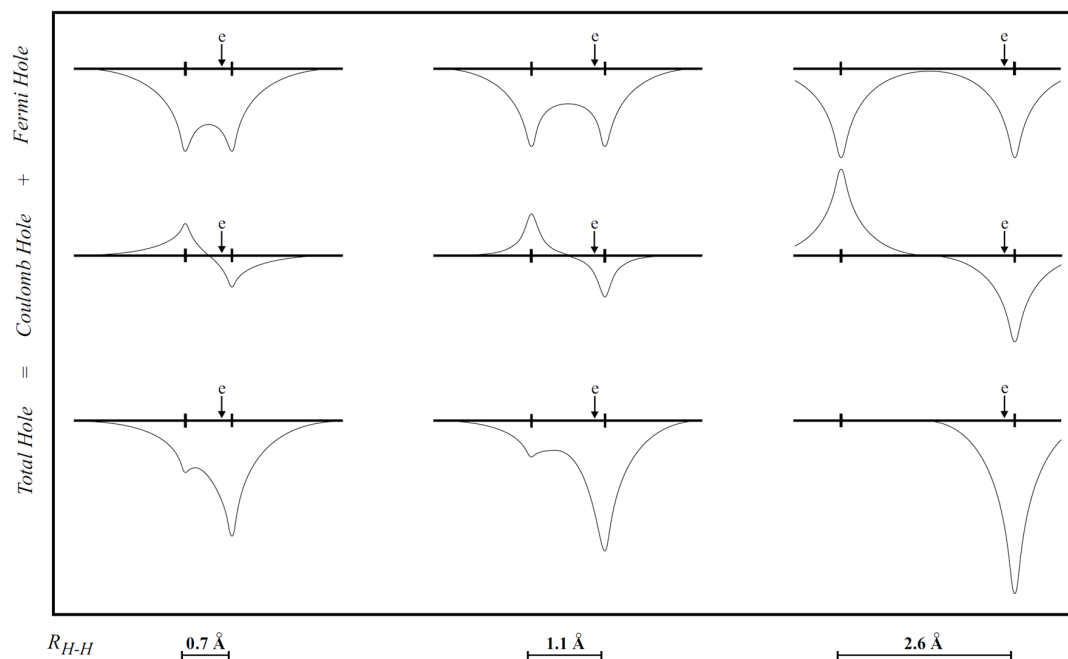
Figure 2.1: Fermi, Coulomb and total holes for H_2 at different distances, $R_{H\text{-}H}$, with the position of the electron marked with an arrow (the depth of peaks and troughs is representative of the density of states). Figure adopted from source.[23]

The inclusion of both components, each representing delocalised reference electrons, results in a poor representation of the whole, as the total hole is relatively localised, especially at large inter-nuclear distances.[23] Hence, an alternative approach to describe systems with a hybrid functional, is through the implementation of contributions from the exact exchange energy of HF and a combination of exchange and correlation energies from LSDA and GGA functionals. Due to the primary contributions from GGA functionals to this method, the resulting functionals are referred to as hybrid-GGA.

One of the most exploited hybrid functional in research, also used in the present work, is the B3LYP.[18, 19] This includes a mixture of LDA and GGA functionals from DFT and HF methods, resulting in the exchange-correlation functional equation below:

$$E_{\text{xc}}^{B3LYP} = E_{\text{xc}}^{\text{LDA}} + \alpha_0 \left(E_x^{\text{HF}} - E_x^{\text{LDA}} \right) + \alpha_x \left(E_x^{\text{GGA}} - E_x^{\text{LDA}} \right) + \alpha_c \left(E_c^{\text{GGA}} - E_c^{\text{LDA}} \right) \tag{2.38}$$

Where, $\alpha_0 = 0.20$, $\alpha_x = 0.72$, $\alpha_c = 0.81$, constituting the three empirical parameters derived from fitting the predicted values to atomization energies, ionization potentials, proton affinities and total atomic energies.[18] The E_x^{GGA} and E_c^{GGA} components are from the B88 and LYP functionals, respectively.

2.1.4 Basis Sets

Basis sets are used in quantum mechanics, through the application of LCAO, for calculation of the electronic structure of molecules. The primary objective when using basis sets, is the representation of the molecular wave function, through expansion of single-electron wave functions into matrices. This, in turn, is used to compute an approximation for the Schrödinger equation. Basis sets comprise of basis functions, which are generally centred on the atoms, hence why

sometimes they are referred to as atomic orbitals. An unknown function, such as a molecular orbital, may be expanded using a set of known functions. However, the only way for that function to not be considered an approximation, is if the basis set is complete; a practically impossible feat, as a complete basis set (CBS) would entail an infinite amount of basis functions. From this statement we infer how the larger the basis set is, the better and more accurately represented the unknown functions will be. By extension, the better a single basis function is in representing the unknown function, the fewer are necessary for enhancing the accuracy. For the description of atomic orbitals, two different types of basis functions may be used: Slater-Type Orbitals (STOs)[24] and Gaussian-Type Orbitals (GTOs).[25] Plane-wave (PW) basis sets may also be used in certain cases, such as when implementing DFT calculations. These work by considering multiple PW pseudopotentials for the individual elements in a system, generating electron density functionals, used in the calculations. The concepts are described in the sections that follow.

2.1.4.1 Atomic Orbitals

Atomic orbitals are generally represented by two types of basis functions, STOs and GTOs. These may be presented by the functionals with general form presented in the equations that follow, where Eq. 2.39 may be used for the STO, while Eq. 2.40 and Eq. 2.41 express GTOs in terms of polar and Cartesian coordinates, respectively. Figure 2.2 illustrates the different plots that may be derived from each of the basis functions.

$$\chi_{\zeta,n,l,m}^{\text{STO}}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{n-1}\,e^{-\zeta r} \tag{2.39}$$

$$\chi_{\alpha,n,l,m}^{\text{GTO}_p}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{2n-2-l}\,e^{-ar} \tag{2.40}$$

$$\chi_{\alpha,i,j,k}^{\text{GTO}_C} = Nx^l y^m z^n e^{-\alpha r^2} \tag{2.41}$$

Where, N = normalising constant, n = natural number which is used as the principal quantum number, $Y_{l,m}$ = spherical harmonic functions, exponents ζ and α are used to define the radial size of the basis functions, values x, y, z describe the local cartesian coordinates, while quantum numbers l, m and n signify the angular shape and direction of the orbital ($l + m + n = \ell$ [orbital quantum number]) and r represents the radial distance to the atomic centre.
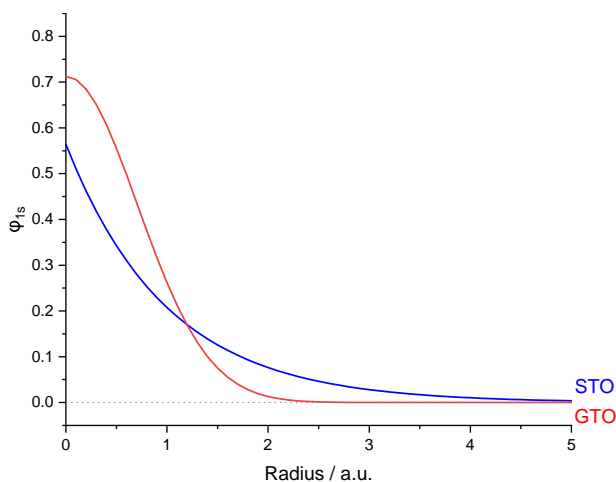
Figure 2.2: Plot of STO vs GTO functions, using an exponent value of 1.0.

From the equations given above, it could be argued that STOs offer better accuracy, compared to GTOs, due to the r^2 dependence of the functionals. Generally, either of the functionals may be implemented to construct a CBS, however about three times more GTOs are required to emulate the accuracy achieved when using STOs.[2] Nevertheless, the reduced integrals required by a GTO-focused basis set, more than compensate for the need of more basis functions. Hence, based on computational efficiency, GTOs are generally preferred over STOs. The computational economy of the GTOs may be increased with a linear combination of primitive GTOs (PGTOs), which may be used to form a basis set contraction, with the basis functions referred to as contracted GTOs (CGTOs). The energy of contracted basis sets is always higher, due to variational parameter restrictions, decreasing the flexibility of the basis set. Hence, such basis sets are usually only employed when the loss of accuracy is compensated by the gain of computational efficiency; and also at points where the effect may be masked from more significant calculations. After determining the basis functions that will be used for constructing a basis set, the number of basis functions that will be used must be considered. These are hence explored in the proceeding sections.

2.1.4.1.1 Minimal Basis Sets   Minimal basis sets, also referred to as single ζ (Sζ), utilise the smallest amount of basis functions required for a basis set. An example of a commonly used Sζ basis set is STO-3G, which may be employed to offer rough approximations of core and valence orbitals. The designation of the basis set usually gives an indication of the functionals used. For this particular one, STO stands for the orbital function, while 3G suggests the use of a linear combination of three Gaussian functions for the characterisation of the STOs. STO-3G works by employing the least-square representation of STOs as a sum of GTOs. An issue that arises from the use of such basis sets, is the assumption of a fixed ζ exponents, where all orbitals are thought to occupy identical spatial spread.[9] Hence, a more elaborate type of basis sets may be used for accountancy of the variance in the space occupied by core and valence orbitals.

2.1.4.1.2 Split-Valence Basis Sets   The split-valence basis sets are an improvement on the Sζ, that work by splitting the basis functions, allowing for the valence electrons to be described using two GTOs, each with an optimised ζ value, and a single GTO for the core electrons. Hence, permitting the inclusion of multiple functions, utilised to denote individual orbitals, by variation of the ζ exponent. The 6-31G is an example of a double ζ (Dζ) split-valence basis set. The way this basis set is expressed may be broken down to deduce how the functions are implemented: where the core orbitals are a contraction of six PGTOs, the inner and outer parts of the valence orbitals are a contraction of three and one PGTOs, respectively. The use of two numbers after the hyphen signify a Dζ basis set. Diffuse and polarisation functions may also be implemented in the basis sets, and are denoted with the use of plus signs (+) and asterisks (*), respectively, after G. The diffuse functions consist only of *s*- and *p*-functions, with one + signifying their use on non-H atoms, and ++ their use on H-atoms as well. The polarisation functions employ *d*-type polarisation (shown in Figure 2.3) on non-H and H-atoms, denoted by one and two asterisks, respectively – these may also be represented with (d) and (d,p) notation, for * and **.
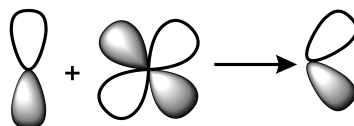
Figure 2.3: d-type polarisation function added to p-orbital.

In the present work, the B3LYP exchange-correlation functional is employed with the 6-31G* basis set, for the parametrisation of the transition metal site.

2.2 Semiempirical Quantum Mechanics

While molecular mechanic methods of simulation make it possible to simulate large systems, without the time limitation encountered when implementing quantum mechanical methods, they often fall short when dealing with metal ions, protonation states, polarisation and proton transfer. Thus, semiempirical quantum mechanical (SQM) methods have been developed, as a way to simulate large systems, benefitting from the quantum mechanical treatment of electrons within realistic timeframes.

Semiempirical methods work by introducing simplifications and empirical parameters to QM methods, on which they are based. This translates to a significantly noticeable increase in the length of a dynamically simulated trajectory that can be obtained using these methods, as opposed to QM.

2.2.1 Tight-Binding Density Functional Theory

The tight-binding DFT (DFTB) model,[26] includes approximations to the DFT quantum mechanical method, with the aim of improving the efficiency of calculations, with minimal loss of accuracy. DFTB manages to only model the valence electrons explicitly, with the implementation of a minimal basis set. Since most of the integrals are either neglected, further approximated or subsituted by parameters, the bulk of the computational cost is used for the diagonalization of the Fock matrix – an approximation to the Hamiltonian – though considering the use of a minimal basis set, it is still way less computationally demanding than having to perform a complete quantum mechanic calculation with a basis set.

Derivations of DFTB are primarily based on a form of LDA approximations of DFT. Most of these methodologies lack a definition of long-range interactions. A recently developed model expanding on the framework of the highest-level variant, DFTB3, seeks to improve upon the current semiempirical methods, through the addition of electrostatic and exchange-correlation functionals, whilst incorporating the handling of dispersion energies, through the implementation of the D4 dispersion model.[27–29] This is permitted through the use of the Vydrov and van Voorhis non-local density-dependent functional kernel (VV10),[30] along with a general approximation to expand the Kohn-Sham equation, such that:

$$E[\rho] = \int \rho\,(r) \left[T[\rho\,(r)] + V_n\,(r) + \epsilon_{xc}^{LDA}[\rho\,(r)] + \frac{1}{2} \int \left(\frac{1}{|r - r'|} + \Phi_C^{NL}\,(r, r') \right) \rho\,(r')\,dr' \right] dr + E_{nn}$$

$$(2.42)$$

Where, Φ_C^{NL} is the VV10 for the long-range correlation, and E_{nn} is the nuclear repulsion energy. The D4 dispersion correction method makes use of the Atomic Mulliken charges (i.e. electronic charge distribution of the atoms) set out in the GFN-xTB semiempirical method,[31] which parameterizes elements at Z≤86; thus improving the interpretation of polar, organometallic and ionic systems from the D3 dispersion coefficient method, through the addition of atomic charge information.[27] The GFN2-xTB method[29] was later developed to include the D4 dispersion corrections, allowing for better handling of the electronic structure effects on the dispersion energies.

2.3   Molecular Mechanics

The concepts below aim to describe the fundamental theories implemented in molecular dynamics models.[a] Explanations of the concepts described were mainly derived from computational chemistry textbooks.[2, 32–34]

2.3.1   Force Fields

Molecular dynamics models aim to simulate the dynamical behaviour of a system at a finite temperature. Force fields are implemented in molecular mechanic methods as a way to define the interactions between atoms. Molecular dynamics simulations are then used to propagate the dynamics of a system. Force fields aim to define bonding information of atoms, by parametrising the nuclear coordinates to obtain a potential energy surface. Electrons are not considered in these calculations, thus there is no need to solve the electronic Schrödinger equation. The quantum considerations of nuclear motion are also neglected, meaning the dynamics of the system are controlled solely by classical mechanics – Newton's second law. The force field energy is calculated from a compilation of individual energy components, that need to be considered when assuming molecular distortion:

$$E_{\text{FF}} = E_{\text{vdw}} + E_{\text{elec}} + E_{\text{str}} + E_{\text{tors}} + E_{\text{bend}} + E_{\text{oop}} + E_{\text{cross}} \tag{2.43}$$

Where, E_{vdw} and E_{elec} = non-bonded (van der Waals and electrostatic) interactions, E_{str} = stretching energy of a bond, E_{tors} = torsional energy for rotation about a bond, E_{bend} = energy required to bend an angle, E_{oop} = out-of-plane bending, E_{cross} = coupling between E_{bend}, E_{oop}, E_{tors} and E_{str}. Each of these energy functions are detailed in the sections that follow. The equations used to describe each of these functions may differ between force fields. Figure 2.4, illustrates the fundamental force field energies:

---

[a]A note on semantics: *molecular mechanics* refer to the methodology and parametrisation used in modelling the force fields and bonds of the molecules within a system. The sole consideration being the atomic positions, while ignoring electron correlation. *Molecular dynamics* constitute the algorithmic calculations and approximations used to solve the Newtonian equations. Here, the molecular motion is simulated within a time step, permitting the system to converge inside this set limit.
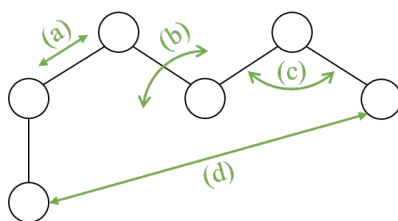
Figure 2.4: Sketch of the fundamental force field energies: (a) stretch, (b) torsional, (c) bend and (d) non-bonded (either electrostatic, when dipoles are present, or van der Waals).

2.3.1.1   Stretch energy

Ideally, the Morse function is used to define the bonded interactions, as shown in Eq. 2.44.

$$E_{\text{str}} = D_e \left\{ 1 - \exp\left[-a \left(R - R_0 \right) \right] \right\}^2 \tag{2.44}$$

$$a = \omega \sqrt{\frac{\mu}{2D_e}} \tag{2.45}$$

$$\omega = \sqrt{\frac{k}{\mu}} \tag{2.46}$$

Where, D_e = depth of potential energy minimum, μ = reduced mass of interacting atoms, k = force constant of interested bond and R_0 = reference bond lengths. The computational cost of the Morse function, however, compels biomolecular force fields to use an approximation. Hence, it is replaced by a Harmonic potential, which utilises Hooke's law to describe the stretching energy:

$$E_{\text{str}} = \frac{k}{2} \left(R - R_0 \right)^2 \tag{2.47}$$

This approximation works relatively well near the equilibrium region, as seen in Figure 2.5, when compared to the Morse function. Some force fields (MMFF, MM3/MM4 and CFF) use anharmonicity corrections to increase the precision for bond lengths of small molecules.
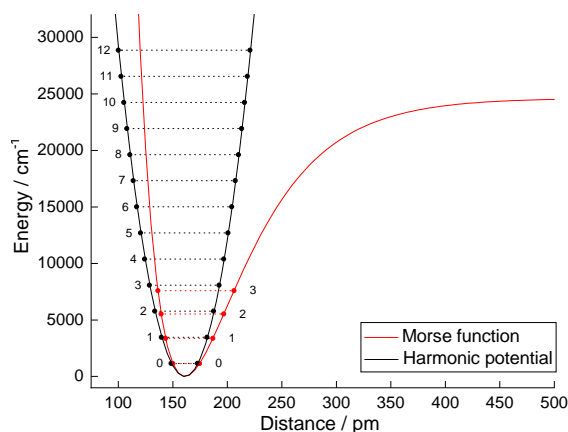
Figure 2.5: Plot of the relative potential energy of the bonded interaction of two atoms, against the distance, R, between them. At E = 0 cm^{-1}, the equilibrium length, R_0, is achieved.

2.3.1.2 Bending energy

The energy required to bend an angle between three atoms A-B-C, is given by the equation
below:

$$E_{\text{bend}} = \frac{k}{2}(\theta - \theta_0)^2 \tag{2.48}$$

This function utilises Hooke's law in the same way as for the stretching energy. The preci-
sion may also increase, in some cases, in the same way as for the stretching energy, using
anharmonicity corrections.

2.3.1.3 Torsional energy

The torsional energy is a result of the rotational barriers around chemical bonds. Figure 2.6
illustrates the torsional angle in a 4-atom A-B-C-D molecule. The equation for the energy
function, defines the energetics of twisting atoms A-D, with V_n = barrier height (i.e. barrier to
rotation) and γ = phase factor (where the cosine function has its minimum). The cosine series
expansion is used here, as it accounts for the periodicity in the torsional potential, owing to the
repeating motion of the angles.

$$E_{\text{tors}}(\omega) = \sum_{n=1}^{N} \frac{V_n}{2}\left[1 + \cos(n\omega - \gamma)\right] \tag{2.49}$$
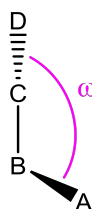
Figure 2.6: Representation of torsional angle (ω).

2.3.1.4 Out-of-plane bending

In the case where three atoms are arranged on the same plane, around a central sp^2-hybridised
atom, one would need to account for the energy cost of pyramidalisation (i.e. distortion from
trigonal planar arrangement to tetrahedral geometry). Instead of implementing large force con-
stants on the angles of the constituting atoms, which would inadvertently lead to rigid in-plane
angles, the out-of-plane bend energy term (E_{oop}) is used. This takes care of the energy compen-
sation for pyramidalisation, while maintaining the in-plane angle force constants to pragmatic
values. The following expression is thus used, to define the out-of-plane bending energy:

$$E_{\text{oop}} = \frac{k}{2}\theta^2 \tag{2.50}$$

Angle θ is defined as the angle between a bond from the central atom and the plane it forms, as
illustrated in Figure 2.7.
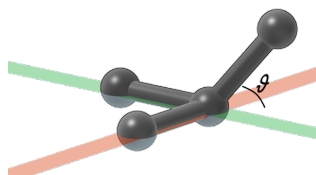
Figure 2.7: Out-of-plane bending in a trigonal planar environment, with labelled θ angle.

2.3.1.5   Cross terms

The cross term energy, E_{cross}, permits interactions between stretching, torsional and bending energy functions. This is done by regarding the deviations of the bond lengths and angles from their reference values, as a result of the atoms interacting. For instance, in the case of a -C(OH) group where the CO and OH bonds are stretched, the distance between the C and H is increased, making it easier to bend the angle. Hence, a stretch-bend cross term may be used to account for this deviation in values. An example of an equation used to describe the derivation of such a cross term, is given below:

$$E_{cross} = \frac{k}{2} \left[(R_{CO} - R_{CO(0)}) + (R_{OH} - R_{OH(0)}) \right] (\theta - \theta_0) \tag{2.51}$$

Where, R_{CO} and $R_{CO(0)}$= new and reference bond length between C and O (and accordingly for R_{OH} and $R_{OH(0)}$), θ and θ_0 = new and reference C-O-H bond angle.

2.3.1.6   Electrostatic interactions

The electrostatic energy function is calculated from the sum of the electrostatic interactions between atoms at a 1,4 or greater relation. The equation that describes this is given below:

$$E_{elec} = \sum_{i<j}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \tag{2.52}$$

Where, N = number of point charges, q_i and q_j = partial atomic charges of atoms i and j, ε_0 = dielectric permittivity of vacuum, ε = dielectric constant, r_{ij} = distance separating the charges. The atomic charges are assigned according to the force field used. In AMBER, for instance, the charges are usually calculated by fitting to HF/6-31G* electrostatic potentials of associated smaller molecules.

2.3.1.7   Van der Waals interactions

The van der Waals function is taken as the sum of all the atom-pair interactions, involving atoms at 1,≥4. The 1,2 and 1,3 interactions (both vdW and electrostatic) are implicitly included in the bond bending and stretching parameters. Each vdW term contains the total of London dispersion and Pauli repulsive forces. These repulsive forces become increasingly strong with decreasing distance between the atoms, in quantum mechanical terms, due to the overlap of electron clouds, which result in repulsion between the electrons because of Coulomb and exchange interactions. E_{vdW} is positive at small distances, negative when the distance between the atoms is less than the sum of the vdW radii and approaches zero at large distances. The Lennard-Jones 12-6 potential function,[35] is used to describe the interactions; using R_{ij} for the distance between atoms i and j, σ_{ij} for the collision diameter, calculated as the sum of the vdW

radii of the two atoms (where E_{vdW} passes a minimum), and ε_{ij} for the potential well depth of the pair. Eq. 2.53 describes these interactions, where the term with an exponent of 12 representing the repulsive forces and the one with the exponent of 6 the attractive. These terms are depicted in Figure 2.8, with -/+ to signify the direction of the repulsive and attractive forces, respectively, and F=0 the equilibrium separation distance between atoms i and j. In order to obtain the interaction energy between different atoms, allowing the estimation of non-bonded interactions, the potentials are calculated using the Lorentz-Berthelot mixing rules.[36] This allows for a less computationally demanding calculation of the interactions, by using the geometric mean of the vdW radii and the arithmetic mean of the well depths. Otherwise, the potentials would have to be calculated for each atom pair, by considering each size, charge and potential well, significantly increasing the computational cost of simulations.

$$E_{vdW} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}}\right)^{6}\right] \tag{2.53}$$
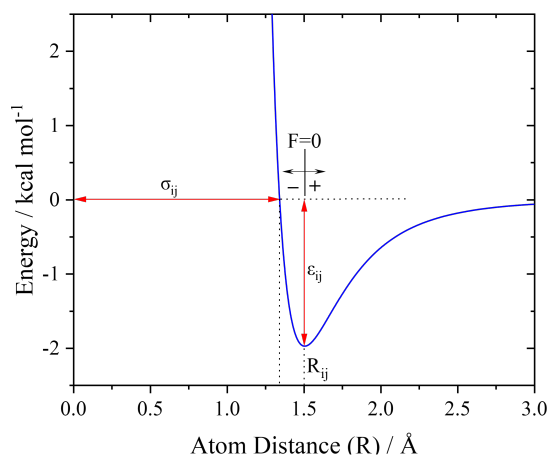
Figure 2.8: Plot based on Eq. 2.53, in accordance with the Lennard-Jones potential. Values used: $\sigma_{ij} \approx 1.34$ Å, $\varepsilon_{ij} = 2.0$ kcal mol^{-1}, resulting at the minimum for $R_{ij} = 1.5$ Å.

2.3.2   AMBER Force Fields

The first introduction of the Assisted Model Building and Energy Refinement (AMBER) force fields, came in 1981, with the establishment of the empirical method as a way of substituting the Born-Oppenheimer approximation for the energy surface of a system, with an analytical function.[37] This approach was put in practise with the development of the united atom force field, ff84,[38] which was later extended to an all-atom force field, ff86.[39] The greatest contribution to the development of the AMBER force fields came in 1995, with the introduction of new charged models, determined using quantum mechanical calculations with the 6-31G* basis set and RESP charge fitting; as well as improved vdW parameters, developed from liquid simulations.[40]

The ff94 force field laid the foundation for the development of more optimised force fields, ff96 and ff99,[41, 42] which aimed to address the secondary structure deficiencies encountered in the ff94, specifically the over-stabilisation of α-helices. This was done through refitting of backbone dihedrals, although it was later observed that the force fields still resulted in over-stabilisation of secondary structural characteristics, arising from incorrect conformational bias

for Gly. The flaw in the two force fields, came from overlooking the sidechain dihedral angles of Gly, branching from Cβ (φ', ψ').[43] Hence, the ff99SB was developed in 2006,[43] as a way to address the incorrect fitting of the dihedral angles, with the introduction of improved dihedrals. The ff03 force field was developed on the ff99 framework, changing the way in which the point charges are derived, through the use of continuum dielectric solvent models in the quantum mechanical calculations.[44] This was done with the intention of emulating solvent polarization, as a way to develop the backbone dihedrals with a reduced preference for helical conformations.

The ff99SB force field saw several modifications over the years, focusing on the way the backbone dihedrals are expressed, either through the inclusion of experimental data in the derivation of the parameters,[45] or through optimisation of the dihedrals to improve the perturbation of the system in explicit solvent.[46] In the present work, the ff99SB-ILDN force field was evaluated, which includes corrections for the side-chain torsional potentials, specifically on the four residues that give it its name – Ile (I), Leu (L), Asp (D), Asn (N).[47]

Most of the force fields developed for AMBER, had been evaluated against the TIP3P water model,[48] although in some cases these fail to reproduce the phase diagram of water. As a way to combat this, the ff03w force field had been developed,[49] through the combination of an AMBER force field derived from ff03 (ff03*),[50] and TIP4P/2005 water model.[51] In the present work, the ff03ws[52] force field was used for the MD simulations of α-Synuclein. It includes modifications to the short-range protein-water interactions, increasing them by a scale of 1.10, which showed an improvement on the R_g of intrinsically disordered systems.

Today, the most widely implemented force field is arguably ff14SB, which is a continuation of the ff99SB, by incorporating small empirical adjustments, to improve the backbone and side chain definitions.[53] This was done by carrying out QM calculations on complete amino acids, contrasted with the small organic compounds used in ff99 and ff99SB. The same study introducing ff14SB, also issued force field ff14SBonlysc, which includes only the updated side-chain dihedrals, as opposed to both backbone and side-chain, in ff14SB.[53] Another attempt at creating a force field able to accurately describe intrinsically disordered proteins, came with ff14IDPSFF.[54] It incorporates grid-based energy correction maps (CMAP), as a way of minimising the distribution difference in dihedral angles. This was done by optimising iteratively the systems, until the molecular dynamics distributions were within 0.10% of the benchmark distributions.

Among the newly developed force fields is fb15,[55] which aimed to increase the conformational space sampled, by improving the estimation of the steepness potential wells. The force field is based on ff99SB and it has been developed by fitting of QM calculations to intramolecular bonds, angles and dihedrals, through potential energy scans. The ForceBalance software was used for parameter optimisation,[56] and the force field has been reported to work best with the TIP3P-FB water model.[56]

2.3.2.1   Non-standard Force Field Parameters

For the parameterisation of the metal ion binding sites, the MCPB tool was used to aid in the derivation of bond, angle and charge parameters,[57] which were combined with the force fields used during the MD simulations. This works by creating small and large models of the

metal ion binding pockets, to ensure that the QM calculations are accurate and come at a low computational cost. The small model is used to obtain angle and bond parameters, while the large calculates the partial charges of the atoms involved in the binding site. When building the small model, there are three different means by which the residues in the metal-containing region are substituted, to reduce the computational cost. For sidechain atoms bound to the metal ion, CH_3R groups are used, in the case of backbone atoms, the residues are replaced with $CO-CH_2R$ or $NH-CH_2R$. Lastly, if both backbone and sidechain atoms of a residue are bound to the metal, then $CO-CH_2R$ and $NH-CH_2R$ groups are used to mimic the residues. When calculating the partial charges of the atoms, the restrained electrostatic potential (RESP)[58–60] fitting scheme is used on the large model (*vide infra*), while the force constants for the newly derived parameters were calculated using the Seminario method,[61, 62] Section 2.3.6.

The QM calculations on the two models constructed via MCPB.py, were performed on Gaussian09.[63] The script then generates a parameter modification script, which is used to incorporate the new terms in the standard force field parameters of choice, using the LEaP[64] program, which outputs the coordinate and topology files used to initiate the MD simulations.

2.3.3 Ligand Field Molecular Mechanics

The first attempt at describing metal complexes, in the context of molecular mechanics, came with the Crystal Field Theory (CFT),[65] which defines metal ions in coordination complexes through their electronic structure. The d-orbital splitting of the metal ions is induced by the electrostatic field, whose symmetry is dependent on the ligands surrounding it. The splitting removes the degeneracies of the orbitals, permitting the evaluation of the orbitals' orientations, by referring to their energies. An illustration of the d-orbital splitting in octahedral arrangements, is given in Figure 2.9:
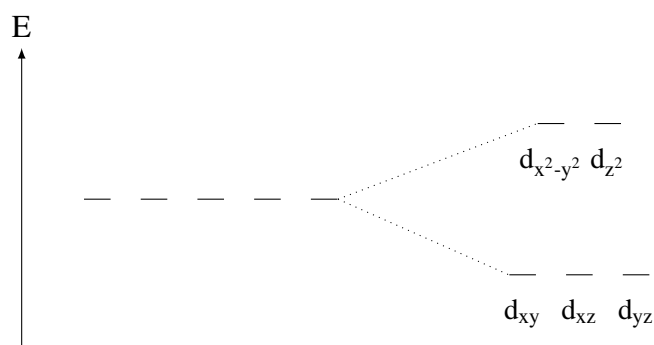
Figure 2.9: Molecular orbital diagram of d-orbital splitting in an octahedral arrangement.

Where CFT is lacking, is in describing the bonding between atoms and transition metals, for this the Ligand Field Theory (LFT) had been developed,[66] as a combination of Crystal Field and Molecular Orbital theories, describing both electronic structure and covalent bonding. The way metals are coordinated to the surrounding ligands or chelating agents, is characterised by the tendency of the overall complex to attain an optimal geometry. These often depend on the coordination number (CN), repulsion between point charges and stereochemically active lone pairs, with the prevailing natural geometries being: tetrahedral (tet; CN=4), trigonal bipyramidal (tbp; CN=5), octahedral (oct; CN=6). In order to define the d-orbital moieties within the MM framework, a Ligand Field Stabilisation Energy (LFSE)[67] term is incorporated in

the force field equation (Eq. 2.43). For the generalisation of this term, the angular overlap model (AOM) was implemented.[*67–70*] The method works by isolating the M-L bonds into local orbitals, simplifying their expression to simple diatomic molecules. The d-orbital energy contribution in each of these is thus expressed, through the diagonalization of a (5×5) matrix. This describes the d-orbitals with respect to the symmetric modes, including one σ- and two π-bonding (in two perpenticular dimensions), angular coordinates and energy parameters as a function of bond length.[*70*]

2.3.4   Restrained Electrostatic Potential (RESP) Charges

One of the key elements in producing a well-defining force field is the correct characterisation of the atoms' partial charges. The idea here is to relate the polarity of a molecule to the charge accumulation and depletion of the constituting atoms. This concept permits the use of the electronic distribution of atoms, disregarding the electrons' wave characteristics. In the community, there is not one universal methodology for the procurement of said charges, instead, they vary with the context in which they are applied.[*34*] The approach to calculating the partial charges used in the present work, follows the methodology set out for the procurement of electrostatic potential (ESP) charges, where points on a grid surrounding the vdW surface are sampled, through the implementation of an algorithm based on the expansion below, where q_k is the charge of the nucleus of atom k, r is the vector of the point in space where ESP is calculated, and r_k is the vector of the nucleus of the atom k [with the absolute difference between the two describing to the distance between the two point vectors]:

$$V_{\text{ESP}}(r) = \sum_{k}^{\text{nuclei}} \frac{q_k}{|r - r_k|} \tag{2.54}$$

A limitation that comes with the ESP charge fitting method, is the differentiation of the hydrogens based on their conformational state. In simple words, for instance, where a methyl group is concerned, the three bonded hydrogens cannot all have the same partial charges. This becomes an issue, where molecular dynamics simulations are concerned, due to the dynamic nature of the method, changing the conformation with each iteration, thus the hydrogens should not maintain the same charge they start with. The RESP approach attempts to account for this through the use of a two-step fit, where initially the fit to the potential is performed with weak restraints and no forced symmetry on methyl groups, followed by a refit of those groups both with strong restraints and forced symmetry. At the first stage, the magnitude of the charges is reduced and the polar areas are well-defined, giving way to the second stage, where forced symmetry is applied to the polar areas, imposing fixed charges to them, while permitting the readjustment of charges of non-polar regions (i.e. methyl groups) to the forced symmetry.[*40, 58–60*]

2.3.5   Optimising General Functions: Geometry Optimisation

Computational chemistry methods tend to rely on the discovery of the minima of functions to conduct dynamical simulations of systems. The simple approach to go about doing this, is by treating each variable independently until the function reaches a local minimum in a unitary matrix, U. The computational cost of this approach, however, makes it a highly inefficient way to do this.

An alternative to this, is the *Simplex* method proposed by Dantzig et al.,[*71, 72*] which has

been implemented as a way of propagating the energy convergence, while handling all the variables at once. This is achieved through the construction of an irregular polyhedron using the function values, and shifting it towards the minimum, while allowing its contraction or expansion to enhance convergence. While this method is better than the simple approach described above, it still falls short when implemented for many-dimensional functions.

Hence, the methods described below have been used by MD simulation methods to aid in this respect. A common aspect of the methods is how the first derivative, the gradient vector ∇U (U = potential energy), is calculated analytically and not as a numerical differentiation, by stepping the variables. This is achieved by assuming a set of 3N-6 first partial derivatives of U, with respect to its variables. The points observed at a local minimum $\nabla U=0$, are referred to as stationary.

2.3.5.1   Steepest Descent Algorithm

The steepest descent algorithm[73] commences through calculation of U and ∇U, from the initial geometry. The vector ∇U points in the direction where the U function displays the most increase. Hence, the value of U can be lowered by stepping in the opposite direction. Considering this, the first step is taken in the direction of $-\nabla U$. After each step, minimisation continues through a new direction, orthogonal to the previous one. After each step, the points on $-\nabla U$ are fitted onto a polynomial and hence its minimum calculated, propagating the algorithm to the next step, where the same method is used until the global minimum is reached – though it is practically never achieved using the steepest descent method, instead the rate of convergence will slow down as the values decrease. Thus, when both the gradient and predicted step size attain a negligible value, then the minimum is assumed to have been reached. This concept is loosely demonstrated in Figure 2.10.

Even though the steepest descent method is a good way of reaching a minimum, it becomes more ineffective as the steps approach this, partly due to the discount of gradient components not found in the engaged steps. Hence, if a disregarded gradient had a lower value, then the energy would be lowered further, however this is never sampled. This results in the algorithm to only ever locate function minima. Thus, gaining efficiency as it only stores the gradient vector. This method is often only used at the beginning of the conformation search, where the minimum point is still not reached and the vector gradient is large. The method of local minima identification then switches to the conjugate-gradient algorithm, described below.

2.3.5.2   Conjugate Gradient Algorithm

The conjugate gradient method[74] follows the same principles set out by the steepest descent algorithm, in that the second geometry (g_2) optimised structure is computed through the deviation of the initial geometry (g_1) from the product of the vector ∇U and eigenvalue λ_1, calculated from a line search, such that:

$$g_2 = g_1 - \lambda_1 \nabla U_1 \tag{2.55}$$

Rationalised for $i > 1$, to:

$$g_{i+1} = g_i + \lambda_i q_i \tag{2.56}$$

Where:

$$q_i \equiv -\nabla U_i + \beta_i q_{i-1} \tag{2.57}$$

$$^b\beta_i \equiv \left(\frac{\nabla U_i}{\nabla U_{i-1}}\right)^2 \tag{2.58}$$

The equations used in conjugate gradient method, permit the assumption of a new step in a direction conjugate to those already used in the previous steps, thus maintaining the minimisation of the previous steps. The technical interpretation of conjugacy is not explored here, but roughly it describes how the initial search direction is obtained from the steepest descents, and the newly constructed directions are orthogonal to all the previous directions.[76]
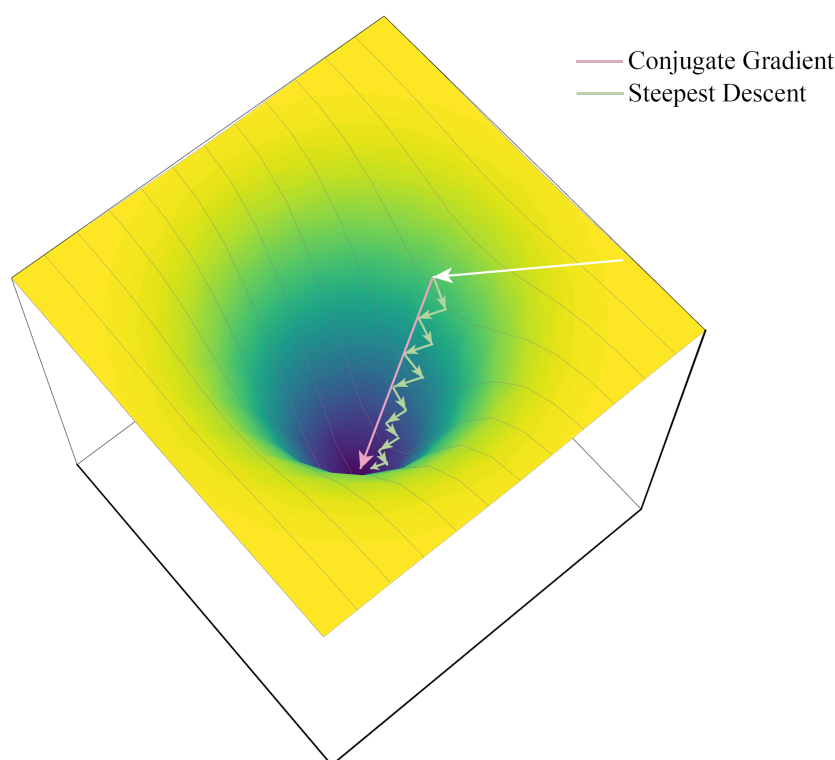
Figure 2.10: Step direction during the conjugate gradient and steepest descent minimisation process.

2.3.6   Seminario Method

The parametrisation of the peptide chain is accomplished through the use of the aforementioned force fields. This is achieved by fitting the molecular mechanic force field parameters to quantum mechanics or experimental data. When modelling a peptide with transition metals bound onto it, the force fields often lack in describing this region. A common approach for transition metal parametrisation is the use of fitting methods specifically designed for this purpose. Although these may have inherent limitations in terms of accuracy, they address the challenges related to the transferability and interdependency of force field parameters, as they can be adjusted specifically for the transition metal and surrounding environment of interest.

---

^{b}Note that the equation for β shown here, is specific to the Fletcher-Reeves associated method[75] and does not represent the correct way of calculating it for all the different methods.

The MCPB.py script[57] has been used here to derive updated force field parameters for the metal ions, which uses the Seminario method,[61] which constitutes one of the most widely integrated non-fitting methods to date. Seminario makes use of QM Hessian matrix projections, for determining force constants for the molecular mechanic force fields.

2.3.6.1   Background

The primary purpose of the Seminario method, developed in 1996, is to parametrise harmonic bonds and angles, from the quantum mechanical Hessian matrix of the biomolecule.[77] This permits the calculation of intramolecular parameters from QM data. Firstly, the reaction force $\delta \mathbf{F}$, with 3N components, occurring from a displacement $\delta \mathbf{p}$, in a system containing N atoms, is expressed to second order, as:

$$\delta \mathbf{F} = -[k]\,\delta \mathbf{p} \tag{2.59}$$

Where, [k] represents the Hessian as a tensor with dimensions $3N \times 3N$, and defined as:

$$[k] = k_{\mathrm{ij}} = \frac{\partial^2 E}{\partial \mathbf{p}_i \partial \mathbf{p}_j} \tag{2.60}$$

Hence, permitting the reiteration of Eq. 2.59, explicitly, as:

$$
\begin{bmatrix}
\delta \mathbf{F}_1 \\
\delta \mathbf{F}_2 \\
\delta \mathbf{F}_3 \\
\vdots \\
\delta \mathbf{F}_{3N}
\end{bmatrix}
= -
\begin{bmatrix}
\frac{\partial^2 E}{\partial \mathbf{p}_1^2} & \frac{\partial^2 E}{\partial \mathbf{p}_1 \partial \mathbf{p}_2} & \frac{\partial^2 E}{\partial \mathbf{p}_1 \partial \mathbf{p}_3} & \cdots & \frac{\partial^2 E}{\partial \mathbf{p}_1 \partial \mathbf{p}_{3N}} \\
\frac{\partial^2 E}{\partial \mathbf{p}_2 \partial \mathbf{p}_1} & \frac{\partial^2 E}{\partial \mathbf{p}_2^2} & \frac{\partial^2 E}{\partial \mathbf{p}_2 \partial \mathbf{p}_3} & \cdots & \frac{\partial^2 E}{\partial \mathbf{p}_2 \partial \mathbf{p}_{3N}} \\
\frac{\partial^2 E}{\partial \mathbf{p}_3 \partial \mathbf{p}_1} & \frac{\partial^2 E}{\partial \mathbf{p}_3 \partial \mathbf{p}_2} & \frac{\partial^2 E}{\partial \mathbf{p}_3^2} & \cdots & \frac{\partial^2 E}{\partial \mathbf{p}_3 \partial \mathbf{p}_{3N}} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 E}{\partial \mathbf{p}_{3N} \partial \mathbf{p}_1} & \frac{\partial^2 E}{\partial \mathbf{p}_{3N} \partial \mathbf{p}_2} & \frac{\partial^2 E}{\partial \mathbf{p}_{3N} \partial \mathbf{p}_3} & \cdots & \frac{\partial^2 E}{\partial \mathbf{p}_{3N}^2}
\end{bmatrix}
\begin{bmatrix}
\delta \mathbf{p}_1 \\
\delta \mathbf{p}_2 \\
\delta \mathbf{p}_3 \\
\vdots \\
\delta \mathbf{p}_{3N}
\end{bmatrix}
\tag{2.61}
$$

The potential energy function of the system, in terms of its internal variables, is expressed below:

$$E = \sum_{\text{bonds}} \frac{1}{2} k_r (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{1}{2} k_\varphi (\varphi - \varphi_0)^2 + \sum_{\text{oop}} \frac{1}{2} k_\omega (\omega - \omega_0)^2 \tag{2.62}$$

Where, r, θ, φ and ω signify the bond lengths, bond angles, dihedral angles and out-of-plane dihedrals, respectively, along with their equilibrium values expressed by r_0, θ_0, φ_0 and ω_0. The corresponding k values represent the force constants of each of the variables. For force fields that include cross-terms, the value of the component is taken as the covariance of the two variables multiplied by the cross-term force constant.

The use of DFT or *ab initio* methods permits the derivation of tensor [k]. The eigenvalues of this correspond to the 3N force constants in the direction of the eigenvectors; these are used to describe the 3 translational, 3 rotational and 3N-6 vibrational modes of the system.

2.3.6.2   Methodology

The interactions between the atom pairs are analysed, with the aim of determining the ones which are pairwise stable. An analogy is used in this section to illustrate the way in which

these interactions are examined. For atom A, with a force felt due to the displacement of atom B, the interaction is described by Eq. 2.59, such that:

$$\delta \mathbf{F_A} = -[k_{AB}]\,\delta \mathbf{p_B} \tag{2.63}$$

A (3×3) matrix may then be constructed, by employing Eq. 2.61, after considering the reaction force $\delta \mathbf{F_A} = (\delta \mathbf{F}_{xA},\ \delta \mathbf{F}_{yA},\ \delta \mathbf{F}_{zA})$ on atom A, from the displacement $\delta \mathbf{p_B} = (\delta \mathbf{x_B},\ \delta \mathbf{y_B},\ \delta \mathbf{z_B})$ of atom B:

$$[k_{AB}] = -\begin{bmatrix} \frac{\partial^2 E}{\partial \mathbf{x}_A \partial \mathbf{x}_B} & \frac{\partial^2 E}{\partial \mathbf{x}_A \partial \mathbf{y}_B} & \frac{\partial^2 E}{\partial \mathbf{x}_A \partial \mathbf{z}_B} \\ \frac{\partial^2 E}{\partial \mathbf{y}_A \partial \mathbf{x}_B} & \frac{\partial^2 E}{\partial \mathbf{y}_A \partial \mathbf{y}_B} & \frac{\partial^2 E}{\partial \mathbf{y}_A \partial \mathbf{z}_B} \\ \frac{\partial^2 E}{\partial \mathbf{z}_A \partial \mathbf{x}_B} & \frac{\partial^2 E}{\partial \mathbf{z}_A \partial \mathbf{y}_B} & \frac{\partial^2 E}{\partial \mathbf{z}_A \partial \mathbf{z}_B} \end{bmatrix} \tag{2.64}$$

The three eigenvalues, λ_i^{AB}, resulting from $[k_{AB}]$, constitute the force constants in the direction of the eigenvectors, $\vec{v}_i^{AB}$. When interpreting these, a positive value signifies a parallel direction, for the reaction force on A from a displacement of B in the $\vec{v}_i^{AB}$ direction. If all eigenvalues are positive, then for any displacement of B there exists a reaction force on A that aims to sustain the initial AB separation, portraying a pairwise stable combination of atoms. If, however, at least one negative eigenvalue exists, then A will not experience a restoring reaction force, to preserve the interatomic separation, hence describing a pairwise unstable situation, where A and B are not bonded. An eigenvector that is in the direction of A and B, described by the unit vector $\vec{u}^{AB}$, and formed in an orthogonal fashion to the other two eigenvectors, may be defined as:

$$k_{AB} = \lambda_i^{AB} \tag{2.65}$$

When describing eigenvectors that are not orthogonal to each other, then the above equation may be altered to include projections of the eigenvectors onto the direction of the unit vector:

$$k_{AB} = \sum_{i=1}^{3} \lambda_i^{AB} \left| \vec{u}^{AB} \times \vec{v}^{AB} \right| \tag{2.66}$$

2.4 Molecular Dynamics

2.4.1 Newtonian Motion and Integrators

As mentioned in Section 2.3, the dynamical behaviour of the nuclei in a simulated system, are governed by Newton's second law:

$$F = m_i \alpha_i = m_i \frac{\partial^2 r_i}{\partial t^2} \tag{2.67}$$

Where, F is the force acting on particle i with mass m_i and acceleration α_i. The right-most part in the equation is the second derivative of coordinate r_i with respect to time. The Verlet algorithm [78] can hence be used to integrate the equation of Newtonian motion, through the use of Taylor expansions. These are implemented to describe the movement of particles at small forward, $+\Delta t$, or reverse, $-\Delta t$, time steps from positions r_i, given by Eq. 2.68 and Eq. 2.69, respectively.

$$r_i(t + \Delta t) = r_i(t) + v_i(t)(\Delta t) + \frac{1}{2}a_i(t)(\Delta t)^2 + \frac{1}{6}b_i(t)(\Delta t)^3 + ... \qquad (2.68)$$

$$r_i(t - \Delta t) = r_i(t) - v_i(t)(\Delta t) + \frac{1}{2}a_i(t)(\Delta t)^2 - \frac{1}{6}b_i(t)(\Delta t)^3 + ... \qquad (2.69)$$

Where, velocities v_i are the first derivatives of the coordinates with respect to time, $\frac{\partial r_i}{\partial t}$, accelerations a_i are the second derivatives, $\frac{\partial^2 r_i}{\partial t^2}$, jerks b_i the third derivatives, and so on.[2] The Verlet algorithm is hence derived, through the summation of the two Taylor expansions, omitting every odd-order derivative:

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + a_i(t)(\Delta t)^2 \qquad (2.70)$$

Thus, the current and previous positions of a particle, along with its acceleration, can be used to determine each subsequent position. The Verlet algorithm is useful in molecular mechanics when the property of interest is independent of momentum, i.e. when looking at just the position coordinates of the phase space. This is due to the algorithm not calculating velocity explicitly. In turn, it cannot be used to generate ensembles at a constant temperature.[34] To remedy this, a leapfrog integrator is used to propagate the position and velocity vectors, through the use of Taylor expansions, analogous to those in Eq. 2.68 and Eq. 2.69, now performed at half a time step and *subtracted*, to give the following expression:

$$r_i(t + \Delta t) = r_i(t) + v_i\left(t + \frac{1}{2}\Delta t\right)\Delta t \qquad (2.71)$$

With similar expansions for velocity:

$$v_i\left(t + \frac{1}{2}\Delta t\right) = v_i\left(t - \frac{1}{2}\right) + a_i(t)\Delta t \qquad (2.72)$$

The above equations can be used to define the modifications introduced by the leapfrog integrator, with the position and velocity vectors out of phase by half a time step. The explicit appearance of velocities in the leapfrog integrator, permits the coupling to an external bath, allowing the implementation of thermostats to control the temperature of the system.[79, 80] Some of the drawbacks that come with the leapfrog algorithm, include the lack of third-order terms in Taylor expansions, and the fact that the positions and velocities of a system cannot be known at the same time, as there will always be a displacement of half a time step. The velocity Verlet algorithm may be used,[81] however, to remove this variance in time steps, through the propagation of the particles by the following equations:

$$r_i(t + \Delta t) = r_i(t) + v_i(t)(\Delta t) + \frac{1}{2}a_i(t)(\Delta t)^2 \qquad (2.73)$$

$$v_i(t + \Delta t) = v_i(t) + \frac{a_i(t) + a_i(t + \Delta t)}{2}\Delta t \qquad (2.74)$$

Other algorithms have also been developed to combat the drawbacks of the leapfrog and Verlet algorithms, through solution of differential equations in a time-reversible manner, improving the energy conservation. These, however, do not come without their own limitations, specifically when it comes to efficiency; these are beyond the scope of this thesis, hence not detailed here. Nevertheless, a different approach to utilising the integrators described here, and increasing the stability of the numerical integration has been devised, by removing the high-frequency

motions in a system. To this end, the SHAKE algorithm is used to constrain the bonds of heavy atoms to hydrogens, maintaining them at a constant length, increasing the integration stability.[82] A more in-depth description of this algorithm is given below.

2.4.2 SHAKE Algorithm

The SHAKE algorithm is used in biomolecular system simulations to impose holonomic constraints on the motion of atoms. The first introduction of the algorithm came in 1977 by Ryck-aert et al.[82] – the derivation of the algorithm is briefly discussed here. Initially, the position vector $\mathbf{r}_i$, of atom i is considered in terms of the vector of the positional constraint, k, of atoms 1 and 2, where k_1 and k_2 are implemented in calculating the distance that separates the two atoms, $\mathbf{r}_{k_1 k_2}$; while $\mathbf{d}_{k_1 k_2}$ signifies the constraint distance. The terms are hence combined in Eq. 2.75, where l is used to describe the holonomic constraints:

$$\sigma_k\left(\{r_i\}\right) = r_{k_1 k_2}^2 - d_{k_1 k_2}^2 = 0, \quad k = 1, \ldots, l \tag{2.75}$$

The force acting on the constraints may therefore be separated in two components: the force arising from the potential energy $\boldsymbol{F}_i$ and the force of constraint $\mathbf{G}_i$ developed from all the constraints σ_k. $\mathbf{G}_i$ is then described by the equation below:

$$G_i = -\sum_{k=1}^{l} \lambda_k(t)\, \nabla_i \sigma_k \tag{2.76}$$

Where, Lagrange's method of undetermined multipliers is implemented for l, using λ_k dependant on time. The constraints are considered in succession, with the algorithm correcting the position of atom i as it does so. Either a leapfrog or velocity Velvet algorithm is used to integrate the equations of motion, resulting in displacement due to the constraint forces. The Lagrange multipliers, along with the displacements, are solved iteratively by SHAKE.

2.4.3 Time step Selection

One of the key parameters that must be considered in MD simulations is the time step, which dictates how frequently the position of the atoms are updated. The time step should be fast enough to capture the fastest motions in the system (e.g. bond vibrations), but not too fast to increase the computational cost. The Nyquist–Shannon sampling theorem,[83] is a simple 'rule of thumb' to estimate a value for this. The theorem states that the sampling frequency should be at least twice the highest frequency of the signal being sampled. In the case of MD simulations, this means that the time step should be at least half of the period of the fastest bond vibration (i.e. highest frequency component). For instance, in the case of a C-H bond, where the vibrational frequency is around $3000\,\text{cm}^{-1}$ (8.99×10^{13} Hz), which translates to *ca.* 11 fs, an appropriate time step would be <5 fs. It should be noted that this does not guarantee stable and accurate simulations, in fact it has been found that a more appropriate time step is in the order of one-tenth the period of the fastest bond vibration.[84] When using the SHAKE algorithm to impose constraints on these high frequency vibrations (i.e. those involving hydrogen atoms), the majority of all-atom simulations have opted for a 2 fs time step.

2.4.4   Cut-off Distance and Ewald Summation

In molecular mechanics, the description of non-bonded interactions in a system, constitute the most time-consuming part to be calculated. This comes as a result of the bonded interactions developing linearly with the system size, as opposed to the non-bonded, E_{elec} and E_{vdW}, which grow exponentially in the order of N^2, for a system with N particles.[85, 86] On that note, a conjecture is introduced in molecular mechanics, where the vdW interactions are smoothly set to zero. This transition, set to occur at a user-specified cut-off distance, incorporates a simple 'switching function,' that is 1 at short distances and reaches 0 at the cut-off distance. The issue that arises from the use of cut-off distances, is that it has been shown that the vdW's energy converges at a distance of *ca.* 20 Å. The most commonly used cut-off, is in the order of 10 Å, as an increase to 20 Å significantly reduces the computational efficiency.[2] An alternative to the use of cut-off distances, is the Ewald summation technique for the evaluation of long-range contributions in periodic systems.[87] A scaling is applied to the selected level of accuracy, the 'Ewald sum limit,' resulting in the Particle-mesh Ewald (PME) method, reducing the energy calculations to $Nlog(N)$. In the present work, PME has been utilised in all the molecular dynamics simulations performed in explicit solvent, owing to their periodic nature.

2.4.5   Solvent Models

When simulating biomolecular interactions, an important thing to consider is the solvated environment, in which they are modelled. Water has long been used to account for the solvation effects a peptide may experience. Arguably, the most reliable approach to simulate a solvated environment, would be using quantum mechanical methods on a large number of solvent molecules, however this is currently not feasible, due to the computational cost that it would entail. Implicit and explicit solvent models are hence employed as a way to simulate such environments, while maintaining low computational costs. Upon doing so, the Gibbs free energy of solvation ΔG^o_{sol} is considered, with a limit set to ensure the behaviour is that of an 'ideal solution.'[34]

$$\Delta G^o_{\text{sol}}(A) = \lim_{[A]_{\text{sol}} \to 0} \left\{ -RT \ln \left. \frac{[A]_{\text{sol}}}{[A]_{\text{gas}}} \right|_{\text{eq}} \right\} \tag{2.77}$$

2.4.5.1   Implicit Solvent Models

Implicit solvent models have been the primary choice for estimating the solvation free energies of simulated systems. Most of these models aim to reproduce the energies obtained from explicit solvent simulations of large systems. This is mostly due to their computational efficiency, at a considerably minimal loss of accuracy, when compared to the water being explicitly modelled.

Implicit solvents may also be referred to as continuum models, as the charge distribution is substituted by a continuous electric field, which surrounds the solute. The simplest way to describe the interactions taking place is using the solvent-accessible-surface-area (SASA) model. The van der Waals surface, created from intersecting vdW radii of the solute, is encompassed by such models, with SASA tracing the outer layer of the vdW surface. The most commonly used radius for the SASA's water molecules is 1.4 Å.[33] This value is combined with the vdW radius of atom (or group) i, to calculate the portion of SASA centred at i. The sum of the values

may then be combined with the atomic surface tension (or atomic solvation parameters), constant σ, found after fitting known solvation free energies, measured under standard conditions. Ultimately, the solvated system generates a charge distribution, which polarizes the electric field.

An assessment of the accuracy of simulations, when using implicit or explicit water models is also studied in the chapters that follow, and evaluated against the different force field parameters.

2.4.5.1.1   The Poisson Equation   Implicit solvent models rely extensively on the Poisson equation, which permits the expression of the electrostatic potential (φ), through the implementation of the charge density of the solute (ρ) and dielectric constant (ε).[34]

$$\nabla^2 \varphi(r) = -\frac{4\pi\rho(r)}{\varepsilon} \tag{2.78}$$

The equation is reconstructed to account for the solute charge distribution, which is considered to be inside a cavity, displacing the dielectric medium. This results in the separation of the equation in two parts, one inside and the other outside the cavity:

$$\nabla\varepsilon(r)\,\nabla\varphi(r) = -4\pi\rho(r) \tag{2.79}$$

The above equation, stands true for systems with zero ionic strength. To account for mobile electrolytes in the solvent, the Poisson-Boltzmann (PB) equation is used instead:[88, 89]

$$\nabla\varepsilon(r)\,\nabla\varphi(r) - \varepsilon(r)\,\lambda(r)\,\kappa^2 \frac{k_B T}{q} \sinh\left[\frac{q\varphi(r)}{k_B T}\right] = -4\pi\rho(r) \tag{2.80}$$

Where, λ = switching function (either zero or one, for inaccessible and accessible regions by the electrolyte), q = charge of ions, and κ^2 = Debye-Hückel (DH) parameter:

$$\kappa^2 = \frac{8\pi q^2 I}{\varepsilon k_B T} \tag{2.81}$$

Where, I = ionic strength. Eq. 2.80 is often adapted for systems with low ionic strength, simplifying it by employing a power series expansion of the hyperbolic sine function, resulting in the linearised PB equation:

$$\nabla\varepsilon(r)\,\nabla\varphi(r) - \varepsilon(r)\,\lambda(r)\,\kappa^2 \varphi(r) = -4\pi\rho(r) \tag{2.82}$$

The equation above solves exactly for the electrostatic field of a charge distribution, however it comes at a high computational cost, thus approximations must be made to permit its integration in MD simulations.

2.4.5.1.2   Generalised Born Model   The simplest approximation to the linearised PB equation comes with the Generalised Born (GB) approach,[90, 91] also used in the current work, which describes the solvation of monatomic ions, by calculating the solvation free energy (G_{solv}).

$$G_{\text{solv}} = -\frac{1}{2}\left(1 - \frac{exp(-\kappa f_{\text{ij}}^{\text{GB}})}{\varepsilon}\right)\sum_{\text{ij}}^{N} \frac{q_i q_j}{f_{\text{ij}}^{\text{GB}}} \tag{2.83}$$

Where, ε represents the dielectric constant, q are the partial charges of particles i and j, and f_{ij}^{GB} is the GB function, defined below:[92]

$$f_{ij}^{GB} = \sqrt{r_{ij}^2 + R_i R_j \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right)} \tag{2.84}$$

Where, r_{ij} = distance between the two atoms, R = effective Born radii of atoms. For an isolated ion, the effective Born radius is equivalent to its vdW radius (ρ). Considering how the DH screening parameter (κ) is assumed to be zero for pure water, Eq. 2.83 may be simplified to:

$$\Delta G_{sol} = -\frac{q_i^2}{2\rho_i}\left(1 - \frac{1}{\varepsilon}\right) \tag{2.85}$$

The effective radii depend on the conformation of the solute, hence they are calculated throughout the trajectory of the system. An approximation to increase the computational efficiency of this process is incorporated using the Coulomb field approximation (CFA), that acts by replacing the electric displacement around the atoms by the Coulomb field. The equation of charge distribution (G) shown below, is expressed in terms of the electric displacement (D) and the scalar product of the electric field (E), with a boundary (Ω) separating the solute from the solvent.

$$G = \frac{1}{8\pi} \int_{\Omega} E \cdot D \, dx \tag{2.86}$$

$$D \approx \frac{q_i r}{r^3} \tag{2.87}$$

Hence, Eq. 2.87 expresses the CFA, through the approximation that the electric displacement is in Coulombic form.

This method tends to fall short when used in macromolecules, where the effective radii may be underestimated, owing to the treatment of vacuum-filled crevices as being filled with water, especially where 'buried' atoms (i.e. atoms in the interior of macromolecules, not accessible to solvent) are concerned. This is an issue that had to be addressed here, when looking at α-Synuclein (αS), due to the extensive structure of the peptide. An approach to solve this shortcoming, involves implementing rescaling parameters proportional to the degree of burial (I_i),[93] such that:

$$R_i^{-1} = \widetilde{\rho}_i^{-1} - \rho_i^{-1} \tanh\left(\alpha\Psi - \beta\Psi^2 + \gamma\Psi^3\right) \tag{2.88}$$

Where, $\widetilde{\rho}_i = \rho_i - offset$, $\Psi = I_i\widetilde{\rho}_i$, and α, β, γ are adjusted accordingly. The implicit solvent that employs this rescaling in AMBER and used in the simulations of αS in the present thesis, is the Onufriev, Bashford, Case (OBC) model.[93]

2.4.5.2 Explicit Solvent Models

In contrast to the implicit solvation methods, where the solvent is treated as a continuous medium, explicit water models rely on the atomistic description of the molecules which constitute the solvent. Explicit models depend on the partition function to describe the statistical thermodynamics of the solvent. For a proper partition function calculation, a large number of solute/solvent configurations must be taken into consideration. MD methods reduce the configurations needed, by exploiting the equations of motion to propagate a system.

Three-site models are characterised by three interaction points, with the nuclear sites used for partial charges. Four-site models aim to improve the electrostatic distribution of the solvent molecules, through the incorporation of a negatively charged dummy atom near the oxygen, illustrated in Figure 2.11(B). In both cases, the oxygen atom carries the Lennard-Jones parameters, in AMBER force field sets.
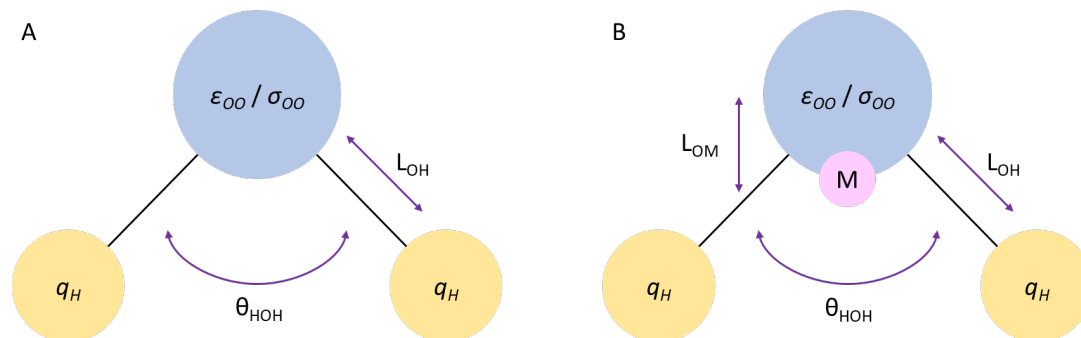
Figure 2.11: Schematic of (A) three- and (B) four-site water models.

Table 2.1: Optimised parameters of the explicit water models used in the present thesis, from the original publications at standard conditions (298.15 K, 1.0 atm).

Water model	ρ (g/cm$_3$)	ε_0	ε_{OO} (kJ/mol)	σ_{OO} (Å)	q_H (e)	θ_{HOH} (deg)	L_{OH} (Å)	ΔH_{vap} (kcal mol^{-1})
Experimental	0.997	78.5	–	–	–	104.45	0.9584	10.52
TIP3P[48, 94]	1.002	94	0.6364	3.15061	0.417	104.52	0.9572	10.41
TIP4P[48, 94]	1.001	53	0.648	3.15365	0.52	104.52	0.9572	10.66
TIP4P/2005[51]	0.9979	60	0.7749	3.1589	0.5564	104.52	0.9572	10.89
TIP3P-FB[56]	0.995	81.3	0.65214	3.178	0.42422	108.15	1.0118	10.71

Where, ρ = density, ε_0 = static dielectric constant, q_H = hydrogen charge, θ_{HOH} = H-O-H angle, L_{OH} = O-H bond length, ΔH_{vap} = heat of vaporisation. Terms ε_{OO} and σ_{OO}, refer to the Lennard-Jones (LJ) potential functions for oxygen, also explored in Section 2.3.1.7, with ε representing the depth of the potential well (dispersion force – the deeper the potential well, the greater the interaction), and σ the intermolecular distance where the potential energy is 0 (vdW radius).

2.4.6   Molecular Dynamics Ensembles

Molecular dynamics simulations are often characterised using quantities, such as temperature (T), volume (V), pressure (P), number of particles (N), total energy (E), and chemical potential (μ). A breakdown of the typical ensembles that may be applied in MD, is given in Table 2.2, with a graphical representation in Figure 2.12.

The state of a system may be described after specifying the particles' positions and momenta. Each particle associates six coordinates to the three-dimensional space, hence a system of N

particles is defined by *6N* coordinates. The 'phase space' of a system is hence described as the *6N*-dimensional space, resulting from these coordinates.

Table 2.2: Breakdown of common ensembles, with coloured blocks indicating the regulated parameters in each case.

N	*T*	*V*	*E*	*P*	*μ*	Acronym	Equilibrium	Name
		█	█			*NVE*	Maximised S^{c}	Microcanonical
█	█	█				*NVT*	Minimised F^{c}	Canonical
	█	█			█	*μVT*	Minimised *PV*	Grand canonical
█	█			█		*NPT*	Maximised G^{c}	Isothermal-isobaric^{d}
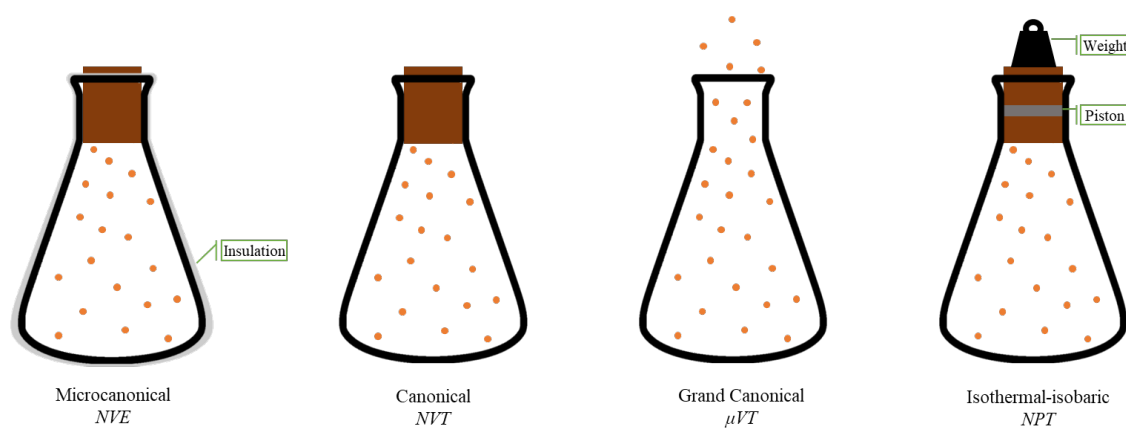
Figure 2.12: Representation of statistical ensembles.

A more detailed description of the two ensembles employed in the MD simulations of this thesis (i.e. *NVT* and *NPT*), is given in the sections that follow.

2.4.6.1   Canonical Ensemble

The canonical ensemble aims to maintain control over the volume and temperature, through direct temperature scaling at the beginning and temperature bath coupling. The partition function is described in Eq. 2.89, and is obtained after integrating the Hamiltonian of the interacting particles, over the momentum contribution.

$$Q(N, V, T) = \frac{1}{\Lambda^{3N}N!} \int \exp\left[-\beta U\left(s^{N}; L\right)\right] ds^{N} = \frac{1}{\Lambda^{3N}N!} \widetilde{Z} \tag{2.89}$$

Where, the set of coordinates scaled by the box length, *L*, are expressed by s^{N}, with the de Broglie wavelength defined by: $\Lambda = \frac{h}{\sqrt{2\pi mkT}}$ and the Boltzmann constant (k_{B}) used to formulate β, such that: $\beta = 1/k_{B}T$. Term $\left(s^{N}; L\right)$ may also be expressed as r^{N} for the positional coordinates.

---

^{c}*S* = entropy, *F* = Helmholtz free energy, *G* = Gibbs free energy.
^{d}Also known as Gibbs ensemble, owing to its state function.

The thermodynamic potential of the ensemble is related to the partition function and defined by the Helmholtz free energy (F):

$$F(N, V, T) = -k_B T \ln \left[\widetilde{Z} \right] \tag{2.90}$$

The probability distribution is described using Boltzmann's law, the Hamiltonian of the system $\widehat{H}$, and relates to the partition function using the normalisation factor $\widetilde{Z}$:

$$p^{\text{NVT}} \propto \frac{1}{\widetilde{Z}} \exp \left[-\frac{\widehat{H}}{k_B T} \right] \tag{2.91}$$

The Langevin thermostat is employed in the simulations here, to maintain the temperature of the system at a constant, set value. A damping coefficient, or collision frequency, γ, is used to define the rate of collisions between two species in a set volume. This value is given in number of collisions per picosecond (ps^{-1}) – the physical collision frequency of liquid water in room temperature is *ca.* 50 ps^{-1}.[95] AMBER simulations have shown that the collision frequency for a system, doesn't need to be this high, as a more mild value has been observed to result in an increased stability of symplectic integrators of the system.[96] A leapfrog integrator is used in order to propagate the system dynamics, adjusting the kinetic energy to come in accordance with the harmonic oscillator, resulting in an accurate virial calculation – i.e. half the product of the stress of repulsion or attraction with the distance between two particles.[97]

2.4.6.2 Isothermal-isobaric Ensemble

The *NPT* ensemble samples a constant pressure and temperature phase space. Variations of the volume, in which the molecules are confined, are performed and evaluated based on the enthalpy change, hence determining if these changes will be accepted or rejected. The piston's function, shown in Figure 2.12, is to control the response time of volume fluctuations. Initially the NVT ensemble is expanded, to account for volume fluctuations, such that:

$$\Delta(N, P, T) = \frac{\beta P}{\Lambda^{3N} N!} \int V^N \exp(-\beta PV) \, dV \int \exp \left[-\beta U \left(r^N \right) \right] ds^N \tag{2.92}$$

Defining the Gibbs free energy, as:

$$G = -k_B T \ln \left[\Delta(N, P, T) \right] \tag{2.93}$$

With probability distribution:

$$\begin{aligned} p^{\text{NPT}} &\propto V^N \exp(-\beta PV) \exp \left[\left(-\beta U \left(r^N \right) \right] \right] \\ &= \exp \left\{ -\beta \left[U \left(r^N \right) + PV - N\beta^{-1} \ln(V) \right] \right\} \end{aligned} \tag{2.94}$$

In the simulations carried out here, the Monte-Carlo (MC) barostat was implemented to maintain a constant pressure.[98, 99] This works by sampling individual particles and performing a trial move from r^N to r'^N. The energy difference is then evaluated and a trial volume change is applied after the particle coordinates are scaled. The trial volume is accepted based on:

$$acc(V \to V') = min \left\{ 1, exp \left[-\beta \left(Ur'^N - Ur^N \right) + P \left(V' - V \right) - N\beta^{-1} ln \left(\frac{V'}{V} \right) \right] \right\} \tag{2.95}$$

The particles sampled in MD methods, are bound to grid positions. The r^N values are taken as the distance to the fixed grid positions.

2.4.7   Ergodic Hypothesis

The ergodic theory aims to describe how the composites of a dynamical system will, eventually, visit all the sites in the available space, uniformly and randomly. Hence, permitting the calculation of ensemble mean values by time averages.

$$\overline{E} = \lim_{T \to \infty} \frac{1}{T} \int_0^T E(r_t) \mathrm{dt} \tag{2.96}$$

For the equilibrium states, the ensemble averages are calculated as shown below:

$$\langle E \rangle_{\mathrm{eq}} = \lim_{t \to \infty} \int E(r) \rho_t(r) d\mu \tag{2.97}$$

Where, $r = 6N$ phase space variables, μ = constant associated with r, ρ_t = probability distribution at $(6N + 1)$ dimensions. The ergodic theory holds that $\langle E \rangle_{\mathrm{eq}} = \overline{E}$. Hence, assenting that molecular dynamics are concerned with time averaging; as for sufficiently long times, the initial conditions are ignored when calculating the time average.[100] The issue that is faced, however, is with the infinite time limit, which is required to assume the macroscopic thermodynamic properties of the system. This only permits the use of systems in equilibrium; as for non-equilibrated systems its properties depend on time.[101] Generally, while a long enough trajectory may be enough to obtain equilibrium properties, it is not the optimal way of carrying out molecular dynamics simulations. This is due to the lack of mixing of probabilistic dynamics, which are the deterministic trajectories in phase space that switch at random times to other trajectories.[102] Thus, not accounting for neighbouring trajectories, the equilibrium is not represented adequately, as not all the different types of global behaviour are sampled. Instead of having one single trajectory, the ensemble average is hence preferably computed through multiple, not necessarily long, simulations.

Having constructed the system, replicas are prepared for running multiple simulations. The difference between the replicas comes from the random seed number, assigned by the AMBER software, based on the time the simulations are performed, which dictates the initial velocities of each of the constituting atoms.

2.4.8   Molecular Dynamics Simulations

When examining the dynamical behaviour of biomolecular substances, conventional molecular dynamics (cMD) simulations are often employed, generating trajectories of the motions in the system. The method constitues a combination of the theorems discussed in the preceding sections.Below, a brief discussion is given on the enhanced sampling methods implemented in the present work.

2.4.8.1   Accelerated Molecular Dynamics

Accelerated molecular dynamics (aMD) simulations, employ an advanced sampling technique that aims to flatten the potential energy surface. The manipulation of different variables, enables the bias of sampling during the simulations. This is often done through the addition of a time-independent positive boost potential, for when the system's potential is lower than a reference energy, allowing for that energy barrier to be transcended.[103] This is achieved through the destabilisation of the local energy minima below the threshold energy, permitting the sampling of trajectories of the phase space that wouldn't be explored otherwise.[2] The equations below

help illustrate this concept, with V = potential energy, E = threshold energy, U = biasing potential and α the acceleration parameter that expresses how shallow the altered energy surface is for values below the threshold energy. As the α parameter decreases, the transitions at the low energy states are increased and the free energy landscape smoothed, Figure 2.13.[103]

$$V_{\text{aMD}}(r) = \begin{cases} V(r) & ; V(r) \geq E \\ V(r) + U(r) & ; V(r) < E \end{cases} \tag{2.98}$$

$$U(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))} \tag{2.99}$$

$$E = V_{mean}(r) + \alpha_{tot} \tag{2.100}$$

Where, $\alpha_{tot} = \alpha \times N_{atom}$.
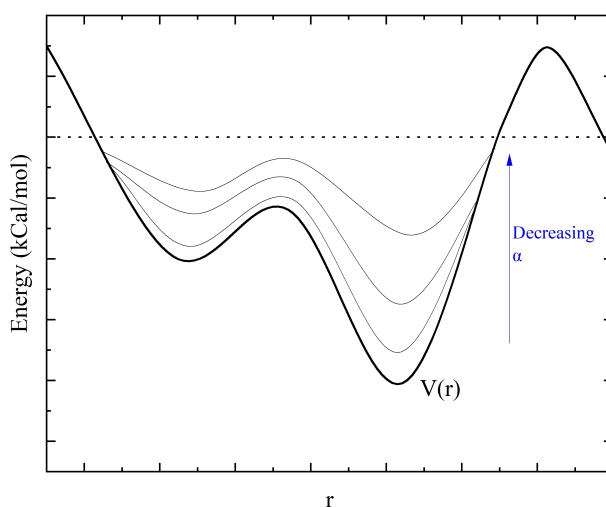
Figure 2.13: Illustration of the accelerated molecular dynamics method. The potential energy surface is altered by the addition of a biasing potential, which is a function of the potential energy. The acceleration parameter, α, determines the steepness of the biasing potential, which is added to the potential energy surface for energies below the threshold energy, E.

