

Studying membrane-protein interplay by modeling realistic conditions

THÈSE N° 7934 (2017)

PRÉSENTÉE LE 28 AOÛT 2017

À LA FACULTÉ DES SCIENCES DE LA VIE

UNITÉ DU PROF. DAL PERARO

PROGRAMME DOCTORAL EN CHIMIE ET GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Martina AUDAGNOTTO

acceptée sur proposition du jury:

Prof. V. Hatzimanikatis, président du jury

Prof. M. Dal Peraro, directeur de thèse

Prof. S. Vanni, rapporteur

Prof. R. Amaro, rapporteuse



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

*“We are still far from the time
when people will understand
the curious relationship between
one fragment of nature and another,
which all the same explain each other
and enhance each other. “*

- Vincent van Gogh

Abstract

Realistic models of cellular environments have long captured the imagination of biologists and physical scientists. More than 20 years ago, Goodshell's inspiring rendering of cellular environments provided the foundation for the idea of capturing the full complexity of biological cells. Since then, increasingly realistic simulations of subcellular models have appeared, showing that we are rapidly moving towards comprehensive, as well as physically and biologically accurate cellular models. The major challenge in developing such models is the necessity to cover a wide range of scales. Indeed, atomistic details of molecular processes are on the length scale of 0.1 nm, while cellular dimensions range from 300 nm (for the smallest bacterial cells) up to 100 μm (for large eukaryotic cells). The biological range also encompasses the internal dynamics of individual molecules (from ns to ms) as well as entire biological processes (ms to hours). Although an atomistic representation of an entire cell is conceivable, the cost for achieving biologically significant timescales with most advanced computers is very high or nearly impossible currently. Therefore, the combination of *in silico* approaches with *in vivo* and *in vitro* experimental data, allows gaining further insights into the biophysics of the cell. In this thesis, a multi-scale approach was applied to investigate the interplay between proteins and their respective biological environments. In particular, two specific systems were the objects of my Ph.D. project: (i) the amyloid precursor protein (APP) and its interactions at the synaptic plasma membrane, and (ii) the human acyl-protein thioesterase I (hAPT1) and its role in protein depalmitoylation cycle across the Golgi and plasma membranes.

Combining atomistic and coarse-grained molecular dynamics (MD) simulations, the stability of the dimeric conformation of APP in a realistic model of the synaptic plasma membrane was investigated. Within this approach, the stable dimeric conformation of the APP as a function of lipid membrane composition was identified, highlighting the important role of lipids unsaturation in protein stability and function.

In detail, this indicated not only the preferential dimerization motif of the protein, but also the possible recruiting mode of cholesterol, which has previously been experimentally observed to correlate with Alzheimer's disease. Furthermore, the role of the synaptic plasma membrane and its ability to form raft microdomains was explored for the interactions of APP with γ -secretase, which specific catalytic cleavage is responsible for the production of physiological or toxic amyloid β -secretase which is specific for the Alzheimer's disease onset.

This same multi-scale approach was applied to understand the role of S-acylation in the trafficking and function of hAPT1. In particular, the combination of experimental techniques (i.e., cell biology, biochemistry, microscopy and structural biology) with molecular modeling and simulation disclosed the substrate/enzyme interaction mode, the enzyme/membrane interface, and the role for hAPT1 for a key post-translational modification as S-palmitoylation. Based on our findings, we provided for the first time an atomistic representation of the hAPT1 enzyme mechanism, suggesting that hAPT1 is active in the monomeric form.

In conclusion, these studies demonstrate once more the importance of modeling the molecular details of the biological environment at the most accurate level for acquiring a deeper understanding of cellular processes. Molecular models and simulations able to cover relevant spatial and temporal scales have the ability to generate both physically and biologically accurate cellular representations, which eventually can provide broader practical applications for understanding biological function and for addressing therapeutics development.

Keyword: molecular dynamics, molecular modeling, biological membrane, post-translational modifications, X-ray crystallography, Amyloid Precursor Protein (APP), Alzheimer's disease (AD), S-acylation, Acyl Protein Thioesterase (APT)

Riassunto

I modelli realistici delle cellule hanno catturato a lungo l'immaginazione di fisici e biologi. 20 anni fa, le rappresentazioni degli ambienti cellulari ad opera di Goodshell hanno suggerito l'idea di poter carpirne la complessità e da allora è stato un susseguirsi di esempi sempre più realistici dimostrando che possiamo fornire modelli cellulare accurati dal punto di vista fisico e biologico. La sfida più grande nello sviluppo di questo tipo di modelli risiede nella necessità di includere diverse dimensioni. I processi biologici avvengono in scale dell'ordine dei 0.1 nm mentre le dimensioni delle cellule sono comprese in un range tra 300 nm (nel caso di piccoli batteri) fino a 100 μm (per grandi cellule eucariote). La dinamica dei processi biologici include sia il movimento delle singole molecole (da ns fino a ms) che gli interi processi cellulari che spaziano da ms fino a ore. Sebbene una rappresentazione atomistica della cellula sia immaginabile, il costo computazione per raggiungere scale di tempo che siano biologicamente rilevanti è alto o addirittura quasi impossibile anche con le moderne potenze di calcolo di cui disponiamo. Per questo motivo, la combinazione di tecniche computazionali e dati sperimentali permette di conoscere meglio il comportamento biofisico delle cellule.

In questa tesi, è stato adottato un approccio a più livelli al fine di investigare l'interazione tra proteine e i loro rispettivi ambienti biologici. In modo particolare, due specifici sistemi sono stati oggetto dei miei studi: (i) l'amyloid precursor protein (APP) e la sua interazione con la membrane sinaptica, e (ii) l' human acyl-protein thioesterase 1 (hAPT1) e il suo ruolo nel ciclo di depalmitoilizzazione attraverso Golgi e la membrana plasmatica.

Tramite simulazioni di dinamica molecolare sia a livello atomistico che coarse-grained, la stabilità della conformazione dimerica dell'APP è stata investigata in presenza di un modello realistico di membrane sinaptica identificandone la struttura più stabile e in particolar modo sottolineando l'importante ruolo dei doppi legami dei

lipidi nell'influencare la stabilità e la funzione delle proteine. Inoltre questo studio ha evidenziato non solo il più probabile sito di dimerizzazione dell'APP ma anche la possibile interazione con il colesterolo, la quale è stata in precedenza sperimentalmente correlata con la malattia dell'Alzheimer. Inoltre, il ruolo della membrana sinaptica e la sua abilità nel formare raft è stata esplorata anche per l'interazione tra l' APP e l'enzima responsabile per il suo cleavage e conseguente formazione degli A β : la γ -secretase.

Il medesimo approccio a più livelli è stato adottato per studiare il ruolo dell' S-acilazione nella localizzazione e funzione dell'hAPT1. In particolare la combinazione di tecniche sperimentali (quali biochimica, microscopia, biologia strutturale e biologia cellulare) con modelli molecolari e simulazioni hanno rivelato non solo il modo di interazione tra substrato ed enzima ma anche l'interazione tra enzima e membrane e il ruolo cruciale dell' hAPT1 nel processo di S-acilazione. Sulla base dei nostri risultati abbiamo proposto per la prima volta una rappresentazione atomistica del meccanismo di azione dell' hAPT1 dimostrando che questo enzima è attivo quando monomero.

In conclusione, questi studi rivelano ancora una volta l'importanza di accurati modelli biologici al fine di ottenere una comprensione profonda dei sistemi che li governano. Modelli molecolari e simulazioni che coprono scale di tempo e spazio biologicamente rilevanti permettono di generare rappresentazioni della cellula accurate sia dal punto di vista fisico che biologico ed eventualmente essere applicati al fine di studiare funzioni biologiche e proporre nuove cure.

Parole chiave: simulazioni di dinamica molecolare, modelli molecolari, membrane biologiche, modificazioni post-translazionali, cristallografia, Amyloid Precursor Protein (APP), Alzheimer, S-acilazione, Acyl Protein thioesterase (APT)

Table of Contents

ABSTRACT.....	4
RIASSUNTO.....	6
ACKNOWLEDGEMENTS.....	10
1 INTRODUCTION.....	12
1.1 BIOLOGICAL MEMBRANES AND MEMBRANE PROTEINS: A SUCCESSFUL MARRIAGE 12	
1.2 POST-TRANSLATIONAL MODIFICATIONS: HOW PROTEINS DIVERSIFICATE THEMSELVES.....	16
1.3 MOLECULAR SIMULATIONS: A COMPUTATIONAL MICROSCOPE.....	19
1.4 OBJECTIVES OF THE THESIS.....	21
2 MOLECULAR MODELING AND SIMULATION METHODS.....	30
2.1 MOLECULAR MECHANICS.....	31
2.2 BONDED INTERACTIONS.....	32
<i>BOND STRETCHING</i>	32
<i>ANGLE BENDING</i>	33
<i>TORSIONAL ANGLES</i>	34
<i>IMPROPER TORSIONAL ANGLES</i>	34
2.3 NON-BONDED INTERACTIONS.....	35
<i>ELECTROSTATIC INTERACTIONS</i>	35
<i>VAN DER WAALS INTERACTIONS</i>	35
2.4 COARSE-GRAINED REPRESENTATION OF BIOMOLECULES.....	36
2.5 COMPUTATIONAL TOOLS FOR SETTING UP COMPLEX BIOLOGICAL SYSTEMS.....	38
<i>LIPIDBUILDER: HOW TO SET UP YOUR FAVOURITE MEMBRANE AT ATOMISTIC RESOLUTION</i>	38
<i>INSANE: LIPIDOMICS AT THE COARSE-GRAINED LEVEL</i>	39
<i>HOW TO DEAL WITH POST-TRANSLATIONAL MODIFICATIONS</i>	39
2.6 CLASSICAL MOLECULAR DYNAMICS.....	41
2.7 ERGODIC HYPHOTHESIS.....	43
2.8 PERIODIC BOUNDARY CONDITIONS.....	44
2.9 MD SIMULATIONS IN DIFFERENT ENSEMBLES.....	45
3 AMYLOID PRECURSOR PROTEIN AND Γ-SECRETASE: THE TWO MAIN CHARACTERS OF ALZHEIMER'S DISEASE.....	52
3.1.....	53
THE EFFECTS OF LIPIDS COMPOSITION ON STABILITY AND FUNCTION OF AMYLOID PRECURSOR PROTEIN.....	53
INTRODUCTION	54
RESULTS AND DISCUSSION	55
CONCLUSIONS	65
COMPUTATIONAL MATERIAL AND METHODS	66
3.2 Γ -SECRETASE MEMBRANE AND APP INTERPLAY: HOW THE TWO MAIN CHARACTERS OF ALZHEIMER DISEASE GO ALONG TOGHETER.....	68
RESULTS AND DISCUSSION	71
COMPUTATIONAL MATERIAL AND METHODS	78
4 INSIGHTS INTO THE MOLECULAR MECHANISM OF DEPALMITOYLATION CATALYZED BY HUMAN ACYL-PROTEIN THIOESTERASE 1.....	88
INTRODUCTION	88
RESULTS AND DISCUSSION	89
CONCLUSION	101
COMPUTATIONAL MATERIAL AND METHODS	102

5	CONCLUSIONS AND OUTLOOKS.....	110
6	CURRICULUM VITAE.....	115

Acknowledgements

I would like to sincerely thank Professor Matteo Dal Peraro, my PhD advisor, for his encouragement and for giving me the freedom in exploring all the scientific aspects that tickled my mind during these years.

For the friendly and quite noisy environment I am grateful to all my lab mates. A special thanks to Luciano, with whom I spent five years in the same office talking about scientific topics, while playing the drum. How can I forget Giulia and Alessio, who revealed to me the modern society's behaviors through their precious videos? It was a great pleasure to collaborate with Professor van der Goot, who together with Professor Dal Peraro encouraged me to directly contribute to the experimental part.

As Woody Allen said at the beginning of one of his movies, *"I would never wanna belong to any club that would have someone like me for a member."* Well, I would do the same but fortunately my friends are not of the same opinion. Therefore let's start to thank the members of this "club". In primis, I want to thank Chiara and Cate for all the wonderful moments we spent this year and for being always there for me no matter what. Helenita, my dear friend, thanks for all the laughs and chocolate you provide to me, without your *"Marty, que pasa?"* my Monday morning would not be the same. Alek, it was by chance that we started hanging out together and since then we have collected a series of funny stories.

The club also includes friends from the other side of the mountains: my dear Josephine, thanks a lot for supporting me every time I needed it. Cate, since the first time we met up at the college enrolment (almost 10 years ago) several things changed in both of our lives but I know I can always count on your friendship.

Last but not least, I want to thank my family for the constant support and encouragement.

1 Introduction

Life takes place in very complex environments, as *van Gogh* believed “*we are still far from the time when people will understand the curious relationship between one fragment of nature and another*”. Although we are still at the beginning of our exploration on the complex biological world that is the cell, great progress in the fields of physics, chemistry and biology has allowed moving from a reductionist approach where the system of interest (e.g., a protein, a small molecule) is isolated from its biological environment to more integrative and holistic approaches.

Nowadays, the advances of experimental high-throughput technologies and computational modeling changed our perception of nature, revealing new levels of biological organization. In particular, the architecture of the cell occupies an important position since it is considered the simplest characteristic unit of life. Therefore, lipids composition of the cellular organelles, protein-protein and membrane-protein interactions as well as crowding effects have become important aspects of research, which are starting to be taken into account to surpass our simplistic models. The necessity to raise the level of complexity of our studies began from the point at which the simple model fails in describing the behavior of the actual biological system. That is the case especially for membrane proteins, whose study is majorly hampered by the lack of experimental techniques capable of controlling and directly observing at the molecular level their real lipids composition and distribution.

1.1 BIOLOGICAL MEMBRANES AND MEMBRANE PROTEINS: A SUCCESSFUL MARRIAGE

Proteins and lipids are the principal components of biological membranes. Lipids are the main constituents of the characteristic membrane bilayer, which is

formed by the entropically favored interaction of the hydrophobic lipid acyl tails and the tendency of the hydrophilic lipids head to interact with the aqueous solution. The amphipathic nature of lipids enables thus the construction of a cellular envelope that provides a physical separation of the internal constituents from the external environment of every cell. The subcellular compartmentalization allows then the segregation of specific chemical reactions increasing their biochemical efficiency and confining their reaction products¹. In particular, the physiochemical properties of lipids are essential in several important biological processes in the cell. For instance, they can act as first and second messengers in signal transduction² or they are implicated in budding³, scission and fusion⁴, which are essential for cell division⁵, membrane trafficking⁶ and biological reproduction⁷.

Lipidomics analyses have shown that the variation in headgroups and aliphatic chains allows the existence of more than thousand different lipids species in any eukaryotic cell^{8,9}. Lipids can be grouped in three main categories: (i) glycerophospholipids, (ii) sphingolipids and (iii) sterols. Glycerophospholipids form the major structural class of lipids in eukaryotic membranes. Their head can vary as well as the length and number of unsaturations in their hydrophobic tails. This class includes phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylinositol (PI), phosphatidylserine (PS) and phosphatidic acid (PA) lipids. In particular, PC accounts for more than 50% of the phospholipids in most eukaryotic membranes^{10,11}. In mammalian cells sphingolipids are mostly observed especially, sphingomyelin (SM) and glycerophospholipid (GSLs) that contains mono-, di- or oligo-saccharides^{12,13}. Sterols are the main non-polar lipids of cell membranes: cholesterol in mammals¹⁴ while ergosterol in yeast^{15,16}. Interestingly, the distribution of phospholipids and sterols is not homogeneous throughout the leaflets of the main mammals and yeast organelles. The endoplasmic reticulum (ER), plasma membrane (PM) and Golgi display an asymmetric lipid distribution with SM and GSLs present on the exo-leaflet, while PE and negatively charged PS are located on the cyto-leaflet^{1,17}. This asymmetric distribution is due to an intrinsic capacity of the lipids to cross spontaneously the bilayer and to the presence of transporters that assist lipids translocation.

In synthetic membrane, lipids can exist in multiple and different phase states, which depend on membrane composition. In particular, long, saturated hydrocarbon chains tend to adopt solid-like phases while unsaturated hydrocarbon chains are

principally observed in liquid phases ¹. On the contrary, in biological membrane the existence of different phases, which may form for instance, ordered assemblies characterized by sterol and sphingolipids, is still under evaluation. Although in the latest 10 years it was hypothesized that these metastable states called “lipid rafts” can be activated by the presence of specific lipid-lipid, protein-lipid and protein-protein interactions ¹⁸⁻²⁰, clear and definitive in vivo evidences of their existence are still missing.

Sterol molecules alone do not form any phase, but they can in turn influence the behavior of other lipids in mixture bilayers leading to a liquid-order phase²¹. This state has the high order of a solid phase, but the high translational mobility of a liquid phase. Much more complex phase behavior has been observed for a mixture of saturated and unsaturated lipids, where the liquid-order (l_o) and liquid-disorder (l_d) phase coexist. Moreover, at physiological temperature a gel phase rich in saturated and well-ordered lipids usually exists²². Therefore, the general hypothesis is that the plasma membrane is not uniform but composed by lipid rafts, which are rich in saturated lipids and cholesterol and are supplied by the Golgi²³⁻²⁵.

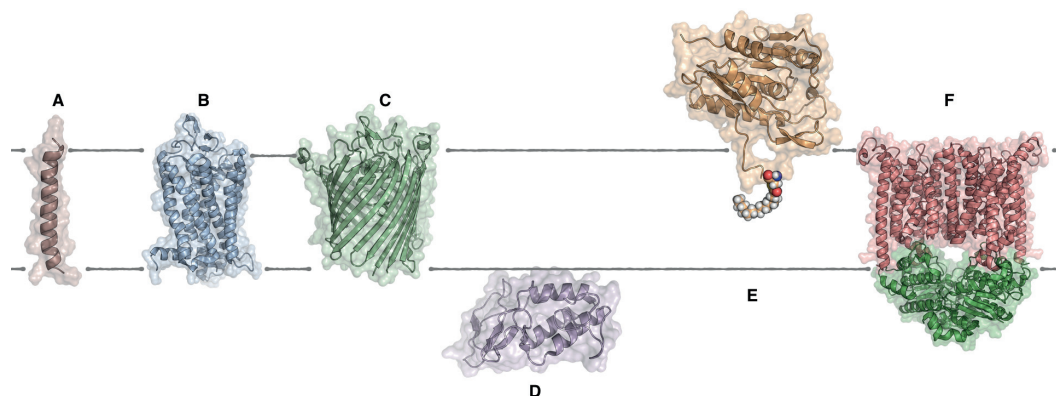


Figure 1.1. Membrane-protein interaction modes. Most membrane proteins are thought to extend across the bilayer as (A) single α -helix (types I and II), (B) multiple α -helices or as (C) rolled-up β -sheet (β -barrel). Other membrane proteins are exposed at only one side of the membrane (D) or attached to the lipid bilayer solely by a covalently attached lipid linker (E). Finally, many proteins are attached to the membrane only by non-covalent interactions with other membrane proteins (F).

Proteins are the other dominant component of biological membranes. Membrane proteins constitute 20-30% of all proteins in both eukaryotic and prokaryotic organisms^{26,27} representing the portal through which cells and organelles communicate with the external environment. The amounts and types of membrane

proteins are a function of the specific membrane considered. For instance, in myelin membranes, which are responsible for the electrical insulation of nerve cells axons, the total amount of proteins is less than 25% of the total membrane mass²⁸. On the contrary, in the PM and mitochondrial membranes, proteins account for roughly 50% and 75% of the total membrane mass, respectively²⁹.

As for lipids, membrane proteins are amphipathic entities that have a hydrophilic region exposed to the solution and a hydrophobic part, which instead is embedded into the bilayer. Based on the nature of membrane-protein interactions, membrane proteins can be classified in two main categories: integral and peripheral membrane proteins. Integral membrane proteins have one (**Fig. 1.1A**) or more hydrophobic regions (**Fig. 1.1B**) embedded into the bilayer, while the hydrophilic regions are exposed to water solution on both sides of the membrane. The great majority of the membrane-spanning regions form α -helices in order to maximize the H-bond network between peptides in absence of water. These single passing proteins can be classified as type I or type II, depending if the cytoplasmic tail is the C- or N-terminus, respectively. If membrane proteins have multi-passing spanning regions an alternative way for peptide bonds to satisfy their H-bonding requirement is also the β -barrel organization (**Fig. 1.1C**). Specific functions are associated with these three different folds. Single-spanning membrane proteins (**Fig. 1.1A**) are generally involved in metabolism and structure/adhesion function³⁰, while the helices of multiple-spanning membrane proteins (**Fig. 1.1B**) allow conformational changes necessary for opening/closing of ion channels, solute transport and transduction of extracellular/intracellular signals. The rigid β -barrel organization (**Fig. 1.1C**) instead often limits the function of these membrane proteins to receptor and enzyme purposes³¹. On the contrary, peripheral membrane proteins are entirely exposed at both sides of the cell surface interacting with the membrane mainly in three possible ways: (i) through their hydrophobic patch (**Fig. 1.1D**), (ii) by covalent fatty acid linkage (**Fig. 1.1E**) or (iii) by interacting with other membrane proteins (**Fig. 1.1F**)³². The function of these proteins is related to how they interact with the membrane. In particular, it was observed that the cumulative effects of curvature and lipid compositions of biological membranes are responsible for directing the recruitment of many peripheral proteins to specific cellular organelles³³. Indeed, peripheral proteins involved in various functions (e.g. lipid transfer or nuclear pore formation) seem to detect the packing

defects derived from lipid compositions and interact with positively curved membranes by the insertion of long acyl chains³⁴⁻³⁷.

The hydrophobicity, as well as the instability and flexibility, of membrane proteins leads to some extra challenges at all levels of experimental investigation, including protein preparation (expression, purification, solubilization) as well as detection on their native conditions (i.e., membranes composed by different lipid species). These experimental issues have so far limited the number of membrane protein structures solved to date (only 1.7% of the total PDB) hampering a detailed molecular understanding of membrane proteins and their interaction with lipids. Recently, a new crystallographic approach has emerged in the field of structural biology. Solvent contrast modulation, a technique widely used in small angle neutron and X-ray scattering³⁸, was used for resolving the structure of Ca²⁺-ATPase along with its first layer of phospholipids³⁹. The X-ray solvent contrast modulation appears as a promising technique for overcoming crystallographic limitations and obtaining an atomistic representation of membrane-protein interactions, helping also for an accurate setup of initial conditions for molecular simulations.

Due to their essential role in several diseases membrane proteins represent the preferred molecular target for over 40% of all FDA approved drugs⁴⁰. Therefore, the creation of robust molecular-detail models is essential for better predicting the function of membrane proteins, in particular the perturbations from their static representation usually provided by canonical structural biology approaches. Importantly, molecular models of this kind can help to understand and rationalize the actual influence of the physiological membrane environment and the effect of post-translational modifications, just to name few key ingredients that might affect significantly protein function.

1.2 POST-TRANSLATIONAL MODIFICATIONS: HOW PROTEINS DIVERSIFY THEMSELVES

(Adapted from: "Protein post-translational modifications: in silico prediction tools and molecular modeling", M Audagnotto and M Dal Peraro, Computational and Structural Biotechnology Journal, 2017, 15: 307-319)

Post-translational modifications (PTMs) introduce another layer of complexity in biological function within the cell. PTMs consist in a covalent modification of amino acids of the primary protein sequence⁴¹ and have the effect to create a much larger array of possible protein species. In response to specific physiological requirements, PTMs play a crucial role in regulating many biological functions⁴², such as protein localization in the cell^{43,44}, protein stability⁴⁵, and regulation of enzymatic activity⁴⁶. To date more than 90'000 individual PTMs have been detected using biochemical and biophysical analyses⁴⁷. In particular, it was observed that almost 5% of the human genome encodes enzymes in charge of catalyzing reactions leading to PTMs⁴⁸, highlighting once more the importance of these chemical modifications of the proteome.

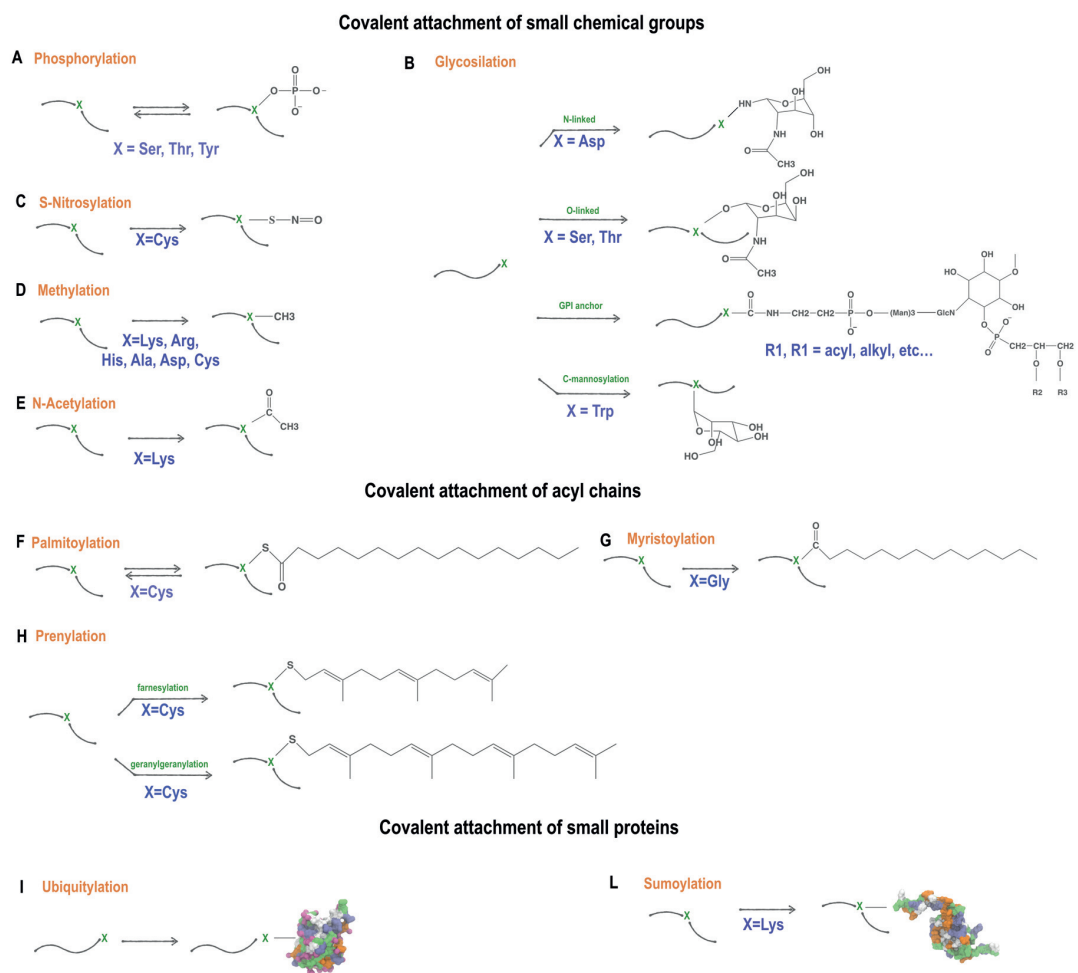


Figure 1.2. Schematic representation of PTMs discussed in this paragraph.

Generally, PTMs can be classified based on the covalent attachment of (i) small chemical groups (**Fig. 1.2A-E**), (ii) lipids (**Fig. 1.2F-G**) or (iii) small proteins (**Fig. 1.2I-L**) to the main peptide chain. In particular, enzymes often are responsible for regulating these chemical modifications in proteins, as in the case of phosphorylation, acetylation, methylation, carboxylation or hydroxylation⁴⁹. For instance, protein kinases can phosphorylate a given protein target to induce a signaling cascade, while this PTM can be further removed by specific protein phosphatases. These enzymes are found indeed in important signaling pathways, like G-protein^{49,50} and Wnt signaling^{51,52}. On the other side, PTMs not induced by specific enzymes (e.g., carbonylation or oxidations) were observed to be responsible of non-specific protein damage involved in neurodegenerative diseases, cancer and diabetes⁵³⁻⁵⁵.

Among the classes of PTMs, protein lipidation is a unique post-translational modification, which has the result to directly control the interaction of soluble protein with biological membranes affecting in turn cellular organization and trafficking. Among lipidations, palmitoylation is the only reversible modification (**Figure 1.3**). It consists in the attachment of a 16-carbon acyl chain to cysteine residues via a thioesteric bond^{56,57}. Palmitoylation can dynamically regulate protein function, as in the case of H-Ras and N-Ras^{43,58}. Two families of enzymes regulate the palmitoylation/depalmitoylation process: palmitoyltransferases (PATs), which catalyze the attachment of a palmitate from CoA to specific cysteines, and acyl protein thioesterases (APT_s), which remove the palmitate acyl chain. Palmitoylation occurs both in soluble and membrane proteins playing a critical role in the regulation of key biological processes, such as protein membrane trafficking, signaling, cell growth and development⁵⁹. Aberrant palmitoylation is associated to a variety of human diseases including neurological disorders (e.g., Huntington disease's⁶⁰ or Alzheimer's disease⁶¹) and cancer⁶²⁻⁶⁷. Mass spectrometry has allowed the identification of several palmitoylated proteins in cells and tissues, which can be further experimentally characterized using acyl biotin exchange (ABE) or acyl resin assisted capture (Acyl-RAC) techniques⁶⁸. Metabolic labeling and click chemistry probes^{69,70} were developed to recognize palmitoylation sites in order to shed light on the molecular mechanism and dynamics of palmitoylation. All these experimental techniques are time consuming and expensive, thus computer-aided methods are attractive alternatives for predicting palmitoylation sites.

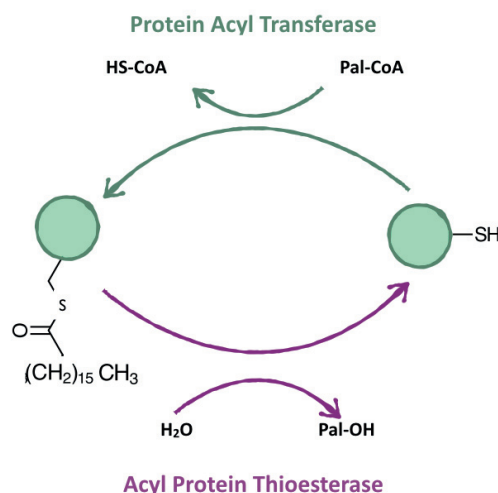


Figure 1.3. Palmitoylation-depalmitoylation cycle. This reversible process is regulated by the opposing action of DHHC protein acyl transferases, which catalyze the palmitoylation step and acyl protein thioesterases, as APT1, which are responsible for the depalmitoylation of the substrates.

1.3 MOLECULAR SIMULATIONS: A COMPUTATIONAL MICROSCOPE

Up to date, most of the current knowledge on biomolecular functional recognition result from *in vitro* characterization where crowding transient interactions and anisotropic environments (e.g., membrane) are not properly estimated at a molecular level, mainly due to technical issues. In the last sixty years, the accuracy of the description of biological environments increased exponentially: thanks to the development of advanced techniques that cover several layers of resolution (e.g., X-ray crystallography, NMR and cryo-electron microscopy) coupled with mass spectrometry (MS)-based high-throughput proteomics techniques we can now have better insights into the organization of large protein complexes as, for instance, the proteasome⁷¹, the ribosome^{72,73}, and the nuclear pore complex⁷⁴, not only in isolation but also within their cellular environment. However, the increment in system size is always coupled to the increment in system complexity. Indeed, for many years, the main biological components of cells, i.e. proteins and nucleic acids, were studied individually, neglecting other important cellular components such as carbohydrates, lipids and small metabolites, as well as specific and non-specific interactions with other protein partners.

Nowadays, the new challenge is to combine together structural, biophysical and biochemical data into more dynamic and realistic models for studying proteins in their physiological environments. In addition to the range of experimental structural and biophysical techniques, molecular modeling and molecular dynamics (MD) simulations have become complementary and powerful tools able to fill the resolution gap in the study of proteins behaviors in complex cellular environments^{36,75-77}.

