CONSTRUCTION AND SCREENING OF CUSTOM PEPTIDE LIBRARIES DERIVED FROM SYNTHETIC DNA POOLS

by

Saadet Albayrak

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Engineering)
in The University of Michigan
2011

Doctoral Committee:

Professor Erdogan Gulari, Chair
Professor Henry Y. Wang
Associate Professor Michael Mayer
Assistant Professor Xiaoxia Lin

Aileme ve Askima

ACKNOWLEDGMENTS

I would like to first thank my research advisor Prof. Erdogan Gulari for granting me the opportunity to pursue my career goals and for the continued support he has provided during my work. I have endless thanks to Dr. Jean-Marie Rouillard for all the help he has provided me in the last six years. He has been a great mentor to me and his motivation and passion has served as an example which I have tried to follow.

I would like to thank Prof. Henry Wang, Prof. Michael Mayer and Prof. Nina Lin for their guidance and feedback towards completing my work. I thank past and current members of Gulari research group for their help and support. I also want to thank all of my friends and peers in the department who have made this a great experience.

My deepest appreciation goes to my parents Yalçın and Nilüfer Albayrak for always supporting every decision I make and providing everything I need, and my sister Güzide for being my best friend. I thank Ata for being the best companion to me all these years. Lastly, I would like to thank my husband Orgun, who is the reason why I came to Michigan, for being the most caring, loving and supporting person I have ever known. He is the most important thing in my life, and I would not be where I am today without him.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The latest technologic advancements in chemical synthesis of DNA and peptides led to a tremendous increase in protein engineering research. High throughput, economically feasible as well as design-flexible platforms are in much demand to construct and screen large peptide libraries for numerous applications. By combining light-directed *in situ* parallel oligonucleotide synthesis with biological expression systems, a novel method for the construction of large custom designed peptide libraries has been developed in this study.

In this method, fully defined collections of peptides are reverse-translated into oligonucleotides for DNA synthesis, expressed in a cellular host, and then screened for the selection of peptide analogs with desired properties. Each step of this process, from design to screening, was examined and optimized to maintain initial library complexity. A variety of cellular systems, both eukaryotic and prokaryotic, were investigated for their potential as an expression host.

An application to discovery of antimicrobial peptides has been demonstrated by constructing two mutant libraries coding for derivatives of Pediocin PA-1 and Plantaricin-423. Screening of these libraries resulted in identification of mutant peptides with greater activities as well as substantial knowledge on sequence-related activity that can be used to generate subsequent libraries to further optimize these peptides against selected targets. Therefore, this method presents great potential for countless applications from discovery of therapeutic peptides to gaining fundamental understanding of their biological function and characteristics.

CHAPTER 1

INTRODUCTION AND MOTIVATION

Recent developments in synthetic biology and high throughput technologies combined with advanced molecular biology techniques have resulted in substantial advances in the field of genomics, proteomics, and related disciplines. The ability of chemically synthesizing biological molecules, such as DNA and peptides, in large quantities allowed researchers to construct collections of engineered molecules and isolate the ones with desirable properties. These combinatorial libraries can be constructed using various biological and synthetic techniques, and screened by both *in vivo* and *in vitro* assays for the discrimination of library members on the basis of their desired biological properties.

Nearly all biological processes are controlled by various kinds of molecular recognition, and peptides and proteins are a major part of this control process due to their immense diversity and unique biochemical properties. Many naturally occurring peptides function as hormones, immunity barriers, neurotransmitters, growth factors, and so on. These physiological roles make them a very important class of molecules for investigating their potential as therapeutics and biomedical compounds. Peptides can be tailored based on their structural and biophysical

characteristics to improve their biological activity and stability, target specificity and affinity, and reduce their toxicity against animal cells.

A peptide library is a collection of systematic combinations of peptides in large numbers. Peptide libraries have numerous applications in drug discovery, medicinal chemistry, and physical sciences. They are often employed in generation of synthetic vaccines, identification of bioactive peptides, characterization of ligand-binding activities, discovery of antimicrobial peptides, and epitope mapping. The sequence of each library peptide can either be randomly generated or custom-designed based on its intended use and the technique that is used to construct the library.

Peptide libraries can be constructed by using both synthetic chemistry tools and biotechnological approaches, and screened by various high-throughput techniques for a wide variety of functions. Libraries may contain from thousands to millions of peptides, however, the size of the library is often inversely related with the control over each peptide sequence. The techniques that are suitable for large library constructions are much less flexible in designing the sequence of each entity in the library.

This introductory chapter first gives an overview of both biological and synthetic library construction methods that are available today with their advantages and limitations, and then continues with explaining the motivation and objectives of the current study. Next, background information on the peptide family that was chosen to demonstrate the applicability of the developed method is provided, and the chapter is concluded with project description.

1.1 Current Methods for Peptide Library Construction

1.1.1 Biotechnological Methods

A prerequisite for screening large peptide libraries is the ability to subsequently identify those members of the library with desired properties. One solution to this problem involves the creation of a linkage between genotype and phenotype, so that after the selection process, the isolated peptide can be identified and characterized via its encoding DNA or RNA. In nature, this link is formed by cellular compartmentalization. Each microbial cell produces one type of peptide while containing the genetic information within its DNA inside the cell. Besides cellular expression and secretion systems, techniques based on the same theory such as phage display, plasmid display, and ribosome display have been commonly used for peptide library screening. Biological construction systems rely on generating diverse libraries at DNA level coding for random peptides or analogs of a peptide of interest, and various strategies based on polymerase chain reaction (PCR) are discussed below.

Random Mutagenesis

This method relies on error-prone PCR (EP-PCR) for introducing random point mutations- substitutions, deletions, and insertions-, at random positions of a defined segment of DNA when performed under reduced replication fidelity conditions [1]. The error rate of the DNA polymerase used during amplification can be artificially increased by using unequal dNTP concentrations, changing buffer composition, and adding unnatural nucleotide analogs [1, 2]. The only user-controlled parameter in this method is the average number of mutations per DNA fragment by fine-tuning reaction conditions and the number of EP-PCR cycles performed [3].

While EP-PCR is very useful in applications where no information is available for direct mutagenesis of particular sequence positions that are linked to the phenotype, the mutations introduced by EP-PCR are biased at the protein level. EP-PCR mostly introduces one nucleotide mutation per codon; however, most amino acid changes require at least two nucleotide mutations. Wong et al. (2006) showed that only between three to seven amino acid substitutions per amino acid occur (depending on the wild-type codon) during EP-PCR, and harmless multiple mutations are detected very rarely since increasing mutagenesis rate also increases the risk of detrimental mutations in the same clone [4]. One application of random mutagenesis to construct peptide libraries is the study where Thie et al. (2009) described a method for antibody maturation based on random mutagenesis and phage display. A mutant antibody gene library was generated by performing several rounds of error-prone PCR, and screened by a panning procedure to select for increased affinity of recombinant scFv antibody fragments against the target [5].

Site-saturation Mutagenesis

In this method, peptide of interest is mutated at selected positions using degenerate oligonucleotides during PCR amplification. A set of oligonucleotides are synthesized *in vitro* where the wild-type codon to be mutated is replaced by a degenerate codon described as NNN, NNK or NNS [6, 7]. These degenerate codons represent equimolar mixtures of nucleotides where "N" stands for A, T, G, and C, "K" stands for G and T, and "S" stands for G and C. While each of these codon patterns represents all 20 naturally occurring amino acids, NNK and NNS encode for only one translational stop codon as opposed to three stop codons by NNN, which is advantageous for reducing the proportion of truncated peptides in the library. Although site-saturation mutagenesis allows for positional control of user-defined

mutations and represents all amino acids, it still presents some codon bias due to the degeneracy of the genetic code as several amino acids are encoded by more than one codon. Since NNK and NNS represent 32 codons and NNN represents all 64 codons, amino acids are encoded with different probabilities; hence, some of them become overrepresented in the final peptide library [8].

The major limitation of site-saturated mutagenesis is the size of the physical library to be screened. The number of clones to be analyzed exponentially increases with the number of mutated positions, and becomes overwhelming since any library should be over-sampled to statistically ensure that the entire library is captured in the screening process. Nevertheless, this method is commonly used for directed evolution studies where different approaches are employed to limit the number of diversified positions in the library [9, 10].

In a study by Tominaga et al. (2006), a mutant library coding for analogs of an antimicrobial peptide, Pediocin PA-1, was generated by NNK scanning to determine the essential and nonessential residues responsible for peptide's activity. The DNA fragment coding for Pediocin PA-1 was amplified using PCR primers with partial random sequences where desired native codon was replaced with the NNK triplet oligonucleotide to generate mutagenized DNA library. Each PCR product was separately expressed in *E. coli*, and 35 clones per position were selected and analyzed by colony overlay assay and DNA sequencing. The authors were able to identify both essential and variable residues of Pediocin based on the mutant pool that they screened, however, only 11 to 15 different amino acids were represented at selected positions to be sequenced due to codon degeneration [11].

Hulley et al. (2010) developed an automated anaerobic procedure for rapid screening of enzyme variants generated by single site-saturated mutagenesis. Several selected residues of PETN (pentraerythritol tetranitrate) reductase, which is

a target enzyme in industrial biocatalysis, were mutated by PCR using NNK degenerate oligonucleotides, and library products were cloned in *E. coli* for expression. From each library of transformants, 96 clones were individually picked and assayed in 96-deep well plates under anaerobic conditions to screen for improvements in steady state reaction rates against a variety of typical OYE (old yellow enzyme family) substrates. Authors were able to select for two mutants with a switch in substrate enantiopreference against the substrates that were used [12].

DNA Shuffling

DNA shuffling is a recombination-based library construction method where multiple homologous genes are fragmented by DNase I and then reassembled using PCR in which the fragments are utilized as primers for each other [13, 14]. Hence, the final library consists of chimeric genes that contain sequence regions from multiple sources displaying different phenotypes from their parents. The parent genes can either be related genes of the same family or often products of EP-PCR generated mutant library.

In most studies, DNA shuffling and EP-PCR are used in combination for more global search of mutant libraries and the mutants identified by these techniques are subjected to site-directed mutagenesis for an extensive search for the best-fit residue. For example, Miyazaki and co-authors (2006) applied all three mutagenesis methods to improve thermal stability of a xylanase enzyme (glycosyl hydrolase family-11 of *Bacillus subtilis*) by direct evolution. First, error-prone PCR was carried out to mutagenize entire xylanase M169I gene. Several hundred clones were screened for thermostability and five thermostable variants were identified and used for recombination experiments by DNA shuffling in two additional separate libraries. After another round of thermostability screening, three positions were

selected to be further optimized by site-saturated mutagenesis which led to identification of a single variant with much improved thermal stability [15].

1.1.2 Synthetic Peptide Libraries

When full control over the entire sequence of the library is required, chemical synthesis methods are often employed to generate custom collections of peptides. To be able to link assayed peptides to their biophysical properties, several spatially addressable approaches have been developed where the synthesis and screening of libraries are performed in parallel. These methods eliminate the need for subsequent sequence and structure determination of the positive peptide [16]. However, maintaining this linkage limits the number of peptides in each library that can be synthesized and screened using these *in vitro* platforms which is usually around 10^3 to 10^4 in most high-throughput systems, compared to 10^7 to 10^{10} with biological systems.

Despite being totally flexible in sequence design, and allowing for the incorporation of unnatural amino acids into peptide sequences, there are two main limitations of synthetic methods: first, the synthesis efficiency is limited by the peptide length which is often shorter than 20 amino acids, and second, synthesizing a large library can become quite costly, especially when conventional solid phase synthesis is used. Spatially addressable techniques such as multi-pin synthesis [17], digital photolithography [18], tea bag synthesis [19], and SPOT synthesis on cellulose [20, 21] are all relatively cost-effective and more rapid alternatives to standard synthesis, however SPOT synthesis has become more popular in recent studies and will be explained in more detail within this context.

SPOT Synthesis

SPOT technology, developed by Frank, R. (1992) follows standard Fmoc chemistry used in the conventional solid phase synthesis, but instead of using resin as the solid support, the synthesis occurs on more economical and stable cellulose filter sheets [20]. The optimum range of peptide length that can be synthesized using SPOT technology is 6-18 amino acids and up to 1000 peptides per 20x29cm membrane can be synthesized in an addressable manner for solution-phase assays using a pipetting robot [22, 23]. However, as the number of peptides per sheet increases, the volume of each peptide solution decreases, hence more sensitive screening methods are required for analysis.

Hancock and his group (2005) combined SPOT synthesis with a sensitive screening assay based on luminescence inhibition for the generation of short cationic antimicrobial peptides with improved activity. A complete substitution library of a 12 amino acid-long bactenecin peptide, a total of 228 peptides, was generated on cellulose membranes and the activity of each variant was screened by incubating the peptides with a luminescence-producing indicator bacteria. The degree in luminescence inhibition was correlated to an increase in antimicrobial activity of the mutant peptides. The authors of this study estimated that production and screening of 50,000 peptides per pipetting robot per year is possible using this technique [22].

While the ability to generate a synthetic library that contains thousands of peptides may present great potential in many studies where peptides of interest are short in length and require specific sequence design, it is not nearly close to the number of peptides that can be screened using biological construction systems.

1.2 Project Introduction

As explained above, current methodologies used for the construction of peptide libraries present both advantages and disadvantages when it comes to sequence design, peptide length, or the library size. In general, biological techniques are inexpensive, each library can contain as high as 10^{10} entities, and the length of peptides is not restricted in most systems; however, peptide sequences are generated randomly and only L-amino acids can be incorporated into biological libraries. Synthetic methods, on the other hand, allow for full control over each peptide sequence, there is more variety to screening assays, but the peptide length as well as the library size is limited due to high cost and labor associated with complex synthetic chemistry.

A platform that offers most advantages of both technologies; flexibility of the sequence design with the ability of producing large libraries inexpensively, would allow researchers to study fully defined libraries in a high-throughput manner. The motivation for this project, therefore, was to develop a novel method for the construction and screening of large custom peptide libraries. The main objective of the study was to combine light-directed *in situ* parallel oligonucleotide synthesis with biological expression and screening systems.

The parallel oligonucleotide synthesis technology (developed by Gao *et al* 2001) allows for each entity of the library to be custom-designed and is suitable for the maskless synthesis of tens of thousands of oligonucleotides on a single array in a very cost-effective way. Libraries of oligonucleotides can be cloned in a variety of engineered microbial expression systems to construct custom peptide libraries where each host cell works as a separate "bioreactor" to produce a particular peptide. These peptides can be individually screened for their specific properties,

such as antimicrobial activity or affinity, and selected ones can easily be identified by sequencing of the coding DNA for further studies.

To demonstrate the feasibility of the developed method, custom-designed peptide libraries encoding for antimicrobial peptides (AMPs) were constructed and screened for the discovery of novel peptides with improved activities. AMPs are a very special family of peptides with their unique physical properties and essential functions in living organisms. There is tremendous amount of research focused on understanding their role in innate immunity and interaction with cellular systems, studying their structure-function relations, and gaining more insight to their mechanisms of antimicrobial activity with one ultimate goal; developing novel AMPs for therapeutic, biomedical or biotechnological applications. Therefore, this peptide class was selected to be investigated in this project.

1.3 Antimicrobial Peptides Background

Antimicrobial peptides are major components of the innate immune system in multicellular organisms. These gene-encoded and ribosomally synthesized peptides are generally 10 to 50 amino-acids long with a broad-spectrum native microbicidal activity and a range of immune-modulatory functions [24, 25]. They play an important role as a first line of defense virtually in every life form. AMPs show extreme diversity in their sequence, size and structure, but they all share two functionally important properties: an overall positive charge (up to +9), and a high proportion of hydrophobic residues [26]. Due to their diversity in nature, they are broadly classified on the basis of their secondary structure: linear, α-helical peptides with no cysteines, such as cecropins and magainins, β-sheet peptides that contain two or three disulfide bridges, such as defensins, loop peptides with one disulfide

bridge such as bactenecin, and peptides that are rich in certain residues (tryptophan, proline, glycine and arginine), such as indolicidin and drosocin [27].

1.3.1 Mechanism of Action of AMPs

AMPs are able to kill a variety of microbes including both gram positive and gram negative bacteria, fungi, and enveloped viruses mostly via non-receptor mediated mechanisms [28]. The minimal inhibitory concentrations of most effective peptides range between 0.1-10μg/ml [27]. Antimicrobial activity mostly relies on the interactions between the peptide and surface molecules of the target organism. Their net positive charge permits for initial electronic attractions to negatively charged membrane molecules and leads to accumulation of peptides at the anionic cell surface. This is followed by penetration through the cell membrane and displacement of the phospholipids, hence, disruption of the physical integrity of the membrane, leakage of internal cell components, and eventual cell death. This process, along with several other mechanisms, is depicted in Figure 1.1. [27-29].

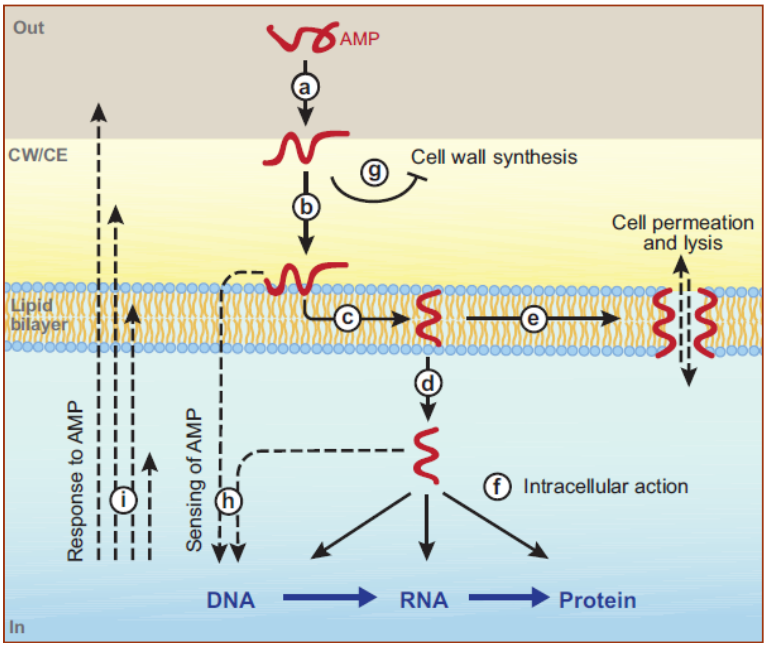

Figure 1.1- Mechanisms of action of AMPs. [29]

The interactions between target cell membrane and AMPs have been studied using variety of techniques, such as solid-state NMR spectroscopy, ion channel formation analysis, circular dichorism, and microscopy, to assess the mechanisms of antimicrobial activity [28]. There are three main mechanisms proposed for membrane permeation and cell lysis by AMPs (Step "e" in Figure 1.1).

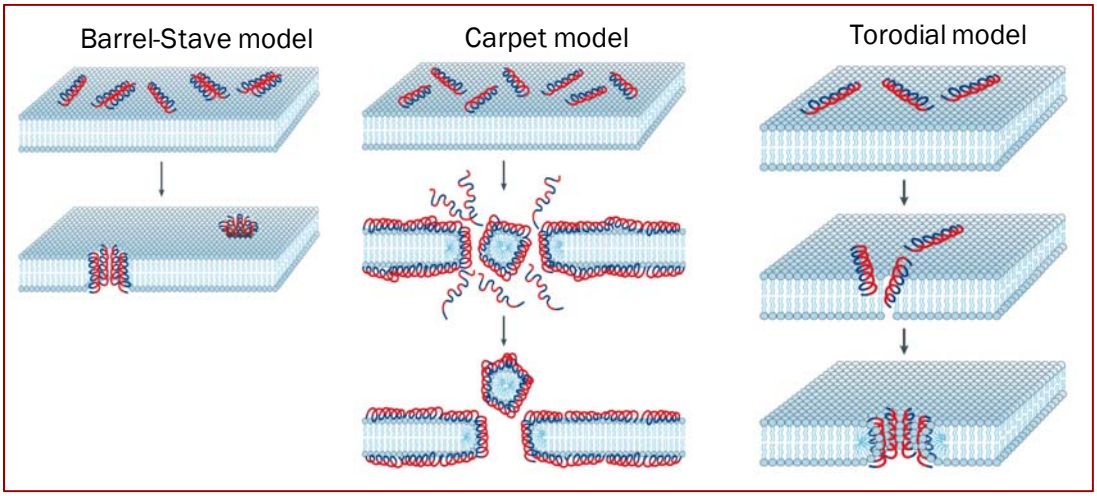

Figure 1.2- Three proposed pore formation mechanisms of AMPs. (Adapted from [28]).

In the barrel-stave model, peptides aggregate and insert into the membrane bilayer to form a transmembrane channel by aligning their hydrophobic regions with the lipid core while their hydrophilic regions are facing inside the pore. These pores can contain 3-11 parallel helical peptide aggregates. The carpet model describes the accumulation of peptides on the bilayer surface in a carpet-like manner with an orientation that is parallel to the membrane. At critical peptide concentrations, the integrity of membrane collapses due to formation of transient holes and disruption of the bilayer curvature. The torodial pore model proposes that antimicrobial peptides insert into the membrane and affect the local curvature of the bilayer by interacting with the polar head groups of lipids. The lipids in these

holes opened by peptides form a continuous bend from the top to the bottom of the membrane forming torodial pores [27, 28].

1.3.2 Bacterial Resistance to Antimicrobial Peptides

To date, studies on AMPs have brought researches to the consensus that the mode of the action of these peptides has played a key role in the success of AMPs during evolution [27, 30-32]. Cationic AMPs target bacterial membrane at micromolar concentrations, and impair its functions and membrane-associated biological processes. Amphipathic properties of AMPs allow them to interact with numerous targets both inside and outside of the membrane leading to the breakdown of its integrity and eventually the loss of cell viability. However, most conventional antibiotics target a specific molecule with a single mode of action, and function at low concentrations, allowing bacteria to develop resistance. It has been argued by Sahl et al. (2006) that the innate immune systems has evolved in a way that AMPs with multiple antimicrobial activities are able to inhibit many biological processes with low potency, so it is not favorable for bacteria to generate resistance, rather than blocking a single target or a specific pathway. For example, Nisin with five different modes of antimicrobial action has been used as a food preservative for many years, and yet no significant bacterial resistance has been generated[33].

Recent work has indicated that AMPs have a broad range of immuno-modulatory functions such as chemo-attraction of immune cells to the infection site, induction of gene expression and protein secretion in epithelial cells, promotion of wound healing, and anti-endotoxin (LPS) activity during infection. Hancock and his colleagues (2006) speculated that AMPs being both antimicrobial and modulator of the innate immune system has been the key strategy that nature has followed to prevent pathogens from developing resistance. Since drugs that boost innate

immunity would not put a direct selective pressure on microorganisms, and the immuno-modulatory properties of AMPs would not be affected by antimicrobial resistance, these peptides offer a very promising approach to combat antibiotic resistance, and are worth further investigations as potential antibiotic-replacements for future [26].

1.3.3 Therapeutic Potential of Antimicrobial Peptides and Other Applications

There is an urgent need for the discovery of novel anti-infective drugs due to the increasing problem of multidrug-resistance in pathogenic bacteria. The broad activity spectrum of AMPs and their immuno-modulatory properties make these peptides very attractive candidates for the development of novel human therapeutics. The potential therapeutic applications of antimicrobial peptides include: **1**- as single anti-infective agents, **2**- as anti-tumor agents, **3**- in combination with conventional antibiotics to increase the potency by facilitating the penetration of drug molecules through the cell wall into the target cell, **4**- as immuno-modulatory agents that enhance natural innate immunity, and **5**- as endotoxin-neutralizing agents [27, 34-36].

In addition to their therapeutic usage, AMPs also can be used in other applications such as for lining of synthetic polymeric materials to prevent microbial colonization and growth (magainin), usage as imaging probes for fungal and bacterial infections (ubiquicidin), as food preservatives (nisin), and in agricultural industry to provide disease resistance to plants and animals [29, 37].

1.4 Project Description

The method developed in this study involves five main steps: peptide and DNA library design, oligonucleotide synthesis, amplification of the library, cloning and expression in biological host systems, and finally screening for the desired properties of the peptides. Each step of this process directly affects the "success" of the library, meaning that the library being screened at the last step actually contains the peptides that were designed initially, and were investigated separately to maintain the complexity of the peptide library.

All peptide analogs in the library are custom-designed and then their amino acid sequences are reverse-translated into codon-optimized oligonucleotide sequences. After oligonucleotide synthesis and purification, the library is amplified by polymerase chain reaction to prepare sufficient material for subsequent cloning experiments. To preserve library complexity during amplification, an alternative PCR technique was evaluated and the methodology was further optimized to minimize the formation of PCR artifacts.

Next, amplified library is cloned into a host cell, which is determined based on the application and type of the peptides being studied, and the library oligonucleotides are translated into peptides and often secreted out of the cells for screening. A variety of host expression systems were investigated and engineered to improve their expression yields in this study; yeasts *Saccharomyces cerevisiae* and *Pichia pastoris* as eukaryotic hosts, and *Escherichia coli* as a bacterial host. The final step in this process is the screening of the library, and it is completely application-dependent like the expression host. Assays like colony-overlay method to screen for antimicrobial activity were re-designed accordingly and performed for the screening of the libraries constructed during this project.

The next chapter of this document discusses the first three steps of the method and outlines the experimental procedures developed during this work. These steps, library design, synthesis, and amplification, can be applied for the construction of any peptide library with little modifications based on the application. This is followed by a chapter describing two yeast expression systems that were employed for the production of antimicrobial peptides and explains the findings of this study on their expression capabilities. Chapter 4 discusses all the results gathered from the construction and screening of two mutant AMP libraries in a bacterial expression system, which is followed by in-depth analysis of the effects of the mutations on antimicrobial activity, and the behavior of the selected peptides in a high-salt environment. Finally, the last chapter provides a summary of conclusions and an outline of future work.

CHAPTER 2

CONSTRUCTION OF PEPTIDE-ENCODING OLIGONUCLEOTIDE LIBRARIES

2.1 Introduction

Construction of any peptide library originated from synthetic DNA sequences involves three major steps; library design, oligonucleotide synthesis, and amplification by PCR. Preparation of a library by following these three steps is independent of the specific application that the peptides are designed for. This chapter describes in detail how custom-designed peptide libraries studied in this project were translated into oligonucleotide libraries, and prepared for expression and screening in cellular hosts.

2.2 Peptide and DNA Library Design

A peptide library to be constructed using parallel DNA synthesis and cellular expression systems may contain any custom-designed peptide that is of interest to study. The peptide sequences of the library can either be designed one by one or custom mutations can be systematically introduced to a peptide of interest. The number of peptides in the library can vary from hundreds to hundreds of thousands based on the scope of the study. Once the peptide library is custom-designed, the amino acid sequences are reverse-translated into nucleotide sequences using in-house developed software (by Dr. Jean-Marie Rouillard) following two optimization

parameters as explained below. The details about the design of the peptide libraries studied in this project were separately explained in the related chapters.

2.2.1 Codon optimization for Expression

Expression of functional proteins in heterologous hosts often starts with the optimization of the transcript gene using codon usage bias of the host organism. This is because the translational yield of a protein is directly related to the balance between the codons present in the gene and the abundance of the cognate tRNAs available in that organism [38, 39]. The degeneracy of the genetic code allows for many amino acids to be encoded by more than one codon, in fact, 3 amino acids (Arg, Leu, and Ser) are encoded by six synonymous codons. The frequencies of synonymous codon usage in coding DNA vary significantly between species [40, 41], and this difference in the frequencies is referred to as 'codon usage bias' of that particular organism. Fast-growing organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae* present the most pronounced codon usage bias (as a result of natural selection for translational optimization) that reduce the diversity of tRNA species to minimize the metabolic load under rapid growth conditions [42].

Generally, if the gene of interest contains codons that are rarely used by the expression host, translational yield of the heterologous protein will be much lower than reasonable levels [43]. To prevent such limitations due to non-optimal codons present in the nucleotide sequence of the protein, the target gene is optimized by altering the rare codons to more closely reflect the codon usage of the host organism. This is usually done according to the codon usage table of the host that contains average frequency values of each 64 codons (Table 2.1) [44].

Table 2.1- Codon Usage in *E. coli* and *S. cerevisiae*. **RSCU**, relative synonymous codon usage, is the value of observed frequency of a codon divided by the frequency expected under the assumption of equal usage of the synonymous codons for an amino acid. **W**, relative adaptiveness of a codon, is the frequency of the optimal codon for that amino acid [44].

