UAB
Universitat Autònoma de Barcelona

Computational Study of the Interaction of Metals with Biological Scaffolds:

Neurodegenerative Diseases and Catalysis

Lorena Roldán Martín

A Dissertation Presented to the
Universitat Autònoma de Barcelona
in Candidacy for the Degree of
Doctor in Bioinformatics

Doctoral Programme of Bioinformatics

Advisers
Professors Jean-Didier Maréchal & Mariona Sodupe

December 2023

Computational Study of the Interaction of Metals with Biological Scaffolds:

Neurodegenerative Diseases and Catalysis

Lorena Roldán Martín

Recommended for Acceptance By Advisers

Jean-Didier Maréchal   Mariona Sodupe Roure

15th December 2023

"Nothing in life is to be feared,
it is only to be understood.
Now is the time to understand more,
so that we may fear less."

Marie Curie

Abstract

Metalloproteins represent one-third of the currently known proteome. Their implications range from several functional roles, such as catalysis; to disease in the case of metal dyshomeostasis, as is the case of neurodegenerative diseases. A great effort is placed into their study by the scientific community since their proper description can lead both to new-to-nature functions and new druggable targets.

While the experimental study of metalloproteins is a daunting challenge due to the technical requirements and biological determination, their computational study is nowadays a suitable alternative. In the last decades, the improvement in computational approaches such as Quantum Mechanics and Molecular Mechanics has exponentially increased the capacity of exploring protein conformational changes and reactivity, reaching meaningful biological time scales. Despite such an advance, a series of challenges have still to be dealt with when metal ions are included in the equation. The present Ph.D. Thesis has focused on developing computational pipelines suitable for overcoming the challenges that the study of metallopeptides represents, placing a special effort on the treatment of metal ions. Moreover, a new computational software is proposed, which could be further applied for the determination of metal ion channels in preorganized proteins.

The first part of this work reports the application of an integrative computational approach to the study of metal ion implication in $A\beta_{42}$ aggregation. In order to shed light on the conformational changes that $A\beta_{42}$ undergoes upon metal binding, monomeric and fibrillar structures have been selected. For the former, metal binding spheres were already determined, while new binding sites had to be searched for the fibrillar ones. Besides, this work also aims to describe the differences among familial

forms of the Aβ_{42}, both on monomer and fibrillar structures. With this study, it is demonstrated that metal binding has a great impact on the Aβ_{42} peptide, regardless of the aggregation state, though each metal ion affects each variant in a particular manner.

Besides, amyloid-like aggregates can also perform catalytic reactions. The second part of this work applies computational techniques, including QM/MM approaches, in the study of two examples of catalytic heptapeptides, whose catalytic function requires the coordination of Zn(II). The catalytic performance of both heptapeptide structures has been experimentally demonstrated to be tuned just by the modification of a single residue of the sequence. Since the molecular reasoning behind such modification in catalysis is unknown, the aim of this project is to describe the conformational changes or interactions that may interfere with and cause the differential catalytic activity. The first study is centred on the hydrolysis of p-nitrophenolate, in which we demonstrate that the 6th residue directly interacts in the catalytic process. For the second project, focused on the enantioselectivity of ethyl diazoacetate, it is assessed that differences in the 2nd residue of the fibre result in different binding poses of the catalyst and, hence, in different enantioselectivity.

The last part of this Thesis is a successful demonstration of a new computational tool which aims at describing metal ion channels in preorganized molecular scaffolds, whose potential is effectively demonstrated by its application on the already crystallographically determined Cu(II) channel in Tyrosinase. Furthermore, its application is also tested in the case study of Human Tyrosinase, whose metal channel is unknown. With its application, a possible Cu(II) channel is proposed, which could ultimately represent new druggable targets for several diseases.

List of Abbreviations

Aβ: amyloid-β peptide

AD: Alzheimer's Disease

APP: amyloid precursor protein

EDA: ethyl diazoacetate

GA: Genetic Algorithm

GaMD: Gaussian accelerated Molecular Dynamics

MCPB: Metal center parameter builder

MM: Molecular Mechanics

MD: Molecular Dynamics

NMA: Normal mode analysis

QM: Quantum Mechanics

QM/MM: Quantum Mechanics / Molecular Mechanics

PCA: Principal component analysis

PD: Parkinson's Disease

PDB: Protein Data Bank

pNPA: para-nitrophenyl acetate

RESP: Restrained Electrostatic Potential

RMSD: Root Mean Square Deviation

x

Contents

1

Introduction

1.1 Metalloproteins

By far, proteins are the most sophisticated known molecules whose structures have been fine-tuned by the evolution process over millions of years. Proteins are built on 20 types of amino acids, each with unique chemical properties. The amino acids are bonded through covalent interactions to form a polypeptide backbone, constituting the primary structure (Figure 1.1). The hydrogen bonding of the protein backbone induces another layer of structural organisation and recurrent folding pattern of the amino acid chain: the secondary structure. α-helix forms a rigid cylinder, with a complete turn every 3.6 amino acids, usually present in cell

membrane proteins due to its hydrophobic character. If not in transmembrane proteins, α-helices tend to wrap around themselves in a globular disposition, forming a strong stable structure. On the contrary, β-sheet regions tend to form either parallel or antiparallel β-sheets, connecting neighbouring regions of the polypeptide chain through hydrogen bonds, resulting in very rigid structures. Besides, such β-sheet structure is typically related to aggregated proteins, which are also highly stable due to the hydrogen bonds formed between the different chains or strands. The formation of either α-helix or β-sheet secondary structures is inherent to the backbone of the amino acid chain, meaning they can be adopted by many combinations of the protein sequence.

A higher folding stage depends on the hydrogen and ionic bonds and van der Waals attractions formed among amino acids and the polar character of the residues' sidechain. All these interactions determine the three-dimensional arrangement of the amino acid chain, called tertiary structure, typically corresponding to the one with the lowest free energy. Such three-dimensional conformations are unique to each group of proteins, determining protein families.

The last organisation state encountered in protein structures is the quaternary structure, as domains with consistent functions or quaternary arrangements can be formed if a protein is composed of more than one polypeptide chain. Such conformations, though, can be altered if an external element interacts with the molecule, as is the case of other proteins or metals.[1]

When a metal ion is part of a protein, the concept of metalloprotein arises, representing from one-third to half of the whole proteome.[2] Early in evolution, metal ions have been demonstrated indispensable for life. Not only do they interact with a wide range of biomolecules, but are also involved in an extensive range of functions such as enzymatic activities,

protein folding, signalling, and catalysis.[3,4] Notwithstanding, metal ions are also widely known for having a determining role in several important diseases, most remarkably in neurodegenerative ones.

α-helix

β-sheet

Primary Structure   Secondary Structure   Tertiary Structure   Quaternary Structure
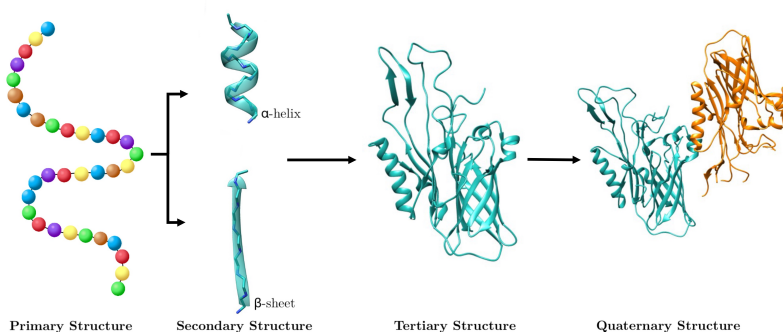
Figure 1.1: Folding steps of an aminoacid chain.

1.1.1 Metals in Health

Twenty known elements are essential for the proper function of most organisms, with carbon (C), oxygen (O), hydrogen (H), and nitrogen (N) making up 96% of the human body. Other major or trace elements include sodium, potassium, calcium, magnesium, manganese, iron, cobalt, copper, molybdenum, and zinc (Table 1.1). Despite these species being necessary for life and to perform a wide range of biological functions (Table 1.1); either their deficiency or excess are toxic for life. Therefore, it is required to maintain a proper balance between the two sides of the coin, a process called homeostasis.[5]

On the functional side, many catalytic and structure-stabilizing roles have already been attributed to metal ions. Around 40% of the proteins with at least one structure available in the protein data bank in 2008 were metal-dependent, known as metalloproteins, which include oxidoreduc-

Table 1.1: Binding properties, mobility, mass of metals in the average human body and biological function of essential metals.

Metal ion	Binding Properties	Mobility	Mass of metals (g)	Biological function
Na^+	Weak	High	65-115	Regulatory function: osmotic pressure, membrane potential, enzyme activity Nerve action, osmotic pressure
K^+	Weak	High	155-195	Regulatory function: osmotic pressure, membrane potential, enzyme activity Nerve action, osmotic pressure
Mg^{2+}	Intermediate	Semi-mobile	30	Energy metabolism ATP to ADP, chlorophyll
Ca^{2+}	Intermediate	Semi-mobile	1100	Signalling, enzyme regulation, muscle contraction, skeletal systems (Bones, Teeth)
Zn^{2+}	Intermediate	Intermediate	2.5	Lewis acid (Carbonic anhydrase, carboxypeptidase), structural roles - zinc fingers, repair enzymes
$Fe^{2+,3+}$	Strong	Low	4.2	Oxygen storage (haemoglobin, hemerythrin, electron transfer Fe-S proteins, cytochromes)
$Co^{2+,3+}$	Strong	Low	2.5	Vitamin B12 coenzyme
$Cu^{+,2+}$	Strong	Low	0.075	Electron transfer (copper blue proteins) oxygen storage, transport proteins, ceruloplasmin
$Mn^{2+,3+,4+}$	Strong	Low	0.013	Enzyme phosphatase, mitochondrial Mn-SOD, photoredox activity PS II
$Mo^{2+,3+,4+,5+,6+}$	Strong	Low	0.005	Enzymes (nitrogenase, reductases, hydroxylases)

tases, isomerases and ligases, to name just a few. In such systems, metals can act as i) an essential element for the proper structural organization of the protein; ii) independent cofactors, such as Mg^{2+} in DNA polymerase; or iii) be a cofactor as a part of a higher complex, like Fe^{2+} in the heme group, an essential co-substrate to perform redox reactions in haemoglobin.

Regarding their structural roles, metal ions are able to modify the content of secondary structure upon binding, as well as help in the configuration of tertiary - even quaternary – proteins' structures. In some cases, such as the copper protein plastocyanin, the metal-binding site is so preorganized that the removal of the metal ion goes unnoticed. On the contrary, zinc finger domains exist as unfolded peptides until a metal ion is present, inducing the proper folding of the structure. An intermedia situation happens with calmodulin, which is already partially folded until Ca^{2+} binds and promotes its final mature form.[6] Thus, metal ions can impact proteins' structure at different organization stages, as they can either bind to an already preorganized binding site or provoke an induced fit upon the unfolded receptor molecule.

Besides, the roles of metal ions are very diverse, ranging from redox catalysis, activation of reactive species, to electrostatic stabilization of structural intermedia states. As seen in Figure 1.2, different metals can perform similar functions in distinct enzymes; as well as the same metals can carry out different functions. Such an event is due to the evolution process itself since living organisms have adapted to whichever was most easily accessible in the environment.[7] The fact that metal ions can bind to many structures also paves the way for developing artificial enzymes, inspired by mimicking biological systems with metal cofactors. Metal ions can provide new catalytically active systems by binding artificially to systems such as DNA architectures or amyloid fibres, tuning their reaction.
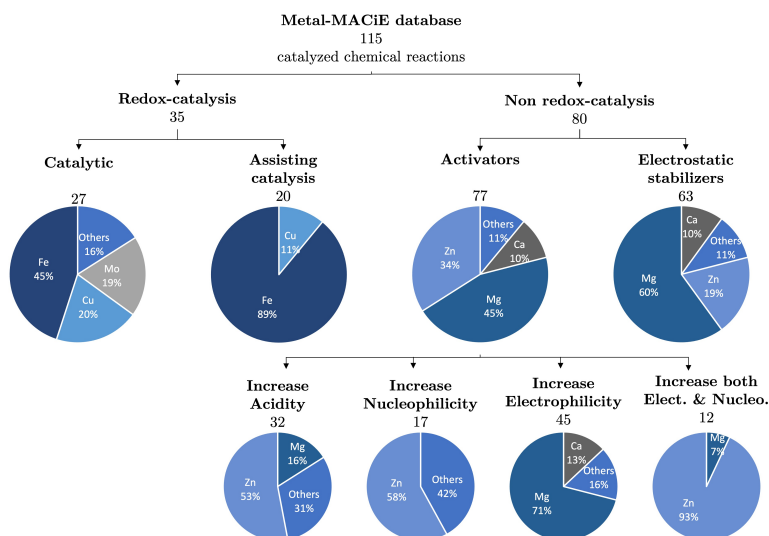
Figure 1.2: Modified from [7], represents the percentage of catalytic reactions performed by each metal ion considering the information in the Metal-MACiE database.

Although metal homeostasis is a tightly controlled process, metal ions can become toxic to the living organism if their concentration happens to be imbalanced or if they are placed in an inappropriate location.[8,9] The presence of a metal ion in an unexpected site may cause competition with other biologically active metal ions of the specific site, maybe even leading to undesired or uncontrolled reactions. Some other metal ions may not even be active, but their mere presence may interfere with functional processes. Besides, the disruption of metal ions' balance typically triggers harmful cell processes, being the production of high amounts of reactive oxygen species (ROS) one of the main consequences. As a result, DNA damage, formation of reactive aldehydes, and protein misfolding might happen.[10] Increasing evidence supports that metal ions dyshomeostasis is linked to several diseases, being neurodegenerative the most recurrent

ones due to the physiological importance of metal ions in a wide range of brain processes (Table 1.2).[10] Ataxia, Huntington's, Parkinson's, and Alzheimer's disease are just some of the extensive number of examples in which metal dyshomeostasis has been demonstrated to cause a detrimental effect.[11] Considering that most of these diseases still lack effective treatment, the importance of investing efforts in studying metal ions and their relation with neurological disorders is reinforced.

One of the main therapeutic strategies when homeostasis is lost is based on metallopharmacology, aiming at either restoring or chelating out the lack or excess of metals, respectively. Chelators are organic or inorganic compounds designed for capturing metal ions, which must be either i) capable of sequestering toxic metal ions, resulting in a reduction of free toxic compounds; or ii) capable of freeing metal ions in a concrete environment once a certain region is reached. Therapies with chelators have been successfully applied for a comprehensive series of human alterations, such as anticancer drugs or Wilson's disease. Moreover, several studies also support that polyphenols, with divalent metals chelation and antioxidant activities, demonstrate neuroprotective effects.[12] However, chelators' therapy also presents serious side effects, mainly arising from chelators' inability to access intracellular sites, affecting exclusively those metal ions located in the interstitial space. Besides, the lack of specificity may result in excessive removal of the target metal ion or unspecific capturing of other essential metal ions, causing the organism to suffer from metal ion deficiency.[13,14] Thus, it is essential to carefully define the role of each metal ion in the organism and the effects of its dyshomeostasis to verify if metal chelating could represent a therapeutic option; then encounter a specific chelator molecule, with the proper chemical features, for its capturing or location-sensitive releasing.

Table 1.2: Neurodegenerative disorders with possible metal-associated pathology.[10]

Disorder implicated	Metal implicated	Metalloproteins or enzymes
Alzheimer disease	Cu, Fe, Zn, Al ↑	Aβ, APP, tau protein ↑
Parkinson disease	Fe, Zn ↑	α-Synuclein↑, neuromelanin↓, ferritin↓
Lewy body dementia	Fe↑	α-Synuclein↑
Multisystem atrophy	Fe, Cu↑	α-Synuclein↑
Progressive supranuclear palsy	Fe↑	Tau protein↑
Creutzfeldt-Jakob disease	Cu, Fe↑	Prion protein
Amyotrophic lateral sclerosis	Cu, Fe, Zn↑	SOD1↑, TDP-43↑
Friedreich ataxia	Fe↑	Frataxin↑
Multiple sclerosis	Fe↑	
Wilson disease	Cu↑	Ceruloplasmin deficiency
Menkes disease	Cu↓	Cytochrome oxidase↑
PKAN/NBIA type 1	Fe↑	Pantithenate kinase deficiency
Huntington disease	Fe, Ca, Mn↑	Huntingtin↑
Infantile neuroaxon dystrophy	Fe↑	
Neuroferritinopathy	Fe↑	Ferritin light-chain gene (FTL1)
Down syndrome	Cu, Fe, Zn ↑	Aβ ↑, tau protein↑

Metal ions can be chemically defined as Lewis acids, which frequently bind to donor ligands accomplishing the rules of the hard-soft theory of acids and bases (HSAB). The "soft" accounts for large and polarizable
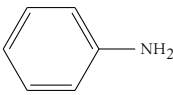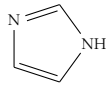
species, typically non-essential metal ions, which would preferably interact weakly with soft ligands as sulphur atoms. In contrast, "hard" accounts for those small, less polarizable species to which essential metal ions belong, preferring to strongly interact with hard, base-like ligands (Table 1.3). Certain hard metal ions can adopt different oxidation states, allowing them to catalyse redox reactions. Besides, their acidic character allows them to selectively bind physiological meaningful ligands, such as O_2, related to respiration processes. Hence, metal ions would prefer to bind to ligands belonging to the same category in the classification, following the distribution available in Table 1.3.[6]

The 20 amino acids conforming proteins are composed of atoms which indeed accomplish the criteria of Lewis ligands, being suitable receptors for metal ion binding. Some of those residues, such as Asp or Glu are even able to bind two metal ions each or bind in a bidentate mode. Then, if the three-dimensional structure of a protein is correctly -or partially-preorganized, metal ions can access the binding site and coordinate with a concrete region. The coordination depends both on the oxidation state of the metal ion and the number of coordinating residues, arranging specific geometries. Figure 1.3 depicts some of the most common geometries for 3 to 6 coordinating residues, the most typical range for metal ions, though substantial distortions from these ideal structures are expected in real systems. Thus, different geometries are feasible even with the same number of residues involved in metal coordination, one of which will be preferred according to the electronic features of binding ligands and electron *d* orbitals of the metal ion itself.

The study of affine environments and the result of metal dyshomeostasis should be then individually assessed for each metal ion. In the next sections, a summary of relevant metal ions to the present Thesis is provided, focusing on their physiological activities but also on their impli-

cation in neurodegenerative diseases. For each metal, its chemical traits and unique biological functions will be presented in individual sections.

Table 1.3: Metal ions and ligand classification according to the hard-soft theory of acids and bases (HSAB).[6]

Metals			Ligands		
Hard					
H^+	Mn^{2+}	Cr^{3+}	H_2O	$CO_3{}^{2-}$	NH_3
Na^+	Al^{3+}	Co^{3+}	OH^-	$NO_3{}^-$	RNH_2
K^+	Ga^{3+}	Fe^{3+}	$PO_4{}^{3-}$	R_2O	RO^-
Mg^{2+}	Ca^{2+}	Tl^{3+}	$ROPO_3{}^{2-}$	$(RO)_2PO_2{}^-$	Cl^-
Borderline					
Fe^{2+}	Ni^{2+}	Zn^{2+}	NO_2^-		
Co^{2+}	Cu^{2+}		N_2	(aniline, NH_2)	
			$SO_3{}^{2-}$		
			Br^-	(imidazole, N NH)	
			$N_3{}^-$		
Soft					
Cu^+	Pt^{2+}	Pt^{4+}	R_2S	R_3P	
Au^+	Tl^+	Hg^{2+}	RS^-	CN^-	
Cd^{2+}	Pb^{2+}		RSH	RNC	
			$(RS)_2PO_2{}^-$	$(RO)_2P(O)S^-$	
			SCN^-	CO	
			H^-	R^-	

Coordination number 3   **Coordination number 4**   **Coordination number 5**   **Coordination number 6**

120°

Trigonal planar   Square planar   Square pyramidal   Octahedral

120°

109°

90°

120°

Pyramidal   Tetrahedral   Trigonal bipyramidal

T-shaped

Figure 1.3: Possible geometries for each coordination number. Image modified from [6].

1.1.1.1   Iron biochemistry

Iron corresponds to the fourth most abundant element on Earth, the second metal behind aluminium. Besides, it is also the fifth most abundant metal in the human body (5 g in a person of 70 kg weight).[15] Along the evolution process, due to its high environmental abundance and versatility as a transition metal, it has become essential for all known living species, a crucial part of multiple biological functions.

Chemical features

In biological systems, iron can be encountered in three different oxidation states: Fe(II), Fe(III) and, though less frequent, Fe(IV). Each component demonstrates differential traits: while Fe(II) is soluble; Fe(III) precipitates as iron hydroxide in contact with water molecules. The rapid interconversion between Fe(II) and Fe(III) facilitates its main biological function, allowing electron transfer and acid-base reactions in an efficient manner. Hence, iron is responsible for single-electron transfer reactions, a core process in the metabolism of oxygen free radicals.

According to its electronic configuration, octahedral and tetrahedral coordination geometries are mostly expected, conformed by up to six coordinating compounds. Indeed, both Fe(II) and Fe(III) present high adaptability towards the coordination environments, retrieving either full or partially complete geometric spheres. Consequently, vacant sites are commonly generated, providing Fe with its main characteristic: the capacity of transporting and activating molecular O_2. Curiously though, proteins that form an incomplete iron coordination sphere, such as haemoglobin and hemerythrin, are related to the transport or activation of molecules. On the contrary, if no vacant sites are observed, as in cytochromes and ferredoxins, those proteins likely correspond to electron transfer processes.[16]

Despite the similarities in Fe(II) and Fe(III) coordination geometry, the binding ligands differ between the oxidation states. As Fe(II) belongs to the borderline category of HSAB, N and S atoms are preferred, meaning that His, Cys and Met residues usually build up the binding sphere. On the contrary, Fe(III) is a hard metal ion, favouring the interaction with O atoms belonging to Tyr or carboxylate groups -Asp or Glu-.[16]

Biological properties

Iron homeostasis is one of the most sophisticated processes in biology with several molecular systems involved. In mammals, proteins such as ferritin or transferrin participate in the solubilization, capture, transport, and release of iron, depending on the region of the organism. Being part of heme proteins (Figure 1.4), the essential cofactor for haemoglobin function, iron is an indispensable element in respiration processes.[17] A heme group is composed of a protoporphyrin IX molecule bound to an iron ion. In fact, 95% of functional iron in the human body is bound to a heme group, displaying an octahedral geometry with two vacant sites. One of such vacant sites is usually completed with molecular oxygen, as haemoglobin is involved in O_2 transport. Moreover, heme cofactor is the most versatile natural catalyst, capable of catalysing many other reactions, such as cyclopropanation.[18]
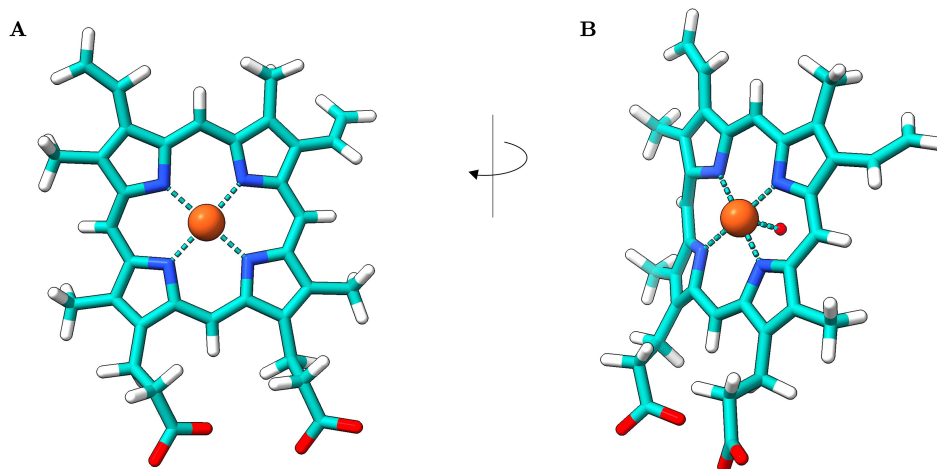
Figure 1.4: Heme protein with an iron atom. A. Frontal image of heme. B. Lateral view, where the ligand in the X axis is a O_2 molecule. The vacant site is typically completed with a ligand from the host protein.

Iron is also linked to several serious illnesses, most remarkable for its huge impact on brain dysfunction, but not exclusively. For example, iron excess is linked to the formation of reactive oxygen species (ROS), promoting the formation of hydroxyl radicals and, therefore, inducing neurotoxicity. Besides, iron has recently been described as essential also for cell death caused by ferroptosis, a pathophysiological process related to many diseases, whose exact mechanism is still under research.[19]

As seen in Table 1.2, iron is involved in many of the most prevalent neurological diseases, such as Alzheimer's and Parkinson's Disease. Experimental studies demonstrate that iron levels are increased in brain tissues with age, which raises suspicion of its relationship with PD and AD since both diseases are age-related.

At the first stage, before developing PD, α-synuclein seems likely to be part of physiological iron metabolism, as *in vitro* α-synuclein binding to iron ion has been demonstrated to reduce the insoluble iron form Fe(III) to Fe(II).[20] Nonetheless, in post-mortem analysis of PD human brains, the *substantia nigra* presents a disbalance in Fe(II)/Fe(III) ratio, favouring the latter. Besides, iron inappropriate retention in PD has been classically reported to be the causative factor of increased toxicity, not only by generating oxidative stress and DNA damage but also by increasing the aggregation rate of α-synuclein.[21]

In AD, increased iron concentration may also affect different levels. Regarding the Tau protein, Fe(III) can directly bind to its hyperphosphorylated form, increasing its aggregation rate. Concerning β-amyloid (Aβ), iron seems to be involved in several processing steps. First, it favours γ-secretase activity, which is involved in the amyloidogenic processing of APP, the precursor of Aβ.[12] Then, iron is also capable of binding Aβ in its N_{Ter} region, promoting oligomer formation and delaying the fibrillar one.[11]

Though with a less remarkable impact, iron also presents a role in Huntington's disease, multiple sclerosis, and inflammatory autoimmune disease, most of them related to the production of hydroxyl radicals. Hence, considering that iron overload is related to several diseases, therapeutic strategies considering chelators seem achievable. The chelator desferrioxamine (DFO) has already been demonstrated to down-regulate α-synuclein aggregation in PD individuals. However, further research is still to be done, as although these chelators are capable of distinguishing between Fe and Cu or Zn, they are still highly unspecific between Al and Fe.[22]

1.1.1.2 Copper biochemistry

Despite copper not being one of the most abundant metal elements on Earth, it performs indispensable reactions in organisms, with similar functions to iron. Both metal ions are essential for the physiological processes of the brain, with roles in protein stabilization, enzyme catalysis, and energy production. However, Cu is far less abundant in human bodies, with an average amount of 100 mg per 70 kg of weight.[15] Besides, Cu disbalance is also linked to severe detrimental outcomes, like oxidative stress, protein misfolding and aggregation.

Chemical features

As iron, Cu ion also presents two main oxidation states, Cu(I) and Cu(II), with different chemical properties. Since Cu(II) belongs to both soft and borderline categories (Table 1.3), it opts for coordinating with N and S atoms contained in His, Cys and Met residues, in agreement with the HSAB principle.[6] Besides, Cu(I) electronic configuration defines a d^{10} valence shell, which makes the coordination sphere highly flexible. It can be composed of either 1, 3 or 4 ligands; in linear, trigonal or tetrahedral disposition, being the latter the most recurrent one. Instead, Cu(II) is of-

ten found bound to 4 ligands, but opting for either square planar, square pyramidal or non-perfect octahedral -due to Jahn-Teller effect- geometries.[23,24]

Biological properties

Cu homeostasis is controlled by a complex network of chaperones, membrane transporters and cuproenzymes. Some examples are albumin and chaperones, responsible for Cu transport along the human body. Besides, metallothioneins act as storage of Cu, with the liver as the main reservoir in the human system.[25] Among all organs, the human brain is extensively studied in terms of Cu localization and transporters, due to its unique high intracellular levels. On the contrary, the rest of the organism tends to keep intracellular Cu concentration at very low ranges, as a moderate increase rapidly causes cytotoxicity and cell death. In fact, a novel copper-dependent form of cell death has been recently discovered, named cuproptosis.[26] Cuproptosis is different to any other form of cell death, as it relies on Cu binding to components in the tricarboxylic acid cycle (TCA), inducing their aggregation, causing proteotoxic stress and ultimately, cell death.

Regarding its physiological roles, Cu is essential for the brain's normal function, as it is not only present in several synaptic transmission processes and neuronal plasticity, but also involved in the homeostasis of other indispensable elements, such as Fe. Cu is also systemically meaningful, acting as a cofactor for several functional enzymes in different processes. One example explored in this PhD Thesis. is copper tyrosinase, a multifunctional dimeric enzyme found in melanocytes. Tyrosinase converts a tyrosine ligand to DOPA, whose polymerisation results in melanin, the skin's natural pigment. Other examples are catechol oxidases, which are similar to tyrosinase both in function and structure, though they contain three Cu centres in the active site.[5] Another example is cytochrome

c oxidase (CcO), the last enzyme in the electron transport chain, which contains two copper sites.

Since Cu is widely used by organisms, with an extensive variety of biological functions depending on it, several diseases are also related to Cu dysregulation. In fact, Cu role in cancer has been a long topic of study and a constant focus for anticancer drugs.[27] The excess of Cu results in cellular harm, catalysing the production of ROS, and leading to severe disorders such as Menkes and Wilson's disease. Moreover, it is also linked to AD and PD.[28] Regarding PD, Cu binding to α-synuclein still retrieves contradictory results, as it has been reported to cause a detrimental increase in hydrogen peroxide production but it may also generate some protective effects against ROS production. Hence, Cu role in PD still requires further study. More well-established is its role in AD, as Aβ contains a high-affinity patch for Cu binding, formed by His and Tyr residues on its N_{Ter} region. Once Cu binds to Aβ, it increases the production of hydroxyl radicals in the vicinity of Aβ plaques, spreading the damage. Besides, Cu has been demonstrated to provoke conformational changes in the Aβ protein, promoting fibrillar formation.[10] Accordingly, Cu is a double-edged sword, as in the proper concentration it is an essential cofactor for several enzymes, but its dysregulation leads to several diseases and, eventually, cell death.

Hence, cu chelators are a promising therapeutic strategy considering the broad spectrum of Cu-related diseases. Indeed, several chelators are being tested nowadays, such as TTM or D-penicillamine. Those chelators have retrieved successful results in delaying cancer metastasis in mouse models, mainly due to their anti-angiogenic effects.[25] Unfortunately, minor success has been obtained in copper chelating strategies for PD and AD, mainly due to the still limited knowledge of the mechanisms regulating copper imbalance and the consequent effect on the brain.[29] Most

recently, since the description of cuproptosis, its pharmacological induction has also been proposed as a potential therapeutic strategy for cancer, inducing the own cancer cell mechanisms to cause death. It is feasible to consider cuproptosis as a therapeutic approach for other diseases, though deep comprehension of its mechanism and applicability is still under study.

1.1.1.3   Zinc biochemistry

Zinc is the second most abundant essential metal, behind iron, with an average amount of 2 g in an adult human body.[15] Around 90% of the zinc is found in the musculoskeletal system, with only 0.1% of this element present in plasma. One in ten human proteins requires zinc, participating in critical physiological processes such as providing structure, catalysis and signalling.[30] Like other metals ions, illness can occur either from excess or deficiency, although in this case the latter is more problematic and can even lead to death in developing countries.[31]

Chemical features

Physiologically, Zn is mostly found in a single oxidation state, Zn(II). As Zn(II) does not contain unpaired electrons, with a d^{10} valence shell, it is a diamagnetic redox-inert metal. According to HSAB theory, Zn(II) belongs to the borderline category, being typically coordinated to the S atom of Cys or N from His, both soft ligands, and to carboxylate anions of Asp and Glu residues, which are hard ligands (Table 1.3). It is also a very flexible metal ion in coordination numbers and geometries, with four to six being the most common, either in tetrahedral, square pyramidal or octahedral geometries.[32]

Biological properties

Zn homeostasis needs to be tightly controlled as it naturally acts as an intracellular and intercellular messenger, with more than ten types of Zn transporters and fourteen different ZIP membrane transporters in charge of the intake and excretion of Zn from cells, respectively.[33] Since free Zn is known to cause serious health consequences due to its regulatory role of several physiological processes, such form is present in extremely small concentrations. Hence, most of the Zn concentration is bound to metallothioneins or stored in zincosomes, which are intracell vesicles.[34]

In physiological terms, Zn is essential for a proper biological function, being involved in many biochemical processes and mechanisms, such as signalling, neuronal modulation, and immune and reproductive processes.[35] A classic example of a Zn-dependent protein is the Transcription Factor MTF-1, which contains a Zn finger domain. Upon metal binding, MTF-1 acts as a sensor of zinc concentration and controls the expression of required genes for Zn homeostasis, such as metallothioneins.[35] Zn is also consistently present in enzymatic proteins, whose coordination sphere typically contains one site occupied by a water molecule, which a substrate can easily displace to produce an active catalytic site. A prototypical example is carbonic anhydrase, in which the Zn active site contains a water molecule whose pKa is reduced from 15.7 to 7 upon binding, generating a hydroxide ion. The hydroxide ion is then capable of attacking the carbon dioxide and generating bicarbonate as a product.[36] Finally, Zn has also been related to antioxidant properties, but not due to its electronic configuration: it binds to SH groups of proteins, protecting them against ROS oxidation activity. Besides, it also competes with Cu and Fe for redox metal-binding regions, thus reducing oxidation activity thanks to its inert character.[32]

Regarding the pathophysiological implication of Zn, it must be remarked

that deficiency is more detrimental than overload, contrary to the previous metal ions detailed here. Such deficiency has been long related to retarded growth and development, skin inflammation, and autoimmune diseases (alopecia).[37] It is also linked to increasing ROS formation and high levels of oxidative damage to important biomolecules, including DNA, as low Zn quantities reduce the expression of antioxidant molecules. In the brain, Zn concentration is higher than in other body regions, whose deficiency has also been severely linked with neurological problems. For example, since Zn is an inhibitory modulator of glutamate release and antagonises the activity of redox metals, its deficiency in the brain leads to enhanced susceptibility to glutamate accumulation in the intercellular space and oxidative stress.[12] While Zn deficiency adverse effects have been extensively reported, the consequences of increased concentrations are not clear.

It has been suggested that an excess of Zn could trigger neuronal death, either by itself or by higher concentrations of Zn inducing protein aggregation and precipitation.[10] Indeed, Zn is known to affect the processing of the AD related-protein $A\beta$, as it inhibits γ-secretase and directly binds to $A\beta$ N_{Ter} region, in the same His residues as Cu(II) does, probably leading to rapid aggregation of $A\beta$ into oligomers.[11] Nonetheless, the toxicity of $A\beta$ has been proved to be inhibited by implementing micromolar levels of Zn, as it blocks the calcium channel pore formed by aggregated $A\beta$; but also by competing with Cu(II), preventing the formation of hydroxyl radicals that Cu typically cause.[5] Besides, Zn has also been observed bound to the other AD hallmark, the Tau protein, via Cys and His residues, accelerating its fibrillar aggregation.[11] Accordingly, the effect of Zn in AD can be both beneficial and prejudicial, depending on the extremely sensitive balance of its concentration. In the same sense, higher levels of Zn positively correlate with a decreased risk of PD, as Zn binding to α-synuclein induces fibrillar formation at the expense of reducing oligomers. Coun-

terintuitive as it may seem, such a fact retrieves a protective effect, as oligomers are more neurotoxic than fibres.

As demonstrated in this section, Zn(II) therapeutic effect may rise from its supplementation rather than its sequestering. Some therapeutic strategies with Zn supplementation have already been demonstrated as protective in the treatment of age-related macular degeneration and have beneficial properties for the elderly. However, such supplementation should be carefully applied due to specificity and its relationship with other metal ions homeostasis, as a side-effect consists of Cu deficiency and, in turn, reduced Fe transport, causing anaemia.[22]

1.1.1.4 Aluminium biochemistry

Despite Al being the most abundant metal element on the Earth's crust, species have evolved without including this metal in their physiological processes. Strangely as it may seem, the reason behind lies in the fact that Al was not originally available in a free form, but most of it was interacting with silicic acid. It was not until human intervention that Al started to be present in its soluble form. Al intake happens from multiple sources, seeming to be higher from drinking water than dietary consumption, but also Al-containing drugs are an important income source. Despite there being little -though growing- knowledge about the implications of human exposure to this element, it is expected to cause important effects according to its highly charged nature.

Chemical features

Al is also a transition metal, with a single oxidation state. As HSAB theory states (Table 1.3), Al(III) belongs to the hard metal category, likely to bind hard ligands such as oxygen atoms, especially if negatively charged as in Asp and Glu carboxylates. Typically, the coordination sphere corresponds

to 6-atom ligands in an octahedral geometry, with most vacant sites easily being filled up with water molecules.

Biological properties

A striking difference with the previously described metal ions is the fact that neither homeostasis processes nor physiological roles are attributed to Al(III). Otherwise, most of the studies on Al(III) presence in the body demonstrate its toxicity, being quite remarkable its effect on the brain.[38] Despite how aluminium manages to get into cells is still unknown, it has been demonstrated that it may occupy binding sites meant to be for other metal cations in healthy individuals. Eventually, Al(III) can compete with Fe(III) and Mg(II) binding sites, as Al(III) shares identical ligands.[39] Al(III) toxicity in living organisms can be related to the competition it causes with other metal cations, either by blocking the normal function of bound proteins or by increasing the free amounts of the metal ions it displaces. Thus, its presence in the body is a potential risk without any benefit.

Al(III) is indeed linked to several dysregulations, for example hampering the normal production of red blood cells and leading to microcytic anaemia.[40] Besides, it interferes with several enzymatic activities and neurotransmission signals, promoting hyperphosphorylation of neurofilaments and causing their degeneration. Though the human brain contains minimal amounts of Al at first, it tends to accumulate with age, with late-onset diseases such as PD and AD being strongly related to Al(III).[41] Al was linked to AD for the first time in the earliest 70s[42] and, though not without controversy, further studies have supported the original finding.[43–45] AD patients demonstrate a strong imbalance in Al despite not being yet proven as a causative factor of the disease. Nonetheless, it is undeniable that Al(III) can interact with Aβ through Asp, Glu, Ser and Tyr-rich peptide fragments, which could induce its oligomerisation and

favour toxicity by preventing the formation of fibres.[46] Moreover, such binding is expected to induce a conformational change, which may cause the exposure of hydrophobic patches to the surface of the protein.[22]

According to the highly deleterious role of Al(III), chelator molecules have been supported as a suitable therapy for Al(III) excess-related diseases. Indeed, antioxidants have been demonstrated to form complexes with both Al(III) and Fe(III), whose chelating activity has led to an improvement in those diseases caused by Al overdose.[39] Still, though, much research could be carried on primarily in discovering new chelating strategies, since designed molecules with high Al selectivity would represent a considerable advance, reducing the harmful side-effects.[47]

1.1.2   Structural Study of Metal Ions in Biological Systems

To understand the biological role of metals, it is essential to first probe metal binding to proteins. Several experimental techniques allow for determining the structure of metalloproteins, such as NMR or X-ray crystallography. However, the exact atomic interaction of the metal ion with its protein ligands is often required. Atomistic determination of the metal binding sphere can be obtained with UV-visible spectra, circular dichroism (CD) or electron paramagnetic resonance (EPR) techniques –for paramagnetic metal centres, such as Cu^{2+}- as these approaches are sensitive to the metal coordination geometry and the nature of the ligands.[48] Despite significant advances in such experimental techniques, metalloprotein complexes are usually challenging to resolve, e.g. due to the lability of the interactions or the requirement of high resolution. Some approximations appeal to use proteinomimetics to represent the natural protein, trying to reduce the flexibility that typically hinders their crystal determination. However, such peptides are chemically unstable and prone to

degradation, apart from binding metal ions with lower affinity because of missing part of the real, complete protein scaffold.[49] To overcome such limitations and provide three-dimensional models, computational methods, such as those gathered in the bioinformatics field, can complement experimental data. The next section introduces the daunting challenges that computation faces when dealing with metal ion presence in protein systems. Further technical details and theoretical background of such methodologies will be extensively provided in Chapter 3.

1.2 Dealing with metals: a challenge in Structural Bioinformatics

1.2.1 What is Bioinformatics?

Bioinformatics is an interdisciplinary field that combines biology and computer science to analyse and interpret biological data. The key areas of bioinformatics research include genomics, which consists of sequence analysis and comparison among species; and proteomics, which includes the study of 3D protein structures to analyse the interactions and conformational changes. With the knowledge gathered in both fields, bioinformatics techniques are beneficial for drug discovery and the development of personalised medicine. Considering the objectives of the present Thesis, the focus will be placed on the computational tools applied in the proteomics field.

Proteomics has experienced accelerated growth over the last years, offering excellent opportunities for learning about those diseases that depend on proteins' proper function. Such a field has significant advantages compared to genomics approaches. The structural study of proteins can reveal overlooked traits in their sequence, whose effect may be considerable

for diseases, such as the role of specific mutations or the changes caused by dimerization or interaction with other proteins.

Molecules' structural behaviour, interactions or catalytic reactions can be accurately described using mathematical methods and approximations, which together with their resolution by computer means lead to computational chemistry. Computational chemistry has existed for the last 60 years, since the beginning of computation. Current improvements in computers' performance have allowed an exponential increase in the complexity of the calculations that can be performed, allowing at the same time an increase in the size of the systems that can be tackled and the speed of the calculations. Computational chemistry is essential to Bioinformatics, as proteins' behaviour relies on chemical properties.

Despite the capacity of computers does not stop increasing, the involvement of the human eye is still fundamental in molecular research. The initial description of the system must be performed based on pre-existing knowledge, say, which are the atoms that form the protein and/or how they are distributed. The former is indispensable, so the desired protein sequence must be experimentally known. The second is relative but recommended, since if the position of the atoms is not determined by experimental means -X-ray or NMR techniques-, bioinformatics predictors, such as Modeller[50] or the recently developed AlphaFold,[51] can be applied to predict the secondary and tertiary structure, though at the expense of a certain margin of error. Once the initial structure is properly defined by either procedure, the corresponding theoretical method can be applied according to the aim of the study.

The study of protein complexes' difficulty scales if metals are introduced into the equation, since their delicate electronic structure demonstrates different spin states or open-shell electronic structures. Accordingly, the computational study of metal ions implies the need for including further

adaptations, such as Quantum Mechanics calculations, to carefully assess their binding sphere and coordination mode; while bigger systems require less computationally demanding techniques, like Molecular Mechanics.

1.2.2   Computational methods for biometallic systems

The computational study of biometallic systems can involve different molecular aspects like i) the identification of where and how a metal interacts with the protein, ii) which are the chemical determinants of these interactions or iii) what is the dynamical behaviour of the metal-bound system, eventually compared with the apo one. In the study of many of these magnitudes, computational chemistry faces a challenge.

In this section, an overview of the different aspects computationally dealt with in this Thesis will be provided, specially oriented towards biometallic species. Altogether, these traits set the grounds for the different pipelines designed for each system under investigation.

1.2.2.1   Prediction & description of metal binding sites

The structure of a protein and its dynamical properties primarily determine its biological function, whose ligand binding regions are typically conserved among protein families. The key and lock theory,[52] or the more recent hand-glove hypothesis,[53] states that proteins tend to be preorganised to accept a ligand, though both molecules, the protein and the ligand, may adapt to a certain extent upon binding.

Finding the location of metal binding in proteins throughout blind approaches requires the exploration of a very wide conformational space. This is not achievable with explicit models of the physical behaviour of atoms and molecules (e.g., force field). In such cases, it is essential to

first identify the possible binding sites, either experimentally, which is often a tedious process, or through computational methods. Software programs like MIB[54] or BioMetAll[55] help in this process by performing a preliminary screen of the protein tertiary structure to detect pre-organized metal-compatible binding sites. Several binding sites can be found along the protein system, requiring further exploration to determine their characteristics for metal binding.

Once the binding region is detected, the coordination mode between the ligand and the surrounding residues has to be described. For that, Molecular Dockings represent an essential technique,[56] since its algorithm relies on the protein structure to predict the most likely poses for ligand compounds, accounting for the interactions between both moieties. However, additional adaptations are required for metal ions, as most Docking software, such as GOLD[57] or AutoDock Vina[58], do not allow modelling metal-protein interactions by themselves. Besides, metal ligands usually lack proper parametrization in the scoring functions applied to rank the best binding poses. To deal with such limitations, a dummy atom approach was proposed in the work of Sciortino et al.,[59] providing parameters for several metal-protein interactions (Figure 1.5).

1.2.2.2 Computational simulations of metalloproteins

Once the binding site is identified, the behaviour of the metalloprotein system can be explored. For that, different levels of theory can be applied depending on the aim of the study. On the one side, molecular mechanics (MM) calculations allow the exploration of the conformational landscape of large systems, as the interactions among atoms are described through force fields. On the other, the description of chemical reactions requires quantum mechanical calculations (QM) to describe the changes in the electronic structure along the reaction. These calculations, however, are
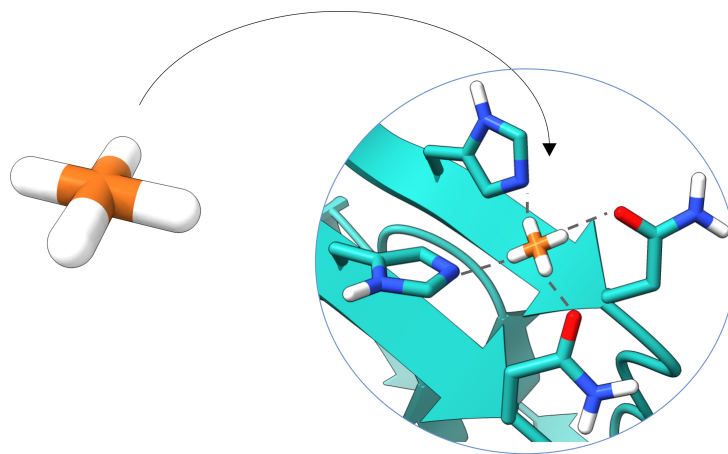
Figure 1.5: Example of dummy-atom strategy. H atoms are bound to the Cu atom in a square planar distribution, so the docking search can look for the specified geometry in the protein structure.

much more computationally costly and are limited to small systems (hundreds of atoms). Thus, multiscale approaches are commonly used when analyzing the reactivity of metalloproteins.

In case one wants to treat the whole system with MM calculations, which represent a very low computational cost, it has to be considered that force field selection is a critical step in Molecular Dynamics (MD) simulations, as parameters strongly determine the outcome of the simulation. Nowadays, several force fields are at disposal, such as AmberFF19SB[60] or Charmm36m[61], which aim at describing general complexes; or ff14IDPFF[62], specifically designed for intrinsically disordered proteins (IDP) -or peptides-, according to their characteristic folding pattern. Accordingly, a particular force field must be chosen in agreement with the system's needs. However, it should be highlighted that most forcefields lack metal ions parametrizations, a drawback for the study of metalloproteins. Besides, transition metal ions present multiple oxida-

tion states, which cannot be reproduced in forcefields if the same atom should present different states along the simulation. Therefore, several approaches have been developed (see Chapter 3) to overcome such limitations, which include i) bonded models, ii) non-bonded models and iii) dummy atoms models. All those methods are based on creating parameters for each metal ion and the coordination sphere under study employing QM calculations, so that the electronic effects of a particular coordination are encoded in the force field.

Once the metal ion is included in the force field parametrization, the impact of metal binding on the protein structure can be evaluated with molecular dynamics simulations MDs. With such a technique, an exhaustive exploration of the conformational landscape is performed to detect the lowest energy structures and thus, the most stable populations.

It should be highlighted that some protein systems -especially metallopeptides- may show highly flexible conformations, implying that the MD conformational exploration can easily get stuck in a local minimum, instead of exploring the whole conformational landscape and finding the real minima. Several techniques have been developed aiming at enhancing the exploration and overcoming local barriers, such as the widely known Gaussian accelerated Molecular Dynamics (GaMD)[63] or Replica Exchange Molecular Dynamics (REMD)[64]. In GaMD, a boost potential is applied to the potential energy, forcing the system to escape from local minima and continue the exploration of the conformational landscape; while REMD runs parallel replicas of the same system at different conditions (temperature, pressure...), interchanging their coordinates at specified times. With the implementation of enhanced exploration techniques, the conformational impact of metal binding to peptides or protein systems can be successfully determined.

While MM methods are capable of describing the conformational changes

that a system may undergo upon metal binding, they lack the capacity to describe electronic changes, such as those happening in chemical reactions. In those cases, QM calculations are successfully applied in the study of metal binding to protein systems, since the complexity of metal properties, and in particular transition metals containing systems, require a fine electronic description.[65] However, it is impractical to consider large systems with QM, as it is limited to handling hundreds of atoms. For that, hybrid methods appear to be a cost-effective strategy, in which the region involved in the chemical reaction is described with QM methods, while the rest of the system is treated with MM. When studying the binding of metal ions into a protein complex, the QM part is typically limited to just the metal ion and its ligand residues - the coordination sphere-; while the rest of the system is described in MM terms. With such a QM/MM strategy, the challenge lies in how to describe the boundary region (see chapter 3 for further details). Altogether, the combination of all these strategies -Docking, QM and MD- along the different stages of a project is regarded as an integrative approach (Figure 1.6). In the present Thesis, different integrative approaches have been applied to several relevant biological processes.

1.3   Where do we place the focus on?

In this Chapter, the importance of metal ions in human health and the challenges they represent for their computational study have been highlighted. As demonstrated, the topic of metal ions is so extensive and with such broad implications that it can be tackled from several perspectives. In the present Thesis, the intention was to learn and apply different techniques in order to describe not only the detrimental side of metal ions but also both their possible beneficial roles and to help in the design of new techniques that could further help in their study. Accordingly, three dif-

ferent relevant studies involving metalloproteins have been performed.

1. **Role of metal ions on Aβ_{42} aggregation.** The aggregation of the Aβ peptide has been classically related to Alzheimer's Disease. Though how and why the aggregation starts is still widely unknown, the metal ion hypothesis highlights the role that metal binding to Aβ may have regarding the aggregation process. Considering that changes in the sequence of the Aβ also lead to different aggregation patterns, which could provide valuable information about the behaviour of this peptide, familial mutations, with and without metal binding, have also been studied.

2. **Catalytic metallofibers.** Experimental studies in the last years have shown that amyloid-like minimalist fibres retain the capacity to catalyze several chemical reactions in the presence of metal cations. Accordingly, two different catalytic roles of heptapeptide fibrillar structures are studied. First, the hydrolytic reaction of pNPA mediated by Zn on different hydrophobic and prion-like heptapeptide fibres has been studied, in collaboration with the groups of Korendovych et al. and Ventura et al.. Furthermore, since it has been experimentally demonstrated that modifying the 6th residue alters the efficiency of the catalytic process, the role of such position has also been evaluated.[66] Second, the cyclopropanation of aromatic styrene mediated by Zn binding is studied, as different peptide sequences have been proven to retrieve different enantioselectivity.[18]

3. **Metal diffusion pathways**. Finally, we would not like to miss the opportunity of providing a new tool that paves the way for further studies involving metal moieties. Considering that the activation of several catalytic metalloenzymes depends on the final maturation of the protein upon metal binding, entrance channels should be available in the already preorganized protein. Hence, a pipeline

for detecting such metal ions pathways is developed, which would generate useful information for the design of alternative therapeutical approaches to chelators, as blocking such entrances would also avoid enzymatic activity.
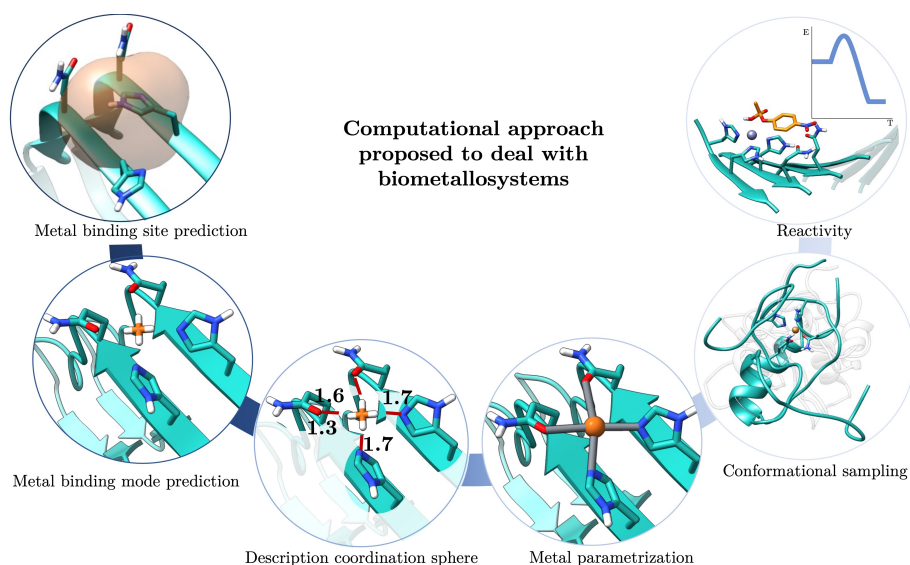
Figure 1.6: Computational pipeline proposed in the present Thesis to deal with metallosystems, consisting of six steps: metal binding site prediction, metal binding mode prediction, description of the coordination sphere, metal parametrization, conformational sampling and reactivity.

2

Objectives

Understanding how metal interacts with biological systems is essential because of their physiological and technological impacts. Still, much remains unknown about the structural effects that metal ions induce upon binding, considering the experimental challenge that their study represents. The main objective of this PhD Thesis is focused on developing and applying multi-scale protocols for the study and design of biometallic systems, hence overcoming the current capacities of modelling. The work performed can be divided into three different topics.

1. **Metal binding to monomeric and fibrillar forms of Aβ_{42}**

   - Observe the effect that Cu and Al have upon binding to monomeric Aβ_{42}.

   - Predict the binding mode of Cu and Al to the fibrillar form of Aβ_{42}, as well as determine their conformational impact.

   - Describe the effect of familial AD variants (E22Q, E22G, E22K and D23N), with and without bound metal ions, both in the monomeric and fibrillar structures.

2. **Catalytic metallofibres**

   - Provide a structural prediction of hydrophobic (IH7, IHQ7, IHY7, IHR7) heptapeptide fibres and Zn(II) interaction.

   - Study the catalytic mechanism of para-nitrophenyl acetate (pNPA) hydrolysis by IH7 fibre, and compare it to that of the SY7 prion-like fibre, in the presence of Zn(II).

   - Determine the role of the 6th residue in the fibrillar heptapeptide (IHQ7) on the pNPA hydrolysis.

   - Discern the determining role of the heptapeptide sequence on the cyclopropanation enantioselectivity.

3. **Metal recruitment in metalloenzymes**

   - Set up of a novel computational approach to predict metal ion migration paths in protein structures.

   - Test the applicability of the novel computational approach on the Tyrosinase enzyme and investigate the evolutive adaptation for metal recruitment.

3

Computational Background

This Chapter presents a theoretical overview of the computational methods used in this thesis. It focuses on the key concepts and sufficient details to comprehend their meaning inside the entire framework as well as their limitations. Instead of providing all the details, particular information will be described in the corresponding chapters of each research topic.

Molecular Modelling

Combining mathematical methods with fundamental physics allows the prediction and description of molecular behaviour under a series of approaches whose unified term is Molecular Modeling. There are mainly two levels of theory that can be applied: Quantum Mechanics (QM) and Molecular Mechanics (MM). QM calculations allow describing the electronic structure of the system, i.e. the functions associated with the behaviour of the electrons are explicitly computed. This is essential for any system requiring changes in chemical states, such as reactivity, redox potential, etc. In contrast, MM adopts a classical description of the systems without an explicit calculation of the electronic wave function. Instead, MM methods consider the atom as a whole with a potential associated with the bonded and unbounded network of interactions. These methods are suitable for studying global movements and conformational changes for large systems, even reaching physiologically relevant times at the atomic level. Hence, the different levels of theory need to be carefully chosen according to the chemical phenomena under research.

3.1 Quantum Mechanics

With QM methods, the nuclei and electrons of each atom are explicitly treated, allowing the description of those properties related to electronic distribution. On the counterpart, they are highly computationally demanding and limited to small systems, usually up to hundreds of atoms. Besides, QM has been extensively used in the last years for the study of metal properties, since subtle electronic effects are indispensable for the proper definition of its coordination sphere and geometry. The most common methods to describe molecular systems at the QM level are divided into two groups, based on the *wave function* or on the *Electron Density*.[67,68]

3.1.1 Wave Function Methods

The first principle of QM is based on the fact that the state of a system can be described by a wavefunction ψ dependent on the position of the particles at a particular time. To accurately describe the electron distribution of a certain particle, in most quantum chemical applications, QM methods need to solve the time-independent non-relativistic Schrödinger equation:

$$H\psi(r,R) = E\psi(r,R)$$

where H is the Hamiltonian operator of the system, E is the total energy and ψ is the wave function of the system, which depends both on the position of the nuclei and the electrons.

The Hamiltonian is composed of several terms, observed in Equation 3.1, where the first two terms are the kinetic energy of the electrons (T_E) and the nuclei (T_N). The remaining ones are the potential energy resulting from the attraction between electrons and nuclei (V_{EN}), the repulsion between electrons (V_{EE}), and the repulsion between nuclei (V_{NN}).[69]

$$H = T_E + T_N + V_{EN} + V_{EE} + V_{NN} \tag{3.1}$$

This equation, though, cannot be solved exactly for molecular systems, so several approximations can be applied. The first one is the Born-Oppenheimer approximation[70], which is based on the different time scale motions between nuclei and electrons. Due to the mass of an atomic nucleus being 1836 times larger than the one of electrons, its movement is much slower, thereby allowing it to treat the two motions (nuclear

and electronic) separately. Thus, the Born-Oppenheimer approximation states that the wavefunction can be expressed as the product of the electronic wavefunction (ψ_{el}) and the nuclear wavefunction (ψ_{nuc}) (Equation 3.2).

$$\psi_{tot}(r,R) = \psi_{el}(r;R)\psi_{nuc}(R) \tag{3.2}$$

Within this approximation, the nuclear motion is considered slow enough for the electrons to perceive the nuclei as fixed positions. Thus, the nuclei's kinetic energy is equated to 0, while the repulsion between nuclei is constant. Therefore, the computational effort is placed on just solving the electronic part of the Schrödinger equation, which explicitly depends on the electronic coordinates (r), at a particular fixed geometry (R).

$$(H_e + V_{NN}(R))\psi_e(r;R) = U(R)\psi_e(r;R) \tag{3.3}$$

being U(R) is the potential energy.

If many different nuclear coordinates of the same system are singularly computed, the Potential Energy Surface (PES) is obtained. Relevant points in the PES are the minima, which correspond to reactants, products, intermediates, and first-order saddle points, with one imaginary frequency, which relate to transition structures that connect different minima during a reaction. With PES information, a reaction's profile can be properly described.[71]

Despite the Born-Oppenheimer approximation, it is still impossible to exactly solve the Schrödinger equation if more than one electron is considered simultaneously. Accordingly, further approximations are required, such as the Hartree-Fock method.

The **Hartree-Fock**(HF)[72] method, also known as the self-consistent field (SCF), is the simplest *ab initio* method. It considers the wave function as an antisymmetrized product of spin-orbitals expanded as a linear combination of atomic basis functions; that is, as a Slater determinant, and computes the energy variationally. This method assumes that each electron in the system interacts with the average (self-consistent) field created by all other electrons. An infinite number of base functions would then be necessary to properly describe the electron density, known as the Hartree-Fock limit. Nonetheless, the Hartree-Fock limit is unfeasible, which requires limiting the expansion to a certain number of base functions.

The smallest set of base functions is known as the minimum basis, which only includes a single basis function per atomic orbital. Extended basis sets, however, should be included to get accurate results. These are named double $- \zeta$, triple-ζ, etc. depending on whether each atomic orbital is described by two, three, etc. basis functions, respectively. Further improvements including higher angular momentum functions, called polarization functions, are also necessary to describe the distortion/polarization of atomic orbitals as a result of bonding. Thus, choosing the proper base for the system under study is strictly mandatory, as it may clearly influence the outcome of the computation.

Two types of functions are the most used to compute the electronic feature of molecules: the Slater-type Orbital (STO) and the Gaussian-type Orbital (GTO). While STO properly describes the behaviour of electrons around the nuclei, they are expensive in terms of computation. On the contrary, GTO is less accurate when describing the electronic movement around the nuclei, but is far cheaper to compute. For that reason, GTO is typically linearly combined until a similar result to STO is obtained. With the described functions, canonical atoms can be suitably described, but special attention must be given to transition metals. Those elements

contain a remarkable number of internal electrons, which can be considered inert in terms of chemistry activity and can be represented through *Effective Core Potential* (ECP), allowing to simplify the computation and represent just their valence electrons explicitly and include scalar relativistic effects. However, though increasing the size of the basis set is expected to improve the results, an exact result would not be achieved even if an infinite basis set were used ($E_{HFlimit}$), since the Hartree-Fock method neglects the electron correlation.

In order to account for electron correlation in wavefunctions, different approaches have been developed, which can be classified as i) those trying to correct the single Slater determinant approximation by expanding the wavefunction (post-Hartree-Fock methods) and ii) those based on the density functional theory (DFT). Density functional methods (DFT) represent an efficient alternative to post-Hartree-Fock methods for recovering electron correlation and thus, they are the ones used in the present Thesis. Accordingly, a brief explanation of DFT is introduced in the next section.

3.1.2   Density Functional Methods

The current Density Functional Theory (DFT)[73] is based on two theorems proposed by Hohenberg and Kohn in 1964. Such theorems state that the energy of the ground electronic state is a unique function of the electron density in the fundamental state and that any given trial electron density may be variationally optimized towards the real electron density, provided the exact functional is known. Therefore, the electron density depends only on the three spatial coordinates, independently of the electron number, greatly simplifying the complexity of the wave function computation: instead of exponentially increasing with the number of electrons, DFT is independent of the size of the system. The problem then lies in

finding the functional that effectively relates the density with the energy.

For this latter fact, the posterior Kohn and Sham (KS) method reintroduces Slater determinants of spin-orbitals, for which a functional relation to electron kinetic energy is known, $T_S[\rho]$. The Kohn and Sham formulation involves the use of a fictitious noninteracting electron system that has the same density as the real system. Since the kinetic energy of the reference system with non-interacting electrons is different from the real kinetic energy, such difference is added to the E_{xc} exchange-correlation functional, which includes the potential energy contribution corresponding to the difference between the real electron-electron interaction potential and that of the reference system. Hence, the success of the DFT calculation depends on finding suitable functionals for E_{xc}, for which several approximations are available: a) Local Density Approximation (LDA), which only depends on the electron density r; b) Generalized Gradient Approximation (GGA), which improves LDA as it not only considers the electron density but also the gradient of the electron density along the space; c) Meta-GGA Approximation, which incorporates a kinetic energy operator and includes the Laplacian of the electron density $\nabla^2(\rho)$ or d) the Hybrid functional, that accounts for non-locality by introducing a certain percentage of exact exchange in the functional. The popular B3LYP is part of this latter group, known for its suitability to explain most chemical reactions, extensively applied for organometallic systems.

In summary, DFT approaches are a middle point between HF and post-HF methods, with high dependency on the functional. In the present Thesis, DFT calculations with B3LYP functional are applied to obtain force field parameters. However, the DFT method is deficient in dispersion terms, which are relevant in biological systems, hence Grimme's correction for dispersion has also been applied in the present work.

