DESIGNER PEPTIDES WITH ANTIMICROBIAL AND ANTICANCER ACTIVITY

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZÜRICH

(Dr. sc. ETH Zürich)

Presented by

Yen-Chu Lin

Master of Sciences, National Taiwan University

born on February 4th, 1985

Citizen of Taiwan

Accepted on the recommendation of

Prof. Dr. Gisbert Schneider, examiner
Prof. Dr. Cornelia Halin, co-examiner
Prof. Dr. Paul Wrede, co-examiner

2014
Parts of this material have been published in *ChemBioChem*
(doi: 10.1002/cbic.201402231R1)
Table of Contents

Acknowledgement ..................................................................................................................III
Summary .......................................................................................................................................V
Zusammenfassung ..........................................................................................................................VIII
Abbreviation ...............................................................................................................................XI

I. Introduction
1. Peptides as therapeutic agents .............................................................................................1
2. Antimicrobial peptides ...........................................................................................................2
   2.1 Unmet medical need for antimicrobial agents .................................................................2
   2.2 Characterization of AMPs ............................................................................................3
   2.3 Mechanisms of AMPs ..................................................................................................5
   2.4 Current clinical development of AMPs and outlook ......................................................9
3. Anticancer peptides ..............................................................................................................9
   3.1 Cancer is the leading cause of death worldwide ...........................................................9
   3.2 Anticancer peptides as innovative therapeutics ............................................................11
   3.3 Current clinical development of ACPs and outlook .....................................................14
4. Computer-assisted drug design ...........................................................................................15
   4.1 Computer-assisted small molecules design ................................................................15
      4.1.1 Receptor-based de novo drug design .....................................................................15
      4.1.2 Ligand-based de novo drug design .......................................................................17
   4.2 Computer-assisted peptide design ................................................................................19
   4.3 Computer-assisted AMP design ...................................................................................21
      4.3.1 Sequence-based approaches ...............................................................................21
      4.3.2 Molecular dynamics simulation ...........................................................................23
      4.3.3 AMP QSAR models ............................................................................................24
         4.3.3.1 Descriptors .....................................................................................................24
         4.3.3.2 Linear modeling methods ..............................................................................25
         4.3.3.3 Machine-learning methods ..........................................................................26
   4.4 Computer-assisted ACP design .....................................................................................30

II. De novo design of antimicrobial peptides
1. Aim ........................................................................................................................................31
2. Hypotheses ..........................................................................................................................31
### Table of Contents

3. **Materials and Methods**..................................................................................................................32  

4. **Results and Discussion**......................................................................................................................41  
   4.1 Choosing AMP templates and scrambled the templates.................................................................41  
   4.2 Characterization of Protonectin and Decoralin and their derivatives..............................................45  
   4.3 Secondary structure of peptides and peptide-membrane interaction..............................................49  
   4.4 Self-organizing maps for the *de novo* design of AMPs.................................................................50  
   4.5 Characterization of SOMs designed peptides.................................................................................53  

5. **Conclusions...........................................................................................................................................59

### III. Sequence-based multi-objective optimization of anticancer peptides

1. **Aim......................................................................................................................................................61  

2. **Hypotheses...........................................................................................................................................61  

3. **Materials and Methods........................................................................................................................65  

4. **Results and Discussion........................................................................................................................73  
   4.1 Computing Support Vector Machine QSAR models........................................................................73  
   4.2 Cytotoxicity of Decoralin and TAT(48-60)......................................................................................74  
   4.3 Identification of critical positions on Decoralin and TAT(48-60) for their activities.........................75  
   4.4 Optimization of Decoralin with better cytotoxicity by improving its cell-penetrating activity...........78  
   4.5 Designed ACPs permeabilize cells to lesser extents and induce cell apoptosis...................................83  
   4.6 AC2 displays better synergism with doxorubicin and epothilone C against A549 but not MCF7 cells...85  
   4.7 Optimization of Decoralin with better selectivity by building into THP features............................88  
   4.8 Optimization of TAT(48-60) with anticancer activity.....................................................................92  

5. **Conclusions...........................................................................................................................................97

### IV. Conclusions and Outlook....................................................................................................................99

### V. Appendix.............................................................................................................................................103

### VI. Cumulative References.....................................................................................................................109
Acknowledgement

Compared with previous studying career, three-year Ph.D. study at ETH is not the longest period to me. However, I believe it is the most venturesome experience in my study career. In the first year, the culture shock, entrance test as well as visa problems were with me to explore the research path. In the second year, I finished the AMPs project but at the same time started the ACPs project with my bold hypothesis and complicated biologic experiments. The final year, undoubtedly, I spent almost every weekend in the lab solving troublesome experimental problems and fought till last minutes before submitting this thesis. Fortunately, I had many friends accompanying me through the journey and provided me with lots of help and encouragement so that I could finish the task on time.

First of all, I would like to thank Prof. Gisbert Schneider for giving me opportunity to work in his lab. Gisbert always gives me unlimited freedom to do whatever I want to do. His enthusiasm in science along with brilliant creativity and productivity in drug discovery are esteemed by me forever. I would also acknowledge the help from Dr. Petra Schneider in ITC experiments. I appreciate Dr. Jan Hiss for leading me in research and in life when I was just here. I would like to express my sincere gratitude to Dr. Tiago Rodrigues for providing me with lots of help in research and for generously sharing with me your vision in science and the life experience. I learned a lot from you.

In addition, I would like to thank every colleague in the lab for providing me so much help and joy during the time. I appreciate Max Pillong and Anna Perna for accompanying me the entire doctoral study life. We started the study at the same time and we had much unforgettable wonderful time in the lab. I would like to thank Katharina Stutz for translating the summary for me and supporting me in peptide synthesis from time to time. Particularly I would like to thank Sarah Haller for helping me order the reagents and organizing the peptide
synthesis. I extremely appreciate Daniel Reker, Jens Kunze, Dr. Nickolay Todoroff, Dr. Johannes Kirchmair and former colleagues, Dr. Michael Reutlinger, Dr. Matthias Rupp, Dr. Tim Geppert and Dr. Felix Reisen for helping me computational problems and providing me useful opinions in research and life.

Moreover, I would like to express my greatest gratitude to my every collaborator. I thank Prof. Silja Wessler and Peter Thelesklaf for conducting bacterial assays for me. I acknowledge Prof. Cornelia Halin and her team, Erica Russo, Simone Häner, Dr. Alvaro Sanchez and Dr. Maria Iolyeva for helping me setting up cell-based experiments and giving me numerous constructive opinions. I express thanks to Prof. Petra Dittrich and her team for teaching me preparing LUVs. I appreciate Dr. Fabian Köhler for helping me perform the AFM measurement. Finally, I would like to acknowledge the kindness of Prof. Dario Neri and Prof. Michael Detmar to provide the cell lines for my project.

Particularly, I would like to appreciate my lovely family in Taiwan. Because of my parents’ constant encouragement and support, I can realize my dream to study abroad and persist to the end. I appreciate my sister for taking care of our parents for me when I am abroad.

Last but not least, I would like to dedicate my work to my lovely wife, Yi Fan. I understand how difficult the decision she made to marry me right before moving to this beautiful but unfamiliar country. We sustained a difficult time here at the beginning and finally managed to survive in Switzerland. She always confronts my genuine feeling and suffers from my bad temper resulting from frustration in research, however in the end she always can comfort me in nice ways. Yi Fan is not only my supporter in life but also a good collaborator in the lab. Without her endurance and fortitude, I would not have chance to accomplish the mission.

Zürich, June, 2014

Yen-Chu Lin

Modlab, ETH
Summary

The spread of resistant bacteria, which results in untreatable infections, is a public health problem worldwide. According to the WHO 2014 report on global surveillance of antimicrobial resistance, antibiotic resistance is no longer a prediction for the future; antibiotic resistance is currently occurring across the world. New antibiotic discovery is urgently required, and antimicrobial peptides (AMPs) represent a promising class of antibiotics.

Although progress has been made in clinical cancer therapies, including chemotherapy, radiation, surgery and target therapies, cancer remains one of the leading causes of death globally. Moreover, the high recurrence risk of this disease and the emergence of resistance to current therapies in progressing tumors are major problems. Anticancer peptides (ACPs) have been suggested as a new strategy to combat cancer because ACPs possess selective and rapid membrane-compromising cytotoxic mechanisms of action against cancer cells. This ACP activity has the potential to be used alone or as an adjuvant with current chemotherapies.

In this thesis, we present computational approaches for designing new AMPs and ACPs. First, we present a computational de novo design approach that results in AMPs possessing low nanomolar direct membrane-lytic activity, antimicrobial activity, and a high therapeutic index. This method is based on a self-organizing map (SOM) for prioritizing amino acid sequences bearing the desired function. For SOM training, we analyzed the activities of two template AMPs, Protonectin and Decoralin from wasp venom, and demonstrated that residue scrambling yields a diverse set of potent derivatives. The results highlight both the N- and C-terminal portions of these AMPs as critical for membrane activity and suggest a relation between their inducible helical content and membrane binding. Isothermal titration
calorimetry reveals an entropy-driven mechanism of peptide-membrane interaction.

For the ACP study, we adopted computer-assisted methods to rationally design peptide sequences with improved activity and selectivity compared with known ACPs. We combined features of cell-penetrating peptides (CPPs) and tumor-homing peptides (THPs) in an ACP template. Therefore, first, we performed glycine scanning on Decoralin and TAT(48-60) and identified residue positions relevant for the inherent anticancer activity of Decoralin and for the cell-penetrating activity of TAT(48-60). We computationally generated peptides mutated at these positions and used support vector machine (SVM) models of ACP, CPP, and THP activities to predict the respective properties of the computer-generated peptides. We successfully obtained Decoralin derivatives with two-fold improved anticancer activity in each tested cancer cell line and reduced toxicity against primary endothelial cells. This improved activity is due to the enhanced cell-penetrating ability in the targeted cancer cells, resulting in induced cell apoptosis. Some of the designed peptides with improved therapeutic indices show tumor-homing-like features. We used the same peptide design strategy to add anticancer activity to TAT(48-60). We successfully obtained mutated TAT(48-60) peptides cytotoxic for MCF7 cells but not for primary endothelial cells. Apparently, this additional anticancer activity is not caused by membrane-lytic activity but may result from the significantly improved cell-penetrating ability in MCF7 cells.

Overall, we demonstrate the application of computational-aided peptide design approaches in both de novo sequence design and multi-objective optimization. We successfully utilized sequence-based peptide design methods in combination with chemical, biophysical, biological and pharmacological experiments to address the design tasks. We obtained de novo AMPs with nanomolar membrane-lytic activities and bacteriostatic effects against both Gram-positive and Gram-negative pathogens. In the ACP project, we designed peptides with
improved activity in the tested cancer cell lines but with reduced toxicity against primary endothelial cells. We also obtained peptides that selectively target cancer cell lines and that possess better therapeutic indices than the design template. The study demonstrates the sustained potential of advanced computer-assisted methods for designing peptides with desired properties and activity.
Zusammenfassung


In dieser Arbeit präsentieren wir computergestützte Ansätze, um AMPs und ACPs zu entwickeln. Zunächst präsentieren wir einen computergestützten de novo Design-Ansatz, aus welchem sich neue AMPs ergeben haben, die bereits in niedrig-nanomolaren Konzentrationen direkte membranlytische Aktivitéät, antimikrobielle Aktivitéität und einen hohen therapeutischen Index besitzen. Die Methode basiert auf einer selbstorganisierenden Karte (SOM), welche die Aminosäuresequenzen mit der gewünschten Funktion priorisiert. Für das Training der SOM haben wir die Aktivitéiten von den zwei aus Wespengift stammenden


MCF7 Zellen, aber nicht für primäre Endothelzellen sind. Diese zusätzliche Antikrebsaktivität wird offensichtlich nicht durch eine membranlytische Aktivität, sondern möglicherweise durch die verbesserte Zellpenetrations-Fähigkeit in MCF7 Zellen verursacht.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACP</td>
<td>Anticancer peptide</td>
</tr>
<tr>
<td>AFM</td>
<td>Atomic force microscopy</td>
</tr>
<tr>
<td>AMP</td>
<td>Antimicrobial peptide</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial neural network</td>
</tr>
<tr>
<td>CADD</td>
<td>Computer-aided drug design</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>CPP</td>
<td>Cell penetrating peptide</td>
</tr>
<tr>
<td>CS</td>
<td>Chondroitin sulfate</td>
</tr>
<tr>
<td>CT</td>
<td>Chemotherapy</td>
</tr>
<tr>
<td>DCM</td>
<td>Dichloromethane</td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic light scattering</td>
</tr>
<tr>
<td>DMF</td>
<td>Dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>EA</td>
<td>Evolutionary algorithm</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>FITC</td>
<td>Fluorescein isothiocyanate</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic algorithm</td>
</tr>
<tr>
<td>GAG</td>
<td>Glycosaminoglycan</td>
</tr>
<tr>
<td>GAIN</td>
<td>Generating Antibiotic Incentives Now</td>
</tr>
<tr>
<td>HCTU</td>
<td>$1\text{-H-benzotriazolium1-}[\text{bis(dimethylamino)methylene}]\text{-5-chloro-}$, hexafluorophosphate (1-)3-oxide</td>
</tr>
<tr>
<td>HNP-1</td>
<td>Human neutrophil peptide-1</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HS</td>
<td>Heparin sulfate</td>
</tr>
<tr>
<td>HTS</td>
<td>High-throughput screening</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal titration calorimetry</td>
</tr>
<tr>
<td>LBDND</td>
<td>Ligand-based <em>de novo</em> drug design</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>LUV</td>
<td>Large unilamellar vesicle</td>
</tr>
<tr>
<td>MCC</td>
<td>Matthews correlation coefficient</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>MDR</td>
<td>Multi-drug resistance</td>
</tr>
<tr>
<td>MHC-1</td>
<td>Major histocompatibility complex-1</td>
</tr>
<tr>
<td>MIC</td>
<td>Minimum inhibitory concentration</td>
</tr>
<tr>
<td>MTT</td>
<td>3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide</td>
</tr>
<tr>
<td>NCE</td>
<td>New chemical entity</td>
</tr>
<tr>
<td>NMM</td>
<td>N-methylmorpholine</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PG</td>
<td>Proteoglycan</td>
</tr>
<tr>
<td>PI</td>
<td>Propidium iodide</td>
</tr>
<tr>
<td>PLS</td>
<td>Partial least squares</td>
</tr>
<tr>
<td>POPC</td>
<td>1-palmitoyl-2-oleoyl-<em>sn</em>-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>POPE</td>
<td>1-palmitoyl-2-oleoyl-<em>sn</em>-glycero-3-phosphoethanolamine</td>
</tr>
<tr>
<td>POPG</td>
<td>1-palmitoyl-2-oleoyl-<em>sn</em>-glycero-3-phospho-(1'-rac-glycerol)</td>
</tr>
<tr>
<td>PS</td>
<td>Phosphatidylserine</td>
</tr>
</tbody>
</table>
(Q)SAR  (Quantitative) structure-activity relationship

RBDND  Receptor-based de novo drug design

RF        Random forest

RT        Radiation therapy

SBDND  Structure-based de novo drug design

SD        Standard deviation

SOM       Self-organizing map

SPR       Surface plasmon resonance

SVM       Support vector machine

TAT       Trans-activator of transcription

TFA       2,2,2-trifluoroacetic acid

TFE       2,2,2-trifluoroethanol

THP       Tumor homing peptide

TIPS      Triisopropylsilane

WHO       World Health Organization

7-AAD     7-Aminoactinomycin
I. Introduction

1. Peptides as therapeutic agents

In the 20th century, most drug discovery and development were generated from small molecules, either based on high-throughput screening or on a rational design process including ligand-based and receptor-based molecular design.\(^1\) A retrospective analysis of successful marketed drugs concluded that ‘the rule of 5’ features predicting small molecules have a higher risk of poor absorption or permeation if these molecules possess: (1) more than 5 hydrogen-bond donors, (2) more than 10 hydrogen-bond acceptors, (3) a molecular weight over 500 Da and (4) an octanol-water partition coefficient log\(P\) greater than 5.\(^2\) Numerous guidelines of this type have been devised ever since.\(^3\) The preference for small molecules as drugs is because of their good oral bioavailability, which is advantageous for both efficacy and patient compliance. However, due to their small sizes, the major drawbacks of small molecules are their miscellaneous off-target activities, which are often ultimately manifested as side effects.\(^4\) Along with the advances in recombinant protein expression and better protein purification techniques, ‘biologics’ have emerged as an extremely successful class of therapeutics.\(^5,6\) These molecules, including proteins and antibodies, disobey every rule-of-five parameter; thus, these molecules are usually not suitable for oral delivery.\(^7\) However, their often better target specificity compared with small molecules has catapulted several biologics in the blockbuster category. The disadvantages of biologics, including high production cost, lack of membrane permeability, low oral bioavailability, and poor metabolic stability, have also triggered the drug discovery community to search for possible alternative therapeutic approaches.\(^8\) Therefore, peptides represent a promising class of drug that possess the advantages of both small molecules and proteins or antibodies. Peptides are generally
classified as molecules containing less than 50 amino acids. Peptides have high specificity and potency, similar to protein drugs, but are more easily accessible via chemical synthetic methods.9 Peptides have better tissue-penetrating activity compared with proteins and antibodies. The generic short half-life of peptides causes few of these molecules to accumulate in tissues, which reduces the risk of systemic toxicity.10 Although oral bioavailability and metabolic stability remain concerns in peptide drug development, recent progresses in formulation, delivery and chemistry have sparked much interest in peptide therapeutics.11 Thus far, more than 100 peptide-based drugs have reached the market, and the market share of peptide-based drugs is growing faster than that of other pharmaceutics.10

2. Antimicrobial peptides

2.1 Unmet medical need for antimicrobial agents

The spread of antibiotic-resistant bacteria, which results in untreatable infections, is a major worldwide public health problem. According to the WHO 2014 report on global surveillance of antimicrobial resistance, antibiotic resistance is no longer a prediction for the future; antibiotic resistance is currently occurring across the world.12 High proportions of resistance to common treatments for bacteria have been reported for all regions, causing infections in both healthcare settings and the community. Antimicrobial resistance has negative effects on patient outcomes and on health expenditures. In 2007, antimicrobial resistance resulted in 25,000 deaths in the European Union and caused 2.5 million extra hospital days, causing approximately €1.5 billion extra societal costs.12 However, the pace of antibiotic discovery has been slowing since 1999, when Roche, among the first large pharmaceutical companies, withdrew from antibiotic research, and other large companies followed suit.13 Nearly all antibiotics we use today were compounds discovered from the 1940s to 1960s by the platform
I. Introduction: 2. Antimicrobial peptides

introduced by Selman Waksman. After the 1960s, an excellent of class of synthetic antibiotics emerged, the fluoroquinolones, which are broad-spectrum compounds. The last new class of antibiotic, daptomycin, which is a lipopeptide that depolarizes the cell membrane of Gram-positive bacteria, was approved in 2003. In 2012, the narrow spectrum antibiotic bedaquiline was approved for tuberculosis treatment. Clearly, the tempo of bacteria developing and spreading resistance to current antibiotics has outpaced new antibiotic discovery. Developing therapies that effectively combat multidrug-resistant bacteria is urgently required. Several additional approaches to antibiotic discovery have been pursued, except for employing new strategies to facilitate antibiotic discovery, such as reviving high-throughput screening (HTS) campaigns with focused libraries designed for cell penetration or conducting rational design of pro-drugs of antimicrobial agents. Among these possibilities, antimicrobial peptides represent a promising class of antibiotics.

2.2 Characterization of AMPs

Antimicrobial peptides (AMPs) are molecules produced in many tissues and cell types in a variety of invertebrate, plant and animal species. These peptides contribute to the first line of defense mechanisms to kill or to slow down the growth of invading microorganisms and ally with other innate and adaptive immune systems. AMPs have the following generic features:

1. Length: AMPs range from five to 60 amino acid residues.

2. Sequence composition: AMPs often contain the basic amino acid residues lysine or arginine and a high percentage of hydrophobic residues, including alanine, leucine, phenylalanine, tryptophan, isoleucine, tyrosine and valine. Some AMPs also contain the acidic amino acid residues glutamic acid and aspartic acid.
3. Charge: Many AMPs are cationic peptides rich in arginine and lysine. A few AMPs are anionic and rich in aspartic and glutamic acid residues. Highly cationic peptides or anionic peptides, which form complexes with zinc, are often more active than neutral peptides.18

4. Structure: The above-mentioned properties permit these peptides to fold into an amphiphilic structure that form separate patches rich in positively charged and hydrophobic amino acids upon contact with membranes.22 Differently folded structures have been discovered in AMPs: α-helix (e.g., LL-37 and Magainins), β-sheet structures stabilized by two to four disulfide bridges (e.g., Protegrin), extended structures (e.g., Indolicidin), loops stabilized by one disulfide bridge (e.g., Thanatin) and mixed structures (e.g., β-Defensin-1) (Figure 1).

AMPs have great sequence diversity, and sequence similarity is observed for only defined groups of AMPs among closely related species. Nevertheless, AMPs even vary substantially among mammalian species.23 The rapid evolution of antimicrobial peptides and the innate immune system may represent a natural strategy against pathogen resistance and virulence.24
I. Introduction: 2. Antimicrobial peptides

2.3 Mechanisms of AMPs

Anionic phospholipids and phosphate groups on lipopolysaccharides (LPS) on the outer envelope of Gram-negative bacteria, as well as polysaccharides, teichoic acids and lipoteichoic acids on the outer capsule of Gram-positive bacteria, are the first binding targets of AMPs. The net negative charge of bacterial surfaces results in the major selectivity of AMPs for bacteria over neutrally charged eukaryotic cell surface. Once AMPs are attracted to microbial surfaces and traverse the outer capsular polysaccharide of Gram-positive and -negative bacteria, AMPs often bind in a parallel orientation to the bacterial cell membranes. As the peptide concentrations on the lipid bilayer increase, peptides are orientated
perpendicularly and inserted into the lipid bilayer, thereby permeabilizing the cell membrane via different mechanisms. Several models have been developed to explain the interactions between AMPs and cell membranes (Figure 2).