	E.coli		Yeast			E.coli		Yeast	
	RSCU	w	RSCU	w		RSCU	w	RSCU	w
Phe UUU	0.456	0.296	0.203	0.113	Ser UCU	2.571	1.000	3.359	1.000
UUC	1.544	1.000	1.797	1.000	UCC	1.912	0.744	2.327	0.693
Leu UUA	0.106	0.020	0.601	0.117	UCA	0.198	0.077	0.122	0.036
UUG	0.106	0.020	5.141	1.000	UCG	0.044	0.017	0.017	0.005
Leu CUU	0.225	0.042	0.029	0.006	Pro CCU	0.231	0.070	0.179	0.047
CUC	0.198	0.037	0.014	0.003	CCC	0.038	0.012	0.036	0.009
CUA	0.040	0.007	0.200	0.039	CCA	0.442	0.135	3.776	1.000
CUG	5.326	1.000	0.014	0.003	CCG	3.288	1.000	0.009	0.002
Ile AUU	0.466	0.185	1.352	0.823	Thr ACU	1.804	0.965	1.899	0.921
AUC	2.525	1.000	1.643	1.000	ACC	1.870	1.000	2.063	1.000
AUA	0.008	0.003	0.005	0.003	ACA	0.141	0.076	0.025	0.012
Met AUG	1.000	1.000	1.000	1.000	ACG	0.185	0.099	0.013	0.006
Val GUU	2.244	1.000	2.161	1.000	Ala GCU	1.877	1.000	3.005	1.000
GUC	0.148	0.066	1.796	0.831	GCC	0.228	0.122	0.948	0.316
GUA	1.111	0.495	0.004	0.002	GCA	1.099	0.586	0.044	0.015
GUG	0.496	0.221	0.039	0.018	GCG	0.796	0.424	0.004	0.001
Tyr UAU	0.386	0.239	0.132	0.071	Cys UGU	0.667	0.500	1.857	1.000
UAC	1.614	1.000	1.868	1.000	UGC	1.333	1.000	0.143	0.077
ter UAA	--	--	--	--	ter UGA	--	--	--	--
ter UAG	--	--	--	--	Trp UGG	1.000	1.000	1.000	1.000
His CAU	0.451	0.291	0.394	0.245	Arg CGU	4.380	1.000	0.718	0.137
CAC	1.549	1.000	1.606	1.000	CGC	1.561	0.356	0.008	0.002
Gln CAA	0.220	0.124	1.987	1.000	CGA	0.017	0.004	0.008	0.002
CAG	1.780	1.000	0.013	0.007	CGG	0.017	0.004	0.008	0.002
Asn AAU	0.097	0.051	0.100	0.053	Ser AGU	0.220	0.085	0.070	0.021
AAC	1.903	1.000	1.900	1.000	AGC	1.055	0.410	0.105	0.031
Lys AAA	1.596	1.000	0.237	0.135	Arg AGA	0.017	0.004	5.241	1.000
AAG	0.404	0.253	1.763	1.000	AGG	0.008	0.002	0.017	0.003
Asp GAU	0.605	0.434	0.713	0.554	Gly GGU	2.283	1.000	3.898	1.000
GAC	1.395	1.000	1.287	1.000	GGC	1.652	0.724	0.077	0.020
Glu GAA	1.589	1.000	1.968	1.000	GGA	0.022	0.010	0.009	0.002
GAG	0.411	0.259	0.032	0.016	GGG	0.043	0.019	0.017	0.004

In this project, codon optimization based on the expression host organism was done using Table 2.1. The oligonucleotide libraries were generated by translating the amino acid sequences of the library peptides into nucleotide sequences using the most abundant codon for each amino acid. This process is illustrated in Figure 2.1 for the short antimicrobial peptide IsCT-[L_6K_{11}], where its DNA sequence was codon-optimized to be expressed in two different expression hosts.

	I	**L**	**G**	**K**	**I**	**L**	**K**	**G**	**I**	**K**	**K**	**L**	**F**
S. cerevisiae :	ATC	TTG	GGT	AAG	ATC	TTG	AAG	GGT	ATC	AAG	AAG	TTG	TTC
E. coli :	ATC	CTG	GGT	AAA	ATC	CTG	AAA	GGT	ATC	AAA	AAA	CTG	TTC

Figure 2.1- Optimization of IsCT-[L_6K_{11}] coding sequence for expression in *S. cerevisiae* and *E. coli*. Most abundant codons for Leu and Lys are different between these two organisms.

2.2.2 Codon Modification for DNA Synthesis

When the oligonucleotide sequences were generated by modifying the codons, selection of the replacement codon was performed according to two criteria; first, using the most abundant codons of the host organism for optimum expression (as explained in the previous section), second, avoiding same base repetitions in the sequence to minimize errors during synthesis as well as downstream processes. If the most abundant codon for a particular amino acid leads to a nucleotide repetition in the sequence, then the second most abundant codon is used in that position. This rule applies mostly for guanine; previous studies showed that oligonucleotides containing guanine-rich repeats can self-associate in aqueous

solutions to form four-stranded tetraplex structures [45]. This particularly becomes an issue once the oligonucleotides are cleaved off the chip after synthesis where they freely interact with each other and can form highly ordered structures [46].

Once the oligonucleotides were optimized both for expression and synthesis, each library was synthesized on glass arrays as explained in the following section.

2.3 Light-directed Parallel Synthesis of Oligonucleotides

Oligonucleotide libraries were synthesized on microarrays using parallel DNA synthesis technology developed by Gao et al. (2001) and adapted to open substrates by Gulari et al. [47]. Massively parallel oligonucleotide synthesis was carried out on the surface of amino linker-derivatized glass slides inside a cartridge by using the conventional phosphoramadite chemistry with one modification where TCA (Tricholoroacetic acid) at the deprotection step is replaced by photogenerated acid (PGA) for the light-directed removal of the DMT (dimethoxytrityl) terminal protective group (Figure 2.2) [48]. Activation of the photogenerated reactions at each location is controlled by digital photolithography using a digital micromirror device (DMD) made of a matrix of 768 x 1024 individually addressable mirrors. Use of digital photolithography allows for control at each reactive site, extreme flexibility and feasibility for the synthesis of tens of thousands of custom-designed oligonucleotides on the same array at a reduced cost compared to mask-based conventional synthesis chemistries [49].

Synthesis of oligonucleotides is carried out by a stepwise addition of monomers to the 5'-terminus of the growing chain until the desired sequence is completed. Each synthetic cycle consists of four chemical reactions: Detritylation (or deprotection), coupling, oxidation, and capping. The deprotection of the protected monomer is initiated by filling the reaction cell with PGA-precursor solution. A

digital light pattern that was generated according to pre-determined array layout is then reflected at the selected sites to induce PGA production, hence, removal of the terminal protective group from the monomer. A solution of appropriate monomer is added to the array, and the free OH groups are coupled with monomers. After oxidation and capping of unreacted monomers, addition of the new residue to the growing chain is completed. The cycle of four reactions is repeated until the full size oligonucleotide is synthesized. This technology allows for the synthesis of oligonucleotides up to 200mer-long with 99.5% step-wise yield and minimum depurination.

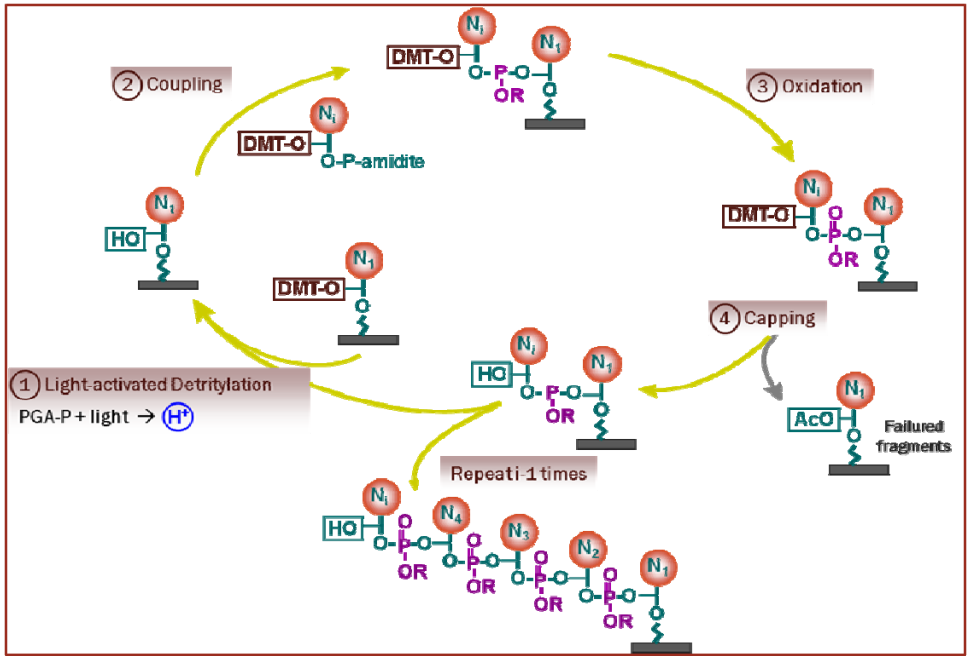

Figure 2.2- Light-directed parallel oligonucleotide synthesis.

The spot uniformity achieved during the synthesis is shown in Figure 2.3. The upper right panel of the image shows a cross-section of a spot. As seen in that image, spots have a highly uniform morphology with very sharp edges. The area

between red lines on the graph represents the full-size oligos synthesized on the chip, while the small areas outside both sides of these lines corresponds to a few oligos that were partially-synthesized on each spot.

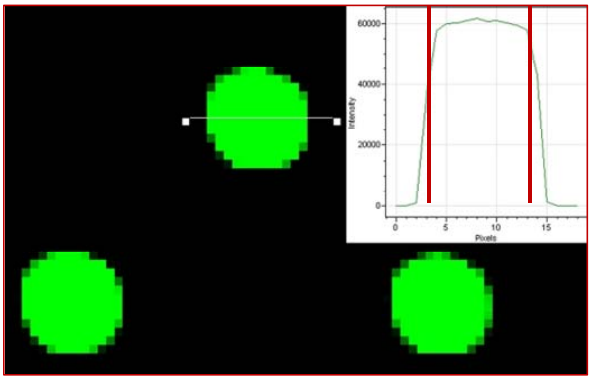

Figure 2.3- Spot uniformity achieved during oligonucleotide synthesis. 45-mer probes were coupled with a synthetic Cy3-labeled target.

Presence of deficient oligos in the synthesis products affects two downstream experiments; first, these shorter oligos are also amplified along with complete oligos during emulsion PCR, and depending on the synthesis yield they may lead to smears when the PCR products are visualized on an agarose gel, second, after cloning of the amplified library products in an expression host, some of the transformed colonies carry these shorter oligos, coding for inactive peptides, and should be considered during screening experiments.

2.4 Amplification of Oligonucleotide Libraries by Emulsion PCR

The final step in the construction of oligonucleotide libraries is the amplification of the synthesis products to generate enough double-stranded DNA material for the following cloning experiments. Uniform amplification of complex DNA libraries by conventional PCR has always been challenging due to formation of PCR-generated artifacts such as chimeric molecules, point-mutations, and

heteroduplexes [50, 51]. When a library of sequences with homologous regions is used as a template for PCR, the amplification efficiency is affected by two factors: first, short fragments are preferentially-amplified over longer fragments, and second, chimeric fragments are generated as a result of recombination between homologous regions of template DNA molecules [50, 52]. These PCR-mediated recombinations occur when a proportion of incompletely extended primers anneal to another template sequence, and amplified during a later cycle [50].

To prevent cross-recombination events between template DNA fragments and reduce competition between fragments of different length, an alternative PCR method, called emulsion PCR, has been developed based on the compartmentalization of each PCR reaction [52, 53]. In theory, a single DNA molecule is the minimum amount required for PCR, and if only one template fragment is present in each compartment, a single amplification will occur without interacting with other template molecules. By following this theory in emulsion PCR, individual fragments are distributed in aqueous droplets of water-in-oil emulsions and amplified by PCR in isolation. Each emulsion droplet behaves like an independent "microreactor", and contains all the reagents required for PCR, such as dNTPs, primers, buffer, and the enzyme along with a single, or at most a few, template DNA molecules [54]. The number of DNA molecules in each droplet can be modeled by Poisson distribution which provides a model for the probability of a number of random occurrences in a fixed period of time or volume [53].

Emulsion PCR reduces the formation of artifactual molecules and enables equal-amplification of each DNA fragment, thus preserving library complexity. Therefore, a modified version of the emulsion PCR protocol was employed for the amplification of highly-homologous oligonucleotide libraries used in this study.

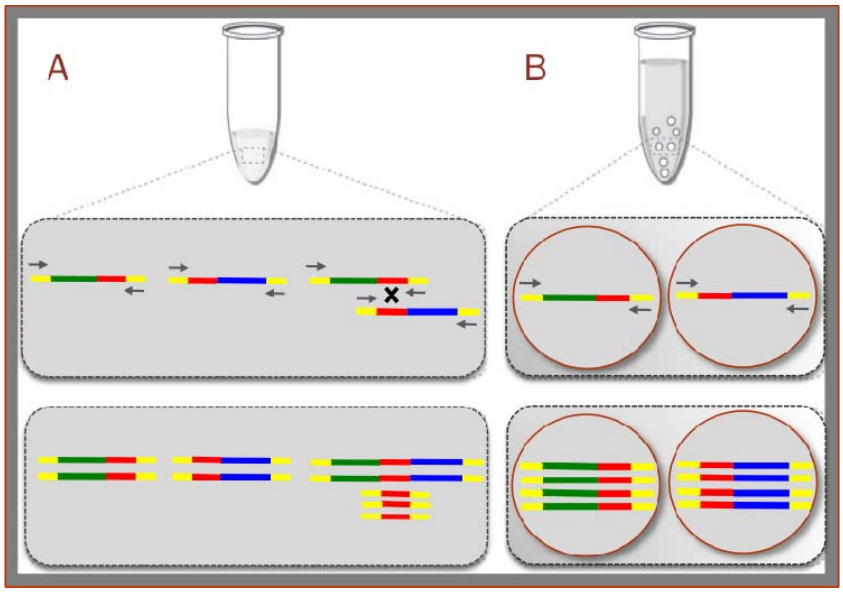

Figure 2.4- Amplification of complex DNA libraries (adapted from Ref.-[52]). **A.** Conventional PCR. **B.** Emulsion PCR.

2.4.1 Emulsion PCR Procedure and Results

Library oligonucleotides were amplified by emulsion PCR by following the protocol developed by Williams et al. (2006) with some modifications. The complete step-by-step procedure is provided in Appendix 1.1. Briefly, an aliquot of the oligonucleotide library was mixed with a solution containing 100pmoles of each primer, 6mM $MgCl_2$, 2mM dNTPs, 0.5g/l BSA, and 10 units of Phusion HF DNA Polymerase in a final volume of 100µl. The PCR mix was emulsified by addition to 600µl oil-surfactant mixture, and stirring for 10min at 1000rpm on a magnetic stirrer in an ice-cooled glass vial. Emulsified PCR mix was recovered from the glass vial, and distributed to small PCR tubes in 50µl aliquots. This procedure described above generates emulsion droplets in the 8-10 microns diameter range. Figure 2.5 below shows an agarose gel image of an oligonucleotide library amplified by both conventional PCR and emulsion PCR. As clearly shown, amplification of an oligonucleotide library by conventional PCR resulted in formation of nonspecific products, that appear as a big smear on the gel. However, when the same library

was amplified by emulsion PCR, the band on the gel corresponding to PCR products was very sharp and at the expected size with respect to the DNA ladder, indicating specific amplification of the library fragments.

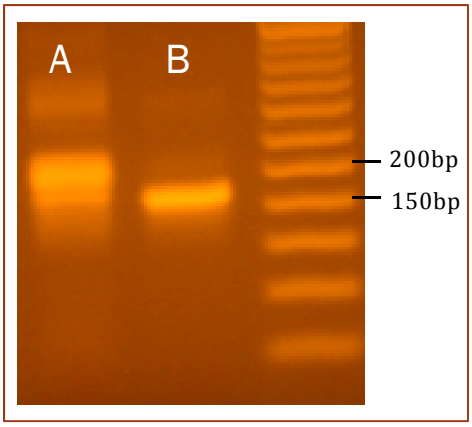

Figure 2.5- Agarose gel image of DNA library amplification. The length of the oligonucleotides was 170bp. **A.** Products from conventional PCR. **B.** Products from emulsion PCR.

The number of template fragments per droplet depends on the concentration of the DNA library that was used to prepare emulsions. Due to impurities (i.e. partially synthesized oligos) that come with the full-length oligonucleotides after synthesis, it is difficult to estimate the exact DNA concentration of the library, thus the amount of initial template DNA required for a desired number of templates per droplet is calculated based on the approximation of the actual DNA concentration present in the synthesis products. It should be noted here that for all emulsion PCR experiments performed during this project, an assumption of 100-fold difference in the measured DNA concentration versus the actual full-size oligonucleotides present in the library aliquot was applied to the calculation of the template amount for each reaction. Therefore, the amount of template DNA shown in the results of this work were most likely in picogram levels rather than in nanograms as measured by spectrophotometry and included in the text. Based on the calculations using

reaction volume, droplet volume, and total number of template fragments in a given library concentration (using an in-house developed excel program), the number of template fragments per droplet was estimated to be between 1 and 10.

Optimization of the Emulsion PCR Parameters

Initial experiments on screening of AMP-encoding oligonucleotide libraries (oligonucleotide size ranged between 141bp to 162bp) showed that more than half of the clones that were sequenced were actually carrying the wild-type sequence. The mutant libraries were designed to have peptides with a maximum of two mutations, hence, the members of these libraries present 95.7% (dual mutations as a result of 6-base change) to 99.2% (single mutation as a result of one base change) similarities at DNA level. Therefore, since the wild-type sequence itself was not included in the library, cross-recombinations between highly homologous library oligonucleotides resulted in the formation of wild-type sequences at the end of the emulsion PCR. This indicated that the template concentration used for the PCR was too high and too many fragments were encapsulated in the same droplet leading to cross-hybridizations between oligonucleotides.

Determination of template DNA concentration is a tradeoff between using the minimum amount that gives amplification and yielding enough PCR products for downstream cloning experiments. The initial value regarding the template concentration used for emulsion PCR (as well as other parameters shown in Table 2.2) was around 40ng of oligonucleotide library per 50µl reaction volume which produced enough material for cloning when the products of two separate emulsion preparations (hence, 100µl total reaction) were combined. However, as explained in the previous paragraph, this template concentration led to generation of wild-type sequences at a ratio of 56% of all sequenced clones.

To reduce the number of template molecules in each droplet, hence the occurrence of wild-type formation, the template concentration was reduced to 15ng of library per 50µl reaction. At this concentration, the amplification yield was very low for downstream applications; however, use of hot start DNA polymerase (which is inactive at room temperature) allowed for the multiple preparations of emulsions for the same PCR and prevented any non-specific amplification that would otherwise occur at room temperature when a standard polymerase was used. By using much lower template DNA concentration, in combination with Phusion hot start polymerase and slight modifications to reaction settings, the ratio of the wild-type sequences to all analyzed sequences was successfully decreased from 56.0% to 2.0%. These findings agree with other studies in literature. Qui et al. 2001 showed that the percentage of total PCR artifacts decreased as the template concentration decreased and elongation time increased [51]. Results of this study were obtained from sequencing of total of 50 clones in each experiment after cloning the amplified library in *E. coli* and extracting the plasmid DNA for analysis. Summary of these findings is given in Table 2.2, and the optimized parameters shown on the table apply to a library containing oligonucleotides that are around 150bp in length. Of course, these parameters should be optimized for each specific library.

Table 2.2- Summary of emPCR optimization for the amplification of AMP-coding libraries.

	Standard	**Optimized**
Template Conc. (in 50µl)	40ng	15ng
Primer Conc. (in 50µl)	50pmol	50pmol
Enzyme	Phusion HF Standard	Phusion Hot Start
Conditions	95 ºC → 20s, *T*→ 20s, 72 ºC → 20s	98 ºC → 15s, *T*→ 20s, 72 ºC → 30s
% of WT to all clones	**56.0**	**2.0**

2.5 Chapter Conclusion

The process of the construction of the peptide-encoding oligonucleotide libraries was explained in this chapter. This process involves, **1-** peptide library design and generation of optimized oligonucleotide sequences, **2-** synthesis of the oligonucleotide library, **3-** amplification of the synthesis products by emulsion PCR, and can be applied to preparation of any peptide library to be expressed in a biological host.

Library design starts with the custom-design of peptides, then reverse-translating the amino acid sequences into DNA sequences with the consideration of two criteria, codon optimization for peptide expression and codon modification to improve synthesis yield. Although modification of the codon-optimized oligonucleotides is kept minimal to prevent alterations to the peptide activity, one should analyze the consequences by comparing the activities of peptides with and without modifications. Oligonucleotides that are both codon-optimized for bacterial expression and subjected to modifications for synthesis should be separately expressed in the host organism, and the produced peptides should be tested for their activities. Therefore, when analyzing peptide analogs of a library by comparing their activities to that of parent peptide, both WT and MWT versions of the starter peptide should be included as positive controls in screening plates.

The oligonucleotide library synthesis is performed by using light-directed parallel oligonucleotide synthesis based on the conventional phosphoramadite synthesis. Use of photo-generated acid deprotection and digital photolithography in this technology allowed for the highly efficient synthesis of oligonucleotides in a very cost-effective way. The main advantage of this synthesis method is that there is no limit to the number of oligonucleotides present in the library, as tens of

thousands of custom-designed oligonucleotides can be synthesized in a single run at a very high yield.

After the synthesis of the library, cleaved and purified oligonucleotides were amplified by emulsion PCR to prevent the generation of chimeric fragments and non-specific products. Initial experiments with AMP-coding libraries showed that the template DNA concentration is very critical in maintaining library complexity as cross-recombination events between highly-homologous library fragments led to formation of wild-type sequences eliminating any mutation that was introduced during synthesis. Optimized conditions to amplify a library of ~150bp-long oligonucleotides were determined to be using less than 20ng template DNA /50µl reaction to minimize the number of template fragments per droplet, and performing the PCR with a hot-start polymerase to prevent any non-specific amplification during emulsification. This optimized emulsion PCR decreased the ratio of wild-type sequences to all clones from 56.0% to 2.0%. Following emulsion PCR, purified products were ready to be cloned in an expression host for the translation of the oligonucleotides into a peptide library.

Preparation of the oligonucleotide library is the most critical step for a successful peptide library construction; because the characteristics of the library that is being screened at the final step of the project solely depends on what was initially cloned into the host cells. From oligonucleotide design to library amplification, each step of this process should be carefully executed for the designed peptide library to coincide with the final library.

CHAPTER 3

CONSTRUCTION AND SCREENING OF CUSTOM PEPTIDE LIBRARIES IN YEAST

3.1 Introduction to Yeast Expression System

Yeast systems have been extensively utilized for the production of recombinant proteins for many years. Yeast can be grown rapidly on a simple media to high cell density, and are often considered safe host organisms for the production of pharmaceutical proteins. The accessibility of the yeast genome for genetic manipulations, available techniques to introduce exogenous DNA into yeast cells, and the broad knowledge of yeast biochemistry, has led to the development of new methods for analyzing and preparing proteins not only from yeast itself, but also from other organisms [55].

Several post-translational processing mechanisms available in yeast, such as glycosylation, particle assembly, amino terminal acetylation, and proteolytic processing, have allowed the expression of proteins and peptides derived from mammalian species with appropriate authentic modifications. In addition, heterologous proteins can be secreted from specially engineered strains via eukaryotic secretory pathway, and are easily harvested from yeast culture media. The use of either homologous or heterologous signal peptides has allowed authentic

maturation of secreted proteins, as being correctly folded and cleaved, by the endogenous yeast apparatus [55].

There are many yeast strains used for recombinant protein expression, however, *Saccharomyces cerevisiae* and *Pichia pastoris* are the most widely used hosts for the production of small proteins due to their high expression levels, promoter strength, secretion efficiency, and the availability of molecular biology tools and commercial supplies .

This chapter describes the development of yeast expression systems for the production and screening of antimicrobial peptide libraries. The main advantage of using a eukaryotic expression host is that a wide range of antimicrobial peptides can be screened for their activities against gram positive and gram negative bacteria. Both *S. cerevisiae* and *P. pastoris* systems were investigated for their suitability for producing AMPs. The expression levels of functional AMPs, peptide secretion efficiencies, and different activity screening assays were evaluated to identify the most feasible eukaryotic host to screen peptide libraries. At the time of this project, our parallel oligonucleotide synthesis was limited to produce 80-mer oligonucleotides which can be translated to 15-20 amino acid long peptides. Therefore, the expression system was required to be able to produce peptides as short as 15 amino acids at sufficient levels for activity screening. For this reason, the feasibility of both yeast systems to produce such short peptides was investigated by cloning and expressing previously-identified AMPs, and is explained in detail in the following sections.

3.2 Expression of Antimicrobial Peptides in *Saccharomyces cerevisiae*

3.2.1 Introduction and Literature Review

The yeast *Saccharomyces cerevisiae* has become the most sophisticated eukaryotic model for recombinant DNA technology, and therefore, it has been used as a host for the expression of heterologous proteins and antimicrobial peptides in numerous studies.

Schoeman et al. (1999) developed a bactericidal strain of *S. cerevisiae* producing biologically active Class II-a Bacteriocin peptide, Pediocin PA-1. The *pedA* gene was successfully expressed and secreted under the control of *ADH1* promoter and *MFα1* secretion signal. Although they could not detect pediocin activity with single copy constructs, yeast clones transformed with multiple copies of *PED1* cassette formed clear growth inhibition zones when overlaid with *L. monocytogenes* and *L. fermentum* [56].

In another study by Reenen et al. (2003), 37 amino acid long peptide Plantaricin-423 was produced by a similar *S. cerevisiae* expression system [57]. Their findings showed that antimicrobial activity could be detected after 3 day-old yeast colonies when overlaid with *L. monocytogenes,* however, bacteriocin activity in the concentrated culture supernatant was very low, which is similar to the findings of Schoeman et al. (1999).

Destoumieux et al. (1999) demonstrated that yeast can be used as an expression host for the large-scale production of functional peptides. They successfully produced two antimicrobial peptides, Penaeidin-2a and 3a (50aa and 62aa long, respectively), in *S. cerevisiae* TGY strain by using yeast *MFα1* promoter in fusion to its secretion signal. Recombinant peptides were purified from culture supernatant by reversed-phase HPLC, and used for activity analysis and MIC

calculations against a number of gram negative and gram positive bacteria as well as yeast strains [58].

A 97amino acid-long sweet protein Monellin was also produced by Chen et al. (2010) under the control of strong-inducible galactose promoter *GAL1*, and the secretion signal *MFα1*. Positive clones were selected on uracil selective medium, and cultured for the secretion of the monellin to the supernatant. High expression and secretion levels, as well as very few host proteins being secreted to the supernatant allowed for successful purification of recombinant proteins by ion exchange chromatography [59].

In a study by Basanta et al (2009), authors tried to develop bacteriocinogenic *S. cerevisiae* strains producing Enterocin L50A and L50B antimicrobial peptides (44 and 43aa, respectively). Peptide coding genes were fused in frame to *MFα1* secretion signal containing a Kozak translation initiation sequence at the 5' termini, and galactose-induced *GAL1* promoter to control the protein expression. Positive yeast clones that were spotted on both rich and minimal media and incubated for 30 days did not show any antimicrobial activity, however, functional recombinant peptides were detected in the concentrated supernatants after 24h of incubation, and the maximum antimicrobial activity was detected after 144h of incubation [60].

Saccharomyces cerevisiae Expression System

A variety of host strains and vectors are available to direct production of recombinant proteins in *S. cerevisiae*. Most expression vectors are based on multi-copy 2μ plasmid, and contain specific regions for propagation both in bacteria and yeast as well as promoter and terminator sequences for efficient transcription of the foreign gene. Other components of an expression vector include a secretion signal sequence, and selection markers to maintain transformants. When constructing an

expression cassette, each component is selected based on the protein of interest to be produced [55, 61].

S. cerevisiae offers many different conditional promoters for tightly regulated and high-level gene expression. *GAL1, CUP1, ADH1,* and *MET25* are most commonly used for conditional expression of heterologous proteins, because these promoters show very low basal expression levels under repressed conditions. Gene expression can easily be regulated by addition of metabolites and ions to the growth medium; etc. galactose for *GAL1*, or copper for *CUP1* [62].

Galactose-inducible promoters (denoted as P_{GAL1}), such as *GAL1* (galactokinase), *GAL7* (galacto-transferase) and *GAL10* (galacto-epimerase), have been successfully used for overproduction of heterologous proteins. P_{GAL1} can rapidly induce the expression of downstream fused-genes about 1000-fold when cells are grown in a media containing galactose [63], and is turned-off when glucose is present in the growth media. Galactose-induced protein expression can be carried out in three ways: 1- by growing the culture on non-repressing and non-inducing sugar such as raffinose, then inducing the expression by galactose addition, 2-by growing on glucose to desired cell density, and replacing glucose media with galactose media, and 3- by growing in a media containing both glucose and galactose; induction starts after glucose is preferentially metabolized.

The peptide leader of the mating pheromone α-factor (α-MF) is the most commonly used signal peptide for the secretion of recombinant proteins in yeast. α-MF is a 13 amino acid peptide, secreted by haploid α mating type cells, and is synthesized as a part of 165aa precursor protein containing a leader and several spacer regions. There are two enzymes involved in proteolytic processing and maturation of α-MF, an endopeptidase encoded by *KEX2,* and an aminopeptidase encoded by *STE13*. Many studies have shown that α-MF leader sequence that is

fused to protein of interest is efficiently removed from the target protein by *KEX2* peptidase, and recombinant proteins are successfully transported out of the cells [61, 64]

Another important component of an expression cassette is the selection marker for the transformant cells. There are a few criteria that need to be considered when choosing the selection marker; adaptability of the transformed cells, ease of transformation, background growth levels, and the stability of the expression cassette. Yeast selection markers consist of two groups; auxotrophic markers, and positive selection markers (by antibiotic-resistance). Auxotrophic selection is carried out by complementation of an auxotrophic mutation in the yeast genome. The strain genome is mutated for an auxotrophic gene in a metabolic pathway, making this strain dependent on finding the metabolite in the medium to survive. A WT copy of the gene present on the plasmid can complement this genomic mutation. For example, *TRP1* and *URA3* genes are used in strains that are *trp1* and *ura3*, respectively. Positive selection markers, such as *KAN1* which confers resistance to G418 and Cm^r for chloramphenicol resistance, can be used when the host strain lacks auxotrophic mutations or when more rapid growth rates are desired, since the cells are plated on rich media rather than on minimal media as for auxotrophic selection [61].