In particular, thanks to the advances in the computational power and force fields development we are now able to reach the level of complexity necessary to describe the biological processes with a very high level of accuracy. This approach turned out to be fundamental in membrane modeling where it is required to zoom at different levels of resolution to cover the wide range of time and length scales associated with membrane biophysics³⁶. Indeed, compared to experimental microscopy techniques, the MD approach has in principle a much higher resolution that can reach atomistic resolution (1 Å) or even higher if electronic degrees of freedom are also considered in quantum mechanics approaches, for instance. The resolution of this ‘computational microscopy’⁷⁸ can be tuned allowing to “zoom in” at atomistic resolution and examine detailed interactions of membrane proteins with water, ions and lipids, as well as to “zoom out” to lower resolution using coarse grained (CG) approaches in order to achieve longer timescale (up to milliseconds) and size scale (up to 100 nm³), while unfortunately losing some chemical details in modeling interatomic interactions. Moreover, this computational microscope permits in general for the precise control of the environment and exact, at least in principle, reproducibility compared to real experiments and/or cellular conditions.

The ongoing advances in software and hardware for simulating complex systems enable to reach today sizes of several hundreds of nm, allowing direct comparisons with cell membrane imaging by cryo-electron tomography and super resolution optical microscopy. On the other side, force fields improvements have allowed approaching another important aspect for cellular functioning, namely the role of post-translational modifications (PTMs) for proteins structure and stability. The study of these effects was hampered by the labile transient nature of most of these modifications and by the lack of experimental techniques able to detect and characterize them. Therefore, molecular modeling and especially MD simulations, which are based on empirical atomistic force fields, represent a powerful way to investigate their role with molecular resolution. The recent availability of more

comprehensive experimental data along with accurate force field parameters have thus allowed to investigate protein properties taking into account also PTMs in the global representation of the physiological protein environments⁷⁹⁻⁸¹.

The diversity of models for lipids and proteins and the accuracy with which the interactions are described have demonstrated the efficacy of multiscale modeling and simulation approaches and the possibility for extension to more complex membrane systems. For instance, the advances in lipidomics analyses have provided a much more detailed view of membrane composition, allowing the construction of more realistic membrane models^{82,83} to better investigate membrane biophysical properties and their interplay with integral and peripheral membrane proteins^{84,85}. Several computational tools have been developed with this aim in mind, such as CHARMM-GUI⁸⁶ and LipidBuilder⁸⁷. Along the same lines, it is also clear that protein PTMs are another important layer of complexity that integrally defines and modulates protein function and, for this reason, needs to be considered at all levels of both experimental and computational investigation. Therefore, the computational advances of bioinformatics and physics-based molecular modeling and simulation techniques appear as a fundamental requirement to complement the experimental investigation of PTMs' impact on cellular processes.

1.4 OBJECTIVES OF THE THESIS

One of the current challenges in biology is to develop realistic models of the cellular organization, in particular the architecture of membranes and internal organelles. The integration of structural, biophysical and biochemical data with multiscale modeling and simulation approaches represents a valuable way for studying the mechanism of proteins in their biological environments. During my thesis, I combined experimental and computational techniques to investigate the effects of the realistic environment on the function of membrane proteins. In particular, I was interested by two specific systems: (i) the amyloid precursor protein (APP) and its interactions at the synaptic plasma membrane, and (ii) the human acyl-protein thioesterase 1 (hAPT1) and its role in protein depalmitoylation cycle across the Golgi and plasma membrane.

(i) The amyloid precursor protein and its interplay with the synaptic plasma membrane and the amyloid

APP is a type-I transmembrane glycoprotein that is present in the synaptic plasma membrane (SPM). Its cleavage by γ -secretase produces β -amyloids ($A\beta$), which are major components of the plaques observed in patients affected by Alzheimer's disease (AD). In the last years, APP has been proposed to have both monomeric and homodimeric structures. In particular, two possible dimerization interfaces were proposed: namely, $G_{700}XXXG_{704}XXXG_{708}$ and $G_{709}XXXA_{713}$. The latter, as suggested by NMR and EPR experiments, is the cholesterol binding motif key for the stabilization of APP transmembrane domain (APP-TM). However, up to date only the structure of the $G_{709}XXXA_{713}$ dimer is available on the Protein Data Bank (PDB: 2LZ3) so it is difficult to discriminate between the roles of these two motifs in the structural and dynamic properties of APP. Experimentally it was observed that the APP-TM homodimer state could be influenced by the detergent used during the NMR experiments, the specific composition of the biological membrane as well as the cholesterol concentration. Our molecular dynamics simulations of the APP embedded in a model of a realistic SPM revealed that the $G_{700}XXXG_{704}XXXG_{708}$ is the most stable species and likely the most biological relevant while the highly unsaturated lipids, that constitute the SPM, perturb the $G_{709}XXXA_{713}$ dimerization motif observed by NMR. A detailed analysis of the cholesterol distribution around the dimerization motifs showed that, independently from the dimer, both motifs interact with the cholesterol. This shed light on the regulation of cholesterol turnover in synaptic function explaining the important interaction with sterol regulatory element binding protein 1 (SREBP1) (Chapter 3).

Another important aspect in the study of AD is the role of γ -secretase in APP processing and $A\beta$ formation. γ -secretase is a membrane-embedded protease that cleaves single transmembrane helical domains of various integral membrane proteins. The enzyme is composed by four domains namely presenilin (PS1 and PS2), presenilin enhancer 2 (PEN-2), nicastrin (NCT), and anterior pharynx-defective 1 (APH-1). The catalytic center is located on the PS domain and excluded from the external surface of the enzyme. Of particular interest is the sequential cleavage of the amyloid precursor protein (APP) into variable length peptides ($A\beta_{40}$ and $A\beta_{42}$). Recently, a 4.2 Å

resolution cryo-EM structure (PDB ID: 5FN2) was published shedding light on the atomistic domains organization. Biochemical experiments showed that this process can occur at position unique from the active site suggesting that several residues in PS1 are involved in substrate binding. Despite the new insights from these studies, neither the mechanism of cleavage is fully understood nor a co-crystal structure of γ -secretase in complex with any substrate has been solved. Moreover, results from previous studies suggested that the relatively small amount of active γ -secretase complex present at the cell surface could be residing in raft membrane domain. Therefore, based on these finding I investigated the interplay between APP and γ -secretase when embedded in a realistic model of the SPM. Compared to previous studies, the membrane lipid composition was increased in complexity considering sphingolipids as well as glycerophospholipids for a total of 32 lipid species, which differ for head types and degree of unsaturation. This new membrane model allowed us to monitor the formation of lipid rafts, which have been experimentally observed to be made up of cholesterol and saturated lipids. Through long time coarse-grained MD simulations, lipid-raft domains appeared to be localized around the NCT domain, one of the regions observed to be in charge for APP recruitment. Although the formation of G₇₀₀XXXG₇₀₄XXXG₇₀₈ dimer was also observed in this SPM model, the APP interacted with the enzyme always as a monomer, suggesting that APP might be more likely processed as a monomer by γ -secretase. Moreover, the high affinity of the cholesterol showed for both γ -secretase and APP proposed its key role in enhancing the γ -secretase-APP binding and consequently A β production. Altogether our findings suggest new insights into γ -secretase processing, providing future directions for AD therapeutic development.

(ii) The catalytic and inhibition mechanism of the human acyl protein thioesterase I

Post-translational modifications play a crucial role in regulating the function of many biological molecules. Lipid modifications, in particular, via prenylation, acylation, myristoylation and palmitoylation help to direct membrane localization, protein-protein interactions, cell signaling, subcellular trafficking and vesicle transport. Among them, palmitoylation is the only reversible process. This modification consists in the addition of a palmitic acid forming a thioester bond with

cysteine residues of both soluble and membrane proteins. The thioester linkage is hydrolyzed by acyl protein thioesterase enzymes (APTs), which belong to the α/β The first soluble thioesterase to be characterized was the human APT1, which catalyzes depalmitoylation of signaling regulators like G α and H/N-Ras. This structure revealed a non-symmetric homodimeric organization where the catalytic pocket is occluded by the dimer interface.

In collaboration with the van der Goot lab at EPFL, I have solved new X-ray structures for wild-type human APT1, as well for the catalytic inactive form (S119A) and the palmitoylation-deficient mutant (C2S). These structures revealed a dimeric interface that suggests a novel mode of interaction with the biological membrane and substrates. Moreover, size-exclusion chromatography followed by multi-angle laser light scattering (SEC-MALLS) experiments combined with molecular dynamics simulations suggested that the biological dimeric form is rather weak favoring a possible monomeric structure as an active enzyme. Our studies further revealed a novel pocket adjacent to the APT1 catalytic site able to accommodate a palmitic acid. This result was confirmed with ethanol extractions of its ligand followed by ultra-performance liquid chromatography (UPLC) quadrupole-time-of-flight mass-spectrometry (Q-TOF-MS). This new evidence allowed us to better understand the catalytic and inhibition mechanism with molecular modeling and simulations, and showed the existence of a previously uncharacterized druggable pocket.

The work described in this thesis is consistent with the new emerging trend in biophysics, where the main aim is to model cellular environments in a comprehensive and realistic fashion. Indeed, the study of the effects of lipids composition and cholesterol distribution, as well as the role of PTMs on protein function and stability both remark the importance of modeling realistic and therefore complex biological environments. Nowadays, with the development of high-resolution and high-throughput experimental techniques and the parallel increment in computer power and modeling accuracy we are very well positioned to reach the level of complexity necessary for a thorough understanding of the biological processes that characterize life.

BIBLIOGRAPHY

1. Van Meer, G. et al. Membrane lipids: where they are and how they behave. *Nature reviews Molecular cell biology* **9**, 112-124 (2008).
2. Spiegel, S. et al. Signal transduction through lipid second messengers. *Current opinion in cell biology* **8**, 159-167 (1996).
3. Hurley, J.H. et al. Membrane budding. *Cell* **143**, 875-887 (2010).
4. Gautreau, A. et al. Function and regulation of the endosomal fusion and fission machineries. *Cold Spring Harbor perspectives in biology* **6**, a016832 (2014).
5. Atilla-Gokcumen, G.E. et al. Dividing Cells Regulate Their Lipid Composition and Localization. *Cell* **156**, 428-439 (2014).
6. Huijbregts, R.P. et al. Lipid metabolism and regulation of membrane trafficking. *Traffic* **1**, 195-202 (2000).
7. Clarke, A. et al. Lipid biochemistry and reproductive biology in two species of Gammaridae (Crustacea: Amphipoda). *Marine Biology* **88**, 247-263 (1985).
8. Sud, M. et al. Lmsd: lipid maps structure database. *Nucleic acids research* **35**, D527-D532 (2007).
9. van Meer, G. Cellular lipidomics. *The EMBO journal* **24**, 3159-3165 (2005).
10. Wenk, M.R. The emerging field of lipidomics. *Nature reviews Drug discovery* **4**, 594-610 (2005).
11. Shevchenko, A. et al. Lipidomics: coming to grips with lipid diversity. *Nature reviews Molecular cell biology* **11**, 593-598 (2010).
12. Slotte, J.P. Biological functions of sphingomyelins. *Progress in lipid research* **52**, 424-437 (2013).
13. Wiederschain, G.Y. Essentials of glycobiology. *Biochemistry (Moscow)* **74**, 1056-1056 (2009).
14. Maxfield, F.R. et al. Cholesterol, the central lipid of mammalian cells. *Current opinion in cell biology* **22**, 422-429 (2010).
15. Axelsson, B.-O. et al. Determination of ergosterol in organic dust by gas chromatography-mass spectrometry. *Journal of Chromatography B: Biomedical Sciences and Applications* **666**, 77-84 (1995).
16. Matile, P. et al. Yeast cytology. *The yeasts* **1**, 219-302 (1969).
17. Marquardt, D. et al. Asymmetric lipid membranes: towards more realistic model systems. *Membranes* **5**, 180-196 (2015).
18. Hancock, J.F. Lipid rafts: contentious only from simplistic standpoints. *Nature reviews. Molecular cell biology* **7**, 456 (2006).
19. Lingwood, D. et al. Lipid rafts as a membrane-organizing principle. *science* **327**, 46-50 (2010).
20. Simons, K. et al. Revitalizing membrane rafts: new tools and insights. *Nature reviews. Molecular cell biology* **11**, 688 (2010).
21. Marsh, D. Liquid-ordered phases induced by cholesterol: a compendium of binary phase diagrams. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1798**, 688-699 (2010).
22. Putzel, G.G. et al. Phenomenological model and phase behavior of saturated and unsaturated lipids and cholesterol. *Biophysical journal* **95**, 4756-4762 (2008).
23. Simons, K. et al. Functional rafts in cell membranes. *Nature* **387**, 569 (1997).
24. Edidin, M. The state of lipid rafts: from model membranes to cells. *Annual review of biophysics and biomolecular structure* **32**, 257-283 (2003).
25. Munro, S. Lipid rafts: elusive or illusive? *Cell* **115**, 377-388 (2003).

26. Liszewski, K. Dissecting the Structure of Membrane Proteins. (2015).
27. Krogh, A. et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**, 567-580 (2001).
28. Morell, P. et al. Characteristic composition of myelin. *Basic Neurochemistry: Molecular, Cellular and Medical Aspects* (1999).
29. Alberts B, J.A., Lewis J, et al. *Molecular Biology of the Cell.*, (2002).
30. Hubert, P. et al. Single-spanning transmembrane domains in cell growth and cell-cell interactions: More than meets the eye? *Cell adhesion & migration* **4**, 313-324 (2010).
31. Alberts B, J.A., Lewis J, et al. Membrane Proteins. in *Molecular Biology of the Cell*, (Garland Science, New York, 2002).
32. Branden, C.I. *Introduction to protein structure*, (Garland Science, 1999).
33. Vanni, S. et al. A sub-nanometre view of how membrane curvature and composition modulate lipid packing and protein recruitment. *Nature communications* **5**(2014).
34. Bigay, J. et al. Curvature, lipid packing, and electrostatics of membrane organelles: defining cellular territories in determining specificity. *Developmental cell* **23**, 886-895 (2012).
35. Barrett, P.J. et al. The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol. *Science* **336**, 1168-1171 (2012).
36. Wahrle, S. et al. Cholesterol-dependent γ -secretase activity in buoyant cholesterol-rich membrane microdomains. *Neurobiology of disease* **9**, 11-23 (2002).
37. Vanni, S. et al. Amphipathic lipid packing sensor motifs: probing bilayer defects with hydrophobic residues. *Biophysical journal* **104**, 575-584 (2013).
38. Pebay-Peyroula, E. *Biophysical analysis of membrane proteins: investigating structure and function*, (John Wiley & Sons, 2008).
39. Norimatsu, Y. et al. Protein-phospholipid interplay revealed with crystals of a calcium pump. *Nature* (2017).
40. Overington, J.P. et al. How many drug targets are there? *Nature reviews Drug discovery* **5**, 993-996 (2006).
41. Walsh, C. *Posttranslational modification of proteins: expanding nature's inventory*, (Roberts and Company Publishers, 2006).
42. Walsh, C.T. et al. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angewandte Chemie International Edition* **44**, 7342-7372 (2005).
43. Rocks, O. et al. An acylation cycle regulates localization and activity of palmitoylated Ras isoforms. *Science* **307**, 1746-1752 (2005).
44. Fairbank, M. et al. RING finger palmitoylation of the endoplasmic reticulum Gp78 E3 ubiquitin ligase. *FEBS letters* **586**, 2488-2493 (2012).
45. Maeda, A. et al. Palmitoylation stabilizes unliganded rod opsin. *Proceedings of the National Academy of Sciences* **107**, 8428-8433 (2010).
46. Hunter, T. Tyrosine phosphorylation: thirty years and counting. *Current opinion in cell biology* **21**, 140-146 (2009).
47. Khoury, G.A. et al. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific reports* **1**, 90 (2011).

48. Venne, A.S. et al. An improved workflow for quantitative Ne Ntion statistics: frequency analysis and curation of the shaFRADIC) to study proteolytic events in Arabidopsis thaliana. *Proteomics* **15**, 2458-2469 (2015).
49. Premont, R. et al. Protein kinases that phosphorylate activated G protein-coupled receptors. *The FASEB Journal* **9**, 175-182 (1995).
50. Gurevich, E.V. et al. G protein-coupled receptor kinases: more than just kinases and not only for GPCRs. *Pharmacology & therapeutics* **133**, 40-69 (2012).
51. Gao, C. et al. Regulation of Wnt/ β -catenin signaling by posttranslational modifications. *Cell & bioscience* **4**, 1 (2014).
52. Jiang, W. et al. Histone H3K27me3 demethylases KDM6A and KDM6B modulate definitive endoderm differentiation from human ESCs by regulating WNT signaling pathway. *Cell research* **23**, 122-130 (2013).
53. Ren, R.-J. et al. Proteomics of protein post-translational modifications implicated in neurodegeneration. *Translational neurodegeneration* **3**, 1 (2014).
54. Krueger, K.E. et al. Posttranslational protein modifications current implications for cancer detection, prevention, and therapeutics. *Molecular & Cellular Proteomics* **5**, 1799-1810 (2006).
55. McLaughlin, R.J. et al. Where, How, and When: Positioning Posttranslational Modification Within Type 1 Diabetes Pathogenesis. *Current diabetes reports* **16**, 1-9 (2016).
56. Emsley, P. et al. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126-2132 (2004).
57. Won, S.J. et al. Molecular mechanism for isoform-selective inhibition of acyl protein thioesterases 1 and 2 (APT1 and APT2). *ACS Chem Biol* (2016).
58. Cornilescu, C. et al. Structural analysis of the N-terminal domain of the human T-cell leukemia virus capsid protein. *Journal of molecular biology* **306**, 783-797 (2001).
59. Vartak, N. et al. The autodepalmitoylating activity of APT maintains the spatial organization of palmitoylated membrane proteins. *Biophysical journal* **106**, 93-105 (2014).
60. Sharon, M. et al. Alternative conformations of HIV-1 V3 loops mimic β hairpins in chemokines, suggesting a mechanism for coreceptor selectivity. *Structure* **11**, 225-236 (2003).
61. Ahearn, I.M. et al. FKBP12 binds to acylated H-ras and promotes depalmitoylation. *Molecular cell* **41**, 173-185 (2011).
62. Robinson, J.A. β -hairpin peptidomimetics: design, structures and biological activities. *Accounts of chemical research* **41**, 1278-1288 (2008).
63. Lo, S.C. et al. Structure of the Keap1: Nrf2 interface provides mechanistic insight into Nrf2 signaling. *The EMBO journal* **25**, 3605-3617 (2006).
64. Gavenonis, J. et al. Comprehensive analysis of loops at protein-protein interfaces for macrocycle design. *Nature chemical biology* **10**, 716-722 (2014).
65. de Jong, D.H. et al. Molecular view on protein sorting into liquid-ordered membrane domains mediated by gangliosides and lipid anchors. *Faraday Discuss* **161**, 347-363 (2013).
66. Dekker, F.J. et al. Small-molecule inhibition of APT1 affects Ras localization and signaling. *Nature chemical biology* **6**, 449-456 (2010).
67. Duncan, J.A. et al. A cytoplasmic acyl-protein thioesterase that removes palmitate from G protein α subunits and p21RAS. *Journal of Biological Chemistry* **273**, 15830-15837 (1998).

68. Forrester, M.T. et al. Site-specific analysis of protein S-acylation by resin-assisted capture. *Journal of lipid research* **52**, 393-398 (2011).
69. Bovigny, C. et al. LipidBuilder: a framework to build realistic models for biological membranes. (ACS Publications, 2015).
70. Bulacu, M. et al. Improved angle potentials for coarse-grained molecular dynamics simulations. *Journal of Chemical Theory and Computation* **9**, 3282-3292 (2013).
71. Kurimoto, E. et al. Crystal structure of human proteasome assembly chaperone PAC4 involved in proteasome formation. *Protein Science* **26**, 1080-1085 (2017).
72. Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **108**, 557-572 (2002).
73. Bai, X.-c. et al. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* **2**, e00461 (2013).
74. Beck, M. et al. The nuclear pore complex: understanding its function through structural insight. *Nature Reviews Molecular Cell Biology* **18**, 73-89 (2017).
75. Botan, A. et al. Toward atomistic resolution structure of phosphatidylcholine headgroup and glycerol backbone at different ambient conditions. *The Journal of Physical Chemistry B* **119**, 15075-15088 (2015).
76. Vogel, A. et al. Headgroup conformations of phospholipids from molecular dynamics simulation: sampling challenges and comparison to experiment. *The Journal of membrane biology* **245**, 23-28 (2012).
77. Pan, J. et al. The molecular structure of a phosphatidylserine bilayer determined by scattering and molecular dynamics simulations. *Soft Matter* **10**, 3716-3725 (2014).
78. Lee, E.H. et al. Discovery through the computational microscope. *Structure* **17**, 1295-1306 (2009).
79. Li, H. et al. Aggregation of lipid-anchored full-length H-Ras in lipid bilayers: simulations with the MARTINI force field. *PloS one* **8**, e71018 (2013).
80. Chiappori, F. et al. A novel molecular dynamics approach to evaluate the effect of phosphorylation on multimeric protein interface: the α B-Crystallin case study. *BMC bioinformatics* **17**, 57 (2016).
81. Olausson, B.E. et al. Molecular dynamics simulations reveal specific interactions of post-translational palmitoyl modifications with rhodopsin in membranes. *Journal of the American Chemical Society* **134**, 4324-4331 (2012).
82. van Eerden, F.J. et al. Characterization of thylakoid lipid membranes from cyanobacteria and higher plants by molecular dynamics simulations. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1848**, 1319-1330 (2015).
83. Ingólfsson, H.I. et al. Lipid Organization of the Plasma Membrane. *Journal of the American Chemical Society* **136**, 14554-14559 (2014).
84. Audagnotto, M. et al. Effect of the Synaptic Plasma Membrane on the Stability of the Amyloid Precursor Protein Homodimer. *The Journal of Physical Chemistry Letters* **7**, 3572-3578 (2016).
85. van Eerden, F.J. et al. Molecular Dynamics of Photosystem II Embedded in the Thylakoid Membrane. *The Journal of Physical Chemistry B* (2016).
86. Jo, S. et al. CHARMM - GUI: a web - based graphical user interface for CHARMM. *Journal of computational chemistry* **29**, 1859-1865 (2008).
87. Bovigny, C. et al. LipidBuilder: A Framework To Build Realistic Models for Biological Membranes. *Journal of Chemical Information and Modeling* **55**, 2491-2499 (2015).

2 Molecular modeling and simulation methods

Biological relevant processes as well as the majority of chemical reactions occur in solution environments at specific pH and ionic conditions. Molecular models are therefore intimately related with describing solute-solvent interactions and their molecular properties. Due to the dynamic nature of the molecules, a dynamic approach is necessary in order to explain the wide range of thermally accessible states of the systems and thereby connect structure and function.

In the last years, we have seen remarkable improvements in the field of molecular modeling and simulation applied to the study of biological systems. In particular, the advance in the theoretical description of complex systems was possible thanks to the increased computational resources as high-performance computing, which have become available to the mass at commodity price. For instance, with the development of faster and bigger parallel architectures the accessible time and length scales have increased at the point that was possible to simulate the HIV capsid unit (1.5 M of atoms) for 1.2 μs ¹. Moreover, algorithms and force field developments have also impacted the field significantly. A multi-scale description of biological systems has allowed reaching an unparalleled level of detail, which, combined with the advances in experimental techniques, has allowed moving further in the study of proteins' behavior. The dynamic approach provided by molecular simulations allows to bypass the static view provided for instance by techniques such as X-ray crystallography, which, although extremely valuable, is still insufficient to fully understand biological activity.

While during this thesis, I also used several experimental techniques (directly or in collaboration with other laboratories), which will be explained further in the relative chapters, my work has been mainly driven by computational modeling and

simulation, thus that I will present in this chapter the main theoretical framework of the computational techniques applied during this project.

2.1 MOLECULAR MECHANICS

Molecular Mechanics (MM) force fields are based on a classical approximation of the intrinsic quantum mechanical energy potential describing interactions within biomolecular systems. This approximation has the virtue of decreasing the cost of simulations on large systems by orders of magnitude, while preserving a reasonable accuracy. This classical treatment of the potential energy allows dealing with the many-body problem represented by the study of complex biological systems, which are generally characterized by thousand to millions of atoms. Force fields based approaches calculate the energy of the system as a function of the nuclear positions. They are based on the assumption of transferability, so that a set of parameters developed and tested on small systems for a small number of cases can be applied to a wide range of problems. In this approach, the electronic energy is written as a parametric function of the nuclear coordinates and the parameters are fitted to experimental or higher level computational data. In MM, the quantum aspects of the nucleus are neglected and the dynamics of the atoms is treated by classical mechanics.

The interactions between a system composed by N particles is describe by a potential energy U , written as a sum of bonded and non-bonded interactions, each describing the energy required for perturbing specific degrees of freedom of a given biomolecule:

$$\begin{aligned}
 U(\mathbf{r}^N) &= U_{bonded}(\mathbf{R}^N) + U_{non-bonded}(\mathbf{R}^N) \tag{1} \\
 U_{bonded}(\mathbf{r}^N) &= \sum_i^{bonds} U_{bond}(R_i) + \sum_i^{angles} U_{angle}(\theta_i) \\
 &+ \sum_i^{torsions} U_{torsion}(\phi_i) + \sum_i^{impropers} U_{impropers}(v_i)
 \end{aligned}$$

$$U_{non-bonded}(R^N) = U_{vdW}(r_{ij}) + \sum_i^N \sum_j^N U_{Coulomb}(r_{ij})$$

By this energy function (Eq. 1), the geometries and relative energies can be calculated, where stable conformations are represented by the minima of $U(r^N)$, while the conformational transitions can be described by detecting transition structures on the potential surface energy.

2.2 BONDED INTERACTIONS

The bonding potential, $U_{bonded}(R^N)$, describes the intramolecular bonding contributions considering 4 types of bonded interactions: bond-stretching terms ($U_{bond}(R_i)$), angle bending terms ($U_{angle}(\theta_i)$), torsional terms ($U_{torsion}(\phi_i)$) and improper torsional angles ($U_{torsion}(\phi_i)$).

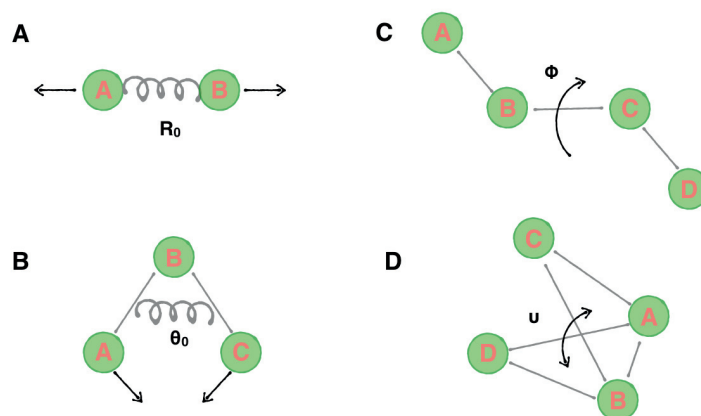


Figure 1.2 Bonded interactions. Bonded contributions of the potential energy function in the molecular mechanics approximation. The molecule are treating as consisting of balls (the atoms) connected by springs (the bonds). (A) Bond stretching, (B) Angle bending, (C) Torsional angle and (D) Improper torsional angle.

Bond stretching

$U_{bond}(R_i)$ is the energy necessary for stretching a bond between two atom types A and B (**Fig. 1.2A**). It is usually written as a Taylor expansion truncated at the second order as deviation around an equilibrium bond length, R_0 :

$$U_{bond}(R^{AB} - R_0^{AB}) = U(0) + \frac{dE}{dR}(R^{AB} - R_0^{AB}) + \frac{1}{2} \frac{d^2E}{dR^2}(R^{AB} - R_0^{AB})^2 \quad (2)$$

The derivatives are evaluated at $R = R_0$ and $U(0)$ term is normally set to zero since it is the zero point for the energy scale. When the expansion is around the equilibrium value the second term is zero. Thus, in its simplest form, the bond stretching energy can be written as:

$$U_{bond}(R_i) = \frac{1}{2} k^{AB} (R - R_0)^2 \quad (3)$$

where k^{AB} is the force constant for the AB bond. Hence, the potential energy varies with the square of the displacement from the equilibrium position R_0 (Hooke's law). The complex interplay between the various components in the force field may well deviate slightly from its reference value in order to compensate for other contributions to the energy. It is also important to remember that "real" molecules undergo vibrational motion, even at absolute zero degree.

Angle bending

$U_{angle}(\theta_i)$ is the energy required for bending an angle formed by three atoms A-B-C characterized by a bond between A and B, and B and C (**Fig. 1.2B**). As for the $U_{bond}(r_i)$ also $U_{angle}(\theta_i)$ is usually expanded as a Taylor series around the equilibrium angle θ_0 and terminating at the second order:

$$U_{angle}(\theta^{ABC} - \theta_0^{ABC}) = \frac{1}{2} k^{ABC} (\theta - \theta_0)^2 \quad (4)$$

The energy required for distorting an angle away from equilibrium is less than to stretch or compress a bond and the force constants are usually proportionately smaller. Indeed, for the same deviation from the equilibrium (e.g. 10%) the difference between the U_{angle} and U_{bond} is roughly 5 k_BT (at 300 K). As in the case of bond-stretching terms, it is possible to improve the accuracy of a force field by the incorporation of higher order terms.

Torsional angles

$U_{torsion}(\phi_i)$ is very important among the intramolecular terms of a force field. Indeed, most of the variation in structural and relative energies is due to the interplay between the torsional and non-bonded contributions: both the van der Waals and electrostatic contributes (see paragraph 2.3) to the rotational barriers affecting the torsional energy.

This potential describes the contribution from each bonded quartet of atoms A-B-C-D in the systems (**Fig. 1.2C**). Compared to $U_{bond}(R_i)$ and $U_{angle}(\theta_i)$, the torsional potential $U_{torsion}(\phi_i)$ is not expanded as a Taylor series because the torsional angle can go far from equilibrium. A Fourier series is instead applied:

$$U_{torsion}(\phi^{ABCD}) = \sum_{n=1} k_d^{ABCD} [1 + \cos(n\phi - \phi_0)] \quad (5)$$

The multiplicity term n determines the periodicity of the function defining the minima of the system, for instance $n=1$ describes a rotation that is periodic by 360° , $n=2$ periodic by 180° , $n=3$ periodic by 120° , and so on. The k_d^{ABCD} determines the size of the barrier for rotation around the B-C bond, which is usually few kcal/mol at 300 K.

Improper torsional angles

For an A-B-C-D torsional angle characterized by a central sp^2 atom (e.g., cyclobutanone) the experimentally equilibrium conformation observed is a coplanar arrangement, which maximize the π -bonding energy (**Fig. 1.2D**). A force field that contains only the standard bond-stretching and angle-bending terms would provide a wrong conformation for this system placing the sp^2 atom out of the plane. Therefore, it is commonly necessary to consider an additional term to the force field to maintain the chirality or the planarity of molecules. The simplest way to realize this is to introduce a so-called improper torsional angle term, $U_{impropers}(v_i)$, which has usually the form of a harmonic function of the *out-of-plane* angle v_i :

$$U_{impropers}(v^{ABC} - v_0^{ABC}) = \frac{1}{2} k^{ABCD} (v - v_0)^2 \quad (6)$$

where k^{ABCD} is the force constant and v_0 is the phase.

2.3 NON-BONDED INTERACTIONS

Independent molecules, as well as atoms in the same molecule but not covalently bonded, experience non-bonded interactions. In force fields, this term is usually the sum of two contributions: the electrostatic ($U_{Coulomb}(r_{ij})$) and the van der Waals interactions ($U_{vdW}(r_{ij})$).

Electrostatic interactions

$U_{Coulomb}(r_{ij})$ describes the internal electron distribution, which defines the positive and negative parts of the molecules. The electrostatic interaction between two different charged point of the system is calculated as a sum of interactions between pairs of point charges, which are retrieved from quantum-mechanics calculations, using the Coulomb's law:

$$U_{Coulomb}(r_{ij}) = \sum_{i=1}^{N_A} \sum_{j>i}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (7)$$

where N_A and N_B are the total number of point charges (q_i and q_j) separated by a distance r_{ij} , while ϵ_0 is the dielectric permittivity of vacuum.

Van der Waals interactions

The van der Waals potential, $U_{vdW}(r_{ij})$, describes the repulsion or attraction between non-bonded atoms. This contribute is zero at infinite distance while becomes very repulsive for short distances passing through a minimum at around 2.5-4 Å. This potential term arises from a balance between attractive and repulsive forces. The attractive contribution is due to dispersive London forces, which come from the generation of instantaneous dipoles in biomolecules immersed in an electric field, while the repulsive forces take into account the Pauli principle by which electrons can not overlap.

