A MOLECULAR GENETIC APPROACH TO STABILIZING BIOACTIVE PEPTIDES VIA PROTEIN-BASED MOTIFS

by

JAMES W. WARREN, III

(Under the Direction of Timothy Hoover)

ABSTRACT

Instability of bioactive peptides represents a major challenge to the development of these molecules as drugs. The purpose of this study was the investigation of protein-based motifs and their potential for protecting peptides from enzymatic degradation. A novel, highly regulable expression vector, pLAC11, was constructed so that potent inhibitory peptides could be isolated from an *in vivo* genetic screen designed to generate randomized bioactive peptides that inhibit the growth of bacteria. Compared to other commonly utilized expression vectors that are known not to be tightly regulable, pLAC11 was demonstrated to possess the ability to regulate potent inhibitor peptides which could prove lethal to the cells generating them. The three protein-based motifs that were examined by this screen were the fusion of randomized peptides to the highly stable Rop protein, the generation of α-helical peptides via the random incorporation of a set of hydrophilic helix-forming amino acids, and the use of oppositely charged residues at the termini of randomized peptides. All three methods were observed to yield an increase in the frequency at which potent inhibitors were isolated during the *in vivo* screen as compared to unprotected peptides, presumably due to greater stability. To further investigate the possible role of secondary structure in peptide stability, a group of α-helical peptides based on Rop and other

helical proteins as well as putative helical peptides designed *de novo* were examined by secondary structure prediction algorithms, CD spectroscopy, and *in vitro* plasma degradation assays. Two highly stable helical peptides were identified and found to have half-lives much greater than other representative peptides and similar to that of small stable proteins.

INDEX WORDS: Peptide, Stability, Degradation, Plasmid, Helical, Rop

A MOLECULAR GENETIC APPROACH TO STABILIZING BIOACTIVE PEPTIDES VIA PROTEIN-BASED MOTIFS

by

JAMES W. WARREN, III

B.S., University of South Carolina, 1993

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

A MOLECULAR GENETIC APPROACH TO STABILIZING BIOACTIVE PEPTIDES VIA PROTEIN-BASED MOTIFS

by

JAMES W. WARREN, III

Approved:

Major Professor: Timothy Hoover

Committee: Elliot Altman
Mark Eiteman
Juergen Wiegel

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2004

DEDICATION

This work is dedicated to my parents, Jim and Nancy, and my sister, Joy, with gratitude and appreciation for their unending love and support which helped me see a long journey through to the end. I could not have done it without them.

Human subtlety will never devise an invention more beautiful, more simple, or more direct than does Nature, because in her inventions, nothing is lacking and nothing is superfluous...

-Leonardo da Vinci

ACKNOWLEDGEMENTS

I would like to thank my good friends for helping me live life between all the experiments over the years; to Neal and John for late-night phone calls and pep talks, and to Winston and Warren for walking the same path and knowing what it means. Thanks to Ron Seidel for assistance with the CD studies in Chapters 4 and 5. I would like to offer my sincere appreciation to the members of my doctoral committee, Dr. Mark Eiteman and Dr. Tim Hoover, for their insight and advice on my research. Finally, I sincerely thank Dr. Elliot Altman for his guidance and mentorship throughout the ups and downs of a long graduate career, and for teaching me what it is to be a scientist.

If we knew what it was we were doing, it would not be called research, would it?

-Albert Einstein

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

Peptides are sequence- and length-specific oligomers composed of between 2-50 amino acids. These important biomolecules are ubiquitous in living systems where they assume a myriad of roles. They are signal molecules that are involved in the control of critical biological functions such as cell division, mating, chemotaxis, pain perception, growth and immunity. They are transport molecules that facilitate the passage of ions through cell membranes. They are protective agents that function as antibiotics which protect eukaryotic organisms as diverse as mold and man from their common enemy, the bacterium. They are important digestive intermediates and play a crucial role in nutrition of both cells and intact organisms. Peptides have become an increasingly important class of molecules in biochemistry, medicinal chemistry, and physiology. Many naturally occurring bioactive peptides function as hormones, neurotransmitters, cytokines, and growth factors. Peptide substrates of proteases, kinases, phosphatases, and aminoacyl or glycosyl transferases are used to study enzyme kinetics, mechanism of action, biochemical and physiological roles and to aid in the isolation of enzymes and the design of inhibitors. Peptides are also used as synthetic antigens for the preparation of polyclonal or monoclonal antibodies targeted to specific sequences. Epitope mapping with synthetic peptides can be used to identify specific antigenic peptides for the preparation of synthetic vaccines, to determine protein sequence regions that are important for biological activity, and to design small molecule peptide mimetics of protein structure or function.

Application of bioactive peptides

It is clear that at the beginning of the 21st century peptide therapeutics have enormous potential in such diverse areas as growth control, blood pressure management, neurotransmission, hormone action, pain management, digestion, reproduction, antimicrobial, and so forth. Human disease states for which peptide-based drug therapies currently exist include diabetes (insulin and glucagon), osteoporosis (calcitonin), prostate cancer and endometriosis (gonadotropin-releasing hormone), acromegaly and ulcers (somatostatin), and hypothyroidism (thyrotropin-releasing hormone) (Lien and Lowman, 2003). Through millions of years of evolution nature has discovered that it can control nearly all biological processes by various types of molecular recognition, and that proteins and peptides are ideally suited for this role due to their almost limitless potential for diversity and their unique biochemical and biophysical properties. This can be readily understood if one considers that given only the 20 normal amino acids, the number of unique chemical forms for a hexapeptide is 64,000,000 (20^6), for a heptapeptide 1,280,000,000 (20^7), and so on. From this perspective, perhaps it is not so unexpected that nature has discovered that peptides and proteins can indeed do it all, from providing structure and motion, to catalysis, to information transduction, to growth and maturation, and more. The immune system's ability to recognize literally millions of foreign materials, both natural and man-made, and to neutralize them as part of its survival strategy is just one example illustrating the vast potential inherent to peptide-based therapeutics.

Over the past decade, levels of bacterial resistance to antibiotics have risen dramatically and "superbugs" resistant to most or all available agents have appeared in hospitals worldwide. Antimicrobial peptides represent ancient host defense effector molecules present in all species of life, ranging from plants and insects to animals, including mollusks, crustaceans, amphibians,

birds, fish, and mammals, including humans (Hancock and Lehrer, 1998; Hancock and Chapple, 1999). Fundamental differences exist between prokaryotic and eukaryotic cells that may represent targets for antimicrobial peptides (Hancock, 2001). These significant distinctions include membrane composition and architecture, energetics such as transmembrane potential and polarization, and structural features including sterols, lipopolysaccharide and peptidoglycan (Devine and Hancock, 2002). In general, although more than 500 antimicrobial peptides have now been characterized as having widely diverse sequences, these molecules have been classified into relatively few conformational structures. Most are cationic amphipathic molecules due mainly to the presence of arginine and lysine residues and can be grouped into four or five different structural categories (Hwang and Vogel, 1998; Epand and Vogel, 1999). These include: (a) amphipathic α-helical peptides without cystine, such as magainins, cecropins and buforin (Zasloff, 1987; Gudmundsson *et al*., 1991; Yi *et al*., 1996); (b) cystine-rich, β-sheet peptides with internal disulfide bonding, such as α- and β-defensins, protegrins and tachyplesins (Lehrer *et al*., 1993; Kokryakov *et al*., 1993; Kawano *et al*., 1990); (c) cystine–disulfide loop peptides with or without amphiphilic tails, such as brevinins and thanatin (Morikawa *et al*., 1992; Mandard *et al*., 1998); and (d) extended linear peptides with one or two predominant amino acids, such as histatin, indolicidin and tritrpticin (Brewer *et al*., 1998; Selsted *et al*., 1992; Lawyer *et al*., 1996). The primary target of action for the majority of antimicrobial peptides has been assumed to be the bacterial cytoplasmic membrane. They all have affinity for membrane lipids and their specificity for microbial membranes has been demonstrated in many cases to be related to the positive charge on the peptide interacting with the exposed anionic lipids of microorganisms. However, the exact mechanism by which these peptides exert their bacteriocidal effect is not well understood. The amphipathic α-helical peptides may form pores

in the membrane allowing for leakage of ions and other materials from the cell (Powers and Hancock, 2003). Alternatively, the β-sheet peptides may disrupt membranes by a "carpet-like" mechanism, in which the aggregation and association of the peptides with lipid head groups leads to an overall destabilization of the membrane, as peptides align across the surface of the negatively charged membrane and once a high enough local concentration (a "carpet") is achieved, an electrostatically-driven collapse of the cytoplasmic membrane occurs (Shai, 1995; Wu *et al*., 1999). Regardless of which model is correct, the net result of membrane disruption would be the rapid depolarization of the bacterial cell leading to rapid cell death, with total killing occurring within 5 minutes for the most potent peptides (Friedrich *et al*., 1999). In addition, alternative mechanisms have been proposed to explain the cytotoxic action of these peptides, including their stimulation of autolytic enzymes, interference with DNA and/or protein synthesis (Subbalakshmi and Sitaram, 1998; Patrzykat *et al*., 2002), or their binding to DNA (Park *et al*., 1998).

Synthesis of peptides

In 1902, Emil Fischer published the first paper in which the term "peptide" was used (Fischer, 1902). In the century since that initiation, the field of scientific endeavor has grown enormously. Fischer himself gave birth to the study of peptide chemistry a year earlier when he prepared the first unprotected synthetic peptide, glycylglycine (Fischer and Fourneau, 1901). In fact, his initial approaches highlighted strategies that remain in place to this day. The Nobel laureate's efforts culminated in the preparation of an octadecapeptide, Leu-Gly_3-Leu-Gly_3-Leu-Gly_9 (Fischer, 1907). Even though it consisted of only two different amino acids, this was an important accomplishment, as peptides of a given sequence with similar lengths were not synthesized again until the 1950s. The nagging problem of chain elongation at the amino

terminus was successfully addressed by the development of the easily removable benzyloxycarbonyl (Boc) protecting group (Bergmann and Zervas, 1932). This led to a new era of peptide chemistry by opening the way to the incorporation of multifunctional amino acids in synthesis. As a result of this discovery, the synthesis of small naturally occurring peptides such as carnosine (Sifferd and du Vigneaud, 1935) and glutathione (Harrington and Mead, 1935) were soon achieved. During the same period that Watson and Crick were elucidating the structure of the DNA double helix, Frederick Sanger developed a method for the determination of the primary sequence of polypeptides and was awarded the Nobel Prize for using this technique to sequence insulin (Sanger, 1952; Watson and Crick, 1953). This new knowledge created the driving force for peptide chemists to develop methods to synthesize peptides more efficiently for comparison with natural products and as models to gain a better understanding of biochemical and biophysical properties. In 1953, a milestone was achieved with the structural determination and synthesis of an active peptide hormone, oxytocin (du Vigneaud *et al*., 1953). This began a new era in which bioactive peptides were synthesized for use in structure-activity studies, and these target molecules increased the need for improved synthetic methods. The months of work that were needed to accomplish the synthesis of oxytocin achieved a yield of less than 1%, but nonetheless earned du Vigneaud the Nobel Prize. In 1963, Bruce Merrifield published a landmark paper describing the development of solid-phase peptide synthesis (SPPS), in which peptides attached to copolymer resins are chemically lengthened from the carboxyl terminus by the addition of one protected one amino acid residue at a time (Merrifield, 1963). This technique was responsible more than anything else for opening the way to widespread use of synthetic peptides as reagents in chemical and biomedical investigations, and Merrifield was awarded the Nobel Prize for this revolutionary work. Indeed, shortly after the introduction of this procedure,

it was used to synthesize insulin successfully (Marglin and Merrifield, 1966). The use of high performance liquid chromatography in the 1970s provided peptide researchers with an important and powerful tool which was sensitive enough to distinguish among the impurities that were expected from SPPS, and allowed mixtures to be purified to homogeneity. In the late 1970s, the search for a stable protecting group which could be removed by base rather than the strong acids (hydrofluoric and trifluoromethanesulfonic) that were predominantly used at the time, led to the discovery of the 9-fluorenylmethoxycarbonyl (Fmoc) group which was rapidly cleaved under basic conditions and readily incorporated into solid phase synthetic strategies (Meienhofer, 1979). By the beginning of the 1980s, the impact of molecular biology was beginning to be felt by the field of peptide chemistry, as reports of the efficient synthesis of peptides and small proteins were appearing. It is rather ironic that during most of the 20th century peptide researchers exerted tremendous effort to isolate pure, bioactive peptides from synthesis mixtures, while beginning in the 1990s many scientists developed what are now termed combinatorial methods to create peptide libraries containing hundreds to billions of molecules. One of the first approaches, the multipin method, involves the rapid parallel solid phase synthesis of microgram quantities of hundreds of peptides bound to the tips of polypropylene pins, and was quickly adapted to 96-well plates (Geysen *et al*., 1984). Biological screening was performed by an enzyme-linked immunosorbant assay (ELISA) on the pins. In the "tea bag" method, the solid support consists of resin beads separated into polypropylene bags which are then immersed in a reaction mixture containing a single or multiple activated amino acids, ultimately yielding 10-20 milligrams of each peptide (Houghten, 1985; Houghten *et al*., 1991). Another version of multiple solid-phase peptide synthesis is the split-couple-mix method in which the resin-bound peptide is divided into equal fractions before each new coupling reaction, each fraction is

coupled with a different amino acid, the fractions are then mixed together and the cycle is repeated (Furka *et al*., 1991). This method was further refined into the "one-bead, one-peptide" which can be utilized for the generation of very large random libraries of 10^5-10^6 peptides (Lam *et al*., 1991). A combinatorial approach which utilized recombinant DNA technology involved the use of filamentous phage to display a random oligopeptide at the amino terminus of the viral coat protein, pIII (Parmley and Smith, 1989; Scott and Smith, 1990). This "phage-display" method was accomplished by the insertion of a sequence of randomized DNA into the pIII gene of the fd phage (Cwirla *et al*., 1990). The major advantage of this molecular biological approach is that the size of the molecules is not limited by the constraints of synthetic peptide chemistry. Furthermore, the biological approach enables researchers to take advantage of known protein secondary structures, such as the immunoglobulin fold or the Rop helix, by fusing random oligopeptides to such structures.

As more and more bioactive peptides began to be targeted as lead compounds, it became apparent that the serum stability, membrane permeability, and bioavailability of the final molecule were of primary importance for drug design. Natural bioactive peptides began to be extensively modified to take these factors into account. It was also during the latter half of the 20^{th} century that impressive advances in understanding the three-dimensional structure of peptides, due largely to the development of X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, were achieved. In addition to the α-helical and β-sheet structures discovered by Pauling and Corey, peptides were shown to assume specific turn structures as well as a variety of other helices (Pauling *et al*., 1951; Pauling and Corey 1951). As a result, the concept that the most potent drugs mimic the three-dimensional structure of the native bioactive peptide bound to its receptor, became an important driving force in the field of

peptide chemistry. Stabilization of these structures would result in increased potency and more effective pharmaceuticals.

Stability and protection of peptides

As already discussed, peptides play important roles in many diverse biochemical processes. With the convergence of molecular biology and biotechnology, many of these bioactive molecules can be produced today on a mass scale. Most of these peptides have therapeutic potential, and considerable interest exists in designing systems to achieve effective delivery of these molecules to various *in vivo* targets. One of the major obstacles to this goal is the large number of endogenous peptidases and proteases that are present in most cells, tissues, organs, and circulating fluids which can rapidly degrade these bioactive molecules. Peptidases and proteases play essential roles in protein activation, cell regulation and signaling, as well as in the generation of amino acids for protein synthesis or utilization in other metabolic pathways. Peptidases can cleave either specific amide bonds (limited proteolysis), depending on the amino acid sequence of a polypeptide, or completely degrade a polypeptide to its constituent amino acids (unlimited proteolysis). Although peptide degradation by peptidases is important for producing bioactive molecules and for inactivating them after their function has been performed, this highly efficient metabolic process represents a major challenge to the therapeutic use of peptides.