3.2.2 Construction of the AMP coding expression plasmids for *S. cerevisiae*

For the expression and secretion of peptide libraries in *S. cerevisiae*, we have developed a vector called pYeF2Kα (or pYeF2Tα when containing *TRP1* instead of *KAN*), which was derived from a yeast/*E.coli* shuttle plasmid "pYeFc-2T" [65]. This multicopy parent plasmid contained 2μ yeast origin of replication, *GAL10-CYC1* hybrid promoter [66] and *PGK* transcription terminator as well as two selection

markers; *Amp* conferring resistance to ampicillin for the selection of *E. coli* cells, and *TRP1* for the selection of yeast cells. To construct pYeF2Kα, auxotrophic marker *TRP1* was replaced by *KAN* for G418 resistance, and α-MF secretion signal was cloned downstream of the *GAL10-CYC1* promoter (as a part of the expression cassette coding for AMPs). Plasmids, pYeF2Kα and pYeF2Tα, were both employed in this study, and will be discussed in following sections.

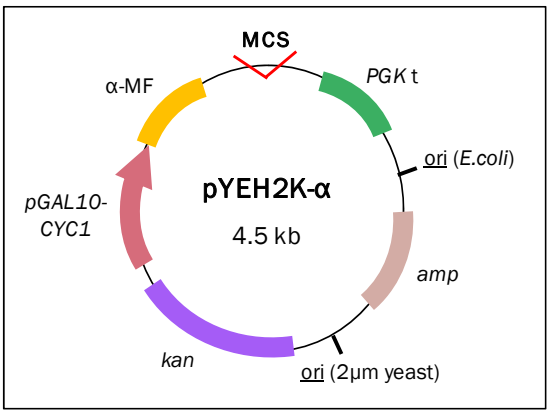

Figure 3.1- Map of the *S. cerevisiae* expression plasmid containing *KAN1* selection marker.

The cloning cassette for *S. cerevisiae* expression system consisted of AMP coding sequence, α-MF secretion signal, and small fragments of *GAL10-CYC1* promoter and *PGK* terminator for the integration into the expression plasmid by homologous recombination mechanism in yeast. Homologous recombination relies on the gap repair mechanisms of yeast, where flanking ends of a linear fragment are replaced with homologous regions of a gapped plasmid and, the "gap" is repaired by recombining the linear fragment with the plasmid DNA. This approach, proposed first by Ma et al. (1987), simplifies conventional cloning procedures by eliminating restriction digestion and ligation reactions, and allows us to clone PCR products at any desired region in the plasmid as long as the target sequence is known and a gap can be introduced [67]. This is particularly advantageous for expression studies

where the gene of interest needs to be in the correct translational phase with the promoter sequence.

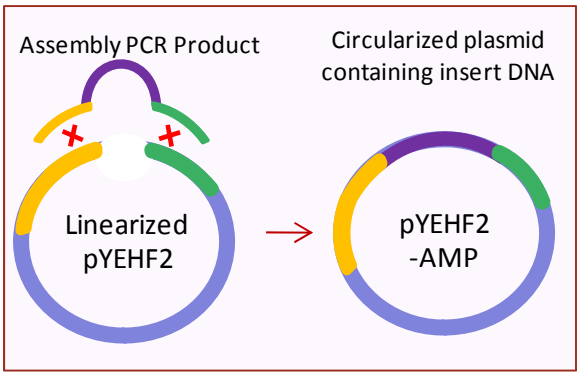

Figure 3.2- Cloning of PCR products into expression plasmid by homologous recombination.

The peptide-coding sequence was designed to have extra 15 nucleotides on both ends of the oligonucleotide corresponding to the last 5 codons of alpha-MF prosequence and the first 4 codon of the transcription terminator with a stop codon to terminate the peptide translation. This oligonucleotide was first amplified using two long primers to introduce 60bp sequence homologous to α-MF and the terminator sequences. The secretion signal α-MF was extracted from another plasmid in–frame to a 150bp-length fragment of the promoter. The PCR products of AMP coding oligonucleotide and the α-MF/P-Gal fragment were combined by assembly PCR to construct the cloning cassette. The final cloning cassette contained 150bp homology at the 5' end, and 60bp homology at the 3' end to the promoter and the terminator regions of the plasmid, respectively, to achieve a high number of recombination transformants. The steps of the construction are depicted in Figure 3.3.

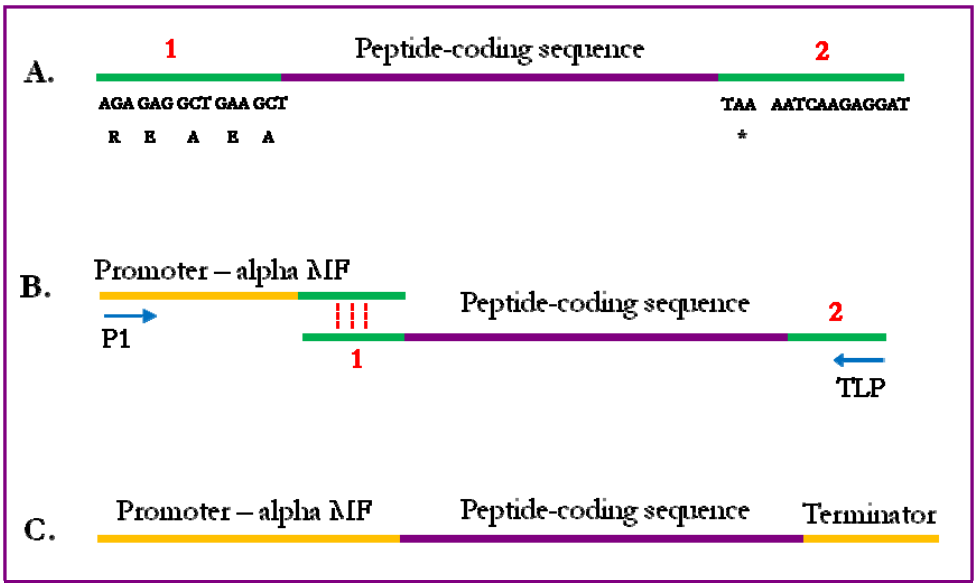

Figure 3.3- Assembly PCR Strategy. A. Peptide-coding oligonucleotide is synthesized with two flanking regions (marked as 1 and 2) homologous to α-MF and terminator. B. Homologous regions of the peptide-coding gene and α-MF fragment hybridize during PCR (region 1), and assembled fragments are amplified by primers P1 and TLP. C. Final expression cassette.

A standard lithium acetate-heat shock method was employed for all the transformations in *S. cerevisiae*-BMA64α strain (the protocol is provided in the appendix). For the cloning of (P-Gal)-αMF-AMP cassette by homologous recombination, plasmid DNA was digested with *BamH*I and *Bsu36*I to create a gap, and mixed with the insert DNA at 1:4 vector:insert molar ratio. The DNA mix was transformed into competent yeast cells by introducing a heat shock for 15 min at 42 ^{0}C, and the insertion occurred at the homologous regions of the cloning cassette and the promoter-terminator genes of the plasmid. Transformants were selected on either YPDG (1% yeast extract, 2% peptone, 2% dextrose and 50µg/ml G418) plates or MMD plates (Minimal Media Dextrose, 0.67% YNB, 0.192% Yeast Drop-out without *tryptophan*, and 2% Dextrose).

3.2.3 Peptide Expression and Activity Screening Assays

Peptide expression on solid media

For the screening of AMP production by transformants, selected colonies were re-suspended in ringer solution, diluted to 50-1000 cells/plate, and spread over the galactose/raffinose plates to induce peptide expression. After 4-6 days of incubation at 30^0C, plates containing patches of selected clones were overlaid with 10ml of soft-agar MRS media containing indicator strain cells (final concentration of OD_{600}=0.03-0.05). *Lactobacillus plantarum* ATCC 8401 and *Escherichia coli* K12 were used as indicator strains in the activity testing assays. After overlaying the plates with the indicator strain and incubating overnight at 30^0C, the plates were examined for the presence of bacterial growth inhibition zones around yeast colonies, which indicates active AMP expression and secretion by the host colony in the middle.

Alternatively, the colonies on transformation plates were directly transferred to galactose/raffinose plates as patches for peptide expression. In this case, the incubation on inductive plates was reduced to 2-3 days, and then the plates were sprayed with MRS media containing indicator bacteria cells.

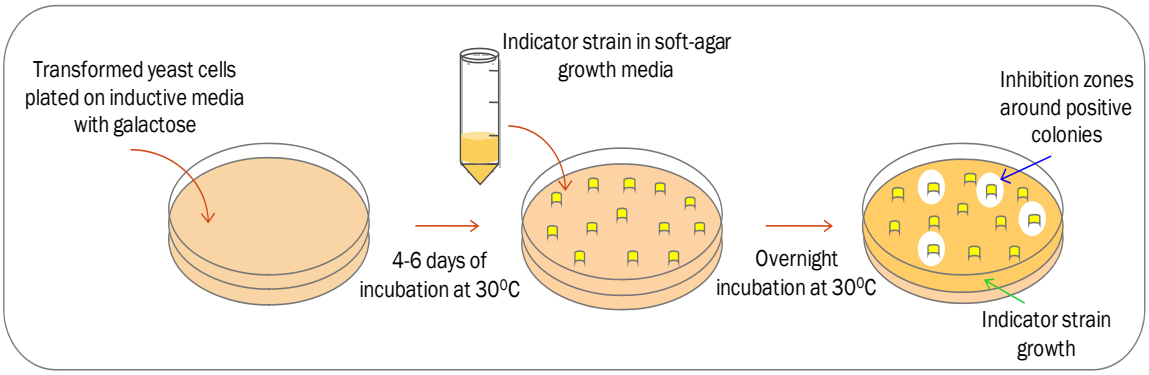

Figure 3.4 - Agar diffusion assay for the screening of AMP activity.

Peptide expression in liquid media

Selected recombinant clones were inoculated in 5ml yeast media (minimal media or YP) containing 2% glucose (and 50µg/ml G418 when using pYeF2Kα for expression), and grown for 24h at 30^0C and 200rpm. Cells were pelleted and washed with sterile water several times to remove any residual glucose and antibiotic. The final pellet was re-suspended in yeast media containing 2% Galactose and 1% Raffinose, and the cultures were incubated for 2-4 days for peptide production. In some experiments, instead of preparing initial cell mass in glucose media, recombinant clones were inoculated directly in a media containing the same amounts of galactose and raffinose. In this case, the adaptation of *S. cerevisiae* cells to galactose as a carbon source generally took about a day. After 24 hours of incubation the cells started growing more rapidly and the peptide expression increased accordingly.

To test the antimicrobial activity of the supernatant, yeast cultures were first centrifuged to pellet the cells and then applied to a 0.2µm-pore size membrane to eliminate any yeast cells from the supernatant. Aliquots of the clarified supernatant were loaded into a multi-well plate and a sample of indicator strain suspension was added to each well. After incubation overnight at 30^0C, absorbencies at 600nm were measured using a micro-plate reader and the bacterial growth inhibition was determined on the basis of optical densities of each sample tested.

Two negative controls were included in all experiments: untransformed *S. cerevisiae* BMA64α culture (i.e. does not contain any expression plasmid) and a culture of the same strain transformed with the expression plasmid coding for "polyglycine" peptide, which should not show any antimicrobial activity.

3.2.4 Expression and Secretion of IsCT and Pediocin in *S. cerevisiae*

To investigate the feasibility of using *S. cerevisiae* expression system for the production and screening of short antimicrobial libraries, two previously identified AMPs with different properties were chosen to be tested in the preliminary studies of this project. The first AMP that was tested in our system was IsCT-[L_6K_{11}], which is a derivative of a 13 amino–acid long, α-helical peptide IsCT with an activity against both gram positive and gram negative bacteria. In a study by Lee et al. (2004), the antimicrobial activity of IsCT and its five derivatives was tested against several bacteria, and the minimal inhibitory concentration of IsCT-[L_6K_{11}] against *E.coli* was found as 2μM being the most potent of all [68]. The second peptide was called Pediocin PA-1 and it is a 44 amino-acid long Class II-a peptide that is only active against gram positive bacteria (discussed in more detail in Chapter-4) [69]. Pediocin PA-1 was previously produced by a different *S. cerevisiae* vector at sufficient levels [56]; therefore it was selected to verify that the expression system developed in this study is suitable for producing small peptides. Both peptide and DNA sequences of Pediocin PA-1 and IsCT-[L_6K_{11}] are shown in Table 3.1.

Table 3.1- DNA and Amino Acid Sequences of IsCT and Pediocin PA-1

IsCT-[L_6K_{11}] Amino acid Sequence: ILGKILKGIKKLF DNA sequence optimized for yeast: ATCTTGGGTAAGATCTTGAAGGGTATCAAGAAGTTGTTC
Pediocin PA-1 Amino acid Sequence: KYYGNGVTCGKHSCSVDWGKATTCIINNGAMAWATGGHQGNHKC DNA sequence optimized for yeast: AAGTACTACGGTAACGGTGTTACCTGTGGTAAGCACTCTTGTTCTG TTGACTGGGGTAAGGCTACCACCTGTATCATCAACAACGGTGCTAT GGCTTGGGCTACCGGTGGTCACCAAGGTAACCACAAGTGT

The peptide-coding oligonucleotides were designed using the codon bias in yeast, chemically-synthesized, and amplified by primers ALP and TLP to introduce extra tails for the following assembly PCR. PCR products of peptide-coding genes and α-MF secretion signal sequence were combined by assembly PCR using primers P1 and TLP, and cloned into the expression vector by homologous recombination.

3.2.5 Results

Expression on Minimal Media

S. cerevisiae cells transformed with the plasmid pYeF2Tα, carrying either IsCT-[L_6K_{11}] or Pediocin-PA1 coding genes, were selected on glucose minimal media lacking tryptophan. After incubation on glucose media for 3 days, the cells were transferred on minimal media containing galactose and raffinose for the induction of peptide expression and secretion. Control plates were also prepared where an aliquot of transformed cells, grown on glucose, were re-plated on non-inductive media containing glucose instead of galactose. The plates were incubated for 4-6 days at 30^0C and then overlaid with MRS and LB media containing the indicator strain *L. plantarum* and *E. coli* for the activity screening of Pediocin and IsCT, respectively.

The first AMP that was cloned in *S. cerevisiae* expression vector was IsCT-[L_6K_{11}] to investigate whether peptides shorter than 20 amino acids can be efficiently produced in this system. When *S. cerevisiae* colonies, transformed with pYeF2Tα-IsCT, were grown on inductive media and overlaid with the indicator strain, no bacterial growth inhibition zones could be observed around yeast colonies indicating lack of active AMP production. To find out whether the absence of these inhibition zones were due to low levels of IsCT-[L_6K_{11}] expression as a result of slower growth on galactose, yeast colonies were first grown on glucose media for

about 5 days, then selected colonies were washed and re-suspended in liquid media to obtain high density cell suspensions. Droplets of these cell suspensions were placed on inductive media containing galactose and the plates were incubated for additional 3-6 days. Each day, one plate was overlaid with the indicator strain to evaluate peptide secretion. On day 6, very small faint zones formed around several droplet colonies (Figure 3.5-A).

To verify the toxicity of IsCT-[L_6K_{11}] on the indicator strain used in this study, IsCT-[L_6K_{11}] peptide was chemically synthesized and samples of the peptide solution was loaded into the wells on yeast minimal agar media that was overlaid with *E. coli* K12. As shown in Figure 3.5-B, all the wells, except the middle control well that was filled with water, had clear inhibition zones around them confirming the activity of IsCT-[L_6K_{11}] against this particular *E. coli* strain.

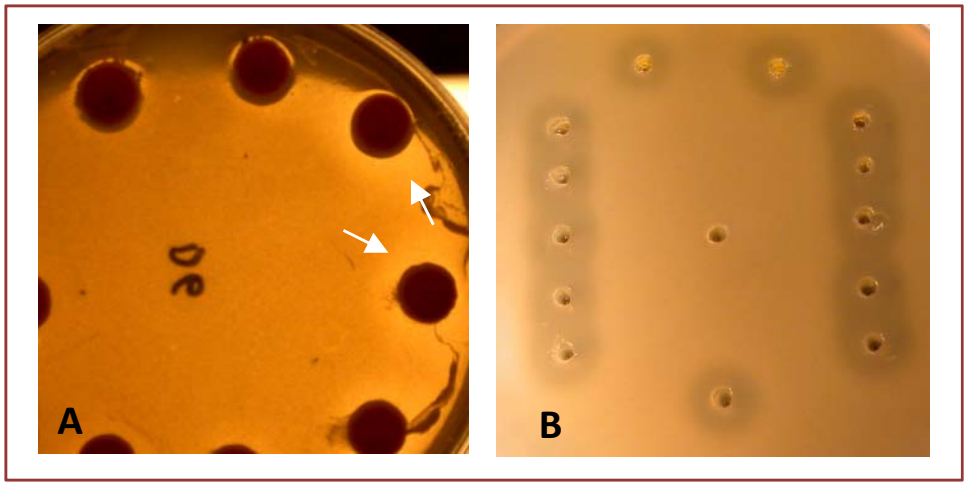

Figure 3.5- Screening of IsCT-[L_6K11] expression and activity in *S. cerevisiae*. A. IsCT expression in *S. cerevisiae* on inductive media Galactose, white arrows point to less bacterial-growth areas around yeast colonies. B. Clear bacterial inhibition zones around the holes filled with chemically synthesized IsCT

Due to problems with the production of IsCT-[L_6K_{11}], a longer AMP Pediocin PA-1 was cloned into the same expression vector to ensure that these problems did not originate from the expression system itself, and that it is able to express and secrete recombinant peptides efficiently. As seen in Figure 3.6, *S. cerevisiae* colonies transformed with pYeF2Tα-PED and then incubated on inductive media had clear growth inhibition zones where *L. plantarum* cells failed to grow due to secreted active Pediocin by these colonies. However, when yeast cells are grown on non-inductive media, *L. plantarum* cells were able to grow and such inhibition zones did not form around the colonies.

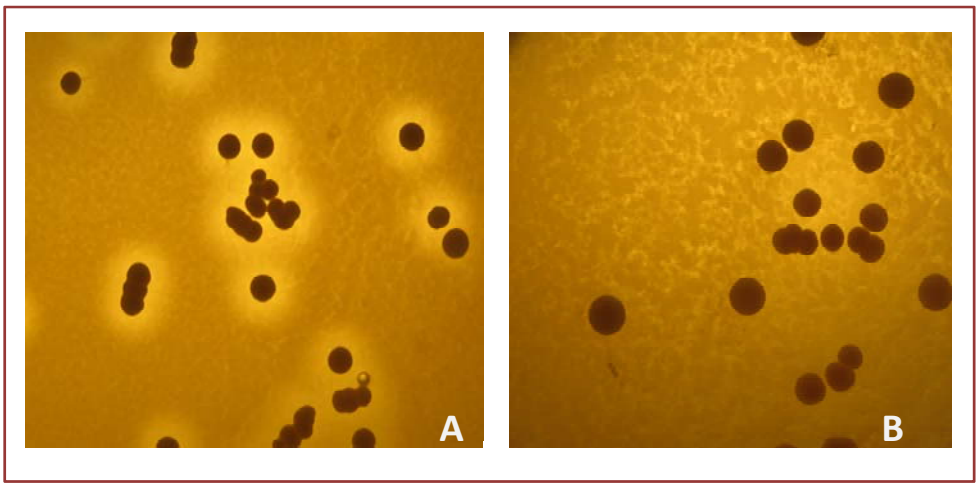

Figure 3.6- Pediocin PA-1 expression in *S. cerevisiae*. A. Pediocin expression on the inductive media Galactose. B. Negative control on non inductive media Glucose.

The absence of distinct growth inhibition zones around IsCT-producing yeast colonies could be due to number of reasons, such as inefficient expression and/or secretion of the peptide, degradation by the host proteases, and/or instability of the peptide in the yeast media. Since yeast growth, hence the peptide production, on minimal media is a slow process, accumulation of the peptide to reach the inhibitory

concentration takes longer, which may also result in increased peptide degradation. IsCT-[L_6K_{11}] being only 13aa-long, and having a linear structure, makes it highly prone to degradation by cellular proteases, and its survival in the expression media decreases with prolonged incubation times.

Expression on Rich Media

To achieve rapid growth rates in a shorter period of time, transformed yeast cells can be grown on rich media instead of minimal media. In this case, auxotrophic selection marker, i.e *TRP1*, in the expression plasmid needs to be replaced with a dominant selection marker, such as *KAN* for G418 resistance, to be able to select the transformants by simply adding the selective agent G418 (Geneticin) to the growth media. Using a dominant selection marker allows for a rapid growth, but also, cells start to lose their plasmid in the absence of the selective pressure. Since the expression, secretion, and activity of AMPs are investigated by overlaying the plates with an indicator strain, the screening plates should not contain any antibiotic as it would kill the non-resistant bacteria cells. For this reason, the yeast colonies grown on glucose media, containing geneticin for selection, needed to be transferred to galactose media lacking the antibiotic, rather than re-suspending the cells and re-growing on the inductive media, to minimize plasmid loss.

S. cerevisiae cells transformed with the plasmid pYeF2Kα carrying AMP coding genes and *KAN1* selection marker were selected on YP-Glucose media containing 50µg/ml Geneticin. To induce AMP expression, colonies were removed by sterile loops and placed on fresh YP plates containing galactose and raffinose and incubated for 2-3 days. After overlaying with the indicator strain, the plates were inspected for clear growth inhibition zones as an indicator of active AMP production. The plate photo (Figure 3.7) below shows colonies containing pYeF2Kα-

PED1 for Pediocin PA-1 expression along with negative control yeast colonies transformed with poly-glycine coding fragment instead of an AMP coding gene.

Figure 3.7- Pediocin PA-1 expression by *S. cerevisiae* on rich media.

Pediocin PA-1 was expressed and secreted at very high levels when yeast colonies are incubated on rich media. Out of six colonies that were randomly picked from the transformation plate, three of them (No: 1, 4, and 6 on the plate shown in Figure 3.7) had very clear and large growth inhibition zones indicating high levels of peptide production by these yeast clones. Especially No: 6 showed the highest expression yield with the least amount of host cell mass compared to other positive clones. On the other hand, colonies expressing poly-glycine (shown as "PG" on the image) did not inhibit *L. plantarum* growth as expected. The production of Pediocin PA-1 was also verified by liquid activity assay where the cell-free supernatants of several Pediocin-expressing yeast cultures were mixed with *L. plantarum* cell suspension and bacterial growth was measured by a spectrophotometer and compared to the growth in the samples with no AMP expression and fresh media. These results confirmed that AMPs can be expressed and secreted successfully on rich media using the plasmid pYeF2Kα with *KAN1* selection marker.

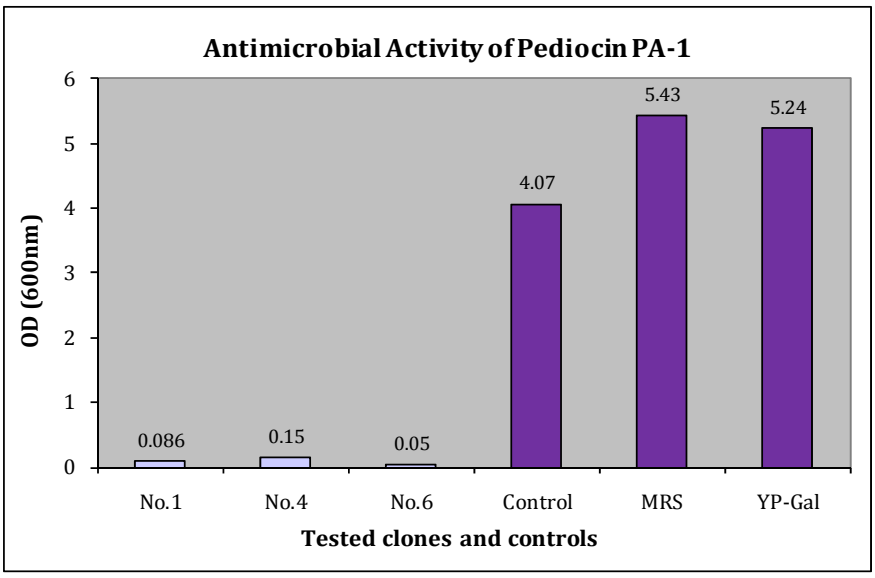

Figure 3.8- Antimicrobial activity of Pediocin-PA1 against *L. plantarum* ATCC 8401 in cell-free culture supernatants. Samples were taken at 48h after induction from *S. cerevisiae*-PED1 clones No.1, 4 and 6, as well as control clone expressing polyglycine. Data labels are the average of three repeats.

After the new expression vector containing *KAN1* was tested by expressing Pediocin PA-1, the next step was to try to produce IsCT-[L_6K_{11}] by *S. cerevisiae* on rich media at sufficient levels for activity screening. Even though bacterial growth inhibition was observed in clarified supernatants of *S. cerevisiae*-IsCT cultures (about 40% inhibition compared to controls), IsCT-[L_6K_{11}] secretion and its activity could not be detected on solid media by agar diffusion assay. Since Pediocin PA-1, which is 44aa-long peptide, was expressed with high yields both on minimal media and rich media, it can be argued that the small size of IsCT-[L_6K_{11}] is the limiting factor in succeeding in expression of this peptide by the eukaryotic host *S. cerevisiae*.

3.2.6 Conclusion

The feasibility of constructing antimicrobial peptide libraries in the eukaryotic host *S. cerevisiae* was investigated in this part of the project. *S. cerevisiae* offers a variety of expression tools that allow successful expression and secretion of heterologous proteins as shown in many studies. A preliminary study was

conducted by cloning and expressing two known antimicrobial peptides with different lengths to identify the size of the peptides that can be efficiently produced in our *S. cerevisiae* expression system. A 13 amino-acid long peptide IsCT-[L_6K_{11}], and 44 amino acid long peptide Pediocin PA-1 were both cloned in two different expression plasmids, one with an auxotrophic marker, the other with a dominant selection marker, and their expression and secretions levels on minimal and rich yeast media were studied. Antimicrobial activity of the recombinant peptides was analyzed by an agar diffusion assay on solid media and a liquid activity assay using cell-free culture supernatants.

Despite numerous attempts, IsCT-[L_6K_{11}] was not produced *by S. cerevisiae*, neither by expression on minimal media nor on rich media, at sufficient levels required for antimicrobial activity screening. Since a longer peptide Pediocin PA-1 was efficiently produced using both expression plasmids, the low yield of recombinant IsCT-[L_6K_{11}], if there was any, was likely due to its small size. Such small peptides often get degraded by cellular proteases either in the secretory pathway or in the extracellular environment by other secreted enzymes. Even though bacterial growth inhibitions in IsCT-producing yeast culture media were observed in several experiments, these results were usually not reproducible, which points to peptide instability in the expression media as well as inefficient production yields.

However, as stated above, Pediocin PA-1 was successfully produced by both *S. cerevisiae* expression constructions, efficiently secreted to the extracellular environment, and its activity against *L. plantarum* in the yeast culture media was verified. These findings suggest that a library of small peptides that are in the same size range with Pediocin PA-1 (i.e. 40-45aa) can potentially be constructed in *S. cerevisiae* and their activities can be analyzed by a simple screening assay.

3.3 Expression of Antimicrobial Peptides in *Pichia pastoris*

3.3.1 Introduction and Literature Review

The methylotrophic yeast *Pichia pastoris* has been successfully used as a cellular host for the production of many antimicrobial peptides. Cintas and co-workers used *P. pastoris* X33 strain as an expression host in three different studies for the production of bacteriocins Enterocin P (44aa) (2005), Enterocin L50 (47aa) (2010) and Hiracin (44aa) (2008) under the control of *AOX1* promoter and α-MF secretion signal [70-72]. In another study by Wang et al. (2009), human α-defensin 5 (33aa) was produced, with its intact N-terminal sequence, at high levels using the same *AOX1-αMF* expression system in *P. pastoris* GS115, and the antimicrobial activity of the purified peptide was tested against several gram positive and gram negative bacteria [73]. Numerous other antimicrobial peptides were successfully expressed and secreted by *P. pastoris*, such as LL37 (37aa) by Kim et al. (2009), Perinerin (51aa) by Zhou et al. (2009), Hepcidin derivatives (20-25aa) by Koliaraki et al. 2008, Penaeidin-5 (79aa) by Kang et al. (2007), CA-MA (22aa) by Jin et al. (2006), and Pleuricidin (27aa) by Burrowes et al. (2005) [74-79].

Several factors have contributed to its popularity; the availability of strong and tightly regulated promoters to drive the expression of foreign genes at high levels, the ability to produce proteins both intracellularly and extracellulary, the capability of performing post-translational modifications such as glycosylation, phosphorylation and disulfide bond formation required for correctly folded and fully-functional recombinant proteins, the strong preference for respiratory growth versus fermentative, technical ease for genetically manipulation, and finally the production of large amounts of desired proteins at a lower cost compared with most other eukaryotic systems [80-82].

