

ELUCIDATING SELF-ASSEMBLY
AND ANTIMICROBIAL STRATEGIES
OF SYNTHETIC PEPTIDES:
AN IN SILICO INVESTIGATION

Irene Marzuoli

RANDALL CENTRE OF CELL AND MOLECULAR BIOPHYSICS
KING'S COLLEGE LONDON



THIS DISSERTATION IS SUBMITTED FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

SEPTEMBER 2019

Declaration

This dissertation describes work I have carried out between October 2016 and September 2019 at the Randall Centre of King's College London, under the supervision of Professor Franca Fraternali (first supervisor) and Dr. Chris D. Lorenz (second supervisor).

This dissertation contains material appearing in the following articles:

- ...

In addition to the above, I have contributed to the following publications during the course of my PhD:

- ...

This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and acknowledgements. It has not been submitted in whole or in part for any degree or diploma at this or any other university.

Irene Marzuoli
September 2019

Acknowledgements

...

Summary

Elucidating self-assembly and antimicrobial strategies of
synthetic peptides an in silico investigation

Irene Marzuoli
King's College London

... ...

Contents

1	Introduction	10
1.1	Antimicrobial resistance	12
1.2	Alternative antibiotic strategies: antimicrobial peptides	18
1.3	Gene therapy	33
1.4	Closing the circle: an antimicrobial drug delivery vehicle	36
1.5	A computational approach to understand capzip	43
2	Methods	45
2.1	Algorithms for Molecular Dynamics	46
2.2	The force field problem	49
2.3	The search problem	58
2.4	The ensemble problem	61
2.5	The experimental problem	62
2.6	MD simulations: successes	64
3	Capzip simulations	77
3.1	Modelling the assembly	78
3.2	Simulations details	82
3.3	Analysis	88
3.4	Results: capsule in solution	90
4	Lipid parametrisation	99
	Appendices	100
A.1	On the derivation of the GP predictive distribution	100
	Bibliography	101

List of Figures

1.1	Graphical abstract of introduction	11
1.2	Mechanisms of antimicrobial resistance to small drugs	14
1.3	Antimicrobial peptides	19
1.4	Principles of gene therapy	34
1.5	Cazip molecule	37
1.6	Experimental results on capzip	41
3.1	Hydrogen bonds between RRWTWE β -sheet	78
3.2	Building blocks of capzip assembly	79
3.3	Cohesion measures on the pentagonal subunit	81
3.4	Snapshot of simulations of capzip on model membranes	82
3.5	Atomistic run of buckyball in solution: final configuration	90
3.6	Structural measures on buckyball in solution	91
3.7	Correlation of motion between molecules of the buckyball	92
3.8	Contacts between molecules during simulations of the buckyball	93
3.9	Hydrogen bonds in the buckyball molecule	95
3.10	SASA per residue of a buckyball in solution	96

Chapter 1

Introduction

“**P**HILosophical INTRODUCTION” to be finished/modified when the work is finished...

(...thus) This introduction is meant to give an overview of the many different challenges the fields of medicine and bioengineering have faced in recent years. These challenges have promoted the research on self-assembling antimicrobial peptides, despite they were not a primary source of interest in these fields, as other materials and concepts were deemed more suitable to solve the tasks coming along the way. It is therefore important to clarify the landscape of such other solutions and approaches to understand and value why a change in the research focus has come to age. Figure 1.1 provides a work flow of this introductory chapter to help the reader in identifying the sections of interest.

Motivations of the work: a graphical abstract

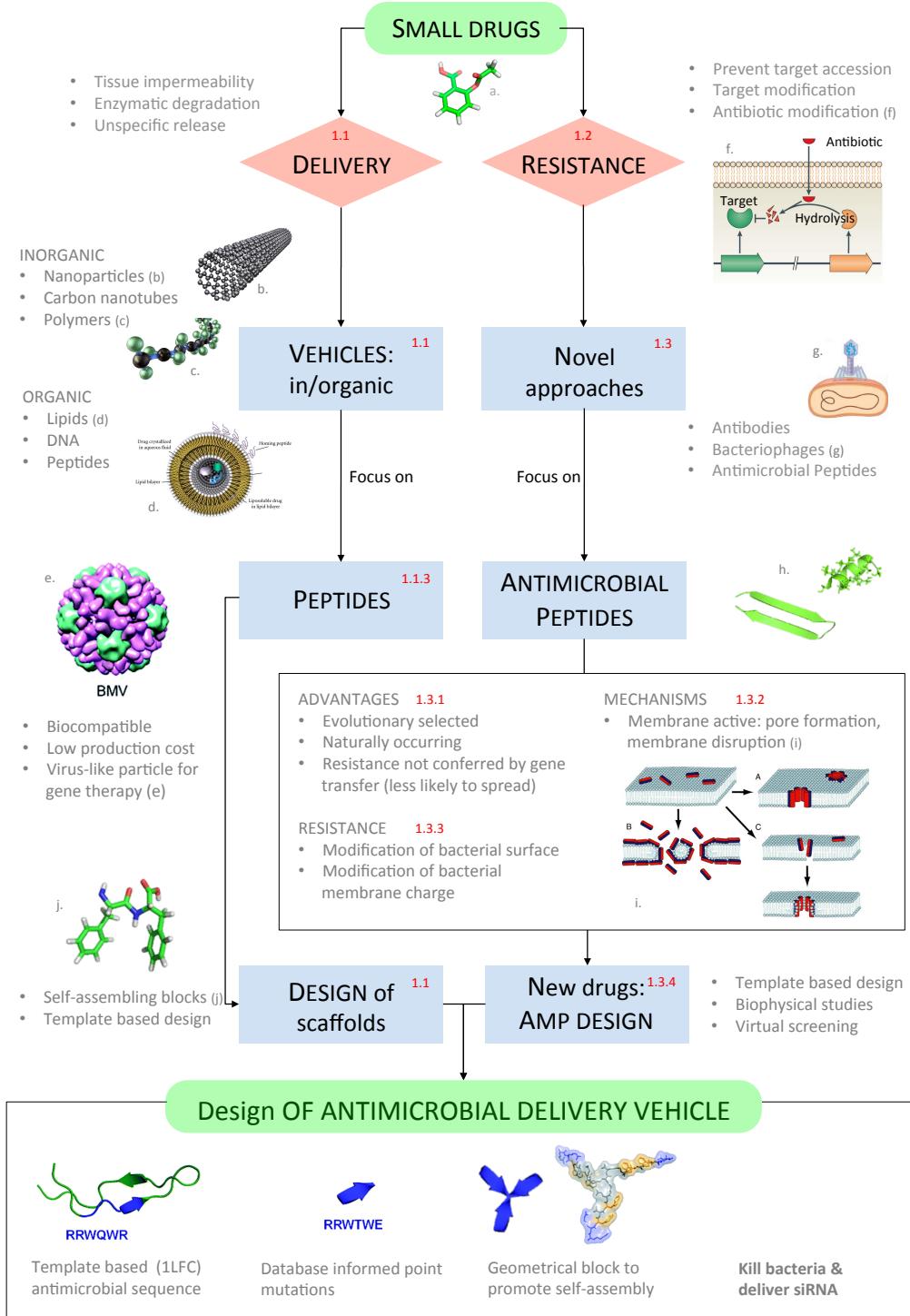


Figure 1.1: Figures a. (acetylsalicylic acid) and j. (diphenyl-alanine) in bond representation. Remaining figures adapted from: b. [-]; c. [1]; d. [2]; e. [3], f. [4]; g. [5]; h. [6]; i. [7]; k. [8]

1.1 Antimicrobial resistance

For most of the last century, the development of new drugs and of delivery agents for such drugs rotated around the paradigm that a drug is a small inorganic compound (of mass up to 900 Da) which targets a specific molecule in a specific target of a mammal or bacterial cell. In its journey to its target, it is necessarily helped by some delivery vehicle which carry the compound to the site of action where it can interfere with the process it is assigned to, least it gets degraded beforehand. Very often the target of interest of small drugs are proteins: out of the 695 small drugs approved by FDA (the American Food and Drug Administration agency) to target human molecules, 667 acts on proteins. Similarly, 189 of the 198 small drugs approved to treat pathogens have a protein as their target [9] (with all the caveats coming from the difficulties that identification of an unambiguous drug target poses, especially when the drug binds to a protein complex or to a number of closely related gene products [9]).

In presenting the aforementioned figures, the data were naturally split among the drugs which target human molecules, “repairing” some faulty process in the human body, or the ones active against bacteria, which “disrupt” the bacterium life cycle in order to kill or prevent the reproduction of the pathogen, commonly named as antibiotics. It appears evident that the pool of drugs available to the second purpose are in consistently lower number than the ones addressing human molecules. This comes from the nature of the action they perform: molecules targeting human proteins need to be highly specific to avoid interference with other proteins or with healthy cells, and in a sufficient number to address the variety of diseases affecting the human body. Antibiotic must be non-toxic for human cells as well, i.e. their target must not be shared between mammal and bacterial cells, but there is a less stringent requirement on their selectivity against different bacterial species. On the contrary, it is often useful to have a broad-spectrum compound. This cross-species efficacy and non-toxic property is obtained thanks to the evolutionary relationship among bacterial species, and between bacteria and humans: while the first are closely related, and therefore share homologous proteins with very similar structures, humans have less architectures in common with them, allowing for a resilience against bacteria-targeting drugs. Nevertheless, the set of bacterial

species is very diverse and the cross-species effectiveness of some drugs does not extend to the whole bacterial population. This is actually demonstrated to be a positive feature, given the large amount of beneficial bacteria that live in symbiosis with the human body (especially in the gut) and that must be preserved for an optimal wellness.

In the framework described above, it is understandable that first-time research on antibiotics was satisfied with the development of a handful of potent, broad-spectrum compounds. Penicillin, the first of them to be synthetically produced, was isolated from a mould in 1928 by Alexander Fleming. It acts inhibiting the formation of peptidoglycan cross-links in the bacterial cell wall and preventing its complete formation, through binding to the enzyme DD-transpeptidase responsible for the catalysis of such cross-link [10] (for further details on bacterial cell membrane structure see Section 1.2.1). As foreseen from Fleming himself in his Nobel Prize acceptance speech, bacteria can become immune to penicillin, and this is achieved in many ways: either by production of penicillase, an enzyme that degrades penicillin, or by subtle changes in the structure of the penicillin-binding proteins to prevent penicillin binding, or again by removal of the drug outside of the cell through specially re-purposed efflux pumps [?].

The mechanisms just outlined are not an exceptional characteristic of penicillin, and many drugs lost their effectiveness against some bacteria since their discovery till nowadays. By now, a broad knowledge has been gathered on how bacterial escape the action of a drug: this understanding helps interpreting the pitfalls of existing drugs and identifying the characteristics sought in the developments of new compounds.

1.1.1 Mechanisms of antimicrobial resistance to small drugs

Antimicrobial resistance can manifest through many different mechanisms, which can be grouped in three main classes, in line with the three processes mentioned in the example of the penicillin resistant bacteria.

Prevention of access to target A first class of resistance mechanisms aim at minimising the intracellular concentration of the antibiotic preventing its penetration or maximising its efflux in the eventuality it has entered the cell (Figure 1.2(a)). Not all the molecules can enter the cell permeating the mem-

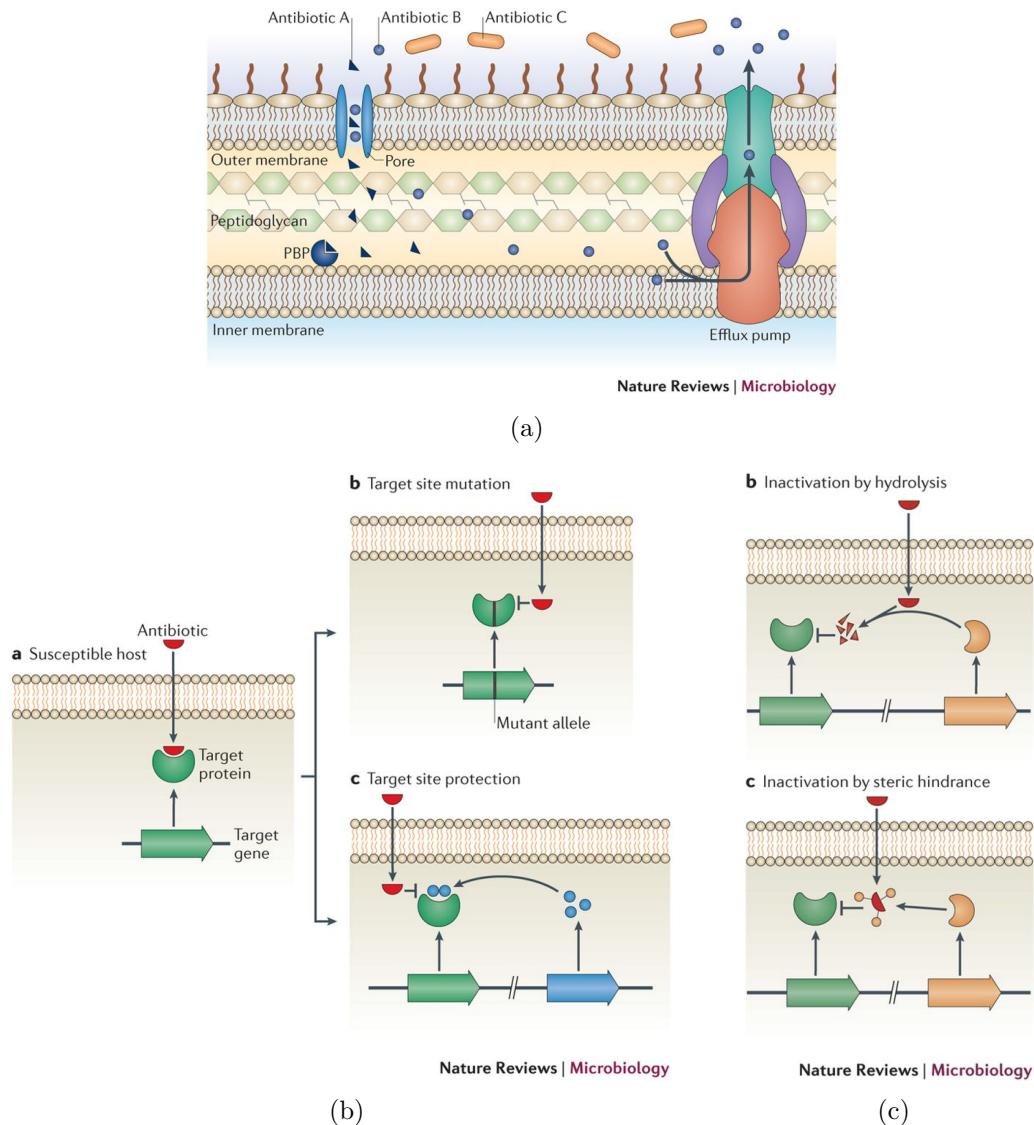


Figure 1.2: Mechanisms of antimicrobial resistance to small drugs. (a) Intrinsic mechanisms of resistance (removal of antibiotic B by efflux pump and inaccessibility of antibiotic C to the PBP target because of membrane impermeability). (b) Target site change via mutation or protection. (c) Direct interactions with antibiotics causing its disruption or structural modification. Reproduced from [4].

brane, and this holds particularly for hydrophilic antibiotics tackling Gram-negative bacteria which are intrinsically less permeable than Gram-positive ones because of the additional presence of the outer membrane [11] (for further details on the bacterial membrane structure, see Section 1.2.1). These

molecules must then be imported into the cell through outer-membrane porin proteins [12, 13]. Resistance arise when porins are either replaced with more selective channels, which prevent the antibiotic penetration, or down regulated so the internal concentration of the drug does not reach a critical concentration [14]. Porin-coding genes can also accumulate multiple mutations, to acquire the selectivity they lack by nature [15].

A complementary strategy to prevent drug influx is to employ bacterial efflux pumps. Some of them are denominated multidrug resistance (MDR) efflux pumps for their effectiveness in the task and are produced by many bacteria [16, 17], but, additionally, can be transferred via plasmids to other bacterial species [18]. Indeed, bacteria are able to exchange genetic material with other individuals via small rings of DNA in a process called conjugation [?], so that the advantageous resistant genotypes can spread quickly across species. Over-expression of such efflux pump is observed in multidrug-resistant bacteria, triggered by exposition to the drug, and proceeding via mutation in the relative regulatory network, [19], or simply as a response to environmental signals [20].

Change or modification of the antibiotic target The second class of resistance mechanisms works modifying the antibiotic target: most antibiotics bind to their substrate with high affinity and specificity, thus small modifications in the target structure can disrupt an efficient binding of the antibiotic, still allowing the target to maintain its normal function (Figure 1.2(b)).

Mutations of some residues in the binding pocket (upon mutation in the gene coding for it) or post-translational protection of the target via addition of chemical groups are equally wide spread mechanisms. Notable examples of the first include the development of methicillin and linezolid-resistant strains of *S. aureus* [21, 22]. Again, it is interesting to notice that several of these mutations are acquired by horizontal gene transfer from other bacterial species, so that resistance development in one specie promotes quickly the insurgence of resistance in other ones. The most relevant mechanism of chemical group addition is methylation, which for example is very common when the target are rRNA subunits [23]).

Direct modification of antibiotics Finally, bacteria can destroy drugs, usually by hydrolysis, or modify them by transfer of a chemical group (Figure 1.2(c)).

The first example historically discovered of drug-degrading enzyme is penicillinase (a β -lactamase) which destroys penicillin [24]. Since then, thousands of similar enzymes have been identified that can modify antibiotics of different classes [25, 26]: these enzymes co-evolve with newly developed drugs, to include in their spectrum of action new compounds of similar composition to the ones they were originally effective on [27].

Antibiotics constituted by large molecules with many exposed hydroxyl and amide groups are particularly susceptible to modification by addition of chemical groups instead. Many enzymes are responsible for this, and according to the chemical moiety added they are grouped in acetyltransferases, phosphotransferases and nucleotidyltransferases [28].

All together, the recent progress in understanding the mechanisms of antimicrobial resistance has helped in directing the development of new drugs, in particular the modification and the improvement of existing compounds to escape the resistance developed by bacteria. However, to ultimately employ the existing drugs at best, it is important to understand also the dynamics of AMR, beyond its molecular mechanisms, to devise the most effective strategies of drug administration.

1.1.2 Course of antimicrobial resistance

In the first stages of the insurgence of antimicrobial resistance (AMR) against a given drug, some strains of bacteria are not damaged by the standard doses of the drug as they came to possess some natural occurring mutations in their genome which promote an escape mechanism invalidating the drug effectiveness [4, 29]. Usually, only a small population of bacteria is resistant, in the first moments, however the resistant population will replicate faster than the peers of the same species because it is more fit in an environment challenged by the presence of the drug. It is noteworthy that this fitness might not be optimal in a natural drug-free environment - and indeed the wild population has not been selected for that genotype - but under the pressure derived from the treatment, other characteristics result more advantageous. In the short

time scale it is usually sufficient to increase the doses of a drug to re-gain efficiency against the target, but it has been observed that the resistant species can usually adapt to higher doses of the same [?]. Moreover, high drug doses are not always applicable due to the severe side effects they are connected to.

As already mentioned, the spread of resistance between bacterial cells and even between species is very effective as bacteria exchange genetic material through conjugation. Therefore, despite AMR is an evolutionary mechanism, the fast pace at which bacteria replicates, their enormous population (in terms of individuals), and the relative easy horizontal gene transfer through conjugation, place the insurgence of resistance well within the human lifespan time scale [?].

It is then clear that this complex problem depends on many variables: the casual appearance of resistant individuals, the transfer of information between them, the gain in fitness of resistant individuals but also the dosage and time line of the drug administration. Many mathematical models have been implemented to understand the issue [30, 31], but it is known that some particular strategies of drug administration are worse than other, favouring the proliferation of so called “super” bugs. One example of bad administration strategy is the underdosage of antibiotics: a low drug load is likely to harm but not kill pathogens, and rather promote the fitness of resistant ones. In a sort of “gym” or “vaccination” process for bacteria, an underdosed drug would kill the weakest individuals but strengthen the resistant population, which would now be fitted to the challenges of higher doses [?]. Similarly, the abuse of antibiotics puts an high pressure on the pathogenic populations, which is desirable but at the same time can induce a faster emergence of escape mechanisms [?]. In this context it must be noticed that many drugs are bacteriostatic agent as opposed to bactericidal: i.e. they prevent the bacterium growth rather than kill it, as they are meant to control the bacteria reproduction and slow down the damage while host defence mechanisms eradicate them. Thus if an high dosage of a bactericidal agent may extinguish the bacterial population and eradicate the disease, for bacteriostatic drugs, once they are removed, bacteria start again the reproduction cycle.

In this context, it is clear that the antibiotic landscape is a dynamic entity in which newly discovered ones enter, while others exit after having been exploited for years, and must be monitored carefully and kept populated: it is

noteworthy that the severity of the AMR issue is such that it has been raised to the status of national emergency in several countries, including UK. Indeed, abuse or misuse of antibiotic can take many forms and strict regulations on the health, agricultural and food industry sector must be taken to prevent its loss of efficacy, as we are leaving the century in which antibiotics were discovered, to enter a phase in which we count the number of the ones loosing efficacy [32].

1.2 Alternative antibiotic strategies: antimicrobial peptides

In the landscape sketched above, it is evident that the development of novel drugs is of crucial importance. Even more beneficial would be to have at disposal a new paradigm for their design, in order to attack pathogens in a completely novel way, avoiding to target pathways which are known to lead easily to the development of antimicrobial resistance. Several novel materials have been developed for the task, not to rely on small molecules and exploit different mechanisms of action, for example antibodies, bacteriophages or antimicrobial peptides [33].

While the use of pathogen-specific antibodies relies on the mechanisms of the host immune system, bacteriophages therapy employ viruses which infect bacteria and archea rather than eukarya. But are antimicrobial peptides the focus of this thesis: indeed peptides can have an active role against bacteria when their sequence possesses some specific characteristics. Such active sequences, capable of damaging and/or killing pathogens, are referred to as antimicrobial peptides. The following subsections will explore their characteristics, modes of action and the response of bacteria against them. It is crucial to understand the knowledge available versus the questions that are still open in order to direct the efforts of future research. This holds in particular when the investigation proceeds by the use of simplified models, as meaningful results can proceed only if such modelling is performed in a sensible and informed fashion.

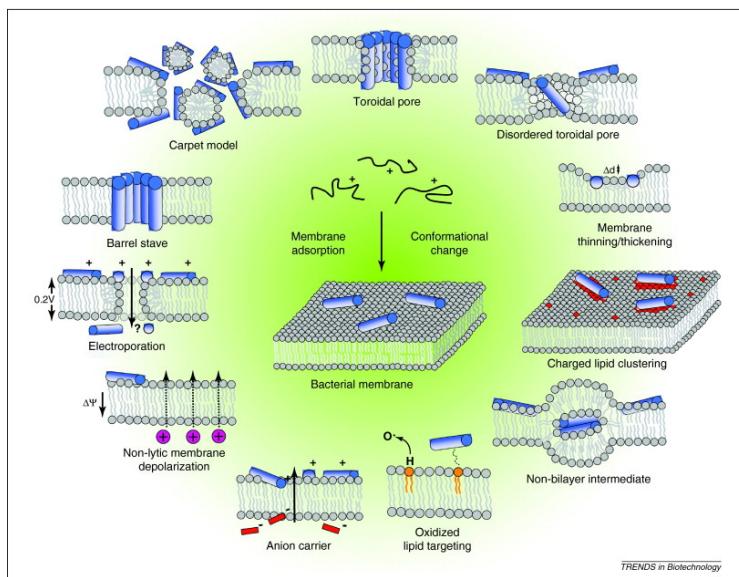


Figure 1.3: Events occurring at the bacterial cytoplasmic membrane following initial antimicrobial peptide (AMP) adsorption. Reproduced from [7].

1.2.1 Membrane active peptides

Antimicrobial peptides (AMPs) are naturally produced by eukarya, either as stand-alone sequences or embedded in larger proteins, as a first, weak, and broad-spectrum defence against bacteria [7]. This pool of molecules has been selected through evolution to be active against pathogens, suggesting that they are weakly prone to provoke resistance reactions in the microbes they attack.

To exploit their potential and engineer AMP-like molecules, a careful characterisation and classification of such peptides must be done. This task has been carried on throughout the past decades but it is complex, so that up to date there are many peptides with ascertained antimicrobial activity for which the mode of action is still not fully understood [34]. However, some general characteristics of these sequences and some of the mechanisms they employ have emerged. Unsurprisingly, AMPs are heterogeneous in shape, targets and mode of action, to tackle the different challenges bacteria pose. Their size can vary between 6 and 59 amino acids [35]: despite being small with respect to the average size of a protein in the human body, these macromolecules are hundreds of times larger than small molecule drugs and as such they penetrate and act on bacteria differently with respect to small compounds.

The most common target of AMPs is the bacterial membrane. Many of them cause disruption of the physical integrity of the microbial membrane while others translocate into the cytoplasm to act on intracellular targets, and the combination of the two is not uncommon either [36] (Figure 1.3). In general, it is widely accepted that membrane interaction is a key factor for the direct antimicrobial activity of AMPs [7].

As such, we propose a brief overview of the structure of such membrane [?], and of its differences with the mammalian cell one.

Structure of bacterial membrane The determinant driving the interaction between AMPs and bacterial membranes is the positive charge that many AMPs present, opposed to the negative charge of the latter. It is striking that such simple mechanism, based on the presence of a certain number of negatively charged lipids, holds across many bacterial species despite the great variability found in their membrane composition. Indeed, based on the differences in their cell envelope structure, bacteria are classified into two macro families, Gram-positive and Gram-negative. In Gram-positive bacteria, the cytoplasmic membrane is surrounded by a thick peptidoglycan layer, while for Gram-negative bacteria this membrane (which assumes the name of internal one) is surrounded by a thin peptidoglycan layer as well as an outer membrane [37].

Starting from the inside and proceeding outwards, the cytoplasmic membranes of both Gram-positive and Gram-negative bacteria are rich in phospholipids like phosphatidylethanolamine, which is neutral, and phosphatidylglycerol, cardiolipin, and phosphatidylserine, which have negatively charged headgroups, highly attractive for positively charged AMPs. This is often sufficient to promote the preferential interaction between this membrane and the peptides. Perturbation of this membrane is highly disruptive because many functions are associated to it: as bacteria do not possess organelles, all the membrane related proteins reside and perform their function on the inner membrane.

In the case of Gram-negative bacteria, between the inner membrane, together with the outer one, delimit the periplasm space, an aqueous cellular compartment, which allow to sequester harmful substances and allow nutrient transportation. Proceeding outward, inside the periplasm is situated the pep-

tidoglycan cell wall. Repeated units of this substance, made of a disaccharide cross-linked by penta-peptide side chains (from which the name), constitute the rigid skeleton of the cell. Damage in the peptidoglycan layer results usually in living but viable cells, therefore it is fundamental for the cell life.

Finally, the outer membrane is grafted to the skeleton by Braun's lipoproteins. This membrane presents an asymmetric structure: phospholipids are present in the inner leaflet, while the outer one is composed of glycolipids, mainly lipopolysaccharide (LPS). This complex molecule consists of lipid A, which presents multiple fatty acids, and a polysaccharide, made of an inner core, covalently bonded to the lipid, an outer core attached to the inner and finally a repetitive glycan polymer (O-antigen). The O-antigen is the molecule exposed by Gram-negative to the external environment and thus is the target of antibodies recognition.

Given the complexity of the Gram-negative cell envelope, and especially the presence of the LPS layer, these bacteria are particularly impermeable to hydrophilic molecules, which are usually imported through porins and similar transmembrane proteins.

For Gram-positive bacteria, the inner membrane is enveloped in a thick peptidoglycan layer. If its thickness in Gram-negative bacteria is about a few nanometers, in Gram-positive it spans from 30 to 100 nm. Its composition is similar to what already stated, with some variations present among different bacteria, for example in the nature of the peptidic linker or in its precise position. This thick layer is threaded by long anionic polymers (the teichoic acids), mainly composed by glycerol phosphate, glucosyl phosphate, or ribitol phosphate repeats. Finally, disseminated in this layer there are several surface proteins with various functions, among which adhesins, which attach to components of the host extracellular matrix.

Structure of bacterial membrane The fact that AMPs tackle negatively charged membranes is crucial for their selectivity, i.e. the fact that they are harmless for the mammalian cells they are produced from [38]. Indeed, mammalian cells have a different membrane composition: they present a single membrane, rich in protein. The lipidic component of the membrane is composed by many different lipids, and is abundant in zwitterionic phospholipids such as phosphatidylethanolamine, phosphatidylcholine, and sphingomyelin,

providing a neutral net charge [39, 40]. Strictly speaking, some negatively charged lipids are present in a few mammal cell types, however they are located in the inner leaflet, while the zwitterionic phospholipids are more abundant in the outer leaflet, in an asymmetric composition [?]. This structure promotes weaker interactions between AMPs and the mammalian cell membrane with respect to the bacterial one, as the former is driven mainly by hydrophobic interactions, while the latter by electrostatic ones. Furthermore, the mammalian cell membrane has a high content of cholesterol [41, 42], a sterol fat, which is proposed to stabilise the membrane regulating its fluidity across the differences in physiological temperatures, but it is also thought to favour a better accommodation of the perturbations caused by AMPs [43]. Finally, the mammalian membrane contains a small percentage of carbohydrates, mainly embedded in glycoproteins, which promote the cell-cell recognition.