2.4.8.1.1 Reweighting The accelerated molecular dynamics method works well in sampling a greater spread of the conformational space, although it does so by altering the system potential, and by extension the canonical distribution remains unsampled. In order to retrieve the canonical ensemble, as well as the free energy profile of the system, reweighting is applied on the aMD trajectories.[103] The canonical distribution, $p(r)$, is related to the probability (or bias) distribution, $p^*(r)$, where r is the atomic positions, by:

$$p(r) \propto p^*(r) e^{\beta U(r)} \tag{2.101}$$

Where, $\beta = 1/k_B T$, and $U(r)$ as the boost potential of each frame. Hence to recover $p(r)$, the equation is adapted to:

$$p(r_j) = p^*(r_j) \frac{\langle e^{\beta U(r)} \rangle_j}{\sum_{j=1}^{M} \langle e^{\beta U(r)} \rangle_j} \tag{2.102}$$

Where, M is the total number of bins, and $\langle e^{\beta U(r)} \rangle_j$ is the average of the boost potential for each bin, j. The equation above defines the exponential average expansion of the bias distribution, which is accurate but it comes with high statistical noise. Two adaptations of the reweighting factor have thus been developed to reduce this noise, the first being the Maclaurin series expansion,[104] and the second the cumulant expansion (which is equivalent to the Maclaurin expansion on the first order).[105, 106] The former has been implemented in the reweighting of the aMD trajectories in the present work, and is given by:

$$\langle e^{\beta \Delta V} \rangle = \sum_{k=0}^{\infty} \frac{\beta^k}{k!} \langle \Delta V^k \rangle \tag{2.103}$$

The free energy surface of the coordinates is defined by the potential of mean force (PMF), which is calculated by $F\left(r_j\right) = -k_B T ln \left[p\left(r_j\right) \right]$.

2.4.8.2   Metadynamics

Much like the accelerated molecular dynamics method, seen above, metadynamics (MTD) is an enhanced sampling technique that aims to improve conformational sampling through the introduction of bias that forces the system out of local minima, and into low-probability states.[107] This is achieved through the addition of adaptive "Gaussian-hills" – given long enough time the conformational space will be a flat surface, having filled the potential wells with these Gaussian functions. In the context of the present work, MTD simulations were performed in the QM/MM framework, with the use of the GNF2-xTB method. The Gaussian functions, are a construct of collective variables, i.e. descriptors of the simulated system, in this case from the root-mean-square-deviation (RMSD) in Cartesian space.[108]

2.4.8.3   Steered Molecular Dynamics

Experimental techniques, such as optical tweezers,[109] have long been used for the assessment of binding properties of biomolecules. The development of computational methods to study protein-ligand interactions has certainly advanced the use of molecular mechanics, especially in the drug discovery process. Steered molecular dynamics (SMD) allows the exploration of the conformational space of such a system, through an irreversible perturbation of the system's potential energy surface and binding affinity. The perturbation is achieved by the addition of a harmonic potential, which is centred on the desired conformation, and is applied to the ligand atoms, pulling them away from the receptor. This has been analogised to the force, F, exerted by a harmonic "spring," and related to the stiffness of the restraint K, and the initial position of the restraint x_0 moving at a constant velocity v:[110]

$$F = K \left(x_0 + vt - x \right) \tag{2.104}$$

The work of the pulling force in the increasing distances between the ligand and receptor, is used to construct a binding affinity plot, by relating the PMF energies to the distance between the two. In the present work, the SMD method has been applied to find the distance where the energy minimum is achieved between the ligand and binding pocket, and use that distance to explore the conformational space of the complex.

2.4.8.3.1   Jarzynski Equality   In adaptive steered molecular dynamics (ASMD),[111] after a set distance variation, the system from each of the simulation repeats, which is closest

to the Jarzynski average,[112] is taken as the starting point for the subsequent transitions. This is calculated through the incorporation of the work done between the start and end of the trajectory,$W_{0 \to t}$, as shown in the following equations.[113]

$$Jar_{avg} = -\frac{log\langle exp(-\beta W_{0 \to t})\rangle}{\beta} \qquad (2.105)$$

Where, β is the inverse temperature, such that $\beta = 1/k_BT$. The Hamiltonian of the system, where the Jarzynski equality is applied, needs to be updated, as $W_{0 \to t}$ is performed on the system undergone the addition of an external potential. This prerequisite to the simulations, accounts for the spring introduced between the two centres of mass, where SMD is applied. This incorporation of a guiding potential, $h(r; \lambda)$, constrains the reaction coordinate, $\xi(r)$, to be within the set proximity to the external parameter, λ:

$$h(r; \lambda) = \frac{k}{2}[\xi(r) - \lambda]^2 \qquad (2.106)$$

Hence, the Hamiltonian, H, is updated to $\widetilde{H}$, where r = coordinates and p = momentum, as follows:

$$\widetilde{H}(r, p; \lambda) = H(r, p) + h(r; \lambda) \qquad (2.107)$$

The potential of mean force (PMF) of the trajectory, which describes the free energy as a function of coordinates, is calculated from the work at the different distances, of the trajectories closest to the Jarzynski average. The PMF is hence defined by the updated work function shown below:

$$W(r) = -k_BT\,ln[g(r)] \qquad (2.108)$$

Where $g(r)$ is the radial distribution function, describing the probability that the coordinate will take a particular value.

2.4.8.4 Analytical Tools

The resulting trajectories from molecular dynamics simulations were primarily analysed using the CPPTRAJ extension program in ambertools 19,[64, 114] except where otherwise specified.

2.4.8.4.1   Root Mean Square Deviation and Radius of Gyration   The root mean square deviation (RMSD) of a trajectory helps illustrate to what degree the system has changed over time, with reference to a selected point. The radius of gyration (R_g), provides information on the molecular shape and compactness of the system over time. This is achieved by considering the root mean square distance of the atoms, comprising the peptide, from their centroid.[115] This results in an overall understanding of the peptide's dimensions. The tensors from the radius of gyration give a good understanding of the distribution of residues in three dimensions, and their resulting shape.

2.4.8.4.2   Diagonalization of Gyration Tensor   The tensor data was used to illustrate the spread of the peptide chains using ellipsoids, although a numerical approach is utilised here for better analysis. The plots were created through diagonalization of the data, after constructing matrices from the tensor coordinates. The matrices allow for the extraction of eigenvalues, which are diagonalized to calculate the radii of the ellipsoids in (x, y, z), and eigenvectors

which are used to illustrate the rotation of the peptide. The diagonalized eigenvalues (λ), relate to the R_g, as shown in Eq. 2.109:

$$R_g^2 = \lambda_x + \lambda_y + \lambda_z \tag{2.109}$$

Hence, permitting the calculation of certain descriptors: globularity, asphericity, acylindricity and anisotropy. The globularity of a system is represented by a value between 0 and 1, with 1 signifying a perfect sphere and values closer to 0 describing a one or two-dimensional object. Asphericity is used to represent the distribution of atoms, by relating the three eigenvalues to each other, where a value of 0 illustrates a spherically symmetric object, with respect to the three coordinates. The acylindricity is also used to describe the distribution of atoms, however, contrary to asphericity, this only uses λ_x and λ_y, hence defining the cylindric symmetry of the system by virtue of these two principal components. As in the case of asphericity, a value of 0 for acylindricity illustrates a cylindrically symmetric distribution along the two axes. Lastly, anisotropy is used to represent how all the points are distributed in the system, with a value of 0 when all points are spherically symmetric and a value of 1 when they lie on a line. The equations for each of these descriptors are defined below:

$$Globularity = \frac{\lambda_x}{\lambda_z} \tag{2.110}$$

$$\text{Asphericity } (b) = \lambda_z - \frac{1}{2}\left(\lambda_x + \lambda_y\right) \tag{2.111}$$

$$\text{Acylindricity } (c) = \lambda_y - \lambda_x \tag{2.112}$$

$$\text{Anisotropy } \left(k^2\right) = \frac{b^2 + \frac{3}{4} \times c^2}{R_g^4} \tag{2.113}$$

2.4.8.4.3 Secondary Structure   The secondary structure of the systems was analysed using the DSSP (Dictionary of Secondary Structure of Proteins) algorithm,[116] implemented within CPPTRAJ, by analysing the backbone conformation, and assigning the secondary structure elements to the residues. The algorithm first identifies the hydrogen bonds within the structure, and then assigns the geometry of those interactions to the folding characteristics. This approach aims to simplify the analysis of the secondary structure elements, by considering a single hydrogen bond parameter, rather than the multiple variables when taking a dihedral angle approach. These hydrogen bond patterns are assessed with regards to the distance between a CO (i) and an NH group ($i+n$), where n-turns are used to define the characteristic, for instance 4-turns for α-helices, relating to Figure 1.3 at the beginning of Chapter 1. Hydrogen-bond 'bridges' between distanced residues, are used to define the β-sheet structure, with alternate-angled hydrogen bonding associated with parallel β-sheet, and parallel hydrogen bonds in antiparallel β-sheet, Figure 1.5.

2.4.8.4.4 Salt Bridges   Salt bridges may be found in peptides, where oppositely charged residues are close enough, for electrostatic interactions to form. The presence of salt bridges contributes to the overall protein structure, without increasing the free energy of unfolding. In reality, salt bridges may often destabilise the folded peptide, due to Coulombic charge interactions being opposed by the desolvation of charges.[117] In molecular dynamics, the salt bridges are defined by the distance between the centre of mass of the oxygen in an acidic side chain and the nitrogen in a basic side chain. Distances for the peptides used in the present thesis above 3.2 Å were disregarded, with the rest constituting the significant interactions, which were thus

turned into a percentage after assuming the sum of the distances and the number of frames that gave those distances.

2.4.8.4.5 Hydrogen Bond Plots   Hydrogen bond plots serve to demonstrate the presence of interactions between an acceptor and a donor residue. These interactions occur when an electronegative atom (acceptor), approaches a hydrogen atom bonded to another electronegative atom (donor). The MDplot package was used to plot the hydrogen bond interactions between the residues of the simulated biomolecules.[118] The persistence of these interactions is presented in relation to the frames where they appear. Multiple hydrogen bonds may be present between two residues. The percentage value used to plot these interactions, is assumed from the most prevalent one; while marking the particular interaction with a black border, to distinguish them from the ones that only share a single hydrogen bond. Interactions with a significance of <5% were disregarded from these plots, for a better evaluation of the most prevalent hydrogen bonds.

2.4.8.4.6 Chemical Shifts   In NMR spectroscopy, the chemical shift is a measure of the resonant frequency of a particular nucleus, relative to a magnetic field. This is influenced by the electron density around the nucleus, along with the electronegativity and, by extension, the local magnetic environment of nearby atoms (i.e. higher electron density, from a highly electronegative atom, shields nucleus from external magnetic field, moving the chemical shift to higher frequencies). The difference between the resonant frequency of an atom and a reference value, defines the chemical shift. In the context of this thesis, secondary chemical shifts are also considered. These are calculated from the difference between an observed chemical shift and the chemical shift expected in random coil conformation. In the case of α-helices, a positive chemical shift value is expected, with β-sheets exhibiting downfield shifts.[119] In the present work, the chemical shifts were predicted using SPARTA+,[120] a software that uses a neural network, trained on 580 proteins, for which high-resolution x-ray structures, and backbone and sidechain ^{13}C chemical shifts are available. Predicted chemical shift values are thus assigned, based on how closely triplets of residues match the structure and sequence of a moiety within the training set.

2.4.8.4.7 Cartesian PCA Clustering   The clustering algorithm employed in the succeeding chapters, was implemented from the carma software package.[121, 122] For efficiency, only the Cα of the constituting residues were used. The program works by sorting the three principal components (PC) of the frames in a trajectory, into a density distribution map. The axes of the map correspond to each of the PC ranks, with the value at each point representing the structures with PC values closest to it. The map is then smoothed using nearest-neighbour averaging, and it is then clustered using a peak-picking algorithm, set to a variable threshold, based on the percentage of the map's variance, from the mean density, aiming at a PC map variance of $\geq$ 80%.

2.4.8.4.8 Markov State Models   The Markov State Model (MSM) is a mathematical model that describes the dynamics of a system in terms of a Markov process.[123] This is a stochastic process that satisfies the Markov property, which states that the future state of the system depends only on the current state, and not on the sequence of events that led to it. The model can be described in terms of a discrete-time Markov chain, where the state space is discretised into a finite number of states, and the transition probabilities between them are calculated. The

MSM can be used to generate synthetic data, which can be used to train other models, such as the Hidden Markov Model (HMM), *vide infra*. Ultimately, the Markov processes can be used to calculate the probability of a sequence of states and events, and thus predict the most favourable future state of the system.

2.4.8.4.8.1 Time-lagged Independent Component Analysis Considering the exceedingly complex features constituting an MD trajectory, a way to reduce the dimensionality of the data is required to better identify discrete states. A common approach is to use Principal Component Analysis (PCA),[124] which is a linear transformation that projects the data onto a lower-dimensional space, where the variance of the data is maximised. While PCA does a good job of efficiently discriminating the coordinates of the atoms in a system, the Time-lagged Independent Component Analysis (TICA) approach improves on previous algorithms by relating the degrees of freedom to the time-correlation for a given lag time.[125]

2.4.8.4.8.2 State Transition Matrix The state transition matrix (STM) describes the probability of transitioning between different states. The STM is calculated by counting the number of times a trajectory transitions between each state, and dividing it by the total number of transitions. The diagonal elements of the square matrix represent the probability of a trajectory remaining in the same state, while the off-diagonal elements represent the probability of a trajectory transitioning from one state to another. In turn, this permits the incorporation of the time parameter, in our case in the analysis of the state-space, *ipso facto* devising the kinetic transformations that take place in the dynamic system.

2.4.8.4.8.3 Hidden Markov Model The Hidden Markov Model (HMM) permits the identification of unobserved phenomena, through the use of a Markov process model. The model works in a similar way to the STM of microstates, only now the transition matrix is estimated from hidden states. The HMM is generally better than MSM at avoiding discretization errors, which is why it has been implemented here for the analysis of the macrostates in the protein folding process.[126]

2.4.8.4.8.4 Chapman-Kolmogorov Test In order to determine whether a given MSM is a good representation of the system, within the chosen lag time, the Chapman-Kolmogorov (CK) test may be used to predict the probability distribution of a system within that lag time. Given a point in time t, the CK test checks the probability distribution of the system at time $t + n$ (i.e. future). The predicted transition probability between the metastable states is then compared to the estimated values from the MD trajectory. If the predicted probability distribution is within a set confidence interval to the estimated probability, then the lag time is considered to be a good representation of the transitions, thus validating the model and ensuring that it accurately represents the dynamics of the system studied.[127]

2.4.8.4.8.5 Mean First Passage Times The mean first passage time (MFPT) is the average time it takes for a stochastic event to occur.[128] In the context of the MSM, the MFPT is the average time it takes for a trajectory to transition from one state to another. The MFPT is calculated by summing the transition probabilities of the trajectory, and dividing it by the total number of transitions. The metric can hence be used for the analysis of the kinetic properties of the system, as it calculates the probability of a trajectory transitioning from one state to another, and thus predict the most favourable future state of the system.

Bibliography

(*1*) Szabo, A., and Ostlund, N. S., *Modern Quantum Chemistry*; Dover Publications, Inc.: Mineola, New York, 1982.

(*2*) Jensen, F., *Introduction to Computational Chemistry Computational Chemistry*; 7, 2017; Vol. 132, p 629.

(*3*) Koch, W., and Holthausen, M. C. In *A Chemist's Guide to Density Functional Theory*, 2nd; Wiley: 2001; Vol. 3; Chapter 4, pp 33–40.

(*4*) Thomas, L. H. (1927). The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society 23*, 542–548.

(*5*) Hohenberg, P., and Kohn, W. (1964). Inhomogeneous electron gas. *Physical Review B 136*, B864–B871.

(*6*) Kohn, W., and Sham, L. J. (1965). Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review A 140*, A1133–A1138.

(*7*) Fermi, E. (1927). Statistical method to determine some properties of atoms. *Rend. Accad. Naz. Lincei 6*, 602–607.

(*8*) Jensen, F. (2013). Atomic orbital basis sets. *Wiley Interdisciplinary Reviews: Computational Molecular Science 3*, 273–295.

(*9*) Hehre, W. J., Stewart, R. F., and Pople, J. A. (1969). Self-consistent molecular-orbital methods. I. Use of gaussian expansions of slater-type atomic orbitals. *The Journal of Chemical Physics 51*, 2657–2664.

(*10*) Berland, K. et al. (2015). van der Waals forces in density functional theory: a review of the vdW-DF method. *Reports on Progress in Physics 78*, 066501.

(*11*) Tkatchenko, A., and Scheffler, M. (2009). Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Physical Review Letters 102*, 6–9.

(*12*) Roothaan, C. C. J. (1951). New Developments in Molecular Orbital Theory. *Reviews of Modern Physics 23*, 69–89.

(*13*) Hall, G. G. (1951). The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 205*, 541–552.

(*14*) Weizsäcker, C. F. (1935). Zur Theorie der Kernmassen. *Zeitschrift für Physik 96*, 431–458.

(*15*) Martin, R. M., *Electronic Structure*; Cambridge University Press: 2004.

(*16*) Jones, R. O., and Gunnarsson, O. (1989). The density functional formalism, its applications and prospects. *Reviews of Modern Physics 61*, 689–746.

(*17*)   Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A., and Joannopoulos, J. D. (1992). Iterative minimization techniques for ab initio total-energy calculations: Molecular dynamics and conjugate gradients. *Reviews of Modern Physics 64*, 1045–1097.

(*18*)   Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics 98*, 5648–5652.

(*19*)   Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B 37*, 785–789.

(*20*)   Pauli, W. (1925). Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik 31*, 765–783.

(*21*)   Buijse, M. A., and Baerends, E. J. In *Density Functional Theory of Molecules, Clusters, and Solids*; Springer Netherlands: Dordrecht, 1996, pp 1–46.

(*22*)   Baerends, E. J., and Gritsenko, O. V. (1997). A quantum chemical view of density functional theory. *Journal of Physical Chemistry A 101*.

(*23*)   Koch, W., and Holthausen, M. C., *A Chemist 's Guide to Density Functional Theory*, 2001; Vol. 3, p 294.

(*24*)   Slater, J. C. (1930). Atomic Shielding Constants. *Physical Review 36*, 57–64.

(*25*)   Boys, S. F. (1950). Electronic wave functions - I. A general method of calculation for the stationary states of any molecular system. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 200*, 542–554.

(*26*)   Wahiduzzaman, M. et al. (2013). DFTB parameters for the periodic table: Part 1, electronic structure. *Journal of Chemical Theory and Computation 9*, 4006–4017.

(*27*)   Caldeweyher, E., Bannwarth, C., and Grimme, S. (2017). Extension of the D3 dispersion coefficient model. *Journal of Chemical Physics 147*, 0–7.

(*28*)   Caldeweyher, E. et al. (2019). A generally applicable atomic-charge dependent London dispersion correction. *Journal of Chemical Physics 150*, DOI: `10.1063/1.5090222`.

(*29*)   Bannwarth, C., Ehlert, S., and Grimme, S. (2019). GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation 15*, 1652–1671.

(*30*)   Vydrov, O. A., and Van Voorhis, T. (2010). Nonlocal van der Waals density functional: The simpler the better. *Journal of Chemical Physics 133*, DOI: `10.1063/1.3521275`.

(*31*)   Grimme, S., Bannwarth, C., and Shushkov, P. (2017). A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1-86). *Journal of Chemical Theory and Computation 13*, 1989–2009.

(*32*)   Gasteiger, J., and Engel, T., *Chemoinformatics: A Textbook*; WILEY-VCH: 2003.

(*33*)   Levine, I. N., *Quantum Chemistry*, 7th Ed.; Pearson: 2013.

(*34*)   Cramer, C. J., *Essentials of Computational Chemistry: Theories and Models*, 2nd Editio, Chichester, 2004.

(*35*)   Jones, J. E. (1924). On the determination of molecular fields. II. From the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 106*, 463–477.

(*36*) Lorentz, H. A. (1881). Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Annalen der Physik 248*, 127–136.

(*37*) Weiner, P. K., and Kollman, P. A. (1981). AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *Journal of Computational Chemistry 2*, 287–303.

(*38*) Weiner, S. J. et al. (1984). A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *Journal of the American Chemical Society 106*, 765–784.

(*39*) Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry 7*, 230–252.

(*40*) Cornell, W. D. et al. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society 117*, 5179–5197.

(*41*) Kollman, P. A. (1996). Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. *Accounts of Chemical Research 29*, 461–469.

(*42*) Wang, J., Cieplak, P., and Kollman, P. A. (2000). How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *Journal of Computational Chemistry 21*, 1049–1074.

(*43*) Hornak, V. et al. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics 65*, 712–725.

(*44*) Duan, Y. et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry 24*, 1999–2012.

(*45*) Li, D. W., and Brüschweiler, R. (2010). NMR-based protein potentials. *Angewandte Chemie - International Edition 49*, 6778–6780.

(*46*) Nerenberg, P. S., and Head-Gordon, T. (2011). Optimizing protein-solvent force fields to reproduce intrinsic conformational preferences of model peptides. *Journal of Chemical Theory and Computation 7*, 1220–1230.

(*47*) Lindorff-Larsen, K. et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics 78*, 1950–1958.

(*48*) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics 79*, 926–935.

(*49*) Best, R. B., and Mittal, J. (2010). Protein simulations with an optimized water model: Cooperative helix formation and temperature-induced unfolded state collapse. *Journal of Physical Chemistry B 114*, 14916–14923.

(*50*) Best, R. B., and Hummer, G. (2009). Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *Journal of Physical Chemistry B 113*, 9004–9015.

(*51*)   Abascal, J. L., and Vega, C. (2005). A general purpose model for the condensed phases of water: TIP4P/2005. *Journal of Chemical Physics 123*, DOI: `10.1063/1.2121687`.

(*52*)   Best, R. B., Zheng, W., and Mittal, J. (2014). Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *Journal of Chemical Theory and Computation 10*, 5113–5124.

(*53*)   Maier, J. A. et al. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation 11*, 3696–3713.

(*54*)   Song, D., Luo, R., and Chen, H.-F. (2017). The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling 57*, 1166–1178.

(*55*)   Wang, L. P. et al. (2017). Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *Journal of Physical Chemistry B 121*, 4023–4039.

(*56*)   Wang, L. P., Martinez, T. J., and Pande, V. S. (2014). Building force fields: An automatic, systematic, and reproducible approach. *Journal of Physical Chemistry Letters 5*, 1885–1891.

(*57*)   Li, P., and Merz, K. M. (2016). MCPB.py: A Python Based Metal Center Parameter Builder. *Journal of Chemical Information and Modeling 56*, 599–604.

(*58*)   Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *Journal of Physical Chemistry 97*, 10269–10280.

(*59*)   Besler, B. H., Merz, K. M., and Kollman, P. A. (1990). Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry 11*, 431–439.

(*60*)   Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995). Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry 16*, 1357–1377.

(*61*)   Seminario, J. M. (1996). Calculation of intramolecular force fields from second-derivative tensors. *International Journal of Quantum Chemistry 60*, 1271–1277.

(*62*)   Lin, F., and Wang, R. (2010). Systematic derivation of AMBER force field parameters applicable to zinc-containing systems. *Journal of Chemical Theory and Computation 6*, 1852–1870.

(*63*)   Frisch, M. J. et al. Gaussian 09, tech. rep., Wallingford, CT: Gaussian, Inc., 2009.

(*64*)   Case, D. et al. AMBER 2019, tech. rep., San Francisco: University of California, 2019.

(*65*)   Bethe, H. (1929). Termaufspaltung in Kristallen. *Annalen der Physik 395*, 133–208.

(*66*)   Griffith, J. S., and Orgel, L. E. (1957). Ligand-field theory. *Quarterly Reviews, Chemical Society 11*, 381.

(*67*)   Deeth, R. J. (2001). The ligand field molecular mechanics model and the stereoelectronic effects of d and s electrons. *Coordination Chemistry Reviews 212*, 11–34.

(*68*)   Woolley, R. G. (1981). The angular overlap model in ligand field theory. *Molecular Physics 42*, 703–720.

(69) Burton, V. J., Kemp, C. M., Gilbert, P. J., and Deeth, R. J. (1995). Molecular Mechanics for Coordination Complexes: The Impact of Adding d-Electron Stabilization Energies. *Journal of the American Chemical Society 117*, 8407–8415.

(70) Deeth, R. J., Fey, N., and Williams-Hubbard, B. (2005). DommiMOE: An implementation of ligand field molecular mechanics in the molecular operating environment. *Journal of Computational Chemistry 26*, 123–130.

(71) Dantzig, G. B., Ordern, A., and Wolfe, P. (1955). The Generalized Simplex Method. *Pacific Journal of Mathematics 5*, 183–195.

(72) Dantzig, G. B., and Wolfe, P. (1961). The Decomposition Algorithm for Linear Programs. *Econometrica 29*, 767.

(73) Debye, P. (1909). Der Lichtdruck auf Kugeln von beliebigem Material. *Annalen der Physik 335*, 57–136.

(74) Hestenes, M., and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards 49*, 409.

(75) Fletcher, R., and Reeves, C. (1964). Function minimization by conjugate gradients. *The Computer Journal 7*, 149–154.

(76) Shewchuk, J. R. An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, tech. rep., Pittsburgh, PA: Carnegie Mellon University, 1994.

(77) Allen, A. E., Payne, M. C., and Cole, D. J. (2018). Harmonic Force Constants for Molecular Mechanics Force Fields via Hessian Matrix Projection. *Journal of Chemical Theory and Computation 14*, 274–281.

(78) Verlet, L. (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review 159*, 98–103.

(79) Hoover, W. G., Ladd, A. J., and Moran, B. (1982). High-strain-rate plastic flow studied via nonequilibrium molecular dynamics. *Physical Review Letters 48*, 1818–1820.

(80) Evans, D. J. (1983). Computer "experiment" for nonlinear thermodynamics of Couette flow. *The Journal of Chemical Physics 78*, 3297–3302.

(81) Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics 72*, 2384–2393.

(82) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics 23*, 327–341.

(83) Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal 27*, 379–423.

(84) Leach, A. R., *Molecular modelling: principles and applications*, 2001, pp 1–744.

(85) Friedrichs, M. S. et al. (2009). Accelerating molecular dynamic simulation on graphics processing units. *Journal of Computational Chemistry 30*, 864–872.

(86) Eastman, P., and Pande, V. S. (2010). Efficient nonbonded interactions for molecular dynamics on a graphics processing unit. *Journal of Computational Chemistry 31*, 1268–1272.

(87) Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics 98*, 10089–10092.

(88)   Gouy, M. (1910). Sur la constitution de la charge électrique à la surface d'un électrolyte. *Journal de Physique Théorique et Appliquée 9*, 457–468.

(89)   Chapman, D. L. (1913). LI. A contribution to the theory of electrocapillarity. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 25*, 475–481.

(90)   Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters 246*, 122–129.

(91)   Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996). Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *The Journal of Physical Chemistry 100*, 19824–19839.

(92)   Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990). Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society 112*, 6127–6129.

(93)   Onufriev, A., Bashford, D., and Case, D. A. (2000). Modification of the generalized born model suitable for macromolecules. *Journal of Physical Chemistry B 104*, 3712–3720.

(94)   Jorgensen, W. L., and Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences of the United States of America 102*, 6665–6670.

(95)   Widmalm, G., and Pastor, R. W. (1992). Comparison of Langevin and molecular dynamics simulations. Equilibrium and dynamics of ethylene glycol in water. *Journal of the Chemical Society, Faraday Transactions 88*, 1747–1754.

(96)   Izaguirre, J. A., Catarello, D. P., Wozniak, J. M., and Skeel, R. D. (2001). Langevin stabilization of molecular dynamics. *The Journal of Chemical Physics 114*, 2090–2098.

(97)   Pastor, R. W., Brooks, B. R., and Szabo, A. (1988). An analysis of the accuracy of langevin and molecular dynamics algorithms. *Molecular Physics 65*, 1409–1419.

(98)   Wood, W. W. (1968). Monte Carlo Calculations for Hard Disks in the Isothermal-Isobaric Ensemble. *The Journal of Chemical Physics 48*, 415–434.

(99)   McDonald, I. (1972). NpT -ensemble Monte Carlo calculations for binary liquid mixtures. *Molecular Physics 23*, 41–58.

(100) Frenkel, D., and Smit, B., *Understanding Molecular Simulation*, 2nd; Elsevier: 2002.

(101) Coveney, P. V., and Wan, S. (2016). On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Physical Chemistry Chemical Physics 18*, 30236–30240.

(102) Devooght, J., and Smidts, C. In *Reliability and Safety Assessment of Dynamic Process Systems*; Springer Berlin Heidelberg: Berlin, Heidelberg, 1994, pp 85–100.

(103) Miao, Y. et al. (2014). Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation 10*, 2677–2689.

(104) Pierce, L. C., Salomon-Ferrer, R., Augusto F. De Oliveira, C., McCammon, J. A., and Walker, R. C. (2012). Routine access to millisecond time scale events with accelerated molecular dynamics. *Journal of Chemical Theory and Computation 8*, 2997–3002.

(*105*) Hummer, G. (2001). Fast-growth thermodynamic integration: Error and efficiency analysis. *Journal of Chemical Physics 114*, 7330–7337.

(*106*) Eastwood, M. P., Hardin, C., Luthey-Schulten, Z., and Wolynes, P. G. (2002). Statistical mechanical refinement of protein structure prediction schemes: Cumulant expansion approach. *Journal of Chemical Physics 117*, 4602–4615.

(*107*) Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America 99*, 12562–12566.

(*108*) Grimme, S. (2019). Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *Journal of Chemical Theory and Computation 15*, 2847–2862.

(*109*) Svoboda, K., and Block, S. M. (1994). Biological applications of optical forces. *Annual Review of Biophysics and Biomolecular Structure 23*, 247–285.

(*110*) Izrailev, S. et al. (1999). Steered Molecular Dynamics. *4*, 39–65.

(*111*) Ozer, G., Valecv, E. F., Quirt, S., and Hernandez, R. (2010). Adaptive steered molecular dynamics of the long-distance unfolding of neuropeptide y. *Journal of Chemical Theory and Computation 6*, 3026–3038.

(*112*) Jarzynski, C. (1992). Diffusion equation for energy in ergodic adiabatic ensembles. *Physical Review A 46*, 7498–7509.

(*113*) Park, S., Khalili-Araghi, F., Tajkhorshid, E., and Schulten, K. (2003). Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of Chemical Physics 119*, 3559–3566.

(*114*) Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation 9*, 3084–3095.

(*115*) Lobanov, M. Y., Bogatyreva, N. S., and Galzitskaya, O. V. (2008). Radius of gyration as an indicator of protein structure compactness. *Molecular Biology 42*, 623–628.

(*116*) Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577–2637.

(*117*) Bosshard, H. R., Marti, D. N., and Jelesarov, I. (2004). Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings. *Journal of Molecular Recognition 17*, 1–16.

(*118*) Margreitter, C., and Oostenbrink, C. (2017). MDplot: Visualise Molecular Dynamics. *The R journal 9*, 164–186.

(*119*) Wishart, D. S., and Case, D. A. In *Methods in Enzymology*; 1994; Elsevier Masson SAS: 2002; Vol. 338, pp 3–34.

(*120*) Shen, Y., and Bax, A. (2010). SPARTA+: A modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of Biomolecular NMR 48*, 13–22.

(*121*) Glykos, N. M. (2006). Carma: A molecular dynamics analysis program. *Journal of Computational Chemistry 27*, 1765–1768.

(*122*)   Baltzis, A. S., Koukos, P. I., and Glykos, N. M. (2015). Clustering of molecular dynamics trajectories via peak-picking in multidimensional PCA-derived distributions. DOI: `10.48550/arXiv.1512.04024`.

(*123*)   Husic, B. E., and Pande, V. S. (2018). Markov State Models: From an Art to a Science. *Journal of the American Chemical Society 140*, 2386–2396.

(*124*)   Jollife, I. T., and Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374*, DOI: `10.1098/rsta.2015.0202`.

(*125*)   Schultze, S., and Grubmüller, H. (2021). Time-Lagged Independent Component Analysis of Random Walks and Protein Dynamics. *Journal of Chemical Theory and Computation 17*, 5766–5776.

(*126*)   Noé, F., Wu, H., Prinz, J. H., and Plattner, N. (2013). Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. *Journal of Chemical Physics 139*, DOI: `10.1063/1.4828816`.

(*127*)   Prinz, J. H. et al. (2011). Markov models of molecular kinetics: Generation and validation. *Journal of Chemical Physics 134*, DOI: `10.1063/1.3565032`.

(*128*)   Polizzi, N. F., Therien, M. J., and Beratan, D. N. (2016). Mean First-Passage Times in Biology. *Israel Journal of Chemistry 56*, 816–824.

Chapter 3

Zn^{II} – Amyloid-$\beta_{1\text{-}16}$

This chapter aims to evaluate the characteristics of a metal-bound biomolecular system, using different force fields, solvents and MD techniques. Amyloid-β (Aβ) has long been associated with the development of Alzheimer's disease (AD). The effects of Zn(II) on this system is of particular interest, as metal ions have been shown to associate with the N-terminus of amyloid-β (A$\beta_{1\text{-}16}$).[1] The 16-amino acid fragment of the amyloid-$\beta_{1\text{-}42}$ peptide, is the main component of the amyloid plaques found in the brains of AD patients. This plaque formation is believed to be the main cause of the neuronal deterioration that occurs in AD. The peptide is known to increase toxicity in the brain,[2] with Zn(II) binding believed to be in-part responsible,[3] while also increasing the aggregation propensity of the peptide and the formation of fibril derivatives. The system studied in this chapter was obtained from experimentally determined structures of the N-terminally acetylated A$\beta_{1\text{-}16}$ peptide interacting with a Zn(II) ion (PDB ID: 1ZE9), through the Nδ of the imidazole rings of His6, His13 and His14, and the carboxylate from Glu11.[4] The ensemble of NMR-defined structures comprises of 20 different conformations, which were used to obtain secondary structure and R_g estimations, permitting comparisons to the computational results obtained here. For the molecular dynamics simulations, the first model was used, as it was determined by the authors to be the closest to average.

3.1 Molecular Mechanic Computational Methods

Different AMBER force field parameters for the amino acids not involved in the metal-binding, were tested throughout the study. The metal ion parameters were obtained using the B3LYP/6-31G(d) basis set,[5] after reworking the system through the MCPB.py program,[6] creating small and large models of the pdb file, as explained in Section 2.3.2.1. In order to exclude atoms not directly involved in the Zn(II) binding, a 3 Å cut-off distance was used when constructing the small model, as a way to limit the computational cost to under 2 hours, using a 32-core CPU. The distance and angle parameters of the metal ion to the coordinating atoms, is given in Tables A.1 and A.2.

For the systems modelled using implicit solvents, sander (simulated annealing with NMR-derived energy restraints) was used to carry out these simulations, while for the explicit solvent ones, pmemd (particle mesh Ewald molecular dynamics) was utilised instead. The distinction was made, having considered the computational efficiency when running MD simulations on periodic systems, with pmemd on GPU nodes. The MD parameters were kept in accordance to similar computational examinations of the system, described in the literature.[7–9]

3.1.1   Implicit Solvent

The generalised Born surface area (GBSA)[10, 11] model was used for most of the force fields (ff99SB-ILDN,[12] ff14SB, ff14SBonlySC,[13] fb15[14]); for ff14SBonlySC, the GB-Neck2 (*igb8*) model was used,[15] as well as the pairwise GBSA. Simulations run with ff03ws,[16] were performed using the OBC modification to the GBSA model.[17] The OBC implicit solvent is mostly used in cases where the system has a more complex structure, which may result to an unoptimized solvent underestimating the effective radii, owing to the treatment of vacuum-filled crevices as being filled with water, especially where 'buried' atoms, in the central region of the peptide, are concerned,[17] this concept has been explored in Section 2.4.5. Despite the relatively small peptide used here, it was still employed for the ff03ws force field, after considering the documented performance of the ff03/OBC combination (and the larger system we will be looking at in the chapters that follow).[18–23] The system was minimised using 1000 steps of steepest descent and subsequent 1000 steps of conjugate gradient algorithms.[24, 25] A cut-off of 12 Å was used, neglecting electrostatic interactions beyond that point.

For the simulations, the NVT ensemble was implemented with the Langevin thermostat,[26] for temperature regulation at 310 K. The SHAKE algorithm[27] was used to impose holonomic constraints to hydrogens, restraining them to their equilibrium values, as specified in the prmtop file, generated by LEaP. The collision frequency (γ) was set to 1 ps^{-1}. The simulations started from the minimised structure coordinates, with random velocities, sampled from a Maxwell-Boltzmann distribution plot. To ensure a well sampled phase space, the conventional molecular dynamics (cMD) simulations were set to run in repeats of three at 125 million steps (time intervals) and 2 fs time step, simulating 250 ns ($\times$3). Having evaluated the RMSD data to estimate the time needed for the systems to equilibrate, looking for consistent fluctuations of the distance between backbone atoms, 50 ns were discounted from each of them and combined into 600 ns trajectory files. A representation of the fluctuations in the RMSD from the cMD runs using ff14SB/GBSA, is shown in Figure A.1.

Having concluded the cMD simulations, the data from each of the force fields was evaluated, in order to select the system more closely resembling the experimental values of secondary structure and R_g, and use that to expand the conformational space sampled, by performing accelerated molecular dynamics (aMD) simulations. These, started from the cMD-finished structure, with the pre-established velocities and coordinates. The system total potential and dihedral energies were boosted, by incorporating the EthreshP (Ep), alphaP(αP), EthreshD (Ed) and alphaD (αD) terms. These were calculated as shown below, and are used to define the threshold energies for the dihedral (Ed) and potential energies (Ep), as well as the total acceleration parameter used in each of them, αP (α=0.16 *kCal mol^{-1} atom^{-1}*) and αS (α=3.5 *kCal mol^{-1} residue^{-1}*), respectively:

$$EthreshP = E_{\text{Total}} + \left(0.16\ kCal\ mol^{-1}\ atom^{-1} \times N_{\text{ATOM}}\right) \tag{3.1}$$

$$alphaP = \left(0.16\ kCal\ mol^{-1}\ atom^{-1} \times N_{\text{ATOM}}\right) \tag{3.2}$$

$$EthreshD = E_{\text{DIHED}} + \left(3.5\ kCal\ mol^{-1}\ residue^{-1} \times N_{\text{RES}}\right) \tag{3.3}$$

$$alphaD = \frac{1}{5} \times \left(3.5\ kCal\ mol^{-1}\ residue^{-1} \times N_{\text{RES}}\right) \tag{3.4}$$

The aMD simulations were performed for 150 million steps, simulating 300 ns. The rest of the parameters were kept identical to those used when running the cMD simulations.

3.1.2   Explicit Solvent

The explicit solvent MD simulations were separated into four processes (heating, two equili-bration steps (one in the NPT and the other in the NVT ensemble), and the production run). The force fields selected to carry out the explicit solvent simulations comprised of ff14SB, fb15, ff99SB-ILDN, and ff03ws. The 'default,' most implemented explicit model is the 3-site transferable intermolecular potential (TIP3P)[28] solvent model, thus it was decided to main-tain that as the standard to examine the changes between the force fields, before moving to more specialised solvent models, some specifically designed for the force fields they are used with. For ff14SB and ff99SB-ILDN, the TIP4P[29] model was tested as well, while for fb15 the TIP3P-FB[30] model was used, and the ff03ws was coupled with TIP4P/2005.[31] The de-cision for which solvent models to use in these simulations was based on examining already reported data and observations from the literature.[12, 14, 32, 33] Irrespective of the solvent model, a periodic box was placed around the peptide, at 10 Å between the atoms and the box edge.

Having obtained the topology and coordinate files for the systems, they were subsequently minimised with identical parameters to those used in the implicit solvent systems. Following that, the systems were heated from 0 to 310 K in the NVT ensemble. The heating parameters included the Langevin thermostat, for temperature regulation, with a collision frequency of 5 ps^{-1}. The higher γ value was chosen in this case to ensure smooth temperature increase, by reg-ulating the transitions in potential energy, which could increase the momentum of the system and spike the temperature above the desired value. The SHAKE algorithm was used to impose holonomic constraints on hydrogens. The cut-off energy of the system throughout the MD simulations was set at 10 Å. The wrapping algorithm was used in all of the MD calculations, to keep the coordinate output from overflowing the trajectory and restart files. This works by "wrapping" the coordinates into a box, allowing for a periodic image of the coordinates to be used (i.e. those which are closest in the middle of the box). The heating step of the simulations was allowed to run for 3 ns (1.5 million steps at 2 fs time step).

Following the heating step, the systems were first equilibrated in the NPT ensemble, using the coordinates and velocities generated in the previous step. Isotropic position scaling was used, with a pressure relaxation time of 1 ps. The target pressure was set at 1 bar and the Monte Carlo barostat was employed to control it. The second equilibration step was in the NVT ensemble, for 20 ns (2 fs time step), with identical parameters to the heating step, but now maintaining the temperature constant at 310 K. The Langevin thermostat was used again during equilibration, to maintain the temperature at 310 K, with the collision frequency now set at 2 ps^{-1}.

The final step for the conventional MD simulations was the production run. The parameters used were fundamentally identical to those for the second equilibration step, but using more MD steps – 250 million, resulting in 500 ns worth of data. Two individual cMD runs were performed for each force field, resulting in a total of 1 μs production data. Accelerated MD simulations were also carried out for the force field and explicit solvent combination, that re-sulted in data most comparable to experimental values. The parameters were kept identical to those for the production run, except for the added terms for the energy-boosting. The aMD simulations resulted in a total of 150 million trajectory steps (300 ns). The simulation proto-col used here, aligns with the setup implemented in previous computational assessments of the

metal-bound and metal-free Aβ system within our research group,[9] for better reproducibility and cross-validation of key findings.

3.2   Semiempirical Computational Methods

The semiempirical simulations in the present work were performed using the GFN2-xTB approach.[34] Before commencing the calculations, model 1 from 1ZE9 was geometry optimised using the GBSA water model. Simulations on the system were thus performed for 2 ns, in the NVT ensemble, maintaining the temperature at 310 K, using the Berendsen thermostat, while the SHAKE algorithm was used again, to restrain the bonds to their equilibrium position.[27, 35] The time step for propagation was set at 4 fs, with an increased mass of 4 amu for hydrogen atoms.[36] The increased time step is allowed here, due to the estimated decrease in the vibrational frequency of the bond between atoms coordinated to the hydrogens, now with a fictious higher mass. The simulations were also repeated using the analytical linearized Poisson-Boltzmann (ALPB) solvent model.[37]

The same parametrisation described above was used to examine the system using meta-dynamics (MTD), with RMSD as the collective variable.[38] The scaling factor (kpush) was set at 0.02 and the width of the gaussian potential (alp) at 1.2. Similarly to the simulations above, the dynamics were performed for 2 ns.

3.3   Experimental Data

To start off, an initial assessment of the system was performed on the 20 experimental structures (PDB ID: 1ZE9) of Aβ$_{16}$-Zn(II).[4] Figure 3.1 displays the structure of the peptide used here, with labelled residues involved in Zn(II)-coordination – three N from the imidazole ring of His6, His13 and His14, and one O from the carboxyl group of Glu11.
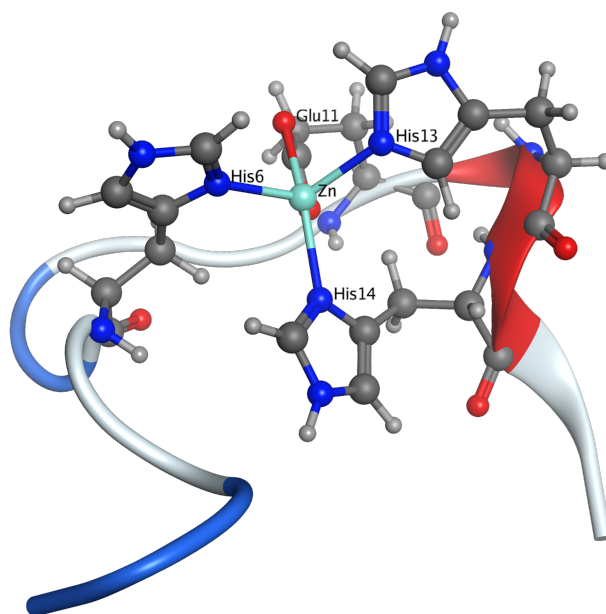
Figure 3.1: 3D illustration of Zn(II)-Aβ$_{16}$ (first model from 1ZE9).
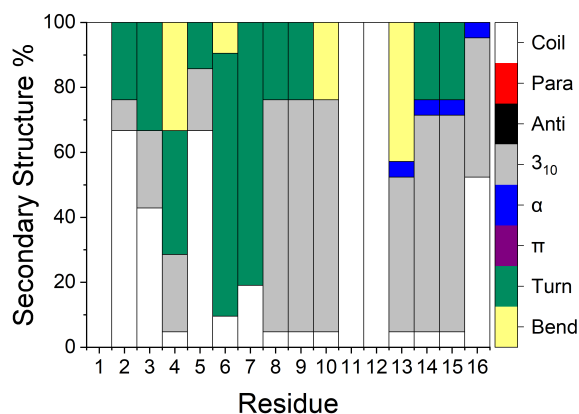
Figure 3.2: Secondary structure distribution per residue, from the ensemble of 20 experimental structures of Zn(II)-Aβ_{16}. The β-sheets are denoted with red (parallel) and black (antiparallel), and helices with grey (3_{10}), blue (α) and purple (π).
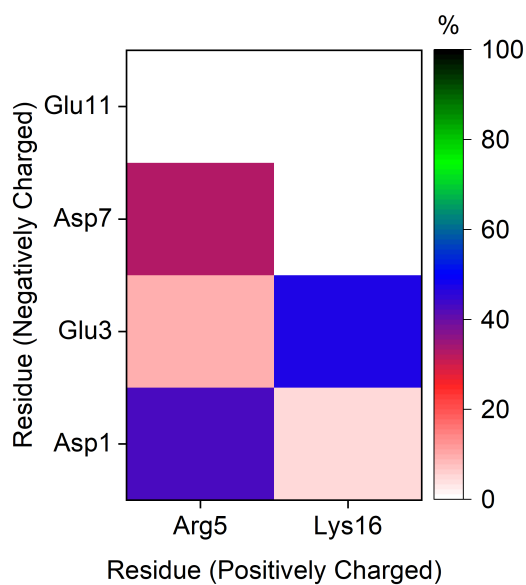
Figure 3.3: Salt bridges of the experimental data.
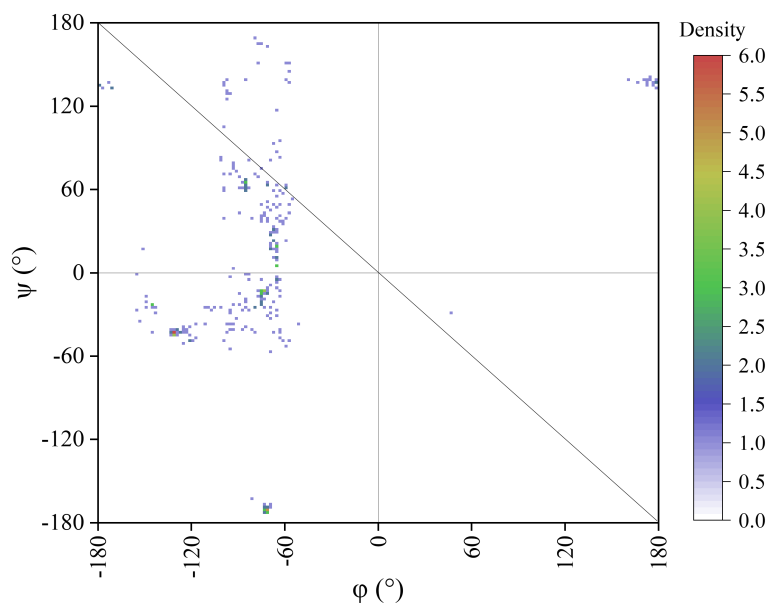
Figure 3.4: Ramachandran plot of Zn(II)-Aβ_{16} complex, from the 20 structures of the experimental data (1ZE9).
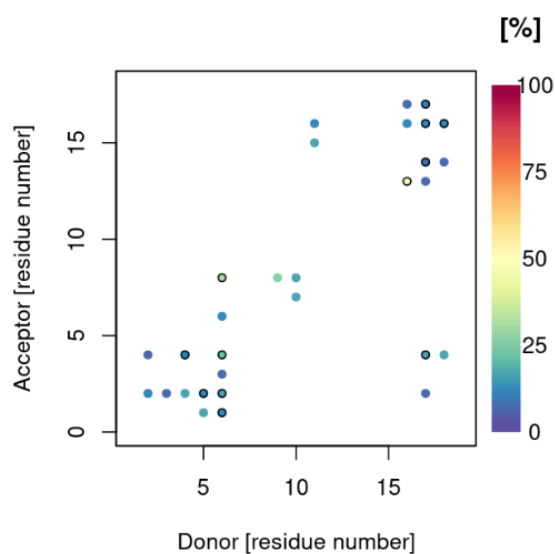
Figure 3.5: Hydrogen bond plot from the experimental peptide data.

Looking at the Ramachandran plot of the experimentally reported structures, Figure 3.4, suggests an increased population of α-helical character (bottom of the diagonal line; $\phi \approx -60°$, $\psi \approx -45°$), along with some presence of β-sheet structure (top left triangle; $\phi \approx -120°$, $\psi \approx 120°$). Although, considering the DSSP values, which average the strand population at 0% and α-helices at 33.33%, raises questions on the suitability of the dihedral angles to estimate the secondary structure. Nevertheless, the torsion angles suggest the geometry and flexibility of the backbone, which influences the peptide's ability to adopt, or be inclined towards, a particular folding conformation. The hydrogen bond plot, Figure 3.5, shows the most prevalent interaction between the oxygen from Val12 (acceptor) and a hydrogen atom bonded to nitrogen from Gln15 (donor), at 45%. These residues also interact through the same oxygen atom from

Val12 and another hydrogen atom bonded to the sidechain N in Gln15, at 25%; hence the black border around that point, to represent the double interaction.

3.4 Conventional Molecular Dynamics

This section presents the data from the conventional molecular dynamics (cMD) simulations on Zn(II)-Aβ_{16}, starting from the minimised structure of the first model in 1ZE9, using different force fields (ff99SB-ILDN,[12] ff14SB, ff14SBonlySC,[13] fb15,[14], ff03ws[16]).

3.4.1 Implicit Solvent

As discussed in the methodology of this chapter, for most of the force fields, the systems were solvated using the 'standard' generalised Born model (GBSA). For ff14SBonlySC, the GB-Neck2 (*igb8*) water model was used, as well as the GBSA one. The decision to use another implicit model to solvate the system parameterised with ff14SBonlySC, was made after surveying studies,[39–41] that used the force field and evaluated its efficiency when coupled with that particular water model, owing to its QM-based sidechain dihedral parameters. For ff03ws, the OBC modification to the GBSA implicit model was used, after considering the water model's use with the ff03 force field,[18–23, 42] which ff03ws is based on.

Table 3.1: Mean, standard deviation, max. and min. R$_g$ of implicit solvent cMD simulations of the Zn(II)-Aβ_{16} complex, using different force fields.

Zn(II)-Aβ_{16}	Avg. R$_g$ (Å)	SD (Å)	Max (Å)	Min (Å)
Experiment[4]	7.42	0.18	7.74	7.11
ff14SB/GBSA	7.58	0.24	9.05	6.78
ff99SB-ILDN/GBSA	7.66	0.42	9.93	6.64
ff14SBonlySC/GBSA	7.62	0.42	10.09	6.66
ff14SBonlySC/GBneck2	8.42	0.39	10.36	7.04
fb15/GBSA	7.54	0.32	9.95	6.68
ff03ws/OBC	7.77	0.36	10.20	6.88

The gyration tensor data were analysed using the descriptors defined in Section 2.4.8.4.2. The resulting values from these calculations are shown in Table 3.2. From these descriptors we get a general understanding of the global distribution of the atoms within the peptide chain. The experimental average from the 20 NMR structures for the globularity of the system, suggests an elongated, prolate shape. The globularity values from the implicit solvent simulations, broadly reproduce this shape, with certain exceptions, such as ff14SBonlySC/GBneck2, where the structure appears to be more linear than what the experiment proposed. The standard deviation values for globularity suggest that the peptide chain visits several conformations. For instance, in the case of ff14SB these values range from 0.78 to 0.09. The asphericity and acylindricity of the systems, emphasise how the diagonalized eigenvalues, Figure A.3, vary in each frame of the different parametrisation scenarios, resulting in values ranging from 45.69 to 2.58 for asphericity, and from 23.65 to 0.05 for acylindricity, using ff14SB. The anisotropy of the

systems, gives a better understanding of how all the points relate to each other and the over-all structure of the peptide. The closer this value is to 0, the more spherically symmetric the points are (in contrast to globularity, where as the system becomes more spherical, the value approaches 1). From this data, we can validate the assumption made above, of how ff14SBonlySC presents a more linear configuration than the rest of the force fields.

Table 3.2: Analysis of the diagonalized tensor's eigenvalues, for the cMD simulations in implicit solvent.

Zn(II)-Aβ_{16}	Globularity		Asphericity (Å)		Acylindricity (Å)		Anisotropy	
	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD
Experiment[4]	0.24	0.04	19.22	4.77	9.41	1.72	0.15	0.05
ff14SB/GBSA	0.27	0.12	16.48	3.30	8.14	4.61	0.15	0.07
ff99SB-ILDN/GBSA	0.27	0.10	21.55	7.96	7.82	5.33	0.16	0.07
ff14SBonlySC/GBSA	0.27	0.10	21.86	8.59	7.11	3.29	0.16	0.08
ff14SBonlySC/GBneck2	0.20	0.06	29.94	7.90	11.14	5.78	0.19	0.06
fb15/GBSA	0.26	0.09	17.85	6.79	10.95	4.38	0.14	0.06
ff03ws/OBC	0.28	0.08	22.58	7.86	6.45	2.99	0.15	0.07

Table 3.1 summarises the data collected from the analysis of the R_g. The overall conclusion that may be drawn from the above, is that the system remains relatively as compact as the peptide in the experimental data, with a mean R_g close to 7.5 Å for all force field configurations; except for the one using the GB-Neck2 implicit solvent model. The variance in the overall shape during the simulation is reflected in the standard deviation values, and the difference between the maximum and minimum. The effect of this change in the compactness in the systems, can be better evaluated by also considering the secondary characteristics in each case. The existence of β-sheets appears to constrain the system to a greater extend, compared to the effects of increasing population of α-helices. This is particularly evident when comparing the two runs of ff14SBonlySC, where the system associated with the higher percentage of β-character (i.e. solvated using GBSA, at 3.41%), gave a more constrained peptide with an R_g at 7.62 Å, compared to 8.42 Å, from the system solvated with GB-Neck2. The population of α-characteristics in the different scenarios, do not appear to influence the compactness of the systems to such a great extent, as the peptide with the highest α-character (i.e. using ff14SB) still yielded the more extended system. The effect of the solvent, in the compactness, will become more apparent further below, when looking at the peptide's dynamics in explicit solvent. The average secondary characteristics in implicit solvent, using different force fields, are given in Figure 3.6.

Figure 3.6: Secondary structure characteristics of the peptide, using (A) ff14SB, (B) ff99SB-ILDN, (C) ff14SBonlySC, (D) fb15 force fields with pairwise GBSA implicit model, (E) ff14SBonlySC with GB-Neck2 model and (F) ff03ws with OBC-GBSA. The β-sheets are denoted with red (parallel) and black (antiparallel), and helices with grey (3_{10}), blue (α) and purple (π).

The secondary structures, illustrated in Figure 3.6, further support the observations made from the radius of gyration data. The majority of the folding characteristics in all systems, appear to populate regions Asp1-Arg5 and Val12-Lys16. Hence, suggesting the regions where possible inter-peptide interactions are likely to form. The percentages of the different structures found in the residues, in each case, are summarised in Table 3.3.

Table 3.3: Mean secondary structure percentages for the complex implicit simulation using different force fields.

Zn(II)-Aβ_{16}	β character (%)	α character (%)	Other (%)
Experiment[4]	0.00	33.33	66.67
ff14SB/GBSA	0.50	27.42	72.08
ff99SB-ILDN/GBSA	1.58	16.09	82.33
ff14SBonlySC/GBSA	3.41	19.67	76.93
ff14SBonlySC/GBneck2	2.03	22.55	75.42
fb15/GBSA	1.46	19.51	79.03
ff03ws/OBC	0.85	16.70	82.45

From Table 3.3, it is apparent that the configuration most analogous to that observed in the experimental data, is the one simulated with ff14SB, as the system there presented the lowest β-sheet content (0.50%), as well as a 27.42% helical content, closely matching the 0% and 33.33% observed in the experimental structures. As discussed above, the data still speaks to the fact that there is no direct correlation between the percentage of α- and β-character present in the system and its overall compactness. A compilation of data, instead, must be considered when reviewing the phase space sampled in each case. Below, the data from the torsion angles is presented, Figure 3.7.
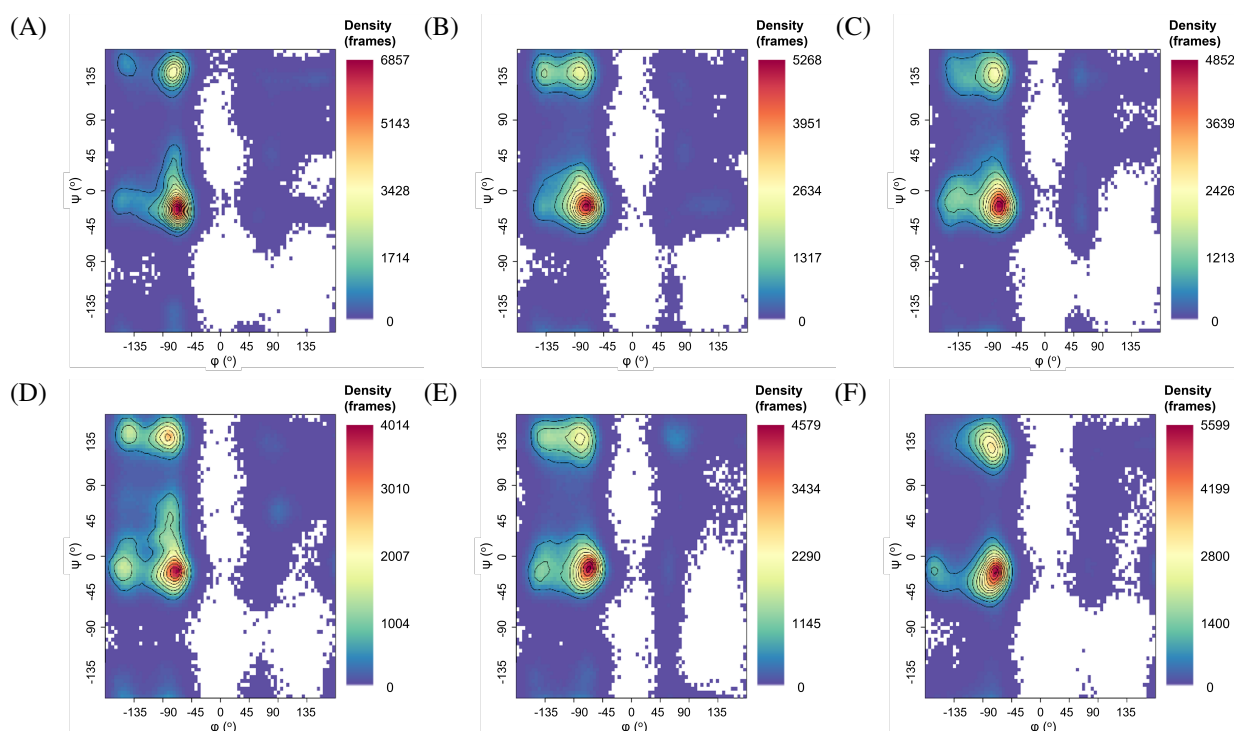
Figure 3.7: Ramachandran plots for Zn(II)-Aβ$_{16}$ cMD simulations, using (A) ff14SB, (B) ff99SB-ILDN, (C) ff14SBonlySC, (D) fb15 with pairwise GBSA model, (E) ff14SBonlySC with GB-Neck2 and (F) ff03ws with OBC-GBSA.

Compared to the experimental data, Figure 3.4, the force field that displays the closest similarity to that is ff03ws, with a subdued presence of left-handed α-helical character (top right quadrant) – a characteristic that appears to be almost non-existent in the experimental data. Nevertheless, given the great similarity in the regions where the dihedral angles are exhibited, amongst all six scenarios, it is difficult to draw any confident conclusions as to which parameters are most suitable for the simulation of the peptide, just from the Ramachandran plots.

The hydrogen bond plots, depicted in Figure 3.8, represent the most notable interactions between the residues in the peptide. From the experimental data, Figure 3.5, the most prominent interaction was between Val12 and Gln15, at 45%. In the case of the MD simulations, the longest-lasting interactions in each system, came from the following residues: ff14SB Glu11-His14 (77%); ff99SB-ILDN Glu11-His14 (54%); ff14SBonlySC(GBSA) Glu11-His14 (53%); fb15 Glu11-His14 (76%); ff14SBonlySC(GBneck2) Glu11-His14 (77%); ff03ws(OBC) Asp1-Val12 (48%). The Glu11-His14 hydrogen bonding in the majority of the force fields, is in all likelihood facilitated by the harmonic bonds between these ligands and Zn(II), decreasing and maintaining the distance between them. Looking at the hydrogen bonds formed between Val12

and Gln15 (i.e. the most prominent from the experimental data), in the case of ff14SB, Val12 now forms an interaction with Lys16 (7%), while Gln15 forms a hydrogen bond with Glu11 (22%), through the H bonded to the backbone N of the former, which also exhibits an intra-residue interaction with its sidechain O (7%), and another one with Arg5 (8%). For the rest of the systems, the following changes, involving Val12 and Gln15, are observed: ff99SB-ILDN Glu11-Gln15 (13%), Gln15-Arg5 (double interaction: 10% and 7%); ff14SBonlySC(GBSA) Glu11-Gln15 (13%); fb15 Glu11-Gln15 (11%), Gln15-Arg5 (double interaction: both at 8%); ff14SBonlySC(GBneck2) Glu11-Gln15 (16%), intra-residue Gln15 (7%), Val12-Lys16 (6%); ff03ws(OBC) Asp1-Val12 (double interaction: 48% and 25%), Glu11-Gln15 (double interaction: 18% and 10%). Looking at the experimental structures, the hydrogen bonds identified from both the experiment and MD simulations are given in Table 3.4, with the systems parameterised with ff14SB and ff99SB-ILDN exhibiting the greatest similarities to the reference.
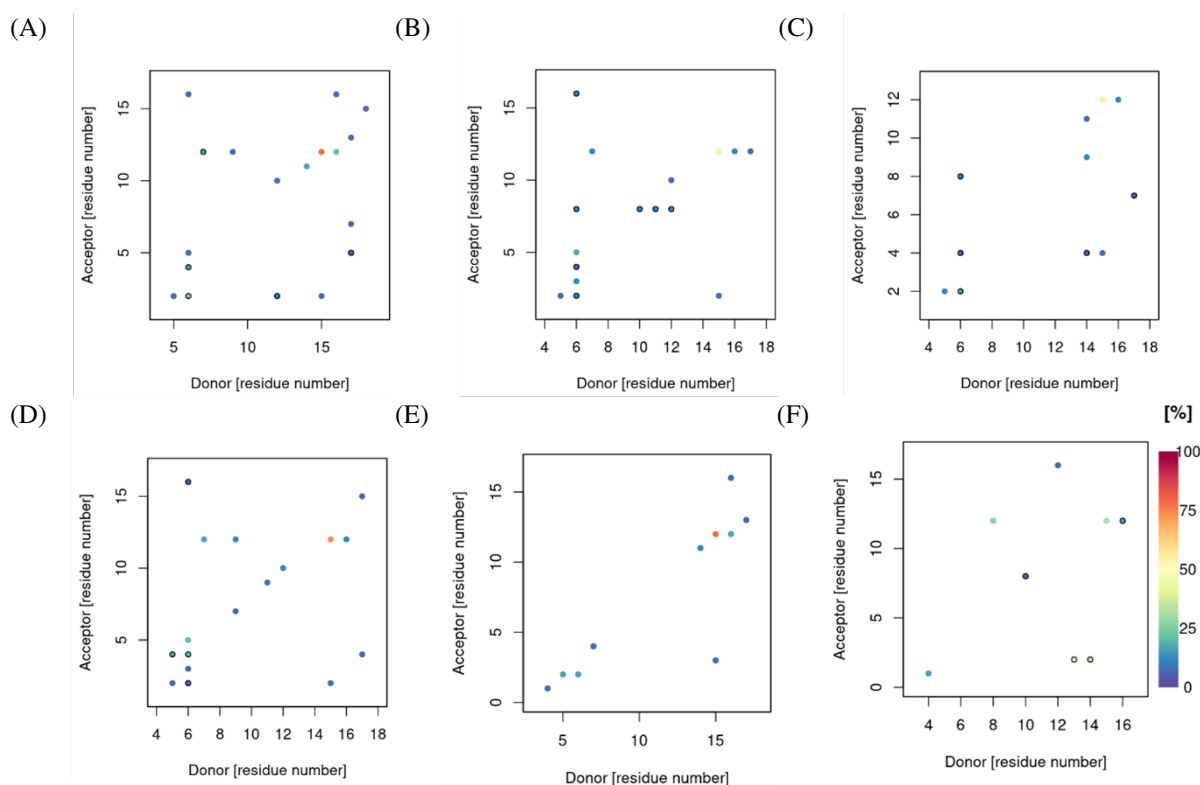
Figure 3.8: Hydrogen bond plots for Zn(II)-Aβ_{16} cMD simulations, using (A) ff14SB, (B) ff99SB-ILDN, (C) ff14SBonlySC, (D) fb15 with pairwise GBSA model, (E) ff14SBonlySC with GB-Neck2 and (F) ff03ws with OBC-GBSA.

Table 3.4: Hydrogen bond interactions (Acceptor – H-donor) shared between experimental data (1ZE9) and MD simulations using different force fields and implicit solvent models. When multiple interactions involving the same residues are present, only the most frequent one is shown. Values given in %.

System	Asp1-Phe4	Asp1-Arg5	Glu3-Arg5	Asp7-Arg5	Asp7-Gly9	Val12-Lys16
Experiment[4]	10	10	20	30	15	5
ff14SB/GBSA	9	26	15	-	-	7
ff99SB-ILDN/GBSA	9	12	5	10	11	-
ff14SBonlySC/GBSA	14	16	7	13	-	-
ff14SBonlySC/GBneck2	18	15	-	-	-	6
fb15/GBSA	5	6	22	-	-	-
ff03ws/OBC	-	-	-	-	6	-
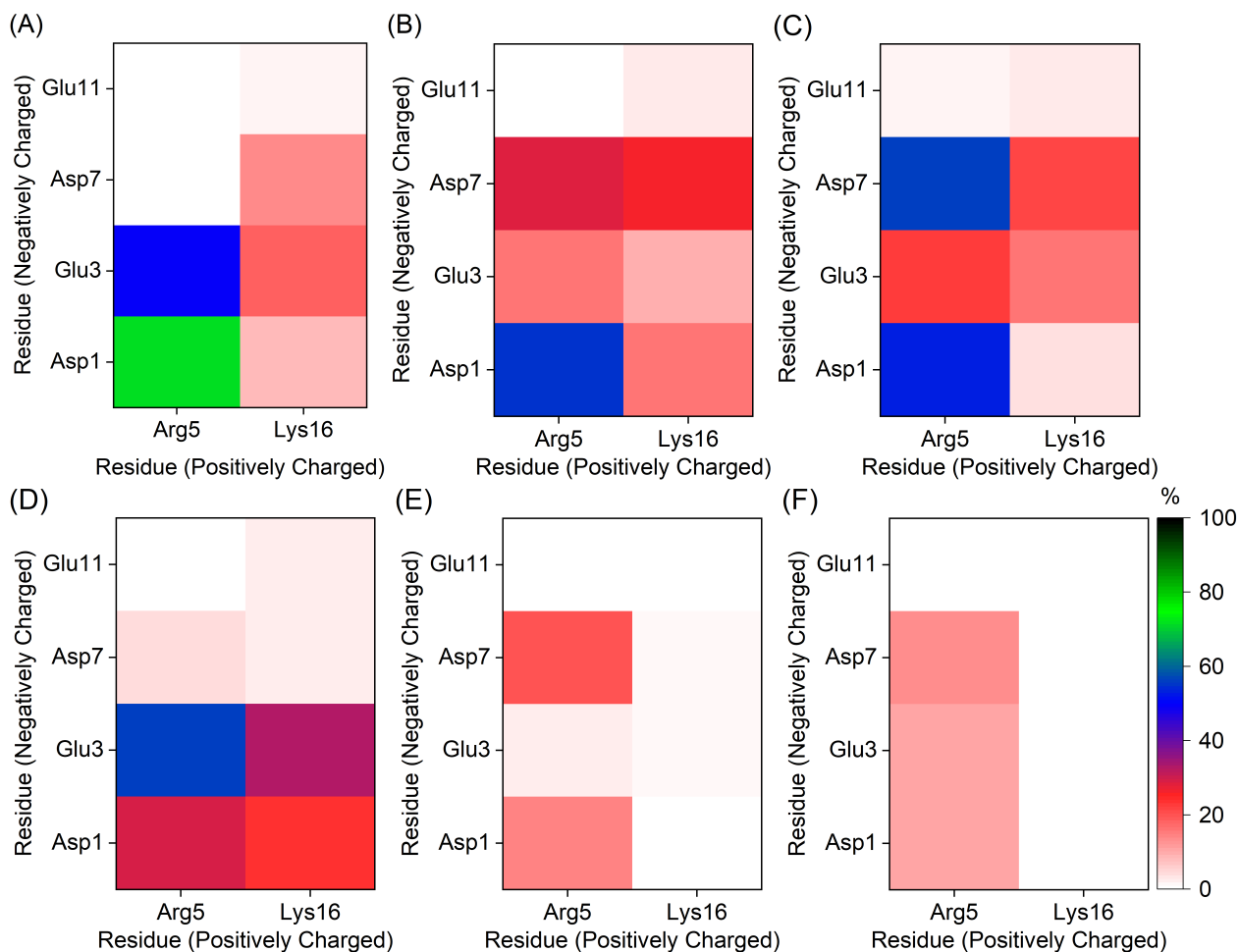
Figure 3.9: Salt bridge contact maps for Zn(II)-Aβ_{16} cMD simulations, using (A) ff14SB, (B) ff99SB-ILDN, (C) ff14SBonlySC, (D) fb15 with pairwise GBSA model, (E) ff14SBonlySC with GB-Neck2, and (F) ff03ws with OBC-GBSA.

Table 3.5: Salt bridge values for the force fields used in obtaining the cMD data in implicit solvent. Values given in %.

Zn(II)-Aβ_{16}	Asp1-Arg5	Asp7-Arg5	Glu3-Arg5	Glu3-Lys16
Experiment[4]	43	33	10	48
ff14SB/GBSA	72	1	49	18
ff99SB-ILDN/GBSA	55	29	16	9
ff14SBonlySC/GBSA	52	56	23	16
ff14SBonlySC/GBneck2	14	20	3	2
fb15/GBSA	29	4	56	33
ff03ws/OBC	11	14	10	-

The salt bridges present, after running the cMD simulations, are illustrated in Figure 3.9. For a clearer evaluation of the results, the data is presented in Table 3.5 – a numerical representation of the most significant interactions. The most substantial salt bridge for force fields ff14SB and ff99SB-ILDN, is the one between Asp1 and Arg5. This comes in agreement with the experimental observations, shown in Figure 3.3. The overestimation of the interactions in ff14SBonlySC with GBSA is a known issue, resulting from the use in polar systems, with that particular implicit solvent model.[43, 44] This is remedied with the use of a different implicit solvent model (GB-Neck2), where the results reproduce the experimental data much better. The underestimation of salt bridges, in some of the force fields, most likely arises from competing side-chain/backbone interactions. A complete numerical evaluation of the salt bridges present in each case is given in Appendix A. The force field that results in the closest agreement with the experimental data appears to be ff99SB-ILDN, as the population in each of the interactions match the values seen in experiment, except for the one between Glu3-Lys16. The force fields chosen to be examined further, in explicit solvent, were ff14SB, ff99SB-ILDN, fb15 and ff03ws; given the results from ff14SBonlySC presented a greater population of β-characteristics, a more linear distribution, and greater size of the chain (using GB-Neck2).

3.4.2 Explicit Solvent

When simulating biomolecular systems, the community generally relies on the explicit modelling of water molecules within the confines of a solvent box, the rationale being the charge incorporation and LJ parameterisation of the environment offering conditions more comparable to real-world experiments, than when incorporating a continuum model. The obvious limitation that has to be considered, however, is the computational cost such a system requires, and whether the supposed improved accuracy, outweighs the efficiency gain from using continuum models. For each of the force fields, two different explicit solvent models were used – TIP3P, which is the standard and most studied model, was the control used with all the force fields. Then, for a more complete assessment of the systems, TIP4P was used for ff14SB[45] and ff99SB-ILDN,[33, 46] TIP3P-FB for fb15,[14, 30, 31] and TIP4P/2005 for ff03ws.[31] This decision was based on literature surveys, comparing the water models in their ability to simulate a system when coupled with the aforementioned force fields. The cMD simulations in explicit solvent, averaged a total of 1 μs worth of data, by simulating each system with two individual runs of 250 million steps (i.e. 500 ns/run).

Looking at Table 3.6, we find the average R_g values for each of the systems, simulated with the different force fields and explicit solvent models. What is apparent before assessing the force fields amongst them, is the general loss of compactness compared to the simulations using implicit solvents. This may be predicated on an assumption, that will be further addressed below, where modelling the solvent explicitly, makes the complex more susceptible to intermolecular electrostatic forces, hence weakening the interactions encountered within the peptides. This holds for the majority of the force field – solvent combinations, except in the case of ff14SB/TIP3P. Even though the standard deviation and difference between the maximum and minimum values, suggest a great variance in the compactness of the amino acid chain, the average R_g (7.73 Å) comes in close agreement to the system simulated with the same force field in implicit water (GBSA). Furthermore, the fact that the explicit solvent cMD simulations were run for 500 ns (x2), following the equilibration steps at the beginning, compared to the 250 ns runs for the implicit solvent, provided longer timescales to sample the phase space, hence more time may have been spent exploring suboptimal local minima. Lastly, a known problem with dynamics in implicit solvent is the overestimation of α-helices, which may contribute to the overall system's compactness.[44, 47] Hence, even though the system may appear more condensed in those kinds of simulations, this does not necessarily corroborate the superiority of that method.

Table 3.6: Mean, standard deviation, max. and min. R_g of explicit solvent cMD simulations of the Zn(II)-Aβ_{16} complex, using different force fields and water models.

Zn(II)-Aβ_{16}	Avg. R_g (Å)	SD (Å)	Max (Å)	Min (Å)
Experiment[4]	7.42	0.18	7.74	7.11
ff14SB - TIP3P	7.73	0.41	10.14	6.85
ff14SB - TIP4P	7.95	0.38	9.58	7.10
fb15 - TIP3P	8.22	0.63	10.47	6.74
fb15 - TIP3P-FB	8.31	0.68	10.55	6.95
ff99SB-ILDN - TIP3P	7.94	0.68	10.40	6.63
ff99SB-ILDN - TIP4P	7.88	0.54	10.42	6.83
ff03ws - TIP3P	8.20	0.70	10.69	6.86
ff03ws - TIP4P/2005	7.86	0.57	10.68	6.89

Upon further examination of the gyration tensor, we find that the overall spherical symmetry of the peptides in explicit solvent, resembles more closely the experimental structure, compared to the implicit simulations, with fewer variations in the average data amongst the majority of force field/water model combinations. The combination with the greatest deviation from experimental values appears to be ff03ws/TIP4P/2005, where the system exhibits a more spherical character than the rest. Nevertheless, no confident conclusion may be drawn from these values, on which method is more accurate, as the results illustrated here do not consistently outperform those seen in Table 3.2.

Table 3.7: Analysis of the diagonalized tensor's eigenvalues, for the cMD simulations in explicit solvent.

Zn(II)-Aβ$_{16}$	Globularity		Asphericity (Å)		Acylindricity (Å)		Anisotropy	
	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD
Experiment[4]	0.24	0.04	19.22	4.77	9.41	1.72	0.15	0.05
ff14SB - TIP3P	0.27	0.08	19.79	7.86	9.83	4.11	0.14	0.06
ff14SB - TIP4P	0.23	0.07	26.34	9.54	7.89	4.62	0.20	0.09
fb15 - TIP3P	0.29	0.13	24.43	12.05	6.59	3.83	0.16	0.10
fb15 - TIP3P-FB	0.28	0.10	20.24	9.04	10.30	4.68	0.13	0.07
ff99SB-ILDN - TIP3P	0.26	0.11	28.95	12.74	6.38	3.68	0.19	0.10
ff99SB-ILDN - TIP4P	0.25	0.11	30.08	14.59	7.40	4.16	0.20	0.11
ff03ws - TIP3P	0.27	0.13	25.37	13.08	9.59	4.76	0.16	0.10
ff03ws - TIP4P/2005	0.38	0.14	18.45	11.14	6.27	3.60	0.10	0.08

Table 3.8: Mean secondary structure percentages for the cMD simulations of the complex in explicit solvent, using different force fields.

