Politecnico di Torino

Master's degree in Biomedical Engineering

Academic year 2023/2024

December 2024 graduation session

Molecular mechanics simulations aimed at understanding the effects of amyloid precursor protein mutations on Alzheimer's disease pathogenesis

Thesis Supervisor:

    Prof. Jacek Adam Tuszynski

Co-supervisor:

    Dr. Lorenzo Pallante

Candidate:

    Davide Messina

Contents

Abstract

Alzheimer's disease (AD), the most common cause of dementia in older adults, despite decades of research, drug candidates and clinical trials, remains a challenging condition, both from a biological and clinical perspective. Apart from palliative drugs, which only slow symptoms, there is no approved reversal therapy; the impact of AD is therefore enormous on healthcare systems.

While most of AD cases are idiopathic, i.e. without a clear causative factor, the production and following the accumulation of aberrant amyloid beta peptides in patients' brains are considered as cornerstones of the disease and linked to the complex biochemical cascade, which is believed to start with excessive proteolytic cleavage by beta secretase (BACE1) enzyme.

In 2012, a mutation at site 673 (alanine to threonine) in the amyloid precursor protein (APP) was found in an Icelandic population sample; carriers of this mutation have a lower risk of AD up to 40% and a reduction in amyloid beta production was observed. However, a precise mechanistic explanation on its neuroprotective effect is still lacking, with evidence pointing to different cleavage or resulting amyloid oligomers behavior.

The A673V and the double Swedish mutations are conversely associated with early onset AD and thought to enhance cleavage by BACE1. Interestingly, the protective and these two pathogenic mutations reside near the beta cleavage site.

The aim of this M.Sc. thesis is therefore to study the APP-BACE1 complex using molecular docking and dynamics simulations, elucidating differences among mutated APP.

In this work, analyses of surface properties, contact probabilities, residue fluctuations and binding free energy estimation were performed using molecular mechanics modelling and simulations, including more than 1 microsecond of dynamics simulation for every system.

Wild type APP results in the most stable structure, while mutations alter its interactions pattern with the BACE1, especially in its flap region. MM-GBSA calculations showed that APP expressing the valine substitution (pathogenic) possesses higher binding affinities compared to wild type. In a comparable way, the protective A673T mutation appears to be a less favored ligand.

1 Introduction

The present chapter provides a general introduction of this master thesis work, summarizing the clinical features, the biological background, the aims of the research and the organization of the dissertation.

Neurodegenerative diseases are a heterogeneous group of debilitating disorders characterized by progressive neuronal dysfunction and loss, leading to a decline in cognitive (dementia) and motor (ataxia) abilities. Among these, Alzheimer's disease (AD) is the most common form of dementia, accounting for up to 70% of cases worldwide. As the global population continues to age, the societal and economic burden of AD is expected to increase in the future, making it a critical public health concern. Moreover, there are no therapies available that can reverse or halt the progression of the disease, and existing pharmacological interventions only offer palliative relief. Despite decades of intensive research and countless clinical trials, the development of effective disease-modifying therapies has faced significant challenges, mainly because of a poor understanding of AD molecular bases.

AD is usually divided into two main categories: late-onset AD (LOAD) and early-onset familial AD (EOAD). LOAD, also called "sporadic" AD, which constitutes the majority of cases, typically occurs after the age of 65 and has a poorly understood etiology involving a combination of genetic susceptibility, environmental, and lifestyle factors. LOAD accounts for up to 90% of AD cases. In contrast, EOAD is often associated with autosomal dominant genetic mutations and occurs in young people, even in their 30's. Mutations in these genes are believed to alter production, aggregation, or clearance of amyloid-beta (Aβ) peptides, which many argue is a central player in disease initiation and progression.

Although the precise causative factors of AD remain unclear, a much discussed but important hypothesis regards the accumulation of cytotoxic Aβ peptides as a key event in the disease process. These peptides are generated through sequential proteolytic cleavage of the amyloid precursor protein (APP) by β-secretase (BACE1) and γ-secretase. Aberrant cleavage and subsequent aggregation of Aβ peptides into oligomers and fibrils are believed to disrupt cellular homeostasis, leading to synaptic dysfunction and finally neurodegeneration. The identification of specific mutations in EOAD patients within the APP gene altering Aβ production and aggregation has been seen as one of the most solid proofs of Aβ involvement in AD.

In 2012, researchers identified a rare protective mutation, an alanine to threonine substitution at site 673 (A673T) of APP, in an Icelandic population sample, which was associated with up to a 40% reduction in the risk of developing AD and a more general higher cognitive level in the elderly. The A673T variant, which is located near the β-secretase cleavage site of APP, was also found to decrease the production of Aβ by approximately 50%, suggesting a direct link between amyloid enzymatic processing and AD. This mutation represented the first and only reported case of a protective mutation in the APP gene, resulting in extensive biochemical research -in order to understand its effects- and raising the possibility of gene therapy as a preventative measure of AD development.

Interestingly, residue 673 appears to be a critical site, as other mutations in this region, such as A673V and the Swedish double mutation (KM670/671NL), located at the BACE1 cleavage site, are associated with EOAD and are considered pathogenic. For these mutations, we have proofs that the Aβ production and processing profile is different. This suggests that changes in the amino acid sequence at or around this site can have contrasting impacts on AD risk, making possible a comprehension of otherwise unknown protective factors and then moving forward an exploitation of these effects to alter AD pathogenesis. Although the precise mechanistic effects of the A673T mutation remain unclear, it is hypothesized that A673T may influence APP processing by altering its interaction and affinity with the BACE1 enzyme, changing the biophysical properties, such as aggregation kinetics of the resulting Aβ oligomers, or even affecting downstream cellular pathways involved in amyloid metabolism.

Over the past years, advances in computational power and molecular modeling techniques have improved the study of biomolecular systems. Methods such as molecular docking and molecular dynamics simulations, combined with approaches like the Molecular Mechanics Generalized-Born Surface Area (MM-GBSA) method for binding energy calculations, have become used countless times for investigating protein-ligand interactions and the dynamic behavior of biomolecules. The so-called in-silico approach is also used in almost every step of drug design and allows for detailed analysis of molecular conformations and interaction energies; it is therefore considered as a valuable tool in biological and clinical research.

The aim of this M.Sc. thesis is to investigate the protective and pathogenic effects of different APP mutations through a molecular modeling approach. This study will employ molecular docking and dynamics simulations to explore the structural and functional implications of mutations at residue 673 and for the double Swedish mutation near the BACE1 cleavage site,

followed by MM-GBSA calculations to assess their impact on binding affinity and stability. The results of this study can contribute to a better understanding of the molecular mechanisms causing or protecting AD, providing a framework for in-silico investigation of AD related mutations and potentially aid in the development of therapeutic strategies.

The work is organized as follows.

Chapter 1 is the present general introduction.

Chapter 2 deals with a presentation of the clinical features of AD, the biological background including the hypotheses of disease initiation, the role of APP and its mutations possessing a protective or pathogenic potential.

Chapter 3 provides an overview of the computational methods used in this work. An introductory description of molecular modelling is provided, followed by a presentation of molecular docking, molecular mechanics and binding free energy calculation methods.

Chapter 4 is devoted to investigating the effects of APP mutations (A673T, A673V, Swedish mutation) occurring near the beta secretase site. Molecular docking is exploited to generate different APP-BACE1 complexes; then, molecular dynamics simulations are analyzed, focusing on structural changes, residue fluctuations and distance analysis. Finally, an estimated affinity for mutated APP is provided using binding free energy methods.

Chapter 5 serves as the conclusion of the dissertation, including some insights into future developments of this work.

2 Biological background

The present chapter provides a general overview of clinical, epidemiological and biological aspects of Alzheimer's disease. In greater detail, the common biochemical theories of disease pathogenesis are presented; moreover, a focus on amyloid precursor protein structure, functions, involvement in disease and known mutations is provided.

2.1 Alzheimer's disease overview

The concept of age-related cognitive decline has been known since ancient times. Thinkers such as Pythagoras and Galen described a condition known as "senility," which was believed to be an inevitable decline in cognitive function due to advanced age [1]. However, it wasn't until the early 20th century that this condition was recognized as a distinct entity, not present in most elders: thousands of years after Greek/Roman medicine findings, in 1906, the German neuroanatomist Alois Alzheimer identified a unique set of pathological brain changes (Figure 1) during the post-mortem examination of a patient named Auguste Deter. He observed the presence of abnormal protein aggregates, called "senile plaques," along with intracellular bundles known as "neurofibrillary tangles" (Figure 2). These two hallmarks of the disease were linked to the patient's clinical symptoms: progressive severe memory loss, language impairment, and disorientation.

Today, AD is recognized as the most prevalent form of dementia, accounting for up to 70% of all cases. The global prevalence of AD is staggering, resulting in more than 50 million individuals currently affected [2]. As a progressive neurodegenerative disorder, AD is characterized by profound cerebral atrophy and cognitive decline, ultimately resulting in the loss of autonomy and functional capacity.
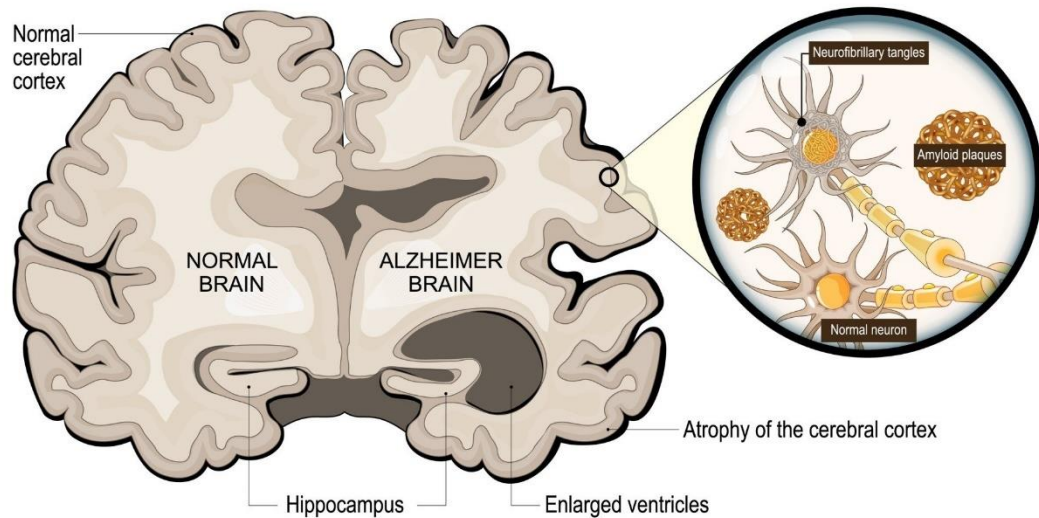
Alzheimer's disease

Figure 1 – A representation of changes in AD brain, in the circle on the right there is a depiction of neurofibrillary tangles and amyloid plaques [https://www.scmp.com/lifestyle/health-wellness/article/3279147/can-you-delay-alzheimers-how-new-dementia-screening-test-and-drug-may-help].

The natural course of the disease can be divided into several stages, each associated with distinctive clinical symptoms.

The prodromal stage, referred to as amnestic mild cognitive impairment (MCI), is marked by subtle memory deficits that exceed typical age-related forgetfulness but do not cause a significant impairment in daily functioning [3]. It is worth noting that MCI is not always followed by AD and offers a window of intervention to "shift" the course to the conservation of cognitive abilities. As the disease progresses into the early stage, patients experience memory loss and difficulties with language, the vocabulary shrinks, and confusion is present. In the middle stage, cognitive impairments become more severe, and neuropsychiatric symptoms such as irritability, wandering, and even aggressiveness may manifest. Usually, long-term memory is now affected, and patients could develop delusions or hallucinations. The late stage is characterized by apathy and the loss of bodily functions, patients are bedridden and entirely dependent on caregivers. Terminal lucidity, a temporary return of mental clarity (mostly, memory) before death, may occur in some cases [4]. Although the proximal cause of death is often due to secondary complications such as falls or pneumonia or other acquired infections, the underlying cause is the progressive inability of the patient to care for himself or herself.
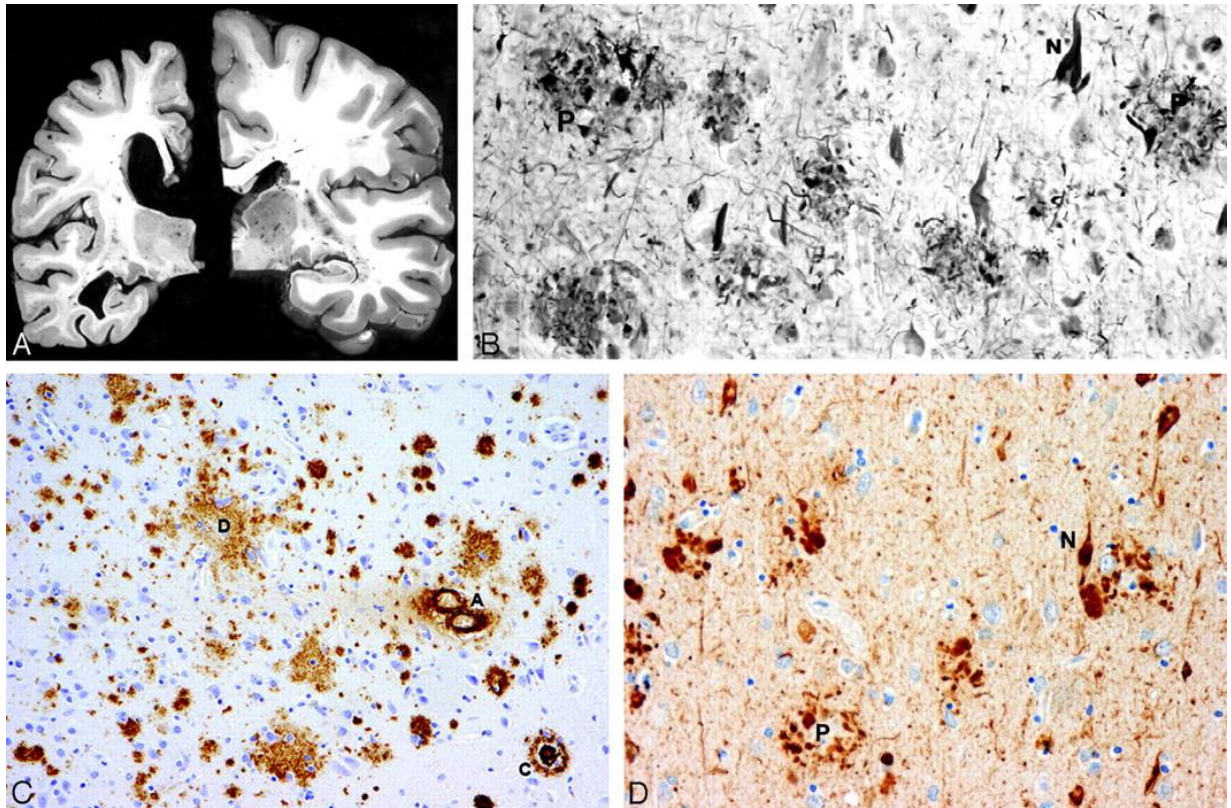
Figure 2 – Brain histopathological signs of AD. A: gross brain atrophy with enlarged ventricles compared to hemi-brain of an aged control case. B: senile plaques (P) and neurofibrillary tangles (N) in the hippocampus. C: immunohistochemistry of amyloid deposits in the frontal lobe. D: phosphorylated immunohistochemistry of plaques and tau clumps [https://www.ajnr.org/content/29/1/18].

The diagnosis of AD is complex and involves a combination of clinical expert evaluations and supportive diagnostic tools, while it should be considered there is no definitive consensus, and clinical practice is still standardizing the diagnostic methods [3]. Herein, we only describe the general guidelines, referring to external works for an in-depth analysis of AD diagnostics and clinical state-of-art.

According to the fifth edition of Diagnostic and Statistical Manual of Mental Disorders, (DSM-V), a diagnosis of *probable Alzheimer's disease dementia* can be made when the patient is a carrier of a well-known genetic mutation linked to AD, either from family history or genetic testing. Otherwise, a *possible Alzheimer's disease dementia* may be inferred when there is a steady and progressive decline in memory and cognitive abilities, verified through repeated standardized cognitive assessments like the Mini-Mental State Examination (MMSE) [5].

In addition to clinical evaluation, biomarkers are increasingly gaining importance in the diagnosis and understanding of AD. Amyloid PET scans and the measurement of amyloid and tau protein levels in cerebrospinal fluid (CSF) or plasma are minimally invasive procedures that help diagnosis and are used in research of AD before dementia onset. Tau PET scan is capable

of visualizing phosphorylated tau protein clumps in the brain, and it was demonstrated to better correlate with disease severity and progression than amyloid based biomarkers [6].

However, despite advances in imaging and molecular diagnostics, a definitive diagnosis of AD is only possible post-mortem during autopsy: the presence of phosphorylated neurofibrillary tangles, widespread amyloid plaques, brain atrophy, and enlargement of the brain ventricles is considered pathognomonic [3].

With a current prevalence of approximately 3 to 4% among individuals aged 60 years or older, 1% considering the whole population, AD is a high impact condition, with prevalence increasing with age, reaching more than 20% of people in their 80's [7]. The incidence of AD varies widely according to different reports, ranging from 2 to 16.8 new cases per 1,000 person-years, depending on demographic and regional factors [3]. What is more, as the world's population continues to age (Figure 3), both the prevalence and incidence of AD are projected to increase significantly in the coming decades, posing a substantial burden on healthcare systems worldwide. It is also worth to consider that, at present, dementia already ranks as the fifth leading cause of death, with a median survival of seven years from diagnosis. This translates to more than one million people affected by dementia in France and Italy alone, while the United States reach 5.8 million cases.

As stated above, the demographic "transition", i.e. the progressive population aging, stresses the urgency of addressing a possible AD "epidemic". By 2050, the global population is expected to reach 9.7 billion, with 15.9% of people aged 65 years or older, nearly doubling from 9% in 2019 [8]. Consequently, the number of AD cases in the United States is projected to nearly double, reaching 13.8 million by 2050; obviously, this will place unprecedented pressure on healthcare resources, given that AD is considered one of the costliest diseases: in 2022 alone, the healthcare costs associated with AD treatment were estimated at $321 billion, a figure expected to exceed $1 trillion by 2050 [9]. These economic costs are without doubt paralleled by societal and personal costs, as caregivers—often female relatives of patients—experience a diminished quality of life and productivity, mostly caused by the burden of long-term care and a lack of proper assistance.

Population pyramids, EU-27, 2019 and 2050
(% share of total population)
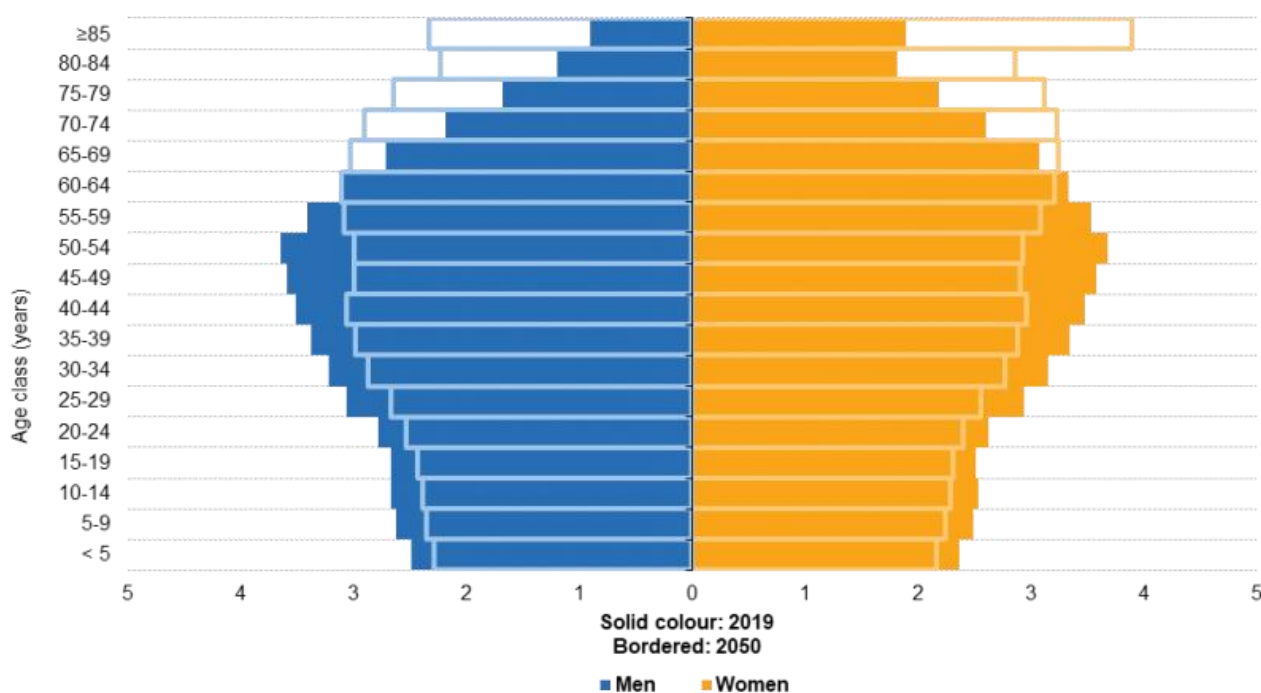
Note: all data as of 1 January. 2019: estimates and provisional. 2050: population according to the 2019 projections, baseline variant (EUROPOP2019).
Source: Eurostat (online data codes: demo_pjangroup and proj_19np)

Figure 3 – The diagram shows the population distribution in the EU-27 region in 2019 in solid color. Projected population in 2050 is depicted with bordered boxes [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Ageing_Europe_-_statistics_on_population_developments].

Decades of research have identified a range of risk factors associated with Alzheimer's disease, with age and genetics being the most significant [10]: thus, age is considered the strongest risk factor, with the probability of developing AD doubling approximately every five years after the age of 65. The role of genetics, particularly the presence of the apolipoprotein E (APOE) ε4 allele but also a more subtle polygenic influence, have been shown to be of primary importance, with a report stating that up to 70% of AD cases are of genetic nature [11]. Individuals carrying one copy of the ε4 allele have in fact a threefold increased risk, while those with two copies face up to a 15-fold higher risk of developing AD compared to non-carriers; another noteworthy figure is the presence of at least one copy in two thirds of AD patients [3], making this gene involved with cholesterol transport and processing not negligible.

Other non-modifiable risk factors include gender, as the women to man ratio is 2 to 1; this difference is partially attributed to longer female life expectancy, but evidence suggests that hormonal differences and psychosocial factors may also play a role [12].

On the other hand, modifiable risk factors (Figure 4) offer a window for intervention, via primary or secondary preventative measures: cardiovascular diseases, for instance, have been closely associated with cognitive decay and vascular/metabolic disfunctions such as hypertension, diabetes, midlife obesity, and hypercholesterolemia have been associated with an increased risk of developing AD and all act as independent additive risk factors [13]. Also, inactivity, both physical and mental is a modifiable risk factor and there is proof that a higher educational attainment is linked to a lower dementia prevalence; this was explained by the "cognitive reserve", the beneficial effect of education on brain coping mechanism, resulting in a delayed onset of thinking and behavioral deficits.

Finally, traumatic brain injuries, especially when repeated, represent a strong risk factor and therefore should be prevented [14]. A link between repeated head concussion and cognitive decline is also evident in the so-called *dementia pugilistica*, a condition in which a pattern of Alzheimer's and Parkinson's symptoms are present in *élite* combat sports athletes.

While irrefutable effectiveness of any intervention to prevent AD must be demonstrated, also given its multifactorial risk profile, it is clear that a holistic approach is necessary, encompassing different life domains such as diet, activity patterns and pharmacological treatments. As a relevant example, the Lancet Commission on Dementia Prevention, Intervention, and Care has recently discussed the potential effect of addressing 14 modifiable risk factors over the life course on dementia cases, speculating a reduction of up to 40% [15].

Figure 4 – Graphical summary of dementia modifiable risk factors [https://www.alzint.org/resource/dementia-risk-factors-infographic/].

The biochemistry of Alzheimer's disease (AD) remains elusive, despite lots of research efforts, and currently a coherent explanation of intra and extra-cellular changes leading to dementia is still lacking; in fact, different hypotheses have been proposed, each offering a distinct molecular mechanism causing dementia [16]. While several AD biochemical hypotheses are being made, and the observed hallmarks of the pathology range from protein misfolding to dysfunctions in mitochondria, here only the discussion of the cholinergic, amyloid and tau will be assessed. It is however important to acknowledge that the consensus is moving to the recognition of AD disease hallmarks, which are believed to better explain the complexity of cellular-tissual alterations, instead of proposing comprehensive cause-effect relationship theory [17].

The **cholinergic hypothesis** was historically the first attempt to link AD dementia (its "phenotypic" macroscopic feature) to biochemical changes in neuronal cells: emerged in the late 20th century, it postulated that a dysfunction of cholinergic pathway—especially in memory related brain areas such as the basal forebrain and hippocampus—was the primary cause of AD and could explain the main clinical symptoms and progressive disease course (Figure 5) [18]. Observations of decreased acetylcholine (ACh) levels - which is one of the main neurotransmitters- and extensive loss of cholinergic neurons in post-mortem AD brains are considered the strongholds of this theory; moreover, the development of cholinesterase inhibitors, which prevent ACh degradation after its release in the synaptic cleft, can temporarily rescue memory deficits and it is regarded as one of the few successes of AD understanding.

Despite their initial promise, these treatments cannot reverse or halt disease progression, and the cholinergic hypothesis is considered insufficient to coherently explain the causative factor of this neurodegenerative disease: the brain experiences significant cholinergic neurons loss in advanced states of disease and this is probably an effect of underlying alterations beginning much earlier [18].
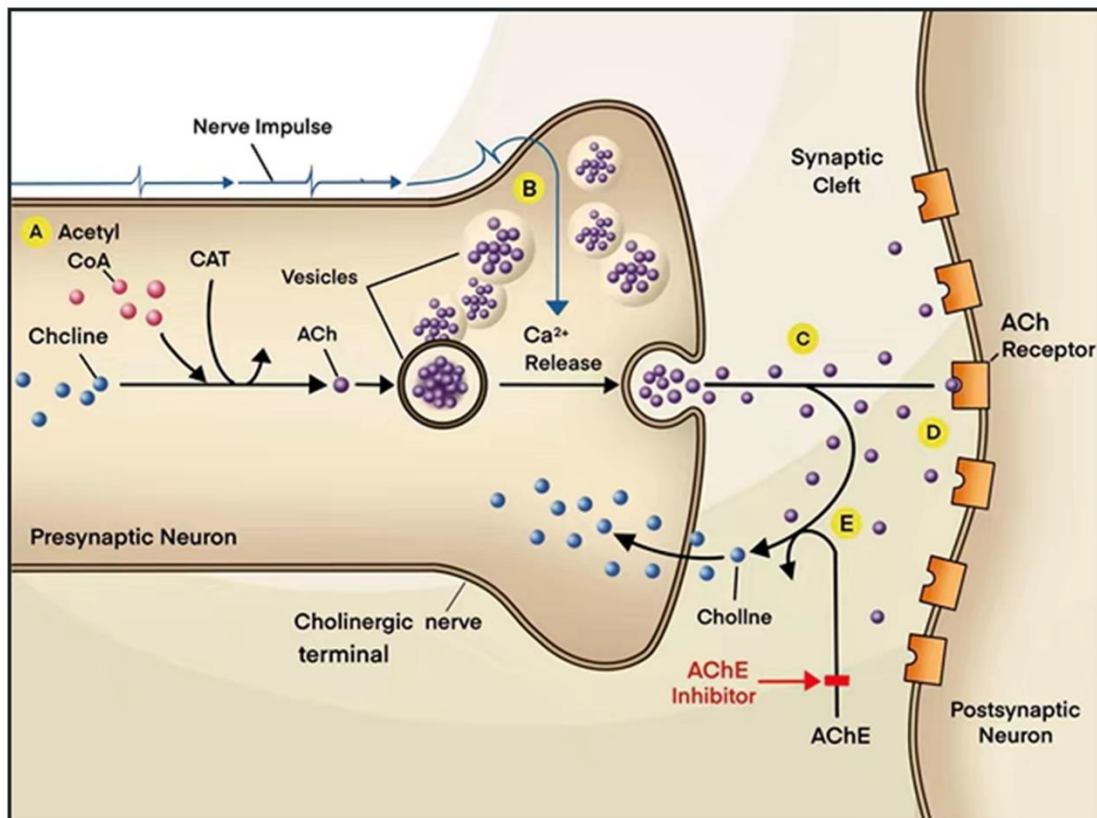

Figure 5 – Depiction of synaptic cleft showing the Ach role in neuronal transmission and the effect of cholinesterase inhibitor[https://www.mdpi.com/1420-3049/27/6/1816].

The limitations of the cholinergic hypothesis became visible, and the **amyloid hypothesis** rose to prominence at the start of the XXI century, postulating that the accumulation and aggregation of amyloid-β (Aβ) peptide species in the brain are the primal cause driving AD pathogenesis. These peptides, starting from a helical shape, misfold and aggregate, forming insoluble plaques deposited extracellularly, causing a disruption of neuronal environment and synaptic transmission (Figure 6); the toxicity was attributed to either the monomeric form, the oligomers and even the resulting macroscopic "senile" plaques [19].

Aβ peptides are generated through cleavage by BACE1 and γ-secretase of APP: the imbalance between Aβ production and clearance was therefore seen as the trigger of a cascade of downstream events explaining most of AD hallmarks, including oxidative stress, inflammatory

responses, synaptic dysfunction and neurons loss [17].

The majority of neurodegenerative diseases, although diverse in clinical presentation and course, share the widespread misfolding and aggregation of one of more protein: the term "proteinopathy" was then coined to describe pathological conditions where the incapability of the organism (the brain in this case) to correctly face the misfolded "stream" of proteins resulted in neuronal sufferance and loss. The prion-like behavior - to cause an unfolding cascade of correctly folded proteins of the same type- is observed in almost every neurodegeneration implicated protein [20].

In addition to this, autosomal dominant mutations in the APP gene and in the ones encoding the presenilin proteins (components of the γ-secretase complex) were associated with EOAD, further increasing the amyloid hypothesis validity.
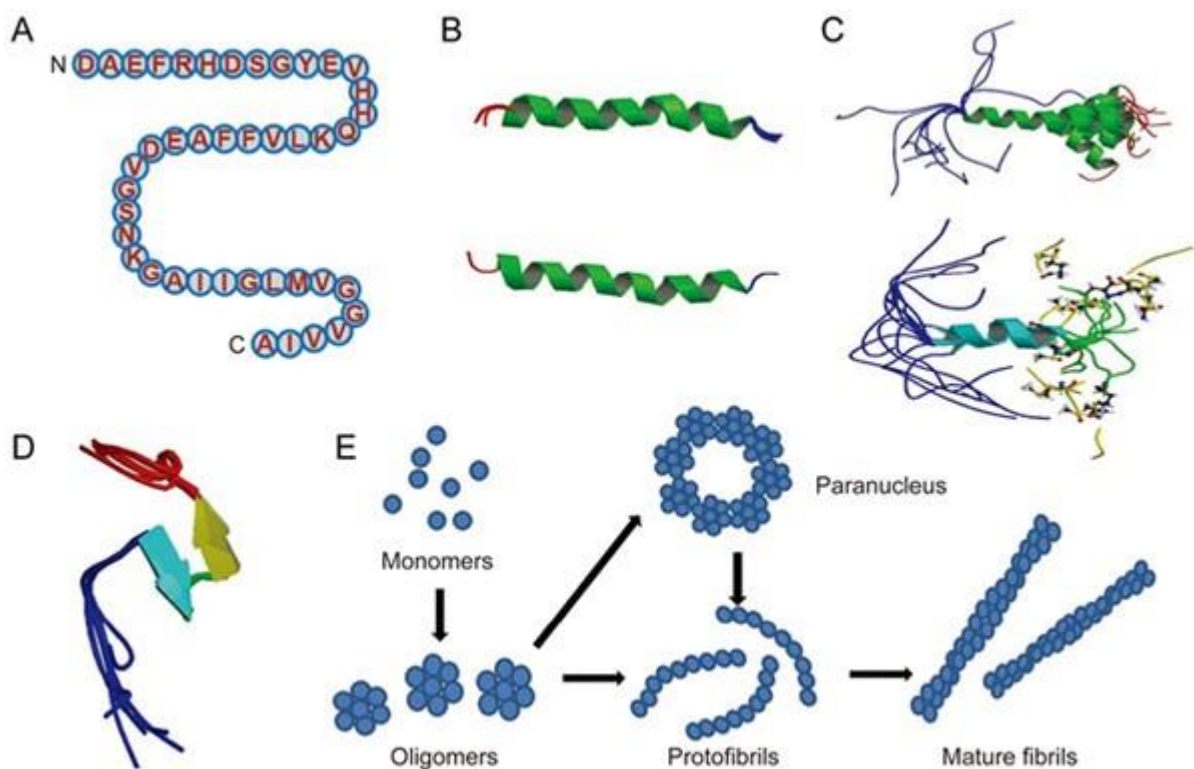

Figure 6 – The amyloid beta peptide sequence (A), its native secondary structure (B), the solution NMR, the transition from helix to beta sheet (D) making the peptide insoluble and its progressive aggregation into mature fibrils (E)[https://www.nature.com/articles/aps201728].

However, not all findings have supported the amyloid hypothesis and the precise causal relationship between deposition of Aβ and AD dementia is yet to be demonstrated: firstly, some patients with extensive amyloid plaque burdens do not manifest clinical dementia (i.e., they are amyloid-positive but cognitively normal), whereas others with cognitive impairment may exhibit minimal plaque deposition, as demonstrated in carriers of the Osaka deletion [21];

nevertheless, these observations could be compatible with the theory if the actual toxicity was ascribed to smaller, soluble Aβ oligomers, rather than the plaques themselves (See Figure 7 for a proposed biochemical *cascade* leading to macroscopic effects). Another weak point is the failure of most clinical trials with amyloid targeting drugs, aimed primarily at reducing production or increasing clearance, in providing clear cognitive benefits, while at the same time succeeded in the reduction of Aβ levels. An exception is the Aβ-targeting monoclonal antibody Aducanumab (Figure 8) [22].
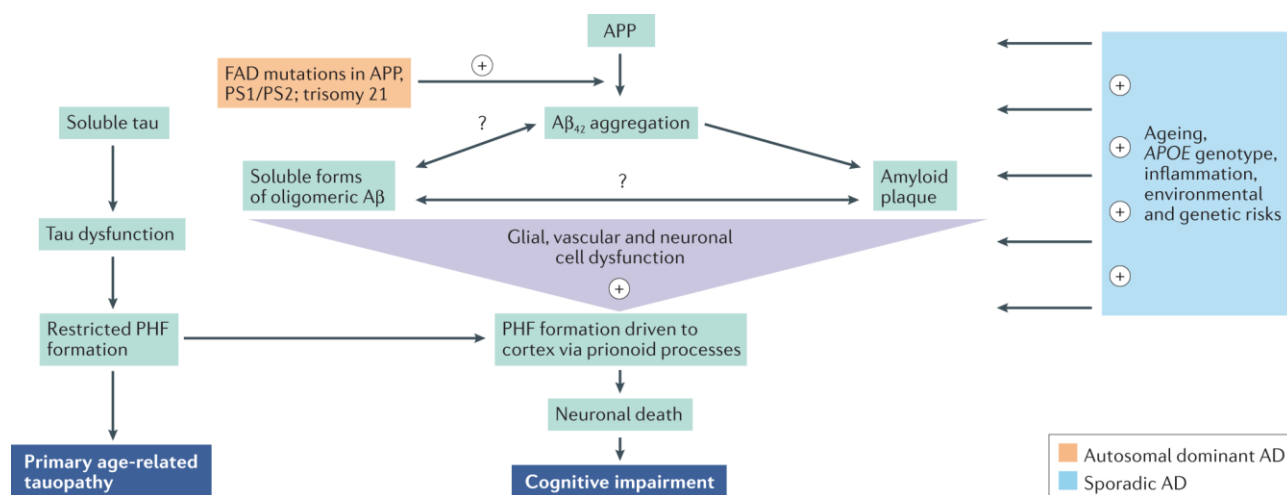
Figure 7 – A graphical summary of amyloid hypothesis: Aβ aggregation-clearance equilibrium is altered by genetic mutations (in EOAD) or several factors (in sporadic AD), resulting in neuronal dysfunctions leading to hyperphosphorylated tau deposition in brain cortex. The aberrant tau phosphorylation is speculated to proceed in parallel driven by other mechanisms[https://www.nature.com/articles/s41573-022-00391-w].
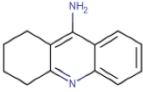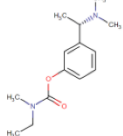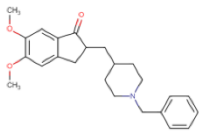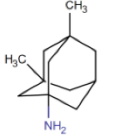

Figure 8 – Molecular formulas of approved AD drugs, belonging to the three classes: cholinesterase inhibitors, NMDA antagonist "Memantine" and the monoclonal antibody Aducanumab [https://www.mdpi.com/1999-4923/14/6/1117#].

Complementary to the amyloid hypothesis, the **tau hypothesis** has emerged as a parallel explanation for AD pathology, following some reports that studied the microtubule associated protein tau (MAPT) and its aberrant behavior in AD: MAPT is a stabilizing protein which was observed to become hyperphosphorylated during the course of AD; those changes diminishes its affinity for microtubules and causes the former to aggregate into neurofibrillary tangles (NFTs) inside neurons [23]. NFTs were demonstrated to disrupt intracellular vesicular transport, impair synaptic transmission and contribute to cell death (Figure 9).

The progressive spatial distribution of NFTs in the brain, as described by Braak stages (see below and Figure 10), correlates strongly with cognitive decline and disease progression, linking in a more reliable way than senile plaques the protein misfolding and aggregation with macroscopic cognitive impairment [24].
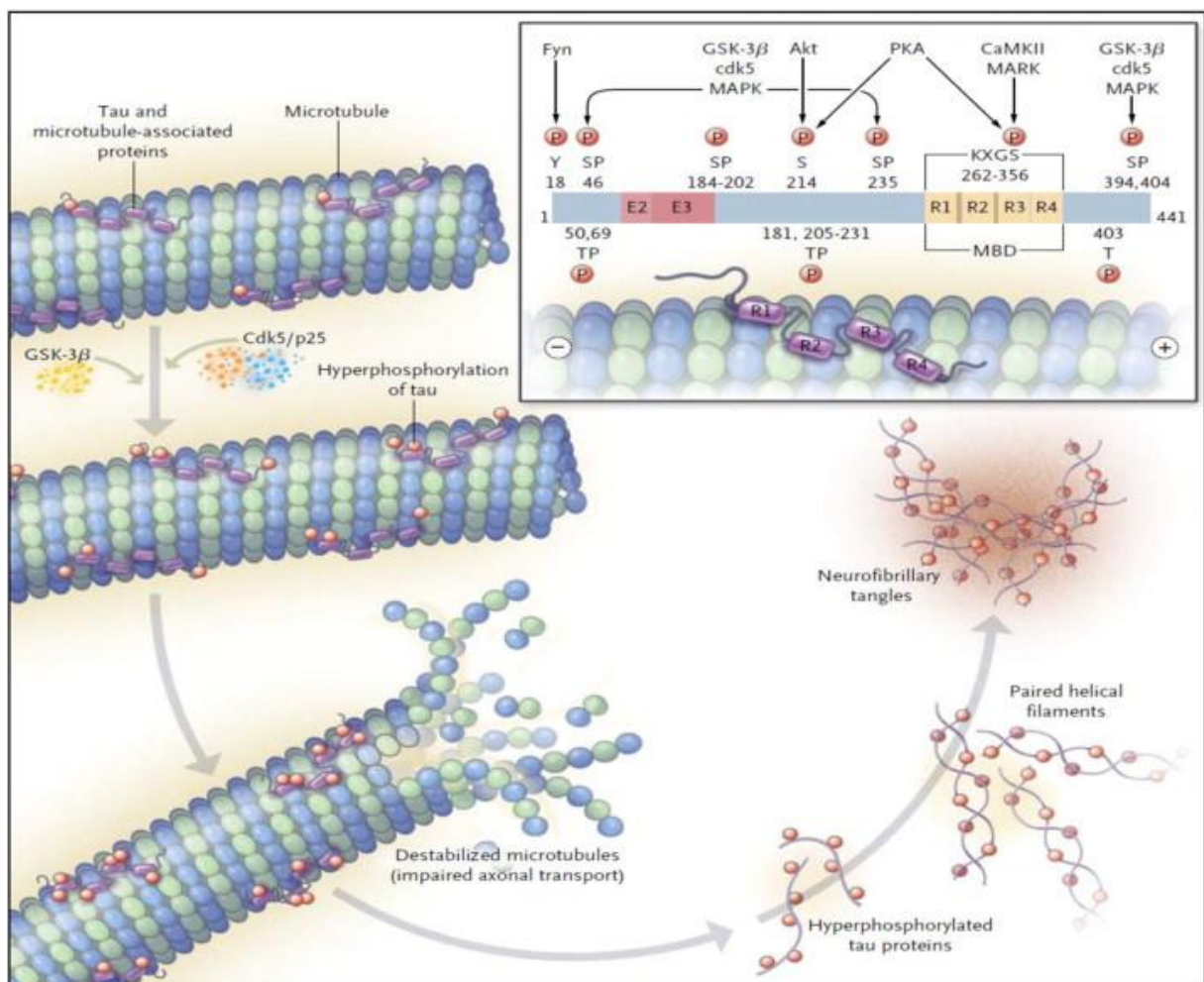

Figure 9 – Graphical depiction of excessive tau protein phosphorylation, which destabilizes microtubules leading to impaired axonal transport and aggregation of tau into neurofibrillary tangles; in the box on the right main sites of tau phosphorylation, as well as involved kinase, are shown [https://www.sciencedirect.com/science/article/abs/pii/S0143417915000657].