Another relevant difference between bacterial and mammal cells is that the first ones have a typical transmembrane potential - the difference of electrostatic potential between the inside and the outside environment - between -130 and -150 mV, while mammalian cells between -90 and -110 mV [41, 44, 45]. Given that a potential generates an electric field across the membrane, the higher the potential, the higher the electric field pointing from outside to inside the cell. A field in such direction pushes cationic compounds on the outside of the membrane toward the membrane itself. Therefore the stronger bacterial transmembrane potential may promote an enhanced - and thus disruptive - interaction of AMPs with the cell, contributing to the selectivity of AMPs between bacteria versus mammals [41].

1.2.2 Common mechanisms of action of AMPs

Investigating the perturbation and disruption of a bacterial membrane by antimicrobial peptides is a key point of this work, therefore it is important to highlight the mechanisms known so far through which AMPs reach this outcome. As already mentioned, many AMPs have a positive charge which facilitates the binding to the membrane via charge-charge recognition; accordingly, Arginine and Lysine residues are usually abundant in AMPs sequences. However, the disruptive action takes place through the interaction of the AMP with the hydrophobic core of the membrane, therefore their sequence contains also hydrophobic aromatic residues, especially Tryptophan, which favours the

anchoring to the lipid core [46]. Overall, AMPs resort often to adopt an amphiphatic structure to segregate the hydrophilic from the hydrophobic amino acids and thus act at the interface between membrane and solution. It is interesting to notice that some of them fold into the active structure only nearby the membrane, as they can expose their hydrophobic components to face its core, while in solution these ones are preferentially buried inside to be screened from the solvent [7]. Common folds adopted by AMPs are both α -helix or β -sheet rich structures. Amphiphatic α -helices present a charged side which is tailored to face towards the phospholipid head groups and an hydrophobic ones which is favourably buried into the acyl chains core, and a similar arrangement is found for structures rich in β -sheets include β -hairpins.

Membrane disruption Several models have been proposed to describe the exact mechanisms of AMPs penetration after they bind to the cytoplasmatic membrane, and how their combined action leads to membrane permeabilization (Figure 1.3) [7, 35].

For example, for a single copy of a amphiphatic helical AMP, the proposed mechanism of action suggests that initially the peptide is attracted with its charged side to the membrane and lies parallel to its plane, with the hydrophobic side unfavourably exposed in solution. Then the helix rearranges to have the two faces in the respective favourable regions. Subsequently, the helix axis starts to form an angle with the membrane plane, and finally inserts deeper into the lipid core, often spanning the full membrane thickness. Similarly, for β -sheet rich structures, it is suggested insertion within the membrane after a first flat approach. The final insertion arrangement depends on the peptide characteristics and length, the presence of kinks in its structure (in case of helices), and the interactions with other copies of the peptide.

The picture becomes more complex for oligomer-mediated insertion, i.e. when the action is triggered by the combined action many copies of the peptide. At low peptide to lipid ratio, the favourable configuration is represented by peptides lying parallel to the membrane plane as described previously [47], but an increase in peptide concentration triggers the transition to an inserted state where the main axis of the AMP is perpendicular to the membrane. The organisation of AMPs inside the membrane core can assume different configurations, as described below.

The “barrel-stave” model proposes that AMPs insert perpendicularly into the bilayer. Recruitment of peptides in the same area results in the formation of a transmembrane pore with a central lumen. The walls of the pore are constituted by the hydrophilic face of the peptides, while their hydrophobic side is interacting with the lipid tails around the pore. This model is adopted for example by the α -helical AMP alamethicin, which forms voltage-dependent ion channels by aggregation of four to six molecules [48].

In the “toroidal” pore model instead, the insertion of peptides forces the phospholipid to bend continuously from one leaflet to the other, resulting in a pore defined by both peptides and phospholipids head groups. The toroidal model differs from the barrel-stave model as the peptides are always associated with the lipid head groups even when they are perpendicularly inserted in the lipid bilayer. Toroidal pores are induced by α -helical magainins, protegrins and melittin [47, 49, 50], and lead to membrane perturbation which extends further away from the pore than in the barrel-stave case, as lipids must rearrange around them.

As a comparison, alamethicin induced barrel-stave pores have an inner and outer diameters of 1.8 nm and 4.0 nm respectively [48], while magainin-induced toroidal pores are larger and can vary in their size, with an inner diameter of 3.0-5.0 nm and an outer diameter of 7.0-8.4 nm, involving about 4 to 7 magainin monomers and about 90 lipid molecules [51].

Finally, in the “carpet” model, the accumulation of AMPs on the surface of the membrane, laying parallel to it, causes tension in the bilayer and the membrane is then disrupted by peptides in a detergent-like manner, leading to the formation of micelles. The critical threshold concentration triggers a cascade effect, in which formation of the first disruption allows the penetration of AMPs in the inner side of the bilayer. The cooperation between peptides on both sides of the lipid membrane enhances the AMP-induced curvature on the membrane causing accelerated disruption. The “carpet” model mechanism is observed for peptides presenting an α -helical structure (like melittin [52]) or two or more helices connected by short loops (ovispirin [53]).

The prevalence of examples with an helical structure for the above models derives from the fact that the understanding of how helical AMPs function is often easier than the one of β -sheet rich structures. Indeed, helices have a well defined fold (at least nearby the membrane environment), a compact

structure, and often a clear segregation of complementary patches that can attract other copies of the peptide and thus promote the self-assembly process necessary for the pore formation.

On the contrary, many β -sheet AMPs have a more flexible structure, and more diversified mechanisms of action [?]. AMPs rich in β -sheets can be divided into β -hairpins and peptides from the defensin family [7]. Many representative of the former class disrupt bacterial membranes via formation of toroidal pores: as an example, porcine peptide protegrin I triggers the toroidal pore formation assembling into a β -barrel structure when in contact with anionic membranes (while it folds into β -sheet aggregates on the surface of cholesterol containing membranes, thus acting selectivity on bacterial membranes only [54]).

In the case of defensins, many mechanisms are known according to the specific member of the family [55]. Some form transmembrane pores on planar bilayer when a physiologically relevant negative potential is applied to the membrane [56], while others form oligomers in phospholipid vesicles [57]. Although various descriptions of membrane damage have been reported, and include ion channels, transmembrane pores and extended rupture of the membrane, they are likely related, being a modulation of a similar acting principle.

Alternative mechanisms of action Finally, many non-lytic mechanisms are suggested for AMPs, especially for β -sheet structures: defensin A from *P. terramovae* reduces the cytoplasmic potassium concentration [35], partially depolarising the inner membrane; tachyplesin from horseshoe crabs is able to bind to the minor groove of DNA, interfering DNAprotein interactions [58], and bovine lactoferricin can act synergistically with other antimicrobial agents by affecting the transmembrane potential and proton-motive force, resulting in inhibition of ATP-dependent multi-drug efflux pumps [59]. Moreover, after translocation within the cell, bovine lactoferricin can also inhibit DNA, RNA and protein synthesis. Section 1.4.1 will treat in detail the functioning of this AMP, distinguishing its role as membrane active peptide versus intra-cellular targeting compound: indeed, many works have focussed on lactoferricin antimicrobial processes versus locating the section of the sequence performing the membrane disruptive activity [60, 61], to understand whether it retains the efficacy regardless of the fold.

These and similar strategies of investigations, conducted on several AMPs [?], provided the discovery of first minimal functioning antimicrobial blocks, which promoted the understanding of how AMPs work in general, boosting the design of synthetic tailored AMPs from specific sequences.

1.2.3 Mechanisms of resistance to AMPs

Antimicrobial peptides are introduced here as a class of new drugs and a possible solution to the crisis of antimicrobial resistance. Any new drug entering the pool of the clinically approved compounds is (at least temporary) a solution to the problem of resistance to known antibiotics, but it must be clarified that bacteria can develop resistance to AMPs too. Never the less, the resistance to their action is generally not based on dedicated genes that are conferred by horizontal gene transfer, as in the case of many antibiotics resistance mechanisms [62]. Because of that, a certain increase of resistance after exposure to the drug is to be expected ('MIC creep'), but it is less likely to spread quickly to other species.

Some of the mechanisms of AMPs resistance are similar to the ones employed by bacteria to counteract small molecule drugs, for example over-expression of efflux pumps to dispose of AMPs, and degradation of the peptide by extracellular enzymes and sequestration by the bacterial or biofilm matrix to prevent accession to the target [62].

Differently with respect to antibiotics hydrolysis, AMPs proteolitic degradation is operated by proteases, proteins themselves secreted on the extracellular side of the membrane to destroy other proteins. Linear AMP are more prone to this type of degradation [63], as opposed to the ones presenting disulfide bonds [62], such as defensins, which nevertheless can be hydrolysed by more specific proteases enzymes [64].

Bacteria can also enhance their resistance to AMPs organising into specialized structures known as biofilms. These structures are formed by sessile bacteria adhering to a surface in an organized manner that allows the circulation of nutrients [65]. Biofilm bacteria secrete an extracellular matrix with adhesion and protection functions [66], which effectively repels and/or captures AMPs through exopolysaccharid or capsular polysaccharides. For example polysaccharide intercellular adhesin (PIA) produced by *S. aureus* and a variety of other bacteria is responsible for the resistance to both cationic

AMPs (HBD-3, LL-37) and the anionic one dermcidin [67]: deacetylation of PIA increases its positive net charge, thus repelling more efficiently cationic AMPs, and increasing sequestration of the anionic ones at the same time, as well as forming a mechanical barrier for both of them [68].

It's worth noting that AMPs are nevertheless a promising alternatives in the treatment of biofilm-associated infections: indeed, in this type of infections (where bacteria are growing slowly) it is advantageous to have bactericidal agents such as AMPs, as opposed to bacteriostatic ones which target fast-growing bacteria only, as the majority of traditional antibiotics are [69].

But the most specific mechanism of AMPs resistance concerns modifications of the bacterial cell envelope: bacteria modify the characteristics of their surface to prevent the efficient action of an AMP, even in the eventuality that the peptide reaches the bacterial envelope intact. The target of such modifications are different for Gram-positive and Gram-negative bacteria, according to their distinct cell envelopes Gram-positive bacteria change teichoic acids (TA): for example, D-Alanylation of TA observed in *S. Aureus*, adds a positive charge to it, reducing the attraction of cationic AMPs and in turn increasing the cell wall density, so reducing the surface permeability [70]. Alternatively they can modify the bacterial peptidoglycan precursor, lipid II, which has a key role in the formation of the cell wall and is often targeted by AMPs. For example, the replacement of its terminal D-alanine with D-lactate or D-serine [71] avoids the action of the glycopeptide vancomycin as the functioning of this molecule proceeds by binding to the D-Ala-D-Ala terminal moieties of the precursor.

In Gram-negative bacteria a positive charge can be added to lipopolysaccharides (LPS) by addition of amine-containing molecules [72] or by removing phosphate lipids (which have a negative charge) from lipid A, one of the constitutive moieties of LPS [73], to obtain the same effects as for the increased charge in TA. They can also enhance the rigidity of the outer membrane to reduce permeability to AMPs via addition of extra acyl chains into lipid A [74], and finally they can act at the cytoplasmic membrane level, as this is the final target of many antimicrobial peptides. In the eventuality that AMPs successfully pass the cell wall and reach this membrane, they are attracted to its surface by the negative charge of the lipids composing it, in particular, phosphatidyl-glycerol (PG) and diphosphatidylglycerol (DPG, also called cardiolipin). Their negative charge can be masked by amino-acylation of the PG

head group, so that the final compound repels AMPs through electrostatic interaction [75], or the overall rigidity of the cytoplasmic membrane can be enhanced, by an increase in saturated acyl chains which has been proven to confer resistance [76].

Finally, resistant bacteria often employ many of the aforementioned strategies at the same time, for example modification of the surface charge together with modification of other membrane components for a decreased recognition and augmented rigidity [77].

1.2.4 Principles of AMP design

The study and classification of AMPs provide knowledge on the characteristics a sequence must have to perform an antimicrobial function. As discussed in Section 1.2.2, there are some features which, comprehensively, help in discriminating AMPs against non antimicrobial peptides. The constantly increasing amount of data available is gathered in several curated databases [78–82], which catalogue AMPs (or subclasses of them, like membrane active, biofilm active or haemolytic peptides) based on such features to promote future data-driven prediction of the antimicrobial character of a sequence.

We recapitulate below a few key characteristics which are peculiar of AMP sequence. To be noticed that while some are easily retrievable from the sequence of the peptide, other imply direct experimental measures, and thus are difficult to implement into a simple sequence-based method to predict activity:

- **Structure:** both α -helical and β -sheet rich AMPs exist, as well as mixed structures. Short helix (~ 22 amino acids) and short β -sheet (~ 10 amino acids) are particularly common, acting through slightly different mechanisms (when known). When screening a potential AMPs, it must be considered that the fold close to a membrane environment might be different with respect to the one in solution.
- **Charge:** AMPs are charged moieties. Usually cationic (up to $\sim +10\text{ e}$), with fewer anionic examples (like dermcidin). Among cationic ones, not all the positive amino acids have equal role, for example Arginines are more effective than Lysins [46]. The potency, but also their haemolytic activity, are often directly related to the amount of charge [?].

- **Hydrophobicity:** AMPs contain also hydrophobic residues, usually with abundance of aromatic chains and specifically Tryptophan, as they must insert and anchor into the lipid core of membranes, which is an hydrophobic environment.
- **Amphipathicity:** to host both the charged and hydrophobic residues, most AMPs organise themselves in an amphiphatic structure, i.e. the two types of amino acids side chains are located on the opposite side of the peptide.
- **Solubility** AMPs need good solubility to prevent aggregation in the aqueous environment they float in before arriving to the membrane. Aggregation would most likely impede their optimal interaction with the membrane.

A part from predictions regarding existing sequences, the knowledge of AMPs sequence-activity and structure-activity relationships is beneficial to find new, better performing ones. The design of new AMP sequences aims at producing peptides with improved characteristics:

- **specificity** against particular bacterial species;
- **stability** against the action of proteases, thus allowing a longer residence time in the body;
- **low cytotoxicity** at the therapeutic dose required (so an high therapeutic index).

The need for such improved peptides lies in the fact that their natural form constitutes a first broad spectrum defence our body employs against infectious bacteria and thus AMPs are often of mild potency. However, foreseeing their application as future drugs, it is desirable to tailor them to fulfil different criteria according to the infection to treat. At the present state of the art, a golden rule for the design of such sequences is still missing, however several methodological approaches to AMP design have been explored, and they can be grouped in three main lines: template based studies, biophysical studies and virtual screening [83].

Template based studies The main idea behind template based methods consists in modifying existing antimicrobial sequences in the direction of the desired characteristics. The most widely explored templates are helical peptides, for their short sequences and because several of them (cecropin, magainin and protegrin) have been well characterised [84].

Ideally, an amino acid scanning of all the residues in an AMP provides information on the role of each of them, prompting at the most suitable mutations. High-throughput methods allow nowadays for such thorough investigation in the case of short AMPs [85], but similar, less resource consuming Alanine scannings can still point at the most important residues for the antimicrobial activity [86]. Alternatively, simpler approaches aim at designing peptides with enhanced charge and amphiphilicity, as these characteristics are deemed crucial for their effectiveness (see the paragraph above) [84].

However the above methods focus on single amino acids and can not take into account the interplay between residues, nor the three dimensional structure of the peptide. Without such information, it is difficult to extract general rules on why some mutations work better, and often the results of these studies give indeed enhanced AMPs, but cannot be generalised to other sequences. Therefore, recent works sought to integrate structural information on template based models, designing peptides active against many bacterial lines at the same time [87] or to enhance selectivity for bacterial membranes [88].

A complementary approach to single point mutations on known peptides consists in designing minimal antimicrobial blocks: several investigations proved the importance of single residues and their intercalated pattern in natural and designed AMPs. For example, natural AMPs are rich in Tryptophan and Arginine residues [46], and synthetic ones have been produced with only Lysines-Leucine, or Arginine-Valine combinations to produce amphipathic helices [89]. An effort to extract principles from these examples is represented by text based model where amino acids constitute the letters and patterns occurring in natural AMPs are the grammar rules [90].

In general, the advantage of template based methods is in the reduced number of sequences to test, with decreased cost, as only a subspace of them is explored, namely the ones close to the original template.

Biophysical studies Biophysical studies aim at understanding the functioning of AMPs investigating their structure. Free energy perturbation, Molecular Dynamics (MD) simulations and thermodynamics calculations can all provide knowledge on how the three dimensional arrangement of residues is important to allow their functional role. Contrary to sequence based methods, these techniques give an insight into the mechanism of action of an AMP but their drawback lies in the high computational cost. For example simulations can approach systems with a limited size, and short (microseconds) time scales only preventing the reproduction of phenomena of the order of millisecond (a detailed overview of the state of the art, advantages and drawback of MD simulations will be given in Chapter 2). For these reasons, such techniques have been applied to fewer systems in comparison to sequence based screenings, and only few mutations have been tested and compared *in silico*.

The strength of biophysical studies lay instead in the fact that they exploit the whole information available on a system (sequence, structure, chemistry), so they can single out the interactions that are crucial for a mechanism, clarifying whether they can be transferred to a different environment, or discriminate cases in which similar sequences behave differently due to the environment around them. In this respect, they provide a generalisable knowledge applicable to different systems and thus to the design of novel AMPs at the atomistic level.

Examples of studies where MD simulations shed light on AMPs mechanism helping evaluating a mutant structure is the case temporin [91] sequences, where a mutant with improved activity and decreased haemolytic activity was synthesised on a computational basis.

Virtual screening Contrary to biophysical assay, virtual screening methods are employed to analyse a large number of sequences, when an experimental or computational test of all of them would be prohibitive. The concept of these methods consists in the identifying descriptors which allow to predict the potency of the sequence: from the analysis of a database of AMP with known activity, a model is created and used to score novel synthetic sequences.

The recent evolution of machine learning (ML) techniques, artificial neural network in particular, gave a great impulse to virtual screening of AMPs (for a historically informed review see Table 1 in Ref. [83]). Machine learning appears

particularly suitable to the task as the potency of AMPs is determined by the combination of many factors, the relative weight of which can be difficult to identify. Moreover, they can help in the identification of relevant features traditionally overlooked.

Practically, ML algorithms are trained on a set of AMPs labelled by their potency and characterised by different properties (features) along the line of the ones mentioned at the beginning of the section: sequence, partial charge, hydrophobicity, etc; but also experimental measures (pK measures, nuclear magnetic resonance data, octanol-water partition fraction and so on), together with theoretically computed features such as the van der Waals surface area. The more the input properties to consider, the more expensive is to train the model, but the higher the accuracy that can be meet. At the same time, the output descriptor is likely complex (i.e. combinations of many features) and thus of difficult interpretation. In principle, this is not a problem (provided there is no overfitting) as, rather than guide the design of AMPs from first principles, descriptors can be used to scores combinatorial sequences of the desired length to identify the best ones, but in practice, this is difficult because of their exponentially growing number.

The second obstacle to ML procedures is given by the fact that the more features one wants to consider, the more sequences need to be given as input to the algorithm, i.e. need to be experimentally tested. Modern high-throughput synthesis methods, together with surrogate measures of bacterial killing are allowing it nowadays, as shown by Cherkasov et al. [92], who assessed the antimicrobial properties of thousands of 9 residues sequences and trained a neural network on the outcome, to then score novel sequences with good accuracy.

As already mentioned, peptide design has proven successful in producing sequences with improved potency or selectivity; however, it is still a case-dependent procedure, rather than a general, automated protocol easily applicable to enhance any sequence of choice.

1.2.5 Clinical applications

Antimicrobial peptides have been studies for many years, however the push to capitalise them to get compounds viable for the clinical stage has been delayed by many factors, including production costs, and lack of interest in

the face of more potent small molecules which were deemed more economically advantageous by pharmaceutical companies. The constant creeping of AM resistance though has focussed more effort on this class of compounds, mainly from small biopharmaceutical companies, and at present several of them are in clinical trials, in phase 1 or 2 [93].

The two major problems encountered so far for AMPs sequences in trial are the liability to proteolytic degradation, and the unknown toxicology profile when administered systemically [36]. For the last reason in particular, many of them in trial are for topical use against skin infections only, while they are deemed unsuitable for internal administration. Design of novel AMPs can be tailored to improve the liability to degradation, for example introducing D-amino acids, non natural amino acid analogues of opposite chirality, which, with appropriate formulations, are mimetic to the immune system [94]. Moreover, machine Learning protocols can help in pre-screening their toxicity through virtual screening methods.

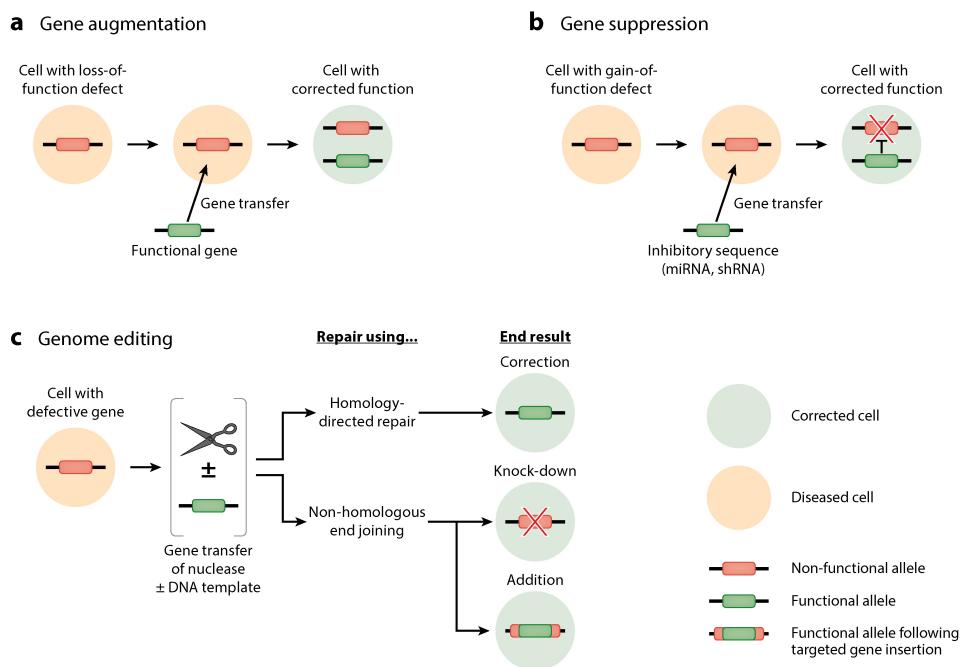
Overall, antimicrobial peptides remain a promising tool to counteract infections and, as their design is still - comparatively - in its infancy, there is room to explore novel applications and synthesise improved sequences apt to get to the clinical stage.

1.3 Gene therapy

Alongside the new compounds used to counteract bacterial infections, we want to bring the reader's attention to another class of therapies developed in the last decades for the treatment of non infectious diseases and is relevant for the work of this thesis as well: gene therapy. In recent years it has greatly evolved and gained attention for the treatment of tumours, genetic diseases and complex acquired disorder [95].

The key concept is the delivery of genetic material to sick cells which possess a faulty copy of a gene, to influence its expression. Such fault can result in lack of synthesis of the protein of interest or in its misfold and/or malfunction. The correction can be performed in three different ways (Figure 1.4).

Augmentation gene therapy introduces an healthy gene copy to restore the normal functionalities of the protein of interest and thus of the cell: it usually consists in the delivery of a DNA strand, which in turn can be internalised in



AR Anguela XM, High KA. 2019. Annu. Rev. Med. 70:273–88

Figure 1.4: Principles of gene therapy. Reproduced from [95].

the genome and thus spread when the cell replicates, or not internalised and thus can influence the functionalities of the particular transfected cells only. Suppression gene therapy suppresses a detrimental gene, and this is particularly useful in the case of cancer, to impede cancer cells replication. Usually this strategy employs RNA interference, delivering miRNA (microRNA) or siRNA (small interfering RNA) strands which repress the transcription of the problematic RNA sequence. Finally, gene-editing, the most recent advance in the field, overlooks the possibility of correcting base pairs mutations to restore the original healthy sequence and is often done through the functionalisation of the CRISPR-Cas9 technology, a mechanism found in prokaryotic organism as defence against viruses [96]: CRISPR (clustered regularly interspaced short palindromic repeats) is a library of DNA fragments from viruses that have previously infected the prokaryote, and the Cas9 enzyme (“CRISPR-associated protein 9”) uses these sequences to recognize and cleave strands of DNA complementary to the CRISPR sequence to blocks the reproduction of viruses if a following infection occurs. Research has been able to engineer the CRISPR-Cas9 technology to edit (rather than simply cleave) genes within eukaryotic

organisms [97], thus performing a therapeutic role.

Despite the challenge posed by the development of genome editing tools, and the risk associated to them (for example the possibility of deleterious insertional mutagenesis or deleterious immune responses), at present six gene therapies have received approval in the Western world [95], with many more undergoing regulatory review.

One of the main challenges in the development of such therapies lies in the identification of a suitable vector: delivery of free genome in solution results in poor internalisation and low therapeutic effect. Vectors instead allow the DNA/RNA to enter effectively into the cell: viruses can be used, modifying their genome to include the necessary sequence and remove the ones promoting viral replication [101, 102]. Nowadays, the outlook of gene therapy research lies not only in improving specific cargos to cure at the molecular lever more diseases, but also in the research of appropriate vectors with low toxicity, low induced immune response and high delivery efficiency. In that respect synthetic vectors started to be investigated for a virus-free delivery strategy. The system studied in this thesis proposes, among its other functions, to deliver genetic material into human cells.

1.4 Delivery of therapeutic material

The problem of delivery sets a parallel with the small drug one, introduced at the beginning of the chapter. Indeed small molecules needs as well some delivering agents to be efficiently internalised in the cells, as well as genes do.

...

1.5 Closing the circle: an antimicrobial drug delivery vehicle

Twice in this introduction peptide design has been brought to the reader's attention. First, design can engineer self-assembling building blocks for the formation of delivery scaffolds. Second, it can produce antimicrobial peptides with improved potency or selectivity, or reduced toxicity. As design is not bound to natural rules, it can foresee and imagine multifunctional materials

which are not observed in nature. In particular, the introduction above poses the question of whether it is possible to engineer peptides able to perform both an antimicrobial and a delivery function at once (either of drugs or genetic material).

Such self-assembling, antimicrobial compounds would have a twofold interest for medical applications. First of all, self-assembly is functional to the antimicrobial activity: many AMP sequences have a weak potency, and only a high (critical) concentration can trigger the bactericidal mechanism. This is intuitive in the case of the carpet model strategy (see Section 1.2.2), where AMPs lay homogeneously on the surface of the bacterial membrane and breaks it upon sufficient coverage of its area. Also the barrel-stave and toroidal pore models rely on the mutual interactions between peptides to maintain the pore edges. Generally, as AMPs are positively charged, the localised presence of many copies of a sequence enhances the local electric field and charge imbalance, which are critical to the membrane stability. Second, in order for the assembly to be able to perform the additional delivery function, it must be able to either organise in a tailored structure (for example a capsule able to host a drug), rather than an amorphous aggregate, or to co-assemble with the cargo of interest.

Out of all the possible applications, the most promising is perhaps the use of such vehicles to deliver drugs to treat metabolic or genetic diseases: while the cargo tackles a defect of the host system, the vehicle can counteract the proliferation of bacteria. This is particularly important in situation where the host immune response is weakened and thus infections normally harmless can spread and cause damage. However, it must be noticed that the cargo is not bound to be a small molecule, as long as it can effectively co-assemble with the peptidic carrier. As mentioned in the previous section, gene therapy is also an actively expanding field which looks with interested at the development of vehicles for genetic material. Given that viruses have been the first choice for DNA/RNA delivery so far, peptidic carriers seem a natural evolution of them.

Given the above premises, it is evident the importance of pursuing the research on novel multifunctional peptidic materials. As mentioned when discussing AMPs design, to understand such systems, each of them must be characterised by itself, as a generalised knowledge is still lacking. With this aim, this thesis proposes to elucidate the behaviour of a specific synthetic



Figure 1.5: Capzip molecule scheme and bond representation. [TO BE IMPROVED] Adapted from [8].

self-assembling peptide, suitable for antimicrobial activity and gene delivery strategies. Its full characterisation will complete the knowledge on its mechanisms of action and complement the broader information already known on the class of such functional building blocks. This will be crucial to engineer new synthetic blocks with improved characteristics, either regard their antimicrobial activity, assembly performances, or tailored cargo delivery.

