EPFL

Characterization of Protein-Membrane Interfaces through a Synergistic Computational-Experimental Approach

Alessio PRUNOTTO

École
polytechnique
fédérale
de Lausanne

2020

"...if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

– Richard Feynman, The Feynman Lectures on Physics

Acknowledgements

First of all, I would like to express my foremost gratitude to my advisor, Prof. Matteo Dal Peraro, for giving me the opportunity to do a PhD in his lab. Matteo is not only a great scientist, but also an ideal boss. Throughout these years, he has always encouraged me to work independently, and to think outside the box: he provided me the tools to perform great science, while at the same time granting me freedom to manage autonomously my scientific projects. His suggestions and feedbacks were always passionate but calm, rigorous but encouraging. If my PhD journey has been an experience that I will always remember with pleasure, I mainly owe it to him.

For critical reading of this thesis, and for their appreciated feedbacks, I would like to thank the members of my PhD committee, specifically Prof. Julian Charles Shillcock, Prof. Syma Khalid and Prof. Adrian Mulholland in quality of jury experts, and Prof. Bruno Correia as the jury President.

I would also like to thank my experimental collaborators from the University of Rosario, with whom I learned uncountable new notions regarding wet lab techniques: without this very fruitful partnership, my PhD would have not been the same. In particular, I would like to thank Prof. Alejandro Vila for sharing his infinite knowledge about metalloproteins, and for hosting me in his lab; Dr. Guillermo Bahr and Dr. Lisandro Gonzalez for the countless discussions about science (and politics, and football) and for being perfect hosts during my time in Argentina. In general, I would like to thank all the people from Alejandro's group in Rosario, for welcoming me in the warmest possible way and for making me feel like I never left home.

Another advantage of having Matteo as a boss is that, by will or by chance, he is always able

Acknowledgements

who could not see this moment, but which I'm sure would have made him proud. I would not have reached this important life goal without their support and their sacrifices, and this thesis is mainly dedicated to them.

Lausanne, August 7, 2020 Alessio Prunotto

Abstract

The characterization of biological interfaces is widely recognized as one of the main challenges for modern biology. In particular, biological membranes are nowadays known to be an active environment that allows membrane proteins to perform their work and modulates their function. Integral and peripheral membrane proteins constitute 1/3 of the human proteome, and account for about 50% of the targets of modern medicinal drugs. Despite their remarkable role, their interplay with the membrane is often poorly characterized, mainly because of the limits that the currently available experimental techniques encounter when treating hydrophobic environments. In particular, peripheral membrane proteins are often studied in their soluble version: this approach is highly limiting, as the interaction with the membrane is essential for the activity of these biomolecules.

Here, we show the potential of a combined computational-experimental approach in order to overcome the aforementioned limits. In particular, we use molecular modeling to study two peripheral membrane proteins of interest, and successively design ad hoc wet lab experiments to verify the outcomes and predictions of the simulations. This approach allows to bypass the technical limits and high costs of the wet lab techniques, by guiding the experiments with the data of the computational simulations.

We focused our attention on the following peripheral membrane proteins:

New Delhi metallo-β-lactamase (**NDM-1**). NDM-1 is a bacterial enzyme that causes antibiotic resistance. Within the class of metallo-β-lactamases, it represents the most serious threat to global health. The larger resistance of NDM-1 with respect to other proteins of the same class, has been linked to its post-translational modification, which connects it to the outer bacterial membrane of Gram-negative bacteria: this event can significantly increase the chances of NDM-1 to spread through the infection through vesicles excretion. In the present work, we elucidated the mechanistic aspects of the NDM-1/bacterial membrane interaction, and

identified the features that contribute to the efficiency of this mechanism.

Golgi phosphorylated protein 3 (**Golph3**). Golph3 is a peripheral membrane protein present at the Golgi apparatus of most eukaryotic cells. Its normal function consists in binding glycosylating enzymes, and transport them through the Golgi *cisternae*. In humans, Golph3 has been found to be overexpressed in several forms of cancer: however, no Golph3 inhibitors are currently present in the pharmaceutical market. This is mainly due to the lack of structural information regarding the molecular mechanism of Golph3. Here, we clarify the features of Golph3 that allow it to bind to the Golgi, and elucidate the mechanism of membrane binding. We also propose a recognition mechanism between Golph3 and the glycoenzymes, based on events predicted by the computer simulations.

Overall, in the present work we demonstrate the potential of computational-experimental approaches in structural biology, and in particular in the study of peripheral membrane proteins. We show that a combined approach constitutes the best way of overcoming the limits of each technique, and we discuss the repercussions on the study of systems of biological interest.

Keywords: biological interfaces, peripheral membrane proteins, molecular modeling, molecular dynamics, New Delhi metallo-β-lactamase, Golgi phosphorylated protein 3

Résumé

La caractérisation des interfaces biologiques est largement reconnue comme l'un des principaux défis pour la biologie moderne. En particulier, les membranes biologiques sont aujourd'hui connues pour être un environnement actif qui permet aux protéines membranaires d'effectuer leur travail et de moduler leur fonction. Les protéines membranaires intégrales et périphériques constituent 1/3 du protéome humain et représentent environ 50% des cibles des médicaments modernes. Malgré leur rôle remarquable, leur interaction avec la membrane est souvent mal caractérisée, principalement en raison des limites que les techniques expérimentales actuellement disponibles rencontrent lors du traitement des environnements hydrophobes. En particulier, les protéines des membranes périphériques sont souvent étudiées dans leur version soluble : cette approche est très limitante, car l'interaction avec la membrane est essentielle pour l'activité de cettes biomolécules.

Ici, nous montrons le potentiel d'une approche combiné computationnelle-expérimentale afin de surmonter les limites susmentionnées. En particulier, nous utilisons la modélisation moléculaire pour étudier deux protéines membranaires périphériques d'intérêt, et concevons des expériences en laboratoire pour vérifier les résultats et les prévisions des simulations. Cette approche permet de contourner les limites techniques et coûts élevés des techniques de laboratoire, en guidant les expériences avec les données des simulations computationnelles. Nous avons concentré notre attention sur les protéines membranaires périphériques suivantes :

New Delhi métallo-β-lactamase (**NDM-1**). Le NDM-1 est une enzyme bactérienne qui provoque résistance aux antibiotiques. Dans la classe de métallo-β-lactamases, il représente la menace la plus grave pour la santé mondiale. La plus grande résistance de NDM-1 par rapport aux autres protéines de la même classe, a été liée à sa modification post-traductionnelle, qui la relie à la membrane bactérienne externe de bactéries Gram-négatif : cet événement

peut augmenter considérablement les chances de propagation du NDM-1 par l'infection, par l'excrétion des vésicules. Dans cette thèse, nous avons élucidé les aspects mécanistiques de l'interaction entre NDM-1 et la membrane bactérienne, et identifié les caractéristiques qui contribuent à l'efficacité de ce mécanisme.

Golgi protéine phosphorylatée 3 (**Golph3**). Golph3 est une protéine membranaire périphérique présente dans l'appareil de Golgi de la plupart des cellules eucaryotes. Sa fonction normale consiste à lier les enzymes glycosylantes et à les transporter à travers les *cisternae* du Golgi. Chez l'homme, Golph3 s'est révélé surexprimé dans plusieurs formes de cancer : cependant, aucun inhibiteur de Golph3 n'est actuellement présent sur le marché pharmaceutique. Cela est principalement dû au manque d'informations structurelles concernant le mécanisme moléculaire de Golph3. Ici, nous clarifions les caractéristiques de Golph3 qui lui permettent de se lier au Golgi et nous élucidons le mécanisme de liaison membranaire. Nous proposons également un mécanisme de reconnaissance entre Golph3 et les glycoenzymes, basé sur les événements prédits par les simulations computationnelles.

En general, dans cette thèse, nous démontrons le potentiel des approches computationnelles/expérimentales en biologie structurale, et en particulier dans l'étude des protéines membranaires périphériques. Nous montrons qu'une approche combinée constitue le meilleur moyen de dépasser les limites de chaque technique, et nous discutons les répercussions sur l'étude des systèmes biologiques d'intérêt.

Mots-clés : interfaces biologiques, protéines membranaires périphériques, modélisation moléculaire, dynamique moléculaire, New Delhi métallo-β-lactamase, Golgi protéine phosphorylée 3

Contents

Contents

List of Figures

1 Introduction

1.1 Biological interfaces and biological membranes

The definition of biological interface is strongly dependent on the scientific field of interest: in general, biological interfaces are defined as the contact region between two objects, where at least one of the two is constituted by organic material.[1]

In molecular biology, a biological interface is defined as the contact between two biomolecules (proteins, phospholipid constructs, DNA, etc.). The characterization of biological interfaces is of capital importance not only in biology and physiology, but also in engineering and materials science. The scientific community has indeed grown an interest in replicating the mechanisms that have been refined by billions of years of Darwinian evolution, in order to achieve the so-called biomimetic materials.[2,3] The study of biological interfaces is therefore a very multidisciplinary field, involving several different branches of science.

In biology, the characterization of the mechanisms by which biological interfaces act at a molecular level, is of capital relevance. In particular, the aim is both to understand the emergent physiological phenomena that are related, for example, to the rise of new diseases, and to provide useful information that can be developed in novel technologies. The challenges in this field are multiple, and are related both to the characteristic instability of biological interfaces[4] and to the fact that relevant events generally happen at time- and space-scales that are arduous to access through wet lab experiments (order of magnitude of microseconds and nanometers, respectively). In the present work, we tackled these issues

through a combined computational/experimental approach. In particular, we made ample use of molecular modeling, followed by *ad hoc* experimental setups aimed at verifying the predictions of computer simulations. Thanks to this approach, the experiments can be aimed at measuring emerging properties predicted by the simulations, rather than to reproduce the simulations output themselves. This way, there is no need to access the time- and space-scales that are often prohibitive to experimental methods.

One peculiar case of biological interface that is particularly relevant in biology is present in the interplay between biological membranes and proteins. In the present work, we focused on exploiting the features and functions of two peripheral membrane proteins. In order to proceed with the definition of peripheral membrane proteins, we first need to provide an exhaustive definition of biological membranes.

Biological membranes are phospholipid constructs present in cells, often in the form of bilayers.[5] They are ubiquitous in the cytoplasm and in their absence, life as we know it, would not be possible. Biological membranes define the limits of the cell, and protect the organelles and the cell machinery from the pressure of the outer environment. Biological membranes represent indeed one of the best isolating material known in nature, with an electric conductivity equal to ~10^{-24} S/m[6] and a thermal conductivity equal to ~0.25 W/M*K.[7] To provide a term of comparison, Teflon and silica aerogels, which are the best electric and thermal isolators developed by human technology, are characterized by an electric and thermal resistance equal to about 10^{-17} S/m and 0.6 W/M*K, respectively.[8,9]

Several theories have been proposed to explain how the first biological membranes formed, and how they evolved to become one of the main actors in cell biology.[10,11] The matter is still object of discussion, but it seems likely that current membranes evolved from small vesicles, which were formed in the prebiotic broth as a simple chemical reaction to the pressure of the surrounding environment. In particular, phospholipids, which are amphipatic molecules, tend to react to the presence of polar solvent by arranging into ordered structures, which minimize the exposition of their hydrophobic surfaces (which, in phospholipids, are represented by the acyl tails) and maximize the exposition of the hydrophilic ones (in phospholipids, polar heads with a phosphate group). Successively, these vesicles encapsulated

(probably by chance) pieces of self-replicating material (RNA, the first form of life). Darwinian mechanisms positively selected these constructs, as the encapsulated RNA was protected from the outer environment, and therefore more likely to perform its self-replicating activity. As a consequence, the most replicating (and hence successful) RNA molecules were the ones whose features made them most likely to "recruit" phospholipid vesicles as a screen from the harsh conditions of the surrounding, prebiotic environment (Figure 1.1a). These primordial structures constituted the first prototype of what would have later evolved into proper cells.

Biological membranes were also responsible for one of the most relevant events in the history of life on Earth, that is the differentiation of eukaryotic cells. Together with viruses, bacteria and *Archaea* were the first forms of life present on Earth: despite lacking any internal compartmentalization, they disposed of a biological membrane to isolate and protect themselves from the surrounding environment. Their capability to produce more membranes was highly limited by the lack of tools to transform and use chemical energy in order to perform useful work. A dramatic change in this sense happened when the first mitochondria appeared, an event which marked the beginning of a new era in the history of life.[12] It is very challenging for scientists to estimate both when and how this occurred. Recent fossil discoveries found eukaryotic cells that are 2.2 billion years old.[13] The way mitochondria appeared is not clear either, but the most credited theories, based on genetic studies, state that mitochondria were originally separated prokaryotic cells that developed oxidative mechanisms, and that were later acquired from the first eukaryotic cells as endosymbionts (endosymbiotic theory[14]). Mitochondria, which are considered the powerhouse of cells, allowed the first eukaryotic cells to make use of chemical energy, coming from ATP hydrolysis, to build a huge extra number of molecules, among which the lipids that are the building blocks of membranes. This increased availability of lipids allowed to the eukaryotic cells to develop differentiated compartments, each with a precise task, such as protecting the genetic material for the nucleus, assembling proteins for the endoplasmic reticulum, molecular transport for the Golgi, etc. This brought to immense evolutionary advantages, which subsequently allowed the formation of the first multicellular organisms. Indeed, biological membranes not only define the limits of eukaryotic cells, but are also ubiquitous within the cytoplasm: indeed, several key organelles are made of phospholipids in different forms. For example, mitochondria, Golgi apparatus, endoplasmic
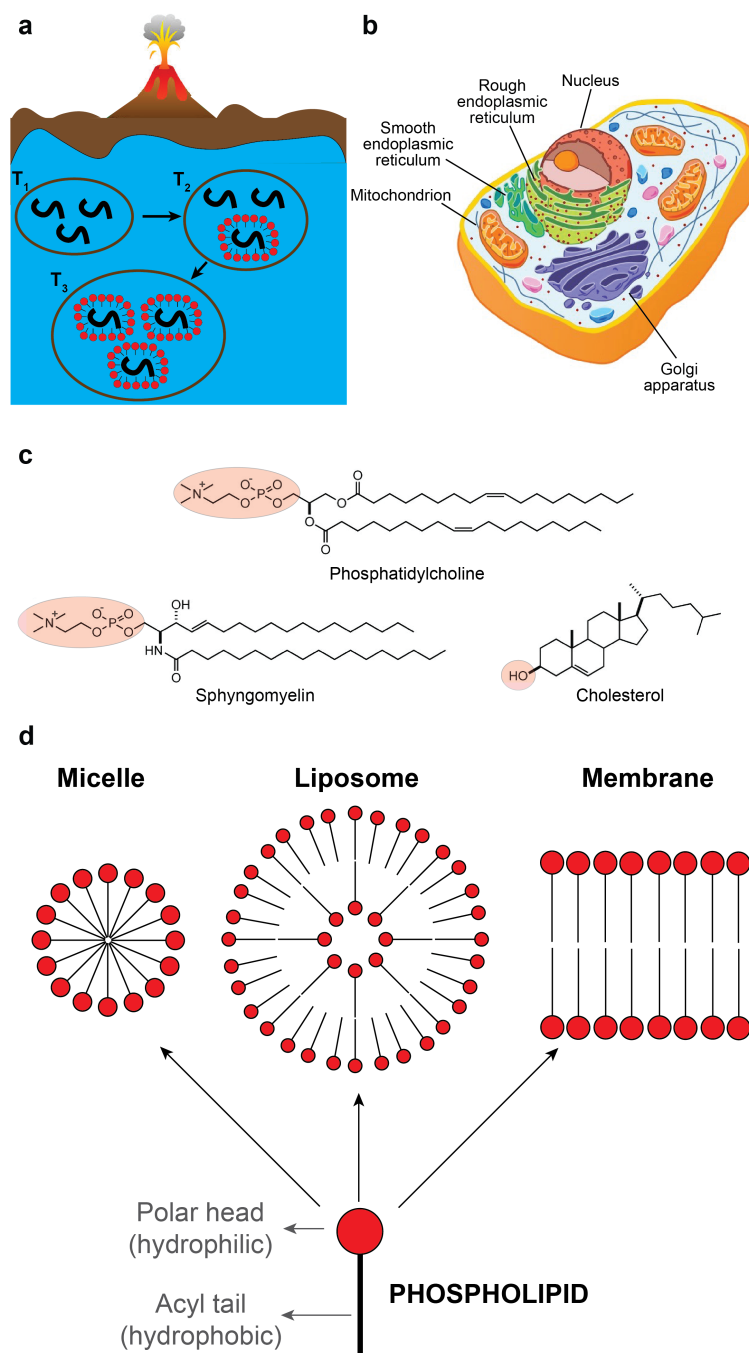
Figure 1.1 – (a) Formation and evolution of biological membranes: initially (T_1), the first forms of self-replicating material (RNA) were not surrounded by any lipid layer. At a later stage (T_2), some RNA would have been incapsulated within lipid vesicles. Having an evolutionary advantage, the self-replicating material that was more prone to be encapsulated, would have survived and replicated faster and better (T_3). (b) Schematic representation of a cell, with some of the organelles that are primarily constituted by lipid layers (Figure by *Encyclopedia Britannica*). (c) 2D chemical structure of some of the main cell lipid components. The polar head is highlighted in red. (d) Schematic representation of a phospholipid and of the lipid constructs that are found in a cell.

reticulum and the nuclear membrane are all mainly constituted by phospholipids[15] (Figure 1.1b). These membranes are not identical, but they differ for a large variety of features: for example, they can be composed by different lipid types,[16] or contain a different amount of membrane proteins or carbohydrates (mainly glycolipids and glycoproteins).[17] These differences directly influence a large number of mechanical and physical properties in membranes, such as the viscosity,[18] the mechanical stability[19] or the ion conductance.[20,21] Moreover, these features also have a strong impact on the geometrical features of membranes, such as size and curvature (for example, the smooth endoplasmic reticulum is characterized by a much wider curvature, if compared to the rough endoplasmic reticulum).

Although several different kinds of lipids compose biological membranes, it is possible to identify 3 types that are present in a dominant way (Figure 1.1c): glycerophospholipids, sphingolipids and sterols.[22] Chemically speaking, glycerophospholipids are glycerol-based molecules, characterized by a polar head which contains a phosphate, and by fatty acids that are hydrophobic, hence making these molecules amphipatic. The number of fatty acids that are linked to the polar head can vary, but is generally in a number of 2. Sphingolipids are also amphipatic and their structure is characterized by the presence of a sphingoid based backbone. Sterols are a subgroup of steroids: among them, the most important molecule is cholesterol, which is present in a relevant amount in the plasmamembrane, and has a significant impact on its viscosity.[23]

Biological membranes gained world-wide notoriety within the scientific community in 1972, when Singer and Nicolson first introduced the *fluid mosaic* model[24]: this fundamental study highlighted that the membranes were not just a passive and static environment where proteins could perform their activity, but rather an active entity which directly participates in the modulation of proteins properties and activity. This discovery drove most of the successive studies on biological membranes: it was then highlighted how membranes can actively communicate with proteins and the cytoskeleton through hydrophobic and Coulomb forces, thus bringing messages and therefore representing a key-element for the life cycle of cells.[25]

The lipid composition strongly depends on the lipid synthetizing enzymes that the cell disposes. Interestingly, it has been found that the lipid composition of eukaryotes share many

more similarities with bacteria, rather than with archaea.[26] Since the evolutionary division between bacteria and archaea happened at an earlier stage with respect to the appearance of eukaryotes, it is reasonable to assume that the eukaryotes likely evolved from bacteria, rather than from archaea.

The chemical and physical features of the lipids define the properties of the membrane itself. For example, lipids can differ in their electrical properties (in particular their polar head can be zwitterionic, like phosphatidylcholines and phosphatydilethanolamines or anionic, like cardiolipins and phosphatydilserines); in the length (i.e. number of carbons) of their acyl chains; or in the number and position of insaturations present in the acyl chain. These properties influence parameters of the membrane such as the curvature, thickness and viscosity.

Since lipids are central for regulating protein activity, it is reasonable that imbalances from their physiological proportions can lead to severe pathologies: indeed, changes in the metabolism of lipids has been linked to serious diseases, such as cancer,[27] metabolic diseases,[28] neurological[29] and immunological[30] disorders. These are mainly due to a loss of membrane properties that can significantly affect cell division, cell migration, synaptic processes and other functions.[31] In recent years, some therapeutic techniques have been proposed, which aim to modify the lipid composition in membranes. For example, alterations of some enzymatic activities can be employed to modify lipid levels and restore physiological values.[32]

Phospholipid constructs in cells are not limited to bilayers: on the contrary, they can be found in different architectures (Figure 1.1d). For example, they can assume the form of liposomes or micelles.[33,34] Liposomes and micelles are present not only within the cells, but also in the extracellular environment, as they are responsible for transporting cargoes of biological material from one cell to another, or simply to remove materials that the cell does not need anymore.[35] This efficient transport mechanism is essential for alleviating pressure from the cells and for allowing the correct prosecution of their life cycle.[36]

1.2   Membrane proteins

Integral and peripheral membrane proteins represent more than one-third of the human proteome,[37] and are the target for more than 50% of modern medicinal drugs.[38,39] Despite their remarkable role in biology and pharmacology, only 3% of the Protein Data Bank entries refer to membrane proteins.[40] This is mainly due to their hydrophobicity, high flexibility, and lack of stability, which make it extremely challenging to characterize these proteins and their interactions with lipids by means of techniques such as X-ray crystallography, NMR or cryo-EM.[41] A growing number of lipid-related proteins are associated with newly identified disease-related genes,[42] which calls for a functional elucidation in order to improve diagnostics and develop new therapeutic interventions.

Membrane proteins are classified according to the way they interact with the bilayer (Figure 1.2): if they span it completely, hence interacting not only with the polar surface of the membrane, but with its hydrophobic core as well, they are defined as integral membrane proteins. If, on the contrary, the protein interacts with the bilayer from the outside, without spanning the hydrophobic core, we generally refer to them as peripheral membrane proteins. Integral membrane proteins are defined according to the number of domains that span the bilayer, and to their main secondary structure.[43] In particular, a protein that spans the membrane only once, is called bitopic membrane protein. Bitopic membrane proteins generally consist of one single α-helix. If the polypeptide chain, after spanning once the bilayer, makes turns and spans it again with other domains, it assumes the name of polytopic membrane protein. Polytopic membrane proteins are subdivided in two extra groups, depending on the dominant secondary structure. In details, they can be made mainly of α-helices (the most notorious example of protein class of this type is probably constituted by the G-protein coupled receptors family), or of β-sheet (in this case, they are commonly called β-barrels proteins).

The second class of membrane proteins is represented by peripheral membrane proteins (PMPs). PMPs adhere to the membrane through a limited external portion of their surface, while maintaining their main body into the solvent. The adhesion to the membrane can occur through non-specific interactions (i.e. hydrophobic or electrostatic), or through covalent
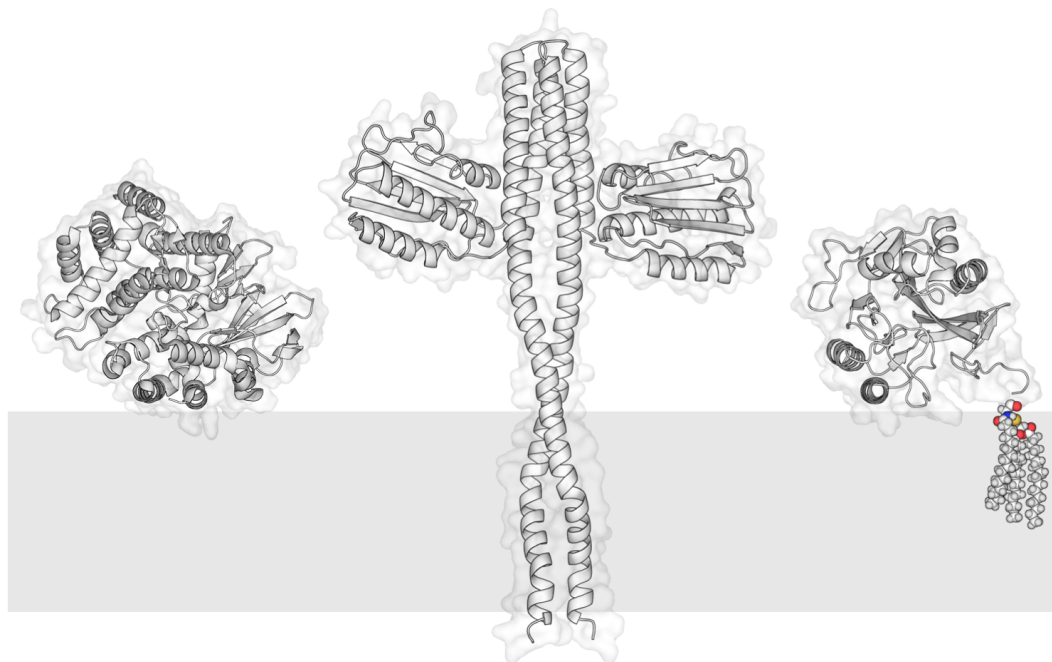
Figure 1.2 – A membrane protein can interact with the lipid bilayer through non-covalent forces from the outer side (left); through hydrophobic forces, by spanning the bilayer completely (integral membrane protein, center); or through a post-translational modification (right).

bonds (i.e. post-translational modifications, such as palmitoylation[44]). In the first case, it is possible to differentiate between PMPs that interact with the membrane through amphipatic domains (typically α-helices or turn-coils), or through electrostatic-mediated interactions. PMPs include a large variety of different proteins, which are involved in a wide range of functions, including transferring small molecules,[45–47] messaging[48] and providing structural stability.[49] Common types of PMPs are for example lipases, whose aim is to catalyze the hydrolysis of lipids; all proteins belonging to the annexin family, which are fundamental in exocytosis and endocytosis events, are also PMPs. With respect to integral membrane proteins, the relationship between PMPs and phospholipid bilayers is even more poorly characterized. This is mainly due to the fact that their adhesion to the membrane is frequently transient, and therefore PMPs are generally studied as soluble proteins.[50]

The Orientation of Proteins in Membrane (OPM) database[51] collects information present in literature concerning how membrane proteins (including PMPs) orient with respect to phospholipid bilayers. If we perform a systematic search to investigate the nature of the PMPs

residues that mediate the interaction with the membrane, we see that they mainly include hydrophobic residues (Figure 1.3). This measure does not significantly differ from the domains in the proteome that do not interact with phospholipids. However, we can also observe that the peripheral domains present a lower amount of acidic residues, with respect to the rest of the proteome. This can be explained in terms of electrostatics, as all biological membranes are negatively charged.[52] Fuglebakk et al.[53] showed that peripheral membrane domains are determined not only by the nature of their amino acids, but also by their shape. After analyzing 300 families of proteins that are classified as peripheral membrane binders, they found that peripheral domains are often characterized by a peculiar protruding shape, that is particularly prone to induce interaction with lipids.
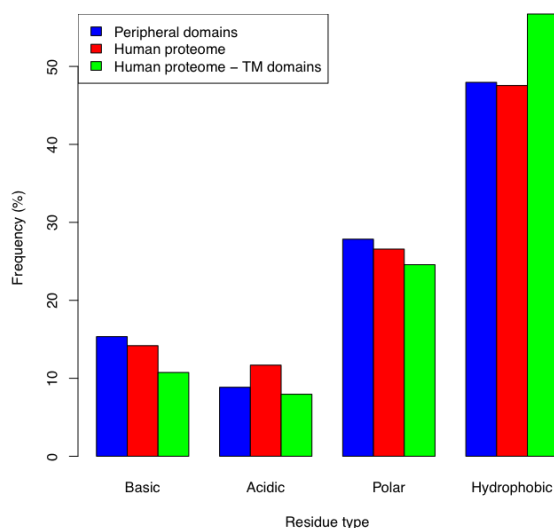
Figure 1.3 – Physical features of residues composing trans-membrane domains in proteins.

1.3 State of the art of experimental methods

In recent times, the scientific community saw the uprising of several, novel experimental techniques, that are considerably transforming the way protein/membrane interfaces are studied.[54] Among them, we can cite cryo-electron microscopy and tomography (cryo-EM/ET),[55] state-of-the-art lipidomics,[56] mass spectrometry (MS),[57] membrane imaging by high-speed atomic force microscopy (AFM)[58] (Figure 1.4). Despite their potency and the undeniable

impact that they represented for structural biology, these techniques still provide just a partial view of the systems of interest: in order to get a more general and complete understanding of protein/membrane interactions, it is essential to combine the information that come from experimental methods, to data from modelling and computational techniques. Among the aforementioned experimental methods, the scientific community widely recognizes that cryo-EM/ET is the most promising one[59]: for this special technique, the term "resolution revolution" was coined, since it is likely to provide unique advancements to the field in the upcoming years. Unlike X-ray crystallography, cryo-EM/ET allows to obtain 3D atomistic structures of protein complexes within their native environment, since it does not require any crystallization procedure.[60] However, the difficulties to treat small proteins with cryo-EM/ET, contribute in making X-ray crystallography and NMR still essential to obtain 3D protein structures with atomistic resolution.[61,62] However, in the case of membrane proteins, it is noteworthy that the usage of X-ray crystallography and NMR in presence of protein/membrane interfaces is often limited by the intrinsic difficulty of these techniques in treating hydrophobic and flexible surfaces.[63] At the same time, X-ray crystallography and NMR share a common and fundamental limitation, that is, they are generally reliable in providing one single "snapshot" of the protein of interest, but they rarely offer useful information concerning the dynamics of the molecular events (the B factors of the X-ray structures refer to small, local fluctuations, whereas the conformational data given by NMR is often affected by artifacts). Cryo EM/ET is more promising from this point of view, as the molecule of interest can be caught in several different biological conformations, which may eventually allow to estimate the free energy difference between one state and the other, by statistical sampling.[64,65] However, the technical and economical (high cost) limits of this technique still constitute the bottleneck for its wide diffusion among structural biologists. In this sense, computer simulations still represent the only reliable and cheap possibility to get an exhaustive idea about how the molecular systems described by cryo-EM/ET, X-ray crystallography or NMR evolve in time.

Another aspect has to be taken into account, which is the fact that the previously described approaches mainly focus on the protein, providing little details on what is happening at the membrane level. This aspect cannot be neglected, as the structure and function of membrane proteins, including PMPs, greatly depend on specific interactions with the lipids
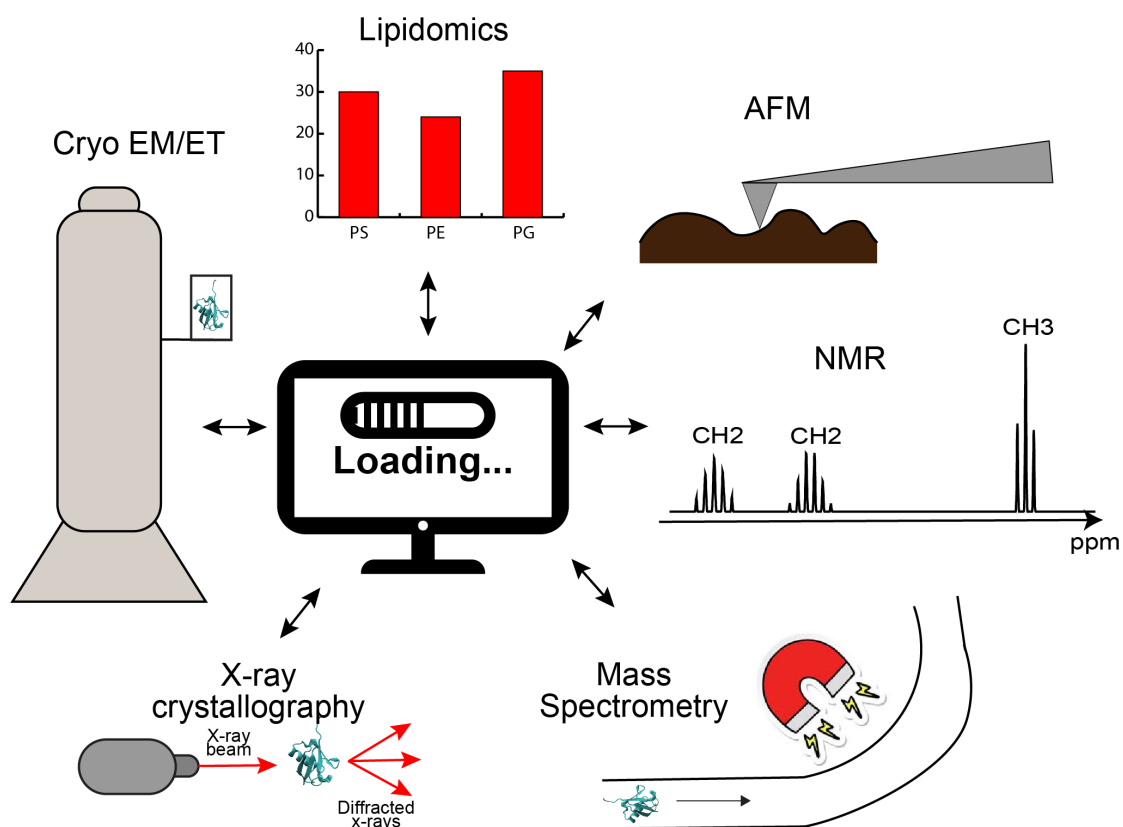
Figure 1.4 – Schematic representation of the main experimental techniques whose output can be employed in integrative modelling.

that compose the substrate where they are located. This implies that the membrane does not only constitute a static, passive, fluid, hydrophobic scaffold where the protein can perform its function: on the contrary, it represents an active environment which is concurring in providing inputs to the protein, shaping its structure and guaranteeing its correct functioning.[66] MS analyses revealed how the lipids are able to tune the structure and function of proteins[67]; at the same time, molecular simulations showed that every membrane protein has a specific and unique "lipid fingerprint",[68] which needs to be conserved in order to guarantee the correct functioning of the protein. Other techniques that can provide insights into the membrane properties are state-of-the-art lipidomics[69] and high-speed AFM.[70] However, even these methods cannot monitor in real time the evolution of a phospholipid bilayer with atomistic or nearly atomistic resolution.[71]

In summary, it is undeniable that the experimental methods developed in the last

decade, allowed to make a big step forward in the full comprehension of protein/membrane interfaces: however, it is also clear that they still carry several limitations, and even these groundbreaking techniques cannot provide a complete description of the molecular events that govern the protein/membrane interplay. Within this context, computational modelling, and in particular molecular dynamics, can provide a unique perspective, thanks to the so-called integrative modeling techniques.[72] Integrative modeling recently surged as a key tool to analyze biological interfaces in general, and protein/membrane complexes in particular. Integrative modeling aims at collecting data from different sources, both experimental and theoretical, and combine them in order to obtain detailed descriptions of the biological events that occur at a molecular level. In the present work, we used integrative modeling approaches to reveal structure-function relationships between PMPs and the lipid environment where they reside.

1.4   State of the art of computational methods

Molecular simulations are a tool that has been present since the '50s of XX century, and lays its bases on the theoretical discoveries in physics that characterized the beginning of the century (see Chapter chapter 2). However, it is only relatively recently that the increasing potency of computational tools contributed to take this instrument out of pure theoretical science and to make it effective for addressing practical biological issues.[73,74] Molecular simulations aim at reproducing *in silico* (i.e. on a computer) the behaviour of biological systems, starting from their building blocks, i.e. atoms, and from the physical laws that describe their behavior. The ergodic hypothesis[75] states that, by simulating the evolution in time of a system of particles, it is possible to infer its thermodynamic properties (such as the free energy of binding between two molecules), or to evaluate the so-called "emerging properties" (such as protein folding).

   In the last decade, significant advancements have been made in the realization of membrane models that can reproduce as close as possible the physico-chemical properties of real phospholipid bilayers. In particular, significant results have been accomplished in the realization of lipids topologies: nearly all the most diffused atomistic and coarse-grained force fields today dispose of reliable lipids parameters. Among them, it is worth to cite CHARMM (the first atomistic force field to introduce an extensive collection of lipids parameters[76]), Amber (which

developed its first lipid force field in 2012,[77] and extended the number of molecules in 2014[78]), Gromos[79] and, among the coarse-grained ones, MARTINI (which was originally proposed as a force field to simulate lipids systems only[80]). Moreover, several tools became recently available, that allow to pack lipid models into 3D membrane models. Most of these tools consists of repositories of lipids parameters from different force fields, coupled with an algorithm that is able to generate the 3D membrane model with atomistic (or coarse-grained) resolution, basically solving a geometric problem. Among them, we can cite Charmm-gui,[81] the "insane" framework of the MARTINI force field,[82] Lipidbook[83] and others,.[84,85] Our lab developed a tool, called LipidBuilder,[86] which, rather than being a repository of already known lipids, allows the users to generate parameters for lipids of virtually any nature, and successively pack them into a bilayer. This goal has been achieved thanks to the modular nature of the CHARMM force field, which allows to generate small modules of acyl tails, together with eventual insaturations. This new philosophy allows the user to personally develop the parameters for the lipids of interest in an easy and straightforward way, rather than relying on pre-built repositories. The fact that the user can autonomously select the number of insaturations and the length of the acyl tails, allows to mimic in a closer way the real biological membranes, which are generally composed of several different lipid types. As a matter of fact, the lipidome includes around 40'000 different kinds of lipids,[16] and it is virtually impossible, for online repositories, to dispose of parameters for each of them. In general, force fields developers tend to privilege lipids (such as phosphatidylcholines, phosphatidylethanolamines, sphyngomyelins) and acyl tails (mainly 18:1, 18:2 or 16:0 configuration) that are most abundant in nature.[52] This approach is highly limiting, since some lipids, despite their scarcer abundance, present a privileged role in the interplay with the membrane proteins, and drive the protein performance, as we will demonstrate in the following chapters.

1.5 New Delhi metallo-β-lactamase 1 and Golgi phosphorylated protein 3

In the present work, we focused on two PMPs which, despite being very different from a biological point of view, share similar structural characteristics (Figure 1.5). Specifically, we studied the **New Delhi metallo-β-lactamase 1** (**NDM-1**), a bacterial enzyme that provides resistance

towards last-resort antibiotics[87,88]; and the **Golgi phosphorylated protein 3** (**Golph3**), a protein present in the Golgi apparatus that is overexpressed in several malignant tumors.[89] These two proteins differ both for the organism where they reside (bacteria for NDM-1, mammals for Golph3) and for their function. However, they are both PMPs, and in both cases the interaction with a biological membrane is relevant for the protein activity, although not well characterized yet.
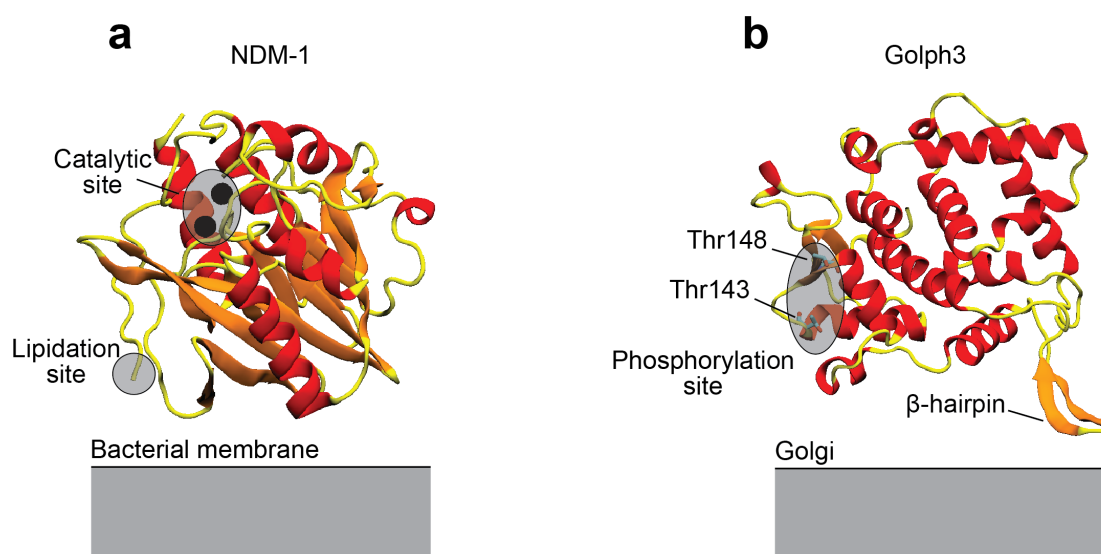
Figure 1.5 – Structures of the proteins of interest. (a) New Delhi metallo-β-lactamase 1 (the catalytic zinc ions are represented as grey spheres, while the red sphere represents the lipidated N-terminal region). (b) Golgi phosphorylated protein 3.

NDM-1 is a bacterial enzyme which confers resistance to last resort antibiotics.[87,88] In general, metallo-β-lactamases (MBLs) constitute a class of proteins that are able to disrupt antibiotic molecules, and specifically the β-lactam ring that characterizes all penicillin-like drugs.[90,91] These enzymes were selected by evolution in the billion-years long war between bacteria and fungi[92,93]: indeed, most of the anti-microbic drugs that are currently approved and on market, are extracted (or mimic) molecules that are present in fungi.[94,95] Penicillin-like molecules bind and inactivate penicillin binding proteins (PBPs).[96] PBPs are essential for the bacterial life cycle, as they build peptidoglycans, which are the building blocks of the peptidoglycan layer. The peptidoglycan layer confers mechanical stability to the bacterial wall, and in its absence, the bacterium collapses and loses its ability to reproduce.[97]

MBLs use metal ions, present within a specific active site, to initiate a catalytic reaction

that will reduce the penicillin-like molecules to their inactive forms (penicilloic acid).[98] This constitutes a formidable aid to the action of bacteria, which can prosecute the infection as their peptidoglycan layer is not any more affected by the antibiotic. However, the presence of MBLs is generally not critical for the treatment of bacterial infections, as the response of the human immune system includes a metal starvation action, often mediated by proteins known as calprotectins.[99, 100] This action is not directly aimed at targeting MBLs, but rather at removing metal ions from the infection site, as metal ions are essential for bacterial activity.[101] However, as a positive side effect, the metal starvation mechanism supports the action of antibiotics, by inactivating (or at least depotentiating) MBLs, which represent their most threatening opponent.

NDM-1 recently came to the fore because it is able to perform its enzymatic action even in presence of metal starvation conditions.[102] As a result, NDM-1 caused a serious outbreak of infections that cannot be treated by ordinary anti-microbic drugs (the first outbreaks exploded in India, from which came the name of the enzyme).[103–106] Being considered a serious treat to the public health, this enzyme gained notoriety not only within the scientific community, but also within the general public.[107, 108] It was recently discovered that the main difference between NDM-1 and other MBLs lays in its location within the bacterium: unlike other MBLs, NDM-1 is post-translationally modified (lipidated) in its N-terminal region.[109] This lipidation and consequent membrane anchoring were found to be associated to increased resistance to antibiotics, mainly thanks to the higher probability for NDM-1 to be secreted into vesicles and to spread outside of the bacterial wall.[110] In this work, we used molecular dynamics to study the dynamics of interaction between NDM-1 and the outer bacterial membrane: in particular, we simulated both NDM-1 and another bacterial enzyme of the same class, namely VIM-2, which did not develop the capability to anchor to the membrane.[110] The aim was triplex: 1) understand why NDM-1 only, among MBLs, has developed the membrane anchoring mechanism; 2) provide a description of the mechanism with which membrane anchoring occurs, from a molecular point of view (for example, which residues are involved, and whether the membrane properties are influenced by the protein anchoring); 3) evaluate possible actions to inhibit NDM-1/membrane anchoring. We also evaluated the potential of NDM-1 to be secreted into vesicles, and compared it to other MBLs.

Golph3 is a PMP present at the Golgi apparatus in mammalian cells.[111] Its physiological role has been extensively clarified, and consists in interacting with glycosylating enzymes to promote their recycling.[112] In particular, Golph3 regulates the retrograde transport of gly-coenzymes to the inner layers of the Golgi. This helps the cell to recycle these useful enzymes (which require a large amount of energy to be synthesized) and to avoid unnecessary wastes of useful resources. However, Golph3 has also been associated to several forms of cancer, in which it results to be overexpressed. Although the overexpression of Golph3 is clearly related to arising forms of cancer, it is still unclear why this happens. A few hypotheses have been made, which however are still lacking a definitive confirmation. Among these, we can cite the fact that Golph3 is known to get phosphorylated by DNA-dependent protein kinases, following DNA damage events. Its overexpression may be thus related to resistance to DNA damage, possibly causing genome instability.[113] It has also been hypothesized that Golph3 may sustain secretions of pro-tumoural factors that mediate extracellular matrixremodelling and metastatization.[114] As the interaction between Golph3 and the Golgi is not well charac-terized, we aimed at evaluating the mechanistic behind this interplay, in order to provide a detailed description of its functioning. The aim is to provide useful information for driving the research of Golph3-inhibiting molecules.

The two systems that we chose are representative examples of the impact of PMPs for global health and for the challenges of modern medicine: in particular, these two systems have a direct relationship with the main present (cancer) and future (antibiotic resistance) sanitary emergencies. The statistics show that about 20 million people die every year from cancer[115]: 13% of these represent juvenile cancers, hence people dying at significantly young ages, which represent particularly important losses, both from the human and from the economic point of view.[116] It is estimated that around 14 billions are currently being spent worldwide in cancer research, to improve the current medical treatments or to design new ones, with an increasing trend.[117] Cancer is a very complex disease, whose etiology depends on a very large number of factors, which all need to be accounted in order to derive a coherent explanation. Thus, the lack of valuable models for PMPs represents a major issue for researchers, as only one missing block can significantly affect the reliability of the models that are used to fight this lethal disease.

A similar argument applies for antibiotic resistance: bacterial infections were the main cause of death at the beginning of the XX century, accounting for about 400 deaths every 100'000 people around the world.[118] The medical revolution represented by the introduction of vaccines and antibiotics allowed a dramatic decrease in this number, lowering it to less than 50 deaths per 100'000 people. However, as the number of antibiotic-resistant bacteria increase, it is estimated that this number will consistently grow in the upcoming future, until bacterial infections will be the main cause for 10 million deaths around 2050, according to estimates of the World Health Organization.[119] This would mean that bacterial infections would come back to be one of the main causes of death, hence depriving humanity of one of the main scientific conquests achieved in the industrial era. In terms of resources, it is estimated that about 2 billions Euros are currently being spent in the European Union only, in order to fight a comeback of bacterial infections.[120]

The present work is therefore covering two of the main global sanitary emergencies, and this states the importance of developing new methods and models for the structural characterization and functional elucidation of PMPs systems. This will indeed provide a big advantage in understanding their contribution to relevant human pathologies: moreover, it will also be helpful to propose new treatments that are able to fight such pathologies with better results in terms of life extension and improved quality of life of the patients.

Overall, the aim of the present work is to provide an original and complete characterization of the two aforementioned systems, whose structural and dynamic association with the membrane is poorly characterized up to date. We did so by using a synergistic approach which combines the most advanced modelling techniques with well assessed experimental methods, which can validate the simulations predictions. The results highlighted aspects of the systems that were previously unknown. For example, we observed that the presence of a protein patch with specific electrostatic characteristics is necessary in order to induce a stable contact with the bilayer; on the membrane side, the protein attachment occurs only if clusters of specific lipid types (such as cardiolipins) accumulate in proximity of the protein itself. Removal of these lipid types is sufficient to partially or totally inhibit the protein/membrane interaction. Surprisingly, the results show common traits between the two systems, despite the differences that we previously listed (different function, different protein family, different membrane,

even different organism). This suggests that the conclusions and findings that we reached may be common to several other PMPs systems.

On a larger extent, our work highlights how the combination of experimental and computational techniques can provide a unique perspective that neither of them could offer alone. A full integration of these techniques is the key to deepen our understanding of protein-membrane interplay, which still remains elusive for many systems nowadays, despite its pharmacological and biological relevance.

2 Molecular Modelling Methods

"We are perhaps not far removed from the time
when we shall be able to submit the bulk of
chemical phenomena to calculation."
– Joseph Gay-Lussac, 1808

2.1 Historical background

2.1.1 The evolution of medicine and physics

Since the dawn of history, mankind has sought to better understand the operating mode of human organisms, with the aim to cure diseases and in general to elongate human life.[121–124] This comprehensible attitude is confirmed by numerous archaeological findings, which date back to the beginning of civilization: the Edwin Smith Papyrus, a surgical treatise on trauma which was written around 1600 BC, is probably the most antique and remarkable example of ancient medicine.[125]

However, for thousands of years (from Ancient History to Middle Age), the medical sciences were affected by a major mistake: physicians of the time strongly believed in *Vitalism*, a principle which states that living and non-living organisms are essentially different, because of a supposed "vital principle" that would pervade the living matter. This principle implies that organic and inorganic objects are made of different substances, and consequently respond to different rules.[126] Vitalism was made obsolete only in 1828, when German chemist Friedrich

Wohler was able to synthesize urea (an organic, and therefore supposedly "*vital*" compound) starting from inorganic ("*non-vital*") elements.[127] This simple experiment constituted the final proof that the scientific community had been waiting, in order to put aside Vitalism once and for all.

The fall of Vitalism had strong repercussions not only in philosophy, but especially in science. It was finally world-wide accepted, that living organisms were made of the same matter as non-living ones, and therefore subject to the same rules: the rules of physics and chemistry. This fundamental notion brought to a reformulation of the concept of organic chemistry. Until that moment, organic chemistry was indeed known as "*the chemistry of living things*": upon the decline of Vitalism, this definition was turned into "*the chemistry of compounds which contain carbon in covalent bonding*".[127,128] Following Wohler's experiment, the scientific community finally recognized that the study of living things would have ultimately coincided with the study of the building blocks (atoms) that compose them, and that are no different from those that compose inanimate objects. These notions were included in the new theory of *Mechanism*, a doctrine stating that all natural objects (and in particular the living ones) respond to the laws of physics, and can be described in these terms.[126]

Molecular modelling is a technique that aims at reproducing the behaviour of nature at a nanoscopic scale, in order to study the so-called emerging properties (such as thermodynamic properties). In the present work, we made ample use of molecular modelling, and in particular of one of its most popular branches, that is ***Bio***molecular modelling. Biomolecular modelling relies on the principles of Mechanism, as it aims to study the building blocks of life (i.e. atoms and molecules) by encompassing all the known notions from chemistry and physics.