According to the **barrel-stave model**, peptides interact laterally with each other and aggregate as α-helical bundles in the membrane, with a central lumen that is reminiscent of an ion channel. The hydrophobic peptide regions align with the bilayer hydrocarbon core, using this core as a template for peptide assembly, whereas the hydrophilic peptide regions form the interior region of membrane-protein-like aqueous channels. The classic example for this type of transmembrane pore is Alamethicin, which is a 20-residue fungal peptide adopting an amphipathic α-helix. Alamethicin can only insert into the membrane and form barrel-stave pores when their concentrations exceed a lipid-dependent critical value. These pores consist of 3-11 helical molecules, with an inner diameter of approximately 1.8 nm, as detected by neutron in-plane scattering.

In the **toroidal-pore model**, no specific peptide-peptide interactions similar to the barrel-stave pore exist. Instead, once peptides insert into the membrane, these peptides induce lipid translocation across the membrane and bend continuously through the pores. Inserted peptides provide alternate interaction surfaces between lipid hydrocarbon chains and head groups. Inserted peptides and lipid head groups line the pores. Magainin 2 (23-residue AMP from *Xenopus* skin) and Melittin (26-residue peptide from honey bee venom) are two examples of toroidal pore-forming peptides. The flip-flop rate of phospholipids induced by Magainin 2 was investigated using fluorescent lipids. The results demonstrated that these peptides could accelerate the flip-flop rates of lipids and that the flip-flop was coupled with membrane permeabilization and with peptide translocation. As determined by neutron off-plane scattering, Magainin-induced pores have an inner diameter of 3.0-5.0 nm, which is
larger than Alamethicin-induced barrel-stave pores. Additionally, a molecular dynamics (MD) study demonstrated that Melittin might form disordered toroidal pores in which only one or two peptides lined the inside the pore and that the disordered toroidal pore could either represent a thermodynamically stable state or be an intermediate in the carpet or ordered toroidal pathway.

In the carpet model, which is the most frequently cited AMP mechanism of action, peptides accumulate in a parallel fashion on the bilayer surface in a carpel-like manner, electrostatically attracted to the anionic phospholipid head groups. At critical peptide concentrations, peptides are thought to disrupt the membranes in a detergent-like manner. This carpet model was actually observed for Cecropin P1 (31-residue peptide from roundworms) and in other short AMPs consisting of less than 20 amino acids that did not have sufficient lengths to span the cell membrane and to form pores.

In addition to cell membranes as major targets, AMPs also have intracellular targets. Some AMPs can translocate into cell membranes without permeabilization. Once peptides penetrate the cell membranes, these peptides accumulate in the cytoplasm and exert their inner cytotoxicity, such as inhibiting bacterial cell wall synthesis, nucleic acid synthesis, protein synthesis or enzyme activity. Buforin II is a synthetic 21-residue AMP containing a helix-hinge-helix structure. This peptide did not lyse the bacterial membranes, even at 5 times the minimal inhibitory concentration (MIC=4 µg/ml) without lysing the cell. Buforin II also shows 20-fold higher affinities for DNA and RNA than Magainin 2 in vitro. PR-26, which is a proline-arginine-rich peptide derived from the porcine neutrophil peptide PR-39, has been shown to inhibit cell division. AMPs have also been shown to exhibit multifaceted immunomodulatory activities, including profound anti-infective and selective anti-inflammatory properties. For example, human neutrophil peptide (HNP-1) is a potent
antimicrobial agent against a variety of bacteria in vitro; however, in vivo studies have demonstrated that its activity is mediated by leukocyte accumulation at the site of infection. Notably, HNP-1 lost its activity in leukocytopenic mice.

It is reasonable to assume that most AMPs have multiple simultaneous mechanisms. Forming transmembrane pores and extensive membrane rupture may represent continuous AMP-mediated bacterial cytotoxicity. Considering the physicochemical similarity between AMPs and cell-penetrating peptides (CPP), although rarely studied, some membrane-active AMPs will also likely have cell-penetrating activity. To conclude, multiple mechanisms of AMPs may explain the broad-spectrum activity of AMPs and rare cases of AMP-induced bacterial resistance.

Figure 2. Mechanisms of action of antimicrobial peptides. Once peptides attach to the bacterial cell membrane, peptides can adopt to specific secondary structures with amphipathic features capable of inducing membrane-permeabilization by (1) the barrel-stave pore model, (2) the carpet model, and (3) the toroidal-pore model. When peptides translocate into cells, these peptides can exert their inner cytotoxicity in different mechanisms, which are summarized in these figures. Adapted from ref 46.
2.4 Current clinical development of AMPs and outlook

Currently, no AMPs have been approved by the U.S. Food and Drug Administration (FDA) as drugs. Pexiganan, which is a broad-spectrum 22-amino acid membrane-lytic peptide derivative of Magainin, is among the closest to AMPs on the market. Pexiganan was turned down by the FDA in 2000 for the topical treatment of diabetic foot ulcers because its efficacy had not been sufficiently demonstrated. Several other AMPs are in either the early phase of clinical trials or preclinical development. Nevertheless, AMPs are widely used as food preservatives because of their potency and are easily broken down when ingested. The reasons for the lack of successful AMP marketing stories are primarily due to a lack of potency and due to pharmacokinetic problems. The cost of goods and potential systematic toxicity are also major hurdles in AMP development. In this regard, with a better understanding of their mechanisms of action and structure-activity relations, scientists may be able to rationally design and optimize AMPs with improved therapeutic profiles. Encouraged by the Generating Antibiotic Incentives Now (GAIN) act proposed by the FDA grants, additional years of market exclusively for antimicrobials agents and innovative trial designs, this initiative provides an incentive to reinvest in antibiotic research, which may simultaneously facilitate the growth of AMP development.

3. Anticancer peptides

3.1 Cancer is the leading cause of death worldwide

Although progress has been made in clinical cancer therapies, including chemotherapy, radiation, surgery or target therapies, cancer remains one of the leading causes of death globally. Twenty-six million new cancer cases and 17 million cancer deaths are estimated to occur annually within the next 20 years. Moreover, the high reoccurrence risk of this disease
and the emergence of resistance to current therapies in progressing tumors are major problems. It is believed that only by adopting a global prevention approach will it become feasible to slow down and ultimately reverse the worldwide burden of cancer. Conventional cancer managements are radiation therapy (RT), chemotherapy (CT) and surgery. However, both forms of therapy have severe side effects and low therapeutic indices. Most chemotherapeutics do not exhibit sufficient selectivity against healthy mammalian cells. The most common side effects are observed on rapidly dividing healthy cells, which is also the primary properties of tumor cells, such as myelosuppression, mucositis and alopecia. Additionally, both forms of therapy can cause delayed neurotoxicity. Another limitation for RT and CT is the development of resistance against treatment. Cancer cells can develop multiple drug resistance (MDR) via several mechanisms, including increased expression of drug detoxifying enzymes and drug transporters that eliminate the drugs from the cells before these drugs reach their intracellular targets. Overexpression of the MDR1 gene, which is responsible for the expression of those drug transporters, is often observed in patients relapsing after CT. New chemotherapies that can target specific differences between tumors and healthy cells have drawn attention. Antibody-based therapy is now one of the most successful and important strategies for treating cancers. Antigens on tumor cell surfaces have been revealed as targets selectively expressed or overexpressed compared with nontransformed tissue. The modes of anticancer action of monoclonal antibodies include receptor blockade, immune-mediated cell killing mechanisms, payload delivery, and specific effects on tumor vasculature and stroma. Currently, 14 antibody-based drugs are approved by the FDA for various oncology indications, and all of these drugs are immunoglobulin G (IgG) molecules or their conjugates. Many more antibodies, including engineered antibodies and novel
antibody-like variants, are currently in clinical trials. More recent strategies to combat cancers, including cancer immunotherapy and cancer vaccines, are also being actively discussed and show promising anticancer activities.

### 3.2 Anticancer peptides as innovative therapeutics

Anticancer peptides (ACPs) represent a promising class of anticancer agents. ACPs share molecular characterization and modes of action similar to AMPs; however, not all AMPs are ACPs, and the factors that determine ACP selectivity are not fully understood. The negative surface charge of the cancer cell membrane is a characteristic shared by bacterial cells but not by healthy mammalian cells (Figure 3). Anionic phospholipid phosphatidylserine (PS), sialylated gangliosides, o-glycosylated mucins, and heparin sulfate are widely distributed on the membranes of cancer cells, conferring a net negative charge, whereas the membranes of healthy mammalian cells are zwitterionic. The PS density on the surface of several cancer cell lines correlates with their sensitivity to AMPs. Sialic acid residues linked to glycoproteins and glycolipids are overexpressed in many breast carcinomas and in other cancer cells, such as ovarian, lung, colon and pancreas, and the extent of surface sialylation correlates with the metastatic potential of cancer cells. Although Buforin IIB primarily exerts its cytotoxicity against cancer cells via a mitochondria-dependent pathway, peptide-glycosphingolipid interactions are crucial for cellular peptide entry. Proteoglycans (PG) are glycosylated proteins bearing negatively charged glycosaminoglycan (GAG) side chains. Heparin sulfate (HS) and chondroitin sulfate (CS) are the two primary sulfate groups on the side chains of repeated disaccharides. A recent study has found that lactoferricin B was more active in melanoma cells that had higher levels of GAG expression than in colon cancer cells with lower levels of GAG expression; however, the inhibitory effect of the
peptide was enhanced when the sulfation levels of GAG in melanoma cells decreased. These authors suggested that the inhibitory effects of ACPs are not only influenced by GAG expression but also by the sulfation type on GAG side chains.\textsuperscript{79} Some ACPs bind with high affinity to HS, in contrast to CS, which sequesters ACPs away from the membrane.\textsuperscript{79,80}

**Figure 3.** Cartoon of specific characteristics of cancer cells that account for ACP selectivity. Widely exposed negatively charged lipids on the surface, PS (1) and increased levels of glycoproteins and lipids (2) can attract positively charged ACPs to cancer cells. The fluidity of the cell membrane (3) and the local pH (4) influence the susceptibility to and activity of ACPs. Microvilli (5) on the cancer cell membrane provide ACPs with extra contact surfaces, facilitating their specific interaction with cancer cells. Once ACPs enter cells, these peptides can target the negatively charged mitochondria and, consequently, induce cell apoptosis (6). Adapted from ref\textsuperscript{75}. 
In addition, a higher amount of cholesterol, along with other sterols in the cell membrane, decreases the fluidity of the membrane, which prohibits the lytic ability of ACPs. The cell surface area is also a factor contributing to ACP activity and selectivity. Elevated numbers of microvilli on malignant cells provide a large surface area to interact with ACPs molecules and to attract ACPs to cancer cells. The acidic microenvironment around cancer cells also affects their susceptibility to peptides, particularly in the form of an "activation switch" for peptides composed of the amino acid histidine, which is only positively charged at a low pH.

In general, the negatively charged molecules on the cancer cell surface render cells susceptible to ACPs. Once ACPs are electrostatically attracted to the cell membrane, the same membrane-lytic models used to describe the interaction between AMPs and the bacterial cell membrane have also been applied in describing ACP modes of action. The carpet and barrel-stave mechanisms can account for the membrane-lytic properties of some of the known ACPs. Following membrane permeation, ACPs induce cell death resulting from necrosis and/or apoptosis. Lactoferricin B has been reported to induce necrosis and apoptosis in vitro and to have a direct antitumor effect in vivo. Mitochondria possess a negatively charged membrane and can be targeted by ACPs once these peptides enter the cell. ACPs can penetrate mitochondria, inducing the swelling of mitochondria followed by releasing of cytochrome c and apoptosis. Notably, the modes of action of ACPs are not restricted to membrane-related mechanisms. In a study with HNP-1, which is a potent AMP that belongs to the α-Defensin group and that has 30 amino acid residues, the expression of mature HNP-1 in breast and colon mice tumor models could not only inhibit tumor growth but also increase cytotoxic T cell infiltration. This study also demonstrated that HNP-1 could mediate host immune responses to tumors in situ through recruiting and activating immature dendritic
cells. Its immune modulation activity was also demonstrated by treating \textit{in vivo} bacterial infection. Special ACP targets have been reported for some peptides. The N-terminal arginine residue of PR-39, which is rich in proline and arginine residues, has been shown to bind to the SH3-domain, subsequently inducing syndecan-1 production by a mechanism related to cell invasion inhibition and preventing cancer cell metastasis. Other studies also revealed the potential therapeutic benefit of conjugating lytic ACPs with tumor-targeting peptides or with CPPs to treat cancers.

### 3.3 Current clinical development of ACPs and outlook

Although some ACPs show promising activity and selectivity for cancer cells in \textit{in vitro} and \textit{ex vivo} assays, some \textit{in vivo} assays have demonstrated the systemic toxicity of ACPs. Those studies indicate the lack of therapeutic potential of the current ACPs, and thus far, only a few peptides from this class have entered clinical trials. LTX-315 is a synthetic membrane-active host defense peptide derived from Lactoferricin B that was developed against transdermal-accessible tumors. This peptide can not only eradicate cancer cells immediately but also induce an immune reaction targeting cancer cells. This peptide is currently being developed by Lytix Biopharma from Norway and is in a phase 1 clinical trial. Nevertheless, a detailed structure-activity relationship of ACPs is required for the rational design of more selective and active ACPs. Considering the specific features of each type of cancer cells to obtain selective ACPs will be an efficient strategy to obtain useful ACPs in the future.
4. Computer-assisted drug design

4.1 Computer-assisted small molecules design

The use of computers to design new active drug compounds dates back to the 1980 when quantitative structure-activity relationship (QSAR) studies were widely used in the drug design process to analyze the correlations between molecular structure and biological activity. However, QSAR studies were not sufficient in *de novo* lead molecule generation because of limited chemical space in the models, which resulted in primarily receptor-based ("structure-based") *de novo* drug design concepts introduced in the 1990s. De novo design aims at generating biologically active molecules from scratch and the obtained chemotypes represent new chemical entities (NCEs) that not only provide new insight into the atomic scale of molecular-receptor interactions but ideally are also patent-free. However, it is not practical to expect to immediately obtain molecules that are highly active, non-toxic, and orally bioavailable from computational approaches. Instead, these *de novo* molecules can serve as templates for further optimization. Structure-based *de novo* drug design (SBDND), or receptor-based *de novo* drug design (RBDND), and ligand-based *de novo* drug design (LBDND) are the two most commonly used computational approaches.

4.1.1 Receptor-based *de novo* drug design

RBDND aims at developing new molecules that bind to a specified biological target by assembling atoms and fragments according to molecular construction rules. An X-ray structure or reasonably valid homology models of desired targets are a prerequisite for structure-based drug design. The first critical step is to accurately identify binding pockets on the receptor models. Numerous programs have been developed for identifying potential ligand binding sites. For example, the PocketPicker software developed by our group
calculates the concavity values of grid probes installed in areas close above the protein surface and clusters the adjoining grid probes around deep regions to identify potential binding pockets and their shapes.\textsuperscript{101} In the next step of RBDND, ligands are generated by joining atoms or fragments so that the resulting product fits into the desired pocket, followed by employing filters or scoring functions to evaluate the "fitness" of the generated molecules. Many systematic approaches are used to generate novel ligand structures. Atom-based methods add the atoms one by one based on the receptor site, whereas the fragment-based approaches search fragment libraries and use the fragments to build new molecules by fragment growth and linkage. Pharmacophore-based methods, which is another widely used concept, generate structurally similar or complementary molecules to the pharmacophore retrieved from defined pockets.\textsuperscript{102}

Considering the almost unlimited chemical structures able to be generated \textit{de novo}, the RBDND programs tend to either include limited numbers of fragments and/or to employ sophisticated search strategies to sample hit compounds from chemical space. Evolutionary algorithms (EAs) are one of the most widely used methods to iteratively generate a population of molecules through mutation and crossover operations, and the generated offspring are evaluated by fitness functions.\textsuperscript{103} Typically, the offspring with high fitness scores will be retained to the next generation and will generate the next offspring. Virtual screening methods adopt pharmacophore-based similarity searching in libraries with many chemicals and couple the search with suitable fitness functions to identify potential hits. Structure-based \textit{de novo} approaches often use molecular docking and molecular dynamics methods to evaluate the fitness of the compounds.\textsuperscript{104-106}

The improvements of computational power in recent decades, with more available X-ray structures, have fostered a boom of receptor-based drug design studies. However, the reliable
and practical prediction of binding affinities remains unsatisfied. One primary reason for this insufficiency is that the binding of the ligand to its receptor is not only driven by the enthalpy contribution but also the entropy contribution. Entropy is contributed by the entire system, including proteins, ligands and solvent molecules surrounding the ligand and the receptor. However, in practical applications, we usually neglect the influence of the free-energy cost of displacing ordered or partially ordered water. SAR studies have shown that the binding of highly potent ligands containing approximately 30 or more atoms is primarily entropy-driven, which suggests that new fast scoring techniques are required to consider the entropy contribution in ligand-receptor binding.

4.1.2 Ligand-based de novo drug design

High-resolution X-ray crystal structures of desired targets are not always available for receptor-based drug design. In this case, the ligand-based design approach is an option. Ligand-based de novo design aims at creating new chemical entities by identifying fragments with similar pharmacophoric features to template scaffolds and by replacing the template with new fragments. This fragment-based molecular design follows the concept of scaffold-hopping, stating that newly designed molecules should possess property profiles similar to or superior to the template but should contain different scaffolds. NCEs are assembled by either commercially available building blocks or building blocks dissected by pseudo-retrosynthesis of known drugs. Our group has developed the software tool DOGS (Design of Genuine Structures), featuring reaction-based de novo design of drug-like molecules. This software was implemented for a stock of 25,144 available building blocks and 58 common reaction principles that enabled this software to suggest a synthesis route for each compound.
Adaptive design is another key feature of the ligand-based drug design approach.\textsuperscript{113,114} Coupling evolutionary algorithms or other nature-inspired algorithms with proper QSAR models or similarity indices as fitness functions to continuously \textit{in silico} optimize the designed compounds has been successfully applied in ligand-based optimization problems.\textsuperscript{115-117} Successful application of compound optimization must be performed on a smooth fitness landscape so that small changes in molecular structures only result in small changes in biological activity. The fitness landscape is defined by representing molecules and their underlying SAR features. Algorithms such as EAs can control the search process on a smooth fitness landscape and can design focused libraries covering chemical space with pharmacophoric features similar to template ligands, which may be exploited by the \textit{de novo} design process.\textsuperscript{113} Recently, Reutlinger \textit{et al.} presented a computational ligand-based design approach to this process. These authors developed a predictive quantitative polypharmacology model for 640 human targets and considered reductive amination a privileged reaction to create virtual drug-like products that clustered densely with known drugs from a drug database. Employing the nature-inspired concept of an ant colony to optimizing the building block selection, the authors successfully obtained novel subtype-selective and multitarget-modulating dopamine D4 antagonists and selective sigma-1 receptor binders.\textsuperscript{118} A paradigm in drug design has been to create target-specific ‘magic bullet’ drugs to remove undesirable toxic side effects. However, the increase in attrition rates in clinical trials, resulting from a lack of efficacy and safety, has challenged this dogma. Indeed, the ‘off-target’ activities of drugs were shown to decisively determine their efficacy and side effects, although the drugs were designed for a specific target.\textsuperscript{4} Understanding the polypharmacological nature of drugs and their effects on biological networks and phenotypes will be essential in drug discovery.\textsuperscript{119} Besnard \textit{et al.} developed a ligand-based multi-objective
automated approach aimed at designing drugs with specific multi-target profiles. These authors built Bayesian models from known drugs targeting 784 molecular targets as the scoring function and iteratively evolved the template molecule toward the desired polypharmacological profile or selective profiles for GPCRs.\textsuperscript{120}

4.2 Computer-assisted peptide design

Similar to small-molecule drug design, peptide design can be distinguished as template-based or receptor-based. In 1994, Schneider \textit{et al.} pioneered the use of physicochemical property descriptors of amino acids to code \textit{Escherichia coli} leader peptidase (SP1) and used artificial neural networks (ANNs) to recognize the characteristic cleavage site features. \textit{De novo} peptides were generated by an EA, taking the ANNs as the scoring function to optimize the sequences; this process was termed \textit{simulated molecular evolution}.\textsuperscript{121,122} Since this pioneer work, many studies have succeeded in designing different classes of peptides by building QSAR models based on machine-learning methods and by employing adaptive peptide strategies.\textsuperscript{123-125} Additionally, a new term, ‘computational peptidology’, has emerged to describe the use of computational and theoretical approaches to treat peptide-related problems.\textsuperscript{126} In a recent study, our group employed nature-inspired algorithms as sequence space exploration tools and a jury of feedforward ANNs as scoring functions to generate novel peptides that stabilize major histocompatibility complex I (MHC-I).\textsuperscript{125} A further study followed similar procedures and identified MHC-1 binders from genome-derived proteomes of vesicular stomatitis virus.\textsuperscript{127}

The contact surface between two interacting proteins is often flat and large, with only shallow pockets that are difficult for small molecules to target.\textsuperscript{128} The rational design of therapeutic peptides and peptidomimetics targeting protein-protein interactions with high affinity has
emerged as a new and promising tool in the discovery of potential drug targets.\textsuperscript{129} With the increase in high-resolution structures of protein-peptides complexes, large-scale structural studies have attempted to describe the key features of peptide binding.\textsuperscript{130} Akin to small molecule design, docking algorithms combined with a genetic algorithm were employed to design tetrapeptides against $\alpha$-Synuclein, which is a protein associated with Parkinson’s disease.\textsuperscript{131} However, this approach can only work for short peptide designs and when the structure of the target is known. Indeed, structural information is available for only 50% of the current drug targets; specifically, the structures of membrane proteins are lacking.\textsuperscript{132} Consequently, sequence-based peptide design approaches are commonly used by deriving sequence-binding motifs from databases of protein-peptide interactions.\textsuperscript{133,134}

One study employed an algorithm iteratively optimizing the predicted energy gap between peptide-target and peptide-off-target energy states to identify peptides of optimal stability and specificity. Then, protein arrays were used to characterize the designed peptides, and the results demonstrated that the designed peptides selectively bind to their representative member of bZIP proteins in the family in which these proteins shared strong sequence and structure similarities.\textsuperscript{135} Another study reported a method for the computational design of peptides targeting transmembrane (TM) helices. Motivated by the consensus sequence signatures of structural motifs with similar modes of interaction, first, the authors chose the helix-helix pairs from the database based on sequence compatibility with the target TMs. Then, the sequence of the target TM helix was threaded on one helix of the helix pair, followed by employing a Monte Carlo repacking algorithm that considered different combinations of side chains in low-energy rotamer states. Peptides specificity targeting TM helices of two different integrins, $\alpha_{\text{IIb}}\beta_3$ and $\alpha_\text{v}\beta_3$ were successfully designed, which computational designed helixes, anti-$\alpha_{\text{IIb}}$ and anti-$\alpha_\text{v}$, were able to activate the target integrin
and to induce human platelet aggregation.\textsuperscript{136} These studies have demonstrated that specifically binding peptides can be designed \textit{in silico} even when a high degree of sequence and structural similarity exists between target and non-target molecules.

