Politecnico di Torino

Politecnico di Torino

Master's degree in Biomedical Engineering

Computational investigation of Amyloid Precursor Protein Familial mutants and their interaction with F-spondin in Alzheimer's Disease

Supervisor:
Prof. Jacek Adam Tuszynski

Candidate:
Rosaria Rita Arrotuto

Academic year 2024/2025

July 2025

Contents

List of Figures

List of Tables

ABSTRACT

Alzheimer's disease (AD) is a neurodegenerative disease and represents the most common form of dementia, characterized by a progressive deterioration of memory and cognitive functions. It mainly affects the elderly population (late onset) but can also occur at a younger age (early onset), especially in the presence of genetic mutations. According to the amyloid hypothesis, a central role is played by the amyloid precursor protein (APP), which is cleaved by β- and γ-secretase through the amyloidogenic pathway, generating the Aβ protein, in particular Aβ42, a form more prone to aggregation and neurotoxic. Familial APP mutations, such as the Swedish (KM670/671NL), Iowa (D694N) and Iberian (I716F) mutations, alter APP processing, favoring the pathological accumulation of Aβ. A possible strategy to counteract this mechanism is the use of ligands capable of binding to APP and preventing the initial cleavage by β-secretase. In this context, F-spondin has proven to be a potential natural inhibitor: it interacts with the extracellular E2 domain of APP and negatively regulates its amyloidogenic processing.

In this thesis, computational molecular modeling methods were used to study the interactions between different variants of the APP protein and the F-Spondin protein, with the aim of investigating the molecular mechanisms involved in AD. Starting from the amino acid sequences obtained from the UniProt database, AlphaFold Server was employed to predict the three-dimensional structure of APP, its mutated variants (Swedish, Iowa and Iberian) and the F-spondin segment with reelin and spondin domains. To improve their stability and optimize their conformation, preliminary molecular dynamics (MD) simulations were performed on all protein models. Molecular Docking was then performed, via the ClusPro server, between F-spondin and each APP variant, generating four distinct molecular complexes: wild-type APP/F-spondin, Swedish APP (KM670/671NL)/F-spondin, Iowa APP (D694N)/F-spondin and Iberian APP (I716F)/F-spondin. The four complexes obtained through docking were subjected to molecular dynamics simulations. A comparative analysis was then conducted by examining the global stability (RMSD), the system's flexibility (RMSF), and the interface stability between the two proteins in each complex, based on the number of hydrogen bonds and the contact surface. Finally, the binding free energy was estimated to assess the binding affinity across the different complexes.

1

INTRODUCTION

1.1 What is Alzheimer's disease?

Alzheimer's Disease (AD) is a neurodegenerative disorder that represents the most common form of dementia, and it causes neuronal death principally in the cortex and hippocampus.[1] It is characterized by progressive loss of memory and cognitive abilities, leading to dementia, especially in the elderly (late onset) and, less frequently, in young adults (early onset). The progression of the disease severely impairs daily life, with personality changes, problematic behavior, and difficulty with social interaction. Symptoms include agitation, isolation, and loss of memory and skills, often culminating in mortality caused by complications such as infection or malnutrition.[2]

AD is traditionally defined by the combined presence of extracellular deposits of amyloid β (Aβ) and intracellular aggregates of hyperphosphorylated tau, elements that form amyloid plaques and neurofibrillary tangles, respectively. In recent decades, numerous experimental evidences have confirmed that these accumulations are closely associated with the neurodegenerative processes typical of the pathology. However, more recent studies have highlighted the complexity of the disease, suggesting that it cannot be explained by a simple linear causal model. In fact, the role of multiple factors, genetic, environmental and related to aging, which interact with the main molecular mechanisms of the pathology is increasingly recognized. The classic amyloid hypothesis considered the accumulation of Aβ as the initial and triggering event, followed by alterations of the tau protein. Today, however, it is plausible that Aβ and tau act through parallel pathways, influencing each other and jointly contributing to neurodegeneration.

The clinical diagnosis of AD is based on anamnesis, neuropsychological assessment and observation of the evolution of symptoms over time. The typical form manifests itself with disturbances in memory and executive functions, while atypical forms, often associated with early onset, may begin with visual, linguistic or motor symptoms. In recent years, the introduction of biomarkers in the cerebrospinal fluid (CSF) and advanced imaging techniques, such as PET and magnetic resonance imaging (MRI), has

revolutionized the diagnostic criteria, allowing the pathology to be identified even in the preclinical phase.

The main risk factors include genetic ones, together with elements related to lifestyle, age and vascular conditions.

- The main genetic risk factor for sporadic Alzheimer's disease is the APOE $\varepsilon 4$ allele. The APOE4 variant is associated with increased accumulation of beta-amyloid (Aβ) in the brain. APOE plays a key role in the maintenance and repair of the central nervous system by interacting with several brain receptors, including VLDL-R, LDL-R, and LRP. Structural alterations in these receptors, particularly in LRP, can compromise the metabolism of the APP protein, favoring excessive production of Aβ. Mutations in the APP, PSEN1, and PSEN2 genes are responsible for familial forms of Alzheimer's disease. Other genes, such as TREM2, are also involved in regulating the immune response and in the removal of amyloid, contributing to the modulation of disease risk.
- Another area of great interest concerns the role of infections and neuroinflammation. Chronic brain infections can trigger a persistent inflammatory response, which contributes to the onset and progression of the disease.
- Finally, several modifiable lifestyle factors have a significant impact on the risk of developing dementia. These include hypertension, diabetes, physical inactivity, poor diet, low education, and social isolation. Addressing these issues through prevention and health promotion strategies could significantly reduce the incidence of the disease in the general population.

Globally, dementia affects approximately 40 million people, mostly over 60, and this number is expected to double every 20 years until 2050, with a faster increase in developing countries. Early-onset forms are rare, and under the age of 50 the prevalence is less than 1 in 4000; approximately 30% of these cases are attributable to Alzheimer's. European and US studies have shown a decline in the age-specific incidence of dementia in recent decades, probably due to better control of vascular risk factors. In Asia, a shift from vascular dementia to Alzheimer's disease as the prevalent form is observed, reflecting demographic and diagnostic changes.

In this context, prevention emerges as a top priority. Acting on modifiable factors such as obesity, inactivity, depression, smoking and poor diet could reduce the incidence of the disease by up to 30%. However, a significant proportion of cases remains that will require the development of more effective therapeutic strategies. Currently, treatments have only symptomatic effects (e.g. cholinesterase inhibitors and memantine), while experimental therapies aim to intervene in the early phase, integrating pharmacological, cognitive, nutritional and cardiovascular approaches. [3, 4]

There are several hypotheses for AD development: cholinergic hypothesis, tau hypothesis, amyloid cascade hypothesis, and recently a vascular hypothesis has been developed.[1]

1.1.1 The amyloid cascade hypothesis

The amyloid cascade hypothesis has been the predominant hypothesis on AD for over 25 years.[5] According to this hypothesis, the Alzheimer's Disease is mainly caused by the aggregation of the protein beta-amyloid into amyloid plaques, triggering a cascade leading to the formation of neurofibrillary tangles, neuronal death, vascular damage and dementia.[6] In addition to the evidence of amyloid plaques in the brains of AD patients, mutations in the APP gene, which encodes the amyloid precursor protein, have been identified in families with hereditary forms of AD (FAD); these mutations alter the metabolism of APP and favour the production of toxic forms of Aβ. Furthermore, it has been observed that trisomy 21 (Down syndrome), which involves an extra copy of the APP gene, leads to the neuropathology of AD.[7]

The amyloid precursor protein (APP) is processed through two pathways: a non-amyloidogenic and an amyloidogenic one.

- The non-amyloidogenic pathway is mediated by α-secretase and does not produce Aβ.
- The amyloidogenic pathway involves β-secretase and γ-secretase and leads to the production of Aβ. There are several isoforms of Aβ, including Aβ40 and Aβ42. Normally, about 90% of secreted Aβ peptides are of the Aβ40 type, a soluble form that slowly transforms into an insoluble β-sheet configuration, making it easily eliminated from the brain. In contrast, about 10% of secreted Aβ peptides are of the Aβ42 type, which are highly fibrillogenic and lead to the formation of toxic plaques.[8]

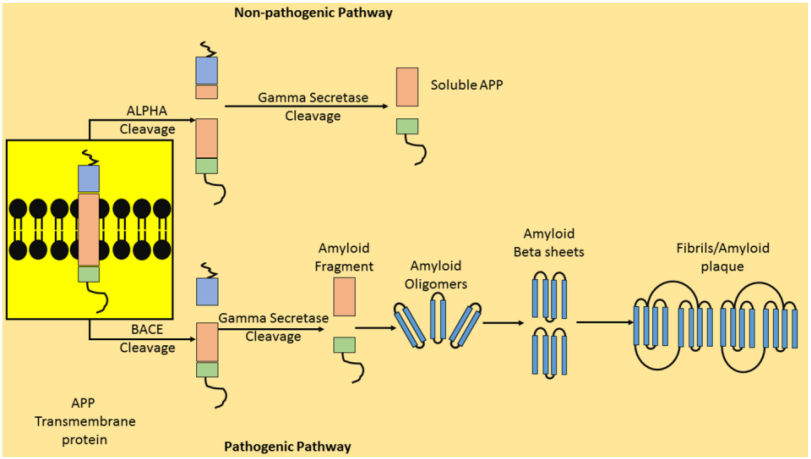


Figure 1: *The Amyloid Cascade Hypothesis. The two processing pathways of APP: the non-amyloidogenic pathway, which prevents Aβ formation, and the amyloidogenic pathway, leading to Aβ production, aggregation, and plaque formation. Source: Behl T et al, 2020. https://doi.org/10.3390/ijms21207443*

The amyloid hypothesis has stimulated research into therapies for AD that aim to reduce the production, aggregation or deposition of Aβ. Among therapeutic approaches, those based on decreasing Aβ production include β-secretase inhibition, γ-secretase inhibition or α-secretase activation. Approaches based on inhibition of Aβ oligomerisation or fibrillation and approaches for Aβ degradation have also been developed:

- Inhibition of β-secretase: β-secretase is a key enzyme in the production of Aβ. Several β-secretase inhibitors have been developed and tested in clinical trials, but few of these have reached phase I clinical trials due to problems such as poor ability to cross the blood-brain barrier and lack of specificity for β-secretase.

- Inhibition of γ-secretase: γ-secretase is another enzyme involved in the production of Aβ. Inhibitors of γ-secretase present significant problems, including interference with other important physiological pathways, such as the Notch signalling pathway.

- Activation of α-secretase: α-secretase cleaves APP in a way that does not allow Ab production. Increasing α-secretase activity is therefore considered a promising therapeutic approach, but the lack of clear information on the cellular mechanisms of α-secretase cleavage presents a challenge for the development of effective therapies.

- Inhibitors of Aβ oligomerisation or fibril formation: several molecules that interfere with the process of oligomer and fibril formation have been identified. However, many of these compounds have shown limited clinical efficacy.

- Degradation and clearance of Aβ: several enzymes are capable of degrading Aβ, including endothelin-converting enzyme (ECE), angiotensin-converting enzyme (ACE) and neprilysin. Increasing the activity of these enzymes or administering monoclonal antibodies that bind to Aβ have been proposed as therapeutic strategies. However, even these approaches have shown limited clinical results.[9]

Although the amyloid hypothesis has been the focus of research for years, all attempts to develop Aβ-targeted drugs to treat AD have failed. It has been observed that in mouse models genetically modified to overexpress Aβ, senile plaques formed but not neurofibrillary tangles due to tau accumulation and no neuronal death was observed. This led to the idea that extracellular accumulation of Aβ fibrils is not intrinsically cytotoxic and that Aβ does not induce tau accumulation. Recent studies indicate that the main factor underlying the development and progression of AD is tau, not Aβ.

The amyloid hypothesis has been reconsidered and Kametani and Hasegawa (2018) proposed that the initiation of AD is closely linked to alterations in APP metabolism and the accumulation of C-terminal fragments of APP. It has been observed that mutations in Presenilin-1 (PS1), a constitutive protein of the γ-secretase complex, reduce γ-secretase activity, resulting in decreased production of Aβ40 and increasing the Aβ42/Aβ40 ratio. The same thing happens with different types of mutations in the APP gene. At the same time, the C-terminal fragments of APP that should be cleaved by γ-secretase are not cleaved and accumulated in the cell membrane. This accumulation has been linked to neurotoxicity and altered vesicular trafficking. Through experiments on transgenic mice expressing the C-terminal intracellular domain of APP, it was observed that they developed AD-like symptoms, such as the accumulation of phosphorylated tau and memory impairment. [5]

Changes in Aβ metabolism
Increase in total Aβ production Increase in the Aβ42/Aβ40 ratio Reduced Aβ clearance
Oligomerization of Aβ42 and initial deposition
Subtle effects of soluble Aβ42 oligomers on synaptic function
Inflammatory response (microglial and astrocytic activation) and amyloid plaque formation
Progressive synaptic/neuronal injury
Altered neuronal ionic homeostasis and oxidative injury
Aberrant oligomerization and hyperphosphorylation of tau
Widespread neuronal dysfunction and cell death
Dementia with plaque and tangle pathology

Table 1: *The amyloid cascade hypothesis (Haass and Selkoe, 2007).*

1.1.2 The tau hypothesis

The tau hypothesis focuses on tau protein, a microtubule-associated protein (MAP) that plays a crucial role in stabilizing microtubule assembly within neurons, through its isoforms and phosphorylation. According to the tau hypothesis, Alzheimer's disease is caused by abnormal hyperphosphorylation of the tau protein. This process leads to an accumulation of the protein by forming neurofibrillary tangles (NFTs), which accumulate within neurons. Hyperphosphorylated tau loses the ability to bind to microtubules and thus stabilize them, also impairing the function of the cytoskeleton. Microtubule dysfunction interferes with axonal transport, impairing communication between neurons and leading to neuronal death.[8]
Scientific research has led to the development of therapeutic strategies that could slow down or even reverse disease progression. These therapeutic strategies are mainly focused on:

- inhibition of tau phosphorylation, to block the activity of kinases, i.e. the enzymes responsible for tau phosphorylation;
- inhibition of tau aggregation, preventing the formation of toxic tau oligomers and filaments;
- disaggregation of tau filaments, using molecules, such as quinoline derivatives, capable of binding and disaggregating tau oligomers.

In addition, research is also focusing on the development of specific biomarkers to visualise tau aggregates, so that they can be used for the early diagnosis of AD and to monitor the effectiveness of treatments.[10]

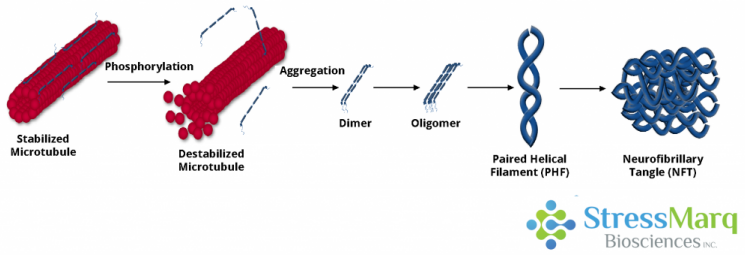


Figure 2: *Tau hypothesis. Tau dissociates from microtubules, microtubule destabilization, tau aggregates into oligomers, paired helical filaments, neurofibrillary tangles formation (Brenda Knox, 2022, stressmarq.com)*

1.1.3 The cholinergic hypothesis

The cholinergic hypothesis was born in the mid-1970s and is the first theory on the pathogenesis of Alzheimer's disease. According to the cholinergic hypothesis, degeneration of cholinergic neurons in the brain, with a reduction in cholinergic neurotransmission, plays a crucial role in the cognitive decline characteristic of this disorder. Such degeneration occurs mainly in the basal forebrain, resulting in neurotransmission deficits in the cerebral cortex and other brain regions. Several therapeutic approaches have been developed, such as cholinesterase inhibitors and choline precursor. In recent years, "second-generation" cholinesterase inhibitors have been developed, such as donepezil, rivastigmine, metrifonate, and galantamine. Through clinical trials, it has been shown that these cholinesterase inhibitors can improve cognition, global function, and slow symptomatic decline in patients with Alzheimer's disease; they can alleviate behavioral symptoms such as agitation, apathy, hallucinations, and abnormal motor behaviors.[11] However, it should be emphasized that the cholinergic hypothesis does not completely explain Alzheimer's disease. Other factors, such as beta-amyloid (Aβ) accumulation, hyper phosphorylation of tau protein, neuroinflammation, and vascular dysfunction, contribute significantly to the development and progression of the disease.[12]

1.1.4 The vascular hypothesis

The vascular hypothesis of Alzheimer's disease (VHAD) suggests that the trigger of Alzheimer's disease is a chronic reduction in cerebral blood flow (CBF). The hypothesis proposes that the initial pathological event leading to a neuronal energetic crisis, and thus to neurodegeneration, is chronic cerebral hypoperfusion. Consequently, what may also increase the risk of Alzheimer's disease are vascular risk factors such as hypertension, atherosclerosis, type 2 diabetes, smoking, obesity and heart disease (all of which may contribute to cerebral hypoperfusion). According to this hypothesis, there is a critical threshold of cerebral hypoperfusion, called CATCH. CATCH is a precursor to cognitive decline, oxidative stress and neurodegeneration, which ultimately leads to Alzheimer's disease. CATCH is the onset of CBF insufficiency that reaches a critical threshold, in which cerebral haemodynamic deterioration increases due to an imbalance between CBF supply and neuronal demand; thus, a slow and progressive decrease in neuronal efficacy occurs. Achieving CATCH impairs neuron-astroglial metabolism, limiting the supply of high-energy nutrients to the brain, such as oxygen and glucose. This leads to an ischaemic-hypoxic state, which can also induce Aβ formation in the brain.[13]

2

BIOLOGICAL BACKGROUND

2.1 APP: Amyloid Precursor Protein

The amyloid precursor protein (APP) is a type I transmembrane protein, and plays a crucial role in Alzheimer's disease [14] since it's considered the main protein involved in the disease according to the amyloid hypothesis.[6] This protein is encoded by a single gene, and three major isoforms resulting from alternative splicing have been characterized: APP695, APP751, and APP770. [15]
APP can be cleaved by enzymes as α and β-secretase, generating the extracellular fragments, respectively sAPPα and sAPPβ, and the carboxyterminal fragments (CTF), α-CTF (CTF83) and β-CTF (CTF99). Subsequently γ-secretase cleaves the C-terminal fragments, within the transmembrane domain, leading to the generation of the amino-terminal APP intracellular domain (AICD), and to the release of the small peptide p3 (from CTF83) or Aβ (from CTF99).[14, 16]

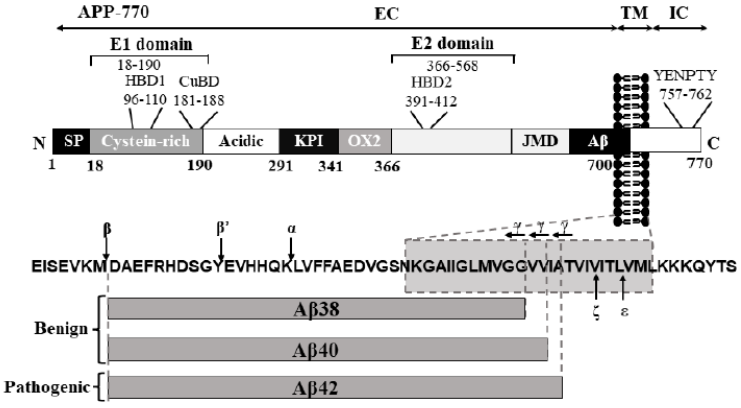


Figure 3: *Structure of APP: APP770 structure with its three domains and Aβ peptide fragment. [15]*

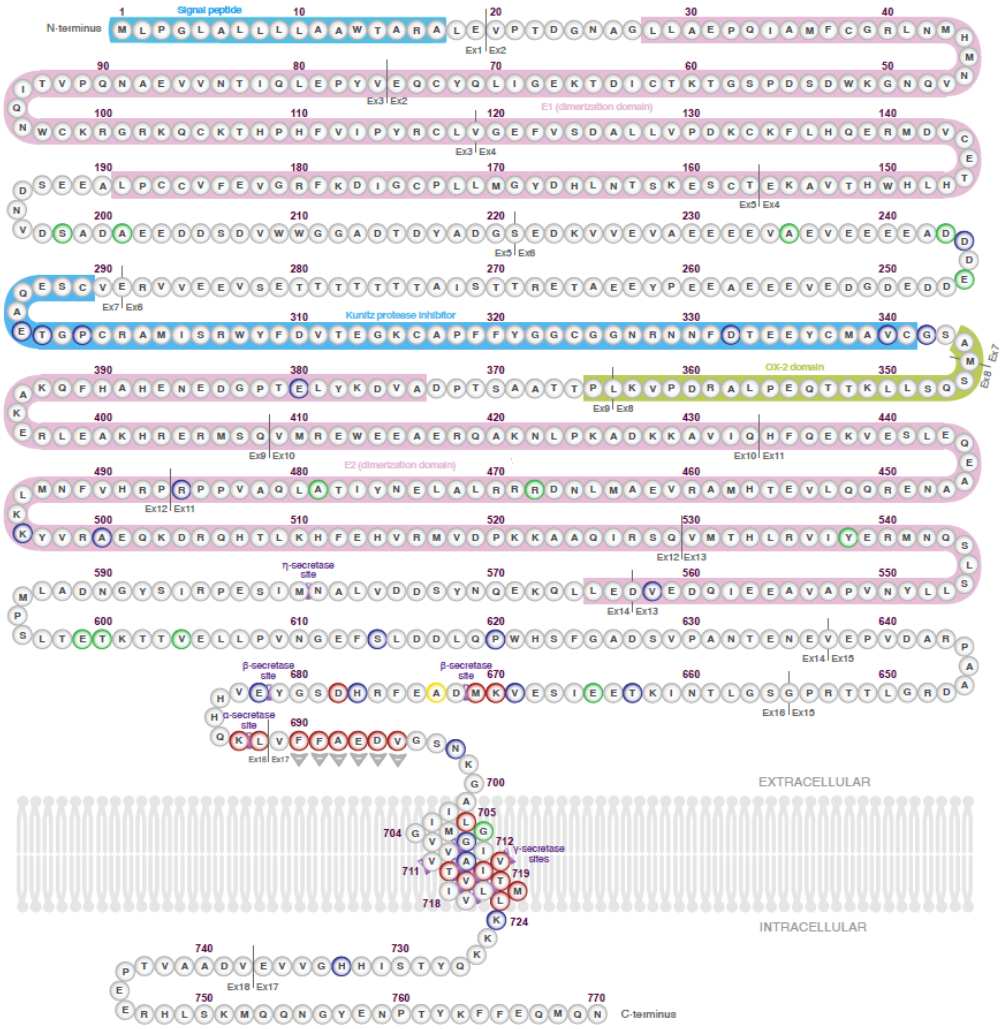


Figure 4: *Schematic of Amyloid Precursor Protein (APP). Variants named according to amino acid positions in isoform 1 (Uniprot: P05067), with 770 amino acids https://www.alzforum.org/mutations/app*

APP is composed of three main domains: extracellular domain (EC), transmembrane domain (TM) and intracellular domain (IC).