Zn(II)-Aβ$_{16}$	β character (%)	α character (%)	Other (%)
Experiment[4]	0.00	33.33	66.67
ff14SB - TIP3P	0.94	20.50	78.56
ff14SB - TIP4P	0.33	23.38	76.28
fb15 - TIP3P	4.20	8.24	87.56
fb15 - TIP3P-FB	2.31	15.92	81.77
ff99SB-ILDN - TIP3P	3.72	8.36	87.92
ff99SB-ILDN - TIP4P	3.89	6.06	90.05
ff03ws - TIP3P	1.65	12.23	86.13
ff03ws - TIP4P/2005	8.06	4.31	87.63

The secondary structure characteristics of each of the systems, using the different force fields and water models, are illustrated in Figure 3.10 with a numerical representation in Table 3.8. From the data presented in these figures, it appears that cMD simulations with ff14SB in the TIP4P water model, yield the secondary structure closest to the experimental data. The same force field with the TIP3P model, nevertheless, still yielded a promising average R_g. The β-characteristics of the systems support, once more, the idea of how they relate to the compactness of the system. This is evident, when comparing the same force fields, with the different water models; as in every case, the system with the highest percentage of β-sheets, resulted in the more compact peptide chains. Furthermore, the data reinforces the theory of how implicit solvent simulations result in an overestimation of α-helices. Nonetheless, an evaluation of

the interactions should yield a more comprehensive assessment of the effect of the simulation parameters, on the structure of the amino acid chain.
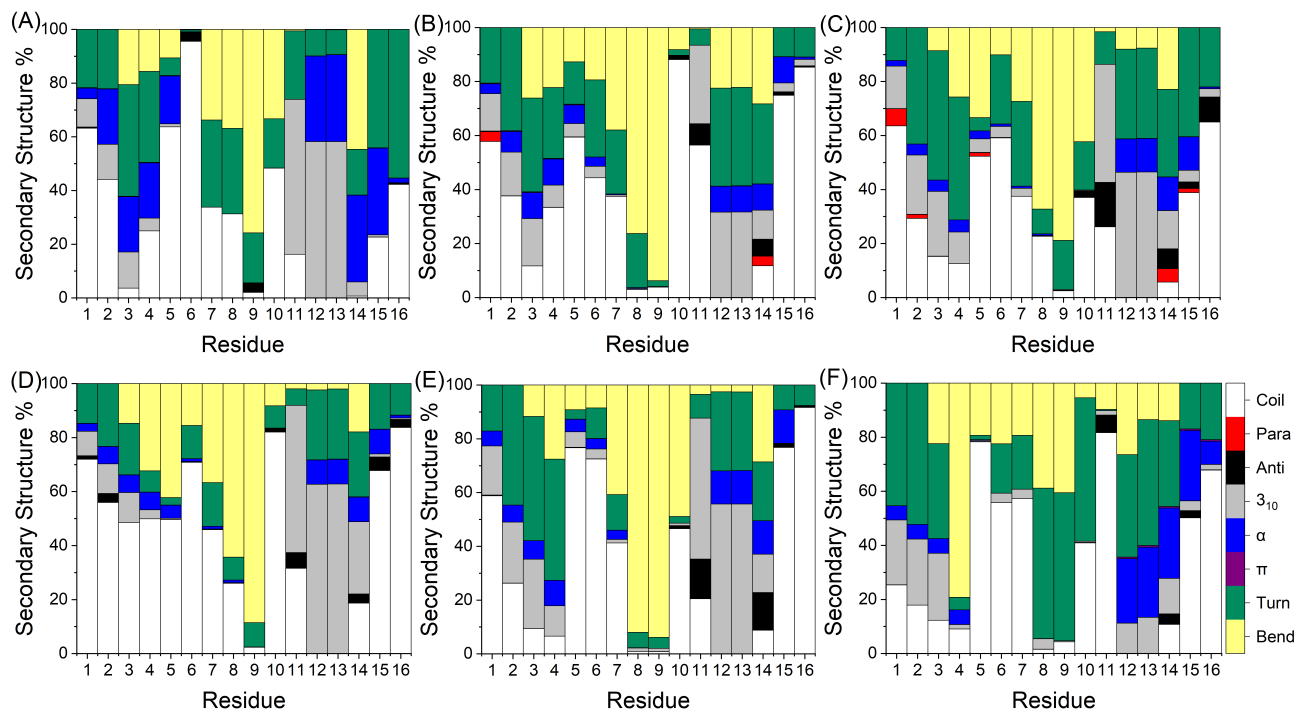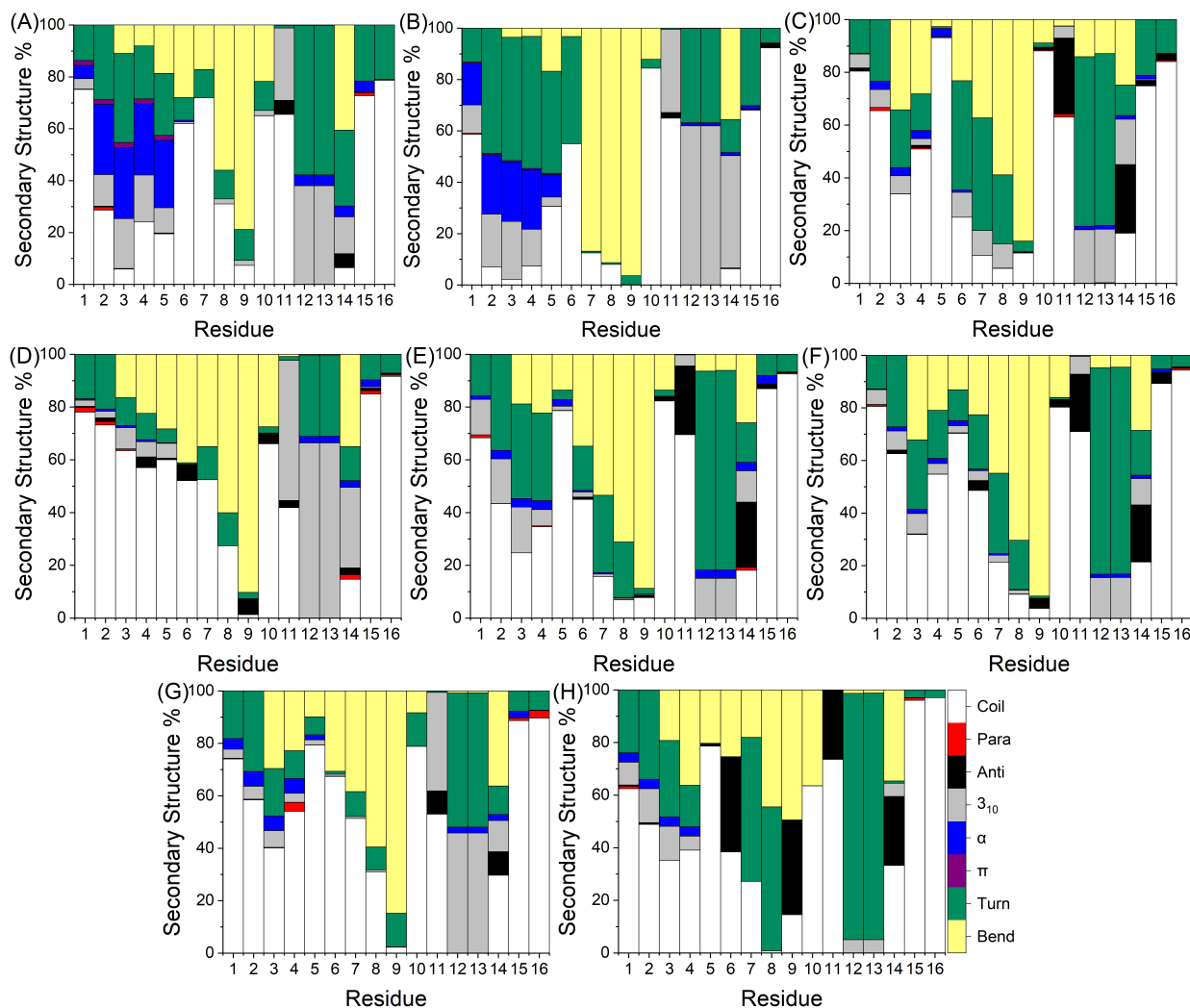

Figure 3.10: Secondary structure characteristics of the peptide, using (A) ff14SB, (C) fb15, (E) ff99SB-ILDN, (G) ff03ws force fields with TIP3P explicit water model, (B) ff14SB and (F) ff99SB-ILDN with TIP4P, (D) fb15 with TIP3P-FB, and (H) ff03ws with TIP4P/2005. The β-sheets are denoted with red (parallel) and black (antiparallel), and helices with grey (3_{10}), blue (α) and purple (π).

The Ramachandran plots from the explicit solvent simulations depict the conformations already hinted at by the secondary structure plots. The greatest α-helical character is present in Figure 3.11(B), which corresponds to the system parameterised using ff14SB and solvated with TIP4P. Overall, the plots that appear to resemble the experimental values, come from ff03ws (Figure 3.11(G/H)), although considering the density of the frames, which are much higher than the rest of the systems, it is difficult to make confident comparisons.
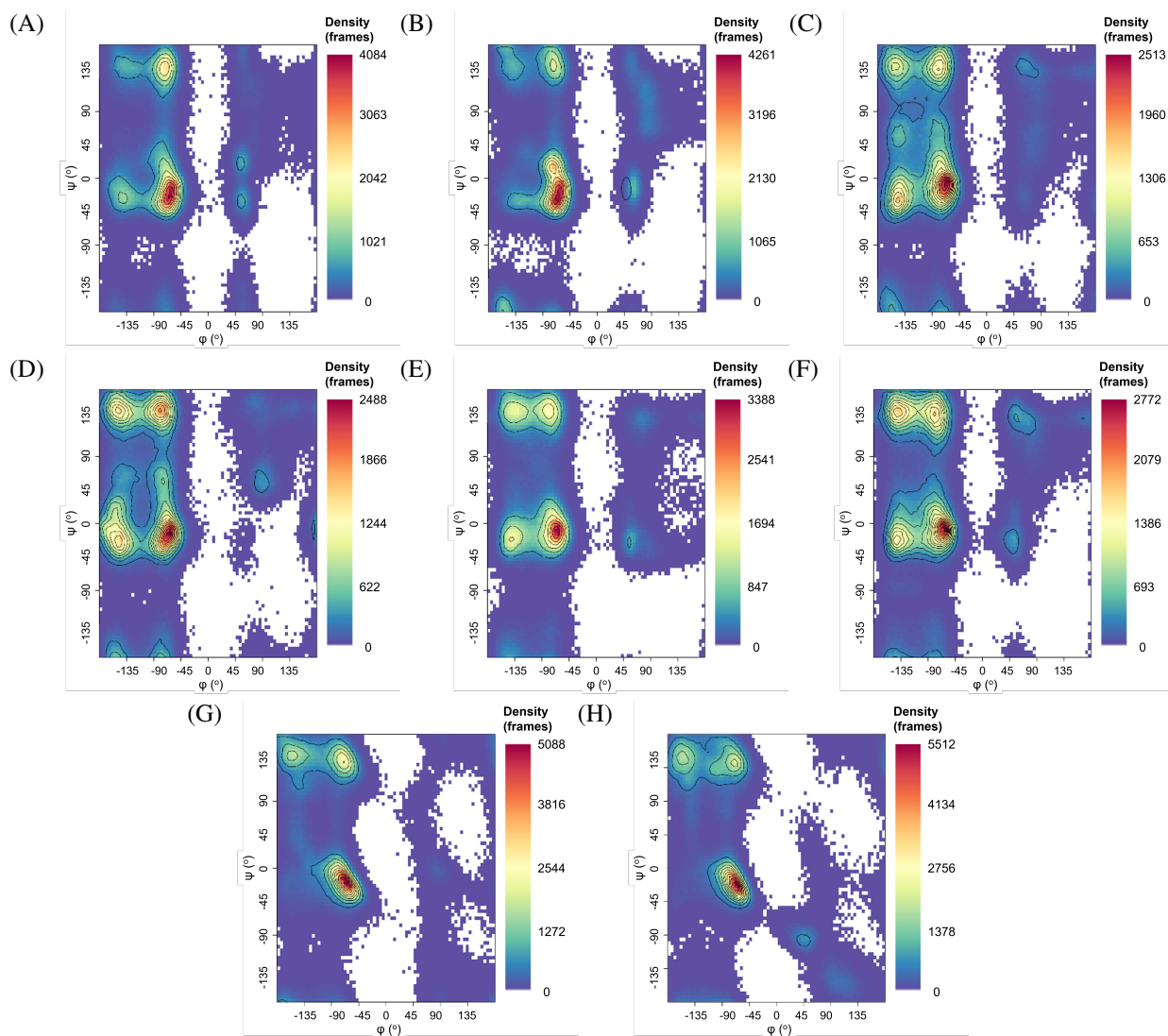
Figure 3.11: Ramachandran plots for Zn(II)-Aβ$_{16}$ cMD simulations, using (A) ff14SB, (C) fb15, (E) ff99SB-ILDN, (G) ff03ws force fields with TIP3P explicit water model, (B) ff14SB and (F) ff99SB-ILDN with TIP4P, (D) fb15 with TIP3P-FB, and (H) ff03ws with TIP4P/2005.
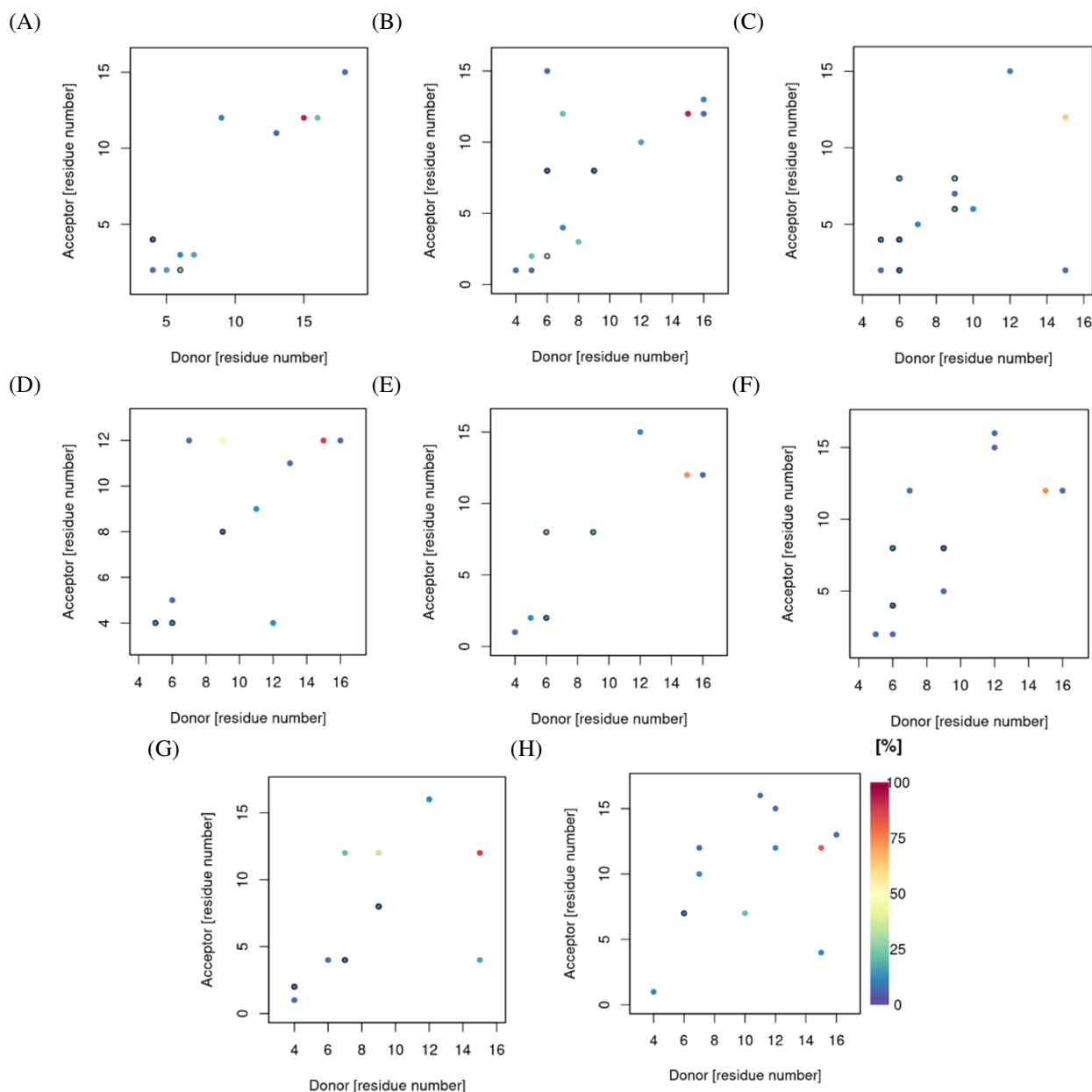
Figure 3.12: Hydrogen bond plots for Zn(II)-Aβ$_{16}$ cMD simulations, using (A) ff14SB, (C) fb15, (E) ff99SB-ILDN, (G) ff03ws force fields with TIP3P explicit water model, (B) ff14SB and (F) ff99SB-ILDN with TIP4P, (D) fb15 with TIP3P-FB, and (H) ff03ws with TIP4P/2005.

The most significant hydrogen bond interactions, observed in Figure 3.12, come from the same two residues, seen in the majority of the implicit simulations – between Glu11 and His14 (i.e. two of the Zn(II)-binding residues), now also found using the ff03ws force field. The different percentages observed in each case are: ff14SB/TIP3P (91%), ff14SB/TIP4P (93%), fb15/TIP3P (65%), fb15/TIP3P-FB (90%), ff99SB-ILDN/TIP3P (76%), ff99SB-ILDN/TIP4P (74%), ff03ws/TIP3P (86%), ff03ws/TIP4P/2005 (83%). These appear to validate the compactness suggested from the R$_g$ values; if the secondary structure characteristics are also taken into consideration, just for the systems that were simulated using ff14SB. In other words, even though the strongest interactions are present in ff14SB/TIP4P, the additional β-character in ff14SB/TIP3P results in a more compact peptide, compared to the former. Furthermore, the rest of the force fields, appear to follow the same trend, with the systems exhibiting a greater

population of β-character, resulting in a greater compactness overall. The data also supports the observations made in the implicit solvent simulations, where the α-helical character did not appear to significantly influence the R_g. Table 3.9, displays the hydrogen bonds shared between the experimental conformations and the ones from the explicit simulations. From the data here, the systems exhibiting interactions found in the experiment are ff14SB/TIP4P, fb15/TIP3P, and ff99SB-ILDN/TIP4P.

Table 3.9: Hydrogen bond interactions (Acceptor – H-donor) shared between experimental data (1ZE9) and MD simulations using different force fields and explicit solvent models. When multiple interactions involving the same residues are present, only the most frequent one is shown. Values given in %, with standard deviation from block averaging every 5,000 frames, given in brackets.

System	Asp1-Phe4	Asp1-Arg5	Glu3-Arg5	His6-Gly9	Asp7-Arg5	Asp7-Ser8	Val12-Gln15
Experiment[4]	10	15	20	15	30	25	45
ff14SB - TIP3P	14 (8)	24 (23)	-	-	-	-	-
ff14SB - TIP4P	17 (15)	20 (18)	-	-	-	5 (6)	-
fb15 - TIP3P	7 (7)	7 (9)	7 (16)	-	19 (12)	23 (13)	-
fb15 - TIP3P-FB	-	-	14 (14)	-	-	8 (15)	-
ff99SB-ILDN - TIP3P	14 (8)	14 (8)	-	-	22 (17)	15 (10)	-
ff99SB-ILDN - TIP4P	7 (6)	5 (6)	6 (5)	-	12 (10)	7 (6)	-
ff03ws - TIP3P	-	-	5 (5)	-	-	6 (5)	-
ff03ws - TIP4P/2005	-	-	-	21 (18)	-	-	6 (7)
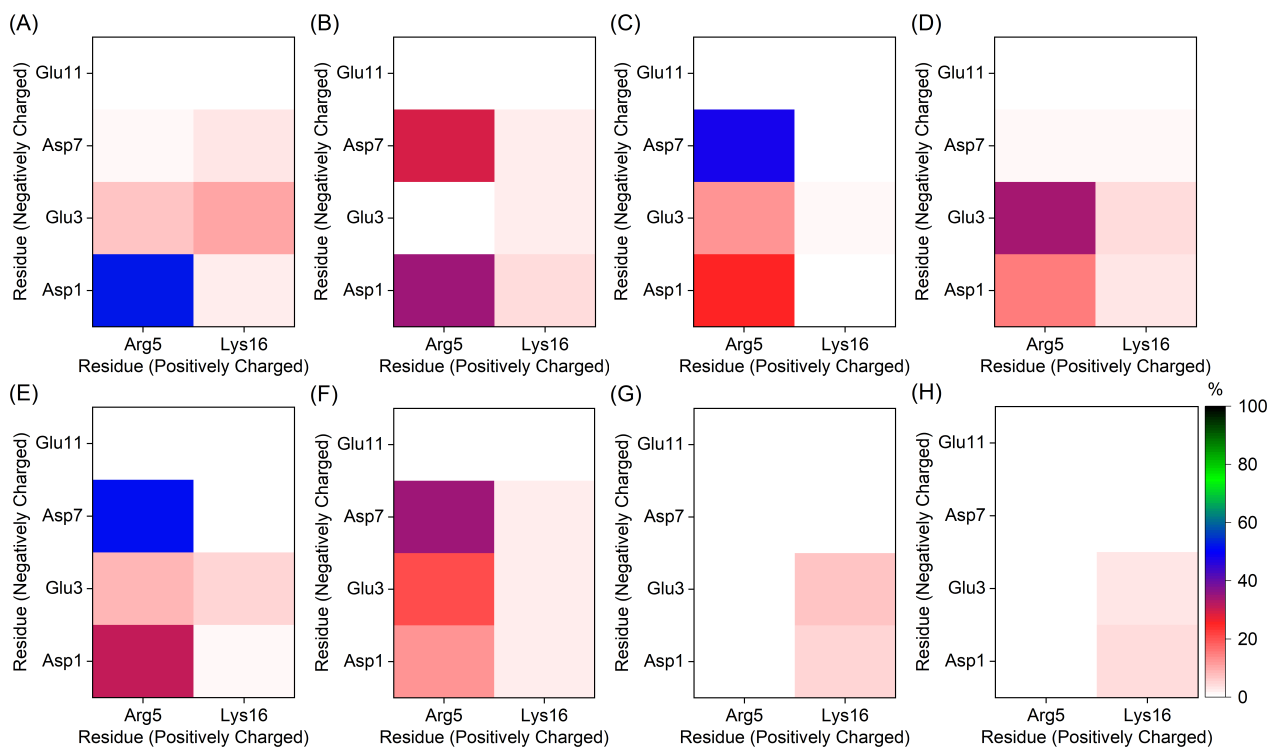
Figure 3.13: Salt bridge contact maps for Zn(II)-Aβ$_{16}$ cMD simulations, using (A) ff14SB, (C) fb15, (E) ff99SB-ILDN, (G) ff03ws force fields with TIP3P explicit water model, (B) ff14SB and (F) ff99SB-ILDN with TIP4P, (D) fb15 with TIP3P-FB, and (H) ff03ws with TIP4P/2005.

Table 3.10: Salt bridge values for the force fields used in obtaining the cMD data in explicit solvent. Values given in %.

Zn(II)-Aβ$_{16}$	Asp1-Arg5	Asp7-Arg5	Glu3-Arg5	Glu3-Lys16
Experiment[4]	43	33	10	48
ff14SB - TIP3P	52	1	7	10
ff14SB - TIP4P	35	29	-	3
fb15 - TIP3P	25	47	13	2
fb15 - TIP3P-FB	15	1	33	4
ff99SB-ILDN - TIP3P	31	52	9	5
ff99SB-ILDN - TIP4P	13	35	21	2
ff03ws - TIP3P	-	-	-	7
ff03ws - TIP4P/2005	-	-	-	3

The configurations that overall exhibit salt bridges found in the experimental data, come from the systems simulated with the TIP3P solvent model, despite the greatest similarities in the population of Asp7-Arg5 coming from the TIP4P solvents. The salt bridges formed between the residues give a good understanding of the conformational stability within the peptide. A conclusion that may be drawn on the differences of the systems, when simulated with implicit and explicit solvents, is the overestimation of salt bridges in implicit solvent and the difference

in secondary structure distributions. This is in part due to exaggeration of α-helices, when using the GBSA implicit solvent models, compared to explicit solvent simulations.[44, 47]

The analysis of the results from the cMD simulations using explicit solvent, point towards the ff14SB/TIP3P combination, to be the one resulting in the most representative average conformations, when considering the intramolecular interactions, R_g and secondary characteristics. Hence, it was decided to further the assessment through an enhanced sampling technique, with the introduction of a bias in the system, permitting access to a greater number of conformations by increasing the likelihood of escaping local minima.

3.5   Accelerated Molecular Dynamics

Following the cMD simulations using the different parameters, both in explicit and implicit solvents, the total and dihedral energies of the simulated systems were used to boost the system. Hence, the accelerated molecular dynamics (aMD) data, obtained from the introduction of a bias in the systems, is displayed and analysed in this section. The force field used in the parametrisation of the peptide is ff14SB, with the chain solvated both in implicit (GBSA) and explicit (TIP3P) water models. The aMD runs, simulated a system at 150 million steps using a 2 fs time step, resulting in 300 ns worth of data.

The average R_g values displayed in Table 3.11 illustrate how, overall, the cMD simulations resulted in a more compact system compared to aMD. This was something to be expected, considering the former approach may result in the system settling within a conformational basin that is harder to escape without an energy bias to propel it. The aMD simulations, especially using the explicit solvent resulted in the greatest variance in data. The more restricted nature of the peptide using cMD is further evident from the secondary characteristics, where it resulted in more populated β-character compared to aMD, both in implicit and explicit solvents, hence implying a more constrained structure in those simulations.

Table 3.11: R_g of the Zn(II)-Aβ$_{16}$ complex, using ff14SB, in different solvent models and simulation protocols.

Zn(II)-Aβ$_{16}$	Avg. R_g (Å)	SD (Å)	Max (Å)	Min (Å)
Experiment[4]	7.42	0.18	7.74	7.11
ff14SB - GBSA - cMD (600 ns)	7.70	0.29	9.55	6.86
ff14SB - GBSA - aMD (300 ns)	7.73	0.30	9.40	6.79
ff14SB - TIP3P - cMD (1 μs)	7.73	0.41	10.14	6.85
ff14SB - TIP3P - aMD (300 ns)	8.29	0.61	10.48	6.71

The globularity and anisotropy of the systems, suggest that the chains in explicit water resemble the sphericity proposed by the experiment. Looking at the explicit solvent results, a conclusion that may be drawn about the difference between the two MD methods is that the z tensor-diagonalized eigenvalue, Table A.7, is the reason for the poorer performance of the aMD simulations. This is evident from the asphericity and anisotropy values which, when compared to the cMD data fall short of the experimental results, while the rest (globularity/acylindricity)

outperform them. These two descriptors (asphericity/anisotropy) rely heavily on the λ_z, which is why they are the ones most impacted by this.

Table 3.12: Analysis of the diagonalized tensor's eigenvalues, for the aMD simulations using ff14SB.

Zn(II)-Aβ₁₆	Globularity		Asphericity (Å)		Acylindricity (Å)		Anisotropy	
	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD
Experiment[4]	0.24	0.04	19.22	4.77	9.41	1.72	0.15	0.05
ff14SB - GBSA - cMD	0.20	0.06	27.10	5.74	6.89	2.23	0.22	0.06
ff14SB - GBSA - aMD	0.19	0.05	26.98	6.16	7.68	2.64	0.22	0.06
ff14SB - TIP3P - cMD	0.27	0.08	19.79	7.86	9.83	4.11	0.14	0.06
ff14SB - TIP3P - aMD	0.23	0.10	28.75	12.32	9.43	9.43	0.20	0.10

Table 3.13: Mean secondary structure percentages for the cMD and aMD simulations of the complex in implicit and explicit solvent, using ff14SB.

Zn(II)-Aβ₁₆	β character (%)	α character (%)	Other (%)
Experiment[4]	0.00	33.33	66.67
ff14SB - GBSA - cMD (600 ns)	0.65	33.95	65.41
ff14SB - GBSA - aMD (300 ns)	0.60	33.95	65.44
ff14SB - TIP3P - cMD (1 μs)	0.94	20.50	78.56
ff14SB - TIP3P - aMD (300 ns)	0.88	15.67	83.45
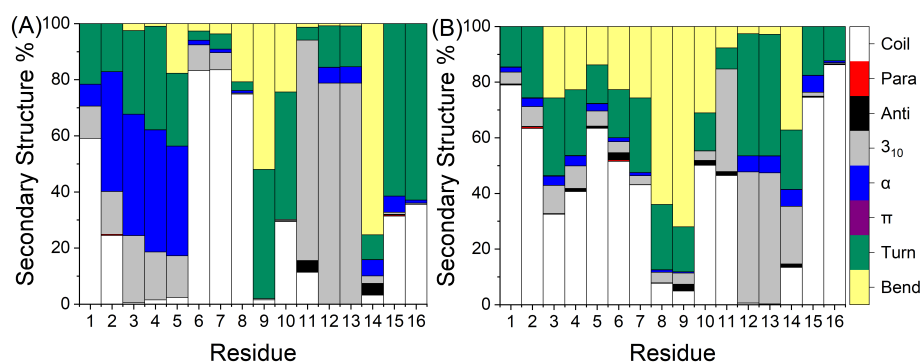
Figure 3.14: Secondary structure characteristics of the peptide for the aMD simulations, using ff14SB in (A) implicit and (B) explicit solvent. The β-sheets are denoted with red (parallel) and black (antiparallel), and helices with grey (3₁₀), blue (π) and purple (α).

The Ramachandran plots for the aMD simulations using implicit and explicit solvent are shown in Figure 3.15. The data extracted from these plots illustrate how the implicit solvent simulations are more in line with the experimental data, possibly due to the increased friction when

simulating the peptide in explicit solvent. Furthering the observations from the aMD simulations, the cMD results, Figure 3.7(A) and Figure 3.11(A) for implicit and explicit solvents, again propose a greater agreement between the implicit simulations and experiment. The dihedrals, in the case of the explicit solvent, cover a wider spread than the ones observed both in implicit water and experiment. Granted, that overall the densities from the aMD simulations are lower than for cMD, there is no particular difference in the areas where these are observed.
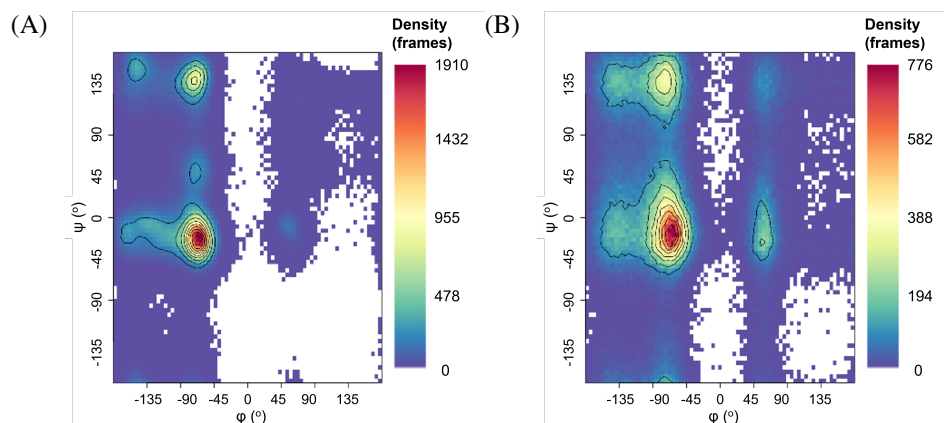
Figure 3.15: Ramachandran plots for Zn(II)-Aβ$_{16}$ aMD simulations, using ff14SB in (A) implicit and (B) explicit solvent.

The hydrogen bond plots, again, maintain the residues where the most prominent interactions are present in the cMD data – Glu11 and His14, both at *ca.* 70% occupancy. The hydrogen bonds for the simulations in explicit solvent, as expected, are fewer than what is observed from implicit simulations and experimental results. This is due to the increased interactions with the solvent, reducing the tendency of the peptide to develop these in an intramolecular fashion. The only shared interactions between the explicit/aMD simulations and the experiment, are between Glu3-Arg5 at 6% (1ZE9: 20%), while in the implicit/aMD simulations the interactions maintained, include: Asp1-Phe4 (1ZE9: 10%; ff14SB: 18%), Asp1-Arg5 (1ZE9: 10%; ff14SB: 22%), Asp7-Arg5 (1ZE9: 30%; ff14SB: 10%), Val12-Gln15 (1ZE9: 45%; ff14SB: 5%). In view of the implicit solvent results here, the aMD simulations are in better agreement with experimental results than cMD simulations. Although both aMD and cMD maintain four interactions seen in the experiment, the former accurately capture the ones with the highest occupancy, between Val12-Gln15 and Asp7-Arg5, as opposed to Glu3-Arg5 and Val12-Lys16, in the cMD.

(A)   (B)
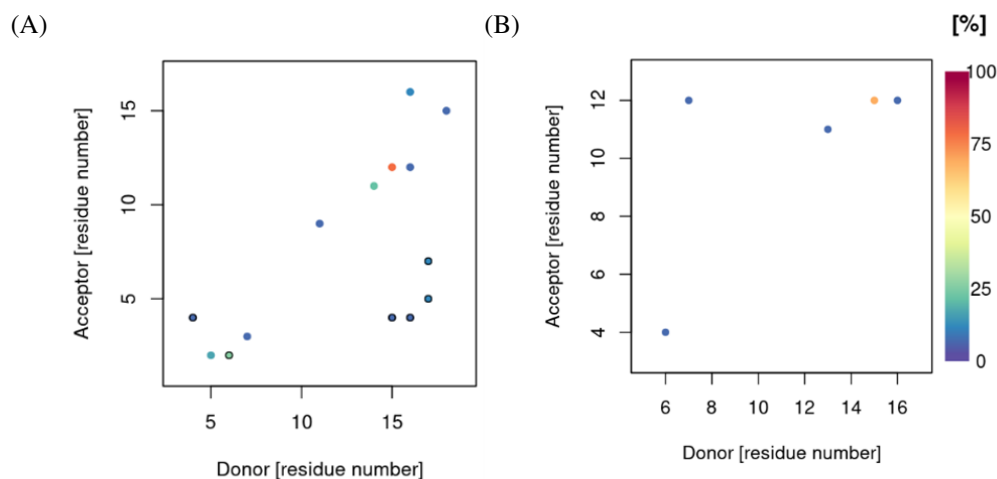
Figure 3.16: Hydrogen bond plots for Zn(II)-Aβ_{16} aMD simulations, using ff14SB in (A) implicit and (B) explicit solvent.

From the salt bridges formed within the peptide, the observation that remains true, from the previous assessment of the data, is their overestimation when using an implicit solvent. The combination that yields the most practical salt bridge configuration, appears to be when simulating the peptide using ff14SB in explicit solvent with aMD. This holds for most of the interactions, except for the Glu3-Lys16, which is greatly underestimated. Nevertheless, not one approach seems to greatly surpass the other in the whole.

(A)   (B)
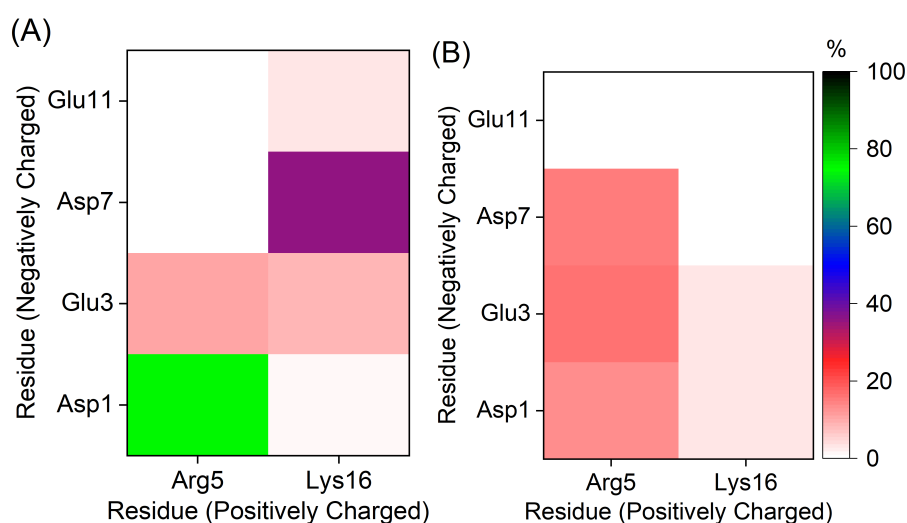
Figure 3.17: Salt bridge contact maps for Zn(II)-Aβ_{16} aMD simulations, using ff14SB in (A) implicit and (B) explicit solvent

Table 3.14: Salt bridge values for the simulation data, obtained using ff14SB in implicit and explicit solvent. Values given in %.

Zn(II)-Aβ$_{16}$	Asp1-Arg5	Asp7-Arg5	Glu3-Arg5	Glu3-Lys16
Experiment[4]	43	33	10	48
ff14SB - GBSA - cMD (600 ns)	74	12	7	3
ff14SB - GBSA - aMD (300 ns)	76	-	11	9
ff14SB - TIP3P - cMD (1 μs)	52	1	7	10
ff14SB - TIP3P - aMD (300 ns)	13	15	16	3

The accelerated MD allow the construction of the system's free energy landscape. Reweighting the aMD trajectories, permit the procurement of the corrected ensemble average and the original free energy profile of the system.[48] The reweighted trajectories allow the representation of certain properties in one and two-dimensional plots, against the potential energy surface. This data provides a representation of the stable, metastable and unstable states in a system.

The 1-dimensional free energy profiles of the hydrogen bonds found in the systems, parameterised with ff14SB and using implicit and explicit solvents, are shown in Figure 3.18. Here, we can see how the system in the implicit model exhibits an energy minimum at five hydrogen bonds, while the peptide in explicit solvent has a preference for three to four. The number of hydrogen bonds presented here is consistent with those seen in the hydrogen bond plots, Figure 3.16, where most of the hydrogen bonds are present in the implicit solvent simulations, owing to the absence of explicit interactions with solvent molecules.
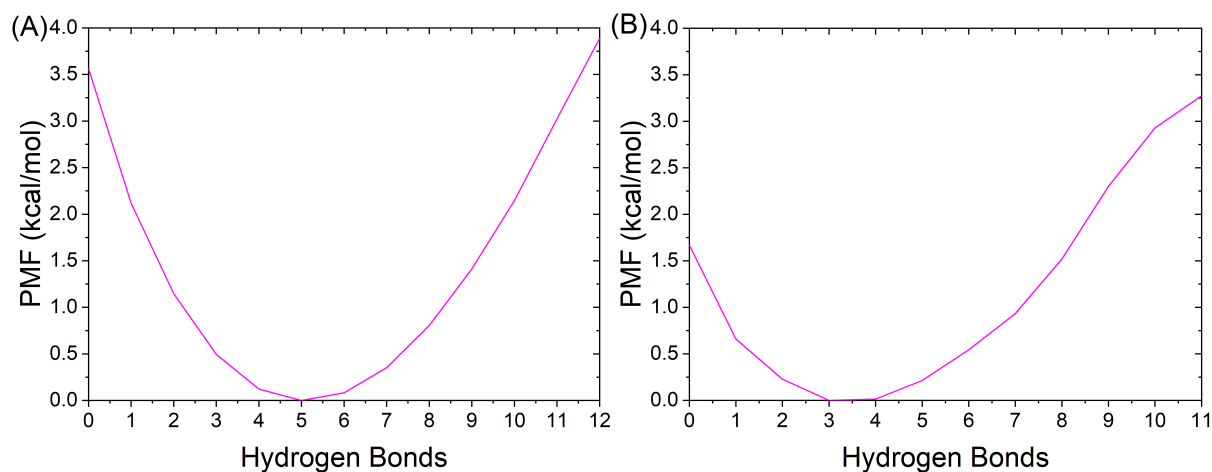
Figure 3.18: 1D free energy plots of hydrogen bonds, formed during aMD simulations using (A) implicit and (B) explicit water models.
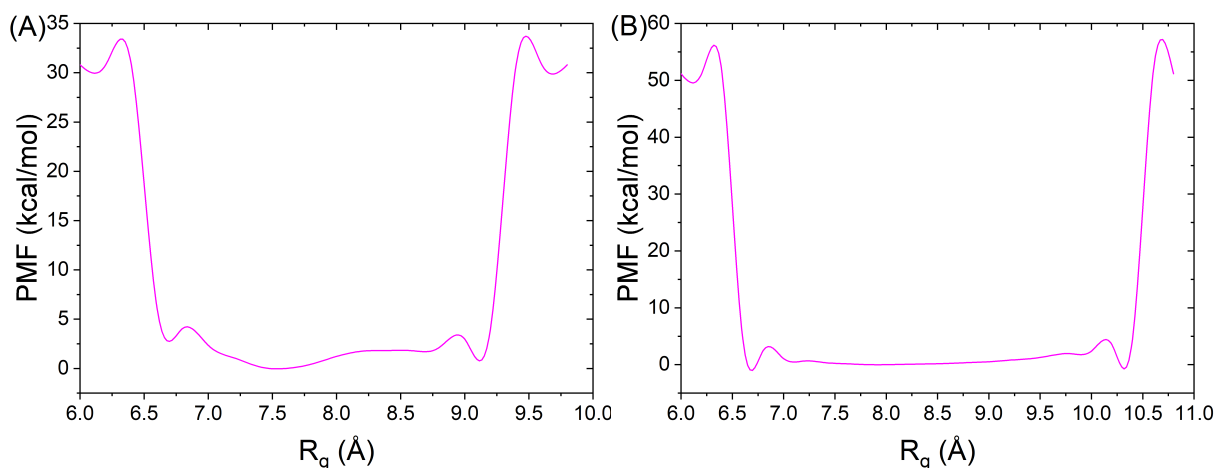
Figure 3.19: 1D free energy plots of R_g, during aMD simulations using (A) implicit and (B) explicit water models.

The free energy profiles for the radius of gyration, also confirm the observations made from the average and standard deviation values calculated above. The high standard deviation in the aMD simulations of the system in explicit solvent is verified from the pictorial representation of it in Figure 3.19, where the peptide appears to spend time between various conformations ($R_g \approx 7.5 - 9$ Å). The free energy profile of the system in implicit solvent shows how the most stable conformation of the peptide is at $R_g = 7.5$ Å, although coming at a lower PMF than the system in explicit, suggesting the more well-defined minima to be present in the explicit environment.

The end-to-end distances give an indication of the compactness of the peptide, by considering the distance between the two chain ends. When these are plotted against the radius of gyration, it becomes more apparent at what distances the most stable configuration of the peptide occurs. Looking at Figure 3.20, it comes at no surprise that the system in explicit solvent covers a more dispersed area of the phase space, Figure 3.20(B) – something already hinted at by the standard deviation values from the gyration tensor calculations. The energy minima each of the systems reach are relatively similar, with the chain in implicit solvent displaying two: one at end-to-end distance between 11 and 13 Å and R_g at *ca.* 7.6 Å; and another at end-to-end distance of *ca.* 6 Å and R_g at *ca.* 7.5 Å. The system in explicit solvent, on the other hand, has a more dispersed energy profile, with a less defined minimum at end-to-end distance between 9 and 14 Å and R_g between 7.8 and 8.5 Å. These are in accordance with the average R_g values, calculated above, Table 3.11.

(A)

(B)
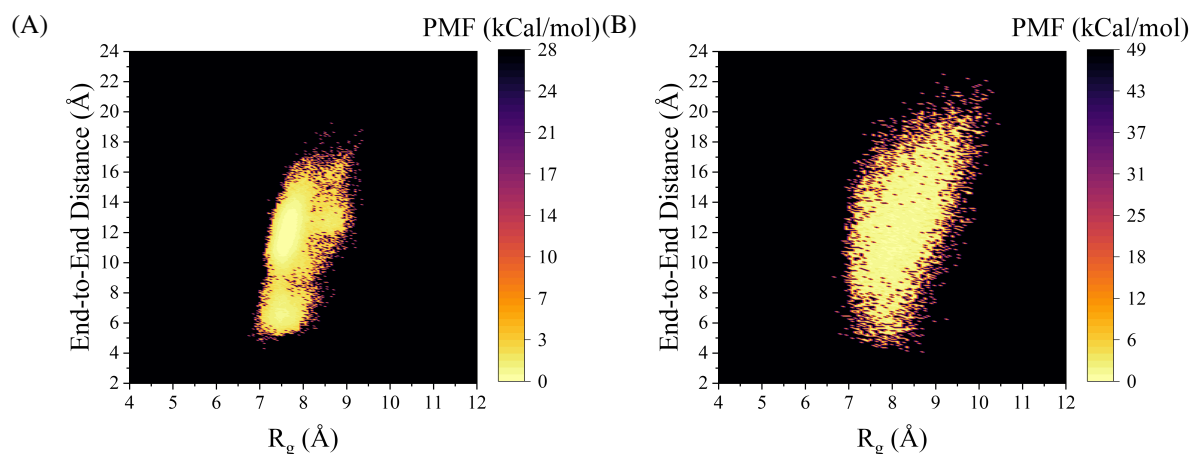
Figure 3.20: 2D free energy plots showing the end-to-end distance (Å) against the R_g, for aMD simulations using (A) implicit and (B) explicit water models.

3.6 GFN-xTB Simulations

The semiempirical tight-binding, GFN2-xTB approach, was also employed as a way to simulate the dynamics of the system. The results, however, showed that this method does not bode well when put against the data obtained from the molecular mechanics approach. The main issue faced when implementing this method, was the indication of a 5th coordination site, either Gly9 or Arg5, with the classical dynamics and metadynamics (MTD), respectively. These two residues have not been reported before as Zn(II) binding sites and hence come in contrast with the reported conformations in the literature,[49] except a study employing proton and nitrogen Heteronuclear Single Quantum Coherence (HSQC) spectroscopy, which reported these residues to exhibit peak broadening upon Zn(II) binding to nearby sites, signalling towards a potential chemical exchange or decrease in the peptide motion from binding.[50] Furthermore, Arg5 has been hypothesised as a potential fourth residue for Zn(II) binding, but lacks evidence to support its presence as a 5th binding site.[4, 51, 52] The fluctuations in distances of the interacting atoms in each of the binding residues as a function of time, are given in Figure 3.21.
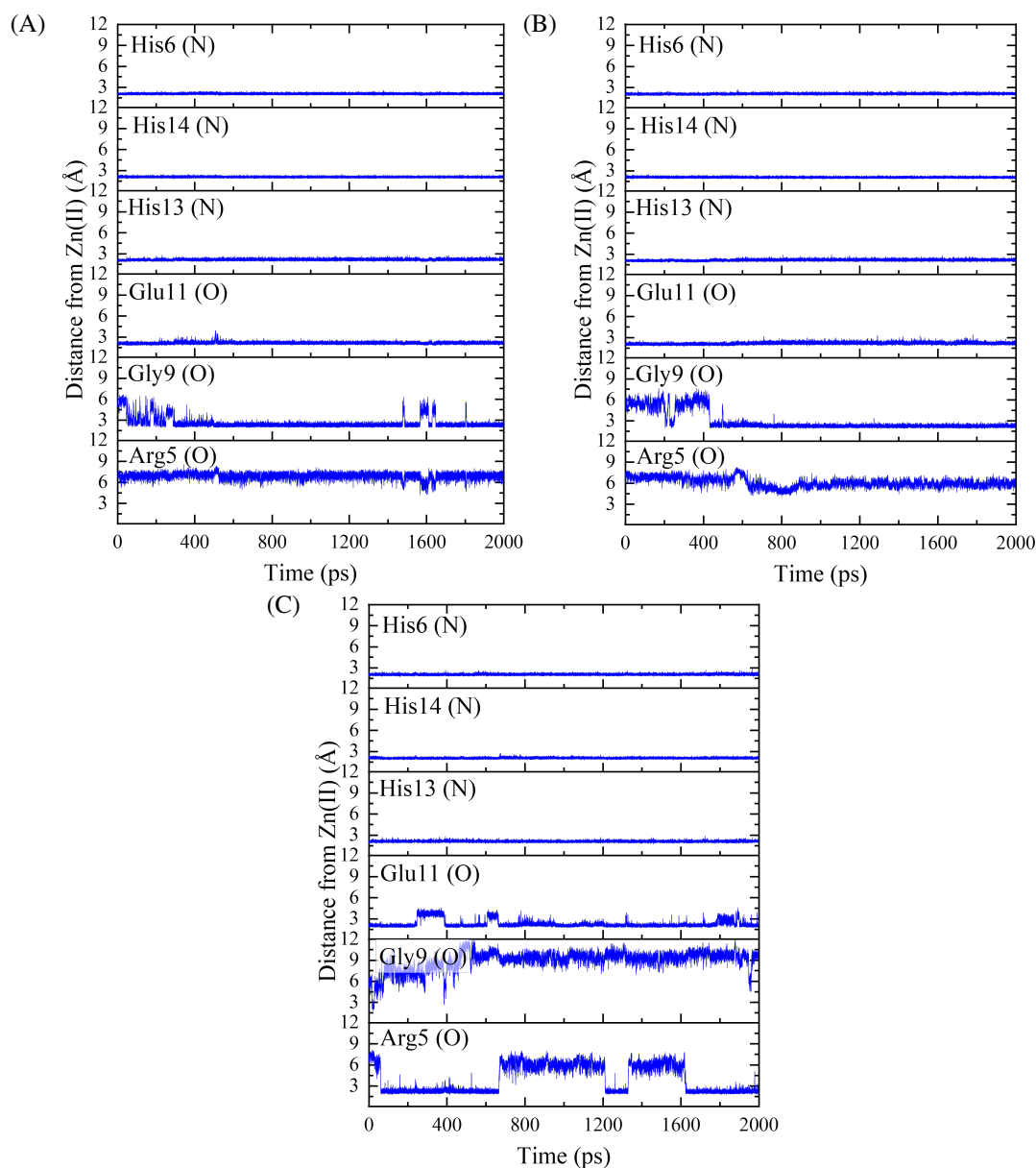
Figure 3.21: Plots of distance of interacting atoms to Zn(II), using classical dynamics in (A) GBSA and (B) ALPB, and (C) MTD in GBSA.

3.7   Conclusions

This chapter evaluated the performance of different AMBER force fields and solvent models, from simulations of the dynamics of Aβ_{16}. The findings were compared to the NMR-derived structure of the peptide in complexation with Zn(II).[4] For this, three different approaches were employed: classical and accelerated molecular dynamics, and semiempirical metadynamics, using GFN2-xTB.[34]

We find that the ff14SB force field resulted in the trajectories with the mean structure most comparable to experiment. Even though the data may suggest that explicit models solve some of the issues faced when carrying out implicit solvent simulations, such as an overestimation of salt bridges and α-character, the efficiency of the implicit models, at a considerably minimal

loss of accuracy, makes them the preferred choice – especially for larger systems. Thus, when used for computationaly costly systems, the implicit solvent setup is the preferred choice. In some cases, such as when using fb15, these provide an even better description of the native state of the peptide, compared to explicit models. The results from the semiempirical approach proved less reliable, with regards to the interactions of the peptide with the metal ion, as it was found to interact with residues that have not been reported before as Zn(II) binding sites. To conclude, the results presented here, may thus be applied to the study the structure of larger systems, such as the full-length Aβ peptide, to study ligand-protein and protein-protein inter-actions, as a way of understanding the mechanisms behind the aggregation of the peptide.

Bibliography

(1) Lermyte, F. et al. (2019). Metal Ion Binding to the Amyloid β Monomer Studied by Native Top-Down FTICR Mass Spectrometry. *Journal of the American Society for Mass Spectrometry 30*, 2123–2134.

(2) Reiss, A. B., Arain, H. A., Stecker, M. M., Siegart, N. M., and Kasselman, L. J. (2018). Amyloid toxicity in Alzheimer's disease. *Reviews in the Neurosciences 29*, 613–627.

(3) Lee, M. C. et al. (2018). Zinc ion rapidly induces toxic, off-pathway amyloid-β oligomers distinct from amyloid-β derived diffusible ligands in Alzheimer's disease. *Scientific Reports 8*, 1–16.

(4) Zirah, S. et al. (2006). Structural Changes of Region 1-16 of the Alzheimer Disease Amyloid β-Peptide upon Zinc Binding and in Vitro Aging. *Journal of Biological Chemistry 281*, 2151–2161.

(5) Scott, A. P., and Radom, L. (1996). Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Møller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *Journal of Physical Chemistry 100*, 16502–16513.

(6) Li, P., and Merz, K. M. (2016). MCPB.py: A Python Based Metal Center Parameter Builder. *Journal of Chemical Information and Modeling 56*, 599–604.

(7) Kennedy-Britten, O. D., Al-Shammari, N., and Platts, J. A. (2020). Molecular dynamics simulations of copper binding to N-terminus mutants of amyloid-β. *Journal of Biomolecular Structure and Dynamics 0*, 1–11.

(8) Kennedy-Britten, O. D., Alshammari, N., and Platts, J. A. (2021). Accelerated Molecular Dynamics to Explore the Binding of Transition Metals to Amyloid-β. *ACS Chemical Neuroscience 12*, 4065–4075.

(9) Al-Shammari, N., Savva, L., Kennedy-Britten, O., and Platts, J. A. (2021). Force-field evaluation and accelerated molecular dynamics simulation of Zn(II) binding to N-terminus of amyloid-β. *Computational Biology and Chemistry 93*, 107540.

(10) Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters 246*, 122–129.

(11) Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996). Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *The Journal of Physical Chemistry 100*, 19824–19839.

(12) Lindorff-Larsen, K. et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics 78*, 1950–1958.

(*13*)   Maier, J. A. et al. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation 11*, 3696–3713.

(*14*)   Wang, L. P. et al. (2017). Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *Journal of Physical Chemistry B 121*, 4023–4039.

(*15*)   Nguyen, H., Roe, D. R., and Simmerling, C. (2013). Improved generalized born solvent model parameters for protein simulations. *Journal of Chemical Theory and Computation 9*, 2020–2034.

(*16*)   Best, R. B., Zheng, W., and Mittal, J. (2014). Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *Journal of Chemical Theory and Computation 10*, 5113–5124.

(*17*)   Onufriev, A., Bashford, D., and Case, D. A. (2000). Modification of the generalized born model suitable for macromolecules. *Journal of Physical Chemistry B 104*, 3712–3720.

(*18*)   Voelz, V. A., Dill, K. A., and Chorny, I. (2011). Peptoid conformational free energy landscapes from implicit-solvent molecular simulations in AMBER. *Biopolymers 96*, 639–650.

(*19*)   Patel, S., Ramanujam, V., Srivastava, A. K., and Chary, K. V. (2014). Conformational propensities and dynamics of a $\beta\gamma$-crystallin, an intrinsically disordered protein. *Physical Chemistry Chemical Physics 16*, 12703–12718.

(*20*)   Shell, M. S., Ritterson, R., and Dill, K. A. (2008). A test on peptide stability of AMBER force fields with implicit solvation. *Journal of Physical Chemistry B 112*, 6878–6886.

(*21*)   Lei, H., Wu, C., Wang, Z.-X., Zhou, Y., and Duan, Y. (2008). Folding processes of the B domain of protein A to the native state observed in all-atom ab initio folding simulations. *The Journal of Chemical Physics 128*, 235105.

(*22*)   Rungnim, C., Rungrotmongkol, T., Hannongbua, S., and Okumura, H. (2013). Replica exchange molecular dynamics simulation of chitosan for drug delivery system based on carbon nanotube. *Journal of Molecular Graphics and Modelling 39*, 183–192.

(*23*)   Gallego-Villar, L. et al. (2013). Functional characterization of novel genotypes and cellular oxidative stress studies in propionic acidemia. *Journal of Inherited Metabolic Disease 36*, 731–740.

(*24*)   Debye, P. (1909). Der Lichtdruck auf Kugeln von beliebigem Material. *Annalen der Physik 335*, 57–136.

(*25*)   Hestenes, M., and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards 49*, 409.

(*26*)   Izaguirre, J. A., Catarello, D. P., Wozniak, J. M., and Skeel, R. D. (2001). Langevin stabilization of molecular dynamics. *The Journal of Chemical Physics 114*, 2090–2098.

(*27*)   Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics 23*, 327–341.

(*28*)   Jorgensen, W. L. (1981). Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *Journal of the American Chemical Society 103*, 335–340.

(*29*) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics 79*, 926–935.

(*30*) Wang, L. P., Martinez, T. J., and Pande, V. S. (2014). Building force fields: An automatic, systematic, and reproducible approach. *Journal of Physical Chemistry Letters 5*, 1885–1891.

(*31*) Huang, J., and MacKerell, A. D. (2018). Force field development and simulations of intrinsically disordered proteins. *Current Opinion in Structural Biology 48*, 40–48.

(*32*) Dick, T. J., and Madura, J. D. (2005). Chapter 5 A Review of the TIP4P, TIP4P-Ew, TIP5P, and TIP5P-E Water Models. *Annual Reports in Computational Chemistry 1*, 59–74.

(*33*) Zapletal, V. et al. (2020). Choice of Force Field for Proteins Containing Structured and Intrinsically Disordered Regions. *Biophysical Journal 118*, 1621–1633.

(*34*) Bannwarth, C., Ehlert, S., and Grimme, S. (2019). GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation 15*, 1652–1671.

(*35*) Berendsen, H. J., Postma, J. P., Van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics 81*, 3684–3690.

(*36*) Spicher, S., and Grimme, S. (2020). Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angewandte Chemie - International Edition 59*, 15665–15673.

(*37*) Sigalov, G., Fenley, A., and Onufriev, A. (2006). Analytical electrostatics for biomolecules: Beyond the generalized Born approximation. *The Journal of Chemical Physics 124*, 124902.

(*38*) Grimme, S. (2019). Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *Journal of Chemical Theory and Computation 15*, 2847–2862.

(*39*) Huang, H., and Simmerling, C. (2018). Fast Pairwise Approximation of Solvent Accessible Surface Area for Implicit Solvent Simulations of Proteins on CPUs and GPUs. *Journal of Chemical Theory and Computation 14*, 5797–5814.

(*40*) Shao, Q., and Zhu, W. (2018). Assessing AMBER force fields for protein folding in an implicit solvent. *Physical Chemistry Chemical Physics 20*, 7206–7216.

(*41*) Tao, P., and Xiao, Y. (2020). Using the generalized Born surface area model to fold proteins yields more effective sampling while qualitatively preserving the folding landscape. *Physical Review E 101*, 62417.

(*42*) Duan, Y. et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry 24*, 1999–2012.

(*43*) Okur, A. et al. (2006). Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. *Journal of Chemical Theory and Computation 2*, 420–433.

(*44*) Geney, R., Layten, M., Gomperts, R., Hornak, V., and Simmerling, C. (2006). Investigation of salt bridge stability in a generalized born solvent model. *Journal of Chemical Theory and Computation 2*, 115–127.

(*45*) Sullivan, S. S., and Weinzierl, R. O. (2020). Optimization of molecular dynamics simulations of c-MYC1-88—an intrinsically disordered system. *Life 10*, 1–20.

(*46*) Ollila, O. H., Heikkinen, H. A., and Iwaï, H. (2018). Rotational Dynamics of Proteins from Spin Relaxation Times and Molecular Dynamics Simulations. *Journal of Physical Chemistry B 122*, 6559–6569.

(*47*) Okur, A., Wickstrom, L., and Simmerling, C. (2008). Evuation of salt bridge structure and energetics in peptides using explicit, implicit, and hybrid solvation models. *Journal of Chemical Theory and Computation 4*, 488–498.

(*48*) Miao, Y. et al. (2014). Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation 10*, 2677–2689.

(*49*) Alies, B. et al. (2016). Zinc(II) Binding Site to the Amyloid-β Peptide: Insights from Spectroscopic Studies with a Wide Series of Modified Peptides. *Inorganic Chemistry 55*, 10499–10509.

(*50*) Rezaei-Ghaleh, N., Giller, K., Becker, S., and Zweckstetter, M. (2011). Effect of zinc binding on β-amyloid structure and dynamics: Implications for Aβ aggregation. *Biophysical Journal 101*, 1202–1211.

(*51*) Danielsson, J., Pierattelli, R., Banci, L., and Gräslund, A. (2007). High-resolution NMR studies of the zinc-binding site of the Alzheimer's amyloid β-peptide. *FEBS Journal 274*, 46–59.

(*52*) Zirah, S. et al. (2003). Zinc binding properties of the amyloid fragment Aβ(1-16) studied by electrospray-ionization mass spectrometry. *International Journal of Mass Spectrometry 228*, 999–1016.

Chapter 4

Cu$^{\text{II/I}}$ – α-Synuclein

In this chapter, the differences in the structure and dynamics, between the copper-bound and metal-free α-Synuclein (αS) are described. The peptide has been found to accumulate in senile plaques, specifically through the non-Aβ component (NAC) found in the central region of the peptide, between residues 61 and 95, Figure 4.1.[1] The metal-bound αS, is believed to increase the aggregation propensity of the peptide, resulting in an accelerated development of Parkinson's disease[2] The structural alterations that contribute to this proliferation of the aggregated form, are still elusive from wet-lab experiments, owing to the highly disordered nature of the protein.[3] The free-αS was simulated using two starting conformations, an experimentally defined structure from an NMR study on the free-peptide confined in a micelle,[4] and a fully extended conformation modelled in MOE, shown after an initial minimisation in Figure B.1.[5] Having obtained the preliminary results of the study, the free-peptide simulated from the extended conformation proved to be marginally better in replicating observations made in wet lab experiments, pertaining to the secondary structure and intra-peptide interactions, than the NMR-obtained structure.[4, 6–8] The effects of Cu(II)-binding to αS, have been attributed to two possible mechanisms. One where the Cu(II)-αS complexes stabilize intramolecular interactions, resulting in self-assembly of the peptide into fibrils. The other, where redox-mediated reactions induce oxidation of αS through electron donors, resulting in oligomerisation.[9–12] The Cu(I)-αS system is a result of the redox-mediated catalytic release of reactive oxygen species (ROS), thought to be a seed for αS aggregation and fibril formation.[13] Both of the copper-bound systems were modelled in their extended conformation in MOE, where they were initially minimised and subsequently used to obtain the metal site parameters, using quantum mechanical calculations with the aid of the MCPB package included with AMBER.[14]

[1]MDVF[M]KGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKEGVV[H]GVATVAEKTK[60]
[61]EQVTNVGGAVVTGVTAVAQKTVEGAGSIAAATGFV[95]
[96]KKDQLGKNEEGAPQEGILED[M]PV[D]P[DNE]AYE[M]PSEEGYQDYEPEA[140]

Figure 4.1: Primary structure of αS, separating the three main regions of the peptide: N-terminal (residues 1-60), non-amyloid β component (NAC – residues 61-95) and C-terminal (residues 96-140). The sites involved in Cu(II) binding include: M1, D2, H50, D119, D121, N122, E123 (red). The ones for the Cu(I) binding include: M1, M5, M116, M127 (underlined).

4.1   Computational Methods

The methodology used here, resembles the one described in the previous chapter. The choice of force fields to examine under the current system, was made after considering the results from the implicit solvent simulations, described in the previous section, showing great reproduction of experimental data using the ff14SB/GBSA combination.[15] The ff03ws force field was also chosen in view of its performance with IDPs, accurately reproducing experimental observations from small-angle X-ray scattering (SAXS),[16, 17] and nuclear magnetic resonance (NMR) studies.[18] Although the studies that implemented the ff03ws for simulations of αS, were done using explicit solvent, the performance of the ff03 force field,[19] coupled with the OBC implicit solvent, has been well documented in the literature, both in free and ion-containing systems.[20–25] The ff03ws force field has been refined against ensemble-averaged experimental observations for peptide systems in aqueous solution, and has been found to address deficiencies relating to secondary structure bias, seen in ff03, particularly important when looking into IDPs.[26] Acknowledging the backbone corrections on the amino acids, made to refine the force field for use with the TIP4P/2005 explicit solvent in ff03w, and additional scaling in ff03ws, the force field was kept the same, both in the explicit and implicit simulations. The LEaP[27] function was used to solvate the systems with the GBSA solvent model for ff14SB, and TIP4P/2005,[28] as well as the Onufriev, Bashford, Case (OBC) modification to the generalised Born (GB) model, for the ff03ws force field.[29–32]

An initial assessment was performed on the starting conformation, where the free peptide was simulated with two starting conformations – an experimentally defined structure from an NMR study on the free-peptide confined in a micelle,[4, 8] and a fully extended conformation modelled in MOE.[5] The two metal sites, where the copper ions interact with the peptide, were parametrised using the metal centre parameter builder (MCPB) program,[14] using angle, bond and charge parameters obtained through Gaussian09[33] using B3LYP/6-31G(d).[34] The Seminario[35] method and restrained electrostatic potential (RESP) fitting scheme,[36–38] were utilised to obtain harmonic force constants and atomic charges from DFT calculations. After minimisation, conventional molecular dynamics (cMD) simulations were performed on the unbound system for 150 million steps, at a 2 fs time step, resulting in 300 ns worth of data. The NVT ensemble was employed at 310 K, with the Langevin thermostat.[39] SHAKE[40] holonomic constraints were imposed on bonds to hydrogen, restraining them to their equilibrium length. After examining the root mean square deviation (RMSD) and radius of gyration (R_g) plots from the whole trajectory, 50 ns were discounted from the beginning of the trajectory to account for equilibration of trajectories.

Having completed these simulations, the final, folded conformation of the metal-free peptide was used to run explicit solvent MD simulations, so as to avoid the significantly larger water box required to simulate the dynamics starting from an extended conformation. The explicit solvent used with the chosen force field was TIP4P/2005,[28] after considering its performance in biological systems,[41] as well as the force field's optimisation around that specific water model.[42] A periodic solvent box was built using LEaP, with a 50 Å distance between the atoms of the peptide and the box edge. The large distance of the water box was chosen, having tested 20 and 35 Å, to avoid interaction with neighbouring simulation boxes.[43] The explicit simulations here, were again performed on a single 300 ns run. Heating of the systems was performed within 3 ns, using a collision frequency of 5 ps^{-1}, a Lennard-Jones cut-off of 10 Å, and 2 fs time step. Equilibration of these simulations first employed the NPT ensemble for 8 ns

at the same cut-off and time-step, with the Monte Carlo barostat at a pressure of 1 bar. This was followed by NVT simulation for 100 ns at a cut-off of 10 Å and 2 fs time step, with Langevin dynamics for temperature scaling at 310 K and a collision frequency $\gamma = 2$ ps^{-1}, before moving to the production run, in the NVT ensemble.

The MD simulations performed here may be split into four simulation scenarios:

1. For the evaluation of the starting conformation and force field, conventional MD were performed for 300 ns, with the initial 50 ns discarded as equilibration steps;

2. Explicit solvent simulations of the free peptide were performed on a single 300 ns cMD trajectory;

3. Having performed the initial assessment, conventional and accelerated MD simulations[44] for the free and Cu(II)-bound peptide, were performed in sets of 3 individual runs of 600 ns, with the initial 50 ns from the cMD simulations discounted for equilibration;

4. Accelerated MD simulations on the Cu(I) system were performed for 3×600 ns runs, having completed three initial 100 ns cMD simulations, to get the boost potentials. The aMD simulations were initiated from the final conformation of conventional MD simulations, providing equilibrated systems and three individual velocities.

The implicit solvent conventional MD simulations were extended to three individual 600 ns trajectories, for both the unbound and Cu(II)-bound systems, with the rest of the parameters same as above. Having obtained the data from the conventional MD simulations in implicit solvent, an enhanced sampling of the phase space was performed using accelerated MD.[44] The mean total potential energy was used to impose a bias on the system, using a boost potential, pushing the system to unexplored local minima positions. The aMD simulations were a continuation on the initial three cMD trajectories, each with their own final velocity and conformation. The enhanced sampling achieved using aMD, permits the sampling of trajectories that would remain uncharted by standard MD. Free energy landscape plots were constructed through reweighting,[44] and carma[45] was used to obtain clusters from principal component analysis (PCA) of the cartesian coordinates.The rest of the analysis was performed using the cpptraj[46] tool, acquiring data on secondary structure, root mean square fluctuation (RMSF), salt bridges, hydrogen bonding, RMSD and R_g.

Experimental studies, looking at the Cu(II)-αS complex, implicate regions [1]MDVFMKGLS[9], [48]VAHGV[52] and [119]DPDNEA[124] as the metal ion-binding sites.[47–49] In the present chapter, the macro-chelated form (i.e. interaction with distanced residues) of the N-terminal coordination is explored, which involves the interaction of Cu(II) with residues M1, D2 and H50, Figure 4.2, both including and excluding the C-terminal binding mode, between residues D119, D121, N122 and E123. Cu(I) has been reported to interact with αS through S in methyl amino acids found in the two chain ends, again shown in Figure 4.2.[13, 49–55]

(A)

(B)

Figure 4.2: Schematic of the two metal ion binding sites, in the (A) Cu(II) and (B) Cu(I) systems.
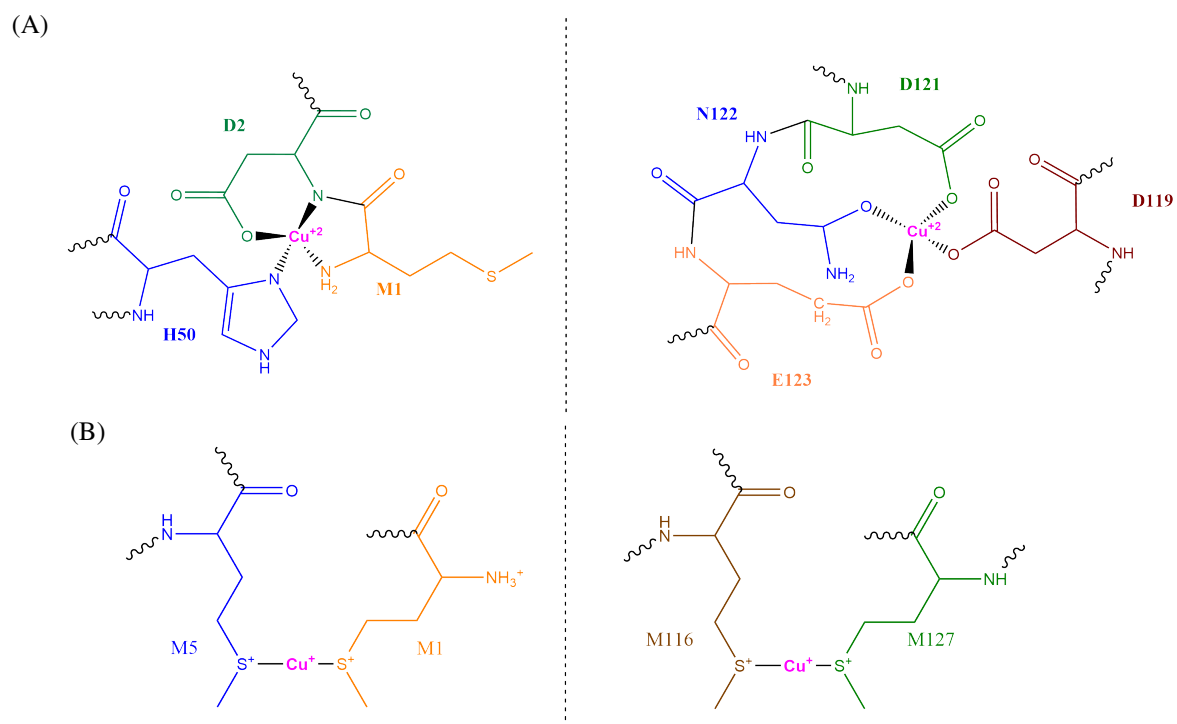
4.2 Evaluation of Methodology

4.2.1 Starting conformation and force field in implicit solvent

The free peptide was initially simulated in implicit solvent, with two starting conformations – one already folded in a horseshoe-like configuration, from an NMR experiment where the peptide was studied in a micelle,[4] and the other in an extended form, Figure B.1. Having performed 300 ns conventional MD simulations on the two systems, the first 50 ns were discounted, being regarded as equilibration steps. The decision on the length of the trajectory that needed to be removed was made after looking at the R_g and RMSD data, shown in Figure 4.3. Henceforth the data analysed only includes values after the equilibration steps have been removed.
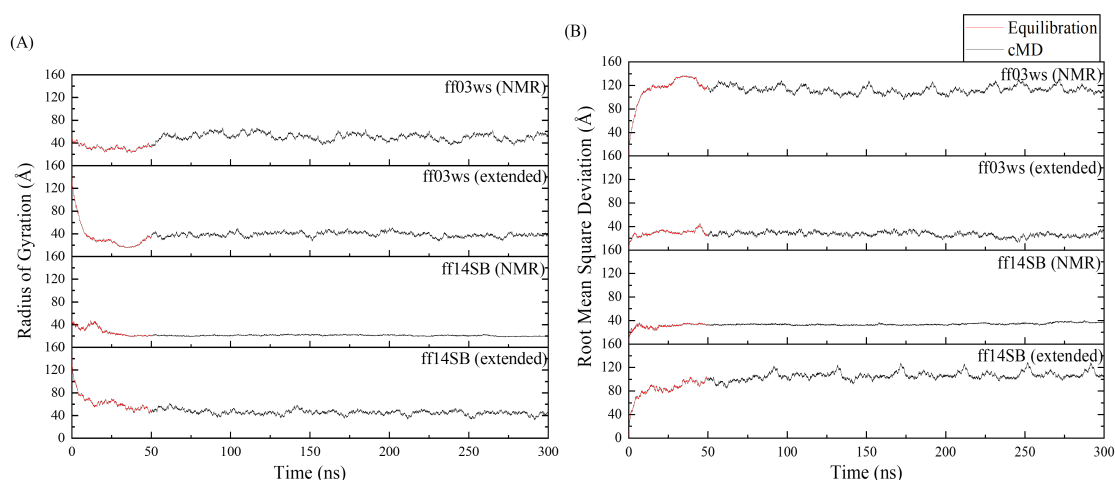
Figure 4.3: (A) Radius of gyration and (B) root mean square deviation (RMSD) plots, showing the change from equilibration to conventional MD, for the different systems studied.

The secondary structural characterisation of α-Synuclein has been elusive in experimental studies, owing to its highly disordered nature. The general consensus amongst researchers involving this peptide, has been that the lack of a defined structure, makes it hard to identify a structure-function relationship.[56, 57] The majority of studies surrounding the characterisation of the peptide, primarily rely on results from circular dichroism (CD) spectroscopy. A survey of the reported characteristics from past studies is given in Table 4.1.

Table 4.1: Literature survey of reported secondary structural characteristics from different experimental methods.

Method	Reference	Conditions	α-helix (%)	β-strand (%)	Other (%)
Circular Dichroism	[6]	Untreated αS (0.074 mg/mL), 293.15 K, pH 7.4.	2±3	11±7	86±22
	[58]	αS (0.2 mg/mL), Tris-HCl (25 mM), NaCl (50 mM), 293.15 K, pH 7.4.	~0	31	68
	[59]	Purified αS from mouse brain (7.5 μM), PBS[a] (25%), 293.15 K, pH 7.4.	22.5±1.5	n/a	46.5±12.5
	[60]	αS (2.0 mg/mL), PBS[a] (100 mM), NaCl (100 mM), pH 7.4.	19±1	n/a	n/a
Raman Spectroscopy	[61]	αS (200 μM), Tris-HCl (20 mM), pH 7.5.	49	10	41
	[56]	αS (300 μM), PBS[a] (20 mM), pH 7.5.	48	15	37
Surface Enhanced Raman Spectroscopy (SERS) – Optical Tweezers	[62]	Supernatant αS (2 mM) {from cells grown in lysogeny broth with 50 μg/mL Kanamycin and 0.3 mM IPTG, incubated at 277.15 K and centrifuged for 45 min (+60 min after boiling to remove *E.coli* proteins)}, gradient elution using NaCl, Tris-HCl (25 mM), pH 7.4.	13.3	8.4	78.3
		Supernatant αS (250 μM), *same method as above*.	15.1	9.5	75.4
Attenuated Total Reflectance Fourier-Transform Infrared (ATR-FTIR) Spectroscopy	[63]	Supernatant αS {from lyophilized sample (2-5 mg) dissolved in ddH$_2$O, NaOH (100 mM), pH 10±0.5, incubated at 293.15 K and centrifuged for 30 min.}, HCl (1 mM), Tris-HCl, pH 7.4.	35	3	62

[a]Phosphate-buffered saline

The variability in the characteristics of the peptide, along with the general uncertainty that surrounds the reported secondary structure of αS in the literature, makes it hard to use such values to assess the validity of simulation methodologies. Despite the discrepancies encountered from the use of different methods, the changes in the secondary structure, from alterations in the conditions and the environment in which the peptide is modelled, can still be evaluated.

