

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

To my parents, for letting me go

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	Drug delivery: challenges and solutions	12
1.2	Antimicrobial resistance	17
1.3	Alternative antibiotic strategies: antimicrobial peptides	25
1.4	Gene therapy	43
1.5	Closing the circle: an antimicrobial drug delivery vehicle	46
1.6	A computational approach to understand capzip	53
2	Methods	56
2.1	The force field problem	57
2.2	The search problem	68
2.3	The ensemble problem	71
2.4	The experimental problem	72
2.5	MD simulations: successes	74
	Appendices	76
A.1	On the derivation of the GP predictive distribution	76
	Bibliography	77

List of Figures

1.1	Graphical abstract of introduction	11
1.2	Materials for drug delivery vehicles	14
1.3	Mechanisms of antimicrobial resistance to small drugs	22
1.4	Antimicrobial peptides	27
1.5	Principles of gene therapy	44
1.6	Cazip molecule	47
1.7	Experimental results on capzip	51

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

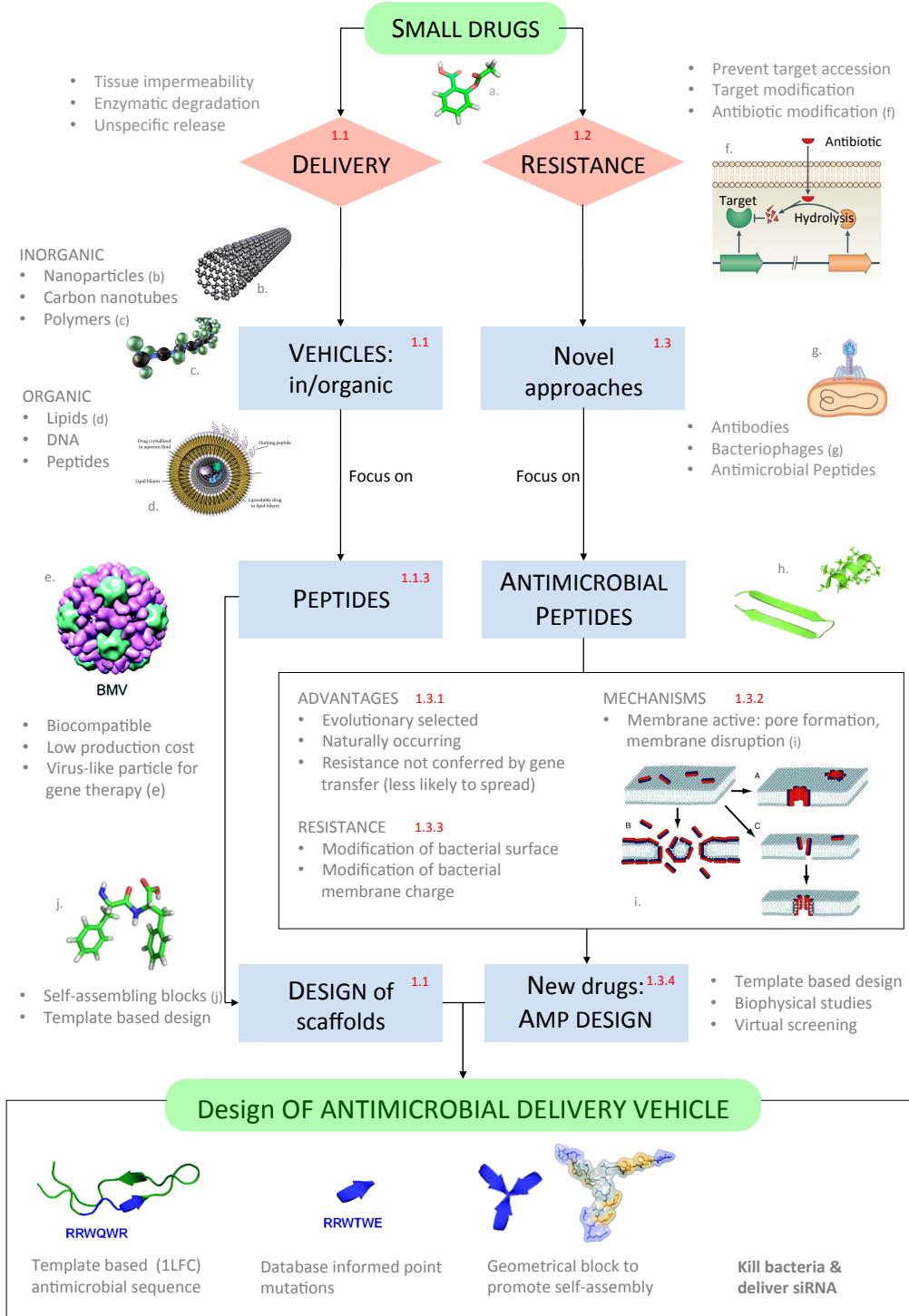


Figure 1.1: Figures a. (acetylsalicylic acid) and j. (diphenyl-alanine) in bond representation. Remaining figures adapted from: b. [-]; c. [1]; d. [2]; e. [3], f. [4]; g. [5]; h. [6]; i. [7]; k. [8]

1.1 Drug delivery: challenges and solutions

1.1.1 Environmental challenges of drug delivery

The problem of drug delivery is an excellent example of the hurdles existing between a theoretical model and nature: a new drug is usually designed to affect a specific target. Even if *in silico* experiment can prove the efficacy, its usefulness is bound to the ability to cross the barriers dividing the inoculation site from the specific target inside the human body. To reach the aimed organ, drug molecules must be compatible with the different cellular environments they cross, but be preferentially retained from, and act only on, the ones they are designed for. This implies a subtle balance between a disruptive activity on one side, and harmlessness on the other, least the compound is recognised as dangerous and disposed of by the efficient immune and reticuloendothelial systems of the body, which aim at neutralise exogenous substances.

As an example, the trip of an orally administered small molecule “free” drug, i.e. an active molecule without any aiding delivery agent, passes through the digestive system, with its challenging acidic environment and limited permeation across the intestinal epithelium, and from there to the blood stream [9, 10]. The drug then diffuses in the tissues flanking the blood vessels naturally depleting its concentration downstream [11], so that regions further away in the line have less chances of getting a sizeable dose. This implies that high drug concentrations might be needed at the starting point to efficiently target every organ.

However, this naive picture of drug diffusion is complicated by the impermeability of specific tissues: the brain for example, one of the most delicate organs in the body, is well protected from the attack of external agents by the blood brain barrier (BBB), which allows the passage of small molecules only (< 400-500 Da, while standard “small molecule drugs” weigh up to 900 Da), and only if they have high lipid solubility [11, 12]. Other tissues, like tumoral ones, are instead poorly vasculated, reducing the chances of delivery at their interior [12].

Moreover, as already hinted, during its journey to reach the receptor, enzyme or organelle it is meant for, a drug must not be sequestered by the immune systems. Many inorganic small molecules are not mimetic by themselves, i.e. often they do not resemble the ones naturally present in the body,

and this brings uncertainty on how they would interact with the, for example, the components of the blood stream. Generally, as soon as they reach it, small molecules are coated by a protein corona based on their shape and charge [11]. Such modifications are difficult to predict and can disrupt or decrease significantly the efficacy of a compound as they modify the way drugs are recognised and absorbed by the target.

For all the above reasons, research has focussed on developing systems to assist the delivery of drugs [10, 12, 13]. A mimetic carrier can not only improve delivery, but also be designed to selectively bind to particular tissues or to trigger a delayed drug release time or upon changes in environmental variables (for example pH) to reduce drug concentration in non targeted regions. A stand alone field of research has then focussed on the development of delivery vehicles irrespective from the quest for new drugs. The optimised products of the two separate efforts can then be paired according to the condition to address, to give a successful therapy.

At present, many molecules have been successfully employed to build drug vehicles: inorganic metals, polymers, lipids and proteins are all suitable for the aim and offer a range of different physico-chemical characteristics useful to target different cells [14] (Figure 1.2). A brief (and non exhaustive) overview of some of them is meaningful to point out the broad variety and exotoxicity of structures which are useful, sometimes unexpectedly, to the medical world.

1.1.2 Inorganic materials for small drugs delivery

Metal nanoparticles In the range of inorganic compounds, golden nanoparticles demonstrated to be remarkably useful for tumour treatment: their structure can be customised in shape and size (down to a 10 nm radius), made less visible to the immune system by coating with biologically active moieties or conjugation to a poly-ethyleneglycol (PEG) polymer layer [15]. From their metallic nature, golden nanoparticles inherit optical properties that allow to track them inside the body and to thermally stimulate them to trigger drug release, favouring the penetration through cell membrane or disrupting the nearby cells [16]. This is important for tumour treatment because of the difficulty with which tumoural cells are reached by drugs and the selectivity required when delivering a highly damaging chemotherapeutic.

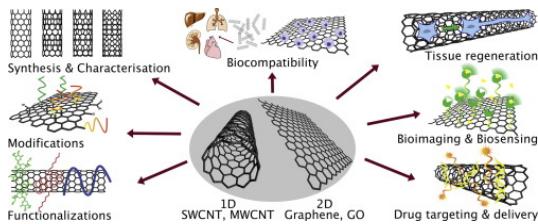


Figure 1.2: INCLUDE ONE EXAMPLE IMAGE (FROM PAPERS) FOR EACH MATERIAL (nanoparticles, carbon, polymers, lipids, DNA, proteins).

At present, there are mixed evidence about their toxicity [16] and doubts have been raised on the long term effects of metallic fragments in the body. For that reason, only a few golden nanoparticle based compounds have made to the clinical stage so far [15] but, given their high and still unexplored potential, they continue to be a primary interest for the medical community and a very active research field.

Carbon nanotubes Similarly, carbon nanotubes have been used for biomedical applications as they have a high loading efficiency thanks to their high surface area and easy interaction with biomolecules through van der Waals interactions, π - π stacking or hydrophobic effect [17]. They are easy to functionalise through conjugation to extra organic groups to increase their biocompatibility and have potential for targeted drug release upon change in environmental pH [18].

Similarly to metallic nanoparticles, their applications are still at the experimental stage, with more verifications to be performed to give a viable product.

Polymers Polymers are another large class of inorganic molecules functionalised for the benefit of medicine with several representatives already clinically approved as drug excipients. For example, PEG, already mentioned as aid to make golden nanoparticles bio-compatible, it is widely used, thanks to its high hydrophilicity, to mimetise structures (e.g. inorganic nanoparticles, peptides) which in turn carry a drug [19]; or as a stand alone carrier system which has a high drug payload [20]. The great strength of polymers is their flexibility: as each of their constituent monomers can be either hydrophilic or hydrophobic,

they can be engineered to assemble in many different structures [21]. Moreover, they can trigger a sustained drug release by swelling slowly in water [22], or undergo sol-gel phase transition upon specific changes in the environment [20]. Finally, research has also focussed on improving their biodegradability [23] and in making polymers a bioactive compound itself [24].

1.1.3 Organic materials for small drugs delivery

A somehow opposite approach for designing drug vehicles consists in using molecules similar to the ones present in the body, in an effort to exploit already available biocompatible materials and reduce toxicity [25]. In this category fall lipids, DNA and peptides.

Lipids Lipids are the main constituents of the cell membrane. They come with a great variety, broaden by the many species produced synthetically. The components selected for drug delivery are usually taken from the biological lipidome [?], but the composition of the final assembly differs from the one of cellular membranes, and possibly includes synthetic molecules, to tune the release properties and enable them to survive the delivery journey [26]. For example, as mentioned for other materials, a change in pH can dissolve the lipid structure exposing its content. Lipids can encapsulate efficiently both hydrophobic or hydrophilic drugs, arranging themselves respectively in micelles structures (monolayer spheres with the hydrophobic tails facing the interior) or in liposomes (bilayer spheres with a water filled core) [27]. By now, many of them overcame the clinical trials and are currently approved as delivery agents for cancer and infections drugs [28, 29].

DNA scaffolds Similarly, many DNA scaffolds have been tested for smart delivery: DNA origami is nowadays an established technique to build three dimensional customised solids [30], and the nanometric knowledge about their constituents allows to fine tune them for a triggered release of the content [31]. First studies proved them successful in delivering anticancer agents [32, 33], however they are very sensitive to different cellular environments which challenge their stability. This, united with high production costs and the relative young developments in their manipulation, prevented them to constitute a

viable class of carriers so far, but at the same time holds promise for future improvements and applications.

Peptidic scaffolds Another widely used and trustworthy mimetic vehicle comes, quite surprisingly, from the world of pathogens: viruses have co-evolved with humans, to be able to penetrate into cells where they complete their reproductive cycle [34]. Therefore their capsid, the peptidic shell encapsulating the genome, is highly suitable for cell penetration. The first application sought historically was to employ genome free viruses to stimulate and train the natural immune response against the respective genome-loaded ones, creating viral vaccines - in a concept similar to the inoculation of dead bacteria to counteract the infections caused from them [35]. Later in the history, their potential as cargo carrier was pursued by first modifying the original genetic material to include sequences beneficial for the host cell, and inactivate the ones promoting the infectious duplication at the same time. In particular adeno-associated virus (AAV) has been widely studied [36] as it triggers a low immune response [37], and the first AAV viral therapy has been approved a few years ago [38]. To fully exploit the potential of a peptidic carrier many efforts have focussed on synthesising *in vitro* gene-free capsids, either as they appear in nature [39] or designing artificial building blocks, which assemble in so called Virus-Like particles (VLPs). This helps overcoming the reaction stimulated by specific viral capsids to which the immune system is sensible to. Similarly to other delivery vehicles, the surface of VLPs can be functionalised with additional molecules to improve the target selectivity and increase biocompatibility, while the capsid peptidic scaffold grants robustness to the structure. Therefore, VLPs loaded with drugs can be tuned for an efficient intra cellular release [40].

A step further in engineering peptidic structures is represented by the design of self-assembling functional structures from first principles, exploiting the physico-chemical characteristics of peptides, regardless their resemblance of viral capsids. Indeed, self-assembling peptides can form nanostructures ranging from nanoparticles to nanotubes, nanofibers, nanorods and hydrogels [41, 42]. Among their advantages, peptides present biocompatibility, a low production cost and a tunable bioactivity thanks to their chemical diversity, which helps in tailor the assembly toward the target of interest [41]. Moreover, the variety of amino acid available makes possible to load peptidic structures with both

hydrophilic and hydrophobic drugs, according to their amino acid composition [40, 42]. The peptidic self-assembly is modulated by the peptide length and its hydrophobic or hydrophilic character, given by its amino acid composition: on one end of the length scale, phenylalanine dipeptides were designed with inspiration from a pathogenic pathway of molecular self-assembly [43] and were shown to self-assemble in a multi-scale process producing nanotubes able to load drug molecules [44]. The relatively small diphenylalanine building block is non the less complex as it bears two charged termini (as the process is observed at neutral pH), and two aromatic hydrophobic rings, so that the dipeptide is driven towards assembly by the hydrophobic forces acting on the phenylalanine side chains and the complementary charges of the termini.

In a different approach, longer sequences can be employed to guide the formation of the local structure, as they organise spatially in well studied motives (the secondary structure) with known interactions among themselves. The two typical secondary structures, α -helices and β -sheets, appear in sequences of about 20 or more amino acids length and both can be amphiphatic, thus promoting the assembly between the hydrophobic faces of different copies of the same structure. With the appearance of a secondary structure, more complex building blocks can be designed, to tune the shape into the ones needed for the supramolecular organisation of interest [45]. The easy manipulation of peptides is made possible because proteins are a fundamental, well studied component of the human body: thus, the knowledge of many of their structures[46] give an insight in how the small ones can hierarchically assemble into larger units. Moreover, the vast literature on their interactions with membranes, cell receptors and in general biological components, can inspire the design of building blocks sensible to particular triggers within the body. From this background, the outlook of protein design often goes in the direction of surpassing natural limitations, synthesising exotic, non natural, geometries [47, 48] for multifunctional materials.

1.2 Antimicrobial resistance

The previous brief review on drug carriers rotates 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 this

light, the ultimate goal of the delivery vehicle is to carry the drug to the site of action where it can interfere with the process it is assigned to. 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 [49]. (It must be noticed however that the identification of an unambiguous drug target poses challenges in many cases, especially when the drug binds to a protein complex or to a number of closely related gene products [49].)

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. 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 [?]. Of course, 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 actually demonstrates 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 po-

tent, broad-spectrum compounds. Penicillin, the first of them, 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 (for further details on bacterial cell membrane structure the reader can refer to Section 1.3.1 and the relative references). This inhibition is achieved through binding to the enzyme DD-transpeptidase responsible for the catalysis of such cross-link [50]. 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 [?].

1.2.1 Course of antimicrobial resistance

The mechanism just outlines is not an exceptional characteristic of penicillin, and many drugs lost their effectiveness against some bacteria since their discovery till nowadays. Indeed the antibiotic landscape is a dynamic entity in which newly discovered ones enter, while others exit after having been exploited for years.

