DISS. ETH NO. 27505

Massively-parallelized discovery and optimization of antimicrobials

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

PHILIPP KOCH

M.Sc., Technical University of Munich

born March 26th, 1990

citizen of Germany

accepted on the recommendation of

Prof. Dr. Sven Panke (ETH Zurich, Switzerland), examiner

Dr. Martin Held (ETH Zurich, Switzerland), co-examiner

Prof. Dr. Urs Jenal (University of Basel, Switzerland), co-examiner

Prof. Dr. Jörn Piel (ETH Zurich, Switzerland), co-examiner

2021

Abstract

The number of newly approved antimicrobial compounds has been steadily decreasing over the past 50 years emphasizing the need for novel antimicrobial substances. Besides, current antimicrobial therapies become increasingly ineffective due to the rapid spread of drug-resistant pathogens, and even the most potent small-molecule antibiotics may eventually fail to cure infections with highly resistant bacteria. Peptides are a promising class of potential drug substances, as they are still underexplored, offer high structural complexity, and at least partly cover the chemical space left open between biologics and small molecule drugs. Ribosomally produced antimicrobial peptides are of particular interest, as they already act as effective antimicrobials in the defense against invading pathogens of most organisms. Consequently, they are currently being investigated for potential use to fight infectious diseases in humans. The fact that antimicrobial peptides are produced using transcription and translation makes them particularly attractive as they can be produced recombinantly and their modification can be simply achieved by straightforward manipulation of the synthesis template at the DNA level.

In this thesis, we developed a high-throughput method for the discovery of antimicrobial peptides which is based on the principle of self-screening: Different peptides are expressed in a recombinant *Escherichia coli* strain, and their effect on growth rate is recorded. In our case, this is done by next-generation sequencing of expression plasmids in the bacterial culture, enabling the recording of massive numbers of growth curves for single strains in a single flask. We termed this technology massively parallelized growth assays (Me^x). We applied this method to discover novel candidates for antimicrobial peptides by screening a library of ~12'000 naturally occurring peptides with a length between 5 and 42 amino acids and diverse properties. Analysis of thousands of growth curves allowed us to identify more than 1,000 previously unknown[*] antimicrobials. Additionally, by incorporating the kinetics of growth inhibition, we were able to obtain a first indication of the mode of action, with important implications for the ultimate usefulness of the peptide in question. We chemically synthesized the most promising peptides of the screen and determined their activity when applied externally. Notably, the results indicated that 10 out of 15 investigated peptides efficiently eradicated bacteria at a minimal inhibitory concentration in the upper nM / lower µM range. We think that this work represents a step-change in the high-throughput discovery of functionally diverse antimicrobial peptides.

[**]Not present on the antimicrobial peptide database

Next, we applied a simplified version of Me^x to optimize a single antimicrobial peptide in high-throughput. As a model peptide, we optimized the 23 amino acid version of the well-researched and highly active antimicrobial peptide Bac7$_{1-23}$ using deep mutational scanning, consisting of a first random and then a semi-rational approach. The random library of ~600,000 different Bac7$_{1-23}$ variants allowed us to derive a fitness landscape of the peptide and to identify residues that are essential for growth inhibition and residues with potential for activity optimization. A smaller semi-rational library of ~160,000 Bac7$_{1-23}$ variants enabled us to extract the most beneficial amino acid combinations, thereby generating an antimicrobial peptide that, if synthesized chemically, is non-toxic and superior to Bac7$_{1-23}$ against a large panel of bacterial pathogens. We thus created a new-to-nature peptide lead with a great potential to be further developed in pre-clinical stages.

To our best estimation, these novel methods exceed competing approaches in terms of throughput, hit-rate and sensitivity, while also offering an opportunity for the direct functional characterization of large libraries. These methods will accelerate the discovery and optimization of antimicrobials drastically, and may thus provide a path forward to master one of today's most urgent challenges, the antimicrobial resistance crisis.

Zusammenfassung

Die Anzahl neu zugelassener antimikrobieller Wirkstoffe ist in den letzten 50 Jahren stetig gesunken, was den Bedarf an neuen antimikrobiellen Substanzen verdeutlicht. Ausserdem werden derzeitige antimikrobielle Therapien immer ineffektiver aufgrund der schnellen Verbreitung antibiotika-resistenter Pathogene und selbst die wirkungsstärksten niedermolekularen Antibiotika könnten schlussendlich nicht mehr in der Lage sein, Infektionen mit hoch-resistenten Bakterien zu heilen. Peptide stellen eine vielversprechende Klasse möglicher Arzneimittelsubstanzen dar, da sie noch unerforscht sind, eine hohe strukturelle Komplexität aufweisen und zumindest teilweise den chemischen Raum zwischen Biologika und niedermolekularen Arzneimitteln ausfüllen. Vom Ribosom hergestellte antimikrobielle Peptide sind dabei von besonderem Interesse, da sie in den meisten Organismen schon zur Abwehr eintretender Pathogene als effektive antimikrobielle Agenzien wirken. Daher werden sie derzeit auf ihren potentiellen Nutzen im Kampf gegen Infektionskrankheiten im Menschen untersucht. Antimikrobielle Peptide gelten als besonders attraktiv da sie via Transkription und Translation hergestellt werden und sie somit rekombinant hergestellt werden und sie unkompliziert auf der Ebene des Synthese Musters, der DNA, modifziert werden können.

In dieser Arbeit entwickelten wir eine Hochdurchsatzmethode für die Entdeckung von antimikrobiellen Peptiden, die auf dem Prinzip des „Selbst-Screenings" basiert: verschiedene Peptide werden in einem rekombinanten *Escherichia coli* Stamm exprimiert und ihr Effekt auf die Wachstumsrate wird aufgezeichnet. In unserem Fall wurde diese Aufzeichnung mittels 'Next-Generation Sequencing' der Expressionsplasmide in der bakteriellen Kultur durchgeführt, was es ermöglicht, eine grosse Anzahl an Wachstumskurven einzelner Stämme in einer einzelnen Kulturflasche aufzuzeichnen. Wir nennen diese Technologie „massively parallelized growth assays" (Me^x). Wir verwendeten diese Methode um eine Bibliothek von ~12'000 natürlich vorkommenden Peptiden mit einer Länge zwischen 5 und 42 Aminosäuren und vielfältigen Eigenschaften zu screenen um neue Kandidaten für antimikrobielle Peptide zu entdecken. Die Analyse tausender Wachstumskurven erlaubte es uns, mehr als 1,000 bisher unbekannte[†] antimikrobielle Peptide zu identifizieren. Wenn wir zusätzlich die Kinetiken der individuellen Wachstumshemmung miteinbezogen, konnten wir zudem einen ersten Hinweis auf den jeweiligen Wirkmechanismus ableiten, was wichtige Implikationen für die endgültige Nutzbarkeit des jeweiligen Peptides beinhaltet. Wir synthetisierten die vielversprechendsten Peptide aus dem Screen chemisch und bestimmten ihre Aktivität bei

---

[†] Nicht vorhanden in der Datenbank 'antimicrobial peptide database'

Zugabe zu Bakterien von aussen. Bemerkenswerterweise deuteten unsere Ergebnisse darauf hin, dass 10 von 15 getesteten Peptiden Bakterien mit einer minimalen inhibitorischen Konzentration im oberen nM / unterem µM Bereich abtöteten. Wir glauben, dass diese Arbeit einen Sprung in der Hochdurchsatzentwicklung von funktionell vielfältigen antimikrobiellen Peptiden darstellt.

Als nächstes verwendeten wir eine vereinfachte Variante von Me^x, um ein einzelnes antimikrobielles Peptid im Hochdurchsatz zu optimieren. Als ein Modell-Peptide optimierten wir die 23 Aminosäuren lange Version des stark beforschten und hochaktiven antimikrobiellen Peptids $Bac7_{1-23}$ zunächst mittels eines zufälligen und dann mittels eines halb-rationalen Ansatzes des „deep mutational scannings". Die zufällig erzeugte Bibliothek aus ~600,000 verschiedenen $Bac7_{1-23}$ Varianten erlaubte es uns, eine Fitness-Landschaft des Peptides zu generieren und dabei einerseits Aminosäurereste zu identifizieren, die essential für die Wachstumsinhibition sind und andererseits Reste zu finden, die das Potential für eine Optimierung der Aktivität des Peptides haben. Eine kleinere, halb-rational designte Bibliothek von ~160,000 $Bac7_{1-23}$ Varianten ermöglichte es uns, die für die Aktivität des Peptides vorteilhaftesten Aminosäurekombinationen zu identifizieren und somit ein antimikrobielles Peptid zu generieren, das nach chemischer Synthese nicht toxisch gegenüber eukaryotischen Zellen ist und dem ursprünglichen $Bac7_{1-23}$ in der Abtötung einer grosse Liste an bakteriellen Pathogenen überlegen ist. Somit generierten wir eine bisher nicht in der Natur auftretende Peptid-Leitstruktur mit einem hohen Potential für eine Weiterentwicklung in präklinischen Phasen.

Unserer Einschätzung nach übertrumpfen diese neuartigen Methoden konkurrierende Ansätze im Hinblick auf Durchsatz, Treffer-Rate und Sensitivität und bieten zudem die Möglichkeit für eine direkte funktionelle Charakterisierung grosser Peptid-Bibliotheken. Diese Methoden werden die Entdeckung und Optimierung von antimikrobiellen Agenzien drastisch beschleunigen und können somit den Weg ebnen, um eine der dringlichsten Herausforderungen der heutigen Zeit, die Krise der antimikrobiellen Resistenz, zu bewältigen.

Table of contents

CHAPTER 1 | Introduction

1.1 Antimicrobials and antimicrobial resistance

At the beginning of the 20th century, infectious diseases accounted for roughly half of all deaths in the United States of America[1]. Since then there has been a sharp decline in infectious disease mortality for which, next to continued improvements in hygiene measures, the discoveries of antibiotics played a crucial role. Especially the discoveries of salvarsan, penicillin, and streptomycin in the first half of the 20th century revolutionized medicine and resulted from decades of research lead by some of the most influential (anti) microbiologists to date: Paul Ehrlich, Alexander Fleming, Howard Florey, Ernst Chain, and Selman Waksman. Paul Ehrlich initially discovered the organoarsenic compound salvarsan[2] and for the first time publicly described the concept of a 'Zauberkugel' or 'magic bullet': a compound with high target-selectivity and minimal deleterious side effects, which in the case of antibiotics attacks pathogens but remains harmless for healthy tissues in humans[3]. But it was not until the discovery of the β-lactam antibiotic penicillin in 1928[4] by Alexander Fleming and its subsequent characterization and purification by Howard Florey and Ernst Chain in 1940[5] that the medicine was applied at a global scale[‡]. The discovery of the first broad-spectrum[§] antibiotic, the aminoglycoside streptomycin in 1944[6], was made by Selman Waksman in the course of systematic cultivation and isolation of soil bacteria and their secreted compounds. It was at that point when scientists around the world including large pharmaceutical and chemical companies joined in, starting large screening campaigns and ushering in what is now known as the golden era of antibiotic discoveries[7]. During the following two decades the majority of additional antibiotic classes were discovered, such as the chloramphenicols, tetracyclines, macrolides, and glycopeptides. However, from the end of that period until today, there has been a steady decline of novel antibiotic matter derived from natural origin. More specifically: until the 1960s, 24 classes of antibiotics had been discovered. Since then, only 8 new classes of natural origin have been added[8].

Even though there was a strong decline in the discovery of natural antibiotic classes, a plethora of synthetic antibiotic analogs have been developed[7]. However, there are two major biological issues why this replenishment did not solve all problems and we still struggle to treat bacterial infections today: (i) the bacterial cell envelope is highly restrictive of what can pass the barrier it represents[9], considerably limiting the chemical space of drug-like substances that can be applied, and (ii) the emergence of extensive antimicrobial resistance (AMR)[10]. Especially the latter is believed to be one of the major medicinal challenges humanity has to master in the upcoming decades[11]. Simplified, there are four common mechanisms by which bacteria can

---

[‡]Big drivers were World War II and Pfizer's ability to produce it at large scale.
[§]Compared to narrow-spectrum antibiotics, broad-spectrum antibiotics are active against a wide range of bacterial pathogens, often against both Gram-positive and Gram-negative bacteria.

enhance their resistance to antibiotics: reduction of cell permeability, increased efflux, alteration of the target, or inactivation of an antibiotic through enzymatic processing[12]. These mechanisms mainly result from millions of years of exposure to antibiotic substances produced from competing or defending organisms, but also from excessive use of antibiotics in human or animal therapeutics[12]. Many of the emerging multidrug resistant bacterial pathogens such as _Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa_ and the _Enterobacter spp._, the so-called ESKAPE panel, cause great concern as they have become difficult to treat with available antibiotics, and are responsible for most hospital-acquired infections[13]. Especially carbapenem-resistant Gram-negative _P. aeruginosa, A. baumannii,_ and Enterobacteriaceae are currently listed as 'critical priority' by the WHO according to criteria such as mortality, treatment, health-care burden, or prevalence of resistance[14]. Many strategies have been implemented that emphasize antibiotic stewardship and target hurdles from the regulatory, clinical, or development side[**] to delay or even prevent a potentially emerging 'post-antibiotic era' in which hardly any available antibiotic is effective to treat infections. Antibacterial drug discovery still plays a central role in those strategies and thus establishing novel methods to discover or optimize promising new compounds that end up in clinical pipelines is crucial to success.

As most of the currently used medicines, including antibiotics, are either natural products or derived from natural products[15], it is potentially rewarding to cast a fresh look at nature as our best engineer and chemist[16]. Even though most natural sources, such as the actinomycetes, seem to have been exhaustively investigated[7], novel omics-based technologies and screening methods are highly promising to find new leads that might have remained hidden until now, like teixobactin[17] or albicidin[18], two recently discovered non-ribosomal peptide (NRP) antibiotics. Many hopes are also pinned on antimicrobial peptides (AMPs), a well-characterized but clinically not yet used group of antibiotic compounds. Compared to most small-molecule antibiotics, which are usually built from various building blocks by an enzymatic cascade or enzyme complexes, AMPs are ribosomally translated. These compounds are produced by most life forms, and some representatives have shown already very promising results against the most threatening multidrug-resistant pathogens today[19].

[**]https://carb-x.org/; https://amractionfund.com; https://gardp.org/; https://www.finddx.org/amr/; EU One Health Action Plan against AMR; https://www.repair-impact-fund.com/

1.2 Antimicrobial peptides

1.2.1 Definitions

Seven years before Alexander Fleming first discovered penicillin, he noticed strong antibacterial properties when treating staphylococci with nasal mucus of patients suffering from rhinitis, and named the responsible compound lysozyme[20]. It was later discovered that lysozyme is a 139 amino acid long enzyme, widely distributed in nature, that hydrolyzes the linkages in the peptidoglycan layer of Gram-positive bacteria[21]. Following this discovery, many other peptides and proteins with antibacterial properties were isolated such as nisin, phagocytin, cecropin or magainin[22–25], establishing the research field of AMPs[26]. During that time many important NRPs, such as polymyxin, gramicidin, or vancomycin[27–29], were discovered and have since been broadly used to treat infectious diseases. Even though sometimes these compounds are also referred to as AMPs, they are generally excluded from the group of AMPs as I do in this thesis, because the primary peptide sequence does not appear as template on genes and is not produced ribosomally.

The terminology of ribosomally produced antibacterial molecules has been widely debated and in addition to AMPs, they are often referred to as host defense peptides (HDPs)[30], ribosomally synthesized and post-translationally modified peptides (RiPPs)[31], or in the case of bacterial origin, bacteriocins[32]. In this thesis, I will consistently use the term AMP for all ribosomally produced peptides that are associated with antimicrobial activity including all HDPs, RiPPs, and bacteriocins. Moreover, there have been multiple suggestions on how to classify AMPs: by their taxonomy[33] (bacteriocins, plant AMPs, animal AMPs, fungal AMPs) with further subclassifications; by their 3D-structure[30] (α-helices, β-sheets, αβ, loop, extended structure); or by their bonding pattern[34] (linear, sidechain-linked, backbone-linked, sidechain-backbone linked). In this thesis, I will use the classification based on their taxonomy, with further subclassifications into smaller AMP groups such as proline-rich AMPs within animal AMPs.

AMPs are indispensable to the defense processes of higher organisms[35]. In addition to their ability to directly kill invading pathogens, they play an important role in immune modulation. This includes stimulation and regulation of the host immune response to an infection or involvement in functions such as inflammation control or wound healing[30]. As the focus of this work is the antimicrobial activity, I will only focus on the direct killing function of AMPs in the entire work. The broader context, including most of the known biological functions of AMPs, can be found in a recent review by Haney *et al.*[36].

1.2.2 Biosynthesis and function of antimicrobial peptides

AMPs are ribosomally produced and usually between 5 and 50 amino acids in length and many are synthesized as a long precursor peptide. In the simplest case, only a leader sequence is present in addition to the AMP core sequence, that aids in transport and final activation of the AMP via its own enzymatic cleavage[31]. These peptides adopt a secondary structure without post-translational modifications (PTMs)[30]. In the most complex case, the precursor peptide is recognized by a co-expressed enzymatic modification machinery, which chemically modifies amino acids of the core peptide after translation resulting in highly cross-linked and modified peptide coils[26]. These PTMs restrict the conformational flexibility of the peptide for improved target recognition, increase chemical and proteolytic stability of the molecule, or introduce specific chemical functionalities such as enhanced membrane permeability[37]. While PTMs modulate peptide properties, the extent to which their contribution is essential for antimicrobial activity is not always fully understood[38]. In fact, there are plenty of highly active, unmodified AMPs in nature[39], and the omission of PTMs does not necessarily result in decreased bioactivity[40].

The large majority of AMPs are of amphipathic nature resulting from the overrepresentation of positively charged and hydrophobic residues in their primary structure[30]. This property supports their biological selectivity towards bacteria. Contrary to negatively charged bacterial membranes, eukaryotic membranes contain phospholipids with a net charge of zero in the outer leaflet. Eukaryotic cell membranes also have a lower transmembrane potential[41] and incorporate cholesterol, which increases membrane stability or repulses the peptide[35]. They are therefore chemically quite distinct from their bacterial counterparts. Following electrostatic attraction to the bacterial membrane, AMPs associate with the lipid cores of these membranes using their hydrophobic portions. AMPs eventually kill bacteria by membrane perturbation, by crossing the membrane and inhibiting vital intracellular processes, or by a combination thereof (Fig. 1.1). The most common mechanism of action (MoA) is by membrane perturbation, also called membrane permeabilization, and can be subcategorized by the barrel starve, the carpet (or detergent-like), or the toroidal pore model (Fig. 1.1)[26]. However, the detailed molecular mechanism, as well as the influence of environmental variables such as pH, ionic strength, temperature or peptide concentration, are still not fully elucidated[42]. The second route to direct bacterial killing is via membrane translocation and killing of bacteria without apparent membrane damage. Here, AMPs traverse the cell barriers either via direct penetration or via pore- or receptor-mediated uptake[43–45] and then inhibit vital cellular processes such as DNA or RNA synthesis or protein translation[46]. These mechanisms however are even less studied than membrane permeabilization. In fact, within the group of 3,000 natural AMPs that are reported

in the antimicrobial peptide database (APD)[34], an intracellular MoA has been assigned to less than 50 AMPs[46].
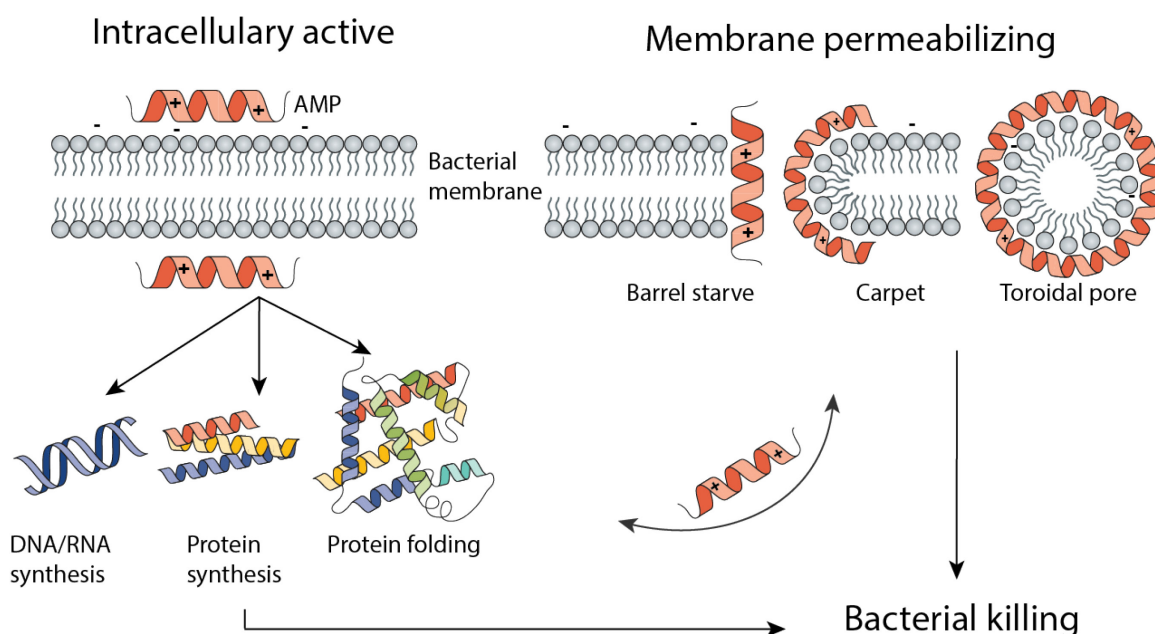

Fig. 1.1 | Antibacterial MoAs of AMPs. AMPs kill bacteria in two different ways. Intracellularly active AMPs translocate across the membrane and interfere with processes involved in protein folding, protein synthesis, or DNA/RNA synthesis. Membrane permeabilizing AMPs kill cells by using different mechanisms: barrel starve, carpet, or toroidal pore. AMPs can also bind to multiple intracellular targets or combine membrane permeabilization and binding to intracellular targets. Adapted from Mookherjee *et al*.[47].

While small molecule-antibiotics often have a single target which they bind with high affinity, some AMPs disturb multiple biological functions with moderate potency[30]. Having multiple targets led to the classification as 'dirty' drugs but in fact might represent a smart way to decrease the formation of resistance in nature, usually caused by the selective pressure from high-affinity antibiotics[48]. However, mounting evidence shows that a few AMPs in fact also show a high target specificity and work in an orchestrated and synergistic fashion together with other AMPs[49]. For example, the bumblebee AMP abaecin, which targets an intracellular protein, is enabled to enter the cytosol by pores created from the co-occurring membrane permeabilizing peptide hymenoptaecin[50]. This illustrates that since the discovery of the lysozyme about 100 years ago, our understanding of the role AMPs play in natural defense, of their route of synthesis, and of the various functions and MoA has increased considerably. AMPs currently represent a promising alternative to conventional small-molecule antibiotics as a starting point for the development of human therapeutics.

1.2.3 Challenges for antimicrobial peptides in clinical pipelines

Thousands of natural AMPs or AMP peptidomimetics have been isolated or generated, displaying excellent *in vitro* properties to kill bacterial pathogens[34]. In fact, a great effort has been made to bring AMP-based antimicrobials from the research bench through clinical

development pipelines. However, to date, even though many AMP-based antibiotics have entered clinical trials, none[††] have achieved regulatory approval and failed during pre-clinical development[36]. A prominent example of failure is pexiganan, a magainin (frog AMP)-derived membrane-permeabilizing AMP that showed excellent in vitro activity[51], but failed in clinical phase III due to inferiority to currently used antibiotics (clinical trial identifier: NCT01590758). Another more recent example is murepavadin, a peptidomimetic derived from the animal AMP protegrin I that targets the outer membrane biogenesis of *P. aeruginosa*[52]. It showed excellent in vitro activity[19], but failed in clinical phase III due to indications of kidney injury (NCT03409679).

Generally, AMPs show crucial benefits over small-molecule antibiotics such as a slower emergence of resistance, a broader spectrum of activity, and, in some cases, also the potential to positively modulate the host immune response[53]. However, they face disadvantages such as higher cost-of-goods, as well as low in vivo stability, higher toxicity, and often low activity under physiological conditions[30]. Some of these disadvantages have been addressed over the past years. Chemical synthesis of smaller peptides (<50 amino acids) has become easier with increasing automation. Larger peptides can be competitively produced recombinantly[54]. Optimizations of biological synthesis routes using bacteria, plant, mammalian, or yeast cells, as summarized by Pachón-Ibáñez *et al.*[55], have increased the feasibility of large-scale production. In contrast, translational research of AMPs still suffers from many unknowns such as pharmacological profiles, formulation, or pharmacokinetics[53]. In the past, these unknowns led to a strong focus on the development of AMP-based antibiotics for topical (body surface) applications[56]. However, with increasing numbers of AMPs being discovered and put into clinical pipelines, late-stage AMP commercialization efforts have adapted and are now experiencing an increase of AMPs destined for other application[47]. In fact, synthetic and natural AMPs are part of the group of antibacterials focusing on 'new targets', which is the largest group of molecules in the current global preclinical antibacterial pipeline (72% of all molecules)[57].

There is a great chance to discover a number of suitable and novel AMPs if technical, biotechnological, and state-of-the-art computational advances are combined with smartly designed AMP discovery, engineering, and screening approaches. This should increase the number of AMPs in preclinical pipelines, fill knowledge gaps, and eventually bring a number of AMP-based therapeutics to the patients.

[††]The 13 amino acid long human-derived AMP called P113 is currently sold as part of a non-prescription antibacterial mouth rinse solution in Taiwan (https://www.oh-care.com). Please note that this statement refers only to AMPs as strictly ribosomally synthesized peptides.

1.3 Discovery and screening of antimicrobial peptides

1.3.1 Overview of bioactive peptides in drug discovery

Over the past 50 years, drug discovery has become a slower and more difficult process, an effect termed 'Eroom's Law'[‡‡], which refers to a steady decrease in the number of drugs released on the market per money spent[58]. Concomitantly, thanks to scientific advances in the past decades, there has been a significant boost in numbers of recombinant biopharmaceuticals such as proteins, especially antibodies, and on a smaller scale also peptides[59]. In fact, 28 new peptide drugs have been approved for various medicinal indications in the period from 2000 to 2018[60]. Peptides are traditionally between 2 and 100 amino acids in length, and lie at the interface of small molecule drugs (<500-900 Da) and proteins (>10 kDa) but according to their composition, size, and route of administration are rather similar to proteins[59]. Even though peptides still represent a small share of all therapeutics (around 5% in 2015), they are increasingly applied in the fields of metabolic diseases, cancer, cardiovascular diseases, and antimicrobials, and their market share in the coming decades is predicted to increase[59].

In the past, bioactive peptides were obtained directly from natural sources. The first example was the extraction and purification of insulin from pancreas cells and its use to treat diabetes in humans[61]. However, as ribosomally synthesized peptides are encoded on DNA, the advent of recombinant DNA technologies allowed for peptides to be synthesized by biological production hosts such as bacteria or yeast. Insulin for example was first commercially produced in *E. coli* by Genentech in 1978[61]. A giant step-change for peptide synthesis and discovery was marked by the development of more economical synthetic methods. Here, the invention of solid-phase peptide synthesis[62] strongly increased production yields and combinatorial possibilities for the generation of peptide libraries.

In general, the ability to generate combinatorial peptide libraries for drug discovery has been greatly advanced and can roughly be divided into chemical and biological peptide libraries[63]. For chemical libraries, solid-phase peptide synthesis, as previously mentioned, and adjusted methods such as the split-and-mix[64] or tea-bag[65] approach are most often used. In more recent approaches, peptides generated by these methods can even be coupled to single or double-stranded DNA, which can be used as identifiers after affinity selection of the peptides[66]. Such libraries are called DNA-encoded libraries (DEL) and allow the screening of millions of peptides at the same time in a target-based fashion. In biological peptide libraries, DNA (genotype) is

---

[‡‡]"Eroom's law" refers to 'Moore's Law', which describes the increase of transistors per integrated circuit over time, read backwards.

used for library generation, creating the link to the encoding peptide (phenotype). Peptides from these genetic libraries are finally produced by cells, for example, using phage display[67] or bacterial systems such as two-hybrid[68], or 'split-intein circular ligation of peptides and proteins' (SICLOPPS)[69]. Other biological methods do not need cells for the production of peptides, such as ribosome display[70], mRNA display[70], or in vitro compartmentalization techniques[71].

In this section, I will focus on the emerging technologies employed for the discovery and activity screening of AMPs.

1.3.2 Natural sources for antimicrobial peptides

The richest source to discover bioactive peptides has been natural products. In fact, the majority of all therapeutic agents approved for clinical use between 1981 and 2019 are natural products or are at least derived from a natural product or its pharmacophore[72]. It has been postulated that because naturally-derived peptides have been preselected for biological activity and as a substrate for biological transporters that are present in their natural host, they cover the relevant 'drug-like' space much better than synthetic compounds[73]. Natural products are important for drug discovery as they already provide the required diversity for covering a wide range of chemical scaffolds and complex structures[74]. Correspondingly, AMPs were originally isolated from natural sources such as plants, animals, and bacteria – a procedure described as bioprospecting[75]. For animal AMPs (75% of all discovered AMPs are animal-derived[34]) this often involved the detection of an antibacterial activity within the animal under study, mostly by first infecting it, and then purifying and further characterizing the antibacterial agent. Cecropins for example were first purified from the hemolymph of moths[76] and then subsequently characterized using Edman sequencing and bacterial susceptibility testing[24].

The potential of bioprospecting has been greatly expanded by substantial progress in the field of genomic sequencing of all organisms, including increasing amounts of metagenomic data, including DNA sequence data from uncultivable bacterial species. Here, the number of individual sequenced organisms that are deposited on GenBank, a genetic database organized by the National Center for Biotechnology Information (NCBI)[77], has vastly increased. Since 1982, the number of bases deposited has doubled every 18 months (Fig. 1.2a). Most of the raw genetic information available (bases deposited) is derived from the animal kingdom, but in terms of the number of deposited genomes, bacteria are representing the largest group (Fig. 1.2b/c). Both peptide sequences and the corresponding coding DNA sequence can be discovered in these databases, a process called genome mining. While genome mining can generally be applied to peptides from different types of organisms[78], it is successfully carried

out using bacterial genomes. Bacteria produce a variety of antimicrobials, including secondary metabolites (e.g. the aminoglycoside streptomycin) or NRPs (e.g. the glycopeptide vancomycin) produced by NRP synthetases (NRPSs) or polyketide synthetases (PKSs). In addition, many bacteria are capable of producing AMPs. A famous example of a bacterial AMP is nisin[25], a lantibiotic produced by *Lactococcus lactis*, which has been used as a food preservative for many years. Moreover, bioinformatics tools such as BAGEL[79] or antiSMASH[80] have helped in mining bacterial genomes for AMPs and their biosynthetic gene clusters (BCGs), groups of genes involved in the production of AMPs, which collectively has yielded a high number of novel compounds[81].

In summary, bioprospecting for AMPs by direct isolation from its source but also through genome mining brought remarkable successes with currently more than 3,200 AMPs deposited on the antimicrobial peptide database (APD, https://wangapd3.com or aps.unmc.edu/AP/main.php)[34].

a

b

c
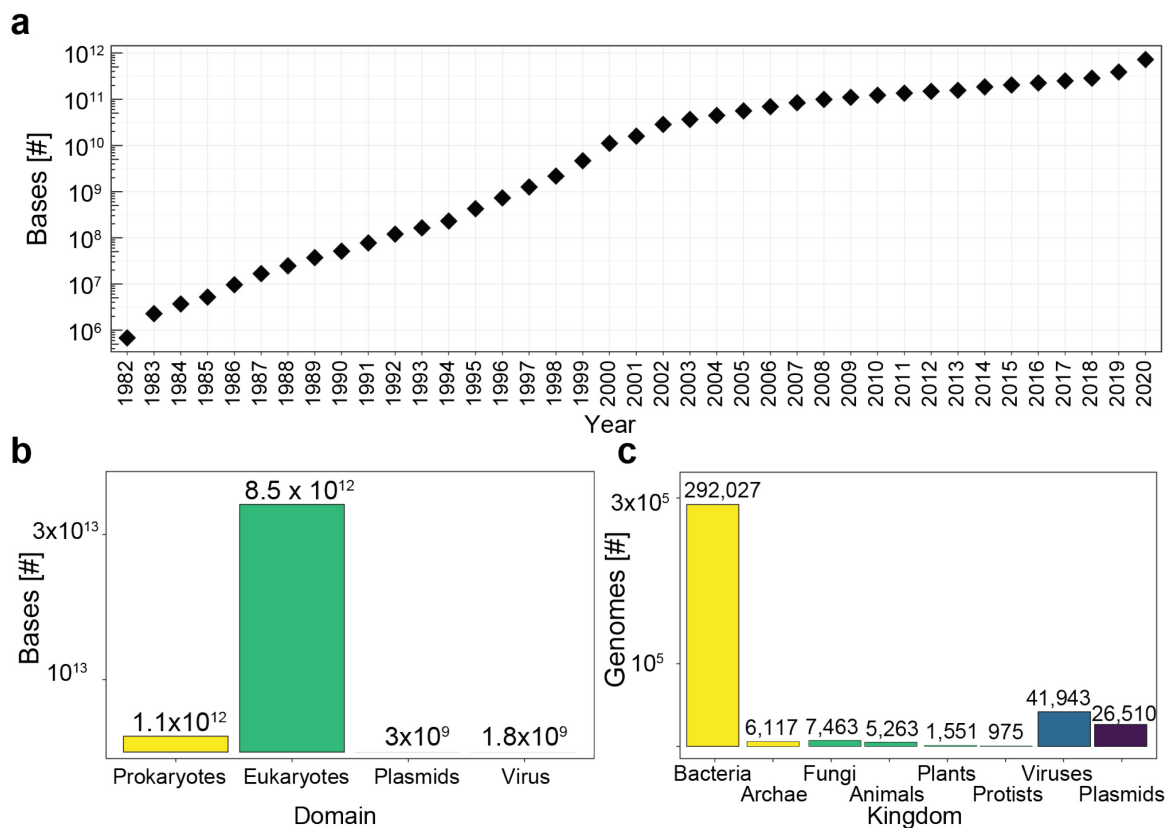
Fig. 1.2 | Genetic information on GenBank a) The total number of available bases on GenBank from 1982-2020. **b)** Number of bases deposited on GenBank distributed over the main domains of life (yellow = prokaryotes; green = eukaryotes; purple = plasmids; blue = viruses). **c)** Number of sequenced genomes deposited on GenBank separated by kingdom including viruses and plasmids. There can be multiple genomes from the same species, e.g. there are ~23,000 *E. coli* genomes available on GenBank. Color code corresponds to domains of c). Data were obtained from Genbank (https://www.ncbi.nlm.nih.gov/genbank) on January 2021.

1.3.3 Screening of antimicrobial peptide libraries

The rich source of natural AMP sequences stored on databases has guided numerous AMP screening campaigns aiming at investigating structure-activity relationships (SARs), improve antibacterial activity, or reduce unwanted toxicity towards eukaryotic cells. Since 2009, more than 1,000 publications and 300 patents on AMPs have appeared each year[82]. Most of these investigations are based on rational design approaches to improve antibacterial activity. Often, only minor alternations of the natural AMP sequences are made, resulting in collections of ~5-10 chemically produced peptide variants that may have the desired improved features[83]. In total, these screening campaigns resulted in the generation of tens of thousands of synthetic AMPs, some of which are also stored on databases such as DRAMP[84], covering both synthetic and natural AMPs.

However, most of these studies only investigate a small part of the overall AMP fitness landscape, which maps a target peptide sequence to its fitness as measured in the experiment[85], measured as antimicrobial activity against one pathogen while altering the peptide sequence (Fig. 1.3). As the primary sequence of a peptide encodes its biological function[83], identifying the most potent antimicrobial peptide of a length of 20 amino acid residues would require in theory the screening of 20^{20} different peptides for antimicrobial activity (20 positions, 20 canonical amino acids possible at each position, $>10^{26}$ different sequence possibilities). This would however be impossible with any currently available screening method. Even fully randomizing five amino acids would require analysis of 3.2 million (20^5) peptides. This problem is additionally amplified as antimicrobial fitness landscapes, such as the one displayed Fig. 1.3, might not align with other desirable fitness landscapes[36]. For example, screening for better antimicrobial activity against specific bacteria might increase the toxicity towards eukaryotic cells or decrease the antimicrobial activity against other bacteria[36]. In order to cover large parts of AMP fitness landscapes, methods have to be developed to strongly increase experimental throughput.
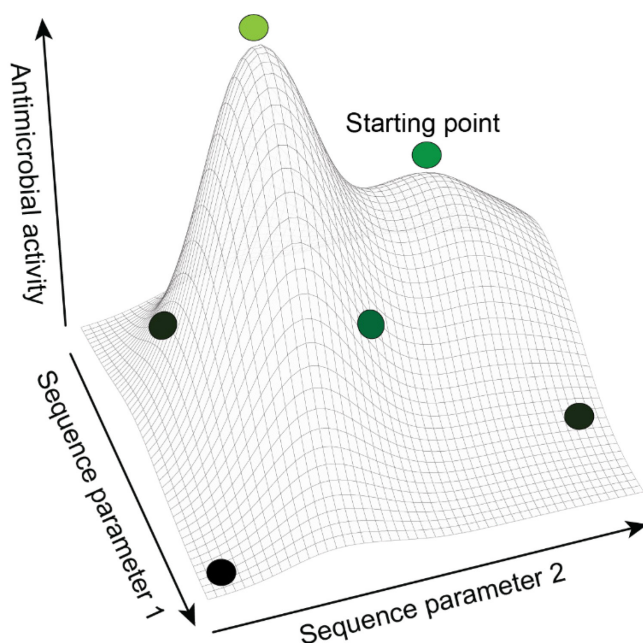
Fig. 1.3 | Fitness landscape of an AMP. Fitness landscapes can be created by altering the chemical space of an AMP and measuring the antimicrobial activity of the resulting AMP. Changing sequence parameters can include modifying to specific amino acids to drastically change physicochemical properties such as overall charge or hydrophobicity, but can also include substituting to all other canonical amino acids. When changing sequence parameters of a natural AMP ("Starting point"), the resulting AMPs may become more active (a lighter green) but may also decrease in antimicrobial activity (darker green). Fitness landscapes become more complicated (multidimensional) if more sequence parameters are included. This is a hypothetical example.

One option to circumvent large screening campaigns and still analyze AMP fitness landscapes is the use of computational methods. Such tools usually exploit the information stored on AMP databases such as the amino acid sequences, physicochemical parameters, or annotated biological functions. To predict a specific function of a peptide or to make design suggestions for AMP optimization, they apply methods using artificial intelligence such as quantitative structure-activity relationships (QSAR)[86] or machine learning[87] including neural network-based deep learning[88]. If we implement increasing levels of mechanistic and functional understanding of AMPs into improved computational methods, it can be expected that the AMP-based drug discovery process will be considerably accelerated.

Next to computational methods, also experimental techniques have been optimized to improve the throughput of AMP screenings. As technologies involving DNA manipulation, synthesis, and sequencing experienced huge advances, methods involving the recombinant production of peptides are increasingly applied for AMP library screenings. Hit identification is straightforward in these screens as the input genotype (DNA) is linked to the peptide output, and the DNA can be easily recovered and read out after the peptide has displayed its activity. Many of these high-throughput approaches are based on display technologies. For example, phage display was used to screen for a strong membrane damaging peptide using a 12-mer random peptide library that was sampled for ligands with increased affinity for the cell surface of *E. coli*[89]. Other attempts are based on bacterial self-screening - a negative selection process in which the expression of an active AMP limits the growth of the bacterium that is synthesizing the AMP[90]. In a recent approach, next-generation sequencing (NGS) was used to study peptide populations in a self-killing approach in which a potentially membrane-damaging peptide was tethered to the cell surface of the producing bacterial host[91]. Most of these recombinant high-

throughput methods rely on the production of linear peptides using the regular synthesis machinery of *E. coli*. A recent method developed at the ETHZ Bioprocess Laboratory demonstrated that also the antimicrobial activity of AMPs with an intricate PTM can be screened in high-throughput. In this method producer cells and sensor bacteria were co-localized in alginate beads ("nanoFleming" platform), and a library of 6,000 different and post-translationally modified lantibiotics was screened for antimicrobial activity[92].