The extracellular domain of APP is composed of several subdomains with distinct structural features:

- N-terminal signal peptide (SP).
- E1 domain is broken into 2 regions: HBD1 and CuBD.
  - HBD1 (heparin-binding domain): this region can bind to heparin and other molecules called glycosaminoglycans. It contains a β-sheet and a flexible loop. HBD1 helps promote neurite growth (the development of neuron extensions) and has a strongly positive surface, which makes it suitable for interacting with negatively charged glycosaminoglycans.

  - CuBD (copper-binding domain): this section is able to bind copper ions. Structurally, it includes a small β-sheet and an α-helix. It forms a hydrophobic pocket located right next to HBD1, which could act as a site for protein interaction or dimer formation. Besides copper, it might also interact with other metal ions.

- The KPI (Kunitz protease inhibitor) domain is present in specific APP isoforms, such as APP751 and APP770. In Alzheimer's disease, its expression is increased, and this may influence the production of Aβ peptides. This increased expression has been linked to disruption of metabolic enzymes and mitochondrial dysfunction, observed in AD. Moreover, mutations in the KPI domain of APP751 can lead the protein to be trapped in the endoplasmic reticulum (ER), which promotes increased Aβ production.

- The OX-2 domain is believed to participate in cell surface binding and recognition processes.

- The E2 domain includes both a heparin-binding site (HBD2) and a metal-binding site. It also features the REMRS motif, which promotes neuronal outgrowth and cell proliferation, and allows interaction with membrane-bound heparan sulfate proteoglycans (HSPGs).

- The juxtamembrane domain (JMD), located just before the transmembrane region of APP, contains a GxxxG motif, which facilitates APP dimerization and interactions with other transmembrane proteins.

- The transmembrane domain of APP is a single-pass α-helix that spans the lipid bilayer. This helix is amphipathic, with hydrophobic residues oriented toward the membrane and hydrophilic residues exposed to both the cytoplasm and extracellular space. It plays a key role in APP's membrane stability and positioning, and is also important for its interactions with other membrane proteins.

- The Aβ sequence is distributed across both the extracellular/juxtamembrane region and the transmembrane domain of APP. γ-secretase cleaves APP within the membrane, generating Aβ peptides ranging from 39 to 42 amino acids. Among these, Aβ42 is less soluble and more prone to aggregation. Aβ can form oligomers, protofibrils, fibrils, and eventually plaques, which are key pathological features of Alzheimer's disease (AD). Although Aβ is constantly produced in the brain, its aggregation and deposition typically begin in the hippocampus and entorhinal cortex, marking the early stages of AD pathology.

- The APP intracellular domain (AICD) is situated next to the Aβ sequence and contains phosphorylation sites and a YENPTY motif, which mediates interactions with intracellular signaling proteins. AICD is generated following γ-secretase cleavage and results from both amyloidogenic and non-amyloidogenic processing. It has been shown to promote cell death, increase tau phosphorylation, suppress neuronal activity, and disturb calcium homeostasis, all of which contribute to neuronal dysfunction. Elevated AICD levels found in the postmortem brains of AD patients suggest that AICD accumulation may be involved in AD pathogenesis. [15]

After APP is synthesized, the N-terminal signal peptide (SP) fragment is removed. It is then processed in the Golgi and transported to the plasma membrane, where it is mainly localized. The intracellular fate of APP depends on its subcellular localization, as it is processed by different enzymes depending on where it is located.

As mentioned in the section of amyloid cascade hypothesis, there are two main pathways:

- The non-amyloidogenic APP processing pathway involves cleavages by α- and γ-secretases, with the generation of sAPPα, p3 and AICD50 (cleaved from CTF83);
- The amyloidogenic APP processing pathway involves cleavages by β- and γ-secretases, with the generation of sAPPβ, Aβ and AICD (cleaved from CTF99). [16]

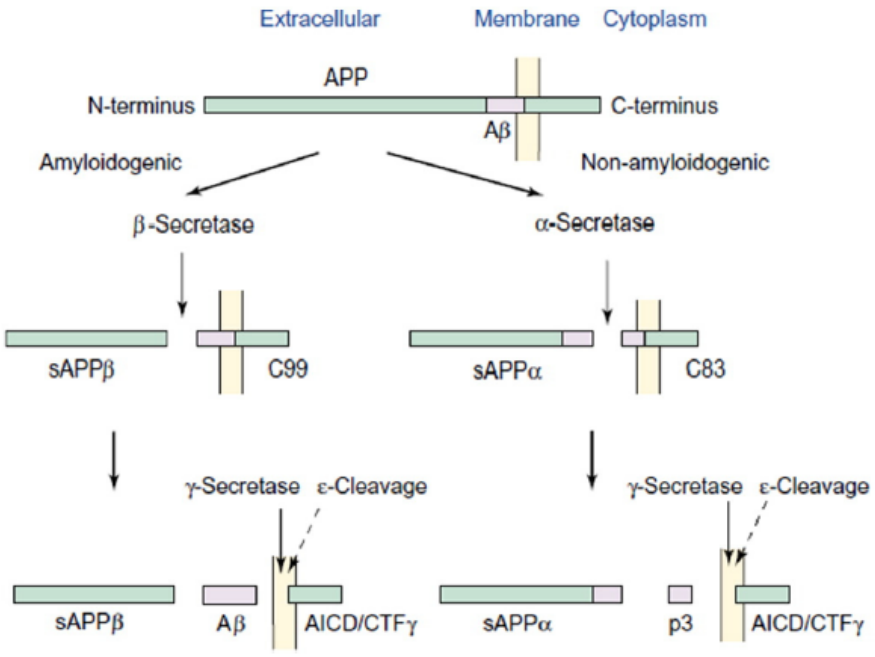


Figure 5: *Processing of APP: amyloidogenic and nonamyloidogenic pathways. [9]*

2.1.1 Non-amyloidogenic process

The non-amyloidogenic pathway is favored when APP accumulates at the cell surface, where it is cleaved by α-secretase within the Aβ sequence, specifically after Lys687 in the APP770 isoform. This cleavage, which occurs primarily at the plasma membrane where α-secretase is localized, precludes the formation of Aβ peptides. Instead, it produces two fragments: a soluble ectodomain called sAPPα, which is released extracellularly, a membrane-bound C-terminal fragment of 83 amino acids (CTFα or CTF83). CTFα can be internalized and further processed by γ-secretase within endosomes, generating the p3 peptide and the APP Intracellular Domain (AICD). [15, 16]

sAPPα

The physiological functions of sAPPα are not yet fully understood, but it is thought to have positive effects on neurons. In vitro studies have shown that it can protect them from harmful situations, such as oxygen and glucose deficiency or excessive stimulation by neurotransmitters (excitotoxicity). This protective action can occurs thanks to its ability to regulate ionic currents: sAPPα reduces the flow of calcium and increases the flow of potassium, helping to maintain stable resting membrane potential. Furthermore, it promotes processes that are fundamental for the development and communication between neurons, such as neurite growth, synapse formation and cell adhesion.

The beneficial functions of sAPPα depend on a site located at the carboxy-terminal end. This region begins just before the β-secretase cleavage site and extends to the end of the protein, which includes a heparin-binding motif, which is considered essential for binding to other molecules and for the activity of sAPPα.

From in vivo studies in mice, sAPPα has been shown to promote cell growth and survival and play an important role in brain development and cognitive function, although the precise mechanisms are not yet fully understood. When administered intracerebroventricularly, it can improve memory and learning, effects associated with increased long-term potentiation (LTP) and NMDA receptor activity, thus positively impacting synaptic plasticity. In addition to its effects on mature neurons, sAPPα also acts as a growth factor, stimulating the proliferation of neural stem cells (both embryonic and adult).

It has been noted that in some APP gene mutations there is impairment of the α-secretase cleavage site, as in the Dutch and Flemish mutations (APP E693Q and APP A692G, respectively). In the APP Swedish mutation (APP 670/671), a reduction in the concentrations of sAPPα in cerebrospinal fluid (CSF) was observed, in contrast to cases of sporadic AD in which there seems to be an unchanged or even higher concentration of sAPPα.[15, 16]

CTF83

CTF83 is the carboxyterminal fragments generated by the cleavage of APP by α-secretase, in the non-amyloidogenic pathway. subsequently this fragment will undergo γ-secretase cleavage obtaining the p3 and AICD fragments. Currently, no relevant biological roles are known for the carboxy-terminal fragment CTF83. [16]

AICD

AICD is the amino-terminal APP intracellular domain. In the non-amyloidogenic pathway, AICD is generated by the cleavage of the CTF83 fragment by γ-secretase. In addition to the classic cuts that produce AICDs of 57 or 59 amino acids, an alternative cut at the ε site, closer to the carboxy-terminal end, generates a shorter variant, AICD50. It has been proposed that the specific production of AICD50 depends on CTF83 and not on CTF99 derived from β-secretase.

Some authors hypothesize that this dependence may contribute to a reduction in AICD levels in Alzheimer's disease. However, it remains uncertain whether the pathology is associated with an excess of AICD or a loss of its physiological function. The role of AICD in Alzheimer's disease is still unclear, but it is thought to be involved in various cellular processes, including gene regulation, apoptosis, neuronal development, and cytoskeletal dynamics. [17]

2.1.2 Amyloidogenic process

From the cell membrane, APP can be internalized through clathrin-mediated endocytosis, a process involving the YENPTY motif in its cytoplasmic tail. Once in early endosomes, APP can follow three distinct pathways:

- Recycling to the plasma membrane, allowing renewed interactions and functional activity.
- Retrograde transport to the Trans-Golgi Network (TGN) via the retromer complex, which enables the reuse of proteins and lipids.
- Degradation in lysosomes.

Importantly, the internalization of APP into acidic compartments, like early endosomes, promotes the amyloidogenic pathway. This is due to the presence of β-secretase on endosomal membranes, where its colocalization with APP in the acidic environment facilitates amyloidogenic cleavage.

At the endosomal membrane, β-secretase (BACE) cleaves APP after Met671 in the APP770 isoform. This cleavage generates two fragments: a soluble ectodomain (sAPPβ) and a membrane-bound C-terminal fragment (CTFβ or CTF99).

CTFβ is subsequently cleaved by γ-secretase within its transmembrane (TM) domain. This proteolytic activity, which occurs in acidic compartments of living neurons, generates Aβ peptide monomers and the APP intracellular domain (AICD). [15, 16]

sAPPβ

sAPPβ is the soluble fragment generated by the cleavage of APP by β-secretase (BACE1). It shares most of its sequence with sAPPα, but differs in the last 16 C-terminal amino acids, which are absent in sAPPβ. This difference is crucial, since sAPPβ is much less effective in neuroprotective functions and synaptic plasticity.

Unlike sAPPα, sAPPβ does not participate in long-term potentiation (LTP) and is 50 to 100 times less effective in protecting neurons against stressors such as excitotoxicity, glucose deprivation, or β-amyloid toxicity. Furthermore, it is unable to prevent cell death in stressful situations.

Nevertheless, sAPPβ has some specific functions. It is involved in synapse pruning during nervous system development, a process important for correctly shaping connections between neurons. It also binds to the DR6 receptor, which activates a molecular cascade (caspase-6) responsible for the destruction of axons, but not of the cell body, suggesting a role in selective axonal death.

Other research shows that sAPPβ promotes the growth of neurites (the extensions of neurons) and can promote neuronal differentiation into stem cells, in some cases better than sAPPα. It also stimulates microglia — the brain's immune cells — as efficiently as sAPPα, thanks to the shared N-terminal domain. However, this effect, if prolonged, could have harmful consequences, because microglial activation is linked to inflammation and neurodegeneration.

In summary, although sAPPβ does not have the same neuroprotective properties as sAPPα, it promotes neural growth, regulates some protective proteins, and plays a role in development. Its neurotrophic capacity suggests that the production of Aβ, from which it derives, is not in itself harmful. However, considering that it stimulates microglia, any therapeutic strategy aimed at increasing sAPPβ levels should be applied in the very early stages of Alzheimer's disease, before inflammatory processes appear.[16, 18]

CTF99

As for the CTF83 fragment, also for CTF99, produced by the action of β-secretase on APP, a biologically significant function has not yet been clarified.[16]

Aβ

The Aβ peptide is a key element in the pathogenesis of Alzheimer's disease (AD). It derives from the sequential cleavage of the amyloid precursor protein (APP) by β- and γ-secretases, mainly at the level of endosomes and trans-Golgi network.

Figure 6: *Schematic of amyloid β (Aβ) peptide, from Alzforum. https://www.alzforum.org/mutations/app*

The resulting peptide has variable lengths, from 37 to 49 amino acids. The most common isoforms of Aβ are Aβ40 and, above all, Aβ42, the form most prone to aggregation and the one most involved in pathological processes. In hereditary cases and animal models of AD, an increase in total Aβ production or an increase in the Aβ42/Aβ40 ratio is often observed. In patients, however, a reduction of Aβ42 is detected in the cerebrospinal fluid, considered a good biomarker for AD, probably related to its greater deposition in amyloid plaques. In monomeric form, Aβ is considered to be structureless and nontoxic. However, it can self-assemble into more complex structures: soluble oligomers, protofibrils, and insoluble amyloid fibrils, which are deposited in the brain as senile plaques, characterized by a typical "crossed β-sheet" structure. In particular, oligomers are considered the most toxic species, as they inhibit NMDA-mediated synaptic transmission and contribute to synapse loss. According to the amyloid hypothesis, the accumulation of Aβ, together with aberrant phosphorylation of tau protein, gives rise to a neurotoxic environment that leads to synapse loss, neuronal damage, and progressive impairment of brain function, especially in areas involved in memory, learning, and emotion, such as the hippocampus, entorhinal cortex, and amygdala.

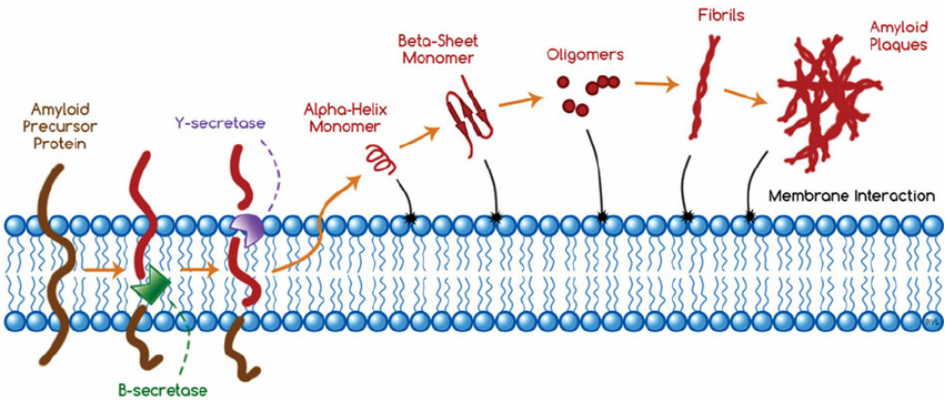


Figure 7: *Aβ origin and the mechanism of the amyloid plaques formation. Drolle E et al.(2014)*

Aβ also interacts with various cellular receptors (including PrPc, LRP, mGluR5, and NMDA), activating signaling pathways that alter synaptic balance and promote neurotoxicity. In addition, Aβ aggregates can interfere with neuronal metabolism, cause oxidative stress through the production of reactive species, and promote neuroinflammation through the activation of microglia and astrocytes, with release of proinflammatory cytokines.

Despite its involvement in Alzheimer's disease, Aβ is not intrinsically toxic: during brain development, its production is physiological and plays neuroprotective and neurotrophic roles. At low concentrations that do not allow the formation of oligomers, Aβ monomers promote neuronal survival, protect against cell death, and promote the proliferation and differentiation of neural progenitor cells. Aβ also participates in a negative feedback loop on synaptic activity, contributing to the physiological regulation of neuronal function. Furthermore, it has shown activity similar to that of antimicrobial peptides (AMPs), suggesting a possible defensive role in the innate immune system.

In summary, Aβ is a peptide with a double face: physiological and protective at low concentrations, but potentially toxic when it accumulates in oligomeric forms and insoluble aggregates, contributing substantially to the neurodegeneration observed in Alzheimer's disease.[16, 18, 19]

AICD

In the amyloidogenic process, AICD is released by the cleavage of CTF99 by γ-secretase. Regardless of the cleavage site and therefore its length, AICD retains the consensus motif YENPTY, which is essential for binding to adaptor proteins such as Fe65 and for the activation of cellular pathways. It can be further modified post-translationally, for example by phosphorylation. The best-known signaling pathway of AICD involves its binding to the protein Fe65, which allows the recruitment of the enzyme TIP60. This complex can then enter the cell nucleus and activate the transcription of specific genes.

In transgenic mice, co-overexpression of AICD and Fe65 induces EEG alterations, increased susceptibility to seizures, tau hyperphosphorylation, and neuronal loss in old age. These phenotypes do not appear in models with mutations that prevent caspase cleavage of AICD. Experiments performed on APP-deficient mice, in which the protein was reintroduced without the YENPTY domain, showed a marked reduction in Aβ production, probably due to altered APP trafficking in endosomes. This suggests that AICD, or more generally the YENPTY domain, may contribute to the control of cellular APP trafficking.[16]

Thus, it can be concluded that AICD, independently of the pathway through which it is generated, has been hypothesized to contribute to the pathophysiology of Alzheimer's disease. However, its role remains highly controversial, and it is hypothesized that alterations in AICD levels, related to γ-secretase activity, may be involved in the early mechanisms of the disease. [17]

2.2 APP processing enzymes: the role of secretases

2.2.1 α-secretase

α-secretase is an enzyme involved in the processing and cleavage of APP. Its importance derives from the fact that its cleavage site is located within the sequence of the Aβ peptide, between residues 687-688. Consequently, this cleavage, also called shedding, prevents the formation of Aβ. The products of this cutting are the secreted N-terminal fragment called sAPPα and a C-terminal fragment (CTF) called CTFα or CTF83.

α-secretase activity is attributed to one or more members of the ADAM (A Disintegrin And Metalloproteinase) family, of which ADAM9, 10, 17 and 19 are the leading candidates. In particular, the overexpression of ADAM10 in a mouse model of Alzheimer's disease showed a reduction in Aβ production, plaque deposition and cognitive deficits. These findings support the role of ADAM10 as a key α-secretase in the pathogenesis of AD, although other ADAM family members may also contribute to the generation of sAPPα.
The activity of α-secretase is strongly influenced by its subcellular localization. On the surface of the plasma membrane, the enzyme is considered constitutively active and is the main cleavage pathway for APP to reach the membrane. In contrast, in the trans-Golgi network, its activity is regulated by intracellular signals, as Protein Kinase C (PKC). PKC is a family of enzymes involved in cellular signal transduction that when activated, for example in response to an increase in intracellular calcium, can stimulate several processes, including vesicular trafficking and the activity of enzymes such as α-secretase. Thus, in this compartment, α-secretase competes with β-secretase for APP cleavage, but PKC activation may favor the nonamyloidogenic pathway.

It has been observed that neuronal activity can increase α-secretase activity, and this effect depends on the presence of calcium inside the cell and involves NMDA receptors. However, some studies have reported the opposite effect, suggesting that neuronal depolarization may also reduce α-secretase activity.[15, 16]

2.2.2 β-secretase

BACE1 is the major β-secretase enzyme in the brain and plays a central role in the formation of the Aβ peptide, which is implicated in Alzheimer's disease. BACE1 acts by cleaving APP into two fragments: sAPPβ, which is released outside the cell, and CTF99, which remains in the membrane and is further processed by γ-secretase to generate the Aβ peptide. In some conditions, BACE1 can cleave APP at a slightly different location, producing CTF89 and leading to the formation of a variant of the Aβ peptide (Aβ11–40).
APP cleavage by BACE1 occurs predominantly within endosomal vesicles, rather than at the cell membrane. BACE1 is mainly localized in the trans-Golgi network and endosomes, although it can temporarily reach the plasma membrane via vesicular trafficking, from which it is rapidly recycled. Neuronal stimulation, such as depolarization and increased vesicular turnover, promotes APP cleavage by BACE1, but also by α-secretase, in a competitive manner.

BACE1 is initially produced as a proenzyme in the endoplasmic reticulum and must undergo some modifications to become fully active: the union of two molecules (homodimerization), the cleavage of the prodomain by enzymes such as furin in the trans-Golgi network and glycosylation. Furthermore, it can be released from the membrane through a process called shedding, probably mediated by proteases of the ADAM family, the same ones involved in the action of α-secretase.