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, 51]. 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 [?].

The spread of resistance between bacterial cells and even between species is very effective as bacteria are able to exchange genetic material with other individuals via small rings of DNA in a process called conjugation [?]. In this way the advantageous characters spread across individuals, and species with an innate resistance can transfer to other ones their mechanisms of resilience to a particular drug. 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 [52, 53], 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, in particular to 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.

Finally, it is noteworthy that abuse of antibiotics can take many forms: the agricultural and breeding sectors are constantly using antibiotics to keep their products secure from illness. This results in large quantities of drugs to be released in the soil and water, which ultimately reach humans in underdosed

quantities. Diseases of plants and animals are different from the ones affecting humans, however some drugs are effective on many bacteria including the one harming humans. Therefore the widespread use of antibiotic for animals or plants can ultimately train resistant bacteria in humans [?]. Additionally, diseases can cross species: this means that an extra care must be taken in the treatment of non human bacteria least to promote resistant ones which can at a point affect us [?].

The complexity and severity of the AMR issue is such that it has been raised to the status of national emergency in several countries, including UK, 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 [54].

1.2.2 Mechanisms of antimicrobial resistance to small drugs

Antimicrobial resistance can manifest through many different mechanisms, as highlighted in the example of the penicillin resistant bacteria. In particular, resistance mechanisms fall into three main groups: a first group minimises intracellular concentration of the antibiotic preventing penetration or maximising efflux; a second one modifies the antibiotic target by genetic mutation or post-translational modification; finally a third group inactivates the antibiotic by hydrolysis or modification of the drug molecule (Figure 1.3) [4]. We give here a brief review of them, to help understanding the pitfalls of existing drugs and the characteristics sought in the developments of new compounds.

Prevention of access to target One possible mechanism of defence bacteria employ against antibiotics is to prevent the access to the target. This is performed either blocking the drug influx or promoting its quick efflux in the eventuality it has entered the cell.

Regarding drug influx, not all the molecules can enter the cell permeating the membrane, and this holds particularly for hydrophilic antibiotics tackling Gram-negative bacteria: indeed, compared with Gram-positive ones, Gram-negative bacteria are intrinsically less permeable because of the structure of the additional outer membrane [55], therefore hydrophilic molecules are imported into the cell through outer-membrane porin proteins [56, 57] (for further details on the bacterial membrane structure, see Section 1.3.1). The major porins of most Enterobacteriaceae are thought to be non-specific channels [58], therefore

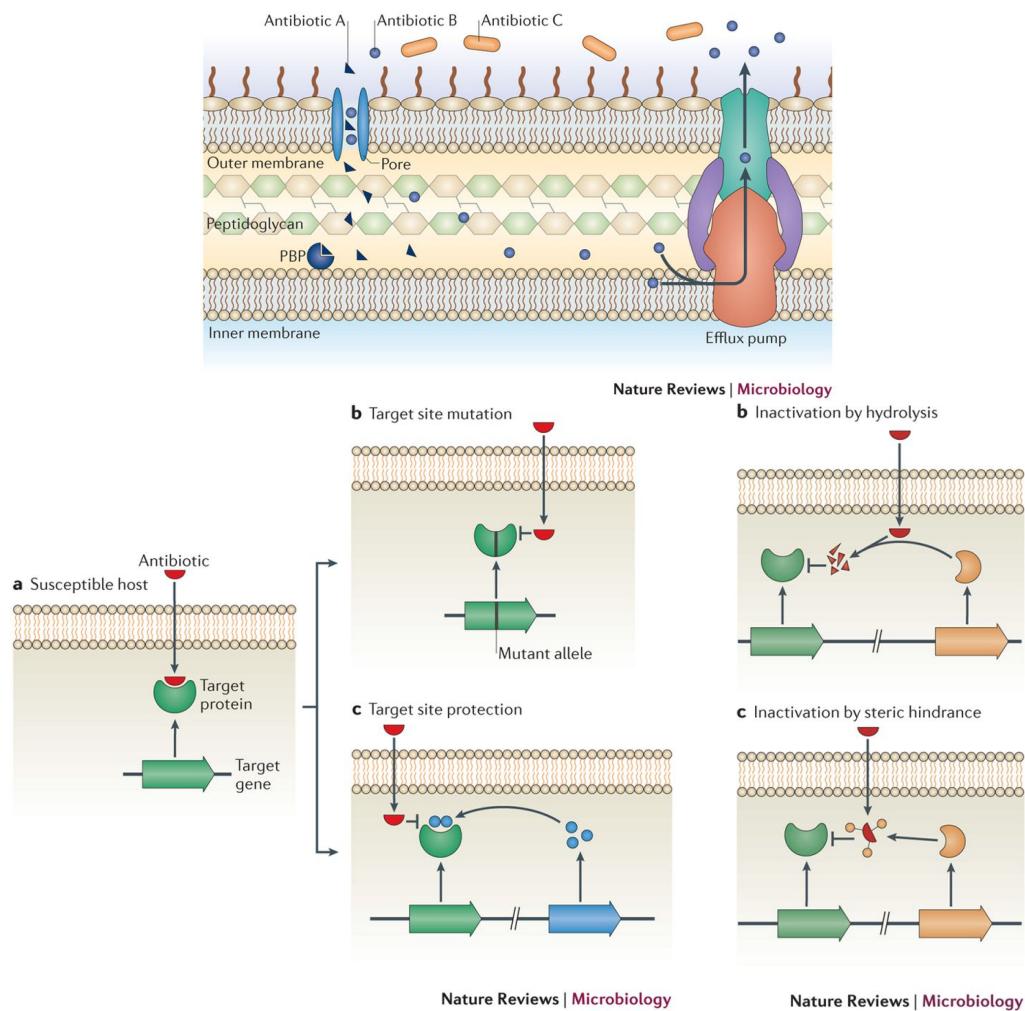


Figure 1.3: Mechanisms of antimicrobial resistance to small drugs. a) Intrinsic mechanisms of resistance (removal of antibiotic B by efflux pump and inaccessibility of antibiotic C to the PBP target because of membrane impermeability). b) Target site change via mutation or protection. c) Direct interactions with antibiotics causing its disruption or structural modification. Reproduced from [4].

replacing porins with more selective channels or down regulating their expression would limit the intake of the drug. This last mechanism is well established and contribute to resistance to many different drugs in Gram-negative bacteria, including newer drugs such as carbapenems and cephalosporins, for which resistance is otherwise mediated by enzymatic degradation [59–63] (see the related paragraph below). Alternatively, as happens in *E. coli* exposed to car-

bapenems, not only the porin expression is down-regulated, but also the genes coding for them are heavily mutated, suggesting that changes in the porin structure can enhance their selectivity and reduce the drug influx [61, 64, 65].

A strategy complementary to prevent drug influx is to dispose the drug efficiently once it has invaded the cell. Bacterial efflux pumps transport many antibiotics out of the cell, and they constitute a major hurdle for the treatment of Gram-negative bacteria as opposed to Gram-positive ones. Indeed, many of the drugs effective on the latter are evacuated by the formers through efflux pumps; in particular, multidrug resistance (MDR) efflux pumps can transport a wide range of structurally dissimilar substrates. All bacteria can produce their own MDR pumps [66–69], but it has also been shown that some of the genes encoding for them have been transferred to plasmids and thus can be transferred to other bacterial species, disseminating resistance [70].

In general, the over-expression of efflux pump seen in multidrug-resistant bacteria is often due to mutation in the regulatory network controlling it, [71], but it can also occur as a result of induction in response to environmental signals and in conditions in which their function is required [72–74].

Change or modification of the antibiotic target Most antibiotics bind to the target with high affinity and therefore specificity. Small modifications of the target structure can disrupt an efficient binding of the antibiotic, still allowing the target to maintain its normal function. These modifications can be reached by either mutation or protection of the binding site.

In the first case, a casual mutation in the genome would provide such minimal required change in the protein structure, and the resistant population would spread according to its improved fitness. An example is the development of resistance to linezolid in *S. pneumoniae* and *S. aureus*: this drug targets the 23S rRNA ribosomal subunit of Gram-positive bacteria which is encoded by multiple, identical copies of its gene. The use of linezolid has selected first a population with a mutation in one of the copies, which has afterwards passed to the other copies via recombination, generating a resistant population [75, 76]. Other mutations occurred by transformation, i.e. uptake of DNA from the environment: in the case of penicillin resistant *S. pneumoniae*, a mutated penicillin-binding protein gene is included in the genome by recombination with DNA from the closely related species *Streptococcus mitis*. Similarly, the

acquisition of a gene homologous resulted in a methicillin-resistant strain of *S. aureus* [77]: this gene allows the synthesis of the PBP2 (penicillin binding protein 2) protein which enable cell wall synthesis despite the native PBP is inhibited by the antibiotic [78].

The second modification mentioned consists in protecting the target from the binding of the drug via addition of chemical groups to the target after its synthesis; as such, these modifications do not require mutations at the genetic level. Among them, methylation is an important process: for example, under the pressure of macrolides, lincosamines and streptogramins, the 16S rRNA subunit is methylated and thus the drug-binding site altered [79]. Similarly, specific methylation of a base (A2503) in the 23S rRNA subunit confers resistance to many drugs that target nearby regions (phenicols, pleuromutilins, streptogramins, lincosamides and oxazolidinones) [80]. In a different mechanism, quinolone resistance can be conferred by a gene coding for a pentapeptide repeat proteins (PRPs), which binds to topoisomerase IV and DNA gyrase promoting the release of the drug and rescuing the normal function of topoisomerase [81].

Direct modification of antibiotics Finally, bacteria can modify or destroy drugs to prevent their action, usually by either hydrolysis or by transfer of a chemical group. Enzyme-catalysed modification of antibiotics is a major mechanism of antibiotic resistance: the very first example being penicillinase (a β -lactamase) which destroy penicillin [82]. Since this discovery, thousands of enzymes have been identified that can degrade and modify antibiotics of different classes, such as β -lactams, aminoglycosides, phenicols and macrolides [83–86]. These enzymes co-evolved together with the newly developed drugs which bacteria are exposed to, to include in their spectrum of disruptive action new compounds of similar composition: for example the first β -lactamases evolved to be active against the new β -lactams antibiotics developed, up to the emergence of isolates resistant to all the drugs in the β -lactam class [87]. This localised emergence of resistance is a particularly serious problem because of the effectiveness with which these mechanisms spread to the whole bacterial population in a short period of time [85, 87, 88], as mentioned before.

The addition by bacterial enzymes of chemical groups (for example acyl, phosphate, nucleotidyl or ribitoyl [89]) to vulnerable sites on the antibiotic

molecule is another mechanism to block the action of the drug, as it prevents its binding to the target protein due to steric hindrance. Antibiotics constituted by large molecules with many exposed hydroxyl and amide groups are particularly susceptible to these modifications. An example of such antibiotics is the aminoglycoside class (in which streptomycin is included), which can be modified by three classes of enzymes, grouped according to the chemical moiety added: acetyltransferases, phosphotransferases and nucleotidyltransferases [90]. A recent development reports the discovery of a genetic island in *Campylobacter coli* isolated from broiler chickens in China coding for six of these enzymes, including members of all three classes: the expression of such genes would then confer resistance to many antibiotic of the aminoglycoside class at once [91].

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. This in turn has highlighted the effectiveness of some clinical strategies, such as the use of combined therapies, to counteract an early development of resistance. However, the problem persists and more knowledge needs to be gathered for a complete understanding and the possible development of resistance-free compounds.

1.3 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. There are several possible solutions to this: antibodies, bacteriophages or antimicrobial peptides instead of small molecules [92].

Regarding antibodies, the development of pathogen-specific monoclonal antibodies (mAb) is an emerging area of research. For example, they can be employed for immunisation through serum therapy, i.e. exposing the patient to

the serum of an individual already immunised. Such passive immunization has been used for the treatment of bacterial infections well before the discovery and development of antibiotics, but has since then been overshadowed by the use of small-molecule compounds and is now regaining relevance.

The second class mentioned, bacteriophages, are formed by viruses which infect bacteria and archaea rather than eukarya. They are effective as they can be used both in natural environmental reservoirs and in humans and are usually highly specific against one bacterial strain. Both these strategies have been only partially explored so far, bringing potential for new therapies. Phage therapy is promising also in terms of promoting less resistance development: indeed phages and bacteria have been coexisting since a long time - in evolutionary scale - and the formers are never the less effective against the latter, suggesting that their mechanism of attack is weakly prone to provoke defence mechanisms [?].

But are antimicrobial peptides the focus of this thesis: we have already highlighted the importance of peptides as tunable structural elements of drug delivery vehicles. However, they can have a role against bacteria as drug themselves when their sequence possesses some specific characteristics: such active sequences, capable of damaging and/or killing bacteria, are referred to as antimicrobial peptides. The following subsections will explore their characteristics, modes of action and the response of bacteria against them: indeed, it is crucial to understand the knowledge available versus the questions that are still open. 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.

1.3.1 Host-defence, membrane active peptides

Antimicrobial peptides (AMPs) are naturally produced by mammals, either as stand-alone sequences or embedded in larger proteins, as a first, weak, and broad-spectrum defence against bacteria [7, 93–95]. Similar to phages, 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

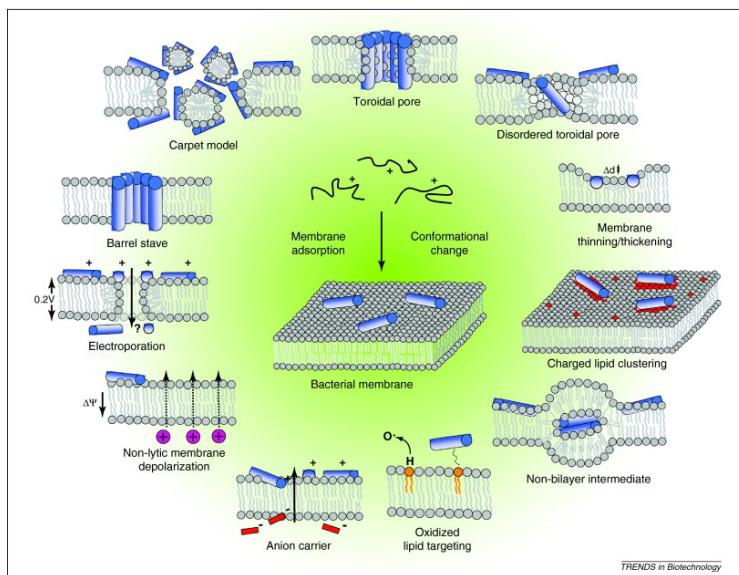


Figure 1.4: Events occurring at the bacterial cytoplasmic membrane following initial antimicrobial peptide (AMP) adsorption. Reproduced from [7].

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 [96]. 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 [97]: 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 [98] (Figure 1.4). In general, it is widely accepted that membrane interaction is a key factor for the direct antimicrobial activity of AMPs [7, 99]. The determinant driving this interaction is the positive charge that many AMPs present, opposed to the negative charge of bacterial membrane [100, 101]. It is striking that such simple mechanism, based on the presence of a certain number of neg-

atively 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 [102]. The cytoplasmic membranes of both Gram-positive and Gram-negative bacteria are rich in phospholipids like phosphatidylglycerol, cardiolipin, and phosphatidylserine, which have negatively charged headgroups, highly attractive for positively charged AMPs, and this is often sufficient to promote the preferential interaction between this membrane and the peptides.

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 [103]. Indeed, mammalian cells have a different membrane composition, in particular their membrane is rich in zwitterionic phospholipids such as phosphatidylethanolamine, phosphatidylcholine, and sphingomyelin, providing a neutral net charge [104, 105]. 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 cholesterol content [106, 107], which is proposed to stabilise the membrane enhancing its fluidity, so that it is more able to accommodate the perturbations caused by AMPs [108].

Finally, bacterial cells 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 [106, 109, 110]. 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 [106].

1.3.2 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 [111]. 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 1.4) [7, 97, 112].

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 [110]. 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 [113], 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 [113–116].

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 [113, 117, 118], and lead to membrane perturbation which extends further away from the pore than in the barrel-stave case, as lipids must rearrange around them [114].

As a comparison, alamethicin induced barrel-stave pores have an inner and outer diameters of 1.8 nm and 4.0 nm respectively [116, 119], 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 [120, 121].

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 [122, 123]. 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 [124]. The “carpet” model mechanism is observed for peptides presenting an α -helical structure with two or more helices connected by short loops (like cecropin [125] or ovispirin [126]).

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 [127]).

In the case of defensins, their mechanisms are not as well explored [108, 128, 129]. Some members of the family form transmembrane pores on planar bilayer when a physiologically relevant negative potential is applied to the membrane,[130] while others form oligomers in phospholipid vesicles [131]. 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 [132].