1.4 Scope of this thesis

So far, most biosynthetic high-throughput approaches have relied on random amino acid residue substitutions or low-throughput diversification strategies and resulted in AMPs with mediocre activity or only minor activity improvement[91,93,94]. Eventually, we want to increase the number and most importantly, the quality of promising AMP in clinical pipelines. As high-throughput screening methods are an essential part of drug discovery, we aim to evolve methods for the massively-parallelized discovery and optimization of AMPs.

In Chapter 2 of this thesis, we aim to develop an AMP discovery method that exploits the growing mountain of data available on genetic and AMP databases. As naturally-derived molecules have been valuable starting points in the past, we aim to mine all available genetic sequences for naturally encoded peptides with a certain degree of sequence similarity to already known AMP sequences and subsequently screen them for antimicrobial activity. A created library consists of thousands of highly diverse peptide sequences that cannot be economically chemically synthesized or generated using molecular biological tools. We thus aim to take advantage of recent advances in microarray-based synthesis technology for their template synthesis and recombinant *E. coli* for their production. The antimicrobial activity of all peptides can be assessed by combining *E. coli* self-screening and deep-sequencing, thereby circumventing large-scale AMPs syntheses and functional analyses. Next to AMP discovery, the combination of *E. coli* self-screening and deep-sequencing can be implemented to accelerate sequence to function studies or activity optimizations of AMPs. In Chapter 3, we aim to optimize the antimicrobial activity of an intracellularly active AMP and improve its prospects to be used in clinically relevant settings. The greatest challenge is to functionally assess a large number of possible peptide variants, which arise due to the combinatorial explosion when mutating multiple amino acids simultaneously. We hence aim to perform a two-step mutational approach. First, we screen large numbers of randomly diversified AMP variants to get a broad overview of the antimicrobial fitness landscape of the initial AMP. In a second step, these results can guide a rational peptide-engineering approach that targets selected residues of the peptide. Screening this focused set of peptides generates a second

fitness landscape that allows the identification of peptide candidates with amino acid combinations improving their antimicrobial activity. Such optimized variants have a great chance to enter the clinical development path. Lastly, in Chapter 4, we describe additional methods that need to be developed to successfully analyze AMP candidates derived from Chapter 2 and Chapter 3.

CHAPTER 2 | Discovery of antimicrobials by Me^x: massively parallelized growth assays

Philipp Koch[1], Steven Schmitt[1], Mathias Cardner[2,3], Niko Beerenwinkel[2,3], Sven Panke[1], Martin Held[1]

[1]Bioprocess Laboratory, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland.

[2]Computational Biology, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland.

[3]SIB Swiss Institute of Bioinformatics, 4058 Basel, Switzerland.

2.1 Abstract

High-throughput methods are fundamental for the discovery and characterization of bioactive peptides. Employing recombinant microbes, we developed Me^x to address two major limitations in the field: peptide synthesis and functional analysis. Using Me^x, the activity of 10,663 naturally occurring peptides including nearly all currently known AMPs was profiled. The method provides unparalleled insight into functionally diverse and previously unknown antimicrobial peptides.

2.2 Introduction

Natural compounds are fundamental for drug discovery as they provide the biological relevance and structural diversity required to identify drug-like pharmacophores[73]. Owing to their high structural complexity and their ability to penetrate tissues and membranes, peptides are becoming increasingly important for many therapeutic areas[59]. Especially antimicrobial therapies have a very strong demand for novel compounds due to rising antimicrobial resistance[7]. Although about 3,000 antimicrobial peptides (AMPs) have already been discovered[34], advances in genome sequencing and mining[95] provide an ever-increasing number of peptides with elusive functions[78].

Large peptide libraries can be screened for antimicrobial activity using bacteria self-screening protocols. Here, peptides are expressed from their encoding DNA template and then accumulate either in the cytosol, the periplasm or at the bacterial surface[90,91]. If antimicrobial, their expression negatively impacts the proliferation rate or survival of the expressing cell. Sequencing of the peptide-encoding DNA of such impaired cells allows for the identification of antimicrobials in large pools of uncharacterized peptides. However, previous self-screening approaches failed to deliver large fractions of highly active peptides[91,93], or were unsuited for the screening of big libraries[90].

We gathered naturally-encoded peptides from peptide and genomic sequence databases and assayed them for antimicrobial activity using massively parallelized growth assays (Me^x). Combined, the method delivered a rich collection of functionally diverse and highly active AMPs.

2.3 Results and discussion

We first designed a library of naturally-encoded peptides. For this, we collected the amino acid sequences of 3,063 peptides with experimentally validated activity ("PARENTS" from here on) from the AMP database (APD)[34] (Fig. 2.1a). Notably, PARENTS differed considerably with respect to the host from which they were derived, length, physiochemical properties, chemical modifications, 3D-structure, and sequence (Fig. 2.1a/b). Next, we applied *tblastn* on the translated nucleotide databases accessible through the NCBI[77] using the amino acid sequence of the PARENTS as queries. This search yielded 36,898 amino acid sequences with a similarity of ≥ 21.1% to the PARENTS ("SIMILARS" from here on). Unlike the PARENTS, few SIMILARS have been synthesized or experimentally tested. However, owing to their natural origin and similarity to the PARENTS, a fraction of the SIMILARS is likely to display antimicrobial properties[96,97]. For technical reasons (Methods), we applied a cut-off of 42 amino acids in peptide chain length and selected SIMILARS with at least 62.2% sequence similarity. In this way, a library of 2,122 PARENTS and 10,300 SIMILARS (Fig. 2.1b) was obtained. Examination of the final library indicated net charges from -10 to +15 and hydrophobicity of -3.5 to 2.9 (GRAVY scale; Supplementary Fig. 2.1). Additionally, we were able to allocate the origin of 7,497 of these peptides to the kingdom animalia, 74 to fungi, 678 to bacteria, and 2,485 to plantae (Supplementary Fig. 2.2).

For Me^x, we converted the peptides into corresponding oligonucleotides (Supplementary Fig. 2.3), retrieved the latter as a pool after chemical synthesis on a microarray, and ligated the sequences into a plasmid on which their expression was controlled by the tightly regulated P$_{BAD}$ promotor (Fig. 2.1c). We then transformed the model organism, *E. coli* TOP10, with the peptide-encoding DNA library. Due to a high sequence bias in the initial oligonucleotide pool (Supplementary Fig. 2.4), we only identified 10,663 different peptide-encoding DNA sequences (listed by ID in Supplementary Table 2.1) in *E. coli* using next-generation sequencing (NGS).

a

Antibacterial · Antimammal · Antifungal

1436 · 182 · 66 · 224 · 729 · 6 · 152

Peptide length [aa]: Fungi, Plantae, Bacteria, Animalia

Charge: 20, 10, 0, -10

Chemical modifications: SS- A U C T D W J L Q G K O
Count: 1000, 100, 10, 1

3D-structures: Unknown, Bridge, Helix, Helix-Beta, Rich, Beta
Count: 1000, 100, 10, 1

b

2,112 APD derived PARENTS

10,300 NCBI derived SIMILARs (green towers)

c

Design & Optimization, Synthesis, Cloning, Transformation, Growth experiment, NGS

4.5 h, 3 h, 1.5 h, 0 h

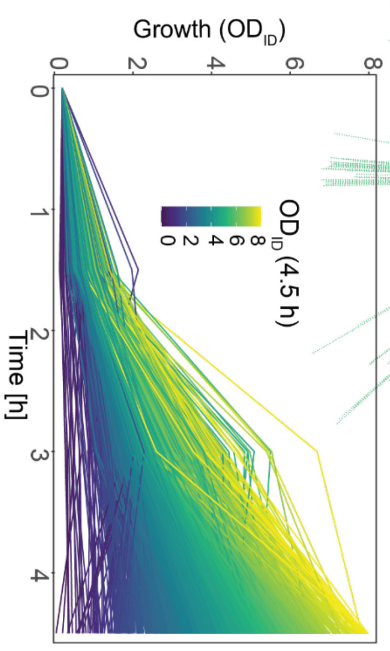**d**

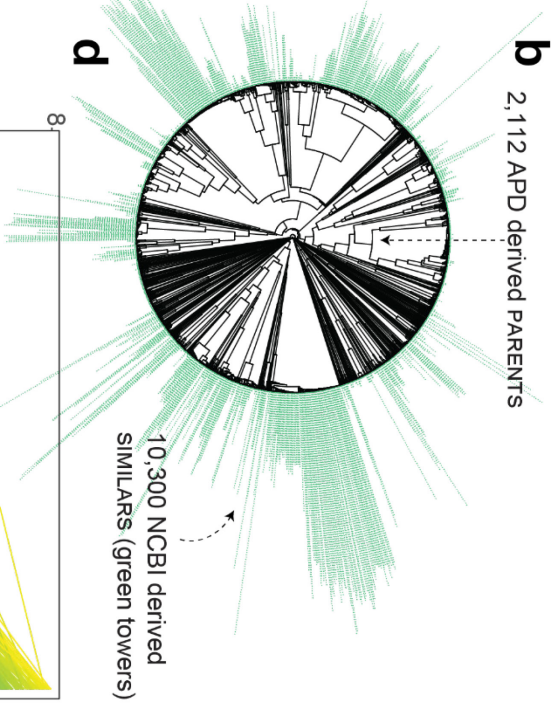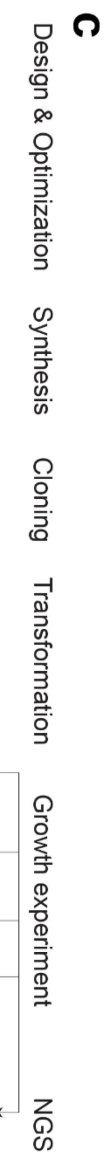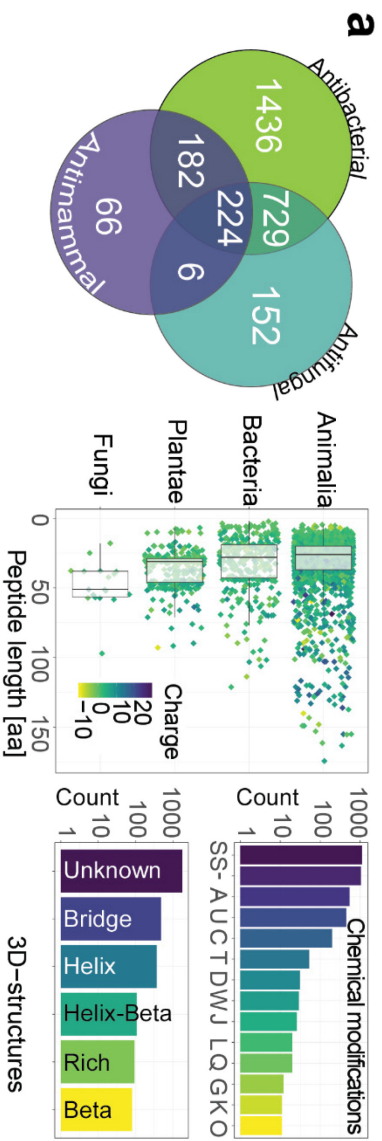Growth (OD$_{ID}$) vs Time [h]

OD$_{ID}$ (4.5 h): 0, 2, 4, 6, 8

Fig. 2.1 | Screening of AMPs using Meˣ. a) Biological diversity of PARENTS. PARENTS are derived from the AMP database (APD). They have experimentally proven biological activity, e.g. antibacterial (Gram-negative and/or Gram-positive bacteria), antifungal, or antimammal (hemolytic or anticancer). They originate from species of various kingdoms of life and differ considerably by length, charge, chemical modification (among others: SS = disulfide bridges, A = amidation, U = terminal Rana box, C = backbone cyclization, T = thioether bridges, D = d-amino acids, W = dehydration, J = sidechain cyclization, L = lipidation, Q = terminal glutamate, E = acetylation, G = glycosylation, K = hydroxylation, - = no modififcation reported), and 3D-structure (Beta = beta-sheet, Bridge = disulfide bond, Helix = alpha-helix, Helix-Beta = alpha-helix and beta-sheet, Rich = rich in unusual amino acids, Unknown = no reported structure). **b)** Sequence distances of the complete peptide library. Pairwise sequence distance between 2,112 PARENTS (BLOSUM62) as a basis for hierarchical clustering. SIMILARS found using *tblastn* for each PARENTS' search query are stacked as towers on the tips of the dendrogram. **c)** Meˣ workflow: Design & Optimization: Peptide sequences are reverse translated into *E. coli* codon-optimized nucleotide sequences. Synthesis: All peptide-encoding sequences are synthesized as oligonucleotides. Cloning: The sequences are inserted into plasmids. Transformation: *E. coli* TOP10 is transformed with the generated peptide-encoding DNA library. Growth: Strains are incubated in shaking flasks, peptide expression is induced and plasmids are isolated. NGS: peptide-encoding DNA sequences are counted at four time points using NGS. **d)** Growth curves of all 10,663 peptide-expressing strains, expressed as OD for a specific peptide-expressing strain (OD_{ID}; average of n=3). Coloring from yellow to dark blue indicates higher growth inhibitory effects based on OD_{ID} of last sampling point. Curves reaching a higher OD_{ID} than eight (0.7 %) are omitted for clarity.

To assess the antimicrobial activity of the DNA-encoded peptides, we performed Meˣ and generated growth curves for each of the 10,663 peptide-expressing *E. coli* strains. To do so, we inoculated three liquid cultures each with 500 million transformed cells, and induced peptide synthesis after four cell doublings (Supplementary Fig. 2.5). Because the expression of an AMP should inhibit the growth of the expressing host, the propagation rate of the peptide-encoding DNA will also be reduced. Hence, we harvested bacteria at the time of induction as well as 1.5 h, 3.0 h, and 4.5 h post-induction and used NGS to count reads for each peptide-encoding DNA. To derive growth curves (Fig. 2.1d), we calculated the abundance of each strain (ID) using the respective NGS read counts and multiplied these with the measured cell concentration of the entire liquid cultures (OD) thereby obtaining an approximation of the strain-specific concentrations (OD_{ID}) at each sampling point. Comparing OD_{ID} of all peptide-expressing strains after 4.5 h, we found that intracellular expression of 1,240 peptides (11.6%) significantly inhibited the growth of their host ("Meˣ-actives" from here on; Wald's test, p-value (*p*)<0.05, adjusted for multiple testing (adj.); Supplementary Fig. 2.6). The remaining peptides did not show growth inhibition in Meˣ, likely because they are not antimicrobial at all or require chemical modifications not introduced in *E. coli*, could not access their (e.g. extracellular) target, or did not reach inhibitory concentrations due to limited mRNA or peptide stability.

Next, we confirmed that the intracellularly synthesized peptides also inhibited growth if the strains were grown individually. For this, we selected 110 peptide-expressing strains experiencing different levels of growth inhibition in Meˣ and measured their growth in microtiter plate wells (Supplementary Fig. 2.7a/b). As the growth curves recorded in Meˣ and microtiter plates were comparable (Supplementary Fig. 2.7c), we concluded that the complex dynamic of the Meˣ-culture did not bias the results.

Screening 10,663 peptides at once allowed us to address several research questions. Firstly, we sought to confirm that our approach of exploiting sequence similarities to known AMPs indeed allowed us to identify antimicrobials. In fact, 1,035 out of 1,240 Me^x-actives (83 %) were SIMILARS, i.e. peptides whose functions were not reported on the APD. A closer look revealed that for 310 inactive PARENTS we found at least one active SIMILAR. As an example, PARENT Apo5 APOC1$_{667\ APD}$ (nomenclature: name of PARENT on APD $_{ID\ Origin}$), itself inactive, yielded 27 SIMILARS of which one showed eight amino acid differences to the PARENT and displayed antimicrobial activity (Supplementary Fig. 2.8). We argue that the amino acids by which the inactive PARENT and the active SIMILAR differed were of high importance for activity and necessary for evading the abovementioned reasons for failed growth inhibition in Me^x. Furthermore, 47 PARENTS spawned an overrepresentation of active SIMILARS (Fisher's exact test, adj. $p<0.05$; Supplementary Fig. 2.9). Examples include Myticin-B (21/31), which spawned 31 SIMILARS, of which 21 were active, and PepG1 (11/11). This indicates that the respective peptide sequences have considerable plasticity and can accommodate multiple amino acid exchanges without losing activity. We argue that these peptides might be well suited for additional modifications performed for instance in the course of lead optimization programs[98].

Secondly, we evaluated the phylogeny of the hosts from which the inhibitory peptides were derived. For this, all peptides of the library were grouped taxonomically based on their natural host. We then calculated the fraction (%) of Me^x-actives within the ranks Kingdom and Class (Fisher's exact test; Supplementary Fig. 2.10). Me^x-actives were significantly underrepresented ($p<0.05$) among bacteria (8.5%), amphibians (7.7%), and mammals (10.3%) but overrepresented ($p<0.05$) in insects (13.4%), birds (25%), ray-finned fishes (15.6%) and bivalves (31.8%). Since insects contain by far the most species in the animal kingdom, this indicates a huge and so far undiscovered pool of antimicrobials in insects.

Thirdly, as cationic and hydrophobic peptides generally display antimicrobial activity[30], we wondered whether growth inhibition in Me^x was biased by the physiochemical properties of peptides. However, linear regression analysis indicated no correlation of growth inhibition with hydrophobicity (correlation = 0.04) and charge (correlation = -0.01; Supplementary Fig. 2.11a). Furthermore, among the 47 PARENTS with overrepresented active SIMILARS, there was no clear relationship between charge or hydrophobicity and growth inhibition (Supplementary Fig. 2.11b). We thus conclude that growth inhibition in Me^x-actives is driven by the specific antimicrobial activity of a peptide either damaging the cytoplasmic membrane or binding and inhibiting other cellular components.

Lastly, we characterized the 50 most growth inhibitory (38 SIMILARS, 12 PARENTS) peptides in Me^x using different assays (Supplementary Fig. 2.12). We first used two biosensor constructs, containing the *cspA* and *recA* promoters, which upon activation are indicative of translation impairment[99] and DNA damage[100], respectively. The results indicated translational impairment for 11 and DNA damage for 12 peptide-expressing strains (one-sided t-test, adj. *p*<0.05; Fig. 2.2b; Supplementary Fig. 2.13), which suggests intracellular targets for these peptides. For example, Metalnikowin IIA$_{8984\ APD}$, Metalnikowin III$_{9011\ APD}$, known ribosomal inhibitors, and Pyrrhocoricin$_{7122\ NCBI}$, whose PARENT is also a ribosomal inhibitor, caused the strongest indication for translational impairment[39]. Next, we measured membrane damage by quantifying propidium iodide (PI) uptake. Expression of 11 peptides resulted in membrane damage, with the strongest damages observed for Delta Lysin I SIMILARS whose PARENT is a membrane pore inducing bacteriocin from *Staphylococcus* (Supplementary Fig. 2.13, Supplementary Table 2.3). Interestingly, all 11 peptides that caused membrane damage significantly inhibited growth already after 1.5 h in Me^x (Wald's test, *p*<0.05) (Fig. 2.2a; Supplementary Fig. 2.14) while for 25 peptides, and especially for those with putative intracellular targets, growth inhibition started only after 4.5 h (Fig. 2.2a; Supplementary Fig. 2.14). Noteworthy, delay of the growth inhibition onset is indicative for peptides that interact with an intracellular target[101]. We concluded that this effect could be observed in Me^x and hence reanalyzed all growth curves recorded for the Me^x-actives. Growth was significantly inhibited after 1.5 h for 806 peptides (65%) suggesting membrane damage (Supplementary Table 2.1) but only after 4.5 h in the case of the remaining 434 peptides, suggesting interaction with an intracellular target. As the transition from the discovery pipeline to the patient is often hampered by the general toxicity of membrane damaging peptides[53], Me^x could hence be a valuable tool for the high-throughput discovery of AMPs that do not predominantly rely on this mechanism.

Growth (OD)

Time [h]

b

Growth (OD_ID)

c
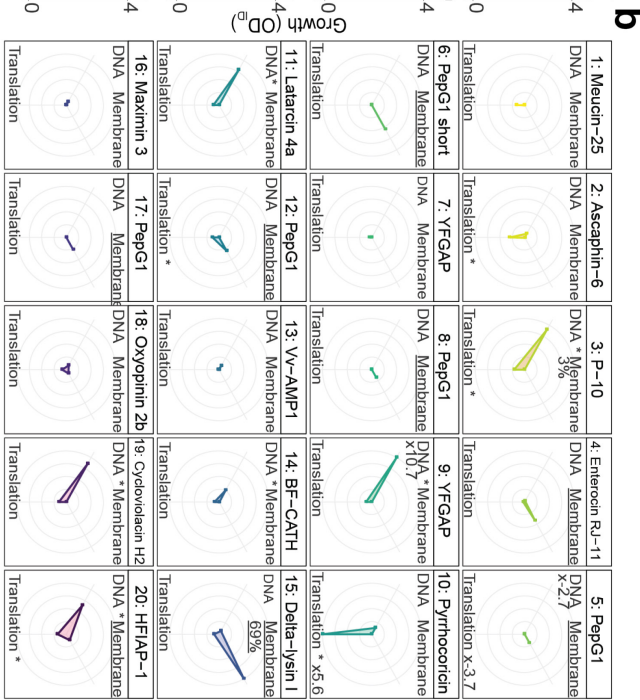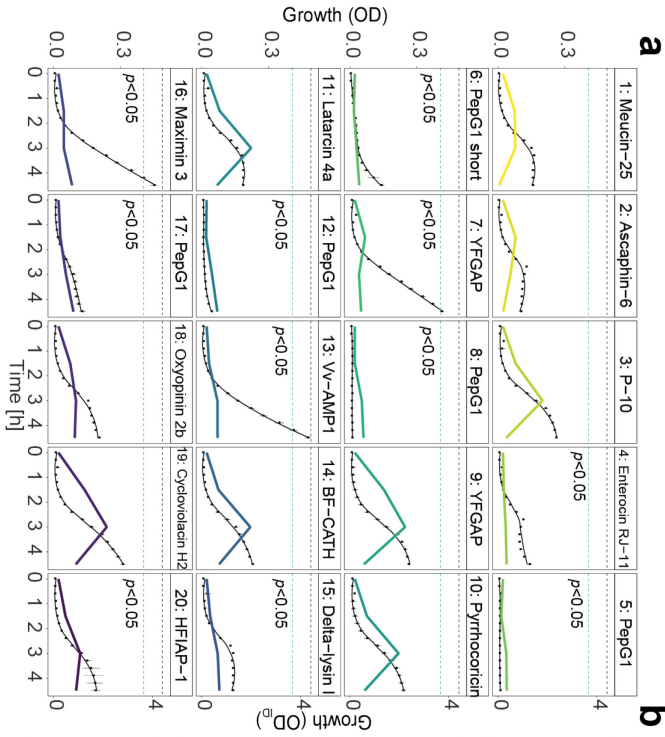
Rank by OD_ID (4.5 h)	Peptide	Origin	ID	*MIC [µM] in MHB 25%MHB	MIC [µM] in MHB
1	Meucin-25	APD	11598	>65	>65
2	Ascaphin-6	APD	9286	0.8	1.4
3	P-10	APD	8942	>70	0.7
4	Enterocin RJ-11	NCBI	3780	0.5	1
5	PepG1	NCBI	11834	1	1
6	PepG1	NCBI	11828	p.f.	p.f.
7	YFGAP	NCBI	8112	>72	>72
8	PepG1	NCBI	11833	>60	p.f.
9	YFGAP	NCBI	8135	>60	p.f.
10	Pyrrhocoricin	NCBI	7122	20	15
11	Latarcin 4a	APD	5147	8	5
12	PepG1	NCBI	11836	p.f.	p.f.
13	Vv-AMP1	NCBI	8053	>68	>68
14	BF-CATH	APD	9639	1	1
15	Delta-lysin I	NCBI	3458	>75	>75
16	Maximin 3	NCBI	5468	>71	>71
17	PepG1	NCBI	11827	p.f.	p.f.
18	Oxyopinin 2b	APD	9690	0.5	0.5
19	Cycloviolacin H2	NCBI	10889	p.f.	p.f.
20	HFIAP-1	NCBI	4545	0.5	0.5

HFIAP-1 NCBI 4545 tested against	ATCC number	MIC [µM] in MHB
Staphylococcus aureus	29213	1.4
Pseudomonas aeruginosa	47085	0.7
Enterococcus faecalis	29212	5.6
Klebsiella pneumonia	13883	0.4
Escherichia coli	25922	0.7

Fig. 2.2 | Characterization of the 20 most active peptides in Me^x a) Growth curves of the 20 most active peptides. Colored lines are Me^x-recorded growth curves (average of n=3) determined via OD$_{ID}$ approximation (header: 'rank: PARENT name'). Black lines are growth curves (n=3, error bars: 2σ) determined via OD measurement in microtiter plates of individually grown strains. Horizontal dashed lines, in black (OD) or colored in green (OD$_{ID}$), show final values measured 4.5 h post-induction of a strain synthesizing the inactive control peptide HNP-1$_{3425\ APD}$ (obtained from Supplementary Fig. 2.14). In each facet, we state if we obtain a $p<0.05$ (Wald's test) for significant growth inhibition after 1.5 h in Mex. **b)** Potential mechanisms of action. Each radar plot shows the mean SOS-response (DNA; activation of the *recA* promoter; n=3), translation inhibition (Translation; activation of the *cspA* promoter; n=3), and membrane-damage (Membrane; PI stained cells in percent; n=2) obtained after peptide expression in *E. coli* TOP10. Only the maximum and minimum values are reported in digits. The center represents values measured for the negative control peptide HNP-1$_{3425\ APD}$. Lower values are scaled to the center. Membrane damage is attributed if more than 10% of cells were PI-positive (underlined). For SOS and Translation, signals are reported relative to the signal obtained for the inactive control peptide HNP-1$_{3425\ APD}$. A significant increase (one-sided t-test, adj. $p<0.05$) compared to the inactive control is indicated by an asterisk (*). **c)** Antimicrobial activity in clinically relevant assays. Mean MIC-values are recorded (n=3) in microtiter plate assays using chemically synthesized peptides. (*) = against the screening strain *E. coli* TOP10; p.f. = purification failed.

Finally, we chemically synthesized 15 out of the 20 most growth inhibitory peptides in Me^x and determined their minimal inhibitory concentrations (MIC; Fig. 2.2c) for *E. coli* TOP10. For five, no MIC was obtained; however, as four of these five were either PARENT or derived from PARENTS known to not deliver a MIC with *E. coli* (Supplementary Table 2.3), we believe that these peptides exerted activity in the cytosol (directly at the place of their synthesis) but could not reach their target when applied externally. Remarkably, ten of the 15 peptides for which MICs were recorded, very efficiently inhibited the growth of *E. coli* (MICs: 0.4 - 20 μM; mean = 3.7 μM; median = 1 μM), a concentration range that qualifies as a starting point for drug development[19]. These results indicate that even though synthesized cytosolically, Me^x-active peptides also strongly inhibited growth when added to cells externally. Additionally, we selected the most active SIMILAR HFIAP-1$_{4545\ NCBI}$ and measured activity against other clinically relevant Gram-negative and -positive bacteria. SIMILAR HFIAP-1$_{4545\ NCBI}$ inhibited growth of these strains (MICs: 0.4 - 5.6 μM; Fig. 2.2c), which suggests a broad activity spectrum even though Me^x had been performed using *E. coli*.

Taken together, Me^x is suitable for the rapid discovery of naturally-occurring, and functionally-diverse AMPs. We argue that Me^x will enable de novo design or optimization of peptides by directed evolution approaches and we envision its application also in drug resistant (e.g. *Pseudomonas aeruginosa* or *Acinetobacter baumannii*) or recombinant strains used as reporters for microbial models relevant

2.4 Methods

Chemicals and reagents

Unless otherwise stated, all chemicals, reagents, and primers were obtained from Sigma Aldrich (Buchs, CH). Restriction enzymes and their buffers were obtained from New England Biolabs (Ipswich, USA). Synthetic genes were obtained from Integrated DNA Technologies (Leuven, BE) or Twist Bioscience (San Francisco, USA). Kits for plasmid isolation and DNA purification were obtained from Zymo Research (Irvine, USA). Peptides in either purified (>90%) or crude format were obtained from Pepscan (Lelystad, NL). Sanger-sequencing was done at Microsynth (Balgach, CH).

Bacterial strains and cultivations

Unless otherwise stated, all experiments were performed using *Escherichia coli* TOP10 (F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L(Str^R) *end*A1 *nup*G; Thermo Fisher Scientific, Waltham, USA). In this study, all cultivations were performed either in 14 ml polypropylene tubes (Greiner, Kremsmuenster, AT), filled with 5 ml of lysogeny broth (LB) medium (Difco, Becton Dickinson, Franklin Lakes, USA), or in 96-deep-well polypropylene plates (Greiner, Kremsmuenster, AT) filled with 500 µl of LB-medium. All samples were incubated at 37°C with agitation on a shaker (Kuhner, Birsfelden, CH) operated at 200 r.p.m. and 25 mm amplitude. All media were supplemented with the appropriate antibiotic for plasmid maintenance (50 µg ml⁻¹ kanamycin; 100 µg ml⁻¹ carbenicillin) and 1 % (w/v) D-glucose for repression of gene expression from catabolite-repression sensitive promoters such as P$_{BAD}$. In the case of peptide expression experiments, cultures were incubated without D-glucose and 0.3 % (w/v) of the inducer L-arabinose was used for induction. For all cultivations on solid medium, 15 mg ml⁻¹ agar (Difco) was added to the broth, and incubation was performed without shaking in an incubator (Kuhner) at 37°C. If not indicated differently, the optical densities (OD) of bacterial cultures were determined by measuring light scattering at 600 nm using a UV/VIS spectrophotometer (Eppendorf, Hamburg, DE).

In silico generation of peptide library

We collected all peptide sequences (called "PARENTS") available on the 'AMP Database' (APD)[34] in May 2017 (http://aps.unmc.edu/AP/main.php, or https://wangapd3.com/main.php). These sequences were used as input queries to find sequence-similar peptide sequences in the NCBI non-redundant nucleotide collection (nr/nt)[77], a collection that holds sequences from GenBank, European Molecular Biology

Laboratory (EMBL), DNA Databank of Japan (DDBJ), and Reference Sequence database (RefSeq), as well as translated protein information from the protein database (PDB). By applying *tblastn*[102], 170,300 additional peptide sequences (called SIMILARS) were found. Because we were limited to 12,412 different peptides with a maximum length of 42 amino acids (the chosen platform for the synthesis of the peptide-encoding oligonucleotides allowed 12'412 different sequences with a maximal length of 170 bases), we discarded SIMILARS with sequence similarity to the respective parent of less than 62.2%. The following parameters were used for the *tblastn* search: maximum sequences = 100; matrix = BLOSUM62; gap cost = 11.1; word size = 6; active low complexity filter; adjustment = conditional compositional score matrix adjustment.

Sequence distance among PARENTS and SIMILAR

To visualize sequence diversity among PARENTS, we created a sequence-based phylogenetic tree. We performed pairwise global alignment of all PARENT sequences using the Needleman-Wunsch algorithm, as implemented in the R Bioconductor package 'Biostrings' (https://bioconductor.org/packages/release/bioc/html/Biostrings.html). The BLOSUM62 substitution matrix was used to compute the alignment scores, which were converted into pairwise distances following the method Scoredist[103]. Based on the pairwise distances between PARENTS, we used hierarchical clustering with average linkage to compute a dendrogram of sequences reflecting their similarities. PARENTS and their *tblastn*-derived SIMILARS were consolidated into groups, which were named after the PARENT from the APD (http://aps.unmc.edu/AP/main.php, or https://wangapd3.com/main.php). In the sequence-based phylogenetic tree, each SIMILAR was stacked on top of its PARENT at the tip of the dendrogram. A SIMILAR may appear multiple times if it was found multiple times in the *tblastn* search using different PARENTS.

Peptide-encoding DNA architecture

The corresponding oligonucleotide sequences of the peptide library were synthesized using microarray technology supplied from CustomArray Inc. (now GeneString, Piscataway, USA). The chosen platform allowed 12'412 different oligonucleotides with a maximal length of 170 bases. A generic oligonucleotide design employing four functional units was created (Supplementary Fig. 2.3): A coding unit, a filler unit, and two universal units for amplification. This process was automated for each sequence by using an in-house written script in R. The coding unit contained the reverse translation of the peptide amino acid sequence into a codon-optimized DNA for *E. coli*. We always chose the most abundant codon for each amino acid. In cases in which restriction sites had been introduced that could potentially interfere with subsequent manipulations, the crucial codon

was replaced by the second most abundant one for this amino acid. The filler sequence was added to compensate for the various lengths of peptide genes (shortest coding sequence = 15 nucleotides, longest coding sequence = 126 nucleotides) and adjust the total of filler and coding unit to 129 nucleotides for all members of the library. To do so, we first added a UAA stop codon to the end of the coding sequence and then added downstream a semi-random sequence, ensuring a GC content of 40% for the filler sequence and limiting the number of identical nucleotides following each other to three. By adding this filler sequence we maximized sequence disparity at the DNA level (many coding sequences are homologs) thereby potentially increasing both synthesis and, later, sequencing quality. Two amplification units, of 23 and 18 bases, respectively, were appended upstream and downstream of the coding sequence and filler unit and contained the ribosomal binding site and restriction sites for the enzymes PstI and HindIII. Two amplify the peptide-encoding DNA, primer 1: CTGCACAAAGCTTACGTG, complementary to the upstream amplification unit, and primer 2: CACGTAAGCTTTGTGCAG, reverse complementary to the downstream amplification unit were used. The final 170 bases long oligonucleotide sequences as synthesized are listed by ID in Supplementary table 2.2 (erroneous sequences were discarded).

Peptide-encoding DNA cloning

The chemically synthesized and single-stranded oligonucleotides were separated from their array and we received them as a pool. This pool was aliquoted in 10 mM Tris-Cl, 1 mM EDTA, pH 8 and deep-frozen at -80°C. The pool was amplified by polymerase chain reaction (PCR) in a 50 µl reaction using 5 ng of the template and 10 µM HPLC-purified primer 1 and primer 2, complementary to the amplification sites, and 25 µl of Phusion® High-Fidelity PCR Master Mix with HF buffer. The amplification was performed using 25 cycles of 98°C for 15 s, 55°C for 20 s, and 72°C for 5 s. The now double-stranded peptide-encoding DNA sequences were purified using a DNA purification kit. DNA concentration was measured using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific) and 500 ng of the purified product was used for a restriction digest using enzymes HindIII-HF and PstI-HF in Cutsmart buffer. The digested product was again purified using a DNA purification kit and ligated to plasmid pBAD[104] (Thermo Fisher Scientific) digested with the same enzymes. This plasmid harbored the tightly controllable P_{BAD} promoter for peptide gene expression, a pBR322 replication of origin, and a resistance gene encoding for beta-lactamase. For ligation, pBAD was purified using a 1% agarose gel and a DNA gel recovery kit after digestion. Next, T4 ligase (800 units) was used to ligate 100 ng of cut pBAD vector and 10 ng peptide-encoding DNA sequences in T4 ligase buffer (molar ratio of 7:1 insert:vector). The ligation mix was incubated for 14 h at 16°C. The ligation product

was dialyzed in deionized water using filters (MilliporeSigma, Burlington, USA) and 1 µl of the mix was used to transform 20 µl of CloneCatcher™ Gold DH5G Electrocompetent *E. coli* (Genlantis, Burlington, USA) cells using electroporation. Recovered cells were plated and incubated overnight on LB agar plates supplemented with carbenicillin. Afterward, ~500,000 colonies were washed off the plates using LB medium, and the plasmids containing the peptide-encoding DNA sequences were extracted from $2.5*10^9$ cells using a plasmid isolation kit. An aliquot of 5 ng of these plasmids was used to transform *E. coli* TOP10 cells using the protocol from the transformation above. A total of 1'000'000 colonies were recovered from the plates after overnight incubation by washing with LB medium, the suspension was diluted to OD = 1 with LB-medium, glycerol was added to a final concentration of 20% (v/v), and aliquots of 500 million cells were stored at -80°C.

Growth experiment

Three aliquots of *E. coli* TOP10 harboring the peptide-encoding DNA sequences on the pBAD plasmid were thawed and added to three 1 l baffled shake flasks containing 100 ml of LB medium + 100 µg ml^{-1} carbenicillin. The cultures were grown for roughly 7.5 h at 37°C. When the OD reached 0.2, the cultures were supplemented with L-arabinose to a final concentration of 0.3 % (w/v) to induce peptide expression. Cell samples were taken from each biological replicate at the point of induction and 1.5 h, 3 h, and 4.5 h post-induction. The plasmids were extracted from all samples using a plasmid isolation kit.

NGS

For the generation of Me^x growth curves, peptide-encoding DNA sequences on plasmids, collected from the three replicates across four time points during the growth experiment, were sequenced by NGS. Additionally, the abundance of peptide-encoding DNA sequences in the original oligonucleotide pool and after transformation of the assay strain *E. coli* TOP10 was assessed by NGS as well. Peptide-encoding DNA sequences were amplified by primer 1 and primer 2 using 100 ng of plasmid and the PCR-amplification protocol mentioned before, but only for 10 cycles to avoid amplification bias. The amplification product was purified using an agarose gel. Single Index PentAdapters from Pentabase were used to prepare PCR-free libraries with the KAPA HyperPrep Kit (now Roche, Basel, CH) according to the manufacturer's specifications. Libraries were quantified using the qPCR KAPA Library Quantification Kit. Libraries were pooled and sequenced PE 2x151 with an Illumina HiSeq 2500 using v4 SBS chemistry. Roughly 10% genomic PhiX library as spike-in to increase sequence diversity. Basecalling was done with bcl2fastq v2.20.0.422. The resulting fastq files were processed using in-house software written in R and C. This software aligns each sequence to our reference table of

12'412 sequences linking peptide-encoding DNA sequences and peptide sequence, identifies mismatches and sequencing errors, and counts how often each peptide-encoding DNA sequence was sequenced in each sample. NGS read counts for each sequence analyzed in Me^x were listed with a unique identifier (ID) in Supplementary Table 2.2.

Generation of Me^x growth curves

We used the standard workflow of DESeq2[105] (NGS read count normalization, dispersion estimates, and Wald's tests) to analyze NGS read counts. Only sequences that passed independent filtering were included in further analyses (= 10,633). To draw growth curves for each peptide-expressing strain, we calculated the log2-fold changes of NGS read counts (listed for each ID in Supplementary table 2.2) between the time of induction and all other time points (1.5 h, 3.0 h, and 4.5 h post-induction). A Bayesian shrinkage estimator was employed to shrink the log2 fold-change for each ID (lfcShrinks$_{ID}$) between all time points using the R/Bioconductor package 'apeglm'[106]. To draw the Me^x growth curves, we calculated a strain-specific OD$_{ID}$ at each time point according to equation (1). OD values at the specific time points were averaged values from all three biological replicates (Supplementary Fig. 2.5). The OD$_{ID}$ (0 h) for each peptide-expressing strain was set to 0.2 at the time of induction as lfcShrink$_{ID}$ (0 h) = 0 and OD = 0.2. This enabled us to compare peptide-expressing strains of different abundancies (see Supplementary Fig. 2.6). OD$_{ID}$ values can be interpreted as the OD values that would have been measured when incubating the respective strain individually in the same experiment, i.e. in this case in LB medium in a 100 ml shake flasks.

$$OD_{ID}(t) = OD(t) \times 2^{lfcShrinkID(t)} \qquad (1)$$

To find Me^x-active peptides, we also performed a one-sided Wald's test, with the alternative hypothesis that the expression of a given peptide leads to a reduced OD$_{ID}$ 1.5 h and 4.5 h post-induction. We rejected the null hypothesis at significance level alpha = 0.05. Peptides with a $p<0.05$ (after adjustment for multiple testing using the Benjamini-Hochberg method) after 4.5 h are considered Me^x-active peptides. Peptides with $p<0.05$ after 1.5 h do significantly inhibit growth already after 1.5 h. All values and results are reported in Supplementary table 2.1.