Perhaps a better approach for assessing the possible structural characteristics of αS is through backbone and sidechain ^{13}C NMR chemical shifts, which provide more easily reproducible predictions on the local covalent interactions within the peptide. The values for the chemical shifts were obtained through SPARTA+.[64] The predicted chemical shifts derived from deviations of Cα and Cβ from their random coil values are given in Figure 4.4, with the predicted Cα chemical shift values obtained from the simulations, plotted against experimental data, given in Figure 4.5 (full data in Appendix B).[65] The plots point towards the ff03ws/OBC combination, to give the structure that more closely resembles experimental NMR predictions, particularly where the two termini are concerned, closely matching residues 10-21 and 114-131, in the negative region which they appear, Figure 4.4, indicating a propensity for β-sheets.[66] These observations correlate with the predicted Cα chemical shift values, showing a close resemblance to the experimental predictions throughout, but in particular in the two termini, with mean percentage error for the N- and C-termini at 1.73% and 1.98%, respectively, and 2.35% for the NAC region, totalling a 1.97% mean error for the whole peptide. These values increase all through the peptide simulated with the ff14SB/GBSA parameterisation, raising the mean error to 2.62%, with the three individual regions averaging values between 2.52% and 2.74%, Table B.2.

Further key characteristics, that correlate to experimental findings, include the presence of a combination of positive and negative shifts, representing α- and β-characteristics, respectively, in the N-terminal region of the peptide, with the C-terminal consisting mainly of β-character, and the NAC region of the peptide presenting a combination of the two structures.[66, 67] The lack of serial negative chemical shifts in the core regions, as well as the N-terminus, of the ff14SB plot, brings into question the assessment of β-structures using this force field.
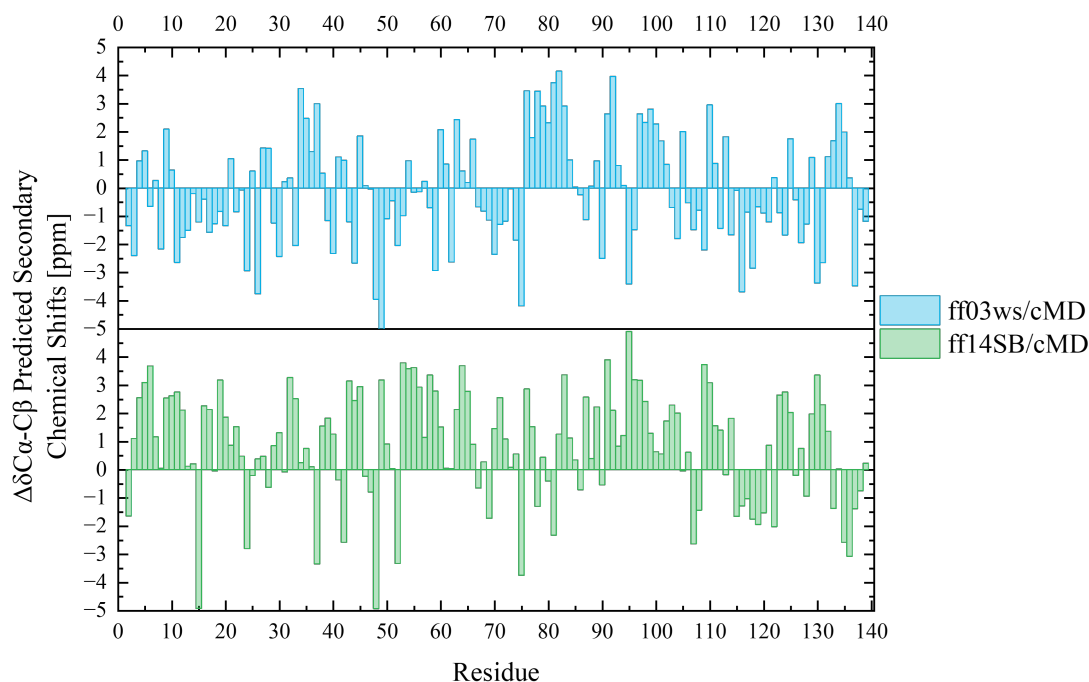
Figure 4.4: Deviation of Cα-Cβ chemical shifts from their random coil values. The results consist of 1250 frames, taken from conventional MD simulations of the free-αS, starting from extended conformation.
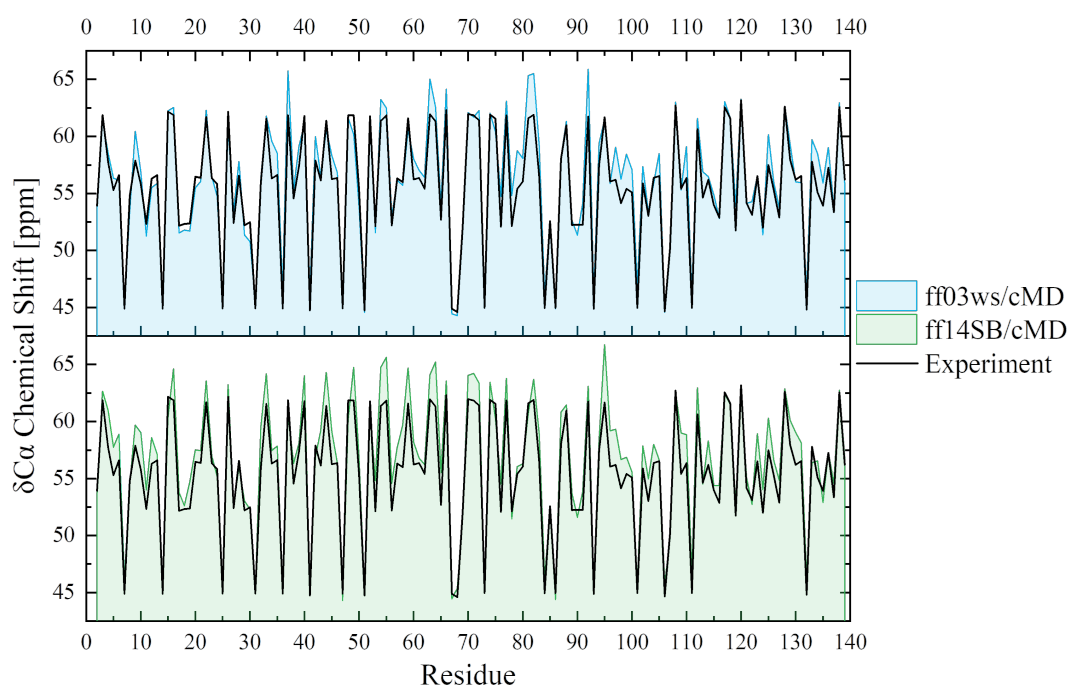
Figure 4.5: Predicted Cα chemical shift values per residue, from 1250 frames taken from the conventional MD simulations on αS, using different force field/solvent combinations, starting from the extended conformation. Experimental data obtained from source.[65]

Specific motifs that appear in residue repeats have also been used here, alongside radius of gyration, to corroborate experimental and computational observations. The greatest contribution of secondary structural characteristics, and in particular β-sheets, has been found in or around the KTK[EQ][QG]V repeats (between residues 21-26, 32-37 and 43-48); starting from the N-terminal region and extending into the NAC, through [58]KTKEQV[63], which has been reported to have a great influence on the stability of the NAC region.[68] These regions have also been reported in the past to be involved in the compact topological arrangement in α-Synuclein, while also being present in other forms of synuclein.[69, 70] The contribution of these repeats, along with the residues surrounding them, in the overall folding of the peptide can also be seen from the salt bridges formed between the Lys and Glu residues contained within these regions (*vide infra*, Figure 4.9). These repeats have been found to result in the formation of α-helices and β-hairpin structures, through the presence of anti-parallel β-sheets, Figure 4.6 and Figure 4.7. Similar β-hairpin structures have also been observed here between residues 63-72, existing in a more permanent fashion. These locations, have mainly been taken into consideration when evaluating the efficacy of the force fields to simulate trajectories that come in agreement with experimental findings; as they constitute the most well defined characteristic of synucleins.
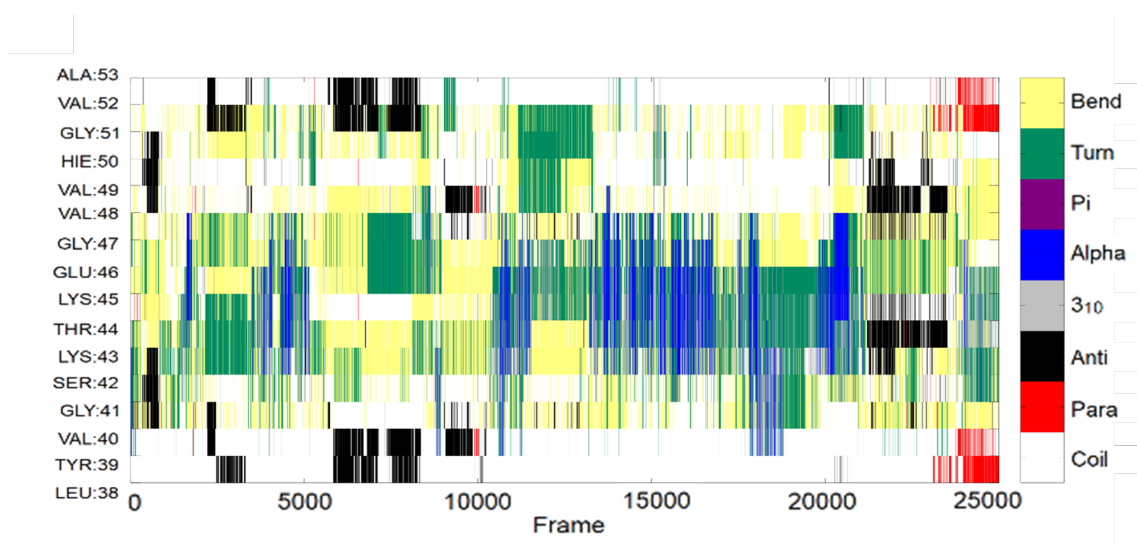
Figure 4.6: Evolution of secondary structure characteristics of each of the residues in the hairpin region, found within the free peptide over the length of the conventional MD trajectory, using **ff03ws/OBC**. The antiparallel (black) β-sheets with turn/bend (yellow/green) between indicate possible presence of β-hairpin.
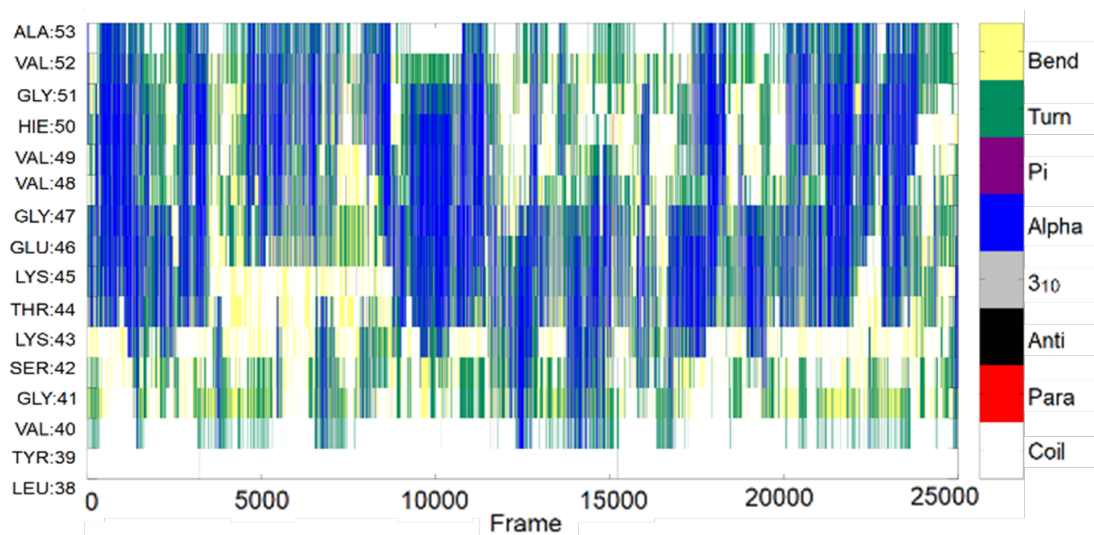
Figure 4.7: Evolution of secondary structure characteristics of each of the residues in the hairpin region, found within the free peptide over the length of the conventional MD trajectory, using **ff14SB/GBSA**. The lack of antiparallel β-sheets suggests the absence of β-hairpins in this system.

Secondary structure characteristics as a function of residue are shown in Figure 4.8, with a numerical representation of the overall percentages in Table 4.2, and a breakdown of the secondary structure contribution of the two termini and central NAC regions in Table 4.3. From the data presented here, it is evident that the central region of the peptide accounts for the highest percentage of β-character in the overall system. The most discernible observation that may be made on the difference of the four systems, is the lack of β-characteristics with ff14SB, especially where the extended starting conformation is concerned where it only accounts for 0.12% of the secondary characteristics, preferring the helical and coiled structural arrangement.

Table 4.2: Secondary structure percentages for the free peptide cMD simulations.

Forcefield	Starting point	β character (%)	α character (%)	Other (%)
ff03ws/OBC	NMR	3.73	25.57	70.69
ff03ws/OBC	Extended	5.68	24.25	70.06
ff14SB/GBSA	NMR	1.63	47.20	51.17
ff14SB/GBSA	Extended	0.12	40.57	59.31
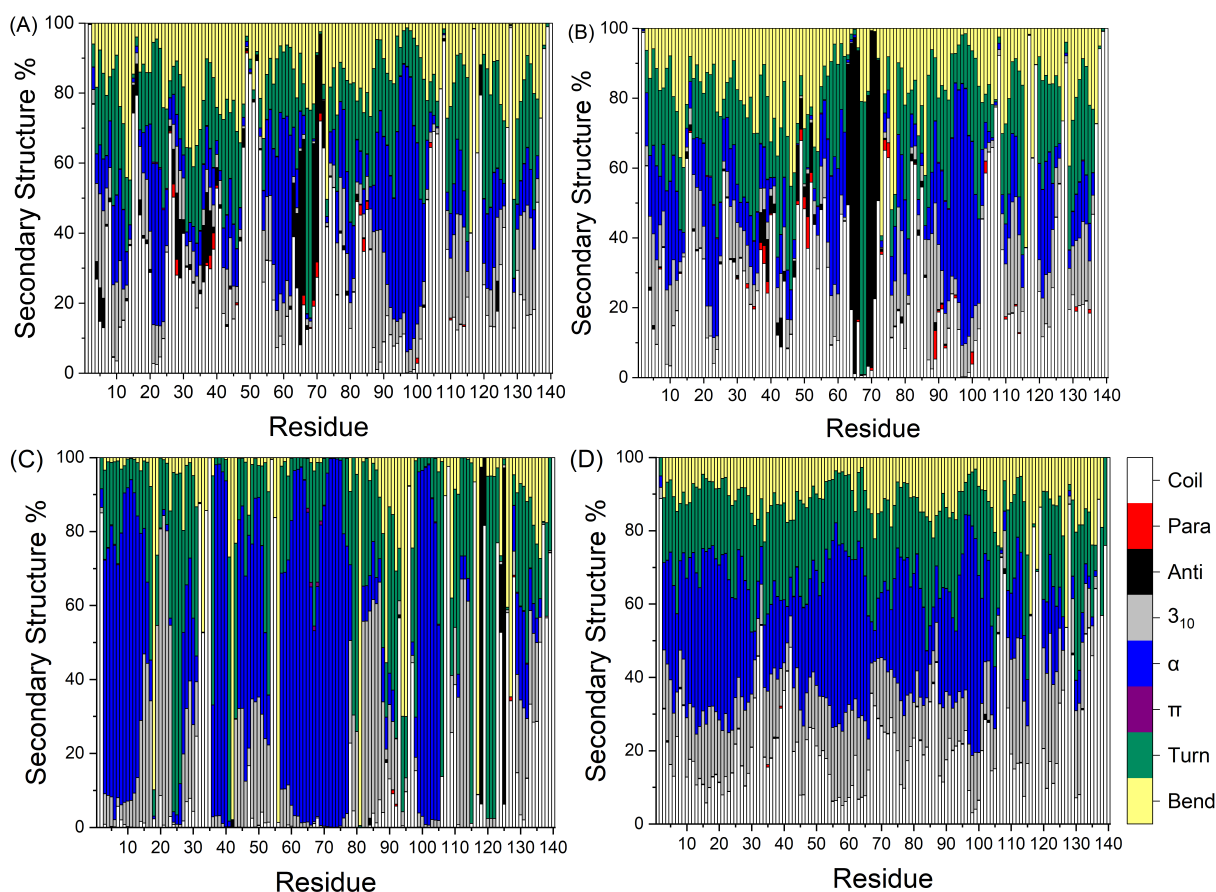
Figure 4.8: Secondary structure distribution per residue after 250 ns cMD using ff03ws with (A) NMR-defined and (B) extended starting conformations, and ff14SB with (C) NMR-defined and (D) extended starting conformations. The β-sheets are denoted with red (parallel) and black (antiparallel), and helices with grey (3_{10}), blue (α) and purple (π).

The folding characteristics expressed using ff03ws appear to better correlate with experimental observations on this system, Table 4.1. This is largely due to the lack of β-character in the ff14SB-simulated system, which evidently is not the case in any experimental data. In particular, looking at the ATR-FTIR and SERS spectroscopic results,[62, 63] acceptable reproduction of the experimental values was achieved with the ff03ws simulations, with β-sheet percentages ranging between 3.5-6% and α-helices between 24-26%, compared to 3% and 8.4-9.5% β-sheet, and 35% and 13.3-15.1% α-helical characteristics reported from the experiments, respectively. The overestimation of α-helices in the ff14SB simulations is in accord with the Raman spectroscopic results,[56, 61] although the near-null value of the β-sheet character in the system with the extended starting conformation, brings the validity of the force field into question.

The secondary structures found in the three main regions of the peptide, are presented in Table 4.3. The systems simulated with ff03ws/OBC appear to follow the same trend, where the β-structure is concerned, as the central NAC region displays the greatest percentage contribution to the overall peptide's character, followed by the N- and finally the C-terminal. This is in agreement with experimental observations, where the NAC region, encompassing the most hydrophobic residues, has been shown to promote protein aggregation, in contrast to the C-terminus which evidently decreases that propensity.[71]

Table 4.3: Detailed secondary structure percentages of the three main regions of αS, after cMD simulations.

Region	β character (%)	α character (%)	Other (%)
ff03ws/OBC (NMR)			
N-terminal	3.22	24.04	72.74
NAC	8.74	27.84	63.42
C-terminal	0.52	25.86	73.62
ff03ws/OBC (extended)			
N-terminal	2.75	24.25	72.99
NAC	17.64	21.84	60.52
C-terminal	0.30	26.12	73.59
ff14SB/GBSA (NMR)			
N-terminal	0.03	54.02	45.95
NAC	0.15	60.44	39.41
C-terminal	4.90	27.83	67.27
ff14SB/GBSA (extended)			
N-terminal	0.11	47.13	52.76
NAC	0.10	43.11	56.79
C-terminal	0.15	29.84	70.01

The R_g data, shown in Table 4.4, was compared to experimental measurements using Guinier analysis of small-angle X-ray scattering (SAXS) curve on the free peptide. The two references found, report R_g of 40 ± 1 Å,[7] and 35.5 ± 0.5 Å.[72] Starting from the extended conformation resulted in mean R_g of 38.8 Å, closely resembling the mean of two experimental measurements. The systems with the NMR-derived starting configuration, either greatly overestimate or underestimate that value, for both ff03ws and ff14SB. This was somewhat expected, as previous studies indicate that equilibration of αS in explicit solvent requires simulations in the order of several μs to escape conformational basins.[73] The systems simulated with ff03ws, from both NMR and extended starting structures, and ff14SB with extended starting conformation, display a relatively unrestricted flexibility in water, a known property of IDPs.[74] Where the ff03ws force field is concerned, although convergence to experimental values wasn't achieved with the NMR starting conformation, the extended peptide reached that point within the 300 ns trajectory, apparently due to the lack of bias towards any secondary structure at the outset, allowing the simulation to approach equilibrium in a reasonable timescale. The NMR-derived trajectory with ff03ws is expected to eventually arrive at the desirable values, but significantly longer simulation times would be required to escape the bias from the starting point. The system simulated with ff14SB and started from the NMR-derived conformation, results in R_g values quite far from experimental observations. Considering the overestimation of α-helices in this system, and how it appears to have already settled within a possible local minimum, Figure 4.3, the ergodicity here is not believed to improve by increasing simulation time.

Table 4.4: R_g of cMD simulations of the free peptide.

Forcefield	Starting point	Mean R_g (Å)	SD (Å)	Max (Å)	Min (Å)
Experiment		40 ± 1[7] 35.5 ± 0.5[72]			
ff03ws/OBC	NMR	50.97	6.13	66.98	31.87
ff03ws/OBC	Extended	38.80	3.96	50.90	26.55
ff14SB/GBSA	NMR	20.94	1.05	24.14	17.99
ff14SB/GBSA	Extended	45.53	4.20	61.57	31.91

Despite the disparities encountered above, from the different parameterisation of the systems, the salt bridge interactions observed in three of the four instances, Figure 4.9, appear in a similar pattern, with a difference on the percentages in which they appear. This is especially evident when looking at ff03ws, with the two different starting configurations. There, a similar distance between positively and negatively charged residues is observed, permitting similar electrostatic attractions to form between them, maintaining those interactions at no more than 20%. The systems simulated with the ff14SB force field, present more permanent occupancy of those interactions, even in the peptide with the extended starting conformation, which sustains the same diagonal pattern observed in the ff03ws simulations, reaching percentages as high as 43%. The peptide starting from the NMR-derived structure, and simulated with ff14SB illustrates a much more unstructured pattern in the electrostatic interactions, with occupancies reaching to values >96%, bearing further evidence of a system well confined within a local minimum. A similar pattern is not observed in the hydrogen bonding between the ff03ws systems and the peptide simulated from the extended conformation using ff14SB, as illustrated in Figure 4.10. The latter system gave significantly weaker hydrogen bonds, compared to the other three systems. The lack of off-diagonal hydrogen bonding here, hints towards a linear arrangement of the residues within the peptide, also backed by the lack of β-sheets and overestimation of α-helices. The most significant hydrogen bond interactions in each of the four cases are shown in Table 4.5. The significant degree of hydrogen bonds between distanced residues, in the system simulated with ff14SB starting from NMR-derived conformation, illustrates the reason behind the greatly constrained peptide, as seen from the radius of gyration.
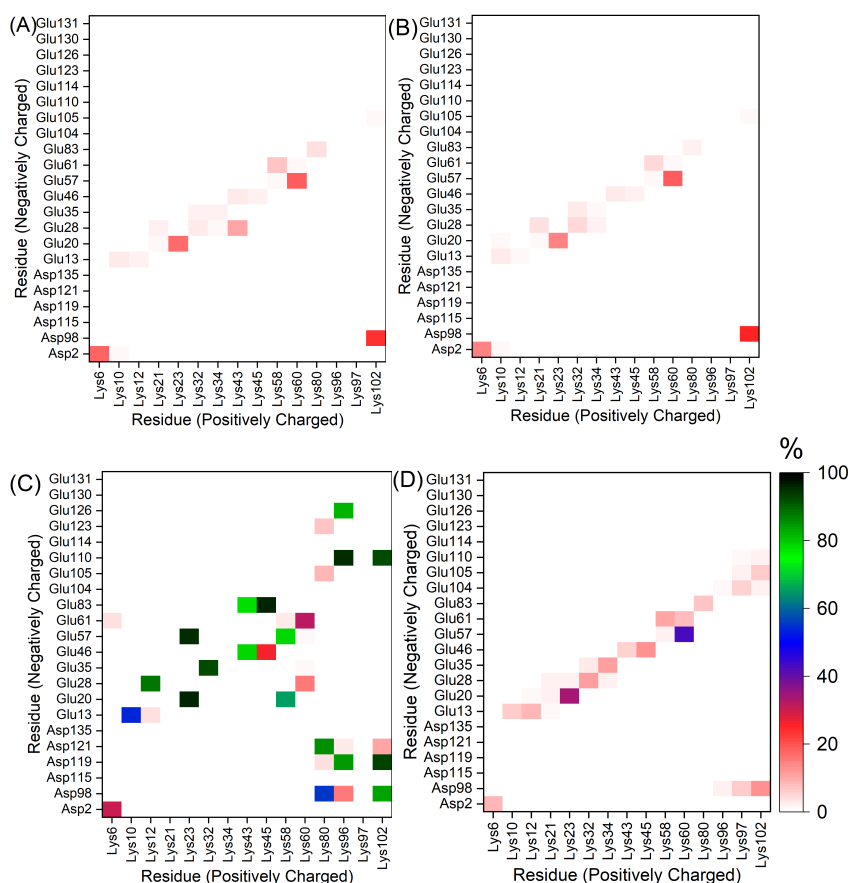
Figure 4.9: Salt bridges formed between negatively and positively charged residues from cMD using ff03ws with (A) NMR-defined and (B) extended starting conformations, and ff14SB with (C) NMR-defined and (D) extended starting conformations.

Table 4.5: Most significant hydrogen bond interactions found in the four systems with different starting conformations, simulated in implicit solvent. The percentage corresponds to the number of frames the hydrogen bond appears.

Forcefield	Starting point	Hydrogen Bond (%)
ff03ws	NMR	Thr59-Val63 (49)
ff03ws	Extended	Thr64-Val71 (78)
		Ala69-Val66 (78)
ff14SB	NMR	Ser42-Thr44 (75)
ff14SB	Extended	Val63-Gly67 (14)
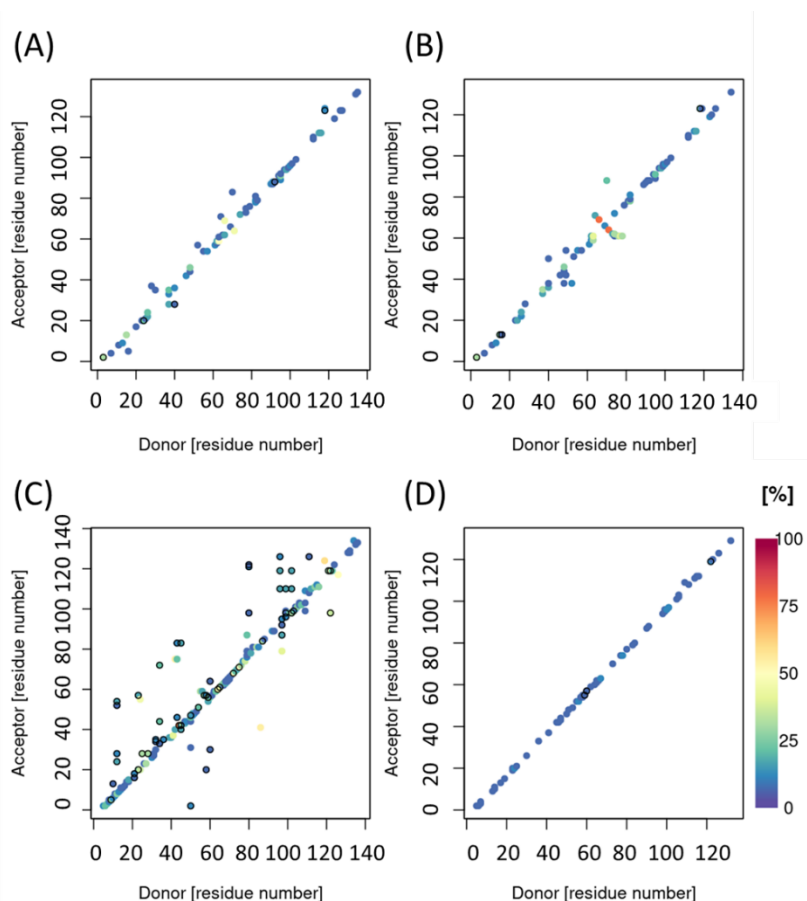
Figure 4.10: Hydrogen bonds present after cMD simulation, using ff03ws with (A) NMR-defined and (B) extended starting conformations and ff14SB with (C) NMR-defined and (D) extended starting conformations.

In spite of the documented performance of ff14SB/GBSA when used with more ordered peptide systems,[15] it is quite evident from the data collected here, that it is not adequate for use with αS. Considering the overestimation and difference in pattern of salt bridges, and the highly inconsistent hydrogen bond characteristics among the different starting conformations; along with the *ca.* 25 Å difference between the R_g values, it is clear that ff14SB/GBSA cannot be trusted to yield an accurate trajectory in this case. On the other hand, the simulations with ff03ws/OBC have proven to be much more reliable in providing results that are both attainable, to a degree, with different starting conformations, but more importantly close to experimental findings. This particular force field has long attracted attention for use with highly disordered systems, but since its inception it has only be tested in explicit solvent systems, with a proven track record of reliable results, especially where accurate reproduction of R_g is concerned, compared to other force fields.[42, 75–77] The system that was selected to be studied further, was the one with the extended starting conformation. In order to validate the efficiency of the implicit solvent in obtaining results with minimal loss of accuracy, the system has also been evaluated in explicit solvent.

4.2.2 Explicit Solvent Simulations

Having performed implicit solvent simulations on the two systems, the findings were sought to be corroborated using explicit solvent. Systems of this size, however, could not be studied starting from an extended conformation, as above, due to the very large explicit solvent box that would require, to permit the system to fold into a desirable conformation, whilst avoiding interactions with periodic images. The computational cost such setup entails would be too great to justify. Hence, the already folded structure for the system was acquired from the final frame of conventional MD simulation in implicit solvent, and a TIP4P/2005 periodic water box was built, using LEaP,[27] around the chain at a 50 Å distance from the atoms of the solute and the solvent box edge.

The R_g data in Figure 4.11 and Table 4.6, show a *ca.* 11 Å decrease in the size of the peptide, when simulating the peptide in explicit solvent. It is further clear that simulations in TIP4P/2005 solvent lead to much more compact conformations, than those observed either in experiment or in implicit solvent. In fact, the simulations performed in explicit solvent are so restricted that the *maximum* R_g sampled is smaller than the experimental average. This shortcoming of standard water models, especially TIP3P and TIP4P, in describing αS has been noted before.[72] Even models designed specifically for IDPs, such as TIP4P-D, are found to give overly compact ensembles, only approaching experimental values after reweighting against SAXS data.[72] Nevertheless, similar studies on αS have also reported R_g values within this range for explicit solvent simulations using ff03ws, as well as other force fields, with most of them giving values even less than the ones obtained with this particular force field.[75]
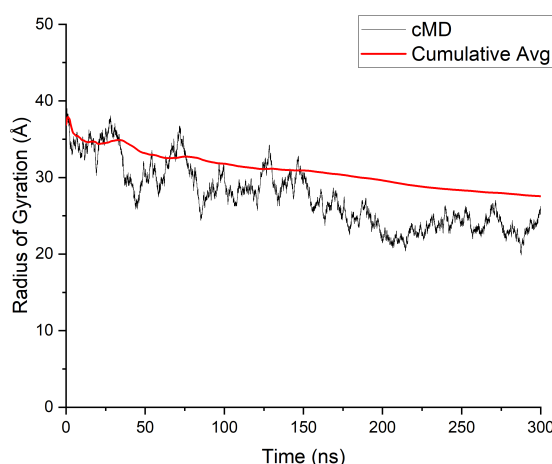
Figure 4.11: Radius of gyration plot from explicit solvent simulations on the free peptide, with the cumulative average shown in red.

Table 4.6: R_g of cMD simulations of the free peptide in explicit solvent. The data presented is from the production run (after 108 ns of equilibration in the NPT and NVT ensembles).

	Mean R_g (Å)	SD (Å)	Max (Å)	Min (Å)
free-αS (Experiment)	40 ± 1[7] 35.5 ± 0.5[72]			
free-αS (OBC)	38.80	3.96	50.90	26.55
free-αS (TIP4P/2005)	27.56	4.16	39.01	19.85

The secondary characteristics of the system simulated using the explicit solvent show greater population of β-sheets, with largely reduced α-helices, now only accounting for *ca.* 6% of the characteristics of the peptide, compared to the *ca.* 24% from the implicit solvent simulations, more closely matching the majority of experimental values, Table 4.1. A breakdown of the changes in the secondary characteristics of each of the residues, as a function of time is given in Appendix B, Figure B.2.
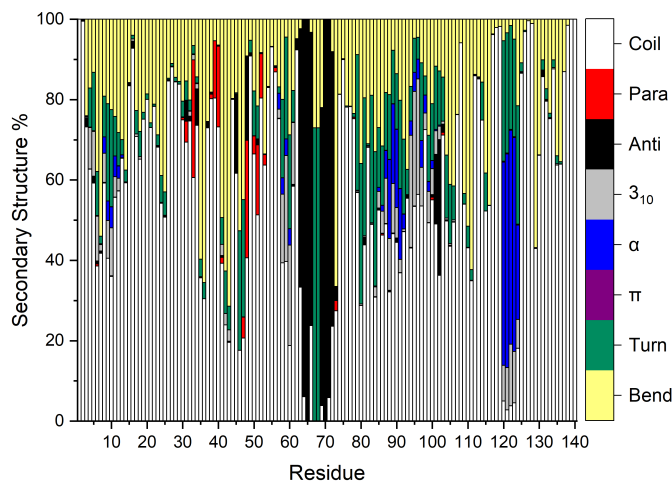
Figure 4.12: Secondary structure distribution per residue after cMD simulations of the free-αS in TIP4P/2005. The β-sheets are denoted with red (parallel) and black (antiparallel), and helices with grey (3_{10}), blue (α) and purple (π).

The contact maps of alpha carbons (Cα) and salt bridges between the residues comprising the synuclein chain are shown in Figure B.3 and Figure B.4, respectively. In both cases, the explicit water molecules appear to have a major effect on the interactions between formally distant residues. This comes in conflict with our previous findings on explicit solvent simulations using a smaller peptide, which did not show a large difference in the interactions found in implicit and explicit simulations.[15] The most likely reason behind this, is the effect of explicit solvent on the hydrophobic residues of the peptide, found especially in the NAC-region (residues 61-95).[78] The packing of these residues in explicit solvent, increases the rigidity of the peptide in the regions where these are found. The overall assessment of the above findings indicate that explicit solvent simulations do not adequately sample conformations reported by experimental investigations – at least with the specific solvent model used here. Radius of gyration and secondary characteristics obtained with this approach are in stark contrast to experimental findings on the peptide's conformational assembly.[7, 61] Similar observations have also been

reported in previous studies on the free α-Synuclein using ff03ws with TIP4P/2005, with the R_g reported to deviate from experimental findings due to insufficient sampling.[72, 79–81]

4.2.3   Parameterisation of metal sites

The metal sites of the modelled peptide were parameterised using the MCPB.py tool,[14] after establishing the metal ion binding sites from literature survey of experimental *in vitro* and *in silico* studies on the coordination of Cu(II)[47, 48, 71, 82, 83] and Cu(I).[13, 49–55] DFT calculations were then performed to assign harmonic bonds between the metal ions and the atoms involved in their coordination, with the aim of maintaining the distances and angles in a square planar arrangement during the MD simulations, by incorporating them into the force field parameters. In this way, the conformation of the metal-binding sites is expected to remain stable around the equilibrium determined by the QM calculations, throughout the simulations, owing to the high-energy force constants.

Table 4.7: Force constants and equilibrium distances of coordinating atoms to the **Cu(II)** metal centres, as calculated from B3LYP/6-31G(d) optimisation of the metal sites.

Metal site	Ligating atoms	Force constant (kcal mol^{-1}·Å^{-2})	Equilibrium distance (Å)
N-terminal	MET1 (N)	110.6	1.910
	ASP2 (N)	128.6	1.881
	ASP2 (O)	90.70	1.929
	HIS50 (N)	127.0	1.913
C-terminal	ASP119 (O)	93.20	1.905
	ASP121 (O)	117.8	1.879
	ASN122 (O)	1.400	2.298
	GLU123 (O)	54.20	2.030

Looking at the values from the QM calculations on the Cu(II) binding sites, Table 4.7 and Table B.4, a relatively consistent force is imposed on the ligating atoms in the N-terminus. This is not the case, however, in the C-terminal metal site, where smaller force constants are observed for N122 and E123. This allowed greater flexibility of the ligating atoms in these residues, yielding closer distances to the metal centre during the MD simulations, albeit with more fluctuations in the bond distance, compared to the other coordinating atoms, Table B.8. The stability in the distance of atoms in the metal-coordination sites, can be seen from Figure B.7, where the distances are maintained within bonding length; even for the O from Asn122, where the lowest force constant is seen, the bond distance is fluctuating between 1 and 2 Å.

Table 4.8: Force constants and equilibrium distances of coordinating atoms to the **Cu(I)** metal centres, as calculated from B3LYP/6-31G(d) optimisation of the metal sites.

Metal site	Ligating atoms	Force constant (kcal mol^{-1}·Å^{-2})	Equilibrium distance (Å)
N-terminal	MET1 (S)	79.7	2.146
	MET5 (S)	87.2	2.158
C-terminal	MET116 (S)	94.7	2.133
	MET127 (S)	94.5	2.133

The force constants obtained for the Cu(I) binding sites, Table 4.8 and Table B.5, are much more consistent for all the atoms involved in the coordination of the metal centres. Considering the linear binding mode of these sites, this is in the order of what was expected, with the distances and angles of the two binding sites, given in Table B.10 and Table B.11, depicting a fairly standard linear configuration of the atoms.

Analysis of the Cu(II) coordination sites (Figure B.7, Table B.8 and Table B.9) confirm that Cu-L distances and L-Cu-L angles are stable over the course of the entire trajectory. For the assessment of the most prominent geometry expressed by the peptide, clusters were created using the cartesian coordinates of the Cα – the different clusters are given in Table B.12. The coordination of ligating atoms in each of the Cu(II) and Cu(I)-binding sites, from the average cluster structure obtained from cartesian PCA analysis, is shown in Figure 4.13.

(A)

(B)

(C)

(D)
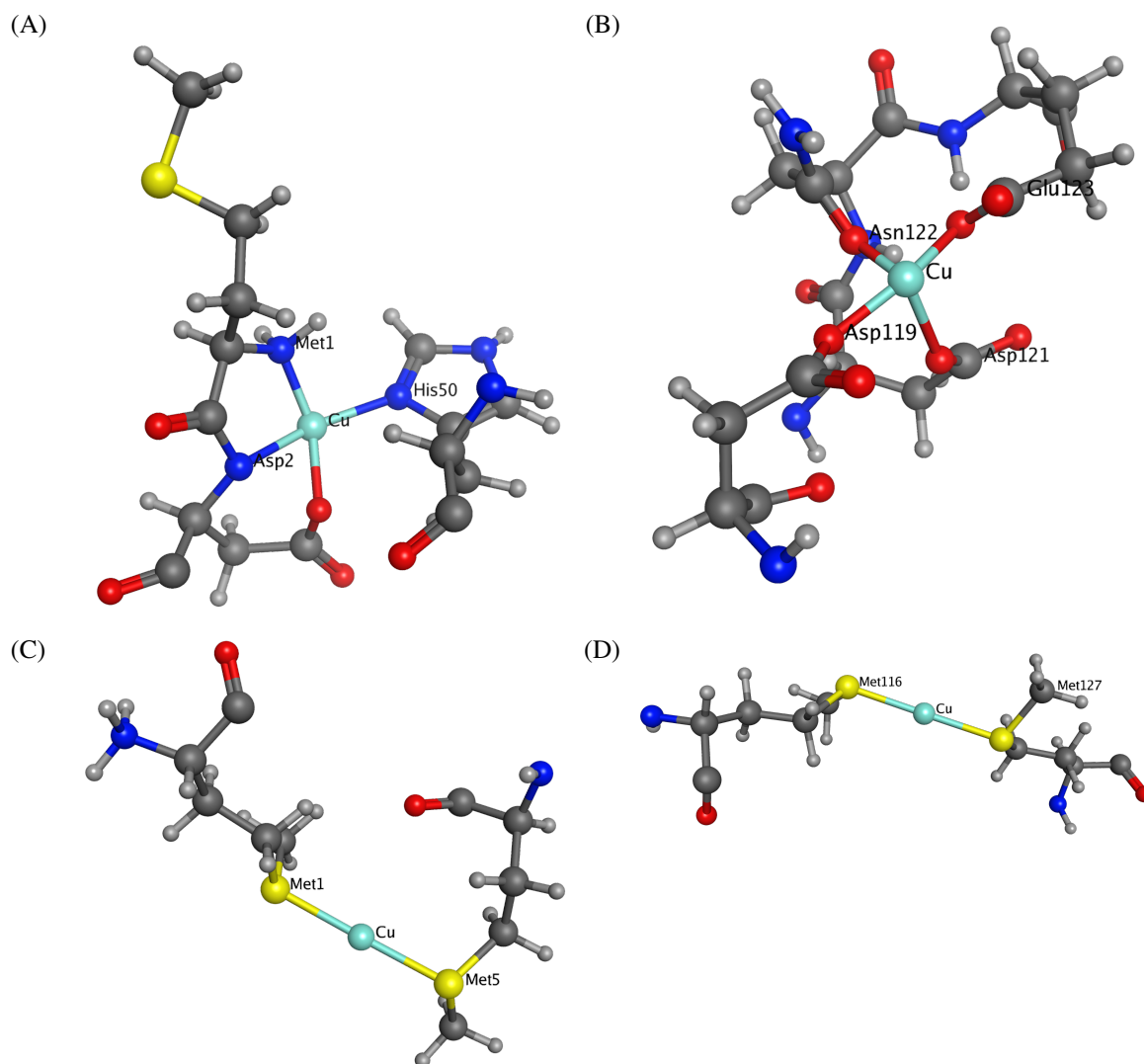
Figure 4.13: Coordination of ligating atoms in each of the (A – B) Cu(II) and (C – D) Cu(I)-binding sites, from the mean cluster structure.

The conformational assembly of the metal sites, seen in Figure 4.13, correlates with experimental observations, detecting a distorted square planar arrangement of atoms.[48, 84, 85] Interaction with atoms neighboring the ligands, distorts the geometry in each of the sites, exerting repulsion on the equatorial positions. These atoms are from a sidechain oxygen of Glu35, and the second oxygen in the Glu123 sidechain. The coordination of Asp2 in a bidentate fashion, adds to the strain of the structure of the first metal site, further distorting the square planar geometry.

4.3   Accelerated MD Simulations of Free and Cu^{II}-αS

Having analysed the conventional MD trajectories, the most promising system – ff03ws with extended starting conformation in implicit solvent – was chosen to extend the trajectory to 3×600 ns, as well as perform enhanced sampling simulations using aMD, again for 3×600

ns. Below, the results from the MD simulations on the metal-free and Cu(II)-bound αS (Figure 4.2), are reported. Experimental values for secondary structure and radius of gyration of the metal-bound peptide could not be found in the literature, hence the values are contrasted against the simulations on the free peptide.

The convergence of the system during each of the runs, was evaluated by plotting RMSD as a function of simulation time (Figure 4.14), as well as calculating cumulative average, which settle into constant values over the course of each trajectory. R_g values over the cMD and aMD trajectories, are shown in Figure 4.15, along with the cumulative average. The rolling standard deviation of the simulations against time were also plotted, Figure B.6, providing further evidence of equilibration over the course of each run. From these plots, it is evident that the system fluctuates between a range of R_g values, with a mean SD for the free peptide from cMD: 4.28 Å and aMD: 4.58 Å, and for the copper-bound one at cMD: 4.05 Å and aMD: 4.31 Å. The distribution plots of the R_g for each of the runs, are given in Figure B.5, where all runs are seen exhibiting peak values close to the average.
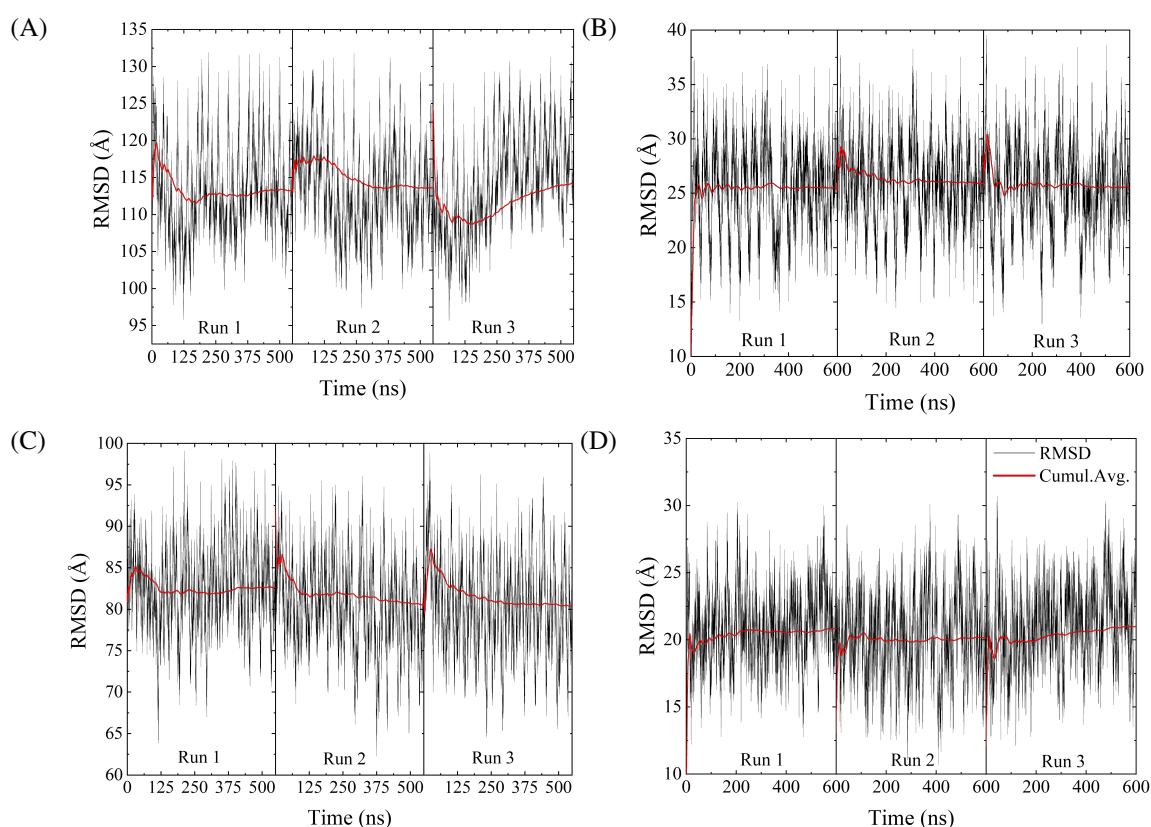
Figure 4.14: RMSD plots of (A) conventional and (B) accelerated MD of the free, and respectively for (C – D) Cu(II)-αS, with the cumulative average shown in red.
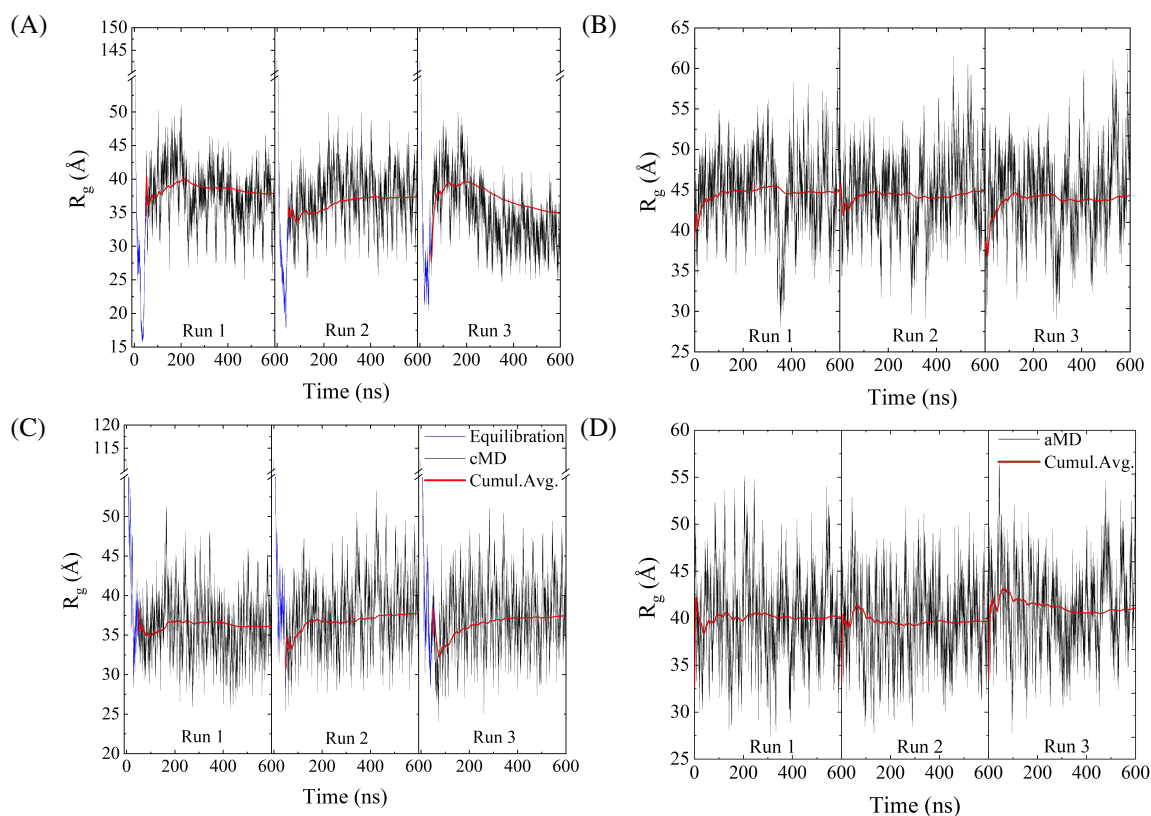
Figure 4.15: Radius of gyration plot, showing the (A) change from equilibration to cMD and (B) aMD simulations on the free and respectively (C – D) copper-bound peptide, with the equilibration steps in blue and cumulative average in red.

Table 4.9: R_g of conventional and accelerated MD simulations of the free and copper-bound peptides.

	Mean R_g (Å)	SD (Å)	Max (Å)	Min (Å)
free-αS (Experiment)	40 ± 1[7] 35.5 ± 0.5[72]			
free-αS (cMD)	36.65	4.28	50.90	24.67
free-αS (aMD)	44.26	4.58	61.50	28.07
Cu(II)-αS (cMD)	37.10	4.05	53.34	24.01
Cu(II)-αS (aMD)	39.81	4.31	56.53	25.48

Upon comparison of the R_g data, the change in the compactness between the free and Cu(II)-αS, is not as clear when looking at the conventional MD data, as the mean difference only comes at 0.45 Å. However, the accelerated MD simulations, give a much more substantial difference among the two systems, with the Cu(II)-αS being 4.5 Å more compact than the free-αS. Overall, the MD results of the free and Cu(II)-αS present differences, depending on the simulation protocol employed, also evident from a comparison of the secondary structure %, Table B.6, where a drop in the β-characteristics is seen in the aMD trajectories. For a better evaluation of the effect of metal ions on the peptide, one of the two methods has to be trusted that it provides the most reliable results over the other. In addition to the secondary structure

propensities reported in the experiments cited in Table 4.1, a comparison of the Cα chemical shifts from the accelerated MD simulations, Figure 4.16, further substantiate the effectiveness of the ff03ws/OBC parameters, in combination with the aMD methodology, in providing results that reflect experimental findings. The mean error seen between simulations and experiment in the chemical shifts, under these conditions is at 1.42%, Table B.1, hinting towards a great similarity in the local covalent interactions of the experimental and simulated systems. Considering the evidence presented here, the analysis of the simulations in implicit solvent, henceforth, focus on the data obtained from the accelerated MD trajectories.
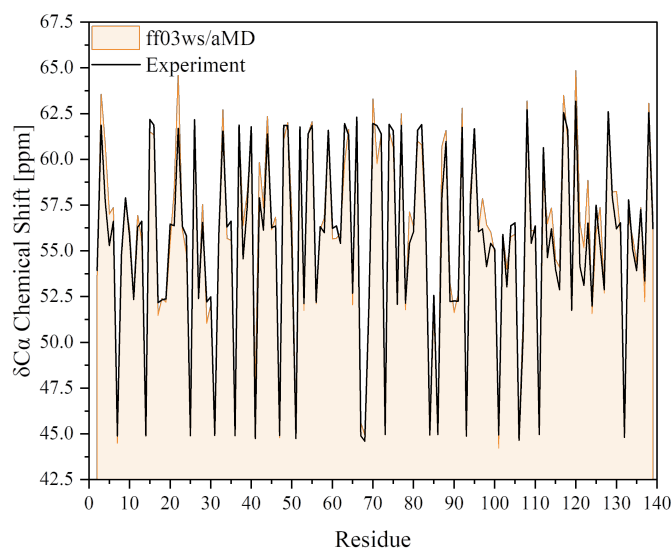
Figure 4.16: Predicted Cα chemical shift values per residue, from 1200 frames taken from the accelerated MD simulations on αS. Experimental data obtained from source.[65]

The standard deviation in R_g, as well as maximum and minimum values, are smaller in the Cu(II)-bound peptide, indicating more compact and less flexible structures in the presence of two metal ions. This comes as a result of the closer contacts developed within the peptide, especially from the macro-chelation formed with the coordination of the first two residues and His50. This increase in the intramolecular interactions is seen not only in the Cα contact maps (*vide infra* Figure 4.26), but also in the slightly increased sphericity of the copper-coordinated system, Table B.7 [b]. This is also seen at the free energy landscape of R_g against globularity, Figure 4.17, where the free peptide, although visiting conformations up to 0.5, on average displayed lower globularity values. The copper-bound peptide appears to have a more constrained sampling of the conformational space, owing to the increased ordering of residues bound to metal centres, decreasing flexibility of the overall system.

---

[b]The sphericity of the system was assessed from measurements of the globularity, after dividing the smallest by the largest diagonalized eigenvalues of the R_g tensor, whereby *Globularity* = λ_x/λ_z.
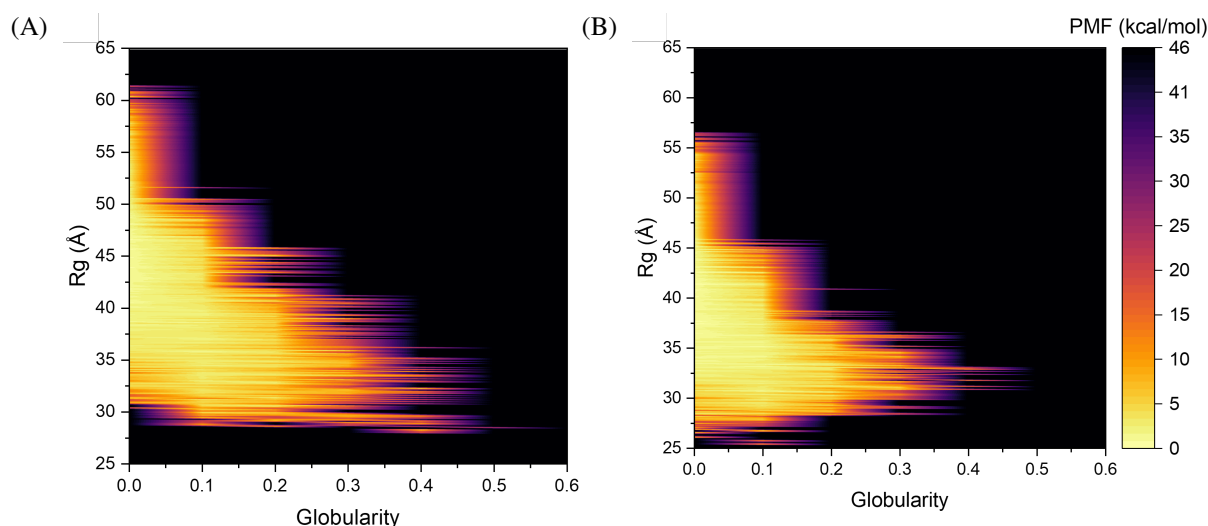
Figure 4.17: 2D free energy plot of the radius of gyration (R_g) against globularity, showing the potential of mean force (PMF) in the phase space, from the aMD simulations on the (A) free and (B) Cu(II)-bound peptide.

The secondary structure characteristics of the aMD trajectories change by quite a significant degree compared to conventional MD, owing to the increased range of conformations sampled using this method. These are detailed in Figure 4.19, Table 4.11 and Table 4.12. A significant contribution to the secondary structural characteristics of αS comes from motifs in or around repeating residue regions KTK[EQ][QG]V, found between residues 32-37, 43-48 and 58-63. These repeats start in the N-terminus and extend into the NAC region of αS, and have been implicated before with the ordered arrangement of the peptide.[*68–70*] Within these repeats, β-hairpin structures have been found in aMD trajectories for both the unbound and metal-bound peptide, corroborating the results found by Yu et al.,[*86*] on the region where these are observed (residues L38-A53) through the formation of anti-parallel β-sheets; Figure 4.18 shows the presence of these folding elements from the clustered structures. An experimental study looking at the nucleation capacities of different regions within α-Synuclein, has reported the region encompassing residues 37-61 to act as a nucleation-promoter, possibly as a result of the β-hairpin assemblies.[*87*] More recently, Y39 has been the focus of an experimental study, that concluded in the importance of the aromaticity in the folding mechanics of that region of the peptide.[*88*] Here, these β-hairpin structures appear more frequently in the Cu(II)-bound α-Synuclein trajectory compared to the unbound one, Figure B.8 and Figure B.9. Another region where β-hairpin structures have been found here, is between residues 63-72, expressed almost in twice as many frames as in the hairpin found between residues L38-A53. The maximum time these have been found to last in each of the cases, are given in Table 4.10, with the copper-bound system exhibiting the greatest persistency in both of these regions. Potential intramolecular interactions, Lysine (K) and Glutamine (E) residues within these repeats, are discussed in more detail below.
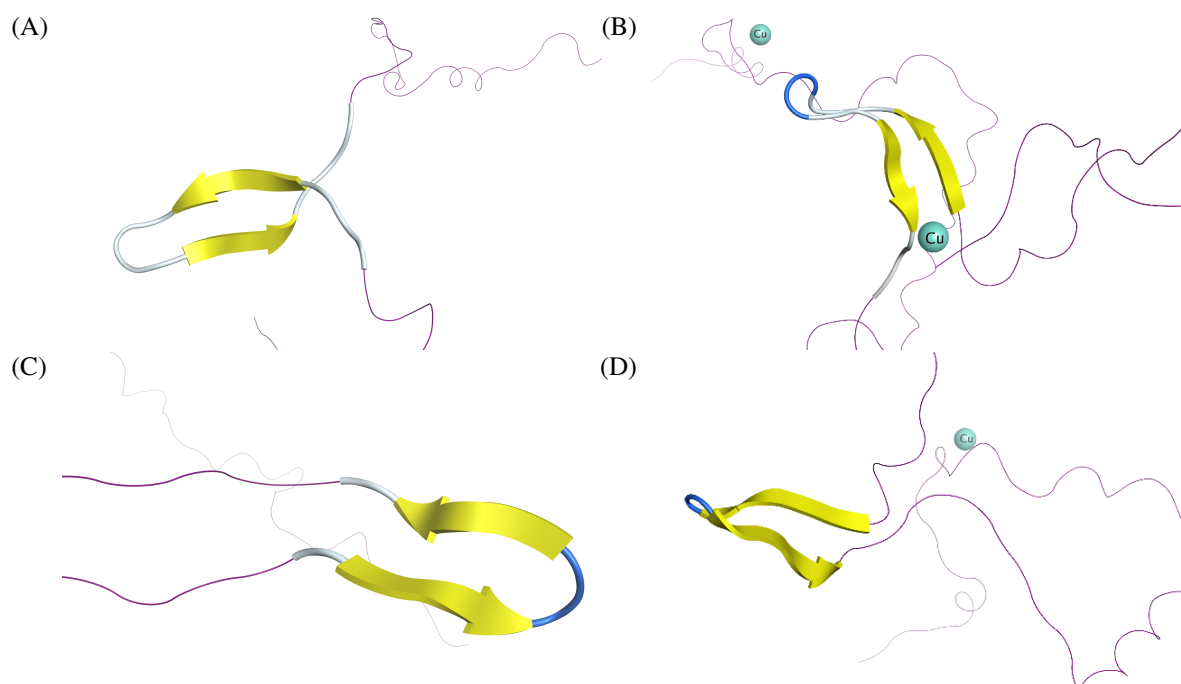
Figure 4.18: β-hairpin structure found between residues L38-A53 in the (A) unbound and (B) Cu(II)-bound αS and respectively (C-D) the same secondary structure between residues 63-72.

Table 4.10: Maximum time β-hairpins are maintained in the two residue ranges of L38-A53 and 63-72.

	Maximum Time (ns)	
	[38]LYVGSKTKEGVVHGVA[53]	[63]VTNVGGAVVT[72]
free-αS	42	54
Cu(II)-αS	74	60

The amounts of the different structures in the systems indicate the ratio of β-characteristics in the different regions of the peptide remains consistent with experimental observations, with the highest percentage of sheets present in the NAC region.[89] Despite the lack of experimental data on the secondary characteristics of the copper-bound peptide, an in-depth evaluation of the free peptide, with experimental findings from CD,[6, 58–60] Raman,[56, 61] optical trapping-assisted SERS,[62] and ATR-FTIR[63] spectroscopic techniques, and NMR data on the chemical shifts of the Cα within this system,[65] was given in previous sections. The simulated findings, here, showed agreement with values reported from the ATR-FTIR study (α-helices: 35%; β-sheets: 3%),[63] and α-helical percentages reported from the SERS (α-helices: 13.3-15.1%; β-sheets: 8.4-9.5%),[62] and CD experiments (α-helices: 19±1%[60] and 22.5±1.5%[59]). The level of agreement between the free-peptide secondary structure found here and the experimental values, permits the assessment of the differences between the two systems, with relative confidence in their reliability. The NAC region of the peptide has long been thought to be involved in the pathogenesis of PD, owing to the formation a hydrophobic β-sheet intermediate in that particular region.[90] The secondary structure in NAC is almost unaffected by binding of Cu(II), with a decrease in β-sheet, possibly owing to the pull exerted

on the residues comprising the NAC region, from coordination of His50 to Cu(II), but also from the increased preference for long-range interactions, upon binding of the metal ion, seen both in the increased compactness of the system, Figure 4.26 and Table 4.9, but also in the lack of 3_{10}-helices, as opposed to the metal-free peptide, where such structures make up most of the helical character in the system.

Conversely, a decrease in the helical character is observed in the residues involved in the copper interactions, in both N- and C-termini. This decreased helicity, may in turn influence the membrane binding affinity of the peptide, especially considering the higher affinity lipid membrane binding region is in the N-terminus.[91] Experimental evidence have also reported oxidation of Met residues in that region of the peptide results in a decreased membrane affinity,[92, 93] as well as the possible modulation of αS, as a result of protein interactions in that region.[94] This drop in the % of α-helices, particularly in the N-terminus of the Cu(II)-αS system, therefore suggests the possible hampering of the binding affinity with lipid membranes upon coordination of the metal-ion, as this region has generally been linked with the ability of αS to form such interactions,[95] aiding in the physiological activity of αS and as a way of balancing between the normal and aberrant forms.[96] The dampening of these interactions in the Cu(II)-coordinated system, may therefore act as a mechanism for the formation of toxic oligomers. The community is yet still divided on the possible effects of membrane coordination, with arguments on both sides: regulation of misfolding and oligomerization upon membrane binding,[97, 98] and promotion of aggregation.[99] Seeing as how the membrane interactions of these systems are not examined here, as well as the documented effect of membrane curvature,[100] these can only act as speculations on the possible effects when these systems do in fact bind. A recent experimental report on the possible mechanisms that take place upon interaction of αS with lipid membranes in the presence of Cu(II), showed two possible hypotheses: (1) an increased affinity of Cu(II) interactions to the N-terminal of monomeric αS, thus increasing oligomerization in-solution and decreasing upon membrane-binding; (2) free-αS membrane binding results in extended helical formation increasing the affinity of Cu(II) association with the C-terminal binding site.[91]
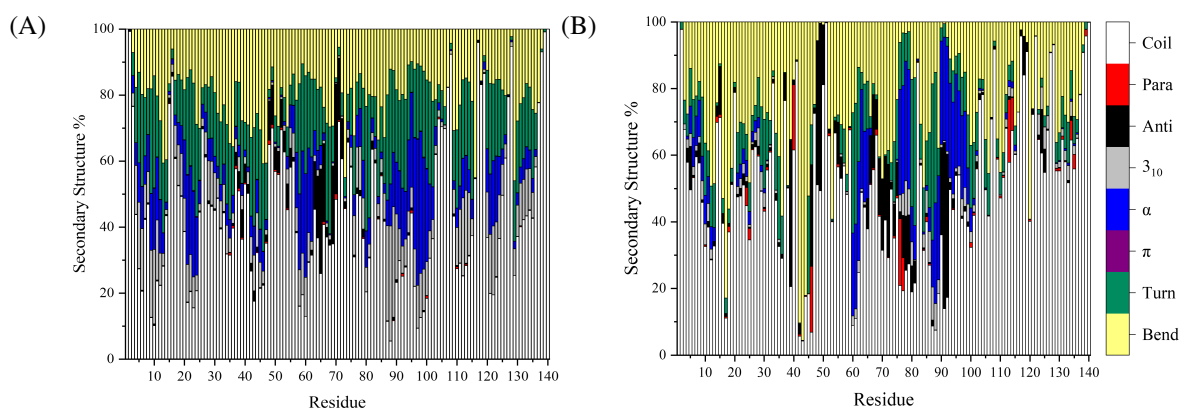
Figure 4.19: Secondary structure distribution per residue after 1.80 μs aMD of the (A) free and (B) Cu(II)-αS. The β-sheets are denoted with red (parallel) and black (antiparallel), and helices with grey (3_{10}), blue (α) and purple (π).

Table 4.11: Detailed secondary structure percentages of the three main regions of αS, after aMD simulations using ff03ws.

Region	β character (%)	α character (%)	Other (%)
free-αS			
N-terminal	2.12	16.30	81.58
NAC	6.84	14.56	78.61
C-terminal	0.20	16.98	82.82
Cu(II)-αS			
N-terminal	2.45	9.73	87.82
NAC	5.10	14.73	80.18
C-terminal	0.20	12.65	87.15

Table 4.12: Secondary structure percentages for the free and copper-bound peptide aMD simulations.

Forcefield	β character (%)	α character (%)	Other (%)
free-αS	2.47	16.55	80.98
Cu(II)-αS	2.23	12.43	85.33

A closer evaluation of helical characteristics implicates residues Glu57-Gln62 and Ser87-Gly101 as the regions with most helical population, in both the unbound and Cu(II)-bound peptides. The latter region was found to exhibit the greatest mean α-character occupancy both in the free, at 38.31%, and Cu(II)-αS, at 38.61%, while the former region presented mean occupancies of 34.55% and 33.18%, respectively. Representations of the structure in each of these regions is given in Figure 4.21. A thing to note here is that the region between residues Ser87-Gly101 is split between residues Ala91-Phe94, by coiled structures, resulting in two short helices between residues Ser87-Ala90 and Val95-Leu100. This happens for the majority of the conformations, despite instances of a long continuous amphipathic helix, steric hindrance restricts the conservation of such structures. The observations here are in line with experimental results, suggesting the importance of the central NAC region in fibrillation of αS, but also in the membrane-binding capacities of the protein, with both these regions existing in the SL2 membrane binding domain.[101] Below, Figure 4.20, the Ser87-Gly101 region has been extended to include residues Gly86-Glu105, allowing the construction of a helical wheel, with the sphere sizes proportional to the average helix occupancy; the dotted line above residues A90, G101 and F94 indicates the hydrophobic face of the helix, while the arrow points towards the polar face (residues N103, T92 and Q99) – together resulting in the formation of an amphipathic helix.[c] Three domains between residues Met1-Lys102 have been characterised by their ability to individually bind the protein to lipid vesicles. A significant contribution to this comes from

---

[c]The scale of the image does not allow for any meaningful representation of differences between the metal-free and metal-bound systems, thus only one of the two systems is shown, but can be used to represent both, with a maximum of *ca.* 12% error for the diameter of the residues in free-αS.

the presence of amphipathic helices with Thr residues sitting in the core of the polar faces. Two of those polar faces involve residues Thr59 and Thr92, included in the two helical regions discussed here.[*102*]
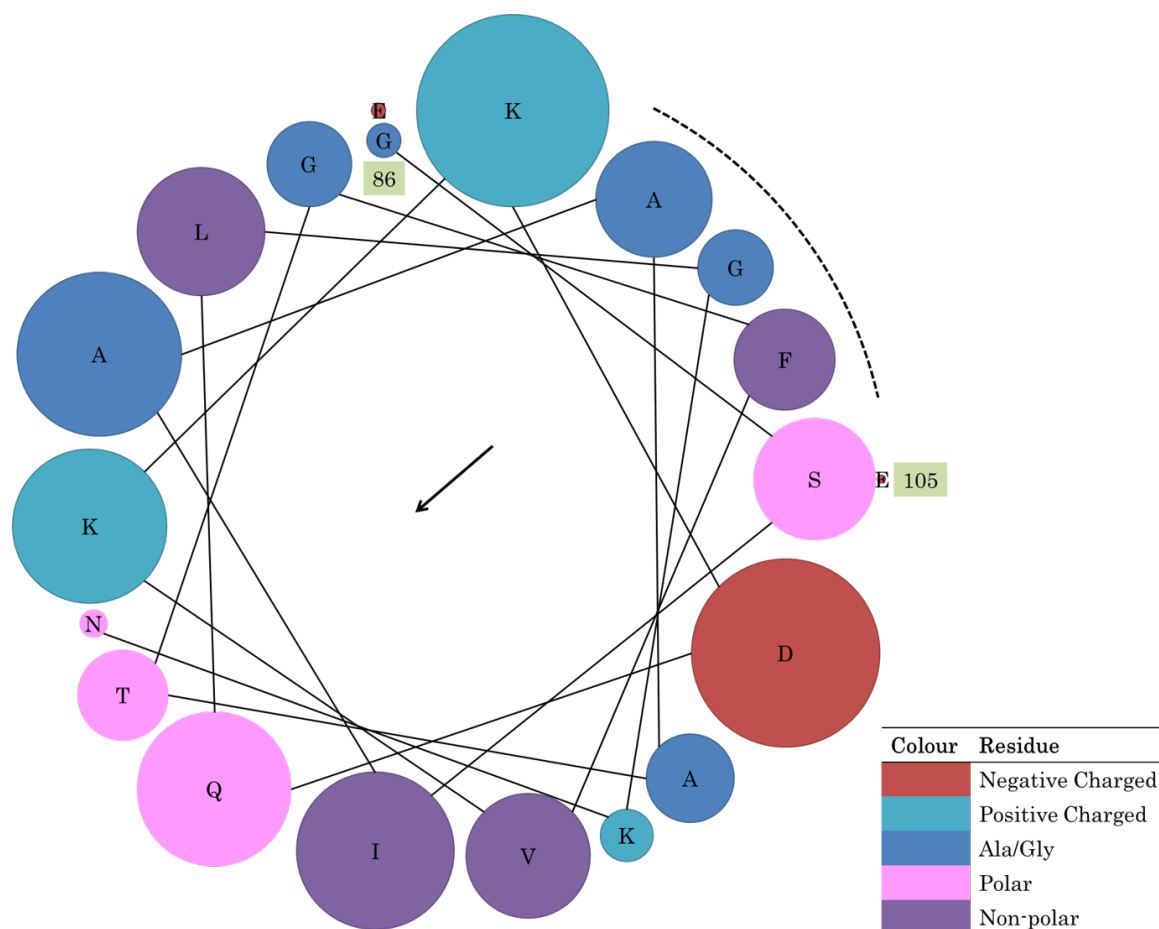
Figure 4.20: Helical wheel of residues Gly86-Glu105, with the sphere sizes proportional to the mean α-helical occupancy in Cu(II)-αS. The dotted line represents the hydrophobic face of the helix, with the arrow pointing towards the polar face.
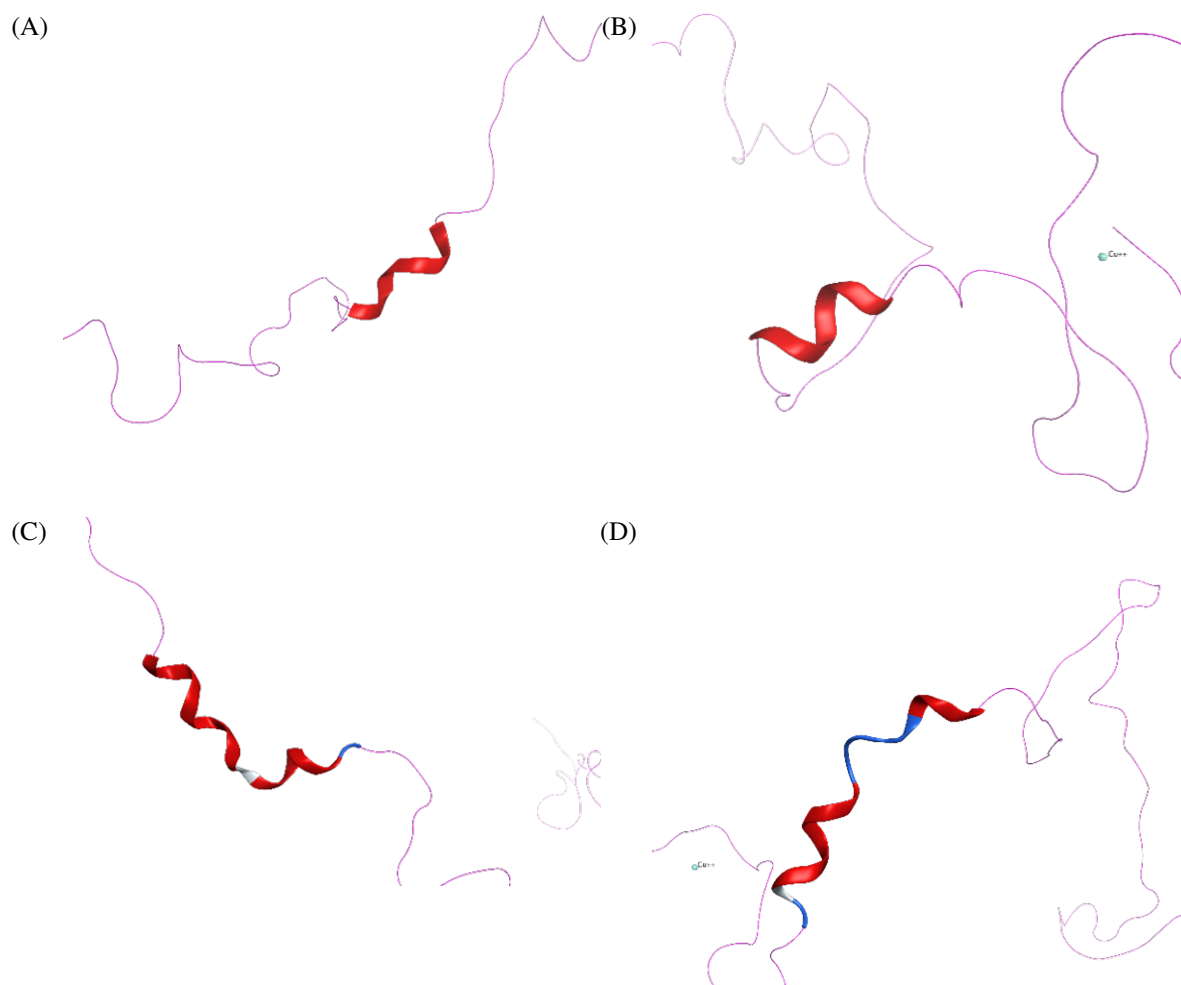
Figure 4.21: Representation of the most populated α-helical regions (red), between residues 57-62 in the (A) unbound and (B) Cu(II)-bound αS and respectively (C-D) between residues 87-101.

The two regions reported above, represent the two most populated helical regions, occurring around at least 1/3rd of the total trajectory, and consist at least of 6 residues. A further four regions where notable helices occur between 4 residues are found in: Ser9-Lys12[d] (free-αS: 35.50%; Cu(II)-αS: 24.13%), Glu20-Lys23 (free-αS: 35.00%; Cu(II)-αS: 22.83%), Gly111-Glu114 (free-αS: 29.36%; Cu(II)-αS: 28.66%), Pro120-Glu123 (free-αS: 35.07%; Cu(II)-αS: 0.48%). Thus, the data suggests an increased presence of helices in the N-terminal, compared to the other regions of the protein, but an overall decrease of these structures, upon coordination of Cu(II).