At the pathological level, it has been shown that genetic ablation of BACE1 completely prevents the formation of Aβ and amyloid pathology. However, BACE1 deletion is not without consequences: alterations in the brain have been observed, such as the reduction of synaptic spines in hippocampal neurons and behavioral disorders similar to schizophrenia, including hyperactivity, cognitive and social recognition problems. These effects could arise because BACE1 also acts on other substrates in addition to APP.
Finally, it is known that BACE1 expression increases under conditions of cellular stress, such as hypoxia, ischemia, oxidative stress and energy deficiency, and that this increase also depends on γ-secretase activity.

Although BACE1 is considered the main β-secretase in the pathogenesis of Alzheimer's disease, recent studies have questioned its exclusivity, showing that other proteases, such as cathepsins, can also have β-secretase activity.[16]

2.2.3 γ-secretase

γ-secretase is an enzyme involved in both non-amyloidogenic and amyloidogenic processing of APP. In the first case, it acts on the CTFα fragment, generating the peptides p3 and AICD, while in the second case, it acts on the CTFβ fragment, producing Aβ and AICD. [16]
γ-secretase processes APP in a three-step cleavage process: first the ε cleavage, quickly followed by the ζ cleavage, and finally the γ cleavage. This pattern, which proceeds from the membrane boundary toward the interior of the transmembrane domain, suggests an ordered mechanism in which each cleavage depends on the previous one.

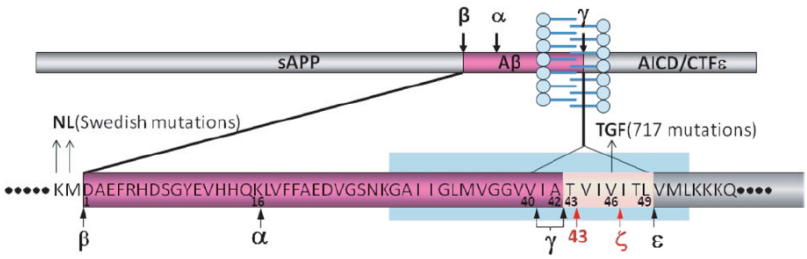


Figure 8: *Schematic representation of the ε, ζ and γ cleavage sites of γ-secretase.[20]*

- ε site (Aβ49/APP720): it is the first to be cleaved, located close to the membrane boundary. From this cleavage the intracellular domain of APP (AICD) originates. The proximity of the ε cleavage site to the membrane boundary may facilitate the access of water molecules, necessary for peptide bond hydrolysis, to the active site of gamma secretase located in the hydrophobic environment of the membrane.

- ζ site (Aβ46/APP717): it represents an intermediate step. It was identified thanks to the detection of Aβ46, whose formation is blocked by gamma secretase inhibitors.

- γ site (Aβ40/42): generates the most common secreted forms of Aβ. Aβ40 is the most abundant, while Aβ42, more hydrophobic and aggregating, is implicated in the formation of amyloid plaques.

From the formation of the Aβ46 peptide, the cleavage process can follow two main pathways. In the predominant pathway, Aβ46 is cleaved at the level of Aβ43, which in turn is processed to Aβ40 and in some cases, Aβ40 can be further shortened to Aβ37.
Alternatively, although with less efficiency, Aβ46 can be cleaved directly at the level of Aβ42. Aβ42 can also be further processed to generate shorter peptides, such as Aβ38 or Aβ39, or be released in its active form. [20]

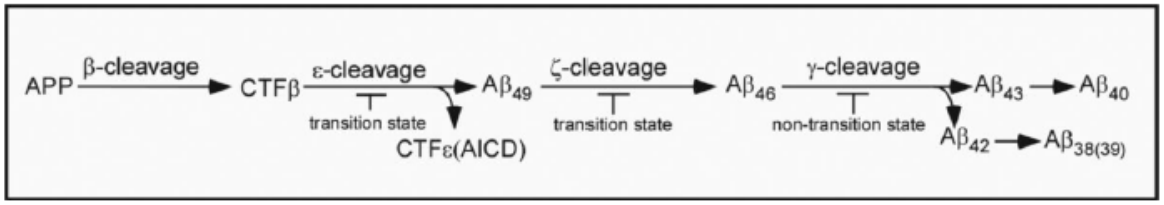


Figure 9: *Schematic illustration of sequential cleavages of APP mediated by γ-secretase. After cleavage of APP by β-secretase, the CTFβ fragment is processed by γ-secretase with successive cuts: ε-cleavage (Aβ49) and ζ-cleavage (Aβ46). Two pathways branch off from Aβ46: a main pathway Aβ46 → Aβ43 → Aβ40 → Aβ37, and an alternative (less efficient) pathway Aβ46 → Aβ42 → Aβ38/39.[20]*

γ-secretase is an enzymatic complex composed of four essential subunits: presenilin (PS), nicastrin (NCT), APH-1 and PEN2. It is an aspartyl protease involved in the intramembrane cleavage of over 90 proteins, including APP, Notch, ErbB4, N-cadherin and p75NTR.

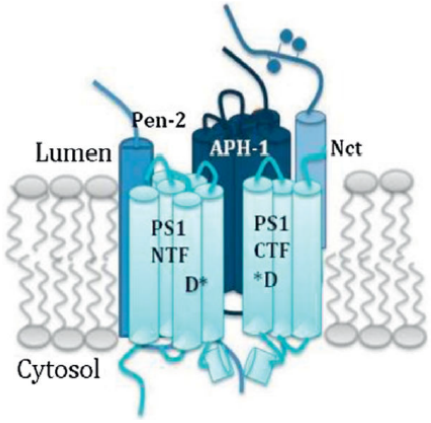


Figure 10: *γ-secretase complex [21]*

PS constitutes the catalytic center of the complex and has nine transmembrane domains. To become functional, it must undergo endoproteolysis between the sixth and seventh domains, where the two catalytic aspartyl residues are located, generating two fragments (N- and C-terminal) that remain associated in the membrane and form the active enzyme. In the absence of this process, full-length PS is rapidly degraded. Mutations in PS associated with the familial form of Alzheimer's disease alter the production of Aβ peptides, favoring more amyloidogenic species or increasing the total amount of Aβ. The other subunits have structural and functional roles: NCT is involved in substrate recognition, APH-1 acts as a scaffold for the initial assembly, while PEN2 is involved in the maturation of the complex and in the proteolysis of PS.

The assembly of γ-secretase occurs in an orderly manner: in the endoplasmic reticulum NCT and APH-1 form a scaffold to which PS binds; finally PEN2 completes the complex and induces the cleavage of PS. The complex is then transported to the Golgi, where it is glycosylated, becoming fully active only at the end of these steps.

The activity of γ-secretase against APP is strongly influenced by its subcellular localization. Possible interactions with lipids and proteins play a crucial role in APP trafficking. For example, LRP1 promotes APP endocytosis, favoring the production of Aβ. Blocking this interaction shifts APP to the cell surface, reducing amyloidogenic processing. Furthermore, since LRP1 is also a substrate of γ-secretase, modulation of LRP1 cleavage by γ-secretase can modify APP trafficking. Numerous studies in murine models have shown that a partial inhibition of γ-secretase can significantly reduce Aβ production and amyloid pathology, without severely compromising neuronal function. For example, a moderate reduction of γ-secretase activity led to cognitive benefits in APP transgenic mice. However, complete inhibition is often associated with neurodegeneration or cognitive deficits, suggesting that γ-secretase activity is essential for maintaining neuronal function over time.

Several variants of γ-secretase (γ-secretase) have been identified, with different subunits determining its specificity for APP or other substrates such as Notch. This has allowed the development of more selective inhibitors to reduce Aβ without altering other important functions. However, even selective inhibitors must be used with caution, as they can cause side effects, especially in the adult or elderly brain.[16, 21]

Pathways	Secretases	Cleavage sites	Fragments	Notes
Non-amyloidogenic	α-secretase	APP 687-688 (Aβ16-17)	sAPPα + CTF83	Cleavage of APP (within the Aβ domain) Prevents Aβ formation
	γ-secretase	ε-site: APP720-721 ζ-site: APP717-718 γ-site: APP711-712/APP713-714	p3 + AICD	Cleavage of CTF83
Amyloidogenic	β-secretase	APP 671-672	sAPPβ + CTF99	Cleavage of APP
	γ-secretase	ε-site: Aβ49 ζ-site: Aβ46 γ-site: Aβ40/42	Aβ40/42 + AICD	Cleavage of CTF99 Aβ formation

Table 2: *APP cleavage pathways and resulting products from α-, β-, and γ-secretases [16, 20]*

2.3 Familial Alzheimer's Disease Mutations

Alzheimer's neurodegenerative disease can be sporadic or familial. The sporadic form is the most common and is usually not associated with any genetic cause but with age and environmental factors. The familial form is rarer and leads to an earlier onset of the disease, often before the age of 65.
FAD is mainly caused by mutations in three genes: presenilin 1 (PSEN1), presenilin 2 (PSEN2), and amyloid protein precursor (APP). PSEN1 and PSEN2 are both reported to be part of the γ-secretase complex. To date, approximately 230 mutations have been identified in these three genes. These mutations influence a common pathogenetic pathway in the synthesis and proteolysis of APP, leading to excessive production of Aβ. In particular, these kind of mutations lead to an increased Aβ42 production, supporting the importance of the amyloid cascade theory.

Clinically, compared to sporadic AD, FAD is distinguished by an earlier age of onset, a positive family history, a variety of non-cognitive neurological symptoms and a more aggressive course. Initial symptoms often involve episodic memory, followed by other cognitive deficits. The clinical features depend on the genotype. For example, PSEN1 mutations tend to have an earlier age of onset and may present with non-cognitive symptoms more frequently, including an 'AD variant' with spastic paraparesis and CWP (cotton wool plaques). PSEN2 mutations are less common and tend to have a later age of onset and a more variable course. APP mutations are often associated with more severe CAA (cerebral amyloid angiopathy), which can lead to cerebral hemorrhage and stroke. APP gene duplication is also sufficient to cause FAD with CAA. [22]

Amino acid mutations associated with familial Alzheimer's disease (FAD) in the APP gene are located close to the cleavage sites of α-, β- and γ- secretases:

- A double missense mutation was identified at the β-secretase cleavage site, found in a Swedish family: the Swedish KM670/671NL mutation, which is associated with a 4- to 7-fold increase in total Aβ production.
- Near the α-secretase cleavage site, the APP Flemish (A692G, A21G), APP Arctic (E693G, E22G) and APP Iowa (D694N, D23N) mutations were identified, known to increase the tendency of Aβ to aggregate, contributing to toxicity to cerebral vessel cells and the development of cerebral amyloid angiopathy (CAA).
  - Previous studies noted that APP Flemish led to a doubled production of Aβ40 and Aβ42.
  - For APP Arctic, a significant reduction in Aβ42 levels was observed, while the amount of Aβ40 remained unchanged. In vitro experiments showed that the mutated Aβ1-40 variant forms protofibrils faster and in greater quantity than the wild-type Aβ1-40.

  - In APP Iowa, the production of Aβ40 and Aβ42 was not altered. However, there was evidence of reduced sAPPα production and marked accumulation of the protein in early endosomes, consistent with reduced α -secretase activity.

- Around the γ-secretase cleavage site the APP Iberian variant (I716F, I45F) was identified, in which a 34-fold increase in the Aβ42/Aβ40 ratio was noted, mainly due to a drastic reduction in Aβ1-40 production. This mutation also leads to a decrease in synaptogenic activity.

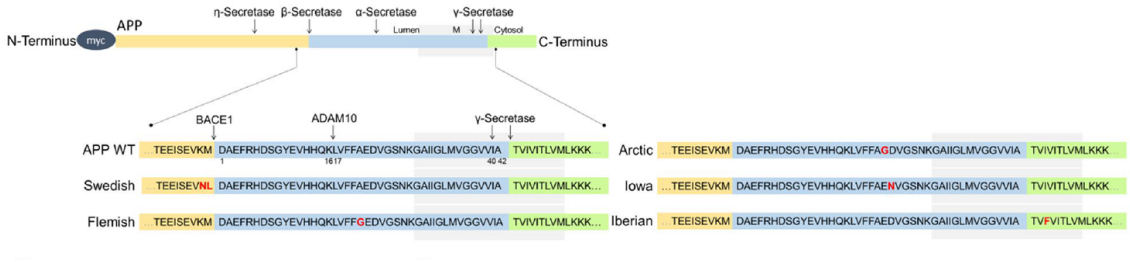


Figure 11: *Schematic representation of APP695 WT and FAD mutants with indicated Aβ region, transmembrane domain, and secretase cleavage sites [14]*

Recent studies have indicated an essential APP function at the synaptic level. This depends on both the production of the sAPPα fragment and the full-length APP protein, which appears to act through trans-synaptic signalling mechanisms. However, specific FAD mutations affect intracellular trafficking, proteolytic processing and synaptogenic activity of APP, revealing that mutations located at the cleavage sites of α -, β- and γ-secretases alter these functions through distinct pathogenic mechanisms. [14]

2.3.1 The Swedish mutation

The cleavage site of β-secretase (BACE1) on the APP is located between residues Met671 and Asp672 (referred to the APP770 isoform). The cut made by BACE1 generates an N-terminal fragment which begins with the amino acid Asp672, corresponding to the first residue of the β-amyloid peptide (Aβ) sequence. Upstream of the β-amyloid sequence are residues involved in the Swedish mutation, named KM670/671NL. In this mutation, the lysine residue (Lys, K) at position 670 is replaced by asparagine (Asn, N), while methionine (Met, M) at position 671 is replaced by leucine (Leu, L).

The main effects observed are:

- Increased Aβ production: The Swedish mutation is known to increase the affinity of BACE1 for its cleavage site, resulting in a 4- to 7-fold increase in overall Aβ production compared with wild-type APP.
- Alteration in proteolytic processing: cleavage by BACE1 leads to increased accumulation of C-terminal β fragments (β-CTF), also observing increased production of sAPPβ.
- Peptide profiles of modified Aβ: Analysis by mass spectrometry (MALDI-TOF) showed that although the Aβ1-42/Aβ1-40 ratio remained similar to that of wild-type APP, APP Swedish generated increased levels of shorter N-terminal fragments, such as Aβ1-14, Aβ1-15, Aβ1-16, Aβ1-17, Aβ1-19 and Aβ1-20.
- Effects on cellular localization and trafficking: The mutation does not appear to alter the synaptogenic activity of APP, nor its localization on the cell membrane, the ability to increase protein complexes or the rate of endocytosis. However, it was observed that accumulation of the β-CTF fragment leads to an increase in endosome volume, suggesting a possible effect on endosomal trafficking.
- Evidence from murine models: On a Tg2576 transgenic mouse model expressing human APP with Swedish mutation, pathological changes typical of AD were observed, including cognitive deficits, Aβ accumulation, amyloid plaque formation, and loss of synapses.
- Clinical manifestations in patients: Familial AD subjects with the Swedish mutation show a picture of brain atrophy, characterized by enlargement of cortical furrows and mild dilation of cerebral ventricles.

In conclusion, the Swedish mutation acts as a potent amplifier of the amyloidogenic pathway, promoting an increased affinity for β-secretase and thus leading to a marked increase in Aβ production. [14]

2.3.2 The Iowa mutation

The cleavage site of α-secretase in the amyloid precursor protein (APP) is located within the Aβ sequence, between residues Lys687 and Leu688 of the APP770 form, corresponding to residues 16 and 17 of the Aβ form. Near this site is the Iowa mutation (D694N according to APP770 or D23N in Aβ sequence), a genetic mutation associated with the familial form of Alzheimer's disease (FAD). In this mutation, the Aspartic acid residue (Asp, D) at position 694 is replaced by the Asparagine residue (Asn, N). Several studies have shown functional and biochemical alterations associated with the Iowa mutation, involving both the subcellular localization of the protein and its processing by secretases:

- The Iowa form of APP tends to accumulate more in early endosomes than the wild-type form. This does not appear to be due to changes in the rate of endocytosis, which remains unchanged, but probably to an alteration in the regulation of intracellular trafficking of the protein.
- Cleavage by α-secretase appears to be reduced in the Iowa form, resulting in a reduction in the soluble sAPPα fragment. This finding is consistent with the increased localization of mutated APP in endosomes, which are compartments less favorable to cleavage by α-secretase and more favorable to cleavage by β-secretase.
- Accessibility to the β-cleavage site remained unchanged, as no significant differences were found in processing by β-secretase.
- Regarding γ-secretase, analysis by mass spectrometry (MALDI-TOF) indicated a slight reduction in Aβ1-40 peptide, resulting in a slight increase in the Aβ1-42/Aβ1-40 ratio, although not statistically significant.
- Spectrometric analysis revealed a significant increase in Aβ1-19 fragment production and an increase in Aβ1-33 peptide. In addition, a slightly reduced signal for Aβ1-38 was observed. Aβ peptides truncated at the N-terminal end from position 5 (Aβ5-29 and Aβ5-33) were also identified, suggesting activation of alternative processing pathways.
- The Iowa mutation does not significantly alter the synaptogenic activity of APP nor the formation of protein complexes or the localization of the protein on the cell surface.
- Regarding pathological implications, the Iowa mutation is associated with cerebral amyloid angiopathy (CAA), a condition in which Aβ peptide accumulates in the blood vessels of the brain, causing small infarcts and hemorrhages of brain tissue.
- At the molecular level, it appears that this mutation promotes the formation of Aβ fibrils, contributing to the increased toxicity.

In summary, the Iowa mutation leads to altered localization of APP in early endosomes, resulting in decreased processing by α-secretase and specific changes in the levels of certain Aβ peptides, including an increase in Aβ1-19 and the appearance of N-terminal truncated forms (Aβ5-x).[14]

2.3.3 The Iberian mutation

The γ-secretase cleavage occurs between residues Val711 and Ile712 (for Aβ40) and between residues Ala713 and Thr714 (for Aβ42) of APP770 protein. In the vicinity of these cleavage sites is the Iberian mutation. The Iberian mutation (I716F according to APP770 or I45F in Aβ numbering) is a genetic FAD mutation, in which the Isoleucine residue (Ile, I) at position 716 is replaced by the Phenylalanine residue (Phe, F).

- In the Iberian mutation, synaptogenic activity is impaired.
- No significant changes in processing by α-secretase or β-secretase were detected compared with wild-type APP, as no significant alterations in the production of sAPPalfa, sAPPβ, α-CTF or β-CTF were revealed.
- The Iberian APP mutation has a very marked effect on the way γ-secretase cuts the protein, significantly altering the production of Aβ peptides. In particular, it causes a sharp increase in the Aβ42/Aβ40 ratio, which is 34 times higher than normal. This is relevant because Aβ42 is much more prone to aggregate than Aβ40 and is closely linked to the pathogenesis of Alzheimer's disease. Analysis by mass spectrometry (MALDI-TOF) confirmed these changes: there is a significant increase in Aβ1-38 and Aβ1-42 peptides; at the same time, the level of Aβ1-40 is significantly reduced. These changes explain why Aβ1-42/Aβ1-40 and Aβ1-38/Aβ1-40 ratios are much higher than in samples without the mutation. In addition, the Aβ1-17/Aβ1-40 ratio is also found to be increased, indicating a further alteration in the way APP is cut. Finally, the fact that Aβ1-40 and Aβ1-42 show such marked differences in their levels suggests that they might be produced in two separate ways or through different cutting lines by γ-secretase, rather than being the result of the same sequential pathway.
- In the case of the Iberian mutation, there were no major differences from the wild-type form in the distribution of APP in the cis-Golgi compartment and on the cell surface. The presence of APP in early endosomes and the rate of endocytosis were also not significantly altered, suggesting that the mutation does not substantially alter intracellular trafficking or internalization of the protein.
- In addition to typical symptoms of Alzheimer's disease, the Iberian mutation has been associated with motor deficits and the presence of Lewy bodies in the amygdala, suggesting broader

neuropathological involvement.In addition, the Iberian mutation was observed to exacerbate synapse loss and decline in cognitive function. The significant reduction in Aβ40 levels, typical of this mutation, could contribute to these negative effects, considering that Aβ40 also plays a modulatory role on Aβ42 aggregation and in synaptic stability.

In summary, the Iberian mutation has a major impact on APP processing by γ-secretase, leading to a pronounced increase in the Aβ42/Aβ40 ratio and reduced synaptogenic activity. It does not significantly affect the processing by α- and β-secretases or the cellular localization of APP. Its association with complex pathology, including potential links to synucleinopathy, distinguishes it from other mutations studied. [14]

2.4 F-spondin: a ligand to modulate APP cleavage

F-spondin is a secreted neuronal glycoprotein identified through subtractive hybridization due to its abundant expression in the floor plate during spinal cord development.[23]

This protein consists of 807 aminoacids and contains three domains:

- N-terminal reelin domain (residues: 29-194);
- Central spondin domain (residues: 195-388);
- C-terminal thrombospondin domains, that consists of six repeats (residues type1: 442-495, type 2: 501-555, type 3: 558-611, type 4: 614-666, type 5: 668-721, type 6: 754-806).[24]

Its expression is high during embryonic development but decreases after birth, though it can be upregulated following neuronal injury. F-spondin is also widely expressed outside the brain.

Functionally, F-spondin plays a role in axonal pathfinding, neural regeneration, and cell–cell interactions. It promotes neurite outgrowth, inhibits the adhesion of neural crest cells and endothelial cells to the extracellular matrix, and impairs endothelial cell migration. In adult animals, sciatic nerve axotomy significantly increases F-spondin expression distal to the lesion, suggesting its involvement in neuronal repair.[23]

2.4.1 F-spondin interaction with APP

As already mentioned, according to the cascade amyloid hypothesis, Aβ aggregation in plaques is the main cause of Alzheimer's Disease. Aβ originates from the amyloid precursor protein (APP) through a series of sequential cleavages mediated by BACE1 and γ-secretase. [6]

One of the various therapeutic approaches that has been studied is the inhibition of BACE1 enzyme, in order to reduce Aβ production. The problem is that BACE1 is not only responsible for initiating amyloidogenesis, but it also plays a role in various physiological processes in the brain, including the regulation of sodium currents, synaptic transmission, and myelination. Consequently, its inhibition could cause dysfunction in normal brain activity.