Heiko Braak was the first to describe this predictable pattern in 1991 and proposed a stadiation of AD in six stages, comparing the severity of dementia with the phosphorylated MAPT deposition

in the brain. Early-stage NFTs appear in the entorhinal regions, progressing to limbic areas, especially the hippocampus and, eventually, to cortical areas in the advanced stages (Figure 10) [25].

The relationship between Aβ and tau pathologies remains a topic of active research, and evidence suggests an integrated approach in the AD biochemical description, not considering those two theories mutually exclusive. Some studies suggest that amyloid plaques may precede and drive tau aggregation, while others point an "independence" of tau pathology, which can occur independently of amyloid presence in the form of Tau only cognitive decay, which is believed to be a distinct form of dementia [23].

While the most common studied and discussed theories on AD pathogenesis have been described above, a multitude of other molecular mechanisms have been implicated, including neuroinflammation with microglia activation, oxidative stress (mitochondrial damage), and vascular dysfunction; these AD hallmarks are again shared amongst neurodegenerative disease such Parkinson's disease and Huntington's Corea, raising the possibility that different clinical symptoms follow neurobiological changes and damages in particular brain areas [16].
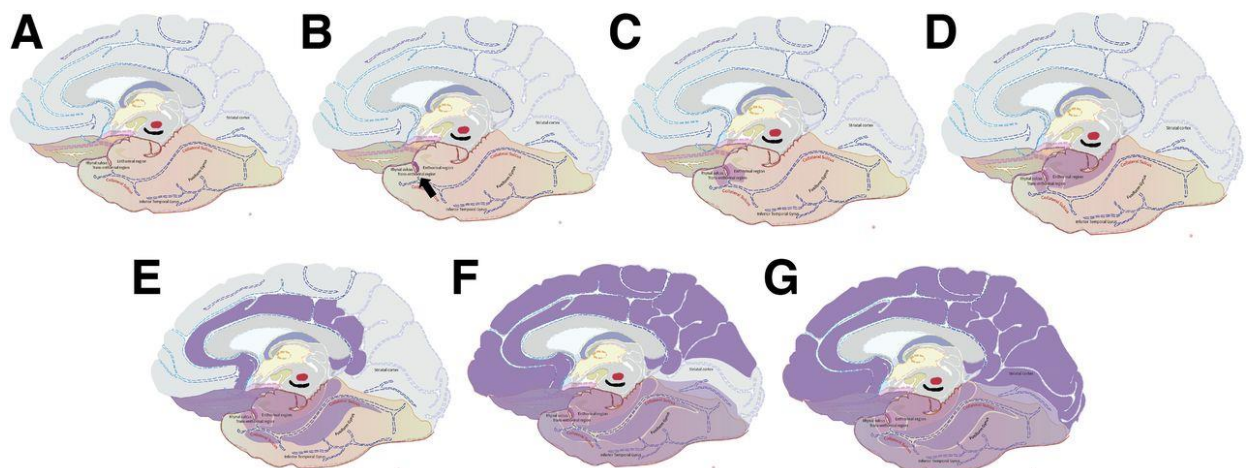
Figure 10 – Proposed neurofibrillary tangles diffusion into brain areas by Braak: starting from entorhinal regions, aggregates are observed in limbic areas such as hippocampus and finally spreading to the cortex with staging correlated to cognitive impairment [https://jnm.snmjournals.org/content/64/8/1171].

Recent data are consistent with progressive divergence of some biomarkers during the course of AD, showing alterations occurring long before the overt dementia diagnosis [26]. A study with Chinese individuals observed a decrease of Aβ42 levels in cerebrospinal fluid (CSF)

diverging up to 18 years before cognitive symptoms as well as a decrease in Aβ42/ Aβ40 ratio compared to cognitively intact matched controls; tau CSF biomarkers (phosphorylated tau 181, total tau and neurofilament light chain concentrations) increased at around 10 years before diagnosis. Magnetic resonance revealed hippocampal volume differences at 8 years and subtle cognitive impairment at 6 years, progressing to dementia (Figure 11) [27].

The observed biomarker trajectories, confirmed by various longitudinal studies, is redefining the approach to AD intervention, emphasizing the need for early diagnosis tools and clinical trials design, before cognitive deficit -maybe irreversible- appears.
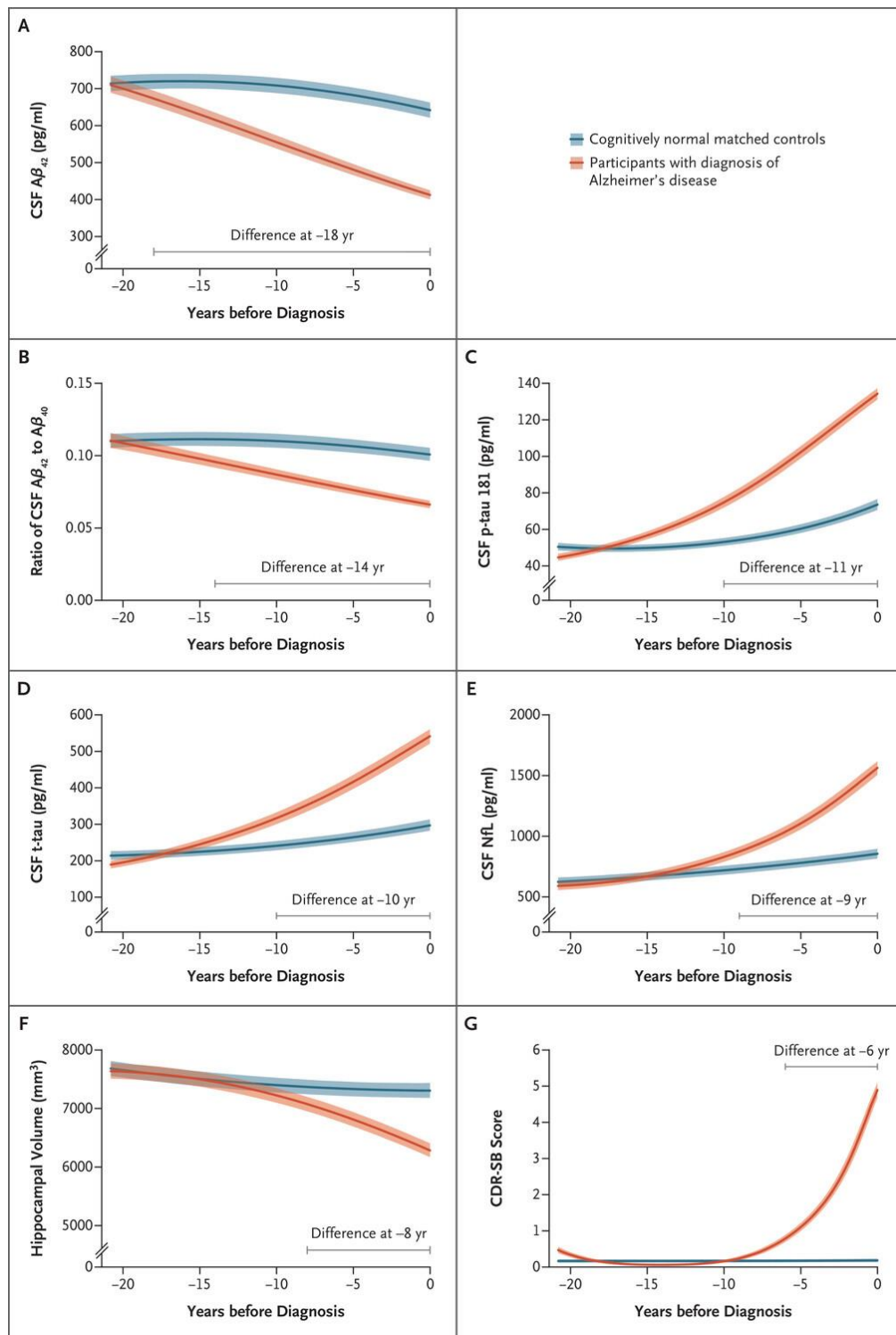
Figure 11 – Graphs showing divergence amongst AD patients' biomarkers and control subjects: decrease in CSF Aβ42 concentration (A) and Aβ42/ Aβ40 ratio (B), increase in phosphorylated CSF tau 181(C), total tau (D) and neurofilament light chain concentration (E); decrease in hippocampal volume (F) and higher CDR score (G), the latter starting from 6 years before AD diagnosis [https://www.nejm.org/doi/full/10.1056/NEJMoa2310168].

2.2 Amyloid precursor protein as a converging point

2.2.1 Amyloid precursor protein physiology

The Amyloid Precursor Protein (APP) is an integral transmembrane protein encoded by the APP gene located on chromosome 21. It is a member of a highly conserved family of proteins that include APLP1 (Amyloid Precursor-Like Protein 1) and APLP2 (Amyloid Precursor-Like Protein 2). Interestingly, the APP, along with its homologs, exhibits high sequence similarity in their extracellular and intracellular domains, while the Aβ sequence is present exclusively in human APP and not in APLP1 or APLP2 [28].

APP, APLP1, and APLP2 are in fact conserved across various species, ranging from mammals to simpler organisms like *Caenorhabditis elegans* (C. elegans), thus indicating their essentiality and conservation following natural selection forces; however, the sequence encoding the Aβ peptide within APP (residues 672-713 in the standard sequence) lacks the same conservation pattern.

APP consists of several functional domains (Figure 12), including extracellular domains such as E1 and E2, a transmembrane region, and a cytoplasmic tail. Although the complete structure of the full-length APP protein has not been fully characterized due to its inherent flexibility, high dimensions and disordered regions, the crystal structures of individual domains were solved [29]. These domains are connected by flexible linkers, allowing for the dynamic interaction of APP with its binding partners. Below a description of the key APP domains is provided:

- **E1 domain**: located at the N-terminus (residues 18-190), this globular cysteine-rich domain consists of a region binding heparin and one binding copper. This domain is putatively involved in APP dimerization and could bind to extracellular matrix proteins such as proteoglycans or acts as a growth factor.
- **Kunitz protease inhibitor domain (KPI)**: this domain is mainly expressed in non-neuronal cells such as platelets and could have a role in wound healing by inhibiting serine proteases.
- **E2 domain**: positioned between the E1 domain and the transmembrane region in the most expressed APP isoform, the E2 domain is a helix rich region; it engages in heparin binding, possibly favoring dimerization and some metal-binding sites suggest a role in iron homeostasis and export.

- **Aβ region**: this segment, unique to APP and relevant to AD, comprises residues 672-713 in the standard sequence and includes residues from both ectodomain and transmembrane region; several enzymes called secretase take part in APP processing and degradation via proteolytic cleavage.
- **APP Intracellular Domain (AICD)**: the intracellular domain is conserved and plays a key role in intracellular localization and endocytosis of APP; it contains the **YENPTY** motif which is also present in other receptor proteins such as tyrosine kinases and integrins.

The biological importance of APP and its homologs is highlighted by knockout (KO) studies, showing that triple KO mice, lacking APP, APLP1 and APLP2, exhibit postnatal lethality due to severe hypoglycemia, as well as mice lacking both APP and APLP2; single KO models tend to have milder, often behavioral deficits, indicating a partially overlapping function of the three members of this family. It is noteworthy that the knock-in of the extracellular domain in double KO could rescue the lethal phenotype, suggesting a role in growth [28].

APP is primarily observed in the Golgi apparatus, endosomes, and at the cell surface, although at a smaller amount: it undergoes a complex intracellular processing and can be recycled back to the cell surface or be subjected to processing within acidic endosomes and lysosomes after interaction with secretases (see 2.2.2 for a description of these enzymes) [29].

The standard isoform (UniProt P05067), also known as APP770 because of its length, is poorly expressed in neurons. The major brain isoform is APP695 (Supplementary material) but residue positions are referred to the standard sequence.

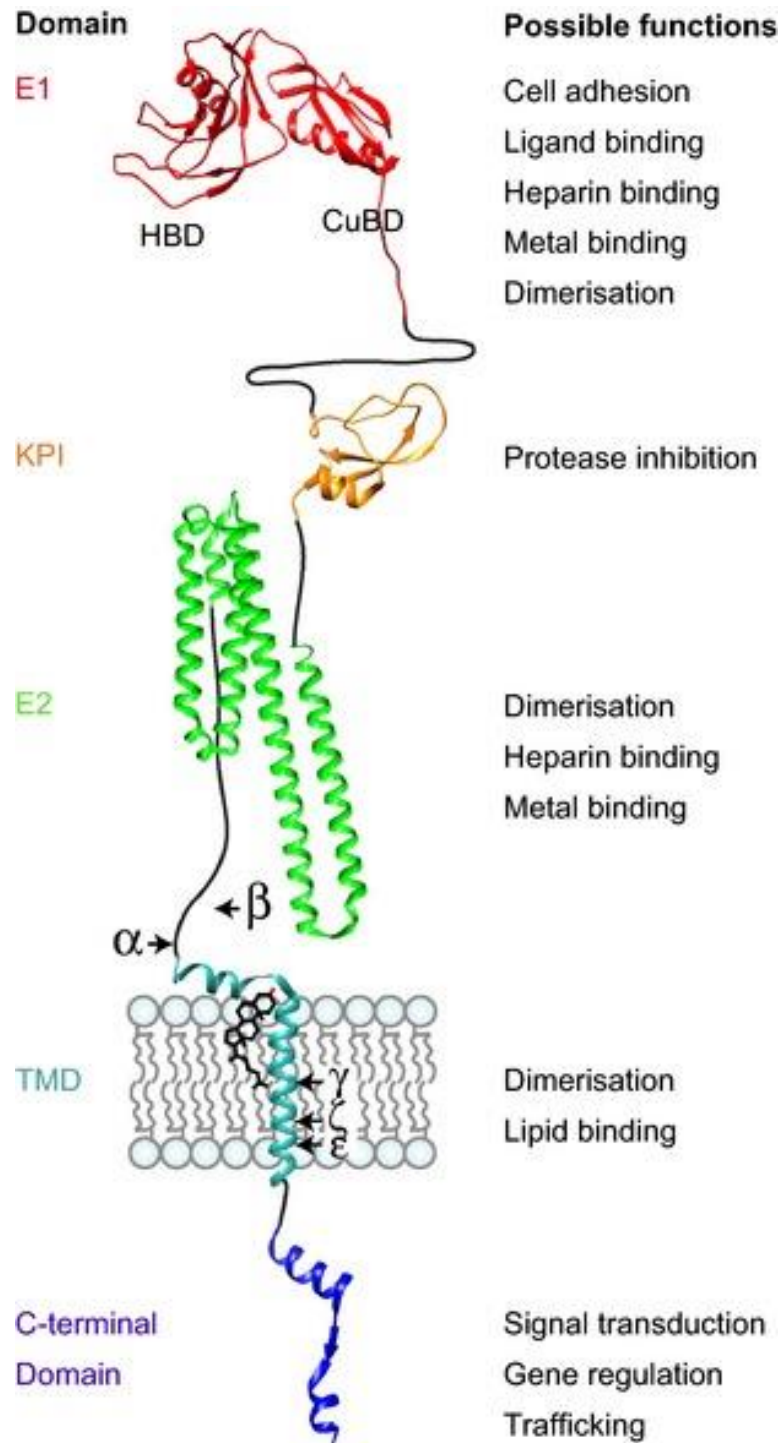*Figure 12 – Proposed structure of APP with putative physiological functions on the right. Main secretase cleavage sites are indicated with black arrows and Greek letters. The three-dimensional structure of domains is taken from solved structure (PDB) while the black segments are linkers with secondary structure poorly defined and considered flexible [https://pmc.ncbi.nlm.nih.gov/articles/PMC4314671/].*

APP can be processed through two major proteolytic pathways (Figure 13), whose biological significance is currently unclear: the non-amyloidogenic and the amyloidogenic pathway [30].

- **Non-Amyloidogenic Pathway**: in this pathway, which is believed to be predominant in young, healthy individual, as well as cognitively intact elderly, APP is first cleaved by α-secretase, producing a soluble domain (APPsα) which is then excreted and a membrane-bound fragment (C83); C83 is processed by γ-secretase, resulting in the release of the P3 peptide and the AICD. There are proofs that APPsα is neurotrophic and neuroprotective, activating cellular repair mechanism, in particular following brain injuries.

- **Amyloidogenic Pathway**: an alternative metabolic route, cleavage by β-secretase (BACE1) generates a shorter soluble fragment (APPsβ) and a membrane-bound fragment (C99). Further processing of C99 by γ-secretase leads release of Aβ peptides which can be of various length, prone to aggregation and involved in AD [29].
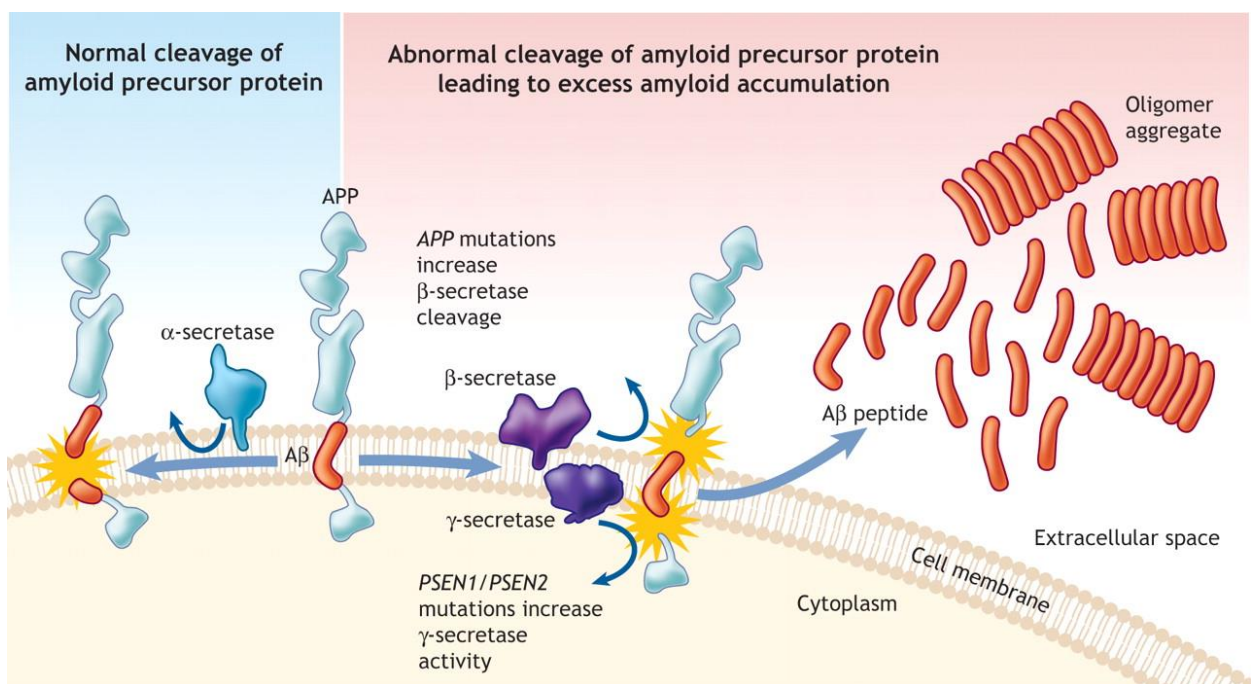

Figure 13 – Graphical summary of normal and aberrant catabolism of APP, in the first case by alpha secretase mediated cleavage and in the second by subsequent beta and gamma secretase, the latter resulting in Aβ production and aggregation per the amyloid hypothesis[https://www.cmaj.ca/content/178/5/548].

2.2.2 The secretases

The secretases are defined as transmembrane proteolytic enzymes possessing the ability to process APP, "secreting" fragments such as Aβ.

Beta-site APP Cleaving Enzyme 1 (BACE1), also named **beta secretase**, is a type I transmembrane aspartyl protease composed of 501 amino acids [31]: it is classified as an anchored aspartyl protease due to its transmembrane domain, which plays a crucial role in its localization and function; the globular active domain is extracellular (Figure 14). The catalytic mechanism of BACE1 is based on a pair of essential conserved aspartic acid residues, Asp32 and Asp228, defined as the "aspartate dyad": thus, it is responsible for the hydrolysis of peptide bonds of its substrates, following activation of a water molecule. Another structural important domain is the antiparallel β-hairpin structure (residues 67-77), also referred to as the "flap", which covers the active site and is necessary for substrate recognition and enzymatic activity; reports described conformational changes switching to an "open" to a "closed" state during catalysis to allow substrate binding and product release [32].

BACE1 mediated cleavage is predominantly localized to acidic subcellular compartments such as the late Golgi apparatus and endosomes, where the catalytic activity is optimal at a pH around five. At the cell surface, BACE1 activity is inferior, as the neutral extracellular pH is not favorable for its enzymatic function [33].
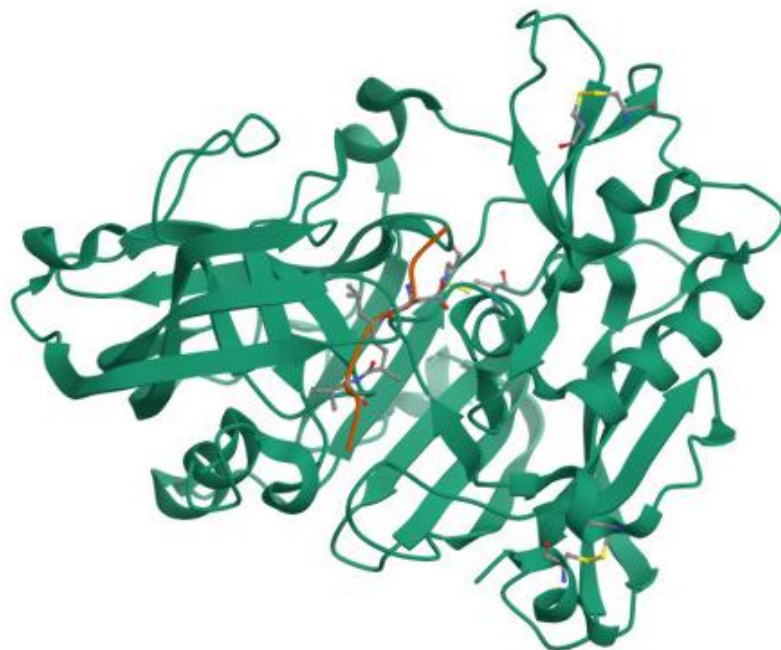

Figure 14 – Crystal structure of BACE1 extracellular domain (green) in complex with the

Beta secretase is widely expressed in the central nervous system, with a particularly high abundance in neurons and in peripheral tissues, especially in liver: beyond its well-known role in amyloidogenic metabolism of APP, BACE1 is also recognized to be involved in a variety of physiological processes through the cleavage of nearly 70 identified substrates, inside and outside nervous system; various functions are ascribed to those substrate, including axonal growth, synaptic transmission, and myelination, this latter influenced by cleavage of Neuregulin-1. Recent evidence underscores BACE1 implications in inflammatory processes and metabolic disease (i.e. obesity) [34].

The pathological role of BACE1, for the sake of this work, primarily lies in a dysbalanced processing of APP, the amyloidogenic pathways, instead of an alpha secretase mediated cleavage: the activation of a water molecule triggers the scission of the peptide bond between Met671 and Asp672 APP residues (beta site); the products, APPsβ and C99, the latter is subsequently cleaved by γ-secretase, are involved in AD cascade per the amyloid hypothesis. While there is no consensus, it is believed that either overexpression or an increase in BACE1 activity are at the core of AD and therefore drug design efforts were aimed at inhibiting beta cleavage [35].

Alpha secretases are actually a subfamily of ADAM protein class, responsible for the "physiological" processing of APP, cleaving it between residue 687-688 in the transmembrane region; they are multicomponent transmembrane proteases and a detailed description of their structure, function and drug targeting can be found as a reference [36].

Gamma-secretase is a multi-subunit transmembrane protein complex responsible for cleavage of a wide range of single pass membrane proteins, such as Notch, Neural cadherin and CD44 [37]; its catalytic activity is conferred by the presenilin (PSEN-1) subunit, an aspartyl protease in complex with nicastrin (NCT), the latter essential for substrate recognition, anterior pharynx defective 1 (APH-1), and presenilin enhancer 2 (PEN-2). After BACE1 mediated cleavage, the C99 fragment is cleaved following activation of a water molecule, resulting in various amyloid peptide lengths, most notably Aβ40 and Aβ42, usually in a 1:10 ratio.

Each subunit possesses distinct structural and regulatory roles that contribute to the overall stability and function of the enzyme (Figure 15); despite the precise assembly and maturation is yet to be understood, the active γ-secretase usually colocalizes with BACE1 in and it is

possible that the two secretases build a beta-gamma complex to sequentially cleave APP in acidic cellular compartments [38].

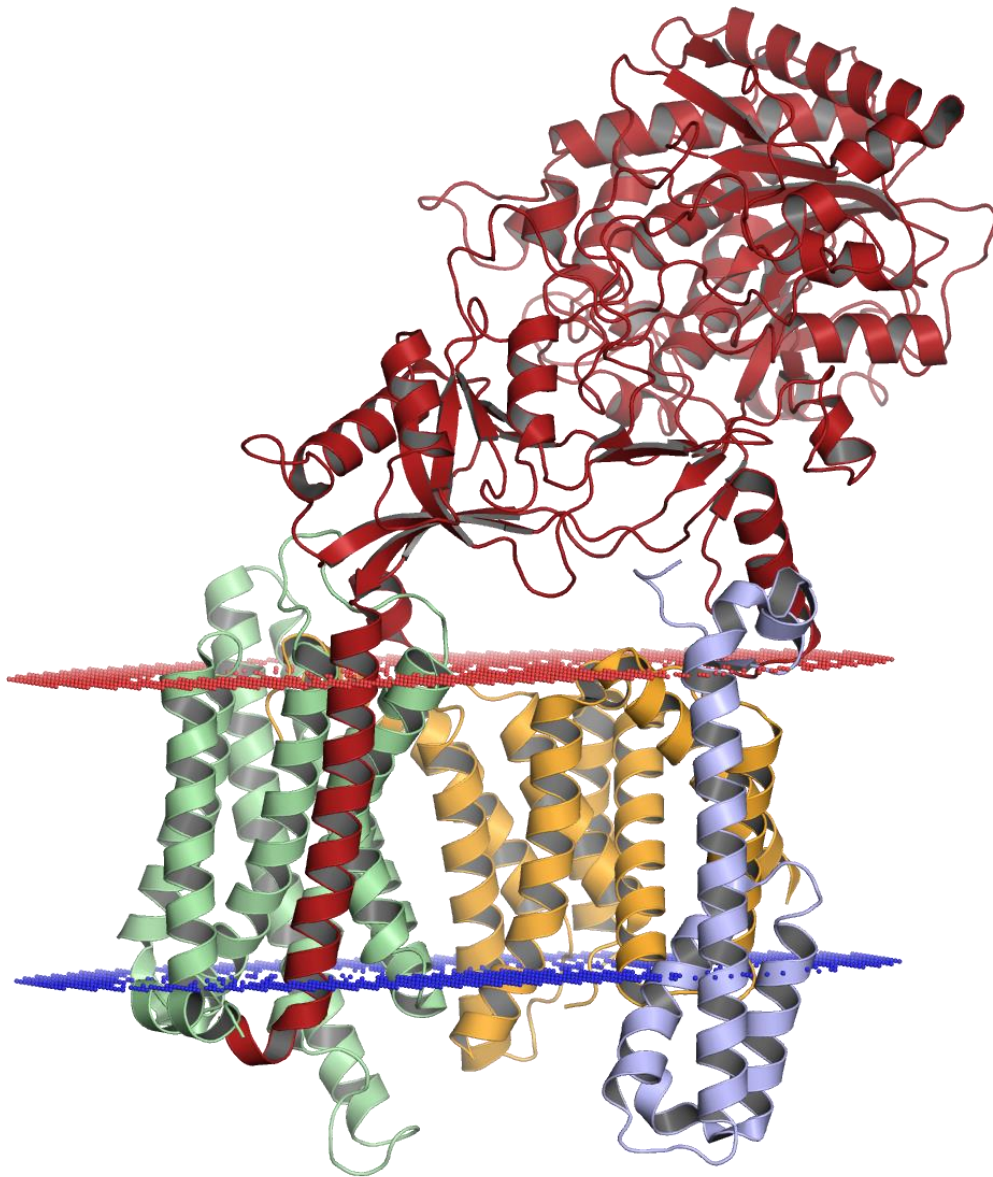*Figure 15 – Cryo-electron microscopy derived structure of gamma secretase multi-subunit complex: in red nicastrin, in green APH-1, in orange presenilin-1 and in light blue PEN-2; the red and blue surfaces represent respectively the lumenal and cytosolic membrane boundaries. APP gamma cleavage occurs in the membrane, presenilin-1 acting as the catalytic subunit [https://pmc.ncbi.nlm.nih.gov/articles/PMC4568306/].*

2.2.3 Mutations in the amyloid precursor protein as a sufficient condition for Alzheimer's disease

As highlighted before, late onset AD is considered sporadic, and in spite of some genetic variants altering the risk profile, most cases are of idiopathic cause [3]. A minority, around 10%, of new AD cases are genetically inherited, mostly believed to be caused by single missense mutations in genes involved in APP processing pathways: APP itself, PSEN-1 or PEN-2 [39]. To this date, in the Alzforum database more than 500 mutations have been collected, with a hundredth in APP gene [40].

Pathogenic mutations of γ-secretase's presenilin subunits share amyloid related changes, for example increase of Aβ42 compared to Aβ40 or faster cleavage kinetics, resulting in overproduction of amyloid peptides.

Considering APP, it is possible to state that some mutations are a causative factor for AD: the discovery of autosomal dominant genetic missense variants are without doubt a leap forward to a better comprehension of AD mechanisms, as well as other conditions such as APP gene duplication.

Down syndrome, i.e. trisomy of chromosome 21, is the commonest aneuploidy: apart from characteristic symptoms, the presence of three copies of APP gene and the related protein overproduction is speculated to be the causative factor for a brain pathology which is indistinguishable from AD by age 40, with more than 60% of the over 70 having an AD type dementia. Even in people without Down syndrome, though possessing more APP copies (copy number variations), AD was present in an autosomal dominant fashion [29], [41].

The majority of pathological mutations in APP occur in the amyloid region and near the cleavage sites or in the immediate proximity of it, from residue 670 to 724 (Figure 16) [29]; nowadays, more than 30 mutations are thought to be pathogenic and observed in families with early onset AD. In Supplementary material, the complete sequence of APP and the mutating sites are provided.
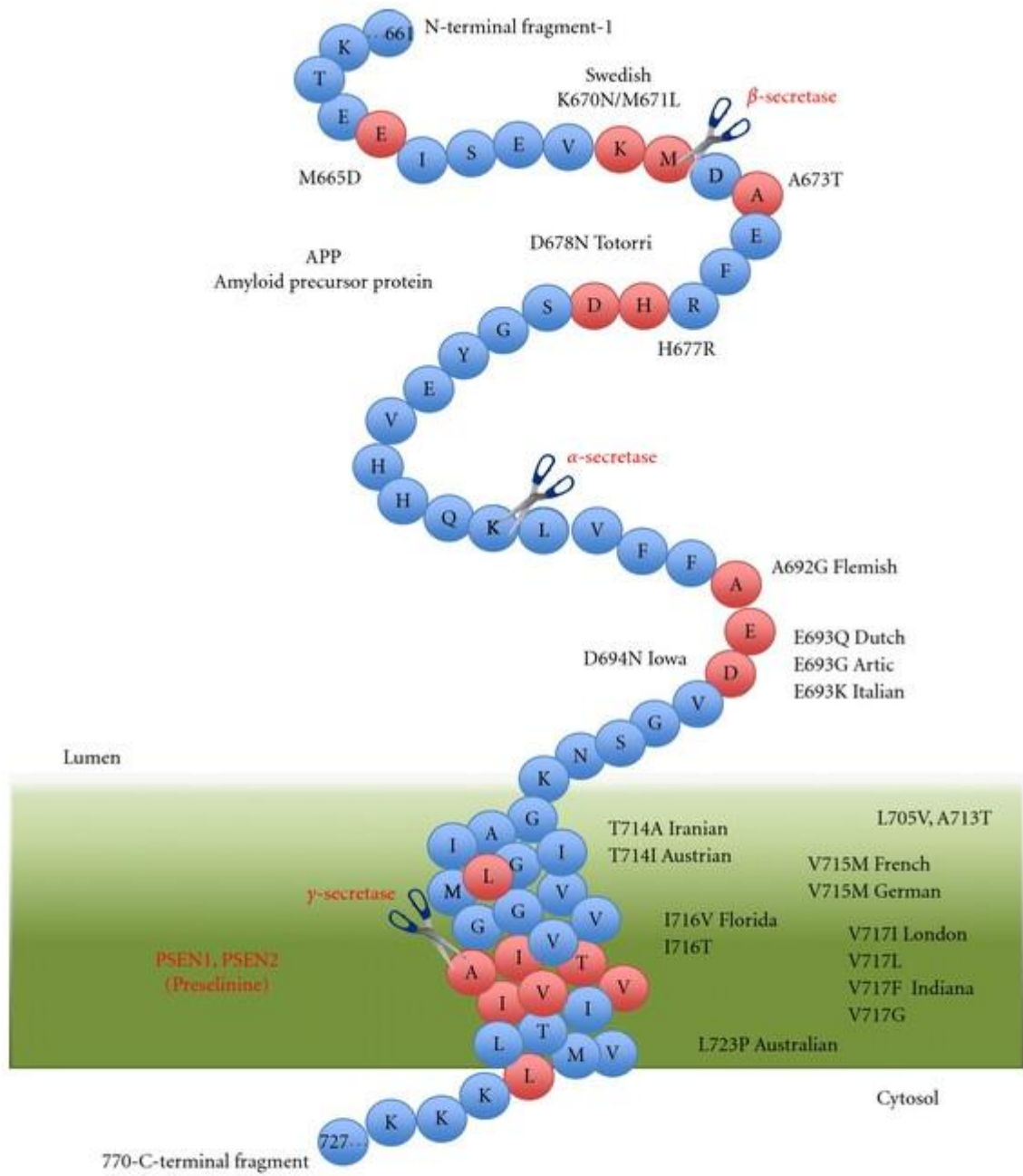
Figure 16 – APP sequence detail from residue 661 to 727 (juxtamembrane plus transmembrane domains) with indication of cleavage sites; known APP mutations are highlighted in red and written following the convention "wild type residue-residue number-mutated residue" [https://onlinelibrary.wiley.com/doi/10.4061/2011/929042].

The main hypothesized mechanisms for AD pathogenesis in patients carrying APP mutations with some examples are considered:

- The alteration of secretase equilibrium; increasing the beta secretase (A673V or Swedish mutation) or in an opposite fashion hampering the physiological alpha pathway (K687N)
- A shift towards the production of more toxic Aβ42 peptide (K724N and M722K)
- Accelerated kinetics of oligomers and fibril formation (E693del, D678H and L688V)

It is therefore evident that even a single aminoacidic variant can contribute to AD pathogenesis, with nearly complete penetrance, in otherwise healthy individuals: the involvement of APP in higher cerebral functions such as memory and cognition is widely accepted; in addition, it is clear that an imbalance in its processing can alter brain homeostasis leading to macroscopic deficits such as neuronal loss and dementia [30].

Near the APP beta secretase cleavage site, located between Met671 and Asp672, we observe two pathogenic mutations.

The Swedish mutation, consisting in a double substitution of lysine and methionine to asparagine and leucine at position 670-671 is shown to increase the amyloidogenic cleavage at beta site (M671-D672) and subsequently increasing the total production of beta amyloid species [42]; the Aβ40/Aβ42 is unchanged but alterations of smaller peptides are reported. It is believed that an increased affinity of mutated APP for BACE1 is responsible for the observed effects in silico [43], in vitro and in vivo.

Other biochemical effects include a disruption of axonal transport machinery, impairing vesicular movement inside neuronal bodies, while metabolic and subcellular location anomalies are still controversial.

The A673V mutation [44], consisting of an alanine to valine substitution, was first observed in two young Italian siblings with progressive dementia; this mutation is believed to be recessive rather than dominant, as several relatives were heterozygous and cognitively intact, while the two cases were homozygous. Analysis showed that this missense mutation alters the beta amyloid processing pathway, increasing cleavage at the beta site, producing more Aβ40; another putative effect, given the mutated residue lies inside the Aβ sequence, is an increase aggregation propensity, a different aggregates structure and cytotoxicity. Interestingly, a mixture of wild type and A2V (position 2 of Aβ) peptides hampers the aggregation rates, and it

is argued that this is the cause of AD pathogenesis only for homozygous carriers [45]. At the same site, a protective mutation was also discovered (Figure 17).
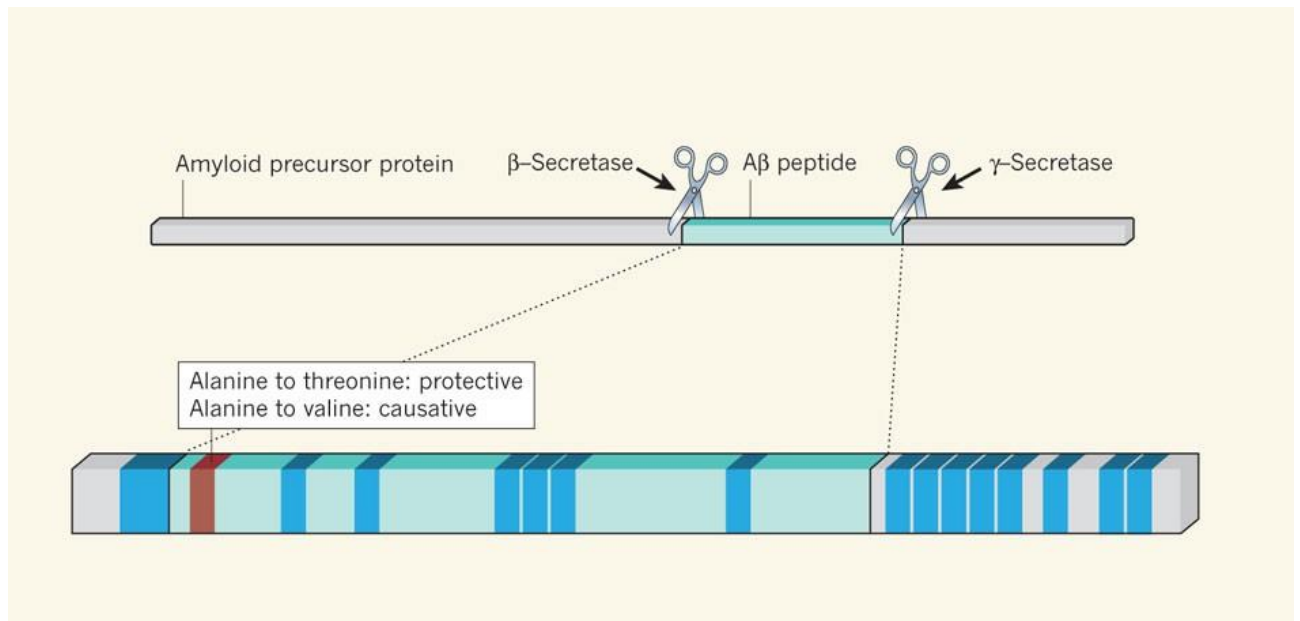

Figure 17 – Graphical representation of APP cleavage site and magnification of the Aβ sequence at the bottom, showing the proximity of residue 613 mutations (alanine to threonine protective while valine pathogenic) to the beta secretase cleavage site [https://www.nature.com/articles/488038a].

2.2.4 The protective A673T "Icelandic" mutation

Described as a protective mutation, the first of this kind APP mutations, it is a rare substitution of an alanine to a threonine at position 673 which was reported to be five time more prevalent in cognitively intact Icelandic than AD patients [46]. In addition, the variant was associated to longevity and higher cognition, suggesting effects beyond amyloid production; from a neuropathological point, patients' biopsies showed no senile plaques or phosphorylated tau [47].

This mutation is very uncommon outside European Nordic countries [48], [49], [50].

Obviously, the discovery of the first protective point mutation for AD prompted extensive research aimed at elucidating its biological effects (Figure 18).

Different mechanistic explanations were made in order to link dementia-free phenotype and biochemical behavior [51]: first of all, A673T seems to favor the non-amyloidogenic cleavage pathway [52], either by increasing cleavage at a different beta site (between residues Y681-E682) or generally decreasing affinity for BACE1 [51]: in fact, a reduction of Aβ and amyloid related products such as sAPPβ were observed in vitro and in vivo; moreover, the amyloid 42 to 40 ratio was reported to decrease. Apart from Aβ production, which is evidently reduced,

studies revealed a less aggregation prone amyloid peptide, dodecamer formation was inhibited as well as >50kDa fibrils build-up (Figures 19 and 20) [53].