A popular function that describes these general requirements is the Lennard-Jones (LJ) potential, which was used in the earliest studies of the properties of liquid argon^{2,3}:

$$U_{vdW}(r_{ij}) = \sum_{pairs} 4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \quad (8)$$

This function contains two adjustable parameters: (i) the collision diameters σ , i.e. the separation for which the energy is zero, and (ii) the depth ε . The attractive part varies as r^{-6} and the repulsive part as r^{-12} , which is found to be quite reasonable for rare gas phase, but too steep for other systems. However, the 6-12 LJ potential is broadly used given it is a reasonable approximation that is also fast to compute for big systems, since r^{-12} can be rapidly obtained by squaring the r^{-6} term,

The determination of the van der Waals parameters for a system containing N different types of atoms is a difficult and time-consuming process. Therefore, it is usually assumed that the parameters for the cross interactions can be obtained from the parameters of the pure atoms using the Lorentz-Berthelot mixing rules:

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \quad (9)$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii} \varepsilon_{jj}} \quad (10)$$

van der Waals interactions are a case of short range interactions, which decay at large r like $1/r^6$. Therefore, the computational cost for the van der Waals interactions is reduced considering only the pairwise interaction within a cut-off distance r_c and a switching-function is applied on $U_{vdW}(r_{ij})$ to avoid a discontinuity in the potential. Typically values of cut-off are between 10 and 12 Å.

2.4 COARSE-GRAINED REPRESENTATION OF BIOMOLECULES

Coarse-grained (CG) representations of complex biomolecular systems permit to explore biological processes (e.g. protein folding mechanisms⁴⁻⁶, membrane self-assembly⁷ or protein-membrane interaction^{8,9}), which are usually precluded at atomistic level. Usually, compared to an atomistic representation, a coarse-grained description unwinds the energy landscape helping to avoid local energy minima traps. Moreover, CG affects the balance between enthalpy and entropy of a modeled system. Indeed, the reduction of the degrees of freedom affects the entropy of the system,

which is compensated by reducing the enthalpic term. Therefore, CG simulations could accurately reproduce free energy differences, but the enthalpic and entropic contribution may be inaccurate.

The typical energy functional form of the CG resembles the all-atom potential energy function:

$$V^{CG} = V_{bonds}^{CG} + V_{bendings}^{CG} + V_{dihedrals}^{CG} + V_{vdw}^{CG} + V_{ele}^{CG} \quad (11)$$

where:

$$V_{bonds}^{CG} = \sum_{bonds} k_{ij} (b - b_0)^2 \quad (12)$$

describes the harmonic oscillation of two consecutive beads between the equilibrium value b_0 with a force constant k_{ij} ;

$$V_{bendings}^{CG} = \sum_{bendings} k_{ijk} (\omega - \omega_0)^n \quad (13)$$

is used to describe pseudo-bending angles among three beads with ω_0 equilibrium value and k_{ijk} constant force;

$$V_{dihedrals}^{CG} = \sum_{dihedral} k_{ijkl} [1 + \cos(n\chi - \chi_0)] \quad (14)$$

defines the pseudo-dihedral having χ_0 equilibrium value, multiplicity n and k_{ijkl} force constants.

The non-bonded interactions are also generally defined by two contributions:

$$V_{vdw}^{CG} = \sum_{pairs} 4\sigma_{ij} \left[\left(\frac{\lambda_{ij}}{r_{ij}} \right)^m - \left(\frac{\lambda_{ij}}{r_{ij}} \right)^6 \right] \quad (15)$$

that is the generalized Lennard-Jones potential having an exponent m between 12 and 8. This repulsive part is used to mimic the apparent softness of coarse-grained beads.

The other contribution is a simple Coulombic potential used for describing electrostatic interaction between beads of charge Q_i and Q_j . Sometimes, a relative

dielectric constant ϵ_r , is used in order to screen charge-charge interaction due to the coarse representation of the solvent

$$V_{ele}^{CG} = \sum_{pairs} \frac{Q_i Q_j}{\epsilon_0 \epsilon_r r_{ij}} \quad (16)$$

In this thesis, I used the well-known MARTINI force field, which originally was developed for lipids^{10,11} and furthermore extended to peptides, proteins^{12,13} and other small biomolecules^{13,14}. The granularity of the MARTINI force field is based on a one-to-four mapping where one bead is roughly composed by four heavy atoms including associated hydrogens. The chemical nature of the atoms is properly reproduced by the definition of four main types of CG particles: polar (P), nonpolar (N), apolar (C) and charged (Q). These four types of beads are additionally divided based on hydrogen-bonding capabilities (donor, acceptor, both or none) and polarity (from 1= low to 5=high polarity) providing a total of 18 unique fundamental units. This approach allows a simple mapping procedure, which generate a unified set of parameters and topologies for different biological systems. Nowadays, MARTINI provides a vast selection of parameters not only for lipids, but also for sterols, sugars, polymers and peptides allowing a straightforward modeling of complex systems. However, its simplistic backbone representation as well as the nature of the protein coarse-graining hampered the changes in protein secondary structure and therefore the study of folding proteins, for instance.

2.5 COMPUTATIONAL TOOLS FOR SETTING UP COMPLEX BIOLOGICAL SYSTEMS

Several different tools were used in order to build up the realistic biological systems studied in this thesis.

Lipidbuilder: how to set up your favourite membrane at atomistic resolution

*LipidBuilder*¹⁵ is a useful tool to model complex biological membranes. This software recently developed in our laboratory enables the creation of templates and topology files centered on the CHARMM force field. *LipidBuilder* is an online platform

(<http://lipidbuilder.epfl.ch/home>) that, based on SMILE structure of the acyl chains, allows the creation of the corresponding phospholipid. The lipid's topology is created from a built-in library of structure and acyl chains. In particular, the acyl chains have been parametrized exploiting the “plug and play” philosophy developed in the CHARMM force field. QM-based parameters are used to describe better the cyclic moiety observed in different class of bacterial membranes phospholipids. The psfgen algorithm and the generated topology file were combined to build the PSF and PDB files for the membrane patch. With this tool it is possible to built and store biologically relevant models characterized by acyl tails of heterogeneous compositions.

Insane: lipidomics at the coarse-grained level

*Insane*¹⁶ is a standalone program, which allows an on-the-fly generation of coarse-grained lipid types by specifying their headgroup, linker, and lipid tails on the command line. As for *LipidBuilder*¹⁵, this tool permits the generation of membranes of arbitrary composition and therefore it can be used for exploring properties and interactions of lipids *in silico*. In addition, it is possible to insert proteins in the bilayer creating simultaneously the membrane-protein system. Although you can in principle create your own lipids, *Insane* does not provide a graphical interface as *LipidBuilder*, but a basic knowledge of Python scripting language is required in order to create a new species.

How to deal with post-translational modifications

(Adapted from: “Protein post-translational modifications: *in silico* prediction tools and molecular modeling”, M. Audagnotto and M. Dal Peraro, *Computational and Structural Biotechnology Journal*, 2017, 15: 307-319)

Despite the important role played by PTMs, their structural and dynamic effects on protein function remain poorly understood from a molecular point of view, due to the labile transient nature of most of these modifications and the lack of adequate experimental techniques able to detect and characterize them and the underlying chemical mechanism of formation. There are several online tools recently summarized¹⁷ that allow to overcome some of this experimental limitations and

predict putative PTMed sites, but they do not usually provide any information about the impact of post-translational modifications from a mechanistic point of view.

Molecular modeling and molecular simulation (such as molecular dynamics, MD), which are based on empirical atomistic force fields¹⁸⁻²¹, are a powerful strategy for studying biological systems at single-molecule resolution and nanosecond-to-millisecond time scales. Although this computational approach allows nowadays the study of protein processes and properties that are not easily accessible experimentally, there are still some apparent limitations regarding the availability of accurate parameters that would allow the investigation of PTM proteins. In the past years, several improvements have been made in order to expand this approach also to non-standard biomolecules. Within the AMBER force field²² atomic charges and parameters were developed for phosphorylated residues as phosphoserine, phosphothreonine, phosphotyrosine, phosphohistidine²², and S-nitrosylated residues (S-nitrocysteine²³) and methylation (trimethyllysine^{24,25}). Similarly, within the CHARMM force field there are parameters for methylated lysines and arginines, as well as acetylated lysines and palmitoylated cysteines²⁶. There are also ad hoc comprehensive atomistic force field parameters for treating the description of the link between carbohydrates and proteins such as in GLYCAM for AMBER²⁷/CHARMM²⁸ and modified version of GROMOS^{29,30}. In theory, within these schemes, there are existing strategies to parameterize virtually any kind of non-standard amino acids, as for the case of PTMs; in practice, the development of new force field models always involves time-consuming parameterization protocols and rigorous a posteriori validations of the quality and robustness of the new models.

Moving to lower resolution, coarse-grained force fields can be also very useful to study the impact of PTMs on protein function. In this domain there are no specific parameters for the description of PTMed residues. The MARTINI force field¹⁴, for instance, provides parameters for treating non-covalently bound sugars molecules or phosphate groups but a complete general representation of modified residues is not yet available. However, a recent work described new parameters for modeling palmitoylated cysteines³¹ that were used to study H-Ras, and contributed to show the influence of this PTM in regulating the partition of the protein with the membrane (*parameters used for the study of human acyl protein thioesterase described in Chapter 4*).

Answering to the need of new and better molecular models to more realistically describe proteins, some automatic tools for generating force field parameters for new molecular species have become available, such as ParaChem or SwissParam³² compatible with the CHARMM force field, q4md-forcefieldtools for AMBER³³ and ATB for GROMOS³⁴. However, none of them directly focus on the parameterization of PTMs, likely because of the complexity of parameters development required for most PTMs. Therefore, the necessity of having computational tools allowing an automatic parameterization of PTMed protein structures to be used in MD simulations resulted in the development of some new web-servers, such as FF_PTM (<http://selene.princeton.edu/FFPTM/>) and Vienna-PTM (<http://vienna-ptm.univie.ac.at>). FF_PTM focuses on expanding the existing AMBER force field (i.e., ff03) including parameters for 32 PTMs. In particular, it is characterized by parameters that describe the attachment of small molecules (e.g., phosphorylation, methylation or acetylation) and the covalent interaction with acyl chains such as palmitic acid (palmitoylation) and geranylgeranyl pyrophosphate (geranylgeranylation). On the other hand, Vienna-PTM is a web platform designed for introducing PTMs on PDB structures to run simulations using the GROMOS 54A7 and 45A3 force fields.

2.6 CLASSICAL MOLECULAR DYNAMICS

Classical molecular dynamics (MD) aims at predicting the temporal evolution of a molecular system using a molecular mechanics based potential. The integration of the Newton's law of motion ($\mathbf{F} = m\mathbf{a}$) results in a trajectory that denotes how the positions and velocities of the particles in the system vary with time. Among the several algorithms for integrating the equations of motion, the Verlet algorithm^{35,36} is probably the most widely used in molecular dynamics simulations. It retrieves the new positions $\mathbf{r}(t + \delta t)$ from the positions and acceleration at time t and $t - \delta t$:

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) \quad (17)$$

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) \quad (18)$$

Adding these two equations gives:

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \delta t^2 \mathbf{a}(t) \quad (19)$$

The velocity is not explicitly treated in the Verlet integration algorithm but it can be calculated in several ways. For instance, a simple approach is to divide the difference in positions at times $t + \delta t$ and $t - \delta t$ by $2\delta t$:

$$\mathbf{v}(t) = \frac{[\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)]}{2\delta t} \quad (20)$$

A critical parameter in the integration procedure is the choice of the timestep, which is typically one order of magnitude smaller than the fastest event of the system. This is a quite severe restriction since these high-frequency motions (e.g. the stretching vibrations of hydrogen atoms, e.g. the C-H bond stretching is on the order of 10 fs and therefore the time step in normal conditions should be about 0.5-1 fs) have usually less impact on many of the system properties. Therefore, one solution to this problem is to ‘freeze’ all such vibrations, which allows for a longer timestep and consequently longer simulation times at the same computational cost. This is normally done by either the SHAKE or RATTLE algorithms³⁷ that incorporate the distance constraints on bond lengths typically allows the timestep to be increased by a factor of 2 or 3.

The most time-consuming part of molecular dynamics simulations is the electrostatic calculation. Indeed, in the treatment of the electrostatic interactions the atomic charges are independent and the total charge of the system may be simply represented as the sum of individual atom-atom electrostatic interactions. Therefore, for a model of N particles, $N(N-1)/2$ pair interactions have to be calculated. Thus, it is necessary to truncate the evaluation of these long-range interactions introducing a cut-off distance. For highly polar systems the choice of the cut-off radius is critical since an entry/exit of charge in the exclusion area could bring to significant error. As first approximation, a big cut-off could be applied to avoid this error but the cost of computation will increase with a complexity of $O(N^2)$, where N is the number of particle system.

2.7 ERGODIC HYPHOTHESIS

The idea behind MD is that we can study the average behavior of many-body systems solving numerically the equation of motion and averaging the quantity of interest over a sufficient long time. Based on statistical mechanics, the average value of an observable ϕ is defined as an ensemble average:

$$\bar{\phi} = \int_p^N \int_x^N \phi(x, p) f(x, p; t) dx dp \quad (21)$$

where $f(x, p; t)$ indicates the probability of finding the phase point representative of its dynamical state (x, p) in the volume $dx dp$. In MD simulation, for a perfect isolated system of N atoms, in a volume V , at a constant total energy E (i.e., in the microcanonical ensemble) the time-average of observable ϕ is:

$$\phi_\tau = \frac{1}{\tau} \int_0^\tau \phi[p(p^0, x^0, t), x(p^0, x^0, t)] dt \quad (22)$$

For a τ sufficiently long we can assume that the time average ϕ_τ becomes independent of τ :

$$\phi_\tau(t_0) = \int \phi_\tau(x^0, p^0) f(x^0, p^0; t_0) dx^0 dp^0 \quad (23)$$

where $f(x^0, p^0; t_0)$ is the distribution of function at time t_0 . Eq. 23 shows that in general ϕ_τ depends on the time t_0 at which the time average starts and the necessary and sufficient condition for ϕ_τ to be independent of time is that the distribution function f be independent of time. Thus, $\left(\frac{\partial f}{\partial t}\right) = 0$

From the Liouville's theorem we know that:

$$f(p, q; t_0 + s) = f(p^0, q^0; t_0) \quad (24)$$

since f is independent from time $\left(\frac{\partial f}{\partial t}\right) = 0$, we can rewrite eq. 24 as:

$$f(p, x) = f(p^0, x^0) \quad (25)$$

Combining the eq.18 with eq.19 and based on the assumption of eq.25, we obtain:

$$\overline{\phi_\tau} = \int f(x^0, p^0; t_0) \left\{ \frac{1}{\tau} \int_0^\tau \phi[p(p^0, x^0, t), x(p^0, x^0, t)] dt \right\} dx^0 dp^0 \quad (26)$$

$$\overline{\phi_\tau} = \int f(p, q) \left\{ \frac{1}{\tau} \int_0^\tau \phi[p, x] dt \right\} dx dp \quad (27)$$

Interchanging the order of integration we obtain that the time average is equal to the ensemble average (*Ergodic hypothesis*):

$$\overline{\phi_\tau} = \frac{1}{\tau} \int_0^\tau [\int f(p, q) \phi[p, x] dx dp] dt = \frac{1}{\tau} \int_0^\tau \overline{\phi} dt = \overline{\phi} \quad (28)$$

since the ensemble average of ϕ ($\overline{\phi}$) is independent of time.

This assumption is at the basis of MD simulations and consents to calculate the thermodynamic properties from the trajectories. Although the trajectory in MD are in principles never long enough, the ergodic hypothesis is satisfied if the time scale explored by simulations is longer than the specific observable under investigation, practically speaking if the sampling time is at least one order of magnitude longer than the event of interest.

2.8 PERIODIC BOUNDARY CONDITIONS

An additional important aspect in MD simulations is the concept of boundary conditions and how to treat the edges of the simulated systems in order to avoid surface effects during the measuring of the systems properties. There are different options, but the most common one is to use a rectangular box with periodic boundary conditions (PBC). PBCs mimic the presence of an infinite bulk surrounding a N-particle model system by virtually replicating the simulation box in every direction. This means that when an atom leaves the unit cell crossing the boundary, an image enters to replace it in order to conserve the total number of particles. However, the PBCs do not eliminate the effects of a finite box size and it can influence properties as diffusion, bilayer undulation and lipid domain dynamics. Therefore, it is crucial to set

systems that are large enough to avoid artifacts due to the perturbations of the potential of mean force and consequent over-stabilization of the system in exam.

PBCs can be used in conjunction with Particle Mesh Ewald (PME)^{38,39}, a method for computing long-range electrostatic interactions. This method has a complexity of $O(N\log N)$ and requires that the system is embedded in a periodic boundary conditions in order to suppress the finite sample size surface effects.

2.9 MD SIMULATIONS IN DIFFERENT ENSEMBLES

In MD simulations, like in real experiments, it is possible to choose under which thermodynamic conditions to perform the *in silico* experiments selecting different *ensembles*. The concept of *ensemble* derives from statistical mechanics where it is used for identifying a collection of N macroscopic replicas of the identical thermodynamic system. Among the possible ensembles proposed, the canonical and the isothermal-isobaric ensemble are the most used in MD. In a canonical ensemble (also denoted as NVT ensemble) the number of particles (N) is kept constant, the box size is fixed (V) and external thermostat controls the temperature (T). In particular, the control of the temperature is a crucial issue in any molecular dynamics simulations. Therefore, several methods were proposed to fix the temperature during the simulations. For instance, the differential thermostat introduced by Woodcock⁴⁰ proposed to scale the velocities to p_i :

$$p_i = \sqrt{T_0/T} p_i \quad (29)$$

where T_0 is the reference temperature and T is the actual one obtained from the velocity of the particles. However, this method leads to discontinuities in the momentum part of the phase space trajectory due to rescaling procedure. Therefore, in alternative to this approach the deviation to the actual temperature T is corrected by multiplying the velocities of a certain factor λ . This method allows the fluctuation of the temperature rather than fixing it to constant value as in the case of the differential control. The Berendsen thermostat is an example of proportional thermostat where the momenta p_i are modified as λp_i with λ :

$$\lambda = \left[1 + \frac{\delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right) \right]^{1/2} \quad (30)$$

The constant τ_T is called coupling time constant that determines the time scale on which the desired temperature is reached. In particular, for $\tau_T = \delta t$, the fluctuation in the kinetic energy vanish and the configurational part of the phase space density has the canonical form:

$$\rho(\mathbf{q}, \mathbf{p}) = \delta(T - T_0) e^{-\beta U(\mathbf{q})} \quad (31)$$

Another class of thermostats is stochastic, where all or a subset of the degrees of freedom of the system are subject to collision with virtual particles. Based on the Langevin stochastic differential equation, the motion of a particle due to the thermal agitation of a bath is describe as:

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial U}{\partial \mathbf{q}_i} - \gamma \mathbf{p}_i + \mathbf{F}^+ \quad (32)$$

where γ is a friction constant and \mathbf{F}^+ a Gaussian random force which amplitude is determined by the second fluctuation dissipation theorem:

$$\langle \mathbf{F}_i^+(t_1) \mathbf{F}_j^+(t_2) \rangle = 2\gamma k_B T \delta_{ij} \delta(t_1 - t_2) \quad (33)$$

Additional degrees of freedom into the Hamiltonian of the system are introduced by the method invented by Nosé⁴¹. The idea proposed is to reduce the effect of an external system acting as heat reservoir, to one additional degree of freedom. The interaction between the thermal bath and the particles in the systems results in a change of the particles velocity. Two different variables were introduced: (i) the real and the (ii) virtual ones. The resulting Hamiltonian in virtual coordinates is defined as following:

$$H^* = \sum_{i=1}^N \frac{\pi_i^2}{2m_i s^2} + U(\mathbf{p}) + \frac{\pi_s^2}{2M_s} + g k_B T \ln s \quad (34)$$

where g is the $3N + 1$ degree of freedom for a system of N particles. This Hamiltonian actually represents the probability density in phase space for a canonical ensemble.

In this thesis, the Bussi ⁴² was applied for controlling the temperature during the simulations. This method is an extension of the Berendsen thermostat where a properly constructed random force is applied in order to force the correct distribution for the kinetic energy. Bussi thermostat proposed a different way for calculating the rescaling factor: instead of forcing the kinetic energy to be exactly equal to the target temperature $\bar{K} = \frac{N_f}{2\beta}$ (N_f is the number of degrees of freedom and β is the inverse temperature), a target value K_t is selected with a stochastic procedure aiming at obtaining the desired ensemble. Therefore, the factor α becomes:

$$\alpha = \sqrt{\frac{K_t}{K}} \quad (35)$$

Living systems are usually experiencing conditions of constant temperature and pressure, therefore in MD it is important to be able to describe this *ensemble* in order to have a direct comparison with the experimental observables. In the isothermal-isobaric ensemble (also called NPT) it is necessary to keep constant the temperature while allowing for volume variations. The most adopted barostat in MD is the Langevin approach, which is based on the simple model of a box coupled with a piston, which controls the pressure. Since the energy is transferred to or from the external system, the conserved quantity monitored during the simulation will be the enthalpy rather than the energy as in the case of NVT. In the proportional thermostat the variation in volume is considered as scaling of particles positions:

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\mathbf{p}_i}{m_i} + \alpha \mathbf{q}_i \quad (36)$$

Since a change in pressure is related to the isothermal compressibility K_T as

$$\dot{P} = -\frac{1}{K_T V} \frac{\partial V}{\partial t} = -\frac{3\alpha}{K_T} \quad (37)$$

that is approximated

$$\frac{(P_0 - P)}{\tau_P} = -\frac{3\alpha}{K_T} \quad (38)$$

Therefore the eq. 31 can be written as:

$$\frac{\partial q_i}{\partial t} = \frac{p_i}{m_i} - \frac{K_T}{3\tau_P} (P_0 - P) \quad (39)$$

which corresponds to a scaling of the box length sL and coordinates $s\mathbf{q}$ where

$$s = 1 - \frac{K_T \delta t}{3\tau_P} (P_0 - P) \quad (40)$$

The K_T value is $44.6 \cdot 10^{-6} \text{ bar}^{-1}$ for water and the reference pressure usually applied in the MD simulation is 1 atm, which usually corresponds to the pressure experience by the system *in vivo* and *in vitro* experiments.

BIBLIOGRAPHY

1. Perilla, J.R. et al. All-atom molecular dynamics of virus capsids as drug targets. *The journal of physical chemistry letters* **7**, 1836-1844 (2016).
2. Rahman, A. Correlations in the motion of atoms in liquid argon. *Physical Review* **136**, A405 (1964).
3. Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review* **159**, 98 (1967).
4. Sterpone, F. et al. Protein simulations in fluids: coupling the OPEP coarse-grained force field with hydrodynamics. *Journal of chemical theory and computation* **11**, 1843-1853 (2015).
5. Barducci, A. et al. Assessing the quality of the OPEP coarse-grained force field. *Journal of chemical theory and computation* **7**, 1928-1934 (2011).
6. Chebaro, Y. et al. The coarse-grained OPEP force field for non-amyloid and amyloid proteins. *The Journal of Physical Chemistry B* **116**, 8741-8752 (2012).
7. Marrink, S.J. et al. Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *Journal of the American Chemical Society* **125**, 15233-15242 (2003).
8. Sharma, S. et al. An atomistic model for assembly of transmembrane domain of T cell receptor complex. *Journal of the American Chemical Society* **135**, 2188-2197 (2013).
9. Carpenter, T. et al. Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel. *Biophysical journal* **95**, 3790-3801 (2008).
10. Marrink, S.J. et al. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B* **108**, 750-760 (2004).
11. Baron, R. et al. Comparison of Thermodynamic Properties of Coarse-Grained and Atomic-Level Simulation Models. *ChemPhysChem* **8**, 452-461 (2007).
12. de Jong, D.H. et al. Improved parameters for the martini coarse-grained protein force field. *Journal of Chemical Theory and Computation* **9**, 687-697 (2012).
13. Monticelli, L. et al. The MARTINI coarse-grained force field: extension to proteins. *Journal of chemical theory and computation* **4**, 819-834 (2008).

14. Marrink, S.J. et al. The MARTINI force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry B* **111**, 7812-7824 (2007).
15. Bovigny, C. et al. LipidBuilder: a framework to build realistic models for biological membranes. (ACS Publications, 2015).
16. Wassenaar, T.A. et al. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J Chem Theory Comput* **11**, 2144-2155 (2015).
17. Audagnotto, M. et al. Protein post-translational modifications: In silico prediction tools and molecular modeling. *Computational and Structural Biotechnology Journal* **15**, 307-319 (2017).
18. Cornell, W.D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117**, 5179-5197 (1995).
19. Wang, J. et al. Development and testing of a general amber force field. *Journal of computational chemistry* **25**, 1157-1174 (2004).
20. Oostenbrink, C. et al. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry* **25**, 1656-1676 (2004).
21. Kaminski, G.A. et al. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B* **105**, 6474-6487 (2001).
22. Homeyer, N. et al. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *Journal of molecular modeling* **12**, 281-289 (2006).
23. Han, S. Force field parameters for S-nitrosocysteine and molecular dynamics simulations of S-nitrosated thioredoxin. *Biochemical and biophysical research communications* **377**, 612-616 (2008).
24. Lu, Z. et al. Importance of charge independent effects in readout of the trimethyllysine mark by HP1 chromodomain. *Journal of the American Chemical Society* **131**, 14928-14931 (2009).
25. Machado, M.R. et al. Isoform-specific determinants in the HP1 binding to histone 3: insights from molecular simulations. *Amino acids* **38**, 1571-1581 (2010).
26. Grauffel, C. et al. Force field parameters for the simulation of modified histone tails. *Journal of computational chemistry* **31**, 2434-2451 (2010).
27. Woods, R.J. et al. Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. *The Journal of Physical Chemistry* **99**, 3832-3846 (1995).
28. Jo, S. et al. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry* **29**, 1859-1865 (2008).
29. Schmid, N. et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European biophysics journal* **40**, 843-856 (2011).
30. Soares, T. et al. Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme. *Journal of Biomolecular NMR* **30**, 407-422 (2004).
31. de Jong, D.H. et al. Molecular view on protein sorting into liquid-ordered membrane domains mediated by gangliosides and lipid anchors. *Faraday discussions* **161**, 347-363 (2013).

32. Zoete, V. et al. SwissParam: a fast force field generation tool for small organic molecules. *Journal of computational chemistry* **32**, 2359-2368 (2011).
33. Vanquelef, E. et al. RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic acids research* **39**, W511-W517 (2011).
34. Malde, A.K. et al. An automated force field topology builder (ATB) and repository: version 1.0. *Journal of chemical theory and computation* **7**, 4026-4037 (2011).
35. Frenkel, D. et al. *Understanding molecular simulation: from algorithms to applications*, (Academic press, 2001).
36. Grubmüller, H. et al. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation* **6**, 121-142 (1991).
37. Ryckaert, J.-P. et al. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).
38. Darden, T. et al. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics* **98**, 10089-10092 (1993).
39. Petersen, H.G. Accuracy and efficiency of the particle mesh Ewald method. *The Journal of chemical physics* **103**, 3668-3679 (1995).
40. Woodcock, L.-V. Isothermal molecular dynamics calculations for liquid salts. *Chemical Physics Letters* **10**, 257-261 (1971).
41. Shuichi, N. Constant temperature molecular dynamics methods. *Progress of Theoretical Physics Supplement* **103**, 1-46 (1991).
42. Bussi, G. et al. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **126**, 014101 (2007).

3 Amyloid precursor protein and γ -secretase: the two main characters of Alzheimer's disease

Described for the first time in 1907 by Alois Alzheimer, who observed pathological alterations in the brain of female patients, Alzheimer's disease (AD) has become one of the first causes of dementia among elderly people. It is an age related, chronic debilitating neurodegenerative condition characterized by progressive neuronal loss. In 2010, it was estimated that 5.3 million of people were affected by AD¹ in the US and more than 26 million patients worldwide with a projection for 2050 of 106 million new cases².

The hallmark of this disease was initially defined by Alzheimer as a "*peculiar substance*", which generates extracellular deposits in specific brain regions, nowadays called senile plaques. Almost 100 years later, the search for genetic connections brought to a major hint: missense mutations in the amyloid precursor protein (APP) transmembrane regions cause early-onset (familial) AD³⁻⁵. These results, together with the observation that amyloid-beta ($A\beta$) peptides form neurotoxic fibrils⁶⁻⁹ supported the view that the deposition of $A\beta$ leads to neuronal dysfunction and clinical manifestation of the disease (the amyloid pathway).

The APP is an integral membrane protein processed by several different secretases (**Fig. 3.1**). The β -secretase cleaves the N-terminal APP leaving the C-terminal embedded in the membrane. On the other side, α -secretase cuts within the $A\beta$ region producing soluble APP (sAPP α) and an 83-residue fragment embedded in the membrane. Both the membrane-embedded products of α - and β -secretase

cleavage are substrates for γ -secretase enzyme, which proteolysis produces a 3-kDa peptide (P3) and 4-kDa A β , respectively.

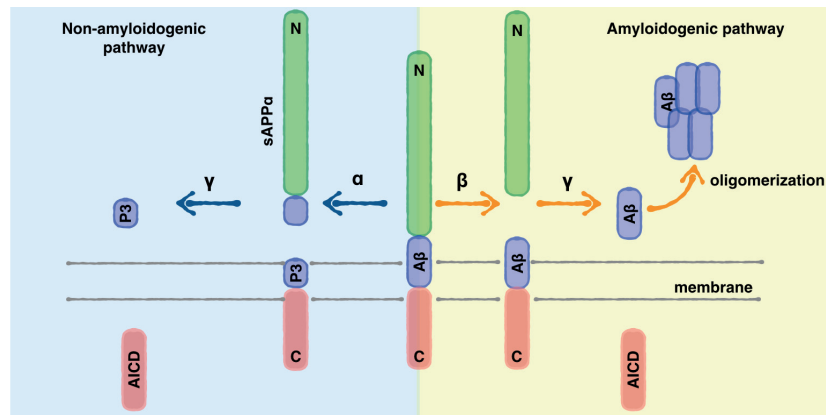


Figure 3.1 APP catalytic cleavage and the production of A β peptides. APP is a type I transmembrane protein that undergoes cleavage at either α - or β -secretase cleavage sites in order to release large ectodomains (sAPP α or sAPP β). The product of β -secretase cleavage is the membrane-embedded C-terminal fragments, which are substrates for γ -secretase. Proteolysis of C99 by γ -secretase promotes the release of A β while the proteolysis of C83 (product of α -secretase) generates an N-terminally truncated and non-toxic A β protein (P3).

The interplay between γ -secretase and APP is one of the critical issues in the field of AD. The recruitment mechanism and the consequent catalytic process, which leads to the formation of A β peptides, are still poorly understood at the molecular level and many are the open scientific questions. Recently, another important parameter emerges in the AD's equation: the role of membrane lipids composition in the function and stability of APP and γ -secretase.

3.1 THE EFFECTS OF LIPIDS COMPOSITION ON STABILITY AND FUNCTION OF AMYLOID PRECURSOR PROTEIN

*This chapter is adapted from the paper: "Effect of the Synaptic Plasma Membrane on the Stability of the Amyloid Precursor Protein Homodimer", Audagnotto, M., Lemmin, T., Barducci, A., & Dal Peraro, M., *The Journal of Physical Chemistry Letters*, 2016, 7.18: 3572-3578.*

The proteolytic cleavage of the transmembrane (TM) domain of the Amyloid Precursor Protein (APP) releases amyloid- β (A β) peptides, which accumulation in the brain tissue is an early indicator of Alzheimer's disease. We used multiscale molecular dynamics simulations to investigate the stability of APP-TM dimer in more realistic models of the synaptic plasma membrane (SPM). Between the two possible dimerization motifs proposed by NMR and EPR, namely G₇₀₉XXXXA₇₁₃ and G₇₀₀XXXG₇₀₄XXXG₇₀₈, our study revealed that the dimer promoted by the G₇₀₉XXXXA₇₁₃ motif is not stable in the SPM due to the competition with highly unsaturated lipids that constitute the SPM. Under the same conditions, the dimer promoted by the G₇₀₀XXXG₇₀₄XXXG₇₀₈ motif is instead the most stable species and likely the most biologically relevant. Regardless of the dimerization state, both these motifs can be involved in the recruitment of cholesterol molecules.