Regardless of their source organism, peptidases can be categorized into two main groups, exopeptidases and endopeptidases, and these groups can be further subdivided into additional classes based on their substrate specificity and catalytic mechanism. Exopeptidases that hydrolyze peptides from their amino terminus are termed aminopeptidases, an example of which is aminopeptidase N which removes the amino terminal amino acid from dipeptides, tripeptides

and a variety of oligopeptides (McDonald and Barrett, 1986). Exopeptidases that hydrolyze peptides from their carboxyl terminus are called carboxypeptidases, an example of which is carboxypeptidase P which removes the carboxyl terminal amino acid from peptides where this residue is proline, alanine, or glycine (Erickson *et al*., 1989). Endopeptidases are defined as those that hydrolyze the interior peptide bonds of peptides, and endopeptidase-24.11 (Endo-1) which was identified by its ability to hydrolyze the insulin β chain is one example (Kerr and Kenny, 1974). Beyond these basic categories, peptidases and proteases are classified according to their catalytic mechanism into four groups: serine proteinases, cysteine proteinases, aspartic proteinases, and metallo proteinases (Rawlings and Barrett, 1993).

The class of serine proteinases comprises two distinct families. The chymotrypsin family which includes the mammalian enzymes such as chymotrypsin, trypsin, and elastase, and the substilisin family which includes bacterial enzymes such as subtilisin. The cysteine proteinase family includes the plant proteases papain, actinidin and bromelain, several mammalian lysosomal cathepsins, the cytosolic calpains, as well as several parasitic proteases. Most of the aspartic proteinases belong to the pepsin family. The pepsin family includes the digestive enzymes pepsin and chymosin, lysosomal cathepsins D, processing enzymes such as renin, and certain fungal proteases including penicillopepsin, rhizopuspepsin and endothiapepsin. A second family of aspartic proteinases is comprised of viral proteinases such as the protease from HIV-1, which is also called retropepsin. The metallo proteinases may be one of the oldest classes of proteinases and are found in bacteria and fungi, as well as in higher organisms. They differ widely in their sequences and their structures but the great majority of enzymes contain a zinc atom as part of the catalytic site. In some cases, zinc may be replaced by other metals such as cobalt or nickel without loss of the activity.

For most naturally occurring bioactive peptides, the serum half-life and biological activity is often limited by the rapid proteolytic hydrolysis of susceptible amide bonds, which result in typical half-life values for peptides in the seconds-to-minutes time frame (Table 1.1). To make these peptides physiologically or pharmacologically useful it is necessary to increase their serum half-life by increasing their resistance to enzymatic degradation. Protection from proteolysis, therefore, has become the predominant focus for research directed at the design of peptide drugs. The use of amide bond replacements has been successful on occasion, but the products frequently suffer from decreased potency due to the loss of critical binding interactions or altered ligand conformation (Fincham *et al*., 1992). Multiple replacement strategies have been used to avoid the susceptibility of the peptide amide bond to cleavage by peptidases and proteases. For example, D-amino acid substitution, N-terminal acetylation, C-terminal amidation, cyclization, amide bond reduction, and unnatural amino acid substitution have all shown promise in increasing peptide half-lives, as in the case of amino-terminally substituted analogs of glucagon-like peptide 1 (GLP-1). GLP-1 is a 29-amino acid peptide with great potential as a therapeutic agent in the treatment of diabetes, but its 2-minute half-life due primarily to rapid degradation by dipeptidyl peptidase IV severely limits its usefulness. Following a seemingly inconsequential substitution at position 8 with glycine, serine, or threonine, GLP-1 analogs were found to have plasma half-lives of 159, 174 and 197 minutes, respectively (Deacon *et al*., 1998). Five-amino acid β-sheet breaker peptides have shown promise in the inhibition of amyloid plaque deposition, which characterizes Alzheimer's disease, but are hindered by 5-minute half-lives in human plasma. However, end-protected derivatives that incorporate both acetylation of the amino terminus and amidation of the carboxyl terminus exhibit *in vivo* half-lives of 37 minutes (Permanne *et al*., 2002). Likewise, the *in vivo* half-life of

a somatostatin analogue increased from 3 minutes for the unmodified peptide to over 400 minutes following N-terminal acetylation (Benuck and Marks, 1976). Although terminal protection has been widely used to improve peptide stability it rarely prevents degradation completely, and often the increased stability is accompanied by a reduction in bioactivity. For example, when the pentapeptide analogue of the 49-amino acid hormone thymopoietin, thymopentin, is modified by either amino-terminal acetylation or carboxyl-terminal amidation, a complete loss of biological activity is observed (Heavner *et al*., 1986).

Another classical approach to circumventing enzymatic degradation and increasing plasma half-life of bioactive peptides has been the covalent attachment of polyethylene glycol (PEG) to polypeptide amino groups (Abuchowski *et al*., 1977; Chen *et al*., 1981; Davis *et al*., 1981). Polyethylene glycol conjugation (PEGylation) to biomolecules has been a long exploited strategy to improve the properties of many drugs. It was in fact often found to improve physical and chemical stability, to increase aqueous solubility, to provide protection against enzymatic degradation by steric hindrance, to prolong *in vivo* half-life, and to decrease clearance of conjugated drugs when compared to the unmodified parent molecules. Reduced antigenicity, immunogenicity and toxicity are other positive properties often observed with PEGylated molecules. PEGylated calcitonin has demonstrated strong resistance against enzymatic degradation in rat nasal mucosa, with a 56-fold increase in half-life compared to unmodified peptide (Na *et al*., 2004). However, modifications such as these that reduce susceptibility to proteolytic cleavage can have a profound effect on the *in vivo* potency of bioactive peptides. In some cases *in vitro* bioactivity is reduced but is nevertheless accompanied by an increase in the biological activity *in vivo*, such as in the case of human growth hormone-releasing factor (hGRF). hGRF is a 29-amino acid peptide involved in the regulation of the expression and

release of growth hormone from the anterior pituitary gland that has been clinically used for the therapeutic treatment of growth hormone deficiency in children as an alternative treatment to growth hormone therapy (Low, 1993). However, one of the major limitations in hGRF clinical use is related to its 10-20 minute *in vivo* half-life, due mainly to rapid cleavage by dipeptidyl peptidase IV (Frohman *et al*., 1986). PEGylated analogs of hGRF were demonstrated to have *in vitro* plasma half-lives of from 4-10 hours, versus unmodified hGRF peptide which had a half-life of ~30 minutes, and were also found to be highly bioactive, with a potency range 12 to 55-fold higher than that of native hGRF (Campbell *et al*., 1997).

Transport of peptides and drug delivery

The concept of transmembrane transport of peptides has been implicit in cell physiology for nearly a century, and its history has been the subject of extensive reviews (Matthews, 1977; Matthews, 1987). Peptide transport has been recognized not only in mammalian species, but also in fungi (Naider and Becker, 1987), plants (Walker-Smith and Payne, 1984), and bacteria (Payne and Smith, 1994). Efficient delivery of therapeutic and diagnostic agents across the plasma membrane of cells is crucial in developing novel therapies. Only compounds within a narrow range of molecular size, net charge, and polarity are able to diffuse effectively through the lipid bilayers of cell membranes and thus reach their pharmacological targets (Lipinski *et al*., 2001). By their very nature, infective organisms such as bacteria and viruses possess proteins endowed with properties to penetrate living cells indiscriminately (Pugsley, 1996). For example, the HBV VP22 protein and pore-forming toxins such as tetanus toxin, anthrax toxin, cytolysins and diphtheria toxin can gain entry into cells in essentially receptor-independent fashion (Boquet and Duflot, 1982; Stenmark *et al*., 1991; Ballard *et al*., 1996; Rossjohn *et al*., 1997). Membrane permeating peptides derived from such proteins, in particular HIV-1 tat peptide and its

derivatives, have enabled the intracellular delivery of molecules of various sizes and physicochemical properties (Frankel and Pabo, 1988). HIV-tat peptide and its derivatives have recently received much attention, primarily because of their efficiency of transport and short sequence. The 12-amino acid Tat peptide was originally derived from the HIV-1 tat protein, an 86-amino acid transactivation protein involved in the replication of HIV-1 virus (Fawell *et al*., 1994).

Recent studies have confirmed the ability of the insect-derived antimicrobial peptide pyrrhocoricin to enter various bacterial, fungal and mammalian cell types, suggesting that the proline-rich, cationic peptide might be used as a drug-delivery vehicle (Kragol *et al*., 2002). Researchers are also exploring the ability of pyrrhocoricin to deliver antigen to immune cells in an effort to explore whether the peptide could be used in a novel vaccine approach. Both pyrrhocoricin in its native state and synthetic analogs have demonstrated the capability of entering and killing a number of strains of bacteria by binding to the 70-kD heat shock protein DnaK and inhibiting protein folding, including antibiotic-resistant strains of Gram-type negative bacteria such as *Escherichia coli* and *Salmonella enterica* serovar Typhimurium (Cudic *et al*., 2002). Initially interested in pyrrhocoricin and its derivatives because of their antimicrobial properties, researchers are now exploring whether these peptides could be employed as delivery vehicles for other peptide or peptide-based drugs that would not normally be able to penetrate mammalian cell membranes and tissues.

Research has also focused on new delivery methods to complement the development of novel bioactive peptides. Some of these include transdermal patches, transdermal iontophoresis, aerosol inhalation and oral formulations (Henry *et al*., 1998; Suzuki *et al*., 2002; Bot *et al*., 2001; Still, 2002)

Purpose of study

Bioactive peptides such as calcitonin, glucagon, and insulin are widely utilized in the field of medicine and show great promise for the design of new therapeutic agents. Undesirable pharmaceutical and biopharmaceutical properties, which include low aqueous solubility, poor stability, and low permeability through biological membrane barriers, often hinder the clinical development of potential drug candidates. This is particularly true with peptide and peptidomimetic compounds. Finding solutions to these problems is a very contemporary and complex issue in the clinical development of the potentially vast number of biologically active peptides and peptidomimetics as potential new drugs. Based on what has been learned about peptide function there is also increasing interest in using novel synthetic peptides developed through combinatorial synthesis or phage display to generate new therapeutic agents. However, while often promising during *in vitro* trials, many lead compounds are usually inherently unstable and rapidly degraded by endogenous proteases and peptidases upon introduction to their target organisms, with *in vivo* half-lives typically on the order of minutes.

The purpose of this thesis was to study the problem of peptide degradation and investigate alternative methods by which bioactive peptides might be stabilized. A molecular genetic approach was devised in *E. coli* to study the problem of peptide instability, and relied on the use of a novel highly regulable expression vector. Using this *in vivo* approach, protein-based motifs which could be used to stabilize peptides were identified and characterized. Because peptidases and proteases are highly conserved across species what has been learned from these bacterial studies is directly relevant to therapeutic applications in humans.

References

Abuchowski, A., McCoy, J.R., Palczuk, N.C., van Es, T., and Davis, F.F. (1977). Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase. J Biol Chem 252, 3582-3586.

Ardaillou, R., Sizonenko, P., Meyrier, A., Vallee, G., and Beaugas, C. (1970). Metabolic clearance rate of radioiodinated human calcitonin in man. J Clin Invest 49, 2345-2352.

Aronin, N., Carraway, R.E., Ferris, C.F., Hammer, R.A., and Leeman, S.E. (1982). The stability and metabolism of intravenously administered neurotensin in the rat. Peptides 3, 637-642.

Ballard, J.D., Collier, R.J., and Starnbach, M.N. (1996). Anthrax toxin-mediated delivery of a cytotoxic T-cell epitope *in vivo*. Proc Natl Acad Sci USA 93, 12531-12534.

Benuck, M. and Marks, N. (1976). Differences in the degradation of hypothalamic releasing factors by rat and human serum. Life Sci 19, 1271-1276.

Bergmann, M. and Zervas, L. (1932). Ber Dtsch Chem Ges 65, 1192-1201.

Boarder, M.R. and McArdle, W. (1986). Breakdown of small enkephalin derivatives and adrenal peptide E by human plasma. Biochem Pharmacol 35, 1043-1047.

Boquet, P. and Duflot, E. (1982). Tetanus toxin fragment forms channels in lipid vesicles at low pH. Proc Natl Acad Sci USA 79, 7614-7618.

Bot, A.I., Smith, D.J., Bot, S., Dellamary, L., Tarara, T.E., Harders, S., Phillips, W., Weers, J.G., and Woods, C.M. (2001). Receptor-mediated targeting of spray-dried lipid particles coformulated with immunoglobulin and loaded with a prototype vaccine. Pharm Res 18, 971-979.

Brewer, D., Hunter, H., and Lajoie, G. (1998). NMR studies of the antimicrobial salivary peptides histatin 3 and histatin 5 in aqueous and nonaqueous solutions. Biochem Cell Biol 76, 247-256.

Campbell, R.M., Heimer, E.P., Ahmad, M., Eisenbeis, H.G., Lambros, T.J., Lee, Y., Miller, R.W., Stricker, P.R., and Felix, A.M. (1997). PEGylated peptides. V. Carboxy-terminal PEGylated analogs of growth hormone-releasing factor (GRF) display enhanced duration of biological activity *in vivo*. J Pept Res 49, 527-537.

Chen, R.H., Abuchowski, A., Van Es, T., Palczuk, N.C., and Davis, F.F. (1981). Properties of two urate oxidases modified by the covalent attachment of poly(ethylene glycol). Biochim Biophys Acta 660, 293-296.

Cudic, M., Condie, B.A., Weiner, D.J., Lysenko, E.S., Xiang, Z.Q., Insug, O., Bulet, P., and Otvos, L. Jr. (2002). Development of novel antibacterial peptides that kill resistant isolates. Peptide 23, 2071-2083.

Cwirla, S.E., Peters, E.A., Barrett, R.W., and Dower, W.J. (1990). Peptides on phage: a vast library of peptides for identifying ligands. Proc Natl Acad Sci USA 87, 6378-6382.

Cyr, M., Lepage, Y., Blais, C. Jr., Gervais, N., Cugno, M., Rouleau, J.L., and Adam, A. (2001). Bradykinin and des-Arg(9)-bradykinin metabolic pathways and kinetics of activation of human plasma. Am J Physiol Heart Circ Physiol 281, H275-283.

Danser, A.H., Koning, M.M., Admiraal, P.J., Derkx, F.H., Verdouw, P.D., and Schalekamp, M.A. (1992). Metabolism of angiotensin I by different tissues in the intact animal. Am J Physiol 263, H418-428.

Davis, S., Abuchowski, A., Park, Y.K., and Davis, F.F. (1981). Alteration of the circulating life and antigenic properties of bovine adenosine deaminase in mice by attachment of polyethylene glycol. Clin Exp Immunol 46, 649-653.

Deacon, C.F., Nauck, M.A., Toft-Nielsen, M., Pridal, L., Willms, B., and Holst, J.J. (1995). Both subcutaneously and intravenously administered glucagon-like peptide I are rapidly degraded from the NH_2-terminus in type II diabetic patients and in healthy subjects. Diabetes 44, 1126-1131.

Deacon, C.F., Knudsen, L.B., Madsen, K., Wiberg, F.C., Jacobsen, O., and Holst, J.J. (1998). Dipeptidyl peptidase IV resistant analogues of glucagon-like peptide-1 which have extended metabolic stability and improved biological activity. Diabetologia 41, 271-278.

Devine, D.A. and Hancock, R.E.W. (2002). Cationic peptides: distribution and mechanisms of resistance. Curr Pharm Des, 8, 703-714.

Domschke, S., Domschke, W., Bloom, S.R., Mitznegg, P., Mitchell, S.J., Lux, G., and Strunz, U. (1978). Vasoactive intestinal peptide in man: pharmacokinetics, metabolic and circulatory effects. Gut 19, 1049-1053.

du Vigneaud, V., Ressler, C., and Trippet, S. (1953). The sequence of amino acids in oxytocin, with a proposal for the structure of oxytocin. J Biol Chem 205, 949-957.

du Vigneaud, V., Ressler, C., Swan, J.M., Roberts, C.W., Katsoyannis, P.G., and Gordon, S. (1953). The synthesis of an octapeptide amide with the hormone activity of oxytocin. J Am Chem Soc 75, 4879-4880.

Epand, R.M. and Vogel, H.J. (1999). Diversity of antimicrobial peptides and their mechanisms of action. Biochim Biophys Acta 1462, 11-28.