Recently acquired plasma and CSF of A673T carriers were obtained and confirmed the reduced production, around 40% [54].
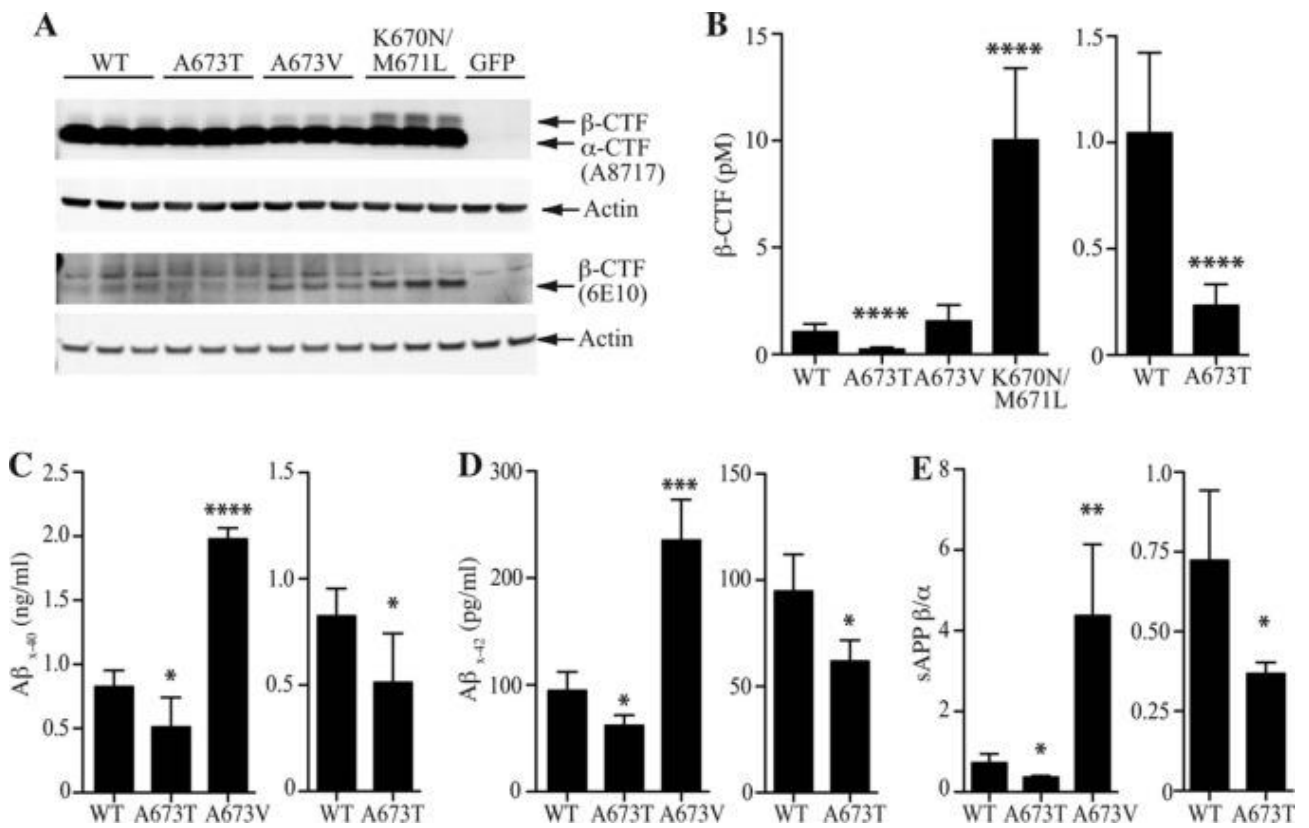

Figure 18 – Effects of pathogenic (A673V and K670N/M671L), protective (A673T) mutations compared to wild type (WT) APP: blotting of alpha and beta CTF from two cell lines expressing APP (A); picomolar concentrations of beta CTF, reduced in A673T and increased in Swedish mutation (B); Aβ40 (C) and Aβ42 (D) concentrations are reduced as well in cells expressing the protective mutation; ratio of beta to alpha CTF is also different for A673T and A673T compared to wild type [https://www.sciencedirect.com/science/article/pii/S0021925820333329].
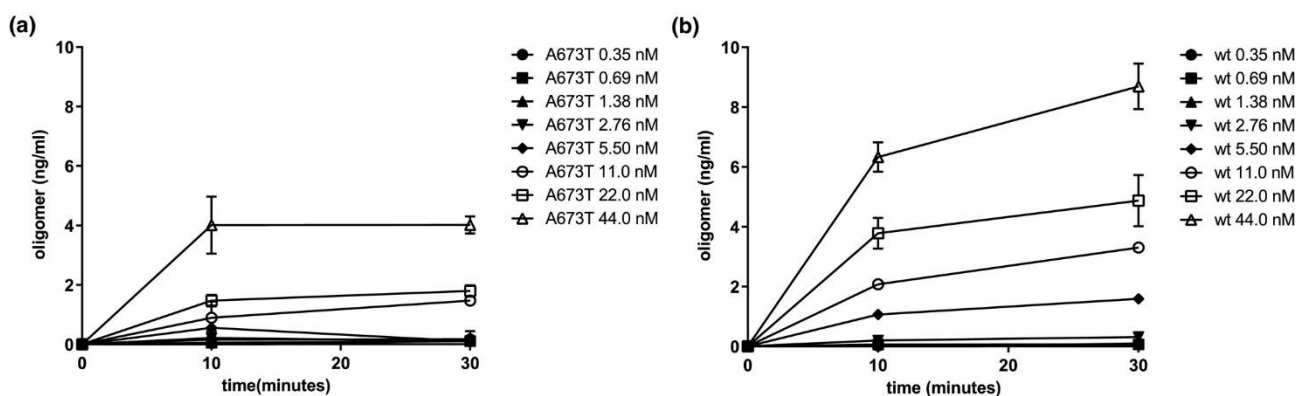

Figure 19 – A673T Aβ42 (a) forms fewer oligomers than wild type (b). Different concentrations are showed in a 30-minute aggregation assay [https://onlinelibrary.wiley.com/doi/10.1111/jnc.15212].
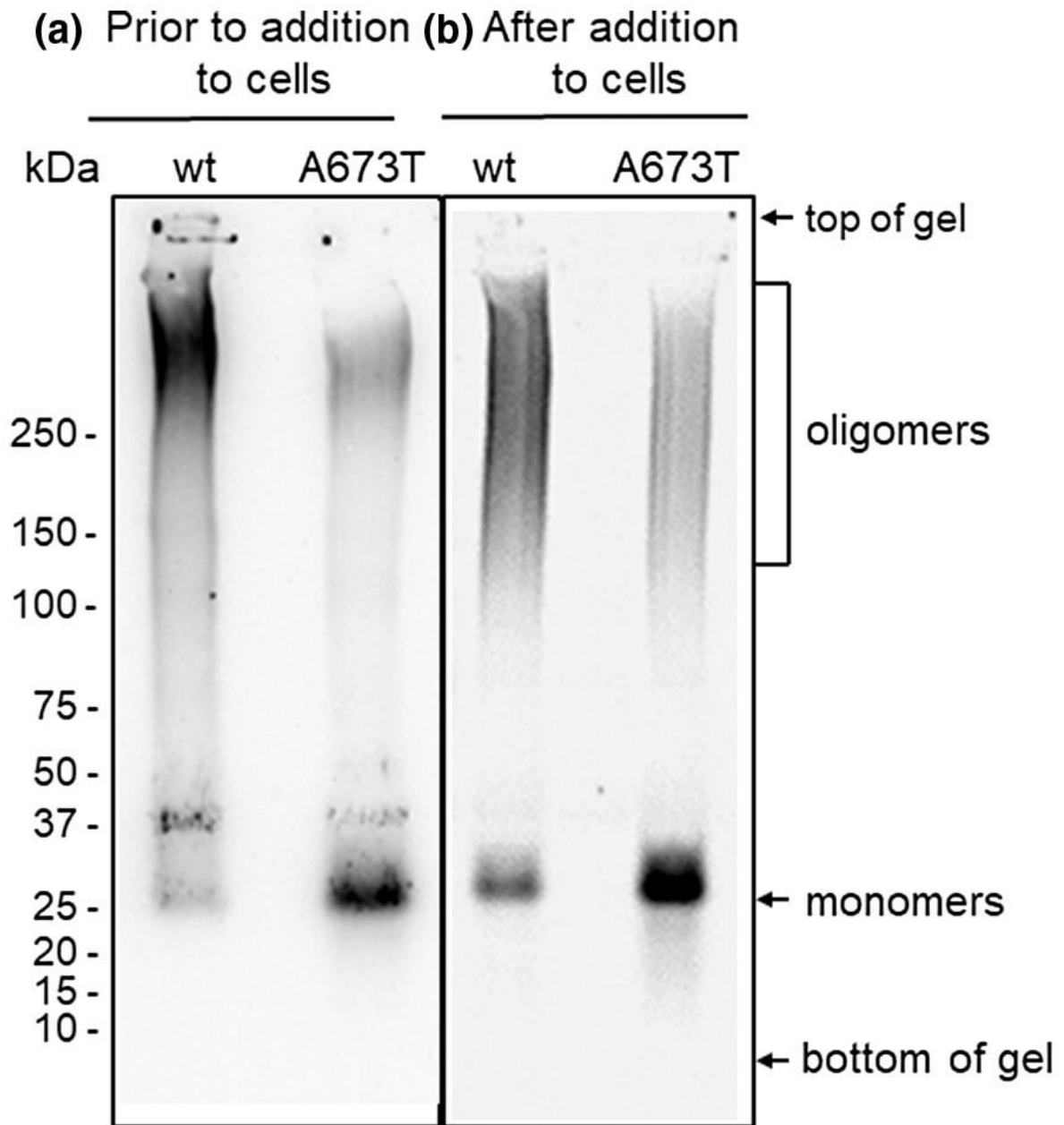
Figure 20 – Western blotting of wild type (wt) and mutant (A673T) Aβ preparations. Oligomers expressing A673T mutations are at lower concentration, mostly in the monomeric form compared to wt [https://onlinelibrary.wiley.com/doi/10.1111/jnc.15212].

In recent years, active scientific interest has led to the development of cell lines (e.g., UEFi003-A, derived from a skin biopsy of a heterozygous carrier [55]) and a mouse model expressing the A673T mutation [56]. Researchers are also exploring gene therapy as a preventative approach for AD, with the use of base substitution strategy being evaluated in mice [57], [58]. However, challenges such as low transfection rates in post-mitotic cells such as neurons and ethical concerns continue to hinder the advancement of those therapies.

2.2.5 Secretase inhibitors design, a failure?

Today's AD approved drugs are palliative and aimed at easing the symptoms interfering with daily life but incapable of slowing down the natural course of dementia. In clinical practice, an acetylcholinesterase inhibitor such as donepezil is administered in the early stages of the disease to counter the destruction of Ach secreting neurons, thus improving mental functions in some individuals [3]. As symptoms severity progresses, memantine, a pore blocker of N-methyl-D-aspartate receptor, is given to decrease excitotoxicity and the following neuronal cells destruction, with variable efficacy.

The recently approved monoclonal antibody Aducanumab is controversial [22]: while the decrease in amyloid aggregation was evident and, from a biochemical point of view, the clinical trials were successful, cognitive improvement was not observed and nowadays it seems to be rarely prescribed [59]; as highlighted before, the failure of amyloid targeting agents such as antibodies could be caused by the wrong attribution of toxic properties to Aβ plaques or to the wrong timing of interventions, with a brain already compromised by diffuse synaptic loss.

However, the APP processing pathway has been regarded as a promising drug target for decades and the clear role of its dysfunction in AD acknowledged (Figure 21). In particular, the inhibition of gamma and beta secretase was attempted with small molecules or synthetic peptides.
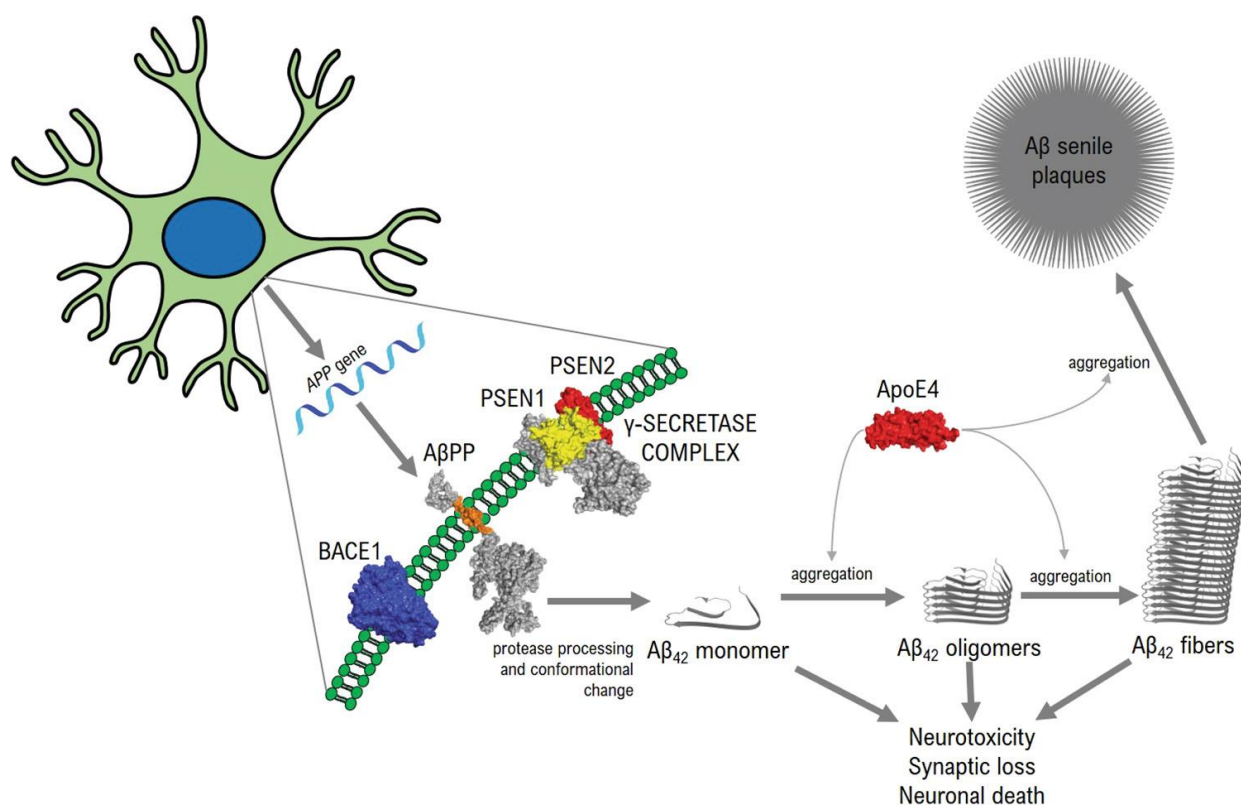
Figure 20 – Schematic representation of the complex interplay of APP: gene translation results in APP expression; cleavage by BACE1 and gamma secretase complex generates different Aβ species, of which Aβ42 is believed to cause neuronal damage whether in its monomeric or aggregated form; finally, apoliproteins binding Aβ, particularly ApoE allele 4, are involved in oligomers clearance [https://www.laboratoryinvestigation.org/article/S0023-6837%2822%2902603-4].

Gamma inhibitors were developed during the 90's when most clinical trials failed, because of serious side effects [60]: NOTCH signaling was impaired and the inhibition of an important brain enzyme, reported to cleave more than a hundred substrates is no longer considered feasible; this drug class is currently being repurposed in oncology for its anti-proliferative and anti-angiogenic properties [61].

The development of BACE1 inhibitors, with considerable hopes because of the amyloid "paradigm" of AD pathogenesis, was not capable of demonstrating any beneficial effects in clinical trials (Figure 22) [31]: moreover, patients in the intervention groups saw an accelerated cognitive decline; while the "loss" in cognitive test was in general regained after stopping drug administration [35]. It is now considered dangerous to inhibit BACE1, probably because its involvement in cleaving many substrates or because non-specific action on BACE2, a closely related secretase expressed in peripheral tissues and implicated in metabolism. Another reason has been hypothesized to lie in the excessive underproduction of amyloid species, counterintuitively leading to side effects [62]; to this regard, carriers of the protective A673T mutations show only a "modest" reduction in Aβ production.

On one hand, it could be thought that targeting those essential secretases is futile and even dangerous to brain homeostasis and therefore moving forward new molecular targets, while others suggested that drug design efforts should be made to reconsider gamma and beta secretase, fine-tuning their usage or develop alpha secretase targeting agents to test possible clinical benefits [63].
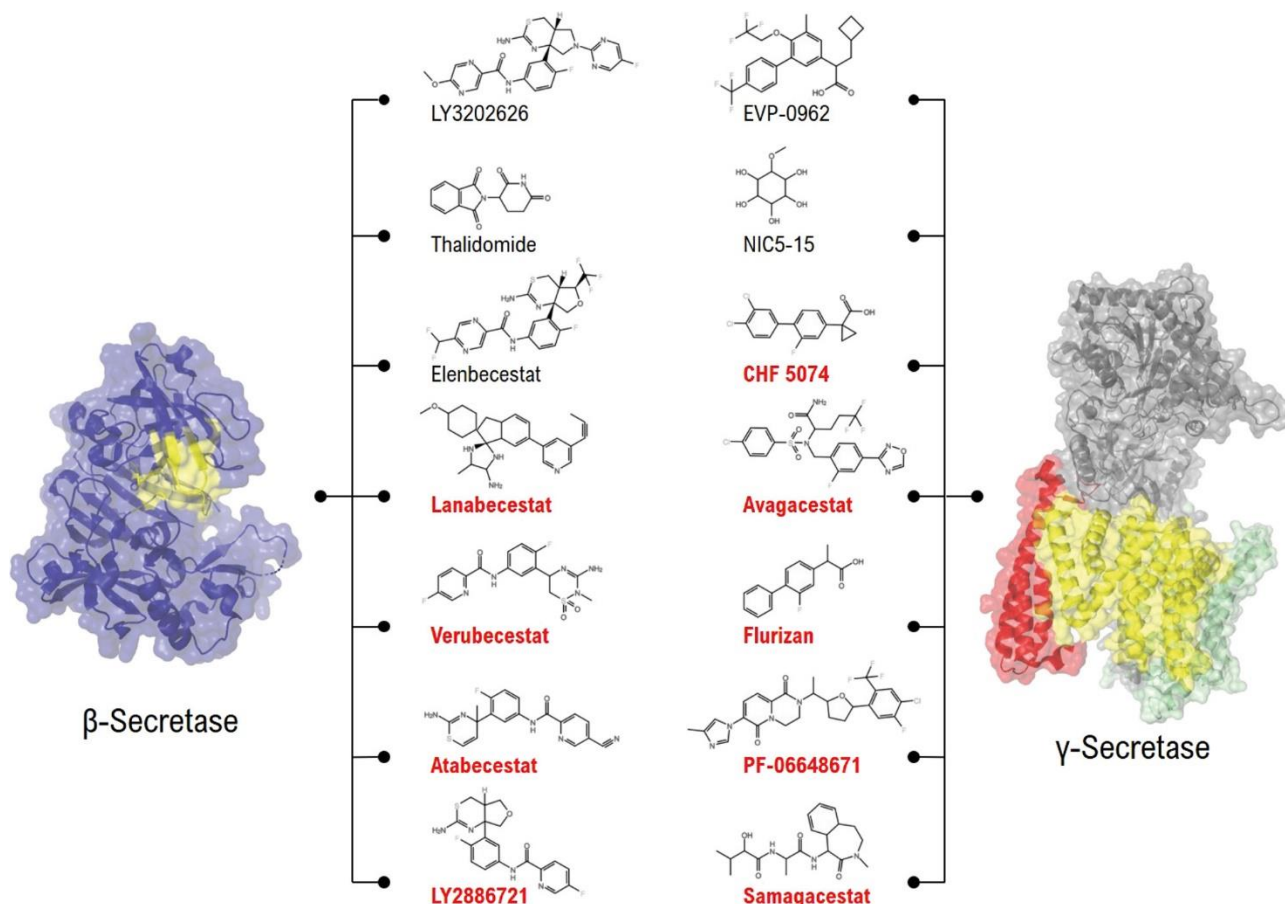

Figure 21 – Three-dimensional representations of beta and gamma secretase with depiction of chemical formulas and names of some drug candidates (secretase inhibitors) reaching Phase III clinical trials for AD treatment. None of these candidates resulted in improved cognition versus placebo [https://www.laboratoryinvestigation.org/article/S0023-6837(22)02603-4/].

Nowadays, despite less enthusiasm and obviously at a smaller scale, the research on secretases is shifting to characterization of gamma modulators [60], with reduced inhibition of Notch signaling; alpha secretase activity enhancers, which have not reached yet the clinic [63]; and BACE1 specific inhibition to reduce off-targets, particularly BACE2 [35]. Finally, re-timing (starting before symptoms onset) and re-dosing (decrease the inhibition potency) of previous tested compounds could be a feasible option, in the light of acquired knowledge on secretases' role beyond AD.

3 Materials and methods

This chapter offers a description of computational methods used in this work: first, a general introduction highlighting the importance of a multidisciplinary approach in understanding biological systems; then, modelling of biomolecules, molecular docking, dynamics and binding free energy calculation methods are described.

3.1 An integrated approach to understanding biological systems

The recent advances in biotechnology, *on par* with increase in computational power, have transformed the understanding of biological systems: it is therefore necessary an innovative approach, following an "explosion" of biological data, for effective analysis and interpretation. A multiscale perspective is crucial for comprehensive characterization of biological phenomena [64]: as an example, a single amino acid substitution (in autosomal Alzheimer's disease or beta thalassemia) can have a huge impact on the whole functioning of the human organism, with effects spanning across length scales spanning ten orders of magnitude and up to sixteen for timescales (Figure 22).

Another important field of study is certainly structural biology [65], defined as "the study of the molecular structure and dynamics of macromolecules": usually, the three-dimensional architecture of biomolecules is intrinsically linked to their function and alteration in the correct structure (misfolding) is a well-documented phenomenon, leading to various pathological conditions such as prion diseases and cystic fibrosis.
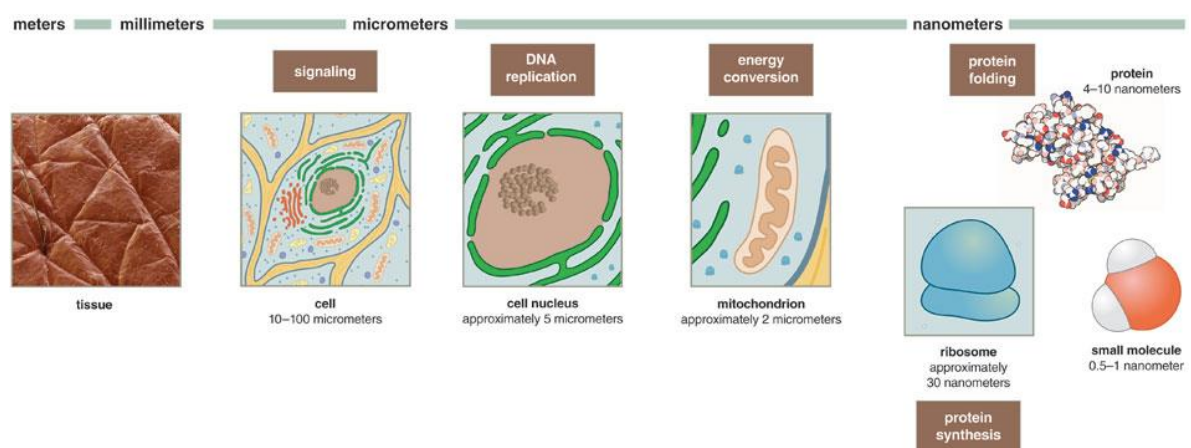

Figure 22 – Characteristic length scales in biology [https://www.americanscientist.org/article/multiscale-modeling-in-biology].

The study of complex biological systems necessitates converging efforts from several scientific domains: biochemistry, biomechanics, biophysics, engineering and computer science; the resulting holistic understanding is fundamental in addressing complex questions from multiple perspectives, shifting from reductionistic claims to a more integrative approach (Figure 23) [64].

A notable achievement in this field is the recent development of AlphaFold [66], with two scientists from its team awarded with the 2024 Nobel Prize in Chemistry. This artificial intelligence (AI) program, based on deep learning, performs the prediction of protein three-dimensional structures, starting from a sequence of amino acids; its latest version, AlphaFold3, is capable of predicting protein interactions with other proteins as well as nucleic acids or ions [67].

The in-silico simulation of macromolecules has become a crucial "tool" for the researchers in the biomedical field: by allowing the visualization of the molecular behavior with atomistic resolution, these simulations significantly reduce the costs associated with wet lab experiments. Additionally, they facilitate the characterization of molecular motion, providing invaluable insights that help experimental design and hypothesis testing.

In this chapter, the theoretical foundations of molecular modeling, along with the principles underlying molecular docking, molecular dynamics and binding free energy calculations, are discussed.
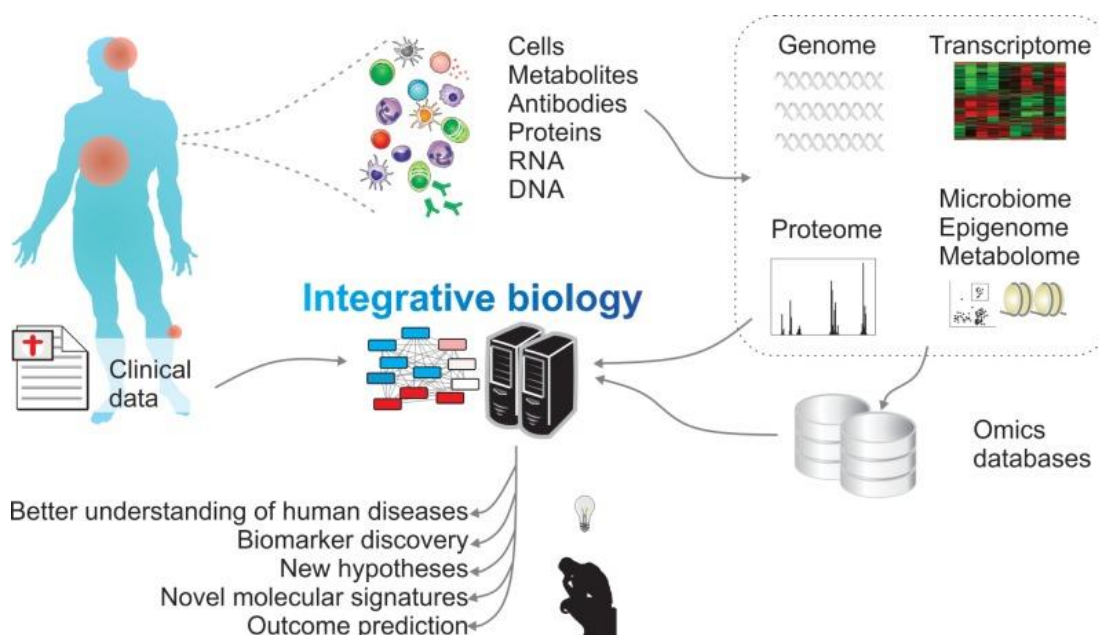
Figure 23 – Summary of integrative biology multidisciplinarity, merging the knowledge from omics databases and clinical data to improve understanding of biological systems.
[https://rupress.org/jcb/article/206/6/695/54581/Systems-cell-biologySystems-cell-biology]

3.2 Modelling interacting molecules

Molecular mechanics can be defined as the application of classical laws of mechanics to describe the behavior of molecules, especially in the biomedical field. While quantum mechanics governs the interactions at atomic length scales, needing a resolution of Schrödinger's equation, a simplified view of molecular systems is often preferred: the Born-Oppenheimer approximation posits that electron clouds adjust instantaneously to changes in nuclear position, being the electron's mass negligible with respect to nuclei [68].

Following the simplified model, atoms are represented as spheres with appropriate radii and assigned specific charges; bonds between these atoms are modelled as springs, reflecting the potential energy associated with their relative positions. The application of Newton's laws of motion is then used to describe the system behavior.

Molecular mechanics is a computational approach that applies Newton's equations of motion to describe the behavior of a system of atoms.

The potential energy V of a molecular system is calculated considering the contribution of bond-dependent and bond-independent terms (Figure 24). The former encompasses bond, angle, and dihedral energies, while the latter includes electrostatic and Van der Waals interactions [68].

This relationship is expressed as:

$$V = V_{bonded} + V_{non-bonded} \qquad (Equation\ 1)$$

The potential energy depends on the relative positions of the atoms within the system and can vary according to the chosen force field. The force field is the set of parameters used in calculations while each interaction term is composed of specific equations.

$$U(R) = \sum_{bonds} k_r (r - r_{eq})^2 \qquad \text{bond}$$
$$+ \sum_{angles} k_\theta (\theta - \theta_{eq})^2 \qquad \text{angle}$$
$$+ \sum_{dihedrals} k_\phi (1 + \cos[n\phi - \gamma]) \qquad \text{dihedral}$$
$$+ \sum_{impropers} k_\omega (\omega - \omega_{eq})^2 \qquad \text{improper}$$
$$+ \sum_{i<j}^{atoms} \varepsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^6 \right] \qquad \text{van der Waals}$$
$$+ \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \text{electrostatic}$$

Figure 24 – Potential energy function terms with graphical representations [https://www.mdpi.com/2073-4344/6/6/82].

The first term, bonded interactions, accounts for the energy changes associated with variations in bond lengths, bond angles, and dihedral angles; covalent bonds are usually modeled using harmonic potential, penalizing deviation from the equilibrium length as observed experimentally.

$$V_{bond} = \sum_{bonds} k_b (r_{ij} - r_0)^2 \qquad (Equation\ 2)$$

Where:

- r_{ij} is the bond length between atoms i and j
- k_b is the bond stiffness constant
- r_0 is the equilibrium bond length, depending on the atom type and the chosen force field

Additionally, energy variations due to changes in bond angles formed by bounded atom triplets are also described using harmonic approximation:

$$V_{\text{angle}} = \sum_{\text{angles}} k_\theta \big(\theta_{ijk} - \theta_0\big)^2 \qquad (Equation\ 3)$$

Where:

- θ_{ijk} is the bond angle formed by atoms i, j and k
- k_θ is the angle stiffness constant
- θ_0 is the equilibrium bond angle

Concerning rotations around bonds involving four atoms, not every configuration is favored, and the potential energy variation is modeled by the dihedral term:

$$V_{\text{dihedral}} = \sum_{\text{dihedrals}} k_\phi \big[1 + \cos\big(n\phi_{ijkl} - \phi_0\big)\big] \qquad (Equation\ 4)$$

Where:

- ϕ_{ijkl} is the dihedral angle formed by atoms i, j, k and l
- k_ϕ is the dihedral stiffness constant
- Φ_0 is the equilibrium dihedral angle
- n is the periodicity of the torsional potential

Non-bonded interactions are forces between not covalently bonded atoms, generally categorized into electrostatic and Van der Waals potentials.

For two particles i and j separated by a distance r with charges q_i and q_j, the electrostatic potential is given by Coulomb's law:

$$V_{\text{coul}} = \sum_{\text{pairs}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \qquad (Equation\ 5)$$

Where:

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

Short-range (vanishing after 1 nm) Van der Waals forces are commonly modeled using the Lennard-Jones potential (Figure 25) with 12 and 6 as indexes respectively for the repulsive and attractive terms:

$$V_{\mathrm{LJ}} = \sum_{\mathrm{pairs}} 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \qquad (Equation\ 6)$$

Where:

- r_{ij} is the distance between atoms i and j,
- σ is the collision diameter (the distance at which the potential is zero),
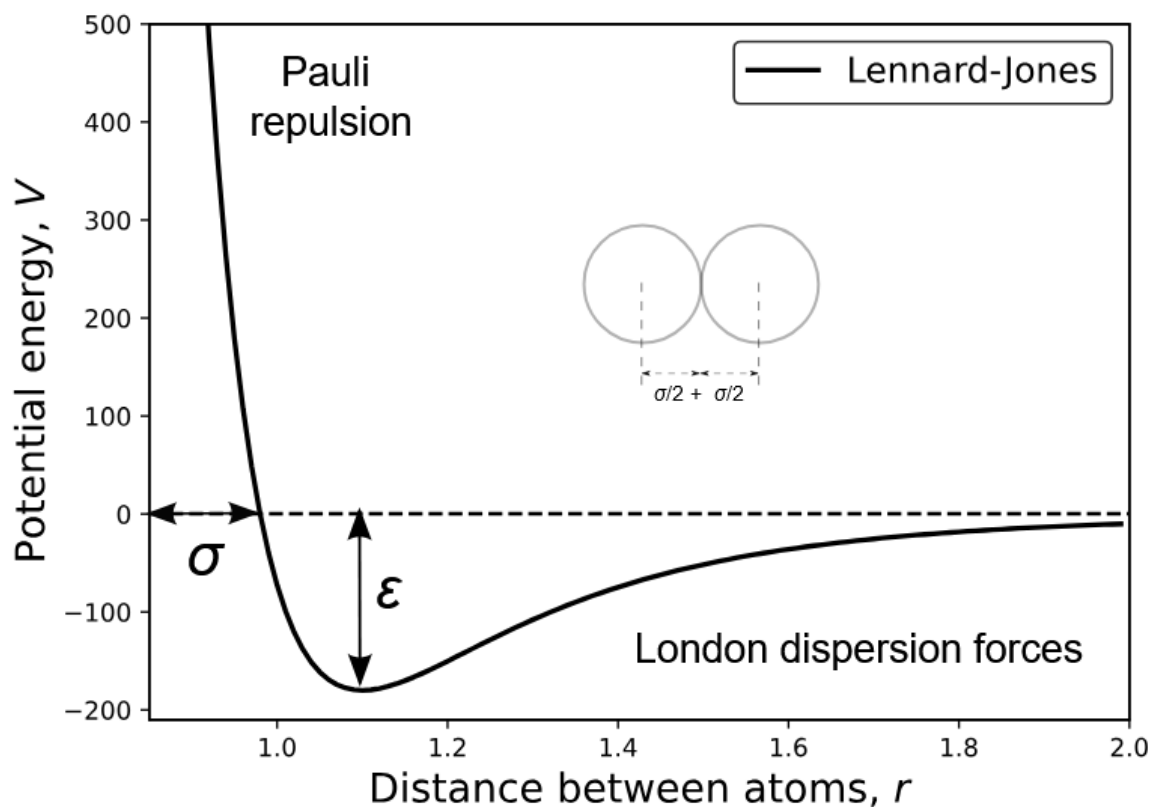- ϵ is the well depth (the minimum potential energy).

Figure 25 – Diagram of potential energy between two atoms as a function of distance (in nanometers). [https://computecanada.github.io/molmodsim-md-theory-lesson-novice/aio/index.html]

Non-bonded interactions are computationally expensive, due to the number of these interactions scaling with N^2, where N represents the number of atoms in the system; several strategies could be employed to mitigate this burden such as cut-off distances, scaling functions or Particle Mesh

Ewald (PME) [69], the latter considered to be an efficient and precise solution for non-bonded interactions treatment. From a computer architecture perspective, the rise of potent Graphics Processing Units (GPU) accelerates non-bonded interactions calculations because of their high parallelizability.

The comprehensive final potential energy function that incorporates both bonded and non-bonded terms is expressed:

$$V = V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}} + V_{\text{coul}} + V_{\text{LJ}} \qquad (Equation\ 7)$$

This equation represents the total potential energy and is the basis to molecular mechanics calculations; to correctly simulate the system, every force field defines the parameters such as stiffness constants, equilibrium radii etc.

In molecular simulations, all atoms are situated within a virtual box, which can possess various geometrical shapes and is filled with solvents, typically water. To minimize edge effects and subsequently artifacts, Periodic Boundary Conditions (PBCs) are applied, simulating an infinite environment surrounding the box with identical replicas (Figure 26); the inaccuracies introduced with this method are usually negligible with respect to the artifacts arising from employing artificial boundaries with vacuum [68].
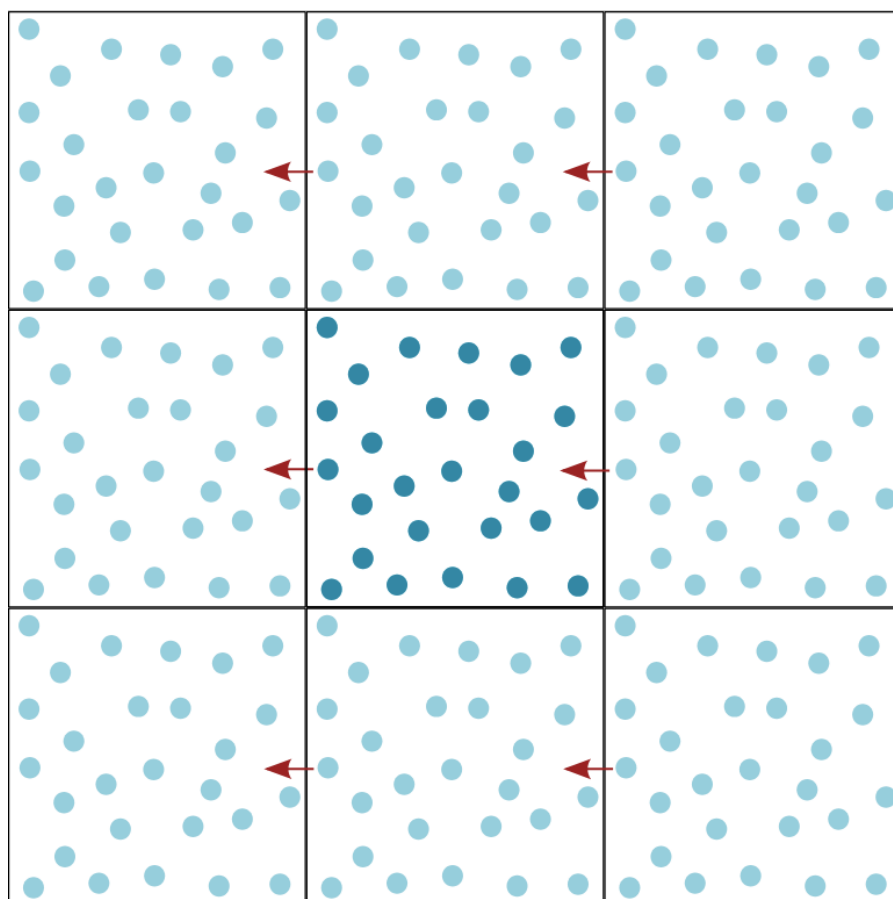
Figure 26 – Representation of Periodic Boundary Condition (PBC): the central box is replicated in copies of itself, and the particle exiting is replaced with a replica from the other side [https://computecanada.github.io/molmodsim-md-theory-lesson-novice/aio/index.html].

The potential energy function can be thought of as a multidimensional surface, defined as the Potential Energy Surface (PES); changes in the coordinates in the systems lead to potential energy variations, which can be visualized as movements on the PES. This surface presents various stable states, and various local minima may be identified. The further the structure is from a minimum, the less stable the simulation will be: therefore, energy minimization (EM) is a crucial process aimed at reducing the potential energy of the system, ending in a stable state (Figure 27). Simulations starting from non-minimized structures often experience failures because of the high forces and energies, resulting in unphysical clashes [70].

The approaches to energy minimizations are derivative methods, classified as first or second order: first order methods (such as the steepest descent) use the gradient of the potential energy function to find a minimum while the second order methods such as LBFGS exploit the second derivative, using an estimate of the inverse Hessian matrix. Second order methods are usually more computational demanding methods, providing more accurate results [68].
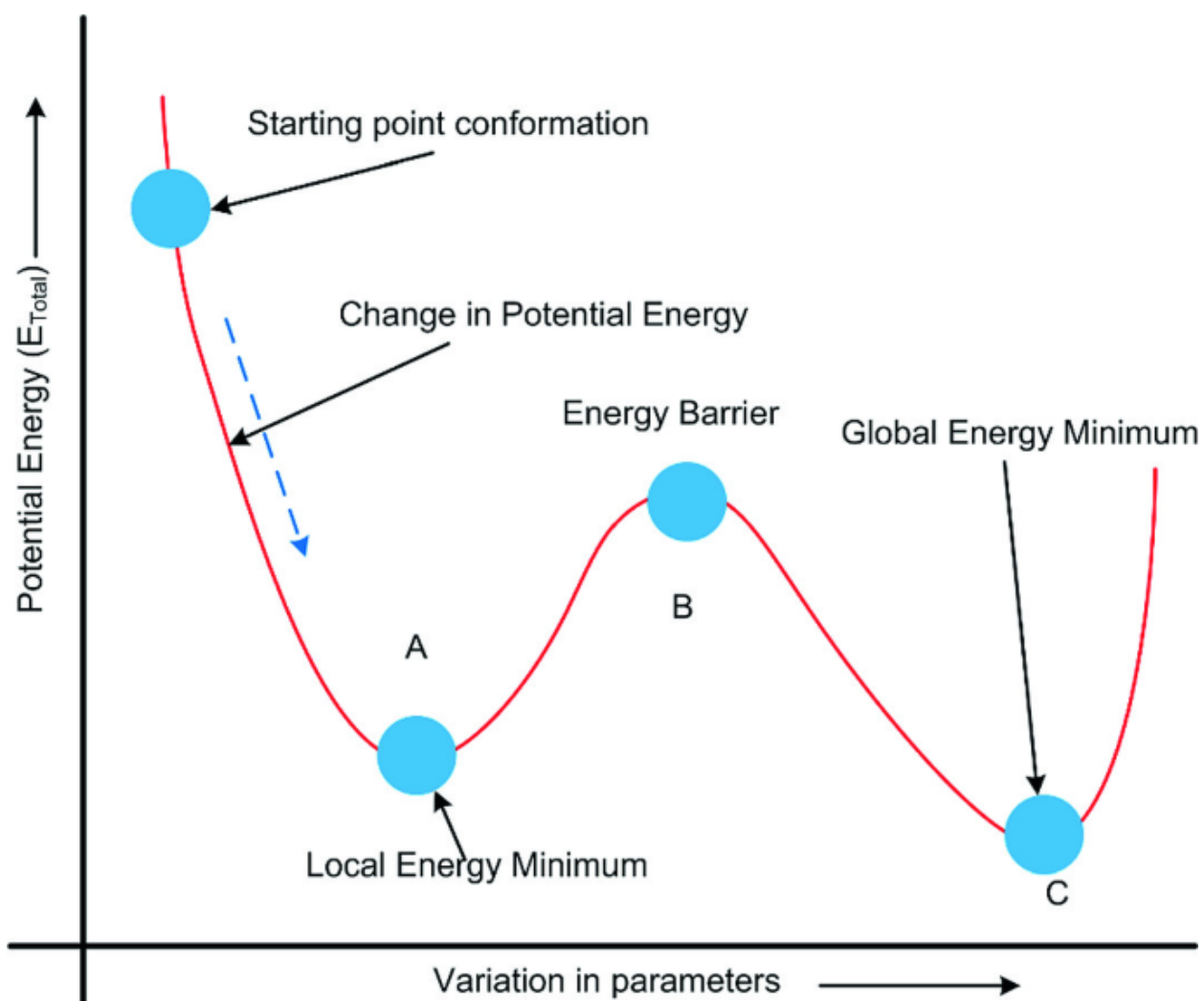
45

Figure 27 – Depiction of energy minimization process: from the starting point at high potential energy the system reaches a local energy minimum A. The energy barrier in B keep the system far from the global energy minimum in C [https://link.springer.com/chapter/10.1007/978-981-19-3092-8_1].