Introduction

The transmembrane (TM) 99-amino-acid long APP-C99 protein is a critical intermediate in Alzheimer's disease¹⁰. Its proteolytic cleavage by γ -secretase releases amyloid- β (A β) peptides, whose neurotoxicity is defined by the precise position of the cleavage site. Moreover, several studies reported that the trafficking and/or the activity of the membrane proteins, such as APP, BACE1 and presenilins, responsible to control the A β level are regulated by the neuronal lipid composition^{11,12}. It is therefore key to obtain a characterization of the APP transmembrane domain (hereafter called APP-TM). While the first models of APP-TM obtained by NMR suggested a monomeric conformation^{13,14} which is perturbed by the local membrane environment^{15,16}, subsequent solid-state NMR^{17,18} experiments and computational studies^{19,20} supported the existence of an APP-TM homodimeric state^{14,17,21-23}. Indeed, the APP-TM possesses two distinct possible dimerization motifs at opposite sides of its helical arrangement: G₇₀₀XXXG₇₀₄XXXG₇₀₈ and G₇₀₉XXXXA₇₁₃, while right and left-handed conformations have been observed with NMR spectroscopy only for the G₇₀₉XXXXA₇₁₃ APP-TM dimer^{23,24}. It has been observed that the variation in detergents used in structural biology experiments may affect helix stability and dimerization^{25,26}. Therefore, also the APP-TM homodimer state could likely be influenced by detergents used during NMR experiments, the specific composition of the biological membrane where it resides, and the concentration of cholesterol therein.

Results and discussion

Here, to investigate the effect of the membrane composition on the stability of the APP-TM dimers, we performed a set of molecular dynamics simulations at different molecular resolutions of APP in the synaptic plasma membrane (SPM). To this aim we first built a model of the SPM including the 9 major lipids species observed in male mice brain²⁷. Our SPM model, produced both at atomistic resolution using the CHARMM force field and at coarse-grained (CG) using the MARTINI force field, was thus composed of phosphatidylcholine (PC, counting for a 50% of the total lipid content), phosphatidylethanolamine (PE, 33%), negatively charged phosphatidylserine (PS, 17%), and cholesterol (up the remaining 40% of the SPM). The lipid tails included saturated palmitoleic, polyunsaturated arachidonic and docosahexaenoic acids (**Table 3.1**). Using these more complex models of the SPM, we studied APP-TM dimer assembly; within this approach the representation of the membrane environment is more faithful to native physiological conditions, on the other side current limitations affecting both force field accuracy and sampling to treat complex lipids mixtures at CG resolution might have to be taken into account²⁸⁻³⁰.

Table 3.1. Lipids distribution of the Synaptic Plasma Membrane models. The relative abundance of the different lipids species is given for both leaflets.

	young-SPM		old-SPM	
	% inner	% outer	% inner	% outer
POPC (16:0/18:1)	4 (14)	11 (8)	5 (12)	8 (9)
PAPC (16:0/20:4)	4 (13)	10 (8)	5 (10)	7 (9)
HDPC (16:0/22:6)	4 (14)	11 (8)	5 (12)	8 (8)
SOPC (18:0/18:1)	6 (16)	13 (12)	6 (15)	10 (11)
POPE (16:0/18:1)	4 (6)	9 (7)	4 (9)	7 (6)
SAPE (18:0/20:4)	2 (6)	5 (4)	3 (6)	4 (6)
SDPE (18:0/22:6)	5 (14)	11 (6)	5 (11)	9 (8)
ODPS (18:0/22:6)	5 (7)	6 (9)	6 (6)	5 (10)
SOPS (18:0/18:1)	1 (3)	2 (2)	2 (2)	1 (4)
CHOL	64 (28)	23 (133)	59 (62)	41 (111)

All the CG molecular dynamics (MD) simulations were performed within the DAFT approach that allows for the characterization of membrane helix interactions as previously done for GpA and polyleucine helices³¹. We set up two series of 500 MD

simulations, where two APP-TM monomers (built from 2LZ3²⁴ and initially separated by 35 Å) were inserted in a CG model of the SPM and the POPC bilayer. These two series run, respectively, for up to 500 ns and 1000 ns.

Time evolution of the non-bonded interaction energy between pairs of monomers (**Figure 3.2A-B**) showed that in 52% of the cases a dimer species is formed in SPM and POPC membrane models, in both cases representing a mixture of the 2 possible dimer states. In particular, we observed that in SPM monomers are readily formed, while in POPC more sampling is required before dimerization occurs.

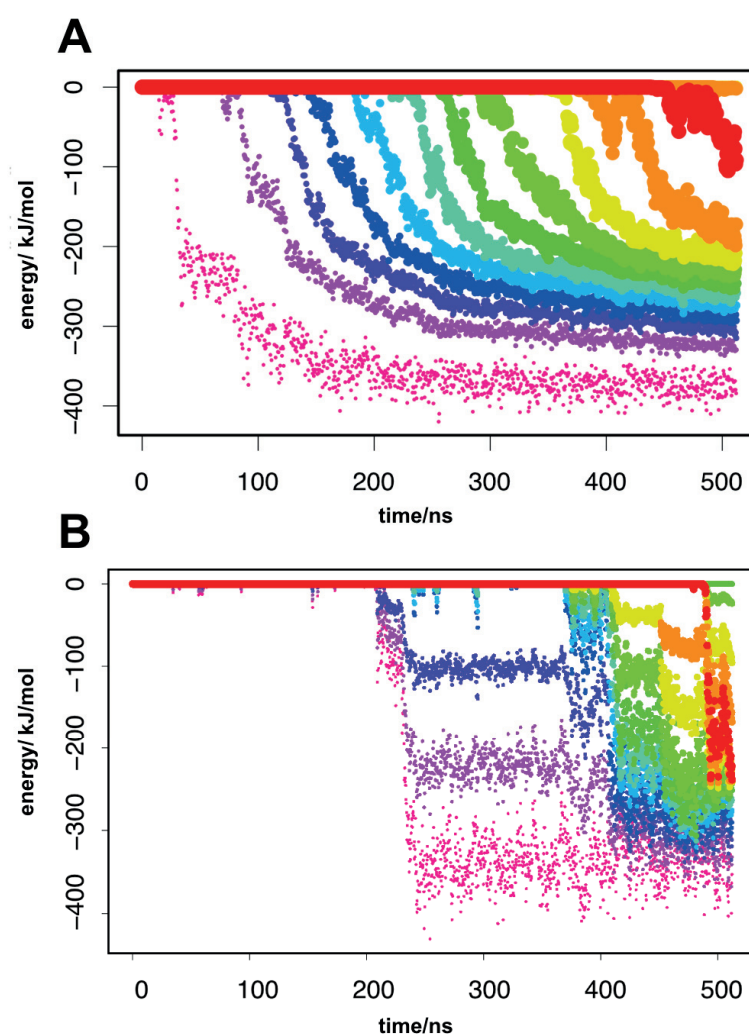


Figure 3.2. DAFT interaction energy between monomers during the MD time evolution. The graphs show the time evolution of the non-bonded interaction energy of the APP-TM monomers in **(A)** SPM and **(B)** POPC bilayer. At each time, the interaction energy is characterized by marking the 5% points (i.e., vigintile). In the case of SPM and POPC we observed that 11 out of 21 vigintile bands are below zero, indicating that in both systems 52% of the total simulations represent an interaction between the two monomers.

The structural characterization of the dimer ensemble based on the orientation of the residues with respect to the axis between the two helices (i.e., Crick angle) showed that in a pure POPC bilayer (**Figure 3.3A**) there is equilibrium slightly shifted toward the $G_{700}XXXG_{704}XXXG_{708}$ conformation (76% of the total dimeric population). Within the differences of the approach and analysis parameters, this result appears compatible with what reported by a previous computational work³² where these 2 dimer species can be selected within the local environment of POPC membrane model. On the contrary, in SPM the $G_{700}XXXG_{704}XXXG_{708}$ dimer resulted much more predominant over the alternative conformation (92%), hinting to some extra stabilization produced by the SPM environment (**Figure 3.3B-D**). Consistently, this result is in agreement with the dimeric arrangement proposed by solid-state NMR¹⁷ and by mutagenesis experiments^{21,22,33,34}.

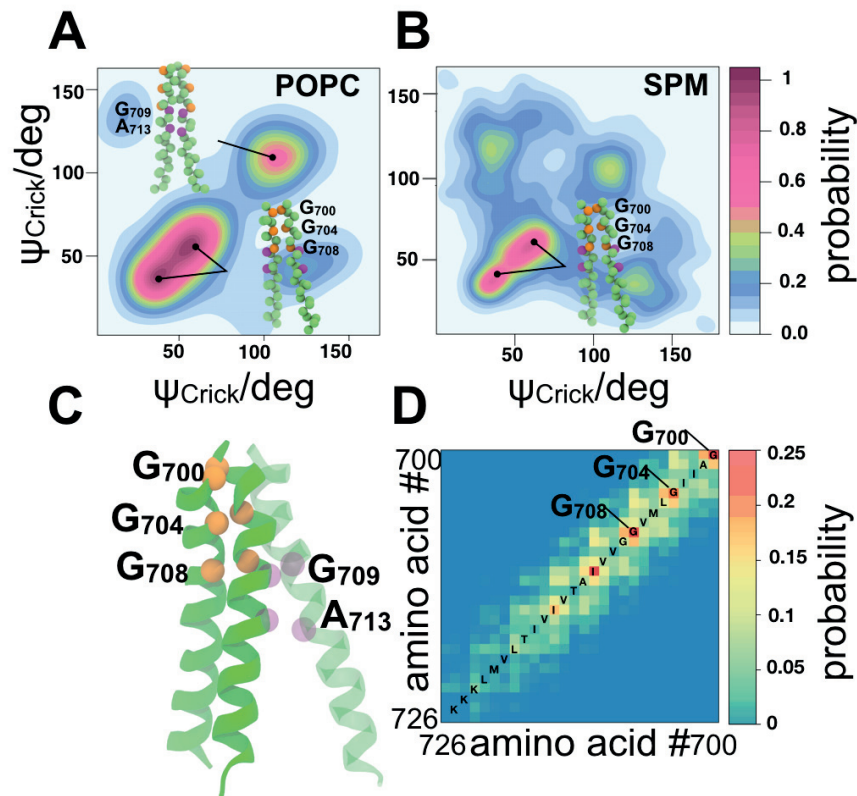


Figure 3.3. Influence of the membrane composition for the formation of the $G_{709}XXXA_{713}$ and $G_{700}XXXG_{704}XXXG_{708}$ dimer. The orientation of the residues with respect to the axis between the two APP-TM helices (Crick angle) in (A) the POPC bilayer and (B) the SPM model; (C) representative MD structure of the $G_{700}XXXG_{704}XXXG_{708}$ dimer superimposed to the $G_{709}XXXA_{713}$ dimer; (D) APP-TM contact map over time calculated for each pair of residues in the SPM during CG MD. In the SPM the bound state is confined to the $G_{700}XXXG_{704}XXXG_{708}$ dimeric conformation, while in the pure POPC bilayer is able to explore also the $G_{709}XXXA_{713}$ binding motif.

To further characterize the effects of membrane composition on the stability of the two APP-TM dimers, atomistic MD simulations were performed in SPM with different cholesterol concentrations. Indeed, it was observed that cholesterol made up the 40% of SPM composition. The ratio of cholesterol between the exo-leaflet and cyto-leaflet has been observed to change as a function of age³⁵. Of the total amount of cholesterol, young mice (3-4 months) were reported to have 15% in the exo-leaflet and 85% in the cyto-leaflet, while at a more mature stage (24-25 months) the exo/cyto-leaflet distribution becomes roughly equal³⁶. To capture these membrane features we prepared 2 atomistic SPM models (hereafter called young-SPM and old-SPM, Table S1) featuring different cholesterol distribution that mimics membrane aging. All these atomistic SPM models were assembled using *LipidBuilder*³⁷.

The NMR APP-TM dimer (2LZ3)²⁴ interacting through the G₇₀₉XXXA₇₁₃ motif (**Figure 3.4A**) was inserted in these membranes and the resulting systems were simulated with atomistic MD. 5 independent replicas with different randomized and previously equilibrated lipids conditions were simulated for each SPM model for a cumulative time of 3 μ s (MD simulations details are reported in SI). These MD simulations showed that the G₇₀₉XXXA₇₁₃ dimer, while stable in micelle (**Figure 3.4A**), is not stable in SPM models, regardless their cholesterol distribution. All atomistic MD replicas showed a preference for the monomeric conformation of G₇₀₉XXXA₇₁₃ APP-TM (**Figure 3.4B-D**), confirming the picture obtained by CG MD simulations (**Figure 3.4B**). The dimeric and monomeric states were discriminated based on their inter-helical distances between Gly709 (d_{GG}) and Ala713 residues (d_{AA}), which define the dimer interface. The experimental NMR structures of the dimer in micelles are characterized by $d_{GG} \approx 7$ Å and $d_{AA} \approx 5$ Å. The simulated homodimer distributions in old- and young-SPM rapidly diverged to $d_{GG} \approx 11$ Å and $d_{AA} \approx 10$ Å (**Figure 3.4C-D**), showing an invariant instability of the dimer. The dimer breaking events occurred spontaneously in MD on time scales of $\sim 10^2$ ns, suggesting that the sampling achieved with atomistic models is representative of the dimer-monomer equilibrium.

Since for the G₇₀₀XXXG₇₀₄XXXG₇₀₈ dimer no experimental structures were available, we obtain a representative conformation from our previous CG MD simulations clustering the APP-TM dimers in SPM based on principal component analysis (see SI). These models were back-mapped into atomistic structures, inserted into the young- and old-SPM bilayers and simulated by MD (with 5 replicas). In all the

simulations the dimer was stable (**Figure 3.3E-F**) keeping the same distances between Gly704 ($d_{GG(704)} \approx 5 \text{ \AA}$), and Gly708 ($d_{GG(708)} \approx 5 \text{ \AA}$) residues observed for CG APP-TM dimeric structure. Interestingly, cholesterol did not influence the structural stability of the $G_{709}XXXA_{713}$ and $G_{700}XXXG_{704}XXXG_{708}$ dimer, since dissociation events were indistinctively observed for MD simulations carried out in young- and old-SPM.

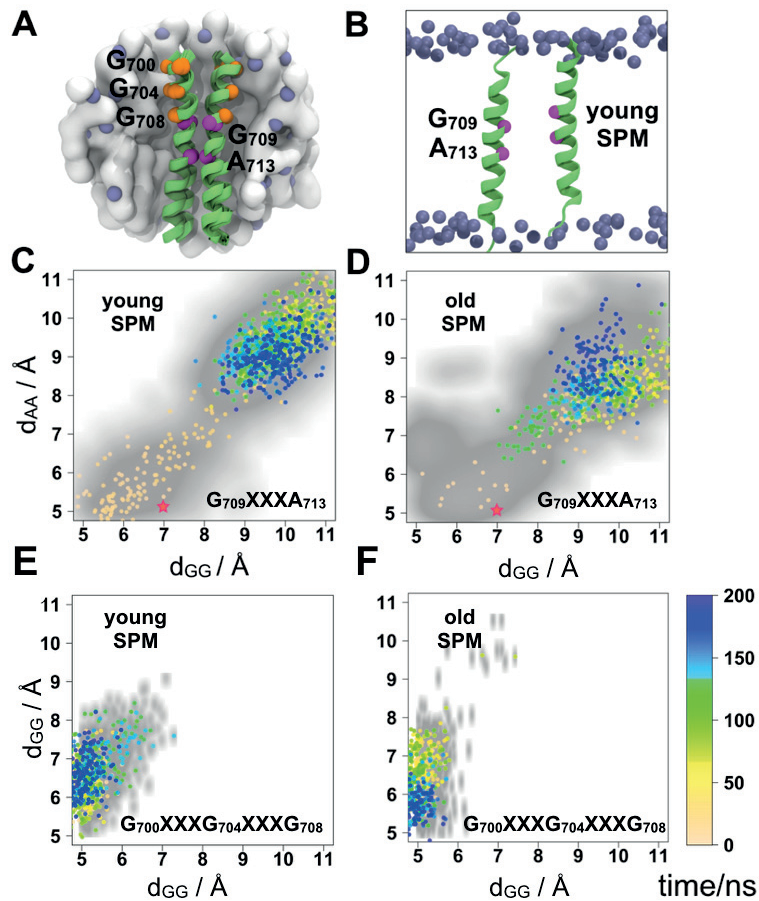


Figure 3.4. The $G_{709}XXXA_{713}$ and $G_{700}XXXG_{704}XXXG_{708}$ dimer in SPM. (A) APP-TM NMR ensemble (2LZ3) embedded in a dodecylphosphocholin micelle²⁴; (B) representative MD snapshot of the $G_{709}XXXA_{713}$ dimer dissociated in young-SPM; averaged $G_{709}XXXA_{713}$ APP-TM dimer evolution during MD in different membrane models: (C) young-SPM and (D) old-SPM; averaged $G_{700}XXXG_{704}XXXG_{708}$ APP-TM dimer evolution in (E) young-SPM and (F) old-SPM. The color scale represents the MD time evolution. The grey surface represents the global conformational space explored by each system, while the red star indicates the initial NMR values of d_{GG} and d_{AA} ²⁴.

In particular, analysis of the $G_{709}XXXA_{713}$ dimer in a SPM model without cholesterol confirmed these results, showing dissociation within the same timescale (**Figure 3.5**).

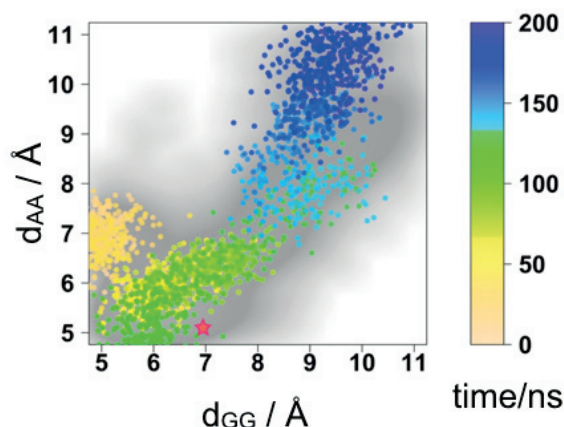


Figure 3.5. APP-TM dimer conformation in SPM without cholesterol. The color scale shows the time evolution. The grey surface represents the averaged conformational space explored by the dimer at different membrane lipids compositions in all the replicas. The red star is the value of d_{GG} and d_{AA} from the experimentally derived structure of APP-TM $G_{709}XXXA_{713}$ dimer (PDB: 2LZ3)²⁴. Note that the APP-TM in SPM without cholesterol explored the monomeric conformation characterized by $d_{GG} \approx 10 \text{ \AA}$ and $d_{AA} \approx 11 \text{ \AA}$.

In order to understand the effect of unsaturated lipids on the APP-TM dimer stability we measured the lipid-protein interactions for young-SPM as well as old-SPM. The averaged lipids distribution among all the simulated trials (**Figure 3.6A-B**) revealed that in all MD simulations, the lipids with high unsaturations compete to interact with the $G_{709}XXXA_{713}$ dimerization motif (**Figure 3.6B**), while unsaturated lipids tend to remain farther from the $G_{700}XXXG_{704}XXXG_{708}$ dimer (**Figures 3.6A**). In particular, unsaturated bonds at the end of the polyunsaturated lipids tail (for example in PAPC) interact with the $G_{709}XXXA_{713}$ region, disrupting the APP-TM dimer interface (**Figure 3.6C**). The presence of several highly unsaturated lipids thus increases the probability of breaking the APP-TM dimer. Indeed double bonds concurrently bordered by saturated bonds increase the acyl tail flexibility (i.e., lower rotational energy barrier³⁸⁻⁴⁰) allowing the polyunsaturated acyl chains to quickly adapt their conformation eventually interfering with the $G_{709}XXXA_{713}$ binding motif. We also observed that lipids responsible of breaking the dimeric interface mainly belong to the cyto-leaflet (**Figure 3.4C and 3.5C-D**).

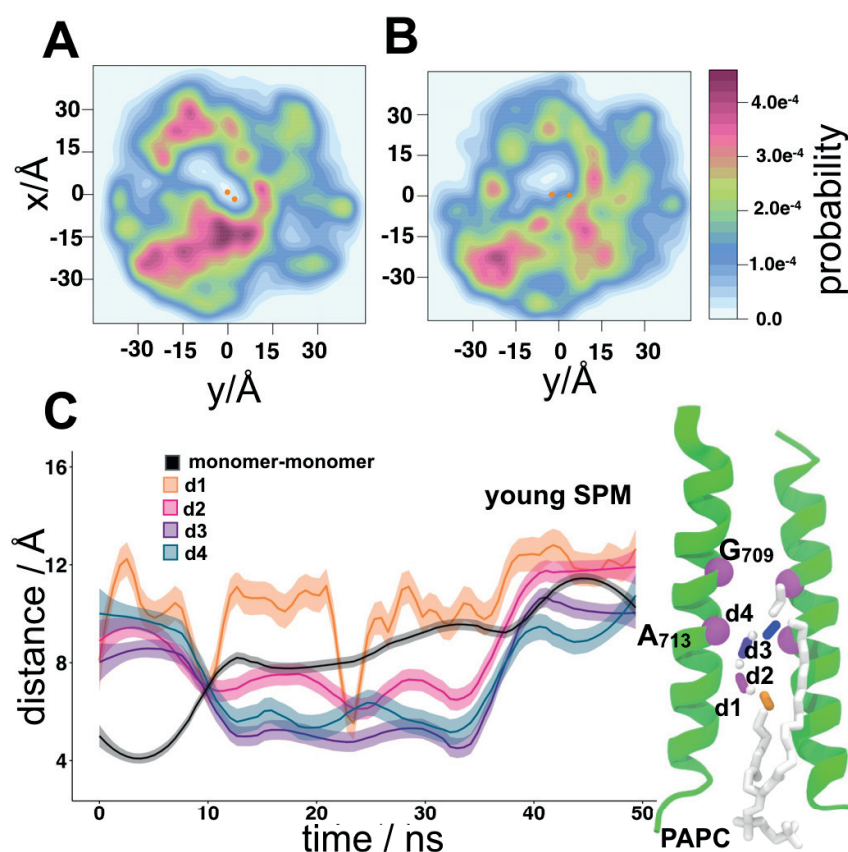


Figure 3.6 Unsaturation lipids distribution for $G_{709}XXXA_{713}$ and $G_{700}XXXG_{704}XXXG_{708}$ dimers. Averaged distribution of unsaturated lipids in young-SPM among all the MD replicas for (A) $G_{700}XXXG_{704}XXXG_{708}$ and (B) $G_{709}XXXA_{713}$. The orange dots represent the averaged positions of the helices in all the simulated replicas. (C) Representative lipid-protein interactions, based on a single replica, as a function of time for the young-SPM (similar graphs are reported for old-SPM and POPC/PAPC bilayer, Figure 3.5C-D). Distance between each lipid unsaturation and the $G_{709}XXXA_{713}$ motif residues is colored by the unsaturation position. The semi-transparent represent lines represent the standard deviation.

MD simulations (for an aggregate time of $2.5 \mu\text{s}$) were performed in a simple bilayer and in POPC/PAPC mixed membrane model (Figure 3.7A-B) confirming our results: the mere presence of lipids characterized by 4 unsaturation (PAPC) favoured the monomeric form of the $G_{709}XXXA_{713}$ dimer rather than the dimeric structure observed in POPC.

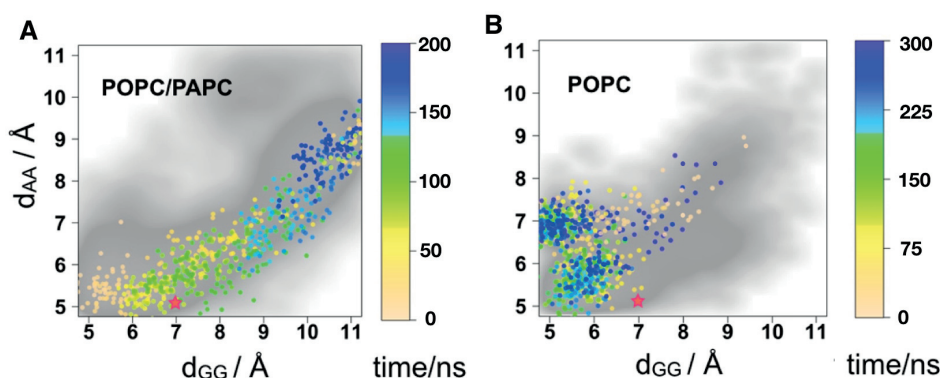


Figure 3.7. The $G_{709}XXXA_{713}$ dimer is not stable in POPC/PAPC membrane. APP-TM dimer evolution during MD in different membrane models: (A) POPC/PAPC and (B) POPC. The grey surface represents the global conformational space explored by each system, while the red star indicates the initial NMR values of d_{GG} and d_{AA} .

In summary, within the multi- μ s sampling achieved by our MD simulations at both CG and atomistic level of resolution, we observed that the $G_{700}XXXG_{704}XXXG_{708}$ dimer is the most stable and predominant species in complex SPM models, while the $G_{709}XXXA_{713}$ dimer is not stable in SPM as it is within a simple POPC bilayer, as suggested by our previous CG simulations (**Figure 3.3A**) and previous studies^{10,11,32,23}.

APP is a large transmembrane protein with a high avidity for cholesterol molecules⁴¹. Recently FRET and EPR experiments in POPC/POPG identified in the $G_{700}XXXG_{704}XXXG_{708}$ motif the amino acid residues physically involved in cholesterol binding⁴². In particular, it was highlighted the critical role played by $G_{700}XXXG_{704}XXXG_{708}$ in inducing flexibility to the membrane spanning region of the C99⁴³ and increasing the binding affinity for cholesterol molecules⁴⁴. The APP-TM/cholesterol binding energy was recently computational estimated in POPC bilayer showing also the importance of N-terminal residues⁴⁵. Nevertheless, the APP-TM/cholesterol interactions likely depend strongly on the local molecular environment adopted in a given experiment. For instance, in a mixed lysomyristoylphosphatidylglycerol (LMPG) micellar system the N-terminal residues (Gly696 and Lys699) are critical for binding water-soluble cholesterol analogue⁴⁶, but in dimyristoylphosphatidylcholine (DMPC) micelles cholesterol has a strong interaction with Gly700 and Gly704 and only weak with Gly696 and Lys699⁴⁴, therefore, in the more physiological and heterogeneous environment of the biological membrane the nature of these interaction could be even more diverse.

When the $G_{700}XXXG_{704}XXXG_{708}$ dimer was not formed in our MD simulations, cholesterol indeed bound to this motif with an angle $\theta \approx 40^\circ$ (**Figures 3.8A and 3.9**), as also previously observed for α -synuclein⁴⁷ and HIV gp41⁴⁸. Moreover, we observed in all atomistic MDs that the $G_{709}XXXA_{713}$ motif was also involved in cholesterol interactions (**Figure 3.7B**).

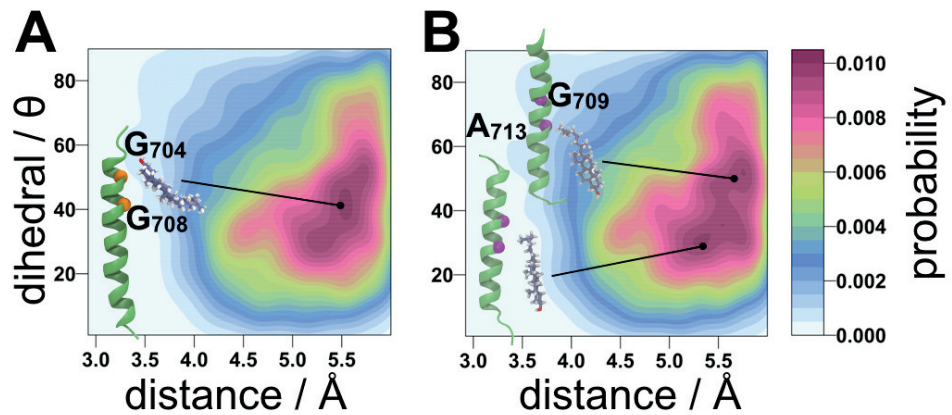


Figure 3.8. Cholesterol interactions with both APP-TM GXXXG dimer interfaces. MD-averaged cholesterol orientations around the $G_{700}XXXG_{704}XXXG_{708}$ (**A**) and $G_{709}XXXA_{713}$ (**B**) motif.

For example, in 40% of MD trajectories cholesterol bound to the $G_{709}XXXA_{713}$ motif, which remained exposed either when the $G_{700}XXXG_{704}XXXG_{708}$ dimer is formed (**Figures 3.4E-F**) or when the $G_{709}XXXA_{713}$ dimer dissociated (**Figure 3.4B-C-D**). The geometrical analysis of the cholesterol interactions with the $G_{709}XXXA_{713}$ motif showed two preferential orientations characterized by an angle $\theta \approx 27^\circ$ and $\theta \approx 57^\circ$ associated with binding from the SPM cyto-leaflet (**Figure 3.8B**).

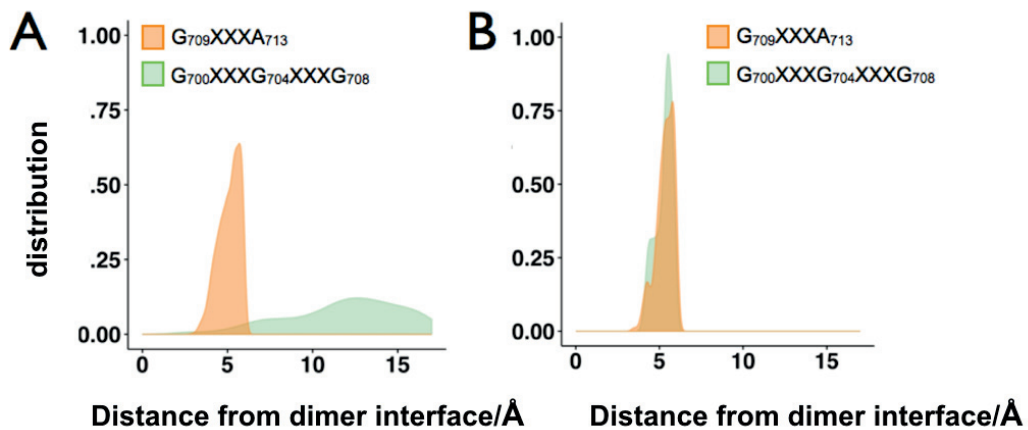


Figure 3.9. Distribution of cholesterol around the $G_{709}XXXA_{713}$ and $G_{700}XXXG_{704}XXXG_{708}$ motifs. Cholesterol distribution around the $G_{700}XXXG_{704}XXXG_{708}$ APP-TM dimer (**A**) and $G_{709}XXXA_{713}$ dimer (**B**), which then dissociate during MD simulations. In the case of $G_{700}XXXG_{704}XXXG_{708}$ dimer we observed that cholesterol has a strong preference for $G_{709}XXXA_{713}$ since the $G_{700}XXXG_{704}XXXG_{708}$ motif is involved in dimerization. In the case of $G_{709}XXXA_{713}$ dimer MD simulations showed that the dimer is not stable and cholesterol showed an indiscriminate preference for both $G_{700}XXXG_{704}XXXG_{708}$ and $G_{709}XXXA_{713}$ motifs

These results shed light on the possible mechanism of interaction between the APP-TM dimer and cholesterol molecules. It has been recently reported that APP and APP-C99 are responsible for the regulation of cholesterol turnover in synaptic function. Particularly, that the direct binding of the APP with sterol regulatory element binding protein 1 (SREBP1, a regulator factor responsible of cholesterol biosynthesis) occurs via the $G_{700}XXXG_{704}XXXG_{708}$ motif⁴⁹. If SREBP1 binds directly to APP-TM or the APP-TM/SREBP1 complex is activated when the cholesterol is already bound to the APP-TM remains an open question. In this context, our results suggest that APP-TM, in either the monomeric or dimeric state, can recruit cholesterol through the $G_{709}XXXA_{713}$ motif, thus possibly contributing to regulate the formation of the APP-TM/SREBP1 complex through the $G_{700}XXXG_{704}XXXG_{708}$ motif.