1.5.1 Capzip

The molecule capzip has been designed to perform the functions mentioned above at once. To recapitulate, the properties it possesses are:

1. assembly into nanoscale virus-like capsules with and without nucleic acids. This ensures that the vector can autonomously form and thus there is flexibility in the choice of the cargo;
2. antimicrobial activity of the molecule itself and of the capsule on a time scale useful for therapeutic applications;
3. promotion of gene transfer into mammalian cells when the peptide is co-assembled with the RNA strands, without causing cytotoxic and haemolytic effects.

Furthermore, the design effort aimed at building a template structure of minimal complexity, in order to reduce the synthesis effort to a short sequence. Arguably, short sequences are also more flexible in their assembly: it is thus important to explore them and prove whether even small blocks can form ordered structures.

Based on the above requirements, two design principles emerged: first the employment of a non-linear structure. There is indeed some evidence suggesting that non linear peptides are more prone to assemble in three dimensional structures, opposed to planar ones [?], and this holds in particular for short sequences which do not fold into a defined secondary structure. The second principle consists in the use of a template antimicrobial sequence which is short and has proved potency. Given that AMPs are usually anionic, the co-assembly with anionic RNA sequences is arguably inherited by consequence.

To satisfy the above guidelines, a short peptidic scaffold constituted by a β -Alanine and two Lysins has been engineered. Three identical copies of the antimicrobial sequence of choice are covalently bonded to the N-terminus of the scaffold sequence and to the nitrogen atom of the Lysin residues side chains (Figure 1.5). Regarding the antimicrobial sequence selected, it has been derived from the antimicrobial peptide bovine lactoferricin, which is in turn a portion of the Lactoferrin protein.

Lactoferrin Lactoferrin is an iron binding protein present in milk (in which it is most abundant, hence its name), saliva and other secretions, as well as in polymorphonuclear leukocytes. It works as an iron binder and provides a natural defence against bacteria and fungi [103–107], constituting a first defence for infants.

Lactoferrin contributes to bacterial suppression in several ways. At present, its known modes of action fall in three categories: first, thanks to its iron sequestering capabilities, it removes essential substrate required for bacterial growth [108]; second, it interacts with bacterial membranes and in particular binds to the lipopolysaccharides of bacterial walls, oxidising them and affecting the membrane permeability with consequent cell lysis [108]; finally it is implicated in the stimulation of different immunological cells (killer cells [109], polymorphonuclear leukocytes, and macrophages [110]). The peptide fragment responsible for binding of lactoferrin to the bacterial membrane, named lactoferricin (Lfcin), has been identified near its N-terminus and found to have a more potent bactericidal effect than intact lactoferrin on a wide range of bacteria [59, 60, 111, 112]. Similarly, a synthetic short peptide derived from a subsequence of human lactoferricin has been proven effective against bacteria as it depolarises the cytoplasmic membrane decreasing the pH gradient [113].

The bovine homolog of lactoferricin (LfcinB) has a higher bactericidal potency than human lactoferricin on several bacteria [114] and therefore has been more extensively studied. Its active core LFC is a 25-amino acid sequence which adopts a helical conformation in the full structure but, once isolated, crystallises in a β -hairpin with a disulfite bridge nearby the terminals which stabilises the fold, but was shown to be not essential for bactericidal activity [114]. Further experiments on LfcinB subsequences identified an even shorter antimicrobial core, constituted by the six amino acids RRWQWR [61]. This core presents a characteristic Tryptophan zipper motif WTW, which appears very often in nature in β -turn and β -sheet conformations, paired to another copy of the same motif, so that Tryptophan rings from facing strands are packed tightly against each other in an alternated way [114] (Figure –). In general, the six amino acid sequence contains both charged and hydrophobic residues, in line with the usual composition of antimicrobial peptides. Accordingly, its antimicrobial action is likely derived from the interaction with biological membranes through charge recognition first and aromatic rings insertion in a second moment.

To further elucidate this mechanism, several experimental investigations have been carried both on LfcinB and its subsequences. First, the structure of LfcinB in solution has been investigated by NMR (Nuclear Magnetic Resonance), resulting very flexible [115]. Then, the binding of its antimicrobial core to sodium dodecyl-sulfate micelles was studied [61], suggesting a favourable interaction of aromatic residues with the micelles surface. Similar experiments were performed on large unilamellar vesicles, constituted by lipids modelling biological membranes: ePE:ePC was chosen as a model of a mammal membrane, and ePE:ePG or ePC:ePG for a bacterial one. The experiments showed preferential binding to the latter ones, based on Tryptophan fluorescence [116], suggesting a selective antimicrobial action on anionic membranes. Additional experiments have been performed on the full sequence or mutated subsequences [117, 118] to investigate the binding to other different model membranes but, as the systems investigated are slightly different, as well as the experimental conditions, it is difficult to relate them and give a unified interpretation of the modes of action of lactoferricin derived peptides.

Finally, an alanine scanning has attempted to clarify the role of each amino acid in the antimicrobial activity of the LFC peptide [119]. The results sug-

AFM/TEM	cryo-em
fluo hollow capsule	fluo RNA uptake

Figure 1.6: ... Reproduced from [8]

gested a binding function for the Tryptophan residues, in line with one of the roles Tryptophan can assume in antimicrobial peptides [46]. Other possible roles involve its propensity to form hydrogen bonds, in which case the residue would position itself at the interface between solution and membrane, rather than inside the latter (which happens instead when Tryptophan residues have a binding role).

The designed block From the active core of LFC (of sequence RRWQWR), a mutated sequence was obtained to comply the design criteria of a self-assembling building block. Two mutations were introduced to favour the assembly of arms belonging to different molecules in an antiparallel fashion. Specifically, given that the original RRWQWR sequence is found in a β -sheet (at least in the crystal lattice), the mutations aim at promoting a similar structure. Therefore, the Glutamine residue and the C-terminal Arginine of the lactoferrin motif were replaced with Threonine and Glutamic acid residues to have a self-complementary sequence RRWTWE: the pairing is promoted by the attraction of opposite charges at the ends of the sequence. Three copies of this sequence were thus covalently bonded to the scaffold described previously and shown in Figure 1.5, to obtain a self-assembling molecule in a three dimensional shape, hosting multiple copies of an actively antibacterial sequence.

1.5.2 A viable systems: experimental background and question

Many experiments have been performed to verify that capzip had the characteristics it was designed for. The set of experimental results obtained on the molecule has been published in Reference [8], while more recent results extend and consolidate the previous findings.

Experimental results First, the assembly ability have been tested: the peptide does not show assembly in pure water (as verified by Dynamic Light

Scattering), while in biological buffer (MOPS, 150 mM) at physiological pH of 7.0 it forms capsules with dominating size range of 20–200 nm. This is confirmed by images of the capsules obtained with multiple techniques, namely transmission electron microscopy (TEM) (Figure 1.6), atomic force microscopy (AFM), and cryo-scanning electron microscopy (SEM). The fine structure of these assemblies appears irregular to the resolution power of such techniques. Some insight into the details of the assembly is given by Circular Dichroism (CD) spectra, which show a profile characteristic of β -turns and contain elements of a β -sheet structure and of indole rings, with minima at $\lambda \sim 200$ nm and 214 nm. Complementary evidence about the overall shape of the assembly was provided by the cross-sectional analysis of the assembled capsules by fluorescence microscopy using fluorescein to label capzip. The signal comes from the wall of the capsule only, showing an inner cavity (Figure 1.6). Finally, small angle X-ray scattering (SAXS) measurements were consistent with compact capsules interfacing with solvent.

The assembly process is also tested and monitored in combination with small interfering RNAs (siRNA): as mentioned in section 1.3, these sequences are a promising tools for RNA interference techniques which aim at inhibiting the expression of specific genes, however, they are easily degradable and thus difficult to deliver to the target cell without an appropriate vehicle. The co-assembly of a 21 base pairs duplex with the peptide shows the formation of structures similar to the ones formed by the stand alone peptide only: CD spectra highlight the helical signal from RNA together with the features proper of the peptide.

These co-assembled structures were tested for siRNA delivery in HeLa cells, showing that the presence of the peptide favours the internalisation of the genetic material with respect to the transfection results of a pure siRNA control. The delivery of fluorescent siRNA (Figure 1.6) showed that the internalisation occurred within the first hours from the transfection in localised regions of the cytoplasm, suggesting an endocytic uptake. This distribution was stable over the first five hours of incubation after which the fluorescence signal decayed. Flow cytometry assays quantified the increase in siRNA uptake levels due to the presence of capzip, confirming that this molecule is competitive with other commercial transfection reagents (like Lipofectamine[®], unpublished results).

To further quantify the level of RNA internalisation, a mRNA knockdown

experiment was performed on a HeLa cell line with two housekeeping genes, ACTB (β -actin, targeted) and GAPDH (reference) [120]. The silencing of β -actin mRNA was detected 22 ± 2 hours after transfection; and its knockdown “fitness” was expressed relative to cells treated with siRNA alone (background) and normalised against viable cell counts (Figure 1.6). Capzip fitness was lower than Lipofectamine[®] one, however cells treated with capzip remained viable after 24 or 48 hour, resulting in higher cell counts than the samples treated with the commercial reagent, suggesting that capzip has little cytotoxicity. The experiment above was performed at neutral to positive charge ratio close to one (where each siRNA molecule has a -42 e charge and capzip a +6 e charge), as test experiments performed at higher peptide-to-siRNA ratio showed no improved uptake.

Finally, the peptide does exert an antimicrobial function: the non-assembled peptide has shown to be effective against both Gram positive and negative bacteria (E. coli, P. aeruginosa and S. aureus), with no haemolytic effects and minimum inhibitory concentrations typical of other antimicrobial agents. On Supported Lipid Bilayer with negative total charge (mixed DLPC and DLPG, 3:1 ratio), the capsules create localized pores within minutes, as proven by AFM experiments repeated in time. The pore depth ranges between 1.4 and 2.2 nm, which is smaller than the radii of the capsules, however it is sufficient to disrupt the structure of the membrane. Finally, to prove the viability of capzip as antimicrobial agent *in vivo*, it was used to counteract methicillin-resistant S. aureus (MRSA) infections in G. mellonella larvae. The particular bacterial strain used was susceptible to vancomycin, which could be used as control: the larvae treated with capzip showed survival rates significantly higher than the untreated control, and comparable to those treated with high doses of vancomycin (unpublished results).

Open questions Despite the success of the experiments mentioned above, there is much information still to be uncovered on the precise mechanism of action of such peptide.

Specifically, both the assembly process and the antimicrobial mechanisms contain some unknown: regarding the former, it is important to understand which amino acids or sub-structures allow the pairing of molecules, whether such pairing is specific or not, how reversible it is, and how rigid the final

structure is. Regarding the latter, it must be highlighted what molecules in the membrane the peptide binds to, and how this binding affects the full membrane structure. Finally, as there is evidence that the assembled molecule is a more powerful antimicrobial compound than the single molecule, it is interesting to understand whether any cooperative action is taking place or the enhanced antimicrobial power of the assembly is due only to the localised higher concentration.

Even if further experiments or future improvements in the techniques already employed might tackle some of the aspects above in a near future, arguably no experimental outcome can provide an atom-by-atom knowledge of the processes of interest in any time soon. Ideally though, one would like to track each of them, i.e. the processes happening in any the environments capzip has been exposed to (physiological solution, supported lipid bilayers, bacterial extracellular matrix, mammal cell membrane and cytoplasm) both in space and time with the finest level of details, and the impossibility of that leaves large gaps in the understanding of the system.

1.6 A computational approach to understand capzip

The gaps mentioned in the characterisation of the systems prompts for new investigations in order to complement the knowledge already provided.

Beside the quest to enrich the fundamental knowledge on self-assembling peptides and antimicrobial ones, the understanding of this very system is crucial for its further development. We outlined already in Section 1.2.4 how antimicrobial peptide design can proceed from already viable templates and empirical principles, when first principles are not available. Similar rules hold for designing self-assembling peptidic materials, to obtain tailored delivery vehicles (see Section ??). Therefore, a full knowledge of the interactions between peptides and between their assembled structures and the membrane, i.e. of the mechanism of its functions, will drive the engineering of new likewise peptides. A knowledge-driven design would hopefully provide new blocks suitable for a double action as the one capzip performs, and this in a shorter amount of time than a research based on less information or on a trial-and-error procedure of mutations in the chemical composition or in the architecture of the molecule. A few examples of possible knowledge-related improvements

include the following:

- the knowledge of capzip binding mode to the bacterial membrane might suggest its suitability as a broad range spectrum compound or the possibility of tuning its action against specific pathogens;
- understanding the molecule-molecule interactions classifies the robustness of the assembled structure and the possibility of designing block which disrupts under particular chemical conditions only;
- querying the electrostatic profile of the assembled structure suggests which type of molecules, other than siRNA, could be efficiently co-assembled and thus delivered.

In recent years, computational techniques are stepping forward to complement incomplete experimental knowledge and complete the picture of how biological systems work. For this reason, it seems natural to query such techniques to study the capzip system as well. Zooming into the details of the interactions can be performed via a theoretical modelling of the system in time, and thus through the simulation of its evolution, starting from few basic principles and the knowledge of the chemical composition of its parts. The technique this work focusses on is Molecular Dynamics simulations, which aims at reproducing the behaviour of a system of atoms in a semi-classical description using basic physical laws, as it will be described in details in the next chapter.

Thus it is the aim of this thesis to prove that Molecular Dynamics simulations can clarify further details on the assembly mechanisms of capzip and on its interactions with biological membranes, in order to gather more information on the system and contribute in the future to the designed of new molecules with enhanced functional capabilities.

Chapter 2

Methods

MOLECULAR DYNAMICS is a computational method which has gained popularity and significance in the past few decades in the fields of biology, biological chemistry, and biophysics. The increasing amount of data available from experiments on biomolecular materials and processes, united with the increasing computational power, has made possible an analysis of such experimental data tailored to implement theoretical models of the systems studied. The resulting models can be simulated on a computer so that the dynamical properties of the processes in exam are uncovered at a molecular level which would be inaccessible to experiments.

The general idea of simulating biological processes consists in describing both the components of a system and their mutual interactions, so that the laws of physics provide the natural evolution of the system. In principle every atoms should be present in the picture, and the evolution rules should be derived by the principles of quantum mechanics. To facilitate the task, several simplified descriptions are possible, which differ in the choice of spatial resolution, degrees of freedom and evolution laws; and each description is most suitable to address particular questions and investigate particular systems.

Some models (including all the ones we will focus on) opt for classical mechanics laws to move atoms in space, and a subgroup of these pushes the simplification further by mapping small groups of atoms into a single bead - to update less positions at each time instant. For increasing sizes of the system simulated and longer time spans described, the approximations due to a classical approach will be less and less relevant, as classical mechanics represents well the evolution of large systems, for which the atomistic quantum behaviour

and in general some fine grain details are of minor importance. There are certainly biological processes for which a quantum mechanics description is more suitable - such as photosynthesis, DNA mutation processes or particular enzymatic activities - and to model them precisely, hybrid techniques have been developed, to gain the accuracy of a quantum description in the region of interest and the speed up of a classical one in the surrounding areas [121].

Later on in this chapter we will discuss three different models used in this work to simulate biological molecules, together with relevant examples of how simulations have been successfully applied in the field of biophysics and in particular to the study of assembling or antimicrobial peptides. But first, strengths and limitations of Molecular Dynamic simulations in general will be briefly outlined. In doing this, the discussion focuses on four problems simulations have to face and solve: the force-field problem, the search problem, the ensemble problem and the experimental problem. This schematic follows the excellent review by van Gunsteren [122], which identifies in these four issues the interpretative key with which MD simulations must be designed, run and interpreted.

2.1 Algorithms for Molecular Dynamics

Although most of the content in this section highlights key features applicable to many computational techniques employed in biophysics, it is written with Molecular Dynamics (MD) simulations as its focus. Therefore, we first outline more in detail the core algorithm which allows MD simulations to run, as it sets the ground for the approximations to follow: indeed, as much accurately the system can be modelled, in a classical MD framework it will always be processed by an engine based on classical mechanics rules and finite steps approximations, which inevitable influences the outcome.

2.1.1 The Newton's law

In a classical MD framework, Newton's second law of motion rules the dynamics, stating that the acceleration \mathbf{a} that a particle is subject to, at each moment, depend on the total force \mathbf{F} acting on the particle itself and on its

mass m (bold denotes vectorial quantities):

$$\mathbf{F}(t) = m \cdot \mathbf{a}(t). \quad (2.1)$$

As the acceleration $\mathbf{a}(t)$ is the second derivative of the position $\mathbf{r}(t)$ with respect to time, given the position and the velocity of the particle at the initial time $(\mathbf{r}(t_0), \mathbf{v}(t_0))$, their temporal evolution can be computed integrating the acceleration (and thus $\mathbf{F}(t)/m$) as follow:

$$\mathbf{v}(t) = \mathbf{v}(t_0) + \int_{t_0}^t \frac{1}{m} \mathbf{F}(t') dt'; \quad (2.2)$$

$$\mathbf{r}(t) = \mathbf{r}(t_0) + \int_{t_0}^t \mathbf{v}(t') dt' + \int_{t_0}^t \int_{t_0}^{t'} \frac{1}{m} \mathbf{F}(t'') dt'' dt'. \quad (2.3)$$

Several analytical techniques have been devised in the previous centuries to solve this particular problem in a number of cases, feeding the fields of analytic and rational mechanics. Indeed, out of all the possible second order differential equations, the equations of motion represent a peculiar subset for which it is proved that an analytical solution exists. Moreover some properties of the motion can be derived even when the solution itself is too complicated to compute: for example, quite often the positions visited by the particle can be obtained, albeit the exact time instant at which they are explored is not known [123].

However, in the case of large, complex systems, analytical approaches are almost hopeless in facing the task. The case of biomolecular modelling falls in this category because of the large number of degrees of freedom present and the complexity of the forces acting on each of them. Such forces derive from chemical bonds, electrostatic interactions, and Pauli repulsion between atoms, all at once. In the impossibility of solving (in the analytical sense) the problem, a different, feasible approach consists in discretising the equations of motion. The idea is to consider very short time steps of length Δt , so that in such interval the forces are (almost) constant, and thus the integration of Equation 2.2 becomes trivial:

$$\mathbf{v}(t_0 + \Delta t) = \mathbf{v}(t_0) + \frac{\mathbf{F}(t)}{m} \Delta t; \quad (2.4)$$

$$\mathbf{r}(t_0 + \Delta t) = \mathbf{r}(t_0) + \mathbf{v}(t_0) \Delta t + \frac{\mathbf{F}(t)}{m} \Delta t^2. \quad (2.5)$$

This procedure, the Euler algorithm, clearly contains some approximation (of the order of $(\Delta t)^2$) that will accumulate step after step. To obviate to that, several different algorithms have been designed to integrate Newton's equation, mainly playing with the choice of the velocity to be integrated during each time step: one possibility is to take its value at time t_0 as in Equations 2.4 and 2.5, but another legitimate choice is given by its value at time $t_0 + \Delta t/2$. The leap-frog algorithm is based on this, giving:

$$\mathbf{v}\left(t_0 + \frac{\Delta t}{2}\right) = \mathbf{v}\left(t - \frac{\Delta t}{2}\right) + \frac{\mathbf{F}(t)}{m} \Delta t; \quad (2.6)$$

$$\mathbf{r}(t_0 + \Delta t) = \mathbf{r}(t_0) + \mathbf{v}\left(t_0 + \frac{\Delta t}{2}\right) \Delta t. \quad (2.7)$$

This scheme is more precise than the Euler (its error is of the order of $(\Delta t)^4$), and it is the algorithm used by the vast majority of MD engines.

An engine based on such approximation can thus “solve” every possible Newton equation, at the expenses of some precision. Once the equations have been set up, the next challenge is represented by modelling the forces in a suitable way to represent the phenomena observed in nature.

2.1.2 Rescuing the approximation limit

For completeness, we mention here a few algorithms that have been developed to alleviate the influence of the approximations performed in the integration algorithms.

First, some bonds in the molecules are constrained to a particular length or angle. To ensure this property, after the update of all the atoms positions, a constraint algorithm is applied to bring the positions of mutually constrained atoms back to a value which satisfies the constraint itself. Several algorithms have been developed to do so, and the ones used in this work are LINCS (Linear Constraint Solver) [124] and SETTLE [125]. As the problem of constraints is hardly solvable analytically, most of these algorithm proceed iteratively, finding the best solution within an approximation tolerance. SETTLE however is an exact implementation of the solution for rigid bodies of three elements, and as such is useful for the treatment of water molecules (in their atomistic description).

The approximation in the computation of the velocity for the different

particles makes the overall average velocity drift away from its initial value, despite the dynamics is conducted in conditions which should preserve it. This has consequences on the average kinetic energy and thus temperature: to ensure that the same temperature is maintained throughout the simulation, several thermostat algorithms have been devised, which aim at releasing the excess kinetic energy or absorbing it if necessary. The general idea consist in rescaling the velocity of one or few selected particles at fixed interval of times, to bring the kinetic energy to its initial value. Throughout this work the velocity rescale thermostat has been employed [126].

When simulating a system in NPT conditions, the simulation box must be rescaled during the run to ensure the pressure is maintained correctly: indeed the average quantity of motion exchanged from the particles with the walls of the box depends on their velocity and on the frequency of collision, determined by the extent of the box. To be noticed that an external algorithm to keep pressure is necessary as, implementing periodic boundary conditions, the box does not have hard walls against which atoms bounce, rather, they trespass them and are virtually put back in the box entering from the opposite side. Also for pressure coupling several algorithms are used, which approach the optimal dimensions of the box with different behaviours: the two employed in this work are the Berendsen [127] or Parrinello-Rahman barostat [128]. The first approaches the correct value with an exponential behaviour, and it is recommended in equilibration phases, the second with an oscillatory one, and is suggested for the production phase.

2.2 The force field problem

The modelling of force fields to be used in conjunction with a classical description of the dynamics usually relies on the breakdown of the interactions between atoms into several, independent terms, identified on an empirical physical basis. We report here the functional form adopted for the GROMOS force field [129, 130] as implemented in the GROMACS MD engine [131–133], explaining what each term represents. Other force fields can have slightly different implementations, or miss some terms if the level of details investigated is too coarse to necessitate all of them. However, the general classification of interactions and the type of functional forms used to describe them are similar.

Covalent (bonded) interactions Covalent interactions are modelled with potential energy terms representing bond-stretching, bond-angle bending, improper and proper dihedral-angle torsion. The equilibrium values of such quantities and the fluctuations they can withstand are determined by either molecular orbital theory, quantum mechanics calculations, or fitting the results of simulations to some macroscopic quantities as the free energies of solvation of given compounds. The GROMOS force field is based on the latter, while others like CHARMM [134, 135] and AMBER [136] use a quantum mechanics approach. In many cases, the parametrisation procedure is performed for small moieties only, assuming that the values obtained for them can be transferred when a moiety is included in a larger compound. This assumption limits the number of parameters needed in the force field to describe biomolecular systems.

The functional form of the potential-energy for bonded interactions aims at a simplified, semi-classical description of the sub atomic motion of molecules, assuming harmonic-like vibrations around the equilibrium position of the bond, angle or dihedral in exam.

Specifically, in the GROMOS force field, a bond between atoms i and j is described by a fourth power potential, which is similar to a harmonic form, but computationally more efficient. The forces acting on the atoms when the bond is stretched are obtained from the derivative of the potential in space:

$$V_b(\mathbf{r}_{ij}) = \frac{1}{4} k_{ij}^b (|\mathbf{r}_{ij}|^2 - b_{ij}^2)^2 \quad (2.8)$$

$$\mathbf{F}_i(\mathbf{r}_{ij}) = k_{ij}^b (r_{ij}^2 - b_{ij}^2) \mathbf{r}_{ij} \quad (2.9)$$

where the force constant k_{ij}^b is given in kJ/mol/m² and b_{ij} is the equilibrium position of the bond between atom i and j .

The preferred angle between three atoms i , j and k , and the stiffness with which its value can deviate from the preferred one (θ_{ijk}^0) are implemented through a cosine based angle potential:

$$V_a(\theta_{ijk}) = \frac{1}{2} k_{ijk}^\theta (\cos(\theta_{ijk}) - \cos(\theta_{ijk}^0))^2 \quad (2.10)$$

$$\text{with: } \cos(\theta_{ijk}) = \frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{kj}}{r_{ij} r_{kj}} \quad (2.11)$$

with k_{ijk}^θ in kJ/mol.

Improper dihedrals are used to ensure ring planarity and control the chirality of some tetrahedral centres. They are described through a harmonic potential:

$$V_{id}(\xi_{ijkl}) = \frac{1}{2} k_{ijkl}^\xi (\xi_{ijkl} - \xi_{ijkl}^0)^2 \quad (2.12)$$

where the ξ values are given in degrees and the force constant in kJ/mol/rad². By convention, the improper dihedral for a set of four atoms i, j, k and l , is taken as the angle between the plane defined by atoms (i, j, k) and the one defined by atoms (j, k, l) .

Finally, the last bonded interaction is represented by proper dihedrals, described though a periodic potential:

$$V_d(\phi_{ijkl}) = k_{ijkl}^\phi (1 + \cos(n\phi_{ijkl} - \phi_{ijkl}^0)) \quad (2.13)$$

following the convention that ϕ_{ijkl} is the angle between the (i, j, k) and (j, k, l) planes, with i, j, k , and l four subsequent atoms (for example along a protein backbone). A value of zero for a proper dihedral corresponds to a *cis* configuration; n denotes the number of equally spaced minima available for the dihedral in a 360° turn. k_{ijkl}^ϕ is expressed in kJ/mol.

It must be noticed that potentials can not model the rupture of a bond: for this, more sophisticated descriptions are needed.

Non bonded interactions Non bonded interactions includes the short range Pauli repulsion, the “mid-range” van der Waals attraction between atoms, and finally the long range electrostatic term.

The first two can be modelled together by a Lennard-Jones potential. Its functional form, describing the interaction between two neutral atoms at distance r , models the long range dispersion with a r^6 behaviour typical of the dipole-dipole interactions found in noble gases (London dispersion forces), while the Pauli term is represented by a r^{12} behaviour to ease the computation in relation with the previous term:

$$V_{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]. \quad (2.14)$$

Two parameters, ϵ and σ , tune the interaction strength and the equilibrium

distance between the two particles. They are fitted against experimental data and are specific of each pair of atoms species.

The Coulomb energy between two charges q_1 and q_2 at distance r is represented by the Coulomb law itself:

$$V_C(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_r r_{ij}} \quad (2.15)$$

with ϵ_0 the dielectric constant of vacuum and ϵ_r the relative dielectric constant, introduced to properly take into account the screening provided by the material surrounding the object.

The treatment of non-bonded interactions requires particular care because of their intrinsically long range nature. The van der Waals forces are usually weak and decay fast, therefore the tail of their functional can be cut after a threshold distance with little impact on the overall computation of the forces; Coulomb interactions however must be taken into account throughout the whole extension of the simulated system, as a simple cut-off approach would impact the simulation severely. Many algorithms have been devised to efficiently compute them, like the Particle Mesh Ewald [137] technique, or the Reaction Field [138] approach; when choosing the former, the GROMACS MD engine dedicates a fraction of the overall computer resources exclusively to it, because of its high computational cost.

Finally, it must always be considered that a simulation outcome is determined by the combination of all the non-bonded interactions (together with the bonded ones). As they come in great number (in principle proportional to N^2 , with N the number of particles in the simulations), their collective result is often difficult to predict based on the action on a single atom or on scaling reasoning, and small shift in the parameter choice can give very different “macroscopic” results.

Because of these reasons, parameterising biomolecular force fields is a challenging problem: however the assumption that parameters can be transferred across different molecules when they describe bonds between the same atoms, in similar chemical context, allows to contain how many of them are necessary for the simulation. Moreover, biomolecular systems evolve at room temperature, or at temperatures very close to it, so that force fields are calibrated

against experimental values obtained in such conditions. On one hand, this means they might be unsuitable to reproduce the behaviour at very high or very low temperatures, but at the same time, this allows to reduce the complexity of the description, as more convoluted ones would be needed to properly take into account the changes in behaviour due to temperature shifts.

Before moving on to the other goals and problems of MD simulations, we give here a brief description of the three force fields employed in this work. Each of them adopts a functional form equal or similar to the one described above. Their difference lies in the number of degrees of freedom modelled, in a hierarchy of descriptions proceeding from detailed to coarse. Coarse-graining is a common procedure to reduce the number of degrees of freedom to sample, which allows for a quicker exploration of the system energy landscape (see Section 2.3). What this specifically corresponds to would be explained for each of the coarse grain force fields considered.

2.2.1 The GROMOS force field

The GROMOS force field is a united atom description of biological systems. This means that each atom is modelled as an independent entity (a sphere) a part from non polar hydrogens, which are incorporated in the heavy atom they are bonded to. For example, alongside a lipid chain, there can be (among others) CO, CH, and CH₃ groups (see Figure –). The last two are modelled as unique atoms, with mass equal to the mass of the carbon plus the masses of the hydrogens bonded to it. Accordingly, they are treated as different carbon atom types, which in turn are different with respect to the carbon type used to model the “bare” C atom in the CO group.