The majority of heterologous protein production in *P. pastoris* is based on the methanol metabolism present in these cells. *P. pastoris* has the ability to grow on methanol, as the sole carbon and energy source, and the enzymes required to metabolize methanol are only present at high levels when the cells are grown on methanol. Alcohol oxidase 1 (*AOX1*) is one of the key enzymes involved in methanol utilization pathway and has been the most commonly utilized promoter for controlling the foreign gene expression in *P. pastoris*. The presence of methanol in the media is essential to induce high levels of *AOX1* transcription while it is strongly repressed by other alternate carbon sources such as glucose. When *P. pastoris* cells grow on methanol, the promoter regulates the *AOX1* gene which allows producing alcohol oxidase enzyme for the oxidation of methanol and the level of *AOX1* enzyme can reach up to 30% of total cell protein, therefore, this strong promoter can be used to drive high-level expressions of heterologous genes introduced downstream in the expression cassette [81, 82].

The second promoter that has been commonly used is the glyceraldehydes 3-phosphate dehydrogenase (*GAP*) which provides strong constitutive expression on glucose at a level comparable with *AOX1* on methanol [83]. The main advantage of *GAP* promoter over *AOX1* is that cells can be grown without need of methanol so it is not necessary to shift cultures from glucose to methanol, which makes this promoter particularly useful for large-scale production.

Protein expression in fusion to SUMO gene tag

Gene fusion tags have been commonly used in proteomics to enhance protein expression, increase protein solubility and folding, decrease proteolytic degradation, and simplify purification; however, removal of these tags with high efficiency has been a main drawback [84]. Recently, an advanced fusion system,

SUMO (small ubiquitin related modifier), has been developed that has all the advantages of traditional fusion tags, and in addition, is much more efficient at cleaving fusion proteins due to highly robust SUMO-specific proteases [85, 86]. The main advantage of this system is that SUMO protease recognizes the tertiary sequence of SUMO protein, never cleaving within the fusion protein of interest, which yields in native-N terminus of the target protein (except when proline is present at the amino terminus) [85].

In a study by Butt et al. (2005), the SUMO fusion system was compared to other traditional gene fusion tags and proteases including maltose-binding protein (MBP), glutathione S-transferase (GST), NUS A and AcTEV protease, and it was shown that SUMO and NUS A were the best tags for enhanced protein expression and solubility, but SUMO protease was approximately 25 times more active than other proteases, which suggests that the SUMO fusion system can be employed for the expression of difficult-to-express recombinant protein [87].

This part of the project investigated whether *P. pastoris* could be used as an expression host for the production of small antimicrobial peptides. Due to difficulties in producing IsCT-[L_6K_{11}] in *S. cerevisiae*, the *P. pastoris* expression system carried a SUMO gene fusion tag to improve the solubility and stability of the recombinant IsCT-[L_6K_{11}], which is discussed in detail in the following section.

3.3.2 Construction of the AMP coding expression plasmid for *P. pastoris*

For the expression and secretion of tagged IsCT-[L_6K_{11}] in *P. pastoris*, the vector pP-αSUMOstar (Life-sensors Inc.) that contains α-MF secretion signal and a gene coding for human SUMO3 protein was employed. The DNA fragment coding for IsCT-[L_6K_{11}] was amplified by PCR using primers IsCT-F (forward, 5'-GCGCGTCTCAAGGTTTCTTGAAGAAGATC-3') and IsCT-R (reverse, 5' GCGTCTAGATC

AGATCAAACCCTTGATCAA-3') to introduce *BsmB*I and *Xba*I restriction sites (underlined), respectively. The PCR products were digested with *BsmB*I, and ligated into pP-αSUMO3 plasmid (3.8 kb) that was linearized with the same enzymes. This allowed the insertion of the fragment in-frame with the SUMO3 tag, downstream of the α-MF secretion signal, resulting in SUMO3-IsCT fusion protein construct (Figure 3.9). The ligation products were transformed into *E. coli* XL_1Blue competent cells by electroporation, and recombinant clones were selected on LB-Ampicillin plates (following standard ligation and electroporation protocols: Sambrook and Russell, 2001) [88]. Plasmid DNA was extracted from a selected clone, and subjected to DNA sequencing to confirm the cloning of IsCT-[L_6K_{11}] fragment in the correct frame with the gene tag.

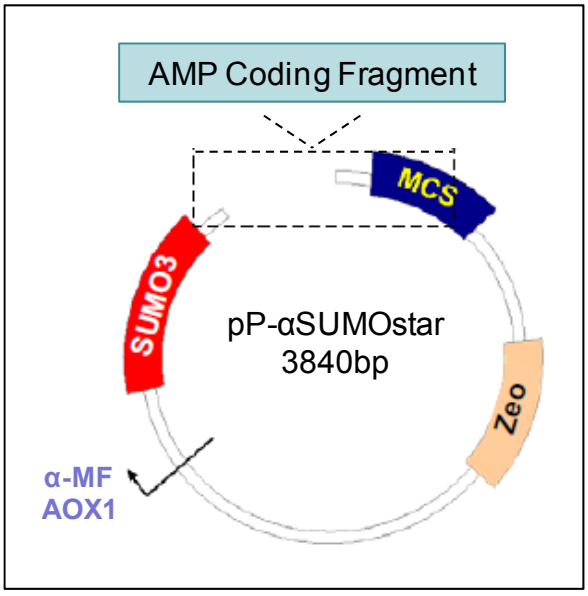

Figure 3.9- *P. pastoris* secretory expression plasmid (adapted from Life Sciences, Inc. manual for SUMO Expression System for Pichia pastoris).

To allow integration of the plasmid DNA at the AOX1 promoter locus of the *P. pastoris* genome by homologous recombination, pP-αSUMO3-IsCT was first linearized with SacI. *P. pastoris*-GS115 competent cells were transformed with 2-4μg of linearized vector DNA by electroporation (Electroporator 2510, Eppendorf) according to the manufacturer's instructions (SUMO3 Expression Systems for *Pichia pastoris*, Life Sensors Inc). Transformants were selected on YPDS plates (1% yeast extract, 2% peptone, 2% glucose, and 1M sorbitol) containing 100μg/ml Zeocin.

3.3.3 Peptide Expression and Purification

For the methanol-induced expression in *P. pastoris*, the selected clone was first grown in 100ml BMGY medium (1% yeast extract, 2% peptone, 1.34% YNB, $4x10^{-5}$% biotin, 100mM potassium phosphate, and 1% glycerol) in a 1-liter flask for 16-20 hours at 28^0C and 250rpm until the OD_{600}=2-6. The pre-culture was centrifuged at 5000rpm for 3 min, and the cell pellet was re-suspended in 20ml BMMY medium (1% methanol instead of 1% glycerol in BMGY). The culture was supplemented with 100% methanol to a final concentration of 1% (v/v) every 24 hour for 4-days at 28^0C. Two milliliter samples of expression medium were taken at 24h intervals, the supernatants were quick-frozen in liquid NH_2 and stored at -70^0C to be used for the analysis of secreted protein expression by Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and immobilized metal affinity chromatography (IMAC). SUMO3-IsCT-[L6K11] fusion proteins secreted by *P. pastoris* into expression medium were analyzed by 14% SDS-PAGE. Two-milliliter culture samples were either precipitated by cold acetone, or purified by IMAC before loading. The gels were stained with coomassie blue (Simply Blue Safestain, Invitrogen) for 1 hour and then washed in 4% NaCl solution for 3 hours for maximum sensitivity.

TALON Metal Affinity Resin (Clontech) was used for the purification of 6xhis-tagged fusions by IMAC. The recombinant proteins were recovered by imidazole elution (200mM imidazole, 50mM sodium phosphate, 300mM NaCl) at 500ul fractions.

3.3.4 Activity Screening Assay

The antimicrobial activity of IsCT-[L_6K_{11}] was tested against *Escherichia coli* strain K12 by the following procedure: The recombinant *P. pastoris* clone was grown in 2ml BMGY for 16 hours at 28 ^{0}C and the cell pellet was re-suspended in 5ml BMMY. After overnight of expression, 2ml cultures were removed for the AMP activity assay. The cells were re-suspended in 250µl fresh inductive medium resulting in 8-fold concentrated culture and loaded in 96-well microplates for 3 more days of incubation at 28^{0}C. The supernatants were recovered by centrifugation and treated with 0.5ul SUMO Protease II (Life Sensors) for 2 hours at 28^{0}C to remove the SUMO fusion tag from the recombinant peptide. After protease treatment, 50ul supernatants were mixed with 50µl LB media containing *E. coli* cells (10ul overnight culture, at OD_{600}=2-3, mixed with 5ml LB media) and loaded in 96-well microplates. The negative controls were fresh BMMY and the supernatants from untransformed *P. pastoris* culture mixed with the same amount of bacteria suspension. The microplates were incubated at 37 ^{0}C overnight and the samples were analyzed by measuring the absorbance at 600nm to evaluate the *E. coli* growth depending on the presence or absence of peptides and their antimicrobial activities on *E. coli*.

3.3.5 Results

Generation of the recombinant yeast clones

For the generation of recombinant *P. pastoris* clones, the oligonucleotide encoding for IsCT-[L_6K_{11}] was PCR-amplified and transformed into pP-αSUMO3 vector downstream of the human SUMO3 gene fusion tag. Restriction enzyme digestion and DNA sequencing confirmed the integration of IsCT-[L_6K_{11}] coding gene into the plasmid. Electrocompetent *P. pastoris* GS115 cells were transformed with the *Sac*I-linearized pP-αSUMO3-IsCT plasmids. Transformants were selected on YPGS-Zeocin plates, and analyzed by whole-cell PCR. Recombinant SUMO3-IsCT fusion proteins produced by the positive clones were detected by SDS-PAGE after 6xHis-tag purification. Only one *P. pastoris* clone was used for further studies.

Inducible expression in *P. pastoris*

The expression level of SUMO3-IsCT fusion protein in *P. pastoris* GS115 was investigated. The pre-cultures were prepared by cultivating the selected positive clone and a control 'no-insert' clone in 100ml BMGY medium for 16 hours at 28 ^{0}C. The cell pellets were then re-suspended in 20ml BMMY for methanol-induced protein expression. Aliquots of culture supernatants, collected at 24h intervals, were purified on an IMAC column and analyzed by SDS-PAGE. The following figure shows the SDS-PAGE image of the supernatant samples taken at the 96th hour of the induction, before and after treatment with commercial SUMO protease.

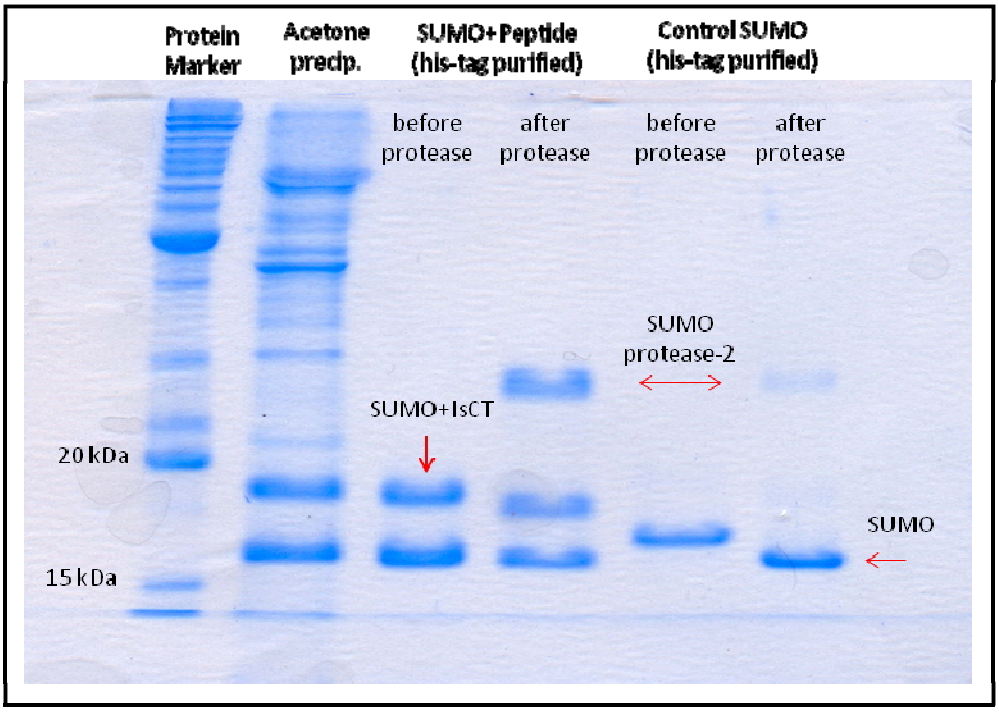

Figure 3.10- SDS-PAGE gel of purified proteins from *P. pastoris* cultures. First lane: Protein Marker. Second lane: Acetone precipitated culture supernatant at 96h of induction. Third and fourth lanes; Purified SUMO-IsCT products before and after protease treatment, upper band corresponds to fusion protein, lower band corresponds to cleaved SUMO3. Fifth and sixth lanes: Purified SUMO3 protein expressed by a control colony, before and after protease treatment.

The band for SUMO3-IsCT fusion products was observed around 19kDa on the gel along with a secondary band around 16kDa (Figure 3.10). The lower band corresponds to SUMO3 protein separated from IsCT-[L_6K_{11}] by the yeast endogenous SUMO protease during expression and secretion. Presence of the histidine-tag and the size of the control SUMO3 protein (lane 5) that was treated with commercial protease verified that the second band corresponds to cleaved SUMO3 protein. Further treatment of the purified fusion products with SUMO proterase-2 did not increase the amount of cleaved SUMO tag (lane 4), and the accuracy of the cleavage stayed around 50% of the total fusion products.

Antimicrobial activity

The antimicrobial activity of recombinant IsCT-[L_6K_{11}] was tested against *E. coli* K12 by a microplate assay. The supernatants collected from 8-fold concentrated yeast cultures after 3 days of induction at 28^0C were first treated with SUMO-Protease 2, then mixed with *E. coli* cell suspension and incubated at 37^0C to allow bacteria cells to grow. The microplates were analyzed by taking absorbance measurements at 600nm. The results showed that, when culture supernatants are concentrated by 8-fold (Figure 3.11), the bacterial growth was inhibited by mature IsCT-[L_6K_{11}], which was produced by some of the clones tested, compared to the controls of fresh media and culture of no-insert *P. pastoris* clone. It should be noted that the antimicrobial activity decreased in less concentrated supernatants.

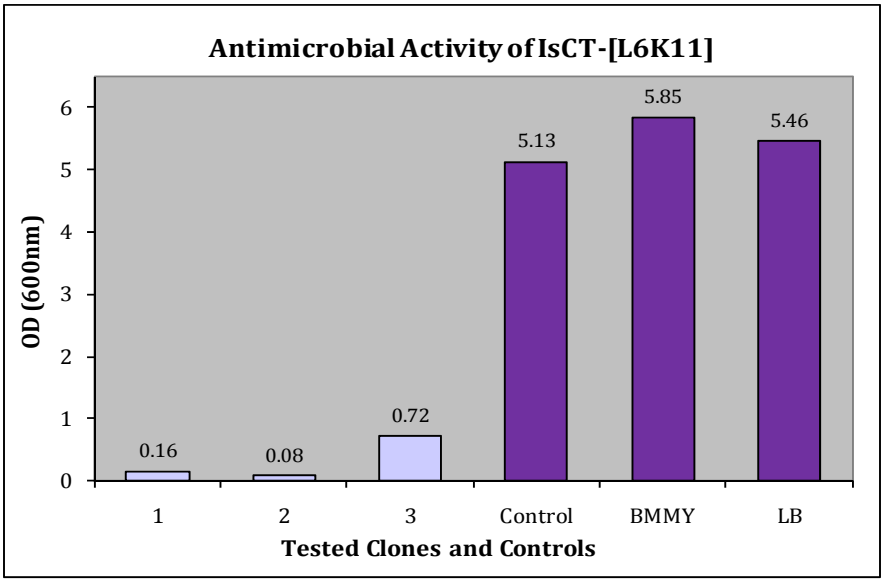

Figure 3.11- Antimicrobial activity of IsCT-[L_6K_{11}] in 8-fold concentrated *P. pastoris* culture supernatants. Positive clones expressing an active AMP are numbered as 1, 2 and 3, No.2 being the most active clone. 'Control' corresponds to a P.pastoris clone carrying the expression cassette

3.3.6 Conclusion

This part of the study investigated the efficiency of using gene fusion tags on the expression yield of small antimicrobial peptide IsCT-[L_6K_{11}] by *P. pastoris*. IsCT-encoding fragment was cloned in fusion to SUMO gene tag in pP-αSUMO-3 and the linearized vector DNA was transformed into *P. pastoris* genome by homologous recombination. Selected clones were grown in expression media and cell-free supernatants were analyzed for recombinant products of SUMO-IsCT fusions.

SDS-PAGE analysis of purified culture supernatants showed that IsCT was successfully expressed and secreted in fusion to SUMO protein tag. This is the first report, to our knowledge, of recombinant IsCT-[L_6K_{11}] production in a cellular host. Approximately 50% of the tags were cleaved by yeast's endogenous SUMO protease, and treating the un-cleaved fusion products with commercial protease did not result in further separation of the tag. This suggests that the activity of SUMO protease can be affected by the target protein's sequence, and thus the cleavage of the tag can be limited. The complete separation could not be reached in this experiment; however, the preserved antimicrobial activity of IsCT-[L_6K_{11}] suggests that N-terminus of the recombinant peptides was not altered due to fusion tag removal.

These results demonstrate the potential of using SUMO fusion tags to enhance expression and secretion of small peptides with less protease degradation, and increased stability and solubility. However, the antimicrobial activity of recombinant IsCT-[L_6K_{11}] could only be observed in concentrated culture supernatants. This is not an issue for individual protein expression studies, but in the case of a library screening, it requires each colony be grown and analyzed separately which is very difficult to do in a high-throughput fashion. Therefore, it can be concluded that this particular *P. pastoris* expression system is not suitable to construct and screen large libraries that contain thousands of candidate peptides.

3.4 Chapter Conclusion and Discussion

Yeast expression systems offer versatile and accessible tools for recombinant protein production. High expression and secretion yields, ease of recombinant techniques, and availability of numerous post-translational modifications make these organisms favorable hosts for the economical production of heterologous peptides and proteins. Two yeast strains have been investigated in this study, *S. cerevisiae* and *P. pastoris,* for their suitability to construct peptide libraries. Due to limitations with oligonucleotide synthesis at the time of this project, the expression systems were expected to be able to produce peptides as short as 15 amino acids. Production of such short peptides efficiently in a cellular host is often challenging due to proteolytic degradation and instability of peptides. Therefore, before constructing libraries, both yeast strains were tested for their ability to produce small peptides by expressing two known antimicrobial peptides, IsCT-[L_6K_{11}] and Pediocin PA-1.

Expression in *Saccharomyces cerevisiae*

Protein expression in yeast can be done either on minimal media using auxotrophic selection markers, or on rich media using either dominant selection markers or counter-selection markers. Both expression methods were tested in *S. cerevisiae* using *TRP1* and *KAN1* selection markers, but despite numerous attempts, IsCT-[L_6K_{11}] could not be expressed efficiently enough for activity screening, most probably due to degradation by cellular proteases at the yeast secretory pathway. Even though its antimicrobial activity was detected at several experiments, it was very difficult to reproduce the same results. Therefore, it was concluded that *S. cerevisiae* expression systems used in this study were not feasible to produce peptides that are similar to IsCT-[L_6K_{11}] in length.

However, Pediocin PA-1 was successfully produced in both *S. cerevisiae* constructions. Antimicrobial activity of the recombinant peptide was verified against *L. plantarum* by agar diffusion and liquid activity assay as explained above. The expression levels on rich media were much higher compared to minimal media due to rapid growth and shorter duration of the experiment which was reduced from 9-10 days to 5 days. The main drawback of doing protein expression on rich media, where transformants are selected based on their resistance to a certain antibiotic, is the necessity of removal of any residual antibiotic that can be carried over to the screening experiment. Bacteria cells are sensitive to even trace amounts of G418 (Geneticin), and thus, when antimicrobial peptides are tested for their activity in yeast media, the bacterial growth inhibition may not only be due to toxicity of these peptides but also carry-over antibiotic present in growth media. To prevent any G418 contamination in the inductive media, excessive washings with large volumes of fresh media should be performed between culture growth and peptide expression steps of the experiment. Proper controls must be included to verify that the entire antibiotic is removed and that the bacterial growth is inhibited only by the antimicrobial peptides secreted by yeast cells.

Expression on rich media can also be done using co-lethal markers such as *URA3* and *FUR4* genes rendering the yeast cell unable to either produce uracil or use the uracil from the growth medium [89]. The only way for the cell to survive is to bear a plasmid carrying either one of the two genes. This approach has the advantage of high expression yields on rich media, and also eliminates the use of toxic selection agents in inductive media simplifying the screening experiments. Therefore, by modifying the selection or screening methods based on specific applications, *S. cerevisiae* expression plasmids that were developed in this study can be used for the construction of libraries with peptides longer than 40 amino acids.

Expression in *Pichia pastoris*

Since the expression of IsCT in *S. cerevisiae* was problematic due its small size and linear structure, *P. pastoris* expression plasmid carried the SUMO gene tag to express IsCT-[L_6K_{11}] in fusion to SUMO3 protein, hence, enhance secretory expression and stability of target peptide while protecting it from proteolytic degradation.

When his-tag purified supernatants of concentrated *P. pastoris* cultures were loaded on SDS-PAGE, fusion products were clearly present along with an equal amount of SUMO3 protein cleaved by yeast endogenous protease in the same sample. Fusion proteins were treated with commercial protease, but SUMO3 protein could not be cleaved from IsCT-[L_6K_{11}] any further, indicating that the protease recognition site could be altered post-translation, or the protease activity can be limited depending on the target sequence. The activity of cleaved IsCT-[L_6K_{11}] in culture supernatants was verified against *E. coli*. Cleavage of the fusion tags by endogenous proteases presents a great advantage to screening experiments as it eliminates extra treatment of purified samples with the commercial enzyme. It should be noted, this argument is merely based on the results with IsCT-[L_6K_{11}], and the behavior of this system may vary with a different peptide.

Production of functional IsCT-[L_6K_{11}] was accomplished using *P. pastoris*, demonstrating that this system is efficient for secretory expression of such short peptides; hence, it can be used to construct small libraries with an automated screening system where each colony is picked, cultured, and transferred from one multiwell plate to another by robotic tools. However, it is not suitable for screening large libraries where thousands of peptides need to be analyzed, since the concentration of recombinant peptides are too low on solid media for activity assays.

CHAPTER 4

CONSTRUCTION AND SCREENING OF CUSTOM PEPTIDE LIBRARIES IN BACTERIA

4.1 Introduction to the *Escherichia coli* Expression System

Escherichia coli has been the choice of host organism for recombinant protein expression primarily due to ease of genetic manipulations, and its rapid growth in simple and inexpensive media, hence, highly efficient protein production. The knowledge gained in genetics and biochemistry of *E. coli* resulted in development of many translational and transcriptional control elements that can be employed in expression of foreign genes [90]. The most common approach begins with growing transformed cells under repressive conditions until exponential growth phase is reached, then the expression of recombinant gene is induced using chemical agents that activate the promoter region of the plasmid. The expression strategy involves selection of host strain, promoter, selectable marker, fusion proteins/tags, and secretion signal sequence which can be determined based on the physical and biochemical characteristics of the protein of interest, desired production yields and downstream applications. Recombinant proteins can either be expressed intracellularly or secreted to the periplasmic space between the cytoplasmic and outer membranes. Although the production yield of cytoplasmic expression is generally higher, purification of recombinant proteins from the periplasmic space is a simpler process [91].

Periplasmic expression is usually preferred for the production of properly folded and functional proteins that are naturally-secreted. The secretion is mediated by a secretion signal sequence that is removed from the recombinant protein after translocation in the periplasmic space. To improve translational efficiencies and protein solubility, fusion proteins (such as maltose-binding protein and thioredoxin) are often incorporated downstream of the secretion signal sequence and in-frame to the N-terminal of the gene of interest. Smaller gene fragments, called fusion tags, can also be added either to the N-terminal or C-terminal end of a protein to simplify purification and allow for detection [91] . Although use of fusion proteins and tags enhances expression and purification, the protein of interest carries additional sequences that may alter its activity, hence, should be selected carefully.

For the construction of antimicrobial peptide libraries in *E. coli,* the expression system should be suitable for direct screening of peptide activities. This means that recombinant peptides should be secreted /released to the media where they can freely diffuse around the host colony in active form and screened for their antimicrobial activities.

The bacterial expression system employed in this study consists of a periplasmic-leaky *E. coli* strain and a plasmid suitable for controlled periplasmic expression of recombinant proteins. This special *E. coli* strain is an *lpo* mutant where both bound and unbound forms of murein lipoprotein are absent in its membrane, thus, it leaks periplasmic enzymes and proteins into the culture medium [92, 93]. The plasmid used in this study carries *Ptac* promoter, *lacI* repressor binding site, and OmpA secretion signal for the periplasmic expression of proteins. *Ptac* is a strong hybrid promoter composed of -35 region of the *trp* promoter and -10 region derived from the *lac* UV5 promoter. It carries two consensus sequences that are in close contact with RNA polymerase during RNA synthesis, increasing

promoter efficiency greatly. *Ptac* is repressed by LacI protein during the growth phase, however, addition of de-repressor IPTG inactivates the *lacI* gene, hence, inducing protein expression from *Ptac* promoter [94]. Thus, when AMP encoding sequences are cloned downstream of OmpA under the control of *Ptac* promoter, expression and secretion to the periplasmic space can be induced by addition of IPTG, and recombinant peptides are released into the culture medium by periplasmic-leaky host cells where they can be screened for their activities.

Use of a periplasmic-leaky *E. coli* strain for the expression and screening of antimicrobial peptides was first introduced by Miller et al. (1998). In this study, Pediocin AcH was expressed in-fusion to C-terminus of maltose-binding protein (a 43-kDa secretory protein), and released into the culture medium in an active form displaying bactericidal activity against *Listeria innocua* Lin1 strain [95].

Tominaga and Hatakeyama (2006) applied the same approach where a library of mutated Pediocin-PA1 encoding fragments, generated by site-saturation PCR mutagenesis (using NNK scanning method), were cloned in pFLAG-ATS plasmid that carries OmpA secretion signal and an N-terminal FLAG gene tag, and expressed by a periplasmic-leaky *E. coli* strain. Mutant peptides that were released into the medium by the host cells were screened for their altered antimicrobial activities against gram positive *P. pentosaceus* JCM2026. By mutating each of 44 residues of the wild-type peptide, they were able to identify both essential and nonessential residues of Pediocin PA-1 [11].

Since the host organism used for the peptide expression in this system is a gram negative strain *E. coli*, the peptides that are being produced should not possess activities against *E. coli* to ensure that the growth of the host organism is not affected by the toxicity of the recombinant AMPs. Bacteriocins are a class AMPs that display activities only against gram positive bacteria (with few exceptions),

specifically against food-borne pathogens and food-spoilage bacteria, making this group ideal to be studied using *E. coli* expression/screening system.

4.1.1 Class-IIa Bacteriocins

Bacteriocins are a group of ribosomally synthesized antimicrobial peptides mainly produced by lactic acid bacteria (LAB) [96]. LAB-bacteriocins have been a great interest for researchers due to their ability to kill food-borne bacterial pathogens, hence, their great potential as biopreservatives. Two bacteriocins, Nisin and Pediocin PA-1, are presently used as food preservatives [97, 98], and more research is focused on exploring other bacteriocins for their potential for medical applications.

LAB-bacteriocins are divided into 3 main classes: Class I contains post-translationally modified small peptides containing unusual amino acid lanthionine, commonly referred to as lantibiotics; Class II includes unmodified small heat-stable non-lantibiotics and Class III contains large heat-labile peptides. Class II is further divided into four sub-classes, class-IIa consists of anti-listerial peptides, also referred as pediocin-like bacteriocins, class-IIb contains two-peptide bacteriocins, class-IIc includes sec-dependent peptides, and class-IId are the cyclic bacteriocins with no sequence similarity to the members of previous groups [99, 100].

Class-IIa bacteriocins are by far the most thoroughly studied bacteriocins, due to not only their antimicrobial activities against gram positive food spoilage bacteria, particularly *Listeria monocytogenes*, but also the fact that they are produced by other food-grade lactic acid bacteria which makes them relatively safe to be used in a number of food products as preservatives. The first peptide identified in this group was Pediocin PA-1 (from which the term 'Pediocin-like

bacteriocins' has been derived for Class IIa members) produced by *Pediococcus acidilactici* strains, and it is the most studied member of this group [69, 101].

Pediocin-like bacteriocins vary from 37 to 48 amino acids in size with up to 80% sequence similarity at the amino acid level. There are more than 20 members in this group, and they all contain the "Y-G-N-G-V/L" consensus sequence (pediocin box) and two cysteine residues in their highly conserved N-terminal region [100]. Sequence alignment of these peptides shows that they consist of a highly conserved, cationic and hydrophilic N-terminal region (up to residue 17), and a less conserved but more hydrophobic/amphiphilic C-terminal region with a flexible hinge at the conserved aspartic acid (or asparagine in some peptides) residue 17, which allows these two domains to move in accordance with each other (Figure 4.1) [99].