### 4.3 Computer-assisted AMP design

Unlike designing small molecules for kinase inhibitors or peptides inhibiting protein-protein interactions with concrete structures of targeting receptors, structure-based drug design approaches are not applicable to generating membrane-active AMPs or ACPs. Instead, sequence-based approaches or quantitative structure-activity relationship (QSAR) models with machine-learning statistical methods are the most common strategies used in computer-assisted AMP design.\textsuperscript{137,138} A few studies also demonstrate the application of molecular dynamics simulations in AMP design.\textsuperscript{139} Except for computational design strategies, the choice of descriptors for modeling AMPs somewhat differs from the choice for modeling small-molecules. Compared with small molecules, peptides are of higher molecular weight and flexibility and have more repeated pharmacophoric features (amide bonds). Therefore, one will tend to choose descriptors for AMPs based on the prior understanding of physicochemical properties contributing to peptide activity rather than on representing all atoms in peptides in the same manner as in the design of small molecules.\textsuperscript{139,140}

#### 4.3.1 Sequence-based approaches

Starting from a known AMP template allows one to optimize the sequence to obtain improved activity and less toxicity. The conventional approaches rely on randomly or rationally mutating one or more positions in the AMP template so that one can realize the sequence-activity relationship of the template AMP to optimize a sequence with greater
activity and selectivity. Several types of amino acid substitution methods can be considered for optimizing peptide activity; for example, alanine and glycine scans can identify whether a particular side chain contributes to interactions with targets. D-amino acid scans can provide information regarding the importance of turn conformation; aspartic acid and lysine scans probe the charge contributions to the activity; leucine and tyrosine scans probe the importance of hydrophobic regions for activity; and other non-natural amino acid scans are also considered. In the case where considerable diversity in the sequence is tolerated or where functional roles of particular residues are poorly defined, these scans can quickly establish structural and functional preferences for peptide optimization. An efficient peptide synthesis method combining robust high-throughput activity assays opened a new phase in sequence-based AMP design. Hilpert et al. replaced each amino acid in Bac2A with all of the other 19 conventional amino acids, creating 228 unique peptides on a cellulose support. Antimicrobial activity was assessed by the extent of luciferase constitutively expressed by a Pseudomonas aeruginosa strain. These authors successfully optimized Bac2A activity by combining the most favorable substitutions of amino acids in the template. Another representative study generated a combinatorial peptide library based on a known membrane-spanning β-sheet-forming peptide by synthesizing the peptides on microbeads. High-throughput lipid-vesicle rupturing assays were used to screen for membrane-active peptides. Of a 10,000 peptides library, the authors obtained 16 highly membrane-active peptides with β-sheet structures. Overall, with the help of high-throughput peptide synthesis and assays, the sequence-based peptide design approach is well suitable for peptide optimization.

Computer-aided approaches can assist scientists in conducting the labor- and resource-intensive sequence-based AMP design from many aspects. Loose et al. treated
peptide sequences as strings where each amino acid in the sequence was equal to word in a sentence. The authors employed the Teiresias pattern discovery tool to identify a ‘grammar’ from natural known AMPs, resulting in a set of 684 ten-residue expressions. The grammar, e.g., [IVL]K[TEGDK]V[GA]K[AELNH]-[VA][GA]K, represented 600 sequences due to the option of choosing one of the amino acids in each bracketed position that correspond to an antimicrobial peptide pattern. The grammar set produced three million 10-mers, and the authors used these 10-mers to screen all possible 20-mer sequences for which each window of 10 amino acids corresponded to the three million 10-mers. The study successfully identified de novo AMPs showing activity against Gram-positive bacteria comparable to the natural AMP Melittin. Another project took advantage of computational calculation power by integrating a sequence alignment and feature selection method to build up an AMP prediction model.

4.3.2 Molecular dynamics simulation

Molecular dynamics (MD) simulations have been applied to understand the conformations and mechanisms of AMPs. In the context of a membrane-like environment, MD simulation can model the adsorption process of AMPs to the membrane, followed by spontaneous assembly of multiple peptides and, subsequently, membrane lysis. However, MD simulations often demand high computational power and long simulation times to simulate on the time scale of microseconds and to observe the entire process of AMP-membrane interaction. MD simulations are affected by many factors; thus, a slightly different parameter initialization can easily produce different results. Due to our limited knowledge of the atomic interactions in real world applications, the current MD simulations, which use consistent force fields to calculate molecular interactions, may oversimplify
Nevertheless, MD simulations have been useful in designing peptides with the desired activity. Tsai et al. performed MD simulations with a total simulation time of up to 4 µs, modeled the mechanism of the process of Indolicidin adsorption onto a model membrane and observed the insertion into the zwitterionic and negatively charged membranes. These authors identified the amino acids that were most responsible for the adsorption onto the negatively charged membrane and related the antimicrobial activity and the insertion into zwitterionic membrane as critical for hemolytic activity. Therefore, Indolicidin derivatives were designed by replacing amino acids at certain positions with the desired amino acids, and peptides with improved activity and selectivity were successfully obtained.

4.3.3 AMP QSAR models

4.3.3.1 Descriptors

Molecular descriptors are used to describe quantitative properties of peptides. The objective is to choose the descriptors that correlate the most to the peptides' activity and, consequently, to use different strategies to build up QSAR models in an attempt to design more potent peptides. The following two different types of descriptors exist: descriptors that can be calculated (e.g., peptide net charge, mean hydrophobic moment, and the most probable three-dimensional (3D) structure of peptides) and descriptors obtained from empirical measurements (e.g., HPLC retention time and aqueous solubility). These descriptors can be used to characterize either individual amino acids or entire peptides.

In general, chemical descriptors can be divided into different dimensionality groups, from 0D-4D. The complexity and high molecular mass of peptides restrict the use of 0D-2D descriptors that encode from the bond counts, atom counts, hydrogen-bond acceptor or donor to molecular connectivity and topology. Although 3D and 4D descriptors carry a higher
information level than the simpler descriptors, current energy minimization algorithms for predicting peptide 3D structure are insufficiently accurate to provide reliable QSAR models. Therefore, the mainstream descriptors for modeling AMP QSAR studies are combinations of individual residue descriptors generated to capture the features of entire peptide sequences. The pioneering work in AMP QSAR studies was performed in 1987, when Hellberg et al. first implemented 29 different physicochemical properties describing the lipophilicity, hydrophobicity, charge and size, HPLC retention times at various pH and elution conditions, and NMR data of the amino acids. Through principal component analysis (PCA) and orthogonal transformation, this extensive set of parameters was condensed to three descriptors (the z-scales) and successfully applied to model AMP activities. This z-scale approach was later revised to include some common non-coded amino acids and led to five principal components (z1-z5). Each component represented lipophilicity (z1), size, shape and polarizability (z2), polarity and electronic properties (z3), electro-negativity, the heat of formation, electrophilicity and hardness for the z4 and z5 components. Since the initial z-scale was introduced in 1987, a continually growing number of amino acid scales have been developed, including those scales aimed at capturing whole molecule features and topology. However, most descriptors can only be applied to relatively short peptides or to a small set of AMPs, and the initial z-scale remains considered one of the best performers. Nevertheless, descriptors for whole peptides are required for next-generation AMP design because interpreting peptides as a combination of amino acids is a strong simplification of their actual characteristics.

4.3.3.2 Linear modeling methods

Principal component analysis (PCA) and partial least squares projection (PLS) are two
common linear algorithms that have been successfully used for modeling antimicrobial activity.\textsuperscript{165} These linear methods project the high-dimensional data onto a lower dimensional space. This projection is achieved by transforming the original data to a new set of uncorrelated variables, the principal components, which are ordered so that the first few new variables retain most of the variation present in all of the original variables.\textsuperscript{166} Component analysis helps find the related features (descriptors) from the data, and linear regression is performed using small numbers of descriptors to fit the measured and predicted activities.\textsuperscript{113} Lejon \textit{et al.} used the theoretical variables of all natural amino acids and derived the variables via PCA. The antibacterial activity of the peptide could be modeled using three variables (alpha-helicity, hydrophobicity/hydrophilicity and charge) per amino acid position.\textsuperscript{160,165} QSAR analysis of peptides using PCA modeling has also been shown as an effective method to explain and to predict the anticancer and antiviral activities of peptides.\textsuperscript{167,168}

### 4.3.3.3 Machine-learning methods

Non-linear machine learning methods can fit hundreds of descriptors simultaneously. However, these complex prediction models often require a large number of training examples, in this case, the minimum inhibitory concentration (MIC) values of AMPs, to fit a large set of parameters. Traditional solid-phase peptide synthesis and MIC determination may not be efficient for generating sufficiently large training datasets within a reasonable period. Peptide synthesis on a cellulose support plus surrogate measures of bacterial killing, such as lipid vesicle rupturing assays, and monitoring luciferase changes constitutively expressed by bacteria can fulfill these tasks.\textsuperscript{148} By combining cellulose peptide synthesis technology and bacterial luciferase assays, two random nine-amino acid peptide libraries (Sets A and B) were assembled with amino acid compositions of highly active AMPs. Set B was further iteratively
designed by adjusting its amino acid composition toward the amino acid occurrence of the best performance peptides in Set A. The peptides in Set B had similar or better antimicrobial activity compared with the template and compared with Set A. The authors subsequently utilized the amino acid composition of Set B to virtually generate 100,000 nine-mers. A QSAR descriptor, which considered all of the peptides' atoms and characters of individual amino acids, was used to code Set A and Set B, and QSAR models were trained by ANNs. The QSAR models were shown to be sufficient for predicting the activity of the virtual peptides, and two of the selected de novo peptides showed high in vitro and in vivo antibacterial activity (MIC = 0.3-11 µM against most of the tested Superbugs) but neither hemolytic activity nor systemic toxicity.$^{138,169}$

ANNs are one of the most effective pattern recognition techniques utilizing artificial intelligence. A seminal publication in the field introduced the backpropagation algorithm for training large feed-forward networks.$^{170}$ In peptide QSAR model building, peptides in a training dataset are often encoded in the form of descriptors (as real-value vectors) and measured activity (0 for inactive, 1 for active), and the network training algorithm attempts to extract a pattern from these training data. Figure 4 shows the representative feed-forward network architecture used for modeling sequence (structure)-activity relationship.

The descriptor values are fed to input layers, and then the calculation propagates from the input layer to the connected hidden layer as weighted sums. The output value of a hidden unit is a nonlinear combination of nonlinear functions from the previous layer:

$$o = \sigma(\overline{w} \cdot \overline{x})$$

where $\sigma(y) = \frac{1}{1+e^{-y}}$ (1)

where $o$ is output, $\overline{w}$ is the weight vector of the input vectors $\overline{x}$, and $\sigma$ is often called the sigmoid function or logistic function.
Figure 4. Feed-forward neural network model. Input values, which are often real-value vectors of peptide representations, are fed to the input layer. Then, the values are then propagated to the hidden layer and non-linearly transformed to the output layer. Adapted from Ref\textsuperscript{171}.

The ANNs usually employ the backpropagation algorithm to learn the weights for a multilayer feedforward network. The ANNs employ gradient descent to minimize the squared error between the computed network output values and the expected target values for these outputs:

$$E(\vec{w}) \equiv \sum_{d \in D} \sum_{k \in \text{outputs}} (t_{kd} - o_{kd})^2$$

where $E(\vec{w})$ defines the training error of a hypothesis (weight vector) relative to the training examples, $D$ is the set of training examples, and $t_{kd}$ and $o_{kd}$ are the target and output values associated with the $k^{th}$ output unit and training example $d$, respectively. Although ANNs learning times can be relatively long, its evaluation process is fast for the learned target function. Additionally, ANN learning methods can be robust to noise in the training data, although interpreting the weights learned by ANNs is difficult.\textsuperscript{172}

Combining a sophisticated activity estimator with techniques enabling the stochastic optimization of peptides has recently been used for AMP design. Following the earlier
research by Schneider et al.,\textsuperscript{173} Fjell et al. demonstrated the use of genetic algorithms (GAs) to heuristically generate candidate sequences and employed trained ANNs as fitness functions to identify \textit{de novo} active AMPs. In this study, the authors demonstrated that a heuristic GA search method could more efficiently identify active peptides \textit{in silico} than searching a large-biased random library,\textsuperscript{137} which is perfectly consistent with earlier studies.\textsuperscript{173} GAs are examples of evolutionary algorithms that attempt to solve complex problems by mimicking the process of Darwinian evolution.\textsuperscript{103} GAs generate offspring solutions by repeatedly mutating and recombining parts of the best currently known hypotheses, e.g., for peptide design, the peptides with the highest fitness (for example, the highest probability as AMPs). At each step, a collection of hypotheses (the current population) is updated by replacing some fraction of the population with the offspring of the best fit of current hypotheses. GAs can search the space of hypotheses containing complex interacting parts where the effect of each part on overall hypothesis fitness is difficult to model.\textsuperscript{172}

Another example combining machine-learning methods with stochastic peptide optimization for AMP design was designing AMPs by evolutionary multiobjective optimization. Maccari \textit{et al.} constructed QSAR models of AMPs using random forest tree (RF) methods with global and topological descriptors. These authors encoded two sets of peptides: Dataset A, which represented the functional requirement for AMP activity, and dataset B, which contained alpha-helix peptides from a well-defined non-redundant set of proteins. The goal was to optimize random peptides via an evolutionary algorithm to obtain \(\alpha\)-helix AMPs. The study also demonstrated the possibility to computationally design peptides with non-natural amino acids.\textsuperscript{174}

Non-linear machine-learning techniques are more complex than linear models; thus, these techniques are expected to capture realistic non-linear SARs of AMPs. However, complex
machine-learning methods often require many samples for training. High quality training data are critical for the robustness of the QSAR models. Therefore, an iterative machine-learning peptide design cycle that combines high-throughput reproducible experimental read-outs has a good chance to enhance the utilization of computer-aided approaches in AMP discovery.

4.4 Computer-assisted ACP design

Unlike AMP design, currently, only few exercises in computational ACP design exist. In 2003, Yang et al replaced aromatic amino acids on a peptide L5 with three other amino acids to probe the SAR of the peptide’s anti-cancer activity and selectivity using PCA methods. Tyagi et al established an ACP database and built up a webserver with a support vector machine (SVM) QSAR model using the amino acid composition descriptor for virtual screening of de novo generated ACPs. Another study also applied a SVM QSAR model with alignment kernel algorithms to predict the anticancer activity of peptides derived from the HIV-1 p24 protein. Computational design in ACPs remains in its infant stage and represents a challenging field for further investigation.
II. De novo design of antimicrobial peptides

1. Aim

AMPs are widely found in nature from bacteria to eukaryotes. Most of these peptides known so far are either discovered from organism or modified from the known AMPs. These peptides are diverse in their sequences, structures and function, which their complex SAR impede our progress in AMP development.

Computer-assisted peptide design approach may provide a resolution to the adversity. Some studies as mentioned in the Introduction utilized computational approaches for de novo design AMPs,\textsuperscript{137-139,158,164} however most of them required a large amount of training dataset from robust but exhausted biochemical assays to build up reliable QSAR models. Our aims are as follows:

- *De novo* design of AMPs.
- Avoid exhausted synthesis and test cycle as traditional sequence-based peptides design approach.
- Conduct small-scale iterative peptide design cycle.

2. Hypotheses

- There is sequence ‘context’ preference for membrane-active peptides.
- Self-organizing map (SOM) is sufficient to cluster only a few tested peptides based on their activity and is able to prioritize the novel peptides with AMP-like properties.
3. Materials and Methods

Peptide synthesis and analytics. DMF (dimethylformamide), DCM (dichloromethane), diisopropylether, pyrolidine and TIPS (trisopropylsilane) were purchased from Sigma-Aldrich (St.Louis, MO, USA). NMM (N-methylmorpholine) and TFA (2,2,2-trifluoroacetic acid) were acquired from Fisher Scientific (Waltham, MA, USA); Fmoc-protected Wang-resins, Fmoc-protected amino acids, and HCTU (1H-benzotriazolium 1-[bis(dimethylamino)methylene]-5-chloro-hexafluorophosphate (1-),3-oxide) were obtained from AAPPTEC (Louisville, KY, USA). A Symphony robotic solid phase peptide synthesizer (Protein Technologies, Tucson, AZ, USA) was used to synthesize peptides utilizing 10-fold excess of Fmoc-protected amino acids relative to the Fmoc-Wang-resin. Coupling was performed with a ratio of 1:1:2 (amino acid, HCTU and NMM) in DMF. Automated cleavage was applied for 2 h with Reagent B (TFA-phenol-H_{2}O-TIPS (88:5:5:2)). The overall reaction is summarized below:

Activation of coupling amino acids:

\[
\text{FMOC-AA} + \text{HCTU} \rightarrow \text{C-terminal activated FMOC-AA}
\]
De-protection of amino acids linked on resin (growing peptides):

Coupling amino acids on growing peptides:

Peptides were washed and precipitated by using ice-cold diisopropylether. All peptide products were analyzed on an LC-20A HPLC instrument (Shimadzu, Kyoto, Japan) using an rpC18, 110 Å, 5 µm, 150 x 3 mm column (Macherey-Nagel), with a linear gradient of 5-70% ACN/H₂O (0.1% FA) over 25 min with a flow rate of 0.5 ml/min. Masses were detected between 300-1500 Da with a Shimadzu LCMS-2020 single-quad mass spectrometer (ESI+). UV²¹⁰ purity > 90%.

**Preparation of large unilamellar vesicles (LUVs).** We performed lipid vesicle rupturing assays as surrogate screening assays to determine the membrane-lytic activity of peptides. They allowed us to study the direct membrane preferences of the peptides without disturbance by other membrane-bound macromolecules.

LUVs were prepared from 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoehthanolamine (POPE), and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1’-rac-glycerol) (POPG) sodium salt (Avanti Polar Lipids, Alabaster, USA). LUVs were composed either from 100% POPC or
POPE:POPG (70:30) to mimic eukaryotic or bacterial cell membrane. Phospholipids dissolved in chloroform were dried under vacuum for at least 2 h. The dried lipid film was hydrated with 0.2 M Tris buffer containing 50 mM carboxyfluorescein (Sigma-Aldrich) and subjected to 10 freeze/thaw cycles. The lipid mixture was drawn into a syringe and pushed through an extruder (Avanti Polar Lipids, Alabaster, USA) with 0.1 µm or 1 µm polycarbonate membranes for 10 cycles. Dye-entrapped LUVs were separated by gel filtration on a Sephadex G25 column, and the lipid concentration was determined by a total phosphorus assay.

**Lipid vesicle assay.** Carboxyfluorescein-loaded LUV suspension was incubated with different concentrations of test peptide to yield a final lipid concentration of 10 µM. Fluorescence emission intensity was measured using Tecan Infinity M1000 spectrophotometer (Tecan, Switzerland) over five min ($\lambda_{\text{ex}} = 492$, $\lambda_{\text{em}} = 517$ nm), and the emission intensity after three min was used in the analysis. The percentage of peptide-induced membrane lysis was compared to 0.1% Triton X-100 (set to 100%). EC$_{50}$ values were determined by fitting the Hill equation to the experimental data ($n = 3$) using the software GraphPad Prism (GraphPad Software, Inc., La Jolla, USA).

**Hemolysis assay.** Synthesized peptides were tested on fresh blood cells from C57BL/6 mice to determine their hemolytic activity as previously described with slight modification. Briefly, fresh mouse red blood cells were collected and mixed 1:1 with Alsever’s solution (NaCl 4.2 g, D-glucose 20.5 g, sodium citrate 8 g, citric acid 0.55 g in 1 l solution). The mixture was washed twice with 10 mM PBS (pH = 7.4), centrifuged for 10 min (600 g) at 4°C, and re-suspended in PBS to attain a 1:10 (v:v) dilution of the erythrocyte volume initially
collected. 90 µl of blood solution was incubated with 10 µl peptide in serial dilution concentration dissolved in water for 1 h at 37°C. Intact erythrocytes were pelleted by centrifugation for 10 min (800 g) at 4°C, and the supernatant was transferred to a 96-well plate. Release of haemoglobin was measured as absorbance at 405 nm. The percentage of peptide-induced haemolysis was compared to 0.1% Triton X-100 (set to 100%). The minimal hemolytic concentration (MHC) was defined as the lowest peptide concentration, which induced 25% hemolysis compared to Triton treatment.

**Circular dichroism (CD) spectroscopy.** CD is defined as the unequal absorption of left-handed and right-handed circularly polarized light. CD spectra of peptides were measured on a Chirascan™ instrument (Applied Photophysics, Leatherhead, UK). The peptides were dissolved either in pure water or 50% 2,2,2-trifluoroethanol (TFE) (v/v) yielding a final concentration of 100 µM. The solution was loaded into a 0.1 cm rectangular quartz cell and spectra were recorded between 180-260 nm with a step size of 1 nm. The secondary structure of peptides were estimated by on-line analysis with DichroWeb\textsuperscript{181} using CONTIN methods.\textsuperscript{182,183}

**Surface plasmon resonance (SPR).** SPR is a charge-density oscillation that may exist at the interface of two media with dielectric constants of opposite signs, e.g., metal and water. It can be excited by light beam with same momentum of the metal. Changes in the measured refractive index at the interface, given in response units, are proportional to the amount of material in the immediate vicinity of the sensor surface.\textsuperscript{184} SPR was used to quantify peptide-membrane interaction for selected peptides. Experiments were carried out on a SierraSensors SPR-2 instrument (Sierra Sensors, Hamburg, Germany)
II. De novo design of antimicrobial peptides: Materials and Methods

using HPA sensor chips (Sierra Sensors). Lipid vesicles were immobilized on the HPA sensor chips, modified from published protocols.\textsuperscript{185-187} Briefly, 20 mM sodium phosphate with 150 mM NaCl (pH 7.4) was always used as a running buffer, and the system was running at a constant flow rate of 25 µl/min. The surface of the HPA chip was cleaned by 25 µl of cleaning buffer containing 2-propanol and 50 mM NaOH (4:6, v/v). After rinsing the system by two buffer injections, LUV (200 µl, 500 µM) was injected immediately to immobilize lipid vesicles onto the sensor chips. Then NaOH (20 µl, 10 mM) was applied to remove multilamellar lipids on the surface. The second injection of LUV was needed to ensure satisfactory coverage. The negative control BSA (0.1 mg/ml in running buffer) was injected to confirm complete coverage of the nonspecific binding sites. Peptide solutions (75 µl in different concentrations) were then injected and allowed a 10 min dissociation time to monitor their interaction with the lipid surface. Sensorgrams were analyzed using Scrubber 2.0 software (BioLogic Software Pty Ltd, Campbell, Australia).