Alternative mechanisms of action Finally, many non-lytic mechanisms are suggested for AMPs, especially for β -sheet structures: defensin A from *P. terranova* reduces the cytoplasmic potassium concentration, partially depolarising the inner membrane; tachypleasin from horseshoe crabs is able to bind to the minor groove of DNA, interfering DNAprotein interactions [133], 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 [134]. Moreover, after translocation within the cell, bovine lactoferricin can also inhibit DNA, RNA and protein synthesis. Section 1.5.1 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 [135–138], 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.

1.3.3 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. As such, AMPs might not be a definitive solution to the problem; never the less, the resistance to their action is generally not based on dedicated resistance genes that are conferred by horizontal gene transfer, as in the case of many antibiotics resistance mechanisms [139, 140]. 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 resistance to AMPs are similar to the ones employed by bacteria to counteract small molecule drugs, for example overexpression of efflux pumps to dispose of AMPs, proteolytic degradation of the

peptide by extracellular enzymes, or sequestration by the bacterial biofilm matrix which prevents accession to the target. Others instead tackle the specific action of AMPs on the cell membrane, and prevent it by modifying the composition of the surface or of the cytoplasmic membrane. Table 1.1, from Ref. [141] lists these mechanisms, offering examples for each, in both Gram-positive and Gram-negative bacteria. In the following paragraphs some of them are explained in more details, with the omission of efflux pump, for which the principle is practically identical to what explained in Section 1.2.2.

Mechanism	Gram-positive bacteria	Gram-negative bacteria
Extracellular proteins	Proteolytic degradation Sequestration	Proteolytic degradation
Exopolymers	PIA, PGA	Alginate, polysialic acid
Surface modification	Repulsion (D-alanylation of TA) Steric hindrance (L-rhamnosylation of WTA) Lipid II modification	Repulsion (lipid A phosphate modification) Increased OM rigidity (lipid A acylation) O-antigen of LPS
Cytoplasmic membrane alteration	Charge repulsion (PG amino-acylation)	Increased IM rigidity (PG acylation)

Table 1.1: Overview of bacterial resistance mechanisms against antimicrobial peptides. Adapted from Ref. [141]

Proteolitic degradation and sequestration [TAKE OFF?] The first defence of bacteria against AMPs are proteins secreted on the extracellular side of the membrane, the proteases, as they are able to degrade peptides, and thus AMPs. For example, staphylococci secrete many of these proteins (such as the metalloproteases aureolysin and SepA, or the serine endopeptidases V8), which are known to degrade linear AMPs as the human cathelicidin LL-37 [142, 143]. In general, linear AMPs are more easily degraded than the ones with non-linear structures containing disulfide bonds [139] such as defensins [148]. However, some bacteria have evolved proteases able to degrade even these AMPs with increased stability; for example group A Streptococcus produces

a cysteine protease able to disrupt LL-37 and beta-defensins [144–147], and the OmpT protein contributes to resistance in *E. coli* by degrading the AMP protamine [149] which is thought to have a non-linear structure involving three disulfide bonds [150]. In an adaptation to such resistance mechanism, AMPs can escape the action of proteases by binding to proteins such as extracellular actin, preventing the access of degradative proteases while still maintaining the activity [151].

To be noticed that also host immune response related proteins can have AMP-degrading activity [155].

Another process relevant for the neutralisation of AMPs at the extracellular environment level is the sequestration of the peptides: as an example, staphylokinase (one of the most prominent extracellular AMP-sequestering molecules [156, 157]) inactivates α -defensin binding to them and preventing their interaction with the designed target.

Biofilms Bacteria can resist AMPs by 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 [158]. Bacteria in a biofilm secretes an extracellular matrix with adhesion and protection functions. This matrix includes various compounds as cellulose, teichoic acids, proteins, lipids and nucleic acids [159] and confers resistance to antibiotics and AMPs, in some cases 1000 times as great as the one developed by bacteria in their planktonic form [160, 161]. This is achieved by repulsion and/or capture of AMPs by mainly 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 anionic dermcidin [162, 163]: deacetylation of PIA increases its positive net charge, thus repelling more efficiently cationic CAMPs, and increasing sequestration of the anionic AMP dermicidin at the same time, as well as forming a mechanical barrier for both of them [164]. In other cases, structural hindrance and electrostatic trapping prevent cationic AMPs to penetrate bacterial biofilms, while they are effective against their planktonic counterpart (as observed for polymyxin B, HNP-1, HBD-1, lactoferrin and protamine on *K. pneumoniae*, *S. pneumoniae* or *P. aeruginosa*)

[165, 166].

Moreover AMPs are nevertheless promising as alternatives to traditional antibiotics 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 as opposed to bacteriostatic ones which target fast-growing bacteria, as the majority of traditional antibiotics [167, 168]. Therefore, biofilm-intrinsic AMP resistance constitutes a great challenge in a field already depleted of efficient treatments [169, 170].

Surface remodelling As mentioned in the previous two paragraphs, the bacterial cell envelope environment constitutes a major impediment for AMPs activity. Even if a peptide reaches the bacterial envelope intact, bacteria can modify the characteristics of their surface to prevent its efficient action. Gram-positive and Gram-negative bacteria put in place different strategies to do that, according to their distinct cell envelopes. In particular, the target of such modifications are the teichoic acids (TA) in Gram-positive cell wall, and lipopolysaccharides (LPS) in the Gram-negative outer membrane. For example, D-Alanylation of TA, observed in *Staphylococcus*, adds a positive charge to it, reducing the attraction of cationic AMPs [171–173]. In turn, this increases the cell wall density, reducing the surface permeability [173]. Similarly, for Gram-negative bacteria (like *P. aeruginosa*), the LPS positive charge is increased by addition of different amine-containing molecules [174, 175] or by removing phosphate lipids, which have a negative charge, from lipid A, one of the constitutive moieties of LPS [176, 177].

Another target of AMPs in Gram-positive bacteria is the bacterial peptidoglycan precursor, lipid II, which has a key role in the formation of the cell wall. Many bacteria started using a modified version of it, the best known case being the replacement of its terminal D-alanine with D-lactate or D-serine [178] to avoid the action of the glycopeptide vancomycin. This molecule works binding to the D-Ala-D-Ala terminal moieties of the precursor, preventing cross linking of molecules between them and thus the cell wall synthesis [179].

Gram-negative bacteria can instead enhance the rigidity of the outer membrane to reduce permeability to AMPs via addition of extra acyl chains into lipid A [180, 181]. The long polysaccharide chain of LPS (called O-antigen) makes this class of bacteria particularly resilient to the action of AMPs [182] as

both the LPS core and the O-antigen were proven to promote AMP resistance (in *B. cenocepacia* and *Brucella abortus* [183, 184]).

Surface modification to counteract the AMPs activity occurs very often also 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 [185]. Usually the group added is a Lysine [186], but Alanine is commonly chosen as well [187].

The rigidity of the cytoplasmic membrane can be enhanced as well, by an increase in saturated acyl chains which has been proven to confer resistance [188, 189], though the precise mechanisms underlying the connection are still unclear.

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 [190].

1.3.4 Principles of AMP design

It was already highlighted in Section 1.3.2 that a classification of AMPs would provide knowledge on the characteristics a sequence must have to perform an antimicrobial function. At the present state of the art, several databases exist gathering AMPs and subclasses of them, like membrane active, biofilm active or haemolytic peptides [191?]. Based on the increasing amount of data, it is now possible to identify features which, comprehensively, discriminate AMPs with respect to non antimicrobial peptides:

- **Structure:** as mentioned before, 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 among AMPs, and their structural difference is reflected in the different mechanisms of actions (when known). When screening a peptide to

identify its AMP-likeness, it must be considered that some sequences may rearrange in proximity of the membrane, thus their structure in solution does not reflect their active conformation.

- **Charge:** AMPs are charged moieties. Usually they present positive charge (up to $\sim +10\text{ e}$), but there are examples of anionic ones [?]. Their potency is often related to the amount of charge each unit possesses, however an excessive charge may promote haemolytic activity as well. [?]
- **Hydrophobicity:** together with charged amino acids, AMPs contain also hydrophobic residues, usually with abundance of aromatic chains and specifically Tryptophan. Indeed membrane active peptides must insert into the lipid core of membranes, which is an hydrophobic environment, therefore having such residues help them in anchoring in the lipid tail region.
- **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. The usefulness of this segregation in the anchoring and penetrating mechanism has been explained in Section 1.2.2.
- **Solubility** AMPs need good solubility to prevent aggregation in the aqueous environment they float in before arriving to the membrane, as aggregation would most likely impede their optimal interaction with the membrane.
- **Sequence motifs:** a long debate exists on whether the effectiveness of AMPs is related to particular sequence motifs or only to the overall amino acid composition. Statistical methods are trying to extract relevant pattern from the databases available, however the details of the structure-activity relationship are still uncertain.

Data-driven knowledge based on the above features, make possible to predict whether an amino acid sequence is antimicrobial or not. Several online servers (APD3, dbAMP, DBAASP, antiBP2, amPEP [191–195]) are available to host curated databases and, based on statistical or machine learning methods elaborated on those, evaluate the antimicrobial properties of user provided

sequences: the output is generally a score of how likely the peptide is to have such function, with sometimes an optional indication on whether the activity might be stronger against Gram-positive or Gram-negative bacteria.

At present, it is still impossible to foresee the precise efficacy and mechanism of action of an AMP from its sequence only (while more information can be derived if a structure is present as well). The knowledge of structure-activity relationships for AMPs would be beneficial to find new, better performing ones. Indeed the design of new AMP sequences aims at improving some specific 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. Several methodological approaches to AMP design are possible, and they can be grouped in three main lines: template based studies, biophysical studies and virtual screening [196].

Template based studies The main idea behind template based methods consists in employing existing antimicrobial sequences and modifying them in the direction of the desired characteristics. The most widely explored templates are cecropin, magainin and protegrin for their short sequences and because their action and structure has been well characterised [197–200].

Ideally, an amino acid scanning of all the residues in an AMP provides information on the role of each of them thus prompting at the most suitable mutations. High-throughput methods allow nowadays for such thorough investigation in the case of short AMPs [201, 202]; similarly, a less resource consuming Alanine scanning points at the most critical residues for the antimicrobial activity, on which all the mutations can be tested [203–205].

In the absence of such resources, earlier studies focused on simpler approaches to enhance charge and amphiphilicity of the peptides, as these characteristics are deemed crucial in their effectiveness (see the paragraph above) [197–200]. Similarly, the addition of acyl moieties have been shown empirically to improve the performances of AMPs, as these can provide the necessary hydrophobic domains that, together with charged amino acids, allow an amphiphatic structure in short peptides [206–208].

However the above methods focus on the single amino acid level and can not take into account the interplay between residues, while the paired mutation of some of them at the same time can give optimal results with respect to a single intervention. Furthermore there is little to no information on the three dimensional structure of the mutated 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.

Only recently a structure based approach has been developed to integrate structural information on template based models to design peptides active against many bacterial lines at the same time [209]. Similarly, a template design method combining chemical and case specific structural information [210] has recently been designed, producing AMPs with improved selectivity for bacterial membranes. Starting from a synthetic broad spectrum AMP with high toxicity, the positioning of positively charged residues at the centre of the non polar face of the amphipathic α -helix reduced its haemolytic activity while improving its therapeutic index. This proves that charge and structure features do affect the antimicrobial activity, but again does not provide very generalised design rules.

A complementary approach consists in focussing on minimal antimicrobial blocks: several investigations proved the importance of single residues and their intercalated pattern in natural and designed AMPs. In particular, natural AMPs are rich in Tryptophan and Arginine residues [111], while synthetic ones have been produced with only Lysines-Leucine, or Arginine-Valine combinations to produce amphipathic helices [211]. Furthermore, polyarginine are long known for being cell penetrating peptides [212].

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 [213]. This approach can benefit of the improving size of peptide databases, together with the advancement in text mining technology and dedicated machine learning algorithms, bringing to the streamlined selection of promising sequences to investigate further [214].

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 by 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: for example, free energy perturbation allows to pinpoint the interactions stabilising a structure, while Molecular Dynamics simulations can show why a particular residue binds favourably to the membrane. Moreover, structural information is crucial to discriminate cases in which similar sequences behave differently due to the environment around them, as this information is necessarily lost in a sequence only approach.

The drawback of such techniques lays in their computational cost. All of them can approach systems with a limited size, and simulations can access short (microseconds) time scales 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*.

As such, the strength of biophysical studies does not lay in the power of analysing large datasets, but rather in the fact that, as they exploit the whole information available on a system (sequence, structure, chemistry), they can single out the interactions that are crucial for a mechanism, clarifying whether they are peculiar of a given local environment or they can be transferred elsewhere. In this respect, they provide a generalisable knowledge applicable to different systems and thus to the design of novel AMPs at the atomistic level.

An example of how MD simulations shed light on AMP-membrane interactions is given by the protegrine peptide: porcine protegrin is a β -hairpin AMP which is thought to act through pore formation. A model for the detailed steps of such mechanism was obtained by MD simulations, and proposed a non trivial process of electrostatic attraction to the anionic membrane, followed by dimerisation and subsequent insertion into the membrane. Inserted peptides finally form large aggregates that lead to transmembrane pores formation [215]. Further steps were taken for ovispirin [216], indolicidin [217] and temporin [218] sequences, designing and testing mutants of the original sequence with improved activity and decreased haemolytic activity on a computational basis, which was confirmed experimentally later on.

Virtual screening Contrary to biophysical assay, virtual screening methods are employed to analyse a large number of sequence, when an experimental or computational test of all of them would be prohibitive. The key concept of these methods consists in the identifications of some 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 aforementioned text-based method falls in this category, but it can be generalised to include more information than the sequence composition only, namely chemical or structural information.

These methods are witnessing new popularity due to the recent evolution of machine learning (ML) techniques: if originally they relied on regression methods, in the past three decades artificial neural network have been extensively applied to the problem (for a historically informed review see Table 1 in Ref. [196]). Machine learning appears particularly suitable to the task as the potency of AMPs is certainly determined by the combination of many factors, and it is difficult to properly weight them and identify the predominant ones in each context.

Therefore, ML algorithms are trained on a set of AMPs labelled by their potency, with each of them characterised by many different properties (features): sequence, partial charge, hydrophobicity, amphiphilicity and molecular weight are the most intuitive ones; but also experimental measures of pK (the logarithm of the dissociation constant), pI (the isoelectric point), nuclear magnetic resonance data, chromatographic indices retention time in a given

chromatography column, octanol-water partition fraction or circular dichroism data. Finally, other theoretically computed features as van der Waals surface area and hardness (the energy required to remove the outermost electron) can be considered as well. The more the input properties to consider, the more expensive is to train the model, but an higher accuracy can be meet: the power of such approach is indeed in the identification of relevant features traditionally overlooked. At the same time though, the output is likely providing complex descriptors (i.e. combinations of many features) of difficult interpretation. This is why the step of features selection is important and more than one model is trained to identify the minimum set which gives satisfactory agreement with the data [219, 220]. In principle, having complex descriptors is not a problem as, rather than guide the design of future AMPs from first principles, they can be used to scores all possible combinatorial sequences of the desired length to identify the best ones. In practise though, this can be difficult even for such an in silico screening because of their exponentially growing number (as a meter, all the possible amino acid combinations on a 10 residues sequence are of the order of 10^{11}). A possible solution is constituted by evolutionary search methods in the sequence space: analogously to an energy minimization process in space, single amino acid mutations are attempted and the fitness of the new sequence computed, based on the model generated via ML. The move is accepted if it proceeds toward an improved fitness, i.e. a maximum in the fitness landscape, and rejected otherwise.

Another 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: nowadays, high-throughput synthesis methods, together with surrogate measures of bacterial killing (such as lipid vesicle experiments [221] or the diminished energy dependent luminescence of bacteria constitutively expressing luciferase 90, rather than Minimum Inhibitory Concentration assays [222]), allow for quick screening of many of those. This procedure was employed by Cherkasov et al. [223], as they 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 proved subsequently by experiments.

Overall, peptide design has proven successful in producing sequences with improved potency or selectivity. However, it is still a case-dependent proce-

dure, rather than a general, automated protocol easily applicable to enhance any sequence of choice.