The parametrisation of the GROMOS force field relies on the accurate reproduction of free enthalpies of solvation of different compounds in many solvents, and aims at reproducing thermodynamic properties such as the density and the heat of vaporization of small molecules in the condensed phase at physiological temperatures and pressures. As mentioned before, the parameters used for such small molecules are employed to represent the same moieties when they appear in larger molecules. What can, and must, change accordingly is the distribution of the charges inside a molecule: as atoms are represented by spheres, no electrons are included for the sake of efficiency, and their redistribution across atoms which are inter-bonded is modelled through

fractional charges assigned to each atom (while the total charge of a molecule must clearly sum to an integer).

In such simulations, the description of water is clearly important. Out of the many water model proposed, the GROMOS parametrisation has been performed with a flexible simple point charge (SPC [139]) description. Intuitively, this model represents water as a three atoms molecule, placing a negative charge on the oxygen and a positive complementary one on the two hydrogen atoms, and allowing the bonds to vibrate (thus they are not rigid). This model is able to reproduce correctly the density and dielectric permittivity of water. Computationally wise, water represents the vast majority of the particles involved in a simulation and thus a significant fraction of the computer time is spent in updating water molecule positions. As mentioned before, atomistic water has a dedicated algorithms for the renormalisation of its bond lengths (SETTLE rather than LINCS), which can solve the constraint problem in an analytical way for this three bodies problem.

The improvement of computational techniques and reparametrisation strategies prompts the periodical release of newer versions of the force field. In the present work, we employed version 53a6 [129] for the set of simulations involving peptidic assembly in solution, while we switched to 54a8 [130] for the simulations involving biological membranes. While it is advisable to have a coherent set of parameters across simulations, to compare their outcome in a consistent manner, we deemed the 54a8 parameter set more suitable for lipid simulations because of the improvements introduced in the phosphocholine head parametrisation (see Chapter – for a complete discussion on lipid parametrisation in GROMOS). For this reason, we performed the update, still being able to compare the set of simulations of peptide in solution among themselves, and the ones involving lipids as well.

2.2.2 The SIRAH force field

The idea behind coarse-grain force fields is to group together in one unique bead a few atoms, to reduce the number of particles to displace during the simulation. The clustered atoms are such that their mutual distances are expected to vary little: the coarse-grain approximation is overlooking such details, while still maintaining information on the ample movements of the components of the system far away from each other.

While coarse graining a description, two approaches are possible: bottom-up and top-down. In the first case, the parameters are developed fitting the coarse-grain simulations results to the ones from atomistic simulations (so from a more detailed description), while in the latter they are chosen to fit directly global quantities derived from experimental data - as it is performed for example in the GROMOS atomistic force field parametrisation.

The coarse-grain force field SIRAH [140, 141] is a top-down generic force field derived to fit structural properties. It aims at reducing the complexity of an atomistic description while still being able to reproduce the correct secondary structure of proteins across a wide variety of folds contained in the PDB, and their evolution in time.

To obtain this, it opts for a non-uniform granularity, i.e. according to the region of interest a different number of heavy atoms is grouped in a bead, from a minimum of two up to four. Regarding proteins, it maintains the backbone flexibility by grouping NH, $C_\alpha H$ and CO in three different beads, while the side chains are represented with less details, generally grouping three atoms together. A schematic of the mapping for each amino acid is shown in Figure -. Contrary to force fields where the amino acid backbone is mapped to one bead only, the SIRAH description allows to reproduce secondary structures without recurring to additional constraints. The dual granularity approach is based on physico-chemical intuition, and is more difficult to generalise than a uniform one. Never the less, the force field has been recently (and successfully) extended to lipids, while it comprised a parametrisation for DNA molecules since its infancy.

The modelling of water in a coarse-grain force field is also critical: usually, a few water molecules are grouped together in one bead. This has two implications: water particles are large and thus cannot solvate very narrow pockets; moreover, collapsing the molecules in one single point in space removes the separation of charges and thus the characteristic dipole every water molecule has. The dipole of water is responsible for hydrogen bonds formation and for the electrostatic screening observed in an aqueous solution. Such screening can be roughly modelled tuning the relative dielectric constant, but as this is a mean field approach, it cannot account for local effects. To partially obviate to that, SIRAH force field maps four waters to a tetrahedral molecule, with one bead on each vertex: all the bonds are rigid, and the structure serves

the purpose of having a repartition of plus and minus charges, by assigning a positive charge to two vertices and the opposite charge to the other two, giving a polarisable structure. The geometrical arrangement reproduces the tetrahedral network of water molecules observed in its liquid state, which is characteristic of this fluid and tunes its remarkable properties.

Based on the above premises, SIRAH force field simulations of different peptides and proteins in solution proved to match the relative NMR results, showing a good reproduction of secondary structures; simulations of lipids randomly oriented in water showed the formation of an organised bilayer, and finally the expected behaviour of a few transmembrane proteins in model membranes was correctly reproduced. More details on some of the systems simulated will be given in Section 2.6, where simulations of assembling peptides and antimicrobial ones will be reviewed in function of the objectives of this thesis.

2.2.3 The MARTINI force field

The MARTINI force field is another very popular description of biological molecules [142–144]. It was developed much prior to the SIRAH force field, and since its first description, it has been refined and extended to include proteins, small ligands and DNA/RNA molecules besides lipids, which were its initial focus.

As a general rule, MARTINI opts for a four-to-one approach, i.e. four heavy atoms are grouped in one bead, resulting in a uniform graining and a coarser description than the SIRAH one. Moreover, the panel of possible beads has been kept to the minimum necessary number to take into account the variability of moieties found in biological systems, and it is organised in a systematic way: beads are classified as polar, non-polar, apolar, or charged, and each of these type has a number of subtypes, representing “shades of polarity” to represent accurately the chemical nature of the different underlying atomistic structures. The advantage of this systematic approach is its transferability: beads capture general properties of the structures represented, and as such they can be used for the parametrisation of new compounds containing chemically similar moieties, without the need to introduce new bead types for each new compound.

This logic is analogous to what pursued in GROMOS, where the description of different chemical groups was optimised against global properties such as their solvation free energies and then transferred to the description of large molecules composed of these chemical groups. Similarly, the MARTINI force field chooses this top-down approach to parametrise non-bonded interactions of the beads, tuning them against experimental partitioning free energies between polar and apolar phases. On the other hand, bonded interactions are derived from reference all-atom informations, in a bottom-up approach. Specifically, they were designed to match the structural data of the underlying atomistic geometry (for example bond lengths of rigid structures), derived either from available structures or atomistic simulations. For the second case, each frame in the atomistic simulation is converted (“mapped”) to its coarse-grain description and the distribution of a specific property (e.g. a bead-bead bond length) is computed over the mapped trajectory. This is compared with the distribution obtained directly from the coarse-grain simulation and the coarse-grain parameters are systematically changed in an iterative way until the two overlap.

To be noticed that the four-to-one approach implies that the backbone of amino acids is represented by one bead only. This prevents the description of directional hydrogen bonds, which are key to reproduce the secondary structure of proteins. The bonded parameters partially account for this, favouring for each amino acid the backbone conformation in which it is most likely found (based on the distribution of bond lengths, angles, and dihedrals calculated from the Protein Data Bank - PDB). When this is not sufficient, to constrain the protein to a particular state, an elastic network model approach is used (ElNeDyn [145]), together with the standard force field parameters. Both the backbone parametrisation and the possible use of ElNeDyn imply that large conformational changes in the secondary structure are penalised and therefore not well sampled in MARTINI simulations.

Some molecules obviously require a deviation from the general four-to-one approach: in ring-like molecules, two heavy atoms are mapped to a bead, to preserve the circular topology. While all the other beads are represented with the same value of the mass, regardless the composition of the atomistic structure they refer to, ring beads have a lower mass, according with the fact that they include less heavy atoms.

The MARTINI force field provides two water model. The first one (historically) groups four water molecules in a bead and therefore suffer from the non-polarisability problem mentioned above. MARTINI simulations employing this water model thus opt for a high dielectric constant to reproduce the solvent screening. Later on a polarisable water model was designed [146]: it maps four water molecule to a single “inflated” water, i.e. a three-beads molecule with the same geometry and charge splitting of a single molecule, but expanded. This model allows to revert the dielectric constant back to a value closer to 1 (an exact value of 1 corresponds to no mean-field correction to the electrostatic interactions, meaning they are correctly modelled by the collective action of the atoms/beads described).

Overall, the MARTINI force field pushes the limits of simplification to enhance the simulations speed-up, with considerable gain in efficiency with respect to atomistic simulations. Despite it can not capture some fine details of the system studied, it has been successfully applied to describe the behaviour of many biological membranes, lipid self-assembly, peptide-membrane binding, and protein-protein recognition. The (re)introduction of a more detailed water model allowed the description of electroporation processes and translocation of ions through bilayers.

Whenever one wants to investigate long time processes, coarse-grain descriptions are more effective in achieving the required time scale; and to retrieve the details of such processes, backmapping techniques have been designed to obtain atomistic configurations from the coarse-grain ones visited in the simulations. These backmapped structures can in turn be simulated at the atomistic level to explore the short time scale movements around such interesting conformation, in a by now consolidated multi scale approach.

2.3 The search problem

Very often the aim of Molecular Dynamics simulations, or other computational techniques which investigate biosystems, consists in characterising the energy landscape and in particular in finding the configurations of minimal energy. For example in the case of a protein, to find all the folds which are energetically favoured.

Biomolecular systems have thousands of strongly interdependent degrees

of freedom, therefore their energy landscape is complex and rough, meaning that many local minima of energy are present. Ideally, the full landscape needs to be explored as the properties of the system are determined by the ensemble of conformations visited and how often each of them is adopted. However, statistical mechanics teaches that the configurations with lower energy have an higher contribution to the system, according to the Boltzmann weight:

$$P(x) \propto \exp(-V(x)/k_B T) \quad (2.16)$$

with k_B the Boltzmann constant, T the absolute temperature and $V(x)$ the position-dependent potential energy. Therefore the importance of investigating energy minima. At this stage, we voluntarily omit the entropic contribution, which will be discussed later: indeed, conformations of non-minimal energy can be important as well if a large number of microstates corresponds to them, i.e. many different rearrangement of the internal degrees of freedom give the same macroscopic outcome (which is the definition of high entropy).

At the core of every energy landscape exploration lays the potential energy function, as modelled in the force field, but the initial configuration plays an important role as well. Indeed, many techniques perform a local search of the landscape in the vicinity of the starting conformation, and regions further away are sampled only in much longer runs. Very often in the simulations of proteins the initial structure is derived from X-ray crystallography, however it is well known that this might not represent the native state of the protein in solution nor the functional form of interest, making the convergence toward the desired structure a long process.

Different techniques have been developed to sample efficiently the energy (and thus conformation) space, and a non exhaustive list comprises:

- generating a series of independent configurations for the system to cast the search problem into a distance-based form (in the so-called distance-geometry metric-matrix method [147, 148]);
- building a system configuration from the configurations of its fragments in a stepwise manner (for example in the Monte Carlo chain-growing methods [149]);
- using step methods, where a new configuration is derived from the pre-

vious one. Energy minimization and Metropolis Monte Carlo are step methods [150]. Molecular Dynamics falls as well in this category, and for this technique the step is intuitively associated to time.

MD simulations are particularly interesting as they propose to reproduce the “true” relaxation of a structure toward its energy minimum, as it would be observed in nature. However, they struggle in investigating large systems and reproducing processes undergoing slow transitions because of their computational cost, making MD a somewhat poor techniques for the full characterisation of the energy landscape. For this reason, many techniques have been designed to overcome such impediment, giving rise to the field of enhanced MD, and many expedients are put in place to limit the search to interesting area of the phase space.

Only coarse graining would be considered in this work among the enhanced MD techniques, but it is interesting to understand the flexibility and possibilities of MD facing the search problem. Possible include: 1) smoothing and deforming the potential energy surface, 2) enhancing the pace at which the space is explored or 3) forcing the exploration of new/interesting regions only.

The first can be achieved for example using long range distance bonds based on experimental results (e.g. Nuclear Overhauser effect - NOE - data) [151]; softening geometric restraints derived from NMR or X-ray data through time averaging [152]; or finally using “soft-core” atoms, thus reducing the Pauli repulsion among them [153]. An enhanced exploration pace can be obtained using higher temperatures to overcome energy barriers thanks to the acquired kinetic energy [154], scaling the mass to reduce inertia [155], or combining multiple simulations together (for example in replica-exchange algorithms [156] some configuration are extracted from simulations held at different conditions and used to feed a new set of simulations). Finally, avoiding the re-sampling of energy minima can be reached through local potential-energy elevation [157, 158]; while constraining the high-frequency degrees of freedom (for example non-polar hydrogens) avoids spending time computing non interesting fine-details [159].

Coarse-graining of the model to reduce the number of interaction sites is another widely employed and effective technique to speed up the sampling (two examples of coarse-grain force field have been given in Section 2.2): a coarse-grain potential discards the high-frequency or less interesting degrees

of freedom, and at the same time gives a smoother energy surface, so that the search is not trapped into local minima due to the landscape roughness.

Alongside the aforementioned techniques, a set of expedient allows to reproduce at best the natural conditions while keeping the complexity low: for examples periodic boundary conditions approximate an infinite system even simulating a small portion of space (Figure –); moreover, as often done in this work, the initial conditions are chosen carefully to sample the regions of interest, based on some prior knowledge or to test some hypotheses.

Thus, the outcome of MD simulation is a (local and incomplete) sampling of the configuration space. If on one hand the search problem if further complicated by the fact that many different conformations can be equally important (the ensemble or entropic problem), in the case of biomolecular system it is never the less alleviated by the common knowledge accumulated on them, and such knowledge is coded in the energy functional of the force field commonly employed: for example, only a few rotamers of the common amino acids are favoured, according to the informations gained from X-ray crystallography, thus avoiding the sampling of high-energy, unfavourable conformations.

2.4 The ensemble problem

In the previous paragraph the exploration of the energy landscape was indicated as the major goal of MD simulations. Despite energy (U) is often the reference quantity for the investigation of biomolecular systems, it is the combination with entropy S in the form of free energy ($F = U - TS$) that drives the evolution. Many states of the system can have the same free energy while having different energetic and entropic contributions, and while some processes are dominated by the variation in the first, others are governed by the second. This also means that a configuration with an energy higher than the minimal possible one can still determine the behaviour of the system if such configuration can be obtained by more microstates (entropic contribution).

For example, the preferred state of a solvent is highly governed by its entropic contribution, as many molecules contribute to it. Accordingly, among all the possible folds of a peptide in solution, the presence of solvent molecules can shift the preferred fold to a conformation of non-minimal energy for the protein itself, because it results in a lower free energy for the whole protein-solvent

system. As the entropic contribution in free energy is weighted by temperature, this means that the preferred conformations adopted by the peptide are temperature-dependent.

Such pool of equally relevant conformations is the so-called ensemble: if at the beginning of structural biology the development of X-ray crystallography pushed forward the idea that a protein is fixed in one particular shape, in recent years the concept of ensemble has re-emerged, supported by techniques such as NMR. Their results can be correctly interpreted only assuming the protein adopts an ensemble of shapes, each visited for a given amount of time, while no one single conformation can explain the overall results by itself. In such context, MD simulations can characterise these conformations and estimate the time of residency in each, uncovering their relative importance.

Finally, it must be noticed that MD simulations are successful in computing free energy differences between states, as it is sufficient to sample extensively the region of the phase or configurations space where the two states differ. In contrast, to compute entropy differences requires the correct evaluation of the full Hamiltonian operator in both states and not only of the terms which are distinct. Some contributions (for example the solvent entropy) are very hard to compute as they require a prohibitively extended sampling for their correct evaluation. Even if some techniques have been developed to address the problem, at present they are still difficult to apply to the calculation of ligand-protein binding entropy or polypeptide folding entropy [160]. Thus, up to now, entropy computations remain under-represented with respect to free energy ones in the landscape of computational biology, diminishing the accuracy with which relevant biological processes, as the ones just mentioned, are modelled.

2.5 The experimental problem

The validation of MD simulations is performed by comparison with experiments: the properties obtained experimentally are computed from the MD trajectory as well, and the latter compared with the former. If these are correctly reproduced, it is usually assumed that the simulation is sampling the correct ensemble of states. This holds if the properties of the simulation are not drifting away, namely the system has reached equilibration and it is

thus in a stationary state. Once the simulation has been validated, one can identify, from the conformations in the trajectory, the details of the processes responsible for the experimental outcome of interest, as such information is not accessible by the experiment itself.

In such procedure it is not unusual that the measured and computed quantity do not match or that the interpretation of the comparison is hard. This can be due to several factors, which can be grouped into three classes. First, the average problem: the quantity measured by an experiment is almost always an average in time and/or space. For example, Circular Dichroism spectra and SAXS profiles of a peptide in solution are the convolution of the profiles cast by every conformation adopted by the protein in the time window of the measurement, averaged over all the copies of the peptide present in the sample. As such, even knowing the pool of possible peptide configurations from MD simulations, many different combinations can produce the same results: there is uncertainty in the weight each conformation is assigned, as well as the possibility that some conformations are missing in the pool computed. Directly from this arise the second challenge: the under-determination of the problem itself. Indeed, the experimental information is limited in comparison with the many degrees of freedom involved in the system, and with the ones handled by MD simulations. It is thus impossible to obtain experimental evidence proving the existence of each conformation in the ensemble or each detail in a particular process - and exactly in this lays the value of MD as integrative technique. Finally, the accuracy of the experimental data can be a limiting factor as MD resolution is usually higher than experimental one, suggesting the importance of mechanism not reachable by experimental verification. This problem will be likely alleviated in the future as experimental techniques get better and better.

From the examples above, it is clear the importance of MD simulations in accessing details of systems which are beyond the experimental reach, but it is also crucial to validate the simulations set up against experimental properties before using them for predictions. In such validation is important to have a critical attitude both when the results agree and when they do not. Indeed, agreement may arise from either a simulation that reflects correctly the experimental system; but also from a “wrong focus” of the attention, e.g. the property examined is insensitive to the details of the simulated trajectory and

thus always agrees with experiments; or finally from a compensation of errors, which happens more easily for systems with a high number of degrees of freedom. Similarly, disagreement may hint at an error in the simulation (either in the theory behind it, the model, the implementation or simply the simulation is not converged yet) or an error in the experiment (either in the result itself or its interpretation), so that both must be carefully checked to finally improve the agreement.

2.6 MD simulations: successes

Despite all the caveats listed above, Molecular Dynamics simulations proved to be a valuable tool to interpret experimental results, clarify biomolecular mechanisms and suggest new focus of attention for further research efforts. In the following, we want to highlight some success of MD simulations, with a special focus on the simulations of antimicrobial peptides and how they can aid peptide design, as well as on simulations of self-assembling blocks.

2.6.1 Simulations of antimicrobial peptides

MD simulations of antimicrobial peptides are quite well documented since the first developments of the technique. Such peptides are a suitable system for a computational investigation as, in most of the cases, their mechanisms of action are not completely understood from the experimental information available (see Section 1.2.2). As experiments prove that even the mutation of one single residue in short AMPs can change remarkably the antimicrobial activity of the sequence (see Section 1.2.4), it is then clear that their action is governed by subtle atomic interactions, so that MD simulations, with their atomistic resolution, can help in understanding this aspect.

Systems As mentioned in the previous chapter, it has been proposed that most AMPs act through a process of attraction to the bacterial membrane, possible aggregation with other copies of the same sequence, insertion, and membrane lysis. The time scales of the overall process are accessible to coarse grain techniques, but not - or rarely - to atomistic ones. For this level of description instead, the different steps are usually investigated separately, based on prior hypothesis: for example, the peptide can be positioned close to the

membrane surface with an orientation known to promote binding (from experiments or based on energetic assumptions), or again can be placed directly within the membrane core with different insertion depths and tilt angles to verify which configurations are the most disruptive ones. In this case, the full insertion process can only be reconstructed from a “stepwise” knowledge combining the different states sampled and further exploring the intermediate regions if necessary. For these reasons, the choice of the conformation to simulate, i.e. the initial conditions in terms of the mutual position of peptide and membrane, is crucial, as it likely biases the simulation towards the sampling of a particular subset of configurations, and this must be considered in the interpretation of the results. Recent advances are making possible the simulations of the full process even at the atomistic resolution for simple enough systems, as it will be shown in the following, nevertheless the “stepwise” approach is still common and the preferred one in case of complex AMP systems.

Model membranes The second important choice in the setup of a simulations of antimicrobial peptides concerns the model of the membrane to simulate. In an effort to keep complexity low, bacterial and mammal membranes can be modelled with a minimal number of lipids. Very often, models of bacterial membrane retain as only key characteristic an overall negative charge, with about 25% of the lipids being anionic ($-1e$ charge) and the rest being zwitterionic, i.e. neutral but with positively and negatively charged regions separate in space (see Chapter 4) [161–164]. For a model mammal membrane instead, only zwitterionic lipids are employed, with the occasional inclusion of cholesterol, as it is deemed important in achieving the flexibility typical of mammal membranes [161–165]. Because of their simplicity, very similar or identical systems are used also in experiments, when the use of cells is prevented by the experimental conditions necessary for the technique chosen, or to test the working hypothesis on a system easy to interpret [8, 38, 54]. Therefore, even if these simple membranes don’t model accurately the structure of the cellular envelope, simulations and experiments of these systems can provide a first explanation of some steps of the antimicrobial activity, with the two techniques complementing and validating each other.

Nevertheless, attempts to model more accurately cell membranes have been pursued. This can be performed at the atomistic level [166] but the task is

especially suited for a coarse grain description, as the inclusion of all the elements of the cell membranes results in quite large systems for which atomistic computations started only in recent years to be affordable. Accordingly, coarse grain (MARTINI) simulations have been incorporating more and more components into model membranes, describing the bacterial inner membrane, the bacterial wall, and finally the combination of the two [167]. These large scale, coarse grain simulations provide information on the mechanic characteristics of the system: for example, simulation of the outer membranes of Gram negative bacteria combined with the peptoglycan layer (which, in bacteria, is positioned between the two membranes) elucidated how the distance between the two is variable, thanks to the presence of Braun's lipoproteins which act as a bridge between the two, and can bring them closer by bending and tilting. On the other hand, the permeability of membranes to ions and small compounds needs to be assessed at the atomistic level, and to access informative simulation time scales, smaller and simpler systems must be chosen for the task (e.g. the inner membrane only), often together with enhanced MD techniques such as umbrella sampling [166, 168].

In the context of assessing the antimicrobial activity of a sequence, the choice of a minimally simple membrane might even have some advantages in terms of the investigation pipeline: ideally, simple membranes are tested first, and then their complexity is increased to verify which element triggers the antimicrobial activity and the selectivity for bacteria versus mammals. Usually, the limitations in the computational time available have precluded such methodical analysis, directing the choice toward very few simple systems to be tested at once, but the improvements witnessed in the last years are allowing for a more extended sampling. For example, an *in silico* experiment simulated a dermicidin channel inserted into patches of membranes composed by different phospholipids and with variable cholesterol content [169]. Nine membrane compositions were tested overall resulting in different membrane thickness, thus in a different orientation of the dermicidin channels inserted into it, with consequent variance in the conductance of the channel itself. This structure-function relationship shows the importance of an accurate membrane model to fully capture all the aspects of AMPs activity. Notably, the simulations were performed with the coarse grain model MARTINI, showing that a supra atomistic view retains enough details to investigate such systems.

Force field comparison Finally, simulations of the peptide interaction with a model membrane are clearly determined by the parametrisation of the force field employed for protein and lipids (and by their mutual consistency). There are multiple evidence suggesting that different force fields produce very different outcomes when simulating the same system, under the same conditions. This is valid for simulations of pure lipid patches (see Chapter 4), resulting in incompatible values of area per lipid, organisation of the tails and energetic profiles across the membrane, and thus has an impact in the simulations of AMPs interacting with a membrane.

For example, Wang et al. [170] run simulations of the antimicrobial peptide melittin with different force fields, namely CHARMM27 and 36 (for protein and lipids respectively) [134, 135], OPLS all atoms (for protein) and united atoms (for lipids) [171] and GROMOS 53a6 [129] (and the TIP3P water model for all of them [172]). Despite these parameterisation have similar values of partial charges on the different atoms, and similar bonded interactions at the protein level, the level of unfolding of melittin in the membrane was significantly different (with NMR experimental results on melittin bound to a membrane showing an almost completely folded state). Most likely this can be attributed to the fact that lipid and protein parametrisations are obtained separately and might present some inconsistencies. The most evident example is the OPLS case, for which an all atom description of lipids is not available and a mixed resolution description has been adopted. In the case of GROMOS, both components are parametrised at the united atom level, but their mutual consistency might be questioned as well, as fully explained in Chapter 4. In general, a united atom description is clearly less accurate than an all atom one, and in the case of lipids it has been postulated that it is not able to represent faithfully the dynamical processes happening in the hydrophobic tail region.

A similar investigation has been proposed by Bennett et al., proving that the propensity of the synthetic AMP CM15 to form pores strongly depend on the force field used but also on some extent - at least at the time of the work - on the MD engine used (GROMACS compared to NAMD [173]). This should not come as a surprise because the membrane characteristics emerge from the collective behaviour of lipids, so that a small difference in the way their interactions are treated might be amplified resulting in different macroscopic

outcome for the simulations [174].

A more systematic investigation on the topic has been performed by Sandoval-Perez et al. [175], focussing on the reproduction of membrane-protein interactions in different force fields. Transmembrane protein positioning, amino acid side chains insertion depth and reproduction of free-energies of adsorption of Wimley-White peptides were tested, showing that all the force field considered (GROMOS 54a7 [130], CHARMM36, Amber14SB/Slipids [176] and Amber14SB/Lipid14 [177]) were able to reproduce the overall experimental results for the first two criteria. Each parametrisation had different points of strength, for example performed better in reproducing the insertion of some amino acids with respect to others, but this was observed in a non systematic manner, leading to the conclusion that for every particular system tested, the comparison with at least one experimentally measured quantity would be the only way to assess the simulation performance accurately.

Simulations of membrane-peptide interaction: examples Even in the context of a simplified model scenario and with the caveats coming from the chosen parametrisation, simulations of antibacterial peptides on a membrane have been successful in elucidating some of their mechanisms. The first important contribution consists in the introduction of the disordered toroidal pore concept: as explained in the previous chapter (Section 1.2.2), the models of membrane poration due to AMPs consist in quite ordered structures, with peptides contouring the pores, either being in contact with the hydrophobic tails of lipids (barrel-stave model), or with their head, as these ones bends around the pore to keep their tails screened from the outside environment (toroidal pore model). However, simulations of the short helical peptide magainin MG-H2 [178], among others, showed that a single copy of this helix was sufficient to make the lipid rearrange and form a water-filled pore, with the helix in none of the positions suggested in the two above models. This does not disprove them, but proposes a novel mechanism which is clearly less disruptive, as it produces smaller pores, but at the same time statistically more favourable to happen, as it requires the insertion of one copy of the peptide only.

Regarding possible rearrangements of the antimicrobial peptide structure when interacting with a membrane, simulations of cathelicidin LL-37 on pure POPG (anionic) and POPC (zwitterionic) lipid patches showed that LL-37

has a propensity to bind to the former, as expected due to the opposite charge that membrane has with respect to the cationic peptide [163]. However the simulations highlighted also that, in contact with POPC, the helical secondary structure was lost, while the interaction with POPG preserved it, suggesting that the spacial arrangement of the residues, and not only the overall chemical character, is important for their action. Such type of information is hardly available to experiments or through a theoretical reasoning.

Further insight into the role of the secondary structure was obtained simulating the helical antimicrobial peptide CM15 nearby a POPC membrane, starting from a fully structured helix or from a coil configuration: Wang et al. [162] proved that the interaction with the lipids is stronger when the peptide approaches the membrane in its disordered form rather than in a fully formed helix. The two have similar electrostatic and van der Waals energy contributions, however the larger flexibility of a coil arrangement allows for more residues to come in contact with the membrane at once, triggering a faster adsorption. It is important to be aware of the existence of such biases in the setup of simulations, not to incur in unfunded conclusions: as pointed out by the authors of this study, a low propensity for the α -helical fold to bind to the membrane would have result easily in the inference that the peptide does not bind to it at all, if a few repeated simulations of 100 ns length would have been used only - as it was common procedure (and the maximum simulation time available) up to a few years ago.