**Bacteria and real-time growth monitoring.** GFP-transformed *Escherichia coli* (K12 derivate DH5α) were grown on LB agar plates containing 30 µg/ml chloramphenicol. *Staphylococcus aureus* SH1000 expressing GFP was a gift from Phil Hill (The University Nottingham, UK) and was cultured on nutrient agar containing 30 µg/ml chloramphenicol. For monitoring bacterial growth, 9 x 10^6 bacteria were incubated in 200 µl LB medium or nutrient broth, respectively, containing 30 µg/ml chloramphenicol and 100 µg/ml peptide solution in transparent Nunclon™ Edge 96-well plate (SIFIN GmbH, Germany) for 24 h at 37°C in a M200 PRO Quad4 Monochromators™-based multimode reader (Tecan, Anif, Austria). Loading the plate moats with 1.5 ml of 0.1% agarose reduced evaporation of the culture medium. Growth was monitored by detecting the fluorescence signal of
GFP-transformed bacteria at 600 nm.

**Dynamic light scattering.** DLS of vesicles was performed at a wavelength of 660 nm and scattering angle of 90° using a linear spacing of the correlation time with a Brookhaven digital autocorrelator and analyzed by digital autocorrelation software (Brookhaven Instruments Corporation, New York, USA). Mean and standard deviation for the size distribution were calculated by assuming a lognormal distribution of intensity-weighted size distribution.

**Isothermal titration calorimetry (ITC).** Isothermal titration calorimetry is a highly sensitive tool that provides binding affinities, binding isotherm, and entire thermodynamic parameters of peptide-vesicle interaction. ITC measures the energetics of peptide-lipid interaction at constant temperature. In this case the concentration of reactant is therefore the independent variable under experimental control, and the heat change associated with the binding is the direct thermodynamically observable. By calculating the degree of binding (i.e., the fraction of bound peptides per mole of total lipid) and the corresponding free peptide concentration, one can fit these values into an empirical-derived binding model (e.g., one-site binding model) and obtain $K_d$ value, which is the equilibrium dissociation constant of between the peptide and the lipid. The binding between the lipid membrane and the peptide can thus be characterized thermodynamically by its $K_d$ and $\Delta H$ measured by ITC. The corresponding free energy of binding, $\Delta G$, can be calculated using the standard formula:

$$\Delta G = -RT \ln \left( \frac{55.5}{K_d} \right)$$

We can then calculate the binding reaction entropy $\Delta S$ from the equation:
II. *De novo* design of antimicrobial peptides: Materials and Methods

\[
\Delta S = \left( \frac{\Delta H - \Delta G}{T} \right)
\]

(4)

Thermal characterization of peptide-vesicle interaction was performed on a high-sensitivity nanoITC instrument (TA Instruments, Newcastle, DE, USA) with a measuring cell volume of 177 µl and a syringe volume of 50 µl. Data were acquired by NanoAnalyze computer software (Nano ITCRun Software v2.2.5) developed by TA Instruments. Peptide and vesicle solutions were prepared in 10 mM PBS buffer (pH = 7.4) to yield the appropriate concentrations. The measuring cell was filled with 300 µl of de-gassed peptide solution, and the vesicle solution was filled into the syringe. Titration was initialized with an aliquot of 1 µl and the recorded data was omitted from data analysis. For productive runs, we titrated 20 aliquots of 2.5 µl vesicle solution (10 mM) into the measuring cell containing 100 µM peptide solution. In addition, titrations of vesicles into buffer were performed to correct the measurements for heat of dilution effects. We analyzed the corrected data with the "independent" model available in the NanoAnalyze Software v3.1.2.

**Atomic force microscopy.** Topographic AFM images of immobilized vesicles on HPA sensor chips for SPR (Sierra Sensors, Hamburg, Germany) were recorded using the Cypher AFM system from Asylum Research (Santa Barbara, USA) with Nanosensor SSS-NHCR super sharp cantilevers. Measurements were conducted in AC mode at 290 kHz, 0.8 V amplitude and a set point of 0.7 V. The scanning frequency was 0.5 Hz. Images were processed with Gwyddion (www.gwyddion.net) and the Asylum Research software.

**Peptide representation by correlation vectors.** Amino acid sequences were converted to a sequence length-independent, real-valued vector representation using a cross-correlated
pharmacophore feature approach. Briefly, all pairs of features separated one to six residue positions apart (correlation distance: 1-6 positions) were counted and the resulting occurrences divided by the length of the peptide (here: 11 residues). We used the pharmacophore feature types suggested by Schneider and Baringhaus, which resulted in a 126-dimensional real-valued vector for each peptide sequence.

**Hydrophobic moment.** Hydrophobic moment calculations were carried out using an automated python script. In adaption to the definition of the hydrophobic moment given by Eisenberg in 1982, the script first calculates a bundle of vectors originating from the peptides center of mass. Every vector \( v_i \) points towards the center of exactly one amino acid and is scaled to a length correlated to its relative hydrophobicity \( H_i \) - Normalized consensus values. The absolute length of the vector sum resulting from adding the individual vectors and scaling them with the total number of amino acids represents the molecules hydrophobic moment \( \mu_H = \frac{\| \sum^n_{i=1} (H_i \cdot \frac{v_i}{\|v_i\|}) \|}{N} \). Hydrophilic amino acids receive negative scores in terms of hydrophobicity reversing their respective vectors’ directions. Thus, amphiphilic peptides exhibiting both a hydrophilic and a hydrophobic surface receive greater hydrophobic moment values than uniformly hydrophobic (or hydrophilic) ones. The three-dimensional peptide structures required for the hydrophobic moment calculations were generated using a Python script pipeline incorporating PyMOL (The PyMOL Molecular Graphics System, version 1.5.0.4, Schrödinger, LLC, Portland, USA) and the Molecular Operating Environment v2012.10 (Chemical Computing Group Inc., Montreal, Canada). Input sequences were first projected onto idealized helical topologies using PyMOLs mutation wizard. Secondly, in order to prevent side chain clashes and irrational conformations due to potential intramolecular repulsion, structures were then energy-minimized in MOE. Minimizations
were carried out with the AMBER99 forcefield using an implicit (Generalized Born) solvation model with a gradient of $10^{-5}$ RMS kcal/mol/Å.<sup>192</sup>

**Self-organizing map (SOM).** A self-organizing map is a type of machine-learning systems that is trained using unsupervised learning to produce a low-dimensional representation of the high-dimensional input space of the training samples. The model was first described as an artificial neural network by Teuvo Kohonen in 1982.<sup>193</sup> A SOM consists of components called neurons that are usually arranged as rectangular grids. Each node is associated with a weight vector of the same dimension as the input data vectors and a position in the map space. The initial weight vectors of neurons may be given randomly. The training starts with feeding a training example to the network followed by determining a neuron closet to the data points. Various methods for distance computation have been utilized and the Euclidian distance metric is commonly used. The neuron with weight vectors most similar to the input and its neighbor neurons in the SOM lattice are adjusted towards the data point. The update formula for the weight vectors $W_v(s)$ of the neuron is the following:

$$W_v(s+1) = W_v(s) + \Theta(u,v,s)\alpha(s)(D(t) - W_v(s)),$$  \hspace{1cm} (5)

where $s$ denotes the current iteration, $t$ is the index of target input in dataset $D$, $\alpha(s)$ is a monotonically decreasing learning coefficient, $D(t)$ is the input data vector, and $\Theta(u,v,s)$ is neighborhood function depending on the lattice distance between the closest neuron $u$ and its neighbor neuron $v$. The process is repeated for each input sample for a defined cycle of $\lambda$.<sup>194</sup> We used an in-house SOM implementation (*molmap* software tool)<sup>195</sup> with a $20 \times 15$ toroidal map layout, Gaussian neighborhood function, linear decay of the learning rate ($\tau^{ini} = 1$) and adaptive neighborhood width ($\sigma^{ini} = 10$). Training samples were presented in random order.
The complete SOM algorithm can be formulated as follows:\textsuperscript{196}

Step 1. Initialize the self-organizing map $A$ to contain $\sum N = N_1 \times N_2$ neurons with reference vectors $\mathbf{W} \in \mathbb{R}^n$. The vector is randomly assigned from $\mathcal{D}$. Initialize the time parameter $t = 0$.

Step 2. Generate at random an input signal from the training set.

Step 3. Determine the winner neuron according to the Euclidian distance metric.

Step 4. Adapt the neuron vectors according to Equation 5.

Step 5. Increase the time parameter: $t = t + 1$.

Step 6. If $t < t_{\text{max}}$ then continue with Step 2, otherwise terminate.

\section*{4. Results and Discussion}

\subsection*{4.1 Choosing AMP templates and scrambled the templates}

We first chose the template peptides to work on. Since there were several types of AMP with different structures and modes of action, we decided to choose short peptides, which were easily handled in automated peptides synthesizers. We chose seven peptides from Antimicrobial Peptide Database 2 (APD2, http://aps.unmc.edu/AP/main.php)\textsuperscript{197} with additional filtering criteria, such as without methionine and cysteine in the sequence (two amino acids are easily oxidized), less toxicity in mammalian cells, and rarely studied sequences (Table II.1). Chosen peptides were synthesized and their membrane-lytic activities were tested by a lipid vesicle-rupturing assay. LUVs served as a surrogate for bacterial and eukaryotic cell membranes. They allowed us to study the direct membrane preferences of the peptides without disturbance by other membrane-bound macromolecules. We prepared two sizes of LUVs (target diameter: 0.1 $\mu$m or 1 $\mu$m) with lipid compositions of (i) 70\% POPE and 30\% POPG – this lipid mixture is referred to as "PEPG" herein –, or (ii) 100\% POPC.
The effective size distribution of the LUVs was determined by dynamic light scattering and showed 143±2 nm for the 0.1 µm preparations (*polydispersity* = 0.07±0.01), and 278±5 nm (*polydispersity* = 0.20±0.01) for the 1 µm preparations (values given as *mean ± standard error*). There was no statistically significant difference in size distribution between different batches of lipid vesicles and lipid compositions. Indolicidin, Decoralin, Protonectin and Temporin-1PRb showed membrane-lytic activity, and Decoralin and Protonectin were the most active peptides (Figure II.1A, B). As a result, we picked Decoralin and Protonectin as the seed peptides.
II. *De novo* design of antimicrobial peptides: Results and Discussion

Table II.1 Summary of peptides tested for template sequences. Information adapted from APD2. N.A.: not available. *: Articles in PubMed as of August, 2011.

<table>
<thead>
<tr>
<th>Name, Sequence</th>
<th>Net charge, hydrophobic residues %</th>
<th>MIC Gram + (<em>S. aureus</em>)</th>
<th>MIC Gram - (<em>E. coli</em>)</th>
<th>MIC Fungi (<em>C. albicans</em>)</th>
<th>Eukaryotic cell toxicity</th>
<th>Structure</th>
<th>Origin</th>
<th>First published</th>
<th>Related articles*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indolicidin_{CONH2} ILPWKWPWWPWRR</td>
<td>3, 53%</td>
<td>10 µg/ml</td>
<td>Active</td>
<td>Active</td>
<td>Toxic</td>
<td>Extended coil</td>
<td>Bovine neutrophil</td>
<td>1992</td>
<td>135</td>
</tr>
<tr>
<td>Decoralin_{CONH2} SLLLSLRLKLIT</td>
<td>2, 54%</td>
<td>40 µM</td>
<td>Active</td>
<td>40 µM</td>
<td>No toxic</td>
<td>Helix</td>
<td>Solitary, <em>Oreumenes decoratus</em> (wasp)</td>
<td>2007</td>
<td>1</td>
</tr>
<tr>
<td>Substance P_{CONH2} RPKPQQFFGLM</td>
<td>2, 36%</td>
<td>0.007%</td>
<td>0.06%</td>
<td>0.25%</td>
<td>No toxic</td>
<td>Helix</td>
<td>Human</td>
<td>2002</td>
<td>2</td>
</tr>
<tr>
<td>Bradykinin_{CONH2} LYENKPRRPYL</td>
<td>2, 25%</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.25%</td>
<td>No toxic</td>
<td>Helix</td>
<td>Human</td>
<td>2002</td>
<td>2</td>
</tr>
<tr>
<td>Myxinidin_{CONH2} GIHDILKYGKPS</td>
<td>2, 25%</td>
<td>2 µg/mL</td>
<td>2 µg/mL</td>
<td>Active</td>
<td>No toxic</td>
<td>Unknown</td>
<td>Epidermal mucus, <em>Myxine glutinosa</em> L. (Fish)</td>
<td>2009</td>
<td>1</td>
</tr>
<tr>
<td>Protonectin_{CONH2} ILGTLILGLLKGL</td>
<td>1, 58%</td>
<td>10.33 µM</td>
<td>20.66 µM</td>
<td>N.A.</td>
<td>No toxic</td>
<td>Unknown</td>
<td>Venom <em>Agelaia pallipes</em> (Wasp)</td>
<td>2004</td>
<td>5</td>
</tr>
<tr>
<td>Temporin-1PRb_{CONH2} ILPILGNLLNSLL</td>
<td>0, 61%</td>
<td>&gt;100 µM</td>
<td>&gt;100 µM</td>
<td>&gt;100 µM</td>
<td>No toxic</td>
<td>Unknown</td>
<td>The Hokkaido frog, <em>Rana pirica</em>, Asia</td>
<td>2004</td>
<td>151</td>
</tr>
</tbody>
</table>
II. De novo design of antimicrobial peptides: Results and Discussion

Figure II.1 Lipid vesicle-rupturing assay results of eligible templates in (A) POPC and (B) POPE: POPG (7:3) vesicles (Ø = 0.1 µm) vesicles. The superimposed scatter plots are represented as the mean ± standard deviation (SD) (n=3) for different tested concentrations. The lytic effect of 0.1% Triton X-100 was defined as 100%.

For the sake of investigating SAR for these two templates, we synthesized a series of their derivatives by mutating amino acids at the certain positions of the sequences to their most similar amino acids in terms of physicochemical properties.\footnote{196} Besides, for the sake of investigating the hypothesis of potentially conserved residue patterns for membrane-lytic peptides, we synthesized and tested twenty randomly scrambled derivatives of the Protonectin and Decoralin templates.

Hilpert \textit{et al.} scrambled Bac2A peptides for investigating the flexibility of amino acid arrangements for creating active AMPs. Sequence scrambling of Bac2A led to activities ranging from superior or equivalent to Bac2A to inactive, indicating that good activity was not only dependent on the composition of amino acids or the overall charge or hydrophobicity, but rather on particular linear sequence patterns. Moreover these authors used the scrambled sequences as templates for the optimization process.\footnote{198} The study indicated the possible preference of conserved residue patterns for activity of AMPs, however the authors failed to build a robust QSAR model for \textit{de novo} design AMPs.
4.2 Characterization of Protonectin and Decoralin and their derivatives

Overall, Protonectin and Decoralin showed stronger membrane-lytic activity on POPC vesicles than on PEPG vesicles, and both peptides bearing C-terminal amidation exhibited greater activity than peptides carrying a free C-terminus, which is in accordance with previous studies on cationic helical AMPs (Table II.2). Deletion of either C-terminal or N-terminal amino acids abolished or strongly reduced the membrane-lytic activity of the two natural AMPs. For Decoralin, deletion of the N-terminus had a much stronger effect on the activity on POPC vesicles than deleting the C-terminus. We also observed modified membrane-lytic activities when we conservatively replaced C-terminal residues (Table II.2). The C-terminal sections of several AMPs, *e.g.* from the marine AMP Pardaxin, have actually been shown to be particularly critical for membrane preference and insertion. Some of these C-terminal regions adapt α-helical conformation only in the presence of negatively charged lipids. Apparently, even subtle differences of hydrophobicity and the positioning of charged residues critically affect membrane insertion.

To investigate the influence of potentially conserved residue patterns, we synthesized and tested twenty randomly scrambled derivatives of the Protonectin and Decoralin templates. 45% (9/20) of the scrambled sequences exhibited sustained activity in the vesicle rupture assay but showed different membrane selectivity than the respective mother peptide. It is noteworthy that the majority of the scrambled peptides exhibited deviating activity against Gram-positive and Gram-negative pathogens compared to the templates (Table II.2), and presented diverse hemolytic potential. These results indicate that the mechanisms leading to actual antimicrobial or hemolytic activity are more complex than lipid vesicle rupture, which is in line with a recent study by Wimley and coworkers. Importantly, our analysis suggests that antimicrobial activity may be anticipated for a peptide with an appropriate overall amino
acid composition, which is perfectly in line with earlier reports. Residue permutation of naturally occurring membrane-interacting peptides may thus be considered a viable strategy for probing the relevance of residue patterns and structural integrity for general membrane activity, keeping in mind that differences in membrane lipid compositions, e.g. between Gram-positive and Gram-negative bacteria and eukaryotic cell lines, will most likely not be rationalized by this concept.
II. *De novo* design of antimicrobial peptides: Results and Discussion

(Table II.2 extends over 2 pages)

**Table II.2** Assay results and properties of Protonectin, Decoralin, their variations (prefix P, D), scrambled derivatives (PS, DS), and the *de novo* designed peptides (SOM). Bacterial growth inhibition was performed with 100 µg/ml peptide.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sequence</th>
<th>Vesicle Ø / µm</th>
<th>Peptide concentration 100 µM</th>
<th>Peptide concentration 10 µM</th>
<th>Peptide concentration 1 µM</th>
<th>MHC / µM</th>
<th>E.coli (GFP)</th>
<th>S.aureus (GFP)</th>
<th>µH</th>
<th>Fold inducible helix content</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProtonectinNH2</td>
<td>ILGTILGLLKG</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>ProtonectinOH</td>
<td>ILGTILGLLKG</td>
<td>0.1</td>
<td>+</td>
<td>+++</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>inactive</td>
</tr>
<tr>
<td>P1</td>
<td>ILGTILGLK-</td>
<td>0.1</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>P2</td>
<td>-LGTILGLKGL</td>
<td>0.1</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>P3</td>
<td>ILGTILGLKG</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>P4</td>
<td>ILGTILGLKSL</td>
<td>0.1</td>
<td>+</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
<td>-</td>
<td>+</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>PS1</td>
<td>LGGILLTLKGI</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>PS2</td>
<td>LGLILGLGTKLI</td>
<td>0.1</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>PS3</td>
<td>GKLTLILGLILL</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>++</td>
<td>-</td>
<td>10</td>
<td>+</td>
</tr>
<tr>
<td>PS4</td>
<td>ILLGGKLTGLI</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>PS5</td>
<td>TLILGLGLGILG</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>inactive</td>
</tr>
<tr>
<td>PS6</td>
<td>KITLLGGLLIG</td>
<td>0.1</td>
<td>+++</td>
<td>+</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>+</td>
</tr>
<tr>
<td>PS7</td>
<td>ITGLLLLGGILGK</td>
<td>0.1</td>
<td>+++</td>
<td>+</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>+</td>
</tr>
<tr>
<td>PS8</td>
<td>LILGTLGGLKL</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>PS9</td>
<td>TGGLTLGLLKL</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>+</td>
</tr>
<tr>
<td>PS10</td>
<td>KGILGLLGLTGL</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>+</td>
</tr>
<tr>
<td>Peptide</td>
<td>Sequence</td>
<td>MHC</td>
<td>GFP</td>
<td>MHC</td>
<td>GFP</td>
<td>n.d.</td>
<td>n.d.</td>
<td>Result</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------</td>
<td>----------------</td>
<td>-------</td>
<td>---------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>-----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decoralin NH2</td>
<td>SLLSLIRKLIT NH2</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>+++</td>
<td>10   +   +   0.45 6.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decoralin OH</td>
<td>SLLSLIRKLIT OH</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100  -   +   0.45 3.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>SLLSLIRKLI</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>+++</td>
<td>-</td>
<td>10   +   +   0.54 4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>-LLSLIRKLIT</td>
<td>0.1</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>100  -   +   0.51 4.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3</td>
<td>SLLSLIRKLIS</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>+++</td>
<td>-</td>
<td>10   +   +   0.45 4.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D4</td>
<td>SLLSLIRKLVT</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>+++</td>
<td>+</td>
<td>100  +   +   0.45 4.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS1</td>
<td>LLLSKIRSLIT</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100  -   -   0.32 3.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS2</td>
<td>SILLRISLTK</td>
<td>0.1</td>
<td>+++</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100  +   +   0.29 4.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS3</td>
<td>ILSSLIRKLLT</td>
<td>0.1</td>
<td>+</td>
<td>+++</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>100  -   +   0.53 5.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS4</td>
<td>ILRSTKSLIL</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100  -   -   0.38 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS5</td>
<td>ISTRKLLSILL</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>10   +   +   0.52 4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS6</td>
<td>TLILRLSSKLI</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>inactive        -   -   0.47 3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS7</td>
<td>LILLSLKSTIR</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>inactive        +   +   0.28 2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS8</td>
<td>LLSLILRTIKS</td>
<td>0.1</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100  -   +   0.51 4.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS9</td>
<td>LLIRSTIKSLL</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100  -   -   0.41 2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS10</td>
<td>SLIRKTSILL</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>inactive        -   -   0.34 2.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOM8</td>
<td>RPVINLACKLW</td>
<td>0.1</td>
<td>+</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100  n.d. n.d. 0.47 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOM5</td>
<td>YLLTIILRLKM</td>
<td>0.1</td>
<td>+</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
<td>100  +   +   0.44 10.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOM6</td>
<td>KKLCLLIINWL</td>
<td>0.1</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>10   +   +   0.40 11.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MHC: minimal hemolytic concentration; GFP: green fluorescent protein; n.d.: not determined. ++++, ++, + in the LUV assays represent peptide-induced membrane lytic activity > 75%, 50-75%, and 25-50% compared to triton treatment. + in the bacterial growth inhibition assays represent peptides that are active in 100 µg/ml concentration.
4.3 Secondary structure of peptides and peptide-membrane interaction

Using CD spectroscopy we estimated the secondary structure content of the peptides in water and in 50% tetrafluoroethylene (TFE). We observed a relationship between the TFE-inducible helix content (coil to helix transition) and the peptides' activities in vesicle and bacterial growth assays (Table II.2). For Protonectin and its derivatives with activity against *Escherichia (E.) coli* and *Staphylococcus (S.) aureus* the average inducible helix content was 3.0-fold for active and 2.4-fold for inactive peptides. This trend is also prominent for Decoralin and derivatives, (4.6-fold vs. 3.5-fold). With only 0.9-fold inducible helix content sequence PS5 lacked direct membrane-lytic activity, as expected, but still showed antibacterial activity (Table II.2), which might be attributed to a different mechanism or cellular target.