1.3.5 Clinical applications

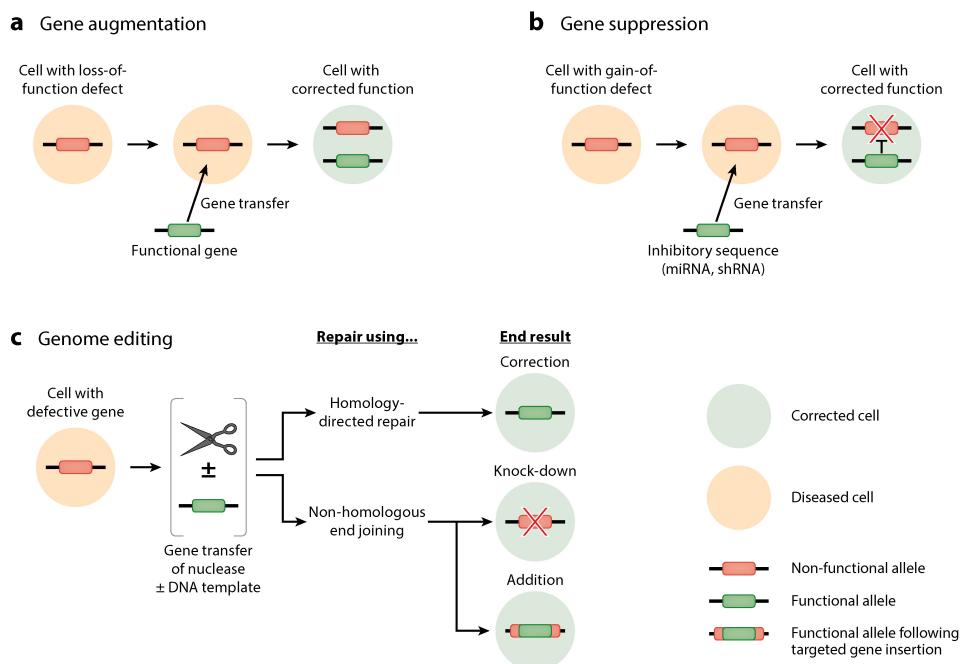
Antimicrobial peptides have been studies for many years, however the push to capitalise them to get compounds viable for the clinical stage has been delayed by many factors, including production costs, and lack of interest in the face of more potent small molecules which were deemed more economically 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 [225].

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 [98]. For the last reason in particular, many of them in trial 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 [226, 227]. 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.

1.4 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. Briefly mentioned in Section 1.1.3 when introducing Adeno Associated Viruses, in recent years it has greatly



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

Figure 1.5: Principles of gene therapy. Reproduced from [228].

evolved and gained attention for the treatment of tumours, genetic diseases and complex acquired disorder [228].

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: augmentation gene therapy introduces an healthy gene copy to restore the normal functionalities of the protein of interest and thus of the cell; suppression gene therapy suppresses a detrimental gene (this is particularly useful in the case of cancer, to impede cancer cells replication); gene-editing, the most recent advance in the field, overlooks the possibility of correcting base pairs mutations to restore the original healthy sequence (Figure 1.5).

For the first strategy mentioned above, the therapy itself usually consists in the delivery of a DNA strand, which in turn can be internalised in the genome and thus spread when the cell replicates, or not internalised and thus can influence the functionalities of that particular cell only. The first approach is used mainly for ex vivo administration (in cultured cells taken from the pa-

tient that are subsequently transplanted back), while the second for in vivo one (direct injection into the patient). The second strategy, gene repression, employs RNA interference and in vivo therapy, therefore aims at delivering miRNA (microRNA) or siRNA (small interfering RNA) strands which repress the transcription of the problematic RNA sequence. Finally, gene-editing is often done through the functionalisation of the CRISPR-Cas9 technology, a mechanism found in prokaryotic organism as bacteria and archea as defence against viruses [229]. CRISPS (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. In doing so, it 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 [230, 231], thus performing a therapeutic role. More complex strategies are possible combining gene and drug therapy, such as the delivery of suicide genes to increase the sensitivity of tumour cells to cytotoxic drugs [232], or the use of oncolytic viruses (OVs) that selectively replicate in tumour cells only, disrupting them [233].

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 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 [234, 235].

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 [228], with many more undergoing regulatory review.

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 delivery genetic material into human cells.

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 1.3.2), 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



Figure 1.6: Capzip molecule scheme and bond representation. [TO BE IMPROVED] Adapted from [8].

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 interest 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 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 1.6). 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 [236–240], 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 [241]; 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 [241]; finally it is implicated in the stimulation of different immunological cells (killer cells [242], polymorphonuclear leukocytes, and macrophages [243]). 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 [134, 135, 244, 245]. 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 [246].

The bovine homolog of lactoferricin (LfcinB) has a higher bactericidal potency than human lactoferricin on several bacteria [247] 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 [247]. Further experiments on LfcinB subsequences identified an even shorter antimicrobial core, constituted by the six amino acids RRWQWR [137]. 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 [247] (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, resulting very flexible [136]. Then, the binding of its antimicrobial core to sodium dodecyl-sulfate micelles was studied [137], 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 [138], suggesting a selective antimicrobial action on anionic membranes. Additional experiments have been performed on the full sequence or mutated subsequences [248, 249] 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 [250]. The results suggested a binding function for the Tryptophan residues, in line with one of the roles Tryptophan can assume in antimicrobial peptides [111]. 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 1.6, to obtain a self-assembling molecule in a three dimensional shape, hosting multiple copies of an actively antibacterial sequence.

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

Figure 1.7: ... Reproduced from [8]

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 1.7), 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 1.7). 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 1.4, 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 1.7) 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) [251]. 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 1.7). 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 1.3.4 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 1.1.3). 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 blocks 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 descrip-

tion 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 the components of a system and their mutual interactions, so that the natural evolution of the system is obtained simply implementing the laws of physics. 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, a classical approximation will be less and less relevant, as classical mechanics represents well the evolution of

large systems, for which in turn some atomistic 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.

Later on in this chapter will be discussed 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 [?], which identifies in these four issues the interpretative key with which MD simulations must be designed, run and interpreted.

2.1 The force field problem

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 this sets the ground for the approximations to follow: as much accurately the model of the system can be built, in a 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 MD algorithms

Specifically, 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 sequations of motion represent a peculiar subset for which it is proved that an analytical solution exists. This allows to derive some properties of the motion even when the solution itself is too complicated to compute: for example often the positions visited by the particle can be obtained, albeit the exact time instant at which they are explored is not known [?].

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 component of the system. Such forces derive from chemical bonds, electrostatic interactions, and Pauli repulsion, 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 force to be integrated during each time step: one possibility is to take its value at time t_0 as in 2.4, but another legitimate choice is given by its value at time $t_0 + \Delta t/2$. Based on this is the leap-frog algorithm::

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

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

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, included GROMACS (which however has additional options).

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 Functional form of force-fields

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 GROMOS96 force field [?] as implemented in the GROMACS MD engine [?], explaining what each term represents. Other force fields can have slightly different implementations, or can 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 or analogous.

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 molec-

ular orbital theory, quantum mechanics calculations, or fitting some macroscopic quantities as the free energy of solvation. The GROMOS force field is based on the latter, while others like CHARMM [?] and AMBER [?] use a quantum mechanics approach. In both cases, the parametrisation procedure is usually performed for small moieties, 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.

Specifically, in the GROMOS96 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.

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 arrangements. 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 through a periodic potential:

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

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 the proper dihedral corresponds to the *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 for ease of 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:

$$V_C(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_i j r} \quad (2.15)$$

with ϵ_0 the dielectric constant of vacuum and ϵ_r the relative dielectric constant, 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 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 [?] technique, or the Reaction Field [?] approach; for the former, the GROMOS MD engine even dedicates a fraction of the overall computer resources to exclusively perform the PME calculations.

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 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: 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 force field, as more convoluted descriptions would be needed to properly take into account the change in behaviour due to temperature.

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 indeed a common procedure to reduce the number of degrees of freedom to sample, which allows for a quicker exploration of the system (see Section 2.2). What this specifically corresponds to would be explained for each of the force fields considered.

2.1.3 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 CO groups, CH, CH₂, and CH₃ ones (see Figure –). Each one of these last three is modelled as a unique atom, having mass equal to the mass of the carbon plus the masses of the hydrogens bonded to it. As such, they are treated as three different carbon atom types, in turn different from the carbon type used to model the “bare” atom in the CO group.

The parametrisation of this 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, and no electrons are included for the sake of efficiency, their redistribution across atoms which are inter-bonded is modelled through fractional charges assigned to each atom (while the total charge 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) description. Intuitively, this represents water as a three atoms molecule, placing a complementary charge on the oxygen (negative) and the two hydrogen (positive), and allowing for bonds vibration. This model is able to reproduce correctly the density and dielectric permittivity of water. Water has a dedicated algorithms for the renormalisa-

tion of its bond lengths (SETTLE rather than LINCS) - a procedure which take place after each MD step for all the bonded interaction: SETTLE can analytically solve this constrained problem for 3 degrees of freedom (as the ones present in every water molecule), while for molecules with more atoms one must adopt approximate methods as the LINCS algorithm.

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 [?] for the set of simulations involving peptidic assembly in solution, while we switched to 54a8 [?] for the simulations involving biological membranes. While it is advisable to have a coherent set of parameters across simulation, to compare their outcome in a consistent manner, we deemed the 54a8 parameter set more suitable for lipid simulations because of the improvements 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 them, and the ones involving lipids as well.

2.1.4 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 form experimental data - as it is performed for example in the GROMOS atomistic force field parametrisation.

The coarse-grain force field SIRAH [?] 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 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_αH 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. Such dual granularity approach is based on physico-chemical intuition, and is more difficult to generalise than a uniform granularity one. Never the less, the force field has been recently extended to lipids, while it comprised a parametrisation for DNA molecule 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. This feature is responsible for hydrogen bonds formation and for the electrostatic screening observed in an aqueous solution. The second aspect 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 to other the other two. Moreover, the geometrical arrangement reproduce 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 the secondary structures they adopt; simulations of lipids randomly oriented in water showed the formation of an organised

bilayer; and finally the force field reproduced the expected behaviour of a few selected transmembrane proteins in model membranes. More details on the systems simulated will be given in Section 2.5, where simulations of assembling peptides and antimicrobial ones will be reviewed in function of the work presented in this thesis.

2.1.5 The MARTINI force field

The MARTINI force field is another very popular description for biological molecules. It was developed much to the SIRAH force field, and since its first description, it has been refined and extended to include lipids (its initial focus), proteins, small ligands and DNA/RNA molecules.

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 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 to represent accurately the chemical nature of the different underlying atomistic structures. The advantage of this systematic approach is its transferability: the 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, so that there is no need for 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 such chemical groups. Similarly, the MARTINI force field chooses this top-down approach to parametrise the 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 one 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 an amino acid 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 [?]), together with the standard bonded interactions. Both approaches 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 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 a unique 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 of 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: 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 as 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.

2.2 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 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 conformation visited by the system 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 it is important to investigate 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 they can be obtained by a large number of microstates,

i.e. many different rearrangement of the internal degrees of freedom give the same macroscopic outcome (or, more technically, they have an 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 [? ?]);
- building a system configuration from the configurations of its fragments in a stepwise manner (for example in the Monte Carlo chain-growing methods [? ?]);
- using step methods, where a new configuration is derived from the previous one. Energy minimization, Metropolis Monte Carlo, Molecular Dynamics[?] are all step methods. For MD in particular, the step is intuitively associated to the time.

MD simulations are 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.

Enhanced MD will not be the focus of this work, but it is interesting to understand the flexibility of the technique in facing the search problem. Many

approaches are possible and include smoothing and deforming the potential energy surface, enhancing the pace at which the space is explored or 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. NOE data) [?]; softening geometric restraints derived from NMR or X-ray data through time averaging [? ?]; or finally using “soft-core” atoms, thus reducing the Pauli repulsion among them [?]. The enhanced exploration pace is obtained using higher temperatures to overcome energy barriers thanks to the acquired kinetic energy [?], or scaling the mass to reduce inertia [?], as well as by combining multiple simulations together (for example in replica-exchange algorithms [?] some configuration are extracted from simulations held at different conditions and used to feed a new set of simulations). Then, avoiding the re-sampling of energy minima can be reached through local potential-energy elevation [? ?]; while constraining the high-frequency degrees of freedom (for example non-polar hydrogens) avoids spending time computing non interesting fine-details [90].

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 2.1.2): 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 hypothesis.

Thus, the outcome of MD simulation is a (local and incomplete) sampling of the configuration space. If 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 knowl-

edge is coded in the energy functional: for example, only a few rotamers of the common amino acids are favoured in the force field commonly employed, according to the informations gained from X-ray crystallography, thus avoiding the sampling of high-energy, unfavourable conformations.

2.3 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 energy term, others are governed by changes in entropy. 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, entropy plays a relevant role when considering the state of the solvent: being constituted by many molecules, its entropic contribution can be significant. When considering the fold 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 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 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, they are hardly applicable to the calculation of ligand-protein binding entropy or polypeptide folding entropy [?]. 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.

2.4 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, so that there is uncertainty in the weight each conformation is assigned, as well as the possibility that some conformations are missing. 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 biomolecular systems, 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; 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.

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 self-assembling

ones, as well as on exploratory simulations performed as an aid for peptide design.

2.5 MD simulations: successes

2.5.1 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 their mechanisms of action are not completely understood in most of the cases (see Section 1.3.2). As it is proven by experiments that the change of even one residue in short AMPs can change remarkably the antimicrobial activity (see Section 1.3.4), it is clear that their action is governed by subtle atomic interactions, and MD simulations are a suitable tool for understanding this aspect.

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), but the choice of the system to simulate is crucial as well.

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 description, the different steps are usually investigated separately, based on prior hypothesis: for example, the peptide can be positioned closed to the membrane surface with the correct face (if known or based on energetic assumptions), or directly into the membrane with different insertion depths and tilt angles to verify the most disruptive configurations. Therefore, the full process can be reconstructed from a “stepwise” knowledge combining the different states and sampling intermediate regions if necessary.

The second important choice concerns the model of the membrane to simulate. In an effort to keep complexity low, bacterial and mammal membrane have been modelled with a minimal number of lipids. Very often, the only key characteristic retained is a negative charge for models of bacterial membrane,

with around 25% of the lipids presenting a $-1e$ charge and the rest being zwitterionic (i.e. overall neutral but with a positively and a negatively charged region separate in space, see Chapter [4 lipid parametrisation]); while for model mammal membranes only zwitterionic lipids are employed, and cholesterol can be included as well, as it is considered important in allowing the flexibility typical of mammal membranes. Because of their simplicity, very similar or identical systems are used also in experiments, where the use of cells is prevented by the experimental conditions necessary. For example... . Therefore, even if they don't model accurately the structure of cellular membranes, simulations and experiments of these systems can provide a first explanation of the antimicrobial activity, with the two techniques complementing and validating each other.

Never the less, attempts to model more accurately cell membrane, and especially the bacterial inner membrane, bacterial wall, and the combination of the two have been pursued. This can be efficiently performed especially at the coarse grain level, as the inclusion of all the elements of the cell membranes result in quite large systems.

It has been proposed that AMPs act to kill bacteria through a series of steps: attraction, aggregation, penetration, and lysis (14). Simulations of this process would provide the ultimate level of detail in understanding the mechanism of action of AMPs; however, the timescales on which these processes occur exceed what is currently possible with all-atom MD simulations. Formation of pores has been observed in coarse-grained MD simulations (15,16), but such simulations do not provide the same resolution as atomistic approaches, and may not accurately reflect system properties and behavior. To accommodate the current limitations on computational power, we can explore various steps of this process individually (Fig. 17.2)

Many of the simulations of AMP, especially in the first stages of the development of the technique, focussed on short helical peptides

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. URL https://www.mt.com/au/en/home/applications/L1_AutoChem_Applications/L2_ReactionAnalysis/L2_Polymerization.html#publications.
 - [2] Wikipedia. Liposome. URL <https://en.wikipedia.org/wiki/Liposome>.
 - [3] Schoonen, L. & van Hest, J. C. M. Functionalization of protein-based nanocages for drug delivery applications. *Nanoscale* **6**, 7124–7141 (2014). URL <http://xlink.rsc.org/?DOI=C4NR00915K>.
 - [4] A Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & V Piddock, L. J. Molecular mechanisms of antibiotic resistance. *Nature Publishing Group* **13** (2014). URL <https://www.nature.com/nrmicro/journal/v13/n1/pdf/nrmicro3380.pdf>.
 - [5] Kim, J. Phage as a therapeutic agent. URL <https://medium.com/@thryve/phage-as-a-therapeutic-agent-ed4c466302e5>.
 - [6] Torres, M. D., Sothiselvam, S., Lu, T. K. & de la Fuente-Nunez, C. Peptide Design Principles for Antimicrobial Applications. *Journal of Molecular Biology* (2019). URL <https://www.sciencedirect.com/science/article/pii/S0022283618312890>.
 - [7] Nguyen, L. T., Haney, E. F. & Vogel, H. J. The expanding scope of antimicrobial peptide structures and their modes of action. *Trends in Biotechnology* **29**, 464–472 (2011). URL <https://www.sciencedirect.com/science/article/pii/S0167779911000886?via%23Dihub>.
-

- [8] Castelletto, V. *et al.* Structurally plastic peptide capsules for synthetic antimicrobial viruses. *Chem. Sci.* **7**, 1707–1711 (2016). URL <http://xlink.rsc.org/?DOI=C5SC03260A>.
- [9] Masaoka, Y., Tanaka, Y., Kataoka, M., Sakuma, S. & Yamashita, S. Site of drug absorption after oral administration: Assessment of membrane permeability and luminal concentration of drugs in each segment of gastrointestinal tract. *European Journal of Pharmaceutical Sciences* **29**, 240–250 (2006). URL <https://www.sciencedirect.com/science/article/pii/S0928098706001709?via%23Dihub>.
- [10] Mitragotri, S., Burke, P. A. & Langer, R. Overcoming the challenges in administering biopharmaceuticals: formulation and delivery strategies. *Nature Reviews Drug Discovery* **13**, 655–672 (2014). URL <http://www.nature.com/articles/nrd4363>.
- [11] Krol, S. Challenges in drug delivery to the brain: Nature is against us. *Journal of Controlled Release* **164**, 145–155 (2012). URL <https://www.sciencedirect.com/science/article/pii/S0168365912003999>.
- [12] Pattani, B. S. & Torchilin, V. P. Targeted Drug Delivery Systems: Strategies and Challenges. 3–38 (Springer, Cham, 2015). URL http://link.springer.com/10.1007/978-3-319-11355-5{_}1.
- [13] Jain, S. & Edwards, M. J. Advances and challenges in the development of drug delivery systems - A European perspective (2016). URL <https://www.semanticscholar.org/paper/Advances-and-challenges-in-the-development-of-drug-Jain-Edwards/51a16060d641078aac675e3a9c2ab8e7b1effbfa>.
- [14] Hughes, G. A. Nanostructure-mediated drug delivery. *Nanomedicine: Nanotechnology, Biology and Medicine* **1**, 22–30 (2005). URL <https://www.sciencedirect.com/science/article/pii/S1549963405000122?via%23Dihub>.
- [15] Singh, P. *et al.* Gold Nanoparticles in Diagnostics and Therapeutics for Human Cancer. *International journal of molecular sciences* **19** (2018). URL <http://www.ncbi.nlm.nih.gov/pubmed/29986450>.