Monoseptic growth experiments

Taking the OD$_{ID}$ (4.5 h) of each peptide-expressing strain, we could rank all peptides by their growth inhibitory effect. We selected 110 peptides (Ranks 1-50, 100-119, 1000-1019, and 10,000-10,019) and then generated an identical copy of the strain previously used in

Me^x for its expression. First, the corresponding peptide-encoding DNA-sequences were synthesized as gene fragments. An aliquot of 400 ng of each gene fragment was directly used for a restriction digest using enzymes HindIII-HF and Pst-HF in Cutsmart buffer. The product was purified using a DNA purification kit. Next, T4 ligase (800 units) was used to ligate 50 ng of identically digested pBAD vector and 10 ng of purified gene fragment in T4 ligase buffer for 14 h at 16°C. The ligation product was purified using a DNA purification kit. An aliquot of 5 µl of the purified ligation product was then used to transform chemically competent *E. coli* TOP10 cells. From the resulting colonies, we isolated one strain, sequence-verified the correct assembly of the expression plasmid, and stored it after overnight growth in glycerol at -80°C. For the growth experiment, we first re-isolated single colonies on solid media and then picked three clones, incubated them separately overnight, and inoculated them into 200 µl fresh LB medium containing 0.3 % (w/v) L-arabinose to a final OD of 0.01 into 96-well microtiter plates (Greiner). Growth was recorded by measuring OD in a Tecan Infinite 200 PRO (Tecan, Männedorf, CH) for 4.5 h (37°C, 1.5 mm orbital shaking).

Enrichment analyses

We used Fisher's exact test to assess the over- or underrepresentation of Me^x-actives in various groups. This amounts to a hypergeometric test to assess the significance of drawing n active peptides in a group of k, from a population of size N containing K active peptides. We rejected the null hypothesis at significance level alpha = 0.05. Groups with a $p<0.05$ had a significantly different representation of active peptides compared with the overall population. When adjusting for multiple testing, we used the Benjamini-Hochberg method.

Peptide classifications

The physicochemical parameters of the peptides were calculated at pH 7 using the R package 'Peptides' (https://cran.r-project.org/package=Peptides). For charge, we used the method by Lehninger[107]. For hydrophobicity, we used the calculations by KyteDoolittle[108]. The information for each PARENT such as the name, chemical modification, activity, 3D-structure, was extracted from the APD website (http://aps.unmc.edu/AP/main.php, or https://wangapd3.com/main.php) using an in-house R script. The information on the species from which a specific peptide sequence originated, was extracted from the *tblastn* search and the APD website. The entire taxonomic classifications (kingdom, phylum, class) for each species were extracted, if available, from the Global Biodiversity Information Facility Data Portal (https://gbif.org) using the R package 'taxize'

). The results are summarized in Supplementary Table 2.1.

Purification of chemically synthesized peptides

Peptides were obtained from Pepscan (Lelystad, NL) in >90% purity or in crude format and subsequently purified to >90% purity in-house. For the latter, crude peptides were dissolved in 5 ml DMSO and 15 ml 0.1% aqueous trifluoroacetic acid, TFA. HPLC-purification of the dissolved crude peptides was performed on an ÄKTAexplorer chromatography system (GE Heathcare, SE). The entire peptide sample was loaded onto a RP C18 column (PRONTOSIL 120 C18 10 μm, 250 x 20 mm, 50 x 20 mm precolumn, Bischoff, Leonberg, DE), heated to 30°C and operated at a flow rate of 10 ml min^{-1} using 0.1% aqueous TFA as solvent A and acetonitrile supplemented with 0.1% TFA as solvent B. The ratios of A to B were adapted for each peptide and typical values are given below. The column was equilibrated with the peptide-specific mixture of solvent A and solvent B 0-20%) prior to injection. After injection and an initial wash step of 6 min a gradient was imposed with the same mixture, and then a gradient was applied, in the course of which the amount of solvent B was increased to 50-90 % in 40 min. The column was washed with 95 % solvent B for 8 min and equilibrated with the specific solvent A/solvent B mixture for the next run for 13 min. Peptide elution was monitored spectrophotometrically at 205 nm, and generally the main peptide peak was collected. The sample was frozen at -80 °C for >2 h and lyophilized (approx. 18 h) using a freeze-dryer (Alpha 2-4 LDplus, Christ, DE), connected to a vacuum pump (RC6, Vacuubrand, DE). The lyophilized peptides were dissolved in 1 ml DMSO and stored at -20 °C. The concentration of the peptide stocks was determined via HPLC using an Agilent 1200 series HPLC system. Each peptide stock was analyzed as a 1:100 dilution in water. An aliquot of 10 µl of the peptide stock was injected onto an RP-C18 column (ReproSil-Pur Basic C18, 50 x 3 mm, Dr. Maisch, Germany) operated with water supplemented with 0.1 % TFA as solvent A and acetonitrile supplemented with 0.1 % TFA as solvent B. Separation was performed using the same concentration profile previously used for purification. The concentration was measured using the integrated peak area at 205 nm and then calculated using peptide-specific absorption properties[109,110].

Measurement of the MIC

On the same day at which MIC assays were executed, purified peptides were thawed and the concentration was determined by HPLC as described before. *E.coli* TOP10 cells were grown in Mueller Hinton Broth (MHB) or diluted MHB (25 % of the original strength) overnight to stationary phase. Diluted MHB has been frequently used to assay AMPs[111].

The cultures were then supplemented with 20 % glycerol, aliquoted, and frozen at -80°C. For MIC measurements, a frozen stock of the cells was thawed, resuspended in MHB or 25% MHB to adjust to a density of $5*10^5$ cells ml^{-1} in the experiment, and distributed to microtiter plate wells by an automated liquid handling system (Hamilton, Bonaduz, CH). Then the peptides were added by the liquid handling system in 2-fold dilutions using minimum of 100 µg ml^{-1} as the highest concentration. MICs were determined as broth microdilution assay in 384-well flat bottom polypropylene plates (Falcon® 96-Well Flat-Bottom Microplate) adapted from the protocol of Wiegand *et al.*[112]. The plates were sealed airtight and incubated for 18 h without shaking at 37°C before reading the OD using a Tecan Infinite 200 PRO plate reader. The MIC value corresponded to the concentration at which no growth of the bacterial strain was observed (< 5% of the OD value of the growth control). MIC experiments were performed at least in triplicate.

Membrane damage assay

We selected the peptide-expressing strains of rank 1-50 in Me^x that we had previously constructed for the monoseptic growth assay. Additionally, we selected the strain expressing the inactive control peptide HNP-1$_{3425\ APD}$, a peptide known to be inactive if expressed in *E.coli*[91]. Each strain was re-isolated on solid media from frozen stock and incubated overnight. Then, two colonies were picked and incubated overnight in 96-deep-well polypropylene plates. These cultures were used to inoculate fresh media containing 0.3 % (w/v) L-arabinose to a final OD of 0.01 into 96-well microtiter plates. The plates were then incubated on for 4.5 h (37 °C, 1.5 mm orbital shaking). After 4.5 h, an aliquot of 50 µl of cell suspension a Tecan Infinite 200 PRO plate reader was added to 150 µl of phosphate-buffered saline into a fresh 96-well microtiter plate. Propidium iodide (PI) was added to a final concentration of 1 µg ml^{-1}. PI is a DNA-intercalating dye that cannot pass an intact cytoplasmic membrane[113]. For each sample, PI fluorescence (λ_{Ex}= 579 nm / λ_{Em}= 616 nm) of ~10,000 cells were analyzed using a flow cytometer LSR Fortessa (BD Biosciences, Allschwil, CH). To determine the membrane damaging properties of each of the expressed peptides, we calculated the fraction of cells in percent for which a PI uptake was measured using the software FlowJo V10 (BD Biosciences).

Stress response assay

We selected peptide-expressing strains of rank 1-50, previously generated for the monoseptic growth assay. Additionally, we selected the strain expressing the inactive control peptide HNP-1$_{3425\ APD}$. Moreover, two plasmids (cloning vector: puA66) containing either the promoter of the gene for recombinase A (P$_{recA}$) or for the gene for cold shock

protein A (P_{cspA}) were purified from the *E. coli* Alon collection[114]. Both plasmids contained a transcriptional fusion of their promoter with a downstream gene for green fluorescent protein (*gfp*), an additional kanamycin resistance cassette, and a pSC101 origin of replication. We transformed each of the 51 peptide-expressing *E. coli* strains with each of the two plasmids to generate 102 different strains and incubated them overnight on solid media. Then, three colonies were picked and incubated overnight. These cultures were used to inoculate fresh media containing 0.3 % (w/v) L-arabinose to a final OD of 0.05 into 96-well microtiter plates. We recorded OD and GFP expression (λ_{Ex} 488 nm/λ_{Em} 530nm) after 1.5 h and 4.5 h using a Tecan Infinite 200 PRO (37 °C, 1.5 mm orbital shaking). For each strain, we calculated the specific fluorescence change between the two time points (GFP/OD (4.5 h) - GFP/OD (1.5 h)). Statistical significance was calculated by one-sided t-tests, adjusted for multiple testing by Benjamini-Hochberg, using the signal of HNP-1$_{3425}$ $_{APD}$ as null distribution. We rejected the null hypothesis at significance level alpha = 0.05.

Data availability

All relevant information for each peptide analyzed in M*e*^{x} (amino acid sequences, physiochemical properties, taxonomical information, p-values, OD$_{ID}$-values, link to APD) is provided in Supplementary Table 2.1. NGS read counts and all peptide-encoding DNA sequences are provided in Supplementary Table 2.2. Tables are available at https://polybox.ethz.ch/index.php/s/2t4a886WXCOZqaA using the password BPL2021+MeX. The computational workflow to reproduce the analysis of Supplementary Table 2.1 and generate growth curves of each peptide expressing-strain is available on GitHub (https://github.com/derpkoch/MeX) and in the Supplementary code section. NGS data are available at the NCBI Sequence Read Archive (SRA) under accession number PRJNA686958. Additional data that support the findings of this study are available from the corresponding author upon reasonable request.

2.5 Supplementary figures
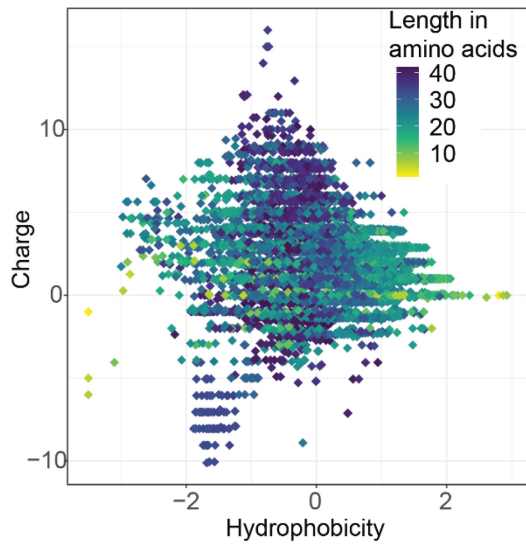
Supplementary Fig. 2.1 | Predicted physiochemical properties of peptides in the library. All peptides of the library are plotted according to their charge and hydrophobicity at pH 7 and colored by their length. Mean charge = +2.3; mean hydrophobicity (GRAVY scale) = 0.0; mean length = 27 amino acids.
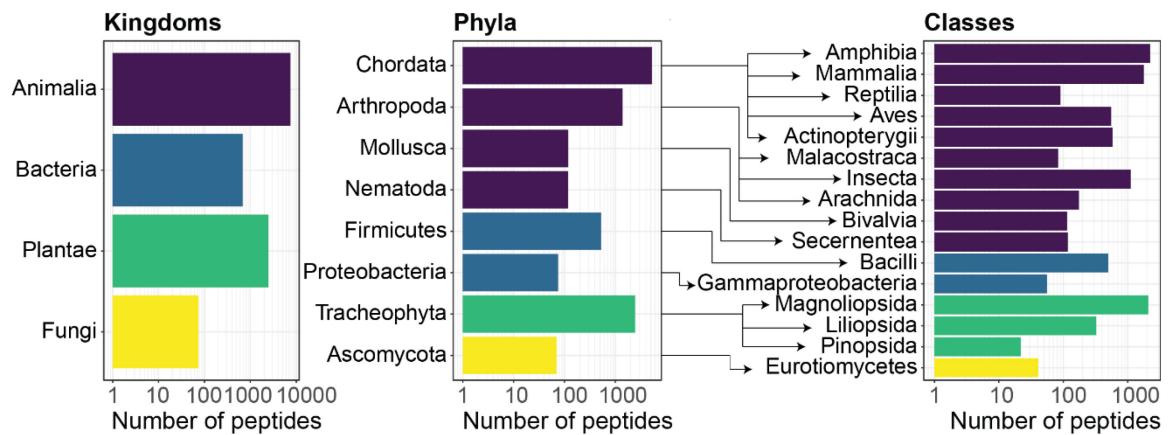
Supplementary Fig. 2.2 |Taxonomical classification of the peptide library. All peptides of the library are grouped by their taxa in rank kingdom, phylum, and class of the host from which their sequences had been derived. Only groups comprised of at least 20 peptides are displayed. Phyla and classes are colored by their kingdom (left).

Supplementary Fig. 2.3 | Sequence architecture of peptide-encoding DNA sequences as used for synthesis. Each DNA sequence contains a peptide coding sequence (green), a unique filler sequence (black) used for standardizing sequence length to 170 nucleotides, and two universal amplification sites (yellow) used for both, cloning and amplification. The coding sequences are generated by reverse translation of the respective peptide amino acid sequence followed by codon optimization for expression in *E. coli*. Amplification sites at the 5' and 3' end are the same for all inserts and contain restriction sites for subsequent integration into the multiple cloning site of the expression plasmid.
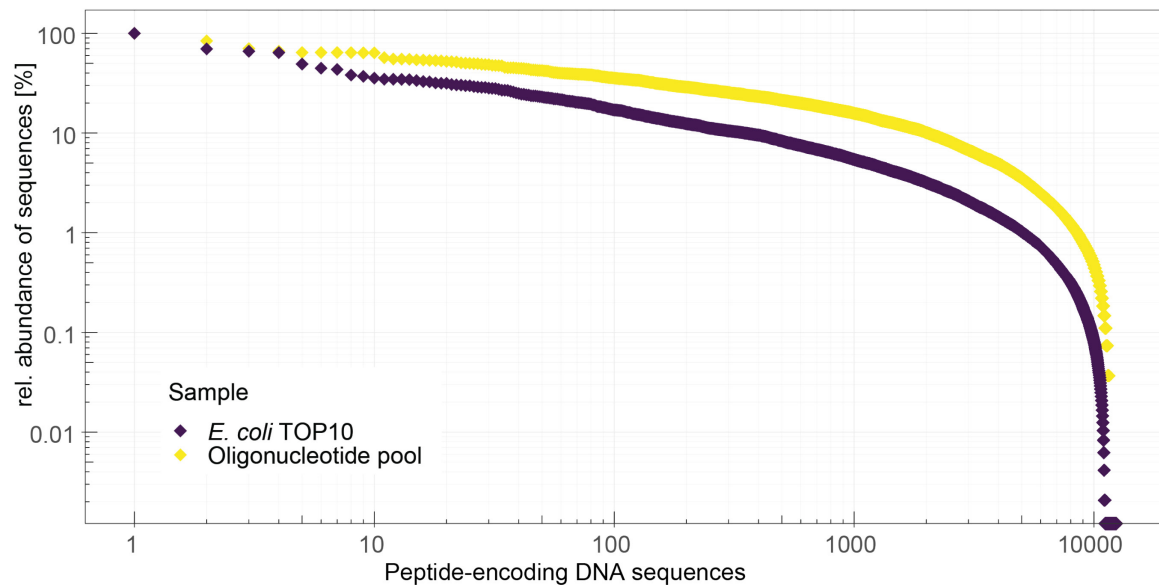
Supplementary Fig. 2.4 | Relative abundance of peptide-encoding DNA sequences before and after cloning. Peptide-encoding DNA sequences in the original synthesized oligonucleotide pool (yellow) and after insertion into a plasmid and transformation of *E. coli* TOP10, as used for the growth experiment (purple) are counted by NGS. All counts are relative to the most abundant peptide-encoding sequence Chensinin-1CEb$_{2720 \, NCBI}$, which appears 2,720 times (oligonucleotide pool) and 5,466 times (growth experiment), respectively.
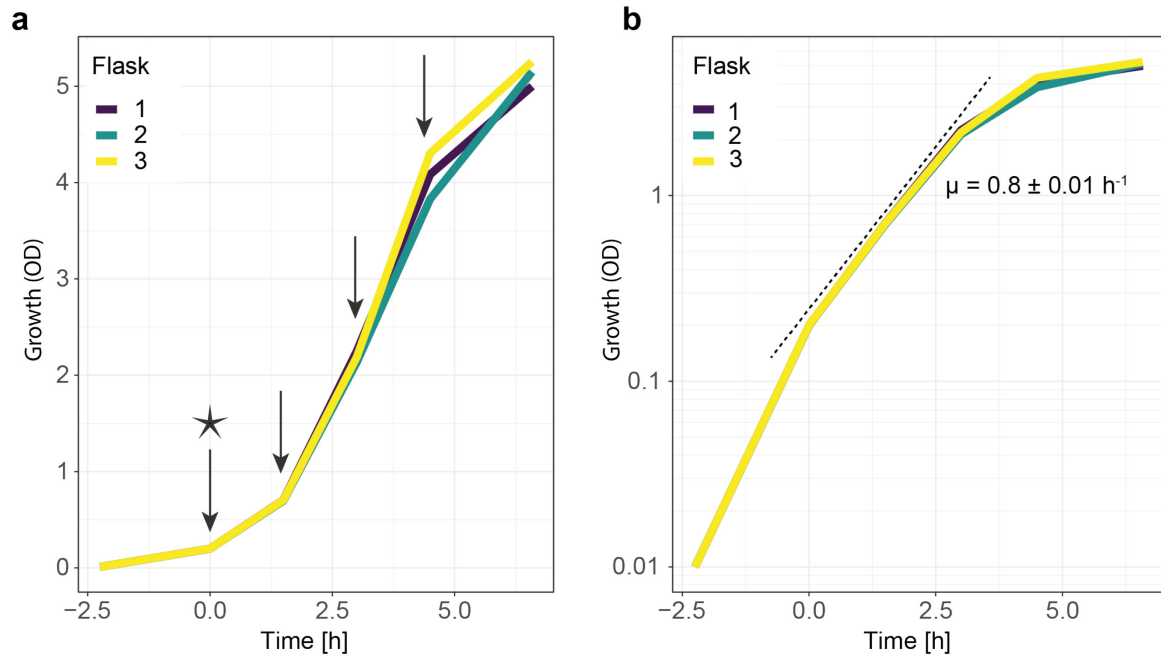
Supplementary Fig. 2.5 | Growth of *E. coli* TOP10 expressing the peptide-encoded DNA library. **a)** Optical density at 600 nm (OD) is recorded over 8 h. Three 1 liter shake flasks containing 100 ml of LB-medium each are inoculated with 500 million cells of *E. coli* TOP10 carrying the peptide-encoded DNA library at -2.5 h (time reported relative to the time of induction). Peptides are expressed after 4 generations (0.0 h; OD~0.2) by adding L-arabinose (asterisk). Cell samples for NGS are isolated from each replicate at the time of induction, and 1.5 h, 3.0 h, and 4.5 h post-induction (arrows). **b)** Log10 transformed data of (a) for calculation of the specific growth rate (μ).
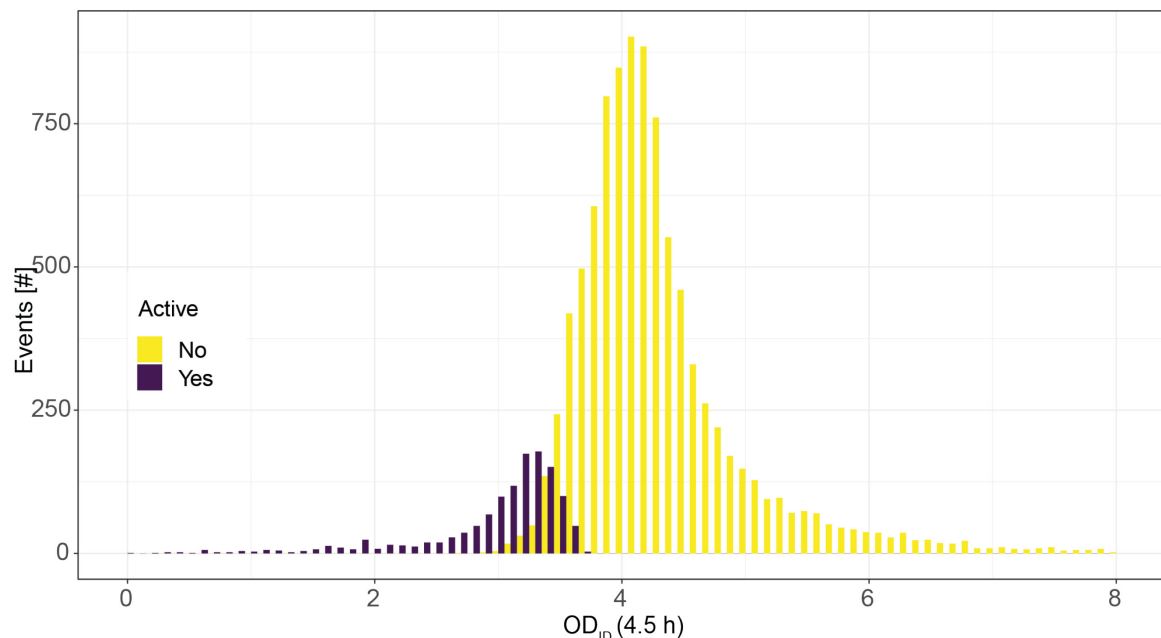
Supplementary Fig. 2.6 | Growth inhibition as determined by OD_{ID} at 4.5 h recorded for 10,663 peptide-expressing strains. OD_{ID}-values are recorded for each of the peptide-expressing strains and are averaged from the three replicates. Me^x-active peptides (purple) significantly (Wald's test, adj. $p<0.05$) reduce the OD_{ID} of their expressing strain after 4.5 h while Me^x-inactive peptides (yellow) fail to do so. Note that some candidates may also fail to reach statistical significance in the performed Me^x-activity test due to low NGS read counts or high variance between replicates (Supplementary Table 2.2).

Supplementary Fig. 2.7 | Growth of peptide-expressing *E. coli* strains in Me^x and in monoseptic cultures. **a)** All peptide-expressing strains are ranked by their growth inhibition after 4.5 h (= ranked by their OD$_{ID}$ (4.5 h)) and selected representatives were then subdivided into four groups: purple = rank 1-50 (average OD$_{ID}$ (4.5 h) = 1.0); blue = rank 100-119 (average OD$_{ID}$ (4.5 h) = 2.0); green = rank 1,000-1,019 (average OD$_{ID}$ (4.5 h) = 3.3); and yellow = rank 10,000-10,019 (average OD$_{ID}$ (4.5 h) = 5.4) and HNP-$_{13425\ APD}$ (negative control, rank 2172, OD$_{ID}$ (4.5 h) = 3.7). **b)** Overlay of growth curves recorded by Me^x (colored lines; average of n=3) and during monoseptic growth of the same strains in microtiter plates (black lines; n=3, error bars = 4s). Numbers on top of each recording correspond to Me^x-derived activity rank. **c)** Boxplot of OD$_{ID}$(4.5 h) and OD (4.5h) of the peptide-expressing strains in the different subgroups identified in (a). Note the different scales on the x-axis. Rank 1-50: average OD (4.5h) = 0.2. Rank 100-110: average OD (4.5h) = 0.35. Rank 1,000-1,019: average OD (4.5h) = 0.4. Rank 10,000-10,019: average OD (4.5h) = 0.42.

a

| ID	Sequence	Class	Species
677	FSTKTRNWFSEHFKKVKEKLKDTFA	Reptilia	*Alligator sinensis*
9989	I TTKTRNWFSEAFSKMKEKLQSAFS	Mammalia	*Ochotona princeps*
691	. . TKTRTWFSETFGKLKEKLKTTF.	Mammalia	*Mesocricetus auratus*
695	LSAKTRDWFSETFRKVKEKLK.	Mammalia	*Pongo abelii*
699	. . KTRSWFSETLNKMKEKLKTTFA	Mammalia	*Rattus norvegicus*
701	LPAKTRDWFSETFRKVKEKLK.	Mammalia	*Cercocebus atys*
681	. . TKARNWFSETYGKVKEKIKTTFA	Mammalia	*Nannospalax galili*
697	FSEKTQNWFSELFHKVKEKFETTF.	Mammalia	*Pteropus vampyrus*
693	. . KTRNWFSETFNKVKERFKTAFS	Mammalia	*Ailuropoda melanoleuca*
702	. . TKTRTVLSETFDKVKEKFKTTFA	Mammalia	*Peromyscus maniculatus*
9988	FATKIKNWFSEQYEKVKQKLSESF.	Actinopterygii	*Hemibarbus mylodon*
704	LSAKMEWFSETFCKVKEKLK. . . .	/	/
9990	. STKTRNWFSETLCKVKEKLK. . . .	Mammalia	*Carlito syrichta*
709	. . . . RAWFSEAFGKVKEKLKTTFS	/	/
686	I TTKTRNWFSEITSKVKEKFKSTFS	Mammalia	*Oryctolagus cuniculus*
700	. . TKTRNWFSETFKKIKEKV.	Amphibia	*Xenopus laevis*
683	. . KTRNWFSETFNKVKEQLKTAFS	Mammalia	*Ursus maritimus*
706	LSAKMEDWFSETFCKVKEKLK. . . .	Mammalia	*Gorilla gorilla*
690	. . TKTRAWFSETFCKVKEKLK. . . .	Mammalia	*Mus caroli*
679	. . TKTRAWFSEAFGKVKEKLKTTFS	Mammalia	*Mus musculus*
9993	I PTKTRNWFSETLCKVKEKL.	Mammalia	*Callithrix jacchus*
9991	FPAKTRNWFSETFRKVKEKLK.	Mammalia	*Cebus capucinus*
696	I ATKTRDWFSETFKKVKEK.	Mammalia	*Papio anubis*
703	FPSKTRNWFSETYSKVKEKFKTTFS	Mammalia	*Propithecus coquereli*
684	. . KTRNWFSETFKKVKEQLKTAFS	Mammalia	*Rhinolophus sinicus*
680	. . KTRNWFSETHKVKEQLKTAFS	Mammalia	*Odobenus rosmarus*
694	. . KTRNWFSETFKKVKEQLKTAFS	Mammalia	*Leptonychotes weddellii*
682	I PAKTRNWFSETHKVKEQLKTAFS	Mammalia	*Odobenus rosmarus*
698	. . KTRNWFSETFKKVKEQLKTAFS	Mammalia	*Leptonychotes weddellii*
692	FSTKTRNWFKHNFCKVKDKMDATFS	Mammalia	*Phascolarctos cinereus*
689	FSEKTRNWFSEQFCKMKEKFNEQFS	Reptilia	*Chrysemys picta*
685	. . KTRNWFSETLNKMKEKLKTTFA	Mammalia	*Rattus norvegicus*
687	I PAKTRNWFSEAFKKVKEHLKTAFS	Mammalia	*Canis lupus*
688	. . KTRNWFSETFCKVKEKLK.	Mammalia	*Otolemur garnettii*
707	FADKTRHVLSETFKKIKEKLE.	Mammalia	*Galeopterus variegatus*
708	FPAKTWDWFSETFRKVKEKLK.	Mammalia	*Mandrillus leucophaeus*
678	FATKRTWFSETFTKVKEKLKDTFS	Mammalia	*Marmota marmota*
9992	I PAKTRNWFSETLCKVKEKLR.	Mammalia	*Aotus nancymaae*

Legend: non conserved; ≥ 50% conserved; all match

b

Apo5 APOC1 677 APD

Apo5 APOC1 9989 NCBI

Growth (OD); Growth (OD$_{ID}$); Time [h]

Supplementary Fig. 2.8 | Analysis of Apo5 APOC1-derived SIMILARS. a) Amino acid sequence alignment of all 36 SIMILARS of the Apo5 APOC1$_{677\ APD}$ PARENT. The inactive PARENT, derived from Chinese alligator (*Alligator sinensis*), and the only Me^x-active similar (Apo5 APOC1$_{9989\ NCBI}$) derived from American pika (*Ochotona princeps*) differ by nine amino acids. The top shows the consensus sequence plot. **b)** Overlay of growth curves recorded by Me^x (colored line, an average of n=3) and via monoseptic growth in microtiter plates (black line; n=3, error bars = 2σ), of *E. coli* TOP10 cells expressing Apo5 APOC1$_{677\ APD}$ and Apo5 APOC1$_{9989\ NCBI}$. Horizontal dashed lines, in black (OD) or colored in green (OD$_{ID}$), show final values measured 4.5 h post-induction of a strain synthesizing the inactive control peptide HNP-1$_{3425\ APD}$ (obtained from results displayed in Supplementary Fig. 2.14).
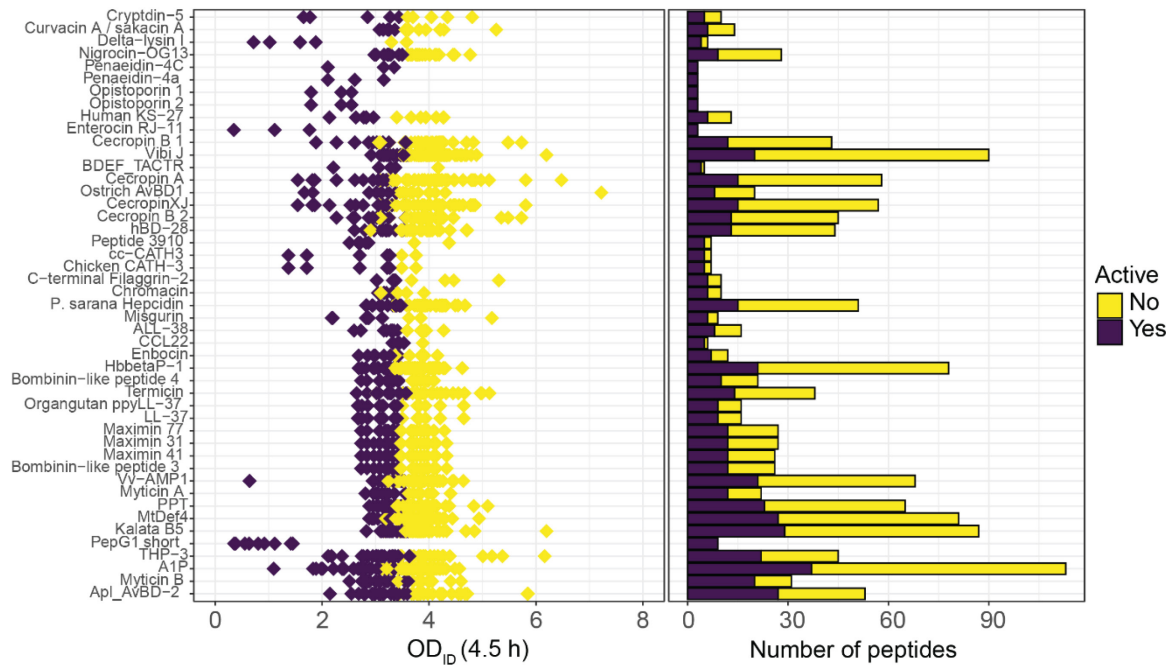
Supplementary Fig. 2.9 | Overrepresentation of active SIMILARS derived from 47 PARENTS. For 47 PARENTS (names on the left), Me^x-actives were significantly overrepresented among the SIMILARS identified in the similarity search (Fisher's exact test, adj. $p<0.05$). The OD$_{ID}$ (4.5 h) values for the individual peptide-expressing strains within a group of PARENTS and SIMILARS are shown as dots (left) and the total number of active and non-active peptides for each of the 47 parents as bars (right).
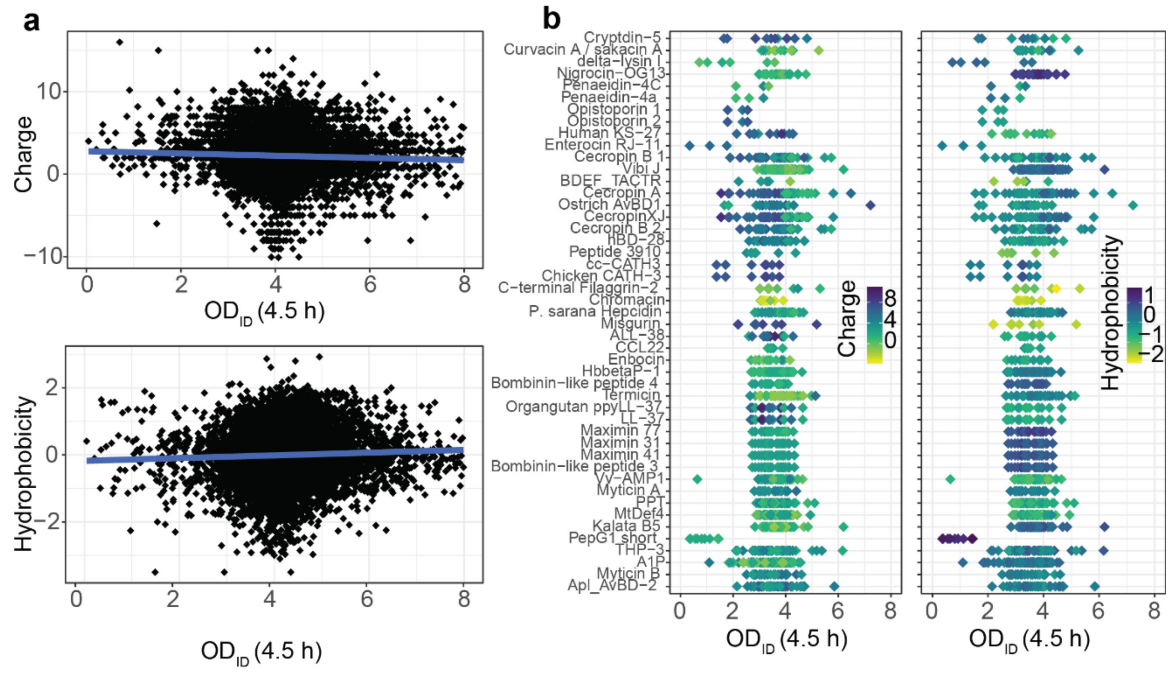
Supplementary Fig. 2.10 | Influence of the physicochemical properties of peptides on Me^x-activity. a) Charge (top) and hydrophobicity (bottom) of each peptide assayed by Me^x are plotted against OD$_{ID}$ (4.5h). Linear fits (both $p<0.01$, $R^2<0.001$, 10,661 DF) are displayed for the entire peptide library (blue line) **b)** Charge (left) and hydrophobicity (right) are displayed for the 47 groups of PARENTS and their SIMILARS containing an overrepresentation of Me^x-positives (Fisher's exact test, adj. $p<0.05$) as shown in Supplementary Fig. 2.9.
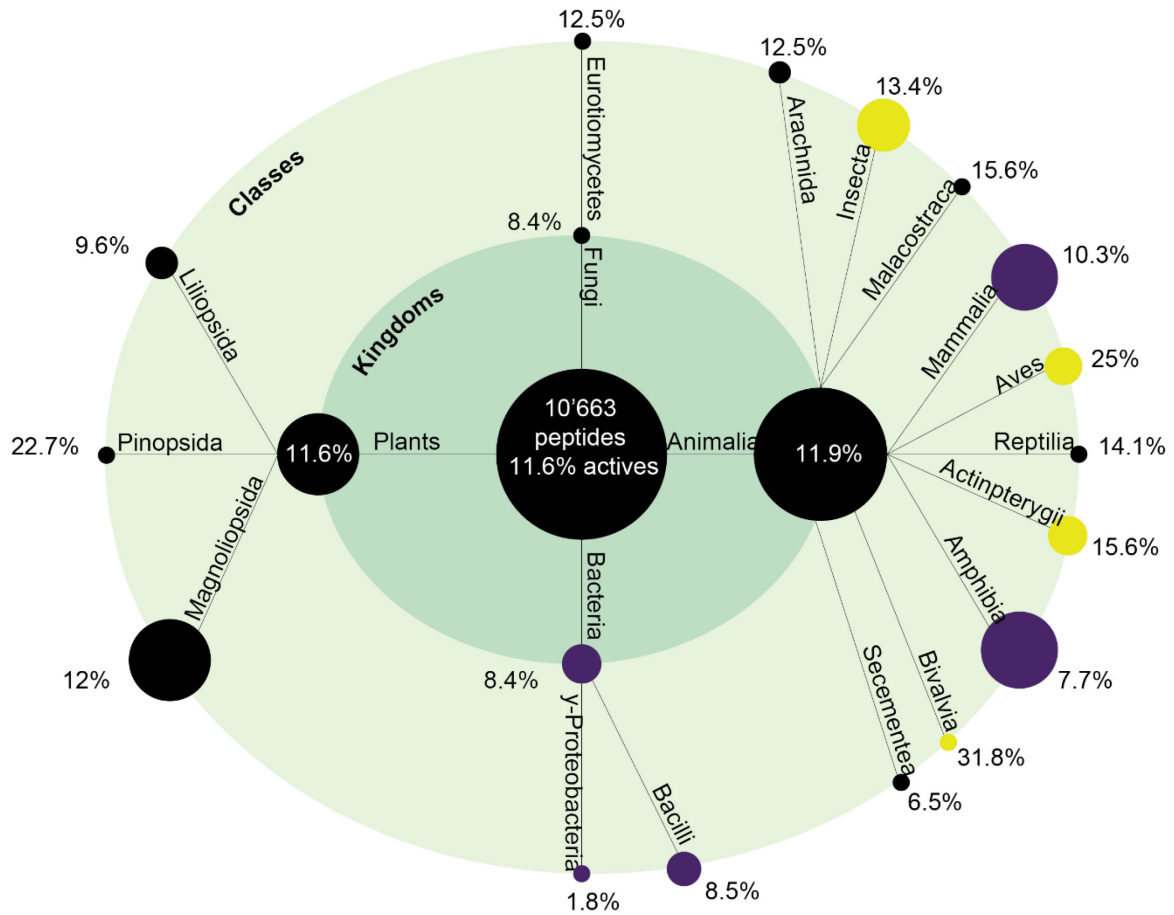
Supplementary Fig. 2.11 | Taxa-specific Me^x-activity in rank kingdom and class. All peptides tested by Me^x are clustered taxonomically according to the host from which they were derived. The percentage of Me^x-actives in the cluster is written next to (or inside) each circle. Compared to the 11.6% Me^x-actives in the entire library, taxa in which Me^x-actives are over- (yellow) or underrepresented (purple) are highlighted ($p<0.05$). Only clusters with more than 20 peptides are displayed. The circle area is representative of the total number of peptides within each cluster.
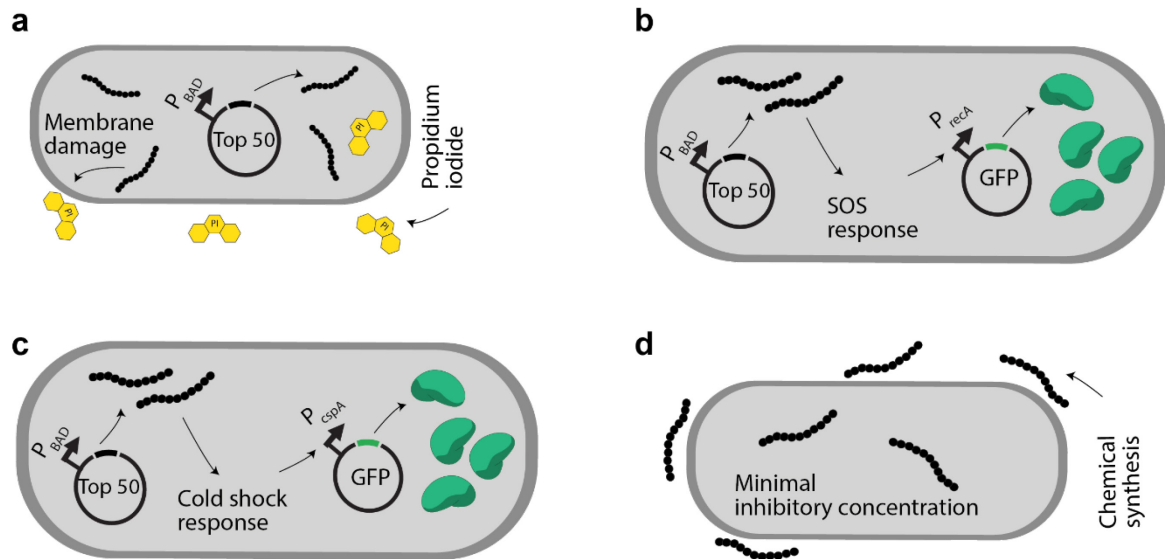
Supplementary Fig. 2.12 | Assays to characterize the antimicrobial effect. (a-c) The 50 most active peptides in Me^x (ranks 1-50; by OD$_{ID}$ (4.5 h)) are expressed intracellularly in a strain of *E. coli* TOP10 with or without a plasmid that expresses the gene for a reporter protein (green fluorescent protein, *gfg*) under the control of a promoter whose activity has been linked before to a specific stress response. **a)** Analysis of membrane damage by quantification of propidium iodide (PI) uptake. **b-c)** Quantification of GFP indicating the interference with intracellular targets by eliciting an SOS stress response indicative of DNA damage (readout via P$_{recA}$ (promoter of *E. coli*'s *recA* gene)) or a cold shock response indicative of translation inhibition (readout via P$_{cspA}$ (promoter E. coli's *cspA* gene)). **d)** Method to determine the minimal inhibitory concentration (MIC). The 20 most Me^x-active peptides are synthesized chemically, purified, and added to cultures of *E. coli* TOP10 and other pathogens.
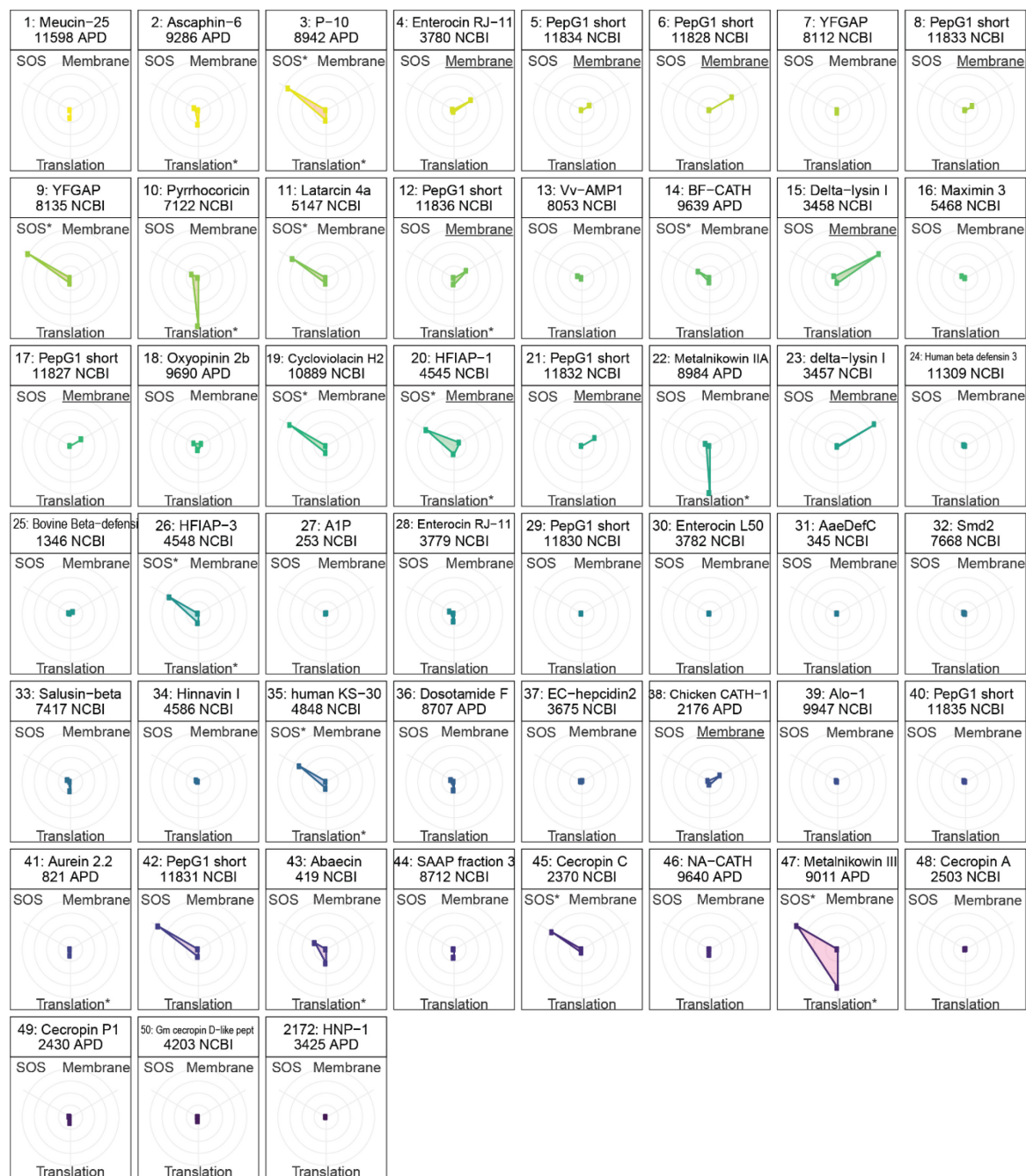
Supplementary Fig. 2.13 | Characterization of the 50 most active peptides in Me^x as addition to Fig. 2.2. Potential mechanisms of action. Each radar plot shows the mean SOS-response (DNA; activation of the *recA* promoter; n=3), translation inhibition (Translation; activation of the *cspA* promoter; n=3) and membrane-damage (Membrane; PI stained cells in percent; n=2) obtained after peptide expression in *E. coli* TOP10. Only the maximum and minimum values are reported in digits. The center represents values measured for the negative control peptide HNP-1$_{3425\ APD}$. Lower values are scaled to the center. Membrane damage is attributed if more than 10% of cells were PI-positive (underlined). For SOS and Translation, signals are reported relative to the signal obtained for the inactive control peptide HNP-1$_{3425\ APD}$. A significant increase (one-sided t-test, adj. *p*<0.05) compared to the inactive control is indicated by an asterisk (*).