From the results of our CD measurements we conclude that membrane-rupturing activity of a peptide seems to be more likely with a higher inducible helix content. This finding is in perfect agreement with previous studies, in particular with work by Bechinger and coworkers who observed helix formation and orientation of some AMPs upon contact with lipid bilayers by solid-state NMR spectroscopy. Other biophysical studies confirmed differential interaction of amphipathic AMPs with zwitterionic versus anionic lipids. Using deep-UV Raman spectroscopy Cooley and coworkers provided first evidence of lipid-dependent secondary structure (helix) formation and aggregation of ME1 peptide in the presence of micelles. Recently, Torrent *et al.* observed interdependencies between aggregation-promoting, amyloid-like properties and antimicrobial activity, with positively charged residues playing a function-determining feature of AMPs. The authors concluded that, apparently, antimicrobial activity of amphipathic peptides depends on the contact with an appropriately lipophilic environment. Following this hypothesis, we tested the influence of
computed peptide amphiphilicity, expressed as the hydrophobic moment $\mu_H$ of an energy-minimized helical structure, with their measured inducible helix content (Table II.2). Of note, we did not observe strong correlation ($r = 0.36$, $r^2 = 0.13$). This finding might relativize the usefulness of the hydrophobic moment as a decisive molecular descriptor for modeling quantitative structure-activity relationships for AMPs, and certainly motivates future descriptor development.

### 4.4 Self-organizing maps for the de novo design of AMPs

For de novo design of AMPs, we adopted the Self-Organizing Map (SOM) that has found wide acceptance and successful application in small molecules hit finding and lead discovery. We first computationally generated 10,000 11-mer peptides with a balanced amino acid frequency (5% chance for each residue). For further computational analysis, we represented all 11-mers from our study (sequences presented in Table II.2) and the virtually generated sequences as real-valued vectors using a cross-correlation pharmacophore feature approach (PepCATS descriptor). We chose this molecular representation because it exhaustively captures residue neighborhoods, which was shown to be important for the activity of some AMPs with direct membrane lytic activity. We had previously successfully employed this approach for classifying membrane targeting peptides and designing MHC-I binding peptides. Briefly, this method results in a sequence-length independent, numerical molecular representation of the pharmacophoric feature distribution (hydrogen-bonding, lipophilic, aromatic, charged interaction potential) along the amino acid sequence. Then we employed the unsupervised self-organizing map (SOM) algorithm for clustering the peptides according to these properties, and obtained a two-dimensional map for visual inspection (Figure II.2). The peptide density was approximately equally distributed.
over the map without revealing empty areas devoid of sequence data (Figure II.2A). Each cluster may be considered as a local feature model of peptide sequence space. Peptides within a cluster are more similar to each other than to any other peptide in the training set. We observed a distinct area on the SOM (cluster 4/11), which contained five bacterial membrane-active sequences from the training set including Decoralin, D3, D4 and DS2. It also contained one membrane-inactive peptides DS8. (Figure II.2B) We specifically chose the peptide closest to the cluster centroid (RS1), and peptides from the vicinity of the known actives.

To our surprise, none of seven selected peptides possessed membrane-lytic activity. Though SOM is expected to cluster peptides with similar property according to the representation of descriptors, however in terms of sequence space, 10,000 peptides only represent a very little portion of the total sequence space \(20^{11}\) 11-mer peptides). SOM is therefore forced to cluster membrane-active peptides in the training set with in fact dissimilar virtual generated peptides. This finding also points out that a more focused peptide library is required to this end.

**Figure II.2** Two-dimensional self-organizing map with \(20 \times 15\) clusters (Voronoi fields). (A) Overall peptide density in each cluster; neurons in darker red contained more peptides. (B) Location of the seven synthesized peptides (RS1-RS7), cluster (4/11)). White fields represent empty clusters.
Motivated by the high number of membrane-lytic peptides that were successfully obtained from scrambling the Protonectin and Decoralin sequences, we computationally generated 10,000 11-mer peptides with a residue frequency that mimics the amino acid composition of the template AMPs (cf. histogram in Figure II.3). The overall residue bias was the only design constraint; we did not apply additional design patterns (e.g. positional preferences) during \textit{ab initio} sequence generation. The most densely populated "activity island" (cluster 11/8) contained a total of 49 peptides including Decoralin and its derivatives D3, D4, and DS3. Apparently, the pharmacophoric features of these sequences share a common pattern that allowed for an approximate distinction between active and inactive peptides. From this set we selected, synthesized and tested eight of the computer-generated sequences (Figure II.3B; peptides labeled SOM1-SOM8). Because the sequence generation was guided by the amino acid composition of the template AMPs, there is an expected prevalence of isoleucine, leucine, and positively charged residues in these peptides. Of note, the selected designer peptides completely lack negatively charged amino acids, which emerged as a feature during SOM training. We again specifically chose the peptide closest to the cluster centroid (SOM1), and peptides from the vicinity of the known actives.
II. De novo design of antimicrobial peptides: Results and Discussion

Figure II.3 Two-dimensional self-organizing map with 20×15 clusters (Voronoi fields). (A) Overall peptide density in each cluster (left), and residue frequency used for peptide design (right); (B) location of the eight synthesized membrane-rupturing peptides (SOM1-SOM8, cluster (11/8)), which also contained the most active designer peptide (SOM6) among other active training peptides. White fields represent empty clusters.

4.5 Characterization of SOMs designed peptides

For one of the automatically designed sequences (SOM8) we observed medium and for two peptides (SOM5, SOM6) strong membrane-lytic effects, with SOM6 exhibiting low nanomolar activity (EC$_{50}$ = 59±1 nM for PEPG; 7±2 nM for POPC vesicles) (Figure II.4, Table II.3). The designer peptides SOM5 and SOM6 possess a ten-fold inducible helix content, as determined by CD spectroscopy, which correlates with their strong membrane-lytic activity on vesicles and exceeds the value measured for the natural templates Decoralin (6.1-fold) and Protonectin (2.8-fold). In bacterial growth assays, SOM5 and SOM6 exhibited strong bacteriostatic effects against E. coli (Figure II.4C) and S. aureus (Figure II.4D). Of note, while SOM6 showed mild hemolytic activity against mouse red blood cells (minimal hemolytic concentration, MHC ≥ 10 µM), SOM5 lacked hemolytic activity up to a concentration of 100 µM (Table II.2).
Table II.3 Membrane-lytic activity of the design templates and the de novo designed peptides \((mean±stderr; n = 3)\). Lipid vesicles \(Ø = 0.1 \mu m\).

<table>
<thead>
<tr>
<th>ID</th>
<th>EC(_{50}) / µM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PEPG</td>
</tr>
<tr>
<td>Protonectin(_{NH2})</td>
<td>6.8±1.3</td>
</tr>
<tr>
<td>Decoralin(_{NH2})</td>
<td>7.4±1.4</td>
</tr>
<tr>
<td>DS2</td>
<td>17±1.2</td>
</tr>
<tr>
<td>SOM6</td>
<td>0.059±0.001</td>
</tr>
</tbody>
</table>

Figure II.4 PEPG (A) and POPC (B) vesicle \((Ø = 0.1 \mu m)\) ruptured by SOM6 peptide \((mean±stderr; n = 3\) technical replicates). The lytic effect of 0.1% Triton X-100 was defined as 100%. The antibacterial effect of peptides SOM5 and SOM6 (100 µg/ml) are shown in (C) and (D) \((mean±stddev; n ≥ 3\) independent experiments). The bacterial strains used in the growth experiments were GFP-transformed, and GFP fluorescence intensity served as a measure of cell density.
Direct peptide-membrane binding

We used SPR to assess the direct interaction between peptides (Protonectin, Decoralin, DS2, SOM6) and lipid membranes. The sensorgrams reveal that signal intensity (peptide binding) increases as a function of the peptides' concentration (Figure II.5). Evidently, all four tested peptides directly interacted with the membrane as a prerequisite for their activity seen in the vesicle rupture assay. The membrane binding measured for Protonectin and Decoralin reached similar levels on immobilized PEPC and POPC vesicles implying nonselective activity of these two AMPs toward zwitterionic and negatively charged membranes. Of note, the sensorgrams for SOM6 suggest slow off-rates for this peptide (Figure II.5D), which is not observable for Protonectin and Decoralin (Figure II.5A, B). Scrambled Decoralin derivative DS2 features an intermediate behavior (Figure II.5C). To exclude artifacts caused by improper vesicle immobilization, we probed the membrane integrity in the SPR experiments by non-contact AFM to visualize the surface structure of vesicle-coated SPR chips. The resulting image clearly shows that the sensor surface bears intact LUVs with only isolated patches of immobilized planar bilayers (Figure II.6). One may thus conclude that both SPR and the vesicle rupture assay probe peptide interaction with intact vesicles. The AFM measurement is a strong argument for allowing direct comparison of results between the two techniques. We further conclude that the designed peptides DS2 and specifically SOM6 might actually exert their activity by a different mechanism of action than the other AMPs tested in our study.
II. *De novo* design of antimicrobial peptides: Results and Discussion

**Figure II.5** Results of the surface plasmon resonance experiments with Protonectin (A), Decoralin (B), DS2 (C), and SOM6 (D). Sensor surfaces were coated with either POPC (upper panels) or PEPG (lower panels) lipid vesicles ($\Omega = 0.1 \mu m$).

**Figure II.6** (A) AFM topographic image of an HPA SPR sensor chip with immobilized POPC vesicles. (B) Magnified 3D phase image of a selected region of the sensor surface (upper left quadrant in (A)). (C) Cross section measurement along the white line shown in (A).
**Thermodynamic of peptide-vesicle interaction**

Consequently, we investigated the thermodynamics of peptide-vesicle interaction by ITC. For this purpose we followed an established experimental procedure with 10 mM vesicles as injectant and 100 µM peptide solution in the measuring chamber of the calorimeter. Results of a preliminary study with varying concentrations of Decoralin (5, 50, 100 µM) confirmed this setting to ensure a satisfactory signal-to-noise ratio. For Protonectin and Decoralin we observed a concentration-dependent decrease of heat rates per injection (Figure II.7A, B), which is in perfect agreement with earlier ITC results for cationic helical AMPs. The measured binding enthalpy difference ($\Delta H$) and derived entropy change ($T\Delta S$) clearly indicate an entropically driven interaction between these peptides and the lipid vesicles. Their $k_D$ values confirm their preference for POPC vesicles observed in the vesicle assay (Table II.4). The measured binding enthalpy difference ($\Delta H$) and derived entropy change ($T\Delta S$) clearly indicate an entropically driven interaction between these peptides and the lipid vesicles. $\Delta H$ is virtually zero for DS2 and SOM6 (Figure II.7C, D), so that it is safe to conclude that these peptides are entropy-driven membrane ligands. For now, it remains unknown if the observed differences reflect a different mode of action. The ITC traces show that the measured absolute heat rates of cationic helical AMPs tend to be very small (0.3–4 µJ/s; Figure II.7A, B) or undetectable (Figure II.7C, D). One may speculate whether the pronounced membrane activity of SOM6 might be owed to its N-terminal twin-lysine motif, which was obtained by chance during random peptide generation. Futaki and coworkers recently demonstrated that substituting lysine by arginine in the model AMP [KLAKLAK]$_2$ significantly increased its membrane permeation. The mutated sequence [RLARLAR]$_2$ did no longer posses membrane-lytic activity but efficiently translocated across the membrane into the cytoplasm. In fact, the observation of a slow off-rate of SOM6 (Figure II.5D) compared to Protonectin
and Decoralin (Figure II.5A, B) is in agreement with the known effect of arginine-lysine substitutions on the membrane residence times of cell-penetrating peptides. Diminished membrane translocation of lysine-rich AMPs has been attributed to the weaker ionic interaction of the ammonium (lysine) than the guanidinium (arginine) side chain with the lipid headgroups, which might actually explain the ITC trace recorded for SOM6.

**Figure II.7** ITC measurements of peptide-vesicle (Ø = 0.1 μm) interaction for Protonectin (A), Decoralin (B), and the designed peptides DS2 (C) and SOM6 (D). Vesicles (10 mM) were titrated to 100 μM peptide solution.

The first injections were omitted from data analysis.
III. De novo design of antimicrobial peptides: Conclusions

Table II.4 Thermodynamic and kinetic values for the interaction of Protonectin and Decoralin with lipid vesicles (Ø = 0.1 µm) obtained by isothermal titration calorimetry. $T \Delta S$ given for 298 K; $n$ is the stoichiometric factor.

<table>
<thead>
<tr>
<th>Peptide+Lipid</th>
<th>$\Delta H$ (kJ/mol)</th>
<th>$T \Delta S$ (kJ/mol)</th>
<th>$k_A$ (mol$^{-1} \times 10^3$)</th>
<th>$k_D$ (µM)</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protonectin+PEPG</td>
<td>19±19</td>
<td>43±17</td>
<td>27±28</td>
<td>88±70</td>
<td>2±1</td>
</tr>
<tr>
<td>Protonectin+POPC</td>
<td>4±2</td>
<td>32±2</td>
<td>100±145</td>
<td>20±24</td>
<td>3±1</td>
</tr>
<tr>
<td>Decoralin+PEPG</td>
<td>4±3</td>
<td>28±1</td>
<td>15±10</td>
<td>98±69</td>
<td>3±1</td>
</tr>
<tr>
<td>Decoralin+POPC</td>
<td>1±0</td>
<td>32±1</td>
<td>380±123</td>
<td>3±1</td>
<td>4±1</td>
</tr>
</tbody>
</table>

5. Conclusions

The traditional approach to sequence-based computational peptide design requires a large set of training examples to develop a reliable quantitative structure-activity relationship (QSAR) model in a supervised fashion. Pioneered in the 1990s$^{122,171}$ this concept has recently been applied to generating novel AMPs$^{137,169,174}$ and MHC-I ligands$^{127}$. Supervised machine-learning requires a large set of training examples to develop a reliable QSAR model. In fact, many AMPs are readily available from public databases,$^{197,224,225}$ but unless their targets and modes of action are annotated and individual models are generated for each subclass, there is a risk of obtaining poor predictors. Here we have validated the unsupervised, data-driven SOM technique for de novo peptide design. The SOM algorithm works with large and small sets of training data and has been widely applied to amino acid sequence analysis.$^{195,226}$ We proved our first hypothesis and corroborated literatures that there was conservative sequence context of membrane-active peptides. The second hypothesis held true as well that we used only 11 active sequences during SOM training, which turned out to be
sufficient and allowed us to prioritize and select novel peptides with AMP-like properties. In principle, the method also works with a single known active template, as the SOM algorithm identifies local clusters of sequences that share mutual features with the template without modeling an explicit mathematical QSAR function. Picking candidate peptides from these local applicability domains represents a valuable complementary approach to full-fledged supervised model building, specifically in the absence of a large set of training data. In combination with residue-frequency guided sequence generation, we immediately obtained peptides with the desired nanomolar activity. Our computational approach is broadly applicable to any sequence-based peptide design task, without requiring large training data sets and running the risk of working with over-trained, non-predictive QSAR models.
III. Sequence-based multi-objective optimization of anticancer peptides

1. Aim

Anticancer peptides (ACPs) represent a promising class of anticancer agents. Although ACPs share molecular characterization and modes of action similar to AMPs, not all of the AMPs are ACPs, and the factors that determine the selectivity of these peptides are not fully understood. In addition, less than 200 peptides are annotated as ACPs in the APD2 database, making it difficult to build up reliable SAR models for either de novo design or optimization purposes. Under this circumstance, we only considered the pharmacological mode of actions of the peptides and the property we wanted to optimize. Our aims are as follows:

- Utilizing computational approaches to optimize ACPs with better activity and selectivity.
- Combining computational and biochemistry experiments for heuristic peptide design.
- Setting up pharmacological experiments in-house to characterize all designed peptides.

2. Hypotheses

Hypothesis-1

- The activity of ACPs can be improved via enhancing their cell-penetrating activity.

**Background:** AMPs consisting of less than 18 amino acids are too short to span across the membrane and to form pores. Therefore, the primary mechanisms of action of these short AMPs are as follows: 1) the carpet model and/or 2) intracellular cytotoxicity. The peptides with the latter mechanism of action can penetrate target cells without lysing the membrane.
ACPs are believed to present mechanisms of action similar to AMPs; however, unknown factors determine their specificity. As we discussed in the Introduction, peptides follow the carpet model only when the critical peptide concentration is reached, allowing membrane disruption in a detergent-like fashion. Consequently, we created the hypothesis of improving ACP activity by conferring cell-penetrating ability at low concentrations so that the ACPs are able to penetrate cells and trigger their inner cytotoxicity before reaching the critical membrane-lytic concentration on the membrane.

Cell-penetrating peptides (CPPs), which are a group of cationic peptides, are able to penetrate the cell membrane at low concentrations via either an energy-dependent or -independent pathway without using any receptors or permeabilizing membranes. Endocytosis, which is an energy-dependent pathway, consists of several pathways requiring protein or receptor assistance for membrane invagination, whereas direct penetration is an energy-independent pathway involving the transient destabilization of the membrane, allowing peptides to cross the membrane directly into the cytoplasm and nucleus. CPPs can be amphipathic and short, with high cationic amino acid contents, and their structural versatility and the interaction force when interacting with lipid membranes are crucial for their cell-penetrating ability. In this regard, it is believed that these two classes of peptides, AMPs (or ACPs) and CPPs, cannot be differentiated because of their structure and sequence but because of their different mechanisms of action in penetrating cells and different applications. Indeed, several studies have demonstrated that well-known CPPs (e.g., TAT peptides) show potent antimicrobial activity, whereas AMPs, such as LL-37, have potential uses in drug delivery. Although these two classes of peptides overlap somewhat in their properties, modifying the sequence of the peptide can magnify the cell-penetrating or antimicrobial effect.
III. Sequence-based multi-objective optimization of anticancer peptides: Aim and Hypotheses

In this study, we aimed to modify the sequence of an ACP template to enhance its cell-penetrating activity. We hypothesized that once modified, the designed peptides could penetrate cancer cells at concentrations lower than the critical concentration for lysing the membrane. The peptides could exert their potential inner cytotoxicity by preferentially targeting the mitochondrion, another organelle with a net negatively charged membrane, and induce cell apoptosis. Coupling Magainin II, which is a pore-forming peptide, with a known CPP, Penetratin, suggests that the anticancer activity may be improved by 30-fold, further supporting our hypothesis. Instead of the coupling strategy yielding a high mass peptide, we aimed to design a short peptide with these features using a computational approach combined with biochemistry experiments.

Hypothesis-2

- Conferring a tumor-homing character on ACPs can improve their selectivity.

**Background:** Most ACPs are cationic peptides that can specifically target cancer cell membranes with a net negative charge. However, not all ACPs with antitumor activity are sufficiently selective against cancer cells. A systemic analysis of a database comprising 158 amphipathic α-helix peptides with annotated activities against cancer cells and normal cells revealed that net positive charge, hydrophobicity and hydrophobic moments were insufficient for use as selectivity predictors. Such an observation suggests the requirement for a more detailed knowledge of their structure-activity relationship to design selective and potent ACPs.

Since phage display technology was first reported in 1985, numerous peptides specifically targeting tumors or its vasculatures have been discovered. These tumor-homing peptides (THPs) have approximately 10 residues and can either transport their cargo to the cell surface
without internalization or have the ability to penetrate into cells. THPs have a relatively higher average composition of arginine and cysteine, and the abundance of the former amino acid results from the presence of common motifs, such as RGD and NGR, whereas the abundance of the latter amino acid is due to many cyclic peptides. The RGD (Arg-Gly-Asp) and NGR (Asn-Gly-Arg) motifs are among the most studied of tumor-homing peptides. Peptides with these two motifs home to tumor blood vessels and appear to be independent of the tumor types. The RGD peptides have high affinity toward αv integrins in the angiogenic vasculature, whereas NGR peptides can recognize Aminopeptidase N (APN), which is overexpressed in angiogenic blood vessels. Although identifying specific homing peptides is straightforward, identifying their receptors remain a hurdle in this field.

THPs have recently drawn much attention and have been used as a cancer drug delivery vehicle. The RGD tumor-homing sequence has been conjugated to doxorubicin and inhibits tumor growth while prolonging the lifespan of tumor-bearing animals. Several THPs have also been conjugated to ACPs, and these chimeric peptides have been shown to selectively induce apoptosis in cultured cancer cells and to inhibit tumor growth in mice. In this study, instead of traditional conjugation methods, we adopted a new strategy to improve ACP selectivity. We aimed to employ computational approaches to build THP features onto an ACP template. We proposed that our designed peptides with ACP and THP features would have better therapeutic indices than the templates.
3. Materials and Methods

Computing Support Vector Machines. A SVM classifies data by determining the best hyperplane that separates all data points of one class from those data points of the other class. The best hyperplane for a SVM is defined as the one with the maximum optimal margin between two classes. Data points with the smallest margins are exactly the ones closest to the decision boundary, and these points are called the support vectors. The number of support vectors can be much smaller than the size of the training set. Consider the computing SVM for ACP as shown in the figure below:

We collect known ACPs as positive training sets, and the other peptides classes, AMPs, CPPs and THPs, construct the negative training sets. Intuitively, suppose that given a training set, we manage to determine a decision boundary that all data points, except support vectors, are far from so that we would be confident to make correct predictions.