Erickson, R.H., Song, I.S., Yoshioka, M., Gulli, R., Miura, S., and Kim, Y.S. (1989). Identification of proline-specific carboxypeptidase localized to brush border membrane of rat small intestine and its possible role in protein digestion. Dig Dis Sci 34, 400-406.

Fawell, S., Seery, J., Daikh, Y., Moore, C., Chen, L.L, Pepinsky, B., and Barsoum, J. (1994). Tat-mediated delivery of heterologous proteins into cells. Proc Natl Acad Sci USA 91, 664–668.

Fincham, C.I., Higginbottom, M., Hill, D.R., Horwell, D.C., O'Toole, J.C., Ratcliffe, G.S., Rees, D.C., and Roberts, E. (1992). Amide bond replacements incorporated into CCK-B selective "dipeptoids". J Med Chem 35, 1472-1484.

Fischer, E. (1902). Chem-Ztg 26, 939.

Fischer, E. (1907). Ber Dtsch Chem Ges 40, 1754-1767.

Fischer, E. and Fourneau, E. (1901). Ber Dtsch Chem Ges 34, 2868.

Frankel, A.D. and Pabo, C.O. (1988). Cellular uptake of the tat protein from human immunodeficiency virus. Cell 55, 1189–1193.

Friedrich, C., Scott, M.G., Karunaratne, N., Yan, H., and Hancock, R.E. (1999). Salt-resistant alpha-helical cationic antimicrobial peptides. Antimicrob Agents Chemother 43, 1542-1548.

Frohman, L.A., Downs, T.R., Williams, T.C., Heimer, E.P., Pan, Y.C., and Felix, A.M. (1986). Rapid enzymatic degradation of growth hormone-releasing hormone by plasma *in vitro* and *in vivo* to a biologically inactive product cleaved at the NH2 terminus. J Clin Invest 78, 906-913.

Fuchs, A.R. and Dawood, M.Y. (1980). Oxytocin release and uterine activation during parturition in rabbits. Endocrinology 107, 1117-1126.

Furka, A., Sebestyen, F., Asgedom, M., and Dibo, G. (1991). General method for rapid synthesis of multicomponent peptide mixtures. Int J Pept Protein Res 37, 487-493.

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

Goltzman, D., Bennett, H.P., Koutsilieris, M., Mitchell, J., Rabbani, S.A., and Rouleau, M.F. (1986). Studies of the multiple molecular forms of bioactive parathyroid hormone and parathyroid hormone-like substances. Recent Prog Horm Res 42, 665-703.

Gudmundsson, G.M., Lidholm, D.A., Asling, B., Gan, R., and Boman, H.G. (1991). The cecropin locus. Cloning and expression of a gene cluster encoding three antibacterial peptides in *Hyalophora cecropia*. J Biol Chem 266, 11510-11517.

Harrington, C.R. and Mead, T.H. (1935). Synthesis of glutathione. Biochem J 29, 1602-1611.

Hancock, R.E.W. (2001). Cationic peptides: effectors in innate immunity and novel antimicrobials. Lancet Infect Dis 1, 156-164.

Hancock, R.E.W. and Chapple, D.S. (1999). Peptide Antibiotics. Antimicrob Agents Chemother 43, 1317-1323.

Hancock, R.E.W. and Lehrer, R. (1998). Cationic peptides: a new source of antibiotics. Trends Biotechnol 16, 82-88.

Heavner, G.A., Kroon, D.J., Audhya, T., and Goldstein, G. (1986). Biologically active analogs of thymopentin with enhanced enzymatic stability. Peptides 7, 1015-1019.

Henry, S., McAllister, D.V., Allen, M.G., and Prausnitz, M.R. (1998). Microfabricated microneedles: a novel approach to transdermal drug delivery. J Pharm Sci 87, 922-925.

Holst, J.J., Bersani, M., Hvidberg, A., Knigge, U., Christiansen, E., Madsbad, S., Harling, H., and Kofod, H. (1993). On the effects of human galanin in man. Diabetologia 36, 653-657.

Houghten, R.A. (1985). 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 USA 82, 5131-5135.

Houghten, R.A., Pinilla, C., Blondelle, S.E., Appel, J.R., Dooley, C.T., and Cuervo, J.H. (1991). Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. Nature 354, 84-86.

Hwang, P.M. and Vogel, H.J. (1998). Structure-function relationships of antimicrobial peptides. Biochem Cell Biol 76, 235-246.

Iversen, E. (1988). Intra- and extravascular turnover of thyrotrophin-releasing hormone in normal man. J Endocrinol 118, 511-516.

Kawano, K., Yoneya, T., Miyata, T., Yoshikawa, K., Tokunaga, F., Terada, Y., and Iwanaga, S. (1990). Antimicrobial peptide, tachyplesin I, isolated from hemocytes of the horseshoe crab (*Tachypleus tridentatus*). NMR determination of the beta-sheet structure. J Biol Chem 265, 15365-15367.

Kerr, M.A. and Kenny, A.J. (1974). The purification and specificity of a neutral endopeptidase from rabbit kidney brush border. Biochem J 137, 477-488.

Kokryakov, V.N., Harwig, S.S., Panyutich, E.A., Shevchenko, A.A., Aleshina, G.M., Shamova, O.V., Korneva, H.A., and Lehrer, R.I. (1993). Protegrins: leukocyte antimicrobial peptides that combine features of corticostatic defensins and tachyplesin. FEBS Lett 327, 231-236.

Kolts, B.E. and McGuigan, J.E. (1977). Radioimmunoassay measurement of secretin half-life in man. Gastroenterology 72, 55-60.

Kragol, G., Hoffmann, R., Chattergoon, M.A., Lovas, S., Cudic, M., Bulet, P., Condie, B.A., Rosengren, K.J, Montaner, L.J., and Otvos, L. Jr. (2002). Identification of crucial residues for the antibacterial activity of the proline-rich peptide, pyrrhocoricin. Eur J Biochem 269, 4226-4237.

Lam, K.S., Salmon, S.E., Hersh, E.M., Hruby, V.J., Kazmierski, W.M., and Knapp, R.J. (1991). A new type of synthetic peptide library for identifying ligand-binding activity. Nature 354, 82-84.

Lawyer, C., Pai, S., Watabe, M., Borgia, P., Mashimo, T., Eagleton, L., and Watabe, K. (1996). Antimicrobial activity of a 13 amino acid tryptophan-rich peptide derived from a putative porcine precursor protein of a novel family of antibacterial peptides. FEBS Lett 390, 95-98.

Lehrer, R.I., Lichtenstein, A.K., and Ganz, T. (1993). Defensins: antimicrobial and cytotoxic peptides of mammalian cells. Annu Rev Immunol 11, 105-128.

Lien, S. and Lowman, H.B. (2003). Therapeutic peptides. Trends Biotechnol 21, 556-562.

Liotta, A.S., Li, C.H., Schussler, G.C., and Krieger, D.T. (1978). Comparative metabolic clearance rate, volume of distribution and plasma half-life of human beta-lipotropin and ACTH. Life Sci 23, 2323-2330.

Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46, 3-26.

Low, L.C. (1993). The therapeutic use of growth-hormone-releasing hormone. J Pediatr Endocrinol 6, 15-20.

Mandard, N., Sodano, P., Labbe, H., Bonmatin, J.M., Bulet, P., Hetru, C., Ptak, M., and Vovelle, F. (1998). Solution structure of thanatin, a potent bactericidal and fungicidal insect peptide, determined from proton two-dimensional nuclear magnetic resonance data. Eur J Biochem 256, 404-410.

Marglin, A. and Merrifield R.B. (1966). The synthesis of bovine insulin by the solid phase method. J Am Chem Soc 85, 5051-5052.

Matthews, D.M. (1977). Memorial lecture: protein absorption--then and now. Gastroenterology 73, 1267-1279.

Matthews, D.M. (1987). Ther Clin Nutr 17, 6-53.

McDonald, J.K. and Barrett, A.J. (1986). Mammalian Proteases: A glossary and bibliography; Vol. 2, Exopeptidases (Harcourt Brace Jovanovich: London), pp. 1-6.

McGregor, A., Richards, M., Espiner, E., Yandle, T., and Ikram, H. (1990). Brain natriuretic peptide administered to man: actions and metabolism. J Clin Endocrinol Metab 70, 1103-1107.

Meienhofer, J., Waki, M., Heimer, E.P., Lambros, T.J., Makofske, R.C., and Chang, C.D. (1979). Solid phase synthesis without repetitive acidolysis. Preparation of leucyl-alanyl-glycyl-valine using 9-fluorenylmethyloxycarbonylamino acids. Int J Pept Protein Res 13, 35-42.

Merrifield, R.B. (1963). Solid phase peptide synthesis I. The synthesis of a tetrapeptide. J Am Chem Soc 85, 2149-2154.

Morikawa, N., Hagiwara, K., and Nakajima, T. (1992). Brevinin-1 and -2, unique antimicrobial peptides from the skin of the frog, *Rana brevipoda porsa*. Biochem Biophys Res Commun 189, 184-190.

Na, D.H., Youn, Y.S., Park, E.J., Lee, J.M., Cho, O.R., Lee, K.R., Lee, S.D., Yoo, S.D., DeLuca, P.P., and Lee, K.C. (2004). Stability of PEGylated salmon calcitonin in nasal mucosa. J Pharm Sci 93, 256-261.

Naider, F. and Becker, J.M. (1987). Curr Top Med Mycol 2, 170-198.

Park, C.B., Kim, H.S., and Kim, S.C. (1998). Mechanism of action of the antimicrobial peptide buforin II: buforin II kills microorganisms by penetrating the cell membrane and inhibiting cellular functions. Biochem Biophys Res Commun 244, 253-257.

Parmley, S.F. and Smith, G.P. (1989). Filamentous fusion phage cloning vectors for the study of epitopes and design of vaccines. Adv Exp Med Biol 251, 215-218.

Patrzykat, A., Friedrich, C.L., Zhang, L., Mendoza, V., and Hancock, R.E.W. (2002). Sublethal concentrations of pleurocidin-derived antimicrobial peptides inhibit macromolecular synthesis in *Escherichia coli*. Antimicrob Agents Chemother 46, 605-614.

Pauling, L. and Corey, R.B. (1951). The pleated sheet, a new layer configuration of polypeptide chains. Proc Natl Acad Sci USA 37, 251-256.

Pauling, L., Corey, R.B., and Branson H.R. (1951). The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci USA 37, 205-211.

Payne, J.W. and Smith, M.W. (1994). Peptide transport by microorganisms. Adv Microb Physiol 36, 1-80.

Permanne, B., Adessi, C., Saborio, G.P., Fraga, S., Frossard, M.J., Van Dorpe, J., Dewachter, I., Banks, W.A., Van Leuven, F., and Soto, C. (2002). Reduction of amyloid load and cerebral damage in a transgenic mouse model of Alzheimer's disease by treatment with a beta-sheet breaker peptide. FASEB J 16, 860-862.

Pezalla, P.D., Lis, M., Seidah, N.G., and Chretien, M. (1978). Lipotropin, melanotropin and endorphin: in vivo catabolism and entry into cerebrospinal fluid. Can J Neurol Sci 5, 183-188.

Pontiroli, A.E., Calderara, A., Perfetti, M.G., and Bareggi, S.R. (1993). Pharmacokinetics of intranasal, intramuscular and intravenous glucagon in healthy subjects and diabetic patients. Eur J Clin Pharmacol 45, 555-558.

Powers, J.S. and Hancock, R.E.W. (2003). The relationship between peptide structure and antibacterial activity. Peptides 24, 1681-1691.

Pugsley, A.P. (1996). Bacterial toxins deliver the goods. Proc Natl Acad Sci USA 93, 8155-8156.

Rawlings, N.D. and Barrett, A.J. (1993). Evolutionary families of peptidases. Biochem J 290, 205-218.

Rens-Domiano, S. and Reisine, T. (1992). Biochemical and functional properties of somatostatin receptors. J Neurochem 58, 1987-1996.

Rosenblum, N.G. and Schlaff, S. (1976). Gonadotropin-releasing hormone radioimmunoassay and its measurement in normal human plasma, secondary amenorrhea, and postmenopausal syndrome. Am J Obstet Gyncol 124, 340-347.

Rosenzweig, A. and Seidman, C.E. (1991). Atrial natriuretic factor and related peptide hormones. Annu Rev Biochem 60, 229-255.

Rossjohn, J., Feil, S.C., McKinstry, W.J., Tweten, R.K., and Parker, M.W. (1997). Structure of a cholesterol-binding, thiol-activated cytolysin and a model of its membrane form. Cell 89, 685-692.

Sanger, F. (1952). The arrangement of amino acids in proteins. Adv Protein Chem 7, 1-67.

Schurmeyer, T.H., Avgerinos, P.C., Gold, P.W., Gallucci, W.T., Tomai, T.P., Cutler, G.B. Jr., Loriaux, D.L., and Chrousos, G.P. (1984). Human corticotropin-releasing factor in man: pharmacokinetic properties and dose-response of plasma adrenocorticotropin and cortisol secretion. J Clin Endocrinol Metab 59, 1103-1108.

Scott, J.K. and Smith, G.P. (1990). Searching for peptide ligands with an epitope library. Science 249, 386-390.

Selsted, M.E., Novotny, M.J., Morris, W.L., Tang, Y.Q., Smith, W., and Cullor, J.S. (1992). Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils. J Biol Chem 267, 4292-4295.

Shai, Y. (1995). Molecular recognition between membrane-spanning polypeptides. Trends Biochem Sci 20, 460-464.

Share, L. (1962). Rate of disappearance of arginine vasopressin from circulating blood in the dog. Am J Physiol 203, 1179-1181.

Sifferd, R.H. and du Vigneaud, V. (1935). A new synthesis of carnosine, with some observations on the splitting of the benzyl group from carbobenzoxy derivatives and from benzylthio ethers. J Biol Chem 108, 753-761.

Sodoyez-Goffaux, F., Sodoyez, J.C., Koch, M., De Vos, C.J., and Frank, B.H. (1988). Scintigraphic distribution of 123I labeled proinsulin, split conversion intermediates and insulin in rats. Diabetologia 31, 848-854.

Stenmark, H., Moskaug, J.O., Madshus, I.H., Sandvig, K., and Olsnes, S. (1991). Peptides fused to the amino-terminal end of diphtheria toxin are translocated to the cytosol. J Cell Biol 113, 1025-1032.

Still, J.G. (2002). Development of oral insulin: progress and current status. Diabetes Metab Res Rev 18, S29-S37.

Subbalakshmi, C. and Sitaram, N. (1998). Mechanism of antimicrobial action of indolicidin. FEBS Microbiol Lett 160, 91-96.

Suzuki, Y., Nagase, Y., Iga, K., Kawase, M., Oka, M., Yanai, S., Matsumoto, Y., Nakagawa, S., Fukuda, T., Adachi, H., Higo, N., and Ogawa, Y. (2002). Prevention of bone loss in ovariectomized rats by pulsatile transdermal iontophoretic administration of human PTH(1-34). J Pharm Sci 91, 350-361.

Walker-Smith, D.J. and Payne, J.W. (1984). Planta 162, 166-173.

Walsh, J.H., Isenberg, J.I., Ansfield, J., and Maxwell, V. (1976). Clearance and acid-stimulating action of human big and little gastrins in duodenal ulcer subjects. J Clin Invest 57, 1125-1131.

Watson, J.D. and Crick, F.H. (1953). A structure for deoxyribose nucleic acid. Nature 171, 737-738.

Wu, M., Maier, E., Benz, R., and Hancock, R.E. (1999). Mechanism of interaction of different classes of cationic antimicrobial peptides with planar bilayers and with the cytoplasmic membrane of *Escherichia coli*. Biochemistry 38, 7235-7242.

Yi, G.S., Park, C.B., Kim, S.C., and Cheong, C. (1996). Solution structure of an antimicrobial peptide buforin II. FEBS Lett 398, 87-90.

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

Table 1.1 *In vivo* half-lives of bioactive peptides.