Molecular modelling was first theorized in the third decade of the XX century, when physicists were able to obtain accurate mathematical descriptors of the events that occur at atomistic and molecular scales. It was immediately clear that, if it was possible to solve the equations that regulate the motions of all particles present in a certain system, then it would have been feasible to predict the future evolution of such system, and to calculate with extreme accuracy its thermodynamic properties, without the need to perform experimental measures, which are often unpractical and expensive, in terms of both time and resources.

The foundations to this theory were laid in 1900, when Max Planck first discovered that in nature, certain physical quantities were not continuous, but could only assume well defined and discrete values.[129] The popular Latin motto "*Natura non facit saltus*" ("*Nature does not make jumps*"), coined in the XVII century by Leibniz, who denied the existence of atoms,[130] had been finally proven wrong. This discovery represented such a huge revolution in physics, that a new term had to be introduced in order to describe the new field: *Quantum Mechanics*. From that moment on, quantum mechanics would have been juxtaposed to (and somehow complemented by) classical physics, where the latter refers to the theories of physics that do not make use of the quantisation paradigm.[131] (A few years later, another huge breakthrough, led by the work of Albert Einstein, would have added a third field to this list: the *General Relativity*, which can accurately describe the behaviour of infinitely heavy objects, such as black holes and quasars[132]).

It is important to notice that, although quantisation represents a fundamental trait of the reality that surrounds us, it is possible, and sometimes even desirable, to neglect it: this allows to avoid a useless increase in the complexity of the mathematical descriptors of the physical event of interest. For example, the laws of classical physics (and in particular classical mechanics), can describe with great accuracy the trajectory undertaken by objects such as bullets or billiard balls, whose motions are only influenced by Earth's gravity and by other objects, whose weight is of their same order of magnitude.

However, at the dawn of XX Century, classical physics had been proven several times to provide a poor description of submolecular events. Physicists like Heinrich Hertz (who discovered the photoelectric effect in 1887[133]) and Robert Kirchoff (who exposed the black body radiation issue in the second half of XIX century[134]) paved the ground to Planck's studies. The new-born field of quantum mechanics grew exponential interest within the scientific community, hence attracting the greatest minds of the time. Scientists whose names belong today to popular culture, such as Niels Bohr,[135] Louis De Broglie,[136] Werner Karl Heisenberg,[137] Erwin Schrödinger[138] and others, contributed to the refinement of the quantum mechanics theories, and ended up in providing an accurate mathematical description of the behavior of atomistic systems (Figure 2.1).
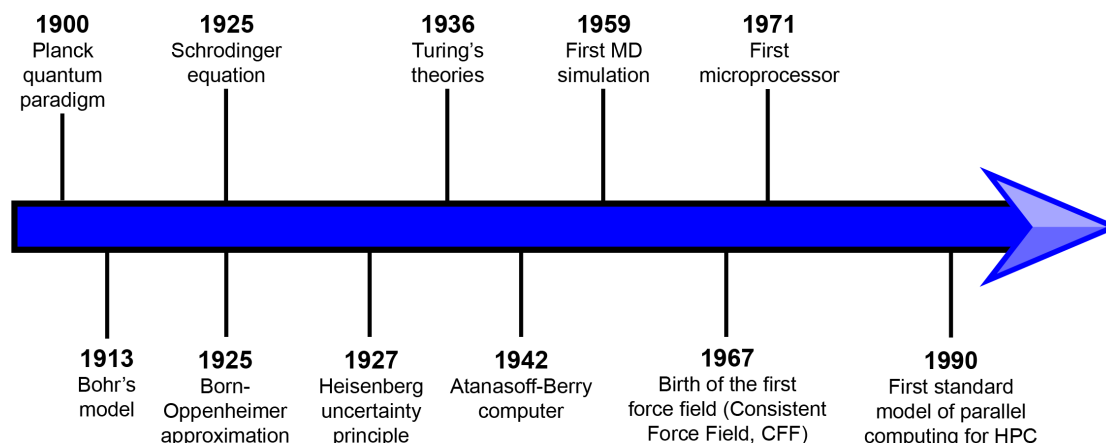
Figure 2.1 – The main historical events that laid the foundations of molecular dynamics methods.

In particular, the final stage of this long trip was represented by the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) = -\frac{\hbar^2}{2m}\nabla^2\Psi(\mathbf{r}, t) + V(\mathbf{r}, t)\Psi(\mathbf{r}, t) \qquad \text{Schrödinger Equation}[138] \qquad (2.1)$$

The Schrödinger equation represents the ultimate dream of all philosophers and physicists of history, from Democrito and Aristotle to Galileo and Newton: indeed, this equation represents the most accurate way to mathematically describe the evolution in time of any atomistic system made of an arbitrary number of particles, provided that their initial positions and velocities are known (for the moment we will not consider the necessary uncertainty defined by the Heisenberg uncertainty principle[137]). The scientific community had ultimately found a way to mathematically describe the behavior of the building blocks of reality, and to predict their future evolution. A huge philosophical problem had long been associated with the potentiality of this knowledge. Scientists were now asserting that it was possible to predict the evolution of any system of particles (knowing their initial positions and velocities): but as we saw, almost 100 years earlier, Wohler had proven that any living organism is made of atoms, just like the rest of inanimate matter in the universe (principle of *Mechanism*). This would necessarily imply that it was theoretically possible to predict the future actions of any biological system. But this, in turn, would mean that the free will, that humans had believed

to possess, was just apparent. In particular, living organisms would not "decide" to undertake specific actions: on the contrary, any decision taken by sentient beings (animals in general, and humans in particular) would inevitably be the consequence of environmental pressure. If we consider a synapse, that is one of the main actors in the arising of decisional events within the brain, we can see how the transmission of the electrical signal would be only driven by the surrounding atoms, and therefore, ultimately, by the environment. The behavior of conscient beings would therefore not be "free", but rather the logical consequence of events that happen around them.

However, this somehow dystopic (and disturbing) vision of the world is critically hindered by the fact that the Schrödinger equation is not just a simple equation of motion (like, for example, Newton's equations): as a matter of fact, the Schrödinger equation states that any particle can be at the same time in different places, with a certain probability. This statement goes along with the eminent principle that Heisenberg had expressed a few years prior the enunciation of the Schrödinger equation: the Heisenberg principle states that, due to the perturbation that is introduced by an observer, it is impossible to know with accuracy both the position and the velocity of a single particle.[137] In particular, the margins of error with regards to position and velocity determinations are related through the following equation:

$$\Delta x \Delta p \quad \geq \quad \frac{\hbar}{2} \qquad\qquad \text{Heisenberg Uncertainty Principle}^{[137]} \qquad\qquad (2.2)$$

In conclusion, the very same nature of Schrödinger equation, together with Heisenberg uncertainty principle, assesses that it is impossible to know the future state of a system of quantum particles, and therefore predict its macroscopical consequences, including the events that drive the neuronal functioning: the principle of free will is safe.

Another relevant reason for which it is not possible to predict with total accuracy the future evolution of a system of particles is that, despite its theoretical feasibility, it is unrealistic to solve the Schrödinger equation. The Schrödinger equation is indeed characterized by a very high complexity, that makes it impossible to be solved analytically. Provided that some

approximations are made, it would still be feasible to use automatic calculating machines which, however, were little more than a dream at the time of the enunciation of Schrödinger equation.

2.1.2 The evolution of computing

Alongside with medicine, another issue had populated the dreams of scientists from the beginning of history: building machines that could provide a framework for automatic calculation. In this sense, some very sophisticated machines were built even in Ancient History. The most astonishing example is probably represented by the so-called *Antikythera Mechanism*, a mechanical tool that was discovered in the early XX century on a Greek ship sunk a few miles off the coast of the Antikythera's island (Figure 2.2a). The Antikythera mechanism is basically a planetary, which makes ample use of cogwheels, in order to predict events like eclipses, lunar phases, equinoxes and even the correct dates for the Olympic Games. According to some estimates, the construction of this mechanism dates back to 150 BC.[139]

Despite not reaching the perfection of ancient Greeks, mechanical calculators were extensively built and used also during the Middle Age and up to the Modern Era (Figure 2.2b). In addition to their obvious technological limits, at the beginning of the XX century the debate was still on, about whether calculators could theoretically solve problems of high complexity, such as the Schrödinger equation.

During the same years that saw the rise of quantum mechanics, the theories of Alan Turing paved the way to the creation of modern computers. In particular, in 1936 Turing demonstrated that any possible calculation could be performed by a mechanical device, provided that enough time and storage space were granted.[140] This would imply the possibility to transfer the burden of calculations from humans to machines, hence allowing, among other things, the resolution of the complex equations of motion that characterize molecular systems.

Just a short time after Turing's theories were published, the first electronic computer was built in the United States, by Atanasoff and Berry.[141] The revolution, with respect to the previous calculators, consisted in the usage of electronic elements to perform the calculations, rather than mechanical switches (Figure 2.2c).

Despite these great theoretical and technological advances, the available computational capacity was still very far away from allowing the resolution of the equations of motions of molecular systems (at least for significant space- and time-scales). However, as years passed, it became clear that the novel technology had the potentiality to support the creation of more and more powerful computers. Eventually, with the invention of microprocessors by Federico Faggin,[142] and the consequent massive growth in the number of transistors that could be integrated on one single chip (growth that was exponential for several years, as described by Moore's law[143]), some digital computers became available, whose computing power would have been unimaginable only a few years earlier (Figure 2.2d).
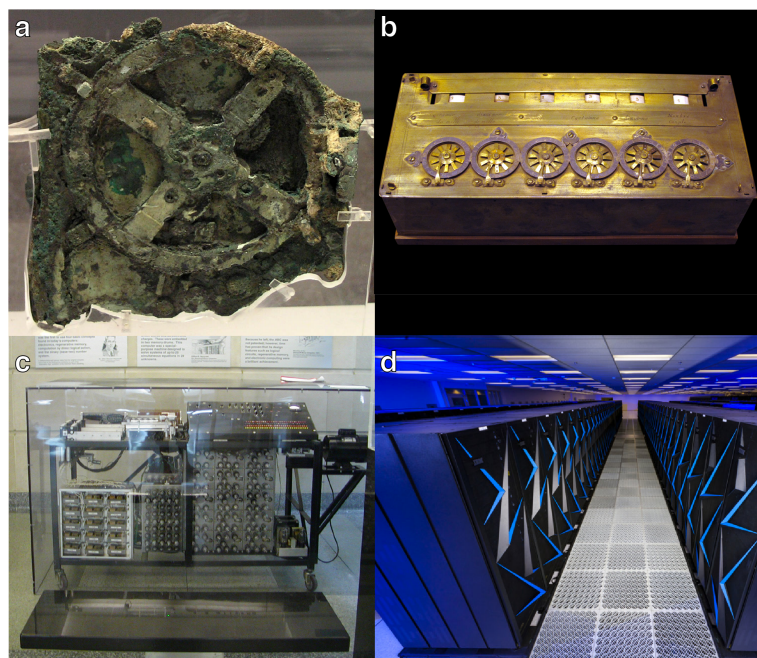
Figure 2.2 – Evolution of computing machines throughout history. (a) The Antikythera Mechanism. (b) A Pascal's calculator (mechanical calculator used in the XVII and XVIII centuries). (c) A reproduction of the Atanasoff-Berry machine (first digital computer). (d) A modern supercalculator.

Appropriate approximations, such as the Born-Oppenheimer approximation,[144] were introduced in order to decrease the complexity of Schrödinger equation for systems of atoms, and make it suitable to be solved by electronic calculators. This paved the way to the generation of new algorithms which, despite some successive refinements, are still used today (Figure 2.3), and that have been intensively utilized within the present thesis. In particular, in the next sections we will discuss in details the *Monte Carlo* and *Molecular Dynamics* methods.
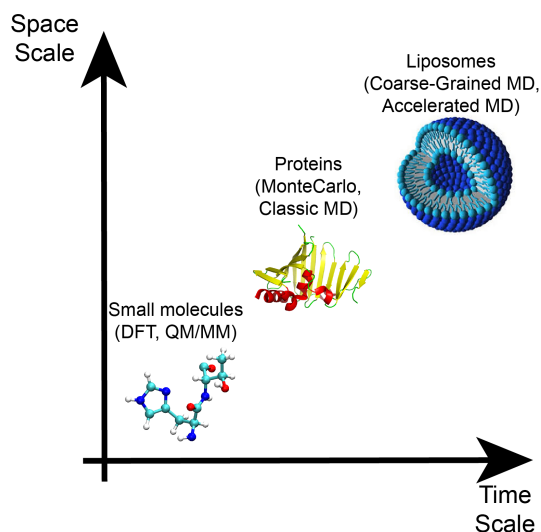
Figure 2.3 – Some examples of the appropriate molecular modeling techniques to be used at specific time- and space-scales.

2.2 Monte Carlo methods

Monte Carlo methods were introduced by Enrico Fermi (who made ample use of them during the Manhatthan project, although he never published[145]) and by Nicholas Constantine Metropolis (who first provided the theoretical framework[146]). Monte Carlo methods assess that, if the probability distribution of an event is known *a priori*, then the macroscopic (thermodynamic) properties of the system can be estimated by sampling the state of the system multiple times, and measuring the quantity of interest that is associated to the sampled state.

A classic example that shows the potentiality of Monte Carlo methods is represented by the estimation of π (Figure 2.4). If we consider a unit square and we inscribe a circle in it, it is well known from classical geometry that the areas of the two figures are related according to the equation:

$$Area_{circle} = Area_{square} * \frac{\pi}{4} \tag{2.3}$$

Monte Carlo methods offer a way to estimate the value of π. By selecting random

points within the square, with X and Y coordinates picked randomly according to a uniform distribution, we have that each point can belong (or not) to the circle, according to the equation that defines the area of the circle itself, that is:

$$x^2 + y^2 < 1 \tag{2.4}$$

If the points are randomly taken according to a normal, uniform distribution, then the probability to pick a point that belongs to the area of the circle is equal to $\frac{\pi}{4}$. This implies that, as the number of points that are picked increase, the ratio between the number of points that belong to the circle and the total number of points picked, should tend to this number, according to the law of large numbers and to the equation:

$$\lim_{\#\to\infty} \frac{\#points_{circle}}{\#points} = \frac{\pi}{4} \tag{2.5}$$

π ~ 2.66   π ~ 3   π ~ 3.2
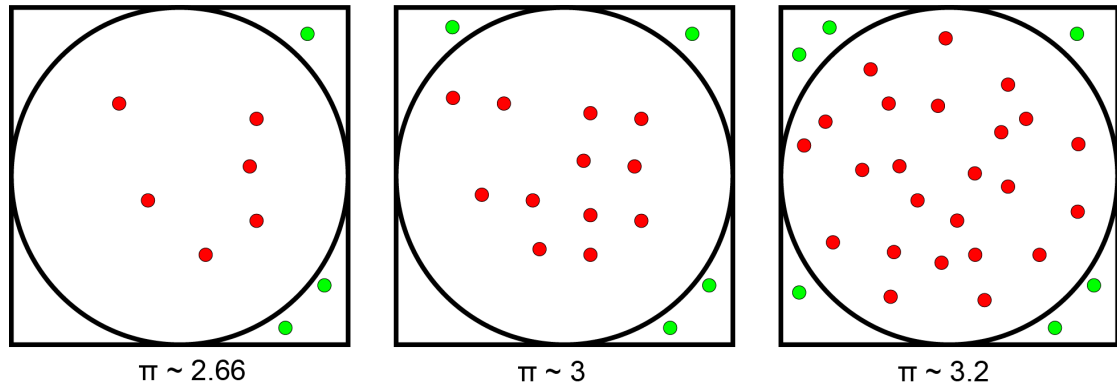
Figure 2.4 – Estimation of π through Monte Carlo methods.

Monte Carlo methods are widely used in a number of different fields, such as weather forecast,[147] computer graphics,[148] telecommunications,[149] and biology[150] as well. Together with molecular dynamics, Monte Carlo methods represent one of the main tools for simulation of biomolecular systems. However, compared to molecular dynamics simulations, Monte Carlo methods are less used in biology, because they do not reveal any information regarding

the temporal evolution of molecular systems. Indeed, all Monte Carlo moves are randomly chosen, and therefore do not carry a physical meaning. As a consequence, dynamical properties, such as a diffusion coefficient, cannot be inferred from a Monte Carlo simulation.[151]

2.3 Molecular dynamics simulations

2.3.1 All-atom approach

Molecular Dynamics (MD) is a powerful, theoretical tool that allows to analyze and predict the future evolution of a system of particles.

One of the main assumptions that lays at the basis of MD is the following: when we perform an experimental observation (measurement) of a macroscopic property of a system (such as the temperature), we are not measuring the real state of the system in a specific instant (microstate), but rather an average of the states of all the particles that compose the system itself. For example, since the temperature of a system of particles (such as a gas) is correlated to the kinetic energy of all its particles, when we measure the temperature for that system, we are actually measuring an average of all the kinetic energies of the particles that compose it (*ensemble average*). Since the systems of interests are generally made of a number of particles that is on the order of the Avogadro number ($6.023*10^{23}$), it would be totally unfeasible to simulate them. With MD, it is possible to infer the same properties of a system made of a large number of particles, through the simulation of a much smaller (if compared to the Avogadro number) group of particles, for a defined period T of time.

In particular, there are two ways of calculating an average:

- Average across a trajectory in the phase space (*ensemble average*)

$$A_{obs} = \frac{1}{N} \sum_{n=1}^{\infty} A(\Gamma_i) \tag{2.6}$$

- Average across a trajectory in time (*time average*)

$$A_{obs} = \frac{1}{T} \int^{T} A(\Gamma(t)) \, dt \equiv \overline{A} \tag{2.7}$$

In the first case, making a measure would be equivalent to calculating the property of interest (such as the kinetic energy) for all the particles composing the system, at a given moment: the kinetic energy of the system is the average of the kinetic energies of all its components.

In the second case, the measure would be equivalent to following a single particle, and evaluate all the states that the particle assumes: the measured property would correspond to an average *in time* of the property of the particle, measured across a defined interval of time T.

In particular, MD simulations rely on the *Ergodic Hypothesis*: this rule implies that all the microstates that belong to the phase space are equally probable. As a consequence, when the number of particles that compose the system is very high (N -> ∞) and the observation time T too (T -> ∞), the two equations reported above become equivalent, i.e. an average in time is equivalent to an average in the phase space (or ensemble average).

This approach has the considerable advantage that, taken any system of particles, its thermodynamical properties can be inferred from a simulation that can predict the evolution *in time* of its components. In other words, a simulation in time can potentially sample every possible microstate of the phase space. The ergodic hypothesis is essential because MD aims to define the properties of systems whose number of particles is generally on the order of the Avogadro number ($6.023 *10^{23}$). It is virtually impossible to solve the equations of motion for systems of that size: however, the ergodic hypothesis states that, for ergodic systems, following the time-evolution of one single particle (time average) is equivalent to sampling a group of particles at a specific moment (ensemble average).[75, 152] As a consequence, for long simulations (time T -> ∞), a MD simulation of a relatively small-sized system would end up in providing information that is significant at macroscopic levels. Of course, a real simulation can only be run for a finite amount of time: however, when the simulated time T is fairly larger (10-fold) than the typical-scale of the event of interest, MD provides results that are in line with the experimental data.

As stated in the previous sections, the proper way to evaluate the evolution in time of a system of particles that respond to quantum mechanical laws, would be to solve the Schrödinger equation for every single particle (considering nuclei and electrons as separate

entities). However, it is impossible to solve the Schrödinger equation for systems with a high number of particles. Quantum Mechanics / Molecular Mechanics (QM/MM) methods introduce some approximations to simplify the required calculations, however the region that is currently treatable with the quantum mechanical theory is not larger than a few hundreds of atoms.[153]

In order to treat larger systems, such as a protein, a biological membrane or a DNA molecule (which are systems on the order of 10^4 atoms), we can adopt an approximation, which has been known since 1927, when Max Born and Robert Oppenheimer first theorized it, and that is indeed known as the Born-Oppenheimer approximation.[144]

The Born-Oppenheimer approximation states that it is possible to decouple the motion of the nuclei from the one of the electrons. In particular, this is doable thanks to two fundamental assumptions: nuclei and their electrons are subject to the same forces (as they concur in composing one specific atom), and the vibrational motions of electrons are much faster than the ones that characterize the nuclei. In particular, an electron vibrates with a frequency that is on the order of 10^6 ms^{-1}, while nuclei vibrate at about 10^2 ms^{-1}. This is a direct consequence of the difference in the mass between the two entities (the mass of an electron is about 2000-fold lower than the mass of an atomic nucleus), and therefore to their different inertia and capability to respond to the acting force. This evident disproportion makes it reasonable to suppose that, whatever change in motion the nucleus undergoes, all the electrons of the atom will follow it, almost instantly, because of their lower mass, and hence higher velocity.

In a mathematical fashion, the Born-Oppenheimer approximation can be written as:

$$\Psi_{total} = \psi_{electronic}\psi_{nuclear} \tag{2.8}$$

where Ψ_{total}, $\psi_{electronic}$ and $\psi_{nuclear}$ are the wave functions of the whole atom, of its electrons, and of its nucleus, respectively.

This allows to represent a system of atoms as a system of interacting spheres. The motions of particles of this size (> 1 Å) and weight (> 10 Da) are quite accurately described by Newton's equations of motion. The solution of these equations represents the core of MD.

The first MD-like simulation of history was conducted for a system of hard spheres by the group of Berni Alder,[154] which first realized an algorithm that could be run on an electronic calculator for a significant amount of simulation time. (Interesting fact: the algorithm was actually written by Mary Ann Mansigh Karlsen, a woman that was working as a programmer in Alder's group. Back in those times, it was very common that all programmers positions would be covered by women. Nevertheless, they were receiving very little credit for that, to the point that programmers' names were generally not included as authors in the scientific articles. The same fate was indeed attended by Mrs. Mansigh, who does not appear among the authors of the first MD simulation paper, despite being the one who materially wrote the code. A few years ago, I had the luck to attend a talk given by Mrs. Mansigh. She admitted that the environment in which she grew up as a scientist was not very fair to women, and she declared that she is very happy for all the positive changes that have occurred in the scientific community since then. However, she also acknowledged Berni Alder for always providing her with all the credit that, at the time, was considered "fair enough" for programmers. As a matter of fact, Alder even included her name in the authors list of a paper that his group published some years later, in the 1970s. This positively surprised Mrs. Mansigh, as it was an extremely rare event for programmers to be included in scientific publications at that epoch).

At the time of the first MD simulation performed by Alder's group, only a simple system, such as a liquid of argon, could be simulated. In that case, the system of interest was a simple group of point particles, that could interact with each other via a specific potential, given by the equation:

$$V(r) = \begin{cases} \infty & r < \sigma_1 \\ -\epsilon & \sigma_1 < r < \sigma_2 \\ 0 & r > \sigma_2 \end{cases} \qquad (2.9)$$

this potential was a simplified version of the Lennard-Jones equation.

Despite being very rudimentary, this first MD simulation marked the beginning of a new era: it showed that it was actually possible to perform on a computer something that,

until then, had only been theorized.

From that moment on, many things started to change: as stated by Moore's law,[143] more and more powerful computers became available. Moreover, huge improvements occurred in the definition of force fields. In particular, new experimental data became available, that could be used for setting up empirical force fields. The force field, defined as the definition of all the forces that are acting on a specific particle of the system, can indeed be of two different kinds:

- *ab initio*

- *semiempirical*

Ab initio force fields present functional forms that are directly derived from quantum mechanics-based first principles, and therefore they do not require fitted parameters.[155, 156] These force fields can be very accurate, but they require heavy calculations in order to cover all the possible states of different atoms. As an example, a carbon in an aliphatic chain or a carbon in an aromatic ring will behave in very different ways. It is technically very challenging to achieve these data for all the atoms that compose biomolecules of interest (such as all aminoacids, a reasonable number of phospholipids, ions, etc.). Because of these reasons, currently no *ab initio* force fields are available, which can be utilized for addressing issues involving biomolecules.[157]

Because of this, the most popular sets of force fields (Amber,[158] CHARMM,[159] Gromos,[160] etc.) are *semi empirical* force fields: this means that, for certain parameters, they rely on experimental, empirical observations, coming from experiments such as X-ray crystallography and gas phase electron diffraction.[161]

The definition of force fields underwent significant improvements in the decades after the first MD simulation of Alder took place. The pioneering work of Shneior Lifson paved the way to the definition of modern force fields.[162] In particular, Lifson was the first to realize that hydrogen bonds could be described as simple electrostatic interactions, and to develop a procedure to derive energy parameters. Subsequently, two of his students, Michael Levitt[163] and Arieh Warshel,[164] together with Martin Karplus,[165] incorporated a huge amount of data to develop the first force fields with the ability to treat molecules as big as proteins. For their

studies, Levitt, Warshel and Karplus were awarded the Nobel prize for Chemistry in 2013.

By relying on the principles of molecular mechanics that were previously exposed, and in particular on the Born-Oppenheimer approximation, we can model the atoms as particles that respond to Newton's laws of classical mechanics. In details, if we take a time-step that is small enough (order of ~1-2 fs), we can say that the motion of each particle (atom) is uniformly accelerated during the time that goes from $step_j$ to $step_{j+1}$, and therefore its position at $step_{j+1}$ can be calculated as $r_{j+1} = r_j + v(\Delta t) + \frac{a\Delta t^2}{2}$ where Δt is the time step. The acceleration can be calculated as:

$$\mathbf{a}_i(t) = \frac{d^2 \mathbf{r}_i(t)}{dt^2} = \frac{\mathbf{F}_i(t)}{m_i} \qquad \text{Newton's second law} \qquad (2.10)$$

where the force on each particle i is equivalent to the vectorial sum of all forces that are acting on particle i at step$_j$. These forces are indeed defined by the force field, and in particular can be inferred as:

$$\mathbf{F}_i(t) = -\frac{dU(\mathbf{r})}{d\mathbf{r}_i(t)} \qquad (2.11)$$

A general equation for atomistic force fields was proposed by Lifson within the Consistent Force Field (CFF) project,[162] which was the first research project in history aimed at force fields development. This general equation has not undergone significant modifications since then, and it has the following form:

$$U(r) = \sum_{bonds} K_r(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\phi[1 + cos(n\phi + \delta)] + \sum_{improper} K_\varphi(\varphi - \varphi_0) +$$

$$+ \sum_{i<j} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] + \sum_{i<j} \frac{q_1 \cdot q_2}{4 \cdot \pi \cdot \varepsilon \cdot \varepsilon_0 \cdot r^2} \qquad (2.12)$$

Where the first 4 terms represent the so-called "bonded" interactions and provide a mathematical description for the covalent bonds of the system (Figure 2.5). All the K terms are parameters that work as elastic constants: this means that the higher these parameters are, the

more stiff and rigid the interaction will be (i.e. more energy is required to displace the atom from the equilibrium position). The last 2 terms are "non-bonded" interactions, and represent the non-covalent interactions, in particular van der Waals and electrostatic interactions.

We provide here a detailed description for each term:

$K_r(r - r_0)^2$ harmonic potential acting between two covalently bonded atoms. When the relative position between the two particles gets out of the equilibrium (r_0), a force intervenes, which tends to bring them back to the equilibrium position (Figure 2.5a).

$K_\theta(\theta - \theta_0)^2$ harmonic potential describing the oscillation that takes place among 3 consecutive atoms that are covalently bonded (Figure 2.5b).

$K_\phi[1 + cos(n\phi + \delta)]$ term representing the *dihedral* angle. The dihedral angle is the angle between two planes defined by two sets of 3-consecutively bonded atoms, with two atoms in common. When a system of 4-consecutively bonded atoms is present, an oscillation around the dihedral angle takes place, as described by this equation (Figure 2.5c). Unlike the previous harmonic terms, the term for the dihedral angle is periodic, which implies that the potential energy does not increase indefinitely, but can go back to 0 at certain displacement values (in other words, the dihedral angle has multiple equilibrium positions). To describe this periodicity, a cosine function is used, which requires two extra parameters, i.e. n and δ, which represent the multiplicity and the phase shift, respectively.

$K_\varphi(\varphi - \varphi_0)$ term representing the *improper dihedral*. The improper dihedral is used in presence of a set of 4 atoms, 3 of which are covalently bonded to a central one. In this case we can define two intersecting planes, one identified by i,j,k and the other one by j,k,l. The improper dihedral term describes the oscillation of the atoms around the angle between these two planes (Figure 2.5d).

$4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$ Lennard-Jones term, representing the typical curve of the van der Waals potential, i.e. characterized by a well of energy (with a minimum at a specific atom-atom distance), a tendency to 0 at infinite distance, and a tendency to an infinite amount of energy for a distance that approaches 0 (Figure 2.6). The behavior of this graph for long inter-atomic distances is inversely proportional to the sixth power of the distance itself. This
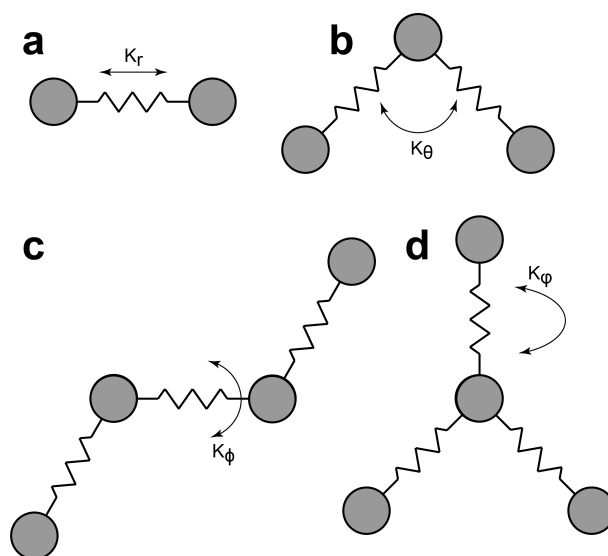
Figure 2.5 – Representation of the bonded terms in molecular mechanics force fields. (a) Bond. (b) Angle. (c) Proper dihedral. (d) Improper dihedral.

is a model of the London dispersion forces, which are part of the van der Waals forces and are predominant in this region of the potential. Being a non-bonded term, the Lennard-Jones should in principle apply to all pairs of atoms in the system, regardless of their distance. However, since this interaction is short-ranged (it is inversely proportional to the sixth power of the atomic distance), it is possible to neglect pairs of atoms that are placed very far away (and whose contribution to the Lennard-Jones potential is therefore close to zero), without significant loss of information. To do this, a cutoff distance value is generally applied, above which the Lennard-Jones term is considered null; sometimes, rather than bringing the Lennard-Jones term to zero with a step at the threshold, the potential is rescaled, in order to slowly approach the zero value. This method allows to have a smoother transition and avoid possible instabilities in the simulation. The cutoff approach allows to save significant amounts of computational power.

$\frac{q_1 \cdot q_2}{4 \cdot \pi \cdot \varepsilon \cdot \varepsilon_0 \cdot r^2}$ Coulomb term, representing the electrostatic interactions. It applies to all pairs of atoms that are electrically charged. Unlike the Lennard-Jones potential, the Coulomb interaction is a long-range one, as it is inversely proportional to the first power of the distance. Introducing a reasonable cutoff value would be unfeasible, as the required computational resources scale up to N^2, where N is the number of interactions; however, shortening too much the cutoff value would cause unacceptable approximations, that are correlated to artifacts
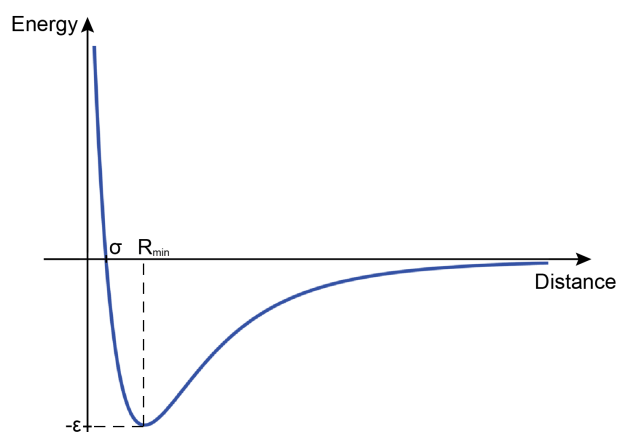
Figure 2.6 – Energy-distance relationship in Lennard-Jones potential. The Lennard-Jones potential is characterized by a well of energy (-ϵ at a specific distance R$_{(}$min)) and by a distance at which the interaction energy is 0 (σ).

and instabilities. The solution is provided by the Particle Mesh Ewald approach,[166] which contemplates a cutoff distance to separate the short-range Coulomb contributions from the long-range ones: while the former are calculated directly in the real space, the latter are calculated using a Fourier transform. This method allows a high accuracy in the calculation of the electrostatic contributions, while at the same time maintaining a reasonable computing speed: in particular, the computational resources that are required for systems of increasing size, scale like N*ln(N), rather than N^2 (where N is the number of particles). It is important to notice that maintaining a good level of accuracy in systems characterized by periodic boundary conditions, requires the system to be, overall, electrically neutral. Indeed, in MD simulations, it is a good rule to assure that the amount of ions present in the solvent guarantees an overall neutral box.

Integration of the laws of motion

All the parameters that are required for the force field definition (such as K_r, K_Θ etc.), come from empirical or semiempirical data. Once these parameters have been defined, all the above equations become functions that are exclusively dependent on time and space. It is therefore possible to solve the force field equations and calculate the resulting force that is acting, at a certain moment, on each atom that composes the system. Once the force is known, it is possible to compute the acceleration that every particle undergoes, through the Newton's

equations of motion: it is useful to remind that it is possible to apply the Newton's equations of motion on atomistic systems, thanks to the Born-Oppenheimer approximation, which decouples the nuclei from the electrons, hence neglecting the quantum effects.

Several algorithms are available to integrate the Newton's equations of motions that represent the core of MD simulations. The two most popular methods are the Verlet algorithm[167] (which is a special case of the Verlet integration[168]) and the leapfrog algorithm.[169] These two algorithms share similar aspects, as they both consist in calculating the position and velocity of particles at different moments. In particular, the equations that characterize these methods are the following (a bold character indicates a vectorial quantity):

Verlet algorithm:

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_n \Delta t + \frac{1}{2} \frac{\mathbf{F}_n}{m} \Delta t^2 \tag{2.13}$$

$$\mathbf{r}_{n-1} = \mathbf{r}_n - \mathbf{v}_n \Delta t + \frac{1}{2} \frac{\mathbf{F}_n}{m} \Delta t^2 \tag{2.14}$$

from which we have the final formula: $\mathbf{r}_{n+1} = 2\mathbf{r}_n - \mathbf{r}_{n-1} + \frac{\mathbf{F}_n}{m} \Delta t^2$

Leapfrog algorithm:

$$\mathbf{v}(t + \tfrac{\Delta t}{2}) = \mathbf{v}(t - \tfrac{\Delta t}{2}) + \frac{\mathbf{F}_n}{m} \Delta t \tag{2.15}$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t + \tfrac{\Delta t}{2}) \Delta t \tag{2.16}$$

Despite their similarity, the Verlet algorithm is more widely used within the scientific community, because of its higher computational performance.[170]

Periodic Boundary Conditions

The systems that can be studied through MD simulations are necessarily characterized by a finite size: thus, the treatment of the boundary conditions becomes of crucial importance. In particular, if no boundary conditions were applied, MD simulations would become meaningless, as all particles would indefinitely diffuse throughout the space, until no interaction would be present anymore.

To avoid this, MD simulations implement periodic boundary conditions. This means

that the system is virtually repeated on an infinite scale, and that the particles that are close to the boundaries interact with the particles that are on the opposite side of the system. On one side, this method has the considerable advantage to prevent indefinite diffusion of particles; on the other side, it constitutes a limit, because the entities that interact with themselves produce an artifact. Because of this, it is important to make sure that, along the MD trajectory, no molecules get at a distance from their copy that is shorter than the cutoff threshold of the Coulomb and Lennard-Jones interactions. If that happens, an artifact is present, and it is appropriate to repeat the simulation increasing the size of the box.

2.3.2   Coarse-grained approach

Coarse-grained (CG) methods allow to explore wider time- and space- scales. The idea of CG force fields consists in lowering the total number of degrees of freedom of the system, in order to be able to compute, with the same computational resources, longer time scales and larger systems, if compared to atomistic force fields. Generally, the decrease in the total number of degrees of freedom is achieved by grouping 2 or more atoms into one single entity. For example, the simplest approach would consist in using one point particle to represent two atoms, using the center of mass of the two atoms as the spatial coordinate for this new entity. This would imply that the information about the relative motion of the two atoms, would be lost: however, the decrease of the total number of particles in the system (from N to N/2), would also allow to employ fewer computational resources to simulate the same system. In addition, the usage of larger "pseudo" atoms implies that the particles vibrations get significantly slower. As a consequence, it is possible to increase the time-step of the simulation, without inducing instabilities: this further decreases the total amount of computational power that is required to perform the simulations.

Several CG force fields have been developed in the last years[171–175]: the MARTINI force field[80] has been proven among the most accurate ones to simulate systems that include biological molecules such as proteins, DNA and lipids.

MARTINI force field

In MARTINI,[80] atoms are grouped with the following rule (Figure 2.7): every MARTINI bead represents 3-4 heavy atoms (by "heavy atoms", we mean all atoms but hydrogens, which are neglected); moreover, each MARTINI bead is characterized by a mass (36 or 72 Da), a charge (0 or ±1 elementary charge) and a van der Waals radius (0.23 or 0.26 nm). Additionally, each particle is characterized by three different parameters, respectively describing: a general degree of polarization (polar, apolar, non polar, charged); a finer degree of polarization (5 different levels, from low polarity to high polarity); and the hydrogen-bonding capability (donor, acceptor, donor-acceptor, none).

For the rest, the structure of the MARTINI force field is similar to the atomistic ones. Like for atomistic force fields, we can indeed define bonds, angles and dihedrals (bonded interactions); Lennard-Jones and Coulomb interactions (non-bonded interactions). In this case however, the MARTINI bonds do not represent real covalent bonds between atoms in a classic sense, but rather an ideal link present between two groups of atoms (each of which is represented by a MARTINI bead), that cannot be separated if not at a big energetic cost. Similar considerations can be applied to Lennard-Jones and Coulomb interactions. Despite these differences, the equations that describe the force of these interactions share the same general form as the atomistic force fields. The van der Waals radiuses and the particle type are used to determine the Lennard-Jones parameters.

Despite its simplicity, the MARTINI force field can accurately describe several biological systems.[176] Among its limits, we have to cite the fact that it is not possible to capture changes in secondary and tertiary structure of proteins (being many atomic bonds fixed and not subject to fluctuations). To prevent artifacts, elastic constraints are often employed, that maintain fixed the secondary and tertiary structures of proteins.[177]

Different solvent models exist in MARTINI: the original MARTINI water was a one bead model representing 4 water molecules.[80] This model however lacked any polarization effect, and therefore a novel model was introduced, to reproduce the polarizability of water[178]: this new water model consists of 3 beads, and is therefore more computationally expensive, with respect to the original MARTINI water. However, the gain in accuracy is significantly higher.
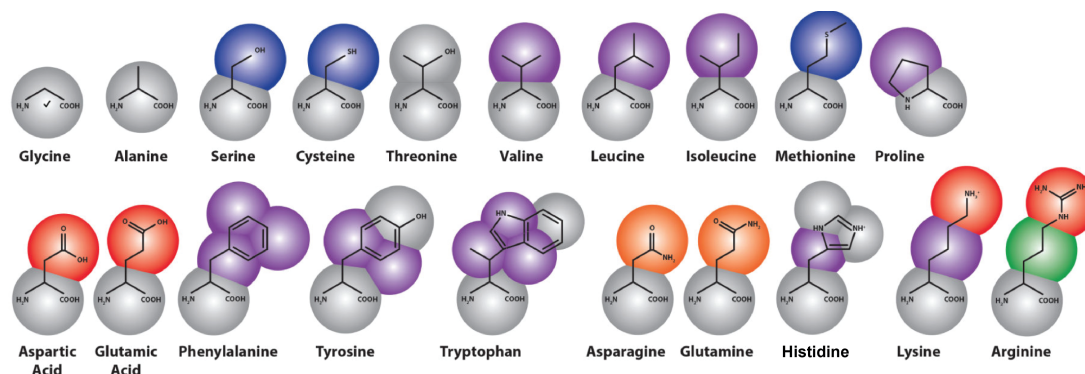
Figure 2.7 – The MARTINI coarse-grained force field approach (Figure by Bradley et al.[179]).

To estimate the computational savings of MARTINI, we need to take into account different parameters. First of all, the reduction in the total number of particles, which corresponds to about $N_{CG}=(N_{AA}-50\%)/3.5$, where N_{CG} is the total number of MARTINI beads and N_{AA} is the total number of atoms in the system. This reduction is due to the fact that MARTINI neglects all the hydrogen atoms (which in proteins constitute about 50% of all atoms[180]), and to the fact that each MARTINI bead represents 3-4 of the remaining particles; for what concerns the solvent (which constitutes the biggest portion of particles in MD simulations), considering to use the polarizable water model, we have a reduction in the total number of particles that is equivalent to $N_{CG}=N_{AA}/4$. Overall, we can approximate the reduction in the total number of particles to about $N_{CG}=N_{AA}/4$. This reduction is particularly significant, because the time that is required by a processor to compute all interactions, does not scale linearly according to the number of particles N, but according to N*ln(N). This means that the reduction in the number of particles, brings a MARTINI simulation to be about 6-fold faster than an atomistic one. Moreover, as previously stated, the CG simulations can achieve a substantial increase in the time-step, due to the slower vibrations that its particles undergo, with respect to atomistic force fields: for MARTINI, we have an increment from ~ 2 fs (AA force fields) to ~ 15-20 fs (MARTINI force field), depending on the system.

Overall, the total speed-up when going from an atomistic to a MARTINI representation is on the range of nearly 2 orders of magnitude.

40

2.4 Molecular dynamics analysis

Molecular dynamics can be used to infer a number of relevant information from systems of biological interest. From a pharmacological perspective, the property that is mainly investigated is the free energy of interaction between a ligand and its target. This information gives important indications about the pharmacodynamic properties of potential drug molecules, and can therefore be evaluated to screen and design drugs with high potency. The free energy can be estimated from MD simulations: however, the sampling is often critical, and therefore multiple simulations are generally required, at a high computational cost. Given its importance, the scientific community has invested considerable energies in development of mathematical models which can be used to solve this problem in the most accurate and cheap way. Among several computational techniques that can be used to estimate the free energy of interaction, we can cite thermodynamic integration,[181] the weighted histogram analysis method[182] and free energy perturbation methods.[183]

Another study that can be done with MD simulations is diffusion of small molecules: in biological systems, a typical example is provided by lipid diffusion through the membranes. A number of mathematical models are available which can relate the trajectory of a molecule throughout a MD simulation with its diffusion coefficient: one example is provided by the Green-Kubo equation.[184]

Another field of study for which MD simulations can be helpful is aggregation, in particular of proteins. The evaluation of protein aggregation is an issue that typically involves neurodegenerative diseases, as they are often caused by aggregation phenomena.[185] However, all-atom approaches cannot typically reach the time scales that are required to investigate aggregation events: because of this, MD has been mainly utilized for the study of peptides aggregation, as the limited size of these molecules allows to reach interesting time scales.[186–189]

The usage of CG methods would therefore be desirable, in order to reach more significant time-scales: however, CG force fields suffer a major limit, that is the tendency to overestimate the electrostatic features of proteins. This can be highly limiting for aggregation studies, as electrostatics is one of the main driving forces of this phenomenon. A possible solution would consist in first running CG MD simulations to identify surfaces of the proteins

that can likely drive the interaction; and successively, to refine the study by running AA MD simulations that are driven by the information that have been collected in the previous step.

3 Membrane association of New Delhi metallo-β-lactamase 1

The first part of this chapter (membrane affinity of NDM-1, comparison with VIM-2 and N-VIM, role played by cardiolipins) is adapted from the following paper submitted for publication and available in bioRxiv (URL: https://doi.org/10.1101/2020.06.01.126664):

"Molecular Bases of the Membrane Association Mechanism Potentiating Antibiotic Resistance by NDM-1." **Prunotto A**, Bahr G, Gonzàlez LJ, Vila AJ, Dal Peraro M

Author contributions: A.P., G.B., L.J.G., A.J.V. and M.D.P. conceived the project and designed the experiments. A.P. performed the computational experiments. G.B. performed the wet lab experiments. A.P., G.B., L.J.G., A.J.V. and M.D.P. wrote the manuscript. A.J.V. supervised the wet lab experiments. M.D.P. supervised the computational experiments. A.J.V. and M.D.P. provided the financial funding.

The second part of the chapter (secretion of MBLs into outer membrane vesicles), is adapted from a paper in preparation.

Resistance to last-resort carbapenem antibiotics is an increasing threat to human health, as it critically limits therapeutic options. Metallo-β-lactamases (MBLs) are the largest family of carbapenemases, enzymes that inactivate these drugs. Among MBLs, New Delhi metallo-β-

lactamase 1 (NDM-1) has experienced the fastest and largest worldwide dissemination. This success has been attributed to the fact that NDM-1 is a lipidated protein anchored to the outer membrane of bacteria, while all other MBLs are soluble periplasmic enzymes. By means of a combined experimental and computational approach, we show that NDM-1 interacts with the surface of bacterial membranes in a stable, defined conformation, in which the active site is not occluded by the bilayer. Although the lipidation is required for a long-lasting interaction, the globular domain of NDM-1 is tuned to interact specifically with the outer bacterial membrane. In contrast, this affinity is not observed for VIM-2, a natively soluble MBL. We identified key residues involved in the membrane interaction of NDM-1, which constitute potential targets for developing therapeutic strategies able to combat resistance granted by this enzyme. Finally, we compare the NDM-1 potential to be secreted into vesicles with other MBLs.

3.1 Introduction

The emergence of bacterial resistance has led to a worldwide healthcare crisis.[190, 191] The fight of medicine against bacterial infections is thousands of years old, as this has historically been one of the main causes of human death.[118] Unlike other common lethal diseases that we need to face today, such as cancer and hearth diseases, bacterial infections have the potential to kill indistinguishably the young and the old, the sick and the healthy. As a matter of fact, the removal of bacterial infections from the main causes of human death has contributed to reach the most consistent increase in life expectancy in human history.[192] Despite its destructive potential, it is only relatively recently (last 200 years) that medicine developed the knowledge to fight bacterial infections. It is estimated that the medical treatments developed in the last two centuries (mainly vaccines and antibiotics) allowed to lower the number of deaths due to bacterial infections from 400 per 100'000 people to less than 50.[118] Hungarian physician Ignàc Semmelweis was the first to hypothesize that the high death rate among women during childbirth may be related to microorganisms coming from the external environment. By simply imposing to all physicians in his department to wash hands before examining the patients, he was able to reduce the death rate of women in labor from 11% to only 1%.[193] However, the strong opposition that his theories encountered across the medical community,

brought him to be ostracized and isolated, an event which shocked Dr. Semmelweis to the point that he got recovered in a mental institute. One of his strongest opponents was Rudolf Virchow, who is nowadays considered the father of modern pathology[194] (a situation that reminds us that being an expert and a visionary in one scientific field, does not necessarily bring to having correct opinions in all fields of science). Ironically and sadly, Dr. Semmelweis died from an infection following a surgical operation which, is easy to presume, was conducted without washing hands.[195] Despite this sad event, the fight against bacterial infections had only been delayed. The studies of Louis Pasteur[196] and Robert Koch[197] demonstrated that microorganisms were indeed the cause of common infections. These studies paved the way to the huge decrease in the number of deaths caused by bacterial infections, thanks to the spread of common hygienic practices (such as washing hands before surgical operations) and to the creation of the first vaccines.[198] (To provide an idea of how science can influence the course of history, it is useful to remind that the French-Prussian war of 1870, an event which according to all historians led to the settlement of a strong nationalistic sentiment in Germany, which would have later been the cause of two world wars in the XX century, was strongly influenced by vaccines. The French army was indeed devastated by a smallpox epidemic, which killed almost 25,000 soldiers. On the Prussian side, as a consequence of mass vaccination in the army, only 400 soldiers died of smallpox, and this allowed Prussia to defeat the French army which, at the time, was considered invincible[199]). Later, in 1928, Alexander Fleming discovered by serendipity penicillin, the first drug that could be used as a treatment to ongoing bacterial infections.[200] Chemically, penicillin is a molecule that possesses a β-lactam ring, and thanks to this can bind and inhibit the so-called penicillin-binding proteins. These proteins are fundamental for the creation of the peptidoglycan layer, which is in turn essential for the bacterial life cycle: indeed, the peptidoglycan layer provides mechanical stability to the bacteria, and without it, the microorganisms cannot survive.[201] Despite this potent treatment (penicillin is basically active towards any kind of bacterial infection, since the penicillin binding proteins are present both in gram-negative and in gram-positive bacteria), it was clear from the beginning that bacteria had the potential to find a way of surviving this treatment, thanks to Darwinian evolution. In particular, casual mutations in bacteria are able to develop certain characteristics of the microorganisms that may make them resistant to antibiotics. Since all the non-resistant bacterial strains are destroyed by the antibiotic, the bacterial strains

which develop antibiotic resistance, would have no opponents in spreading their offspring, and therefore they will find an extra-favourable environment where to proliferate. Antibiotic resistance is a phenomenon that was indeed predicted by Fleming himself, who once stated that "The time may come when penicillin can be bought by anyone in the shops. Then there is the danger that the ignorant man may easily under-dose himself and by exposing his microbes to nonlethal quantities of the drug make them resistant",[202] a statement that is more current than ever nowadays. Antibiotic resistance is indeed increased by misbehaviours of the community, including clinical misuse[203] (the unnecessary usage of antibiotics, for example to treat viral pathologies, for which they are totally useless) and farming misuse[204] (huge quantities of antibiotics are currently fed to livestock all over the world, both to prevent infections and to promote muscular growth).