3.2   Molecular Mechanics

3.2.1   Introduction

When larger systems such as biomolecules are the target of study, QM approaches are too demanding in terms of computation for their description, since the number of atoms extends the limit that could be handled. For those cases, Molecular Mechanics (MM) calculations represent a more appropriate tool, as electrons are not considered explicitly but simplified to classical particles. Hence, particles' behaviours are described through 1) their inner physical traits, such as mass, charge, and radius; and 2) interactions among them, described by parameters. Thus, as the physical traits can be treated with classical laws of physics, the computational cost is greatly reduced, allowing a significant increase in the size of the system under study, even being capable of dealing with whole biomolecules. However, the direct consequence of such simplification is that as bonds are defined at the very initial moment, chemical reactions – which imply the formation and breakage of bonds – are unable to be described.[74,75]

Molecular Mechanics relies on the accurate description of the structural properties of the system's molecules contained in force fields. By solving the potential energy equation, force fields describe the energy of a system depending on its conformation, aiming at reproducing the structural properties of molecules already defined, either empirically, by quantum calculations, or a mixture of both. To compute the total potential energy of the system (Equation 3.4), a sum of bonded and non-bonded energy contributions is evaluated for each atom building the system:

$$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{cross} + E_{Hbond} + E_{VdW} + E_{elec} \qquad (3.4)$$

Where bonded interactions account for those atoms directly connected, generally composed by E_{str}, the energy for stretching a bond; E_{bend}, the energy for bending an angle; and E_{tors}, the torsional energy of dihedral atoms separated by three bonds. E_{str} and E_{bend} energies are usually represented as harmonic potentials (Equation 3.5 and Equation 3.6), while E_{tors} uses a cosine series expansion form to reflect its periodicity (Equation 3.7).

$$E^{stretch}(r) = \frac{1}{2}k^{stretch}(r - r^{eq})^2 \qquad (3.5)$$

$$E^{bend}(\alpha) = \frac{1}{2}k^{bend}(\alpha - \alpha^{eq})^2 \qquad (3.6)$$

$$E^{torsion}(\omega) = k^{torsion}[1 - cos3(\omega - \omega^{eq})] \qquad (3.7)$$

where the ^{eq} represents the equilibrium values; that is, the ideal values for each bond length and bond angle, while k components are force constants. r stands for the distance between two atoms, α corresponds to the angle between two atoms, and ω to the torsion angle among three atoms.

On the other hand, non-bonded interaction terms account for the interactions between not-connected atoms, resulting in the sum of hydrogen bonds (E_{Hbond}), Van der Waals (E_{VdW}) and electrostatics contribution (E_{elec}).

Van der Waals interactions are a sum of both repulsive and attractive terms, which are commonly reproduced by Lennard-Jones potential, describing them as a short-range repulsion zone in the closest area between atoms and a slowly decreasing attraction zone, ultimately reaching a negligible value (Equation 3.8). On the other side, Coulombic terms account for the interaction between charges. Such electrostatic interactions are treated with the Coulomb potential (Equation 3.9), which depends on the

point charges of two given atoms q^i and q^j at a certain distance r.

$$E^{VdW}(r) = \varepsilon \left[\left(\frac{r^0}{r} \right)^{12} - 2 \left(\frac{r^0}{r} \right)^6 \right] \tag{3.8}$$

$$E^{Coulombic}(r) = \frac{qq'}{r} \tag{3.9}$$

As atomic charges are not physically observable, the Restrained Electrostatic Potential (RESP)[76] is usually applied, which fits the q^i to reproduce the electrostatic potential computed with QM. In either approximation, a cut-off value is usually set to discriminate the long-range interactions beyond a determined value, remarkably reducing the computational cost.[75] A representation of each energetic term and its correspondence in a molecular example is provided in Figure 3.1.
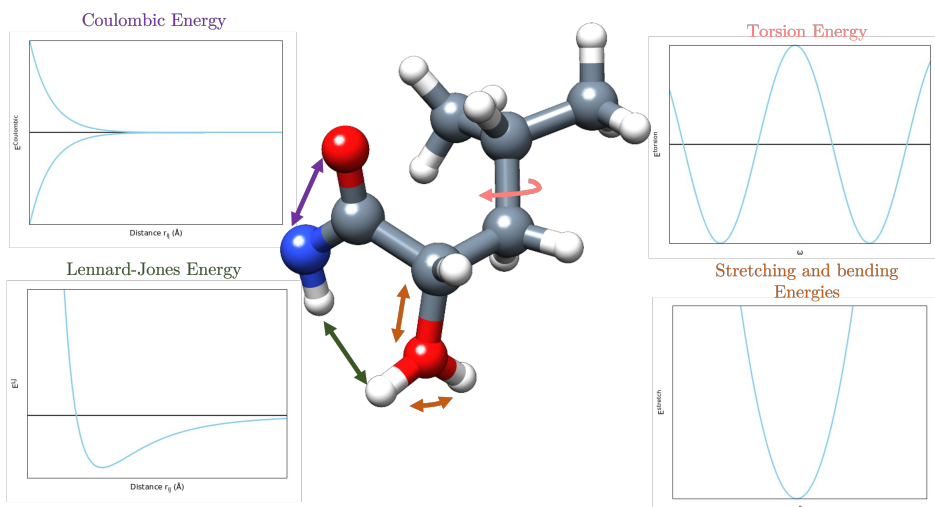

Figure 3.1: Example of energetic terms' evolution along their x-axis and their correspondence in a molecular example.

Some additional terms to the already described ones can be further included in the equation, such as improper torsions and cross terms. While impropers account for torsion angles among unbounded atoms, either from the same molecule or between interacting regions, in order to represent those interactions that could move out of the plane; cross terms correct the interrelation between E_{str}, E_{bend} and E_{tors} (e.g. stretch-stretch, stretch-torsion, and bend-bend interactions).

The estimation of the potential energy computed by adding the whole set of functions above mentioned constitutes the *force fields*. In those functions, parameters are needed for each atom, though it is considered that the same atom types maintain constant structural properties when located in a similar environment. Therefore, there are appropriate force fields depending on the targeting system, as their parameters are designed specifically for similar sets of molecules.

Nowadays there are many force fields available, some with generic purposes, others focused mostly on biomolecular systems. Such is the case of GROMOS,[77] OPLS,[78] AMBER,[79] and CHARMM,[80] which support a great variety of molecules including proteins, lipids, and nucleic acids. There are more specific ones aiming at certain types of proteins, such as intrinsically disordered proteins,[62,81] which include improvements for those peptides with a marked tendency to disorganization. In the last years, even machine learning approaches have been applied to develop more versatile force fields.[82]

Once the proper force field is selected it can be applied in many techniques, such as the conformational exploration of biomolecules with Molecular Dynamics simulations, the study of Membrane Contacts, or the search of binding poses with Molecular Dockings, as it will be detailed in the next sections. However, it should be mentioned that the applicability of the technique is still limited in several aspects, for example, both

in describing bond-breaking processes and by the range of the force field parametrization.[83] In the frame of this last limitation, most force fields lack an appropriate parametrization for systems with metal moieties or ligands, which represents a drawback for bioinorganic applications.

3.2.2   Metals in Force Fields

Due to the increasing importance of metals in biomolecular systems, several alternatives can solve the unavailability of metal parameters in force fields, including i) bonded models, ii) non-bonded models, and iii) dummy atoms models.

Bonded models[84] are the most similar approach to force fields, as fixed bonds are constructed between the metal ion and the protein ligands, in the proper coordination geometry, through QM calculations. This is a suitable approach for studying the effect of metal ions on the structural disposition of the protein, as they are kept attached to the protein while it undergoes conformational changes. On the contrary, as the covalent bonds are predefined, these models fail at allowing ligand exchange or interconversion of coordination geometries along the simulation.

A totally opposite strategy is the non-bonded model, either with 12-6 Lennard-Jones[85] or the improved 12-6-4 LJ,[86] which includes an additional parameter to account for dipoles' interactions. With these models, the metal ion is represented as a set of compressible spheres, with LJ potential properties. The metal is weakly bound to the protein through VdW and electrostatic interactions, allowing flexibility of the coordination sphere, representing a highly used approach due to its computational efficiency and its easy transferability between metal ions. Nonetheless, transition metal ions require a coordination geometry and present different affinities for protein ligands, which are misrepresented in this model.

Dummy atoms models are an intermedia solution between the previous methods.[87,88] In this case, artificial atoms are attached around a central atom - the metal ion -, in a disposition that represents the appropriate coordination geometry and allows weak interactions with the corresponding protein ligands. Despite dummy atoms being adjusted to the desired metal coordination, atom ligands are interchangeable during the simulation, combining the advantages of both bonded and non-bonded models. Recently, polarizable dummy-model approaches have also been reproduced,[89] introducing the possibility that charges of the metal ion and dummy atoms fluctuate in response to the environment. The main disadvantage of this model lies in its intricate parametrization process, as many empirical parameters are required.

In the present Thesis, since the aim is to maintain a metal-binding interaction that lasts along the simulation and allows to study its impact on the protein conformation, bonded models are applied. The parameters for such models can be derived using tools based on Seminario's method,[90] such as FUERZA or the MCPB.py (Metal Center Parameter Builder)[91] script by Amber suite, which has been used in the present work.

MCPB.py software creates two models from the input structure. First, a small model is built to obtain just the bond and angles associated with the metal binding sphere. To build the small model, the real chemical environment of the metal is mimicked by simpler representations to reduce the computational cost. Second, a larger model including the backbone of the involved residues is generated to obtain, with a larger basis set, the partial charges of the atoms with RESP charge fits. In order to simplify the large model calculation, the residues in the metal site are capped by ACE and NME groups or alternatively, if the residues are bound between them for less than 5 residues, the non-interacting intermediates are simplified to Glycine residues. Once both models are generated, they

undergo Gaussian calculations. Then, the parameters file (frcmod) is obtained from the QM calculation on the small model, which describes the bonds, angles, torsions, and impropers of the metallic geometric sphere; while mol2 files, with atom charges, are obtained from the RESP calculation over the large model. The final parameters are then integrated into the user-chosen force field, where the parameters for the rest of the system are contained, and parameters and coordinates files for the whole system are finally achieved.

The basis of Molecular Mechanics explained here is applied to further develop exploration techniques. The most popular ones are Molecular Docking, which allows one to predict the interaction between two molecular moieties; while others, such as Molecular Dynamics, allow a more complex exploration of the systems, even over time. It is worth mentioning, though, that for all these techniques it is required to have an initial structure of the system under study to start the calculation. Nowadays there are already more than 200,000 3D protein structures disposable in the Protein Data Bank (PDB), but in the absence of the desired 3D structures, Homology Modelling techniques or the recently released Artificial Intelligence structure predictors, such as AlphaFold, can be useful to model the systems. Those tools, in short, rely on the correlation among sequences and secondary structures to approximately predict the structure of the unknown sequence by comparison to template structures, providing a suitable model from which calculations can be applied.

3.3   Molecular Docking

3.3.1   Introduction

Molecular docking aims at providing a fast and accurate prediction of the binding mode and position/orientation (*poses*) between two molecules,

corresponding to a metal ion and a protein/peptide in the current work. Their most extended use, though not exclusively, consists of modelling the interaction between a small molecule and a protein at the atomic level. Dockings' applications are extensive, being an increasingly important tool for drug discovery, especially regarding Virtual Screening computations (VS). In VS, every compound of a large library is tested against a target/receptor in a short period of time, distinguishing among binding and non-binding molecules, exclusively ranking the first ones. In order to properly perform their function, docking programs need not only to maintain a good balance between computational cost and accurate predictions of the docking results but also reproduce the experimental binding modes of the ligands. To do so, the docking process involves two basic steps: first, the prediction of the small molecule –*ligand* hereafter- conformation as well as its pose within the receptor protein, performed by search algorithms; second, the assessment of the binding affinity for each predicted pose, determined by a scoring function.[92]

As mentioned in chapter 1, the 3D structure of the receptor should be achieved by any of the methods available, either experimentally or through computation. Then, several types of dockings can be performed depending on the existing knowledge of the binding site. The most computer-demanding approach is *blind docking*, in which the binding site is searched at every possible position of the receptor, without any preliminary assumption at all. For certain structures, though, the binding site location could have been already determined by experimental means or a concrete area may show higher chances of binding the ligand, significantly increasing computational efficiency. Once the search area is defined, the protein-ligand docking procedure can be followed in two different manners:

1. Rigid docking is the least computer-demanding approach, as both the ligand and the receptor are rigid, and only translational and rotational freedoms are allowed. However, this is the least accurate approximation, providing a rough prediction which should then be submitted to subsequent refinement.

2. Flexible dockings can consider just ligand or both ligand and receptor's flexibility, but the limitation of computer resources still commonly forces to further constrain the technique and reduce the mobility of the host system. As subtle protein movements impact severely on molecular docking results, several strategies are available to introduce protein flexibility up to a certain extent:[93]

   (a) Soft docking strategy relaxes the repulsive term in the Lennard-Jones potential in order to allow certain overlapping between protein and ligand atoms. Though rapid, this method should be applied only to local motions.

   (b) Side-chain flexibility is explored by modifying torsional angles and freezing the backbone of the residues in the active site. Such conformations can be included either by the use of rotamer libraries, with classified probabilities for each residue's sidechain position, or a total free rotation movement of the angles.

   (c) Molecular relaxation is performed to optimise the protein-ligand conformation obtained through rigid-protein methods, usually as a post-processing step to evaluate the stability of the docked complexes.

   (d) Ensemble docking tests the ligand over a set of protein conformations, which can be obtained either by experimental means

or computational samples. In this case, the difficulty is placed on the scoring function that should be applied to analyse and compare the outcome structures, as the binding environment may not be the same. Besides, the computational cost increases with the ensemble size.

(e) Collective degrees of freedom consider the full protein flexibility but reduce the protein motion to just the dominant ones, similar to the approximation applied in normal mode analysis or principal component analysis. In this case, inaccuracies may be introduced as the movements considered are not native ones but derived from collective movements.

Once the search space is defined and the docking flexibility chosen, space exploration can be performed. The algorithms that carry the space exploration can be classified from lower to highest complexity in i) shape matching, which is the simplest sampling algorithm as it places the ligand in a complementary region of the protein's molecular surface; ii) systematic search, which is highly exhaustive and typically used for flexible-ligand dockings to generate as many binding conformations as possible, then test each of them; iii) stochastic algorithms, in which random changes are made over the ligand both in the conformational and translational space at successive runs, reducing the poses to be searched and substantially speeding up the computation; and iv) deterministic methods, as the state of the system directs the modifications that should be done to obtain the next state, a technique that is strongly limited by the input structure and the impossibility to explore other local minima.[93,94]

Once the poses are found, the next key element of receptor-ligand docking is their accurate evaluation with a scoring function, the most challenging part of the algorithm as rather than calculating the real binding

affinity, it is estimated through simplifications. As no universal scoring function can be applied to molecular docking studies, the one that best adapts to the system under investigation should be chosen. Therefore, there are a great number of scoring functions, which can be grouped into three categories:[94]

1. Force Field Scoring Functions: based on decomposing the ligand binding energy into the non-bonded energetic terms that define force fields. They consider each receptor-ligand atom pair interaction, though they fail at accounting for solvent and entropic effects. GOLDScore (Equation 3.10) is an example of this scoring function.[95]

$$GOLDFitness = S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int} \qquad (3.10)$$

   where S_{hb_ext} is the protein-ligand hydrogen-bond score and S_{vdw_ext} is the protein-ligand Van der Waals score. S_{hb_int} is the contribution of the intramolecular hydrogen bonds of the ligand and S_{vdw_int} is the intramolecular strain in the ligand.

2. Empirical scoring functions: the binding energy is computed relying on weighted energy terms obtained from empirical data, simplifying the previous approximation. Its accuracy, though, depends on the training set used for their correct determination. ChemScore is an example of this scoring function.

3. Knowledge-based scoring functions: a statistical analysis is applied to obtain pairwise contact frequencies and distances, assuming that more favourable interactions demonstrate higher occurrence frequency. PMF is an example of a knowledge-based function.

In the present Thesis, stochastic algorithms have been used by applying the GOLD software (Genetic Optimization for Ligand Docking).[57] In GOLD, the binding search area has to be determined by the user as a search sphere, inside which the solvent-accessible surface area and cavities of a certain wide are considered. Within this space, all hydrogen-bond donors and acceptors atoms are identified according to their atom type. A maximum of 10 residues inside the search space can be defined as flexible sidechains, either by introducing a rotamer library -to limit the flexibility to those poses already observed experimentally- or by total free rotamers. Thus, GOLD software performs flexible-ligand docking, though accepting certain flexibility in the receptor. The ligand is also pre-processed, detecting hydrogen-bond donors and acceptors, and minimizing its structure. Once the molecules are prepared and the space search is determined, a Genetic Algorithm (GA) is employed for space exploration. In GA, each location and conformation of the ligand is considered a *chromosome*, which can be mutated (changed randomly) and crossover (exchanged with other chromosomes) in order to generate new random poses. In each run, the best poses will survive to the next generation, until the number of determined generations is completed and the best results are evaluated with the chosen scoring functions, giving a ranked list of results. For the present work, the force field-based GOLDscore function has been preferred.

3.3.2   Metals in Molecular Docking

The docking procedure is well-established and straightforwardly applied for most ligand molecules. However, dealing with transition metals represents a daunting challenge, as they are not the typical ligands for which docking software is designed. Over the last decades, there has been an increasing interest in modelling metallosystems, also favouring the development of some strategies to tackle the search for metal binding regions.

First, the experimentally determined metal-binding sites are mostly defined for those systems in which the metal ion is fundamental for their function or folding, but it is unknown for those in which the metal may not be present in canonical forms (i.e. Aβ fibrils). For those cases, prior to Molecular Docking calculations, the prediction of the preorganized binding areas for certain metal ions is extremely useful, as a reduced and feasible number of search areas would then be input to Molecular Docking simulations.

For those cases, BioMetAll software[55] represents a major advance in the previous steps to docking computations. With proteins as receptors, BioMetAll scans the 3D structure of the protein to find compatible binding sites either for generic transition metals or, if a motif and number of residues are specified, for a concrete metal ion (Figure 3.2). Besides, BioMetAll only considers the backbone of the protein, determining an area as adequate for metal binding if in the surrounding space is possible to find the sidechain of compatible residues. Hence, even though the protein remains fixed during the computation, as the binding distance of each residue is computed from the backbone and not the sidechain, it implicitly considers residues as completely flexible, retrieving extremely fast computations (in the range of seconds).

As mentioned, just some of the preferred geometry for metallic interaction are predefined in force fields, implying the necessity of drawing on various strategies to expand the exploration range and include missing types. One of these strategies compatible with GOLD docking software is the H-bond trick,[59] which assumes that hydrogen bond functions can model the coordination bonds in directionality and polarity. Then, the desired binding geometry is provided by the user by distributing Hydrogen atoms around the metal ion, which is successfully implemented in GOLD and accurately evaluated with GOLDscore scoring function (Figure 3.2).
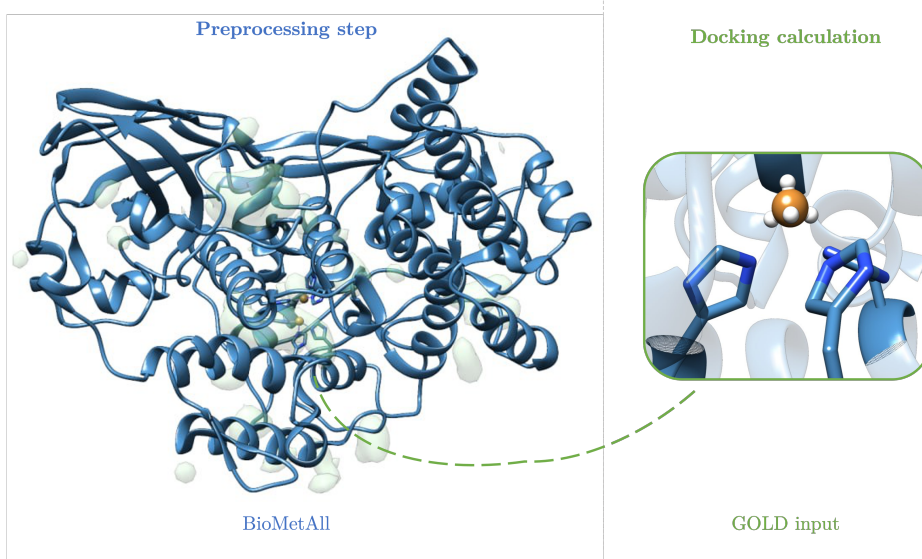
Figure 3.2: BioMetAll output, highlighted areas are metal-binding compatible. Preprocessing step: In this example (Labster Hemocyanin, PDB code 6L8S) the most populated areas correspond to the actual metal binding site, which is already determined by experimental means. Docking calculations can be performed once the most probable metal-binding regions are detected. GOLD docking software requires the H-bond approximation to deal with metals, in which the desired metal binding geometry is provided by the user by distributing H atoms in the proper geometry around the metal ion. In this image, a Cu metal ion with tetrahedral coordination is observed.

3.4 Molecular Dynamics

First developed in the 70s, molecular dynamics simulation (MDs) is a widely applied technique which predicts protein atoms' movement over time based on molecular mechanics. Since its origin, MDs have significantly improved from simulating just several hundred atoms to full biological complexes, with present simulation times close to biologically relevant ones. A variety of biomolecular processes can be simulated, such

as protein folding, conformational change and biomolecular response to perturbations.[83,96] In real-life conditions, protein changes occur along different time scales, usually being extremely challenging to observe them through experimental methods, or may need to be observed under exact conditions, which are not entirely replicable in a laboratory. Therefore, MD simulations can overcome such limitations, depicting proteins' intrinsic flexibility and function-related conformational changes under certain conditions and over a period of time.

3.4.1   Integration Schemes

In order to use MD, an initial structure with coordinates and interparticle relations is necessary. The subsequent positions of the particles shaping the system are then calculated as a function of time, integrating Newton's motion equation (Equation 3.11) in discrete time steps until the desired simulation time is reached.[97,98]

$$F = ma \tag{3.11}$$

The evolution of the system (Equation 3.12) is determined by its potential energy (V), which should be computed at each time step for the given set of Cartesian coordinates (r). The accuracy of the simulation depends critically on the potential $V(r)$ calculation, which at the same time relies on the parameters described by force fields.

$$\frac{d^2\vec{r}}{dt^2} = -\nabla V(\vec{r}) \tag{3.12}$$

The position of atoms in subsequent time steps should be calculated by integration of Newton's equations of motion. Several integration schemes

accomplish the criteria for solving the equation in a suitable manner for simulations, usually by a Taylor expansion, being the Verlet Algorithm the most common.[99]

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2 + O\delta t^4 \tag{3.13}$$

$$v(t) = (r(t + \delta t) - r(t - \delta t))/2\delta t \tag{3.14}$$

The Verlet algorithm is based on Taylor Expansion for the atom's position at a certain time t. The equations allow calculating the updated position and velocity, using the position and acceleration at time t but also position at the previous time step ($r(t - \delta t)$) (Equation 3.13 & Equation 3.14). Verlet algorithm allows predicting the system coordinates both backward ($-\delta t$) and forward (δt); though numerical precision is scarce as large and small numbers are added in the same equation. To overcome these difficulties, the leap-frog algorithm eliminates the inaccuracy of adding different scale numbers by computing velocities at half-time-step, though introducing a mismatch between positions and velocities (Equation 3.15).

$$v(t) = (v(t + \frac{1}{2}\delta t) + v(t - \frac{1}{2}\delta t))/2 \tag{3.15}$$

However, in both cases, it is mandatory to compute the velocity in a previous state of the system to compute the following step. The improvement of the algorithm to the self-starting Velocity – Verlet algorithm permits to compute simulations with no prior knowledge of the system's state, being the preferred integrative scheme nowadays (equations 3.16 & 3.17).

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2m}F(t)\delta t^2 \tag{3.16}$$

$$v(t + \delta t) = v(t) + \frac{1}{2m} \left[F(t) + F(t + \delta t) \right] \delta t \qquad (3.17)$$

The Velocity-Verlet algorithm accomplishes all the criteria to perform optimum simulations: i) accuracy on how the trajectory will differ from the exact trajectory at every time step; ii) energy conservation along the trajectory, fundamental for a proper kinetic energy evaluation; iii) time-reversibility, so back and forward movements are possible; and iv) symplecticity, meaning that the algorithm maintains global stability.[98]

Besides deciding the integration scheme, the integration time step has also to be chosen, which is determined by the smallest vibrational degrees of freedom involved. At longer integration time steps, the lower the computational cost is at the expense of higher inaccurate results. A smaller integration step would retrieve more exact results, though it could be reaching unsustainable computational costs. In general, bonds including hydrogen atoms vibrate the fastest, limiting the integration time step to a minimum of 1 fs. If larger time steps are desired to ease the computational requirements, bonds with H atoms can be constrained with certain algorithms, such as the SHAKE algorithm.[100]

3.4.2   Thermostats and Barostats

The algorithm discussed so far is suitable for a *Microcanonical ensemble* (NVE), where the total energy of the system is a constant of motion as the system is isolated. To approximate the real situation of biomolecular systems, it is advisable to consider *canonical ensembles* (NVT) or *isobaric-isothermal ensembles* (NPT), which is generally the real case in a laboratory. Hence, systems are coupled to a constant heat or pressure bath, respectively. The thermostats modify Newtonian dynamics to generate a statistical ensemble, avoiding numerical errors as energy is added and removed from the boundaries of the system, not keeping the temperature

constant but rather ensuring a desired average one. For example, a heat bath adapts to the proper average system by modifying the velocity of its particles; whereas pressure baths scale their volume along the simulation to ensure a globally appropriate energetic value. The election of NVT or NPT ensembles depends on the aim of each simulation, though little differences are expected between them. Nowadays there are different thermostats at disposal, such as Velocity (re)scaling, Berendsen, Andersen, Nosé-Hoover or Langevin. For barostats, there are two main categories: those assuming either a constant box shape or a constant box volume. Berendsen and Andersen Barostats are examples of the first group, while Parrinello-Rahman and Nosé-Hoover belong to the second.[98]

In the present Thesis, Berendsen thermostat (NPT) has been applied in the equilibration phases of the system, say, during the first steps of the simulations to relax possible misconnections and steric clashes that may be present in the input system, while heating until the gold standard 300 K. Then, the production phase has been simulated either by applying Monte Carlo barostat and Langevin thermostat (NPT) for MD or Andersen-like temperature coupling scheme (NVT) for GaMDs, which will be explained in the next section.

3.4.3 Gaussian accelerated Molecular Dynamics

The major challenge of classical MD lies in the fact that their maximum simulations' time range between hundreds of nanoseconds to microseconds, usually an insufficient length to explore conformational transitions that imply larger energy barriers. To achieve more exhaustive explorations, enhanced sampling and free energy calculations can be applied. Those biasing simulation methods introduce a potential or force bias at certain moments of the simulation to ease transitions across high-energy barriers, determined by reaction coordinates as atom distances,

torsional dihedrals or eigenvectors, depending on the method. Among enhanced sampling methods are accelerated Molecular Dynamics (aMD), characterized by adding a boost potential to smooth the potential energy barrier, easing transitions between low-energy states. Therefore, rare barrier-crossing events are made accessible, whereas with classical MDs the chances of overcoming them are extremely complicated.[63] There are different possibilities with aMD, such as replica exchange molecular dynamics (REMD),[64] in which the system runs in parallel replicas under different conditions (either pressure, temperature, etc) and cross their coordinates among them every certain steps; or Gaussian accelerated Molecular Dynamics (GaMD),[63] being the ones applied in the present Thesis.

GaMDs are proposed as feasible even for large proteins, with its boost potential adapted to the biomolecular potential energy by using harmonic functions. Though greatly favouring conformational sampling, boost potentials tend to introduce energetic noise in the simulations, which should be carefully reweighted back to the original free energy landscape. In GaMD, simulation parameters are used to adjust the magnitude and width of the boost potential, following a Gaussian distribution and allowing accurate reweighting of the simulations via cumulant expansion to the second order. To enhance the exploration, GaMD relies on a boost potential that is added when the potential of the system $V(r)$ is lower than a threshold energy E. The modified potential $V*(r)$ is computed as:[63]

$$V*(r) = V(r) + \Delta(V)(r)$$

$$\Delta V(r) = \begin{cases} \frac{1}{2}k(E - V(r))^2, V(r) < E \\ \\ 0, V(r) \geqslant E \end{cases}$$

In this equation, k and E are adjustable parameters based on several principles:

1. Two potential values V1 and V2 should keep their relative order (say, if V1 < V2, V1* < V2*).

2. The potential difference between two arbitrary potential values should be smaller after applying the boost potential.

3. The energy threshold should be set between the system minimum (V_{min}) and maximum (V_{max}) potential energy, adding a correction to the minimum potential called k.

4. k has to be a value between 0 and 1, satisfying that $k < 1/(V_{max} - V_{min})$. The greater k, the higher boost potential is added to the potential energy surface, further reducing the energy barriers.

However, the standard deviation (σ) of ΔV needs to be small enough to obtain a narrow distribution and ensure energetic reweighting, following that:

$$\sigma_{\Delta V} = k(E - V_{avg})\sigma_V \leqslant \sigma_0$$

where V_{avg} and σ_0 refer to potential energy, and $\sigma_{\Delta V}$ to the difference between potentials. σ_0 is a user-defined limit to allow proper reweighting.

Currently, GaMD simulations can be performed on several suits, being the AMBER one of the most applied. There is a well-established protocol in AMBER, composed of several steps:

1. Conventional MD as a preparatory stage is run, with the aim of equilibrating the system and reducing potential clashes.

2. Conventional MD as a statistical stage, to determine V_{max}, V_{min}, V_{avg} and $\sigma_{\Delta V}$.

3. GaMD prequilibration stage to apply the boost potential without updating the parameters, equilibrating the system once the boost is applied.

4. GaMD equilibration stage to apply both the potential and update boost parameters.

5. GaMD production with the application of the boost potential and without updating boost parameters.

The parameters of each step should be correctly implemented for each study, as the computation of the acceleration parameters may differ between systems. Furthermore, in AMBER suit it is possible to restrict the boost potential on different parameters of the energy, being them i) the total potential, ii) the dihedral, iii) both the total and the dihedral, iv) non-bonded energy, or iv) dihedral and non-bonded energy. Generally, applying dual-boost enhances the exploration acceleration, though in some cases energy thresholds may be compromised.

Overall, the implementation of GaMD techniques has allowed substantial improvement in the quality of explorations, opening a completely new field of possibilities. Further adaptations of GaMD algorithms have been recently developed focusing on specific search projects, such as LiGaMD and Pep-GaMD, aiming at simulating the binding of small-molecule ligands and folding of highly flexible peptides in the scale of microseconds, respectively. The design of such GaMD techniques points to the fact that the full potential of GaMD simulations is still to be discovered.

3.5   Solvation Effects

Up to this point, it has already been mentioned the necessity of embedding the system in realistic environments to perform suitable simulations, regarding mostly temperature and pressure conditions. However, most biomolecules are surrounded by a solvent, typically water, which may change substantially the properties of the molecule and interfere with natural processes, such as folding or catalytic reactions, due to electrostatics and free energy changes. Accordingly, the solvation-free energy should be carefully accounted for to obtain an accurate simulation.

Generally, solvation effects can be computed in two different manners. A first strategy includes explicit models, in which solvent molecules are treated explicitly, say, all the interactions between solute and solvent are considered, requiring high computational resources. The alternative is implicit treatments, which opt for approximating the discrete solvent as a continuum medium, characterized by its dielectric constant, which significantly speeds up the computational sampling.[101] Though arguably less accurate, implicit models are the preferred alternative for MD simulations due to their computational efficiency not having to process individual interactions and allowing a faster conformational space exploration.[102] Despite several solvents can be modelled, water solvent is the most common and the one on which the following techniques are focused.

Implicit Solvent

In implicit solvent, the water environment is replaced by an infinite continuum corresponding to the dielectric properties of water. With this strategy, the sample of the conformational space is enhanced as solvent viscosity is reduced, so there is no need for long equilibration simulations of water molecules as the dielectric response is instantaneously obtained.

Therefore, lower computational resources are required, and free energies can be straightforwardly estimated. Nonetheless, all these advantages depend critically on the adequation of the model to the system, which can be divided into those based on the Poisson-Boltzmann equation, such as solvation models based on density (SMD); or those based on the Generalized Born model (GB). The latter is commonly applied in Molecular Dynamics, as it allows to calculate electrostatics of the molecular solvation-free energy approximately.[103]

Explicit Solvent

Explicit solvent models may become prohibitive for molecular systems that undergo significant structural transitions and require longer equilibration times for water molecules, also introducing the difficulty of calculating relative free energies due to a large number of degrees of freedom. For those cases, the implicit solvent framework becomes a more suitable alternative. However, as they account for solute-solvent interactions, explicit methods are the most accurate description of the solvation effect and are required when solvent molecules are an indispensable part of a reaction or a key element of the metal coordination sphere. In fact, explicit models are further used in MM simulations than QM, as the former uses a force field description of water molecules that allows introducing a high amount of solvent molecules within a reasonable computational cost; while for QM the possibility of introducing explicit molecules is highly limited and rapidly increases the computational demand.

Explicit solvent relies on realistic water models to reproduce experimentally measured water properties and describe the solute-solvent interactions. The most common models are TIP3P and SPC/E, which simplify the molecule to eight parameters regarding partial charges of O and H, Lennard-Jonnes parameters, OH bond distance and H-O-H angle.

Though based on the same properties, each model reproduces differently each quality, thus it should be chosen according to the need of the aim system. More advanced models, such as TIP4P or TIP5P, significantly improve the reproduction according to experimental properties but they imply a significantly increased computational cost.[103]

In explicit models, the solute is embedded in a finite box filled with modelled water molecules, forming a solvation cell, whose shape is typically cubic though any polyhedral that suits the disposition of the system could be chosen. However, the interaction of solute atoms with the walls of the solvation cell may retrieve misleading results, as unless the goal is to study the behaviour near those walls, they would be introducing distorted interactions. To overcome such interferences, Periodic Boundary Conditions (PBC) are applied, which consist of eliminating the barriers among different solvent cells, building an infinite space-filling array with copies of the simulation space in all directions. Therefore, if an atom goes out of the simulation through any of the walls, it will enter again into the simulation space on the opposite side. With the same idea, atoms close to the limits would interact with a copy of the atoms of the opposite wall, simulating a globe and making the computation of short-range interactions feasible.[104] The main challenge of PBC representations is long-range interactions, which could extend along the whole system, even affecting adjacent cells. The most straightforward solution may be its truncation, though it is not an option as it would introduce artefacts. Hence, to avoid truncation, Ewald methods apply interaction resummation, meaning that each replica of the system includes contributions from the other replicas in terms of long-range interactions. As Ewald summation scales rapidly with the number of unit cell repetitions, Particle Mesh Ewald (PME) is usually recurred for simulation purposes. PME differs from Ewald as Fast Fourier Transformations are introduced, splitting short and long-range interactions in differential properties. The former is

computed in a particle-particle way; while for long-rage interactions the system is divided as a grid, defining first the potential for each grid and then applying Fourier Transformations to unravel the global interaction among grids, greatly decreasing the computational cost.

In the present Thesis, different solvation methods have been applied for QM and MD calculations. For the former, implicit solvent calculations with SMD, which does not depend on partial atomic charges and is optimized for B3LYP/6-31G* basis set, have been applied.[105] For MD and GaMD, the explicit model with TIP3P has been selected, in PBC.

3.6   QM/MM: a Hybrid Method

Up to this point, QM and MM procedures have been clearly differentiated and identified as optimal for certain specific topics. However, on some occasions, it is advisable to mix both theoretical approaches to describe processes at different levels and time scales, such as a bond breaking during an MD simulation. For those cases, a hybrid method would be desired, known as QM/MM.

With QM/MM the system is divided into two different terms: those general, collective movements are described through MM techniques; while in those regions in which the chemical processes take place, QM is applied (Figure 3.3).

The total energy of the system, then, consists of a sum of three interactions: those in the QM region, those in the MM region and the interactions between QM and MM atoms. As mentioned in the previous sections, the description of interactions at every single level demonstrates plenty of alternatives, though computing the interactions in the boundaries between the two subsets is a great challenge. There are two sets of approaches to
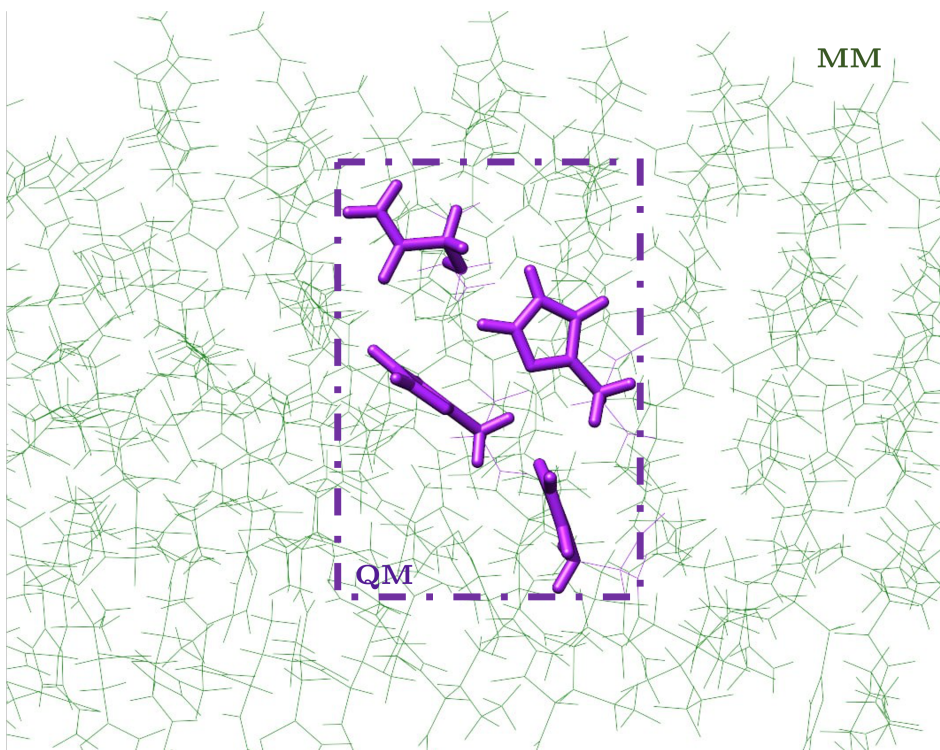
Figure 3.3: Representation of QM/MM calculation. The purple region is included in the QM, while the green one corresponds to MM calculations. The boundary region is represented in a dotted square.

tackle the interrelation between QM and MM systems: subtractive and additive QM/MM methods.[67]

In the subtractive QM/MM coupling (Equation 3.18), the energy of the whole system (MM and QM regions) is evaluated at the MM level. Then, the QM energy of the QM region is added. Finally, the MM energy of the QM system is subtracted, avoiding introducing QM interactions twice. Preventing the interaction between the regions with different descriptions leads to a straightforward approach, though a force field is neces-

sary for the QM region. One of the most common subtractive methods is the ONIOM scheme.[106] However, those systems for which solute-solvent charge transfer processes should be computed require a more realistic description, as the electronic interactions of the QM part with their environment, in the MM region, are missing.

$$V_{QM/MM} = V_{MM}(MM + QM) + V_{QM}(QM) - V_{MM}(QM) \qquad (3.18)$$

In additive QM/MM coupling (Equation 3.19, the energy is computed for the QM and MM regions, treating explicitly the interactions between the two sets. Those interactions can be described most simply with force fields, with a mechanical embedding approach, though the electrostatic potential at the QM surface should be treated with care. More realistic approaches including electrostatic effects can be obtained with an electrostatic embedding scheme, where electrostatic interactions are described with an electronic wave function, despite being less commonly applied due to its underneath complexity. The most sophisticated level is polarization embedding, which includes the mutual polarization of MM and QM atoms but requires a substantially higher computational resource.

$$V_{QM/MM} = V_{QM}(QM) + V_{MM}(MM) + V_{QM-MM}(QM + MM) \qquad (3.19)$$

As described, QM/MM methods combine the best of both strategies in order to describe molecular properties. QM/MM techniques could not be applied in the study of Aβ aggregation (chapter 4) as i) several coordination spheres are tested for each complex, which would have reached unfeasible computational cost under QM/MM and ii) the aim was either to obtain general tendencies in the protein behaviour or to

describe the chemical reaction, but not a combination of both at the same time. For that study, the observed experimental coordination is parametrised, allowing the observation of the metalloproteins' behaviour upon each coordination. On the contrary, QM/MM techniques are suitable for the study of pNPA catalytic fibres (chapter 5), in order to discern the role of the sequence's 6th residue in reactivity.

In this Chapter, a series of computational techniques have been described and pointed to their specific advantages, drawbacks, and applications. However, studying a complex system requires not just a single methodology, but a reasonable and sequential application of them. Multi-scale protocols are designed to tackle a complex problem by dividing it into concrete, individual parts, which are solved separately before joining them to obtain the overall description of the process. In the present Thesis, an integrative protocol with the application of a broad range of techniques - from QM to Docking, MD and GaMD - has been designed, aiming at an accurate description of the systems under study at different chemical levels.

4

Computational Study of Amyloid-β Aggregation

This Chapter presents the results obtained from the daunting challenge of exploring Amyloid-β conformational changes upon metal coordination. Two main structures, the monomeric and fibrillar structures, have been aimed, either in their canonical sequence or considering familial mutations.

4.1 General Overview

4.1.1 What is Alzheimer's Disease?

Neurodegenerative diseases are one of the main focuses of attention of the scientific community due to their increasing incidence in society. For the last few years, they have gathered interdisciplinary research, uniting the knowledge of medicine, biochemistry, genetics, and chemistry. Though extremely valuable knowledge about their onset, causes, and related genes has already been gained, many remaining questions should be resolved to achieve appropriate treatments.

A common cause among the most prevalent neurodegenerative diseases is protein misfolding. In misfolding, proteins change their conformation to aberrant forms, accumulating in aggregates until a size that cannot be cleared through physiological processes is reached. Such aggregation is lethal for neurons of the affected area, producing the symptomatic disease. For each disease, the misfolded protein is different as it also is the affected area: amyloid-β and Tau proteins are related to Alzheimer's Disease, α-synuclein to Parkinson's Disease or the prion protein to Creutzfeldt-Jakob disease, to name just a few. Though specific genetic mutations have been related to the aggregation process in a minority of patients, which tend to be well-characterized and related to familiar cases; most neurodegenerative disease cases still have an unknown onset.

Among all neurodegenerative diseases, Alzheimer's Disease (AD) is the most prevalent one, causing 60-70% of dementia cases,[107] meaning 10 million new cases worldwide each year.[108] AD originates in the entorhinal cortex and the hippocampus brain areas, which are memory-related. Nonetheless, many other areas get damaged as it spreads, further affect-

ing language, reasoning, social behaviour, and cognitive skills. In the last stages, the patient loses the capacity to live independently; and eventually, the disease is fatal.

Early detection is fundamental for adequately treating the disease, but AD is only diagnosed once it is symptomatic, implying that the neuronal mass has already been irreversibly affected. Besides, the drugs available for AD are either palliative or aim at slowing the progression of the disease. There is no effective cure once the aggregation process has started, and nothing can stop or reverse it. Therefore, describing the mechanism behind the aggregation onset is crucial for understanding and developing a cure. Two types of accumulations are related to AD, as hyperphosphorylated Tau protein aggregates form intraneuronal fibrillar tangles; while Aβ fragments form extracellular fibrillar peptides, called senile plaques. Even though several hypotheses have been raised in the last few years about the molecular mechanism of Aβ and Tau accumulation, it has not been determined yet.

4.1.2   Aβ peptide

The Aβ peptide is an intrinsically disordered peptide and the main component of the extracellular amyloid plaques or senile plaques. The Aβ is produced by the cleavage of the membrane protein Amyloid Precursor Protein (APP), which produces Aβ fragments of different lengths, usually between 39 to 43 residues. Such cleavage may happen through two different routes. On the one side, there is the predominant nonamyloidogenic pathway involving α and γ-secretases, which also happens in healthy individuals. With the nonamyloidogenic pathway, the fragments produced access signalling pathways and help against oxidative damage.[109] Conversely, the amyloidogenic route happens when β and γ-secretases cleavage the APP, which acts in the extracellular and intracellular domains, re-

spectively. This latter process mostly produces the Aβ form of 40 amino acids, but there is a 10% production of $A\beta_{42}$. This 10% is normally cleared through proteases, such as neprilysin, which degrades the monomers and small oligomers (Figure 4.1). However, when this subtle mechanism is disrupted and the production of $A\beta_{42}$ is imbalanced, the peptide may start its accumulation in the brain, forming the senile plaques and causing the disease.[110]

The accumulation of Aβ has been classically considered the central event of the disease, especially due to its prion-like character. This implies that the aggregates of monomeric forms attract other monomers, growing to form oligomers, which then bind among them and form protofibrils. Finally, the protofibrils can be rearranged and form the fibrillar structure, which is observable in the brain tissue through the microscope. The aggregation unit could break in any of these phases, though, forming small subunits that may travel through neurons to new areas. There, such subunits can act like seeds, altering the $A\beta_{40}/A\beta_{42}$ ratio and spreading the aggregation process all around the brain mass.[111] That is the reason why, even though the disease tends to start in the memory area of the brain -hippocampus and entorhinal cortex-, several functions are altered in the last stages of the disease, once the complete brain mass has been affected. Surprisingly though, senile plaques are not unique to AD, as 20-40% of healthy individuals also present plaques in post-mortem analysis, which suggests that further unknown elements are involved in the disease.[110]

The evolution of the peptide conformation along this process and the sequential energy barriers that are overcome until the fibrillar form is adopted are still elusive. Consequently, the detailed study of each step of the aggregation process is of high interest, as the traits of the Aβ peptide along these phases could represent potential targets to achieve an early stop in the progression of the disease.

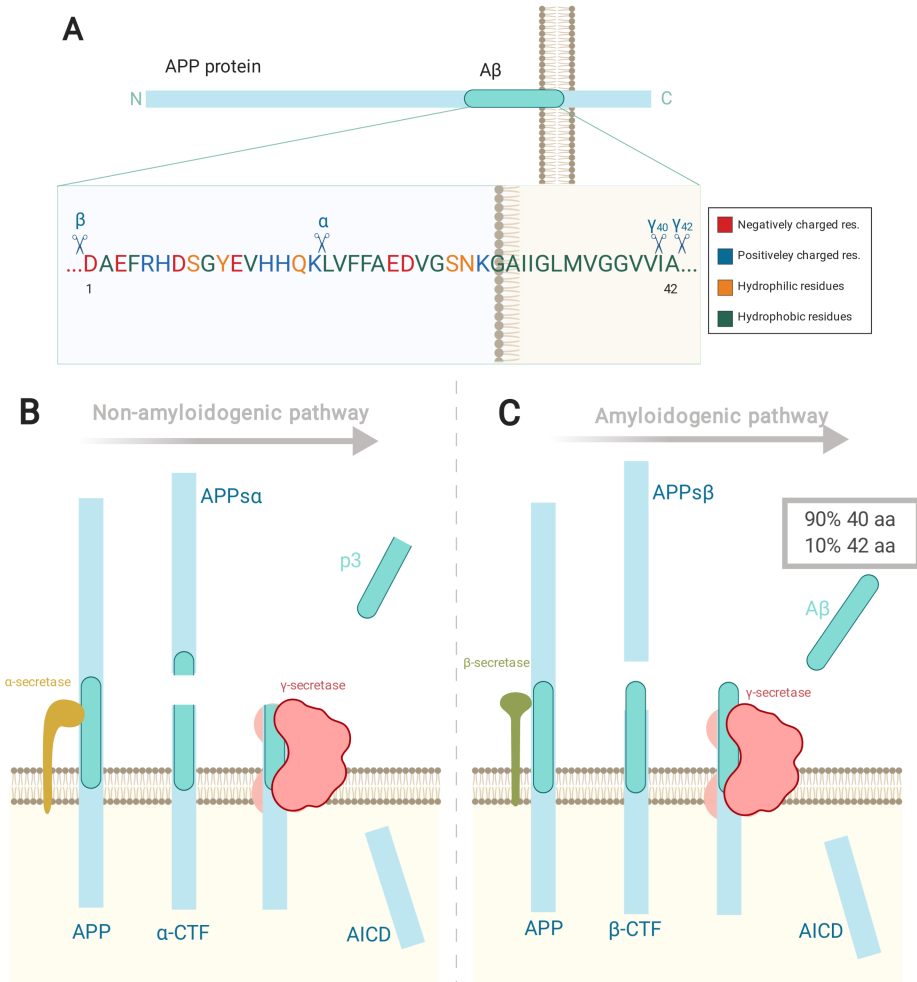**Figure 4.1:** Cleavage of the Amyloid Precursor Protein (APP). A) APP is a transmembrane peptide with extracellular N_{Ter} and intracellular C_{Ter} regions. The fragment corresponding to Aβ is represented in aquamarine. B) Non-amyloidogenic cleavage pathway, in which α and γ-secretases are involved. C) Depending on the γ-secretase, the Amyloidogenic cleavage pathway produced the Aβ fragment. Aβ fragments can be either 40 or 42 amino acids long. Created with BioRender, modified from [46].

4.1.2.1   Monomer form

The monomeric form is the smallest subunit involved in the aggregation process. The $A\beta_{40}$ form has already been crystallized[112] (PDB code 2LFM) as a central helical region with a hydrophobic core formed by residues 16 to 40. The N_{Ter} region, mainly polar, is characterized by being unstructured. The monomer form is therefore highly flexible, whose secondary structure may strongly depend on the environment. Monomers are not considered toxic *per se*, but the alteration of their secondary structure, from α-helix to β-sheet conformation, seems to be key for the aggregation process, as it is associated with higher aggregated forms.[113] Thus, $A\beta$ monomers aggregate into several types of higher structures, being oligomers the first step of this aggregation process.

4.1.2.2   Oligomeric form

The aggregation process of oligomers can happen either through i) a fibril-independent pathway, by which the monomeric species assemble to form dimeric or trimeric forms; or ii) a fibril-dependent process, where the fibrils' surface acts as inducing spaces for aggregation, in a prion-like manner.[114] Most oligomers, though, have been observed to easily dissociate back to monomeric species, instead of forming fibrils. However, such fibrils do originate from oligomers, highlighting the need for structural study of these intermedia species. Therefore, the relationship between oligomers and fibrils has yet to be established.

Besides, plenty of evidence has already highlighted that $A\beta$ oligomers are highly cytotoxic *per se*, triggering neuronal death, even more than the relatively inert fibrils.[115,116] Their cytotoxic activity is caused by interfering with several physiological processes, such as producing pore structures in the membrane, dysregulating Ca^{2+} homeostasis or binding to neurons' receptors, affecting several signalling pathways.[117,118] Besides,

soluble oligomers can easily spread through the brain and act as seeds, being strongly related to the progression of the disease.

Nonetheless, both the biological activity and structural characterization of the Aβ oligomers are still scarce, as their transient nature represents a great challenge for their structural study, with just a few studies available about it.[113]

4.1.2.3 Fibrillar form

The fibrillar form of the Aβ peptide is large and insoluble, whose assemble finally produces the amyloid plaques. It has been typically considered that amyloid plaques cause the characteristic neuronal loss of Alzheimer's, though some recent studies suggest that there is not a direct correlation between amyloid plaques and the loss of synapses and neurons in AD.[113] Such fact could point to the hypothesis that, even though the fibrils represent the last stage of the aggregation process and are observed in the brain of AD patients, they may not be the main cause of neuronal loss.[118]

Due to such structures presenting constrained flexibility, the study of the structural form of the fibrils is far more extensive than the oligomeric conformation. Several studies[119–123] have already crystallized distinct fibrillar states of the Aβ (PDB codes 2LMN, 2LMP, 2M4J, 2MXU, 2NAO, 5KK3). All these fibrillar structures show a β-cross conformation, with individual β-strands disposed perpendicularly to the axis of the fibril, forming lateral inter-sheet hydrogen-bond interactions among them, with an unstructured N_{Ter} region. There is also a fundamental salt bridge between Asp23 and Lys28,[118] which helps in the maintenance of the turn. Despite these common traits, the monomer can lead to different fibrillar morphologies (Figure 4.2), since the aggregation process is affected by subtleties such as the length of the Aβ (40 vs 42) or growth conditions.
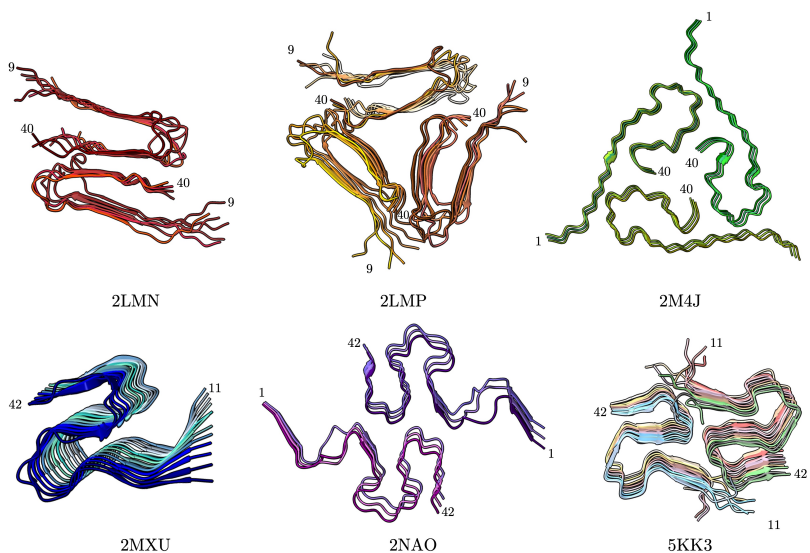
2LMN   2LMP   2M4J

2MXU   2NAO   5KK3

Figure 4.2: Crystallographic structures for the Aβ fibre. A β-cross conformation perpendicularly to the axis of the fibril, with flexible N_{Ter} (commonly missing in crystal structures) is predominant.
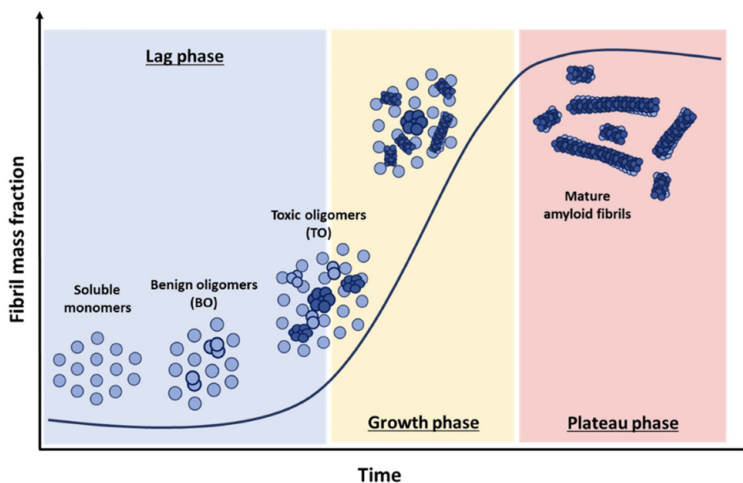

Figure 4.3: Schematic representation of amyloid formation process according to the nucleation–conversion–polymerization model. Image from [124].

The study of any of these structures (monomeric, oligomeric and fibrillar) (Figure 4.3) is still of high interest, particularly concerning the unique dynamical behaviour of each fibril and the critical elements related to its stabilization. If the contacts that lead to the fibril formation are appropriately identified, it would also be possible to prevent them from happening and hamper the aggregation process.

4.1.2.4   Aβ mutations or alternative forms

The previously described structures correspond to Wild Type (WT) Aβ peptide, meaning, the canonical sequence. However, as with many other diseases, the APP protein presents several mutations related to the familiar forms of the disease, in which this work has concentrated on due to their remarkable biological interest. Moreover, the usually considered form for all the studies is the Aβ_{42} fragment, broadly known for being the most aggregation-prone,[110] though shorter forms could also be of interest.

AD has two distinct genetic forms: familial AD (fAD), whose onset is detected usually before the patients reach their sixties; and late-onset AD, which is considered sporadic and does not demonstrate an inheritance pattern. Though fAD represents less than 1% of AD cases, its consideration is of great importance as it casts light on the critical regions/residues of the Aβ peptide for the aggregation process and its proper function.

A wide range of mutations has already been reported in such hereditary forms of AD, located not only in the APP but also in the Aβ itself. Most of the mutations are placed on the former, affecting its splicing process and later cleavage events, increasing the production of the Aβ_{42} peptide.
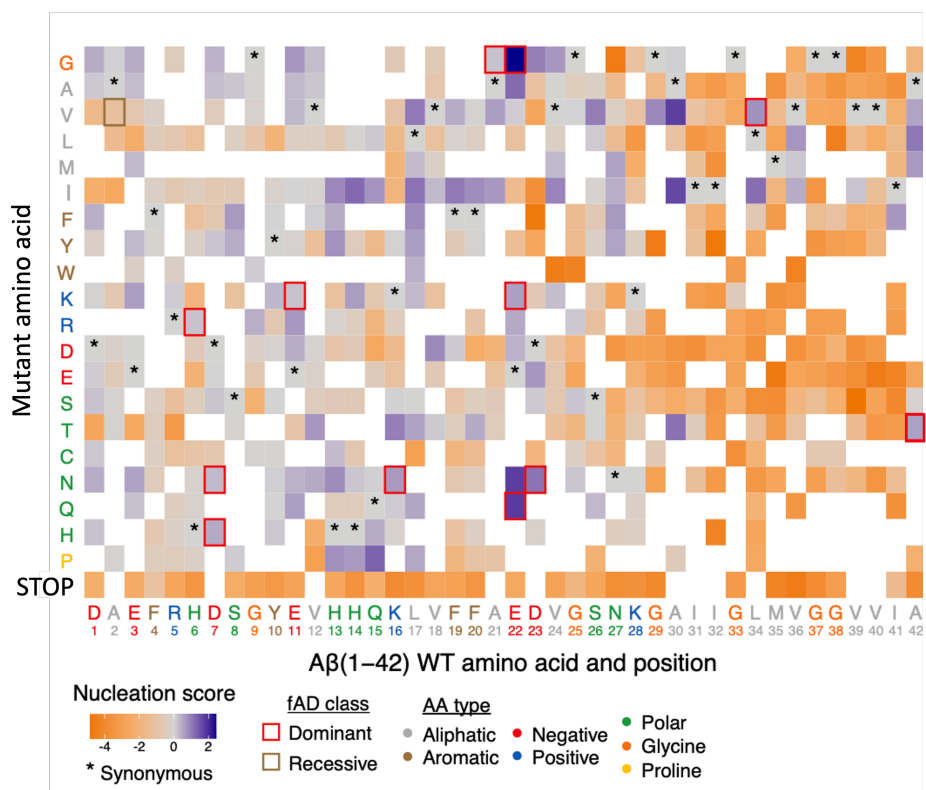
Figure 4.4: Heatmap of nucleation scores for single aa mutants. The WT aa and position are indicated on the x-axis and the mutant aa on the y-axis, both coloured by the aa class. Variants not studied are represented in white. Synonymous mutants are indicated with '*' and familial Alzheimer's Disease (fAD) mutants with a box, red if dominant and brown recessive. Orange areas correspond to reduced nucleation score, more frequent in the C_{Ter} region; while purplish zones correspond to increased mutation score, more frequent in the N_{Ter} end. Image from [125].

Regarding the $A\beta_{42}$ peptide itself, the effect of mutations demonstrates a biased pattern, with mutations that decrease nucleation being enriched in the C_{Ter} end, while the ones increasing nucleation are preferably situ-

ated in the N_{Ter} end. Besides, there are some residues in which mutations frequently increase nucleation (D1, E3, D7, E11, E22, D23, L17 and A42) (Figure 4.4).[125] In fact, five of these seven residues are also encountered among the ten well-described pathogenic familial mutations: D7N (Tottori), E11K, A21G, E22Q (Dutch), E22K (Italian), E22G (Arctic), E22Δ (Osaka), D23N (Iowa), L34V (Piedmond) and A42T.[126] These mutations are of remarkable interest for the scope of the present project, since they highlight those residues whose function is key for the aggregation process.

4.1.2.5   The aggregation mechanism

As mentioned before, though the aggregation process is central to the development of AD, the most relevant question is why the aggregation process even starts. Several risk factors have been proposed to be related to AD, such as age and environmental factors, which may significantly affect the onset of the disease. Taking all these possibilities into account, three hypotheses have been stated:[110]

- The **amyloid cascade hypothesis** suggests that the imbalance between $A\beta$ production and removal could cause AD, which starts aggregation due to the increased $A\beta_{42}/A\beta_{40}$ ratio. In this hypothesis, the amyloid peptide's aggregation is considered neurotoxic. Nonetheless, such initial accumulations have also been detected in healthy individuals' samples without leading to higher aggregation forms or causing the disease.

- The **oxidative stress hypothesis** is based on the fact that oxidative stress increases the appearance of gene defects and disrupts normal mitochondrial function, again causing $A\beta$ imbalance, ultimately inducing neuronal death and functional impairment.

- The **metal ion hypothesis** considers a metal imbalance as the AD cause. In this latter case, metals, such as Zn, Cu and Fe, are suggested to affect the Aβ peptide equilibrium, increasing its aggregation capacity. The increased concentration of such metals could come either from the alteration of physiological processes or environmental pollution.

As noted, all these hypotheses are based on an imbalance of a certain component, either the proper Aβ peptide or secondary processes. Nowadays, all of them are still on the table, with an open discussion about their implication, even suggesting a mixing of all three factors. Besides, since the study of the metal ion hypothesis is one of the most scarce and challenging, the application of cutting-edge computational tools may help in its understanding.

4.1.2.6 The Role of Metals in Alzheimer's Disease

For years, the target of AD drugs has been the elimination of the Aβ peptide. However, the lack of clinical success of these drugs forces us to raise our sights and include further possibilities, such as the metal ion hypothesis, in which the scope is not exclusively on the peptide but on the effect of its interaction with metal ions.

Disruption of metal homeostasis has been already related to multiple neurodegenerative disorders. Indeed, it has been almost 30 years since metals were first related to AD.[127] One of the main reasons to include metals in the onset of the disease is the fact that the process of APP cleavage and Aβ balance is substantially controlled by metal ions. Besides, it has been already observed that dyshomeostasis of Fe, Zn and Cu are correlated to AD.[110]

Apart from being involved in the processing process of the APP protein, some metal ions also present potential binding sites along the Aβ peptide. Metal binding to such regions could indeed modify both its structure and properties.[113] Once formed, metallic Aβ complexes have been linked to increased oxidative stress damage, as they produce reactive oxygen species, a previous step to the finding of Aβ deposits.[110] Therefore, as the Aβ peptide presents possible metal binding areas, it is of interest to study every stage of the aggregation process upon metal-peptide interaction, as metal implications could differ along the process.

Though several metal ions are important for AD, the present study has been limited to copper, a redox-active element; and aluminium, which is a non-natural metal ion. Besides, their coordination sphere with the Aβ peptide was already established in previous works.[128–130] It would also be interesting to expand the present analysis to Zn and Fe ions, whose concentrations have also been extensively reported to be altered in AD. However, Zn ion is already widely studied,[131] whose effect is expected to be similar to copper due to their similarities; while Fe is known to also produce nanoparticles,[132] whose study is out of the scope of the present Thesis.

Copper in AD

As mentioned in Chapter 1, copper (Cu) is vital for several biological processes, especially in the brain. It is involved in a high number of redox-active proteins, like tyrosinase or ceruloplasmin. Its homeostasis is also extremely critical, as both its absence and increase are dramatic: lack of copper can cause neuronal degeneration as in Menkes disease; while its increase causes oxidative damage, which finally affects several biological processes.[110,133] In fact, copper dysregulation is involved in the two proteins linked to AD. On one side, it is involved in the hyperphosphory-

lation of Tau protein, whose aggregation increases and forms neurofibrillary tangles. On the other side, Cu(II) can also directly bind to the APP protein, enhancing the production of Aβ peptide, which is another cause for the amyloid imbalance.[134] It is also suggested that Cu(II) binding to Aβ leads to amorphous aggregates, preventing fibrillar aggregation.[135] Besides, the Cu(II)-Aβ complex is capable of producing reactive oxidative species (ROS), increasing the toxicity of the peptide.[110,130] Nowadays, it is still not clear if the levels of Cu are increased or reduced in AD, as there are contradictory reports. With age, one of the most known risk factors for presenting the disease, Cu levels have been demonstrated to rise. However, other evidence indicates that lower copper levels lead to a more aggressive pathology, as Cu(II) binding to Aβ disrupts the β-sheet structure, thereby slowing down the aggregation process[136] and induces the degradation of Aβ in collaboration with anti-oxidants.[137] Accordingly, the copper effect on the Aβ peptide is still to be discovered.

Aluminium in AD

Aluminium (Al) is a non-natural metal, meaning that its presence is not essential for any living organism, opposite to Cu's function. Indeed, the uptake of Al in substantial amounts is known to have detrimental effects on the skeleton, and nervous and hemopoietic systems. The intake of Al has remarkably increased in the last years due to human activity, mainly industrialization, the ingestion of food that naturally contains aluminium, plus additives and utensils.[138] Nowadays, exposure to Al is considered a risk factor for AD, though its implication is still unknown.[43] Even though there has been certain controversy over this fact in the last years, there is growing evidence of the increase of Al concentration in AD patients' brain tissue,[139,140] which is usually co-located with Aβ.[141] Such a fact points to the possibility that Al(III) could increase Aβ toxicity in comparison to other metals, such as Cu or Zn, but further studies are necessary to properly define its relation with the disease.[110]

4.1.3   Aims of Aβ study

Alzheimer's Disease investigation is an extensive field with an important implication for human health, not only extremely attractive for the scientific community but also for society. There is growing urgency to unravel the process behind Alzheimer's Disease onset, to find effective treatments. In the present study, the conformational impact of Cu(II) and Al(III) metal cations binding on different stages of the aggregation process of the Aβ peptide has been tackled. The specific objectives have been:

1. Understand the variability of the monomeric unit of the Aβ_{42}, in its free form and when bound to Cu(II) and Al(III) metal cations.