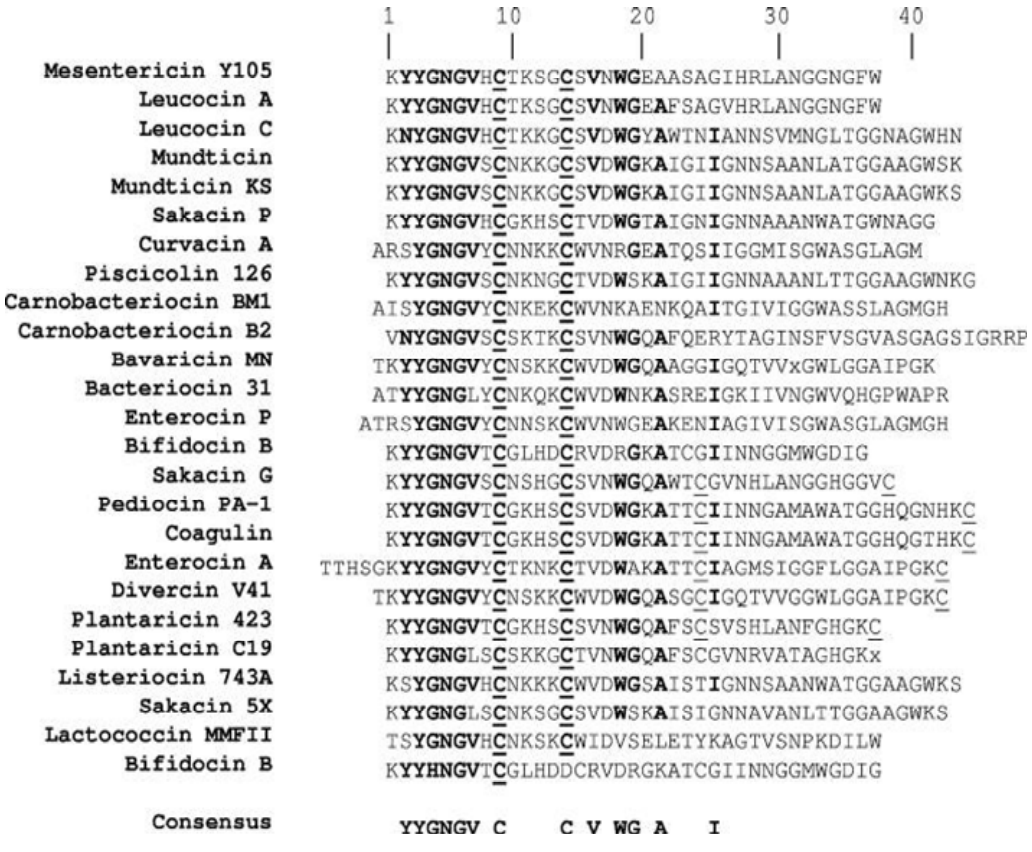

Figure 4.1- Sequence alignment of Class IIa bacteriocins based on their similarities at N-terminal [99].

The N-terminal region of these peptides forms a three-stranded β-sheet-like domain, which is stabilized by a conserved disulfide bridge, while the C-terminal domain forms into a structure that contains one or two amphiphilic α-helices (from residue 17 to residue 33). The C-terminal domain is often followed by an extended tail that folds back onto the central helix, where in some peptides such as Pediocin PA-1/AcH, Plantaricin-423 and Enterocin A, an additional disulfide bridge forms between the cysteine residue in the helix and the cysteine at the end of the C-terminus. This additional disulfide bridge plays an important role in peptide stability and often in antimicrobial activity at higher temperatures. However, most Class-IIa peptides carry only two cysteine residues and an additional two tryptophan residues which help stabilize the hairpin structure of the peptides in the membrane-water interface [102, 103].

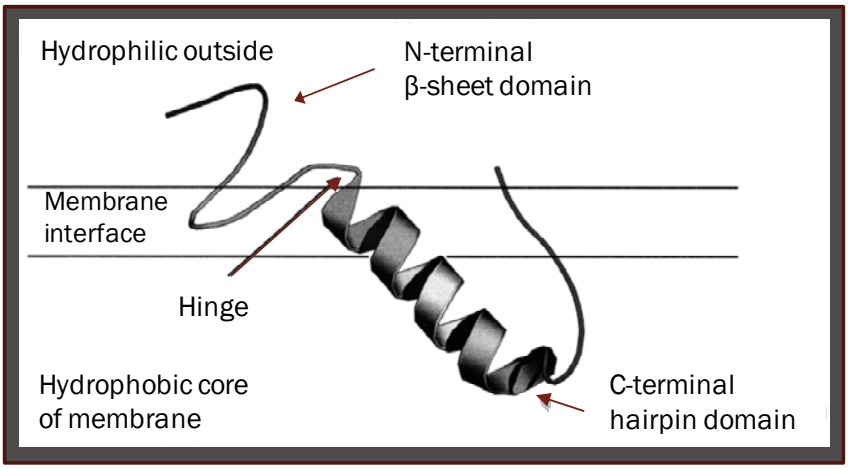

Figure 4.2- Proposed domain structure of pediocin-like bacteriocins [102].

Despite the extensive sequence similarity between pediocin-like bacteriocins, they notably differ in their antimicrobial activities [104]. Site directed mutagenesis studies indicate that the cationic N-terminal domain is responsible for

binding to the target cell surface by positioning itself in the membrane interface region, whereas the amphiphilic C-terminal domain penetrates into the hydrophobic core of the cell membrane causing disruption and leakage, leading to eventual cell death. The N-terminal region being highly conserved across the group, numerous studies investigated the role of the C-terminal domain on target cell specificity by creating hybrid peptides (combining N-terminal region of one peptide with the C-terminal region of another), and introducing mutations at the C-terminal region. The results of these studies showed that hybrid peptides have similar specificities to that of the parent bacteriocin from which the C-terminal region was derived, while mutated peptides often differed from the wild type in their target cell specificity [103]. All of these characteristics of Pediocin-like bacteriocins make this group of great interest for structure-function relationship analysis and especially for mutagenesis studies to investigate the role of each residue on antimicrobial activity.

4.2 Construction of the AMP Mutant Libraries in *E. coli*

In this study, custom designed libraries of two bacteriocins, Pediocin PA-1 and Plantaricin-423, were constructed in *E. coli* and screened for the altered antimicrobial activity of mutant peptides. Selected clones from the library were sequenced to identify the mutations and relate them to the overall effect on peptide structure and activity.

4.2.1 Library design and synthesis

Sequence modification of AMPs that are found in nature is generally the primary strategy in the design of novel antimicrobial peptides. Based on the fact that C-terminal region of Class-IIa bacteriocins is much more diverse compared to their N-terminal region, and believed to be responsible for antimicrobial activity,

only C-terminal region of wild-type peptides was mutated in this study. A single mutation was introduced at each position, starting at the 18th amino acid, by replacing the wild-type residue with a random amino acid selected from each of six groups. A second mutation was introduced at each remaining position again with one amino acid selected from the same amino acid groups shown in Table 4.1. These groups were formed based on the biophysical properties of amino acids to ensure that each residue of the wild-type peptide is mutated with at least one amino acid representing each group. Thus, one set of single mutations and one set of dual mutations resulted in a total of approximately 23K peptides in the pediocin PA-1 library and 13K peptides in the plantaricin-423 library.

Table 4.1-Amino acid groups used in the library design.

Positive and Hydrophilic	Lysine, Arginine, Histidine
Negative and Hydrophilic	Aspartic acid, Glutamic acid
Polar hydrophilic	Serine, Threonine, Tyrosine, Asparagine, Glutamine
Hydrophobic	Valine, Leucine, Isoleucine, Methionine, Phenylalanine
Aliphatic/Small	Glycine, Alanine
Structural (hydrophobic)	Cysteine, Proline, Tryptophan

Peptide sequences were generated following these design guidelines and converted to oligonucleotide sequences. Each oligonucleotide sequence contained 20bp primer-binding sites at both ends carrying two restriction enzyme sites, HindIII and EcoRI, and a stop codon. Bases “ct” was also inserted right after the HindIII site to maintain the reading phase. The sequence of the library oligonucleotide is shown below.

5’- GGCATAGAGGATCC**AAGCTT**CT–peptide coding sequence-TAA**GAATTC**TGAGCTACGGCCA -3’

The oligonucleotides of each library were synthesized, cleaved off the chips, and then purified by using CentriSpin-20 DNA purification columns (Princeton Sep. Inc., NJ) for downstream applications.

4.2.2 Construction of the AMP-encoding expression plasmids

For the expression and secretion of antimicrobial peptides, a periplasmic-leaky *E. coli* strain JE5505 was obtained from the Yale University *E. coli* Genetic Stock Center and the expression plasmid pFLAG-CTS was purchased from Sigma Aldrich (Figure 4.3). This plasmid carries a C-terminal FLAG gene tag for the detection of recombinant proteins by anti-FLAG antibody. In this study, peptide expression was carried out without a FLAG tag to maintain the native form of peptides and ensure that the peptide activity is not altered by the tag, thus, a stop codon was introduced between the peptide-coding region and the FLAG tag during oligonucleotide synthesis.

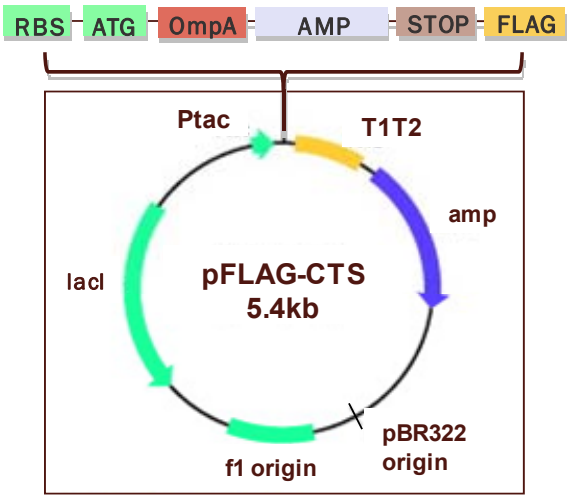

Figure 4.3- Map of the expression plasmid pFLAG-CTS.RBS, ribosome binding site; ATG, start codon; OmpA, secretion signal; AMP, peptide coding fragment; amp, Ampicillin resistance; Ptac, promoter; and TIT2, terminator.

Cloning of peptide coding fragments into the expression plasmid and transformation of plasmid DNA in *E. coli* JE5505 cells was performed by following standard molecular biology techniques [88]. The oligonucleotide library was amplified by emulsion PCR using the same method explained in Chapter 2. Primer sequences were as follows: FWD, 5'-GCATCATAGACATAGTCAGGCATAGAGGATCCAA GCTTCT-3' and RVS, 5'-GCAATACGT TAGCGTTCATGGCCGTAGCTCAGAATTCTTA-3'. After amplification, DNA extraction, and gel purification, PCR products were digested with HindIII and EcoRI in 100µl of 1X reaction buffer using 20units of each enzyme for 2h at 37 ^{0}C, and then ligated into pFLAG-CTS expression vector linearized with the same enzymes. Ligation reaction mix was incubated at room temperature overnight and ligated DNA was precipitated by ethanol. After transformation into electrocompetent *E. coli JE5505* cells, transformants were selected on LB-Ampicillin plates. Plasmid DNA was extracted from selected clones and the cloning of the library oligonucleotides was confirmed by DNA sequencing.

4.2.3 Peptide Expression and Activity Screening

Primary screening for peptide activity is often based on *in vitro* growth inhibition analysis of target cells and calculation of the inhibitory concentrations of the peptide that is being studied. The screening method used in this study was a modified version of the standard colony-overlay method as previously described (Figure 4.4) [105]. This method is commonly used for the screening of antimicrobial compounds because of its simplicity and ease of result analysis. Although this technique is very efficient in rapidly detecting positive hits from a library, it is not suitable for the determination of the minimum inhibitory concentration (MIC) of an active peptide. Hence, the results obtained using this method are only close-approximations of the actual activity of the peptides tested in this study, which is

enough for the purpose of screening libraries and gathering sequence information for further studies.

Basically, transformed *E. coli JE5505* cells were washed with 1ml LB broth, and diluted to 2000-5000cfu in 25ml melted LB (0.8% agar). Cell suspensions were poured into large petri dishes (150mm in diameter), overlaid with 10ml melted LB soft agar, and incubated at 37^{0}C for 24h for the development of colonies. The next day, these plates were overlaid with 10ml 0.8% TSB soft agar (or MRS soft agar for *L.plantarum*) containing 1mM IPTG, and the indicator strain cells (diluted to OD_{600}=0.03 from 4h-culture). After overnight incubation at 30^{0}C, the plates were inspected for clear growth inhibition zones around colonies. Two gram-positive indicator strains were used in this study: *Listeria innocua* ATCC 33090 in all library screening experiments and *Lactobacillus plantarum* ATCC 8104 for testing the activity range of positive mutant peptides.

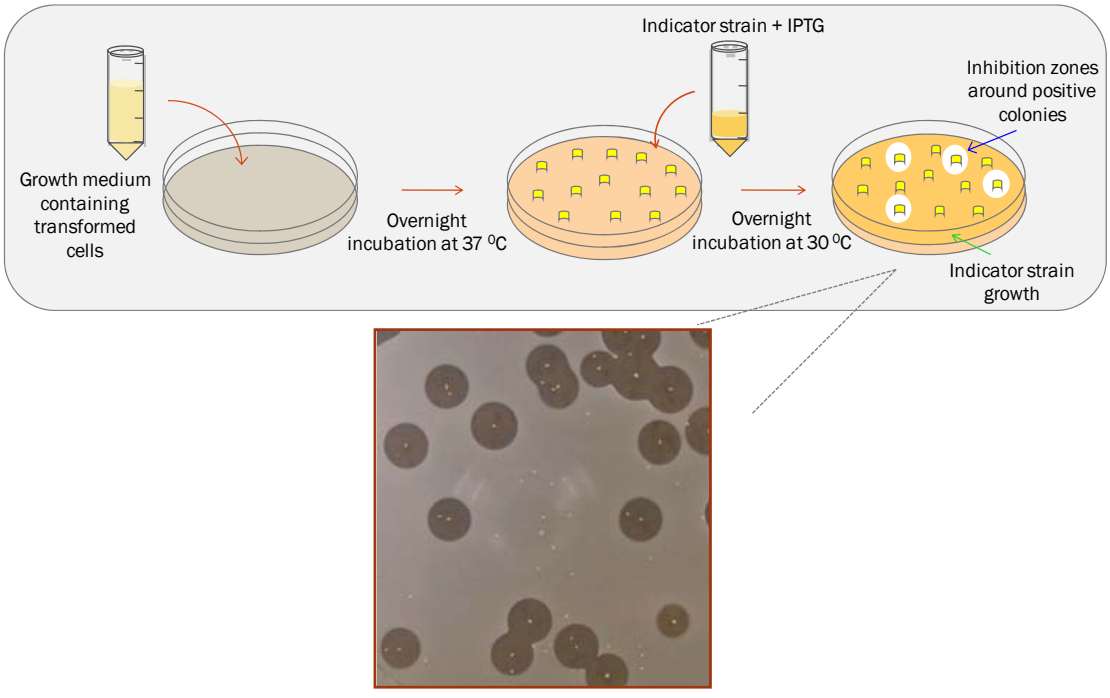

Figure 4.4- A. Agar-diffusion assay used for library screening. **B.** *E. coli* JE5505 colonies expressing wild-type pediocin PA-1.

For the verification of active peptide production, positive colonies with growth inhibition zones (selected at the library screening) were picked with sterile toothpicks, placed on LB plates, and incubated for 5h at 37^0C. These plates were then overlaid with LB-soft agar containing 1mM IPTG and the indicator strain (OD_{600}=0.03), and incubated at 30^0C for 18h for the development of inhibition zones [11]. Once the peptide activity is verified, plasmid DNA is extracted from the positive colonies, and sent for DNA sequencing for the identification of the peptide sequence.

Salt concentration in screening plates

Numerous studies have investigated the effect of salt ions on the activity of antimicrobial peptides. Short peptides, especially those with high positive charge and a linear structure, are often inhibited by NaCl at physiological concentrations [106, 107]. However, the activity of longer peptides with more amphiphilic structures varies in the presence of NaCl based on the pH and temperature of the system [108, 109]. Of course, when determining the effect of salt by the viability of the target cells, one should consider not only the conformational changes in peptides induced by salt ions, but also the sensitivity of target cells to high salt concentrations.

For instance, sensitivity of *L. monocytogenes* to Nisin (Class-I bacteriocin) was considerably increased by the addition of 2% or more NaCl to the growth media [110, 111]. Similarly, Enterocin As-48 (Class-IIc bacteriocin) was most effective against *S. aureus* in combination with 6-7% of NaCl when incubated at +4°C[112]. In another study, the thermal stability of a bacteriocin-like peptide P34 was improved at high concentrations of NaCl [108]. However, in another study, the adsorption of plantaricin-423 to *Enterococcus sp.* was reduced when bacteria cells were treated

with 1% NaCl prior to the incubation with peptide solution [109]. Thus, the specific interaction of salt ions with peptides and the overall effect on the antimicrobial activity vary significantly for each peptide-target cell pair.

To determine the minimum NaCl concentration to use in all screening experiments, LB agar plates were prepared with increasing concentrations of NaCl (0.1% to 1.0% w/v). These plates were spotted by individual colonies producing Pediocin PA-1, and incubated for 5h at 37^0C. After overlaying the plates with *L.plantarum* 8104, and incubation overnight at 30^0C, the size of the growth inhibition zones was examined. It was observed that the radius of the *L. plantarum* inhibition zones decreased by increasing NaCl concentrations. Based on the results of this particular experiment, 0.1% NaCl (~0.02M) was used in the following screening experiments. Using such low concentration allows for the selection of both salt-sensitive and salt-favoring peptides.

4.3 Pilot Study with Pediocin-Mutant Library

To demonstrate the use of a periplasmic-leaky *E. coli* strain as a host to construct and screen AMP libraries, a well-studied bacteriocin Pediocin PA-1 was selected as the starter peptide for the pilot study AMP library. Pediocin PA-1 is a 44 amino acid long peptide possessing two disulfide bonds, a net charge of +6 (including histidine residues) and 34% hydrophobic residues. Pediocins produced by various strains of *Pediococcus acidilactici* were shown to display antimicrobial activities against *Listeria, Leuconostoc,* and *Lactobacillus* species, hence, have been commonly used as food preservatives [69, 101]. There are numerous studies related to Pediocin, such as analysis of structure-function relationships, determination of critical residues, and characterization of chimeric variants with improved and broader range antimicrobial activities [11, 95, 102, 113].

Amino acid Sequence of Pediocin PA-1.

4.3.1 Results

E. coli JE5505 cells, transformed with Pediocin-mutant library, were added to melted soft agar LB, and poured in petri dishes for activity screening by agar diffusion assay. In this pilot study, approximately $5x10^4$ colonies were screened against *L. innocua* 33090 which is about 2 fold the library and several clones with various antimicrobial activities were selected for further analysis.

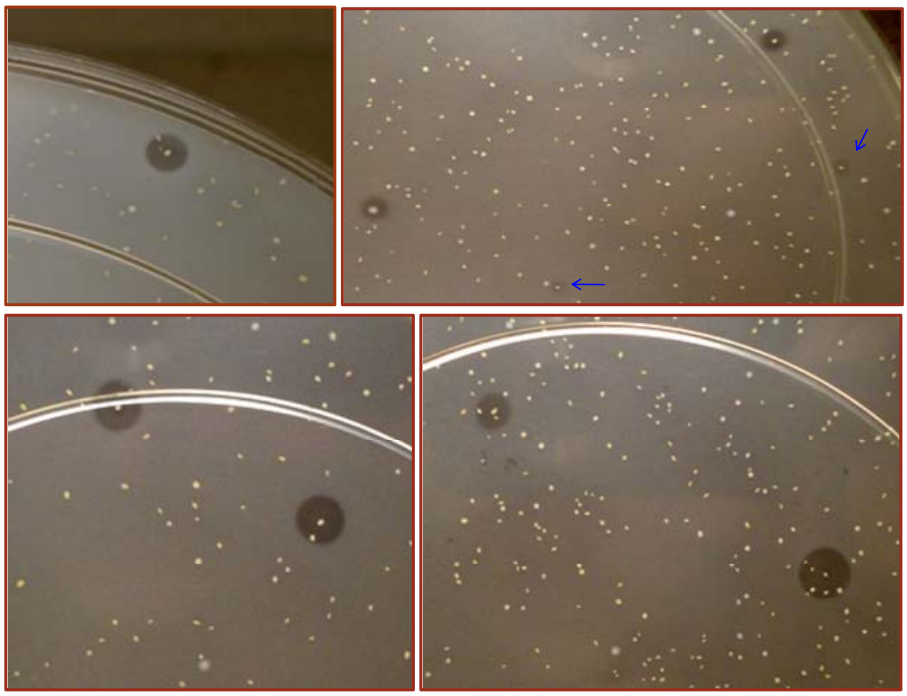

Figure 4.5- AMP library screening plates.

These positive clones were isolated, their antimicrobial activities against *L. innocua* were verified, and their sequences were identified by DNA sequencing. Examples of selected peptides with single and dual mutations are shown in Figure 4.6 and Figure 4.7, respectively. The antimicrobial activities of these mutated peptides were compared to the activity of wild type peptide (WT) to study the effect of corresponding mutations on peptide activity. Their amino acid sequences, clone names used in this text, as well as their scientific names (showing the specific mutation present in that peptide) are provided in Table 4.2 and Table 4.3.

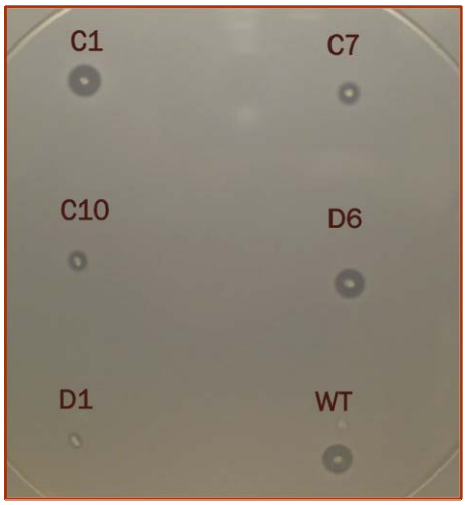

Figure 4.6- Activities of Pediocin variant peptides with a single mutation against *L. innocua* 33090.

Table 4.2- Amino acid sequences of Pediocin PA-1 mutants shown in Figure 4.6. Only C-terminal of each peptide (residue 18 to 44) is shown in the table. Mutations are shown in red.

CLONE NAME	SCIENTIFIC NAME	SEQUENCE 18 25 30 35 44	ACTIVITY
WT	Ped-PA1	**W G K A T T C I I N N G A M A W A T G G H Q G N H K C**	
C1	Ped-PA1 [T27]	W G K A T T C I I **T** N G A M A W A T G G H Q G N H K C	↑
C7	Ped-PA1 [K28]	W G K A T T C I I N **K** G A M A W A T G G H Q G N H K C	↓
C10	Ped-PA1 [V19]	W **V** K A T T C I I N N G A M A W A T G G H Q G N H K C	↓
D1	Ped-PA1 [I29]	W G K A T T C I I N N **I** A M A W A T G G H Q G N H K C	↓
D6	Ped-PA1 [P19]	W **P** K A T T C I I N N G A M A W A T G G H Q G N H K C	wt

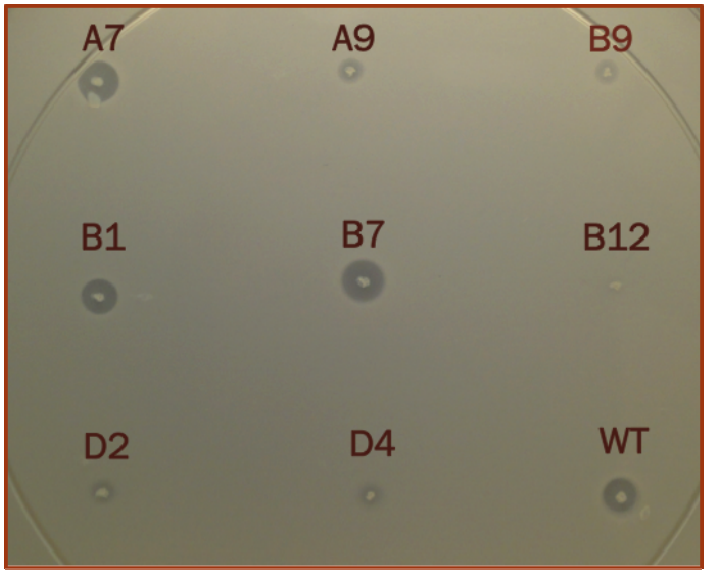

Figure 4.7- Activities of Pediocin variant peptides with two mutations against *L. innocua* 33090.

Table 4.3- Amino acid sequences of Pediocin PA-1 mutants shown in Figure 4.7. Only C-terminal of each peptide (residue 18 to 44) is shown in the table. Mutations are shown in red.

CLONE NAME	SCIENTIFIC NAME	SEQUENCE 18 25 30 35 44	ACTIVITY
WT	Ped-PA1	**WGKATTCIINNGAMAWATGGHQGNHKC**	
A7	Ped-PA1 [E27, L35]	WGKATTCII**E**NGAMAWA**L**GGHQGNHKC	↑
A9	Ped-PA1 [L31, I34]	WGKATTCIINNGA**L**AW**I**TGGHQGNHKC	↓
B1	Ped-PA1 [H20, T41]	WG**H**ATTCIINNGAMAWATGGHQG**T**HKC	wt
B7	Ped-PA1 [A20, G23]	WG**A**AT**G**CIINNGAMAWATGGHQGNHKC	↑
B9	Ped-PA1[F26, A29]	WGKATTCI**F**NN**A**AMAWATGGHQGNHKC	↓
B12	Ped-PA1 [V20,R28, S42]	WG**V**ATTCIIN**R**GAMAWATGGHQGN**S**KC	↓
D2	Ped-PA1 [A33, Y34]	WGKATTCIINNGAMA**A**Y**T**GGHQGNHKC	↓
D4	Ped-PA1 [D19, I36]	W**D**KATTCIINNGAMAWAT**I**GHQGNHKC	↓

The mutations in this document are coded as follows: single letter amino acid code for WT residue/amino acid position in WT peptide/single letter amino acid code for the mutation (for example: N28E in peptide A7).

As seen in Figure 4.6 and Figure 4.7, peptides C1, A7, and B7 were slightly more active than WT Pediocin, while B1 and D6 maintained wild type activity. A single mutation of N28K in C7 resulted in 40-50% loss of activity, while peptides A9, B9, D4 with dual mutations displayed 20-40% relative activities. For peptide D2, replacement of tryptophan at position 33 with alanine in combination with the mutation A34Y had a detrimental effect on the activity. This is consistent with other findings in literature [11] where W33 was shown to be an essential residue for pediocin activity. The role of this tryptophan on peptide structure is believed to be a stabilizer for the hairpin-like structure by positioning in the membrane-water interface, hence critical to diffusion into the hydrophobic core of the membrane. For peptide B12, substitution of arginine at N28 in combination with valine at K20 with an extra mutation of H42S caused complete activity loss, which is probably due to disruption in the alpha-helix domain of the peptide.

Replacement of glycine with valine at position 19 decreased peptide activity (C10), however, when the same residue is replaced with proline (D6), peptide activity was not altered. Although these three amino acids are all non-polar, valine is known to be more hydrophobic and bulky compared to small amino acids glycine and proline. The fact that proline is commonly found in tight turns of proteins and residue 19 is next to the hinge between two regions of the peptide can account for preserving the wild-type activity. In a similar case, substitution of isoleucine at G29 (clone D1) diminished the antimicrobial activity almost completely. Besides being highly hydrophobic, both valine and isoleucine are also C-beta branched amino acids meaning that there is more bulkiness near the peptide backbone, which makes it more difficult for these amino acids to adopt an alpha-helical confirmation [114]. Since the C-terminal of pediocin is an alpha-helix, disturbance in the structure may

explain the activity loss seen in peptides C10 and D1, where glycine was replaced with either valine or isoleucine.

4.4 Construction and Screening of the Plantaricin-423 Mutant Library

Plantaricin-423 (or Pln-423) is a small and heat-stable bacteriocin produced by *Lactobacillus plantarum* 423 isolated from sorghum beer [115]. It is a 37 amino acid long peptide with bactericidal activities against many foodborne pathogens and spoilage gram positive bacteria including *Listeria spp., Leuconostoc spp., Staphylococcus spp., and Lactobacillus spp.* [57, 115]. The amino acid sequence of Pln-423 is shown below; the N-terminal region is marked from the 1st to 17th amino acid, and C-terminal region is marked from the 18th to 37th amino acid, where the hinge between two regions is shown in the color orange.

Amino acid Sequence of Plantaricin-423.

The crystal structure of Pln-423 is not yet available on protein databases, however, due to conserved sequence regions across all Class II-a members, its structure may be very similar to the proposed structure for the peptides belonging to the same group (as explained in Section 4.1.1). In fact, when pln-423 residues are placed on the anticipated structure, the characteristics of the residues align very well with structural domains based on the positions of the disulfide bridges and the hinge between two domains. For the interest of this project, Figure 4.8 shows only certain C-terminal residues to demonstrate how they fit in the α-helix formation and

disulfide bridges of the peptide. These residues were also used to draw a helical wheel diagram for the helix domain of the Pln-423. All helical wheel diagrams included in this document were generated by an online drawing program at http://rzlab.ucr.edu/scripts/wheel/wheel.cgi [116].