3.3 Molecular docking and dynamics

3.3.1 Molecular docking

Molecular docking is a computational strategy to assess the interactions between two or more molecules; in particular, one of the binding partners is usually a protein, while the other could be another protein, a peptide, or a small molecule [71]. The aim of molecular docking is to provide a reasonable interaction profile, resembling in-vivo binding. Applications of this methodology are numerous: the efficacy of a drug is tightly connected to its affinity for the target, and in-silico screening of promising compounds is a central part of the drug design process. Moreover, protein-protein interactions (PPI) are involved in several pathophysiological processes, and there is a growing interest in targeting these interface regions with small ligands. Finally, peptides have increasingly been included in clinical trials to treat many diseases; the complexity of their experimental characterization is another key advantage of computational methods to elucidate binding mechanisms (Figure 28).
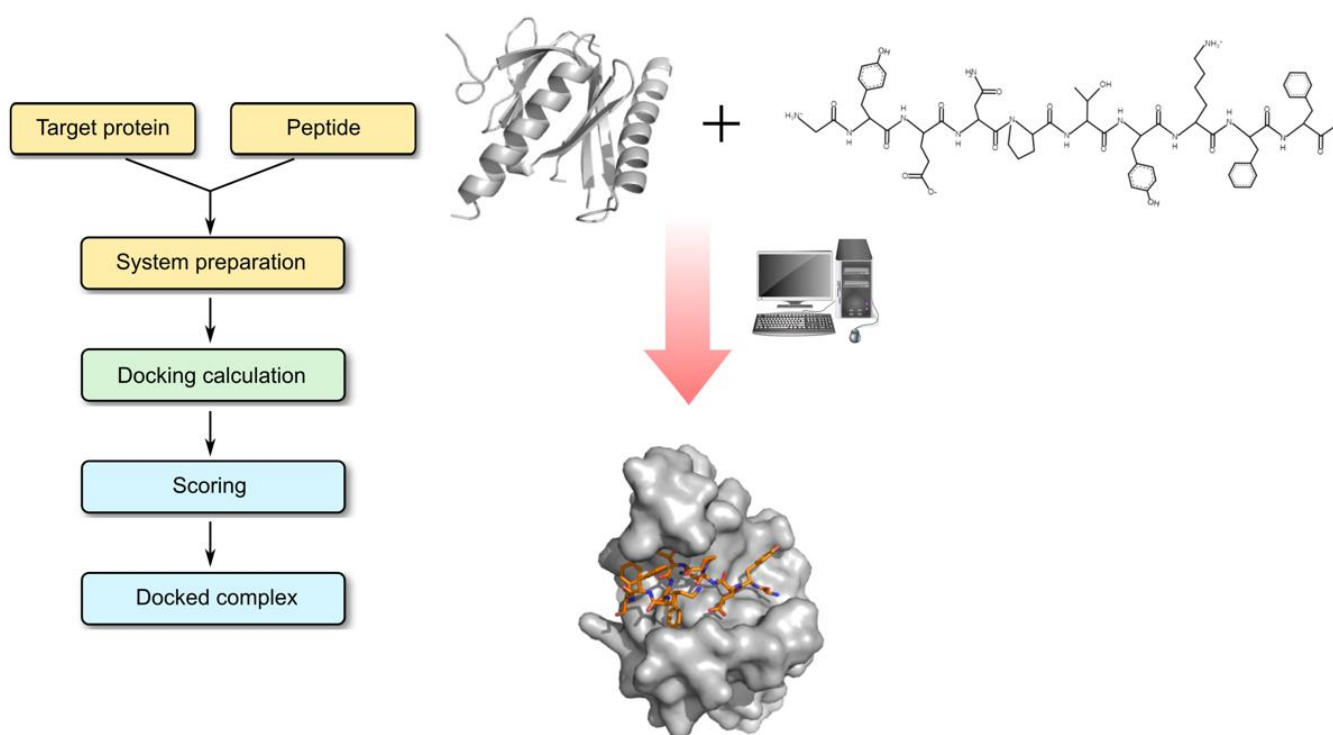

Figure 28 – On the left, the flowchart of protein-peptide docking and, on the right, a graphical representation starting from a protein three-dimensional structure and peptide sequence [https://www.profacgen.com/protein-peptide-docking.htm].

The algorithms employed for docking aim to explore the conformational space of the ligand, usually giving as a result a list of binding poses sorted using a docking score.

In every docking program we distinguish between a searching function, aimed at finding the best binding pose, and a scoring function, to assess interaction strength (Figure 29). The algorithm should yield results closely aligned with experimental results, using wisely computational resources [71].
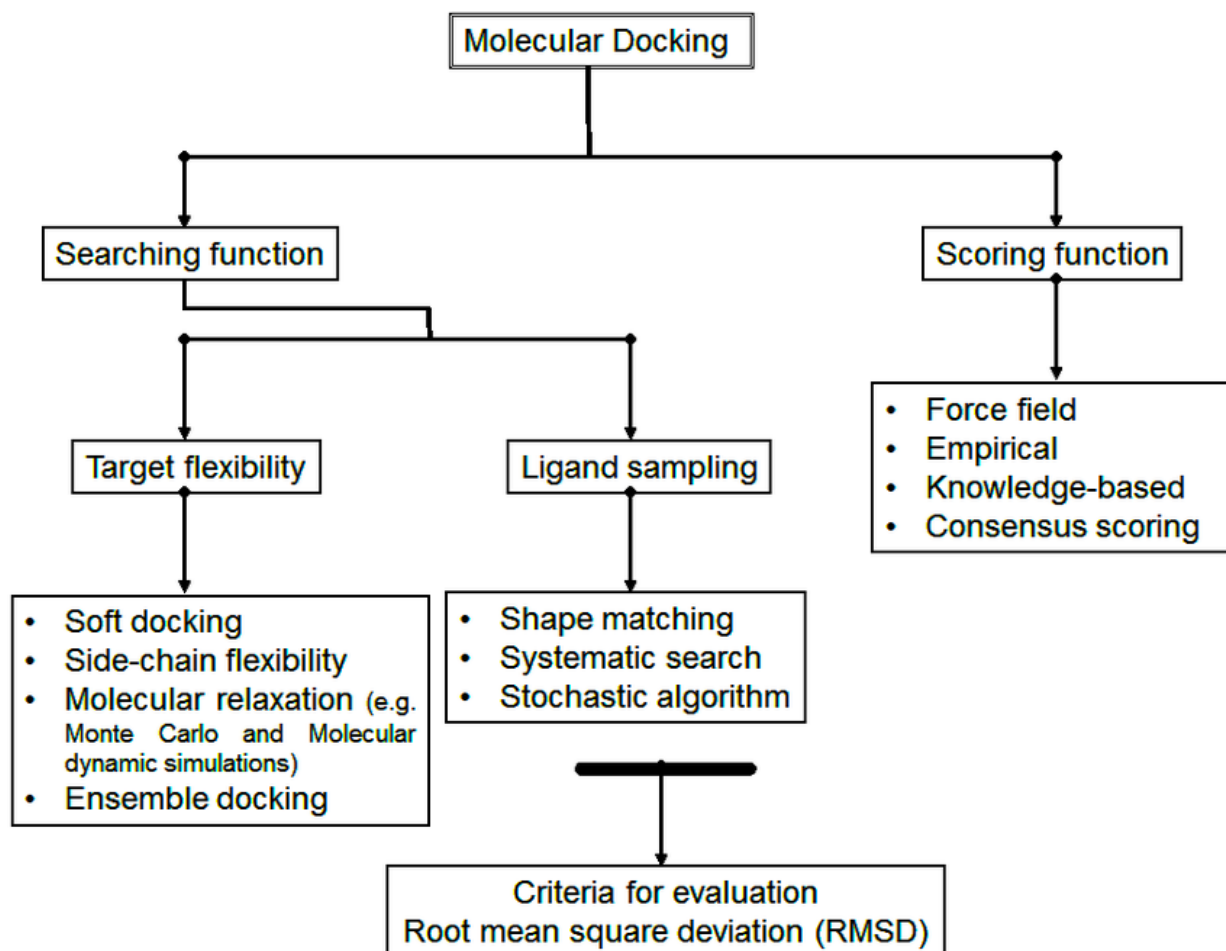

Figure 29 – Molecular docking scheme, listing different searching and scoring functions [https://www.intechopen.com/chapters/44790].

The first approach is a systematic search that explores the ligand's conformational space by applying a set of changes to its roto-translational degrees of freedom; this group can be divided into exhaustive or fragmentation methods. The first is a complete sampling that iteratively modifies bond rotations, though it suffers from efficiency issues due to the rapid explosion of possible conformations. The latter starts with decomposing the ligand into fragments, followed by proper placement and linking to explore sub-conformations.

Another approach is the matching algorithm, which compares physically active portions—such as hydrogen bond donors and acceptors—of the ligand and the target. In this strategy, the three-dimensional arrangement of these active groups is also taken into account.

The third group addresses the enormous number of different binding poses by performing a stochastic search, introducing random variations of bond parameters. This approach is thus suitable for experimental settings involving a high number of ligands (virtual screening) or for cases of increased complexity, such as peptide-protein docking.

Usually, starting from experimental structures obtained in static conditions (e.g., X-ray diffraction of crystals), the interactions between ligand and receptor can be calculated while considering them as rigid. However, it is well known that conformational changes occur upon binding (induced fit); therefore, neglecting intrinsic flexibility results in a severe approximation, often leading to incorrect results.

Different methods are employed to perform flexible docking, namely, semi-flexible or fully flexible docking strategies (Figure 30). In semi-flexible docking, the ligand is flexible while the receptor is considered rigid, improving correlation with experimental data. Flexible docking, on the other hand, allows for both intra-ligand and intra-receptor movements [71].
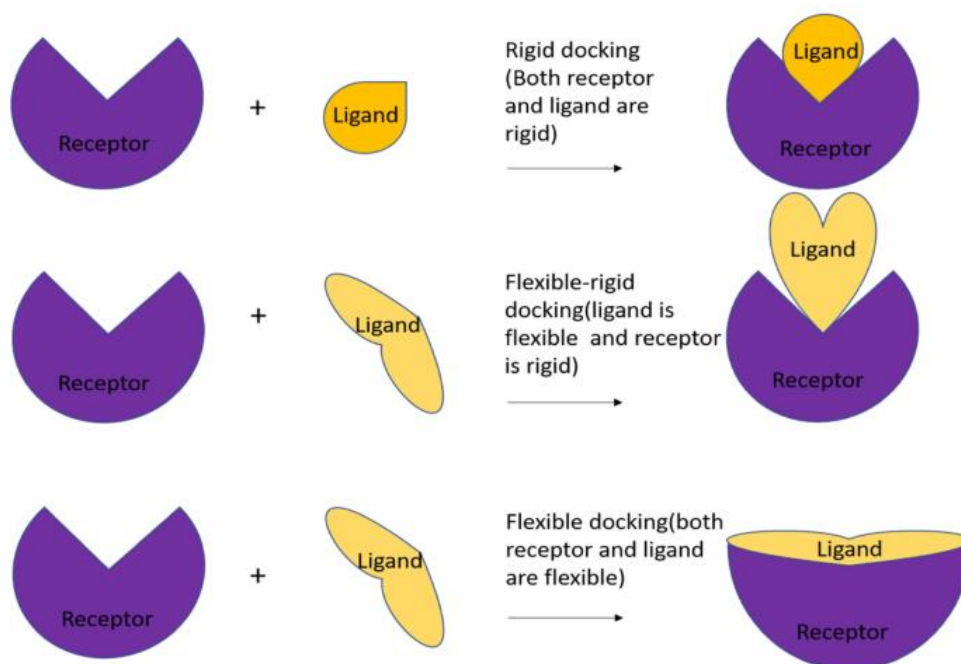

Figure 30 – Different types of molecular docking, considering the system rigid or with ligand or receptor flexibility [https://link.springer.com/article/10.1007/s00706-023-03076-1].

To properly sort docking results, scoring functions are used, with the most common strategies being force field scoring, knowledge-based functions, and empirical scoring functions. Docking software packages implement various search methods and scoring functions, and the best choice of program depends on the system under investigation.

AutoDock is a widely used docking platform that combines force field scoring with a stochastic search method, specifically a genetic algorithm [72]. However, **AutoDock Vina** is now often preferred due to its refined accuracy and approximately twofold increase in speed [73].

Concerning peptide-protein docking, **CABS-dock** is a popular choice: it offers a simple web-based interface and considers the peptide fully flexible while allowing for limited receptor's backbone arrangement. Another advantage is that the algorithm does not require previous knowledge of the binding site, making possible a "blind" docking approach [74].

In protein-protein docking, AI-based tools have made significant advancements: **AlphaFold3** leverages deep learning technology to specifically model interactions between multiple protein chains. It rose to popularity after winning several times the Critical Assessment of protein Structure Prediction challenge (CASP), providing the folding of protein without experimental solved homologs [66]. Apart from this astounding success in elucidating the correct three-dimensional structure starting from the amino acid sequence, it is particularly useful for predicting multimeric protein structures, recently permitting also inclusion of nucleic acids, lipids and ions [67].

Another tool for protein-protein docking is **ClusPro2**. This platform uses a combination of rigid body docking and clustering algorithms to predict the most likely binding poses for protein pairs. ClusPro2 performs calculations based on shape complementarity, electrostatic interactions, and other parameters, followed by clustering of the top poses to identify likely binding modes; in addition, submitters can indicate regions of attractions or repulsions and removing terminal unstructured regions [75].

3.3.2 Molecular dynamics

Molecular dynamics (MD) is a computational technique used to describe the dynamic evolution of complex chemical and biological systems. Using MD, it is possible to calculate the trajectories of interacting particles during the simulation, allowing for the derivation of average properties by numerically solving Newton's equation of motion [68].

Following the recent increase in computational power and parallelization algorithms, it is now possible to model and simulate systems like viral capsids and organelles (Figure 31).

A **statistical ensemble** defines all accessible physical states of a molecular system at a specific temperature. It is a mathematical abstraction, a collection of system configurations, or microstates, that share the same macroscopic (thermodynamic) state. To correctly describe the system's state in phase space, 6N values are needed, three position coordinates and three momentum components for each of N atoms [68].

The definition of the ensemble is necessary to determine the system's macroscopic properties, and various ensembles are used in MD simulations:

- **Micro-Canonical Ensemble (NVE)** simulates an isolated system with fixed energy, volume, and number of particles.
- **Canonical Ensemble (NVT)** is a closed system with fixed temperature, volume, and number of particles.
- **Grand Canonical Ensemble (μVT)** considers the system as open with fixed temperature, volume, and chemical potential.
- **Isobaric Ensemble (NPT)** can be thought of as a system with fixed pressure, temperature, and number of particles.
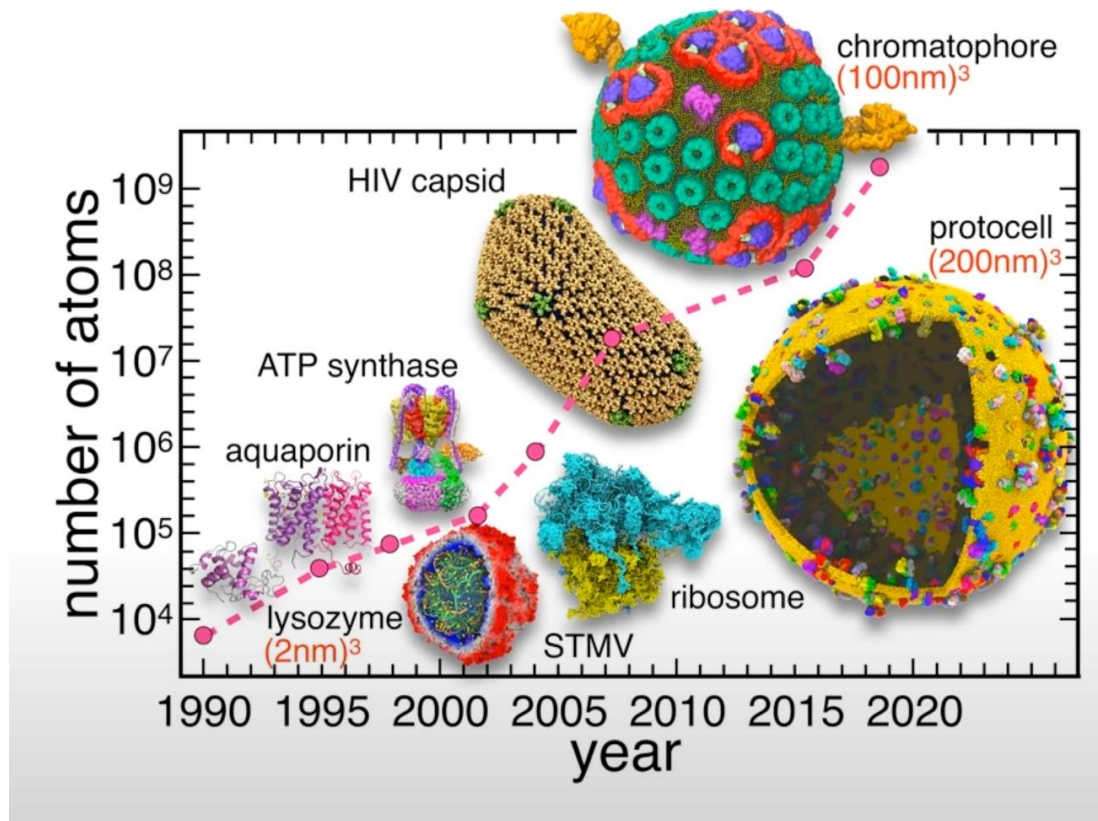
Figure 31 - [https://computecanada.github.io/molmodsim-md-theory-lesson-novice/aio/index.html]

Individual microstates are useless *per se,* and average macroscopic properties are therefore needed. The ensemble average of a property A is calculated by integrating over all possible configurations:

$$\langle A \rangle = \int A(r,p) \; \rho(r,p) \; dr \; dp \qquad (Equation\ 8)$$

Where:

- $\langle A \rangle$ is the ensemble average of property A,
- r and p are the positions and momenta of the particles,
- $\rho(r,p)$ is the probability density function of the ensemble.

The probability density function is given by:

$$\rho(r,p) = \frac{e^{-\frac{H(r,p)}{k_B T}}}{Q} \qquad (Equation\ 9)$$

Where:

- H(r,p) is the Hamiltonian of the system,
- k_B is Boltzmann's constant,
- T is the temperature,
- Q is the **partition function**, expressed as:

$$Q = \int e^{-\frac{H(r,p)}{k_B T}} \, dr \, dp \qquad (Equation\ 10)$$

The partition function sums the Boltzmann factors over all microstates and links microscopic thermodynamic variables to macroscopic properties. However, this equation is not solvable analytically due to the integration over all states. To resolve this, the **ergodic hypothesis** is applied: over a long period, the ensemble average equals the time average. From an MD point of view, the system would eventually reach all the accessible microstates and provided a long enough simulation, the time spent in each state would be proportional to the accessible states [68]. The time-average of property A is:

$$\langle A \rangle = \frac{1}{M} \sum_{m=1}^{M} A(t_m) \qquad (Equation\ 11)$$

Where:

- M is the number of steps in the simulation,
- $A(t_m)$ is the instantaneous value of the property at time t_m.

As said before, given sufficient sampling the ergodic hypothesis allows for the calculation of ensemble-averaged properties from the time average; it has to be noted that the necessary and sufficient conditions for ergodicity to be a reasonable assumption are a field of active research. Moreover, the "classical" sampling (MD) could not reach all the states because of high energetic minima to overcome during the course of a typical nanosecond length simulation, resulting in the development of "enhanced" techniques to augment the sampling of the potential surface area [76].

The setting up of an MD simulation is similar to real world experiments. A starting structure, usually a macromolecule, is adequately prepared and put in a solvated box, with counterions to neutralize the charge (Figure 32); the minimization of potential energy function is needed to
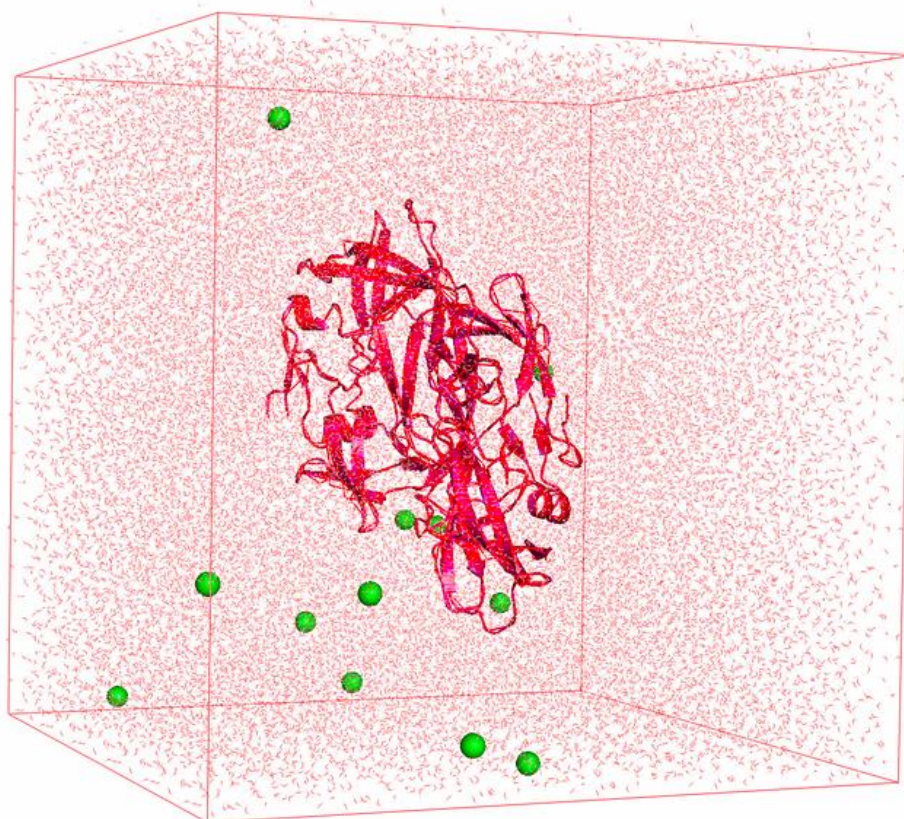
begin with dynamics calculations.

Figure 32 – A typical starting system for molecular dynamics calculation: the protein (red) is in the center of the cubic box; solvent molecule can be appreciated, and green spheres are chlorine ions neutralizing the total charge [https://www.compchems.com/setting-up-a-molecular-dynamics-simulation/].

The objective of MD simulations is to solve Newton's equation of motion and obtain particle trajectories (positions and velocities). The acceleration a_i of particle i is the derivative of the potential energy V with respect to the particle's position r_i is:

$$a_i = \frac{d^2 r_i}{dt^2} = -\frac{1}{m_i}\frac{\partial V}{\partial r_i} \qquad (Equation\ 12)$$

Where:

- a_i is the acceleration,
- r_i is the position of the particle,
- m_i is the mass of the particle,
- V is the potential energy.

The consensus prescribes selecting an appropriate time step for integrating equations, ideally less than half the period of the fastest oscillation according to Nyquist's theorem, often in the range of femtoseconds [68]. This choice is an optimization problem: a smaller time step increases computational cost without adding significant information, reaching convergence in an unfeasible amount of iterations, while a larger one can destabilize the system, leading to crashes.

The system obtained after energy minimization does not possess any velocity indication, therefore the starting velocities are assigned randomly, according to the Maxwell-Boltzmann distribution law (Figure 33) [77].
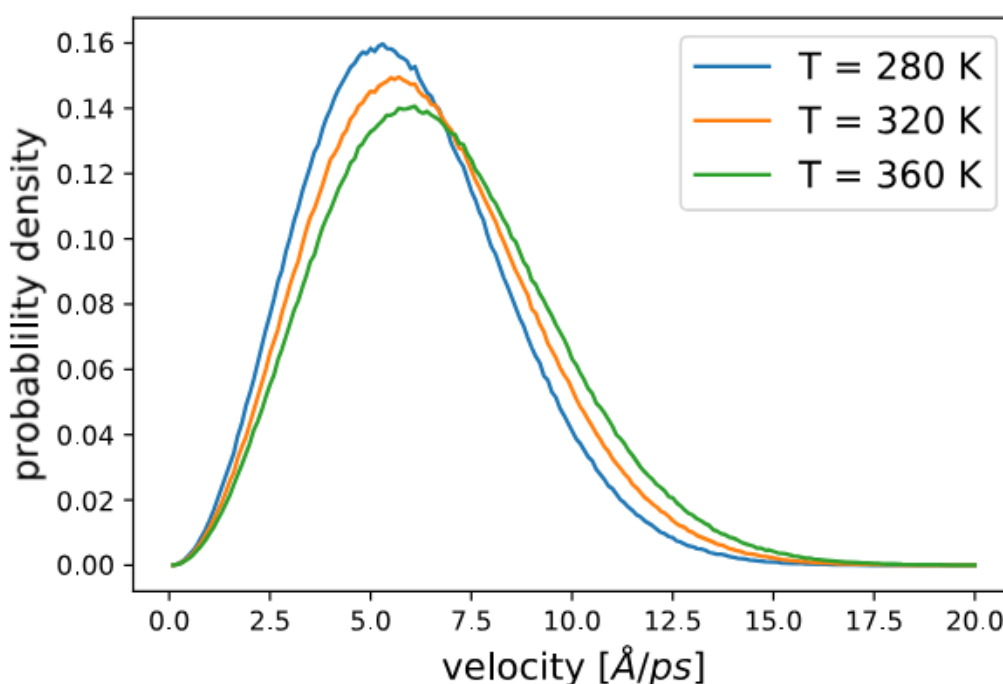

Figure 33 – Velocities of oxygen atoms plotted at different temperatures, following the Maxwell-Boltzmann distribution law [https://computecanada.github.io/molmodsim-md-theory-lesson-novice/aio/index.html].

Simulated systems are kept at desired pressure and temperature, by varying the equations of motion with the inclusion of respectively a barostat or a thermostat; the analogy with respect to real life experiments is evident using this nomenclature. Usually, the coupling with an "external" heat reservoir (for NVT and NPT ensembles), also known as "weak coupling" is preferred because the system can oscillate around the equilibrium temperature [68].

The most common used thermostats are Berendsen and V-rescale, while examples of barostats are C-rescale and Parrinello-Rahman.

A flowchart summarizing the main step of MD calculations is provided below (Figure 34). The relevant information (positions, velocities and forces) is stored in the output file for further analysis.
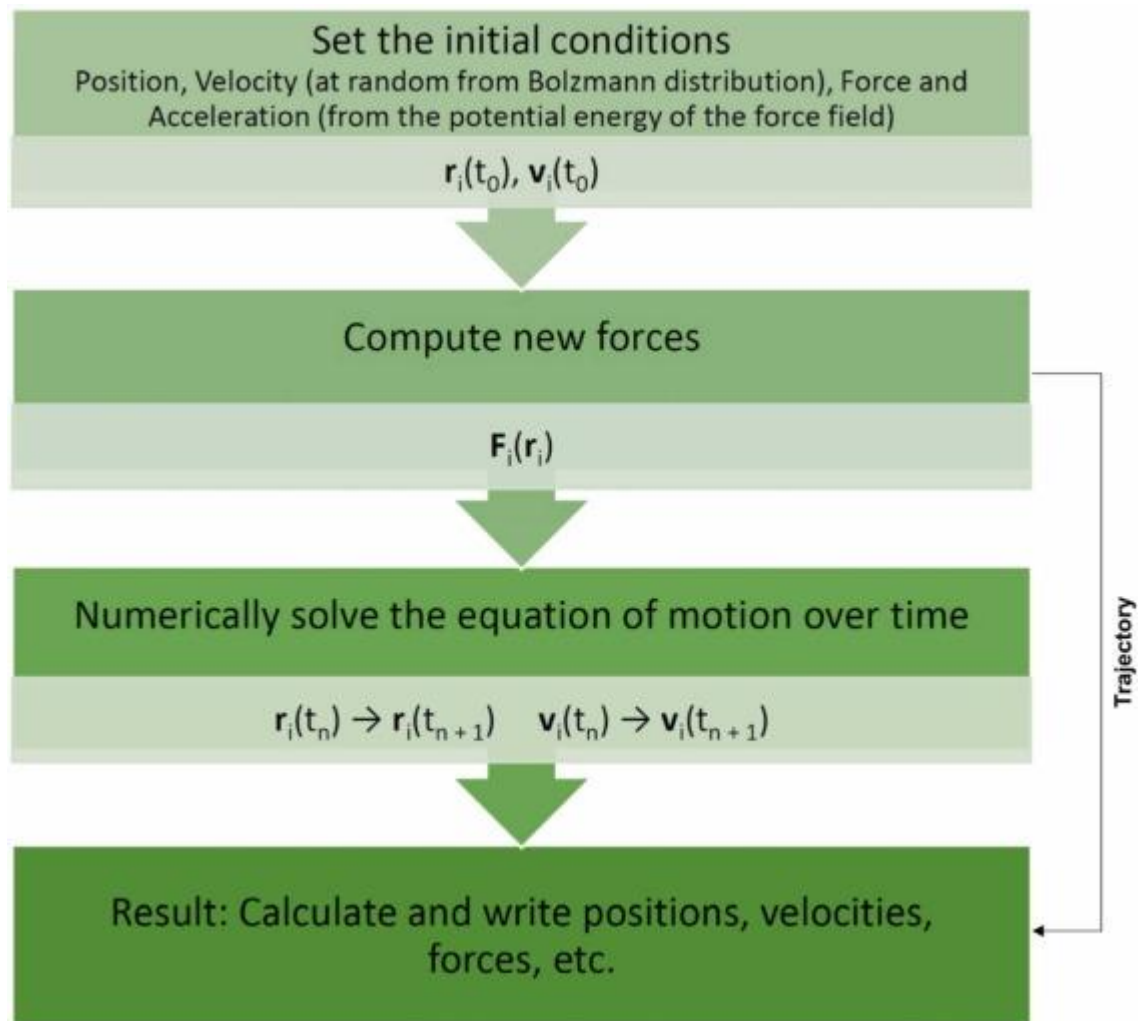

Figure 34 – MD algorithm scheme
[https://www.sciencedirect.com/science/article/pii/S0731708523006052?via%3Dihub].

3.4 Binding free energy predictions

The assessment of the affinity between a ligand and a receptor is fundamentally important in health-related applications such as drug design, as a ligand with potent and selective affinity for its target could become a promising drug candidate. While wet-lab procedures remain prevalent, they are increasingly complemented by in-silico assays aimed at speeding up the process and reducing costs, which are estimated to be around $1 billion for a single drug candidate [78]. The standard binding free energy for a non-covalent ligand-protein complex is related to the association constant K_a by the equation:

$$\Delta G° = -RT \ln K_a = \Delta H - T\Delta S \qquad (Equation\ 13)$$

where H is the enthalpy, S is the entropy, T is the absolute temperature, R is the gas constant, and the superscript ∘ indicates that the binding free energy is evaluated at standard temperature and pressure conditions. Therefore, to evaluate the binding energy, enthalpic and entropic effects must be taken into account (Figure 35). Enthalpic contributions consider standard terms from molecular mechanics theory, such as electrostatic and Van der Waals interactions, and the contribution to the solvation free energy, while the entropic contributions account for the flexibility of the molecules [79].

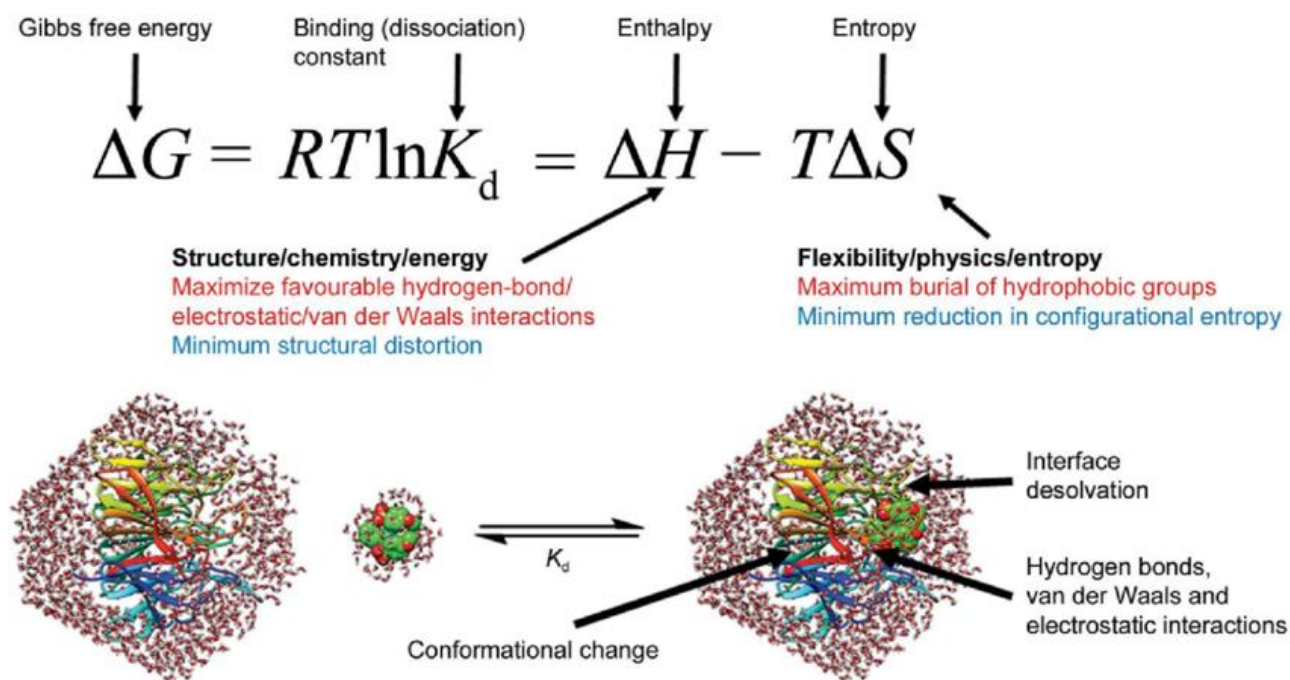A rough estimate of binding affinity is provided by scoring functions used in docking.

Figure 35 – Graphical representation of receptor-ligand binding with standard free energy formula explanation at the top
[https://www.sciencedirect.com/science/article/pii/S0731708523006052?via%3Dihub].

The theoretically precise alchemical perturbation (AP) methods suffer from high computational demands and the introduction of unphysical intermediate states. Consequently, approximate methods are typically used to calculate binding energy, balancing computational efficiency with good correlation to experimental data. Methods such as Molecular Mechanics Generalized Born/Poisson-Boltzmann Surface Area (MMGB/PBSA) achieve this balance, offering intermediate levels of computational burden and precision [80].

The binding energy is calculated as the difference between the energy term of the complex (PL) and the ones of the unbound receptor (P) and free ligand (L):

$$\Delta G_{\text{bind}} = G_{\text{PL}} - G_{\text{P}} - G_{\text{L}} \qquad (Equation\ 14)$$

Normally, only the complex is simulated (single trajectory approach), and the ensemble averages of the free receptor and ligand are calculated by removing the uninterested atoms from the entire structure, obtaining:

$$\Delta G_{\text{bind}} = \langle G_{\text{PL}} \rangle - \langle G_{\text{P}} \rangle - \langle G_{L} \rangle \qquad (Equation\ 15)$$

The one trajectory approach counterintuitively usually provides better results than simulation of the three states, mainly because of "filtering" of spurious ligand conformations not contributing to binding.

The energy in each state -P, L, and PL- is calculated according to the equation:

$$G = E_{\text{MM}} + G_{\text{solv}} - TS \qquad (Equation\ 16)$$

The first terms are standard terms from molecular mechanics potential energy function, such as bonded (bond, angle …), electrostatic, and Van der Waals interactions. $G_{\text{solv}}^{\text{polar}}$ and $G_{\text{solv}}^{\text{nonpolar}}$ are the contributions to the solvation free energy of polar and nonpolar terms. The last negative term is the entropic contribution (see below).

The solvent has a huge impact in protein-ligand binding and must be considered; implicit solvation methods are therefore preferred to reduce computational burden [81].

The polar contribution is calculated using a finite-difference solution equation in the MM-PBSA method (being a partial differential of the second order) or a Generalized Born (GB) pairwise approximation in the MM-GBSA approach [80]. The Poisson-Boltzmann equation (PBE) is defined by:

$$-\nabla \cdot [\varepsilon(r)\nabla\phi(r)] = 4\pi\rho(r) \qquad (Equation\ 17)$$

Where:

- $\varepsilon(r)$ is the dielectric distribution function,
- $\phi(r)$ is the potential distribution function,
- $\rho(r)$ is the fixed atomic charge density.

In the Generalized Born equation, the electrostatic contribution is given by:

$$\Delta G_{\text{GB}} = -\frac{1}{2}\left(\frac{1}{\varepsilon_{\text{solv}}} - \frac{1}{\varepsilon_{\text{solute}}}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{f_{ij}} \qquad (\textit{Equation } 18)$$

Where:

- $q_i\ q_j$ are the atomic charges,
- $\varepsilon_{\text{solv}}$ is the solvent dielectric constant,
- $\varepsilon_{\text{solute}}$ is the solute dielectric constant,
- f_{ij} depends on inter-particle distances and Born radii.

The nonpolar term is calculated using a linear relation to the solvent-accessible surface area (SASA):

$$\Delta G_{\text{nonpolar}} = \gamma \times \text{SASA} + b \qquad (\textit{Equation } 19)$$

Where:

- γ is the microscopic surface tension,
- b is a constant representing the cavitation term,
- SASA is the accessible area (Figure 36), measured by a probe molecule,

The sum of polar and nonpolar terms gives the solvation energy:

$$G_{\text{solv}} = G_{\text{solv}}^{\text{polar}} + G_{\text{solv}}^{\text{nonpolar}} \qquad (\textit{Equation } 20)$$
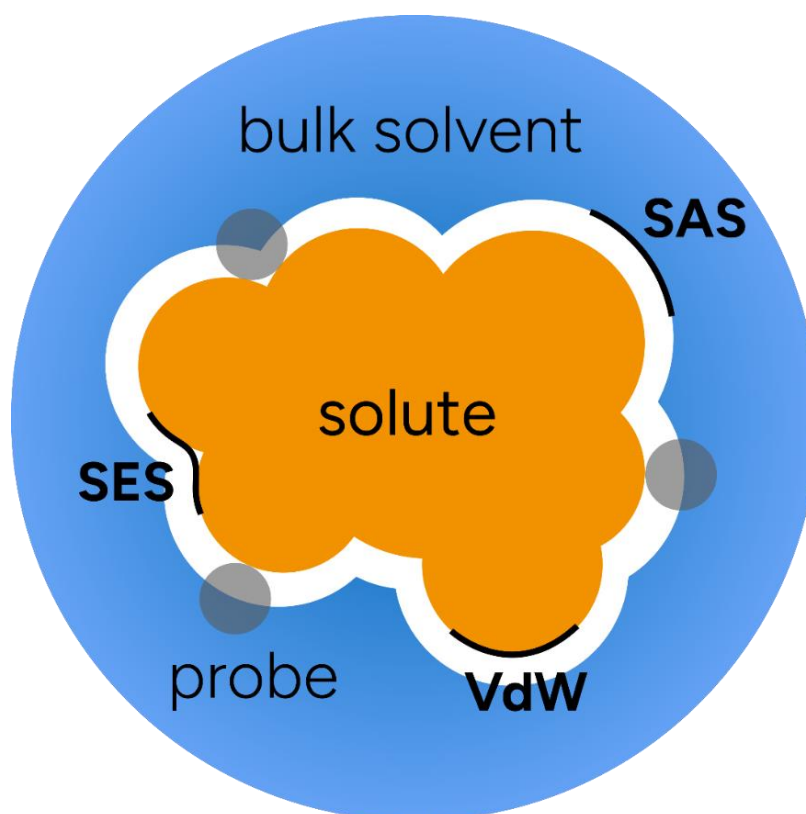
Figure 36 – Idealized representation of solute using Van der Waals (VdW) radii and calculation of Solvent Accessible Surface (SAS) using a spherical probe; the white region is excluded to solvent molecules (SES)[https://www.faccts.de/docs/orca/5.0/tutorials/prop/cpcm.html].