Below, Figure 4.22, displays the reweighted free energy landscape of R$_g$ vs. two elements of secondary structure characteristics in the two systems. The potential of mean force (PMF) energies display the lowest values at 0% β-character, with *ca.* 16% and 10% α-character, for the free and copper-bound systems, respectively. The free peptide may also be seen sampling a greater conformational space, as explained before, possibly due to the increased flexibility of the system. The local minima with the lowest energy values appear to be between R$_g$ values 35-45 Å in both cases, something to be expected considering the experimental and average R$_g$

---

[d]This region may be further extended to include residues Phe4-Leu8, resulting in an overall helical population of 29.65% for the free-αS and 16.92% for the Cu(II)-αS system.

values, shown in Table 4.9.
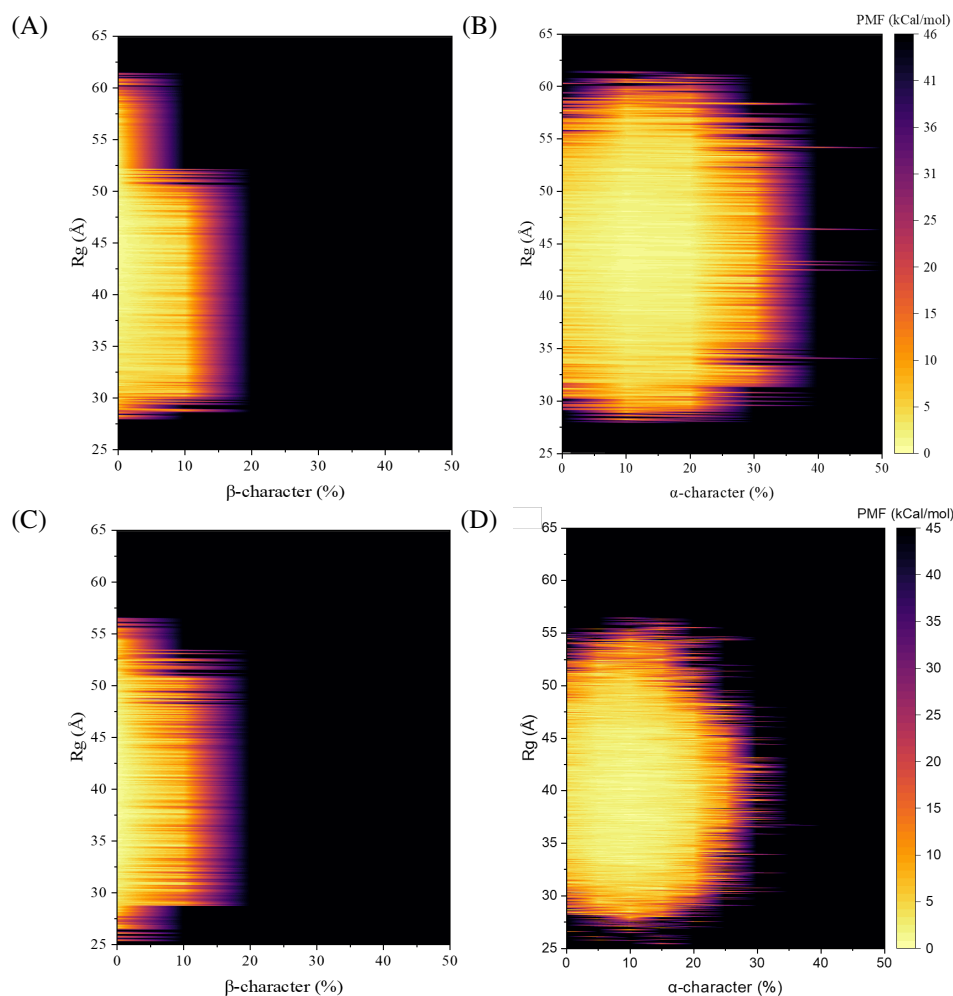
Figure 4.22: Free energy landscape plots of R_g against α- and β-characteristics present in the (A-B) unbound and (C-D) Cu(II)-bound peptide.

The flexibility of each residue is shown by their root mean square fluctuation (RMSF), Figure 4.23. From that, it is evident that adding the metal ions to the peptide restricted the motion of the residues throughout the chain, but especially in the regions around the N-terminal binding site. This appears to be due to the macro-chelation of Cu(II) with M1, D2 and H50. The C-terminal mobility is also greatly affected by Cu(II) coordination, with the RMSF reduced from *ca.* 37 Å in the unbound system, to 27 Å in Cu(II)-αS.
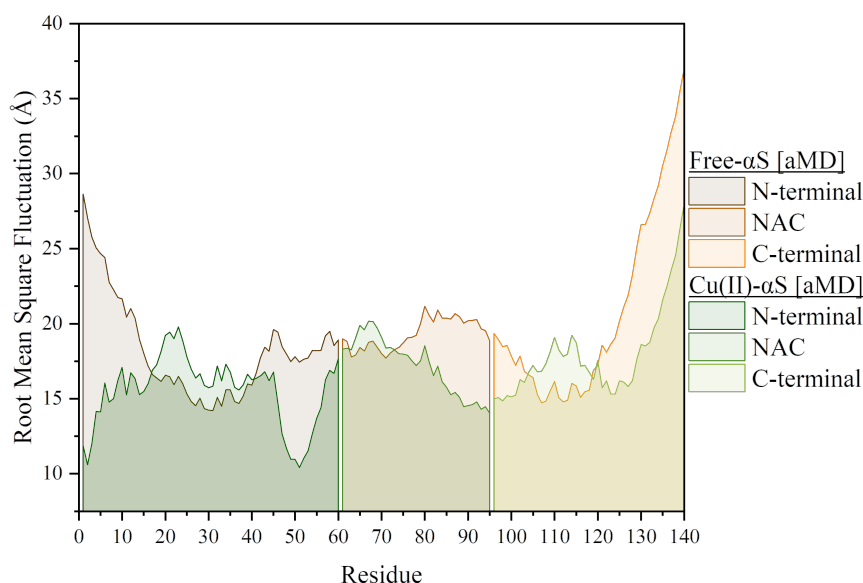
Figure 4.23: Root mean square fluctuation, of the individual residues in each of the systems.

The free energy landscape associated with a combination of R_g and end-to-end distance, Figure 4.24, illustrates the smaller size found in the Cu(II)-αS, when compared to the unbound peptide, with the former exploring conformations where the end-to-end distance was maintained below 165 Å, with an average of 97.5±20.6 Å, as opposed to the free peptide, which went as high as 210 Å, averaging 117.1±27.2 Å.
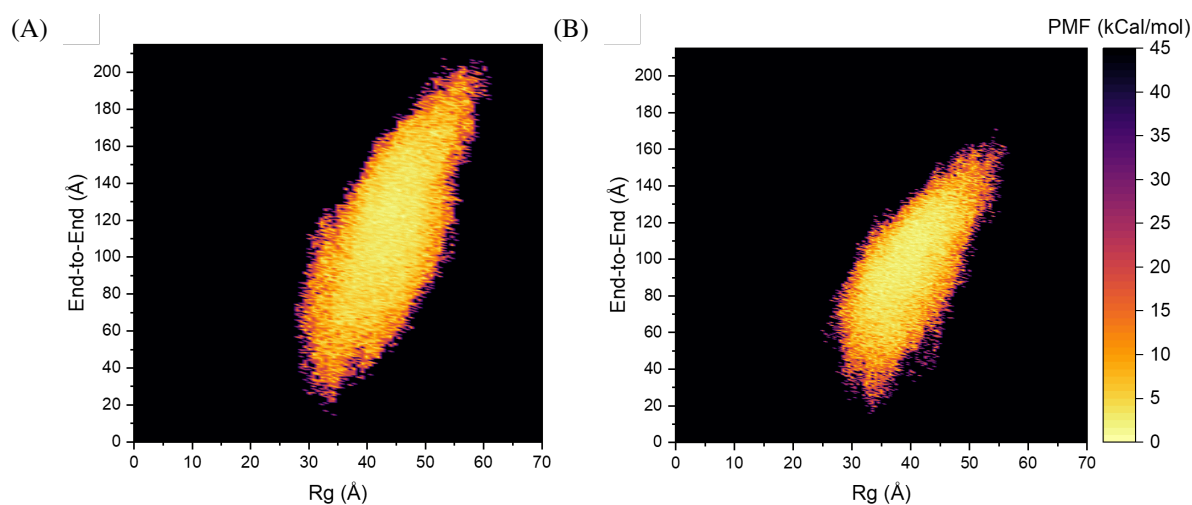
Figure 4.24: 2D free energy plot (FEP) of the end-to-end distance against the radius of gyration (R_g), from the aMD simulations on the (A) free and (B) copper-bound peptide.

In the intramolecular interactions of the two systems, binding of two Cu(II) ions does not strongly alter the pattern of salt bridges within the peptide, Figure 4.25. The sole exception is Asp2, which is bound directly to copper and so is not available for interaction with Lys6. This is true even for the repeat sequences where hairpins were observed, highlighting the transience of these elements of secondary structure. Hydrogen bonds are more strongly affected by metal ions (Figure 4.27), especially in the N-terminal region, where a significant number of "off-diagonal" H-bonds are found in the Cu(II)-bound form. This again appears to be related to the macro-chelation of Cu(II), which brings His50 and neighbouring residues into close contact

with N-terminal ones, closely interacting with residues Glu35 and Glu13. As a result of these contacts, neighbouring residues to His50 are also seen forming H-bonds with more distanced ones, such as Val49-Glu35 and Met1-Glu35. The most significant hydrogen bond formed in the copper-bound system, is seen intra-residue in Glu123 (at *ca.* 83% of frames). This is also evident in maps of close contacts between residues (Figure 4.26): the free peptide only shows contact close to the diagonal, but the Cu(II)-bound peptide has close contact throughout the N-terminal region, extending as far as residue 70, *i.e.* into the NAC. It should be noted, however, that the intrinsically disordered nature of α-Synuclein means that occupancy of all salt-bridge and hydrogen bond contacts is low, typically under 10% of the overall trajectory.
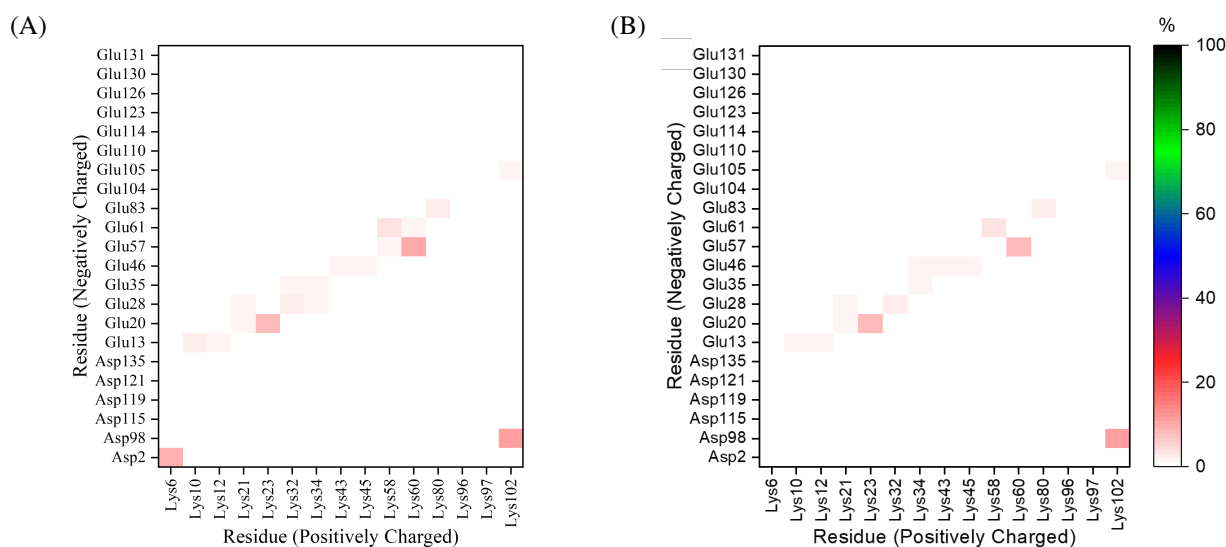
Figure 4.25: Salt bridges formed between negatively and positively charged residues from accelerated MD simulations on (A) free and (B) copper-bound α-Synuclein.
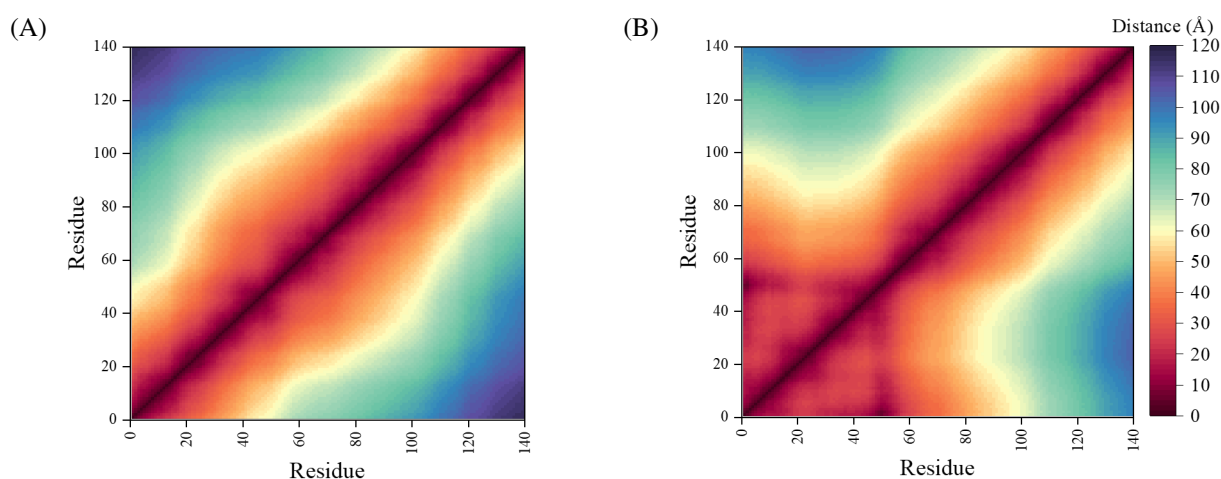
Figure 4.26: Contact maps of the distance between the backbone-C from the dynamics of the (A) free and (B) Cu(II)-bound peptide using aMD simulations.
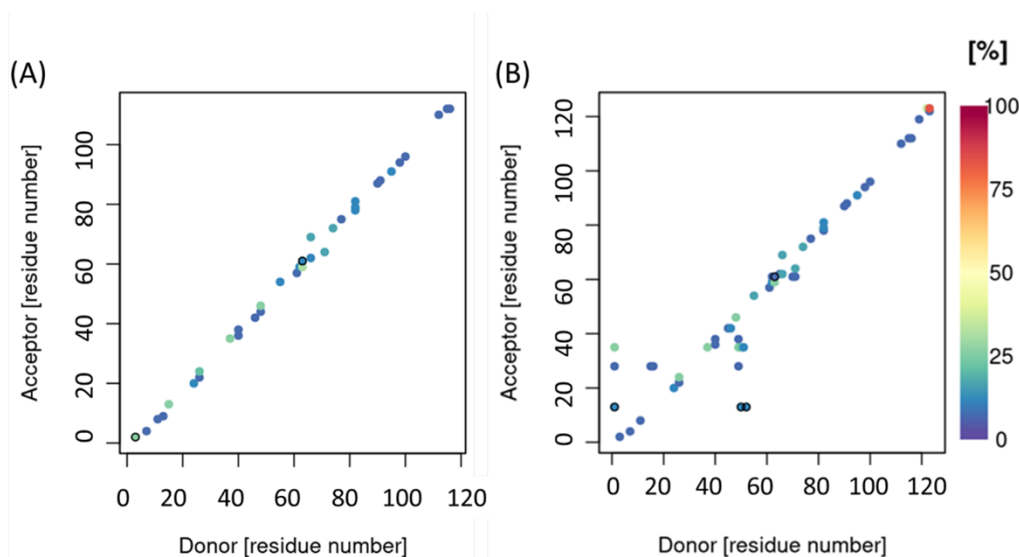
Figure 4.27: Hydrogen bonds present after aMD simulations of the (A) free and (B) copper-bound peptide.

Research on the behaviour of α-Synuclein in its monomeric form has shown that it adopts no lasting secondary characteristics, instead they are rather transient, owing to its natively disordered nature, favoring unfolded and extended conformations.[61, 103] The role of copper ions in the aggregation propensity of α-Synuclein has been studied extensively in the scientific community, with evidence suggesting that it induces aggregation of the peptide.[104] This could be attributed, in part, to restriction of the peptide's structure, allowing it to maintain the folded conformation when interacting with the metal ions, possibly owing to the molecular crowding introduced.[105] Evidence for this may be seen in the increased contacts between the residues, Figure 4.26, presenting a more populated region where the Cu(II)-interactions occur. The R_g data from the accelerated MD simulations, provide further evidence of the crowding introduced into the system on binding of two Cu(II) ions, showing a decrease in the R_g of 4.5 Å. The secondary structure characteristics exhibit a decrease in defined characteristics going from the unbound peptide to Cu(II)-αS, show that the compactness gain in the bound peptide is a direct result of the binding to the metal ions, increasing the intramolecular interactions, as evident by the increased presence of hydrogen bonds, Figure 4.27.

An evaluation of the Cu(II)-bound system, where only the N-terminal binding mode was considered, is given in Appendix B. The findings did not present any significant differences from the results presented here, including the C-terminal binding site. The most significant similarity between the two is in the time the β-hairpin, between residues L38-A53, is maintained for. In particular, the 74 ns expressed in one of the three trajectories, when Cu(II) was interacting with both sites, was increased to 90 and 93 ns in two of the trajectories when only the N-terminal site was considered, Figure B.12. The α-helical character in the N-terminal region (11.03%), Table B.14, was within good agreement to that observed in the system with both binding sites (9.70%), suggesting that the two systems have similar binding affinity to lipid membranes. The greatest difference between the single and double Cu(II)-coordinated systems, appears to be the R_g of the peptide, which is now closer to the average value observed in the unbound system (44.26 Å), at 43.68±5.73 Å, suggesting the macro-chelation cannot be the sole binding site responsible for the compactness observed above.

4.4   Accelerated MD Simulations of Cu^{I}-αS

The results from the aMD simulations of the Cu(I)-αS system are discussed here. Based on the secondary structure % displayed by this system, Table 4.13 and Figure 4.28, it could be classified as an intermediate step between the unbound and Cu(II)-bound forms, seeing as how they fit right in the middle of the folding characteristics displayed in those systems. Upon closer inspection of the per-residue average % of these structures, Figure 4.29 and Figure 4.32, the most significant change in the characteristics comes between residues 110-130, where a great increase in the β-sheets is observed, also hinting towards a well formed β-sheet structure in that region that exists for *ca.* 30% of the total simulation time. Looking at the percentage distribution of the secondary structure, Table 4.13, the % of the β-structures in the C-terminal, is indeed at odds with the other two systems examined, Table 4.11, which only displayed a mean 0.20% β-characteristics, in both cases. This is in line with experimental observations, suggesting a generally unstructured C-terminus,[106, 107] it does, however, raise questions about the possible functional effects of Cu(I)-binding, considering the added folding characteristics now present. This β-sheet region forms from the residues surrounding the metal ion binding site (M116 and M127), hinting towards a facilitation of this secondary structure formation by the metal ion, bridging the strands on the two sites, Figure 4.30. The increased β-characteristics in this region, appear to mask the loss of the β-hairpin that was present in the NAC domain, between residues L38-A53, now only appearing very transiently throughout the trajectory, Figure 4.31.
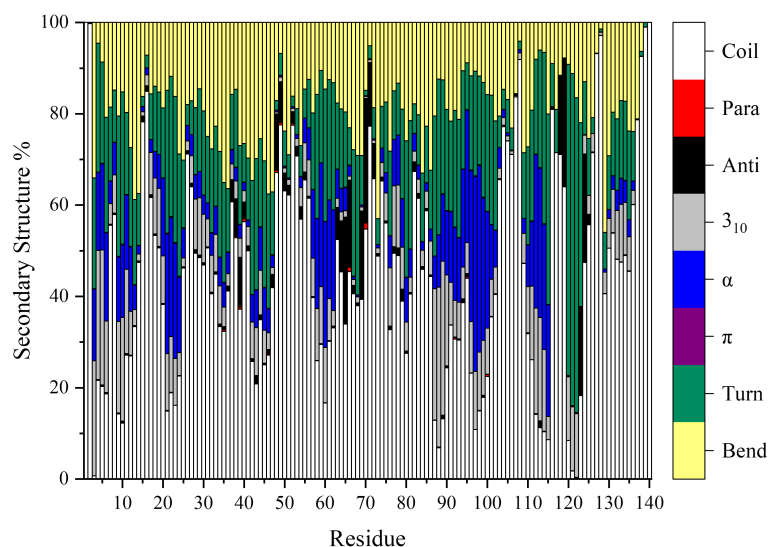
Figure 4.28: Secondary structure distribution per residue from the Cu(I)-αS aMD simulations. The β-sheets are denoted with red (parallel) and black (antiparallel), and helices with grey (3$_{10}$), blue (α) and purple (π).

Table 4.13: Detailed secondary structure percentages of the three main regions of Cu(I)-αS.

Region	β character (%)	α character (%)	Other (%)
N-terminal	1.53	16.50	81.97
NAC	4.35	15.88	79.77
C-terminal	2.15	15.08	82.78
Total	**2.43**	**15.89**	**81.68**
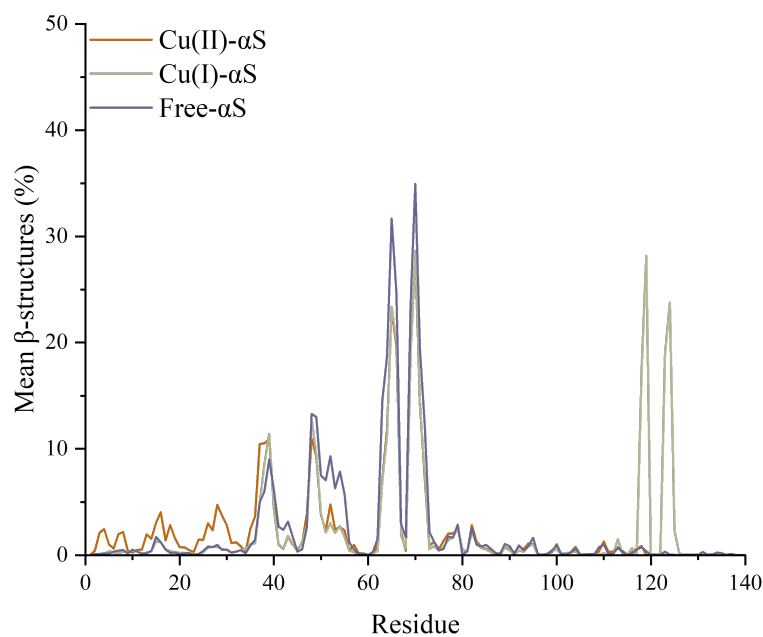
Figure 4.29: Average β-characteristics per residue, from the Cu(II) and Cu(I) bound systems.
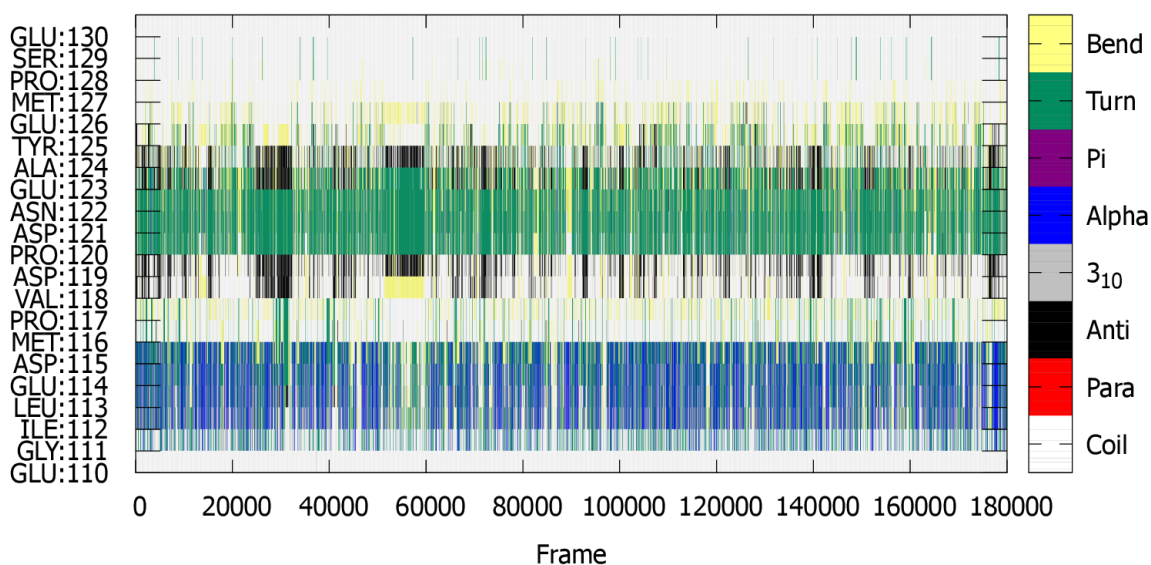
Figure 4.30: Evolution of secondary structure characteristics between residues E110-E130 in the Cu(I)-bound peptide.
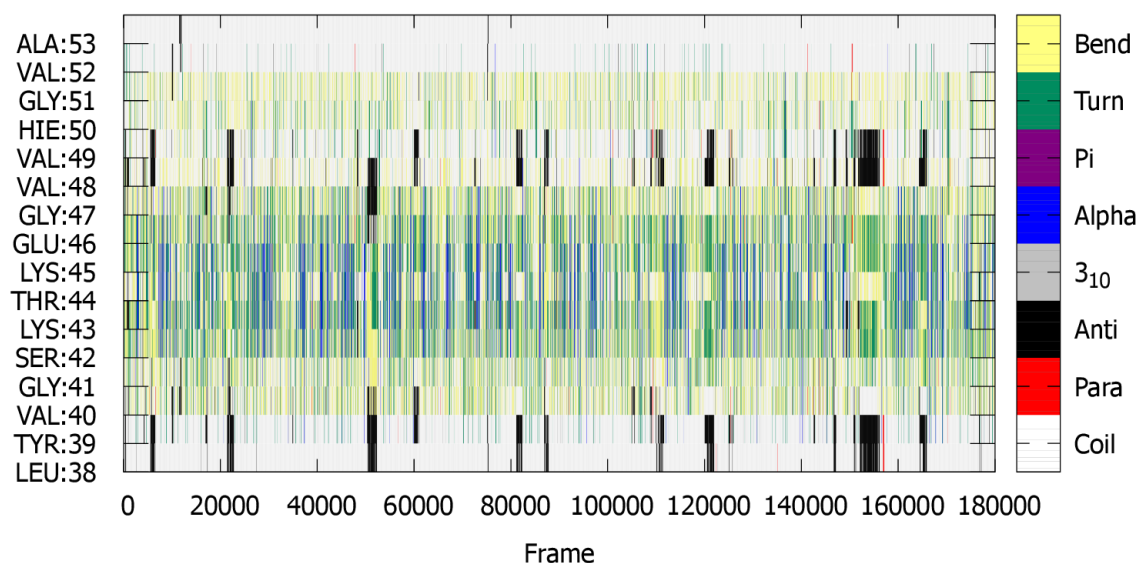
Figure 4.31: Evolution of secondary structure characteristics between residues L38-A53 in the Cu(I)-bound peptide.

Overall, the secondary characteristics in the Cu(I)-system appear to follow the pattern seen in the unbound one, except for the C-terminal strand population, which was near null in the Cu(II)-bound and unbound systems. This facilitation of folding from the coordination of Cu(I) in the C-terminal, is justified by the formation of a bridge between two β-strands, through the coordination of the copper ion, stabilising a well formed β-sheet structure. It would thus appear the two Cu(I), in both binding pockets, are strategically interacting so as to permit the formation of a stable α-helix and β-sheet, in the N- and C-termini, respectively. The C-terminal exhibits a more stable helical structure right before the methionine involved in the binding of the copper ion in that region (i.e. between residues 110-115). Obviously, the presence of a stable helical region between residues 116-127, would be highly unlikely, considering the already established β-sheet that forms in that region. Considering the relatively flexible environment surrounding the metal ion binding sites for Cu(I), where coordination happens linearly with the residues in the peptide, and without the macro-chelation in the N-terminal, as in the Cu(II)-bound peptide, the changes in the structure are rationalised by the reconditioned metal-ion binding sites.
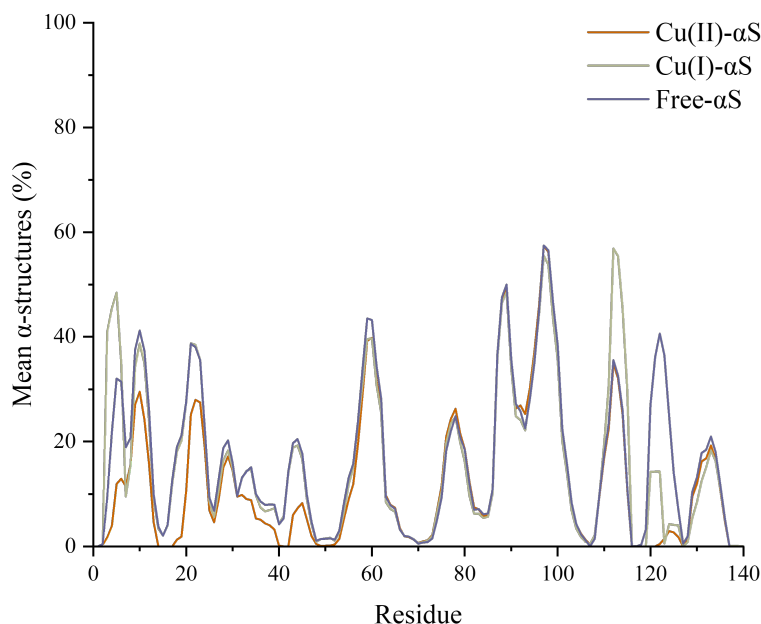
Figure 4.32: Average α-characteristics per residue, from the Cu(II) and Cu(I) bound systems.

Table 4.14 shows the R_g values for the unbound and copper-bound systems. The compactness exhibited in the Cu(I)-system appear to closely resemble that found in the unbound system, averaging at 2 Å above the free-αS. Of note is the significantly higher SD value, exhibiting a >2.5 Å increase from the other two systems. This is very pronounced in the plots of R_g distribution, Figures 4.33 – 4.34, where in both cases the Cu(I)-bound system is clearly sampling a much wider conformational space than the other two systems. Figure 4.33 elucidates a system with no clear local minimum, but with an overall highest peak population around the mean. Figure B.10 shows a similar distribution of R_g for the three runs on Cu(I)-αS, suggesting this may be something inherent to the system, rather than a sampling issue.

Table 4.14: R_g of the copper-bound systems, from aMD simulations.

	Mean R_g (Å)	SD (Å)	Max (Å)	Min (Å)
free-αS (Experiment)	40 ± 1[7] 35.5 ± 0.5[72]			
free-αS	44.26	4.58	61.50	28.07
Cu(II)-αS	39.81	4.31	56.53	25.48
Cu(I)-αS	46.26	7.16	68.19	28.35
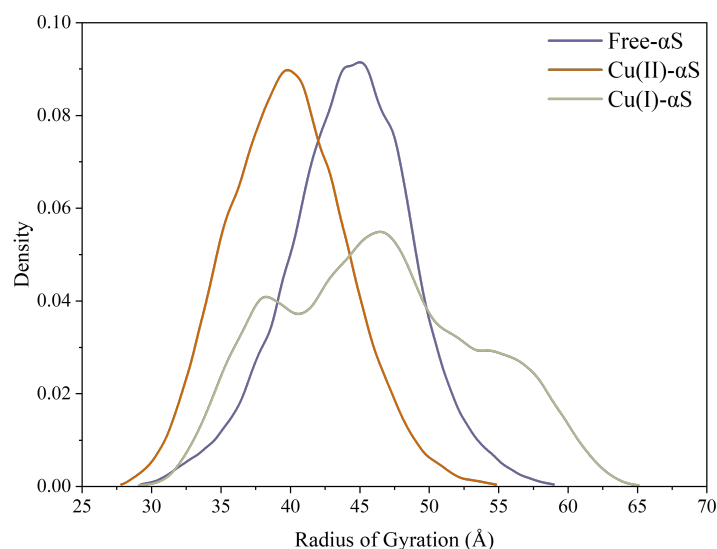
Figure 4.33: Distribution of R_g from the unbound, and Cu(II) and Cu(I) bound systems.
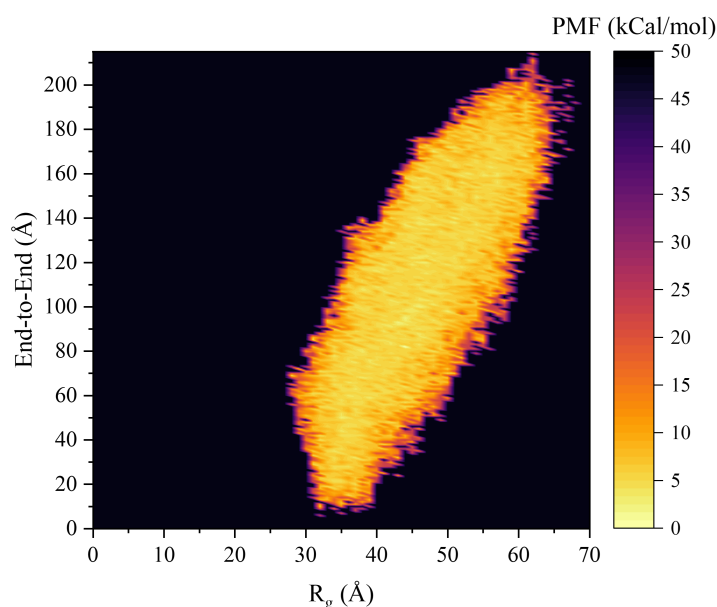
Figure 4.34: Free energy landscape plot of the end-to-end distance against the R_g from the aMD simulations of the Cu(I)-αS system.

Looking at the intramolecular interactions in the system, the off-diagonal hydrogen bonds in the N-terminal of the Cu(II)-bound system, were not replicated in the Cu(I)-case, owing to the loss of macro-chelation in that region. Overall, the hydrogen bonding in this system appears to be in accordance with the one seen in the metal-free peptide, where most of the interactions occupy the diagonal. The most significant of these interactions, that are shared between all three systems, along with the percentage of frames out of the whole trajectory these are expressed, are: Thr59-Val63 (free: 30%; Cu(II): 28%; Cu(I): 29%), Glu46-Val48 (free: 26%; Cu(II): 27%; Cu(I): 26%), Glu35-Val37 (free: 26%; Cu(II): 26%; Cu(I): 26%), and Gln24-Val26 (free: 22%; Cu(II): 24%; Cu(I): 22%). From these percentages, a pattern emerges, where the Cu(I)-bound system is more similar to the free-αS than the Cu(II)-bound system. This is in accordance with the less restricted nature of Cu(I)-αS, resembling the structure observed in the metal-free αS.
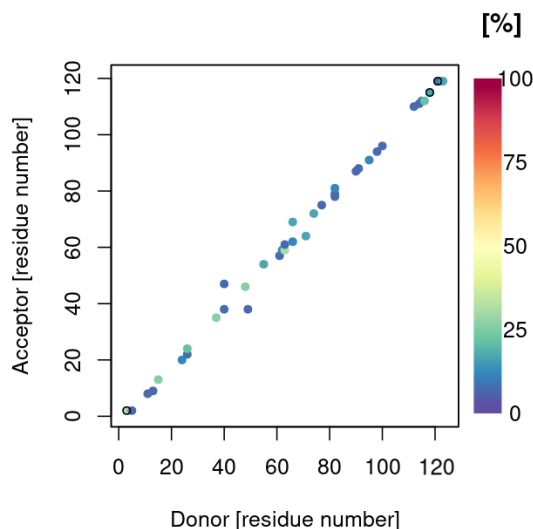
Figure 4.35: Hydrogen bonds present after aMD simulations on Cu(I)-αS.

4.5 Markov State Models of α-Synuclein

Markov state models (MSM) aid in the analysis of dynamical systems through discretisation of the state space, and the construction of a transition matrix, which is used to estimate the probability of a transition between states. The kinetics of state transitions have been analysed to assess the simulated time required for the conformational transitions to take place.

4.5.1 Analysis Setup

The MSM analysis was performed using the PyEMMA package.[108] Initially, the R_g was added as the collective variable by which the trajectories would be compartmentalised, through the MDTraj package.[109] Time-lagged independent component analysis (TICA) was then performed on the aMD trajectories, analysed in the previous sections. This works in a similar way to principal component analysis (PCA), but is more robust to noise and is able to capture slow processes in molecular systems.[110] The TICA projection was then used to cluster the trajectories into 5 states using the kmeans algorithm. The implied timescales (its) were subsequently calculated for a lag time of up to 100 ns. From this data, the lag time for the MSM was chosen to be 70 ns. Having performed Chapman-Kolmogorov (CK) tests under the set conditions for 3, 4 and 6 states, it was determined that 3 were enough for transitions to take place between the states, seeing how the model-estimated lag time fitted well within the 95% uncertainty, Figure 4.36. The convergence of the transition probabilities (not considering the self-transitions) to estimated values, within an *i+1* relationship, are at the very least proof that the systems studied here are Markovian.

Following the construction of the MSM, the Robust Perron Cluster Analysis (PCCA+) method was used, to cluster the states based on the eigenvectors of the transition matrix.[111] These were then used to class the states into a three-state Hidden Markov Model (HMM), by coarse graining the MSM. The use of HMMs allowed for a more confident derivation of the kinetic models. The HMM was then used to calculate the stationary distribution of the different states and the flux between them (Table 4.15). The transition times, derived from the flux and distri-

bution of the macrostates, were calculated using mean first passage times (MFPT) analysis – described in more detail in Chapter 2.
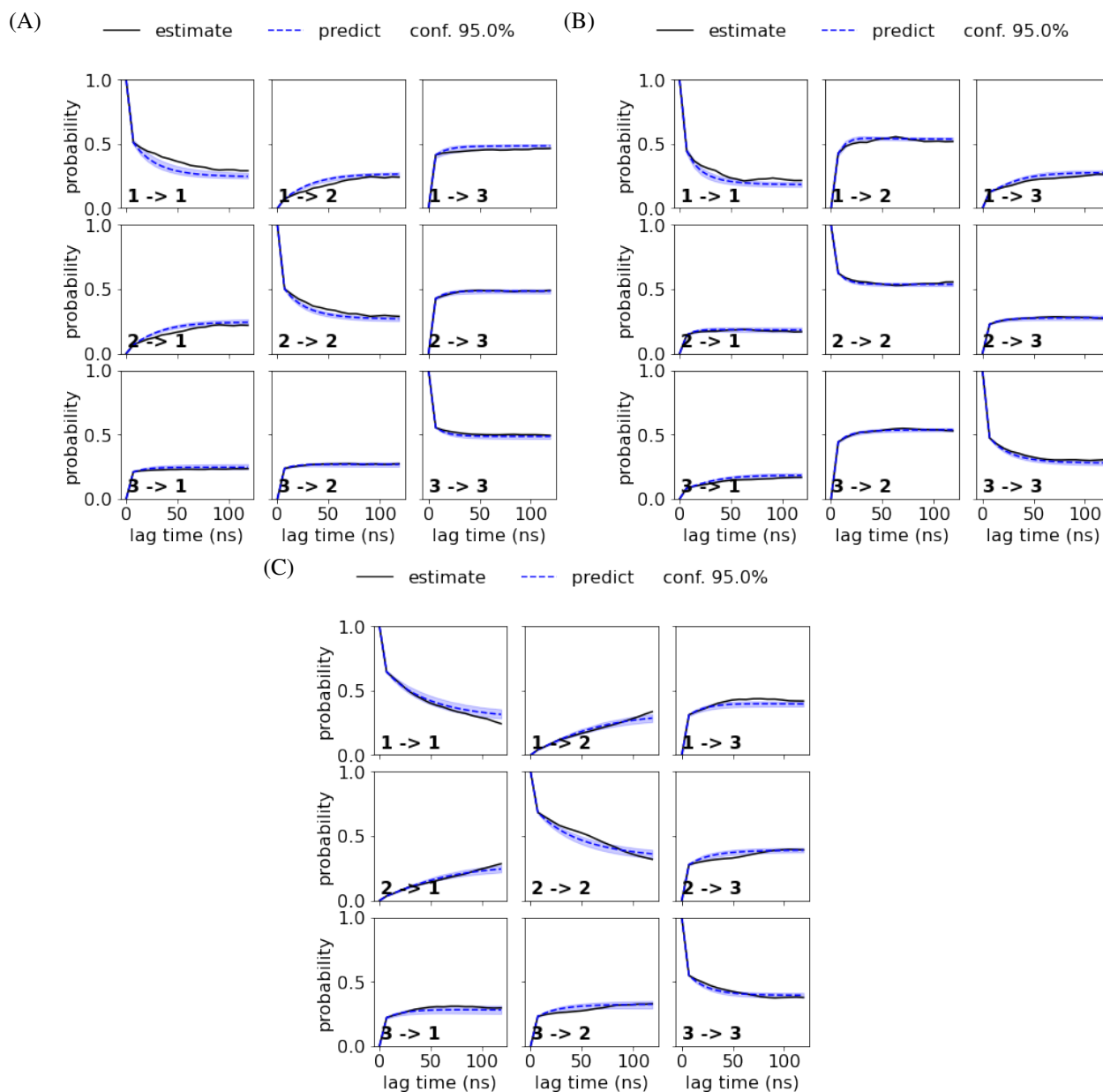
Figure 4.36: CK test of the 3 states for the (A) unbound, (B) Cu(II)-bound, and (C) Cu(I)-bound systems.

4.5.2 Findings

Considering the large structure of αS, the macrostates created here cannot be reduced to single structural elements. Hence, the decision was made to use R_g as a means to classify the key bulk transitions that take place in each of the systems, but also to help in assessing the kinetic adjustments that occur. Table 4.15 shows the key characteristics in each of these macrostates in the three systems examined above, with a pictorial representation of the transition times between the three macrostates given in Figure B.11. The population in each of these states was based on the stationary distribution of the macrostates, while the flux calculated from the

transition matrix was implemented to the calculation of the transition times. The mean R_g and secondary structure propensities in each of the states, were calculated from 20 representative conformations extracted from each of the PCCA+ distributions.

Table 4.15: Characteristics of the ensemble created for each of the macrostates for the unbound, Cu(II)-bound, and Cu(I)-bound systems.

MSM Macrostate	Population (%)	R_g (Å)	α-helices (%)	β-sheets (%)	Flux[e]	Transition time (ns)
Free-αS						
1 (C)	29.34	40.69	14.52	2.96	C → B: 0.114	C → B: 0.331
2 (A)	22.36	48.00	16.50	2.14	A → B: 0.166	A → B: 0.319
3 (B)	48.29	43.69	15.17	1.97	B → C: 0.069	B → C: 0.500
					B → A: 0.077	B → A: 0.441
Cu(II)-αS						
1 (A)	25.85	45.13	11.39	2.24	A → B: 0.171	A → B: 0.307
2 (B)	43.78	40.16	10.71	2.86	B → A: 0.101	B → A: 0.492
					B → C: 0.125	B → C: 0.390
3 (C)	30.37	37.26	9.86	2.79	C → B: 0.180	C → B: 0.342
Cu(I)-αS						
1 (A)	28.27	54.37	15.54	1.29	A → B: 0.118	A → B: 0.539
2 (C)	28.78	38.80	14.12	3.61	C → B: 0.073	C → B: 0.645
3 (B)	42.95	47.98	14.56	1.77	B → A: 0.078	B → A: 0.786
					B → C: 0.049	B → C: 0.664

In the three scenarios examined, all systems transition through an intermediate state before reaching either the higher or lower R_g states. For the better assessment of the variations in the states between the three systems, these have been renumbered to reflect the high (A), mid (B) and low (C) R_g states. The changes in the peptide compaction of the systems is given in Table B.13. The close difference in the R_g of the macrostates in the Cu(II)-bound and metal-free systems is expected, considering the similarity in standard deviations, and by extension the maximum and minimum values of the conformations explored in the phase space, compared to the Cu(I)-αS system, Table 4.14. The two copper-bound systems show identical trends in the transition times between the macrostates of large (A) and medium (B) R_g – with the fastest transition happening from A to B and the slowest from B to A, thus kinetically these two systems exhibit a tendency to adopt a relatively folded state. In the case of the unbound system, however, the slower transition appears to be from the medium (B) to the low (C) R_g states, indicating that the lack of a metal ion increases the strain of reaching more compact states, and by extension the tendency of the unbound system towards a partially unfolded state. In fact,

---

[e]Probability of a transition from one macrostate to another.

it would appear all three systems are more inclined to the intermediate helical folding state, exhibiting the greatest transition times when trying to escape from that state. The transition times seen in the Cu(I)-bound peptide are the highest of the three, suggesting that the system may exhibit the least flexibility, although looking at the free energy landscape plots above, the conformational space explored by the Cu(I)-αS is considerably more expansive than the other two. The more likely reason for the increased transition times, is the difference in R_g between the macrostates, going from *ca.* 5 and 3 Å in the free and Cu(II)-systems, to *ca.* 6 and 9 Å for Cu(I) [for high to medium, and low to medium R_g], Table B.13. It should be noted that while the timescale differences between the three macrostates, appear to be relatively similar, given the results here are from trajectories of accelerated MD, the transition times without the added energy bias are expected to be longer. Furthermore, considering the largely unstructured nature of αS, we expect to see fast transitions in the size of the peptide, in parallel with the <1 ns times seen in Table 4.15. Nevertheless, these can still be used to compare the kinetic paths between the three systems under the same conditions. The use of MSM as a complementary tool to enhanced sampling techniques has been documented in the literature as a way to reconstruct the kinetics of the systems, while maintaining the improved sampling of the conformational space.[*112*]

4.6   Conclusions

Having examined the different force field – solvent combinations in the previous chapter on Amyloid-β, here we sought to apply those findings to simulate α-Synuclein; a protein related to the onset and development of PD. The ff14SB/GBSA combination, which was found to work best from the implicit solvent combinations tested for Aβ, proved to be less capable in describing the characteristics of αS, reported from experiments.[*62, 63, 65*] Considering the intrinsically disordered nature of the protein, the ff03ws/OBC combination was tested and proved to better represent the metal-free system, with regards to the secondary characteristics, R_g and chemical shifts. This parameterisation protocol was therefore applied to study the changes appearing as a result of the peptide's complexation with Cu(II) and Cu(I). These metal ions were selected after considering the reported increased aggregation propensity of αS, as a response to Cu(II), as well as the expected reduction to Cu(I), following the release of reactive oxygen species from fenton-like reactions.[*113*]

The simulation results on the systems, presented a decreased α-helical content in the N-terminal of the Cu(II)-bound peptide, suggesting an increase in the aggregation capacities of the protein, as a result of a predicted reduction in the membrane-binding affinity of the peptide increasing the availability of monomers for oligomerisation. This is in accordance with experimental evidence, from mutation studies on the system, which showed a correlation between a reduction of helical characteristics, membrane affinity and aggregation propensity, discussed in detail in Section 1.4.4.[*114–118*] Stabilised β-character regions were also found to be maintained for longer periods, allowing the peptide to more readily associate with other chains. The Cu(I)-αS system, on the other hand, showed parallels with the unbound peptide, with secondary characteristics and R_g in close agreement to the values observed there. This could be indicative of a system less prone to aggregation, although the transient nature of Cu(I)-binding, and the lack of experimental evidence on the system, make it difficult to draw any conclusions.

Markov state models were created for the evaluation of transitions between macrostates from

clusters, created using R_g as the collective variable. The transition times suggested the fastest progression in the systems to occur going from large to mid-sized R_g values. Looking at the slowest transitions, the copper-bound systems took the longest to move from the medium to large R_g, as opposed from medium to small-R_g, in the metal-free case. This suggests the increased flexibility in the unbound system, from the absence of a metal ion, makes it more taxing to reach the lower-sized states. It should also be noted that the small to large-R_g transitions are prohibited, instead proceeding through the intermediate mid-sized macrostate.

Bibliography

(*1*) Uéda, K. et al. (1993). Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc. Natl. Acad. Sci. USA 90*, 11282–11286.

(*2*) Uversky, V. N., Li, J., and Fink, A. L. (2001). Metal-triggered structural transformations, aggregation, and fibrillation of human α-synuclein: A possible molecular link between parkinson's disease and heavy metal exposure. *Journal of Biological Chemistry 276*, 44284–44296.

(*3*) Bisi, N. et al. (2021). α-Synuclein: An All-Inclusive Trip Around its Structure, Influencing Factors and Applied Techniques. *Frontiers in Chemistry 9*, 1–28.

(*4*) Ulmer, T. S., Bax, A., Cole, N. B., and Nussbaum, R. L. (2005). Structure and dynamics of micelle-bound human α-synuclein. *Journal of Biological Chemistry 280*, 9595–9603.

(*5*) Deeth, R. J., Fey, N., and Williams-Hubbard, B. (2005). DommiMOE: An implementation of ligand field molecular mechanics in the molecular operating environment. *Journal of Computational Chemistry 26*, 123–130.

(*6*) Rekas, A. et al. (2010). The structure of dopamine induced α-synuclein oligomers. *European Biophysics Journal 39*, 1407–1419.

(*7*) Uversky, V. N., Li, J., and Fink, A. L. (2001). Evidence for a Partially Folded Intermediate in α-Synuclein Fibril Formation. *Journal of Biological Chemistry 276*, 10737–10744.

(*8*) Ulmer, T., Bax, A., Cole, N., and Nussbaum, R. Human micelle-bound alpha-synuclein (PDB: 1XQ8), 2005.

(*9*) Kowalik-Jankowska, T., Ruta-Dolejsz, M., Wiśniewska, K., and Lankiewicz, L. (2001). Cu(II) interaction with N-terminal fragments of human and mouse β-amyloid peptide. *Journal of Inorganic Biochemistry 86*, 535–545.

(*10*) Kowalik-Jankowska, T., Ruta, M., Wiśniewska, K., and Łankiewicz, L. (2003). Coordination abilities of the 1-16 and 1-28 fragments of β-amyloid peptide towards copper(II) ions: A combined potentiometric and spectroscopic study. *Journal of Inorganic Biochemistry 95*, 270–282.

(*11*) Lucas, H. R., and Lee, J. C. (2011). Copper(ii) enhances membrane-bound α-synuclein helix formation. *Metallomics 3*, 280.

(*12*) Guilloreau, L. et al. (2006). Structural and thermodynamical properties of CuII amyloid-β16/28 complexes associated with Alzheimer's disease. *Journal of Biological Inorganic Chemistry 11*, 1024–1038.

(*13*) Gentile, I. et al. (2018). Interaction of Cu(i) with the Met-X3-Met motif of alpha-synuclein: binding ligands, affinity and structural features. *Metallomics 10*, 1383–1389.

(*14*)   Li, P., and Merz, K. M. (2016). MCPB.py: A Python Based Metal Center Parameter Builder. *Journal of Chemical Information and Modeling 56*, 599–604.

(*15*)   Al-Shammari, N., Savva, L., Kennedy-Britten, O., and Platts, J. A. (2021). Force-field evaluation and accelerated molecular dynamics simulation of Zn(II) binding to N-terminus of amyloid-β. *Computational Biology and Chemistry 93*, 107540.

(*16*)   Shrestha, U. R. et al. (2019). Generation of the configurational ensemble of an intrinsically disordered protein from unbiased molecular dynamics simulation. *Proceedings of the National Academy of Sciences of the United States of America 116*, 20446–20452.

(*17*)   Zheng, W. et al. (2016). Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment. *Journal of the American Chemical Society 138*, 11702–11713.

(*18*)   Bellaiche, M. M., and Best, R. B. (2018). Molecular Determinants of Aβ42 Adsorption to Amyloid Fibril Surfaces. *Journal of Physical Chemistry Letters 9*, 6437–6443.

(*19*)   Duan, Y. et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry 24*, 1999–2012.

(*20*)   Voelz, V. A., Dill, K. A., and Chorny, I. (2011). Peptoid conformational free energy landscapes from implicit-solvent molecular simulations in AMBER. *Biopolymers 96*, 639–650.

(*21*)   Patel, S., Ramanujam, V., Srivastava, A. K., and Chary, K. V. (2014). Conformational propensities and dynamics of a $\beta\gamma$-crystallin, an intrinsically disordered protein. *Physical Chemistry Chemical Physics 16*, 12703–12718.

(*22*)   Shell, M. S., Ritterson, R., and Dill, K. A. (2008). A test on peptide stability of AMBER force fields with implicit solvation. *Journal of Physical Chemistry B 112*, 6878–6886.

(*23*)   Lei, H., Wu, C., Wang, Z.-X., Zhou, Y., and Duan, Y. (2008). Folding processes of the B domain of protein A to the native state observed in all-atom ab initio folding simulations. *The Journal of Chemical Physics 128*, 235105.

(*24*)   Rungnim, C., Rungrotmongkol, T., Hannongbua, S., and Okumura, H. (2013). Replica exchange molecular dynamics simulation of chitosan for drug delivery system based on carbon nanotube. *Journal of Molecular Graphics and Modelling 39*, 183–192.

(*25*)   Gallego-Villar, L. et al. (2013). Functional characterization of novel genotypes and cellular oxidative stress studies in propionic acidemia. *Journal of Inherited Metabolic Disease 36*, 731–740.

(*26*)   Zerze, G. H., Zheng, W., Best, R. B., and Mittal, J. (2019). Evolution of All-Atom Protein Force Fields to Improve Local and Global Properties. *Journal of Physical Chemistry Letters 10*, 2227–2234.

(*27*)   Case, D. et al. AMBER 2019, tech. rep., San Francisco: University of California, 2019.

(*28*)   Dick, T. J., and Madura, J. D. (2005). Chapter 5 A Review of the TIP4P, TIP4P-Ew, TIP5P, and TIP5P-E Water Models. *Annual Reports in Computational Chemistry 1*, 59–74.

(*29*)   Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters 246*, 122–129.

(*30*) Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996). Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *The Journal of Physical Chemistry 100*, 19824–19839.

(*31*) Onufriev, A., Bashford, D., and Case, D. A. (2000). Modification of the generalized born model suitable for macromolecules. *Journal of Physical Chemistry B 104*, 3712–3720.

(*32*) Savva, L., and Platts, J. A. (2022). Evaluation of implicit solvent models in molecular dynamics simulation of α-Synuclein. *Journal of Biomolecular Structure and Dynamics 41*, 1–16.

(*33*) Frisch, M. J. et al. Gaussian 09, tech. rep., Wallingford, CT: Gaussian, Inc., 2009.

(*34*) Scott, A. P., and Radom, L. (1996). Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Møller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *Journal of Physical Chemistry 100*, 16502–16513.

(*35*) Seminario, J. M. (1996). Calculation of intramolecular force fields from second-derivative tensors. *International Journal of Quantum Chemistry 60*, 1271–1277.

(*36*) Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *Journal of Physical Chemistry 97*, 10269–10280.

(*37*) Besler, B. H., Merz, K. M., and Kollman, P. A. (1990). Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry 11*, 431–439.

(*38*) Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995). Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry 16*, 1357–1377.

(*39*) Izaguirre, J. A., Catarello, D. P., Wozniak, J. M., and Skeel, R. D. (2001). Langevin stabilization of molecular dynamics. *The Journal of Chemical Physics 114*, 2090–2098.

(*40*) Kräutler, V., and Gunsteren, W. F. V. A. N. (2001). A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular. *22*, 501–508.

(*41*) Somavarapu, A. K., and Kepp, K. P. (2015). The Dependence of Amyloid-β Dynamics on Protein Force Fields and Water Models. *ChemPhysChem 16*, 3278–3289.

(*42*) Best, R. B., Zheng, W., and Mittal, J. (2014). Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *Journal of Chemical Theory and Computation 10*, 5113–5124.

(*43*) Coskuner-Weber, O., and Uversky, V. (2018). Insights into the Molecular Mechanisms of Alzheimer's and Parkinson's Diseases with Molecular Simulations: Understanding the Roles of Artificial and Pathological Missense Mutations in Intrinsically Disordered Proteins Related to Pathology. *International Journal of Molecular Sciences 19*, 336.

(*44*) Miao, Y. et al. (2014). Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation 10*, 2677–2689.

(*45*) Glykos, N. M. (2006). Carma: A molecular dynamics analysis program. *Journal of Computational Chemistry 27*, 1765–1768.

(46) Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation 9*, 3084–3095.

(47) Binolfi, A. et al. (2006). 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 128*, 9893–9901.

(48) Rasia, R. M. et al. (2005). Structural characterization of copper(II) binding to α-synuclein: Insights into the bioinorganic chemistry of Parkinson's disease. *Proceedings of the National Academy of Sciences of the United States of America 102*, 4294–4299.

(49) Camponeschi, F. et al. (2013). Copper(I)-α-Synuclein Interaction: Structural Description of Two Independent and Competing Metal Binding Sites. *Inorganic Chemistry 52*, 1358–1367.

(50) Binolfi, A. et al. (2011). Exploring the structural details of Cu(I) binding to scopy. *Journal of the American Chemical Society 133*, 194–196.

(51) De Ricco, R. et al. (2015). Differences in the Binding of Copper(I) to α- and β-Synuclein. *Inorganic Chemistry 54*, 265–272.

(52) De Ricco, R. et al. (2015). Copper(I/II), α/β-Synuclein and Amyloid-β: Menage à Trois? *ChemBioChem 16*, 2319–2328.

(53) Carboni, E., and Lingor, P. (2015). Insights on the interaction of alpha-synuclein and metals in the pathophysiology of Parkinson's disease. *Metallomics 7*, 395–404.

(54) Abeyawardhane, D. L. et al. (2018). Copper Induced Radical Dimerization of α-Synuclein Requires Histidine. *Journal of the American Chemical Society 140*, 17086–17094.

(55) González, N. et al. (2019). Effects of alpha-synuclein post-translational modifications on metal binding. *Journal of Neurochemistry 150*, 507–521.

(56) Maiti, N. C., Apetri, M. M., Zagorski, M. G., Carey, P. R., and Anderson, V. E. (2004). Raman Spectroscopic Characterization of Secondary Structure in Natively Unfolded Proteins: α-Synuclein. *Journal of the American Chemical Society 126*, 2399–2408.

(57) Cai, Y. et al. (2015). Changes in secondary structure of α-synuclein during oligomerization induced by reactive aldehydes. *Biochemical and Biophysical Research Communications 464*, 336–341.

(58) Lin, X.-J. et al. (2006). Secondary structural formation of α-synuclein amyloids as revealed byg-factor of solid-state circular dichroism. *Biopolymers 83*, 226–232.

(59) Burré, J. et al. (2013). Properties of native brain α-synuclein. *Nature 498*, 107–110.

(60) Ulrih, N. P., Barry, C. H., and Fink, A. L. (2008). Impact of Tyr to Ala mutations on α-synuclein fibrillation and structural properties. *Biochimica et Biophysica Acta - Molecular Basis of Disease 1782*, 581–585.

(61) Apetri, M. M., Maiti, N. C., Zagorski, M. G., Carey, P. R., and Anderson, V. E. (2006). Secondary Structure of α-Synuclein Oligomers: Characterization by Raman and Atomic Force Microscopy. *Journal of Molecular Biology 355*, 63–71.

(62) Dai, X. et al. (2021). Optical tweezers-controlled hotspot for sensitive and reproducible surface-enhanced Raman spectroscopy characterization of native protein structures. *Nature Communications 12*, 1–9.

(63)   Silva, B. A., Einarsdóttir, Ó., Fink, A. L., and Uversky, V. N. (2013). Biophysical char-
acterization of α-synuclein and rotenone interaction. *Biomolecules 3*, 703–732.

(64)   Shen, Y., and Bax, A. (2010). SPARTA+: A modest improvement in empirical NMR
chemical shift prediction by means of an artificial neural network. *Journal of Biomolec-
ular NMR 48*, 13–22.

(65)   Bermel, W. et al. (2006). Protonless NMR experiments for sequence-specific assign-
ment of backbone nuclei in unfolded proteins. *Journal of the American Chemical Soci-
ety 128*, 3918–3919.

(66)   Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006). Sensitivity of sec-
ondary structure propensities to sequence differences between α- and γ-synuclein: Im-
plications for fibrillation. *Protein Science 15*, 2795–2804.

(67)   Kim, D. H., Lee, J., Mok, K. H., Lee, J. H., and Han, K. H. (2020). Salient features of
monomeric alpha-synuclein revealed by NMR spectroscopy. *Biomolecules 10*, 1–15.

(68)   Xu, L., Nussinov, R., and Ma, B. (2016). Coupling of the non-amyloid-component
(NAC) domain and the KTK(E/Q)GV repeats stabilize the α-synuclein fibrils. *Euro-
pean Journal of Medicinal Chemistry 121*, 841–850.

(69)   Tuttle, M. D. et al. (2016). Solid-state NMR structure of a pathogenic fibril of full-
length human α-synuclein. *Nature Structural & Molecular Biology 23*, 409–415.

(70)   Mor, D. E., Ugras, S. E., Daniels, M. J., and Ischiropoulos, H. (2016). Dynamic struc-
tural flexibility of α-synuclein. *Neurobiology of Disease 88*, 66–74.

(71)   Bisaglia, M., Tessari, I., Mammi, S., and Bubacco, L. (2009). Interaction between α-
synuclein and metal ions, still looking for a role in the pathogenesis of Parkinson's
disease. *NeuroMolecular Medicine 11*, 239–251.

(72)   Ahmed, M. C. et al. (2021). Refinement of α-Synuclein Ensembles Against SAXS
Data: Comparison of Force Fields and Methods. *Frontiers in Molecular Biosciences 8*,
1–13.

(73)   Chwastyk, M., and Cieplak, M. (2020). Conformational Biases of α-Synuclein and
Formation of Transient Knots. *Journal of Physical Chemistry B 124*, 11–19.

(74)   Duong, V. T., Chen, Z., Thapa, M. T., and Luo, R. (2018). Computational Studies of
Intrinsically Disordered Proteins. *Journal of Physical Chemistry B 122*, 10455–10469.

(75)   Robustelli, P., Piana, S., and Shaw, D. E. (2018). Developing a molecular dynamics
force field for both folded and disordered protein states. *Proceedings of the National
Academy of Sciences of the United States of America 115*, E4758–E4766.

(76)   Rauscher, S. et al. (2015). Structural Ensembles of Intrinsically Disordered Proteins
Depend Strongly on Force Field: A Comparison to Experiment. *Journal of Chemical
Theory and Computation 11*, 5513–5524.

(77)   Hoffmann, K. Q., McGovern, M., Chiu, C. C., and De Pablo, J. J. (2015). Secondary
structure of rat and human amylin across force fields. *PLoS ONE 10*, 1–24.

(78)   Zhou, R., and Berne, B. J. (2002). Can a continuum solvent model reproduce the free
energy landscape of a -hairpin folding in water? *Proceedings of the National Academy
of Sciences 99*, 12777–12782.

(*79*)   Shrestha, U. R., Smith, J. C., and Petridis, L. (2021). Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. *Communications Biology 4*, 1–8.

(*80*)   Bhattacharya, S., Xu, L., and Thompson, D. (2020). Long-range Regulation of Partially Folded Amyloidogenic Peptides. *Scientific Reports 10*, 1–17.

(*81*)   Chan-Yao-Chong, M., Durand, D., and Ha-Duong, T. (2019). Molecular Dynamics Simulations Combined with Nuclear Magnetic Resonance and/or Small-Angle X-ray Scattering Data for Characterizing Intrinsically Disordered Protein Conformational Ensembles. *Journal of Chemical Information and Modeling 59*, 1743–1758.

(*82*)   Binolfi, A., Quintanar, L., Bertoncini, C. W., Griesinger, C., and Fernández, C. O. (2012). Bioinorganic chemistry of copper coordination to alpha-synuclein: Relevance to Parkinson's disease. *Coordination Chemistry Reviews 256*, 2188–2201.

(*83*)   Ramis, R., Ortega-Castro, J., Vilanova, B., Adrover, M., and Frau, J. (2017). Copper(II) Binding Sites in N-Terminally Acetylated α-Synuclein: A Theoretical Rationalization. *Journal of Physical Chemistry A 121*, 5711–5719.

(*84*)   Drew, S. C. et al. (2008). Cu2+ binding modes of recombinant α-synuclein - Insights from EPR spectroscopy. *Journal of the American Chemical Society 130*, 7766–7773.

(*85*)   Kowalik-Jankowska, T., Rajewska, A., Jankowska, E., and Grzonka, Z. (2006). Copper(ii) binding by fragments of α-synuclein containing M 1-D2- and -H50-residues; A combined potentiometric and spectroscopic study. *Dalton Transactions*, 5068–5076.

(*86*)   Yu, H., Han, W., Ma, W., and Schulten, K. (2015). Transient β -hairpin formation in α -synuclein monomer revealed by coarse-grained molecular dynamics simulation. *The Journal of Chemical Physics 143*, 243142.

(*87*)   Gallardo, J., Escalona-Noguero, C., and Sot, B. (2020). Role of α-Synuclein Regions in Nucleation and Elongation of Amyloid Fiber Assembly. *ACS Chemical Neuroscience 11*, 872–879.

(*88*)   Buratti, F. A. et al. (2022). Aromaticity at position 39 in α-synuclein: A modulator of amyloid fibril assembly and membrane-bound conformations. *Protein Science 31*, e4360.

(*89*)   Moons, R. et al. (2020). Metal ions shape α-synuclein. *Scientific Reports 10*, 1–13.

(*90*)   Li, H. T., Du, H. N., Tang, L., Hu, J., and Hu, H. Y. (2002). Structural transformation and aggregation of human α-synuclein in trifluoroethanol: Non-amyloid component sequence is essential and β-sheet formation is prerequisite to aggregation. *Biopolymers 64*, 221–226.

(*91*)   Wang, H. et al. (2022). Cu2+ ions modulate the interaction between α-synuclein and lipid membranes. *Journal of Inorganic Biochemistry 236*, 111945.

(*92*)   Maltsev, A. S., Chen, J., Levine, R. L., and Bax, A. (2013). Site-specific interaction between α-synuclein and membranes probed by NMR-observed methionine oxidation rates. *Journal of the American Chemical Society 135*, 2943–2946.

(*93*)   Miotto, M. C. et al. (2014). Site-specific copper-catalyzed oxidation of α-synuclein: Tightening the link between metal binding and protein oxidative damage in Parkinson's disease. *Inorganic Chemistry 53*, 4350–4358.

(*94*) Horvath, I. et al. (2019). Interaction between Copper Chaperone Atox1 and Parkinson's Disease Protein α-Synuclein Includes Metal-Binding Sites and Occurs in Living Cells. *ACS Chemical Neuroscience 10*, 4659–4668.

(*95*) Jao, C. C., Hegde, B. G., Chenb, J., Haworth, I. S., and Langen, R. (2008). Structure of membrane-bound α-synuclein from site-directed spin labeling and computational refinement. *Proceedings of the National Academy of Sciences of the United States of America 105*, 19666–19671.

(*96*) Runfola, M., De Simone, A., Vendruscolo, M., Dobson, C. M., and Fusco, G. (2020). The N-terminal Acetylation of α-Synuclein Changes the Affinity for Lipid Membranes but not the Structural Properties of the Bound State. *Scientific Reports 10*, 1–10.

(*97*) Pineda, A., and Burre, J. (2017). Modulating membrane binding of α-synuclein as a therapeutic strategy. *Proceedings of the National Academy of Sciences of the United States of America 114*, 1223–1225.

(*98*) Kurochka, A. S., Yushchenko, D. A., Bouř, P., and Shvadchak, V. V. (2021). Influence of lipid membranes on α-synuclein aggregation. *ACS Chemical Neuroscience 12*, 825–830.

(*99*) Lee, H. J., Choi, C., and Lee, S. J. (2002). Membrane-bound α-synuclein has a high aggregation propensity and the ability to seed the aggregation of the cytosolic form. *Journal of Biological Chemistry 277*, 671–678.

(*100*) Dikiy, I., and Eliezer, D. (2014). N-terminal Acetylation stabilizes N-terminal Helicity in Lipid- and Micelle-bound α-Synuclein and increases its affinity for Physiological Membranes. *Journal of Biological Chemistry 289*, 3652–3665.

(*101*) Bodner, C. R., Maltsev, A. S., Dobson, C. M., and Bax, A. (2010). Differential Phospholipid Binding of α-Synuclein Variants Implicated in Parkinson's Disease Revealed by Solution NMR Spectroscopy. *Biochemistry 49*, 862–871.

(*102*) Perrin, R. J., Woods, W. S., Clayton, D. F., and George, J. M. (2000). Interaction of human α-synuclein and Parkinson's disease variants with phospholipids: Structural analysis using site-directed mutagenesis. *Journal of Biological Chemistry 275*, 34393–34398.

(*103*) Eliezer, D., Kutluay, E., Bussell, R., and Browne, G. (2001). Conformational properties of α-synuclein in its free and lipid-associated states. *Journal of Molecular Biology 307*, 1061–1073.

(*104*) Wang, X., Moualla, D., Wright, J. A., and Brown, D. R. (2010). Copper binding regulates intracellular alpha-synuclein localisation, aggregation and toxicity. *Journal of Neurochemistry 113*, 704–714.

(*105*) Uversky, V. N., M. Cooper, E., Bower, K. S., Li, J., and Fink, A. L. (2002). Accelerated α-synuclein fibrillation in crowded milieu. *FEBS Letters 515*, 99–103.

(*106*) Kim, T. D., Paik, S. R., and Yang, C. H. (2002). Structural and functional implications of C-terminal regions of α-synuclein. *Biochemistry 41*, 13782–13790.

(*107*) Farzadfard, A. et al. (2022). The C-terminal tail of α-synuclein protects against aggregate replication but is critical for oligomerization. *Communications Biology 5*, DOI: `10.1038/s42003-022-03059-8`.

(*108*) Scherer, M. K. et al. (2015). PyEMMA 2: A Software Package for Estimation, Valida-
tion, and Analysis of Markov Models. *Journal of Chemical Theory and Computation
11*, 5525–5542.

(*109*) McGibbon, R. T. et al. (2015). MDTraj: A Modern Open Library for the Analysis of
Molecular Dynamics Trajectories. *Biophysical Journal 109*, 1528–1532.

(*110*) Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G., and Noé, F. (2013). Iden-
tification of slow molecular order parameters for Markov model construction. *Journal
of Chemical Physics 139*, DOI: `10.1063/1.4811489`.

(*111*) Röblitz, S., and Weber, M. (2013). Fuzzy spectral clustering by PCCA+: Application
to Markov state models and data classification. *Advances in Data Analysis and Classi-
fication 7*, 147–179.

(*112*) Biswas, M., Lickert, B., and Stock, G. (2018). Metadynamics Enhanced Markov Mod-
eling of Protein Dynamics. *Journal of Physical Chemistry B 122*, 5508–5514.

(*113*) Bacchella, C. et al. (2023). Copper Binding and Redox Activity of α-Synuclein in
Membrane-Like Environment. *Biomolecules 13*, DOI: `10.3390/biom13020287`.

(*114*) Jensen, P. H., Nielsen, M. S., Jakes, R., Dotti, C. G., and Goedert, M. (1998). Binding
of α-synuclein to brain vesicles is abolished by familial Parkinson's disease mutation.
Journal of Biological Chemistry 273, 26292–26294.

(*115*) Winner, B. et al. (2011). In vivo demonstration that α-synuclein oligomers are toxic.
Proceedings of the National Academy of Sciences of the United States of America 108,
4194–4199.

(*116*) Lemkau, L. R. et al. (2012). Mutant protein A30P α-synuclein adopts wild-type fib-
ril structure, despite slower fibrillation kinetics. *Journal of Biological Chemistry 287*,
11526–11532.

(*117*) Ghosh, D. et al. (2013). The parkinson's disease-associated H50Q mutation accelerates
α-synuclein aggregation in vitro. *Biochemistry 52*, 6925–6927.

(*118*) Xu, C. K. et al. (2022). The Pathological G51D Mutation in Alpha-Synuclein Oligomers
Confers Distinct Structural Attributes and Cellular Toxicity. *Molecules 27*, DOI: `10.3
390/molecules27041293`.

Chapter 5

Cu^{II} Dummy Atom Model: Binding Affinity and α-Synuclein Dimer

In previous chapter, the structural characteristics of the N-terminal macro-chelated Cu(II)-bound αS were evaluated. Despite the evidence presented for that macro-chelate form of copper coordination, the intermediary, near-range binding sites involving the same residues in the N-terminal, are still significant, with experimental observations on both the M1-D2-H$_2$O and V49-H50-H$_2$O docking sites, Figure 5.1,[1] while the M1-D2-H50 binding mode has also been related to the dimerisation of αS, with the copper ion bridging two peptide chains through residues M1 and D2 from one monomer and H50 from the other.[2–6] The present chapter can be broken down into three key objectives: (1) create an alternative non-bonded cationic dummy atom model (CaDAM) for Cu(II); (2) assess the binding affinity of the near-range M1-D2-H$_2$O and V49-H50-H$_2$O docking sites, through steered MD and semiempirical simulations on 7-residue fragments for each site; (3) implement the CaDAM to study the structure of the dimeric αS. Here, only the binding free energy of the near-range binding modes was considered, as a way of estimating the binding affinity to each of them, and elucidating the anchoring residue for the metal ion. The macro-chelate could not be assessed owing to the size of the system and the lack of control on the direction for the increasing distance, that leaves the CaDAM interacting with residues along the whole peptide.
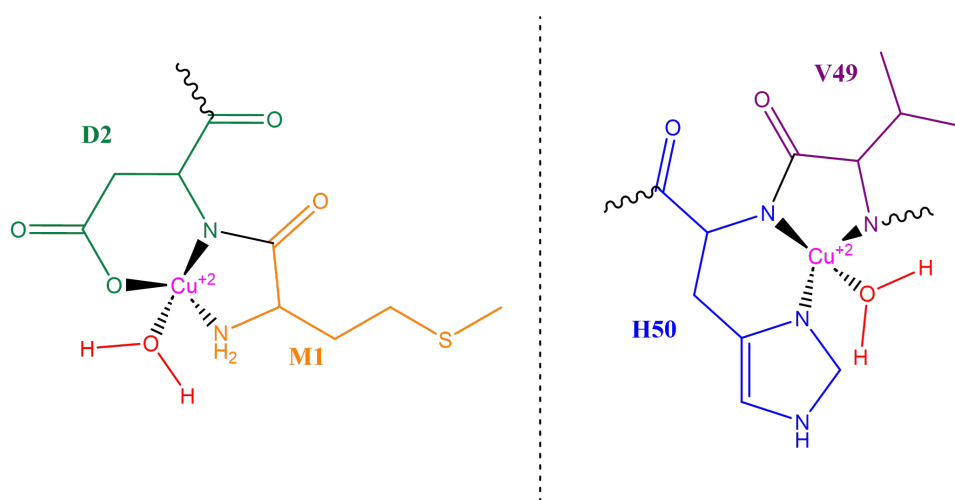
Figure 5.1: Schematic of the two metal ion binding sites, in the N-terminal, between residues M1-D2 and V49-H50.

5.1   Computational Methods

In order to assess the binding affinity of Cu(II) to each of the docking sites, a cationic dummy atom model (CaDAM) was constructed, as a way to circumvent the harmonic bonds between the coordinating atoms and the metal ion, and permit the unrestricted interaction of the CaDAM with the full length of the chain. This new model incorporates four dummy atoms placed in a regular square planar arrangement around the uncharged metal centre at a distance of 1 Å,[7, 8] each carrying +0.5 charge, to account for the +2 charge of the metal centre, thereby transferring the metal ion's charge to the dummy atoms. This model has already been described in the literature as a valid approach to simulating transition metal ions.[9] To assess the capacities of this model to interact with the reported residues, the macro-chelate system studied in Chapter 4 was used, replacing the metal ion with the CaDAM, and examining the distance between the residues of interest. The simulation protocol involved three classical MD runs, each at 300 ns, with the first 100 ns disregarded as equilibration. Considering the nature of the accelerated MD simulations, the macro-chelate structure of the binding pocket and the non-bonded model used here, it was decided to concentrate on the classical MD simulations, as a bias on the potential energy could push the metal ion out of the binding site. Nevertheless, aMD simulations were also performed, for three runs of 500 ns, to examine the stability of the dummy model interactions within the reported docking site. The binding affinity of Cu(II) to the two docking sites was estimated, through a series of steered MD simulations, pushing the CaDAM away from the docking site.

When estimating the relative binding free energy, adaptive steered MD (ASMD) simulations were performed,[10] utilising 100 runs, where the distance between the metal ion and the atoms involved in the bonding, using fragments M1-G7 and G47-A53, was increased by 4 Å at each run[a]. This step size represents the cumulative distance that the metal ion is moved in each run, achieved through numerous smaller steps during the simulation, increasing the distance between the metal and the binding pocket, at a pulling speed of 10 Å/ns, and a 7.2 kcal mol^{-1}/Å force constant.[11, 12] Simulations on the macro-chelate were also performed, although given the lack of a defined direction to pull the CaDAM out of the binding pocket, left the metal interacting with residues far from the site, thus no reliable estimates could be obtained. After modelling the systems in MOE and minimising the structure using the LFMM approach with harmonic bonds between the metal ion and the ligating atoms, the metal ions were thus replaced with CaDAM. This allowed for a quicker equilibration of the system during the MD simulations, after minimising the systems again now using the CaDAM. The systems were minimised, before performing 0.2 ns (200 thousand steps) of steered MD (SMD) for every 4 Å increase. Having completed the first set of simulations, with the distance between the metal ion and the ligating atoms increased from 1 to 5 Å, the Jarzynski average (Jar$_{avg}$) was obtained after looking at the work performed in all the runs and getting the average. The repeat that was closest to Jar$_{avg}$ was used as the restart point to extend the distance from 5 to 9 Å, again looking at 100 runs. This was repeated until the distance between the metal ion and the coordinating atoms was increased to 29 Å, to ensure there were no more interactions between the metal and the amino acid chain. The ff03ws force field and OBC implicit solvent were used throughout. The differences between explicit and implicit models using SMD have previously been studied, where similar PMF energies between the two solvation approaches were reported.[13]

---

[a]The distance in ASMD simulations was between the CaDAM and the centre of mass of atoms in M1-G7: backbone N of Met1 and Asp2, and O from carboxyl group in Asp2; and in G47-A53: backbone N of V49 and H50, and Nδ from imidazole ring in H50.