The improvements in computational resources is slowly removing some of these obstacles, pushing the extent of simulation time to the microsecond timescale. In a recent example, the translocation of the helical PGLa peptide through the membrane has been observed as a rare event on the multi microsecond timescale without the formation of an organised pore [179]. The *in silico* experiment still benefited of an enhanced sampling in the form of a higher temperature used for the simulations, but no pre-insertion of the peptide was performed. In this work, the insertion of the peptide and consequent switch to the opposite side of the membrane was observed more often when many peptides were on the surface: even if only one copy was inserted at once, the presence of other copies at nearby locations helps in destabilising the membrane, favouring the translocation of the first. This study shed light on a possible mechanism of permeabilisation which is usually overlooked in favour

of processes involving organised channels and pores. The fact that no organised neither disorganised pore is observed matches the experimental results which can not identify such structures for the peptide considered. Thus, these simulations can not exclude that other mechanisms of translocation proceed via pore or intra-membrane oligomer formation, as they might be observed on a longer time scale, but prove the existence of other competing processes resulting in the penetration of the peptide.

Similarly, simulations were able to shed light on the mechanism of translocation of Arginine-rich peptides, proposing a mechanism of action on an otherwise puzzling problem [180]. These sequences have high positive charge, but despite this, possess a high propensity to penetrate membranes, overcoming the hydrophobic region represented by the lipid tails. A commonly used explanation considers polyarginine translocation a quasi-equilibrium process, so that the occasional penetration rate is governed by the free energy cost of pore or lipid defect formation. However, very similar peptides where the Arginines were swapped with Lysins showed no significant penetration, while the above quasi-equilibrium reasoning would hold for them as well, as they possess a similar energetic profile. After extensive simulations of the two different systems (multiple, hundreds of nanosecond long runs), the proposed mechanism involves dynamical considerations on the spontaneous formation of thermal pores: such events are rare, but still observed on the time scale of the simulations performed. In some of these events, the transient pore would be occupied by a peptide (a precursor), slowing down its dynamics and thus closure. In such situation, the translocation of other copies of the peptide is highly favoured and follows shortly after if the concentration on the membrane and thus the number of peptides around the precursor is sufficiently high. Indeed, other copies of the peptide are driven to aggregate with the precursor inside the membrane and are then pushed toward the opposite side as there is a lower charge density in that region. Differently from polyarginines, polylysins have a much lower aggregation propensity, so that the presence of a precursor peptide inside the membrane does not induce an enhanced insertion of further peptides, explaining why only the former ones show a high penetration propensity. These types of mechanisms would be hardly understandable without the atomistic insight of MD.

The last two examples mentioned bring the attention back to the discus-

sion on whether and for which peptides oligomerisation is necessary for an efficient antimicrobial activity. MD simulations can offer an insight as well on this aspect, which appear not to have a unique, simple answer. Contrary to oligomerisation in solution, which can happen on shorter time scale, the spontaneous aggregation of peptides on a membrane surface requires a long time, as the structures must diffuse on the membrane to meet each other, and many other competing processes (such as insertion) are happening at the same time.

A recent example of how MD elucidated oligomerisation mechanisms comes from simulations of maculatin (an helical AMP), which showed that the pores it forms can include a variable number of helices and thus assume many different conformations [181]. The suggested process of pore formation proceeds via insertion of a single residue, closely followed by other ones which are able to penetrate the membrane thanks to the lipid defects already created by the first peptide. This is at odds with other models proposed beforehand, which include the insertion of already pre-formed peptide aggregates and suggest a unique “rigid” form of the pore, but it is actually consistent with the experimental findings.

Similar investigations can be carried on also for other cell penetrating peptides, which are not antimicrobial: as such, some of them aim at inserting within cells without necessarily disrupting the membrane. One example is constituted by the influenza fusion peptides, which have been extensively studied with a simulation set up similar to the one mentioned for AMPs: a few copies of the peptide were positioned on a model membrane and their oligomerisation and insertion processes were followed in time, showing the formation of aggregates of different sizes which perturbed locally the membrane [182, 183].

To be noticed that, when investigating oligomerisation, the size of the system must necessarily be increased to include all the copies necessary to form the aggregates observed experimentally. As such, not all the systems can be investigated from unbiased initial conditions yet. In the case of protegrin, a β -hairpin antimicrobial peptide which has been long thought to act through the formation of transmembrane β -barrels, many variables can influence the outcome of the final structure, which is unknown. Following from the previous example, it is also likely that many conformations are equally possible, or that some might arise as rare events. Even with the computer power available

now, it is unlikely to sample all the possible conformations resulting in stable or transient β -barrel pores in a free simulations starting from a few peptides scattered on the surface, and this hinders the understanding of their relative importance. To overcome such problems, a semi-systematic investigation has been carried on by Lazaridis et al. [161], simulating different organisation of multiple copies of the peptide at the interface with the membrane or inserted into it. In particular, the simulated assemblies varied in the number of protegrin employed to form a structure, in the orientation with which protegrin copies were assembled (e.g. parallel or antiparallel), and in the geometry of the structure (disorganised bundle, barrel, sheared barrel). Microsecond long simulations discriminated which ones of these initial configuration formed stable pores for the whole length of the simulations, and the ones which were disrupted. As in the previous example, several different possibilities were found stable in solution, suggesting that single AMPs might have multiple mechanisms of insertion into membranes.

Most of the examples above employ atomistic descriptions of the system. Similar investigations have been carried on also using the MARTINI force field, indeed the coarse grain description does capture the pore-forming behaviour of some AMPs. As an example, simulations of maculatin and aurein on POPC membranes showed different propensities for pore formation versus aggregation, showing that the model retains enough details and chemical information to reproduce different membrane perturbing behaviours [184] Nevertheless, the developers of the MARTINI model themselves pointed out how some aspects of pore formation might not be captured in a satisfactory way [185], e.g. the formation of pores only under particularly favourable conditions (thin membranes or high peptide concentration) or the fact that they were not filled with water as can be intuitively expected from a model which clusters four water molecules together (as some pore conformations allow for the passage of fewer water molecule if not one at the time).

In general, the outlook of simulations of antimicrobial peptides interacting with membranes goes in the direction of reproducing longer time scales thanks to the enhanced computational power available, trying to match the experimental findings showing that many antimicrobial related processes happen at the microsecond scale or beyond. This enhanced power would also reduce the need to use biased initial conditions or higher temperatures to speed up

the simulations. Moreover, gathering the contribution of the whole community, simulations will likely go in the direction of modelling more accurately the bacterial membrane, and while this is already at an advanced stage for coarse grain simulations, it is still an ongoing process for atomistic ones. Finally, the force field issue must be solved in collaboration with experimentalists, finding new tests and experimental quantities to compare the computational outcome with and make the different parameters sets converge toward a similar description of the phenomena observed, which is consistent with the experimental results.

Simulation-aided AMPs design The role of simulations in aiding AMPs design has been briefly sketched in Section 1.2.4. As pointed out, MD simulations are hardly a tool to analyse large dataset, therefore a systematic analysis can be performed for very small systems only, or, alternatively, the investigation can focus on a few selected sequences.

As already mentioned in Section 1.2.4, when classifying AMPs, simulations can be helpful in integrating structural information which is otherwise lacking, when no crystal structure of the peptide is available. Such approach was followed by Liu et al. [87] to complement the chemical information available on a dataset of short AMPs, and the overall information was used to feed a predictor of AM activity of novel sequences. Preliminary results showed that such structural information of minimal AM sequences improved significantly the ability of the predictor to discriminate whether a new sequence was suitable for antimicrobial activity or not.

Another commonly followed approach consists in using simulations to elucidate the reasons why a particular mutation is important and effective in terms of increased activity or decreased toxicity. Indeed, for short sequences, such mutation screenings can be afforded experimentally, thus there is little need to predict whether they would be beneficial. Rather, once assessed they are, it is interesting to understand why: for example, simulations of ovispirin and a mutant peptide with reduced toxicity showed that the bend in the helix in the latter was responsible for mitigating the interaction with mammal membranes and thus reducing haemolysis [186]. Again, the mechanism of lipid disordering and insertion by indolicidin was assessed through MD, and the amino acids responsible for each of them separately were identified, so that mutants could be designed with either reduced toxicity or enhanced potency [187]. Finally,

temporin and a derived sequence were investigate to discover that the mutant improved activity derived from a reduced aggregation propensity of the peptide in water, so that more copies were ready to bind to the membrane and thus disrupt it [91].

Many more examples can be listed, each with a slight different focus: the protocol of integrating simulations and design is usually customised according to the system in exam, as the field has not reached yet a systematic organisation. However, it is clear that simulations are in the positive feedback loop to help understanding better the already available AMP sequences, and thus contribute to device design rules for the creation of synthetic sequences with tailored properties.

2.6.2 Simulations of self-assembling peptides

Self-assembling peptides are another fascinating and challenging topic that MD simulations can help investigating. Simulating such systems implies different challenges with respect to the ones faced when simulating AMPs on membranes.

In theory, the set up of the system is quite straightforward: only the solvent characteristics and optionally the experimental salt concentration need to be matched, then a random initial configuration of the molecules - in the desired concentration - would allow the simulation of the process of interest. In reality, reproducing the experimental conditions often implies working with very large systems: with the level of dilution of the solutions employed in the experiments, a considerable volume needs to be simulate to host enough copies of the peptide to grant the assembly. However with this approach the time scale useful to witness a spontaneous assembly would greatly exceed the computational time available. For that reason, two main strategies have been adopted: coarse grain simulations or pre-assembled structures. Other routes include the choice of an implicit solvent model, or the use of other techniques such as Monte Carlo (a probabilistic exploration of the phase space rather than a dynamical algorithm) which can sometimes be less time consuming. Here, as we focus on MD simulations, we give a few examples of the first two strategies mentioned, which can be adopted in this framework.

Coarse grain models allow to track a reduced number of degrees of freedom during the simulations, and thus can achieve time scales sufficient to reproduce

the spontaneous assembly of peptides. Many works have been performed with the MARTINI force field to witness assembly of surfactants [188], polymers [189, 190] and lipids [191, 192], and a few focussed on peptides as well [193, 194]. With such approach, the systems which perform best with this approach are short sequence, for which a good sampling of many conformations for each of the copies can be reached in a relatively short time, so that the favourable shape for the assembly (if one is present) is adopted often enough to trigger the process.

For example, the assembly in water of peptide amphiphiles (PAs) into cylindrical fibers has been simulated at the coarse grain level [195], showing a transition from small micelles to longer fibres. This example of minimal PA structures is particularly interesting for the study of AMPs as well, as it share with them the amphiphatic character, so that having a general knowledge on how similar sequences assemble together would help in tuning their aggregation properties in water prior to the delivery to the membrane. In the work mentioned, pre-assembled fibres have been simulated as well at the atomistic level, to confirm their stability in solution.

The latter approach consists in preparing the system in a particular shape and using MD simulation to verify whether the conformation is kept or it is disrupted and which one out of many is the one most energetically favoured. It has been widely employed in cases where the final assembly was hypothesised to have a high degree of order, achievable only with a long sampling.

A similar dual approach has been used to prove first that a branched peptide can self-assemble in bilayers, and then that a larger hypothetical structure assembled in the shape of a capsule was stable in the run time of the coarse grain simulations employed. Specifically, the capsule has been build to match the peptide density on the self-assembled double layer and to respect the constraints derived from the presence of the curvature [196].

Similar approaches have been crucial in elucidating the assembly process of viral capsids: capsids are very large systems and the assembly of their protein subunits is mediated by energy barriers. For such reasons, already pre-assembled systems have been simulated to understand the interaction between the components and thus the first mechanisms of the assembly. This has been done recurring to ultra coarse grain or elastic network models [197] first, and only in the most recent advances to atomistic simulations [198, 199]. Ad-

ditionally, smaller portions of a capsid can be simulated, to obtain a minimal information on the cohesion of its blocks [200].

The examples above show how Molecular Dynamics simulations have been employed for the investigation of many different systems during the years, adapting their resolution, set up and the techniques related to better query the system of interest. Such overview suggests then that simulations would be a suitable tool to investigate the system of interest of this work, namely the self-assembling antimicrobial peptide capzip. The two aspects of its behaviour will be studied separately, adopting the necessary approximations and strategies to make the simulations efficient and to query the related questions at each time.

The details of the systems simulated and specific parameters used for each of the system simulated in the following chapters can be found in the relative sections, together with an extensive explanation of the motivation of the choices made.

Chapter 3

Capzip simulations

In Chapter 1 we introduced the molecule capzip and its properties, highlighting the unknowns of its mechanisms of action. In Chapter 2 we presented a review on Molecular Dynamics simulations, proving their successes in elucidating the behaviour of self-assembling and antimicrobial peptides in the past. Now, we employ this technique to understand better our system of interest. Given the exoticism of the unit, and the little atomistic information at disposal, modelling such peptide must proceed in a stepwise manner.

The first aim is to elucidate which structures it forms in solution and what interactions are keeping the molecules together. To understand the latter ones it is important to retain the highest level of detail possible, and for this reason we resorted to atomistic simulations first. This description has an high computational cost though, preventing the simulation of very large systems for a very long time, as it would be required to reproduce the natural assembly from a dispersed solution. Thus, we simulated increasingly complex pre-assembled blocks, verifying each time their behaviour in solution and inferring whether they are suitable to form a stable supramolecular structure. This approach, fully explained in Section 3.1, lead to the model of a minimal capsule, which has been subsequently investigated at coarser levels to explore its behaviour on longer time scales.

Of the pre-built structures a few selected ones were simulated in contact with model membranes to understand the determinants of their antimicrobial activity. Details on these simulations and the specific techniques employed to enhance the sampling are given in Section 3.2.

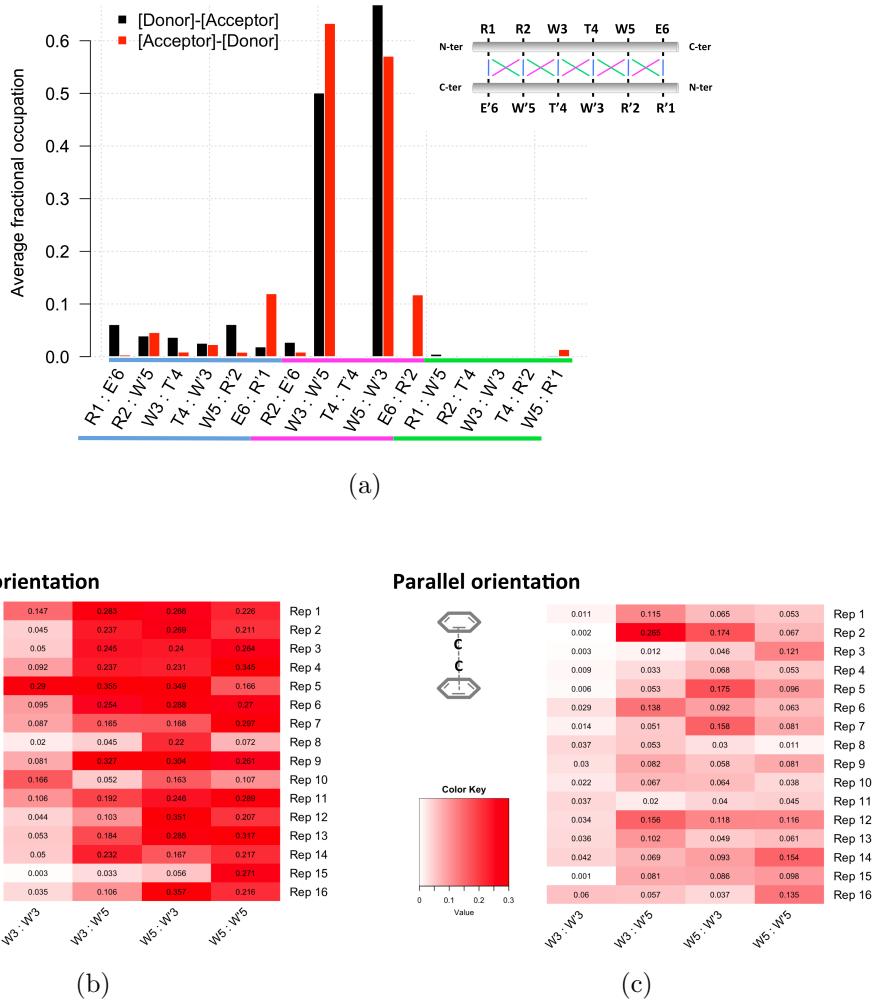


Figure 3.1: (a) Presence of backbone hydrogen bonds between amino acids in two facing antiparallel RRWTWE chains. The top right inset shows a scheme of the initial configuration. All the pairs highlighted in blue, green and pink are reported in the histogram; the same color code appears in the bar labels. Occupancy is averaged over 16 simulations of 20 ns. (b, c) For each replica and possible pair of facing Tryptophan residues in the β -sheet, the map gives the fraction of time for which a parallel or perpendicular π -stacking interaction has been observed. The second and fourth column correspond arrangement with the most populated hydrogen bonds.

3.1 Modelling the assembly

As previously mentioned, the antimicrobial sequence of capzip is designed with opposite charges at its extremes to favour an antiparallel β -sheets pairing with other copies of itself. MD simulations of two RRWTWE sequences paired in

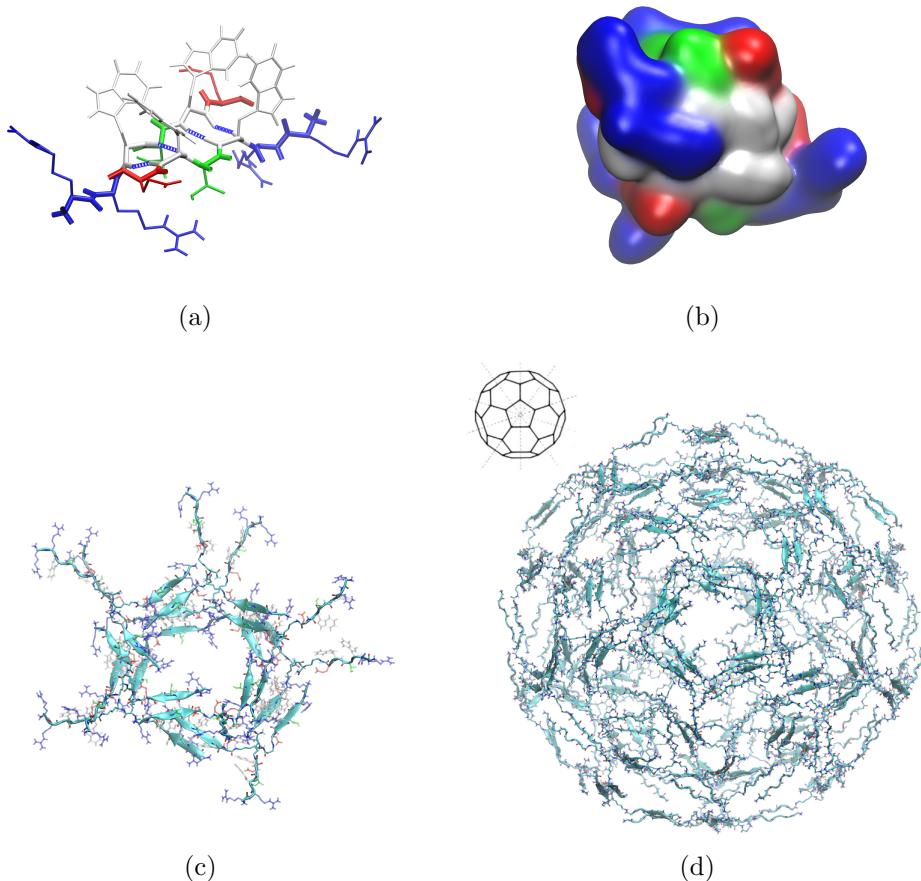


Figure 3.2: (a) Detail of β -sheet pairing with facing Tryptophan residues forming hydrogen bonds between their backbone atoms (bonds representation coloured by residue type, and hydrogen bond representation). (b) Two stacking β -sheets in surface representation, coloured by residue type. In white the partially buried hydrophobic patch. (c) A pentagonal subunit: ten antimicrobial molecules arranged in two stacking pentagons. Chains are paired in antiparallel β -sheets within each pentagon, and the two are interfacing with their Tryptophan residues in contact. (d) Atomistic structure of the buckyball simulated (bonds and cartoon representation) and geometrical model for comparison.

this fashion confirm that the assembly is stabilised by opposite charge interactions (with statistics gather over 16 replicas, each run for 20 ns). Moreover, backbone hydrogen bonds form between Tryptophan residues of facing strands, after a rearrangement of the mutual position of the backbones (Figure 3.1(a)). Finally, π -stacking contributes to the interaction as well, albeit in minor measure (Figure 3.1(b), (c)).

The favourable hydrophobic interactions between Tryptophan residues re-

sult in the creation of a hydrophobic patch which includes four of them (in white in Figure 3.2(a) on one side of the β -sheet plane. This creates an amphiphilic structure where the hydrophobic core is segregated from the remaining charged residues distributed at the other positions. The combination of two stacking β -sheets, paired to match their hydrophobic patches, constitutes an effective supramolecular assemblies to reduce solvent exposure of such residues.

This pairing strategy, however, needs to be applied in the context of full molecules assembly. The quasi three-fold symmetry of capzip suggests a regular geometric arrangement. The best examples of organised protein structures can be found in viral capsids, which are composed by the regular repetition of highly symmetric protein subunits. Inspired by this, we tested whether a geometrical organisation can represent a stable capsule, choosing as representative geometry a truncated icosahedron (buckyball).

Preliminary atomistic simulations (100 ns) were run on a pentagonal subunit formed by ten molecules arranged in two stacking pentagons (as in Figure 3.2(c)), proving the cohesion between molecules belonging to the subunit. Specifically, the number of contacts between backbone C_α s did not decrease in time but augmented slightly at the beginning, due to the compaction of the unpaired arms toward the core of the structure (Figure 3.3(a)). Moreover, for each pair of facing chains, it is computed the distance between their centres of mass. Figure 3.3(b) reports the variance of this distance over its average value, as a measure of the cohesion of the subunit, showing that in the majority of the cases less than 2% of variability is observed.

The pentagonal subunit respects the building principles of a) β -sheet pairing between antimicrobial sequences and b) double layer structure to screen the hydrophobic patches, and will constitute a face of the icosahedron. Each capzip molecule is centred in one vertex of the polygon, with the branches laying alongside the edges departing from it. On each edge two branches coming from opposite sides meet in an antiparallel fashion. When possible, a β -sheets with paired Tryptophan residues is organised (as in Figure 3.2(a)). The stacking pentagons interact through the hydrophobic patches of their β -sheet which are arranged in stacked positions.

The full truncated icosahedron was assembled from twelve pentagonal subunits (Figure 3.2(d)). As each subunit is formed by two stacked pentagons, the resulting structure has two concentric layers, for a total of 120 molecules,

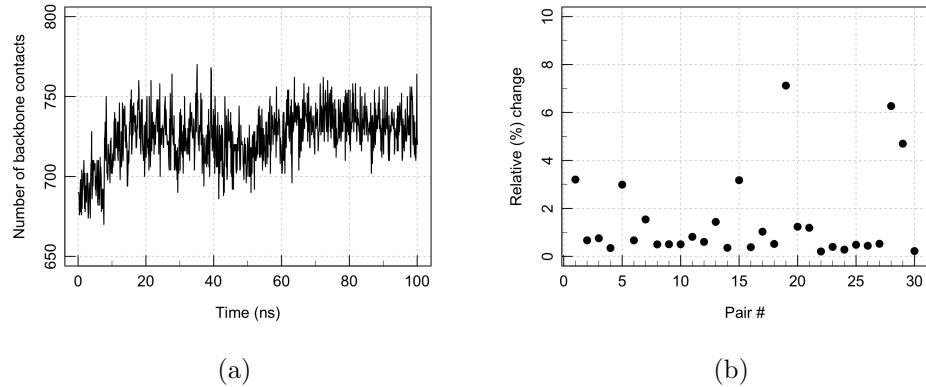


Figure 3.3: (a) Number of backbone contacts during a simulation of a pentagonal subunit. (b) Variability of the inter chain average distance between facing chains. The 30 pairs defined as facing are the chains belonging to the same β -sheet (5 for each of the two stacking pentagons), and for each stacking β -sheet the 4 possible inter-pentagon (inter-layer) pairs of chains.

and initial radius of 7.7 nm. This geometry represents a minimal model of the possible structures of capzip assembly in solution as, given the flexibility of the molecule, different geometries are possible, though proceeding from analogous interactions between the components. The final structure was simulated at atomistic and coarse grain levels, respectively with the GROMOS 53a6 [129], SIRAH [140] and MARTINI [142, 143] force fields (with both standard and polar water [146] to compare the two models). From the final configurations of the MARTINI coarse grain model (standard water), atomistic coordinates were obtained and simulated, to be compared with the original atomistic dynamics. Moreover, additional simulations were run at all the coarse grain levels on a structure made of one layer only (i.e. build from pentagonal block make by one pentagon only), to prove whether the bilayer structure was more energetically favoured.

A multiscale analysis is needed also to investigate the antimicrobial activity. Being highly costing to simulate a full truncated icosahedron on a bilayer patch at the atomistic level, the pentagonal subunit employed to build the complete structure (Figure 3.2(c)) was taken as representative of the latter. It was simulated close to the membrane plane, parallel to it, to avoid spending time in sampling non bound conformations (Figure 3.4(a)). This, together with a tailored use of an applied electric field (see Section 3.2), will speed up

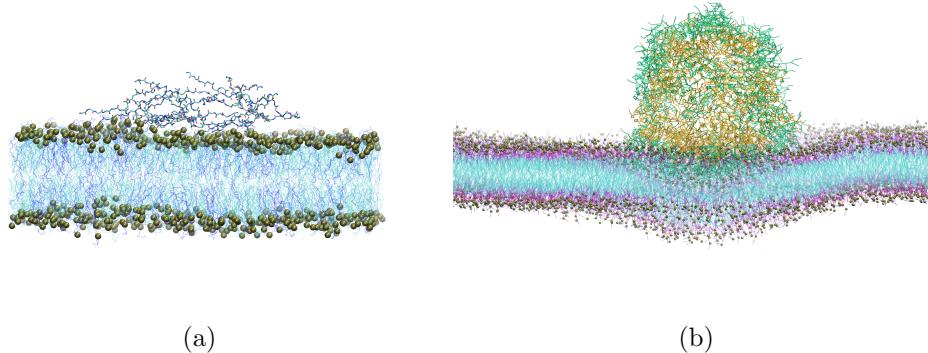


Figure 3.4: (a) Atomistic structure of a pentagonal subunit on a 740 lipid bacterial model membrane of composition DLPC:DLPG 3:1 (initial configuration). Peptide backbone in line and cartoon representation; lipid in cyan lines (DLPC) and blue ones (DLPG), all lipids phosphate in golden van der Waals beads. (b) Coarse grain (MARTINI) representation of the buckyball on a 2880 lipids bacterial model membrane (final configuration of the trajectory). Protein in bonds representation: green outer buckyball layer, yellow inner one. Lipid in line representation, coloured by bead type, and lipids phosphate in golden van der Waals beads.

simulations considerable.

To observe the natural binding of the peptide to the membrane, the process of the full buckyball approaching a model membrane was simulated with a MARTINI coarse grain description (Figure 3.4(b)).

Two membrane patches were simulated for both resolutions, a model bacterial and a model mammal membrane, to identify the different interactions with the peptide. The first one presents 25% of anionic lipids (DLPG), and the rest are zwitterionic (DLPC), while the second has only DLPC lipids.

3.2 Simulations details

Atomistic simulations All simulations were performed with the GROMACS software, version 5.5 and 2016 [131–133].

The atomistic coordinates for the peptidic supramolecular assemblies described in Section 3.1 were built combining GROMACS tools and the MOE software [201]. Simulations were run with the GROMOS 53a6 force field [129]. Parameters for the central residue connected with the peptidic chains are com-

puted with the ATB software [202, 203]; the ones for the joining bonds are derived from tabulated values of analogous peptide bonds.

The systems were solvated with SPC water [139] and counter ions were added (Na^+ or Cl^-); further ions are introduced to reach the concentration of 150 mM, to reproduce the experimental conditions. For simulations of a single β -sheets and of the pentagonal subunit, the systems were energy minimised with a steepest descent algorithm, then equilibrated in the NVT ensemble with decreasing positional restraints at increasing temperatures (100 K, 200 K, 250 K, 300 K and respectively 1000, 1000, 500, 250 $\text{kJ/mol}\cdot\text{nm}^2$ restraints, 100 ps each); then in the NPT ensemble, without restraints, at the same temperatures steps and for the same time. Production was followed for 100 ns.

For the truncated icosahedron structure the above equilibration was still insufficient. Due to the construction procedure, two thirds of the branches are not properly paired along the edges. Therefore, after an NVT equilibration as above, strong flat-bottom restraints (1000 $\text{kJ/mol}\cdot\text{nm}^2$) were placed between the center of mass of imperfectly aligned branches throughout the NPT heating, to penalise their mutual separation with respect to their initial distance (100 ps runs at 100 K, 200 K and 250 K and 35 ns at 300 K). This was followed by a series of 10 ns runs at 300 K with decreasing restraints strength (750, 500 and 250 $\text{kJ/mol}\cdot\text{nm}^2$) and by a free production run (100 ns). Three different replicas were run, generated from the final configuration of the 300 K NPT run with 1000 $\text{kJ/mol}\cdot\text{nm}^2$ restraints.