To solve this drawback, a strategy to inhibit APP processing could be to use potential ligands that could bind to APP to prevent the beginning of amyloidogenesis. Recent findings indicate that F-spondin is a potential ligand for APP. It binds to the extracellular APP domain CAPPD, also called E2 domain, and inhibits the initial β-cleavage of APP by BACE1. This suggests that F-spondin can regulate its cleavage and consequently may serve as regulator for Aβ formation.[25]

Studies carried out by Angela Ho and Thomas C. Sudhof in 2004 showed that the presence of F-spondin significantly decreases the CTFs of APP, by approximately 70-80%. The decrease in CTFs caused by F-spondin is not simply a reflection of a small decrease in APP levels but is due to a large decline in APP cleavage by relatively low levels of F-spondin.[23]

2.4.2 F-spondin interaction with ApoEr2

Recent genome studies have identified apoE as the primary genetic risk factor for late-onset Alzheimer's disease. In affected brains, apoE is a component of most plaques and it influences Aβ production, aggregation, and clearance and has been described as a pathological chaperone.[26] A chaperone is a protein that helps other proteins to fold correctly or prevent the formation of toxic aggregates. However, ApoE (especially the ApoE4 isoform) can instead promote abnormal aggregation of Aβ, promoting the formation of the amyloid plaques that characterize Alzheimer's disease.[27, 28]

Its normal functions are mediated by the low-density lipoprotein receptor (LDLR) family. ApoE interacts with several LDLR family members, including apoE receptor 2 (apoEr2), which facilitate the cellular endocytosis of apoE-containing lipoproteins.

Recent research has discovered structural and functional links between APP and this receptor family. Apolipoprotein E receptors and APP share important features: they are type I transmembrane proteins with large extracellular and small cytoplasmic domains, and both undergo γ-secretase cleavage, generating intracellular domains that associate with common adaptor proteins.

APP and apoE receptors seem to interact through extracellular and intracellular proteins. In the extracellular domain, F-spondin interacts with APP and with some of the apoE receptors, including Apoliprotein E receptors 2 (ApoEr2):[26, 29]

- The interaction between F-spondin and ApoEr2 occurs through the thrombospondin domain of F-spondin and the ligand binding domain of ApoEr2 (containing ligand binding domain repeats 1, 2, 3, 7 and 8). The thrombospondin domain has six repeats, and only the first four repeats interact with ApoEr2.
- The last two thrombospondin repeats (5-6) of F-spondin, which aren't involved in the interaction with ApoEr2, binds the extracellular matrix.
- The interaction between F-spondin and APP takes place through the reeler and spondin domains of F-spondin and the central APP domain (or ectodomain E2).

Some experiments showed that:

- F-spondin treatment promoted cell surface levels of APP and ApoEr2.
- Full-length F-spondin increased cleavage of ApoEr2, consequently secreted ApoEr2 and ApoEr2 CTF, and increased levels of secreted APP and APP CTF. This doesn't occur with only single F-spondin domains.
- Full-length F-spondin increased levels of secreted APP and APP CTF, but decreased β-CTF, suggesting that F-spondin inhibits β-secretase cleavage.
- There is a decrease in secreted Ab40 levels (by 7%) and Ab42 levels (by 50%).
- The inhibitor of apoE receptor family (RAP) prevented F-spondin from influencing APP proteolysis, this means that the effects on APP proteolysis are mediated by interactions with apoE receptors.
- F-spondin enhances the levels of α-CTF of APP and CTF of ApoEr2 in transfected cells, primary neurons, and brain.

There is also an intracellular interaction between APP and ApoEr2. F-spondin enhances their extracellular binding, while intracellularly, adaptor proteins like Fe65 may link them. In the intracellular region, adaptor proteins such as JIB and Dab1 may influence downstream signaling pathways.[26]

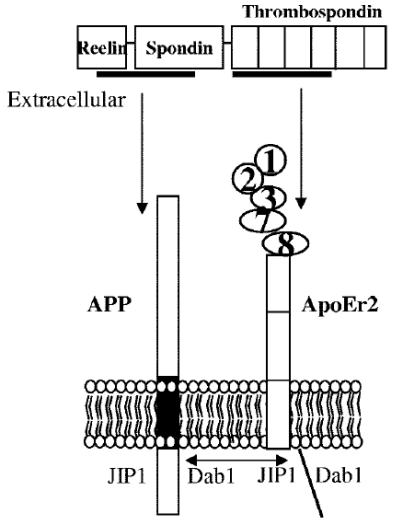


Figure 12: *Model of APP, ApoEr2, and F-spondin interaction [26]*

3

MATERIALS AND METHODS

3.1 Introduction

In this thesis work, computational molecular modeling methods were used to investigate the interactions between different variants of the APP protein and the F-Spondin protein, with the aim of better understanding the molecular mechanisms involved in Alzheimer's disease.

The adopted approach is part of a multidisciplinary context that combines bioinformatics, computational chemistry and structural biology. In particular, starting from the amino acid sequences obtained from the UniProt database, three-dimensional (3D) structures of the proteins of interest, wild-type APP, its mutant variants (Swedish, Iowa and Iberian) and the F-Spondin segment containing the Reelin and Spondin domains, were generated using AlphaFold. Subsequently, preliminary molecular dynamics (MD) simulations lasting 5 ns were performed on all the protein structures, aimed at structural stabilization and conformational optimization. The mutated APP structures were then aligned to the wild-type form to allow a consistent comparison of interactions in subsequent molecular docking studies.

Molecular Docking was performed between each APP variant and F-Spondin, thus generating the following molecular complexes:

- Wild-type APP + F-Spondin;
- Swedish APP (KM670/671NL) + F-Spondin;
- Iowa APP (D694N)+ F-Spondin;
- Iberian APP (I716F) + F-Spondin.

Through these techniques it was possible to model the interaction between proteins at the atomic level, enabling more in-depth analyses, such as the estimation of binding energies and the exploration of the mechanical and dynamic properties of the formed complexes. In this way, molecular simulation confirms itself as an innovative and powerful method to investigate the structure, dynamics and biological functions of macromolecules involved in pathological processes.

3.2 Multiscale Modeling of Biological Systems

In recent years, systems biology has changed the way to study biological processes, proposing an integrated and multidisciplinary approach to understand the complexity of living organisms. By combining knowledge from genomics, proteomics, biochemistry, physiology, mathematics, and computer science, this field aims to model and analyze intra- and intercellular dynamics across different spatial and temporal scales. A central objective is the development of mathematical and computational models capable of describing biological phenomena from the molecular to the organ level. In this perspective, molecular and multiscale modeling play a key role, especially in structural biology, where the three-dimensional structure of macromolecules, such as proteins and nucleic acids, is essential to understanding cellular function. These models allow to realistically simulate biomolecular interactions, predict the effects of structural mutations and study the mechanical and dynamic properties of complex biological systems. To understand the emergent behavior of living systems, both in physiological and pathological conditions, the convergence of different disciplines and a constant interaction between experimental and computational approaches is necessary.

To fully understand biological functions, it is necessary to integrate information across various levels of biological organization, each of which operates on specific spatial and temporal scales:

- Atomic/Subatomic: ab initio quantum simulations that consider electron-electron interactions, typically on systems composed of 100-1000 atoms.
- Molecular: classical atomic dynamics is studied, based on Newton's laws, applied to larger systems with 10k-100k atoms, with simulation times in the order of 100–1000 ns.
- Macromolecular: to manage more complex systems simplified models are used, called Coarse Graining, which allow longer simulations but with reduced precision.
- Subcellular: a continuous approach is adopted, often based on partial differential equations (PDEs), to model the flows and diffusion of substances on a mesoscopic scale.
- Cellular: The cell, the basic unit of physiology, is modeled through ordinary differential equations that describe its interactions and functions.
- Tissue: Groups of cells coordinated with each other are considered; space-dependent processes are represented with partial differential equations.
- Organ: different tissues integrate to form a functional structure; modeling also includes geometric aspects, such as the use of finite element models.

- System/Organism: different interconnected organs work together to perform complex functions; modeling must take into account the feedback mechanisms that regulate these interactions.
- Environment: the mutual influence between organism and environment is considered, which includes signals, other organisms and various external factors that can impact the biological system.

Since many of these interactions are not directly observable experimentally, computational models are essential for capturing the nonlinear dynamics of biological systems. Depending on the scale, different physical laws and computational methods apply:

- Quantum mechanics for electronic phenomena
- Molecular dynamics for atomic interactions
- Mesoscale models for molecular assemblies
- Continuum models for tissues and organs

Multiscale modeling is therefore essential to connect properties at the molecular level with behaviors observable at the macroscopic level. In the biological field, spatial scales can be defined based on the hierarchy of biological organization, while temporal scales depend on the speed of physiological processes.[30]

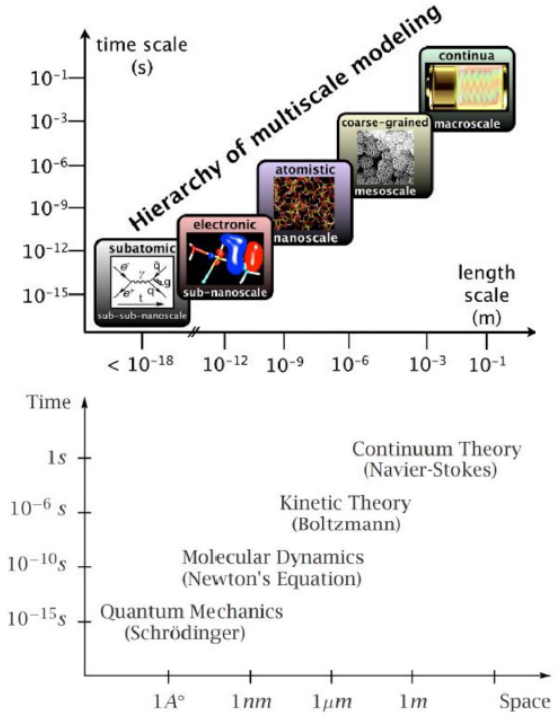


Figure 13: *Hierarchy of multiscale modelling*

3.3 Structural predictions using the AlphaFold Server

AlphaFold Server is a free online platform (for non-commercial use only) that allows to predict with great accuracy the 3D structure of many biological molecules, such as proteins, DNA, RNA, small ligands, ions and even chemically modified residues. The service is based on the new AlphaFold 3 (AF3) model, developed by Google DeepMind and Isomorphic Labs (2024), a very advanced evolution of the previous AlphaFold 2 (AF2). [31]

AF2 has greatly expanded the structural knowledge of the human proteome, providing models with an accuracy comparable to experimental accuracy and facilitating techniques such as cryo-EM, X-ray crystallography and nuclear magnetic resonance (NMR). [32]

AF3 represents a breakthrough in molecular modeling, accurately handling heterogeneous biomolecular complexes that include almost all types of molecules present in the Protein Data Bank, particularly improving predictions of difficult interactions. This makes it a valuable tool for both basic research, to better understand cellular mechanisms, and drug development, where an accurate structural information is essential.

AF3 introduces a new architecture based on a diffusion process. The system receives protein sequences, ligands and chemical bonds as input, and uses specific modules to process this information: one of these (Pairformer) focuses on the relationships between pairs of atoms, reducing the use of multiple sequence alignments. The diffusion module directly predicts the basic atomic coordinates, without having to go through more complex representations such as angles or twists of side chains. This module is trained to correct the noise in the atomic coordinates, learning to recognize both local details and the global structure of the complex. During the prediction, it starts from a random configuration and gradually cleans it up until it obtains the final structure.

AF3 is optimized to quickly learn local structures, but requires more time to model complex interactions such as those between different proteins. Finally, to ensure reliable results, especially on difficult complexes such as antibody-antigen ones, it generates many different structures and selects the one with the highest confidence.

Despite its potential, AF3 still has some limitations. It can sometimes violate the rules of stereochemistry, generating structures with overlapping atoms or wrong chirality. It can also produce disordered regions that do not resemble typical structures seen in experiments. Furthermore, the predicted structures are static and do not reflect the dynamic behavior of molecules in solution. In some cases, it can only predict one conformational state even if more than one exists. Finally, for very difficult complexes, such as those between antibodies and antigens, it may be necessary to generate many predictions to obtain reliable results. [33]

3.4 Molecular Mechanics

Molecular mechanics (MM) is a discipline developed in the 1970s that applies the principles of classical (Newtonian) mechanics to model the structure and behavior of molecules, from the simplest to complex systems such as proteins, nucleic acids or materials made up of millions of atoms. In this context, atoms are treated as point masses connected by springs that represent chemical bonds, according to a model called mass-spring. The interactions within the molecule are described by force fields, which are mathematical functions that include parameters related to bonds, angles, torsions and non-bonded interactions. MM is widely used in theoretical, computational and experimental fields (for example in structural analyses using X-rays or NMR). Simulations allow us to build realistic molecular models, optimize them from an energetic point of view and predict the behavior of new substances before their synthesis, with applications ranging from molecular biology to drug design, up to advanced materials and nanotechnologies. [34]

Molecular mechanics is based on some simplifying assumptions, including the fundamental Born-Oppenheimer approximation, which allows the motion of electrons to be separated from that of nuclei. Although electrons are not represented explicitly, their properties are incorporated into the parameters of force fields. Since they adapt rapidly to changes in the position of nuclei, the energy of the system can be treated as an exclusive function of nuclear coordinates. Energy variations that arise from atomic motion, such as rotations of bonds or more complex rearrangements, are described on a potential energy surface, where stationary points identify stable configurations, i.e. energy minima. The force fields used in MM, even if based on relatively simple functions, produce reliable results. A key property is transferability, i.e. the ability to apply parameters developed on small molecules to larger systems without losing their accuracy.
MM force fields consider four main components: bond stretching, angle variation, torsion around single bonds, and non-bonding interactions. Each term is associated with specific internal coordinates, simplifying both the analysis of molecular behavior and the parameterization step.[35] The total potential energy depends on the relative positions of the atoms in the system and varies according to the force field adopted. The force field represents the set of parameters and mathematical functions used to describe atomic interactions. Each interaction term is modeled by specific equations, aimed at accurately reproducing molecular geometries, energies and dynamic properties.[36]

3.4.1 Potential Energy Surface and Energy Minimization

The Potential Energy Surface (PES) is a multidimensional function that describes the potential energy of a system as a function of its atomic coordinates. A system consisting of N atoms is defined by 3N Cartesian coordinates and 3N-6 internal coordinates, and the PES represents how the energy varies as these coordinates vary. It is defined in a high-dimensional space; the commonly used three-dimensional representation is actually a simplified projection of the Potential Energy Surface, referred to as energy landscape.

The potential energy function is expressed as the sum of the contributions from bonded and non-bonded interactions:

$$V = V_{bond}(r_1 + r_2 + \cdots + r_n) + V_{non\text{-}bond}(r_1 + r_2 + \cdots + r_n) \qquad \text{(Equation 1)}$$

Bonded interactions include:

- Covalent bond stretching
- Angle deformation
- Dihedral angle
- Improper dihedral

Non-bonded interactions include:

- Van der Waals interactions
- Electrostatic interactions
- Hydrogen bonds

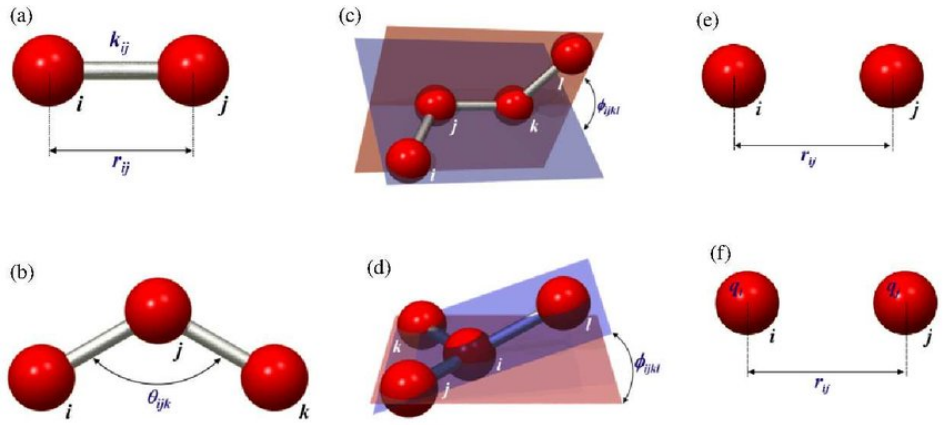


Figure 14: *Schematic representation for (a) covalent bonding , (b) bond angle interactions, (c) dihedral angles, (d) improper dihedral, (e) long range Van der Waals and (f) electrostatic interactions.*
from: Kouza, Maksim. (2013). Numerical Simulation of Folding and Unfolding of Proteins.

Bonded interactions refer to pairs of atoms joined by a covalent bond and, in molecular mechanics (MM), are modeled as harmonic potentials that depend on the bond length.

$$V(l) = \sum_{bonds} \frac{1}{2} K_l (l - l_0)^2 \qquad \text{(Equation 2)}$$

- K_l: force constant.
- l_0: reference bond length, assumed when all the other forcefield terms are zero.
- l: equilibrium bond length, that minimizes V when all the terms are considered.

The chemical bond is modeled as an ideal spring with a potential $V(l)$ that is zero at the equilibrium length l_0. Since the energy dependence is quadratic, the potential assumes a parabolic shape, increasing symmetrically both for compression ($l < l_0$) and for extension ($l > l_0$) of the bond. However, the atoms are not static: the bond length oscillates around an average value due to thermal vibrations. The harmonic model is valid only near l_0, i.e. near the energy minimum, and does not correctly describe the formation or breaking of covalent bonds. In compression, the energy should rapidly tend to infinity to avoid the overlapping of the atoms, in accordance with the Pauli exclusion principle; in extension, the energy should instead decrease beyond a certain threshold, until it vanishes with the breaking of the bond.

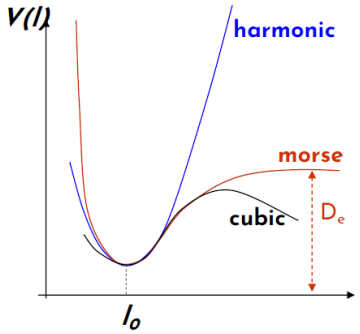


Figure 15: *Different models for bond terms: harmonic, Morse, and cubic potentials*

The relative motion of three atoms arranged in a plane defines an angle, which represents an interaction among the particles. This interaction can often fluctuate unpredictably, for instance due to molecular collisions. Such angular variations are typically modeled as harmonic interactions.

$$V(\theta) = \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \qquad \text{(Equation 3)}$$

- K_θ: force constant
- θ_0: reference bond angle, assumed by the bond without considering the other forcefield terms.
- θ: equilibrium bond angle, that minimizes V when all the terms are considered.

In reality, the angle between three atoms is subject to vibrational motions that depend on temperature, causing potential experimental errors in the determination of the reference length.

The dihedral angle describes the rotation between two planes defined by a system of four consecutively bonded atoms. In particular, it represents the rotation of the fourth atom with respect to the plane formed by the first three. The energy contribution associated with this torsion takes into account the steric effects between the atoms and is generally modeled as a sum of cosine terms.

$$V(\phi) = \sum_{dihedrals} K_\phi[1 + cos(n\phi - \delta)] \qquad \text{(Equation 4)}$$

- K_ϕ: energetic cost related to the angle deformation.
- n: multiplicity, the number of energetic minima along a 360° rotation.
- δ: phase, determines the minimum position for the torsional angle.

In the case of planar molecules, a deformation of the bonds with respect to the plane defined by the structure can occur. This motion is described by the improper dihedral angle, an additional term in the potential that energetically penalizes deviations from the expected planarity. The associated energy is generally modeled by a harmonic function.

$$V(\zeta) = \sum_{\substack{\text{improper} \\ \text{dihedrals}}} \frac{1}{2} K_\zeta[\zeta]^2 \qquad \text{(Equation 5)}$$

When two particles approach each other, a balance is established between attractive and repulsive interactions that influence the stability and behavior of the system. The study of these interactions falls within the scope of intermolecular forces, whose associated potential energy depends not only on the distance between the particles, but also on their mutual spatial configuration.
Non-bonded interactions are generally divided into electrostatic and Van der Waals potentials, and are usually modeled as functions inversely proportional to the distance between two atoms.

Van der Waals interactions can occur between neutral particles and are generally short-range, becoming significant only when atoms are close enough to interact without forming chemical bonds. If the atoms get too close, the overlap of their electron clouds generates repulsive forces that prevent them from interpenetrating. Conversely, as the distance increases, the interaction strength gradually decreases until it vanishes. However, there exists an intermediate equilibrium distance corresponding to an energy minimum, toward which the particles naturally tend. These interactions affect all atoms in the system and become negligible beyond approximately 1 to 1.5 nm.
Van der Waals forces are classified into:

- Short-range repulsive interactions, responsible for the steric exclusion of atoms;
- Long-range attractive interactions (up to 1.5 nm), including London dispersion forces.

The Lennard-Jones potential is a widely adopted model for describing the physical behavior of these interactions.

$$V(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \quad \text{(Equation 6)}$$

σ and ε are parameters of the forcefield:

- σ: collision diameter, the minimum interaction distance at which the potential is zero.
- ε: the minimum value of the potential energy at equilibrium.

There are force fields, such as AMBER, that use a specific Lennard-Jones function with 12-10 exponents to model hydrogen bond interactions.