2. Study the possibility and effect of Cu(II) and Al(III) metal binding to the highest aggregation form, the fibril.

3. Determine the different behaviour that familial forms of Aβ_{42} may present, with and without metal presence.

Here, an integrative computational approach has been applied to study the Aβ dynamical behaviour, both in monomeric and fibrillar structures; adding the challenge that metal ions can singularly bind to any of those states. In previous works from the group (see [128–130]) the binding of Al(III) and Cu(II) to the Aβ monomeric structure was already described and models for Cu(II)-Aβ_{16} and Al(III)-Aβ_{40} were built through computational methods, which represented the starting point of the present Thesis. However, the binding of metals in higher aggregation states has been largely unknown, which requires applying scanning techniques such as BioMetAll and Molecular Docking to detect possible binding areas. Apart from the difficulties that dealing with metals represent *per se*, Aβ_{42} monomer is a short, flexible peptide whose secondary and tertiary conformation can be easily affected by the presence of metal cations and,

besides, its conformational exploration with MD is remarkably biased by the starting point of the simulation. Hence, accelerated MD techniques are also required to ensure a proper exploration. The computational details regarding the present studies are specified in the following section.

4.2 Computational Details

4.2.1 Initial Models

The first step requires designing the initial models from which calculations would be performed. Here, the monomeric peptide corresponds to the Protein Data Bank 2LFM structure elongated to 42 residues using the UCSF Chimera Software, as the NMR structure comprises only up to 40 amino acids.

Several fibrillar structures of the Aβ fibril are available in the Protein Data Bank. The most complete PDB structure, with code 2NAO,[122] corresponds to the sequence 1-42 of the amyloid for each strand, though it retrieves poor punctuation in all the PDB parameters (clashes score 22, Ramachandran 12.4% and Sidechain 26.7%). On the other hand, the PDB structure 2MXU,[121] retrieves far better punctuation (clashes score 0, Ramachandran 3.8% and sidechain 6.22%) at the expense of residues 1 to 11 being missing, an observation consistent with the high flexibility expected for the N_{Ter} region of the peptide. Therefore, the last structure was chosen for the fibre, elongating each strand with the 11 missing residues using homology modelling through the UCSF Chimera Software Modeller tool[142] to obtain the N_{Ter} missing region. Chimera minimization was then performed to reduce the clashes, with Amber ff14SB force field[143] and 100 steps of Steepest Descent.

For the familial variants, all the mutations -E22G, E22Q, E22K, D23N- were performed in the corresponding structure, monomer or fibre here mentioned, using UCSF Chimera Software.

4.2.2 Metal-binding areas

For the monomers, the metal binding spheres and initial models were set according to the structures proposed by Alí-Torres et al. [130] and Mujika et al. [128] for Cu(II) and Al(III) binding respectively (see details below), considering only the seven most stable models of those reported in their studies and elongating them into 42 residues scaffolds.

In the fibres, putative metal binding areas were detected by applying the BioMetAll software[55] to the $A\beta_{11-42}$ crystallographic structure (PDB code 2MXU), configuring the search to include backbone atoms and a minimum of 3 residues as possible coordinating groups. Then, protein-ligand docking calculations were performed with the GOLD software[57] using the GOLDScore scoring[144] function, whose parameter file was modified to include atom types for metal ions and their possible coordinating amino acids.[59] Genetic algorithm (GA) parameters were set to 50 GA runs and a minimum of 100,000 operations each. The remaining parameters, including pressure, number of islands, crossovers, or mutations, were set to default. Finally, docking solutions were analyzed through GaudiView,[145] an in-house developed GUI tool built as an extension in UCSF Chimera. For the variants of the fibre, new binding sites had to be searched for the Al(III) binding, which is directly affected by mutations. Hence, a new binding site was predicted with BioMetAll only for E22Q and tested again with GOLD software, including Gln22 and Asp23 residues; while for mutations E22G, E22K and D23N, no compatible binding site was predicted.

4.2.3   Classical Molecular Dynamics simulations

Once the initial models were set up, Classical Molecular Dynamics simulations were performed to relax the original coordinates of the system and reduce possible inaccuracies of the complex. For that, Cu(II) and Al(III) metal parameters for the bonded model were obtained using the already presented Amber tool MCPB.py,[91] from quantum mechanical calculations with DFT(B3LYP) and adding Grimme's D3 correction for dispersion. For Cu-bound complexes, calculations were done with the 6-31+G(d,p) basis for C, H, N and O atoms, and the SDD basis-set complemented with f-polarization function and pseudopotential for Cu, whereas for Al-bound complexes, the 6-31+G(d,p) basis set was applied for all atoms. Solvent effects were accounted for with the solvent-polarizable dielectric continuum model (SMD) in water. Force constants and equilibrium parameters for those atoms coordinating the metal ion were obtained using Seminario's method,[90] while point charges were derived using the RESP (restrained electrostatic potential) model. The complexes were solvated within a cubic box of pre-equilibrated TIP3P water molecules and Cl^- or Na^+ ions to balance the total charge according to the needs of each system.

As already mentioned in Chapter 3, the selection of the force field is a determinant issue in the development of MD. Accordingly, three force fields were tested for the present study: Charmm36m,[61] Amber14SB[146] and ff14IDPSFF,[62] whose characteristics are summarized in Table 4.1.

The three force fields were tested in a well-characterised short-peptide, Melittin,[147] whose NMR crystal demonstrates a consistent α-helix structure. Hence, the Melittin peptide was submitted to $1\mu s$ GaMD simulation using either Charmm36m, Amber ff14SB or FF14IDPSFF force fields (Figure 4.5). With Charmm36m, the α-helix structure is just conserved in a

Table 4.1: Comparison of Charmm36m, Amber ff14SB and ff14IDPSFF force field properties.

Charmm36m	Amber ff14SB	ff14IDPSFF
Designed for ID peptides and proteins	Recommended for general proteins and peptides. Improves side chain rotamers (vs ff99SB)	Optimized backbone dihedrals designed for ID proteins
Does not stabilize helical content	Increases helical content	Does not stabilize helical content
Understimates β-hairpins	No reported information on β-sheets	Favours β-sheet structures

short region near the N_{Ter}. With ff14SB, despite the α-helix structure being partially disrupted in the intermedia region, obtaining two helices with an interhelical loop, its secondary structure was mostly maintained. On the contrary, ff14IDPSFF totally disrupts the secondary structure towards a random coil conformation.

Charmm36m was selected in a preliminary phase of the present study since it has been proven suitable both for intrinsically disordered proteins and folded systems, as long as metal moieties are not present. Though some advances have been recently reported in Amber Tutorials, Charmm36m did not accept bonded models of metals in its force field at the moment of starting the present Thesis, hence it was not suitable for our purpose. On the contrary, both Amber force fields can be straightforwardly tuned to accept bonded models for metals. Besides, even though ff14IDPSFF is recommended for intrinsically disordered proteins, it does not retrieve suitable results if the structure is likely to present a certain extent of secondary structure pattern. In agreement with these results,

ff14SB was preferred for the present Thesis work, as Aβ_{42} is expected to demonstrate secondary structure patterns rather than a totally disorganized behaviour. Besides, the combination of different force fields and explicit solvent models has also been recently discussed,[148] suggesting that the ff19SB/OPC combination is the most suitable for Intrinsically Disordered Peptides. Such a combination has also been tested on the monomeric Aβ_{42} system, with no significant differences observed between the structural results of ff19SB/OPC and ff14SB/TIP3P, presumably as Aβ_{42} cannot be considered a pure IDP, since it presents a strong α-helix propensity.
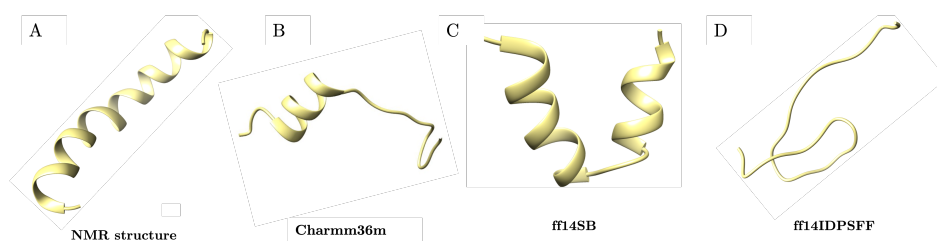
Figure 4.5: Images of A) NMR structure of Melittin (PDB code 6DST), and most populated clusters after 1μs GaMD simulation with B) Charmm36m, C) ff14SB and D) ff14IDPSFF force fields.

Accordingly, Molecular Dynamics simulations of the Aβ systems were performed with AMBER18[79] using the AMBER ff14SB[146] force field in the NPT ensemble, with an integration time step of 1 fs. Constant temperature and pressure were set by coupling the system to a Monte Carlo barostat at 1.01325 bar and a Langevin thermostat at 300 K. The SHAKE algorithm was used to constrain bonds involving hydrogen atoms.

For the monomeric systems, three independent MD simulations were run for at least 300ns, starting from the same initial structure but with initial random velocities. For the fibrillar ones, a single 10ns MD simulation was

performed. Then, the coordinates obtained after 10ns of conventional MDs were applied as a starting point for GaMDs in all systems.

4.2.4   Gaussian accelerated Molecular Dynamics

For GaMD simulations, the AMBER ff14SB force field[146] in the NVT ensemble was also applied, constraining the bonds involving H atoms with SHAKE, but with an integration time step of 2 fs. A dual boost was performed on both dihedral and total potential energy (igamd=3).

For WT monomeric systems, GaMDs up to 4μs were performed for each system; while for the monomeric variants, the simulation time was reduced to 1μs, as analyses demonstrated a sufficient conformational exploration. For the fibrillar systems, the simulation time was set to a total of 3μs, distributed in three replicas of 1μs each for every complex, to ensure proper exploration.

4.2.5   Analysis

The resulting trajectories were analysed based on several criteria, both energetic and structural. First, the proper exploration of the conformational space has been confirmed by performing a Principal Component Analysis (PCA),[149] which considers global movements to perform a 2D representation and demonstrate the complete conformational exploration; computing also the RMSD against the first frame, the RMSD against all the other frames and the cluster countering, which asses that the simulation has reached a stable plateau. Then, the energy state of the systems along the simulation was computed using the Generalized Born Implicit model,[150] stripping water molecules, and performing a short minimization (maxcyc = 500 steps) for each point before the energy extraction. The Radius of Gyration (RoG), which is defined as the root mean square distance of each

atom to the centroid of the protein and informs about the compactness of the system, was also computed over the same frames and overlapped to the same graphic, to evaluate the correspondence between compactness of the system and its energy.

Monomers

The folding pattern of the peptides was evaluated through the Frequency Contact Maps obtained with the Contact Map Explorer package for MD-Traj, which analyses the internal contacts of the peptide and their maintenance along the trajectory and provides information on the most prevalent contacts and the flexibility of the peptide. Finally, analysis of secondary structure and contacts in the second part of the peptide (Lys16-Ala42 residues) has been performed in all the trajectories to detect patterns. Only residues from Lys16 to Ala42 have been considered both for the α-helix content computation and the Hydrogen Bond (HB) contacts, since both metal ions here considered coordinate to the 1-15 N_{Ter} region, preventing it from adopting any secondary structure or participating in any relevant contacts with the rest of the peptide. In addition, the mean Radius of Gyration (RoG) has been computed as a global definition for compactness.

For the WT studies, with Cu and Al binding, analyses were performed on the GaMD trajectories. For the familial variants, with and without metal binding, a second MD of 100ns was performed on the lowest energy well from the GaMD, since no crystal structures of the familial variants are available and the development of the techniques demonstrated more reliable results with such protocol.

Fibres

Fibrillar PCA movements along trajectories were analysed with the VMD tool NMWiz and counterchecked by Normal Mode Analysis with WEB-nma software,[151,152] to study their global low energy internal motions.

For all the fibrillar systems, a second MD of 100ns was performed on the lowest energy well extracted from the implicit solvent calculation of the GaMD. All distance analyses for fibre measurements shown in the following Results section are performed over such 100ns MD trajectories.

4.3 Monomers: Impact of Cu(II) and Al(III) binding on the conformational landscape of Aβ_{42}

4.3.1 Introduction

As mentioned, the first step of the aggregation process is the misfolding of monomeric Aβ_{42} forms. As in every computational study, an initial 3D structure of the protein under study is required, and in this case, the crystallographic structure of the free form is already available (PDB code 2LFM). Regarding its capacity for binding metal ions, preliminary knowledge obtained from previous studies of the group allowed us to build plausible models for Cu(II) and Al(III) bound to Aβ_{42} (Scheme 4.1).

For Cu(II)-Aβ, initial models were built based on electron paramagnetic resonance (CW-EPR) spectroscopy experimental data, which revealed square planar coordination and two main species at physiological pH: one referred to as Component I at low pH ($<$7), and another one, named Component II, at higher pH ($>$7) (Scheme 4.1).[153] These experimentally proposed coordination spheres were used to build 3D plausible models for Cu(II)-Aβ_{16} by combining quantum chemical methods with homology modelling techniques.[129,130,154]

In the case of Al(III), quantum chemical calculations and molecular dynamics simulations, using a template-free strategy based on the preor-

ganization hypothesis, showed that the preferred coordination mode involves up to three carboxylate groups, generally located among the first 11 residues of the Aβ sequence.[128]
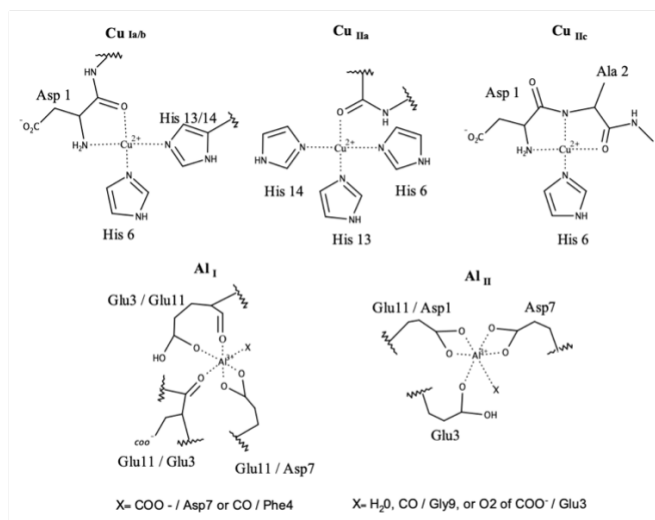
Scheme 4.1: Coordination species for Cu(II) and Al(III) with Aβ. Cu presents two different species, which can be grouped into three categories: Cu_I, at pH <7, Cu_{IIa} at pH >7 with single backbone coordination and Cu_{IIc}, also at pH >7 but with double backbone coordination. Al can be split into two categories: Al_I, which includes at least two backbone atoms; and Al_{II}, with one or no backbone coordination.

In the last years, the great improvement of molecular dynamics simulations together with a variety of force fields approximations that better describe protein behaviour, already allowed the study of $Aβ_{42}$ binding to Cu(II),[155–159], Al(III),[160] as well as Zn[161] and Na[162] systems. Although those works provided key insights on the effect of monomer metal binding, none of them afforded an intensive comparison between all the possible low-energy coordination spheres previously reported for Cu(II) and Al(III), which could shed light on the impact of the metal coordination mode. Furthermore, these studies tend to suffer from the intrinsic limi-

tations of standard Molecular Dynamics and can easily get trapped into local minima, influenced by the initial structure.[63]

Here, an extensive analysis of the $A\beta_{42}$ structure bound to Cu(II) and Al(III) metals is presented, considering all the most likely metal coordination spheres and implementing enhanced conformational sampling MD approaches. For Cu(II), seven coordination modes were selected, separated into species Cu_I and Cu_{II} according to the expected pH. For Al(III), five coordination modes were chosen, separated in Al_I, if more than one backbone atom is present in the coordination sphere; or Al_{II} species, if just one or no backbone atoms are present in the coordination sphere (see Table 4.2 for details of the coordination spheres).

Table 4.2: Metal coordination spheres of considered models.

Metal	Coordination		Reference
Al_{I_1}	$(COO^-,CO)_{Glu3}$; $(COO^-)_{Asp7}$;κ^3-$(COO^-,CO)_{Glu11}$		1 in ref.[128]
Al_{I_2}	$(CO)_{Glu3}$; κ^2-$(COO^-,CO)_{Glu11}$	$(CO)_{Phe4}$;κ^2-$(COO^-)_{Asp7}$;	2 in ref.[128]
Al_{II_1}	$(COO^-)_{Glu3}$; κ^2-$(COO^-)_{Glu11}$	κ^2-$(COO^-)_{Asp7}$;$(CO)_{Gly9}$;	3 in ref.[128]
Al_{II_2}	κ^2-$(COO^-)_{Glu3}$; κ^2-$(COO^-)_{Asp7}$;κ^2-$(COO^-)_{Glu11}$		5 in ref.[128]
Al_{II_3}	κ^2-$(COO^-)_{Asp1}$;$(COO^-)_{Glu3}$;κ^2-$(COO^-)_{Asp7}$;OH_2		6 in ref.[128]
Cu_{I_1}	$(NH_2,CO)_{Asp1}$; $(\delta$-$N)_{His6}$;$(\delta$-$N)_{His13}$		$I_{a\_\delta\delta}$ in ref.[130]
Cu_{I_2}	$(NH_2,CO)_{Asp1}$; $(\varepsilon$-$N)_{His6}$;$(\varepsilon$-$N)_{His13}$		$I_{a\_\varepsilon\varepsilon}$ in ref.[130]
Cu_{I_3}	$(NH_2,CO)_{Asp1}$; $(\delta$-$N)_{His6}$;$(\delta$-$N)_{His14}$		$I_{b\_\delta\delta}$ in ref.[130]
Cu_{I_4}	$(NH_2,CO)_{Asp1}$; $(\varepsilon$-$N)_{His6}$;$(\delta$-$N)_{His14}$		$I_{b\_\varepsilon\delta}$ in ref.[130]
Cu_{IIa_1}	$(CO)_{Ala2}$; $(\varepsilon$-$N)_{His6}$;$(\delta$-$N)_{His13}$;$(\varepsilon$-$N)_{His14}$		$II_{a\_\varepsilon\delta\epsilon}$ in ref.[130]
Cu_{IIa_2}	$(CO)_{Ala2}$; $(\varepsilon$-$N)_{His6}$;$(\varepsilon$-$N)_{His13}$;$(\varepsilon$-$N)_{His14}$		$II_{a\_\varepsilon\varepsilon\varepsilon}$ in ref.[130]
Cu_{IIc_1}	$(NH)_{Asp1}$; $(NH,CO)_{Ala2}$;$(\epsilon$-$N)_{His6}$		II_{c_ϵ} in ref.[130]

4.3.2   Results

In Table 4.3, the charges, α-helix content and residues in such secondary structure, the residues involved in the turn and HB formation, the percentage of U-shape and the radius of gyration for the thirteen systems under study (7 copper-bound, 5 aluminium-bound and the free peptide) are reported. Figure 4.6, Figure 4.8 and Figure 4.9 show the resulting energy profiles obtained from GaMD simulations, along with RoG, the timeline analysis and Frequency Contact Maps for the WT and for a few Cu(II) and Al(III) representative systems, one from each category according to the coordination sphere. Results for the remaining ones are given in Appendix A.1.

As general trends for all complexes, energy fluctuations approximately correlate with RoG, showing that the more compact the system is, the more stable it becomes. In all cases, a series of low-energy wells and transition structures are identified, which shows a substantial dynamism of these systems, as expected for peptides in GaMD simulations. The structural variability observed is much higher than the one obtained with preliminary conventional MD simulations, even though simulation lengths and replicas get to almost 1μs for most of the complexes. Therefore, it was demonstrated that accelerated MD approaches are essential for the proper evaluation of these kinds of flexible systems; otherwise, the complexes are likely to get trapped in the closest position to the initial state and not reach further stable structures. In the following sections, particular differences between the FP and the metal-bound systems will be analysed in detail.

Free Peptide

To properly understand the effect of metals in the lower aggregation state, it should be first observed how the free peptide behaves by itself. It

Table 4.3: Secondary and tertiary structure characteristics for β-amyloid complexes.

System	Q^a	α-hel.^{b}	α-helix res.	Turn res.	HB (16-42)^c	U-shape (%)^e	R$_g$ (Å)^f
FP	-3	44.5%	10-15; 20-27; 35-40	–	None	6.7%	34.9 ± 23.2
Cu$_{I\_1}$	-2	60.8%	15-22; 30-40	22-27	**D23-S26** (af 43.0%)	8.5%	20.2 ± 5.7
Cu$_{I\_2}$	-2	62.6%	15-26; 31-37	27-31	**E22-L28** (af 20.4%)	74.8%	18.3 ± 5.9
Cu$_{I\_3}$	-2	46.2%	16-20; 26-38	24-26	< af 10%	78.1%	18.4 ± 5.1
Cu$_{I\_4}$	-2	51.1%	15-25; 30-37	25-30	**D23-S26** (af 10.1%)	58.0%	19.3 ± 4.6
Cu$_{IIa\_1}$	-1	49.9%	15-27; 33-39	27-33	**D23-S26** (af 10.9%)	29.5%	20.3 ± 6.3
Cu$_{IIa\_2}$	-1	49.5%	17-25; 31-37	25-30	**E22-L28** (af 14.5%)	61.9%	19.9 ± 5.9
Cu$_{IIc\_1}$	-3	55.3%	11-20; 30-38	22-30	**D23-S26** (af 32.9%)	69.0%	19.0 ± 5.5
Al$_{I\_1}$	0	34.5%	17-25; 30-40	25-30	**D23-S26** (af 26.4%)	34.8%	19.4 ± 4.7
Al$_{I\_2}$	0	67.7%	17-25; 31-38	25-31	**D23-S26** (af 22.8%)	19.7%	20.6 ± 7.0
Al$_{II\_1}$	0	40.3%	18-27; 31-36	27-30	**D23-Q15** (af 16.5%)	94.6%	18.9 ± 3.8
Al$_{II\_2}$	0	43.6%	13-26; 28-37	26-28	**D23-S26** (af 55.1%)	3.5%	20.0 ± 5.4
Al$_{II\_3}$	0	37.9%	15-25; 31-38	25-30	**D23-S26** (af 26.0%)	27.4%	19.4 ± 5.5

^{a} Electrostatic charge of the complex. ^{b} α-helix percentage along the trajectory. ^{c} HB contacts with an average frequency of at least 10%, which are suggested to participate in the U-shape maintenance. ^{e} Percentage of U-shape structure presence along the trajectory, counted by adding up the percentage of clusters with such structure. ^{f} Average Radius of Gyration along the GaMD trajectory and its standard deviation.

was expected to observe a very flexible behaviour and conformational changes, and indeed the simulation shows that it can fold in up to three short α-helix regions: Tyr10-Gln15, Asp23-Ala30 and Leu34-Gly37 (Figure 4.6). Those regions are visually observable in the Frequency Contact Map (Figure 4.6C), with values around 1.0, indicating that they are present almost all along the simulation. The Frequency Contact Map also shows a large number of transient contacts (contacts below 1.0), particularly for regions Phe4-Asp7 and Ala30-Met35, the former being the one where metals would bind. However, the peptide exhibits large conformational shifts in which the helices constantly fold and unfold along the simulation or reorient between them, arranging several different conformations, some of them leading to highly exposed situations to the solvent (Table 4.3, mean RoG of 34.9 Å) and with very varied conformations (RoG standard deviation $\pm$ 23.2 Å)(Figure 4.6A). The timeline analysis also supports previous observations (Figure 4.6B), as the same three regions with a recurrent α-helix structure are represented, while a very short, both in time extension and residues implication, β-sheet structure is formed at around 400ns of the simulation.
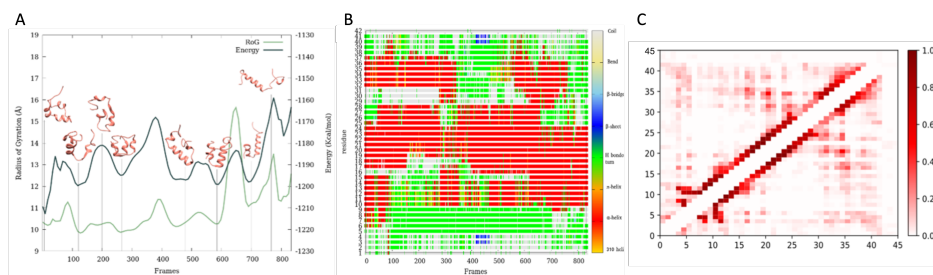
Figure 4.6: Free Peptide analyses. Implicit Solvent Energy and Radius of Gyration (A), Timeline Analysis (B) and Frequency Contact Map (C) are represented. A very flexible structure, with three main α-helix regions, is observed.

Altogether, it could be stated that the free peptide is very flexible, adopting several poses and reorganizations along the trajectory with a clear propensity towards forming recurrent helical regions, in agreement with the crystal structure of Aβ (PDB code 2LFM).[112] Notwithstanding, β-sheet regions also appear, which are usually linked to higher aggregation propensity.

Metal Coordination

Upon metal coordination, general trends are observed in comparison to the free peptide simulation. The first one is a dramatic increase in the compactness of the metallopeptides, as the RoG decreases to a value around 20 Å, against the 34.9 Å retrieved by the free peptide system. As expected, the Asp1-Lys16 region completely loses the α-helix structure previously observed, with a dramatic decrease in its flexibility, due to the binding of the metal in that region. The metal binding in the N$_{Ter}$ region also induces a refolding that propagates along the remaining residues (Leu17-Ala42), forcing the structuring of the C$_{Ter}$ region. Hence, the general profile of metal-Aβ includes two -instead of three- α-helices regions, a first one comprising from Gln15 to Gly25 and another one ranging from Ala30 to Val40. Those two helices tend to orientate in an antiparallel way, with an inter-helical loop between residues Gly25 and Ala30, which we named U-shape. Such a U-shape is stabilized by H-bond contacts, usually observed either between Glu22 or Asp23 and Ser26 or Lys28.

The formation of those H-bond contacts is especially remarkable, as they have been previously reported in monomeric Aβ forms[156,159,161] and remind the ones typically found in reported structures of Aβ fibrils (Figure 4.7).[118,121,163] The synergy between H-bond contacts in the loop together with inter-helical hydrophobic contacts leads to several minima energy wells with a U-shape structure, a conformation that is in agreement with previous works.[160]
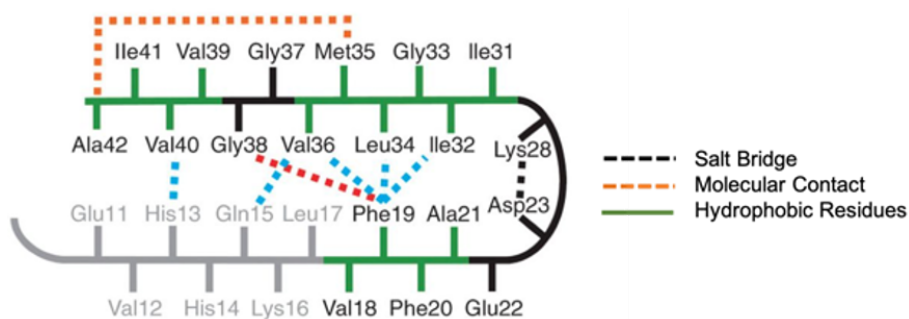
Figure 4.7: Intramonomeric interactions observed in fibrillar structures, reported by Ahmed et al.[118].

Nonetheless, remarkable differences depending on the metal ion and its coordination mode are also observed, which highlights the necessity of not only considering just a single coordination mode but all the possible variants.

Copper-bound systems

The binding of Cu(II) to Aβ (Cu(II)-Aβ$_{42}$) induces a systematic increase of the α-helix content regardless of the coordination mode, from 46 to 62% as compared to the 44% observed in the free peptide (Table 4.3). Besides, the lowest energy wells in all the cases always depict the aforementioned U-shape structure (see Figure 4.8).

Indeed, except Cu$_{I\_1}$, all complexes present a substantially higher amount of such structure concerning the free peptide, suggesting that the reorganization of the secondary structure induced by Cu(II) binding tends to constantly retrieve the U-shape conformation. Detailed analysis of Cu$_{I\_1}$, the exception regarding the U-shape, points out the formation of strong interactions between the side chains of Glu22 and Asn27 with their environment. These transitory interactions ultimately hamper the formation of the U-shape as both Glu22 and Asn27 are required for forming the

inter-helical loop and orientating the α-helix regions in an antiparallel way. This also suggests that coordination of histidine amino acids by δ or ε nitrogen atoms may have a significant impact on the structural arrangement of the peptide, as further sustained by the different U-shape percentage between Cu_{IIa_1} to Cu_{IIa_2}, which only differ in the N-His involved in coordination (Table 4.2).
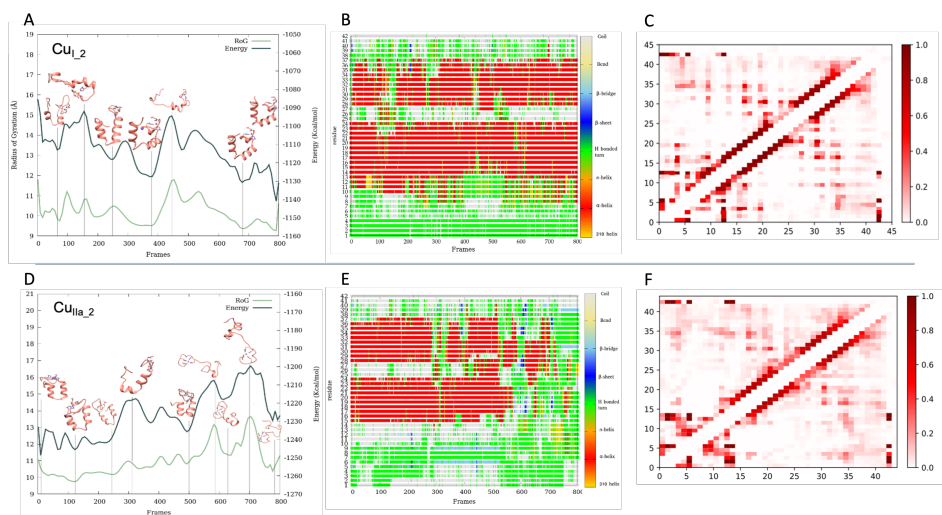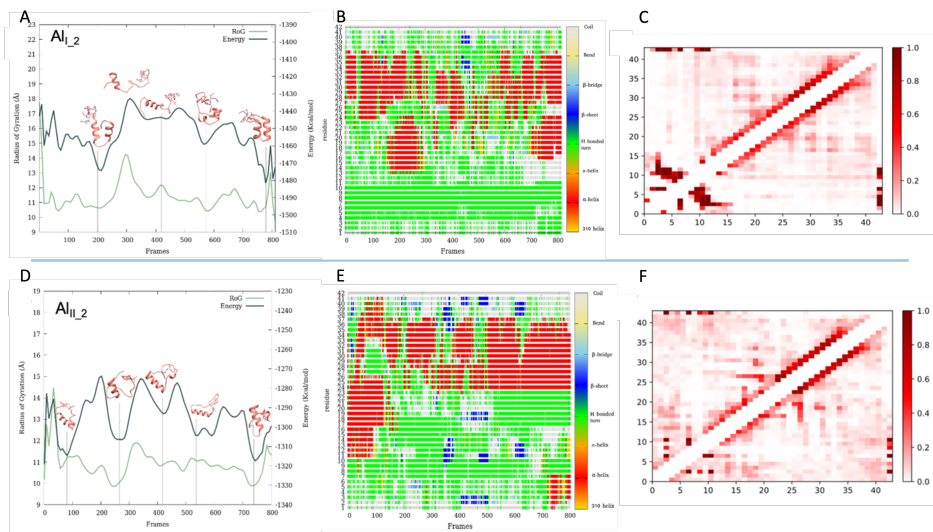

Figure 4.8: Two cases of Cu(II) coordination. Implicit Solvent Energy and Radius of Gyration (A,D), Timeline Analysis (B,E) and Frequency Contact Map (C,F) for coordinations Cu_{I_2} and Cu_{IIa_2} respectively, are represented. Two main α-helix regions are observed, which are quite stable over time, and U-shape structures are found in the lowest energy wells.

These analyses also point at a differential behaviour of the $A\beta_{42}$ once Cu(II) is bound, usually leading to a U-shape conformation. However, it remains unclear if such changes have any impact at all on the aggregation process. On the one hand, the formation of the U-turn reorganizes the monomer in a structure that can be considered fibril-like due to the creation of similar contacts to fibrillar structures; but on the other hand,

Cu(II) binding also increases the helical content, which is thought to prevent aggregation. Thus, no sound conclusions can be drawn just from these observations, so further work is required.

Aluminium-bound systems

Among all Al(III) binding complexes, only Al_{I_2} shows a significant increase of the α-helix content with respect to the free peptide (+23%, see Table 4.3); while the other Al(III) complexes even show a slight decrease in it (up to 10%, Table 4.3). This would agree with the experimental results of Exley et al. [164], who demonstrated that the partial helical conformation of $A\beta_{40}$ can be disrupted in the presence of physiologically relevant concentrations of aluminium. This also complements the results reported by Turner et al. [160], who concluded that Al(III)-$A\beta$ shows an increment towards helical propensity, as indeed there is an increase in the α-helix content in one of the coordination modes, while the other ones retrieve opposite results. The difference between both studies could be due to two main reasons. First, all models here considered are extracted from our previous multi-scale study on Al(III)-$A\beta$, while [160] focus on a coordination mode resulting from a short non-bonded simulation, obtaining a completely different coordination sphere to either Al_I or Al_{II} in our study. Second, in the present study, the conformational landscape is explored with accelerated MD techniques, while [160] used conventional MD, whose disadvantage in the present case has already been highlighted.

Another remarkable trait of Al(III) bound complexes is the emergence of β-strands motifs, characteristic of $A\beta$ fibrils (Figure 4.9B,D). Such motif is relevant for Al_{II_2} and Al_{II_3} complexes, for which a significant enhancement of β-strand formation is observed compared to the FP, both further time-lasting and with more implicated residues. In these cases, β-strands are transiently formed both in the C_{Ter} (Gly38-Ala42) and cen-

tral regions (Tyr10-Lys16), corresponding to the beginning and end of α-helix regions.

Figure 4.9: Two cases of Al(III) coordination. Implicit Solvent Energy and Radius of Gyration (A,D), Timeline Analysis (B,E) and Frequency Contact Map (C,F) for coordinations Al$_{I\_2}$ and Al$_{II\_2}$ respectively are represented. Two main α-helix regions are observed, with less compact structures. They also tend to revisit β-sheet structures along their trajectories.

Finally, it is also worth mentioning that the tertiary structure of these complexes greatly depends on the coordination mode. The percentage of U-shape structure observed in all the Al(III) simulations is highly variable, from 3.5 to 94.6%. However, for Al$_{I\_1}$, Al$_{I\_2}$ and Al$_{II\_1}$, which have at least one backbone coordination, the U-shape is recursively adopted in the lower energy conformational states; while for Al$_{II\_2}$ and Al$_{II\_3}$, with exclusively side-chain coordination, such conformation is transitory or not reached at all (Figure 4.9D). Hence, those Al(III) coordinations that include backbone residues seem to induce a higher constraint both on the

103

secondary and tertiary structure, leading to the formation of the U-shape structure, especially in the Al_{II_1} complex. The opposite result is obtained with Al_{II_2} coordination, which differs from Al_{II_1} in just one atom in the coordination sphere. Surprisingly, Al_{II_2} shows the smallest propensity of adopting the U-shape geometry, even less than the free peptide (Figure 4.9), as a folding towards a unique and longer α-helix is preferred.

4.3.3 Conclusions

This first study on the monomeric structure shows that coordination of Al(III) and Cu(II) to Aβ reduces the flexibility and increases the peptide's compactness, regardless of the coordination mode and the metal ion considered. However, significant differences are observed between both metal ions.

As a general trend, Cu(II) binding leads to an increase of α-helix content along with the formation of two helices that tend to organize in a U-shape. In recent experimental studies, it is supported that Cu(II) binding to monomers produces aggregation inert complexes, favouring the formation of amorphous aggregates rather than fibrillar structures.[135,165]

On the contrary, most Al(III) complexes induce a decrease in the helical content, in agreement with early experimental studies, with a smaller propensity to reach a U-shape structure, leaning towards more extended structures. This also favours the appearance of transitory, longer β-sheet regions, suggesting more aggregation-prone structures in the presence of Al(III). Another trait which has not been mentioned is the fact that Al(III) complexes exhibit a zero metal charge, thereby further enhancing their aggregation propensity as compared to those of Cu(II) or the free peptide.

Even though these results on the monomeric structure already give a hint about the metal impact, studying these metals in the following steps of the aggregation process is of high importance to fully comprehend their implication on $A\beta_{42}$ aggregation. While their study on oligomeric structures was attempted, oligomers' high structural variability hampered obtaining any sound conclusions. For that reason, in the following section, the focus is placed on fibrillar conformations.

4.4 Fibres: Computational Assessment of Cu(II) and Al(III) impact - Binding Sites and Structural Stability

4.4.1 Introduction

The last step in the aggregation process corresponds to the formation of fibrillar structures, which requires the switch of the secondary structure from α-helix to parallel β-sheet strands. In the last decades, numerous studies have focused on decoding the folding and aggregation process of $A\beta$ systems.[166–170] Despite these efforts, the molecular process behind the aggregation mechanism has not been reached yet, especially challenging when it comes to larger aggregated forms such as fibrillar $A\beta$ complexes.[171]

Contrary to monomers, aggregate forms are less flexible and several fibrillar structures have been recently reported by X-ray and ss-NMR techniques,[121,122,172,173] providing a solid ground for computational studies to unravel the effect of metal ions binding to amyloid fibrils.[157,158,174] Experiments lack the techniques to precisely identify the binding site of metal ions, with scarce experimental details on plausible metal binding

sites in fibrils.[175] Therefore, computational approaches represent a suitable tool to discern the impact of metal binding on Aβ fibril behaviour.

To shed light on those questions, a multi-scale protocol has been designed and applied to Aβ_{42} fibril, focusing on the conformational impact of Cu(II) and Al(III) binding. The work stands on BioMetAll application together with protein-ligand dockings compatible with metallic ligands, and both classical and accelerated Molecular Dynamics. Since the number of strands necessary to reproduce a fibrillar behaviour is unknown, the first part of this study focuses on establishing the minimum number of strands that an Aβ_{42} system should contain to reproduce it. Results for possible metal binding sites in the Aβ_{42} fibril and their impact on fibrillar conformation upon binding are reported in this section.

4.4.2   Results

Setting up the system

The complete fibrillar structure was obtained from the PDB entrance 2MXU and completed with Homology Modelling in Chimera, obtaining a 12-strand system. However, one may wonder how many strands were necessary to keep the fibrillar structure in a stable conformation. Hence, systems composed of 2, 4, 6, 8 and 10 strands were built and submitted to simulation. In such simulations with an increasing number of strands, it is observed that at least 6 strands are required to obtain a fibril-like conformation (RMSD of 10 in the free peptide), though at least 10 strands (RMSD of 6) should be considered to reach a stable core and a similar structure to the crystallographic one. As a consequence, from now onwards, the study is based on the fibrillar structure composed of 10 strands to optimize computational resources.

Metal-free fibril

Once the minimum number of strands for a fibrillar structure is determined, GaMD simulations on the metal-free 10-strands $A\beta_{42}$ system can be performed.

The GaMD simulation spread over 3 replicas of $1\mu s$ each, which exhibit similar behaviours in terms of global breathing motions and exhaustive exploration of the conformational space, as can be seen in the circular pattern described in PCA analysis (Figure 4.10B, Figure A.8). The corresponding potential energy profile clearly demonstrates a series of low energy wells explored along the trajectory of the simulation, with a slowly but continuously raising RoG – meaning a geometric expansion of the system (Figure 4.10C). This is in line with the results previously reported by Millet et al.[176], which suggests that fibres do not show a unique lower energy state but different conformers with barriers between them, which are expected to be more or less favoured upon metal binding.

The trajectories of the free-metal system present a series of interesting structural features. First, only the 6 central strands maintain the fibrillar conformation while the two at each extreme tend to unfold (4 in total) (Figure 4.10D). This is related to the fact that these extremes are more exposed to the solvent and naturally start to adopt a more hydrophobic exposed conformation with increased helical content, reminiscent of monomeric peptides. The second remarkable trait is that the terminal regions of each strand, from residue 1 to 10, are disorganized as also observed in the PDB structure, which does not contain atomic information of such region. Nonetheless, the overall geometry of the amyloid is barely affected by N_{Ter} region, as its random conformations do not destabilize the core structure at all. Based on this, the N_{Ter} region as well as the 4 strands at the extremes are discarded for the rest of the analyses, focusing just on the core formed by the 6 central strands from residues 11 to 42,

aiming at reducing the inherent dispersion that 1-10 residues introduce in the analysis. Indeed, suppressing those residues directly before running the simulation was a bare option, but it was preferable to generate a system as similar as possible to the real situation and opt for avoiding their inclusion afterwards in the analyses.
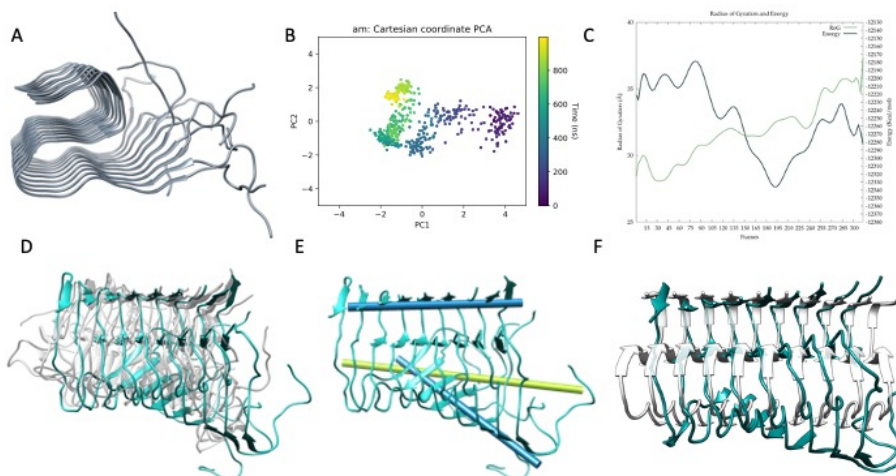
Figure 4.10: Results of the metal-free system. A) $A\beta_{42}$ model built from $A\beta_{11-42}$ (2MXU)[121] after adding the missing 1–11 residues. B) Principle Component Analysis (PCA) of GaMD simulation. C) Energy profile and Radius of Gyration (RoG) along the GaMD simulation. D) Overlap of representative structures of most populated clusters obtained from GaMD simulations. E) In blue, metal-free fibril axis representation of 39–41 (top) and 11–15 β-sheets (bottom); in green, 11–15 β-sheets axis of NMR-structure. F) Overlap between the solid-state NMR (ss-NMR) structure and the lowest energy structure of the GaMD simulation.

Regarding the conformational behaviour of the free peptide, the major motion observed is a fan-like movement of the $A\beta_{11-20}$ β-sheet with respect to the rest of the core, which appears to be the natural breathing

movement of the system as counterchecked by Normal Mode Analysis performed with WEBnma software.[177] This fan-like motion could be described as a hinge, numerically characterized by the angle formed between two principal axes: one conformed by residues 11 to 15 from all the strands; the second, by 39 to 41 (Figure 4.10E). The angle obtained between these two axes for a representative structure of the most populated cluster reaches 48.5º against the 7.8º of the ss-NMR structure, describing the fan-like movement as an opening between those two regions. For the lowest energy structure of the whole simulation, the same analysis shows a hinge of 46.7º, which still represents the main distortion with respect to the ss-NMR structure. In summary, these simulations and analyses show that the metal-free fibre mostly retains the S-shape, experimentally determined geometry, though demonstrating a characteristical fan-like movement (Figure 4.10F).

Predicting metal binding sites in amyloid aggregated structures

The previous part of this study further supports the intrinsically disordered protein (IDP) behaviour of the fibre N_{Ter} region. Hence, the binding of the metal ions to this region would be similar to what occurs in monomeric species, with single strands forming the binding spheres. However, the core of the aggregate is quite rigid, whose pre-organization may generate unidentified sites for metal binding composed of more than a single strand.

To detect those new metal binding sites in the aggregate, BioMetAll calculations were performed accounting for site-specific binding areas for copper and aluminium. Then, the search was first performed looking for a minimum of 3 coordinating amino acids, considering any amino acid type known to participate in the coordination sphere of transition metals. In this screening, BioMetAll finds geometries compatible with the simultaneous coordination of two His residues in the N_{Ter} region of the

Aβ fibril as identified in the monomers.[129,130] However, this only occurs in a few strands due to the intrinsic flexibility of that region, and a saturated model was desired for this study. Since other metal-binding sites are more frequently found in the core region of the fibril, the N_{Ter} monomer-like binding sites were discarded and the focus was placed on the core region, favouring Asp and Glu residue binding for Al(III)[128] and His and Gln for Cu(II). For the latter, it is important to point out that Gln residues have been included in the search since, although rare, its coordination to Cu(II) is observed in a series of X-ray structures from the PDB. Besides, it has also been proposed as a suitable candidate for Cu(II) binding in Aβ fibres.[175]

Hence, two differentiated areas were found for metal binding in the core of the fibril: one consistent with copper and the other with aluminium (Figure 4.11). The former, named Cu_F, stands at the linker between the N_{Ter} loop and the β-strand region (from residue 12-15) and is composed of histidines His13 of two adjacent strands and two glutamines Gln15 of the same strands. The latter, named Al_F, is located at the beginning of the coil with residues Glu22 and Asp23 of adjacent strands. This site indeed agrees with the work reported by [178], which experimentally shows that this area is likely to be bound to a metal ion. With these two sites identified, realistic three-dimensional models of the metal-aggregate systems were pursued by carrying out protein-ligand dockings. Those were performed with GOLD, assuming a square planar environment for Cu(II) and an octahedral one for Al(III). Docking results further support that Cu(II) can achieve a square planar coordination with two His and two Gln from adjacent strands. However, for Al(III), docking results only find an incomplete octahedral coordination in which two Glu and two Asp of adjacent strands are involved, thereby leaving vacant sites that would probably be occupied by solvent molecules.

Once the metal-Aβ_{42} models are determined, each model is parametrized and submitted to triplicated 1μs-long Gaussian accelerated Molecular Dynamics simulations, aiming at comparing them with metal-free fibre tendencies.
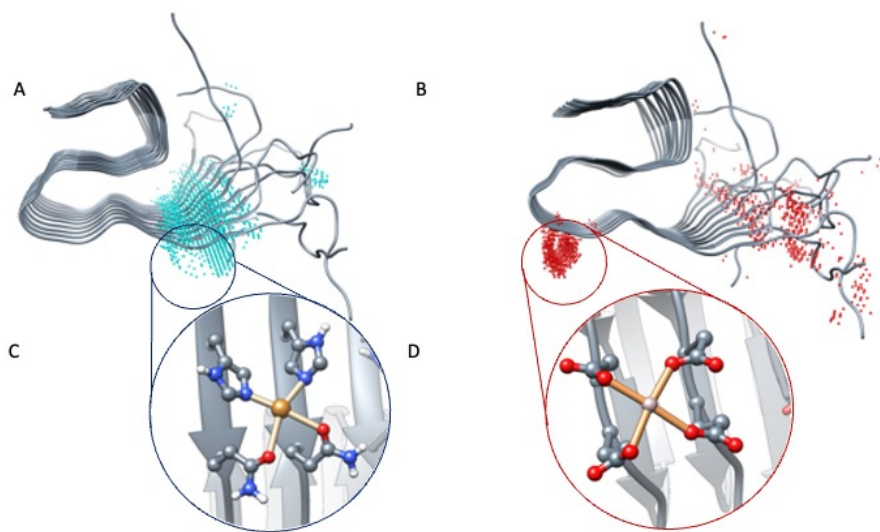
Figure 4.11: BioMetAll results of Cu(II) (A) and Al(III) (B) coordination. Docking results with GOLD software were obtained for Cu(II) (C) and Al(III) (D) in the previously predicted areas.

Copper

For metal-bound systems, simulations were also performed in triplicates, whose individual analyses are provided in Figure A.9. In all cases, the trajectories tend to converge around 400ns and show the presence of a series of minima, reinforcing the fact that amyloids are highly polymorphic.

Along the 3μs of the GaMD simulations (Figure 4.12), the same fan-like motion observed in the metal-free fibre is depicted, with the hinge between the core of the fibre and the N_{Ter} end. Strikingly though, the geometric variability of the N_{Ter} loop (residues 1-11) is significantly more

pronounced in Cu_F, as those metal-bounded pairs of strands uncouple from the movement of the adjacent pairs. Consequently, the β-sheet H-bonds between the non-metal bounded adjacent pairs are significantly weakened, changing their secondary structure towards a higher α-helix content in the N_{Ter} ends, particularly for the first strands. Such behaviour is even more pronounced than in the metal-free system, with α-helix regions being formed in the same position as in the monomeric structures. Regardless of the mentioned traits, the hinge movement previously observed in the metal-free fibre is reduced for Cu_F (Figure 4.12E), as the angle between the two principal axes (considering just the 6 central strands as done previously for the metal-free fibre) in a representative structure of the most populated cluster is 6.1°. That could be related to the fact that the fan-like movement is not globally spread all over the fibril as before, since there is a disruption of the fibrillar form due to a higher displacement of the first strands with respect to the remaining ones, probably as a result of the repulsion between adjacent Cu(II) ions. Such observations are highly remarkable, as they may indicate that Cu(II) binding tends to disrupt the fibrillar form in favour of lower-aggregation structures or amorphous aggregates.

Aluminium

For Al(III) the preferred binding site in the fibre involves residues Glu22 and Asp23 of adjacent strands, at the very core of the fibre. In this last case, the GaMD explores a series of minima until finding a well it cannot escape during the last 400ns, obtaining a very stable cluster of structures (Figure 4.13, Figure A.10).

Accordingly, the whole trajectory displays a very stable fold with minor motions in the Radius of Gyration, despite the energy profile highlighting a remarkable minimum. Recalling the fan-like movement previously observed, its amplitude is far more limited in Al_F than in the other two
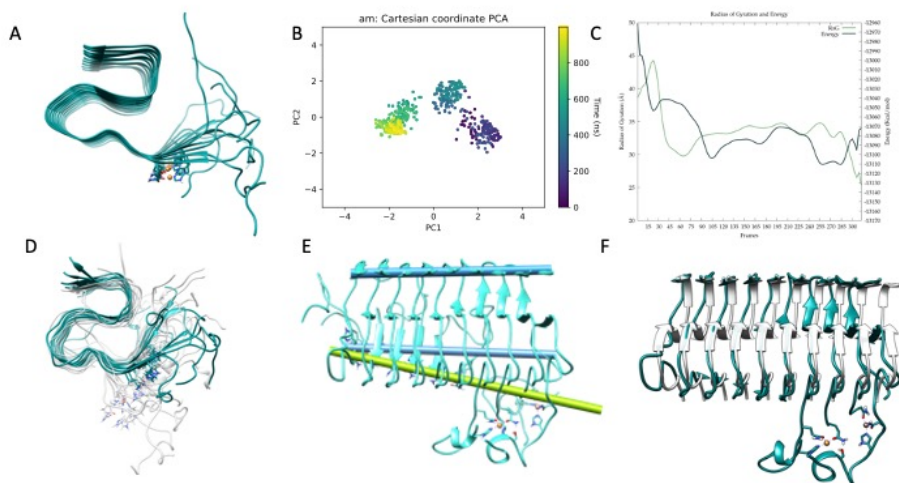
Figure 4.12: A) Cu(II)-Aβ_{42} model. B) Principle Component Analysis (PCA) of GaMD simulation. C) Energy profile and Radius of Gyration (RoG) along the GaMD simulation. D) Overlap of representative structures of most populated clusters obtained from GaMD simulations. E) In blue, axis representation of 39–41 residues (top) and 13–16 β-sheets (bottom); in green, 13–16 β-sheets axis of NMR-structure. F) Overlap between the solid-state NMR (ss-NMR) structure and the lowest energy structure of the GaMD simulation.

systems. For example, one of the most representative structures presents an inter-axial angle of only 2.7°, similar -even lower- than the one present in the ss-NMR structure, which stands for 7.8°. Therefore, GaMD simulations demonstrate that the aluminium-bound aggregate retrieves the most compact structure, even more than both free-metal and copper-bound ones and remains closer to the static view provided by experimental ss-NMR geometry. That may be an indication that Al(III) participates in further stabilising the conformation of the fibrillar structure.
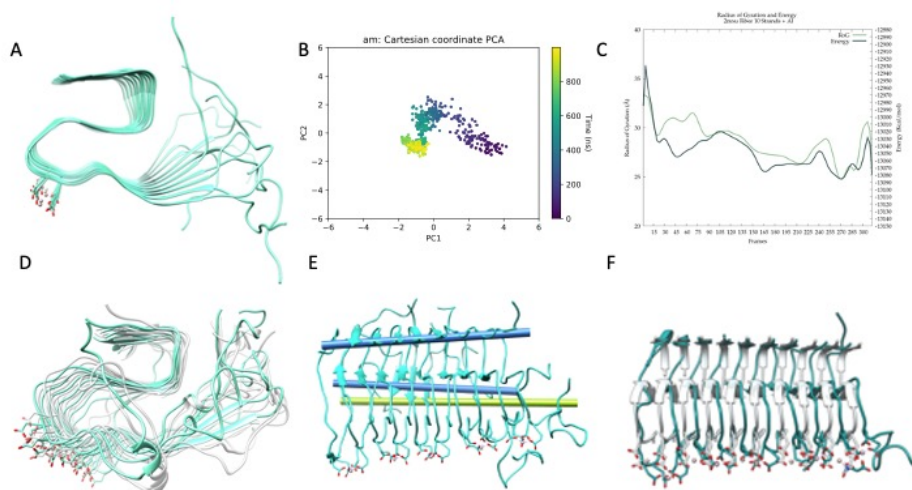
Figure 4.13: A) Al(III)-Aβ$_{42}$ model. B) Principle Component Analysis (PCA) of Gaussian accelerated Molecular Dynamics (GaMD) simulation. C) Energy profile and Radius of Gyration (RoG) along the GaMD simulation. D) Overlap of representative structures of most populated clusters obtained from GaMD simulations. E) In blue, axes representation of 39–41 residues (top) and 13–16 β-sheets (bottom); in green, 13–16 β-sheets axis of NMR-structure. F) Overlap between the solid-state NMR (ss-NMR) structure and the lowest energy structure of the GaMD simulation.

Further analysis

To further analyse how the metal ions impact the conformation landscape and ensure the stability of the lowest energy aggregates obtained from the GaMD simulations, classical MDs of 100ns were carried out. To test the exhaustiveness of the exploration and their stability, RMSD and PCA analysis were performed on the trajectories. In the previous paragraphs, special emphasis is given to how metal binding alters the fibre's low collective modes and breathing motions, with the called fan-like movement, but an exhaustive numerical description of such movement would be appropriate. For that, five distances were evaluated, restricting them to the

Cα of residues from the 6 central strands of the fibril. To ease the labelling and the reading, we use the X-ResNum identification format where X is the number of the strand the residue belongs to, Res is the type of the residue, and Num is its number in the sequence.

1. Horizontal measure in the core region ($\mathbf{D}_{HC}$): distance between 3-Gly33 and 8-Gly33 (Figure 4.14A) shows the fibre opening of the core region in the X-axis and indicates if the β-sheet interaction is weakened -longer distance- or strengthened -shorter distance-.

2. Vertical measure in the core region ($\mathbf{D}_{VC}$): the average distance from Phe19 to Val39 (Figure 4.14B) of each core strand shows the Y-axis opening, measuring the compactness of the S-shape supramolecular structure.

3. Horizontal measure in the external region ($\mathbf{D}_{HE}$): the distance between 3-Gln15 to 8-Gln15 (Figure 4.14C) indicates how the fibre opens horizontally with respect to the N_{Ter} region, a representation of the β-sheet interaction at this region.

4. Vertical measure in the external region ($\mathbf{D}_{VE}$): the average distance from Glu11 to Val39 (Figure 4.14D) of each core strand describes how the N_{Ter} tail separates from the core region of the fibre.

5. Interstrand measure ($\mathbf{D}_{IS}$): the average distance between His14 of each core strand (Figure 4.14E) differentiates between collective or individual fibre movements.

With these descriptors, it can be evaluated that the three complexes behave very similarly both in the core horizontal and vertical axes ($\mathbf{D}_{HC}$ and D_{VC}). For D_{HC}, the metal-free fibre shows a distance of 24.64 Å ($\pm$5.6), while Al(III) and Cu(II) exhibit values around 25 Å, very similar to the 24.6 Å of the ss-NMR structure. Distances of D_{VC} are also in a similar
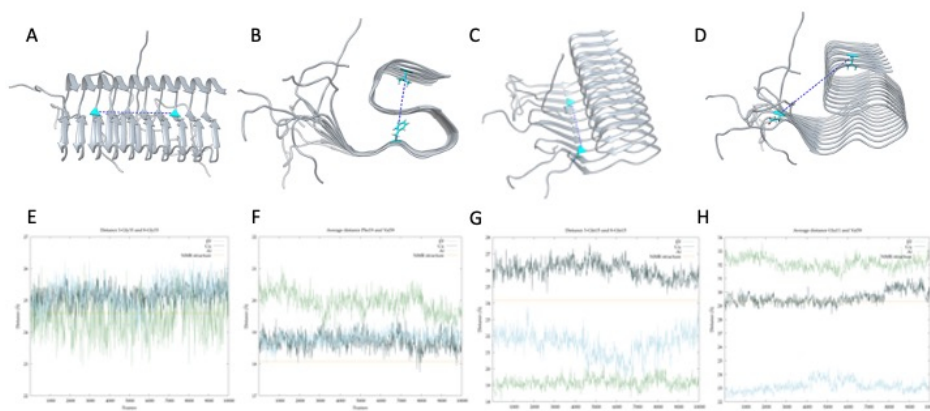
Figure 4.14: Atoms chosen for the horizontal distance (A), vertical distance (B), 11–15 β-strand intersheet distance (C) and opening of the 1–18 region (D) represented in the solid-state NMR (ss-NMR) model. Results of the horizontal (E), with an average distance of 24.6 Å (±5.6), 25.2 Å (±1.6) and 25.5 Å (±10.9) for metal-free, Cu(II) and Al(III) systems; vertical (F), with an average distance of 19.9 Å (±0.3), 18.7 Å (±0.2) and 18.8 Å (±0.2) for metal-free, Cu(II) and Al(III) systems; intersheet (G); with an average distance of 19.3 Å (±5.2), 26.06 Å (±1.1) and 21.5 Å (±0.7) for metal-free, Cu(II) and Al(III) systems; and opening (H), with an average distance of 32.1 Å (±0.4), 29.6 Å (±0.5) and 23.1 Å (±0.4) for metal-free, Cu(II) and Al(III) systems; along the MD simulation, with ss-NMR distance represented in yellow line.

range, with values of 19.9 Å (±0.3) for the metal-free fibre, 18.7 Å (±0.3) and 18.8 Å (±0.2) for the Cu and Al bound complexes respectively, and 18.3 Å for the ss-NMR structure.

Despite the similarity between complexes with general measurements, D_{HE} introduces significant differences as it relates to the disruption of the fibrillar structure in the N_{Ter} region. It is observed that Cu(II) coordination produces an increase in this distance, from 24.3 Å in the ss-NMR to 26 Å (±1.1). On the contrary, both the metal-free fibril and Al(III) bound

structure show shorter distances, of 19.93 Å ($\pm$0.3) and 21.5 Å ($\pm$0.7) respectively, in comparison to the ss-NMR structure. D_{VE}, which accounts for the fan-like movement, is remarkably increased in the metal-free fibril (32.1 Å ($\pm$0.4)) with respect to the ss-NMR (29.3 Å); whereas Cu(II) system maintains a similar value to the ss-NMR structure, with a distance of 29.6 Å ($\pm$0.5). Note, however, that this distance is an average of the six central strands and the displacement of the terminal chains is blurred by the shorter distance retrieved by the remaining. Indeed, the two most external strands included in the average exhibit values of 35 and 42 Å, respectively, while the four inner ones show values that range between 20-25 Å, again even lower than the ss-NMR. Last, the Al(III) system is characterised by being more compact than the ss-NMR structure, with an average distance of 23.1 Å ($\pm$0.4), supporting the fact that aluminium binding increases the stability of the fibrillar conformation.

The distances between each pair of strands (D_{IS}) were also measured to qualitatively define the repulsion between Cu(II) bound strands (Figure 4.15). The metal-free fibril has approximately the same distance between each pair of strands. Though such a system has higher flexibility on the vertical axis, as D_{VC} and D_{VE} also support, the horizontal axis is more compact and shows a collective movement of all the strands which corresponds to the synchronised hinge fan-like movement. However, upon metal coordination, the distance between the strands differs.

First, Cu(II) bound complex is the single one which increases the distance between each pair of strands, even between those bound to the same metal ion. The most remarkable is the increase in the 4-His14 - 5-His14 distance, which corresponds to the disruption in the collective movement of the strands.
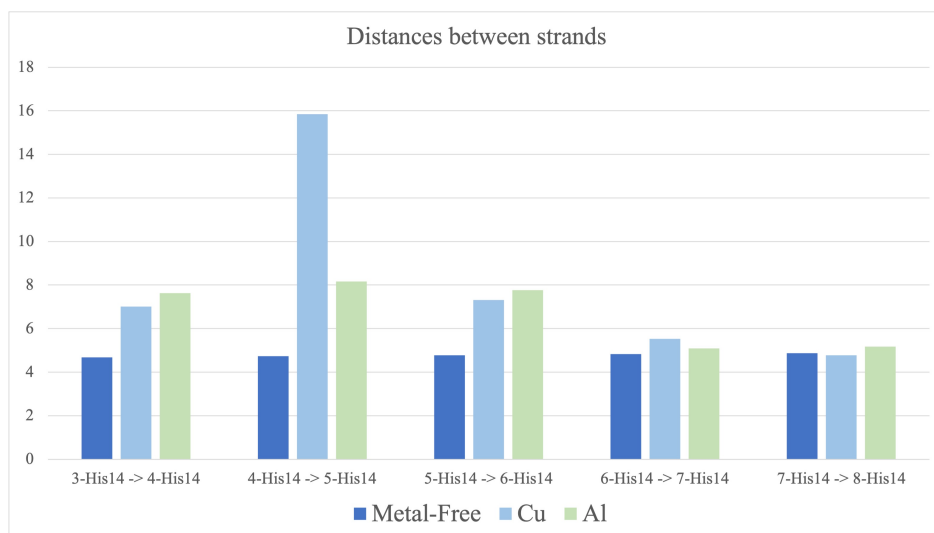
Figure 4.15: Average distance between His14 from each core strand to the adjacent one along 100ns molecular dynamic (MD) simulation for metal-free fibril (dark blue), Cu(II) (light blue) and Al(III) (green). Standard deviation ranges from 0.1 to 0.7. Strands 3–4, 5–6, and 7–8 are bound by metal coordination in Cu(II) and Al(III) systems.

For Al(III), the distance between the pairs of strands is also slightly increased, especially in strands 3 to 6, but the collective movement is highly preserved. This points to the fact that, although Al(III) binding increases the compactness and stability of the core region, as D_{HC} and D_{VC} support, the N_{Ter} region is still highly flexible.