Throughout all the simulations, the temperature was maintained by independently coupling the protein and the solvent (plus ions) to two external temperature baths using a velocity rescale thermostat [126] with coupling constant τ_T of 0.1 ps. The pressure was kept at 1 bar by Berendsen [127] or Parrinello-Rahman barostat [128] (for the equilibration phases and the production run respectively) using an isotropic coupling, with isothermal compressibility $4.5 \times 10^{-5} \text{ bar}^{-1}$ and coupling constant τ_P of 1 ps. Electrostatic interactions were treated using the smooth Particle Mesh Ewald (PME) algorithm [137], with a short-range cutoff of 0.9 nm. The van der Waals interactions were treated with a plain 0.9 nm cutoff. A SPC water model [139] is used. All atomistic dynamic runs were performed using a 2 fs time step. An overview of simulations of peptide assembly in solution is given in Table 3.1.

The atomistic coordinates for the bacterial model membrane patch were

Capzip in solution simulations

System (Nr peptides)	Model	Time (ns)	Replicas
β -sheet (Fraction)	GR	100	32
Pentagon-bilayer (10)	GR	100	1
Buckyball bilayer (120)	GR	100	3
Buckyball bilayer & monolayer (120)	SI	1000	3
Buckyball bilayer & monolayer (120)	MA	1000	3
Buckyball bilayer & monolayer (120)	MA_P	1000	3

Table 3.1: Table of simulations of capzip assembly in water. GR = united atom GROMOS 53a6 force field [129], SI = coarse grain SIRAH force field [140], MA = coarse grain MARTINI force field [142, 143], MA_P = coarse grain MARTINI force field with polar water model [146].

built with the PACKMOL software [204], from pdb files of a single DLPC [205] and DLPG [206] molecule. Two patches were built, made respectively of 512 and 740 lipids, with composition DLPC:DLPG (3:1). The initial area per lipid was set to 7 nm^2 , above the values found experimentally for either lipid species ($0.608 \pm 0.012 \text{ nm}^2$ for DLPC [207] and $0.656 \pm 0.012 \text{ nm}^2$ for DLPG [208]). The correct area per lipid of the mixture was reached during a 400 ns equilibration, the final configuration of which was used for simulations with the peptide on the membrane. A similar procedure was held for DLPC, producing a patch of 748 lipids with an initial area per lipid of $0.656 \pm 0.012 \text{ nm}^2$. This patch was used for control simulations against the 740 lipids bacterial one.

The initial configuration of atomistic simulations of the peptide on the membrane was generated from the equilibrated lipid bilayer (after 400 ns run) and the equilibrated pentagonal subunit (after 100 ns run with positional restraints on the C_α), placing the pentagon plane parallel to the membrane one and close to it (Figure 3.4). The inflategro [209] script was used to solve the partial overlap of the peptide side chains with the lipid molecules, removing the ones overlapping. The sizes of the two patches fit the pentagonal subunit with respectively 3.5 nm and 5.4 nm distance between its periodic boundary images (along both x and y).

For simulations involving membranes, the version 54a8 [210, 211] of the

GROMOS force field was employed. Lipid parameters were taken from [205] for DLPC, while for DLPG they were built from the ones available in the literature for POPG [206].

The simulations set-up parameters are as above, except for the use of three thermal coupling groups (peptide, membrane, water plus ions), a semi-isotropic pressure coupling, and a larger cut off radius for both Coulomb and van der Waals interactions (1.2 nm). Additionally, for the 512 lipids membrane, a Reaction Field [138] was used instead of PME long range electrostatic treatment. Control simulations on membrane patches without peptide showed that the results in terms of area per lipid are compatible with the ones obtained from PME.

Each membrane patch was first equilibrated for 50 ps in NPT conditions at 50 K, then the temperature was gradually increased up to 300 K in 500 ps, and finally a 400 ns production was run. The final configuration was used as initial structure for the peptide-membrane simulations. A similar equilibration procedure was followed for peptide-membrane systems.

Additional simulations were performed applying an external electric field to the membrane, from the side hosting the peptide to the opposite one, to mimic the membrane potential and verify how the presence of the peptide affects the membrane response to external stimuli. In a first test performed on the 512 lipids bacterial patch, the field was increased by 20 mV/nm steps every 200 ns, until reaching the electroporation value, which resulted to be 130 mV/nm. For the bacterial membrane patches of both sizes, other simulations were performed starting from the unperturbed membrane configuration and the threshold field, in three replicates each. Three control runs were performed on the 512 patch without peptide, at the threshold value of the field, and one at the higher value of 140 mV/nm.

To be noticed that membrane potential is estimated around 2 mV/nm (from a -70 mV potential [212] and an estimate membrane thickness of 4 nm), but previous computational work explored the effects of fields up to 500 mV/nm [166, 213, 214], to witness poration within the simulations time, according to the resources available at the time.

An overview of simulations of peptide-membrane systems is given in Table 3.2 (control simulations on pure membrane are listed in Table ??).

Capzip on membrane simulations

	System	Model	Time (ns)	Replicas
Bacterial	10 peptides, 512 lipids	GR	500	2
	10 peptides, 740 lipids	GR	500	1
	Buckyball, 2880 lipids	MA	10000	2
	Buckyball, 2880 lipids	MA_P	10000	1
Mammalian	10 peptides, 740 lipids	GR	500	1
	Buckyball, 2880 lipids	MA	10000	1
	Buckyball, 2880 lipids	MA_P	10000	1

Electroporation simulations

	System	Model	Time (ns)	Replicas	E (mV/nm)
Bact.	10 peptides, 512 lipids	GR	75, 20, 71	3	-130
	10 peptides, 740 lipids	GR	60, 50, 70	3	-130
Bact.	Buckyball, 2880 lipids	MA_P	??	1	??
	Buckyball, 2880 lipids	MA_P	??	1	??s

Table 3.2: Table of simulations of peptide-membrane complexes. All the mentioned ones run at 150 mM concentration of NaCl. GR = united atom representation (GROMOS force field), MA = coarse grain MARTINI force field, MA_P = coarse grain MARTINI force field with polar water model. For electroporation simulations, the time refer to the disruption of the membrane. For electroporation simulations on pure membranes, see SI Table ??.

SIRAH coarse grain simulations SIRAH coarse grain simulations are run with the SIRAH force field [140]. Peptide coordinates for the buckyball geometry were obtained from the atomistic ones using the converter distributed with force field. Parameters for the central residue were built from comparison with similar chemical moieties.

For these simulations, the temperature coupling was performed with a velocity rescale thermostat [126] and coupling constant τ_T of 0.1 ps, and the pressure coupling at 1 bar pressure, with 4.5×10^{-5} bar $^{-1}$ isothermal compressibility, using a Parrinello-Rahman barostat [128] with a τ_P of 6 ps. Electrostatic

interactions were treated using the PME algorithm [137], with a short-range cutoff of 1.2 nm and relative dielectric constant of 1. The van der Waals interactions are treated with a plain 1.2 nm cutoff.

After energy minimization, a 4 ns NVT equilibration was run at 300 K, followed by a 10 ns NPT run, both with positional restraints ($1000 \text{ kJ/mol}\cdot\text{nm}^2$) on the solute. Two 10 ns run (NPT ensemble, 300 K) were then performed with backbone restraints of 1000 and $100 \text{ kJ/mol}\cdot\text{nm}^2$, respectively. Similar to the procedure adopted for the atomistic simulations, during the latter, flat bottom positional restraints ($1000 \text{ kJ/mol}\cdot\text{nm}^2$) were enforced on the unpaired branches during the NPT. Three additional 10 ns equilibrations were run at 300 K, with no backbone restraints and decreasing flat bottom ones (respectively 750, 500 and $250 \text{ kJ/mol}\cdot\text{nm}^2$). Finally the production run was carried on for 1 μs . All runs were performed with a 20 fs time step.

MARTINI coarse grain simulations For the MARTINI [142, 143] coarse grain simulations, peptide coordinates and parameters are obtained from the atomistic ones using `martinize.py` [144], and `pycgtool.py` [215] for the central residue. Parameters for the joining bonds are derived from tabulated values of analogous ones.

The bacterial and mammal model membranes, hosting 2880 lipids each, are built with `insane.py` [216], with composition DLPC:DLPG 3:1 and pure DLPC respectively. The simulations parameters used for lipids are consistent with Ref. [217]. The peptide-membrane systems are built placing the buckyball at a minimum distance of 1 nm from the membrane surface.

For simulations performed with the standard water model, the temperature coupling is performed with a velocity rescale thermostat [126] with a coupling constant τ_T of 1 ps. An isotropic or semi-isotropic pressure coupling is applied (for peptide in solution or peptide on membrane simulations) at 1 bar pressure, with $4.6 \times 10^{-5} \text{ bar}^{-1}$ isothermal compressibility, using a Berendsen [127] or Parrinello-Rahman barostat [128] (equilibration and production phase respectively) with a τ_P of 12 ps. Coulomb interactions are treated with a Reaction Field scheme [138] and cut off radius of 1.1 nm, van der Waals interaction with a cut off scheme and the same cut off radius. The relative dielectric constant is set to 15.

After energy minimization, a 500 ps equilibration is run with a 10 fs time

step at 300 K, followed by the dynamic run with a 20 fs time step. No restrained equilibration has been performed for the buckyball because the characteristics of the force field (and water model) grant a good cohesion between different peptidic branches and a compact structure is reached even with a standard short equilibration. The membranes used in the simulations are equilibrated for 1 μ s and the final configuration used to build the peptide-membrane system, which is simulated for 10 μ s.

Simulations performed with the polar water model are run with the parameters above, except the relative dielectric constant set to 2.5, and the choice of a PME scheme for the long range Coulomb interaction (1.2 nm cut off radius).

From the final configurations of MARTINI simulations of the buckyball in solution (standard water model), atomistic coordinates are obtained using the MARTINI backward tool (version 5) [218] and run for additional 200 ns with set up for atomistic simulations, to compare the original dynamic to one at a later time (as obtained with coarse grain simulations). This procedure is performed for two out of the three MARTINI-standard water replicas available.

3.3 Analysis

Simulations in solution Several structural analysis were performed on the outcome of the simulations of the buckyball in solution. Standard measures like the Radius of gyration (R_g), the Root Mean Square Deviation with respect to the initial configuration (RMSD) were computed (with GROMACS). To get the average distribution of the mass of the capsule, the Radial Distribution Function (RDF) of the protein masses around their center of mass was computed (with GROMACS). The profile could be fitted with a Gaussian function (with the R [?] software) as they presented negligible skewness, so that the position of its maximum can be taken as an average radius of the capsule, and its Full Width at Half Maximum (FWHM) as an estimate of the thickness of the bilayer.

Another structural measure concerns the pairing of the branches. Two branches are defined as paired if their center of mass is closer than a cut off distance of 1.2 nm. This simple measure discards any more precise information on the orientation of the chains with respect to each other, and aims at checking whether the network of molecules present in an ideal buckyball structure is

maintained. In the ideal buckyball, contacts within the same layer sum up to 90 for each layer. This measure can be easily applied to any description (atomistic or coarse grain) without disagreement in the interpretation. The computational pipeline combine GROMACS tools and a post-processing in R language.

The dynamical character of the structure was assessed computing the correlation of motion between the molecules. The central atom (or bead) from which the branches depart was taken as reference. For all the pairs i and j of such reference positions, the covariance of motion $\sigma^2(i, j)$ was computed (with GROMACS), where the covariance is the sum of the components along each axis: $\sigma^2(i, j) = \sigma^2(i, j)_x + \sigma^2(i, j)_y + \sigma^2(i, j)_z$. Then for each pair, this measure was normalised as:

$$\text{corr}(i, j) = \frac{\sigma^2(i, j)}{\sqrt{\sigma^2(i, i) \cdot \sigma^2(j, j)}}. \quad (3.1)$$

To characterise the chemical determinants that promote the assembly, we investigate the interactions between amino acids of different types, computing the number of contacts between backbone and side chains of single amino acids, filtered for the ones present at least 50% of the simulation time, and classified them by amino acid type. We define contacts between amino acids backbones if the C α of two residues (or the corresponding coarse grain bead) are closer than a cutoff distance of 0.6 nm; and between side chains if selected reference atoms in the side chain are closer than the same cut off; finally mixed ones if the proximity is between a C α and the side chain reference atom. As side chains reference, we took the heavy atom or bead farthest away from the backbone (respectively for GROMOS, SIRAH and MARTINI: CZ/SC2/BCZ for Arg, CZ2/SC4/BNE for Trp, OG1/SC1/BPG for Thr and CD/SC1/BCD for Glu). The functions to perform the analysis were built on the ones implemented in the MDAnalysis software (see Appendix –).

Further investigation has been carried on for atomistic simulations computing the hydrogen bonds between amino acids and grouping them by amino acid type and by region of occurrence (e.g. between two backbones, side chains or connecting a backbone atom and a side chain one).

Finally, the Solvent Accessible Surface Area was computed for each amino acid (with GROMACS). Its value was averaged in time and over all the residues

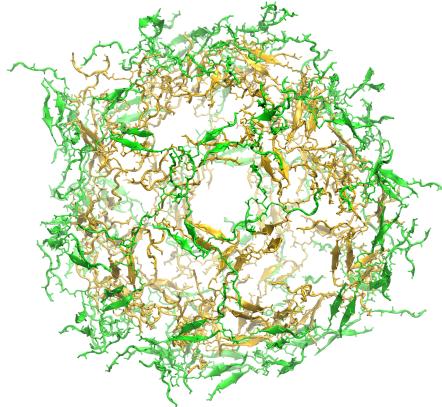


Figure 3.5: Final snapshots (100 ns) from the simulation of a buckyball in solution (replica 3): bonds and cartoon representation, backbone only, green external layer and yellow internal one.

of the same type. It was normalised over its reference value for each residue type X, obtained as the measure of the SASA of X from a Gly-X-Gly tripeptide. The resulting measure (named Q_{SASA}) takes into account the size of the side chain of each amino acid, giving a measure of exposure which can be compared between different residues. This normalisation is somewhat inappropriate for coarse grain models, however we employed it as it provides nevertheless a coarse regularisation for size effects.

Simulations on a membrane

3.4 Results: capsule in solution

We list here the results for the capsule in solution first: starting from the atomistic simulations, we then proceed to compare the outcome with the ones obtained by the different coarse grain models.

3.4.1 Atomistic simulations

Atomistic simulations of the buckyball in solution show a consistently equilibrated structure across the three replicas (Figure 3.5 and SI movie –). This is proven by both the stable value of the protein R_g and the almost plateauing backbone RMSD (Figure 3.6(a) and (b)). It is interesting to notice that previous simulations performed with a shorter equilibration, without the phase

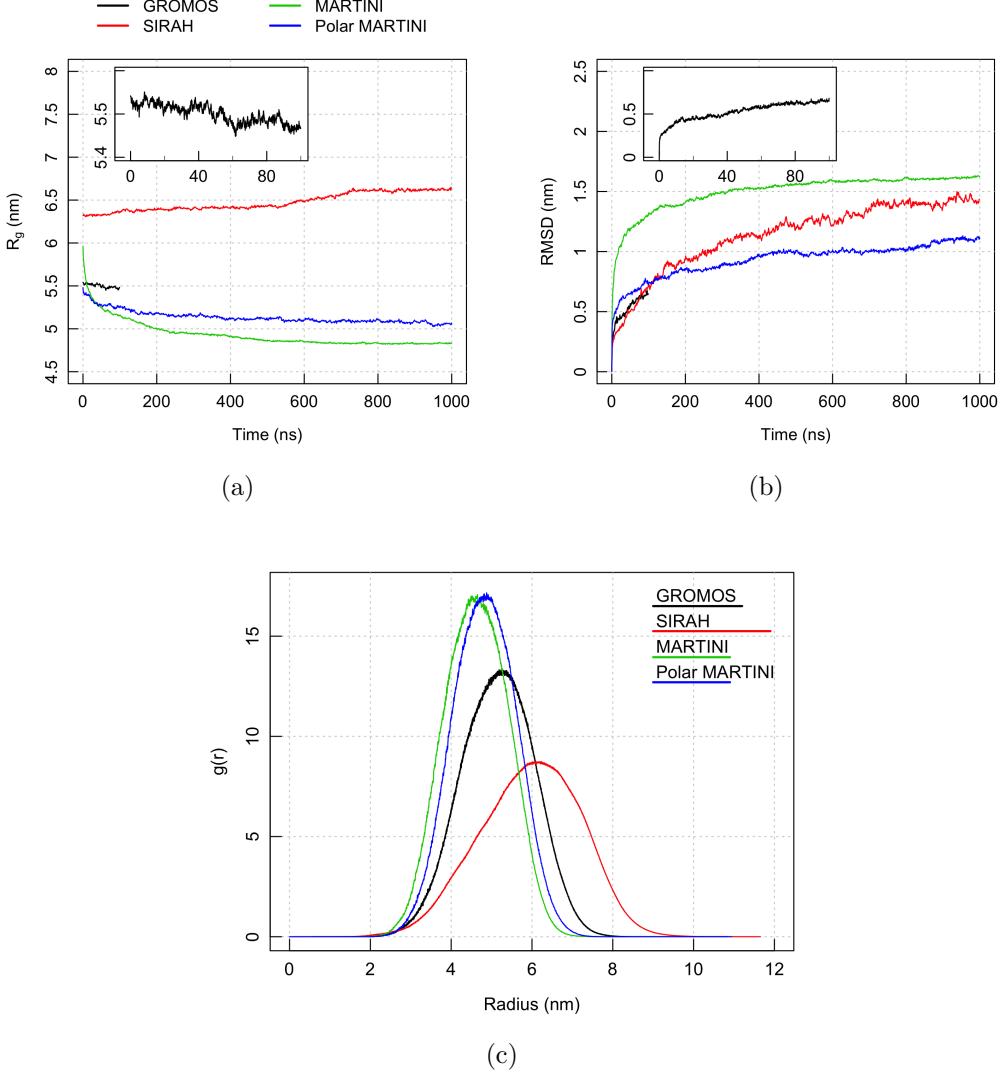


Figure 3.6: (a) R_g and (b) RMSD computed respectively on the Protein and its backbone. Results are displayed for simulations performed in GROMOS (100 ns), SIRAH, MARTINI and MARTINI with polar water (all 1 μ s). Inset: zoom on the GROMOS values. (c) RDF of Protein masses around their center of mass, computed on both layers, displayed for the same simulation set up as in (a,b). For each label of the legend, the bar has length of the respective RDF FWHM (thickness estimate). All results are showed for Replica 1 of each simulation set up.

employing flat bottom restraints, resulted in the immediate disruption of several connections in the buckyball network, with resulting larger R_g . This suggest that the structure pairing present in the structure can form only when the chains are in close contact. For this to happen, a sampling of many confor-

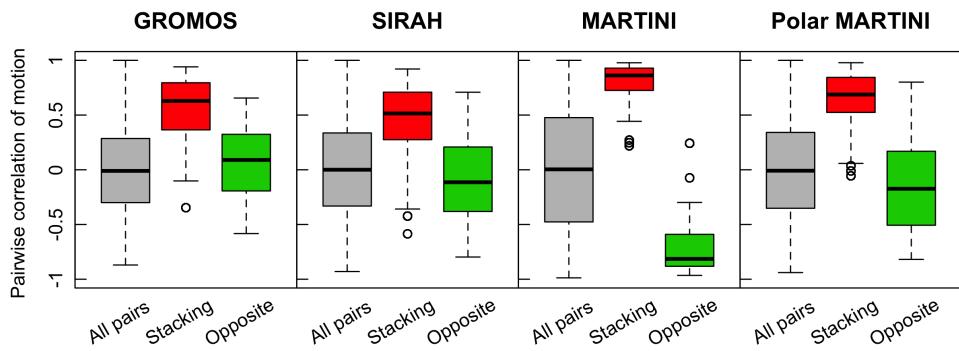


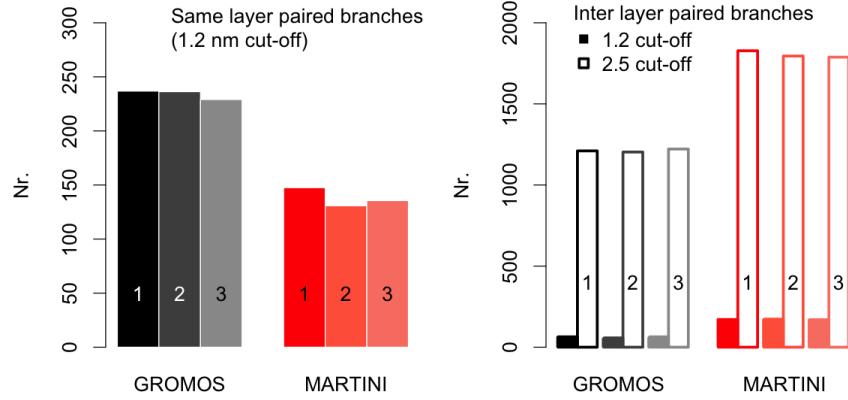
Figure 3.7: Distribution of the correlation of motion between different molecules in the buckyball simulations. Black band: median of the distribution; box: first and third quartiles; whiskers: maximum and minimum, outliers excluded (hollow dots). Results are shown for Replica 1 of each simulation set up.

mations needs to be performed, and this is compatible with the long time of assembly observed experimentally (up to 7-10 hours). The RDF of the protein masses around the buckyball centre of mass shows a Gaussian profile (Figure 3.6(c)). The fact that no matter is observed nearby the origin means that the molecules do not collapse to the center and the central cavity is maintained. To be noticed that the value of RDF at a given radius is already normalised by the volume of the spherical shell at that position, so that a uniformly full object would display a flat distribution. A fit of the RDF to a Gaussian curve returns a mean value of 5.1 nm and a FWHM of 2.2 nm, which gives an estimate of the bilayer thickness. A similar computation is repeated for the inner and outer layer separately, providing 1.1 nm of distance between the two distributions means. This interlayer distance is compatible with the distance between the backbones of stacking β -sheet in structures like densely packed amyloids with cross- β sheet quaternary structure (1.0 nm [?]).

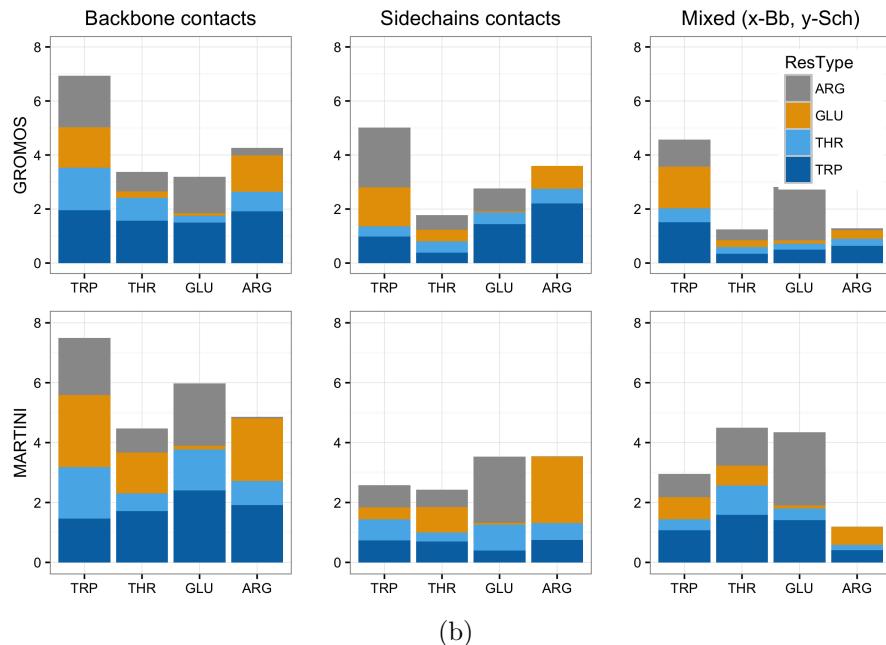
This thickness value thus hints at the fact that the two layers are closely packed. This is confirmed by the analysis of the pairwise correlations of motion: molecules at the same polar coordinates (i.e. stacking radially one on the other) have a positive correlation and so move coherently, while the ones at opposite poles do not show particular correlation, as well the ensemble of all possible pairs (Figure 3.7).

The measure of paired branches in the buckyball network, shown in Fig-

ure 3.8(a), reports an average of 240 pairings between arms belonging to the same layer (summing over inner and outer), and around 100 only for inter-layer ones. The first figure means that each molecule establishes on average



(a)



(b)

Figure 3.8: TAKE OUT 2.5NM INTER CONTACTS?, REPUT BACKMAPPING?, BETTER (a) Number of paired branches within the same layer and between layers, as defined in the main text. Results are shown for Replica 1 of each simulation set up. (b) Contacts with persistence greater than 50% between C_α s (top row), or between side chains (bottom row) and different simulations set up (Replica 1 for each of them). A similar plot for mixed C_α -side chain contact in shown in Figure – of Appendix –.

contacts with four neighbours. This value is larger than the one predicted for a perfect buckyball arrangement (which would be three), likely due to the fact that the structure contracts slightly with respect to the initial size and some branches deviate from the original position, locating themselves close to multiple neighbours. On the contrary, few inter layer contacts are observed within 1.2 nm distance cut off because of the steric hindrance of the side chains which are located in between the two layers, which keep their average position at a distance greater than the cut off chosen.

To look more into detail at the network of interactions between arms of the molecule, we computed the contacts between backbone and/or side chains of amino acid which survive more than 50% of the time of the simulations, grouped by residue type (Figure 3.8(b)) and normalised by the number of molecules present. The number of backbone contacts per molecule is around 3 for Glutamic acid and Threonine, while is between 6 and 7 for Tryptophan residues, proving that on average each residue is well paired with another one, except for Arginines: only two thirds of them are paired on average, likely because of their terminal position. At the same time however the bar plot shows also that there is no rigid arrangement between branches. Indeed, for example, Tryptophan residues are not chiefly paired to Tryptophan ones, as the optimal arrangement would be, suggesting a flexibility in the structure. Nevertheless, the analysis highlights that Trp residues are key to form contacts with the neighbours and (from panel b) that the interaction with Arginine through their side chains (and thus cation- π) is an important element of the structure.

Some of these contacts are mediated by an hydrogen bond interaction: the number of them present during the simulations is computed, and divided by amino acid type. We identify the ones occurring between backbones, side chains, and mixed, in this case as well. It is shown that Tryptophan contributes to a large number of backbone hydrogen bonds (Figure 3.9 top), especially with other Tryoptophan residues, consistently with what found in the analysis of general contacts. Arginin side chains are the most prone to establish H-bonds as a donor with many different amino acid side chains, but especially Glutammic acid (in black in Figure 3.9, bottom) as expected from the facing positions they occupy in the molecules arrangement.

Finally, it is important to understand what residues are exposed at the

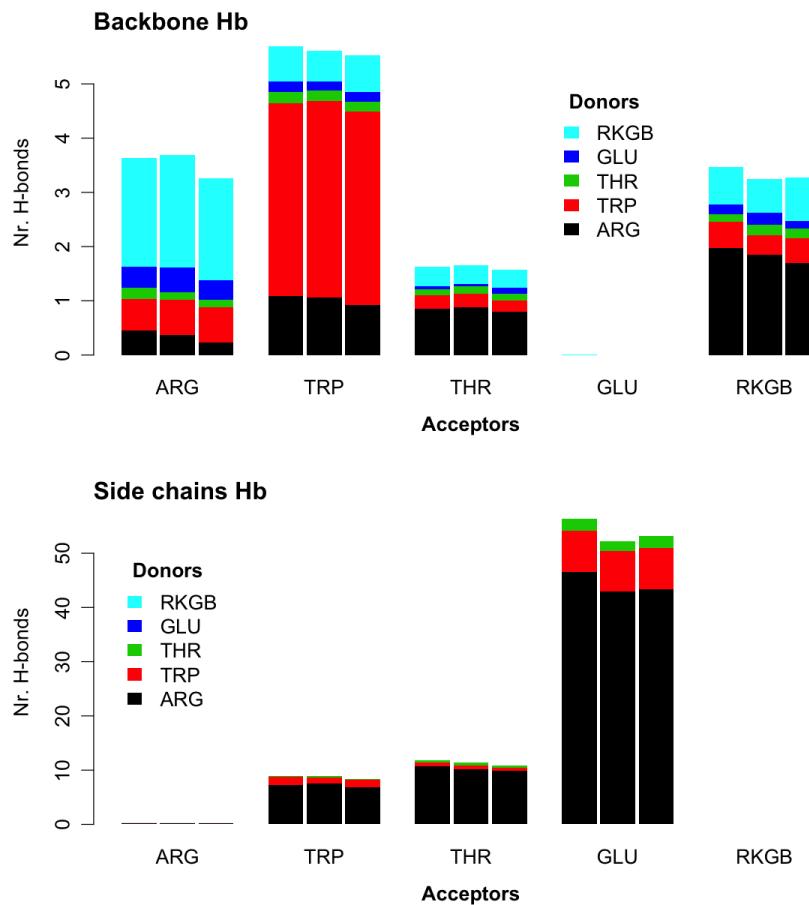


Figure 3.9: [REDO BETTER, put "Nr. Hbonds/molecule" in ylab]
 Average number of hydrogen bonds per molecule occurring between backbone of molecules in a 100 ns atomistic simulation of the buckyball in solution (top) and between its amino acid side chains (bottom). Result shown for Replica 1.

surface of the structure, especially for future applications: in order to make the peptide co-assemble with other products, the two must have a compatible chemical character. To understand what surface the peptide exposes to the solution, we compute the Solvent Accessible Surface Area (SASA) and break it down by amino acid type. Figure 3.10 shows that half of the accessible surface is represented by the charged residues Arginine, while Tryptophan contribute to less than one quarter to it, despite having bulky side chains. For each arm (of sequence RRWTWE), the sum of SASA from charged and polar residues (Arg, Thr and Glu) constitute the 79% of the total SASA of an

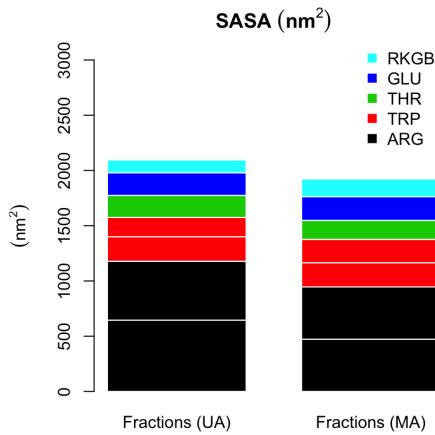


Figure 3.10: [REDO BETTER AND PUT PANEL LETTER] Break down of the total Solvent Accessible Surface Area (SASA) by residue types. Results for Replica 1 of each simulation set up.

arm, while summing their nominal areas (the SASA each residue would have within a Gly-X-Gly tripeptide) it would constitute the 66% of a RRWTWE sequence SASA. The screening of the hydrophobic residues is a consequence of the double layer structure and it is crucial in the perspective of binding mechanisms to membranes, in particular to the anionic bacterial one.