The emergence of multi-resistant and pan-resistant microbes, for which treatment options are very limited, is also compounded by the lack of development of new antibiotics in the last decades.[205, 206] As a matter of fact, following the introduction of penicillin, a huge number of novel antibiotic molecules were identified and patented. In particular, the period which goes from 1950s to the 1970s is considered the 'golden era' for antibiotics, as they saw the rise of 12 different classes of antibiotics.[207] However, in the last 40 years, only 2 new classes of antibiotics have been introduced (Figure 3.1). The reason for this dramatic downgrade is not totally clear, but most of the observers attribute it to economical reasons[208]: in other words, it is getting too expensive for pharmaceutical companies to develop and patent novel antibiotic molecules. This is partially due to the stricter measurements that have been imposed over the years by agencies such as the Food and Drug Administration or the European Medical Agency, and partially to the increase in life expectancy, which correlates to the emergence of a larger number of chronic diseases, for which treatments are more lucrative, because they need to be administered for the life-time of the patient.[208] It would be therefore required that national governments take decisive actions, either in terms of incentives to the pharma-industry to develop novel antibiotics, or in terms of massive public investments for this specific research line. Amid calls from the WHO for urgent action in addressing this complex issue, estimates predict that infections from drug-resistant bacteria will become the major cause of death by 2050, if the current trend is not reversed.[119] The rise in carbapenem-resistant bacteria

is particularly concerning, since this class of β-lactam antibiotics is reserved as a last resort option for life-threatening infections.[209, 210] Carbapenems belong to the β-lactam class of antibiotics and therefore, like penicillin, are able to bind and inhibit the penicillin-binding proteins. The most prevalent cause of this resistance is the production of carbapenemases, hydrolases that degrade and inactivate these antibiotics.[211] Metallo-β-lactamases (MBLs) are Zn(II)-dependent enzymes that represent the largest family of carbapenemases.[87, 88] MBLs can also hydrolyze other β-lactam antibiotics such as cephalosporins and penicillins.[90, 91] At the moment, there are no clinically approved inhibitors for MBLs, so that bacterial resistance mediated by these proteins cannot be countered.[212]

1953: Glycopeptides, Nitroimidazoles, Streptogramins ◄ ► 1955: Cycloserine, Novobiocin
1952: Macrolides ◄ ► 1957: Rifamycins
1950: Pleuromutilins ◄ ► 1961: Trimethoprim
1948: Cephalosporins ◄ ► 1962: Quinolones, Lincosamides, Fusidic acid
1947: Polymyxins, Phenicols ◄ ► 1969: Fosfomycin
1946: Nitrofurans ◄ ► 1971: Mupirocin
1945: Tetracyclines ◄ ► 1976: Carbapenems
1943: Aminoglycosides, Bacitracin (topical) ◄ ► 1978: Oxazolidinones
1932: Sulfonamides ◄ ► 1979: Monobactams
1928: Penicillins ◄ ► 1987: Lipopeptides

1920 1930 1940 1950 1960 1970 1980 1990 2000 2010

DISCOVERY VOID

© ReAct Group 2015

Figure 3.1 – Antibiotics discovery timeline (figure from the ReAct group, adapted from Silver et al.[213]).

New Delhi metallo-β-lactamase 1 (NDM-1) is a MBL first identified in 2008[214] which has experienced a remarkably fast spread worldwide, having been detected in more than 100 countries distributed in all continents.[215, 216] This enzyme is not only widely disseminated in healthcare settings, but it also displays an unprecedented prevalence in the environment, with its coding gene being present in soil and water samples worldwide.[217, 218] The cellular localization of NDM-1 is unique among clinically relevant MBLs: while all others are soluble periplasmic enzymes, NDM-1 is a lipoprotein anchored to the inner leaflet of the outer membrane in Gram-negative bacteria.[109, 110] This post-translational modification is due to the presence of a lipidation signal within the signal peptide of the enzyme, which is recognized by a widely conserved lipoprotein biogenesis machinery located in the cell envelope of bacteria.[219, 220] The lipid moiety is covalently bound to a Cys residue (Cys26, Figure 3.2a) present in the signal peptide, which becomes the N-terminal residue of the mature protein, and is responsible for membrane anchoring.

Membrane anchoring endows NDM-1 with unique features within the bacterial host that have favored its worldwide dissemination and persistence.[110] First, this cellular localization stabilizes the enzyme upon conditions of metal starvation that occur at the infection sites. Second, and most important, membrane binding favors secretion of NDM-1 into outer-membrane vesicles (OMVs). These lipid-enclosed particles are released by all Gram-negative bacteria,[221] and provide a mechanism for the dissemination of NDM-1. On one hand, these vesicles expand the spatial domain of antibiotic hydrolysis beyond the bacterial cells. On the other hand, OMVs also protect populations of bacteria that would otherwise be sensitive to antibiotics. Thus, membrane anchoring provides several evolutionary advantages to NDM-1.

The biochemical and biophysical characterization of NDM-1 has been restricted to its soluble domain. Indeed, all available crystal structures of the protein have been obtained with truncated soluble forms lacking the lipid group.[109,222,223] In addition, computational studies on NDM-1 that are present in literature,[224–226] generally focus on the catalytic action of the enzyme, regardless of its membrane anchoring. As a result, the molecular details of the interaction of NDM-1 with the membrane surface are unknown. There are several unaddressed questions regarding the membrane association mechanism of NDM-1: (1) how does the soluble domain of NDM-1 (in particular, the active site) orient with respect to the bacterial membrane? (2) is lipidation the only mechanism contributing to membrane association, or are there specific interactions of the bilayer with the soluble domain? (3) is NDM-1 tuned to interact more favorably with the outer membrane environment rather than with the inner membrane? Answering these questions could assist in designing drugs that thwart these interactions, thus hampering the distinct advantages conferred by membrane anchoring.

By combining experimental and computational approaches, we show that NDM-1 adopts a defined orientation with respect to the membrane, characterized by the synergistic anchoring action of the lipidated cysteine and a molecular surface with specific affinity for the outer bacterial membrane. As a result, the NDM-1 active site faces the periplasmic space and is exposed to the solvent, hence without spatial restrictions that limit its hydrolytic action. We have identified residues that are critical for this interaction, as well as the effect of the membrane composition on the protein-membrane interaction. Overall, this picture reveals that the soluble domain contributes favorably to the interaction with the outer membrane of

Gram negative bacteria, potentiating the effect of enzyme lipidation.

Moreover, we compared the membrane affinity of NDM-1 to other known MBLs, in order to evaluate their potential to be secreted into vesicles.

3.2 Results and Discussion

3.2.1 NDM-1 presents restricted mobility in contact with the membrane

As a first step to characterize the interaction of NDM-1 with the bacterial membrane, we aimed to obtain information about the proximity and relative mobility of NDM-1 in its anchored form with respect to the membrane bilayer surface. We decided to use proteoliposomes as a model system amenable to in vitro characterization. Purified NDM-1 was introduced into liposomes with a phospholipid composition mimicking the inner leaflet of the outer membrane of E. coli, which is formed by neutral PhosphatidylEthanolamines (PEs, 91%), and by two anionic lipids: Cardiolipins (CDLs, 6%) and PhosphatidylGlycerols (PGs, 3%).[227] This was achieved by using liposomes with 91% POPE (1-palmitoyl-2-oleoyl-phosphatidylethanolamine), 6% tetraoleoyl cardiolipin, and 3% POPG (1-palmitoyl-2-oleoyl-phosphatidylglycerol). NDM-1 bound to proteoliposomes may be oriented outwards (with the lipid group inserted within the outer leaflet of the membrane), inwards (with the protein encapsulated within the liposome), or with the presence of mixed populations exhibiting each orientation. Treatment of proteoliposomes with proteinase K resulted in full degradation of NDM-1 (Figure 3.2b), indicating that all protein molecules were located on the outer surface of the liposomes.

We then sought to evaluate whether NDM-1 assumes a fixed orientation with respect to the membrane or if it is sampling different conformations. Fluorescence anisotropy measurements enable determining the tumbling rates of molecules in solution, as higher rotational correlation times cause increased anisotropy in the emitted fluorescence. We measured the fluorescence anisotropies of lipidated NDM-1 in liposomes, and of the soluble form of the enzyme, NDM-1 C26A. In this mutant, the lipidated Cys26 residue is replaced by an alanine, resulting in a soluble periplasmic enzyme 19. Since NDM-1 possesses four tryptophan residues distributed along the protein structure (Figure 3.2a), we exploited the intrinsic anisotropy of
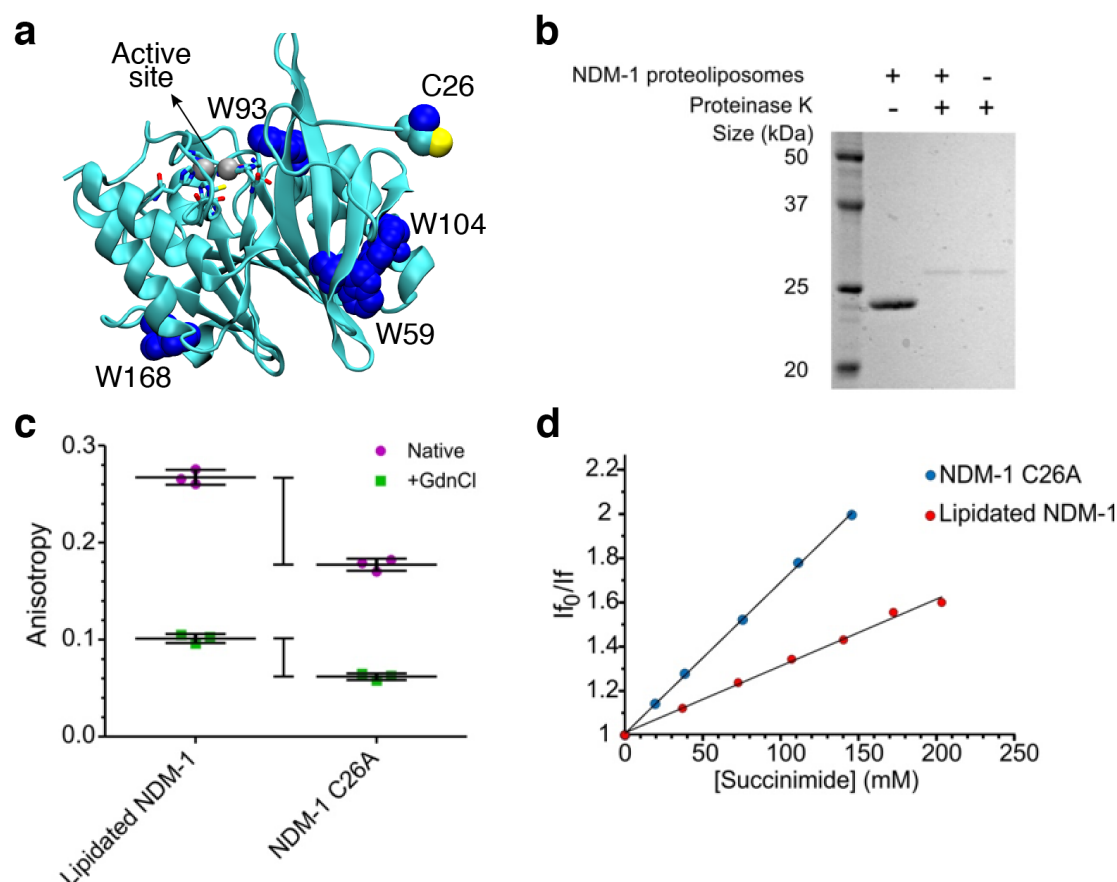
Figure 3.2 – NDM-1 mobility at the membrane interface. (a) Location of tryptophan residues, Cys26 (highlighted in the lipidated form) and active site in the NDM-1 crystal structure (PDB: 5ZGE 30). (b) Analysis of NDM-1 orientation in proteoliposomes, using proteolysis by proteinase K. (c) Tryptophan fluorescence anisotropy values for lipidated NDM-1 in liposomes and soluble NDM-1 C26A, in absence (Native) or in presence (+GdnCl) of 4.5 M guanidinium chloride. Data for each protein are presented as mean ± st.dev. of 3 biological replicates, individual data points are presented in color. (d) Stern-Volmer plots of fluorescence quenching by succinimide of lipidated NDM-1 in liposomes and soluble NDM-1 C26A.

the protein upon excitation at 298 nm. The membrane-anchored form of the protein presents a much higher (0.267 ± 0.008) anisotropy than the soluble NDM-1 C26A variant (0.177 ± 0.006), suggesting that NDM-1 assumes a fixed orientation with respect to the membrane (Figure 3.2c). A different scenario was found when measuring fluorescence anisotropy of the unfolded states of both variants (by addition of 4.5 M guanidinium chloride). Indeed, the difference in anisotropy between lipidated and soluble NDM-1 was much smaller compared to the folded proteins (0.039 vs. 0.090) (Figure 3.2c). Thus, despite the proximity to the liposome indeed induces a restricted mobility, the soluble domain seems to have a specific interaction with the

bilayer that is only present in the folded protein.

Since three out of the four Trp residues in NDM-1 are partially exposed to the solvent (Figure 3.2a), we sought to confirm that the protein contacts the membrane surface by determining if any of these residues becomes occluded in the anchored protein. We carried out Trp fluorescence quenching assays with lipidated NDM-1 in liposomes and soluble NDM-1 C26A, using the non-ionic quencher succinimide. Since fluorescence quenching requires almost direct contact of the fluorophore and quencher, if any of the Trp residues is directly contacting the surface of the membrane, it should change its accessibility to the quencher in the membrane bound form of NDM-1 with respect to NDM-1 C26A in solution. Quenching data fit to a linear Stern-Volmer plot, indicating that all Trp residues both in NDM-1 C26A and in the liposome-bound NDM-1 are susceptible to quenching by succinimide. However, the two samples presented differences, as the lipidated protein displayed a significantly smaller Stern-Volmer constant (KSV) with respect to the soluble form (Figure 3.2d). A smaller KSV correlates to a restricted diffusion constant of the quencher towards the protein. As a consequence, the same degree of quenching in the lipidated protein can be achieved with a higher concentration than that required for the soluble NDM-1 C26A. We attribute this finding to the proximity of the liposome in the case of lipidated NDM-1.

In summary, fluorescence anisotropy and fluorescence quenching experiments point to a close interaction of NDM-1 with the membrane surface, in which the protein adopts a stable conformation with respect to the bilayer.

3.2.2 The soluble domain of NDM-1 contributes to membrane association

In order to characterize in further detail the interaction of NDM-1 with the membrane, we conducted coarse-grained (CG) molecular dynamics (MD) simulations using the MARTINI force field. We generated models of NDM-1 in the lipidated (wild type) and soluble forms (C26A), considering both their holo and apo forms, i.e. with and without zinc ions at the catalytic site (Figure 3.3). The membrane bilayer was modeled with the same lipid composition used in the liposomes mimicking the inner leaflet of the bacterial outer membrane.

The enzyme was located at 4 nm at least from the bilayer in the starting geometry in
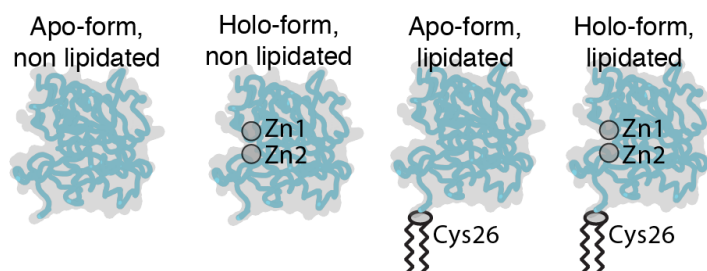
Figure 3.3 – Summary of the protein states that were studied with CG-MD simulations.

order to avoid initial protein-membrane interactions that may bias the simulations. After exploring several conformations, lipidated NDM-1 spontaneously binds to the membrane by insertion of the lipid moiety into the bilayer. NDM-1 C26A also reaches a stable binding conformation to the membrane within the same timescale (Figure 3.4a,b), showing a similar number of binding events compared to the lipidated variant (Figure 3.4c). Indeed, we observed membrane binding in all the CG-MD replicas for both protein forms. The membrane-bound forms, which adopt similar orientations in all the different replicas, reveal that the active site is not occluded by the interaction with the membrane (see following section). Remarkably, the soluble and lipidated NDM-1 display the same surface of interaction with the membrane. The CG-MD trajectories show that the tip of the N-terminal tail has a natural propensity to interact with the membrane even in absence of lipidation.

The Zn(II) content at the active site does not affect the interaction with the membrane, as we observed a similar binding propensity in both the apo and the holo form, (Figure 3.5a) despite their different net charges. This can be attributed to the fact that the active site is oriented on the opposite side of the surface anchored to the membrane. Additionally, the Zn(II) ions are not located at the protein surface, reducing their impact in surface electrostatics.

These results suggest that NDM-1 has the tendency to interact with the outer bacterial membrane regardless of its metal content and lipidation at Cys26. Despite lipidation is fundamental for membrane association and in vivo requires a complex biogenesis machinery 21, these simulations suggest that the soluble domain of NDM-1 has native electrostatic features that favor its interaction with the membrane bilayer. The effect of lipidation probably acts on reinforcing and stabilizing membrane anchoring on a longer time scale.
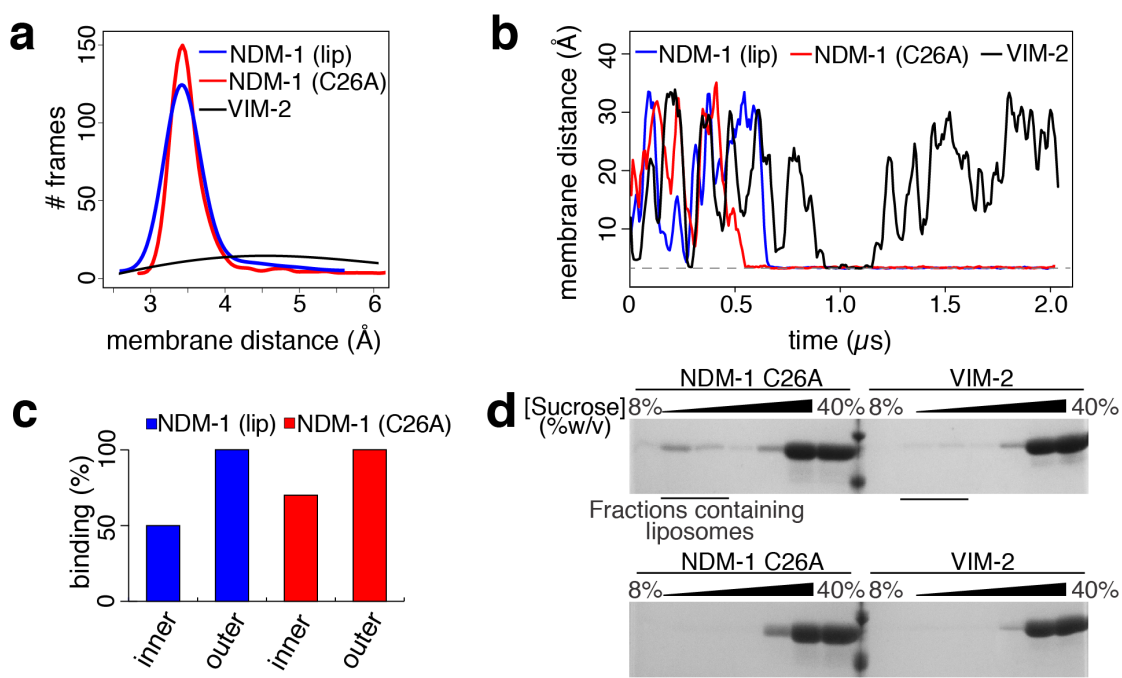
Figure 3.4 – Membrane affinity of NDM-1 in different conditions. (a) Distribution of CG MD trajectory frames, collected every 750 ps, with respect to the protein/membrane distance of lipidated NDM-1, soluble NDM-1 (NDM-1 C26A) and VIM-2. (b) Protein-membrane distance evolution for lipidated and soluble NDM-1, and for VIM-2. One representative replica for each protein is reported. (c) Percentage of binding events observed during the CG MD simulations for the tested systems. (d) SDS-PAGE analysis of sucrose gradient fractions from liposome flotation assays of NDM-1 C26A and VIM-2. The flotation assays were carried out using liposomes made with an outer membrane composition (top) or E. coli Polar Lipid Extract (bottom).

We also simulated the enzyme VIM-2 as a model of a soluble periplasmic MBL in its native state. In this case, no interaction with the membrane was observed (except for some transient ones, <200 ns, Figure 3.4a,b and Figure 3.5b), in contrast to soluble NDM-1 C26A. It appears that different MBLs with high degree of structural homology can present distinct affinities towards the bacterial membrane. Therefore, we can conclude that the soluble domain of NDM-1 is specifically adapted to interact with the membrane, in contrast to VIM-2.

To test the predictions of the molecular simulations, we carried out liposome flotation assays by incubating NDM-1 C26A with liposomes mimicking the outer membrane composition of *E. coli*. Samples were loaded at the bottom of a discontinuous sucrose gradient and ultracentrifuged. Under these conditions the liposomes tend to float along the gradient towards lower sucrose concentrations, carrying with them bound proteins, while free protein

remains at the bottom of the gradient. We used SDS-PAGE to analyze the distribution of NDM-1 C26A after ultracentrifugation, and clearly observed that this soluble enzyme indeed binds the vesicles (Figure 3.4d, top), with a portion of the protein having migrated with the liposomes to a lower sucrose concentration. Instead, no binding of VIM-2 could be detected (Figure 3.4d, top).

These results confirm that NDM-1 has the ability to bind to the membrane regardless of the presence of the lipidated cysteine residue. The low amount of bound NDM-1 that we observed suggests a transient interaction, leading to a high proportion of the protein being located in its initial position at the bottom of the gradient. The presence of the lipid group in the wild-type protein is required to fully stabilize the interaction, granting a more defined and long-lasting contact between the protein and the membrane.

In order to explore the effect of membrane localization and composition, we used CG-MD simulations to probe NDM-1 association to the inner bacterial membrane. In *E. coli*, the inner bacterial membrane is characterized by a significantly higher content in charged lipids, in particular PGs (67% PEs, 28% PGs, and 5% CDLs) 32. The simulations revealed that NDM-1 has a lower affinity for the inner membrane compared to the outer bilayer: the binding events of all protein states (i.e., apo/holo form and with/without lipidation) represent 60% of the CG-MD replicas, in contrast to the 100% of replicas showing binding events to the outer membrane (Figure 3.4c, Figure 3.5a, Figure 3.6). Analogous simulations on the soluble MBL VIM-2 revealed no affinity for the inner bacterial membrane model.

These predictions were also validated by performing liposome flotation assays for NDM-1 C26A with liposomes produced with the *E. coli* polar lipid extract, which contains a lipid composition mimicking the inner membrane, resulting in no detectable binding (Figure 3.4d, bottom). Similar experiments with VIM-2 also resulted in no binding, confirming the behavior predicted by CG-MD simulations. Based on these results, we then decided to identify the molecular features favoring the interaction of NDM-1 with the outer bacterial membrane.

a

NDM-1 form	Membrane	Inner/Outer (MD length)	Replica				
			1	2	3	4	5
Apo / No lipidation	Bacterial	INNER (2 μs)	■	□	■	■	■
		OUTER (2 μs to 10)	■	■	■	■	■
	Bacterial / No CDL	INNER (2 μs)	□	□	□	□	□
		OUTER (2 μs)	■	□	■	■	■
Apo / With lipidation	Bacterial	INNER (2 μs)	■	□	■	■	□
		OUTER (2 μs to 10)	■	■	■	■	■
	Bacterial / No CDL	INNER (2 μs)	□	□	□	□	□
		OUTER (2 μs)	□	■	□	■	□
Holo / No lipidation	Bacterial	INNER (2 μs)	■	■	■	□	■
		OUTER (2 μs to 10)	■	□	■	■	□
	Bacterial / No CDL	INNER (2 μs)	□	■	■	□	■
		OUTER (2 μs)	■	□	■	■	□
Holo / With lipidation	Bacterial	INNER (2 μs to 10)	■	□	■	□	□
		OUTER (2 μs to 10)	■	■	■	■	□
	Bacterial / No CDL	INNER (2 μs)	□	□	□	□	□
		OUTER (2 μs to 10)	□	■	□	□	■
R45E / R52E	Bacterial	OUTER (2 μs)	■	□	□	□	■

b

VIM-2 form	Membrane	Inner/Outer (sim. length)	Replica				
			1	2	3	4	5
Holo / No lipidation	Bacterial	INNER (2 μs)	□	□	□	□	□
		OUTER (2 μs to 10)	□	□	□	□	□

c

N-VIM form	Membrane	Inner/Outer (sim. length)	Replica				
			1	2	3	4	5
Apo / No lipidation	Bacterial	OUTER (2 μs)	□	□	□	□	□
Apo / With lipidation	Bacterial	OUTER (2 μs)	□	□	□	□	□
Holo / No lipidation	Bacterial	OUTER (2 μs)	□	□	□	□	□
Holo / With lipidation	Bacterial	OUTER (2 μs to 10)	□	□	■	□	□

Figure 3.5 – List of all CG MD simulation and the relative binding events observed for (a) NDM-1, (b) VIM-2 and (c) N-VIM, in different protein and membrane states. A black box means that the protein/membrane binding event has happened, while a grey box means that no binding was observed.

55
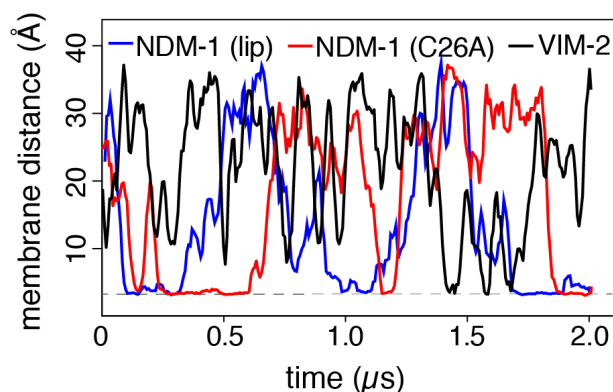
Figure 3.6 – Protein-membrane distance evolution for lipidated NDM-1, soluble NDM-1 and VIM-2 in presence of the inner bacterial membrane model. One representative replica for each protein is reported.

3.2.3 Specific interaction of the soluble domain of NDM-1 with the outer membrane

The analysis of the CG-MD trajectories reveals that NDM-1 interacts with the membrane by means of a specific patch at the protein surface. During the initial stages of the simulation, NDM-1 explores different orientations, but only one of them elicits a neat binding event (Figure 3.7a). While the lipidated Cys26 inserts directly into the membrane, the globular domain interacts with the membrane by means of the -strand domain spanning from Thr41 to Val58 and the segment between Asn103 and Pro112 (which includes a portion of -helix, Asn103-Ile109, and a portion of a loop, Asn110-Pro112, Figure 3.7a). These two regions include six charged residues: Arg45, Arg52 and Lys106, Asp43, Asp48 and Glu108 (Figure 3.8). Particularly the three basic amino acids might be key residues in providing the affinity of the globular domain of NDM-1 towards the membrane. While the overall charge of NDM-1 is negative (-5e in the holo form and -7e in the apo form), these positively charged residues could provide a favorable electrostatic interaction with the membrane. It is important to observe that the identified configuration orients the active site of NDM-1 exposed to the solvent: the enzyme is therefore still able to capture, host and cleave -lactam molecules, without being hindered by the presence of the membrane. In addition, we analyzed the variation in the tilt angle of the protein across the simulations, which is the angle formed by the catalytic site, the center of mass of the protein, and its projection on the membrane. NDM-1 adopted an

orientation with a constant tilt angle (~150º, Figure 3.7b), reflecting that the protein adopts a well-defined conformation with respect to the membrane surface.
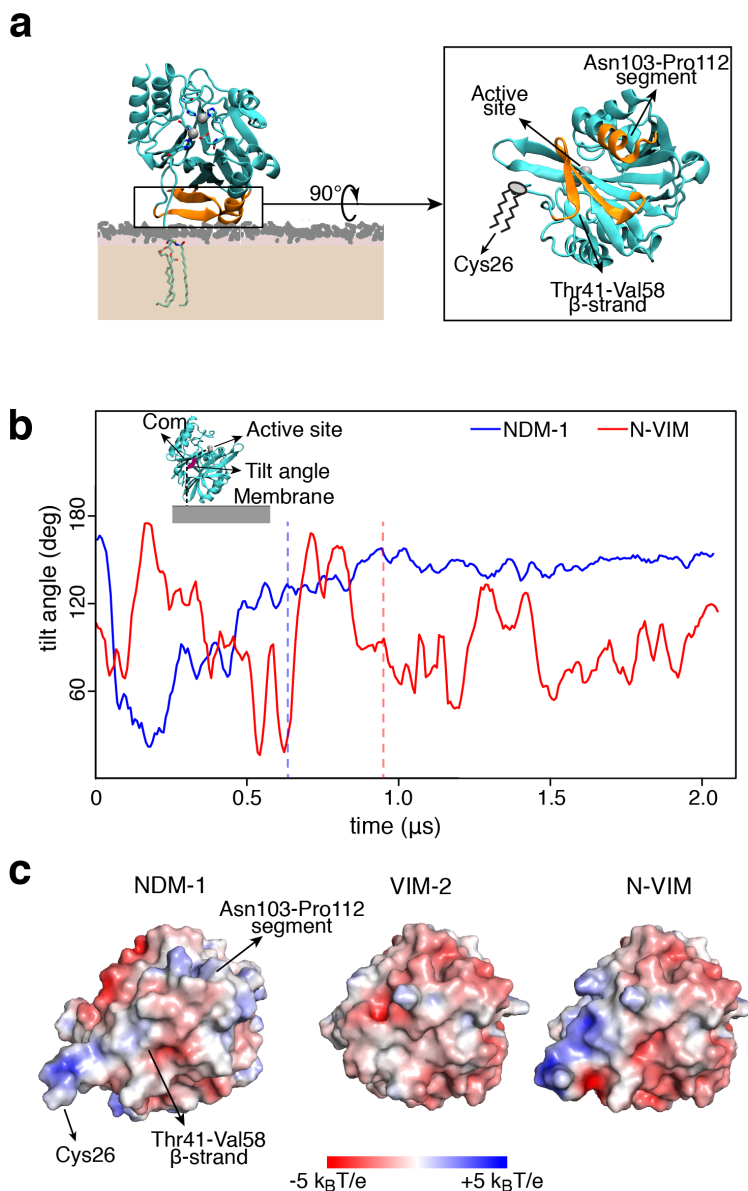
Figure 3.7 – A patch in the NDM-1 soluble domain drives the interaction with the membrane. (a) Final and stable orientation of NDM-1 with respect to the membrane, and protein domains that are involved. The active site is exposed to the solvent and not hindered by the membrane. (b) Tilt angle definition and evolution of the tilt angle vs. time, for NDM-1 and N-VIM at the outer bacterial membrane. The tilt angle is defined as the angle formed by the active site, the center of mass of the protein and the projection of the center of mass of the protein on the plane of the membrane. The vertical dashed lines represent the moment in which the anchoring occurs. (c) Electrostatic potential of the protein-membrane surface of interaction (seen from the membrane surface) for NDM-1, VIM-2 and N-VIM.
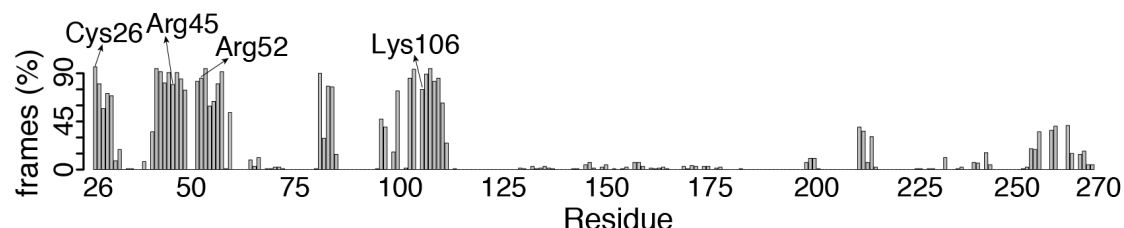
Figure 3.8 – Contacts with the outer bacterial membrane, expressed in percentages of total number of frames (collected every 750 ps) for each NDM-1 residue across the CG MD simulations.

To further evaluate the contribution of the soluble domain to this interaction, we tested the behavior of a chimera between NDM-1 and VIM-2. This protein, named N-VIM, contains the main core of the soluble enzyme VIM-2 with the N-terminal region of NDM-1 (including the lipidation site Cys26), resulting in a membrane-bound protein.[110] In the CG-MD simulations, the chimeric protein N-VIM did not show a significantly improved binding with respect to the native, soluble VIM-2 protein (Figure 3.5c). Analysis of the tilt angle shows that N-VIM did not achieve a stable membrane anchoring, in contrast with wild type NDM-1 (Figure 3.7b). Clustering analysis of the MD trajectories confirms that the orientation of N-VIM with respect to the membrane covers a broad range of conformations, in contrast to the stable and defined interaction observed for NDM-1 (Figure 3.9).
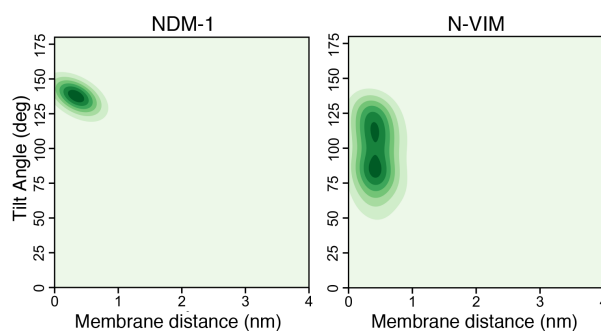
Figure 3.9 – Tilt angle distribution vs. protein/membrane distance across the CG MD simulations for NDM-1 and N-VIM, respectively.

The diverse interaction of NDM-1, VIM-2 and N-VIM can be accounted for by analyzing the electrostatic potential of the 3 enzymes (Figure 3.7c). NDM-1 shows a larger surface of positive electrostatic potential with respect to VIM-2 in the region that drives membrane anchoring to the negatively charged phospholipid bilayer. Comparing the interaction surface

of these proteins, we observed that VIM-2 is almost completely negatively charged, whilst N-VIM is characterized by an electrostatic profile that is intermediate between NDM-1 and VIM-2 (Figure 3.10). The behavior of N-VIM is probably caused by the lack of most of the surface of interaction identified in NDM-1. In particular, the surface of NDM-1 is more positive than N-VIM in the region corresponding to the Asn103-Pro112 segment, which is reasonable since in this area N-VIM has the same primary sequence as VIM-2.
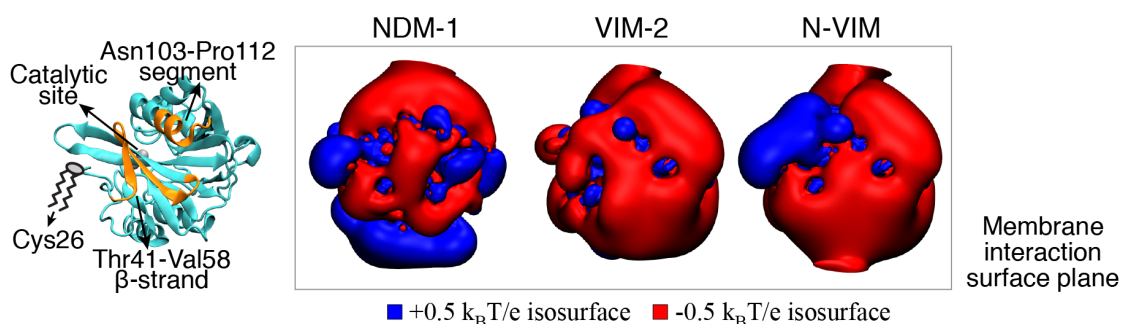
Figure 3.10 – Electrostatic potential isosurfaces at ± 0.5 $k_B T/e$ for NDM-1, VIM-2 and N-VIM, respectively.

In order to better define the molecular interaction at the protein-membrane interface, we performed all-atom (AA) MD simulations based on the CG models. These calculations confirmed the affinity between NDM-1 and the bacterial membrane. The behavior of NDM-1 largely coincided in both types of simulations, although we observed a partial reorientation of the protein on the membrane surface (Figure 3.12a). This suggests a more prominent role for the -strand Thr41-Val58, with respect to the segment Asn103-Pro112, as confirmed by the analysis of the residues contacts (Figure 3.11). Despite these minor differences, both CG- and AA-MD simulations reveal that NDM-1 adopts a stable orientation with respect to the membrane, in excellent agreement with the fluorescence anisotropy results (Figure 3.2).
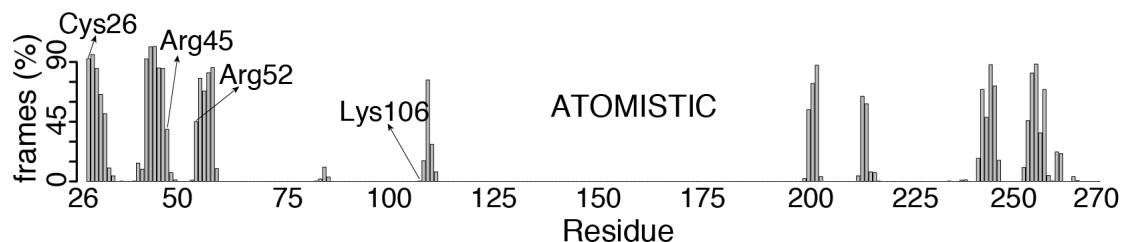
Figure 3.11 – Contacts with the membrane, expressed in percentages of total number of frames, collected every 750 ps, for each NDM-1 residue across the AA MD simulation.

These simulations also agree in pinpointing the Thr41-Val58 stretch as the main contributor to the interaction between NDM-1 and the outer membrane. To challenge this hypothesis, we ran CG-MD simulations of an NDM-1 variant in which the positively charged residues Arg45 and Arg52 were replaced by glutamates. This in silico mutation resulted in a reduction of the total number of binding events to the outer membrane with respect to wild type NDM-1, that dropped from 100% to 20% (Figure 3.5a). The tilt angle of the bound NDM-1 double mutant R45E-R52E resembled that of N-VIM (Figure 3.12b), suggesting that these mutations indeed destabilize the protein-membrane interaction.

To validate these predictions, we experimentally evaluated the impact of these mutations in the interaction of the soluble form of the protein with the membrane. Flotation assays with liposomes mimicking the outer membrane showed that these mutations indeed reduced the binding of NDM-1 to the liposomes, with a behavior closer to that of VIM-2 than to wild type NDM-1 (Figure 3.12c). Overall, these data provide compelling evidence of the role of the electrostatic interaction between the soluble domain of NDM-1 and the outer membrane in maintaining membrane association. We conclude that the soluble domain of NDM-1, in contrast to VIM-2, has been selected during evolution to better interact with the membrane, and that the presence of the N-terminal lipid group is not sufficient on its own for a stable interaction with the membrane.

3.2.4   Role of cardiolipin in NDM-1 binding to the outer bacterial membrane

CG MD simulations showed a clear preference of NDM-1 for the outer bacterial membrane model compared to the inner one (100% vs. 60% of replicas showing binding, Figure 3.4c, Figure 3.6 and Figure 3.5a), despite its lower content in charged lipids (9% vs. 33% of anionic lipids). This is against the contention that electrostatics is expected to be the main driving force of the interaction between phospholipid bilayers and peripheral binding proteins. The role of charged lipids, and cardiolipins (CDLs) in particular, has been reported as fundamental for the correct functioning of bacterial and mitochondrial membrane proteins.[66,68,228,229] In order to evaluate the role of CDLs in the interaction between the bacterial membrane and NDM-1, we conducted CG MD simulations in absence of this specific lipid, but keeping fixed the amount of PGs. This resulted in a significant reduction in the number of binding events
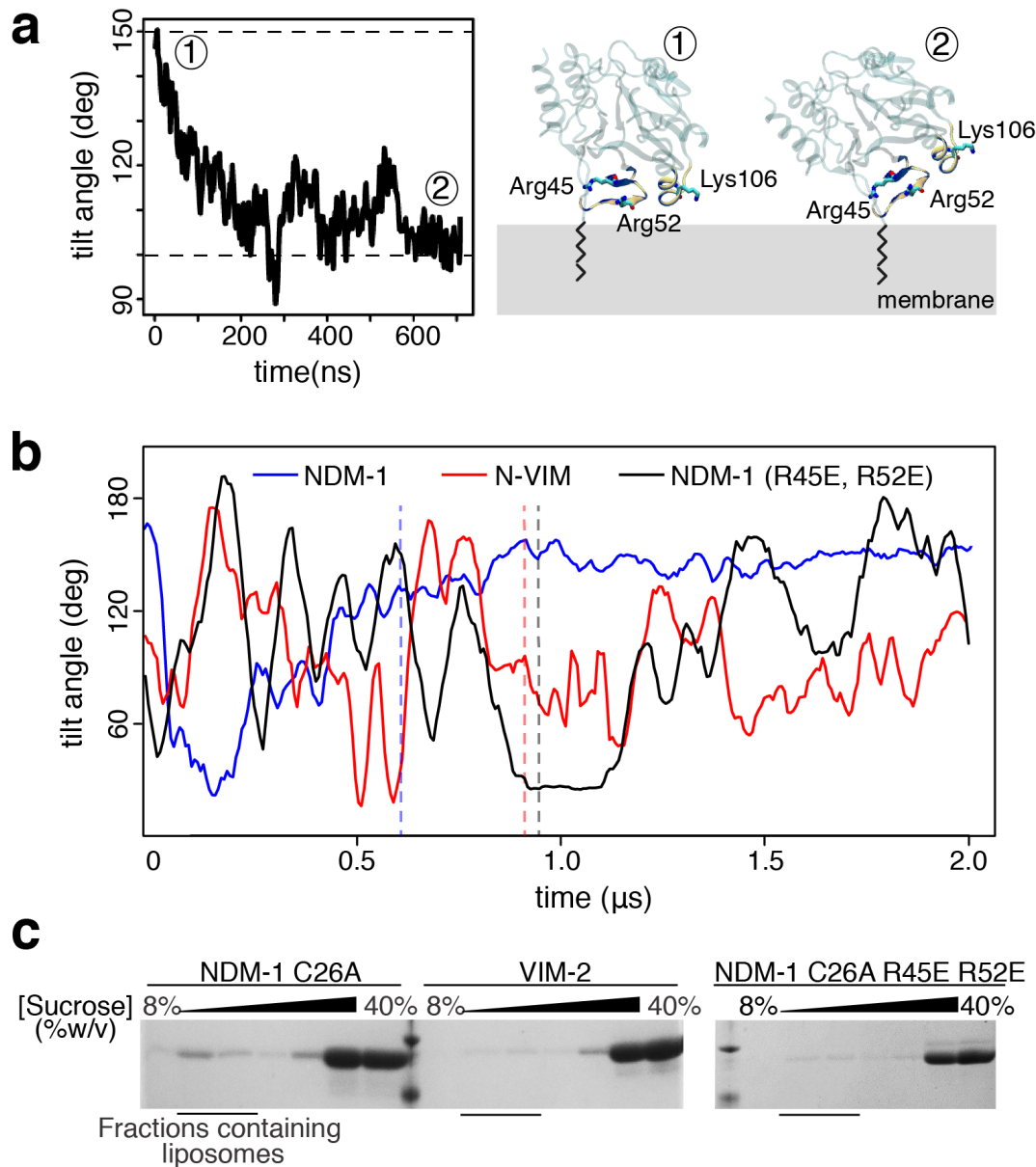
Figure 3.12 – Analysis of residues involved in NDM-1/membrane interaction. (a) Tilt angle evolution of NDM-1 during all-atom MD simulations (left) and comparison between the surface of interaction identified during the CG (~150°) and all-atom (~100°) MD simulations (right). (b) Comparison of tilt angle evolution during MD simulations for wild type NDM-1, N-VIM and mutant NDM-1 R45E-R52E. The vertical dashed lines represent the moment in which the membrane anchoring occurs. (c) SDS-PAGE analysis of sucrose gradient fractions from liposome flotation assays of NDM-1 C26A, VIM-2 and NDM-1 C26A R45E-R52E, using liposomes with an outer membrane composition.

(40% of replicas compared to 100% in presence of CDLs, Figure 3.5a). The same trend was observed in simulations performed with an inner bacterial membrane model without CDLs (Figure 3.5a). In both cases, NDM-1 approaches and explores the bilayer surface without ever establishing a stable contact (Figure 3.13a), resembling the situation previously described for VIM-2. These findings support the idea that the nature of the lipids present in the membrane plays also a fundamental role in NDM-1 recognition and association.

Driven by these results, we further analyzed the average occupancy of lipids in proximity of membrane-bound NDM-1. Clusters of charged lipids (in particular CDLs) form close to the NDM-1 binding surface (Figure 3.13b). In contrast, it was not possible to identify a specific pattern of lipid contacts for VIM-2, since in this case the contacts are transient and never turn into stable and long-term interactions. However, the lipid arrangement within the membrane patch seems to be completely non-specific (Figure 3.14). This suggests that the presence of NDM-1 in close proximity with the membrane is fundamental in order to drive the assembly of lipid clusters.

In order to further understand whether the physical proximity to the membrane (regardless of a specific affinity) could drive the recruitment of charged lipids, we tested the behavior of N-VIM. In the only replica in which membrane binding occurred for N-VIM, we could not detect any specific, more dominant, contact with lipid species (Figure 3.13b). We conclude that recruitment of charged lipids – specifically CDLs – is a fundamental step for the generation of a stable and strong association with the membrane, which needs to be driven by a specific protein surface. Lipid distribution analysis on the AA MD trajectories indicates that clusters of CDLs are still present and contribute to maintain a strong contact between NDM-1 and the phospholipid bilayer (Figure 3.15). In particular, we observed two CDL molecules generating long-lasting interactions (i.e. for the whole length of the simulation) with the residues that we identified in the previous section, i.e. Arg45 and Arg52 (Figure 3.16). The proximity to CDLs confirms the importance of these aminoacids in generating a strong and stable contact with the membrane, and explains why the NDM-1 R45E-R52E mutant is characterized by a smaller interaction with the bilayer (Figure 3.12b,c).
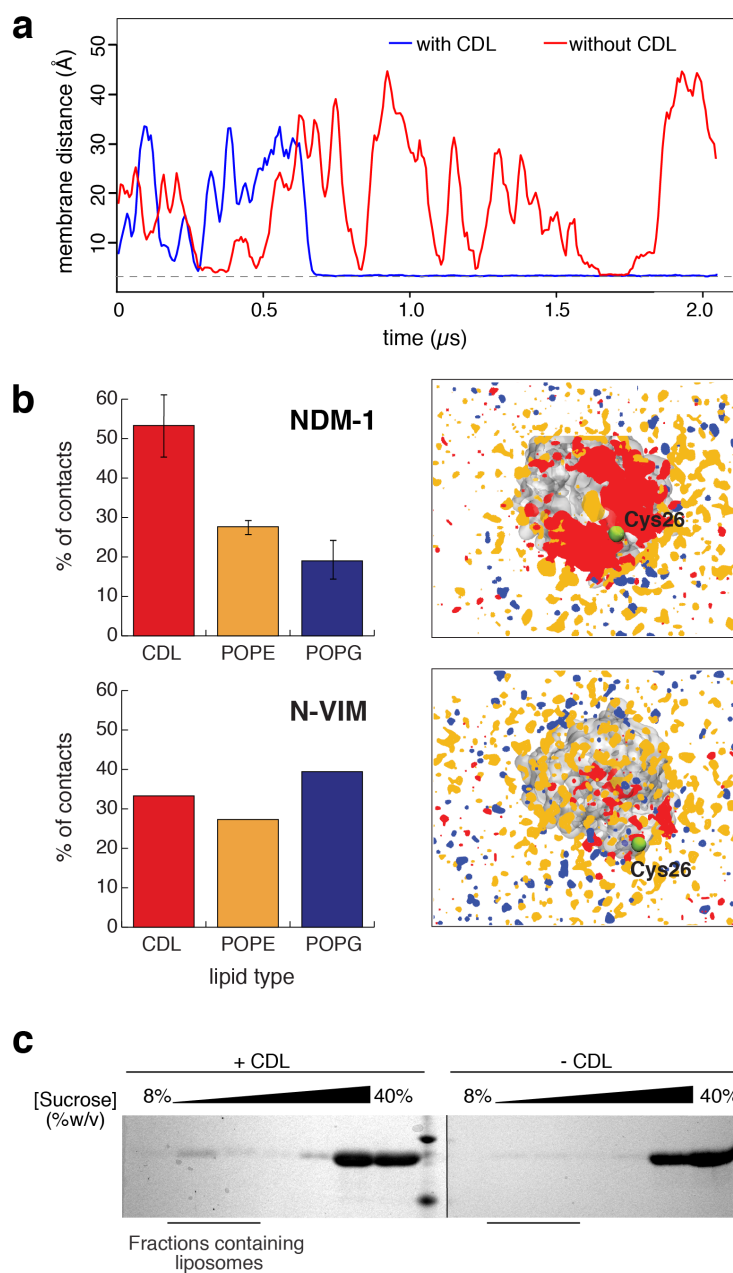
Figure 3.13 – Role of cardiolipins in NDM-1/membrane interaction. (a) Distance evolution of NDM-1 vs. outer membrane distance during CG MD simulations, with and without CDLs. One representative replica for each membrane model is reported. (b) Lipid contacts between the enzymes and the different lipid types that compose the bilayer, normalized according to lipid composition of the membrane (left); lipid distribution in proximity of NDM-1 and N-VIM, averaged over all the MD trajectories (right); (c) SDS-PAGE analysis of sucrose gradient fractions from liposome flotation assays of NDM-1 C26A, using liposomes with an outer membrane composition (+CDL) or liposomes with an equivalent composition but lacking cardiolipin (-CDL).
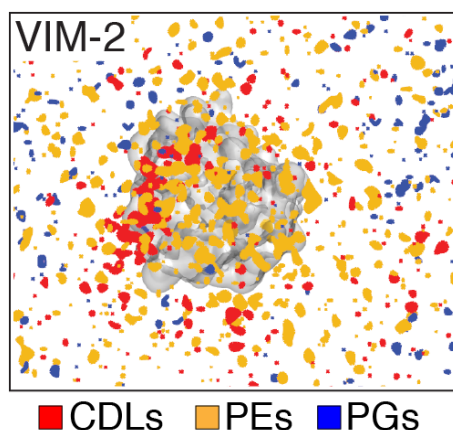
Figure 3.14 – Lipid distribution in proximity of VIM-2 during CG MD simulations.
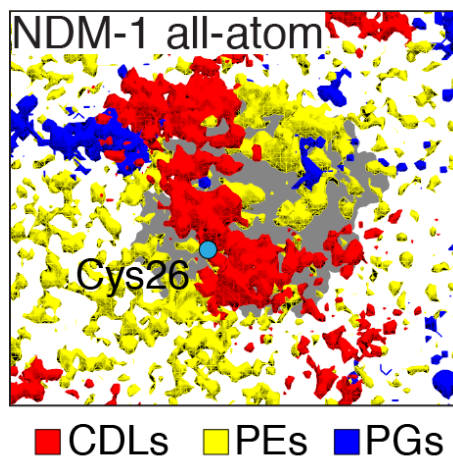
Figure 3.15 – Lipid distribution in proximity of lipidated NDM-1 in holo-form during AA MD simulations.