The mentioned changes in metal coordination are also easily observed just by overlapping the systems against the ss-NMR structure (Figure 4.16), demonstrating that the metal-free fibre has a collective hinge fan-like movement, increasing the flexibility of the system but without implying its disruption. On the contrary, Cu(II) leads to the partial dismantling of the fibrillar conformation in the N_{Ter} region by breaking the

collective movement of the fibre. Lastly, Al(III) bound system is more compact in the core region, though the N_{Ter} region is still highly flexible.
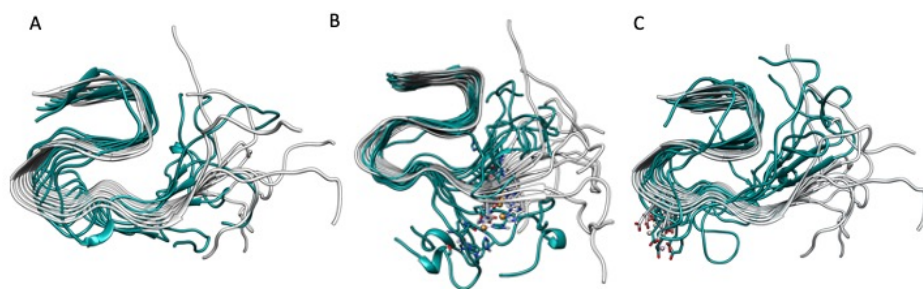
Figure 4.16: Overlapping of the lowest energy structure obtained from the GaMD simulation of Metal-Free Fibril (A), Cu(II) (B), and Al(III) (C) against the $A\beta_{42}$ crystal model (grey - PDB code 2MXU).

4.4.3   Conclusions

In the present work, an integrative computational study with BioMetAll, Molecular Docking and MD simulations has been demonstrated essential for those structures with preorganized conformations whose metal binding region is not determined, and their effect is not described.

Metal-free simulations allowed us to characterise collective low-energy vibrational modes that constitute the natural breathing motions of the S-shape structure. The main one could be described as a hinge fan-like movement due to the opening-closing tendency of the N_{Ter} region of the fibre, which could be linked to the dynamic process of aggregating and

dismantling of the Aβ fibres. Nonetheless, metal-bound complexes show a differential behaviour that depends on the metal ion. On the one side, Cu(II) bind at the N_{Ter} region with two His13 and Gln15 from adjacent strands, at a very close site to the one identified in the monomeric species. The binding of copper ion introduces charges in the N_{Ter} region, and the analyses of the simulations show that the system tends to disrupt the fibrillar structure of the most extreme strands, favouring the appearance of α-helix configurations, resembling the monomeric structure of the Aβ peptide. This fact supports the hypothesis that copper binding is more prone to lead to amorphous aggregates,[159] as experimentally shown.[179]

On the contrary, Al(III) binding takes place in the first coil region of the S-shape, on residues Glu22 and Asp23 from adjacent strands, which compensates for the Al(III) charge. In this case, aluminium ion binding reinforces the stability of the system in the core region, even reducing the hinge movement observed in the metal-free fibre complex and obtaining structures with the lowest difference with respect to the ss-NMR structure. Such fact reinforces the idea that Al(III) further stabilises neurotoxic species,[176] and a larger stability of the fibril has been related to a higher aggregation rate.[180] However, mature fibril deposits are less relevant than oligomers with respect to AD's aetiology and severity, due to their inert seeding activity.[176]

Overall, the computational study carried out in this work shows that the binding of metal ions to Aβ fibre indeed affects its dynamical behaviour, though in a different manner depending on the metal ion involved: while Cu(II) leads to less organised tertiary structures than the unbound system; Al(III) retrieves the opposite behaviour, with a high stable S-shape structure. This is in agreement with the study of Bolognin et al. [22], which demonstrates that Cu(II) prevents the formation of fibrillary aggregates, whereas Al(III) induces the aggregation of fibrillar oligomers.

Until here, the canonical form of the Aβ, either in monomeric or fibrillar form, has been studied as a reference for the aggregation effect of metal ions. Furthermore, it is of high interest and biological significance the study of some variants of the Aβ peptide sequence, since they demonstrate unique aggregation patterns. Hence, the following section aims at describing familial variants, in synergy with Cu(II) and Al(III) metal ions.

4.5 Aβ₄₂ Variants: Familial Mutations

4.5.1 Introduction

The most biologically meaningful mutations of Aβ peptide are those related to familial forms of Alzheimer's Disease. In this work, mutations in residues 22 and 23 have been studied due to their well-known impact on aggregation propensity. Concerning E22 mutations, the Italian mutation (E22G) presents an enhanced small Aβ protofibril and oligomers formation but with a lower aggregation rate; the Dutch mutation (E22Q) induces the formation of amorphous fibrils and oligomers faster, also with a faster aggregation rate than the wild-type; while the Italian mutation (E22K) presents an accelerated aggregation rate to oligomers but with fewer fibrils.[116,181] The second key position is D23, whose Iowa mutation (D23N) causes a differential bend motif which could affect folding and aggregation, also reducing the electrostatic repulsion between monomers as a consequence of a reduced charge.[182,183] In fact, all these mutations reduce the negative charge of the peptide. Nonetheless, such mutations may not only affect the monomeric structure but also higher aggregation states. As an example, D23N mutation is believed to form antiparallel β-sheet fibrils rather than the in-register β-sheet conformation adopted by WT Aβ.[181] Besides, nucleation on the surface of existing fibres may also be affected by any of these mutations due to the differential charge or

side chain size.[184] Hence, unveiling the exact mechanism by which each mutation modifies the aggregation pattern is highly interesting because it may lead to a better understanding of AD development.

Concerning the familial mutations, the changes in the secondary structure of the peptides derived from the E22 mutations have already been explored, though mostly centred on the comparison between the mutated $A\beta_{42}$ and $A\beta_{40}$ forms.[185,186] The dimerization of the E22Q mutated form has also been computationally explored, though only in the 16-22 fragment.[187,188] Some experimental studies also report the effect of the mutations in $A\beta_{40}$ fibrillar forms.[189,190] However, to our knowledge, no studies on the influence of mutations on the complete fibrillar form and the interrelation between mutations and metal binding, two of the main accepted aggregation hypotheses, have been reported.

Despite all those works setting the basis for the alternative forms of $A\beta$, a comparative study of the different effects that each of these mutations may produce at different steps of the aggregation process is still missing. Besides, none of them considered merging the mutated forms with metal binding, which would represent the synergy between two of the main aggregation hypotheses. Accordingly, here we report an integrative computational study which covers the previous situations. The monomeric and fibrillar structures of $A\beta_{42}$ with the aggregation-prone familial mutations E22Q, E22G, E22K, and D23N, simultaneously bound to Cu(II) or Al(III) binding, are analysed. For that, an extensive conformational analysis is performed by means of GaMD and cMD.

4.5.2   Results

All the familial mutations in this study have been reported to highly impact the aggregation profile of the peptide.[181,182] Although our

work aims at providing molecular insights throughout computation, simulations of the entire metal-dependent aggregation process that consider the total flexibility of the monomers, as well as their assembly to form metal-containing amorphous or fibrillar structures, is still a far-reaching challenge of molecular modelling. As an alternative, addressing monomeric and fibrillar structures separately can lead to relevant insights. Accordingly, the impact of the mutations on monomeric and aggregate structures and their relationship with metal-binding have been studied. For monomeric species, the mutations are expected to primarily impact the secondary structure, which could further alter the subsequent aggregation process. For aggregates, the mutations may affect both the stability of the tertiary structure and the metal-binding patterns.

Monomeric species

Structural descriptors for the monomeric species of the metal-free peptide, Cu(II), and Al(III) systems are summarised in Table 4.4, Table 4.5, and Table 4.6, respectively. These descriptors are i) the α-helix and ii) β-sheet contents obtained along the trajectories, which inform about the secondary structure adopted by the peptides, iii) the U-shape percentage, related to the tertiary structure of the monomers, and linked to higher aggregation structures,[119] iv) the Hydrogen Bond (HB) contacts, primarily situated in the turn (Glu22-Lys28) of the U-shape, since they have been reported to be essential for the nucleation of $A\beta$,[191] and v) the radius of gyration, related to the flexibility of the peptide. Such parameters are expected to be modified upon introducing the familial mutations in the $A\beta$ sequence. Results for the WT $A\beta_{42}$ peptide obtained in the previous sections have also been included for comparison. Figure 4.17, Figure 4.18, and Figure 4.19 show the timeline analysis and the representative structure of the MD simulation corresponding to the GaMD most populated cluster of the metal-free peptide and the Cu(II) and Al(III) bound sys-

tems, respectively. Stability analysis including the energy profile, contact map, RMSD to all, principal component analysis (PCA) and cluster countering for each complex is provided in Appendix A.3. It is worth pointing out that the analyses do not correspond to the complete GaMD simulation, as in the first section of this Chapter, but in the subsequent MD of 100ns. This is because no crystal structures are available neither for the monomeric of fibrillar structures with familial mutations, hence the GaMD simulation provided an exhaustive exploration of the conformational space and the MD ensured the finding of a minima.

For the metal-free peptide structures, all four mutations either dramatically reduce (E22Q, E22K, and D23N) or maintain (E22G) the α-helix content (Table 4.4, Figure 4.17 and Appendix A.3). The β-sheet content is increased in mutations E22G and E22K. This is particularly striking for E22K as the β-sheet content rises to 38.9% with three consistent regions (residues 6-12, 17-21, and 30-33) (Figure 4.17), which induces the disappearance of the α-helix content (0%).

Table 4.4: Results for the metal-free peptide monomeric system.

System	α-hel.[a]	β-sheet.[b]	HB (16-42)[c]	U-shape (%)[d]	RoG(Å)[e]
FP WT	48.8%	0%	None	0%	23.1±10.7
FP E22Q	12.8%	0%	D23-S26 (af 47.0%)	0%	23.0 ± 8.2
FP E22G	48.2%	6.8%	None	0%	23.5 ± 8.5
FP E22K	0%	38.9%	D23-S26 (af 19.7%)	0%	25.7 ± 9.2
FP D23N	30.9%	0%	N23-S26 (af 91.8%)	0%	25.6 ± 9.8

[a]α-helix percentage along the trajectory. [b]β-sheet percentage along the trajectory. [c]HB contacts with an average frequency of at least 10%, suggested to participate in the U-shape maintenance. [d]Percentage of U-shape structure presence along the trajectory, counted by adding up the percentage of clusters with such structure. [e]Average Radius of Gyration along the GaMD trajectory and its standard deviation.
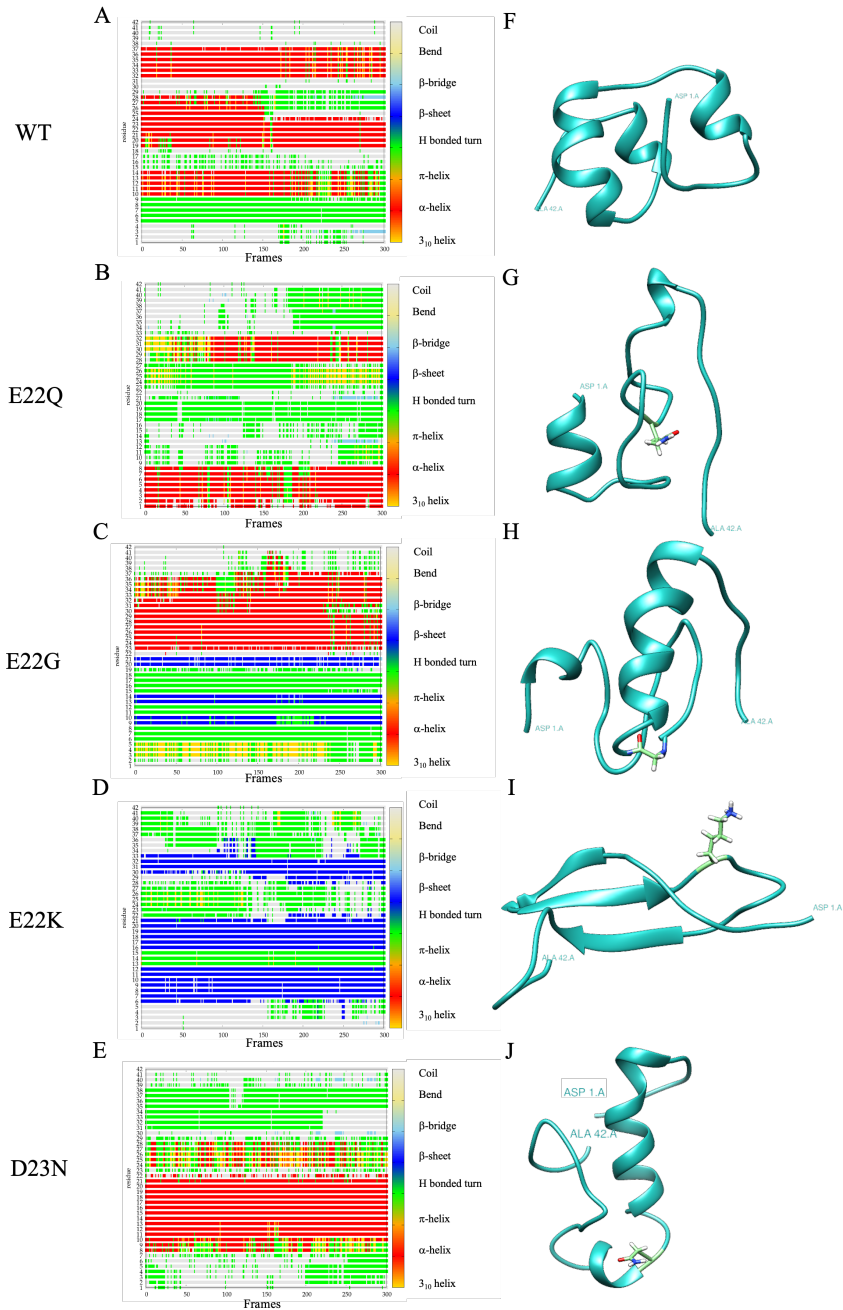
Figure 4.17: A-E Timeline analysis of the MD trajectory for FP WT, E22Q, E22G, E22K and D23N mutations. The β-sheets correspond to the blue regions. F-J Representative structure of the most populated cluster from the MD of each variant. The mutated residue, when present, is represented in light green.

However, the mutated systems' flexibility is similar to the WT's, according to the similar Radius of Gyration (23.1 Å for WT and between 23.0 Å and 25.7 Å for the variants) and the significant standard deviation. Furthermore, hydrogen bonds (HBs) in the U-turn, known for their role in promoting aggregation,[191] appear with all mutations except E22G (47.0% in E22Q, 19.7% in E22K, 91.8% in D23N), while in the WT they were not formed. The appearance of HBs in such systems does not correlate with the percentage of the U-shape as initially expected. Thus, the data show that mutations have a clear impact on the general fold of the metal-free peptide, which could already suggest significant changes in how the system pre-organizes for metal-binding and/or aggregation processes.

Upon metal coordination, the effects observed correspond to those already reported in the previous section: both Cu(II) and Al(III) coordination reduce the Radius of Gyration, producing less flexible structures, and increasing their α-helix content. Indeed, metal binding redistributes the helical regions leading to U-shape structures that are stabilised by hydrophobic interhelical contacts. Here, comparisons are performed among the complexes with the same coordinated metal ion to describe the effect of the mutations.

Cu-WT (Table 4.5, Figure 4.18), exhibits a high α-helix content (69.1%), involving residues from 13 to 24 and 28 to 37, and a 100% U-percentage. Regarding Cu-E22Q, Cu-E22K and Cu-D23N mutated complexes, minor differences are observed in comparison to the Cu-WT complex, obtaining a similar α-helix content (69.2%, 66.3% and 74.5%) and a 100% percentage of U-shape structure. On the contrary, Cu-E22G reduces the α-helix content and abolishes the U-shape structure. Besides, only Cu-E22G produces several long-lasting β-sheet regions (7.7%), involving residues 5-6, 9-10, 13-14 and 35-36 (Figure 4.18) in agreement with glycine being a helix-breaker. Overall Cu-E22G keeps similar values as those of the

metal-free system E22G. It is remarkable that the Cu-E22K system does not retrieve any β-sheet content at all, while such mutation exhibited a substantial β-sheet increase in the metal-free system.

Table 4.5: Results for Cu(II)bound monomeric system.

System	α-hel.[a]	β-sheet.[b]	HB (16-42)[c]	U-shape (%)[d]	RoG(Å)[e]
Cu-WT	69.1%	0%	E22-L28 (af 23.7%)	100%	15.0 ± 0.7
Cu-E22Q	69.2%	0%	**None**	100%	18.3 ± 1.8
Cu-E22G	33.8%	7.7%	**None**	0%	17.5 ± 1.0
Cu-E22K	66.3%	0%	**None**	100%	18.3 ± 3.0
Cu-D23N	74.5%	0%	**None**	100%	16.7 ± 1.7

[a] α-helix percentage along the trajectory. [b] β-sheet percentage along the trajectory. [c] HB contacts with an average frequency of at least 10%, suggested to participate in the U-shape maintenance. [d] Percentage of U-shape structure presence along the trajectory, counted by adding up the percentage of clusters with such structure. [e] Average Radius of Gyration along the GaMD trajectory and its standard deviation.

For Al(III) systems (Table 4.6, Figure 4.19), the Al-WT form shows both high α-helix content (68.3%) and U-shape percentage (96.7%), with transitory β-sheet regions (0.3%) appearing along residues 7-8 and 40-41. Al-E22Q and Al-E22G show both a smaller α-helix (55.2% and 51.9%, respectively) and U-shape content (71.4% and 76.3%) compared to Al-WT. For Al-E22G variant there is also a slight increase of β-sheet content (2.4%), as two stable β-sheet regions are identified at residues 12-13 and 39-40. Both Al-E22K and Al-D23N produce a slight increase in the α-helix content (71.0% and 70.5%) and in the U-shape (100% and 99.7%), abolishing the β-sheet content (Figure 4.19). In general, the timeline analysis (Figure 4.19) indicates that all Al-bound systems exhibit a less organized secondary structure pattern than the copper-containing systems, in agreement with the larger Radius of Gyration values obtained for the formers (Table 4.6). Such differences may be linked to Al(III) coordination introducing a higher charge than Cu(II), further affecting distal areas.

Table 4.6: Results for Al(III)-bound monomeric system.

System	α-hel.[a]	β-sheet.[b]	HB (16-42)[c]	U-shape (%)[d]	RoG(Å)[e]
Al-WT	68.3%	0.3%	None	96.7%	18.9 ± 1.8
Al-E22Q	55.2%	0%	None	71.4%	22.6 ± 5.1
Al E22G	51.9%	2.4%	None	76.3%	19.6 ± 1.7
Al-E22K	71.0%	0%	None	100%	17.6 ± 1.8
Al-D23N	70.5%	0%	E22-L28 (af 73.7%)	99.7%	20.3 ± 1.5

[a]α-helix percentage along the trajectory. [b]β-sheet percentage along the trajectory. [c]HB contacts with an average frequency of at least 10%, suggested to participate in the U-shape maintenance. [d]Percentage of U-shape structure presence along the trajectory, counted by adding up the percentage of clusters with such structure. [e]Average Radius of Gyration along the GaMD trajectory and its standard deviation.

Overall, mutations at residues E22 and D23 and metal binding have different impacts on the structure and dynamics of the Aβ$_{42}$ peptide. Without metal, all mutations reduce the amount of α-helix and increase, in some cases, the β-sheet content. This may correlate with experimental observations showing faster aggregation for E22 mutants.[181] The most striking case is the E22K variant, for which stable β-sheet motifs (up to 39%) are observed during the trajectory.

In the presence of Cu(II) and Al(III) metal ions, the peptide becomes less flexible, the β-sheet content decreases in favour of forming α-helix motifs. Such helices are displaced toward the C$_{Ter}$ end compared to the metal-free system, further stabilised by interhelical hydrophobic patterns, promoting the formation of U-shape configurations; even for E22K, for which stable β-sheet motifs were observed in the non-metallic system. The differences between variants, though, are not so evident in the presence of the metallic ions, the effect of metal coordination being more important than the mutation itself. The only exception is Cu-E22G, which compared to the metal-free system, no major changes are observed.
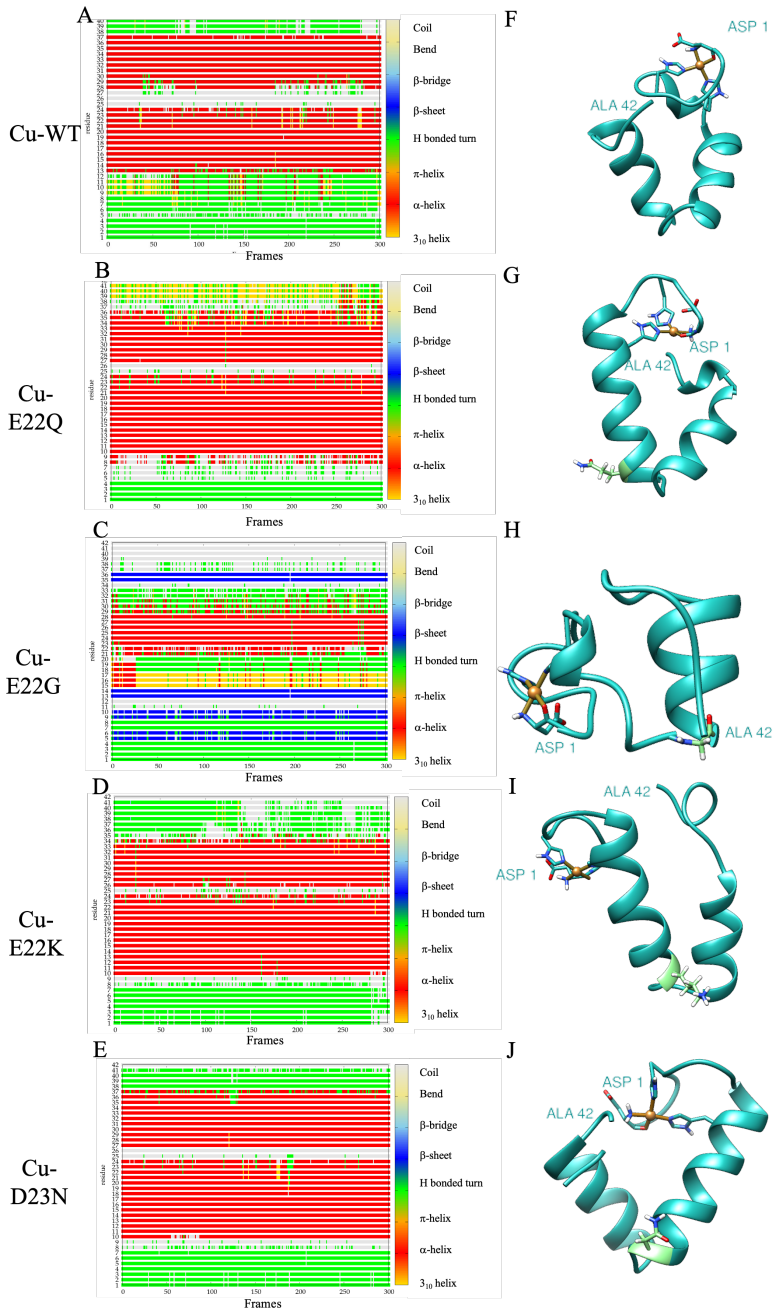
Figure 4.18: A-E Timeline analysis of the MD trajectory for Cu-WT, Cu-E22Q, Cu-E22G, Cu-E22K and Cu-D23N mutations. The β-sheets correspond to the blue regions. F-J Representative structure of the most populated cluster from the MD of each variant. The mutated residue, when present, is represented in light green.

129
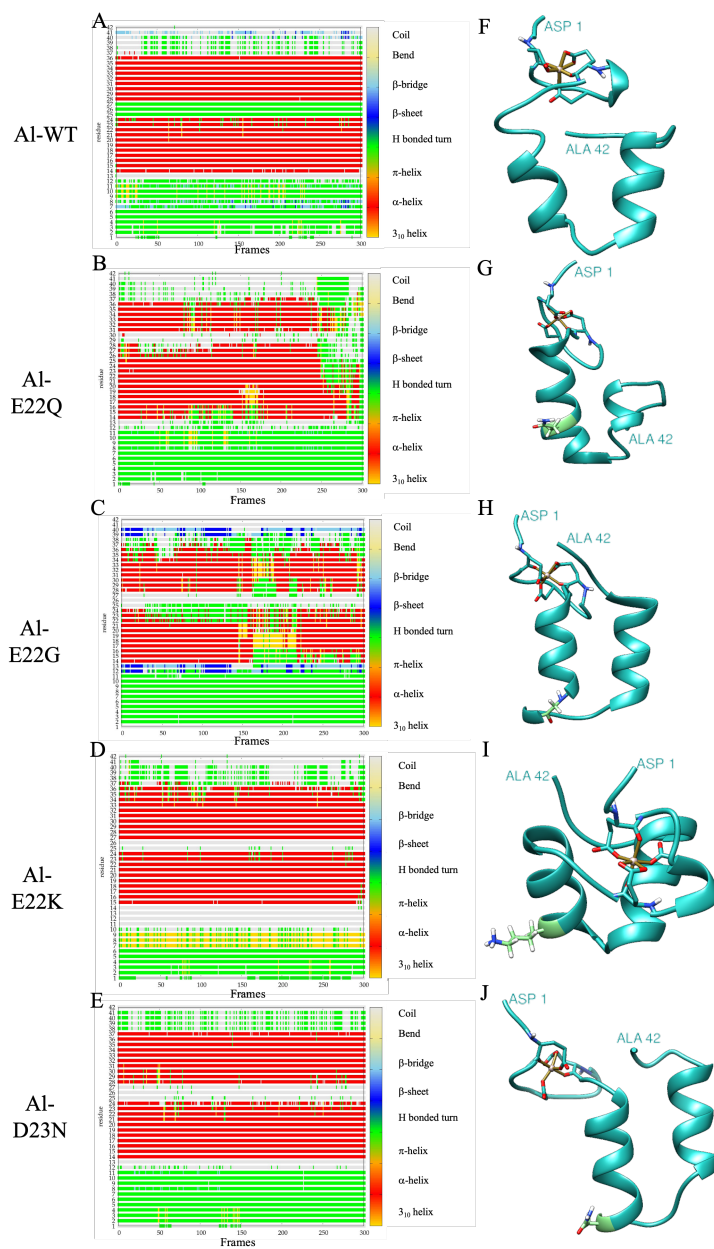
Figure 4.19: A-E Timeline analysis of the MD trajectory for Al-WT, Al-E22Q, Al-E22G, Al-E22K and Al-D23N mutations. The β-sheets correspond to the blue regions. F-J Representative structure of the most populated cluster from the MD of each variant. The mutated residue, when present, is represented in light green.

Fibrillar species

Our study's next step aims to assess familial mutations' and metal binding impact on fibrillar species. Since no crystallographic structures for the mutated Aβ fibrillar forms are available, the WT crystallographic structure has been selected as a starting conformation. For the metal-free and the Cu(II)-bound systems, it has been possible to study the four variants (E22Q, E22G, E22K and D23N), maintaining the coordination sphere defined in our previous section (Figure 4.20). Otherwise, for the E22K, E22G and D23N mutations, BioMetAll was not able to find a compatible configuration for Al(III) binding since the mutated E22 and D23 residues, previously found in the metal coordination sphere, are substituted by Gly, Lys or Asn, which do not allow a proper coordination. For E22Q, however, a feasible binding site which includes the mutated Q22 and D23 residues was predicted by BioMetAll (Figure 4.20). Hence, simulations with Al(III)-bound to the fibrillar system were only possible for the E22Q variant.

In the study of WT metal-free and metal-bound Aβ₄₂ aggregates, a collective motion in the form of a fan-like displacement was identified. Present simulations for the four familial variants in metal-free forms show that the fan-like movement is maintained. However, the calculation shows a lower amplitude, particularly for the E22G mutant (Figure 4.21) thereby indicating that the fibril is more rigid and less prone to disaggregate.

To further analyse the changes induced by mutations in the tertiary structure of the fibres, the same five-distances analysis applied in the previous section (Figure 4.14) was used to measure the compactness of the fibre. The analysis of all distances indicates that those corresponding to the core of the fibre ($\mathbf{D}_{HC}$ and $\mathbf{D}_{VC}$) do not change significantly upon mutation. However, those involving residues at the most external part of the fibre suffer more significant changes ($\mathbf{D}_{HE}$ and $\mathbf{D}_{VE}$), mainly when bound
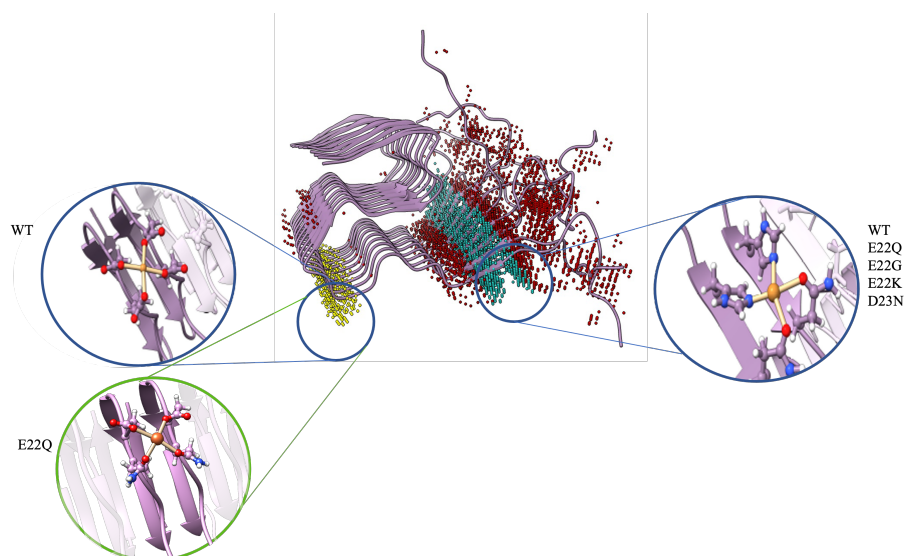
Figure 4.20: Representation of BioMetAll results. On the left, coordinations obtained for Al(III) bound to WT and E22Q variants are represented. On the right, coordination is found for all the Cu(II) bound complexes.

to metal ions, and thus, only values for these descriptors are reported in Figure 4.22. The remaining ones are given in the Appendix (Figure A.31, Figure A.37 and Figure A.40).

For metal-free fibres, the most remarkable result is observed for the most representative distance in the fan-like movement, $\mathbf{D}_{VE}$ (Glu11-Val39), which shows a significant decrease in E22K (22.6 Å), E22G (23.8 Å), E22Q (28.7 Å) and D23N (30.4 Å) compared to the WT (32.1 Å) (Figure 4.22). Thus, the four mutations tend to reduce the system's flexibility and stabilise the S-shape of each monomer by reducing the fan-like movement. This reduction of the breathing movement could be attributed to the fact that the mutations substitute a negatively charged residue (E) with a neutral (Q and G) or a positively charged (K) one, which reduces the repulsion with D23. Furthermore, E22K allows the formation of strong
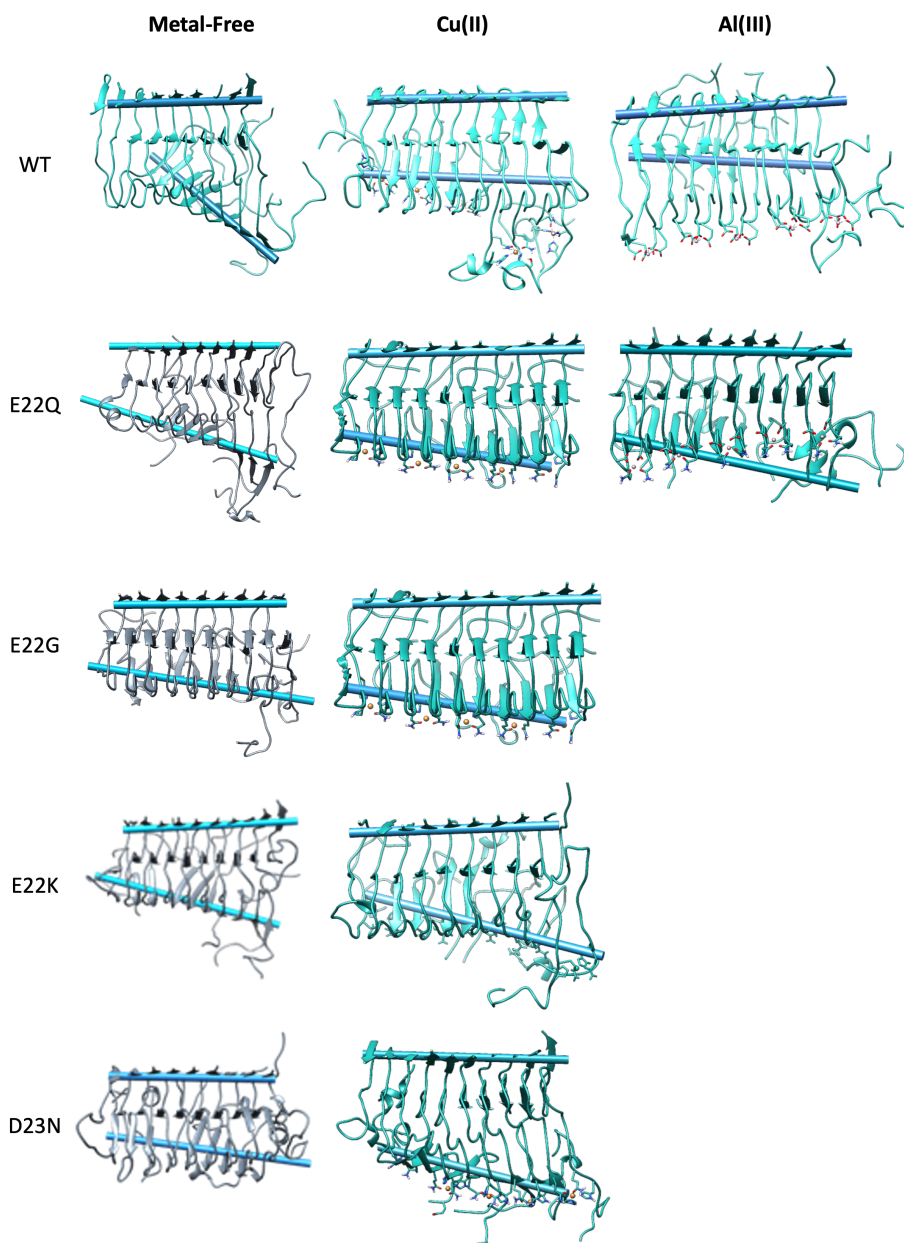
Figure 4.21: The most populated cluster obtained from the GaMD simulation for the Metal-Free, Cu(II), and Al(III) bound forms in WT, E22Q, E22G, E22K, and D23N variants. In blue, the two axes comprised of residues 39-41 (top) and residues 13-15 in β-sheets (bottom) are represented.

hydrogen bonds between Lys22 and Asp23 (mainly of the same strand), and E22Q the formation of mild hydrogen bonds between Gln22 and Asp23, which hinder such movement.

With respect to the Cu(II)-bound aggregate, both the fan-like movement and the intrinsically disordered behaviour of the N_{Ter} region (residues 1 to 11) are observed among the variants (Figure 4.21). However, the fan-like movement is reduced for all mutants compared to the WT (see Figure 4.22) since the $\mathbf{D}_{VE}$ values (21-28 Å) for Cu(II)-mutants are smaller than that of the Cu(II)-WT (29-30 Å). Indeed, the formation of hydrogen bonds for Cu-E22K and Cu-E22Q systems is also observed, as in the metal-free systems. On the other hand, for Cu(II)-E22Q and Cu-E22K, the horizontal breathing motion increases the $\mathbf{D}_{HE}$ distance up to 35 and 31 Å, respectively (Figure 4.22), which is much larger than that observed for Cu(II) bound WT, E22G and D23N variants (24-27 Å). Interestingly, in our previous study, we observed that under Cu(II) binding to WT, some strands start to disassociate in the course of the simulation, thereby suggesting a possible disruption of the aggregate, as the $\mathbf{D}_{IS}$ distance between 4-His14 and 5-His14 increases significantly (Figure 4.23). This effect is absent in the Cu(II)-E22G systems, and is slightly reduced for Cu(II)-E22Q, Cu(II)-E22K and Cu(II)-D23N systems, suggesting that the mutations' effect compensate for the perturbation caused by metal binding. Remarkably, Cu(II)-E22Q and Cu(II)-E22K variants show large distances both for $\mathbf{D}_{HE}$ and $\mathbf{D}_{IS}$, while Cu(II)-WT demonstrates an increased $\mathbf{D}_{IS}$ magnitude. This is related to the fact that the disruption observed between strands 4 and 5 is not considered by the $\mathbf{D}_{HE}$ measure, which measures the global distance between strands 3 and 8.

Altogether, the analysis of the four familial variants shows the mutations to impact at different levels on the aggregate structure. Associated with the binding of Cu(II), mutations tend to improve the stability of the
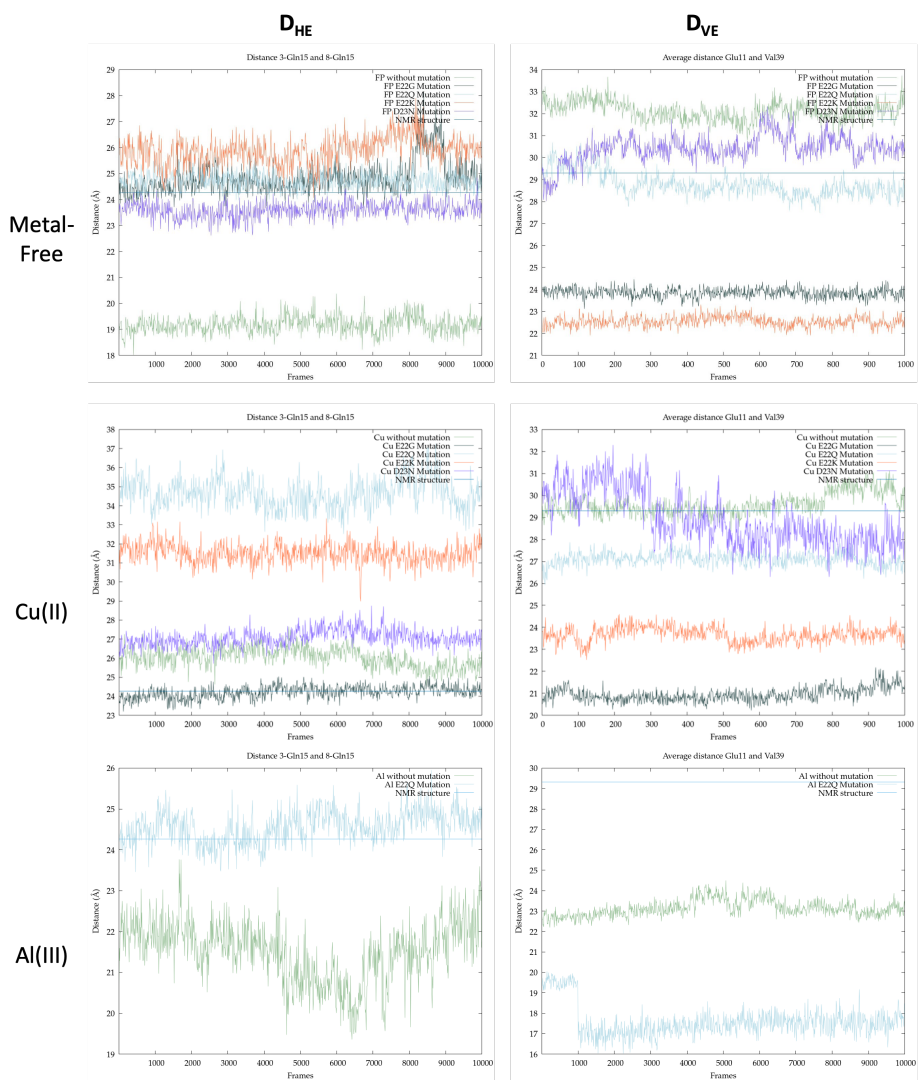
Figure 4.22: Metal-free, Cu(II) and Al(III) bound distances of the 3-GLN15 - 8-GLN15 measure (**D**$_{HE}$) and GLU11-VAL39 measure (**D**$_{VE}$) for WT (light green), E22Q (light blue), E22G (dark green), E22K (orange) and D23N (purple) variants.

S-shape form of the fibre, though destabilizing the interstrand interactions, except for Cu(II)-E22G, which retrieves a more compact system. The higher interstrand stability caused by Cu(II)-E22G may be related to the fact that changing a Glu by a Gly reduces the size of the lateral chain of the residues and, hence, the hydrophobic repulsion.
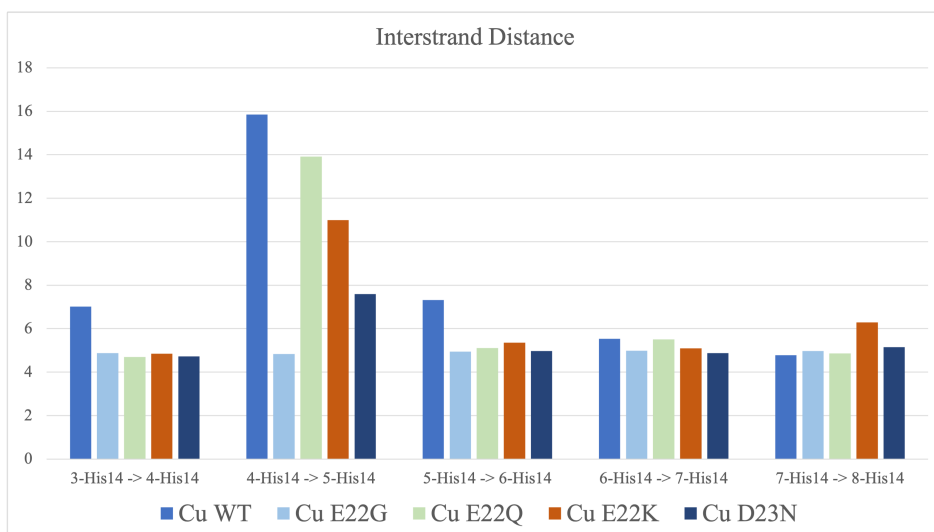
Figure 4.23: Interstrand distance ($\mathbf{D}_{IS}$) in Cu(II) bound systems.

Regarding the Al(III) bound systems, only one mutation could be studied, corresponding to the Al(III)-E22Q variant, since metal coordination occurs with residues 22 and 23, and in all other variants, the introduced residues are not capable of coordinating Al(III). For Al(III)-E22Q, molecular dynamics simulations show a slight disruption of the lateral strands related to the fan-like movement, also visible in the Al(III)-WT form (Figure 4.21). Concerning the measures, the effect of the E22Q mutation on the Al(III) bound systems is consistent with the one already observed for Cu(II), increasing $\mathbf{D}_{HE}$ up to 25 Å but reducing $\mathbf{D}_{VE}$ up to 17 Å (Figure 4.22). Thus, such a variant consistently increases the breathing movement in the horizontal axis while the S-shape is further stabilized. Be-

sides, the structural organization of the N_{Ter} region in the Al(III)-E22Q complex is maintained similarly to the WT system.

4.5.3 Conclusions

This study analyses the dynamical behaviour of four of the most relevant familial mutations of AD patients in both monomeric and aggregate forms of the $A\beta_{42}$, in the absence and presence of Cu(II) and Al(III) metal cations.

Regarding the metal-free monomers, all the mutations have a similar effect, leading to a decrease in the α-helix content. Besides, the β-sheet content substantially increases for E22K and E22G (39% and 7%, respectively). Since the β-strand structure is a significant marker of the fibril, these results suggest that these mutations could favour the formation of well-organised aggregates. When considering familial mutations and metal binding, the metal effect seems prevalent in mutations, inducing the formation of U-shape structures stabilised by interhelical contacts. The only exception is Cu(II)-E22G, for which the U-shape structure does not appear due to the formation of β-sheets, which are maintained throughout the 1μs of the GaMD simulation. The presence of β-sheet content is also observed for Al(III)-E22G, though with less prevalence. This agrees with glycine being a known helix breaker, which leads to lower percentages of U-shape among E22G structures. Differences between Cu(II) and Al(III) systems, though, can be related to the differences in metal coordination, which affects the distribution of α-helices in the monomer.

Regarding the impact of the familial mutations on the fibrillar structures, the observed effect of metal binding is consistent with that reported for WT-$A\beta_{42}$ fibrillar form, since Cu(II) interactions at the N_{Ter} region of the

fibre disrupt the characteristic tertiary structure of the fibre. In contrast, Al(III) binding further stabilises the S-shape. However, the effect of the amino acid substitutions seems to prevail over those related to metal-binding, as consistent trends are observed among systems regardless of the metal-binding state. As observed, E22Q and E22K mutations decrease the opening of the fibre in the vertical axis. Such reduction of the fan-like breathing movement could be attributed to the formation of strong and mild hydrogen bonds between Lys22 and Gln22 with Asp23, respectively, together with the reduction of the overall charge of the system. The E22G mutation seems to further stabilise the fibrillar structure, primarily by reducing its disorganisation and leading to more compact fibres, both in the vertical and horizontal axes, which is attributed to the reduction of the negative charge and the absence of lateral chain at residue 22. These changes could be responsible for the differential aggregation rates observed for E22 mutations.[116,181,192] Lastly, the D23N mutation reduces the system's stability, likely due to the abolishment of the D23-K28 salt bridge, leading to the most flexible complexes of the entire set and involving motions both in the horizontal and vertical axes of the aggregates. Such results correlate well with experimental observations in which the D23N fibre is less thermodynamic and kinetic stable.[189]

5

Catalytic fibres

This Chapter presents the studies performed on the functional side of amyloid aggregates regarding catalysis. Two main reactions, hydrolysis and cyclopropanation, have been described upon metal-mediated catalysis both on artificial minimalist classical and prion-like amyloid fibres, formed by heptapeptide strands.

5.1   General Overview

Most of the biological processes that happen inside living organisms are chemical reactions, usually facilitated by enzymes. Enzymes act as biological catalysts, accelerating the rate of chemical reactions and allowing them to occur at biological conditions, without altering themselves along the process. The enzyme's protein sequence is deterministic of its folding pattern, hence also of its function and specificity. Depending on its specificity, an enzyme interacts with a certain substrate -or various if the specificity is low- to catalyse a concrete reaction. To facilitate the reaction, the catalyst lowers the energy barrier by forming a transition state with the substrate. Once the intermediate is broken down into products, the unchanged enzyme is freed to react with further substrate molecules.[193]

The complex mechanism by which proteins have evolved to achieve specific enzymatic activity remains unknown, though it has been suggested that primitive enzymes could have been formed by short amyloid-like fibres or polymers. Up to this point, it should go without saying that amyloid fibrils are typically associated with protein misfolding, aggregation and neurodegenerative diseases, such as AD. However, recent studies have also demonstrated the potential functional role of amyloid fibres in biological processes[194] and their capacity as building blocks for new supramolecular structures.[195] Even $A\beta_{42}$ has been demonstrated to also perform catalytic functions.[196] Due to their stability, robustness, and ease of being tuned, amyloid fibres may provide a favourable environment for chemical reactions to occur. Hence, artificial amyloids with short peptides have been synthesised as an attractive platform for synthetic chemistry. Several studies have demonstrated the astonishing capacity of amyloid fibrils to perform plenty of chemical reactions, such as hydrolysis, O_2 activation or enantioselective reactions.[194,196–200]

Regarding their structure, those amyloid fibrillar assemblies are characterised by adopting a cross-β structure in which β-strands are perpendicular to the fibril axis. In such assemblies two β-sheets are in contact forming a steric zipper, which is mainly composed of hydrophobic residues in classical amyloids, or enriched in polar residues in prions, these latter ones often referred to as *polar zippers* (Figure 5.1). These architectures are further stabilised by hydrogen bonds formed either among β-strands or β-sheets.
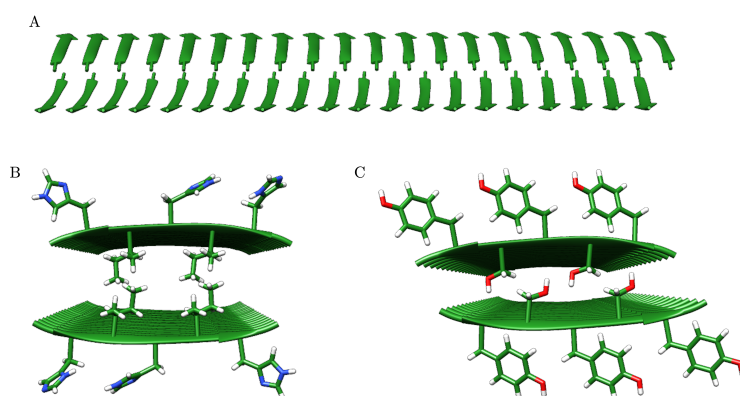
Figure 5.1: Example of fibrillar structure, with β-sheets forming a steric zipper (A). Classical amyloids present hydrophobic residues in the steric zone (B), while polar zippers present polar residues (C).

Short amyloid fibrils usually exhibit polymorphism, as the same sequence may adopt different quaternary structures and, therefore, different properties. One of the most interesting amyloid fibrillar sequences are the minimalist heptapeptides, which with only seven amino acids retain the self-assemble capacity and form stable aggregates, either classical[66] or prion-like.[195,201] Thus, their inexpensive synthesis allows the creation of suitable surfaces for chemical reactions, opening a completely new field for nanomaterial synthesis.[202] However, identifying the most favourable fibrillar structure among all the possibilities -and, conse-

quently, its function- represents a great challenge, since the same peptide can yield different supramolecular structures. The stability of such structures is determined by a delicate balance between hydrophobic, van der Waals, hydrogen bonds, and electrostatic interactions. These architectures are classified into eight classes of steric zippers[203,204] according to i) intra-sheet organisation (parallel or antiparallel), ii) how side chains are packed at the interface (face-to-back, face-to-face or face = back) and iii) the relative orientation of the β-sheets (up–up, up–down or up = down) (Figure 5.2). Since experimental atomic resolution determination by ssNMR and cryoTEM technologies is challenging, computational simulations can provide important complementary information to elucidate preferred architectures.[205,206]

Metal ions are known for playing a central role in the catalytic activity within amyloid fibres. Though few cofactor-free fibre catalysts have been described,[207] metal ions typically influence fibres' folding through their interactions with residues and are part of the proper catalytic reaction as metallic cofactors. Experimental[66,208,209] studies have already demonstrated the capacity of metal ions to enhance the catalytic activity of amyloid structures, especially Zn(II) and Cu(II). Despite some advances in understanding the catalytic function of these fibres by means of computational simulations,[210] much remains unknown about catalytic mechanisms or how their structure adapts to cofactors' presence, both being highly important to continue enlarging the whole range of applications for which catalytic fibres could be applied.

The aim of this Chapter is to shed light on the role of the fibre structure, derived from a particular peptide sequence, behind two of the most relevant reactions, hydrolysis and cyclopropanation, both performed by heptapeptides amyloid fibres in the presence of metallic cofactors. To do so, the computational approaches already mastered in the $A\beta_{42}$ aggregation

study will be appropriately adapted, including Quantum Mechanics and QM/MM for the detailed description of the hydrolysis catalytic mechanism.
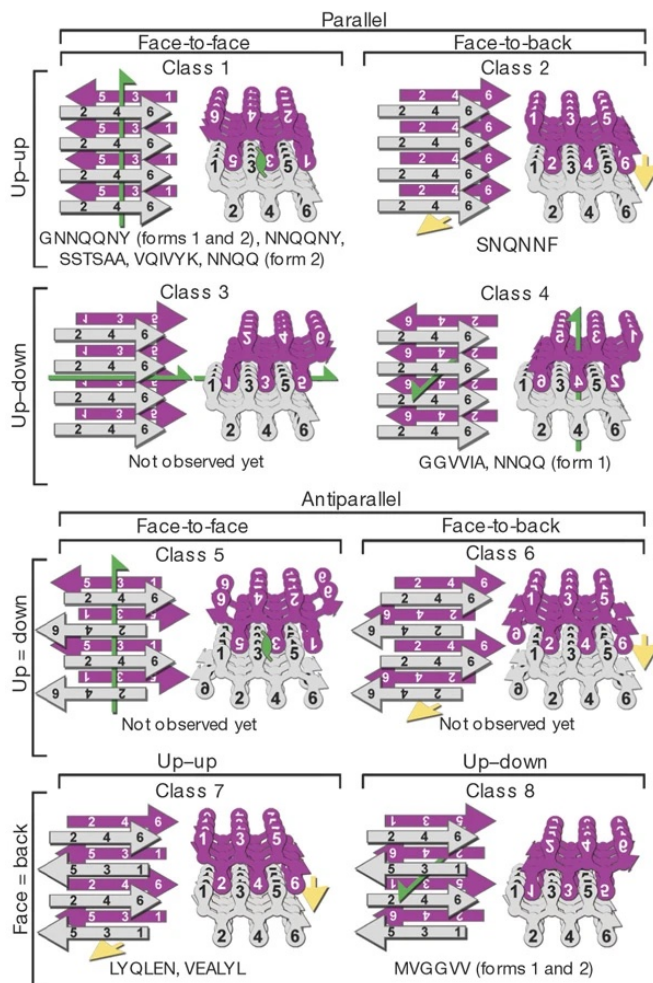
Figure 5.2: Two identical sheets can be classified by: the orientation of their faces (either 'face-to-face' or 'face-to-back'), the orientation of their strands (with both sheets having the same edge of the strand 'up', or one 'up' and the other 'down'), and whether the strands within the sheets are parallel or antiparallel. Image from [204].

5.2   Hydrolysis study

5.2.1   Overview

Among all the possible reactions that enzymes can catalyse, the pioneering study of Rufo et al.[66] demonstrated that seven-residue amyloid peptides with an amphiphilic β-strand, i.e. alternating hydrophobic residues, associate via hydrophobic interactions to form extended amyloid structures that in the presence of Zn(II) metal ions exhibit esterase activity and catalyse the hydrolysis of p-nitrophenylacetate substrate (pNPA from now on, Scheme 5.1). In their work, different hydrophobic cores formed either by Ala, Leu, Iso or Val were designed to evaluate their capacity to form stable fibrillar structures. As a result, the replacement of Leu with Ile or Val, which demonstrate higher β-sheet propensity, produce fibres with a remarkable increase in activity (Sequences 1, 2 and 3 in Table 5.1). On the contrary, the replacement to Ala, the smallest hydrophobic residue, showed lower β-sheet formation and resulted in an inactive catalyst. In the study of Rufo et al.[66], the fibrillar-forming capacity was demonstrated severely linked to the catalytic activity.
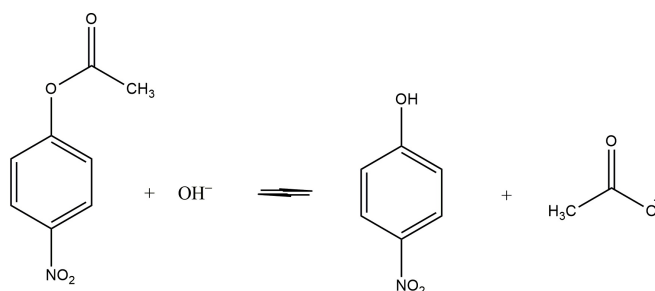
Scheme 5.1: pNPA hydrolysis, with acetate and p-nitrophenol as products.

Table 5.1: Catalytic results obtained by Rufo et al.[66] for hydrophobic-core catalytic fibres.

Reference	Sequence	K_{cat}/K_M $(M^{-1}s^{-1})$
1	LHLHLQL	30 ± 3
2	IHIHQI	62 ± 2
3	VHVHQV	32 ± 2
4	IAIHIRI	0.36 ± 0.16
5	LHLHLHL	0.2 ± 0.1
6	LHLHLYL	13 ± 5
7	LHLHLRL	18 ± 4

Regarding the hydrophilic side, His residue is conservatively selected to be exposed to the solvent, taking advantage of its natural capacity of binding Zn(II) ion and, thus, providing the catalytic activity. Indeed, the coordination of the Zn(II) ion was demonstrated to be essential since the absence of the metal ion considerably decreased the catalytic activity. Not only must the ion be present but a proper coordination mode with the fibre must be encountered, as mutating His by Ala resulted in the loss of Zn(II) coordination and a 60-fold decrease in activity (Sequence 4 in Table 5.1). Moreover, further alternatives of the aforementioned peptide, with Leu in the hydrophobic core, were designed by replacing the 6th position of the fibre (His) with Asp, Glu, Lys, Arg, Gln and Tyr, in order to explore their assisting role along the catalytic process (Sequences 1 and 5 to 7 in Table 5.1). In fact, the combination of the most stable hydrophobic core, formed by Iso; with the amyloid-promoting residue Gln in the sixth position, resulted in the most active catalyst (Sequence 2 in Table 5.1).

Regarding the fibre structure, the work by Lee et al.[211] studied Zn(II) coordination to the IHVHLQI heptapeptide sequence by NMR techniques, which is morphologically similar to IHIHIQI. The parallel face-to-back geometry (Figure 5.2) was determined to be the preferred ori-

entation, with His and Gln residues orientated towards the solvent (Figure 5.3). The Zn(II) coordination was also detailed in their study, observing very interesting traits as Zn(II) was coordinated in a 1:1 molar ratio, with half of the histidines double coordinating to two different Zn(II).
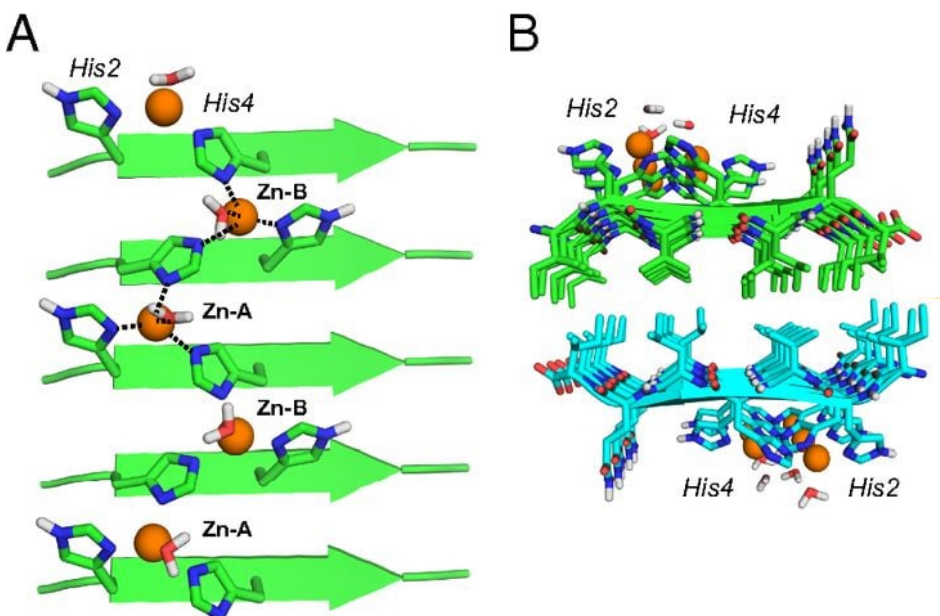
Figure 5.3: Coordination structure of Zn(II)-bound HHQ fibrils. A. Energy-minimized structure of one β-sheet. Each Zn(II) (orange) is coordinated by three His Nitrogens from two neighbouring strands. Half of the histidines bridge two Zn(II) ions. B. Two β-sheets stack with the hydrophobic residues facing each other and with the strands in the two sheets having antiparallel orientations. Image from Lee et al. [211].

The work of Navarro et al.[201] demonstrates that polar zippers can also build up prion-like fibres despite their hydrophilicity, and also perform hydrolysis. Nonetheless, their catalytic activity is substantially lower (Table 5.2) than the one reported for hydrophobic steric zipper fibres (Table 5.1).[66]

Table 5.2: Catalytic results obtained by Navarro et al.[201] for steric zipper fibres.

Reference	Sequence	K_{cat}/K_M ($M^{-1}s^{-1}$)
1	NYNYNYN	2 ± 1
2	QYQYQYQ	5 ± 2
3	SYSYSYS	6 ± 2
4	GYGYGYG	6 ± 2

Despite the great potential that amyloid fibres can present, to our knowledge only one computational study[210] has been performed to determine the esterase catalytic mechanism on amyloid fibres. Besides, this work just considers a two-strand model to study the catalytic mechanism, the effect of the local structure of the fibre being introduced by imposing geometry restrictions. Thus, the effect of the 6th position is overlooked, though it has been already reported to be highly relevant in catalysis.[66]

Hence, considering the possibility of building fibrillar structures from scratch with the in-house developed software FibTool,[205] a computational approach combining QM and MM approximations is performed in the present work. The objectives are i) to assess if differences in the hydrophobic core of the steric zipper fibrils influence the structure of the fibre and, consequently, its catalytic activity; and ii) which interactions are formed between the substrate and the 6th position variants, aiming at deciphering the role that such position may play along the hydrolysis of pNPA. Besides, their catalytic profile will be compared with the SY7 Tyr-containing fibres from Navarro et al.[201], in order to compare their catalytic efficiency.

5.2.2   Computational Details

5.2.2.1   Initial Models

Due to the lack of initial crystallographic structures comprising the heptapeptide sequences under interest, the recently designed FibTool[205] has been used to build up different models. This tool allows to build initial fibre structures just by introducing the desired sequence, specifying i) the number of strands, ii) the separation between strands, iii) the number of layers and iv) the separation in Å among layers.

Once the parameters are set, all the possible orientations of the desired backbone sequence are retrieved, being either parallel or antiparallel inside each sheet or between sheets, which leads to 11 different possible incomes. Each of the possible orientations obtained should ideally be submitted to MD simulations and posterior analysis on stability. Nonetheless, based on already reported works where self-assembly was already experimentally assessed,[211] the fibre with parallel intrasheet (P), with the same residues packed at the interfase (face-to-face) and antiparallel intersheet (up-to-up) orientation (P-FF-UU orientation in Figure 5.2), with His residues exposed to the solvent and Iso forming the zipper, was the only one considered for the present study.

The reference system is a fibre composed of IHIHIHI residues (IH7), for which we generated three different fibre models, with different interstrand distances (6 and 7 Å) and different interlayer separation (9, 11 and 14 Å), to see whether different initial structures affect the morphology of the fibre. The three possibilities were submitted to a 300ns MD simulation and energies were computed with the Generalized Born implicit solvent model, obtaining almost a perfect overlap between fibres. Accordingly, the customised distances between the strands and the sheets are not a determinant parameter for obtaining differential structures, as the same

energetic structures are obtained once the conformational exploration is stable, regardless of the initial point. Once the orientation and distances were determined, hydrophobic cores formed either Iso (IH7), Leu (LH7), Val (VH7) or Pro (PH7), were generated to study their impact on the fibrillar twist; while 6th position variants with Gln (IHQ7), Arg (IHR7) and Tyr (IHY7), were generated in IH7 fibres to unravel their effect on the catalysis.

For SY7, however, no experimental information was available. Previous works from the group already explored the 11 architectures and concluded that the most stable one exhibits a parallel β-sheet arrangement with Tyrosines exposed to the solvent.[205,206] This is because parallel arrangements allow the formation of stronger hydrogen bond contacts at the zipper, due to a better interdigitation of side chains. Indeed, Serine residues at the zipper establish H-bond contacts either among them or with the backbone. Besides, since the tetralayer supramolecular systems show the same preferred assembly as the bilayer,[206] the latter models were preferred for this study to reduce the computational cost. Hence, all models comprise a double-layer system, each layer containing 20 β-strands, with a total of 40 strands per model (Figure 5.1).

5.2.2.2 Molecular Dynamics simulations

After initialising the models, classical Molecular Dynamics simulations were carried out, originally to define the effect of the 6th position on the structures' twist angle. Once a stable structure for each sequence was obtained, Zn(II) and pNPA dockings were performed. The obtained coordinated complexes were again submitted to MD simulations, in this second round to analyse the most recurrent position of the substrate with respect to the environment.

Zn(II) metal parameters were obtained using the MCPB.py Amber tool,[91] from quantum mechanical calculations with DFT(B3LYP) and adding Grimme's D3 correction for dispersion.[212] Optimization of the coordination structure was done with the 6-31+G(d,p) basis for C, H, N and O atoms, and the SDD basis set for Zn(II). Solvent effects were accounted for with the solvent-polarizable dielectric continuum model (SMD) in water. Force constants and equilibrium parameters for those atoms coordinating the metal ion were obtained using the Seminario method,[90] while point charges were derived using the RESP (restrained electrostatic potential) model. The complex was solvated within a cubic box of pre-equilibrated TIP3P water molecules and Na^+ ions to balance the total charge according to the needs of each system.

Molecular Dynamics simulations were done for a minimum of 300ns, first to analyse the twist and then for the distances of pNPA to the 6th position, with the AMBER ff14SB force field[146] in the NPT ensemble, using an integration time step of 1fs. Constant temperature and pressure were set by coupling the system to a Monte Carlo barostat at 1.01325 bar and a Langevin thermostat at 300 K. The SHAKE algorithm was used to constrain bonds involving hydrogen atoms. In this case, GaMD approximations could not be applied as the potential boost resulted in the dismantling of the fibre.