The entropy term in binding energy calculations can be evaluated using normal mode analysis (NMA), which requires analyzing the vibrational modes of the system to estimate the changes in entropy upon binding; this process is computationally expensive and can increase statistical uncertainty. As a result, many studies omit the entropy term unless calculating absolute free energy [80].

From a practical point of view, to obtain meaningful results, it is preferred to simulate a large number of short independent replicas -up to 100- and perform MM-GBSA calculations with 20-250 ns sampling to lower the high standard deviations which are typical of this method [82]. Another weak point of this technique is the overestimation of binding affinities differences between experimentally similar behaving ligands.

4. Molecular mechanics simulations of mutated BACE1-APP complexes

4.1 Abstract

Mutations of amyloid precursor protein (APP) are a well-known cause of early-onset Alzheimer's disease (AD). Pathogenic mechanism is believed to lie in altered processing of APP by enzymes called "secretases", resulting in overproduction of a putative neurotoxic peptide species, especially amyloid beta (Aβ) of 40 and 42 amino acid length.

In 2012, an APP mutation, "Icelandic", consisting in an alanine to threonine substitution at residue 673 (A673T), demonstrate protective effects, lowering AD and other age-related dementia risks. At the same site, the valine substitution (A673V) is pathogenic and linked to early-onset dementia. Moreover, the Swedish mutations, consisting in a double substitution at sites 670 and 671, is pathogenic.

The proximity of those mutations to the beta cleavage site (671-672) by beta amyloid cleaving enzyme 1 (BACE1) resulted in extensive research to understand their effects, as well as the development of BACE1 inhibitors to stop Aβ production. However, beta secretase targeting failed to show improved cognition, and the mechanistic effect of A673T mutations are far from being completely elucidated: some authors showed a decrease in Aβ production or a different aggregative behavior, reducing oligomers building-up.

Performing docking of an APP helical fragment to BACE1 enzyme, four systems were generated (wild type, A673T, A673V and Swedish) and each one was simulated for 1.8 microseconds, considering the sum of three replicas.

Trajectory analysis showed a decrease in complex stability for every mutated system, particularly for APP interactions with flap domain of BACE1, involved in substrate recognition. MM-GBSA method was used to give an estimate of binding affinity, revealing that APP expressing the protective A673T mutation is a worse BACE1 binder, while the opposite effect is observed in the case of A673V.

4.2 Introduction

Alzheimer's disease (AD) is the most common cause of dementia worldwide and despite decades of research, drug candidates and clinical trials it remains a challenging condition. Apart from palliative drugs, there is a lack of disease reversing therapies; the impact of AD is enormous on healthcare systems, and it is projected to grow with a progressively ageing population [3].

While most AD cases are without a clear proximal cause, the production and accumulation of aberrant amyloid beta (Aβ) peptides -in particular Aβ42 and Aβ40- in patients' brains are considered as cornerstones of the disease and linked to the complex biochemical cascade which can start up to 20 years before diagnosis [27].

Amyloid precursor protein (APP) is an integral membrane protein with several putative functions and undergoes processing by enzymes called secretases: the "physiological" catabolic pathways is believed to start with alpha secretase mediated cleavage at residues 681-682, followed by gamma cleavage resulting in a soluble APP alpha, P53 fragment and an intracellular fragment [36]. Another pathway is described as pathogenic, with first cleavage by beta secretase (BACE1) at residues 671-672, followed by gamma cleavage: the secreted amyloid beta peptides become insoluble following misfolding and aggregates in oligomers and then fibrils [31]. Those aggregates are believed to cause harm to neural homeostasis, via disruption of vesicular transport, synaptic transmission and inflammation.

Moreover, mutations of APP are pathogenic, causing early onset AD: the proposed mechanism is either the alteration of secretases' cleavage or an increase in aggregation propensity of mutated amyloid peptides [30].

Near the beta secretase site, a double substitution K670M-N671L is a well-known familial AD mutation: the resulting APP is a more favorable substrate for BACE1 cleavage and production of beta amyloid is increased [42].

In addition to this, at site 673, the alanine to valine substitution (A673V) is believed to act in the same way, shifting APP processing from the alpha mediated cleavage to beta, but also increasing toxicity and aggregation potential of mutated Aβ. This mutation appears to cause AD in a recessive fashion and in-vitro analysis with a mixture of wild type and mutated Aβ demonstrated a slower aggregation, because of unfavorable intermolecular interactions [45].

Interestingly, at the same site, an alanine to threonine substitution confers protection against AD and is correlated with longevity and higher cognition in later years [46]. Several A673T carriers showed little to no amyloid deposition and were cognitively intact, cerebrospinal fluid (CSF) Aβ concentration was 40% lower [56].

From its discovery in 2012, a precise characterization of its biological effects is lacking, with mixed results from literature: reduced affinity for BACE1 [46]; shifting of beta cleavage from beta to the alternative beta' (between residues Tyr681-Glu682) site, resulting in a shorter secreted peptide [52]; lowering of Aβ aggregation, with dodecamer formation inhibited; and lower toxicity of mutated Aβ [53].
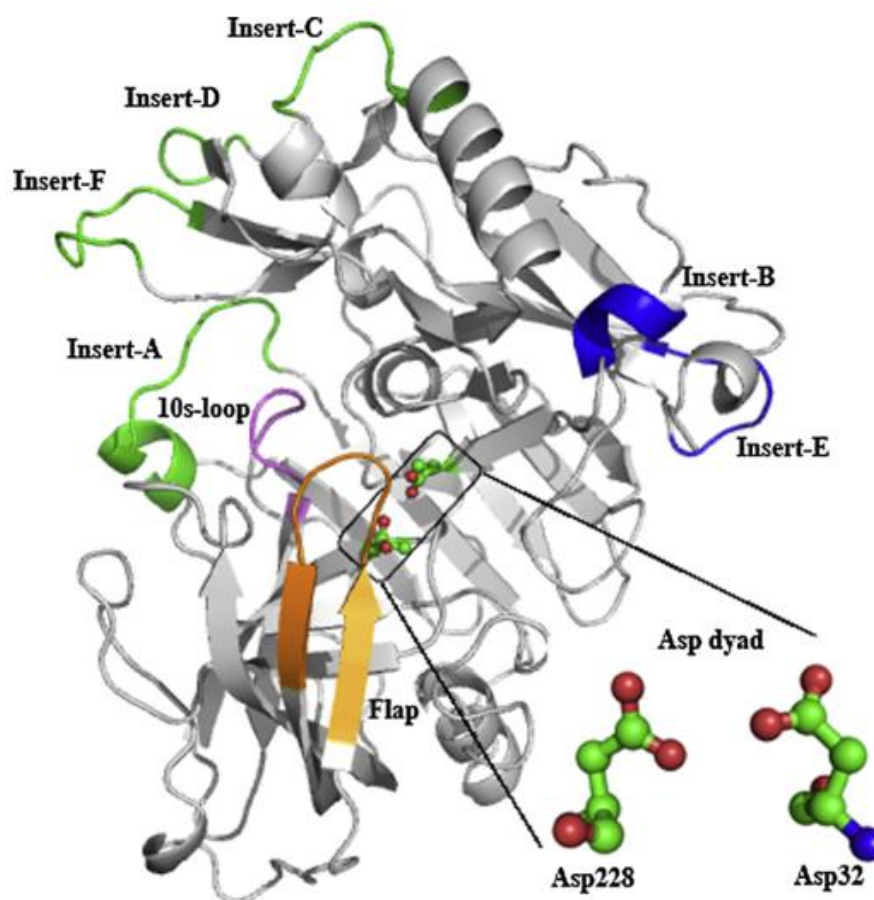

Figure 37 – BACE1 structure with highlighted relevant regions of APP binding. The catalytic dyad composed of Asp residues is magnified in the bottom right corner [http://dx.doi.org/10.1016/j.jmgm.2012.12.010].

In this context, molecular modelling could be a useful strategy for elucidating the mechanisms lying behind the phenotypic effects (dementia causing mutations versus protective allele): the atomistic resolution and the trajectory analysis, paired with tools to estimate binding affinity, offer an interesting framework [68].

In this work, wild type and mutated APPs in complex with BACE1 will be simulated using molecular docking and dynamics, providing also an estimate of the affinity, with the aim of characterizing the effects of APP mutations on BACE1 binding.

4.3 Materials and methods

4.3.1 Modelling of APP-BACE1 complex

The starting structure of BACE1 was retrieved from the Protein Data Bank. An X-ray diffraction structure with 1.75 Å of resolution was chosen (PDB: 1W50), representing BACE1 in its apo form, in the flap open position [83]. The missing residues (158-167) were modeled as a loop with MOE software package. Following the experimental and theoretical consensus, the aspartate dyad was considered monoprotonated, with Asp32 and Asp228 respectively in the protonated and unprotonated (negative) form [35], [84].

The APP was not modeled in its entirety because of the dimensions and the presence of the transmembrane and intracellular domains: only a fragment, from residue 655 to 689, was considered in this work, as done before [43], [85], [86].

The APP fragment was modeled with MOE protein builder tool choosing a helical secondary structure [85], starting from the FASTA sequence of interest (UniProt identifier P05067). The result was compared with AlphaFold3 prediction of the same fragment.
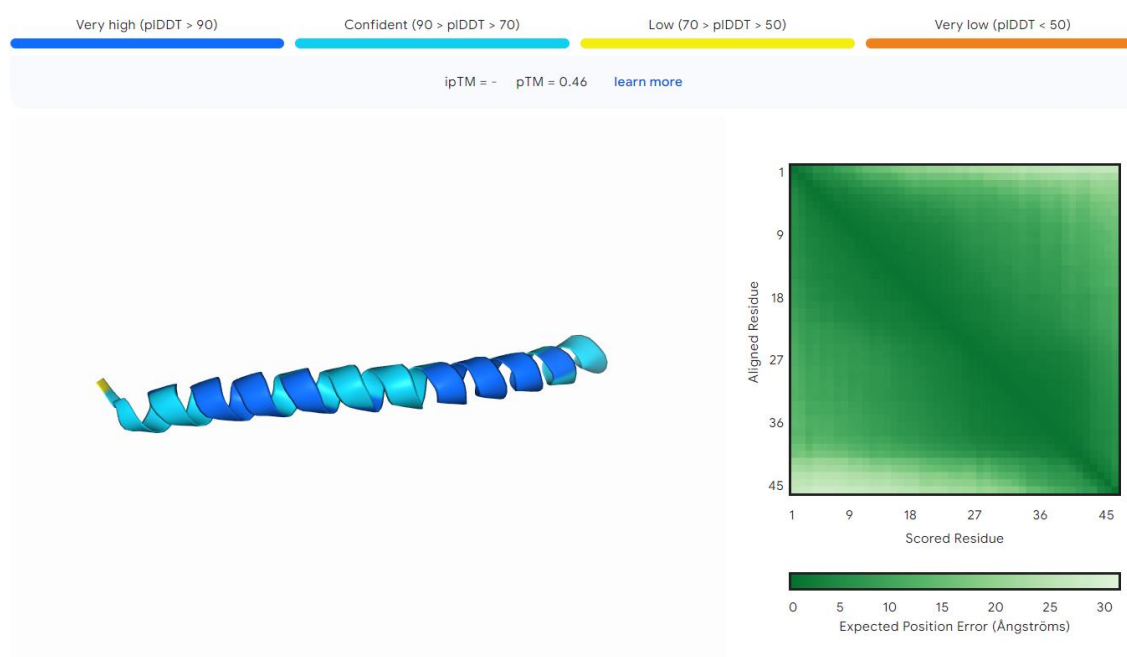

Figure 38 – AlphaFold 3 predicts the APP fragment as helical with high confidence scores (plDDT) [Query: GSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLV on https://alphafoldserver.com].

Additionally, solved structures of Aβ, starting from residue 672, show a helical shape in solution (Figure 38), consistent with the chosen model [87], [88].
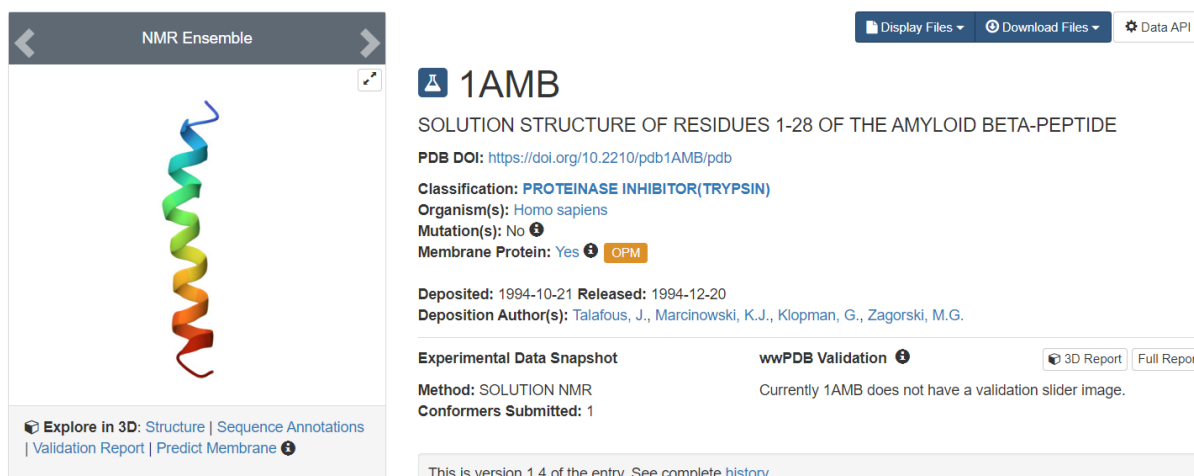

Figure 38 – NMR structure of Aβ peptide residues 1-28 (672-699 in standard APP sequence), showing helical secondary structure [https://www.rcsb.org/structure/1AMB].

4.3.2 Molecular docking and mutations

ClusPro server [75] was used to dock the APP to the prepared BACE1, with default settings, selecting the structure with the scissile bond (Met671-Asp672) in proximity of the Asp32 and Asp228 residues to mimic physiological cleavage process [52]. Modifying settings, such as imposing ligand-receptor residue attraction (Asp dyad and scissile bond) did not improve results.

From this docked structure, four systems were considered and mutant peptides generated using MOE:

- APP wild type – BACE1 was the original docked structure (APPwt)
- APP A673T – BACE1 (APPa2t)
- APP A673V – BACE1 (APPa2v)
- APP K670L-M671N (APPsw)

4.3.4 Simulation set up

Molecular dynamics simulations were performed using GROMACS [89]. The AMBER ff99SB-ILDN force field was chosen [90], and the systems were solvated using explicit water model (TIP3P) in a cubic box with minimum distance between solute and box of 1 nm; periodic boundary conditions (PBC) were applied [68]. Ions, sodium and chlorine, were added to

neutralize the charge and maintain a physiological saline concentration of 150 mM. Systems were at first minimized using the steepest descent algorithm for 5000 steps with maximum force of 100 kJmol^{-1}nm^{-1}.

Each system was then equilibrated in NVT and NPT ensemble with position restraints applied on protein atoms: the NVT equilibration at 300K for 500 ps using velocity-rescale thermostat with constant tau equal to 0.1; the NPT equilibration for 1000 ps using the Berendsen barostat to maintain the pressure at 1 atm. Particle Mesh Ewald (PME) method was utilized for electrostatic non bonded calculations with 1.0 nm cutoff, Fourier spacing of 0.12 nm and interpolation order of 4 because of its efficiency using GPU [91].

The production run was 600 ns without any position restraint with a time step of 2 fs, using the leap-frog algorithm.

To ensure repeatability of the results, three replicas per system were created with random velocity generation before starting the equilibration phase.

4.3.5 Analysis

To assess the stability of each system, root-mean-square deviation (RMSD) of BACE1 backbone from the initial configuration was calculated. Following visual inspection, the last 400 ns of each trajectory were used for the analysis.

In addition, the flexibility of the system was evaluated by calculating the root-mean-square fluctuation (RMSF) of the C-alphas, for both ligand and receptor.

The distance between BACE1 residue Thr72 and Ser325 was computed and plotted as a distribution (Kernel Density Estimation). The number of hydrogen bonds (h-bonds) between APP and BACE1 was calculated for every system. Mean and standard deviations for these quantities were calculated from the three replicas. Distances among APP 671-673 and BACE1 were also computed.

To perform the above-mentioned analyses, GROMACS have been used and the plots were created with custom scripts and python libraries (mostly matplotlib) [92].

Binding free energy calculation was performed using the gmx_MMPBSA tool [93], selecting the Generalized Born model GB-Neck2 and maintaining default options. According to literature [94], to improve results, we performed a windowing of the trajectory considering 10 ns every 100 ns, starting from 200 ns (four windows per replica). The windows were then concatenated and analyzed with sampling every 200 ps.

MOE, VMD [95] and ChimeraX software were used to visualize the systems and take snapshots.

In particular, ChimeraX built-in function to calculate surface electrostatic potential and hydrophobicity were employed [96].

4.4 Results

4.4.1 Docking validation

The first step was the validation of the model generated with docking of BACE1 and APPwt. In Figure 39 we observe the APP, modeled as helix, correctly positioned inside the active pocket. Moreover, the BACE1 flap domain is in close contact with APP (see also Supplementary material).
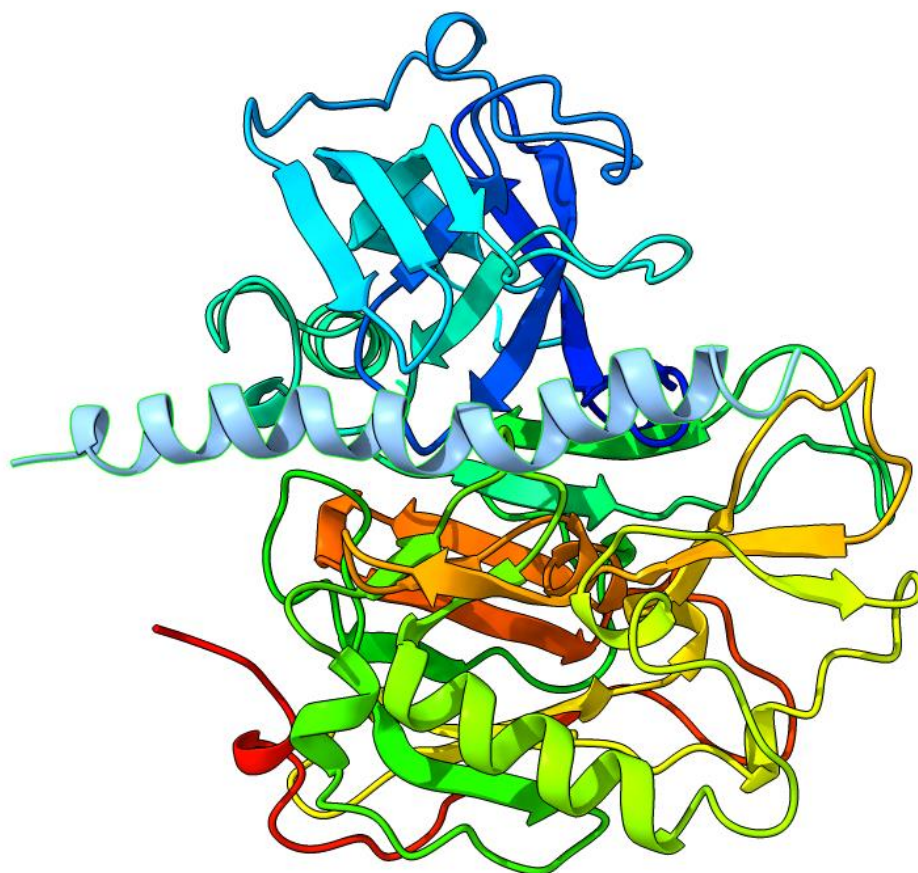
Figure 39 – Snapshot of the APPwt-BACE1 docked system taken with ChimeraX; APP is the light blue helix.

The catalytic dyad (Asp32 and Asp228) is in the proximity of the APP residues Met671-Asp672 (Figure 40), corresponding to the scissile bond which undergoes cleavage. This indicates that docking was able to correctly predict this three-dimensional arrangement, which is physiologically relevant; moreover, this is a consistent starting point for subsequent dynamics simulations.
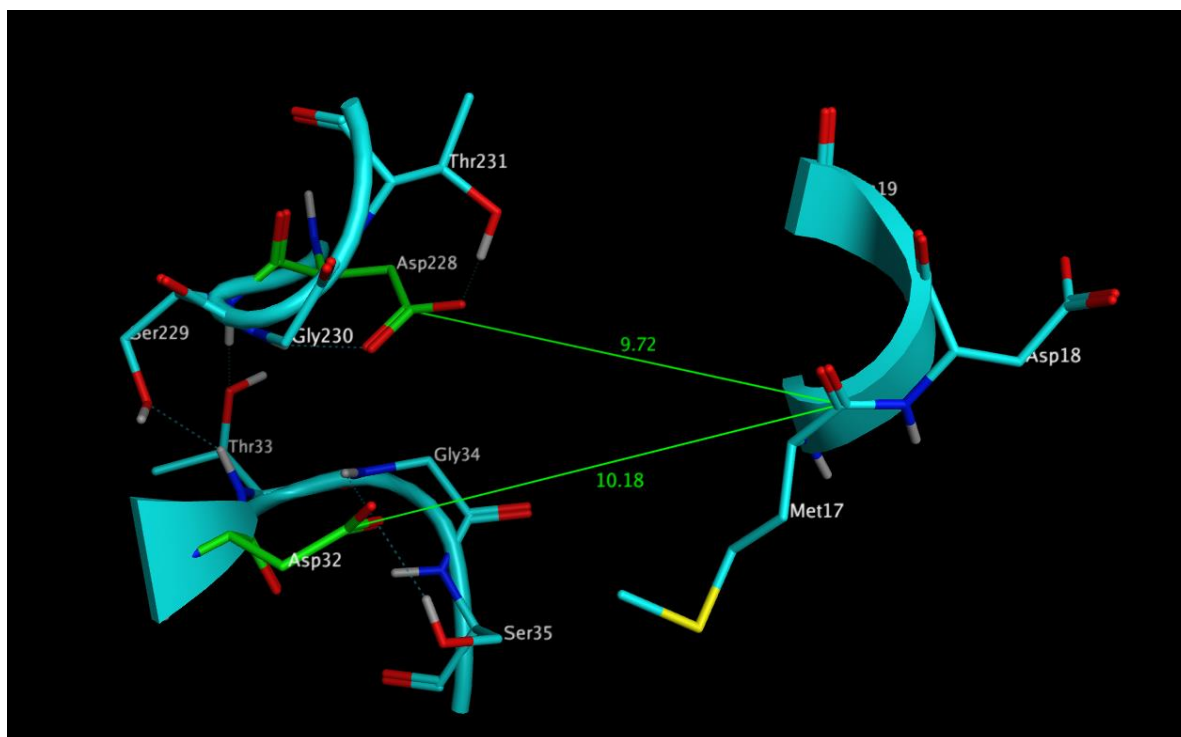
Figure 40 – Detailed view of the docked system, showing on the left the aspartate dyad (Asp32 and Asp228) and on the right the scissile residue Met671, with distances in angstroms; the snapshot was taken with MOE.

The Ramachandran plot (Supplementary material), showing the backbone dihedral angles Phi and Psi, reveals a structure with a really limited number of unallowed torsional angles; nevertheless, those ones are far from the binding pocket and were present in the starting crystallographic structure.

The analysis of the electrostatic surfaces revealed that BACE1 active site is predominantly positively charged, while APP possesses electron excess (Figure 41). In particular, APP residues Glu668, Asp672 and Glu674 (negatively charged sidechain) are located near the scissile bond and associated with the positive environment of BACE1 active pocket.

See also Supplementary material for the effect on the electrostatic surface of the Swedish mutation.

Rendering of molecular lipophilicity potential surfaces reveals that BACE1 active site is mostly hydrophilic (Figure 42), apart from a hydrophobic region deeply buried.
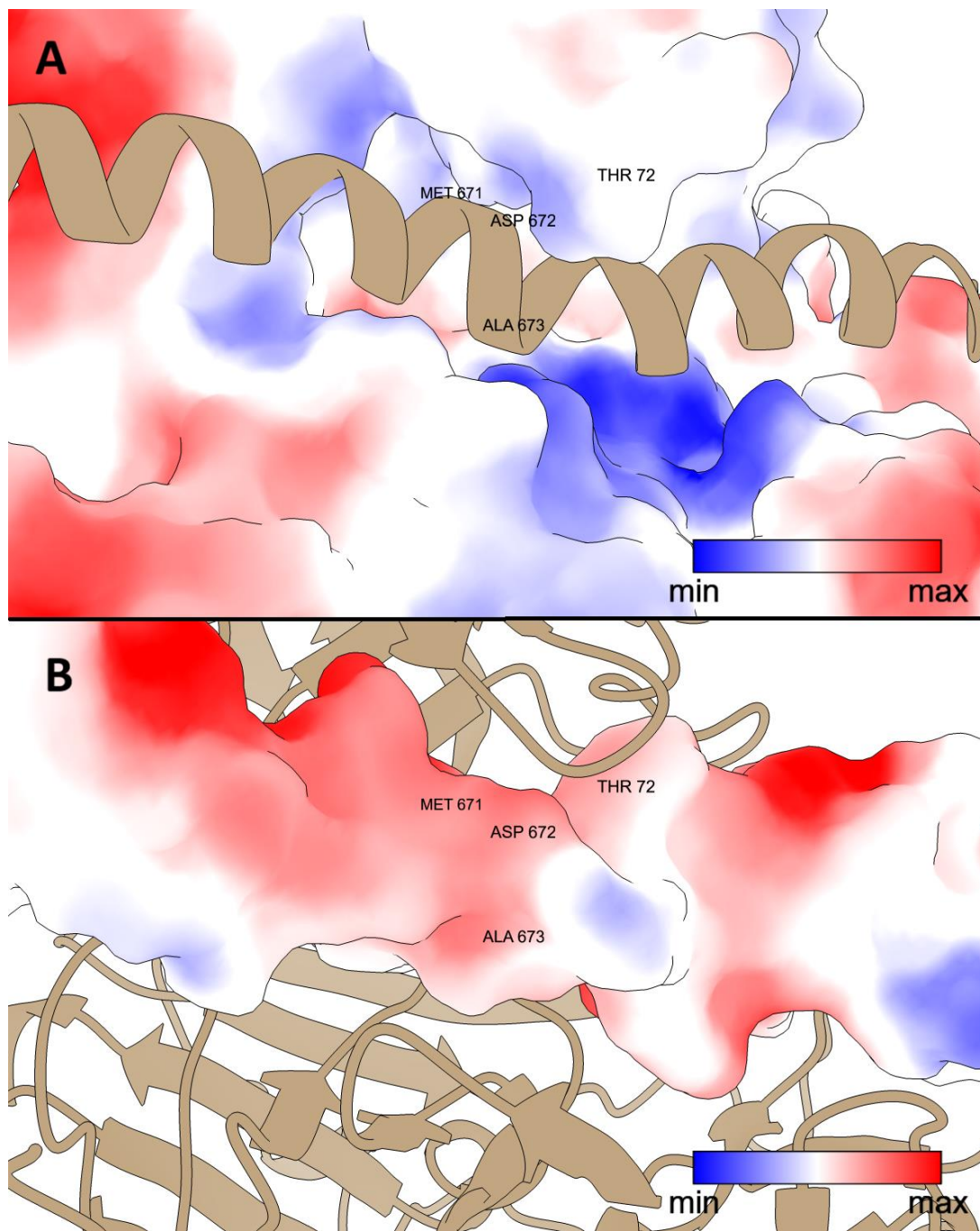
Figure 41– Coulombic electrostatic potential surfaces of BACE1 (A) and APP (B), with blue indicating regions with electron deficit and red negatively charged zones. Snapshot taken with ChimeraX.
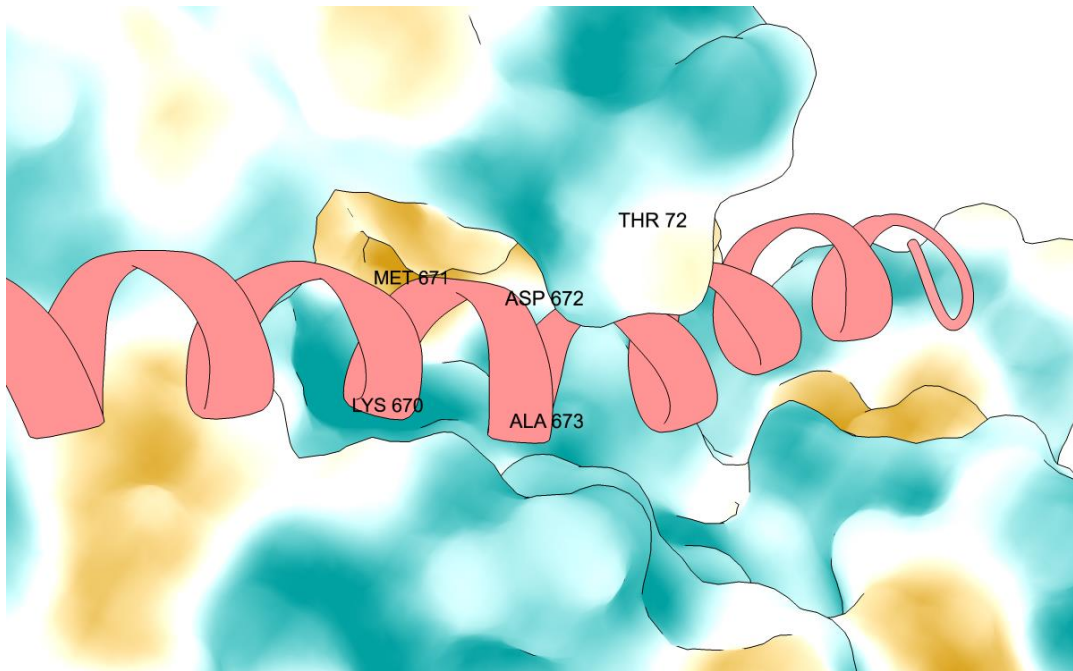
Figure 42 – Hydrophobicity surface of BACE1, APP rendered as a pink helix. Light blue zones are hydrophilic while yellow are hydrophobic.

Mutated systems displayed different surfaces (Figure 43): Val and Thr are respectively hydrophobic and polar amino acids, resulting in deviations of hydrophobic surfaces compared to the wild type, expressing Ala.

The Swedish double mutation similarly alters the landscape, showing a more hydrophobic surface (Supplementary material). In fact, the double substitution results in an APP expressing two more hydrophobic amino acid residues.
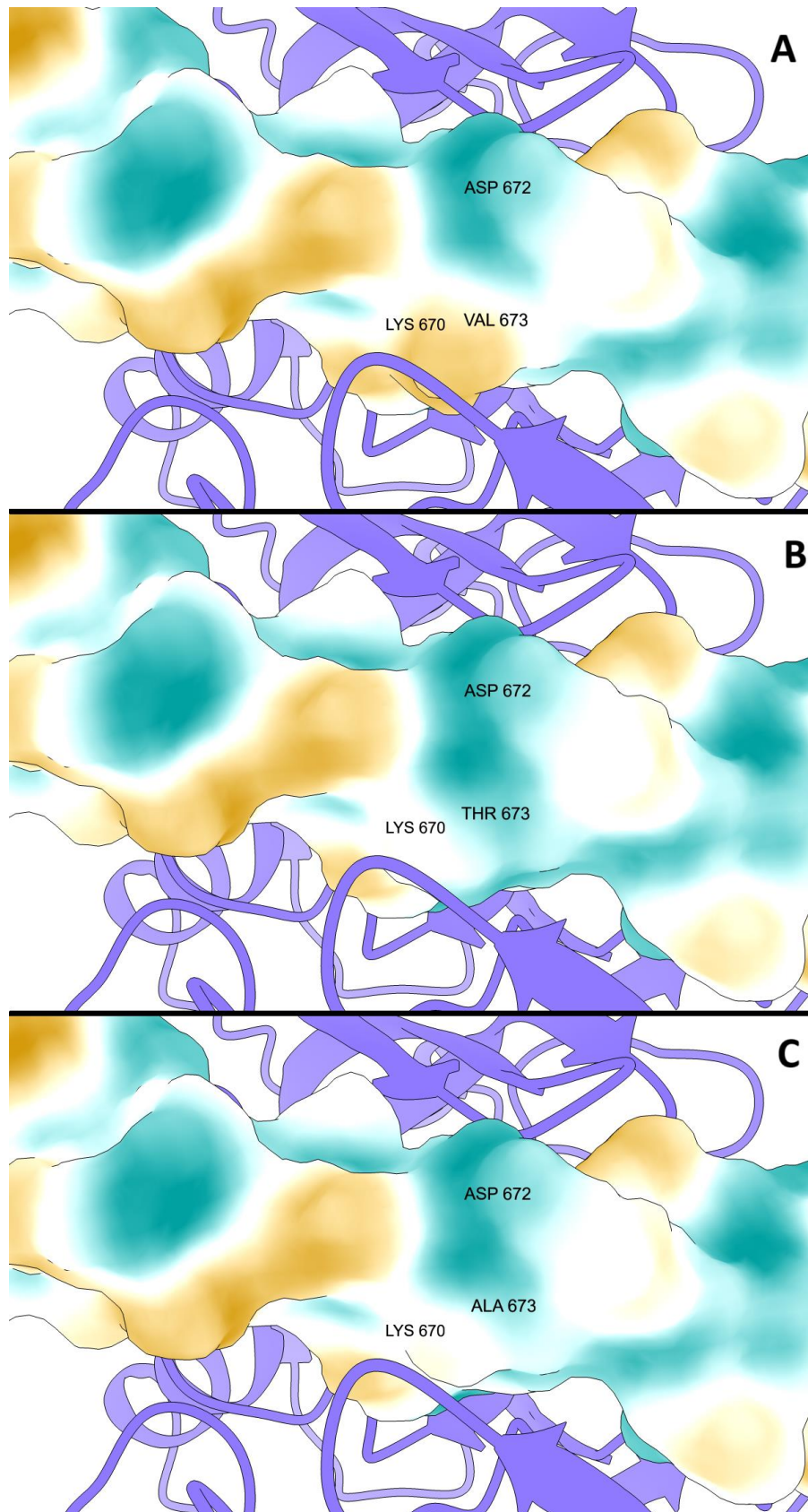
Figure 43 – APP hydrophobic surface (from hydrophilic in blue to yellow) for APPa2v (A), APPa2t (B) and APPwt (C). Image created with ChimeraX.

4.4.2 Trajectory analysis

Minimization and equilibrations phases were correctly performed (see Supplementary material).

The trajectories were first analyzed to assess the stability of the complex: the backbone RMSD of BACE1 showed convergence from 200 ns for every simulated replica (Figure 44). The deviations from the starting structure were under 0.35 nm in every case.
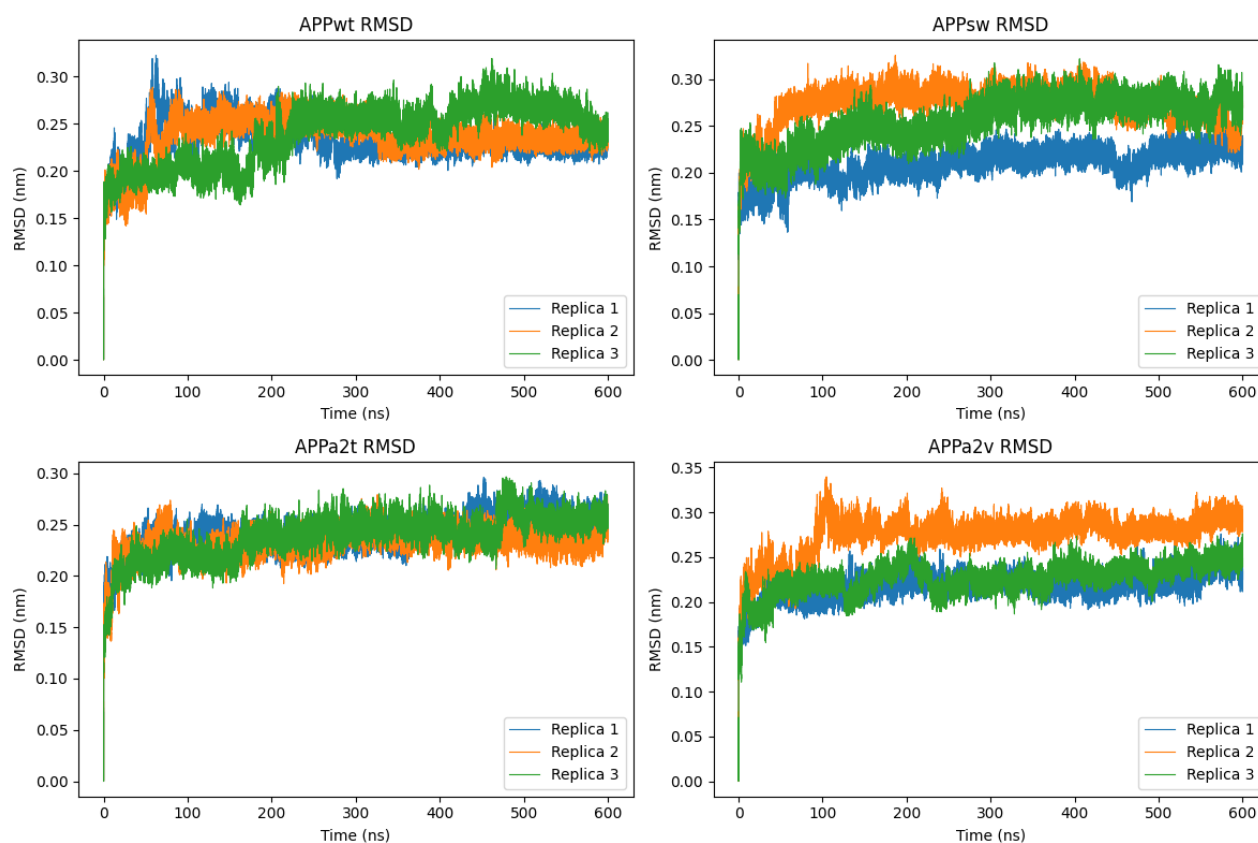

Figure 44 – Backbone RMSD evolution of BACE1 for the four systems as a function of time.

Visual inspection of trajectories and structures extracted at 300 and 600 ns was performed to verify structure stability and correct peptide placement (data not shown). In some replicas the APP termini, outside the pocket, underwent unfolding, while the central residues (667-676) were more stable. Therefore, the APP RMSD was far from convergence, because of those boundary effects, and was not considered in the following as an indicator of stability.

A 600 ns simulation with BACE1 without APP was also performed as control. RMSD was moderately higher (around 0.3 nm) and did not steadily reach convergence (see Supplementary material): visual inspection and RMSF analysis (Figure 45) of free BACE1

compared to APPwt system revealed widespread higher fluctuations. In fact, previous studies reported the increased stability of BACE1 upon APP binding [85], [86].
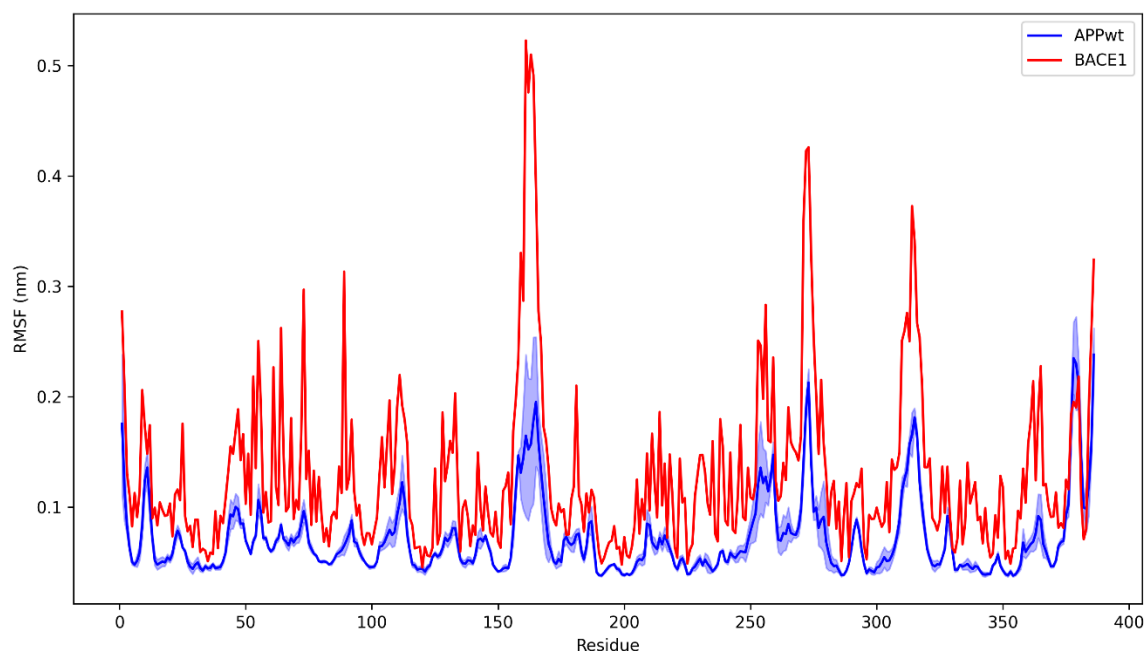

Figure 45 – RMSF comparison of free BACE1 (red line) and APPwt complex, shaded blue area is the standard deviation.