To experimentally test these predictions, we carried out liposome flotation assays with NDM-1 C26A using liposomes containing 97% PE and 3% PG (without CDLs). Removal of CDLs indeed abolishes binding of soluble NDM-1 to the liposomes (Figure 3.13c), in agreement with the weaker interaction predicted by simulation studies.

These results pinpoint a crucial role of CDLs in the recruitment and binding of NDM-1 to the membrane, confirming both the impact of the membrane composition for membrane-protein association and the role of electrostatics in the interaction of the soluble domain of NDM-1 to the membrane.
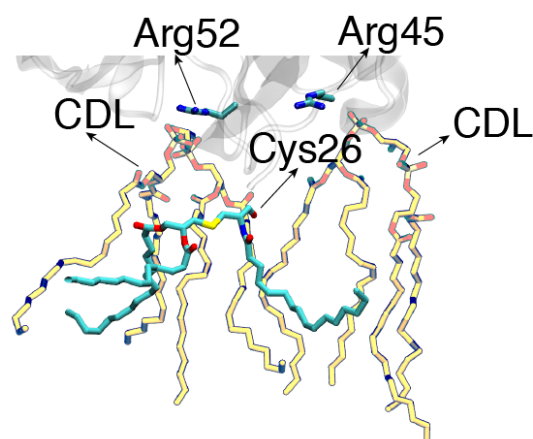
Figure 3.16 – CDLs interacting with Arg45 and Arg52 of NDM-1.

3.2.5 Secretion of MBLs into outer membrane vesicles

As previously reported, the facilitated secretion into outer membrane vesicles (OMVs) is the main reason why the NDM-1 membrane anchoring makes this peculiar MBL more resistant with respect to other enzymes of the same class. OMVs are spherical phospholipid constructs that extrude from the bacterial membrane, more frequently from Gram-negative bacteria. They generally participate in bacterial signalling, and in particular they can carry different biological material, such as enzymes, DNA/RNA or virulence factors.[230] OMVs are generally characterized by the same lipid composition of the membrane from which they extruded. Often, they would also comprehend an external layer of lipopolysaccharides.[231]

In addition to bacterial communication and to the expansion of their virulence power, OMVs secretion brings another fundamental advantage to bacteria: they allow to release toxic proteins from the periplasm, thus alleviating the envelope stress, which guarantees a longer bacterial life length.[232] However, the stress that a protein can cause is strongly dependent on the microorganism family. Recent studies have indeed correlated the secretion into OMVs with the MBL type and with the bacterial family. In particular, tests were performed on *E. coli*, *P. aeruginosa* and *A. baumannii*, in order to measure the quantity of 3 different MBLs, namely NDM-1, VIM-2 and SPM-1, within the secreted OMVs: the results showed that NDM-1 is secreted in OMVs from all the 3 tested bacterial families, whereas VIM-2 and SPM-1 are secreted in *E. coli* and *A. baumannii*, but not in *P. aeruginosa*[232] (Figure 3.17a/b). The same
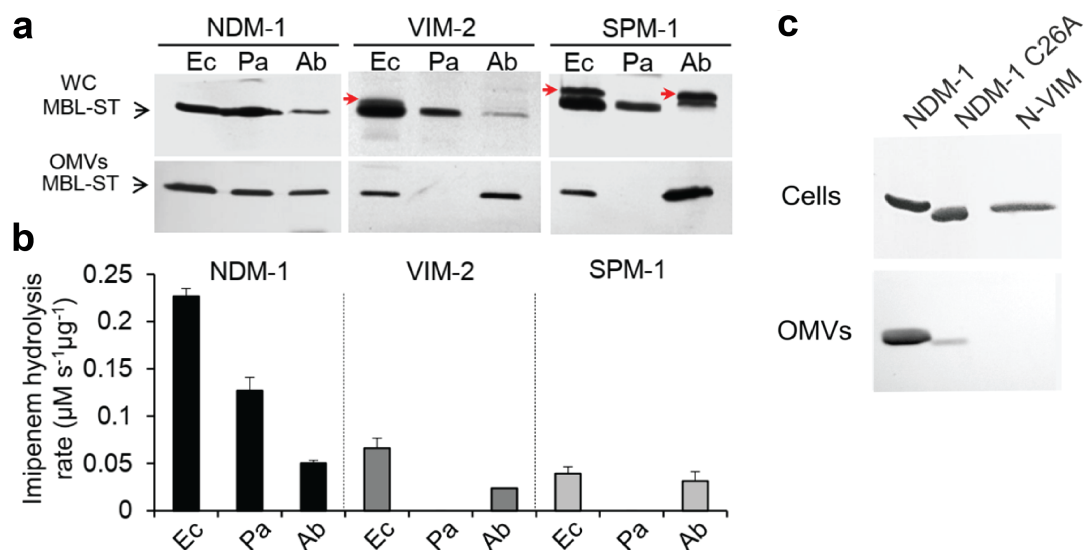
Figure 3.17 – (a) Concentration levels of three different MBLs (NDM-1, VIM-2 and SPM-1) into OMVs secreted by *E. coli*, *P. aeruginosa* and *A. baumannii.* (b) Antibiotic hydrolysis potential of the OMVs secreted by the bacteria represented in (a). (c) Concentration levels of lipidated NDM-1, soluble NDM-1 and N-VIM chimera in *E. coli.* (Figure of Lopez et al.[232])

tests were performed on the systems that we previously characterized, i.e. lipidated NDM-1, soluble NDM-1 and the N-VIM chimera. The results correlate with the membrane affinity as it was computationally predicted: in particular, high levels of lipidated NDM-1 were observed in OMVs; for what concerns the two soluble enzymes, a small amount of soluble NDM-1 was still found in OMVs. On the contrary, no trace of the N-VIM chimera was spotted (Figure 3.17c). This is in agreement with the molecular simulations presented in the previous sections, which predict the membrane affinity of soluble NDM-1 as stronger than the one of N-VIM. It would indeed be reasonable to assume that the capability of a protein to be secreted into OMVs is proportional to its affinity for the membrane, as the protein would spend more time in proximity of the bilayer, and therefore would have statistically more probabilities to be enveloped when a vesicle is formed.

Running MD simulations of OMVs containing multiple copies of MBLs would provide precious information in this sense: however, these systems are characterized, *in vivo*, by a size that is not possible to be investigated through CG MD simulations, because of its prohibitive computational cost (about 70 nm of diameter, which would bring to systems with several

millions particles). As an alternative, we decided to run simulations with systems similar to the ones that we previously used for NDM-1: despite these systems will not provide any information regarding the influence of the OMVs curvature on the proteins, it is reasonable to assume that, due to the very large curvature of OMVs (~ nm), compared to the small size of the proteins (~ 4 nm), this effect would be minimal. In summary, our strategy is to run CG MD simulations of different MBLs in presence of a membrane patch, in order to evaluate their affinity, and to compare this ranking with the natural tendency of MBLs to be secreted. This would also bring the possibility to identify specific residues that drive membrane recognition in different MBLs, and perform mutagenesis experiments to validate the simulations outcomes. This protocol may serve as a computational predictor for MBLs capability to be secreted into OMVs.

We therefore collected CG MD simulations for 4 different MBLs, namely lipidated NDM-1 and VIM-2 (whose simulations were performed in the previous sections) and, additionally, SPM-1 and IMP-1. Despite the lack of experimental evidence for IMP-1, we chose to include it in the analysis because this MBL is reportedly very positively charged, and therefore more likely to interact with the negatively charged bacterial membrane. The results clearly indicate that IMP-1 has a very high tendency to bind to the bacterial membrane, possibly even higher than NDM-1. On the other side, SPM-1 presents a behaviour from this point of view that is intermediate between NDM-1 and VIM-2. This tendency can be inferred both in terms of binding events and in terms of simulation time spent in proximity of the membrane (Figure 3.18).

The SPM-1 enzyme interacts with the bilayer through a very limited portion of its surface (Figure 3.19a), which goes from Lys176 to Asn185: although two positive amino acids (Lys176 and Lys179, Figure 3.19b) are present that might drive the interaction, this surface is not large enough to establish a strong and stable contact with the membrane, which in fact results to be significantly smaller than NDM-1.

We then focused our analysis on IMP-1, because of its very high membrane affinity: like NDM-1, we could identify a specific protein patch that interacts with the membrane in all CG MD replicas, and goes from Gly67 to Ser76 and from Thr115 to Pro135 (Figure 3.20a).

a

	Replica 1	Replica 2	Replica 3	Replica 4	Replica 5
NDM-1	O	O	O	O	O
VIM-2	X	X	X	X	X
SPM-1	O	X	O	X	O
IMP-1	O	O	O	O	O

b
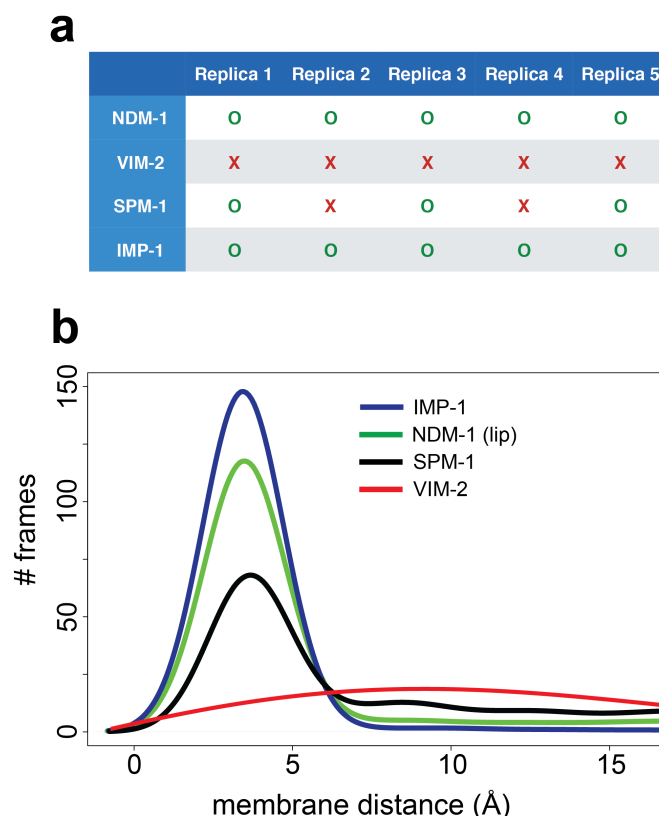
Figure 3.18 – (a) Comparison of number of binding events (marked with a green circle) observed for NDM-1, VIM-2, SPM-1 and IMP-1. (b) Distribution of CG MD trajectory frames, collected every 750 ps, with respect to the protein/membrane distance of lipidated NDM-1, VIM-2, SPM-1 and IMP-1.

Through molecular superimposition, we noticed that this patch does not coincide with the interacting surface that we identified for NDM-1. However, a characteristic in common with NDM-1 is the presence of positively charged residues: in particular, we identified 4 lysines (Lys69, Lys71, Lys127 and Lys129) which the simulations show to mediate the interaction with the bilayer, similarly to Arg45 and Arg52 in NDM-1. Driven by these results, we also performed CG MD simulations with a quadruple-mutant form of IMP-1, replacing these 4 lysines with 4 alanines: in this case, we observed a radical decrease both in the number of binding events (3 out of 5 CG MD replicas did not present a binding event) and in terms of general affinity to the membrane (Figure 3.20b/c), hence confirming that these residues are likely to play a central role in IMP-1/membrane interaction.
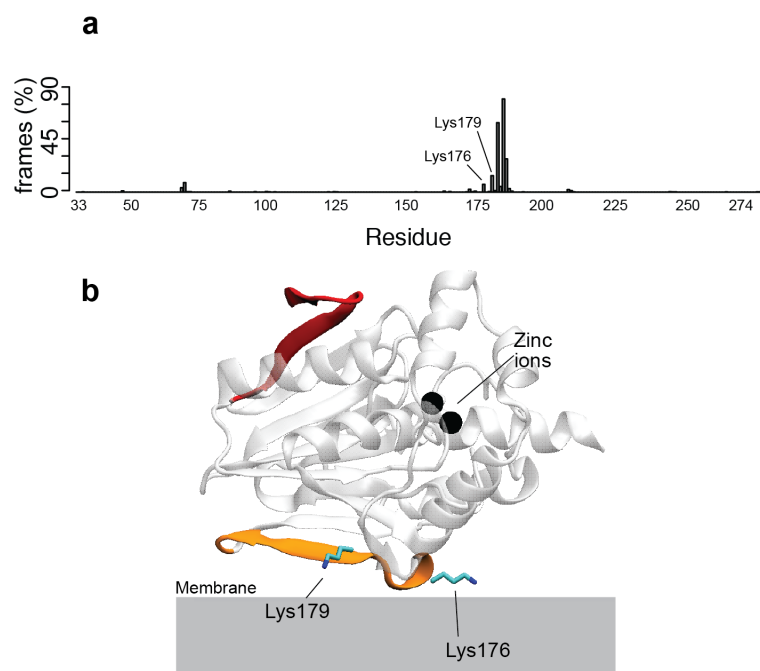
Figure 3.19 – (a) Contacts with the outer bacterial membrane, expressed in percentages of total number of frames (collected every 750 ps) for each SPM-1 residue across the CG MD simulations. (b) Identified surface of interaction between SPM-1 and the bacterial membrane. The surface of membrane interaction of SPM-1 is represented in orange, while the corresponding surface of interaction in NDM-1 is depicted in red. The two do not overlap.
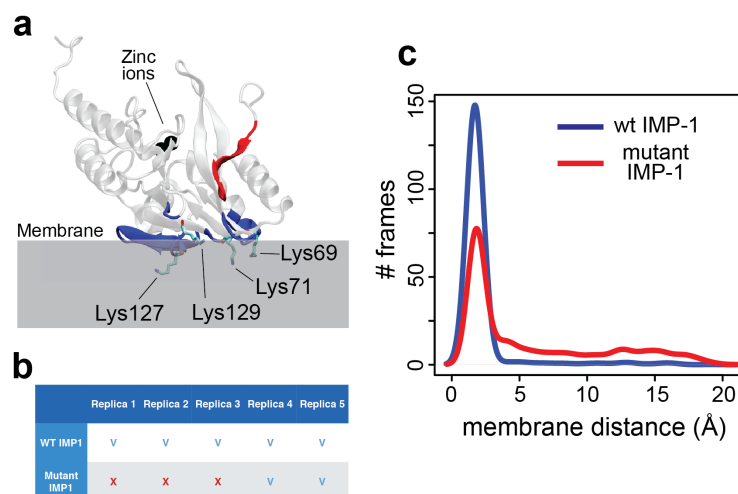
Figure 3.20 – (a) Identified surface of interaction between IMP-1 and the bacterial membrane. The surface of membrane interaction of IMP-1 is represented in blue, while the corresponding surface of interaction in NDM-1 is depicted in red. The two do not overlap. (b) Number of binding events observed for wt IMP-1 and quadruple-mutant IMP-1. (c) Distribution of CG MD trajectory frames, collected every 750 ps, with respect to the protein/membrane distance of IMP-1 and quadruple mutant IMP-1.

3.3 Conclusions

NDM-1 has been one of the major molecular causes of carbapenem resistance in several outbreaks of opportunistic and pathogenic bacteria.[233] The hydrolytic abilities of this enzyme are similar to those of other clinically relevant MBLs,[214] and cannot explain its success as a resistance determinant. The unique worldwide dissemination of the zinc-dependent lactamase NDM-1 has been attributed to the fact that this enzyme is bound to the inner leaflet of the outer membrane of Gram-negative bacteria,[110] in contrast with the rest of MBLs that are soluble, periplasmic proteins.

Lipidation allows membrane anchoring of NDM-1, being a unique characteristic of this enzyme. In the present work, we have characterized the mechanistic details of the association of NDM-1 with the bacterial membrane, demonstrating that the enzyme maintains a stable orientation with respect to the bilayer, while exposing its active site to the solvent, therefore preserving its catalytic activity. A defined protein surface contacts the membrane and contributes to the strong electrostatic interaction that acts synergistically with lipidation at Cys26. These results are strongly supported by molecular simulations at different levels of resolution, as well as by fluorescence studies and liposome flotation assays in a model mimicking the outer bacterial membrane. This membrane-protein interaction is characteristic of NDM-1, since in the case of the soluble MBL VIM-2, no interactions with the membrane are predicted by the simulations nor observed in the interaction with liposomes. Instead, a soluble version of NDM-1 is still able to interact with the membrane despite lacking the lipid moiety.

These findings allow us to conclude that, while the lipid group attached to the N-terminal Cys26 of NDM-1 is of crucial importance in maintaining membrane association, the globular domain of the protein possesses a native affinity towards the bacterial membrane. This phenomenon has been reported for other peripheral membrane proteins.[229,234] The presence of the lipid group is essential in order to stabilize the protein/membrane interaction on a long-term timescale, resulting thus in robust anchoring. Lipidation by itself is not sufficient, since the chimeric protein N-VIM (a membrane-anchored version of VIM-2) has a poorer membrane association than NDM-1. Both the lipidation and the recruitment of

CDLs are essential for generating a strong and stable anchoring to the membrane. The anchoring mechanism of NDM-1 is therefore specific for the outer bacterial membrane, and is strongly compromised in presence of inner bacterial membrane or CDL-free bilayers. Thus, evolution has optimized the ability of NDM-1 to interact with the outer membrane milieu that is determined by its lipidation signal and the bacterial lipoprotein biogenesis machinery.

We have identified a positively charged patch in the surface of NDM-1 that is responsible for this additional mechanism of association. However, the presence of positively charged residues does not seem to be the only responsible for driving strong and stable association between the protein and the membrane. Indeed, the positive residues must pack in a proper way, in order to be able to attract specific clusters of lipids (such as CDLs) that stabilize the interaction with the protein. The induction of lipid clusters formation by the protein is crucial to guarantee a long-lasting interaction with the bilayer, as suggested by the CG-MD simulations of the N-VIM chimera, which is not able to attract CDL aggregates.

Finally, this contact surface contains two basic residues (i.e., Arg45 and Arg52) that are important in defining a stable interaction with the bilayer, which is disrupted by the R45E-R52E substitutions. Notably, these two Arg residues are conserved in all reported NDM variants (28 so far). The first 16 NDM variants have been shown to be bound to the membrane.[235] While the other 12 still lack a biochemical characterization, all variants have a conserved lipobox in the signal peptide, suggesting that all of them are effectively membrane proteins. Thus, the conservation of the two positively charged Arg residues confirms the essentiality of this patch for the membrane-protein interaction, and reveal that the globular domain of NDM variants has been selected by evolution with features that favor its cellular localization, and ultimately, secretion to vesicles. New therapeutic strategies targeting these positions could inhibit the association of NDM variants with the membrane surface, aiding in the fight against this resistance determinant by negating the advantages conferred by NDM unique cellular localization.

Even without considering the lipidation, NDM-1 is therefore particularly prone to interact with the membrane. It is reasonable to assume that this feature played an essential role in helping to grow the membrane-anchoring mechanism, which does not result to be

present in other MBLs. However, we also found that IMP-1 is even more affine to the bacterial membrane with respect to non-lipidated NDM-1, and is therefore a candidate to develop the same mechanism. In this case the membrane interaction seems to be mediated by 4 lysines, which are conserved in 27 out of 31 reported IMP variants. Lipidation on bacterial proteins occurs at cysteines in the N-terminal region, provided that a specific signal sequence is present. The signal sequence is required because it will be recognized by a lipoprotein signal peptidase, which will initiate the cascade that will end up in the lipidation of the cysteine.[236] This signal sequence has specific characteristics and can be divided in 3 regions: the n-region, characterized by the presence of positive amino acids (lysines or arginines); the h-region, hydrophobic; and the c-region (also called cleavage region or lipobox), which is located before the cysteine and possesses a very conserved sequence, specifically L-A-(G/A)-C.[237,238] IMP-1 possesses several of the features that lead to lipidation, i.e. a n-region (Lys3 in the sequence), a h-region (rich in phenylalanines) and a cysteine at the N-terminal tail (Cys13). However, IMP-1 does not possess the lipobox, which causes it to be a periplasmic protein. However, if random genetic mutations caused IMP-1 to get a lipobox, then our study predicts that IMP-1 is likely to develop the same mechanism that caused NDM-1 to become a serious threaten to global health. On the contrary, VIM-2 and SPM-1 do not present a particular high affinity for the membrane according to our simulations: this is reflected on their reduced capability to be secreted into OMVs with respect to NDM-1.

3.4 Materials and Methods

Cloning and generation of MBL mutants. The full length NDM-1 gene was amplified by PCR from the pMBLe NDM-1 plasmid[110] using the NDM-1 NdeI Fw (5' TATACATATGGAATTGC-CCAATATTATGCACC 3') and NDM-1-TEV-TwinST-HindIII Rv1 (5' GAACCACCACCCTTTTC-GAATTGTGGGTGAGACCAGCCCTGAAAATACAGGTTTTCGCGCAGCTTGTCGGCCATGC 3') primers. This product was then used as a template for another round of PCR, using the NDM-1 NdeI Fw and NDM1-TwinST-HindIII Rv2 (5' GACGTAAGCTTCTACTTTTCGAATTGTGGGTGAGAC-CACGCAGAACCACCAGAACCACCACCAGAACCACCACCCTTTTCG 3') primers. The product was cloned into the NdeI and HindIII sites of the pET-26 vector, obtaining the pET-26 NDM-1 TEV TST plasmid. From this plasmid, the pET-26 NDM-1 C26A TEV TST plasmid was obtained

through site directed mutagenesis by plasmid amplification as previously described 40, using primers C26A Fw (5' CATTGATGCTGAGCGGGGCGATGCCCGGTGAAATC 3') and C26A Rv (5' GATTTCACCGGGGCATCGCCCCGCTCAGCATCAATG 3').

Liposome preparation. Pure lyophilized phospholipids (1-palmitoyl-2-oleoyl-phosphatidyl ethanolamine, tetraoleoyl cardiolipin, and 1-palmitoyl-2-oleoyl-phosphatidylglycerol) and *E. coli* Polar Lipid Extract were purchased from Avanti Polar Lipids. The lipids were dissolved in chloroform and after mixing the required proportions of each pure lipid, the lipid mixtures were dried under a nitrogen atmosphere and then kept under vacuum for 2h. The dried lipid film was hydrated with 50 mM HEPES pH 7, and heated at 65ºC for 1h with periodic vortexing. Lipid suspensions were frozen in liquid nitrogen and then thawed at 65ºC, for a total of 5 cycles, and afterwards were passed through a 400 nm polycarbonate filter using an Avanti Miniextruder apparatus (Avanti Polar Lipids) at 65ºC, with >20 passes through the device.

Purification of lipidated NDM-1 and soluble NDM-1 C26A. Full length NDM-1, including its signal peptide and lipidation signal, was overexpressed in E. coli BL21 (DE3) cells using the pET-26 NDM-1 TEV TST plasmid. The NDM-1 protein is produced from this vector with a fusion to its C-terminus of a TEV protease cleavage site followed by the Twin-StrepTag peptide,[239] which allows affinity purification by binding to a StrepTactin Sepharose resin (GE Healthcare). Cells were grown in LB medium at 37ºC with agitation to OD600nm = 0.8, and protein expression was induced by addition of 0.5 mM IPTG. Cultures were then grown at 20ºC for 16 h. The cells were collected by centrifugation, resuspended in 50 mM HEPES pH 7.5, 200 mM NaCl, and ruptured at 15000 psi using an Avestin Emulsiflex C3 high pressure homogenizer. Cell debris was spun down by centrifugation at 14000g and 4ºC for 20 min, and membranes were isolated by ultracentrifugation for 1h at 4ºC and 125000 g in a Beckman SW Ti90 rotor. The lipid-anchored protein produced in this way was extracted from membranes by solubilization with 1% w/v of the non-ionic detergent Triton X-100 in 50 mM HEPES pH 7.5 200 mM NaCl, followed by ultracentrifugation using the same conditions as before to remove non-solubilized material. Protein was purified by affinity chromatography with StrepTactin Sepharose resin, eluted with 2.5 mM desthiobiotin in 10 mM HEPES pH 7.5 200 mM NaCl 0.033% w/v Triton X-100, and the affinity tag was removed by cleavage with TEV protease. The protein was dialyzed for 16h versus >100 volumes of 10 mM HEPES pH 7.5 200 mM NaCl

0.033% w/v Triton X-100, 0.6 mM βME, and then for 4h versus >100 volumes of 10 mM HEPES pH 7.5 200 mM NaCl 0.033% w/v Triton X-100, ZnSO4 0.1 mM. Finally, the protein was passed a second time through the StrepTactin Sepharose resin to remove the residual non-processed protein.

Proteoliposomes containing NDM-1 were then prepared by incubation of the purified protein with preformed liposomes and 0.7% w/v Triton X-100, and detergent removal was carried out with SM2 Bio-beads (Biorad). Proteoliposomes were separated from non-incorporated protein by ultracentrifugation on a discontinuous sucrose gradient (8% w/v, 25% w/v and 55% w/v), with protein incorporation into proteoliposomes being around 25% of the total NDM-1 added.

NDM-1 C26A was purified from *E. coli* OverExpress C43 (DE3) cells transformed with the pET-26 NDM-1 C26A TEV TST plasmid. Cells were grown in LB medium at 37ºC with agitation to OD600nm = 0.8, and protein expression was induced by addition of 0.5 mM IPTG. Cultures were then grown at 20ºC for 16 h. The cells were collected by centrifugation and resuspended in 50 mM HEPES pH 7.5, 200 mM NaCl. The cells were then ruptured at 15000 psi using an Avestin Emulsiflex C3 high pressure homogenizer, and cell debris was spun down by centrifugation at 14000g and 4ºC for 20 min. Protein was purified by affinity chromatography with StrepTactin Sepharose resin, eluted with 2.5 mM desthiobiotin in 10 mM HEPES pH 7.5 200 mM NaCl, and the affinity tag was removed by cleavage with TEV protease. The protein was dialyzed for 16h versus >100 volumes of 10 mM HEPES pH 7.5 200 mM NaCl 0.033% w/v Triton X-100, 0.6 mM ME, and then for 4h versus >100 volumes of 10 mM HEPES pH 7.5 200 mM NaCl 0.033% w/v Triton X-100, ZnSO4 0.1 mM. Finally, the protein was passed a second time through the StrepTactin Sepharose resin to remove the residual non-processed protein.

Fluorescence anisotropy determinations. Determinations of tryptophan fluorescence anisotropy were carried out in a Varian Cary Eclipse Spectrofluorometer using a 10x2 mm optical path quartz cuvette. Sample excitation was performed at 298 nm, to take advantage of the higher fundamental anisotropy of tryptophan at this wavelength, and fluorescence emission was collected at 370 nm. The excitation and emission slit widths were set to 5 nm and 10 nm, respectively, and a 360-1100 nm band pass filter was used in the emission optical path

to remove scattered light. An integration time of 5 s was used for each measured fluorescence signal, and 5 replicate anisotropy determinations were acquired for each sample and averaged. To evaluate whether there is an appreciable contribution from light scattering to the anisotropy values previously determined, we added an equivalent amount of liposomes (not containing any protein) to samples of NDM-1 C26A. The observed anisotropies were similar to those of the soluble protein in absence of liposomes, indicating no interference from scattered light in our analysis.

Fluorescence quenching. Tryptophan fluorescence quenching determinations were carried out in a Varian Cary Eclipse Spectrofluorometer using a 5x5 mm optical path quartz cuvette. Sample excitation was performed at 280 nm, and fluorescence emission was collected from 300 to 450 nm, at 60 nm/s scan speed. Emission and excitation slit widths were set to 5 nm each, and photomultiplier voltage to 600V or 700V depending on sample emission intensity. Protein concentrations in the cuvette were typically 3 μM.

Liposome flotation assay. Samples containing 55 μM protein in 50 mM HEPES ph 7 were incubated with liposomes for 30 min at room temperature. Sucrose was added to 40% w/v, the samples were loaded in an ultracentrifuge tube, and sucrose 25% w/v and sucrose 8% w/v (both buffered with 50mM HEPES pH 7) were layered on top, forming a discontinuous sucrose gradient. Afterwards, samples were ultracentrifuged for 1h at 4ºC and 125000 g in a Beckman SW Ti90 rotor, and fractions along the gradient were analyzed by SDS-PAGE to assess the final distribution of the MBL protein.

Determination of protein orientation in liposomes by proteolysis with proteinase K. Proteoliposome samples were incubated overnight at 45°C after addition of 30 μg/mL proteinase K and 1 mM CaCl2. The proteolysis reaction was stopped by addition of 5 mM PMSF, and samples were analyzed by SDS-PAGE.

Coarse-grained molecular simulations. In CG MD simulations, the enzyme was located at 4 nm at least from the bilayer, in order to avoid an early protein/membrane sensing event that might bias the subsequent interaction. The MARTINI2.2p (polarizable) force field[80] was used for all the CG MD simulations. The crystal structure of NDM-1, VIM-2, SPM-1 and IMP-1 were taken from the Protein DataBank (PDB codes: 5ZGE[240] for NDM-1; 1KO3[241]

for VIM-2; 5NDE[242] for SPM-1; 5EV6[243] for IMP-1) and turned into a coarse-grained model with the Martinize tool provided by the MARTINI team. All the lipid bilayers were generated with the Insane tool of MARTINI,[82] with a lipid composition of 91% PEs, 6% CDLs and 3% PGs for the outer membrane, and 67% PEs, 5% CDLs, 28% PGs for the inner one, according to lipidomics analyses present in literature.[227,244] For what concerns the acyl chains, since no reliable data are present in literature, we used the motifs that appear most frequently in nature, that are: one 16:0 and one 18:1 (1-palmitoyl-2-oleoyl) for PEs and PGs, and four 18:1 for CDLs. All the systems were solvated with the polarizable MARTINI water model[178] and ionized in 150 mM of NaCl. Each system (NDM-1 in apo/holo form; with/without post-translational modification; in presence of inner/outer bacterial membrane; with/without cardiolipins. VIM-2 in holo form; in presence of inner/outer bacterial membrane. N-VIM in apo/holo form; with/without post-translational modification; in presence of outer bacterial membrane. NDM-1 R45E-R52E in presence of outer bacterial membrane) was repeated in 5 distinct replicas. Each replica was simulated for 2 μs; the systems for which we evaluated the lipid distribution were elongated to 10 μs, as reported in Figure 3.5. These trajectories, taken together, sum up to a total simulation time of 550 μs. Frames for the analysis were collected every 750 ps. A binding event (Figure 3.5) is considered to have occurred when the protein settles at 3 Å distance from the membrane, and this distance remains constant for the rest of the simulation (i.e. no detachment occurs). For each system, the equilibration procedure was run as follows: first, the system went through 5000 steps of minimization using the steepest descent algorithm; successively, it was equilibrated with 5 ns of MD in NVT conditions, using Particle Mesh Ewald (PME) for the electrostatic contributions and velocity rescale algorithm for temperature coupling at 310 K. The production phase was conducted in NPT ensemble, using a Parrinello-Rahman semi-isotropic coupling algorithm[245] for maintaining the pressure constant at 1 bar. The post-translationally modified Cys26 was built using the parameters for a similar lipid present in literature.[246] The linker between the cysteine and the lipid was modelled, following the guidelines provided by the MARTINI developers. The two zinc ions in the catalytic site were represented as one single MARTINI bead of Qa type, connected to the 6 coordinating residues (His120, His122, Asp124, His189, Cys208, His250) through harmonic potentials. Such particle was charged with +2e, that is the net charge of the catalytic site (2 zinc ions charged +2e each; 1 hydroxyl ion charged -1e; the deprotonated Cys225 charged -1e).

The decision to model the two zinc ions with one single particle was taken because all the available MARTINI beads are characterized by a van der Waals radius that is too large to fit two particles within the limited free volume that is present in the catalytic site. The electrostatic potential generated by our model was compared to the atomistic representation and, within the limits of the CG modelling, the two are shown to have a good agreement (Figure 3.21).
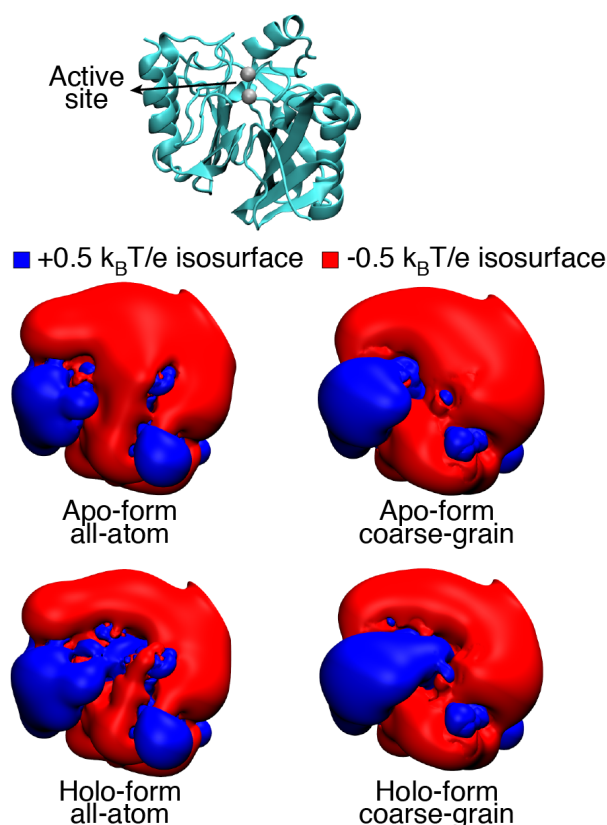
Figure 3.21 – Electrostatic potential isosurface for the apo/holo forms of NDM-1 in the AA and CG model.

All-atom molecular simulations. Out of all the CG MD trajectories, 3 replicas for each condition of NDM-1 (apo/holo form, with/without post-translational modification) were selected to be back-mapped into atomistic systems. In particular, the lipids were back-mapped using the Backward MARTINI tool.[247] The same tool is not efficient enough to reproduce reliable atomistic representations of proteins: therefore, in order to add the enzyme to the atomistic system, we took the crystal structure available at the Protein DataBank and, for each replica, we aligned it to the CG structure of the same protein in the last frame of the CG-MD simulation. With this procedure, it was possible to preserve the lipid distribution obtained

during the CG simulation, without compromising the quality of the atomistic structure of the proteins. We used Amber ff99SB-ILDN[248] to parameterize the protein, while the parameters for the lipids were taken from the Lipid14[78] repository of Amber. Lipid14 allows to generate lipid models in a modular way, that is by combining the polar heads with appropriate acyl chains. The CDL polar head is still unavailable in Lipid14, therefore we used an Amber parameterization obtained by Lemmin et al.[249] The catalytic site was modelled according to a parameterization from Merz et al.[250, 251] In this model, the charge distribution was updated for all the atoms of the six residues that coordinate the two zinc ions, to model the charge transfer due to the presence of the metal ions.

The parameters for the post-translationally modified Cys26 were generated ex-novo following the Amber guidelines. All the AA systems were solvated with a TIP3P solvent model and ionized in 150 mM NaCl. Since the atomistic systems were coming from a back-mapped CG system, the equilibration procedure had to be particularly accurate, in order to avoid clashes: initially, we followed the procedure suggested by the MARTINI team, which consists in performing a first minimization step after inactivating all the non-bond interactions (Coulomb and Lennard-Jones); non-bond interactions are then gradually restored in successive MD equilibration stages. After having reached a reasonable equilibrium, we performed another stage of minimization (5000 steps of steepest descent algorithm), followed by 5 different stages of MD equilibration in NVT ensemble (for a total of 500 ps), with harmonic constraints applied both on the protein backbone and on the lipids phosphates, which were gradually released. The production phase was run in NPT ensemble, using the Verlet algorithm[252] for the neighbor search, velocity rescale temperature coupling algorithm[253] for maintaining the temperature constant at about 310 K and Parrinello-Rahman pressure coupling algorithm[245] in semi-isotropic conditions at a pressure of 1 bar. The electrostatic interactions were treated with Particle Mesh Ewald method.[254] Each AA MD replica was simulated for 800 ns, for a total of 9.6 μs (12 replicas). Frames for analysis were collected every 100 ps.

Both in the CG and in the AA simulations, the lipid occupancy was evaluated through the VolMap Tool plugin of VMD, in particular by calculating the average occupancy of each lipid phosphate throughout the trajectory. The electrostatic potentials calculations were performed through the APBS software,[255] which solves the Poisson-Boltzmann equations to compute the

electrostatic potentials of biomolecules. The dielectric constant was set at 78 for the solvent, and 2 for the protein; the ionic concentration was set at 150 mM NaCl; protonation states of the protein aminoacids were assigned with PropKa.[256]

4 Golph3/Golgi interaction mechanism

The Golgi phosphorylated protein 3 (Golph3) is a peripheral membrane protein of the Golgi apparatus that regulates the glycosphingolipid synthetic pathway. Golph3 has been found to be overexpressed in several forms of malignant tumors. Although its gene is regarded among the most oncogenic ones, the role of the corresponding protein is poorly understood from a molecular point of view. In particular, it is not clear how the peripheral interaction with the Golgi influences the work of the protein, and how the recognition between Golph3 and the glycoenzymes occurs.

Here, we study the interplay between Golph3 and the Golgi, by means of molecular simulations. We show the characteristics of the phospholipid bilayer that drive the interaction with the protein and identify local properties of Golph3 that drive its nature as a peripheral membrane protein, and that make it particularly suitable to interact specifically with the Golgi, rather than with other bilayers. We show the difficulties that arise in the experimental evaluation of peripheral membrane proteins by reporting results from nuclear magnetic resonance and tryptophan fluorescence assays experiments. Moreover, we propose a molecular mechanism of recognition between Golph3 and the glycoenzymes, and suggest a strategy for experimental verification.

4.1 Introduction

The Golgi apparatus (often referred to simply as "the Golgi") is an organelle present in most eukaryotic cells.[257–259] It was one of the first organelles to be discovered, as it was observed at the microscope by Dr. Camillo Golgi in 1898.[260] It is localised at the centre of the cell, and is separated from the nucleus by the presence of the endoplasmic reticulum (ER), with which it has continuous interactions (Figure 4.1).[261–265]The Golgi is constituted by a series of phospholipid bilayers that form closed compartments, known as *cisternae.*[266,267] The Golgi performs several key activities within the cell life cycle, some of which are still object of debate.[268,269] Its main role consists in accommodating the proteins that are synthesized by the ribosomes present in the ER, and to distribute them within the cellular space.[270,271] In details, after having been synthesized by the ribosomes, the proteins are encapsulated into vesicles: thanks to a fusion mechanism, these vesicles are incorporated within the Golgi, and later excreted outside of the organelle with the mediation of other vesicles. The protein-loaded vesicles that are excreted from the Golgi can either be allocated within the cytoplasm or excreted by the cell membrane into the extra-cellular environment. Because of its role as a protein dispatcher, the Golgi is sometimes referred to as the "post office of the cell".
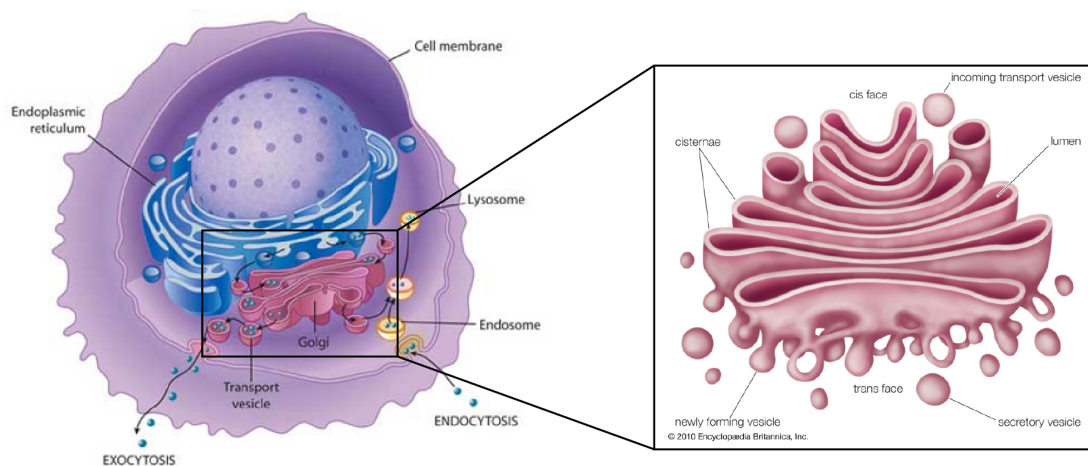
Figure 4.1 – The Golgi apparatus: location within the cell (left, figure by *Encyclopedia Britannica*) and detail (right, figure by *Encyclopedia Britannica*).

The Golgi phosphorylated protein 3 (Golph3) is a peripheral membrane protein that is localized at the Golgi apparatus.[111] Golph3 is involved in the regulation of transport of

glycosylating enzymes (glycoenzymes) through the Golgi. While the normal flow of proteins within the Golgi proceeds from the ER towards the outer *cisternae*, the glycoenzymes proceed backward, in order to be recycled and utilized multiple times by the Golgi. This mechanism allows to save a considerable amount of energy, and is therefore fundamental for the cell economy.[112] Golph3 is however better known for its aberrant behaviour, which is linked to the formation of several malignant forms of tumor. In particular, Golph3 has been found to be overexpressed in a significant number of lung, breast, ovarian, pancreatic and colon cancers.[89] The oncogenic action of Golph3 has been recently linked to glycoenzymes that regulate the glycosphingolipid (GSL) synthetic pathway.[269, 272, 273] In particular, the overexpression of Golph3 reduces the degradation of these glycoenzymes: as a consequence, the GSL production is increased, an event which drives uncontrolled cell proliferation.

Several glycoenzymes have been shown to be actively, retrogradely transported in the Golgi by Golph3: in particular, the glycoenzymes of the GSL synthetic pathway, which are of greater medical relevance, were found to interact with Golph3 through their cytosolic tail.[274] Previous results from collaborators showed that most of the cytosolic tails that interact with Golph3 share a common amino acid sequence (unpublished data), that is the form x-L-xx-R/K, i.e. where the second amino acid is a leucine, and the fifth is positively charged (arginine or lysine). Ad hoc mutations performed on the lactosylceramide synthase, which is one of the enzymes that has been shown to interact with Golph3, confirmed that these residues are fundamental to drive the molecular recognition. In the present work, we will exploit the relationship between Golph3 and the cytosolic tail of glycoenzymes, with a particular focus on the lactosylceramide synthase, for consistency with the available experimental evidence. We will also evaluate the role played by the Golgi in this mechanism.

From the structural point of view, the primary sequence of Golph3 counts a total of 298 aminoacids,[111] for a molecular weight of about 34 kDa. The first 58 residues constitute the signal peptide, and need to be cleaved by a specific protease for the protein to be activated. The structure of Golph3 has been reported to be very similar to Vsp74, a protein from yeast to which Golph3 is an ortholog.[111] In particular, Golph3 is characterized by the presence of 4 core α-helices, which are mainly hydrophobic; 6 surface α-helices, which are amphipatic; and 4 β-strands. The phosphorylation sites of Golph3 are at Thr143 and Thr148. Golph3

possesses a peculiar β-hairpin domain, which is particularly hydrophobic, due to the presence of two leucines (Leu195 and Leu196) and two phenilalanines (Phe194 and Phe197) at its tip (Figure 4.2a). Because of its hydrophobic nature, the β-hairpin is suspected to play a role in the interaction between Golph3 and the Golgi. This β-hairpin is also highly conserved, which strengthens the hypothesis that it may be fundamental for the protein action. Golph3 also interacts with a non-conventional myosin: this interaction induces the flattening of the Golgi, and Golph3 may therefore be indirectly responsible for the peculiar flat shape of this organelle.[275] Moreover, Golph3 has been linked to cell proliferation, growth and survival, in relation to its interaction with mTOR, a member of serine/threonine protein kinases.[89]

Despite the role of Golph3 in cancer has been confirmed several times, and for nearly all the most frequent and lethal forms of tumor, its oncogenic mechanism still remains partially elusive: its role in the GSL synthetic pathway might indeed be only one of the factors that promote tumor formation. In addition to that, it has also been proposed that Golph3 may be linked to the promotion of mitosis,[276] or to the release of pro-tumoral factors that favour metastasis growth by the remodelling of extracellular matrix.[114] Moreover, it has been observed that the phosphorylation of Golph3 happens more frequently following an event of DNA damage: this seems to suggest that its overexpression may be linked to induction of DNA instability, as the answer of the cell to DNA damage may be significantly affected.[113]

One of the factors that will be exploited here is the role played in the Golph3 activity by a specific lipid type, i.e. phopsphatidylinositol-4-phosphate (PI4P). PI4P is an anionic lipid, present in a relevant amount (10%) in the Golgi (Figure 4.2b).[277–279] The Golph3 activity has been shown to be strongly dependent on the presence of PI4Ps,[280] although the molecular mechanisms that define their interaction are still partially unclear. In particular, the presence of PI4 kinases have been identified as regulators of the whole Golgi function. Removal of this enzyme implies a relocation of Golph3 (and of Vsp74, its ortholog in yeast). An obvious deduction of this event would therefore be that Golph3 is likely to interact with the Golgi via PI4P lipids. Indeed, the crystal structure revealed the presence of a putative binding site for PI4Ps, which involve residues Trp81, Arg90, Arg171 and Arg174.[111] Successive mutagenesis experiments showed that the elimination of these residues do not cause alterations in the structure of the protein, but affect significantly its function *in vivo*. This evidence confirms

the theory according to which PI4Ps binding is crucial for Golph3 activity.
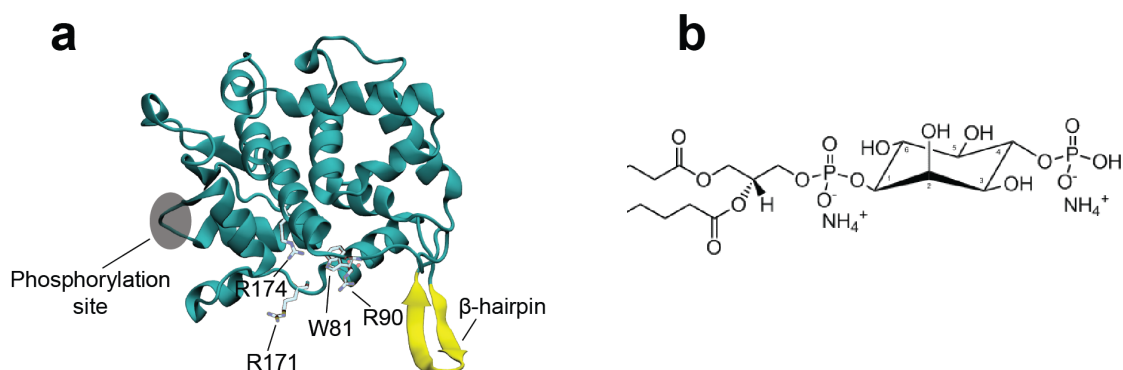
Figure 4.2 – (a) Structure of the Golgi phosphorylated protein 3. The β-hairpin that supposedly interacts with the Golgi is marked in yellow; residues that were reported in literature as fundamental for Golph3 action are in licorice. (b) 2D chemical structure of phopsphatidylinositol-4-phosphate.

A clear molecular characterization of the interaction between Golph3 and the Golgi is still missing up to date. The present work has a double aim: 1) apply molecular modeling techniques to describe in details the molecular events that drive the interaction of Golph3 with the Golgi, and therefore its functioning; 2) characterize the recognition mechanism between Golph3 and the glycoenzymes.

This knowledge will allow to elucidate the molecular mechanism of functioning of Golph3, in particular in its interaction with the Golgi and the glycoenzymes, which is crucial to drive a new generation of drugs that specifically target *ad hoc* sites of the protein, hence inhibiting the oncogenic activity of Golph3.

4.2 Results and Discussion

4.2.1 PI4Ps drive the association between Golph3 and the Golgi apparatus

We performed Coarse-grained (CG) molecular dynamics (MD) simulations to evaluate the association mechanism between Golph3 and the Golgi. The lipid composition of the bilayer was based on lipidomics data of the Golgi apparatus present in literature.[277–279] These lipidomics analyses report an accurate description of the lipid composition of the Golgi, as they include not only the polar heads present in the bilayer, but also the acyl chains composi-

tion. This is particularly relevant, as the data concerning the acyl chains composition are not often found in lipidomics analyses, although they can significantly affect the behavior of the membrane.[281, 282] The presence of this information guarantees that the model is particularly realistic and representative of the physico-chemical features of the Golgi. In particular, the composition that we used was: 30% POPCs, 20% DOPCs, 12% POPEs, 7% DOPEs, 5% POPSs, 10% PI4Ps, 16% cholesterol. The protein was placed at an initial distance of 60 Å from the membrane, in order to prevent initial sensing between Golph3 and the Golgi, that would bias the association in favour of the initial orientation of the protein. With respect to NDM-1 (see chapter 3), Golph3 is more positively charged (in particular, it is neutral, while NDM-1 is charged -5e), while at the same time the Golgi has the same electrical charge as the bacterial membrane (15% content of charged lipids, if we consider that each CDL in the bacterial membrane has two phosphate groups). Nevertheless, as we assessed in the case of NDM-1 (see chapter 3), the simple presence of favourable electrostatics does not guarantee a quick and stable association between the protein and the membrane, as specific lipid/amino acid interactions are required (Figure 3.7, 3.13, 3.16).