- [16] Boisselier, E. & Astruc, D. Gold nanoparticles in nanomedicine: preparations, imaging, diagnostics, therapies and toxicity. *Chemical Society Reviews* **38**, 1759 (2009). URL <http://xlink.rsc.org/?DOI=b806051g>.
- [17] Erol, O. *et al.* Recent Advances in Bioactive 1D and 2D Carbon Nanomaterials for Biomedical Applications. *Nanomedicine: Nanotechnology, Biology and Medicine* (2017). URL <http://linkinghub.elsevier.com/retrieve/pii/S1549963417300898>.
- [18] Depan, D., Shah, J. & Misra, R. Controlled release of drug from folate-decorated and graphene mediated drug delivery system: Synthesis, loading efficiency, and drug release response. *Materials Science and Engineering: C* **31**, 1305–1312 (2011). URL <https://www.sciencedirect.com/science/article/pii/S0928493111001159>.
- [19] Lammers, T. *et al.* Simultaneous delivery of doxorubicin and gemcitabine to tumors in vivo using prototypic polymeric drug carriers. *Biomaterials* **30**, 3466–3475 (2009). URL <https://linkinghub.elsevier.com/retrieve/pii/S0142961209002324>.
- [20] Liechty, W. B., Kryscio, D. R., Slaughter, B. V. & Peppas, N. A. Polymers for drug delivery systems. *Annual review of chemical and biomolecular engineering* **1**, 149–73 (2010). URL <http://www.ncbi.nlm.nih.gov/pubmed/22432577>.
- [21] Kawakatsu, T. *Statistical Physics of Polymers : an Introduction* (Springer Berlin Heidelberg, 2004).
- [22] Nicolas, J., Mura, S., Brambilla, D., Mackiewicz, N. & Couvreur, P. Design, functionalization strategies and biomedical applications of targeted biodegradable/biocompatible polymer-based nanocarriers for drug delivery. *Chem. Soc. Rev.* **42**, 1147–1235 (2013). URL <http://xlink.rsc.org/?DOI=C2CS35265F>.
- [23] Nair, L. S. & Laurencin, C. T. Biodegradable polymers as biomaterials. *Progress in Polymer Science* **32**, 762–798 (2007). URL <https://www.sciencedirect.com/science/article/pii/S0079670007000664>.

- [24] Rao, S. H., Harini, B., Shadamarshan, R. P. K., Balagangadhara, K. & Selvamurugan, N. Natural and synthetic polymers/bioceramics/bioactive compounds-mediated cell signalling in bone tissue engineering. *International journal of biological macromolecules* **110**, 88–96 (2018). URL <http://www.ncbi.nlm.nih.gov/pubmed/28917940>.
- [25] Yoo, J.-W., Irvine, D. J., Discher, D. E. & Mitragotri, S. Bio-inspired, bioengineered and biomimetic drug delivery carriers. *Nature Reviews Drug Discovery* **10**, 521–535 (2011). URL <http://www.nature.com/articles/nrd3499>.
- [26] Yingchoncharoen, P., Kalinowski, D. S. & Richardson, D. R. Lipid-Based Drug Delivery Systems in Cancer Therapy: What Is Available and What Is Yet to Come. *Pharmacological reviews* **68**, 701–87 (2016). URL <http://www.ncbi.nlm.nih.gov/pubmed/27363439>.
- [27] Bunker, A., Magarkar, A. & Viitala, T. Rational design of liposomal drug delivery systems, a review: Combined experimental and computational studies of lipid membranes, liposomes and their PEGylation. *BBA - Biomembranes* **1858**, 2334–2352 (2016). URL <http://dx.doi.org/10.1016/j.bbamem.2016.02.025>.
- [28] Pattni, B. S., Chupin, V. V. & Torchilin, V. P. New Developments in Liposomal Drug Delivery. *Chemical Reviews* **115**, 10938–10966 (2015). URL <http://pubs.acs.org/doi/10.1021/acs.chemrev.5b00046>.
- [29] Jain, S. & Pillai, J. Bacterial membrane vesicles as novel nanosystems for drug delivery. *International journal of nanomedicine* **12**, 6329–6341 (2017). URL <http://www.ncbi.nlm.nih.gov/pubmed/28919737>.
- [30] Linko, V., Ora, A. & Kostiainen, M. A. DNA Nanostructures as Smart Drug-Delivery Vehicles and Molecular Devices. *Trends in Biotechnology* **33**, 586–594 (2015). URL <https://www.sciencedirect.com/science/article/pii/S0167779915001614>.
- [31] Douglas, S. M., Bachelet, I. & Church, G. M. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* **335**, 831–

- 834 (2012). URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1214081>.
- [32] Zhang, Q. *et al.* DNA Origami as an *< i > In Vivo </ i >* Drug Delivery Vehicle for Cancer Therapy. *ACS Nano* **8**, 6633–6643 (2014). URL <http://pubs.acs.org/doi/10.1021/nn502058j>.
- [33] Jiang, Q. *et al.* DNA Origami as a Carrier for Circumvention of Drug Resistance. *Journal of the American Chemical Society* **134**, 13396–13403 (2012). URL <http://pubs.acs.org/doi/10.1021/ja304263n>.
- [34] Lobo, F. P. *et al.* Virus-Host Coevolution: Common Patterns of Nucleotide Motif Usage in Flaviviridae and Their Hosts. *PLoS ONE* **4**, e6282 (2009). URL <http://dx.plos.org/10.1371/journal.pone.0006282>.
- [35] Lauer, K. B., Borrow, R. & Blanchard, T. J. Multivalent and Multipathogen Viral Vector Vaccines. *Clinical and Vaccine Immunology* **24**, e00298–16 (2017). URL <https://cvi.asm.org/content/24/1/e00298-16>.
- [36] Daya, S. & Berns, K. I. Gene therapy using adeno-associated virus vectors. *Clinical microbiology reviews* **21**, 583–93 (2008). URL <http://www.ncbi.nlm.nih.gov/pubmed/18854481>.
- [37] Büning, H. & Schmidt, M. Adeno-associated Vector Toxicity-To Be or Not to Be? *Molecular therapy : the journal of the American Society of Gene Therapy* **23**, 1673–1675 (2015). URL <http://www.ncbi.nlm.nih.gov/pubmed/26606658>.
- [38] Smalley, E. First AAV gene therapy poised for landmark approval. *Nature Biotechnology* **35**, 998–999 (2017). URL <http://www.nature.com/articles/nbt1117-998>.
- [39] Wu, W., Hsiao, S. C., Carrico, Z. M. & Francis, M. B. Genome-free viral capsids as multivalent carriers for taxol delivery. *Angewandte Chemie (International ed. in English)* **48**, 9493–7 (2009). URL <http://www.ncbi.nlm.nih.gov/pubmed/19921725>.

- [40] Ma, Y., Nolte, R. J. & Cornelissen, J. J. Virus-based nanocarriers for drug delivery. *Advanced Drug Delivery Reviews* **64**, 811–825 (2012). URL <https://www.sciencedirect.com/science/article/pii/S0169409X12000087>.
- [41] Fan, T., Yu, X., Shen, B. & Sun, L. Peptide Self-Assembled Nanostructures for Drug Delivery Applications. *Journal of Nanomaterials* **2017**, 1–16 (2017). URL <https://www.hindawi.com/journals/jnm/2017/4562474/>.
- [42] Habibi, N., Kamaly, N., Memic, A. & Shafiee, H. Self-assembled peptide-based nanostructures: Smart nanomaterials toward targeted drug delivery. *Nano Today* **11**, 41–60 (2016). URL <https://www.sciencedirect.com/science/article/pii/S1748013216300081>.
- [43] Yan, X., Zhu, P. & Li, J. Self-assembly and application of diphenylalanine-based nanostructures. *Chemical Society Reviews* **39**, 1877 (2010). URL <http://xlink.rsc.org/?DOI=b915765b>.
- [44] Silva, R. F., Araújo, D. R., Silva, E. R., Ando, R. A. & Alves, W. A. l-Diphenylalanine Microtubes As a Potential Drug-Delivery System: Characterization, Release Kinetics, and Cytotoxicity. *Langmuir* **29**, 10205–10212 (2013). URL <http://pubs.acs.org/doi/10.1021/la4019162>.
- [45] King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510** (2014). URL <https://www.nature.com/nature/journal/v510/n7503/full/nature13404.html>.
- [46] Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Research*, 28: 235-242 (2000). URL <http://www.rcsb.org/>.
- [47] Yeates, T. O. Protein assembles into Archimedean geometry. *Nature* **569**, 340–342 (2019). URL <http://www.nature.com/articles/d41586-019-01407-z>.
- [48] Malay, A. D. *et al.* An ultra-stable gold-coordinated protein cage displaying reversible assembly. *Nature* **569**, 438–442 (2019). URL <http://www.nature.com/articles/s41586-019-1185-4>.

- [49] Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery* **16**, 19–34 (2017). URL <http://www.nature.com/articles/nrd.2016.230>.
- [50] Gordon, E., Mouz, N., Duée, E. & Dideberg, O. 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* **299**, 477–485 (2000). URL <https://www.sciencedirect.com/science/article/pii/S0022283600937409?via%23Dihub>.
- [51] Kapoor, G., Saigal, S. & Elongavan, A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of anaesthesia, clinical pharmacology* **33**, 300–305 (2017). URL <http://www.ncbi.nlm.nih.gov/pubmed/29109626>.
- [52] Birkegård, A. C., Halasa, T., Toft, N., Folkesson, A. & Græsbøll, K. Send more data: a systematic review of mathematical models of antimicrobial resistance. *Antimicrobial resistance and infection control* **7**, 117 (2018). URL <http://www.ncbi.nlm.nih.gov/pubmed/30288257>.
- [53] Niewiadomska, A. M. *et al.* Population-level mathematical modeling of antimicrobial resistance: a systematic review. *BMC Medicine* **17**, 81 (2019). URL <https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1314-9>.
- [54] O ’neill, J. TACKLING DRUG-RESISTANT INFECTIONS GLOBALLY: FINAL REPORT AND RECOMMENDATIONS THE REVIEW ON ANTIMICROBIAL RESISTANCE (2016). URL https://amr-review.org/sites/default/files/160525{_}Finalpaper{_}withcover.pdf.
- [55] Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1794**, 808–816 (2009). URL <https://www.sciencedirect.com/science/article/pii/S1570963908003592?via%23Dihub>.
- [56] Vargiu, A. V. & Nikaido, H. 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*

- 109**, 20637–42 (2012). URL <http://www.ncbi.nlm.nih.gov/pubmed/23175790>.
- [57] Kojima, S. & Nikaido, H. Permeation rates of penicillins indicate that *Escherichia coli* porins function principally as nonspecific channels. *Proceedings of the National Academy of Sciences* **110**, E2629–E2634 (2013). URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1310333110>.
- [58] Tran, Q.-T., Williams, S., Farid, R., Erdemli, G. & Pearlstein, R. The translocation kinetics of antibiotics through porin OmpC: Insights from structure-based solvation mapping using WaterMap. *Proteins: Structure, Function, and Bioinformatics* **81**, 291–299 (2013). URL <http://doi.wiley.com/10.1002/prot.24185>.
- [59] Tamber, S. & Hancock, R. E. W. On the mechanism of solute uptake in *Pseudomonas*. *Frontiers in bioscience : a journal and virtual library* **8**, s472–83 (2003). URL <http://www.ncbi.nlm.nih.gov/pubmed/12700103>.
- [60] Baroud, M. *et al.* Underlying mechanisms of carbapenem resistance in extended-spectrum β -lactamase-producing *Klebsiella pneumoniae* and *Escherichia coli* isolates at a tertiary care centre in Lebanon: role of OXA-48 and NDM-1 carbapenemases. *International Journal of Antimicrobial Agents* **41**, 75–79 (2013). URL <https://linkinghub.elsevier.com/retrieve/pii/S0924857912003470>.
- [61] Lavigne, J.-P. *et al.* An adaptive response of *Enterobacter aerogenes* to imipenem: regulation of porin balance in clinical isolates. *International Journal of Antimicrobial Agents* **41**, 130–136 (2013). URL <https://linkinghub.elsevier.com/retrieve/pii/S0924857912004219>.
- [62] Poulou, A. *et al.* 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* **51**, 3176–3182 (2013). URL <http://jcm.asm.org/cgi/doi/10.1128/JCM.01244-13>.
- [63] Wozniak, A. *et al.* Porin alterations present in non-carbapenemase-producing *Enterobacteriaceae* with high and intermediate levels of car-

- bapenem resistance in Chile. *Journal of Medical Microbiology* **61**, 1270–1279 (2012). URL <http://jmm.microbiologystreasearch.org/content/journal/jmm/10.1099/jmm.0.045799-0>.
- [64] Novais, Â. *et al.* Spread of an OmpK36-modified ST15 Klebsiella pneumoniae variant during an outbreak involving multiple carbapenem-resistant Enterobacteriaceae species and clones. *European Journal of Clinical Microbiology & Infectious Diseases* **31**, 3057–3063 (2012). URL <http://www.ncbi.nlm.nih.gov/pubmed/22706513><http://link.springer.com/10.1007/s10096-012-1665-z>.
- [65] Tangden, T., Adler, M., Cars, O., Sandegren, L. & Lowdin, E. Frequent emergence of porin-deficient subpopulations with reduced carbapenem susceptibility in ESBL-producing Escherichia coli during exposure to ertapenem in an in vitro pharmacokinetic model. *Journal of Antimicrobial Chemotherapy* **68**, 1319–1326 (2013). URL <http://www.ncbi.nlm.nih.gov/pubmed/23478794>.
- [66] Floyd, J. L., Smith, K. P., Kumar, S. H., Floyd, J. T. & Varela, M. F. LmrS Is a Multidrug Efflux Pump of the Major Facilitator Superfamily from *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* **54**, 5406–5412 (2010). URL <http://aac.asm.org/cgi/doi/10.1128/AAC.00580-10>.
- [67] Hu, R.-M., Liao, S.-T., Huang, C.-C., Huang, Y.-W. & Yang, T.-C. An Inducible Fusaric Acid Tripartite Efflux Pump Contributes to the Fusaric Acid Resistance in *Stenotrophomonas maltophilia*. *PLoS ONE* **7**, e51053 (2012). URL <https://dx.plos.org/10.1371/journal.pone.0051053>.
- [68] Kim, C. *et al.* The Mechanism of Heterogeneous Beta-Lactam Resistance in MRSA: Key Role of the Stringent Stress Response. *PLoS ONE* **8**, e82814 (2013). URL <http://dx.plos.org/10.1371/journal.pone.0082814>.
- [69] Ogawa, W., Onishi, M., Ni, R., Tsuchiya, T. & Kuroda, T. Functional study of the novel multidrug efflux pump KexD from *Klebsiella pneumoniae*. *Gene* **498**, 177–182 (2012). URL <https://linkinghub.elsevier.com/retrieve/pii/S0378111912002107>.