Steered MD simulations in implicit solvent for the study of protein-ligand interactions were subsequently reported by another research group.[*14*]

Having obtained the binding free energy of the two docking sites, another approach was used to validate the findings, through the implementation of a semiempirical method, GFN2-xTB.[*15*] The calculations involved a stepwise distance increase, between the metal ion and one of the atoms invovled in the coordination[b], from 1 to 30 Å within 600 steps, minimising at each increment and obtaining the energy.

Lastly, for the simulations on the metal-free dimeric system of αS, two chains were modelled in the extended conformation and placed at a *ca.* 12° angle to each other, crossing at H50 of both chains, as it has been found to be the region populated with folded secondary structural elements, and the likely area for inter-peptide interactions. The angle between the two chains was chosen after considering the increasing distance between the residues with wider angles, where the system did not have enough time to form interactions, with the two monomers instead drifting away from each other, sometimes within 50 ns. Several runs of conventional MD simulations were in fact performed for the metal-free dimer, with only 5 out of a total 12 maintaining interactions between the two chains for the entirety of the simulations, Table C.5. In the case of the copper-bound dimer, the two extended chains were bridged by two CaDAM, each placed between residues M1 and D2 of one chain and H50 of the other, Figure 5.2, which proved enough to sustain interactions between the two chains, without the need for multiple runs. Having constructed the two systems, five individual classical MD runs of 400 ns (200 million steps at 2 fs time step) were simulated, with the first 100 ns disregarded for equilibration.

Figure 5.2: Minimised structure of the copper-coordinated αS dimer, with CaDAM (blue-pink) bridging residues M1 and D2 from one chain with H50 from the other.

5.2   Cationic Dummy Atom Model of Cu[II]

As mentioned above, in order to assess the ability of the CaDAM to interact with the residues of interest, the macro-chelate bonded model, was used as a reference, with the metal ions replaced with CaDAM on the two binding sites (M1-D2-H50 and D119-D121-N122-E123), Figure 5.3. The focus of this section is mainly on the distance between the non-bonded metal ion model and the centre of mass (COM) of the residues reported to interact with copper. The results below are from 3×200 ns of classical MD simulations, after discounting 100 ns from the beginning for equilibration, along with results from accelerated MD simulations, which were performed for 3×500 ns.

---

[b]The GFN2-xTB approach only allowed the control of distance between two atoms. The terminal N of Met1 for the first site and the N of the imidazole ring of His50 were therefore selected, being the most flexible atoms in the two coordination sites, due to their sidechain locale.
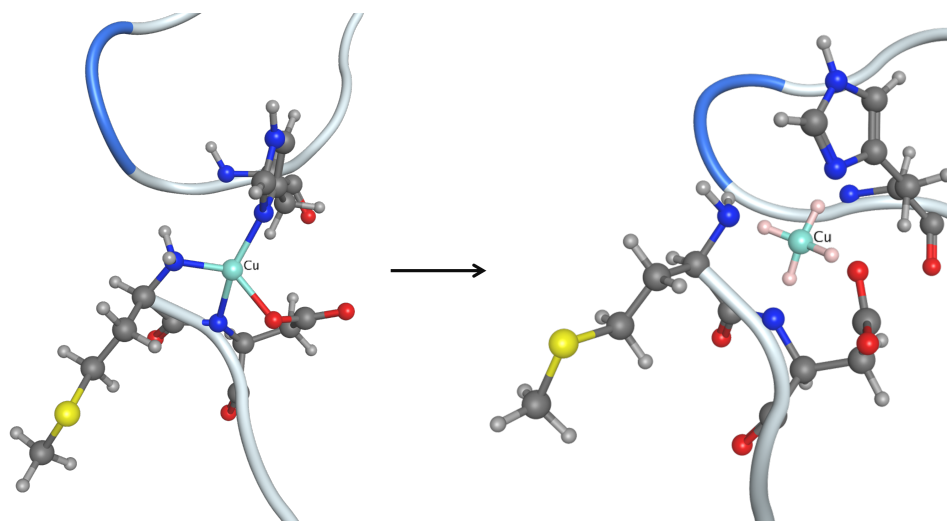
Figure 5.3: Arrangement of atoms involved in the N-terminal macro-chelate binding site, with dummy atoms (pink) in a square planar order around the metal centre, creating the cationic dummy atom model.

The contact map of the distance between the residues and the two CaDAM, after the three classical MD runs is shown in Figure 5.4. There, it is clear that the CaDAM remains on average within a close distance to the coordinating residues, especially where the C-terminal binding site is concerned, where the interspace is very much constrained around residues D119-E123. The interactions are not as clear when looking at the N-terminal binding site, given the macro-chelation which brings together two remote sites, although the shorter distances still appear to exist around the coordinating residues. A closer evaluation of these separation values between the COM of residues M1, D2 and H50 and the CaDAM is shown in Figure C.1, where D2 displays the closest distance to the metal ion, in all three runs, while all three residues remain within an interacting distance, except H50, which exhibits an increased distance (*ca.* 7 Å) in the first run after 135 ns (excluding the equilibration steps). Looking at Figure 5.4, this could be because of an interchange between the carboxyl group of Glu35 and the imidazole ring of H50, as a result of the decrease in the distance between the former and the CaDAM from the initial macro-chelate conformation. Nevertheless, of note is that when looking at the the accelerated dynamics of the first run, this is reduced to a range where interactions are feasible, within the first 30 ns, suggesting that this metal ion model is able to interact with the residues of interest, even when it is biased to explore the wider conformational space.
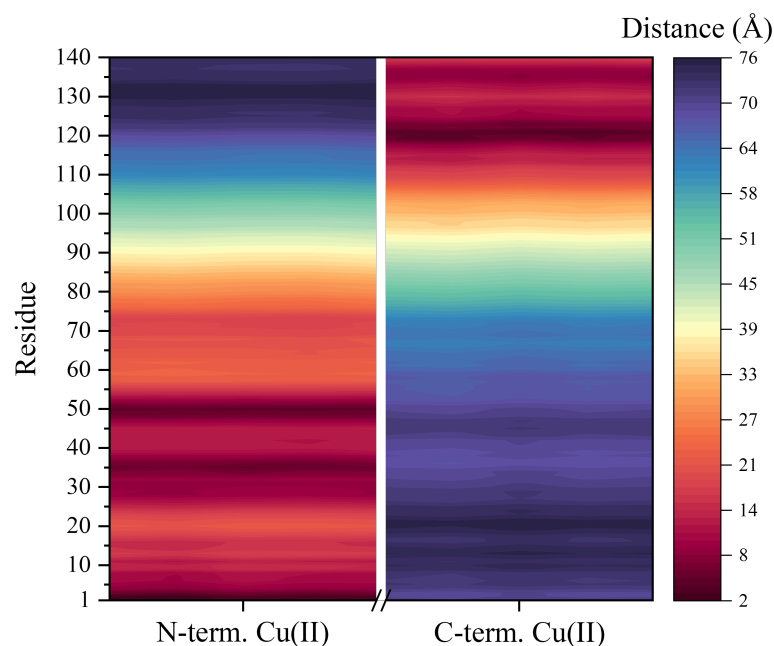
Figure 5.4: Contact map of the average distance between the residues in the system and the four dummy atoms of the two CaDAM, from all three runs using classical MD.

At this point, the CaDAM proved to be a good model for the metal ion, maintaining an interaction with the residues reported in the literature. Given the performance of accelerated MD simulations, it was decided to also perform a series of enhanced sampling runs, with a bias on the potential energy, to further investigate how permanent these interactions are. Figures 5.5 and C.2 show that despite the bias on the potential, the CaDAM remains in close proximity to the residues of interest, again with H50 displaying the greatest fluctuations. Looking at the change in the distance between the CaDAM and H50, Figure C.2, quite unexpectedly the third run, where the metal ion showed the most stable interactions with that residue in the conventional MD, now appears to oscillate between the bound and unbound state, until it eventually detached completely, and left from the vicinity of the H50 binding site after *ca.* 450 ns.
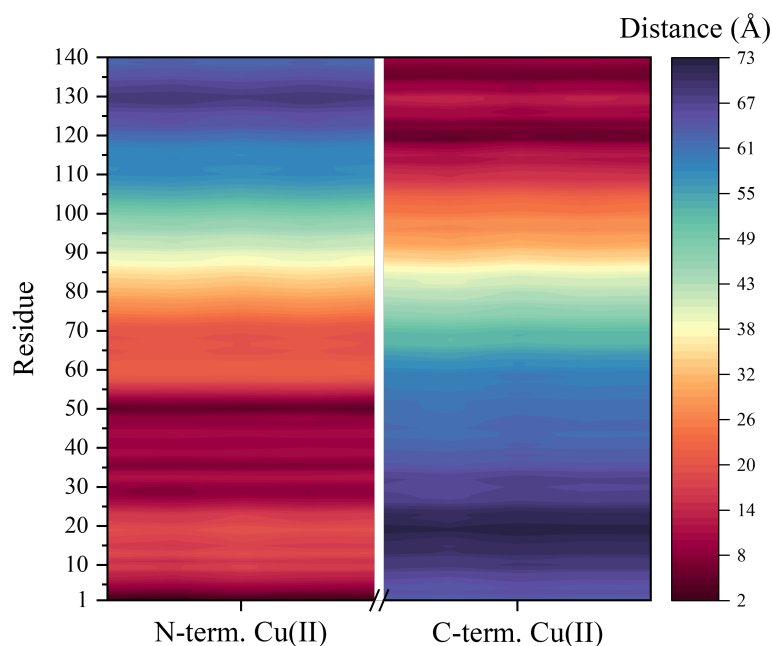
Figure 5.5: Contact map of the average distance between the residues in the system and the four dummy atoms of the two CaDAM, from all three runs using accelerated MD.

Taken together, these findings suggest that the CaDAM is a good model for the metal ion, maintaining an interaction with the residues reported in the literature, even when the potential energy is biased to increase conformation sampling. The fact that the metal ion is able to maintain a close distance, especially with H50, where it only coordinates with one atom and the rest are from the distanced residues M1 and D2, further substantiates the capacity of this model to form interactions. Therefore, the CaDAM was used in the sections that follow, to assess the binding affinity of the metal ion to the near-range binding sites, as well as to study the αS-dimer.

5.3   Binding Affinity of Cu^{II} to α-Synuclein

In order to assess the binding affinity of the CaDAM to αS, two different methodologies were implemented. Figure 5.6, shows the relative energies from the ASMD and GFN2-xTB simulations. It should be noted that while the values reported from the SMD simulations can be used to describe the binding free energy, the ones from the GFN2-xTB method lack an entropic term, and thus can only be evaluated relative to the other binding site, using the same approach. The zero point energy (ZPE) term, relating to the internal vibrational motion at absolute zero, was calculated for the minimum and dissociated states, as well as pV = pressure × volume, and TS = temperature × entropy, as an approximation for the missing entropic term,[16] whereby $G = E + pV + ZPE - TS$, with G = Gibbs free energy, and E = binding energy. The energies presented in Figure 5.6, are rescaled with respect to the dissociated state, acting as the reference energy, thus providing the relative energy when the metal ion is interacting with the residues in the peptide fragments.
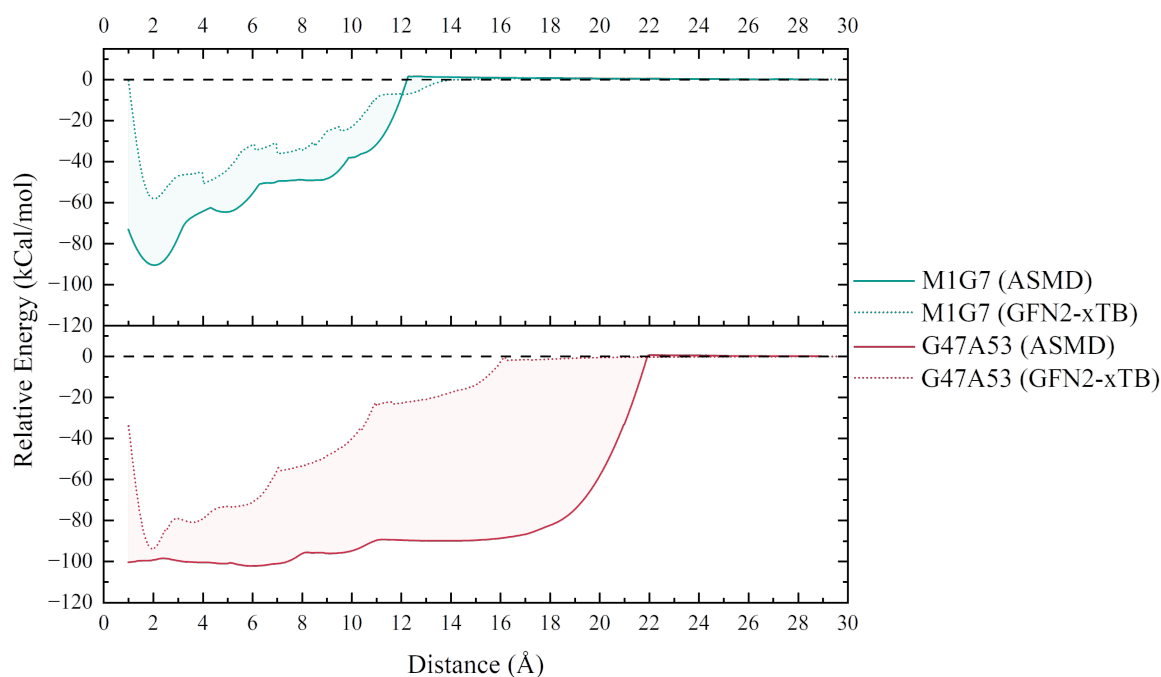
Figure 5.6: Relative binding free (ASMD) and binding (GFN2-xTB) energies for the two near-range N-terminal binding sites, at the different distances between the metal ion and the coordinating atoms.

In both simulation scenarios, the lowest energy was found for the G47-A53 system, at relatively good agreement in the energies from the two methods, with the natural state of the binding mode at -100.6 kcal mol^{-1} (relative to the unbound geometry) from the ASMD simulations, and -93.7 kcal mol^{-1} from GFN2-xTB, Table 5.1. The adjustment for the entropic term raised the energy to -85.4 kcal mol^{-1}. The difference in energies between the ASMD and GFN2-xTB methods was expected, considering the difference in the level of theory and the effects of the pulling speed and force on the spring in SMD. It should also be noted that the discontinuity in the slope near the distance where the metal ion dissociated from the fragment, is a potential result of a nonoptimal parameterisation of the electrostatic effects in the system. The distance between the metal ion and the binding pocket where the minima are observed is also different, with the lowest energy from the GFN2-xTB calculations found at an average distance, between the ligating atoms and metal ion, of 2.0 Å (i.e. at the binding pocket), while in the case of the ASMD simulations, the lowest energy was found at 5.9 Å, Table C.1. Despite the coordination observed for the G47-A53 system, considering the energy from having the metal in the binding site is only 1.5 – 2 kcal mol^{-1} higher than the lowest energy distance, and well within the potential range of deviation, we can assume that although this site is presented as the most stable, had we not applied any bias into the system (i.e. spring pulling the metal ion), the CaDAM would remain in the binding pocket. This is evident from the distances between the H50 and G51 atoms from the simulations on the macro-chelate structure, described in the previous section, where despite transient interaction with the O in H50, the majority of the interactions are maintained with the Nδ of the imidazole ring, and no apparent affinity to G51, Figure 5.7.
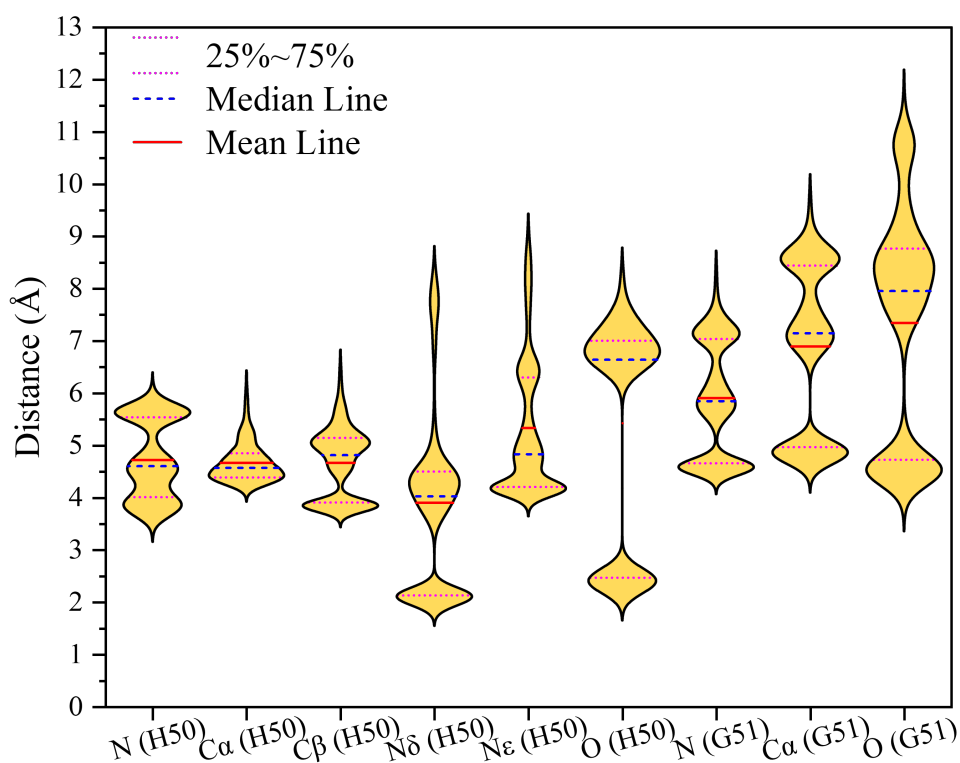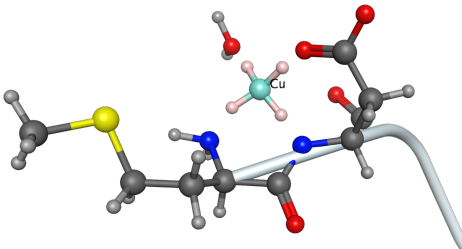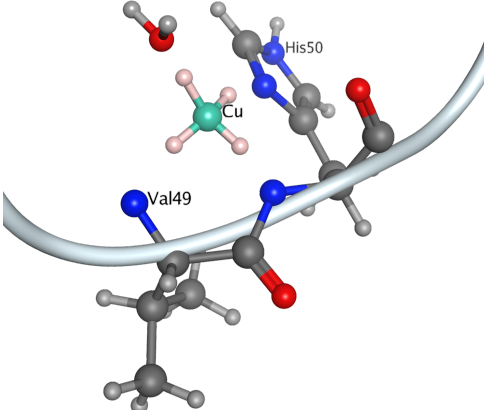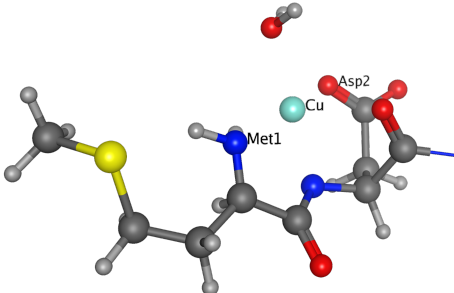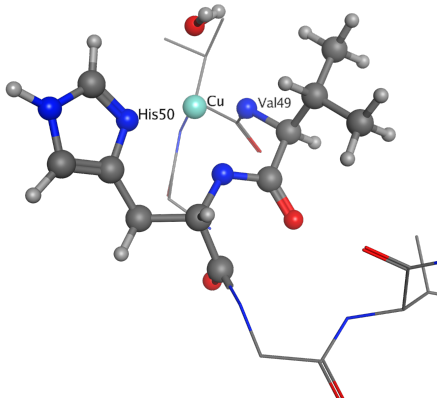
Figure 5.7: Violin plot of the distance between the metal centre of the CaDAM and backbone (Cα, N, O) and sidechain atoms (Cβ, Nδ, Nε) of H50 and G51, from the classical MD simulations presented in Section 5.2. The width of each 'violin' is proportional to the number of frames that fall within the distance range.

In the M1-G7 system we find quite a significant difference in the energies between the methods, coming at -90.6 kcal mol^{-1} from the ASMD simulations, and -58.4 kcal mol^{-1} from GFN2-xTB, which is increased to -48.1 kcal mol^{-1} after adjusting for the entropy. Considering the atoms interacting with the metal ion here, Table 5.1, this could potentially be a result of the partial charges on the modelled atoms, with a more negative charge on the oxygen atoms in the MM approach (-0.730, compared to -0.514 in GFN2-xTB), increasing its affinity to the metal ion. This higher affinity to the oxygen atoms, is also evident from the conformation with the lowest energy in G47-A53, with two of the interactions replaced with oxygen atoms, Table C.1. In the case of the GFN2-xTB approach, which is a self-consistent-charge density-functional tight-binding method (SCC-DFTB), the partial charges are calculated iteratively until they converge, thus resulting in more reliable estimates of the interactions between the coordinating atoms and the metal. In the case of the MM method, the partial charges are pre-assigned based on the force field used, however, for the residues interacting with the metal ion, these were recalculated using QM, as described in previous chapters. This was necessary especially for the methionine where, being a terminal residue, coordination through the terminal amine group requires it to be modelled as -NH$_2$ rather than the default -NH$_3^+$ [c].

---

[c]From a charge of +0.159 for N in -NH$_3^+$ to -0.871 in -NH$_2$.

Table 5.1: Binding energies from the ASMD simulations (top), and the GFN2-xTB calculations (bottom), for the two near-range N-terminal binding sites – along with the approximated free energy for the GFN2-xTB values, given in parentheses.

M1 – G7 (ASMD)   G47 – A53 (ASMD)

Distance: 2.06 Å   Distance: 1.02 Å

Relative Free Energy: -90.58 kcal mol^{-1} Relative Free Energy: -100.62 kcal mol^{-1}

M1 – G7 (GFN2-xTB)   G47 – A53 (GFN2-xTB)

Distance: 2.06 Å   Distance: 1.96 Å

Relative (Free) Energy: -58.39 (-48.14) kcal mol^{-1} Relative (Free) Energy: -93.72 (-85.39) kcal mol^{-1}

From the evidence presented here, we find the highest affinity binding site to be the G47-A53 system, at *ca.* 10% lower energy than the M1-G7 fragment, looking at the ASMD simulations, and *ca.* 80% lower in the semiempirical calculations, after the energy corrections. This is in agreement with the experimental findings, where His50 has been described as the anchoring residue for the copper ion.[7, 17–19] The difference in the ASMD lower energies, between the two systems, suggests that the MM approach maintains a similar affinity between the two sites although, as discussed above, the SMD simulations are susceptible to error; with the values obtained from the semiempirical calculations more in-line with experiments. Despite the questionable performance of SMD, the CaDAM still manages to maintain the interactions with the ligating atoms, and as seen from the simulations in the previous section, does a good job of conserving the binding site.

Looking at the non-bonded copper-αS simulations in the previous section, the CaDAM exhibited the highest fluctuation in distances at H50, suggesting that while V49-H50 is the most stable interaction, when considering the macro-chelated region, the sole coordination with H50, through the N of the imidazole ring, is not sufficient to maintain the binding affinity seen when considering the near-range binding mode. This, again, appears to follow experimental findings, with the *imidazole* interaction described to have a lower-affinity, compared to those in the N-terminal.[20] It should also be noted that the values reported here appear to be in conflict with experimentally reported values, where the binding energies of the N-terminal sites appear in the order of -7.5 to -12.5 kcal mol^{-1}.[1, 21] This difference has been attributed to the missing deprotonation energy, for the two backbone N in each of the sites. An attempt was made at obtaining the deprotonation energy in each of the systems, using QM optimisation calculations in Gaussian09 (B3LYP/6-31G(d,p)+), on the protonated and deprotonated 3-residue M1-A3 and V48-H50 systems, and using a solvated proton free energy of -273.07 kcal mol^{-1}, from a previous QM study, using the same functional.[22] The deprotonation free energy obtained from these calculations was +41.33 kcal mol^{-1} and +67.19 kcal mol^{-1}, for M1-V3 and V48-H50, respectively. This increased the binding free energies within close agreement to experiment, especially looking at the semiempirical calculations, where the binding free energy is now increased to -6.89 and -18.20 kcal mol^{-1}, for M1-G7 and G47-A53, respectively. This is by no means an extensive assessment of the system, but it is evidence of the effect of deprotonation on the binding energies.

The results presented thus far, suggest that while copper ions may preferentially bind to the H50 near-range binding site, upon initial approach to the peptide, the affinity to that residue is reduced after the macro-chelation, now displaying a higher affinity to the M1-D2 binding site. The mechanism facilitating this binding reposition is still a matter of speculation. In Chapter 4 we encountered the formation of an anti-parallel β-hairpin between residues L38-A53. Although the near-range binding sites were not simulated, the appearance of this structure in the systems examined in the previous chapter, suggests that this is in all likelihood a feature one can expect to be present when these near-range binding sites are engaged. The location and direction of this β-sheet, decreases the distance between H50 and the start of the chain, possibly facilitating the macro-chelate formation.

Given the nature of the methodologies used here, it was not possible to assess the binding affinity of the metal ion to the macro-chelated region, since increasing the distance on a fragment of that size, without control on the direction it can take, leaves the metal ion interacting with other residues along the peptide, rather than escaping into space. This is especially an issue when the CaDAM passes areas of folding characteristics, where it may spend energy to unfold them with increasing distance.

The analysis in these two sections, suggest that the CaDAM can be implemented towards the study of interactions between Cu(II) and the protein. The ensuing section will focus on the use of this dummy atom model to study the dimer of αS, with the use of two CaDAMs, bridging the chains through interaction with residues M1-D2 of one chain and H50 from the other.

5.4   Molecular Dynamics Simulations of αS Dimer

For the dimeric systems, examined here, only conventional MD simulations were performed, in five repeats of 400 ns, out of which only the last 300 ns were regarded in the analysis of the data[d]. Figures C.3 and C.4, displays the RMSD plots for the two systems, where both appear to have equilibrated within the first 100 ns that were removed from the analysis of the trajectories. Accelerated dynamics were also assessed, although the bias on the potential energy pushed the two chains apart within the first 100 ns, Figure C.5, and no meaningful analysis can be presented from these; the discussion is therefore focused on the 1.5 µs combined data from the classical MD simulations. It should also be noted that the chains in the CaDAM-free dimer do not follow the direction of the ones in the CaDAM-bound system, with the ones in the former case placed in the same direction, while in the latter in opposite directions, for the M1-D2-H50 coordination to be possible. This direction remains the same for the CaDAM-bound dimer, although the chains in the metal-free system appear to populate the perpendicular arrangement of chains, with the angle between the two chains, each regarded as a vector between the two terminal residues, peaking at 60° – 100°, Figure C.6. Figure 5.8, shows the average distance between the dummy atoms of the two CaDAMs and the backbone-C of the residues in the system, where it is clear that these remain within a close proximity to the residues they were first placed near, especially in the case of the CaDAM bridging M1-D2 from Chain A and H50 from Chain B.
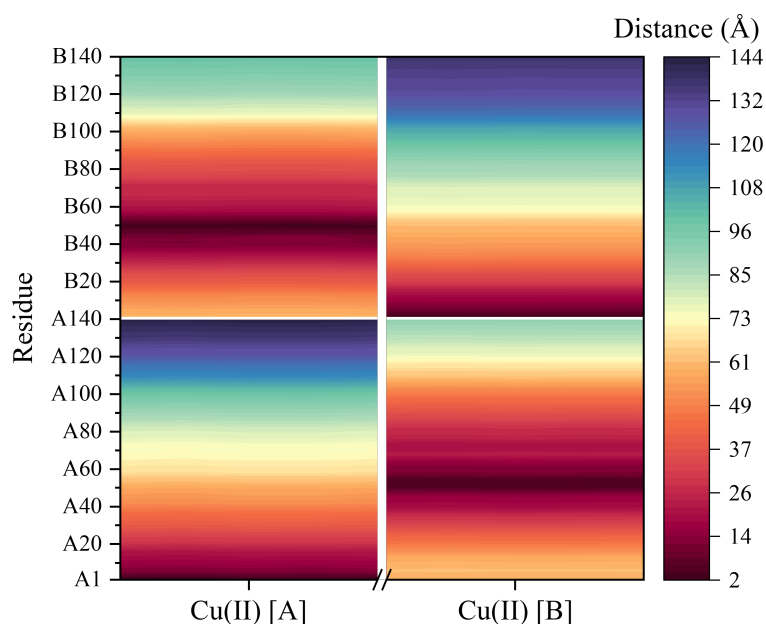
Figure 5.8: Contact map of the average distance between the backbone-C and the CaDAMs in the dimeric system.

The secondary structural elements of each of the chains in the two systems, were also examined, with a breakdown of the mean % in each of the regions given in Tables C.2 and C.3. Considering the Cu(II)-bound system in Chapter 4, where two Cu(II) ions are interacting each with the N- and C-termini of the monomer (NC-Cu(II)-αS), Table 4.11 as well as the system where Cu(II) only interacts with the N-terminal of the peptide (N-Cu(II)-αS), Table B.14, a

---

[d]The 1.5 µs combined trajectories for the two systems can be viewed (in 50-frame increments) at: CaDAM-dimer: doi.org/10.6084/m9.figshare.23552151; metal-free dimer: doi.org/10.6084/m9.figshare.23552475.

clear increase in the β-characteristics can be seen throughout. In particular, the N-terminal of Chain A now exhibits *ca.* 6 times greater β-content, compared to the N-Cu(II)-αS, and *ca.* 3 times greater than the NC-Cu(II)-αS, with a similar but higher increase for Chain B. The NAC-region of both chains also display an increased β-content, *ca.* 5 and 3 times higher than the N-Cu(II)-αS and NC-Cu(II)-αS, respectively. The difference in the percentages in each case are given in Table 5.2. In the case of the metal-free system, the β-content most influenced by the dimer coordination appears to be in the NAC region, where Chain A is twice as populated compared to the monomer. Of note is the fact that the characteristics between the two chains here, are not as alike as those seen in the CaDAM-bound dimer. This is also reflected in the average secondary structure plots below, Figures 5.9 and 5.10, where despite the characteristics following an identical trend, they are on average more pronounced in Chain A, especially where the β-content is concerned.

Table 5.2: Percentage of β-content in the different regions of the monomeric and dimeric systems. NC-Cu(II)-αS and N-Cu(II)-αS are for the N- and C-terminal Cu(II)-bound, and N-terminal Cu(II)-bound monomeric systems, respectively. The values for the monomers are from accelerated MD simulations, while the dimers are from conventional MD. Standard deviations are given in parentheses.

Region	N-terminal (%)	NAC (%)	C-terminal (%)
free-αS	2.12 (3.83)	6.84 (7.55)	0.20 (1.01)
free-dimer (Chain A)	4.77 (4.41)	15.31 (8.12)	0.43 (1.51)
free-dimer (Chain B)	3.84 (3.83)	9.40 (8.53)	0.37 (1.54)
NC-Cu(II)-αS	2.45 (3.70)	5.10 (6.96)	0.20 (1.03)
N-Cu(II)-αS	1.40 (2.75)	2.68 (5.46)	0.14 (0.83)
CaDAM-dimer (Chain A)	7.70 (9.33)	13.98 (9.69)	0.42 (1.40)
CaDAM-dimer (Chain B)	8.67 (8.55)	12.55 (9.77)	0.33 (1.38)

Interestingly, many similarities can be identified between the metal-free and metal-bound systems, with regards to the residues that are part of the most populated characteristics. In particular, the α-helical regions remain almost identical between the two systems, with the greatest content found between residues S87-K102. This region was also found to be the most populated in the monomeric system, and resulted in the formation of an amphipathic helix. The replication of this characteristic in the dimeric system, suggests that the presence of the metal ion, or additional chains, does not affect the folding of this stable region. In fact, the only difference in the helices between the two dimer systems, appears to be in the persistence of the region between residues V55-V63 (where one of the KTKEQV repeats is contained), which is about twice more pronounced in the metal-bound system. This region was also found to be populated with α-helices in the monomer systems, Figure 4.32, although at *ca.* 40-42% occupancy in both the metal-free and Cu(II)-αS. The inter-chain interactions could have aided in that regard here, with the stabilisation of these helices from hydrogen bonds formed between the two chains, especially in the CaDAM-bound dimer where these are more populated, Figure 5.14.

The greatest differences in the secondary characteristics, appear to arise from the β-content

of the two systems. The most populated region involved residues V63-T72, which was also identified in the monomeric systems as one of the locations exhibiting long-lasting β-hairpin character, through the formation of anti-parallel β-strands. This characterstic is replicated here in the two chains of both systems through intra-chain parallel β-strands, Figures 5.11 and C.9. Inter-chain interactions are also formed between frames 90,000-120,000, through the formation of a parallel β-sheet, in the metal-free system. Upon assessing the per-residue secondary characteristics of the other regions in the two systems, with respect to the frames in the trajectories, the β-hairpin that was present between residues L38-A53 in the monomers is now repurposed towards the formation of inter-peptide β-sheets, bridging the two chains and by extension aiding in the conservation of the dimer, Figure C.8. While this region displays a high propensity for this inter-peptide interaction in the metal-free system, the same is not observed in the CaDAM-bound dimer. Upon a closer inspection of the per-residue characteristics, this bridging appears to be instead facilitated by the residues in the extended N-terminal region (between residues V15-V40), Figures C.7 and C.9(B). From the former figure, it is clear that this β-sheet is not maintained in all five repeats, only appearing in three of them – in the rest (between frames 60,000-120,000), the two chains instead appear to only be held together by the two CaDAMs. This is in line with the short-lived nature of secondary characteristics seen in the monomer systems, owing to the intrinsically disordered nature of the protein. Overall, the N-terminal region of both systems give rise to the anchoring characteristics that help sustain the dimer.
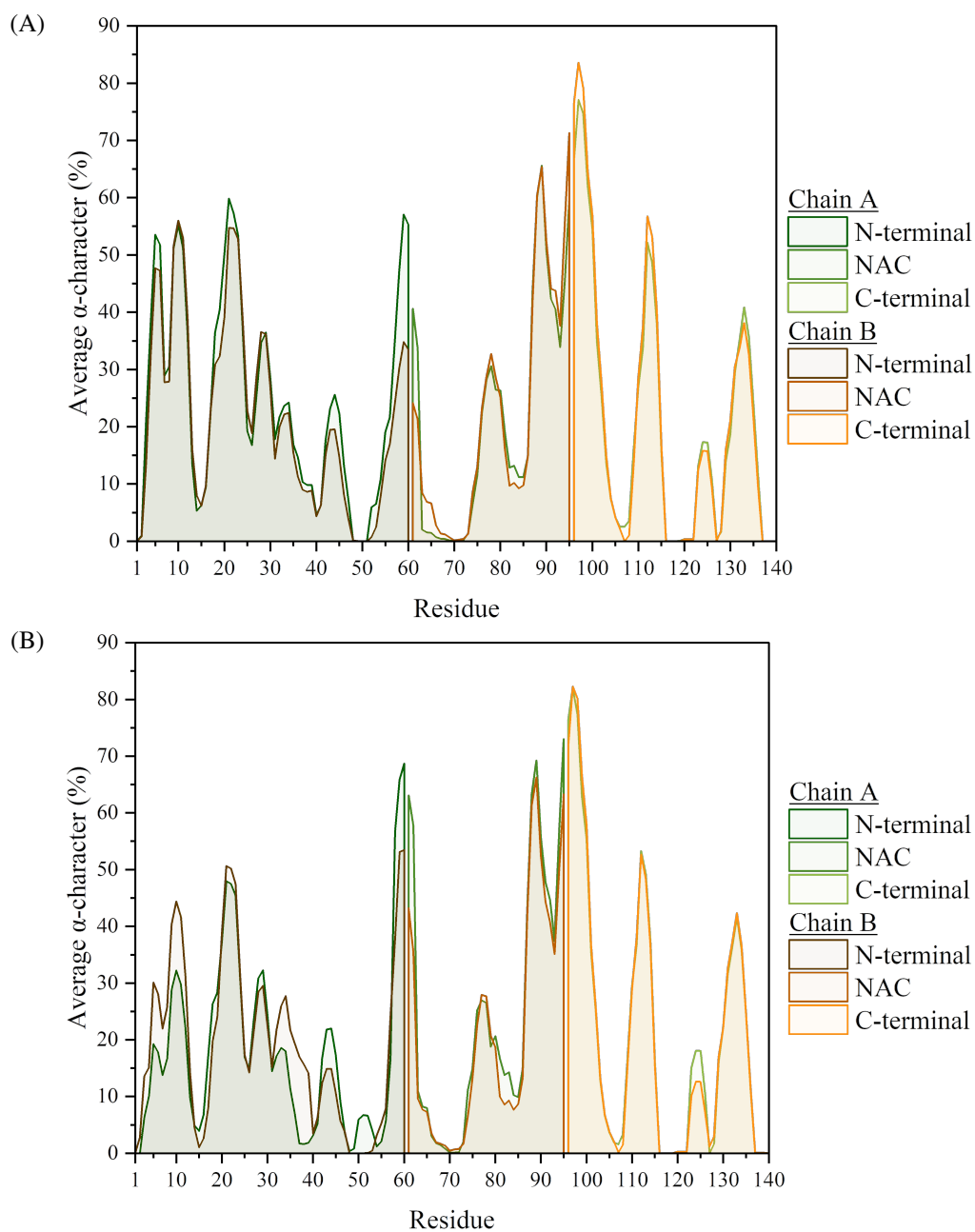
Figure 5.9: Average α-characteristics of the residues in the (A) metal-free and (B) CaDAM-bound dimeric system.

(A)

(B)
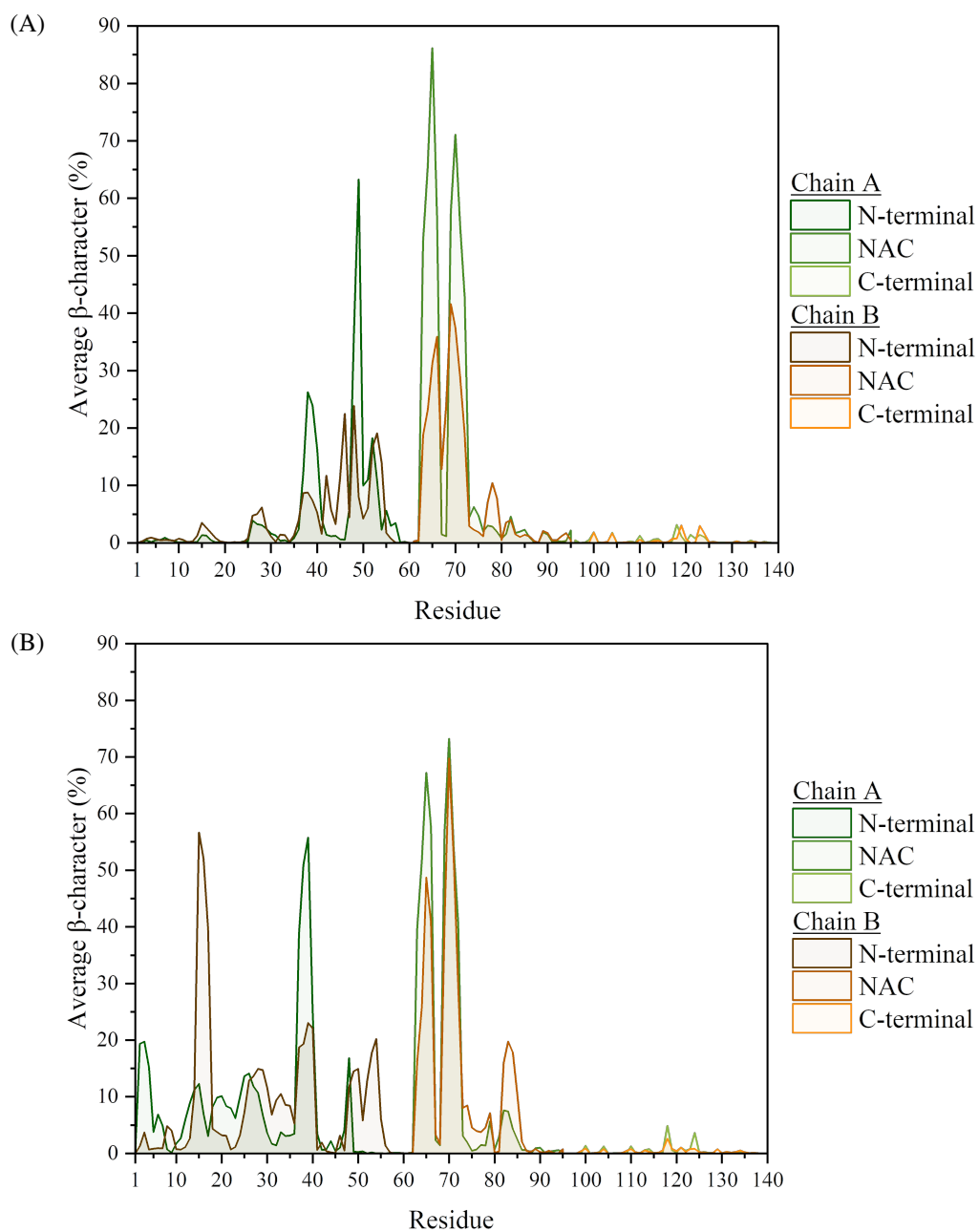
Figure 5.10: Average β-characteristics of the residues in the (A) metal-free and (B) CaDAM-bound dimeric system.
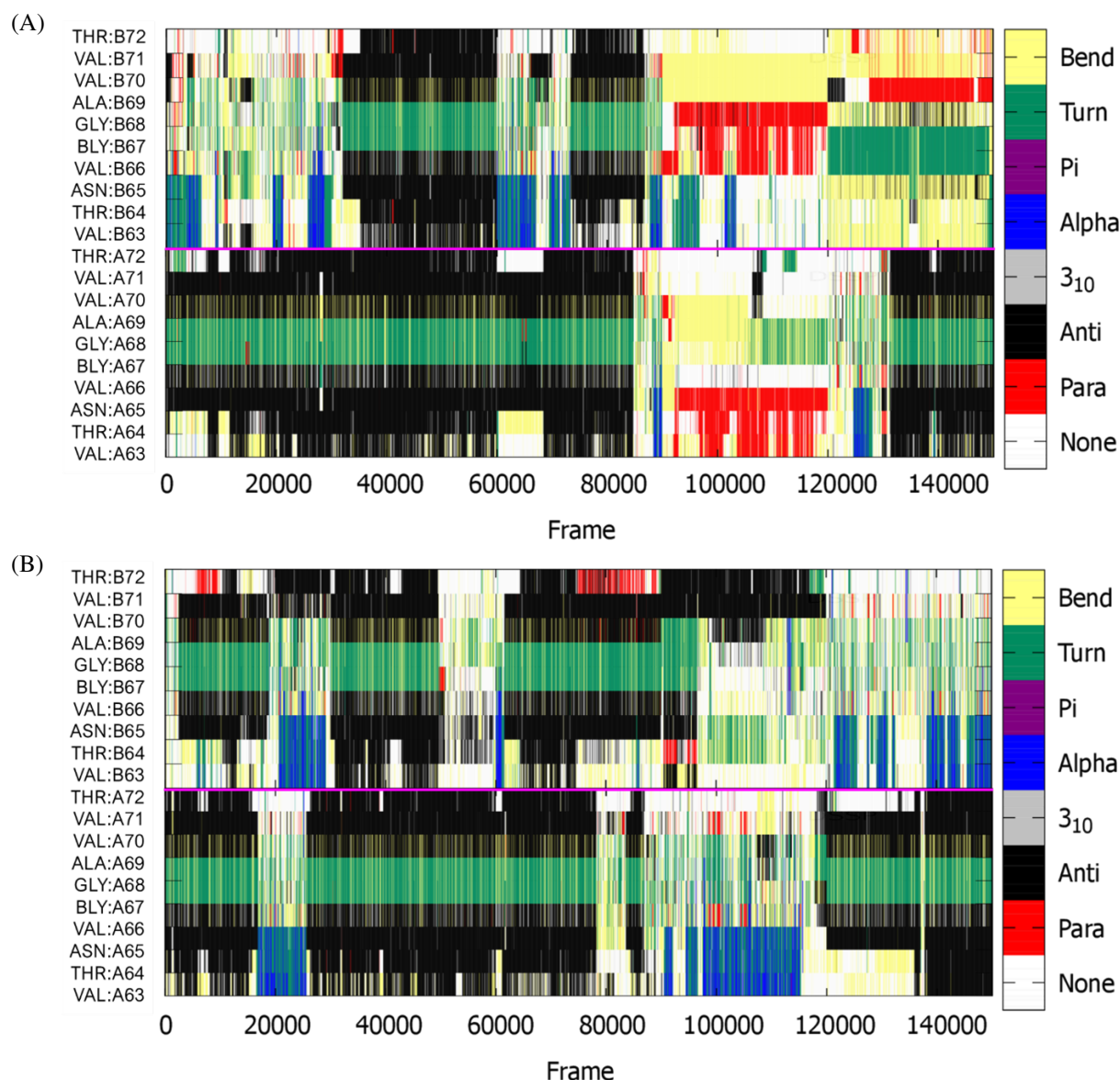
Figure 5.11: Evolution of secondary structural elements of each of residues V63-T72 in the two chains of the (A) metal-free and (B) CaDAM-bound dimers. The pink line separates the two chains.

The snapshots of the systems given in Figure C.9, illustrate the areas where β-sheets can form, although they do not represent the average conformation in these systems. Therefore, clusters were created using PCA of the Cα, detailed in Section 2.4.8.4.7. The results from this analysis are shown in Table C.4, with the average cluster structure for each of the systems given in Figure 5.12. The interactions in the metal-free dimer, now appear to be mainly focused on the intrachain V63-T72 β-hairpins, with the inter-chain interactions formed between the two hairpins. This is in accordance with the observations made above, for frames 90,000-120,000, suggesting cluster 1 may be populated with frames from this region; confirmed from DSSP analysis on these structures, Figure C.10, with *ca.* 16,000 ($\approx$ 36% of the whole) structures exhibiting this character. The CaDAM-bound system's average cluster structure is in accordance with the observations made in Figure C.7, discussed above, with an inter-peptide β-sheet between residues V37-Y39 from Chain A, and V15-A17 from Chain B.

(A)

(B)
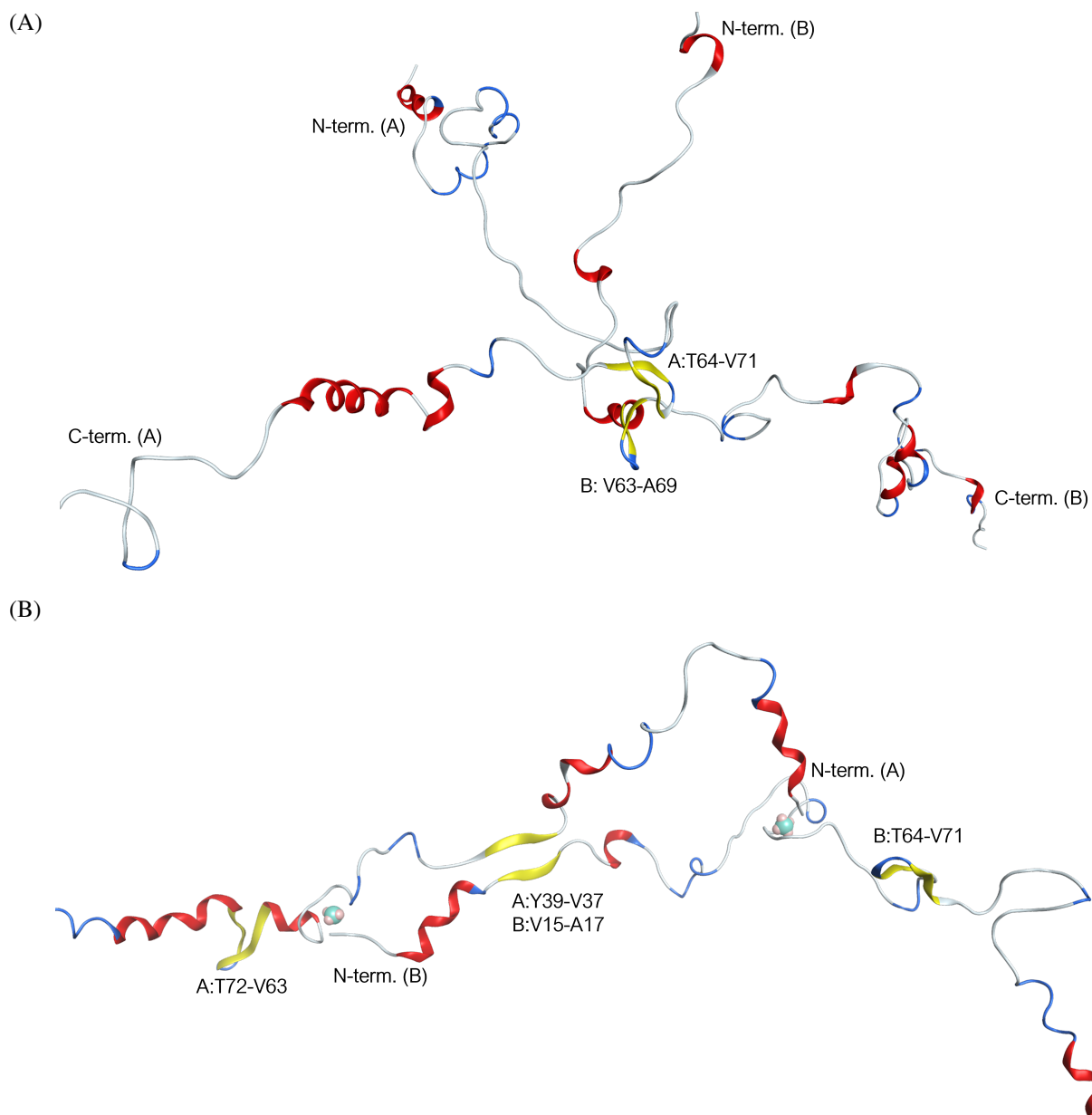
Figure 5.12: Snapshot of the two average cluster structures from the (A) metal-free and (B) CaDAM-bound dimers, with labelled residues involved in the β-strands (yellow) going from top to bottom, and left to right.

Below, the Cα-contact maps for the two systems are shown, Figure 5.13, where the chains in both systems are found to interact through the N-terminal region. This interaction is more pronounced in the CaDAM-bound system, with the metal ions bridging the two chains both at the beginning and the end of the N-termini. In the case of the metal-free system, the inter-peptide interactions are concerted around residues V40-K60. These regions of close proximity between the two chains, unsuprisingly follow the areas of inter-peptide β-sheet formation, discussed above, as well as hydrogen bonds between the two chains, Figure 5.14. These are also both more populated and longer-lasting in the CaDAM-αS, decreasing the flexibility of the dimer. This is very much reflected in the RMSF plots, Figure 5.15, where the CaDAM-bound system is overall at least half as mobile as the metal-free system, especially in the two

chain ends, where the mean fluctuations of the residues range between 60-75 Å, as opposed to 27-30 Å in the metal-coordinated system. The shortest motions in the metal-free dimer are exhibited in the region between residues K45-T75, where the β-content is also the highest, Figure 5.10(A). The difference in the flexibility of the residues in the CaDAM-bound system is much less noticeable, with the fluctuations maintained between 20-33 Å. Despite this, the size of the two chains in the metal-bound dimer is overall more extended than in the metal-free system, Table 5.3, with the R_g 5 Å higher in both chains. This has been attributed to the extended intermolecular interactions in the N-terminal and NAC regions of the CaDAM-bound system, which are more constrained in the metal-free dimer. This is especially evident from the total R_g of the systems, where the CaDAM-bound system is *ca.* 13 Å larger than the metal-free system.
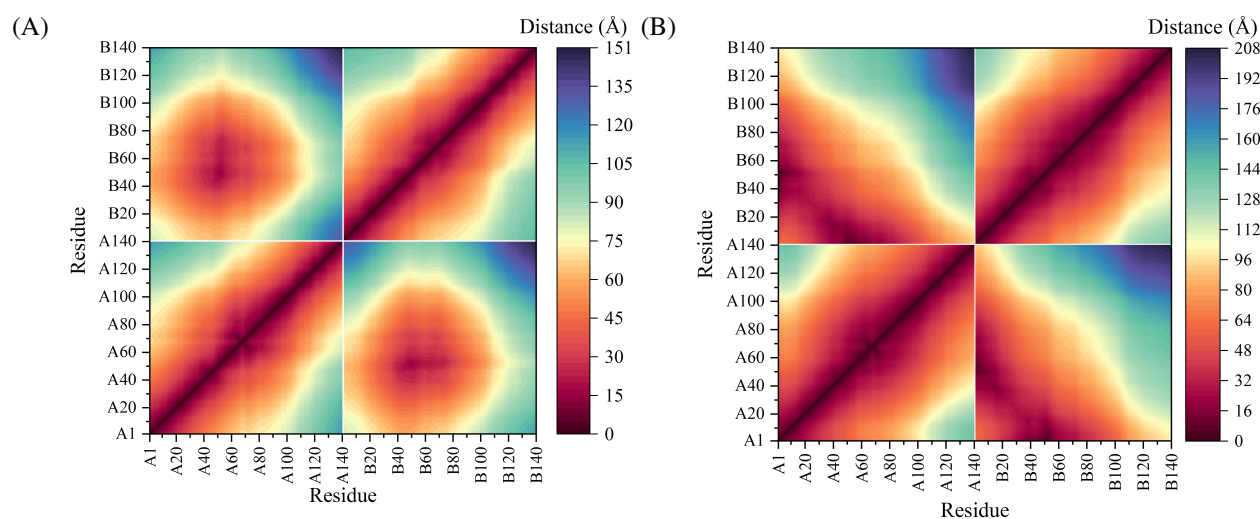
Figure 5.13: Contact maps of the distance between the backbone-C from the dynamics of the (A) metal-free and (B) CaDAM-bound dimeric systems.
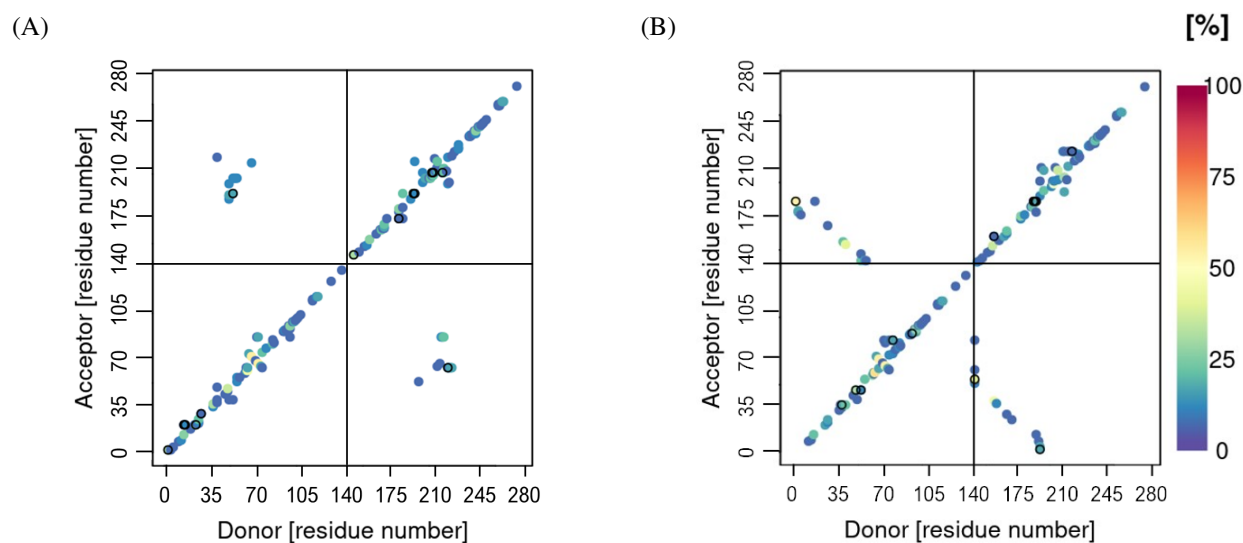
Figure 5.14: Hydrogen bonds in the (A) metal-free and (B) CaDAM-bound dimeric systems.
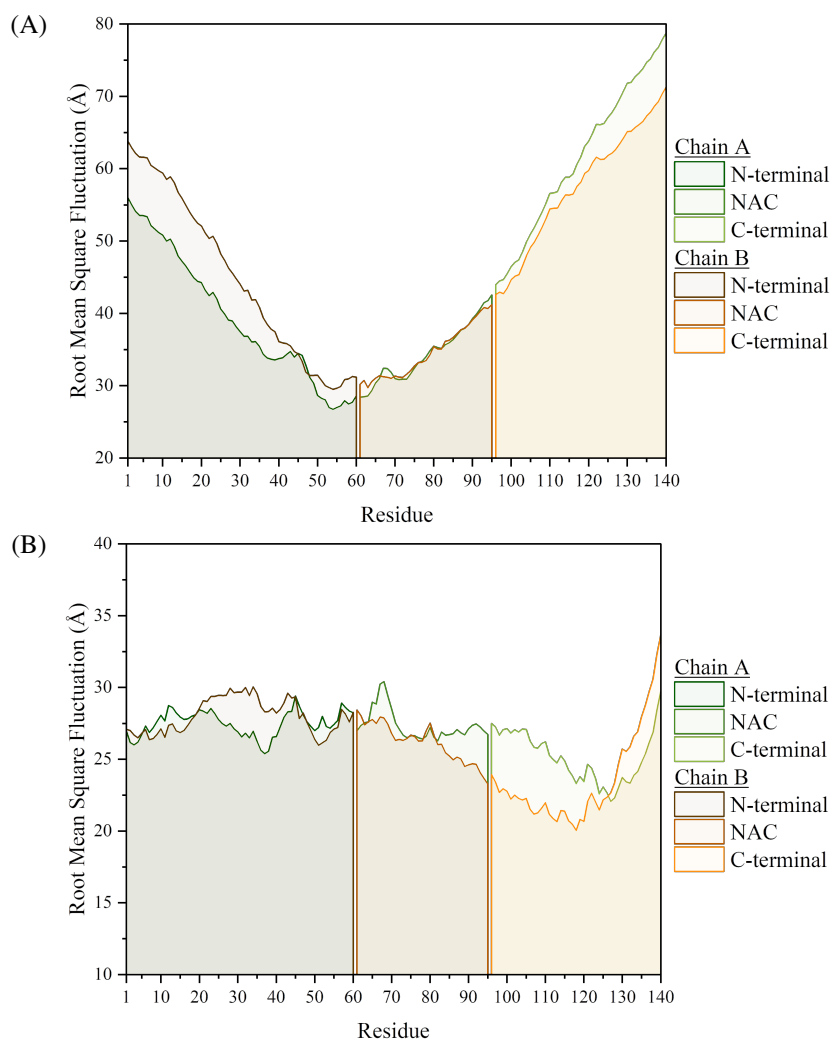
Figure 5.15: Root mean square fluctuation, of the individual residues in the (A) metal-free and (B) CaDAM-bound dimeric system.

Table 5.3: Radius of gyration of the two chains in the dimer systems.

Chain	Avg. R_g (Å)	SD (Å)	Max (Å)	Min (Å)
free-αS				
Chain A	41.96	5.35	61.65	26.47
Chain B	42.27	5.76	63.27	26.00
Total	50.26	4.75	66.12	36.17
CaDAM-αS				
Chain A	46.76	5.47	64.37	31.62
Chain B	47.85	7.77	77.51	28.26
Total	63.53	6.65	90.03	45.63

5.5 Conclusions

This chapter introduced a different way of simulating Cu(II), through the use of charged dummy atoms arranged in a square planar geometry around an uncharged metal centre. This approach enabled a non-bonded description of the metal, allowing interactions with the extended residues in the system, while also permitting the calculation of binding free energies. This was done through the use of steered MD simulations on 7-residue fragments of the near-range binding modes in αS. Section 5.2 validated the CaDAM in its ability to reproduce the experimental binding modes, where it was found to maintain coordination with residues M1, D2 and H50, even when a bias was applied to the potential energy of the system. The model was subsequently implemented for the calculation of the binding free energy with two, experimentally observed near-range binding modes, M1-D2-H_2O and V49-H50-H_2O. The energy profiles pointed towards a higher affinity for the latter, with a 10 kcal mol^{-1} lower energy than the M1-G7 site. The findings were corroborated with a semiempirical minimisation of the systems, with the metal ion simulated at increasing distances from the binding residues with a 37 kcal mol^{-1} lower energy in G47-A53. Despite a *ca.* 17% and 87% increase in the binding free energies (after corrections for entropy) for the M1-G7 and G47-A53 sites, a similar trend was observed in the energy profiles of both systems, with the second site exhibiting higher affinity to the metal ion, similar to the MM approach. This difference in energies was attributed to the change in the level of theory, but also the effects of the pulling speed and force constant on the spring in SMD – given more time this would have been something worth exploring further. The values obtained from the semiempirical approach were also adjusted with an estimate for the deprotonation energy from QM calculations, bringing the free energy values within close agreement to experimental observations,[*1, 21*] at -6.9 and -18.2 kcal mol^{-1}, for M1-G7 and G47-A53.

Having concluded a sound description of the metal ion using the cationic dummy atom model, the next section focused on simulating the αS dimer, both with and without the CaDAM. Two of these ions were placed within interaction distance of residues M1-D2 of one chain and H50 of the other, as described in EPR experiments.[*4, 5*] The findings showed that the metal ions were, on average, able to maintain coordination with the residues they were initially interacting. Stable secondary characteristics were also identified for both the free and CaDAM-bound systems, specifically intramolecular β-hairpin structures between residues V63-T72, also expressed in the monomeric systems. Furthemore, inter-peptide bridging was found to proceed through β-sheets formed between these β-hairpins, as well as between residues L38-A53, in the metal-free; while in the CaDAM-bound system, these interactions were maintained only between residues M1-V40. In the case of the metal-free dimer, the L38-A53 β-sheet has been repurposed from the β-hairpin structure observed in the monomeric system, while in the CaDAM-bound system, a number of β-sheets are formed and maintained throughout the N-terminal, between the two binding sites. These bridging structures have been important in maintaining interactions between the two chains, specifically in the metal-free system, where in the absence of β-sheets shared between the two monomers, the dimer dissociated (Table C.5). In the case of the CaDAM-bound system, however, even in the absence of these β-sheets, the dimer was able to maintain its inter-peptide linkage, through the CaDAM bridges formed between the two chains. This speaks towards the stability of the multimers formed in the presence of the metal ion, which are less prone to dissociation and increase the persistence of interactions between chains.

Bibliography

(*1*)   Valensin, D. et al. (2011). The role of His-50 of α-synuclein in binding Cu(ii): PH dependence, speciation, thermodynamics and structure. *Metallomics 3*, 292–302.

(*2*)   Dudzik, C. G., Walter, E. D., and Millhauser, G. L. (2011). Coordination features and affinity of the Cu2+ site in the α-synuclein protein of Parkinson's disease. *Biochemistry 50*, 1771–1777.

(*3*)   De Ricco, R. et al. (2015). Remote His50 Acts as a Coordination Switch in the High-Affinity N-Terminal Centered Copper(II) Site of α-Synuclein. *Inorganic Chemistry 54*, 4744–4751.

(*4*)   Drew, S. C. (2015). The N Terminus of α-Synuclein Forms Cu II -Bridged Oligomers. *Chemistry - A European Journal 21*, 7111–7118.

(*5*)   Drew, S. C. (2016). Probing the quaternary structure of metal-bridged peptide oligomers. *Journal of Inorganic Biochemistry 158*, 30–34.

(*6*)   Tian, Y. et al. (2019). Copper2+ Binding to α-Synuclein. Histidine50 Can Form a Ternary Complex with Cu2+ at the N-Terminus but Not a Macrochelate. *Inorganic Chemistry*, DOI: `10.1021/acs.inorgchem.9b02644`.

(*7*)   Rasia, R. M. et al. (2005). Structural characterization of copper(II) binding to α-synuclein: Insights into the bioinorganic chemistry of Parkinson's disease. *Proceedings of the National Academy of Sciences of the United States of America 102*, 4294–4299.

(*8*)   Ramis, R., Ortega-Castro, J., Vilanova, B., Adrover, M., and Frau, J. (2017). Copper(II) Binding Sites in N-Terminally Acetylated α-Synuclein: A Theoretical Rationalization. *Journal of Physical Chemistry A 121*, 5711–5719.

(*9*)   Liao, Q., Pabis, A., Strodel, B., and Kamerlin, S. C. L. (2017). Extending the Non-bonded Cationic Dummy Model to Account for Ion-Induced Dipole Interactions. *Journal of Physical Chemistry Letters 8*, 5408–5414.

(*10*) Ozer, G., Valecv, E. F., Quirt, S., and Hernandez, R. (2010). Adaptive steered molecular dynamics of the long-distance unfolding of neuropeptide y. *Journal of Chemical Theory and Computation 6*, 3026–3038.

(*11*) Ozer, G., Quirk, S., and Hernandez, R. (2012). Adaptive steered molecular dynamics: Validation of the selection criterion and benchmarking energetics in vacuum. *Journal of Chemical Physics 136*, DOI: `10.1063/1.4725183`.

(*12*) Park, S., Khalili-Araghi, F., Tajkhorshid, E., and Schulten, K. (2003). Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of Chemical Physics 119*, 3559–3566.

(13) Bureau, H. R., Merz, D. R., Hershkovits, E., Quirk, S., and Hernandez, R. (2015). Constrained unfolding of a helical peptide: Implicit versus explicit solvents. *PLoS ONE 10*, 1–19.

(14) Peng, Y. et al. (2019). DFMD: Fast and Effective DelPhiForce Steered Molecular Dynamics Approach to Model Ligand Approach Toward a Receptor: Application to Spermine Synthase Enzyme. *Frontiers in Molecular Biosciences 6*, 1–12.

(15) Bannwarth, C., Ehlert, S., and Grimme, S. (2019). GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation 15*, 1652–1671.

(16) Menzel, J. P. et al. (2021). Efficient workflow for the investigation of the catalytic cycle of water oxidation catalysts: Combining GFN-xTB and density functional theory. *Journal of Computational Chemistry 42*, 1885–1894.

(17) Kowalik-Jankowska, T., Rajewska, A., Jankowska, E., and Grzonka, Z. (2006). Copper(ii) binding by fragments of α-synuclein containing M 1-D2- and -H50-residues; A combined potentiometric and spectroscopic study. *Dalton Transactions*, 5068–5076.

(18) Villar-Piqué, A. et al. (2017). Copper(II) and the pathological H50Q α-synuclein mutant: Environment meets genetics. *Communicative and Integrative Biology 10*, 1–4.

(19) Ranjan, P. et al. (2017). Differential copper binding to alpha-synuclein and its disease-associated mutants affect the aggregation and amyloid formation. *Biochimica et Biophysica Acta - General Subjects 1861*, 365–374.

(20) Mason, R. J., Paskins, A. R., Dalton, C. F., and Smith, D. P. (2016). Copper Binding and Subsequent Aggregation of α-Synuclein Are Modulated by N-Terminal Acetylation and Ablated by the H50Q Missense Mutation. *Biochemistry 55*, 4737–4741.

(21) Wilkinson-White, L. E., and Easterbrook-Smith, S. B. (2008). A dye-binding assay for measurement of the binding of Cu(II) to proteins. *Journal of Inorganic Biochemistry 102*, 1831–1838.

(22) Dutra, F. R., Silva, C. D. S., and Custodio, R. (2021). On the Accuracy of the Direct Method to Calculate pKa from Electronic Structure Calculations. *The Journal of Physical Chemistry A 125*, 65–73.

Chapter 6

A53T and pS129 Modifications to α-Synuclein

In Chapter 4, the effect of Cu(II)-coordination to wild-type (WT) αS was studied, after validating the MD methodology on the unbound system. An examination of the experimental studies on the nature of αS in diseased brains, revealed that the majority of patients suffering from synucleinopathies, carry an alternate form of WT-αS, phosphorylated at S129 (pS129). Despite being found in 90% of patients with PD, as opposed to the 4% occurrence in normal brains,[1–3] studies on the effect of this phosphorylation in the protein's ability to aggregate, still have not settled on a definite conclusion. In spite of a report suggesting the promotion of fibrillation upon phosphorylation at Ser129,[1, 4] other studies argue the phosphorylated form may be an outcome of the disease itself, as a result of proteolysis impairment acting as a signal for degradation,[5, 6] with no correlation to its ability to form fibrils, or in certain cases inhibiting their expression altogether.[6–8] Notwithstanding the disagreement of studies on the impact of pS129 on the aggregation capacities of αS, it has been shown that oxidative stress induces phosphorylation at S129.[9] Recognizing the effect of copper coordination to αS, catalysing the production of reactive oxidative species (ROS),[10] as well as experimental evidence suggesting an increase in the binding affinity of divalent metal ions upon phosphorylation of αS,[3, 11] the Cu(II)-bound peptide with the aforementioned post-translational modification (PTM), pS129-αS, is studied here. Experimental observations on the copper-bound pS129-αS, report higher binding affinity in the C-terminal binding site,[12–14] although to the best of our knowledge no data has been published on the structural effects that result from such interactions. The GFN2-xTB semiempirical approach was implemented here for the estimation of the binding affinity of the metal ion to that site.

One of the first point mutations to be identified in αS, and related to familial-PD, is the substitution at position 53 of Ala with Thr. This mutation has been found to increase the formation of fibrils, by accelerating the release of protofibrilar intermediates.[15, 16] In turn, it has been related to an increased aggregation of the protein, as well as a change in the neurotoxicity of the resulting structures; although the effects have been shown to be dependent on the conditions, with some suggesting an increased toxicity as a result of the mutation,[17, 18] while others reporting the opposite.[19] The mutation has also been reported to result in post-synaptic degeneration, thus increasing the risk of developing synucleinopathies.[20] When it comes to membrane interactions, the mutation exhibited no particular changes compared to the WT-form, maintaining a similar binding affinity to the lipid bilayer.[21, 22] The coordination of Cu(II) on A53T-αS has already been examined experimentally,[23] where it was found to maintain the

same increased aggregation seen in the WT-form, thus no particular increased aggregation is expected in the A53T form, although the structural changes that take place upon this complexation with Cu(II) are still of interest. The metal-free A53T system has previously been simulated in explicit solvent (conventional MD and REMD) and implicit solvents (REMD);[24, 25] the novelty here is in the simulation of the copper-coordinated system, as well as the MD methodology used.

This chapter aims to dive deeper into the two modifications to the WT-αS, described above, with the goal of elucidating their effects on the aggregation capacities of αS, by looking at the structural changes that accompany them.

6.1 Computational Methods

Having established the methodology that works best in reproducing experimental results of α-Synuclein, as described in Chapter 4, an identical approach was used here, with force field ff03ws and OBC implicit solvent model. For the pS129 peptides, the parameters for the phosphorylated residue were obtained from the phosaa10 force field.[26, 27] The parameterisation for the molecular dynamics simulations were kept the same, the only difference being the simulations comprise of 3 individual 100 ns conventional MD, to obtain the average potential energy, which was used to apply the boost energy for the accelerated MD simulations, which ran for 3×600 ns. The metal binding sites for both the pS129 and A53T systems were kept the same as that of WT-αS, despite a report suggesting an alternative C-terminal coordination site (through D115-M116 and P128-S129),[11] not enough evidence are present to support that binding mode, which is why it was not examined here. The semiempirical approach for estimating the binding free energy, described in the previous chapter, was implemented here for the calculation of the affinity of the metal ion to G47-T53, and the C-terminal of the WT and pS129 systems, using the V118-E130 fragment. In order to assess changes in affinity of the metal ion to the binding sites, the distance of the metal to the G47-T53 fragment was increased from the Nδ of the imidazole ring, and in the V118-E130 fragment, from the O of the carboxyl group of D121. An illustration of the fragments used for the calculation of the binding free energy, after an initial minimisation, is given in Figure 6.1.
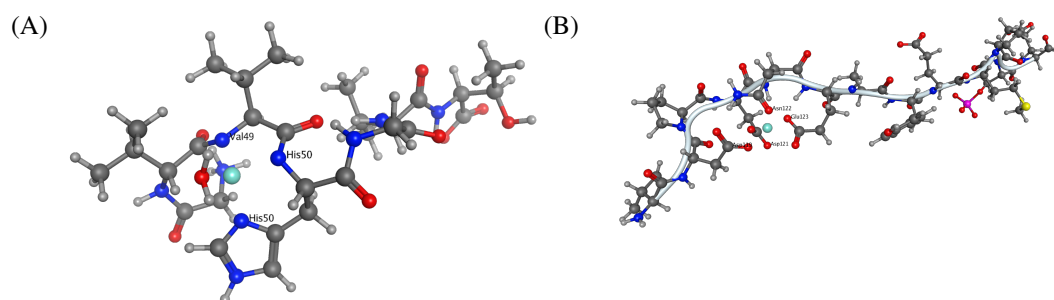
Figure 6.1: Structure of the metal-binding sites used for the calculation of the binding energies in the (A) A53T system, between residues G47-T53, and (B) pS129 system between residues V118-E130.

6.2   A53T Mutation

The most striking differences between the WT and A53T forms of the peptide, is a significant increase in the R_g values that was observed for both the metal-free and metal-bound systems, Table 6.1. This increase is much more pronounced in Cu(II)-αS, exhibiting double the size expansion seen in the metal-free system. Nevertheless, considering the conformational space explored in each case, the latter system samples structures with an R_g as low as 29 Å and as high as 75 Å, with a SD of 9 Å, as opposed to the metal-bound system, where the SD is more in-line with the sampling range seen in the WT-forms, at 5 Å. This is particularly evident in the distribution of R_g, Figure 6.2, where the metal-free system does not exhibit a preferred conformation, instead sampling a wide range of structures, with a slight peak around the mean R_g value of 51 Å. This is not the case, however, for the metal-bound system, where the distribution is much more concentrated around the mean value of 47 Å. This overall increase in the expansion of the peptide, upon the A53T mutation, is in line with another MD study on the mutated system, where an average difference of 6 Å was reported, between the WT and A53T forms.[25]

Table 6.1: R_g in the WT and A53T systems.

System	Mean R_g (Å)	SD (Å)	Max (Å)	Min (Å)
free-αS [WT]	44.26	4.58	61.50	28.07
free-αS [A53T]	50.96	8.94	75.12	29.02
Cu(II)-αS [WT]	34.94	4.06	62.34	21.93
Cu(II)-αS [A53T]	46.83	4.99	64.31	31.63
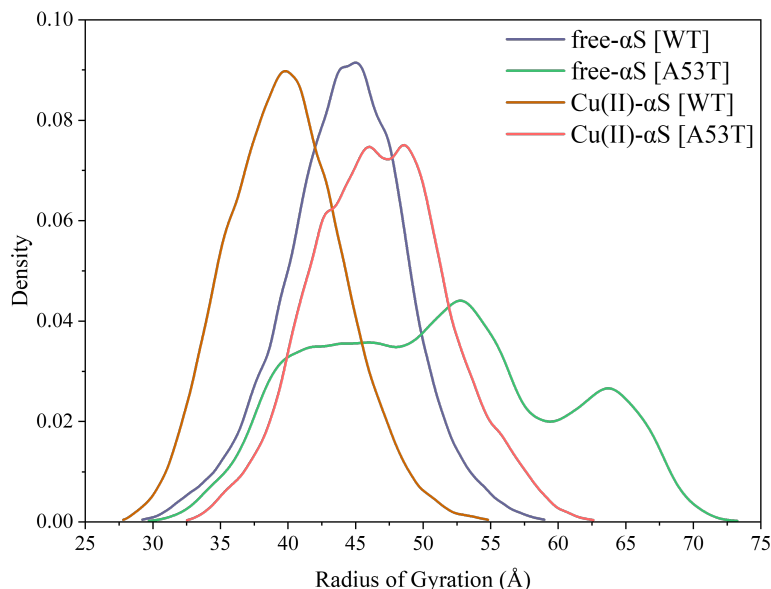
Figure 6.2: Distribution of R_g from the WT and A53T systems.