The simulations showed a very quick association between the Golph3 and the Golgi, if compared to NDM-1. In particular, we assessed that binding events occur on average after 136 ±56 ns for Golph3/Golgi, compared to 675 ±146 ns for NDM-1/bacterial membrane. We also observed a clear preference of the protein to interact with the Golgi through a specific portion of its surface (Figure 4.3). Like NDM-1, this surface of interaction does not coincide with the most positively charged side of the protein. Indeed, the electrostatic potential analysis shows that the protein surface is characterized by two very electrically charged areas, one of which is very positive (Figure 4.3b, mainly due to the presence of residues Arg100, Arg102, Lys123, Lys184, Arg212, Arg234 and Arg235), while the other one, which is placed on the opposite side of Golph3, is characterized by a high negative potential (Figure 4.3c, due to the presence of residues Asp83, Glu154, Glu159, Asp247 and Glu250). The Golph3 surface which interacts with the Golgi lays in between these 2 main surfaces, and interestingly does not preclude the interaction between the Golgi and Trp81, Arg90, Arg171 and Arg174, which were reported as fundamental residues for the protein activity. Moreover, it is noteworthy that the identified surface of interaction also includes the β-hairpin domain. Because of its

hydrophobic nature (it contains two leucines and two phenylalanines at its tip), this domain is one of the main suspect for promoting the interaction with the membrane. In particular, it would be reasonable that the hydrophobic tip contributes to the Golph3/Golgi binding through hydrophobic interactions with the inner core of the lipid bilayer. Although the CG MD simulations did not highlight a behavior of this type, we need to take into account that the MARTINI force field implies the usage of elastic networks to constrain the secondary and tertiary structures of proteins: therefore, we cannot exclude that conformational modifications may occur, that bring at a more relevant role of the β-hairpin in the interaction with the bilayer. The AA MD simulations can therefore provide better insights in this evaluation, as will be discussed in details in the next sections.
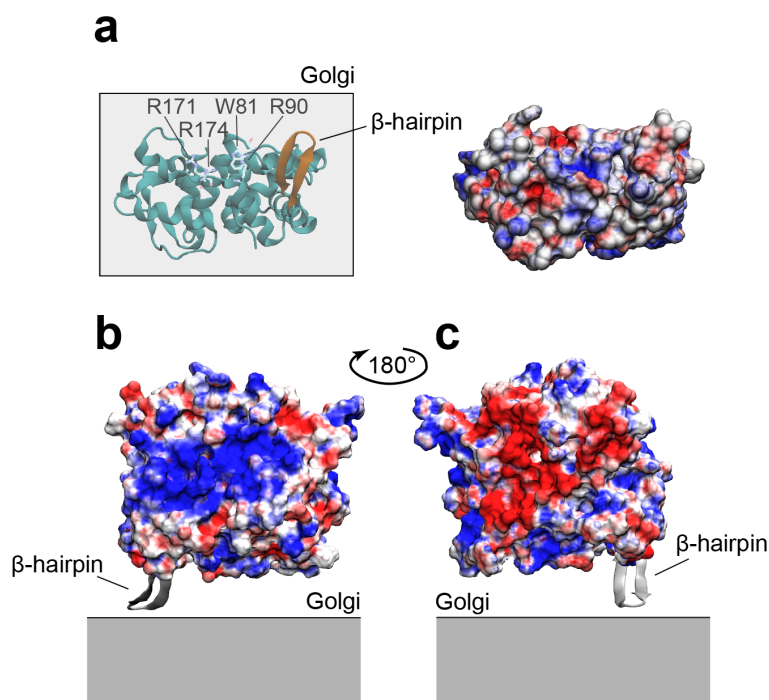
Figure 4.3 – (a) Side of Golph3 that interacts with the Golgi (left) and its electrostatic surface (right). (b) Positively charged surface of Golph3. (c) Negatively charged surface of Golph3.

We analyzed the number of contacts between the protein and each lipid type (Figure 4.4): we observed that, proportionally to the lipid composition of the Golgi, the PI4Ps show a significant majority of contacts with Golph3, with respect to any other lipid. This cannot be entirely explained with the anionic nature of these lipids, because the same does not happen for the other negatively charged lipids present in the Golgi, i.e. PSs. Moreover, the surface that

of Golph3 which interacts with the Golgi is not entirely positively charged, but is amphipatic (Figure 4.3a). In other words, we observe in Golph3 a mechanism of membrane interaction that is similar to the one between NDM-1 and the outer bacterial membrane. Moreover, these observations suggest that the presence of PI4Ps is particularly important for driving the interaction between Golph3 and the Golgi (similar as the role played by CDLs for NDM-1).
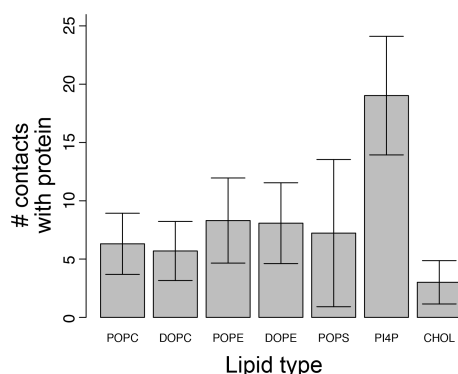
Figure 4.4 – Golph3 interactions with specific lipid types, normalized according to lipid composition of the Golgi.

Driven by these results, and knowing the importance of PI4Ps for the Golph3 activity, we decided to evaluate more accurately the role played by this lipid in Golph3/Golgi association. Specifically, we set and simulated CG systems with 0% and 20% presence of PI4Ps in the Golgi, respectively (the physiological level corresponds to 10%). We found that in absence of PI4Ps, Golph3 can still associate to the Golgi (binding events were observed in 4 replicas out of 5), although the time that is required for the membrane association to occur is higher than in presence of the physiological (10% PI4Ps) Golgi (Figure 4.5a/b). The analysis of the lipid contacts (Figure 4.5c) doesn't report any lipid type that significantly interacts with Golph3, suggesting that PI4Ps are not "replaced" by any other lipid type in their preferential interaction with Golph3, even by other anionic ones (PSs). The interaction between Golph3 and the 0% PI4Ps Golgi is therefore non-specific, and this may explain its slower association dynamics. For what concerns the 20% PI4Ps Golgi, we observed a behavior of Golph3 that is similar to the one with physiological lipid composition (10% PI4Ps), both in terms of time required for protein/membrane association, and in terms of protein surface which mediates the interaction. In particular, we observed that the Golph3 surface which interacts with the
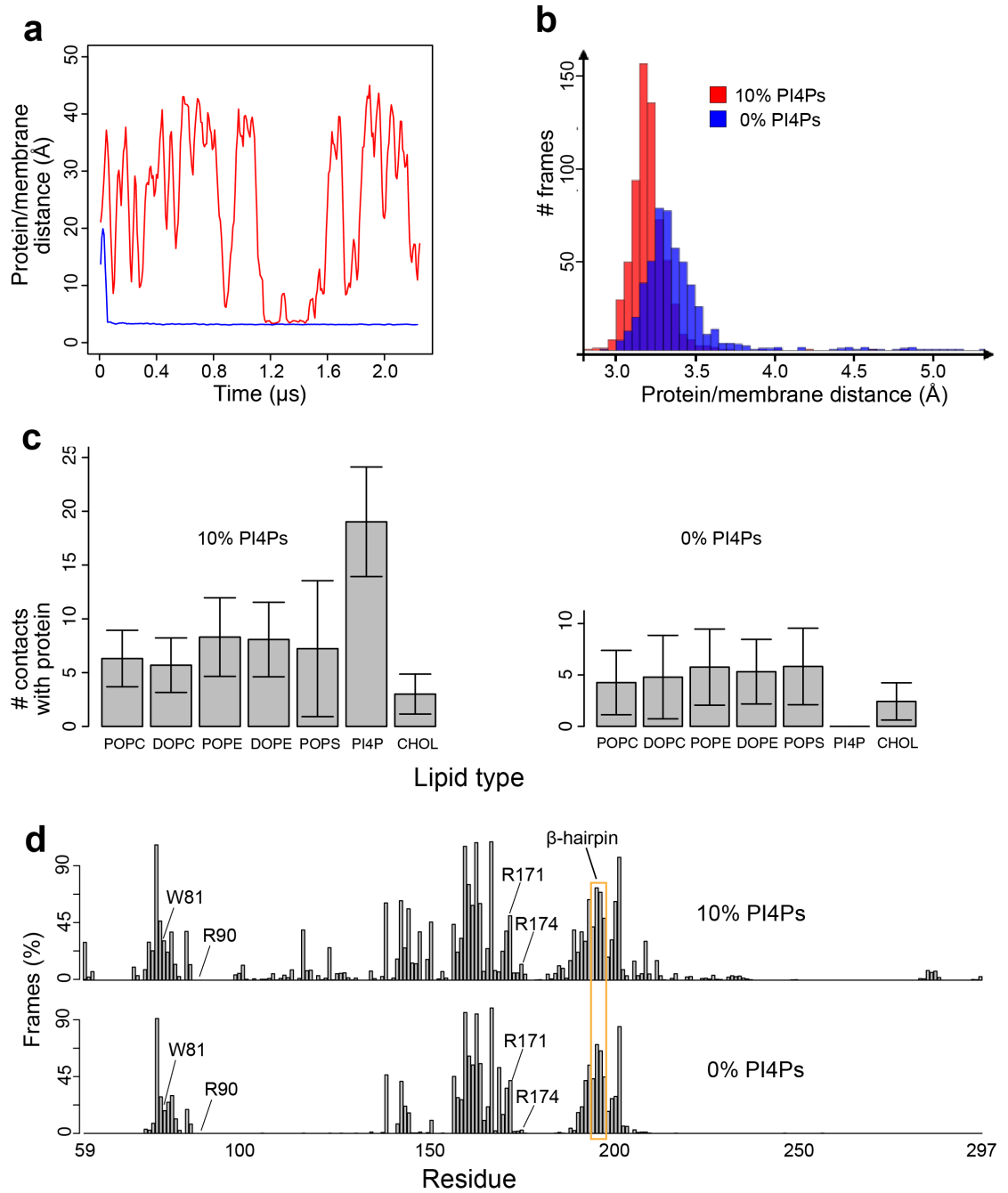
Figure 4.5 – (a) Golph3/Golgi distance evolution for 10% PI4Ps (blue) and 0% PI4Ps (red). (b) Distribution of MD trajectory frames with respect to Golph3/Golgi distances. (c) Lipid types involved in the interaction with Golph3. (d) Golph3 residues involved in the interaction with the Golgi.

Golgi is the same that was identified in the 10% PI4Ps in 3 replicas. However, in 2 replicas, we observed a different surface of interaction (Figure 4.6a). This second interaction surface that we identified corresponds to the very positively charged surface that we described earlier. This observation agrees with the fact that the Golgi model with 20% of PI4P is highly negatively charged. These results suggest that the Golph3/Golgi interplay is sensible to the change in the amount of PI4Ps that constitute the Golgi. Through the analysis of the lipid contacts, we again observed a preponderant role played by the PI4Ps: however, in this case the amount of PI4Ps contacts does not follow proportionally the increased number of PI4Ps in the membrane (Figure 4.6b). These results suggest that a saturation mechanism intervenes: increasing the amount of PI4Ps in the membrane, does not imply a larger number of Golph3/PI4Ps contacts. This is likely due to the fact that the size of the surface of interaction of the protein cannot host more PI4Ps. This also hints to the fact that Golph3 is optimized to specifically interact with the Golgi, which is characterized by a defined and peculiar amount of PI4Ps.
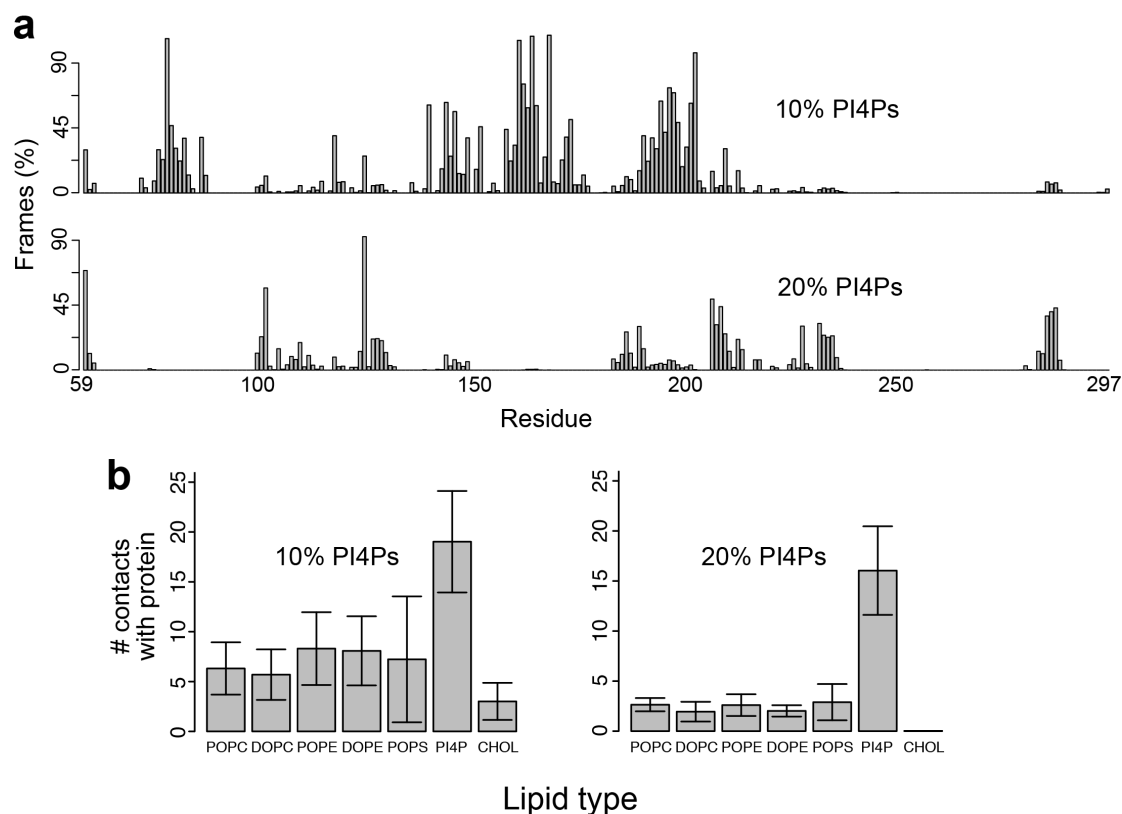
Figure 4.6 – (a) Contacts with the Golgi of each Golph3 residue, for 10% PI4Ps (top) and 20% PI4Ps (bottom). (b) Golph3 interactions with specific lipid types, normalized according to lipid composition of the Golgi, for 10% PI4Ps (c) and for 20% PI4Ps.

4.2.2 Evaluation of the role played by specific residues

As previously stated, some amino acids were reported in literature to have large influence on the activity of Golph3,[111,280] in particular Trp81, Arg90, Arg171 and Arg174. In order to understand whether the relevance of these amino acids could be attributed to their interplay with the Golgi, we ran CG MD simulations of single-point mutated Golph3 in presence of the Golgi. Specifically, we modelled four different single point mutants of Golph3, in which each of the aforementioned residues was replaced by an alanine (W81A, R90A, R171A, R174A). Interestingly, we didn't observe major differences between the wild type and the mutants for what concerns the membrane association (Figure 4.7a), suggesting that single residues are not fundamental for the initial recognition between Golph3 and the Golgi. We therefore decided to test a quadruple-mutant version of the protein, in which all the four amino acids were mutated into alanines at the same time. In this case too, we could not spot statistically significant differences with the wild type (Figure 4.7b). These results suggest that the initial Golph3/Golgi recognition depends on general electrostatics, rather than on specific residues. Nevertheless, all the four reported residues are displaced within the membrane contact area that we identified: this fact seems to suggest that these amino acids might have a role in stabilizing the membrane association through finer interactions, which cannot be depicted by CG MD simulations. To assess this question, we backmapped the last snapshots of CG MD simulations into all-atom (AA) systems, and we ran MD simulations with atomistic resolution. The AA MD would also allow to explore eventual conformational modifications that Golph3 may undergo as a consequence of its interaction with the membrane.
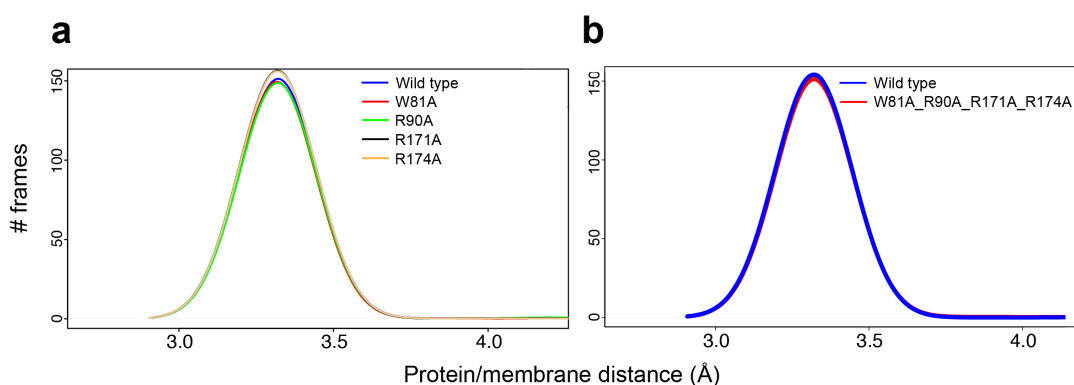
Figure 4.7 – Distribution of MD trajectory frames with respect to Golph3/Golgi distances, for single point (a) and quadruple mutations (b).

The AA MD simulations showed a partial rearrangement of the 3D structure of Golph3. In particular, we observed that the β-hairpin displaced deeper into the membrane (Figure 4.8a); at the same time, a PI4P present within the proximity of the protein got quickly absorbed by a pocket present at the contact surface of Golph3. Interestingly, this pocket is placed at a higher position, on an axis that is transversal to the phospholipid surface: this means that the PI4P needs to be extracted from the surface, and its polar head will be located at a significantly higher position, with respect to the other lipids that compose the membrane (Figure 4.8b). This mechanism was previously described in literature,[111] and was reported to be essential for the protein activity.[280] However, it was never characterized through molecular simulations. The CG MD simulations showed a net prevalence of PI4Ps in proximity of that pocket, but no specific interactions between the pocket and a lipid could be appreciated, likely due to the same nature of the CG force field, which does not allow to reproduce fine interactions with atomistic resolution: in particular, the backbone of the protein is maintained fixed, and therefore the amino acids do not have the possibility to achieve structural rearrangements in order to interact with other molecules. In the the AA MD simulations, the absorption of the PI4P occurs after a few tens of nanoseconds, and this interaction lasts for the whole length of the simulation (800 ns). This behavior also confirms that our model correctly reproduces the behavior of the system.

Interestingly, we found that this PI4P is coordinated by Trp81 and especially by Arg90 and Arg174: in particular, the two arginines establish a stable contact, mediated by a salt bridge between their side chain and the phosphate group of the PI4P (Figure 4.8b). For what concerns the other important residue, i.e. Arg171, in our simulations it did not directly participate to the binding with the PI4P (average distance from the PI4P lipid is 9.7 ±1.7 Å). However, the AA MD simulations showed that Arg171 would directly orient towards the membrane (Figure 4.8a). Although it does not seem to coordinate any specific lipid, its positively charged nature suggests that this residue may play a role in strengthening the interaction between Golgi and Golph3, after the initial recognition has occurred. We can therefore identify two different kinds of lipid/amino acid interactions, one specific (Trp81, Arg90 and Arg174 with regards to the PI4P lipid), and one non-specific (Arg171 with regards to several phospholipids). Both the kinds of interactions seem to be required in order to guarantee the correct functioning of

Golph3.
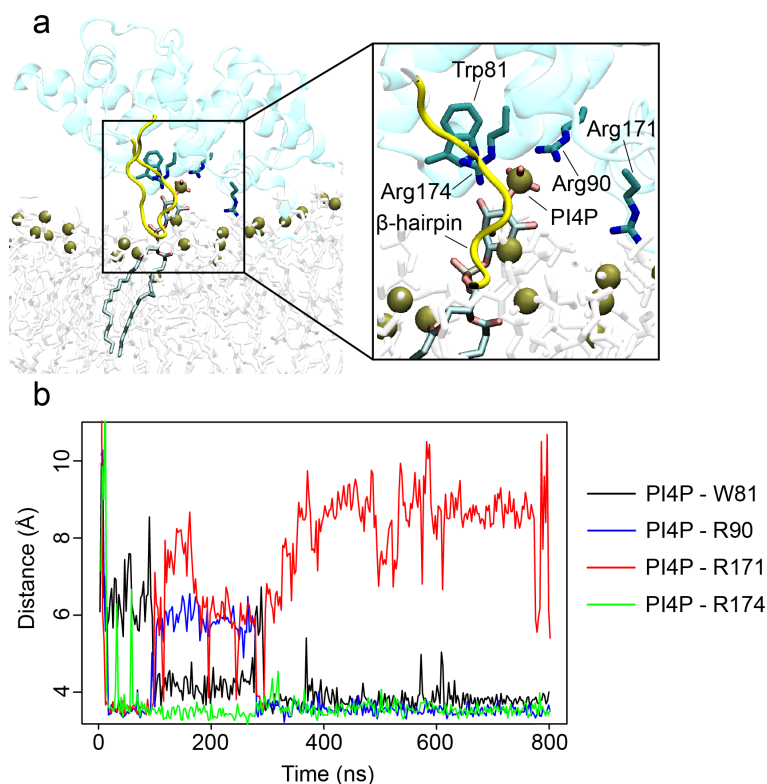
Figure 4.8 – (a) Displacement of Golph3 with respect to the Golgi apparatus, as identified by the AA MD simulations. A PI4P is attracted by the binding pocket of Golph3 which includes residues W81, R90, R171 and R174. (b) Distance evolution of the PI4P with respect to the main residues that compose the binding pocket.

4.2.3 Golph3/LCS peptide interaction: nuclear magnetic resonance

To achieve a complete characterization of the mechanism of action of Golph3, we also aimed at identifying the binding mode between Golph3 and a glycoenzyme C-terminal tail. In particular, we focused on the C-terminal tail of the lactosylceramide synthase (LCS), as our experimental collaborators used this molecule to characterize the features of the peptide recognized by Golph3. In particular, it is known from literature that a specific consensus sequence is required to initiate the peptide interaction.[274] This consensus sequence has the form L-x-x-(R/K), and is present in the cytosolic tail of 47 different glyco-enzymes, most of which belong to the GSL biosynthetic pathway. Mutagenesis experiments confirmed that this domain is crucial for the interaction, and in addition, identified a second region of the

cytosolic tail whose mutation also inhibits the binding. These experiments led to the definition of a binding motif with the following form: (H/R/K)-[x]$_{2-5}$-L-x-x-(R/K). This binding motif constitutes the cytosolic tail of the peptide, and is followed by a trans-membrane domain that spans the Golgi apparatus (Figure 4.9).

LCS peptide sequence:

MRARRGLLRLPRRS LLAALFFFSLSSSLLYFVYVA
Cytosolic tail   *Transmembrane domain*
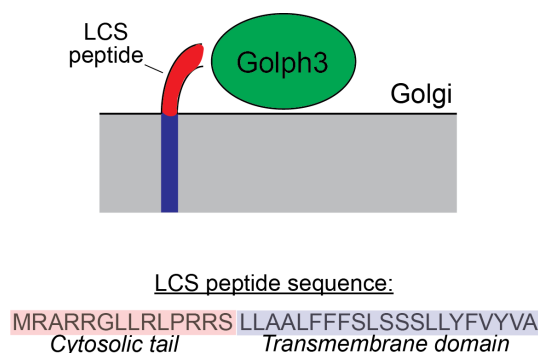
Figure 4.9 – Golph3/LCS peptide binding scheme.

Nuclear Magnetic Resonance (NMR) experiments were conducted in order to identify the Golph3 residues that participate in the interaction with the LCS peptide. In particular, a shift in the relaxation time between the unbound and bound form would imply a conformational rearrangement of that specific residue, which therefore can be assumed to be directly participating to the peptide binding, or at least to be close to the binding site. Unfortunately, although the NMR spectra highlighted significant shifts (Figure 4.10), it was not possible to label the amino acids, and therefore to identify the specific residues that are involved in the peptide binding.

Even without labelling, it is possible to speculate on which residues are represented by the shifts that we see on the NMR spectra. Indeed, some specific residue types generally cover a peculiar region of the NMR spectra, and therefore can be distinguished: for example, glycines are likely to occupy a position in the top-right part of the spectrum. In the spectrum that we have available, we can see that one residue indeed occupies this region of the graph. Moreover, it is also shifting from the unbound to the bound form, meaning that it is likely to play an active role in the peptide binding. However, Golph3 has 9 glycines spread all over its structure, and without labeling, it is virtually impossible to know which of these are actively participating in the binding. One solution would be to realize 9 different mutants, replacing
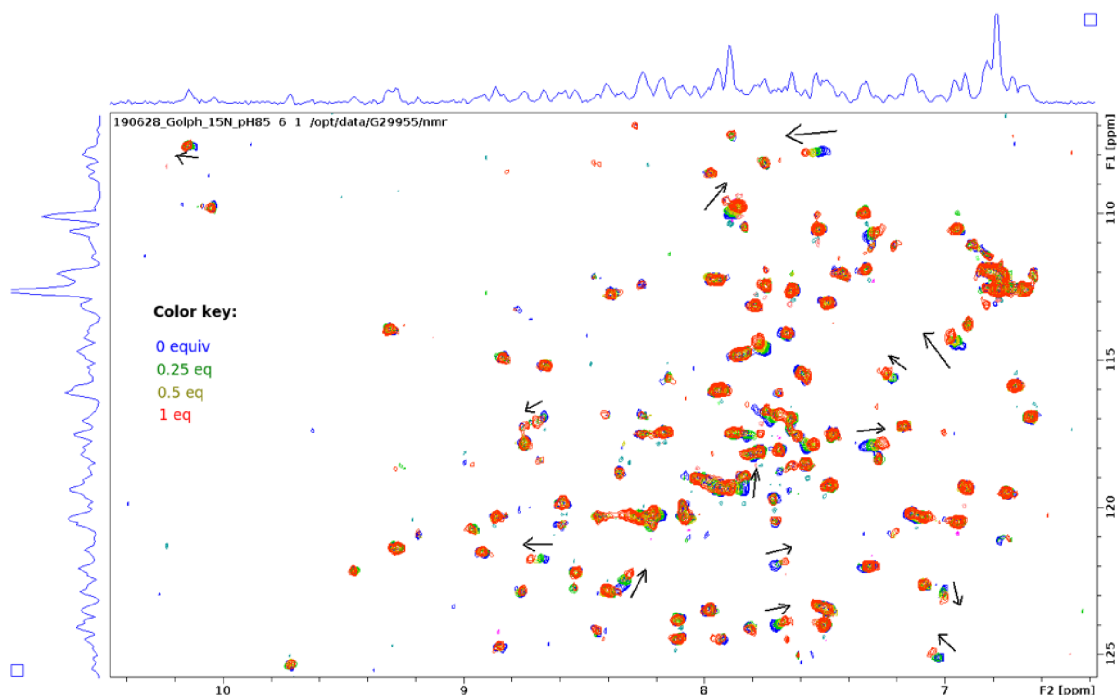
Figure 4.10 – Nuclear magnetic resonance shifts spectra of Golph3 in solution with 0 (blue), 0.25 (green), 0.5 (yellow) and 1 (red) eq of LCS peptide, respectively.

each glycine with an amino acid with opposite characteristics (polar or basic), in order to abolish the work played by the glycine in the peptide interaction. However, the overall stability of the protein should be accurately checked, as even one single point mutation may pose protein folding problems. These experiments are planned to be implemented in the next future.

4.2.4 Golph3/LCS peptide interaction: docking analysis

We therefore decided to apply a computational approach, in particular molecular docking. A visual inspection of the surface of Golph3 did not allow to spot any specific cavity that might be a candidate to host a molecule with the size and characteristics of the LCS peptide. We therefore applied a blind docking procedure: this is generally considered a last-resort technique, as blind docking is characterized by significant limitations. In particular, blind docking does not contemplate the presence of the solvent; moreover, it is biased towards the most exposed surfaces (in contrast with the cavities, which are underestimated); in general, it is less accurate then docking performed on targeted and limited protein volumes. However,

95

given the lack of information present in literature, and the difficulties encountered with the experimental assays (NMR), the blind docking procedure is the only technique that could provide more useful insights. Besides, we were able to add some constraints from geometrical considerations: in particular, given the length of the LCS peptide and knowing the residues that belong to the transmembrane domain, we excluded from the docking box the portion of the protein which is placed farther away from the membrane, and hence unable to interact with the LCS peptide. The excluded domain corresponds to the residues that go from Leu257 to Asn284 and from Leu222 to Met232.

The blind docking did not identify a unique pocket: on the contrary, the identified configurations were distributed over most of the protein surface (Figure 4.11). However, one specific conformation involved the insertion of the LCS peptide within a cavity that is placed close to the β-hairpin domain, and which includes residues Leu71 to Phe80 (Figure 4.11). We evaluated that this cavity could be a good candidate to explain the LCS/Golph3 specific interaction. Moreover, as the molecular simulations predict that the β-hairpin domain drives the interaction of Golph3 with the membrane, we considered plausible that the same domain can also concur in the peptide recognition. In this particular pose, the peptide is predicted to interact mainly with residues Lys72, Asn82, Arg174 and Asn251.

Driven by these results, we analysed the cavity in details, and we identified two key residues, whose single point mutation might potentially inhibit (or reduce) the LCS/Golph3 interaction. The two amino acids are Ala178 and Met199, respectively. In details, we sought to replace the hydrophobic residue Ala178 with a more voluminous hydrophobic residue, specifically a leucine. This mutation would not change the electrostatics of the pocket, however we expect it to cause steric hindrance at the entrance of the pocket, and therefore to prevent the proper binding of the LCS peptide (Figure 4.12a). The second mutation, i.e. M199E, is meant to generate a salt bridge at the entrance of the putative binding pocket. In particular, we speculate that the acidic Glu199 may interact with a close basic residue, i.e. Arg174, hence generating a salt bridge (Figure 4.12b). The orientational change of the sidechains would strongly hinder the putative binding pocket, once again preventing the binding with the LCS peptide. It should also be taken into account that Arg174 is involved in the interaction with the Golgi (see previous sections), and therefore a reorientation of this peptide may significantly
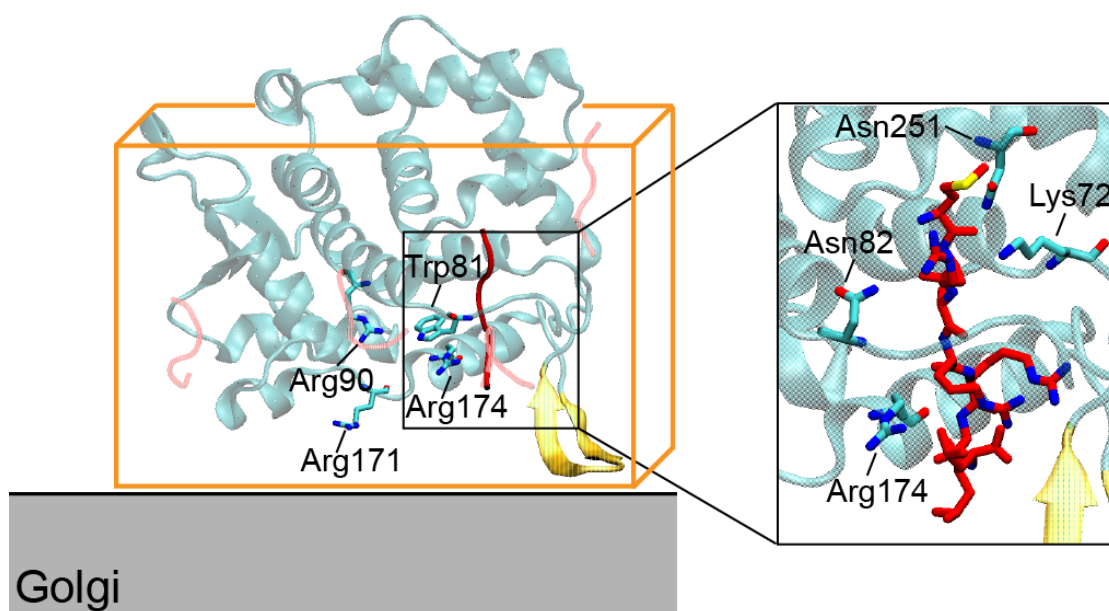
Figure 4.11 – Output of the blind docking calculations (left), and detail of the selected pose (right). The orange square represents the box that has been used for blind docking, which excludes the α10 and α11 helices. The peptide poses identified through blind docking are all in bright red, except for the pose that we selected, which is in magenta. The β-hairpin is marked in yellow.

affect the stability of the Golph3/Golgi contacts.

4.2.5 Proposing a Golph3/LCS recognition mechanism

As a next step, we aimed at identifying the recognition mechanism that drives the interaction between Golph3 and the LCS peptide. The sampling of all possible conformations through MD simulations, until a stable contact between Golph3 and the LCS peptide is reached, would require prohibitive computational resources. We therefore decided to run AA MD simulations separately for the two systems, one including the LCS peptide and the other one including Golph3, respectively (both in presence of a Golgi model). The aim was to evaluate the relationship between the Golgi and the two proteins, in order to identify eventual characteristical patterns that might drive the recognition between them. The atomistic simulations revealed a peculiar and surprisingly similar influence of both the molecules on the Golgi: in particular, we observed a small lipid curvature of the Golgi appearing in proximity of both the peptide and Golph3 (Figure 4.13). This depression is likely due to the presence of residues Lys191 and
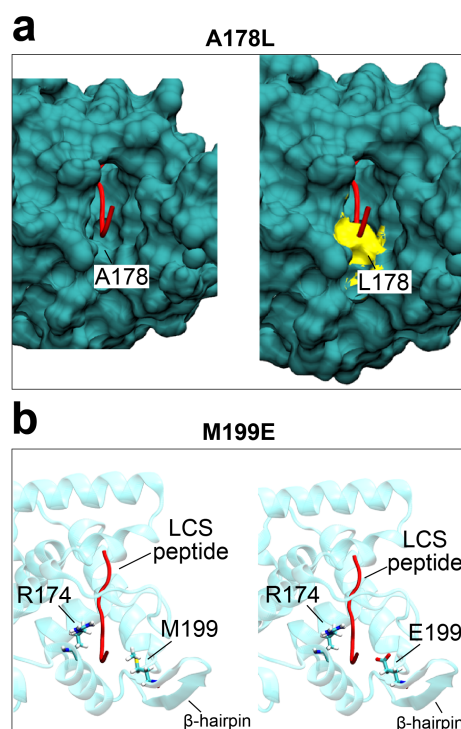
Figure 4.12 – Details of two single point mutations that could potentially disrupt the interaction between Golph3 and the LCS peptide.

His202 which, thanks to their positively charged sidechain, attract the phosphate groups of the surrounding lipids, thus lowering the level of the leaflet. A similar argument applies to the LCS peptide, where the lipid phosphates are pushed towards the hydrophobic core by the multiple arginines that are present at the interface with the membrane, in particular Arg9, Arg12 and Arg13. This observation drives the hypothesis that the β-hairpin is involved in the recognition of the peptide. In particular, we hypothesize that the recognition of the LCS peptide might be driven by the sensing of the membrane depression that is induced in proximity of the LCS peptide. The AA MD simulations conducted in presence of the LCS peptide also reveal that, in absence of Golph3, the peptide tends to assume a bent conformation, with the N-terminus contacting the polar heads of the phospholipid bilayer.

4.2.6 Tryptophan fluorescence assays

Tryptophan fluorescence assays use a form of fluorescence spectroscopy, to get insights of proteins structural features. In particular, this technique takes advantage of the natural
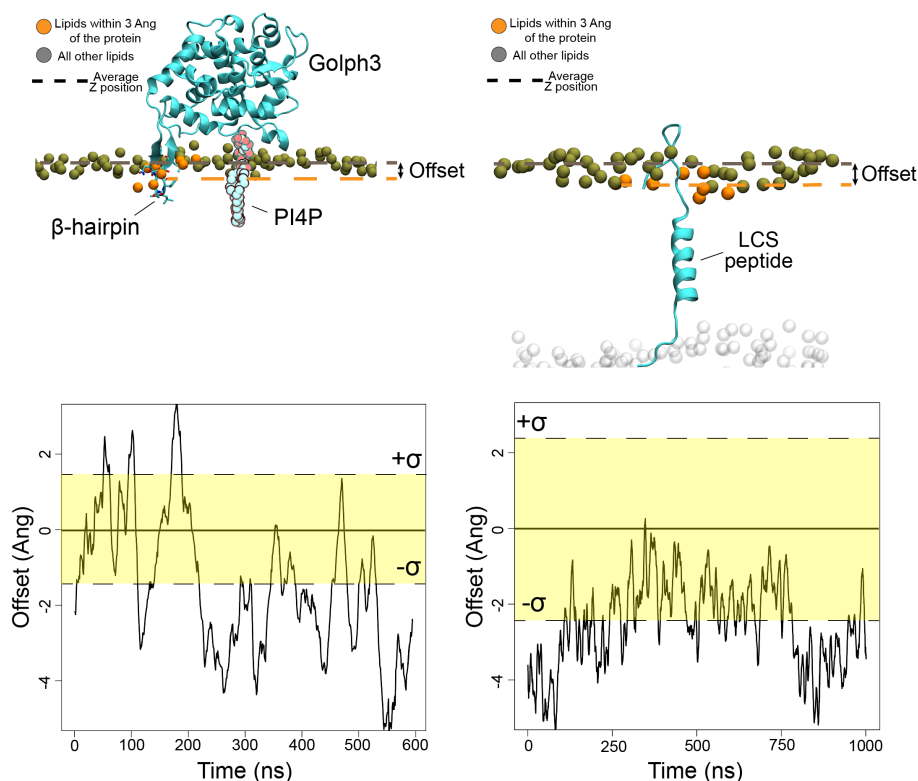
Figure 4.13 – Lipid depression induced by LCS peptide and Golph3. The plots show the offset in the Z coordinate between the lipids around the proteins (β-hairpin domain for Golph3, right, and the LCS peptide, left) and all the other lipids composing the membrane. The plotted Z coordinate is the one of the phosphate group of each lipid. To provide a measure of the significancy of the offset fluctuations, the standard deviations of the average lipid fluctuation is drawn (dashed lines).

fluorescence provided by tryptophans which, when excited by a light at the wavelength of 280 nm, respond with an emission light which is characterized by a variable peak, which should generally be centered at 340 nm, but can range from 308 nm to 350 nm (Figure 4.14a). By comparing the two emission peaks prior and after a event (such as the insertion of a binding molecule in solution), it is possible to evaluate the reaction of the protein, in terms of the capability of its tryptophans to get excited. For example, if a molecule binds the protein in an area that is close to a tryptophan, it would be reasonable to expect that the emission peak of the protein would be affected, since that tryptophan would likely be covered by the binding molecule and therefore less exposed to the light. The structure of Golph3 has 5 tryptophans (Trp81, Trp152, Trp161, Trp225, Trp290), one of which (Trp81) is close to the putative binding pocket that we identified through docking: therefore, it is reasonable to believe that the

fluorescence of Golph3 should be significantly affected by the addition of the LCS peptide in solution. Therefore, collaborators performed the tryptophan fluorescence assays, in order to investigate this aspect.

However, the experimental setup did not highlight any significant difference between Golph3 alone and Golph3 in presence of the LCS peptide (Figure 4.14b). This would suggest that the identified binding pocket is not involved in the peptide binding. However, the fact that the two profiles are very similar, would also suggest that none of the 5 tryptophans present in the structure of Golph3 are involved in the binding. This is highly unlikely, because they cover a large portion of the protein. An alternative explanation may be that the interaction between Golph3 and the LCS peptide is strongly dependent on the presence of the Golgi, and cannot occur in its absence. Although it is not possible to perform tryptophan fluorescence assays in presence of a model of the Golgi, it would be doable to repeat the experiments in presence of PI4P polar heads which, as previously discussed, play an important role in the activation of Golph3. These experiments will be performed in the next future.
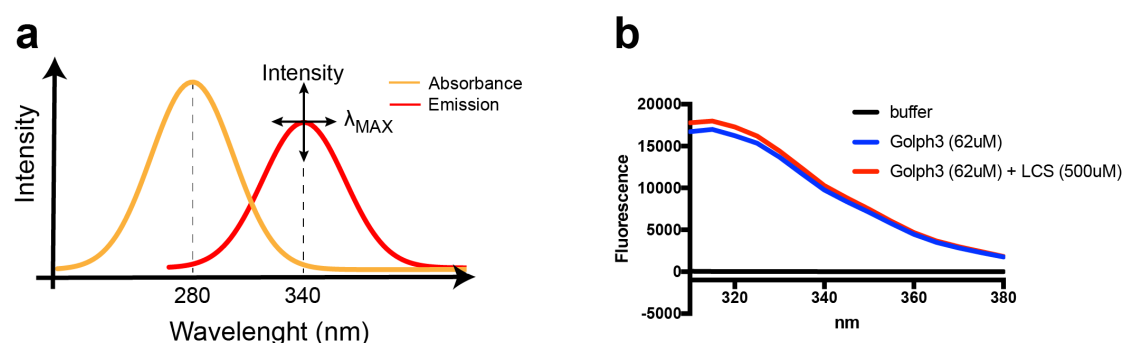
Figure 4.14 – (a) Trytophan fluorescence mechanism. A binding event can influence the emission spectrum, either in its peak (maximum wavelength), or in its intensity. (b) Tryptophan fluorescence assays of Golph3 in absence and presence of LCS peptide.

4.3   Conclusions

The oncogenic activity of Golph3 has been reported in a large variety of studies[89,283–285]: despite the importance of this protein in the formation of several different kinds of malignant tumors, its molecular mechanism of action contains several aspects that are not clarified, especially for what concerns its binding mode with glycoenzymes and with the Golgi. This is

mainly due to the intrinsic difficulties that arise when treating membrane proteins in wet lab experiments[63, 286, 287] (see chapter 1). In the present work, we also experienced difficulties of this type in two wet lab tests (NMR and tryptophan fluorescence assays) that were unable to provide the desired information. In this context, molecular simulations can provide insights into the aspects that are not accessible through experiments. In particular, MD simulations clarified that the absence of PI4Ps in the Golgi does not preclude its interaction with Golph3, but can strongly influence its binding dynamics, eventually causing the association mechanism to become less efficient. We speculate that the structure of Golph3 is optimized to interact with the Golgi: indeed, the simulations conducted in presence of a Golgi model with different amounts of PI4Ps (0%, 10% and 20%) show that the interaction with Golph3 is affected both by a smaller and by a larger presence of PI4Ps. In the latter case, we found that in a few MD replicas, the interaction was mediated by an erroneous surface of interaction. These results suggest that a 10% level of PI4Ps in the membrane might be the ideal amount for Golph3 to bind correctly to the Golgi, and therefore to perform its activity. Overall, the simulations higlight a high sensibility of this protein towards variations in PI4P levels. In order to experimentally verify this hypothesis, we propose the use of liposome floatation assays, which proved to be an efficient tool to evaluate the action of specific lipid types when treating NDM-1 (see chapter 3).

The atomistic simulations also showed in details the role played by the reported key residues Trp81, Arg90, Arg171 and Arg174. In particular, Trp81 and Arg90 are involved in the rapid absorption of a PI4P lipid by Golph3, which previous studies suggest to be essential for its activity.[111, 280] Arg171 and Arg174 do not seem to participate in this absorption, but their orientation suggests that they might strengthen and stabilize the contact with the Golgi. In summary, we can attribute the Golph3/Golgi binding to a synergistic combination of different factors: on one side, general electrostatics is fundamental for driving the initial (and very fast) recognition of the interaction surface, as spotted by the CG MD simulations; in addition, we also have the contributions that were revealed by the AA MD simulations, which are the specific interactions with a PI4P lipid and the role played by the β-hairpin which is able to hydrophobically interact with the inner layer of the membrane, hence promoting a stronger binding between the protein and the bilayer.

The oncogenic activity of Golph3 could be prevented through *ad hoc* molecules that target specific protein sites and inhibit the interaction of Golph3 with the glycoenzymes, and consequently the GSL pathway and cell proliferation. However, no drugs of this type are currently present on the market, mainly because no accurate information is available about the mechanism of action of Golph3.[273] The present study highlighted a putative binding pocket in proximity of the β-hairpin domain, and a possible recognition mechanism was proposed, in which the molecular recognition is mediated by a small depression in the Golgi. The hypotheses here expressed require experimental verification: we proposed the A178L and M199E single point mutations, in order to assess the influence of these specific residues in the interaction of Golph3 with the glycoenzymes.

Moreover, we propose to repeat the tryptophan fluorescence assays described in this chapter in presence of PI4P polar heads: according to the simulations and to previous studies,[111,280] the presence of PI4Ps is fundamental not only for the interaction with the Golgi, but also for the proper binding with the glycoenzymes. Their presence in solution may therefore significantly affect the output of the experimental tests.

If the newly designed experiments will confirm the validity of the identified binding site, a virtual screening campaign will be conducted, in order to identify inhibiting molecules that can compete with the physiological ligands. Successively, the hit compounds will be tested *in vitro*, in order to assess with greater accuracy their potency and efficacy in terms of blocking the cell proliferation induced by Golph3.

4.4   Materials and methods

Coarse-grained molecular simulations. All CG MD simulations were conducted with the MARTINI2.2p (polarizable) force field.[80] In the Golph3/Golgi simulations, the enzyme was located at 6 nm from the bilayer. 5 different replicas for each of the conditions (physiological Golgi with the composition specified in the Introduction, Golgi with 0% PI4Ps, Golgi with 20% PI4Ps) were run, each for a simulation time of 2 μs. All lipid bilayers were generated with the Insane tool of MARTINI.[82] The structure of Golph3 was taken from the Protein DataBank (PDB code: 3KN1[111] and coarse-grained with the Martinize tool. The systems were solvated with the

polarizable MARTINI water[178] and ionized in 150 mM of NaCl. Frames for the analysis were collected every 750 ps. As explained in the previous chapter, a binding event is considered to have occurred when the protein settles at 3 Å distance from the membrane, and this distance remains constant for the rest of the simulation (i.e. no detachment occurs). Each system was minimized with 5000 steps of steepest descent algorithm; successively, it was equilibrated with 5 ns of MD in NVT conditions, using Particle Mesh Ewald (PME) for the electrostatic contributions and velocity rescale algorithm for temperature coupling at 310 K. The production phase was conducted in NPT ensemble, using a Parrinello-Rahman semi-isotropic coupling algorithm[245] for maintaining the pressure constant at 1 bar.

All-atom molecular simulations. The last snapshots of 3 replicas among the Golph3/Golgi CG MD simulations were backmapped to AA. All lipids were backmapped with the Backward tool of MARTINI.[247] The crystal structure of the protein available at the Protein DataBank (PDB code: 3KN1) was aligned to the CG model to provide the starting configuration of the AA MD simulations. The CHARMM36m force field was used for all AA MD simulations.[288] All AA systems were solvated with a TIP3P solvent model and ionized in 150 mM NaCl. The equilibration procedure was as follows: initially, we followed the procedure suggested by the MARTINI team, which consists in performing a first minimization step after inactivating all non-bonded interactions (Coulomb and Lennard-Jones); non-bonded interactions are then gradually restored in successive MD equilibration stages. After having reached a reasonable equilibrium, we performed another stage of minimization (5000 steps of steepest descent algorithm), followed by 5 different stages of MD equilibration in NVT ensemble (for a total of 500 ps), with harmonic constraints applied both on the protein backbone and on the lipids phosphates, which were gradually released. The production phase was run in NPT ensemble, using the Verlet algorithm[252] for the neighbor search, velocity rescale temperature coupling algorithm[253] for maintaining the temperature constant at about 310 K and Parrinello-Rahman pressure coupling algorithm[245] in semi-isotropic conditions at a pressure of 1 bar. The electrostatic interactions were treated with Particle Mesh Ewald method.[254] Each AA MD replica was simulated for 800 ns. Frames for analysis were collected every 100 ps.

Both in the CG and in the AA simulations, the lipid occupancy was evaluated through the VolMap Tool plugin of VMD, in particular by calculating the average occupancy of each lipid

phosphate throughout the trajectory. The electrostatic potentials calculations were performed through the APBS software,[255] which solves the Poisson-Boltzmann equations to compute the electrostatic potentials of biomolecules. The dielectric constant was set at 78 for the solvent, and 2 for the protein; the ionic concentration was set at 150 mM NaCl; protonation states of the protein aminoacids were assigned with PropKa.[256]

Docking. The docking experiments were performed with AutoDock Vina.[289] A box of appropriate size was used as shown in Figure 4.11. The amino acids of Golph3 were kept rigid, while the LCS peptide was flexible, which means that its torsional angles were left free to rotate. The first 5 poses identified by the software according to the AutoDock scoring function were considered for analysis.

5 Conclusions

The molecular characterization of biological interfaces is a prevailing topic in modern biology.[290–294] It also constitutes a very challenging task, when one of the surfaces that are involved is constituted by a biological membrane: this is mainly due to its instability, which makes it particularly arduous to be treated with currently available experimental methods.[63, 286, 287] In this context, molecular modeling comes of help to address the open issues that experimental techniques cannot answer, either because it would be unfeasible or simply because of its high cost.

With respect to experimental methods, molecular simulations have the obvious limit to represent a model, rather than reality. However, the great advancements that have been made in the past 30 years in terms of force fields development, allow to obtain results that are particularly reliable in the description of biomolecular systems.[295–298]

In the present work, we made ample use of molecular simulations (both at an atomistic and at a coarse-grained level) to characterize biological interfaces: in particular, we focused on two systems that still lacked an accurate molecular description, despite representing a hot topic in medicine.