5.2.2.3 Zn(II) Docking on the Fibre

Docking of the Zn(II) ion was performed on the most populated cluster of the fibrils to test if the reported coordination could be encountered.[213] Once Zn(II) coordination was established, docking of the pNPA substrate over Zn(II) was done.

Such docking calculations were carried out with the GOLD program[57] using the GOLDScore scoring[144] function, with modified parameters to

include possible metal ions coordinations.[59] Genetic algorithm (GA) parameters were set to 50 GA runs and a minimum of 100,000 operations each. The remaining parameters, including pressure, number of islands, crossovers, or mutations, were set to default. All histidines were deprotonated for the calculation, in order to allow the system to coordinate the Zn(II) ion with the most suitable nitrogen from His. Docking solutions were then visualized with GaudiView,[145] a built-in in UCSF Chimera.

5.2.2.4   QM/MM calculations

The reaction mechanism was analysed by means of QM/MM calculations. Such simulations were set up using the MolUp plug-in for VMD.[214] The QM region comprised the metal, the ligand and the three coordinated residues until their C_β (HIS for IH7 and IHQ7 and Tyr for SY7). The rest of the system (MM region) consisted of over 5000 atoms. The energy of the QM region was calculated with the B3LYP functional[215] plus the Grimme's correction for dispersion[212] and using 6-31+G(d,p) basis set[216] for main elements, and the Stuttgart ECP[217] pseudopotential and associated basis set for Zn(II). The MM part of the system was treated with AMBER FF19SB,[60] in gas phase. For solvent correction of the reaction energies, a single-point calculation with implicit solvent was performed once the optimised structures were obtained with the CPCM implicit solvent model. The nature of all stationary points was verified by vibrational analysis, ensuring that all frequencies are positive in the minima and that only one frequency is negative in the transition structures.

5.2.2.5   Analysis

Principal Component Analysis (PCA),[149] RMSD all-to-all and clustering were performed over the MD simulations to test if the exploration was exhaustive enough. If required, the energy state of the systems along the simulation was computed using the Generalized Born Implicit model, for

800 frames extracted along each 300ns MD trajectory, stripping water molecules, and performing a short minimization (maxcyc = 500 steps) for each point previously to the energy extraction. The twist angle of each of the fibrillar variants was analysed by computing the vectors of residues 1 to 7 of the first strand and the same residues for the tenth strand; both in the upper and bottom layers. The twist angle of the fibre is thus referred to as the angle formed by both vectors along the last 100 frames of the simulation, averaging the angle of the upper and bottom layers to obtain a single value. Finally, the distances of both NO_2 and CO atoms of the pNPA ligand with N atom of Arg, His and Gln sidechain; O atom of Tyr sidechain and the adjacent Zn(II) were obtained with UCSF Chimera.[142]

5.2.3   Results

5.2.3.1   Classical amyloid fibres

As reported in the introduction, the work of Rufo et al.[66] has shown that minimalist amyloid fibres formed by seven residue amphiphilic peptides are capable of exhibiting esterase activity and hydrolyse pNPA in the presence of Zn(II) ion. However, their results also indicate that the catalysis depends both on the nature of the core hydrophobic residues of the peptide (Iso, Val, Leu, Ala) that conform to the hydrophobic zipper of the fibre; and on the residue present in the sixth position of the fibre.

It has been hypothesised that the reason why changes in the hydrophobic residues modify the catalytic activity is due to changes in the fibrillar twist angle, induced by the different steric effects in the zipper, which could modify the local environment of the active site exposed to the solvent. Indeed, depending on the twist angle, the pNPA substrate might be able to double coordinate to two different Zn(II) ions if suitable distances are obtained. Furthermore, such a twist could induce the physical approximation of the 6th residue to pNPA, modifying the catalytic mech-

anism. For that reason, four fibres with different hydrophobic residues, composed either by Leu (LHLHLQL – LHQ7), Iso (IHIHIHI – IH7 and IHIHIQI – IHQ7), Val (VHVHVQV – VHQ7) and Pro (PHPHPQP - PHQ7) were built, and metal docking calculations performed to set up the metal-fibre system.

Docking calculations were carried out on the initial structures built by FibTool, assuming that all His were double deprotonated to determine which N (ε or δ) would preferably bind to Zn(II) ion. Even though a full-coverage model would be desirable, double deprotonated histidines cannot be parametrized with MCPB.py, since it would require considering the whole fibre for calculating charge distribution, instead of a limited coordination sphere. For that reason, the half-coverage model was assumed, being it a suitable model for the study of catalysis. For such a half-coverage model, docking calculations found a compatible coordination sphere, consisting of three coordinating histidines (2 from strand i, one coordinating through $N\delta$ and another with $N\varepsilon$; and 1 from strand i+1, through $N\delta$), where a water molecule or a hydroxide would complete the tetrahedral coordination environment (Figure 5.4).

Once the metal-fibre systems were built, they were sent for 300ns MD simulations to see whether they could induce different twist angles and, hence, different conformations. For comparison, the twist angle of the metal-free and metal-coordinated forms was computed on the last 100 frames of the simulation, once the fibre is already stable. Results are given in Table 5.3, which demonstrates a constant flattening effect of the Zn(II) ion coordination. In fact, the twist angle ranges from 95° to 118° for the metal-free fibres, while it reduces to 97-40° for the metal-bound forms. The most remarkable flattening effect is observed in the VHQ7 fibre, whose twist angle is reduced by more than 80° (Figure 5.5), which could be due to lower steric repulsion in the zipper by Val. Since Val
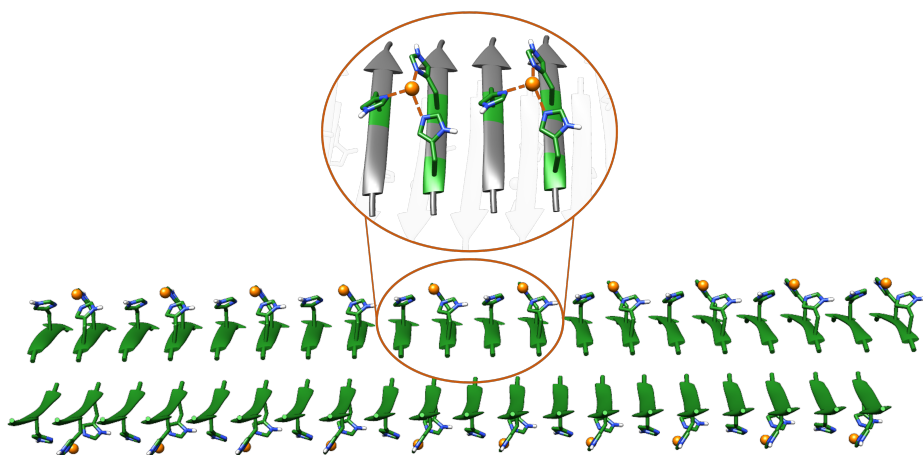
Figure 5.4: Half coverage model of Zn(II) coordination with IH7, with three His forming the coordination sphere (2 with Nδ and 1 with Nε).

presents a short side-chain, with an interlayer distance reduced with respect to other fibres (9.4 Å vs 11.1 Å for IH7), the flattening effect of Zn(II) caused by its coordination with His is predominant to the steric clashes that induce the twist for the other fibres. Moreover, the modification of the sixth position increases the average twist angle by about 10º (95.3º for IH7 vs. 107.2º for the IHQ7). In the IHQ7 variant, though, the flattening effect of Zn(II) coordination is slightly more evident (-10º) than in the IH7 variant, which leads to similar angles for both IH7 and IHQ7. Besides, Pro residue is incapable of maintaining the fibrillar structure due to its low hydrophobic character, leading to the disruption of the fibre.

With these results, it is concluded that the twist angle induced by different hydrophobic cores is not directly linked to the catalytic activity, as no correlation is observed between both parameters. Since the twist angle does not seem to be an indicator of the catalytic activity, further aspects, such as the self-assembly capacity of fibres with different hydrophobic cores, which is out of the scope of this work, should be considered.

Table 5.3: Twist angles observed for the different hydrophobic cores.

Fiber	Metal-free twist (º)	Zn-bound twist (º)	Catalytic Activity $(M^{-1}s^{-1})$
IH7	95.3	93.6	NA
IHQ7	107.2	96.9	62 $\pm$2
LHQ7	118	94.7	30 $\pm$3
VHQ7	113.5	40	32 $\pm$2
PHQ7	Disrupts	Disrupts	NA
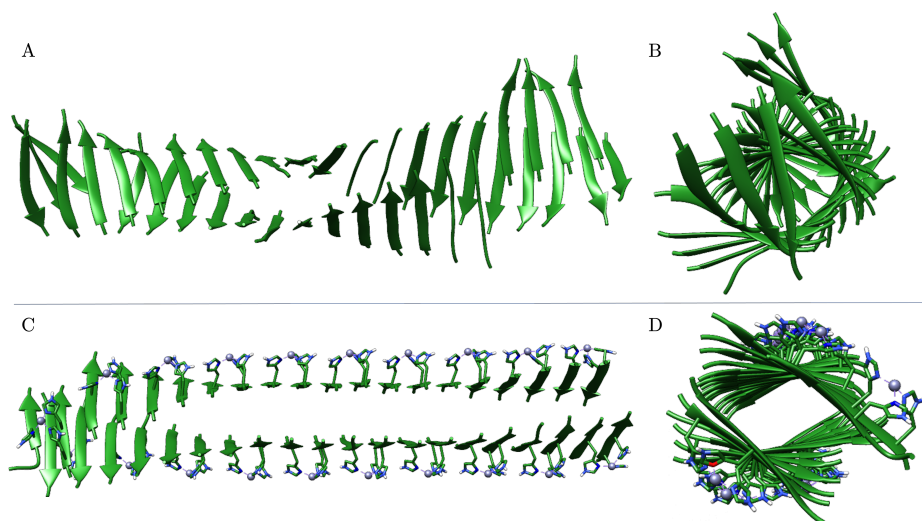
Figure 5.5: VHQ7 Fiber, (A, B) without and (C,D) with Zn(II) coordination. Lateral and front views. The flattening effect of Zn(II) coordination is visually observable (-80º).

The most surprising, still unknown fact is how the 6th position of the fibre affects its catalytic activity, since it is not directly involved in the binding of Zn(II). Indeed, the 6th position is modified by His to Tyr, Arg and Gln in the previous work of Rufo et al.[66], retrieving catalytic activities that follow the trend Tyr $\approx$ Arg < Gln for Leu-core fibres (see Table 5.1).

Regarding the enhancement of the catalytic activity induced by the 6th position, three working hypotheses based on the direct interaction between the sixth position and the pNPA have been applied. Such interaction was expected to take place either through i) the -COO- ester group of the pNPA with the residue in the 6th position, ii) the NO_2 group of the pNPA with the residue in the 6th position or iii) the NO_2 group of the pNPA with an adjacent Zn(II) ion. In the first two hypotheses, the proton donor group of the 6th residue (Gln, Tyr, Arg) side chain may establish H-bond contacts with the NO_2 or -COO- groups of pNPA, which could further activate the substrate and favour the OH nucleophilic attack. On the latter, a double interaction with Zn(II) ions could also affect the catalytic activity, since NO_2-Zn(II) interactions may further activate pNPA for a nucleophilic attack (Figure 5.6).
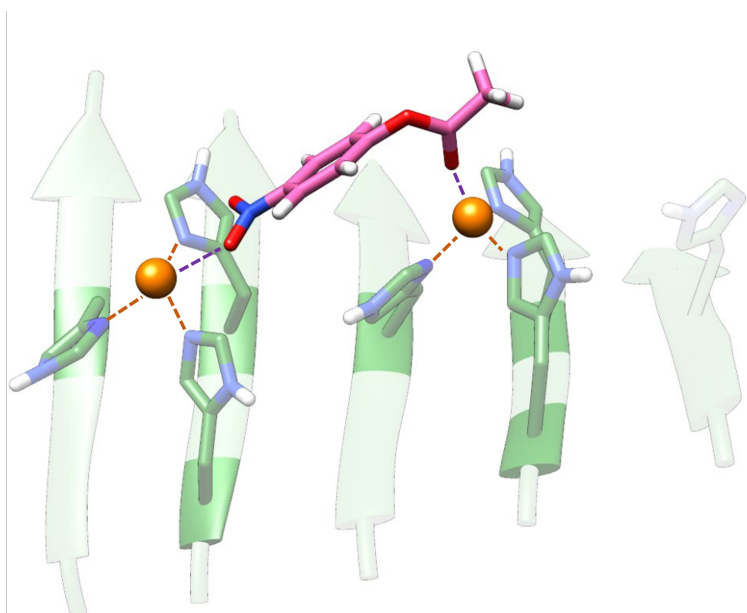
Figure 5.6: Docking of pNPA (pink) on two random Zn(II) of the IHQ7 fiber.

First, MD simulations for fibres whose 6th position mutated to Gln (IHQ7), Arg (IHR7) and Tyr (IHY7), bound to Zn(II), were performed. In the present work, the hydrophobic core is formed by Iso (not Leu), since the work of Rufo et al.[66] reports improved catalysis for Iso-core fibres, while the observed trends upon changing the sixth position are expected to be maintained.

Since the focus of attention is placed on the transitory interactions of pNPA with the surrounding environment and, more concretely, with the 6th residue, the pNPA substrate was docked with GOLD software upon a random Zn(II) ion of the fibre. However, in preliminary QM studies with only the active site, the transitory species oxyanion, derived from the nucleophilic attack of OH to NPA, was identified as an intermediate in the hydrolysis. Hence, in order to analyse whether the 6th position or the adjacent Zn(II) may interact with pNPA, such intermediate was overlapped with the docked position of the pNPA and parametrized in a bonded manner, so MD simulations could be performed. With this strategy, it is possible to simulate the possible interactions that the transitory state may form with the environment.

In order to properly assess the interactions between pNPA and the 6th position, either from the same strand as the coordinating ion or the adjacent ones, the distances to the N atom from the NO_2 group or the oxygen atoms of the -COO- group were measured with respect to the lateral chain of the 6th position residues (Figure 5.7). Distances for the 6th position were obtained considering the heavy atom of the proton donor group, say, O for Tyr and N for Arg and Gln. The threshold has been established at 5 Å for the 6th position, since QM calculations prove a N−N distance between Gln and NO_2 of about 4.4 Å. However, the distances between Zn(II) atoms and NO_2 have to be below 4 Å in order to influence the catalysis.
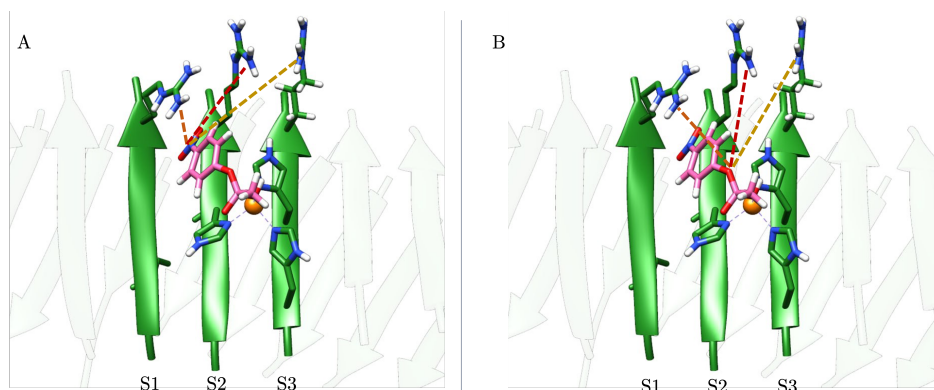
Figure 5.7: Distances from NO_2 (A) or C (B) of the NPA to the lateral chain of the 6th residue from three consecutive strands (S1, S2, S3).

Analyses of MD simulations indicate that no suitable distances between -COO- with the lateral chains of the sixth position were observed, discarding our second hypothesis. On the contrary, the NO_2 group can reach feasible distances in all the alternative 6th positions (see Table 5.4).

Table 5.4: Distances between NO_2 group of the pNPA towards either the donor atom of the 6th residue or adjacent Zn(II) ions for each fibrillar variant.

Distance	IH7	IHQ7	IHR7	IHY7
NO_2-S1 (<5Å)	3.6%	14.5%	31.6%	1.7%
NO_2-S2 (<5Å)	13.6%	25.0%	3.9%	0.5%
NO_2-S3 (<5Å)	1.5%	10.9%	0.2%	2.1%
NO_2-ZN1 (<4Å)	0.3%	0%	6.7%	0.2%
NO_2-ZN2 (<4Å)	0%	0%	0%	0.2%

For the IH7 sequence, the NO_2 of pNPA stays under 5 Å distance from the His lateral chain for 18.7% of the simulation. On the contrary, for IHY7 -the less catalytic fibre-, such interaction with the 6th position drops to just 4.3% along the simulation. The most catalytic fibres, IHR7 and IHQ7, substantially increase such percentage, allowing the substrate to be close to the Arg and Gln 6th residues for more than 35.7% and 50.4% respectively, being the latter the highest result. A direct correlation is observed between IHY7, the least active catalytic fibre; and IHQ7, the most active, having the lowest and the biggest percentage of interaction with the lateral chain of the 6th residue (Table 5.4). Besides, the double coordination of pNPA with two Zn(II) ions, which could be another possibility for improved catalytic activity, is again discarded, since all the alternatives retrieve similar values when considering the interaction of pNPA with adjacent Zn(II) ions (0% for IH7, IHQ7, and IHY7, 7% for IHR7) (Table 5.4).

The hypothesis of the interaction of the NO_2 from the pNPA substrate with the 6th position has been then demonstrated feasible. Consequently, QM and QM/MM calculations were performed to obtain the energy profile of the reaction either for IH7 and IHQ7, in order to describe the impact of the 6th residue on the catalysis. The starting geometry was, in both cases, structures obtained from MD simulations, in which the pNPA was closest to the 6th residue. It should be noted that several possible orientations of pNPA, particularly of the -$OCOCH_3$ ester group with respect to the metal ion, are possible. MD simulations have allowed us to select among them the most representative orientation, in which the CO is pointing toward the Zn(II). The active sites reported at these Zn(II)-fibres mimic those in Zn(II)-metalloenzymes of carbonic anhydrase, which converts CO_2 into bicarbonate and has been extensively studied.[218,219] In such metalloenzymes, the Zn(II) ion is coordinated to three His and one OH^- in a tetrahedral coordination.

The energy profiles obtained for the pNPA hydrolysis either by IH7 or IHQ7 are shown in Figure 5.8. Besides, optimized structures for the reactant, oxanion intermediate, product and two transition structures are given. In our results for the hydrolysis of the pNPA, the first step of the reaction mechanism is the nucleophilic attack of the coordinated OH^- to the C atom of the ester group of pNPA. As a result, a tetrahedric coordinated oxanion is obtained through a pentacoordinated transition state (Figure 5.8, TS1), overcoming an energy barrier of 13.0 (17.3) kcal·mol^{-1} in terms of potential and Gibbs energy, respectively. In this process, the C(Ph)−O bond distance of the NPA moiety significantly increases from 1.37 to 1.46 Å. From this oxanion intermediate, the proton of OH is transferred to the O of the nascent nitrophenolate, through an energy barrier of 8.0 (6.1) kcal·mol^{-1} (Figure 5.8, TS2). Our computed ΔG barrier is smaller than that reported by Song et al.[210] (22 kcal · mol^{-1}). Then, the final products nitrophenol and acetate are obtained, the latter coordinated to Zn(II) (Figure 5.8, P) in an exergonic reaction (−19.5 kcal · mol^{-1}). Interestingly, this mechanism is slightly different to that reported by Song et al.[210] in which the oxanion evolves towards acetic acid and nitrophenolate which, in a subsequent step, converts to nitrophenol and acetate through an additional proton transfer.

For IHQ7, a similar energy profile is obtained (Figure 5.8, blue profile), though with substantially lower energy barriers. In this second case, the energy barrier of the first transition state is 2.7 kcal · mol^{-1} lower, which would explain the enhanced catalytic activity of IHQ7. As for IH7, the transition structure corresponds to the nucleophilic attack of OH (TS1), leading to a pentacoordinate Zn(II) ion with the carbonyl O of the pNPA at 2.38 Å and the O from the OH at 2.03 Å (Figure 5.8). This pentacoordination is lost in the oxanion, as the O from the OH is displaced (2.94 Å) by the O from the pNPA (1.94 Å), obtaining again a tetrahedral coordination. In this case, the barrier of the second step is 5.6 kcal·mol^{-1}, in
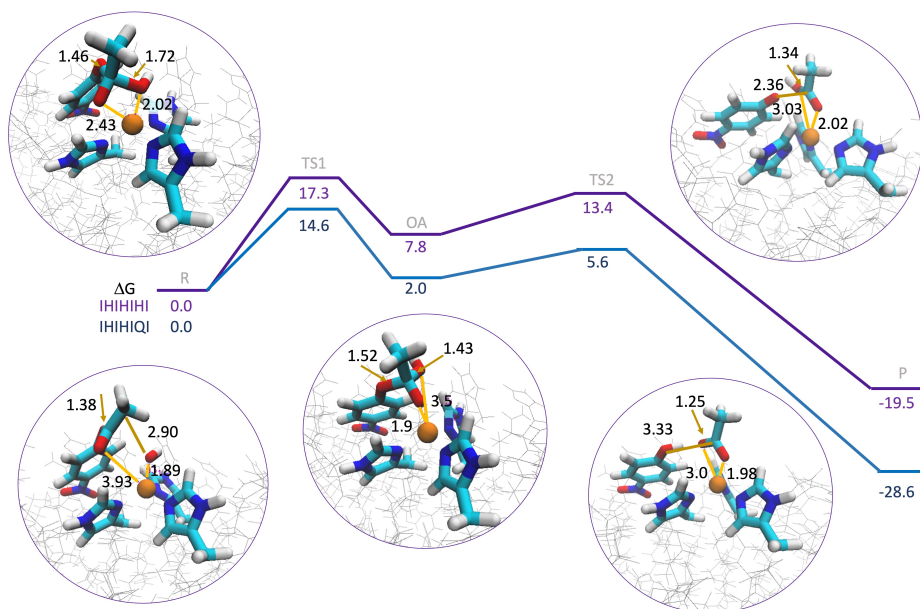
Figure 5.8: Free energy profiles for IH7 (purple) and IHQ7 (blue). Optimized structures for the reactants (R), transition states (TS1 and TS2), oxyanion (OA) and products (P) are shown. Energies are in kcal·mol^{-1} and distances, in Å.

which the phenol and the acetate are identified. This step leads to the final products through a proton transfer in an exergonic reaction (-28.6 kcal·mol^{-1}), even more than for IH7. This is due to the fact that the interaction with the 6th residue allows the formation of a hydrogen bond network (Figure 5.9), reducing the energy barrier and stabilising the products more than the reactants.

As observed, the IHQ7 system demonstrates a smaller energy barrier, with 14.6 kcal·mol^{-1} versus the 17.3 kcal·mol^{-1} observed in the IH7 system. Such difference in the energetic barrier would explain the different catalytic activity observed for such systems, which could be attributed to the polarizing effect of the Gln on the pNPA as well as on its ability to form a stabilizing H-bond network. Hence, it can be observed that the

161

modification of the 6th position helps in the success of the reaction, even when such residue is not considered in the QM part of the reaction.
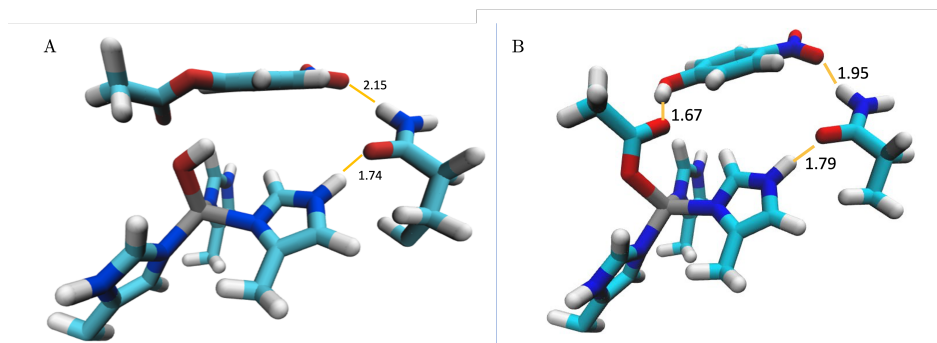
Figure 5.9: H-bonds formed among the side-chain of Gln in the 6th position and the pNPA and coordinating His of IHQ7, in reactants (A) and products (B) optimized structures. Distances are in Å.

5.2.3.2 Prion-like amyloid fibres

As mentioned in the introduction of this section, prion-like fibres, which are constituted by polar residues, have been recently demonstrated to also self-assemble and perform catalytic functions.[195] Besides, in the study of Díaz-Caballero et al.[195], His residues are demonstrated not to be the only ones capable of coordinating metal ions, since heptapeptide amyloid fibres built with three alternate Tyr residues also retain the ability of binding divalent metal cations, such as Co(II), Cu(II) and Zn(II).[202]. Among all their tested alternatives, the most stable heptapeptide fibre was already identified in the work of Peccati et al.[205], which exhibits a parallel intrasheet organization and a face-to-face architecture, with serines in the polar zipper while tyrosines are exposed to the solvent.

In order to analyse the catalytic activity of this prion-like fibres and compare it with the one observed in classical amyloid fibres, docking calculations were performed on the SY7 fibril models previously built. Ac-

cording to the coordination properties of Zn(II) ions, docking simulations were carried out assuming a tetrahedral environment. Results showed that the maximum number of Tyr residues that can simultaneously coordinate is three, coming from two different strands, with the vacant site being filled with a water molecule. Quantum chemical calculations at the B3LYP-D3 level of theory confirmed that the binding mode is stable, with bonding distances being about 2.1 Å, similar to the values $2.2 - 2.3$ Å identified in metalloproteins (Figure 5.10).[201]
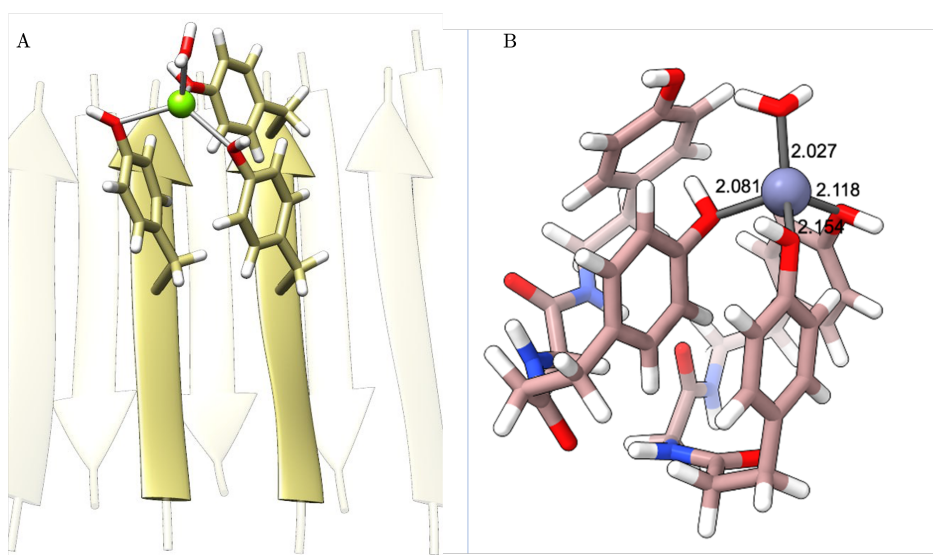
Figure 5.10: Tetrahedral coordination of Tyrosinases of the SY7 fibre with Zn(II) ion. A) Fibrillar view and B) all-atom model with distances in Å.

Nonetheless, the catalytic activity reported for such complex is remarkably less active than that for classical amyloid fibres.[201] For that reason, the same QM/MM protocol as for the classical amyloid fibres was applied, aiming at explaining the difference in the catalytic mechanism mediated by Tyr coordination of Zn(II). In this case, the reaction mechanism is similar to the one reported for classical amyloid fibres, since the first step

of the reaction mechanism is again the nucleophilic attack of the coordinated OH⁻ to the C atom of the ester group of pNPA. However, in this second fibre, the oxyanion is not obtained but directly leads to a nitrophenolate and acetic acid. The final products nitrophenol and acetic acid are obtained through a proton transfer, both still coordinated to the Zn(II).
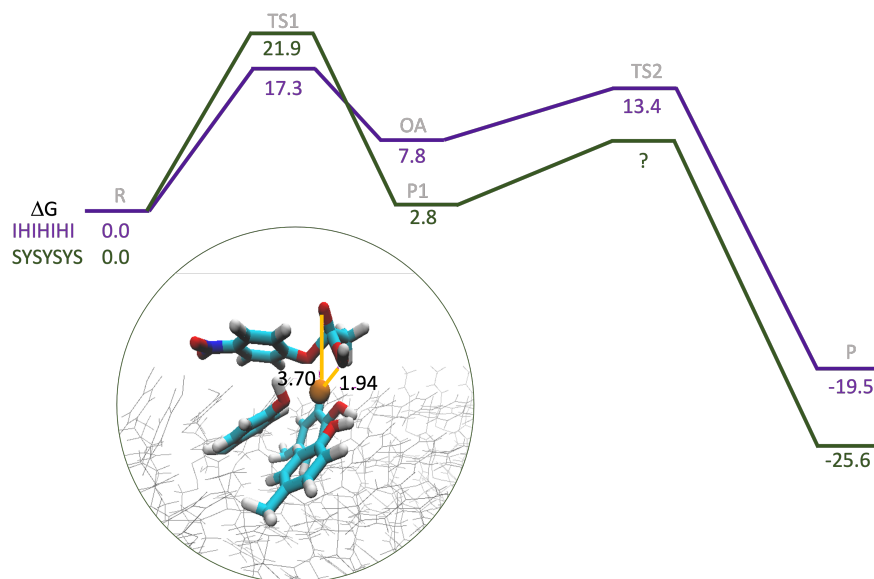
Figure 5.11: Gibbs energy profile for SY7 (green) compared with IH7 one (purple) and transition structure (TS1) for SY7. Energies are in kcal·mol⁻¹ and distances, in Å.

In the SY7 energy profile (Figure 5.11), the energetic barrier of the first TS scales up to 21.9 kcal·mol⁻, demonstrating that the Zn(II) ion coordinated to Tyr is less effective in performing the hydrolysis of the pNPA, in accordance with experimental results. This relates to the fact that the Zn-OH bond seems to be stronger in SY7 than in IH7. Note that the Zn-OH distance for SY7 is 1.84 Å, shorter than that for the IH7 system (2.0 Å). This probably results from the fact that His is a better electron donor than Tyr. As a result, the OH nucleophilic attack in SY7 is hampered compared

to IH7, which leads to a larger energy barrier and a smaller catalytic activity in the case of SY7. Furthermore, the ability to generate an OH^- in the active site of IH7 is more favourable than in SY7, due to the presence of a fourth His capable of accepting the proton.

5.2.4 Conclusions

In the present study, a computational study has been performed in order to determine the catalytic mechanism of pNPA hydrolysis by classical and prion-like amyloid fibres. Moreover, considering the different catalytic activity reported for classical amyloid fibres with different residues in the 6th position, alternative sequences have also been built up to unravel the effect of such position in the catalysis.

For classical amyloid fibres, with His exposed at the surface of the fibre, it has been demonstrated that the effect of the 6th position is due to its direct influence in the catalysis process, rather than the modification of the structural conformation of the fibre. By measuring the distance of the 6th residues with the pNPA, it is demonstrated that transitory interactions take place between the donor group of the 6th residue and the NO_2 group of the pNPA substrate. Besides, with QM/MM calculations for IH7 and IHQ7 fibres, it is supported that the presence of Gln residue near the active site reduces the energy barrier by 2.7 kcal·mol^{-1} due to the fact that it establishes H-bond contacts with the NO_2 group of the pNPA, thus favouring the catalytic process.

For prion-like amyloid fibres, the SY7 structure has been modelled and docking calculations have been performed assuming the tetrahedral coordination geometry of Zn(II) ion with the tyrosinases. The catalytic mechanism and energy profiles are determined, demonstrating that such fibre is also capable of performing the hydrolysis of pNPA though with less

efficiency, since the bond between the Zn(II) ion and the OH seems to be stronger when tyrosines are coordinated, due to their low donation capacity. On the contrary, the bond observed between Zn(II) and OH is longer in IH7, due to histidines being better donors. Then, since the energy required for breaking the Zn-OH bond is higher in SY7 fibres, the energy barrier increases, resulting in less catalytically active fibres.

Overall, the combination of MD, Dockings and QM/MM calculations has retrieved successful results when dealing with catalytic fibres, establishing an interesting pipeline for further study of catalytic systems in which the catalytic process is still to be discovered.

5.3 Cyclopropanation

5.3.1 Overview

There are several reactions of high importance for the scientific community. In particular, the enantioselective cyclopropanation of styrene is crucial for drug discovery, a reaction typically dependent on transition-metal facilitation. Such a reaction may give either trans or cis cyclopropane products, depending on the relative orientation of cyclopropane substituents (Scheme 5.2).[18]

The cyclopropanation process of styrene hemin-mediated is computationally well-detailed in the study of Wei et al.[220]. As seen in Scheme 5.3, the ethyl diazoacetate (EDA) substrate transfers the carbene to the hemin group after displacing the Cl^- ion, which leads to a complex in which the carbene is coordinating the Fe(II) atom. Then, the styrene molecule would interact with the carbene, leading to the cyclopropane product through a transition structure in which the C1 of the carbene interacts with C2 and C3 of styrene in the concerted mechanism, or leads
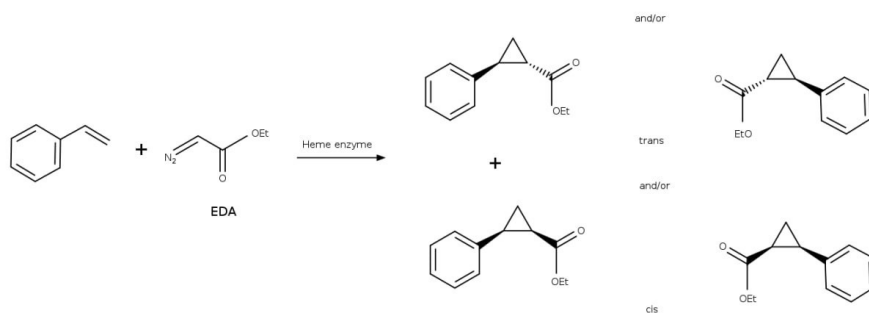
Scheme 5.2: Styrene cyclopropanation reaction. Trans and cis products can be obtained.

to a radical intermediate in the stepwise mechanism. Either in the concerted or stepwise pathways, the metal promotes cyclopropanation, resulting in products whose enantioselectivity is dependent on styrene π-stacking with porphyrin.
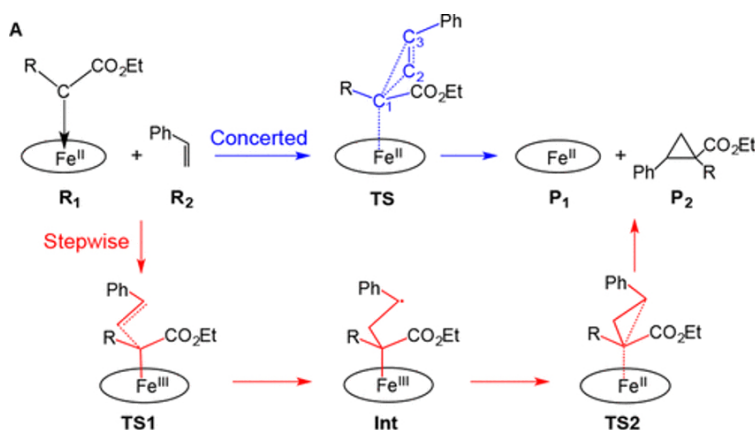
Scheme 5.3: Reaction mechanism for the cyclopropanation process of styrene hemin-mediated, image from Wei et al.[220].

In the previous section of this Chapter of Results, it has been demonstrated the catalytic potential of amyloid fibres in the presence of Zn(II)

metal ions, particularly in the hydrolysis of pNPA, but metal-containing cofactors also present further capacities when performing catalytic processes. Indeed, Zozulia et al.[209] have recently demonstrated the ability of amyloid fibres to bind the hemin group and promote the enantioselective cyclopropanation of aromatic styrenes with EDA, with efficiencies that are comparable to those of engineered hemin proteins. While the most hydrolytic active fibres seen in the previous section (IHIHIHI) are incapable of coordinating hemin, as neither are those formed by Valine core peptides; alternative Leu core-formed fibres can effectively bind hemin. Such coordination is even improved when Phe residues are included in the hydrophilic side, introducing the possibility of forming stacking interactions with the porphyrin subunit. In fact, the resulting fibre, LHLHLFL, effectively binds hemin and promotes enantiospecific cyclopropanation.

Moreover, as a single His is necessary for hemin coordination, being His4 the indispensable one, the rest of the sequence can be mutated seeking enhanced catalysis. Especially relevant is position 2, which has been demonstrated as essential in modifying catalytic activity and enantioselectivity. For that reason, His2 is mutated to different residues, coming up with LILHLFL as the fibre with the highest enantioselectivity; while LALHLFL shows an intermedia enantioselectivity and LTLHLFL, the lowest one (see Table 5.5). Besides, the conformation of the fibre is also a determinant for enantioselectivity, which was demonstrated by the fact that building the (D) conformation of the fibre resulted in the opposite enantiomer to the (S) conformation. Such a fact also points out that the selectivity of the catalyst is linked to the chirality of the monomeric peptides.

Considering that amyloid fibres have been demonstrated to be suitable platforms for hemin binding and capable of performing enantioselective cyclopropanation, a major question arises about how an essentially flat

surface results in any enantioselectivity at all. Besides, slight changes in the sequence retrieve strong differences in catalytic enantioselectivity. Indeed, it is a sound possibility that the way hemin assemblies may be determined by the fibrillar structure, particularly by its twist angle, or the surrounding environment, causing subtle effects on the reaction.

Here, a computational study on the heptapeptide fibrillar structure and hemin binding is presented, aiming at unravelling the role that the second position of the fibre may present in determining both hemin binding mode and fibre folding; consequently, influencing enantioselectivity.

5.3.2   Computational Details

5.3.2.1   Initial Models

As in the previous section, FibTool[205,206] was applied to build up the initial fibrillar structures with the desired sequences. Using the parameters consistent with the earlier study, we built double and tetralayer fibre structures, with 20 β-strands per layer, for LILHLFL, resulting in a total of 40 and 80 strands, respectively. P-FF-UU with hydrophobic residues at the zipper was chosen as the preferred orientation in all instances. In this scenario, we constructed Ala and Thr alternatives at the 2nd position of the Leu-core fiber, to assess the enantioselective capacity of the fiber.

Since the required number of layers of the fibrillar structures is unknown, as no crystallographic studies are obtained, bilayer structures were built for the three variants and submitted to 300ns MD simulations, measuring its twist angle.

However, all three variants opt for a twisted conformation when constructed in the bilayer form (Figure 5.12). One of the hypotheses at this point was that hemin binding -and cyclopropanation reaction- could hap-

pen at the interface between both layers if the twist angle created a sufficiently big pocket inside the fibre. To test such a possibility, docking simulations were performed on the most populated cluster of each of the fibres, but results demonstrate that the interface does not provide enough space to hold the catalyst. Increasing the interlayer distance, which would seem a possibility to fit the complex, is not a feasible option as the steric zipper would not be formed and, consequently, the fibrillar structure would be dismantled. Hence, the bilayer model was discarded as a suitable model for hemin-binding, opting for tetralayer fibre model systems, with an interstrand space of 11 Å.

LILHLFL   LALHLFL   LTLHLFL
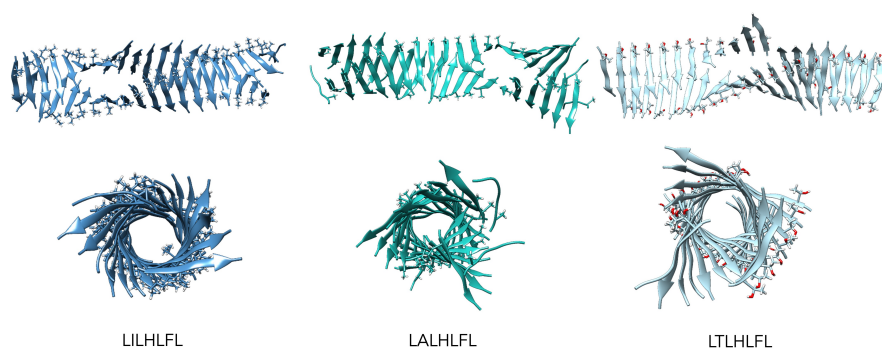
Figure 5.12: LI (A), LA (B) and LT (C) bilayer fibrillar structures. Twist angles correspond to 147°, 131° and 114°, respectively.

5.3.2.2   Molecular Dynamics Simulations

After initialising the models, Classical Molecular Dynamics simulations were performed, both on the bilayer and tetralayer models, to first obtain the most stable conformation. As bilayer systems seemed not in agreement with the experimental data observed by Zozulia et al.[209], which suggested higher aggregation forms, only tetralayer systems were maintained for the following steps of the study. Hence, once a stable structure for each tetralayer system was obtained, hemin dockings were performed.

The obtained coordinated complexes were again submitted to MD simulations, now to observe the interactions of the environment with the catalyst and the carbene substrate.

Fe(III) parameters were obtained using the MCPB.py Amber tool,[91] from quantum mechanical calculations with DFT(B3LYP) and adding Grimme's D3 correction for dispersion. Optimization of the coordination structure was done with the 6-31+G(d,p) basis for C, H, N and O atoms, and the SDD basis set for Fe. Solvent effects were accounted for with the solvent-polarizable dielectric continuum model (SMD) in water. Force constants and equilibrium parameters for those atoms coordinating the metal ion were obtained using the Seminario method,[90] while point charges were derived using the RESP (restrained electrostatic potential) model. The complex was solvated within a cubic box of pre-equilibrated TIP3P water molecules and Na^+ ions to balance the total charge according to the needs of each system.

All Molecular Dynamics simulations were performed for a minimum of 300ns with the AMBER ff14SB[146] force field in the NPT ensemble, using an integration time step of 1fs. Constant temperature and pressure were set by coupling the system to a Monte Carlo barostat at 1.01325 bar and a Langevin thermostat at 300 K. The SHAKE algorithm was used to constrain bonds involving hydrogen atoms.

5.3.2.3   Hemin Docking on the Fibre

Docking of the hemin complex was performed on tetralayer fibril model systems with GOLD software[57] on the most populated cluster of the fibrils. GOLD presents an intrinsic limitation of rotating a maximum of 10 residues during a Docking search, so considering that tetralayer structures may present not a single but several suitable coordination regions, the 12 central β-strands of the fibre were divided into four regions (A, B, C

and D including three strands each). From each region, the top layer was selected as a search area, over which triplicate runs of docking searches were performed (Figure 5.13). Besides, histidine residues were deprotonated, ensuring that both δ and ε could coordinate with the hemin group. With this approximation, the whole fibre surface was explored, allowing free movement of residues' side chains.
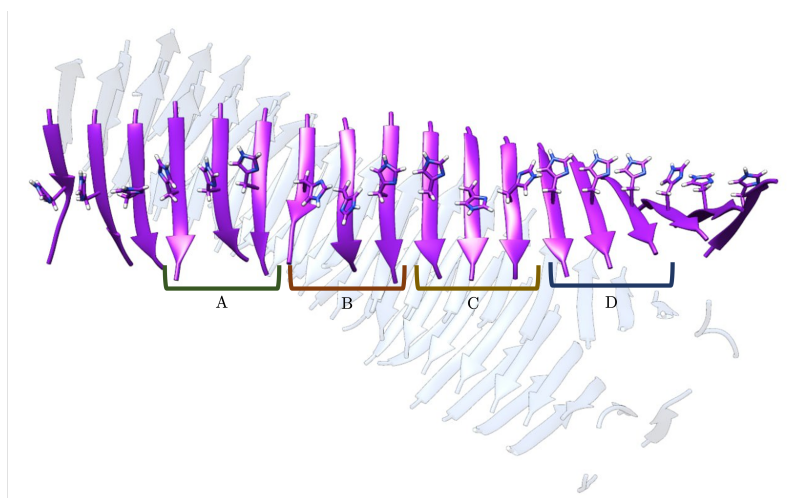
Figure 5.13: Docking areas determined for the hemin binding search. The fibre was divided into four regions, A, B, C and D. The search area was then limited to three consecutive strands of the upper layer.

The GOLD search included modified parameters to account for metal ion binding.[59] Genetic algorithm (GA) parameters were set to 50 GA runs and a minimum of 100,000 operations each. The remaining parameters were set to default, including pressure, number of islands, crossovers, or mutations. Docking solutions were then visualized with GaudiView[145] a built-in in UCSF Chimera, analysing those with the best punctuation and discriminating between Nε and Nδ hemin coordinated complexes.

5.3.2.4 Analysis

Principal Component Analysis (PCA),[149] RMSD all-to-all and clustering were performed over the MD simulations to test if the exploration was exhaustive enough.

The twist angle of each of the variants was analysed by computing the vectors of residues 1 to 7 of the first strand and the same residues for the tenth strand in the top and first intermedia layer for the tetralayer system. The twist angle is thus referred to as the average angle formed by both vectors along the last 100 frames of the simulation.

Finally, the relative poses of hemin+carbene compound with respect to the fibre were assessed by computing the angle formed between i) the plane of the His-coordinating containing strand and three adjacent ones; and ii) the plane of hemin, thus obtaining the angle between the planes. The computation of such angles allowed for numerical comparison among fibre variants.

5.3.3 Results

The target heptapeptide sequences selected for the present study have been based on Rufo et al.[66] work, whose experimental results retrieved remarkable differences in enantioselectivity depending on the sequence (Table 5.5). From their reported results, the most enantioselective sequence is LILHLFL (LI), while LALHLFL (LA) shows an intermedia selectivity and LTLHLFL (LT), the lowest. For comparison purposes, only these three cases are studied in the present work, which should be sufficient for explaining the differential enantioselectivity of the process.

Then, tetralayer models for LI (LILHLFL), LA (LALHLFL) and LT (LTLHLFL) systems were built and sent to MD simulations. Each fibre

Table 5.5: Representative examples of the catalytic activities reported from Rufo et al.[66] for Leu-core amyloid fibres. For remaining fibres, see [66].

Peptide	Yield, %	de, %	ee_{cis}	ee_{trans}
LILHLFL	52	68	-12	40
LALHLFL	28	69	6	7
LTLHLFL	30	72	0	1

Table 5.6: Twist angle obtained for each fibre (LI, LA and LT) either with protonated $N\delta$ (HID) or $N\varepsilon$ (HIE).

Sequence	LILHLFL	LALHLFL	LTLHLFL
Twist HID	146.5º	82.8º	74.3º
Twist HIE	95.0º	48.0º	93.6º

was modelled either with protonated $N\varepsilon$ (HIE) and $N\delta$ (HID) forms of the histidine residues, in order to evaluate potential differences between the two protonation forms.

Table 5.6 and Figure 5.14 show the twist angle for LI, LA and LT when His is protonated at $N\delta$ (HID) or $N\varepsilon$ (HIE). It can be observed that for HID fibres, the twist angle follows the trend (LI > LA > LT); that is, it follows the same enantioselectivity trends experimentally observed (Table 5.5). Thus, the larger the twist, the larger the enantioselectivity. However, for HIE fibres such a tendency is not observed, as the twist angle obtained by the LA system is substantially lower than LT (48º vs 95 and 93.6º obtained for LI and LT, respectively). Besides, HIE systems reduce the angle even more than HID for LI (95º vs 146º) and LA (48º vs 82º) sequences, while the opposite happens for LT (93º vs 74.3º).

The next step consisted of performing docking calculations of Hemin + carbene substrate in the most populated cluster of each fibre to identify

Table 5.7: Maximum GOLDscore for Nε coordinated (HID) and Nδ coordinated (HIE) and percentage of Nε and Nδ coordinations for LI, LA and LT systems.

Sequence	LILHLFL	LALHLFL	LTLHLFL
Max Fitness Nε **coord.**	92.43	51.49	78.43
Max Fitness Nδ **coord.**	78.83	76.26	75.43
% Nε coord	50%	37.1%	92.5%
% Nδ coord	41%	55.1%	4.2%

possible correlations between the twist of the fibres and a particular environment of hemin binding. These calculations were done assuming deprotonated Nδ and Nε of His residues, to avoid constraints when evaluating His coordination. To prevent docking simulations with GOLD from getting stacked in transitory wells, triplicated replicas for each of the four regions in each structure were computed and globally analysed (Table 5.7, Figure 5.15).

Among all the results, the percentage of results considering Nε and Nδ coordinating His has been counted, to detect any preference for a particular protonation state that fibres may show. Strikingly, the LT fibre demonstrates a majority of Nε coordinating the hemin group (92.5 vs 4.2%); while no remarkable differences are observed between LI (50 and 41%) and LA systems (37.1 and 55.1%), pointing to the fact that both orientations seem equally preferred. The best score for each of the alternative protonation states, though, is very similar among the fibres, suggesting that both states are possible.

Observing the maximum fitness structures, it is revealed that LI and LA fibres obtain hemin binding poses oriented towards the Phe6 residue, as hemin Nε coordination (HID systems) retrieves a slightly perpendicular
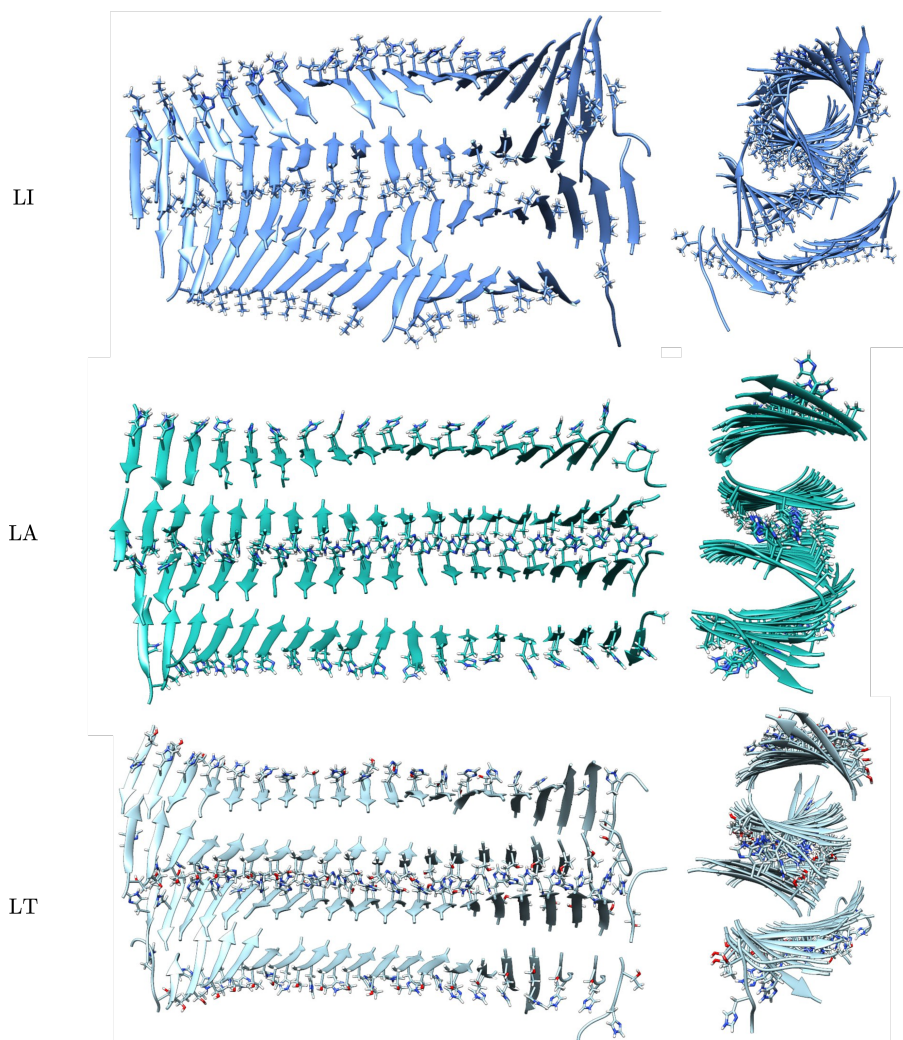
Figure 5.14: Tetralayer forms of LI, LA and LT, front and lateral view for Nε protonated systems (HIE).

coordination to the fibre axis. On the contrary, LT fibres lean towards the Thr in the 2nd position, (Figure 5.15) forming hydrogen bonds with the propionates of the hemin. For hemin Nδ coordination (HIE systems),
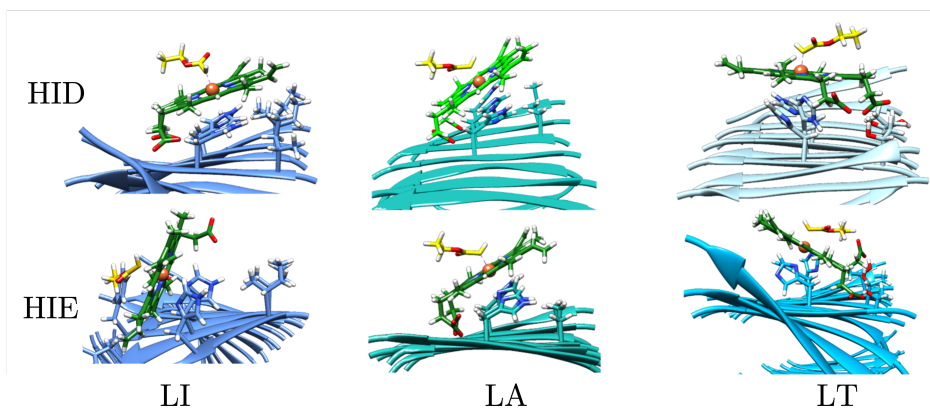
Figure 5.15: Docking of Hemin and carbene substrate in LI, LA and LT sequences, both with His-Nε and His-Nδ coordination.

similar poses are obtained, though LI fibres retrieve a higher angle between hemin and the fibre than for Nε coordination. In order to evaluate the stability of the docked positions, the best-fit poses of the three fibres, both for Nε and Nδ coordinated systems, have been parameterised and submitted to 300ns of MD simulation.

After the simulations, the most populated cluster has been identified considering the positions of just histidine, hemin, carbene, and the 2nd residue of the closest strands, in order to evaluate the most recurrent position of the hemin and substrate with respect to the fibre. The angle formed by the plane of the fibre, composed of the three strands related to the coordinated hemin; and the plane of hemin is computed for comparison purposes (Figure 5.16).

With MD simulations, it was reassured that the initial docked structures are very stable, with insignificant movements along the simulations, though substantially different among them. Both for LI and LA, HIE systems with His-Nδ coordinating to hemin produce leaned confor-
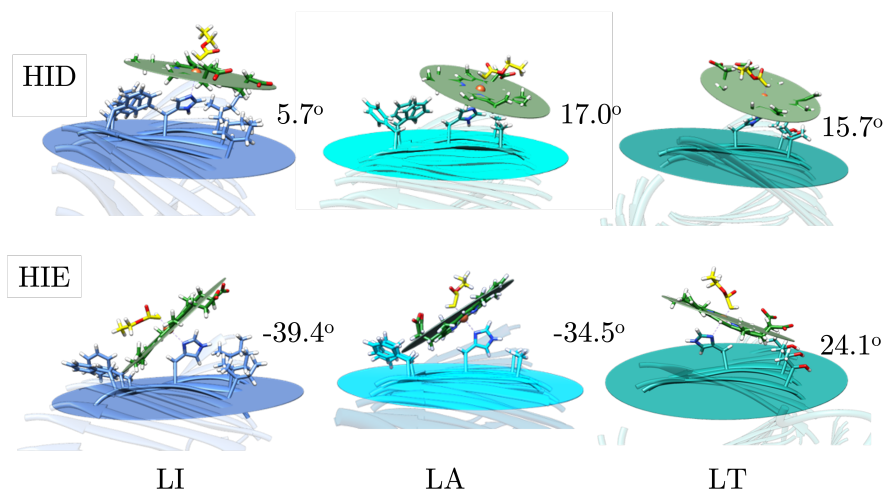
Figure 5.16: Angle formed between the Hemin and fibre's planes, both for Nε (HID) and Nδ (HIE) histidine coordination for LI, LA and LT.

mations towards the Phe2, the inclination being higher for LI (-39.4°), which could be due to the hydrophobic contacts between Ile residue and the heme ring. For LA, the inclination is milder, but also remarkable. Even though the carbene substrate can rotate freely on its Fe(II) coordination, such non-parallel configurations are likely to be responsible for the enantioselectivity, as the styrene substrate could only approach the complex opposite to Phe. On the contrary, Nε coordination in HID histidines for the same systems retrieves almost parallel orientations (5.7 and 17°), just over the 2nd position and totally exposing the carbene to the solvent. Thus, Nε coordination is not expected to provide any stereoselectivity at all. Finally, the LT sequence retrieves interactions with the Thr residue in the 2nd position regardless of the coordinated N of histidine, as hemin's propionates form hydrogen bonds with Thr. In that case, both systems totally expose the carbene to the solvent, which could be related to the lack of stereoselectivity that such a fibrillar system causes.

5.3.4 Conclusions

While amyloid fibres are widely applied as catalytic systems, cyclopropanation has not been typically performed on such kind of self-assembled systems. Rufo et al.[66] published the first article linking both systems, demonstrating the great potential of amyloid fibres at modifying the enantioselective properties of the reaction. However, the mechanism by which fibres could define enantioselectivity is still undetermined. In the present study, its binding has been assessed through a computational study, based on MD simulations and dockings, aiming at discerning the mechanism behind fibrillar structure determining the reaction's stereoselectivity.

Accordingly, extreme catalytic cases, meaning, the most active (LI), an intermedia case (LA) and the least active (LT) fibres have been selected, their models have been built and submitted to docking calculations. Tetralayer models have been designed to identify substantial differences among their binding properties. Simulations on tetralayer systems retrieve different twist angles among them, which result in unique docking poses dependent on the coordinating N atom from His4. Coordination of hemin with His-Nε (HID), regardless of the sequence, exposes both the hemin and the carbene substrate to the solvent, which is expected to allow styrene to approach from any direction, not inducing any stereoselectivity at all. On the contrary, His-Nδ (HIE) coordination in LI and LA sequences forces perpendicular coordination to the fibrillar axis, which blocks styrene approximation from the fibre side and, ultimately, could be the reason for enantioselectivity. Regardless of the nitrogen involved in hemin coordination for sequence LT, hydrogen bonds are formed between Thr in the 2nd position and hemin's propionates prevail. As a result, hemin is parallel to the fibrillar axis, totally accessible to the solvent, which may explain LT's lack of stereoselectivity.

In summary, the fibrillar twist is demonstrated to be linked to stereose-lectivity, as it partially determines hemin binding poses. However, it may not be the single reason, as a combination of histidines' protonation states could also influence the catalytic activity. In fact, His-Nδ coordination, as in the case of LI and LA fibres, tends to situate hemin in a perpendic-ular position, constraining the entrance of the substrate; while His-Nε, preferred by the LT fibre, is more likely to favour its total exposure to the solvent. Thus, further experimental approaches would be desired to confirm fibre helicity and histidine protonation effect on catalytic activ-ity, as blocking certain protonation states could be a suitable approach to generate specific stereoselective products.

6

Metal Diffusion Pathways

During the previous Chapters of the present Thesis, it has been demonstrated that the presence of metal ions in biomolecules is essential. Their role in biological processes is extremely varied, from folding and assembly to molecular interactions, catalysis or recognition. Besides, our understanding of the interaction between metal ions and biomolecules is rapidly increasing as both experimental[221] and computational approaches[55,222] are improving. Thus, it is expected that metal ions study will be a central topic in the next years.

Considering the great importance that upcoming new tools represent, the last Chapter of Results is devoted to illustrating the usefulness of a new computational protocol, aiming at defining the load of metal ions into the active site of preorganized proteins. Until now, extensive studies[223] have assessed the role of metals in folding and stability, but none of them explains metal diffusion into the active site of enzymes at the atomic scale. Our group has shown throughout the first applications of BioMetAll software the possibility of faintly identifying metal ion channels inside proteins, further developing a new computational tool named BioBrigit.

Here, profit was made from such capacity of BioBrigit, proposing a computational protocol to adapt it for metal ion diffusion study. Tyrosinase protein has been selected as a case study, due to its essential role in melanin production and its relation with several diseases, for which a crystal structure is already available with transient copper binding sites.

Metal Difussion in Tyrosinase Systems

6.1   Overview

Nowadays, metalloenzymes are a common target for experimental and theoretical studies due to their involvement in numerous reactions relevant to life. Most of these studies typically concentrate on chemical mechanisms or substrate specificity. However, metalloenzymes require transition metal ions to perform their function, which must be transported to convenient sites. Such a loading mechanism is still poorly understood despite being crucial not only for the proper understanding of the enzyme but also to help in the design of *de novo* enzymes and the treatment of diseases.[224]

Metal-loading in proteins can occur through two mechanisms: i) the metal is retrieved during the folding process to give the protein its final geometry, or ii) the metal ion diffuses from the solvent to the active site in a pre-organized protein structure. The degree of pre-organization may vary from system to system. Moreover, the metal can be directly captured by the protein scaffold or require the assistance of a chaperone, which catches the metal from the solvent and leads it to the target.[225]

Despite intensive works to understand metal recruitment, either involving folding processes[223] or pre-organized systems, have been performed, no complete recruiting paths have ever been reported for metals in protein, experimentally or theoretically. In fact, to the best of our knowledge, only a handful set of X-ray structures display metal ions located in intermediate sites, and only one reports an almost complete diffusion route. This structure corresponds to the Tyrosinase from *Streptomyces castaneoglobisporus* published by Matoba et al.[225] (PDB code 3ax0), in which a molecular mechanism for copper transport into the tyrosinase catalytic centre is proposed after crystallised at different times and soaking conditions.

In recent years, our research group has invested significant efforts in predicting metal-binding modes in proteins. Several software have been released in that frame, as is the case of Gaudi[226] which aims at molecular modelling based on a multi-objective genetic algorithm or an optimized version of the software GOLD.[57] However, these approaches are worthwhile when the binding site of the metals is identified. Blindly identifying binding sites requires exploring the whole protein or set of proteins, which obliges using less computationally demanding approaches than docking but still retaining the predictive capacity.

To solve this problem, we first developed BioMetAll,[55] a metal binding site predictor based on the pre-organization of the protein backbone.

BioMetAll's capacity for detecting metal binding sites has already been demonstrated in previous chapters. As a byproduct, BioMetAll has been shown to be able to detect transient sites and eventually display discrete positions of metals that seem related to diffusive paths. Despite promising results, BioMetAll aims primarily to identify pre-organised metal binding sites and, in a way, focuses on the correct geometry of the first coordination sphere. This is limiting when looking for transient geometries and full pathways for metal recruitment.

Recently, another predictor has been developed: BioBrigit.[227] It merges ML approaches and novel knowledge-based score functions, also part of the new version of BioMetAll. Although BioBrigit can be considered a competitor of BioMetAll, they are complementary. BioBrigit has been trained upon physicochemical variables not included in BioMetAll (e.g., Van der Waals, hydrogen bonding, etc.) and, at the same time, does not consider the BioMetAll pre-organisation concept. As a result, BioBrigit performs better in detecting pathways than identifying discrete binding sites. With this new tool in hand, the present Chapter is devoted to using BioBrigit as a suitable application to better understand the metal recruitment process of tyrosinases along evolution.