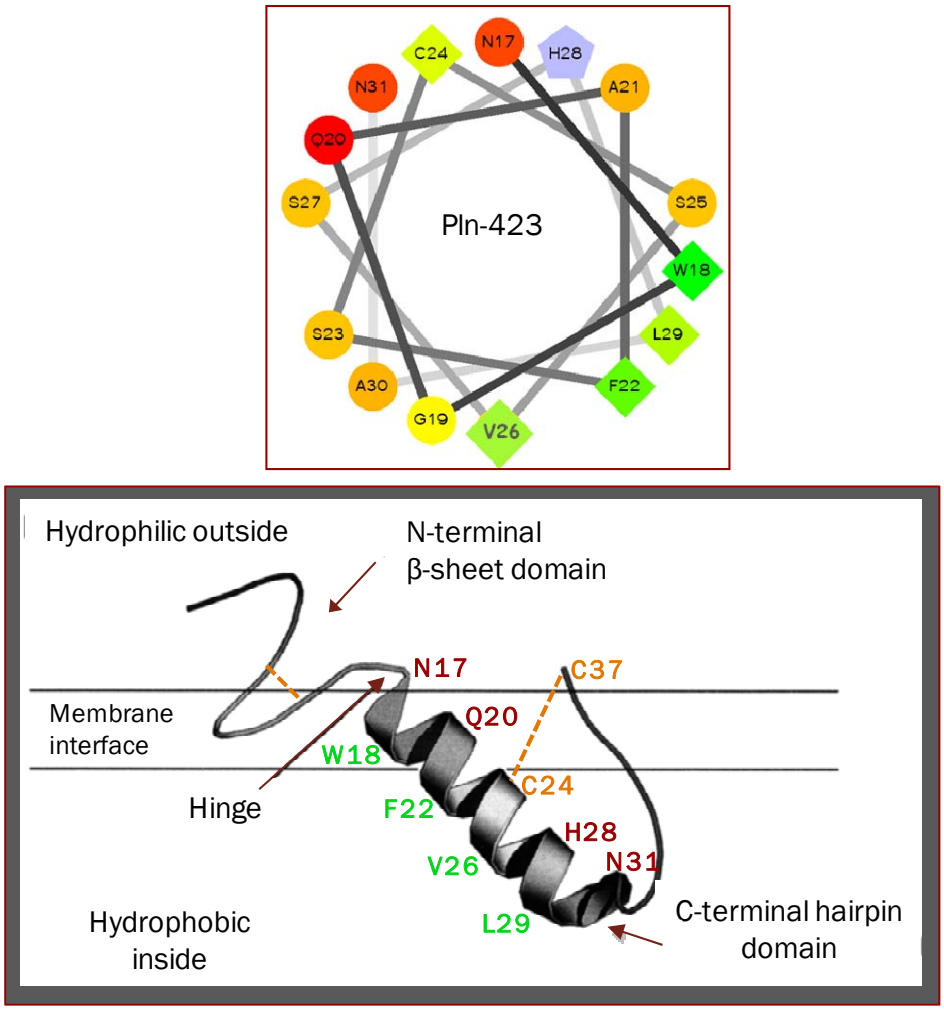

Figure 4.8- Alpha helical wheel diagram of the residues 17-31 and their positions on the proposed structure of Class IIa bacteriocins. Hydrophobic residues (shown in green) are aligned on one side of the helix (facing the hydrophobic core of the membrane) while hydrophilic residues (shown in red) are aligned on the other side of the helix (towards hydrophilic outside). Cysteine residues and their disulfide bridges are shown in orange. (Second figure is modified from Ref [102])

The N-terminal domain of plantaricin-423 is highly homologous to that of pediocin PA-1, but their C-terminal regions are quite different in sequence, accounting for the diversity in their antimicrobial activity spectrum. There are several studies on screening pediocin-PA1 derivatives by introducing random mutations into wild-type peptide [11], or by generating chimeric peptides for the search of candidates with higher activities [113], however, the studies involving Pln-423 are quite limited to characterization of the peptide and fermentation optimization for its production. In this study, a mutant library of Pln-423 was constructed and screened for the determination of critical residues for its function, generation of mutagenesis-activity relation data to better understand its interaction with target bacteria, and hopefully, identification of novel plantaricin derivatives that are more active than the wild-type peptide.

The library containing ~14K peptides was prepared as explained in Section 4.2. Mutations were introduced to only the C-terminal of Pln-423 using the same design guidelines applied to the pediocin PA-1 mutant library. Approximately $1.2x10^5$ clones (~10 fold coverage of the library) were screened against *Listeria innocua* ATCC 33090, and selected positive hits were further analyzed by DNA sequencing.

4.4.1 Results

From the initial library screening, a total of 100 colonies were selected based on the size of their inhibition zones for further analysis. This selection process involved two criteria; **1-** the size of the inhibition zone around each positive colony was compared to that of wild-type Pln-423 as a correlation to anti-listerial activity level, and only the clones displaying different activities were selected, **2-** when several colonies formed inhibition zones that are very similar in size and character

(clarity and distinction), only a few clones from each group were selected. These clones were re-screened to verify and compare their activities and their plasmids were extracted and analyzed by DNA sequencing to obtain peptide sequences. DNA sequencing showed that some of the clones that were selected from the library screening were expressing the same peptides despite the selection criteria that were followed, thus these clones were removed from the following experiments.

Relative activities were calculated by assigning 100% to the radius of the inhibition zone around wild-type Pln-423 and normalizing the values of mutants in percentage based on the difference between radiuses. For example, Pln-2 has a relative activity of 224% which means that its radius of inhibition zone is ~2.2X larger than that of WT; however, Pln-14 has a relative activity of 47% meaning that the inhibition zone of Pln-14 is half the size of the zone around WT. It should be noted here that the percent representation of activities as a function of the radius is only for comparison purposes. Although the size of the radius of the inhibition zone is a good indication of activity and commonly used as a primary screen to detect active antimicrobial compounds, in reality, a peptide with a twice as big inhibition zone is not necessarily a two times more active peptide. The activity cannot be completely inferred from the radius as the radius is a function of activity, diffusion rate, peptide stability, as well as secretion rate. Therefore, all peptide analogs selected at this primary screen based on their inhibition zones should be subjected to a secondary screen where purified samples of each peptide are analyzed separately to determine more precise data on their antimicrobial activity.

Figure 4.9 and Figure 4.10 show activity screening plates of the 34 clones tested in this study. Their corresponding peptide sequences with both their clone names used in this text for simplicity and their scientific names, and their relative activities are given in Table 4.4 and Table 4.5. Each relative activity percentage

shown is a single value from one experiment, and not an average of several reads. However, each experiment was repeated at least three times and very similar values were obtained in all repeats.

Out of 34 clones shown below, 7 of them have a single mutation, 25 clones have dual mutations, while the remaining 2 clones have 3 mutations. Although the library was designed to introduce a maximum of 2 mutations per peptide, additional mutations were introduced due to occasional errors during oligonucleotide synthesis.

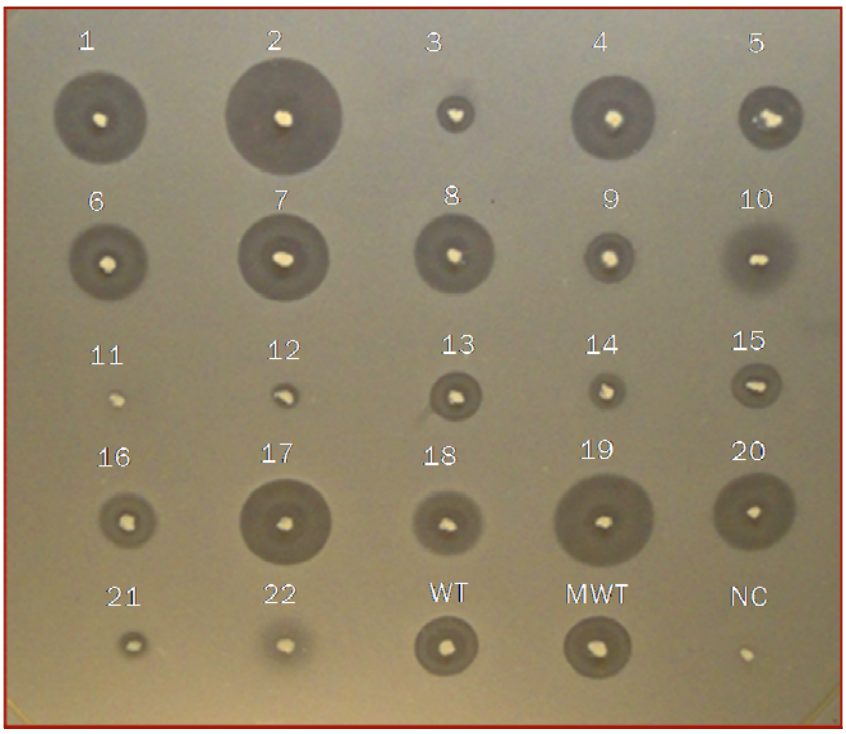

Figure 4.9- Activities of Plantaricin-423 mutants No.1 to 22 against *L. innocua* 33090. Wild-type Pln-423 (WT) and modified wild-type Pln-423 (MWT) was included as positive controls. NC corresponds to the negative control without AMP expression.

Table 4.4- Amino acid sequences of Pln-423 mutants shown in Figure 4.9. Only C-terminal of each peptide (residue 18 to 37) is shown in the table. Mutations are shown in red. Relative activity is calculated based on the radius of the inhibition zones. "NA" means not applicable as the radius of Pln-11 was too small to be accurately measured.

CLONE NAME	SCIENTIFIC NAME	SEQUENCE (C-TERMINAL ONLY) 18 25 30 37	RELATIVE ACTIVITY-%
WT	Pln-423	**W G Q A F S C S V S H L A N F G H G K C**	100
Pln-1	Pln-423 [N25, N36]	W G Q A F S C **N** V S H L A N F G H G **N** C	151
Pln-2	Pln-423 [D27, N36]	W G Q A F S C S V **D** H L A N F G H G **N** C	224
Pln-3	Pln-423 [Y25,W36]	W G Q A F S C **Y** V S H L A N F G H G **W** C	45
Pln-4	Pln-423 [A19, E36]	W **A** Q A F S C S V S H L A N F G H G **E** C	143
Pln-5	Pln-423 [T25, Y28, S36]	W G Q A F S C **T** V S **Y** L A N F G H G **S** C	92
Pln-6	Pln-423 [P19, N36]	W **P** Q A F S C S V S H L A N F G H G **N** C	121
Pln-7	Pln-423 [S30, S36]	W G Q A F S C S V S H L **S** N F G H G **S** C	157
Pln-8	Pln-423 [A34, H36]	W G Q A F S C S V S H L A N F G **A** G **H** C	141
Pln-9	Pln-423 [R28. L34]	W G Q A F S C S V S **R** L A N F G **L** G K C	65
Pln-10	Pln-423 [E34, N35]	W G Q A F S C S V S H L A N F G **E** **N** K C	104
Pln-11	Pln-423 [Y27, C31]	W G Q A F S C S V **Y** H L A **C** F G H G K C	NA
Pln-12	Pln-423 [C25, T31]	W G Q A F S C **C** V S H L A **T** F G H G K C	24
Pln-13	Pln-423 [Y31]	W G Q A F S C S V S H L A **Y** F G H G K C	75
Pln-14	Pln-423 [R25, A31]	W G Q A F S C **R** V S H L A **A** F G H G K C	47
Pln-15	Pln-423 [K31]	W G Q A F S C S V S H L A **K** F G H G K C	72
Pln-16	Pln-423 [K22, R28]	W G Q A **K** S C S V S **R** L A N F G H G K C	89
Pln-17	Pln-423 [E24, R28]	W G Q A F S **E** S V S **R** L A N F G H G K C	160
Pln-18	Pln-423 [E23, R25]	W G Q A F **E** C **R** V S H L A N F G H G K C	108
Pln-19	Pln-423 [M22, D27]	W G Q A **M** S C S V **D** H L A N F G H G K C	206
Pln-20	Pln-423 [A20, D23]	W G **A** A F **D** C S V S H L A N F G H G K C	153
Pln-21	Pln-423 [L26, C35]	W G Q A F S C S **L** S H L A N F G H **C** K C	30
Pln-22	Pln-423 [E25, D35]	W G Q A F S C **E** V S H L A N F G H **D** K C	24

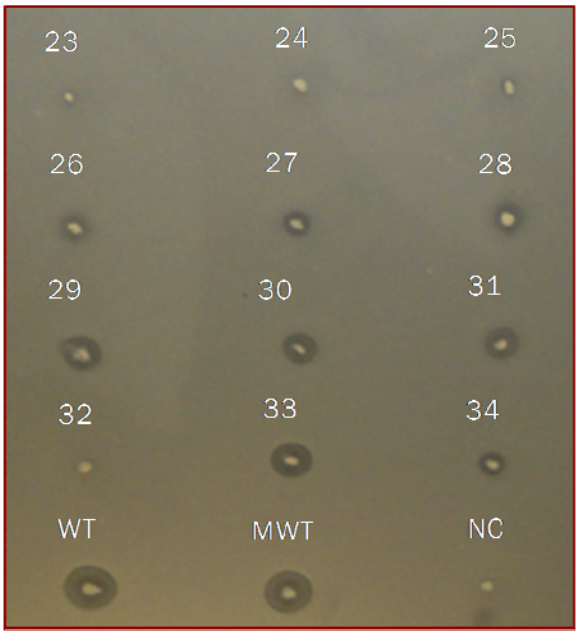

Figure 4.10- Activities of Plantaricin-423 mutants 23 to 34 against *L. innocua* 33090. Wild-type Pln-423 (WT) and modified wild-type Pln-423 (MWT) was included as positive controls. NC corresponds to the negative control without AMP expression.

Table 4.5- Amino acid sequences of Pln-423 mutants shown in Figure 4.10. Only C-terminal sequence of each peptide (residue 18 to 37) is shown in the table. Mutations are shown in red. Relative activity is calculated based on the radius of the inhibition zones. "NA" means not applicable as the radiuses of Pln-23, 24, 25 and 32 were too small to be accurately measured.

CLONE NAME	SCIENTIFIC NAME	SEQUENCE (C-TERMINAL ONLY) 18 25 30 37	RELATIVE ACTIVITY- %
WT	Pln-423	**W G Q A F S C S V S H L A N F G H G K C**	100
Pln-23	Pln-423 [I23]	W G Q A F **I** C S V S H L A N F G H G **W** C	NA
Pln-24	Pln-423 [I23, F25]	W G Q A F **I** C **F** V S H L A N F G H G K C	NA
Pln-25	Pln-423 [F23]	W G Q A F **F** C S V S H L A N F G H G K C	NA
Pln-26	Pln-423 [T22]	W G Q A **T** S C S V S H L A N F G H G K C	29
Pln-27	Pln-423 [A20, S25]	W G **A** A F S C **R** V S H L A N F G H G K C	36
Pln-28	Pln-423 [I21, T25]	W G Q **I** F S C **T** V S H L A N F G H G K C	34
Pln-29	Pln-423 [A19, K26]	W **A** Q A F S C S V S **K** L A N F G H G K C	57
Pln-30	Pln-423 [S19, I26, I27]	W **S** Q A F S C S **I I** H L A N F G H G K C	56
Pln-31	Pln-423 [V19]	W **V** Q A F S C S V S H L A N F G H G K C	56
Pln-32	Pln-423 [H32, W36]	W G Q A F S C S V S H L A N **H** G H G **W** C	NA
Pln-33	Pln-423 [W32]	W G Q A F S C S V S H L A N **W** G H G K C	81
Pln-34	Pln-423 [*18]	***** G Q A F S C S V S H L A N F G H G K C	37

All the mutation-activity analysis below assumes that the difference in radius, hence relative activity percentage, is only due to the activity of the peptide analog and not stability, diffusion, or production rate. Based on the sequencing results of these 34 clones; 10 mutants contained a mutation at Lys-36, 6 of which displayed higher activities, and another 10 mutants contained a mutation at Ser-25, 2 of which were more active compared to wild-type peptide. Similarly, mutations at Gly-19, Ser-23, His-28 and Asn-31 were present in 5 clones displaying high range of activity depending on the particular mutation that was introduced. These results indicate that these 6 residues, especially Lys-36 and Ser-25, play a critical role in peptide activity. However, no mutations were present at Leu-29, Gly-33, and Cys-37 in sequenced clones, which is most likely due to either mutations that did not alter the wild-type activity (probably at Gly-33), or abolished the activity completely (most probably at Leu-29 and Cys-37), thus, these clones were not selected at the library screening for further analysis.

Interestingly, 3 of the most active mutant peptides (Pln-1, 2 and 6) share a mutation of K36N; however, when the same residue was replaced with tryptophan in Pln-3, 23, and 32, the activity was decreased more than 50% or completely terminated. Considering that the rest of the peptides with a mutation at K36 have either E, S, or H, all with equal or higher activities, it can be speculated that when positively-charged Lys at position 36 is mutated with a neutral, or negatively-charged hydrophilic amino acid, the activity is increased, whereas the substitution of Trp, being the largest as well as a very hydrophobic amino acid, most probably disturbed the formation of the disulfide bridge between C24 and C37, hence, the stability of the helix, and consequently decreased the peptide activity. It should be pointed out here that a single mutation of F31W in Pln-33 decreased the activity by only %20 since Trp is an aromatic amino acid like the native residue phenylalanine.

Of course, altered activity of these peptides cannot be attributed to only one mutation; they all have a second mutation at a different position, thus, the interaction between these mutations, their total effect on the peptide structure and position within the membrane should be carefully investigated. For example, Pln-3 and Pln-23 with shared mutation of K36W both carry a second mutation at a serine residue, where serine was replaced with more hydrophobic amino acids tyrosine or isoleucine, respectively. While Pln-3 with S25Y/K36W still displayed a relative activity of 45%, Pln-23 with S23I/K36W lost its activity completely. This is most probably due to isoleucine being very hydrophobic and a bulky amino acid (opposite to serine), hence, affecting the structural interactions between neighboring residues related to serine and disturbing the α-helix conformation, in addition to less stable disulfide bridge induced by tryptophan. Helical wheel diagrams of these peptides (Figure 4.11) show the effects of the mutation type and position on helix formation. Y25 in Pln-3 is positioned on the hydrophobic side of the helix, whereas I23 is closer to the hydrophilic residues.

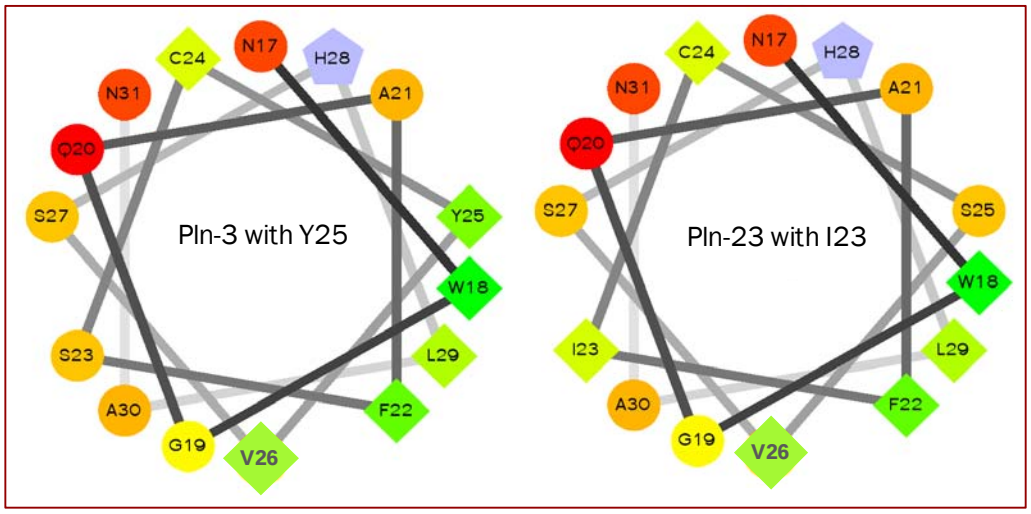

Figure 4.11- Helical wheel diagrams of Pln-3 and Pln-23. Hydrophobic residues are shown in gradient color from green to yellow with decreasing hydrophobicity, and hydrophilic residues are shown in red. Orange residues are partially hydrophobic. Charged and hydrophilic residues are shown in blue.

Replacement of serine residues with negatively-charged and more hydrophilic amino acids, such as Asp and Glu, generally led to increase in peptide activity as seen in Pln-2, 18, 19 and 20 (with the exception of Pln-22 carrying S23E and G35D). In fact, two most active clones (Pln-2 and Pln-19) share a mutation of S27D along with K36N and F22M, respectively, which resulted in more than a 100% increase in peptide activity based on the agar-diffusion assay. This is most likely due to S27 being on the hydrophilic side of the helix (Figure 4.12), thus, the substitution of a much more hydrophilic amino acid, asparagine, improved the helical structure. In addition, secondary mutations of these peptides were very similar to their native residues in terms of hydrophobicity.

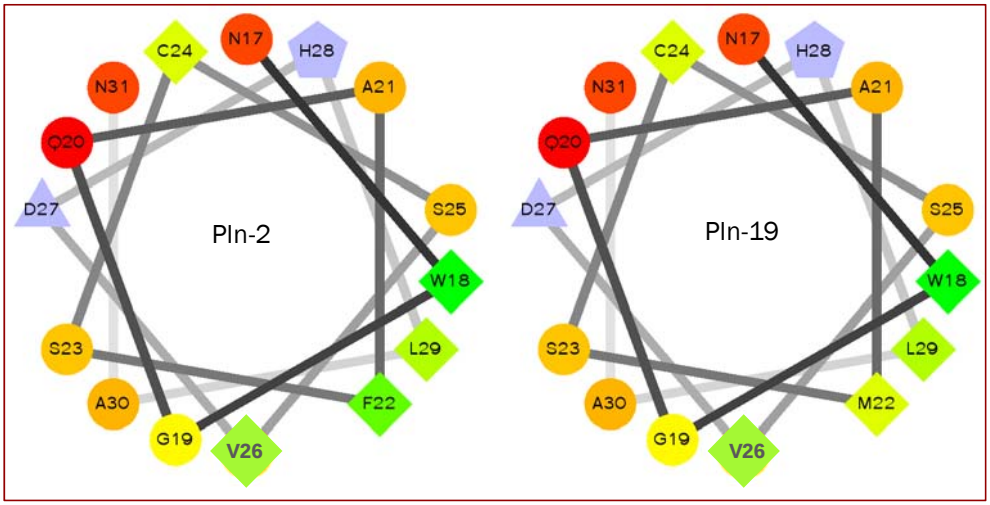

Figure 4.12- Helical wheel diagrams of Pln-2 and Pln-19.

Cysteine substitution in mutants Pln-11, 12, and 21 was always detrimental to peptide activity as expected. Although the replacement of N31 with a cysteine caused complete activity loss, the activity was reserved around 20-30% when substitutions were positionally close to native cysteines, most likely because the new disulfide bridge between two cysteines did not alter the secondary structure of the peptide as much as the bridge between mutation Cy31 and native C37 did. On

the other hand, when one of the cysteines, C24, was mutated with glutamate in Pln-17 in combination with the second mutation of H28R, the activity of the wild-type Pln-423 was improved about 60%. This result is rather surprising since the mutant peptide now lacks one of the disulfide bridges between C24 and C37, which was previously shown to be essential for the activity of a similar peptide pediocin-PA1 [11]. However, it could be speculated that the substitution of very hydrophilic amino acids in both positions improved the amphiphilic character of the helix domain, and maintained the stability of the structure by a hydrogen bond formed between negatively-charged glutamate (position *i=24*) and positively charged arginine (position *i+4=28*) [117]. Another peptide that is worth mentioning here is Pln-16 with mutations of F22K and H28R. This peptide shares the mutation H28R with Pln-17; however, its other mutation replaced a hydrophobic residue, phenylalanine, with a charged hydrophilic amino acid, Lys. Even though Pln-16 carried helix-enhancing arginine at H28 and preserved its disulfide bridge unlike Pln-17, substitution of lysine at the opposite side of the helix, where it faces the hydrophobic core of the membrane, resulted in slight activity loss.

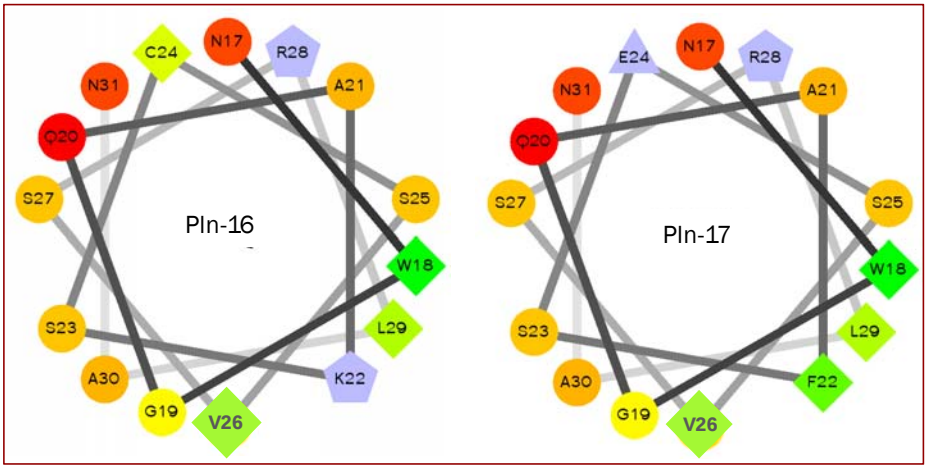

Figure 4.13- Helical wheel diagrams of Pln-16 and Pln-17.

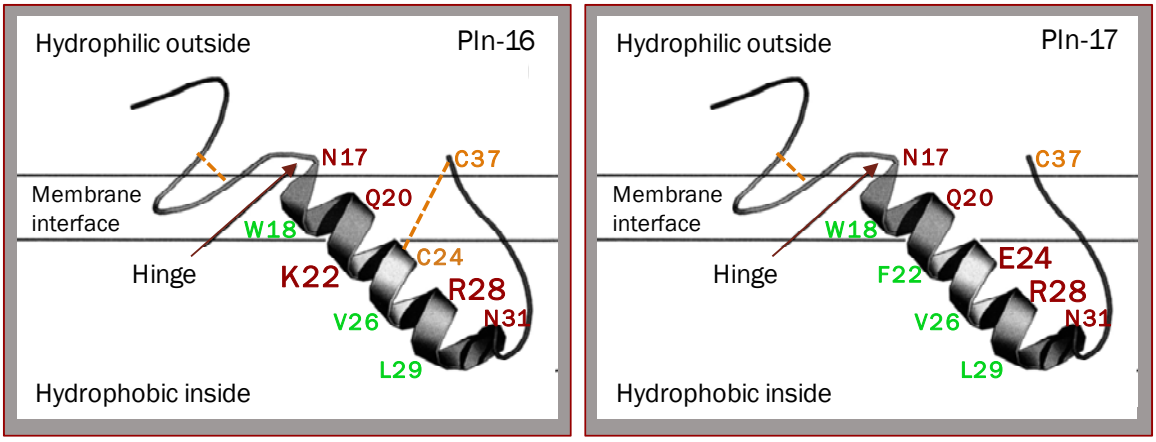

Figure 4.14- Position of the mutations on Pln-16 and Pln-17's proposed structure at membrane interface. Mutated residues are shown in larger font (Green color represents hydrophobic amino acids, red color represents hydrophilic amino acids) (Modified from [102]).

Lastly, Pln-34 carries a stop codon at W18, which means that this peptide consists of only an N-terminal domain, and it displays 37% relative activity compared to wild-type. This particular peptide shows the essential role of the C-terminal domain on WT Pln-423 activity, accounting for 60% of total activity, and also agrees with the predictions on the N-terminal's function for initial cell interaction and binding [103].

4.4.2 Effect of Salt Concentration on Peptide Activity

Effect of NaCl on Wild-Type Bacteriocins

The effect of NaCl concentration (used in the expression media) on the activity of wild-type bacteriocins was investigated. Wild-type pediocin-PA1, plantaricin-423, and leucocin-A (37aa, also member of class-IIa bacteriocins) were studied for their activity against both *Listeria innocua* 33090 and *Lactobacillus plantarum* 8104 on increasing salt concentrations. Colonies expressing wild-type peptides were placed on expression plates containing increasing concentrations of

NaCl from 0.02M up to 1.0M (*L. innocua,* Figure 4.15) or to 0.2M (*L. plantarum,* Figure 4.16). To determine the peptide activity on varying NaCl concentrations based on the size of the inhibition zone around the colonies, the radius of the zones on each plate was measured and these values were compared to each other.