NDM-1 represents a potentially high risk for the public health: the constant drop in the number of annually approved antibiotic molecules, caused the rise of a generation of bacteria that are resistant to all current treatments.[190, 191, 205, 206] In this perspective, NDM-1 represents one of the biggest threats to public health, as it provides resistance to carbapenems, which are considered last-resort antibiotics.[211] This enzyme is present in a wide variety of bacteria,

especially belonging to the family of *Enterobacteriaceae* (such as *Klebsiella pneumoniae* and *Escherichia coli*), but also others (such as *Acinetobacter baumannii*).[215,216] Notably, several other enzymes of the same class (metallo-β-lactamases) exist, which do not represent the same threat to public health, as NDM-1 does. It has been proven that the main difference between NDM-1 and its homologs is represented by the allocation within the cell: NDM-1 is the only MBL to reside in the outer bacterial membrane, as a peripheral membrane protein, rather than in the periplasm, as a soluble one.[110] This peculiar location is granted to NDM-1 by a post-translational modification at its N-terminal region, and confers to the enzyme several evolutionary advantages, such as the capability to be secreted into vesicles and consequently to be more easily spread within the infection site. Based on this unique characteristic, we evaluated the mechanisms which allow NDM-1 to achieve a stable contact with the bacterial membrane. We found that the post-translational modification is not the only element that drives the initial interaction between the protein and the bacterial membrane. This statement was confirmed by experimental verifications, in particular through liposome flotation assays. Overall, we provided a mechanistic characterization of how NDM-1 interacts with the bacterial phospholipid bilayer, and we proposed allosteric sites that may be targeted, in order to block this interplay and reduce NDM-1 to its soluble, less harmful version.

Golph3 is overexpressed in many of the most lethal cancer forms.[89] The mechanism of its carcinogen action is partially unclarified, and no molecules are currently approved which can partially or totally block its expression or activity.[273] We aimed to characterize how this protein relates with its native environment, which is the Golgi apparatus. The simulations showed that Golph3 interacts with the Golgi in a very specific way, i.e. through a well-defined portion of its surface. This is similar to what we saw for NDM-1 and the bacterial membrane. Interestingly, like NDM-1, also in Golph3 the identified portion of the surface does not correspond to the most positively charged area of the protein, as we would have expected, since both the Golgi and the bacterial membrane are negatively charged.

We were also able to provide a molecular description of a mechanism that had been previously proposed in literature, such as the absorption of a PI4P lipid within a pocket of Golph3. Moreover, we proposed a recognition mechanism between Golph3 and the LCS peptide, which will require experimental verification.

In general, we found that a specific membrane composition is essential to guarantee the correct interplay between protein and membrane at the interface. For both the systems that we evaluated, we found that specific lipid types drive the protein/membrane interaction by establishing contacts with specific amino acids: this mechanism is possibly common to other peripheral membrane proteins. Notably, both for NDM-1 and for Golph3, the lipid types that played the most relevant role in protein recognition were present in small amounts (6% CDLs in bacterial membrane, 10% PI4Ps in Golgi). Notably, similar results were obtained in our laboratory for other two PMP systems, specifically the coenzyme Q9 homolog (COQ9) and the acyl protein thioesterase 2 (APT2). COQ9 is involved in the biosynthesis of coenzyme Q, which is a isoprene lipid that serves as a cofactor for several enzymes. COQ9 is a PMP that is present at the inner mitochondrial membrane: this membrane is very rich in CDLs. Molecular simulations demonstrated that the interaction between COQ9 and the inner mitochondrial membrane is strongly dependent on the amount of CDLs.[229] In particular, driven by the presence of the CDLs, COQ9 binds to the bilayer and induces a warping on the membrane itself, hence opening a hole in the polar area, which consents to host the very hydrophobic α-helix 10 into the hydrophobic region of the membrane. *In silico*, this phenomenon does not occur in absence of CDLs, and successive *in vivo* studies assessed how the interaction between COQ9 and the membrane is strongly diminished in absence of CDLs. APT2 is instead an enzyme which guarantees the reversibility of the S-acylation PTM. The enzymatic activity of APT2 strongly depends on its peripheral membrane localization and has been shown to be particularly optimized in presence of specific lipid proportions, i.e. 40% PCs, 40% PSs and 20% PEs (unpublished data). This aspect emphasizes the importance of using realistic models of the membrane when performing molecular simulations, as even minor differences in the initial conditions can lead to inaccurate results. For the same reason, it is once again showed how essential it is for molecular modeling methods to continuously interact with the experimental world, in order to get all the available information for the system of interest, and turn the output data of the experiments into the input data of the simulations. In this specific case, we refer to lipidomics analyses, which provide information about lipid composition in membranes; however, the same argument may apply to other kinds of data, such as protein 3D structures with atomistic resolution (coming from X-ray crystallography, NMR or cryo-EM methods) that represent a fundamental step to set up a molecular system.

We also concluded that for what concerns peripheral membrane proteins, stable protein/membrane configurations are driven by multiple factors, which overall concur in strengthening the interaction. The first factor that needs to be taken into account is general electrostatics, which is crucial for the initial protein recognition: as we saw in the case of NDM-1 vs. VIM-2, the overall electrostatic properties can drive a very different behaviour between two proteins that, overall, are structurally very similar. To make sure that the observed differences could be attributed to general electrostatics, we realized a chimera (N-VIM) with the bulk electrostatic properties of VIM-2 and the surface of interaction of NDM-1. The behaviour of this chimera with respect to the membrane is strongly shifted towards VIM-2, hence confirming that the overall electrostatics is the main feature that drives the initial protein recognition. However, electrostatics needs to be particularly well tuned, in order to obtain a contact with the correct protein surface of interaction. As we saw in the case of both NDM-1 and Golph3, small changes in the amounts of anionic lipids in the membrane, can affect very significantly the interplay with the protein, thus influencing its biological function. Moreover, electrostatics alone is not sufficient to fully explain the PMPs/membrane interplay: as we saw in the case of both NDM-1 and Golph3, also specific amino acid/lipid interactions are required. In particular, it is necessary to attract the correct lipid types both from a global point of view, with clusters of specific lipid types interacting with the PMP in an non-specific way (in our systems, clusters of CDLs and PI4Ps drive the interaction between the membrane and NDM-1/Golph3, respectively); and specific amino acid/lipid interactions that strengthen and stabilize the protein/membrane interplay on a long term time scale (these specific interactions involve R45 and R52 in NDM-1; W81, R90, R174 in Golph3). Finally, the insertion of specific domains seems also to provide a big advantage to the PMP/membrane interaction: let alone the covalent interaction provided by a lipidation, which obviously can add a lot of stability to the contact (as we saw for NDM-1), it is indeed possible to obtain a similar effect through simple hydrophobic interactions, as we saw in the case of the β-hairpin domain for Golph3.

In the present work, we showed how molecular modeling can be used in combination with experimental methods, to provide answers that would not be possible to obtain otherwise. Experimental verification remains central to validate the predictions of the simulations, and to confirm the proposed biological mechanisms. As we saw in the present work, a classical

example of experimental validation is provided by mutagenesis experiments, which can knock out residues that the simulations predicted to be important for the protein activity; or by liposome floatation assays, which allow to adopt a similar approach, but from the membrane perspective (i.e. to knock out a lipid type that is supposedly important for the protein activity). For example, in the case of NDM-1, the liposome floatation assays confirmed that this enzyme is able to bind the bacterial membrane also in its soluble form (i.e. with no post-translational modification), hence confirming the output of the simulations. The synergy between experimental and modelling methods is therefore required to overcome the limits of each of the two approaches.

It may be expected that, in perspective, the advancements in both experimental and computational techniques would end up in making obsolete one of the two fields, while the other becomes more and more powerful in characterizing biomolecular systems. However, this does not seem to be the case, at least in this time and age: on the contrary, synergistic approaches are getting more and more powerful, as both the fields evolve. As a matter of fact, the great advancements achieved in the field of experimental techniques, which led to the so-called "resolution revolution", are still far from being able to provide a full and complete characterization of the molecular events that occur at a biological level, especially for what concerns the dynamic aspects (the static properties are instead characterized quite well by the currently available techniques); on the computational side, despite the big increase in computational resources that characterized the last decades, we are still far from being able to simulate systems at the time-scale and space-scale that are of interest for whole cells, or even whole organisms. Pioneering studies have been performed, to approach the millisecond time-scale,[299–302] and micrometer space-scale.[303–305] However, it is still impossible to combine the two and achieve a simulation of a full cell at a reasonable time-scale. Thus, a combined computational/experimental approach is likely to characterize molecular biology for many years to come. A deep understanding and collaboration between the two approaches is desirable and could potentially exploit the big challenges of modern medicine: first of all, a deeper comprehension of the molecular aspects of life, and second the development of tools to directly interact with the molecular machinery of cells, in order to induce desirable effects such as the inhibition of pathogenic infections or the suppression of cancerogenic pathways.

A Salt enhances the thermostability of enteroviruses by stabilizing capsid protein interfaces

This appendix is adapted from the following paper:

"Salt enhances the thermostability of enteroviruses by stabilizing capsid protein interfaces." Meister S, **Prunotto A**, Dal Peraro M, Kohn T. (2020). *Journal of Virology, 94*(11).

Author contributions: S.M. and T.K. conceived and planned the experiments. S.M. carried out the experiments and analyzed the data. A.P. and M.D.P. planned the molecular modeling. A.P. carried out the molecular modeling. T.K. and S.M. wrote the manuscript in consultation with A.P. and M.D.P.

Although not directly related to peripheral membrane proteins systems, this work is another useful example of the usage of integrative modelling for characterization of biological interfaces, and specifically of protein/protein interfaces.

Enteroviruses are common agents of infectious disease that are spread by the fecal-oral route. They are readily inactivated by mild heat, which causes the viral capsid to disintegrate or undergo conformational change. While beneficial for the thermal treatment of food or water, this heat sensitivity poses challenges for the stability of enterovirus vaccines. The ther-

Appendix A. Salt enhances the thermostability of enteroviruses by stabilizing capsid protein interfaces

mostability of an enterovirus can be modulated by the composition of the suspending matrix, though the effects of the matrix on virus stability are not understood. Here we determined the thermostability of four enterovirus strains in solutions with varying concentrations of NaCl, and pH. The experimental findings were combined with molecular modeling of the protein interaction forces at the pentamer and the protomer interfaces of the viral capsids. While pH only affected thermostability, increasing NaCl concentrations raised the breakpoint temperature of all viruses tested by up to 20 °C. This breakpoint shift could be explained by an enhancement of the van der Waals attraction forces at the two protein interfaces. In comparison, the (net repulsive) electrostatic interactions were less affected by NaCl. Depending on the interface considered, the breakpoint temperature shifted by 7.5 or 5.6 °C per 100 kcal/(mol*Å) increase in protein interaction force.

A.1 Importance

The Enterovirus genus encompasses important contaminants of water and food (e.g., coxsackieviruses), as well as viruses of acute public health concern (e.g., poliovirus). Depending on the properties of the surrounding matrix, enteroviruses exhibit different sensitivities to heat, which in turn influences their persistence in the environment, during food treatment or during vaccine storage. Here we determined the effect of NaCl and pH on the heat stability of different enteroviruses, and related the observed effects to changes in protein interaction forces in the viral capsid. We demonstrate that NaCl renders enteroviruses thermotolerant, and that this effect stems from an increase in van der Waals forces at different protein subunits in the viral capsid. This work sheds light on the mechanism by which salt enhances virus stability.

A.2 Keywords

echovirus 11, coxsackievirus B1, coxsackievirus B5, thermostability, pentamer interface, protomer interface

112

A.3 Introduction

Enteroviruses are common viruses infecting humans, and they can cause a spectrum of diseases ranging from mild to fatal[306](1). Vaccines against several enteroviruses are already available (e.g., poliovirus[307](2), enterovirus 71[308](3)) or are under development (e.g., enterovirus D68[309](4), coxsackievirus A16[310](5)). To enable global immunization programs, such vaccines should be thermostable, to minimize the loss of potency over time[311](6). Enteroviruses are also common contaminants of water[312](7) and food[313](8). In these matrices, a high viral thermostability is detrimental to human health, as it can interfere with efforts at food preservation and consumption, and enhance the environmental persistence of waterborne viruses[314](9).

Enteroviruses have a single-stranded RNA genome, surrounded by an icosahedral capsid composed of 60 protomer repeats composed of four structural proteins (VP1 to VP4), and organized in pentameric subunits (Figure A.1). For some enteroviruses, the mechanism of thermal inactivation involves the disintegration of the viral capsid. Specifically, the dissociation of the capsid into pentameric subunits with increasing temperature was observed for foot-and-mouth disease virus (FMDV)[315](10). This indicates that interaction forces between the capsid pentamers modulate this virus' thermostability. Therefore, one option to enhance the thermoresistance of FMDV is to increase the interaction forces between the pentameric subunits, by selective mutation of the amino acids at the interface. For example, engineering a disulfide bond into the pentameric interface via the mutation of a single amino acid in VP2 efficiently increased the thermostability of FMDV and shifted the capsid melting temperature upward by several degrees[315,316](10, 11). A comparable shift was observed when noncovalent interactions between pentamers were enhanced, by introducing a VP2 mutation that causes hydrophobic stacking of aromatic side chains[317](12). And finally, engineered mutations that reduced electrostatic, carboxylate-mediated repulsion forces at the pentameric interface also enhanced the thermostability of FMDV[318](13).

Other enteroviruses (e.g, poliovirus[319,320](14, 15), enterovirus 71 (EV71)[321](16), coxsackievirus A7[322](17)) have been shown to undergo conformational rearrangement rather than dissociation upon mild heating (50-56 °C). This rearrangement is initiated by capsid expansion, which involves rotation of the protomer and leads to structural disruptions along the pentamer

Appendix A. Salt enhances the thermostability of enteroviruses by stabilizing capsid protein interfaces
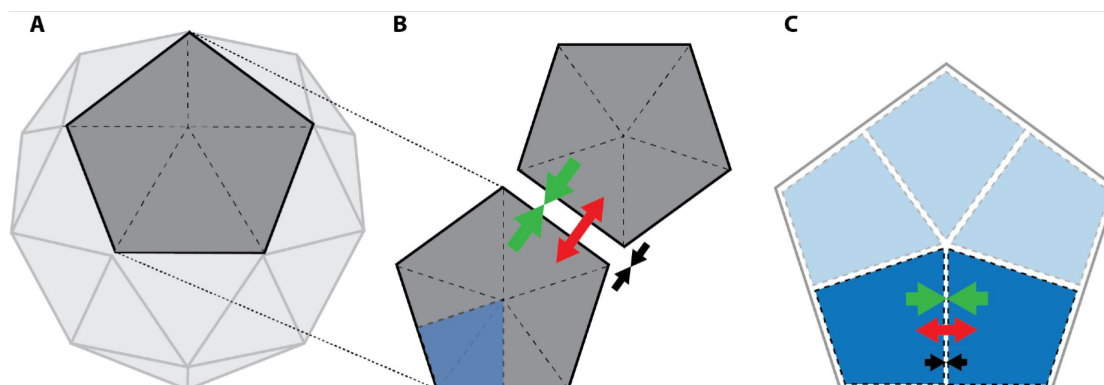
Figure A.1 – Structural features and interaction forces of the virus capsid. In green: Van der Waals forces; in red: electrostatic forces; in black: overall interaction forces. (A) Schematic representation of the icosahedral capsid structure, a single pentamer is displayed in grey. (B) Schematic representation of pentameric interaction forces, a single protomer is displayed in blue. (C) Schematic representation of protomeric interaction forces, each protomer contains a single copy of each structural protein (VP1 to VP4).

and protomer interfaces[319–322] (14–17), and ultimately to the formation of a pore in the 2-fold axis of symmetry through which the viral RNA is released[320, 323, 324] (15, 18, 19). It is reasonable to assume that the initial rotation, as well as the structural disruptions, may be inhibited by stronger interaction forces at the subunit interfaces. Experimental evolution experiments have shown that thermostable poliovirus variants carry mutations in the hydrophobic pocket region of VP1[325–327] (20–22). These mutations are distant from the pentameric interfaces, but close to the interface of the protomeric subunits. They may thus increase capsid stability by enhancing the interaction forces at the protomer interface, though this effect has not been investigated.

Besides changes to the capsid composition, the thermostability of enteroviruses can be affected by the composition of the suspending matrix. Most importantly, it has long been recognized that the thermostability of enteroviruses is enhanced in the presence of dissolved salts, and that the enhancement depends on the salt concentration and identity[328–330] (23–25). The mechanism by which salts achieve this thermostabilizing effect, however, remains unclear. We can reasonably assume that salts modulate the interaction forces at the protein interfaces. Specifically, we expect that shifting the ionic strength of a virus bearing solution results in shielding or promotion of electrostatic and van der Waals forces between protein subunits. Similar effects may be expected for shifts in the solution pH.

114

In this study, we investigated how changes in matrix composition modulate the protein interaction forces and the thermostability of enteroviruses. Specifically, we determined the thermal inactivation kinetics at 55°C, as well as the capsid breakpoint temperatures, in solutions ranging between 0.01 to 3 M NaCl, and over a pH range from 3-9. Hereby we define the capsid breakpoint temperature as the lowest temperature that causes a rapid loss of virus infectivity. Molecular modelling was applied to compute the corresponding electrostatic and van der Waals forces at different protein interfaces. As a representative of the Enterovirus genus, we used coxsackievirus B5 (CVB5). CVB5 is an enteric human pathogen that is frequently detected in the aqueous environment[331](26), and that is notoriously resistant to disinfection[332](27). The most detailed analysis was conducted on the CVB5 Faulkner strain (CVB5-Faulkner). To generalize our findings to other Enterovirus strains, the results for CVB5-Faulkner, were compared to a CVB5 environmental isolate (CVB5-L061815), an environmental isolate of closely related serotype (coxsackievirus B1; and CVB1-L071615), and a more distantly related serotype (echovirus 11 Gregory strain; E11)[333](28).

A.4 Results and Discussion

A.4.1 Effect of salt and pH on breakpoint temperature and thermal inactivation kinetics

To investigate the role of the matrix on virus stability, we determined the breakpoint temperature and inactivation kinetics in buffer solutions with varying salt concentrations and at different pH values. A linear, positive correlation was observed between the breakpoint temperature and the salt concentration up to 1.5 M NaCl (Pearson's $r = 0.98$; Figure A.2A). Specifically, the breakpoint temperature of CVB5-Faulkner increased from 45.0 °C at a NaCl concentration of 10 mM to 64.7 °C at 1.5 M, with an average breakpoint shift of 12.2 °C / M NaCl. Beyond 1.5 M NaCl, the breakpoint continued to increase, albeit at a slower rate, reaching a maximal value of 67.2 °C at 2.5 M.

The shift in breakpoint temperature translated into pronounced changes in the sensitivity of CVB5-Faulkner to thermal inactivation. In a matrix containing 10 mM NaCl, the exposure of CVB5-Faulkner to 55 °C resulted in an inactivation of 3 $\log_{10}$ within 15 seconds (Figure A.3).

Appendix A. Salt enhances the thermostability of enteroviruses by stabilizing capsid protein interfaces
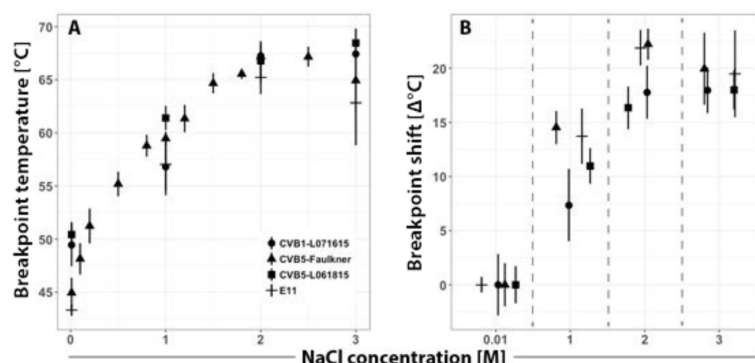
Figure A.2 – Effect of salt on the thermal stability of enteroviruses. (A) Capsid breakpoint temperatures of various Enterovirus B strains and NaCl concentrations. (B) Shift in breakpoint temperature between 10 mM and 1, 2, or 3 M NaCl solutions. Error bars indicate the 95% confidence intervals.

This finding was expected, given that the breakpoint of CVB5-Faulkner in this solution (45.0 °C; Figure A.2A) lay well below the treatment temperature. In contrast, if the matrix contained 1 M NaCl, the virus breakpoint (59.5 °C; Figure A.2A and Figure A.4) was greater than the treatment temperature, and hence the virus remained stable over the timeframe considered (Figure A.3).
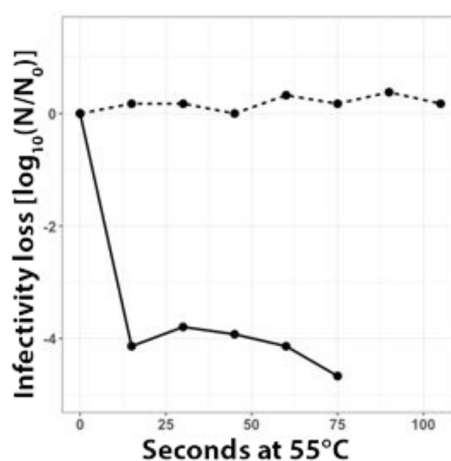
Figure A.3 – Effect of salt on the thermal inactivation kinetics of CVB5-Faulkner. The black line indicates inactivation at low salt condition (10 mM NaCl); the dotted line indicates inactivation at high salt conditions (1M NaCl). N/N_0 is the residual fraction of infectious viruses.

Transmission electron microscopy (TEM) images revealed that in the 10 mM NaCl matrix, exposure of CVB5 to 55 °C resulted in a decrease in the number of intact viral particles per grid area analyzed (162 vs. 84 viral particles per 30 μm^2 in the untreated and heated sample,

116

virus	NaCl [M]	Pentamer Interface										Protomer Interface												
		Overall forces		Electrostatic forces		Van der Waals forces		H Bonds		Salt Bridges		Overall forces		Electrostatic forces		Van der Waals forces		H Bonds		Salt Bridges		Breakpoints	
		Value	SE	Value	SE	Value	SE	Number	SE	Number	SE	Value	SE	Value	SE	Value	SE	Number	SE	Number	SE	Value	SE
CVB5- Faulkner	0.01	-111.2	16.2	180.1	7.1	-291.3	9.2	140	1.1	4	0.2	-97.8	63.2	242.7	2.9	-340.4	64.7	148	1.7	6	0.0	44.96	0.72
	0.5	-166.0	22.9	156.3	9.3	-322.3	13.8	136	1.4	4	0.0	-116.6	40.4	240.6	4.9	-357.2	43.3	144	1.6	6	0.0	55.19	0.59
	1	-184.4	24.5	153.0	10.7	-337.4	14.3	130	4.7	4	0.0	-186.5	45.8	241.5	4.6	-431.4	44.9	142	1.6	6	0.0	59.47	0.32
	2	-263.0	17.5	117.8	6.7	-380.7	11.3	118	4.2	4	0.0	-279.0	22.0	251.9	2.8	-531.9	23.9	140	2.7	6	0.0	67.19	0.18
	3	-352.5	14.7	93.7	2.9	-446.2	12.3	96	1.2	4	0.2	-297.5	63.7	256.8	3.8	-554.3	63.8	128	2.5	6	0.0	64.95	1.54
CVB5-L061815	0.01	-159.5	23.6	174.9	8.2	-334.4	15.8	138	2.1	4	0.1	-86.4	18.3	253.4	1.2	-339.7	17.2	161	2.3	6	0.0	50.43	0.61
	1	-245.0	27.2	134.8	12.5	-379.8	15.2	119	5.3	4	0.0	-176.6	27.7	255.1	1.0	-431.8	26.7	145	3.6	6	0.0	61.40	0.58
	2	-360.3	15.3	92.9	3.1	-453.2	12.3	98	2.0	4	0.0	-294.8	15.7	252.5	0.4	-547.3	15.6	133	2.0	6	0.0	66.78	0.81
	3	-354.5	20.6	97.7	7.1	-452.2	13.6	97	3.1	4	0.1	-304.9	14.1	251.3	1.0	-556.2	13.7	132	2.2	6	0.0	68.45	0.7
CVB1-L071615	0.01	-103.6	20.8	201.4	6.4	-305.0	14.4	150	3.4	4	0.2	-38.7	21.9	237.1	6.2	-275.8	20.5	142	2.3	6	0.1	49.45	1.02
	1	-237.4	29.2	156.2	9.4	-393.6	20.0	127	4.9	4	0.0	-159.7	27.3	227.6	2.0	-387.3	25.6	133	3.3	7	0.2	56.8	1.37
	2	-245.2	26.8	163.0	9.0	-408.2	18.0	125	6.1	4	0.0	-170.3	27.9	226.8	2.0	-397.1	26.7	132	3.4	6	0.2	67.24	0.71
	3	-335.4	12.7	129.1	3.6	-464.6	9.2	104	2.9	4	0.0	-270.2	17.5	221.9	1.6	-492.1	16.6	121	2.4	7	0.0	67.42	0.3
E11	0.01	-125.9	26.5	278.2	2.9	-404.1	23.5	160	2.7	2	0.0	-212.5	51.9	195.9	4.5	-408.4	53.7	162	2.1	4	0.2	43.33	0.26
	1	-203.8	35.3	265.5	7.2	-469.2	28.2	154	2.5	2	0.0	-290.1	64.4	190.2	5.5	-480.3	61.7	147	2.6	4	0.1	57.04	1.28
	2	-341.5	25.0	235.9	4.8	-577.4	20.7	143	0.4	2	0.0	-346.9	31.3	194.2	4.7	-541.1	29.4	139	3.6	4	0.3	65.22	0.80
	3	-372.7	26.5	214.0	5.1	-586.8	22.5	142	0.5	2	0.0	-382.6	30.5	180.4	6.9	-563.0	25.8	136	3.4	4	0.1	62.83	2.03

Figure A.4 – Structural parameters and temperature breakpoints. Summary of the results obtained for the structural parameters and temperature breakpoints of the capsid for several viruses tested at different salt concentration. SE: standard error. The unit of measure for the overall forces and their standard error is [kcal/mol*Å], and for the breakpoints and their standard error is [°C].

respectively). This indicates that a portion of viruses disintegrated upon heating, though a subset of intact viral particles remained clearly visible (Figure A.5). Furthermore, compared to the observed extent of inactivation ($3 \log_{10}$), the decrease in particle numbers was modest. It is therefore reasonable to assume that inactivation is only in part due to capsid disintegration, and also involves conformational rearrangement not visible by the TEM analysis performed.
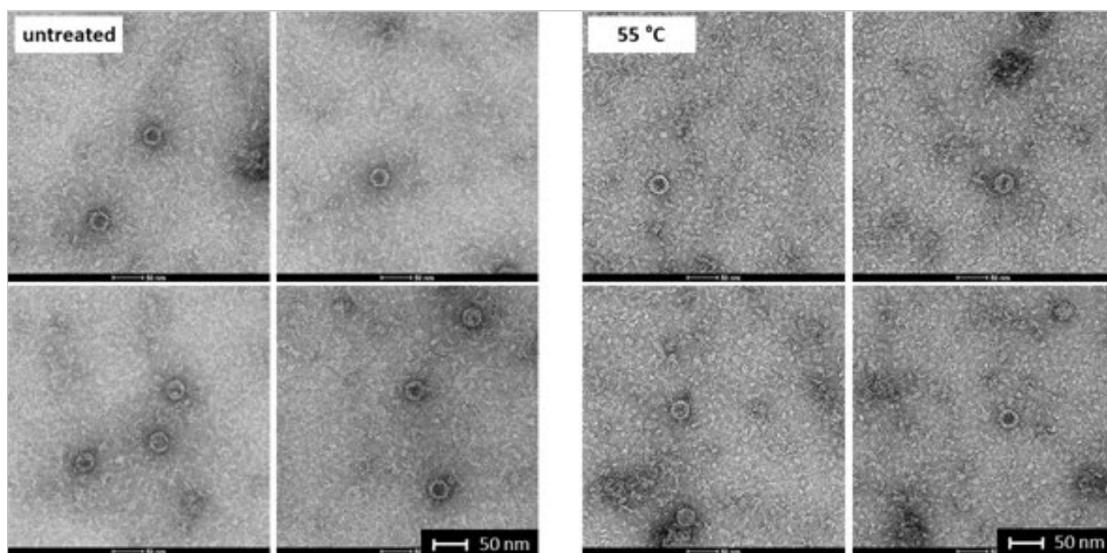
Figure A.5 – TEM images of CVB5-Faulkner prior to and after heating. Samples were held in a 10 mM NaCl PBS matrix. Left: Untreated sample. Right Sample after exposure to 55 °C for one minute. Both samples exhibit intact viral particles, though the overall particle number was about 50% lower in the heat-treated sample.

A stability-enhancing effect of NaCl was observed for all enterovirus strains tested,

though the extent of the enhancement varied between strains. While the breakpoints of E11 and CVB5-Faulkner shifted by approximately 22 °C between 0.01 and 2 M NaCl, the increase was only 16.3 and 17.8 °C for CVB5-L061815 and CVB1-L071615, respectively. At 3 M NaCl, all viruses tested exhibited similar breakpoint shifts of 18 to 20 °C (Figure A.2B and Figure A.4).

Compared to salt, the effect of pH on the breakpoint of CVB5-Faulkner was subtler. At a salt concentration of 10 mM, the breakpoint temperature decreased with increasing pH, with values of 47.2, 45.0, and 40.9 °C for pH 3, 7.4 and 9 respectively (Figure A.6). At all pH values tested, the breakpoint thus remained well below the temperature applied in the thermal inactivation experiment (55 °C). Consequently, changing the pH did not result in a measurable difference in the thermal inactivation kinetics (data not shown). At a NaCl concentration of 1 M, the effects of pH were completely supressed, and no differences in breakpoint were observed over the pH range considered (Figure A.6). The effect of pH on capsid thermostability is not well understood, but may arise from changes in the protonation state of amino acids residues located at the capsid protein interfaces, which in turn modulate the protein interaction forces and hence the thermostability of the capsid. Alternatively, enteroviruses are known to undergo capsid conformational changes as a function of pH[334](29), which appeared to affect the inactivation kinetics under chlorine treatment[335](30). Those conformational changes may also affect the thermostability of the viral capsids. Given the comparatively mild effect of matrix pH on thermostability, however, this solution parameter was not included in our further investigation.

A.4.2 Effect of salt on protein interaction forces

To rationalize the pronounced effects of salt on thermostability, we determined how the forces that act at the protein interfaces are influenced by salt, using the 10 mM NaCl concentration as reference. To this end, we determined the total electrostatic and van der Waals forces at the interface of viral proteins, and we analyzed the number of hydrogen bonds and salt bridges (Figure A.4). Hereby, we focused on the pentamer interface, which previous reports have described as the location of capsid disintegration during thermal inactivation of FMDV[315](10). In addition, we considered the protomer interface, which is affected by conformational rearrangement of poliovirus and EV71[320,321](15, 16) and which is in close vicinity to stability-enhancing
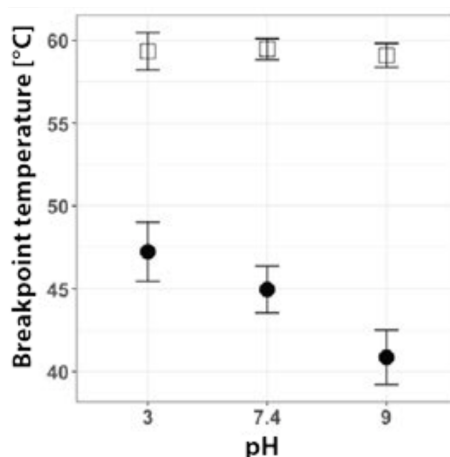
Figure A.6 – Effect of pH on the breakpoint temperature of CVB5-Faulkner. In the presence of 10 mM NaCl (black circles), an increase in pH lowered the breakpoint. In 1 M NaCl (open squares) the breakpoint increase close to 60 °C for all pH conditions considered. The error bars correspond to the 95% confidence intervals.

mutations identified by experimental evolution[325–327] (20–22) (Figure A.1).

The different protein interaction forces of CVB5-Faulkner are shown in Figure A.7. At 10 mM NaCl, the capsid exhibited strong attractive van der Waals forces of -291±9 kcal/(mol*Å) at the pentamer, and -340±65 kcal/(mol*Å) at the protomer interface. Interestingly, however, the net electrostatic interactions at both interfaces were repulsive, with a magnitude of 180±7 and 242±3 kcal/(mol*Å) at the pentamer and the protomer, respectively. The stability of the viral capsid thus arises from the attractive van der Waals forces, which are sufficiently large in magnitude to overcome the destabilizing repulsive forces exerted by the electrostatic interactions at both interfaces. The overall attractive forces at both interfaces are in accordance with the rational fact that the virus needs a stable capsid.

Increasing salt concentrations resulted in a strong increase in van der Waals attraction forces at the pentamer interface of CVB5-Faulkner. Concurrently, the repulsive electrostatic forces decreased, albeit to a lesser extent (Figure A.7A and B). Hence, the presence of salt resulted in an increase of overall attractive forces at this interface (Figure A.7C). The reduced repulsive forces could neither be attributed to a change in the number of salt bridges, which remained stable over all salt concentrations considered, nor to the number of H bonds, which decreased with increasing salt concentration (Figure A.4). Instead, this effect likely stems
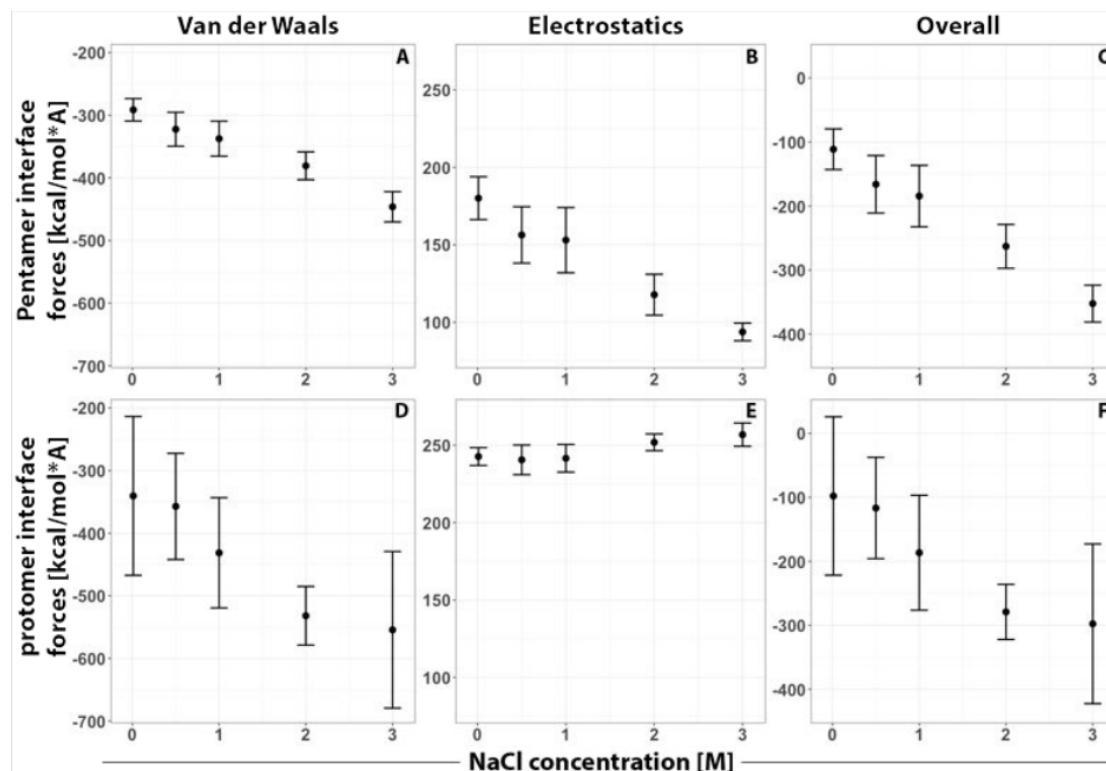
Figure A.7 – Pentamer and protomer interaction forces of CVB5-Faulkner as a function of the salt concentration. (A-C) Pentamer interface interaction forces. (D-F) Protomer interface interaction forces. The error bars correspond to the 95% confidence intervals, based on 7 computational replicas. Exact values are given in Figure A.4.

from the partial screening of the charged residues mediated by NaCl ions, which contributes to reducing the overall weight of the electrostatic interactions. Additionally, the partial neutralization of the charged residues leads to a reorientation of amino acids (in particular of their sidechains) which allows the optimization of the van der Waals contacts and therefore contributes to the reduction of the interaction energy. At the protomer interface, an even greater increase in attractive van der Waals forces with increasing salt concentration was found (Figure A.7D). In contrast, the electrostatic interactions were only minimally affected and remained roughly constant for all salt concentrations considered (Figure A.7E). Overall, the strong increase in van der Waals forces thus resulted in a marked increase of attractive forces at the protomer interface, similar in magnitude to the increase at the pentamer interface (Figure A.7F).

The four viruses tested exhibited some differences in absolute electrostatic and van der

Waals forces (Figure A.4). The largest deviations were found for the pentamer interface of E11, the virus with the least sequence similarity among the viruses tested. At any given salt concentration, E11 exhibited stronger attractive van der Waals forces and repulsive electrostatic forces. Despite these differences, the overall pentamer interaction forces of E11 were equivalent to the other viruses. Furthermore, only minor differences were found among the viruses in the shift in interaction forces when raising the salt concentration from 10 mM to higher values, suggesting comparable salt sensitivities across different strains of the Enterovirus B group.

A.4.3 Protein interaction forces correlate with breakpoint temperature

Across all viruses and salt concentrations considered, the absolute, overall forces at the pentamer interface were strongly correlated with the capsid breakpoint temperatures measured under the corresponding solution conditions (Pearson's $r = 0.87$; Figure A.8A). This correlation indicates that the breakpoint temperature changes by 7.5 °C if the overall pentamer interaction forces change by 100 kcal/(mol*Å). A similar correlation was also found for the protomer interface (Pearson's $r = 0.68$; Figure A.8B). At this interface, the predicted change in breakpoint temperature corresponded to 5.6 °C per 100 kcal/(mol*Å) increase in the protein interaction force. The strength of these correlations further increased if the change in interface forces relative to the values at 10 mM NaCl, rather than the absolute values were considered (Pearson's $r = 0.87$ and 0.92 for the pentamer and protomer interface, respectively). Such correlations may allow to predict how changes in capsid protein interaction forces induced by solution conditions alter the thermostability of an enterovirus. As such, they may serve as tools to guide the design of matrices that preserve or destabilize enteroviruses. Such predictions, however, remain to be validated for additional viruses and solution conditions (e.g., different salts).

Overall, this study demonstrates a major impact of simple matrix variation on virus thermostability, and unravels the underlying changes in protein interaction forces. We demonstrate that even simple matrix modifications, such as changing the pH or the salinity, can have a strong capsid stabilizing effects, and that the main driving force for the stabilization are van der Waals forces at protein interfaces. Finally, our findings were consistent across several Enterovirus strains, indicating that they may also apply to strains of clinical relevance or under vaccine development.

Appendix A. Salt enhances the thermostability of enteroviruses by stabilizing capsid protein interfaces
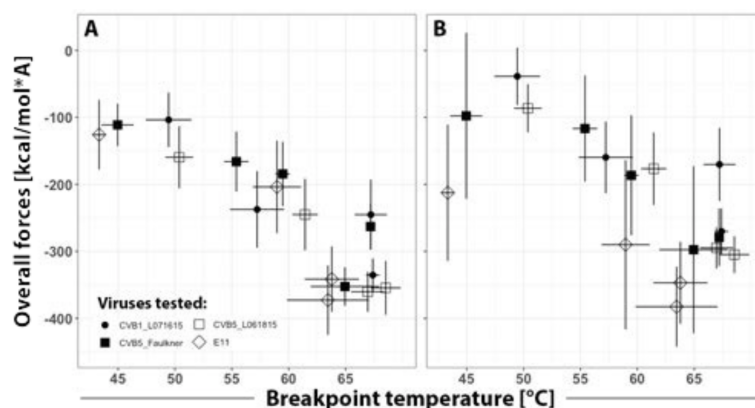
Figure A.8 – Correlation between enterovirus breakpoint temperature and protein interaction forces. Breakpoint temperatures were measured over a NaCl concentration ranging from 0 to 3 M (with each point on the graph representing a different salt concentration), and were correlated against the overall interaction forces at (A) the pentamer interface (Pearson's *r*=-0.87) and (B) the protomer interface (Pearson's *r*=-0.68). Interaction forces correspond to those shown in Figures A.6C and A.6F. The error bars correspond to the 95% confidence interval. The different symbols represent the different viruses tested.

A.5   Materials and Methods

Chemicals. Na2HPO4 (99.0%) was obtained from Fluka (Honeywell International Inc.), Glycine (puriss.) from BioRad (Hercules, USA), HCl (1N) and NaCl (99.5%) from Acros Organics (Geel, Belgium). Cell culture media consisted of modified Eagle medium supplemented with 1% penicillin-streptomycin per ml and 10% (growth medium) or 2% (maintenance medium) heat-inactivated fetal bovine serum (all purchased from Gibco, Frederick, MD).

Cells and viruses. CVB5 Faulkner strain (ATCC VR-185) and echovirus 11 Gregory strain (ATCC VR-41) were obtained from LGC Standards (Molsheim, France). Virus environmental isolates (CVB5- L061815 and CVB1-L071615) were isolated from untreated domestic sewage, as described elsewhere[333](28). Buffalo Green Monkey Kidney (BGMK) cells were kindly provided by the Spiez Laboratory (Switzerland). Virus stock solutions were prepared by amplification in BGMK cells and were purified by PEG precipitation as described previously[333](28), and were stored in phosphate-buffered saline (PBS; 5 mM Na2HPO4, 10 mM NaCl, pH7.4). Virus infectivity was determined by endpoint dilution with Most Probable Number (MPN) statistics[336](31) as described previously[333](28). In brief, virus samples were inoculated on confluent BGMK cells on 96-well plates, with five replicates and eight dilutions for each experimental sample.

The cytopathic effect (CPE) was determined through microscopy five days post-infection and incubation at 37°C with 5% CO_2. The infectivity was then reported as most probable number of cytopathic units per mL (MPNCU mL^{-1}).

Thermal inactivation experiments. Thermal inactivation experiments were performed in aqueous buffers with different pH and ionic strengths. Experiments were conducted in 25 mM glycine-HCl (pH3), in 5 mM Na_2HPO_4 (pH7.4), and in 25 mM glycine-NaOH (pH9). Different buffer compositions had to be used to optimize pH stability at each targeted pH value. The ionic strength was varied by addition of 10 mM to 1 M of NaCl. Kinetic experiments were conducted in a PCR thermocycler (Applied Biosystems, GeneAmp PCR system 9700). 10 μl of virus stock were spiked into PCR tubes containing 90 μl of pre-heated buffer to reach an initial concentration around 10^7 MPNCU mL^{-1}. Samples were maintained at 55°C for varying amounts of time ranging from 0 to 2 minutes, and were then quickly cooled down by placing them on an aluminium PCR cooling block on ice. The residual infectious virus concentration in each sample was enumerated on the same day. Each sample was then mixed with 900 μl of cell culture medium and stored at -20°C prior to enumeration by the MPN assay described above. Each experiment was conducted at least in duplicate.

Determination of the capsid breakpoint temperature. The breakpoint temperature, or melting temperature, of the viral capsids was determined by a thermal-shift assay. The assay was performed in an PCR thermocycler as described above, but each PCR tube was held at a different temperature ranging from 25 to 70°C at 2 degree intervals, and was incubated for one minute at each temperature. The breakpoint was identified as the temperature at which the inactivation rate shifted from slow to rapid. It was determined by fitting a segmental linear regression to a plot of $\ln(N/N_0)$ versus temperature applied, where N/N_0 denotes the residual fraction of infectious virus titers after the one min inactivation period (Figure A.9). The breakpoint temperature, reflected by the intersection of the two linear regression lines, were determined together with their respective standard errors. Each breakpoint was derived based on at least two pooled experimental replicates.

TEM analysis. To investigate if capsid disintegration occurs upon heating, we evaluated the number of intact CVB5 particles after exposed to different heat treatments. A PEG-purified

Appendix A. Salt enhances the thermostability of enteroviruses by stabilizing capsid protein interfaces
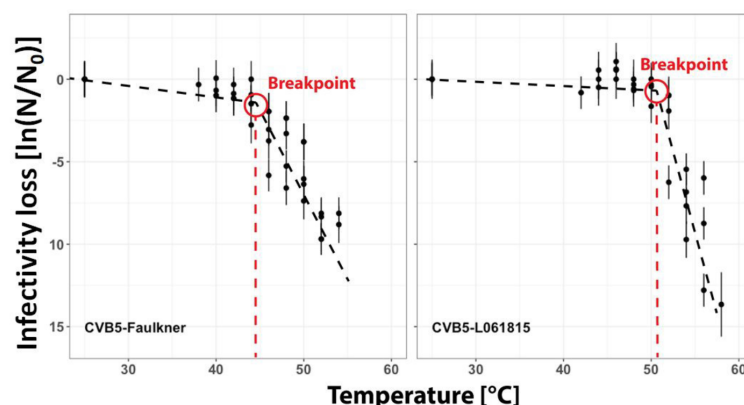
Figure A.9 – Determination of the breakpoint temperature. The figure shows the examples of CVB5-Faulkner and CVB5-L061815 at a NaCl concentration of 0.01 M. The ln of the residual infectivity (N/N$_0$) after 1 minute of treatment is plotted as a function of treatment temperature. The data are fitted to a segmental linear regression (black lines). The intersection of the two linear portions defines the breakpoint temperature of each virus.

solution of CVB5 ($\sim 10^8$ MPNCU mL^{-1} in PBS) was split into two aliquots of 100 μL. One aliquot was held at room temperature, and one was exposed to 55 °C for one minute. For TEM analysis, each sample was adsorbed on a glow-discharged carbon-coated copper grid (400 mesh; EMS, Hatfield, PA, USA) washed with deionized water and stained with 0.75% uranyl formate. Observations were made using an F20 electron microscope (Thermo Fisher, Hillsboro, USA) operated at 200 kV. Digital images were collected using a Falcon III direct detector camera (4098 x 4098 pixels; Thermo Fisher, Hillsboro, USA) at a magnification of 29000x (pixel size = 0.35 nm) and 100000x (pixel size = 0.10 nm), with a defocus range between -1.5 μm and -2.5 μm. For each sample, 16 images were acquired, which jointly covered 30 μm of the grid. Intact viral particles were enumerated visually.

Molecular modelling. 3D capsid models were built using the known crystal structure of coxsackievirus B3 as a scaffold (PDB accession code: 1cov)[337](32). Modeller[338](33) was used to obtain reliable 3D capsid structures (by homology modeling) of strains for which a crystal structure is not currently available in literature. After obtaining the basic unit for each strain, we used the Visual Molecular Dynamics (VMD) software[339](34) to recreate the pentamer, by alignment with the available crystal structure. We also added a portion of the neighbouring pentamers, in order to obtain a full model of the pentamer-pentamer interface (Figure A.1). The protonation states of the amino acids that compose the proteins were assigned with

PROPKA[340](35). All the models were solvated using a TIP3P water model[341](36), and ionized. In order to assess the influence of salt concentration on the structural properties of each strain, we tested different NaCl concentrations (10 mM, 500 mM, 1 M, 2 M, 3 M). The structures were minimized with 1500 steps of conjugate gradient algorithm, in order to remove eventual clashes encountered during the homology modeling and aligning procedures. The presence of hydrogen bonds within the different models was detected using 3.0 Å and 20 degrees as distance and angle threshold; similarly, the salt bridges were assigned based on the distance between the deprotonated oxygen of acidic residues, and the protonated nitrogen of basic ones. Specifically, we assumed the presence of a salt bridge when this distance was below a threshold of 3.2 Å. In both cases, we examined both the total number of bonds within a pentamer, and within two specific interfaces, i.e. the pentamer interface and the protomer interface (Figure A.1). Hydrogen bonds and salt bridges are considered to be part of an interface when the two residues that compose the bond are not part of the same domain (pentamer or protomer), but belong to two separate domains. The NAMD molecular dynamics software engine[342](37) was used to calculate the van der Waals (VdW) and electrostatic forces (ELEC) acting at these interfaces, with the unit of kcal/mol*Å. These forces were decomposed along three main axes: x and z for shearing forces parallel to the interface, and y for the perpendicular forces (Figure A.1). The overall interface force was calculated by calculating the sum of VdWy and ELECy force vectors. These forces were determined separately for the pentamer and protomer interfaces. The calculations for the interaction forces were performed in 7 replicas for each viral strain. For each replica, ions were randomly displaced within the solvent. Afterwards, the system was equilibrated through the minimization procedure described above. After the minimization, interaction forces were calculated for each replica, and the aberrant outlier values were removed ($< \pm 1000$ kcal/mol*Å).

Data analysis. Data handling, MPN calculation and statistical analysis were performed in R[343](38). The following CRAN packages were used: ggplot2[344](39), gridExtra[345](40) and segmented[346](41). The breakpoint temperature was calculated using the segmented function from the segmented package using the linear model function (lm) as object.

A | Molecular basis for recognition of Listeria cell wall teichoic acid by a bacteriophage endolysin

This appendix is adapted from the following paper submitted for publication:

"Molecular basis for recognition of Listeria cell wall teichoic acid by the pseudo-symmetric SH3b-like repeats of a bacteriophage endolysin." Shen Y, Kalograiaki I, **Prunotto A**, Dunne M, Boulos S, Taylor NMI, Sumrall E, Eugster MR, Martin R, Julian-Rodero A, Gerber B, Leiman PG, Menéndez M, Dal Peraro M, Canada FJ, Loessner MJ.

I will here briefly discuss my contribution to this work, in particular the identification of a binding pocket at the cell wall-binding domain of a peptidoglycan hydrolase (validated by mutagenesis experiments) and the evaluation of its dynamic stability.

Endolysins are bacteriophage-encoded peptidoglycan hydrolases targeting the cell wall of host bacteria via their cell wall-binding domains (CBDs). The molecular basis for selective recognition of surface carbohydrate ligands by CBDs remains elusive. Here, we characterize the interaction between the Listeria phage endolysin domain CBD500 and its cell wall teichoic acid (WTA) ligands. We show that 3´O-acetylated GlcNAc residues integrated into the WTA polymer chain are the key epitope recognized by a CBD binding cavity located at the interface

of tandem copies of beta-barrel, pseudo-symmetric SH3b-like repeats. This cavity consists of multiple aromatic residues making extensive interactions with two acetyl groups via hydrogen bonds and van der Waals contacts, while permitting the docking of the diastereomorphic ligands. The multidisciplinary approach described here delineates a previously unknown recognition mechanism by which a phage endolysin specifically recognizes and targets WTA, suggesting an adaptable model for regulation of endolysin specificity.