Consider that the training data are a set of vectors \( x_i (x \in \mathbb{R}^n) \) and that we use parameter \( w, b \) to define the classifier:

\[
h_{w,b}(x) = g(w^T x + b),
\]

where \( w \in \mathbb{R}^n \), \( w^T x \) is the inner dot product of \( w \) and \( x \), \( b \) is real number, and \( h(x) \) represents the predicted results. In this situation, \( g(z) = 1 \) if \( z \geq 1 \), and \( g(z) = -1 \) otherwise,
III. Sequence-based multi-objective optimization of anticancer peptides: Materials and Methods

representing two different classes of peptides.

We define the functional margin of \((w, b)\) with respect to a training example \((x^{(i)}, y^{(i)})\) to be:

\[
\hat{\gamma}^{(i)} = y^{(i)}(w^T x^{(i)} + b)
\]  

(7)

Suppose we are given a training set that is linearly separable. We want to maximize \(\hat{\gamma}\), subject to each training example having a functional margin of at least \(\hat{\gamma}\), that is:

\[
\max_{\gamma \geq b} \frac{\gamma}{\|w\|}
\]

subject to:

\[
y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, i = 1, \ldots, m
\]

(8)

In mathematics, we can add an arbitrary scaling constraint on \(w\) and \(b\) without changing the prediction results for peptide classes; thus, we introduce the scaling constraint that the functional margin of \((w, b)\) with respect to the training set must equal 1:

\[
\hat{\gamma} = 1
\]

(9)

Therefore, we now maximize \(\frac{1}{\|w\|}\) that is the same thing as minimizing \(\|w\|^2\). We now have the following optimization problem:

\[
\min_{\gamma \geq b} \frac{1}{2} \|w\|^2
\]

subject to:

\[
y^{(i)}(w^T x^{(i)} + b) \geq 1, i = 1, \ldots, m
\]

(10)

For a simpler computational calculation, we construct the Lagrangian algorithm for the optimization problem:

\[
L_p(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i \left[ y^{(i)}(w^T x^{(i)} + b) - 1 \right],
\]

(11)

where \(\alpha_i\) is the Lagrange multiplier, and \(\alpha_i \geq 0\). To find the dual form of the problem, we first minimize \(L_p(w,b,\alpha)\) with respect to \(w\) and \(b\), which we will do by setting the derivatives of \(L_p\) with respect to \(w\), and \(b\), to zero. We have:
III. Sequence-based multi-objective optimization of anticancer peptides: Materials and Methods

$$\nabla_w L_p(w, b, \alpha) = w - \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} = 0$$

We obtain:

$$w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} = 0$$

The derivative of $L_p$ with respect to $b$,

$$\frac{\partial}{\partial b} L_p(w, b, \alpha) = -\sum_{i=1}^{m} \alpha_i y^{(i)} = 0$$

Therefore, we obtain:

$$L_p(w, b, \alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle,$$

where we maximize over $\alpha_i \geq 0$. Once we find the optimal $\alpha_i$, then we can use the above equation to obtain $w$ and $b$. Notably, from the KKT dual complementary condition, we will have $\alpha_i > 0$ only when the functional margin is equal one. Support vectors are points with the smallest margins closest to the decision boundary.

To make the algorithm work for non-linearly separable datasets and less sensitive to outliers, SVM can use a soft margin, using a hyperplane that separates many but not all data points:

$$\min_{w \neq b} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{m} \xi_i$$

subject to:

$$y^{(i)} (w^T x^{(i)} + b) \geq 1 - \xi_i, i = 1, \ldots, m$$

$$\xi_i \geq 0, i = 1, \ldots, m.$$

The parameter $C$ controls the relative weighting between the two goals of making the $w$ large and ensuring that most examples have a functional margin of at least one.\textsuperscript{249,250}

In our study, we collected peptides with or less than 16 amino acid residues from databases for further computational analysis. We collected peptides annotated with anticancer activities from APD2\textsuperscript{197} and tested scrambled derivatives of Decoralin,\textsuperscript{251} and glycine-scan derivatives.
of Decoralin and Citropin (GLFDVIKKVASVIGGL) composing ACP-positive datasets (Table III.1).

Table III.1 In-house tested sequences as ACP positive training sets. Peptides were tested against MCF7 cells and demonstrated at least 75% inhibition at 100 µM after 24 h incubation.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protonectin</td>
<td>ILGTILGLLKGL</td>
</tr>
<tr>
<td>P3</td>
<td>ILGTILGLLKGI</td>
</tr>
<tr>
<td>P4</td>
<td>ILGTILGLLKLSL</td>
</tr>
<tr>
<td>PS10</td>
<td>KGILGLLLITGIL</td>
</tr>
<tr>
<td>D1</td>
<td>SLLSLIRKLI</td>
</tr>
<tr>
<td>D2</td>
<td>LLLSRIRKLIT</td>
</tr>
<tr>
<td>D3</td>
<td>SLLSLIRKLIS</td>
</tr>
<tr>
<td>D4</td>
<td>SLLSLIRKLVT</td>
</tr>
<tr>
<td>DS1</td>
<td>LLLSKIRSLIT</td>
</tr>
<tr>
<td>DS3</td>
<td>ILLSKIKRLLT</td>
</tr>
<tr>
<td>DS6</td>
<td>TLLRLSSKLI</td>
</tr>
<tr>
<td>Citropin-A1</td>
<td>ALFDVIKKVASVIGGL</td>
</tr>
<tr>
<td>Citropin-G4</td>
<td>GLFGVIKKVASVIGGL</td>
</tr>
<tr>
<td>Citropin-G11</td>
<td>GLFDVIKKVAGVIGGL</td>
</tr>
<tr>
<td>Citropin-A14</td>
<td>GLFDVIKKVASVIAGL</td>
</tr>
<tr>
<td>Citropin-A15</td>
<td>GLFDVIKKVASVIGAL</td>
</tr>
</tbody>
</table>

We additionally collected AMPs from APD2 as a training set. CPPs and THPs within the
same lengths were collected from these databases.\textsuperscript{243,252} For constructing respective SVMs, we used respective peptides as positive training sets and the three remaining groups of peptides as negative training sets. We assembled 51/1414, 481/928 and 669/740 peptides as positive and negative training sets for training ACP, CPP and THP SVMs respectively. All peptides were represented as real-valued vectors using a cross-correlation physicochemical feature approach (PPCALI descriptor).\textsuperscript{219} PPCALI is a length-invariant auto-correlated descriptors representing peptides in terms of their physicochemical properties. Each amino acid in a peptide sequence is represented by 19 real-number values obtained from a PCA performed on 143 property scales.\textsuperscript{253} We adopted cubic polynomial kernels for computed ACP, CPP and THP SVMs, which were optimized and validated via the MATLAB Statistics toolbox (MathWorks, MA, U.S.A.). For constructing the ACP2 SVM, we additionally included TAT(49-57) (RKKRRQRRR), which was tested in-house and which displayed 15% inhibition to MCF7 cells at 100 µM, and four new ACP sequences from APD2 as ACP-positive examples (LLRHVVKILEKYL, ILPIRSLIKKLL, FLPLKKLRFGGL, RLGDGCTR). We deleted one similar sequence, TAT(48-57) (GRKKRRQRRR), from negative datasets, and the Gaussian kernel was used for constructing the ACP2 SVM model.

**Peptide synthesis, labeling and analytics.** Fmoc peptide syntheses and analytics were conducted as previously mentioned. For coupling FITC at the N-terminus of peptides, first, we coupled β-alanine at the C-terminus, functioning as a linker for FITC coupling:
Cytotoxicity assays. The MCF7 cell line was kindly provided by Prof. Michael Detmar, and the A549 cell line was kindly provided by Prof. Dario Neri. LU1205 and Jurkat cells were kindly provided by Prof. Cornelia Halin. Doxorubicin and epothilone C were kindly provided by Prof. Karl-Heinz Altmann. HDMEC and its medium were purchased from PromoCell, Germany. DMEM, RPMI, F12K Nut Mix, antibiotic and glutamine were purchased from Gibco, CA, U.S.A. The cytotoxic effects of the compounds were evaluated using MTT assays\textsuperscript{254}. MCF7 (human breast adenocarcinoma cells, DMEM with 10% FBS, 1% antibiotic, and 1% glutamine), LU1205 (human epithelial melanoma cells, RPMI-1640, 10% FBS, and 1% antibiotic), Jurkat (human acute T-cell lymphoma cells, RPMI-1640, 20% FBS, and 1% antibiotic), A549 (human lung carcinoma cells, F12K Nut Mix, 10% FBS, and 1% antibiotic) and HDMEC (human dermal microvascular endothelial cells, endothelial cell media MV) lines in their complete medium were seeded in 96-well cell culture plates at a density of $1 \times 10^4$ cells per well and allowed to attach overnight at 37 $^\circ$C. Then, the cells were incubated with different compound concentrations for 24 h or 72 h (synergism experiments), and the medium containing compounds were removed. Next, medium containing 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, 0.5 mg/ml) (ABCR, Germany) were added to cells and removed after 1 h of incubation. Dimethyl sulfoxide (DMSO, Sigma-Aldrich, MO, U.S.A.) was added to dissolve the purple formazan that was reduced from MTT by mitochondria reductase in living cells, and then the absorbance at 540 nm was measured. The percentage of cell viability was compared with groups without peptide treatment.

FACS analysis of cell membrane integrity and cell-penetrating activity of peptides

The cell-penetrating activity of peptides and the cell membrane integrity after peptide
treatments were measured by FACS analysis. Cancer cells were seeded in 12-well culture plates at a density of $2 \times 10^5$ cells per well and incubated overnight at 37 °C. After incubation, for the peptide cell-penetrating activity analysis, first cells were incubated with peptides conjugated with FITC at 1 µM or 10 µM for 4 h. Then, trypsin-EDTA in medium was added to detach the cells, followed by centrifuging the cells. Supernatants were removed, and cells were washed twice with PBS. Then, cells were re-suspended with FACS buffer (PBS with 2 mM EDTA and 2% FBS) and analyzed by flow cytometry (FACSCanto, BD Biosciences, Allschwil, Switzerland).

For cell membrane integrity analysis, cells were detached and incubated with the indicated concentrations of peptides and propidium iodide (PI) (3 µM) in FACS buffer for 10 min before analysis. In total, $2 \times 10^4$ events were recorded for each sample. The data were analyzed offline using FlowJo software (Treestar, Ashland, TN, U.S.A.).

**Cell apoptosis assays.** After detaching and re-suspending the cells with 100 µl 1X Annexin V binding buffer (Life Technologies, CA, U.S.A.), cells were treated with 25 µM of peptides for 20 min, followed by 15 min staining with 5 µl of 50 µg/ml 7-AAD (BioLegend, CA, U.S.A.) and Annexin V-FITC (Life Technologies, MA, U.S.A.). Then, the samples were immediately analyzed by FACS. Annexin V can specifically bind to phosphatidylserine (PS), whose exposure on the outer leaflet of the plasma membrane is often an early event of apoptosis. 7-AAD is a cell-impermeable DNA dye. These two dyes are frequently used together to identify cells in early apoptosis.²⁵⁴

**Light and confocal microscopy analyses**

Cells grown for 24 h on slides with eight chambers (Lab-Tek® Chamber Slide™ System
(177402), PA, U.S.A.) were incubated with the indicated peptides labeled with 1 μM FITC for 4 h. After the incubation, the medium was removed, and cells were washed with PBS and then fixed with 2% paraformaldehyde for ten min. Subsequently, the chambers were removed, and cells were washed with PBS, followed by staining with 1 μg/ml Hoechst 33342 (Sigma-Aldrich, MO, U.S.A.) for 5 min. Then, the cells were washed with PBS and mounted with Mowiol 4-88 reagent (Calbiochem, Germany). For confocal analysis, the samples were stained with cell membrane dye (CellMask™ deep red plasma membrane stain, Life Technologies, CA, U.S.A.) before the fixation step. The cells were either examined on a fluorescence microscope (Zeiss Axioskop 2, Germany) or on a confocal microscope (Zeiss LSM 710-FCS, Germany), and three pictures of each sample were taken from independent areas on the sample.

**Statistical analysis**

The statistical analysis was performed using one-way or two-way ANOVA test, followed by Dunnett’s multiple comparisons test using GraphPad Prism version 6.0b for MAC OS X software (www.graphpad.com). The results are presented as the mean ± standard deviation. Difference were considered statistically significant when $p < 0.05$ in one-way ANOVA test. Confidence intervals are computed for 95% confidence and statistical significance is defined using an alpha of 0.05 in two-way ANOVA test.
4. Results and Discussion

4.1 Compute Support Vector Machine QSAR models

Computer-aided peptide design approaches have been recognized as successful tools for peptide discovery. In the present work, we adopted a supervised machine-learning approach to build QSAR models for target peptides. We collected ACPs, CPPs, and THPs (<16 AAs) from databases and represented all peptides as real-valued vectors using an autocorrelation physicochemical feature approach (PPCALI descriptor) for further computational analysis. The reason for choosing peptides with less than 16 residues as a training set in our study was motivated by the fact that ACP lengths in this range cannot span the membrane to form a pore and because the carpet model and/or inner cytotoxicity are the most common mechanisms of action for these peptides. Thus, we would be able to work on QSAR models corresponding to our hypotheses. Then, we computed a support vector machine (SVM) and built a QSAR model for each class of peptides. The SVM models were able to extract general features of a specific peptide class and to classify the tested peptides accordingly. It was first introduced in 1995 as a potential alternative to ANNs. Byvatov et al. compared the performance of SVM and ANNs on drug/nondrug classification and concluded that SVM outperformed ANNs in terms of MCC values, however these methods were shown to complement each other.

We performed 10-fold cross-validation to estimate the accuracy of the models (Table III.2). Overall, CPP SVM and THP SVM were robust prediction models, particularly when compared with the work by Gautam et al. regarding the development of SVM models for CPPs (maximum MCC=0.82), whereas ACP SVM had a relative higher variance in prediction. This poor performance may result from the lack of positive ACP examples in the training dataset and form an imbalance between positive and negative examples. However,
the high specificity of the ACP SVM model suggested a robust ACP-positive prediction.

Table III.2 Performance of SVM models in 10-fold cross-validation. The data are presented as the mean ± SD.

<table>
<thead>
<tr>
<th></th>
<th>MCC</th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACP SVM</td>
<td>0.59±0.20</td>
<td>0.94±0.04</td>
<td>0.77±0.16</td>
</tr>
<tr>
<td>CPP SVM</td>
<td>0.76±0.06</td>
<td>0.92±0.02</td>
<td>0.81±0.04</td>
</tr>
<tr>
<td>THP SVM</td>
<td>0.72±0.06</td>
<td>0.90±0.01</td>
<td>0.82±0.03</td>
</tr>
</tbody>
</table>

Tyagi et al. also developed SVM models for ACPs based on three positive training sets and created random peptides derived from a protein database as a negative training set. Although these authors utilized different descriptors to encode the peptides, these authors were also suffering from a lack of ACPs in training sets and obtained similar model performance in terms of MCC, sensitivity and specificity. To this end, these authors created a balanced negative training set and obtained SVM models with improved MCC. Nevertheless, the high specificity of our ACP SVM could confidently predict ACPs, which was sufficient for our peptide optimization purpose.

4.2 Cytotoxicity of Decoralin and TAT(48-60)

First, we tested the cytotoxicity of the chosen ACP templates Decoralin (SLLSLIRKLIT\textsubscript{NH2})\textsuperscript{258} and TAT(48-60) (GRKKRRQRRRPQ\textsubscript{NH2})\textsuperscript{259} peptides on the MCF7 (breast adenocarcinoma) cell line and HDMEC (human dermal microvascular endothelial cells) primary cell line. Decoralin\textsuperscript{258} and TAT(48-60)\textsuperscript{44} have been shown to be potent antimicrobial peptides. In this study, we demonstrated that Decoralin was also a potent ACP against MCF7 (Figure III.1a, EC\textsubscript{50} = 19±1 μM) and displayed mild cytotoxicity against
normal primary cells (Figure III.1a, EC₅₀ = 132±1 µM). The TAT peptide did not show any cytotoxicity against either MCF7 or HDMEC cell lines (Figure III.1b). This observation corroborated previous findings that not all AMPs are potent ACPs and that ACPs show basic selectivity over normal healthy cells.

![Figure III.1](image_url)  
**Figure III.1** Cytotoxicity of (a) Decoralin and (b) TAT(48-60) peptides against MCF7 and HDMEC cells (n=3). Cells were treated with different concentrations for 24 h, followed by 1 h MTT treatment. The absorbance at 540 nm was measured and the percentage of cell viability was compared with groups without peptide treatment. The data are presented as mean ± SD from triplicates.

### 4.3 Identification of critical positions on Decoralin and TAT(48-60) for their activities

Consequently, we investigated which amino acids on Decoralin and TAT(48-60) sequences were critical for their activities. For Decoralin, we measured its cytotoxicity against MCF7 cells, whereas for TAT(48-60), we probed its cell-penetrating ability. We performed a ‘glycine scan’ on these two templates by individually substituting every amino acid in the sequences for glycine. The results demonstrated that mutations at three positions (1, 4 and 11 residues from the N-terminus) in Decoralin would not alter its sensitivity to MCF7 cells (Table III.3), whereas four positions (7, 11, 12 and 13 residues from the N-terminus) in
TAT(48-60) were resistant to the glycine mutation in terms of their cell-penetrating ability in MCF7 cells (Figure III.2).

**Table III.3** Cytotoxicity of Decoralin glycine-scan derivatives against MCF7 cells. The data are presented as *mean ± SD* from triplicates. The percentage of cell viability was compared with groups without peptide treatment.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>MCF7 EC$_{50}$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decoralin</td>
<td>SLLSLIRKLIT</td>
<td>19±1</td>
</tr>
<tr>
<td>DecoG1</td>
<td>GLLSLIRKLIT</td>
<td>19±1</td>
</tr>
<tr>
<td>DecoG2</td>
<td>SGLSLIRKLIT</td>
<td>&gt;100</td>
</tr>
<tr>
<td>DecoG3</td>
<td>SLGSLIRKLIT</td>
<td>&gt;100</td>
</tr>
<tr>
<td>DecoG4</td>
<td>SLLGLIRKLIT</td>
<td>27±1</td>
</tr>
<tr>
<td>DecoG5</td>
<td>SLLSGLIRKLIT</td>
<td>&gt;100</td>
</tr>
<tr>
<td>DecoG6</td>
<td>SLLSLGKLIT</td>
<td>&gt;100</td>
</tr>
<tr>
<td>DecoG7</td>
<td>SLLSLGKLIT</td>
<td>&gt;100</td>
</tr>
<tr>
<td>DecoG8</td>
<td>SLLSLIRGLIT</td>
<td>56±1</td>
</tr>
<tr>
<td>DecoG9</td>
<td>SLLSLIRGIT</td>
<td>&gt;100</td>
</tr>
<tr>
<td>DecoG10</td>
<td>SLLSLIRGT</td>
<td>&gt;100</td>
</tr>
<tr>
<td>DecoG11</td>
<td>SLLSLIRGI</td>
<td>24±1</td>
</tr>
</tbody>
</table>

We were surprised that only a few positions were insensitive to the mutation. In addition, the results implied that the side chains of N-terminus and C-terminus on Decoralin or TAT(48-60) were not critical to their cytotoxicity nor cell-penetrating activity on MCF7 cells. Interestingly, previous literatures have demonstrated diverse roles of C-terminus and N-terminus in peptide activity.$^{74,251,260,261}$ Structure-function analysis indicated that the C-terminal $\alpha$-helical region
of an ACP, P18, is important for toxicity against human cancer cells whereas the N-terminal region is not.\textsuperscript{262} In contrast, the amphipathic NH2-terminal $\alpha$-helix of cecropin B, which is believed to interact with anionic membrane components via its basic amino acid residues, is critical to mediate cytotoxic activity against cancer cells, whereas the hydrophobic COOH-terminal $\alpha$-helix is dispensable in this regard.\textsuperscript{263} Collectively, our results corresponded well to early findings that ACPs and CPPs possess complex structure-activity relationship.\textsuperscript{73,139,227} Notably, we did not observe improved activity after the point substitution in both peptides, suggesting ‘optimum convergence’ in the peptide sequence during evolution (Table III.3 and figure III.2).

![Figure III.2](image_url)

**Figure III.2** Cell-penetrating activity of TAT(48-60) derivatives on MCF7. Cells were treated with indicated peptides conjugated with FITC at 10 $\mu$M for 4 h, followed by analyzed by FACS analysis. The Mean fluorescence intensity of each treatment is presented as mean $\pm$ SD from triplicates.

For optimizing the templates with the desired property without losing their original activity, we computationally generated random mutations at the specific amino acid positions identified in the glycine scan. As a result, we generated 8,000 derivatives of Decoralin and 160,000 derivatives of TAT(48-60). These derivatives included every combination of amino acid substitutions at the specific positions.
4.4 Optimization of Decoralin with better cytotoxicity by improving its cell-penetrating activity

Then, we computed SVM models to predict the descriptor-encoded virtually generated peptides. To test our first hypothesis, we used ACP SVM and CPP SVM to classify Decoralin derivatives and chose peptides classified as ACPs and CPPs. In total, 146 of 8000 Decoralin derivatives were classed as ACPs and CPPs simultaneously. We calculated the distance to the hyperplane of classified peptides in SVMs and specifically chose peptides that were farthest to the plane in both SVMs. However, we are aware that SVM is not a probabilistic model and that some mathematic modifications were introduced to produce probabilities from SVM. The distance information could only be used to suggest the possibility of the peptides to be truly allocated in the predicted class but could not directly translate to the activities of peptides because we lacked the activity data in our training set. Eleven Decoralin derivatives predicted as ACPs and CPPs were picked for synthesis (Figure III.3), and their cytotoxicities were tested against four different cancer cell lines and the HDMEC primary cell line as representative normal cells (Table III.4).

Decoralin displayed broad-spectrum activity against various cancer cell lines and moderate toxicity in HDMECs. The results suggest different expression of negatively charged molecules and different membrane fluidity, which may determine the abilities of peptides to kill specific cancer cells and their selectivity over normal cells.
III. Sequence-based multi-objective optimization of anticancer peptides: Results and Discussion

**Figure III.3** Distance to the hyperplane of ACP and CPP SVMs of Decoralin derivatives predicted as ACPs and CPPs. Blue circles indicate selected peptides.