Charged particles are subject not only to Van der Waals interactions but also to an electrostatic potential. This is a long-range interaction that can occur between two particles up to distances of about 10 nm. The electrostatic interaction is calculated using Coulomb's law:

$$V(r_1, r_2, \ldots, r_n) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left[\frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \right] \quad \text{(Equation 7)}$$

- ε_0: vacuum permittivity
- ε_r: relative permittivity (dielectric constant)
- r_{ij}: distance between atoms i and j

Once all the terms have been defined, the total expression of the potential energy function can be written as follows:

$$\begin{aligned} V(r_1, r_2, \ldots, r_N) = & \sum_{bonds} \frac{1}{2} K_l (l - l_0)^2 + \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\phi \big[1 + \cos(n\phi - \delta)\big] \\ & + \sum_{\substack{\text{improper} \\ \text{dihedrals}}} \frac{1}{2} K_\zeta [\zeta]^2 + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left[\frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} + \frac{A(i,j)}{r_{ij}^{12}} - \frac{C(i,j)}{r_{ij}^{6}} \right] \end{aligned} \quad \text{(Equation 8)}$$

This equation consists of bonded and non-bonded terms. It enables the construction of a mechanical model of a molecular system, as the properties of each atom, their types, and the interactions between them are known. [36]

In molecular simulations, all the atoms of the system are contained inside a virtual box, which can assume different geometric shapes and is generally filled with solvent. To simulate a more realistic environment and avoid the artifacts introduced by artificial boundaries, Periodic Boundary Conditions (PBC) are applied, which replicate the box indefinitely in all directions. In this way, boundary effects are significantly reduced: particles that exit from one side of the box automatically re-enter from the opposite side, keeping the number of particles constant and ensuring a stable distribution of the dynamic properties of the system, such as velocity. However, the use of PBC is not without limitations.

One of the main issues concerns long-range interactions, especially when a particle, moving, could find itself interacting with its own periodic image. For van der Waals interactions, it is sufficient to define a sufficiently large box (greater than 1.5 nm) to avoid such overlaps, but in the case of charged particles, electrostatic interactions can still introduce significant errors, such as artificial repulsive forces. For this reason, it is essential to respect the Minimum Image convention, which requires that a particle must never interact with its replica.

Thus, while PBC effectively simulate an infinite and continuous environment, eliminating boundary discontinuities, they require careful definition of box dimensions to avoid introducing artifacts that could compromise the accuracy and reliability of the simulation.

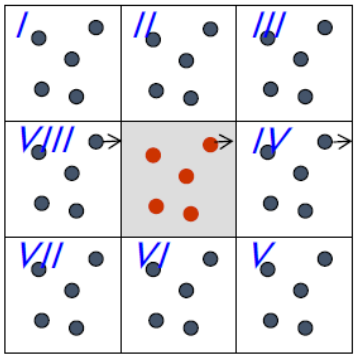


Figure 16: *Periodic box 2D*

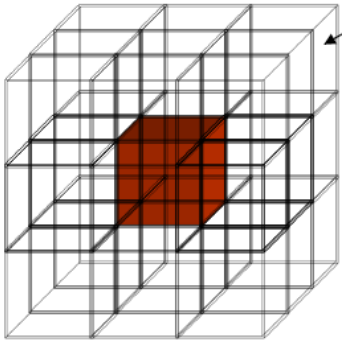


Figure 17: *Periodic box 3D*

In molecular modelling, the solvent, typically water, plays a crucial role in determining the structure and dynamics of biomolecules, helping, for example, to shield electrostatic interactions. To account for these effects, an explicit model can be adopted, directly including water molecules in the system, or an implicit model, which simulates the effect through approximations, such as modifying the dielectric constant or more complex models. The explicit model is more realistic, as it accurately describes phenomena such as solvation and the hydrophobic effect, but it is much more computationally expensive, mainly due to the large number of water molecules and the associated structural constraints. Implicit models, on the other hand, are more efficient but less accurate. [36]

Given the complex topology of the Potential Energy Surface, the system's behavior is strongly influenced by its location within the energy landscape. Regions corresponding to local minima are of particular interest, as they represent relatively stable states, while transition regions between minima may reflect possible structural or functional changes. Since configurations far from an energy minimum are unstable and subject to large forces, it's essential to reduce the potential energy of the system through an energy minimization process.

The energy minimization is a preliminary step that allows to eliminate structural stresses and prepare the system for reliable dynamic simulations. Starting a simulation from a non-minimized structure can cause numerical instability, non-physical collisions and the failure of the simulation itself, due to excessive accelerations induced by too strong forces.

Minimization identifies the local minimum closest to the initial configuration, although not necessarily the global one. Minimization methods, derivative or non-derivative, operate by iteratively modifying the coordinates of the system to progressively reduce the energy, without providing direct information on the thermodynamic properties, but preparing the system for the study of its dynamic evolution.

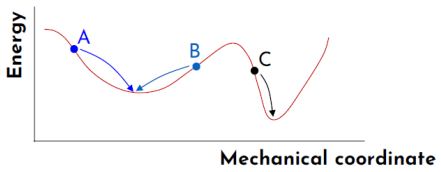


Figure 18: *Two-dimensional representation of a potential energy landscape as a function of mechanical coordinates. Points A and B represent local minima, while point C lies outside a minimum. Arrows indicate possible transitions along the energy surface.*

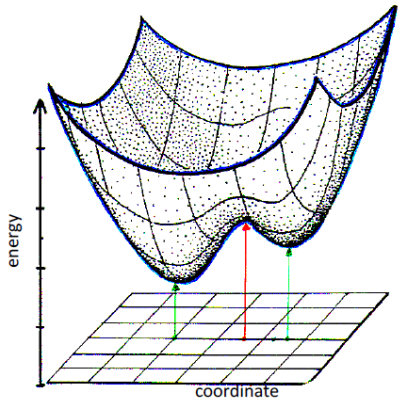


Figure 19: *Three-dimensional representation of a Potential Energy Surface (PES), showing how the potential energy varies with respect to atomic coordinates. The surface illustrates the existence of multiple local minima, separated by energy barriers.*

Derivatives of potential energy provide fundamental information for minimization: the first derivative, i.e. the gradient, indicates the direction of maximum energy variation and its intensity represents its local slope. Since the force acting on each atom is defined as the opposite of the gradient, the system configuration can be improved by moving the atoms along this direction to reduce the energy. Second derivatives, instead, describe the curvature of the surface and allow to identify stationary points. [37]

The Steepest Descent method exploits this property by moving in the direction of the net force, along the steepest descent of the energy landscape. Once the direction is established, it is necessary to determine the amplitude of the displacement. There are two main approaches: the Line Search, which estimates the minimum along the gradient by interpolating three points with a parabola, and a simpler but computationally efficient method, used for example in GROMACS [38], based on arbitrary steps: the position is updated according to the relation $x_{k+1} = x_k + L_k * S_k$, where S_k is the direction and L_k the step length. If the energy increases, it is assumed that the minimum has been exceeded and the step is reduced.

This algorithm is effective in the early stages, when the system is far from the minimum and the slopes are steep. However, near a local minimum it loses efficiency and may not detect intermediate minima if the step is too large. The choice of step length is therefore a trade-off between accuracy and speed: small steps increase precision but slow down convergence, while large steps speed up the process at the expense of stability. Minimization algorithms find single local minima without exploring the surrounding energy surface, thus providing only isolated points. This does not allow the computation of macroscopic properties, which instead require sampling of multiple configurations achievable by molecular dynamics methods. [37]

3.4.2 Molecular Dynamics

Molecular dynamics (MD) is a deterministic computational method used to study the temporal evolution of a system of atoms or molecules, by solving Newton's equations of motion. Starting from initial positions and velocities, the forces acting on the atoms, calculated as the gradient of the potential energy function, allow to determine accelerations, from which, by numerical integration, the trajectories of the system are obtained, i.e. the evolution over time of its atomic configurations.

Unlike minimization algorithms, which provide single stable configurations (local minima) without exploring the surrounding space, molecular dynamics allows continuous sampling of the phase space, making it possible to analyze the entire energy landscape and estimate the average properties of the system.

MD is based on the ergodic hypothesis, according to which, on sufficiently long time scales, the temporal average of a property corresponds to its average on the statistical ensemble. If a system evolves for a sufficiently long time, it will explore all the microstates compatible with its macrostate. In this way, the temporal average of a property along the trajectory of the system coincides with the ensemble average calculated on all the microstates. This equivalence allows to estimate macroscopic properties of the system starting from the dynamic sampling of its configurations.[39]

A statistical ensemble is a mathematical abstraction that represents the set of all microstates accessible to a molecular system under defined thermodynamic conditions. Each microstate is described by 6N variables (position and momentum for each of the N atoms). The choice of the ensemble is fundamental to calculate the macroscopic properties of the system. Different ensembles are used in molecular dynamics simulations, including:

- NVE (microcanonical): constant energy, volume and number of particles (isolated system);
- NVT (canonical): constant temperature, volume and number of particles;
- NPT (isobaric-isothermal): constant pressure, temperature and number of particles;
- μVT (grand canonical): constant temperature, volume and chemical potential (open system).

Individual microstates aren't significant on their own; to obtain meaningful information, one must consider averaged macroscopic properties.
The ensemble average of a property $\langle \mathrm{A} \rangle$ is computed by integrating all possible configurations weighted by their probability distribution:

$$\langle A \rangle = \int \int A(r,p)\, \rho(r,p)\, dr\, dp \qquad \text{(Equation 9)}$$

- r and p: positions and momenta of the particles;
- $\rho(r,p)$: probability density function of the ensemble.

The probability density function is:

$$\rho(r,p) = \frac{e^{\frac{-H(r,p)}{K_B T}}}{Q} \qquad \text{(Equation 10)}$$

- $H(r,p)$: Hamiltonian of the system;
- K_B: Boltzmann constant;
- Q: partition function.

The partition function Q is a dimensionless normalizing sum over the Boltzmann factors of all microstates of the system. It links microscopic thermodynamic variables to macroscopic state functions, providing a complete thermodynamic description of the system: once the partition function is known, the entire state space can be determined.

$$Q = \int \int e^{\frac{-H(r,p)}{K_B T}}\, dr\, dp \qquad \text{(Equation 11)}$$

The integral is generally challenging to compute, as it requires accounting for all possible states of the system. To obtain meaningful results, configurations must be sampled from the entire ensemble.

An alternative approach to computing an ensemble average relies on the Ergodic Hypothesis, which allows the property A to be evaluated as a time average. This involves generating system configurations sequentially over time, enabling the approximation of the ensemble average.

$$\langle A \rangle_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(r(t), p(t))\, dt \approx \frac{1}{M} \sum_{t=1}^{M} A(r, p) \qquad \text{(Equation 12)}$$

- t: simulation time;
- M: number of time steps in the simulation;
- $A(\rho, r)$: instantaneous value of A

The goal of a molecular dynamics simulation is to solve Newton's equations of motion in order to obtain the trajectories of the particles that is the evolution of their positions and velocities over time. This requires three fundamental components:

1. The initial positions of the atoms;
2. An initial distribution of velocities;
3. The acceleration, calculated as the gradient of the potential energy.

The process begins with the definition of the system's initial conditions, typically provided by a pdb file containing atomic coordinates, usually corresponding to those obtained after an energy minimization step. Each atom is then assigned an initial velocity, which can either be set to zero or randomly generated from a Maxwell-Boltzmann distribution at a specified temperature. This distribution gives the probability that an atom i has a velocity v_x in the x direction at a temperature T. [39]

$$p(v_{ix}) = \left(\frac{m_i}{2\pi K_B T}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\frac{m_i v_{ix}^2}{K_B T}} \qquad \text{(Equation 13)}$$

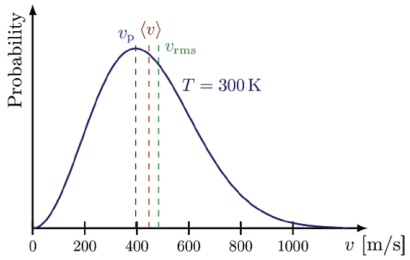


Figure 20: *Maxwell-Boltzmann distribution of particle velocities at 300 K. The curve shows the probability density as a function of velocity, illustrating the probability that particles have a given velocity at this temperature. [39]*

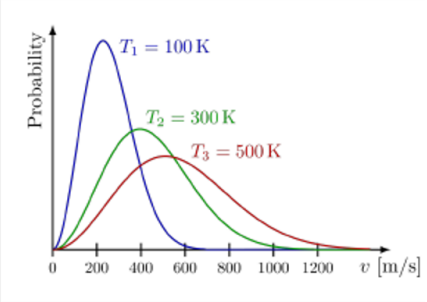


Figure 21: *Maxwell-Boltzmann distribution of particle velocities at temperatures 100 K, 300 K and 500 K. As temperature increases, the curves broaden and flatten, with the peak shifting toward higher velocities, indicating greater average and range of molecular speeds.[39]*

From the potential energy function, the forces acting on the system are computed as the gradient of the energy. Applying Newton's second law, the accelerations of the particles are determined. By integrating these accelerations over time, new velocities and positions are obtained, thereby generating a trajectory that describes the dynamic evolution of the system.

$$\begin{cases} \text{Newton's } 2^{\text{nd}} \text{ law} & F_i = m_i a_i = m_i \dfrac{d^2 r_i}{dt^2} \\ \text{Force as gradient of potential} & F_i = -\nabla_i V \end{cases} \quad \Rightarrow \quad -\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

Acceleration as the derivative of the potential energy with respect to the position r, for each molecule i:

$$a_i = -\frac{1}{m_i}\frac{dV}{dr_i} \qquad \text{(Equation 14)}$$

Due to the complexity of the function describing the system, the equations of motion cannot be solved analytically, but require numerical resolution. It is common to integrate Newton's equations numerically using approximate methods based on discrete increments of time, in a procedure known as integration. Integration algorithms must satisfy some fundamental criteria: they must conserve the energy and momentum of the system, be computationally efficient and allow the use of a sufficiently large time step. These methods are based on the assumption that positions, velocities and accelerations can be approximated by a Taylor series expansion.

Among the most used integration algorithms are:

- Leap-Frog calculates positions and velocities separately, which are staggered in time: this allows good control over energy conservation but does not provide both values simultaneously;
- Position Verlet uses information from previous positions and accelerations to estimate new positions, without directly calculating velocities;
- Velocity Verlet allows to obtain positions, velocities and accelerations simultaneously, ensuring greater precision and efficiency. This makes it one of the most used algorithms in molecular simulations.

The integration step dt, in molecular simulations, is generally between 1 and 2 femtoseconds. This value is chosen to be approximately one tenth of the shortest oscillation period present in the system, which usually corresponds to the vibrations of covalent bonds and the motion of hydrogen atoms. The use of a larger step would result in a significant loss of information and could compromise the stability of the simulation. [39]

Molecular dynamics is traditionally formulated in a Hamiltonian way, which implies the conservation of energy during simulation in microcanonical ensembles (NVE). However, in most cases, to represent a physical system more realistically, it is preferred to use canonical ensembles (NVT) or isobaric-isothermal ensembles (NPT). For this reason, the equations of motion must be modified by introducing thermostats and barostats, devices that allow to control the temperature and pressure of the system.

The thermostat is used to maintain the desired temperature of the system, acting on the control of the velocities of the particles, since the temperature is related to the average kinetic energy. A scaling factor λ is introduced on the velocities, which allows to switch instantaneously from an initial temperature to a new target temperature. However, this direct procedure can introduce dynamic instabilities and compromise sampling, as the particles can move too abruptly, risking losing important information on their positions. To overcome this problem, Berendsen proposed a 'weakly coupled' thermostat, which gradually heats the system by coupling with an external 'thermal bath' (reservoir). The temperature of the system evolves towards that of the reservoir with a dynamics characterized by a coupling time τ. In this way, the velocities are progressively scaled through a factor λ calculated from time to time, allowing the system to fluctuate around the desired temperature with less instability.

Similarly, the barostat controls the pressure by varying the volume of the system. Through coupling with an external reservoir, the volume is scaled to maintain constant pressure, adapting to the natural fluctuations and compressibility of the system. [39, 40, 41]

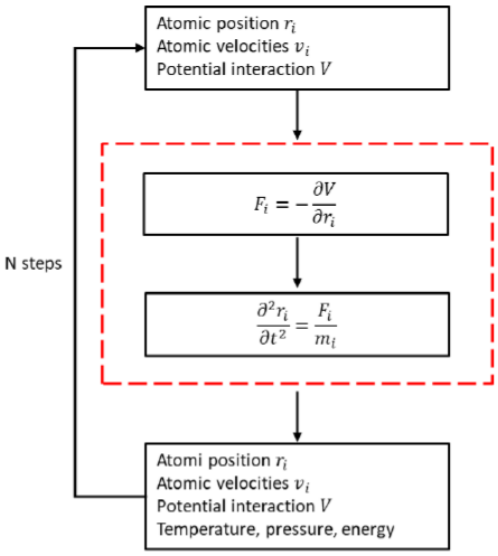


Figure 22: *MD flowchart. Starting from known atomic positions and velocities, the potential energy V is calculated, and the forces F_i are then obtained by deriving the potential energy function. Newton's equations of motion are subsequently integrated to determine the new atomic positions and velocities. This process is repeated iteratively until the system reaches an equilibrium state. [1]*

MD WORKFLOW	
Initial Coordinates	For proteins, the initial atomic configuration is typically derived from experimental data obtained through X-ray crystallography or nuclear magnetic resonance (NMR), and collected in databases such as the Protein Data Bank (PDB). PDB files contain detailed information on the molecular structure of proteins, but do not include any data related to the force field.
Energy Minimization	Once the initial coordinates and force field parameters are defined, the potential energy function of the system can be constructed. An energy minimization step is then performed to eliminate unfavorable interactions that lead to high potential energy and large atomic forces, which could compromise the stability of the simulation. The system iteratively explores the energy landscape by evaluating the gradient of the potential function and moving toward lower-energy configurations until a local minimum is reached.
Assign velocities	The output of the energy minimization is a new PDB file containing the updated atomic positions. Before starting the molecular dynamics simulation, it is necessary to assign initial velocities to the atoms. These can either be set to zero or randomly generated according to a Maxwell–Boltzmann distribution at the desired temperature.
Equilibration	During the equilibration phase, the system is brought to the target temperature (typically 300 K) and/or pressure (typically 1 atm). When an explicit solvent is used, harmonic position restraints are commonly applied to the heavy atoms of the protein. These restraints, modeled as additional potentials, act like springs anchored to fixed reference points to prevent large atomic displacements from the minimized structure. This helps maintain structural integrity and avoids denaturation during heating. The restraint term is added to the potential energy function, allowing the system to gradually reach thermal and structural equilibrium. Once a stable state is achieved, the restraints can be removed.
Production (MD)	Production phase of the MD simulation: Once the restraints are removed, the system undergoes molecular dynamics simulation for the desired time (e.g. 100 ns), and the full trajectory, containing atomic positions and velocities at each time step, is recorded. During this phase, the protein is free to move and explore its conformational space. Throughout the simulation, relevant properties are computed across the sampled states, and the final output is the complete trajectory describing the system's time evolution.
Analysis	Once the simulation is complete, the resulting trajectory can be analyzed to extract both ensemble and time-averaged properties of physical interest: Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration, Distance between terminal residues, Hydrogen bond analysis.

Table 3: *Workflow of Molecular Dynamics Simulation [39]*

3.5 Molecular Docking

Molecular docking is a computational technique widely used to study and predict the interaction between two molecules, typically a biological receptor such as a protein and a ligand, which can be a small molecule, a peptide or another protein. The goal is to identify a plausible binding mode, which reproduces the interactions that occur in physiological conditions, and to estimate key parameters such as binding free energy, affinity and stability of the complex formed. This approach is crucial in rational drug design, allowing to virtually select the most promising compounds before moving on to experimental testing, with consequent reduction in time and costs.
In addition to the pharmaceutical field, docking is also used in the study of protein-protein interactions (PPI), increasingly recognized as targets for small molecules, and in the characterization of peptide binding mechanisms, often difficult to study experimentally.

From a methodological point of view, the docking process is based on two fundamental components:

- a search function that explores the possible orientations and conformations of the ligand with respect to the receptor;
- a scoring function that evaluates the quality of each generated configuration (or "pose"), assigning it a score indicative of the strength and stability of the interaction.

The most common search strategies include:

- The systematic search, which exhaustively explores the conformational space, but is very computationally expensive;
- The geometric matching, which is based on the spatial complementarity of the functional groups between ligand and receptor;
- The stochastic search, which introduces random variations to efficiently explore complex configurations and allows the large-scale screening of molecular libraries.

The scoring functions can be based on physical force fields, empirical models or statistical approaches, and different strategies are often combined to improve precision and reduce false positives, through consensus methods. In relation to molecular flexibility, docking can be classified into:

- Rigid docking, where both ligand and receptor are considered fixed structures;
- Flexible ligand docking, which allows the ligand to conformationally adapt while keeping the receptor rigid, a widely used compromise;
- Fully flexible docking, which allows conformational changes in both molecules, ensuring the best representation of the binding but with high computational costs.

From the application point of view, there are specialized tools for different types of docking. AutoDock and its evolution AutoDock Vina, for example, are among the most used software for small molecule docking, exploiting genetic algorithms and scoring functions based on force fields. For peptide-protein docking, tools such as CABS-dock allow complete peptide flexibility and blind docking approaches without the need to know the binding site a priori. In the field of protein-protein docking, advanced AI-based technologies, such as AlphaFold3, use deep learning to model complex multimeric interactions even without known experimental structures, including other biological components such as nucleic acids, lipids and ions. ClusPro2, on the other hand, applies rigid docking followed by clustering of the best poses, evaluating shape complementarity and electrostatic interactions, also offering the possibility of defining preferential or excluded interaction regions.