A.1   Results

To precisely identify the binding site of the WTA, its repeating unit was docked onto the crystal structure of CBD. Three putative cavities with hydrophobic patches were identified by a blind docking procedure carried out with AutoDock Vina[289] using a grid box comprising the whole CBD surface (Figure A.1a). Collaborators performed a panel of alanine-scanning mutants within residues of these three cavities, and tagged them with GFP, to enable quantitative determination of their relative binding capacity to the WTA by fluorescence spectroscopy. Changing of W198, W242 and N248 to alanine almost completely abolished the binding, closely followed by W254A, F175A and G250A mutants (Figure A.1b). Other mutants showing a significant decrease in bacterial surface binding were Y197A, Y211A and S213A, thus inferring a contribution of the mutated residues to either conservation of the protein conformation or complex formation. All affected residues are within the cavity 1 (Figure A.1c), confirming the output prediction from the docking study.

This docking output was used as input for MD simulations. Throughout the simulation the monomeric ligand remained strongly bound at the interface of the RT loops, while the loops and the overall protein structure did not suffer notable distortions (Figure A.2), therefore advocating for the establishment of a single WTA repeating unit as the minimal epitope recognized by CBD500.
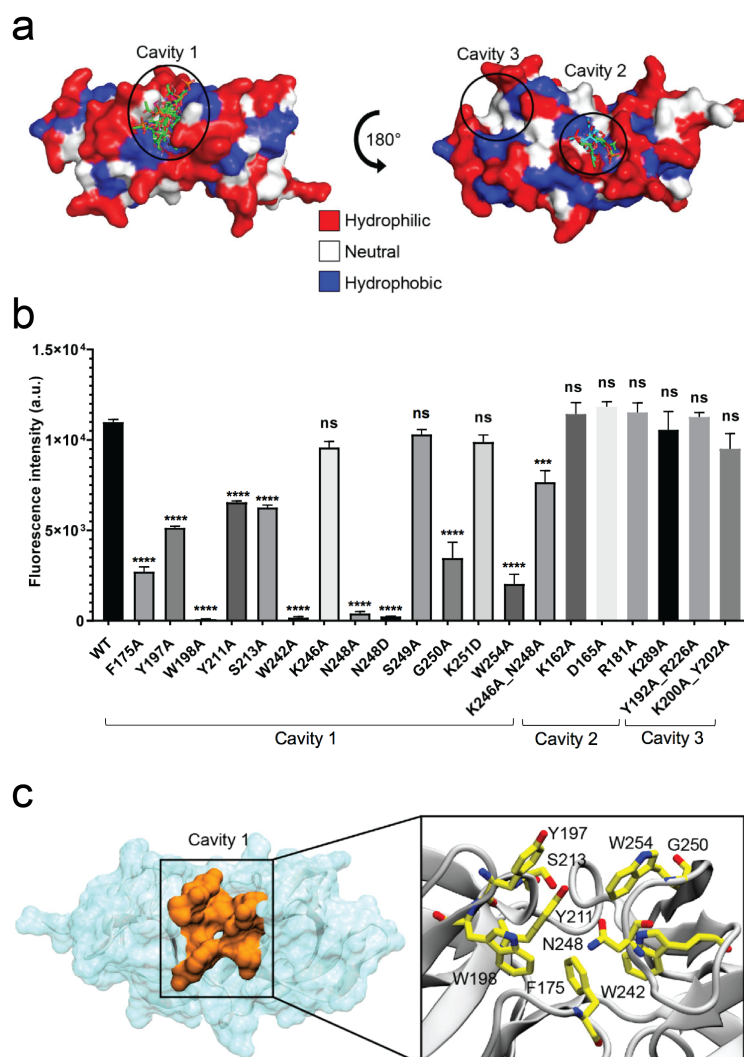
Figure A.1 – Identification of the ligand binding cavity by blind docking and site-directed mutagenesis. (a) Putative binding cavities on the surface of CBD predicted from blind docking analysis. All ligand poses identified through blind docking are represented in sticks. The CBD surface is colored by residue hydrophobicity: hydrophilic (red), neutral (white), and hydrophobic (blue) regions. (b) Relative binding of GFP-CBD mutants to Listeria cells compared to WT, quantified by fluorescence spectroscopy. Results are the mean ± SEM of three experiments with different protein batches. Statistical significance was assessed using one-way ANOVA test (***p<0.001; ****p<0.0001; ns: not significant). (c) Close-up view of the CBD binding cavity 1 (orange). The enlarged view of the square region, shown on the left, depicts the residues (stick representation) shown to be important for binding of CBD to Listeria surfaces, based on mutational studies.
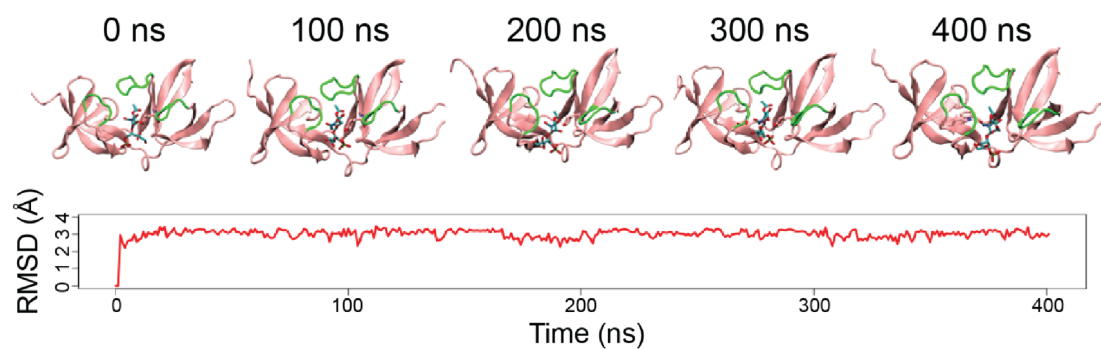
Figure A.2 – Time evolution of the complex of CBD with the WTA repeating unit by molecular dynamics simulations. Representative time-frames (top) and root mean square deviation of the monomeric ligand (bottom). The WTA monomer is represented in sticks, and the CBD loops that conform the binding pocket are colored green.

A Extensive tissue-specific expression variation and novel regulators underlying circadian behavior

This appendix is adapted from the following paper submitted for publication:

"Extensive tissue-specific expression variation and novel regulators underlying circadian behavior." Litovchenko M, Meireles-Filho ACA, Frochaux MV, Bevers RPJ, **Prunotto A**, Anduaga A, Hollis B, Gardeux V, Braman VS, Russeil J, Kadener S, Dal Peraro M, Deplancke B.

I will here briefly discuss my contribution to this work, in particular the evaluation of the influence of specific mutations on the structure of dCRY protein, by means of molecular modeling.

Natural genetic variation affects locomotor activity rhythms across the evolutionary tree, but the underlying molecular mechanisms are poorly understood. Collaborators investigated population-level, molecular circadian clock variation by generating >700 tissue-specific transcriptomes of Drosophila melanogaster (w1118) and 141 Drosophila Genetic Reference Panel (DGRP) lines. Using this comprehensive circadian gene expression atlas, they identified >1,700 cycling genes including putatively new central circadian clock components. They found that >30% of DGRP lines exhibit aberrant circadian gene expression, revealing abundant genetic variation-mediated, inter-tissue circadian expression desynchrony. Genetic analysis of one line with the strongest deviating circadian expression revealed a novel mutation of the Cry protein that, as shown by protein structural modeling and brain immunohistochemistry, disrupts

the light-driven FAD cofactor photoreduction, providing in vivo support for the importance of this conserved photo-entrainment mechanism. Together, the study revealed pervasive tissue-specific circadian expression variation with genetic variants acting upon tissue-specific regulatory networks to generate local gene expression oscillations.

A.1   Results

Sequence analyses of the Cry protein showed a 6 base pair deletion that was associated to a disrupted light input/response pathway. This mutation removes the Met421 and the Trp422 residues, and changes the Val423 to an Ile (V423I). In addition, an extra mutation changes Ser424 to a Pro (S424P). The importance of these residues for Cry function is unclear as they are located between the chain of conserved tryptophan (Trp) residues, which mediates the photoinduced electron transfer activation, and the C-terminal lid domain that changes its conformation to bind the Tim protein.[347] To assess the impact of these amino acid alterations on Cry activity, we generated an atomic 3D model of the mutated (from now on called by the name of DGRP-796) Cry protein by homology modeling based on the published Cry wild type (WT) structure (PDB code: 4JZY[348]) and used molecular dynamics simulations to identify potential conformational differences between the WT and DGRP-796 Cry proteins. Intriguingly, we found that, even though the global structure and dynamic behavior of WT and DGRP-796 Cry proteins are almost identical (RMSD = 1.9 $\pm$ 0.1 Å, Figure A.1A/C), the DGRP-796 cry mutation locally disrupts the secondary structure of the Asp410-Arg430 -helix. Most importantly, our analyses revealed that it causes a dramatic re-orientation of Trp420, which increases the putative distance between Trp420 and the FAD molecule and disrupts the alignment with the adjacent Trp residues (Figure A.1B). This likely impedes the photoactivatable electron transfer chain mechanism necessary for Cry photoactivation,[349] suggesting that the Cry mutation found in DGRP-796, which was initially perceived as being benign, might nevertheless generate a loss-of-function Cry. Moreover, we used molecular docking to evaluate whether FAD binding on the Cry pocket was affected by these mutations. While we were able to recapitulate the crystallographic binding mode of FAD for Cry WT (RMSD = 0.7 Å, Figure A.1D), FAD could not be accommodated in the same binding conformation in the DGRP-796 Cry pocket (RMSD = 7.5 Å, Figure A.1E). Collectively, these modelling and simulation results

suggest that the 6 base pairs deletion found in the DGRP-796 cry allele impairs both the ability to productively bind FAD and the optimal conformation required for the photoactivatable electron transfer cascade mediated by Trp residues.
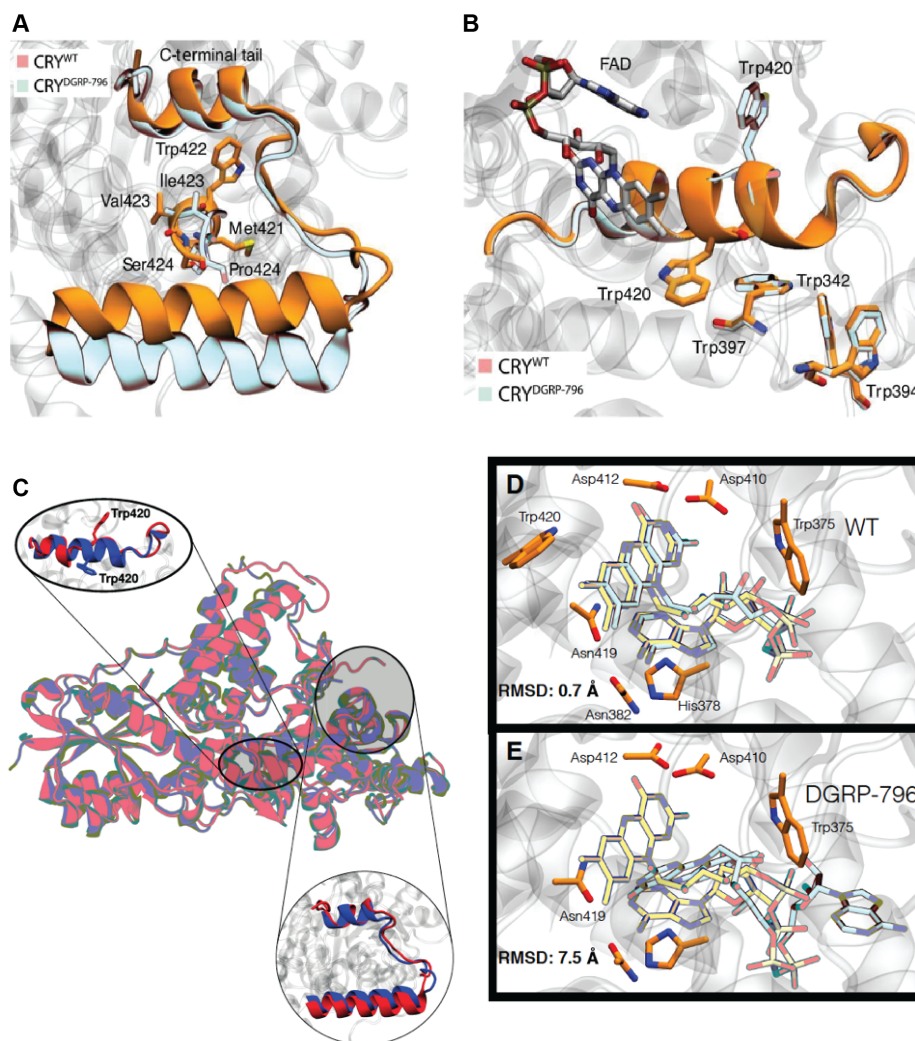
Figure A.1 – (A) WT (yellow) and DGRP-796 (light blue) Cry protein structure models. (B) Enlarged w1118 (yellow) and DGRP-796 (light blue) Cry protein structure models in the Asp410-Arg430 region. The DGRP-796 cry mutation disrupts the secondary structure of the alpha-helix and causes a re-orientation of Trp420. (C) Superimposition between the structures of WT Cry (PDB code: 4JZY, in blue) and DGRP-796 Cry as obtained from homology modeling (in red). The two details represent the Asp410-Arg430 helix and the C-terminal tail, respectively. (D) and (E) results from the docking analysis of the FAD molecule for wild type (D) and DGRP-796 Cry (E) The co-crystallized FAD molecule (PDB code: 4JZY) is in yellow, while the pose of the same molecule as obtained from the docking analysis is represented in cyan. The residues that coordinate the ligand are shown in orange.

133

Bibliography

[1] D. Hutmacher and W. Chrzanowski, *Biointerfaces: Where Material Meets Biology.* No. 10, Royal Society of Chemistry, 2014.

[2] H. Shin, S. Jo, and A. G. Mikos, "Biomimetic materials for tissue engineering," *Biomaterials*, vol. 24, no. 24, pp. 4353–4364, 2003.

[3] J. Patterson, M. M. Martino, and J. A. Hubbell, "Biomimetic materials in tissue engineering," *Materials today*, vol. 13, no. 1-2, pp. 14–22, 2010.

[4] D. S. Dimitrov and R. K. Jain, "Membrane stability," *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes*, vol. 779, no. 4, pp. 437–468, 1984.

[5] J. K. RAISON, "Membrane lipids: structure and function," in *Lipids: Structure and Function*, pp. 57–83, Elsevier, 1980.

[6] A. Goldup, S. Ohki, and J. Danielli, "Recent progress in surface science," *JF Danielli*, 1970.

[7] S. Youssefian, N. Rahbar, C. R. Lambert, and S. Van Dessel, "Variation of thermal conductivity of dppc lipid bilayer membranes around the phase transition temperature," *Journal of the Royal Society Interface*, vol. 14, no. 130, p. 20170127, 2017.

[8] A. Hammoud, E. Baumann, I. Myers, and E. Overton, "Electrical properties of teflon and ceramic capacitors at high temperatures," in *Conference Record of the 1992 IEEE International Symposium on Electrical Insulation*, pp. 487–490, IEEE, 1992.

[9] A. Lamy-Mendes, R. F. Silva, and L. Durães, "Advances in carbon nanostructure–silica aerogel composites: a review," *Journal of Materials Chemistry A*, vol. 6, no. 4, pp. 1340–1369, 2018.

Bibliography

[10] M. Bloom and O. G. Mouritsen, "The evolution of membranes," *Canadian journal of chemistry*, vol. 66, no. 4, pp. 706–712, 1988.

[11] A. Y. Mulkidjanian, M. Y. Galperin, and E. V. Koonin, "Co-evolution of primordial membranes and membrane proteins," *Trends in biochemical sciences*, vol. 34, no. 4, pp. 206–215, 2009.

[12] G. Jékely, "Origin of eukaryotic endomembranes: a critical evaluation of different model scenarios," in *Eukaryotic Membranes and Cytoskeleton*, pp. 38–51, Springer, 2007.

[13] G. J. Retallack, E. S. Krull, G. D. Thackray, and D. Parkinson, "Problematic urn-shaped fossils from a paleoproterozoic (2.2 ga) paleosol in south africa," *Precambrian Research*, vol. 235, pp. 71–87, 2013.

[14] T. Cavalier-Smith, "The phagotrophic origin of eukaryotes and phylogenetic classification of protozoa.," *International journal of systematic and evolutionary microbiology*, vol. 52, no. 2, pp. 297–354, 2002.

[15] J. E. Vance and D. E. Vance, *Biochemistry of lipids, lipoproteins and membranes.* Elsevier, 2008.

[16] O. Quehenberger, A. M. Armando, A. H. Brown, S. B. Milne, D. S. Myers, A. H. Merrill, S. Bandyopadhyay, K. N. Jones, S. Kelly, R. L. Shaner, C. M. Sullards, E. Wang, R. C. Murphy, R. M. Barkley, T. J. Leiker, C. R. H. Raetz, Z. Guan, G. M. Laird, D. A. Six, D. W. Russell, J. G. McDonald, S. Subramaniam, E. Fahy, and E. A. Dennis, "Lipidomics reveals a remarkable diversity of lipids in human plasma„" *Journal of Lipid Research*, vol. 51, pp. 3299–3305, Nov. 2010.

[17] D. Nath, "Membrane biology," 2005.

[18] N. Oppenheimer and H. Diamant, "Correlated diffusion of membrane proteins and their effect on membrane viscosity," *Biophysical journal*, vol. 96, no. 8, pp. 3041–3049, 2009.

[19] A. Engel and H. E. Gaub, "Structure and mechanics of membrane proteins," *Annu. Rev. Biochem.*, vol. 77, pp. 127–148, 2008.

136

[20] L. T. Pearson, J. Edelman, and S. Chan, "Statistical mechanics of lipid membranes. protein correlation functions and lipid ordering," *Biophysical journal*, vol. 45, no. 5, pp. 863–871, 1984.

[21] T. Baumgart, B. R. Capraro, C. Zhu, and S. L. Das, "Thermodynamics and mechanics of membrane curvature generation and sensing by proteins and lipids," *Annual review of physical chemistry*, vol. 62, pp. 483–506, 2011.

[22] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, "Biomembranes: Structural organization and basic functions," in *Molecular Cell Biology. 4th edition*, WH Freeman, 2000.

[23] R. A. Cooper, "Influence of increased membrane cholesterol on membrane fluidity and cell function in human red blood cells," *Journal of supramolecular structure*, vol. 8, no. 4, pp. 413–430, 1978.

[24] S. J. Singer and G. L. Nicolson, "The fluid mosaic model of the structure of cell membranes," *Science*, vol. 175, no. 4023, pp. 720–731, 1972.

[25] M. P. Sheetz, "Cell control by membrane–cytoskeleton adhesion," *Nature Reviews Molecular Cell Biology*, vol. 2, no. 5, pp. 392–396, 2001.

[26] S. B. Gould, "Membranes and evolution," *Current Biology*, vol. 28, no. 8, pp. R381–R385, 2018.

[27] M. Garcia-Gil and E. Albi, "Nuclear lipids in the nervous system: what they do in health and disease," *Neurochemical research*, vol. 42, no. 2, pp. 321–336, 2017.

[28] A. R. Marques and P. Saftig, "Lysosomal storage disorders–challenges, concepts and avenues for therapy: beyond rare diseases," *Journal of cell science*, vol. 132, no. 2, 2019.

[29] A. A. Farooqui, "Lipid mediators in the neural cell nucleus: their metabolism, signaling, and association with neurological disorders," *The Neuroscientist*, vol. 15, no. 4, pp. 392–407, 2009.

[30] M. Radic and D. Pattanaik, "Cellular and molecular mechanisms of anti-phospholipid syndrome," *Frontiers in immunology*, vol. 9, p. 969, 2018.

Bibliography

[31] P. V. Escribá, X. Busquets, J.-i. Inokuchi, G. Balogh, Z. Török, I. Horváth, J. L. Harwood, and L. Vígh, "Membrane lipid therapy: Modulation of the cell membrane composition and structure as a molecular base for drug discovery and new disease treatment," *Progress in lipid research*, vol. 59, pp. 38–53, 2015.

[32] P. V. Escribá, "Membrane-lipid therapy: a new approach in molecular medicine," *Trends in molecular medicine*, vol. 12, no. 1, pp. 34–43, 2006.

[33] J. Lucy, "Globular lipid micelles and cell membranes," *Journal of theoretical biology*, vol. 7, no. 2, pp. 360–373, 1964.

[34] A. Jesorka and O. Orwar, "Liposomes: technologies and analytical applications," *Annu. Rev. Anal. Chem.*, vol. 1, pp. 801–832, 2008.

[35] B. Bingol and M. Sheng, "Deconstruction for reconstruction: the role of proteolysis in neural plasticity and disease," *Neuron*, vol. 69, no. 1, pp. 22–32, 2011.

[36] N. Mizushima, "Autophagy: process and function," *Genes & development*, vol. 21, no. 22, pp. 2861–2873, 2007.

[37] A. Doerr, "Membrane protein structures," *Nature Methods*, vol. 6, no. 1, pp. 35–35, 2009.

[38] J. P. Overington, B. Al-Lazikani, and A. L. Hopkins, "How many drug targets are there?," *Nature reviews Drug discovery*, vol. 5, no. 12, pp. 993–996, 2006.

[39] M. Rask-Andersen, M. S. Almén, and H. B. Schiöth, "Trends in the exploitation of novel drug targets," *Nature reviews Drug discovery*, vol. 10, no. 8, pp. 579–590, 2011.

[40] D. Kozma, I. Simon, and G. E. Tusnady, "Pdbtm: Protein data bank of transmembrane proteins after 8 years," *Nucleic acids research*, vol. 41, no. D1, pp. D524–D529, 2012.

[41] E. P. Carpenter, K. Beis, A. D. Cameron, and S. Iwata, "Overcoming the challenges of membrane protein crystallography," *Current Opinion in Structural Biology*, vol. 18, pp. 581–586, Oct. 2008.

[42] J. T. Marinko, H. Huang, W. D. Penn, J. A. Capra, J. P. Schlebach, and C. R. Sanders, "Folding and misfolding of human membrane proteins in health and disease: from single molecules to cellular proteostasis," *Chemical reviews*, vol. 119, no. 9, pp. 5537–5606, 2019.

[43] G. von Heijne, "Membrane-protein topology," *Nature reviews Molecular cell biology*, vol. 7, no. 12, pp. 909–918, 2006.

[44] A. Whited and A. Johs, "The interactions of peripheral membrane proteins with biological membranes," *Chemistry and physics of lipids*, vol. 192, pp. 51–59, 2015.

[45] T. Szyperski, S. Scheek, J. Johansson, G. Assmann, U. Seedorf, and K. Wüthrich, "Nmr determination of the secondary structure and the three-dimensional polypeptide backbone fold of the human sterol carrier protein 2," *FEBS letters*, vol. 335, no. 1, pp. 18–26, 1993.

[46] C. B. Kallen, J. T. Billheimer, S. A. Summers, S. E. Stayrook, M. Lewis, and J. F. Strauss, "Steroidogenic acute regulatory protein (star) is a sterol transfer protein," *Journal of Biological Chemistry*, vol. 273, no. 41, pp. 26285–26288, 1998.

[47] C. S. Rao, X. Lin, H. M. Pike, J. G. Molotkovsky, and R. E. Brown, "Glycolipid transfer protein mediated transfer of glycosphingolipids between membranes: a model for action based on kinetic and thermodynamic analyses," *Biochemistry*, vol. 43, no. 43, pp. 13805–13815, 2004.

[48] M. Schwab, ed., *Peripheral Membrane Proteins*, pp. 2816–2816. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[49] V. Monje-Galvan and J. B. Klauda, "Peripheral membrane proteins: Tying the knot between experiment and computation," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1858, pp. 1584–1593, July 2016.

[50] M. H. Gelb, "Methods in enzymology enzymology at the membrane interface: Interfacial enzymology and protein-membrane binding preface," 2017.

[51] M. A. Lomize, A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, "OPM: Orientations of Proteins in Membranes database," *Bioinformatics*, vol. 22, pp. 623–625, Mar. 2006.

[52] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "The Lipid Bilayer," *Molecular Biology of the Cell. 4th edition*, 2002.

[53] E. Fuglebakk and N. Reuter, "A model for hydrophobic protrusions on peripheral membrane proteins," *PLOS Computational Biology*, vol. 14, p. e1006325, July 2018.

Bibliography

[54] Y. Cheng, "Membrane protein structural biology in the era of single particle cryo-EM," *Current Opinion in Structural Biology*, vol. 52, pp. 58–63, Sept. 2018.

[55] K. Murata and M. Wolf, "Cryo-electron microscopy for structural analysis of dynamic biological macromolecules," *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1862, no. 2, pp. 324–334, 2018.

[56] K. Yang and X. Han, "Lipidomics: techniques, applications, and outcomes related to biomedical sciences," *Trends in biochemical sciences*, vol. 41, no. 11, pp. 954–969, 2016.

[57] J. R. Bolla, M. T. Agasid, S. Mehmood, and C. V. Robinson, "Membrane protein–lipid interactions probed using mass spectrometry," *Annual review of biochemistry*, vol. 88, pp. 85–111, 2019.

[58] N. Buzhynskyy, L.-N. Liu, I. Casuso, and S. Scheuring, "High-resolution atomic force microscopy of native membranes," *Life at the Nanoscale: Atomic Force Microscopy of Live Cells*, p. 21, 2019.

[59] J. Frank, "Advances in the field of single-particle cryo-electron microscopy over the last decade," *Nature Protocols*, vol. 12, p. 209, Jan. 2017.

[60] H.-W. Wang and J.-W. Wang, "How cryo-electron microscopy and X-ray crystallography complement each other: Cryo-EM and X-Ray Crystallography Complement Each Other," *Protein Science*, vol. 26, pp. 32–39, Jan. 2017.

[61] A. Merk, A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M. I. Davis, R. Pragani, M. B. Boxer, L. A. Earl, J. L. Milne, and S. Subramaniam, "Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery," *Cell*, vol. 165, pp. 1698–1707, June 2016.

[62] J.-P. Renaud, A. Chari, C. Ciferri, W.-t. Liu, H.-W. Rémigy, H. Stark, and C. Wiesmann, "Cryo-EM in drug discovery: achievements, limitations and prospects," *Nature Reviews Drug Discovery*, vol. 17, pp. 471–492, June 2018.

[63] M. S. Hunter, D. P. DePonte, D. Shapiro, R. A. Kirian, X. Wang, D. Starodub, S. Marchesini, U. Weierstall, R. B. Doak, J. C. Spence, *et al.*, "X-ray diffraction from membrane protein nanocrystals," *Biophysical journal*, vol. 100, no. 1, pp. 198–206, 2011.

[64] J. Frank, "New opportunities created by single-particle cryo-em: the mapping of conformational space," 2018.

[65] L. A. Abriata and M. Dal Peraro, "Will cryo-electron microscopy shift the current paradigm in protein structure prediction?," *Journal of Chemical Information and Modeling,* 2020.

[66] R. Phillips, T. Ursell, P. Wiggins, and P. Sens, "Emerging roles for lipids in shaping membrane-protein function," *Nature*, vol. 459, pp. 379–385, May 2009.

[67] A. Laganowsky, E. Reading, T. M. Allison, M. B. Ulmschneider, M. T. Degiacomi, A. J. Baldwin, and C. V. Robinson, "Membrane proteins bind lipids selectively to modulate their structure and function," *Nature*, vol. 510, p. 172, June 2014.

[68] V. Corradi, E. Mendez-Villuendas, H. I. Ingólfsson, R.-X. Gu, I. Siuda, M. N. Melo, A. Moussatova, L. J. DeGagné, B. I. Sejdiu, G. Singh, T. A. Wassenaar, K. Delgado Magnero, S. J. Marrink, and D. P. Tieleman, "Lipid–Protein Interactions Are Unique Fingerprints for Membrane Proteins," *ACS Central Science*, vol. 4, pp. 709–717, June 2018.

[69] K. Yang and X. Han, "Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences," *Trends in Biochemical Sciences*, vol. 41, pp. 954–969, Nov. 2016.

[70] T. Uchihashi, H. Watanabe, S. Fukuda, M. Shibata, and T. Ando, "Functional extension of high-speed AFM for wider biological applications," *Ultramicroscopy*, vol. 160, pp. 182–196, Jan. 2016.

[71] T. Ando, "High-speed atomic force microscopy and its future prospects," *Biophysical Reviews*, vol. 10, pp. 285–292, Apr. 2018.

[72] M. Braitbard, D. Schneidman-Duhovny, and N. Kalisman, "Integrative structure modeling: overview and assessment," *Annual review of biochemistry*, vol. 88, pp. 113–135, 2019.

[73] W. Wang, O. Donini, C. M. Reyes, and P. A. Kollman, "Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions," *Annual review of biophysics and biomolecular structure*, vol. 30, no. 1, pp. 211–243, 2001.

[74] H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole, and S. J. Marrink, "The power of coarse graining in biomolecular simulations," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, no. 3, pp. 225–248, 2014.

[75] A. Patrascioiu, "The ergodic hypothesis," *Proceedings of Los Almos Science Special issue*, pp. 263–279, 1987.

[76] R. W. Pastor and A. D. MacKerell, "Development of the CHARMM Force Field for Lipids," *The Journal of Physical Chemistry Letters*, vol. 2, pp. 1526–1532, July 2011.

[77] A. Skjevik, B. D. Madej, R. C. Walker, and K. Teigen, "LIPID11: A Modular Framework for Lipid Simulations Using Amber," *The Journal of Physical Chemistry B*, vol. 116, pp. 11124–11136, Sept. 2012.

[78] C. J. Dickson, B. D. Madej, A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould, and R. C. Walker, "Lipid14: The Amber Lipid Force Field," *Journal of Chemical Theory and Computation*, vol. 10, pp. 865–879, Feb. 2014.

[79] I. Chandrasekhar, M. Kastenholz, R. D. Lins, C. Oostenbrink, L. D. Schuler, D. P. Tieleman, and W. F. van Gunsteren, "A consistent potential energy parameter set for lipids: dipalmitoylphosphatidylcholine as a benchmark of the GROMOS96 45A3 force field," *European biophysics journal: EBJ*, vol. 32, pp. 67–77, Mar. 2003.

[80] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations," *The Journal of Physical Chemistry B*, vol. 111, pp. 7812–7824, July 2007.

[81] S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: A web-based graphical user interface for CHARMM," *Journal of Computational Chemistry*, vol. 29, pp. 1859–1865, Aug. 2008.

[82] T. A. Wassenaar, H. I. Ingólfsson, R. A. Böckmann, D. P. Tieleman, and S. J. Marrink, "Computational Lipidomics with *insane* : A Versatile Tool for Generating Custom Membranes for Molecular Simulations," *Journal of Chemical Theory and Computation*, vol. 11, pp. 2144–2155, May 2015.

[83] J. Domański, P. J. Stansfeld, M. S. P. Sansom, and O. Beckstein, "Lipidbook: A Public Repository for Force-Field Parameters Used in Membrane Simulations," *The Journal of Membrane Biology*, vol. 236, pp. 255–258, Aug. 2010.

[84] R. Guixà-González, I. Rodriguez-Espigares, J. M. Ramírez-Anguita, P. Carrió-Gaspar, H. Martinez-Seara, T. Giorgino, and J. Selent, "MEMBPLUGIN: studying membrane complexity in VMD," *Bioinformatics*, vol. 30, pp. 1478–1480, May 2014.

[85] C. J. Knight and J. S. Hub, "MemGen: a general web server for the setup of lipid membrane simulation systems: Fig. 1.," *Bioinformatics*, vol. 31, pp. 2897–2899, Sept. 2015.

[86] C. Bovigny, G. Tamò, T. Lemmin, N. Maïno, and M. Dal Peraro, "LipidBuilder: A Framework To Build Realistic Models for Biological Membranes," *Journal of Chemical Information and Modeling*, vol. 55, pp. 2491–2499, Dec. 2015.

[87] T. Palzkill, "Metallo--lactamase structure and function: Metallo--lactamase structure and function," *Annals of the New York Academy of Sciences*, vol. 1277, pp. 91–104, Jan. 2013.

[88] C. Bebrone, "Metallo--lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily," *Biochemical Pharmacology*, vol. 74, pp. 1686–1701, Dec. 2007.

[89] K. L. Scott, O. Kabbarah, M.-C. Liang, E. Ivanova, V. Anagnostou, J. Wu, S. Dhakal, M. Wu, S. Chen, T. Feinberg, *et al.*, "Golph3 modulates mtor signalling and rapamycin sensitivity in cancer," *Nature*, vol. 459, no. 7250, pp. 1085–1090, 2009.

[90] T. R. Walsh, M. A. Toleman, L. Poirel, and P. Nordmann, "Metallo- -Lactamases: the Quiet before the Storm?," *Clinical Microbiology Reviews*, vol. 18, pp. 306–325, Apr. 2005.

[91] M. W. Crowder, J. Spencer, and A. J. Vila, "Metallo--lactamases: Novel Weaponry for Antibiotic Resistance in Bacteria," *Accounts of Chemical Research*, vol. 39, pp. 721–728, Oct. 2006.

[92] W. d. Boer, L. B. Folman, R. C. Summerbell, and L. Boddy, "Living in a fungal world: impact of fungi on soil bacterial niche development," *FEMS microbiology reviews*, vol. 29, no. 4, pp. 795–811, 2005.

Bibliography

[93] M. Bahram, F. Hildebrand, S. K. Forslund, J. L. Anderson, N. A. Soudzilovskaia, P. M. Bodegom, J. Bengtsson-Palme, S. Anslan, L. P. Coelho, H. Harend, *et al.*, "Structure and function of the global topsoil microbiome," *Nature*, vol. 560, no. 7717, pp. 233–237, 2018.

[94] P. Brian, "Antibiotics produced by fungi," *The Botanical Review*, vol. 17, no. 6, pp. 357–430, 1951.

[95] S. F. Brady, M. M. Wagenaar, M. P. Singh, J. E. Janso, and J. Clardy, "The cytosporones, new octaketide antibiotics isolated from an endophytic fungus," *Organic letters*, vol. 2, no. 25, pp. 4043–4046, 2000.

[96] H. T. Clarke, *Chemistry of penicillin.* Princeton University Press, 2015.

[97] E. Sauvage, F. Kerff, M. Terrak, J. A. Ayala, and P. Charlier, "The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis," *FEMS microbiology reviews*, vol. 32, no. 2, pp. 234–258, 2008.

[98] J. R. Knowles, "Penicillin resistance: the chemistry of. beta.-lactamase inhibition.," *Accounts of Chemical Research*, vol. 18, no. 4, pp. 97–104, 1985.

[99] T. E. Kehl-Fie, S. Chitayat, M. I. Hood, S. Damo, N. Restrepo, C. Garcia, K. A. Munro, W. J. Chazin, and E. P. Skaar, "Nutrient metal sequestration by calprotectin inhibits bacterial superoxide defense, enhancing neutrophil killing of staphylococcus aureus," *Cell host & microbe*, vol. 10, no. 2, pp. 158–164, 2011.

[100] S. M. Damo, T. E. Kehl-Fie, N. Sugitani, M. E. Holt, S. Rathi, W. J. Murphy, Y. Zhang, C. Betz, L. Hench, G. Fritz, *et al.*, "Molecular basis for manganese sequestration by calprotectin and roles in the innate immune response to invading bacterial pathogens," *Proceedings of the National Academy of Sciences*, vol. 110, no. 10, pp. 3841–3846, 2013.

[101] G. M. Gadd, "Heavy metal accumulation by bacteria and other microorganisms," *Experientia*, vol. 46, no. 8, pp. 834–840, 1990.

[102] G. Bahr, L. Vitor-Horen, C. R. Bethel, R. A. Bonomo, L. J. González, and A. J. Vila, "Clinical evolution of new delhi metallo-β-lactamase (ndm) optimizes resistance under zn (ii) deprivation," *Antimicrobial agents and chemotherapy*, vol. 62, no. 1, pp. e01849–17, 2018.

144

[103] S. Borgia, O. Lastovetska, D. Richardson, A. Eshaghi, J. Xiong, C. Chung, M. Baqi, A. McGeer, G. Ricci, R. Sawicki, *et al.*, "Outbreak of carbapenem-resistant enterobacteriaceae containing bla ndm-1, ontario, canada," *Clinical infectious diseases*, vol. 55, no. 11, pp. e109–e117, 2012.

[104] J. A. E. Pérez, N. M. O. Escobar, B. Castro-Cardozo, I. A. V. Márquez, M. I. G. Aguilar, L. M. de la Barrera, E. R. B. Barreto, R. A. Marquez-Ortiz, M. V. M. Guayazán, and N. V. Gómez, "Outbreak of ndm-1-producing klebsiella pneumoniae in a neonatal unit in colombia," *Antimicrobial agents and chemotherapy*, vol. 57, no. 4, pp. 1957–1960, 2013.

[105] E. Voulgari, C. Gartzonika, G. Vrioni, L. Politi, E. Priavali, S. Levidiotou-Stefanou, and A. Tsakris, "The balkan region: Ndm-1-producing klebsiella pneumoniae st11 clonal strain causing outbreaks in greece," *Journal of Antimicrobial Chemotherapy*, vol. 69, no. 8, pp. 2091–2097, 2014.

[106] J. Decousser, C. Jansen, P. Nordmann, A. Emirian, R. Bonnin, L. Anais, J. Merle, and L. Poirel, "Outbreak of ndm-1-producing acinetobacter baumannii in france, january to may 2013," *Eurosurveillance*, vol. 18, no. 31, p. 20547, 2013.

[107] K. Kelland, ""super superbug" ndm-1 spreads in europe," Nov 2010.

[108] J. Gallagher, "Ndm-1 superbug enzyme's 'photofit' taken," Sep 2011.

[109] D. King and N. Strynadka, "Crystal structure of New Delhi metallo--lactamase reveals molecular basis for antibiotic resistance," *Protein Science*, vol. 20, pp. 1484–1491, Sept. 2011.

[110] L. J. González, G. Bahr, T. G. Nakashige, E. M. Nolan, R. A. Bonomo, and A. J. Vila, "Membrane anchoring stabilizes and favors secretion of New Delhi metallo--lactamase," *Nature Chemical Biology*, vol. 12, pp. 516–522, July 2016.

[111] C. S. Wood, K. R. Schmitz, N. J. Bessman, T. G. Setty, K. M. Ferguson, and C. G. Burd, "Ptdins4 p recognition by vps74/golph3 links ptdins 4-kinase signaling to retrograde golgi trafficking," *Journal of Cell Biology*, vol. 187, no. 7, pp. 967–975, 2009.

Bibliography

[112] S. Sechi, A. Frappaolo, G. Belloni, G. Colotti, and M. G. Giansanti, "The multiple cellular functions of the oncoprotein golgi phosphoprotein 3," *Oncotarget*, vol. 6, no. 6, p. 3493, 2015.

[113] S. E. Farber-Katz, H. C. Dippold, M. D. Buschman, M. C. Peterman, M. Xing, C. J. Noakes, J. Tat, M. M. Ng, J. Rahajeng, D. M. Cowan, *et al.*, "Dna damage triggers golgi dispersal via dna-pk and golph3," *Cell*, vol. 156, no. 3, pp. 413–427, 2014.

[114] T. Isaji, S. Im, W. Gu, Y. Wang, Q. Hang, J. Lu, T. Fukuda, N. Hashii, D. Takakura, N. Kawasaki, *et al.*, "An oncogenic protein golgi phosphoprotein 3 up-regulates cell migration via sialylation," *Journal of Biological Chemistry*, vol. 289, no. 30, pp. 20694–20705, 2014.

[115] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D. Parkin, M. Piñeros, A. Znaor, and F. Bray, "Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods," *International journal of cancer*, vol. 144, no. 8, pp. 1941–1953, 2019.

[116] G. P. Guy Jr, K. R. Yabroff, D. U. Ekwueme, A. W. Smith, E. C. Dowling, R. Rechis, S. Nutt, and L. C. Richardson, "Estimating the health and economic burden of cancer among those diagnosed as adolescents and young adults," *Health affairs*, vol. 33, no. 6, pp. 1024–1031, 2014.

[117] S. Eckhouse, G. Lewison, and R. Sullivan, "Trends in the global funding and activity of cancer research," *Molecular oncology*, vol. 2, no. 1, pp. 20–32, 2008.

[118] D. S. Jones, S. H. Podolsky, and J. A. Greene, "The burden of disease and the changing task of medicine," *New England Journal of Medicine*, vol. 366, no. 25, pp. 2333–2338, 2012.

[119] "NO TIME TO WAIT: SECURING THE FUTURE FROM DRUG-RESISTANT INFECTIONS," tech. rep., United Nations, 2019.

[120] R. Kelly, G. Zoubiane, D. Walsh, R. Ward, and H. Goossens, "Public funding for research on antibacterial resistance in the jpiamr countries, the european commission, and related european union agencies: a systematic observational analysis," *The Lancet Infectious Diseases*, vol. 16, no. 4, pp. 431–440, 2016.

[121] E. H. Ackerknecht and L. Haushofer, *A short history of medicine*. JHU Press, 2016.

[122] R. Porter, *The Cambridge illustrated history of medicine*. Cambridge University Press, 2001.

[123] A. Castiglioni, *A history of medicine*, vol. 2. Routledge, 2019.

[124] H. E. Sigerist, *A history of medicine: Early Greek, Hindu, and Persian medicine*, vol. 2. Oxford University Press, 1987.

[125] J. T. Hughes, "The edwin smith surgical papyrus: an analysis of the first case reports of spinal cord injuries," *Spinal Cord*, vol. 26, no. 2, pp. 71–82, 1988.

[126] R. S. Lillie, "The philosophy of biology: Vitalism versus mechanism," *Science*, vol. 40, no. 1041, pp. 840–846, 1914.

[127] P. J. Ramberg, "The death of vitalism and the birth of organic chemistry: Wohler's urea synthesis and the disciplinary identity of organic chemistry," *Ambix*, vol. 47, no. 3, pp. 170–195, 2000.

[128] J. R. Partington, *History of chemistry*. Macmillan International Higher Education, 1964.

[129] M. Planck, "On the law of distribution of energy in the normal spectrum," *Annalen der physik*, vol. 4, no. 553, p. 1, 1901.

[130] G. W. Leibniz and G. W. F. von Leibniz, *Leibniz: New essays on human understanding*. Cambridge University Press, 1996.

[131] M. Rudan, "From classical mechanics to quantum mechanics," in *Physics of Semiconductor Devices*, pp. 143–170, Springer, 2018.

[132] A. Einstein, "Does the inertia of a body depend upon its energy-content," *Ann Phys*, vol. 18, pp. 639–641, 1905.

[133] H. Hertz, "Ueber sehr schnelle electrische schwingungen," *Annalen der Physik*, vol. 267, no. 7, pp. 421–448, 1887.

[134] G. Kirchhoff, "Ueber das verhältniss zwischen dem emissionsvermögen und dem absorptionsvermögen der körper für wärme und licht," *Annalen der Physik*, vol. 185, no. 2, pp. 275–301, 1860.

[135] N. Bohr, "The spectra of helium and hydrogen," *Nature*, vol. 92, no. 2295, pp. 231–232, 1913.

Bibliography

[136] L. De Broglie, *Recherches sur la théorie des quanta.* PhD thesis, Migration-université en cours d'affectation, 1924.

[137] W. Heisenberg, "Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik," in *Original Scientific Papers Wissenschaftliche Originalarbeiten*, pp. 478–504, Springer, 1985.

[138] E. Schrödinger, "Quantisierung als eigenwertproblem," *Annalen der physik*, vol. 385, no. 13, pp. 437–490, 1926.

[139] T. Freeth, Y. Bitsakis, X. Moussas, J. H. Seiradakis, A. Tselikas, H. Mangou, M. Zafeiropoulou, R. Hadland, D. Bate, A. Ramsey, *et al.*, "Decoding the ancient greek astronomical calculator known as the antikythera mechanism," *Nature*, vol. 444, no. 7119, pp. 587–591, 2006.

[140] A. M. Turing, "On computable numbers, with an application to the entscheidungsproblem," *J. of Math*, vol. 58, no. 345-363, p. 5, 1936.

[141] N. Ensmenger, "Computer science: bits of history," *American Scientist*, vol. 91, no. 5, pp. 467–469, 2003.

[142] F. Faggin, M. E. Hoff, S. Mazor, and M. Shima, "The history of the 4004," *Ieee Micro*, vol. 16, no. 6, pp. 10–20, 1996.

[143] G. E. Moore *et al.*, "Cramming more components onto integrated circuits," 1965.

[144] M. Born and R. Oppenheimer, "Zur quantentheorie der molekeln," *Annalen der physik*, vol. 389, no. 20, pp. 457–484, 1927.

[145] N. Metropolis, "The beginning of the," *Los Alamos Science*, vol. 15, pp. 125–30, 1987.

[146] N. Metropolis and S. Ulam, "The monte carlo method," *Journal of the American statistical association*, vol. 44, no. 247, pp. 335–341, 1949.

[147] P. Hill, J. Manners, and J. Petch, "Reducing noise associated with the monte carlo independent column approximation for weather forecasting models," *Quarterly Journal of the Royal Meteorological Society*, vol. 137, no. 654, pp. 219–228, 2011.

[148] E. Veach and L. J. Guibas, "Optimally combining sampling techniques for monte carlo rendering," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 419–428, 1995.

[149] C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt, "Fitting of stochastic telecommunication network models via distance measures and monte–carlo tests," *Telecommunication Systems*, vol. 31, no. 4, pp. 353–377, 2006.

[150] M. Milik and J. Skolnick, "Insertion of peptide chains into lipid membranes: an off-lattice monte carlo dynamics model," *Proteins: Structure, Function, and Bioinformatics*, vol. 15, no. 1, pp. 10–25, 1993.

[151] E. Paquet and H. L. Viktor, "Molecular dynamics, monte carlo simulations, and langevin dynamics: a computational review," *BioMed research international*, vol. 2015, 2015.

[152] R. D. Skeel, "What makes molecular dynamics work?," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 1363–1378, 2009.

[153] H. M. Senn and W. Thiel, "Qm/mm methods for biomolecular systems," *Angewandte Chemie International Edition*, vol. 48, no. 7, pp. 1198–1229, 2009.

[154] B. J. Alder and T. E. Wainwright, "Phase transition for a hard sphere system," *The Journal of chemical physics*, vol. 27, no. 5, pp. 1208–1209, 1957.

[155] Y.-P. Liu, K. Kim, B. Berne, R. A. Friesner, and S. W. Rick, "Constructing ab initio force fields for molecular dynamics simulations," *The Journal of chemical physics*, vol. 108, no. 12, pp. 4739–4755, 1998.

[156] P. Carloni, U. Rothlisberger, and M. Parrinello, "The role and perspective of ab initio molecular dynamics in the study of biological systems," *Accounts of Chemical Research*, vol. 35, no. 6, pp. 455–464, 2002.

[157] P. Xu, E. B. Guidez, C. Bertoni, and M. S. Gordon, "Perspective: Ab initio force field methods derived from quantum mechanics," *The Journal of Chemical Physics*, vol. 148, no. 9, p. 090901, 2018.

Bibliography

[158] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *Journal of the American Chemical Society*, vol. 117, no. 19, pp. 5179–5197, 1995.

[159] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *The journal of physical chemistry B*, vol. 102, no. 18, pp. 3586–3616, 1998.

[160] L. D. Schuler, X. Daura, and W. F. Van Gunsteren, "An improved gromos96 force field for aliphatic hydrocarbons in the condensed phase," *Journal of Computational Chemistry*, vol. 22, no. 11, pp. 1205–1218, 2001.

[161] A. D. MacKerell Jr, "Empirical force fields for biological macromolecules: overview and issues," *Journal of computational chemistry*, vol. 25, no. 13, pp. 1584–1604, 2004.

[162] M. Levitt, "The birth of computational structural biology," *Nature structural biology*, vol. 8, no. 5, pp. 392–393, 2001.

[163] M. Levitt and A. Warshel, "Computer simulation of protein folding," *Nature*, vol. 253, no. 5494, pp. 694–698, 1975.

[164] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," *Journal of molecular biology*, vol. 103, no. 2, pp. 227–249, 1976.

[165] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. a. Swaminathan, and M. Karplus, "Charmm: a program for macromolecular energy, minimization, and dynamics calculations," *Journal of computational chemistry*, vol. 4, no. 2, pp. 187–217, 1983.

[166] T. Darden, D. York, and L. Pedersen, "Particle mesh ewald: An n? log (n) method for ewald sums in large systems," *The Journal of chemical physics*, vol. 98, no. 12, pp. 10089–10092, 1993.

150

[167] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten, "Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions," *Molecular Simulation*, vol. 6, no. 1-3, pp. 121–142, 1991.

[168] L. Verlet, "Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules," *Physical review*, vol. 159, no. 1, p. 98, 1967.

[169] W. F. Van Gunsteren and H. J. Berendsen, "A leap-frog algorithm for stochastic dynamics," *Molecular Simulation*, vol. 1, no. 3, pp. 173–185, 1988.

[170] E. Hairer, C. Lubich, and G. Wanner, "Geometric numerical integration illustrated by the störmer–verlet method," *Acta numerica*, vol. 12, pp. 399–450, 2003.

[171] A. Liwo, S. Ołdziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, "A united-residue force field for off-lattice protein-structure simulations. i. functional forms and parameters of long-range side-chain interaction potentials from protein crystal data," *Journal of computational chemistry*, vol. 18, no. 7, pp. 849–873, 1997.

[172] J. Maupetit, P. Tuffery, and P. Derreumaux, "A coarse-grained protein force field for folding and structure prediction," *Proteins: Structure, Function, and Bioinformatics*, vol. 69, no. 2, pp. 394–408, 2007.

[173] T. Bereau and M. Deserno, "Generic coarse-grained model for protein folding and aggregation," *The Journal of chemical physics*, vol. 130, no. 23, p. 06B621, 2009.

[174] M. Pasi, R. Lavery, and N. Ceres, "Palace: A coarse-grain protein model for studying mechanical properties," *Journal of chemical theory and computation*, vol. 9, no. 1, pp. 785–793, 2013.

[175] M. Zacharias, "Protein–protein docking with a reduced protein model accounting for side-chain flexibility," *Protein Science*, vol. 12, no. 6, pp. 1271–1282, 2003.