Considering the increased R_g values in the mutated systems, it is not surprising that the Cα contact maps, Figure 6.3, display increased distances between the residues. Another study looking at the mutated metal-free system, reported the loss of long-range N-terminal–C-terminal and

NAC–C-terminal interactions as a result of the A53T substitution – something also observed here.[25] The metal-free system, now exhibits an increased distance between the backbone-C throughout, without any significant off-diagonal close contacts, unlike the WT-form, Figure 4.26. This is not something that is observed in the metal-bound system, where the overall trend in the Cα contact map is similar to that of the WT-form, except for the increased distances throughout the system. The root mean square fluctuations (RMSF) were thus plotted to examine the extent of motion of these residues, and see how they compare to the WT-αS, Figure 6.4. The evidence here suggest an increase in the fluctuation of the residues in the N-terminal region, between residues M1-E46 in the metal-free and L8-G36 in the Cu(II)-bound systems. Thereafter, the mutated systems display decreased RMSF, except for the free-αS which shows an increase in the motion of residues L100-E126.
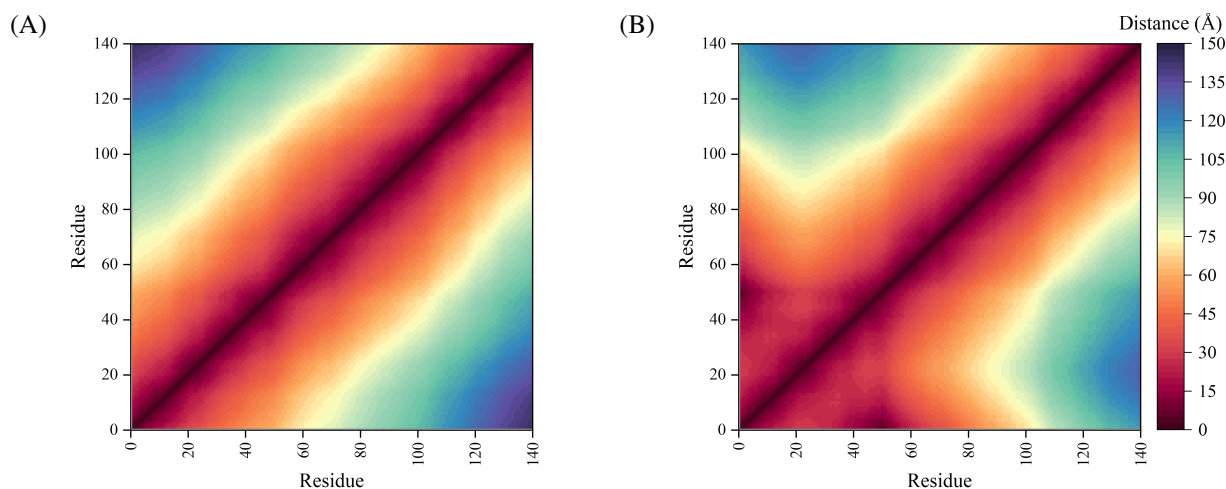
Figure 6.3: Contact maps of the distance between the backbone-C in the (A) metal-free and (B) Cu(II)-bound A53T peptides.
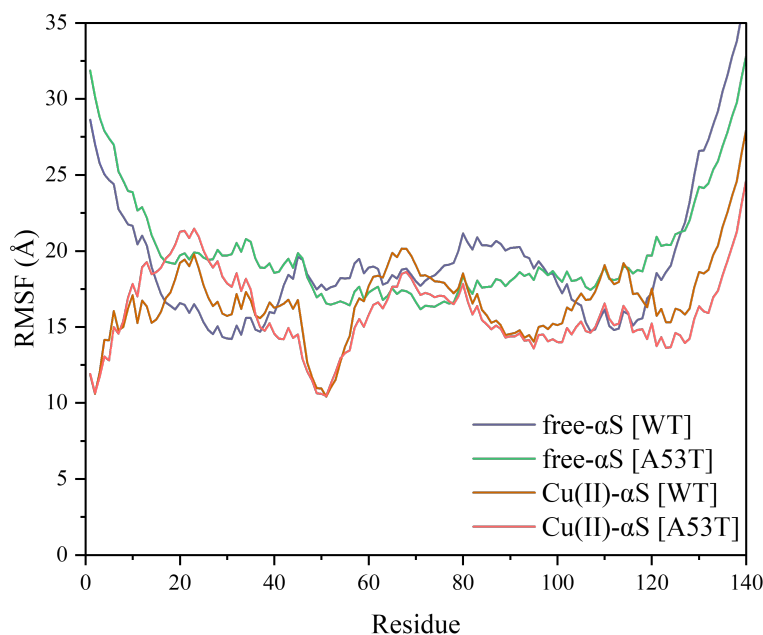
Figure 6.4: RMSF of the residues in the metal-free and Cu(II)-bound A53T peptides.

For the intramolecular hydrogen bonds in the two systems, Figure 6.5, there are great similarities between the WT and A53T systems. To better examine the changes in the hydrogen bonding network, the newly formed, or broken hydrogen bonds (existing for >10% of the trajectory length) are reported in Table 6.2. A thing to note is that the majority of the broken hydrogen bonds in the copper-coordinated system involve residue Glu35, which was responsible for the distortion of the Cu(II)-coordination in the N-terminal of the WT-form, Section 4.2.3. The square planar geometry of the binding site is instead distorted, to a similar degree by Glu46. This is seen from the average bond lengths and angles given in Tables D.1 and D.2, closely matching the ones seen in the WT-αS, Tables B.8 and B.9.
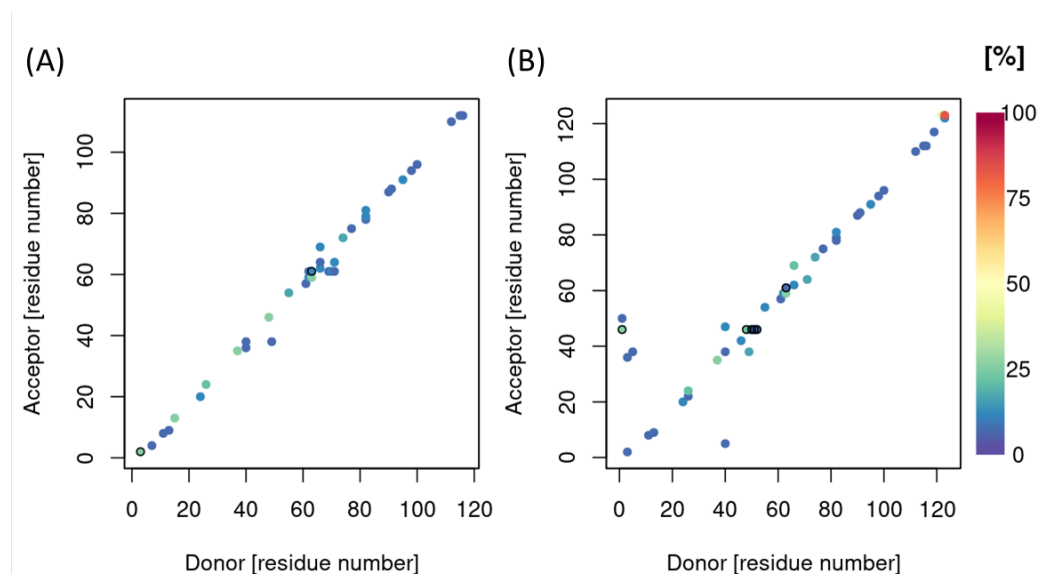
Figure 6.5: Hydrogen bonds in the (A) metal-free and (B) Cu(II)-bound A53T peptides.

Table 6.2: Newly formed or broken hydrogen bonds in the A53T systems (maintained for >10% of the simulations). Brackets have been added to the instances where residues form an additional hydrogen bond through different atoms.

System	Acceptor	Donor	Retention (%)
New Hydrogen Bonds			
free-αS	Glu61	Val70	10.3
	Glu46	Met1	24.2
	Glu46	Val49	19.4
Cu(II)-αS	Glu46	His50	13.3
	Gly47	Val40	11.7
	Glu46	Val48	11.6
Broken Hydrogen Bonds			
free-αS	All broken hydrogen bonds exist for <10%		
	Glu35	Val49	28.2
	Glu35	Met1	27.4
	Glu35	His50	21.2
Cu(II)-αS	Glu13	Met1	12.8 (11.0)
	Glu13	His50	11.8 (10.4)
	Glu13	Val52	11.5

The secondary characteristics in these systems are in line with the WT-αS – especially for the copper-coordinated system – something already expected from experimental observations.[23] The metal-free system presents a drop of 1.14% in the β-sheet content, as a result of the diminished β-hairpin population between residues V63-T72, from 18.4% in the WT-form to 9.1% in the mutated system. Looking at the per-residue secondary characteristics of the copper-coordinated system, Figure 6.7, reveals an increase in the strand content of two of the residues involved in the β-hairpin, between L38-A(T)53, namely Tyr39 and Val48, from 10.9% and 11.0% in the WT-form to 32.3% and 28.4% in the A53T system. The aromaticity present at Tyr39, was recently suggested to be of vital importance to the aggregation of αS.[28] Further considering the implication of the β-hairpin region in the aggregation of αS, and the almost 400 ns expression of that in the Cu(II)-αS system, Figure 6.7, this system potentially demonstrates the greatest aggregation-promoting capabilities tested here.

Table 6.3: Secondary structure percentages from aMD simulations of the different systems.

System	β character (%)	α character (%)	Other (%)
free-αS [WT]	2.47	16.55	80.98
free-αS [A53T]	1.33	15.74	82.93
Cu(II)-αS [WT]	2.23	12.43	85.33
Cu(II)-αS [A53T]	2.48	13.07	84.45
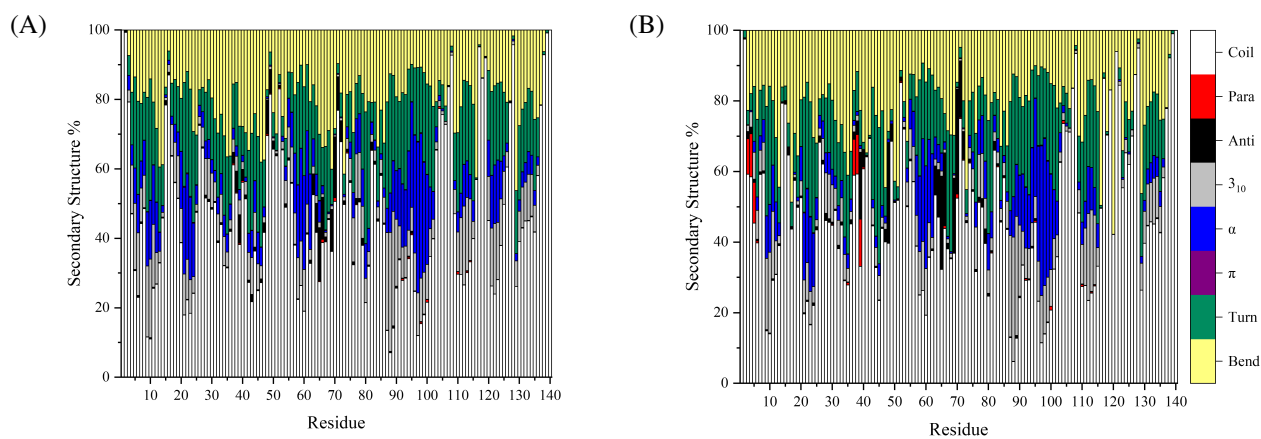
Figure 6.6: Secondary structure distribution per residue for the (A) metal-free and (B) Cu(II)-αS A53T systems. The β-sheets are denoted with red (parallel) and black (antiparallel), and helices with grey (3_{10}), blue (α) and purple (π).

(A)
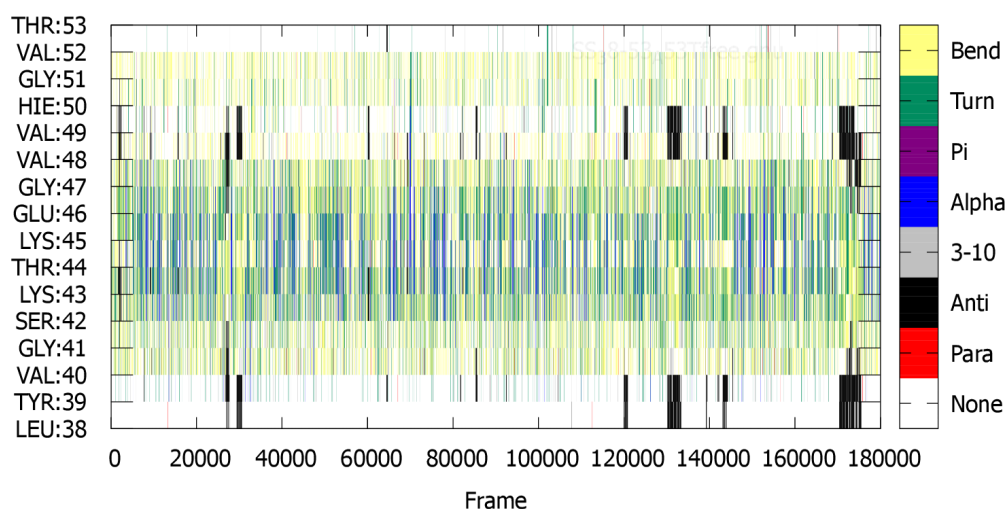
(B)
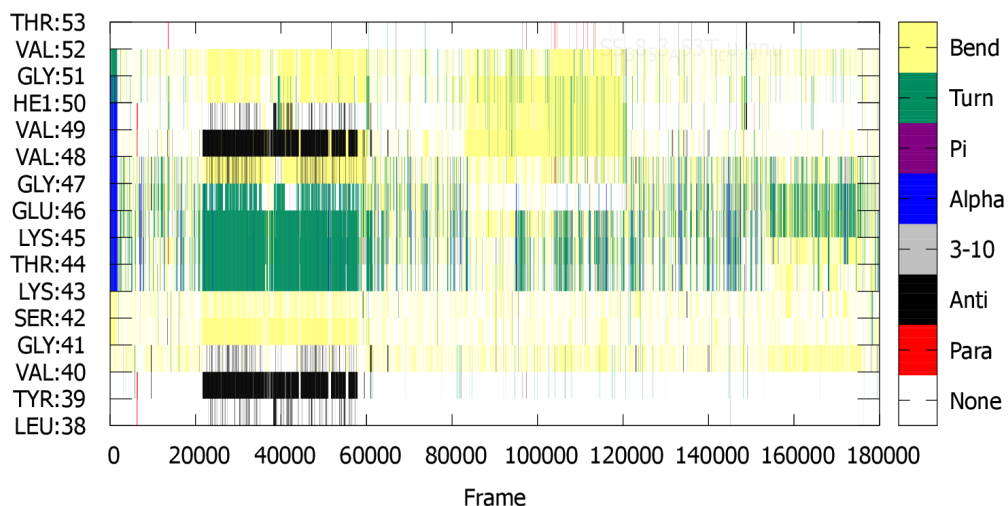
Figure 6.7: Secondary structure in each of the residues (L38-A53) in the primary hairpin region from the (A) metal-free and (B) copper-bound A53T-mutated systems, plotted against the frames in the trajectories.

The binding free energy of the metal ion to the G47-T53 site was also examined here, using implicit solvent semiempirical calculations. However, no particular change in the energy value from the metal ion in the WT site was identified, Figure D.1 – only decreasing from -85.39 kcal mol^{-1} to -85.58 kcal mol^{-1} in G47-T53, after corrections for the entropy. This could be due to the lack of interactions with residue A(T)53, seeing as it is the last residue in the fragment, along with the minimal effect on the distanced residues from the addition of a methyl and a hydroxyl group by the substitution of Ala with Thr; regardless, the mutation does not seem to affect the binding of the metal ion to H50.

Overall, the findings presented here revealed a significant increase in the R_g values of the mutated systems, especially in the presence of copper ions. The contact maps and RMSF presented increased distances between the residues, as well as higher fluctuations in the N-terminal of the mutated systems. The great expression of the β-hairpin region in the copper-bound mutated peptide, could suggest a greater aggregation propensity of this system, especially considering the involvement of this region in inter-chain interactions, as we have seen in the previous chapter.

6.3   Phosphorylated S129

Experimental observations on the phosphorylated αS have proposed parallels between the aggregation capabilities of the WT-αS and its pS129 modification, with evidence suggesting fibril formation is not affected by phosphorylation at S129.[6–8] From what has been reported in the literature, however, phosphorylation appears to increase the binding affinity of the C-terminal to divalent metal ions.[3, 11] The R_g in the metal-free phosphorylated system exhibits a similar distribution to the mutated system, seen above, where the range of conformations sampled is much broader than in the WT-form, Figure 6.8. Overall, an increase in the expansion of the peptide is seen here, with the greatest change from the WT-form found in the copper-coordinated system, with a *ca.* 10 Å increase in the R_g, Table 6.4. In fact, the changes observed in the average R_g here, are in line with those observed in the A53T system, only differing by 1-2 Å. Before exploring further what could constitute the reason for this increase in the size of the systems, we will first consider the changes in the secondary characteristics.
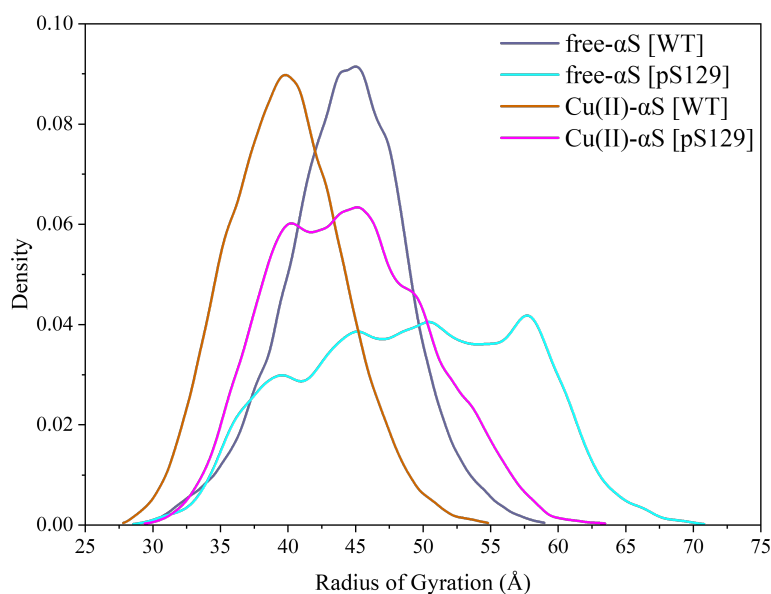
Figure 6.8: Distribution of R_g from the WT and pS129 systems.

Table 6.4: R_g in the WT and pS129 systems.

System	Mean R_g (Å)	SD (Å)	Max (Å)	Min (Å)
free-αS [WT]	44.26	4.58	61.50	28.07
free-αS [pS129]	49.24	8.00	71.39	28.43
Cu(II)-αS [WT]	34.94	4.06	62.34	21.93
Cu(II)-αS [pS129]	44.49	5.75	65.38	27.17

Below, the secondary structure distribution of the WT and pS129 systems is shown, Figure 6.9. The majority of the characteristics here are in line with the those seen in the WT-αS, except for a decrease in the β-sheet content from 2.5% and 2.2% in the metal-free and Cu(II)-bound WT-form, to 1.5% in the pS129-form. This drop in the metal-free system appears to come from the NAC region, similarly to the A53T system. This is not the case, however, for Cu(II)-αS where the decrease is from the N-terminal of the peptide, going from 2.45% in the WT-form to 0.65% in pS129-αS, Table D.3. This overall reduction in the β-content of the systems, could hint towards a decrease in the aggregation propensity of the peptide, as a result of the phosphorylation at S129, restricting the inter-peptide interactions that could form. This is further supported by experimental evidence, using a Thioflavin-T fluorescence binding assay, which showed a decrease in the aggregation of the phosphorylated αS.[8]

Table 6.5: Secondary structure percentages from aMD simulations of the different systems.

| System	β character (%)	α character (%)	Other (%)
free-αS [WT]	2.47	16.55	80.98
free-αS [pS129]	1.57	16.22	82.21
Cu(II)-αS [WT]	2.23	12.43	85.33
Cu(II)-αS [pS129]	1.53	13.22	82.25
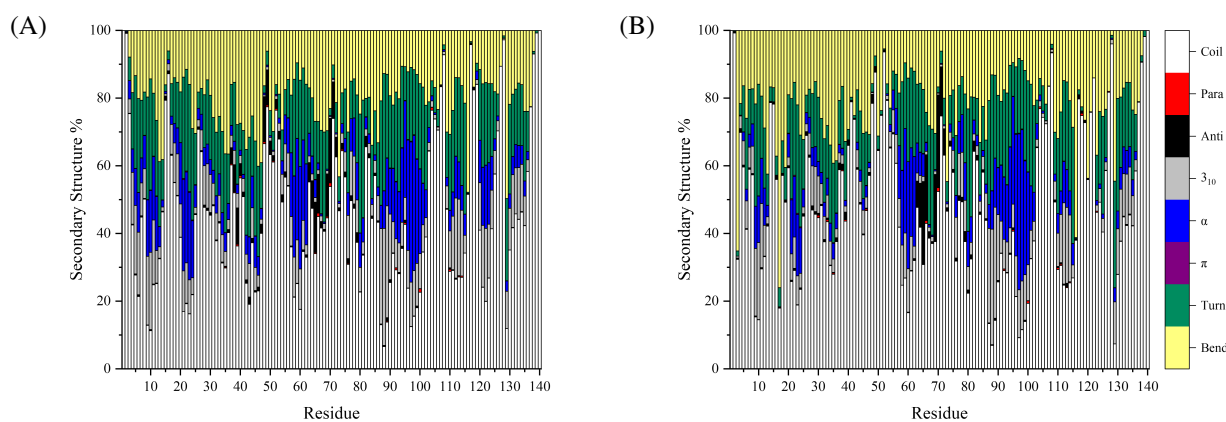
Figure 6.9: Secondary structure distribution per residue for the (A) metal-free and (B) Cu(II)-αS pS129 systems. The β-sheets are denoted with red (parallel) and black (antiparallel), and helices with grey (3_{10}), blue (α) and purple (π).

Looking at the end-to-end distances, generally the changes in the total peptide size are reflected by the end-to-end distance gain/loss, Figure D.2. This is especially true for the metal-free system, where the end-to-end distance increases by 6 Å, in both the A53T and pS129, with the R_g increasing by *ca.* 7 Å in the former and 5 Å in the latter. The Cu(II)-coordinated system appears to follow a similar trend in the A53T form, with an increase of 8 Å translating to a *ca.* 12 Å increase in the R_g. This is not the case, however, in the pS129 system, where despite a 10 Å increase in the R_g, the total end-to-end distance remains the same as the WT form. Even though there is an increase in the N-terminal and NAC regions, a decrease in the end-to-end distance of the C-terminal brings the system back to the value seen in the WT-system. The standard deviation values of these distances, increase the uncertainty of any changes between the systems, with the values for the regions between the different systems, overlapping throughout. Looking at the newly formed hydrogen bonds, Table 6.6, their appearance in the C-terminal is rationalised by the decreased end-to-end distance of that region. The hydrogen bonds that are broken have eased the strain around the Cu(II)-binding site, with the structural geometry now more in line with the bond angles expected in a square planar conformation, Table D.5.

Table 6.6: Newly formed or broken hydrogen bonds in the pS129 systems (maintained for >10% of the simulations). Brackets have been added to the instances where residues form an additional hydrogen bond through different atoms.

System	Acceptor	Donor	Retention (%)
		New Hydrogen Bonds	
free-αS	All newly formed hydrogen bonds exist for <10%		
Cu(II)-αS	Asp119	Val118	12.8
	Glu131	Asn122	11.5 (11.3)
		Broken Hydrogen Bonds	
free-αS	All broken hydrogen bonds exist for <10%		
Cu(II)-αS	Glu35	Val49	28.2
	Glu35	Met1	27.4
	Glu35	His50	21.2
	Glu13	Met1	12.8 (11.0)
	Glu13	His50	11.8 (10.4)
	Glu13	Val52	11.5
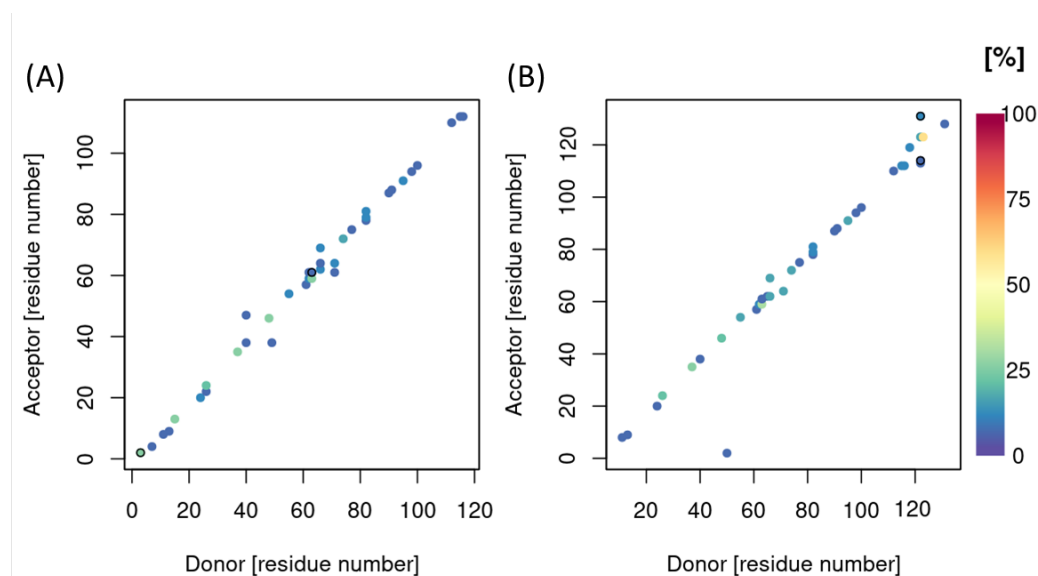
Figure 6.10: Hydrogen bonds in the (A) metal-free and (B) Cu(II)-bound pS129 peptides.

The drop in the C-terminal end-to-end distance of the metal-bound system, indicate a tendency for the residues in this region to 'envelop' more closely around the metal, suggesting the greater submission of the residues to the electrostatic effects from the metal ion. Further evidence comes from the C-terminal binding energy for each of the systems, which was obtained from semiempirical measurements, after increasing the distance between the metal ion and O from the carboxyl group of D121, in V118-E130. Figure 6.11, displays the relative binding energy of the metal ion to this site, where an overall increase in the affinity is observed in the pS129 system. The binding free energy here (including entropy corrections), decrease from -107.85 kcal mol^{-1} to -131.20 kcal mol^{-1}, going from the WT to pS129 system. These results, even though their difference may be within the margin of error, appear to support experimental observations, where a small increase in the binding affinity to the C-terminal is seen, upon phosphorylation of Ser129, from -10.49 kcal mol^{-1} to -10.63 kcal mol^{-1}.[3, 11] The great difference between the binding energies from the semiempirical calculations and the experimental values, was something also observed in Chapter 5; although no further corrections were made to the values calculated from the semiempirical approach here. However, it is unlikely that these would significantly alter the overall trend in the binding energies and the relative increase in the binding affinity of the metal ion to the C-terminal, which admittedly is expected to be larger than what is observed experimentally.
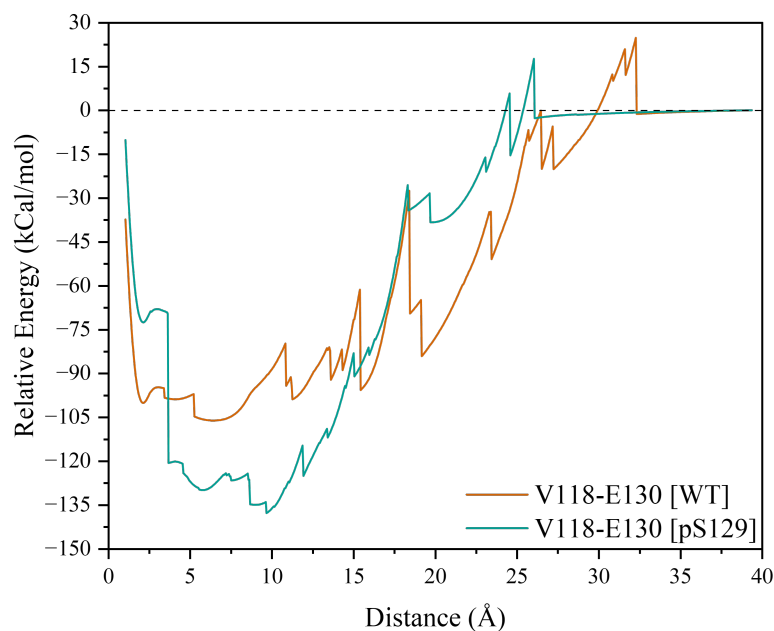
Figure 6.11: Relative binding energy of Cu(II) to the C-terminal binding site of the WT and pS129-αS, calculated on the V118-E130 fragment. The distance is defined between the metal ion and O from the carboxyl group of D121.
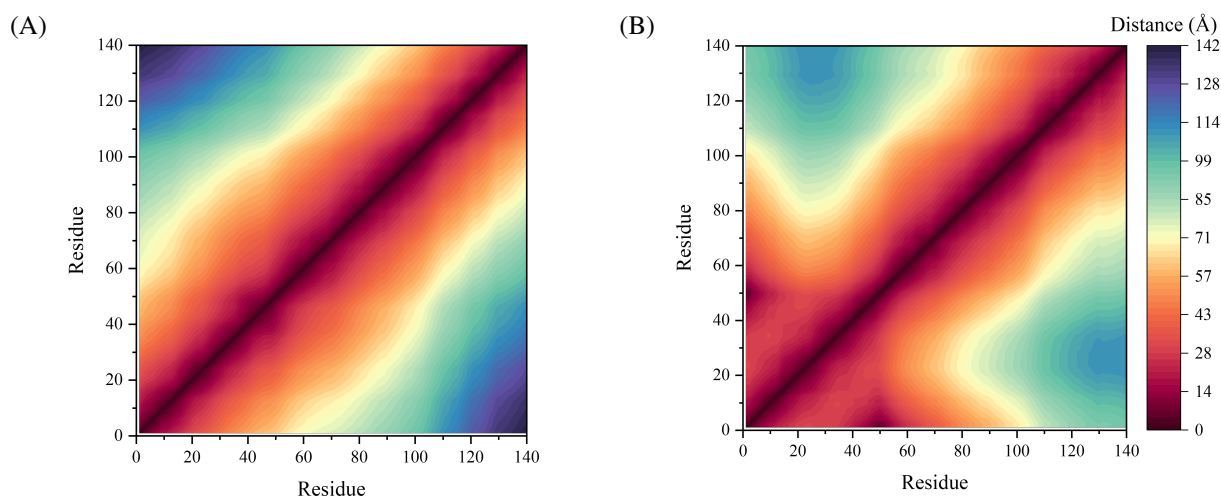
Figure 6.12: Contact maps of the Cα in the (A) metal-free and (B) Cu(II)-bound pS129 systems.

Considering the evidence presented here, phosphorylation of αS at S129 does not alter the conformation and dynamics of the protein in a significant manner to elicit a notable change in fibril formation, with minor reduction in the β-content and overall size of the chains. Nevertheless, an increase in the binding affinity of the C-terminal to Cu(II) is observed, which could potentially lead to stabilisation of the folding structures in that region, also seen by the increase in the % of helix and strand content.

6.4   Conclusions

Post translational modifications and mutations to the WT-αS have long been identified as potential factors in the development of PD. In this chapter, the effects of two such modifications, namely the phosphorylation of S129 and A53T mutation, were examined in the context of the metal-free and Cu(II)-bound αS. These two alterations were selected, as their presence has been recorded in diseased brains, raising interest on their effects in the structure and aggregation properties of the protein. In particular, pS129 has been identified as the primary phosphorylation site, exhibited in the majority of patients with PD,[*3*] while the A53T mutation has been linked with familial cases of PD.[*29*]

Each of the modifications to the WT-αS examined here, introduce their own alterations to the initial state of the system. The most significant of those changes, appear to occur in the A53T Cu(II)-bound system, where the β-hairpin region between residues L38-A(T)53 was found to be maintained for almost the entirety of one of the runs. However, as discussed in the introduction of this thesis, the timescales we observe here are significantly small, when considering the progression of the disease, and the immediate effect these transient conformational changes have on the overall pathology may vary, depending on the auxiliary action from analogous moieties. Significant contribution to the β-strand content in that region, came from residues Y39 and V48, which almost tripled the persistency of β-characteristics observed in the WT-form. The former residue has been reported to greatly contribute in the aggregation capacities of this protein.[*28*] This, along with the expansion of the NAC region, increases the prospect for inter-peptide interactions to form and oligomerisation to occur.

Despite the great expression of the pS129 PTM in patients with PD system, the data presented here suggest a likely decrease in the aggregation capacities of the protein, from a reduced β-sheet content in both the unbound and Cu(II)-bound systems. The C-terminal of the pS129-systems display a reduced size, which has been attributed to an increased affinity for the metal ion in that site, seen from the reduced binding free energy. Alhough, the small differences observed here may also suggest that another method of evaluating the binding affinity of the metal ion to the peptide, may be needed to validate this difference.

Bibliography

(*1*)   Fujiwara, H. et al. (2002). A-Synuclein Is Phosphorylated in Synucleinopathy Lesions. *Nature Cell Biology 4*, 160–164.

(*2*)   Anderson, J. P. et al. (2006). Phosphorylation of Ser-129 is the dominant pathological modification of α-synuclein in familial and sporadic lewy body disease. *Journal of Biological Chemistry 281*, 29739–29752.

(*3*)   Oueslati, A. (2016). Implication of Alpha-Synuclein Phosphorylation at S129 in Synucleinopathies: What Have We Learned in the Last Decade? *Journal of Parkinson's Disease 6*, 39–51.

(*4*)   Uversky, V. N. (2007). Neuropathology, biochemistry, and biophysics of α-synuclein aggregation. *Journal of Neurochemistry 103*, 17–37.

(*5*)   Tenreiro, S., Eckermann, K., and Outeiro, T. F. (2014). Protein phosphorylation in neurodegeneration: Friend or foe? *Frontiers in Molecular Neuroscience 7*, 1–30.

(*6*)   Azeredo da Silveira, S. et al. (2008). Phosphorylation Does Not Prompt, Nor Prevent, the Formation of -synuclein Toxic Species in a Rat Model of Parkinson's Disease. *Human Molecular Genetics 18*, 872–887.

(*7*)   Bell, R., and Vendruscolo, M. (2021). Modulation of the Interactions Between α-Synuclein and Lipid Membranes by Post-translational Modifications. *Frontiers in Neurology 12*, 1–22.

(*8*)   Ghanem, S. S. et al. (2022). α-Synuclein phosphorylation at serine 129 occurs after initial protein deposition and inhibits seeded fibril formation and toxicity. *Proceedings of the National Academy of Sciences of the United States of America 119*, e2109617119.

(*9*)   Smith, W. W. et al. (2005). α-synuclein phosphorylation enhances eosinophilic cytoplasmic inclusion formation in SH-SY5Y cells. *Journal of Neuroscience 25*, 5544–5552.

(*10*) De Ricco, R. et al. (2015). Remote His50 Acts as a Coordination Switch in the High-Affinity N-Terminal Centered Copper(II) Site of α-Synuclein. *Inorganic Chemistry 54*, 4744–4751.

(*11*) Lu, Y., Prudent, M., Fauvet, B., Lashuel, H. A., and Girault, H. H. (2011). Phosphorylation of α-synuclein at Y125 and S129 alters its metal binding properties: Implications for understanding the role of α-synuclein in the pathogenesis of Parkinson's disease and related disorders. *ACS Chemical Neuroscience 2*, 667–675.

(*12*) Binolfi, A. et al. (2006). 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 128*, 9893–9901.

(*13*)   Castillo-Gonzalez, J. A. et al. (2017). Phosphorylated α -Synuclein-Copper Complex Formation in the Pathogenesis of Parkinson's Disease. *Parkinson's Disease 2017*, 1–9.

(*14*)   González, N. et al. (2019). Effects of alpha-synuclein post-translational modifications on metal binding. *Journal of Neurochemistry 150*, 507–521.

(*15*)   Lashuel, H. A. et al. (2002). A-Synuclein, Especially the Parkinson'S Disease-Associated Mutants, Forms Pore-Like Annular and Tubular Protofibrils. *Journal of Molecular Biology 322*, 1089–1102.

(*16*)   Bengoa-Vergniory, N., Roberts, R. F., Wade-Martins, R., and Alegre-Abarrategui, J. (2017). Alpha-synuclein oligomers: a new hope. *Acta Neuropathologica 134*, 819–838.

(*17*)   Giasson, B. I. et al. (2002). Neuronal α-Synucleinopathy with Severe Movement Disorder in Mice Expressing A53T Human α-Synuclein. *Neuron 34*, 521–533.

(*18*)   Karpinar, D. P. et al. (2009). Pre-fibrillar α-synuclein variants with impaired β-structure increase neurotoxicity in Parkinson's disease models. *The EMBO Journal 28*, 3256–3268.

(*19*)   Winner, B. et al. (2011). In vivo demonstration that α-synuclein oligomers are toxic. *Proceedings of the National Academy of Sciences of the United States of America 108*, 4194–4199.

(*20*)   Regensburger, M., Stemick, J., Masliah, E., Kohl, Z., and Winner, B. (2020). Intracellular A53T Mutant α-Synuclein Impairs Adult Hippocampal Newborn Neuron Integration. *Frontiers in Cell and Developmental Biology 8*, 1–9.

(*21*)   Rocha, S., Kumar, R., Horvath, I., and Wittung-Stafshede, P. (2019). Synaptic vesicle mimics affect the aggregation of wild-type and A53T α-synuclein variants differently albeit similar membrane affinity. *Protein Engineering, Design and Selection 32*, 59–66.

(*22*)   Jo, E., McLaurin, J. A., Yip, C. M., St. George-Hyslop, P., and Fraser, P. E. (2000). α-Synuclein membrane interactions and lipid specificity. *Journal of Biological Chemistry 275*, 34328–34334.

(*23*)   Lorentzon, E., Kumar, R., Horvath, I., and Wittung-Stafshede, P. (2020). Differential effects of Cu2+ and Fe3+ ions on in vitro amyloid formation of biologically-relevant α-synuclein variants. *BioMetals 33*, 97–106.

(*24*)   Coskuner, O., and Wise-Scira, O. (2013). Structures and free energy landscapes of the A53T mutant-type α-synuclein protein and impact of A53T mutation on the structures of the wild-type α-synuclein protein with dynamics. *ACS Chemical Neuroscience 4*, 1101–1113.

(*25*)   Losasso, V., Pietropaolo, A., Zannoni, C., Gustincich, S., and Carloni, P. (2011). Structural role of compensatory amino acid replacements in the α-synuclein protein. *Biochemistry 50*, 6994–7001.

(*26*)   Homeyer, N., Horn, A. H. C., Lanig, H., and Sticht, H. (2006). AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *Journal of Molecular Modeling 12*, 281–289.

(*27*)   Steinbrecher, T., Latzer, J., and Case, D. A. (2012). Revised AMBER Parameters for Bioorganic Phosphates. *Journal of Chemical Theory and Computation 8*, 4405–4412.

(*28*)   Buratti, F. A. et al. (2022). Aromaticity at position 39 in α-synuclein: A modulator of amyloid fibril assembly and membrane-bound conformations. *Protein Science 31*, e4360.

(*29*)   Polymeropoulos, M. H. et al. (1997). Mutation in the α-synuclein gene identified in families with Parkinson's disease. *Science 276*, 2045–2047.

Chapter 7

Conclusions and Outlook

Metal ions' interaction with proteins involved in neurodegenerative diseases, has profound effects on the physiological folding and function of such systems. In this thesis, we investigated the effects of divalent metal ions, particularly Cu(II) and Zn(II), on the structure and aggregation of two proteins, α-Synuclein and Amyloid-β, which are associated with Parkinson's and Alzheimer's diseases. These systems are attracting considerable attention due to their propensity to form amyloid fibrils and aggregates that accumulate in the brain of patients. From the research conducted here, we have seen how different approaches, from quantum to molecular mechanics, can be employed towards the study of these interactions, and have identified diverse structural alterations, induced from metal ion coordination, that have the potential to accelerate the rate of aggregation. Understanding these interactions is crucial to uncovering the pathological mechanisms underlying these diseases and developing new therapeutic targets.

Despite the difference in the timescales between the simulated data, and the progression of the diseases in real life, MD simulations offer a way of understanding the mechanisms of folding that may contribute to the long-term progression of these neurodegenerative diseases. The manifestation of symptoms and development of these disorders depends to a significant degree, on the individual and the efficiency of the biological defences against the aggregation of these proteins. The kinetics of the folding processes examined here, assume that the aggregation process is not affected by other factors, such as the presence of other proteins, which is not the case in real life. Therefore, although the timescales of the simulations presented here, are not directly comparable to the progression of the diseases, they can still offer valuable insight into the mechanisms of aggregation, and the structural elements that can contribute to the formation of oligomers.

In this thesis, we begin in Chapter 3 with an evaluation of various widely-used AMBER force fields, in their ability to reproduce the experimental data available for the Amyloid-β system, bound to Zn(II).[1] In that chapter, we see how, not only the force field, but also the water model, as well as the simulation protocol, can have a significant impact on the results obtained. We eventually conclude that the ff14SB force field, in combination with the standard GBSA[2, 3] implicit solvent model, yield a good reproduction of experimental results, through the implementation of an enhanced sampling technique (accelerated MD), that introduces a bias in the system, for when the potential energy is lower than a reference value, calculated from the initial classical MD simulations. The combination of ff14SB with explicit solvents proved to be marginally better in reproducing the experimental data, although the increased computational cost, and the lack of a significant improvement in the results, led to the use of the implicit

solvent model for the remainder of the simulations. For the parameterisation of the metal ion, given the lack of standard force field parameters, QM calculations were performed to obtain bond, angle and charge parameters, in line with the coordination of the ligand. Having obtained these values, the new terms were introduced in the force field definitions, and harmonic constraints applied between the metal ion and the ligating atoms. Results are also presented for a non-bonded model of the metal ion, using a semiempirical approach (GFN2-xTB), which works on the framework of the tight-binding DFT approach, although proved to be less optimal when considering the experimentally-reported coordinations for this complexation.

Shifting the focus to a different system, in Chapter 4, we quickly realise how the same force field, that was previously used to simulate the Amyloid-β system, is not able to reproduce the experimental data available for the WT-α-Synuclein protein, and instead the implementation of a different force field, modified with a scaling parameter to better represent intrinsically disordered proteins, is required. The solvation of the system was also altered here, with the use of the Onufriev-Bashford-Case (OBC) modification to the GBSA solvent, which implement a scaling parameter proportional to the degree of burial for residues inaccessible by the continuum model. The results obtained from the simulations on the unbound and Cu(II)-bound systems, directed us towards a hypothesis that has already started forming in the literature, with regards to the effect of the membrane-binding capabilities of α-Synuclein, on the aggregation of the protein. We see how the binding of Cu(II) to the protein, can induce a conformational change in the N-terminal region, by reducing the helical content.[4] This modification has the potential to decrease the affinity of the protein towards membranes, thus increasing the availability of the monomeric form, and by extension, the rate of secondary nucleation, during the aggregation process. Another structural characteristic that was found to be influenced by the metal ion binding, is the persistence of β-hairpins, which have previously been reported as a key structural feature in the aggregation of α-Synuclein, permitting its interaction with other peptides and the formation of oligomers.[5] Considering the redox activity of metal ions, the effect of the monovalent Cu(I) ion on the aggregation of the protein, was also evaluated. The results from these simulations, showed that the binding of Cu(I) to the protein, replicated features, such as secondary folding structures and R_g, observed in the metal-free peptide. Evaluating the transition between different macrostates, using the R_g as the collective variable, all systems were found to exhibit the fastest transition going from the large to the mid-sized state. In the case of the slowest transition, for the unbound system it was found going from the mid to the short state, while for the two copper-bound systems, this was exhibited when proceeding to the extended size from the mid-sized state, thus suggesting the lack of metal ion decreases the propensity of reaching more compact states.

Chapter 5 introduced a different approach of modelling the metal ion, through the use of a cationic dummy atom model (CaDAM). The suitability of this model was evaluated through conventional and accelerated MD simulations on the monomeric α-Synuclein, through coordination with the binding sites explored in Chapter 4. The results indicated the stable interaction of the CaDAM with these sites, even when a bias was introduced on the potential energy of the system. The model was then used to assess the binding affinity of Cu(II) on the near-range N-terminal binding sites – M1-D2-H_2O and V49-H50-H_2O.[6] This was achieved through the use of adaptive steered MD simulations on 7-residue fragments of each site, as well as semiempirical calculations, using the GFN2-xTB method. The results suggested the lower-energy site to be the one involving H50, with a 15-42 kcal mol^{-1} difference in the binding free energies between the two methods, which has been attributed to the difference in the partial charges, and

the effects of the pulling speed and spring force constant in the steered MD simulations. Having established the suitability of the CaDAM, the dimer systems of α-Synuclein were modelled and simulated, through conventional MD simulations on the unbound and CaDAM-bound dimers. The interactions between the metal ion and the chains were modelled so that residues M1-D2 from one chain interacted with H50 from the other. The primary interactions were found to stabilise dimer formation, through β-sheets between residues L38-A53 and V63-T72 in the metal-free system, and multiple β-sheets between residues M1-V40 in the CaDAM-bound system. The results obtained from these simulations, suggested that the binding of the metal ion to the dimer can act as a stabilising element, maintaining bridging between the two chains, even when β-strand interactions between them were absent.

In Chapter 6, a certain post-translational modification (PTM), along with a widely studied mutation (A53T), were introduced to the WT-α-Synuclein protein, to assess their effect on the aggregation of the protein, when bound to Cu(II). Phosphorylation of S129, has been reported in over 90% of diseased patients, although no clear correlation has been established yet.[7] From the evidence presented here, an increase in the size of the protein, was observed, that was more pronounced in the metal-bound system. This increase in size compared to the WT-αS, suggests that phosphorylation of S129, could allow for inter-peptide interactions to occur more frequently by exposing the central NAC-region, which is known to be involved in the association of the peptide with other monomers. Nevertheless, the decreased population of β-structures could instead point towards a protective effect on the protein, reducing its ability to form stable oligomers, something also observed through an experimental Thioflavin-S binding assay.[8] The changes in the structure of αS from the A53T mutation were also assessed, although no striking changes were observed when compared to the WT-form, other than an extension of the size of the peptides, especially when Cu(II) was coordinated to the system. This has the potential to drive aggregation by exposing the hydrophobic core of the peptide to inter-molecular interactions, permitting the more rapid association with other chains. Perhaps the most significant finding expressed in this mutated system, is the conservation of the L38-A(T)53 β-hairpin for *ca.* 400 ns in the Cu(II)-bound system, which greatly increased the potential for association with other protein chains.

The work presented here has provided a detailed insight into the metal ion-protein interactions of Amyloid-β and α-Synuclein, although there are still many aspects that have not and can be explored further. Work on the simulation of the Amyloid-β system has been presented in the past from ours and other groups, with current efforts to expand upon the knowledge gained from this and past studies looking at other metal ions and coordination sites,[9–11] through changing protonation states and studies on the oligomeric systems. Furthermore, even though the semiempirical approach used here, was not reliable in reproducing the experimental coordination of the metal ion, the use of other semiempirical methods, such as PM6 or PM7 could be explored, although these still come with a high level of uncertainty, requiring refinement with QM methods to achieve reasonable coordinations.[12] The work on the α-Synuclein system, has been focused on the effect of copper ions on the aggregation of the protein, although there are other metal ions that have been reported to be involved in the aggregation process, such as iron and manganese, which have not been studied here and research into their effect on the structural elements of the protein can offer a more complete picture of the underlying mechanics in the aggregation process.[13, 14] The effects of Cu(II) presented here, led to the implication of the membrane-binding capabilities of the protein, on the aggregation process, with the binding of the metal ion believed to reduce the affinity of the protein towards membranes, and

increase the availability of monomers for oligomerisation. This hypothesis has been described in the past from experimental results,[4] although a computational examination of the change in affinity of the protein towards membranes, as a result of the metal ion, can provide valuable insight into the processes driving oligomerisation. Lastly, the work presented here on the dimer systems of α-Synuclein, suggested one of the anchoring characteristics for oligomer formation, exists in the region between residues V63-T72. Although this region has been found to exhibit β-character in previous computational and experimental studies,[15, 16] the incidence of that structure remains relatively unexplored. The evidence presented throughout this thesis suggest an important role for this β-hairpin region that is worth investigating further. Moreover, the CaDAM-free dimer was only simulated with the two chains placed in the same direction, at the starting conformation. While this is a valid model for the oligomerisation of the protein, and we have shown evidence of perpendicular arrangement of the chains, the interactions with the two chains in an opposite-direction starting conformation could also be explored, as this is a structure observed when cross-linking fibrils, but also mirrors the system when copper ions are coordinated with the two chains.[17]

Bibliography

(1)   Zirah, S. et al. (2006). Structural Changes of Region 1-16 of the Alzheimer Disease Amyloid β-Peptide upon Zinc Binding and in Vitro Aging. *Journal of Biological Chemistry 281*, 2151–2161.

(2)   Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters 246*, 122–129.

(3)   Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996). Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *The Journal of Physical Chemistry 100*, 19824–19839.

(4)   Kurochka, A. S., Yushchenko, D. A., Bouř, P., and Shvadchak, V. V. (2021). Influence of lipid membranes on α-synuclein aggregation. *ACS Chemical Neuroscience 12*, 825–830.

(5)   Yu, H., Han, W., Ma, W., and Schulten, K. (2015). Transient β -hairpin formation in α -synuclein monomer revealed by coarse-grained molecular dynamics simulation. *The Journal of Chemical Physics 143*, 243142.

(6)   Valensin, D. et al. (2011). The role of His-50 of α-synuclein in binding Cu(ii): PH dependence, speciation, thermodynamics and structure. *Metallomics 3*, 292–302.

(7)   Oueslati, A. (2016). Implication of Alpha-Synuclein Phosphorylation at S129 in Synucleinopathies: What Have We Learned in the Last Decade? *Journal of Parkinson's Disease 6*, 39–51.

(8)   Ghanem, S. S. et al. (2022). α-Synuclein phosphorylation at serine 129 occurs after initial protein deposition and inhibits seeded fibril formation and toxicity. *Proceedings of the National Academy of Sciences of the United States of America 119*, e2109617119.

(9)   Kennedy-Britten, O. D., Al-Shammari, N., and Platts, J. A. (2020). Molecular dynamics simulations of copper binding to N-terminus mutants of amyloid-β. *Journal of Biomolecular Structure and Dynamics 0*, 1–11.

(10) Al-Shammari, N., Savva, L., Kennedy-Britten, O., and Platts, J. A. (2021). Forcefield evaluation and accelerated molecular dynamics simulation of Zn(II) binding to N-terminus of amyloid-β. *Computational Biology and Chemistry 93*, 107540.

(11) Kennedy-Britten, O. D., Alshammari, N., and Platts, J. A. (2021). Accelerated Molecular Dynamics to Explore the Binding of Transition Metals to Amyloid-β. *ACS Chemical Neuroscience 12*, 4065–4075.

(12) Minenkov, Y., Sharapa, D. I., and Cavallo, L. (2018). Application of Semiempirical Methods to Transition Metal Complexes: Fast Results but Hard-to-Predict Accuracy. *Journal of Chemical Theory and Computation 14*, 3428–3439.

(*13*) Ayton, S., and Lei, P. (2014). Nigral Iron Elevation Is an Invariable Feature of Parkinson's Disease and Is a Sufficient Cause of Neurodegeneration. *BioMed Research International 2014*, 1–9.

(*14*) Xu, C. K. et al. (2022). The Pathological G51D Mutation in Alpha-Synuclein Oligomers Confers Distinct Structural Attributes and Cellular Toxicity. *Molecules 27*, DOI: `10.3 390/molecules27041293`.

(*15*) Li, J., Uversky, V. N., and Fink, A. L. (2001). Effect of familial Parkinson's disease point mutations A30P and A53T on the structural properties, aggregation, and fibrillation of human α-synuclein. *Biochemistry 40*, 11604–11613.

(*16*) Amos, S. B. T. et al. (2021). Membrane Interactions of α-Synuclein Revealed by Multiscale Molecular Dynamics Simulations, Markov State Models, and NMR. *Journal of Physical Chemistry B 125*, 2929–2941.

(*17*) Guerrero-Ferreira, R. et al. (2018). Cryo-EM structure of alpha-synuclein fibrils. *eLife 7*, 1–18.

Appendix A

Chapter 3

Table A.1: Force constants and equilibrium distances of coordinating atoms to the **Zn(II)** metal centre, as calculated from B3LYP/6-31G(d) optimisation of the metal sites.

Ligating atoms	Force constant (kcal mol^{-1}·Å^{-2})	Equilibrium distance (Å)
HIS6(N)-Zn(II)	64.3	2.04
GLU11(O)-Zn(II)	74.9	1.97
HIS13(N)-Zn(II)	69.5	2.03
HIS14(N)-Zn(II)	51.8	2.08

Table A.2: Force constants and equilibrium angles of coordinating atoms to the **Zn(II)** metal centre, as calculated from B3LYP/6-31G(d) optimisation of the metal sites.

Ligating atoms	Force constant (kcal mol^{-1}·Å^{-2})	Equilibrium angle (°)
H6-Zn(II)-E11	39.1	113.08
H6-Zn(II)-H13	40.1	112.57
H6-Zn(II)-H14	29.8	103.60
E11-Zn(II)-H13	45.8	123.17
E11-Zn(II)-H14	52.4	100.07
H13-Zn(II)-H14	28.5	100.27
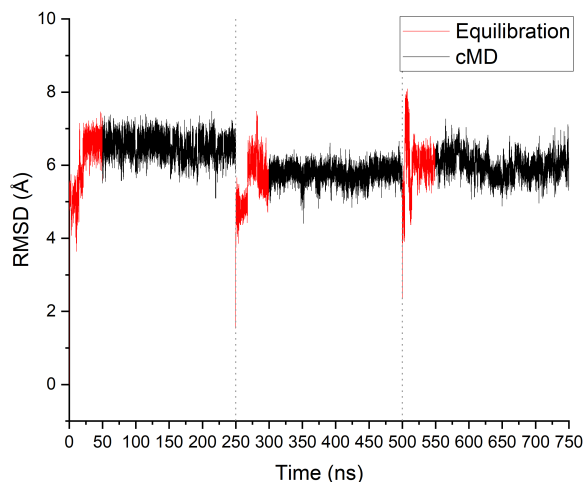
Figure A.1: RMSD plot of cMD using ff14SB/GBSA, showing the steps removed for equilibration (red) and the data used to evaluate the system (black).

Table A.3: Average diagonalized eigenvalues from the gyration tensor coordinates, from the cMD simulations.

Zn(II)-Aβ$_{16}$	λ_x (Å)		λ_y (Å)		λ_z (Å)	
	Avg.	SD	Avg.	SD	Avg.	SD
Experimental[1]	7.27	0.62	16.68	1.61	31.19	4.00
ff14SB - GBSA	7.29	1.34	14.18	1.74	37.84	5.15
ff99SB-ILDN - GBSA	9.19	1.44	16.41	2.59	30.95	4.51
ff14SBonlySC - GBSA	8.05	1.29	16.58	1.34	30.96	3.05
ff14SBonlySC - GB-Neck2	8.00	1.46	18.46	4.29	46.25	7.24
fb15 - GBSA	8.53	1.39	13.73	1.83	38.02	8.21
ff03ws - OBC	9.41	1.22	15.86	2.61	35.22	6.95

Table A.4: Salt Bridge % for each of the systems simulated in implicit solvent.

Salt Bridges in Implicit Solvent	Arg5				Lys16			
	Asp1	**Glu3**	**Asp7**	Glu11	Asp1	**Glu3**	Asp7	Glu11
Experimental	**43**	**10**	**33**	0	5	**48**	0	0
ff14SB (GBSA)	**72**	**49**	**1**	0	9	**18**	14	1
ff14SBonlySC (GBSA)	**52**	**23**	**56**	2	4	**16**	22	3
ff14SBonlySC (GBneck2)	**14**	**3**	**20**	0	1	**2**	2	0
ff99SB-ILDN (GBSA)	**55**	**16**	**29**	0	17	**9**	26	2
fb15 (GBSA)	**29**	**56**	**4**	0	23	**33**	3	2

Table A.4: Salt Bridge % for each of the systems simulated in implicit solvent.

Salt Bridges in Implicit Solvent	Arg5				Lys16			
	Asp1	**Glu3**	**Asp7**	Glu11	Asp1	**Glu3**	Asp7	Glu11
ff03ws (OBC)	**11**	**10**	**14**	0	1	**0**	0	0

Table A.5: Salt Bridge % for each of the systems simulated in explicit solvent.

Salt Bridges in Explicit Solvent	Arg5				Lys16			
	Asp1	**Glu3**	**Asp7**	Glu11	Asp1	**Glu3**	Asp7	Glu11
Experimental	**43**	**10**	**33**	0	5	**48**	0	0
ff14SB (TIP3P)	**52**	**7**	**1**	0	2	**10**	3	0
ff14SB (TIP4P)	**35**	**0**	**29**	0	5	**3**	2	0
ff99SB-ILDN (TIP3P)	**31**	**9**	**52**	0	2	**5**	0	0
ff99SB-ILDN (TIP4P)	**13**	**21**	**35**	0	2	**2**	3	0
fb15 (TIP3P)	**25**	**13**	**47**	0	0	**2**	0	1
fb15 (TIP3P-FB)	**15**	**33**	**1**	0	4	**4**	2	0
ff03ws (TIP3P)	**5**	**7**	**1**	0	0	**0**	0	0
ff03ws (TIP4P/2005)	**4**	**3**	**0**	0	0	**0**	0	0

Table A.6: Salt Bridge % for each of the systems simulated using ff14SB.

Salt Bridges in ff14SB	Arg5				Lys16			
	Asp1	**Glu3**	**Asp7**	Glu11	Asp1	**Glu3**	Asp7	Glu11
Experimental	**43**	**10**	**33**	0	5	**48**	0	0
ff14SB (GBSA) - cMD	**74**	**7**	**12**	0	1	**3**	38	3
ff14SB (GBSA) - aMD	**76**	**11**	**0**	0	1	**9**	36	4
ff14SB (TIP3P) - cMD	**52**	**7**	**1**	0	2	**10**	3	0
ff14SB (TIP3P) - aMD	**13**	**16**	**15**	0	3	**3**	1	0

Table A.7: Average diagonalized eigenvalues from the gyration tensor coordinates, for the aMD simulations using ff14SB.

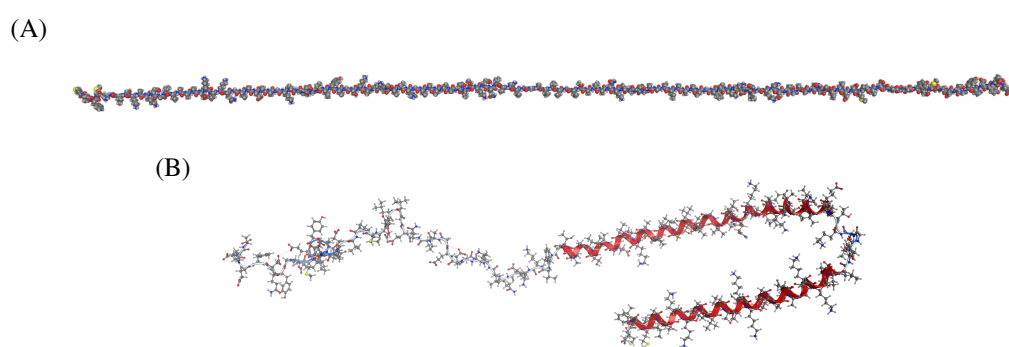| Zn(II)-Aβ$_{16}$	λ_x (Å)		λ_y (Å)		λ_z (Å)	
	Avg.	SD	Avg.	SD	Avg.	SD
Experiment[1]	7.27	0.62	16.68	1.61	31.19	4.00
ff14SB - GBSA - cMD (600 ns)	7.29	1.34	14.18	1.74	37.84	5.15
ff14SB - GBSA - aMD (300 ns)	7.12	1.08	14.80	2.38	37.94	5.45
ff14SB - TIP3P - cMD (1 μs)	8.45	1.46	18.28	3.32	33.16	7.15
ff14SB - TIP3P - aMD (300 ns)	8.71	1.92	18.14	4.09	42.17	11.31

Appendix B

Chapter 4

(A)

(B)

Figure B.1: Starting conformation of the metal-free α-Synuclein, given in its (A) extended and (B) NMR-obtained forms.

Table B.1: Predicted Cα chemical shift values from SPARTA+.[2]

Residue	ff03ws(OBC) cMD 300 ns	ff14SB(GBSA) cMD 300 ns	ff03ws(TIP4P/2005) cMD 300 ns	ff03ws(OBC) aMD 1.8 μs
1	-	-	-	-
2	53.684	53.589	54.243	53.414
3	61.381	62.637	61.894	63.544
4	58.542	60.972	59.216	60.785
5	56.334	57.745	54.363	56.992
6	56.149	58.869	55.881	57.355
7	45.387	46.091	44.886	44.490
8	53.413	54.556	54.075	54.626
9	60.429	59.671	58.211	57.722
10	56.837	59.023	58.437	56.165
11	51.255	53.983	51.106	52.570
12	55.497	58.580	56.354	56.930

Residue	ff03ws(OBC) cMD 300 ns	ff14SB(GBSA) cMD 300 ns	ff03ws(TIP4P/2005) cMD 300 ns	ff03ws(OBC) aMD 1.8 μs
13	55.867	57.104	57.327	55.529
14	44.905	45.298	44.458	45.668
15	62.242	60.303	61.384	61.498
16	62.525	64.603	62.137	61.332
17	51.526	53.772	51.772	51.470
18	51.774	52.602	53.514	52.390
19	51.690	54.712	51.032	52.191
20	55.481	57.502	56.265	55.412
21	56.044	57.449	56.379	58.231
22	62.275	63.539	62.713	64.581
23	56.462	56.979	56.829	56.127
24	54.742	55.153	55.088	54.822
25	45.706	44.892	45.462	45.956
26	60.733	63.211	61.164	61.474
27	53.424	52.534	53.369	52.378
28	57.776	55.999	55.941	57.528
29	51.348	53.032	51.887	51.033
30	50.762	52.306	52.298	52.007
31	45.356	45.002	45.151	44.999
32	56.335	59.440	57.226	55.900
33	61.761	64.170	60.319	62.701
34	59.569	57.446	56.951	55.675
35	58.530	57.837	55.977	55.564
36	46.431	45.143	44.392	45.202
37	65.728	60.578	59.959	61.864
38	55.287	56.306	54.484	56.370
39	59.160	58.092	58.537	58.098
40	61.058	64.005	61.010	61.347
41	46.180	44.752	45.338	47.202
42	59.962	56.913	59.870	59.816
43	56.435	59.175	56.446	57.340
44	60.621	64.283	63.756	62.334
45	58.306	59.107	56.784	56.145
46	56.815	56.355	56.566	56.834
47	45.099	44.299	44.768	44.782

Residue	ff03ws(OBC) cMD 300 ns	ff14SB(GBSA) cMD 300 ns	ff03ws(TIP4P/2005) cMD 300 ns	ff03ws(OBC) aMD 1.8 μs
48	61.602	60.103	62.728	61.054
49	60.144	64.731	60.901	62.006
50	53.780	56.772	56.759	56.926
51	44.565	45.156	43.780	45.259
52	61.611	60.871	60.353	61.568
53	51.552	54.805	50.554	51.749
54	63.210	64.758	62.928	61.394
55	62.475	65.622	63.343	62.053
56	52.962	54.585	51.663	52.109
57	56.133	57.698	55.402	56.169
58	55.699	59.665	56.785	56.773
59	60.578	64.682	60.980	61.591
60	58.006	58.176	56.310	55.641
61	57.006	56.778	56.809	55.677
62	56.404	56.230	57.415	55.952
63	65.008	64.073	59.961	59.758
64	62.569	65.201	60.680	61.625
65	53.939	55.526	54.146	52.053
66	64.131	63.539	61.673	61.655
67	44.429	44.449	45.688	45.599
68	44.279	45.370	44.772	44.923
69	51.971	51.564	50.745	51.707
70	62.084	64.037	62.467	63.301
71	61.685	64.211	59.751	59.756
72	62.251	63.340	63.102	61.294
73	45.077	45.177	44.721	45.326
74	62.030	63.422	62.185	61.510
75	60.397	60.219	63.287	60.627
76	54.743	54.520	52.669	52.142
77	63.082	63.770	62.683	62.481
78	54.820	51.463	52.757	51.782
79	58.731	56.039	55.001	57.112
80	58.069	56.276	56.275	56.301
81	65.323	60.706	62.254	60.982
82	65.499	63.682	61.522	60.780

Residue	ff03ws(OBC) cMD 300 ns	ff14SB(GBSA) cMD 300 ns	ff03ws(TIP4P/2005) cMD 300 ns	ff03ws(OBC) aMD 1.8 μs
83	59.198	58.919	55.477	56.277
84	46.106	46.219	45.258	44.915
85	52.500	52.120	52.255	51.645
86	44.873	44.397	45.560	45.177
87	58.163	60.807	60.185	60.701
88	61.310	61.435	62.949	61.572
89	52.642	53.722	54.659	53.225
90	51.335	51.590	53.674	51.621
91	54.087	53.818	54.231	52.611
92	65.867	63.084	62.827	62.792
93	45.935	45.681	45.449	44.955
94	59.370	58.362	58.992	58.208
95	61.458	66.718	64.508	61.204
96	55.863	59.175	58.371	56.112
97	59.031	59.305	58.261	57.845
98	56.240	56.663	55.101	56.405
99	58.419	56.855	58.106	56.017
100	57.063	55.568	55.675	54.864
101	46.778	45.663	46.261	44.217
102	57.340	57.849	57.004	55.887
103	53.344	54.957	53.016	54.002
104	55.868	57.973	55.185	55.766
105	58.490	56.607	56.091	55.896
106	44.579	45.715	44.581	44.874
107	50.505	49.753	50.763	51.191
108	62.999	62.542	62.511	63.184
109	55.286	58.984	56.733	56.113
110	59.092	58.838	56.627	56.077
111	45.974	46.657	44.867	44.994
112	61.566	62.943	60.184	58.887
113	56.895	54.501	54.706	56.488
114	56.466	58.275	54.922	57.336
115	54.855	54.389	54.254	54.524
116	52.789	54.370	52.016	54.094
117	63.031	62.412	62.660	63.488

Residue	ff03ws(OBC) cMD 300 ns	ff14SB(GBSA) cMD 300 ns	ff03ws(TIP4P/2005) cMD 300 ns	ff03ws(OBC) aMD 1.8 μs
118	61.651	61.511	63.619	60.988
119	53.849	52.724	52.092	52.273
120	63.260	62.417	65.590	64.841
121	54.130	55.068	56.246	56.484
122	54.264	52.728	55.495	55.165
123	56.516	58.917	56.382	58.847
124	51.364	54.020	52.362	51.572
125	60.128	60.277	58.384	55.941
126	55.926	56.572	54.712	57.364
127	53.733	54.814	53.486	52.695
128	62.593	62.852	62.678	62.339
129	59.276	60.090	59.332	58.203
130	56.012	59.032	55.544	58.242
131	55.937	58.131	56.558	55.891
132	46.259	46.213	44.888	45.359
133	59.699	56.369	56.760	56.862
134	58.486	56.535	55.164	55.715
135	55.899	52.899	54.324	54.362
136	59.005	57.277	57.467	57.364
137	53.647	54.584	55.572	52.219
138	62.938	62.726	63.067	63.056
139	55.819	56.803	56.198	56.810
140	-	-	-	-

Table B.2: Percentage error for the three main regions of the peptide, along with total mean % error, with respect to Cα chemical shift values per residue reported in the literature.[3] Values given in %.

Statistics		ff03ws(OBC) cMD 300 ns	ff14SB(GBSA) cMD 300 ns	ff03ws(TIP4P/2005) cMD 300 ns	ff03ws(OBC) aMD 1.8 μs
	Total	**1.97**	**2.62**	**1.58**	**1.42**
Mean %	N-term	1.73	2.74	1.43	1.35
Error	NAC	2.35	2.54	1.92	1.26
	C-term	1.98	2.52	1.50	1.63

Table B.2: Percentage error for the three main regions of the peptide, along with total mean % error, with respect to Cα chemical shift values per residue reported in the literature.[3] Values given in %.

Statistics		ff03ws(OBC) cMD 300 ns	ff14SB(GBSA) cMD 300 ns	ff03ws(TIP4P/2005) cMD 300 ns	ff03ws(OBC) aMD 1.8 μs
Max % Error	**Total**	**6.68**	**8.19**	**4.87**	**5.50**
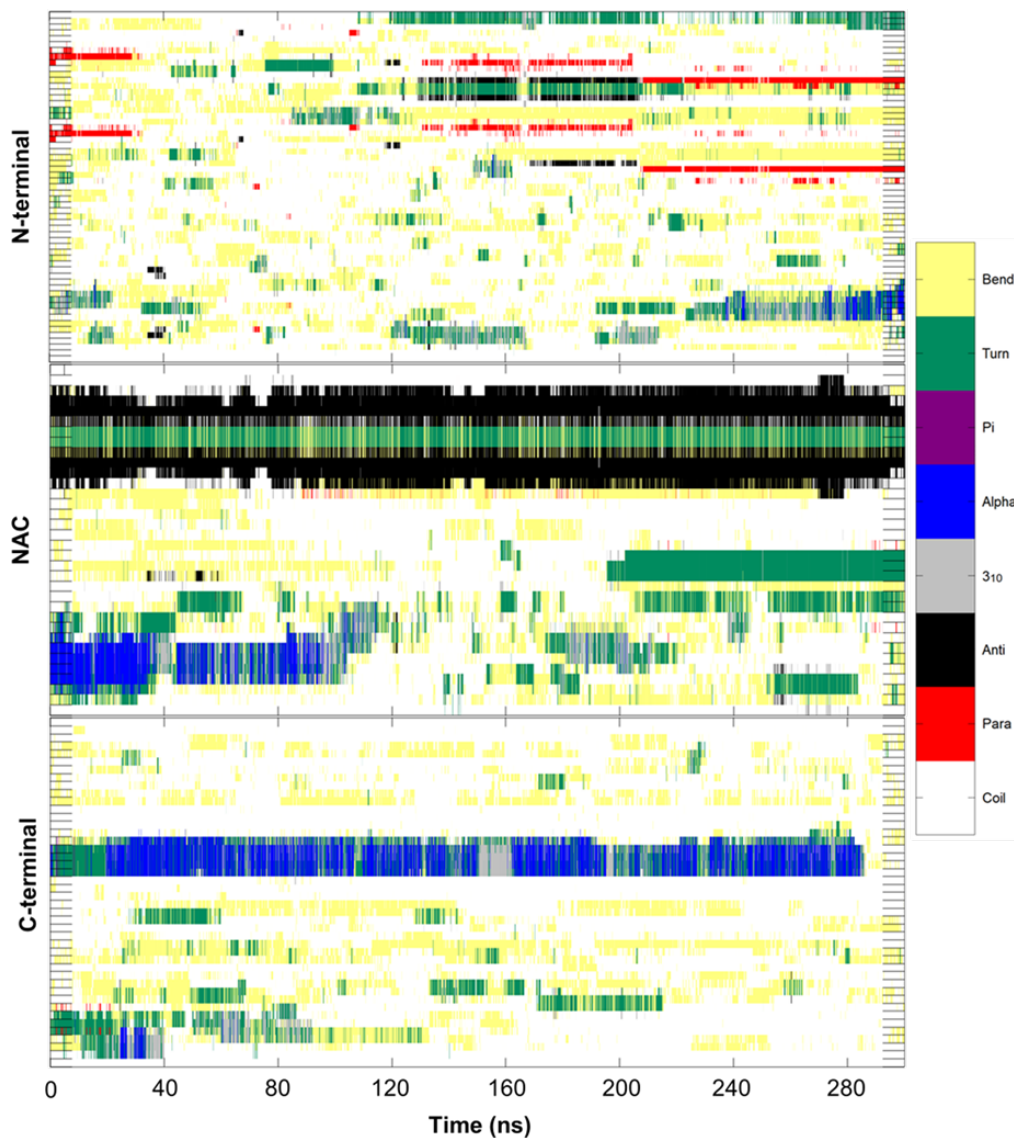
Figure B.2: Evolution of secondary structural elements of each of the residues in the unbound peptide, over the length of the conventional MD trajectory, using explicit solvent.
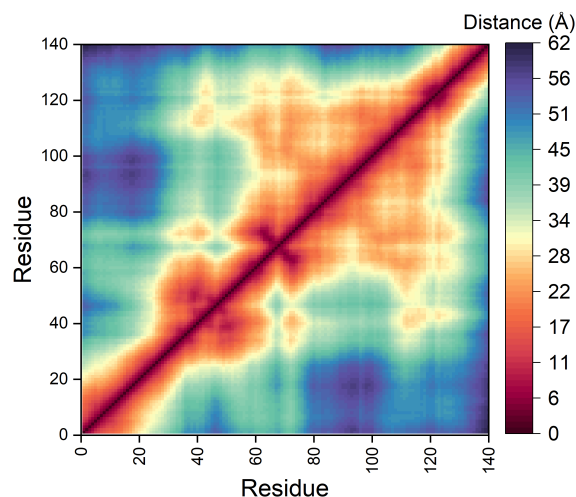
Figure B.3: Contact maps of the alpha-C from the conventional MD simulations in explicit solvent of the free peptide.
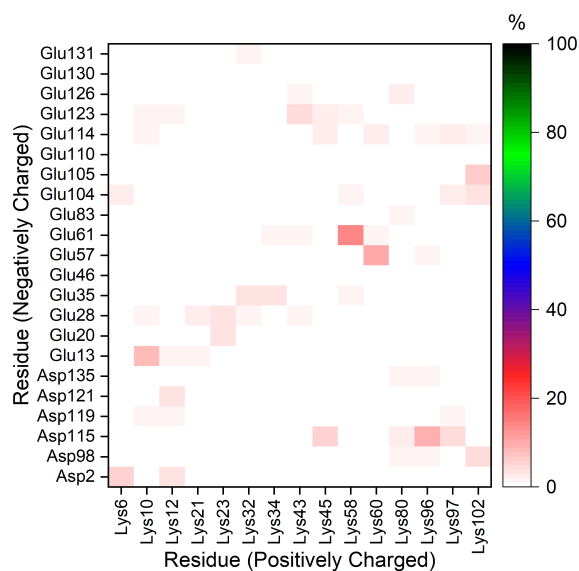
Figure B.4: Salt bridges formed between negatively and positively charged residues from simulations on the free-αS, in explicit solvent.

Table B.3: Detailed secondary structure percentages of the three main regions of αS, after cMD simulations using ff03ws with TIP4P/2005 solvent model.

Region	β-character (%)	α-character (%)	Other (%)
N-terminal	3.59	3.31	93.10
NAC	22.74	4.74	72.52
C-terminal	1.20	9.41	89.39
Total	6.84	5.79	87.36

Table B.4: Angles between the **Cu(II)** and ligating atoms, including force constants, calculated using B3LYP/6-31G(d).

Metal site	Ligating atoms	Force constant (kcal mol^{-1}·rad^{-2})	Angle (°)
N-terminal	MET1 (N) – Cu(II) – ASP2 (N)	89.43	101.79
	MET1 (N) – Cu(II) – ASP2 (O)	79.24	142.50
	MET1 (N) – Cu(II) – HIS50 (N)	62.51	107.33
	ASP2 (N) – Cu(II) – ASP2 (O)	67.53	40.73
	ASP2 (N) – Cu(II) – HIS50 (N)	66.23	149.43
	ASP2 (O) – Cu(II) – HIS50 (N)	54.24	109.55
C-terminal	ASP119 (O) – Cu(II) – ASP121 (O)	55.90	95.33
	ASP119 (O) – Cu(II) – ASN122 (O)	33.94	103.83
	ASP119 (O) – Cu(II) – GLU123 (O)	82.54	152.96
	ASP121 (O) – Cu(II) – ASN122 (O)	26.74	120.78
	ASP121 (O) – Cu(II) – GLU123 (O)	61.37	98.35
	ASN122 (O) – Cu(II) – GLU123 (O)	26.76	88.99

Table B.5: Angles between the **Cu(I)** and ligating atoms, including force constants, calculated using B3LYP/6-31G(d).

Metal site	Ligating atoms	Force constant (kcal mol^{-1}·rad^{-2})	Angle (°)
N-terminal	MET1 (S) – Cu(I) – MET5 (S)	35.36	172.86
C-terminal	MET116 (S) – Cu(I) – MET127 (S)	42.90	166.76

Table B.6: Secondary structure percentages for the free and Cu(II)-bound peptide c/aMD simulations.

Forcefield	β character (%)	α character (%)	Other (%)
free-αS (cMD)	6.63	22.59	70.79
free-αS (aMD)	2.47	16.55	80.98
Cu(II)-αS (cMD)	7.66	18.18	74.16
Cu(II)-αS (aMD)	2.23	12.43	85.33

Table B.7: Globularity values, expressing how spherical (1) or linear (0) the different systems are.