Molecular docking is now a consolidated tool in the pharmaceutical industry and in biological research, used in multiple phases of drug discovery and design, from virtual screening to candidate selection. At the same time, there is a growing need for validated and shared datasets to facilitate comparison between methods and foster innovation.

Among the most relevant current challenges are the improvement of scoring functions to increase the reliability in the evaluation of binding affinity, as well as the more effective integration of protein flexibility and conformational adaptations induced by ligand-receptor interaction. Overcoming these limitations will be critical to fully exploit the potential of molecular docking in the discovery of novel therapeutic agents. [41, 42]

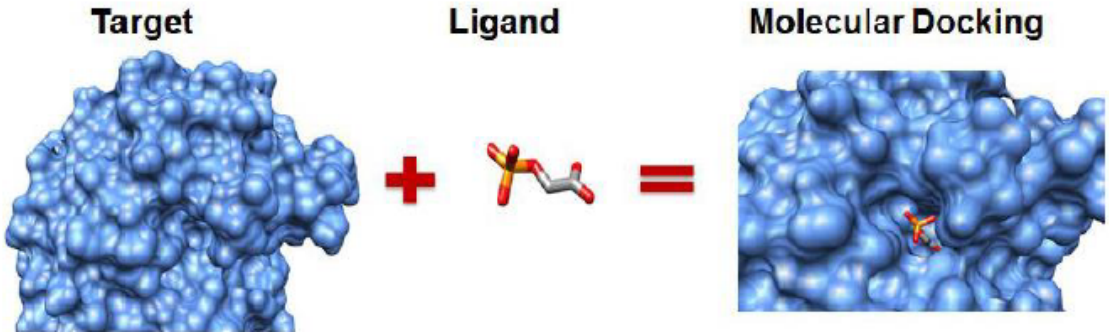


Figure 23: *Schematic representation of molecular docking: the ligand binds to the active site of the target*

3.5.1 ClusPro: protein-protein docking

ClusPro is a widely used web server for protein-protein docking, introduced in 2004 and subsequently enhanced. The simple interface allows basic use by inserting two PDB files and selecting between six different energy functions, depending on the type of interaction.

The docking process on ClusPro is carried out in three main steps:

1. Rigid sampling of billions of conformations using PIPER, a program that uses the Fast Fourier Transform (FFT) to efficiently evaluate the interaction energy between the receptor (fixed) and the ligand (mobile). The center of mass of the receptor is fixed at the origin and the ligand rotates (about 70000 rotations with a resolution of 5°) and translates (step of 1 Å) in space, generating up to 10^9–10^{10} configurations for an average-sized protein.

   The interaction energy is calculated as:

   $$E = w_1 E_{rep} + w_2 E_{attr} + w_3 E_{elec} + w_4 E_{DARS} \qquad \text{(Equation15)}$$

   - E_{rep} and E_{attr}: repulsive and attractive terms of van der Waals interactions;
   - E_{elec}: electrostatic energy;
   - E_{DARS}: structural potential based on the Decoys as the Reference State (DARS) approach, which mainly represents the desolvation energy, i.e. the free energy change associated with the removal of water from the protein interface.

   The coefficients w_1, w_2, w_3, w_4 are empirically optimized weights for different types of docking. ClusPro generates four sets of models with different scoring schemes: balanced (enzyme-inhibitor best), electrostatics-favored (double weight at w_3), hydrophobicity-favored (double weight at w_4), and van der Waals + electrostatics (without E_{DARS}).

2. Clustering of the 1000 lowest energy structures, using the interface root mean square deviation (IRMSD) as a measure. The structure with the largest number of neighbors within 9 ÅIRMSD is selected as the center of the first cluster, formed by all structures within this radius. Cluster members are removed, and the process is repeated with the remaining structures until a maximum of 30 clusters are formed.

3. Refinement: cluster centers undergo energy minimization for 300 steps using only the van der Waals term, with the receptor protein held fixed. This step removes any steric overlap, making only small conformational changes.

ClusPro returns as main results the centers of the 10 most populated clusters, favoring cluster size over the absolute energy values of individual structures. This approach indirectly accounts for entropic effects, since larger clusters indicate stable conformational regions and a higher likelihood of representing the correct conformation. It is assumed that the energies of structures within a cluster are similar and that the cluster population is proportional to the probability of correctness.
Although ClusPro also provides PIPER energy values for the cluster centers and members, these don't reflect the true binding free energy, as they neglect entropic contributions and are optimized to identify native-like structures rather than describe full thermodynamics. Therefore, model ranking is based solely on cluster population.

To evaluate the performance of ClusPro, we consider parameters such as:

- The number of complexes for which there is at least one cluster with structures within 10 ÅIRMSD of the native structure, indicating the ability to generate predictions close to the real.
- The average number of structures among the 1000 lowest energy structures with IRMSDs less than 10 Å, reflecting the density of correct predictions.
- The average value of the lowest IRMSD obtained for each complex, a measure of the improved accuracy achieved.

To calculate the IRMSD, the ligand interface residues (those within 10 Åof at least one receptor atom) are identified, the receptors are aligned, and the RMSD of the C_α atoms of these residues is calculated, thus focusing on the accuracy of the protein interface, the region most important for the function of the complex. [43]

Coefficient set	Energy term weight coefficients			
	E_{rep}	E_{attr}	E_{elec}	E_{DARS}
Balanced	0.40	−0.40	600.0	1.0
Electrostatic-favored	0.40	−0.40	1200.0	1.0
Hydrophobic-favored	0.40	−0.40	600.0	2.0
Van der Waals + electrostatics	0.40	−0.40	600.0	0.0
Others Mode, Set 1	0.30	−0.30	300.0	1.25
Others Mode, Set 2	0.50	−0.20	300.0	0.50
Others Mode, Set 3	0.50	−0.20	300.0	0.0

Figure 24: *Weighting coefficients of PIPER energy terms in various docking modes. [43]*

3.6 Binding free energy

Free energy is a fundamental thermodynamic quantity, particularly useful for predicting the spontaneity of processes. Its two main formulations are:

- Helmholtz free energy (A), suitable for systems at constant volume and temperature (NVT);
- Gibbs free energy (G), used for conditions of constant pressure and temperature (NPT).

However, the calculation of free energy is complex in systems such as liquids or flexible macromolecules, characterized by multiple energy minima separated by low-energy barriers. Even derived quantities such as entropy and chemical potential are difficult to estimate accurately.

In drug discovery, the affinity between a ligand and its receptor is crucial, as strong binding may indicate a promising drug candidate. The standard binding free energy (ΔG°) of a noncovalent complex is directly related to the association constant K_a, evaluated at standard temperature and pressure conditions. The accurate determination of binding free energy requires accounting for both enthalpic contributions,(such as electrostatic interactions, Van der Waals forces, and solvation effects), and entropic contributions, which reflect the conformational flexibility of the system.

$$\Delta G^\circ = -RT \ln K_a = \Delta H - T\Delta S \qquad \text{(Equation 16)}$$

- ΔH: enthalpy;
- ΔS: entropy;
- T: temperature;
- R: gas constant.

In molecular docking, scoring functions are commonly used to estimate the binding affinity through the binding constant K_b, or its inverse the dissociation constant K_d, both of which can be measured experimentally.

$$\Delta G = RT \ln K_d = \Delta H - T\Delta S \qquad \text{(Equation 17)}$$

with the kinetic model: $\mathrm{L} + \mathrm{R} \xrightarrow{k_{\mathrm{on}}} \mathrm{LR}$ $\quad \mathrm{L} + \mathrm{R} \xleftarrow{k_{\mathrm{off}}} \mathrm{LR}$ $\quad K_b = \frac{k_{\mathrm{on}}}{k_{\mathrm{off}}} = \frac{[\mathrm{LR}]}{[\mathrm{L}][\mathrm{R}]}$ $\quad K_d = \frac{1}{K_b}$

To estimate the binding energy between a ligand and a protein efficiently but accurately, approximate methods such as MM-GBSA and MM-PBSA are commonly used. These approaches offer a good compromise between accuracy and computational costs. [41]

In these methods, the binding energy is calculated as:

$$\Delta G_{bind} = G_{PL} - G_P - G_L \quad \text{(Equation 18)}$$

- G_{PL}: free energy of the complex;
- G_P: free energy of the unbound protein;
- G_L: free energy of the free ligand.

For each state (P, L, PL), the free energy is calculated as:

$$\Delta G = E_{MM} + Gsolv - TS \quad \text{(Equation 19)}$$

- E_{MM}: molecular mechanical energy (bonds, angles, torsions, electrostatic and Van der Waals interactions);
- G_{solv}: contribution of the solvation free energy, distinguished in polar and non-polar terms;
- TS: entropic contribution.

The solvent has a crucial impact on the bonding and is generally treated with implicit models to reduce the computational cost.

The polar contribution to solvation is obtained:

- in the MM-PBSA method, by solving the Poisson-Boltzmann equation:

$$-\nabla \cdot [\varepsilon(r)\nabla\phi(r)] = 4\pi\rho(r) \quad \text{(Equation 20)}$$

  - $\varepsilon(r)$:dielectric distribution function;
  - $\phi(r)$: potential distribution function;
  - $\rho(r)$: fixed atomic charge density.

- in the MM-GBSA method, using the generalized Born approximation:

$$\Delta G_{GB} = -\frac{1}{2}\left(\frac{1}{\varepsilon_{solv}} - \frac{1}{\varepsilon_{solute}}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{f_{ij}} \quad \text{(Equation 21)}$$

  - $q_i q_j$: atomic charges;
  - $\varepsilon_s olv$: solvent dielectric constant;
  - ε_{solute}: solute dielectric constant;
  - f_{ij}: depends on the interparticle distances and Born radii.

The nonpolar contribution is estimated via a linear relation with the solvent accessible surface area (SASA):

$$\Delta G_{nonpolar} = \gamma \times \text{SASA} + b \qquad \text{(Equation 22)}$$

- γ: microscopic surface tension;
- b: constant representing the cavitation term;
- SASA: accessible area, measured by a probe molecule.

The total solvation energy is given by polar and non-polar terms:

$$G_{solv} = G_{solv}^{polar} + G_{solv}^{nonpolar} \qquad \text{(Equation 22)}$$

The entropic term can be calculated with normal mode analysis (NMA), which evaluates vibrational modes to estimate the entropic change upon binding. However, this approach is computationally expensive and statistically less stable. Consequently, in many studies it is omitted, except when estimating the absolute free energy. [41]

4

SIMULATION SETUP

4.1 Structural prediction of wild-type APP, its mutants, and F-Spondin

In this thesis work, APP proteins, its mutations associated with Alzheimer's disease, and F-spondin were considered. The amino acid sequences in FASTA format were obtained from the UniProt database. As for APP, the functional domains E2, the juxtamembrane region (JMD), and the segments corresponding to Aβ and AICD fragments were taken into account. Consequently, the selected sequence includes the amino acids from residue 374 to 770, according to the numbering corresponding to the APP770 isoform (UniProt identifier P05067).[44] For F-spondin protein, the analysis focused on the reelin and spondin domains, known for their interaction with the E2 domain of APP. In particular, the residues between positions 29 and 388 were included (UniProt identifier Q9HCB6). [24] The three-dimensional structures of the models were predicted using the AlphaFold server, selected for its high accuracy in structural prediction based on amino acid sequences. [31] For the study of the Swedish (KM670/671NL), Iowa (D694N) and Iberian (I716F) mutations, the APP sequence was appropriately modified by inserting the corresponding point mutations. [14]

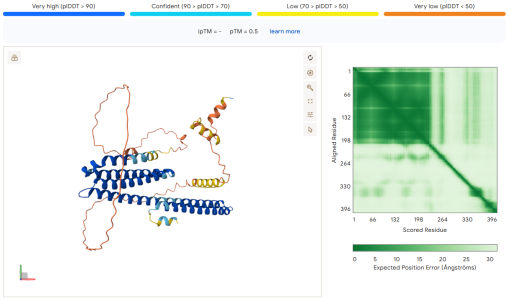


Figure 25: *AlphaFold3 prediction of the APP structure from residues 374 to 770, encompassing the E2 domain and the regions corresponding to the JMD, Aβ, and AICD fragments*

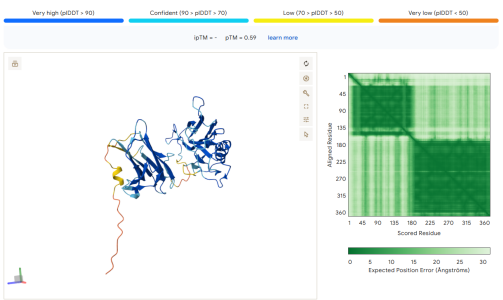


Figure 26: *AlphaFold3 prediction of the F-Spondin structure from residues 29 to 388, containing the Reelin and Spondin domains.*

4.2 Preliminary Molecular Dynamics setup

Preliminary molecular dynamics (MD) simulations were performed on all protein structures, aiming at structural stabilization and conformational optimization before performing docking studies. All simulations were performed on the computing clusters of the Digital Research Alliance of Canada, using the GROMACS software [38] The AMBER99SB-ILDN force field was used to describe molecular interactions. Each system was inserted into a cubic simulation box, with a minimum distance of 1 nm between the protein and the box edges. Periodic boundary conditions were applied in all directions. The system was solvated with an explicit TIP3P water model, and to ensure electrical neutrality, some water molecules were replaced with sodium (Na+) and chlorine (Cl-) ions, up to a concentration of 150 mM (0.15 mol/L), corresponding to the physiological concentration of sodium chloride in biological fluids.

Subsequently, an energy minimization of the system was performed using the Steepest Descent algorithm (integrator = steep). A tolerance on the maximum residual force (emtol) equal to 10 $kJmol^{-1}nm^{-1}$ was set as a criterion for stopping the process. The maximum step per atom (emstep) was set to 0.01 nm, to ensure the numerical stability of the minimization, while the maximum number of steps (nsteps) was set to 5000, in order to avoid an excessive duration of the procedure.

After the minimization phase, two equilibration simulations were performed with position restraint applied to the protein: an initial constant volume and temperature (NVT) equilibration, and a subsequent constant pressure and temperature (NPT) equilibration, both of 200 ps. The NVT phase was performed with a time step of 0.001 ps for 200,000 steps, using the velocity-rescale thermostat (tcoupl = v-rescale) with reference temperature set to 300 K, coupling constant (tau-t) equal to 0.1ps and separation of the thermal coupling groups between water and non-water. In this phase the initial velocities were generated according to a Maxwell-Boltzmann distribution. The NPT equilibration was then performed by keeping the v-rescale thermostat at 300 K, and coupling the pressure with the Berendsen barostat (pcoupl = berendsen), with isotropic coupling (pcoupltype = isotropic), coupling constant (tau-p) of 1.0 ps, compressibility equal to 4.5×10^5 bar^{-1} and reference pressure of 1.0 bar.

Finally, the dynamics simulation was performed for a duration of 5 ns, with a time step of 2 fs (dt=0.002 ps, nsteps = 2,500,000). Long-range electrostatic interactions were treated with the Particle Mesh Ewald method (coulombtype = PME), with a cutoff for Coulomb interactions at 1.0 nm. During this phase, the system was maintained at 300 K using the Nose-Hoover thermostat and at 1 bar using the Parrinello-Rahman barostat, which are both suitable for accurately reproducing the NPT ensemble. [41, 45]

4.3 Molecular Docking between APP, its mutants, and F-spondin

Molecular docking was performed using the ClusPro server. The protein structures in PDB format, obtained from previous molecular dynamics simulations, were provided as input. In particular, the wild-type APP protein and its three mutant variants: Swedish, Iowa and Iberian, were used as receptor. The F-spondin protein was used as ligand, in order to obtain four different protein complexes:

- Wild-type APP / F-spondin
- Swedish APP (KM670/671NL) / F-spondin
- Iowa APP (D694N) / F-spondin
- Iberian APP (I716F) / F-spondin

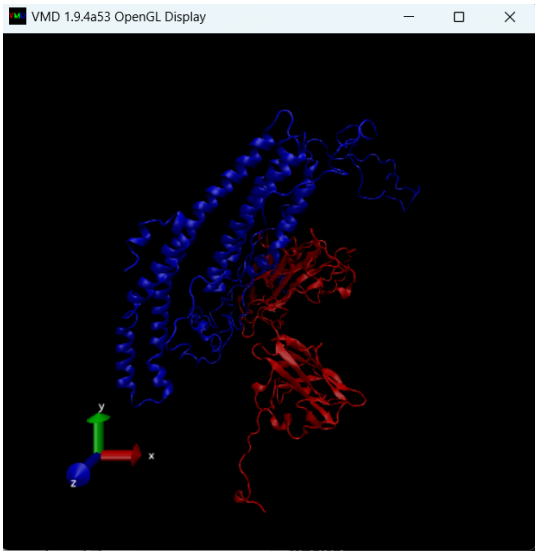
For this phase, a specific binding site was not defined a priori, but a strategy called blind docking was performed, which allows the software to explore the entire surface of the protein to identify possible interaction sites, without predefined constraints.

The ClusPro server returns as main result the centers of the 10 most populated clusters. The final choice of the complexes generated by ClusPro was made taking into account two main aspects. First, the interaction energy was considered, evaluated using the "balanced" weights calculation option, which takes into account in a balanced way the electrostatic, hydrophobic and desolvation contributions. Secondly, a visual inspection of the obtained models was performed, to identify those in which the two proteins showed a closer and more coherent interaction, i.e. with surfaces well-approached and potentially compatible with a biologically realistic interaction.

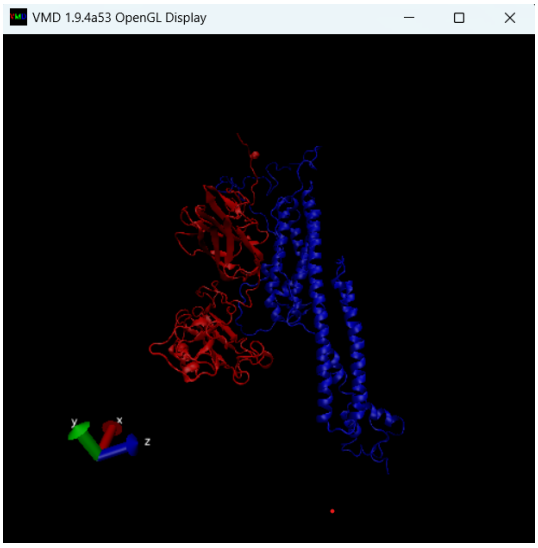
The interaction energy, calculated with the balanced weights provided by ClusPro, was determined according to the following formula, applied to all four protein complexes: [43, 46, 47, 48, 49]

$$E = 0.40E_{rep} - 0.40E_{att} + 600E_{elec} + 1.00E_{DARS} \qquad \text{(Equation 22)}$$

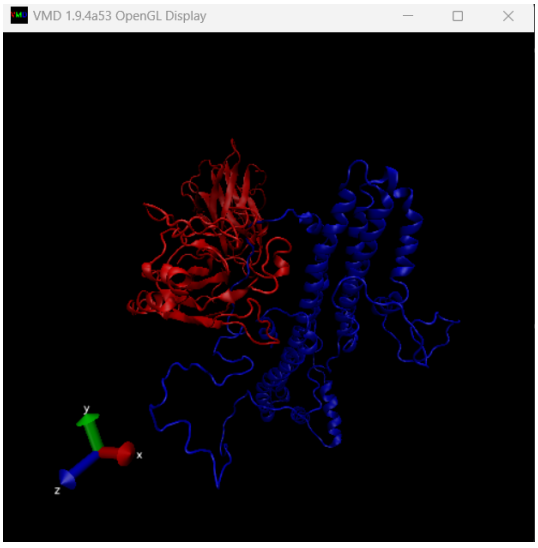
Below are the models chosen for each protein complex.



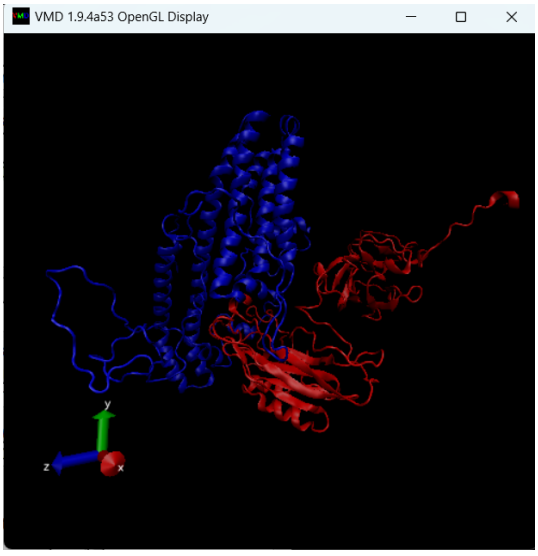
(a): Wild-type APP / F-spondin complex
Lowest Energy = -1074



(b): Swedish APP / F-spondin complex
Lowest Energy = -1034.5



(c): Iowa APP / F-spondin complex
Lowest Energy = -1103.9



(d): Iberian APP / F-spondin complex
Lowest Energy = -1100.8

Figure 27: *Protein complexes obtained by docking with ClusPro and visualized in VMD: in blue the receptor (APP and mutants), in red the ligand (F-spondin). [43, 46, 47, 48, 49]*

4.4 Molecular Dynamics simulation of docked complexes

Once the four protein complexes were obtained by docking, a 50 ns molecular dynamics simulation was performed to stabilize each complex, using GROMACS running on the Digital Research Alliance of Canada cluster.

The force field adopted was amber99sb-ildn. The system was inserted in a rectangular box, adequate to contain the entire complex, and solvated with a TIP3P water model. Neutralization was ensured by the addition of sodium and chlorine ions.

An energy minimization was performed with the steepest descent method (emtol = 10, emstep = 0.01), followed by two equilibration phases with position restraint: first in NVT ensemble for 200 ps with a V-rescale thermostat at 300 K and velocity generation, then in NPT ensemble for 200 ps with a V-rescale thermostat at 300 K and an isotropic Berendsen barostat at 1 bar.