[176] X. Periole and S.-J. Marrink, "The martini coarse-grained force field," in *Biomolecular Simulations*, pp. 533–565, Springer, 2013.

Bibliography

[177] X. Periole, M. Cavalli, S.-J. Marrink, and M. A. Ceruso, "Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition," *Journal of chemical theory and computation*, vol. 5, no. 9, pp. 2531–2543, 2009.

[178] S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta, and S. J. Marrink, "Polarizable Water Model for the Coarse-Grained MARTINI Force Field," *PLoS Computational Biology*, vol. 6, p. e1000810, June 2010.

[179] R. Bradley and R. Radhakrishnan, "Coarse-grained models for protein-cell membrane interactions," *Polymers*, vol. 5, no. 3, pp. 890–936, 2013.

[180] C. TE, "Proteins?structures and molecular properties," *Freeman, New York, 1993) p*, vol. 1, 1993.

[181] M. J. Mitchell and J. A. McCammon, "Free energy difference calculations by thermodynamic integration: difficulties in obtaining a precise value," *Journal of computational chemistry*, vol. 12, no. 2, pp. 271–275, 1991.

[182] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method," *Journal of computational chemistry*, vol. 13, no. 8, pp. 1011–1021, 1992.

[183] W. L. Jorgensen and L. L. Thomas, "Perspective on free-energy perturbation calculations for chemical equilibria," *Journal of chemical theory and computation*, vol. 4, no. 6, pp. 869–876, 2008.

[184] R. E. Jones and K. K. Mandadapu, "Adaptive green-kubo estimates of transport coefficients from molecular dynamics based on robust error analysis," *The Journal of chemical physics*, vol. 136, no. 15, p. 154102, 2012.

[185] E. H. Koo, P. T. Lansbury, and J. W. Kelly, "Amyloid diseases: abnormal protein aggregation in neurodegeneration," *Proceedings of the National Academy of Sciences*, vol. 96, no. 18, pp. 9989–9990, 1999.

[186] E. Paci, J. Gsponer, X. Salvatella, and M. Vendruscolo, "Molecular dynamics studies of the process of amyloid aggregation of peptide fragments of transthyretin," *Journal of molecular biology*, vol. 340, no. 3, pp. 555–569, 2004.

[187] M. Kouza, A. Kolinski, I. A. Buhimschi, and A. Kloczkowski, "Explicit-solvent all-atom molecular dynamics of peptide aggregation," in *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes*, pp. 541–558, Springer, 2019.

[188] M. Carballo-Pacheco, A. E. Ismail, and B. Strodel, "On the applicability of force fields to study the aggregation of amyloidogenic peptides using molecular dynamics simulations," *Journal of Chemical Theory and Computation*, vol. 14, no. 11, pp. 6063–6075, 2018.

[189] J.-E. Shea and Z. A. Levine, "Studying the early stages of protein aggregation using replica exchange molecular dynamics simulations," in *Protein Amyloid Aggregation*, pp. 225–250, Springer, 2016.

[190] M. Frieri, K. Kumar, and A. Boutin, "Antibiotic resistance," *Journal of infection and public health*, vol. 10, no. 4, pp. 369–378, 2017.

[191] C. for Disease Control and P. (US), *Antibiotic resistance threats in the United States, 2019*. Centres for Disease Control and Prevention, US Department of Health and ?, 2019.

[192] M. Roser, E. Ortiz-Ospina, and H. Ritchie, "Life expectancy," *Our World in Data*, 2013.

[193] A. Stewardson and D. Pittet, "Ignác semmelweis?celebrating a flawed pioneer of patient safety," *The Lancet*, vol. 378, no. 9785, pp. 22–23, 2011.

[194] G. A. Silver, "Virchow, the heroic model in medicine: health policy by accolade.," *American journal of public health*, vol. 77, no. 1, pp. 82–88, 1987.

[195] K. C. Carter and B. R. Carter, *Childbed fever: a scientific biography of Ignaz Semmelweis*. Routledge, 2017.

[196] L. Pasteur, J. Joubert, and C. Chamberland, "The germ theory of disease," *CR Hebd Seances Acad Sci*, vol. 86, pp. 1037–52, 1878.

[197] R. Koch, "The etiology of tuberculosis," *Mittheilungen aus dem Kaiserlichen Gesundheit-samte*, vol. 2, pp. 1–88, 1884.

[198] A. M. Stern and H. Markel, "The history of vaccines and immunization: familiar patterns, new challenges," *Health affairs*, vol. 24, no. 3, pp. 611–621, 2005.

Bibliography

[199] J. Rolleston, "The smallpox pandemic of 1870-1874:(section of epidemiology and state medicine)," *Proceedings of the Royal Society of Medicine*, vol. 27, no. 2, p. 177, 1933.

[200] A. Fleming, "On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of b. influenzae," *British journal of experimental pathology*, vol. 10, no. 3, p. 226, 1929.

[201] D. J. Waxman and J. L. Strominger, "Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics," *Annual review of biochemistry*, vol. 52, no. 1, pp. 825–869, 1983.

[202] C. F. Amábile-Cuevas, *Antimicrobial resistance in bacteria.* Horizon Scientific Press, 2007.

[203] C. L. Ventola, "The antibiotic resistance crisis: part 1: causes and threats," *Pharmacy and therapeutics*, vol. 40, no. 4, p. 277, 2015.

[204] K. L. Tang, N. P. Caffrey, D. B. Nóbrega, S. C. Cork, P. E. Ronksley, H. W. Barkema, A. J. Polachek, H. Ganshorn, N. Sharma, J. D. Kellner, *et al.*, "Restricting the use of antibiotics in food-producing animals and its associations with antibiotic resistance in food-producing animals and human beings: a systematic review and meta-analysis," *The Lancet Planetary Health*, vol. 1, no. 8, pp. e316–e327, 2017.

[205] M. A. Fischbach and C. T. Walsh, "Antibiotics for emerging pathogens," *Science*, vol. 325, no. 5944, pp. 1089–1093, 2009.

[206] K. Lewis, "Platforms for antibiotic discovery," *Nature reviews Drug discovery*, vol. 12, no. 5, pp. 371–387, 2013.

[207] J. Conly and B. Johnston, "Where are all the new antibiotics? the new antibiotic paradox," *Canadian Journal of Infectious Diseases and Medical Microbiology*, vol. 16, no. 3, pp. 159–160, 2005.

[208] H. W. Boucher, G. H. Talbot, J. S. Bradley, J. E. Edwards, D. Gilbert, L. B. Rice, M. Scheld, B. Spellberg, and J. Bartlett, "Bad bugs, no drugs: no eskape! an update from the infectious diseases society of america," *Clinical infectious diseases*, vol. 48, no. 1, pp. 1–12, 2009.

[209] F. Codjoe and E. Donkor, "Carbapenem Resistance: A Review," *Medical Sciences*, vol. 6, p. 1, Dec. 2017.

[210] G. Patel and R. A. Bonomo, ""Stormy waters ahead": global emergence of carbapenemases," *Frontiers in Microbiology*, vol. 4, 2013.

[211] A. M. Queenan and K. Bush, "Carbapenemases: the Versatile -Lactamases," *Clinical Microbiology Reviews*, vol. 20, pp. 440–458, July 2007.

[212] C. M. Rotondo and G. D. Wright, "Inhibitors of metallo--lactamases," *Current Opinion in Microbiology*, vol. 39, pp. 96–105, Oct. 2017.

[213] L. L. Silver, "Challenges of antibacterial discovery," *Clinical microbiology reviews*, vol. 24, no. 1, pp. 71–109, 2011.

[214] D. Yong, M. A. Toleman, C. G. Giske, H. S. Cho, K. Sundman, K. Lee, and T. R. Walsh, "Characterization of a New Metallo- -Lactamase Gene, blaNDM-1, and a Novel Erythromycin Esterase Gene Carried on a Unique Genetic Structure in Klebsiella pneumoniae Sequence Type 14 from India," *Antimicrobial Agents and Chemotherapy*, vol. 53, pp. 5046–5054, Dec. 2009.

[215] A. U. Khan, L. Maryam, and R. Zarrilli, "Structure, Genetics and Worldwide Spread of New Delhi Metallo--lactamase (NDM): a threat to public health," *BMC Microbiology*, vol. 17, Dec. 2017.

[216] W. Wu, Y. Feng, G. Tang, F. Qiao, A. McNally, and Z. Zong, "NDM Metallo--Lactamases and Their Bacterial Producers in Health Care Settings," *Clinical Microbiology Reviews*, vol. 32, Jan. 2019.

[217] T. R. Walsh, J. Weeks, D. M. Livermore, and M. A. Toleman, "Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study," *The Lancet Infectious Diseases*, vol. 11, pp. 355–362, May 2011.

[218] B. Wang and D. Sun, "Detection of NDM-1 carbapenemase-producing Acinetobacter calcoaceticus and Acinetobacter junii in environmental samples from livestock farms," *Journal of Antimicrobial Chemotherapy*, vol. 70, pp. 611–613, Feb. 2015.

Bibliography

[219] S.-i. Narita and H. Tokuda, "Bacterial lipoproteins; biogenesis, sorting and quality control," *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1862, pp. 1414–1423, Nov. 2017.

[220] H. Tokuda and S.-i. Matsuyama, "Sorting of lipoproteins to the outer membrane in E. coli," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1693, pp. 5–13, July 2004.

[221] C. Schwechheimer and M. J. Kuehn, "Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions," *Nature Reviews Microbiology*, vol. 13, pp. 605–619, Oct. 2015.

[222] Y. Kim, M. A. Cunningham, J. Mire, C. Tesar, J. Sacchettini, and A. Joachimiak, "NDM-1, the ultimate promiscuous enzyme: substrate recognition and catalytic mechanism," *The FASEB Journal*, vol. 27, pp. 1917–1927, May 2013.

[223] V. L. Green, A. Verma, R. J. Owens, S. E. V. Phillips, and S. B. Carr, "Structure of New Delhi metallo--lactamase 1 (NDM-1)," *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, vol. 67, pp. 1160–1164, Oct. 2011.

[224] R. A. Fritz, J. H. Alzate-Morales, J. Spencer, A. J. Mulholland, and M. W. van der Kamp, "Multiscale Simulations of Clavulanate Inhibition Identify the Reactive Complex in Class A -Lactamases and Predict the Efficiency of Inhibition," *Biochemistry*, vol. 57, pp. 3560–3563, July 2018.

[225] M. Zheng and D. Xu, "New Delhi Metallo--Lactamase I: Substrate Binding and Catalytic Mechanism," *The Journal of Physical Chemistry B*, vol. 117, pp. 11596–11607, Oct. 2013.

[226] C. K. Das and N. N. Nair, "Hydrolysis of cephalexin and meropenem by New Delhi metallo--lactamase: the substrate protonation mechanism is drug dependent," *Physical Chemistry Chemical Physics*, vol. 19, no. 20, pp. 13111–13121, 2017.

[227] E. Lugtenberg and R. Peters, "Distribution of lipids in cytoplasmic and outer membranes of Escherichia coli K12," *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, vol. 441, pp. 38–47, July 1976.

[228] M. Jensen and O. G. Mouritsen, "Lipids do influence protein function—the hydrophobic matching hypothesis revisited," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1666, pp. 205–226, Nov. 2004.

[229] D. C. Lohman, D. Aydin, H. C. Von Bank, R. W. Smith, V. Linke, E. Weisenhorn, M. T. McDevitt, P. Hutchins, E. M. Wilkerson, B. Wancewicz, J. Russell, M. S. Stefely, E. T. Beebe, A. Jochem, J. J. Coon, C. A. Bingman, M. Dal Peraro, and D. J. Pagliarini, "An Isoprene Lipid-Binding Protein Promotes Eukaryotic Coenzyme Q Biosynthesis," *Molecular Cell*, vol. 73, pp. 763–774.e10, Feb. 2019.

[230] A. Kulp and M. J. Kuehn, "Biological functions and biogenesis of secreted bacterial outer membrane vesicles," *Annual review of microbiology*, vol. 64, pp. 163–184, 2010.

[231] J. D. Cecil, N. Sirisaengtaksin, N. M. O'Brien-Simpson, and A. M. Krachler, "Outer membrane vesicle-host cell interactions," *Protein Secretion in Bacteria*, pp. 201–214, 2019.

[232] C. López, J. A. Ayala, R. A. Bonomo, L. J. González, and A. J. Vila, "Protein determinants of dissemination and host specificity of metallo-β-lactamases," *Nature communications*, vol. 10, no. 1, pp. 1–11, 2019.

[233] S.-S. Jean, W.-S. Lee, C. Lam, C.-W. Hsu, R.-J. Chen, and P.-R. Hsueh, "Carbapenemase-producing Gram-negative bacteria: current epidemics, antimicrobial susceptibility and treatment options," *Future Microbiology*, vol. 10, pp. 407–425, Mar. 2015.

[234] B. Rogaski and J. B. Klauda, "Membrane-Binding Mechanism of a Peripheral Membrane Protein through Microsecond Molecular Dynamics Simulations," *Journal of Molecular Biology*, vol. 423, pp. 847–861, Nov. 2012.

[235] G. Bahr, L. Vitor-Horen, C. R. Bethel, R. A. Bonomo, L. J. González, and A. J. Vila, "Clinical Evolution of New Delhi Metallo--Lactamase (NDM) Optimizes Resistance under Zn(II) Deprivation," *Antimicrobial Agents and Chemotherapy*, vol. 62, Oct. 2017.

[236] S. Hayashi and H. C. Wu, "Lipoproteins in bacteria," *Journal of bioenergetics and biomembranes*, vol. 22, no. 3, pp. 451–471, 1990.

[237] G. von Heijne, "The structure of signal peptides from bacterial lipoproteins," *Protein Engineering, Design and Selection*, vol. 2, no. 7, pp. 531–534, 1989.

Bibliography

[238] A. S. Juncker, H. Willenbrock, G. Von Heijne, S. Brunak, H. Nielsen, and A. Krogh, "Prediction of lipoprotein signal peptides in gram-negative bacteria," *Protein Science*, vol. 12, no. 8, pp. 1652–1662, 2003.

[239] T. G. Schmidt, L. Batz, L. Bonet, U. Carl, G. Holzapfel, K. Kiem, K. Matulewicz, D. Niermeier, I. Schuchardt, and K. Stanar, "Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants," *Protein Expression and Purification*, vol. 92, pp. 54–61, Nov. 2013.

[240] H. Zhang, G. Ma, Y. Zhu, L. Zeng, A. Ahmad, C. Wang, B. Pang, H. Fang, L. Zhao, and Q. Hao, "Active-Site Conformational Fluctuations Promote the Enzymatic Activity of NDM-1," *Antimicrobial Agents and Chemotherapy*, vol. 62, Aug. 2018.

[241] I. Garcia-Saez, J.-D. Docquier, G. Rossolini, and O. Dideberg, "The Three-Dimensional Structure of VIM-2, a Zn--Lactamase from Pseudomonas aeruginosa in Its Reduced and Oxidised Form," *Journal of Molecular Biology*, vol. 375, pp. 604–611, Jan. 2008.

[242] M. I. Abboud, M. Kosmopoulou, A. P. Krismanich, J. W. Johnson, P. Hinchliffe, J. Brem, T. D. Claridge, J. Spencer, C. J. Schofield, and G. I. Dmitrienko, "Cyclobutanone mimics of intermediates in metallo-β-lactamase catalysis," *Chemistry–A European Journal*, vol. 24, no. 22, pp. 5734–5737, 2018.

[243] P. Hinchliffe, M. M. González, M. F. Mojica, J. M. González, V. Castillo, C. Saiz, M. Kosmopoulou, C. L. Tooke, L. I. Llarrull, G. Mahler, *et al.*, "Cross-class metallo-β-lactamase inhibition by bisthiazolidines reveals multiple binding modes," *Proceedings of the National Academy of Sciences*, vol. 113, no. 26, pp. E3745–E3754, 2016.

[244] S. Morein, A.-S. Andersson, L. Rilfors, and G. Lindblom, "Wild-type *Escherichia coli* Cells Regulate the Membrane Lipid Composition in a Window between Gel and Non-lamellar Structures," *Journal of Biological Chemistry*, vol. 271, pp. 6801–6809, Mar. 1996.

[245] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied Physics*, vol. 52, pp. 7182–7190, Dec. 1981.

158

[246] Y. Atsmon-Raz and D. P. Tieleman, "Parameterization of Palmitoylated Cysteine, Farnesylated Cysteine, Geranylgeranylated Cysteine, and Myristoylated Glycine for the Martini Force Field," *The Journal of Physical Chemistry B*, vol. 121, pp. 11132–11143, Dec. 2017.

[247] T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink, and D. P. Tieleman, "Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models," *Journal of Chemical Theory and Computation*, vol. 10, pp. 676–690, Feb. 2014.

[248] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins: Structure, Function, and Bioinformatics*, pp. NA–NA, 2010.

[249] T. Lemmin, C. Bovigny, D. Lançon, and M. Dal Peraro, "Cardiolipin Models for Molecular Simulations of Bacterial and Mitochondrial Membranes," *Journal of Chemical Theory and Computation*, vol. 9, pp. 670–678, Jan. 2013.

[250] S. C. Hoops, K. W. Anderson, and K. M. Merz, "Force field design for metalloproteins," *Journal of the American Chemical Society*, vol. 113, pp. 8262–8270, Oct. 1991.

[251] D. Suárez, E. N. Brothers, and K. M. Merz, "Insights into the Structure and Dynamics of the Dinuclear Zinc -Lactamase Site from *Bacteroides fragilis* [†]," *Biochemistry*, vol. 41, pp. 6615–6630, May 2002.

[252] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten, "Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions," *Molecular Simulation*, vol. 6, pp. 121–142, Mar. 1991.

[253] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *The Journal of Chemical Physics*, vol. 126, p. 014101, Jan. 2007.

[254] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems," *The Journal of Chemical Physics*, vol. 98, pp. 10089–10092, June 1993.

[255] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E.

Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M. J. Holst, J. A. McCammon, and N. A. Baker, "Improvements to the APBS biomolecular solvation software suite: Improvements to the APBS Software Suite," *Protein Science*, vol. 27, pp. 112–128, Jan. 2018.

[256] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen, "PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K_a Predictions," *Journal of Chemical Theory and Computation*, vol. 7, pp. 525–537, Feb. 2011.

[257] W. G. Dunphy and J. E. Rothman, "Compartmental organization of the golgi stack," *Cell*, vol. 42, no. 1, pp. 13–21, 1985.

[258] J. Shorter and G. Warren, "Golgi architecture and inheritance," *Annual review of cell and developmental biology*, vol. 18, no. 1, pp. 379–420, 2002.

[259] Y. Wang, "Golgi apparatus inheritance," in *The Golgi apparatus*, pp. 580–607, Springer, 2008.

[260] P. Mazzarello, C. Garbarino, and A. Calligaro, "How camillo golgi became ?the golgi?," *FEBS letters*, vol. 583, no. 23, pp. 3732–3737, 2009.

[261] R. W. Doms, G. Russ, and J. W. Yewdell, "Brefeldin a redistributes resident and itinerant golgi proteins to the endoplasmic reticulum.," *The Journal of cell biology*, vol. 109, no. 1, pp. 61–72, 1989.

[262] R. D. Teasdale and M. R. Jackson, "Signal-mediated sorting of membrane proteins between the endoplasmic reticulum and the golgi apparatus," *Annual review of cell and developmental biology*, vol. 12, no. 1, pp. 27–54, 1996.

[263] S. R. Pfeffer and J. E. Rothman, "Biosynthetic protein transport and sorting by the endoplasmic reticulum and golgi," *Annual review of biochemistry*, vol. 56, no. 1, pp. 829–852, 1987.

[264] V. K. Chiu, T. Bivona, A. Hach, J. B. Sajous, J. Silletti, H. Wiener, R. L. Johnson II, A. D. Cox, and M. R. Philips, "Ras signalling on the endoplasmic reticulum and the golgi," *Nature cell biology*, vol. 4, no. 5, p. 343, 2002.

[265] N. Gomez-Navarro and E. Miller, "Protein sorting at the er–golgi interface," *Journal of Cell Biology*, vol. 215, no. 6, pp. 769–778, 2016.

[266] J. Klumperman, "Architecture of the mammalian golgi," *Cold Spring Harbor perspectives in biology*, vol. 3, no. 7, p. a005181, 2011.

[267] P. A. Gleeson, C. Makhoul, and P. Gosavi, "Golgi dynamics: The morphology of the mammalian golgi apparatus in health and disease," *Frontiers in Cell and Developmental Biology*, vol. 7, p. 112, 2019.

[268] I. Mellman and K. Simons, "The golgi complex: in vitro veritas?," *Cell*, vol. 68, no. 5, pp. 829–840, 1992.

[269] C. Makhoul, P. Gosavi, and P. A. Gleeson, "The golgi architecture and cell sensing," *Biochemical Society Transactions*, vol. 46, no. 5, pp. 1063–1072, 2018.

[270] S. Huang and Y. Wang, "Golgi structure formation, function, and post-translational modifications in mammalian cells," *F1000Research*, vol. 6, 2017.

[271] A. M. Ernst, S. A. Syed, O. Zaki, F. Bottanelli, H. Zheng, M. Hacke, Z. Xi, F. Rivera-Molina, M. Graham, A. A. Rebane, *et al.*, "S-palmitoylation sorts membrane cargo for anterograde transport in the golgi," *Developmental cell*, vol. 47, no. 4, pp. 479–493, 2018.

[272] S. Capasso, L. Sticco, R. Rizzo, M. Pirozzi, D. Russo, N. A. Dathan, F. Campelo, J. van Galen, M. Hölttä-Vuori, G. Turacchio, *et al.*, "Sphingolipid metabolic flow controls phosphoinositide turnover at the trans-golgi network," *The EMBO journal*, vol. 36, no. 12, pp. 1736–1754, 2017.

[273] R. S. Kuna and S. J. Field, "Golph3: A golgi phosphatidylinositol (4) phosphate effector that directs vesicle trafficking and drives cancer," *Journal of lipid research*, vol. 60, no. 2, pp. 269–275, 2019.

[274] L. Tu, W. C. Tai, L. Chen, and D. K. Banfield, "Signal-mediated dynamic retention of glycosyltransferases in the golgi," *Science*, vol. 321, no. 5887, pp. 404–407, 2008.

[275] M. D. Buschman, M. Xing, and S. J. Field, "The golph3 pathway regulates golgi shape and function and is activated by dna damage," *Frontiers in neuroscience*, vol. 9, p. 362, 2015.

Bibliography

[276] K. L. Scott and L. Chin, "Signaling from the golgi: Mechanisms and models for golgi phosphoprotein 3–mediated oncogenesis," *Clinical Cancer Research*, vol. 16, no. 8, pp. 2229–2234, 2010.

[277] G. Van Meer, D. R. Voelker, and G. W. Feigenson, "Membrane lipids: where they are and how they behave," *Nature reviews Molecular cell biology*, vol. 9, no. 2, pp. 112–124, 2008.

[278] E. E. Ambroggio, J. Sillibourne, B. Antonny, J.-B. Manneville, and B. Goud, "Arf1 and membrane curvature cooperate to recruit arfaptin2 to liposomes," *PloS one*, vol. 8, no. 4, 2013.

[279] M. Magdeleine, R. Gautier, P. Gounon, H. Barelli, S. Vanni, and B. Antonny, "A filter at the entrance of the golgi that selects vesicles according to size and bulk lipid composition," *Elife*, vol. 5, p. e16988, 2016.

[280] H. C. Dippold, M. M. Ng, S. E. Farber-Katz, S.-K. Lee, M. L. Kerr, M. C. Peterman, R. Sim, P. A. Wiharto, K. A. Galbraith, S. Madhavarapu, *et al.*, "Golph3 bridges phosphatidylinositol-4-phosphate and actomyosin to stretch and shape the golgi to promote budding," *Cell*, vol. 139, no. 2, pp. 337–351, 2009.

[281] J. Stümpel, H. Eibl, and A. Nicksch, "X-ray analysis and calorimetry on phosphatidylcholine model membranes. the influence of length and position of acyl chains upon structure and phase behaviour," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 727, no. 2, pp. 246–254, 1983.

[282] J. A. Szule, N. L. Fuller, and R. P. Rand, "The effects of acyl chain length and saturation of diacylglycerols and phosphatidylcholines on membrane monolayer curvature," *Biophysical journal*, vol. 83, no. 2, pp. 977–984, 2002.

[283] X.-Y. Li, W. Liu, S.-F. Chen, L.-Q. Zhang, X.-G. Li, and L.-X. Wang, "Expression of the golgi phosphoprotein-3 gene in human gliomas: a pilot study," *Journal of neuro-oncology*, vol. 105, no. 2, pp. 159–163, 2011.

[284] O. Kunigou, H. Nagao, N. Kawabata, Y. Ishidou, S. Nagano, S. Maeda, S. Komiya, and T. Setoguchi, "Role of golph3 and golph3l in the proliferation of human rhabdomyosarcoma," *Oncology reports*, vol. 26, no. 5, pp. 1337–1342, 2011.

[285] Z. Zeng, H. Lin, X. Zhao, G. Liu, X. Wang, R. Xu, K. Chen, J. Li, and L. Song, "Overexpression of golph3 promotes proliferation and tumorigenicity in breast cancer via suppression of the foxo1 transcription factor," *Clinical Cancer Research*, vol. 18, no. 15, pp. 4059–4069, 2012.

[286] E. P. Carpenter, K. Beis, A. D. Cameron, and S. Iwata, "Overcoming the challenges of membrane protein crystallography," *Current opinion in structural biology*, vol. 18, no. 5, pp. 581–586, 2008.

[287] K. M. Merz and B. Roux, *Biological membranes: A molecular perspective from computation and experiment.* Springer Science & Business Media, 2012.

[288] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, "CHARMM36m: an improved force field for folded and intrinsically disordered proteins," *Nature Methods*, vol. 14, pp. 71–73, Jan. 2017.

[289] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of Computational Chemistry*, pp. NA–NA, 2009.

[290] H. Ohshima, *Biophysical chemistry of biointerfaces.* John Wiley & Sons, 2011.

[291] A. Chilkoti and J. A. Hubbell, "Biointerface science," *Mrs Bulletin*, vol. 30, no. 3, pp. 175–179, 2005.

[292] D. Alsteens, H. E. Gaub, R. Newton, M. Pfreundschuh, C. Gerber, and D. J. Müller, "Atomic force microscopy-based characterization and design of biointerfaces," *Nature Reviews Materials*, vol. 2, no. 5, pp. 1–16, 2017.

[293] R. H. Fang, Y. Jiang, J. C. Fang, and L. Zhang, "Cell membrane-derived nanomaterials for biomedical applications," *Biomaterials*, vol. 128, pp. 69–83, 2017.

[294] S. Zalba and T. L. ten Hagen, "Cell membrane modulation as adjuvant in cancer therapy," *Cancer treatment reviews*, vol. 52, pp. 48–57, 2017.

[295] S. Bottaro and K. Lindorff-Larsen, "Biophysical experiments and biomolecular simulations: A perfect match?," *Science*, vol. 361, no. 6400, pp. 355–360, 2018.

Bibliography

[296] P. S. Nerenberg and T. Head-Gordon, "New developments in force fields for biomolecular simulations," *Current opinion in structural biology*, vol. 49, pp. 129–138, 2018.

[297] Z. Jing, C. Liu, S. Y. Cheng, R. Qi, B. D. Walker, J.-P. Piquemal, and P. Ren, "Polarizable force fields for biomolecular simulations: Recent advances and applications," *Annual Review of biophysics*, vol. 48, pp. 371–394, 2019.

[298] D. J. Huggins, P. C. Biggin, M. A. Dämgen, J. W. Essex, S. A. Harris, R. H. Henchman, S. Khalid, A. Kuzmanic, C. A. Laughton, J. Michel, *et al.*, "Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 9, no. 3, p. e1393, 2019.

[299] D. E. Shaw, R. O. Dror, J. K. Salmon, J. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, *et al.*, "Millisecond-scale molecular dynamics simulations on anton," in *Proceedings of the conference on high performance computing networking, storage and analysis*, pp. 1–11, 2009.

[300] L. C. Pierce, R. Salomon-Ferrer, C. Augusto F. de Oliveira, J. A. McCammon, and R. C. Walker, "Routine access to millisecond time scale events with accelerated molecular dynamics," *Journal of chemical theory and computation*, vol. 8, no. 9, pp. 2997–3002, 2012.

[301] T. J. Lane, D. Shukla, K. A. Beauchamp, and V. S. Pande, "To milliseconds and beyond: challenges in the simulation of protein folding," *Current opinion in structural biology*, vol. 23, no. 1, pp. 58–65, 2013.

[302] P. Tiwary and M. Parrinello, "Breaking the millisecond barrier in molecular dynamics simulations of protein-ligand unbinding," in *ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY*, vol. 250, AMER CHEMICAL SOC 1155 16TH ST, NW, WASHINGTON, DC 20036 USA, 2015.

[303] K. Kadau, T. C. Germann, and P. S. Lomdahl, "Large-scale molecular-dynamics simulation of 19 billion particles," *International Journal of Modern Physics C*, vol. 15, no. 01, pp. 193–201, 2004.

[304] K. Kadau, T. C. Germann, and P. S. Lomdahl, "Molecular dynamics comes of age: 320 billion atom simulation on bluegene/l," *International Journal of Modern Physics C*, vol. 17, no. 12, pp. 1755–1761, 2006.

[305] N. Trebesch and E. Tajkhorshid, "Multibillion atom molecular dynamics simulations of cellular membranes," *Bulletin of the American Physical Society*, 2020.

[306] M. Pons-Salort, E. P. Parker, and N. C. Grassly, "The epidemiology of non-polio enteroviruses: recent advances and outstanding questions," *Current opinion in infectious diseases*, vol. 28, no. 5, p. 479, 2015.

[307] C. . C. for Disease Control and Prevention, "Polio vaccination," 2019.

[308] Y. Li, Y. Zhou, Y. Cheng, P. Wu, C. Zhou, P. Cui, C. Song, L. Liang, F. Wang, Q. Qiu, *et al.*, "Effectiveness of ev-a71 vaccination in prevention of paediatric hand, foot, and mouth disease associated with ev-a71 virus infection requiring hospitalisation in henan, china, 2017–18: a test-negative case-control study," *The Lancet Child & Adolescent Health*, vol. 3, no. 10, pp. 697–704, 2019.

[309] W. Dai, C. Zhang, X. Zhang, P. Xiong, Q. Liu, S. Gong, L. Geng, D. Zhou, and Z. Huang, "A virus-like particle vaccine confers protection against enterovirus d68 lethal challenge in mice," *Vaccine*, vol. 36, no. 5, pp. 653–659, 2018.

[310] Q. Mao, Y. Wang, X. Yao, L. Bian, X. Wu, M. Xu, and Z. Liang, "Coxsackievirus a16: epidemiology, diagnosis, and vaccine," *Human vaccines & immunotherapeutics*, vol. 10, no. 2, pp. 360–367, 2014.

[311] U. Kartoglu and J. Milstien, "Tools and approaches to ensure quality of vaccines throughout the cold chain," *Expert review of vaccines*, vol. 13, no. 7, pp. 843–854, 2014.

[312] M. L. Brusseau, I. L. Pepper, and C. Gerba, *Environmental and pollution science.* Academic Press, 2019.

[313] E. P. LARKIN, "Food contaminants–viruses," *Journal of food protection*, vol. 44, no. 4, pp. 320–325, 1981.

Bibliography

[314] I. Bertrand, J. Schijven, G. Sánchez, P. Wyn-Jones, J. Ottoson, T. Morin, M. Muscillo, M. Verani, A. Nasser, A. de Roda Husman, *et al.*, "The impact of temperature on the inactivation of enteric viruses in food and water: a review," *Journal of applied microbiology*, vol. 112, no. 6, pp. 1059–1074, 2012.

[315] A. Kotecha, J. Seago, K. Scott, A. Burman, S. Loureiro, J. Ren, C. Porta, H. M. Ginn, T. Jackson, E. Perez-Martin, *et al.*, "Structure-based energetics of protein interfaces guides foot-and-mouth disease virus vaccine design," *Nature structural & molecular biology*, vol. 22, no. 10, pp. 788–794, 2015.

[316] C. Porta, A. Kotecha, A. Burman, T. Jackson, J. Ren, S. Loureiro, I. M. Jones, E. E. Fry, D. I. Stuart, and B. Charleston, "Rational engineering of recombinant picornavirus capsids to produce safe, protective vaccine antigen," *PLoS pathogens*, vol. 9, no. 3, 2013.

[317] K. A. Scott, A. Kotecha, J. Seago, J. Ren, E. E. Fry, D. I. Stuart, B. Charleston, and F. F. Maree, "Sat2 foot-and-mouth disease virus structurally modified for increased thermostability," *Journal of virology*, vol. 91, no. 10, pp. e02312–16, 2017.

[318] V. Rincón, A. Rodríguez-Huete, S. López-Argüello, B. Ibarra-Molero, J. M. Sanchez-Ruiz, M. M. Harmsen, and M. G. Mateu, "Identification of the structural basis of thermal lability of a virus provides a rationale for improved vaccines," *Structure*, vol. 22, no. 11, pp. 1560–1570, 2014.

[319] D. M. Belnap, D. J. Filman, B. L. Trus, N. Cheng, F. P. Booy, J. F. Conway, S. Curry, C. N. Hiremath, S. K. Tsang, A. C. Steven, *et al.*, "Molecular tectonic model of virus structural transitions: the putative cell entry states of poliovirus," *Journal of virology*, vol. 74, no. 3, pp. 1342–1354, 2000.

[320] H. C. Levy, M. Bostina, D. J. Filman, and J. M. Hogle, "Catching a virus in the act of rna release: a novel poliovirus uncoating intermediate characterized by cryo-electron microscopy," *Journal of virology*, vol. 84, no. 9, pp. 4426–4441, 2010.

[321] X. Wang, W. Peng, J. Ren, Z. Hu, J. Xu, Z. Lou, X. Li, W. Yin, X. Shen, C. Porta, *et al.*, "A sensor-adaptor mechanism for enterovirus uncoating from structures of ev71," *Nature structural & molecular biology*, vol. 19, no. 4, p. 424, 2012.

166

[322] J. J. Seitsonen, S. Shakeel, P. Susi, A. P. Pandurangan, R. S. Sinkovits, H. Hyvönen, P. Laurinmäki, J. Ylä-Pelto, M. Topf, T. Hyypiä, *et al.*, "Structural analysis of coxsackievirus a7 reveals conformational changes associated with uncoating," *Journal of virology*, vol. 86, no. 13, pp. 7207–7215, 2012.

[323] M. Bostina, H. Levy, D. J. Filman, and J. M. Hogle, "Poliovirus rna is released from the capsid near a twofold symmetry axis," *Journal of virology*, vol. 85, no. 2, pp. 776–783, 2011.

[324] K. L. Shingler, J. L. Yoder, M. S. Carnegie, R. E. Ashley, A. M. Makhov, J. F. Conway, and S. Hafenstein, "The enterovirus 71 a-particle forms a gateway to allow genome release: a cryoem study of picornavirus uncoating," *PLoS pathogens*, vol. 9, no. 3, 2013.

[325] O. O. Adeyemi, C. Nicol, N. J. Stonehouse, and D. J. Rowlands, "Increasing type 1 poliovirus capsid stability by thermal selection," *Journal of virology*, vol. 91, no. 4, pp. e01586–16, 2017.

[326] H. Shiomi, T. Urasawa, S. Urasawa, N. Kobayashi, S. Abe, and K. Taniguchi, "Isolation and characterisation of poliovirus mutants resistant to heating at 50 c for 30 min," *Journal of medical virology*, vol. 74, no. 3, pp. 484–491, 2004.

[327] Y. Nguyen, P. R. Jesudhasan, E. R. Aguilera, and J. K. Pfeiffer, "Identification and characterization of a poliovirus capsid mutant with enhanced thermal stability," *Journal of virology*, vol. 93, no. 6, pp. e01510–18, 2019.

[328] M. Majer and R. Thomssen, "Thermal inactivation of 32 p-poliovirus at 37 c and 50 c in the presence of nacl with high molarity," *Archiv für die gesamte Virusforschung*, vol. 17, no. 5, pp. 585–593, 1965.

[329] C. Wallis and J. Melnick, "Stabilization of poliovirus by cations.," *Texas reports on biology and medicine*, vol. 19, p. 683, 1961.

[330] C. Wallis, J. L. Melnick, and F. Rapp, "Different effects of mgcl2 and mgso4 on the thermostability of viruses," *Virology*, vol. 26, no. 4, pp. 694–699, 1965.

[331] K. D. Mena, C. P. Gerba, C. N. Haas, and J. B. Rose, "Risk assessment of waterborne coxsackievirus," *Journal-American Water Works Association*, vol. 95, no. 7, pp. 122–131, 2003.

[332] P. Payment, M. Tremblay, and M. Trudel, "Relative resistance to chlorine of poliovirus and coxsackievirus isolates from environmental sources and drinking water.," *Appl. Environ. Microbiol.*, vol. 49, no. 4, pp. 981–983, 1985.

[333] S. Meister, M. E. Verbyla, M. Klinger, and T. Kohn, "Variability in disinfection resistance between currently circulating enterovirus b serotypes and strains," *Environmental science & technology*, vol. 52, no. 6, pp. 3696–3705, 2018.

[334] M. Butler, A. Medien, and G. Taylor, "Electrofocusing of viruses and sensitivity to disinfection," *Water Science and Technology*, vol. 17, no. 10, pp. 201–210, 1985.

[335] D. C. Young and D. G. Sharp, "Virion conformational forms and the complex inactivation kinetics of echovirus by chlorine in water.," *Appl. Environ. Microbiol.*, vol. 49, no. 2, pp. 359–364, 1985.

[336] Y. Kott, "Estimation of low numbers of escherichia coli bacteriophage by use of the most probable number method," *Appl. Environ. Microbiol.*, vol. 14, no. 2, pp. 141–144, 1966.

[337] J. K. Muckelbauer, M. Kremer, I. Minor, G. Diana, F. J. Dutko, J. Groarke, D. C. Pevear, and M. G. Rossmann, "The structure of coxsackievirus b3 at 3.5 å resolution," *Structure*, vol. 3, no. 7, pp. 653–667, 1995.

[338] B. Webb and A. Sali, "Comparative protein structure modeling using modeller," *Current protocols in bioinformatics*, vol. 54, no. 1, pp. 5–6, 2016.

[339] W. Humphrey, A. Dalke, K. Schulten, *et al.*, "Vmd: visual molecular dynamics," *Journal of molecular graphics*, vol. 14, no. 1, pp. 33–38, 1996.

[340] M. Rostkowski, M. H. Olsson, C. R. Søndergaard, and J. H. Jensen, "Graphical analysis of ph-dependent properties of proteins predicted using propka," *BMC structural biology*, vol. 11, no. 1, p. 6, 2011.

[341] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of chemical physics*, vol. 79, no. 2, pp. 926–935, 1983.

[342] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable molecular dynamics with namd," *Journal of computational chemistry*, vol. 26, no. 16, pp. 1781–1802, 2005.

[343] R. C. Team, "R: a language and environment for statistical computing. r found," *Stat. Comput. Vienna, Austria*, 2017.

[344] H. Wickham, "ggplot2 elegant graphics for data analysis. media (vol. 35)," 2009.

[345] B. Auguie, "gridextra: Miscellaneous functions for ?grid? graphics. r package version 2.3," 2016.

[346] V. M. Muggeo *et al.*, "Segmented: an r package to fit regression models with broken-line relationships," *R news*, vol. 8, no. 1, pp. 20–25, 2008.

[347] A. T. Vaidya, D. Top, C. C. Manahan, J. M. Tokuda, S. Zhang, L. Pollack, M. W. Young, and B. R. Crane, "Flavin reduction activates drosophila cryptochrome," *Proceedings of the National Academy of Sciences*, vol. 110, no. 51, pp. 20455–20460, 2013.

[348] A. Czarna, A. Berndt, H. R. Singh, A. Grudziecki, A. G. Ladurner, G. Timinszky, A. Kramer, and E. Wolf, "Structures of drosophila cryptochrome and mouse cryptochrome1 provide insight into circadian function," *Cell*, vol. 153, no. 6, pp. 1394–1405, 2013.

[349] C. Lin, D. Top, C. C. Manahan, M. W. Young, and B. R. Crane, "Circadian clock activity of cryptochrome relies on tryptophan-mediated photoreduction," *Proceedings of the National Academy of Sciences*, vol. 115, no. 15, pp. 3822–3827, 2018.

Alessio Prunotto

✉ alessio.prunotto@gmail.com   in www.linkedin.com/in/alessioprunotto/
ⓘ 0000-0002-6425-964X

Education

07/2015 - 08/2020	**Ph.D. in Biotechnology and Bioengineering** Swiss Federal Institute of Technology (EPFL, Switzerland) *Advisor: Prof. Matteo Dal Peraro* *Thesis: Characterization of Protein/Membrane Interfaces through* *a Synergistic Computational/Experimental Approach*
10/2010 - 12/2012	**M.Sc. in Biomedical Engineering** Polytechnic University of Turin (Italy) *Advisor: Dr. Marco Deriu* *Thesis: Computational Molecular Modeling to Improve Drug Design and to* *Better Understand Subcellular Dynamics*
04/2012 - 10/2012	**Visiting Student** University of Alberta (Canada) *Advisor: Prof. Jack Tuszynski* *Research topic: Computational modelling and characterization of a viral enzyme*
09/2007 - 10/2010	**B.Sc. in Biomedical Engineering** Polytechnic University of Turin (Italy) *Advisor: Prof. Cristina Bignardi* *Thesis: Creation of a Bone Element 3D Structure from 2D Digital Images*

Employment History

07/2015 - 07/2020	**Swiss Federal Institute of Technology (EPFL, Switzerland)** Doctoral Assistant, Laboratory for Biomolecular Modeling Teaching Assistant, School of Life Sciences
09/2018 - 10/2018	**Institute of Molecular and Cell Biology of Rosario (Argentina)** Visiting researcher, Laboratory for Metalloproteins *Advisor: Prof. Alejandro J. Vila* *Research topic: Structural characterization of a peripheral membrane protein*
01/2015 - 06/2015	**Swiss Federal Institute of Technology (EPFL, Switzerland)** Research Assistant, Laboratory for Biomolecular Modeling *Research topic: Computational studies on protein aggregation*
01/2013 - 12/2014	**University of Applied Sciences and Arts of Southern Switzerland** Research Assistant, Computational Biophysics Group *Advisor: Prof. Andrea Danani* *Research topic: Protein modelling, structure-based drug discovery*
04/2012 - 10/2012	**University of Alberta (Canada)** Graduate Researcher
04/2010 - 06/2010	**Intrauma, Inc., Rivoli (Italy)** Intern at the Department of Biomechanical Engineering

Awards and Fellowships

| 03/2019 | **The Biophysical Society International Travel Award**, *USA* |
| 04/2012 - 10/2012 | **Student Mobility Scholarship for M.Sc. Final Project**, *Italy* |

Publications in Peer-Reviewed Journals

2020 Meister S, **Prunotto A**, Dal Peraro M, Kohn T.
"Salt enhances the thermostability of enteroviruses by stabilizing capsid protein interfaces." *Journal of Virology*. DOI: 10.1128/JVI.02176-19

2017 Ayoub A, Staelens M, **Prunotto A**, Deriu M, Danani A, Klobukowski M, Tuszynski J.
"Explaining the Microtubule Energy Balance: Contributions Due to Dipole Moments, Charges, van der Waals and Solvation Energy." *International Journal of Molecular Sciences*, 18(10), p.2042. DOI: 10.3390/ijms18102042

2015 Gentile F, Deriu M, Licandro G, **Prunotto A**, Danani A, Tuszynski J.
"Structure based modeling of small molecules binding to the TLR7 by atomistic level simulations." *Molecules*, 20(5), pp.8316-8340. DOI: 10.3390/molecules20058316

2013 Barakat KH, Law J, **Prunotto A**, Magee WC, Evans DH, Tyrrell DL, Tuszynski J, Houghton M.
"Detailed computational study of the active site of the hepatitis C viral RNA polymerase to aid novel drug design". *Journal of Chemical Information and Modeling*, 53(11), pp.3031-3043. DOI: 10.1021/ci4003969

Unpublished Manuscripts in Preparation

1. Aydin D*, **Prunotto A***, Abriata LA, Dal Peraro M.
   "Integrative Modeling at the Protein-Membrane Interface."
   (Invited review article for *Accounts of Chemical Research*, *Co-First Authorship)

2. **Prunotto A***, Bahr G*, Gonzalez LJ, Dal Peraro M., Vila AJ.
   "Optimization of Protein Features to Adapt a Resistant Enzyme for Membrane Anchoring."
   (*Co-First Authorship)

3. Meireles-Filho ACA, Litovchenko M, Frochaux M, Bevers R, **Prunotto A**, Hollis B, Gardeux V, Braman V, Julie Russeil J, Dal Peraro M, Deplancke B.
   "Extensive tissue-specific expression variation and novel regulators underlying circadian behaviour."

4. Shen Y, Kalograiaki I, **Prunotto A**, Dunne M, Taylor NM, Sumrall E, Eugster MR, Martin R, Julian-Rodero A, Gerber B, Bulous S, Menéndez M, Leiman P, Dal Peraro M, Cañada FJ, Loessner MJ.
   "Molecular basis for recognition of Listeria cell wall teichoic acid by pseudo-symmetric SH3b-like repeats of a bacteriophage endolysin."

Proceedings

2012 Carbone L, Deriu M, **Prunotto A**, Bignardi C, Morbiducci U.
"Mechanical Properties of Single-wall Carbon Nanotubes.", *GNB 2012*, Rome (IT), June 26 -29. pp. 1–2.

Poster Presentations in International Conferences

2019   **Prunotto A**, Bahr G, Gonzalez LJ, Vila A, Dal Peraro M. "Computer Simulations Give Insights into the Molecular Mechanisms that Lie at the Basis of a Super-Bacterium."
The Collaborative Computational Project for Biomolecular Simulation (CCPBioSim) Annual Meeting, Bristol, UK, September 4-6

2019   **Prunotto A**, Bahr G, Gonzalez LJ, Vila A, Dal Peraro M. "Molecular simulations give insights into the NDM-1/membrane interaction that causes rise of a super-bacterium."
The Biophysical Society (BPS) Annual Meeting, Baltimore, MD, USA, March 2-6

2019   **Prunotto A**, Vila A, Dal Peraro M. "Molecular dynamics simulations shed a light on the relationship between membrane anchoring and the antibiotic resistance mechanism of New Delhi Metallo-β-Lactamase 1."
Multiscale Modeling from Macromolecules to Cell: Opportunities and Challenges of Biomolecular Simulations, Lausanne, Switzerland, February 4-6

2018   **Prunotto A**, Vila A, Dal Peraro M. "Exploiting the role of membrane anchoring on the peculiar biological behaviour of multi-drug resistant enzyme New Delhi Metallo-β-Lactamase 1."
The American Chemical Society (ACS) National Meeting, Boston, MA, USA, August 19-23

2017   **Prunotto A**, Vila A, Dal Peraro M. "Influence of bacterial membranes on the function of New Delhi Metallo-β-Lactamase 1."
Computational Advances in Drug Discovery, Lausanne, Switzerland, September 5-8

2017   **Prunotto A**, Vila A, Dal Peraro M. "Influence of bacterial membranes on the function of New Delhi Metallo-β-Lactamase 1."
The future of Biomembrane Simulations: Hidden Pitfalls and Future Challenges, Lyon, France, June 6-9

2016   **Prunotto A**, Dal Peraro M. "The Role of Cardiolipins in Uncoupling Proteins 1/2 Functions."
The Biophysical Society (BPS) Thematic Meeting on Liposomes, Exosomes and Virosomes, Ascona, Switzerland, September 11-16

2014   **Prunotto A**, Licandro G, Deriu M, Grasso G, Gentile F, Tuszynski J, Danani A. "Drug-conjugation Emerging Properties: Molecular Insights."
From Drug Discovery to Drug Delivery, Athens, Greece, November 13-15

Supervision of Students

03/2015 - 09/2015   **Master's internship project**, Aurelien Severino, *EPFL*
"Molecular dynamics simulations of AACP in mitochondrial membranes."
Supervised the exploitation of the role played by lipids diffusion on the function of a mitochondrial carrier.

06/2017 - 09/2017   **Summer project**, Lu Xiaoping, *EPFL*
"Generation of a reliable atomistic model for the catalytic site of NDM-1."
Supervised the parameterization of the metal ions contained in the catalytic site of New Delhi Metallo-β-Lactamase 1, together with their coordinating residues.

10/2018 - 01/2019   **Master's internship project**, Cem Tekin, *EPFL*
"Computer simulations to exploit the function mechanism of DHHC20."
Supervised the computational simulations of DHHC20 under different conditions, i.e. with/without zinc ions and post-translationally modified residues, to assess their influence on the protein structure and function.

Teaching Activities

Spring 2016 & Spring 2017 & Spring 2018 & Spring 2019	**Graduate Biomolecular Structure and Mechanics**, *EPFL* *Developed exercises and projects; graded assignments; assisted students during practical sessions about computational structural biology; supervised students for their semester project and presentation.*
Spring 2017	**Undergraduate Global Issues: Health**, *EPFL* *Proposed and assigned major global health issues to five groups of students; and supervised the preparation of their semester project and final poster presentation.*
Spring 2016	**Undergraduate Physical Biology of the Cell II**, *EPFL* *Developed weekly exercises and assisted students during practical sessions.*
Fall 2010	**Undergraduate Medical Images**, *Polytechnic University of Turin* *Assisted students during practical sessions and graded weekly assignments.*

Skills

- **Languages**
  English (*Fluent*)
  French (*Advanced*)
  Spanish (*Basic*)
  Italian (*Native*)

- **Computer Skills**
  Computational Chemistry: Gaussian
  Molecular Dynamics Simulations: Gromacs, NAMD, AMBER
  Molecular Docking: AutoDock Vina, Schrödinger Suite (Glide)
  Homology Modeling: Modeller
  Molecular Modeling and Visualization: PyMOL, VMD, Schrödinger Suite (Maestro)
  Computing: Python, Matlab, R, C++, Tcl, Bash, LaTeX
  Computer Graphics: Illustrator