43

Supplementary Fig. 2.14 | Characterization of the 50 most active peptides in Me^x as addition to Fig. 2.2. Growth curves of the 50 most active peptides. Colored lines are Me^x-recorded growth curves (average of n=3) determined via OD_{ID} approximation (header: 'rank: parent name'). Black lines are growth curves (n=3, error bars: 4s) determined via OD measurement in microtiter plates of individually grown strains. Horizontal dashed lines, in black (OD) or colored in green (OD_{ID}), show final values measured 4.5 h post-induction of a strain synthesizing the inactive control peptide HNP-1$_{3425}$ APD. In each facet, we state if we obtain a $p<0.05$ (Wald's test) for significant growth inhibition after 1.5 h in Me^x.

2.6 Supplementary code

```
---

Authors: "Philipp Koch, Mathias Cardner"
Date:"18.12.2020"

---
```

Load packages and data

```
library(tidyverse)
library(DESeq2)
```

Load Supplementary Table 2.2 containing NGS read counts for each peptide-encoding DNA listed by ID.

```
countData <- read.csv("Supplementary Table 2.2.csv") %>%
  column_to_rownames("ID") %>%
  as.matrix()
```

Pre-process

DESeq2 uses the first level of time point as the baseline for DE tests. Since we want to compare against t_0, we relevel it to be the first level of timepoint. (This is also necessary for using lfcShrink later). We leave t_3 as the last level, since this is the final condition of interest.

```
timepoint <- parse_number(colnames(countData)) %>% factor() %>% relevel("0")
batch <- sub(".*\\d", "", colnames(countData))
colData <- data.frame(row.names = colnames(countData),
                      timepoint = timepoint,
                      batch = factor(batch))
```

DESeq2 uses the last term of the design formula for contrasts, which in our case should be timepoint.

```
dds <- DESeqDataSetFromMatrix(countData, colData, ~ batch + timepoint)
dds <- dds[rowSums(counts(dds)) > 1,]
#saveRDS(rownames(dds), "IDs_used_in_DESeq.rds")
```

Differential expression (abundancy) analysis

Fit DESeq model. Data from all time points and replicates are used to estimate dispersion, but only the contrast t_3, corresponding to 4.5 h past induction, and t_1, corresponding to 1.5 h past induction, vs. t_0, the time point for induction will be computed and tested in the end.

```
dds.Wald <- DESeq(dds, parallel = TRUE)
#saveRDS(dds.Wald, "dds_Wald.rds")
```

Since we are interested in growth inhibitory peptides, i.e. those whose log2-fold change at t_3 and t_1 vs t_0 is negative, we set altHypothesis = "less".

```
res.Wald <- results(dds.Wald, name = "timepoint_3_vs_0", altHypothesis = "less",
                    alpha = 0.05, parallel = TRUE)


res.Wald$pvalue.t1 <- results(dds.Wald, name = "timepoint_1_vs_0", altHypothesis = "les
s",alpha = 0.05, parallel = TRUE)$pvalue
```

```
res.Wald$padj.t1 <- results(dds.Wald, name = "timepoint_1_vs_0", altHypothesis = "less"
,alpha = 0.05, parallel = TRUE)$padj
#saveRDS(res.Wald, "res_Wald.rds")
```

Shrink log2-fold changes vs time point 0

Use apeglm to compute shrunken log2-fold changes at time point X vs 0.

```
lfcShrunk1 <- lfcShrink(dds.Wald, coef = "timepoint_1_vs_0", type = "apeglm",
                        parallel = TRUE)
lfcShrunk2 <- lfcShrink(dds.Wald, coef = "timepoint_2_vs_0", type = "apeglm",
                        parallel = TRUE)
lfcShrunk3 <- lfcShrink(dds.Wald, coef = "timepoint_3_vs_0", type = "apeglm",
                        parallel = TRUE)
```

Add shrunken log2-fold changes to DESeq2 result and add the lfcShrink$_X$SE containing the corresponding "standard error".

```
res.merged <- res.Wald
res.merged$lfcShrink1 <- lfcShrunk1$log2FoldChange
res.merged$lfcShrink1SE <- lfcShrunk1$lfcSE

res.merged$lfcShrink2 <- lfcShrunk2$log2FoldChange
res.merged$lfcShrink2SE <- lfcShrunk2$lfcSE

res.merged$lfcShrink3 <- lfcShrunk3$log2FoldChange
res.merged$lfcShrink3SE <- lfcShrunk3$lfcSE
```

Calculate strain-specific concentrations (OD$_{ID}$) at each time point to create growth curves for each peptide-expressing strain

Compute mean optical densities, extracted from the shake flask experiments across time points.

```
log2meanODratios <- tibble(t0A = 0.1995,
                           t0B = 0.2,
                           t0C = 0.1995,
                           t2A = 2.25,
                           t2B = 2.19,
                           t2C = 2.135,
                           t1A = 0.705,
                           t1B = 0.704,
                           t1C = 0.697,
                           t3A = 4.085,
                           t3B = 4.305,
                           t3C = 3.83) %>%
  tidyr::gather(condition, OD) %>%
  # Group ODs by time point, and compute the mean.
  mutate(timepoint = parse_number(condition)) %>%
  group_by(timepoint) %>%
  summarise(meanOD = mean(OD)) %>%
  ungroup()
```

Compute the OD$_{ID}$ of each strain across all time points.

```
# Function for computing the relative number of cells
ODID <- function(x) {
  OD <- log2meanODratios %>%
    filter(timepoint == x) %>%
    pull(meanOD)
  return(2^res.merged[[paste0("lfcShrink", x)]]*OD)
}
```

```
res.merged$ODID_t0 <- 0.2
res.merged$ODID_t1 <- ODID(1)
res.merged$ODID_t2 <- ODID(2)
res.merged$ODID_t3 <- ODID(3)
```

Generate a database file containing all relevant information on the peptides that passed independent filtered in Me^x.

```
dictionary <- read.csv("Supplementary Table 2.2.csv") %>%
  select(ID,Sequence,DNAsequence)

Database <- res.merged %>%
  as.data.frame() %>%
  rownames_to_column("ID") %>%
  as_tibble() %>%
  left_join(dictionary %>%
              mutate(ID = as.character(ID))) %>%
   filter(!is.na(padj)) %>%
  distinct(AAsequence, .keep_all = T)
```

Session info

```
sessionInfo()
```

CHAPTER 3 | Optimization of the antimicrobial peptide Bac7 by deep mutational scanning

Philipp Koch[1*], Steven Schmitt[1*], Alexander Heynisch[1], Anja Gumpinger[2], Marina Gysin[3], Sven Panke[1], Martin Held[1]

[1]Bioprocess Laboratory, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland.

[2]Machine Learning and Computational Biology, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland.

[3]Institute of Medical Microbiology, University of Zurich, Zurich, Switzerland

*these authors contributed equally

3.1 Abstract

Intracellularly active antimicrobial peptides are promising candidates for new antibiotics. However, drug development for those molecules is challenging as their large size spans an enormous sequence space, overwhelming the capabilities of current medicinal chemistry methods by far. We built a high-throughput peptide development platform that incorporates rapid investigation of the fitness landscape of peptides and enables rational optimization of their antimicrobial activity. The platform is based on deep mutational scanning of the DNA encoded peptides and employs highly parallelized bacterial self-screening coupled to next-generation sequencing as readout for bioactivity. As a starting point, we used $Bac7_{1-23}$, a 23 amino acid residues long variant of bactenecin-7, a potent translational inhibitor and one of the most researched proline-rich antimicrobial peptides. Using the platform, we simultaneously analyzed the bioactivity of >600'000 peptide variants exploring the fitness landscape of $Bac7_{1-23}$. This dataset guided the design of a focused library of ~160'000 variants and the identification of Bac7PS as lead candidate with improved bioactivity. Bac7PS showed high activity against multidrug-resistant clinical isolates of *E. coli* and was less dependent on SbmA, a transporter commonly used by proline-rich antimicrobial peptides to reach the intracellular site of activity. Furthermore, Bac7PS displayed low toxicity against eukaryotic cells and demonstrated good efficacy in a murine septicemia model induced by *Escherichia coli*. We demonstrated that our platform can be used to establish a fitness landscape of antimicrobial peptides and furthermore showed its usefulness in the support of hit to lead identification and lead optimization of antimicrobial drug candidates.

3.2 Introduction

The drug development field is in urgent need of novel compounds to deliver the next generation of antibiotics for combating multidrug-resistant (MDR) bacteria[7]. Several promising leads have been identified in the group of antimicrobial peptides (AMPs), some of which are currently in clinical trials[47]. However, most AMP candidates fail in clinical trials[36], often because a majority of these molecules act via disruption of bacterial membranes, a mode of action (MoA) that is prone to cause off-target toxicity against human cells[53].

Proline-rich AMPs usually do not lyse but kill bacteria by interfering with the activity of intracellular targets such as ribosomes[39] and therefore might have the potential for a more specific antimicrobial action. One of the most intensively researched proline-rich AMPs is bactenecin-7 (Bac7). It is a 60 amino acids long linear peptide that was isolated first from bovine neutrophils[115]. For exerting its antimicrobial activity against bacteria, Bac7 first crosses the outer membrane of Gram-negative bacteria via not yet fully elucidated mechanisms and then traverses the inner membrane through the SbmA transporter[116]. Consequently, SbmA loss considerably decreases susceptibly to Bac7 and its absence from other bacterial families restricts the activity spectrum of Bac7 to Gram-negative Enterobacteriaceae[117,118], a family of bacteria currently listed as 'critical priority' by the WHO[14]. In vitro studies on Bac7 truncates indicated that the two N-terminal arginine residues are needed for efficient uptake[119] and its C-terminus can be truncated resulting in peptides with a length of 35, 23, and 16 amino acids that display only a minor loss of activity[118]. Moreover, the activity can be increased towards specific strains via mild modulation of the amino acid sequence[119–121]. The standard workflow to assess the activity of AMP variants relies on chemical methods for their synthesis followed by evaluation of the growth inhibitory effect with bacterial pathogen models and clinical isolates. Due to limited throughput and high peptide synthesis costs, these studies can typically deliver only very few data points generated for mildly modified peptides by single amino acid exchanges performed in the course of alanine or lysine scans[120,98]. This generally provides only a poor coverage of the antimicrobial fitness landscape.

To greatly expand the coverage of such fitness landscapes, deep mutational scanning (DMS) methods can be used, which grant access to millions of variants and data points in single experiments[122]. Most frequently, these methods are used to study the functional consequence of single amino acid residue substitutions in proteins[123], since introducing multiple substitutions leads to a combinatorial explosion. In DMS, first, large protein

libraries are produced by systematically varying the coding DNA sequence via chemical DNA synthesis or error-prone PCR (epPCR)[70]. Then, the protein variants are expressed recombinantly in cells and exposed to a selection pressure such as survival of the host or screening based on the expression of fluorescent reporter proteins. Finally, next-generation sequencing (NGS) is used to quantify the abundance of the protein-encoding DNA before and after selection or sorting[124]. Recently, a study investigated single amino acid residue substitution of the proline-rich AMP apidaecin to determine critical interactions with the ribosome[125]. However, DMS has never been exploited as a platform for hit to lead identification and optimization of antimicrobials. If applied successfully, the coverage of the AMP fitness landscape could be greatly increased, guiding the design of novel and improved AMP drug candidates.

In this study, we modified the 23 amino acid truncate of Bac7, from here on referred to as Bac7$_{1-23}$, in two DMS rounds. In the first round, we screened a Bac7$_{1-23}$ library consisting of 601,551 randomly mutagenized variants and assessed their growth inhibition when expressed intracellularly in *E. coli,* a so-called self-screening[91]. This enabled us to generate a fitness landscape, describing the effect of each amino acid residue substitution on growth inhibition. Guided by these results, we performed a second DMS round on a focused, semi-rationally designed library of Bac7$_{1-23}$ covering 156,779 variants. After assessing the effect on growth inhibition of each variant, we were able to extract the most activity-enhancing amino acid residue combination and built a new-to-nature peptide called Bac7PS with high activity towards a broad-panel of bacterial pathogens, low toxicity against eukaryotic cells and good efficacy in in vivo studies.

3.3 Results

DMS of Bac7$_{1\text{-}23}$ using random mutagenesis

First, we aimed to identify all residues essential for growth inhibition and those with potential for activity optimization of Bac7$_{1\text{-}23}$. For this, we randomly mutagenized the Bac7$_{1\text{-}23}$ coding gene by a highly error-prone version of PCR and ligated the modified fragments into plasmids allowing their expression from the tightly regulated P$_{BAD}$ promotor. We transformed *E. coli* TOP10, serving for intracellular peptide expression and as a microbial pathogen model (Fig. 3.1a). We then recorded growth curves for 94 randomly selected strains in microtiter plates and found that about half of the Bac7$_{1\text{-}23}$ variants suppressed the growth of the respective host upon expression (Supplementary Fig. 3.1). To assess the antimicrobial effect of a much larger proportion of the randomly mutated library at once, we grew around 500 million cells containing the peptide library in a shake flask (n=3; Supplementary Fig. 3.2), induced peptide expression, and counted the abundance of the specific peptide-encoding DNA at the time of induction and 4 h later by NGS. We used the change of DNA abundance as a proxy for the degree of growth inhibition because if peptide expression limits the growth of its host, it also limits the propagation of the peptide-encoding DNA[91]. We calculated the log2 fold-change in DNA abundance and ranked all peptides from most growth inhibitory (lowest negative log2 fold-change) to least growth inhibitory (highest positive log2 fold-change) (Fig. 3.1a). In total, we counted the abundance of 601,551 different peptides covering 398 peptides with one mutation (87% out of 460 possible variants), 21,567 double mutants (21% out of 101,200 possible variants), 185,993 triple mutants (1.3% out of ~14 million possible variants) and 228,433 quadruple mutants (0.01% out of ~1.4 billion possible variants) (Supplementary Fig. 3.3).

Fitness landscape of Bac7$_{1\text{-}23}$

Analysis of the generated data set indicated that 489,520 variants had low read counts (less than 50 read counts at the time of induction), which may result in false abundance values (log2-fold change values) in NGS-based screenings[126]. Therefore, we transformed the log2 fold-change describing the degree of growth inhibition using a Bayesian Shrinkage Estimator[106], yielding a more robust estimation even at low read counts. Additionally, many amino acid substitutions were highly underrepresented in the library; for example, we found the amino acid tryptophan at position 17 or methionine at position 11 in only one peptide, while glycine on position 1 appeared in 2,685 different peptides (Supplementary Fig. 3.4). To overcome these differences and measure the effect of a particular amino acid residue substitution, we investigated whether these substitutions were significantly enriched in higher or lower growth inhibitory peptides using a

permutation scheme, thereby generating a fitness landscape of Bac7$_{1-23}$ (Fig. 3.1b). For 250 out of 483 possible single amino acid substitutions (23 positions x 21 amino acids including truncation by insertion of a stop codon), we could not detect a significant link to growth inhibition (white boxes in Fig. 3.1b), mainly because those amino acids were underrepresented in the library (Supplementary Fig. 3.4 & 5). For 233 amino acid substitutions, we detected a strong influence on growth inhibition. We analyzed this fitness landscape and found that the 'RLPRPR' motif on position 9-14 as well as the arginine residue on position 6 was most crucial for growth inhibition in Bac7$_{1-23}$ variants (yellow boxes in Fig. 3.1b). Additionally, the random incorporation of stop codons in any of the first 16 positions mostly eliminated the growth inhibitory effect, while insertion at one of the following seven positions did not eliminate this effect entirely. This points towards an essential (minimal) length of active Bac7$_{1-23}$ of 16 amino acids (*in Fig. 3.1b).

Amino acid residue substitutions at positions 3, 5, 7, and 15 as well as at positions 17-23 near the C-terminus, seemed to improve the growth inhibiting effect relative to the parental Bac7$_{1-23}$. Thus, next to residues essential for the growth inhibitory effect, we also detected eleven positions that offered room for optimization of the growth inhibitory potential of the peptide. However, extracting guidance for the activity optimization of Bac7$_{1-23}$ from the present activity landscape beyond the identification of critical positions might not be very efficient for two reasons: Firstly, more than 100 amino acid substitutions were too underrepresented in the library to infer statistical significance concerning their influence on growth inhibition (Supplementary Fig. 3.5; p-value>0.1; Benjamini-Hochberg adjusted). Secondly, as Bac7 and truncates thereof are already very active, single or even double substitution may not suffice to considerably increase activity when tested in regular MIC-assays, as already shown previously[120]. As our data hardly covered combinations of amino acid substitutions because those appeared rarely in the error-prone library (e.g. there are >10^9 possible combinations of four amino acid residues in Bac7$_{1-23}$), we decided to create in-depth knowledge on amino acid residue combinations and performed a second round of DMS on a more focused Bac7$_{1-23}$ library, guided by our first Bac7$_{1-23}$ fitness landscape.
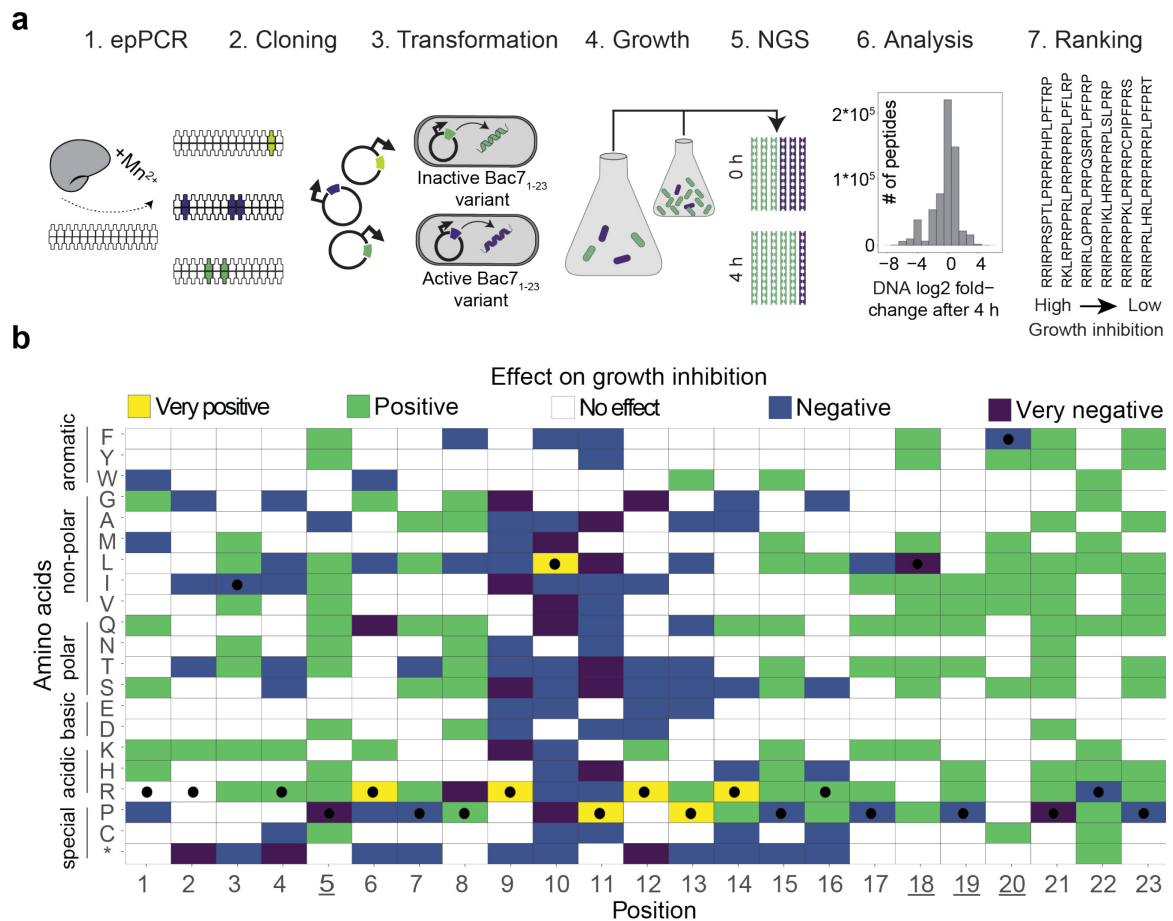
Fig. 3.1 | DMS of Bac7$_{1-23}$ epPCR library. a) DMS workflow. epPCR: The Bac7$_{1-23}$ DNA is amplified at high error rate using an error-prone DNA polymerase in the presence of Mn^{2+}. Cloning: The DNA sequences are inserted into plasmids with inducible promoters. Transformation: *E. coli* TOP10 is transformed with the generated peptide-encoding DNA library. Growth experiment: The transformants are grown together in a single shaking flask (n=3), peptide expression is induced and plasmids are isolated. NGS: peptide-encoding DNA sequences are counted at the time of induction and 4 h later using NGS. Analysis: For each peptide-encoding DNA the log2-fold change is determined (ratio of NGS counts between the two time points). Histogram the log2-fold change of all 601,551 variants. Ranking: Peptide sequences are ranked by the degree of the observed growth inhibition. The more negative a log2-fold change, the higher the observed growth inhibition and vice versa. **b)** Bac7$_{1-23}$ fitness landscape. For each amino acid residue substitution, the enrichment in higher or lower growth inhibitory peptides is determined. Then a z-score (z) is calculated that corresponds to the number of standard deviations by which the calculated enrichment lies above (+) or below (-) a null — distribution that shows no enrichment. z is empirically divided into four groups, corresponding to very positive (yellow; z>=40), positive (green; z>=4), negative (blue; z<=-4), or very negative (purple; z<=-40) effects on growth inhibition. No effect on growth inhibition is detected if the z is close to 0 (white; -4<z<4). Black dots correspond to the Bac7$_{1-23}$ wild-type amino acids. The underlined positions are chosen as targets for the subsequent site-saturation mutagenesis. Black dots correspond to the Bac7 parental amino acid residue at each position.

DMS of Bac7$_{1-23}$ using comprehensive site-directed mutagenesis

For the focused analysis, we first performed saturation mutagenesis at four positions of Bac7$_{1-23}$ (20^4 = 160,000 possible variants). From the eleven positions that seemed to offer promiscuity with regard to the accommodation of amino acid substitutions in the first activity landscape (Fig. 3.1b), we chose to saturate positions 5 and 18-20. Position 5 was selected as it is not part of the crucial N-terminal 'RRIR' motif important for cellular uptake and ribosomal binding, nor of position 6-13 representing a conserved core region among

proline-rich AMPs[127]. Positions 18-20 were selected to increase our knowledge about the effect of C-terminal amino acid residue substitutions, as Bac7 variants longer than the most frequently studied Bac7$_{1-16}$ were shown to have a broader activity spectrum[118]. Furthermore, little is known from crystallization studies about the importance of the activity of residues downstream of position 16[127,128]. The focused library was generated by site-saturation mutagenesis, using one NNK codon for position 18 and a mixture of codons NDT, VMA, ATG, and TGG as described by Tang *et al.*[129] for the remaining positions. This reduced the need for oversampling for cloning and NGS drastically, as it limits the bias in amino acid distribution compared to less restricted codon schemes, such as the NNK codon-scheme[130]. We again grew the entire library in a single flask (Supplementary Fig. 3.6), sequenced the peptide-encoding DNA at the time of induction and 4.5 h later, and used the relative DNA abundance as a proxy for the degree of growth inhibition of each peptide variant (similarly to the DMS workflow display in Fig. 3. 1a). In total, we were able to identify 156,779 Bac7$_{1-23}$ derivatives, covering 98% of all possible variants.

We then quantified the effect on growth inhibition of a particular amino acid residue substitution, similarly to the treatment of the previous library. As each amino acid substitution (e.g. alanine in position 5) appeared in roughly 8,000 different peptides ($1*20^3=8,000$; one residue in a specific position fixed and combined with the entire set of possible substitutions at the remaining three positions), we did not need to use a permutation scheme as for the first library. We first inferred the effect of single amino acid residue substitutions by applying a modified version of the gene-set enrichment analysis (GSEA) proposed by Subramanian *et al.*[131]. Here, enrichment curves for each site and each amino acid (Fig. 3.2a) are drawn after ranking all peptides according to the growth inhibitory effect (log2 fold-change). From those curves, we computed the area under the curve (AUC) giving a value between -1 and 1. Positive AUC-values indicate that the amino acid residue substitution is enriched in peptides with a high growth inhibitory effect, while negative AUC-values indicate enrichment in peptides with low growth inhibitory effect. AUC-values close to 0 indicated no enrichment in either fraction (Fig. 3.2a). As observed during the error-prone library diversification approach discussed above, substituting the wild-type amino acid residues at all four positions could result in peptides with enhanced growth inhibitory effect (Fig. 3.2b). Peptides that had tyrosine and alanine residues at all four positions experienced the strongest increase and decrease in growth inhibition, respectively (Fig. 3.2b). Even though these analyses revealed the effect of each single amino acid residue substitution at all four positions in higher detail, they did not reveal the effect of residue combinations, likely needed to achieve an even higher increase relative to Bac7$_{1-23}$.
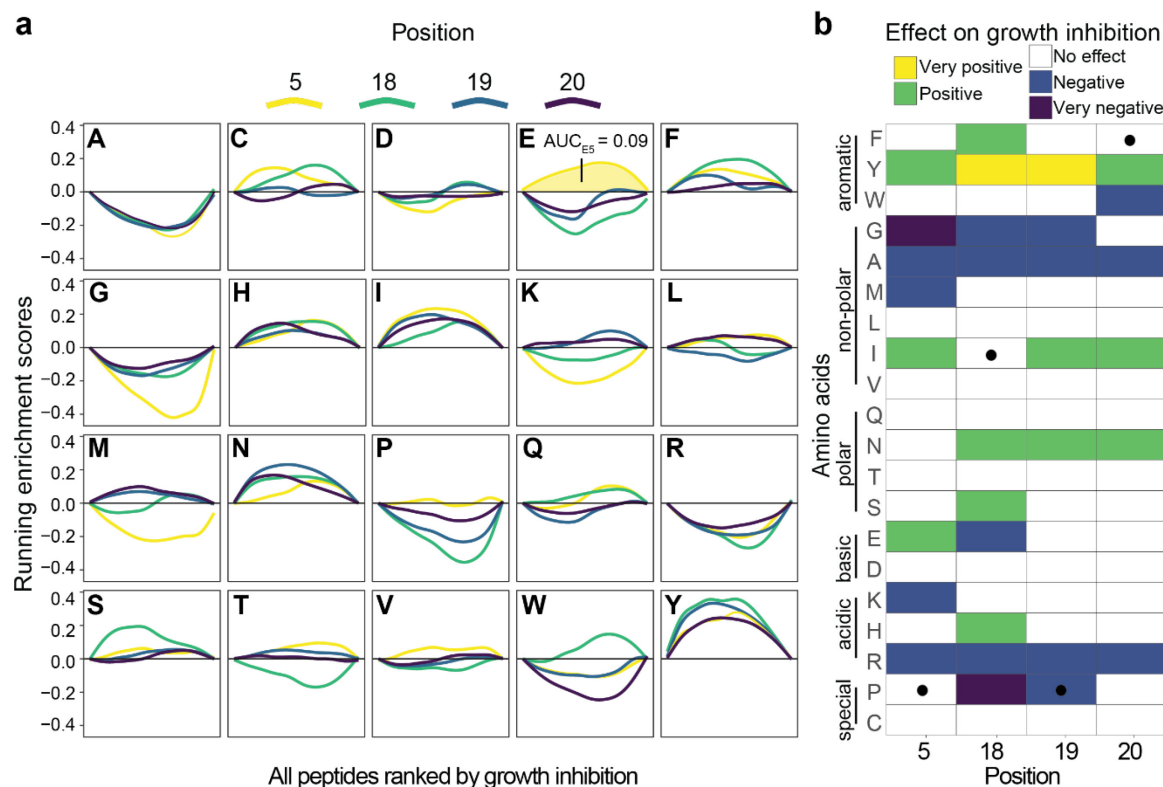
Fig. 3.2 | DMS of Bac7$_{1-23}$ site-saturation mutagenesis library. a) Enrichment curves. Peptides are ranked according to their growth inhibitory effect and a running enrichment score for each amino acid residue at each of the four substitution sites is calculated. Increasing y-values (positive slope) correspond to enrichment in that particular ranking segment. For example, irrespective of the substitution site, alanine (A) is enriched in the peptides in the lower third of the ranked peptides, i.e. those with low(er) growth inhibitory effect. In contrast, tyrosine (Y) is enriched (again, irrespective of the substitution site) in the upper half of the ranked peptides, i.e. those with high(er) growth inhibitory effect. From each curve, the area under the curve (AUC) was extracted. Positive (negative) AUC-values represent an enrichment among growth inhibitory peptides on left side of the x-axis (among not growth inhibitory peptides on the right side of the x-axis). An example is shown for the E residue at position 5 (AUC$_{E5}$). **b)** AUC-values for each amino acid residue substitution extracted from each curve. Effects on growth inhibition are empirically divided into five groups: very beneficial (yellow; AUC≥0.2), positive (green; AUC≥0.1), no effect (white; -0.1<AUC<0.1), negative (blue; AUC≤-0.1) or very negative (purple; AUC≤-0.2). Black dots correspond to the Bac7 parental amino acid residue at each position.

Combination of amino acid residues

Effects of amino acid residue substitutions in proteins or peptides can be non-additive when they appear in combination[83,132], that is, they can become larger (cooperative) or smaller (antagonistic) than the sum of the individual effects[132]. These non-additive effects, from here on referred to as interactions, appear due to conformational changes of the peptide enabling or restricting target binding, introduction or removal of protease cut sites, or changes of the physicochemical properties such as charge, hydrophobicity, or solubility above or below a certain threshold at which biological functions are affected. As we had almost fully saturated position 5, 18, 19, and 20, we can be sure to have most amino acid residue combinations in the library. We thus aimed to extract the most beneficial combinations and then investigate whether interactions contributed to the extent of the growth inhibitory effect.

First, we extracted those amino acid residue combinations which, if combined in a peptide, lead to most and least growth inhibitory peptides. A specific combination of four amino acids (quadruple combination) only appears in one peptide in the investigated peptide library, and as previously mentioned, DNA abundance-based rankings can be slightly error-prone with variants with low read counts. They should thus be regarded as a qualitative instead of a quantitative measure. To create a nevertheless statistically sound dataset, we investigated the effect on growth inhibition of each of the 32,000 different triple combinations ($20^3 \times \binom{4}{3}$; or 8,000 possibilities of combining three amino acid residues at four positions), each appearing in 20 different peptides (20 possible combinations for the fourth amino acid). We used significant pattern mining[133] to extract significantly enriched triple combinations of amino acids among the 10% and 25% (arbitrary thresholds) most and least growth inhibitory peptides (results for all triple combinations in Supplementary Table 3.1). Based on these results, we found that peptides with alanine-proline-proline on positions 5, 18, and 19 were the least growth inhibitory (Table 3.1). Surprisingly, even though peptides with a methionine residue on position 20 or a histidine residue on position 19 did not show beneficial effects on growth inhibition when looking only at single amino acid residue substitutions (Fig. 3.2b), peptides with triple residue combinations of tyrosine-phenylalanine-methionine and asparagine-histidine-asparagine on positions 18-20 were most growth inhibitory (Table 3.1). Additionally, even though peptides with the single residue substitution to tyrosine at each position were most growth inhibitory (Fig. 3.2b), peptides with three tyrosine residues did not rank among the most growth inhibitory ones (Table 3.1). These differences strongly indicated interactions between amino acid residues. However, we were not able to fully quantify an interaction of amino acid residues using significant pattern mining because the analysis used only part of the data set, in our case 20% and 50% of all peptides.

Table 3.1: Subset of relevant triple amino acid combinations obtained by significant pattern mining. The first four columns indicate the positions. "–" indicates that this position has been intentionally kept open. The fifth column indicates the respective p-value for significant enrichment in either the 10% most or 10% least growth inhibitory peptides. Columns six and seven indicate the number of peptides containing the respective combination among the top 25% and top 10% most inhibitory peptides (ranked according to the shrunken log2 fold-change). All possible combinations were ranked from most to least enriched in the top 10% of the most growth inhibitory peptides.