**Table III.4** Cytotoxicity of Decoralin and selected tested derivatives predicted as ACP and CPP. The data are presented as mean ± SD from triplicates. The percentage of cell viability was compared with groups without peptide treatment.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>MCF7 EC$_{50}$ (µM)</th>
<th>A549 EC$_{50}$ (µM)</th>
<th>LU1205 EC$_{50}$ (µM)</th>
<th>Jurkat EC$_{50}$ (µM)</th>
<th>HDMEC EC$_{50}$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decoralin</td>
<td>SLLSLIRKLIT</td>
<td>19±1</td>
<td>28±1</td>
<td>18±1</td>
<td>40±1</td>
<td>132±1</td>
</tr>
<tr>
<td>AC1</td>
<td>RLLRLIRKLII</td>
<td>11±1</td>
<td>21±1</td>
<td>9±1</td>
<td>39±1</td>
<td>141±1</td>
</tr>
<tr>
<td>AC2</td>
<td>RLLRLIRKLIL</td>
<td>9±1</td>
<td>14±1</td>
<td>7±1</td>
<td>25±1</td>
<td>125-250</td>
</tr>
<tr>
<td>AC3</td>
<td>LLLLLIRKLIK</td>
<td>11±1</td>
<td>22±1</td>
<td>16±1</td>
<td>25±1</td>
<td>125-250</td>
</tr>
<tr>
<td>AC4</td>
<td>YLLYLIRKLIL</td>
<td>13±1</td>
<td>29±1</td>
<td>14±1</td>
<td>22±1</td>
<td>256±2</td>
</tr>
<tr>
<td>AC5</td>
<td>RLLRLIRKLIR</td>
<td>14±1</td>
<td>27±1</td>
<td>16±1</td>
<td>27±1</td>
<td>138±7</td>
</tr>
<tr>
<td>AC6</td>
<td>QLLQLIRKLII</td>
<td>22±1</td>
<td>29±1</td>
<td>21±1</td>
<td>35±1</td>
<td>171±1</td>
</tr>
<tr>
<td>AC7</td>
<td>QLLQLIRKLIM</td>
<td>33±1</td>
<td>33±1</td>
<td>24±1</td>
<td>46±1</td>
<td>142±2</td>
</tr>
<tr>
<td>AC8</td>
<td>MLLMLIRKLIL</td>
<td>16±1</td>
<td>58±1</td>
<td>31±1</td>
<td>31±1</td>
<td>262±1</td>
</tr>
<tr>
<td>AC9</td>
<td>QLLQLIRKLIR</td>
<td>53±1</td>
<td>60±1</td>
<td>38±1</td>
<td>75±1</td>
<td>260±1</td>
</tr>
<tr>
<td>AC10</td>
<td>QLLQLIRKLIK</td>
<td>50-100</td>
<td>54±1</td>
<td>39±1</td>
<td>69±1</td>
<td>~250</td>
</tr>
<tr>
<td>AC11</td>
<td>QLLQLIRKLIY</td>
<td>50-100</td>
<td>61±1</td>
<td>29±1</td>
<td>58±1</td>
<td>349±1</td>
</tr>
</tbody>
</table>
Interestingly, all 11 Decoralin derivatives displayed some cytotoxicity against cancer cells. We observed a clear trend of activity correlation in different cell lines, *i.e.*, potent peptides presented broad-spectrum activity. This phenomenon suggests a nonspecific mechanism of action of the peptides, and in this case, the cell membrane might be the primary target of the peptides. Notably, AC1-AC5 displayed better activities against all cancer cell lines compared with Decoralin. Furthermore, AC1-AC5 presented comparable toxicities against HDMECs compared with Decoralin. AC2 was the most promising peptide, displaying a nearly two-fold increase in terms of EC$_{50}$ compared with Decoralin.

Additionally, we investigated whether the improved activity of the designed peptides was due to inner cytotoxicity, benefitting from enhanced cell-penetrating activity at low concentrations. First, we coupled FITC onto the N-terminal of AC1-AC5 peptides and probed their cell-penetrating abilities by microscopy and FACS analyses. We observed that all tested peptides could penetrate MCF7 cells at a low concentration (1 µM) (Figure III.4a). However because the fluorescence microscopic analysis did not allow for an irrevocable localization of the peptides, on the cell surface or inside the cells, next, we performed FACS analysis. The results indicated that AC1, AC2 and AC5 had significantly better cell-penetrating activities than Decoralin in MCF7 cells (Figure III.4b, c). Nevertheless, the designed peptides did not show significantly improved cell-penetrating activity in A549 cells compared with Decoralin (Figure III.5a, b), although we observed a trend of enhanced cell-penetrating activity in AC2.
Figure III.4 Decoraln and its derivatives penetrate MCF7 cells to different extents. MCF7 cells were treated with 1 μM of peptides labeled with FITC for 4 h, followed by fluorescence microscopy (a) and FACS analyses (b, c). (b) Representative results of the FACS analysis. Control represents no peptide treatments. (c) Statistics of three independent experiments. The Mean fluorescence intensity of each treatment is presented as mean ± SD. P values < 0.05 were considered significantly different to Decoraln treatment (one-way ANOVA). *, p<0.05, **: p<0.01.

Decoraln itself showed remarkable cell-penetrating activity in MCF7 and A549 cells at a low concentration, corroborating previous observations that ACPs could have not only membrane-lytic but also cell-penetrating activities. Nevertheless, our machine-learning models were capable of identifying key features of CPPs and allowed us to build these features into the Decoraln template. Moreover, our ACP models successfully allocated...
III. Sequence-based multi-objective optimization of anticancer peptides: Results and Discussion

Figure III.5 Decoralin and its derivatives penetrate A549 cells to different extents. A549 cells were treated with 1 µM of peptides labeled with FITC for 4 h, followed by FACS analysis. (a) Representative of the FACS analysis. Control represents no peptide treatments. (b) The Mean fluorescence intensity of each treatment is presented as mean ± SD. No significant difference was observed between treatments of designed peptides and Decoralin.

designed mutated peptides, and all tested peptides retained anticancer activity. We successfully tuned the Decoralin sequence and conferred the extra cell-penetrating ability in MCF7 cells and slightly in A549 cells. AC1, AC2 and AC5 have similar sequences in which two serine residues of the Decoralin sequence have been replaced by arginine. Our study corroborates a previous work by Nakase et al. demonstrating that introducing an arginine residue in an ACP may improve membrane permeability. However, the arginine-substituted peptides lost cytotoxicity against HeLa cells, suggesting a delicate structure-activity relationship of ACPs and CPPs and complex interweaving chemical spaces between these two classes of peptides.

Nevertheless, our designing strategy combining computational and experimental approaches can adequately address this goal. Additionally, CPPs have different cell-penetrating activities
in different cells, although most of the CPPs lack selectivity. Thus, we were not surprised to observe that optimizing the cell-penetrating activities of peptides against all cell lines simultaneously was almost impossible because this type of CPP did not exist in the training set. Further experiments must be performed to elucidate factors contributing to the improved activity of AC2 against A549 cells.

4.5 Designed ACPs permeabilize cells to lesser extents and induce cell apoptosis

To elucidate the mechanisms of action of the designed peptides, we probed the membrane integrity and cellular apoptosis of cancer cells after peptide treatment. We adopted the propidium iodide (PI) exclusion assay to investigate the effect of peptides on cell membrane integrity. After 10 min of treatment, we observed that Decoralin could permeabilize the cell membranes of MCF7 cells more than the designed peptides in every tested concentration. A significant difference was observed at every concentration (Figure III.6a, 6b). This result suggests that the mechanism of action of Decoralin in MCF7 cells is primarily membrane-lytic. In addition, the PI exclusion assay, together with previous observations, suggested that our designed peptides had different mechanisms of action from Decoralin and that these peptides could kill MCF7 cells not only by lysing the membranes at high concentrations but also by penetrating cells and by exerting its inner cytotoxicity at a low concentration. Furthermore, we performed the 7-AAD and Annexin V-FITC cell-apoptosis assays for Decoralin and for one of designed peptides (AC2) to probe the cell apoptosis level after treatment. We observed that AC2 induced more early apoptotic (1.2% compared with 5.18%, Annexin V-FITC+/7-AAD-) and late-apoptotic and/or necrotic (15% compared with 3.45%, Annexin V-FITC+/7-AAD+) cells than Decoralin did, whereas AC2 and Decoralin
induced a similar level of damaged cells (18.8% compared with 18.6%, Annexin V-FITC/7-AAD⁺) (Figure III.7). Notably, the relative higher percentage of cells in Annexin V-FITC-negative and 7-AAD-positive phases in Decoralin-treated cells indicated that the cell membranes in most dead cells were completely lysed. Therefore, no PS distribution was on concrete cell membranes for proper staining, which is consistent with observations from other membrane-lytic peptides.²⁴⁸ In AC2-treated cells, cells were in all three quadrant phases, demonstrating that AC2 could kill MCF7 cells in both membrane-lytic and inner targeting pathways.

Figure III.6 Decoralin and its derivatives permeabilized MCF7 cell membranes to different extents. MCF7 cells were treated by the indicated concentrations of peptides for 10 min and were stained with PI, and immediately analyzed by FACS. (a) A representative result from two independent experiments. Control represents no peptide treatments. (b) Columns represent relative cell percentages in the pre-defined compromised cell membranes segment (right segment). *: Represents significant difference (two-way ANOVA test) compared with Decoralin treatment in respective concentration.
III. Sequence-based multi-objective optimization of anticancer peptides: Results and Discussion

Figure III.7 MCF7 cells were treated with 25 µM of Decoralin or of AC2 for 20 min and stained with 7-AAD/Annexin V-FITC. The numbers in each quadrant indicate the relative amount of cells as a percentage. This figure is a representative result from three independent experiments. Control represents no peptide treatments. The numbers in each quadrant represent the relative cell amount as a percentage.

4.6 AC2 displays better synergism with doxorubicin and epothilone C against A549 but not MCF7 cells

Furthermore, we investigated the potential synergism of our designed peptides with the conventional chemotherapeutic drugs doxorubicin and epothilone C. Poor tumor penetration of anticancer drugs can be a limiting factor for their efficacy, and co-administration with ACPs or with other tumor-specific penetrating peptides is suggested as a promising solution for clinical use.\textsuperscript{268,269} A549 and MCF7 cells showed different sensitivities to doxorubicin (450±1 and 22±1 nM, respectively) (Figure III.8a, b) and to epothilone C (49±1 and 14±2 nM, respectively) (Figure III.8c, d). We hypothesized that this difference was due to the poor penetration of both drugs into A549 cells.
We co-incubated 10 µM each of Decoralin and AC2 with doxorubicin and epothilone C in A549 and MCF7 cells. After 72 h of incubation, we observed significant synergism with doxorubicin in both Decoralin and AC2 co-treatment groups (260±1 and 127±1 nM, respectively, p<0.0001 from extra sum-of-squares F test) (Figure III.9a). However, we did not observe significant synergism in MCF7 cells (23±1 and 21±1 nM, respectively, p>0.05) (Figure III.9b). AC2 outperformed Decoralin in synergism with doxorubicin in A549 cells in terms of logEC50 (p=0.04). We also observed significant synergism with epothilone C in A549 cells (31±1 and 21±1 nM, respectively, p<0.0001) (Figure III.9c) but not in MCF7 cells (24±1 and 19±1 nM, respectively, p>0.05) (Figure III.9d). Again, AC2 outperformed
Decoralin with epothilone C in terms of logEC\textsubscript{50} (p=0.003). These results confirmed that the poor sensitivity of A549 to doxorubicin was due to a penetration issue; thus, both membrane-permeable Decoralin and AC2 could enhance its efficacy. Moreover, these results clearly demonstrated that AC2 had a better membrane-disturbance effect than Decoralin on A549 cells, although AC2 showed only a minor, but not significant, improvement in cell-penetrating ability in the previous FACS analysis.

![Figure III.9](image-url) Synergism effects of Decoralin and AC2 with doxorubicin on MCF7 and A549 cells. (a), (c) A549 and (b), (d) MCF7 cells were incubated with different concentrations of doxorubicin (a), (b) and epothilone C (c), (d) and 10 µM of the indicated peptides for 72 h. Each data point is presented as the mean ± S.D. from three replicates and fitted with non-linear regression.
4.7 Optimization of Decoralin with better selectivity by building into THP features

Next, we intended to optimize Decoralin for better selectivity. Although ACPs possess some inherent selectivity to cancer cells over normal cells, not all ACPs displaying antitumor activity are sufficiently selective against cancer cells. To test our second hypothesis, we used the ACP SVM and THP SVM to classify Decoralin derivatives and chose peptides classified as ACPs and THPs. In total, 196 of 8000 Decoralin derivatives were classed as ACPs and THPs simultaneously. Again, we selected 7 peptides that were far from the hyperplane in the ACP SVM and THP SVM (Figure III.10). All 7 tested peptides showed improved selectivity over HDMEC compared with the template (Table III.5), and four of these peptides (AT1-AT4) retained remarkable cytotoxicity against a range of cancer cell lines. We specifically selected AT2 and AT4 for further analysis because these two peptides showed selective toxicity against LU1205 and MCF7 cells but less toxicity on HDMEC cells, which advocated for the possession of ACP and THP features. Notably, we did not observe this selective cancer cell toxicity when designing ACPs and CPPs in which all Decoralin derivatives showed all or no potency against the tested cancer cells.

![Figure III.10 Distance to the hyperplane of ACP and THP SVMs of predicted Decoralin derivatives. Blue circles indicate the selected peptides.](image-url)
**Table III.5** Cytotoxicity of Decoralin and selected derivatives predicted as ACPs and THPs. The data are presented as mean ± SD from three replicates. The percentage of cell viability was compared with groups without peptide treatment.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>MCF7 EC_{50} (µM)</th>
<th>A549 EC_{50} (µM)</th>
<th>LU1205 EC_{50} (µM)</th>
<th>Jurkat EC_{50} (µM)</th>
<th>HDMEC EC_{50} (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decoralin</td>
<td>SLLSLIRKLIT</td>
<td>19±1</td>
<td>28±1</td>
<td>18±1</td>
<td>40±1</td>
<td>132±1</td>
</tr>
<tr>
<td>AT1</td>
<td>LLLQLIRKLIL</td>
<td>10±1</td>
<td>30±1</td>
<td>12±1</td>
<td>53±1</td>
<td>154±1</td>
</tr>
<tr>
<td>AT2</td>
<td>TLLLLIRKLIL</td>
<td>38±1</td>
<td>79±1</td>
<td>16±1</td>
<td>30±1</td>
<td>348±1</td>
</tr>
<tr>
<td>AT3</td>
<td>RLLLLIRKLIL</td>
<td>24±1</td>
<td>35±1</td>
<td>18±1</td>
<td>40±1</td>
<td>160±1</td>
</tr>
<tr>
<td>AT4</td>
<td>LLLLLIRKLIL</td>
<td>25±1</td>
<td>&gt;100</td>
<td>48±1</td>
<td>50±1</td>
<td>278±1</td>
</tr>
<tr>
<td>AT5</td>
<td>QLLLLIRKLIL</td>
<td>63±1</td>
<td>&gt;100</td>
<td>44±1</td>
<td>39±1</td>
<td>271±1</td>
</tr>
<tr>
<td>AT6</td>
<td>QLLLLIRKIV</td>
<td>77±1</td>
<td>&gt;100</td>
<td>53±1</td>
<td>&gt;100</td>
<td>339±1</td>
</tr>
<tr>
<td>AT7</td>
<td>CLLLLIRKLIL</td>
<td>&gt;100</td>
<td>&gt;100</td>
<td>76±1</td>
<td>&gt;100</td>
<td>~500</td>
</tr>
</tbody>
</table>

Additionally, we used confocal microscopy to probe the targeting and distribution patterns of AT2 and AT4 in LU1205, MCF7 and A549 cells. At the EC_{50} concentration for their specific sensitive cells (15 µM for AT2 and 25 µM for AT4), FITC-labeled AT2 distributed around the surface of LU120 and MCF7 cells, where AT2 co-localized with the cell membrane dye, whereas no AT2 was found around A549 cells (Figure III.11). Conversely, for AT4, we did not observe any peptides distributed around three tested cells (Figure III.12).
Figure III.11 Distribution of AT2 in tested cancer cells. AT2-FITC was incubated with the indicated cancer cells for 30 min, followed by staining with CellMask Deep Red cell membrane dye and Hoechst 33342 nucleus dye (blue). The images were taken by confocal microscopy. The figures at the left are images in green and blue fluorescence channels, and at the right, these images are overlapping with additional red fluorescence.
Figure III.12 Distribution of AT4 in tested cancer cells. AT2-FITC was incubated with the indicated cancer cells for 30 min followed by staining with CellMask Deep Red cell membrane dye and Hoechst 33342 nucleus dye (blue). The images were taken by confocal microscopy. The figures at the left are images in green and blue fluorescence channels, and at the right, these images are overlapping with additional red fluorescence.

Although more experiments may be required for solid conclusions, for example, animal studies and inhibitor assays, these preliminary results suggested that AT2 could be used as an anticancer peptide with a tumor-homing property. AT2 possessed cytotoxicity against the
specific targeted cancer cells but less cytotoxicity against normal cells. In fact, one study demonstrated that a THP identified by phage-display, LyP-1, was unique among its class because this THP had cytotoxicity on its own.\(^{271}\) When injected into the tail veins of mice with breast cancer xenografts, LyP-1 inhibited tumor growth and reduced the number of lymphatic vessels. Further studies are required to identify AC2 receptors on cancer cell surfaces and mechanisms of action to further explore the targeted delivery of payloads in cancer therapy.\(^{241}\)

### 4.8 Optimization of TAT(48-60) with anticancer activity

We generated 160,000 TAT(48-60) derivatives \textit{in silico} as previous described and employed the ACP SVM to predict the anticancer activities of the derivatives. However, none of the 160,000 derivatives were classified as ACPs. Inspecting the training dataset revealed two major reasons accounting for possible false negative predictions: 1) A lack of diverse ACP sequences in the training set; 2) many similar sequences to TAT(48-60) in the CPP training set, which was used to compose negative training sets of ACPs. In the ACP SVM, we included in-house test results of ten Decoralin scrambled sequences generated for the first AMP project; thus, the ACP SVM was able to make specific and accurate predictions for Decoralin derivatives, although only a restricted number of ACPs were included due to a lack of additional data. To circumvent this problem, we included TAT(49-57) (RKKRRQRRR), which was tested in-house and which displayed 15% inhibition of MCF7 cell growth at 100 \(\mu\)M, and four new ACP sequences from APD2 as ACP-positive examples. We deleted one similar sequence, TAT(48-57) (GRKKRRQRRR), from the negative dataset, which we speculated as also having anticancer activity but was not confirmed. The Gaussian kernel was
used for constructing a new SVM model, ACP2 SVM. We performed 10-fold cross-validation for ACP2 SVM, and the ACP2 SVM performed similar to ACP SVM (Table III.6).

Table III.6 Performance of ACP2 SVM models in 10-fold cross-validation. The data are presented as the mean ± SD.

<table>
<thead>
<tr>
<th></th>
<th>MCC</th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACP SVM</td>
<td>0.59±0.20</td>
<td>0.94±0.04</td>
<td>0.77±0.16</td>
</tr>
<tr>
<td>ACP2 SVM</td>
<td>0.62±0.20</td>
<td>0.94±0.04</td>
<td>0.72±0.20</td>
</tr>
</tbody>
</table>

Then, we utilized the ACP2 SVM, together with the CPP SVM and the THP SVM, to predict classes of TAT derivatives. Ten of 160,000 peptides were classified as ACPs and CPPs (Figure III.13); however, no peptides were classified as ACPs and THPs or THPs alone. We synthesized all 10 peptides for testing, and 5 of the 10 peptides displayed mild cytotoxicity against MCF7 cells (Table III.7).

![Figure III.13](image-url) Distance to the hyperplane of ACP2 and CPP SVMs of TAT(48-60) derivatives. There were ten peptides classified as ACP and CPP. We synthesized all peptides and tested their cytotoxicity.
Table III.7 Cytotoxicities of TAT(48-60) and derivatives predicted as ACPs and CPPs. The data are presented as mean ± SD from three replicates. The percentage of cell viability was compared with groups without peptide treatment. N.A.: no activity.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>MCF7 EC₅₀ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAT(48-60)</td>
<td>GRKKRRQRRRPPQ</td>
<td>N.A.</td>
</tr>
<tr>
<td>TAC-1</td>
<td>GRKKRRKRRQQRRR</td>
<td>N.A.</td>
</tr>
<tr>
<td>TAC-2</td>
<td>GRKKRRKRRRRRQ</td>
<td>182±1</td>
</tr>
<tr>
<td>TAC-3</td>
<td>GRKKRRQRRKKRR</td>
<td>N.A.</td>
</tr>
<tr>
<td>TAC-4</td>
<td>GRKKRRQRRRKKK</td>
<td>N.A.</td>
</tr>
<tr>
<td>TAC-5</td>
<td>GRKKRRKRRRRQR</td>
<td>N.A.</td>
</tr>
<tr>
<td>TAC-6</td>
<td>GRKKRRQRRRKKR</td>
<td>167±1</td>
</tr>
<tr>
<td>TAC-7</td>
<td>GRKKRRRRRRRKRQ</td>
<td>N.A.</td>
</tr>
<tr>
<td>TAC-8</td>
<td>GRKKRRKRRRRQK</td>
<td>216±1</td>
</tr>
<tr>
<td>TAC-9</td>
<td>GRKKRRQRRRRKK</td>
<td>242±1</td>
</tr>
<tr>
<td>TAC-10</td>
<td>GRKKRRRRRRRQKR</td>
<td>180±1</td>
</tr>
</tbody>
</table>

Having displayed moderate activities against the MCF7 cell line, we selected TAC-2, 6 and 10 for further analysis. We tested their cytotoxicities against A549 and HDMEC cells. None of these peptides displayed activities against A549 or HDMEC cells (Table III.8).
Table III.8 Cytotoxicities of TAT(48-60) and selected derivatives. The data are presented as mean ± SD from three replicates. The percentage of cell viability was compared with groups without peptide treatment. N.A.: no activity.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>MCF7 EC_{50}(µM)</th>
<th>A549 EC_{50}(µM)</th>
<th>HDMEC EC_{50}(µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAT(48-60)</td>
<td>GRKKRRQRRPPQ</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>TAC-2</td>
<td>GRKKRRKRRRRQ</td>
<td>182±1</td>
<td>N.A.</td>
<td>&gt;500</td>
</tr>
<tr>
<td>TAC-6</td>
<td>GRKKRRQRRRRKR</td>
<td>167±1</td>
<td>N.A.</td>
<td>&gt;500</td>
</tr>
<tr>
<td>TAC-10</td>
<td>GRKKRRRRRRQKR</td>
<td>180±1</td>
<td>N.A.</td>
<td>&gt;500</td>
</tr>
</tbody>
</table>

In addition, TAC-2, 6 and 10 showed much better cell-penetrating activity than their templates in MCF7 cells (Figure III.14a, b).