In order to assess the flexibility of important protein domains, RMSF of BACE1 and APP were computed on the three replicas and the mean values extracted. The comparison of BACE1 RMSF upon binding with different APP fragments is provided below (Figure 46): the most fluctuating regions are BACE1 termini, loop A (Gly158-Leu167), loop F (Asp311-Asp317) and loop D (Trp270-Thr274), showing mean fluctuations higher than 0.15nm.
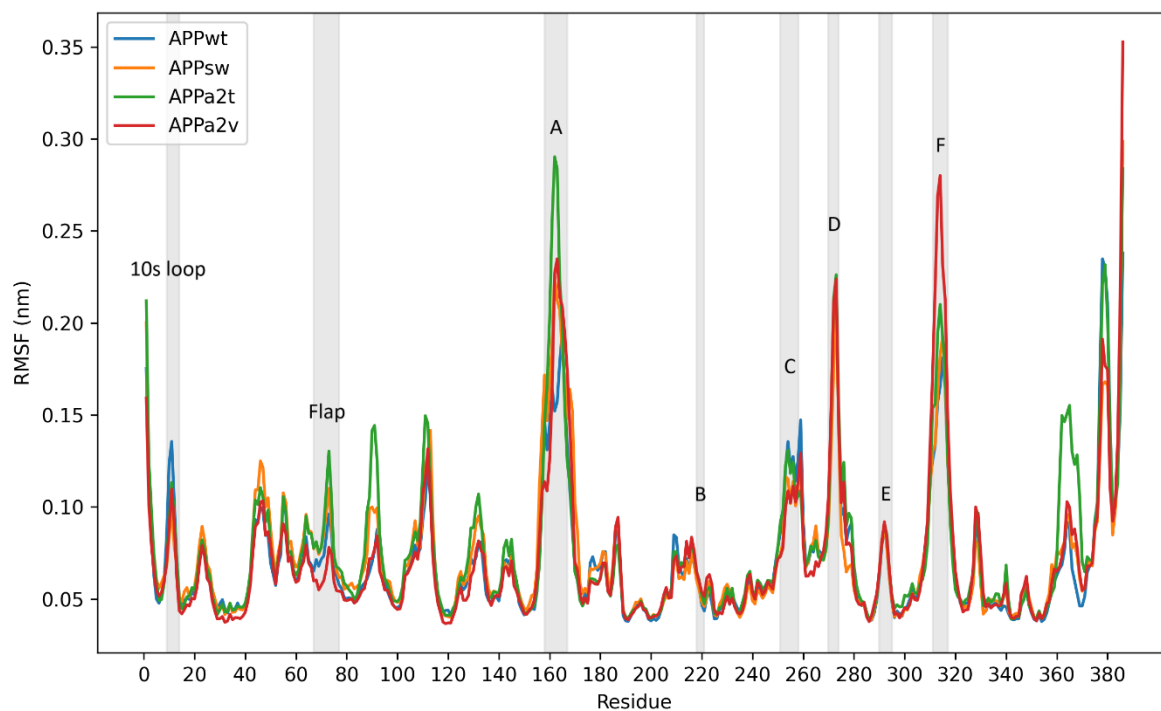
Figure 46 – C-alpha mean RMSF of BACE1

Differences amongst the systems were observed particularly in loop A and flap (Val67-Glu77) domains. RMSF of APPa2t was nearly two times higher than that of APPwt in residues 161 to 163 (Figure 47). The flap residues displayed higher fluctuations upon BACE1 binding with APPa2t than APPwt, while APPa2v resulted in a opposite effect (Figure 48). However, it has to be noted that the standard deviations for the flap regions are elevated, and in this case the difference is not clearly discernible.
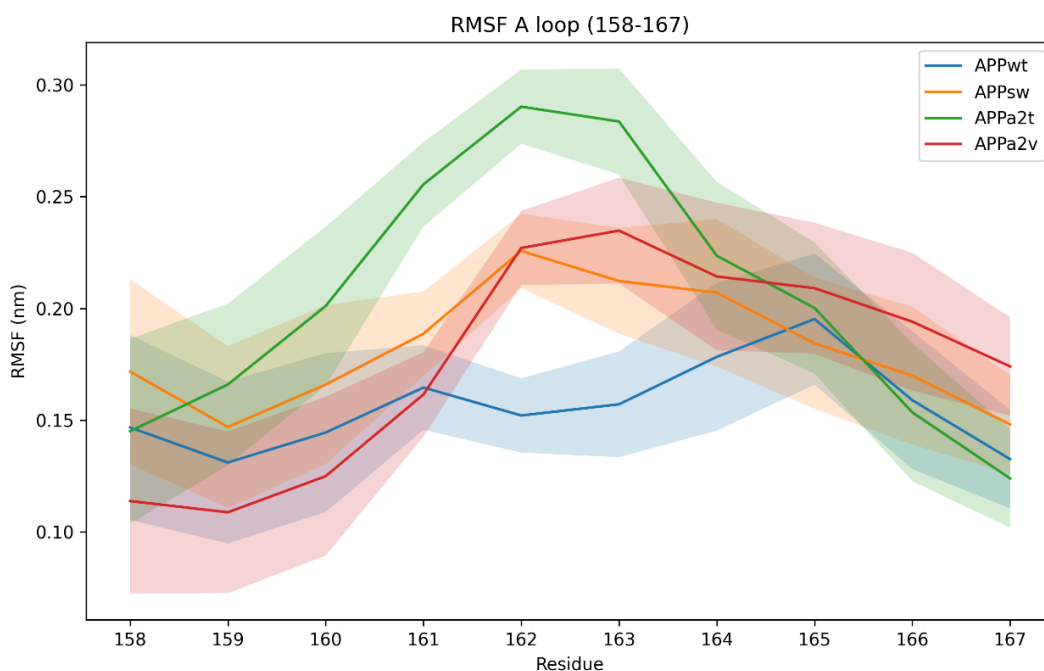
Figure 47 – Detail of RMSF of BACE1 residues 158-167 (A-loop) with standard deviations represented as shady areas.

The RMSF values of residues Thr72, Ser325, Asp32 and Asp228 were then compared (Figure 49): the Asp dyad is stable with small fluctuations (on the order of 0.05 nm) without any appreciable difference among the systems as well as the Ser325 residue. Thr72 showed higher fluctuations in the APPa2t system compared to APPa2v. Interestingly, the flap residue Thr72 deviations between the replicas (the standard deviations) were remarkably higher than those of catalytic Asp residues.

The RMSF of the APP presented higher fluctuations at the termini (as acknowledged during visual inspection) and lower in the scissile region, not revealing any noticeable change from one system to another. The data are reported as Supplementary material.
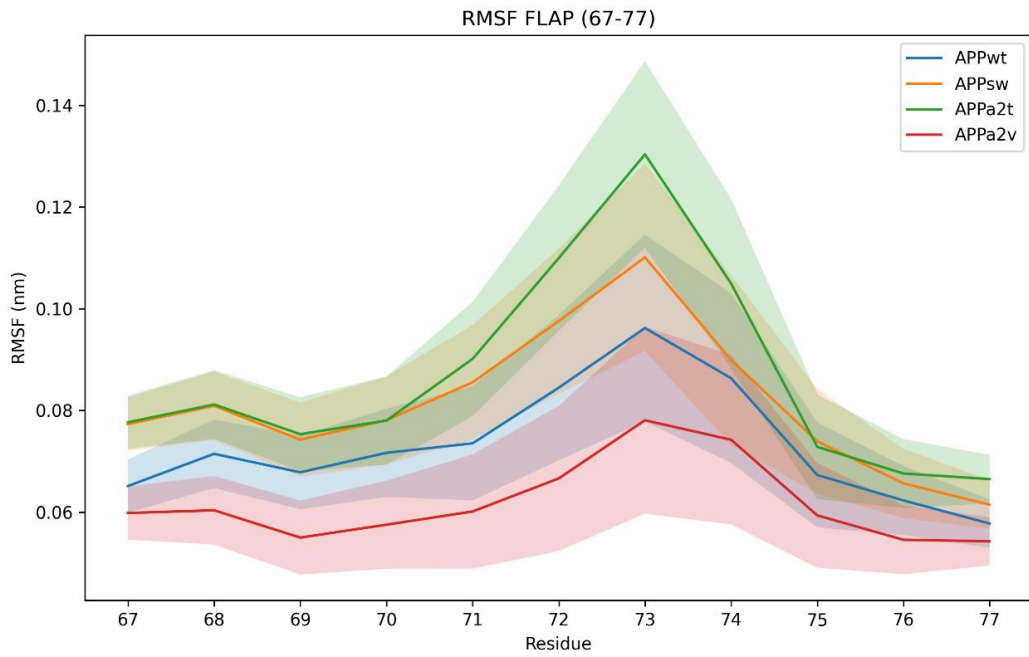
Figure 48 – Detail of BACE1 flap region (residues 67-77) RMSF with standard deviations depicted using shaded regions.
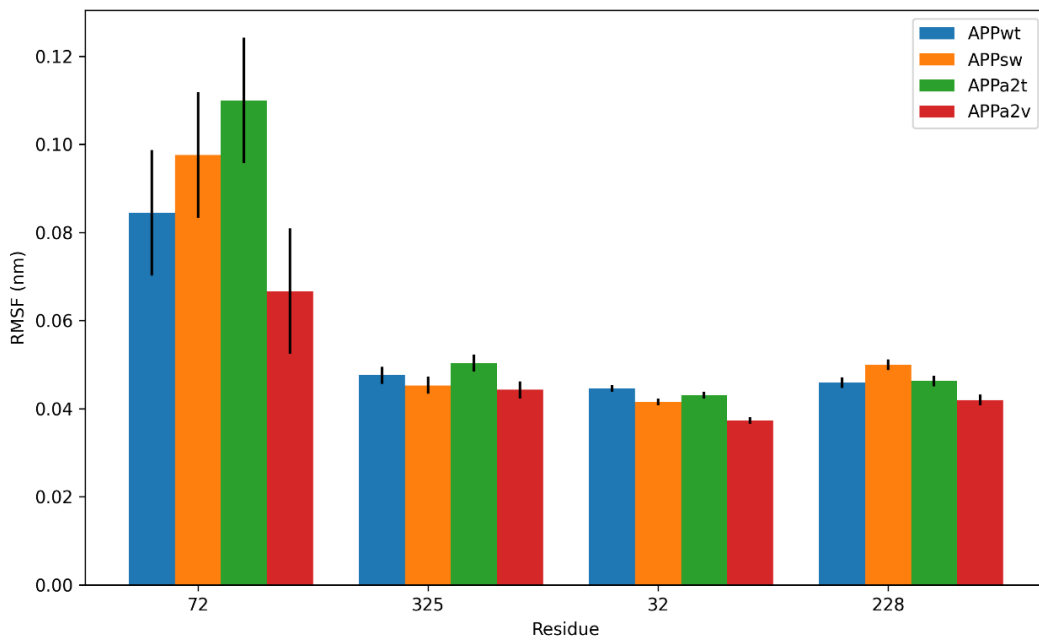

Figure 49 – Bar plot of mean and standard deviation of BACE1 residues Thr72, Ser325, Asp32 and Asp228

The interactions between APP and BACE1 were analyzed at first by looking at relevant trajectory snapshots: while APP extrema exhibited frequent unfolding and high fluctuations, with poor binding to BACE1, the residues around the scissile bond were more stable. The flap residues were in close contact with APP, particularly Thr72, expressing hydrogen bonding with peptide residues 671 to 673. In Figure 50, it is possible to visualize the hydrogen bonds among BACE1 Thr72, Asp672 and Thr673.
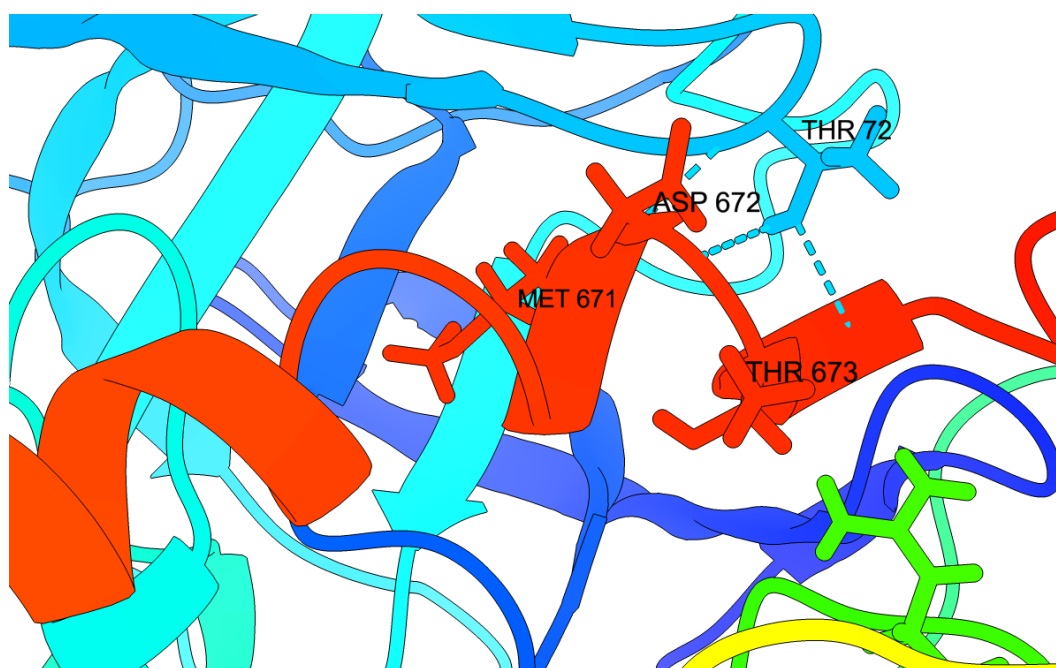

Figure 50 – Detail of BACE1 flap region (blue loop at the top) interaction with APPa2t (red) first replica at 300 ns; dotted blue lines represent the hydrogen bonds of flap residue Thr72. Atoms of Asp672, Met671 and Thr673 were depicted using stick convention. Image created with ChimeraX.

However, having the aim of extracting meaningful patterns, analysis of isolated time frames is insufficient. Therefore, calculation of hydrogen bond count during the course of the trajectory was performed, providing mean and standard deviation values (Figure 51), as well as distributions (Supplementary material).

The number of BACE1-APP hydrogen bonds, considering residues from 667 to 676, was in the range of three to six bonds, probably because of the helix structure reducing the number of potential bonds. A difference was observed, despite the standard deviations are elevated, between APPa2t and APPa2v, expressing respectively a higher and a lower amount of bonds compared to the wild type. This could be caused by the fact that threonine (in APPa2t) is polar while alanine (APPwt) and valine (APPa2v) possess hydrophobic side chains.
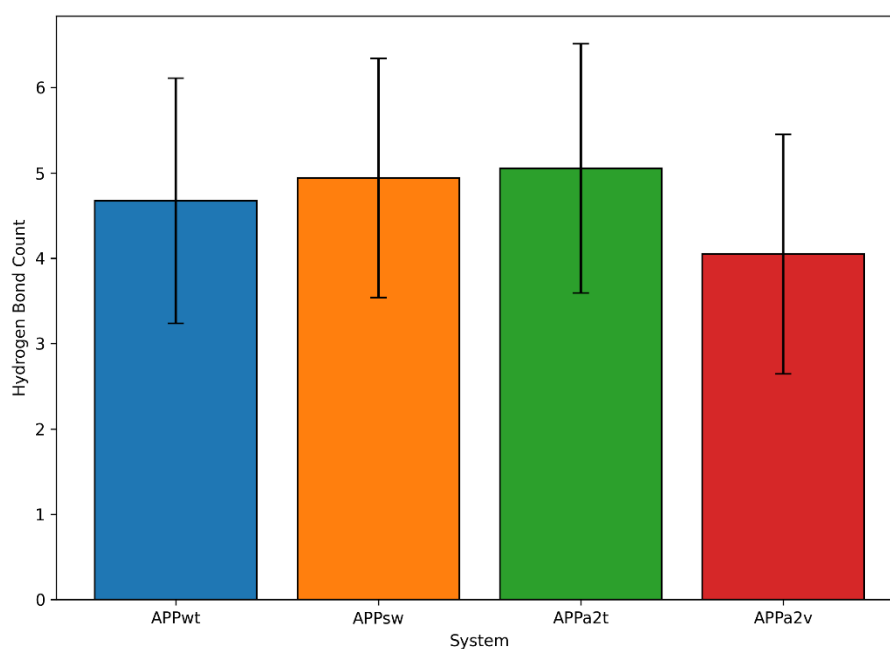
Figure 51 – Bar plot with mean and standard deviation of hydrogen bond counts between BACE1 and APP residues 667-676.

After analyzing hydrogen bonds, contact probabilities were calculated as the ratio of time frames spent under a specific cutoff distance to the total number of frames. Different cutoff distances were selected: 0.6 nm to identify non-bonded interactions and 1 nm as an upper interaction boundary, beyond which van der Waals forces are negligible [68]. Additionally, the distance distribution between BACE1 residue Thr72 and Ser325 was calculated for the four systems and provided as Supplementary material; all the systems, except for the APPwt, present a multi-modal distribution.

To analyze these paired distances meaningfully, the top 20 scoring residues in the APPwt system, with the highest contact probabilities, were ranked and compared against probabilities from the other systems; moreover, only APP residues 673, 671, and 672 were taken into account (respectively in Figure 52, Figure 53 and Figure 54).

Residue Ala673, which is mutated to Val in APPa2v and Thr in APPa2t, exhibits almost continuously a distance under a nanometer (Figure 52) from BACE1 residues 327-329 and 70-73, those groups representing respectively the region at the bottom of the binding pocket and the flap. For these residues, differences between the systems are not observed, apart from a less probable contact in APPsw and APPa2t. However, the heatmap showed an inferior probability for APP residue 673 and BACE1 69, 74 and 198, remarkably lower for APPa2t

substitution: being those residues located in a contiguous space of the pocket, this suggests a deviation from the preferred three-dimensional arrangement of wild type system.
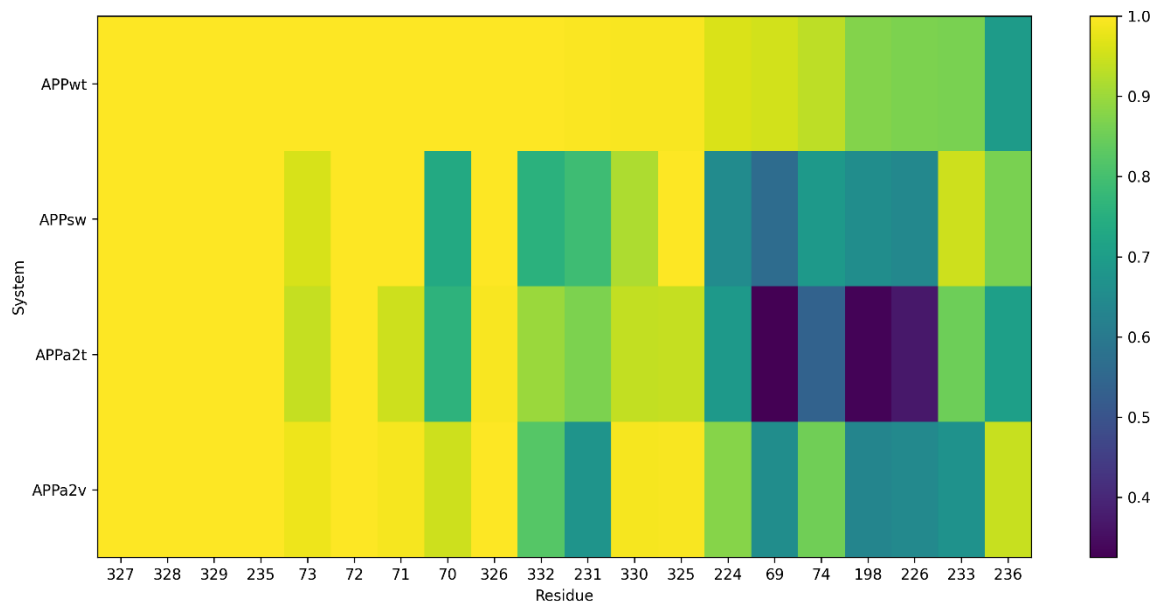

Figure 52 – Contact probability heatmap of APP residue 673 with 1 nm cutoff.

A similar picture (Figure 53) was obtained with distances from residue Asn671, mutated to Leu in APPsw. In this case, the detachment from the flap region is remarkably higher for APPsw, with generalized lower probabilities for BACE1 residues 68-76.
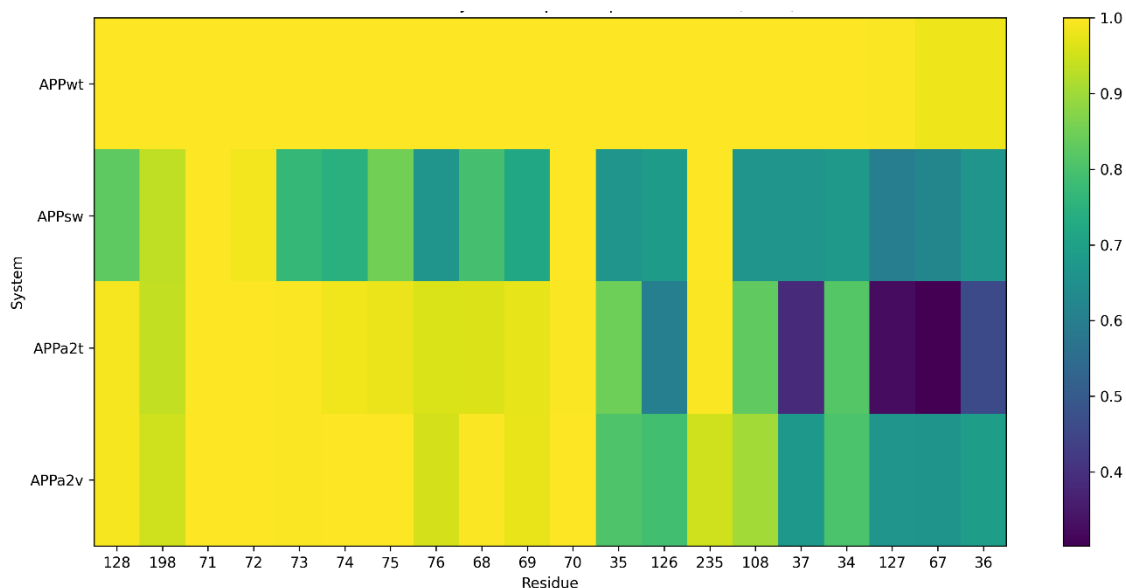

Figure 53 – Contact probability heatmap of APP residue 671 with 1 nm cutoff.

Considering APP residue Asp672, which is conserved for all the simulated systems, the top 20 scoring residues possess inferior contact probabilities compared to the other two heatmaps, with the interactions appearing to be limited to the flap region (Figure 54). Contact with BACE1 residue Val69 is remarkably lower for APPa2t, a finding that is observed also for APP residue 673 (Figure 52); an explanation could be the unfavorable interaction between the polar Thr673 (in comparison to the standard Ala673 of wild type APP) and the apolar Val69.
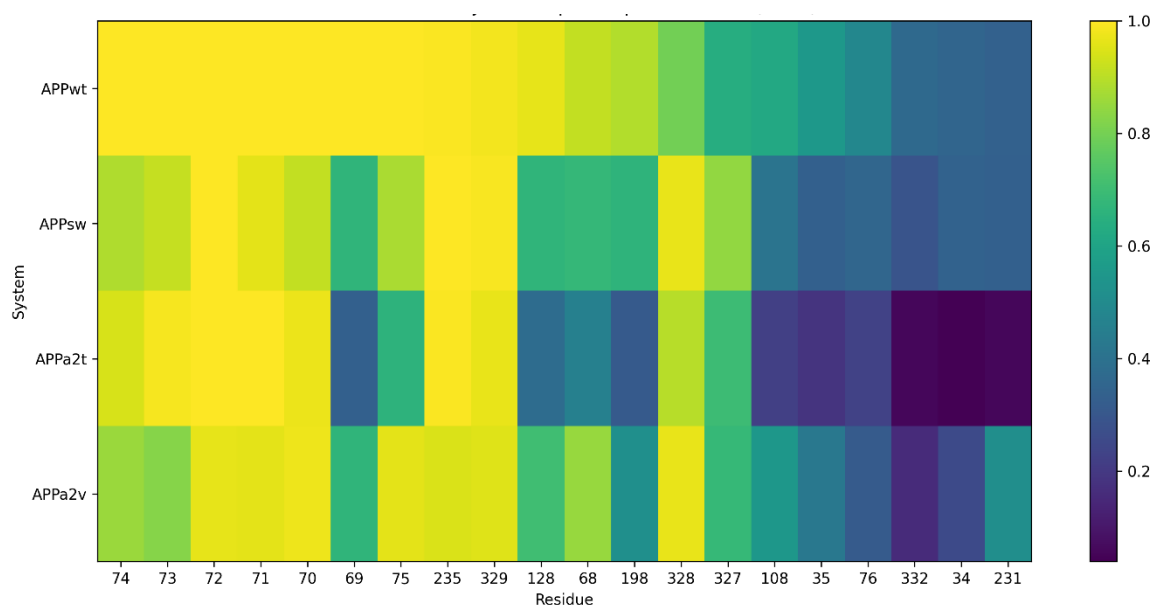

Figure 54 – Contact probability heatmap of APP residue 672 with 1 nm cutoff.

The analysis of distance probabilities under 0.6 nm (Van der Waals forces), considering only the top 5 APPwt scoring residues, shows a notably lower interaction probability among APP residue 672 and the flap region, especially for APPa2t system. Interestingly, interactions between residue 673 and BACE1 Ser328 are favored with both Thr and Val substitution, compared to the standard Ala.

The figures are provided as Supplementary material.

4.4.3 Binding energy calculation

The binding energy calculation results, using the MM-GBSA method, are reported in Figure 55: the low negative values for all the systems indicate that APP is a good BACE1 binder. In detail, the difference between APPa2t and APPa2v binding energy is remarkable high (delta is more than 40 kcal/mol), while intermediate values are reported for APPwt and APPsw. Standard deviations are elevated, particularly for APPsw system, as expected with MM-GBSA method [80].
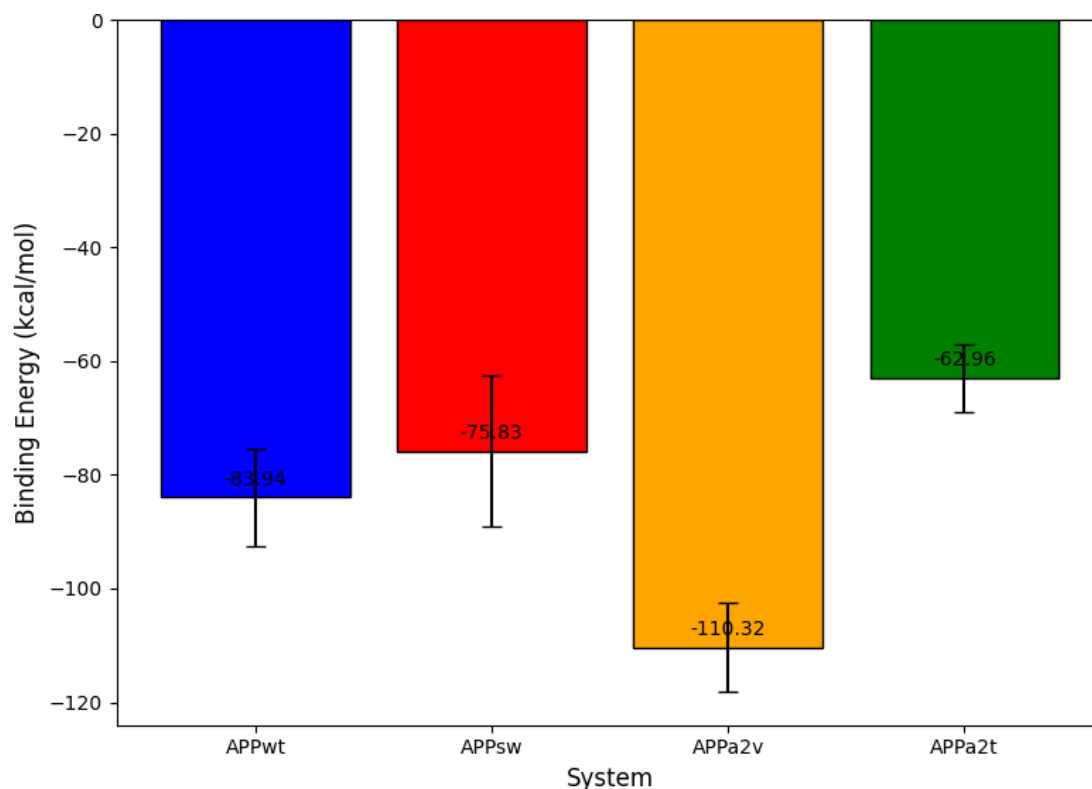

Figure 55 – Comparison of binding energies (mean and standard deviation) for the four systems calculated with MM-GBSA method.

Decomposition analysis was conducted to highlight per-residue contributions to the total energy (Supplementary Material) [93]. Flap residues, primarily Tyr71 and Thr72, exhibit negative values and rank among the top-scoring residues in all systems except APPwt. Residue 671 (Met, except for Leu in APPsw) also contributes significantly to the total energy score, with values around -5 kcal/mol. As shown in Table 1, despite high standard deviations, substitution with Val at residue 673 results in lower energy values compared to Thr mutation.

System	Residue	Mean	SD
APPa2v	VAL:673	-2,7	5,08
APPsw	ALA:673	-1,54	3,69
APPwt	ALA:673	-1,31	3,89
APPa2t	THR:673	-1,26	6,28

Table 1 – Decomposition analysis of APP residue 673, with mean and standard deviation energies expressed in kcal/mol, in ascending order.

Considering the Lys-to-Asn substitution at position 670 in APPsw (Table 2), a negative value is observed exclusively for APPsw, while positive values—indicating poorer affinity—are found in the other systems. This difference is particularly remarkable for the APPa2t system.

System	Residue	Mean	SD
APPsw	ASN:670	-0,15	6,40
APPa2v	LYS:670	0,47	10,68
APPwt	LYS:670	3,79	24,07
APPa2t	LYS:670	4,28	11,37

Table 2 – Mean and standard deviation energy contributions (expressed in kcal/mol), in ascending order, of APP residue 670.

However, it should be noted that these results are affected by high standard deviations, which are expected when using the MM-GBSA approach [94].

4.5 Discussion

The comparison between free BACE1 in solution and bound systems is consistent with previous reports [83], [85], [86], demonstrating gain of stability upon APP binding: widespread higher RMSF are probably the result of active site (occupying the majority of BACE1 volume) interactions with water molecules. Moreover, the flap oscillations between closed and open states, reported to be on the order of 10 ns [86], are another source of decreased stability.

In the present study, performing docking, the elucidation of dynamical APP entry is not possible; the simulations start from an already bound structure (wild type APP bound to BACE1), which is surely a rough approximation, neglecting possible differences in mutated APP fragments in their binding behavior.

However, APP-BACE1 stability is probably granted by electrostatic surface complementarity: the positive binding pocket favors the attraction of mainly negative APP (in particular by its two glutamate residues). Interestingly, both A673T and A673V are uncharged substitution, not perturbing the APP electrostatic surface; on the contrary, APPsw "looses" its standard positive Lys residue for Asn. The Swedish double mutation could therefore increase electrostatic attraction with a positive binding pocket, resulting in more rapid cleavage.

Additionally, a difference is observed in APP hydrophobic surface, with APPsw and APPa2t possessing more hydrophobic residues. The protective A673T is the only polar mutation of the three considered in this work. Given that the binding pocket is mainly hydrophilic, this finding is not consistent with the hypothesis of more tight interactions for the two pathogenic mutations.

Considering global effects of different APP species, fluctuations of flap residue, particularly Thr72, appear higher in the case of A673T, while A673V has the opposite effect. A more rigid flap could increase the catalytic activity of the enzyme, favoring interactions of APP with the active Asp residues, located at the bottom of the binding pocket [31]. Conversely, APPa2t system displays higher flap and loop-A fluctuations.

The short-range interactions (hydrogen bonding) network was not analyzed explicitly, only calculating hydrogen bond count across the trajectories: considering only the "central" APP residues (five before and after the scissile 671-672 bond), the small increase in count observed with A673T was coherent with the polar substitution, as well as a decrease for APPa2v system, expressing a more hydrophobic side chain.

Distance probability calculations, taking as reference the APPwt complex, revealed a widespread decrease in flap proximity of APP residues 671-673; moreover, the pattern is remarkably similar for the three mutated system. This could be interpreted as a limitation of the docking protocol, as the choice of mutating the structure from APPwt could result in sub-optimal stability. Nevertheless, a highly conserved protein, with rare mutations having usually a huge impact on functions such as cognition [29], should considered as the product of natural evolution; therefore, instability could arise when altering its amino acid sequence. Considering that this effect is consistent across the replicas, and the simulations were long enough to let the systems equilibrate to other configurations, docking procedure was regarded as consistent.

The distance analysis using a shorter cutoff (0.6 nm) similarly showed decreased interaction probabilities, with dramatic decrease in APP 671-673 and flap contacts. Once again, this could be explained with APP altered recognition by BACE1, if the former is mutated [85]. The protective effect cannot consequently be ascribed to inferior contact probabilities between BACE1 residues 67-77 and APP scissile region.

In the present work, the top scoring contacts were not calculated for the mutated systems, representing another limitation: the elucidation of interaction types can be performed by using Protein-Ligand-Interaction-Profiler (PLIP) tool, thus enabling a comprehensive evaluation of pattern changes.

Predicted affinities with MM-GBSA methods are in line with spontaneous complex formation, for all the systems simulated [80]. APPwt is a good binder for BACE1, with good agreement with physiological conditions. Interestingly, a remarkably higher affinity was predicted for A673V mutation, in line with experimental evidence of increased Aβ production caused by receptor affinity [44], [45].

In contrast, the less negative energy for APPa2t seems to be congruent with impaired APP-BACE1 affinity as the mechanistic explanation of protective effect [46], [51], [97].

In spite of those experiment correlated differences, APPsw has a binding energy comparable with wild type complex, contradicting the literature consensus of accelerated Aβ generation [29], [39], [42]. A possible explanation could be the following: being the Swedish mutations a double substitution altering the scissile bond (Leu671-Asp672 instead of Met671-Asp672), the effect could not be adequately modeled by molecular mechanics, neglecting bond breaking. A study conducted with a hybrid MM/QM approach calculates energy barriers in bond cleaving comparing wild type and Swedish mutations [43]. In fact, results obtained with MM-GBSA approach are of intermediate precision and strategies such as Alchemical Perturbation or

explicit quantum calculations can offer a better understanding of hydrolytic processing [82], [98].

The decomposition analysis, albeit showing altered contribution of residues 673 and 670 to the total binding energy, in line with protective-pathogenic hypothesized behavior, was affected by high uncertainty. As an expected weak point of MM-GBSA method, per-residue analysis could be improved by running several independent simulations, instead of considering windows (believed as mutually uncorrelated) of a longer simulation [94]. Otherwise, more precise calculation, considering only residues undergoing cleavage, might represent an improvement in results.

In conclusion, molecular mechanics is a useful strategy to rationalize experiment results, given its atomistic resolution and the possibility to calculate time averages of interesting variables. The present work exploited molecular docking to model APP-BACE1 interactions, showing differences among mutations involved in AD pathogenesis/neuroprotection over the course of 600 ns simulations, obtaining one and a half microsecond of analysis per system. While in-silico findings must be replicated in-vitro and in-vivo, the present work could be a useful framework for studying the effect of mutations on neurodegenerative disease risk/protection.

5. Conclusions and future perspectives

This work provided a summary on the clinical and biological bases of AD, a challenging condition without effective treatments, whose biochemistry description is yet to be completely clear. In the following, hypotheses of its pathogenesis were described, from the historical cholinergic neuronal loss to the current consensus: nowadays, evidence points towards a multi-decade long divergence from healthy brain, with progressive unfolding and aggregation of aberrant proteinaceous species, mostly Aβ42 and hyperphosphorylated tau. The observation of biomarker changes offers the opportunity to develop diagnostic tools necessary to tailor clinical trials, aimed at reversing the course of AD, before cognitive deficits appear.

Afterwards, the description of APP functions and structure was provided, clarifying the various catabolic pathways, considered to be of central importance on AD progression. Interestingly, mutations in the APP are one of the causative factors of EOAD [39], with undistinguishable course from the more common late onset type, apart from an earlier beginning; they are located inside the Aβ sequence and/or near the cleavage sites (BACE1 and gamma secretase). The discovery of A673T ("Icelandic") mutations changed this knowledge landscape, extending the effect of APP mutations beyond AD pathology: its association with lower dementia risks is regarded as a possible involvement of APP in higher cerebral functions [56].

The simulation of biological systems is a complex and vast *repertoire* of techniques, for example used to predict the best binding poses of a drug to a receptor or visualize the time evolution of interacting macromolecules. Molecular mechanics theoretical bases were presented, as well as molecular dynamics simulation set-up and methods to assess binding free energy (MM-GBSA).

The application of those computational means on different APP-BACE1, expressing the wild type, two pathogenic and one protective mutation showed a theoretical link among binding affinity, flap interactions and experimental observation of altered Aβ production.

Obviously, more investigations are needed to elucidate the mechanistic effects of APP mutations; for example, a precise characterization of bond type interactions is mandatory to increase the validity of those results. At the same time, more refined methods to assess cleavage energies (QM/MM) could make the picture clearer.

Another hypothesis to be tested, using computational method, is the different behavior of mutated Aβ oligomers. [53]

Interestingly, while the inhibition of BACE1 or gamma secretase activity was plagued by serious side effects and no drug reached FDA approval, the anti-cancer agent Gleevec and some other compounds show reduction in Aβ production without interfering with BACE1/gamma cleavage of other substrate, in a similar way to A673T mutation (Supplementary material) [99], [100]. To my knowledge, a characterization of the effect of those chemicals using molecular mechanics simulation is still lacking.

6. Acknowledgements

7. References

[1] F. Assal, «History of Dementia», in *Frontiers of Neurology and Neuroscience*, vol. 44, J. Bogousslavsky, F. Boller, e M. Iwata, A c. di, S. Karger AG, 2019, pp. 118–126. doi: 10.1159/000494959.

[2] A. Gustavsson *et al.*, «Global estimates on the number of persons across the Alzheimer's disease continuum», *Alzheimers Dement.*, vol. 19, fasc. 2, pp. 658–670, feb. 2023, doi: 10.1002/alz.12694.

[3] «2020 Alzheimer's disease facts and figures», *Alzheimers Dement.*, vol. 16, fasc. 3, pp. 391–460, mar. 2020, doi: 10.1002/alz.12068.

[4] A. Peterson *et al.*, «Is there a difference between terminal lucidity and paradoxical lucidity?», *Alzheimers Dement.*, vol. 18, fasc. 3, pp. 540–541, mar. 2022, doi: 10.1002/alz.12579.

[5] D. Blazer, «Neurocognitive Disorders in DSM-5», *Am. J. Psychiatry*, vol. 170, fasc. 6, pp. 585–587, giu. 2013, doi: 10.1176/appi.ajp.2013.13020179.

[6] V. Mantzavinos e A. Alexiou, «Biomarkers for Alzheimer's Disease Diagnosis», *Curr. Alzheimer Res.*, vol. 14, fasc. 11, ott. 2017, doi: 10.2174/1567205014666170203125942.

[7] K. M. Fiest *et al.*, «The Prevalence and Incidence of Dementia Due to Alzheimer's Disease: a Systematic Review and Meta-Analysis», *Can. J. Neurol. Sci. J. Can. Sci. Neurol.*, vol. 43, fasc. S1, pp. S51–S82, apr. 2016, doi: 10.1017/cjn.2016.36.

[8] G. W. Leeson, «The Growth, Ageing and Urbanisation of our World», *J. Popul. Ageing*, vol. 11, fasc. 2, pp. 107–115, giu. 2018, doi: 10.1007/s12062-018-9225-7.

[9] L. X. Tay, S. C. Ong, L. J. Tay, T. Ng, e T. Parumasivam, «Economic Burden of Alzheimer's Disease: A Systematic Review», *Value Health Reg. Issues*, vol. 40, pp. 1–12, mar. 2024, doi: 10.1016/j.vhri.2023.09.008.

[10] N. Zhao *et al.*, «Alzheimer's Risk Factors Age, APOE Genotype, and Sex Drive Distinct Molecular Pathways», *Neuron*, vol. 106, fasc. 5, pp. 727-742.e6, giu. 2020, doi: 10.1016/j.neuron.2020.02.034.

[11] R. Sims, M. Hill, e J. Williams, «The multiplex model of the genetics of Alzheimer's disease», *Nat. Neurosci.*, vol. 23, fasc. 3, pp. 311–322, mar. 2020, doi: 10.1038/s41593-020-0599-5.

[12] M. A. O'Neal, «Women and the risk of Alzheimer's disease», *Front. Glob. Womens Health*, vol. 4, p. 1324522, gen. 2024, doi: 10.3389/fgwh.2023.1324522.

[13] X.-X. Zhang, Y. Tian, Z.-T. Wang, Y.-H. Ma, L. Tan, e J.-T. Yu, «The Epidemiology of Alzheimer's Disease Modifiable Risk Factors and Prevention», *J. Prev. Alzheimers Dis.*, pp. 1–9, 2021, doi: 10.14283/jpad.2021.15.