The production simulation was run for 50 ns with a time step of 2 fs, using PME for the treatment of Coulomb interactions, Verlet scheme for cut-off, constraints on hydrogen bonds, Nose-Hoover thermostat at 300 K and isotropic Parrinello-Rahman barostat at 1 bar. [38, 45]

4.5 Analysis

Once the 20 ns molecular simulations of the protein complexes were completed, analyses were performed on the obtained trajectories: Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), evaluation of the hydrogen bonds between the two proteins over time, calculation of the distance over time between the two protein interfaces, Ramachandran Plot and study of the secondary structures.

- RMSD is a parameter identifying the shape variation of the structure. It is the root mean square distance of the protein configuration with respect to another protein configuration taken as a reference.

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i(t) - r_{ref}\right)^2} \qquad \text{(Equation 23)}$$

- RMSF represents the root mean square distance that measures how much each residue fluctuates over time. In particular, it evaluates the variation off the position of each residue with respect to its average position calculated over the entire trajectory. The calculation of the RMSF allows to identify the most stable and flexible regions of the protein. For the analysis, the Cα atoms of each residue are considered.

$$RMSF = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(r_i(t) - r_{ref}\right)^2} \qquad \text{(Equation 24)}$$

- A hydrogen bond is a dipole-dipole interaction between a hydrogen atom, covalently bonded to a highly electronegative atom (nitrogen, oxygen, fluorine), and another electronegative atom. In molecular systems, hydrogen bonds are important because they allow the stabilization of structures. In this case, the aim is to evaluate how the number of hydrogen bonds between the two proteins varies along the entire trajectory.

- The Ramachandran Plot is a graph that represents the combinations of the torsional angles φ and ψ of a pair of residues linked by a peptide bond within the protein. There are allowed areas, with combinations of angles actually observed in protein structures, and disallowed areas, that contain combinations that probably generate steric hindrances that prevent the existence of those structures. The Ramachandran Plot provides information on the secondary structure of the residues: based on the combination of angles, helix or beta-sheet structures can be observed. Consequently, the graph allows to obtain an indication of the possible secondary structure of the protein. [41, 50]

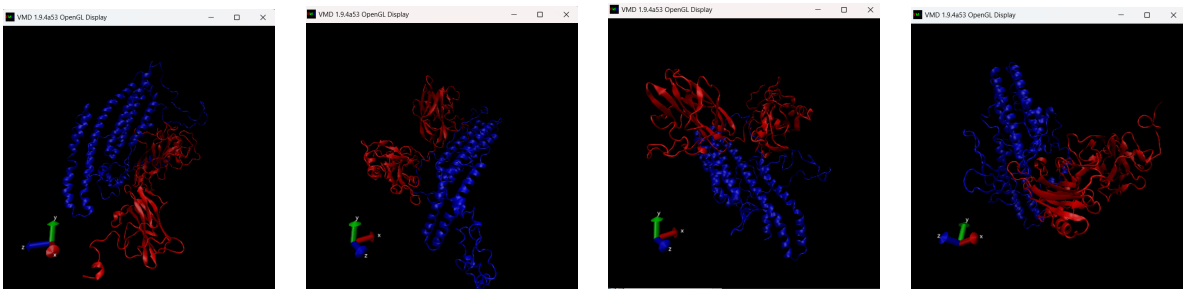
5

RESULTS

5.1 Final configurations of the complexes and analyses performed

The protein complexes considered for analysis are: wild-type APP–Fspondin, Swedish APP–Fspondin, Iowa APP–Fspondin, and Iberian APP–Fspondin. After docking, each complex was subjected to a 20 ns molecular dynamics simulation, with the exception of the Swedish APP–Fspondin complex, for which the simulation was stopped at 15 ns. The final configurations of the complexes at the end of the respective simulations are shown below (Figure 28).

In all four complexes, F-spondin (in red) interacts with the E2 domain of APP (in blue), consistent with what has been reported in the literature.[25, 26] However, a visual comparison of the final configurations reveals differences in the orientation and extent of the interface between the different mutants. Specifically, in the wild-type and Swedish complexes, F-spondin appears to be arranged in a more orderly and aligned fashion along the long α-helix of APP (E2 domain). In contrast, in the Iowa and Iberian complexes, the interface is less defined and F-spondin appears more disorganized, suggesting that the mutations may influence the interaction pattern.



(a): *Wild-type APP* (b): *Swedish APP* (c): *Iowa APP* (d): *Iberian APP*

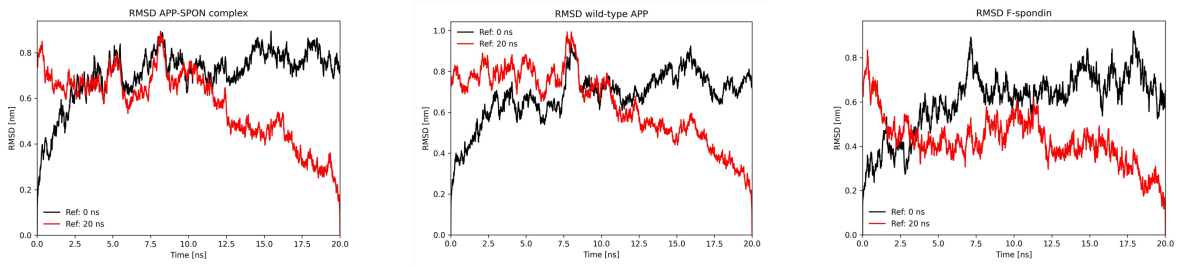
Figure 28: *Final configurations obtained from molecular dynamics simulations. The receptor (APP and its mutants) is shown in blue, while the ligand (F-spondin) is shown in red. Snapshots are visualized using VMD Configurations are taken at 20 ns for APP, Iowa, and Iberian complexes, and at 15 ns for the Swedish complex.*

To assess the structural stability and interactions between the two proteins within each complex, analyses of the simulation trajectories were conducted, including: calculation of the RMSD (Root Mean Square Deviation), the RMSF (Root Mean Square Fluctuation), the number of hydrogen bonds over time, and analysis of the evolution of secondary structures.

5.2 Trajectory analysis

5.2.1 RMSD

The RMSD (Root Mean Square Deviation) analysis of the APP–Fspondin complex shows how the system evolves toward a stable configuration over the course of the simulation. Specifically, the RMSD calculated with respect to the initial structure (reference at 0 ns) progressively increases, reaching a plateau around 0.8 nm, indicating a stable conformational change. The RMSD with respect to the final structure (reference at 20 ns) begins with a plateau at around 0.7 nm and tends to decrease over time, indicating an orderly transition and a progressive approach to the final conformation. This suggests that, even in the initial phases, the complex exhibits a certain structural coherence with the final state. The RMSD calculated for APP alone shows slightly higher values than the complex, with an initial plateau around 0.8 nm (reference at 20 ns), suggesting greater flexibility in the absence of F-spondin. In contrast, F-spondin rapidly reaches a plateau and maintains an ordered and stable behavior throughout the simulation. Overall, despite the structural evolution, the final conformation of the complex does not deviate significantly from the initial one. However, the system does not return to its initial configuration, but rather converges toward a new stable state, suggesting a conformational adaptation favorable to the interaction between the two molecules.

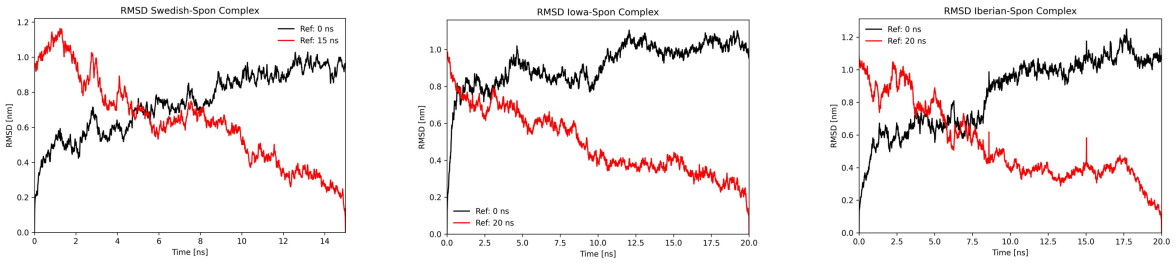


(a): *WT APP-Fspondin complex* (b): *Wild-type APP* (c): *F-spondin*

Figure 29: *Root Mean Square Deviation (RMSD) calculated with respect to the initial configuration (0 ns, black) and the final configuration (20 ns, red).*

RMSD analysis of the complexes with the mutations also shows an initial increase in values compared to the 0 ns structure, indicating a process of structural rearrangement.

- Among the complexes analyzed, Swedish–Fspondin exhibits the most marked conformational rearrangement. The RMSD compared to the initial structure rapidly reaches high values (0.95 nm), which then stabilize, indicating significant structural rearrangement and subsequent stability in the new configuration.
- The Iowa–Fspondin complex also follows a similar behavior, with a gradual increase in RMSD up to approximately 1.0 nm. However, in this case, the system reaches a sharper plateau, suggesting good structural stabilization at the end of the simulation.
- The Iberian–Fspondin complex exhibits an initially high RMSD (1.1 nm), followed by a more irregular and fluctuating behavior compared to the other mutants. Although a slight reduction is observed over time (with values dropping to 0.3–0.4 nm compared to the final structure), the complex appears less stable, with more persistent variations throughout the simulation.
- The wild-type APP–Fspondin complex exhibits more fluctuating and less defined behavior than the mutants, which instead demonstrate greater coherence and stability in the process of adapting to the final configuration.



(a): *Swedish complex* (b): *Iowa complex* (c): *Iberian complex*

Figure 30: *Root Mean Square Deviation (RMSD) calculated with respect to the initial configuration (0 ns, black) and the final configuration (20 ns, red).*

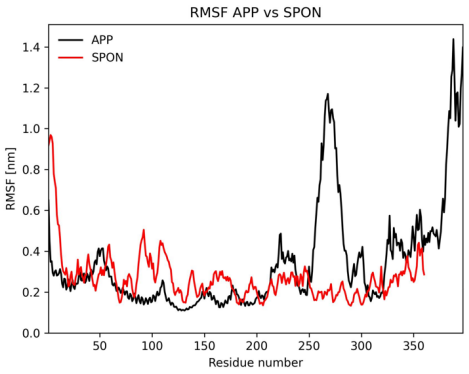
5.2.2 RMSF

Root Mean Square Fluctuation (RMSF) analysis was conducted separately on APP (wild-type and mutant) and SPON and revealed significant differences in structural flexibility profiles.

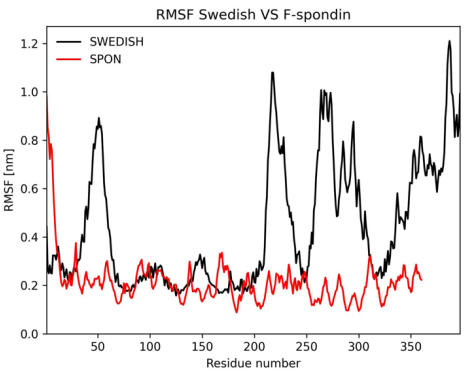
Note on residue numbering: RMSF plots were calculated considering an APP fragment spanning residues 374–770 of the full-length protein. Therefore, residue 1 in the plots corresponds to residue 374 of the original APP. All numerical references below have been corrected to the full-length protein numbering.

- Wild-type APP shows a distributed structural flexibility, with evident peaks at residues 654–674 and 724–754 (which in the plot correspond to the ranges 280–300 and 350–380, respectively, of the analyzed fragment). In particular, the high mobility of the C-terminal domain may be associated with the propensity for pathological aggregation.
- The Swedish mutation exhibits the highest flexibility values overall, with peaks at residues 424–474, 598–618, and 704–744 (equivalent to the 50–100, 230–250, and 330–370 ranges of the fragment). This behavior suggests marked structural destabilization and a potential tendency toward aggregation. The Iowa variant induces structural changes, with peaks observed around residues 574 and 654–674 (equivalent to approximately 200 and 280–300 of the fragment), indicating increased flexibility in new regions compared to the wild-type.
- The Iberian variant appears more stable. It exhibits a single moderate peak at residues 654–674, but with lower values than the wild-type. This behavior suggests local stabilization induced by the mutation.
- F-spondin (SPON) maintains very low flexibility in each complex, with RMSF values below 0.4 nm, confirming that it maintains constant structural rigidity. The Iowa mutation is an exception, where an isolated peak can be observed around residue 360, suggesting a possible disruption of the interaction or an increase in local mobility.

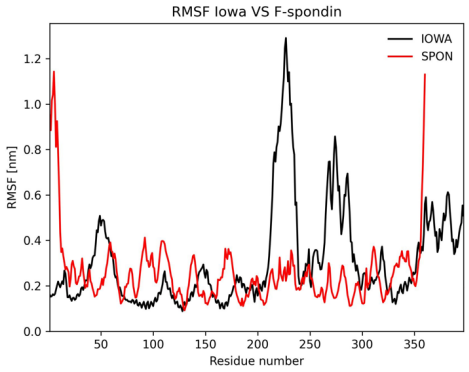
In summary, RMSF analysis highlights how the Swedish and Iowa mutations increase the structural flexibility of APP, with Iowa also appearing to influence SPON dynamics. In contrast, the Iberian variant exhibits more stable behavior. The structural integrity of SPON, preserved under most conditions, suggests a critical role in its molecular function.



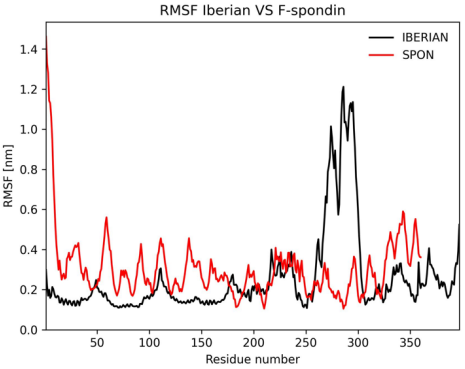
(a) Wild-type APP and SPON



(b) Swedish APP and SPON



(c) Iowa APP and SPON



(d) Iberian APP and spon

Figure 31: *Root mean square fluctuations (RMSF) of the four complexes: wild-type and mutant APP in black, F-spondin in red.*

PROTEIN	RMSF	CONSIDERATION
APP wild-type	High peaks at 280–300 and 350–380	Strong flexibility in C-terminal; prone to aggregation
SWEDISH	Strong peaks at 50–100, 230–250, 330–370	Highest overall flexibility; structurally destabilized
IOWA	Peaks at ∼200 and 280–300	Increased flexibility in new regions
IBERIAN	Moderate peak at 280–300	Lower mobility than WT; potential local stabilization
SPON (all)	Mostly <0.4 nm, stable; one peak ∼360 in Iowa	Consistently stable; structural rigidity may support binding

Table 4: *RMSF analysis and structural considerations of APP and its mutants*

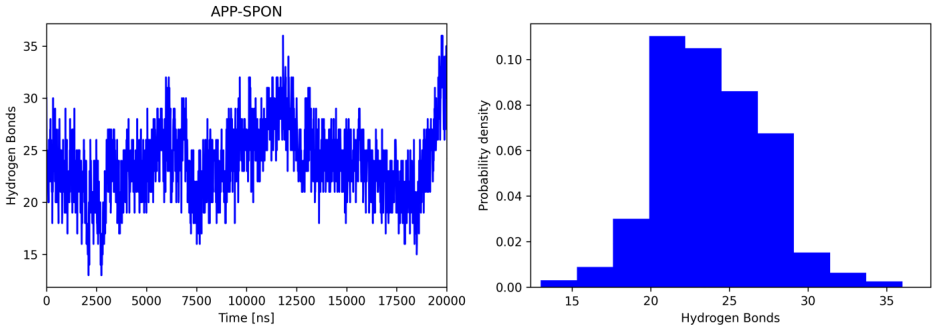
5.2.3 Hydrogen Bonds

Hydrogen bonds between APP, in its various forms (wild-type and mutant), and the F-spondin protein were analyzed.

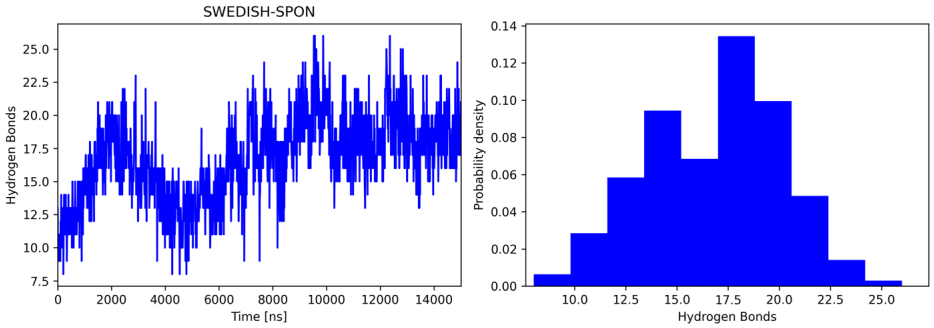
- The wild-type APP–SPON complex appears to be the most stable, with a high and constant number of hydrogen bonds over time, generally between 18 and 35, with a central distribution of around 23–25 bonds. This data indicates a strong and physiologically significant interaction, which can be considered a reference model for the binding between APP and F-spondin.
- The Swedish mutation shows a slight reduction in the number of bonds (between 10 and 25, with a center around 18–20), but the distribution remains regular, suggesting that the interaction with F-spondin, although slightly weakened, is essentially maintained. This suggests a still functional interaction.
- In the Iowa mutation, the hydrogen bonds are less stable and exhibit significant fluctuations. Although the number falls within a range similar to the Swedish mutation (10–25), the central distribution is lower (15–17) and the time trend is more irregular. This behavior reflects a weaker and more localized interaction, less stable overall.
- The Iberian mutation gives rise to the least stable complex: hydrogen bonds are few (between 5 and 18) and highly variable, with a central average of approximately 10–12. This indicates a fragmented and inefficient interaction between APP and F-spondin.

COMPLEX	H-BOND	DISTRIBUTION CENTER	CONSIDERATIONS
WT APP–SPON	18–35	∼23–25	Very stable complex, strong interaction
SWEDISH–SPON	10–25	∼18–20	Mutation slightly reduces interaction
IOWA–SPON	10–25	∼15–17	Noticeable fluctuations, less stable
IBERIAN–SPON	5–18	∼10–12	Weakest and least stable complex

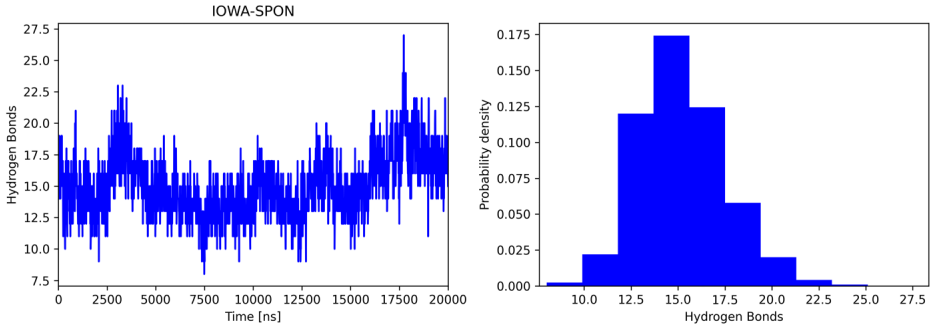
Table 5: *Quantitative analysis of hydrogen bonds between APP (wild-type and mutants) and SPON: observed range, distribution center, and structural considerations.*



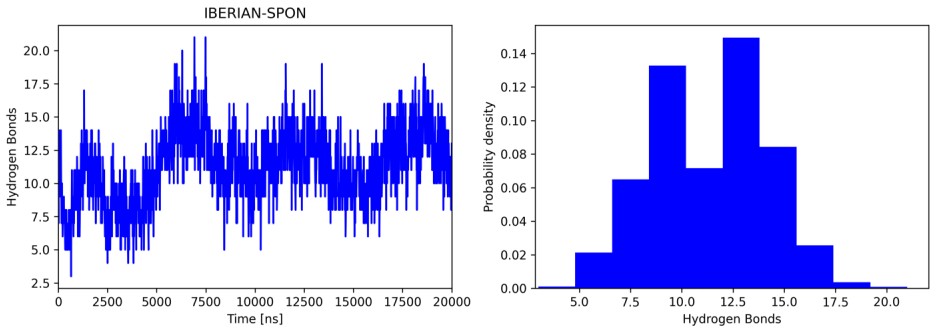
(a): *Wild-type APP*



(b): *Swedish APP*



(c): *Iowa APP*



(d): *Iberian APP*

Figure 32: *Number of hydrogen bonds between APP (wild-type and mutant) and F-spondin during the entire simulation (left graph). On the right, probability distribution of the number of hydrogen bonds..*

5.2.4 Secondary structures

The secondary structures of APP (wild-type and mutants) and F-spondin were analyzed throughout the simulation. Specifically, the APP fragment between residues 665–720 was analyzed, a very important region because it contains the Aβ sequence and the positions of the mutations studied.

- In the case of the wild-type APP protein, the structure remains fairly stable and ordered throughout the simulation. A strong presence of α-helices is observed in the 400–600 region, while the 665–720 region contains mainly coils (flexible regions) and a few short α-helices. This indicates good overall structural stability.
- The Swedish mutation does not cause major changes, but it does make the structure somewhat more flexible: more turns and coils are observed, and a slight reduction in α-helices, indicating a slightly less ordered structure.
- The Iowa and Iberian mutations, however, cause more pronounced effects: the structure loses most of the α-helices and increases the presence of β-sheets, which are often associated with protein aggregation processes, such as those occurring in Alzheimer's disease. In particular, the 665–720 region exhibits great structural variability, with fragmentation and alternation between coils, β-sheets, and other elements.
- Regarding F-spondin (SPON), which has a β-sheet-rich structure according to AlphaFold models, simulations confirm that in the presence of wild-type APP, the structure remains compact and stable. However, when SPON interacts with mutated versions of APP, especially the Iowa and Iberian variants, it displays greater flexibility and slight disorder, while maintaining its structural identity.