The biological relevance of APP-TM homodimer has been a controversial matter of debate. Recently, it has been experimentally observed that while APP-TM in monomeric state promotes an increase in $A\beta$ production, homo-dimeric and -trimeric species decrease $A\beta$ production⁵⁰. In line with the protective role of the homodimer for γ -secretase cleavage, it has been observed that the formation of a covalent APP dimer, obtained using lysine-scanning mutagenesis approach, led to produce only ≈ 10 -30% of $A\beta$, confirming that a dimer species is not easily cleaved by γ -secretase⁵¹,

likely by the difficulty to accommodate it in the catalytic pocket. Unfortunately, in both these studies, mutagenesis experiments were not targeting residues at the dimerization motif (G₇₀₄, G₇₀₈ and A₇₁₃), leaving open the question of whether the dimer formed is promoted by the G₇₀₉XXXA₇₁₃ or the G₇₀₀XXXG₇₀₄XXXG₇₀₈ motif. Moreover it has been reported that mutations on the residues in the APP-TM region influence the γ -secretase cleavage site increasing the ratio A β _{42/40}⁵².

Recently an independent relationship between dimerization and γ -secretase processing was proposed^{34,53,54}: rather than affecting dimerization, mutation at G₇₀₀XXXG₇₀₄XXXG₇₀₈ induces conformational changes that raise the affinity for γ -secretase and alter A β peptides production. Contrarily to what was previously thought, disruptive mutations of glycine residues at the G₇₀₀XXXG₇₀₄XXXG₇₀₈ motif impair the production of A β peptides^{21,34} suggesting that the formation of the APP-TM homodimer can be a control mechanism for APP γ -secretase cleavage. Our results support this latest hypothesis as the G₇₀₀XXXG₇₀₄XXXG₇₀₈ dimer conformation may allow for better accessibility of the cleavage site (i.e., at Val₇₁₁ and Ala₇₁₃ corresponding to Val₄₀ and Ala₄₂ in the A β numbering scheme), which must be exposed to the membrane in this state. Although some experimental evidences highlighted that γ -secretase is localized in the lysosome rather than in the SPM⁵⁵, recent lipidomics analysis on the composition of the late endosome (i.e., precursor of the lysosome) revealed that it is characterized by a high percentage of phosphatidylcholine and low concentration of cholesterol⁵⁶ which closely resembles our models of the SPM (see SI). In particular, it was observed that the total lipid contents of the membrane where APP is localized alone and APP/ γ -secretase are colocalized are similar⁵⁷. Since it is known from experiment that the APP is initially in the SPM⁵⁵, we can hypothesize that the effects of lipids unsaturation on the stability of APP-TM dimer in SPM as observed in this study could also reflect the behavior of the protein in a lysosome membrane environment.

Conclusions

In conclusion, these results reveal new features of APP-TM dimer association in the SPM that can be important for APP's biological function and can foster a more detailed understanding of A β production as well as promote a rational improvement of γ -secretase modulators currently under development. In general, our results support a preferential dimerization of APP-TM through the G₇₀₀XXXG₇₀₄XXXG₇₀₈ motif

in the SPM, whereas the G₇₀₉XXXA₇₁₃ motif could act as an additional cholesterol-recruiting domain. We further showed that the membrane composition has a critical influence on selecting the specific structural association of APP-TM helices. In particular, polyunsaturated lipids abundantly present in the SPM, but absent in the artificial conditions of the micelle or single-lipid bilayers, compete for interacting with the G₇₀₉XXXA₇₁₃ motif influencing in turn APP dimer stability.

Computational material and methods

Atomistic MD simulations

The initial dimer conformation of APP-TM is taken from experimental NMR structure of APP-TM G₇₀₉XXXA₇₁₃ in zwitterionic dodecylphosphocholine (DPC) micelles (PDB: 2LZ3)²⁴. All the membrane systems (young-SPM, old-SPM, POPC, POPC/PAPC) were generated using *LipidBuilder* (<http://lipidbuilder.epfl.ch>)⁵⁸, modeled using the CHARMM36 force field^{59,60} and solvated with the TIP3P water model⁶¹ and equilibrated following the protocol reported in ref. 2. The distribution of the lipids and cholesterol between the leaflets is summarized in Table S1. The NMR structure (PDB: 2LZ3) was inserted in a pre-equilibrated membrane (of 80 x 80 Å² size) and any clashes were removed. 5 independent replicas were run for young- and old-SPM systems as well as the three control bilayers featuring (i) SPM without cholesterol, (ii) and (iii) POPC/PAPC mixture (**Table 3.1**). We run 200 ns MD production for each replica of SPM (young, old and without cholesterol) and for POPC/PAPC system, while we extend up to 300 ns all the replicas of POPC for a total of 5.5 μs aggregate MD time. All atomistic MD simulations were performed using NAMD 2.9⁵⁹ using the CHARMM36 force field. The periodic electrostatic interactions were computed using particle mesh Ewald (PME) summation with grid spacing smaller than 1 Å. Constant pressure of 1 atm was maintained with Langevin piston dynamics¹⁴, a 200-fs decay period, and 50-fs time constant. Constant physiological temperature (310 K) was imposed by using Langevin dynamics⁶² with a damping coefficient of 1.0 ps. All the analysis were performed using as time interval 2 ps.

Coarse-Grained (CG) MD simulations.

The CG simulations are based on DAFT protocol³¹ where the helical models are coarse-grained using *martinize*⁶³ and aligned along the z-axis using PYMOL. 500

starting structures, where the monomers are separated by 35 Å, were generated and the outcome configurations were processed with *insane*⁶⁴, which generate the realistic SPM membrane and solvent. Each system was minimized and equilibrated for 10 ps in NVT MD using 2 fs of time step. Afterward 100 ps NpT MD with a time step of 20 fs was applied in order to bring the temperature and pressure in a range of 310 K and 1 bar respectively. The temperature was controlled using the Bussi thermostat with a coupling time of 1 ps⁶⁵ while the pressure was controlled by weak semi-isotropic coupling with a reference pressure of 1 bar and a compressibility of $3 \times 10^{-4} \text{ bar}^{-1}$. The production simulations were performed for 512 ns with a time step of 20 fs. All the simulations were performed using GROMACS 4.5.6⁶⁶ and the dimerization and relaxation properties were analyzed computing the non-bonded interaction energy over time between the two monomers embedded in the SPM (**Fig. 3.4**).

PCA clustering analysis.

Principal component analysis (PCA) was performed on roughly 44000 frames of MD trajectories. For the PCA, only C α beads coordinate were considered. Eigenvectors describing at least 80% of the protein C α fluctuation were selected to describe the protein flexibility. Following the projection of trajectory frames on relevant eigenvectors, a hierarchical clustering scheme⁶⁷ was applied in order to extract representative protein conformations.

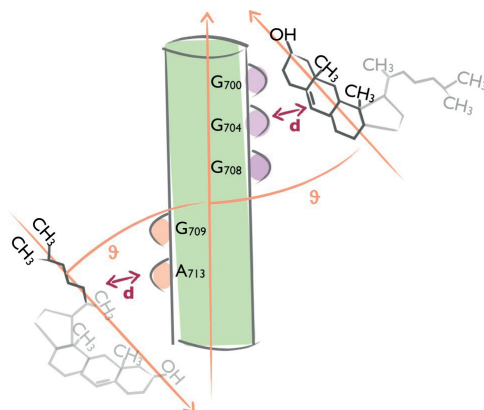
APP-TM/lipid interactions.

The interaction between lipids and APP-TM dimer was characterized by computing the distance between the center of mass of the atoms involved in the two dimerization motifs (G₇₀₀XXXG₇₀₄ XXXG₇₀₈ and G₇₀₉XXXA₇₁₃) and the center of mass of the atoms that form the unsaturation in the lipids.

APP-TM/cholesterol interactions.

The interaction between cholesterol molecules and APP-TM dimer was characterized by computing the (i) angle (θ) between the principal axis of APP-TM dimer and the principal axis of the cholesterol molecules (i.e. aromatic body in the case of the G₇₀₀XXXG₇₀₄XXXG₇₀₈ motif and hydrocarbon tail in the case of G₇₀₉XXXA₇₁₃) and (ii) the distance between the center of mass of all the atoms involved in the two

dimerization motifs (G₇₀₀XXXG₇₀₄ XXXG₇₀₈ and G₇₀₉XXXA₇₁₃) and the center of mass of the cholesterol molecules (see illustration below).



Scheme 3.1. Geometrical characterization of cholesterol/APP-TM interactions.

3.2 γ -SECRETASE MEMBRANE AND APP INTERPLAY: HOW THE TWO MAIN CHARACTERS OF ALZHEIMER DISEASE GO ALONG TOGETHER

γ -secretase is a membrane-embedded protease that cleaves single transmembrane helical domains of various integral membrane proteins. Of particular interest is the sequential cleavage of the amyloid precursor protein (APP) into peptides of variable lengths (A β). This cleavage occurs in two steps and two proteases are involved: the β -secretase first removes the ectodomain of APP and further the γ -secretase cleaves the remaining C-terminal fragment within its transmembrane domain delivering a mixture of 37 to 49 amino acid long A β -peptides. In particular, longer A β peptides (A β 42) are linked closely to Alzheimer's disease (AD) due to their propensity to aggregate and form extracellular senile plaques in the brain⁶⁸. Alternatively, cell surface APP can be cleaved by α -secretase to release the non-amyloidogenic soluble APP α (sAPP α), which was observed to be neuroprotective⁶⁹.

γ -secretase is a 230 kDa complex constituted by 20 TM helices and contains four main components: presenilin (PS1 and PS2), presenilin enhancer 2 (PEN-2), nicastrin (NCT), and anterior pharynx-defective 1 (APH-1) (**Fig. 3.9A**)⁷⁰. The full-length PS is inactive and association with Pen-2 facilitates an autocatalytic cleavage of presenilin between TM6 and TM7, producing two fragments known as N-terminal fragment (NTF) and C-terminal fragment (CTF)⁷¹⁻⁷³.

Subsequently, NTF and CTF bind to form a stable and active nine TMDs heterodimer⁷⁰, which it is actually the PS active catalytic form. At the interface between the NTF and CTF of presenilin is located the catalytic center composed by residues D257 and D385, located on TM6 and TM7 helices, which is excluded from the external surface of the enzyme (**Fig. 3.9**). NCT is a 130 kDa type I TM protein that contains a large glycosylated ectodomain (ECD) and a single TMD. The other two γ -secretase were initially identified through genetic screening in *C. elegans*⁷⁴. PEN-2 directly binds PS (NTF or CTF) and is required for its autocatalytic maturation and protease activity. APH-1 contains seven TMDs and is indispensable for γ -secretase assembly. Moreover, several studies showed that these four components are cross-regulated and down-regulation or deficiency of one given component typically destabilizes the others altering trafficking^{75,76}.

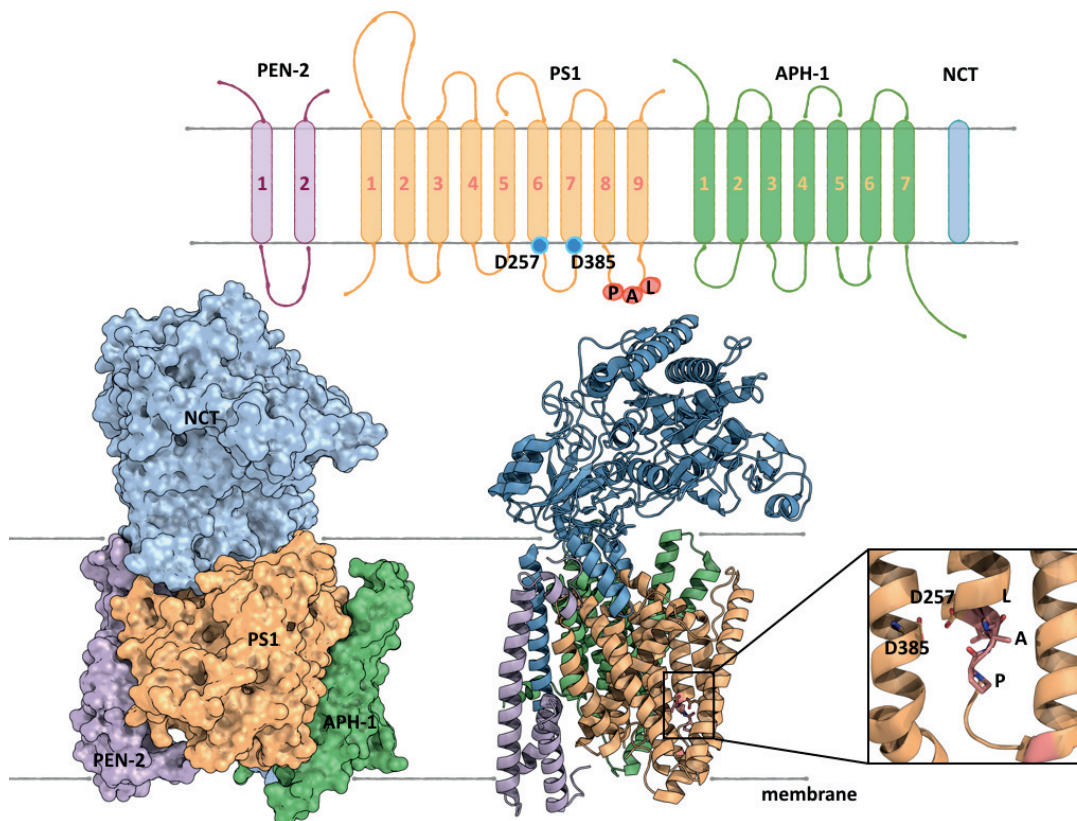


Figure 3.9. Structure of human γ -secretase. The γ -secretase architecture and structure is shown in surface and cartoon representation: PS1 (orange), APH-1 (violet), PEN-2 (green) and NCT (light blue), with focus in the catalytic dyad and adjacent PAL motif. The PAL loop is highlighted in red and the close view show its relative position compare to the catalytic residues.

During the latest years, a large number of investigations on γ -secretase have contributed to the rapid progress of this field of research. Biochemical experiments⁷⁷⁻⁸⁰ shed light on the active site of γ -secretase suggesting an extended binding surface for the initial substrate binding. Advancements in cryo-electron microscopy (cryo-EM) have resulted in high-resolution structures of γ -secretase. Indeed, a 4.2 Å resolution cryo-EM structure deposited in 2015 (PDB: 5FN2) provides an atomistic organization of all four domains of the human γ -secretase as well as the large ECD⁷⁰. In particular, this structure allows moving one step further on the study of Alzheimer's disease investigating on the recognition and recruitment mechanism of APP. It has been shown that the recognition and recruitment of substrate can occur at positions distinct from the γ -secretase active site⁸¹ and the translocation of the substrate to the active site is likely coupled with a large-scale conformational change. In particular, it was suggested that a considerable conformational change in nicastrin (NCT) extracellular domain is required in order to allow the accessibility of the active site at PS. Indeed, the heavily glycosylated ECD of NCT is thought to recognize the N-terminus of substrate protein⁸² and redirect it to PS catalytic pocket for further cleavage. The cryo-EM structure used in this study (PDB: 5NF2) does not show the glycosylated ECD sites on NCT. An earlier cryo-EM structure at 3.4 Å of resolution (PDB: 5A63) does indeed provide density map for 10 out of 16 potential glycosylated sites⁸²; however, it is not currently known the functions of these N-linked glycans and further investigations are required.

Recently, it was observed that the monomeric APP interacted with both NTF and CTF and the flexibility of the APP influences the stability of these interactions⁸³. Moreover, the non-covalent APP homodimer is protected from cleavage of γ -secretase independently from the dimerization motif⁵¹. Therefore, in this study we decided to investigate the interplay between APP and γ -secretase with particular regard to the specific membrane environment, namely, the effect of the lipids membrane composition they are experiencing. In fact, previous results suggested that the relatively small amount of active γ -secretase complexes present at the cell surface could be residing in raft membrane domains⁸⁴. Based on a recent lipidomics analysis of AD human brain tissues⁵⁷, we modeled a coarse-grained representation of the synaptic plasma membrane (SPM) using the MARTINI force field, as done previously for APP studies. In these conditions, the γ -secretase complex surrounded by 6 randomly placed copies of APP was embedded in this complex membrane model,

where we explored the recruitment and recognition of the APP by γ -secretase and the raft formation with respect to the proteins' localization as well.

Results and discussion

Lipid rafts are dynamic and highly ordered membrane micro-domains rich in cholesterol and sphingolipids. They are usually associated to thicker region of the membrane compared to the adjacent areas due to an ordering effect provided by saturated hydrocarbon chains of the lipids and cholesterol molecules⁸⁵. These highly ordered micro-domains are believed to play an important role in cellular function serving as a platform for cellular processes like cell signaling, motility and protein sorting or trafficking^{86,87}. In particular, several studies showed the coexistence of APP and γ -secretase substrates in these micro-domains and the absence or inhibition of γ -secretase activity resulted in the accumulation of APP in lipid raft domains^{88,89}. Altogether, these results strongly suggested that γ -secretase cleavage of APP occurs in raft domains.

Therefore, we first built a model of the SPM including 32 different lipids species observed in the AD brain tissue patients⁵⁷. This model of the SPM reflects the total lipids distribution in the presence of the AD showing an increment in sphingolipids, ganglioside and cholesterol compared to a health brain tissue. These lipids anomalies are potentially linked with AD pathogenesis and therefore this lipidomics analysis represents a valid starting point for studying the interplay between γ -secretase and APP. Our SPM model, produced using the MARTINI force field, was thus composed of phosphatidylcholine (PC, counting for a 21% of the total lipid content), phosphatidylethanolamine (PE, 11%), phosphatidic acid (PA, 0.5%), sphingomyelin (SM, 6%), negatively charged phosphatidylserine (PS, 2.5%) and phosphatidylinositol (PI, 2%), cholesterol (55% of the SPM) and ceramide (CE) monosialodihexosylganglioside (GM3) made up the remaining 2%. The lipid tails included saturated palmitoleic, polyunsaturated arachidonic and docosahexaenoic acids (**Table 2**).

Table 2. Lipids distribution of the synaptic plasma membrane models of AD's brain tissue patients. The total relative abundance of the different lipids species is given for each lipid species used. Each name correspond to the conventional name found in the MARTINI force field.

	SM (%)
PIPC (16:0/18:2)	0.36 (3)
POPC (16:0/18:1)	6.93 (57)
DPPC (16:0/18:0)	2.55 (21)
PAPC (16:0/20:4)	2.92 (24)
PEPC (16:0/20:2)	1.58 (13)
PGPC (16:0/20:1)	3.53 (29)
PUPC (16:0/22:6)	2.07 (17)
PRPC (16:0/24:6)	1.95 (16)
PIPE (16:0/18:2)	0.12 (1)
POPE (16:0/18:1)	0.49 (4)
DPPE (16:0/18:0)	0.12 (1)
PAPE (16:0/20:4)	0.49 (4)
PQPE (16:0/20:3)	0.12 (1)
PGPE (16:0/20:1)	0.97 (8)
PUPE (16:0/22:6)	2.43 (20)
PRPE (16:0/24:6)	6.20 (51)
POPA (16:0/18:1)	0.24 (2)
PGPA (16:0/20:1)	0.24 (2)
PGPS (16:0/20:1)	0.73 (6)
PRPS (16:0/24:6)	1.82 (15)
POPI (16:0/18:1)	0.24 (2)
PVPI (16:0/18:1)	0.24 (2)
PAPI (16:0/20:4)	1.46 (12)
PUPI (16:0/22:6)	0.24 (2)
PPC (16:0/18:0)	0.36 (3)
POSM (18:1/18:1)	0.36 (3)

PVSM (18:1/18:1)	0.36 (3)
DPSM (18:1/18:0)	4.26 (35)
PGSM (18:1/22:1)	0.12(1)
PNSM (18:1/24:1)	1.22 (10)
DPCE (18:1/18:0)	0.36 (3)
DPG3 (18:1/18:0)	0.12 (1)
CHOL	54.74 (450)

In this context, we examined the raft lipid formation with respect to the position of the γ -secretase. The top view and a cut through the planar SPM model (**Fig. 3.10A**) revealed the coexistence of the 2 phases at equilibrium: (i) L_o characterized by the presence of saturated and cholesterol molecules that are parallel to bilayer normal and (ii) L_d , which is composed by the unsaturated lipids. The final composition of the raft-domain (L_o), calculated on the last 500 ns of the MD simulation, is on average 0.51:0:20:0:28 of cholesterol/unsaturated lipids/saturated lipids respectively, while the final composition of the L_d phase is on average 0.35:0.45:0.19. The comparison between these two membrane regions showed roughly 2 times more cholesterol in the L_o phase compared to the L_d as well as an increment in unsaturated lipids in L_d as previously experimentally⁹⁰ and computationally⁹¹ observed for a simple three components lipid bilayer. The difference in lipid composition between these 2 domains is explained by diverse thickness values (**Fig. 3.10B**) and average order parameter distribution (**Fig. 3.10C**), which altogether represent key factors for lipid phase evaluation. In fact, the high thickness values are associated with high order parameters peculiar of a L_o phase, while low thickness value and low order parameters indicate an L_d region. Moreover, the difference in thickness of roughly 6 Å between L_o and L_d is in agreement with atomic force microscopy⁹² and NMR⁹³ data as well as previous molecular dynamics simulations^{91,94}. Although the average size of lipid rafts observed in our MD simulations (9 nm in diameter) is smaller compared to size experimentally estimated (50 nm in diameter), our simulations clearly hinted to the initial step of the formation of raft-like regions.

Interestingly, all our MD simulations showed that a greater thickness is present on the convex side of the γ -secretase structure, namely the region in close proximity of the extracellular part of NCT domain and PS1. These results agree with the previous

experiment where the association of the four domains of the γ -secretase complex has been investigated in neuronal cells that fulfill the criteria of lipid rafts⁹⁵. While the exact thickness distribution varies across all the replicas, this local feature is present in all four of our replicas (**Fig. 3.10**). Liquid order parameters around these regions (**Fig. 3.10C**) mirror the state of the membrane thickness (**Fig. 3.10D**): the membrane has a greater L_0 phase near the convex side of the thickness.

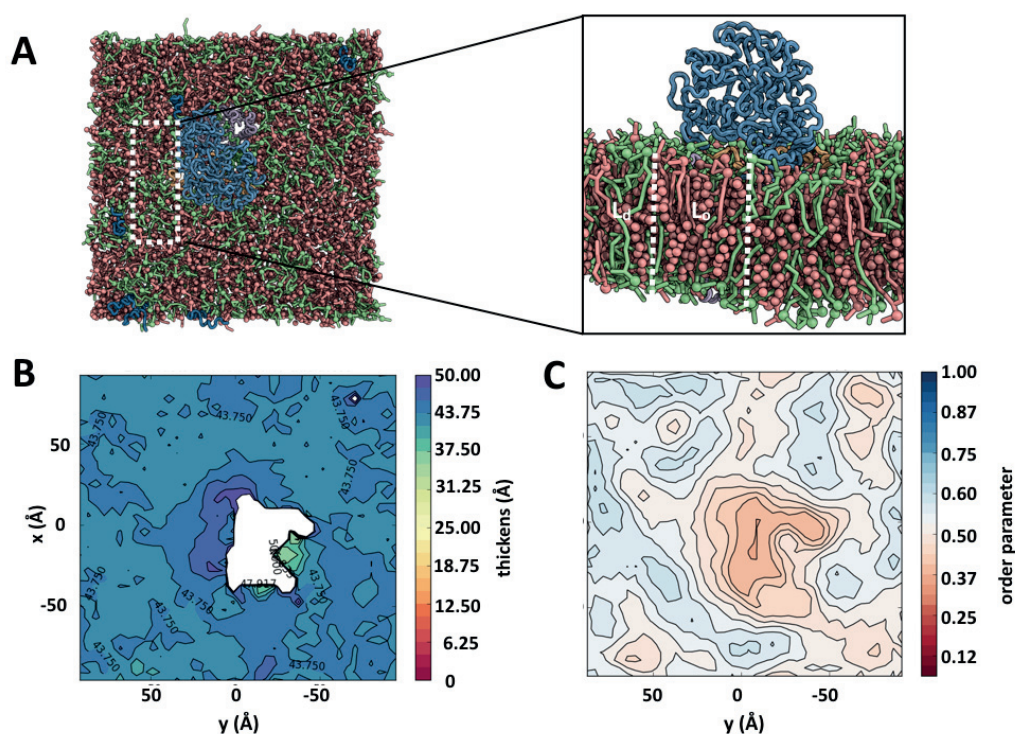


Figure 3.10. The NCT domain of γ -secretase is localized in a raft-like domain. (A) Top-view showing the position of the γ -secretase with respect to the lipid-raft composed by saturated lipids and cholesterol (red). The side-view focus (right) displayed in detail the repartition between L_d phase composed by unsaturated lipids (green) and the L_0 phase composed by cholesterol (represented in van der Waals) and saturated lipids (red). (B) The probability density of the thickness, averaged on the last 500 ns, as well as (C) the averaged order parameter distribution, over the last 500 ns, highlighted the raft lipid formation.

Contrary, APP itself is not generally a raft protein, although a small portion of APP was observed to be localizing into lipid rafts⁹⁶. This APP-raft localization regulation has been suggested to involve an interaction between the APP and cholesterol⁹⁷. Our MD simulations confirmed these experimental data showing that APP is indeed located in 2 different membrane associated pools: either (i) in a raft domain when it is far away from the γ -secretase, or (ii) in a non-raft domain when it

interacts with the enzyme (**Fig. 3.11A**). In fact, the averaged cholesterol distribution with respect to the proteins shows that APP resides in high-cholesterol region when not in the vicinity of γ -secretase (violet dots), while when APP is close to the enzyme the distribution of cholesterol around the APP is lower (orange dot) (**Fig. 3.11B**). Moreover, we observed in our MD replicas that dimerization eventually occurs, but only in the presence of low cholesterol concentration, which is in line with our previous study where we observed that cholesterol interacts with the dimerization motif G₇₀₀XXXG₇₀₄XXXG₇₀₈ APP⁹⁸. Indeed, the dimer detected in this SPM model is again the stable G₇₀₀XXXG₇₀₄XXXG₇₀₈ conformation (Fig 3.12) that we demonstrated to be the preferential conformation of the APP in absence of γ -secretase⁹⁹.

These results suggest a possible relocation mechanism mediated by cholesterol distribution. In fact, the pool of APP in the raft-like domains is stabilized by the interaction with cholesterol, while close to the γ -secretase the APP-membrane stability decreased in order to facilitate the further expulsion of the A β products upon γ -secretase cleavage.

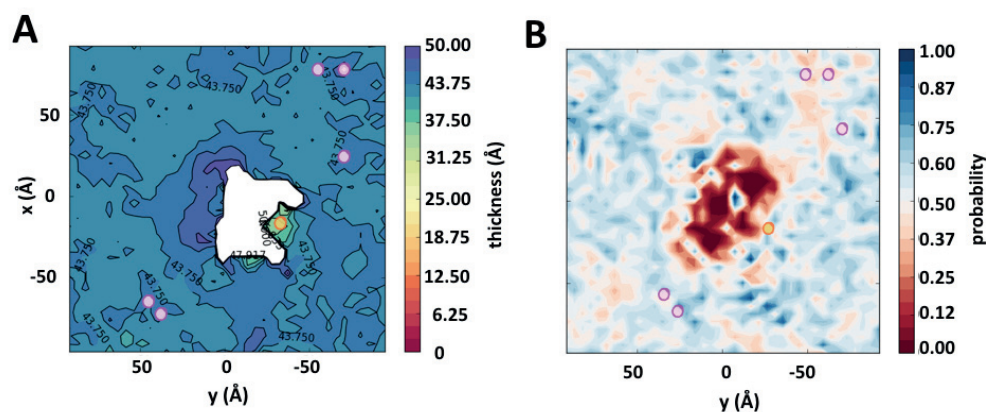


Figure 3.11. The APP localizes in two different regions of the SPM model. (A) Representative replica showed that APP either localized in a raft region (violet dots) or in a non-raft region (orange dots) when APP more closely interacts with the γ -secretase. (B) Average normalized cholesterol distribution showed that APP resides in high density cholesterol when far away from the enzyme.

Another open question in the field is related to the substrate recruitment mechanism by γ -secretase. It has been observed that generally the γ -secretase substrates are type-I membrane proteins that have their extracellular part mostly removed by sheddase, as for instance like α - and β -secretase¹⁰⁰. It has been suggested that the PAL motif in PS1, as well as the extracellular NCT (**Fig. 3.9**), recognize the

substrates and recruits them for catalysis¹⁰¹⁻¹⁰⁵. In particular, it was experimentally observed that NCT had a central role in the APP-PS1 interactions. Indeed, the NCT knock-down by RNAi caused a loss of FRET signal between APP and PS1 in cell ¹⁰⁶. Recently, it was proposed that the NCT binds the ectodomains of the substrates redirecting them to the PS1, the catalytic active site of γ -secretase. This interaction was suggested to be mediated by the residue E333 in the NCT extracellular domain¹⁰⁷. However, the recent high-resolution γ -secretase structure revealed that E333 is actually buried and on the opposite side of the catalytic aspartate residues, D257 and D385^{108,109}. Therefore, the basic mechanism of substrate recognition by γ -secretase remains at the moment controversial and requires further investigation.

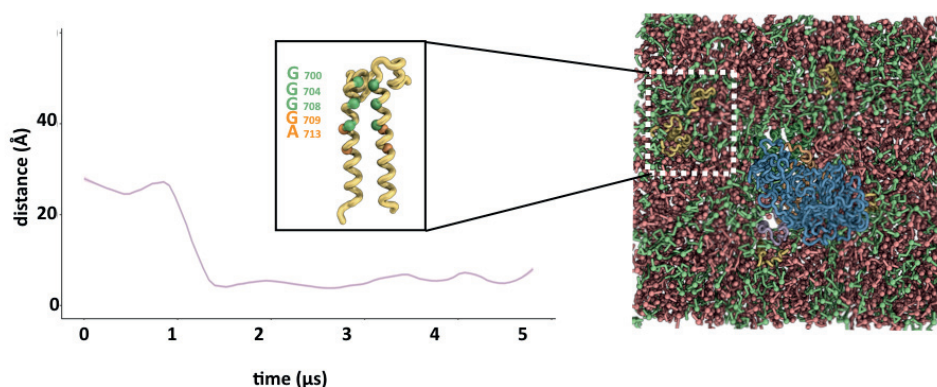


Figure 3.12. Stable G₇₀₀XXXG₇₀₄XXXG₇₀₈ APP dimer where observed in the complex SPM model. Representative snapshot of the dimeric APP conformer observed in a complex SPM model. Distance between the center of mass of the two APP monomers showed the interaction pathway over the MD simulation time.

As mentioned earlier MD simulations revealed that G₇₀₀XXXG₇₀₄XXXG₇₀₈ APP dimeric conformation is the stable one in this complex SPM model (**Fig. 3.13**), confirming our previous study¹¹⁰ reported above. However, we could observe in our MD simulations that when in proximity to the γ -secretase, APP is always found in the monomeric conformation (**Fig. 3.13**). We noticed that the flexibility of the N-terminal of APP seems to be important for the interaction with the enzyme. In the simulations where the APP structure was retained in its NMR based L-shape conformation by harmonic potentials between interacting sites, a substrate-enzyme interaction mode was never detected. A preferred APP interaction was observed for the NCT domain of the γ -secretase with a distance between APP and γ -secretase around 3.5 Å (**Fig. 3.13A-B**). In particular, we observed a preference for the NCT extracellular part that

was neglected in previous studies¹¹¹. On the other side, we never observed the interaction of APP with the catalytic residues D257 and D385 (on average in all simulations the distance between APP and catalytic residues is $\cong 40$ Å). In fact, the active site of PS1 is hardly accessible on the convex site of the TM horseshoe, suggesting a conformational reorganization of the nicastrin domain after substrate recruitment (**Fig. 3.13C**).