3.4.2 Multiscale comparison of model capsule

We performed a multiscale analysis of the capsule structure simulating it with two different coarse grain force fields with a twofold aim: first we wanted to simulate it for a longer time to observe how its structural properties were maintained on the medium time scale. Second, we believe that proving the stability of the capsule with different description strengthen the evidence that the assembly proposed is indeed a possible, favourable arrangement of the molecules in solution.

The final structure proposed by the SIRAH and MARTINI force fields are slightly different with respect to the atomistic one and among themselves, due to the different nature of the interactions between beads and the longer time scale reached. The SIRAH simulated capsule results in a more open structure, with a skewed and broader radial density profile, while MARTINI with the standard water model provides a more compact configuration (respectively 6.0 nm and 4.6 nm average radius, and 2.9 nm and 1.9 nm average thickness)

(Figure 5). MARTINI runs with a polar water model provide a slightly larger, and more solvated, structure with respect to standard MARTINI (4.8 nm average radius), with comparable thickness.

The correlation of motion between molecules for the SIRAH force field and polar MARTINI simulations reproduce the results of the atomistic one, with the appearance of a slight anticorrelation between molecules at the opposite poles. This is due to the contraction or expansion happening at the beginning at the simulation, when the capsule adjusts to the equilibrium size which depend to some extent from the force field. The correlation effects are much more pronounced in standard MARTINI, which results in a smaller object with the two layers strongly correlated in their motion. This is likely due to the poorer properties of solvation of standard MARTINI water, which cause the protein beads to preferentially interact between themselves rather than with the solvent. Indeed, with polar MARTINI, despite the structure has a similar radius with respect to the standard MARTINI outcome, the motion resembles more the atomistic dynamics, with less strong correlations.

The number of chain pairs in the SIRAH simulations show less contacts, in line with the more expanded structure. Standard MARTINI simulation show remarkably less pairing from the same layer, but higher pairing between layers, consistently with the fact that the inner and outer shell are closer in position.

The MARTINI simulations show a lower number of contacts between molecules belonging to the same layer, probably due to the less specific interactions between different beads which do not allow a description of the hydrogen bond network between amino acids, favouring a close backbone proximity. Moreover, this shows that the structure obtained after a longer time scale is less organised, but it does keep its overall shape thanks to the tight packing between chains. Consistently with this and the smaller thickness of the shell in the MARTINI simulations, the number of contacts between the two layers is higher than in the atomistic simulation.

The fact that in both simulations we observe a contraction with respect to the initial configuration, in which all the molecules have an extended shape, and that only a bilayer grants stability of the assembly, suggests that it is necessary to have a minimum density on the surface of the sphere to grant a robust architecture.

Chemical properties To characterise the chemical determinants that promote the assembly we investigate the interactions between amino acids of different types, first identifying the amino acids that are spatially closed, then, for the atomistic simulations, computing the hydrogen bond network.

We define contacts between backbones if the C α of two residues are closer than a cutoff distance of 0.6 nm; side chains ones if reference atoms in the side chains are closer than the same cut off, and finally mixed if the proximity is between a C α and a side chain reference atom. As reference we took the atom or beads farthest away from the backbone, excluding hydrogens (specifically: CZ/SC2 for Arg, CZ2/SC4 for Trp, OG1/SC1 for Thr and CDSC1 for Glu).

Coarse grain simulations give slightly higher figures for the backbone contacts, compatible with a more compact structure. The results for side chains contacts are quite different with respect to the atomistic ones, likely due to the coarse grain model which does not reproduce faithfully the ratio between steric hindrances of the different side chains.

Chapter 4

Lipid parametrisation

Appendices

A.1 On the derivation of the GP predictive distribution

This appendix gives a sketch of the procedure by which Eq. (??) is obtained, which substantially relies on the properties of multivariate Gaussian distributions. For full details on this one can consult the excellent Refs. [?] and [?].

Bibliography

- [1] Mettler Toledo, “Polymerization reactions.”
- [2] Wikipedia, “Liposome.”
- [3] L. Schoonen and J. C. M. van Hest, “Functionalization of protein-based nanocages for drug delivery applications,” *Nanoscale*, vol. 6, pp. 7124–7141, jun 2014.
- [4] J. M. A. Blair, M. A. Webber, A. J. Baylay, D. O. Ogbolu, and L. J. V. Piddock, “Molecular mechanisms of antibiotic resistance,” *Nature Publishing Group*, vol. 13, 2014.
- [5] J. Kim, “Phage as a therapeutic agent.”
- [6] M. D. Torres, S. Sothiselvam, T. K. Lu, and C. de la Fuente-Nunez, “Peptide Design Principles for Antimicrobial Applications,” *Journal of Molecular Biology*, jan 2019.
- [7] L. T. Nguyen, E. F. Haney, and H. J. Vogel, “The expanding scope of antimicrobial peptide structures and their modes of action,” *Trends in Biotechnology*, vol. 29, pp. 464–472, sep 2011.
- [8] V. Castelletto, E. de Santis, H. Alkassem, B. Lamarre, J. E. Noble, S. Ray, A. Bella, J. R. Burns, B. W. Hoogenboom, M. G. Ryadnov, J. Vandesompele, and C. T. Wittwer, “Structurally plastic peptide capsules for synthetic antimicrobial viruses,” *Chem. Sci.*, vol. 7, pp. 1707–1711, feb 2016.
- [9] R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T. I. Oprea, and J. P. Overing-

- ton, “A comprehensive map of molecular drug targets,” *Nature Reviews Drug Discovery*, vol. 16, pp. 19–34, jan 2017.
- [10] E. Gordon, N. Mouz, E. Duée, and O. Dideberg, “The crystal structure of the penicillin-binding protein 2x from *Streptococcus pneumoniae* and its acyl-enzyme form: implication in drug resistance,” *Journal of Molecular Biology*, vol. 299, pp. 477–485, jun 2000.
- [11] A. H. Delcour, “Outer membrane permeability and antibiotic resistance,” *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1794, pp. 808–816, may 2009.
- [12] A. V. Vargiu and H. Nikaido, “Multidrug binding properties of the AcrB efflux pump characterized by molecular dynamics simulations.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 20637–42, dec 2012.
- [13] S. Kojima and H. Nikaido, “Permeation rates of penicillins indicate that *Escherichia coli* porins function principally as nonspecific channels,” *Proceedings of the National Academy of Sciences*, vol. 110, pp. E2629–E2634, jul 2013.
- [14] J.-P. Lavigne, A. Sotto, M.-H. Nicolas-Chanoine, N. Bouziges, J.-M. Pagès, and A. Davin-Regli, “An adaptive response of *Enterobacter aerogenes* to imipenem: regulation of porin balance in clinical isolates,” *International Journal of Antimicrobial Agents*, vol. 41, pp. 130–136, feb 2013.
- [15] A. Poulopoulos, E. Voulgari, G. Vrioni, V. Koumaki, G. Xidopoulos, V. Chatzipantazi, F. Markou, and A. Tsakris, “Outbreak Caused by an Ertapenem-Resistant, CTX-M-15-Producing *Klebsiella pneumoniae* Sequence Type 101 Clone Carrying an OmpK36 Porin Variant,” *Journal of Clinical Microbiology*, vol. 51, pp. 3176–3182, oct 2013.
- [16] J. L. Floyd, K. P. Smith, S. H. Kumar, J. T. Floyd, and M. F. Varela, “LmrS Is a Multidrug Efflux Pump of the Major Facilitator Superfamily from *Staphylococcus aureus*,” *Antimicrobial Agents and Chemotherapy*, vol. 54, pp. 5406–5412, dec 2010.

- [17] W. Ogawa, M. Onishi, R. Ni, T. Tsuchiya, and T. Kuroda, “Functional study of the novel multidrug efflux pump KexD from *Klebsiella pneumoniae*,” *Gene*, vol. 498, pp. 177–182, may 2012.
- [18] M. Dolejska, L. Villa, L. Poirel, P. Nordmann, and A. Carattoli, “Complete sequencing of an IncHI1 plasmid encoding the carbapenemase NDM-1, the ArmA 16S RNA methylase and a resistance-nodulation-cell division/multidrug efflux pump,” *Journal of Antimicrobial Chemotherapy*, vol. 68, pp. 34–39, jan 2013.
- [19] Y. M. Abouzeed, S. Baucheron, and A. Cloeckaert, “ramR mutations involved in efflux-mediated multidrug resistance in *Salmonella enterica* serovar Typhimurium.,” *Antimicrobial agents and chemotherapy*, vol. 52, pp. 2428–34, jul 2008.
- [20] E. Nikaido, I. Shirosaka, A. Yamaguchi, and K. Nishino, “Regulation of the AcrAB multidrug efflux pump in *Salmonella enterica* serovar Typhimurium in response to indole and paraquat,” *Microbiology*, vol. 157, pp. 648–655, mar 2011.
- [21] A. C. Shore, E. C. Deasy, P. Slickers, G. Brennan, B. O’Connell, S. Monnecke, R. Ehricht, and D. C. Coleman, “Detection of Staphylococcal Cassette Chromosome mec Type XI Carrying Highly Divergent *mecA*, *mecI*, *mecR1*, *blaZ*, and *ccr* Genes in Human Clinical Isolates of Clonal Complex 130 Methicillin-Resistant *Staphylococcus aureus*,” *Antimicrobial Agents and Chemotherapy*, vol. 55, pp. 3765–3773, aug 2011.
- [22] D. S. Billal, J. Feng, P. Leprohon, D. Légaré, and M. Ouellette, “Whole genome analysis of linezolid resistance in *Streptococcus pneumoniae* reveals resistance and compensatory mutations,” *BMC Genomics*, vol. 12, p. 512, dec 2011.
- [23] K. S. Long, J. Poehlsgaard, C. Kehrenberg, S. Schwarz, and B. Vester, “The Cfr rRNA Methyltransferase Confers Resistance to Phenicols, Lincomosamides, Oxazolidinones, Pleuromutilins, and Streptogramin A Antibiotics,” *Antimicrobial Agents and Chemotherapy*, vol. 50, pp. 2500–2505, jul 2006.

- [24] E. P. Abraham and E. Chain, “An enzyme from bacteria able to destroy penicillin. 1940.,” *Reviews of infectious diseases*, vol. 10, no. 4, pp. 677–8, 1988.
- [25] D. Livermore, “Defining an extended-spectrum β -lactamase,” *Clinical Microbiology and Infection*, vol. 14, pp. 3–10, jan 2008.
- [26] P. Nordmann, L. Poirel, T. R. Walsh, and D. M. Livermore, “The emerging NDM carbapenemases,” *Trends in Microbiology*, vol. 19, pp. 588–595, dec 2011.
- [27] N. Woodford and A. P. Johnson, “Global spread of antibiotic resistance: the example of New Delhi metallo- β -lactamase (NDM)-mediated carbapenem resistance,” *Journal of Medical Microbiology*, vol. 62, pp. 499–513, apr 2013.
- [28] G. Wright, “Bacterial resistance to antibiotics: Enzymatic degradation and modification,” *Advanced Drug Delivery Reviews*, vol. 57, pp. 1451–1470, jul 2005.
- [29] G. Kapoor, S. Saigal, and A. Elongavan, “Action and resistance mechanisms of antibiotics: A guide for clinicians.,” *Journal of anaesthesia, clinical pharmacology*, vol. 33, no. 3, pp. 300–305, 2017.
- [30] A. C. Birkegård, T. Halasa, N. Toft, A. Folkesson, and K. Græsbøll, “Send more data: a systematic review of mathematical models of antimicrobial resistance.,” *Antimicrobial resistance and infection control*, vol. 7, p. 117, 2018.
- [31] A. M. Niewiadomska, B. Jayabalasingham, J. C. Seidman, L. Willem, B. Grenfell, D. Spiro, and C. Viboud, “Population-level mathematical modeling of antimicrobial resistance: a systematic review,” *BMC Medicine*, vol. 17, p. 81, dec 2019.
- [32] J. O 'neill, “TACKLING DRUG-RESISTANT INFECTIONS GLOBALLY: FINAL REPORT AND RECOMMENDATIONS THE REVIEW ON ANTIMICROBIAL RESISTANCE,” 2016.
- [33] P. Mantravadi, K. Kalesh, R. Dobson, A. Hudson, and A. Parthasarathy, “The Quest for Novel Antimicrobial Compounds: Emerging Trends in

- Research, Development, and Technologies,” *Antibiotics*, vol. 8, p. 8, jan 2019.
- [34] A. Ebbensgaard, H. Mordhorst, M. T. Overgaard, C. G. Nielsen, F. M. Aarestrup, and E. B. Hansen, “Comparative Evaluation of the Antimicrobial Activity of Different Antimicrobial Peptides against a Range of Pathogenic Bacteria,” *PLOS ONE*, vol. 10, p. e0144611, dec 2015.
 - [35] K. A. Brogden, “Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?,” *Nature Reviews Microbiology*, vol. 3, pp. 238–250, mar 2005.
 - [36] R. E. W. Hancock and H.-G. Sahl, “Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies,” *Nature Biotechnology*, vol. 24, pp. 1551–1557, dec 2006.
 - [37] T.-Y. Lin and D. B. Weibel, “Organization and function of anionic phospholipids in bacteria,” *Applied Microbiology and Biotechnology*, vol. 100, pp. 4255–4267, may 2016.
 - [38] E. Glukhov, M. Stark, L. L. Burrows, and C. M. Deber, “Basis for selectivity of cationic antimicrobial peptides for bacterial versus mammalian membranes.,” *The Journal of biological chemistry*, vol. 280, pp. 33960–7, oct 2005.
 - [39] A. A. Spector and M. A. Yorek, “Membrane lipid composition and cellular function.,” *Journal of lipid research*, vol. 26, pp. 1015–35, sep 1985.
 - [40] 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, pp. 112–24, feb 2008.
 - [41] M. R. Yeaman and N. Y. Yount, “Mechanisms of Antimicrobial Peptide Action and Resistance,” *Pharmacological Reviews*, vol. 55, pp. 27–55, mar 2003.
 - [42] Y. Lai and R. L. Gallo, “AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense,” *Trends in Immunology*, vol. 30, pp. 131–141, mar 2009.

- [43] M. Zasloff, “Antimicrobial peptides of multicellular organisms,” *Nature*, vol. 415, pp. 389–395, jan 2002.
- [44] K. Matsuzaki, “Control of cell selectivity of antimicrobial peptides,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1788, pp. 1687–1692, aug 2009.
- [45] T. Ebenhan, O. Gheysens, H. G. Kruger, J. R. Zeevaart, and M. M. Sathekge, “Antimicrobial peptides: their role as infection-selective tracers for molecular imaging.,” *BioMed research international*, vol. 2014, p. 867381, aug 2014.
- [46] D. I. Chan, E. J. Prenner, and H. J. Vogel, “Tryptophan- and arginine-rich antimicrobial peptides: Structures and mechanisms of action,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1758, pp. 1184–1202, sep 2006.
- [47] L. Yang, T. A. Harroun, T. M. Weiss, L. Ding, and H. W. Huang, “Barrel-stave model or toroidal model? A case study on melittin pores.,” *Biophysical journal*, vol. 81, pp. 1475–85, sep 2001.
- [48] A. Spaar, C. Münster, and T. Salditt, “Conformation of peptides in lipid membranes studied by x-ray grazing incidence scattering.,” *Biophysical journal*, vol. 87, pp. 396–407, jul 2004.
- [49] K. Matsuzaki, O. Murase, N. Fujii, and K. Miyajima, “An Antimicrobial Peptide, Magainin 2, Induced Rapid Flip-Flop of Phospholipids Coupled with Pore Formation and Peptide Translocation,” *Biochemistry*, vol. 35, pp. 11361–11368, jan 1996.
- [50] K. J. Hallock, D.-K. Lee, and A. Ramamoorthy, “MSI-78, an analogue of the magainin antimicrobial peptides, disrupts lipid bilayer structure via positive curvature strain.,” *Biophysical journal*, vol. 84, pp. 3052–60, may 2003.
- [51] K. Matsuzaki, K.-i. Sugishita, M. Harada, N. Fujii, and K. Miyajima, “Interactions of an antimicrobial peptide, magainin 2, with outer and inner membranes of Gram-negative bacteria,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1327, pp. 119–130, jul 1997.

- [52] A. S. Ladokhin and S. H. White, “‘Detergent-like’ permeabilization of anionic lipid vesicles by melittin,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1514, pp. 253–260, oct 2001.
 - [53] S. Yamaguchi, D. Huster, A. Waring, R. I. Lehrer, W. Kearney, B. F. Tack, and M. Hong, “Orientation and Dynamics of an Antimicrobial Peptide in the Lipid Bilayer by Solid-State NMR Spectroscopy,” *Bioophysical Journal*, vol. 81, pp. 2203–2214, oct 2001.
 - [54] M. Tang and M. Hong, “Structure and mechanism of β -hairpin antimicrobial peptides in lipid bilayers from solid-state NMR spectroscopy,” *Molecular BioSystems*, vol. 5, p. 317, mar 2009.
 - [55] R. I. Lehrer, “Primate defensins,” *Nature Reviews Microbiology*, vol. 2, pp. 727–738, sep 2004.
 - [56] B. L. Kagan, M. E. Selsted, T. Ganz, and R. I. Lehrer, “Antimicrobial defensin peptides form voltage-dependent ion-permeable channels in planar lipid bilayer membranes.,” *Proceedings of the National Academy of Sciences*, vol. 87, pp. 210–214, jan 1990.
 - [57] K. Takeuchi, H. Takahashi, M. Sugai, H. Iwai, T. Kohno, K. Sekimizu, S. Natori, and I. Shimada, “Channel-forming membrane permeabilization by an antibacterial protein, sapecin: determination of membrane-buried and oligomerization surfaces by NMR.,” *The Journal of biological chemistry*, vol. 279, pp. 4981–7, feb 2004.
 - [58] A. Yonezawa, J. Kuwahara, N. Fujii, and Y. Sugiura, “Binding of tachypleasin I to DNA revealed by footprinting analysis: significant contribution of secondary structure to DNA binding and implication for biological action,” *Biochemistry*, vol. 31, pp. 2998–3004, mar 1992.
 - [59] J. L. Gifford, H. N. Hunter, and H. J. Vogel, “Lactoferricin,” *Cellular and Molecular Life Sciences*, vol. 62, pp. 2588–2598, nov 2005.
 - [60] M. Tomita, M. Takase, W. Bellamy, and S. Shimamura, “A review: the active peptide of lactoferrin.,” *Acta paediatrica Japonica : Overseas edition*, vol. 36, pp. 585–91, oct 1994.
-

- [61] D. J. Schibli, P. M. Hwang, and H. J. Vogel, “The structure of the antimicrobial active center of lactoferricin B bound to sodium dodecyl sulfate micelles,” *FEBS Letters*, vol. 446, pp. 213–217, mar 1999.
- [62] A. Peschel and H.-G. Sahl, “The co-evolution of host cationic antimicrobial peptides and microbial resistance,” *Nature Reviews Microbiology*, vol. 4, pp. 529–536, jul 2006.
- [63] M. Sieprawska-Lupa, P. Mydel, K. Krawczyk, K. Wójcik, M. Puklo, B. Lupa, P. Suder, J. Silberring, M. Reed, J. Pohl, W. Shafer, F. McAleese, T. Foster, J. Travis, and J. Potempa, “Degradation of human antimicrobial peptide LL-37 by *Staphylococcus aureus*-derived proteinases.,” *Antimicrobial agents and chemotherapy*, vol. 48, pp. 4673–9, dec 2004.
- [64] D. C. Nelson, J. Garbe, and M. Collin, “Cysteine proteinase SpeB from *Streptococcus pyogenes* - a potent modifier of immunologically important host and bacterial proteins,” *Biological Chemistry*, vol. 392, pp. 1077–88, jan 2011.
- [65] J. W. Costerton, P. S. Stewart, and E. P. Greenberg, “Bacterial biofilms: a common cause of persistent infections.,” *Science (New York, N.Y.)*, vol. 284, pp. 1318–22, may 1999.
- [66] A. Jolivet-Gougeon and M. Bonnaure-Mallet, “Biofilms as a mechanism of bacterial resistance,” *Drug Discovery Today: Technologies*, vol. 11, pp. 49–56, mar 2014.
- [67] C. Vuong, J. M. Voyich, E. R. Fischer, K. R. Braughton, A. R. Whitney, F. R. DeLeo, and M. Otto, “Polysaccharide intercellular adhesin (PIA) protects *Staphylococcus epidermidis* against major components of the human innate immune system.,” *Cellular microbiology*, vol. 6, pp. 269–75, mar 2004.
- [68] C. Vuong, S. Kocanova, J. M. Voyich, Y. Yao, E. R. Fischer, F. R. DeLeo, and M. Otto, “A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence.,” *The Journal of biological chemistry*, vol. 279, pp. 54881–6, dec 2004.

- [69] N. Strempel, J. Strehmel, and J. Overhage, “Potential Application of Antimicrobial Peptides in the Treatment of Bacterial Biofilm Infections,” *Current Pharmaceutical Design*, vol. 21, pp. 67–84, nov 2014.
- [70] R. Saar-Dover, A. Bitler, R. Nezer, L. Shmuel-Galia, A. Firon, E. Shimoni, P. Trieu-Cuot, and Y. Shai, “D-Alanylation of Lipoteichoic Acids Confers Resistance to Cationic Peptides in Group B Streptococcus by Increasing the Cell Wall Density,” *PLoS Pathogens*, vol. 8, p. e1002891, sep 2012.
- [71] T. D. H. Bugg, G. D. Wright, S. Dutka-Malen, M. Arthur, P. Courvalin, and C. T. Walsh, “Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA,” *Biochemistry*, vol. 30, pp. 10408–10415, oct 1991.
- [72] S. M. Moskowitz, R. K. Ernst, and S. I. Miller, “PmrAB, a two-component regulatory system of *Pseudomonas aeruginosa* that modulates resistance to cationic antimicrobial peptides and addition of aminoarabinose to lipid A.,” *Journal of bacteriology*, vol. 186, pp. 575–9, jan 2004.
- [73] X. Wang, S. C. McGrath, R. J. Cotter, and C. R. H. Raetz, “Expression cloning and periplasmic orientation of the *Francisella novicida* lipid A 4'-phosphatase LpxF.,” *The Journal of biological chemistry*, vol. 281, pp. 9321–30, apr 2006.
- [74] L. Guo, K. B. Lim, C. M. Poduje, M. Daniel, J. S. Gunn, M. Hackett, and S. I. Miller, “Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides.,” *Cell*, vol. 95, pp. 189–98, oct 1998.
- [75] A. Peschel, R. W. Jack, M. Otto, L. V. Collins, P. Staibitz, G. Nicholson, H. Kalbacher, W. F. Nieuwenhuizen, G. Jung, A. Tarkowski, K. P. van Kessel, and J. A. van Strijp, “*Staphylococcus aureus* Resistance to Human Defensins and Evasion of Neutrophil Killing via the Novel Virulence Factor Mprf Is Based on Modification of Membrane Lipids with l-Lysine,” *The Journal of Experimental Medicine*, vol. 193, pp. 1067–1076, may 2001.

- [76] R. Kumariya, S. K. Sood, Y. S. Rajput, N. Saini, and A. K. Garsa, “Increased membrane surface positive charge and altered membrane fluidity leads to cationic antimicrobial peptide resistance in *Enterococcus faecalis.*,” *Biochimica et biophysica acta*, vol. 1848, pp. 1367–75, jun 2015.
- [77] V. Band and D. Weiss, “Mechanisms of Antimicrobial Peptide Resistance in Gram-Negative Bacteria,” *Antibiotics*, vol. 4, pp. 18–41, dec 2014.
- [78] G. Wang, X. Li, Z. Wang, L. J.Z., C. K.Y., L. M.P., P. M., V. K.A., S. J.M., and van Hoek M.L., “APD3: the antimicrobial peptide database as a tool for research and education,” *Nucleic Acids Research*, vol. 44, pp. D1087–D1093, jan 2016.
- [79] M. Pirtskhalava, A. Gabrielian, P. Cruz, H. L. Griggs, R. B. Squires, D. E. Hurt, M. Grigolava, M. Chubinidze, G. Gogoladze, B. Vishnepolsky, V. Alekseyev, A. Rosenthal, and M. Tartakovsky, “DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides,” *Nucleic Acids Research*, vol. 44, pp. 6503–6503, jul 2016.
- [80] J.-H. Jhong, Y.-H. Chi, W.-C. Li, T.-H. Lin, K.-Y. Huang, and T.-Y. Lee, “dbAMP: an integrated resource for exploring antimicrobial peptides with functional activities and physicochemical properties on transcriptome and proteome data,” *Nucleic Acids Research*, vol. 47, pp. D285–D297, jan 2019.
- [81] S. Lata, N. K. Mishra, and G. P. Raghava, “AntiBP2: improved version of antibacterial peptide prediction,” *BMC Bioinformatics*, vol. 11, p. S19, jan 2010.
- [82] P. Bhadra, J. Yan, J. Li, S. Fong, and S. W. I. Siu, “AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest,” *Scientific Reports*, vol. 8, p. 1697, dec 2018.
- [83] C. D. Fjell, J. A. Hiss, R. E. W. Hancock, and G. Schneider, “Designing antimicrobial peptides: form follows function,” *Nature Reviews Drug Discovery*, vol. 11, p. 37, dec 2011.