- [70] Dolejska, M., Villa, L., Poirel, L., Nordmann, P. & Carattoli, A. Complete sequencing of an IncH1 plasmid encoding the carbapenemase NDM-1, the ArmA 16S RNA methylase and a resistance-nodulation-cell division/multidrug efflux pump. *Journal of Antimicrobial Chemotherapy* **68**, 34–39 (2013). URL <https://academic.oup.com/jac/article/68/1/34/672511>.
- [71] Abouzeed, Y. M., Baucheron, S. & Cloeckaert, A. ramR mutations involved in efflux-mediated multidrug resistance in *Salmonella enterica* serovar Typhimurium. *Antimicrobial agents and chemotherapy* **52**, 2428–34 (2008). URL <http://www.ncbi.nlm.nih.gov/pubmed/18443112>.
- [72] Baucheron, S. *et al.* Bile-mediated activation of the acrAB and tolC multidrug efflux genes occurs mainly through transcriptional derepression of ramA in *Salmonella enterica* serovar Typhimurium. *Journal of Antimicrobial Chemotherapy* **69**, 2400–2406 (2014). URL <http://www.ncbi.nlm.nih.gov/pubmed/24816212>.
- [73] Nikaido, E., Shirokawa, I., Yamaguchi, A. & Nishino, K. Regulation of the AcrAB multidrug efflux pump in *Salmonella enterica* serovar Typhimurium in response to indole and paraquat. *Microbiology* **157**, 648–655 (2011). URL <http://mic.microbiologystrearch.org/content/journal/micro/10.1099/mic.0.045757-0>.
- [74] Hirakawa, H., Inazumi, Y., Masaki, T., Hirata, T. & Yamaguchi, A. Indole induces the expression of multidrug exporter genes in *Escherichia coli*. *Molecular Microbiology* **55**, 1113–1126 (2004). URL <http://doi.wiley.com/10.1111/j.1365-2958.2004.04449.x>.
- [75] Billal, D. S., Feng, J., Leprohon, P., Légaré, D. & Ouellette, M. Whole genome analysis of linezolid resistance in *Streptococcus pneumoniae* reveals resistance and compensatory mutations. *BMC Genomics* **12**, 512 (2011). URL <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2164-12-512>.
- [76] Gao, W. *et al.* Two Novel Point Mutations in Clinical *Staphylococcus aureus* Reduce Linezolid Susceptibility and Switch on the Stringent Response to Promote Persistent Infection. *PLoS Pathogens* **6**, e1000944 (2010). URL <http://dx.plos.org/10.1371/journal.ppat.1000944>.

- [77] Shore, A. C. *et al.* Detection of Staphylococcal Cassette Chromosome *< i>mec</i>* Type XI Carrying Highly Divergent *< i>mecA</i>*, *< i>mecI</i>* , *< i>mecR1</i>* , *< i>blaZ</i>* , and *< i>ccr</i>* Genes in Human Clinical Isolates of Clonal Complex 130 Methicillin-Resistant *< i>Staphylococcus aureus</i>*. *Antimicrobial Agents and Chemotherapy* **55**, 3765–3773 (2011). URL <http://www.ncbi.nlm.nih.gov/pubmed/21636525>.
- [78] Katayama, Y., Ito, T. & Hiramatsu, K. A new class of genetic element, staphylococcus cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy* **44**, 1549–55 (2000). URL <http://www.ncbi.nlm.nih.gov/pubmed/10817707>.
- [79] Kumar, N. *et al.* Crystal structure of the transcriptional regulator Rv1219c of *< i>Mycobacterium tuberculosis</i>*. *Protein Science* **23**, 423–432 (2014). URL <http://doi.wiley.com/10.1002/pro.2424>.
- [80] Long, K. S., Poehlsgaard, J., Kehrenberg, C., Schwarz, S. & Vester, B. The Cfr rRNA Methyltransferase Confers Resistance to Phenicols, Lincosamides, Oxazolidinones, Pleuromutilins, and Streptogramin A Antibiotics. *Antimicrobial Agents and Chemotherapy* **50**, 2500–2505 (2006). URL <http://aac.asm.org/cgi/doi/10.1128/AAC.00131-06>.
- [81] Vetting, M. W. *et al.* Structure of QnrB1, a Plasmid-mediated Fluoroquinolone Resistance Factor. *Journal of Biological Chemistry* **286**, 25265–25273 (2011). URL <http://www.jbc.org/lookup/doi/10.1074/jbc.M111.226936>.
- [82] Abraham, E. P. & Chain, E. An enzyme from bacteria able to destroy penicillin. 1940. *Reviews of infectious diseases* **10**, 677–8 (1988). URL <http://www.ncbi.nlm.nih.gov/pubmed/3055168>.
- [83] Livermore, D. Defining an extended-spectrum β -lactamase. *Clinical Microbiology and Infection* **14**, 3–10 (2008). URL <https://linkinghub.elsevier.com/retrieve/pii/S1198743X14604717>.
- [84] Nordmann, P., Poirel, L., Walsh, T. R. & Livermore, D. M. The emerging NDM carbapenemases. *Trends in Microbiology* **19**, 588–

- 595 (2011). URL <https://linkinghub.elsevier.com/retrieve/pii/S0966842X11001776>.
- [85] Voulgari, E., Poulou, A., Koumaki, V. & Tsakris, A. Carbapenemase-producing *< i>Enterobacteriaceae</i>* : now that the storm is finally here, how will timely detection help us fight back? *Future Microbiology* **8**, 27–39 (2013). URL <https://www.futuremedicine.com/doi/10.2217/fmb.12.130>.
- [86] Woodford, N., Turton, J. F. & Livermore, D. M. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. *FEMS Microbiology Reviews* **35**, 736–755 (2011). URL <http://www.ncbi.nlm.nih.gov/pubmed/21303394>.
- [87] Woodford, N. & Johnson, A. P. Global spread of antibiotic resistance: the example of New Delhi metallo- β -lactamase (NDM)-mediated carbapenem resistance. *Journal of Medical Microbiology* **62**, 499–513 (2013). URL <http://www.microbiologyresearch.org/content/journal/jmm/10.1099/jmm.0.052555-0>.
- [88] Lynch, J. P., Clark, N. M. & Zhan, G. G. Evolution of antimicrobial resistance among Enterobacteriaceae (focus on extended spectrum β -lactamases and carbapenemases). *Expert Opinion on Pharmacotherapy* **14**, 199–210 (2013). URL <http://www.tandfonline.com/doi/full/10.1517/14656566.2013.763030>.
- [89] Wright, G. Bacterial resistance to antibiotics: Enzymatic degradation and modification. *Advanced Drug Delivery Reviews* **57**, 1451–1470 (2005). URL <https://linkinghub.elsevier.com/retrieve/pii/S0169409X05000980>.
- [90] Norris, A. L. & Serpersu, E. H. Ligand promiscuity through the eyes of the aminoglycoside *< i>N</i>* 3 acetyltransferase IIa. *Protein Science* **22**, 916–928 (2013). URL <http://doi.wiley.com/10.1002/pro.2273>.
- [91] Qin, S. *et al.* Identification of a Novel Genomic Island Conferring Resistance to Multiple Aminoglycoside Antibiotics in *Campylobacter coli*. *Antimicrobial Agents and Chemotherapy* **56**, 5332 (2012). URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3457361/>.

- [92] Mantravadi, P., Kalesh, K., Dobson, R., Hudson, A. & Parthasarathy, A. The Quest for Novel Antimicrobial Compounds: Emerging Trends in Research, Development, and Technologies. *Antibiotics* **8**, 8 (2019). URL <http://www.ncbi.nlm.nih.gov/pubmed/30682820>.
- [93] Bahar, A. A. & Ren, D. Antimicrobial peptides. *Pharmaceuticals (Basel, Switzerland)* **6**, 1543–75 (2013). URL <http://www.ncbi.nlm.nih.gov/pubmed/24287494>.
- [94] Mahlapuu, M., Håkansson, J., Ringstad, L. & Björn, C. Antimicrobial Peptides: An Emerging Category of Therapeutic Agents. *Frontiers in Cellular and Infection Microbiology* **6**, 194 (2016). URL <http://www.ncbi.nlm.nih.gov/pubmed/28083516>.
- [95] Zhang, L.-j. & Gallo, R. L. Antimicrobial peptides. *Current Biology* **26**, R14–R19 (2016). URL <https://www.sciencedirect.com/science/article/pii/S0960982215014098{#}!>
- [96] Ebbensgaard, A. *et al.* Comparative Evaluation of the Antimicrobial Activity of Different Antimicrobial Peptides against a Range of Pathogenic Bacteria. *PLOS ONE* **10**, e0144611 (2015). URL <http://dx.plos.org/10.1371/journal.pone.0144611>.
- [97] Brogden, K. A. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Reviews Microbiology* **3**, 238–250 (2005). URL <http://www.nature.com/articles/nrmicro1098>.
- [98] Hancock, R. E. W. & Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nature Biotechnology* **24**, 1551–1557 (2006). URL <http://www.nature.com/articles/nbt1267>.
- [99] Malmsten, M. Interactions of Antimicrobial Peptides with Bacterial Membranes and Membrane Components. *Current topics in medicinal chemistry* **16**, 16–24 (2016). URL <http://www.ncbi.nlm.nih.gov/pubmed/26139113>.
- [100] Zhang, L., Rozek, A. & Hancock, R. E. Interaction of cationic antimicrobial peptides with model membranes. *The Journal of biological chemistry*

- 276**, 35714–22 (2001). URL <http://www.ncbi.nlm.nih.gov/pubmed/11473117>.
- [101] Schmitt, P., Rosa, R. D. & Destoumieux-Garzón, D. An intimate link between antimicrobial peptide sequence diversity and binding to essential components of bacterial membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1858**, 958–970 (2016). URL <http://linkinghub.elsevier.com/retrieve/pii/S0005273615003430>.
- [102] Lin, T.-Y. & Weibel, D. B. Organization and function of anionic phospholipids in bacteria. *Applied Microbiology and Biotechnology* **100**, 4255–4267 (2016). URL <http://link.springer.com/10.1007/s00253-016-7468-x>.
- [103] Glukhov, E., Stark, M., Burrows, L. L. & Deber, C. M. Basis for selectivity of cationic antimicrobial peptides for bacterial versus mammalian membranes. *The Journal of biological chemistry* **280**, 33960–7 (2005). URL <http://www.ncbi.nlm.nih.gov/pubmed/16043484>.
- [104] Spector, A. A. & Yorek, M. A. Membrane lipid composition and cellular function. *Journal of lipid research* **26**, 1015–35 (1985). URL <http://www.ncbi.nlm.nih.gov/pubmed/3906008>.
- [105] van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nature reviews. Molecular cell biology* **9**, 112–24 (2008). URL <http://www.ncbi.nlm.nih.gov/pubmed/18216768>.
- [106] Yeaman, M. R. & Yount, N. Y. Mechanisms of Antimicrobial Peptide Action and Resistance. *Pharmacological Reviews* **55**, 27–55 (2003). URL <http://pharmrev.aspetjournals.org/cgi/doi/10.1124/pr.55.1.2>.
- [107] Lai, Y. & Gallo, R. L. AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends in Immunology* **30**, 131–141 (2009). URL <https://www.sciencedirect.com/science/article/pii/S1471490609000052>.
- [108] Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **415**, 389–395 (2002). URL <http://www.ncbi.nlm.nih.gov/pubmed/11807545> <http://www.nature.com/articles/415389a>.

- [109] Matsuzaki, K. Control of cell selectivity of antimicrobial peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1788**, 1687–1692 (2009). URL <https://www.sciencedirect.com/science/article/pii/S0005273608003076>.
- [110] Ebenhan, T., Gheysens, O., Kruger, H. G., Zeevaart, J. R. & Satheskge, M. M. Antimicrobial peptides: their role as infection-selective tracers for molecular imaging. *BioMed research international* **2014**, 867381 (2014). URL <http://www.ncbi.nlm.nih.gov/pubmed/25243191>.
- [111] Chan, D. I., Prenner, E. J. & Vogel, H. J. Tryptophan- and arginine-rich antimicrobial peptides: Structures and mechanisms of action. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1758**, 1184–1202 (2006). URL <http://linkinghub.elsevier.com/retrieve/pii/S0005273606001404>.
- [112] Toke, O. Antimicrobial peptides: New candidates in the fight against bacterial infections. *Biopolymers* **80**, 717–735 (2005). URL <http://doi.wiley.com/10.1002/bip.20286>.
- [113] Yang, L., Harroun, T. A., Weiss, T. M., Ding, L. & Huang, H. W. Barrel-stave model or toroidal model? A case study on melittin pores. *Biophysical journal* **81**, 1475–85 (2001). URL <http://www.ncbi.nlm.nih.gov/pubmed/11509361>.
- [114] Bertelsen, K., Dorosz, J., Hansen, S. K., Nielsen, N. C. & Vosegaard, T. Mechanisms of Peptide-Induced Pore Formation in Lipid Bilayers Investigated by Oriented 31P Solid-State NMR Spectroscopy. *PLoS ONE* **7**, e47745 (2012). URL <http://dx.plos.org/10.1371/journal.pone.0047745>.
- [115] Lee, M.-T., Chen, F.-Y. & Huang, H. W. Energetics of Pore Formation Induced by Membrane Active Peptides . *Biochemistry* **43**, 3590–3599 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/15035629>.
- [116] Spaar, A., Münster, C. & Salditt, T. Conformation of peptides in lipid membranes studied by x-ray grazing incidence scattering. *Biophys-*

- cal journal* **87**, 396–407 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/15240474>.
- [117] Matsuzaki, K., Murase, O., Fujii, N. & Miyajima, K. An Antimicrobial Peptide, Magainin 2, Induced Rapid Flip-Flop of Phospholipids Coupled with Pore Formation and Peptide Translocation . *Biochemistry* **35**, 11361–11368 (1996). URL <http://www.ncbi.nlm.nih.gov/pubmed/8784191>.
- [118] Hallock, K. J., Lee, D.-K. & Ramamoorthy, A. MSI-78, an analogue of the magainin antimicrobial peptides, disrupts lipid bilayer structure via positive curvature strain. *Biophysical journal* **84**, 3052–60 (2003). URL <http://www.ncbi.nlm.nih.gov/pubmed/12719236>.
- [119] He, K., Ludtke, S. J., Huang, H. W. & Worcester, D. L. Antimicrobial peptide pores in membranes detected by neutron in-plane scattering. *Biochemistry* **34**, 15614–8 (1995). URL <http://www.ncbi.nlm.nih.gov/pubmed/7495788>.
- [120] Matsuzaki, K. *et al.* Relationship of Membrane Curvature to the Formation of Pores by Magainin 2 . *Biochemistry* **37**, 11856–11863 (1998). URL <https://pubs.acs.org/doi/10.1021/bi980539y>.
- [121] Matsuzaki, K., Sugishita, K.-i., Harada, M., Fujii, N. & Miyajima, K. Interactions of an antimicrobial peptide, magainin 2, with outer and inner membranes of Gram-negative bacteria. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1327**, 119–130 (1997). URL <https://www.sciencedirect.com/science/article/pii/S0005273697000515?via%20Dihub>.
- [122] Shai, Y. Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by α -helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1462**, 55–70 (1999). URL <https://www.sciencedirect.com/science/article/pii/S000527369900200X?via%20Dihub>.
- [123] Ladokhin, A. S. & White, S. H. Detergent-like' permeabilization of anionic lipid vesicles by melittin. *Biochimica et Biophysica Acta*

- (BBA) - *Biomembranes* **1514**, 253–260 (2001). URL <https://www.sciencedirect.com/science/article/pii/S0005273601003820>.
- [124] Oren, Z. & Shai, Y. Mode of action of linear amphipathic α -helical antimicrobial peptides. *Biopolymers* **47**, 451–463 (1998). URL <http://www.ncbi.nlm.nih.gov/pubmed/10333737>.
- [125] Gazit, E., Boman, A., Boman, H. G. & Shai, Y. Interaction of the mammalian antibacterial peptide cecropin P1 with phospholipid vesicles. *Biochemistry* **34**, 11479–88 (1995). URL <http://www.ncbi.nlm.nih.gov/pubmed/7547876>.
- [126] Yamaguchi, S. *et al.* Orientation and Dynamics of an Antimicrobial Peptide in the Lipid Bilayer by Solid-State NMR Spectroscopy. *Biophysical Journal* **81**, 2203–2214 (2001). URL <https://linkinghub.elsevier.com/retrieve/pii/S0006349501758687>.
- [127] Tang, M. & Hong, M. Structure and mechanism of β -hairpin antimicrobial peptides in lipid bilayers from solid-state NMR spectroscopy. *Molecular BioSystems* **5**, 317 (2009). URL <http://xlink.rsc.org/?DOI=b820398a>.
- [128] Lehrer, R. I. Primate defensins. *Nature Reviews Microbiology* **2**, 727–738 (2004). URL <http://www.nature.com/articles/nrmicro976>.
- [129] Fujii, G., Eisenberg, D. & Selsted, M. E. Defensins promote fusion and lysis of negatively charged membranes. *Protein Science* **2**, 1301–1312 (1993). URL <http://doi.wiley.com/10.1002/pro.5560020813>.
- [130] Kagan, B. L., Selsted, M. E., Ganz, T. & Lehrer, R. I. Antimicrobial defensin peptides form voltage-dependent ion-permeable channels in planar lipid bilayer membranes. *Proceedings of the National Academy of Sciences* **87**, 210–214 (1990). URL <http://www.pnas.org/cgi/doi/10.1073/pnas.87.1.210>.
- [131] Takeuchi, K. *et al.* Channel-forming membrane permeabilization by an antibacterial protein, sapecin: determination of membrane-buried and oligomerization surfaces by NMR. *The Journal of biological chemistry* **279**, 4981–7 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/14630928>.