Rank	Pos. 5	Pos. 18	Pos. 19	Pos. 20	*p*-value	# within 25% most actives	# within 10% most actives
1	-	Y	F	M	$1.9*10^{-6}$	20	20
2	-	N	H	N	$1.9*10^{-6}$	20	20
3	C	N	N	-	$3.8*10^{-6}$	20	19
454	Y	-	Y	Y	$1.0*10^{-3}$	14	11
4,815	Y	Y	Y	-	0.3	3	10
32,000	A	P	P	-	$1.9*10^{-6}$	0	0

To fully quantify interactions between amino acid residue combinations, we investigated if the calculated effect on growth inhibition, the AUC, for each amino acid residue at each position changed upon conditioning the same calculation on a second amino acid residue at another position (AUC conditional). For example, we compared the effect on growth inhibition for the single substitution histidine residue on position 5 (AUC_{H5}) to the effect on growth inhibition of histidine residue on position 5 among peptides that have a cysteine residue introduced at position 18 ($AUC_{H5|C18}$; note that the latter is a subset of the former). As the difference between $AUC_{H5|C18}$ and AUC_{H5} is small, we concluded that the cysteine residue on position 18 does not influence the growth inhibitory effect of the histidine residue on position 5 and that these two amino acid residues behave additively and hence are not interacting (Fig. 3. 3a). We calculated such differences between all possible double combination of amino acid residues across the four sites and found that 96.1% are in the range of [-0.09,0.09] (within interquartile range (IQR) ± 1.5*IQR of all values; covering boxplot and whiskers) and thus effectively did not interact with each other (Fig. 3.3b; all data in Supplementary Table 3.2). However, we detected 187 amino acid residues with interactions: 83 of these displayed a strongly cooperative effect on the growth inhibitory effect of a second amino acid residue, and the remaining 104 residues displayed a strongly antagonistic effect. For example, we observed cooperativity for the residue combinations tyrosine-phenylalanine and asparagine-histidine on positions 18 and 19, which are both part of significantly enriched triple combinations (Table 3.1). On the other hand, we observed antagonistic effects for a combination of two tyrosine residues on positions 18 and 19, on positions 5 and 18, and positions 18 and 20. Additionally, every combination of two proline residues on positions 18, 19, and 20 showed strong antagonistic effects for growth inhibition. In general, aromatic amino acid residues (57), as well as proline (24), and arginine residues (18) caused most interactions, while we saw the least interactions for residues of amino acid serine (0), valine (1), and glutamate (2) (Supplementary Fig. 3.7). Moreover, we measured most interactions among amino acid residues on neighboring positions, e.g. between residues on positions 19 and 20, or 19 and 18 (Fig. 3b). Interestingly, we also discovered cooperativity between the proline and phenylalanine residues on positions 19 and 20, which is part of the Bac7$_{1-23}$ wild type sequence. Ultimately, the most growth inhibitory combination of three amino acid residues that we could identify by this analysis among peptides in this library was the residue combination tyrosine-phenylalanine-methionine on positions 18-20, including a cooperative interaction between a phenylalanine and tyrosine residues, possibly increasing target binding. Among all 20 possible peptides with this combination, the most growth inhibitory peptide had an isoleucine residue at position 5. This specific Bac7$_{1-23}$ variant P5I R18Y L19F P20M is from here on referred to as Bac7PS.

a

Single substitution | Combination

Position — 5 — 19

Running enrichment score of peptides with F19

$AUC_{F19} = 0.050$ — All peptides (~160,000)

$AUC_{F19 \mid Y18} = 0.195$ — **increased** — Peptides with Y18 (~8,000)

Running enrichment score of peptides with H5

$AUC_{H5} = 0.089$ — All peptides (~160,000)

$AUC_{H5 \mid C18} = 0.072$ — **unchanged** — Peptides with C18 (~8,000)

Running enrichment score of peptides with P19

$AUC_{P19} = -0.148$ — All peptides (~160,000)

$AUC_{P19 \mid P18} = -0.355$ — **decreased** — Peptides with P18 (~8,000)

b

Position of the second AA: ◆ 5 ◆ 18 ◆ 19 ◆ 20

Interaction [$AUC_{AA2 \mid AA1} - AUC_{AA2}$]

cooperative

antagonistic

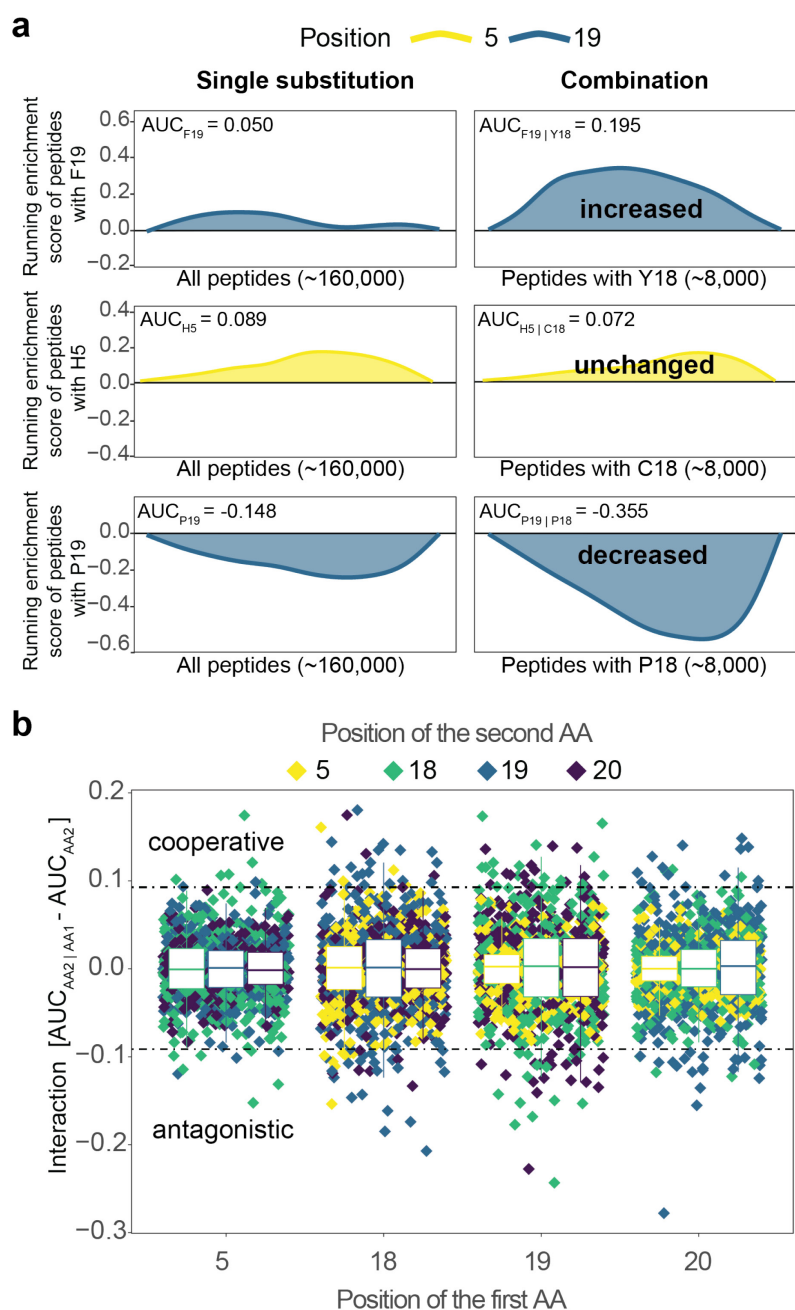Position of the first AA: 5, 18, 19, 20

Fig. 3.3 | Interactions of amino acid residue combinations. a) Examples of amino acid residue combinations (right) resulting in an increased, a decreased, or unchanged AUC when compared to the single substitution (left). Peptides are ranked on the x-axis according to their effect on growth inhibition. AUC_{F19} strongly increases for peptides that have tyrosine residue at position 18. AUC_{H5} remains similar among peptides with a cysteine residue at position 18. AUC_{P19} strongly decreases for peptides with a proline residue at position 18. For easier interpretation of the conditional AUCs, the curves were smoothened using cubic spline interpolation. **b)** Interactions of all 4,800 amino acid residue combinations. For each of the 20 amino acid residues at each of the four positions, there are 60 (20*3) possibilities to interact with a second amino acid residue. Interactions (shown on the y-axis) are calculated by subtracting the AUC calculated for the single substitution = AUC_{AA2} (values for the single substitution from Fig. 3.2b) from the AUC using an amino acid residue (AA2) in combination with another amino acid residue (AA1) = $AUC_{AA2 \mid AA1}$. An interaction is measured when the differences are larger (cooperative) than 0.09 or smaller (antagonistic) than -0.09 (=outliers of a boxplot containing all interaction values; IQR ± 1.5 *IQR).

In vitro activity characterization of Bac7PS and Bac7$_{1-23}$

To investigate if the amino acid residue substitutions suggested by the two rounds of DMS might also translate into the improved antimicrobial performance of Bac7PS when applied extracellularly, we characterized Bac7PS in different in vitro experiments and compared it directly to Bac7$_{1-23}$. Both peptides were synthesized by solid-phase peptide synthesis and purified by reversed-phase high-performance liquid chromatography to 92% (Bac7PS) and 95% (Bac7$_{1-23}$) purity. We first evaluated the in vitro antimicrobial activity against the microbial pathogen model used for DMS (*E. coli* TOP10) by measuring the minimal inhibitory concentration (MIC) of the peptides following CLSI standards[112]. We included the transporter-loss mutant *E. coli* BW25113 Δ*sbmA*[134] which had been previously

described to limit Bac7 uptake and activity[135], together with its parental strain *E. coli* BW25113[134]. The set was complemented by strain *E. coli* ATCC 25922, a quality control strain often used in clinical microbiology. Remarkably, we found an approximately twofold reduction in MIC when applying Bac7PS from outside to the DMS strain *E. coli* TOP10 (MIC of 2.1 µM for Bac7PS, 4.6 µM for Bac7$_{1-23}$), suggesting that our strategy had indeed allowed for a variant with higher antimicrobial activity. Furthermore, externally applied Bac7PS showed activity against the transporter-loss mutant BW25113 Δs*bmA*, while Bac7$_{1-23}$ was not active, even at the highest concentration tested (Table 3.2). Interestingly, the MIC for the quality control strain was similar for both peptides (2.6 µM for Bac7PS, 2.8 µM for Bac7$_{1-23}$). To investigate the potential of Bac7PS against a broader panel of strains we determined the MICs of both peptides against a set of 45 *E. coli* clinical isolates collected from Swiss hospitals, of which 23 strains carried the information for extended-spectrum beta-lactamase (ESBL) or a carbapenemase (CRE) and thus can be considered as multidrug resistant (MDR). For the clinical isolates, the antimicrobial activity of Bac7PS clearly exceeded that of Bac7$_{1-23}$ with an MIC$_{50}$ for Bac7PS of 2.9 µM and for Bac7$_{1-23}$ of 7.5 µM (Fig. 3.4a; Table 3.2). Remarkably, Bac7PS showed good activity even against MDR strains, and its activity was improved by a similar factor for these strains relative to Bac7$_{1-23}$. Thus, the observed improvements in our DMS approach for intracellularly expressed peptides translated well into in vitro properties.

Table 3.2: Susceptibility assays. MIC-values are averaged from quadruplicate experiments in full MHB 2 medium. ">" indicates no activity at the highest measured concentration. Hemolysis assays and toxicity measurement were performed in triplicates. The therapeutic index (TI) is calculated by dividing the IC$_{50}$ values measured against HEK 293 cells by the MIC$_{50}$ obtained from the clinical isolates (see Fig. 3.4a).

	MIC against *E.coli* strains [µM]					Hemolysis of mouse red blood cells [%]		Toxicity IC$_{50}$ [µM]	TI
Peptide	TOP10	ATCC 25922	BW 25113	BW 25113 Δ*sbmA*	Clinical isolates (MIC$_{50}$)	1 x MIC	4 x MIC	HEK 293	Toxicity / MIC
Bac7$_{1-23}$	4.6	2.8	7.4	>33.2	7.5	2.1%	6.4%	1460	195
Bac7PS	2.1	2.6	3.6	19.9	2.9	3.1%	3.8%	521	180

We next investigated whether an increased propensity to damage membranes could be the reason for the improved antimicrobial activity of Bac7PS. Membrane damage is the most frequent way AMPs kill bacteria and is often discussed as reason for high toxicity against eukaryotic cells[53], but Bac7$_{1-23}$ does not damage membranes at its MIC. Consequently, we wanted to investigate if Bac7PS might have acquired a tendency to damage membranes and thus an increased risk for applications in vivo. We followed a protocol using a variant of the green fluorescent protein (GFP) and propidium iodide (PI) as markers for membrane damage in bacteria exposed to AMPs[136]. In our experiment, membrane damage is indicated by the uptake of the (otherwise membrane-impermeable) dye PI and/or by the leakage of intracellularly expressed GFP. For both peptides, Bac7$_{1-23}$

and Bac7PS, we could confirm that the integrity of the membrane is neither affected at the MIC nor at concentrations 8-fold above the MIC (~1% PI-positive cells for Bac7$_{1-23}$ and Bac7PS, ~99% of cells retained GFP levels). In contrast, the membrane-active peptide melittin rapidly induced PI uptake and loss of GFP (Fig. 3.4b). However, at approximately 16-fold above its MIC (~35 µM), we noticed minor membrane damage of cells treated with Bac7PS, indicated by loss of GFP (26% of cells lost GFP) and minor uptake of PI (~3% PI-positive cells). The effects were more pronounced when comparing these peptides in MHB I medium, which is not cation adjusted and therefore potentially less ionic, an effect often taken advantage of to increase membrane interaction of AMPs[30,111] (Supplementary Fig. 3.8). To ensure that this absence of membrane-damaging properties at relevant concentrations was not limited to bacterial membranes, we performed a hemolysis assay using red blood cells from mice. We did not detect strong lysis (<4%) when applying Bac7 and Bac7PS at the MIC measured for *E. coli* ATCC 25922 and only very minor lysis (<7%) when applying a four times higher concentration (Table 3.2). Thus the growth inhibitory activity of Bac7PS was still not based on membrane damage. However, the loss of the dependency on the inner membrane SbmA transporter as seen for increased activity against the SbmA knockout strain (Table 3.1) and possibly the slight increase of membrane damage at higher peptide concentration (Fig. 3.4b & Supplementary Fig. 3.8) could indicate a higher degree of membrane interaction that could potentially lead to increased membrane crossing and thus uptake.

We also investigated the effect of the changes in Bac7PS on the expected main activity, protein synthesis inhibition, and performed a translation inhibition assay. We determined the inhibitory effect on cell-free ribosomal translation by incubating Bac7PS and Bac7$_{1-23}$ (concentration range: 800 µM to 0.08 µM in 2.5-fold dilutions steps) with bacterial (*E. coli* ATCC 25922) and human (HEK-293) lysates together with the mRNA that encoded a luciferase protein. We measured the luminescence of each sample at each concentration (n>6) and extracted the IC$_{50}$-value, the concentration at which half-maximal inhibition of luminescence was achieved. Bac7PS showed a mean reduction of ~10% of the IC$_{50}$ (Fig. 3.4c) compared to Bac7$_{1-23}$ at the *E. coli* ribosome, which did however not reach statistical significance (p-value=0.15, Wilcoxon rank-sum test). Surprisingly, Bac7PS displayed a reduced IC$_{50}$ of ~38% to the HEK ribosome when compared to Bac7$_{1-23}$ (p-value<10^{-5}, Wilcoxon rank-sum test). However, similarly to Bac7$_{1-23}$, the binding of Bac7PS to the eukaryotic ribosome remained much weaker (IC$_{50}$ = 4.23 µM) compared to the bacterial ribosome (IC$_{50}$ = 0.59 µM). Bac7PS thus remained a strong and selective inhibitor of the bacterial ribosome. We hypothesize that the identified cooperative amino acid combination incorporated into Bac7PS increases the binding of the peptide to the ribosomes compared

to Bac7$_{1-23}$, but that this improved binding cannot be significantly detected for the *E. coli* ribosome using the in vitro translation inhibition assay.

While low MIC values are desirable for any antibiotic substance, it is equally important that the substance is not too toxic to eukaryotic cells. We hence tested the influence of Bac7PS on cells of the human embryonic kidney cell line 293 (HEK) by quantifying the reduction of the compound 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT), which only occurs if cells are metabolically active and viable[137]. We measured an increase of toxicity of Bac7PS with an IC$_{50}$ of 521 µM compared to an 1460 µM for Bac7$_{1-23}$ (Table 3.2). However, the therapeutic index (TI) remained larger than 180 for both peptides, suggesting good suitability for the use in mammalian systems.
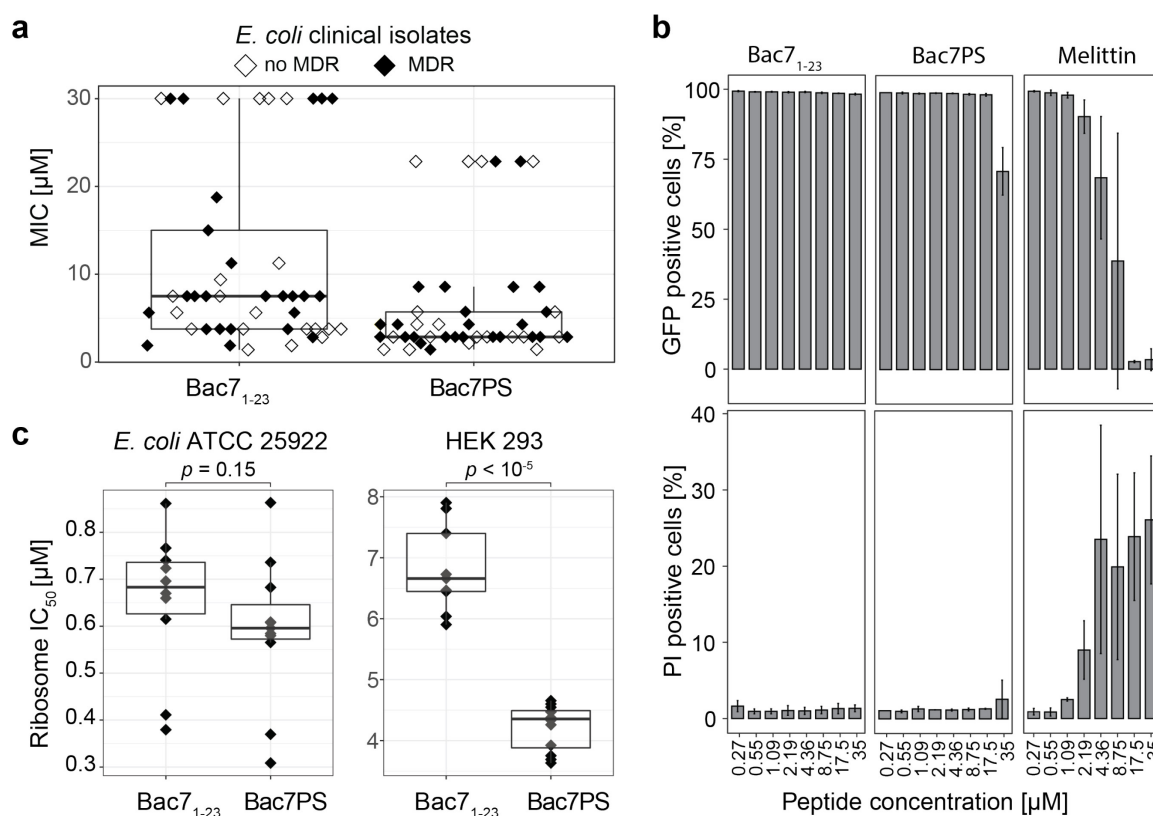

Fig. 3.4 | In vitro activity characterization of Bac7$_{1-23}$ and Bac7PS. **a)** MICs of a panel of clinical isolates of *E. coli* (n=45), including MDR bacteria (ESBL, CRE, n=25). **b)** Membrane damage assays measuring GFP loss (% of cells that lost GFP fluorescence) and PI uptake (% cells that gained PI fluorescence) when incubating *E. coli* TOP10 cells together with peptides added at different concentrations in full MHB II for 30 min. MIC of melittin against *E. coli* TOP10 is 5.0 µM (data not shown). **c)** In vitro translation assays. IC$_{50}$ values are extracted from a luminescence assay translating firefly luciferase mRNA and peptide concentration in a range between 800 µM and 0.08 µM (n=6).

Finally, we investigated the efficacy of Bac7PS in a murine septicemia infection model (Fig. 3.5a). The maximal dose at which all CD-1 mice survived intraperitoneal (IP) Bac7PS treatment for 3 days was 50 mg kg^{-1} (Supplementary Fig. 3.9). Additionally, we saw that a

second administration of 40 mg kg^{-1} Bac7PS was also tolerated by the animals (data not shown). For the efficacy study, we decided to use two concentration of Bac7PS and apply them twice: 30 mg kg^{-1}, a concentration close to the maximum tolerated dose, and 10 mg kg^{-1}. IP infection of CD-1 mice using *E. coli* ATCC 25922 resulted in death of all 10 mice after 72 h if treated using the vehicle control and survival of all mice when treated with 30 mg kg^{-1} ciprofloxacin (CIP) (Fig. 3.5b). Bac7PS showed a dose-dependent efficacy at which 80% of the mice survived at 30 mg kg^{-1} and 60% of mice survived at 10 mg kg^{-1} peptide concentration.
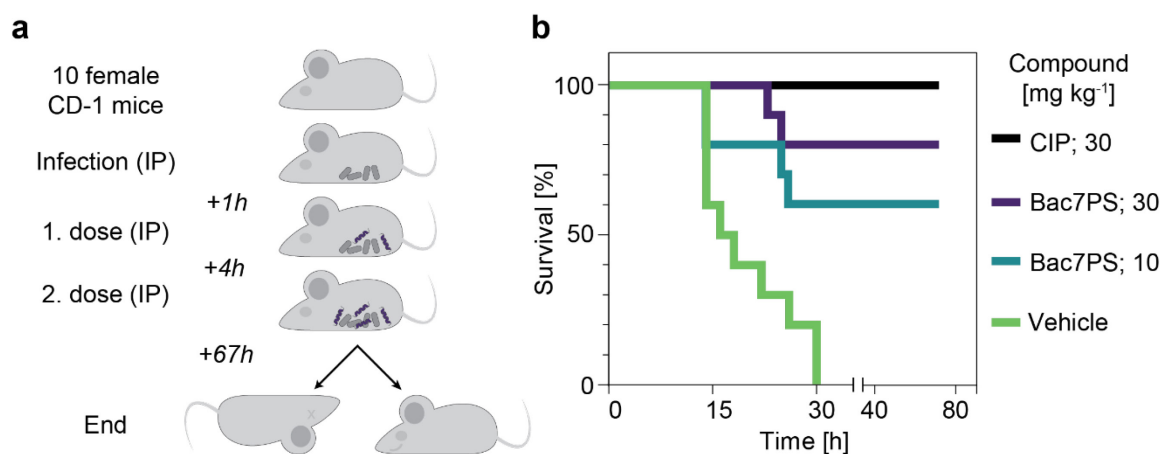

Fig. 3.5 | In vivo efficacy of Bac7PS. a) Efficacy study scheme, applying Bac7PS to mice infected with *E. coli* ATCC 25922. b) Survival rates after IP infection. Study was performed as described in a).

3.4 Discussion

DMS methods have been widely applied to study proteins[138], optimize enzymes[139], or to accelerate the understanding of therapeutic activity (including the study of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) therapeutics)[140]. One of the major advantages of DMS screens is to circumvent large-scale in vitro assessments of protein variants using recombinant expression and purification[124]. This advantage can be even larger when screening AMPs, which would otherwise need to be synthesized using solid-phase peptide synthesis. To the best of our knowledge, this is the first time DMS was applied to study and optimize an AMP.

We present for the first time an in-depth characterization of the importance of growth inhibition of each amino acid residue exchange in Bac7$_{1-23}$, which might guide future Bac7$_{1-23}$ optimization campaigns. Much of the information we acquired using intracellularly synthesized peptides is in high agreement with previously reported data using chemically synthesized peptides. We retrieved the core motif of Bac7$_{1-23}$ between positions 6 and 14 as most important which seems to be evolutionarily conserved among proline-rich AMPs[39]. By the random insertion of stop codons, we found that peptides shorter than 16 amino acids are inactive. As the peptides in our assay are synthesized in the cytosol, we attributed the loss of activity to weaker ribosomal binding, confirming earlier research[141] but contrasting recent claims that the loss of antimicrobial activity of Bac7$_{1-15}$ is due to impaired uptake into the cytosol[39,119]. Strikingly, we found that the C-terminus of Bac7$_{1-23}$ offered a large potential for optimization even though most research currently focuses on substitution of the first 16 residues[120,121]. Unlike the core-motifs of proline-rich AMPs, which well overlapped in the crystal structure when bound to the ribosome, the C-termini of those peptides showed large variations[127]. The higher spatial flexibility of the upper ribosome chamber is illustrated by the fact that also Gram-positive targeting macrolide antibiotics bind in that region[142] and that longer Bac7 variants (60 or 35 amino acids by length) showed a broader (against different Gram-negative bacteria) antimicrobial activity than their shorter counterparts[118]. Surprisingly, we found that insertion of a tyrosine residue is one of the most activity enhancing single amino acid residue substitutions in the site-specific library, especially at positions 18 and 19. Interestingly, even though a tyrosine residue is not present in full-length Bac7 comprised of 60 amino acids, it is common among other proline-rich AMPs[39,143], potentially because, similarly to arginine residues, an aromatic residue can well fill the space and interact with the amino acid residues or ribonucleotides of the surrounding ribosomal exit tunnel[120].

Concerning the applied intracellular screening method, we note that such a method might enable finding better binders to the ribosome or other intracellular targets, variants with increased membrane damaging properties at a lower concentration, or variants that reach higher intracellular concentration because of increased solubility, mRNA stability, or proteolytic resistance. However, uptake across the outer and inner membrane, for example through a transporter such as SbmA, cannot be selected for in screens in which the peptides are synthesized intracellularly. As for the proline-rich peptides used for this study the N-terminal 'RRIR' motif seems to be most critical for cellular uptake[118,141], we decided not to modify this part and accepted the risk for contracting unknown effects regarding uptake when modifying other regions of Bac7$_{1-23}$. However, in a scenario in which a variant is highly active if synthesized intracellularly but not taken up if added to cells from outside, its uptake could be increased post-screen, e.g. by fusing it to motifs of cell-penetrating peptides[144].

Our intricate analysis protocol to extract auspicious amino acid combinations enabled us to design Bac7PS, which excels over the Bac7$_{1-23}$ parent in terms of antimicrobial activity against our screening strain *E. coli* TOP10 even when the peptides are added extracellularly. We observed a broader activity spectrum of Bac7PS against clinical isolates including MDR bacteria compared to Bac7$_{1-23}$ and a decreased dependency on the SbmA transporter. As Bac7$_{1-23}$ was largely inactive against the SbmA transporter knockout strain (Table 3.2), this improvement is a very favorable new property, potentially making resistance development for bacteria harder. We partially attribute the activity gain to the increased ability to penetrate membranes as an alternative route of uptake, which was accompanied by a slight increase in membrane damage at higher concentration compared to Bac7$_{1-23}$. This hypothesis could be supported by the fact that aromatic amino acids, such as the now introduced tyrosine residue at position 18 or the phenylalanine residue at position 19, are known to enhance membrane penetration[145]. Very similar effects were observed by the study of Mardirossian *et al.* relying on the chemical synthesis of 133 Bac7$_{1-16}$ derivatives[120]. The most active candidate mentioned in that study, a peptide called B7-005 incorporating multiple tryptophan residues was also more active against a broader panel of bacteria and a SmbA knockout. However, B7-005 was not more active against the initial screening strain and seemed to show slightly decreased inhibition of protein translation. On the contrary, we hypothesize that part of the gain in activity of Bac7PS results from an increased binding to the ribosome, which might correlate with the observed cooperativity between the phenylalanine and tyrosine residues in our peptide. Even though the small difference (of 10%) in increased binding to the *E. coli* ribosome compared to Bac7$_{1-23}$ was not statistically significant, we measured an increased binding

to the human ribosome, which was however still bound at much lower affinity. Due to the absence of the SmbA transporter in human cells and the high selectivity of AMPs towards bacterial over human membranes, Bac7PS is a very promising candidate as an antimicrobial.

Bac7PS and Bac7$_{1-23}$ were non-toxic to human cells and displayed a far higher TI compared to most other AMP[146]. It is difficult to link the activity increase of Bac7PS compared to Bac7$_{1-23}$ measured in vitro, to in vivo experiments. There has only been a single report examining the efficacy of a Bac7 derivative in vivo[147]. Here, the authors treated mice infected with Gram negative *Salmonella typhimurium* intraperitoneally with 30 mg kg^{-1} of Bac7$_{1-35}$ (a 35 amino acid long version), which increased the mean survival from 10 to 24.5 days compared to untreated mice[147]. In our study, we showed that mortality can be strongly reduced in mice after a lethal infection with *E. coli*, causing death among all untreated mice within 30 h, by applying 10 mg kg^{-1} of our optimized compound Bac7PS.

We provided the first example of successful AMP optimization by DMS resulting in a potential AMP lead candidate. Our generated fitness landscapes might inspire further engineering approaches on Bac7$_{1-23}$ and the entire method could be easily expanded to studying structure-activity-relationships of other intracellularly active or even membrane damaging AMPs if displaying the peptides on the surface[91]. We envision coupling this method to directed evolution approaches to further increase the activity of AMP or using it in drug-resistant strains.

3.5 Methods

Chemicals and reagents

Unless otherwise stated, all chemicals, reagents and primers were obtained from Sigma Aldrich (Buchs, CH). Restriction enzymes and their buffers were obtained from New England Biolabs (Ipswich, USA). Kits for the isolation of plasmid isolation and DNA purification kits were obtained from Zymo Research (Irvine, USA). Peptides in either purified (>90%) or crude format were obtained from Pepscan (Lelystad, NL) or Genescript (Piscataway Township, USA). Sanger-sequencing was done at Microsynth (Balgach, CH).

Bacterial strains and cultivations

Unless otherwise stated, all experiments were performed using *E. coli* TOP10 (F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L(Str^R) *end*A1 *nup*G; Thermo Fisher Scientific, Waltham, USA). In this study, overnight cultivations were performed either in 14 ml polypropylene tubes (Greiner, Kremsmuenster, AT), filled with 5 ml of lysogeny broth (LB) medium (Difco, Becton Dickinson, Franklin Lakes, USA), or in 96-deep-well polypropylene plates (Greiner, Kremsmuenster, AT) filled with 500 µl of LB-medium. All samples were incubated at 37°C with agitation on a shaker (Kuhner, Birsfelden, CH) operated at 200 r.p.m. and 25 mm amplitude. All media were supplemented with the appropriate antibiotic for plasmid maintenance (50 µg ml⁻¹ kanamycin; 100 µg ml⁻¹ carbenicillin) and 1 % (w/v) D-glucose for repression of gene expression from catabolite-repression sensitive promoter P$_{BAD}$. In the case of peptide expression experiments, cultures were incubated without D-glucose and 0.3 % (w/v) of the inducer L-arabinose was used for induction. For all cultivations on solid medium, 15 mg ml⁻¹ agar (Difco) was added to the broth, and incubation was performed in an incubator (Kuhner) at 37°C. If not indicated differently, the optical densities (OD) of bacterial cultures were determined by measuring light scattering at 600 nm using a UV/VIS spectrophotometer (Eppendorf, Hamburg, DE).

Generation of the randomly mutagenized Bac7$_{1-23}$ library

To mutate the Bac7$_{1-23}$ gene randomly, we used an error-prone polymerase chain reaction (epPCR). We amplified the Bac7$_{1-23}$ gene using primer 1 and primer 2 (Table 3.3), which bind upstream of the first codon, including the start codon, and downstream of the last codon, not including the stop codon. For amplification, we used the *Pfu* DNA Polymerase

exo⁻ mutant[§§] (D141A/E143A), *Pfu* reaction buffer including dNTP's and 0.3 mM final concentration of $MnCl_2$ to increase the error rate. The amplification was performed using 30 cycles of 98°C for 10 s, 60°C for 15 s, and 72°C for 10 s. The PCR product (118 bp in length) was purified using a 1.5% agarose gel and a DNA gel recovery kit. 500 ng of the purified product was used for a restriction digest using enzymes HindIII-HF and PstI-HF in Cutsmart buffer. The digested product was again purified using a DNA purification kit and ligated to plasmid pBAD[104] (Thermo Fisher Scientific) previously digested with the same enzymes using T4 ligase (800 units). The ligation product was dialyzed in deionized water using filters (MilliporeSigma, Burlington, USA) and 1 µl was used to transform 20 µl of CloneCatcher™ Gold DH5G Electrocompetent *E. coli* (Genlantis, Burlington, USA) cells using electroporation. Recovered cells were plated and incubated overnight on LB agar plates supplemented with carbenicillin. Approximately 5.6 million colonies were washed off several plates using LB medium and the plasmids containing the peptide-encoding DNA sequences were extracted from $2.5*10^9$ cells using a plasmid isolation kit. An aliquot of 5 ng of these plasmids was used to transform *E. coli* TOP10 cells using the protocol from the transformation above. Approximately 10 million colonies were recovered from the plates after overnight incubation by washing with LB medium, the suspension was diluted to OD = 1 with LB medium, glycerol was added to a final concentration of 20% (v/v), and aliquots of 500 million cells were stored at -80°C.

Generation of the focused Bac7$_{1-23}$ library

A focused, semi-rational strategy was pursued to generate diversity at positions 5, 18, 19, and 20 in the Bac7$_{1-23}$ peptide. These sites were simultaneously randomized on the genetic level by site-saturation mutagenesis using a single NNK codon for position 18 and a mixture of codons as suggested in the small intelligent approach[129] for the remaining three positions (see Table 3.3 for sequences and mixing ratios). Degeneration was introduced by using the QuikChange technique[148], amplifying the pBAD plasmid containing the Bac7$_{1-23}$ gene using a mixture of 20 oligonucleotides at a final concentration of 0.3 µM and Phusion® High-Fidelity PCR Master Mix in HF buffer. The amplification was performed using 20 cycles of 98°C for 10 s, 60°C for 15 s, and 72°C for 10 s. The PCR product was treated with the enzyme Dpn1 to remove the template plasmid and subsequently purified using a DNA purification kit. The purified product was used to transform 20 µl of CloneCatcher™ Gold DH5G Electrocompetent *E. coli* (Genlantis, Burlington, USA) cells using electroporation. Transformation, recovery and storage was performed as explained

---

[§§] Generated, produced, and purified by Dr. Luzius Pestalozzi from the ETHZ BPL

for the random mutagenesis protocol, but this time we recovered approximately 2.3 million (*E. coli* Clonecatcher), and approximately 1.9 million (*E. coli* TOP10) colonies.

Table 3.3 | Primers and genes used

ID	Sequence (5'-3')	Description	Mixing ratios site-saturation
Primer 1	TGTCTGCAGAGGAGATATAAATG	amplification forward	
Primer 2	TGCACAAAGCTTACGTG	amplification reverse	
Primer 3	GTCTGCAGAGGAGATATAAATGCGGAGAAT AAGANDTCGGCCACCTAGACTGCC	forward primer site-saturation	240
Primer 4	GTCTGCAGAGGAGATATAAATGCGGAGAAT AAGAVMACGGCCACCTAGACTGCC	forward primer site-saturation	120
Primer 5	GTCTGCAGAGGAGATATAAATGCGGAGAAT AAGAATGCGGCCACCTAGACTGCC	forward primer site-saturation	20
Primer 6	GTCTGCAGAGGAGATATAAATGCGGAGAAT AAGATGGCGGCCACCTAGACTGCC	forward primer site-saturation	20
Primer 7	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GAHNAHNMNNTGGACGCGGGCGC	reverse primer site-saturation	144
Primer 8	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GTKBAHNMNNTGGACGCGGGCGC	reverse primer site-saturation	72
Primer 9	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GCATAHNMNNTGGACGCGGGCGC	reverse primer site-saturation	12
Primer 10	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GCCAAHNMNNTGGACGCGGGCGC	reverse primer site-saturation	12
Primer 11	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GAHNTKBMNNTGGACGCGGGCGC	reverse primer site-saturation	72
Primer 12	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GTKBTKBMNNTGGACGCGGGCGC	reverse primer site-saturation	36
Primer 13	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GCATTKBMNNTGGACGCGGGCGC	reverse primer site-saturation	6
Primer 14	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GCCATKBMNNTGGACGCGGGCGC	reverse primer site-saturation	6
Primer 15	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GAHNCATMNNTGGACGCGGGCGC	reverse primer site-saturation	12
Primer 16	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GTKBCATMNNTGGACGCGGGCGC	reverse primer site-saturation	6
Primer 17	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GCATCATMNNTGGACGCGGGCGC	reverse primer site-saturation	1
Primer 18	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GCCACATMNNTGGACGCGGGCGC	reverse primer site-saturation	1
Primer 19	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GAHNCCAMNNTGGACGCGGGCGC	reverse primer site-saturation	12
Primer 20	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GTKBCCAMNNTGGACGCGGGCGC	reverse primer site-saturation	6
Primer 21	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GCATCCAMNNTGGACGCGGGCGC	reverse primer site-saturation	1
Primer 22	ACAAAGCTTACGTGCTGACTTAAGGCCGAG GCCACCAMNNTGGACGCGGGCGC	reverse primer site-saturation	1
Bac7$_{1-23}$ gene	ATGCGGAGAATAAGACCTCGGCCACCTAGA CTGCCTAGACCGCGCCCGCGTCCATTACCA TTCCCTCGGCCTTAA	Bac7$_{1-23}$ gene following the P_{BAD} promoter	

Single-strain growth experiments

To assess the antimicrobial effect of single peptides when expressed intracellularly in *E. coli* TOP10, a monoclonal strain carrying the pBAD plasmid containing a single peptide was picked from solid media, incubated overnight, and inoculated into a fresh LB medium containing 0.3 % (w/v) L-arabinose to a final OD of 0.01 into 96-well microtiter plate

(Greiner). Growth of strains was recorded by measuring OD in a Tecan Infinite 200 PRO (Tecan, Männedorf, CH) for at least 4 h (37°C, 1.5 mm orbital shaking).

Parallel growth experiment

To assess the antimicrobial effect of multiple peptides in parallel when expressed intracellularly in *E. coli* TOP10, previously prepared aliquots containing *E. coli* plasmid libraries created by either random mutagenesis or site-saturation mutagenesis were used. For both approaches, three aliquots containing approximately 500 million cells each of *E. coli* TOP10 harboring peptide-encoding DNA sequences on the pBAD plasmid were thawed and added to three 1 l baffled shake flasks containing 100 ml of LB medium + 100 µg ml^{-1} carbenicillin. The cultures were grown for roughly 7 h at 37°C. When the OD reached approximately 0.2, the cultures were supplemented with L-arabinose to a final concentration of 0.3 % (w/v) to induce peptide expression. When analyzing the randomly mutagenized library, cell samples were taken from each biological replicate at the point of induction and 4 h post-induction. When analyzing the library created by site-saturation mutagenesis, cell samples were taken from each biological replicate at the point of induction and 4.5 h post-induction. The plasmids were extracted from all samples using a plasmid isolation kit.