- [84] J. Wang, S. Chou, L. Xu, X. Zhu, N. Dong, A. Shan, and Z. Chen, “High specific selectivity and Membrane-Active Mechanism of the synthetic centrosymmetric α -helical peptides with Gly-Gly pairs.,” *Scientific reports*, vol. 5, p. 15963, nov 2015.
- [85] K. Hilpert, R. Volkmer-Engert, T. Walter, and R. E. W. Hancock, “High-throughput generation of small antibacterial peptides with improved activity,” *Nature Biotechnology*, vol. 23, pp. 1008–1012, aug 2005.
- [86] D. Migoń, M. Jaśkiewicz, D. Neubauer, M. Bauer, E. Sikorska, E. Kamysz, and W. Kamysz, “Alanine Scanning Studies of the Antimicrobial Peptide Aurein 1.2,” *Probiotics and Antimicrobial Proteins*, pp. 1–13, dec 2018.
- [87] S. Liu, J. Bao, X. Lao, and H. Zheng, “Novel 3D Structure Based Model for Activity Prediction and Design of Antimicrobial Peptides,” *Scientific Reports*, vol. 8, p. 11189, dec 2018.
- [88] Z. Jiang, A. I. Vasil, L. Gera, M. L. Vasil, and R. S. Hodges, “Rational Design of α -Helical Antimicrobial Peptides to Target Gram-negative Pathogens, *Acinetobacter baumannii* and *Pseudomonas aeruginosa*: Utilization of Charge, ‘Specificity Determinants’, Total Hydrophobicity, Hydrophobe Type and Location as Design Parameters to Improve the Therapeutic Ratio,” *Chemical Biology & Drug Design*, vol. 77, pp. 225–240, apr 2011.
- [89] B. Deslouches, S. M. Phadke, V. Lazarevic, M. Cascio, K. Islam, R. C. Montelaro, and T. A. Mietzner, “De Novo Generation of Cationic Antimicrobial Peptides: Influence of Length and Tryptophan Substitution on Antimicrobial Activity,” *Antimicrobial Agents and Chemotherapy*, vol. 49, pp. 316–322, jan 2005.
- [90] C. Loose, K. Jensen, I. Rigoutsos, and G. Stephanopoulos, “A linguistic model for the rational design of antimicrobial peptides,” *Nature*, vol. 443, pp. 867–869, oct 2006.
- [91] A. Farrotti, P. Conflitti, S. Srivastava, J. Ghosh, A. Palleschi, L. Stella, G. Bocchinfuso, A. Farrotti, P. Conflitti, S. Srivastava, J. K. Ghosh,

- A. Palleschi, L. Stella, and G. Bocchinfuso, “Molecular Dynamics Simulations of the Host Defense Peptide Temporin L and Its Q3K Derivative: An Atomic Level View from Aggregation in Water to Bilayer Perturbation,” *Molecules*, vol. 22, p. 1235, jul 2017.
- [92] A. Cherkasov, K. Hilpert, H. Jenssen, C. D. Fjell, M. Waldbrook, S. C. Mullaly, R. Volkmer, and R. E. Hancock, “Use of Artificial Intelligence in the Design of Small Peptide Antibiotics Effective against a Broad Spectrum of Highly Antibiotic-Resistant Superbugs,” *ACS Chemical Biology*, vol. 4, pp. 65–74, jan 2009.
- [93] M. A. B. Naafs, “The Antimicrobial Peptides: Ready for Clinical Trials?,” *Biomedical Journal of Scientific & Technical Research*, vol. 7, pp. 001–005, aug 2018.
- [94] P. Wipf, J. Xiao, and C. R. J. Stephenson, “Peptide-Like Molecules (PLMs): A Journey from Peptide Bond Isosteres to Gramicidin S Mimetics and Mitochondrial Targeting Agents,” *CHIMIA International Journal for Chemistry*, vol. 63, pp. 764–775, nov 2009.
- [95] X. M. Anguela and K. A. High, “Entering the Modern Era of Gene Therapy,” *Annual Review of Medicine*, vol. 70, pp. 273–288, jan 2019.
- [96] R. Barrangou, “The roles of CRISPRCas systems in adaptive immunity and beyond,” *Current Opinion in Immunology*, vol. 32, pp. 36–41, feb 2015.
- [97] F. Zhang, Y. Wen, and X. Guo, “CRISPR/Cas9 for genome editing: progress, implications and challenges,” *Human Molecular Genetics*, vol. 23, pp. R40–R46, sep 2014.
- [98] P. D. Hsu, E. S. Lander, and F. Zhang, “Development and applications of CRISPR-Cas9 for genome engineering.,” *Cell*, vol. 157, pp. 1262–78, jun 2014.
- [99] E. H. Oldfield, Z. Ram, K. W. Culver, R. M. Blaese, H. L. DeVroom, and W. F. Anderson, “Gene Therapy for the Treatment of Brain Tumors Using Intra-Tumoral Transduction with the Thymidine Kinase Gene and Intravenous Ganciclovir. National Institutes of Health,” *Human Gene Therapy*, vol. 4, pp. 39–69, feb 1993.

- [100] S. E. Lawler, M.-C. Speranza, C.-F. Cho, and E. A. Chiocca, “Oncolytic Viruses in Cancer Treatment,” *JAMA Oncology*, vol. 3, p. 841, jun 2017.
- [101] L. Naldini, “Ex vivo gene transfer and correction for cell-based therapies,” *Nature Reviews Genetics*, vol. 12, pp. 301–315, may 2011.
- [102] F. Mingozzi and K. A. High, “Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges,” *Nature Reviews Genetics*, vol. 12, pp. 341–355, may 2011.
- [103] L. Sánchez, M. Calvo, and J. H. Brock, “Biological role of lactoferrin.,” *Archives of disease in childhood*, vol. 67, pp. 657–61, may 1992.
- [104] R. Arnold, M. Cole, and J. R. McGhee, “A bactericidal effect for human lactoferrin,” *Science*, vol. 197, pp. 263–265, jul 1977.
- [105] R. R. Arnold, M. Brewer, and J. J. Gauthier, “Bactericidal activity of human lactoferrin: sensitivity of a variety of microorganisms.,” *Infection and immunity*, vol. 28, pp. 893–8, jun 1980.
- [106] C. H. Kirkpatrick, I. Green, R. R. Rich, and A. L. Schade, “Inhibition of Growth of *Candida albicans* by Iron-Unsaturated Lactoferrin: Relation to Host-Defense Mechanisms in Chronic Mucocutaneous Candidiasis,” *Journal of Infectious Diseases*, vol. 124, pp. 539–544, dec 1971.
- [107] S. Jahani, A. Shakiba, and L. Jahani, “The Antimicrobial Effect of Lactoferrin on Gram-Negative and Gram-Positive Bacteria,” *International Journal of Infection*, vol. 2, jul 2015.
- [108] S. Farnaud and R. W. Evans, “Lactoferrin - a multifunctional protein with antimicrobial properties.,” *Molecular immunology*, vol. 40, pp. 395–405, nov 2003.
- [109] H. Shau, A. Kim, and S. H. Golub, “Modulation of natural killer and lymphokine-activated killer cell cytotoxicity by lactoferrin,” *Journal of Leukocyte Biology*, vol. 51, pp. 343–349, apr 1992.
- [110] M. Gahr, C. P. Speer, B. Damerau, and G. Sawatzki, “Influence of Lactoferrin on the Function of Human Polymorphonuclear Leukocytes and Monocytes,” *Journal of Leukocyte Biology*, vol. 49, pp. 427–433, may 1991.

- [111] W. Bellamy, M. Takase, K. Yamauchi, H. Wakabayashi, K. Kawase, and M. Tomita, “Identification of the bactericidal domain of lactoferrin,” *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, vol. 1121, pp. 130–136, may 1992.
- [112] H. Wakabayashi, S. Abe, T. Okutomi, S. Tansho, K. Kawase, and H. Yamaguchi, “Cooperative anti-Candida effects of lactoferrin or its peptides in combination with azole antifungal agents.,” *Microbiology and immunology*, vol. 40, no. 11, pp. 821–5, 1996.
- [113] O. Aguilera, H. Ostolaza, L. Quirós, and J. Fierro, “Permeabilizing action of an antimicrobial lactoferricin-derived peptide on bacterial and artificial membranes,” *FEBS Letters*, vol. 462, pp. 273–277, dec 1999.
- [114] A. G. Cochran, N. J. Skelton, and M. A. Starovasnik, “Tryptophan zippers: Stable, monomeric β -hairpins,” *Proceedings of the National Academy of Sciences*, vol. 98, pp. 5578–5583, may 2001.
- [115] P. M. Hwang, N. Zhou, X. Shan, C. H. Arrowsmith, and H. J. Vogel, “Three-Dimensional Solution Structure of Lactoferricin B, an Antimicrobial Peptide Derived from Bovine Lactoferrin,” *Biochemistry*, vol. 37, pp. 4288–4298, mar 1998.
- [116] L. T. Nguyen, D. J. Schibli, and H. J. Vogel, “Structural studies and model membrane interactions of two peptides derived from bovine lactoferricin,” *Journal of Peptide Science*, vol. 11, pp. 379–389, jun 2005.
- [117] A. Tsutsumi, N. Javkhlanlugs, A. Kira, M. Umeyama, I. Kawamura, K. Nishimura, K. Ueda, and A. Naito, “Structure and orientation of bovine lactoferrampin in the mimetic bacterial membrane as revealed by solid-state NMR and molecular dynamics simulation.,” *Biophysical journal*, vol. 103, pp. 1735–43, oct 2012.
- [118] M. Arseneault, S. Bédard, M. Boulet-Audet, and M. Pézolet, “Study of the Interaction of Lactoferricin B with Phospholipid Monolayers and Bilayers,” *Langmuir*, vol. 26, pp. 3468–3478, mar 2010.
- [119] M. B. Strøm, B. E. Haug, O. Rekdal, M. L. Skar, W. Stensen, and J. S. Svendsen, “Important structural features of 15-residue lactoferricin

- derivatives and methods for improvement of antimicrobial activity,” *Biochemistry and cell biology = Biochimie et biologie cellulaire*, vol. 80, no. 1, pp. 65–74, 2002.
- [120] L. Crombez, G. Aldrian-Herrada, K. Konate, Q. N. Nguyen, G. K. McMaster, R. Brasseur, F. Heitz, and G. Divita, “A New Potent Secondary Amphipathic Cellpenetrating Peptide for siRNA Delivery Into Mammalian Cells,” *Molecular Therapy*, vol. 17, pp. 95–103, jan 2009.
- [121] 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, pp. 227–249, may 1976.
- [122] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu, “Biomolecular Modeling: Goals, Problems, Perspectives,” *Angewandte Chemie International Edition*, vol. 45, pp. 4064–4092, jun 2006.
- [123] V. I. Arnold, *Mathematical Methods of Classical Mechanics*. Springer New York, 1989.
- [124] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, “LINCS: A linear constraint solver for molecular simulations,” *Journal of Computational Chemistry*, vol. 18, pp. 1463–1472, sep 1997.
- [125] S. Miyamoto and P. A. Kollman, “Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models,” *Journal of Computational Chemistry*, vol. 13, pp. 952–962, oct 1992.
- [126] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *The Journal of Chemical Physics*, vol. 126, p. 014101, jan 2007.
- [127] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, “Molecular dynamics with coupling to an external bath,” *The Journal of Chemical Physics*, vol. 81, pp. 3684–3690, oct 1984.

- [128] 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.
- [129] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, “A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6,” *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1656–1676, 2004.
- [130] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. van Gunsteren, “Definition and testing of the GROMOS force-field versions 54A7 and 54B7,” *European Biophysics Journal*, vol. 40, pp. 843–856, jul 2011.
- [131] H. Berendsen, D. van der Spoel, and R. van Drunen, “GROMACS: A message-passing parallel molecular dynamics implementation,” *Computer Physics Communications*, vol. 91, pp. 43–56, sep 1995.
- [132] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers,” *SoftwareX*, vol. 1-2, pp. 19–25, sep 2015.
- [133] M. J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS User Manual version 2016*. 2018.
- [134] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, “All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins ,” *The Journal of Physical Chemistry B*, vol. 102, pp. 3586–3616, apr 1998.
- [135] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O’Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, and R. W. Pas-

- tor, “Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types,” *The Journal of Physical Chemistry B*, vol. 114, pp. 7830–7843, jun 2010.
- [136] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, “ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB,” *Journal of Chemical Theory and Computation*, vol. 11, pp. 3696–3713, aug 2015.
- [137] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, “A smooth particle mesh Ewald method,” *The Journal of Chemical Physics*, vol. 103, pp. 8577–8593, nov 1995.
- [138] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren, “A generalized reaction field method for molecular dynamics simulations,” *The Journal of Chemical Physics*, vol. 102, pp. 5451–5459, apr 1995.
- [139] H. Berendsen, J. Postma, W. Van Gunsteren, and J. Hermans, “Interaction models for water in relation to protein hydration.,” in *Intermolecular Forces* (B. Pullman, ed.), pp. 331–342, Dordrecht: D. Reidel Publishing Company, 1981.
- [140] M. R. Machado, E. E. B. Guisasola, F. Klein, M. Sóñora, S. Silva, and S. Pantano, “The SIRAH force field 2.0: Altius, Fortius, Citius,” *bioRxiv*, p. 436774, oct 2018.
- [141] E. E. Barrera, M. R. Machado, and S. Pantano, “Fat SIRAH: Coarse-grained phospholipids to explore membrane-protein dynamics,” *bioRxiv*, p. 627570, may 2019.
- [142] 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,” *Journal of Physical Chemistry B*, vol. 111, no. 27, pp. 7812–7824, 2007.
- [143] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink, “The MARTINI Coarse-Grained Force Field: Extension to Proteins.,” *Journal of chemical theory and computation*, vol. 4, pp. 819–34, may 2008.

- [144] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, and S. J. Marrink, “Improved Parameters for the Martini Coarse-Grained Protein Force Field,” *Journal of Chemical Theory and Computation*, vol. 9, pp. 687–697, jan 2013.
- [145] 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, pp. 2531–2543, sep 2009.
- [146] 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, jun 2010.
- [147] G. Crippen and T. Havel, *Distance geometry and molecular conformation*. New York: Wiley, 1988.
- [148] T. F. Havel, “The sampling properties of some distance geometry algorithms applied to unconstrained polypeptide chains: A study of 1830 independently computed conformations,” *Biopolymers*, vol. 29, pp. 1565–1585, oct 1990.
- [149] B. Velikson, T. Garel, J. C. Niel, H. Orland, and J. Smith, “Conformational distribution of heptaalanine: Analysis using a new Monte Carlo chain growth method,” *Journal of Computational Chemistry*, vol. 13, pp. 1216–1233, dec 1992.
- [150] W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, eds., *Computer Simulation of Biomolecular Systems*. Dordrecht: Springer Netherlands, 1997.
- [151] W. Braun and N. G, “Calculation of protein conformations by proton-proton distance constraints: A new efficient algorithm,” *Journal of Molecular Biology*, vol. 186, pp. 611–626, dec 1985.
- [152] A. E. Torda, R. M. Scheek, and W. F. van Gunsteren, “Time-averaged nuclear overhauser effect distance restraints applied to tendamistat,” *Journal of Molecular Biology*, vol. 214, pp. 223–235, jul 1990.

- [153] . Thomas Huber, , . Andrew E. Torda, and §. Wilfred F. van Gunsteren*, “Structure Optimization Combining Soft-Core Interaction Functions, the Diffusion Equation Method, and Molecular Dynamics,” 1997.
- [154] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by Simulated Annealing,” *Science*, vol. 220, pp. 671–680, may 1983.
- [155] B. Mao and A. R. Friedman, “Molecular dynamics simulation by atomic mass weighting,” *Biophysical Journal*, vol. 58, pp. 803–805, sep 1990.
- [156] Y. Okamoto, “Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations,” *Journal of Molecular Graphics and Modelling*, vol. 22, pp. 425–439, may 2004.
- [157] T. Huber, A. E. Torda, and W. F. van Gunsteren, “Local elevation: A method for improving the searching properties of molecular dynamics simulation,” *Journal of Computer-Aided Molecular Design*, vol. 8, pp. 695–708, dec 1994.
- [158] A. Laio and M. Parrinello, “Escaping free-energy minima.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 12562–6, oct 2002.
- [159] W. van Gunsteren and H. Berendsen, “Algorithms for macromolecular dynamics and constraint dynamics,” *Molecular Physics*, vol. 34, pp. 1311–1327, nov 1977.
- [160] C. Peter, C. Oostenbrink, A. van Dorp, and W. F. van Gunsteren, “Estimating entropies from molecular dynamics simulations,” *The Journal of Chemical Physics*, vol. 120, pp. 2652–2661, feb 2004.
- [161] R. Lipkin, A. Pino-Angeles, and T. Lazaridis, “Transmembrane Pore Structures of β -Hairpin Antimicrobial Peptides by All-Atom Simulations,” *The Journal of Physical Chemistry B*, vol. 121, pp. 9126–9140, oct 2017.
- [162] Y. Wang, D. E. Schlamadinger, J. E. Kim, and J. A. McCammon, “Comparative molecular dynamics simulations of the antimicrobial peptide CM15 in model lipid bilayers,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1818, no. 5, pp. 1402–1409, 2012.

- [163] L. Zhao, Z. Cao, Y. Bian, G. Hu, J. Wang, and Y. Zhou, “Molecular Dynamics Simulations of Human Antimicrobial Peptide LL-37 in Model POPC and POPG Lipid Bilayers,” *International Journal of Molecular Sciences*, vol. 19, p. 1186, apr 2018.
- [164] C. Chen, C. G. Starr, E. P. Troendle, G. Wiedman, W. C. Wimley, J. P. Ulmschneider, and M. B. Ulmschneider, “Simulation-guided rational de novo design of a small pore-forming antimicrobial peptide,” *Journal of the American Chemical Society*, p. jacs.8b11939, mar 2019.
- [165] H. J. Risselada and S. J. Marrink, “The molecular face of lipid rafts in model membranes.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 17367–72, nov 2008.
- [166] T. J. Piggot, D. A. Holdbrook, and S. Khalid, “Electroporation of the *E. coli* and *S. Aureus* Membranes: Molecular Dynamics Simulations of Complex Bacterial Membranes,” *The Journal of Physical Chemistry B*, vol. 115, pp. 13381–13388, nov 2011.
- [167] S. Khalid, T. J. Piggot, and F. Samsudin, “Atomistic and Coarse Grain Simulations of the Cell Envelope of Gram-Negative Bacteria: What Have We Learned?,” *Accounts of Chemical Research*, vol. 52, pp. 180–188, jan 2019.
- [168] T. S. Carpenter, J. Parkin, and S. Khalid, “The Free Energy of Small Solute Permeation through the *Escherichia coli* Outer Membrane Has a Distinctly Asymmetric Profile,” *The Journal of Physical Chemistry Letters*, vol. 7, pp. 3446–3451, sep 2016.
- [169] C. Song, B. L. de Groot, and M. S. Sansom, “Lipid Bilayer Composition Influences the Activity of the Antimicrobial Peptide Dermcidin Channel,” *Biophysical Journal*, vol. 116, pp. 1658–1666, may 2019.
- [170] Y. Wang, T. Zhao, D. Wei, E. Strandberg, A. S. Ulrich, and J. P. Ulmschneider, “How reliable are molecular dynamics simulations of membrane active antimicrobial peptides?,” *BBA - Biomembranes*, vol. 1838, no. 9, pp. 2280–2288, 2014.

- [171] . William L. Jorgensen, , David S. Maxwell, and J. Tirado-Rives, “Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids,” 1996.
- [172] 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, pp. 926–935, jul 1983.
- [173] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, “Scalable molecular dynamics with NAMD,” *Journal of Computational Chemistry*, vol. 26, pp. 1781–1802, dec 2005.
- [174] W. F. D. Bennett, C. K. Hong, Y. Wang, and D. P. Tieleman, “Antimicrobial Peptide Simulations and the Influence of Force Field on the Free Energy for Pore Formation in Lipid Bilayers,” *Journal of Chemical Theory and Computation*, vol. 12, pp. 4524–4533, sep 2016.
- [175] A. Sandoval-Perez, K. Pluhackova, and R. A. Böckmann, “Critical Comparison of Biomembrane Force Fields: ProteinLipid Interactions at the Membrane Interface,” *Journal of Chemical Theory and Computation*, vol. 13, pp. 2310–2321, may 2017.
- [176] J. P. M. Jämbeck and A. P. Lyubartsev, “Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids,” *The Journal of Physical Chemistry B*, vol. 116, pp. 3164–3179, mar 2012.
- [177] C. J. Dickson, B. D. Madej, Å. A. Skjervik, 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.
- [178] . Hari Leontiadou, , . Alan E. Mark, and . Siewert J. Marrink*, “Antimicrobial Peptides in Action,” 2006.
- [179] J. P. Ulmschneider, “Charged Antimicrobial Peptides Can Translocate across Membranes without Forming Channel-like Pores,” *Biophysical Journal*, vol. 113, pp. 73–81, jul 2017.

- [180] D. Sun, J. Forsman, and C. E. Woodward, “Atomistic Molecular Simulations Suggest a Kinetic Model for Membrane Translocation by Arginine-Rich Peptides,” *The Journal of Physical Chemistry B*, vol. 119, pp. 14413–14420, nov 2015.
- [181] Y. Wang, C. H. Chen, D. Hu, M. B. Ulmschneider, and J. P. Ulmschneider, “Spontaneous formation of structurally diverse membrane channel architectures from a single antimicrobial peptide,” *Nature Communications*, vol. 7, p. 13535, nov 2016.
- [182] N. R. Haria, L. Monticelli, F. Fraternali, and C. D. Lorenz, “Plasticity and conformational equilibria of influenza fusion peptides in model lipid bilayers,” *BBA - Biomembranes*, vol. 1838, no. 4, pp. 1169–1179, 2014.
- [183] F. Collu, E. Spiga, C. D. Lorenz, and F. Fraternali, “Assembly of Influenza Hemagglutinin Fusion Peptides in a Phospholipid Bilayer by Coarse-grained Computer Simulations.,” *Frontiers in molecular biosciences*, vol. 2, p. 66, 2015.
- [184] G. E. Balatti, E. E. Ambroggio, G. D. Fidelio, M. F. Martini, and M. Pickholz, “Differential Interaction of Antimicrobial Peptides with Lipid Structures Studied by Coarse-Grained Molecular Dynamics Simulations.,” *Molecules (Basel, Switzerland)*, vol. 22, oct 2017.
- [185] S. J. Marrink and D. P. Tieleman, “Perspective on the Martini model.,” *Chemical Society reviews*, vol. 42, pp. 6801–22, aug 2013.
- [186] H. Khandelia and Y. N. Kaznessis, “Molecular dynamics simulations of helical antimicrobial peptides in SDS micelles: What do point mutations achieve?,” *Peptides*, vol. 26, pp. 2037–2049, nov 2005.
- [187] C.-W. Tsai, N.-Y. Hsu, C.-H. Wang, C.-Y. Lu, Y. Chang, H.-H. G. Tsai, and R.-C. Ruaan, “Coupling Molecular Dynamics Simulations with Experiments for the Rational Design of Indolicidin-Analogous Antimicrobial Peptides,” *Journal of Molecular Biology*, vol. 392, pp. 837–854, sep 2009.
- [188] D. Wu and X. Yang, “Coarse-Grained Molecular Simulation of Self-Assembly for Nonionic Surfactants on Graphene Nanostructures,” *The Journal of Physical Chemistry B*, vol. 116, pp. 12048–12056, oct 2012.

- [189] H. Wang, H. Zhang, C. Liu, and S. Yuan, “Coarse-grained molecular dynamics simulation of self-assembly of polyacrylamide and sodium dodecylsulfate in aqueous solution,” *Journal of Colloid and Interface Science*, vol. 386, pp. 205–211, nov 2012.
- [190] D. Bochicchio and G. M. Pavan, “From Cooperative Self-Assembly to Water-Soluble Supramolecular Polymers Using Coarse-Grained Simulations,” *ACS Nano*, vol. 11, pp. 1000–1011, jan 2017.
- [191] H. Lee and R. W. Pastor, “Coarse-Grained Model for PEGylated Lipids: Effect of PEGylation on the Size and Shape of Self-Assembled Structures,” *The Journal of Physical Chemistry B*, vol. 115, pp. 7830–7837, jun 2011.
- [192] P. Brocos, P. Mendoza-Espinosa, R. Castillo, J. Mas-Oliva, and Á. Piñeiro, “Multiscale molecular dynamics simulations of micelles: coarse-grain for self-assembly and atomic resolution for finer details,” *Soft Matter*, vol. 8, p. 9005, aug 2012.
- [193] C. Guo, Y. Luo, R. Zhou, and G. Wei, “Probing the Self-Assembly Mechanism of Diphenylalanine-Based Peptide Nanovesicles and Nanotubes,” *ACS Nano*, vol. 6, pp. 3907–3918, may 2012.
- [194] M. Seo, S. Rauscher, R. Pomès, and D. P. Tieleman, “Improving Internal Peptide Dynamics in the Coarse-Grained MARTINI Model: Toward Large-Scale Simulations of Amyloid- and Elastin-like Peptides,” *Journal of Chemical Theory and Computation*, vol. 8, pp. 1774–1785, may 2012.
- [195] O.-S. Lee, V. Cho, and G. C. Schatz, “Modeling the Self-Assembly of Peptide Amphiphiles into Fibers Using Coarse-Grained Molecular Dynamics,” *Nano Letters*, vol. 12, pp. 4907–4913, sep 2012.
- [196] S. Gudlur, P. Sukthankar, J. Gao, L. A. Avila, Y. Hiromasa, J. Chen, T. Iwamoto, and J. M. Tomich, “Peptide Nanovesicles Formed by the Self-Assembly of Branched Amphiphilic Peptides,” *PLoS ONE*, vol. 7, no. 9, 2012.
- [197] J. M. A. Grime, J. F. Dama, B. K. Ganser-Pornillos, C. L. Woodward, G. J. Jensen, M. Yeager, and G. A. Voth, “Coarse-grained simulation

- reveals key features of HIV-1 capsid self-assembly.,” *Nature communications*, vol. 7, p. 11568, may 2016.
- [198] J. R. Perilla, J. A. Hadden, B. C. Goh, C. G. Mayne, and K. Schulten, “All-Atom Molecular Dynamics of Virus Capsids as Drug Targets.,” *The journal of physical chemistry letters*, vol. 7, no. 10, pp. 1836–44, 2016.
- [199] J. A. Hadden and J. R. Perilla, “All-atom virus simulations,” *Current Opinion in Virology*, vol. 31, pp. 82–91, aug 2018.
- [200] A. Abi Mansour and P. J. Ortoleva, “Multiscale Factorization Method for Simulating Mesoscopic Systems with Atomic Precision,” *Journal of Chemical Theory and Computation*, vol. 10, pp. 518–523, feb 2014.
- [201] C. C. G. ULC, “Molecular operating environment (moe), 2013.08,” 2018. 1010 Sherbooke St. West, Suite 910, Montreal, QC, Canada, H3A 2R7.
- [202] A. Malde, L. Zuo, M. Breeze, M. Stroet, D. Poger, P. Nair, C. Oostenbrink, and A. Mark, “An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0,” *Journal of Chemical Theory and Computation*, vol. 7, pp. 4026–4037, dec 2011.
- [203] K. Koziara, M. Stroet, A. Malde, and A. Mark, “Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies,” *Journal of Computer-Aided Molecular Design*, vol. 28, pp. 221–233, mar 2014.
- [204] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, “PACK-MOL: A package for building initial configurations for molecular dynamics simulations,” *Journal of Computational Chemistry*, vol. 30, pp. 2157–2164, oct 2009.
- [205] D. Poger, W. F. Van Gunsteren, and A. E. Mark, “A new force field for simulating phosphatidylcholine bilayers,” *Journal of Computational Chemistry*, vol. 31, pp. 1117–1125, apr 2010.
- [206] A. Kukol, “Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins,” *Journal of Chemical Theory and Computation*, vol. 5, pp. 615–626, mar 2009.

- [207] N. Kučerka, M.-P. Nieh, and J. Katsaras, “Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1808, no. 11, pp. 2761–2771, 2011.
- [208] J. Pan, F. A. Heberle, S. Tristram-Nagle, M. Szymanski, M. Koepfinger, J. Katsaras, and N. Kučerka, “Molecular structures of fluid phase phosphatidylglycerol bilayers as determined by small angle neutron and X-ray scattering,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1818, pp. 2135–2148, sep 2012.
- [209] C. Kandt, W. Ash, and D. Peter Tielemans, “Setting up and running molecular dynamics simulations of membrane proteins,” *Methods*, vol. 41, pp. 475–488, apr 2007.
- [210] C. Oostenbrink, T. A. Soares, N. F. A. van der Vegt, and W. F. van Gunsteren, “Validation of the 53A6 GROMOS force field,” *European Biophysics Journal*, vol. 34, pp. 273–284, jun 2005.
- [211] M. M. Reif, M. Winger, and C. Oostenbrink, “Testing of the GROMOS Force-Field Parameter Set 54A8: Structural Properties of Electrolyte Solutions, Lipid Bilayers, and Proteins,” *Journal of Chemical Theory and Computation*, vol. 9, pp. 1247–1264, feb 2013.
- [212] J. R. Wilson, R. B. Clark, U. Banderali, and W. R. Giles, “Measurement of the membrane potential in small cells using patch clamp methods,” *Channels*, vol. 5, pp. 530–537, nov 2011.
- [213] D. P. Tielemans, “The molecular basis of electroporation.,” *BMC Biochemistry*, vol. 5, p. 10, jul 2004.
- [214] R. A. Böckmann, B. L. de Groot, S. Kakorin, E. Neumann, and H. Grubmüller, “Kinetics, Statistics, and Energetics of Lipid Membrane Electroporation Studied by Molecular Dynamics Simulations,” *Biophysical Journal*, vol. 95, pp. 1837–1850, aug 2008.
- [215] J. A. Graham, J. W. Essex, and S. Khalid, “PyCGTOOL: Automated Generation of Coarse-Grained Molecular Dynamics Models from Atomistic Trajectories,” *Journal of Chemical Information and Modeling*, vol. 57, pp. 650–656, apr 2017.

- [216] 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.
- [217] . Siewert J. Marrink, , Alex H. de Vries, and A. E. Mark, “Coarse Grained Model for Semiquantitative Lipid Simulations,” 2003.
- [218] T. A. Wassenaar, K. Pluhackova, R. A. Bckmann, 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, no. 2, pp. 676–690, 2014. PMID: 26580045.