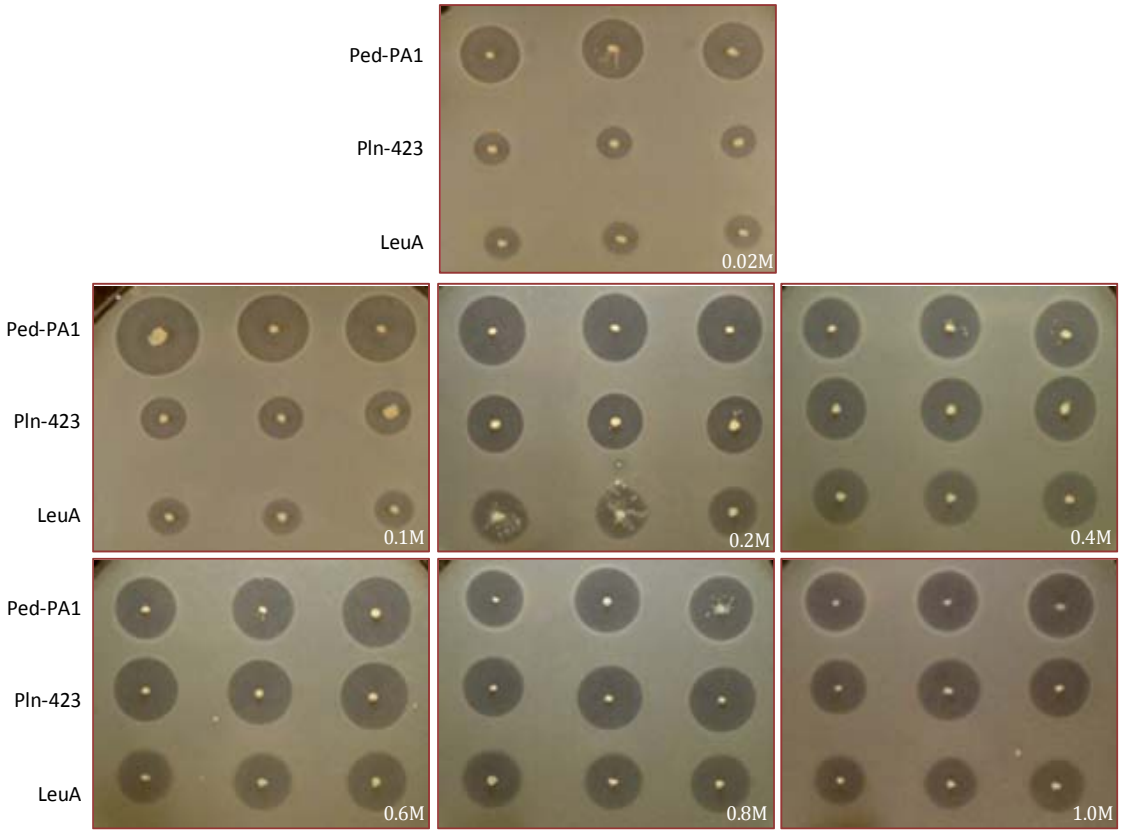

Figure 4.15- Activity of wild-type bacteriocins Pediocin-PA1, Plantaricin-423 and Leucocin-A against *L. innocua* 33090 on increasing NaCl concentrations from 0.02M to 1.0M. Each clone was spotted in 3 replicates per row.

Table 4.6- Average radius (mm) of the inhibition zones shown in Figure 4.15. Values shown in red indicate the radius of the largest zone formed by the peptide.

Peptide	NaCl Concentration						
	0.02M	**0.1M**	**0.2M**	**0.4M**	**0.6M**	**0.8M**	**1.0M**
Ped-PA1	4.61	**5.72**	5.68	5.22	5.23	5.27	5.25
Pln-423	2.40	3.32	4.41	5.24	**5.60**	5.56	5.37
Leu-A	2.62	2.82	3.96	4.56	**4.89**	4.87	4.30

As seen in Figure 4.15, all wild-type peptides displayed higher activities against *L. innocua* on NaCl concentrations of 0.1M or higher. Pediocin-PA1 displayed 24% more activity when expressed on 0.1M than on 0.02M of NaCl, but showed slight decrease with increasing salt concentrations after 0.1M. The activity of both plantaricin-423 and leucocin-A improved with increasing concentrations of NaCl up to 0.6M, then decreased slightly when expressed on 1.0M NaCl. Percent increase in the radius of inhibition zones, from 0.02M to 0.6M plates, was 133% for plantaricin-423, and 86% for leucocin-A.

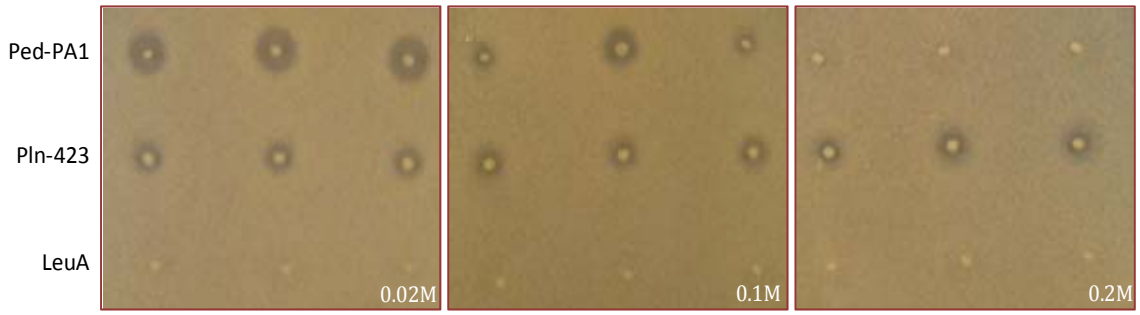

Figure 4.16 Activity of wild-type bacteriocins Pediocin-PA1, Plantaricin-423 and Leucocin-A against *L. plantaricin* 8104 on increasing NaCl concentrations from 0.02M to 0.2M. Each clone was spotted in 3 replicates per row.

In contrast to the results with *L. innocua*, both pediocin PA-1 and plantaricin-423 displayed much lower activities against *L. plantarum* while Leucocin-A was

completely inactive. Confirming the initial findings in this study (Section 4.2.3) and literature [118], pediocin PA-1 lost its activity against *L.plantarum* on higher NaCl concentrations and became inactive after 0.2M NaCl. However, although the activity of plantaricin-423 against *L.plantarum* is quite low, it stayed constant with increasing salt concentrations.

These results together clearly demonstrate how much the behavior of a peptide varies in two different environments. The activity of pediocin PA-1 against *L. innocua* increased more than 20% when expressed on 0.2M NaCl, but it was completely inactive against *L. plantarum* at the same NaCl concentration. The constant activity of plantaricin-423 indicates that the interactions between salt ions and *L. plantarum* membrane did not affect peptide activity, thus, the structural changes induced by NaCl probably caused activity loss for pediocin PA-1 while plantaricin-423 activity was not altered. It should be noted that concentrations of NaCl higher than 0.2M could not be tested due to growth inhibition of *L. plantarum* on high salt-containing media.

Effect of NaCl on Plantaricin-423 Mutant Peptides

For the analysis of plantaricin-423 mutant peptides, the 9 most active peptides displaying 20% or more relative activity compared to wild-type peptide were selected, and studied for their altered activities against *L. innocua* 33090 on increasing NaCl concentrations. The colonies expressing these mutant peptides were placed on expression plates with varying NaCl concentrations of 0.02M to 0.2M, and the experiment was performed as explained in the previous section.

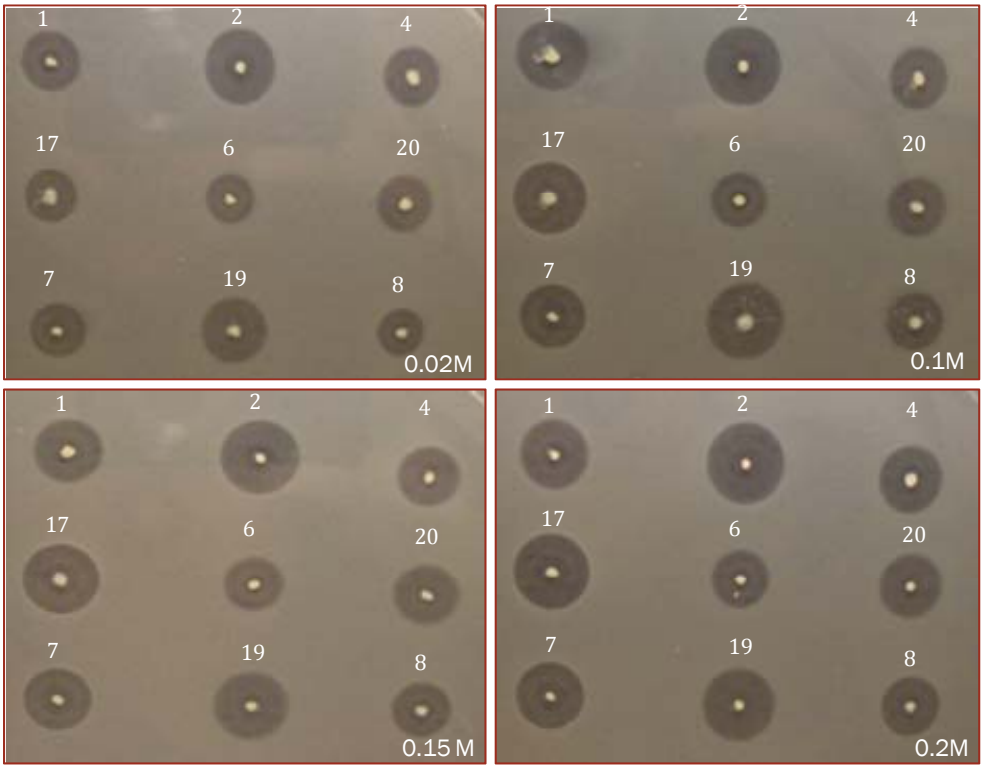

Figure 4.17- Activity of selected Pln-423 mutants against *L. innocua* 33090 on increasing NaCl concentrations from 0.02M to 0.2M. (Peptide numbers are shown above each colony.)

Table 4.7 – Radiuses of the inhibition zones (in mm) and the percent increase.

Peptide Name	NaCl Concentration				
	0.02M	**0.1M**	**0.15M**	**0.2M**	**% increase**
Pln-1	3.67	3.91	4.11	4.14	**12.8**
Pln-2	4.63	4.90	5.10	5.21	**12.5**
Pln-4	3.16	3.79	3.90	3.92	**23.4**
Pln-17	3.06	4.15	4.50	4.88	**59.4**
Pln-6	2.83	3.19	3.45	3.52	**24.3**
Pln-20	3.35	3.77	4.03	4.05	**20.5**
Pln-7	3.49	3.92	4.24	4.36	**24.9**
Pln-19	4.28	4.56	4.66	4.75	**10.9**
Pln-8	2.81	3.41	3.68	3.86	**37.3**

The results of this experiment showed that, the radius of *the L. innocua* inhibition zones increased ~10-60% around all colonies with increasing NaCl concentrations (up to 0.2M). This positive effect of higher ionic strength on *L. innocua* inhibition can be explained by two factors: the increased sensitivity of *Listeria* cells to high NaCl concentrations due to higher osmotic pressure, and the interaction between polar side chains of amino acids in the presence of salt ions. These two factors are interrelated; higher concentrations of NaCl in the growth media may cause changes in the composition of the cell membrane, hence, in addition to salt-induced conformational changes within the peptide, the orientation of the peptides in the membrane and their interaction with membrane components may also vary leading to altered antimicrobial activity.

While bacterial cells are generally sensitive to high concentrations of salt, increase in the *L. innocua* inhibition around the colonies cannot be attributed to only higher sensitivity of target cells, due to the large variance in the % increase of the zones between the peptides (Table 4.7). Especially, Pln-17 showed a 59.4% increase in the inhibition zone radius, becoming the second most potent mutant when expressed on media containing 0.2M NaCl. This clone lacks one of the two disulfide bridges, but carries an extra pair of charged amino acids, glutamate and arginine, forming a helix stabilizing bond. It was previously shown that the helicity of short α-helical peptides increases due to interactions between NaCl ions and charged-side chains [117]. Of course, such interactions are specific to behavior of each peptide in certain environments, and changes in peptide structure and orientation, hence activity, should be studied in a separate manner. In another experiment, these mutant peptides were tested against *L. innocua* on higher NaCl concentrations up to 0.8M. A general decrease in the radius of the inhibition zones was observed for all clones after 0.4M, and they all became inactive on 0.8M NaCl.

4.5 Chapter Conclusion and Discussion

A periplasmic-leaky *E.coli* expression system was employed for the production and screening of antimicrobial peptide libraries in bacteria. This system allows for the release of the expressed peptides to the growth media, thus, recombinant peptides can be directly screened for their antimicrobial activities where the activity of a peptide can easily be linked to its sequence. These peptides can possess activities against both gram positive and gram negative bacteria, hence, to prevent the toxicity of recombinant peptides to the host organism *E. coli*, a special group of AMPs, called bacteriocins, that are active only against gram positive bacteria, were selected to be investigated in this study.

To demonstrate that oligonucleotide libraries can be translated into peptide libraries using this bacterial expression and screening system, custom libraries of two class II-a bacteriocins, Pediocin PA-1 and Plantaricin-423, were constructed. By introducing single and dual mutations into the C-terminal domain of the wild-type peptides, libraries that contain between ten to twenty thousand peptides were generated. A library of ~20K peptides can still be easily screened by 10 fold coverage, while representing all single mutations and a limited number of dual mutations in the library. The oligonucleotide libraries were synthesized, amplified by emulsion PCR, and cloned in *E. coli JE5505* for the screening of mutated peptides for their altered antimicrobial activities.

A pilot study by screening approximately $5x10^4$ clones was performed to analyze pediocin-mutant library, and several clones were selected to identify the mutations responsible for the peptide's activity change. By obtaining these sequences and comparing their activities with the wild-type pediocin PA-1, mutations that lead to complete loss or improvement of the antimicrobial activity were identified.

More extensive screening was performed to study the plantaricin-423 mutant library. Activities of 34 clones were compared to each other and to wild-type pln-423 to understand the role of both native and mutated residues on peptide activity. Nine of these clones displayed 20% or more relative activities than the WT. Mutations such as K36N and S27D were commonly present in most active peptides, while other mutations at the same residues led to a decrease in the activity. Generally, when a native residue was mutated with an amino acid that has an opposite hydrophobicity, polarity or charge, the peptide activity was either decreased or completely vanished. This is most probably due to alterations in the α-helix structure of the C-terminal. Cysteine substitutions were always deleterious to peptide activity, whereas replacement of a native cysteine with glutamate (in combination with the second mutation) improved anti-listerial activity. Finally, when the wild-type peptide was expressed only with its N-terminal domain, due to a stop codon introduced at W18, its activity was diminished by almost 60%.

The last set of experiments performed during this study was on investigating the effects of NaCl on the activity of bacteriocins. Wild-type plantaricin-423 and its derivatives displayed higher activities against *L. innocua* 33090 as the concentration of NaCl increased up to 0.8M and 0.4M, respectively. However, activity of WT pln-423 against *L. plantarum* 8104 stayed constant on increasing salt, while its mutants did not inhibit *L. plantarum* growth at all indicating that these mutated peptides possess *Listeria*-specific antimicrobial activity. The results with pediocin PA-1 showed that, while the activity of this peptide against *L. innocua* is much higher on 0.2M NaCl, it becomes completely inactive against *L. plantarum* on the same NaCl concentration. Thus, it can be concluded that antimicrobial activity depends on the synergetic effects of the interactions between salt ions, target organism and the peptide itself, and it presents great variation for each peptide-target cell pair.

All results obtained during this project are based on the evaluations of the colony-overlay assay used for the detection of the antimicrobial activity. This method allows for rapid screening of libraries by observation of the plates with a naked eye, and the peptide activities can be expressed as a function of the size of the inhibition zones formed around host colonies. However, it should be noted that, since this method relies on many factors, such as the size of the host colonies and their expression levels, physical state of the both producer and indicator organisms, incubation duration, and overlay volumes, it may show great variance between experiments and extra attention should be paid to maintain consistency. Therefore, due to possible variability of colony-overlay assay, data on peptide activity gathered in this study may present slight variations between experiments, thus, more precise data may be obtained by testing high purity solutions of each candidate peptide to quantify their antimicrobial activity and determine their MIC values.

Screening of both pediocin-PA1 and plantaricin-423 libraries resulted in detection of clones with multiple mutations. These libraries were designed to have peptides with single or dual mutations; however, due to synthesis errors or recombination events occuring during emPCR, some peptides carry unintended mutation combinations. This is actually advantageous for screening mutant libraries as it increases the diversity of the clones tested.

This chapter demonstrated that by combining parallel oligonucleotide synthesis and *E. coli* secretory expression system, customized peptide libraries can be constructed for the selection and identification of novel antimicrobial peptides. By employing the method developed in this project, libraries can be screened for the discovery of novel peptides with superior activities and a thorough mutagenesis study can be performed to determine the role of each and every residue on peptide activity by analyzing any desired number of clones and identifying their sequences.

CHAPTER 5

CONCLUSIONS AND FUTURE RECOMMENDATIONS

A novel methodology for the construction and screening of custom peptide libraries has been developed in this study by combining high-throughput parallel oligonucleotide synthesis with biological expression systems. This method will allow future researchers to economically generate fully defined libraries containing hundreds of thousands of peptides and rapidly screen for their desired properties. A successful application of this novel technique was demonstrated by constructing and screening libraries coding for antimicrobial peptides which identified a number of mutants more active than the wild type peptide.

Critical steps of the library preparation process, specifically generation of the oligonucleotide sequences and amplification of the library, were carefully studied and procedures associated with these steps were optimized to maintain library complexity. A variety of host systems, two yeast strains and a bacterial strain, were investigated for their feasibility to express and secrete antimicrobial peptides. Two AMP libraries were constructed and screened in the bacterial host and altered activities of selected mutant peptides were evaluated in depth based on their mutations to better understand the relations between their structural properties and antimicrobial activity.

5.1 Summary of Conclusions

The following summarizes the findings and accomplishments of this work:

Library preparation

- For the generation of oligonucleotide sequences coding for the peptide library, two design criteria; codon optimization for expression and synthesis, were established to reverse-translate the amino acid sequences into DNA sequences.
- Emulsion PCR (emPCR) technique was employed for the amplification of the oligonucleotide library to minimize the formation of PCR artifacts due to high similarity between library fragments. The procedure used for emPCR was optimized by decreasing the template DNA concentration and using a hot-start polymerase, and consequently, the ratio of the wild-type sequence to all sequenced clones was successfully decreased from 56% to 2.0%.

Eukaryotic expression systems

The feasibility of yeast host cells to produce short (<15aa) antimicrobial peptides were investigated by expressing and screening IsCT-[L_6K_{11}] in *Saccharomyces cerevisiae* and *Pichia pastoris*. A longer peptide, Pediocin PA-1, was included as a control in both studies.

Expression in Saccharomyces cerevisiae

- Two expression plasmids carrying *GAL* promoter and α-MF secretion signal were successfully constructed to enable peptide production on both minimal media (using *TRP1* marker) and rich media (using *KAN1* marker).
- IsCT-[L_6K_{11}] expression levels in both systems were below the detectable levels for agar diffusion assay, hence it was concluded that *S. cerevisiae* is not a

suitable host for the construction and rapid screening of a library containing such short peptides without the use of expression enhancers.

- Pediocin PA-1, on the other hand, was successfully produced in both systems (with a significantly higher yield on rich media), and its antimicrobial activity was detected by various screening assays. These results verified that the expression systems that were constructed for *S. cerevisiae* were capable of producing longer AMPs, hence, can be used for the construction of libraries that contain peptides similar to Pediocin PA-1 in length.

Expression in Pichia pastoris

- The *P. pastoris* expression system carried the SUMO fusion tag in-frame with IsCT-[L_6K_{11}] to improve the production yield of the recombinant peptides while protecting it from proteolytic degradation. The results showed that IsCT-[L_6K_{11}] was successfully expressed in fusion to SUMO3 protein at sufficient levels, and cleaved peptides, which were around 50% of the total fusion products, preserved their antimicrobial activity.

- However, it was found that the cleavage of the fusion products were due to endogenous proteases as further separation could not be achieved with commercial SUMO protease treatment. Despite the limitation with separation yield, cleavage of the fusion tags by endogenous proteases during secretion simplifies the screening experiments; therefore, *P. pastoris* expression system has potential for the construction of libraries containing short peptides, as long as the libraries are screened on multiwell plates where each host colony cultured in isolation.

Construction of AMP libraries *Escherichia coli*

Two mutant AMP libraries, coding for derivatives of Pediocin PA-1 and Plantaricin-423, were constructed in *E. coli* and screened by a modified colony

overlay assay to select for mutants with varying activities and study the relations between the mutations introduced and their effects on antimicrobial activity.

- A pilot study with the pediocin PA-1 mutant library demonstrated that the *E. coli* expression system was suitable for the screening of large collections of mutated AMPs for their altered activities. Several clones were analyzed for their mutations in relation to peptide's altered activity and the ones that led to complete loss or improvement of the antimicrobial activity were identified.

- The Plantaricin-423 mutant library was studied more thoroughly where 34 clones were selected and subjected to in-depth analysis of mutation-activity relations. Nine peptides displayed 20% or more relative activity, two of which with a significantly higher activity, whereas five peptides lost their activities as a result of the mutations that were introduced. Effects of these mutations on peptide structure, hence activity, were evaluated by analyzing physical properties of amino acids as well as their α-helical wheel diagrams. Certain positions of plantaricin-423 such as K36, C24, and all serine residues were identified as critical since the peptide activity dramatically increased or completely lost based on the type of the mutations introduced at these positions.

- The effect of high salt concentration on antimicrobial activity was also investigated by testing the activities of selected mutants and wild-type bacteriocins on increasing NaCl concentrations against *L. innocua* and *L. plantarum.* It was shown that all wild type bacteriocins (Ped-PA1, Pln-423 and Leu-A) showed higher activities against *L. innocua* as the NaCl concentration increased up to 0.8M, however, their activities against *L. plantarum* either stayed constant (pln-423) or disappeared after 0.1M NaCl. The radius of *the L. innocua* inhibition zones increased around all Pln-423 analogs with increasing NaCl concentrations up to 0.2M. One interesting finding in this study was that mutant Pln-17 showed an almost 60%

increase in activity (as a function of the radius of the inhibition zone), and became the second most potent mutant when expressed on media containing 0.2M NaCl.

5.2 Recommendations for Future Work

As with any type of successful research, there will always be more avenues to investigate once the initial project is completed. The results of this thesis open up many possibilities for potential new research, and some examples are given below.

5.2.1 Continuation of the present work

- Recent advances made in the parallel oligonucleotide synthesis in our group now allow us to generate oligonucleotides up to 200bp long. This means that libraries containing peptides as long as 50 amino acids (plus universal primer binding sites) can be constructed using either bacterial or yeast expression systems investigated in this work. The *S. cerevisiae expression* system, in particular, now presents a great potential to be used for the construction of AMP libraries, especially for peptides that are toxic to both gram-negative and gram-positive bacteria, hence require a eukaryotic host to be produced.
- By changing the screening assay, the *E. coli* expression system can also be used for the construction of libraries that contain host-toxic peptides. Altosaar and co-workers (2008 and 2009) proposed a screening method that relies on a vital colony staining assay for assessing bacterial viability. Basically, *E. coli* colonies, transformed with a plasmid coding for toxic peptides, are grown on an inductive medium containing a vital dye, and upon expression of toxic peptides, the antagonized colonies become colored allowing for visual selection of peptides with high activity. The theory behind this screening method can be adapted to the *E. coli* expression system and AMP libraries can be screened for their activities against the

host organism. Of course, such an expression and screening system presents few limitations that should be addressed, such as the requirement for tighter promoter control and differentiation of low and high antimicrobial activity.

- One limitation that the current work faced was the capacity of the screening method that was employed for analyzing antimicrobial activity. The data from screening experiments were obtained by growing a maximum of 7000 colonies per plate to maintain individual colony formation and selecting and sequencing each positive hit individually. To increase the number of clones in the screening experiments, one way would be to prepare high density plates where each growth inhibition zone contains multiple host colonies and then subjecting these colonies to mass sequencing to identify their mutations responsible for activity. These multiple sequence pools would allow for the characterization of specific peptide regions that are responsible for its activity. After this primary screen, more detailed analysis can be performed to relate each mutation to its specific effect on peptide structure and function.

- An alternative high-throughput screening method for antimicrobial activity is the compartmentalization of both producer and indicator cells in alginate capsules. In this screening assay, the indicator strain is engineered to express an alginate-degrading enzyme, and co-encapsulated with a single producer cell in each capsule. When host cells produce an active peptide and inhibit indicator strain growth, the alginate capsules stay intact while the ones where the indicator strain was able to grow and degrade the capsule are eliminated. The peptide sequence can subsequently be identified by sequencing of the host cell plasmid from the recovered capsules. This method would allow for the high-throughput screening of millions of clones in isolation.

- Screening of Plantaricin-423 library yielded in identification of nine highly active mutant peptides and the comparative analysis of mutation-activity relations between these peptides produced substantial sequence-related data. The data gathered in this study constitutes a solid knowledge base for the design of highly-defined libraries. These follow-up libraries that contain derivatives of the nine pln-423 mutants can be screened to select for peptides with even higher activities or broader range of activity. Speaking of activity range, these nine Pln-423 analogs can be individually tested for their activities against a variety of gram-positive strains such as *Mycobacterium smegmatis*, *Bacillus cereus*, *Enterococcus faecalis*, *and Lactobacillus* species as well as other pathogenic strains to determine their potential for other applications.

5.2.2 Potential future applications

Some of the potential applications of this method include:

- Screening of other AMP libraries; numerous other AMP classes that are of great interest to many medical studies can be screened using the methods developed in this thesis. These include AMPs with activity against highly pathogenic bacteria *Pseudomonas aeruginosa* and *Staphylococcus aureus* which are particularly responsible for more than 90% of infections in cystic fibrosis patients [119]. Another class of AMPs that can be further explored is the ones that exhibit selective activity against cancerous cells but not healthy mammalian cells, hence can be developed as anticancer agents [120]. Custom-designed derivatives of AMPs, which are already identified for their anticancer activities, can be screened for candidates with improved activity against tumor cells and minimal cytotoxicity on healthy cells.
- Discovery of cell-binding peptides; libraries of designer peptides that are derivatives of selected extra cellular matrix components (such as laminin) can be

screened for their binding affinities to cell surface receptors (i.e integrins and heparins) to explore their potential in tissue engineering applications [121, 122].

- Alternative expression and screening platforms such as the yeast-two hybrid system and phage display can also be employed for to construct peptide libraries for the screening of antibody libraries, discovery of drug lead compounds, studying protein-protein and protein-DNA interactions, epitope mapping, and identification of novel receptors and ligands.

— Phage display has been a powerful method for screening large molecular libraries in many applications. A vast number of randomly generated peptides are fused to coat proteins of the filamentous phage particles and screened by a "biopanning" process to isolate phages that present a peptide with higher affinity or specificity to a certain target from an immense background of nonspecific phage [8, 123, 124]. Most commercially available phage libraries display up to 10^9 randomly designed peptides of variable lengths ranging from 7mers to 12mers. Since the enriched sequence is selected from a large pool of random peptides that were non-specifically generated, there is still much room in most cases to improve its specificity and selectivity to the desired target by screening custom-designed derivatives of the peptide. Such custom-build peptide libraries for phage display can be constructed by implementing the method developed in this study, and peptides with optimum desired properties can be identified.

— Yeast two hybrid system is also commonly used to study protein-protein or protein-DNA interactions, and the selection for such specific interactions relies on transcriptional activation of the reporter gene induced by the interaction between the protein of interest (bait library) and the target sequence (prey). Therefore, customized bait libraries can be screened for their interactions with a selected prey molecule.

The work presented in this dissertation enables the production of large customized peptide libraries in a very cost-effective way starting from a sequence file and off-the-shelve DNA monomers, which allows for unprecedented applications both in advanced research and fundamental science. This thesis has just opened the door for not only designing and discovering novel AMPs and other functional peptides in a high-throughput fashion, but also gaining new insights and understanding towards basic molecular research. Potential future research areas are only limited by our imagination.

APPENDIX

EXPERIMENTAL PROTOCOLS

1.1 Emulsion PCR Procedure for Oligonucleotide Library Amplification

This protocol is adapted from Williams et al. *Nature Methods*. 2006.

1.1.1 Preparation of the oil phase

Prepare oil surfactant mixture by vortexing the following components at room temperature and put on ice for 25-30min. Excess oil phase can be stored at 4 °C for several days.

Component	Final concentration	Vol. (2ml)
ABIL EM90	4% (v/v)	80 µl
Triton X-100	0.05% (v/v)	1 µl
Mineral Oil	96% (v/v)	1919 µl

1.1.2 Preparation of the aqueous phase

Prepare 100 µl of PCR solution by mixing the following components in order. The mix should be prepared on ice and kept on ice until the emulsion is prepared. This will make the aqueous phase of the emulsion. Please note that for oligonucleotide libraries cleaved from microarray, the concentration of amplifiable material is usually 1% of the total DNA concentration determined by UV absorbance at 260 nm. Therefore, a 100 fold excess of template should be used.

Components	Volume (Final Conc.)
5X GC Reaction Buffer	20µl (1X)
50mM $MgCl_2$	5µl (2.5mM)
10mM dNTP	2µl (0.2mM)
FWD primer (50pmol/µl)	2µl (1pmol/µl)
RVS primer (50pmol/µl)	2µl (1pmol/µl)
Template DNA	1 template/droplet
BSA (10µg/µl)	5µl (0.5µg/µl)
Phusion Hot Start (5U/µl)	2µl (10units)

1.1.3 Preparation of the emulsion

It is important to maintain the aqueous:oil ratio at 1:6 and keep the volume between 200 – 600 µl for good emulsification.