Name	Size (aa)	Plasma half-life	Reference
Thyrotropin releasing hormone (TRH)	3	4 min	Iversen, 1988
Enkephalin	5	1-2 min	Boarder & McArdle, 1986
Thymopentin (TP5)	5	30 sec	Heavner *et al*, 1986
Angiotensin II	8	30 sec	Danser *et al*, 1992
Bradykinin	9	30 sec	Cyr *et al*, 2001
Oxytocin	9	1-5 min	Fuchs & Dawood, 1980
Vasopressin	9	10-20 min	Share, 1962
Gonadotropin releasing hormone (GnRH)	10	2-4 min	Rosenblum and Schlaff, 1976
Melanotropin	13	5 min	Pezalla *et al*, 1978
Neurotensin	13	30 sec	Aronin *et al*, 1982
Somatostatin	14	2 min	Rens-Domiano & Reisine, 1996
Gastrin	17	5-6 min	Walsh *et al*, 1976
Secretin	27	3-4 min	Kolts and McGuigan, 1977
Atrial natriuretic peptide (ANP)	28	2-5 min	Rosenzweig & Seidman, 1991
Vasoactive intestinal peptide (VIP)	28	1 min	Domschke *et al*, 1978
Glucagon	29	5-6 min	Pontiroli *et al*, 1993
Galanin	30	4 min	Holst *et al*, 1993
Glucagon-like peptide 1 (GLP-1)	30	1-2 min	Deacon *et al*, 1995
β-endorphin	31	5 min	Pezalla *et al*, 1978
Brain natriuretic peptide (BNP)	32	3 min	McGregor *et al*, 1990
Calcitonin	32	9 min	Ardaillou *et al*, 1970
Parathyroid hormone (PTH 1-34)	34	5 min	Goltzman *et al*, 1986
Adrenocorticotropic hormone (ACTH)	39	10-15 min	Liotta *et al*, 1978
Corticotropin releasing hormone (CRH)	41	4 min	Schurmeyer *et al*, 1984
Insulin (α/β chains)	51	4-6 min	Sodoyez-Goffaux *et al*, 1988

CHAPTER 2

CONSTRUCTION AND CHARACTERIZATION OF A HIGHLY REGULABLE EXPRESSION VECTOR, pLAC11, AND ITS MULTIPURPOSE DERIVATIVES, pLAC22 AND pLAC33[1]

---

[1]Warren, J.W., J.R.Walker, J.R. Roth, and E. Altman. (2000). Plasmid 44:138-151.

Reprinted here with permission of publisher.

Abstract

A number of different expression vectors have been developed to facilitate the regulated overproduction of proteins in *Escherichia coli* and related bacteria. Some of the more popular ones include pKK223-3, pKK233-2, pTrc99A, and the pET family of expression vectors. These vectors were designed to be regulable and can be grown under conditions that repress protein production or under conditions that induce protein production. Unfortunately, however, numerous researchers have found that these vectors produce significant amounts of protein even when grown under repressed conditions. This study describes a new expression vector, pLAC11, which was designed to be more regulable and thus more tightly repressible when grown under repressed conditions. The tight regulation of pLAC11 was achieved by utilizing the O3 auxiliary operator, CAP binding site, promoter, and O1 operator that occur in the wild-type *lac* control region. The pLAC11 vector can be used to conduct physiologically relevant studies in which the cloned gene is expressed at levels comparable to that obtainable from the chromosomal copy of the gene in question. In experiments in which a bacterial cell contained both a null allele in the chromosome and a second copy of the wild-type allele on pLAC11, it was observed that cells grown under repressed conditions exhibited the null phenotype while cells grown under induced conditions exhibited the wild-type phenotype. Two multipurpose derivatives of pLAC11, pLAC22 and pLAC33, have also been constructed to fulfill different experimental needs.

Introduction

Most of the routinely employed expression vectors rely on *lac* control in order to overproduce a gene of choice (Brosius, 1988; Balbás and Bolivar, 1990). With the wild-type *lac* promoter/operator, induction ratios of up to 1000X have been observed between repressed versus induced growth conditions (Beckwith and Zipser, 1970). The *lac* promoter/operator functions as it does due to the interplay of three main components (the wild-type *lac* control region is shown in Figure 2.1; for general reviews, see Glass, 1982 and Müller-Hill, 1996). First, the wild-type *lac* –10 region (TATGTT) is very weak. c-AMP-activated CAP protein is able to bind to the CAP site just upstream of the –35 region which stimulates binding of RNA polymerase to the promoter. Repression of the *lac* promoter is observed when glucose is the main carbon source because very little c-AMP is present and thus low amounts of c-AMP-activated CAP protein are available. When poor carbon sources such as lactose or glycerol are used, c-AMP levels rise, large amounts of c-AMP-activated CAP protein become available, and thus induction of the *lac* promoter can occur. Second, Lac repressor binds to the *lac* operator, which prevents transcription of the *lac* operon. Lac repressor can be overcome by allolactose, which is a natural by-product of lactose utilization in the cell or by the gratuitous inducer IPTG. Third, the *lac* operator can form stable loop structures which prevent the initiation of transcription due to the interaction of the Lac repressor with the *lac* operator (O1) and one of two auxiliary operators, O2 which is located downstream in the coding region of the *lacZ* gene or O3 which is located just upstream of the CAP binding site.

While binding of Lac repressor to the *lac* operator is the major effector of *lac* regulation, the other two components are not dispensable. Unfortunately, most of the routinely used *lac*-regulable vectors contain either mutations or deletions that alter the effect of the other two

components. The pKK223-3 (Brosius and Holy, 1984), pKK233-2 (Amann and Brosius, 1985), pTrc99A (Amann *et al*., 1988), and pET family of vectors (Studier *et al*., 1990) contain only the *lac* operator (O1) and lack both the CAP binding site and the O3 auxilliary operator. pKK223-3, pKK233-2, and pTrc99A use a *trp-lac* hybrid promoter that contains the *trp* –35 region and the *lac*UV5 –10 region which contains a strong TATAAT site instead of the weak TATGTT site. The pET family of vectors uses the strong T7 promoter. Given this information, perhaps it is not surprising that researchers have found it is not possible to tightly shut off genes that are cloned into these vectors. The promoter/operator control regions utilized by these vectors are shown in Figure 2.2.

The purpose of this study was to design a vector that would allow researchers to be able to better regulate their cloned genes in order to conduct physiological experiments. The novel expression vectors described here were designed utilizing the wild-type *lac* promoter/operator to accomplish this purpose and include all of the *lac* control region that is contained between the start of the O3 auxiliary operator through the end of the O1 operator (Figure 2.1). As with all *lac*-based vectors, expression from the pLAC11, pLAC22, and pLAC33 vectors that have been constructed can be turned on or off by the presence or absence of the gratuitous inducer IPTG.

Materials and Methods

Media

Minimal M9 and rich Luria-Bertani media used in this study were prepared as described by Miller (1972). The antibiotics ampicillin, kanamycin, streptomycin, and tetracycline were used in rich media at final concentrations of 100, 40, 200 and 20 μg/mL, respectively. When used in minimal media, tetracycline was added at a final concentration of 10 μg/mL. XGal was added to media at a final concentration of 40 μg/mL, glucose was added to media at a final

concentration of 0.2%, and unless otherwise noted IPTG was added to media at a final concentration of 1 mM.

Bacterial strains and plasmids

The bacterial strains and plasmids that were used in this study are listed in Table 2.1. To construct ALS225, ALS224 was mated with ALS216 and streptomycin-resistant, blue recombinants were selected on a rich LB plate that contained streptomycin, XGal and IPTG. To construct ALS226, ALS224 was mated with ALS217 and streptomycin-resistant, kanamycin-resistant recombinants were selected on a rich LB plate that contained streptomycin and kanamycin. To construct ALS515, ALS514 was mated with ALS216 and streptomycin-resistant, blue recombinants were selected on a rich LB plate that contained streptomycin, XGal and IPTG. To construct ALS527, ALS524 was mated with ALS224 and streptomycin-resistant, tetracycline-resistant recombinants were selected on a rich LB plate that contained streptomycin and tetracycline. To construct ALS533, a P1 lysate prepared from ALS213 was used to transduce ALS224 and tetracycline-resistant transductants were selected. To construct ALS535, ALS533 was mated with ALS498 and tetracycline-resistant recombinants were selected on a minimal M9 glucose plate that contained tetracycline, leucine and thiamine (B_1). To construct ALS611, a P1 lysate prepared from ALS420 was used to transduce ALS410 and tetracycline-resistant white transductants were selected on a rich LB plate that contained tetracycline, XGal and IPTG. To construct ALS749, a P1 lysate prepared from ALS611 was used to transduce ALS221 and tetracycline-resistant white transductants were selected on a rich LB plate that contained tetracycline, XGal and IPTG.

Construction of the pLAC11, pLAC22, and pLAC33 expression vectors

To construct pLAC11, primers 1 and 2 (Table 2.2) were used to PCR amplify a 953-bp fragment from the plasmid pBH20, which contains the wild-type *lac* operon. Primer 2 introduced two different base-pair mutations into the seven-base spacer region between the Shine-Dalgarno site and the ATG start site of the *lacZ* gene that converted it from AACAGCT to AAGATCT, thus placing a *Bgl*II site between the Shine-Dalgarno and the start codon of the *lacZ* gene. The resulting fragment was gel isolated, digested with *Pst*I and *Eco*RI, and then ligated into the 3613-bp fragment from the plasmid pBR322Δ*Ava*I (described below) that had been digested with the same two restriction enzymes. To construct pBR322Δ*Ava*I, pBR322 was digested with *Ava*I, filled in using Klenow, and then religated.

To construct pLAC22, a 1291-bp *Nco*I-*Eco*RI fragment was gel isolated from pLAC21 (described below) and ligated to a 4361-bp *Nco*I-*Eco*RI fragment that was gel isolated from pBR322/*Nco*I (described below). To construct pLAC21, primers 2 and 3 (Table 2.2) were used to PCR amplify a 1310-bp fragment from the plasmid pMS421 which contains the wild-type *lac* operon as well as the $lacI^{Q}$ repressor. The resulting fragment was gel isolated, digested with *Eco*RI, and then ligated into pBR322 that had also been digested with *Eco*RI. To construct pBR322/*Nco*I, primers 4 and 5 (Table 2.2) were used to PCR amplify a 789-bp fragment from the plasmid pBR322. The resulting fragment was gel isolated, digested with *Pst*I and *Eco*RI, and then ligated into the 3609-bp fragment from the plasmid pBR322 that had been digested with the same two restriction enzymes. The pBR322/*Nco*I vector also contains added *Kpn*I and *Sma*I sites in addition to the new *Nco*I site.

To construct pLAC33, a 2778-bp fragment was gel isolated from pLAC12 (described below) that had been digested with *Bsa*BI and *Bsa*I and ligated to a 960-bp fragment from pUC8

that had been digested with *Afl*III, filled in with Klenow, and then digested with *Bsa*I. To construct pLAC12, a 1311-bp *Pst*I-*Bam*HI fragment was gel isolated from pLAC11 and ligated to a 3232-bp *Pst*I-*Bam*HI fragment that was gel isolated from pBR322.

Compilation of the DNA sequences for the pLAC11, pLAC22, and pLAC33 expression vectors

All of the DNA that is contained in the pLAC11, pLAC22, and pLAC33 vectors has been sequenced. The sequence for the pLAC11 vector which is 4547-bp can be compiled as follows: bp 1-15 is AGATCTTATGAATTC from primer 2 (Table 2.2); bp 16-1434 is bp 4-1422 from pBR322 (Accession No. J01749); bp 1435-1442 is TCGGTCGG, caused by filling in the *Ava*I site in pBR322Δ*Ava*I; bp 1443-4375 is bp 1427-4359 from pBR322 (Accession No. J01749); and bp 4376-4547 is bp 1106-1277 from the wild-type *Escherichia coli lac* operon (Accession No. J01636).

The sequence for the pLAC22 vector which is 5652-bp can be compiled as follows: bp 1-15 is AGATCTTATGAATTC from primer 2 (Table 2.2); bp 16-4370 is bp 4-4358 from pBR322 (Accession No. J01749); bp 4371-4376 is CCATGG which is the *Nco*I site from pBR322/*Nco*I; and bp 4377-5652 is bp 2-1277 from the wild-type *E. coli lac* operon (Accession No. J01636), except that bp 4391 of the pLAC22 sequence or bp 16 from the wild-type *E. coli lac* operon sequence has been changed from a C to a T to reflect the presence of the $lacI^{Q}$ mutation (Calos, 1978).

The sequence for the pLAC33 vector which is 3742-bp can be compiled as follows: bp 1-15 is AGATCTTATGAATTC from primer 2 (Table 2.2); bp 16-1684 is bp 4-1672 from pBR322 (Accession No. J01749); bp 1685-2638 is bp 786-1739 from pUC8 (Accession No. L09132); bp 2639-3570 is bp 3428-4359 from pBR322 (Accession No. J01749); and bp 3571-3742 is bp 1106-1277 from the wild-type *E. coli lac* operon (Accession No. J01636). In the maps for these

vectors the *ori* is identified as per Balbás *et al*. (1986), while the *lac*PO is indicated starting with the O3 auxiliary operator and ending with the O1 operator as per Müller-Hill (1996).

Construction of the pLAC11-, pLAC22-, pLAC33-, pKK223-3-, pKK233-2-, pTrc99A-, and pET-21(+)-lacZ plasmids

To construct pLAC11-*lacZ*, pLAC22-*lacZ*, and pLAC33-*lacZ*, primers 6 and 7 (Table 2.2) were used to PCR amplify a 3116-bp fragment from the plasmid pTer7 which contains the wild-type *lacZ* gene. The resulting fragment was gel isolated, digested with *Bgl*II and *Hin*dIII, and then ligated into the pLAC11, pLAC22, or pLAC33 vectors which had been digested with the same two restriction enzymes. To construct pKK223-3-*lacZ* and pKK233-2-*lacZ*, primers 8 and 9 (Table 2.2) were used to PCR amplify a 3138-bp fragment from the plasmid pTer7. The resulting fragment was gel isolated, digested with *Pst*I and *Hin*dIII, and then ligated into the pKK223-2 or pKK233-2 vectors which had been digested with the same two restriction enzymes. To construct pTrc99A-*lacZ* and pET-21(+)-*lacZ*, primers 9 and 10 (see Table 2.2) were used to PCR amplify a 3138-bp fragment from the plasmid pTer7. The resulting fragment was gel isolated, digested with *Bam*HI and *Hin*dIII, and then ligated into the pTrc99A or pET-21(+) vectors which had been digested with the same two restriction enzymes.

Construction of the pLAC11-recA and -xylE plasmids

To construct pLAC11-*recA*, primers 11 and 12 (Table 2.2) were used to PCR amplify a 1086-bp fragment from the plasmid pGE226, which contains the wild-type *recA* gene. The resulting fragment was gel isolated, digested with *Bgl*II and *Hin*dIII, and then ligated into the pLAC11 vector that had been digested with the same two restriction enzymes. To construct pKK223-3-*recA* and pKK233-2-*recA*, primers 13 and 14 (Table 2.2) were used to PCR amplify a 1104-bp fragment from the plasmid pGE226. The resulting fragment was gel isolated, digested

with *Hin*dIII, and then ligated into either the pKK223-3 or pKK233-2 vector, which had been digested with *Hin*dIII and dephosphorylated with alkaline phosphatase. To construct pLAC11-*xylE*, primers 15 and 16 (Table 2.2) were used to PCR amplify a 980-bp fragment from the plasmid pXE60 which contains the wild-type *Pseudomonas putida xylE* gene isolated from the TOL pWWO plasmid. The resulting fragment was gel isolated, digested with *Bgl*II and *Eco*RI, and then ligated into the pLAC11 vector that had been digested with the same two restriction enzymes.

β-Galactosidase and catechol 2,3-dioxygenase assays

β-Galactosidase (LacZ) assays were performed as described by Miller (1972), while catechol 2,3-dioxygenase (XylE) assays were performed as described by Zukowski *et al*. (1983).