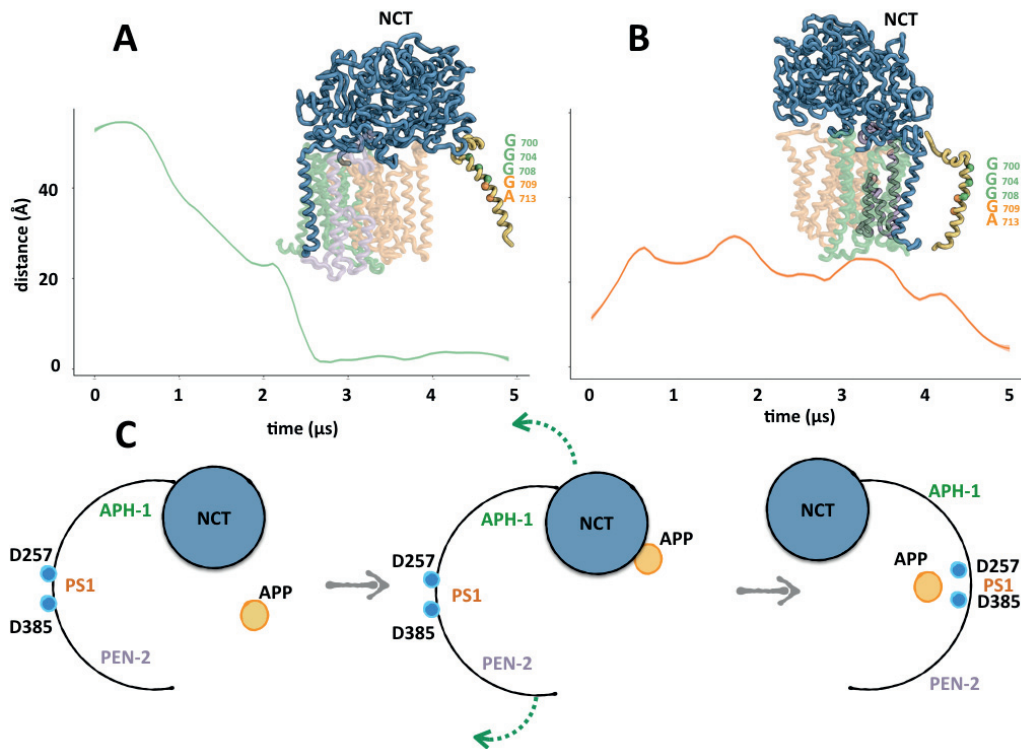


Figure 3.13. The interaction between APP and γ -secretase. Representative APP-enzyme interactions, based on two different MD replicas, as a function of time. The monomeric APP interacts mainly with the NCT domain and in particular with the TM (A) as well as the extracellular (B) regions. (C) Diagram of a possible mechanism for substrate binding at the catalytic site involving a large rearrangement of γ -secretase domains.

In the last 20 years, the knowledge of the AD's neurobiology increased leading to the development of several putative disease-modifying treatments. Particular attention was recently raised for the disease-modifying therapies based on the inhibition or modulation of α -, β - and γ -secretases^{74,112-114}. γ -secretase notably not only cleaves A β peptides but many other transmembrane proteins, for instance Notch protein. This broad range of substrates is therefore the main cause of toxicity in preclinical tests of existing γ -secretase inhibitors. Therefore, exhaustive efforts have

been made to design specific APP/ γ -secretase modulators. For example, it was discovered that some small molecules¹¹⁵ as well as non-steroidal anti-inflammatory drugs (NSAIDs) might modulate the APP cleavage without interfering with the cleavage of other substrates. This initial study shed light on the early stages of APP recruitment and interaction mechanism with γ -secretase, which would be important for further drug design investigations.

In conclusion, these MD simulations performed in a complex model of SPM revealed the importance of lipids rafts in the localization and stability of both the substrates and enzyme. In particular, we observed that the APP is localized in two different regions of the synaptic plasma membrane, which vary mainly by their cholesterol concentration. This in turn suggests the key role of these sterol molecules in stabilizing the substrates before recruitment by the enzyme. Moreover, our computational study reveals insights into the role of NCT domain in substrate recognition endorsing the importance of this extracellular domain. Although our study is still preliminary, has to deal with all the known limitations associated with the approach we chose (i.e., CG MD simulations) and further investigations are surely required, these results provide some new structural information on the substrate-enzyme complex and its localization at the membrane. Altogether, our results allow for a glimpse of the proteolytic processes involved in AD pathology representing also a possible starting point for future investigation on the action of mechanism of γ -secretase inhibitors.

Computational material and methods

Coarse-Grained (CG) MD simulations.

A 4.2 Å resolution cryo-EM structure (PDB ID: 5FN2) was used as the initial model for MD simulations of γ -secretase¹¹⁶. This structure was chosen because it comprised all of four subunits of γ -secretase (Aph-1, Pen-2, PS1, and NCT). APP was taken from an NMR structure (PDB ID: 2LLM) that contains both the transmembrane domain and the extracellular kink¹¹⁷. Both atomic structures were coarse-grained using *martinize*¹¹⁸. γ -secretase and six randomly placed APP were embedded into a synaptic plasma membrane generated using *insane*¹¹⁹. The composition of the plasma

membrane was constructed to be as similar as possible to its experimental composition reported in recent lipidomics data ⁵⁷, similar to the extent of available Martini CG parameters ^{118,120}. The dimension of the membrane model used in all MD replicas was 200 Å × 200 Å.

In the present study, all MD simulations were conducted using the GROMACS package ¹²¹. These simulations were used to examine processes of substrate binding to γ -secretase and to derive information about binding poses. All the systems were solvated in MARTINI CG water boxes, buffered with NaCl at 150 mM concentration. The polarized MARTINI 2.2 parameters were used for the simulations ^{118,120}. All the production runs were preceded by a 5000-step energy minimization followed by a 100 ns NVT and 1 μ s NPT equilibration with positions of protein beads being restrained. The temperature was set to 323 K using the V-rescale thermostat and the pressure was set to 1 bar using a semi-isotropic coupling method. A time step of 20 fs was used, a typical value employed in MARTINI MD simulations. Four replicas of the systems ran for 5 μ s each, the first 500 ns and the last 500 ns of which was used for analysis.

Membrane thickness and order parameter.

Two useful quantities to measure in a MD simulation of a membrane system are the membrane thickness and the order parameter of the lipid chains. In particular, these observables are helpful in determining the presence of lipid rafts.

Membrane thickness is first obtained by splitting the bilayer into its inner and outer leaflet. Each leaflet is overlaid on a grid of nodes on the xy-plane, where each node is equally spaced from each other. Each lipid assigns its z-position to the closest matching node. The thickness is obtained by subtracting each grid of nodes element-wise from each other. A finer grid provides more accurate estimation of the thickness, but is more computationally demanding. In this study, each node is set to have a neighborhood with the dimension of 5 Å by 5 Å (see **Fig.3.10** and **Fig.3.11**).

Popular measurements of order in lipid bilayers are of the type that can be measured by deuterium (D) NMR. Mathematically, this type of order parameter can be defined as

$$P_2 = \frac{1}{3} \langle \cos^2 \theta \rangle - \frac{1}{2}$$

where θ is the angle between a CD-bond (in the experiment) or a CH-bond (in the simulation) and the membrane normal. An order parameter of zero can thus represent either an unordered (isotropic) system or a perfectly ordered system oriented at the magic angle of 54.7 with respect to the magnetic field. A value of -0.5 indicates a perfectly ordered acyl chain in all-trans conformation, rapidly rotating around the bilayer normal. Likewise, a value of 1 indicates perfect alignment with the bilayer normal.

BIBLIOGRAPHY

1. Alzheimer's, A. 2015 Alzheimer's disease facts and figures. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **11**, 332 (2015).
2. Brookmeyer, R. et al. Forecasting the global burden of Alzheimer's disease. *Alzheimer's & dementia* **3**, 186-191 (2007).
3. Goate, A. et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704 (1991).
4. Chartier-Harlin, M.-C. et al. Early-Onset Alzheimer's Disease Caused by Mutations at Codon 717 of the (Beta)-Amyloid Precursor Protein Gene. *Nature* **353**, 844 (1991).
5. Murrell, J. et al. A mutation in the amyloid precursor protein associated with hereditary Alzheimer's disease. *Science* **254**, 97 (1991).
6. Hilbich, C. et al. Aggregation and secondary structure of synthetic amyloid β A4 peptides of Alzheimer's disease. *Journal of molecular biology* **218**, 149-163 (1991).
7. Jarrett, J.T. et al. Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **73**, 1055-1058 (1993).
8. Pike, C.J. et al. Neurodegeneration induced by beta-amyloid peptides in vitro: the role of peptide assembly state. *Journal of Neuroscience* **13**, 1676-1687 (1993).
9. Lorenzo, A. et al. Beta-amyloid neurotoxicity requires fibril formation and is inhibited by congo red. *Proceedings of the National Academy of Sciences* **91**, 12243-12247 (1994).
10. Haass, C. et al. Normal cellular processing of the beta-amyloid precursor protein results in the secretion of the amyloid beta peptide and related molecules. *Ann N Y Acad Sci* **695**, 109-116 (1993).
11. Di Paolo, G. et al. Linking lipids to Alzheimer's disease: cholesterol and beyond. *Nat Rev Neurosci* **12**, 284-296 (2011).
12. Grosvenor, S. et al. Role of amyloid beta in lipid homeostasis. *Biochim Biophys Acta* **1801**, 966-974 (2010).
13. Barrett, P.J. et al. The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol. *Science* **336**, 1168-1171 (2012).
14. Feller, S.E. et al. Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *Journal of Chemical Physics* **103**, 4613-4621 (1995).
15. Lemmin, T. et al. Perturbations of the straight transmembrane alpha-helical structure of the amyloid precursor protein affect its processing by gamma-secretase. *J Biol Chem* **289**, 6763-6774 (2014).

16. Dominguez, L. et al. Transmembrane fragment structures of amyloid precursor protein depend on membrane surface curvature. *J Am Chem Soc* **136**, 854-857 (2014).
17. Sato, T. et al. A helix-to-coil transition at the epsilon-cut site in the transmembrane dimer of the amyloid precursor protein is required for proteolysis. *Proc Natl Acad Sci U S A* **106**, 1421-1426 (2009).
18. Litman, B.J. et al. The role of docosahexaenoic acid containing phospholipids in modulating G protein-coupled signaling pathways: visual transduction. *J Mol Neurosci* **16**, 237-242; discussion 279-284 (2001).
19. Miyashita, N. et al. Structures of beta-amyloid peptide 1-40, 1-42, and 1-55-the 672-726 fragment of APP-in a membrane environment with implications for interactions with gamma-secretase. *J Am Chem Soc* **131**, 17843-17852 (2009).
20. Wang, H. et al. Molecular determinants and thermodynamics of the amyloid precursor protein transmembrane domain implicated in Alzheimer's disease. *J Mol Biol* **408**, 879-895 (2011).
21. Kienlen-Campard, P. et al. Amyloidogenic processing but not amyloid precursor protein (APP) intracellular C-terminal domain production requires a precisely oriented APP dimer assembled by transmembrane GXXXG motifs. *J Biol Chem* **283**, 7733-7744 (2008).
22. Gorman, P.M. et al. Dimerization of the transmembrane domain of amyloid precursor proteins and familial Alzheimer's disease mutants. *BMC Neurosci* **9**, 17 (2008).
23. Nadezhdin, K.D. et al. Dimeric structure of transmembrane domain of amyloid precursor protein in micellar environment. *FEBS Lett* **586**, 1687-1692 (2012).
24. Chen, W. et al. Familial Alzheimer's mutations within APPTM increase Abeta42 production by enhancing accessibility of epsilon-cleavage site. *Nat Commun* **5**, 3037 (2014).
25. Fisher, L.E. et al. Effect of detergents on the association of the glycoporphin a transmembrane helix. *Biophys J* **85**, 3097-3105 (2003).
26. Fisher, L.E. et al. Detergents modulate dimerization, but not helicity, of the glycoporphin A transmembrane domain. *J Mol Biol* **293**, 639-651 (1999).
27. Igbavboa, U. et al. A new role for apolipoprotein E: modulating transport of polyunsaturated phospholipid molecular species in synaptic plasma membranes. *J Neurochem* **80**, 255-261 (2002).
28. Ingolfsson, H.I. et al. Lipid organization of the plasma membrane. *J Am Chem Soc* **136**, 14554-14559 (2014).
29. Piggot, T.J. et al. Electroporation of the E. coli and S. Aureus membranes: molecular dynamics simulations of complex bacterial membranes. *J Phys Chem B* **115**, 13381-13388 (2011).
30. Polley, A. et al. Atomistic simulations of a multicomponent asymmetric lipid bilayer. *J Phys Chem B* **116**, 13403-13410 (2012).
31. Wassenaar, T.A. et al. High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. *J Chem Theory Comput* **11**, 2278-2291 (2015).
32. Dominguez, L. et al. Structural heterogeneity in transmembrane amyloid precursor protein homodimer is a consequence of environmental selection. *J Am Chem Soc* **136**, 9619-9626 (2014).
33. Russ, W.P. et al. TOXCAT: a measure of transmembrane helix association in a biological membrane. *Proc Natl Acad Sci U S A* **96**, 863-868 (1999).

34. Munter, L.M. et al. GxxxG motifs within the amyloid precursor protein transmembrane sequence are critical for the etiology of Abeta42. *EMBO J* **26**, 1702-1712 (2007).
35. Wood, W.G. et al. Brain membrane cholesterol domains, aging and amyloid beta-peptides. *Neurobiol Aging* **23**, 685-694 (2002).
36. Igbavboa, U. et al. Increasing age alters transbilayer fluidity and cholesterol asymmetry in synaptic plasma membranes of mice. *J Neurochem* **66**, 1717-1725 (1996).
37. Bovigny, C. et al. LipidBuilder: A Framework To Build Realistic Models for Biological Membranes. *J. Chem. Inf. Model* **55**, 2491-2499 (2015).
38. Feller, S.E. et al. Polyunsaturated fatty acids in lipid bilayers: intrinsic and environmental contributions to their unique physical properties. *J Am Chem Soc* **124**, 318-326 (2002).
39. Applegate, K.R. et al. Computer-based modeling of the conformation and packing properties of docosahexaenoic acid. *J Lipid Res* **27**, 658-680 (1986).
40. Deguil, J. et al. Modulation of lipid-induced ER stress by fatty acid shape. *Traffic* **12**, 349-362 (2011).
41. Lahdo, R. et al. The amyloid precursor protein interacts with neutral lipids. *Eur J Biochem* **269**, 2238-2246 (2002).
42. Song, Y. et al. Competition between homodimerization and cholesterol binding to the C99 domain of the amyloid precursor protein. *Biochemistry* **52**, 5051-5064 (2013).
43. Di Scala, C. et al. Interaction of Alzheimer's beta-amyloid peptides with cholesterol: mechanistic insights into amyloid pore formation. *Biochemistry* **53**, 4489-4502 (2014).
44. Barrett, P.J. et al. The Amyloid Precursor Protein Has a Flexible Transmembrane Domain and Binds Cholesterol. *Science* **336**, 1168-1171 (2012).
45. Nierzwicki, L. et al. Specific Binding of Cholesterol to the Amyloid Precursor Protein: Structure of the Complex and Driving Forces Characterized in Molecular Detail. *J Phys Chem Lett* **6**, 784-790 (2015).
46. Beel, A.J. et al. Direct binding of cholesterol to the amyloid precursor protein: An important interaction in lipid-Alzheimer's disease relationships? *Biochim Biophys Acta* **1801**, 975-982 (2010).
47. Crowet, J.M. et al. Tilted properties of the 67-78 fragment of alpha-synuclein are responsible for membrane destabilization and neurotoxicity. *Proteins* **68**, 936-947 (2007).
48. Charloteaux, B. et al. The N-terminal 12 residue long peptide of HIV gp41 is the minimal peptide sufficient to induce significant T-cell-like membrane destabilization in vitro. *J Mol Biol* **359**, 597-609 (2006).
49. Grimm, M.O. et al. The role of APP proteolytic processing in lipid metabolism. *Exp Brain Res* **217**, 365-375 (2012).
50. Jung, J.I. et al. Independent relationship between amyloid precursor protein (APP) dimerization and gamma-secretase processivity. *PLoS One* **9**, e111553 (2014).
51. Winkler, E. et al. Homodimerization Protects the Amyloid Precursor Protein C99 Fragment from Cleavage by gamma-Secretase. *Biochemistry* **54**, 6149-6152 (2015).

52. Dimitrov, M. et al. Alzheimer's disease mutations in APP but not gamma-secretase modulators affect epsilon-cleavage-dependent AICD production. *Nat Commun* **4**, 2246 (2013).
53. Jiang, W. et al. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. *J Phys Chem Lett* **2**, 87-92 (2011).
54. Page, R.M. et al. Beta-amyloid precursor protein mutants respond to gamma-secretase modulators. *J Biol Chem* **285**, 17798-17810 (2010).
55. Pasternak, S.H. et al. The role of the endosomal/lysosomal system in amyloid-beta production and the pathophysiology of Alzheimer's disease: reexamining the spatial paradox from a lysosomal perspective. *J Alzheimers Dis* **6**, 53-65 (2004).
56. Hamer, I. et al. Lipids and lysosomes. *Curr Drug Metab* **13**, 1371-1387 (2012).
57. Chan, R.B. et al. Comparative lipidomic analysis of mouse and human brain with Alzheimer disease. *J Biol Chem* **287**, 2678-2688 (2012).
58. Bovigny, C. et al. Lipid Builder: A Framework To Build Realistic Models for Biological Membranes. *Journal of Chemical Information and Modeling* **55**, 2491-2499 (2015).
59. Klauda, J.B. et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* **114**, 7830-7843 (2010).
60. Phillips, J.C. et al. Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802 (2005).
61. Jorgensen, W.L. et al. Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926-935 (1983).
62. Brunger, A. et al. Stochastic Boundary-Conditions for Molecular-Dynamics Simulations of St2 Water. *Chemical Physics Letters* **105**, 495-500 (1984).
63. de Jong, D.H. et al. Improved parameters for the martini coarse-grained protein force field. *Journal of Chemical Theory and Computation* **9**, 687-697 (2012).
64. Wassenaar, T.A. et al. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *Journal of Chemical Theory and Computation* **11**, 2144-2155 (2015).
65. Bussi, G. et al. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101 (2007).
66. Hess, B. et al. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **4**, 435-447 (2008).
67. Lindahl, E. et al. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **7**, 306-317 (2001).
68. Zhang, X. et al. The gamma-secretase complex: from structure to function. *Front Cell Neurosci* **8**, 427 (2014).
69. Greenfield, J.P. et al. Cellular and molecular basis of beta-amyloid precursor protein metabolism. *Front Biosci* **5**, D72-83 (2000).
70. Bai, X.C. et al. An atomic structure of human gamma-secretase. *Nature* **525**, 212-217 (2015).
71. Ratovitski, T. et al. Endoproteolytic processing and stabilization of wild-type and mutant presenilin. *J Biol Chem* **272**, 24536-24541 (1997).
72. Takasugi, N. et al. The role of presenilin cofactors in the gamma-secretase complex. *Nature* **422**, 438-441 (2003).
73. Thinakaran, G. et al. Endoproteolysis of presenilin 1 and accumulation of processed derivatives in vivo. *Neuron* **17**, 181-190 (1996).

74. Levitan, D. et al. Facilitation of lin-12-mediated signalling by sel-12, a *Caenorhabditis elegans* S182 Alzheimer's disease gene. *Nature* **377**, 351 (1995).
75. De Strooper, B. Aph-1, Pen-2, and Nicastrin with Presenilin generate an active gamma-Secretase complex. *Neuron* **38**, 9-12 (2003).
76. Iwatsubo, T. The gamma-secretase complex: machinery for intramembrane proteolysis. *Curr Opin Neurobiol* **14**, 379-383 (2004).
77. Imamura, Y. et al. Inhibition of gamma-secretase activity by helical beta-peptide foldamers. *J Am Chem Soc* **131**, 7353-7359 (2009).
78. Kornilova, A.Y. et al. The initial substrate-binding site of gamma-secretase is located on presenilin near the active site. *Proc Natl Acad Sci U S A* **102**, 3230-3235 (2005).
79. Takagi-Niidome, S. et al. Cooperative roles of hydrophilic loop 1 and the C-terminus of presenilin 1 in the substrate-gating mechanism of gamma-secretase. *J Neurosci* **35**, 2646-2656 (2015).
80. Watanabe, N. et al. Functional analysis of the transmembrane domains of presenilin 1: participation of transmembrane domains 2 and 6 in the formation of initial substrate-binding site of gamma-secretase. *J Biol Chem* **285**, 19738-19746 (2010).
81. Esler, W.P. et al. Activity-dependent isolation of the presenilin- γ -secretase complex reveals nicastrin and a γ substrate. *Proceedings of the National Academy of Sciences* **99**, 2720-2725 (2002).
82. Schedindin (2002).et al. The role of protein glycosylation in Alzheimer disease. *The FEBS journal* **281**, 46-62 (2014).
83. Li, S. et al. Initial Substrate Binding of gamma-Secretase: The Role of Substrate Flexibility. *ACS Chem Neurosci* (2017).
84. Vetrivel, K.S. et al. Association of gamma-secretase with lipid rafts in post-Golgi and endosome membranes. *J Biol Chem* **279**, 44945-44954 (2004).
85. Vidal, A. et al. Transbilayer Peptide Sorting between Raft and Nonraft Bilayers: Comparisons of Detergent Extraction and Confocal Microscopy. *Biophysical Journal* **89**, 1102-1108 (2005).
86. Simons, K. et al. Lipid rafts and signal transduction. *Nature reviews Molecular cell biology* **1**, 31-39 (2000).
87. Helms, J.B. et al. Lipids as targeting signals: lipid rafts and intracellular trafficking. *Traffic* **5**, 247-254 (2004).
88. Vetrivel, K.S. et al. Spatial segregation of γ -secretase and substrates in distinct membrane domains. *Journal of Biological Chemistry* **280**, 25892-25900 (2005).
89. Wahrle, S. et al. Cholesterol-dependent γ -secretase activity in buoyant cholesterol-rich membrane microdomains. *Neurobiology of disease* **9**, 11-23 (2002).
90. Veatch, S. et al. Liquid domains in vesicles investigated by NMR and fluorescence microscopy. *Biophysical journal* **86**, 2910-2922 (2004).
91. Risselada, H.J. et al. The molecular face of lipid rafts in model membranes. *Proceedings of the National Academy of Sciences* **105**, 17367-17372 (2008).
92. Rinia, H.A. et al. Visualizing detergent resistant domains in model membranes with atomic force microscopy. *Febs Letters* **501**, 92-96 (2001).
93. Soni, S.P. et al. Docosahexaenoic acid enhances segregation of lipids between raft and nonraft domains: 2 H-NMR study. *Biophysical journal* **95**, 203-214 (2008).

94. Niemelä, P.S. et al. Assessing the nature of lipid raft membranes. *PLoS Comput Biol* **3**, e34 (2007).
95. Vetrivel, K.S. et al. Association of γ -secretase with lipid rafts in post-Golgi and endosome membranes. *Journal of Biological Chemistry* **279**, 44945-44954 (2004).
96. Parkin, E.T. et al. Amyloid precursor protein, although partially detergent-insoluble in mouse cerebral cortex, behaves as an atypical lipid raft protein. *Biochemical Journal* **344**, 23-30 (1999).
97. Beel, A.J. et al. Direct binding of cholesterol to the amyloid precursor protein: An important interaction in lipid-Alzheimer's disease relationships? *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1801**, 975-982 (2010).
98. Audagnotto, M. et al. Effect of the Synaptic Plasma Membrane on the Stability of the Amyloid Precursor Protein Homodimer. *J Phys Chem Lett* **7**, 3572-3578 (2016).
99. Audagnotto, M. et al. Protein post-translational modifications: In silico prediction tools and molecular modeling. *Computational and Structural Biotechnology Journal* **15**, 307-319 (2017).
100. Bolduc, D.M. et al. Nicastrin functions to sterically hinder γ -secretase-substrate interactions driven by substrate transmembrane domain. *Proceedings of the National Academy of Sciences* **113**, E509-E518 (2016).
101. Sato, C. et al. The C-terminal PAL motif and transmembrane domain 9 of presenilin 1 are involved in the formation of the catalytic pore of the γ -secretase. *Journal of Neuroscience* **28**, 6264-6271 (2008).
102. Wang, J. et al. C-terminal PAL motif and transmembrane domain 9 of presenilin 1 are involved in the formation of the *Journal of neurochemistry* **96**, 218-227 (2006).
103. Wang, J. et al. Conserved "PAL" sequence in presenilins is essential for γ -secretase activity, but not required for formation or stabilization of γ -secretase complexes. *Neurobiology of disease* **15**, 654-666 (2004).
104. Yu, G. et al. Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and β APP processing. *Nature* **407**, 48-54 (2000).
105. Shah, S. et al. Nicastrin functions as a γ -secretase-substrate receptor. *Cell* **122**, 435-447 (2005).
106. Uemura, K. et al. Reciprocal relationship between APP positioning relative to the membrane and PS1 conformation. *Molecular neurodegeneration* **6**, 15 (2011).
107. Dries, D.R. et al. Glu-333 of nicastrin directly participates in γ -secretase activity. *Journal of Biological Chemistry* **284**, 29714-29724 (2009).
108. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology* **180**, 519-530 (2012).
109. Scheres, S.H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* **3**, e03665 (2014).
110. Audagnotto, M. et al. Effect of the Synaptic Plasma Membrane on the Stability of the Amyloid Precursor Protein Homodimer. *The Journal of Physical Chemistry Letters* **7**, 3572-3578 (2016).
111. Li, S. et al. Initial Substrate Binding of γ -Secretase: The Role of Substrate Flexibility. *ACS Chemical Neuroscience* (2017).

112. P Imbimbo, B. et al. γ -secretase inhibitors and modulators for the treatment of Alzheimer's disease: disappointments and hopes. *Current topics in medicinal chemistry* **11**, 1555-1570 (2011).
113. Tomita, T. Secretase inhibitors and modulators for Alzheimer's disease treatment. *Expert review of neurotherapeutics* **9**, 661-679 (2009).
114. Panza, F. et al. Disease-modifying approach to the treatment of Alzheimer's disease. *Drugs & aging* **26**, 537-555 (2009).
115. Petit, A. et al. New protease inhibitors prevent γ -secretase-mediated production of A β 40/42 without affecting Notch cleavage. *Nature Cell Biology* **3**, 507-511 (2001).
116. Bai, X.C. et al. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**(2015).
117. Nadezhdin, K.D. et al. Structural and dynamic study of the transmembrane domain of the amyloid precursor protein. *Acta Naturae* **3**, 69-76 (2011).
118. de Jong, D.H. et al. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J Chem Theory Comput* **9**, 687-697 (2013).
119. Wassenaar, T.A. et al. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J Chem Theory Comput* **11**, 2144-2155 (2015).
120. Marrink, S.J. et al. The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* **111**, 7812-7824 (2007).
121. Abraham, M.J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).

4 Insights into the molecular mechanism of depalmitoylation catalyzed by human acyl-protein thioesterase 1

Cellular protein trafficking is dynamically regulated by the reversible S-palmitoylation of cysteine residues, however the molecular details of the palmitoylation and depalmitoylation mechanism of soluble and transmembrane substrates are still poorly understood. Here, we describe the molecular mechanism of depalmitoylation by human acyl protein thioesterase 1 (hAPT1) on the basis of new X-ray crystallographic structures, supported by biochemical experiments and multiscale molecular simulations. Our results indicate that the monomeric species is the active enzyme, and suggest a novel mode of interaction with the biological membrane and palmitoylated substrates.

Introduction

Reversible S-palmitoylation (or S-acylation) of cysteine amino acids plays a crucial role in regulating the localization of different cellular proteins. Several cycles of acylation and depalmitoylation control the trafficking of proteins as, for instance, H-Ras and N-Ras localization between the cytoplasm and plasma membrane that in turn modulates signaling regulation¹. Not only soluble, but also membrane proteins required palmitoylation²⁻⁷ for triggering conformational changes⁸, interacting with specific lipids (e.g., cholesterol or sphingolipid-based rafts)⁹, stabilizing membrane association¹⁰⁻¹³ and cross-talking with other post-translational modifications (e.g., ubiquitination or phosphorylation)¹⁴.

Palmitoylation is catalyzed by a family of membrane protein acyltransferases (PATs) usually referred to as zDHHC enzymes due to the presence of the DHHC motif within a zinc-finger cysteine-rich domain¹⁵⁻¹⁷. Depalmitoylation is instead mediated by palmitoyl-protein thioesterase (PPTs) in the lysosome¹⁸ and acyl-protein thioesterases (APTs) in the cytosol¹⁹. APT1²⁰ was the first cytosolic acyl-protein thioesterase structure to be solved and showed a near-canonical α/β hydrolase fold similar to the distant homolog APT1-like (homology of 95%). It is characterized by the presence of the typical α/β hydrolase catalytic triad (i.e., Ser119, Asp174 and His208, **Fig. 4.1a**)²¹ and by a central, predominantly parallel β -sheets region (indicated from $\beta 2$ to $\beta 8$ in **Fig. 4.1a**), which is connected to five α -helices (represented by A to F) and four small α -helices (denoted from G1 to G4). However, compared to the typical α/β hydrolase²¹, APT1 is considerable smaller due to the lack of the first β -strand of the core β -sheets. In particular, the atypical structure of APT1 is characterized by an insertion made up of four short antiparallel β strands (S1-S4), which form an irregular sheet structure.

Interestingly APT1 and APT2 are themselves palmitoylated on Cys2. APT1 can depalmitoylated itself as well as APT2²² rendering the system dynamic. APT2, in contrast, cannot depalmitoylated itself, but was shown to deacylated the palmitoyltransferase DHHC6.

The enzymatic mechanisms responsible for the S-acylation and deacylation process are not fully understood. New co-crystals structures of APT1 and APT2 in complex with their inhibitors (i.e., ML348 and ML349, respectively) have been recently published²³, shedding new light into the isoform-specific inhibition mechanism in APTs. However, the enzymatic mechanism responsible for the reversible S-acylation process is still not well characterized. To address the question of how APT1 depalmitoylates its substrates, we used X-ray crystallography to determine the structures of the full-length, wild-type human APT1 and two of its mutants (C2S and S119A). We further characterized the native form of this enzyme using a combination of multiscale molecular modeling and simulation and biochemical/biophysical experiments.

Results and discussion

Novel X-ray structures of the wild type, C2S and S119A mutants, although belonging to different space groups, showed a similar fold for the monomeric APT1 structure, which agrees with the previously reported structure of APT1 (PDB: 1FJ2²⁰).

Indeed, the average RMSD for the 230 C_α atoms in common among these new structures and 1FJ2 is ~1.2 Å. In the new structures we were able to further capture the conformation of the N-terminal, which was absent in the early reported structure. Different crystal packing allows to stabilize this disorder region showing that the Cys2 is exposed to the solvent and accessible for palmitoylation to occur, as recently experimentally observed²⁴. The asymmetric unit of the crystals contains 2 or 6 monomers. EPPIC²⁵ predicts the possible existence of some biological relevant dimers, for each of our crystal structures, as also predicted for the previous APT1 structure (1FJ2). When these predicted dimers are superimposed (**Fig. 4.1a**), wt and C2S new X-ray structures showed a similar dimeric interface (RMSD ~1.1 Å), while the comparison with the previous APT1 structure (**Fig. 4.1b**) and the catalytic inactive APT1 form (S119A) reveals a largely different dimer conformation (RMSD ~14.2 Å).

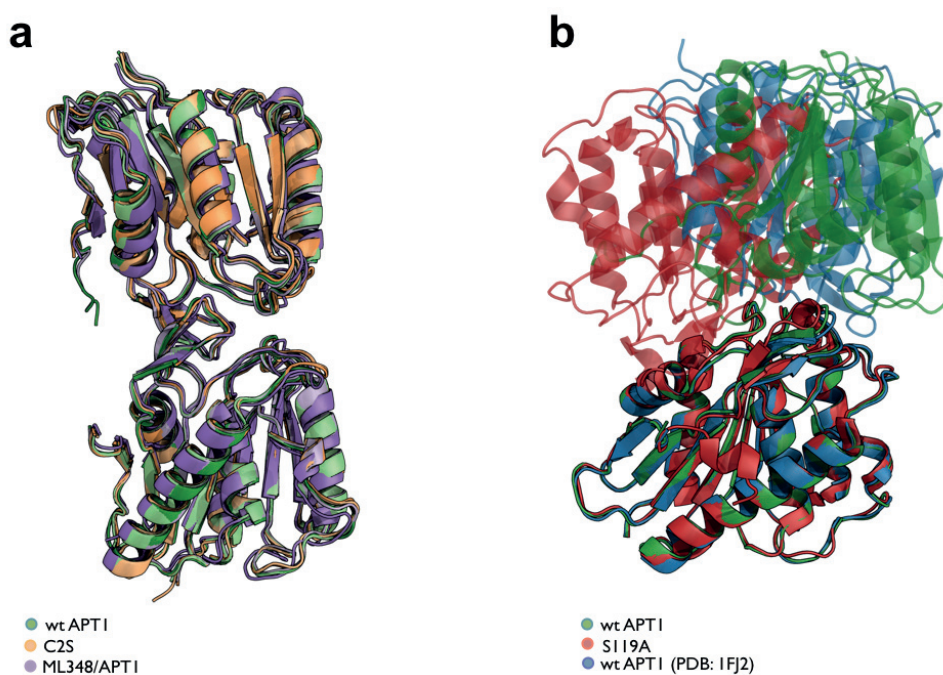


Figure 4.1 | Comparison among dimeric assemblies (a) The comparison among the dimeric structure of wild-type APT1 (green), palmitoylation deficient (orange) and ML348/APT1 (violet) showed the same mutual disposition of the monomers. A different dimeric assembly (b) was instead detected among the previous APT1 structure (PDB: 1Fj2), the wild-type APT1 and the catalytic inactive form presented in this chapter.