Globularity	Average	SD	Max	Min
free-αS (aMD)	0.0612	0.0433	0.513	0.00493
Cu(II)-αS (aMD)	0.0635	0.0386	0.432	0.00704
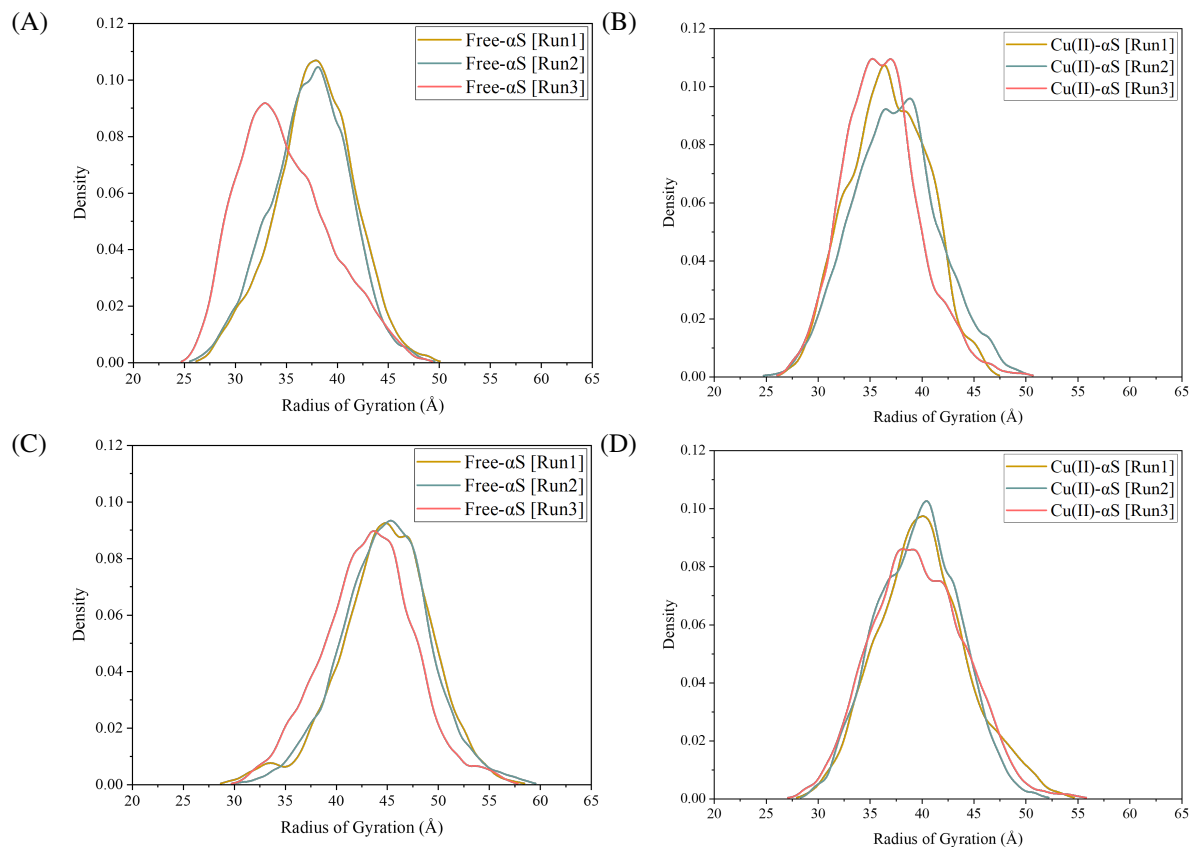
Figure B.5: Distribution of the R_g values from each of the three individual cMD simulations, of the (A) metal-free and (B) Cu(II)-bound αS, and respectively (C – D) for the aMD simulations.
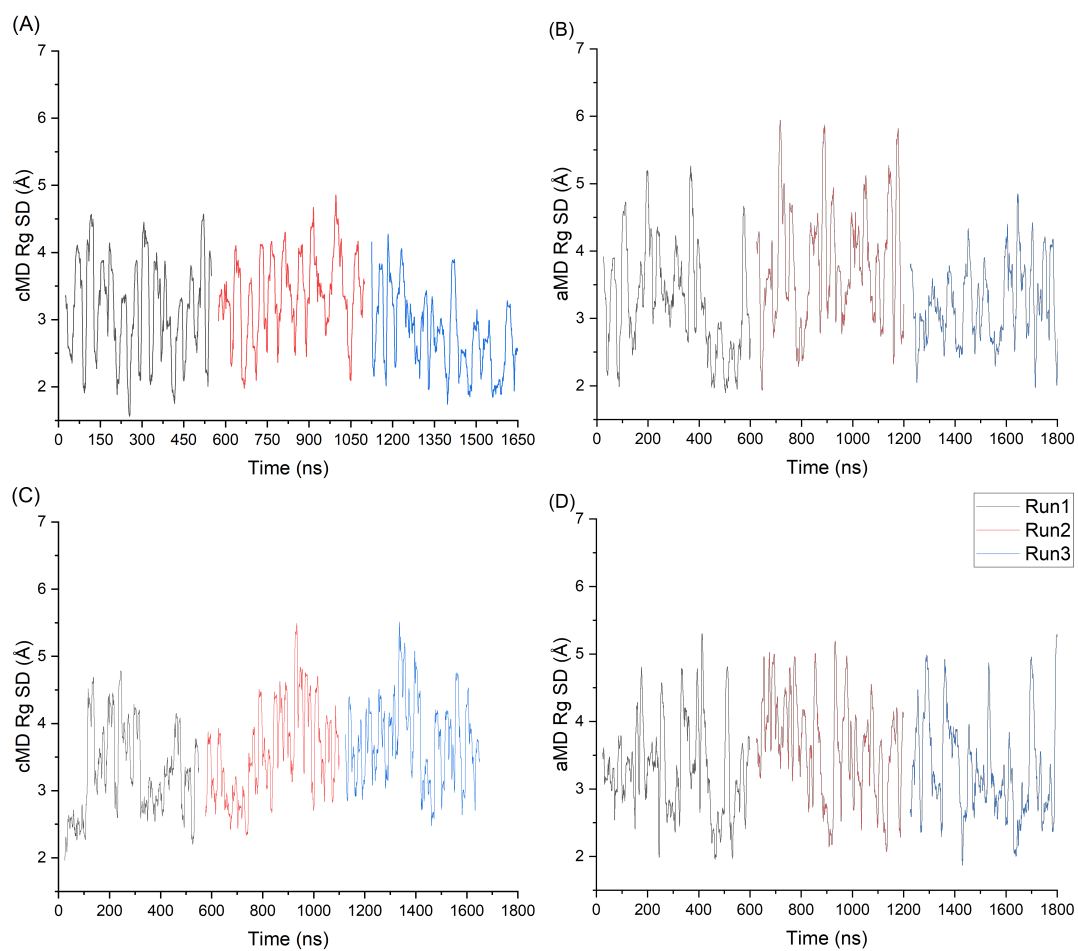
Figure B.6: Rolling standard deviation (25 ns window) from the R_g data, of the three (A) conventional and (B) accelerated MD runs on the free, and respectively (C – D) copper-bound peptide.
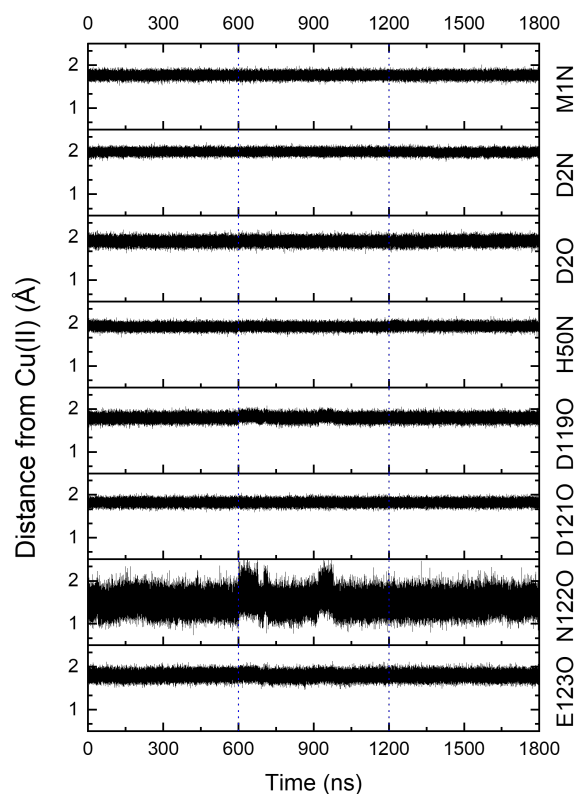
Figure B.7: Distances between the atoms bonded to Cu(II), in the two binding sites, from the accelerated MD simulations in implicit solvent.

Table B.8: Mean distances between atoms in the Cu(II) binding sites.

	Avg. Distance (Å)	SD (Å)
M1N-Cu(II)	1.766	0.053
D2N-Cu(II)	1.986	0.049
D2O-Cu(II)	1.904	0.064
H50N-Cu(II)	1.923	0.053
D119O-Cu(II)	1.798	0.064
D121O-Cu(II)	1.822	0.054
N122O-Cu(II)	1.496	0.196
E123O-Cu(II)	1.797	0.076

Table B.9: Angles between atoms in the Cu(II) binding sites.

	Avg. Angle (°)	SD (°)
M1N-Cu(II)-D2N	90.0	2.5
M1N-Cu(II)-H50N	100.0	5.7

	Avg. Angle (°)	SD (°)
D2O-Cu(II)-H50N	101.5	6.3
D2N-Cu(II)-D2O	60.0	2.6
D119O-Cu(II)-D121O	84.5	4.6
D121O-Cu(II)-N122O	106.0	5.3
N122O-Cu(II)-E123O	65.1	5.7
D119O-Cu(II)-E123O	149.6	3.5

Table B.10: Mean distances between atoms in the Cu(I) binding sites.

	Avg. Distance (Å)	SD (Å)
M1S-Cu(I)	2.191	0.066
M5S-Cu(I)	2.204	0.063
M116S-Cu(I)	2.168	0.061
M127S-Cu(I)	2.172	0.061

Table B.11: Angles between atoms in the Cu(I) binding sites.

	Avg. Angle (°)	SD (°)
M1S-Cu(I)-M5S	168.8	4.7
D116S-Cu(I)-M127S	164.2	4.9

Table B.12: Clusters from the complete aMD trajectory of the Cu(II)-bound αS, created using PCA of the cartesian coordinates of Cα.

Cluster#	Frames in Cluster (out of the total 180,000)
1	13,657
2	6,136
3	8,398
4	850
5	166
6	125
7-18	<100
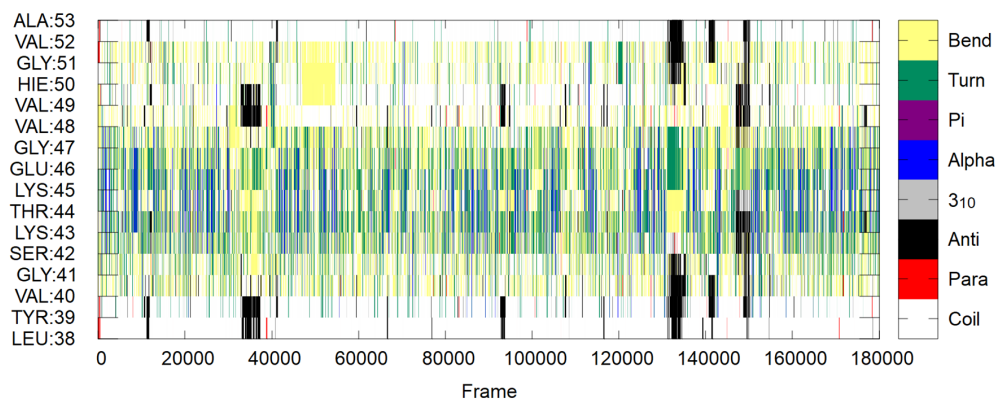
Figure B.8: Evolution of secondary structural elements of each of the residues (L38-A53) in the primary hairpin region, found within the **free peptide** over the length of the accelerated MD trajectory, using implicit solvent. The presence of β-sheets (black/red) with turn/bend (yellow/green) between indicate possible presence of β-hairpin.
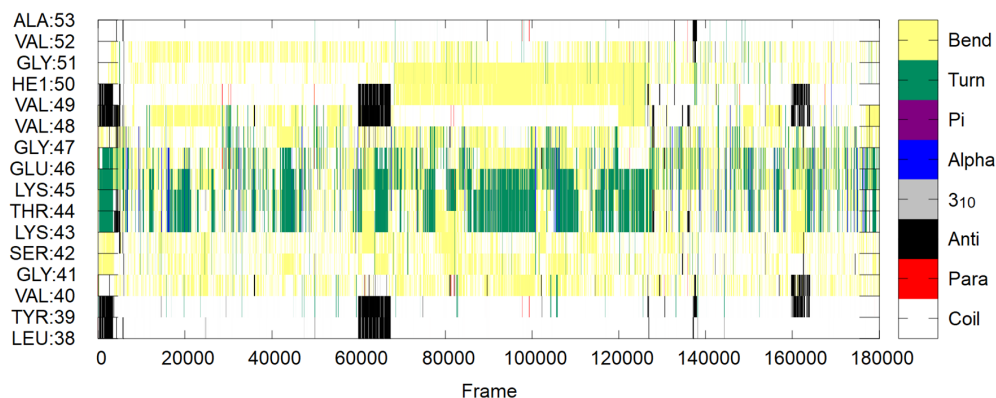
Figure B.9: Evolution of secondary structural elements of each of the residues (L38-A53) in the primary hairpin region, found within the **Cu(II)-bound peptide** over the length of the accelerated MD trajectory, using implicit solvent. The presence of β-sheets (black/red) with turn/bend (yellow/green) between indicate possible presence of β-hairpin.
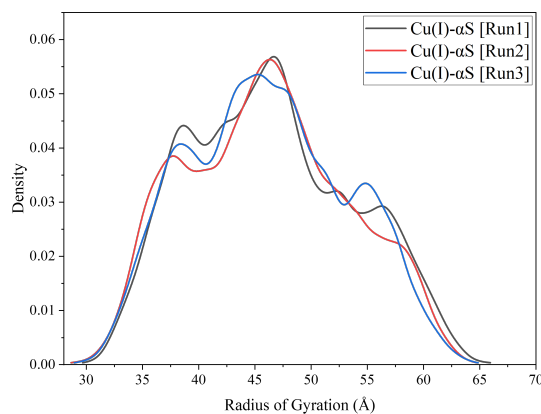
Figure B.10: Distribution of the R_g values from each of the three individual aMD simulations on Cu(I)-αS.
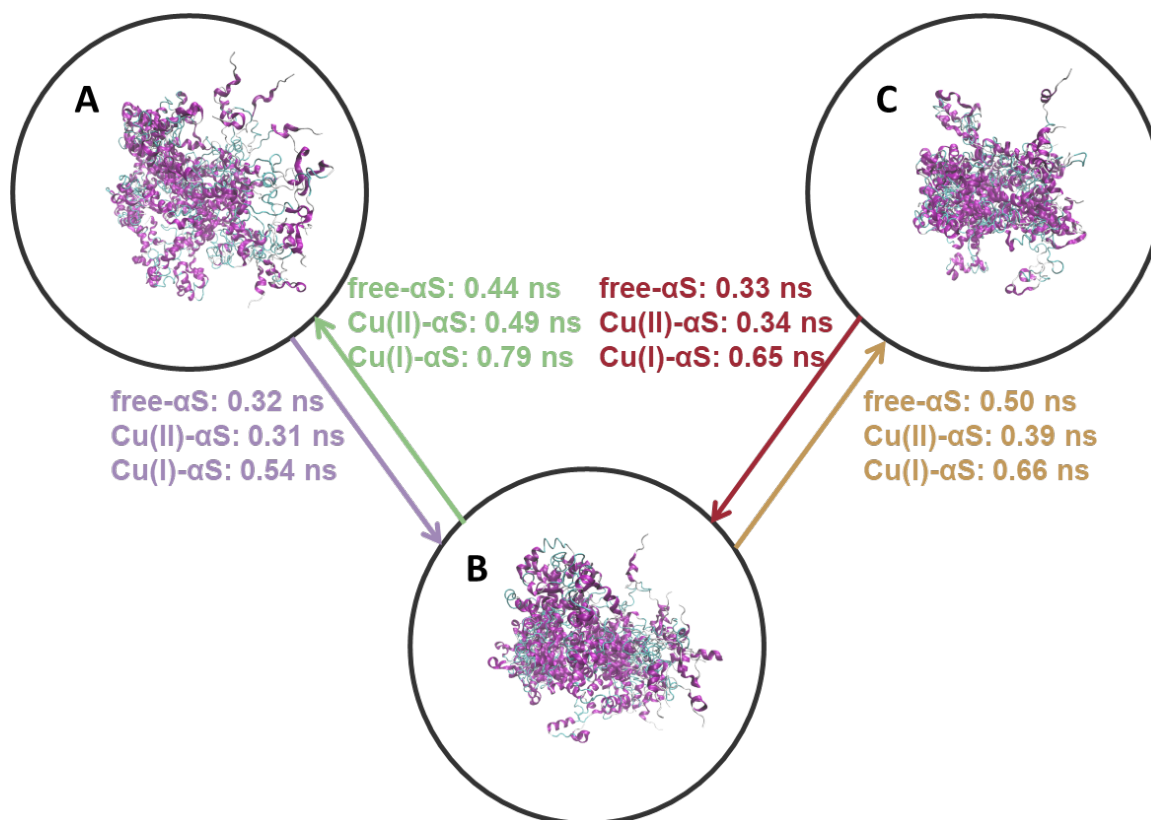
Figure B.11: Transition times between the three macrostates of the MSMs, from (A) large to (B) medium and (C) small R_g. The peptide representations are from the three macrostates of the Cu(I) system.

Table B.13: Difference in R_g between the macrostates – from (A) high to (B) mid, and (C) low to (B) mid compaction. Values given in Å.

System	A → B	C → B
Free-αS	-4.31	+3.00
Cu(II)-αS	-4.97	+2.90
Cu(I)-αS	-6.39	+9.18

Accelerated MD Simulations of N-terminal Cu^{II} – αS

In order to examine the system where a single copper ion interacts with the peptide, the analysis presented below is from three accelerated MD simulations of 600 ns, on the system where only the macro-chelate N-terminal binding site was occupied. As discussed in the main text, the most significant finding is the increased stability of the β-hairpin region between residues L38-A53, to 90 and 93 ns, for two of the three trajectories, Figure B.12(A). This is not the case for the secondary β-hairpin region, between residues V63-T72, Figure B.12(B), where antiparallel β-strands are only maintained for as long as 17.5 ns. The secondary structures, Table B.14, indicate a similarly reduced binding affinity to membranes, as that seen in the system with the two binding sites (9.73%, Table 4.11), while the radius of gyration now more closely resembles that of the free peptide (44.26±4.58 Å), at 43.68±5.73 Å, suggesting the increased compactness in the system with both binding sites occupied, is influenced by the C-terminal coordination, despite its near-range nature.

Table B.14: Detailed secondary structure percentages of the three main regions of N-terminal bound Cu(II).

Region	β character (%)	α character (%)	Other (%)
N-terminal	1.40	11.03	87.58
NAC	2.68	15.77	81.55
C-terminal	0.14	15.70	84.16
Total	1.31	13.71	84.97

(A)
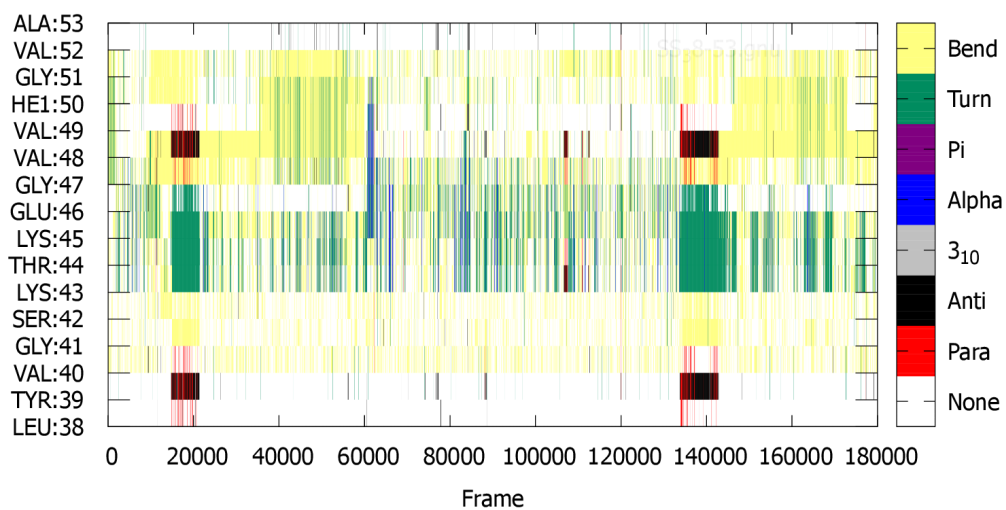
(B)
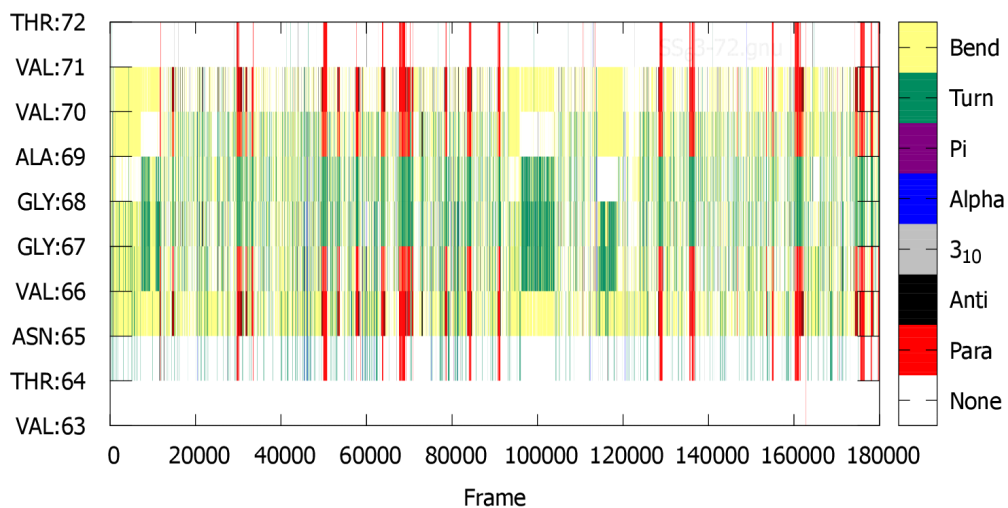
Figure B.12: Evolution of secondary structural elements of each of residues (A) L38-A53, and (B) V63-T72, involved in the β-hairpin regions, for the macro-chelate N-terminal bound Cu(II).

Appendix C

Chapter 5
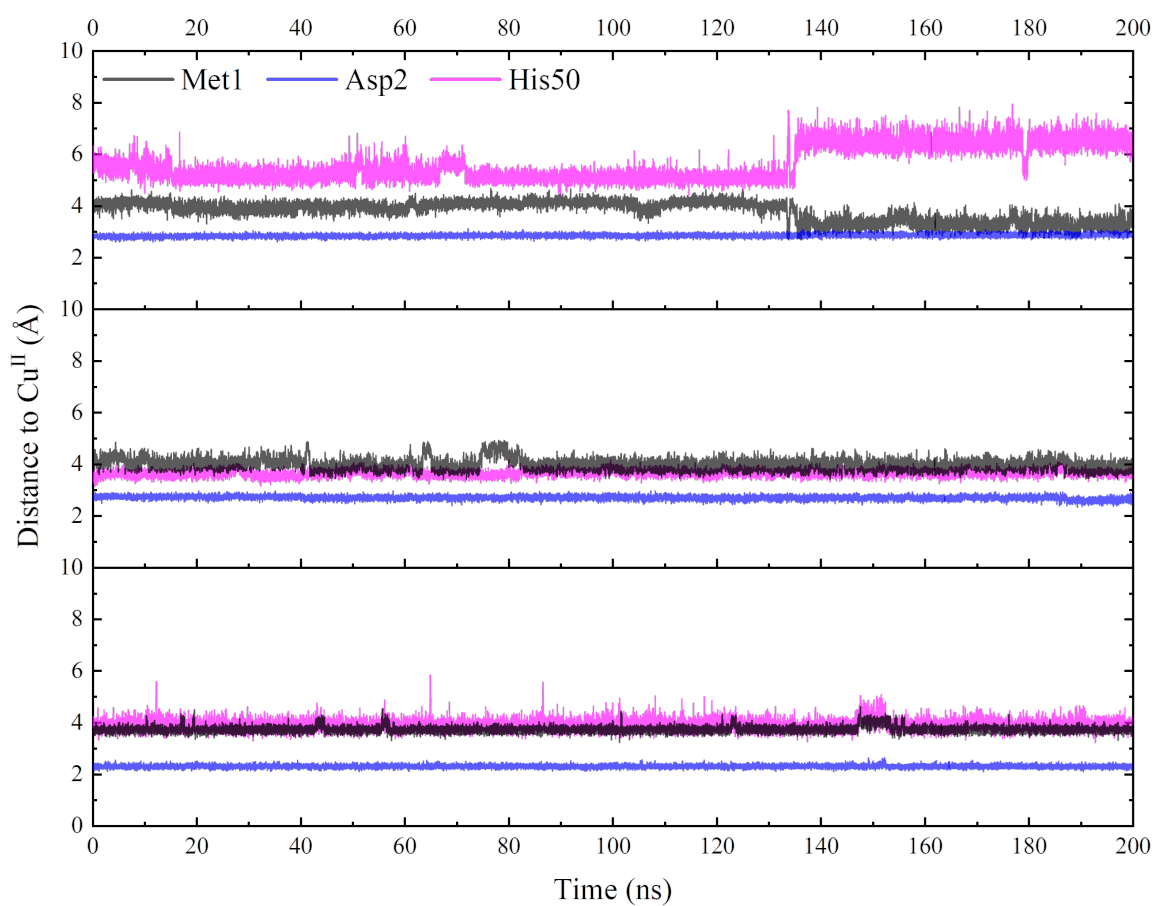
Figure C.1: Plot of the distance between the copper CaDAM and residues involved in the N-terminal macro-chelated binding site, going from the first (top) to the third (bottom) conventional MD run.
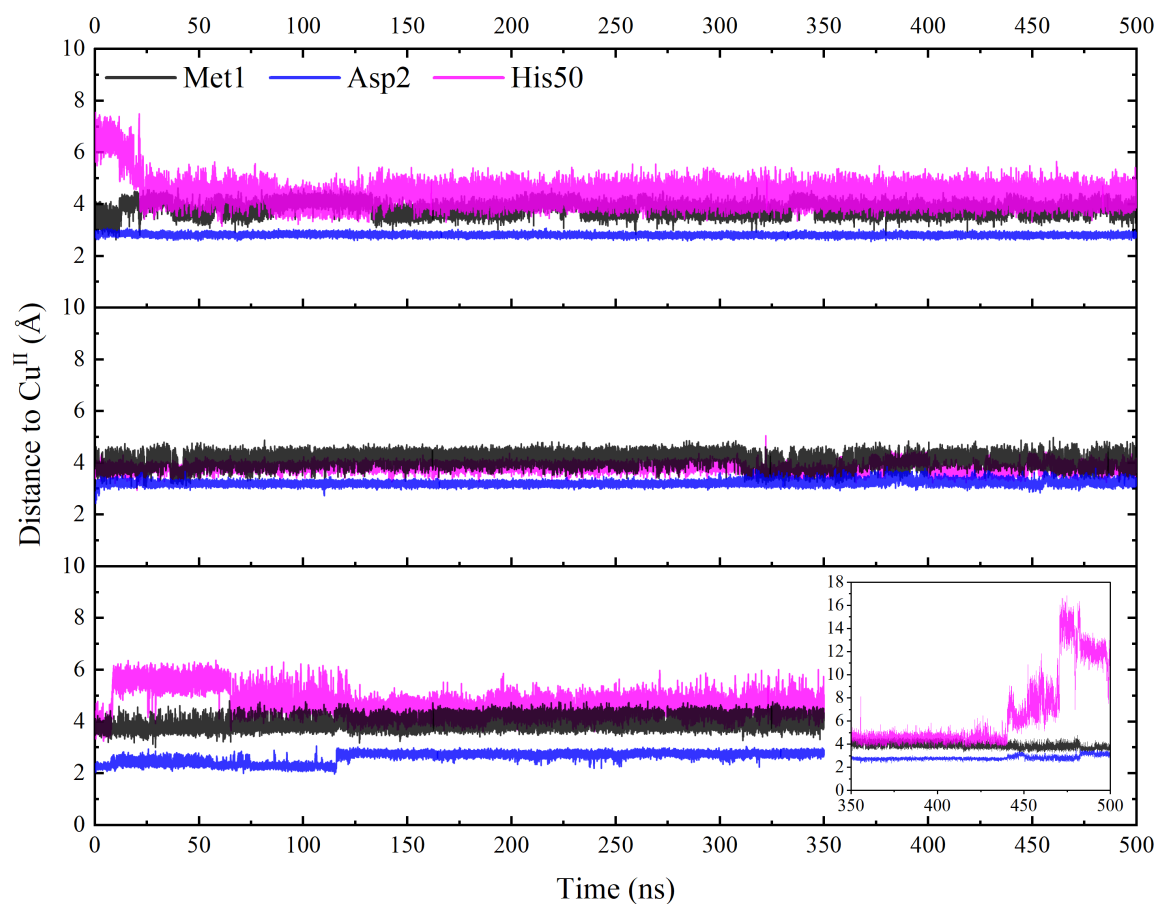
Figure C.2: Plot of the distance between the copper CaDAM and residues involved in the N-terminal macro-chelated binding site, going from the first (top) to the third (bottom) accelerated MD run. The data on the third run, after 350 ns, is displayed on the overlayed plot, to maintain a legible scale for the rest of the values.

Table C.1: Lowest energy conformation from the ASMD simulations, for the G47-A53 binding site.

G47 – A53 (ASMD)

Distance: 5.94 Å

Relative Free Energy: -102.45 kcal mol^{-1}
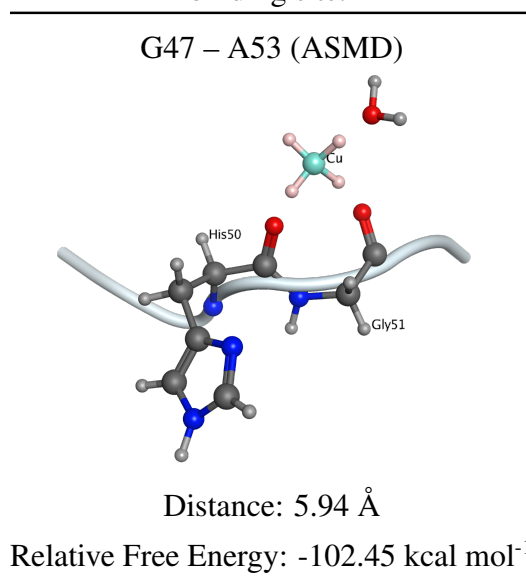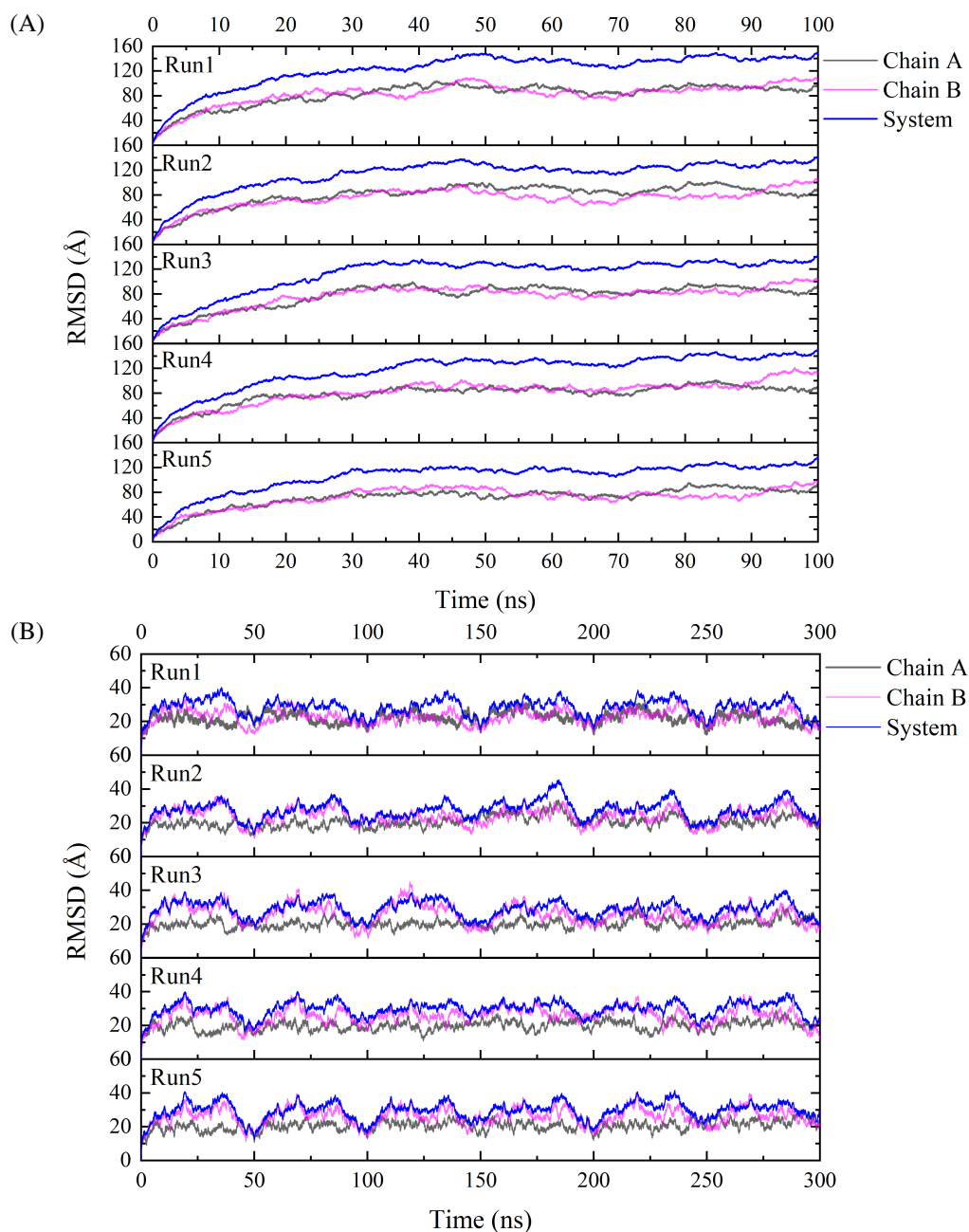
Figure C.3: Plots of the RMSD in each of the runs, for the steps involved in (A) the equilibration of the CaDAM-bound dimer system, and (B) the data used to assess its behaviour. The RMSD of chain A (black) and chain B (pink), is plotted against the system's overall change (blue). The values are calculated with respect to the first frame of each run.
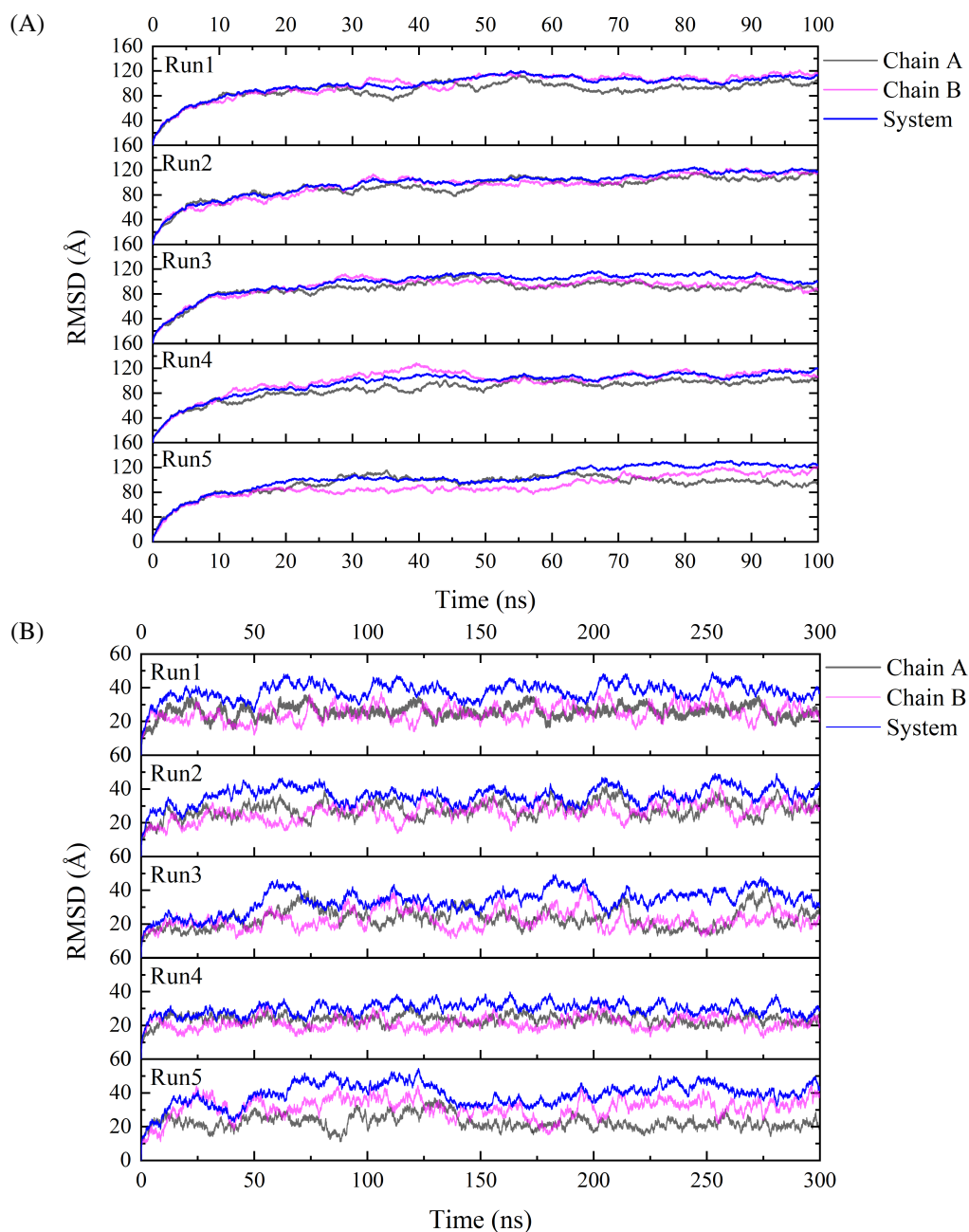
Figure C.4: Plots of the RMSD in each of the runs, for the steps involved in (A) the equilibration of the metal-free dimer system, and (B) the data used to assess its behaviour. The values are calculated with respect to the first frame of each run.
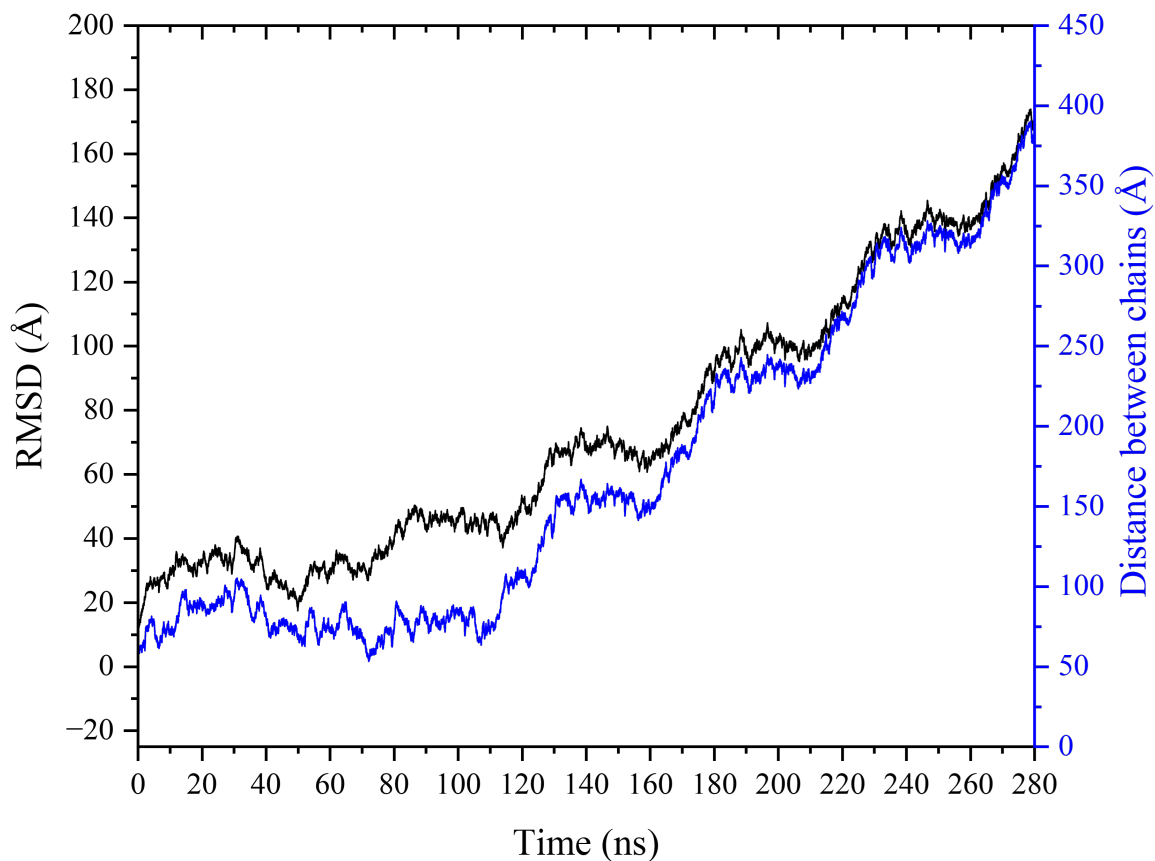
Figure C.5: Plot of RMSD (black) and distance between the centre of mass of the two chains (blue) in the metal-free system, from aMD simulations.
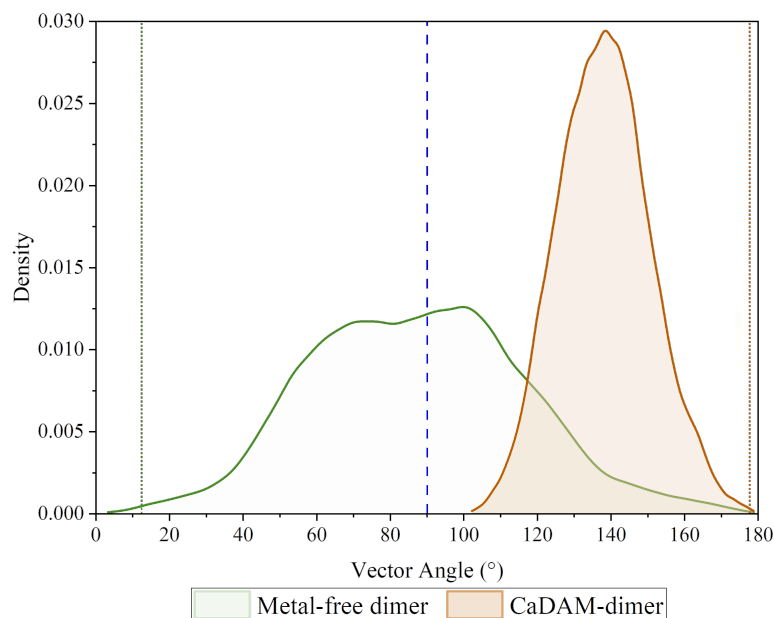
Figure C.6: Density plot of the angle between the two vectors, corresponding to each of the chains in the dimers. The green/orange dotted lines indicate the angle of the starting conformation, with the blue dashed line at the 90° mark, indicating the perpendicular arrangement of the chains.

Table C.2: Detailed secondary structure percentages of the three main regions of the metal-free αS-dimer.

Region	β character (%)	α character (%)	Other (%)
Chain A			
N-terminal	4.77	24.94	70.29
NAC	15.31	21.70	63.00
C-terminal	0.43	20.66	78.91
Total	6.01	22.75	71.24
Chain B			
N-terminal	3.84	22.01	74.14
NAC	9.40	22.20	68.40
C-terminal	0.37	21.37	78.27
Total	4.11	21.85	74.03

Table C.3: Detailed secondary structure percentages of the three main regions of CaDAM-αS dimer.

Region	β character (%)	α character (%)	Other (%)
Chain A			
N-terminal	7.70	17.96	74.34
NAC	13.98	24.92	61.10
C-terminal	0.42	21.48	78.10
Total	6.93	20.83	72.24
Chain B			
N-terminal	8.67	19.27	72.06
NAC	12.55	21.88	65.57
C-terminal	0.33	21.24	78.42
Total	6.96	20.56	72.48
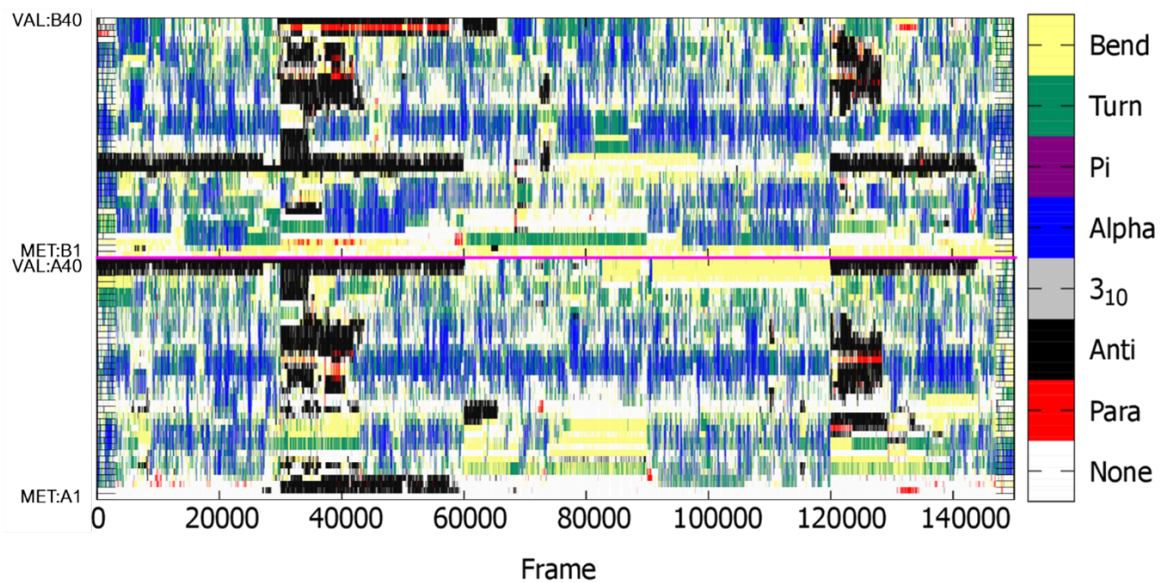
Figure C.7: Evolution of secondary structural elements of each of residues M1-V40 in the two chains of the CaDAM-bound dimer. The pink line separates the two chains.
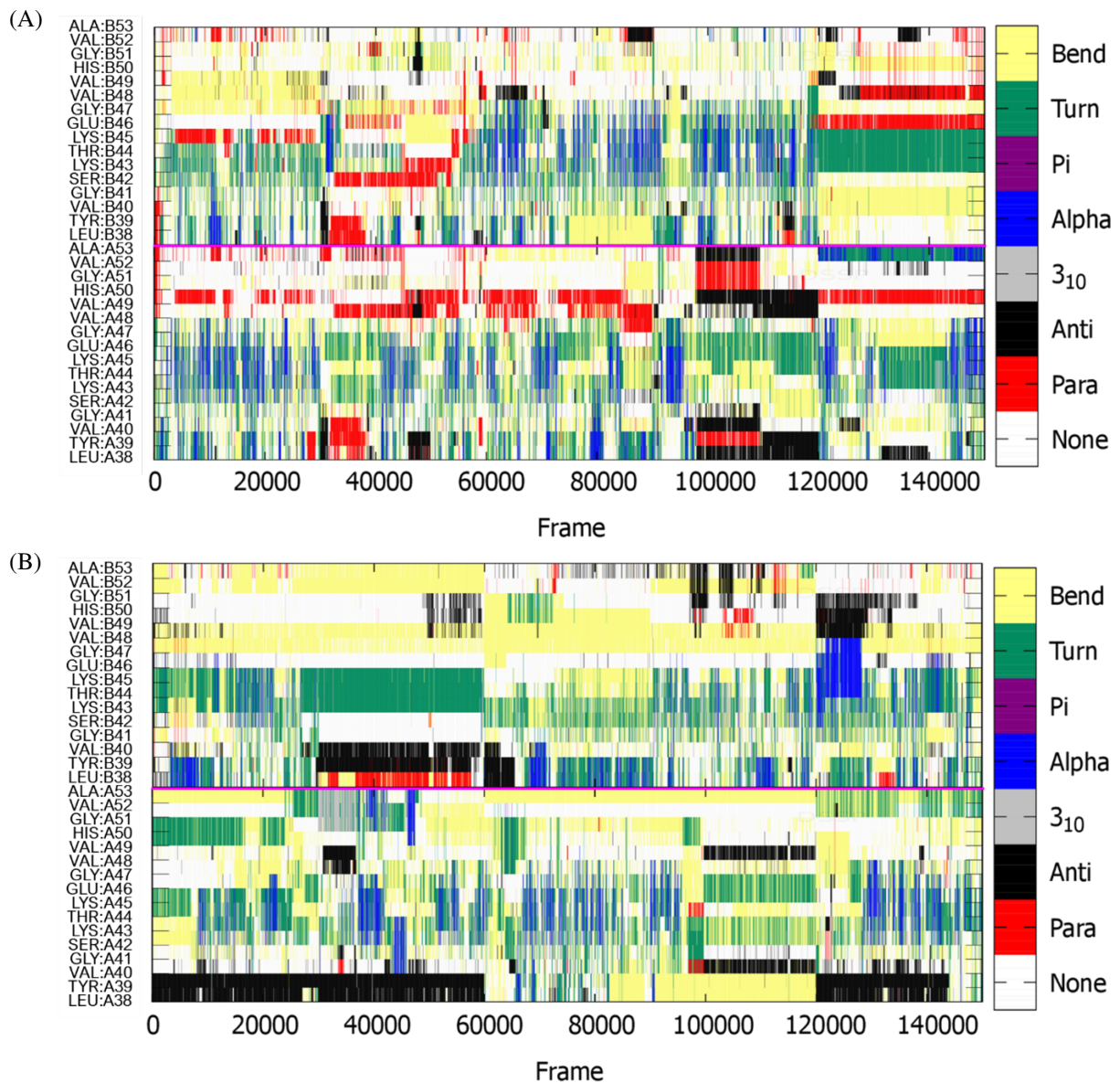
Figure C.8: Evolution of secondary structural elements of each of residues L38-A53 in the two chains of the (A) metal-free and (B) CaDAM-bound dimers. The pink line separates the two chains.
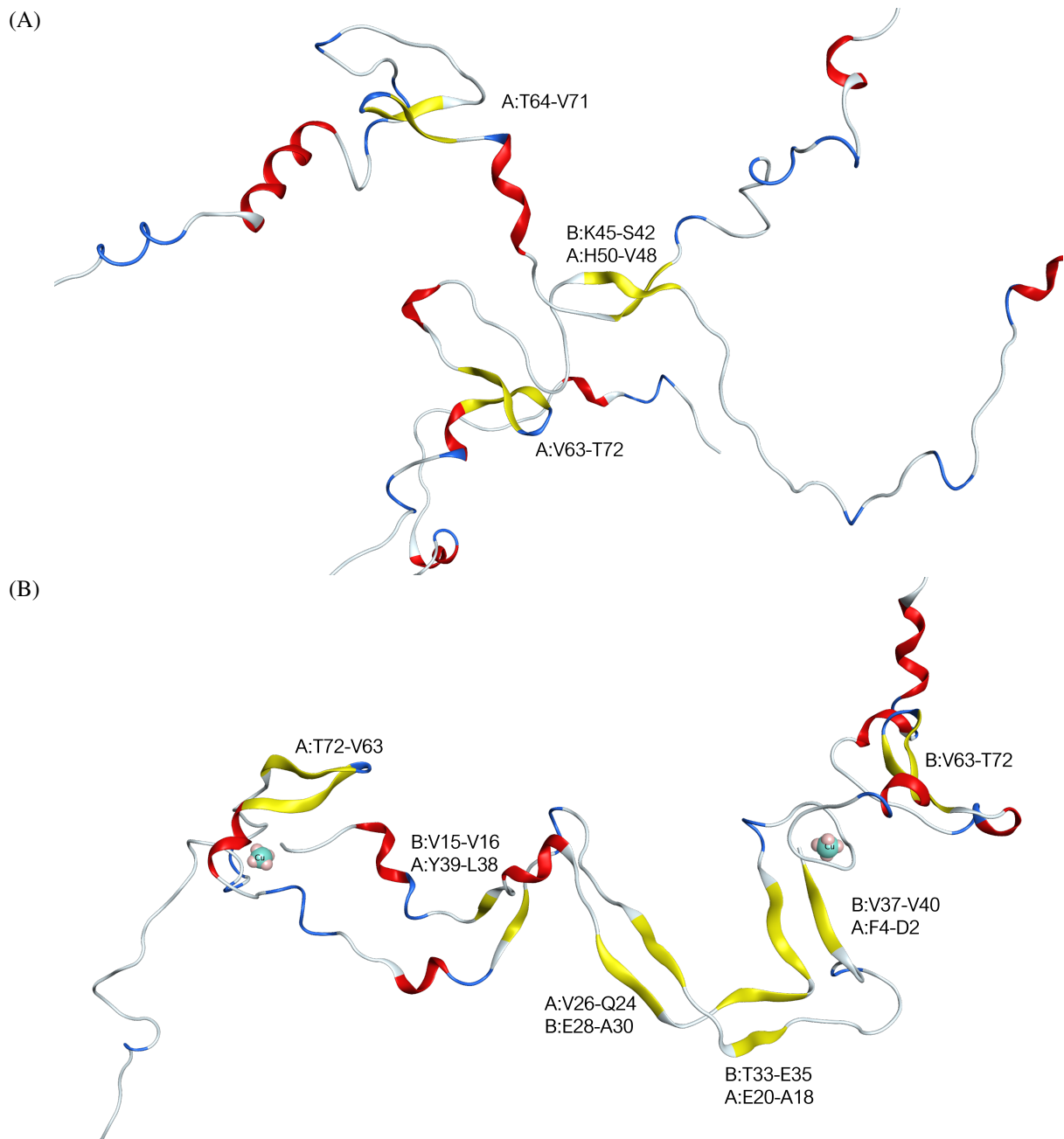
(A)

(B)

Figure C.9: Snapshot of two random frames (#40,000) from the (A) metal-free and (B) CaDAM-bound dimers, with labelled residues involved in the β-strands (yellow) going from top to bottom, and left to right.

Table C.4: Clusters from the two dimer systems, created using PCA of the cartesian coordinates of Cα.

Cluster#	Frames in Cluster (out of the total 150,000)
free-dimer	
1	44,808
2	460
3	41
CaDAM-dimer	
1	39,981
2	3,751
3	2,045
4	52
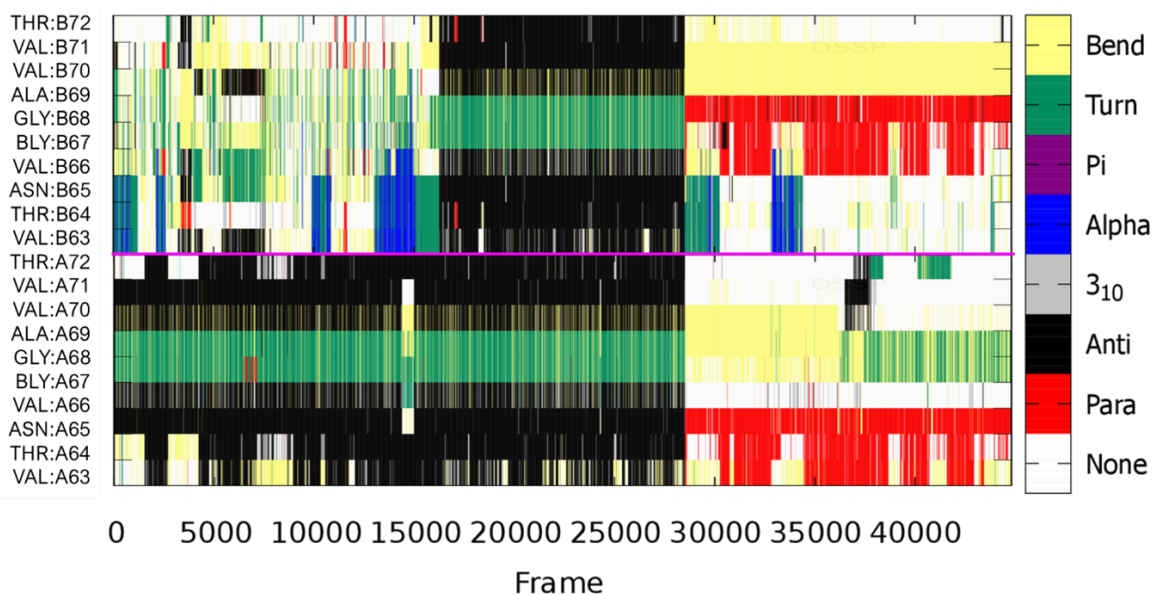
Figure C.10: Evolution of secondary structural elements of each of residues V63-T72, from the most populated cluster, in the two chains of the metal-free dimer. The pink line separates the two chains.

Table C.5: Breakdown of the runs performed on the metal-free dimer, with an indication of the time for the ones that separated and a snapshot of the last frame in each of them.
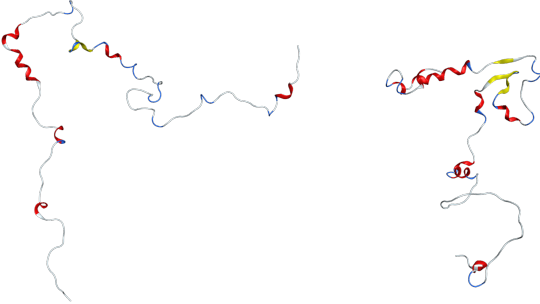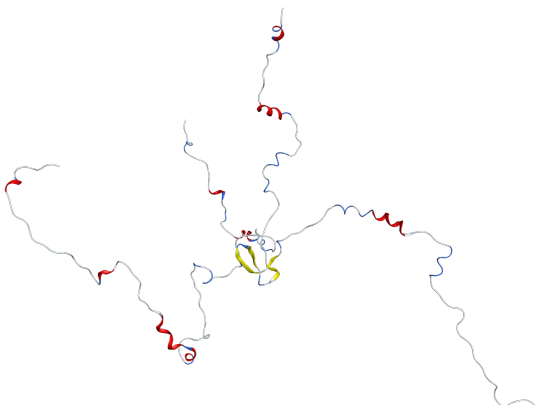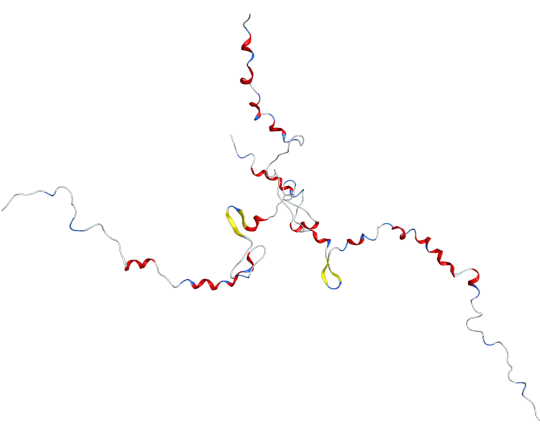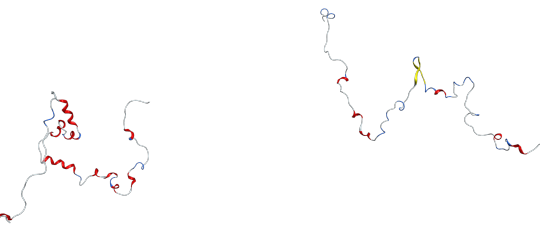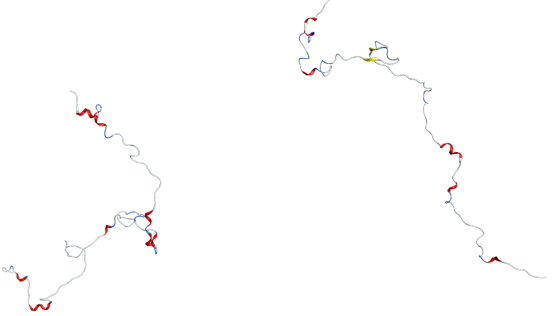
Run#	Separated? [Time (ns)]	Snapshot of last frame
1	Yes [150]	
2	No [400]	
3	No [400]	
4	Yes [150]	

Table C.5: Breakdown of the runs performed on the metal-free dimer, with an indication of the time for the ones that separated and a snapshot of the last frame in each of them.
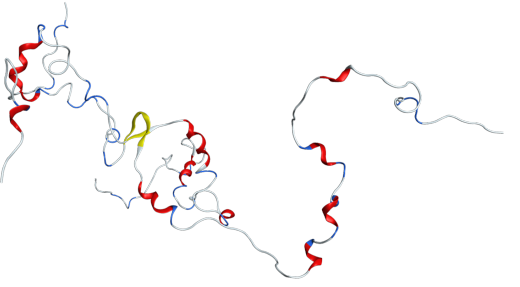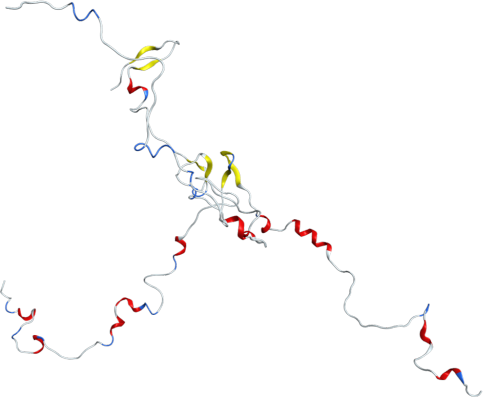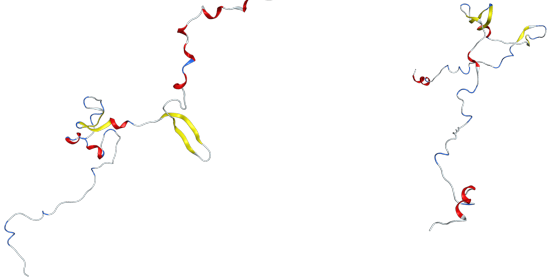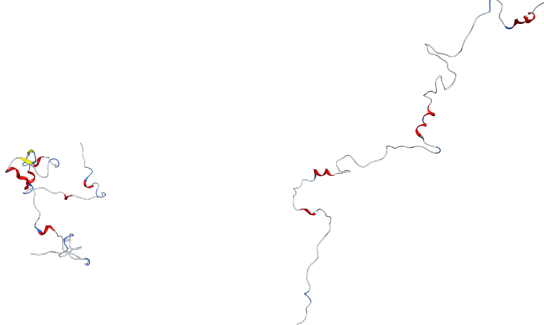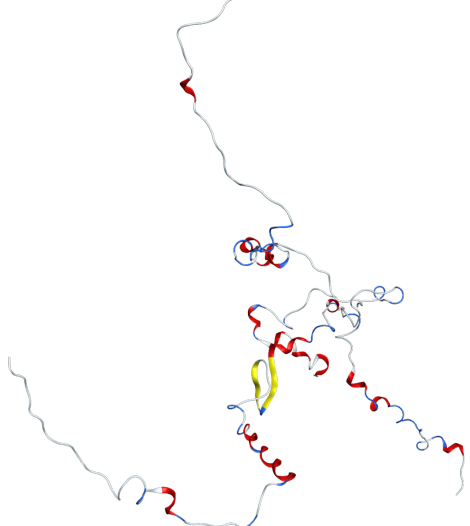
Run#	Separated? [Time (ns)]	Snapshot of last frame
5	Yes [150]	
6	Yes [250]	
7	No [400]	
8	No [400]	

Table C.5: Breakdown of the runs performed on the metal-free dimer, with an indication of the time for the ones that separated and a snapshot of the last frame in each of them.

Run#	Separated? [Time (ns)]	Snapshot of last frame
9	Yes [100]	
10	Yes [200]	
11	No [400]	
12	Yes [100]	

Appendix D

Chapter 6

Table D.1: Mean distances between atoms in the Cu(II) binding sites of A53T-αS.

	Avg. Distance (Å)	SD (Å)
M1N-Cu(II)	1.765	0.054
D2N-Cu(II)	1.982	0.049
D2O-Cu(II)	1.903	0.065
H50N-Cu(II)	1.925	0.053
D119O-Cu(II)	1.808	0.066
D121O-Cu(II)	1.821	0.054
N122O-Cu(II)	1.554	0.224
E123O-Cu(II)	1.802	0.079

Table D.2: Angles between atoms in the Cu(II) binding sites of A53T-αS.

	Avg. Angle (°)	SD (°)
M1N-Cu(II)-D2N	90.2	2.5
M1N-Cu(II)-H50N	102.1	6.3
D2O-Cu(II)-H50N	104.0	7.1
D2N-Cu(II)-D2O	60.7	2.7
D119O-Cu(II)-D121O	85.6	4.9
D121O-Cu(II)-N122O	106.4	5.7
N122O-Cu(II)-E123O	63.6	6.1
D119O-Cu(II)-E123O	150.3	3.7

Table D.3: Detailed secondary structure percentages of the three main regions of αS (A53T and pS129), after aMD simulations using ff03ws/OBC.

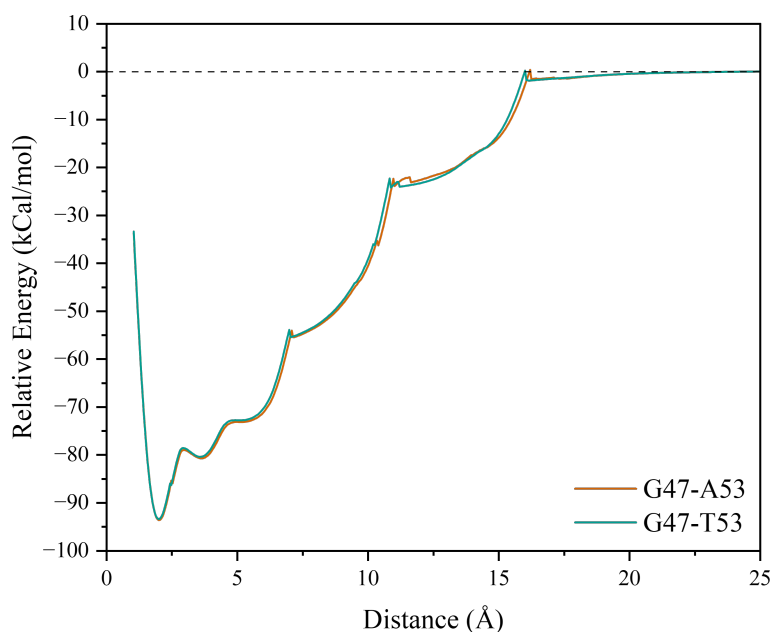| Region	β character (%)	α character (%)	Other (%)
free-αS [A53T]			
N-terminal	1.14	15.32	83.54
NAC	3.18	16.19	80.63
C-terminal	0.14	15.95	83.91
free-αS [pS129]			
N-terminal	1.54	16.05	82.41
NAC	3.41	16.03	80.56
C-terminal	0.17	16.60	83.23
Cu(II)-αS [A53T]			
N-terminal	2.83	11.45	85.72
NAC	4.86	16.08	79.06
C-terminal	0.15	12.90	86.95
Cu(II)-αS [pS129]			
N-terminal	0.65	10.95	88.40
NAC	4.62	16.61	78.77
C-terminal	0.30	13.61	86.09

Figure D.1: Relative binding energy of Cu(II) to fragments G47-A53 (WT) and G47-T53 (A53T). The distance is defined between the metal ion and Nδ from the imidazole ring of H50.
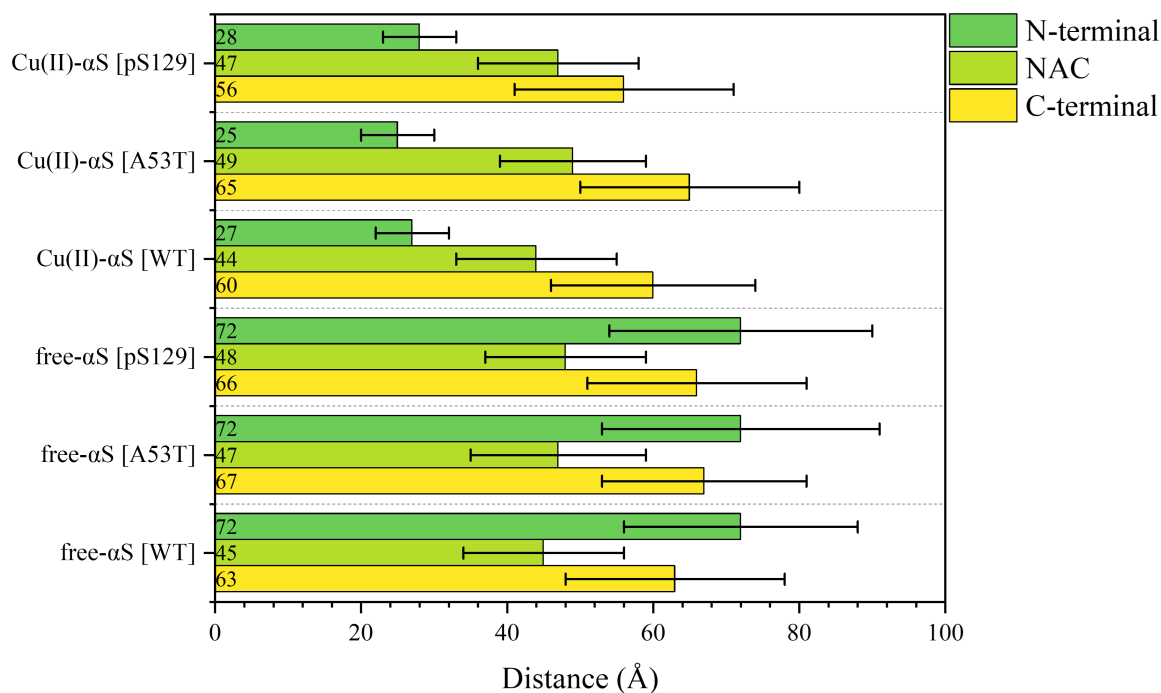
Figure D.2: End-to-end distances of the WT, A53T and pS129 systems.

Table D.4: Mean distances between atoms in the Cu(II) binding sites of pS129-αS.

	Avg. Distance (Å)	SD (Å)
M1N-Cu(II)	1.994	0.055
D2N-Cu(II)	1.902	0.047
D2O-Cu(II)	1.935	0.056
H50N-Cu(II)	1.979	0.052
D119O-Cu(II)	1.805	0.067
D121O-Cu(II)	1.876	0.062
N122O-Cu(II)	1.075	0.167
E123O-Cu(II)	1.875	0.081

Table D.5: Angles between atoms in the Cu(II) binding sites of pS129-αS.

	Avg. Angle (°)	SD (°)
M1N-Cu(II)-D2N	76.7	2.8
M1N-Cu(II)-H50N	95.6	3.2
D2O-Cu(II)-H50N	94.9	2.7
D2N-Cu(II)-D2O	94.5	2.7

	Avg. Angle (°)	SD (°)
D119O-Cu(II)-D121O	78.7	4.1
D121O-Cu(II)-N122O	105.2	4.4
N122O-Cu(II)-E123O	75.7	8.9
D119O-Cu(II)-E123O	152.3	4.1

The RMSF values in the pS129 system, Figure D.3, did not present any significant changes in the fluctuation of the residues, compared to the WT-form, overall maintaining the same trend, with a maximum difference of *ca.* 6 and 4 Å in the metal-free and Cu(II)-bound systems, respectively. These differences are expressed between residues E130-A140 in the former, and A11-V16 in the latter. In the case of the metal-bound system, the difference comes from an increase in the fluctuations, which appears to be a result of the overall breakage of hydrogen bonds in the N-terminal, Table 6.6. The decrease in the fluctuations in the metal-free peptide, are now in line with the motion seen in the phosphorylated metal-bound system. However, the inherently unstructured nature of the C-terminal of αS, makes it difficult to associate the changes in the fluctuations to a specific structural feature in that region, and could instead be a result of the carrying effect from the less noticeable drop in the motion of residues in the NAC region.
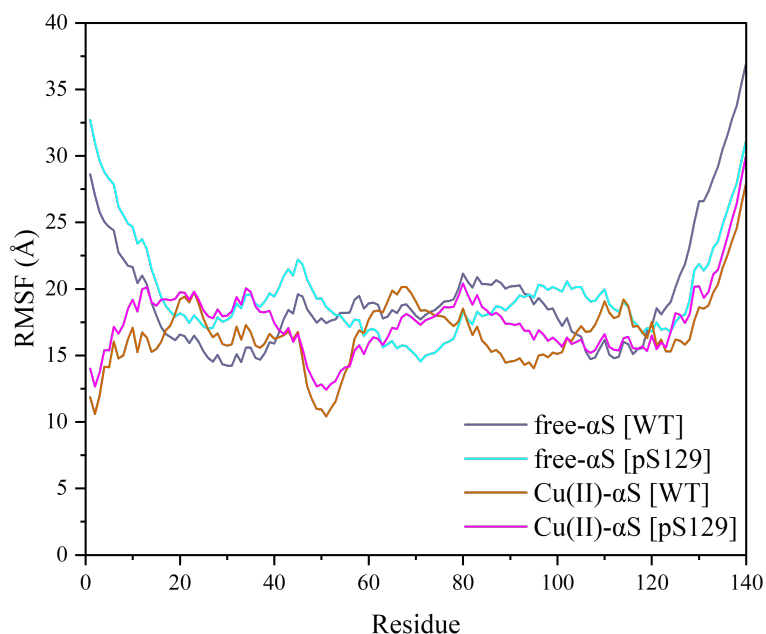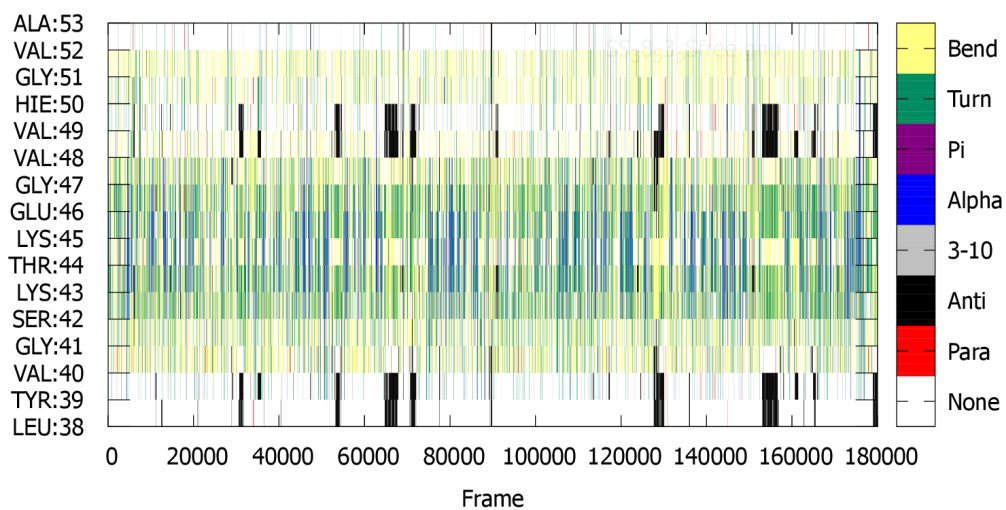

Figure D.3: RMSF of the residues in the metal-free and Cu(II)-bound WT and pS129 systems.
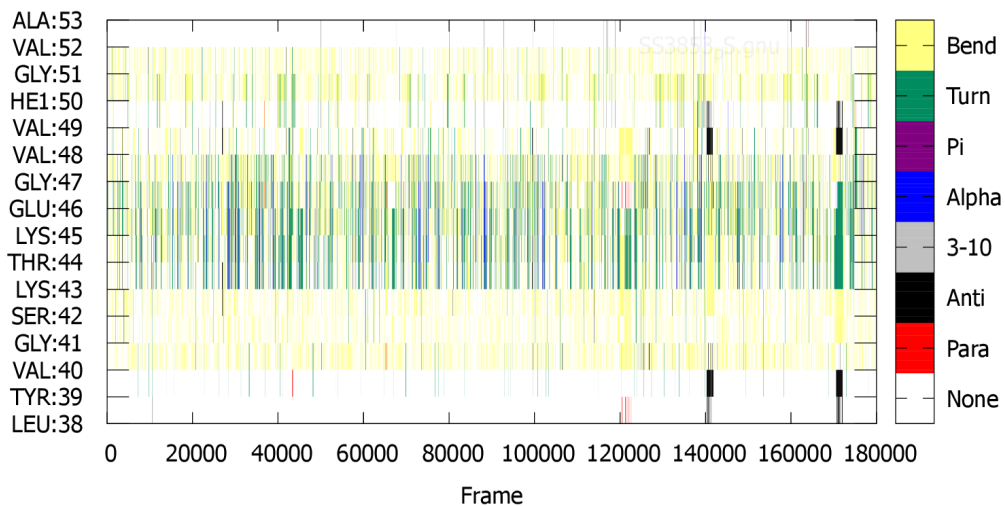
(A)

(B)

Figure D.4: Evolution of secondary structural elements of each of the residues (L38-A53) in the primary hairpin region from the (A) metal-free and (B) copper-bound pS129 systems.

Bibliography

(*1*) Zirah, S. et al. (2006). Structural Changes of Region 1-16 of the Alzheimer Disease Amyloid β-Peptide upon Zinc Binding and in Vitro Aging. *Journal of Biological Chemistry 281*, 2151–2161.

(*2*) Shen, Y., and Bax, A. (2010). SPARTA+: A modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of Biomolecular NMR 48*, 13–22.

(*3*) Bermel, W. et al. (2006). Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *Journal of the American Chemical Society 128*, 3918–3919.