P1 transduction assay

Overnight cultures were prepared from each of the strains to be tested using either rich medium to which glucose was added at a final concentration of 0.2% (repressed conditions) or rich medium to which IPTG was added at a final concentration of 1 mM (induced conditions). The overnights were diluted 1:10 into the same medium that contained calcium chloride added to a final concentration of 10 mM and aerated for 2 h to render them competent for transduction with P1 phage. A 0.1 mL volume of cells at an OD_{550} of 1.0 was transduced with either 0.1 mL of P1 lysate that had been adjusted to yield a maximal number of tranductants or 0.1 mL of a 10^{-2} dilution of the lysate. After a 25-min incubation at 37°C, 0.2 mL of 0.1 M sodium citrate was added to the cell/phage mixtures and 0.2 mL of the final mixtures was plated onto rich kanamycin plates and incubated overnight at 37°C. The total number of kanamycin-resistant colonies was then counted. As reported in Table 2.4, ALS225 *recA*$^{+}$ data points were taken from the transductions which used the 10^{-2} diluted phage, while ALS225 *recA*$^{-}$ data points were taken

from the transductions which used the concentrated phage. The data points for ALS515 *recA*$^-$ strains containing a *recA*$^-$ plasmid were taken from transductions that used the concentrated phage when cells were grown under repressed conditions and transductions that used the 10^{-2} diluted phage when cells were grown under induced conditions.

Chemicals and reagents

When amplified DNA was used to construct the plasmids that were generated in the study, the PCR was carried out using native *Pfu* polymerase from Stratagene (La Jolla, CA, USA). 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (XGal) and isopropyl β-D-thiogalactoside (IPTG) were purchased from Diagnostic Chemicals Limited (Prince Edward Island, Canada).

Results

Construction and features of pLAC11, pLAC22, and pLAC33

The construction of pLAC11, pLAC22, and pLAC33 is described under Materials and Methods and plasmid maps which indicate the unique restriction sites, drug resistances, origins of replication, and other relevant regions that are contained in these vectors are shown in Figure 2.3. Sequences of these three vectors can be compiled as described under Materials and Methods. pLAC11 utilizes the ColE1 origin of replication from pBR322 and Lac repressor is provided in *trans* from either an episome or another compatible plasmid. pLAC22 is very similar to pLAC11; however, it also contains *lacI*^{Q} and thus a source of Lac repressor does not have to be provided in *trans*. pLAC33 is a derivative of pLAC11 which utilizes the mutated ColE1 origin of replication from pUC8 (Lin-Chao *et al*., 1992) and thus the pLAC33 copy number is significantly higher than pLAC11 and is comparable to that of other pUC vectors. Because the cloning regions of these three vectors are identical, cloned genes can be easily

shuffled between the three vectors depending on the expression demands of the experiment in question.

To clone into pLAC11, pLAC22, or pLAC33, PCR amplification is performed with primers that are designed to introduce unique restriction sites just upstream and downstream of the gene of interest (Figure 2.4). Usually a *Bgl*II site is introduced immediately in front of the ATG start codon and an *Eco*RI site is introduced immediately following the stop codon. After amplification, the dsDNA is restricted with the same two enzymes. If the gene of interest contains a *Bgl*II site, then *Bam*HI or *Bcl*I can be used instead since they generate overhangs which are compatible with *Bgl*II. If the gene of interest contains an *Eco*RI site, then a site downstream of *Eco*RI in the vector (such as *Hin*dIII) can be substituted. As shown in Figure 2.4, an additional six bases was included at both ends of the oligonucleotide in order to ensure that complete digestion occurs.

Comparison of pLAC11, pLAC22, and pLAC33 to other expression vectors

In order to demonstrate how regulable the pLAC11, pLAC22, and pLAC33 expression vectors were, the wild-type *lacZ* gene was cloned into pLAC11, pLAC22, pLAC33, pKK223-3, pKK233-2, pTrc99A, and pET-21(+). Constructs that required an extraneous source of LacI for their repression were transformed into strain ALS226, while constructs that contained a source of LacI on the vector were transformed into strain ALS224. pET-21(+) constructs were transformed into strain ALS749 because they require T7 RNA polymerase for their expression. Four independent *lacZ* clones were chosen from each vector. The plasmids pLysE and pLysS which make T7 lysozyme and thus lower the amount of available T7 polymerase were also transformed into each of the four pET-21(+)-lacZ clones. LB + ampicillin overnights were diluted 1:200 in either LB + ampicillin + glucose medium (repressed conditions) or LB +

ampicillin + IPTG medium (induced conditions) and grown until they reached mid-log (OD_{550} = 0.5). Cell extracts were prepared and β-galactosidase assays were performed as per Miller (1972). Table 2.3 shows the results of these studies and also lists the induction ratio that was determined for each of the expression vectors. As the data clearly indicate, pLAC11 is the most regulable of these expression vectors and its induction ratio is close to that which can be achieved with the wild-type *lac* operon. The vector that yielded the lowest level of expression under repressed conditions was pLAC11, while the vector that yielded the highest level of expression under induced conditions was pLAC33.

Demonstrating that pLAC11 constructs can be tightly regulated

To further demonstrate the utility of pLAC11 for physiological experiments, the *recA* gene was cloned into the pLAC11, pKK223-3, and pKK233-2 vectors and transformed into cells that contained a null *recA* allele in the chromosome. The pKK223-3 and pKK233-2 vectors were chosen as controls because aside from pLAC11, they were the most tightly regulable of all the vectors that were examined in the experiments with *lacZ* shown in Table 2.2. As the data in Table 2.4 clearly show, recombination cannot occur in a host strain which contains a nonfunctional RecA protein and thus P1 lysates which provide a Tn10dKan transposon cannot be used to transduce the strain to kanamycin resistance at a high frequency. A *recA*$^-$ strain which also contains the pLAC11-*recA* construct can be transduced to kanamycin resistance at a high frequency when grown under induced conditions but cannot be transduced to kanamycin resistance when grown under repressed conditions. This is not the case for the pKK223-3-*recA* and pKK233-2-*recA* constructs as a significant number of transductants above background can be obtained under repressed conditions.

pLAC11 was designed to provide researchers with an expression vector that could be utilized to conduct physiological experiments in which a cloned gene is studied under completely repressed conditions where it is off or partially induced conditions where it is expressed at physiologically relevant levels. Figure 2.5 demonstrates how a pLAC11-*lacZ* construct can be utilized to mimic chromosomally expressed *lacZ* that occurs under various physiological conditions by varying the amount of IPTG inducer added.

Testing various sources of LacI for trans repression of pLAC11

Because pLAC11 was designed to be used with an extraneous source of Lac repressor, different episomal or plasmid sources of LacI that are routinely employed by researchers were tested. Since one of the LacI sources also contained the *lacZ* gene, a reporter construct other than pLAC11-*lacZ* was required and thus a pLAC11-*xylE* construct was engineered. Table 2.5 shows the results of this study. All of the LacI sources that were tested proved to be adequate to repress expression from pLAC11; however, some were better than others. The basal level of expression that was observed with F's which provided $lacI^{Q1}$ or with the plasmid pMS421 which provided $lacI^{Q}$ at approximately six copies per cell was lower than the basal level of expression that was observed with F's which provided $lacI^{Q}$ all three times that the assay was performed. Unfortunately, however, the *xylE* gene could not be induced as high when $lacI^{Q1}$ on an F' or $lacI^{Q}$ on a plasmid was used as the source of Lac repressor.

Discussion

Most of the routinely employed expression vectors rely on elements of the *lac* control region for their regulation. While these vectors allow for the overexpression of the gene product of interest, they are leaky due to changes that have been introduced into the *lac* control region and gene expression cannot be completely turned off under repressed conditions. Numerous

researchers have noticed this problem with the more popular expression vectors pKK223-3 (Posfai *et al*, 1986; Scrutton *et al*, 1987), pKK233-2 (Beremand *et al*, 1987; Ooki *et al*, 1994), and pTrc99A (Ranie *et al*, 1993; Ghosh and Singh, 1997), as well as the pET series (Eren and Swenson, 1989; Godson, 1991). Described here is a new vector, pLAC11, which relies on the wild-type *lac* control region from the auxiliary *lac* O3 operator through the *lac* O1 operator and thus can be more tightly regulated than the other available expression vectors. In direct comparison studies with pKK223-3, pKK233-2, pTrc99A, and pET-21(+), it was found that the lowest level of expression under repressed conditions was achievable with the pLAC11 expression vector. Under fully induced conditions, pLAC11 expressed LacZ protein at levels that were comparable to what could be achieved with the other expression vectors.

Induction ratios of 1000X have been observed with the wild-type *lac* operon. Of all the expression vectors that were tested, only pLAC11 yielded induction ratios that were comparable to what has been observed with the wild-type *lac* operon. It should be noted that the regulation achievable by pLAC11 is actually better than the data indicated in Table 2.3. Because *lacZ* was used in this test, the auxiliary *lac* O2 operator that resides in the coding region of the *lacZ* gene was provided to the pKK223-3, pKK233-2, pTrc99A, and pET-21(+) vectors which do not normally contain either the O2 or O3 auxiliary operators. Thus the repressed states that were observed in the study in Table 2.3 are lower than what is normally obtainable with the pKK223-3, pKK233-2, pTrc99A, and pET-21(+) vectors. The studies with RecA in Table 2.4 demonstrate this as RecA protein expression could only be completely shut off under repressed conditions using pLAC11. Significant expression of the RecA protein occurred in the pKK223-3 and pKK233-2 vectors under repressed conditions.

To meet the expression needs required under different experimental circumstances, two additional expression vectors that are derivatives of pLAC11 were designed and constructed. pLAC22 provides *lacI*^{Q} on the vector and thus, unlike pLAC11, does not require an extraneous source of LacI for its repression. pLAC33 contains the mutated ColE1 replicon from pUC8 and thus allows proteins to be expressed at much higher levels due to the increase in the copy number of the vector. Of all the expression vectors that were evaluated in direct comparison studies, the highest level of protein expression under fully induced conditions was achieved using the pLAC33 vector. Because the cloning regions are identical in pLAC11, pLAC22, and pLAC33, genes that are cloned into one of these vectors can be subcloned into one of the other two vectors depending on experimental circumstances. For physiological studies, pLAC11 is the best suited of the three vectors. If however, the bacterial strain of choice cannot be modified to introduce elevated levels of Lac repressor protein that can be achieved by F's or compatible plasmids that provide *lacI*^{Q} or *lacI*^{Q1}, the pLAC22 vector can be utilized. If maximal overexpression of a gene product is the goal, then the pLAC33 vector should be utilized.

Numerous experiments call for expression of a cloned gene product at physiological levels, i.e., at expression levels that are equivalent to the expression levels observed for the chromosomal copy of the gene. While this is not easily achievable with any of the commonly utilized expression vectors, these kinds of experiments can be done with the pLAC11 expression vector. By varying the IPTG concentrations, expression from the pLAC11 vector can be adjusted to match the expression levels that occur under different physiological conditions for the chromosomal copy of the gene.

Because the use of Lac repressor is an essential component of any expression vector that utilizes the *lac* operon for its regulation, the ability of different sources of LacI to repress the

pLAC11 vector was also investigated. Researchers have historically utilized either *lacI*^{Q} constructs that make 10-fold more Lac repressor than wild-type *lacI*, or *lacI*^{Q1} constructs that make 100-fold more Lac repressor than wild-type *lacI* (Müller-Hill, 1975). It was found that the greatest level of repression of pLAC11 constructs could be achieved using F's that provided approximately one copy of the *lacI*^{Q1} gene or a multicopy compatible plasmid that provided approximately six copies of the *lacI*^{Q} gene. However, the induction that was achievable using these LacI sources was significantly lower than the induction that could be achieved when F's that provided approximately one copy of the *lacI*^{Q} gene were used to repress the pLAC11 constructs. Thus if physiological studies are the goal of an investigation, then F's that provide approximately one copy of the *lacI*^{Q1} gene or a multicopy compatible plasmid that provides approximately six copies of the *lacI*^{Q} gene should be used to regulate the pLAC11 vector. However, if maximal expression is desired, then F's that provide approximately one copy of the *lacI*^{Q} gene should be utilized. Alternatively, if a bacterial strain can tolerate prolonged overexpression of an expressed gene and overexpression of a gene product is the desired goal, then maximal expression under induced conditions is obtained when a bacterial strain lacks any source of Lac repressor.

References

Altman, E., Bankaitis, V.A., and Emr., S.D. (1990). Characterization of a region in mature LamB protein that interacts with a component of the export machinery of *Escherichia coli*. J Biol Chem 265, 18148-18153.

Amann, E. and Brosius, J. (1985). ATG vectors for regulated high-level expression of cloned genes in *Escherichia coli*. Gene 40, 183-190.

Amann, E., Ochs, B., and Abel, K-J. (1988). Tightly regulated *tac* vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene 69, 301-315.

Balbás, P. and Bolivar, F. (1990). Design and construction of expression plasmid vectors in *Escherichia coli. In* "Methods in Enzymology" (D.V. Goeddel, Ed.), Vol. 185, pp. 14-37. Academic Press, Inc., Harcourt Brace Jovanovich, Publishers.

Balbás, P., Soberón, X., Merino, E., Zurita, M., Hilda, M., Valle, F., Flores, N., and Bolivar, F. (1986). Plasmid vector pBR322 and its special-purpose derivatives - a review. Gene 50, 3-40.

Beckwith, J.R. and Zipser, D. (eds.) (1970). "The Lactose Operon." Cold Spring Harbor Laboratory, N.Y.

Beremand, P.D., Hannapel, D.J., Guerra, D.J., Kuhn, D.N. and Ohlrogge, J.B. (1987). Synthesis, cloning, and expression in *Escherichia coli* of a spinach acyl carrier protein-I gene. Arch Biochem Biophys 256, 90-100.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H., and Falkow, S. (1977). Construction and characterization of new cloning vehicles II. A multipurpose cloning system. Gene 2, 95-113.

Brosius, J. (1988). Expression vectors employing lambda-, *trp*-, *lac*-, and *lpp*-derived promoters. Biotechnology 10, 205-225.

Brosius, J. and Holy, A. (1984). Regulation of ribosomal RNA promoters with a synthetic *lac* operator. Proc Natl Acad Sci USA 81, 6929-6933.

Calos, M.P. (1978). DNA sequence for a low-level promoter of the *lac* repressor gene and an 'up' promoter mutation. Nature 274, 762-765.

Casadaban, M. and Cohen, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J Mol Biol 138, 179-207.

Eren, M. and Swenson, R.P. (1989). Chemical synthesis and expression of a synthetic gene for the flavodoxin from *Clostridium MP*. J Biol Chem 264, 14874-14879.

Ghosh, S. and Singh, M. (1997). cDNA cloning, expression, and rapid purification of a Kunitz-type winged bean chymotrypsin inhibitor. Protein Expr. Purif. 10, 100-106.

Glass, R.E. (1982). "Gene function: *E. coli* and its heritable elements." University of California Press, Berkeley and Los Angeles, California.

Godson, G.N. (1991). An over-expression plasmid for *Escherichia coli* primase. Gene 100, 59-64.

Graña, D., Gardella, T., and Susskind, M.M. (1988). The effects of mutations in the *ant* promoter of Phage P22 depend on context. Genetics 120, 319-327.

Guyer, M.S., Reed, R.R., Steitz, J.A., and Low, K.B. (1980). Identification of a sex-factor-affinity site in *E. coli* as gd. Cold Spring Harbor Symp Quant Biol 45, 135-140.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F., and Boyer, H.W. (1977). Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science 198, 1056-1063.

Lin-Chao, S., Chen, W-T., and Wong, T-T. (1992). High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. Mol Micro 6, 3385-3393.

Miller, H.I. and Friedman, D.I. (1980). An *E. coli* gene product required for lambda site-specific recombination. Cell 20, 711-719.

Miller, J.H. (1972). "Experiments in molecular genetics." Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Müller-Hill, B. (1975). Lac repressor and *lac* operator. Prog Biophys Mol Biol 30, 227-252.

Müller-Hill, B. (1996). "The *lac* operon: a short history of a genetic paradigm." Walter de Gruyter, Berlin, Germany.

Ooki, K., Amuro, N., Shimizu, Y., and Okazaki, T. (1994). High level expression of rat gamma-D-crystallin in *Escherichia coli*. Biochimie 76, 398-403.

Posfai, G., Kiss, A., and Venetianer, P. (1986). Overproduction of the *Bacillus sphaericus* R modification methylase in *Escherichia coli* and its purification to homogeneity. Gene 50, 63-67.

Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W., Elhai, J., and Hanahan, D. (1988). McrA and McrB restriction phenotypes and some *E. coli* strains and implications for gene cloning. Nucleic Acids Res 16, 1563-1575.