These observations suggest that mutations in APP can destabilize both the structure of APP itself and that of SPON, negatively impacting the interaction between the two proteins.

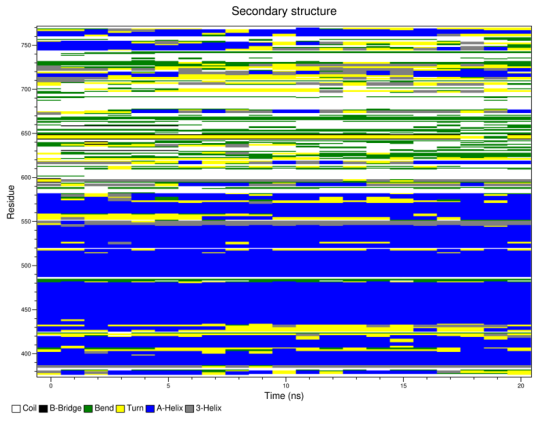
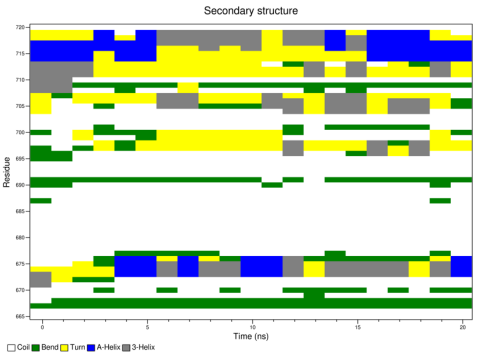
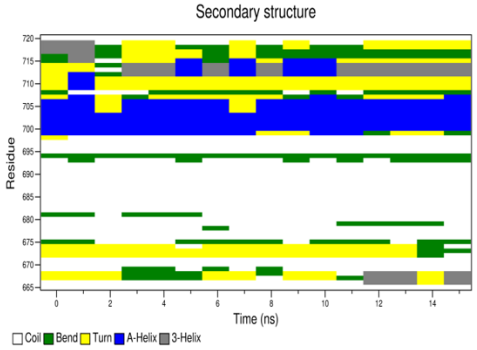


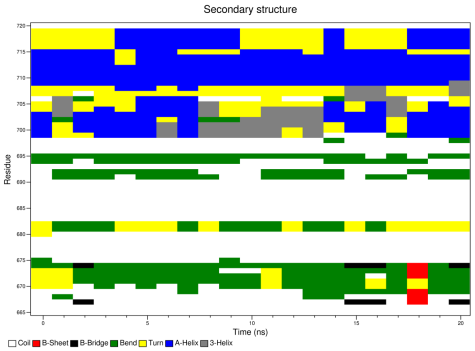
Figure 33: *Secondary structure of the wild-type APP protein analyzed during molecular dynamics simulation, from residue 374 to residue 770.*



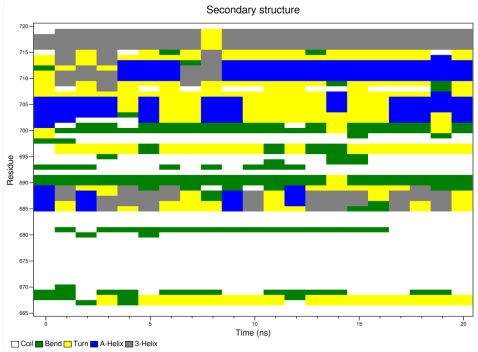
(a): *Wild-type APP*



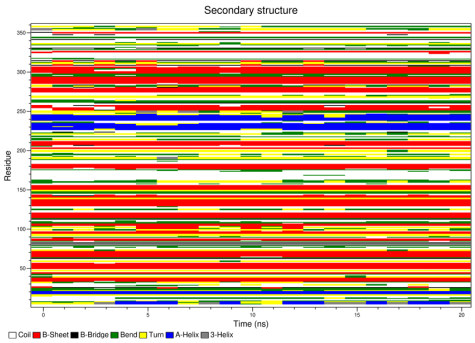
(b): *Swedish APP*



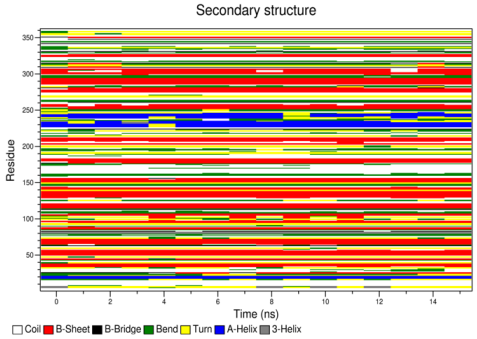
(c): *Iowa APP*



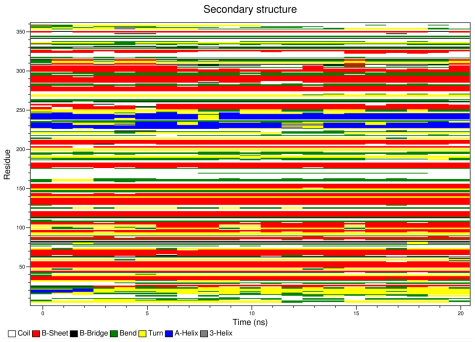
(d): *Iberian APP*



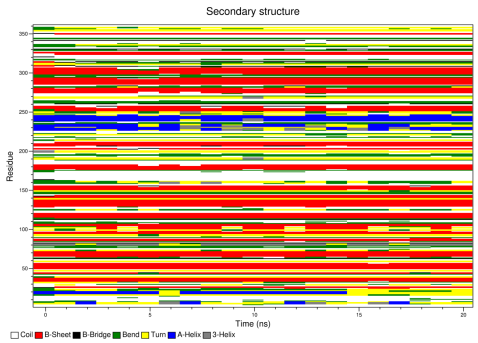
(e): *F-spondin in WT APP complex*



(f): *F-spondin in Swedish complex*



(g): *F-spondin in Iowa complex*



(h): *F-spondin in Iberian complex*

Figure 34: *Secondary structures of the APP fragment and its mutants (residues 665–720) (a,b,c,d), and of the F-spondin protein in the four complexes (e,f,g,h).*

5.3 Discussion

Trajectory analyses, obtained from molecular dynamics simulations, allowed to assess the stability of the interaction between the APP protein and F-spondin, also highlighting the influence of specific mutations associated with familial forms of Alzheimer's disease.

- The APP – F-spondin complex was found to be overall stable throughout the entire simulation. The RMSD values and the number of hydrogen bonds remained constant, suggesting good affinity between the two proteins. For this reason, the wild-type complex was used as a reference model for comparison with the mutated variants.
- The F-spondin protein showed a rigid and ordered structure, with minimal variations in all the complexes analyzed, largely maintaining its β-sheet structural motifs. This conformation is particularly evident in the complex with wild-type APP, where the structure appears more compact and well-defined.
- The Swedish mutation showed increased flexibility of the APP protein, without significantly affecting the interaction with F-spondin, although signs of structural destabilization were observed.
- The Iowa mutation led to a more marked structural reorganization, compromising the stability of both APP and F-spondin and weakening their interaction.
- The Iberian mutation was found to be the least stable complex among those analyzed. Severe structural fragmentation and an unstable interaction with F-spondin were observed.

From the structural analyses conducted, particularly on the APP fragment between residues 665–720 (which includes the Aβ sequence and therefore the positions of the mutations), it emerges that the mutations influence this region, favoring the formation of β-sheet structures, which are more prone to aggregation.

6

CONCLUSIONS

This thesis analyzed the Amyloid Precursor Protein (APP), its main familial mutations, and their interaction with the transmembrane protein F-spondin. The study was based on the amyloid cascade hypothesis, according to which the amyloidogenic processing of APP by β- and γ-secretases leads to the formation of the β-amyloid peptide. Numerous studies have shown that the latter is one of the main culprits in the pathogenesis of Alzheimer's disease, contributing to the formation of amyloid plaques and subsequently to the development of neurofibrillary tangles, neuronal death, vascular damage, and dementia.[6, 7, 9]

Currently, available treatments have a purely symptomatic effect and do not target the molecular mechanisms responsible for the disease. For this reason, research is focusing on strategies that aim to inhibit the production of β-amyloid and the formation of aggregates. A promising approach involves inhibiting β-secretase (BACE1), responsible for the initial amyloidogenic cleavage of APP. However, BACE1 is involved not only in the amyloidogenic process but also in other physiological processes, so its complete inhibition could have other side effects. Therefore, it is essential to identify strategies capable of selectively inhibiting only the binding between BACE1 and APP. Another alternative strategy is to prevent APP cleavage through interaction with specific ligands, which bind to the extracellular domain of APP and inhibit its binding to BACE1. In this context, recent studies have highlighted a possible role for the protein F-spondin as a natural ligand for APP. Specifically, F-spondin binds to the E2 domain of APP and appears to block access of β-secretase, thus inhibiting the amyloidogenic pathway. Furthermore, F-spondin acts as a link between APP and the ApoEr2 receptor: its Reelin and Spondin domains interact with APP, while the Thrombospondin domain binds to the ApoE receptor.[23, 25, 26]

This thesis also considered three familial APP mutations, all located near the cleavage sites:

- Swedish (KM670/671NL): near the β-secretase site;
- Iowa (D694N): near the α-secretase site;
- Iberian (I716F): near the γ-secretase site.

These mutations are associated with early-onset forms of Alzheimer's disease.[14, 22]

The molecular simulations conducted allow us to explore the structural and functional consequences of these mutations at the atomic scale, providing a mechanistic interpretation of the phenomena observed at the cellular level. Computational tools [42, 40, 38] were used to model and simulate the interaction between APP and F-spondin. An APP fragment (residues 374–770) containing the E2 domain, the Aβ sequence, and the intracellular AICD domain was chosen. For F-spondin, the fragment comprising residues 29–388, including the Reelin and Spondin domains, was selected.

The initial three-dimensional structures were obtained through AlphaFold prediction [31], starting from the amino acid sequences downloaded from UniProt [24, 44]. After short preliminary molecular dynamics simulations (5 ns) to verify the stability of the individual proteins, molecular docking was performed using the ClusPro server [46] to obtain protein complexes. Four complexes were analyzed: wild-type APP – F-spondin, Swedish APP – F-spondin, Iowa APP – F-spondin, Iberian APP – F-spondin. Each of these was subjected to longer molecular dynamics simulations (20 ns, except for the Swedish complex, which was simulated for 15 ns).

Structural analyses (RMSD, RMSF, number of hydrogen bonds, secondary structures) were performed on the obtained trajectories to compare stability and mutation-induced changes. [41, 50] The wild-type APP–F-spondin complex showed good stability and a well-defined structure. F-spondin maintained a rigid and ordered conformation in all complexes, with a well-conserved β-sheet organization, particularly in the complex with wild-type APP. The mutations showed significant effects on the secondary structure of APP, especially in the 665–720 region, where an increase in β-sheet structures was observed, associated with a greater tendency to aggregate. The interaction between APP and F-spondin was altered in the mutated complexes. Specifically, the Swedish mutation increases the structural flexibility of APP but maintains a good interaction with F-spondin; the Iowa mutation leads to a marked structural reorganization, which compromises the stability of both APP and F-spondin, weakening the interaction; the Iberian mutation causes significant structural fragmentation and a loss of interaction with F-spondin.

The data obtained therefore show that mutations, particularly Iowa and Iberian, can compromise the physiological interaction between APP and F-spondin. As reported in the literature, this interaction is essential for inhibiting APP cleavage by beta-secretase [23]. Its alteration could promote the activation of the amyloidogenic pathway and the subsequent formation of pathological aggregates, contributing to the neurodegeneration observed in Alzheimer's disease.

Bibliography

[1] Chiara Lionello. "The impact of amyloid beta assembly on membrane conformational stability and dynamics". http://webthesis.biblio.polito.it/id/eprint/10696. Master's Thesis. Politecnico di Torino. 2019.

[2] Panda PK et al. "Mutation-based structural modification and dynamics study of amyloid beta peptide (1-42): An in-silico-based analysis to cognize the mechanism of aggregation." In: *Genomics Data* (2016). doi:10.1016/j.gdata.2016.01.003.

[3] Scheltens P, Blennow K, and Breteler MM. "Alzheimer's disease". In: *Lancet* (2016). doi:10.1016/S0140-6736(15)01124-1.

[4] Liu E, Zhang Y, and Wang JZ. "Updates in Alzheimer's disease: from basic research to diagnosis and therapies". In: *Translational neurodegeneration* (2024). doi:10.1186/s40035-024-00432-x.

[5] Fuyuki Kametani and Masato Hasegawa. "Reconsideration of Amyloid Hypothesis and Tau Hypothesis in Alzheimer's Disease". In: *Frontiers in neuroscience* (2018). doi:10.3389/fnins.2018.00025.

[6] John A. Hardy and Gerald A. Higgins. "Alzheimer's Disease: The Amyloid Cascade Hypothesis". In: *Science* (1992). doi:10.1126/science.1566067.

[7] John Hardy and Dennis J. Selkoe. "The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics". In: *Science* (2002). doi:10.1126/science.1072994.

[8] Mohandas E, Rajmohan V, and Raghunath B. "Neurobiology of Alzheimer's disease". In: *Indian journal of psychiatry* (2009). doi:10.4103/0019-5545.44908.

[9] Sagar H. Barage and Kailas D. Sonawane. "Amyloid cascade hypothesis: Pathogenesis and therapeutic strategies in Alzheimer's disease". In: *Neuropeptides* (2015). doi:10.1016/j.npep.2015.06.008.

[10] Ricardo B. Maccioni et al. "The Revitalized Tau Hypothesis on Alzheimer's Disease". In: *Archives of Medical Research* (2010). doi:10.1016/j.arcmed.2010.03.007.

[11] Francis PT et al. “The cholinergic hypothesis of Alzheimer’s disease: a review of progress”. In: *Journal of neurology, neurosurgery, and psychiatry* (1999). doi:10.1136/jnnp.66.2.137.

[12] Chen Z-R et al. “Role of Cholinergic Signaling in Alzheimer’s Disease”. In: *Molecules* (2022). doi:10.3390/molecules27061816.

[13] de la Torre Jack. “The Vascular Hypothesis of Alzheimer’s Disease: A Key to Preclinical Prediction of Dementia Using Neuroimaging”. In: *Journal of Alzheimer’s disease* (2018). doi:10.3233/JAD-180004.

[14] Schilling S. et al. “Differential effects of familial Alzheimer’s disease-causing mutations on amyloid precursor protein (APP) trafficking, proteolytic conversion, and synaptogenic activity”. In: *Acta neuropathologica communications* (2023). doi:10.1186/s40478-023-01577-y.

[15] Chen J, Chen JS, and Li S. “Amyloid Precursor Protein: A Regulatory Hub in Alzheimer’s Disease”. In: *Aging and disease* (2024). doi:10.14336/AD.2023.0308.

[16] Chow VW et al. “An overview of APP processing enzymes and products”. In: *Neuromolecular medicine* (2010). doi:10.1007/s12017-009-8104-z.

[17] Müller T et al. “The amyloid precursor protein intracellular domain (AICD) as modulator of gene expression, apoptosis, and cytoskeletal dynamics-relevance for Alzheimer’s disease”. In: *Progress in neurobiology* (2008). doi:10.1016/j.pneurobio.2008.05.002.

[18] Chasseigneaux S and Allinquant B. “Functions of Aβ, sAPPα and sAPPβ: similarities and differences”. In: *Journal of neurochemistry* (2012). doi:10.1111/j.1471-4159.2011.07584.x.

[19] Chen GF, Xu TH, and Yan Y. “Amyloid beta: structure, biology and structure-based therapeutic development”. In: *Acta pharmacologica Sinica* (2017). doi:10.1038/aps.2017.28.

[20] Xu X. “γ-secretase catalyzes sequential cleavages of the AβPP transmembrane domain”. In: *Journal of Alzheimer’s disease* (2009). doi:10.3233/JAD-2009-0957.

[21] Carroll CM and Li YM. “Physiological and pathological roles of the γ-secretase complex”. In: *Brain research bulletin* (2016). doi:10.1016/j.brainresbull.2016.04.019.

[22] Liyong Wu et al. “Early-Onset Familial Alzheimer’s Disease (EOFAD)”. In: *Canadian Journal of Neurological Sciences* (2012). doi:10.1017/S0317167100013949.

[23] Ho A and Südhof TC. “Binding of F-spondin to amyloid-B precursor protein. A candidate amyloid precursor protein ligand that modulates amyloid precursor protein cleavage”. In: *Proceedings of the National Academy of Sciences of the United States of America* (2004). doi:10.1073/pnas.0308655100.

[24] In: *https://www.uniprot.org/uniprotkb/Q9HCB6/entry* ().

[25] Park SY et al. “SPON1 Can Reduce Amyloid Beta and Reverse Cognitive Impairment and Memory Dysfunction in Alzheimer’s Disease Mouse Model”. In: *Cells* (2020). doi:10.3390/cells9051275.

[26] Hoe HS et al. “F-Spondin interaction with the Apolipoprotein E Receptor ApoEr2 Affects Processing of Amyloid Precursor Protein”. In: *Molecular and cellular biology* (2005). doi:10.1128/MCB.25.21.9259-9268.2005.

[27] Wisniewski T. and Frangione B. “Apolipoprotein E: a pathological chaperone protein in patients with cerebral and systemic amyloid”. In: *Neuroscience letters* (1992). doi:10.1016/0304-3940(92)90444-c.

[28] Blumenfeld J et al. “Cell type-specific roles of APOE4 in Alzheimer disease”. In: *Nature reviews. Neuroscience* (2024). doi:10.1038/s41583-023-00776-9.

[29] Hoe HS and Rebeck GW. “Functional interactions of APP with the apoE receptor family”. In: *Journal of neurochemistry* (2008). doi:10.1111/j.1471-4159.2008.05517.x.

[30] Deriu Marco A. and Tuszynski Jacek A. “Introduction to Molecular and Multiscale Modeling in Bioengineering”. In: *Corso di Biomeccanica Multiscala, Politecnico di Torino* (2023/2024). Slide del corso.

[31] In: *https://alphafoldserver.com/about* ().

[32] Cramer P. “AlphaFold2 and the future of structural biology”. In: *Nature structural molecular biology* (2021). doi:10.1038/s41594-021-00650-1.

[33] Abramson J, Adler J, and Dunger J. “Accurate structure prediction of biomolecular interactions with AlphaFold 3”. In: *Nature* (2024). doi:10.1038/s41586-024-07487-w.

[34] Poltev V. “Molecular Mechanics: Principles, History and Current Status”. In: *Handbook of Computational Chemistry* (2015). doi:10.1007/978-94-007-6169-$8_9 - 2$.

[35] A R Leach. “Molecular Modelling: Principles and Applications”. In: *Pearson Education* (2001).

[36] Deriu Marco A. and Tuszynski Jacek A. “Molecular Mechanics: bond and non-bond interactions”. In: *Corso di Biomeccanica Multiscala, Politecnico di Torino* (2023/24). Slide del corso.

[37] Deriu Marco A. and Tuszynski Jacek A. “Molecular Mechanics – Energy Minimization”. In: *Corso di Biomeccanica Multiscala, Politecnico di Torino* (2023/24). Slide del corso.

[38] Van Der Spoel D et al. “GROMACS: fast, flexible, and free”. In: *Journal of computational chemistry* (2005). doi:10.1002/jcc.20291.

[39] Deriu Marco A. and Tuszynski Jacek A. “Molecular Dynamics”. In: *Corso di Biomeccanica Multiscala, Politecnico di Torino* (2023/24). Slide del corso.

[40] Vanommeslaeghe K, Guvench O, and MacKerell AD Jr. “Molecular mechanics”. In: *Current pharmaceutical design* (2014). doi:10.2174/1381612811319999060 0.

[41] Davide Messina. “Molecular mechanics simulations aimed at understanding the effects of amyloid precursor protein mutations on Alzheimer’s disease pathogenesis”. https://webthesis.biblio.polito.it/33991/. Master’s Thesis. Politecnico di Torino. 2024.

[42] Agarwal S and Mehrotra R. “An overview of Molecular Docking”. In: *JSM Chemistry* (2016).

[43] Kozakov D et al. “The ClusPro web server for protein-protein docking”. In: *Nature Protocols* (2017). doi:10.1038/nprot.2016.169.

[44] In: *https://www.uniprot.org/uniprotkb/P05067/entry* ().

[45] Brown, Anne M, and David R Bevan. “Molecular Dynamics Simulations of Amyloid -Peptide (1-42): Tetramer Formation and Membrane Interactions”. In: *Biophysical journal* (2016). doi:10.1016/j.bpj.2016.08.001.

[46] Jones G et al. “Elucidation of protein function using computational docking and hotspot analysis by ClusPro and FTMap”. In: *Acta Crystallogr D Struct Biol* (2022). doi:10.1107/S2059798322002741.

[47] Desta IT et al. “Performance and Its Limits in Rigid Body Protein-Protein Docking”. In: *Structure* (2020). doi:10.1016/j.str.2020.06.006.

[48] Vajda S et al. “New additions to the ClusPro server motivated by CAPRI”. In: *Proteins: Structure, Function, and Bioinformatics* (2017). doi:10.1002/prot.25219.

[49] Kozakov D et al. “How good is automated protein docking?” In: *Proteins: Structure, Function, and Bioinformatics* (2013). doi:10.1002/prot.24403.

[50] Daggett V Benson NC. “A comparison of multiscale methods for the analysis of molecular dynamics simulations”. In: *The journal of physical chemistry* (2012). doi:10.1021/jp302103t.