[14] A. Graham, G. Livingston, L. Purnell, e J. Huntley, «Mild Traumatic Brain Injuries and Future Risk of Developing Alzheimer's Disease: Systematic Review and Meta-Analysis», *J. Alzheimers Dis.*, vol. 87, fasc. 3, pp. 969–979, mag. 2022, doi: 10.3233/JAD-220069.

[15] G. Livingston *et al.*, «Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission», *The Lancet*, vol. 404, fasc. 10452, pp. 572–628, ago. 2024, doi: 10.1016/S0140-6736(24)01296-0.

[16] X. Du, X. Wang, e M. Geng, «Alzheimer's disease hypothesis and related therapies», *Transl. Neurodegener.*, vol. 7, fasc. 1, p. 2, dic. 2018, doi: 10.1186/s40035-018-0107-y.

[17] D. M. Wilson, M. R. Cookson, L. Van Den Bosch, H. Zetterberg, D. M. Holtzman, e I. Dewachter, «Hallmarks of neurodegenerative diseases», *Cell*, vol. 186, fasc. 4, pp. 693–714, feb. 2023, doi: 10.1016/j.cell.2022.12.032.

[18] Z.-R. Chen, J.-B. Huang, S.-L. Yang, e F.-F. Hong, «Role of Cholinergic Signaling in Alzheimer's Disease», *Molecules*, vol. 27, fasc. 6, p. 1816, mar. 2022, doi: 10.3390/molecules27061816.

[19] E. Karran e B. De Strooper, «The amyloid hypothesis in Alzheimer disease: new insights from new therapeutics», *Nat. Rev. Drug Discov.*, vol. 21, fasc. 4, pp. 306–318, apr. 2022, doi: 10.1038/s41573-022-00391-w.

[20] L. C. Walker e M. Jucker, «The prion principle and Alzheimer's disease», *Science*, vol. 385, fasc. 6715, pp. 1278–1279, set. 2024, doi: 10.1126/science.adq5252.

[21] T. Tomiyama e H. Shimada, «APP Osaka Mutation in Familial Alzheimer's Disease—Its Discovery, Phenotypes, and Mechanism of Recessive Inheritance», *Int. J. Mol. Sci.*, vol. 21, fasc. 4, p. 1413, feb. 2020, doi: 10.3390/ijms21041413.

[22] A. Rahman *et al.*, «Aducanumab for the treatment of Alzheimer's disease: a systematic review», *Psychogeriatrics*, vol. 23, fasc. 3, pp. 512–522, mag. 2023, doi: 10.1111/psyg.12944.

[23] F. P. Chong, K. Y. Ng, R. Y. Koh, e S. M. Chye, «Tau Proteins and Tauopathies in Alzheimer's Disease», *Cell. Mol. Neurobiol.*, vol. 38, fasc. 5, pp. 965–980, lug. 2018, doi: 10.1007/s10571-017-0574-1.

[24] A. F. T. Arnsten, D. Datta, K. Del Tredici, e H. Braak, «Hypothesis: Tau pathology is an initiating factor in sporadic Alzheimer's disease», *Alzheimers Dement.*, vol. 17, fasc. 1, pp. 115–124, gen. 2021, doi: 10.1002/alz.12192.

[25] A. C. Macedo *et al.*, «The Use of Tau PET to Stage Alzheimer Disease According to the Braak Staging Framework», *J. Nucl. Med.*, vol. 64, fasc. 8, pp. 1171–1178, ago. 2023, doi: 10.2967/jnumed.122.265200.

[26] C. Hadjichrysanthou *et al.*, «The dynamics of biomarkers across the clinical spectrum of Alzheimer's disease», *Alzheimers Res. Ther.*, vol. 12, fasc. 1, p. 74, dic. 2020, doi: 10.1186/s13195-020-00636-z.

[27] J. Jia *et al.*, «Biomarker Changes during 20 Years Preceding Alzheimer's Disease», *N. Engl. J. Med.*, vol. 390, fasc. 8, pp. 712–722, feb. 2024, doi: 10.1056/NEJMoa2310168.

[28] Y. Cho, H.-G. Bae, E. Okun, T. V. Arumugam, e D.-G. Jo, «Physiology and pharmacology of amyloid precursor protein», *Pharmacol. Ther.*, vol. 235, p. 108122, lug. 2022, doi: 10.1016/j.pharmthera.2022.108122.

[29] A. Delport e R. Hewer, «The amyloid precursor protein: a converging point in Alzheimer's disease», *Mol. Neurobiol.*, vol. 59, fasc. 7, pp. 4501–4516, lug. 2022, doi: 10.1007/s12035-022-02863-x.

[30] R. J. O'Brien e P. C. Wong, «Amyloid Precursor Protein Processing and Alzheimer's Disease», *Annu. Rev. Neurosci.*, vol. 34, fasc. 1, pp. 185–204, lug. 2011, doi: 10.1146/annurev-neuro-061010-113613.

[31] H. Hampel *et al.*, «The β-Secretase BACE1 in Alzheimer's Disease», *Biol. Psychiatry*, vol. 89, fasc. 8, pp. 745–756, apr. 2021, doi: 10.1016/j.biopsych.2020.02.001.

[32] H. M. Kumalo, S. Bhakat, e M. E. Soliman, «Investigation of flap flexibility of β-secretase using molecular dynamic simulations», *J. Biomol. Struct. Dyn.*, vol. 34, fasc. 5, pp. 1008–1019, mag. 2016, doi: 10.1080/07391102.2015.1064831.

[33] X. Zhang e W. Song, «The role of APP and BACE1 trafficking in APP processing and amyloid-β generation», *Alzheimers Res. Ther.*, vol. 5, fasc. 5, p. 46, 2013, doi: 10.1186/alzrt211.

[34] H. A. Taylor, L. Przemylska, E. M. Clavane, e P. J. Meakin, «BACE1: More than just a β-secretase», *Obes. Rev.*, vol. 23, fasc. 7, p. e13430, lug. 2022, doi: 10.1111/obr.13430.

[35] N. M. Moussa-Pacha, S. M. Abdin, H. A. Omar, H. Alniss, e T. H. Al-Tel, «BACE1 inhibitors: Current status and future directions in treating Alzheimer's disease», *Med. Res. Rev.*, vol. 40, fasc. 1, pp. 339–384, gen. 2020, doi: 10.1002/med.21622.

[36] V. Vingtdeux e P. Marambaud, «Identification and biology of α-secretase», *J. Neurochem.*, vol. 120, fasc. s1, pp. 34–45, gen. 2012, doi: 10.1111/j.1471-4159.2011.07477.x.

[37] J.-Y. Hur, «γ-Secretase in Alzheimer's disease», *Exp. Mol. Med.*, vol. 54, fasc. 4, pp. 433–446, apr. 2022, doi: 10.1038/s12276-022-00754-8.

[38] L. Liu, L. Ding, M. Rovere, M. S. Wolfe, e D. J. Selkoe, «A cellular complex of BACE1 and γ-secretase sequentially generates Aβ from its full-length precursor», *J. Cell Biol.*, vol. 218, fasc. 2, pp. 644–663, feb. 2019, doi: 10.1083/jcb.201806205.

[39] D. W. Sirkis, L. W. Bonham, T. P. Johnson, R. La Joie, e J. S. Yokoyama, «Dissecting the clinical heterogeneity of early-onset Alzheimer's disease», *Mol. Psychiatry*, vol. 27, fasc. 6, pp. 2674–2688, giu. 2022, doi: 10.1038/s41380-022-01531-9.

[40] J. Kinoshita e T. Clark, «Alzforum», in *Neuroinformatics*, vol. 401, in Methods in Molecular Biology™, vol. 401. , Totowa, NJ: Humana Press, 2007, pp. 365–381. doi: 10.1007/978-1-59745-520-6_19.

[41] J. Fortea, S. H. Zaman, S. Hartley, M. S. Rafii, E. Head, e M. Carmona-Iragui, «Alzheimer's disease associated with Down syndrome: a genetic form of dementia», *Lancet Neurol.*, vol. 20, fasc. 11, pp. 930–942, nov. 2021, doi: 10.1016/S1474-4422(21)00245-3.

[42] C. Haass *et al.*, «The Swedish mutation causes early-onset Alzheimer's disease by β-secretase cleavage within the secretory pathway», *Nat. Med.*, vol. 1, fasc. 12, pp. 1291–1296, dic. 1995, doi: 10.1038/nm1295-1291.

[43] A. Barman e R. Prabhakar, «Elucidating the catalytic mechanism of β-secretase (BACE1): A quantum mechanics/molecular mechanics (QM/MM) approach», *J. Mol. Graph. Model.*, vol. 40, pp. 1–9, mar. 2013, doi: 10.1016/j.jmgm.2012.12.010.

[44] G. Di Fede *et al.*, «Good gene, bad gene: New APP variant may be both», *Prog. Neurobiol.*, vol. 99, fasc. 3, pp. 281–292, dic. 2012, doi: 10.1016/j.pneurobio.2012.06.004.

[45] G. Giaccone *et al.*, «Neuropathology of the recessive A673V APP mutation: Alzheimer disease with distinctive features», *Acta Neuropathol. (Berl.)*, vol. 120, fasc. 6, pp. 803–812, dic. 2010, doi: 10.1007/s00401-010-0747-1.

[46] T. Jonsson *et al.*, «A mutation in APP protects against Alzheimer's disease and age-related cognitive decline», *Nature*, vol. 488, fasc. 7409, pp. 96–99, ago. 2012, doi: 10.1038/nature11283.

[47] Q. Xia *et al.*, «The Protective A673T Mutation of Amyloid Precursor Protein (APP) in Alzheimer's Disease», *Mol. Neurobiol.*, vol. 58, fasc. 8, pp. 4038–4050, ago. 2021, doi: 10.1007/s12035-021-02385-y.

[48] L.-S. Wang *et al.*, «Rarity of the Alzheimer disease-protective APP A673T variant in the United States», *JAMA Neurol.*, vol. 72, fasc. 2, pp. 209–216, feb. 2015, doi: 10.1001/jamaneurol.2014.2157.

[49] Y.-W. Liu *et al.*, «Absence of A673T variant in APP gene indicates an alternative protective mechanism contributing to longevity in Chinese individuals», *Neurobiol. Aging*, vol. 35, fasc. 4, p. 935.e11–12, apr. 2014, doi: 10.1016/j.neurobiolaging.2013.09.023.

[50] J. Mengel-From, B. Jeune, T. Pentti, M. McGue, K. Christensen, e L. Christiansen, «The APP A673T frequency differs between Nordic countries», *Neurobiol. Aging*, vol. 36, fasc. 10, p. 2909.e1–4, ott. 2015, doi: 10.1016/j.neurobiolaging.2015.07.011.

[51] J. A. Maloney *et al.*, «Molecular mechanisms of Alzheimer disease protection by the A673T allele of amyloid precursor protein», *J. Biol. Chem.*, vol. 289, fasc. 45, pp. 30990–31000, nov. 2014, doi: 10.1074/jbc.M114.589069.

[52] A. Kimura, S. Hata, e T. Suzuki, «Alternative Selection of β-Site APP-Cleaving Enzyme 1 (BACE1) Cleavage Sites in Amyloid β-Protein Precursor (APP) Harboring Protective and Pathogenic Mutations within the Aβ Sequence», *J. Biol. Chem.*, vol. 291, fasc. 46, pp. 24041–24053, nov. 2016, doi: 10.1074/jbc.M116.744722.

[53] C. S. Limegrover *et al.*, «Alzheimer's protection effect of A673T mutation may be driven by lower Aβ oligomer binding affinity», *J. Neurochem.*, vol. 157, fasc. 4, pp. 1316–1330, mag. 2021, doi: 10.1111/jnc.15212.

[54] R. Wittrahm *et al.*, «Protective Alzheimer's disease-associated APP A673T variant predominantly decreases sAPPβ levels in cerebrospinal fluid and 2D/3D cell culture models», *Neurobiol. Dis.*, vol. 182, p. 106140, giu. 2023, doi: 10.1016/j.nbd.2023.106140.

[55] T. Rolova *et al.*, «Generation of a human induced pluripotent stem cell line (UEFi003-A) carrying heterozygous A673T variant in amyloid precursor protein associated with a reduced risk of Alzheimer's disease», *Stem Cell Res.*, vol. 48, p. 101968, ott. 2020, doi: 10.1016/j.scr.2020.101968.

[56] S. Shimohama *et al.*, «The Icelandic Mutation (APP-A673T) Is Protective against Amyloid Pathology In Vivo», *J. Neurosci.*, vol. 44, fasc. 47, p. e0223242024, nov. 2024, doi: 10.1523/JNEUROSCI.0223-24.2024.

[57] A. Guyon, J. Rousseau, F.-G. Bégin, T. Bertin, G. Lamothe, e J. P. Tremblay, «Base editing strategy for insertion of the A673T mutation in the APP gene to prevent the development of AD in vitro», *Mol. Ther. Nucleic Acids*, vol. 24, pp. 253–263, giu. 2021, doi: 10.1016/j.omtn.2021.02.032.

[58] G. Tremblay, J. Rousseau, C. H. Mbakam, e J. P. Tremblay, «Insertion of the Icelandic Mutation (A673T) by Prime Editing: A Potential Preventive Treatment for Familial and Sporadic Alzheimer's Disease», *CRISPR J.*, vol. 5, fasc. 1, pp. 109–122, feb. 2022, doi: 10.1089/crispr.2021.0085.

[59] J. L. Heidebrink e H. L. Paulson, «Lessons Learned from Approval of Aducanumab for Alzheimer's Disease», *Annu. Rev. Med.*, vol. 75, fasc. 1, pp. 99–111, gen. 2024, doi: 10.1146/annurev-med-051022-043645.

[60] M. S. Wolfe, «γ-Secretase: once and future drug target for Alzheimer's disease», *Expert Opin. Drug Discov.*, vol. 19, fasc. 1, pp. 5–8, gen. 2024, doi: 10.1080/17460441.2023.2277350.

[61] Y. Pathak, I. Camps, A. Mishra, e V. Tripathi, «Targeting notch signaling pathway in breast cancer stem cells through drug repurposing approach», *Mol. Divers.*, vol. 27, fasc. 6, pp. 2431–2440, dic. 2023, doi: 10.1007/s11030-022-10561-y.

[62] E. McDade *et al.*, «The case for low-level BACE1 inhibition for the prevention of Alzheimer disease», *Nat. Rev. Neurol.*, vol. 17, fasc. 11, pp. 703–714, nov. 2021, doi: 10.1038/s41582-021-00545-1.

[63] A. Miranda, E. Montiel, H. Ulrich, e C. Paz, «Selective Secretase Targeting for Alzheimer's Disease Therapy», *J. Alzheimers Dis.*, vol. 81, fasc. 1, pp. 1–17, mag. 2021, doi: 10.3233/JAD-201027.

[64] L. V. Schaffer e T. Ideker, «Mapping the multiscale structure of biological systems», *Cell Syst.*, vol. 12, fasc. 6, pp. 622–635, giu. 2021, doi: 10.1016/j.cels.2021.05.012.

[65] R. Nussinov, C.-J. Tsai, A. Shehu, e H. Jang, «Computational Structural Biology: Successes, Future Directions, and Challenges», *Molecules*, vol. 24, fasc. 3, p. 637, feb. 2019, doi: 10.3390/molecules24030637.

[66] P. Cramer, «AlphaFold2 and the future of structural biology», *Nat. Struct. Mol. Biol.*, vol. 28, fasc. 9, pp. 704–705, set. 2021, doi: 10.1038/s41594-021-00650-1.

[67] D. Desai, S. V. Kantliwala, J. Vybhavi, R. Ravi, H. Patel, e J. Patel, «Review of AlphaFold 3: Transformative Advances in Drug Design and Therapeutics», *Cureus*, lug. 2024, doi: 10.7759/cureus.63646.

[68] A. R. Leach, *Molecular Modelling: Principles and Applications*. Pearson Education, 2001.

[69] H. G. Petersen, «Accuracy and efficiency of the particle mesh Ewald method», *J. Chem. Phys.*, vol. 103, fasc. 9, pp. 3668–3679, set. 1995, doi: 10.1063/1.470043.

[70] I. T. Christensen e F. S. Jørgensen, «Molecular Mechanics Calculations of Proteins. Comparison of Different Energy Minimization Strategies», *J. Biomol. Struct. Dyn.*, vol. 15, fasc. 3, pp. 473–488, dic. 1997, doi: 10.1080/07391102.1997.10508959.

[71] N. S. Pagadala, K. Syed, e J. Tuszynski, «Software for molecular docking: a review», *Biophys. Rev.*, vol. 9, fasc. 2, pp. 91–102, apr. 2017, doi: 10.1007/s12551-016-0247-1.

[72] D. S. Goodsell, M. F. Sanner, A. J. Olson, e S. Forli, «The AUTODOCK suite at 30», *Protein Sci.*, vol. 30, fasc. 1, pp. 31–43, gen. 2021, doi: 10.1002/pro.3934.

[73] O. Trott e A. J. Olson, «AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading», *J. Comput. Chem.*, vol. 31, fasc. 2, pp. 455–461, gen. 2010, doi: 10.1002/jcc.21334.

[74] M. Kurcinski, A. Badaczewska-Dawid, M. Kolinski, A. Kolinski, e S. Kmiecik, «Flexible docking of peptides to proteins using CABS-dock», *Protein Sci.*, vol. 29, fasc. 1, pp. 211–222, gen. 2020, doi: 10.1002/pro.3771.

[75] S. R. Comeau, D. W. Gatchell, S. Vajda, e C. J. Camacho, «ClusPro: a fully automated algorithm for protein-protein docking», *Nucleic Acids Res.*, vol. 32, fasc. Web Server, pp. W96–W99, lug. 2004, doi: 10.1093/nar/gkh354.

[76] Y. I. Yang, Q. Shao, J. Zhang, L. Yang, e Y. Q. Gao, «Enhanced sampling in molecular dynamics», *J. Chem. Phys.*, vol. 151, fasc. 7, p. 070902, ago. 2019, doi: 10.1063/1.5109531.

[77] J. S. Rowlinson *, «The Maxwell–Boltzmann distribution», *Mol. Phys.*, vol. 103, fasc. 21–23, pp. 2821–2828, nov. 2005, doi: 10.1080/002068970500044749.

[78] A. V. Sadybekov e V. Katritch, «Computational approaches streamlining drug discovery», *Nature*, vol. 616, fasc. 7958, pp. 673–685, apr. 2023, doi: 10.1038/s41586-023-05905-z.

[79] A. De Ruiter e C. Oostenbrink, «Advances in the calculation of binding free energies», *Curr. Opin. Struct. Biol.*, vol. 61, pp. 207–212, apr. 2020, doi: 10.1016/j.sbi.2020.01.016.

[80] S. Genheden e U. Ryde, «The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities», *Expert Opin. Drug Discov.*, vol. 10, fasc. 5, pp. 449–461, mag. 2015, doi: 10.1517/17460441.2015.1032936.

[81] C. J. Cramer e D. G. Truhlar, «Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics», *Chem. Rev.*, vol. 99, fasc. 8, pp. 2161–2200, ago. 1999, doi: 10.1021/cr960149m.

[82] E. Wang *et al.*, «End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design», *Chem. Rev.*, vol. 119, fasc. 16, pp. 9478–9508, ago. 2019, doi: 10.1021/acs.chemrev.9b00055.

[83] S. Patel, L. Vuillard, A. Cleasby, C. W. Murray, e J. Yon, «Apo and Inhibitor Complex Structures of BACE (β-secretase)», *J. Mol. Biol.*, vol. 343, fasc. 2, pp. 407–416, ott. 2004, doi: 10.1016/j.jmb.2004.08.018.

[84] D. A. Sabbah e H. A. Zhong, «Modeling the protonation states of β-secretase binding pocket by molecular dynamics simulations and docking studies», *J. Mol. Graph. Model.*, vol. 68, pp. 206–215, lug. 2016, doi: 10.1016/j.jmgm.2016.07.005.

[85] S. Chakraborty e S. Basu, «Structural insight into the mechanism of amyloid precursor protein recognition by β-secretase 1: A molecular dynamics study», *Biophys. Chem.*, vol. 202, pp. 1–12, lug. 2015, doi: 10.1016/j.bpc.2015.03.006.

[86] A. Barman, S. Schürer, e R. Prabhakar, «Computational Modeling of Substrate Specificity and Catalysis of the β-Secretase (BACE1) Enzyme», *Biochemistry*, vol. 50, fasc. 20, pp. 4337–4349, mag. 2011, doi: 10.1021/bi200081h.

[87] T. K. Karamanos, A. P. Kalverda, G. S. Thompson, e S. E. Radford, «Mechanisms of amyloid formation revealed by solution NMR», *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 88–89, pp. 86–104, ago. 2015, doi: 10.1016/j.pnmrs.2015.05.002.

[88] J. Talafous, K. J. Marcinowski, G. Klopman, e M. G. Zagorski, «Solution Structure of Residues 1-28 of the Amyloid .beta.-Peptide», *Biochemistry*, vol. 33, fasc. 25, pp. 7788–7796, giu. 1994, doi: 10.1021/bi00191a006.

[89] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, e H. J. C. Berendsen, «GROMACS: Fast, flexible, and free», *J. Comput. Chem.*, vol. 26, fasc. 16, pp. 1701–1718, dic. 2005, doi: 10.1002/jcc.20291.

[90] C. Tian *et al.*, «ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution», *J. Chem. Theory Comput.*, vol. 16, fasc. 1, pp. 528–552, gen. 2020, doi: 10.1021/acs.jctc.9b00591.

[91] T. Darden, D. York, e L. Pedersen, «Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems», *J. Chem. Phys.*, vol. 98, fasc. 12, pp. 10089–10092, giu. 1993, doi: 10.1063/1.464397.

[92] E. Bisong, «Matplotlib and Seaborn», in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, Berkeley, CA: Apress, 2019, pp. 151–165. doi: 10.1007/978-1-4842-4470-8_12.

[93] M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente, e E. Moreno, «gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS», *J. Chem. Theory Comput.*, vol. 17, fasc. 10, pp. 6281–6291, ott. 2021, doi: 10.1021/acs.jctc.1c00645.

[94] S. Genheden e U. Ryde, «How to obtain statistically converged MM/GBSA results», *J. Comput. Chem.*, vol. 31, fasc. 4, pp. 837–846, mar. 2010, doi: 10.1002/jcc.21366.

[95] W. Humphrey, A. Dalke, e K. Schulten, «VMD: Visual molecular dynamics», *J. Mol. Graph.*, vol. 14, fasc. 1, pp. 33–38, feb. 1996, doi: 10.1016/0263-7855(96)00018-5.

[96] E. C. Meng *et al.*, «UCSF CHIMERAX : Tools for structure building and analysis», *Protein Sci.*, vol. 32, fasc. 11, p. e4792, nov. 2023, doi: 10.1002/pro.4792.

[97] H. Martiskainen *et al.*, «Decreased plasma β-amyloid in the Alzheimer's disease APP A673T variant carriers», *Ann. Neurol.*, vol. 82, fasc. 1, pp. 128–132, lug. 2017, doi: 10.1002/ana.24969.

[98] C. E. Tzeliou, M. A. Mermigki, e D. Tzeli, «Review on the QM/MM Methodologies and Their Application to Metalloproteins», *Molecules*, vol. 27, fasc. 9, p. 2660, apr. 2022, doi: 10.3390/molecules27092660.

[99] W. J. Netzer *et al.*, «Gleevec inhibits β-amyloid production but not Notch cleavage», *Proc. Natl. Acad. Sci.*, vol. 100, fasc. 21, pp. 12444–12449, ott. 2003, doi: 10.1073/pnas.1534745100.

[100] W. J. Netzer, K. Bettayeb, S. C. Sinha, M. Flajolet, P. Greengard, e V. Bustos, «Gleevec shifts APP processing from a β-cleavage to a nonamyloidogenic cleavage», *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, fasc. 6, pp. 1389–1394, feb. 2017, doi: 10.1073/pnas.1620963114.