Ranie, J., Kumar, V.P., and Balaram, H. (1993). Cloning of the triosephosphate isomerase gene of *Plasmodium falciparum* and expression in *Escherichia coli*. Mol Biochem Parasitol 61, 159-169.

Scrutton, N.S., Berry, A., and Perham, R.N. (1987). Purification and characterization of glutathione reductase encoded by a cloned and over-expressed gene in *Escherichia coli*. Biochem J 245, 875-880.

Singer, M., Baker, T.A., Schnitzler, G., Deischel, S.M., Goel, M., Dove, W., Jaacks, K.J., Grossman, D., Erickson, J.W., and Gross, C.A. (1989). A collection of strains containing

genetically linked altering antibiotic resistance elements for genetic mapping of *Escherichia coli*. Microbiol Rev 53, 1-24.

Studier, F.W. (1991). Use of bacteriophage-T7 lysozyme to improve an inducible T7 expression system. J Mol Biol 219, 37-44.

Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 189, 113-130.

Studier, F.W., Rosenberg, A.H., Dunn J.J., Dubendorff J.W. (1990). Use of T7 RNA-polymerase to direct expression of cloned genes. *In* "Methods in Enzymology" (D.V. Goeddel, Ed.), Vol. 185, pp. 60-89. Academic Press, Inc., Harcourt Brace Jovanovich, Publishers.

Vieira, J. and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19, 259-268.

Weisemann, J.A. and Weinstock, G.M. (1985). Direct selection of mutations reducing transcription or translation of the *recA* gene of *Escherichia coli* with a *recA-lacZ* protein fusion. J Bacteriol 163, 748-755.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33, 103-119.

Zukowski, M.M., Gaffney, D.F., Speck, D., Kauffmann, M., Findeli, A., Wisecup, A., Lecocq, J.P. (1983). Chromogenic identification of genetic regulatory signals in *Bacillus subtilis* based on expression of a cloned *Pseudomonas* gene. Proc Natl Acad Sci USA 80, 1101-1105.

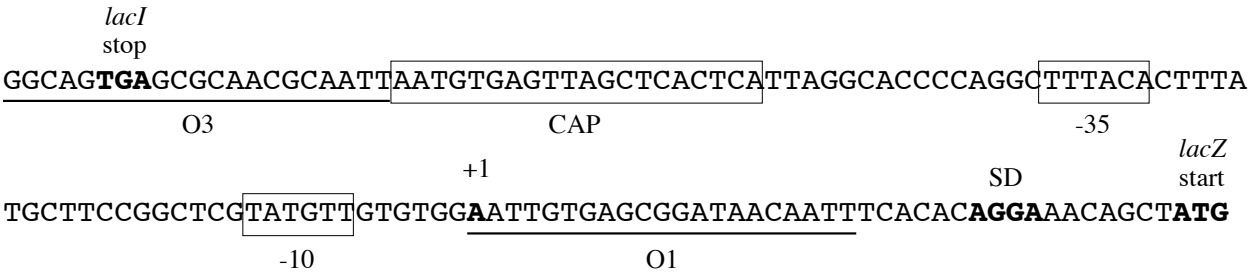


Figure 2.1 Control region of the wild-type *lac* operon. The region from the O3 auxiliary operator through the translational start of the *lacZ* gene is shown. DNA binding sites are indicated below the DNA sequence while important RNA sites are shown above the DNA sequence. The Shine-Dalgarno ribosome binding site for *lacZ* is indicated by SD.

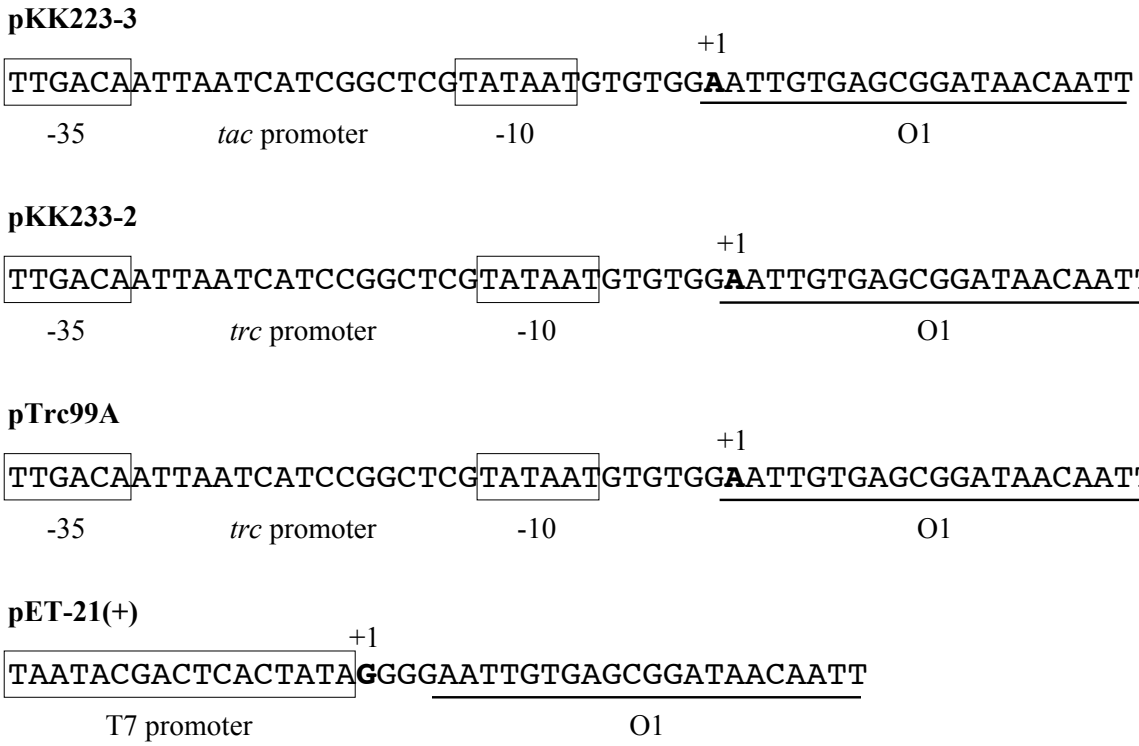


Figure 2.2 Control regions of routinely employed expression vectors. The region from the promoter through the *lac* O1 operator for the pKK223-3, pKK233-2, pTrc99A, and pET-21(+) vectors is shown. DNA binding sites are indicated below the DNA sequence while important RNA sites are shown above the DNA sequence. The pET-21(+) control region is typical of the control region that is found in other pET expression vectors. Note that the pKK223-3, pKK233-2, pTrc99A, and pET-21(+) expression vectors do not contain a CAP binding site and lack an auxiliary *lac* operator.

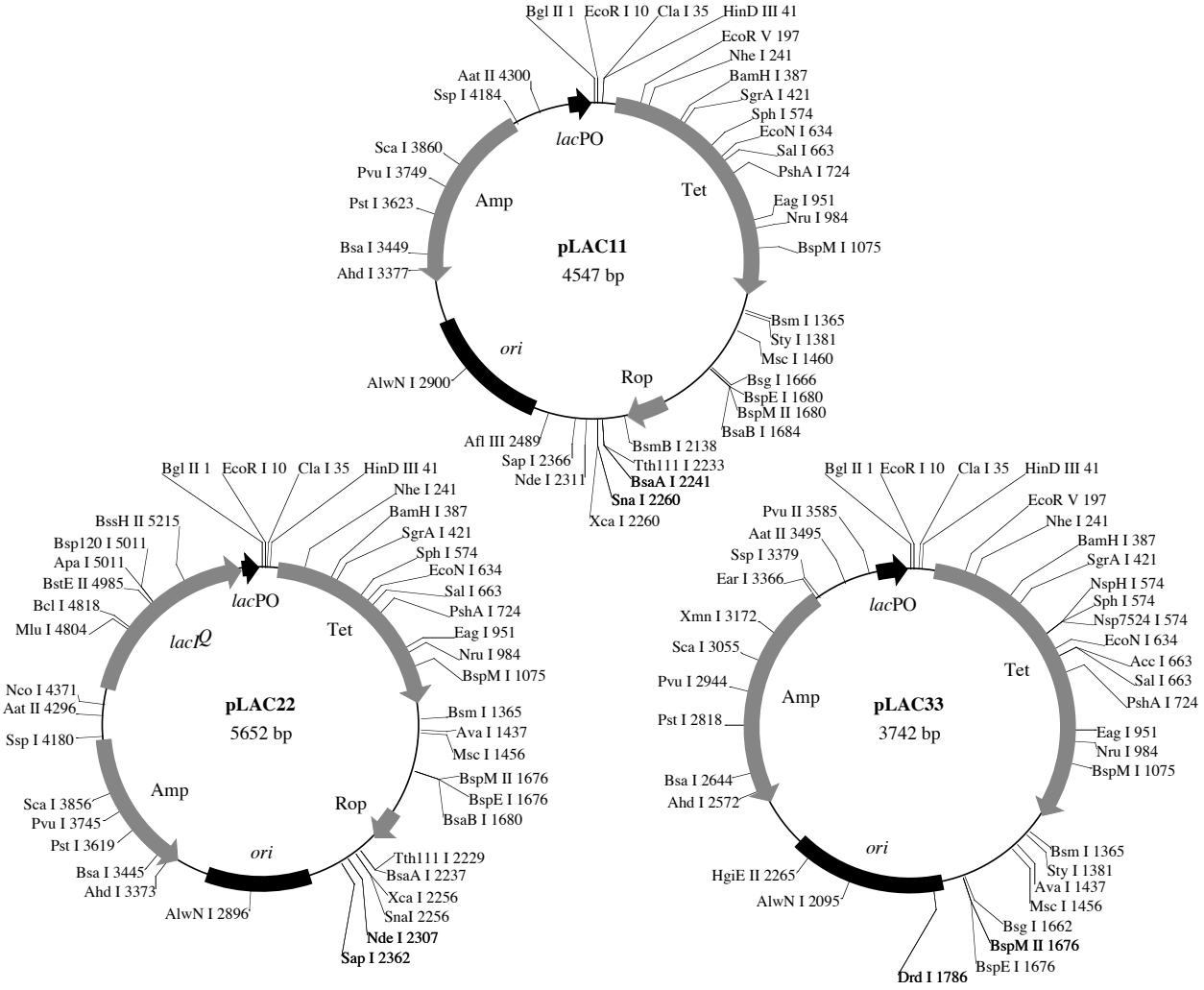


Figure 2.3 Map of plasmids pLAC11, pLAC22, and pLAC33. The unique restriction sites and the base pair at which they cut are indicated. Key sites of interest on these plasmids are also shown: pLAC11; Tet (98-1288), Rop (1931-2122), *ori* (2551-3138), Amp (3309-4169), and *lac*PO (4424-4536), pLAC22; Tet (98-1288), Rop (1927-2118), *ori* (2547-3134), Amp (3305-4165), $lacI^{Q}$ (4452-5536), and *lac*PO (5529-5641), pLAC33; Tet (98-1288), *ori* (1746-2333), Amp (2504-3364), and *lac*PO (3619-3731).

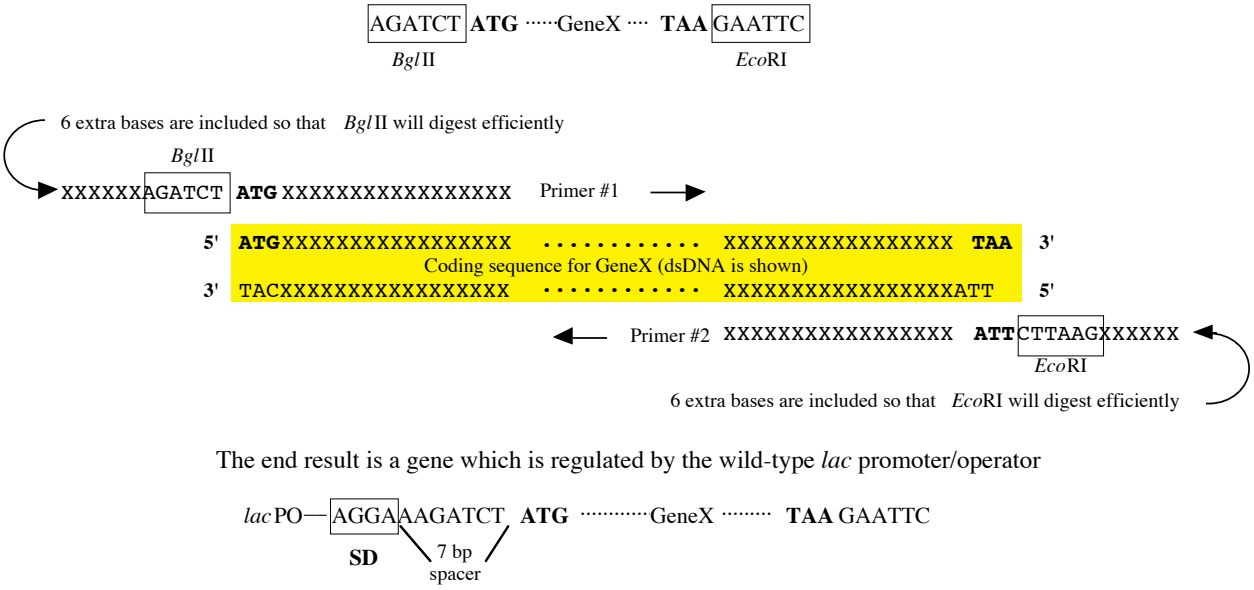


Figure 2.4 Cloning into pLAC11, pLAC22, or pLAC33. This figure shows how the coding region of interest can be PCR amplified, cloned into pLAC11, pLAC22, or pLAC33, and placed under *lac* control.

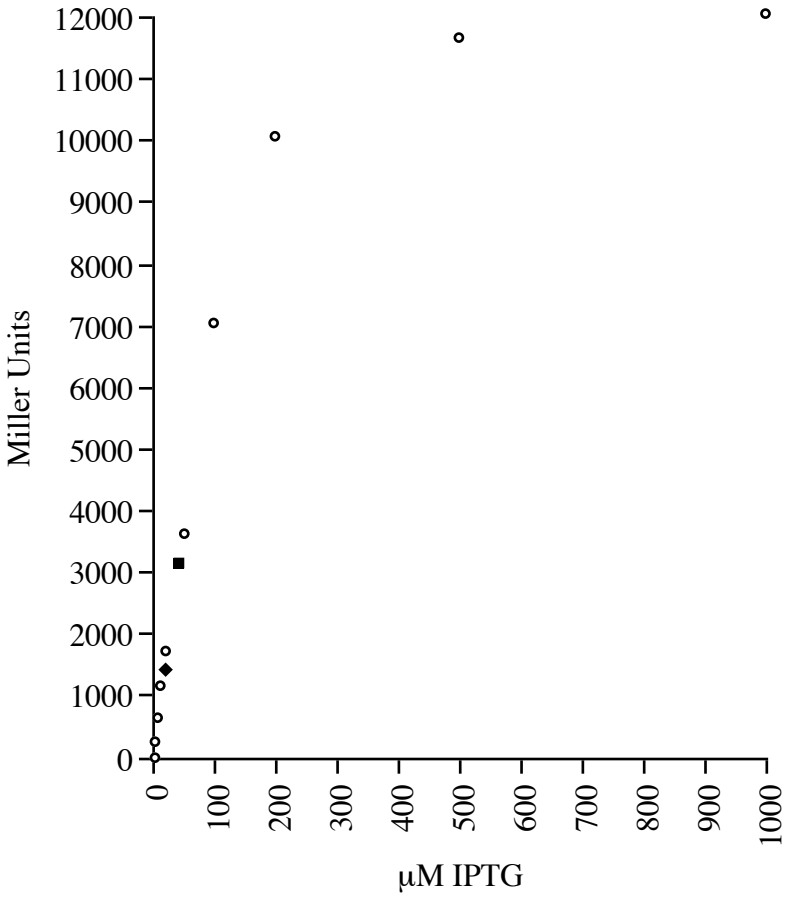


Figure 2.5 Response of the pLAC11-*lacZ* construct to varying amounts of IPTG.