Tyrosinases and tyrosinase-related enzymes are a very broad family of metalloenzymes whose members involve different kinds of cofactors: heme, di-copper and di-zinc.[228] Copper tyrosinases catalyse the hydroxylation of phenol and oxidation of catechol to quinone, finally producing dopaquinone, a melanin precursor (Scheme 6.1). Melanin is a widespread pigment in many species, from bacteria, fungi, plants, and mammals. Melanin is also widespread inside mammal organisms, including the eyes, skin, and brain.[229] Di-copper tyrosinases are found in almost all organisms and mutations have been classically related to albinism and melanoma in humans;[229] while neuromelanin, produced by

the brain form of tyrosinase, has been recently linked to Parkinson's Disease (PD).[230] Strategies aiming at decreasing neuromelanin levels have been proposed as suitable therapeutic approaches for PD. Hence, both tyrosinase inhibition and activation could be regarded as an important target. For the former, the reduction of melanin production could help in the treatment both of melanoma and PD; while for the latter, activation could promote *de novo* enzymes more efficiently in melanin production, being a natural alternative for solar protection.
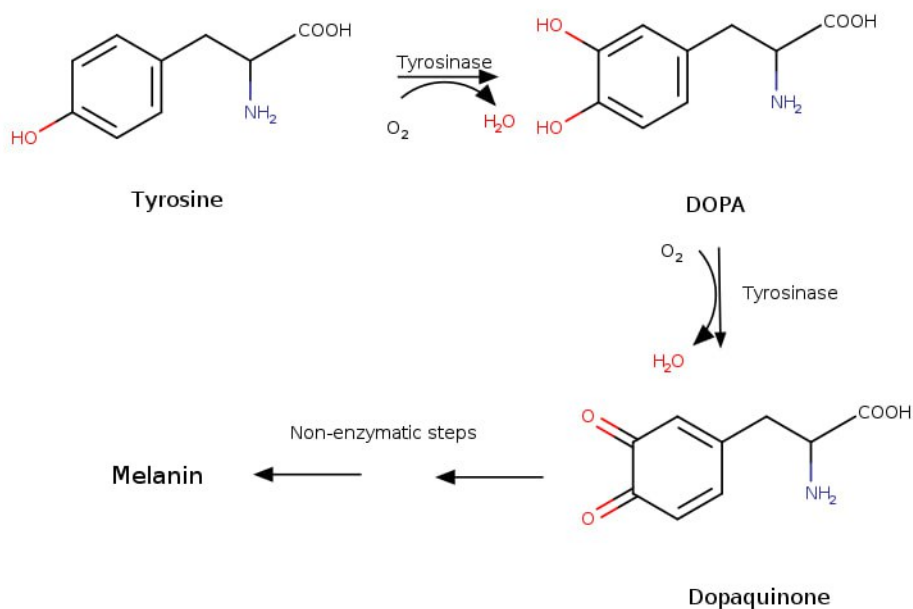
Scheme 6.1: Melanin production from tyrosine. Tyrosinases are involved in DOPA and Dopaquinone production, the first two steps of the reaction.

Copper tyrosinases have been shown to recruit copper in different ways. The *Streptomyces castaneoglobisporus* tyrosinase, however, cannot capture copper ions by itself and requires the aid of a metallochaperone. Metallochaperones are commonly present in biological systems, aiding the

transfer of the metal from the environment to the active site. In the tyrosinase system from *S. castaneoglobisporus*, the metallochaperone is called caddie, whose mutations have been observed to cause a lower activation of the tyrosinase. The tyrosinase-caddie complex is observable in the crystal published by Matoba et al.[225], with several Cu binding sites identified. Even though it is already demonstrated in the crystal that the caddie protein is actively involved in the process of Cu caption and transport to the inside of the tyrosinase, two main questions arise. First, the initial transition site proposed for Cu(II) is already buried inside the chaperone, not exposed, suggesting that a first step is not observed in this prediction. Nonetheless, the caddie crystallographic structure misses an external loop with high flexibility, solvent exposure, and remarkable histidine content, which could be crucial for Cu(II) recruitment. The second issue is to question the need for a metallochaperone itself, as the first transition site inside the tyrosinase seems to be solvent-exposed without the chaperone presence.

Other di-copper tyrosinases do not require metallochaperones. In fact, the Streptomyces species' tyrosinases may not be wholly representative of the bacterial form of these enzymes since the *Bacillus megaterium* tyrosinase has been reported not to require a copper chaperone for activity.[231] Along evolution, the tyrosinase metalloenzyme has consistently conserved the dinuclear metal active site,[232] though gained the ability to proceed without the caddie protein.[229] For mammals and humans more differences are found, though four conserved regions are detected among their tyrosinase structures. One of them is the intramelanosomal domain, which contains i) a tyrosinase-like subdomain with the binuclear metal ion region, typical of the tyrosinase family; and ii) a Cys-rich subdomain, exclusive of mammalian organisms. Such Cys-rich subdomain is said to directly interact with the entrance to the active site,[229] suggesting such region may compensate for the caddie role in loading the active site.

In this Chapter, we rely on a computational pipeline based on BioBrigit to analyse both the metal binding channels in *S. castaneoglobisporus*, and human tyrosinase, providing further analysis that could not be reached by X-ray structure and finally discussing this discovery in terms of evolutive adaptation.

6.2 Computational Details

6.2.1 Initial Models

The caddie-tyrosinase (CTyr) complex crystal structure (PDB 3Ax0),[225] soaked with Cu(II) at different times, has been selected for the setting up of the computational approach. As the caddie subunit is incomplete in the crystallographic model, the missing loops were modelled with Homology Modelling in UCSF Chimera Software.[142]

A homologous tyrosinase human protein (HTyr) model (Uniprot code P14679) was obtained with Modeller v5.0,[50] a Homology Modelling software for its comparison with human tyrosinase. Tyrosinase-related protein 1 (TYRP1) crystal structure was used as a template (PDB code 5M8L), considering its high sequence identity (43.32%) with the tyrosinase protein. AlphaFold was also applied to obtain a model for the HTyr, with an RMSD of 0.98 Å with the model already obtained through Homology Modelling. Then only the Homology Modelling model was considered.

6.2.2 Molecular Dynamics Simulations

Once the two models were obtained, the *S. castaneoglobisporus* and the human tyrosinase were submitted to Classical Molecular Dynamics without Cu ions in the catalytic centre. The complexes were embedded in a cubic box of pre-equilibrated TIP3P water molecules and Na^+ ions to

balance the total charge according to the needs of each system. Simulations were performed with the AMBER ff14SB force field[146] in the NPT ensemble for 300ns, using an integration time step of 1 fs. Constant temperature and pressure were set by coupling the system to a Monte Carlo barostat at 1.01325 bar and a Langevin thermostat at 300 K. The SHAKE algorithm[100] was used to constrain hydrogen atoms bonds.

GaMD simulations were also performed in parallel, with coordinates starting after 10ns of MD simulations. The AMBER ff14SB force field[146] in the NVT ensemble was applied, with SHAKE applied to constrain the bonds involving H atoms and an integration time step of 2 fs. A dual boost was performed on dihedral and total potential energy (igamd=3), finally obtaining a GaMD of 1μs long.

6.2.3   Prediction of metal binding pathways

BioBrigit is a hybrid machine learning knowledge-based approach developed in the group[227] by coworkers. The Machine Learning part of the algorithm consists of a convolutional neural network trained on 3-D point clouds (3D-CNN), representing protein environments to distinguish between metal-binding and non-binding regions. The code is based on two convolutional layers of 512 and 256 neurons each. For the training data set, 103,530 sites from 30,186 proteins were considered, with 28 different metals. This represents the first ML code that considers such a variety of metal ions.

Then, the knowledge-based part of the algorithm consists of a scoring function that has been built specifically for metal-residue interactions. This scoring function will approximate the probability of a certain probe being bound to the protein, considering three different components. The first one is metal-residue binding affinity, considering that certain met-

als are more likely to bind to specific residues. The second is the protein backbone pre-organization, building a geometrical description to determine how close the pre-organization of a certain protein region is to the ideal geometry for metal binding. Finally, a boolean filter is included, to make the search more efficient since unlikely regions are discarded.

Despite the main target of BioBrigit is to predict metal binding sites, it provides flexible results by the fact that several physicochemical properties are considered and a filter can be added. Accordingly, it may also be a suitable tool for the prediction of metal pathways, whose code and applications are still under development. In this section, one possible application for a metal-diffusion path of BioBrigit is explained, according to my role as a user-test in this project.

6.2.4 Dockings

Once the crystallographic pathway in *S. castaneoglobisporus* tyrosinase was properly predicted with BioBrigit, and some of their intermediate steps were submitted to Docking. GOLD software[57] was used, including modified parameters for metal ion binding.[59] Genetic algorithm (GA) parameters were set to 50 GA runs and a minimum of 100,000 operations each. The remaining parameters were set to default, including pressure, number of islands, crossovers, or mutations. Docking solutions were then visualised with GaudiView,[145] a built-in UCSF Chimera.

6.2.5 Analysis

Principal Component Analysis (PCA),[149] RMSD all-to-all and clustering was performed on MD simulations to evaluate exploration exhaustiveness. Electrostatic surface calculation was also performed with Chimera UCSF to evaluate electrostatic complementarity for metal ion binding.

6.3   Results

6.3.1   Caddie on the role of a crucial partner for the metal recruitment of Tyr

In the work of Matoba et al. [225], five discrete binding locations of Cu(II) were identified in the CTyr crystal structure obtained by soaking the protein at different times and leading to a clear metal recruiting channel (Figure 6.1). Two are in the active site and correspond to catalytic metal ions: i) Cu(A) coordinating with His190T, His194T, and His216T; and ii) Cu(B), linked to His38T, His54T, and His63T (T stands for Tyrosinase). The three others represent intermediate locations in the transition of the metal from the solvent-accessible region to the binding site. Site A is the most solvent accessible. There, Cu(II) is bound to residues Glu67C, His68C, and His82C, all belonging to the caddie protein (C stands for the Caddie protein). Site B is still entirely located inside the caddie, which requires the rotation of His82C, involved in Site A, to allow the coordination of Cu by both His82C and His97C. Site C represents the transition between the caddie and the tyrosinase, suggesting that the coordination requires the transport of the Cu ion by a rotameric movement of His97C, which would finally coordinate with His97C, His54T and a water molecule. Interestingly, His54T is part of both the penultimate transition step (Site C) and the coordination sphere of the catalytic site for Cu(B).

A first series of calculations was carried out on the X-ray structure published by Matoba et al.[225] As expected, BioBrigit calculations correctly predict all catalytic and transient copper binding sites observed in the experimental structure (Figure 6.1). Interestingly, the predicted sites show differences in score between the more solvent-exposed regions, which present a score for each probe of 0.5 (blue in Figure 6.1), in the most
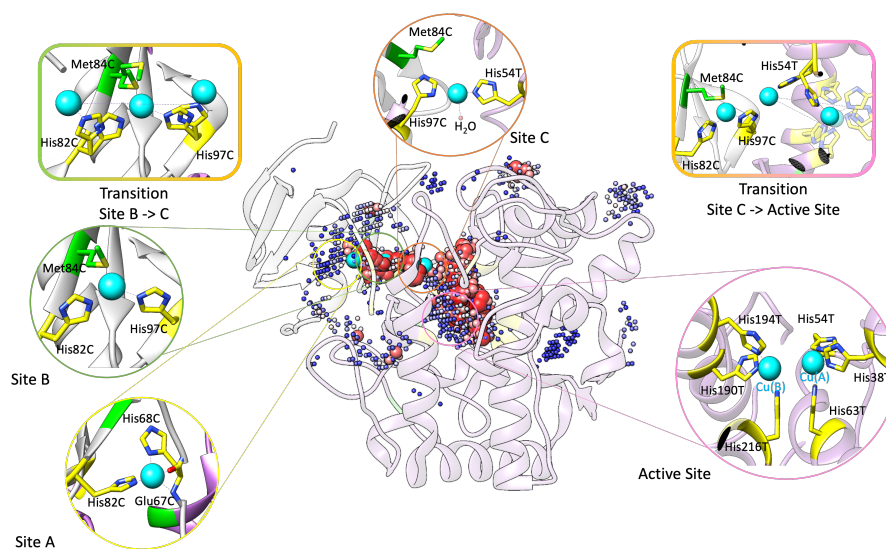
Figure 6.1: Binding sites observed in the crystallographic structure by Matoba et al.[225] (PDB code 3Ax0) overlapped with BioBrigit prediction for metal binding. Red spheres correspond to more likely binding sites (score of 1), while the blue ones correspond to less likely binding sites (score of 0.5).

exposed regions, such as Site C; while the more buried ones, such as the active site with the two catalytic coppers, scores of 1 (red in Figure 6.1) (see Table 6.1). In fact, the 5 best-punctuated regions by BioBrigit effectively correlate with the Active Site (Cu(A) and Cu(B)) (cluster score of 143.5 and 96.7) and detect the residues that comprise sites A (73.6 cluster score), B and C (122.5 and 101.9 cluster score, detecting transition areas). Other sites (clusters 6 to 14 in Table 6.1) with lower affinities (cluster score below 45) are scattered at the protein surface, which cannot be associated with any diffusion path, whose role is yet unknown although allovalency could not be discarded.[233]

However, BioBrigit only provides probes and does not describe the coordination geometry. To further demonstrate how it could be useful in a

Table 6.1: BioBrigit score for cluster of probes in CTyr.

Cluster	Coordinators	Score of the cluster
0	His38T, His63T, His190T, His194T, His216T *(Active Site)*	143.5
1	Met43T, Asp45T, His68T, His82C, Met84C *(Transition Site C - Active Site)*	122.51
2	His54T, Glu182T, His190T, Asn191T, His97C *(Transition Site C - Active Site)*	101.88
3	His38T, His54T, His63T *(Active Site)*	96.7
4	His66C *(Site A)*	73.85
5	Glu67C, His68C, Gln80C, His82C *(Site A)*	59.6
6	Glu175T, His180T, Arg185T	44.34
7	Arg229T, His230T, Asp232T	37.07
8	Asn6T, Asp87T,Asp91T	22.9
9	Arg228T, Asp264T, His265T, Tyr269T	12.19
10	Asp87T, Arg153T, Tyr218T, Lys221T	10.18
11	Arg140T,Glu114C	10.12
12	Arg185T, Asp88C	7.59
13	Glu149T	7.53
14	Asp112T, Tyr138T	6.68

more extensive exercise for metal recruitment, metal-compatible protein-ligand dockings as set up in our group[59] were performed for copper at sites 0 to 5, which show the higher scores (Table 6.1).

Docking calculations reinforce the BioBrigit prediction with sites A, B and C, involving convenient orientation of the residues Glu67C, His68C and His82C for A; His82C and His97C for B; and His97C and His54T for C. A remarkably important feature is the requirement of certain residues, such as His82C, His97C and His54T, to reorientate in order to transport metal ions from one location to the next position. These dockings sustain our hypothesis about the rotameric transition of His97C. Interestingly,

docking calculations demonstrate the indispensable role of Met84C in bridging the metal transport between the chaperone and the tyrosinase itself, which is not identified in the crystallographic structure. The transition between the chaperone and the tyrosinase could not be observed by BioBrigit since Met84C is the only coordinated residue to Cu during the transition, which retrieves a very low score for the software.

Though BioBrigit reaches excellent agreements with the metal binding sites observed in the X-ray structure, one of the remaining questions concerns the identification of the first site for metal recruitment. Indeed, the first copper experimentally and theoretically observed is already buried inside the chaperone, suggesting a first step may be missing. So the question is if there is another metal binding site better exposed to the solvent.

Interestingly, some BioBrigit probes are detected in a slightly more solvent-exposed area of the caddie than in the X-ray structure, though with low affinity, which also include residues from site A (cluster number 5 in Table 6.1). This hints at a possible additional site for metal recruitment as this region corresponds to part of the caddie's metallochaperone missing in the X-ray structure and composed of two additional histidine residues (residues from 60 to 65 with sequence GGGAHH). Such a missing region could be important for Cu(II) recruitment, considering its flexibility, solvent exposure, and histidine content.

To further address the first question and study the caddie metal recruitment capacity at the solvent-exposed region, a Homology Model was built to complete the metallochaperone missing loop. On this model, BioBrigit indeed predicts a metal binding site at the cluster of amino acids His298C, His299C, His300C, His302C and His316C (Figure 6.2), which clearly suggests that the X-ray misses a crucial part of the system in term of metal binding. This new site allows to better understand the role of the chaperone and suggests that the first step in metal binding is a solvent-

exposed, highly flexible and histidine-rich region, entirely composed of amino acids from the metallochaperone.
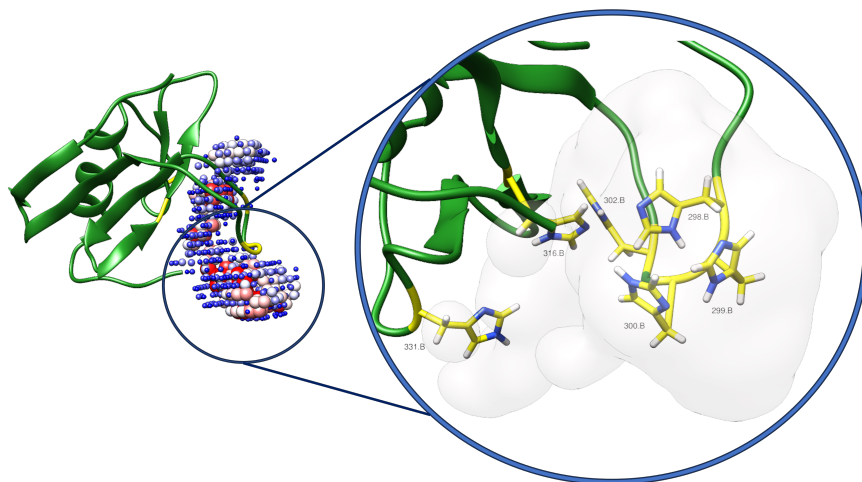
Figure 6.2: Metallochaperone model, with added missing regions. A cluster of His appears (yellow residues), which is predicted by BioBrigit.

The last question regarding CTyr concerns the intrinsic need of the enzyme for a metallochaperone. Indeed, from a purely structural point of view, the enzyme shows a highly solvent-accessible entrance (Figure 6.3). Biobrigit calculation solely at this stage can not answer why an external recruiter would be needed, but further inspection of the CTyr physicochemical properties was expected to provide some pointers. Hints actually were obtained by comparing electrostatic surface analysis between bacterial tyrosinases that require metallochaperone and those that do not. Unfortunately, very limited sequences are available for bacterial tyrosinases and less X-ray structure. Therefore, calculations were carried out on the crystal structure of *S. castaneoglobisporus* as a representative of the former and the AlphaFold model of *B. megaterium* for the latter.

The electrostatic calculations performed on CTyr from *S. castaneoglobisporus* (Figure 6.3) show that, despite space, the entrance area is strongly positively charged, something which would make it highly improbable for Cu(II) to diffuse by itself. For *B. megaterium* the same wide entrance is observed but the surface area of the channel entrance is highly charged, massively in negative values. This would suggest that the caddie protein may then help in bypassing such surface repulsion, capturing Cu(II) from the solvent and transporting it directly to a well connecting with the active site, requiring certain rearrangement of the surface to bind the metal.
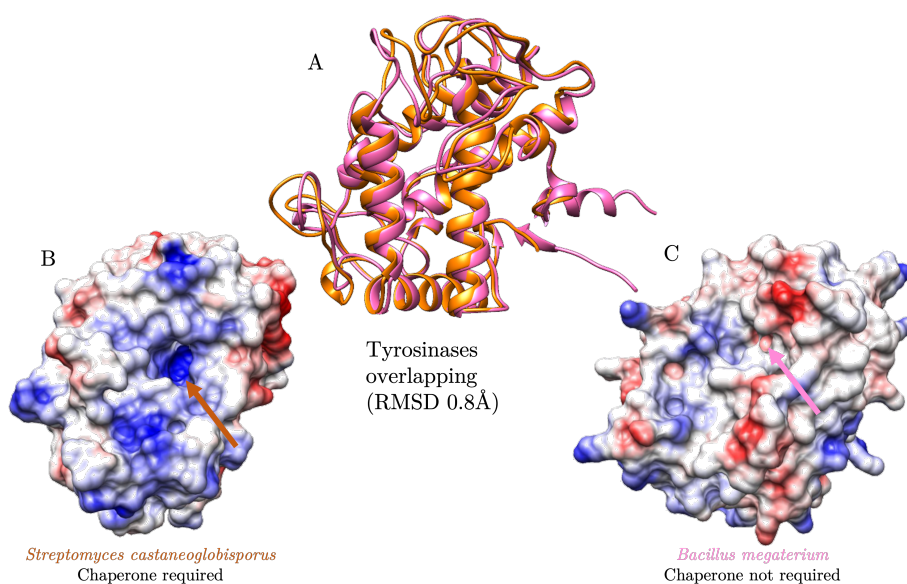
Figure 6.3: A) Overlapping of tyrosinase structures which require (orange – PDB code 3Ax0) and do not require (pink – AF model D5DHA8) a metallochaperone to load de Cu(II). B) Surface of *S. castaneoglobisporus*, whose copper entrance is positively charged (orange arrow). C) Surface of *B. megaterium* tyrosinase, copper entrance is negatively charged (pink arrow).

Finally, although CTyr is known to be folded in the absence of the metal, local rearrangements cannot be overruled easily.[234] On the basis of the analysis performed on the X-ray structure, one could naturally wonder how copper recruitment is coupled with molecular motions of the receptor. To assess this aspect, a GaMD simulation of the apo form of CTyr complex was performed.

The overall analysis of the 1μs GaMD simulations shows that the system is very stable, with no major conformational change observed in the structure, except for the chaperone His-rich loop (His299C, His300C, His302C, His316C), which demonstrates a high flexibility, as expected. To assess how dynamical events affect the copper recruitment path, a representative structure was obtained from the most populated cluster of the simulation and BioBrigit calculation were carried out on this geometry. Results show some new, low-affinity areas (clusters score between 37 and 1) on the surface of the protein that does not lead to the active site. Hence, both the apo structure and the crystal demonstrate the same metal pathway, starting in the His-rich loop of the metallochaperone, with no striking differences between the loaded states.

In summary, this part of the study on the CTyr shows i) that BioBrigit effectively predicts the Cu-metal sites and channel as suggested by the X-ray structure, ii) that to complete the path, several rearrangements of the amino acids from the X-ray structure are required, for example, Met84C, iii) that the apo form of the complex is stable and maintain the same transition path for the metal, iv) that the missing region of the X-ray structure at position 60 to 65 is fundamental for the metal recruitment, which is absent in the experimental structure but it appears as the first step in metal recruitment and v) that the need for a metallochaperone is mostly to reduce possible unfavourable repulsion of the entrance of the binding site to the same metal needed by the protein.

6.3.2 Searching for the origin of the absence of chaperone in mammalian copper tyrosinase

As already observed, the *S. castaneoglobisporus* tyrosinase requires the caddie protein to recruit its metal ions. On the contrary, human tyrosinases have evolved to go without the need of the chaperones. Both tyrosinases actually have substantial changes in their general fold. The human tyrosinase is composed of four conserved regions: a N_{Ter} signal peptide, an intra-melanosomal domain, a transmembrane α-helix and a C_{Ter} cytoplasmic domain. The intramelanosomal domain contains both the catalytic active site, very similar to the one in CTyr, and a Cys-rich subdomain, whose function is still unknown. Besides, the mechanism by which the human tyrosinase recruits the metal ion is still under strong debate.

The unique crystal structure available for human tyrosinase corresponds to TYRP1 (PDB code 5M8L), a tyrosinase-related protein, which contains a di-zinc catalytic site instead of copper. However, it presents a 44% sequence similarity to the human copper tyrosinase, meaning that a high-confidence model for the latter (HTyr) can be obtained either through Homology Modeling or AlphaFold. Both methodologies have been applied to obtain models for HTyr, with very low differences between them (RMSD 0.98 Å). Due to the small differences, only the Homology Modelling model was selected for subsequent analyses.

The overlap of *S.castaneoglobisporus* with the HTyr, retrieves an RMSD of 1.209 Å, reinforcing the conservative character of the structure. Nonetheless, if the surface of both proteins is compared, remarkable differences are observed (Figure 6.4). Similarly to what is seen in bacterial tyrosinases, the caddie binding area of CTyr in the solvent-exposed surface is strongly positively charged, which would hamper Cu(II) diffusion with-

out aid. HTyr, on the contrary, is highly negative in the surface area (Figure 6.4), though no evident entrance to the active site is observed.

Despite the absence of an accessible metal binding site in the HTyr model, BioBrigit calculations identify two zones rich in metal binding probes. One precisely corresponds to the negatively charged surface region overlapping with the location of the metallochaperone in CTyr, while the other corresponds to the Cys-rich zone located at the opposite side of the protein and far from the chaperone binding motive (Figure 6.5).

This suggests that HTyr has a better electrostatic pattern of the same accessible region than CTyr and a similar one to *B. megaterium*. The lack of solvent accessibility observed in the model would be due to conformational restrictions of the Homology Modelling step. Yet, the high-affinity pattern of the Cys-rich region questions if an alternative mechanism could also exist and be the result of evolutive pressure. Indeed, cysteine amino acids are known to be, together with methionine and histidine, the best copper-coordinants both for Cu(II) and Cu(I) oxidation states. Hence, an additional step in the copper recruitment could be hypothesised as a stepwise mechanism for HTyr in which the Cys-rich region represents the first binding area of the Cu. The following steps would involved conformational changes which could bring the copper near the native copper sites of the enzyme and its complete diffusion.

To demonstrate such a possibility, a first analysis of the TYRP1 X-ray structure shows that the cysteines in the Cys-rich region are massively involved in disulphide bonds (Cysteines 42–65, 56–99 and 101–110).[229] Therefore, the stepwise mechanism would only be viable considering an oxido-reduction mechanism during which the copper breaks the disulphide bonds to be recruited.[235]
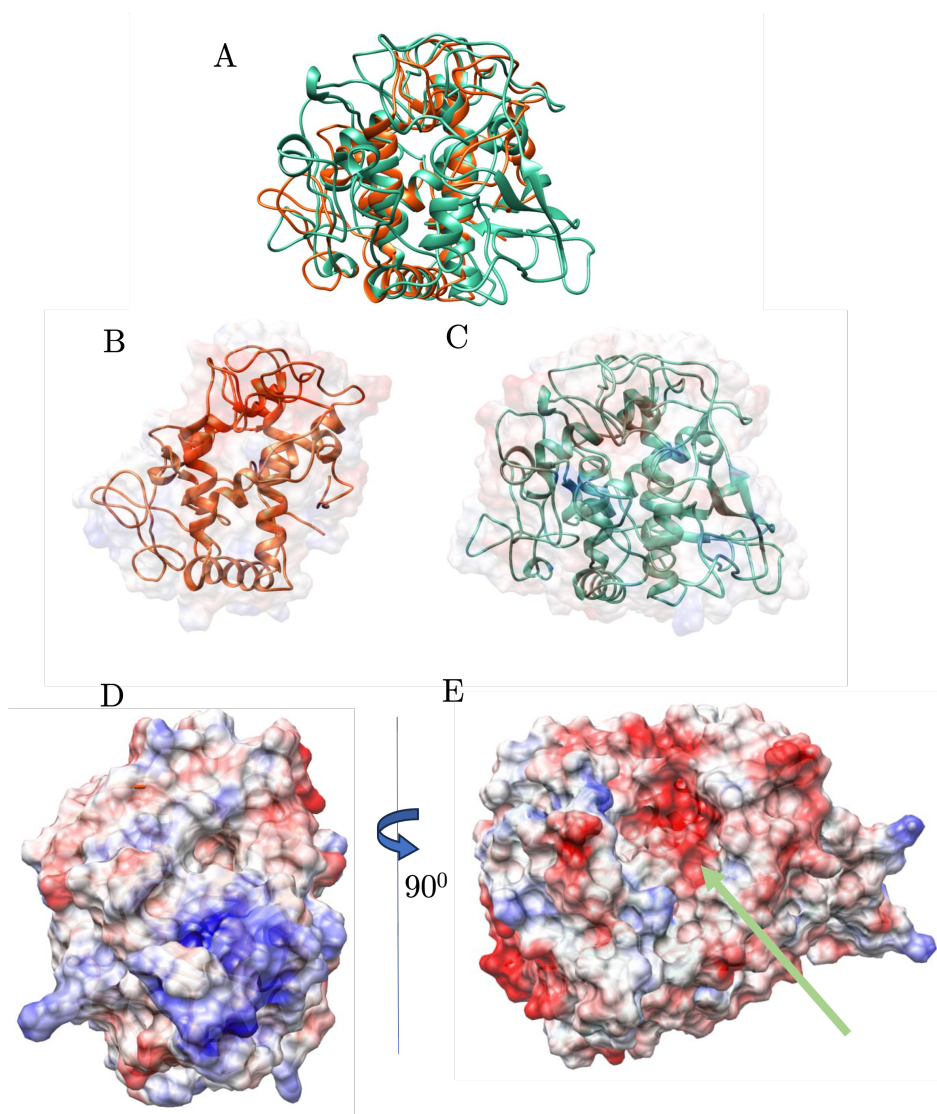
Figure 6.4: Overlapping of crystallographic tyrosinase (orange) with human tyrosinase model (blue). RMSD of 1.209Å A) Coloured surface (red negative charge, blue positive charge) of the crystallographic *S. castaneoglobusporus* (B) and human model (C) tyrosinase. D) and E) correspond to the HTyr in front and 90º flipped view, demonstrating a strong negative patch, though no evident pocket is observed.

Then, a 1μs GAMD simulations were performed on HTyr model without such disulphide bonds, and BioBrigit was applied on 10 representative structures of the simulation after clusterization. When simulating the Cys-rich region without disulphide bridging, the region appears highly flexible and BioBrigit predicts some high-affinity hotspots involving Cys18, Cys28, Cys37, Cys71, Cys73, Cys82, Cys85 and/or Cys94 (Figure 6.5). Despite major flexibility, neither channelling to the catalytic site nor displacement of the loop at the entrance of the cavity was observed. On the contrary, a low-affinity channel was constantly predicted in the same region as in the CTyr, though constituted by Asp168, Arg178, Asp181, Glu185 and Gln360. This suggests that copper recruitment is likely to occur from the same pathway as CTyr. However, it is to highlight that the solvent-exposed amino acids are not histidines, methionines or cysteines, the most common ones; but Arg, Glu or Asp, which show relatively limited metal binding capacity.[236] Yet, the presence of a more negative electrostatic field in HTyr could favour the metal to approach the binding pocket and finally diffuse to a most suitable copper environment, as the active site formed by histidines represents.

6.4 Conclusions

A computational pipeline including Homology Modelling, Docking and BioBrigit approaches has been presented in this last Chapter of Results, demonstrating the capacity of describing metal ion channels in bacterial tyrosinase. The channel, already experimentally defined by Matoba et al.[225] was successfully reproduced, detecting with high affinity those residues in charge of transporting Cu(II) ion from the solvent to the active site. The essential role of the caddie protein, a metallochaperone in charge of recruiting and transporting the metal ion to the entrance of the active site, has also been demonstrated.
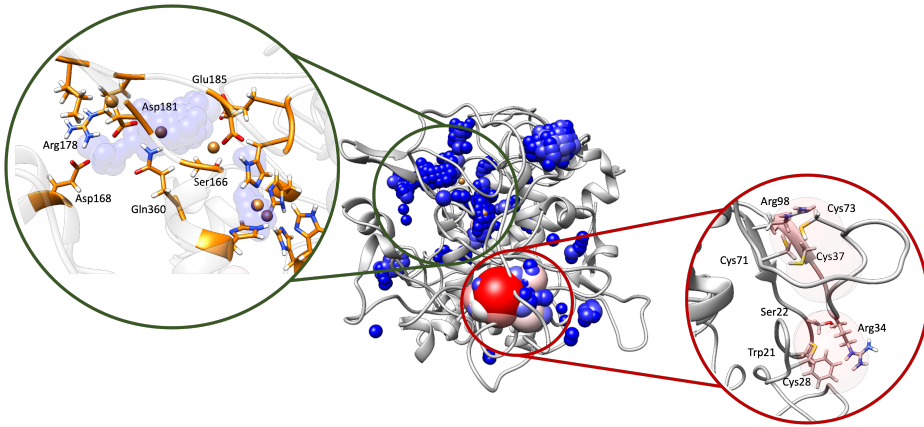
Figure 6.5: BioBrigit results in cluster 7 obtained from the Human Model Tyrosinase GaMD (grey). High-affinity areas are detected both in the Cys-rich area, formed by Cys37, Cys73, Cys71 and Cys28 (bottom right). A complete channel composed of lower-affinity areas is detected in the negatively charged pore (top left), composed of Asp168, Arg178, Asp181, Glu185 and Gln360 before arriving at the active site.

Due to the high importance of several diseases related to tyrosinase, the same pipeline has been applied to a designed model of human tyrosinase. Such a model has been obtained based on the crystal structure of TYRP1, obtaining the human homologous protein to CTyr, with a di-copper active site. Despite the high similarity between both proteins' structures, HTyr presents new regions with respect to CTyr, especially remarkable the Cys-rich N_{Ter} region. Besides, HTyr is independent from the caddie protein, being able to load the Cu(II) by itself, though with a still unknown mechanism. Here, the same protocol as for predicting the channel in CTyr has been applied, suggesting a mechanism in which the Cys-rich zone, not present in CTyr, would be in charge of firstly recruiting the Cu ion from the solvent. Once captured, the Cu ion could be neared to the entrance of the native copper channel, which is composed of more negative-charged residues than CTyr, which would then transport the Cu ion to the active

site. Accordingly, considering that HTyr could be one of the most important druggable targets for melanoma and PD, further studies could base on the present results to i) experimentally confirm the role of Cys-rich area and ii) properly determine -and block- the metal channel of the human tyrosinase.

Overall, the pipeline here presented has been demonstrated useful for those systems in which the loading of metal ions is essential for their function, since metal ion channels can be effectively identified. Its applicability in other metalloproteins, especially those whose druggable target sites are still to be discovered, may represent a major advance in the treatment of diseases.

7

Final remarks and future perspectives

The main goal of this PhD has been to demonstrate how molecular modelling could greatly contribute to understanding the impact of metal ions on proteins and peptides involved in diseases or relevant for biotechnological development. The design of integrative computational approaches, which merges standard approaches such as Quantum Mechanics, Protein-ligand Docking and Molecular Dynamics Simulations, with novel techniques such as BioMetAll or BioBrigit, allows us to tackle the broad world of metallopeptides. Overall, the main objective has been reached by applying multiscale strategies both on the detrimental and

functional side of metallopeptides, besides providing a new strategy to deal with them. The discrete objectives can be summarised in chapters.

Chapter 4 demonstrates that a multiscale strategy can be successfully applied to determine the binding sphere and conformational impact that metal ions represent in the $A\beta_{42}$ aggregation process. First, the monomeric study on metal coordination highlighted the importance of carefully establishing a computational protocol suitable for properly describing $A\beta$ properties. Besides, special attention must be placed on the force field that is going to be applied, since the outcome of the simulation can be highly influenced. With the fibrillar structure, it was necessary to broaden the scope and introduce new analytical tools, as it was not enough just to compare the changes in the secondary structure, but also include tertiary structure adaptability upon metal ion binding. Finally, the gained knowledge and designed computational protocols were applied to also integrate the effect of familial forms of $A\beta$, both on monomeric and fibrillar structures, aiming at describing biologically relevant species.

Along these studies, the complexity of studying a biological system such as the $A\beta$ peptide has been highlighted, mainly due to the conformational variability that is observed for such systems, with many influencing variables. Thus, their proper analyses are extremely challenging, and sometimes it is difficult to find numerical descriptors that fit to describe the conformational behaviour. Second, simulating biological meaningful times is still missing, since our study is focused on the extreme cases of the aggregation process but tackling intermedia structures, such as oligomers, would represent a major advance in understanding Alzheimer's Disease. Besides, further experimental studies would be desired for correctly discerning the differential conformational patterns when considering familial variants.

Altogether, despite the challenge that dealing with experimentally unknown coordination spheres and large conformational variability represents, we firmly believe the results here provided may be useful for further understanding and treatment of Alzheimer's Disease. These results represent further opportunities to continue studying the complex, always exciting, world of amyloid aggregation.

Chapter 5 shows the potential of multiscale strategies for the description of catalytic reactions for catalytic peptides. Focusing on the catalytic role of either Zn(II) or hemin upon binding heptapeptide sequences, this study resulted in the molecular explanation of why slight changes in the sequence of heptapeptide fibrils can cause dramatic changes in the catalytic activity. Through the application of computational demanding techniques such as QM/MM and the use of particular analyses that allow us to determine the subtle differences among fibrils performing the same catalytic reactions. Hence, this work opens the way for further experimental work, since the collaboration between theoretical and experimental is remarkably important in the challenging and versatile world of catalytic fibres.

Chapter 6 aimed at providing a new computational tool to describe metal ion transport into preorganized proteins, considering the challenge that experimental approaches face to determine transitory metal binding sites. Using BioBrigit, a combined strategy standing on BioMetAll concepts and Machine Learning, the metal ion channel for a Tyrosinase case study is properly identified at a reasonable computational cost. The same approach is then applied to predict the metal ion channel of the Human Tyrosinase, providing a new possible druggable target for several diseases. Difficult as it is to determine metal ion channels, this approach effectively allows one to predict metal ion migration in protein structures, without requiring experimental techniques. Such an approach paves the way to

generate valuable knowledge in the promising field of metalloproteins.

This dissertation, though diverse, sticks to the driving line that metal ions are essential in life and that molecular modelling, often considered to be highly challenged in this field, is becoming each time more solid in this area. Despite the challenges their study represents, metalloproteins are not only present in lots of molecular cascades related to diseases but also represent the capacity to perform new-to-nature catalytic activities. Therefore, I expect the work here contained represents a tiny, but forward, step in unravelling the world of metallopeptides.

A

Chapter 4 - Supplementary information

A.1 Monomers

Copper systems
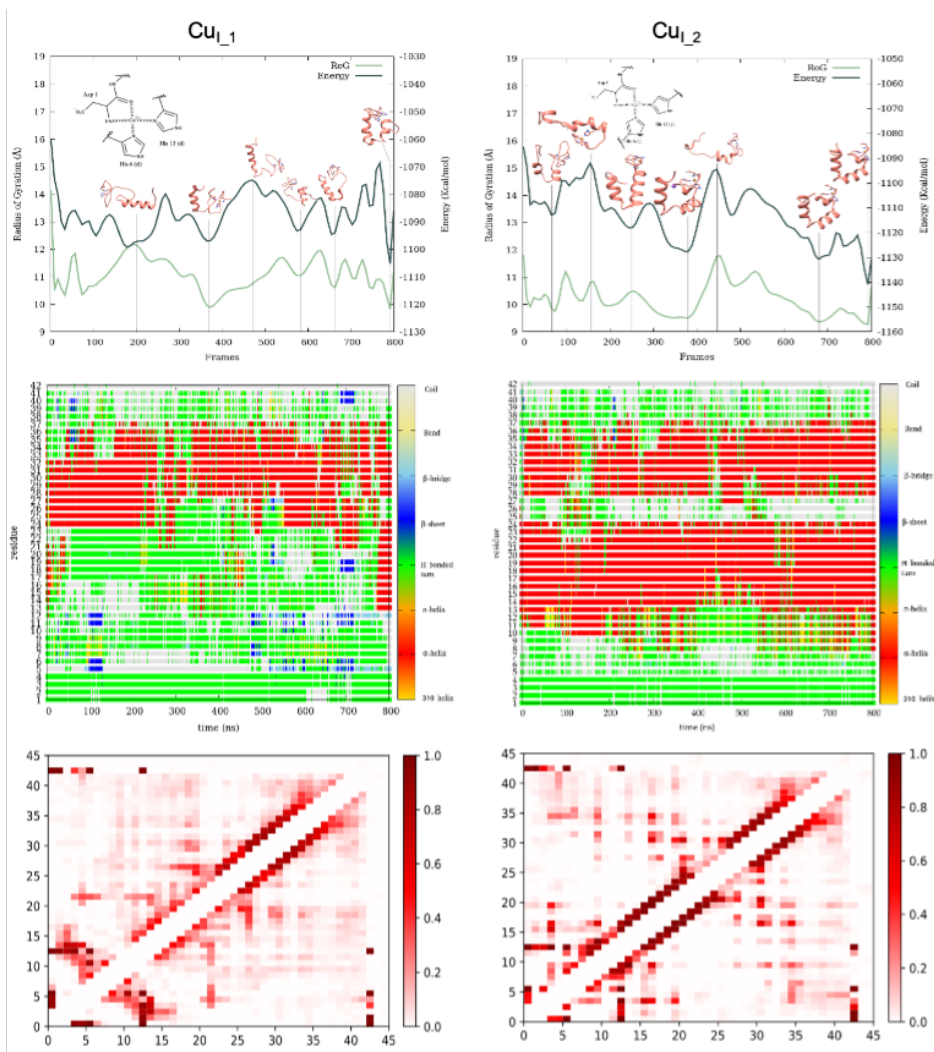
Figure A.1: Energy profile with Radius of Gyration and representative frames, Timeline and Frequency Contact map for Cu$_I$ complexes.
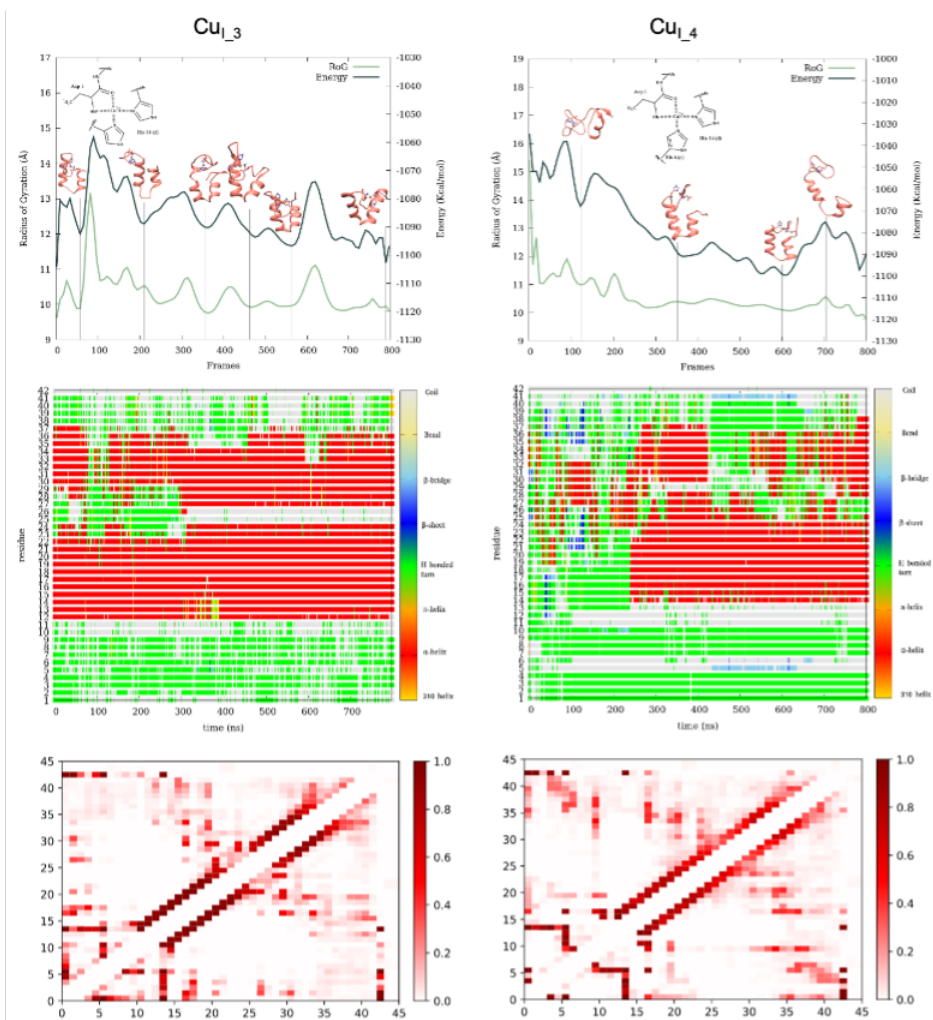
Figure A.2: Energy profile with Radius of Gyration and representative frames, Timeline and Frequency Contact map for Cu$_I$ complexes.
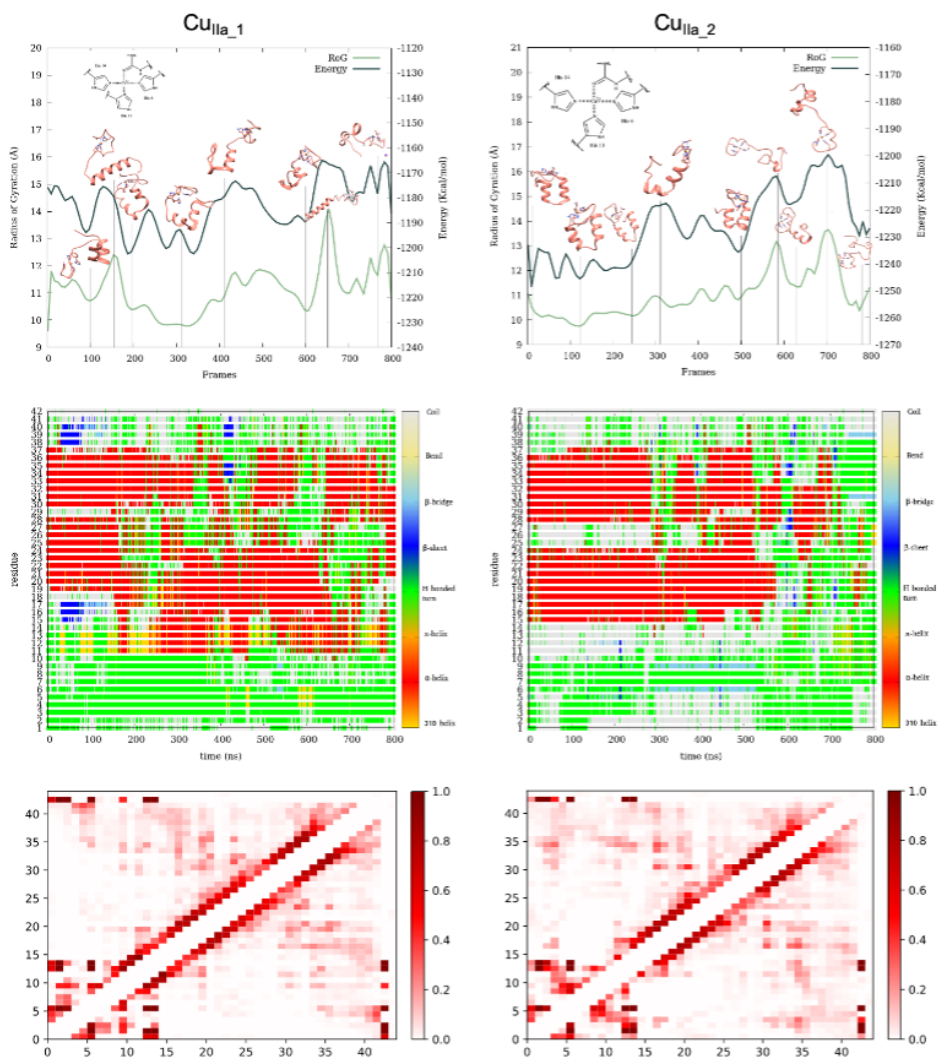
Figure A.3: Energy profile with Radius of Gyration and representative frames, Timeline and Frequency Contact map for Cu_{II_a} complexes.
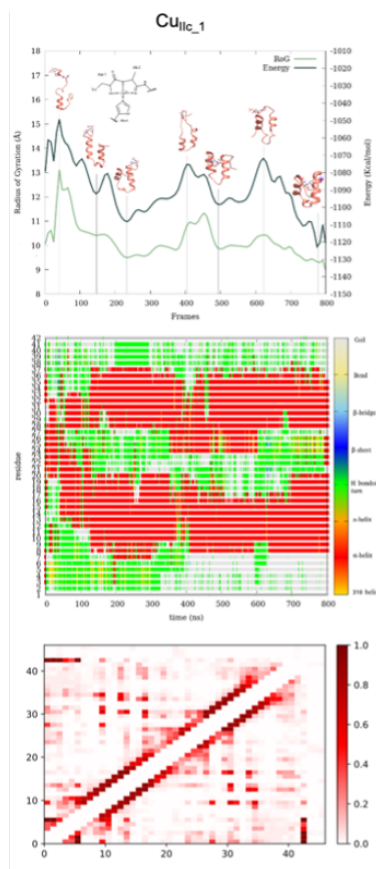
Figure A.4: Energy profile with Radius of Gyration and representative frames, Timeline and Frequency Contact map for Cu$_{II\_c}$ complexes.
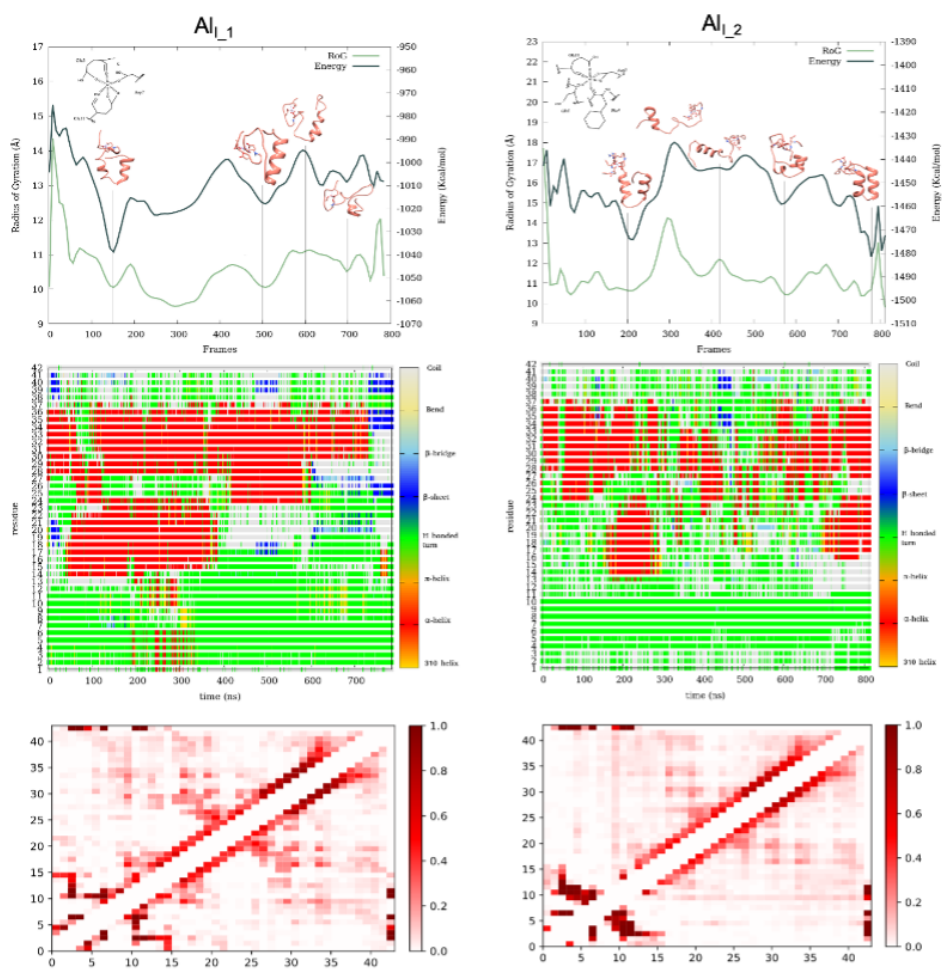
Figure A.5: Energy profile with Radius of Gyration and representative frames, Timeline and Frequency Contact map for Al_I complexes.
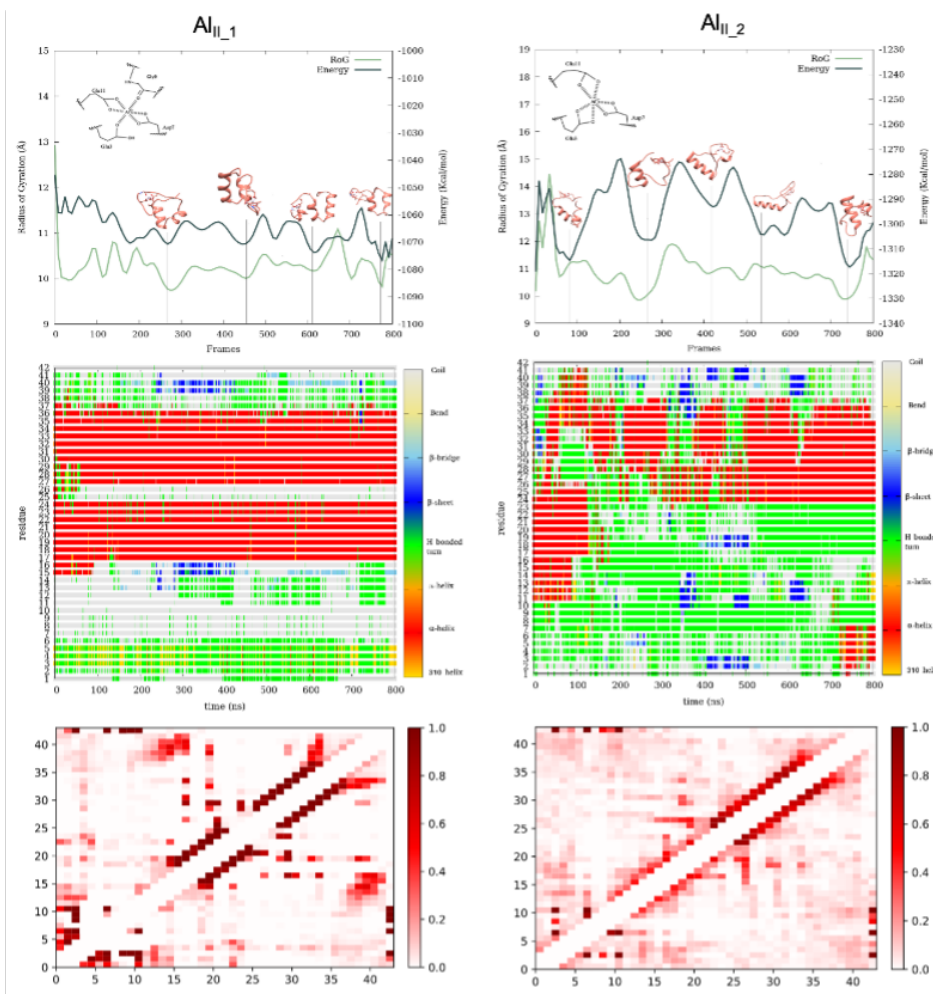
Figure A.6: Energy profile with Radius of Gyration and representative frames, Timeline and Frequency Contact map for Al_{II} complexes.
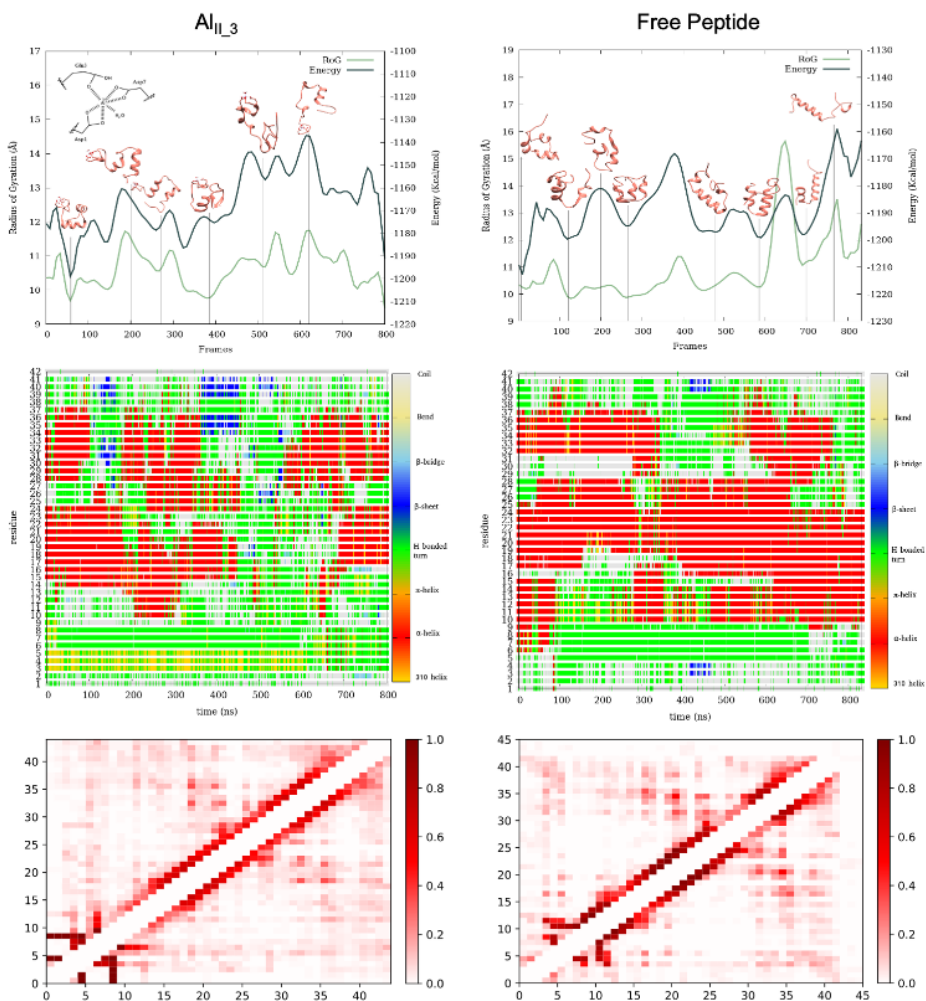
Figure A.7: Energy profile with Radius of Gyration and representative frames, Timeline and Frequency Contact map for Al$_{II\_3}$ complexes and Free Peptide.

A.2 Fibres
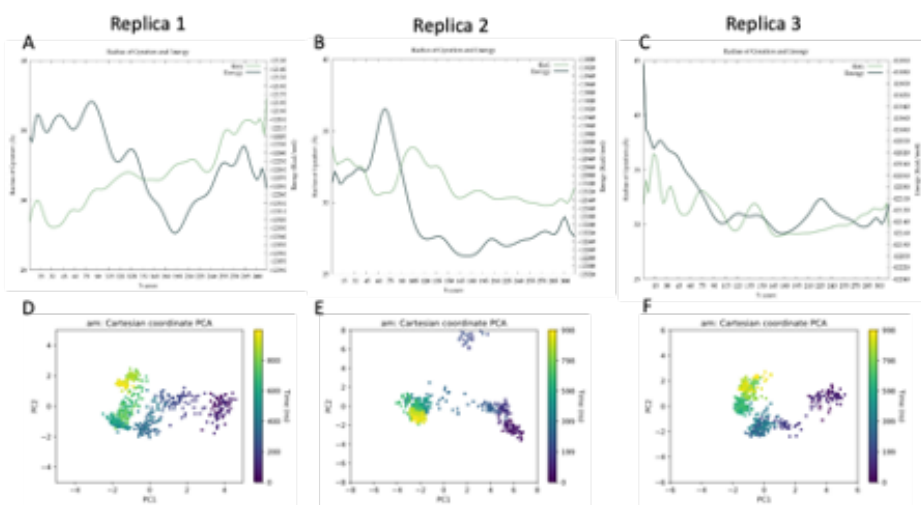
Figure A.8: A, B and C. Energy profile and RoG along the GaMD simulation for the three replicas of the metal-free fibre. D, E and F. Principle Component Analysis (PCA) of the same simulations.
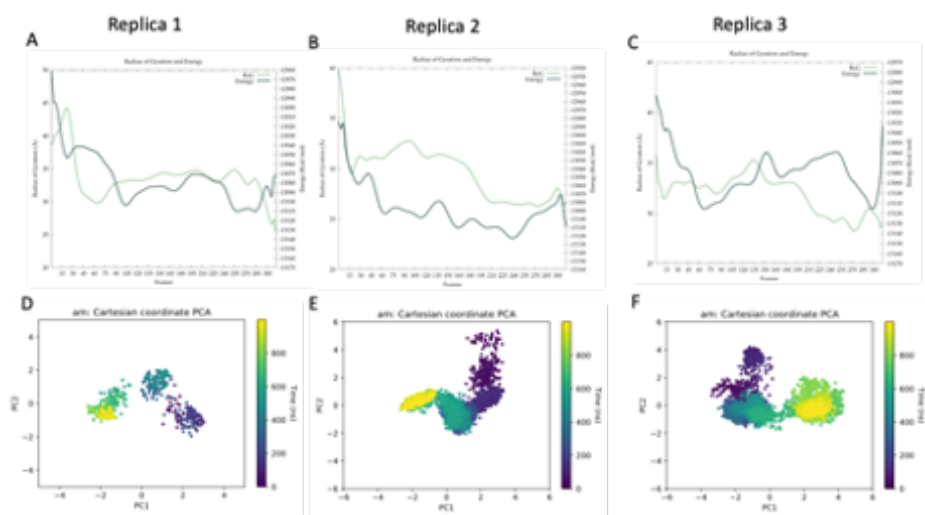
Figure A.9: A, B and C. Energy profile and RoG along the GaMD simulation for the three replicas of the Cu(II)-bound fibre. D, E and F. Principle Component Analysis (PCA) of the same simulations.
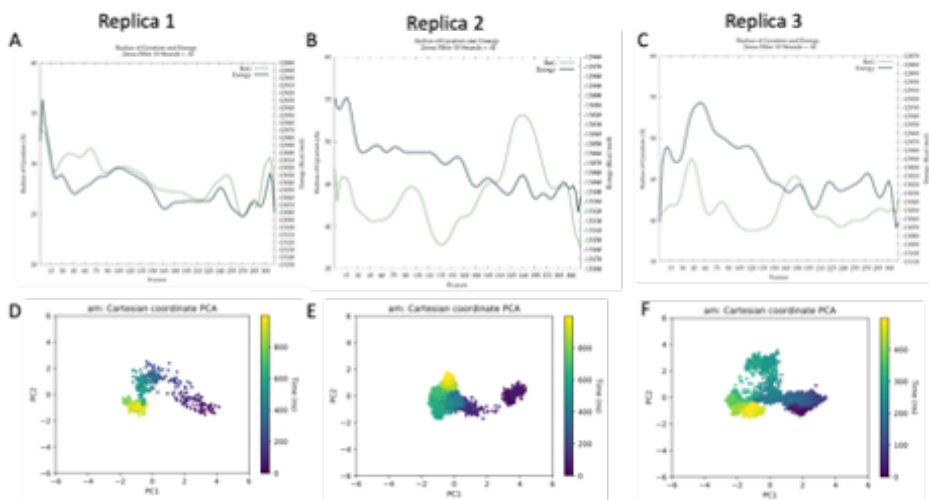
Figure A.10: A, B and C. Energy profile and RoG along the GaMD simulation for the three replicas of the Al(III)-bound fibre. D, E and F. Principle Component Analysis (PCA) of the same simulations.