- [132] Dathe, M. & Weprecht, T. Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells. *Biochimica et biophysica acta* **1462**, 71–87 (1999). URL <http://www.ncbi.nlm.nih.gov/pubmed/10590303>.
- [133] Yonezawa, A., Kuwahara, J., Fujii, N. & Sugiura, Y. Binding of tachyplesin I to DNA revealed by footprinting analysis: significant contribution of secondary structure to DNA binding and implication for biological action. *Biochemistry* **31**, 2998–3004 (1992). URL <http://pubs.acs.org/doi/abs/10.1021/bi00126a022>.
- [134] Gifford, J. L., Hunter, H. N. & Vogel, H. J. Lactoferricin. *Cellular and Molecular Life Sciences* **62**, 2588–2598 (2005). URL <http://link.springer.com/10.1007/s00018-005-5373-z>.
- [135] Tomita, M., Takase, M., Bellamy, W. & Shimamura, S. A review: the active peptide of lactoferrin. *Acta paediatrica Japonica : Overseas edition* **36**, 585–91 (1994). URL <http://www.ncbi.nlm.nih.gov/pubmed/7825467>.
- [136] Hwang, P. M., Zhou, N., Shan, X., Arrowsmith, C. H. & Vogel, H. J. Three-Dimensional Solution Structure of Lactoferricin B, an Antimicrobial Peptide Derived from Bovine Lactoferrin . *Biochemistry* **37**, 4288–4298 (1998). URL <http://www.ncbi.nlm.nih.gov/pubmed/9521752>.
- [137] Schibli, D. J., Hwang, P. M. & Vogel, H. J. The structure of the antimicrobial active center of lactoferricin B bound to sodium dodecyl sulfate micelles. *FEBS Letters* **446**, 213–217 (1999). URL [http://doi.wiley.com/10.1016/S0014-5793\(%\)2899\(%\)2900214-8](http://doi.wiley.com/10.1016/S0014-5793(%)2899(%)2900214-8).
- [138] Nguyen, L. T., Schibli, D. J. & Vogel, H. J. Structural studies and model membrane interactions of two peptides derived from bovine lactoferricin. *Journal of Peptide Science* **11**, 379–389 (2005). URL <http://doi.wiley.com/10.1002/psc.629>.
- [139] Peschel, A. & Sahl, H.-G. The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nature Reviews Microbiology* **4**, 529–536 (2006). URL <http://www.nature.com/articles/nrmicro1441>.

- [140] Juhas, M. Horizontal gene transfer in human pathogens. *Critical Reviews in Microbiology* **41**, 101–108 (2015). URL <http://www.tandfonline.com/doi/full/10.3109/1040841X.2013.804031>.
- [141] Joo, H.-S., Fu, C.-I. & Otto, M. Bacterial strategies of resistance to antimicrobial peptides. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **371** (2016). URL <http://www.ncbi.nlm.nih.gov/pubmed/27160595>.
- [142] Sieprawska-Lupa, M. *et al.* Degradation of human antimicrobial peptide LL-37 by *Staphylococcus aureus*-derived proteinases. *Antimicrobial agents and chemotherapy* **48**, 4673–9 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/15561843>.
- [143] Teufel, P. & Götz, F. Characterization of an extracellular metalloprotease with elastase activity from *Staphylococcus epidermidis*. *Journal of Bacteriology* **175**, 4218 (1993). URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC204852/>.
- [144] Schmidtchen, A., Frick, I.-M., Andersson, E., Tapper, H. & Björck, L. Proteinases of common pathogenic bacteria degrade and inactivate the antibacterial peptide LL-37. *Molecular Microbiology* **46**, 157–168 (2002). URL <http://doi.wiley.com/10.1046/j.1365-2958.2002.03146.x>.
- [145] Barańska-Rybak, W., Sonesson, A., Nowicki, R. & Schmidtchen, A. Glycosaminoglycans inhibit the antibacterial activity of LL-37 in biological fluids. *Journal of Antimicrobial Chemotherapy* **57**, 260–265 (2006). URL <http://academic.oup.com/jac/article/57/2/260/804780/Glycosaminoglycans-inhibit-the-antibacterial>.
- [146] Nelson, D. C., Garbe, J. & Collin, M. Cysteine proteinase SpeB from *Streptococcus pyogenes* a potent modifier of immunologically important host and bacterial proteins. *Biological Chemistry* **392**, 1077–88 (2011). URL <https://www.degruyter.com/view/j/bchm.2011.392.issue-12/bc.2011.208/bc.2011.208.xml>.
- [147] Frick, I.-M. *et al.* Constitutive and Inflammation-Dependent Antimicrobial Peptides Produced by Epithelium Are Differentially Processed and Inactivated by the Commensal *< i>Finegoldia magna</i>* and the

- Pathogen <i>Streptococcus pyogenes</i>. *The Journal of Immunology* **187**, 4300–4309 (2011). URL <http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.1004179>.
- [148] Selsted, M. E. & Harwig, S. S. Determination of the disulfide array in the human defensin HNP-2. A covalently cyclized peptide. *The Journal of biological chemistry* **264**, 4003–7 (1989). URL <http://www.ncbi.nlm.nih.gov/pubmed/2917986>.
- [149] Stumpe, S., Schmid, R., Stephens, D. L., Georgiou, G. & Bakker, E. P. Identification of OmpT as the protease that hydrolyzes the antimicrobial peptide protamine before it enters growing cells of *Escherichia coli*. *Journal of bacteriology* **180**, 4002–6 (1998). URL <http://www.ncbi.nlm.nih.gov/pubmed/9683502>.
- [150] Biegeleisen, K. The probable structure of the protamineDNA complex. *Journal of Theoretical Biology* **241**, 533–540 (2006). URL <https://linkinghub.elsevier.com/retrieve/pii/S0022519305005473>.
- [151] Sol, A. *et al.* Actin enables the antimicrobial action of LL-37 peptide in the presence of microbial proteases. *The Journal of biological chemistry* **289**, 22926–41 (2014). URL <http://www.ncbi.nlm.nih.gov/pubmed/24947511>.
- [152] Groisman, E. A., Parra-Lopez, C., Salcedo, M., Lipps, C. J. & Heffron, F. Resistance to host antimicrobial peptides is necessary for *Salmonella* virulence. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 11939 (1992). URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC50673/>.
- [153] Parra-Lopez, C., Baer, M. T. & Groisman, E. A. Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in *Salmonella typhimurium*. *The EMBO journal* **12**, 4053–62 (1993). URL <http://www.ncbi.nlm.nih.gov/pubmed/8223423>.
- [154] Mason, K. M., Munson, R. S., Jr. & Bakaletz, L. O. A Mutation in the sap Operon Attenuates Survival of Nontypeable *Haemophilus influenzae* in a Chinchilla Model of Otitis Media. *Infection and Immunity*

- 73, 599 (2005). URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC538956/>.
- [155] Taggart, C. C. *et al.* Inactivation of Human -Defensins 2 and 3 by Elastolytic Cathepsins. *The Journal of Immunology* **171**, 931–937 (2003). URL <http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.171.2.931>.
- [156] Bokarewa, M. & Tarkowski, A. Human α -defensins neutralize fibrinolytic activity exerted by staphylokinase. *Thrombosis and Haemostasis* **91**, 991–999 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/15116261>.
- [157] Jin, T. *et al.* Staphylococcus aureus Resists Human Defensins by Production of Staphylokinase, a Novel Bacterial Evasion Mechanism. *The Journal of Immunology* **172**, 1169–1176 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/14707093>.
- [158] Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: a common cause of persistent infections. *Science (New York, N.Y.)* **284**, 1318–22 (1999). URL <http://www.ncbi.nlm.nih.gov/pubmed/10334980>.
- [159] Jolivet-Gougeon, A. & Bonnaure-Mallet, M. Biofilms as a mechanism of bacterial resistance. *Drug Discovery Today: Technologies* **11**, 49–56 (2014). URL <https://linkinghub.elsevier.com/retrieve/pii/S1740674914000043>.
- [160] Nickel, J. C., Ruseska, I., Wright, J. B. & Costerton, J. W. Tobramycin resistance of *Pseudomonas aeruginosa* cells growing as a biofilm on urinary catheter material. *Antimicrobial agents and chemotherapy* **27**, 619–24 (1985). URL <http://www.ncbi.nlm.nih.gov/pubmed/3923925>.
- [161] Mah, T. F. & O'Toole, G. A. Mechanisms of biofilm resistance to antimicrobial agents. *Trends in microbiology* **9**, 34–9 (2001). URL <http://www.ncbi.nlm.nih.gov/pubmed/11166241>.
- [162] Wang, X., Preston, J. F. & Romeo, T. The pgaABCD locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required

- for biofilm formation. *Journal of bacteriology* **186**, 2724–34 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/15090514>.
- [163] Vuong, C. *et al.* Polysaccharide intercellular adhesin (PIA) protects *Staphylococcus epidermidis* against major components of the human innate immune system. *Cellular microbiology* **6**, 269–75 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/14764110>.
- [164] Vuong, C. *et al.* A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence. *The Journal of biological chemistry* **279**, 54881–6 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/15501828>.
- [165] Campos, M. A. *et al.* Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. *Infection and immunity* **72**, 7107–14 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/15557634>.
- [166] Llobet, E., Tomas, J. M. & Bengoechea, J. A. Capsule polysaccharide is a bacterial decoy for antimicrobial peptides. *Microbiology* **154**, 3877–3886 (2008). URL <http://mic.microbiologystrease.org/content/journal/micro/10.1099/mic.0.2008/022301-0>.
- [167] Batoni, G., Maisetta, G., Lisa Brancatisano, F., Esin, S. & Campa, M. Use of Antimicrobial Peptides Against Microbial Biofilms: Advantages and Limits. *Current Medicinal Chemistry* **18**, 256–279 (2011). URL <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=0929-8673&volume=18&issue=2&spage=256>.
- [168] Stremper, N., Strehmel, J. & Overhage, J. Potential Application of Antimicrobial Peptides in the Treatment of Bacterial Biofilm Infections. *Current Pharmaceutical Design* **21**, 67–84 (2014). URL <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1381-6128&volume=21&issue=1&spage=67>.
- [169] Joo, H.-S. & Otto, M. Molecular Basis of In Vivo Biofilm Formation by Bacterial Pathogens. *Chemistry & Biology* **19**, 1503–1513 (2012). URL <https://linkinghub.elsevier.com/retrieve/pii/S1074552112004231>.

- [170] Di Luca, M., Maccari, G. & Nifosì, R. Treatment of microbial biofilms in the post-antibiotic era: prophylactic and therapeutic use of antimicrobial peptides and their design by bioinformatics tools. *Pathogens and Disease* **70**, 257–270 (2014). URL <https://academic.oup.com/femsdp/article-lookup/doi/10.1111/2049-632X.12151>.
- [171] Peschel, A. *et al.* Inactivation of the dlt operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. *The Journal of biological chemistry* **274**, 8405–10 (1999). URL <http://www.ncbi.nlm.nih.gov/pubmed/10085071>.
- [172] Fabretti, F. *et al.* Alanine Esters of Enterococcal Lipoteichoic Acid Play a Role in Biofilm Formation and Resistance to Antimicrobial Peptides. *Infection and Immunity* **74**, 4164–4171 (2006). URL <http://www.ncbi.nlm.nih.gov/pubmed/16790791>.
- [173] Saar-Dover, R. *et al.* D-Alanylation of Lipoteichoic Acids Confers Resistance to Cationic Peptides in Group B Streptococcus by Increasing the Cell Wall Density. *PLoS Pathogens* **8**, e1002891 (2012). URL <http://dx.plos.org/10.1371/journal.ppat.1002891>.
- [174] Moskowitz, S. M., Ernst, R. K. & Miller, S. I. 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* **186**, 575–9 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/14702327>.
- [175] Gunn, J. S. *et al.* PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. *Molecular microbiology* **27**, 1171–82 (1998). URL <http://www.ncbi.nlm.nih.gov/pubmed/9570402>.
- [176] Wang, X., Karbarz, M. J., McGrath, S. C., Cotter, R. J. & Raetz, C. R. H. MsbA transporter-dependent lipid A 1-dephosphorylation on the periplasmic surface of the inner membrane: topography of francisella novicida LpxE expressed in *Escherichia coli*. *The Journal of biological chemistry* **279**, 49470–8 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/15339914>.

- [177] Wang, X., McGrath, S. C., Cotter, R. J. & Raetz, C. R. H. Expression cloning and periplasmic orientation of the *Francisella novicida* lipid A 4'-phosphatase LpxF. *The Journal of biological chemistry* **281**, 9321–30 (2006). URL <http://www.ncbi.nlm.nih.gov/pubmed/16467300>.
- [178] Bugg, T. D. H. *et al.* Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. *Biochemistry* **30**, 10408–10415 (1991). URL <http://pubs.acs.org/doi/abs/10.1021/bi00107a007>.
- [179] Brötz, H. *et al.* Role of lipid-bound peptidoglycan precursors in the formation of pores by nisin, epidermin and other lantibiotics. *Molecular microbiology* **30**, 317–27 (1998). URL <http://www.ncbi.nlm.nih.gov/pubmed/9791177>.
- [180] Guo, L. *et al.* Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell* **95**, 189–98 (1998). URL <http://www.ncbi.nlm.nih.gov/pubmed/9790526>.
- [181] Bishop, R. E. *et al.* Transfer of palmitate from phospholipids to lipid A in outer membranes of gram-negative bacteria. *The EMBO journal* **19**, 5071–80 (2000). URL <http://emboj.embopress.org/cgi/doi/10.1093/emboj/19.19.5071>.
- [182] Silhavy, T. J., Kahne, D. & Walker, S. The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology* **2**, a000414–a000414 (2010). URL <http://www.ncbi.nlm.nih.gov/pubmed/20452953>.
- [183] Loutet, S. A., Flannagan, R. S., Kooi, C., Sokol, P. A. & Valvano, M. A. A complete lipopolysaccharide inner core oligosaccharide is required for resistance of *Burkholderia cenocepacia* to antimicrobial peptides and bacterial survival in vivo. *Journal of bacteriology* **188**, 2073–80 (2006). URL <http://www.ncbi.nlm.nih.gov/pubmed/16513737>.
- [184] Allen, C. A., Adams, L. G. & Ficht, T. A. Transposon-derived *Brucella abortus* rough mutants are attenuated and exhibit reduced intracellular survival. *Infection and immunity* **66**, 1008–16 (1998). URL <http://www.ncbi.nlm.nih.gov/pubmed/9488389>.

- [185] Peschel, A. *et al.* <*i*>Staphylococcus aureus</*i*> 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* **193**, 1067–1076 (2001). URL <http://www.jem.org/lookup/doi/10.1084/jem.193.9.1067>.
- [186] Thedieck, K. *et al.* The MprF protein is required for lysinylation of phospholipids in listerial membranes and confers resistance to cationic antimicrobial peptides (CAMPs) on Listeria monocytogenes. *Molecular Microbiology* **62**, 1325–1339 (2006). URL <http://doi.wiley.com/10.1111/j.1365-2958.2006.05452.x>.
- [187] Klein, S. *et al.* Adaptation of <*i*>Pseudomonas aeruginosa</*i*> to various conditions includes tRNA-dependent formation of alanyl-phosphatidylglycerol. *Molecular Microbiology* **71**, 551–565 (2009). URL <http://doi.wiley.com/10.1111/j.1365-2958.2008.06562.x>.
- [188] Band, V. & Weiss, D. Mechanisms of Antimicrobial Peptide Resistance in Gram-Negative Bacteria. *Antibiotics* **4**, 18–41 (2014). URL <http://www.ncbi.nlm.nih.gov/pubmed/25927010>.
- [189] Kumariya, R., Sood, S. K., Rajput, Y. S., Saini, N. & Garsa, A. K. Increased membrane surface positive charge and altered membrane fluidity leads to cationic antimicrobial peptide resistance in Enterococcus faecalis. *Biochimica et biophysica acta* **1848**, 1367–75 (2015). URL <http://www.ncbi.nlm.nih.gov/pubmed/25782727>.
- [190] Raetz, C. R., Reynolds, C. M., Trent, M. S. & Bishop, R. E. Lipid A Modification Systems in Gram-Negative Bacteria. *Annual Review of Biochemistry* **76**, 295–329 (2007). URL <http://www.annualreviews.org/doi/10.1146/annurev.biochem.76.010307.145803>.
- [191] Wang, G. *et al.* APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research* **44**, D1087–D1093 (2016). URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1278>.
- [192] Pirtskhalava, M. *et al.* DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides.