The new dimeric interface of wt APT1 is established by interactions among G1 and α F- β 8 loop, S3-S4 loop and α E- β 7 loop as well as inter-subunit interaction between

β 8 loop and G1 (**Fig. 4.2a**). By contrast, in the previous APT1 structure the mutual arrangement of the dimer is different producing a dimeric interface that is not compatible with our new crystallographic structures. The β 3 loop shifts away from the interface due to a rotational motion around the main dimer axis, so that the β 4 loop interacts only with the S2 strand, while the β 8 loop interacts with its counterpart in the other protomer and the β 7 loop forms interaction with α A (**Fig. 4.2b**). Principal component analysis (PCA) based on an atomistic molecular dynamics (MD) simulation of the structure of the wt APT1 (see Methods) highlighted a large rotational motion centered at the dimeric interface (**Fig. 4.2c**). Therefore, all in all, it seems that the dimeric arrangement of APT1 is not unique, as revealed by crystallography, and appears to be intrinsically flexible, as emerged by MD simulation. Interestingly, these dimer interfaces are both embracing the APT1 catalytic pocket, which in turn results partially occluded by the presence of the adjacent protomer.

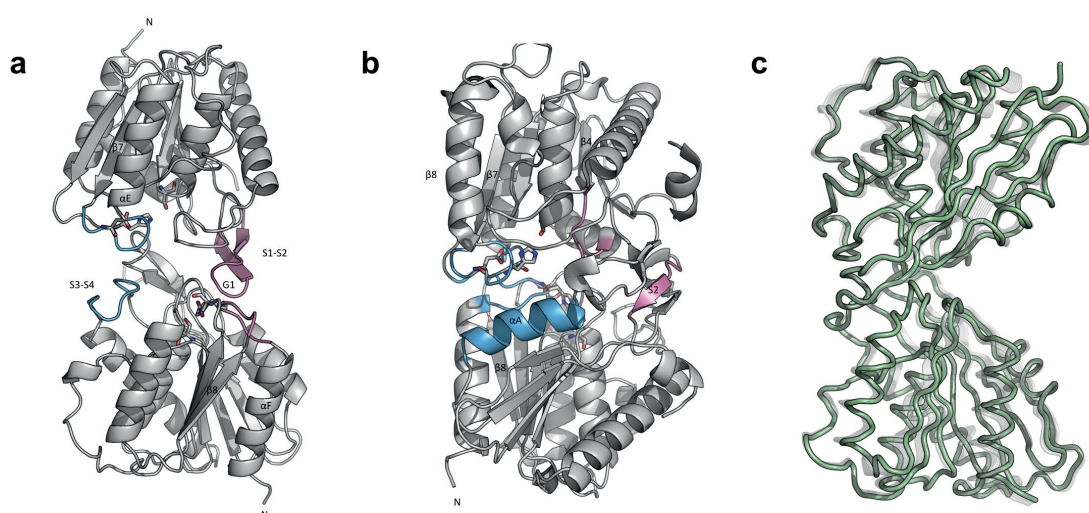


Figure 4.2 | Interface regions analysis Detailed analysis on the residues involved in monomer-monomer interaction for (a) our X-ray structure and (b) the published PDB:1FJ2. The residues at the interface are highlighted in violet and blue. (c) The PCA analysis displayed a rotational motion at the interface, which emphasized the probability of having different dimeric interfaces.

It is unclear whether these alternative conformations are due to the different crystal packing or may be the effect of the presence of the N-terminus in the quaternary structure of APT1 in the new crystals. To gain further insights into the oligomerization properties of APT1 in solution we measured its molecular weight using size-exclusion chromatography (SEC) with multi angle laser light scattering

(MALLS). Although the molecular mass determined by SEC-MALLS is slightly higher (29.3 kDa) than the theoretical molecular weight of an APT1 monomer (~25 kDa), this experiment shows that the enzyme in solution is monomeric at physiological concentration (**Fig. 4.3b**). Thus the dimeric conformations observed in crystals appear to be the result of crystal packing. A monomeric conformation would also be functionally preferable as the active site pocket, otherwise occluded by the dimeric interface (**Fig. 4.3a**), would be accessible for receiving potential substrates to be depalmitoylated.

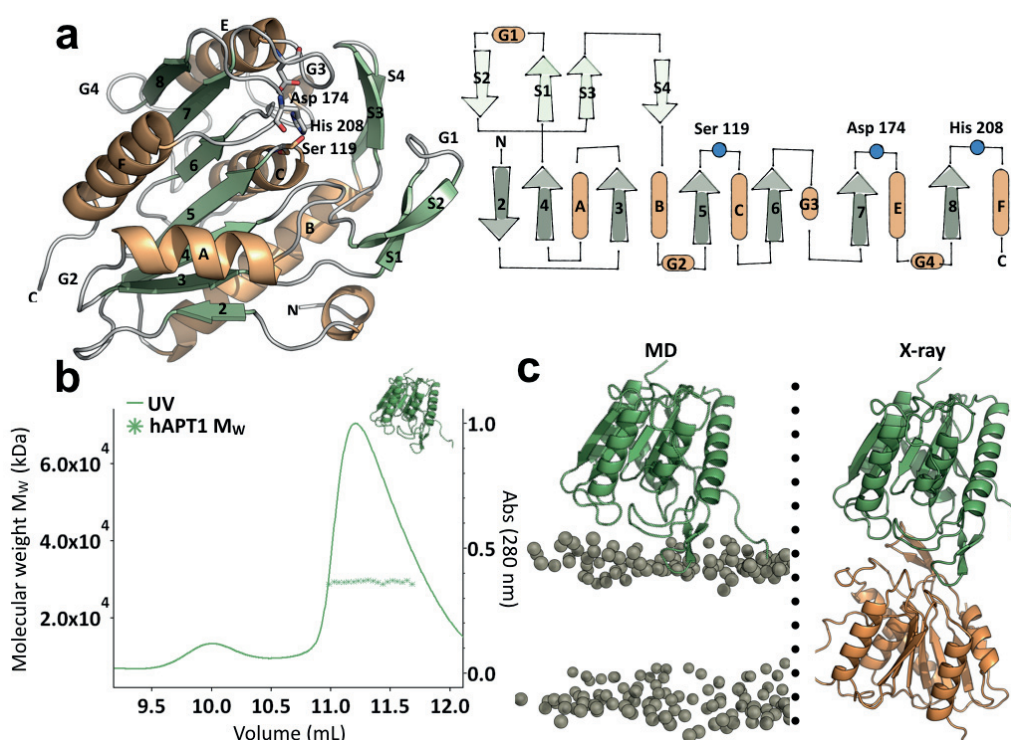


Figure 4.3 | APT1 is a monomer in solution. **(a)** Cartoon representation of the APT1 (left) shows the canonical α/β hydrolase fold. Seven β -strands are numbered sequentially from 2 to 8. Four short irregular β -strands are called S1-S4. Five α -helices are denoted from A to F, while four small α -helices are denoted from G1 to G4. The topology of human APT1 (right) shows the same arrangement and position of the catalytic triad. **(b)** Molar masses calculated from the concentration determined by absorbance measurement ($A_{280\text{nm}}$): one distinct peak, which corresponds to the monomeric form of the APT1 (~25 kDa) was observed. **(c)** Comparison between the average stable conformation at the membrane surface from CG MD simulations (left) with one of the crystallographic dimer conformation (right): the enzyme interacts with the membrane using the same dimeric surface found in crystals.

APT1 is palmitoylated at Cys2 triggering its association with the membrane surface²⁶. To explore this aspect at a molecular level, we performed coarse-grained

(CG) MD simulations of palmitoylated Cys2 (pCys2) APT1 in solution, exploring its interaction with a palmitoyloleoylphosphatidylcholine (POPC) bilayer model (see Methods). The simulations showed that while in solution, pCys2 is protected either in a hydrophobic patch at the protein surface or in proximity of the catalytic pocket. Detailed analysis of the trajectories revealed that the fatty acid, covalently bound to Cys2, interacts with the α B helix of APT1 (**Fig. 4.4a**) and with the catalytic S119 (**Fig. 4.4b**). In particular, the flexible N-terminal segment allows the fatty acid to interact with the hydrophobic patch of the protein composed by α B (i.e., Ile97, Ala99 and Leu100), the parallel β sheets from β 2 to β 4 (Pro12, Leu28 and Pro55, respectively) and some residues in the N-terminal region (Met7, Pro10 and Leu11). These two different ways of protecting the palmitated Cys2 seems to affect also protein-membrane anchoring indicating that pCys2 inserts into the membrane (see **Fig. 4.4d**) only when it is previously protected in the catalytic pocket (see **Fig. 4.4b**).

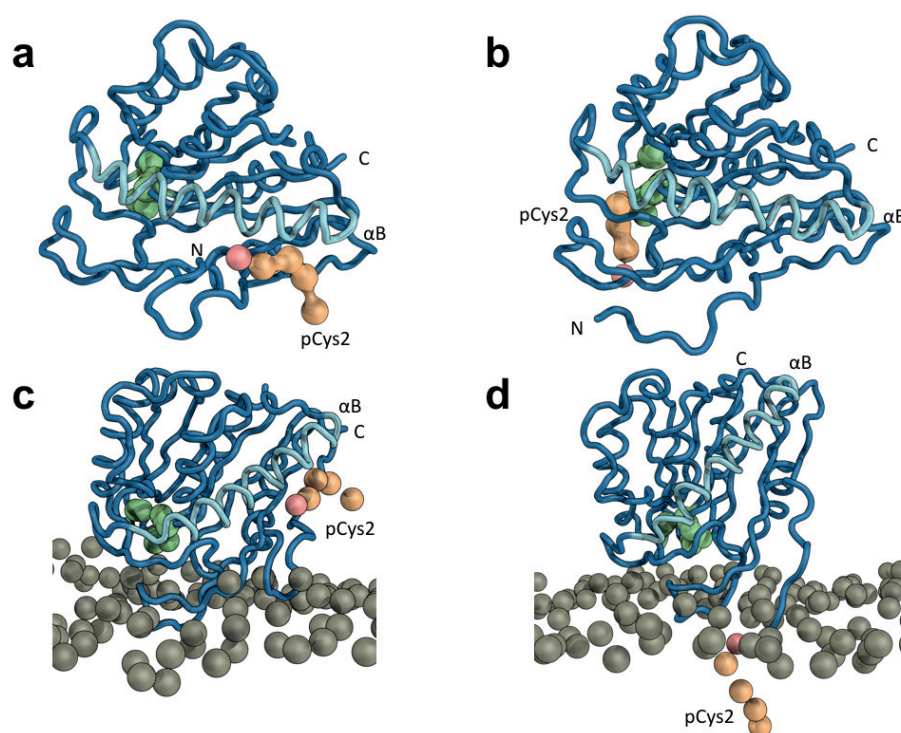


Figure 4.4 | APT1-pCys2 interactions and the resulting membrane anchoring. CG simulations reveal that pCys2 tends to protect itself from the hydrophilic environment in two possible modes: interacting (a) with the hydrophilic helix α B or (b) with the catalytic pocket. These two different kinds on interactions have effects on the protein-membrane anchoring showing the insertion of the pCys2 on the POPC bilayer only when it is previously embedded in the catalytic pocket as show in (c) and (d).

By comparison with other systems, the palmitate protection is reminiscent of what observed for a myristoyl group founded in an allosteric pocket in Abl-kinase²⁷ or the palmitate group in the TEAD transcription factors²⁸. Although in the latter cases, these moieties were deeply buried in a hydrophobic pocket, in the case of APT1, the palmitate tends to protect itself in solution remaining however partially exposed to the solvent. This weak hydrophobic interaction was already observed for the palmitoyl-protein thioesterase 1 (PPT1)²⁹, the equivalent of APT1 in the lysosome.

In five independent CG-MD replicas, APT1 invariably tends to eventually interact with the phospholipid bilayer always via a preferred surface. Interestingly, this surface is the same engaged in the dimeric interface observed in X-ray structures (**Fig. 4.3c**). Independently of pCys2, this membrane-interacting surface presents a highly positively charged that likely promotes the initial association with the membrane by long-range electrostatic attraction (**Fig. 4.5**). Once close to the membrane, APT1 uses a specific, mostly hydrophobic, region defined by β 4- β 5 sheets and G1 helix (hereafter called “ β -tongue”) to anchor further the enzyme to the membrane surface (**Fig. 4.5**), and might have a role in recruiting transmembrane protein substrates for promoting more efficient depalmitoylation. Interestingly, this architecture is not conserved in all the isoforms of APTs: isoform 2 of APT1, for instance, misses the β -tongue region (residue 57 to 72) influencing probably membrane association and, consequently, enzyme activity. The importance of a similar β -tongue motif was also observed for the pore-forming toxin cytolysin A (ClyA), where it was necessary for membrane association³⁰.

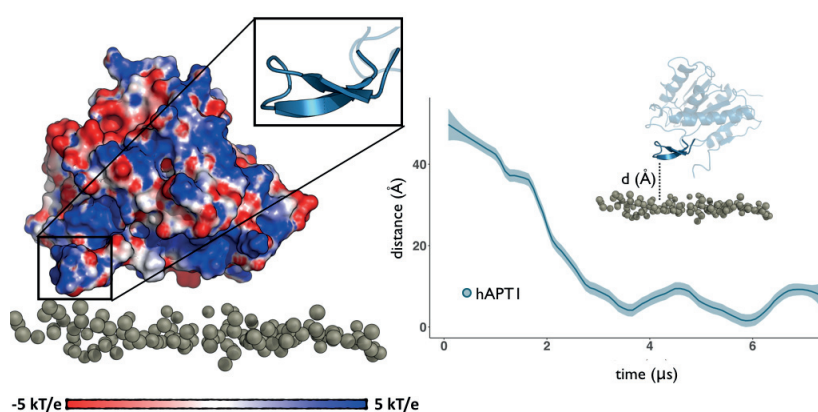


Figure 4.5 | Membrane-protein interactions. The electrostatic surface potential shows the highly positive surface (blue) that promotes the interaction with the lipid bilayer. In particular, the β -tongue has a fundamental role in targeting APT1 to the membrane as shown by distance plot (right) that monitored the β -tongue-membrane interaction over the simulation time.

The most intriguing and novel finding from the crystal structures reported here was that the catalytic pocket is occupied by a fatty acid in about one third of the APT1 proteins in the asymmetric unit (**Fig. 4.6**). Clearly evident in the electron-density maps is a palmitic acid (C16:0) with the carboxylate group facing the catalytic Ser119. Extractions of ligand with ethanol followed by ultraperformance liquid chromatography (UPLC) quadrupole-time-of-flight massspectrometry (Q-TOF-MS) analyses confirm the presence of the palmitic acid in all the three structures (wt APT1, C2S and S119A) (**Fig. 2c**). The fatty acid was not provided during APT1 expression in *E. coli*, purification or crystallization so that the palmitic acid was probably retained during purification as previously observed in literature³¹.

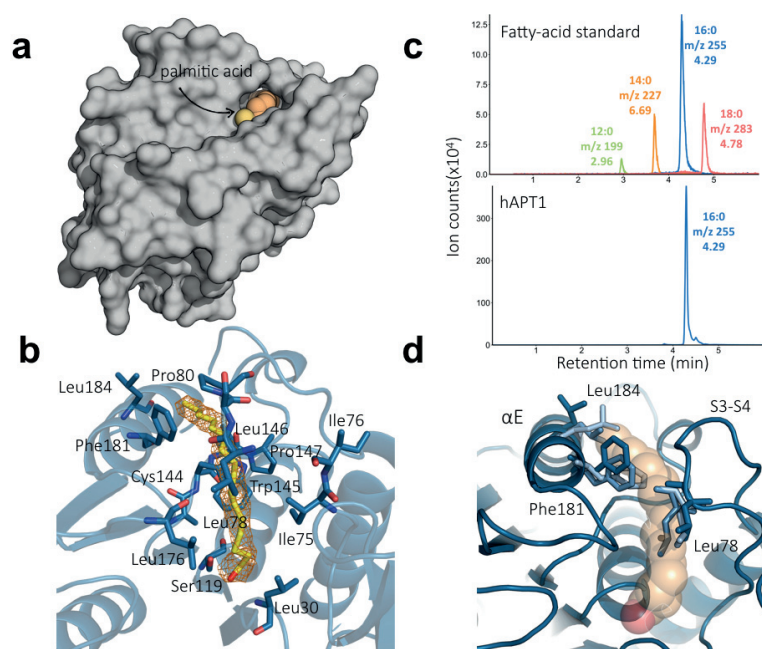


Figure 4.6 | An extended catalytic pocket retains a palmitate moiety in APT1. **(a)** Co-crystal structure of APT1-palmitate. Bound C16:0 is shown in van der Waals representation, where carbons are in orange and oxygens in yellow. **(b)** Close-up view of the hydrophobic channel of APT1 with the experimental electron density of C16:0 superimposed at 1 σ . The side-chain of residues that form the hydrophobic cavity are shown as blue sticks. **(c)** Analysis of the ligands associated with purified APT1 enzyme detected by Q-TOF-MS. **(d)** Close-up view of the structural alignment of apo-APT1 and palmitate-APT1 structures where the C16:0 molecules is represented with van der Waals spheres. Significant conformational changes are observed for Leu184, Phe181 and Leu78.

The binding site of the fatty acid encompasses a largely non-polar cavity that sequesters the aliphatic chain of the palmitic acid. The residues composing the

hydrophobic binding pocket belong to several secondary structural elements. The entrance of the hydrophobic pocket is located close to the loop between S3-S4 (Leu78, Ser79 and Pro80) and is solvent accessible. The top of the channel is formed by residues from αE (Phe181 and Leu184), while the residues Cys144, Trp145, Leu146 and Pro147 contribute to the lowest part of the pocket. Finally, several residues near the active site form the rest of the channel: Leu30, Ile75, Ile76 and Leu176 (**Fig. 4.6c**). This hydrophobic channel shows some relevant conformational changes of key amino acids when compared with the apo form of APT1. There, the hydrophobic channel is not as open due to the movement of the S3-S4 loop and adjacent αE helix. In particular, the rearrangement of the residues Leu78, Phe181 and Leu184 modulates the area of the pocket (**Fig. 4.6d**). Indeed, the difference between the area of the opened (878 Å²) and the closed channel conformation (680 Å²) corresponds roughly to the area occupied by the residues responsible for regulating the access of the lipid. Since all these residues are highly conserved among APTs from different species, this pocket likely contributes to the specificity of APTs for palmitate (**Fig. 4.7**).

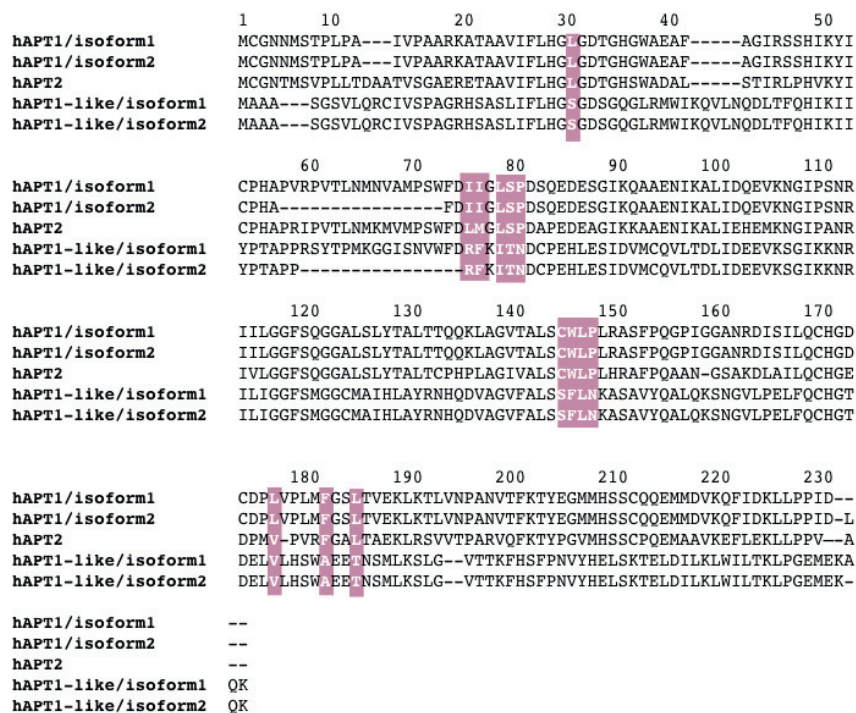


Figure 4.7 | The hydrophobic residues of the extended catalytic pocket are highly conserved in the APT family. Sequence alignment of the 5 members of the APT family suggests that the palmitoylation activity may be conserved among the protein family members.

The presence of a palmitate moiety in the catalytic pocket indirectly reveals novel mechanistic insight into the depalmitoylation reaction. A palmitic acid is one of the final products of depalmitoylation. Thus our structures capture a product-like state, where, after the hydrolytic cleavage catalyzed by Ser119 of a pCys residue of a given substrate, the palmitate moiety is kept in the pocket. Interestingly, as from the analysis of the occupancy in the different APT proteins in the asymmetric unit of all our crystals, the probability of finding a palmitic acid in the hydrophobic pocket is ~33%, hinting to a moderate binding affinity in order for the palmitate to eventually leave the pocket and allow APT1 to start another reaction cycle.

Consistent with this picture, we solved the structure of APT1 in complex with a non-covalent inhibitor, ML348, which is supposed to have a strong affinity for APT1 ($K_i = 280 \pm 33 \text{ nM}^{23}$). First, our structure and therein position of ML348 are consistent with the recently reported structure by Martin and coworkers²³; second, the inhibitor occupies exactly the same hydrophobic pocket of the palmitate moiety in our structures (**Fig. 4.8d**) showing a similar displacement of key residues Leu78, Phe181 and Leu184 previously discussed (**Fig. 4.8c**). Finally, all APT1 proteins in the unit cell are occupied with ML348, suggesting a much larger affinity of the inhibitor compared with the palmitate moiety. The binding energy estimated using MM-PBSA calculations on the ML348/APT1 co-crystal structure is 5 kcal/mol higher for ML348 compared to the palmitic acid, supporting the hypothesis that the palmitic acid provides a mild product inhibition.

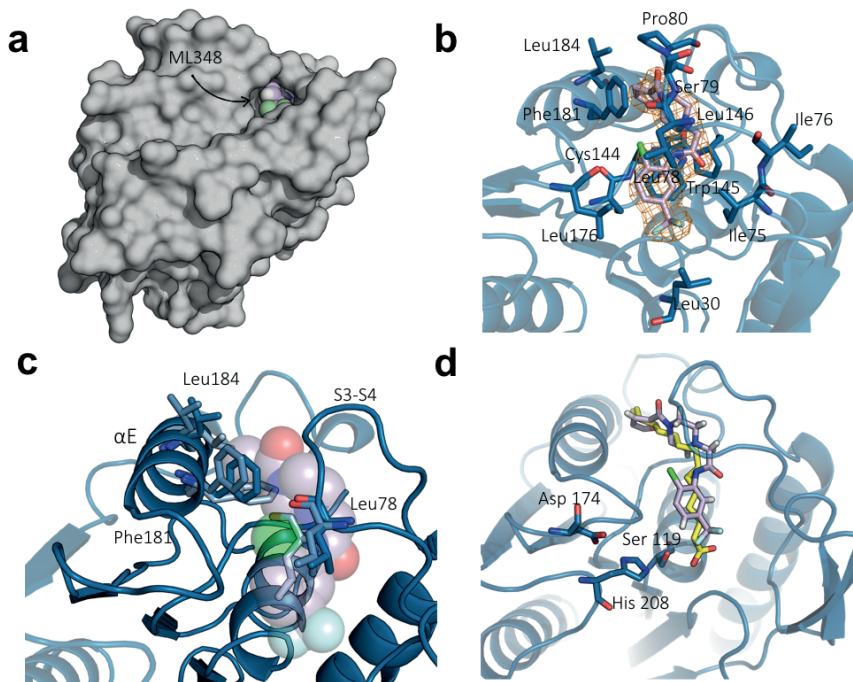


Figure 4.8 | The APT1 inhibitor occupies the extended catalytic pocket. (a) Co-crystal structure of APT1/ML348. The bound inhibitor is shown as van der Waals spheres where carbons are violet and fluoride atoms are green. (b) Close-up view of the hydrophobic channel of APT1 with the experimental electron density of ML348 superimposed at 1σ . The side-chain of residues that form the hydrophobic cavity are shown as blue sticks. (c) Close-up view of the structure alignment of apo-APT1 and ML348-APT1 structures where the ML348 molecules are shown as van der Waals spheres. Significant conformational changes are observed for Leu184, Phe181 and Leu78. (d) Superimposition of APT1/ML348 with APT1/palmitate reveals that the two molecules occupy the same hydrophobic catalytic pocket.

APT1 is a prototypical example of α/β hydrolase, which acts on palmitoylated substrates. The general mechanism for α/β hydrolase catalysis (**Fig. 4.9**) goes through a first step, in which the catalytic serine is activated by deprotonation assisted by a nearby interacting aspartate and histidine in the catalytic pocket (**ES***). Consequently, the nucleophilicity of the catalytic serine is enhanced favoring the nucleophile attack of Ser hydroxyl on the carbonyl of the palmitate and the consequent formation of the acyl-enzyme intermediate (**ES**). Furthermore, a second nucleophile (i.e., H_2O) attacks the acylated enzyme leading to product release and regeneration of the catalytic site (**EP**).

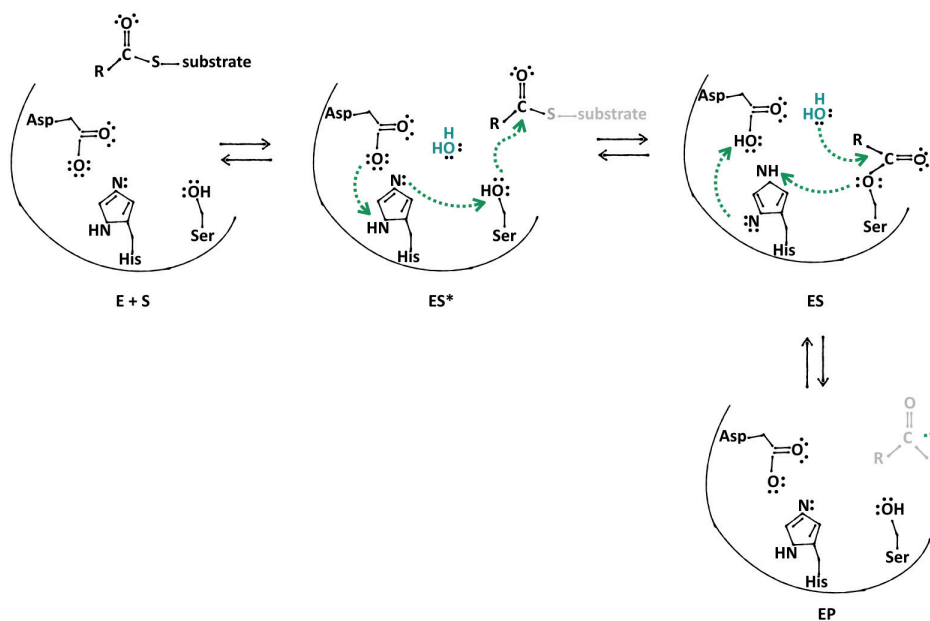


Figure 4.9 | The α/β hydrolases catalytic mechanism. The general mechanism involved during α/β hydrolases catalysis is activated by the deprotonation of Serine assisted by Aspartate and Histidine (ES^*). Consequently, the nucleophilicity of the Serine is enhanced favouring a nucleophile attack of the oxydril of the Serine to the carboxyl of the fatty acid forming an acyl-enzyme adduct (ES). The product is finally released upon nucleophile attack of water and regeneration of the catalytic (EP). R is any possible acyl chains such as palmitate.

Based on this theoretical mechanism and combining our new X-ray structures with molecular modeling and MD simulations, we propose a more detailed molecular description of the depalmitoylation mechanism catalyzed by APT1. Starting from the highest resolution X-ray structure C2S/APT1 (2 Å) in complex with palmitic acid is possible to reconstruct by molecular modeling (MM) followed by MD the consecutive putative steps of the enzymatic reaction (**Fig. 4.10**). Although it is still unknown which transmembrane substrate is specific for APT1, to illustrate a possible enzymatic mechanism we considered calnexin, a palmitoylated ER chaperone. The relative orientation of the APT1 with respect to the membrane was based on the average stable CG-MD conformation, where the β -tongue is located as to direct the enzyme to the membrane surface (**Fig. 4.5**). In the initial stage of the reaction (**Fig. 4.10a**), the orientation of the APT1 hints to a role of the β -tongue as possible substrate recruiting domain. Indeed, this β -tongue is critical in many protein-protein interactions (PPI) and many proteins used it for molecular recognition (e.g., antibodies and T cell receptors)³²⁻³⁴. The palmitate density observed in the C2S/APT1 crystal structure provided us with structural information to model the putative substrate-enzyme state (ES^*), where the palmitate moiety is extended by the covalently linked cysteine in

order to reconstruct the full substrate into the catalytic pocket. It is worth noticing that the structural reconstruction of the **ES*** state from the density of the palmitate from the crystal resulted in an optimal accommodation in the pocket with the catalytic triad perfectly oriented to initiate the enzymatic reaction. In particular, the H δ of His208 points to the Asp174 while the N ϵ of the His208 is turned towards Ser119 hydroxyl (**Fig. 4.10b**). Given this conformation, we propose that APT1 is able to recognize pCys groups at the membrane surface, extract them from the bilayer and direct them inside the extended catalytic pocket to initiate the reaction. This mechanism can likely require the action of the β -tongue and the direct membrane-protein interactions.

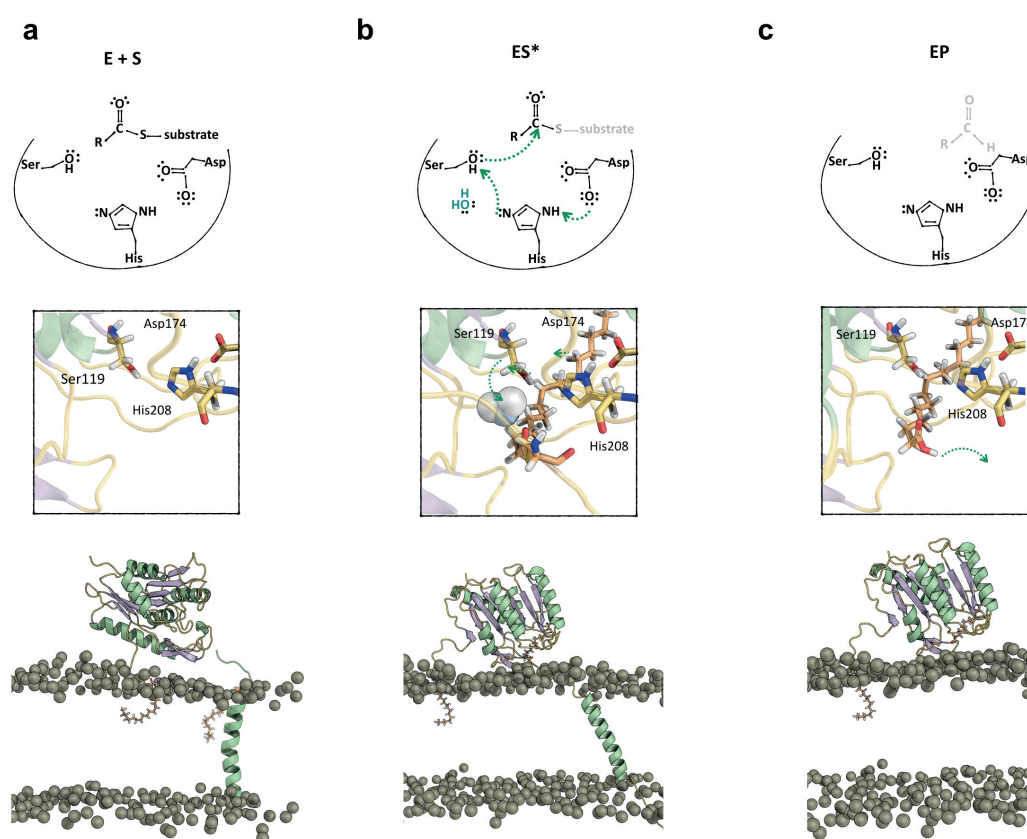


Figure 4.10 | Schematic of the membrane-assisted depalmitoylation mechanism. (a) APT1 anchored by pCys2 at the membrane surface recognizes palmitated membrane protein substrates (**E+S** state). In the inset a close-up of the catalytic pocket with the Ser119, Asp174 and His208 triad properly oriented for accommodating the substrate. (b) Enzyme-substrate complex (**ES***) model based on the relative position of the palmitate in the C2S/APT X-ray structure. The close-up view of the catalytic pocket shows the relative orientation of the catalytic triad as well as the position of the nucleophilic water (drawn in van der Waals representation). (c) Release of the depalmitated substrate into the membrane, while the fatty acid remains in the catalytic pocket forming the products state (**E+P**). Only upon further release of the palmitate from the extended catalytic pocket the enzyme is ready for a new cycle.