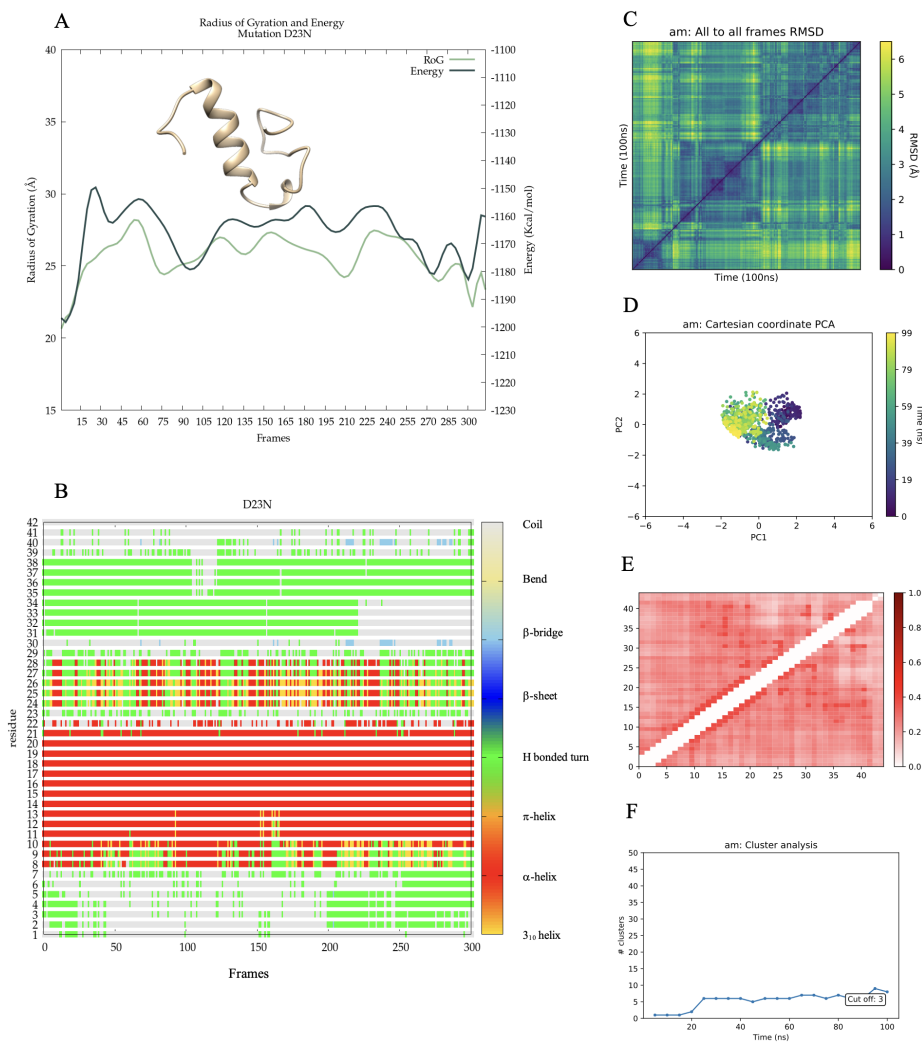# A.3   Aβ_{42} Variants

Monomeric Free Peptide
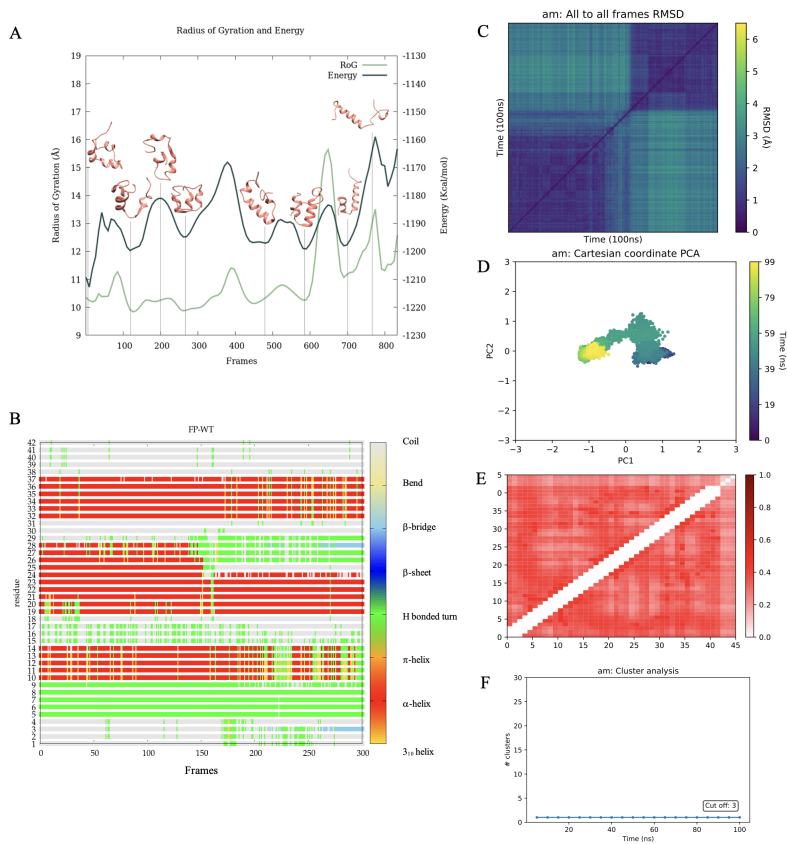
Figure A.11: Convergence analyses for FP-WT. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
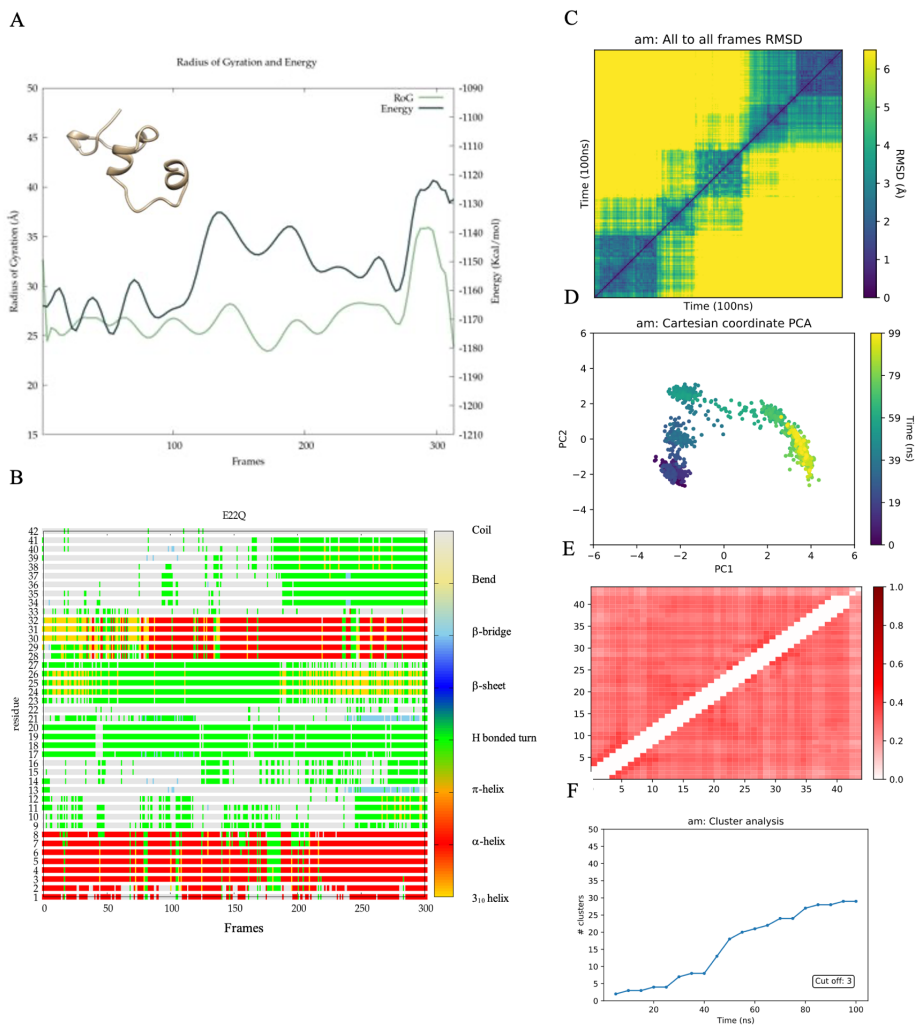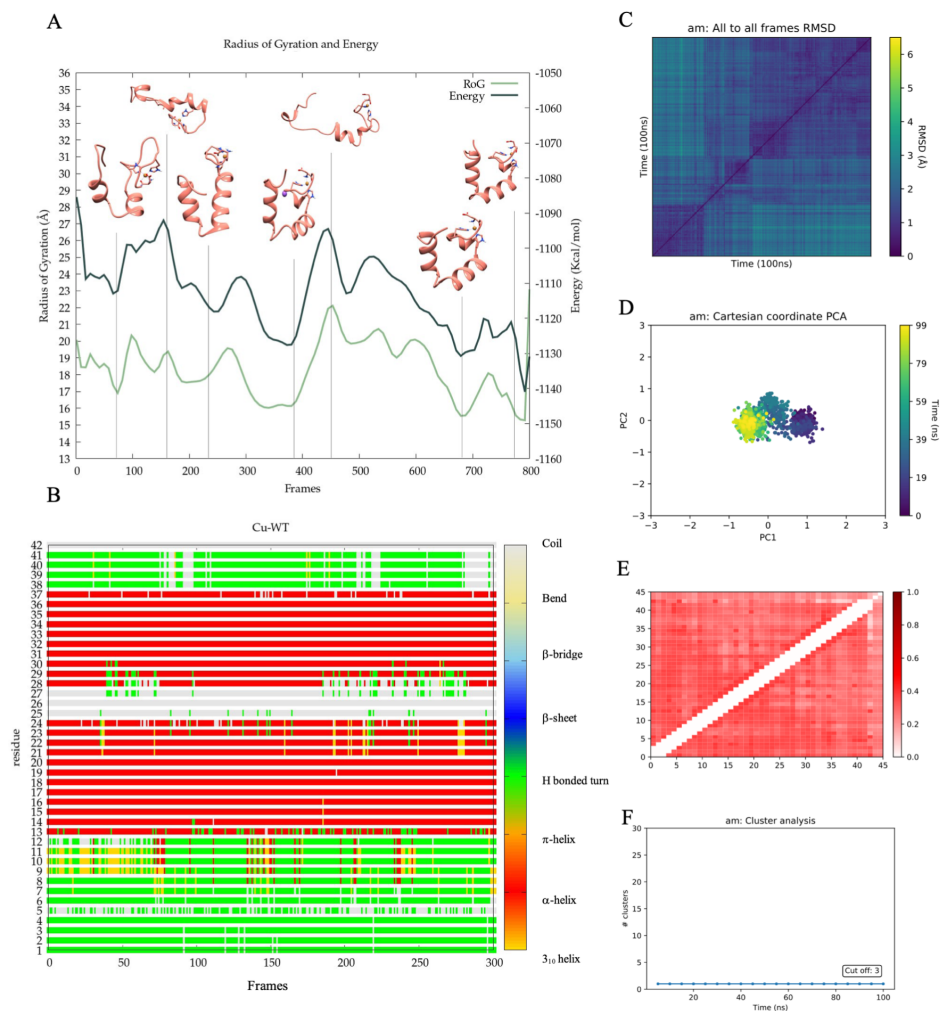
Figure A.12: Convergence analyses for FP-E22Q. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
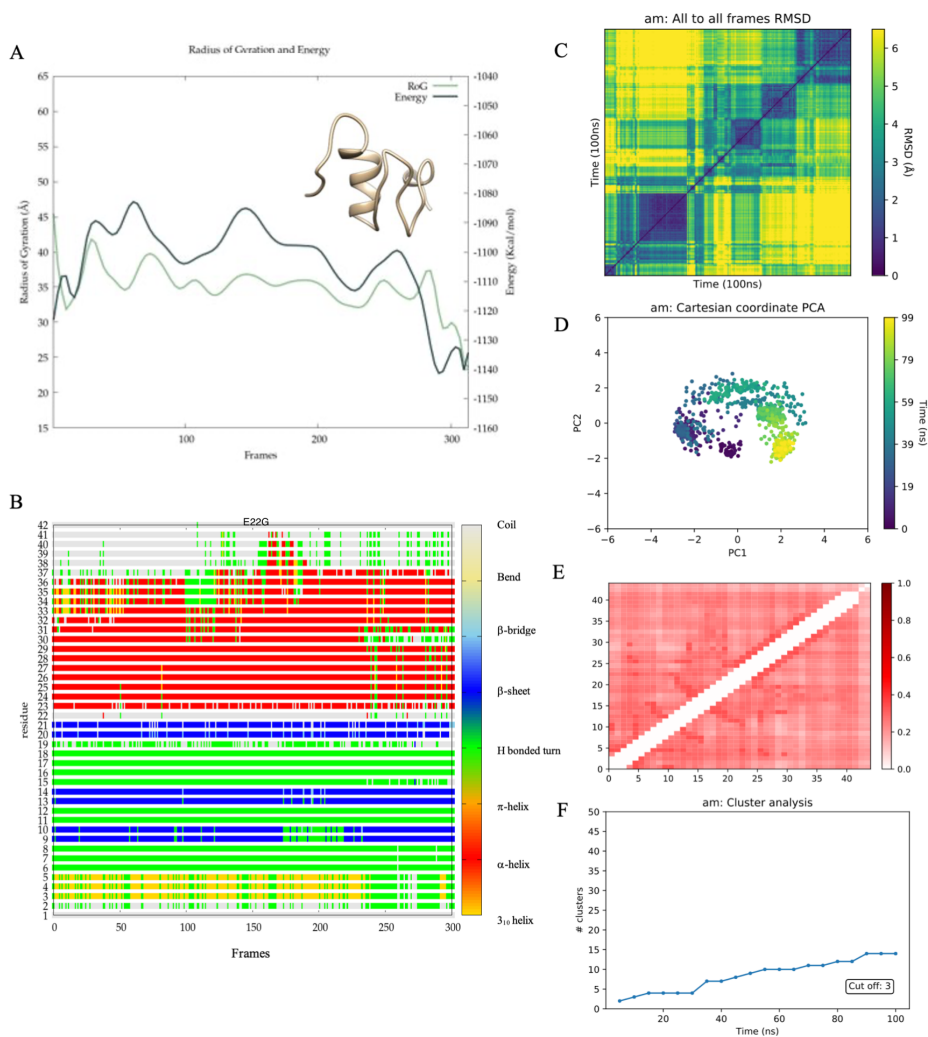
Figure A.13: Convergence analyses for FP-E22G. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
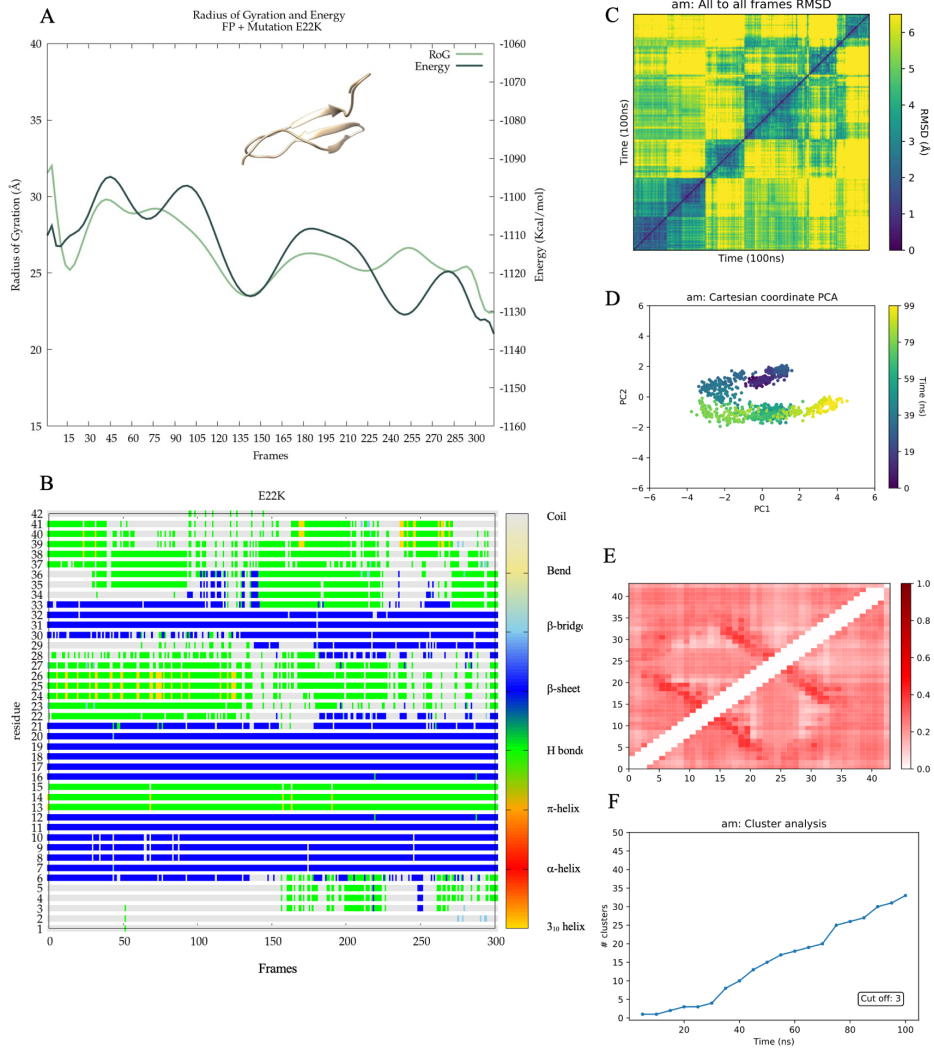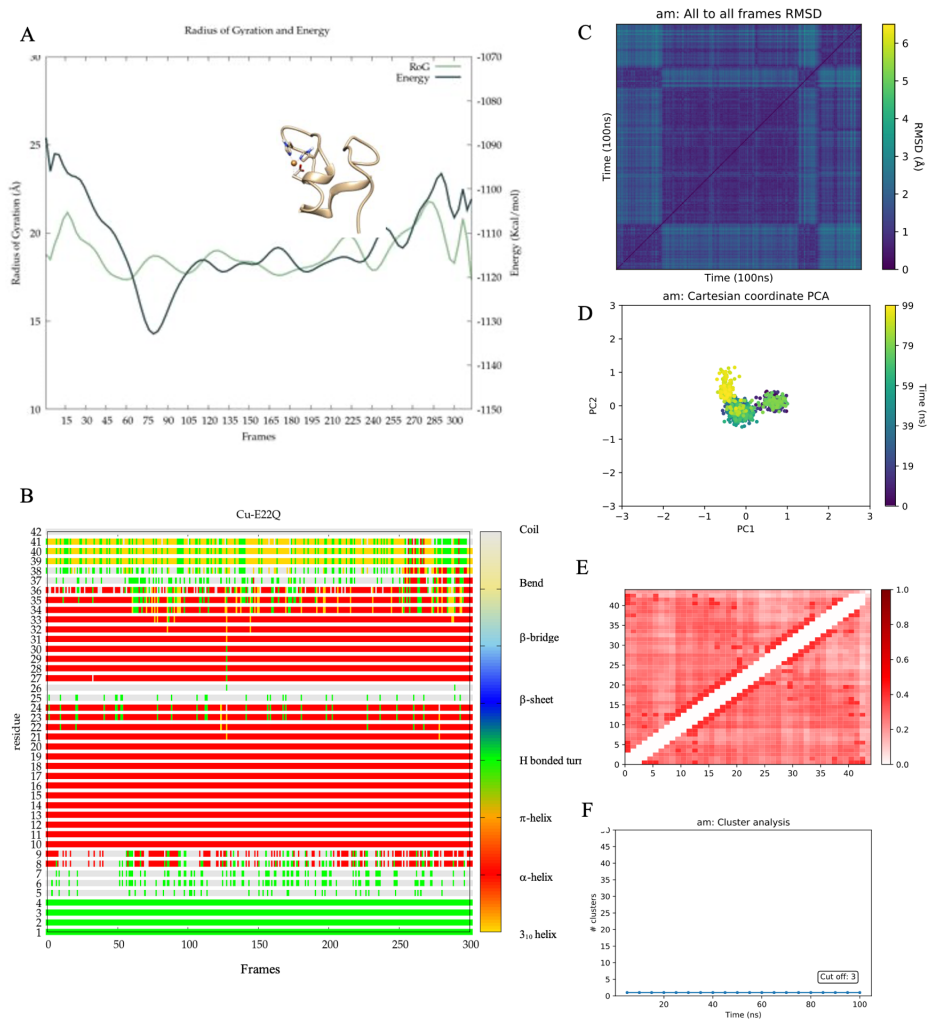
Figure A.14: Convergence analyses for FP-E22K. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
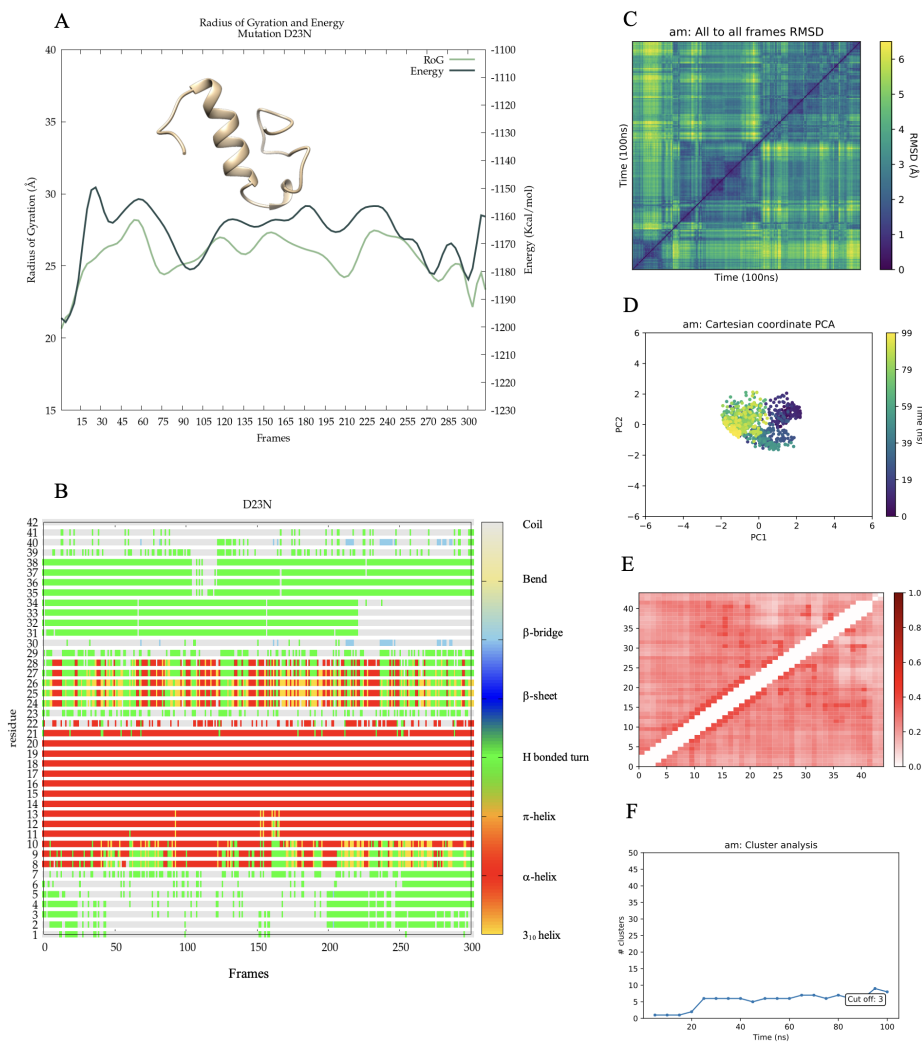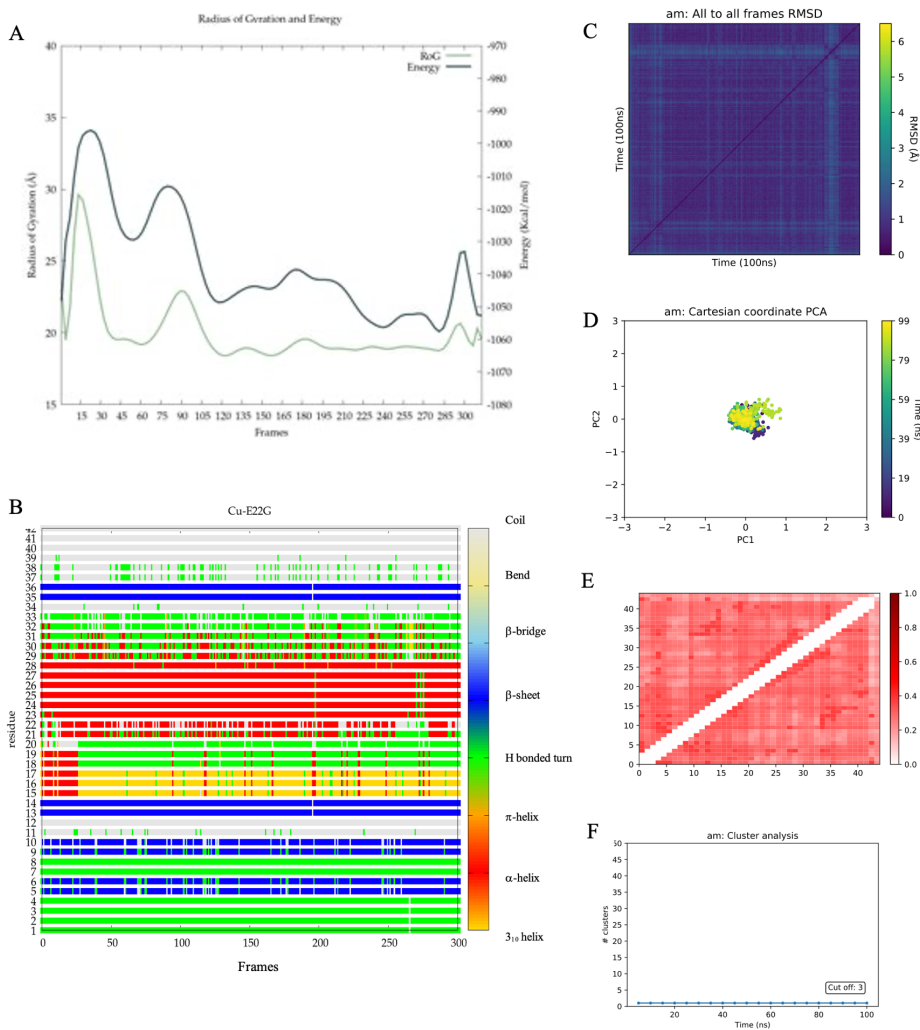
Figure A.15: Convergence analyses for FP-D23N. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.

Figure A.16: Convergence analyses for FP-D23N. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
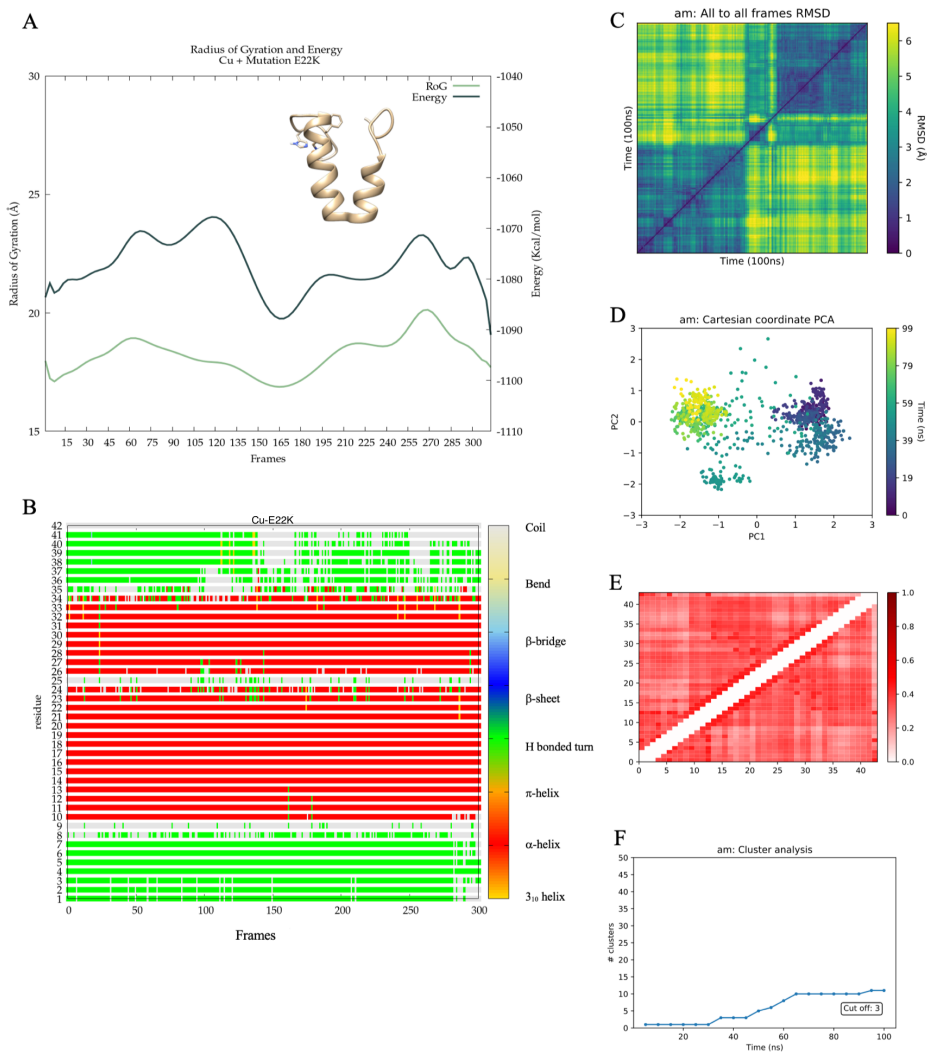
Figure A.17: Convergence analyses for Cu-WT. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
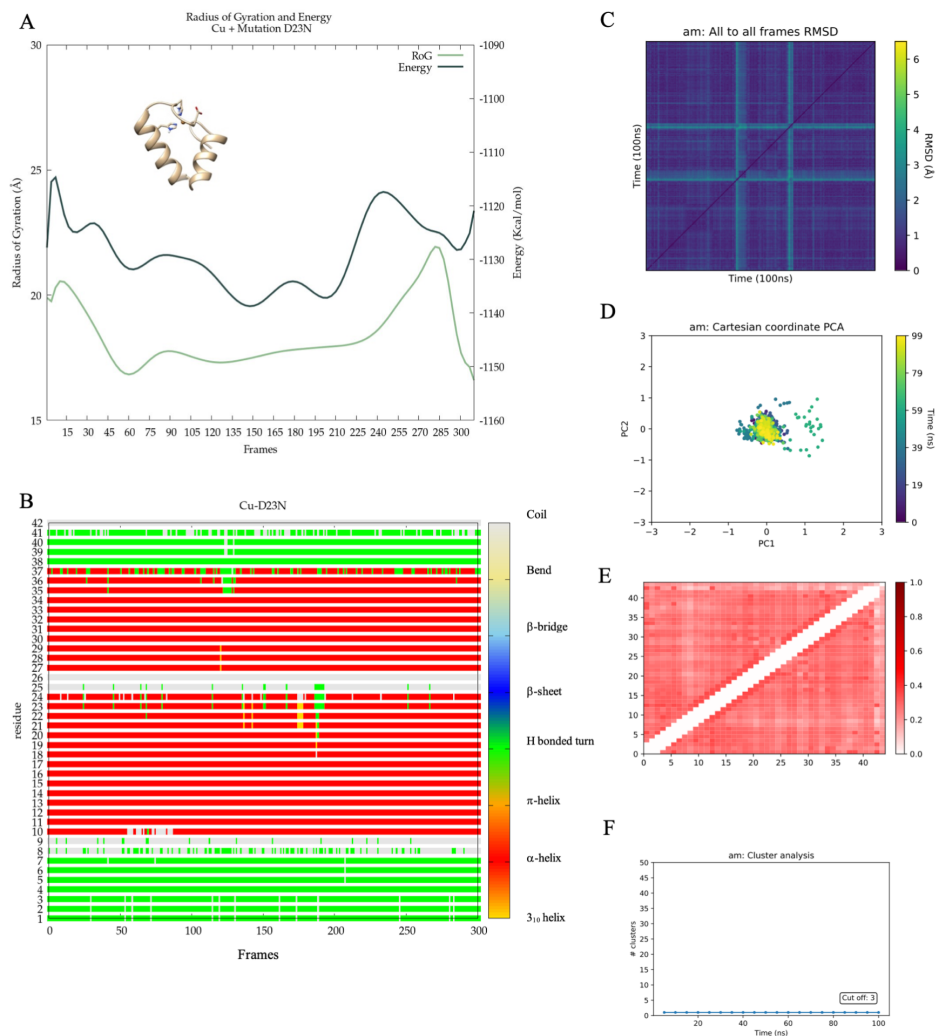
Figure A.18: Convergence analyses for Cu-E22Q. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
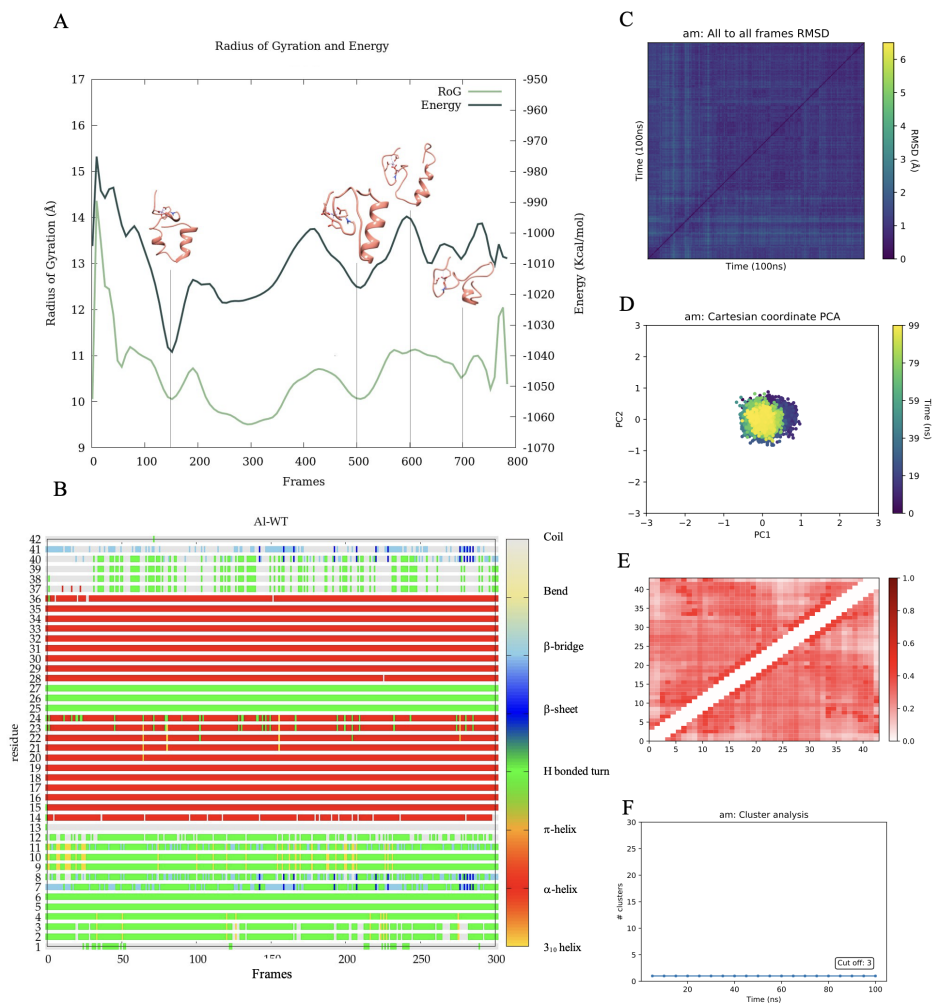
Figure A.19: Convergence analyses for Cu-E22G. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.

Figure A.20: Convergence analyses for Cu-E22K. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
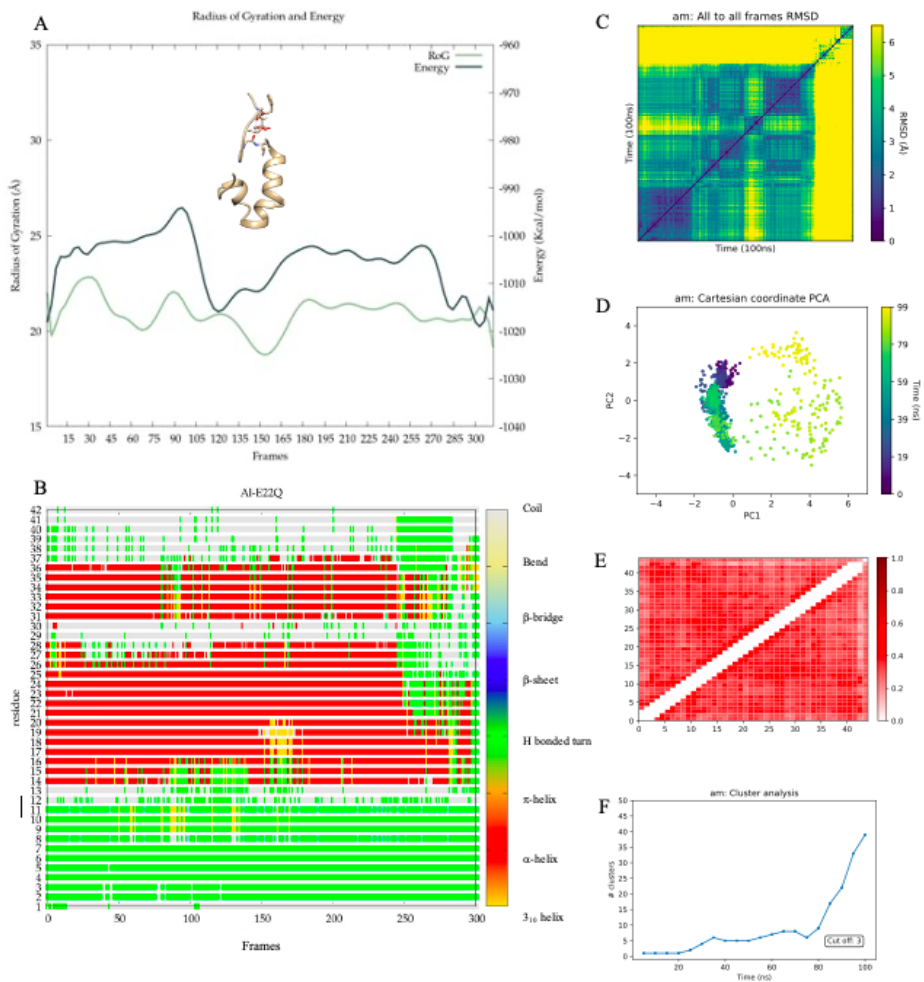
Figure A.21: Convergence analyses for Cu-E22K. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
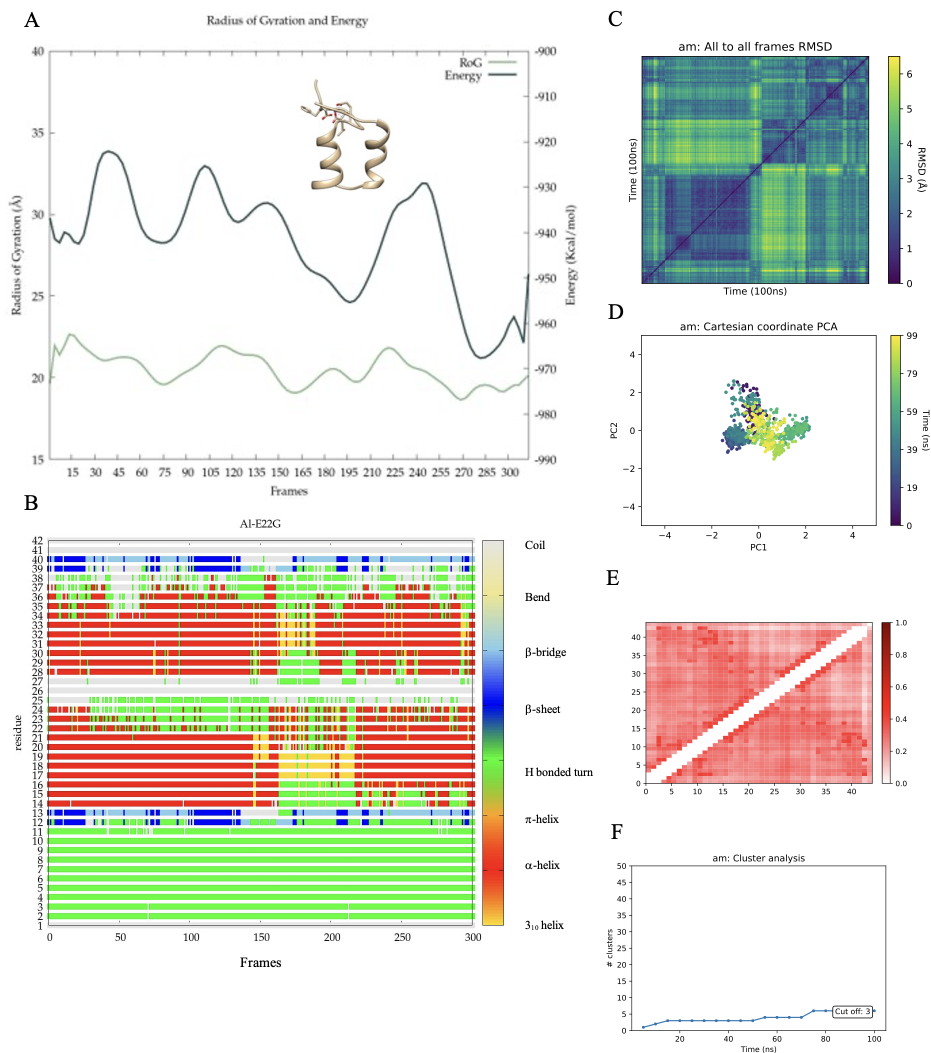
Figure A.22: Convergence analyses for Al-WT. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.

Figure A.23: Convergence analyses for Al-E22Q. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
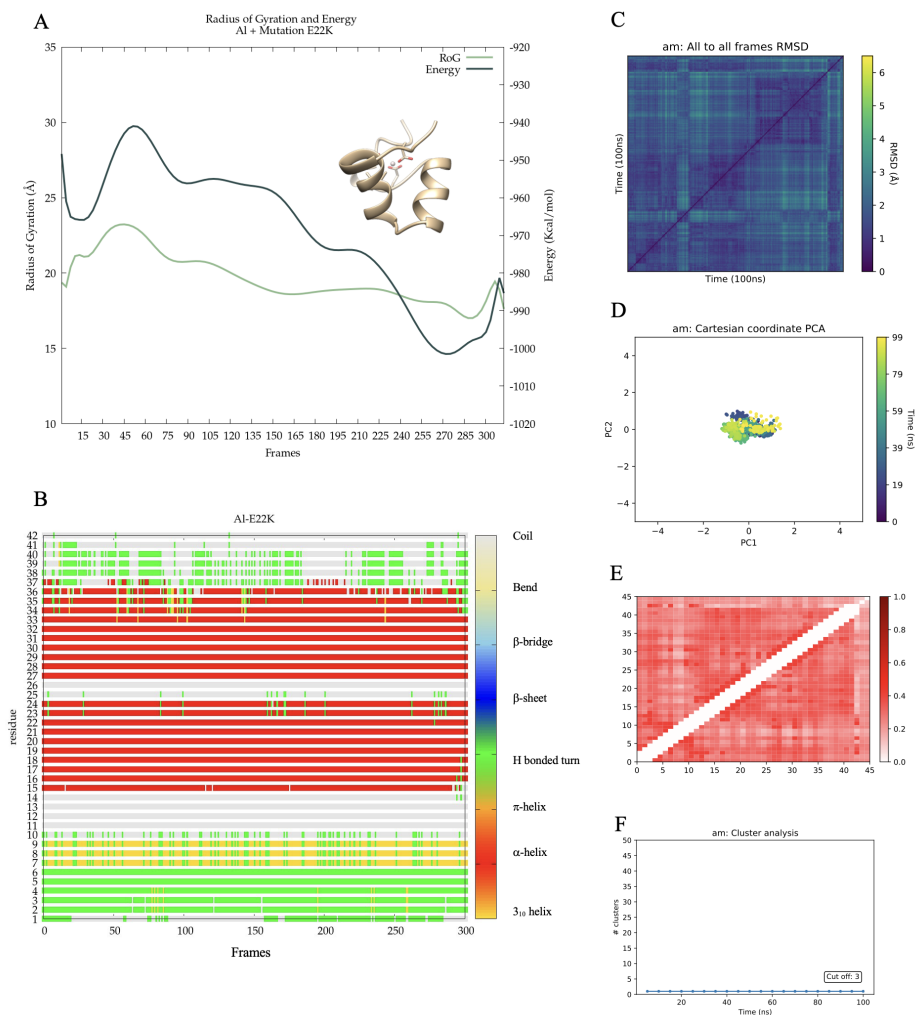
Figure A.24: Convergence analyses for Al-E22G. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
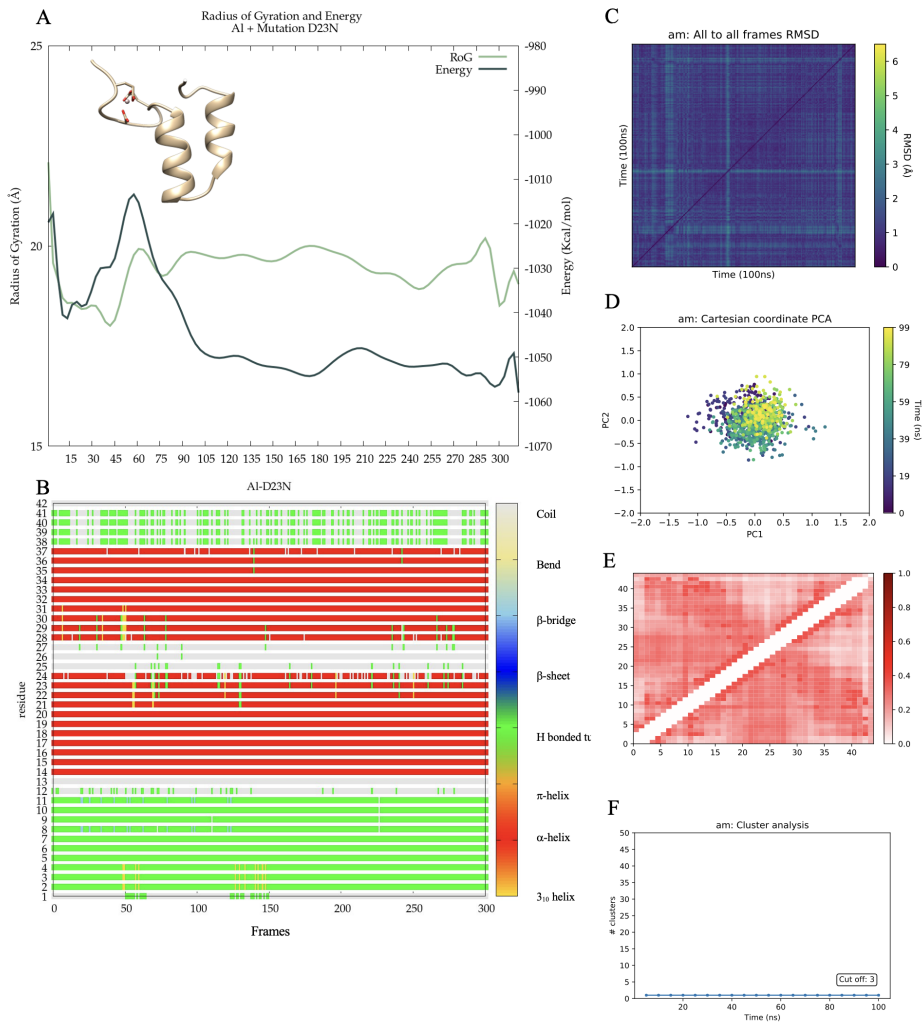
Figure A.25: Convergence analyses for Al-E22K. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.

Figure A.26: Convergence analyses for Al-D23N. Implicit solvent Free Energy and Radius of Gyration analysis, with the lowest energy structure, correspond to the GaMD simulation (A). Timeline Analysis (B), RMSD all to all (C), PCA exploration (D), Contact Map (E) and Cluster Countering (F) correspond to the MD simulation.
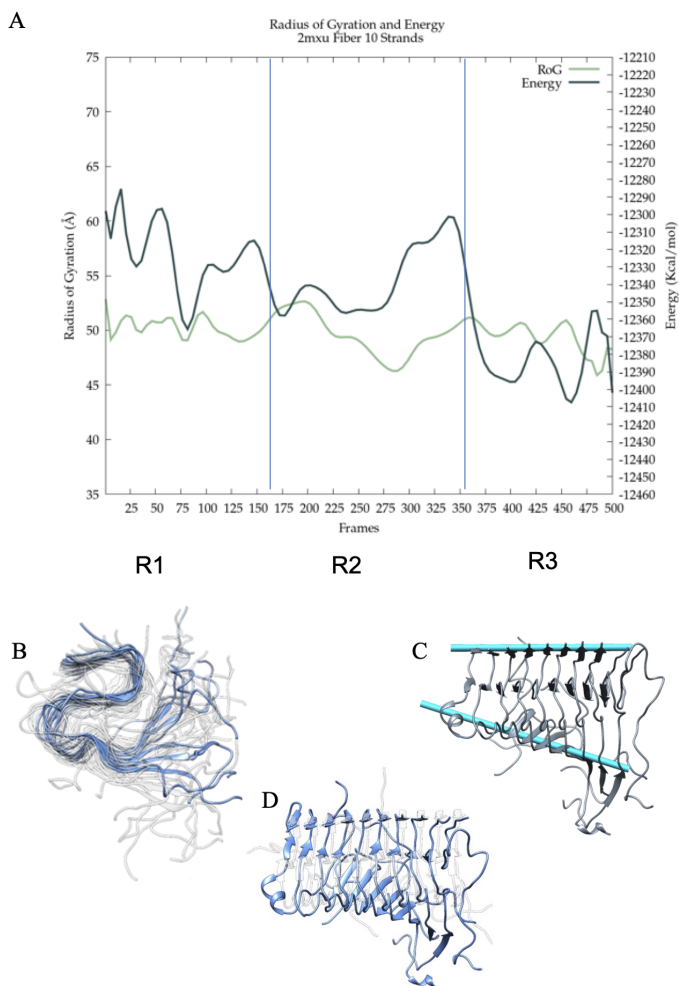
Figure A.27: Trajectory analyses for fibrillar Metal-Free E22Q Complex. A. Implicit solvent Free Energy and Radius of Gyration along the three replicas. B. The most populated cluster (blue) is in superposition with the other 9 most populated clusters. C. In blue, the two axes are comprised of residues 39-41 (top) and residues 13-15 in β-sheets (bottom) of a representative structure of the most populated cluster. D. Overlap of the lowest energy structure with the initial structure.
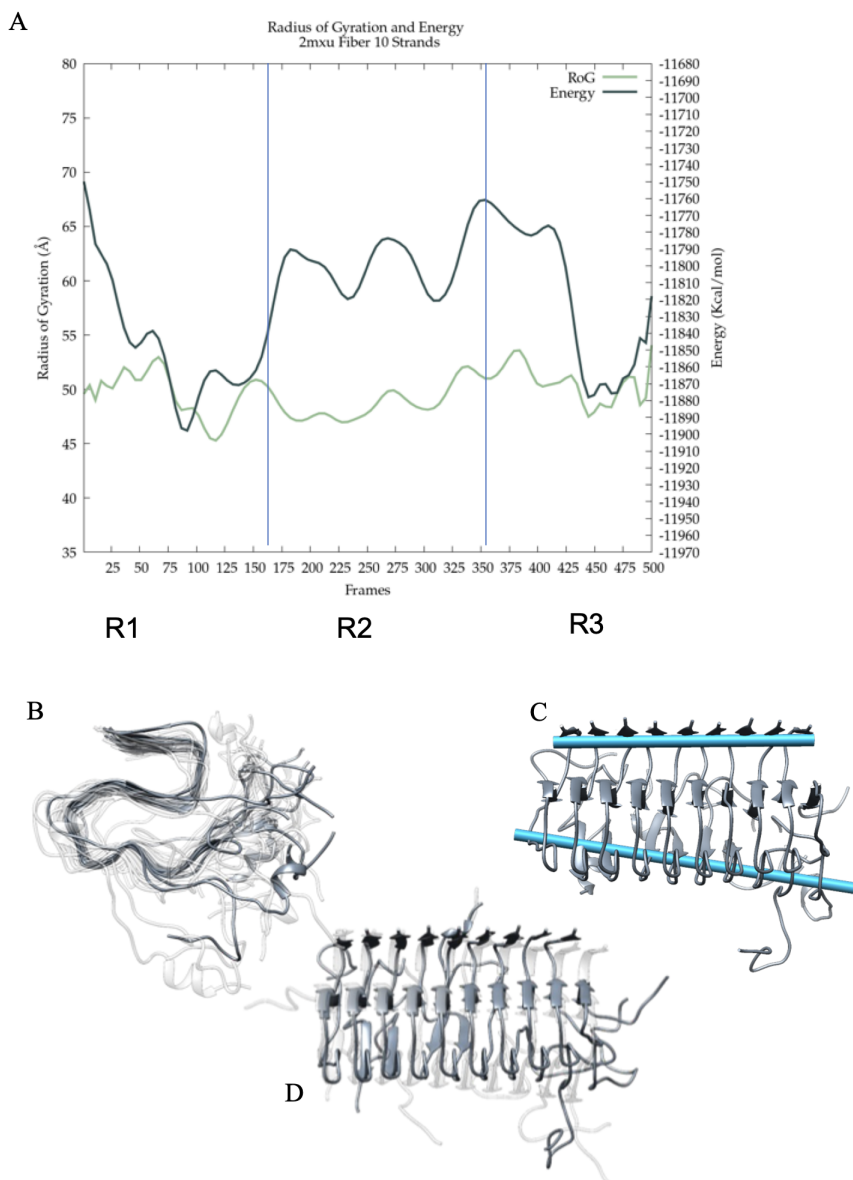
Figure A.28: Trajectory analyses for fibrillar Metal-Free E22G Complex. A. Implicit solvent Free Energy and Radius of Gyration along the three replicas. B. The most populated cluster (blue) is in superposition with the other 9 most populated clusters. C. In blue, the two axes are comprised of residues 39-41 (top) and residues 13-15 in β-sheets (bottom) of a representative structure of the most populated cluster. D. Overlap of the lowest energy structure with the initial structure.

Figure A.29: Trajectory analyses for fibrillar Metal-Free E22K Complex. A. Implicit solvent Free Energy and Radius of Gyration along the three replicas. B. The most populated cluster (blue) is in superposition with the other 9 most populated clusters. C. In blue, the two axes are comprised of residues 39-41 (top) and residues 13-15 in β-sheets (bottom) of a representative structure of the most populated cluster. D. Overlap of the lowest energy structure with the initial structure.
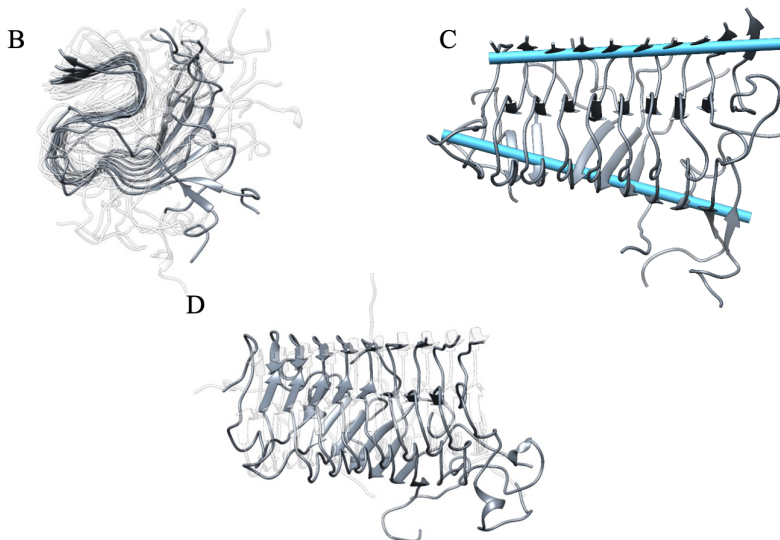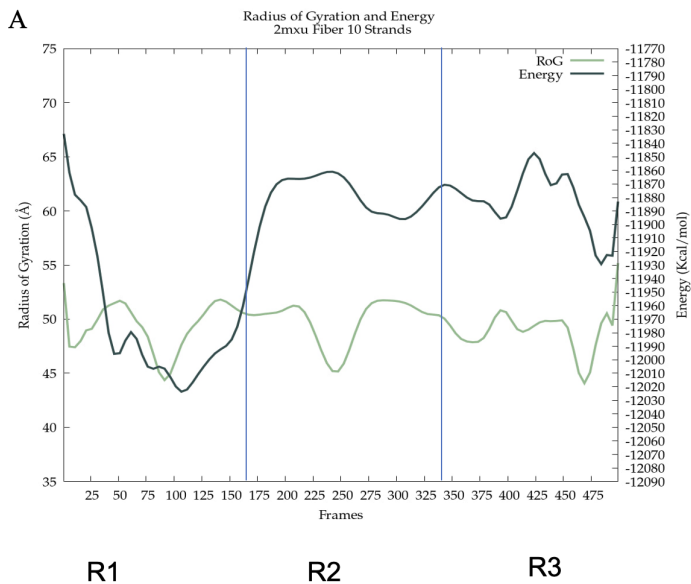
236

Figure A.30: Trajectory analyses for fibrillar Metal-Free D23N Complex. A. Implicit solvent Free Energy and Radius of Gyration along the three replicas. B. The most populated cluster (blue) is in superposition with the other 9 most populated clusters. C. In blue, the two axes are comprised of residues 39-41 (top) and residues 13-15 in β-sheets (bottom) of a representative structure of the most populated cluster. D. Overlap of the lowest energy structure with the initial structure.
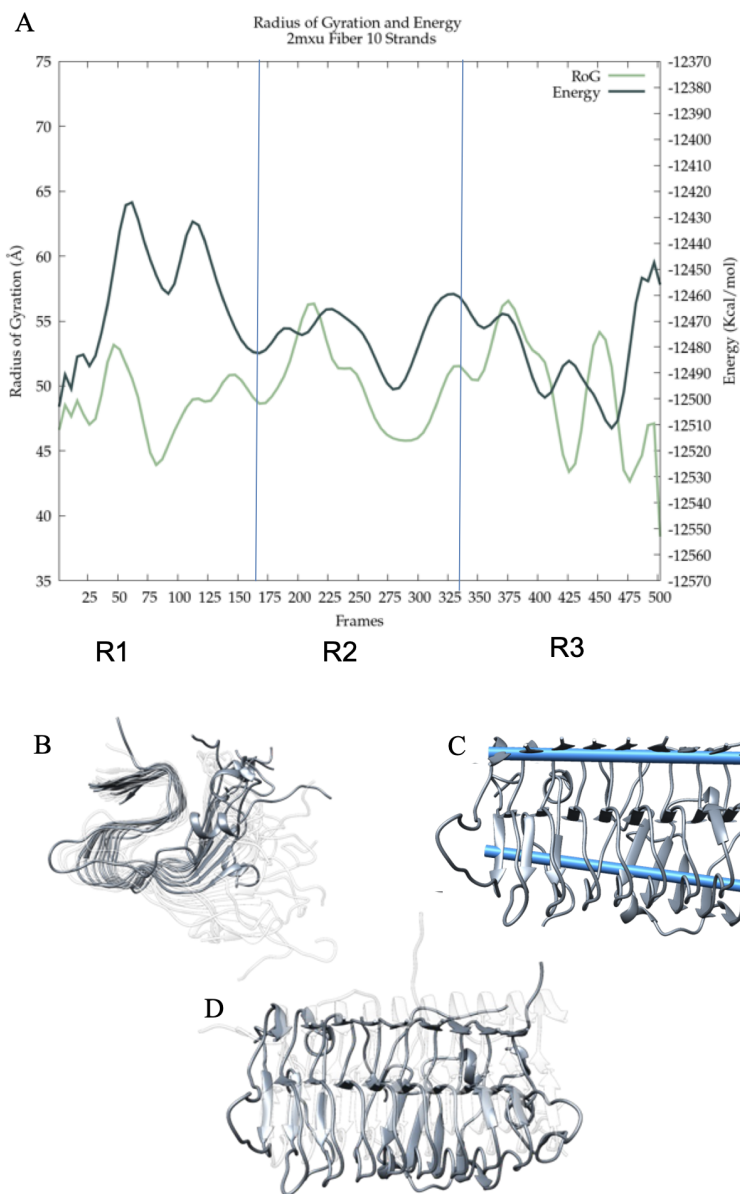
Figure A.31: Metal-Free complexes' measures of A. 3-Gly22 and 8-Gly22 (D_{HC}), B.3-Gln15 and 8-Gln15 measure (D_{HE}), C. Phe19-Val39 (D_{VC}), D. Glu11-Val39 measure (D_{VE}) for WT (light green), E22Q (light blue), E22G (dark green), E22K (orange) and D23N (purple) variants.
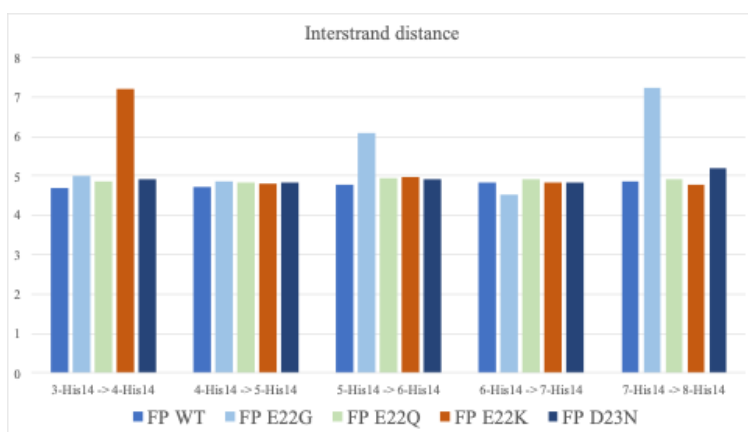
Figure A.32: Interstrand distances (D_{IS}) of metal-free complexes.