1. Add 300 µl of oil mixture to an ice-cooled glass vial containing a magnetic stir bar. Begin stirring at 1000rpm on the magnetic stirrer.
2. Add 50 µl of PCR mix to the oil in the glass vial in a drop-wise manner.
3. Stir for 10 minutes at 1000 rpm to emulsify the mixture.
4. Pipette 50 µl aliquots of emulsion to PCR tubes.
5. Use the following conditions for the PCR amplification where the annealing temperature *t* is set 5 ˚C below the lowest Tm of the primers.

   95˚C – 2 min
   95˚C – 10s, *t* ˚C – 15s, 72 ˚C – 20s, up to 30 cycles
   72˚C – 5 min

1.1.4 Post PCR cleanup

1. Pool the emulsion PCR reactions in a 1.6 ml microcentrifuge tube.

2. Add 1µl of a gel loading buffer to visualize the aqueous phase
3. Add 100 µl of mineral oil and vortex 30s.
4. Centrifuge at 13,000*g* for 10 min.
5. Remove the oil (upper phase).
6. Add 1 ml of water-saturated diethyl ether and vortex 1 minute.
7. Centrifuge at 13,000*g* for 1 min.
8. Remove the diethyl ether (upper phase).
9. Add 1 ml of water-saturated ethyl acetate and vortex 1 minute.
10. Centrifuge at 13,000*g* for 1 min.
11. Remove the ethyl acetate (upper phase).
12. Add 1 ml of water-saturated diethyl ether and vortex 1 minute.
13. Centrifuge at 13,000*g* for 1 min.
14. Remove the diethyl ether (upper phase).
15. Evaporate the residual diethyl ether by incubating the tube at 37 ˚C for 5 minutes with the cap open.

* An aliquot of the PCR product can be analyzed by electrophoresis at this step.

* Perform a Qiagen PCR cleanup following manufacturer's protocol.

1. 2 Transformation of *Saccharomyces cerevisiae* by Liac/PEG Method

This protocol is adapted from Gietz, R.D. and R.A. Woods. *Methods in Enzymology.* v350. 2002.

1.2.1 Preparation of Competent Cells:

Day 1: Inoculate the yeast strain into 5ml YPG medium with a single yeast colony and grow overnight to saturation at 30C and 200 rpm.

Day 2:

1. Determine the OD of the yeast culture at 600nm. For many yeast strains, a suspension containing $1x10^6$ to $1x10^7$cells/ml will give an OD_{600} of 0.5-1.0.
2. Inoculate a 250ml flask containing pre-warmed 50ml YPGA with an appropriate amount of saturated culture to give OD_{600}~0.15 and grow for another 2 generations which should take about 4 hours, until the OD_{600} is around 0.4-0.6.
3. Harvest the cells by centrifugation at 3000g at room temperature for 5min, and wash the cells in 10ml of sterile water, and re-suspend in 1ml of sterile water.
4. Transfer the cell suspension to a 2ml tube, centrifuge for 3 min and discard the supernatant.
5. Re-suspend the cell pellet in 500ul buffered lithium solution, freshly prepared as follows:

   1 vol 10X TE buffer, pH 7.5

   1 vol 10X Lithium acetate stock solution

   8 vol sterile water
6. Incubate at 30C for 15min.

1.2.2 Transformation:

1. Prepare the carrier DNA by repeating the denaturation cycle (boiling and chilling) a few times immediately prior to use (denatured DNA can be used 3-5 times without denaturing again.)
2. Prepare the PEG solution freshly as follows:

   8 vol 50% PEG 4000 (filter sterilized)

   1 vol 10X TE buffer, pH 7.5

   1 vol 10X LiAc stock solution
3. Combine the following in a tube for each transformation:

   50µl competent cells
   5µl carrier DNA (10mg/ml~100ug per transformation)
   350µl PEG solution
   Plasmid DNA + Insert < 20µl
4. Mix by inversion, and incubate at 30C for 30min without shaking.
5. Heat shock exactly **20min** at 42C.
6. Centrifuge for 30s at 5000rpm, and remove the supernatant.
7. Wash in 1ml Ringer solution twice. (removing PEG completely is critical to get high transformation yield)
8. Re-suspend the pellet in 2ml YPGA, and incubate for 1-2 hours at 30^0C and 150rpm.
9. Plate 200µl on YPGA selection medium, centrifuge the remaining cells, re-suspend in 200µl YPG, and plate. Place the plates in the incubator at 30^0C for 3-5 days.

1. 3 Expression and Screening of Antimicrobial Peptides in *E.coli*

Day-1 (afternoon)

Preparation of cell dilutions

1. Wash each transformation plate with 1ml LB, making sure all colonies are lifted from the surface.
2. Transfer cells to sterile 1.6ml tube and vortex it briefly at a very low setting.
3. Pipet up-and-down few times to mix cells well.
4. Depending on the cell density, prepare serial dilutions that will give 500-2500 colonies per plate (usually between 10^7 to 10^{10} times dilutions).

Preparation of screening plates

1. In a sterile 250ml flask, mix LB (equilibrated at 45^0C) and diluted cell suspension:
   For example; to pour 5 plates (150mm diameter) → mix 125ml LB with 500ul cell suspension.
2. Mix it gently but thoroughly.
3. Pour 25ml per plate.
4. Once the first layer of agar is solidified, overlay each plate with additional 10ml LB.
5. Place all the plates at 37^0C, and incubate for 24h.

Day-2 (morning)

Preparation of media and cell culture

1. Prepare TSB media (calculate total volume by 10ml x no of plates):
2. Start *Listeria innocua* culture in 5ml TSB in a sterile 50ml conical tube.
3. Remove IPTG from -20^{0}C freezer, and place it on the bench wrapped in aluminum foil to protect it from light.

Day-2 (afternoon)

Plate Overlay

1. Prepare overlay media:
   Per 100ml TSB-soft agar (at 45^{0}C), add: 100ul IPTG (1M stock, final conc. =1mM), and 100ul *L. innocua* cells (of OD_{600}=1.0-3.0 culture to get OD_{600}=0.01-0.03, adjust volume accordingly)
2. Overlay each LB plate with 10ml TSB+IPTG+cell mixture.
3. Let the media solidify for 10min at room temperature.
4. Place all the plates at 30 ^{0}C for overnight incubation.

Day-3

- Melt/Prepare TSB with 1.2% agar, and place it at 45 ^{0}C.
- Start 5ml *L. innocua* culture.

Verification of activity

1. From each plate, pick all positive colonies and first place on LB (0.1% NaCl), then on LBA (by labeling each clone) plates.

2. Incubate at 37 ^{0}C for 4-5h
3. Overlay with 25ml melted TSB-1.2% agar (at 45^{0}C) +1mM IPTG+ OD_{600}=0.01-0.03 *L. innocua* cells.
4. Incubate O/N at 30 ^{0}C.
5. Inspect the plates for growth inhibition zones.

BIBLIOGRAPHY

1. Cadwell, R.C. and G.F. Joyce, *Mutagenic PCR.* PCR Methods Appl, 1994. **3**(6): p. S136-40.

2. Cirino, P.C., K.M. Mayer, and D. Umeno, *Generating mutant libraries using error-prone PCR.* Methods Mol Biol, 2003. **231**: p. 3-9.

3. Keefe, D.W.a.A.D., *Random Mutagenesis by PCR*, 2000, John Wiley and Sons,Inc. p. 8.3.1-8.3.9.

4. Wong, T.S., D. Zhurina, and U. Schwaneberg, *The diversity challenge in directed protein evolution.* Comb Chem High Throughput Screen, 2006. **9**(4): p. 271-88.

5. Thie, H., et al., *Affinity maturation by phage display.* Methods Mol Biol, 2009. **525**: p. 309-22, xv.

6. Scott, J.K. and G.P. Smith, *Searching for peptide ligands with an epitope library.* Science, 1990. **249**(4967): p. 386-90.

7. Steffens, D.L. and J.G. Williams, *Efficient site-directed saturation mutagenesis using degenerate oligonucleotides.* J Biomol Tech, 2007. **18**(3): p. 147-9.

8. Carlos F. Barbas, D.R.B., Jamnie K. Scott, and Gregg J. Silverman, *Phage Display - A Laboratory Manual*2001: Cold Spring Harbor Laboratory Press.

9. Reetz, M.T., et al., *Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test.* Angew Chem Int Ed Engl, 2005. **44**(27): p. 4192-6.

10. Fellouse, F.A., et al., *Molecular recognition by a binary code.* J Mol Biol, 2005. **348**(5): p. 1153-62.

11. Tominaga, T. and Y. Hatakeyama, *Determination of essential and variable residues in pediocin PA-1 by NNK scanning.* Appl Environ Microbiol, 2006. **72**(2): p. 1141-7.

12. Hulley, M.E., et al., *Focused directed evolution of pentaerythritol tetranitrate reductase by using automated anaerobic kinetic screening of site-saturated libraries.* Chembiochem, 2010. **11**(17): p. 2433-47.

13. Stemmer, W.P., *DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution.* Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10747-51.

14. Stemmer, W.P., *Rapid evolution of a protein in vitro by DNA shuffling.* Nature, 1994. **370**(6488): p. 389-91.

15. Miyazaki, K., et al., *Thermal stabilization of Bacillus subtilis family-11 xylanase by directed evolution.* J Biol Chem, 2006. **281**(15): p. 10236-42.

16. Liu, R., A.M. Enstrom, and K.S. Lam, *Combinatorial peptide library methods for immunobiology research.* Exp Hematol, 2003. **31**(1): p. 11-30.

17. Geysen, H.M., R.H. Meloen, and S.J. Barteling, *Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid.* Proc Natl Acad Sci U S A, 1984. **81**(13): p. 3998-4002.

18. Pellois, J.P., W. Wang, and X. Gao, *Peptide synthesis based on t-Boc chemistry and solution photogenerated acids.* J Comb Chem, 2000. **2**(4): p. 355-60.

19. Houghten, R.A., *General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids.* Proc Natl Acad Sci U S A, 1985. **82**(15): p. 5131-5.

20. Frank, R., *Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support.* Tetrahedron, 1992. **48**(42): p. 9217-9232.

21. Frank, R., *The SPOT-synthesis technique: Synthetic peptide arrays on membrane supports--principles and applications.* Journal of Immunological Methods, 2002. **267**(1): p. 13-26.

22. Hilpert, K., et al., *High-throughput generation of small antibacterial peptides with improved activity.* Nat Biotechnol, 2005. **23**(8): p. 1008-12.

23. Hilpert, K., D.F. Winkler, and R.E. Hancock, *Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers*

of peptides in a parallel and addressable fashion. Nat Protoc, 2007. **2**(6): p. 1333-49.

24. Zasloff, M., *Antimicrobial peptides of multicellular organisms.* Nature, 2002. **415**(6870): p. 389-95.

25. Giangaspero, A., L. Sandri, and A. Tossi, *Amphipathic alpha helical antimicrobial peptides.* Eur J Biochem, 2001. **268**(21): p. 5589-600.

26. Brown, K.L. and R.E. Hancock, *Cationic host defense (antimicrobial) peptides.* Curr Opin Immunol, 2006. **18**(1): p. 24-30.

27. Koczulla, A.R. and R. Bals, *Antimicrobial Peptides: current status and therapeutic potential.* Drugs, 2003. **63**(4): p. 389-406.

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

29. Marcos, J.F., et al., *Identification and rational design of novel antimicrobial peptides for plant protection.* Annu Rev Phytopathol, 2008. **46**: p. 273-301.

30. Yount, Y.N. and M.R. Yeaman, *Mechanisms of antimicrobial peptide action and resistance.* Pharmacological reviews, 2003. **55**: p. 27-55.

31. Brown, K. and R. Hancock, *Cationic host defense peptides.* Current Opinions in Immunology, 2006. **18**: p. 24-30.

32. Zasloff, M., *Antimicrobial peptides of multicellular organisms.* Nature, 2002. **415**: p. 389-395.

33. Pechel, A. and H.-G. Sahl, *The co-evolution of host cationic antimicrobial peptides and microbial resistance.* Nature Reviews-Microbiology, 2006. **4**: p. 529-536.

34. Gordon, Y.J., E.G. Romanowski, and A.M. McDermott, *A review of antimicrobial peptides and their therapeutic potential as anti-infective drugs.* Curr Eye Res, 2005. **30**(7): p. 505-15.

35. Gallo, R.L., et al., *Biology and clinical relevance of naturally occurring antimicrobial peptides.* J Allergy Clin Immunol, 2002. **110**(6): p. 823-31.

36. Lehmann, J., et al., *Antitumor activity of the antimicrobial peptide magainin II against bladder cancer cell lines.* Eur Urol, 2006. **50**(1): p. 141-7.

37. Reddy, K.V., R.D. Yedery, and C. Aranha, *Antimicrobial peptides: premises and promises.* Int J Antimicrob Agents, 2004. **24**(6): p. 536-47.

38. Gustafsson, C., S. Govindarajan, and J. Minshull, *Codon bias and heterologous protein expression.* Trends Biotechnol, 2004. **22**(7): p. 346-53.

39. Lithwick, G. and H. Margalit, *Hierarchy of sequence-dependent features associated with prokaryotic translation.* Genome Res, 2003. **13**(12): p. 2665-73.

40. Ikemura, T., *Codon usage and tRNA content in unicellular and multicellular organisms.* Mol Biol Evol, 1985. **2**(1): p. 13-34.

41. Ikemura, T., *Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system.* J Mol Biol, 1981. **151**(3): p. 389-409.

42. Andersson, G.E. and C.G. Kurland, *An extreme codon preference strategy: codon reassignment.* Mol Biol Evol, 1991. **8**(4): p. 530-44.

43. Kane, J.F., *Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli.* Curr Opin Biotechnol, 1995. **6**(5): p. 494-500.

44. Sharp, P.M. and W.H. Li, *The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications.* Nucleic Acids Res, 1987. **15**(3): p. 1281-95.

45. Poon, K. and R.B. Macgregor, Jr., *Formation and structural determinants of multi-stranded guanine-rich DNA complexes.* Biophys Chem, 2000. **84**(3): p. 205-16.

46. Hardin, C.C., et al., *Monovalent cation induced structural transitions in telomeric DNAs: G-DNA folding intermediates.* Biochemistry, 1991. **30**(18): p. 4460-72.

47. Gulari, E., *Method for forming molecular sequences on an open surface.*

48. Gao, X., et al., *A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids.* Nucleic Acids Res, 2001. **29**(22): p. 4744-50.

49. Gao, X., E. Gulari, and X. Zhou, *In situ synthesis of oligonucleotide microarrays.* Biopolymers, 2004. **73**(5): p. 579-96.

50. Meyerhans, A., J.P. Vartanian, and S. Wain-Hobson, *DNA recombination during PCR.* Nucleic Acids Res, 1990. **18**(7): p. 1687-91.

51. Qiu, X., et al., *Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning.* Appl Environ Microbiol, 2001. **67**(2): p. 880-7.

52. Williams, R., et al., *Amplification of complex gene libraries by emulsion PCR.* Nat Methods, 2006. **3**(7): p. 545-50.

53. Nakano, M., et al., *Single-molecule PCR using water-in-oil emulsion.* Journal of Biotechnology, 2003. **102**(2): p. 117-124.

54. Musyanovych, A., V. Mailander, and K. Landfester, *Miniemulsion droplets as single molecule nanoreactors for polymerase chain reaction.* Biomacromolecules, 2005. **6**(4): p. 1824-8.

55. Romanos, M.A., C.A. Scorer, and J.J. Clare, *Foreign gene expression in yeast: a review.* Yeast, 1992. **8**(6): p. 423-88.

56. Schoeman, H., et al., *The development of bactericidal yeast strains by expressing the Pediococcus acidilactici pediocin gene (pedA) in Saccharomyces cerevisiae.* Yeast, 1999. **15**(8): p. 647-56.

57. Van Reenen, C.A., et al., *Characterization and heterologous expression of a class IIa bacteriocin, plantaricin 423 from Lactobacillus plantarum 423, in Saccharomyces cerevisiae.* Int J Food Microbiol, 2003. **81**(1): p. 29-40.

58. Destoumieux, D., et al., *Recombinant expression and range of activity of penaeidins, antimicrobial peptides from penaeid shrimp.* Eur J Biochem, 1999. **266**(2): p. 335-46.

59. Chen, Z., et al., *Expression and secretion of a single-chain sweet protein, monellin, in Saccharomyces cerevisiae by an alpha-factor signal peptide.* Biotechnol Lett, 2011. **33**(4): p. 721-5.

60. Basanta, A., et al., *Development of bacteriocinogenic strains of Saccharomyces cerevisiae heterologously expressing and secreting the leaderless enterocin L50 peptides L50A and L50B from Enterococcus faecium L50.* Appl Environ Microbiol, 2009. **75**(8): p. 2382-92.

61. Strausberg, R.S.a.S., *Overview of Protein Expression in Sachharomyces cerevisiae*, 1995, Wiley. p. 5.6.1-5.6.7.

62. Ronicke, V.e.a., *Use of conditional promoters for expression of heterelogous proteins in Saccharomyces cerevisiae*, 1997. p. 313-321.

63. Mumberg, D., R. Muller, and M. Funk, *Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression.* Nucleic Acids Res, 1994. **22**(25): p. 5767-8.

64. Brake, A.J., et al., *Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 1984. **81**(15): p. 4642-6.

65. Cullin, C. and L. Minvielle-Sebastia, *Multipurpose vectors designed for the fast generation of N- or C-terminal epitope-tagged proteins.* Yeast, 1994. **10**(1): p. 105-12.

66. Guarente, L., R.R. Yocum, and P. Gifford, *A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site.* Proc Natl Acad Sci U S A, 1982. **79**(23): p. 7410-4.

67. Ma, H., et al., *Plasmid construction by homologous recombination in yeast.* Gene, 1987. **58**(2-3): p. 201-16.

68. Lee, K., et al., *Antibiotic activity and structural analysis of the scorpion-derived antimicrobial peptide IsCT and its analogs.* Biochem Biophys Res Commun, 2004. **323**(2): p. 712-9.

69. Rodriguez, J.M., M.I. Martinez, and J. Kok, *Pediocin PA-1, a wide-spectrum bacteriocin from lactic acid bacteria.* Crit Rev Food Sci Nutr, 2002. **42**(2): p. 91-121.

70. Gutierrez, J., et al., *Production of enterocin P, an antilisterial pediocin-like bacteriocin from Enterococcus faecium P13, in Pichia pastoris.* Antimicrob Agents Chemother, 2005. **49**(7): p. 3004-8.

71. Basanta, A., et al., *Use of the yeast Pichia pastoris as an expression host for secretion of enterocin L50, a leaderless two-peptide (L50A and L50B) bacteriocin from Enterococcus faecium L50.* Appl Environ Microbiol, 2010. **76**(10): p. 3314-24.

72. Sanchez, J., et al., *Cloning and heterologous production of Hiracin JM79, a Sec-dependent bacteriocin produced by Enterococcus hirae DCH5, in lactic acid bacteria and Pichia pastoris.* Appl Environ Microbiol, 2008. **74**(8): p. 2471-9.

73. Wang, A., et al., *High level expression and purification of bioactive human alpha-defensin 5 mature peptide in Pichia pastoris.* Appl Microbiol Biotechnol, 2009. **84**(5): p. 877-84.

74. Kim, S.J., et al., *Antibacterial activity of recombinant hCAP18/LL37 protein secreted from Pichia pastoris.* J Microbiol, 2009. **47**(3): p. 358-62.

75. Zhou, Q., M. Li, and T. Xi, *Cloning and expression of a clamworm antimicrobial peptide perinerin in Pichia pastoris.* Curr Microbiol, 2009. **58**(4): p. 384-8.

76. Koliaraki, V., et al., *Iron regulatory and bactericidal properties of human recombinant hepcidin expressed in Pichia pastoris.* Biochimie, 2008. **90**(5): p. 726-35.

77. Kang, C.J., et al., *Characterization and expression of a new subfamily member of penaeidin antimicrobial peptides (penaeidin 5) from Fenneropenaeus chinensis.* Mol Immunol, 2007. **44**(7): p. 1535-43.

78. Jin, F., et al., *Expression of recombinant hybrid peptide cecropinA(1-8)-magainin2(1-12) in Pichia pastoris: purification and characterization.* Protein Expr Purif, 2006. **50**(2): p. 147-56.

79. Burrowes, O.J., G. Diamond, and T.C. Lee, *Recombinant expression of pleurocidin cDNA using the Pichia pastoris expression system.* J Biomed Biotechnol, 2005. **2005**(4): p. 374-84.

80. Lin Cereghino, G.P., et al., *Expression of foreign genes in the yeast Pichia pastoris.* Genet Eng (N Y), 2001. **23**: p. 157-69.

81. Cereghino, J.L. and J.M. Cregg, *Heterologous protein expression in the methylotrophic yeast Pichia pastoris.* FEMS Microbiol Rev, 2000. **24**(1): p. 45-66.

82. Cregg, J.M., et al., *Expression in the yeast Pichia pastoris.* Methods Enzymol, 2009. **463**: p. 169-89.

83. Wu, J.M., et al., *Combined use of GAP and AOX1 promoter to enhance the expression of human granulocyte-macrophage colony-stimulating factor in Pichia pastoris.* Enzyme and Microbial Technology, 2003. **33**(4): p. 453-459.

84. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2003. **60**(5): p. 523-33.

85. Malakhov, M.P., et al., *SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins.* J Struct Funct Genomics, 2004. **5**(1-2): p. 75-86.

86. Butt, T.R., et al., *SUMO fusion technology for difficult-to-express proteins.* Protein Expr Purif, 2005. **43**(1): p. 1-9.

87. Marblestone, J.G., et al., *Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO.* Protein Sci, 2006. **15**(1): p. 182-9.

88. Russell, J.S.a.D.W., *Molecular Cloning: A Laboratory Manual*. 3rd ed. Vol. 1-3. 2001, NY: Cold Spring Harbor Laboratory Press.

89. Burgers, P.M. and M.B. Klein, *Selection by genetic transformation of a Saccharomyces cerevisiae mutant defective for the nuclear uracil-DNA-glycosylase.* J Bacteriol, 1986. **166**(3): p. 905-13.

90. Baneyx, F., *Recombinant protein expression in Escherichia coli.* Curr Opin Biotechnol, 1999. **10**(5): p. 411-21.

91. LaVallie, E.R., *Production of Recombinant Proteins in Escherichia coli*, in *Current Protocols in Protein Science*1995, Wiley Interscience. p. 5.1.1-5.1.8.

92. Hirota, Y., et al., *On the process of cellular division in Escherichia coli: a mutant of E. coli lacking a murein-lipoprotein.* Proc Natl Acad Sci U S A, 1977. **74**(4): p. 1417-20.

93. Suzuki, H., et al., *Murein-lipoprotein of Escherichia coli: a protein involved in the stabilization of bacterial cell envelope.* Mol Gen Genet, 1978. **167**(1): p. 1-9.

94. de Boer, H.A., L.J. Comstock, and M. Vasser, *The tac promoter: a functional hybrid derived from the trp and lac promoters.* Proc Natl Acad Sci U S A, 1983. **80**(1): p. 21-5.

95. Miller, K.W., et al., *Isolation and characterization of pediocin AcH chimeric protein mutants with altered bactericidal activity.* Appl Environ Microbiol, 1998. **64**(6): p. 1997-2005.

96. Klaenhammer, T.R., *Bacteriocins of lactic acid bacteria.* Biochimie, 1988. **70**: p. 337-349.

97. Paula Santiago-Silva, N.F.F.S., J.E. Nobrega, M.A.W. Junior, K.B.F Barbosa, A.C.P Volp, E.R.M.A. Zerdas and N.J. Wurlitzer, *Antimicrobial efficiency of film incorporated with pediocin (ALTA 2351) on preservation of sliced ham.* Food Control, 2009. **20**(1): p. 85-89.

98. Guerra, N.P., et al., *Antimicrobial activity of nisin adsorbed to surfaces commonly used in the food industry.* J Food Prot, 2005. **68**(5): p. 1012-9.

99. Drider, D., et al., *The continuing story of class IIa bacteriocins.* Microbiol Mol Biol Rev, 2006. **70**(2): p. 564-82.

100. L.M. Cintas, M.P.H., I.F. Nes, P.E.Hernandez, *Bacteriocins of lactic acid bacteria.* Food Science Tech. Int., 2001. **7**(4): p. 281-305.

101. Gonzalez, C.F. and B.S. Kunka, *Plasmid-Associated Bacteriocin Production and Sucrose Fermentation in Pediococcus acidilactici.* Appl Environ Microbiol, 1987. **53**(10): p. 2534-8.

102. Johnsen, L., G. Fimland, and J. Nissen-Meyer, *The C-terminal domain of pediocin-like antimicrobial peptides (class IIa bacteriocins) is involved in specific recognition of the C-terminal part of cognate immunity proteins and in determining the antimicrobial spectrum.* J Biol Chem, 2005. **280**(10): p. 9243-50.

103. Nissen-Meyer, J., et al., *Structure-function relationships of the non-lanthionine-containing peptide (class II) bacteriocins produced by gram-positive bacteria.* Curr Pharm Biotechnol, 2009. **10**(1): p. 19-37.

104. Eijsink, V.G., et al., *Comparative studies of class IIa bacteriocins of lactic acid bacteria.* Appl Environ Microbiol, 1998. **64**(9): p. 3275-81.

105. Miller, K.W., et al., *Production of active chimeric pediocin AcH in Escherichia coli in the absence of processing and secretion genes from the Pediococcus pap operon.* Appl Environ Microbiol, 1998. **64**(1): p. 14-20.

106. Goldman, M.J., et al., *Human beta-defensin-1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis.* Cell, 1997. **88**(4): p. 553-60.

107. Lee, I.H., Y. Cho, and R.I. Lehrer, *Effects of pH and salinity on the antimicrobial properties of clavanins.* Infect Immun, 1997. **65**(7): p. 2898-903.

108. Sant'Anna, V., et al., *Influence of pH and sodium chloride on kinetics of thermal inactivation of the bacteriocin-like substance P34.* J Appl Microbiol, 2010. **110**(1): p. 156-62.

109. Todorov, S.D. and L.M. Dicks, *Parameters affecting the adsorption of plantaricin 423, a bacteriocin produced by Lactobacillus plantarum 423 isolated from sorghum beer.* Biotechnol J, 2006. **1**(4): p. 405-9.

110. F. Ben Hammou, S.N.S., M. Idaomar and J. Abrini, *Combinations of nisin with salt to control Listeria monocytogenes on sheep natural sausage casings stored at 6^0C.* African Journal of Biotechnology, 2010. **9**(8): p. 1190-1195.

111. Linda J. Harris, H.P.F., Todd R. Klaenhammer, *Sensitivity and Resistance of Listeria monocytogenes ATCC 19115, Scott A, and UAL500 to Nisin.* Journal of Food Protection, 1991. **54**(November).

112. Ananou, S., et al., *Effect of combined physico-chemical preservatives on enterocin AS-48 activity against the enterotoxigenic Staphylococcus aureus CECT 976 strain.* J Appl Microbiol, 2004. **97**(1): p. 48-56.

113. Tominaga, T. and Y. Hatakeyama, *Development of innovative pediocin PA-1 by DNA shuffling among class IIa bacteriocins.* Appl Environ Microbiol, 2007. **73**(16): p. 5292-9.

114. Russell, M.J.B.a.R.B., *Amino Acid Properties and Consequences of Substitutions*, in *Bioinformatics for Genetics*2003, Wiley. p. 289-314.

115. van Reenen, C.A., L.M. Dicks, and M.L. Chikindas, *Isolation, purification and partial characterization of plantaricin 423, a bacteriocin produced by Lactobacillus plantarum.* J Appl Microbiol, 1998. **84**(6): p. 1131-7.

116. Zidoveztki, D.A.a.R. *Helical Wheel Projections*. 2001; Available from: http://rzlab.ucr.edu/scripts/wheel/wheel.cgi.

117. Smith, J.S. and J.M. Scholtz, *Energetics of polar side-chain interactions in helical peptides: salt effects on ion pairs and hydrogen bonds.* Biochemistry, 1998. **37**(1): p. 33-40.

118. Jydegaard, A.M., A. Gravesen, and S. Knochel, *Growth condition-related response of Listeria monocytogenes 412 to bacteriocin inactivation.* Lett Appl Microbiol, 2000. **31**(1): p. 68-72.

119. Zhang, L., et al., *Antimicrobial peptide therapeutics for cystic fibrosis.* Antimicrob Agents Chemother, 2005. **49**(7): p. 2921-7.

120. Hoskin, D.W. and A. Ramamoorthy, *Studies on anticancer activities of antimicrobial peptides.* Biochim Biophys Acta, 2008. **1778**(2): p. 357-75.

121. Hozumi, K., et al., *Chain-specific heparin-binding sequences in the laminin alpha chain LG45 modules.* Biochemistry, 2009. **48**(23): p. 5375-81.

122. Hozumi, K., et al., *Cell adhesive peptide screening of the mouse laminin alpha1 chain G domain.* Arch Biochem Biophys, 2010. **503**(2): p. 213-22.

123. Paschke, M., *Phage display systems and their applications.* Appl Microbiol Biotechnol, 2006. **70**(1): p. 2-11.

124. Smith, G.P. and V.A. Petrenko, *Phage Display.* Chem Rev, 1997. **97**(2): p. 391-410.