Interestingly, the C2S/APT1 structure obtained at high resolution allowed us to confidently refine water molecules. In particular, in the catalytic oxyanion hole we observed structured water positioned between Ser119 hydroxyl and the carbonyl of palmitate (**Fig. 4.10b**), optimally located for the second nucleophile attack (**Fig. 4.9**). After completion of the reaction, the depalmitoylated substrate is readily released at the membrane, while the cleaved palmitate moiety remains in the extended catalytic pocket, and will be later released for a new catalytic cycle.

Conclusion

During the past years, two studies have investigated the mechanism that regulates the access of APTs to their substrates at cellular level^{19,35}. In particular, it was observed in HeLa cell that the active APT1 is located in the membrane fraction. This observation is consistent with the active form of the enzyme being acylated³⁶.

We show here that APT1 is a monomer in solution, in contrast to the dimeric organization found in the crystallographic unit cell in which the entry into the hydrophobic pocket is occluded. The comparison of different crystal structures shows that the dimerization surfaces are heterogeneous, consistent with the observation that dimerization is driven by crystallization. MD simulations predict that the dimerization surfaces participate in membrane association.

APT1 features a positive superficial patch that, together with the β -tongue domain, is responsible for the initial association of the protein with the membrane contributing to enhance recognition of membrane-associated substrates. While, APT1 is predicted to have the ability to interact with the membrane, palmitoylation at Cys2 may serve to stabilize membrane-anchoring and alter its kinetics.

The importance of β -tongue domain for protein-protein interactions (PPI) is broadly documented in literature³⁷⁻³⁹. The β -tongue belongs to a class called β -hairpins, which is the most important secondary structure observed in protein epitopes in nature. Residues that compose the β -tongue of APT1 could be thus responsible for substrate hunting and palmitate-membrane extraction. However, the β -tongue is not present in all the APTs. Isoform 2 of APT1 and APT1-like both miss this important motif, possibly affecting membrane interactions and the resulting enzymatic activity.

Furthermore, APT1 new crystal structures revealed the existence of a larger catalytic pocket, which from the canonical catalytic triad extend to the adjacent hydrophobic pocket identified by residues Phe181 and Leu184 on the top, while the residues Cys144, Trp145, Leu146 and Pro147 contribute to the lowest part of the pocket. A palmitate moiety, which partially populates this pocket in the X-ray structures, hints to the importance of this adduct to accommodate pCys of specific substrates. pCys is thus stabilized passing from solution or extracted from the membrane bilayer forming a reactive enzyme-substrate state (Fig. 3). The fact that the putative substrate and the ML348 inhibitor both occupy the same hydrophobic pocket confirms the importance of this region for APT1 catalytic activity. The ML348 acts as a competitive inhibitor preventing the substrate binding by limiting its access to the catalytic triad. In particular, the comparison between the apo and holo form of the enzyme highlighted significant rearrangement of the protein showing a possible gate-mechanism at the base of substrate recruitment.

In conclusion, our findings provide an atomistic description of the activity of APT1 enzyme proposing a catalytic mechanism that can be valid not only for membrane substrates but also for soluble acylated protein (e.g. H-Ras), and might be more largely generalized for other component of the APTs family.

Computational material and methods

Expression and purification of human acyl protein thioesterase 1 (work done by Sylvia Ho, van der Goot Lab (EPFL))

The protein was expressed in *E. coli* BL21(DE3) (Novagen). The cells were grown at 37°C in 1 L of LB broth containing 50 µg/ml ampicillin to an OD₆₀₀ of 0.6–0.8, and expression was initiated by the addition of IPTG to a final concentration of 1 mM. The culture was grown for an additional 4 h at 37°C. The cells were harvested by centrifugation at 6500 x g for 15 min at 4°C. The pellet was resuspended in ice-cold H buffer (5 mL/g) (150 mM NaCl, 50 mM Tris-HCl [pH 8.0]) and sonicated on ice using 0.5 s bursts with a 1 s cooling period between each burst. The cellular debris was pelleted by centrifugation at 10,000 x g for 30 min at 4°C, and supernatant was used for further purification of hAPT1.

The supernatant was incubated with 10 ml of pre equilibrated (H buffer) Ni²⁺-NTA resin over night at 4°C to bind the His₆-tagged hAPT1. The bound protein was

eluted with H buffer containing 2M imidazole (pH 8.0). The His6-tag was cleaved with His6-tagged rTEV protease (Life Technologies) for 36 hr at 4°C in a Slide-a-Lyser cassette (Pierce) during dialysis against 2L of H buffer to remove imidazole and the cut His6-tag sequence. The dialyzed solution was incubated with 5 ml of Ni²⁺-NTA resin for 2 h at 4°C to remove both uncut protein and rTEV protease. DTT, EDTA, and glycerol were added to a final concentration of 5 mM, 1 mM, and 5%, respectively, and protein was concentrated to a final concentration of 5 mg/ml by using an Amicon filters (10,000 MWCO) (Millipore) in a cold room.

Crystallization and structure determination (work done in collaboration with Florence Pojer, PCRYCF (EPFL))

Crystals of wt/APT1, C2S/APT1 and S119A/APT1 were grown by sitting drop vapour diffusion technique at 4°C from 1:1 mixture of protein solution (50 mM Tris-HCl, pH 8.0, 150 mM NaCl) and reservoir solution. The co-crystals ML348/APT1 were obtained from a 4:1 solution of protein and ML348 that was incubated at 4°C for at least one hour before setting crystal trials. The reservoir solution contained 10% of polyethylene glycol (PEG) 8000, 0.2 M potassium bromide, 0.1 M sodium acetate at pH 5.5; 25% of polyethylene glycol (PEG) 2000, 0.3 M sodium acetate, 0.1 M sodium cacodylate pH 6.5; 25% of polyethylene glycol (PEG) 6.000, 0.1 M 2-(N-Morpholino) ethanesulfonic Acid (MES) at pH 6.5. Crystals growing occurred over a period of 5-15 days.

Crystals were flash frozen by immersion in liquid nitrogen after soaking in cryoprotectant solution (reservoir solution supplemented with 25% w/v glycerol in crystallization buffer). X-ray diffraction data were collected from frozen crystals on X06DA of the Swiss Light Source (SLS, PSI, Villigen Switzerland) and indexed, integrated and scaled with XDS⁴⁰ program. Phase determinations were carried out by molecular replacement using Phaser⁴¹ of CCP4 Suite and the published structure of APT1 (PDB: 1FJ2) as a template. Coot⁴² was used for graphical map inspection and manual rebuilding of atomic models.

Analysis of fatty-acid binding (work done in collaboration with Florence Pojer, PCRYCF (EPFL)).

Bound ligands were extracted from protein sample of wt/APT1 by addition of 500 μ L of Ice-cold HPLD grade ethanol to 100 μ L of protein (414 μ M). The solution was incubated at -20°C for 3 days and subsequently was centrifuged (16,000g at 4°C) and evaporated under vacuum. The residual material was re-suspended in 200 μ L of propanol. Chromatographic separations employed an Agilent Zorbax Extended C18 (2.1 mm x 50 mm, 1.8 μ m particle size) reversed-phase column run at a flow rate of 0.4 mL/min. A linear gradient with initial and final mobile phase consisting of 50% water: 50% acetonitrile and 100% acetonitrile, respectively. The presence of fatty acid was determined by mass determination and comparison with a stock solution (Sigma-Aldrich) composed by 4 different fatty-acid standards (Lauric acid (12:0), Myristic acid (14:0), Palmitic acid (16:0) and Stearic acid(18:0))

Size-exclusion chromatography with multi angle laser light scattering (work done in collaboration with my colleague Giulia Fonti)

The mass measurements were performed on a Dionex UltiMate3000 HPLC system coupled with 3 angles miniDAWN TREOS static light scattering detector (Wyatt Technology). A volume of 85 μ L of wt/APT1, C2S/APT1 and S119A/APT1 at 40 μ M were injected into a Superdex 75 10/300 GL column (GE Healthcare) previously equilibrated with 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 2mM TCEP at a flow rate of 0.5 ml/min. The data were further analyzed using ASTRA 6.1 software.

Atomistic MD simulations.

The initial APT1 conformation of the full-length wt/APT1 is taken from our crystallographic structures and Cys2 was palmitoylated (pCys2) as experimentally reported²². The resulting model was fully solvated with TIP3P water models⁴³ in a water box of dimension 70x70x70 \AA^3 and neutralized by the addition of NaCl at a concentration of 150 mM. The CHARMM36 force field was used for the parametrization of the protein (with CMAP correction). MD simulation was performed with NAMD 2.9 software⁴⁴. The system was minimized for 1000 steps, followed by equilibration in the NPT ensemble for 5 ns at 1 atm and 300 K using a time-step of 2 fs. The pCys2/APT1 system was then simulated for 0.3 μ s in the NPT ensemble. Snapshots were taken at 0.1

ns time intervals for structural analysis. The role of the membrane on pCys2/APT1 stability was investigated by embedding the palmitoylated protein in a pre-equilibrated POPC bilayer (of 80 x 80 Å² size) built with CHARMM-GUI⁴⁵. The system was equilibrated following the protocol reported in ref. ⁴⁶ and simulated for 0.3 μs in the NPT ensemble. In both the systems simulated the periodic electrostatic interactions were computed using particle Mesh Ewald (PME) summation with grid spacing smaller than 1 Å.

Coarse-grained MD simulations.

The atomistic structure pCys2/APT1 and APT1 were coarse grained from our X-ray structures by martinize⁴⁷ and initially positioned 60 Å above the membrane using insane⁴⁸, which generates the POPC bilayer and solvent. The pCys parameters were retrieved from a previous CG study⁴⁹. Five independent replicas, where the protein is aligned with different initial conditions with respect to the membrane, were set up and the resulting systems minimized and equilibrated for 10 ps in NVT using a timestep of 2 fs. Afterward 100 ps of NPT MD with a time step of 20 fs was applied for increasing the temperature and pressure in a range of 300 K and 1 bar, respectively. The temperature was controlled using the Bussi thermostat with a coupling time of 1 ps⁵⁰, while the pressure was controlled by a weak semi-isotropic coupling with a reference pressure of 1 bar and a compressibility of 3 X 10⁻⁴ bar⁻¹. Production MD trajectories were collected for 4 μs using a time step of 20 fs. All the simulations were performed using GROMACS 5.1.2.

BIBLIOGRAPHY

1. Fehrenbacher, N. et al. Ras/MAPK signaling from endomembranes. *Molecular Oncology* **3**, 297-307 (2009).
2. Charron, G. et al. Chemical tools for understanding protein lipidation in eukaryotes. *Current Opinion in Chemical Biology* **13**, 382-391 (2009).
3. Hannoush, R.N. et al. Imaging the Lipidome: ω-Alkynyl Fatty Acids for Detection and Cellular Visualization of Lipid-Modified Proteins. *ACS Chemical Biology* **4**, 581-587 (2009).
4. Ivaldi, C. et al. Proteomic Analysis of S-Acylated Proteins in Human B Cells Reveals Palmitoylation of the Immune Regulators CD20 and CD23. *PLoS ONE* **7**, e37187 (2012).

5. Jones, Matthew L. et al. Analysis of Protein Palmitoylation Reveals a Pervasive Role in Plasmodium Development and Pathogenesis. *Cell Host & Microbe* **12**, 246-258 (2012).
6. Kang, R. et al. Neural palmitoyl-proteomics reveals dynamic synaptic palmitoylation. *Nature* **456**, 904-909 (2008).
7. Li, Y. et al. DHHC5 Protein Palmitoylates Flotillin-2 and Is Rapidly Degraded on Induction of Neuronal Differentiation in Cultured Cells. *Journal of Biological Chemistry* **287**, 523-530 (2012).
8. Joseph, M. et al. Interaction of Peptides Corresponding to Fatty Acylation Sites in Proteins with Model Membranes. *Journal of Biological Chemistry* **270**, 16749-16755 (1995).
9. Levental, I. et al. Palmitoylation regulates raft affinity for the majority of integral raft proteins. *Proceedings of the National Academy of Sciences* **107**, 22050-22054 (2010).
10. Maeda, A. et al. Palmitoylation stabilizes unliganded rod opsin. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8428-8433 (2010).
11. Zevian, S. et al. Structure-Function Analysis of Tetraspanin CD151 Reveals Distinct Requirements for Tumor Cell Behaviors Mediated by $\alpha 3 \beta 1$ versus $\alpha 6 \beta 4$ Integrin. *Journal of Biological Chemistry* **286**, 7496-7506 (2011).
12. Delandre, C. et al. Mutation of juxtamembrane cysteines in the tetraspanin CD81 affects palmitoylation and alters interaction with other proteins at the cell surface. *Experimental Cell Research* **315**, 1953-1963 (2009).
13. Scheffer, K.D. et al. Tetraspanin CD151 Mediates Papillomavirus Type 16 Endocytosis. *Journal of Virology* **87**, 3435-3446 (2013).
14. Abrami, L. et al. Palmitoylation and ubiquitination regulate exit of the Wnt signaling protein LRP6 from the endoplasmic reticulum. *Proceedings of the National Academy of Sciences* **105**, 5384-5389 (2008).
15. Lobo, S. et al. Identification of a Ras Palmitoyltransferase in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **277**, 41268-41273 (2002).
16. Roth, A.F. et al. The yeast DHHC cysteine-rich domain protein Akr1p is a palmitoyl transferase. *The Journal of Cell Biology* **159**, 23-28 (2002).
17. Fukata, M. et al. Identification of PSD-95 Palmitoylating Enzymes. *Neuron* **44**, 987-996 (2004).
18. Verkruyse, L.A. et al. Lysosomal targeting of palmitoyl-protein thioesterase. *Journal of Biological Chemistry* **271**, 15831-15836 (1996).
19. Duncan, J.A. et al. A cytoplasmic acyl-protein thioesterase that removes palmitate from G protein α subunits and p21RAS. *J Biol Chem* **273**, 15830-15837 (1998).
20. Devedjiev, Y. et al. Crystal structure of the human acyl protein thioesterase I from a single X-ray data set to 1.5 Å. *Structure* **8**, 1137-1146 (2000).
21. Wang, A. et al. Regiospecificity and catalytic triad of lysophospholipase I. *Journal of Biological Chemistry* **272**, 22030-22036 (1997).
22. Kong, E. et al. Dynamic palmitoylation links cytosol-membrane shuttling of acyl-protein thioesterase-1 and acyl-protein thioesterase-2 with that of proto-oncogene H-ras product and growth-associated protein-43. *Journal of Biological Chemistry* **288**, 9112-9125 (2013).
23. Won, S.J. et al. Molecular mechanism for isoform-selective inhibition of acyl protein thioesterases 1 and 2 (APT1 and APT2). *ACS Chem Biol* (2016).

24. Kong, E. et al. Dynamic Palmitoylation Links Cytosol-Membrane Shuttling of Acyl-protein Thioesterase-1 and Acyl-protein Thioesterase-2 with That of Proto-oncogene H-Ras Product and Growth-associated Protein-43. *The Journal of Biological Chemistry* **288**, 9112-9125 (2013).
25. Duarte, J.M. et al. Protein interface classification by evolutionary analysis. *BMC Bioinformatics* **13**, 334 (2012).
26. Vartak, N. et al. The autodepalmitoylating activity of APT maintains the spatial organization of palmitoylated membrane proteins. *Biophysical journal* **106**, 93-105 (2014).
27. Nagar, B. et al. Structural Basis for the Autoinhibition of c-Abl Tyrosine Kinase. *Cell* **112**, 859-871.
28. Noland, C.L. et al. Palmitoylation of TEAD transcription factors is required for their stability and function in hippo pathway signaling. *Structure* **24**, 179-186 (2016).
29. Bellizzi, J.J. et al. The crystal structure of palmitoyl protein thioesterase 1 and the molecular basis of infantile neuronal ceroid lipofuscinosis. *Proceedings of the National Academy of Sciences* **97**, 4573-4578 (2000).
30. Mueller, M. et al. The structure of a cytolytic α -helical toxin pore reveals its assembly mechanism. *Nature* **459**, 726-730 (2009).
31. Ngaki, M.N. et al. Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature* **485**, 530-533 (2012).
32. Bouamr, F. et al. Structural and dynamics studies of the D54A mutant of human T cell leukemia virus-1 capsid protein. *Journal of Biological Chemistry* **280**, 6792-6801 (2005).
33. Cornilescu, C. et al. Structural analysis of the N-terminal domain of the human T-cell leukemia virus capsid protein. *Journal of molecular biology* **306**, 783-797 (2001).
34. Sharon, M. et al. Alternative conformations of HIV-1 V3 loops mimic β hairpins in chemokines, suggesting a mechanism for coreceptor selectivity. *Structure* **11**, 225-236 (2003).
35. Dekker, F.J. et al. Small-molecule inhibition of APT1 affects Ras localization and signaling. *Nat Chem Biol* **6**, 449-456 (2010).
36. Ahearn, I.M. et al. FKBP12 binds to acylated H-ras and promotes depalmitoylation. *Molecular cell* **41**, 173-185 (2011).
37. Robinson, J.A. β -hairpin peptidomimetics: design, structures and biological activities. *Accounts of chemical research* **41**, 1278-1288 (2008).
38. Lo, S.C. et al. Structure of the Keap1: Nrf2 interface provides mechanistic insight into Nrf2 signaling. *The EMBO journal* **25**, 3605-3617 (2006).
39. Gavenonis, J. et al. Comprehensive analysis of loops at protein-protein interfaces for macrocycle design. *Nature chemical biology* **10**, 716-722 (2014).
40. Kabsch, W. Xds. *Acta Crystallographica Section D: Biological Crystallography* **66**, 125-132 (2010).
41. McCoy, A.J. et al. Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674 (2007).
42. Emsley, P. et al. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126-2132 (2004).
43. Jorgensen, W.L. et al. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926-935 (1983).

44. Klauda, J.B. et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry B* **114**, 7830-7843 (2010).
45. Jo, S. et al. CHARMM/36 all-atom additive force field for lipids: validation against high-resolution crystal structures. *J Comput Chem* **29**, 1859-1865 (2008).
46. Bovigny, C. et al. LipidBuilder: a framework to build realistic models for biological membranes. (ACS Publications, 2015).
47. Bulacu, M. et al. Improved angle potentials for coarse-grained molecular dynamics simulations. *Journal of Chemical Theory and Computation* **9**, 3282-3292 (2013).
48. Wassenaar, T.A. et al. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J Chem Theory Comput* **11**, 2144-2155 (2015).
49. de Jong, D.H. et al. Molecular view on protein sorting into liquid-ordered membrane domains mediated by gangliosides and lipid anchors. *Faraday Discuss* **161**, 347-363 (2013).
50. Bussi, G. et al. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101 (2007).

5 Conclusions and Outlooks

Complexity is the main keyword in this new age of cell biology. All of us who have run a biological experiment, of any kind, had experienced the feeling that biological systems are extremely complex. Pathways are usually branched and interconnected, proteins usually interact with several partners, membrane are much more complex than a simple cluster of lipids equally distributed between leaflets. Last but not least cellular components are dynamic entities that shuttle between cellular compartment exploiting different functions. Therefore, it is clear that a simple and reductionist modeling, despite providing fundamental information, cannot be other than incomplete and inaccurate for representing the ultimate complexity of the cell.

The level of complexity that regulates biological processes is beyond what we can understand only by conventional structural and molecular biological methods. Hence, a multilevel approach is necessary to achieve a global picture of biological systems necessary for interpreting the multitude of interactions and dynamic events that occur in cells. In this scenario, computational methods are our safety net in this sea of complexity. In the past decades, molecular modeling and simulations, as well as online platforms combined with experimental data (lipidomics analysis, X-ray crystallography, NMR or mass spectrometry, to name few), allowed a more accurate modeling of complex systems such as multi-component membrane bilayers, virus capsids or post-translational modified proteins.

The studies presented in this thesis perfectly mirror this new tendency in biology. In fact, the model of the synaptic plasma membrane, composed by several lipids species differing in both polar heads and acyl tail length, so as in unsaturation level, highlighted the importance of describing with great molecular detail the specific membrane environment for better investigating protein function and stability. Our physical models show the preferential conformation adopted by the amyloid precursor protein (APP) in the membrane, which was in turn not in agreement with the NMR structure, this last one obtained in conditions significantly different from the

physiological ones (Chapter 3). In particular, we showed the important interplay between APP and cholesterol, suggesting not only the key role of this protein in cholesterol biosynthesis, but also the critical role of cholesterol in stabilizing the APP upon cleavage. An additional level of complexity was added to our study by considering another important component in the equation that determines the development of Alzheimer's disease, namely the γ -secretase enzyme that is responsible for A β production. Our results, obtained by carefully modeling realistic membrane conditions, suggested a molecular pathway leading to a plausible APP recruitment mechanism, which is in line with the experimentally observed central role of the nicastrin domain.

This multicomponent model could be the starting point for deciphering two correlated important problems in the field of Alzheimer's disease: (i) the cleavage process of APP that lead to the A β production and (ii) the action mechanism of γ -secretase inhibitors. Currently, the hypothesis that the APP cleavage process is based on a large conformational rearrangement of γ -secretase, pivoting around the extracellular region of NCT. Principal component analysis performed on our simulations (data not shown) highlighted indeed a potential movement of the NCT domain that could resemble the rearrangement required for bringing the APP close to the catalytic site in PS1. These different conformations, along with further computational and experimental validation, could provide hints about some possible mechanism of inhibition, where the γ -secretase rearrangement could be either hindered by compounds binding to the active site or by blocking substrate recognition through binding to the NCT domain. While waiting for novel structural data, simulations can help exploring some of these aspects, always given the intrinsic limitations associated with sampling required to better characterize protein conformational space at the membrane. In general, the lateral diffusion coefficient of lipids in membrane ($D \cong 10^{-8} \text{ cm}^2/\text{s}$)¹ is at least three order of magnitude smaller than the diffusion coefficient of water ($D \cong 10^{-5} \text{ cm}^2/\text{s}$)², which results in larger simulation time needed. Moreover, for a proper evaluation of the influence of lipid compositions on proteins dynamics multiple replicas are necessary for adequately sampling relevant configurations of the systems. Therefore, coarse-grained approaches coupled with enhanced sampling techniques (e.g. metadynamics or umbrella sampling) are valid strategies to overcome sampling limitations. For instance, umbrella sampling

and coarse-grained MD simulations have been recently applied for evaluating the protein-lipid-protein interaction for GpA³ and NaNC⁴.

The next layer of complexity introduced in this thesis was represented by the role of post-translational modifications in membrane-protein interactions. Using a multidisciplinary approach, we proposed that the catalytic relevant conformation of the human acyl protein thioesterase 1 (APT1) is monomeric rather than the dimer found in the crystallographic unit. Moreover, the co-crystal of APT1 with a palmitic acid an extended hydrophobic pocket adjacent to the catalytic pocket allowed us to unveil some unrecognized features of the enzyme-substrate interactions (Chapter 4). Finally, by combining X-ray crystallography, *in vitro* experiments (e.g. SEC-MALLS) and molecular dynamics simulations, we were able to suggest new molecular details of the enzymatic mechanism of depalmitoylation, which will be at the basis of a more rational design of new APT1 inhibitors.

Having a better understanding of the membrane-enzyme interactions, as well as of the substrate binding mode, we are now in the position to ask some new question: how does the enzyme retrieve this palmitated substrate embedded in the membrane? We showed in Chapter 4 the importance of the β -hairpin (called also β -tongue) in redirecting the enzyme to the membrane surface, but this motif could be useful for substrates recruitment as well. The positive charge of this β -hairpin could represent a probe for finding the extracellular negative region of palmitoylated substrates, as the protein diffuses on the membrane. Once found, the hydrophobic patch of the enzyme could create a hydrophobic region favoring the release of the acyl chain from the membrane, and allocate it to the hydrophobic pocket that we highlighted in the previous chapter. Unfortunately, we do not know yet the specific APT1 substrates that could be used for supporting this theory. However, the structure of APT2, a close homologue of APT1 (roughly 60% of sequence homology) was recently solved⁵, and a APT2 specific substrate was revealed by our collaborators in van der Goot's lab (data not published). Preliminary CG simulations of ATP2 showed that also this enzyme is directed to the membrane by the β -hairpin, as we observed for APT1. Therefore, we are currently exploring our hypothesis, by more specifically investigating APT2, namely designing mutants with modified β -hairpin, whose substrate recognition and catalytic activity could be tested *in vitro* by our collaborators.

This thesis provides new molecular and mechanistic insights into the behavior of proteins in their complex environments. As Arieh Warshel once said, “*Computational modeling is the best tool we have to see how molecules are working*”. Indeed, computational methods are changing the way we explore Nature. Nowadays, techniques like cryo-electron microscopy (EM) and cryo-electron tomography (ET) are contributing to increase the resolution in cell biology studies. Cryo-ET opens windows into cells in order to investigate on their internal, three-dimensional spatial organization. The integration of structural data into stochastic simulations allowed recently for a tomographic reconstruction of *Saccharomyces cerevisiae* and HeLa cells providing testable prediction for further experiments⁶. These new approaches expanded the observed scale of biomolecules within the cell environment, while physical simulations extended the “frozen” observed states to dynamic entities being an indispensable tool for elucidating the structure and dynamics of cells. In this landscape, our multiscale approach for the study of complex and realistic environments represents a complementary method for improving the level of details of several experimental techniques, offering a powerful way for dealing with multi components systems.

BIBLIOGRAPHY

1. Almeida, P.F. et al. Lipid diffusion, free area, and molecular dynamics simulations. *Biophysical journal* **88**, 4434-4438 (2005).
2. Krynicky, K. et al. Pressure and temperature dependence of self-diffusion in water. *Faraday Discussions of the Chemical Society* **66**, 199-208 (1978).
3. Domański, J. et al. Convergence and sampling in determining free energy landscapes for membrane protein association. *The Journal of Physical Chemistry B* (2016).
4. Dunton, T.A. et al. The free energy landscape of dimerization of a membrane protein, NanC. *PLoS Comput Biol* **10**, e1003417 (2014).
5. Won, S.J. et al. Molecular mechanism for isoform-selective inhibition of acyl protein thioesterases 1 and 2 (APT1 and APT2). *ACS Chemical Biology* (2016).
6. Earnest, T.M. et al. Challenges of Integrating Stochastic Dynamics and Cryo-Electron Tomograms in Whole-Cell Simulations. *The Journal of Physical Chemistry B* **121**, 3871-3881 (2017).

6 Curriculum Vitae

Martina Audagnotto
EPFL Doctoral Assistant

PERSONAL INFORMATION

Address: EPFL SV IBI-SV UPDALPE, AAB 014, Station 19, CH-1015 Lausanne

Phone: +41 21 693 06 85; +41 76 220 68 36 (mobile)

Residence Permit: Permit L (date of expiry: 30.06.2017)

ResearcherID: D-4140-2017

ORCID: <http://orcid.org/0000-0003-0555-4080>

E-mail: martina.audagnotto@epfl.ch

Lab website: <http://lbm.epfl.ch>

EDUCATION

École Polytechnique Fédérale de Lausanne
PhD Chemistry and Chemical Engineering

September 2012 – July 2017

Thesis: "Studying membrane-protein interplay by modeling realistic conditions" under the supervision of Prof. Matteo Dal Peraro

University of Turin (Italy)

M.Sc. Department of Theoretical Chemistry

March 2009 - December 2011

Thesis: "Glycine adsorption at anatase surface: at DFT periodic study" under the supervision of Prof. Anna Maria Ferrari.

110/110

University of Turin (Italy)

B.S. Department of Theoretical Chemistry

September 2006 – March 2009

Thesis: "Quantum mechanics study of glycine aminoacid structures and its polymers" under the supervision of Prof. Anna Maria Ferrari.

103/110

TEACHING ACTIVITIES

Supervised several undergraduate students at EPFL for semester projects and during laboratory activities. Managed high-school students at EPFL for the program "Swiss Youth in Science", which consists in a week project to give them a first-hand laboratory experience.

PRIZES and AWARDS

Education Committee Travel Award 2016

Awarded by the Biophysical Society and Education Committee for scientific merit.

Student Research Achievement Award (SRAA) Poster Competition

Selected among hundreds of participants for best contribution from a PhD student to the study of the mechanism that underlies motility of biological systems

PUBLICATIONS IN PEER-REVIEWED SCIENTIFIC JOURNAL

Martina Audagnotto, Thomas Lemmin, Alessandro Barducci and Matteo Dal Peraro, "Effect of the Synaptic Plasma Membrane on the Stability of the Amyloid Precursor Protein Homodimer" (*J. Phys. Chem. Lett.* 7.18 (2016): 3572-3578;
<http://pubs.acs.org/doi/pdf/10.1021/acs.jpcllett.6b01721>)

Martina Audagnotto, Dal Peraro Matteo "Protein Post-Translational Modifications: *In silico* prediction tools and molecular modeling" (*Biotechnology Journal*, 2017, 15: 307-319;
<http://www.sciencedirect.com/science/article/pii/S2001037016301052>)

SUBMITTED BUT NOT YET ACCEPTED/PUBLISHED PUBLICATIONS

Martina Audagnotto, Sylvia Ho, Giulia Fonti, Maria Jose Marcaida, Florence Pojer, Gisou van der Goot, Matteo Dal Peraro "Insights into the molecular mechanism of depalmitoylation catalyzed by human acyl-protein thioesterase" (*Paper in preparation*)

Martina Audagnotto, Alexander Lorkowski Kengo, Matteo Dal Peraro " γ -secretase and APP interplay: how the two main characters of Alzheimer disease go along together" (*Paper in preparation*)

PEER-REVIEWED CONFERENCE PROCEEDINGS

1. Martina Audagnotto, Sylvia Ho, Patrick Sandoz, Nicole Andenmatten, Gisou van der Goot, Matteo dal Peraro. "New Insight into the Catalytic and Inhibition Mechanism of the Human Acyl Protein Thioesterase."
(*Biophysical Journal* 110.3 (2016): 547a-548a;
<http://www.sciencedirect.com/science/article/pii/S0006349515041120>)

2. Martina Audagnotto and Matteo dal Peraro. "Insights into the Stabilizing role of Cholesterol for the Amyloid Precursor Protein".
(*Biophysical Journal* 108.2 (2015): 547a-548a;
<http://www.sciencedirect.com/science/article/pii/S0006349514013034>)

ORAL CONTRIBUTION TO CONFERENCES

- **EUROQSAR 2016 (Verona, Italy):** "New insights into the catalytic and inhibition mechanism of the human acyl protein thioesterase"

- **Invited talk at UCSF in 2016 (San Francisco, California):** "Experimental and computational studies of the catalytic and inhibition mechanism of the human acyl protein thioesterase"

- **American Chemical Society 2016 (San Diego, California):** "New insights into the catalytic and inhibition mechanism of the human acyl protein thioesterase"

- **CECAM: Computational Advances in Drug Discovery 2015 (Lausanne, Switzerland):** “A new druggable pocket for Acyl Protein Thioesterase inhibitors”

- **Biophysical Society Meeting 2015 (Baltimore, US):** “Insights into the stabilizing role of cholesterol and lipids on the Amyloid Precursor Protein stability by MD simulations”