Figure A.33: Trajectory analyses for fibrillar Cu-bound E22Q Complex. A. Implicit solvent Free Energy and Radius of Gyration along the three replicas. B. The most populated cluster (blue) is in superposition with the other 9 most populated clusters. C. In blue, the two axes are comprised of residues 39-41 (top) and residues 13-15 in β-sheets (bottom) of a representative structure of the most populated cluster. D. Overlap of the lowest energy structure with the initial structure.
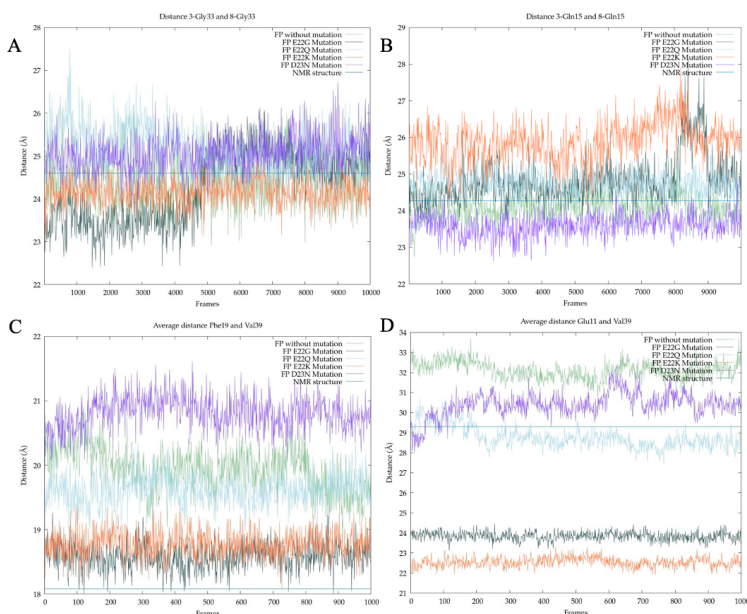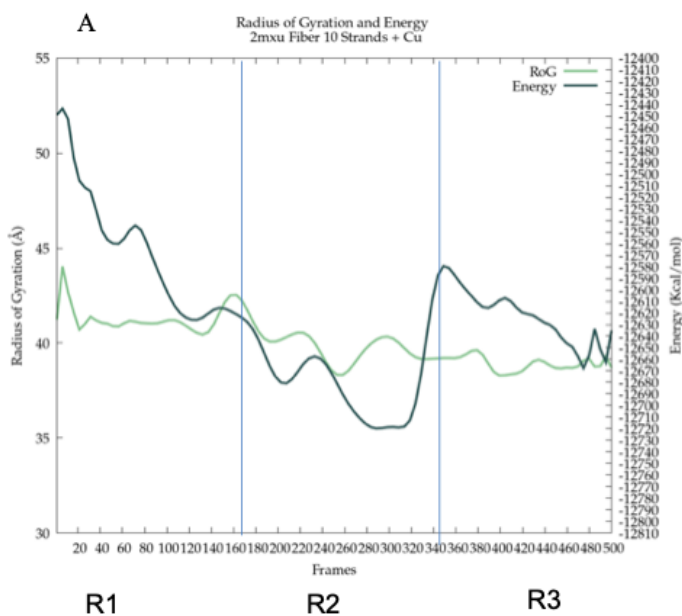
Figure A.34: Trajectory analyses for fibrillar Cu-bound E22G Complex. A. Implicit solvent Free Energy and Radius of Gyration along the three replicas. B. The most populated cluster (blue) is in superposition with the other 9 most populated clusters. C. In blue, the two axes are comprised of residues 39-41 (top) and residues 13-15 in β-sheets (bottom) of a representative structure of the most populated cluster. D. Overlap of the lowest energy structure with the initial structure.
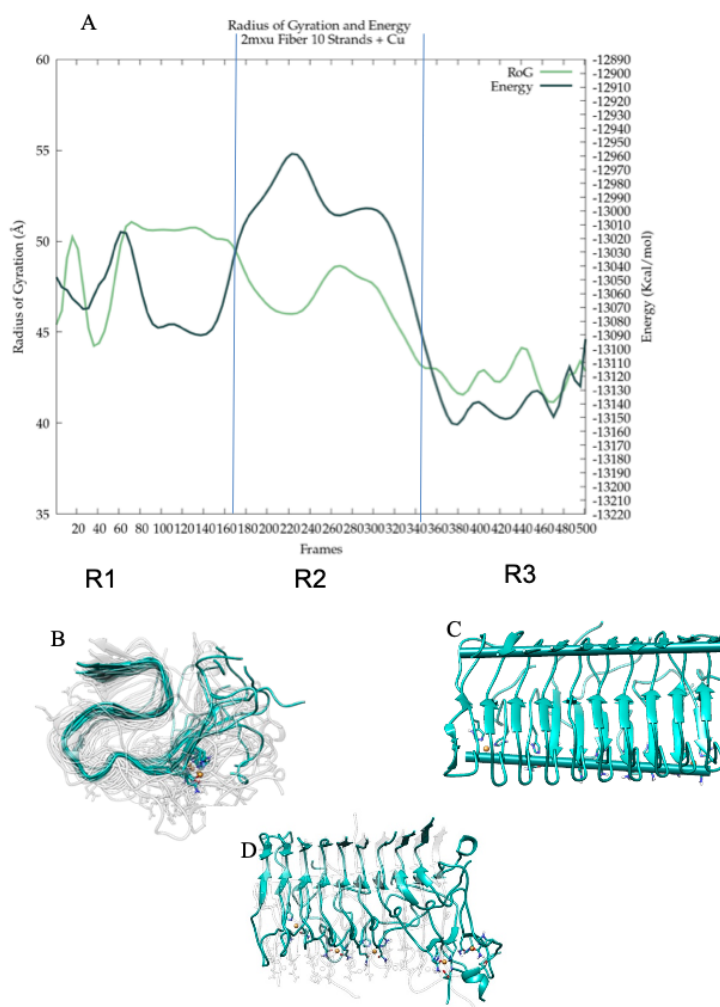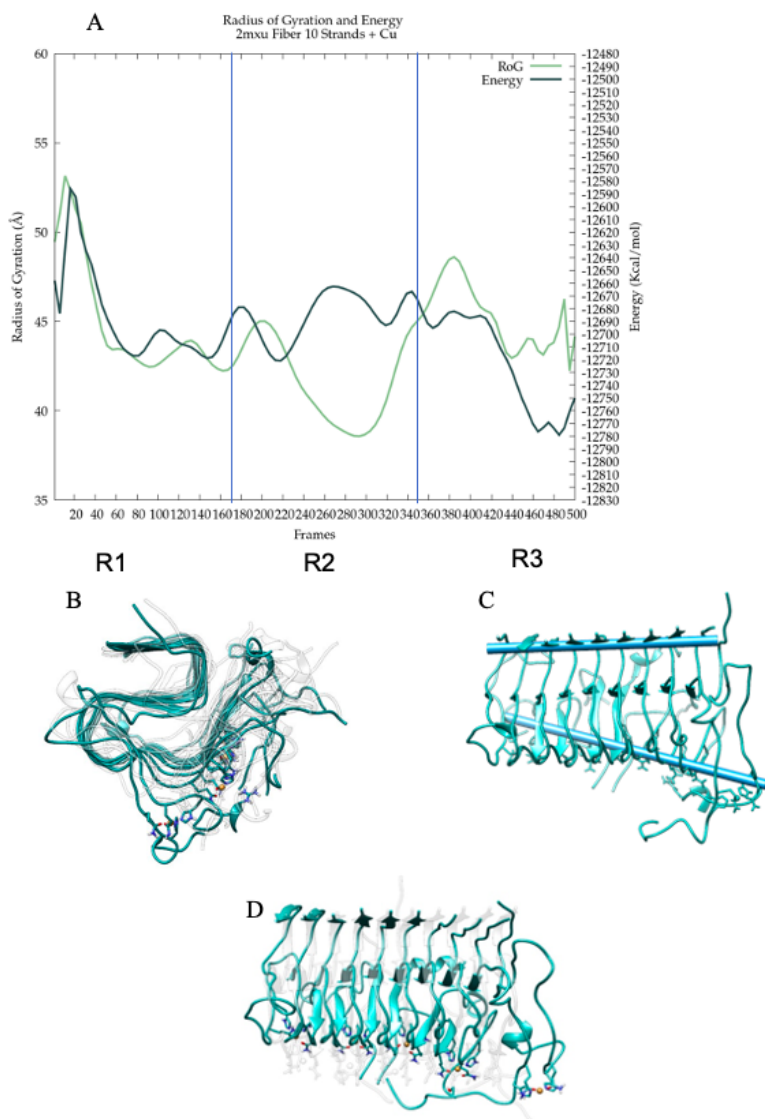
Figure A.35: Trajectory analyses for fibrillar Cu-bound E22K Complex. A. Implicit solvent Free Energy and Radius of Gyration along the three replicas. B. The most populated cluster (blue) is in superposition with the other 9 most populated clusters. C. In blue, the two axes are comprised of residues 39-41 (top) and residues 13-15 in β-sheets (bottom) of a representative structure of the most populated cluster. D. Overlap of the lowest energy structure with the initial structure.
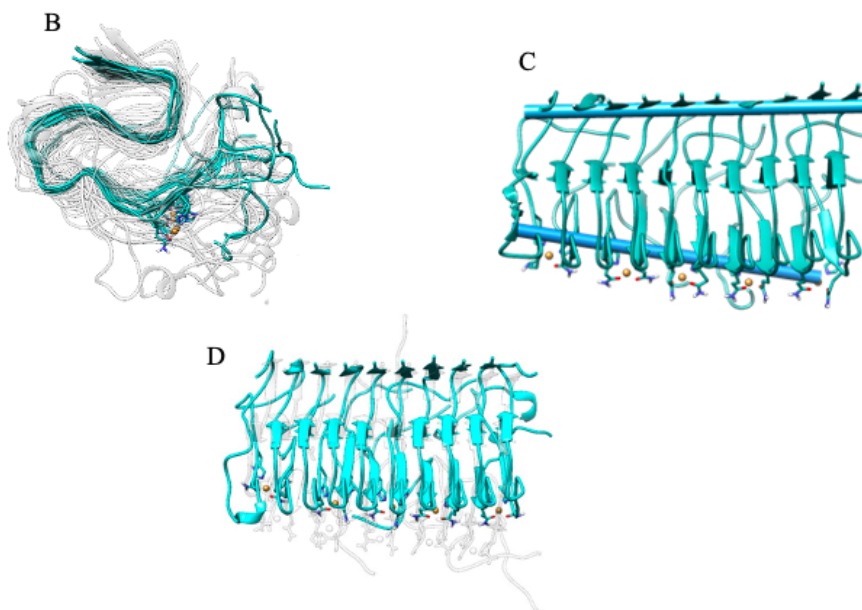
Figure A.36: Trajectory analyses for fibrillar Cu-bound D23N Complex. A. Implicit solvent Free Energy and Radius of Gyration along the three replicas. B. The most populated cluster (blue) is in superposition with the other 9 most populated clusters. C. In blue, the two axes are comprised of residues 39-41 (top) and residues 13-15 in β-sheets (bottom) of a representative structure of the most populated cluster. D. Overlap of the lowest energy structure with the initial structure.

242
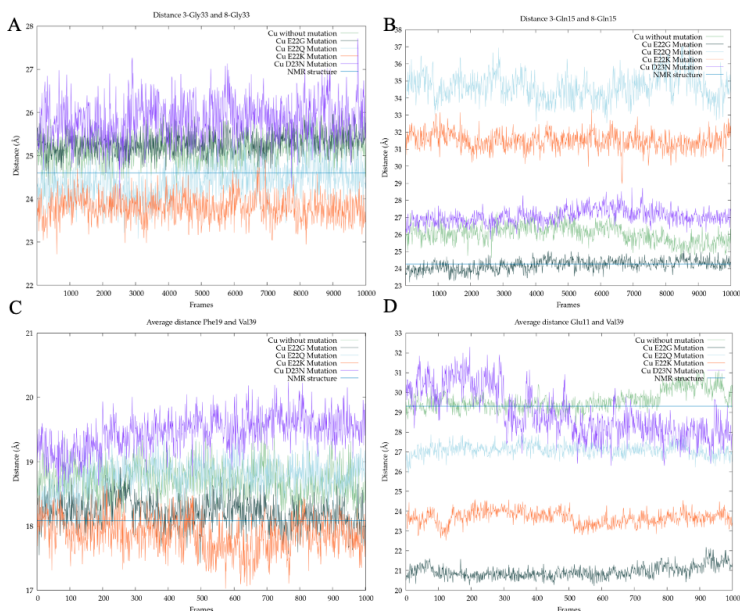
Figure A.37: Cu-bound complexes' measures of A. 3-Gly22 and 8-Gly22 ($\mathbf{D}_{HC}$), B.3-Gln15 and 8-Gln15 measure ($\mathbf{D}_{HE}$), C. Phe19-Val39 ($\mathbf{D}_{VC}$), D. Glu11-Val39 measure ($\mathbf{D}_{VE}$) for WT (light green), E22Q (light blue), E22G (dark green), E22K (orange) and D23N (purple) variants.

Figure A.38: Interstrand distances ($\mathbf{D}_{IS}$) of Cu-bound complexes.
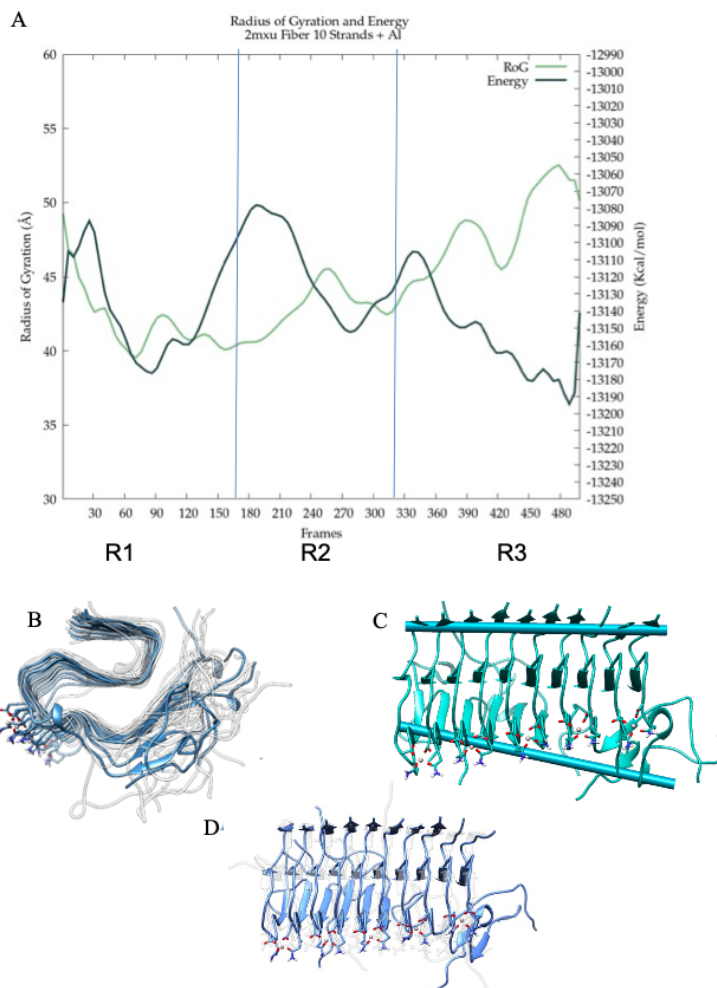
Figure A.39: Trajectory analyses for fibrillar Al-bound E22Q Complex. A. Implicit solvent Free Energy and Radius of Gyration along the three replicas. B. The most populated cluster (blue) is in superposition with the other 9 most populated clusters. C. In blue, the two axes are comprised of residues 39-41 (top) and residues 13-15 in β-sheets (bottom) of a representative structure of the most populated cluster. D. Overlap of the lowest energy structure with the initial structure.
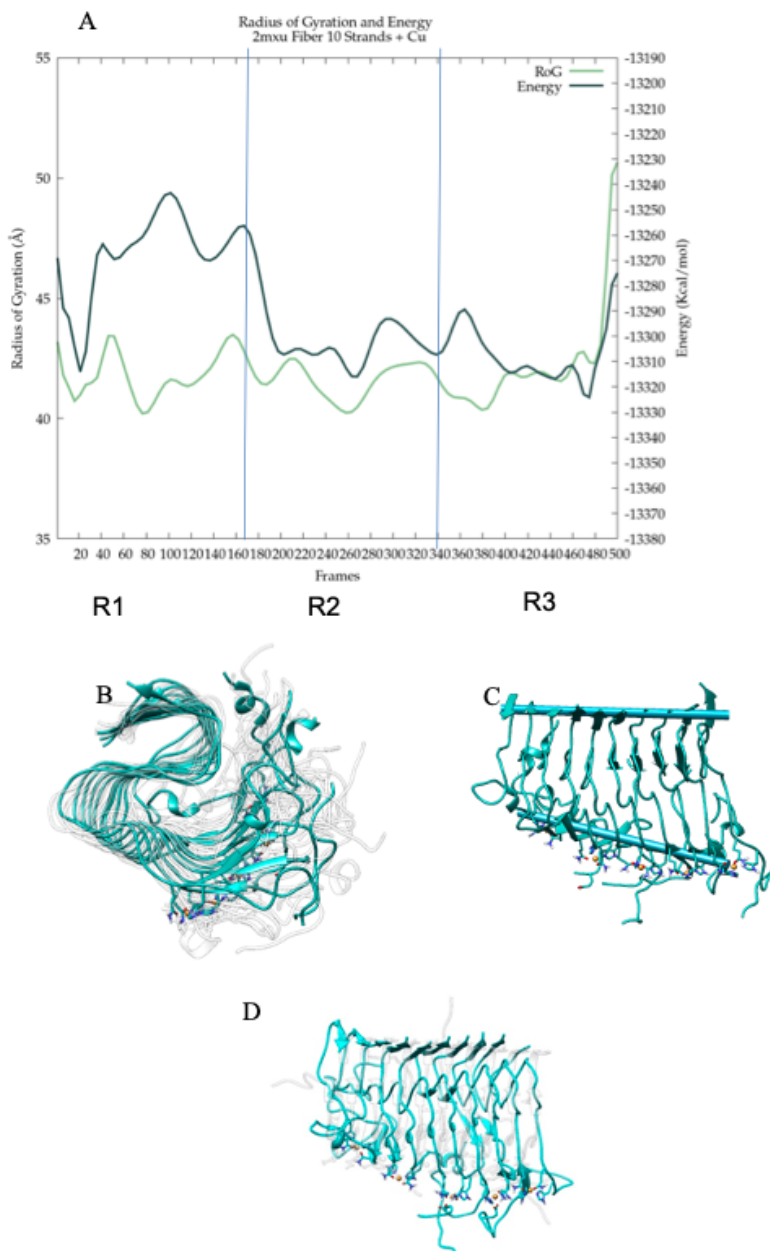
Figure A.40: Al-bound complexes' measures of A. 3-Gly22 and 8-Gly22 ($\mathbf{D}_{HC}$), B. 3-Gln15 and 8-Gln15 measure ($\mathbf{D}_{HE}$), C. Phe19-Val39 ($\mathbf{D}_{VC}$), D. Glu11-Val39 measure ($\mathbf{D}_{VE}$) for WT (light green), E22Q (light blue), E22G (dark green), E22K (orange) and D23N (purple) variants.
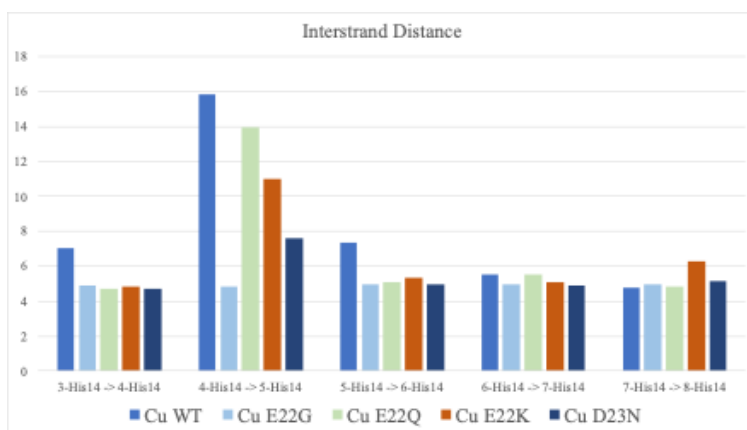
Figure A.41: Interstrand distances ($\mathbf{D}_{IS}$) of Al-bound complexes.

B

List of publications

B.1 Publications Resulting from this Thesis

- **Chapter 4**

  1. Impact of Cu(II) and Al(III) on the conformational landscape of Amyloidβ1-42. Roldán-Martín, L. et al. *PCCP*. (2021) DOI:10.1039/D1CP01561C

  2. Computational assessment of the impact of Cu(II) and Al(III) on β-amyloid42 fibrils: Binding sites, structural stability, and possible physiological implications. Roldán-Martín, L. et al. *Front Neurosci*. (2023) DOI:10.3389/fnins.2023.1110311

  3. Computational study of Aβ_{42} familial mutations and metal-interaction: Impact on monomers and aggregates dynamical behaviours. Roldán-Martín, L. et al. *Inor. Chem.* (Submitted)

- **Chapter 5**

  1. Amyloid Fibrils Formed by Short Prion-Inspired Peptides Are Metalloenzymes. Navarro, S. et al. *ACS Nano*. (2023) DOI:10.1021/acsnano.3c04164
     This work encloses a collaboration with the group of Salvador Ventura, who performed the experimental part.

- **Chapter 6**

  1. MetalPath: a Hybrid machine learning and knowledge-based approach to model metal diffusion in proteins. Fernández-Díaz, R. et al. *Under development*

  2. Metal recruitment mechanism as an evolutive fingerprint in copper tyrosinase. Roldán-Martín, L. et al. *Under development*

B.2   Other collaborations

1. **Dynamic Stereoselection of Peptide Helicates and Their Selective Labeling of DNA Replication Foci in Cells.** **Gómez-González, J.** *Angew. Chem.* **(2020) DOI:**10.1002/anie.202013039

Helicates are chiral species composed of one or more organic ligands which orientate around metal ions. Interesting properties have been attributed to helicates, such as antifreeze agents, amyloid aggregation inhibitors, and their binding capacity to three-way DNA, a common tumoral mark. In this study, experimental work performed by Eugenio Vázquez's group at CIQUS (Universidad de Santiago de Compostela, Spain) is supported by computational calculations performed in our group. In the computational part, MD simulations of the Fe(II) bound helicates were performed to retrieve a representative structure of the complex. Then, the helicates' most representative - likely, the more stable - structure obtained was docked inside the cavity of the three-way DNA, applying GaudiMM software. Two different isomers of the helicate were tested depending on their folding. Thanks to the docking, it could be observed that a certain flexibility of the DNA system is required to allow a successful insertion, which was better for the less voluminous isomer. Altogether, the atomic structure of the helicate was obtained, allowing it to demonstrate that peptide sequences containing bipyridine binding units can fold into three-stranded helicates in the presence of Fe(II). Such a complex can be successfully inserted in the cavity inside the three-way junction DNA, which is a key mark of cancer disease and may point to a potential anticancer effect of the helicates.

2. **BioMetAll: Identifying Metal-Binding Sites in Proteins from Backbone Preorganization.** Sánchez-Aparicio JE. *JCIM.* (2021) DOI:10.1021/acs.jcim.0c00827

Understanding how metallic moieties interact with biological systems is indispensable for properly defining metallosystems. Since the experimental definition of metal environments is still challenging, computational approaches are a suitable alternative. Thus, BioMetAll is a computational predictor of metal binding spheres, based on the geometric preorganization of the protein backbone. Its predictions are performed regardless of the lateral chain of the protein residues, a precious trait to apply to those incomplete or modelled structures. Three case studies were performed to illustrate BioMetAll's potential, being the study of diffusion pathways the one linked to the present Thesis work. Copper-binding proteins, such as hemocyanin, require leading copper from the solvent to the active binding site once the protein structure is organized. A series of hemocyanins were submitted to BioMetAll calculations in the present study, successfully predicting the di-copper binding site in all of them. Besides, other compatible but less affine binding sites were also detected, which may form a channel from solvent-exposed locations to the active binding site, suggesting that BioMetAll may not only be applicable for complete binding sites but also allow the detection of diffusion pathways. In summary, BioMetAll successfully predicts primary and secondary metal-binding sites, which could correspond to diffusion pathways inside proteins. Besides, its low computational demand allows us to easily apply the calculation over a series of structures, accounting for the conformational changes of the protein that could ultimately modify the metal-binding site.

3. **Controlling oncogenic KRAS signalling pathways with a Palladium-responsive peptide. Learte, S.** *Comm. Chem.* (2022) DOI:

RAS oncoproteins are a critical signal of cell proliferation and differentiation, whose mutations are related to the deadliest cancers. In the present study, a bis-histidine peptide derived from the cofactor SOS1 is demonstrated to bind KRAS with high affinity upon Pd(II) coordination. In fact, the Pd(II)-binding is determinant to maintain the required α-helical conformation. In order to support the experimental work performed by the group of Eugenio Vázquez, Gaussian accelerated Molecular Dynamics simulations were run both on the free and Pd(II)-coordinated form of the helical peptide. In the case of the metal-free peptide, the conformational sampling retrieves unfolded structures, completely losing the α-helix structure, as expected for an Intrinsically Disordered Peptide. On the contrary, upon Pd(II)-binding, the α-helical structure is consistently maintained, retrieving stable and organised structures. Accordingly, this work demonstrates that an Intrinsically Disordered Protein can be effectively stabilised upon metal binding. Once the effect of Pd(II) coordination was computationally determined, it was experimentally demonstrated that the short peptide derived from the SOS1, once coordinated to Pd(II), effectively binds and induces an inhibitory effect in KRAS protein.

References

1. Alberts, B. *et al. The Shape and Structure of Proteins* 4th Edition (Garland Science, New York, 2002).

2. Andreini, C., Bertini, I. & Rosato, A. Metalloproteomes: A Bioinformatic Approach. *Accounts of Chemical Research* **42,** 1471–1479 (Oct. 2009).

3. Smethurst, D. G. J. & Shcherbik, N. Interchangeable utilization of metals: New perspectives on the impacts of metal ions employed in ancient and extant biomolecules. *Journal of Biological Chemistry* **297,** 101374 (2021).

4. Carver, P. L. Metal ions and infectious diseases. An overview from the clinic. *Metal Ions in Life Sciences* **13,** 1–28 (2013).

5. Jomova, K. *et al.* Essential metals in health and disease. *Chemico-Biological Interactions* **367,** 110173 (Nov. 2022).

6. Lippard, S. J. & Berg, J. M. *Principles of bioinorganic chemistry* eng (University Science Books, Mill Valley, Calif, 1994).

7. Andreini, C., Bertini, I., Cavallaro, G., Holliday, G. L. & Thornton, J. M. Metal ions in biological catalysis: from enzyme databases to general principles. *JBIC* **13,** 1205–1218 (Nov. 2008).

8. Jaishankar, M., Tseten, T., Anbalagan, N., Mathew, B. B. & Beeregowda, K. N. Toxicity, mechanism and health effects of some heavy metals. *Interdisciplinary Toxicology* **7,** 60 (June 2014).

9. Valensin, D., Dell'Acqua, S., Kozlowski, H. & Casella, L. Coordination and redox properties of copper interaction with α-synuclein. *Journal of Inorganic Biochemistry* **163,** 292–300 (Oct. 2016).

10. Jellinger, K. A. The Relevance of Metals in the Pathophysiology of Neurodegeneration, Pathological Considerations. *Int Rev Neurobiol*, 1–47 (2013).

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

12. Toni, M., Massimino, M. L., De Mario, A., Angiulli, E. & Spisni, E. Metal Dyshomeostasis and Their Pathological Role in Prion and Prion-Like Diseases: The Basis for a Nutritional Approach. *Frontiers in Neuroscience* **11** (Jan. 2017).

13. Kontoghiorghes, G. J. Advances on Chelation and Chelator Metal Complexes in Medicine. *International Journal of Molecular Sciences* **21** (Apr. 2020).

14. Flora, S. J. & Pachauri, V. Chelation in Metal Intoxication. *International Journal of Environmental Research and Public Health* **7,** 2745 (2010).

15. Zoroddu, M. A. *et al*. The essential metals for humans: a brief overview. *J Inorg Biochem* **195** (2019).

16. Sánchez, M., Sabio, L., Gálvez, N., Capdevila, M. & Dominguez-Vera, J. M. Iron chemistry at the service of life. *IUBMB Life* **69,** 382–388 (June 2017).

17. Baker, H. M., He, Q. Y., Briggs, S. K., Mason, A. B. & Baker, E. N. Structural and functional consequences of binding site mutations in transferrin: Crystal structures of the Asp63Glu and Arg124Ala mutants of the N-lobe of human transferrin. *Biochemistry* **42,** 7084–7089 (2003).

18. Roelfes, G. Repurposed and artificial heme enzymes for cyclopropanation reactions. *Journal of Inorganic Biochemistry* **222,** 111523 (Sept. 2021).

19. Li, J. *et al*. Ferroptosis: past, present and future. *Cell Death & Disease 2020 11:2* **11,** 1–13 (Feb. 2020).

20. Chen, B. *et al*. Interactions between iron and α-synuclein pathology in Parkinson's disease. *Free radical biology & medicine* **141,** 253–260 (Sept. 2019).

21. Ayton, S. & Lei, P. Nigral iron elevation is an invariable feature of Parkinson's disease and is a sufficient cause of neurodegeneration. *BioMed Research International* (2014).

22. Bolognin, S. *et al*. Aluminum, copper, iron and zinc differentially alter amyloid-Aβ(1-42) aggregation and toxicity. *The international journal of biochemistry & cell biology* **43,** 877–885 (June 2011).

23. Rubino, J. T. & Franz, K. J. Coordination chemistry of copper proteins: How nature handles a toxic cargo for essential function. *Journal of Inorganic Biochemistry* **107,** 129–143 (Feb. 2012).

24. Veidis, M. V., Schreiber, G. H., Gough, T. E. & Palenik, G. J. Jahn-Teller Distortions in Octahedral Copper(II) Complexes. *Journal of the American Chemical Society* **91,** 1859–1860 (1969).

25. Chen, L., Min, J. & Wang, F. Copper homeostasis and cuproptosis in health and disease. *Signal Transduction and Targeted Therapy* **7,** 378 (Nov. 2022).

26. Wang, Y., Zhang, L. & Zhou, F. Cuproptosis: a new form of programmed cell death. *Cellular & Molecular Immunology* **19,** 867–868 (Apr. 2022).

27. Lelièvre, P., Sancey, L., Coll, J. L., Deniaud, A. & Busser, B. The Multifaceted Roles of Copper in Cancer: A Trace Metal Element with Dysregulated Metabolism, but Also a Target or a Bullet for Therapy. *Cancers* **12,** 1–25 (Dec. 2020).

28. Bandmann, O., Weiss, K. H. & Kaler, S. G. Wilson's disease and other neurological copper disorders. *The Lancet. Neurology* **14,** 103 (Jan. 2015).

29. Baldari, S., Rocco, G. D. & Toietta, G. Current Biomedical Use of Copper Chelation Therapy. *International Journal of Molecular Sciences* **21** (Feb. 2020).

30. Hernández-Camacho, J. D., Vicente-García, C., Parsons, D. S. & Navas-Enamorado, I. Zinc at the crossroads of exercise and proteostasis. *Redox biology* **35** (Aug. 2020).

31. Gupta, S., Brazier, A. K. M. & Lowe, N. M. Zinc deficiency in low- and middle-income countries: prevalence and approaches for mitigation. *Journal of Human Nutrition and Dietetics* **33,** 624–643 (Oct. 2020).

32. Laitaoja, M., Valjakka, J. & Jänis, J. Zinc Coordination Spheres in Protein Structures. *Inorganic Chemistry* **52,** 10983–10991 (Oct. 2013).

33. Bin, B. H., Seo, J. & Kim, S. T. Function, Structure, and Transport Aspects of ZIP and ZnT Zinc Transporters in Immune Cells. *Journal of Immunology Research* (2018).

34. Maywald, M. & Rink, L. Zinc in Human Health and Infectious Diseases. *Biomolecules 2022, Vol. 12, Page 1748* **12,** 1748 (Nov. 2022).

35. Maret, W. Zinc in Cellular Regulation: The Nature and Significance of "Zinc Signals". *International Journal of Molecular Sciences* **18** (2017).

36. Lindskog, S. Structure and mechanism of Carbonic Anhydrase. *Pharmacology and Therapeutics* **74,** 1–20 (1997).

37. Prasad, A. S. Zinc deficiency in human subjects. *Progress in Clinical and Biological Research* **129,** 1–33 (Jan. 1983).

38. D'Haese, P. C. in *Encyclopedia of Metalloproteins* (eds Kretsinger, R., Uversky, V. & Permyakov, E.) 47–53 (Springer, New York, NY, 2013).

39. Malacaria, L. *et al*. A Review on Coordination Properties of Al(III) and Fe(III) toward Natural Antioxidant Molecules: Experimental and Theoretical Insights. *Molecules* **26,** 2603 (Apr. 2021).

40. Rahimzadeh, M. R. *et al*. Aluminum Poisoning with Emphasis on Its Mechanism and Treatment of Intoxication. *Emergency Medicine International* **2022,** 1–13 (Jan. 2022).

41. Puentes-Díaz, N., Chaparro, D., Morales-Morales, D., Flores-Gaspar, A. & Alí-Torres, J. Role of Metal Cations of Copper, Iron, and Aluminum and Multifunctional Ligands in Alzheimer's Disease: Experimental and Computational Insights. *ACS Omega* **8,** 4508–4526 (Feb. 2023).

42. Crapper, D. R., Krishnan, S. S. & Dalton, A. J. Brain Aluminum Distribution in Alzheimer's Disease and Experimental Neurofibrillary Degeneration. *Science* **180,** 511–513 (May 1973).

43. Tomljenovic, L. Aluminum and Alzheimer's Disease: After a Century of Controversy, Is there a Plausible Link? *Journal of Alzheimer's Disease* **23,** 567–598 (Mar. 2011).

44. Islam, F. *et al*. Exposure of metal toxicity in Alzheimer's disease: An extensive review. *Frontiers in Pharmacology* **13,** 3068 (Aug. 2022).

45. Exley, C. Why Industry Propaganda and Political Interference Cannot Disguise the Inevitable Role Played by Human Exposure to Aluminum in Neurodegenerative Diseases, Including Alzheimer's Disease. *Frontiers in Neurology* **5** (2014).

46. O'Brien, R. J. & Wong, P. C. Amyloid precursor protein processing and Alzheimer's disease. *Annual review of neuroscience* **34,** 185–204 (2011).

47. Kiss, T. From coordination chemistry to biological chemistry of aluminium. *Journal of Inorganic Biochemistry* **128,** 156–163 (Nov. 2013).

48. Quintanar, L. & Rivillas-Acevedo, L. Studying metal ion-protein interactions: electronic absorption, circular dichroism, and electron paramagnetic resonance. *Methods in molecular biology* **1008,** 267–297 (2013).

49. Lancelin, J.-M., Grimaldi, M. & Shalev, D. E. Studying Peptide-Metal Ion Complex Structures by Solution-State NMR. *Int. J. Mol. Sci* **23** (2022).

50. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current protocols in bioinformatics* **54,** 1–5 (2016).

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

52. Cramer, F. Biochemical correctness: Emil Fischer's lock and key hypothesis, a hundred years after — an essay. *Pharmaceutica Acta Helvetiae* **69,** 193–203 (Mar. 1995).

53. Koshland, D. E. The Key–Lock Theory and the Induced Fit Theory. *Angewandte Chemie International Edition in English* **33,** 2375–2378 (Jan. 1995).

54. Lin, Y.-F. *et al.* MIB: Metal Ion-Binding Site Prediction and Docking Server. *Journal of Chemical Information and Modeling* **56,** 2287–2291 (Dec. 2016).

55. Sánchez-Aparicio, J. E. *et al.* BioMetAll: Identifying Metal-Binding Sites in Proteins from Backbone Preorganization. *Journal of Chemical Information and Modeling* **61,** 311–323 (Jan. 2021).

56. Doytchinova, I. Drug Design—Past, Present, Future. *Molecules* **27,** 1496 (Feb. 2022).

57. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **267,** 727–748 (Apr. 1997).

58. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and

Python Bindings. *Journal of Chemical Information and Modeling* **61,** 3891–3898 (Aug. 2021).

59. Sciortino, G. *et al.* DFT Protocol for EPR Prediction of Paramagnetic Cu(II) Complexes and Application to Protein Binding Sites. *Magnetochemistry* **4,** 55 (Dec. 2018).

60. Tian, C. *et al.* Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **16,** 528–552 (Jan. 2020).

61. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14,** 71–73 (Nov. 2016).

62. Song, D., Luo, R. & Chen, H. F. IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. *Journal of chemical information and modeling* **57,** 1166 (May 2017).

63. Miao, Y., Feher, V. A. & McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* **11,** 3584–3595 (Aug. 2015).

64. Qi, R., Wei, G., Ma, B. & Nussinov, R. Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example. *Methods in molecular biology (Clifton, N.J.)* **1777,** 101 (2018).

65. Bertini, L. *et al.* Quantum Mechanical Methods for the Investigation of Metalloproteins and Related Bioinorganic Compounds. *Methods Mol. Biol.* **1122,** 207–268 (2014).

66. Rufo, C. M. *et al.* Short peptides self-assemble to produce catalytic amyloids HHS Public Access. *Nat Chem* **6,** 303–309 (2014).

67. Groenhof, G. in *Biomolecular Simulations* (eds Monticelli, L. & Salonen, E.) 43–66 (Humana Press, Totowa, NJ, 2013).

68. Atkins, P. & de Paula, J. *Atkins' Physical chemistry* 8th, 430–468 (Oxford University Press, Great Britain, 2009).

69. Auletta, G., Fortunato, M. & Parisi, G. *Quantum mechanics* 1st (Cambridge University Press, United Kingdom, 2009).

70. Born, M. & Oppenheimer, R. Zur Quantentheorie der Molekeln. *Annalen der Physik* **389,** 457–484 (Jan. 1927).

71. Lewars, E. G. in *Computational Chemistry* 3rd, 9–49 (Springer Cham, 2016).

72. Hartree, D. R. in *Mathematical Proceedings of the Cambridge Philosophical Society* 1, 111–132 (Cambridge University Press, 1928).

73. Orio, M., Pantazis, D. A. & Neese, F. Density functional theory. *Photosynthesis Research* **102,** 443–453 (Nov. 2009).

74. Zhou, K. & Liu, B. Fundamentals of classical molecular dynamics simulation. *Molecular Dynamics Simulation*, 1–40 (Jan. 2022).

75. Hehre, J. *A Guide to Molecular Mechanics and Quantum Chemical Calculations* (Wavefunction, 2003).

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

77. Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry* **25,** 1656–1676 (Oct. 2004).

78. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational en-

ergetics and properties of organic liquids. *Journal of the American Chemical Society* **118,** 11225–11236 (Nov. 1996).

79. D.A. Case *et al*. *Amber 2018* (University of California, San Francisco, 2018).

80. Vanommeslaeghe, K. *et al*. CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* **31,** 671 (Mar. 2010).

81. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **115** (May 2018).

82. Unke, O. T. *et al*. Machine Learning Force Fields. *Chemical Reviews* **121,** 10142–10186 (Aug. 2021).

83. Hospital, A., Goñi, J. R., Orozco, M. & Gelpí, J. L. Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry : AABC* **8,** 37–47 (2015).

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

85. Lang, P. F. Applying the soft sphere model to improve the understanding of bonding in transition metals. *Heliyon* **6,** e03110 (Jan. 2020).

86. Li, P., Song, L. F., Merz, K. M. & Jr. Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water. *The journal of physical chemistry. B* **119,** 883–95 (Jan. 2015).

87. Li, P., Roberts, B. P., Chakravorty, D. K. & Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *Journal of Chemical Theory and Computation* **9,** 2733–2748 (June 2013).

88. Duarte, F. *et al*. Force Field Independent Metal Parameters Using a Nonbonded Dummy Model. *The Journal of Physical Chemistry B* **118,** 4351–4362 (Apr. 2014).

89. Rahnamoun, A. *et al*. A Polarizable Cationic Dummy Metal Ion Model. *J. Phys. Chem. Lett.* **13,** 5334–5340 (June 2022).

90. Seminario, J. Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem* **60,** 1271–1277 (1996).

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

92. Coumar, S. M. *Molecular Docking for Computer-Aided Drug Design* eng (Elsevier Science & Technology, San Diego, 2021).

93. Guedes, I. A., de Magalhães, C. S. & Dardenne, L. E. Receptor–ligand molecular docking. *Biophysical Reviews* **6,** 75 (Mar. 2014).

94. Huang, S. Y. & Zou, X. Advances and Challenges in Protein-Ligand Docking. *International Journal of Molecular Sciences* **11,** 3016 (Aug. 2010).

95. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins: Structure, Function and Genetics* **52,** 609–623 (Sept. 2003).

96. Hollingsworth, S. A. & Dror, R. O. Review Molecular Dynamics Simulation for All. *Neuron* **99,** 1129–1143 (2018).

97. Hug, S. Classical Molecular Dynamics in a Nutshell. *Methods Mol. Biol.* **924,** 127–152 (2013).

98. Fyta, M. Computational approaches in physics. *Computational Approaches in Physics*, 1–137 (Jan. 2016).

99. Kumar, H. & Maiti, P. K. in *Computational Statistical Physics* (eds Santra, S. & Ray, P.) 161–197 (Hindustan Book Agency, Gurgaon, 2011).

100. Elber, R., Ruymgaart, A. P. & Hess, B. SHAKE parallelization. *The European physical journal. Special topics* **200,** 211 (Nov. 2011).

101. Gonçalves, P. F. & Stassen, H. Calculation of the free energy of solvation from molecular dynamics simulations. *Pure and Applied Chemistry* **76,** 231–240 (2004).

102. Anandakrishnan, R., Drozdetski, A., Walker, R. C. & Onufriev, A. V. Speed of conformational change: Comparing explicit and implicit solvent molecular dynamics simulations. *Biophysical Journal* **108,** 1153–1164 (Mar. 2015).

103. Feig, M. in *Modeling Solvent Environments: Applications to Simulations of Biomolecules* 1–29 (John Wiley & Sons, Ltd, Jan. 2010).

104. Rapaport, D. C. in *The Art of Molecular Dynamics Simulation* 2nd, 11–43 (Cambridge University Press, Bar-Ilan University, Israel, July 2004).

105. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *The Journal of Physical Chemistry B* **113,** 6378–6396 (May 2009).

106. Chung, L. W. *et al.* The ONIOM Method and Its Applications. *Chemical Reviews* **115,** 5678–5796 (June 2015).

107. Alzheimer's Association. *2023 Alzheimer's disease facts and figures* tech. rep. 4 (Alzheimer's Association, Apr. 2023), 1598–1695.

108. World Health Organization. *Global action plan on the public health response to dementia 2017-2025* WHO (2017).

109. Lahiri, D. K. & Maloney, B. Beyond the signaling effect role of amyloid-β 42 on the processing of AβPP, and its clinical implications. *Exp Neurol*. **225,** 51–54 (2010).

110. Kepp, K. P. Bioinorganic Chemistry of Alzheimer's Disease. *Chem. Rev.* **112,** 5193–5239 (2012).

111. Cremades, N. & Dobson, C. M. The contribution of biophysical and structural studies of protein self-assembly to the design of therapeutic strategies for amyloid diseases. *Neurobiology of Disease* **109,** 178–190 (Jan. 2018).

112. Vivekanandan, S., Brender, J. R., Lee, S. Y. & Ramamoorthy, A. A partially folded structure of amyloid-beta(1–40) in an aqueous environment. *Biochemical and Biophysical Research Communications* **411,** 312–316 (July 2011).

113. Chen, G. F. *et al*. Amyloid beta: structure, biology and structure-based therapeutic development. *Acta Pharmacologica Sinica* **38,** 1205–1235 (July 2017).

114. Michaels, T. C. *et al*. Dynamics of oligomer populations formed during the aggregation of Alzheimer's Aβ42 peptide. *Nature Chemistry* **12,** 445–451 (May 2020).

115. Törnquist, M. *et al*. Secondary nucleation in amyloid formation. *Chemical Communications* **54,** 8667–8684 (Aug. 2018).

116. Kassler, K., Horn, A. H. & Sticht, H. Effect of pathogenic mutations on the structure and dynamics of Alzheimer's Aβ42-amyloid oligomers. *Journal of Molecular Modeling* **16,** 1011–1020 (2010).

117. Huang, Y. R. & Liu, R. T. The Toxicity and Polymorphism of β-Amyloid Oligomers. *International Journal of Molecular Sciences 2020, Vol. 21, Page 4477* **21,** 4477 (June 2020).

118. Ahmed, M. *et al*. Structural conversion of neurotoxic amyloid-β1–42 oligomers to fibrils. *Nature Structural & Molecular Biology* **17,** 561–567 (May 2010).

119. Paravastu, A. K., Leapman, R. D., Yau, W.-M. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proceedings of the National Academy of Sciences* **105,** 18349–18354 (Nov. 2008).

120. Lu, J.-X. *et al*. Molecular Structure of β-Amyloid Fibrils in Alzheimer's Disease Brain Tissue. *Cell* **154,** 1257–1268 (Sept. 2013).

121. Xiao, Y. *et al*. Aβ(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nature Structural & Molecular Biology* **22,** 499–505 (May 2015).

122. Wälti, M. A. *et al*. Atomic-resolution structure of a disease-relevant Aβ(1-42) amyloid fibril. *Proc Natl Acad Sci U S A* **113,** E4976–E4984 (Aug. 2016).

123. Colvin, M. T. *et al*. Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils. *Journal of the American Chemical Society* **138,** 9663–9674 (Aug. 2016).

124. Polanco, D., Carrancho, A., Gracia, P. & Cremades, N. Characterisation of Amyloid Aggregation and Inhibition by Diffusion-Based Single-Molecule Fluorescence Techniques. *Biophysica* **2,** 506–524 (Nov. 2022).

125. Seuma, M., Faure, A. J., Badia, M., Lehner, B. & Bolognesi, B. The genetic landscape for amyloid beta fibril nucleation accurately discriminates familial Alzheimer's disease mutations. *eLife* **10** (Feb. 2021).

126. Tcw, J. & Goate, A. M. Genetics of β-Amyloid Precursor Protein in Alzheimer's Disease. *Cold Spring Harbor Perspectives in Medicine* **7** (June 2017).

127. Everett, J. *et al*. Biogenic metallic elements in the human brain? *Science Advances* **7,** 6707–6716 (June 2021).

128. Mujika, J. I. *et al.* Elucidating the 3D structures of Al(III)–Aβ complexes: a template free strategy based on the pre-organization hypothesis. *Chemical Science* **8,** 5041–5049 (June 2017).

129. Alí-Torres, J., Maréchal, J.-D., Rodríguez-Santiago, L. & Sodupe, M. Three Dimensional Models of Cu2+-Aβ(1–16) Complexes from Computational Approaches. *Journal of the American Chemical Society* **133,** 15008–15014 (Sept. 2011).

130. Alí-Torres, J., Mirats, A., Maréchal, J.-D., Rodríguez-Santiago, L. & Sodupe, M. 3D Structures and Redox Potentials of Cu2+–Aβ(1–16) Complexes at Different pH: A Computational Study. *The Journal of Physical Chemistry B* **118,** 4840–4850 (May 2014).

131. Rezaei-Ghaleh, N., Giller, K., Becker, S. & Zweckstetter, M. Effect of Zinc Binding on β-Amyloid Structure and Dynamics: Implications for Aβ Aggregation. *Biophysical Journal* **101,** 1202 (Sept. 2011).

132. Yarjanli, Z., Ghaedi, K., Esmaeili, A., Rahgozar, S. & Zarrabi, A. Iron oxide nanoparticles may damage to the neural tissue through iron accumulation, oxidative stress, and protein aggregation. *BMC Neuroscience* **18,** 1–12 (June 2017).

133. Maynard, C. J., Bush, A. I., Masters, C. L., Cappai, R. & Li, Q. X. Metals and amyloid-β in Alzheimer's disease. *International Journal of Experimental Pathology* **86,** 147 (June 2005).

134. Hung, Y. H., Bush, A. I. & Cherny, R. A. Copper in the brain and Alzheimer's disease. *JBIC Journal of Biological Inorganic Chemistry 2009 15:1* **15,** 61–76 (Oct. 2009).

135. Abelein, A. *et al.* Molecular Structure of Cu(II)-Bound Amyloid-β Monomer Implicated in Inhibition of Peptide Self-Assembly in Alzheimer's Disease. *Journal of the American Chemical Society* **2,** 2571–2584 (Nov. 2022).

136. Mold, M., Ouro-Gnao, L., Wieckowski, B. M. & Exley, C. Copper prevents amyloid-β(1-42) from forming amyloid fibrils under near-physiological conditions in vitro. *Scientific reports* **3**, 1256 (2013).

137. Yang, J. *et al.* The double-edged role of copper in the fate of amyloid beta in the presence of anti-oxidants. *Chemical science* **8**, 6155–6164 (Sept. 2017).

138. Stahl, T., Falk, S., Taschan, H., Boschek, B. & Brunn, H. Evaluation of human exposure to aluminum from food and food contact materials. *European Food Research and Technology* **244**, 2077–2084 (Dec. 2018).

139. Mirza, A., King, A., Troakes, C. & Exley, C. Aluminium in brain tissue in familial Alzheimer's disease. *Journal of Trace Elements in Medicine and Biology* **40**, 30–36 (Mar. 2017).

140. Exley, C. & Clarkson, E. Aluminium in human brain tissue from donors without neurodegenerative disease: A comparison with Alzheimer's disease, multiple sclerosis and autism. *Scientific Reports* **10**, 1–7 (May 2020).

141. Mold, M., Linhart, C., Gómez-Ramírez, J., Villegas-Lanau, A. & Exley, C. Aluminum and Amyloid-β in Familial Alzheimer's Disease. *Journal of Alzheimer's Disease* **73**, 1627–1635 (Jan. 2020).

142. Pettersen, E. F. *et al.* UCSF Chimera: A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605–1612 (Oct. 2004).

143. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of chemical theory and computation* **11**, 3696–713 (Aug. 2015).

144. Jones, G., Willett, P. & Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of molecular biology* **245**, 43–53 (1995).

145. Rodríguez-Guerra Pedregal, J., Sciortino, G., Guasp, J., Municoy, M. & Maréchal, J.-D. GaudiMM: A modular multi-objective platform for molecular modeling. *Journal of Computational Chemistry* **38,** 2118–2126 (Sept. 2017).

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

147. Ramirez, L. S., Pande, J. & Shekhtman, A. Helical Structure of Recombinant Melittin. *The Journal of Physical Chemistry B* **123,** 356–368 (Jan. 2019).

148. Coppa, C., Bazzoli, A., Barkhordari, M. & Contini, A. Accelerated Molecular Dynamics for Peptide Folding: Benchmarking Different Combinations of Force Fields and Explicit Solvent Models. *Journal of Chemical Information and Modeling* **63,** 3030–3042 (May 2023).

149. Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **2,** 37–52 (Aug. 1987).

150. Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55,** 383–394 (May 2004).

151. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14,** 33–38 (1996).

152. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27,** 1575–1577 (June 2011).

153. Drew, S. C. & Barnham, K. J. The Heterogeneous Nature of Cu2+ Interactions with Alzheimer's Amyloid-β Peptide. *Accounts of Chemical Research* **44,** 1146–1155 (Nov. 2011).

154. Alí-Torres, J., Mirats, A., Maréchal, J. D., Rodríguez-Santiago, L. & Sodupe, M. Modeling Cu2+-Aβ complexes from computational approaches. *AIP Advances* **5** (Sept. 2015).

155. Liao, Q., Owen, M. C., Bali, S., Barz, B. & Strodel, B. Aβ under stress: the effects of acidosis, Cu¡sup¿2+¡/sup¿-binding, and oxidation on amyloid β-peptide dimers. *Chemical Communications* **54,** 7766–7769 (July 2018).

156. Liao, Q., Pabis, A., Strodel, B. & Kamerlin, S. C. L. Extending the Nonbonded Cationic Dummy Model to Account for Ion-Induced Dipole Interactions. *The journal of physical chemistry letters* **8,** 5408–5414 (Nov. 2017).

157. Mutter, S. T., Turner, M., Deeth, R. J. & Platts, J. A. Metal Binding to Amyloid-β 1–42: A Ligand Field Molecular Dynamics Study. *ACS Chemical Neuroscience* **9,** 2795–2806 (Nov. 2018).

158. Pham, D. Q. H., Li, M. S. & La Penna, G. Copper Binding Induces Polymorphism in Amyloid-β Peptide: Results of Computational Models. *Journal of Physical Chemistry B* **122,** 7243–7252 (July 2018).

159. Huy, P. D. Q., Vuong, Q. V., La Penna, G., Faller, P. & Li, M. S. Impact of Cu(II) Binding on Structures and Dynamics of Aβ42 Monomer and Dimer: Molecular Dynamics Study. *ACS Chemical Neuroscience* **7,** 1348–1363 (Oct. 2016).

160. Turner, M., Mutter, S. T., Kennedy-Britten, O. D. & Platts, J. A. Molecular dynamics simulation of aluminium binding to amyloid-β and its effect on peptide structure. *PLOS ONE* **14** (ed Permyakov, E. A.) e0217992 (June 2019).

161. Boopathi, S., Dinh Quoc Huy, P., Gonzalez, W., Theodorakis, P. E. & Li, M. S. Zinc binding promotes greater hydrophobicity in Alzheimer's Aβ42 peptide than copper binding: Molecular dynamics and solvation thermodynamics studies. *Proteins* **88,** 1285–1302 (Oct. 2020).

162. Khatua, P., Mondal, S. & Bandyopadhyay, S. Effects of Metal Ions on Aβ42 Peptide Conformations from Molecular Simulation Studies. *Journal of Chemical Information and Modeling* **59,** 2879–2893 (June 2019).

163. Wall, M. E., Calabró, G., Bayly, C. I., Mobley, D. L. & Warren, G. L. Biomolecular Solvation Structure Revealed by Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **141,** 4711–4720 (2019).

164. Exley, C., Price, N. C., Kelly, S. M. & Birchall, J. D. An interaction of beta-amyloid with aluminium in vitro. *FEBS letters* **324,** 293–295 (June 1993).

165. Wallin, C. *et al.* Metal ion coordination delays amyloid-β peptide self-assembly by forming an aggregation–inert complex. *Journal of Biological Chemistry* **295,** 7224–7234 (May 2020).

166. Groh, N. *et al.* Age-dependent protein aggregation initiates amyloid-β aggregation. *Frontiers in Aging Neuroscience* **9,** 138 (May 2017).

167. Almeida, Z. L. & Brito, R. M. Structure and Aggregation Mechanisms in Amyloids. *Molecules 2020, Vol. 25, Page 1195* **25,** 1195 (Mar. 2020).

168. Ke, P. C. *et al.* Half a century of amyloids: past, present and future. *Chemical Society Reviews* **49,** 5473–5509 (Aug. 2020).

169. Krishnamurthy, S., Sudhakar, S. & Mani, E. Kinetics of aggregation of amyloid β under different shearing conditions: Experimental and modelling analyses. *Colloids and Surfaces B: Biointerfaces* **209,** 112156 (Jan. 2022).

170. Rahman, A., Saikia, B., Gogoi, C. R. & Baruah, A. Advances in the understanding of protein misfolding and aggregation through molecular dynamics simulation. *Progress in biophysics and molecular biology* **175** (Nov. 2022).

171. Miller, Y. Molecular Insights into the Effect of Metals on Amyloid Aggregation. *Methods in Molecular Biology* **2340,** 121–137 (2022).

172. Kollmer, M. *et al*. Cryo-EM structure and polymorphism of Aβ amyloid fibrils purified from Alzheimer's brain tissue. *Nature Communications* **10,** 1–8 (Oct. 2019).

173. Liberta, F. *et al*. Cryo-EM fibril structures from systemic AA amyloidosis reveal the species complementarity of pathological amyloids. *Nature Communications* **10,** 1–10 (Mar. 2019).

174. Faller, P. Copper and zinc binding to amyloid-beta: coordination, dynamics, aggregation, reactivity and metal-ion transfer. *Chembiochem* **10,** 2837–2845 (2009).

175. Parthasarathy, S. *et al*. Molecular-Level Examination of Cu2+ Binding Structure for Amyloid Fibrils of 40-Residue Alzheimer's β by Solid-State NMR Spectroscopy. *Journal of the American Chemical Society* **133,** 3390–3400 (Mar. 2011).

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

177. Tiwari, S. P. *et al*. WEBnmat v2.0: Web server and services for comparing protein flexibility. *BMC Bioinformatics* **15,** 1–12 (Dec. 2014).

178. Yang, Y. *et al*. Cryo-EM structures of amyloid-b 42 filaments from human brains. *Science* **375,** 167–172 (Jan. 2022).

179. Innocenti, M. *et al*. Trace copper(II) or zinc(II) ions drastically modify the aggregation behavior of amyloid-beta1-42: an AFM study. *Journal of Alzheimer's disease* **19,** 1323–1329 (2010).

180. Thu, T. T. M. & Li, M. S. Protein aggregation rate depends on mechanical stability of fibrillar structure. *The Journal of chemical physics* **157** (Aug. 2022).

181. Lin, Y. S. & Pande, V. S. Effects of familial mutations on the monomer structure of Aβ42. *Biophysical Journal* **103,** L47–L49 (Dec. 2012).

182. Krone, M. G. *et al*. Effects of Familial Alzheimer's Disease Mutations on the Folding Nucleation of the Amyloid β-Protein. *Journal of Molecular Biology* **381,** 221–228 (Aug. 2008).

183. Ngo, S. T., Thu Phung, H. T., Vu, K. B. & Vu, V. V. Atomistic investigation of an Iowa Amyloid-β trimer in aqueous solution. *RSC Advances* **8,** 41705–41712 (Dec. 2018).

184. Yang, X. *et al*. On the role of sidechain size and charge in the aggregation of Aβ42 with familial mutations. *Proceedings of the National Academy of Sciences of the United States of America* **115,** E5849–E5858 (June 2018).

185. Lam, A. R., Teplow, D. B., Stanley, H. E. & Urbanc, B. Effects of the Arctic (E 22 fG) Mutation on Amyloid-Protein Folding: Discrete Molecular Dynamics Study. *JACS* **130,** 17413–17422 (2008).

186. Yoo, B. K., Xiao, Y., McElheny, D. & Ishii, Y. E22G Pathogenic Mutation of β-Amyloid (Aβ) Enhances Misfolding of Aβ40 by Unexpected Prion-like Cross Talk between Aβ42 and Aβ40. *Journal of the American Chemical Society* **140,** 2781–2784 (Feb. 2018).

187. Li, X., Lei, J., Qi, R., Xie, L. & Wei, G. Mechanistic insight into E22Q-mutation-induced antiparallel-to-parallel β-sheet transition of Aβ 1622 fibrils: an all-atom simulation study. *Physical Chemistry Chemical Physics* **21,** 15686–15694 (July 2019).

188. Okamoto, A., Yano, A., Nomura, K., Higai, S. & Kurita, N. Effect of D23N mutation on the dimer conformation of amyloid β-proteins: ab initio molecular simulations in water. *Journal of molecular graphics & modelling* **50,** 113–124 (2014).

189. Rezaei-Ghaleh, N., Amininasab, M., Giller, K. & Becker, S. Familial Alzheimer's Disease-Related Mutations Differentially Alter Stabil-

ity of Amyloid-Beta Aggregates. *Journal of Physical Chemistry Letters* **14,** 1427–1435 (Feb. 2023).

190. Van Nostrand, W. E., Melchor, J. P., Cho, H. S., Greenberg, S. M. & Rebeck, G. W. Pathogenic effects of D23N Iowa mutant amyloid beta -protein. *The Journal of biological chemistry* **276,** 32860–32866 (Aug. 2001).

191. Borreguero, J. M. *et al*. Folding events in the 21-30 region of amyloid β-protein (Aβ) studied in silico. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 6015 (Apr. 2005).

192. Hatami, A., Monjazeb, S., Milton, S. & Glabe, C. G. Familial Alzheimer's Disease Mutations within the Amyloid Precursor Protein Alter the Aggregation and Conformation of the Amyloid-β Peptide. *The Journal of Biological Chemistry* **292,** 3172 (Feb. 2017).

193. Boyle, J. *Lehninger principles of biochemistry* 4th ed (eds Nelson, D. & Cox, M.) **1,** 74–75 (Jan. 2005).

194. Duran-Meza, E. & Diaz-Espinoza, R. Catalytic Amyloids as Novel Synthetic Hydrolases. *International Journal of Molecular Sciences* **22,** 9166 (Aug. 2021).

195. Díaz-Caballero, M., Navarro, S., Fuentes, I., Teixidor, F. & Ventura, S. Minimalist Prion-Inspired Polar Self-Assembling Peptides. *ACS Nano* **12,** 5394–5407 (June 2018).

196. Arad, E., Baruch Leshem, A., Rapaport, H. & Jelinek, R. β-Amyloid fibrils catalyze neurotransmitter degradation. *Chem Catalysis* **1,** 908–922 (Sept. 2021).

197. Akagawa, K. in *Peptide Applications in Biomedicine, Biotechnology and Bioengineering* 513–564 (Elsevier, 2018).

198. Singh, N., Conte, M. P., Ulijn, R. V., Miravet, J. F. & Escuder, B. Insight into the esterase like activity demonstrated by an imidazole

appended self-assembling hydrogelator. *Chem. Comm.* **51,** 13213–13216 (2015).

199. Zheng, L., Marcozzi, A., Gerasimov, J. Y. & Herrmann, A. Conformationally Constrained Cyclic Peptides: Powerful Scaffolds for Asymmetric Catalysis. *Angewandte Chemie International Edition* **53,** 7599–7603 (July 2014).

200. Makhlynets, O. V., Gosavi, P. M. & Korendovych, I. V. Short Self-Assembling Peptides Are Able to Bind to Copper and Activate Oxygen. *Angewandte Chemie International Edition* **55,** 9017–9020 (July 2016).

201. Navarro, S. *et al.* Amyloid Fibrils Formed by Short Prion-Inspired Peptides Are Metalloenzymes. *ACS Nano* **17,** 16968–16979 (Sept. 2023).

202. Díaz-Caballero, M., Fernández, M. R., Navarro, S. & Ventura, S. Prion-based nanomaterials and their emerging applications. *Prion* **12,** 266–272 (Nov. 2018).

203. Nelson, R. *et al.* Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435,** 773–778 (June 2005).

204. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **447,** 453–457 (Apr. 2007).

205. Peccati, F. *et al.* Atomistic fibrillar architectures of polar prion-inspired heptapeptides. *Chemical Science* **11,** 13143–13151 (2020).

206. Peccati, F. & Sodupe, M. Atomistic insights into the structure of heptapeptide nanofibers. *The Journal of Chemical Physics* **155,** 055101 (Aug. 2021).

207. Arad, E. & Jelinek, R. Catalytic amyloids. *Trends in Chemistry* **4,** 907–917 (Oct. 2022).

208. Park, H. *et al.* Designed Amyloid Fibers with Emergent Melanosomal Functions. *Langmuir* **38,** 7077–7084 (June 2022).

209. Zozulia, O. & Korendovych, I. V. Semi-Rationally Designed Short Peptides Self-Assemble and Bind Hemin to Promote Cyclopropanation. *Angewandte Chemie International Edition* **59,** 8108–8112 (May 2020).

210. Song, R. *et al*. Principles Governing Catalytic Activity of Self-Assembled Short Peptides. *Journal of the American Chemical Society* **141,** 223–231 (Jan. 2019).

211. Lee, M. *et al*. Zinc-binding structure of a catalytic amyloid from solid-state NMR. *Proceedings of the National Academy of Sciences* **114,** 6191–6196 (June 2017).

212. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **132,** 154104 (Apr. 2010).

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

214. S. Fernandes, H., Ramos, M. J. & M. F. S. A. Cerqueira, N. molUP: A VMD plugin to handle QM and ONIOM calculations using the gaussian software. *Journal of Computational Chemistry* **39,** 1344–1353 (July 2018).

215. Becke, A. D. Densityfunctional thermochemistry. III. The role of exact exchange Density-functional thermochemistry. III. The role of exact exchange. *Citation: J. Chem. Phys* **98,** 5648 (1993).

216. Petersson, G. A. & Al-Laham, M. A. A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *The Journal of Chemical Physics* **94,** 6081–6090 (May 1991).

217. Fuentealba, P., Preuss, H., Stoll, H. & Von Szentpály, L. A proper account of core-polarization with pseudopotentials: single valence-

electron alkali compounds. *Chemical Physics Letters* **89,** 418–422 (July 1982).

218. Verpoorte, J. A., Mehta, S. & Edsall, J. T. Esterase activities of human carbonic anhydrases B and C. *Journal of Biological Chemistry* **242,** 4221–4229 (Sept. 1967).

219. Lindskog, S. & Silverman, D. N. The catalytic mechanism of mammalian carbonic anhydrases. *EXS*, 175–195 (2000).

220. Wei, Y., Tinoco, A., Steck, V., Fasan, R. & Zhang, Y. Cyclopropanations via Heme Carbenes: Basic Mechanism and Effects of Carbene Substituent, Protein Axial Ligand, and Porphyrin Substitution. *Journal of the American Chemical Society* **140,** 1649–1662 (Feb. 2018).

221. Guo, C., Cheng, M. & Gross, M. L. Protein-Metal-Ion Interactions Studied by Mass Spectrometry-Based Footprinting with Isotope-Encoded Benzhydrazide. *Analytical chemistry* **91,** 1416 (Jan. 2019).

222. Sánchez-Aparicio, J. E. *et al*. GPathFinder: Identification of Ligand-Binding Pathways by a Multi-Objective Genetic Algorithm. *International journal of molecular sciences* **20** (July 2019).

223. Palm-Espling, M. E., Niemiec, M. S. & Wittung-Stafshede, P. Role of metal in folding and stability of copper proteins in vitro. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823,** 1594–1603 (Sept. 2012).

224. Alonso-Cotchico, L., Pedregal, J. R. G., Lledós, A. & Maréchal, J. D. The Effect of Cofactor Binding on the Conformational Plasticity of the Biological Receptors in Artificial Metalloenzymes: The Case Study of LmrR. *Frontiers in chemistry* **7** (2019).

225. Matoba, Y. *et al*. A molecular mechanism for copper transportation to tyrosinase that is assisted by a metallochaperone, caddie protein. *Journal of Biological Chemistry* **286,** 30219–30231 (Aug. 2011).

226. Rodríguez-Guerra, J. P., Alonso-Cotchico, L., Velasco-Carneros, L. & Maréchal, J. OMMProtocol: A Command Line Application to Launch Molecular Dynamics Simulations with OpenMM. *PrePrint* (2018).

227. Fernández-Díaz, R., Roldán-Martín, L., Sánchez-Aparicio, J. & Maréchal, J. Hybrid machine learning and knowledge-based approach to modelling metal diffusion processes in proteins. *Under development*.

228. Wang, Y., Davis, I., Shin, I., Xu, H. & Liu, A. Molecular Rationale for Partitioning between C-H and C-F Bond Activation in Heme-Dependent Tyrosine Hydroxylase. *Journal of the American Chemical Society* **143,** 4680–4693 (Mar. 2021).

229. Lai, X., Wichers, H. J., Soler-Lopez, M. & Dijkstra, B. W. Structure of Human Tyrosinase Related Protein1 Reveals a Binuclear Zinc Active Site Important for Melanogenesis. *Angewandte Chemie* **129,** 9944–9947 (Aug. 2017).

230. Carballo-Carbajal, I. *et al*. Brain tyrosinase overexpression implicates age-dependent neuromelanin production in Parkinson's disease pathogenesis. *Nature Communications* **10,** 1–19 (Mar. 2019).

231. Cabrera-Valladares, N. *et al*. Expression of the melA gene from Rhizobium etli CFN42 in Escherichia coli and characterization of the encoded tyrosinase. *Enzyme and Microbial Technology* **38,** 772–779 (Apr. 2006).

232. Esposito, R. *et al*. New Insights into the Evolution of Metazoan Tyrosinase Gene Family. *PLoS ONE* **7,** 35731 (Apr. 2012).

233. Locasale, J. W. Allovalency revisited: An analysis of multisite phosphorylation and substrate rebinding. *Journal of Chemical Physics* **128** (Mar. 2008).

234. Tiessler-Sala, L. *et al*. Getting Deeper into the Molecular Events of Heme Binding Mechanisms: A Comparative Multi-level Computa-

tional Study of HasAsm and HasAyp Hemophores. *Inorganic Chemistry* **61,** 17068–17079 (Oct. 2022).

235. Maiti, B. K. *et al*. Unusual Reduction Mechanism of Copper in Cysteine-Rich Environment. *Inorganic Chemistry* **57,** 8078–8088 (July 2018).

236. Deschamps, P. *et al*. Copper Complexation by Amino Acid: L-Glutamine-Copper(II)-L-Histidine Ternary System. *Journal of Trace and Microprobe Techniques* **21,** 729–741 (Nov. 2003).