NGS

Peptide-encoding DNA sequences on plasmids, collected from both experiments (three replicates across the different time points) were sequenced using NGS. To first amplify the peptide-encoding DNA we used primer 1 and primer 2 (Table 3.3), 100 ng of plasmid, and using 10 cycles of 98°C for 10 s, 60°C for 15 s, and 72°C for 10 s. The amplification product was purified using an agarose gel. Single Index PentAdapters from Pentabase were used to prepare PCR-free libraries with the KAPA HyperPrep Kit (now Roche, Basel, CH) according to the manufacturer's specifications. Libraries were quantified using the qPCR KAPA Library Quantification Kit. Libraries were pooled and sequenced single read with 101 cycles using an Illumina NovaSeq 6000 SP flow cell. Roughly 10% genomic PhiX library was used as spike-in to increase sequence diversity. Base-calling was done with bcl2fastq v2.20.0.422. The resulting fastq files were processed using the software Geneious Prime 2020 (Biomatters, Auckland, NZ) and an in-house software written in R. For the randomly mutagenized library, we first discarded all sequences that missed the combination of a start codon and 69 bases downstream stop codon. Next, we discarded all sequences that appeared less than five times in at least two replicates. Next, all DNA sequences were translated and the resulting peptide sequences were counted for each

replicate and time point. All NGS counts can be seen in Supplementary Table 3.3. For the site-saturation library, we first discarded all sequences that did not have a start codon and 69 bases downstream stop codon, translated the DNA sequences into peptide sequences, aligned them to our reference table of 160,000 possible peptide variants, and counted them for each replicate and time point. All NGS counts can be seen in Supplementary Table 3.4.

Ranking of peptides based on log2-fold change derived growth inhibition

To analyze the NGS read counts of both libraries, we used the standard workflow of DESeq2[105] (NGS read count normalization, dispersion estimates, and Wald's test). We calculated the log2-fold changes of the NGS read counts (listed for each peptide in Supplementary Table 3.3 and 3.4) between the time of induction and 4.0 h (random mutagenesis) as well as 4. h (focused library) post-induction. A Bayesian shrinkage estimator was employed to shrink the log2 fold-change for each sequence using the R/Bioconductor package 'apeglm'[106]. Finally, the shrunken log2 fold-change was used as a proxy for growth inhibition of each peptide, as the propagation rate of the peptide-encoding DNA would follow the growth rate of the respective host. The most growth inhibitory peptide corresponded to the most negative fold change; the least active peptide corresponded to the most positive fold change. Thus, a ranking from least to most active peptides of both libraries could be established. The ranked peptide list from the randomly mutagenized library can be found in Supplementary Table 3.5. The ranked peptide list from the focused library can be found in Supplementary Table 3.6.

Enrichment curve-derived AUCs to quantify effects on growth inhibition

To determine the effect of amino acid residue substitutions on growth inhibition, we applied a variation of the gene set enrichment analysis (GSEA) proposed by Subramanian *et al.*[131] This adjusted method was based on drawing what we refer to here as *enrichment curve plots* (see Fig. 3.2a). In those plots, each value on the x-axis represented a peptide ranked by their shrunken log2 fold-changes, giving rise to the ranked peptide set L. More active peptides are assigned to the left spectrum of the x-axis, while less active peptides are assigned to the right spectrum of the x-axis. For each single amino acid residue substitution, e.g. alanine at position 1, we defined S to be a set of all peptides that exhibit this substitution. Each y-value indicated whether the corresponding peptide pertains to the peptide-set under study or not. Formally, if a peptide p_i in the ranked list L pertains to the peptide set S, its value is defined as

$$P_{hit}(S, i) = \sum_{p_i \in S; j \leq i} \frac{1}{|S|} \qquad (1)$$

If, on the other hand, peptide p_i is not present in the set S, its value will correspond to

$$P_{miss}(S, i) = \sum_{p_i \notin S; j \leq i} \frac{1}{N - |S|} \qquad (2)$$

where N corresponds to the total number of peptides in the ranked list.

To describe the enrichment concisely, Subramanian *et al.* developed a so-called enrichment score (ES) that is defined as the maximum deviation of $P_{hit} - P_{miss}$ from zero. Here, we proposed a slightly different approach, which we referred to as the *area under the curve* (AUC). Compared to the ES, the AUC describes the complete dynamics of the enrichment curves. We computed the AUCs as follows:

$$AUC(S) = \frac{1}{|L|} \sum_{i=1,...,|L|} [P_{hit}(S, i) + P_{miss}(S, i)]. \qquad (3)$$

Positive AUC-values indicated that the corresponding set S is overrepresented in growth inhibitory peptides (top of the list); while negative AUC values indicate that the corresponding peptide set is overrepresented in less growth inhibitory peptides (bottom of the list). AUC values close to zero indicated that the peptide set was randomly distributed across the list, or exhibited bimodal behaviors.

In cases where the peptide-set S is small, i.e. if there are only few observations for an amino acid residue substitution, the computation of meaningful enrichment curves and AUCs can be misleading. To present a more robust measurement for the randomly mutagenized Bac7$_{1-23}$ library, in which many amino acid residue substitutions were seldom, we resorted to a permutation scheme that allowed us to derive better estimates of the effect of each amino acid residue substitution on growth inhibition. These permutation schemes relied on drawing from the null distribution, i.e. assuming that a single substitution, in the following denoted as S_M, is not overrepresented in either or lower active peptides. The following scheme is executed for each permutation:

1. Randomly permute the peptides in list L, giving rise to permuted list L, which destroys the activity-based ranking

2. Compute enrichment curves for permuted list L and single substitution S_M

3. Compute the permuted $AUC(S_M)$

This scheme is repeated n_{perm} times, where commonly $n_{perm} = 10^3$, or 10^4, giving rise to the null distribution of the AUC values. This distribution can then be used to derive z-scores as

$$z_{S_M} = \frac{auc(S) - mean(auc(S_M))}{std(auc(S_M))} \qquad (4)$$

Z-scores represented how many standard deviations the observed AUC-value differed from the mean of all AUC-values derived from the null distribution. A large positive z-score indicated an enrichment of S_M among higher growth inhibitory peptides, a large negative z-score denote enrichment of S_M among less growth inhibitory peptide. Z-scores close to zero imply that the observed measurement lies close to the mean of the null distribution. We furthermore computed a two-sided p-value (p_{S_M}) to assess the statistical significance of the observed measurements under the null hypothesis. It is defined as

$$p_{S_M} = \frac{\#abs(auc(S_M) \geq auc(S_M))}{n_{perm}} \qquad (5)$$

To account for multiple testing, p-values were adjusted using the Benjamini-Hochberg procedure with a false discovery rate of $\alpha = 0.1$.

Significant pattern mining to rank amino acid residue combinations

Significant pattern mining emerged recently within the field of machine learning[7] and is devoted to finding patterns that occur significantly more often in one group of observations versus another group of observations. Here, we defined a pattern to be any combination of amino acids residues present in our data set. In our specific case, there exist patterns of length one, two, and three, referred to as single, double, or triple combinations in the main text, respectively. To find significant patterns in the data set, we first had to generate two classes. We achieved this by using our activity-based ranking of peptides according to the shrunken log2 fold change and focus our analysis on the most and least growth-inhibitory 10% and 25% of all peptides. To identify patterns that occur significantly more often in either the strongest or the weakest sequences, we applied a tool named fast automatic conditional search (FACS)[8]. It is based on the creation of a 2-by-2 contingency table for each pattern, and a subsequent two-sided Fisher's exact test (enrichment in either 10% or 25% most or least growth inhibitory peptides). Results from all triple combinations can be seen in Supplementary Table 3.1.

Determination of interactions between amino acid residue combinations

The AUC describes the effect of amino acid residue substitutions on growth inhibition and can be calculated as described in equation (1-3). We defined the conditional AUC as the AUC calculated for a single substitution S using a ranked peptide set L' containing only peptides that exhibition a fixed amino acid residue at a specific position. To determine if two amino acid residues interact (behave non-additively), we calculated the difference between the AUC calculated for the conditional ranked peptide set L' (with a size of approximately 8,000 peptides) and the AUC calculated for the entire ranked peptide set L (with a size of approximately 160,000 peptides). See Fig. 3.3a for examples. These differences were calculated for all possible 4'800 (considering both directions of the combinations, as they are not symmetrical) double combinations (see Supplementary Table 3.2). By creating a boxplot of all resulting values, we defined that all amino acid residue combinations which resulted in values that lie in between both whiskers of the boxplot (1.5 times the lower limit of the IQR to 1.5 times the upper limit of the IQR) do not show an interaction, i.e. behave additively. Non-additivity in the form antagonism, i.e. one amino acid residue decreased the effect of the other amino acid residue on growth inhibition, was discovered for combinations with values below -0.09. Non-additivity in the form of cooperativity, i.e. one amino acid residue increased the effect of the other amino acid residue on growth inhibition was discovered for combinations with values above 0.9.

Purification of chemically synthesized peptides

Peptide Bac7$_{1-23}$ (H-RRIRPRPPRLPRPRPRPLPFPRP-OH) and Bac7PS (H-RRIRIRPPRLPRPRPRPYFMPRP-OH) were obtained from Pepscan (Lelystad, NL) or Genscript (Piscatawa, USA) in >90% purity or in crude format and subsequently purified to >90% purity in house. For the latter, crude peptides were dissolved in 5 ml DMSO and 15 ml 0.1% aqueous trifluoroacetic acid, TFA. RP-HPLC-purification of the dissolved crude peptides was performed on an ÄKTAexplorer chromatography system (GE Heathcare, SE). The entire peptide sample was loaded onto a C18 column (PRONTOSIL 120 C18 AQ 10 μm, 250 x 20 mm, 50 x 20 mm precolumn, Bischoff, Leonberg, DE), heated to 30°C and operated at a flow rate of 10 ml min^{-1} using 0.1% aqueous TFA as solvent A and acetonitrile supplemented with 0.1% TFA as solvent B. The ratios of A to B were adapted for each peptide and typical values are given below. The column was equilibrated with the peptide-specific mixture of solvent A and solvent B (0-20%) prior to injection. After injection and an initial wash step of 6 min, a gradient was imposed with the same eluent mixture, and then a gradient was applied, in the course of which the amount of solvent B was increased to 50-90 % in 40 min. The column was washed with 95 % solvent B for 8 min and equilibrated with the specific solvent A/solvent B mixture for the next run for 13 min.

Peptide elution was monitored spectrophotometrically at 205 nm and generally, the main peptide peak was collected. The sample was frozen at -80 °C for >2 h and lyophilized (approx. 18 h) using a freeze-dryer (Alpha 2-4 LDplus, Christ, DE), connected to a vacuum pump (RC6, Vacuubrand, DE). The lyophilized peptides were dissolved in 1 ml DMSO and stored at -20°C. The concentration of the peptide stocks was determined via HPLC using an Agilent 1200 series RP-HPLC system. Each peptide stock was analyzed as a 1:100 dilution in water. An aliquot of 10 µl of the peptide stock was injected onto an C18 column (ReproSil-Pur Basic C18, 50 x 3 mm, Dr. Maisch, DE) operated with water supplemented with 0.1 % TFA as solvent A and acetonitrile supplemented with 0.1 % TFA as solvent B. Separation was performed at a flow rate of 0.6 ml min^{-1} using the same concentration profile previously used for purification. The concentration was measured using the integrated peak area at 205 nm and then calculated using peptide-specific absorption properties[109,110].

Measurement of the MIC

Bacterial cells were grown in cation-adjusted Mueller Hinton Broth (MHB II) overnight to stationary phase. The cultures were then supplemented with 200 g l^{-1} glycerol, aliquoted, and frozen at -80°C. For MIC measurements, an aliquot of the cells was thawed, resuspended in MHB II to a final volume of 750 µl and cell concentration of 1 x 10^6 CFU ml^{-1}. The purified peptides were thawed and the concentration was determined by RP-HPLC as described before. The peptides were diluted with sterile water to 4-fold the desired assays starting concentration and to a final volume of 50 µl. Pipetting of the MIC dilution series was done by a Hamilton Microlab STAR Liquid Handling System (Hamilton, Bonaduz, CH) and in 384-well plates (PP, F-bottom, 781201, Greiner, Kremsmünster, AT) with a final assay volume of 40 µl. The first well of each MIC dilution series was filled with 20 µl of 2-fold concentrated MHB II, wells 2-11 with 20 µl of MHB II, and well 12 with 40 µl MHB II (sterility control). Next, 20 µl of the peptide dilution was added to the first well, mixed, and a log2 serial dilution was performed from well 1-10 (20 µl transfer volume). Well 11 served as growth control (i.e., no peptide added). In the last step, 20 µl of the bacterial cell suspension was added to well 1-11 either using the pipetting robot (*E. coli* TOP10, BW25113, BW25113 Δ*sbmA* and ATCC 25922) or by a manual pipette (*E. coli* clinical isolates). The plates were sealed airtight and incubated for 18 h without shaking at 37°C before reading the OD using an Infinite 200 PRO plate reader (Tecan, Männedorf, CH). The MIC value corresponded to the concentration at which no growth of the bacterial strain was observed (< 5% of the OD value of the growth control) and was evaluated using

a custom-written script in the programming language R. Each MIC value was determined in technical replicate.

Membrane damage assay

For membrane damage assays, the bacterial strain *E. coli* ATCC 25922 [pSEVA271-sfGFP] and the peptide dilutions were prepared as described for the MIC measurements but by scaling all volumes 5-fold and using 96-well plates (PP, U-bottom, 650201, Greiner, Kremsmünster, AT) with a final assay volume of 200 µl. The bacterial strain suspension was furthermore supplemented with 20 µg ml^{-1} propidium iodide (PI, from a 1 mg ml^{-1} stock in DMSO) just before pipetting the assay. After 1 h incubation at room temperature the cell membrane integrity was assessed by flow cytometry using a Fortessa Analyzer (BD Biosciences) and appropriate filters for GFP and PI (488 nm laser with 530/30 nm bandpass filter and 579 nm laser with 610/20 nm bandpass filter). The fractions of PI-positive and PI-negative cells, as well as GFP-positive and GFP-negative cells, were determined with the same gate for all populations using the FlowJo V10 software (BD Biosciences). The membrane integrity assay was performed in triplicate.

In vitro toxicity assay

HeLa epithelioid cervix carcinoma cells (93021013, Sigma Aldrich) were routinely cultivated in Dulbecco's MEM High Glucose (DMEM, with L-Glutamine, without phenol red, 1-26P32, Bioconcept, Allschwil, CH), supplemented with 10% Fetal Bovine Serum (FBS, heat inactivated, sterile filtered, F9665, Sigma Aldrich) and 100 IU ml^{-1}/100 µg ml^{-1} Penicillin/Streptomycin (4-01F00, Bioconcept) at 37°C with 5% CO_2. Cells were split at a confluency of ≤ 90% (every three to four days) and maintained for max. 10 passages before a fresh aliquot of cells were seeded. For the tox assay, cells were cultivated for at least two passages after thawing, detached from the cultivation flask using Trypsin-EDTA (25300054, Gibco, Thermo Fisher Scientific), centrifuged at 200 x g for 4 min, and washed once by resuspending the pellet in an equal volume of Dulbecco's Phosphate Buffered Saline (DPBS, D8537, Sigma Aldrich). The DPBS was removed by another centrifugation step and the cell pellets were resuspended in fresh, prewarmed DMEM. The cell concentration was determined using a Countess 2 device (Thermo Fisher Scientific) and approx. 5000 cells were seeded into wells of a 96-well plate (F-bottom, PS, 655180, Greiner) together with 100 µl DMEM. The last row of each plate was filled with DMEM only. After cell seeding, the plate was incubated for 24 h at 37°C with 5% CO_2. The following day, a log2 dilution series of the peptides were prepared as described for the MIC assays but using a 96-well plate (V-bottom, PP, 651201, Greiner) with a final volume of 50 µl. For

the first well, 2-fold concentrated DMEM medium was used, in wells 2-9 and 11, DMEM medium was used. Well 10 served as killing control (100% DMSO) and well 11 as non-treated control (no peptide added). From the cell culture plate, 50 µl of the medium in each well (except the last row) was removed, discarded, and replaced with the 50 µl of liquid from the equivalent well on the peptide dilution plate. The plate was incubated again for 24 h at 37°C with 5% CO_2. After incubation, 10 µl of (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) solution (from a 5 mg ml^{-1} stock in DPBS) was added to each well. The plate was then incubated for 2 h. After incubation, the cell culture medium containing residual MTT was removed from each well. The formed formazan crystals were dissolved by adding 100 µl of DMSO to each well. The formazan content in each well was determined by measuring the absorbance at 575 nm using an Infinite M1000 PRO plate reader (Tecan) and corrected for light scattering by subtracting the absorbance at 690 nm used as the reference wavelength. For each dilution series, the IC$_{50}$ value was determined by computing a weighted n-parameters logistic regression using the "nplr" package in R (https://CRAN.R-project.org/package=nplr). The in vitro toxicity assay was performed in triplicate.

Hemolysis assay

Mouse blood was obtained from the ETH Phenomics Center. The erythrocytes were isolated by centrifugation at 500 x g for 10 min and removal of the blood plasma. The cells were washed three times by gently resuspending them in an equal volume of DPBS followed by centrifugation. After the last resuspension, the cells were diluted 1:50 in DPBS. For the hemolysis assay, a log2 serial dilution of each peptide was prepared as described for the MIC but using DPBS and a 96-well plate (U-bottom, PP, 650201, Greiner) with a final volume of 100 µl. As lysis control, 2.5% Triton-X100 in DPBS was used in well 10, well 11 served as non-treated control (no peptide added), and well 12 as blank. To each well of the dilution plate, 100 µl of the red blood cells suspension was added. The plate was incubated for 1 h at 37°C. After the incubation, the plate was centrifuged at 500 x g for 10 min and 100 µl of the supernatant was transferred to a clean 96-well plate (F-bottom, PS, 655101, Greiner). The absorbance was measured at 540 nm using an Infinite M1000 PRO plate reader (Tecan) and corrected by the measurements from the untreated wells. The lysis of each peptide concentration was expressed relative to the lysis control (set as 100% lysis). The hemolysis assay was performed in triplicate.

In vitro translation inhibition assay

To measured inhibition of the *E. coli* ATCC 2522 or HEK 293 ribosome, an S30 extract from *E. coli* or HEK 293 cells was purified according to two protocols described elsewhere[149,150]. $Bac7_{1-23}$ and Bac7PS were dissolved in water and 0.3% Tween20 at a concentration of 3 mM. The two peptides were dispensed into white 96-well plates (Eppendorf) using the TECAN D300e digital dispenser using the largest concentration of 800 µM and then in 2.5-fold dilutions steps until a concentration of 0.08 µM was reached. The suspension volume for all drugs was normalized to 1.5 µl (*E. coli* S30) and 2.0 µl (HEK S30) containing 0.3% Tween20. The *E. coli* translation master mix contained 4 µl *E. coli* S30 extract, 0.2 mM amino acid mix, 6 µg tRNA (Sigma), 0.4 µg hFluc mRNA, 0.3 µl protease inhibitor (cOmplete, EDTA-free, Roche), 12 U RNAse inhibitor (Ribolock, Thermo Scientific), 1.3 µl H2O and 6 µL S30 premix without amino acids (Promega) per 15 µl total reaction volume. The HEK translation master mix contained 7 µl HEK S30 extract, 20 mM Hepes-KOH buffer pH 7.4, 95 mM potassium acetate, 10 U RNAse inhibitor (Ribolock, Thermo Fisher), 0.125 mM amino acid mix, 12.5 mM creatine phosphate, 0.25 U creatine phosphokinase, 1.25 mM ATP, 0.25 mM GTP, 0.5 µg hFluc mRNA, 1.875 mM DTT and 2.9 mM magnesium acetate per 20 µl total reaction volume. 13.5 µl (*E. coli*) and 18 µl (HEK) of the translation master mix was added to each well to result in a total reaction volume of 15 µl and 20 µl, respectively. Plates were sealed with transparent foil and incubated for 1 hour at 37°C. After incubation, 75 µl of luciferase assay reagent (Promega) was added to each well, and luminescence was read using the plate reader BIO-TEK FLx800 (Witec AG, Littau, CH). Regression analysis for IC_{50} calculation was performed using Graphpad Prism version 8.3.0 by using the built-in equation: [log(inhibitor) vs. response – Variable slope (four parameters)] with the built-in fitting method: least squares (ordinary) fit. Y=Bottom + (Top-Bottom)/(1+10^((X-LogIC50))).

Animals

CD-1 mice (7 weeks-old, 27-28g female) were used (Charles River, France). These animals were housed for a week of acclimation period before experiment in a protected area in the 'Centre de Zootechnie de l'Université de Bourgogne' (Biosafety level 2 facility) and fed ad libitum according to the current recommendations by the European Institute of Health. Housing took place in rooms where a 12h/12h light/dark cycle is applied, the temperature ranges from 18 to 21°C and the relative humidity from 45 to 65%. Animal facility is authorized by the French authorities (Agreement N° C 21 464 04 EA). Animal housing and experimental procedures were performed according to the French and European Regulations and NRC Guide for the Care and Use of Laboratory Animals. All

procedures using animals were submitted to the Animal Care and Use Committee C2EA agreed by French authorities. Any animal showing poor conditions (20% body weight loss, signs of pain or distress, lack of activity) was humanely euthanized.

In vivo toxicity and efficacy

For animal experiments, Bac7PS was synthesized at Genscript as acetate salt and with a purity of 92%. The peptide was reconstituted in Dulbecco's phosphate-buffered saline (DPBS) at a concentration of 200 mg ml^{-1} and sterile filtered. All animal experiments were performed at Vivexia (Dijon, FR) according to a protocol submitted and approved by the local ethic committee and authorities (Ethics Committee of Burgundy and the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation). First, the maximum tolerated dose (MTD) for the peptide was determined by testing different peptide doses in 8 groups with 5 animals per group. The peptide solution was administered once by intraperitoneal (IP) injection with a volume of between 94 and 240 µl (depending on the weight of the animal and the dosage) and the animals were monitored for 2 h hours following the injection, then 6 to 8 h later, and then once or twice a day, depending on the clinical status, up to 5 days post injection. For each group a different, predefined peptide dose was tested (500, 50, 100, 75, 15, 25, 30, and 40 mg kg^{-1}) and a last group was tested at two injections (each 40 mg kg^{-1}), administered with 4 h time difference. The MTD was defined as dose where no dead animals were observed 2 days after injection. Second, the *in vivo* efficacy of the peptide was tested in a murine septicemia model induced by *E. coli* ATCC 25922. For this, a total of 4 groups with 10 animals per group were infected by IP injection of the bacterial inoculum (1 x 10^6 CFU per animal, +5% mucin) and each group was treated differently: A first group received ciprofloxacin (as positive control) administered IP, once (1 h post infection) at a dose of 30 mg kg^{-1}. A second group received DPBS (as vehicle control) administered IP, once (1 h post infection). The two other groups received the peptide, administered IP, twice (1 and 4 h post infection) either at a dose of 10 mg kg^{-1} or of 30 mg kg^{-1}. As endpoint the wellbeing and survival rate, on a twice daily based evaluation for up to 3 days, was monitored.

Data availability

All Supplementary tables and supplementary data can be accesses using the following link   https://polybox.ethz.ch/index.php/s/FF8FLP7fPK8FSG1   and   password BPL2021+Bac7PS. The computational workflow to reproduce the NGS count data analysis   and   ranking   of   peptides   is   available   on   GitHub (https://github.com/derpkoch/Bac7). NGS data are available at the NCBI Sequence Read

Archive (SRA) under accession number PRJNA730488. Additional data that support the findings of this study are available from the corresponding author upon reasonable request.
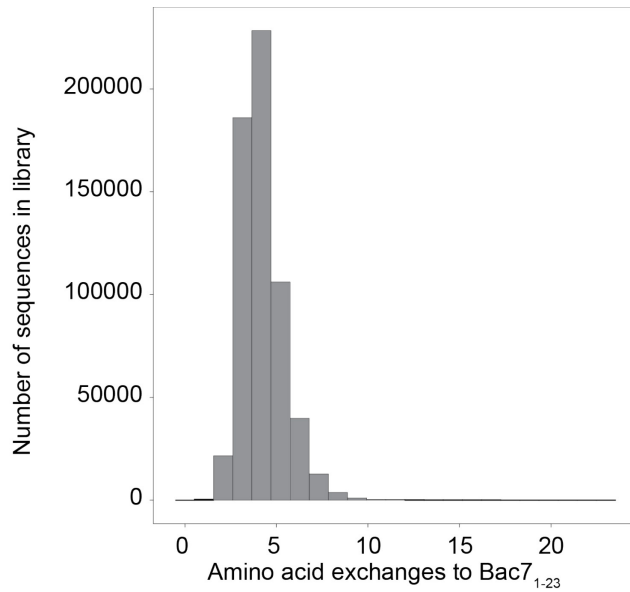
3.6 Supplementary figures
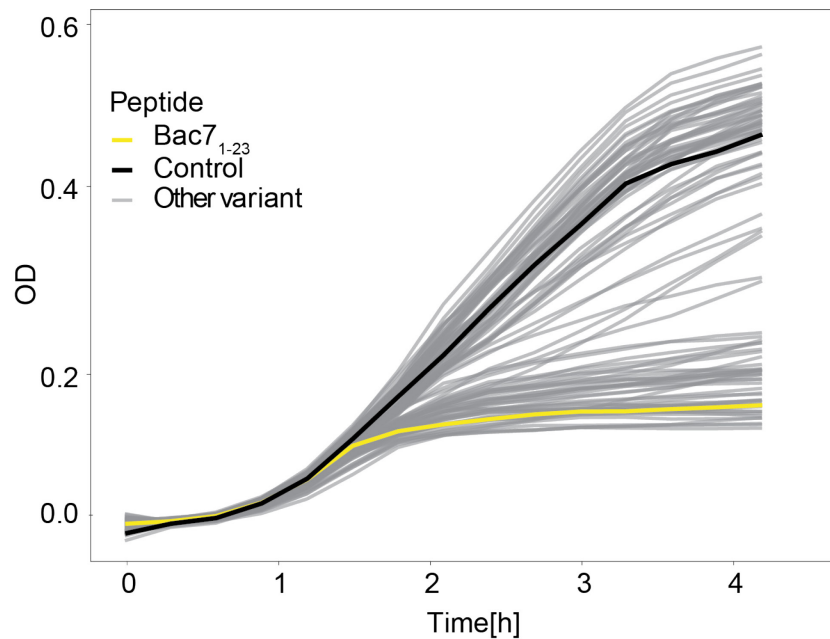
Supplementary Fig. 3.1 | Intracellular expression of randomly mutated Bac7$_{1-23}$ variants. 94 randomly picked *E. coli* TOP10 strains harboring 94 Bac7$_{1-23}$ variants were grown in microtiter plates (yellow lines). Peptide expression was induced at the start of incubation. Bac7$_{1-23}$ wild-type was added as the positive control (purple line). The inactive control peptide HNP-1 was added as the negative control (green line). 45% of all peptides do not reach >50% of the final OD of the negative control.
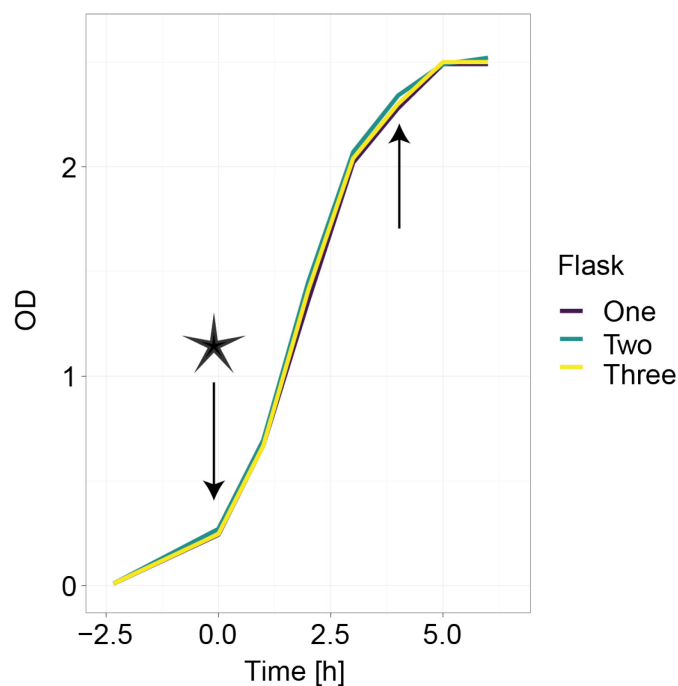

Supplementary Fig. 3.2 | Growth of *E. coli* TOP10 expressing the Bac7$_{1-23}$ error-prone library. OD is recorded over 6 h. Three 1 liter shake flasks containing 100 ml of LB-medium each are inoculated with 500 million cells of *E. coli* TOP10 carrying the peptide-encoded DNA library at -2.5 h (time reported relative to the time of induction). Peptides are expressed after 4 generations (0.0 h; OD~0.2) by adding L-arabinose (0.3% final; asterisk). Cell samples for NGS are isolated from each replicate at the time of induction and 4 h post-induction (arrows).

Supplementary Fig. 3.3 | Histogram of Bac7$_{1-23}$ variants. The amino acid edit distance from each of 601,551 peptides to the wild-type Bac7$_{1-23}$ is calculated. Most peptides (38%; 228,433) of the library are quadrupole mutants with four amino acid residue exchanges to Bac7$_{1-23}$. ~99% of all peptides have between zero (wild-type) and seven amino acid residue substitutions.

|	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
F	17	10	150	12	2244	4	13	2631	7	1193	2348	16	45	325	25	446	32	9879	107	462160	553	11	1404
Y	28	8	7	20	2126	23	15	2343	33	43	2533	26	59	3433	23	2797	38	808	85	7683	674	38	1139
W	36440	82	6	254	15	21851	2	3	75	223	0	81	701	405	87	110	1	20	32	57	1	23255	2
G	2685	6072	18	16478	333	5742	49	225	9157	76	95	9904	284	5444	33	4811	42	44	276	35	35	7677	82
A	48	102	103	99	19522	48	8865	9601	100	354	8530	62	18986	500	5364	279	4705	104	17782	67	3935	31	8326
M	22795	202	9051	163	33	591	6	2	299	24767	1	331	1165	3	1115	0	22	473	42	373	7	447	3
L	541	269	22497	480	25379	10464	12036	37093	469	483039	21669	479	37049	7037	28480	8564	23507	477532	28695	31087	23436	10922	13693
I	5	17905	526354	14955	787	77	679	1107	23836	3205	740	27430	115	241	87	396	1208	33771	1255	20679	839	75	540
V	51	112	11424	282	1761	86	429	1006	240	20213	657	291	1997	89	312	627	416	12242	1413	5434	205	131	445
Q	17773	42	70	406	483	19695	3841	573	55	9909	293	49	16357	1108	13617	455	7573	303	6530	14	437	22852	293
N	2	657	1152	697	293	15	5	288	1081	28	256	592	8	464	23	428	13	59	52	444	278	25	94
T	48	9154	7435	5059	9908	79	12312	9412	8060	346	9789	5350	7836	126	10129	257	11291	355	12038	382	11267	90	8134
S	450	6040	177	8186	52591	687	62161	45293	16245	874	61606	10038	49057	8505	23763	11639	21816	6265	48173	10963	18665	1097	51432
E	44	362	142	566	26	178	37	23	339	311	8	274	412	7	94	4	41	142	161	17	18	240	3
D	3	12	13	57	792	24	5	533	276	14	366	289	16	426	5	276	11	15	26	60	154	34	186
K	111	30573	13124	25613	37	335	46	19	23401	390	12	19527	162	35	178	10	100	633	140	27	5	390	1
H	583	25	4	12	23581	1009	288	28149	7	422	21201	7	723	43111	537	33857	1470	32	1164	194	22663	1009	13486
R	510764	521323	986	498365	8843	503238	1968	13159	476483	3589	6055	477714	10437	433842	3015	429659	4400	81	9181	50	6297	463946	4175
P	7011	36	51	52	422491	3509	465018	415337	14	10405	422902	11	405890	6528	463236	14198	472563	263	414841	241	452074	4081	433232
C	557	91	15	287	753	640	16	936	190	273	573	108	41	39045	23	40766	15	63	58	1758	177	706	361
*	1595	6879	298	20736	45	3703	504	58	7366	693	40	7055	1239	666	528	567	315	6180	1033	325	6	4663	24

Position

Supplementary Fig. 3.4 | Amino acid residue counts per position. For each position, the number of amino acid residues observed among all 601,551 peptides of the library was counted. Only two residues were not observed: tryptophan on position 11 and methionine on position 16. Framed amino acid residues correspond to the Bac7$_{1-23}$ wild-type amino acid residue at each position.
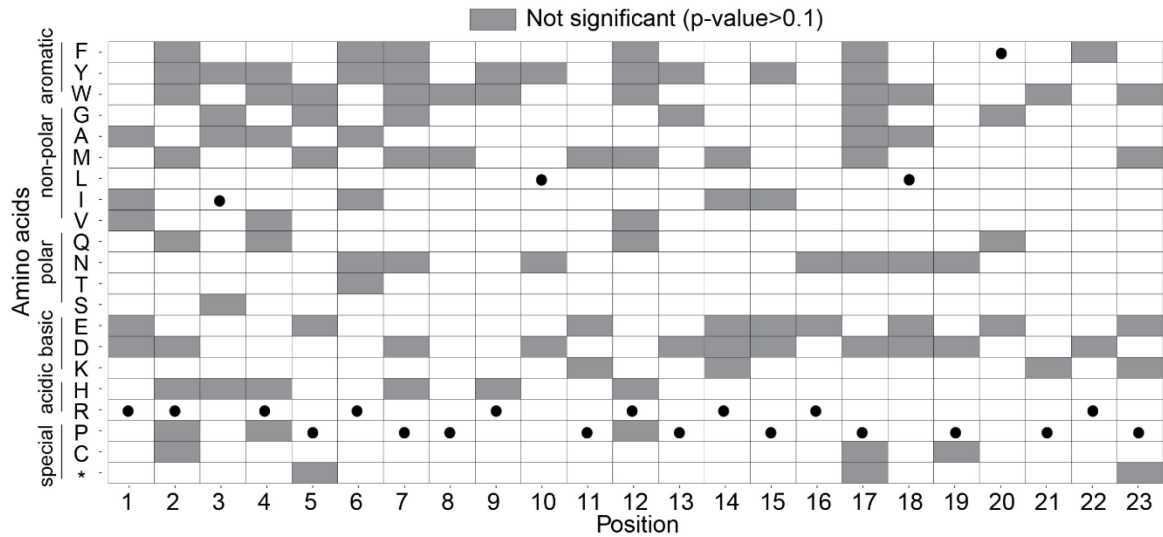
Supplementary Fig. 3.5 | Statistical significance of the observed growth inhibitory measurements for each amino acid substitution. Complementary to calculating the z-score, a two-sided-p-value is calculated to assess the statistical significance of the observed measurements. p-values were adjusted using the Benjamini-Hochberg procedure with a false discovery rate of $\alpha = 0.1$. Grey boxes represent the amino acid residue substitution of which the calculated effect on growth inhibition is statistically not significant. Black dots correspond to the Bac7 parental amino acid residue at each position.
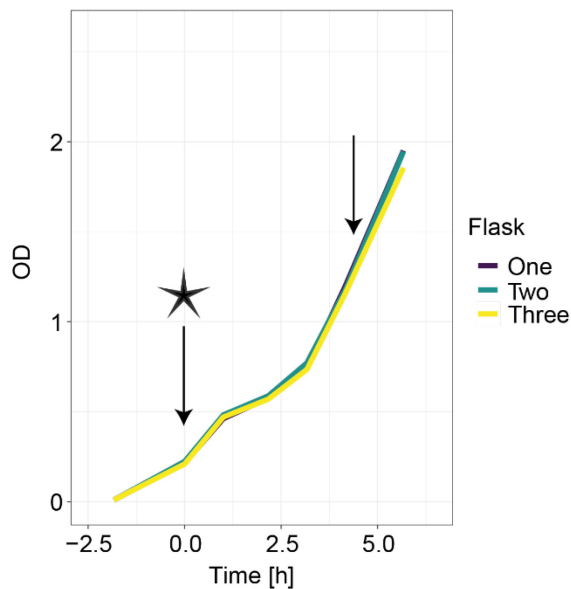
Supplementary Fig. 3.6 | Growth of *E. coli* TOP10 expressing the Bac7$_{1-23}$ focused library. OD is recorded over 5.5 h. Three 1 liter shake flasks containing 100 ml of LB-medium each are inoculated with 500 million cells of *E. coli* TOP10 carrying the peptide-encoded DNA library at -2 h (time reported relative to the time of induction). Peptides are expressed after 4 generations (0.0 h; OD~0.2) by adding L-arabinose (0.3% final) (asterisk). Cell samples for NGS are isolated from each replicate at the time of induction and 4.5 h post-induction (arrows).
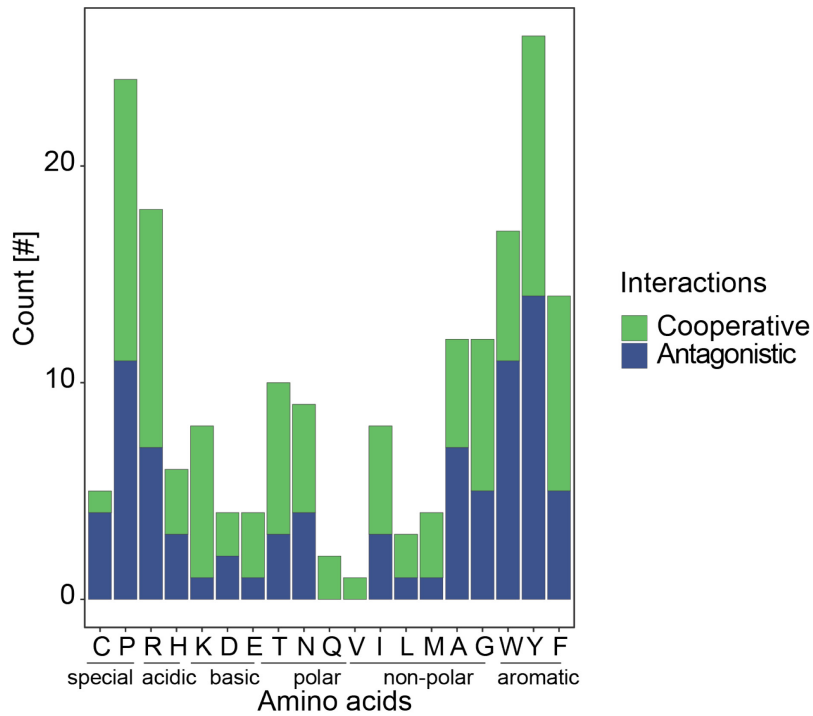
Supplementary Fig. 3.7 | Interactions observed per amino acid residue. For each amino acid residue, the occurrences of interactions at all positions were counted. No antagonistic was effect detected for glutamate and valine residues. No interaction was detected for serine. On average, most interactions were observed for proline, arginine, and aromatic amino acid residues and least interactions for non-polar amino acid residues.
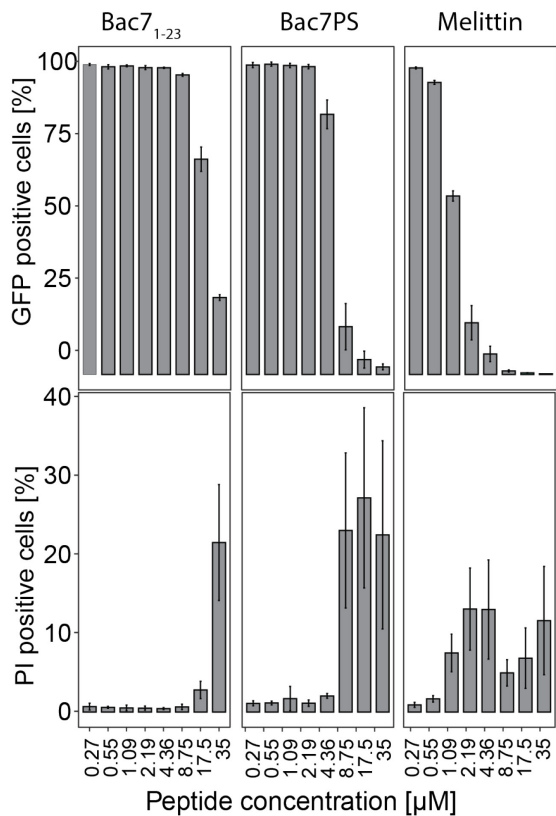
Supplementary Fig. 3.8 | Membrane damage assay. Membrane damage assays measuring PI uptake of the cells (% cells are given that gained PI fluorescence) and sfGFP loss (% of the cells are given that lost sfGFP fluorescence) when incubating *E. coli* TOP10 cells with various peptide concentrations in full MHB I.
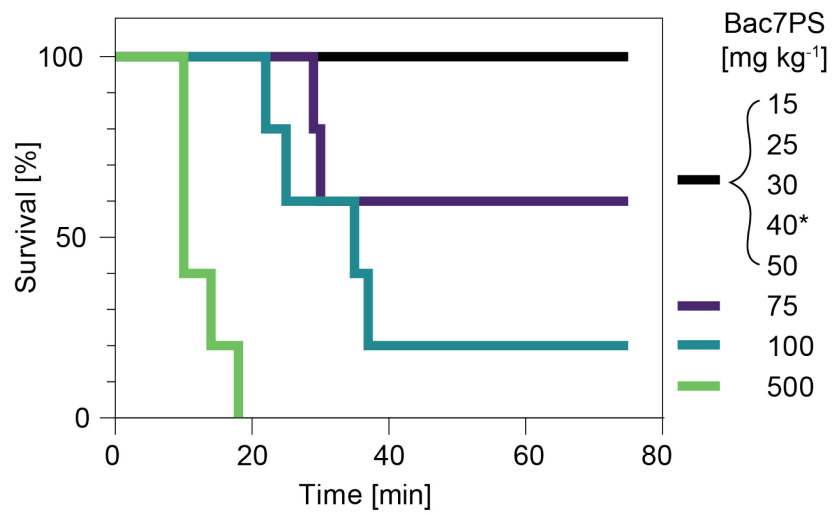
Supplementary Fig. 3.9 | In vivo toxicity. Bac7PS is applied intraperitoneally and survival of five CD-1 mice is measured for each concentration. A repeated administration of 40 mg kg^{-1} 4 h after the first dosing does not influence survival of mice (asterisk). Survival did not change over the course of 2 days.