8. Supplementary materials
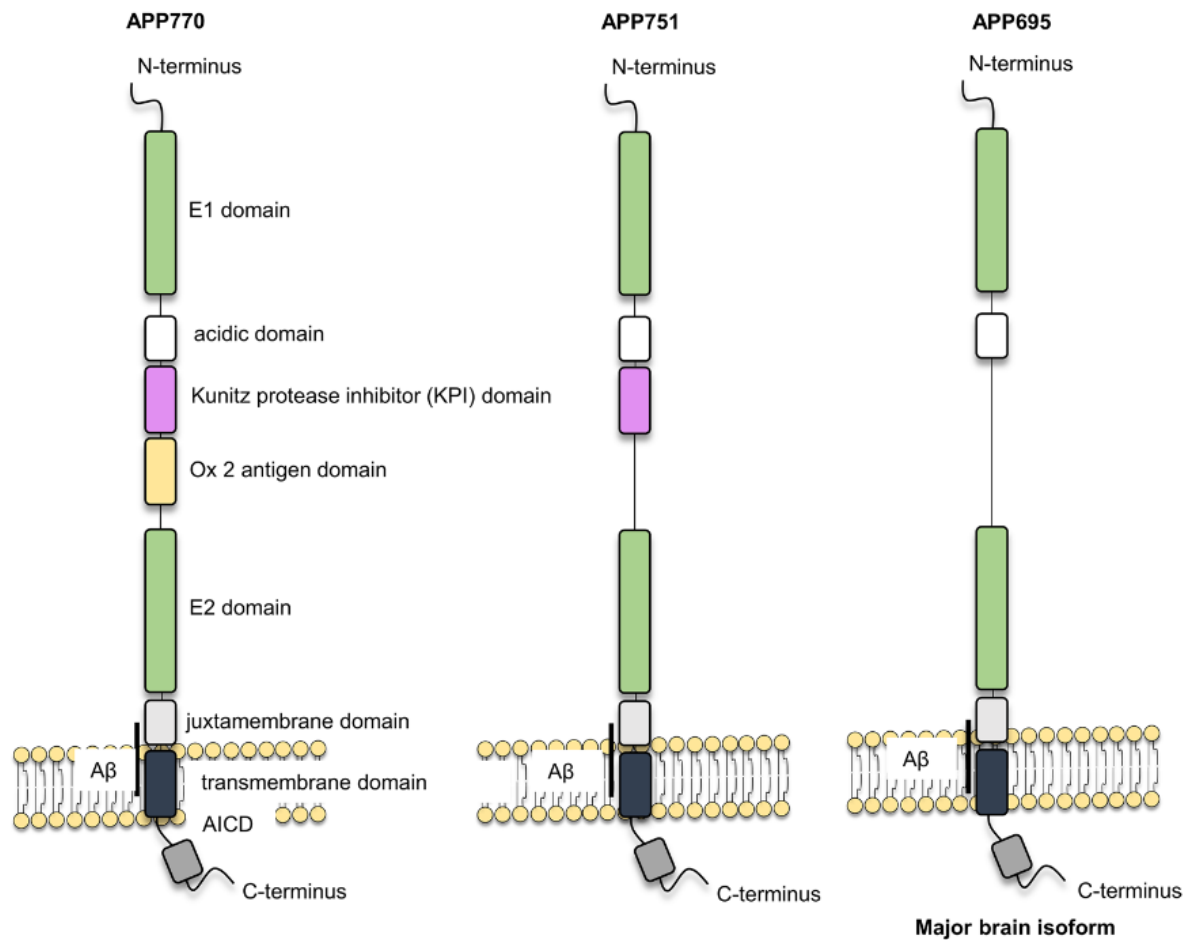
Figure S1 – Schematical representation of the common human APP isotypes: APP770 is recognized as the standard sequence, APP751 lacks the Ox2 antigen domain and APP695 (the most expressed isoform in brain) lacks both Ox 2 antigen domain and KPI domain [https://doi.org/10.1007/s12035-022-02863-x]
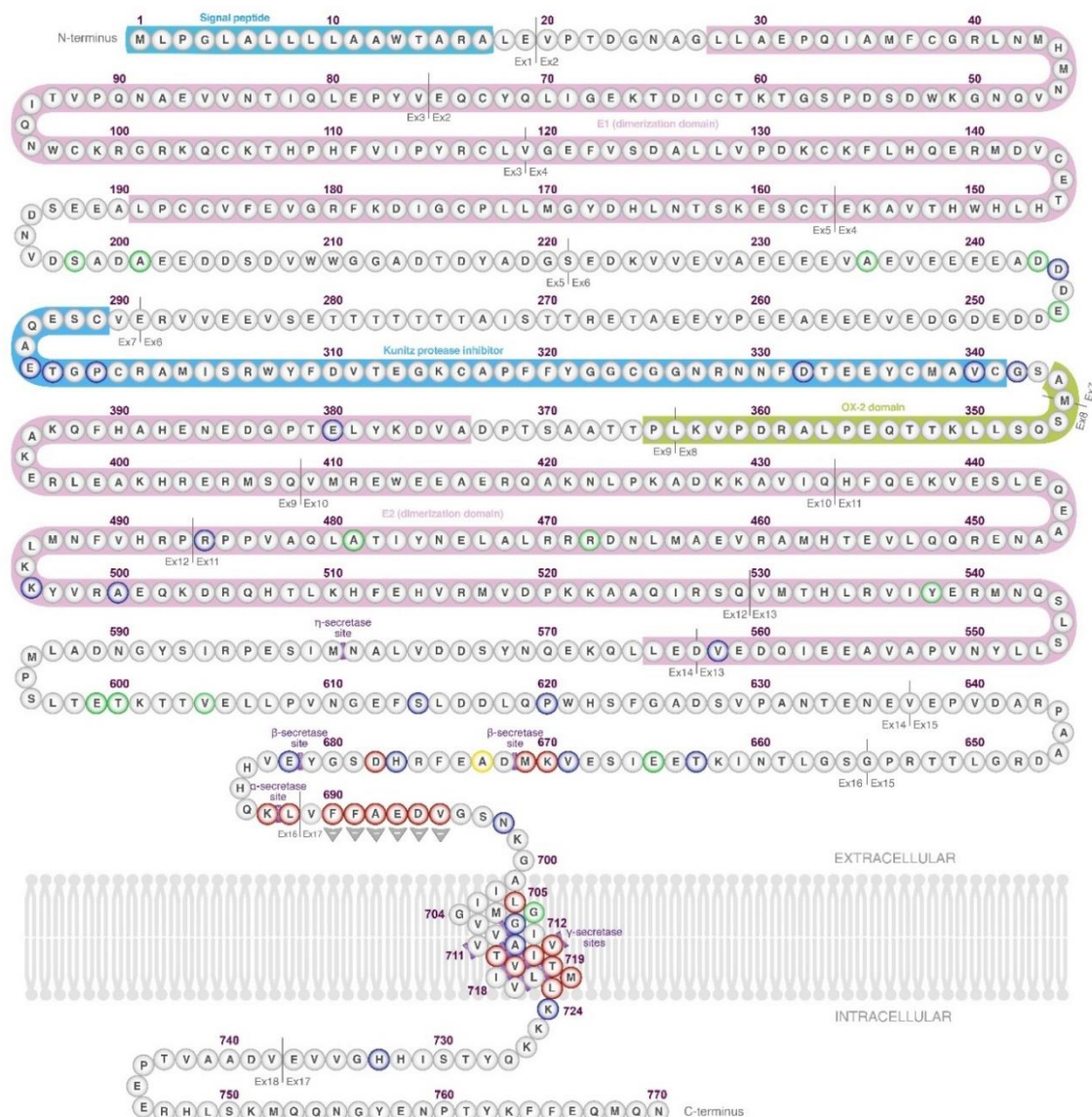
Figure S2 – APP sequence [https://www.alzforum.org/mutations/app]
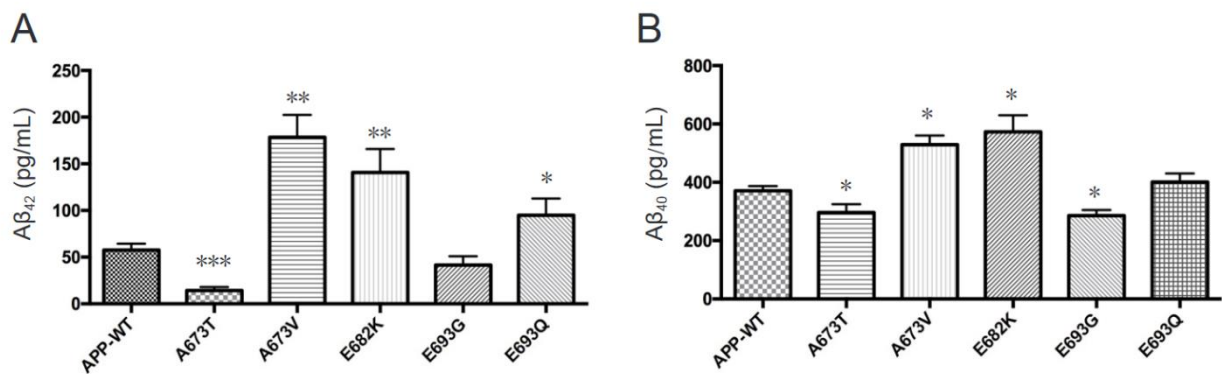
Figure S3 – Amyloid beta secreted levels are inferior in HEK293 cells transfected with A673T mutation compared to wild type. The pathogenic mutations conversely increase amyloid level secretion [doi:10.4103/1673-5374.247469].
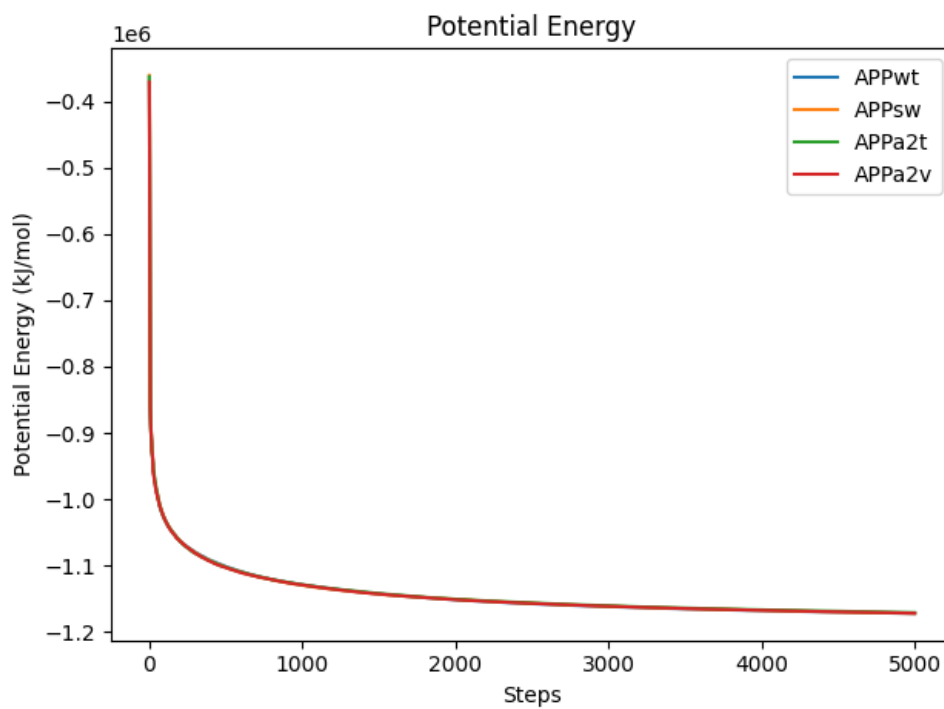
Figure S4 – Plot of different APP-BACE1 complexes' potential energy during minimization showing steady convergence.
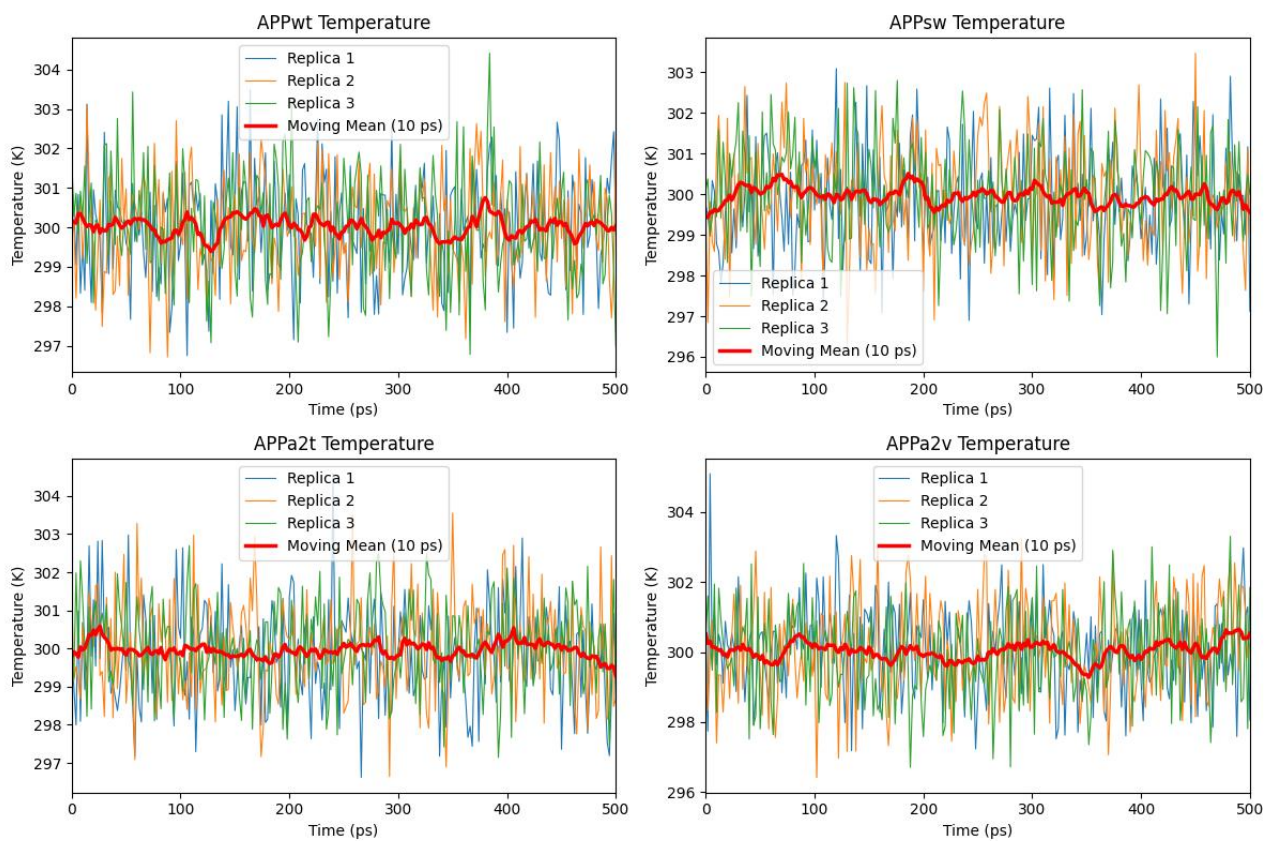
Figure S5 – Temperature plot of NVT equilibration for the four APP-BACE1 complexes. The red line is the computed moving average with 10ps window, oscillating at around 300K.
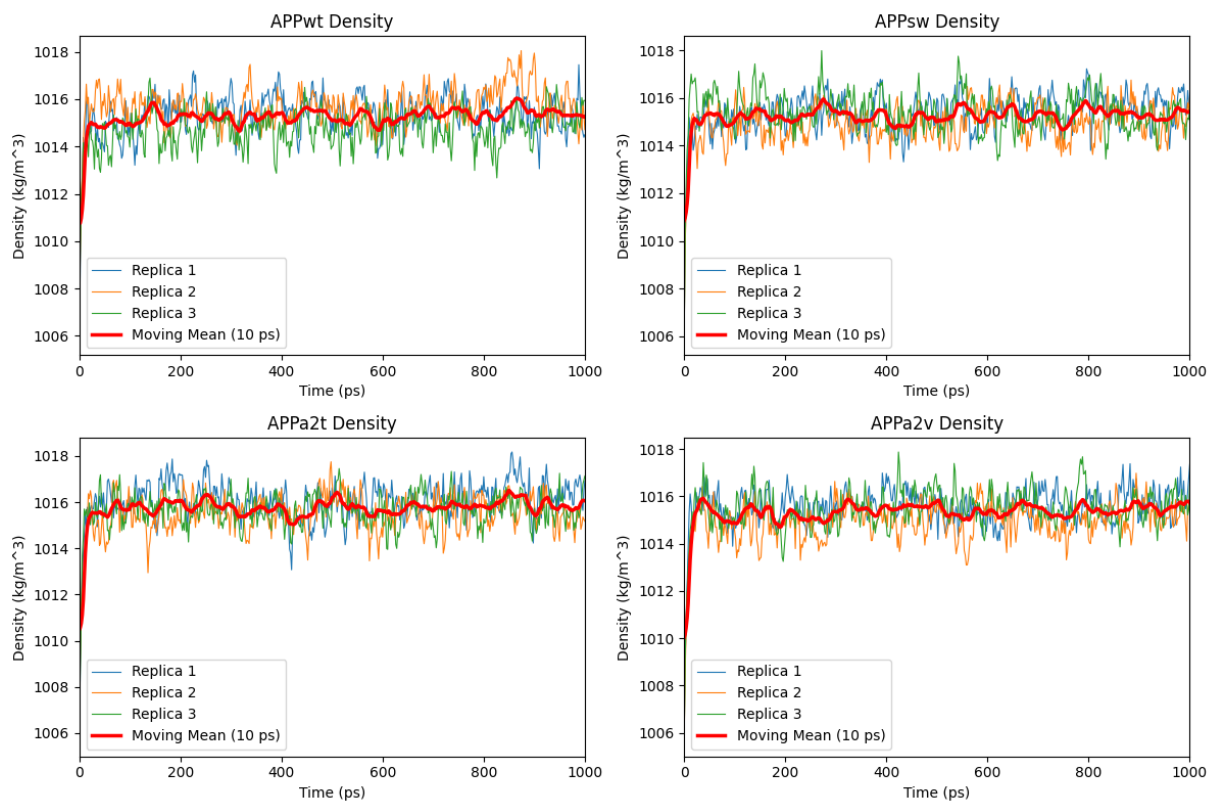
Figure S6 – Density plot of 1 ns NPT equilibration phase.
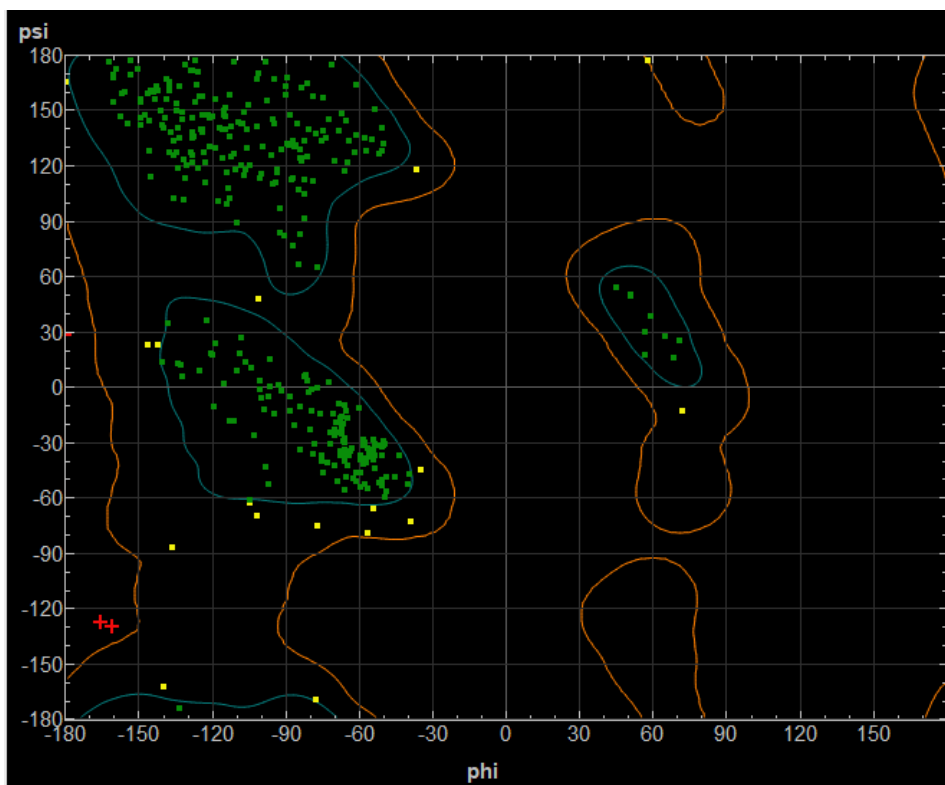
Figure S7 – Ramachandran plot of BACE1 (PDB: 1W50). Image created by MOE.
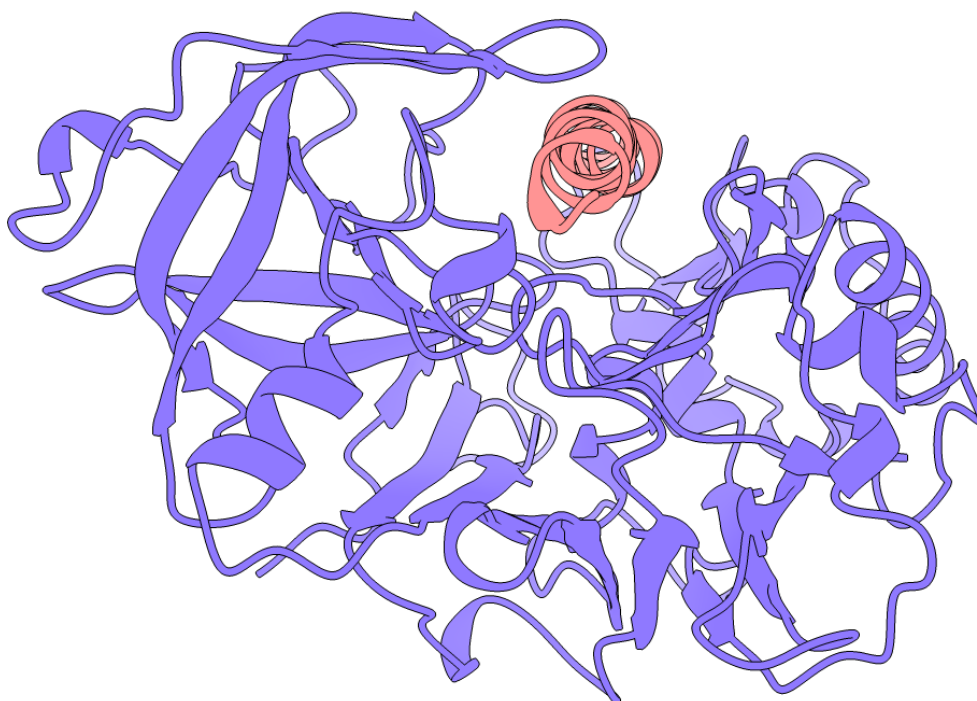
Figure S8 – Side view of BACE1 (violet) in complex with helical APP fragment (pink). Image obtained using ChimeraX.
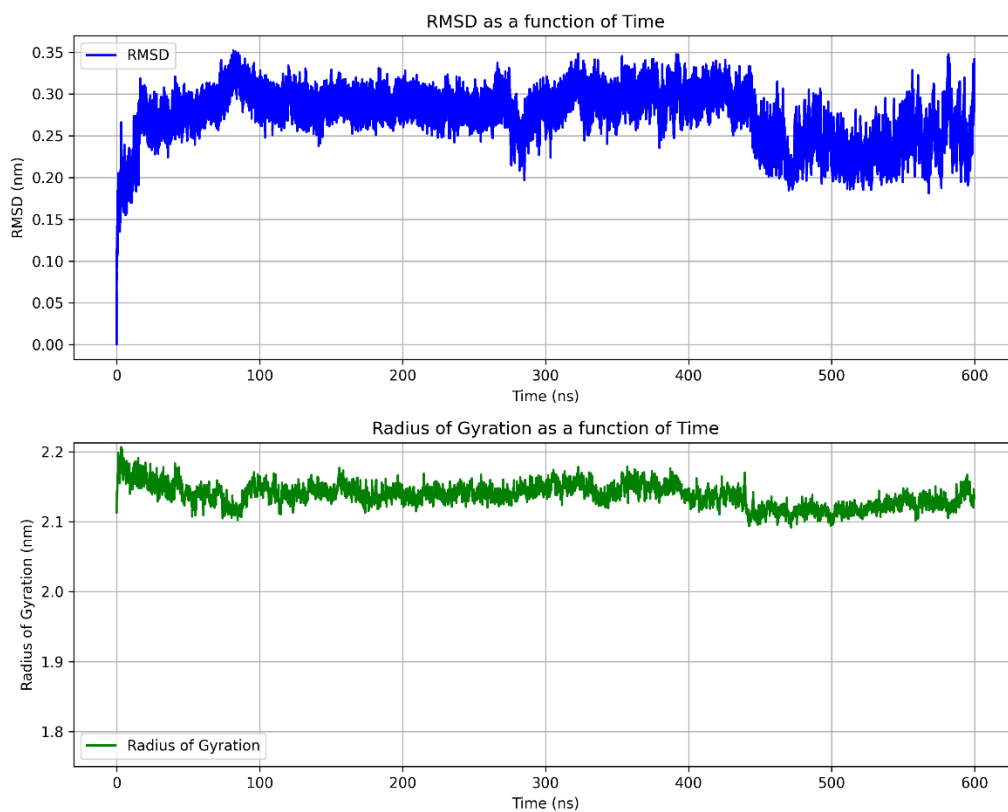
Figure S9 – RMSD (top) and Radius of Gyration (bottom) of unbound BACE1 simulation.
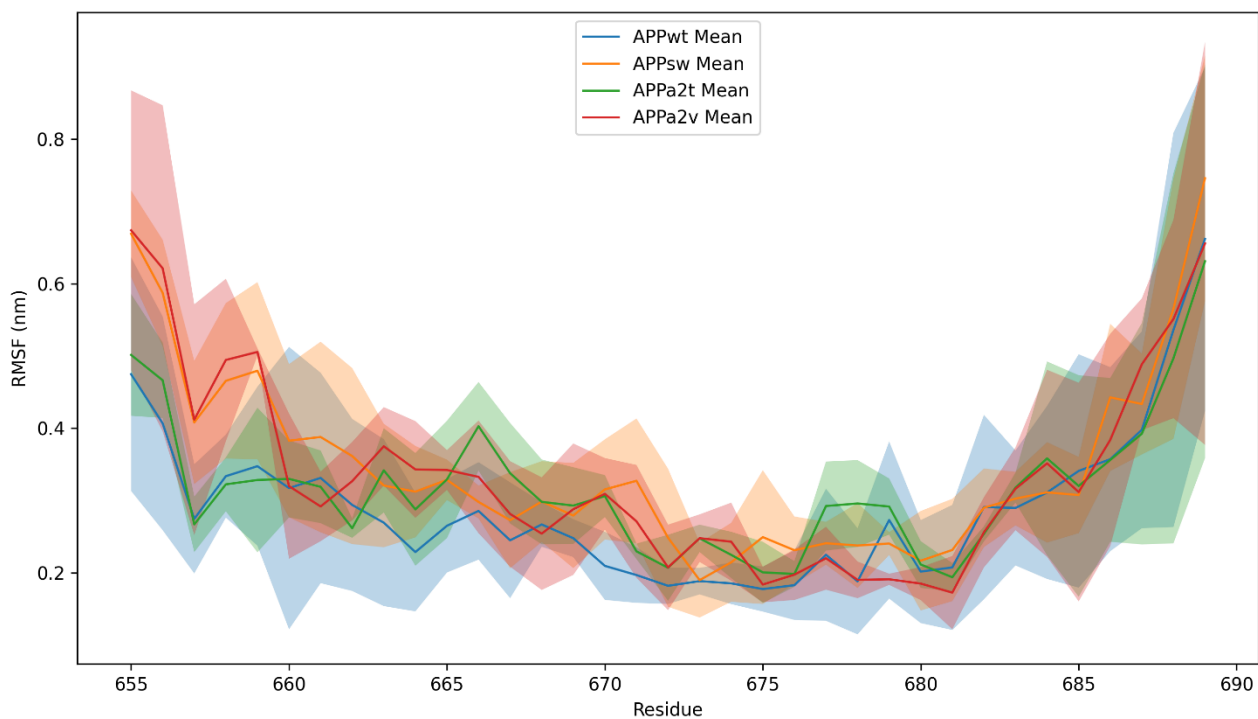
Figure S10 – Mean and standard deviation RMSF of APP.
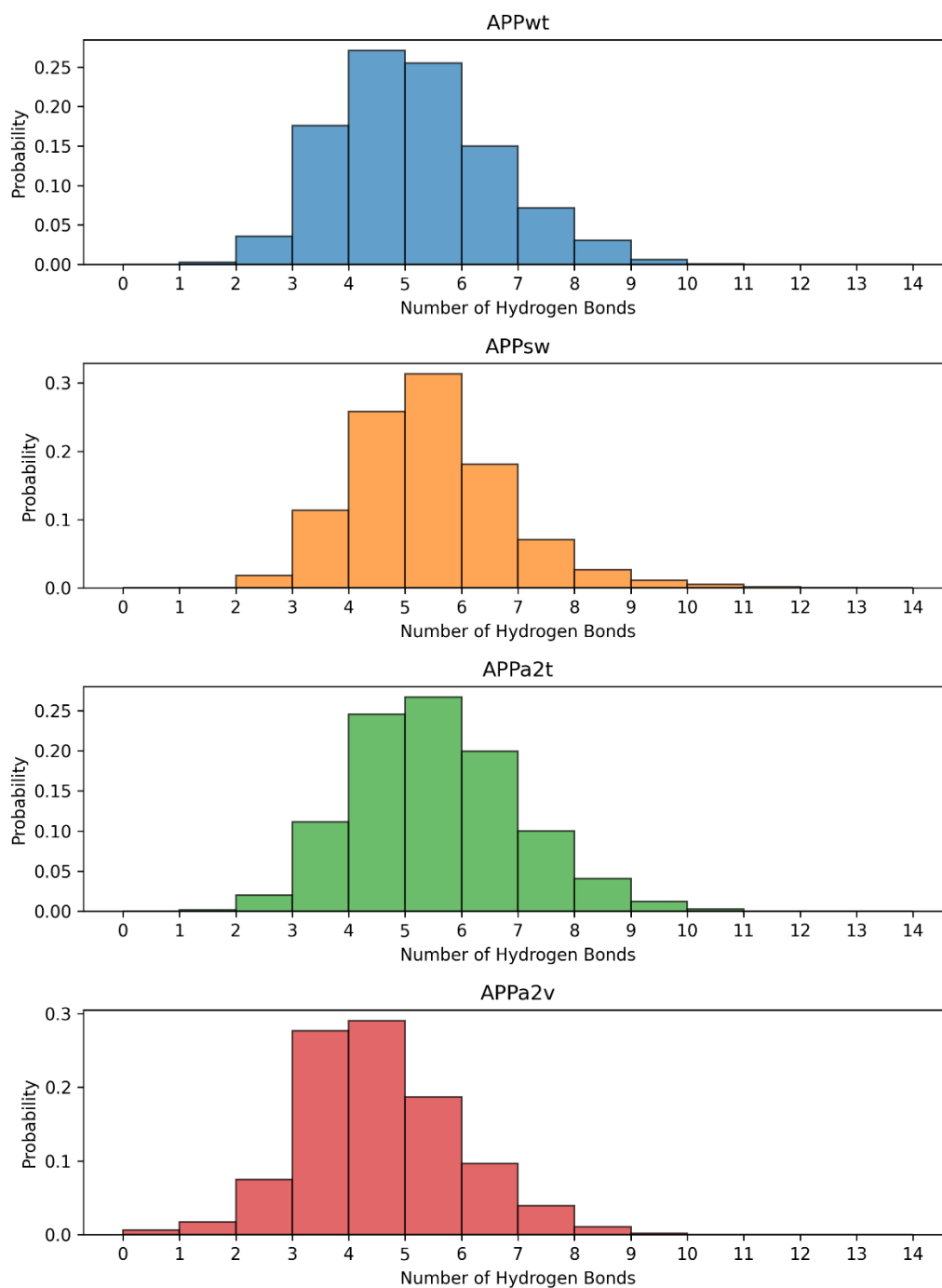
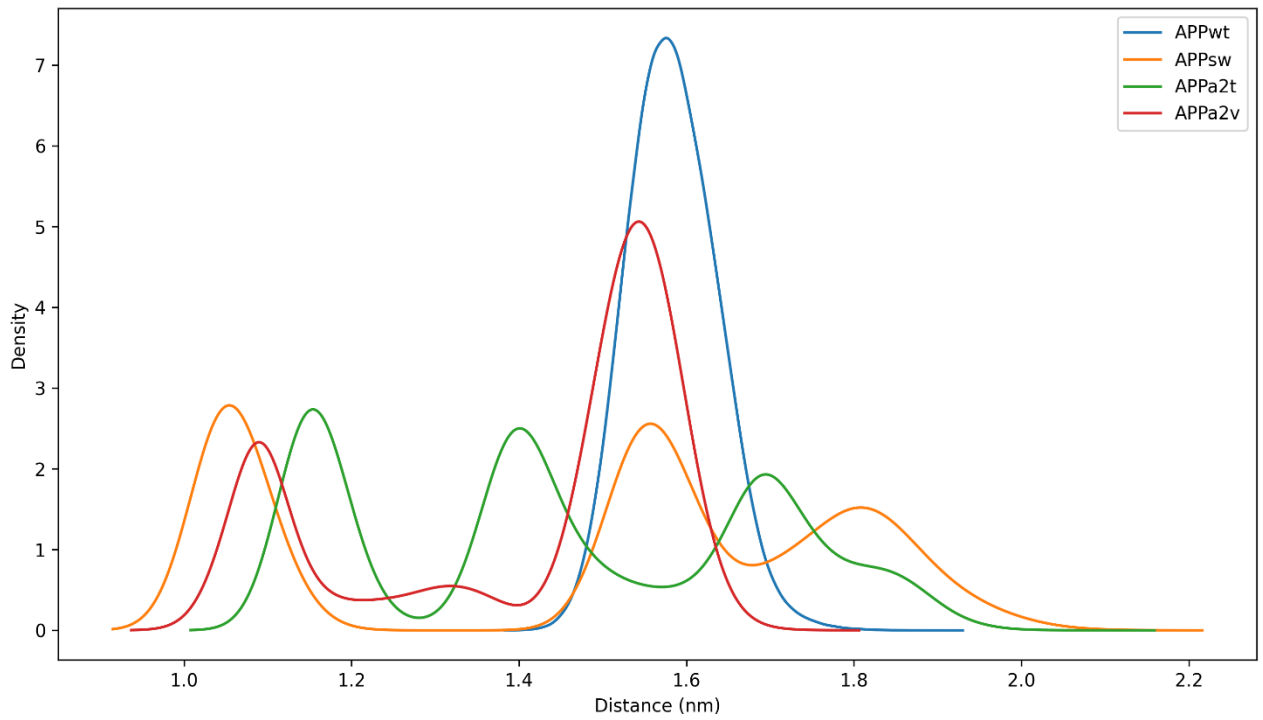*Figure S11 – Histograms showing the hydrogen bond probabilities between APP (residues 667-676) and BACE1.*

Figure S12 – Kernel Density Estimation plot comparing distance distribution between BACE1 Thr72 and Ser325 of different APP-BACE1 complexes.
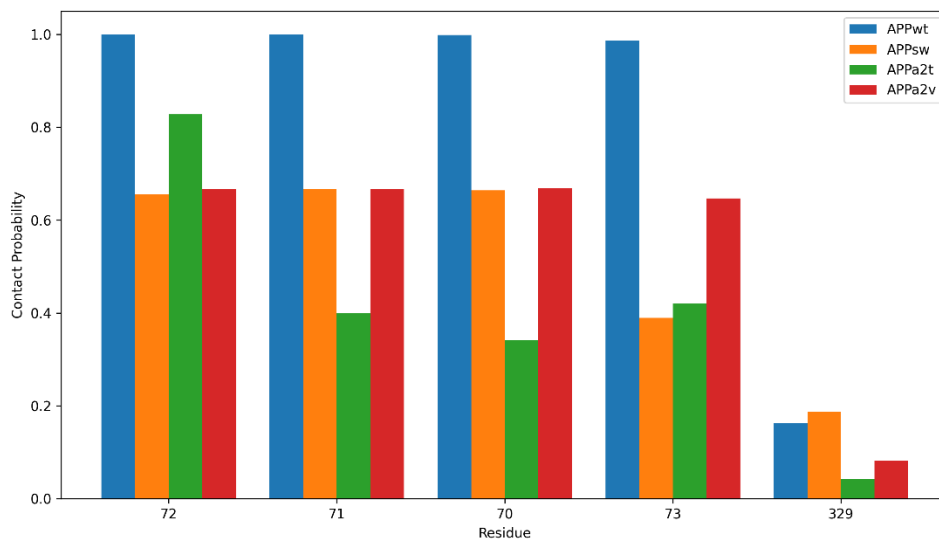

Figure S13 – Bar plot comparing the distance distributions of APP residue 672 and five BACE1 residues, using a 0.6 nm cutoff.
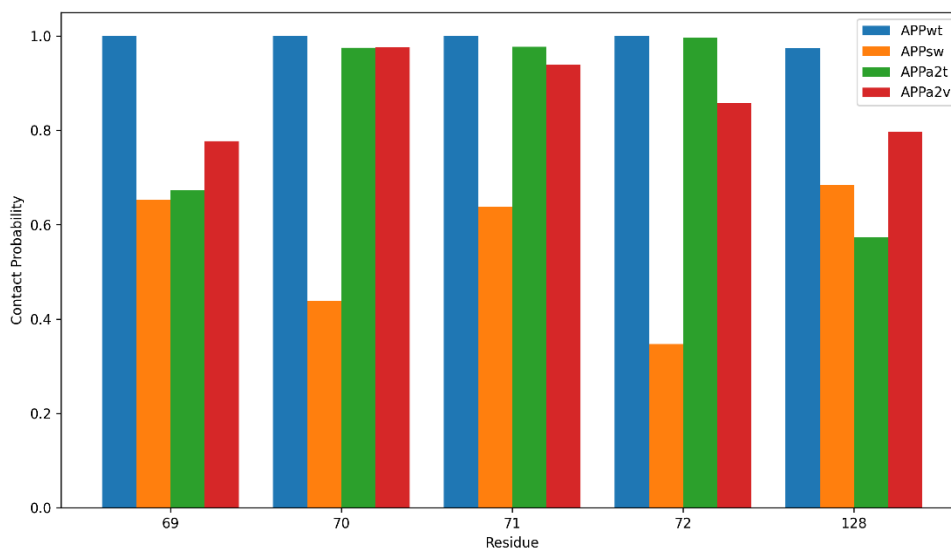
Figure S14 – Bar plot showing the distance distributions of APP residue 671 and five BACE1 residues, for different APP-BACE1 complexes, using a 0.6 nm cutoff.
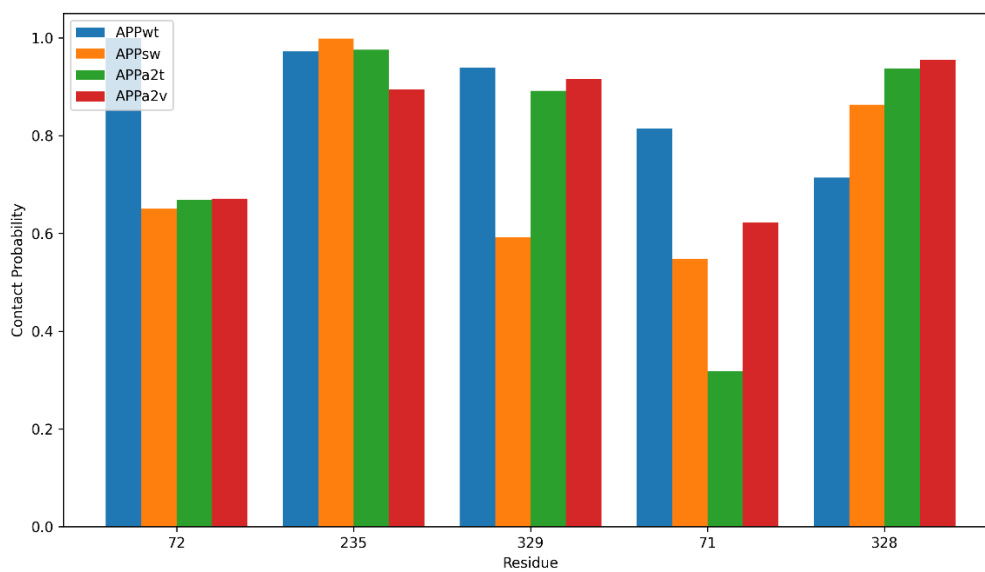

Figure S15 – Comparison of contact probabilities between APP residue 673 and five BACE1 residues using a 0.6 nm cutoff.

	APPsw		APPa2v		APPwt		APPa2t	
	mean	SD	mean	SD	mean	SD	mean	SD
ΔVDWAALS	-127,84	6,75	-148,11	7,61	-145,54	7,45	-83,94	6,43
ΔEEL	-137,07	63,8	-324,34	99,02	-181,87	59,89	-298,07	51,91
ΔEGB	207,65	55,84	383,25	97,33	265,2	55,01	333,93	50,7
ΔESURF	-18,56	0,82	-21,11	0,97	-21,73	0,84	-14,87	0,82
ΔGGAS	-264,92	62,51	-472,46	100,25	-327,42	58,25	-382,01	52,56
ΔGSOLV	**189,09**	**55,29**	**362,14**	**96,91**	**243,48**	**54,63**	**319,05**	**50,28**
ΔTOTAL	**-75,83**	**10,28**	**-110,32**	**7,8**	**-83,94**	**8,54**	**-62,96**	**6,03**

Table S1 – MM-GBSA results for different APP-BACE1 complexes, showing mean and standard deviations expressed in kcal/mol.

APPwt			APPa2v			APPsw			APPa2t		
Residue	Mean	SD	Residue	Mean	SD	Residue	Mean	SD	Residue	Mean	SD
R::ASN:233	-5,50	5,48	R::LYS:321	-8,09	9,88	R::THR:72	-7,26	4,70	R::THR:72	-6,74	4,75
R::SER:325	-5,03	4,21	R::THR:72	-7,20	4,60	L::LEU:671	-5,65	4,77	L::THR:663	-5,15	5,37
L::HIS:677	-4,93	4,96	L::TYR:681	-5,91	5,32	R::ARG:235	-5,13	12,72	L::ASN:660	-4,99	5,55
L::MET:671	-4,13	5,39	L::MET:671	-5,26	5,05	R::ARG:307	-4,36	11,78	L::THR:659	-4,95	5,89
L::THR:659	-3,85	5,38	L::HIS:677	-5,02	5,29	R::TYR:71	-3,62	5,34	L::MET:671	-4,70	4,93
R::ARG:128	-3,68	12,02	L::ILE:666	-4,84	4,94	L::SER:679	-3,47	8,28	R::ARG:128	-4,43	13,42
R::ARG:235	-3,28	10,63	R::LEU:267	-4,44	4,59	L::PHE:675	-3,37	4,90	R::ILE:110	-4,03	5,02
L::ASN:660	-2,99	5,65	R::TYR:71	-3,59	5,23	L::ILE:666	-2,79	4,91	R::ARG:235	-3,27	12,69
L::ARG:676	-2,96	15,94	R::THR:232	-3,49	5,00	R::PRO:129	-2,79	4,23	L::PHE:675	-3,19	4,68
L::TYR:681	-2,86	6,46	L::PHE:675	-3,39	5,24	L::THR:663	-2,54	5,51	R::TYR:71	-3,01	5,54

TableS2 – Decomposition analysis showing the top 10 scoring residues for different BACE1-APP systems (more negative results): R:: indicates BACE1 while L:: APP.
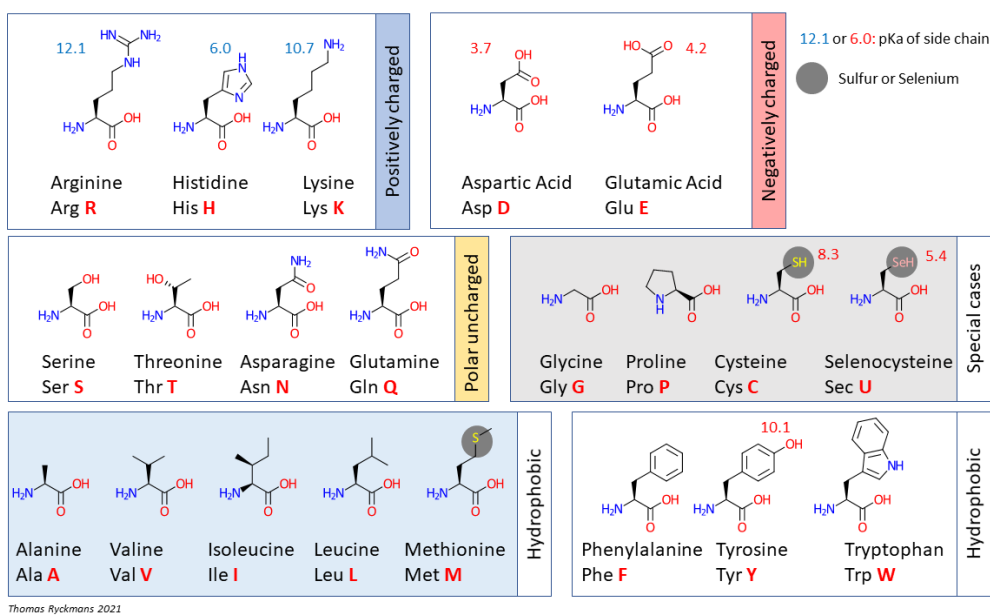
Figure S16 –Amino acid table showing three letter names and codes. Side chain physical properties are also reported. [https://commons.wikimedia.org/wiki/File:Proteinogenic_Amino_Acid_Table.png]
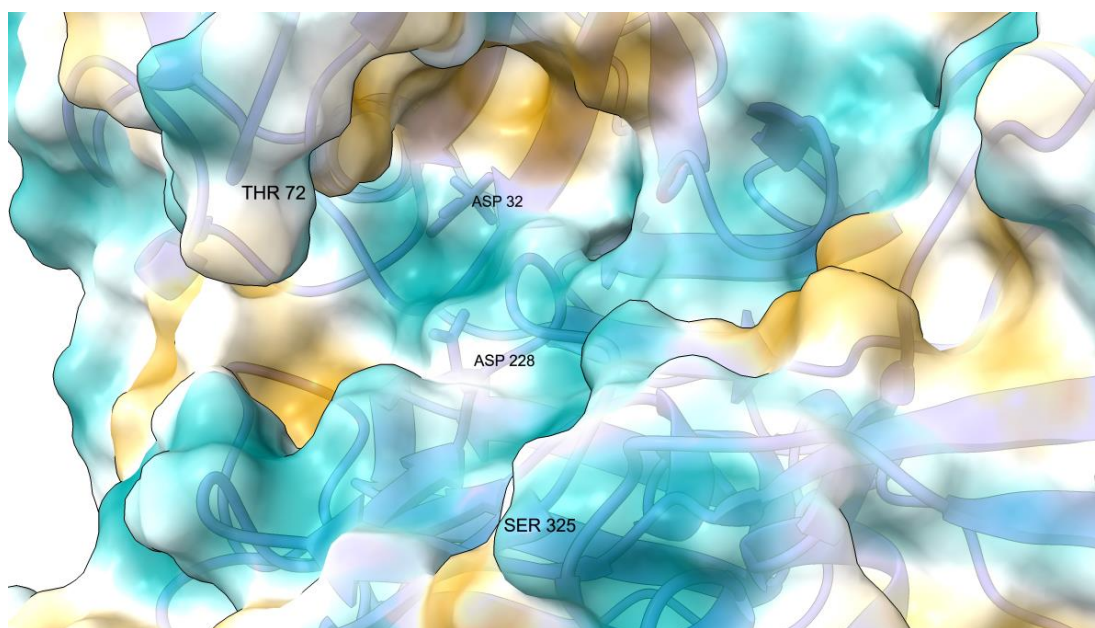

Figure S17 – Magnification of the active site of BACE1, with Asp dyad, flap residue Thr72 and Ser325 explicitly labelled. Hydrophilic zones are rendered in light blue, while hydrophobic ones in pale yellow. Image obtained with ChimeraX.
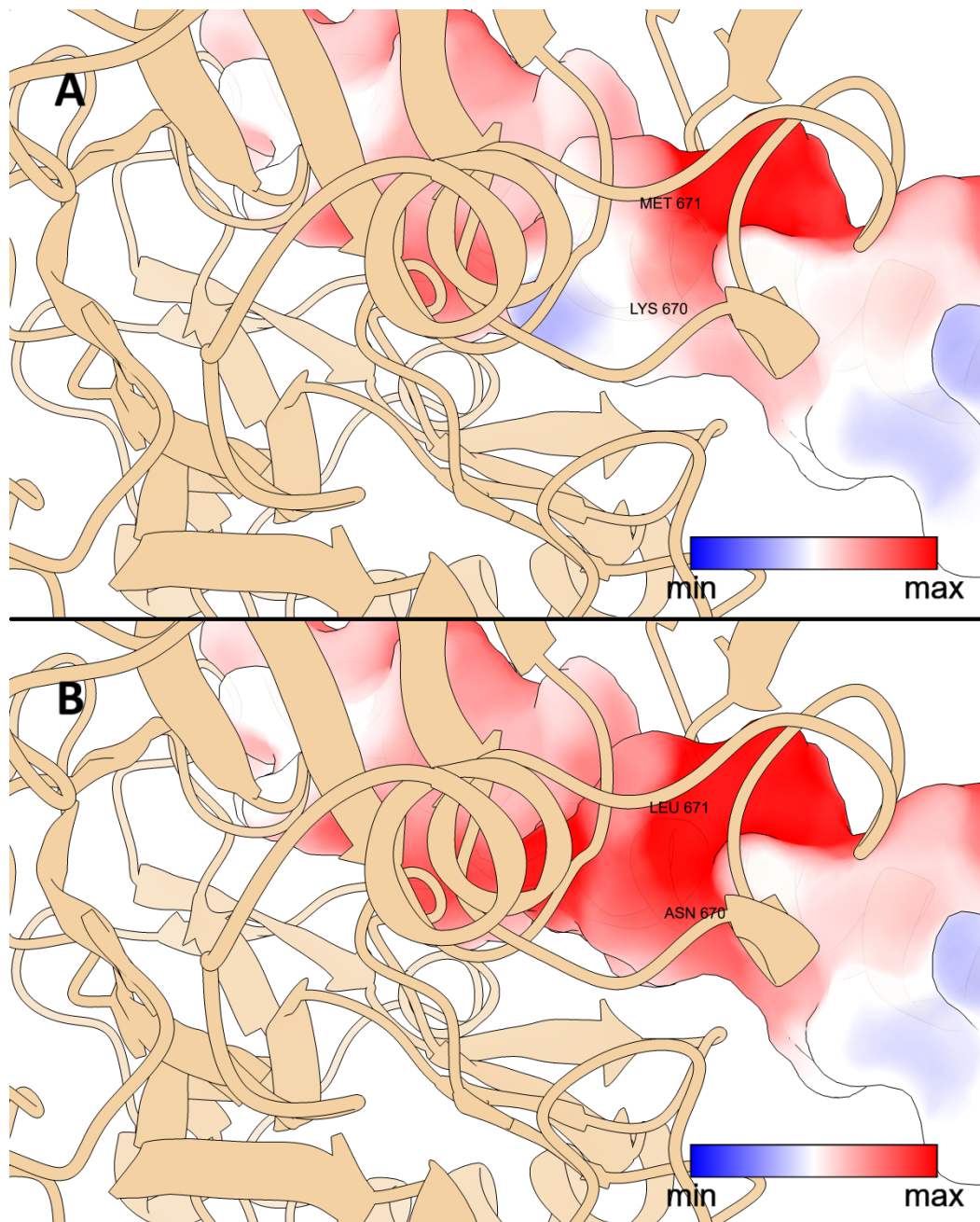
Figure S18 – Electrostatic surfaces of APPwt (A) and APPsw (B). The electron deficit is depicted with blue color while negatively charged regions are in red. Image created with ChimeraX.
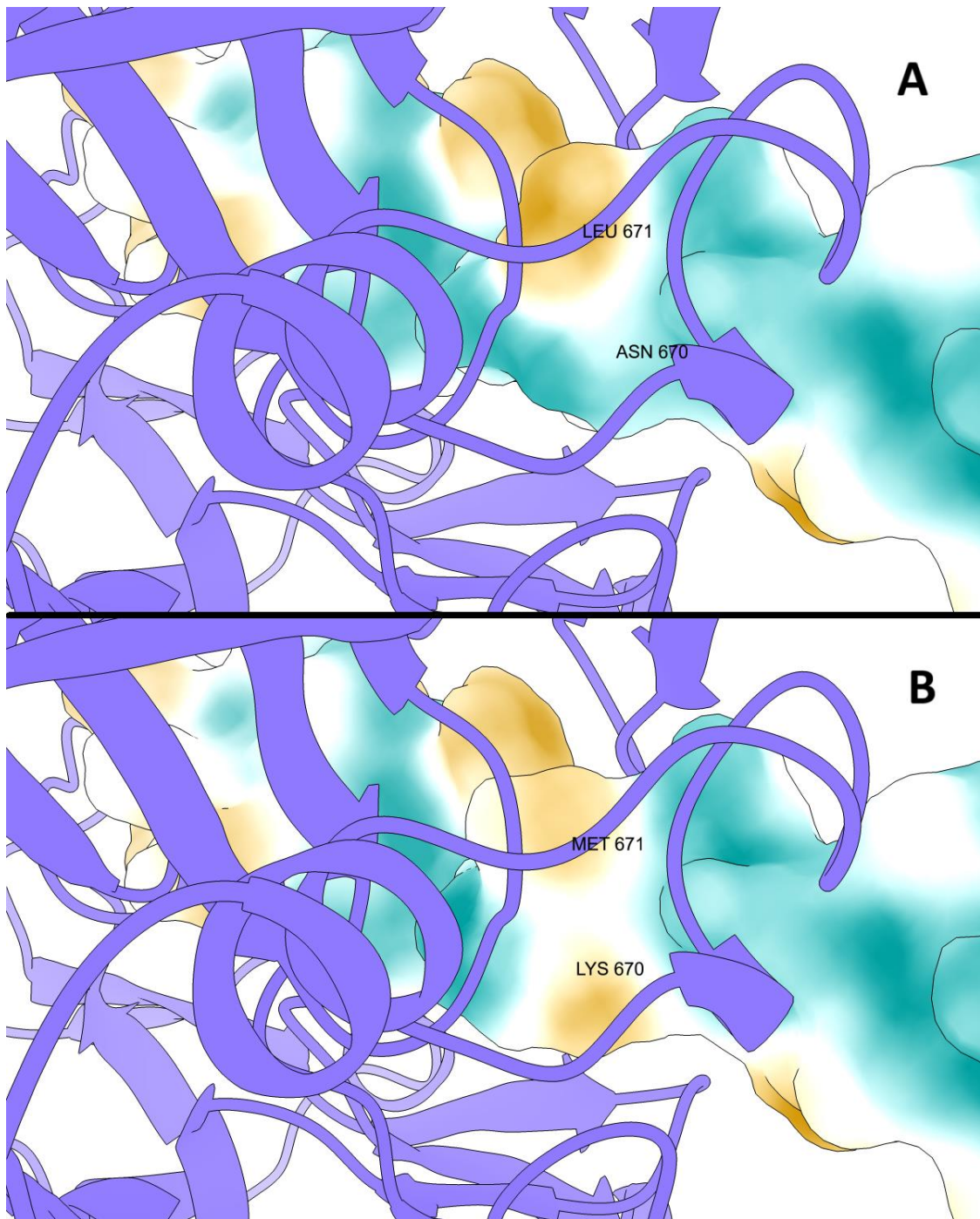
Figure S19 – Hydrophobic surfaces of APP expressing Swedish mutation (A) compared to the wild type (B). More hydrophilic regions are rendered in light blue, yellow for hydrophobic zones. Image obtained with ChimeraX.
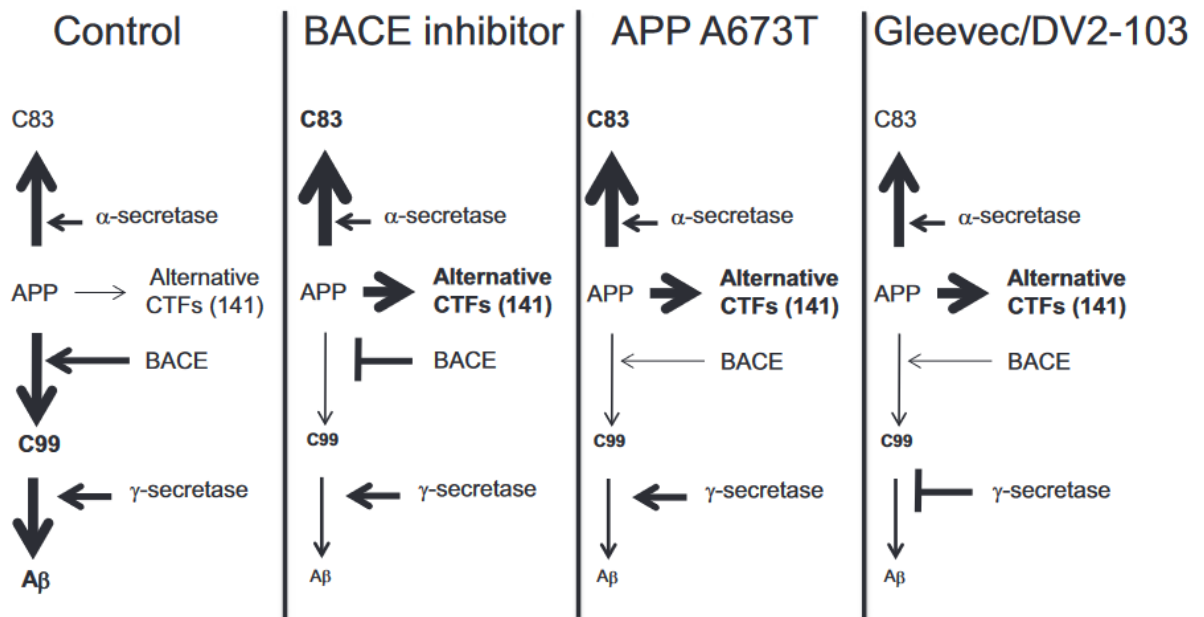
Figure S20 – Schematic representation of APP processing pathways alterations compared to control, from left to right using BACE1 inhibitors, in the case of APP expressing A673T mutation and with Gleevec interaction [http://www.pnas.org/cgi/doi/10.1073/pnas.1620963114].