Figure III.14 TAT(48-60) and its derivatives penetrate MCF7 cells to different extents. MCF7 cells were treated with 10 µM of peptides labeled with FITC for 4 h, followed by FACS analysis. (a) Representative of FACS analysis. Control represents no peptide treatments. (b) Statistics of two independent experiments. The Mean fluorescence intensity of each treatment is presented as mean ± SD. P values < 0.05 were considered significantly different with regard to TAT(48-60) treatments. ***: p<0.001 (one-way ANOVA test).
We also probed the cell membrane integrity after the peptide treatment. After 10 min of incubation with 100 µM of each peptide, no dramatic cell membrane leakage was observed in MCF7 cells (Figure III.15).

![Figure III.15](image)

**Figure III.15** TAT(48-60) derivatives did not permeabilize MCF7 cell membranes. MCF7 cells were treated by 100 µM of the indicated peptides for 10 min and were stained with PI, and immediately analyzed by FACS. The numbers in the graph present the relative percentages of total cells in defined segments. The results from one representative out of three independent experiments are shown. Control represents no peptide treatments.

From our data, we propose that TAT(48-60) derivatives are not membrane-lytic peptides. These peptides exert their activity via inner cytotoxicity. Our SVMs were accurate in extracting physicochemical features of known ACPs related to their activity, including membrane-lytic and/or inner targeting, which allowed us to build anticancer activity into a TAT template. Moreover, this computational approach also built the general selectivity of ACPs into the template, *i.e.*, the designed cancer active peptides did not show cytotoxicity against HDMEC cells up to 500 µM. TAT(48-60) displayed a broad spectrum and potent antimicrobial activity with mechanisms of action other than membrane targeting. Thus, we speculate that their enhanced cell-penetrating abilities in MCF7 cells might account for their
III. Sequence-based multi-objective optimization of anticancer peptides: Conclusions

extra anticancer activities. Nevertheless, the designed TAT derivatives only showed moderate activity \( (EC_{50} = 167\pm1 \mu M) \) against MCF7 cells. One possible reason is the endosomal entrapment of the peptides during cell membrane permeation. TAT-relative peptides use several mechanisms to cross the membrane, including macropinocytosis,\textsuperscript{272} clathrin-mediated endocytosis,\textsuperscript{273} caveolae/lipid-raft-mediated endocytosis,\textsuperscript{273} and endocytosis-independent uptake.\textsuperscript{274} Regardless of the exact mechanisms of entry, most reports have demonstrated that TAT cargo is massively trapped in vesicles.\textsuperscript{275} Considering that cell-penetrating peptides with membrane-lytic properties can efficiently escape from endosomes,\textsuperscript{276} we expect that similar computational approaches can be applied to optimize the designed peptides with membrane-lytic properties to obtain potent anticancer agents.

5. Conclusions

In this study, we first presented a novel strategy of a computational-aided approach to design ACPs with improved activities and selectivities. We built CPP features into Decoralin to enhance its cell-penetrating activity; thus, the peptides might enter cells before reaching the threshold concentration on the membrane and might induce intracellular toxicity. The best designed ACP and CPP peptide AC2 did not lyse the cell membrane as much as Decoralin did, but it might be able to kill cancer cells via inducing cell apoptosis inside the cells. AC2 slightly outperformed Decoralin in synergism with doxorubicin and epothilone C. We also demonstrated methods to include tumor-homing features into Decoralin. The best ACP and THP peptide AT2 retained remarkable cytotoxicity against specific cancer cells but less toxicity against normal cell lines. We also showed that AT2 could be distributed near the surface of targeting cells, suggesting specific surface-targeting features of certain cancers. In addition, we were able to use this computational approach to turn TAT(48-60) into a
cell-penetrating anticancer peptide. Notably, no peptides were classified as ACPs, CPPs and THPs simultaneously. Although some studies have discovered cancer-specific cell-penetrating peptides, using these peptides as vehicles to deliver anticancer drugs or peptides\textsuperscript{270,277} and discovered tumor-homing peptides with cytotoxicity itself,\textsuperscript{271} our results suggest the possibility of different sequence spaces for ACPs and for CPPs/THPs. Our present work corroborates previous findings in ACPs, CPPs and THPs and demonstrates the possibility of combine features in peptide optimization. Our strategy is also suitable for the \textit{de novo} design of peptides with CPPs and THPs properties. Although CPPs have been harnessed to translocate a wide range of molecules, including application in synergism with conventional chemotherapeutic agents, the lack of selectivity of these peptides remains a hurdle in CPP development.\textsuperscript{233,267} Numerous strategies have been adopted in building cell selectivity into CPPs,\textsuperscript{278} \textit{e.g.}, conjugation with a THP\textsuperscript{279,280} or discovering tumor-homing CPPs.\textsuperscript{270} Instead of engineering high mass peptides with potential solubility problems or conducting labor-intensive biologic studies for \textit{de novo} peptide discovery, our computational peptide design strategy can provide great help in optimizing CPP with tumor-homing property and vice versa. In addition, our methods differ from traditional sequence-based computational peptide design because we do not require exhaustive experimental work to build up structure and activity relationship\textsuperscript{139,281}. In contrast, combining a mechanism-guided strategy with a suitable computational approach, we successfully obtained peptides with desired pharmacological properties.
IV. Conclusions and Outlook

In this study, we demonstrated the application of computational-aided peptide design approaches in both de novo sequence design and multi-objective optimization. In general, the sequence-based peptide design approach was demonstrated to be suitable for the discovery of AMPs and ACPs where no specific protein receptor is targeted by these membrane-active peptides. Consistent with previous observations, the sequence-dependency of membrane-active peptides can be used for peptide design. We took advantage of this fact and utilized suitable computational models, specifically the self-organizing map (SOM) algorithm, to prioritize computationally generated membrane-active peptides. Without exhaustively synthesizing and testing peptides to establish a complex supervised QSAR model, we immediately obtained AMPs with profound membrane-lytic and bacteriostatic activity. Although the designed peptides were not sufficiently potent to be developed as antibiotic drugs, these peptides can serve as templates for further optimization because these peptides represent new chemical space (i.e., no similar sequences in the UnitProtKB database according to BLAST). Moreover, the peptides were sensitive to both Gram-positive and Gram-negative pathogens, which provide an opportunity for further study and the potential development of broadly active AMP-based drugs.

In the ACP project, we started from peptide templates and rationally mutated amino acids to optimize their activity and to target cell selectivity. Motivated by the similar physicochemical properties of the ACPs and CPPs, we adopted a supervised machine-learning method to subtly tune the sequences for different purposes. We tuned the ACP template Decoralin toward the CPP sequence space for improved activity. In fact, the designed peptides feature a compromise between membrane-lytic and the newly acquired additional cell-penetrating
abilities, which is reflected in their intracellular cytotoxicity at low concentrations. The CPP template TAT(48-60) was tuned toward the ACP sequence space so that the designed derivatives possess extra cancer cytotoxicity. In contrast, THPs are widely found with cell-penetrating abilities but rarely show cytotoxicity against cancer cells. From our analysis of the SVM models, ACPs and THPs appear to populate different sequence spaces; however, ACPs and THPs may have partial overlaps remaining to be discovered. Our method provides a rationale for exploring this overlapping chemical space. We demonstrated this ability by amalgamating Decoralin features with tumor-homing-like properties; thus, the designed peptides became specifically toxic for certain cancer cell lines. Currently, the exact mechanisms of action of these designer peptides remain unclear. Further studies are required to confirm the tumor-homing properties of these peptides and their targeted receptors/molecules.

Computational methods help in generating (quantitative) sequence- and structure-activity relation QSAR models for automated peptide design.\cite{284,285} In particular, statistical QSAR methods have been employed to obtain new AMPs and other membrane-interacting peptides. For example, Jensen et al. recently presented the application of PCA/PLS to find peptides with activities against antibiotic-resistant \textit{Pseudomonas aeruginosa},\cite{286,287} whereas Maccari et al. used random forest models.\cite{174} Recent examples include studies by Fjell et al., who employed conventional multi-layer feedforward networks for AMP design.\cite{169} In this context, one may distinguish between supervised and unsupervised neural network learning schemes.\cite{250} Although supervised methods generate predictive models by minimizing an error function that requires peptide activity for regression or class labels for classification, unsupervised methods are based on the principle of chemical similarity without the requirement of activity data. These computational techniques are used to group peptides
IV. Conclusions and Outlook

The obtained peptide clusters may be analyzed for common features or identifying peptides for activity testing. A particular advantage of unsupervised approaches over supervised learning is their applicability to small sets of reference data.

High-throughput screening (HTS) represents the mainstream of drug discovery technologies in most pharmaceutical companies. However, the return on investments in terms of marketed products have not met expectations. In the context of intense commercial pressure to increase discovery productivity, new approaches to rapidly develop small molecules and biologics against a wide range of therapeutic targets are always needed.

Computer-aided drug design approaches represent a practical alternative strategy that can perform molecular design cycles in silico and that can efficiently generate hits for further optimization. Although the cost of false-positive hits is not high compared with the total cost of HTS campaigns, clearly, CADD researchers must further improve the hit rate of drug design exercises. To achieve broad application of CADD, a more intensive communication between bioinformaticians, chemists and biologists is required, particularly commuting in a ‘mutually understandable scientific language’ so that each community is able to exchange the experience and unmet research needs.

We foresee that computational drug design approaches combined with automated and fast chemical synthesis and reliable high-throughput biologic testing will become an useful drug discovery tool in the future. In fact, researchers at Roche have recently reported a fully integrated and automated industrial process in which these researchers developed a flow-based biochemical assay integrated with chemical flow synthesis, purification and in-line quantification of compound concentrations. Recently, we have also pioneered a combined computational drug design with microfluidic chemistry. We are optimistic that CADD, in combination with high-throughput readouts, will become a standard technique
in every drug discovery setting, in which CADD will unleash its full effect on the pharmaceutical community. Peptide design is ideally suited for this endeavor.
V. Appendix:

HPLC chromatograms (UV absorbance at 210 nm) of representative peptides

**AC2:**
RLLRLIRKLIL-NH$_2$
MW: 1406.8 g/mol
Retention time: 8.3 min
Purity: 100%

**AC3:**
LLLLLIRKLIK-NH$_2$
MW: 1335.8 g/mol
Retention time: 9.9 min
Purity: 100%

**AC4:**
YLLYLIRKLIL-NH$_2$
MW: 1420.8 g/mol
Retention time: 11.4 min
Purity: 100%

**AC5:**
RLLRLIRKLIR-NH$_2$
MW: 1449.9 g/mol
Retention time: 7.1 min
Purity: 92.2%

**AC6:**
QLLQLIRKLII-NH$_2$
MW: 1350.7 g/mol
Retention time: 10.9 min
Purity: 96.8%
AC7: QLLQLIRKLIM-NH₂
MW: 1368.7 g/mol
Retention time: 11.3 min
Purity: 100%

AC8: MLLMLIRKLIL-NH₂
MW: 1356.8 g/mol
Retention time: 11.4 min
Purity: 100%

AC9: QLLQLIRKLR-NH₂
MW: 1393.7 g/mol
Retention time: 9.6 min
Purity: 100%

AC10: QLLQLIRKLIL-NH₂
MW: 1365.7 g/mol
Retention time: 9.5 min
Purity: 100%

AC11: QLLQLIRKLIL-NH₂
MW: 1400.7 g/mol
Retention time: 10.9 min
Purity: 83.4%

AT1: LLLQLIRKLIL-NH₂
MW: 1335.7 g/mol
Retention time: 11.6 min
Purity: 100%
AT2: TLLLLIRKLIL-NH2
MW: 1308.7 g/mol
Retention time: 11.6 min
Purity: 100%

AT3: RLLLLIRKLIL-NH2
MW: 1363.8 g/mol
Retention time: 9.7 min
Purity: 81.3%

AT4: LLLLLIRKLIL-NH2
MW: 1320.8 g/mol
Retention time: 11.6 min
Purity: 100%

AT5: QLLLLIRKLIL-NH2
MW: 1365.7 g/mol
Retention time: 11.1 min
Purity: 97.6%

AT6: QLLLLIRKLiV-NH2
MW: 1321.7 g/mol
Retention time: 10.6 min
Purity: 82.6%

AT7: CLLLLIRKLIL-NH2
MW: 1310.7 g/mol
Retention time: 11.8 min
Purity: 86.8%
V. Appendix

AC2-FITC: FITC-β-alanine-RLRLIRKLIL-NH2
MW: 1866.3 g/mol
Retention time: 12.1 min
Purity: 100%

AC3-FITC: FITC-β-alanine-LLLLLIRKLKL-NH2
MW: 1837.3 g/mol
Retention time: 13.5 min
Purity: 100%

AC4-FITC: FITC-β-alanine-YLLYLIRKLIL-NH2
MW: 1880.3 g/mol
Retention time: 15.0 min
Purity: 100%

AC5-FITC: FITC-β-alanine-RLRLIRKLIR-NH2
MW: 1909.4 g/mol
Retention time: 10.8 min
Purity: 95.3%

AT2-FITC: FITC-β-alanine-TLLLLIRKLIL-NH2
MW: 1766.2 g/mol
Retention time: 14.8 min
Purity: 100%
V. Appendix

**TAT(48-60):**

- **Sequence:** GRKKRRQRRRPPQ
- **MW:** 1719 g/mol
- **Retention time:** 1.0 min
- **Purity:** 100%

**AT4-FITC:**

- **Sequence:** FITC-β-alanine-LLLLLIRKLIL-NH₂
- **MW:** 1780.3 g/mol
- **Retention time:** 15.1 min
- **Purity:** 100%

**TAC-2:**

- **Sequence:** GRKKRRKRRRRRQ-NH₂
- **MW:** 1837.2 g/mol
- **Retention time:** 0.97 min
- **Purity:** 100%

**TAC-6:**

- **Sequence:** GRKKRRQRRRKR-NH₂
- **MW:** 1837.2 g/mol
- **Retention time:** 0.97 min
- **Purity:** 100%

**TAC-10:**

- **Sequence:** GRKKRRRRRQKR-NH₂
- **MW:** 1837.2 g/mol
- **Retention time:** 0.97 min
- **Purity:** 100%
TAC-2-FITC: FITC-β-alanine-GRKKRRKRRRRRQ-NH₂
MW: 2296.7 g/mol
Retention time: 0.97 min
Purity: 100%

Datafile Name: yelin_22_06_2014_2_18.lcd
Sample Name: yelin_22_06_2014_2_18

Datafile Name: yelin_22_06_2014_2_19.lcd
Sample Name: yelin_22_06_2014_2_19

TAC-6-FITC: FITC-β-alanine-GRKKRRQRRRKR-NH₂
MW: 2296.7 g/mol
Retention time: 0.98 min
Purity: 100%

TAC-10-FITC: FITC-β-alanine-GRKKRRRRRQKR-NH₂
MW: 2296.7 g/mol
Retention time: 0.98 min
Purity: 100%
VI. Cumulative References


Bessalle, R., Gorea, A., Shalit, I., Metzger, J. W., Dass, C., Desiderio, D. M. & Fridkin,


40 Matsuzaki, K., Murase, O., Fujii, N. & Miyajima, K. An antimicrobial peptide, magainin


77 Wang, F. L., Cui, S. X., Sun, L. P., Qu, X. J., Xie, Y. Y., Zhou, L., Mu, Y. L., Tang, W. & Wang, Y. S. High expression of alpha 2, 3-linked sialic acid residues is associated


94 Camilio, K. A., Berge, G., Ravuri, C. S., Rekdal, O. & Sveinbjornsson, B. Complete regression and systemic protective immune responses obtained in B16 melanomas after


97 Funatsu, K. in De novo molecular design, Foreword, (ed Schneider, G.), (Wiley-VCH Verlag GmbH & Co. KGaA), (2013).


110 Schneider, G & Baringhaus, K.-H. in *De novo molecular design, De novo design: from models to molecules*, (ed Schneider, G.), (Wiley-VCH Verlag GmbH & Co. KGaA), (2013).


112 Hartenfeller, M., Zettl, H., Walter, M., Rupp, M., Reisen, F., Proschak, E., Weggen, S.,


144 Witkowska, J., Karpowicz, P., Gaczynska, M., Osmulski, P. A. & Jankowska, E. Dissecting a role of a charge and conformation of Tat2 peptide in allostERIC regulation of


Z., Song, H., Cai, Y. D. & Chou, K. C. Prediction of antimicrobial peptides based on

153 Bond, P. J. & Khalid, S. Antimicrobial and cell-penetrating peptides: structure, assembly
and mechanisms of membrane lysis via atomistic and coarse-grained molecular

154 Khandelia, H. & Kaznessis, Y. N. Molecular dynamics simulations of helical

155 Romo, T. D., Bradney, L. A., Greathouse, D. V. & Grossfield, A. Membrane binding of
an acyl-lactoferricin B antimicrobial peptide from solid-state NMR experiments and

156 Wang, Y., Schlamadinger, D. E., Kim, J. E. & McCammon, J. A. Comparative
molecular dynamics simulations of the antimicrobial peptide CM15 in model lipid

Recent advances in computational modeling of α-helical membrane-active peptides.

Coupling molecular dynamics simulations with experiments for the rational design of

159 Hellberg, S., Sjostrom, M., Skagerberg, B. & Wold, S. Peptide quantitative
structure-activity relationships, a multivariate approach. *J. Med. Chem.* **30**, 1126-1135,

160 Lejon, T., Stiberg, T., Strom, M. B. & Svendsen, J. S. Prediction of antibiotic activity


171 Schneider, G., Schrödl, W., Wallukat, G., Muller, J., Nissen, E., Ronspeck, W.,


179 Pate, M. & Blazyk, J. Methods for assessing the structure and function of cationic


189 Pathak, N., Salas-Auvert, R., Ruche, G., Janna, M. H., McCarthy, D. & Harrison, R. G.


200 Shai, Y., Bach, D. & Yanovsky, A. Channel formation properties of synthetic pardaxin


205 He, J., Krauson, A. J. & Wimley, W. C. Toward the *de novo* design of antimicrobial peptides: Lack of correlation between peptide permeabilization of lipid vesicles and antimicrobial, cytolytic, or cytotoxic activity in living cells. *Biopolymers* **102**, 1-6, (2014).


209 Harzer, U. & Bechinger, B. Alignment of lysine-anchored membrane peptides under


216 Koeppe, R. E., 2nd, Hatchett, J., Jude, A. R., Providence, L. L., Andersen, O. S. & Greathouse, D. V. Neighboring aliphatic/aromatic side chain interactions between

217 Michalek, M., Vincent, B., Podschan, R., Grotzinger, J., Bechinger, B. & Jung, S.


218 Schneider, G., Sjoling, S., Wallin, E., Wrede, P., Glaser, E. & von Heijne, G.


224 Waghu, F. H., Gopi, L., Barai, R. S., Ramteke, P., Nizami, B. & Idicula-Thomas, S.


Crombez, L., Aldrian-Herrada, G., Konate, K., Nguyen, Q. N., McMaster, G. K.,
Brasseur, R., Heitz, F. & Divita, G. A new potent secondary amphipathic
   cell-penetrating peptide for siRNA delivery into mammalian cells. Mol. Ther. 17,

Eiriksdottir, E., Konate, K., Langel, U., Divita, G. & Deshayes, S. Secondary structure of
cell-penetrating peptides controls membrane interaction and insertion. Biochim. Biophys.

Henriques, S. T., Melo, M. N. & Castanho, M. A. Cell-penetrating peptides and

Zhang, X., Oglecka, K., Sandgren, S., Belting, M., Esbjorner, E. K., Norden, B. &
Graslund, A. Dual functions of the human antimicrobial peptide LL-37-target membrane
perturbation and host cell cargo delivery. Biochim. Biophys. Acta 1798, 2201-2208,
(2010).

Splith, K. & Neundorf, I. Antimicrobial peptides with cell-penetrating peptide properties

Dennison, S. R., Harris, F., Bhatt, T., Singh, J. & Phoenix, D. A. A theoretical analysis
   of secondary structural characteristics of anticancer peptides. Mol. Cell. Biochem. 333,

Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned

Laakkonen, P. & Vuorinen, K. Homing peptides as targeted delivery vehicles. Integr.


Gautam, A., Singh, H., Tyagi, A., Chaudhary, K., Kumar, R., Kapoor, P. & Raghava, G.


269 Sugahara, K. N., Teesalu, T., Karmali, P. P., Kotamraju, V. R., Agemy, L., Greenwald, D. R. & Ruoslahti, E. Coadministration of a tumor-penetrating peptide enhances the


287 Jenssen, H., Lejon, T., Hilpert, K., Fjell, C. D., Cherkasov, A. & Hancock, R. E.
Evaluating different descriptors for model design of antimicrobial peptides with

288 Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T.,
Sittampalam, G. S. Impact of high-throughput screening in biomedical research. *Nat.

289 Bleicher, K. H., Bohm, H. J., Muller, K. & Alanine, A. I. Hit and lead generation:

290 Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic advances

291 Werner, M., Kuratli, C., Martin, R. E., Hochstrasser, R., Wechsler, D., Enderle, T.,
Alanine, A. I. & Vogel, H. Seamless integration of dose-response screening and flow
chemistry: efficient generation of structure-activity relationship data of beta-secretase

292 Reutlinger, M., Rodrigues, T., Schneider, P. & Schneider, G. Combining on-chip
synthesis of a focused combinatorial library with computational target prediction reveals

293 Rodrigues, T., Schneider, P. & Schneider, G. Accessing new chemical entities through