CHAPTER 4 | Methods to characterize the mechanism of action of antimicrobials

Philipp Koch[1], Davide Visintainer*[1], Klaus Radlmair*[1], Sven Panke[1], Martin Held[1]

[1]Bioprocess Laboratory, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland.

*these authors contributed equally

Author contributions
P.K. and D.V. performed and analyzed experiments developing the bioreporter systems. P.K and K.R. performed and analyzed experiments developing the membrane damage assay. P.K. wrote the manuscript with input from all authors.

4.1 Introduction

One holdup in the antimicrobial discovery process is the determination of the MoA of antimicrobials[151]. Even for heavily used antibiotics such as chloramphenicol, discovered in 1948, the exact mechanism of action (MoA) is not fully understood[152]. Hence methods are being developed to determine the MoA of antibiotics[151,153–155]. Most of these methods are however not applicable in high-throughput and largely focus on small molecule antibiotics. In this chapter, I briefly summarize our experiments to develop methods for MoA determination that can be applied in high-throughput to determine if an AMP has an intracellular target or if it damages the bacterial membrane. These methods have been briefly introduced already in Chapter 2 and Chapter 3 and their development is reported here in full.

4.1.1 Bioreporter to determine the mechanism of action of antimicrobials

Bacteria have to withstand various environmental stresses such as lack of nutrients, changes in pH or temperature, oxidation, or the encounter of antimicrobials. To adapt to those changes, bacteria regulate pathways involved in transcription, translation, or the stability of proteins or RNA, using so-called stress regulatory networks[156]. The primary mechanism for this adaption occurs on the transcriptional level (RNA synthesis) by controlling the expression of regulons – groups of genes that are controlled by the same regulator[156]. RNA synthesis is directed by the DNA-dependent RNA polymerase (RNAp) in complex with one of a variety of sigma (σ) factors, which bind to RNAp and allow the enzyme to interact with specific classes of promoter sequences[156]. As the extent of the response by these transcriptional regulatory networks, in general, correlates with the level of stress[156], transcriptional bacterial reporter systems (or "bioreporters") have been widely used to study environmental changes. In this regard, a bioreporter is defined as a bacterial cell harboring a genetic circuit (often genetically engineered) that transduces the cellular sensory–regulatory response to environmental factors into a quantifiable signal originating from a reporter protein[157]. This signal can then be used as a measure for the degree of cellular stress in a dose-dependent manner.

Different bioreporter systems have been used to differentiate between MoAs of antibiotics[99,158,159]. A useful collection of suitable bioreporter is the 'Alon collection'[114]. This collection consists of a collection of low-copy plasmids harboring transcriptional fusions a variant of the green fluorescent protein (*gfp*) gene called to one of 1,900 different *E. coli* promoters (75% of all known *E. coli* promoters). This design allows the linking of activity

of the corresponding promoter and thus the cellular regulatory response to the expression of GFP. Thus, the use of such bioreporters allows for fast and parallel analysis and quantification of cellular stresses by, for example, plate readers, flow cytometers, or microscopes. To this end, the 'Alon collection' specifically has been applied for studying promoter activity at a genomic level[160], or as a toxicological pre-screen for pharmaceuticals, among others[100]. However, a thorough analysis of the usability of the 'Alon collection' to detect the MoA of antimicrobials is missing.

4.1.2 Methods to detect antimicrobial-induced membrane damage

All currently known antimicrobial peptides (AMPs) show some degree of membrane activity, the extent of which is dependent on their concentration and the medium or buffer in which they are investigated. The reason for this is that irrespective of their MoA most AMPs interact with membranes to either disrupt them (membrane permeabilizing AMPs) or to translocate across and reach intracellular or periplasmic targets (intracellularly active AMPs; Fig. 1.1). In general, these initial interactions are promoted by the cationic and amphiphilic (both hydrophilic and lipophilic) nature of AMPs, which allow them to accumulate at the anionic surfaces of the cell and the cytoplasmic membrane and eventually interact with fatty acyl chains of membrane phospholipids (reviewed in[30]).

There are several reasons why it is of interest to determine if a peptide exerts activity primarily via membrane rupture. Firstly, membrane damage is often associated with off-target toxicity towards human cells[53] which is assessed in preclinical studies and often expressed as the therapeutic index (TI)[146]. Here, the concentration of peptides required to lyse eukaryotic cells or other cytotoxic damage is divided by their minimal inhibitory concentration (MIC) against bacteria, and the resulting ratio is the TI. A high TI in lead candidates is considered important because it might eventually lead to an increased therapeutic window when developed into a drug. Membrane damaging peptides typically show low TI and are therefore frequently not considered in subsequent drug development steps[53,146]. Secondly, next to having a high TI, peptides that do not lyse membranes but target other components of the cells, so-called intracellularly active peptides, might reveal novel antimicrobial targets. To date however, only a few of those molecules have been discovered and characterized[46].

There are many methods to detect membrane damage inflicted by an antimicrobial, such as fluorescence - or atomic force spectroscopy[161,162], but most of them are very laborious

and therefore unsuited for high-throughput analyses and the characterization of larger peptide libraries. To quantify bacterial membrane damage, some more user-friendly methods were proposed that can also be used in high-throughput studies. In general, these rely on the uptake of propidium iodide (PI), a membrane-impermeable dye that binds to nucleic acids[113] or the release of intracellularly expressed proteins such as GFP or β-galactosidase (βGal)[163,164]. PI uptake is usually measured on the single-cell level via quantification of red PI-fluorescence in cells by flow cytometry. On the other hand, the GFP or βGal system is preferably used on the culture level, as the proteins need to be quantified in supernatants. Here the release of βGal into the medium is measured by quantifying the rate at which *o*-nitrophenyl-β-D-galactopyranoside (ONPG) is converted to the yellow-colored *o*-nitrophenol (ONP) using UV/VIS spectrometry. GFP in the supernatant can be quantified by using fluorescence spectroscopy.

4.1.3 Aim of the study

Firstly, we aimed to exploit the 'Alon collection' to find suitable bioreporters that would allow us to classify antibiotic compounds according to different MoAs. We hypothesized that a promoter, which is part of a stress response network activated upon, for example, membrane damage, DNA damage (SOS-response), or inhibition of protein translation, is specifically upregulated when the cells are treated with compounds causing those damages. Ideally, we could use these bioreporters to discriminate the MoA of added antibiotics and intracellularly expressed AMPs, which, to the best of our knowledge, has never been tried before. For this, we first conducted extensive literature research on stress promoters. Subsequently, we experimentally validated their usability as a bioreporter system. Out of 40 potential *E. coli* bioreporters from the 'Alon collection', two bioreporters including the promoter of cold shock protein A (*cspAp*) or recombinase A (*recAp*) worked best to distinguish between compounds interfering with protein biosynthesis or with the SOS-response, respectively.

Secondly, we aimed to compare different protocols for the detection of antimicrobial-induced membrane damage. Based on our results, we propose to use single-cell analysis based on a dual-staining method using recombinantly produced GFP and externally added PI as reporters as the most suitable method to differentiate between membrane permeabilizing and intracellularly active peptides.

4.2 Results

4.2.1 Experimental validation of 'Alon collection'-derived bioreporters

We first performed literature research on transcriptional stress responses and retrieved 40 different promoters, their corresponding σ factor, and native functions of the genes immediately following the promoter (Table 4.1). We focused on promoters that are available in the 'Alon collection' and that are part of four different stress response networks which seemed particularly appropriate for identifying promoters that respond to antibiotic stress: envelope/membrane, protein biosynthesis, general/metabolic, and SOS-response.

Table 4.1. Promoters investigated in this study. Promoter were empirically separated into four stress response networks.

Stress response network	Required σ factor	Promoter name	Function of expressed gene immediately following the promoter
Envelope/Membrane	σ54	*pspBp*	Important for membrane integrity
	Unknown	*minCp*	Inhibition of cell division
	σ24	*plsBp*	Phospholipid biosynthesis
	σ24	*degPp*	Periplasmic protease
	σ70	*ppiAp*	Periplasmic peptidyl-prolyl cis-trans isomerase
	σ54, 32	*ibpBp*	Hsp, molecular chaperone
	σS, 54, 70, 24	*rpoEp*	Response to periplasmic stress, σE factor of RNA polymerase
	σ24	*skPp*	Periplasmic chaperone
Protein biosynthesis	Unknown	*rpmpB*	50S ribosomal component
	σ70	*rpsBp*	30S ribosomal component
	σ70	*rpsJp*	30S ribosomal component
	σ70	*cspAp*	Cold shock response
	σ70	*ssrAp*	Ribosome rescue
	σ54, 32	*ibpBp*	Heat shock protein, molecular chaperone
	σS, 54, 70, 24	*rpoHp*	σ32, heat shock response
	σ70	*rpsUp*	30S ribosomal component
	unknown	*smpBp*	RNA binding
	σ70, 24	*tufAp*	Elongation factor
General/metabolic	σS, 70	*oxyRp*	Oxidative stress transcription factor
	σ70	*katGp*	Catalase, H_2O_2 degradation
	σ70	*trxCp*	Oxidative (general stress)ductase
	σS, 70	*bolAp*	Activator of morphogenic pathway (general stress)
	σ70	*inaAp*	pH stress response
	σ70	*sodAp*	Superoxide dismutase
	σ70	*marRp*	Multidrug efflux pump
	σ70	*soxSp*	Oxidative and antibiotic stress response
	σS	*tolCp*	Multidrug efflux factor
	σ70	*uspAp*	General stress protein
	σ70	*emrAp*	Multidrug efflux factor
	σ70, 32	*clpBp*	Molecular chaperone
	σ54, 32	*htpGp*	HSP90 family, heat shock
DNA/SOS-response	σS, 70	*gyrBp*	DNA gyrase
	unknown	*yoaAp*	Helicase, post-replication repair
	σ70	*polBp*	Polymerase mutation repair and bypass
	σ70	*recap*	Main SOS response effector
	σ70	*uvrBp*	Nucleotide excision repair subunit
	σ70	*dinGp*	DNA damage-inducible helicase
	σ32	*ybfEp*	DNA damage inducible protein
	σ70	*sulAp*	SOS dependent cell division inhibitor
	σ70	*lexAp*	SOS response regulator

Next, we selected antibiotics and stressors, available at our laboratory at the time, that target processes involved in at least one of the four stress response networks. Polymyxin B (PxB) damages membranes. Chloramphenicol, tetracycline, spectinomycin and streptomycin, as well as the proline-rich AMP bactenecin7 (Bac7) interfere with protein synthesis. Ampicillin (Amp) inhibits cell wall synthesis, eventually leading to cell lysis. We empirically assigned Amp to the general/metabolic stress response network as we speculated Amp might introduce stress on a global level. Hydrogen peroxide (H_2O_2) was assigned to the general/metabolic and to the DNA/SOS network as it causes oxidative stress but also induces damage to DNA. Finally, methylgloxal (MG) and mytomycin C (mC) as well as the AMP dermaseptin were assigned to the DNA/SOS stress response network. A detailed description of their MoA can be found in Table 4.2.

Finally, we performed three assays (Fig. 4.1a) using the antibiotics displayed in Table 4.2 to validate the expected function of bioreporters with the selected promoters. For this, we constructed a panel of bioreporter strains based on *E. coli* MG1655, and each of which carried a reporter plasmid with one of the promoters under study, as suggested by Zaslaver *et. al*[114].

a. Responsiveness: we tested if a given promoter (expected to be part of a specific stress network) indeed responded in an expected way to a specific antibiotic (expected to trigger the same specific stress network).
b. Selectivity: for all promoters that were found inducible in a), we tested if they also responded to antibiotics that elicit stress in at least one other stress response network.
c. Responsiveness upon intracellular synthesis: if a promoter was found responsive AND selective, we tested if it responded to the intracellular expression of a small set of AMPs.

Table 4.2. Antibiotics and stressors used in this study. Antibiotics and other stressors were selected to deploy MoAs, which interfere with at least one of the four considered stress response networks.

Stress response network	Antibiotic or stressor name	MIC* [µg ml^{-1}]	Description of MoA
Envelope/membrane	Polymyxin B (PxB)	2	Permeabilizes the plasma membrane through pore formation.
Protein biosynthesis	Chloramphenicol (Cmp)	5	Irreversibly binds to a receptor site on the 50S subunit of the bacterial ribosome, inhibiting peptidyl transferase.
	Tetracycline (Tet)	1.5	Prevents the attachment of aminoacyl-tRNA to the ribosomal acceptor (A) site.
	Spectinomycin (Spec)	20	Binds to the bacterial 30S ribosomal subunit and interferes with peptidyl tRNA translocation.
	Streptomycin (Strep)	10	Modifies association between 50S and 30S, yielding aberrant initiation complexes and blocking ribosome cycle.

	Bactenecin7 (Bac7)	Intracellular	Inhibits protein translation by blocking the ribosomal exit tunnel
General/metabolic	Ampicillin (Amp)	20	Inhibits transmembrane surface enzymes catalyzing transpeptidation of the peptidoglycans in the bacterial cell wall.
	Hydrogen Peroxide (H$_2$O$_2$)	2505	Leads to different types of oxidative structural damages.
DNA/SOS-response	Mitomycin C (mC)	10	Inhibits DNA replication and transcription.
	Methylglyoxal (MG)	50	Reacts with guanine residues of nucleic acids, interfering with DNA and RNA dynamics.
	Hydrogen Peroxide (H$_2$O$_2$)	2505	Leads to different types of oxidative structural damages.
	Dermaseptin	Intracellular	Binds to DNA. Damages membranes. Disturbs protein translation.

*minimal inhibitory concentration (MIC) values were taken from literature when measured against an *E.coli* strain

We first tested the responsiveness of all bioreporters at the MIC of a specific antibiotic and four and 16-fold above and below the MIC. We did not measure response for 31 out of 40 bioreporters. For example, the *plsBp* bioreporter, which is part of the membrane stress response network (Table 4.1), did not show any response upon addition of the membrane active PxB (Fig. 4.1b). On the other hand, nine bioreporters did show a response, as exemplified for the *cspAp* bioreporter response to Tet (Fig. 4.1b). In addition to the *cspA* bioreporter, the bioreporters containing *rpsUp, recAp, polBp, sulAp, lexAp, katGp, oxyRp,* and *clpBp* were responsive. For all nine promoters, we measured the strongest transcriptional response when testing at subinhibitory concentrations of antibiotics, a phenomenon known as hormesis[165], and hence further used only concentration at MIC or below. We next tested the selectivity of the response of the designated set of bioreporters. Of the nine responsive promoters, only *cspAp*, *rpsUp* (both protein translation), and *recap* (DNA/SOS-response), were selective towards antibiotics targeting a single stress-response network (summary of *cspAp* bioreporter and *recAp* bioreporter results in Fig. 4.1c). All other promoters were responsive to at least one antibiotic or stressor from another stress response network and were thus excluded.

Lastly, we tested the responsiveness upon an intracellular synthesis of AMPs. For this, we transformed bioreporters already harboring reporter plasmids with *cspap, recAp* and *rpsUp* with a second plasmid either carrying the gene for the 23 amino acid long derivative of Bac7 (Bac7$_{1-23}$), a peptide that inhibits protein translation[166], or for dermaseptin, a peptide that, among other MoAs, damages membranes and interferes with DNA synthesis[26]. Expression of Bac7$_{1-23}$ induced a slight *recAp* bioreporter response while strongly inducing the expression of GFP under the control of the *cspAp* and *rpsUp*, which is consistent with Bac7$_{1-23}$ inhibiting protein translation (Fig. 4.1d). Dermaseptin strongly induced the *recAp* and *rpsUp* and did not induce *cspAp*, which is in agreement with dermaseptin disturbing

DNA synthesis. We hence decided to use the *cspAp* and *recAp* bioreporter to detect MoA of both extracellularly as well as intracellularly applied antimicrobials involved in protein synthesis inhibition or DNA damage.
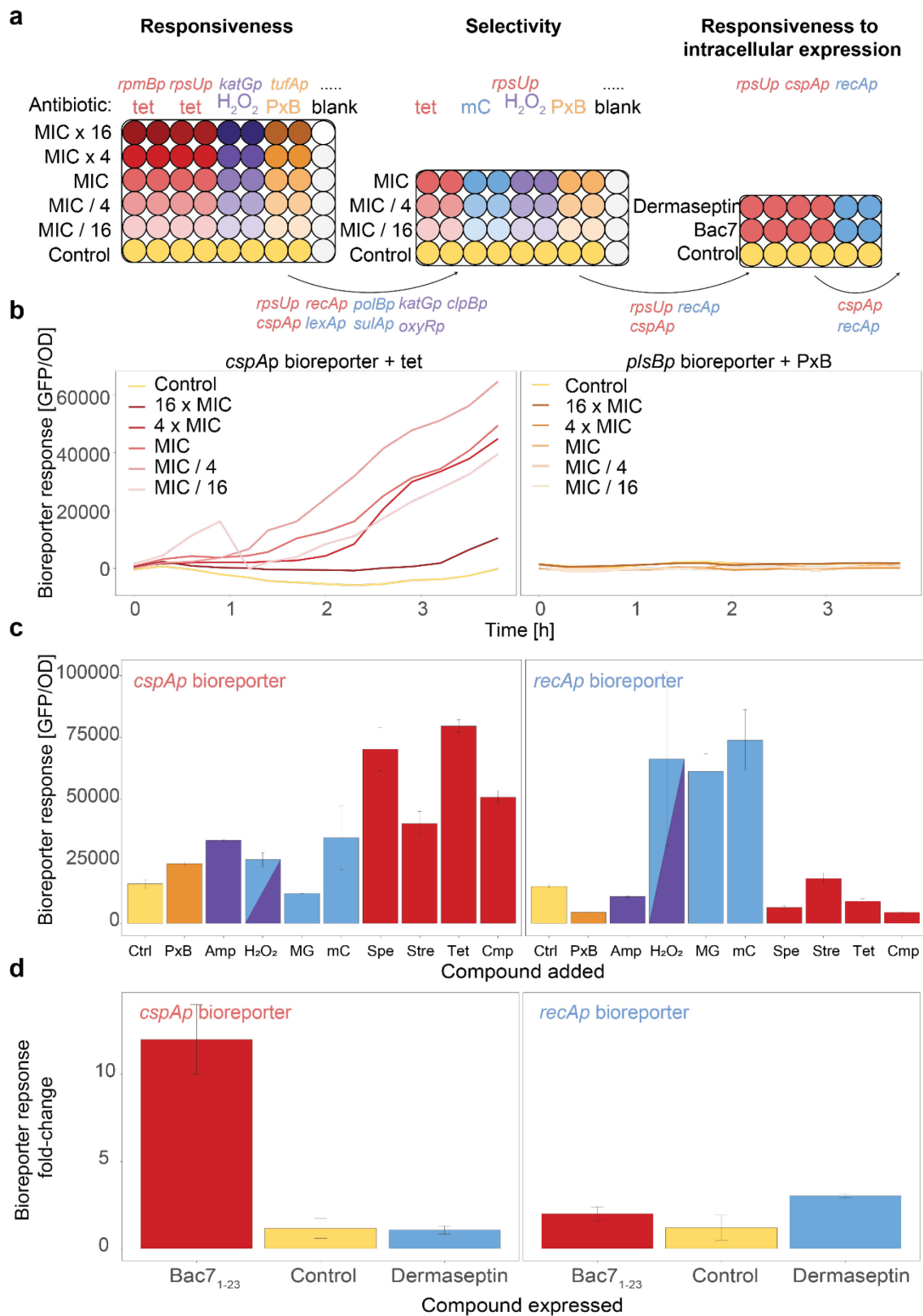

Fig. 4.1. | Characterization of bioreporters derived from the 'Alon collection'

a) Experimental workflow. Responsiveness: all bioreporters from Table 4.1 were tested using antibiotics or stressors from Table 4.2 that target processes involved in the same stress response network. Selectivity: remaining bioreporter were tested against antibiotics that target processes involved in the remaining different stress response networks. Responsiveness to the intracellular expression: responsiveness of bioreporters was measure when AMPs were expressed intracellularly. Colors correspond to different stress response networks (promoters and antibiotics): red (protein biosynthesis), purple (general/metabolic), brown (membrane), blue (DNA/SOS-response), and yellow (water control). **b)** Example of one extracellularly responsive (left) and one non-responsive bioreporter (right). **c)** Summary of *cspAp* and *recAp* bioreporter responses to externally added antibiotics at subinhibitory concentrations. Bars represent the bioreporter response shown as GFP fluorescence (λ_{Ex} 488 nm/λ_{Em} 530 nm) divided by OD (n=6; error bars = two standard deviations) **d)** Summary of *cspAp* and *recAp* bioreporter responses to intracellularly expressed antimicrobials (n=3; σ=1*sd). Control corresponds to the mean response of three intracellularly expressed AMPs that were inactive in previous experiments (results not shown).

4.2.2 Methods to detect membrane damage

To identify a fast and reproducible method for the detection of antimicrobial-induced membrane damage, we optimize an assay relying on the single-cell format. We evaluated these methods by using four antimicrobials: two known for lysis (colistin and melittin) and two know for a non-lytic MoA (Amp, Bac7$_{1-23}$) (see Table 4.3). With these reference compounds, uptake of externally added PI combined with a loss of recombinantly produced GFP from *E. coli* TOP10 quantified by flow cytometry on the other hand.

Table 4.3: Antimicrobial controls for assay validations.

Class	Name	MIC* (ug ml^{-1})	Description of MoA
Non-lytic	Amp	20	Inhibits transmembrane surface enzymes catalyzing transpeptidation of the peptidoglycans in the bacterial cell wall.
Non-lytic	Bac7$_{1-23}$	6	Inhibits protein translation by blocking the ribosomal exit tunnel
Lytic	Colistin	0.1	Disrupts outer and inner membrane
Lytic	Melittin	10-30	Forms pores into outer and inner membrane

*minimal inhibitory concentration (MIC) values were extracted from literature or were derived from our own experiments.

Uptake of PI is a straightforward method to measure bacterial membrane damage in single cells by flow cytometry[113]. However, when assessing PI uptake to detect AMP-induced membrane damage we encountered very poor reproducibility between experiments, especially if PI was incubated for more than 1 h with antimicrobial-treated cells, or when high antimicrobial concentrations had been used (see results of melittin below; Fig. 4.2b). This could for example be attributed to a tendency to achieve complete lysis of cells allowing for leakage of DNA from severely damaged cells. We thus decided to develop a dual-staining method by combining the analysis of PI entry and GFP loss from antimicrobial-treated cells. Such a method had already been proposed before to detect alcohol-induced membrane damage[136]. To do so, we cultured bacteria that constitutively express GFP and treated them with compounds using cell concentrations as proposed for MIC assays[112] (5*10^5 cells*ml^{-1}, equivalent to an optical density at 600 nm (OD) =0.001). To detect membrane damage, we added PI and measured the percentage of PI-positive

and GFP-negative cells among 10,000 cells by flow cytometry. While optimizing the protocol, we noticed an increased resilience of *E. coli* TOP10 towards lytic antibiotics if grown to stationary phase. We hence ensured incubation during the exponential growth phase by adapting the cultivation procedure. The final workflow to detect membrane damage using single-cell analysis can be seen in Fig. 4.2a. We were able to correctly assign $Bac7_{1-23}$, Amp, and colistin to their respective non-lytic and lytic MoAs when using them at their MIC (Fig. 4.2b). $Bac7_{1-23}$ did not lyse cells at its MIC (no GFP release or PI uptake) but showed some membrane activity at >10 times higher concentration, which is in accordance with literature data[117]. Amp did not show membrane damaging properties even if provided at >100 times the MIC (no GFP release or PI uptake). Colistin strongly damaged the membranes even below its MIC and damage increased in a concentration-dependent manner. At its MIC, around 75% of all cells suffered from severe membrane damage. Melittin also strongly damaged membranes at and below its MIC as visualized by almost 100% GFP release. However, PI uptake did not increase at concentrations around the MIC and even seemed to decrease at higher concentrations. We attribute this to severe and complete membrane damage caused by the pore formation of melittin, which would result in leakage of intracellular components such as DNA to which PI should bind. As a result, no cell-specific PI-positive signal could be obtained. This confirmed the need for a dual stain approach. In summary, for the four selected antimicrobial reference compounds, we were able to distinguish between unharmed cells, slight membrane damage, and cells that experienced complete membrane damage (Fig. 4.2c). In fact, GFP as a single-staining method sufficed for reporting the effect of most of the antibiotics used here. However, if membrane-damaging properties of intracellularly expressed AMPs or in strains not readily accessible to recombinant GFP expression are tested, PI remains the preferred alternative and should thus be included.

a

Incoulation | Growth 37 C | Dilute | Growth 37 C | Chill 4 C | Dilute | Incubate RT | Measure | Analyze

~14 h | ~3 h | +PI | MIC*n MIC MIC/n | ~0.5 h

Amp
Bac7₁₋₂₃
Col
Mel

OD=3 OD=.05 | OD=0.5 | OD= 0.001

b

Bac7₁₋₂₃ | Amp | Melittin | Colistin

GFP positive cells [%]

PI positive cells [%]

Compound concentration [µg ml⁻¹]

c

Propidium iodide (PI) → add AMP at MIC →
- No membrane damage
- Slight membrane damage
- Complete membrane damage
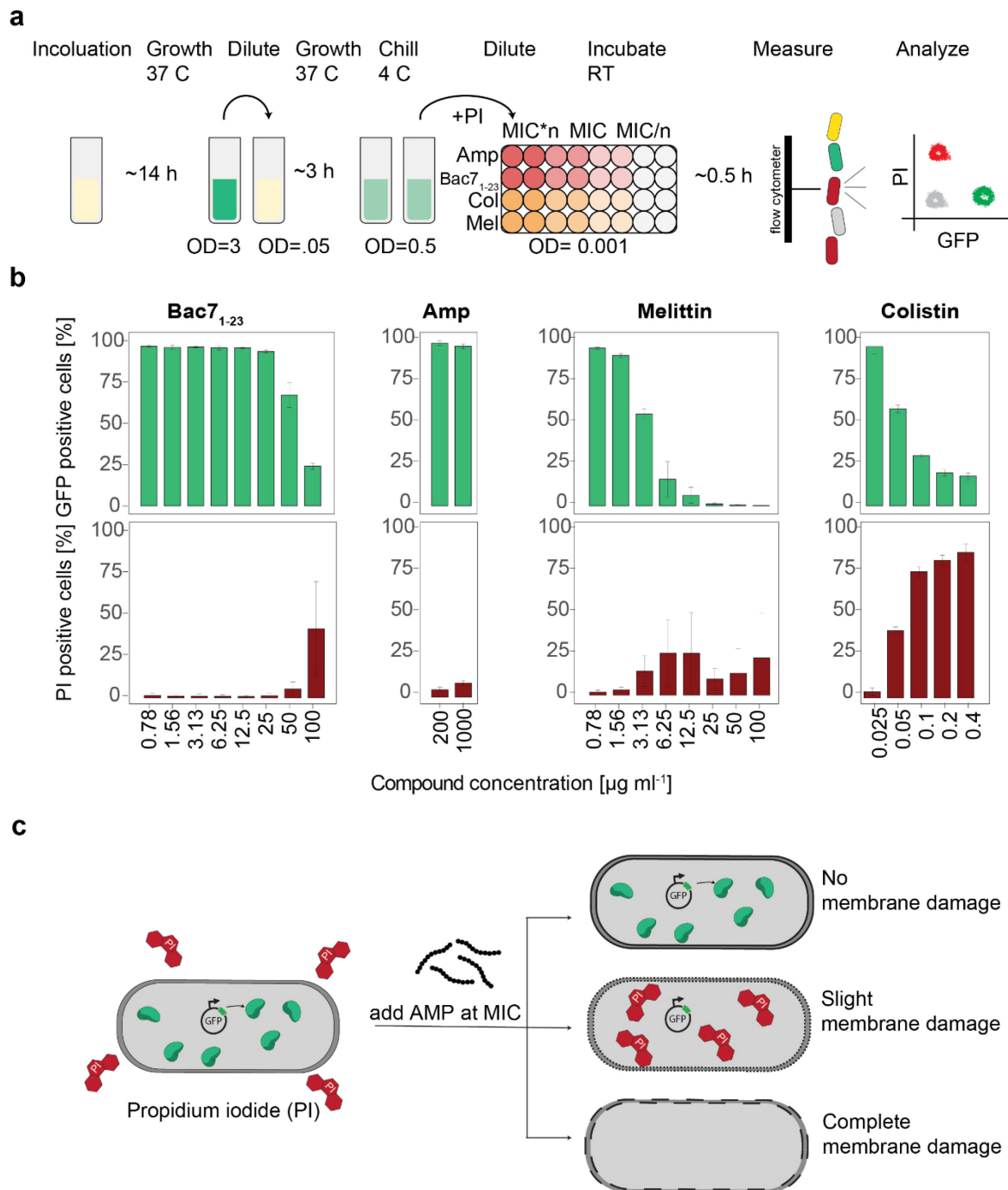
Fig. 4.2 | Flow cytometric single-cell analysis for detection of AMP-induced membrane damage. **a)** Workflow: *E.coli* TOP10 cells harboring a plasmid containing a constitutively expressing *gfp* gene are incubated with antimicrobials, using cell concentration of 5*10⁶ cell ml⁻¹, similarly to standardized MIC assays[112]. The fraction of PI- and GFP-positive cells are determined by flow cytometry and can be compared to a negative control represented by cells grown in the absence of antimicrobials. **b)** Percentage of GFP- and PI-positive cells when using the optimized protocol of a) using the antimicrobial reference compounds of Table 4.3. Bars represent the means of technical triplicates. Error bars = two standard deviations. c) Scheme to illustrate the hypothesis for explaining possible outcomes of the optimized dual staining protocol.

97

4.3 Discussion and conclusion

First, we evaluated 40 different bioreporters of the 'Alon collection' and their usability to discriminate between different MoA of antibiotics and AMPs. Some of the promoters covered in our list had already been applied successfully by others to study the MoA of different antibiotics but failed to deliver the predicted result in our study[159]. One reason might be that we applied more stringent criteria for bioreporter selectivity. Furthermore, some of the authors used bioreporters based on luminescence reporter proteins, or tryptophan auxotrophy-based instead of an antibiotic-based plasmid-selection system, which might interfere with the endogenous stress response to antibiotics[158]. Our results indicated that the *cspAp* and *recAp* bioreporter displayed the highest selectivity towards detecting antibiotic-induced stresses in processes involved in protein translation and DNA synthesis, respectively (Fig. 4.1c). Similar observations were made by others while characterizing the *cspAp* response in a βGal-based reporter system[99].

To the best of our knowledge, we showed for the first time that *cspAp* and *recAp* bioreporter can also be applied to detect the MoA of intracellularly produced AMPs. We showed that stress responses seemed to be similar between AMPs that inhibit the ribosome after their own ribosomal synthesis and externally added ribosomal inhibitors. The assay performance to detect stress response induced by intracellularly expressed AMPs could be optimized by limiting peptide expression levels to reach subinhibitory intracellular peptide concentrations. This could be achieved by including a titratable promoter, or by the use of weaker ribosomal binding sites in front of the gene that encodes the AMP. Lastly, as membrane damage is the most common MoA of AMPs, it would be desirable to design a bioreporter to also detect such damages. However, none of the membrane bioreporters tested by us was responsive to externally added peptides that damage membranes, and therefore we did not test them for membrane damage caused by intracellularly produced AMPs, either. We recently discovered work by others[158], which used a bioreporter harboring the promoter of *micF,* encoding an RNA that regulates the expression of an outer membrane porin[167], to detect membrane damage upon externally added antibiotics. This promoter is not present in the 'Alon collection', but could be synthetically produced, cloned in one of our constructs, and investigate for its suitability to detect membrane damage caused by intracellularly expressed AMPs.

Secondly, we investigated approaches to detect antibiotic- or AMP-induced membrane damage. We propose to incubate strains with antimicrobials at their MIC to differentiate between membrane-lytic compounds and compounds that mainly target intracellular

processes of the cell. For the economical usage of usually limiting research components such as synthetically produced AMPs, we optimized a dual-staining protocol, based on uptake of PI and release of GFP, using flow cytometry-based analysis of single-cells. Because of its simplicity, PI uptake studies are used widely to detect membrane damage of AMPs[91]. However, in the case of AMPs that strongly damage membranes such as meltittin, limited PI uptake might be incorrectly interpreted and would need further dose-dependent or kinetic investigation. However, in these cases, the complete loss of intracellularly produced GFP can be indicative of complete membrane damage and thus the correct assignment toward membrane damaging peptides can still be applied.

4.4 Methods

Chemicals and reagents

Unless otherwise stated, all antibiotics, chemicals, and other reagents were obtained from Sigma Aldrich (Buchs, CH). Kits for plasmid isolation were obtained from Zymo Research (Irvine, USA). Peptides in either purified (>90%) or crude format were obtained from Pepscan (Lelystad, NL) or Sigma Aldrich (Buchs).

Bacterial strains and cultivations

All experiments using 'Alon collection'-derived bioreporter were performed using *E. coli* MG1655 (K-12 F$^-$ λ$^-$ *ilvG*$^-$ *rfb-50 rph-1*) carrying the pUA66 plasmid containing a pSC101 origin of replication, a kanamycin resistance cassette, and the intergenic regions cloned by Zaslaver *et al* containing the respective promoters that control the expression of GFPmut2 reporter[114]. For all other experiments we used *E. coli* TOP10 (F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ$^-$ *rps*L(Str^R) *end*A1 *nup*G; Thermo Fisher Scientific, Waltham, USA). In this study, all cultivations were performed in 14 ml polypropylene tubes (Greiner, Kremsmuenster, AT), filled with 5 ml of lysogeny broth (LB) medium (Difco, Becton Dickinson, Franklin Lakes, USA). All samples were incubated at 37°C with agitation on a shaker (Kuhner, Birsfelden, CH) operated at 200 r.p.m. and 25 mm amplitude. All media were supplemented with the appropriate antibiotic for plasmid maintenance (50 µg ml^{-1} kanamycin; 100 µg ml^{-1} carbenicillin). For all cultivations on solid medium, 15 mg ml^{-1} agar (Difco) was added to the broth, and incubation was performed without shaking in an incubator (Kuhner) at 37°C. If not indicated differently, the OD of bacterial cultures was determined by measuring light scattering at 600 nm using a UV/VIS spectrophotometer (Eppendorf, Hamburg, DE).

Bioreporter-based detection of externally added antimicrobials

We picked strains directly from frozen stocks of the 'Alon collection', incubated them overnight in LB, and inoculated them into fresh medium to an OD of 0.05 on the next morning. We then grew the cells until OD~0.5 (~10^8 cells ml^{-1}) and distributed 100 µl of cells into 96-well microtiter plates (Greiner). In those plates, antimicrobial compounds diluted in 100 µl fresh rich medium were added according to a given experimental layout (see Fig. 4.1a). Commonly, antimicrobial were applied at their MIC as well as at four and 16-times higher and lower concentrations. The plates were incubated for 3-4 h at 37 C with light shaking in a plate reader (Tecan Infinite 200 PRO; Tecan, Männedorf, CH) and OD and GFP fluorescence ($\lambda_{Ex\ 488\ nm}/\lambda_{Em\ 530nm}$) were measured every 1000 seconds. The magnitude of the stress response was determined as the GFP/OD value compared to the

negative control (water added). Bioreporters were not considered for further characterization if their response in all wells (GFP/OD) did not increase by more than 2-fold at compared to the negative control (water added) after 3.8 - 4.3h.

Bioreporter-based detection of intracellularly expressed AMPs

To test the response to intracellularly expressed AMPs, a selection of the 'Alon collection'-derived bioreporter cells were co-transformed with the mid-copy pBAD[104] plasmid, carrying an ampicillin resistance and either the gene encoding for Bac7$_{1-23}$, dermaseptin, or inactive control peptides (tachyplesin, HNP-1, and pleurocidin) under the control of the inducible P$_{BAD}$ promoter. The double transformed strains were incubated identical to the experiment using externally added antimicrobials. 100 µl cells were distributed into 96-well microtiter plates (Greiner) containing 100 µl of rich medium containing 0.6% L-arabinose for induction of peptide expression. The plates were incubated and measured identical to the experiment measuring externally added antimicrobials.

Table 4.4 Genes used in this study that were expressed from the pBAD plasmid

Gene	Sequence (5'-3')
Bac7$_{1-23}$	ATGCGGAGAATAAGACCTCGGCCACCTAGACTGCCTAGACCGCGCCCGCGT CCATTACCATTCCCTCGGCCTTAA
Dermaseptin	ATGGCGTTGTGGAAGACGATGTTGAAAAAGTTGGGGACGATGGCGCTTCACG CAGGCAAAGCGGCGCTTGGCGCCGCCGCAGATACTATTTCCCAGGGCACGC AGTAGTAG
Tachyplesin	ATGAAATGGTGTTTTCGCGTGTGCTACCGTGGCATCTGTTATCGTCGCTGTCG TTGATAG
HNP-1	ATGGCGTGTTACTGCCGCATTCCTGCTTGTATCGCAGGCGAACGCCGTTACG GTACTTGTATTTATCAAGGTCGCCTTTGGGCCTTTTGCTGCTAG
Pleurocidine	ATGGGATGGGGTTCGTTTTTCAAAAAGGCCGCCCACGTAGGGAAACATGTCG GAAAGGCCGCGTTGACTCACTACCTTTGATGA

Membrane damage assay

For flow cytometric assays, *E. coli* TOP10 harboring a plasmid containing GFP under the control of the constitutive promoter J23118 (BioBrick:BBa_K823014), a kanamycin resistance cassette and an pBR322 origin of replication, were grown overnight in rich medium to stationary phase. The cultures were then used to inoculate (1:100) into fresh medium grown until exponential phase (~OD = 0.5), then put on ice for 20 min and finally diluted to $6*10^5$ cells ml^{-1} with medium containing a final concentration of 1 µg ml^{-1} propidium iodide (PI). Next, 50 µl of cell suspension were added to microtiter plate wells containing 50 µl medium with a 2-fold dilution series of the test compound. This was performed in triplicates. Cells and compounds were incubated at room temperature for 30 min without shaking. For each sample, the GFP and PI fluorescence of 10,000 cells was measured using a flow cytometer LSR Fortessa (BD Biosciences), equipped with the appropriate optical configuration (488 nm laser with 530/30 nm bandpass filter and 579 nm

laser with 610/20 nm bandpass filter). To determine the membrane damaging properties, we calculated the percentage of GFP retaining and PI acquired cells using the FlowJo V10 software (BD Biosciences).