- Nucleic Acids Research* **44**, 6503–6503 (2016). URL <http://www.ncbi.nlm.nih.gov/pubmed/27060142>.
- [193] Jhong, J.-H. *et al.* dbAMP: an integrated resource for exploring antimicrobial peptides with functional activities and physicochemical properties on transcriptome and proteome data. *Nucleic Acids Research* **47**, D285–D297 (2019). URL <https://academic.oup.com/nar/article/47/D1/D285/5150231>.
- [194] Lata, S., Mishra, N. K. & Raghava, G. P. AntiBP2: improved version of antibacterial peptide prediction. *BMC Bioinformatics* **11**, S19 (2010). URL <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-S1-S19>.
- [195] Bhadra, P., Yan, J., Li, J., Fong, S. & Siu, S. W. I. AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest. *Scientific Reports* **8**, 1697 (2018). URL <http://www.nature.com/articles/s41598-018-19752-w>.
- [196] Fjell, C. D., Hiss, J. A., Hancock, R. E. W. & Schneider, G. Designing antimicrobial peptides: form follows function. *Nature Reviews Drug Discovery* **11**, 37 (2011). URL <http://www.nature.com/doifinder/10.1038/nrd3591>.
- [197] Wiradharma, N. *et al.* Synthetic cationic amphiphilic α -helical peptides as antimicrobial agents. *Biomaterials* **32**, 2204–2212 (2011). URL <https://linkinghub.elsevier.com/retrieve/pii/S0142961210015097>.
- [198] Huang, Y., Huang, J. & Chen, Y. Alpha-helical cationic antimicrobial peptides: relationships of structure and function. *Protein & Cell* **1**, 143–152 (2010). URL <http://link.springer.com/10.1007/s13238-010-0004-3>.
- [199] Pag, U. *et al.* Analysis of in vitro activities and modes of action of synthetic antimicrobial peptides derived from an α -helical sequence template'. *Journal of Antimicrobial Chemotherapy* **61**, 341–352

- (2008). URL <https://academic.oup.com/jac/article-lookup/doi/10.1093/jac/dkm479>.
- [200] Wang, J. *et al.* High specific selectivity and Membrane-Active Mechanism of the synthetic centrosymmetric α -helical peptides with Gly-Gly pairs. *Scientific reports* **5**, 15963 (2015). URL <http://www.nature.com/articles/srep15963>.
- [201] Hilpert, K., Volkmer-Engert, R., Walter, T. & Hancock, R. E. W. High-throughput generation of small antibacterial peptides with improved activity. *Nature Biotechnology* **23**, 1008–1012 (2005). URL <http://www.nature.com/articles/nbt1113>.
- [202] Hilpert, K. *et al.* Sequence Requirements and an Optimization Strategy for Short Antimicrobial Peptides. *Chemistry & Biology* **13**, 1101–1107 (2006). URL <https://linkinghub.elsevier.com/retrieve/pii/S107455210600336X>.
- [203] Migoñ, D. *et al.* Alanine Scanning Studies of the Antimicrobial Peptide Aurein 1.2. *Probiotics and Antimicrobial Proteins* 1–13 (2018). URL <http://link.springer.com/10.1007/s12602-018-9501-0>.
- [204] Grieco, P. *et al.* Alanine scanning analysis and structure-function relationships of the frog-skin antimicrobial peptide temporin-1Ta. *Journal of Peptide Science* **17**, 358–365 (2011). URL <http://doi.wiley.com/10.1002/psc.1350>.
- [205] Xie, J. *et al.* Potent effects of amino acid scanned antimicrobial peptide Feleucin-K3 analogs against both multidrug-resistant strains and biofilms of *Pseudomonas aeruginosa*. *Amino Acids* **50**, 1471–1483 (2018). URL <http://link.springer.com/10.1007/s00726-018-2625-4>.
- [206] Radzishevsky, I. S. *et al.* Effects of acyl versus aminoacyl conjugation on the properties of antimicrobial peptides. *Antimicrobial agents and chemotherapy* **49**, 2412–20 (2005). URL <http://www.ncbi.nlm.nih.gov/pubmed/15917541>.
- [207] Serrano, G. N., Zhanel, G. G. & Schweizer, F. Antibacterial Activity of Ultrashort Cationic Lipo- -Peptides. *Antimicrobial Agents and*

- Chemotherapy* **53**, 2215–2217 (2009). URL <http://www.ncbi.nlm.nih.gov/pubmed/19237652>.
- [208] Avrahami, D. & Shai, Y. A New Group of Antifungal and Antibacterial Lipopeptides Derived from Non-membrane Active Peptides Conjugated to Palmitic Acid. *Journal of Biological Chemistry* **279**, 12277–12285 (2004). URL <http://www.jbc.org/lookup/doi/10.1074/jbc.M312260200>.
- [209] Liu, S., Bao, J., Lao, X. & Zheng, H. Novel 3D Structure Based Model for Activity Prediction and Design of Antimicrobial Peptides. *Scientific Reports* **8**, 11189 (2018). URL <http://www.nature.com/articles/s41598-018-29566-5>.
- [210] Jiang, Z., Vasil, A. I., Gera, L., Vasil, M. L. & Hodges, R. S. 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 Para. *Chemical Biology & Drug Design* **77**, 225–240 (2011). URL <http://doi.wiley.com/10.1111/j.1747-0285.2011.01086.x>.
- [211] Deslouches, B. *et al.* De Novo Generation of Cationic Antimicrobial Peptides: Influence of Length and Tryptophan Substitution on Antimicrobial Activity. *Antimicrobial Agents and Chemotherapy* **49**, 316–322 (2005). URL <http://www.ncbi.nlm.nih.gov/pubmed/15616311>.
- [212] Schmidt, N., Mishra, A., Lai, G. H. & Wong, G. C. Arginine-rich cell-penetrating peptides. *FEBS Letters* **584**, 1806–1813 (2010). URL <http://doi.wiley.com/10.1016/j.febslet.2009.11.046>.
- [213] Loose, C., Jensen, K., Rigoutsos, I. & Stephanopoulos, G. A linguistic model for the rational design of antimicrobial peptides. *Nature* **443**, 867–869 (2006). URL <http://www.nature.com/articles/nature05233>.
- [214] Cipcigan, F. *et al.* Accelerating molecular discovery through data and physical sciences: Applications to peptide-membrane interactions. *The Journal of Chemical Physics* **148**, 241744 (2018). URL <http://aip.scitation.org/doi/10.1063/1.5027261>.

- [215] Bolintineanu, D. S. & Kaznessis, Y. N. Computational studies of protegrin antimicrobial peptides: A review. *Peptides* **32**, 188–201 (2011). URL <http://www.ncbi.nlm.nih.gov/pubmed/20946928>.
- [216] Khandelia, H. & Kaznessis, Y. N. Molecular dynamics simulations of helical antimicrobial peptides in SDS micelles: What do point mutations achieve? *Peptides* **26**, 2037–2049 (2005). URL <https://linkinghub.elsevier.com/retrieve/pii/S0196978105001610>.
- [217] Tsai, C.-W. *et al.* Coupling Molecular Dynamics Simulations with Experiments for the Rational Design of Indolicidin-Analogous Antimicrobial Peptides. *Journal of Molecular Biology* **392**, 837–854 (2009). URL <https://linkinghub.elsevier.com/retrieve/pii/S0022283609008031>.
- [218] Farrotti, A. *et al.* 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* **22**, 1235 (2017). URL <http://www.mdpi.com/1420-3049/22/7/1235>.
- [219] Walters, W. P. & Goldman, B. B. Feature selection in quantitative structure-activity relationships. *Current opinion in drug discovery & development* **8**, 329–33 (2005). URL <http://www.ncbi.nlm.nih.gov/pubmed/15892248>.
- [220] Gonzalez, M., Teran, C., Saiz-Urra, L. & Teijeira, M. Variable Selection Methods in QSAR: An Overview. *Current Topics in Medicinal Chemistry* **8**, 1606–1627 (2008). URL <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1568-0266&volume=8&issue=18&spage=1606>.
- [221] Wimley, W. C. Describing the Mechanism of Antimicrobial Peptide Action with the Interfacial Activity Model. *ACS Chemical Biology* **5**, 905–917 (2010). URL <http://www.ncbi.nlm.nih.gov/pubmed/20698568>.
- [222] Lewenza, S. Construction of a mini-Tn5-luxCDABE mutant library in *Pseudomonas aeruginosa* PAO1: A tool for identifying differentially regulated genes. *Genome Research* **15**, 583–589 (2005). URL <http://www.genome.org/cgi/doi/10.1101/gr.3513905>.

- [223] Cherkasov, A. *et al.* Use of Artificial Intelligence in the Design of Small Peptide Antibiotics Effective against a Broad Spectrum of Highly Antibiotic-Resistant Superbugs. *ACS Chemical Biology* **4**, 65–74 (2009). URL <http://www.ncbi.nlm.nih.gov/pubmed/19055425>.
- [224] Lee, E. Y., Lee, M. W., Fulan, B. M., Ferguson, A. L. & Wong, G. C. L. What can machine learning do for antimicrobial peptides, and what can antimicrobial peptides do for machine learning? *Interface Focus* **7**, 20160153 (2017). URL <http://rsfs.royalsocietypublishing.org/lookup/doi/10.1098/rsfs.2016.0153>.
- [225] AB Naafs, M. The Antimicrobial Peptides: Ready for Clinical Trials? *Biomedical Journal of Scientific & Technical Research* **7**, 001–005 (2018). URL <https://biomedres.us/fulltexts/BJSTR.MS.ID.001536.php>.
- [226] Wipf, P., Xiao, J. & Stephenson, C. R. J. Peptide-Like Molecules (PLMs): A Journey from Peptide Bond Isosteres to Gramicidin S Mimetics and Mitochondrial Targeting Agents. *CHIMIA International Journal for Chemistry* **63**, 764–775 (2009). URL <http://www.ncbi.nlm.nih.gov/pubmed/20725595>.
- [227] Choudhary, A. & Raines, R. T. An Evaluation of Peptide-Bond Isosteres. *ChemBioChem* **12**, 1801–1807 (2011). URL <http://doi.wiley.com/10.1002/cbic.201100272>.
- [228] Anguela, X. M. & High, K. A. Entering the Modern Era of Gene Therapy. *Annual Review of Medicine* **70**, 273–288 (2019). URL <https://www.annualreviews.org/doi/10.1146/annurev-med-012017-043332>.
- [229] Barrangou, R. The roles of CRISPRCas systems in adaptive immunity and beyond. *Current Opinion in Immunology* **32**, 36–41 (2015). URL <http://www.ncbi.nlm.nih.gov/pubmed/25574773>.
- [230] Zhang, F., Wen, Y. & Guo, X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Human Molecular Genetics* **23**, R40–R46 (2014). URL <http://www.ncbi.nlm.nih.gov/pubmed/24651067>.

- [231] Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–78 (2014). URL <http://www.ncbi.nlm.nih.gov/pubmed/24906146>.
- [232] Oldfield, E. H. *et al.* 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* **4**, 39–69 (1993). URL <http://www.ncbi.nlm.nih.gov/pubmed/8384892>.
- [233] Lawler, S. E., Speranza, M.-C., Cho, C.-F. & Chiocca, E. A. Oncolytic Viruses in Cancer Treatment. *JAMA Oncology* **3**, 841 (2017). URL <http://www.ncbi.nlm.nih.gov/pubmed/27441411>.
- [234] Naldini, L. Ex vivo gene transfer and correction for cell-based therapies. *Nature Reviews Genetics* **12**, 301–315 (2011). URL <http://www.nature.com/articles/nrg2985>.
- [235] Mingozzi, F. & High, K. A. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. *Nature Reviews Genetics* **12**, 341–355 (2011). URL <http://www.nature.com/articles/nrg2988>.
- [236] Sánchez, L., Calvo, M. & Brock, J. H. Biological role of lactoferrin. *Archives of disease in childhood* **67**, 657–61 (1992). URL <http://www.ncbi.nlm.nih.gov/pubmed/1599309>.
- [237] Arnold, R., Cole, M. & McGhee, J. R. A bactericidal effect for human lactoferrin. *Science* **197**, 263–265 (1977). URL <http://www.ncbi.nlm.nih.gov/pubmed/327545>.
- [238] Arnold, R. R., Brewer, M. & Gauthier, J. J. Bactericidal activity of human lactoferrin: sensitivity of a variety of microorganisms. *Infection and immunity* **28**, 893–8 (1980). URL <http://www.ncbi.nlm.nih.gov/pubmed/6772569>.
- [239] Kirkpatrick, C. H., Green, I., Rich, R. R. & Schade, A. L. Inhibition of Growth of *Candida albicans* by Iron-Unsaturated Lactoferrin: Relation to Host-Defense Mechanisms in Chronic Mucocutaneous Candidiasis. *Journal of Infectious Diseases* **124**, 539–544

- (1971). URL <https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/124.6.539>.
- [240] Jahani, S., Shakiba, A. & Jahani, L. The Antimicrobial Effect of Lactoferrin on Gram-Negative and Gram-Positive Bacteria. *International Journal of Infection* **2** (2015). URL <http://intjinflection.com/en/articles/14726.html>.
- [241] Farnaud, S. & Evans, R. W. Lactoferrin—a multifunctional protein with antimicrobial properties. *Molecular immunology* **40**, 395–405 (2003). URL <http://www.ncbi.nlm.nih.gov/pubmed/14568385>.
- [242] Shau, H., Kim, A. & Golub, S. H. Modulation of natural killer and lymphokine-activated killer cell cytotoxicity by lactoferrin. *Journal of Leukocyte Biology* **51**, 343–349 (1992). URL <http://doi.wiley.com/10.1002/jlb.51.4.343>.
- [243] Gahr, M., Speer, C. P., Damerau, B. & Sawatzki, G. Influence of Lactoferrin on the Function of Human Polymorphonuclear Leukocytes and Monocytes. *Journal of Leukocyte Biology* **49**, 427–433 (1991). URL <http://doi.wiley.com/10.1002/jlb.49.5.427>.
- [244] Bellamy, W. *et al.* Identification of the bactericidal domain of lactoferrin. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1121**, 130–136 (1992). URL <https://www.sciencedirect.com/science/article/pii/016748389290346F>.
- [245] Wakabayashi, H. *et al.* Cooperative anti-Candida effects of lactoferrin or its peptides in combination with azole antifungal agents. *Microbiology and immunology* **40**, 821–5 (1996). URL <http://www.ncbi.nlm.nih.gov/pubmed/8985937>.
- [246] Aguilera, O., Ostolaza, H., Quirós, L. & Fierro, J. Permeabilizing action of an antimicrobial lactoferricin-derived peptide on bacterial and artificial membranes. *FEBS Letters* **462**, 273–277 (1999). URL <https://www.sciencedirect.com/science/article/pii/S0014579399015458>.
- [247] Cochran, A. G., Skelton, N. J. & Starovasnik, M. A. Tryptophan zippers: Stable, monomeric β -hairpins. *Proceedings of the National Academy of*

- Sciences* **98**, 5578–5583 (2001). URL <https://www.pnas.org/content/98/10/5578>.
- [248] Tsutsumi, A. *et al.* Structure and orientation of bovine lactoferrampin in the mimetic bacterial membrane as revealed by solid-state NMR and molecular dynamics simulation. *Biophysical journal* **103**, 1735–43 (2012). URL <http://www.ncbi.nlm.nih.gov/pubmed/23083717>.
- [249] Arseneault, M., Bédard, S., Boulet-Audet, M. & Pézolet, M. Study of the Interaction of Lactoferricin B with Phospholipid Monolayers and Bilayers. *Langmuir* **26**, 3468–3478 (2010). URL <http://pubs.acs.org/doi/abs/10.1021/la903014w>.
- [250] Strøm, M. B. *et al.* Important structural features of 15-residue lactoferricin derivatives and methods for improvement of antimicrobial activity. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **80**, 65–74 (2002). URL <http://www.ncbi.nlm.nih.gov/pubmed/11908644>.
- [251] Crombez, L. *et al.* A New Potent Secondary Amphipathic Cellpenetrating Peptide for siRNA Delivery Into Mammalian Cells. *Molecular Therapy* **17**, 95–103 (2009). URL <https://www.sciencedirect.com/science/article/pii/S1525001616314800?via%3Dihub>.
- [252] Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry* **55**, 611–22 (2009). URL <http://www.ncbi.nlm.nih.gov/pubmed/12142370>.
- [253] Elsabahy, M. *et al.* Surface Charges and Shell Crosslinks Each Play Significant Roles in Mediating Degradation, Biofouling, Cytotoxicity and Immunotoxicity for Polyphosphoester-based Nanoparticles. *Scientific Reports* **3**, 3313 (2013). URL <http://www.nature.com/articles/srep03313>.