ALS226 cells containing pLAC11-*lacZ* were grown to mid-log in rich media that contained varying amounts of IPTG and then β-galactosidase activity was assayed. These data points are indicated by open circles. Also indicated in the graph are the average β-galactosidase activities obtained for strains with a single chromosomal copy of the wild-type *lacZ* gene that were grown under different conditions. A filled square indicates the β-galactosidase activity that was obtained when MG1655 or CSH27 cells were grown in rich media induced with 1 mM IPTG, while a filled diamond indicates the β-galactosidase activity that was obtained when MG1655 or CSH27 cells were grown in M9 minimal lactose media.

Table 2.1a Bacterial strains and plasmids used in this work

Bacterial strains (*E. coli*)			
Laboratory name	Original name	Genotype	Reference or source
ALS213	K5096	*proAB*::Tn10	Miller and Friedman, 1980
ALS216	SE9100	*araD139* Δ*(lac)U169 thi f1bB5301 deoC7 ptsF25 rpsE* / F'$lacI^{Q1}$ Z^+ Y^+ A^+	Altman *et al.*, 1990
ALS217	SE9100.1	*araD139* Δ*(lac)U169 thi f1bB5301 deoC7 ptsF25 rpsE* / F' $lacI^{Q1}$ *Z*::Tn5 *Y*+ *A*+	S. Emr
ALS221	BL21(DE3)	*ompT hsdS*(b) (R-M-) *gal dcm* (DE3)	Studier and Moffatt, 1986
ALS224	MC1061	*araD139* Δ*(araABOIC-leu)7679* Δ*(lac)X74 galU galK rpsL hsr- hsm*+	Casadaban and Cohen, 1980
ALS225		MC1061 / F'$lacI^{Q1}$ Z^+ Y^+ A^+	This work
ALS226		MC1061 / F'$lacI^{Q1}$ *Z*::Tn5 Y^+ A^+	This work
ALS269	CSH27	F- *trpA33 thi*	Miller, 1972
ALS410	CSH1	F- *trp lacZ rpsL thi*	Miller, 1972
ALS413	MG1655	*E. coli* wild-type F- l-	Guyer *et al.*, 1980
ALS420	RS1071	*leuB6 fhuA2* zah-281::Tn10 *glnV44*(AS) *gal-6* lambda- *trp-31 hisG1*(Fs) *argG6 rpsL104 malT1*(lambda res) *xylA7 mtlA2 metB1*	R. Simons
ALS498	JM101	*supE thi* Δ*(lac-proAB)* / F' *traD36 proA*$^+$*B*$^+$ $lacI^Q$ Δ*(lacZ)M15*	Yanisch-Perron, 1985
ALS514	NM554	MC1061 *recA13*	Raleigh *et al.*, 1988
ALS515		MC1061 *recA13* / F'$lacI^{Q1}$ Z^+ Y^+ A^+	This work
ALS524	XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* / F' *proAB* $lacI^Q$ Δ*(lacZ)M15* Tn10	Stratagene
ALS527		MC1061 / F' *proAB* $lacI^Q$ Δ*(lacZ)M15* Tn10	This work
ALS533		MC1061 *proAB*::Tn10	This work
ALS535		MC1061 *proAB*::Tn10 / F'$lacI^Q$ Δ*(lacZ)M15 proA+B+*	This work
ALS598	CAG18615	zjb-3179::Tn10dKan lambda- *rph*-1	Singer *et al.*, 1989
ALS611		CSH1 zah-281::Tn10	This work
ALS749		BL21(DE3) *lacZ* zah-281::Tn10	This work

Table 2.1b Bacterial strains and plasmids used in this work

Plasmids		
Plasmid name	Relevant characteristics	Reference or source
pBH20	wild-type *lac* promoter / operator, Amp^R, Tet^R, ColE1 replicon	Itakura *et al.*, 1977
pBR322	Amp^R, Tet^R, ColE1 replicon	Bolivar *et al.*, 1977
pET-21(+)	T7 promoter / *lac* operator, *lacI*, Amp^R, ColE1 replicon	Novagen
pGE226	wild-type *recA* gene, Amp^R	Weisemann and Weinstock, 1985
pKK223-3	*tac* promoter / operator, Amp^R, ColE1 replicon	Brosius and Holly, 1984
pKK233-2	*trc* promoter / operator, Amp^R, ColE1 replicon	Amann and Brosius, 1985
pLysE	T7 lysozyme, Cam^R, P15A replicon	Studier, 1991
pLysS	T7 lysozyme, Cam^R, P15A replicon	Studier, 1991
pMS421	wild-type *lac* promoter / operator, $lacI^Q$, $Strep^R$, $Spec^R$, SC101 replicon	Graña *et al.*, 1988
pTer7	wild-type *lacZ* coding region, Amp^R	R. Young
pTrc99A	*trc* promoter / operator, $lacI^Q$, Amp^R, ColE1 replicon	Amann *et al.*, 1988
pUC8	*lac* promoter / operator, Amp^R, ColE1 replicon	Vieira and Messing, 1982
pXE60	wild-type TOL pWWO *xylE* gene, Amp^R	J. Westpheling

Table 2.2 Primers employed to PCR amplify DNA fragments that were used in the construction of the various plasmids described in this study

pLAC11 and pLAC22	
1 (for)	GTT GCC ATT GCT GCA GGC AT
2 (rev)	ATT GAA TTC ATA AGA TCT TTC CTG TGT GAA ATT GTT ATC CGC
3 (for)	ATT GAA TTC ACC ATG GAC ACC ATC GAA TGG TGC AAA A
pBR322/NcoI	
4 (for)	GTT GTT GCC ATT GCT GCA G
5 (rev)	TGT ATG AAT TCC CGG GTA CCA TGG TTG AAG ACG AAA GGG CCT C
*Bgl*II - *lacZ* - *Hin*dIII	
6 (for)	TAC TAT AGA TCT ATG ACC ATG ATT ACG GAT TCA CTG
7 (rev)	TAC ATA AAG CTT GGC CTG CCC GGT TAT TAT TAT TTT
*Pst*I - *lacZ* - *Hin*dIII	
8 (for)	TAT CAT CTG CAG AGG AAA CAG CTA TGA CCA TGA TTA CGG ATT CAC TG
9 (rev)	TAC ATA CTC GAG CAG GAA AGC TTG GCC TGC CCG GTT ATT ATT ATT TT
*Bam*H1 - *lacZ* - *Hin*dIII (also uses primer #9)	
10 (for)	TAT CAT GGA TCC AGG AAA CAG CTA TGA CCA TGA TTA CGG ATT CAC TG
*Bgl*II - *recA* - *Hin*dIII	
11 (for)	TAC TAT AGA TCT ATG GCT ATC GAC GAA AAC AAA CAG
12 (rev)	ATA TAT AAG CTT TTA AAA ATC TTC GTT AGT TTC TGC TAC G
*Hin*d III - *recA* - *Hin*dIII	
13 (for)	TAC TAT AAG CTT AGG AAA CAG CTA TGG CTA TCG ACG AAA ACA AAC AG
14 (rev)	ATA TAT CCC GGG CAA GCT TTT AAA AAT CTT CGT TAG TTT CTG CTA CG
*Bam*H1 - *xylE* - *Eco*RI	
15 (for)	TAC TAT AGA TCT ATG AAC AAA GGT GTA ATG CGA CC
16 (rev)	ATT AGT GAA TTC GCA CAA TCT CTG CAA TAA GTC GT

The regions of the primers that are homologous to the DNA target template are indicated with a dotted underline, while the relevant restriction sites are indicated with a solid underline. All primers are listed in the 5' to 3' orientation.

Table 2.3 β-Galactosidase levels obtained in different expression vectors grown under either repressed or induced conditions

Host strain	Vector	Source of LacI	No. of Miller units observed		Fold induction
			Repressed conditions	Induced conditions	
ALS224	none	none	3	6	---
ALS226	none	F'	4	7	---
ALS226	pLAC11	F'	19	11209	590X
ALS224	pLAC22	plasmid	152	13315	88X
ALS226	pLAC33	F'	322	23443	73X
ALS226	pKK223-3	F'	92	11037	120X
ALS226	pKK233-2	F'	85	10371	122X
ALS224	pTrc99A	plasmid	261	21381	82X
ALS749	none	none	3	4	---
ALS749	pET-21(+)	plasmid	2929	16803	6X
ALS749	pET-21(+) /pLysE	plasmid	4085	19558	5X
ALS749	pET-21(+) /pLysS	plasmid	1598	20268	13X

The average values obtained for the four clones that were tested from each vector in two different experiments are listed in the table. Standard deviation is not shown but was less than 5% in each case. Induction ratios are expressed as the ratio of enzymatic activity observed at fully induced conditions versus fully repressed conditions. Because pLysE yielded unexpected results, we restriction mapped both of the pLysE and pLysS plasmids to make sure that they were correct.

Table 2.4 The phenotype of a *recA* null mutant strain can be preserved with a pLAC11-*recA* (wild-type) construct under repressed conditions

Strain	Repressed Conditions	Induced Conditions
	Number of Kan^R Transductants	Number of Kan^R Transductants
ALS225 (*recA*$^+$)	178,000	182,000
ALS515 (*recA*$^-$)	5	4
ALS515 (*recA*$^-$ pLAC11-*recA*)	4	174,000
ALS515 (*recA*$^-$ pKK223-3-*recA*)	146	179,000
ALS515 (*recA*$^-$ pKK233-2-*recA*)	143	158,000

The data presented in the table are the number of kanamycin resistant (Kan^R) transductants that were obtained from the different isogenic strains when they were transduced with a P1 lysate prepared from strain ALS598 which harbored a Tn10dKan transposon insertion. Cells were grown in rich media under either repressed or induced conditions, transduced with equal amounts of the P1 lysate, plated onto LB + kanamycin plates and incubated overnight at 37°C as described in the Materials and Methods section. The total number of kanamycin resistant colonies were then counted.

Table 2.5 Catechol 2,3-dioxygenase levels obtained for a pLAC11-*xylE* construct when Lac repressor is provided by various sources

Host strain	Plasmid present	Source of LacI	Catechol 2,3-dioxygenase activity in milliunits/mg	
			Repressed conditions (rich glucose)	Induced conditions (rich IPTG)
ALS224	No	none	0.2	0.2
ALS224	Yes	none	32.7	432.8
ALS535	Yes	F'*lacI*^{Q} Δ*(lacZ)M15 proA+B+* Tn10	0.3	204.4
ALS527	Yes	F'*lacI*^{Q} Δ*(lacZ)M15 proA+B+*	0.3	243.3
ALS227	Yes	pMS421 *lacI*^{Q}	0.2	90.9
ALS225	Yes	F'*lacI*^{Q1} *Z*$^{+}$ *Y*$^{+}$ *A*$^{+}$	0.2	107.4
ALS226	Yes	F'*lacI*^{Q1} *Z*::Tn5 *Y*$^{+}$ *A*$^{+}$	0.2	85.1

The pLAC11-*xylE* construct was transformed into each of the MC1061 derivative strains listed in the table. ALS224, the parental MC1061 strain, without the pLAC11-*xylE* construct served as a control. Rich overnights were diluted 1:200 in either rich glucose or rich IPTG media and grown until they reached mid-log (OD_{550} = 0.5). Cell extracts were then prepared and catechol 2,3-dioxygenase assays were performed as described by Zukowski *et al.*, 1983. The average values obtained in three different experiments are listed in the table. Standard deviation is not shown but was less than 10% in each case. It should be noted that some repression is achieved in ALS224 pLAC11-*xylE* under repressed conditions, because the addition of glucose prevents the accumulation of high levels of c-AMP.

CHAPTER 3

USING PROTEIN-BASED MOTIFS TO STABILIZE PEPTIDES[1]

[1]The research presented in this chapter was published in Walker, J.R., R.K. Altman, J.W. Warren, and E. Altman. (2003). J Peptide Res 62:214-226.

Abstract

While the use of synthetically derived novel bioactive peptides as a source of new therapeutics for medicine remains incredibly promising, there is a major problem with implementing this technology, since many synthetic peptides have proven to be unstable and are degraded by peptidases in the host cell. This study investigated whether randomized inhibitor peptides can be stabilized via fusion to the highly stable Rop protein in order to prevent the action of peptidases. Using an *in vivo* genetic approach developed to screen for synthetic peptides that can inhibit the growth of *Escherichia coli*, it was found that protecting the amino or carboxyl terminus of the peptides via fusion to the Rop protein dramatically increased the frequency at which potent inhibitor peptides could be isolated.

Introduction

Over the past several years there has been an increasing interest in using novel synthetic peptides as a means to generate new therapeutic agents for the pharmaceutical industry (Eichler *et al*., 1995; Lam, 1997; Lowman, 1997). The premise behind this new area of research stems from what has been learned from studying naturally occurring peptides which possess biological activities. These bioactive peptides for the large part appear to act by binding to a specific protein target (Siddle and Hutton, 1990) and in many cases it has been shown that the peptides inactivate the protein target with incredible specificity. Binding constants for the naturally occurring bioactive peptides that have been studied tend to be in the nM range (Rivier and Marshall, 1990; Smith and Rivier, 1992) with binding constants as high as 10^{-12}M having been reported (Bozou *et al*., 1986; Le-Nguyen *et al*., 1990).

Two major strategies have been employed by researchers that have been trying to engineer novel synthetic peptides. The first involves the use of chemically synthesized combinatorial peptide libraries of up to 10 amino acids in length (Eichler *et al*., 1995; Lam, 1997; Lebl and Krchnak, 1997). The second strategy involves the use of fusion phage libraries where randomly encoded peptides, which can be much longer in length, are fused to a coat protein of a filamentous phage that allows the randomized peptide to be displayed on the outside surface of the phage (Smith and Scott, 1993; Lowman, 1997). The libraries that are generated via either chemical synthesis or as fusion phage are then usually mixed with a matrix-bound protein target and peptides which can bind tightly to the protein target are selected. New peptides are generated, either based on individual peptide sequence data or collective consensus data from multiple peptides, and tested for their inhibitory potential. Peptides are initially tested

in vitro by measuring their ability to inhibit the enzymatic activity of the protein target and then promising peptides are tested further in clinical trials.

Although there is enormous potential for the development of synthetic inhibitor peptides using the approaches described above, this technology has not become a mainstay in the pharmaceutical industry due to the problem of peptide stability and the unwanted degradation of the potential peptide drug by peptidases in the host cells (Lauta, 2000). Approaches to solve this problem have included acetylating the amino terminus of the peptide, amidating the carboxyl terminus of the peptide, the use of D-amino acids or modified amino acids as opposed to the naturally occurring L-amino acids, cyclized peptides, as well as the development of enhanced delivery systems which protect the peptides from premature degradation (Sanders, 1990; Pinilla *et al*., 1995; Hruby and Balse, 2000). The purpose of this study was to investigate whether modifying peptides through the use of protein-based stabilizing motifs could circumvent the problem of peptide degradation.

There are three major classes of peptidases that can degrade peptides, the amino and carboxy exopeptidases which act at either the amino or the carboxyl terminus of the peptide, respectively, and the endopeptidases which act on the internal peptide. Aminopeptidases, carboxypeptidases, and endopeptidases have been identified in both prokaryotic and eukaryotic cells, and where they have been extensively characterized, most of these peptidases have been found to function similarly in all cell types (Rawlings and Barrett, 1993). Interestingly, when the abundance of exopeptidases has been examined in either prokaryotic or eukaryotic systems, more aminopeptidases have been identified than carboxypeptidases (Ryan, 1989; Bai and Amidon, 1992; Brownlees and Williams, 1993; Bai *et al*., 1995; Miller, 1996).

In previous work an *in vivo* genetic approach was developed by which novel bioactive peptides that inhibit the growth of *Escherichia coli* can be isolated and identified. In the initial study, two potential protein-based stabilizing motifs were discovered that could increase the stability of peptides (Walker *et al*., 2001). The first motif involved the use of a small stable protein anchor while the second motif involved the use of proline residues which have been argued by numerous researchers to be more resistant to degradation by peptidases (Walter *et al*., 1980; Yaron and Naider, 1993; Vanhoof *et al*., 1995; Cunningham and O'Connor, 1997). In this study, the *in vivo* system was utilized to investigate whether these protein-based stabilizing motifs can be used to more effectively create inhibitor peptides.