CHAPTER 5 | Conclusion and outlook

Traditional drug discovery efforts, that is identifying new chemical or biological entities (NCEs, NBEs) that selectively bind to disease-related targets in the human body, have decreased in output over the last 80 years. These drug discovery models might eventually be replaced by emerging technologies such as cell, gene, or immunotherapy, or even tissue engineering or regenerative medicine. However, even if these technologies succeeded in curing all diseases arising within the human body, there would still be a vast repertoire of ever-changing and evolving pathogens, including both bacteria and viruses, that impose an imminent threat to human health. As the treatment of such infectious diseases mostly depends on the interaction with targets specific to pathogens, it will be pivotal to continue investing in methods for the discovery of NCEs and NBEs as anti-infective drugs, especially as pathogens quickly adapt to established therapies due to evolutionary processes. This is particularly crucial in light of the emerging AMR crisis, which is a global problem of great economic and environmental magnitude. Antimicrobial drug discovery is seminal to combat AMR and I thus aimed to make a contribution to the antimicrobial drug discovery field shown in this doctoral thesis. We developed methods for both the discovery (Chapter 2) and optimization (Chapter 3) of ribosomally produced AMPs. We used the pathogen model *E. coli* TOP10 for intracellular peptide expression and self-screening, and NGS as the assay readout.

For our discovery platform, we used a rationally designed AMP library containing naturally encoded peptides that we discovered on genetic databases by using sequence-similarity to already known AMPs. The peptide-encoding DNA sequences were synthesized on DNA chips (microarray supported oligonucleotide synthesis) and expressed intracellularly in *E. coli*. Other NGS-based self-screening assays use simple endpoint measurement to screen for antimicrobial peptides. Me^x however uses the kinetic recording in order to construct growth curves for each peptide-expressing strain. This enabled us not only to reliably differentiate between active and inactive AMPs, largely exceeding competing approaches in terms of hit-rate but also to determine the speed at which their specific MoA leads to cell death. We attributed fast growth inhibition to peptides that cause stronger membrane damage and delayed onset of a strong growth inhibitory effect to peptides that inhibit intracellular processes. We confirmed this hypothesis by further investigating the MoA of a selected subset of Me^x-active peptides. In my point of view, the ability to identify membrane-damaging peptides via Me^x could greatly accelerate the AMP discovery process: antimicrobials with strong membrane damage have been associated with increased toxicity towards eukaryotic cells, and could thus be eliminated early on in screening campaigns. Furthermore, the resolution of kinetic data can be enhanced by for example running the experiments with increased numbers of replicates, time points, and

higher sequencing depth. Such a rich data set could subsequently be exploited using machine learning algorithms that were trained on growth data of AMPs derived from the APD for which detailed MoA studies had been performed. Such experiments could further strengthen the applicability of a Me^x-based pre-selection of potential peptide leads, or eventually, be used for the tailored discovery of AMPs employing specific MoA.

In general, though, the discovery of AMPs could be massively expanded. The screening capacity of a Me^x is a function of the initial library size, its 'quality' (sequence distribution bias), the transformation efficiency of the pathogen model, the resolution of the kinetic aspect, and available sequencing depth. Within the scope of a Me^x-screening campaign (feasible, timely, economical), the limiting factor is the initial library synthesis platform. From the start of this project until now, the costs for pooled oligonucleotides synthesis have dropped dramatically and commercial platforms are now able to synthesize oligonucleotides with up to 300 bases each, compared to 170 bases about 4 years ago. Also, the 'quality' of these libraries improved, as indicated, for instance, by a minimized sequence distribution bias, which greatly reduces the need for oversampling in all subsequent steps to cover the entire library in the experiment. Thus, taking these improvements into account, current platforms can be expanded to screen libraries of more than 100'000 different peptides of a length of 90 amino acids without large increases in costs or changing any of our downstream analysis methods. I expect further technological advances in the field of nucleotide synthesis in the upcoming years and thus even further increases in Me^x-throughput and scope.

Next to discovering novel active AMPs, we used our bacterial self-screening platform to functionally analyze and optimize the ribosomal inhibitor $Bac7_{1-23}$. Contrary to the discovery pipeline, which was comprised of diverse and naturally-derived peptides, we first randomly and then semi-rationally altered the sequence of $Bac7_{1-23}$ using molecular biological tools. This resulted in hundreds of thousands of synthetic peptide variants, which we screened for antimicrobial activity using Me^x. This allowed us to generate an antimicrobial fitness landscape across $Bac7_{1-23}$, which we exploited to design Bac7PS, a variant with four amino acid residues differences to $Bac7_{1-23}$. Compared to our discovery approach, which was limited by microarray-based synthesis to generate the library, the throughput of our optimization approach was limited by standard molecular biology operations (cloning, transformation) and DNA sequencing. To evaluate the $Bac7_{1-23}$ library, we generated roughly 1 billion sequence reads, which is already about one-tenth of what is currently possible in a single run (using an Illumina NovaSeq 6000 S4 flow cell) and does not leave much room to further increase throughput at this point. With our setup, we

provided one of the largest (the functional consequence of all amino acid residue exchanges) and most thorough (the functional consequence of multiple amino acid residue combinations) experimental analyses of the sequence space of any AMP to date. The optimized peptide variant Bac7PS strongly improved in antimicrobial properties over the already highly active $Bac7_{1-23}$ and is currently being evaluated in initial pre-clinical assessments. Traditional AMP design approaches cannot experimentally assess peptide libraries well when substituting multiple amino acid residues in a peptide due to the exponential increase of possible peptide variants. These approaches thus often suffer from low amounts of data points. I believe that our concept will not only inspire optimizations of other intracellularly active AMPs. In fact, it will open up new possibilities in studying structure-activity-relationship in AMPs and eventually guide advanced machine learning methods for the prediction of tailor-made AMPs.

Our AMP discovery and optimization platform can be easily expanded. For example, screening campaigns can be extended to other transformable pathogens, including resistant or persistent bacteria. Overall, I argue that a transition from standard low-throughput functional assays such as MIC-determinations to modalities that use NGS enables a step-change in terms of throughput, especially when exploiting recent developments in chemical DNA synthesis technology and molecular engineering. Together, these two developments have the power to entirely transform the process of developing therapeutic entities from gene-based templates. Of note, this does not only apply to antimicrobials, but can be extended to any human disease model, which can be reproduced in a microbe, or to the challenging field of protein-protein interactions.

I believe that using our high-throughput methods can greatly accelerate the research and development of antimicrobials, but it still solves only a small piece of the antibiotic-AMR puzzle. For one, it is clear that AMP starting points need further development into lead candidates by chemical optimizations to ensure in vivo efficacy and safety. More strikingly though, there many additional revolutionary concepts that look at the function or the screening of novel antimicrobials from a more integrative perspective, including pathogen metabolism, in vivo behavior, or synergism with human host factors. Especially now, as the world suffers from a pandemic caused by the severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2), I noticed that a single and isolated solution to a global problem can only bring local and temporary relief. Only with a combined effort that includes innovative drug discovery-based approaches in the private or governmental sector as well as efforts in financing, regulation, and the education of the general population, we will be able to emerge successfully from upcoming global challenges, including AMR.

Bibliography

1.   Armstrong, G. L., Conn, L. A. & Pinner, R. W. Trends in infectious disease mortality in the United States during the 20th century. *J. Am. Med. Assoc.* **281**, 61–66 (1999).

2.   Ehrlich, P. & Bertheim, A. Über das salzsaure 3.3′-diamino-4.4′-dioxy-arsenobenzol und seine nächsten verwandten. *Berichte der Dtsch. Chem. Gesellschaft* **45**, 756–766 (1912).

3.   Strebhardt, K. & Ullrich, A. Paul Ehrlich's magic bullet. **8**, 473–480 (2008).

4.   Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. 1929. *Br. J. Exp. Pathol* **10**, 226– 236 (1929).

5.   Chain, E. *et al.* Penicillin as a chemotherapeutic agent. *Lancet* **236**, 226–228 (1940).

6.   Schatz, A., Bugle, E. & Waksman, S. A. Streptomycin, a substance exhibiting antibiotic activity against Gram-positive and Gram-negative bacteria. *Exp. Biol. Med.* **55**, 66–69 (1944).

7.   Lewis, K. The science of antibiotic discovery. *Cell* **181**, 29–45 (2020).

8.   Silver, L. L. Challenges of antibacterial discovery. *Clin. Microbiol. Rev.* **24**, 71–109 (2011).

9.   Richter, M. F. *et al.* Predictive compound accumulation rules yield a broad-spectrum antibiotic. *Nature* **545**, 299–304 (2017).

10.   Gottig, S. *et al.* Detection of pan drug-resistant Acinetobacter baumannii in germany. *J. Antimicrob. Chemother.* **69**, 2578–2579 (2014).

11.   Organization, W. H. Antimicrobial resistance: global report on surveillance. (2014).

12.   Crofts, T. S., Gasparrini, A. J. & Dantas, G. Next-generation approaches to understand and combat the antibiotic resistome. *Nat. Rev. Microbiol.* **15**, 422–434 (2017).

13.   Rice, L. B. Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESKAPE. *J. Infect. Dis.* **197**, 1079–1081 (2008).

14.   Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).

15.   Newman, D. J. & Cragg, G. M. Natural products as sources of New drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).

16.   Arnold, F. H. Innovation by evolution: bringing new chemistry to life (nobel lecture). *Angew. Chemie Int. Ed.* **58**, 14420–14426 (2019).

17.   Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459 (2015).

18.   Hashimi, S. M. Albicidin, a potent DNA gyrase inhibitor with clinical potential. *J. Antibiot. (Tokyo).* **72**, 785–792 (2019).

19.   Luther, A. *et al.* Chimeric peptidomimetic antibiotics against Gram-negative bacteria. *Nature* **576**, 452–458 (2019).

20. Fleming, A. On a remarkable bacteriolytic element found. *Proc. R. Soc. B* **93**, 306–317 (1922).

21. Chipman, D. M. & Sharon, N. Mechanism of lysozyme action. *Science.* **165**, 454–465 (1969).

22. Hirsch, J. G. Phagocytin: a bactericidal substance from polymorphonuclear leucocytes. *J. Exp. Med.* **103**, 589–611 (1956).

23. Zasloff, M. Magainins, a class of antimicrobial peptides from Xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc. Natl. Acad. Sci.* **84**, 5449–5453 (1987).

24. Steiner, H., Hultmark, D., Engström, A., Bennich, H. & Boman, H. G. Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature* **292**, 246–8 (1981).

25. Rogers, L. A. & Whittier, E. O. Limiting factors in the lactic fermentation. *J. Bacteriol.* **16**, 211–229 (1928).

26. Brogden, K. A. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.* **3**, 238–50 (2005).

27. Dubos, R. J. Studies on a bactericidal agent extracted from a soil Bacillus. *J. Exp. Med.* **70**, 1–10 (1939).

28. Stansly, P. G., Shepherd, R. G. & White, H. J. Polymyxin: a new chemotherapeutic agent. *Bull. Johns Hopkins Hosp.* **81**, 43–54 (1947).

29. Geraci, J. E., Heilman, F. R., Nichols, D. R., Wellman, E. W. & Ross, G. T. Some laboratory and clinical experiences with a new antibiotic, vancomycin. *Antibiot. Annu.* 90–106.

30. Hancock, R. E. W. & Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat. Biotechnol.* **24**, 1551–1557 (2006).

31. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).

32. Simons, A., Alhanout, K. & Duval, R. E. Bacteriocins, antimicrobial peptides from bacterial Origin: overview of their biology and their impact against multidrug-resistant bacteria. *Microorganisms* **8**, 639 (2020).

33. Wang, G., Li, X. & Wang, Z. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.* **37**, D933–D937 (2009).

34. Wang, G., Li, X. & Wang, Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **44**, D1087–D1093 (2016).

35. Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **415**, 389–395 (2002).

36. Haney, E. F., Straus, S. K. & Hancock, R. E. W. Reassessing the host defense peptide landscape. *Front. Chem.* **7**, 1–22 (2019).

37. Wang, G. Post-translational modifications of natural antimicrobial peptides and strategies for peptide engineering. *Curr Biotechnol* **1**, 72–79 (2014).

38. Bahar, A. A. & Ren, D. Antimicrobial peptides. *Pharmaceuticals* **6**, 1543–1575 (2013).

39. Graf, M. *et al.* Proline-rich antimicrobial peptides targeting protein synthesis. *Nat. Prod. Rep.* **34**, 702–711 (2017).

40. Mura, M. *et al.* The effect of amidation on the behaviour of antimicrobial peptides. *Eur. Biophys. J.* **45**, 195–207 (2016).

41. Yeaman, M. R. & Yount, N. Y. Mechanisms of antimicrobial peptide action and resistance. *Pharmacological Reviews* vol. 55 27–55 (2003).

42. Guha, S., Ghimire, J., Wu, E. & Wimley, W. C. Mechanistic landscape of membrane-permeabilizing peptides. *Chem. Rev.* **119**, 6040–6085 (2019).

43. Madani, F., Lindberg, S., Langel, Ü., Futaki, S. & Gräslund, A. Mechanisms of cellular uptake of cell-penetrating peptides. *J. Biophys.* **2011**, 1–10 (2011).

44. Jones, A. T. Macropinocytosis: searching for an endocytic identity and role in the uptake of cell penetrating peptides. *J. Cell. Mol. Med.* **11**, 670–684 (2007).

45. Henriques, S. T., Melo, M. N. & Castanho, M. A. R. B. Cell-penetrating peptides and antimicrobial peptides: how different are they? *Biochem. J.* **399**, 1–7 (2006).

46. Le, C.-F., Fang, C.-M. & Sekaran, S. D. Intracellular targeting mechanisms by antimicrobial peptides. *Antimicrob. Agents Chemother.* **61**, 1–16 (2017).

47. Mookherjee, N., Anderson, M. A., Haagsman, H. P. & Davidson, D. J. Antimicrobial host defence peptides: functions and clinical potential. *Nat. Rev. Drug Discov.* **19**, 311–332 (2020).

48. Peschel, A. & Sahl, H. G. The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nat. Rev. Microbiol.* **4**, 529–536 (2006).

49. Lazzaro, B. P., Zasloff, M. & Rolff, J. Antimicrobial peptides: application informed by evolution. *Science (80-.).* **368**, 5480 (2020).

50. Rahnamaeian, M. *et al.* Insect antimicrobial peptides show potentiating functional interactions against Gram-negative bacteria. *Proc. R. Soc. B Biol. Sci.* **282**, 20150293 (2015).

51. Ge, Y. *et al.* In vitro antibacterial properties of pexiganan, an analog of magainin. *Antimicrob. Agents Chemother.* **43**, 782–788 (1999).

52. Srinivas, N. *et al.* Peptidomimetic antibiotics target outer-membrane biogenesis in Pseudomonas aeruginosa. *Science.* **327**, 1010–1013 (2010).

53. Magana, M. *et al.* The value of antimicrobial peptides in the age of resistance. *Lancet Infect. Dis.* **20**, e216–e230 (2020).

54. Lax, R., Ph, D., Development, B., America, N. & Group, P. The future of peptide development in the pharmaceutical industry. 10–15 (2010).

55. Pachón-Ibáñez, M. E., Smani, Y., Pachón, J. & Sánchez-Céspedes, J. Perspectives for clinical use of engineered human host defense antimicrobial peptides. *FEMS Microbiol. Rev.* **41**, 323–342 (2017).

56. Kosikowska, P. & Lesner, A. Antimicrobial peptides (AMPs) as drug candidates: a patent review (2003-2015). *Expert Opin. Ther. Pat.* **26**, 689–702 (2016).

57. Theuretzbacher, U., Outterson, K., Engel, A. & Karlén, A. The global preclinical antibacterial pipeline. *Nat. Rev. Microbiol.* **18**, 275–285 (2020).

58. Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **11**, 191–200 (2012).

59.   Henninot, A., Collins, J. C. & Nuss, J. M. The current state of peptide drug discovery: back to the future? *J. Med. Chem.* **61**, 1382–1414 (2018).

60.   Lau, J. L. & Dunn, M. K. Therapeutic peptides: historical perspectives, current development trends, and future directions. *Bioorg. Med. Chem.* **26**, 2700–2707 (2018).

61.   Vecchio, I., Tornali, C., Bragazzi, N. L. & Martini, M. The discovery of insulin: an important milestone in the history of medicine. *Front. Endocrinol. (Lausanne).* **9**, (2018).

62.   Merrifield, R. B. Solid phase peptide synthesis. I. the synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).

63.   Bozovičar, K. & Bratkovič, T. Evolving a Peptide: library platforms and diversification strategies. *Int. J. Mol. Sci.* **21**, 215 (2019).

64.   Furka, Á., Sebestyen, F., Asgedom, M. & Dibo, G. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res.* **37**, 487–493 (2009).

65.   Houghten, R. A. General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proc. Natl. Acad. Sci.* **82**, 5131–5135 (1985).

66.   Goodnow, R. A., Dumelin, C. E. & Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* **16**, 131–147 (2017).

67.   Smith, G. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science.* **228**, 1315–1317 (1985).

68.   Zhu, W., Williams, R. S. & Kodadek, T. A CDC6 protein-binding peptide selected using a bacterial two-hybrid-like system is a cell cycle inhibitor. *J. Biol. Chem.* **275**, 32098–32105 (2000).

69.   Lennard, K. R. & Tavassoli, A. Peptides come round: using SICLOPPS libraries for early stage drug discovery. *Chem. - A Eur. J.* **20**, 10608–10614 (2014).

70.   Davis, A. M., Plowright, A. T. & Valeur, E. Directing evolution: The next revolution in drug discovery? *Nat. Rev. Drug Discov.* **16**, 681–698 (2017).

71.   Tawfik, D. S. & Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **16**, 652–656 (1998).

72.   Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).

73.   Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* **14**, 111–129 (2015).

74.   Atanasov, A. G., Zotchev, S. B., Dirsch, V. M. & Supuran, C. T. Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* **53**, 1689–1699 (2021).

75.   MacIlwain, C. When rhetoric hits reality in debate on bioprospecting. *Nature* **392**, 535–540 (1998).

76.   Hultmark, D., Steiner, H., Rasmuson, T. & Boman, H. G. Insect immunity. Purification and properties of three inducible bactericidal proteins from hemolymph of immunized pupae of Hyalophora cecropia. *Eur. J. Biochem.* **106**, 7–16 (1980).

77. NCBI Resource Coordinators. Database resources of the national center for niotechnology information. *Nucleic Acids Res.* **45**, D12–D17 (2017).

78. Hao, Y. *et al.* SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. *Brief. Bioinform.* **19**, 636–643 (2017).

79. Van Heel, A. J. *et al.* BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **46**, W278–W281 (2018).

80. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).

81. Montalbán-López, M. *et al.* New developments in RiPP discovery, enzymology and engineering. *Nat. Prod. Rep.* **38**, 130–239 (2021).

82. Koo, H. B. & Seo, J. Antimicrobial peptides under clinical investigation. *Pept. Sci.* **111**, (2019).

83. Fjell, C. D., Hiss, J. A., Hancock, R. E. W. & Schneider, G. Designing antimicrobial peptides: form follows function. *Nat. Rev. Drug Discov.* **11**, 37–51 (2012).

84. Kang, X. *et al.* DRAMP 2.0, an updated data repository of antimicrobial peptides. *Sci. data* **6**, 148 (2019).

85. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).

86. Fjell, C. D. *et al.* Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J. Med. Chem.* **52**, 2006–2015 (2009).

87. Veltri, D., Kamath, U. & Shehu, A. Improving recognition of antimicrobial peptides and target selectivity through machine learning and genetic programming. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **14**, 300–313 (2017).

88. Schneider, P. *et al.* Hybrid network model for "deep learning" of chemical data: application to antimicrobial peptides. *Mol. Inform.* **36**, (2017).

89. Sainath Rao, S., Mohan, K. V. K. & Atreya, C. D. A peptide derived from phage display library exhibits antibacterial activity against E. coli and Pseudomonas aeruginosa. *PLoS One* **8**, 1–11 (2013).

90. Raventos, D. *et al.* Improving on natures defenses: optimization & high throughput screening of antimicrobial peptides. *Comb. Chem. High Throughput Screen.* **8**, 219–233 (2005).

91. Tucker, A. T. *et al.* Discovery of next-generation antimicrobials through bacterial self-screening of surface-displayed peptide libraries. *Cell* **172**, 1–11 (2018).

92. Schmitt, S. *et al.* Analysis of modular bioengineered antimicrobial lanthipeptides at nanoliter scale. *Nat. Chem. Biol.* **15**, (2019).

93. Guralp, S. A., Murgha, Y. E., Rouillard, J.-M. & Gulari, E. From design to screening: a new antimicrobial peptide discovery pipeline. *PLoS One* **8**, 1–7 (2013).

94. Jensen, T. D. *et al.* The PepSeq pipeline. in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* 139–144 (ACM, 2018).

95. Russell, A. H. & Truman, A. W. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Comput. Struct. Biotechnol. J.* **18**, 1838–1851 (2020).

96.   Pearson, W. R. An introduction to sequence similarity ('homology') searching. *Curr. Protoc. Bioinforma.* **42**, 1–8 (2013).

97.   Watt, P. M. Screening for peptide drugs from the natural repertoire of biodiverse protein folds. *Nat. Biotechnol.* **24**, 177–183 (2006).

98.   Torres, M. D. T., Sothiselvam, S., Lu, T. K. & de la Fuente-Nunez, C. Peptide design principles for antimicrobial applications. *J. Mol. Biol.* **431**, 3547–3567 (2019).

99.   Bianchi, A. A. & Baneyx, F. Stress responses as a tool to detect and characterize the mode of action of antibacterial agents. *Appl. Environ. Microbiol.* **65**, 5023–5027 (1999).

100.   Elad, T., Seo, H. Bin, Belkin, S. & Gu, M. B. High-throughput prescreening of pharmaceuticals using a genome-wide bacterial bioreporter array. *Biosens. Bioelectron.* **68**, 699–704 (2015).

101.   Yang, S. T., Shin, S. Y., Hahm, K. S. & Kim, J. Il. Different modes in antibiotic action of tritrpticin analogs, cathelicidin-derived Trp-rich and Pro/Arg-rich peptides. *Biochim. Biophys. Acta - Biomembr.* **1758**, 1580–1586 (2006).

102.   Altschul SF, Gish W, Miller W, Myers EW, L. D. Basic local alignment search tool. *J Mol Bio* **215**, 403–10 (1990).

103.   Sonnhammer, E. L. L. & Hollich, V. Scoredist: A simple and robust protein sequence distance estimator. *BMC Bioinformatics* **6**, 1–8 (2005).

104.   Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–30 (1995).

105.   Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

106.   Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-Tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).

107.   Butterworth, P. J. Lehninger: principles of biochemistry (4th edn) D. L. Nelson and M. C. Cox, W. H. Freeman & Co., New York, 1119 pp (plus 17 pp glossary), ISBN 0-7167-4339-6 (2004). *Cell Biochem. Funct.* **23**, 293–294 (2005).

108.   Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

109.   Anthis, N. J. & Clore, G. M. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci.* **22**, 851–858 (2013).

110.   Pelillo, C. *et al.* Cellular internalization and cytotoxicity of the antimicrobial proline-rich peptide Bac7(1-35) in monocytes/macrophages, and its activity against phagocytosed Salmonella typhimurium. *Protein Pept. Lett.* **21**, 382–390 (2014).

111.   Holfeld, L., Knappe, D. & Hoffmann, R. Proline-rich antimicrobial peptides show a long-lasting post-antibiotic effect on Enterobacteriaceae and Pseudomonas aeruginosa. *J. Antimicrob. Chemother.* **73**, 933–941 (2018).

112.   Wiegand, I., Hilpert, K. & Hancock, R. E. W. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **3**, 163–175 (2008).

113.   Mortimer, F. C., Mason, D. J. & Gant, V. A. Flow cytometric monitoring of antibiotic-

induced injury in Escherichia coli using cell-impermeant fluorescent probes. *Antimicrob. Agents Chemother.* **44**, 676–681 (2000).

114. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat. Methods* **3**, 623–628 (2006).

115. Frank, R. W., Gennarot, R., Schneiderq, K., Przybylskin, M. & Romeo, D. Amino acid sequences of two proline-rich bactenecins. *J. Biol. Chem.* **265**, 18974 (1990).

116. Mattiuzzo, M. *et al.* Role of the Escherichia coli SbmA in the antimicrobial activity of proline-rich peptides. *Mol. Microbiol.* **66**, 151–163 (2007).

117. Podda, E. *et al.* Dual mode of action of Bac7, a proline-rich antibacterial peptide. *Biochim. Biophys. Acta - Gen. Subj.* **1760**, 1732–1740 (2006).

118. Benincasa, M. *et al.* Antimicrobial activity of Bac7 fragments against drug-resistant clinical isolates. *Peptides* **25**, 2055–2061 (2004).

119. Guida, F. *et al.* Effect of size and N-terminal residue characteristics on bacterial cell penetration and antibacterial activity of the proline-rich peptide Bac7. *J. Med. Chem.* **58**, 1195–1204 (2015).

120. Mardirossian, M. *et al.* Peptide inhibitors of bacterial protein synthesis with broad spectrum and SbmA-independent bactericidal activity against clinical pathogens. *J. Med. Chem.* **63**, 9590–9602 (2020).

121. Lai, P., Tresnak, D. T. & Hackel, B. J. Identification and elucidation of proline-rich antimicrobial peptides with enhanced potency and delivery. *Biotechnol. Bioeng.* **116**, 2439–2450 (2019).

122. Fowler, D. M. & Fields, S. Deep mutational scanning : a new style of protein science. **11**, 801–807 (2014).

123. Gray, V. E., Hause, R. J. & Fowler, D. M. Analysis of large-scale mutagenesis data to assess the impact of single amino acid substitutions. *Genetics* **207**, 53–61 (2017).

124. Fowler, D. M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **7**, 741–746 (2010).

125. Baliga, C. *et al.* Charting the sequence-activity landscape of peptide inhibitors of translation termination. *Proc. Natl. Acad. Sci.* **118** (2021).

126. Rubin, A. F. *et al.* A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* **18**, 1–15 (2017).

127. Gagnon, M. G. *et al.* Structures of proline-rich peptides bound to the ribosome reveal a common mechanism of protein synthesis inhibition. *Nucleic Acids Res.* **44**, 2439–2450 (2016).

128. Seefeldt, A. C. *et al.* Structure of the mammalian antimicrobial peptide Bac7(1–16) bound within the exit tunnel of a bacterial ribosome. *Nucleic Acids Res.* **44**, 2429–2438 (2016).

129. Tang, L. *et al.* Construction of 'small-intelligent' focused mutagenesis libraries using well-designed combinatorial degenerate primers. *Biotechniques* **52**, 149–158 (2012).

130. Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the numbers problem in directed evolution. *ChemBioChem* **9**, 1797–1804 (2008).

131. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach

for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

132. Reetz, M. T. The importance of additive and non-additive mutational effects in protein engineering. *Angew. Chemie - Int. Ed.* **52**, 2658–2666 (2013).

133. Llinares-Lopez,Felipe; Borgwardt, K. Machine learning for biomarker discovery: significant pattern mining. in *Analyzing Network Data in Biology and Medicine* 313–368 (Cambridge University Press, 2019).

134. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, (2006).

135. Runti, G. *et al.* Functional characterization of SbmA, a bacterial inner membrane transporter required for importing the antimicrobial peptide Bac7(1-35). *J. Bacteriol.* **195**, 5343–5351 (2013).

136. Lehtinen, J., Nuutila, J. & Lilius, E. M. Green fluorescent protein-propidium iodide (GFP-PI) based assay for flow cytometric measurement of bacterial viability. *Cytom. Part A* **60**, 165–172 (2004).

137. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods* **65**, 55–63 (1983).

138. Sarkisyan, K. S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397–401 (2016).

139. Melnikov, A., Rogov, P., Wang, L., Gnirke, A. & Mikkelsen, T. S. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res.* **42**, 1–8 (2014).

140. Linsky, T. W. *et al.* De novo design of potent and resilient hACE2 decoys to neutralize SARS-CoV-2. *Science.* **1214**, (2020).

141. Sadler, K., Eom, K. D., Yang, J. L., Dimitrova, Y. & Tam, J. P. Translocating proline-rich peptides from the antimicrobial peptide bactenecin 7. *Biochemistry* **41**, 14150–14157 (2002).

142. Kannan, K., Vázquez-Laslop, N. & Mankin, A. S. Selective protein synthesis by ribosomes with a drug-obstructed exit tunnel. *Cell* **151**, 508–520 (2012).

143. Mardirossian, M. *et al.* The dolphin proline-rich antimicrobial peptide Tur1A inhibits protein synthesis by targeting the bacterial ribosome. *Cell Chem. Biol.* **25**, 530–539 (2018).

144. Milletti, F. Cell-penetrating peptides: Classes, origin, and current landscape. *Drug Discov. Today* **17**, 850–860 (2012).

145. Zhang, W., Sato, T. & Smith, S. O. NMR spectroscopy of basic/aromatic amino acid clusters in membrane proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 183–199 (2006).

146. Bacalum, M. & Radu, M. Cationic antimicrobial peptides cytotoxicity on mammalian cells: An analysis using therapeutic index integrative concept. *Int. J. Pept. Res. Ther.* **21**, 47–55 (2015).

147. Benincasa, M. *et al.* The proline-rich peptide Bac7(1-35) reduces mortality from Salmonella typhimurium in a mouse model of infection. *BMC Microbiol.* **10**, (2010).

148. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).

149. Zawada, J. F. Preparation and Testing of E. coli S30 In Vitro Transcription Translation Extracts. in *Ribosome Display and Related Technologies* 31–41 (2012).

150. Mikami, S., Kobayashi, T. & Imataka, H. Cell-Free Protein Synthesis Systems with Extracts from Cultured Human Cells. in 43–52 (2010).

151. Zampieri, M. *et al.* High-throughput metabolomic analysis predicts mode of action of uncharacterized antimicrobial compounds. *Sci. Transl. Med.* **10**, 3973 (2018).

152. Choi, J. *et al.* Dynamics of the context-specific translation arrest by chloramphenicol and linezolid. *Nat. Chem. Biol.* **16**, 310–317 (2020).

153. O'Rourke, A. *et al.* Mechanism-of-action classification of antibiotics by global transcriptome profiling. *Antimicrob. Agents Chemother.* **64**, 1–15 (2020).

154. Jana, B., Baker, K. R. & Guardabassi, L. Macromolecule biosynthesis assay and fluorescence spectroscopy methods to explore antimicrobial peptide mode(s) of action. in 181–190 (2017).

155. Burdine, L. & Kodadek, T. Target identification in chemical genetics. *Chem. Biol.* **11**, 593–597 (2004).

156. Rosenberg, E., DeLong, E. F., Thompson, F., Lory, S. & Stackebrandt, E. The prokaryotes: prokaryotic physiology and biochemistry. *Prokaryotes Prokaryotic Physiol. Biochem.* 1–662 (2013).

157. van der Meer, J. R. & Belkin, S. Where microbiology meets microengineering: design and applications of reporter bacteria. *Nat. Rev. Microbiol.* **8**, 511–522 (2010).

158. Melamed, S. *et al.* A bacterial reporter panel for the detection and classification of antibiotic substances. *Microb. Biotechnol.* **5**, 536–548 (2012).

159. Osterman, I. A. *et al.* Sorting out antibiotics' mechanisms of action: a double fluorescent protein reporter for high-throughput screening of ribosome and DNA biosynthesis inhibitors. *Antimicrob. Agents Chemother.* **60**, 7481–7489 (2016).

160. Zaslaver, A. *et al.* Invariant Distribution of Promoter Activities in Escherichia coli. *PLoS Comput. Biol.* **5**, (2009).

161. Fojan, P. & Gurevich, L. Atomic force microscopy study of the interactions of indolicidin with model membranes and DNA. in *Antimicrobial Peptides* 201–215 (2017).

162. Arias, M. & Vogel, H. J. Fluorescence and absorbance spectroscopy methods to study membrane perturbations by antimicrobial host defense peptides. in *Antimicrobial Peptides* 141–157 (2017).

163. Imura, Y., Nishida, M., Ogawa, Y., Takakura, Y. & Matsuzaki, K. Action mechanism of tachyplesin I and effects of PEGylation. *Biochim. Biophys. Acta - Biomembr.* **1768**, 1160–1169 (2007).

164. Eustice, D. C., Feldman, P. A., Colberg-Poley, A. M., Buckery, R. M. & Neubauer, R. H. A sensitive method for the detection of beta-galactosidase in transfected mammalian cells. *Biotechniques* **11**, 739–40, 742–3 (1991).

165. Davies, J., Spiegelman, G. B. & Yim, G. The world of subinhibitory antibiotic concentrations. *Curr. Opin. Microbiol.* **9**, 445–453 (2006).

166. Graf, M. *et al.* Proline-rich antimicrobial peptides targeting protein synthesis. *Nat. Prod. Rep.* **34**, 702–711 (2017).

167. Delihas, N. & Forst, S. MicF : an antisense RNA gene involved in response of Escherichia coli to global stress factors 1 1Edited by D. Draper. *J. Mol. Biol.* **313**, 1–12 (2001).

Acknowledgements

First of all, I would like to thank my direct supervisor Dr. Martin Held for giving me the opportunity to start working on this truly exciting project. I was challenged and enthusiastic for the entire time. Thank you very much for spending countless hours developing novel ideas and coming up with creative solutions nobody else would have thought of, whenever I ran into problems. Of course, I also want to thank the supervisor of my thesis Prof. Dr. Sven Panke for letting me be part of the bioprocess laboratory, for guidance throughout the entire time, and for creating such a productive, safe and fun work environment. I want to thank both of you for letting me work independently whenever needed and for allowing me to participate in many conferences to interact with other scientists.

Special thanks go to the members of my doctoral and thesis committee Prof. Dr. Jörn Piel and Prof. Dr. Urs Jenal for giving me valuable feedback along the way, taking the time to examine my thesis, and conducting my examination. Thanks to Prof. Dr. Andreas Hierlemann for taking over the chair of the examination.

Many thanks go to *antimicrobiologist* Dr. Steven Schmitt. It is fair to say that without him, my time here would have been a lot harder. Thanks for your selflessness in sharing ideas and knowledge as well as giving your valuable time teaching me and supporting my development as a scientist.

Furthermore, I would like to say thanks to Dr. Mathias Cardner and Dr. Anja Gumpinger for your great input and for helping me figure out the best computational way to analyze my data. Getting help from you greatly accelerated my I work.

I owe a great debt of gratitude to all the students who helped me with all my projects. Thanks to Alexander Heynisch, Davide Visintainer, Klaus Radlmair, Jan Dahinden, Ana Dujmovic, Dominik Schiemann and Max Singer.

Additionally, I would like to thank all members of the BPL for creating a great and productive lab-atmosphere and very fun group retreats. I especially want to say thanks to Dr. Philipp Rottmann, Dr. Markus Jeschek, Simon Hoellerer, Dr. Irene Wüthrich, and Dr. Tania Roberts for all your support and feedback, both professionally and personally.

Thanks for all your help to the people from the Single Cell Facility, the Lab Automation Facility, and Genomics Facility.

I would love to say thanks a lot to my loving mother and father for their constant support!

Last but not least, so many thanks to Verena. You carried much of the weight on your shoulders and supported me the entire time. This would not have been possible without you by my side!

Curriculum vitae

Personal information

Full name: Philipp Koch

Nationality: German

Date of birth: March 26th 1990

Education

Since 10/2016	**Doctoral degree:** ETH Zurich, Switzerland Supervisor: Prof. Dr. Sven Panke, Dr. Martin Held Project: *Massively-parallelized discovery and optimization of anitimicrobials*
08/2016 08/2014	**Master's degree (M.Sc./Dipl.-Ing. (TUM)):** Pharmaceutical Bioprocess Engineering, Technical University of Munich, Germany **Master's thesis:** California Institute of Technology, USA Supervisors: Prof. Dr. Frances Arnold, Prof. Dr. Andrew Buller, Prof. Dr. Volker Sieber Project: *Engineering a coupled enzyme system for a novel route to non-canonical amino acids* **Term paper:** *Microencapsulation of sensitive biomaterial in aerogels*
09/2014 09/2010	**Bachelor's degree (B.Sc.):** Bioprocess Engineering, Technical University of Munich, Germany **Bachelors' thesis:** Rechts der Isar Hospital, Technical University of Munich, Germany Supervisors: Prof. Dr. Ulrike Protzer, Dr. Knud Esser, Dr. Jochen Wettengel Project: *Entwicklung und Validierung eines Hochdurchsatz-Testverfahrens zur Identifizierung von Hepatitis B Virus-Aufnahmeinhibitoren*
04/2013 09/2012	**Exchange semester:** National Taiwan University, Taiwan Chemical engineering program

Conference participations

2020	4th AMR Conference 2020 – Novel Antimicrobials and AMR Diagnostics Basel, Switzerland (Digital conference) **Poster:** *Massively parallelized growth assays (Me^x) for discovery of antimicrobial peptides*
2020	Digitalization and Infectious Diseases: Improving patient outcome in the age of big data Basel, Switzerland
2019	Gordon Research Conference on Antimicrobial Peptides Lucca, Italy **Poster:** *Identification of multifunctional AMPs*
2018	EMBO Workshop: Bacterial persistence and antimicrobial therapy Ascona, Switzerland **Poster:** *Detection of stress induced by antibiotics*
2018	Microbial Stress: from Systems to Molecules and Back Kinsale, Ireland **Poster:** *Visualization of antibiotic stress on the single cell level*
2017	EU project meeting SYNPEPTIDE (Synthetic Biology for the production of functional peptides) Berlin, Germany **Talk:** *Activity profiling of the antimicrobial peptidome*
2017	International Conference on Molecular Systems Engineering Basel, Switzerland
2016	EU project meeting SYNPEPTIDE (Synthetic Biology for the production of functional peptides) Regensburg, Germany

Publications and patents

- **P Koch**, S Schmitt, M Cardner, N Beerenwinkel, S Panke, M Held (2021). Discovery of antimicrobials by Me^x: massively parallelized growth assays. *In review*

- S Schmitt, **P Koch**, M Held, S Panke (2021). Method for identifying target-binding peptides. *Patent application.* EP21182971.8

- C E Boville, S Brinkmann-Chen, A R Buller, **P Koch**, C. Prier, D. K. Romney, R. Scheele. (2020) Methods and enzyme catalyst for the synthesis of non-canonical amino acids. EP3621461A1, US10612056B2

- C E Boville ,R A Scheele*, **P Koch**, S Brinkmann-Chen, A R Buller, F H Arnold (2018) Engineered biosynthesis of beta-alkyl tryptophan analogs. *Angew Chem*, http://doi.org/10.1002/anie.201807998

Extracurricular activity

2018-2021   Main organizer of ETH VMB weekly sport class Tabata

2018-2020   Main organizer of D-BSSE Meets Industry

2016-2018   Member of D-BSSE Meets Industry