Materials and Methods

Bacterial strains and plasmids

ALS225 which is MC1061/F'*lacI*^{Q1}*Z+Y+A+* was the *E. coli* strain used in this study. The genotype for MC1061 is *araD139* Δ*(araABOIC-leu)7679* Δ*(lac)X74 galU galK rpsL hsr-hsm+* (Casadaban and Cohen, 1980). The highly regulable pLAC11 expression vector was used to make the p-Rop(C) and p(N)Rop- fusion vectors (Warren *et al*., 2000).

Media for the in vivo studies

Rich Luria-Bertani and minimal M9 media were prepared as described by Miller (1972). Ampicillin was used in rich media at a final concentration of 100 μg/mL and in minimal media at a final concentration of 50 μg/mL. IPTG was added to media at a final concentration of 1 mM.

Construction of the p-Rop(C) fusion vector

The forward primer 5' TAC TAT AGA TCT ATG ACC AAA CAG GAA AAA ACC GCC 3' and the reverse primer 5' TAT ACG TAT TCA GTT GCT CAC ATG TTC TTT CCT GCG 3' were used to PCR amplify a 558-bp DNA fragment using pBR322 as a template. This

fragment contained a *Bgl*II restriction site which was incorporated into the forward primer followed by an ATG start codon and the Rop coding region. The fragment extended beyond the Rop stop codon through the *Afl*III restriction site in pBR322. The amplified dsDNA was gel isolated, restricted with *Bgl*II and *Afl*III, and then ligated into the pLAC11 expression vector which had been digested with the same two restriction enzymes. The resulting p-Rop(C) fusion vector is 2623 bp in size.

Construction of the p(N)Rop- fusion vector

The forward primer 5' AAT TCA TAC TAT AGA TCT ATG ACC AAA CAG GAA AAA ACC GC 3' and the reverse primer 5' TAT ATA ATA CAT GTC AGA ATT CGA GGT TTT CAC CGT CAT CAC 3' were used to PCR amplify a 201-bp DNA fragment using pBR322 as a template. This fragment contained a *Bgl*II restriction site which was incorporated into the forward primer followed by an ATG start codon and the Rop coding region. The reverse primer placed an *Eco*RI restriction site just before the Rop TGA stop codon and an *Afl*III restriction site immediately after the Rop TGA stop codon. The amplified dsDNA was gel isolated, restricted with *Bgl*II and *Afl*III, and then ligated into the pLAC11 expression vector which had been digested with the same two restriction enzymes. The resulting p(N)Rop- fusion vector is 2262 bp in size.

Generating the randomized peptide libraries

All of the peptide libraries used in this study were constructed as described by Walker *et al.* (2001). To construct the randomized peptide libraries for use with the p-Rop(C) fusion vector, the oligonucleotides 5' TAC TAT AGA TCT ATG $(XXX)_{20}$ CAT AGA TCT GCG TGC TGT GAT 3' and 5' ATC ACA GCA CGC AGA TCT ATG 3' were used. After extension, the resulting dsDNA was digested with *Bgl*II and ligated into the p-Rop(C) fusion vector which had

been digested with the same restriction enzyme and subsequently dephosphorylated using alkaline phosphatase. To construct the randomized peptide libraries for use with the p(N)Rop- fusion vector, the oligonucleotides 5' TAC TAT GAA TTC $(XXX)_{20}$ GAA TTC TGC CAC CAC TAC TAT 3' and 5' ATA GTA GTG GTG GCA GAA TTC 3' were used. After extension, the resulting dsDNA was digested with *Eco*RI and ligated into the p(N)Rop- fusion vector which had been digested with the same restriction enzyme and subsequently dephosphorylated using alkaline phosphatase. To construct randomized 20-amino acid peptide libraries which contained two proline residues at the amino and carboxyl termini of the peptides, the oligonucleotides 5' TAC TAT AGA TCT ATG CCG CCG $(XXX)_{16}$ CCG CCG TAA TAA GAA TTC GTA CAT 3' and 5' ATG TAC GAA TTC TTA TTA CGG CGG 3' were used. After extension, the resulting dsDNA was digested with *Bgl*II and *Eco*RI and ligated into the pLAC11 expression vector which had been digested with the same two restriction enzymes. In the randomized oligonucleotides an X denotes that an equimolar mixture of the nucleotides A, C, G, or T was used.

Chemicals and reagents

Extension reactions were carried out using Klenow from New England Biolabs (Beverly, MA, USA) while ligation reactions were performed using T4 DNA ligase from Invitrogen (Carlsbad, CA, USA). Alkaline phosphatase (calf intestinal mucosa) from Pharmacia (Piscataway, NJ, USA) was used for dephosphorylation. Isopropyl β-D-thiogalactoside (IPTG) was obtained from Diagnostic Chemicals Limited (Prince Edward Island, Canada).

Results

Isolation and characterization of inhibitor peptides that are protected at the carboxyl terminus via fusion to the Rop protein

In initial studies with inhibitor peptides, a completely randomized oligonucleotide library was used to direct the synthesis of up to 20-amino acid peptides in the highly regulable expression vector, pLAC11 (Walker *et al*., 2001). In a screen of 20,000 peptides, 21 were found that could inhibit the growth of *E. coli*. While most of the inhibitors encoded up to 20-amino acid peptides as expected, two of the most potent inhibitors turned out to be fusion peptides in which the carboxyl terminus of the peptide had become fused to the amino terminus of the Rop protein (the *rop* gene is located downstream from where the oligonucleotides are inserted into the pLAC11 expression vector). Because the Rop protein is known to form a very stable structure (Eberle *et al*., 1991; Steif *et al*., 1993), it was postulated that the Rop protein was likely serving as a stable protein anchor which protected the carboxyl terminus of the two inhibitor peptides. Rop is a small 63-amino acid dimeric four-helix-bundle protein whose monomer consists of two antiparallel α-helices that are connected by a sharp hairpin loop. It is a dispensable part of the ColE1 replicon that is used by plasmids such as pBR322 and it can be deleted without causing any ill effect on the replication, partitioning, or copy number of plasmids that contain a ColE1 *ori* (Soberon *et al*., 1980).

To test whether peptides could be stabilized by fusing the carboxyl terminus of the peptides to the amino terminus of the Rop protein, the p-Rop(C) fusion vector, which is a derivative of pLAC11, was constructed as described in Materials and Methods. In order to isolate potential inhibitor peptides which were protected at the carboxyl terminus, a totally randomized oligonucleotide library that encoded up to 20-amino acid peptides was cloned into

the p-Rop(C) fusion vector as shown in Figure 3.1A and transformed into *E. coli* under repressed conditions. In earlier studies of 20,000 anchorless peptides, only one turned out to be a potent inhibitor that could inhibit the growth of *E. coli* for two days on plates. This frequency was used as the basis by which to judge whether protecting the carboxyl terminus of peptides with the Rop protein would increase the number of inhibitor peptides that could be isolated.

10,000 peptides that were protected at the carboxyl terminus were screened using the previously described grid-patching technique and 16 potent two-day inhibitors were isolated. To verify that all of the inhibitors were legitimate, plasmid DNA from each inhibitory clone was prepped, transformed into a fresh background, and then checked that they were still inhibitory on plates and that their inhibition was dependent on the presence of the inducer, IPTG. In order to make a more accurate assessment of how potent the inhibitors were, the first ten inhibitors were subjected to a growth rate analysis in liquid media. Cultures containing either the inhibitor to be tested or p-Rop(C) as a control were diluted 1:100 into new media and induced with 1 mM IPTG. OD_{550} readings were then taken hourly until the control culture had passed log phase. Growth rates were determined as the spectrophotometric change in OD_{550} per unit time within the logarithmic phase of growth. The inhibition of the growth rate was then calculated for the inhibitors using p-Rop(C) as a control. As indicated in Table 3.1, the peptides inhibited the bacterial growth rate at an average of 92%.

Isolation and characterization of inhibitor peptides that are protected at the amino terminus via fusion to the Rop protein

Based on the success in isolating potent inhibitor peptides that were protected at the carboxyl terminus by the Rop protein, it was decided to examine whether protecting the amino terminus of the peptides would increase the frequency of potent two-day inhibitors that could be

isolated. Additionally, when exopeptidases have been characterized in either prokaryotic or eukaryotic systems, more aminopeptidases have been identified than carboxypeptidases (Ryan, 1989; Bai and Amidon, 1992; Brownlees and Williams, 1993; Bai *et al*., 1995; Miller, 1996) and thus an argument could be made that stabilizing the amino terminus of a peptide would be more effective at preventing degradation by peptidases than stabilizing the carboxyl terminus of the peptide.

To test whether peptides could be stabilized by fusing the amino terminus of the peptides to the carboxyl terminus of the Rop protein, the p(N)Rop- fusion vector, which is a derivative of pLAC11, was constructed as described in Materials and Methods. In order to isolate potential inhibitor peptides which were protected at the amino terminus, a totally randomized oligonucleotide library that encoded up to 20-amino acid peptides was cloned into the p(N)Rop- fusion vector as shown in Figure 3.1B and transformed into *E. coli* under repressed conditions. 6,000 of these Rop fusion peptides were screened using the grid-patching technique and 14 two-day inhibitors were isolated. The inhibitors were verified as described in the previous section and the first 10 inhibitors were subjected to growth rate analysis using p(N)Rop- as a control in order to determine their potency. As indicated in Table 3.1, the peptides inhibited the bacterial growth rate at an average of 93%.

Discussion

In the initial studies of novel bioactive peptides that inhibited the growth of *E. coli*, two protective motifs were discovered that might enhance the stability of peptides, a small stable protein anchor such as Rop, or the incorporation of terminal proline residues (Walker *et al*., 2001). In this study it was observed that if either the amino or carboxyl terminus of the peptide was protected by fusing the peptide to the highly stable Rop protein anchor, the frequency at

which potent inhibitors could be isolated increased by as much as 47-fold (Table 3.2). In Walker *et al*. 2003 it was shown that the incorporation of two proline residues at the termini of the peptides also significantly increased the frequency at which potent inhibitor peptides could be isolated, however, the resulting peptides were not as potent as those that could be isolated using the Rop protein anchor. The simplest explanation for why the protection of the termini of peptides via fusion to a small stable protein or the incorporation of two proline residues increases the frequency and potency of the resulting bioactive peptides is that this prevents the action of exopeptidases which degrade the termini of peptides. The fact that Rop appears to be a better stabilizing anchor than two proline residues could be explained simply by their size difference. The significantly larger Rop protein may very well act by hindering the ability of peptidases to gain access to the protected peptide.

The use of APP or a small stable protein anchor such as Rop represents unique possibilities for the design of more effective peptide drugs that are not as susceptible to degradation by peptidases. Because the amino and carboxypeptidases that have been characterized appear to function quite similarly in all cell types (Rawlings and Barrett, 1993), incorporating protein-based stabilizing anchors into peptide drugs should significantly decrease the action of exopeptidases in the host cell. Additionally, the use of protein-based anchors has another distinct advantage over existing technologies that can be used to increase peptide stability such as amidation, acetylation, or the use of modified or D-amino acids that can only be utilized using chemical synthesis. Peptide drugs that are stabilized using APP or a small stable protein anchor could be produced using recombinant protein synthesis in *E. coli* or human cells.

References

Bai, J. P. F. and Amidon, G. L. (1992). Structural specificity of mucosal cell transport and metabolism of amino acids and peptide drugs: implication for oral peptide delivery. Pharm Res 9, 969-978.

Bai, J. P. F., Chang, L. -L., and Guo, J. -H. (1995). Targeting of peptides and protein drugs to specific sites in the oral route. Crit Rev Ther Drug 12, 339-371.

Bozou, J. C., Amar, S., Vincent, J. P., and Kitabgi, P. (1986). Neurotensin-mediated inhibition of cyclic AMP formation in neuroblastoma N1E115 cells: involvement of the inhibitory GTP-binding component of adenylate cyclase. Mol Pharmacol 29, 489-496.

Brownlees, J. and Williams, C. H. (1993). Peptidases, peptides at the mammalian blood-brain barrier. J Neurochem 60, 793-803.

Casadaban, M.J. and Cohen, S.N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J Mol Biol 138, 179-207.

Eberle, W., Pastore, A., Sander, C., and Rosch, P. (1991). The structure of ColE1 Rop in solution. J Biomol NMR 1, 71-82.

Eichler, J., Appel, J. R., Blondelle, S. E., Dooley, C. T., Dorner, B., Ostresh, J. M., Perez-Paya, E., Pinilla, C. and Houghten, R. A. (1995). Peptide, peptidomimetic, and organic synthetic combinatorial libraries. Med Res Rev 15, 481-496.

Goldberg, A. L., Moerschell, R. P., Chung, C. H., and Maurizi, M. R. (1994). ATP-dependent protease La(Lon) from *Escherichia coli*. Methods Enzymol 244, 350-375.

Guyer, M. S., Reed, R. R., Steitz, J. A., and Low, K. B. (1980). Identification of a sex-factor-affinity site in *E. coli* as gd. Cold Spring Harbor Symp Quant Biol 45, 135-140.

Hruby, V. J. and Balse, P. M. (2000). Conformational and topographical considerations in designing agonist peptidomimetics from peptide leads. Curr Med Chem 7, 945-970.

Lam, K. S. (1997). Application of combinatorial library methods in cancer research and drug discovery. Anticancer Drug Des 12, 145-167.

Lauta, V. M. (2000). Pharmacological elements in clinical application of synthetic peptides. Fundam Clin Pharmacol 14, 425-442.

Le-Nguyen, D., Heitz, A., Chiche, L., Castro, B., Boigegrain, R. A., Favel, A., and Coletti-Previero, M. A. (1990). Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. Biochimie 72, 431-435.

Lebl, M. and Krchnak, V. (1997). Synthetic peptide libraries. Methods Enzymol 289, 336-392.

Lowman, H. B. (1997). Bacteriophage display and discovery of peptide leads for drug development. Annu Rev Biophys Biomol Struct 26, 401-424.

Miller, C. G. (1996). Protein Degradation and Proteolytic Modification. In *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*, 2nd Edition (Neidhardt, F.C., Ed.), ASM Press, Washington, D.C., pp-938-954.

Miller, J. H. (1972). *Experiments in Molecular Genetics,* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Pinilla, C., Appel, J., Blondelle, S., Dooley, C., Dorner, B., Eichler, J., Ostresh, J., and Houghten, R. A. (1995). A review of the utility of soluble peptide combinatorial libraries. Biopolymers 37, 221-240.

Porankiewicz, J., Wang, J., and Clarke, A. K. (1999). New insights into the ATP-dependent Clp protease: *Escherichia coli* and beyond. Mol Microbiol 32, 449-458.

Ramachandran, G. N. and Mitra, A. K. (1976). An explanation for the rare occurrence of *cis* peptide units in proteins and polypeptides. J Mol Biol 107, 85-92.

Rawlings, N. D. and Barrett, A. J. (1993). Evolutionary families of peptidases. Biochem J 290, 205-218.

Rivier, J. E. and Marshall, G. R., eds. (1990). *Peptides: Chemistry, Structure, and Biology,* ESCOM Science Publishers BV, The Netherlands.

Ryan, J. W. (1989). Peptidase enzymes of the pulmonary vascular surface. Am J Physiol 257, L53-60.

Sanders, L. M. (1990). Drug delivery systems and routes of administration of peptide and protein drugs. Eur J Drug Metab Pharmacokinet 15, 95-102.

Shagger, H. and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 166, 368-379.

Siddle, K. and Hutton, J. C., eds. (1990). *Peptide Hormone Action,* IRL Press at Oxford University Press, Oxford, England.

Smith, G. P. and Scott, J. K. (1993). Libraries of peptides and proteins displayed on filamentous phage. Methods Enzymol 217, 228-257.

Smith, J. A. and Rivier, J. E., eds. (1992). *Peptides: Chemistry and Biology,* ESCOM Science Publishers BV, The Netherlands.

Soberon, X., Covarrubias, L., and Bolivar, F. (1980). Construction and characterization of new cloning vehicles: IV. Deletion derivatives of pBR322 and pBR325. Gene 9, 287-305.

Steif, C., Weber, P., Hinz, H. -J., Flossdorf, J., Cesareni, G., and Kokkinidis, M. (1993). Subunit interactions provide a significant contribution to the stability of the dimeric four-α-helical-bundle protein Rop. Biochemistry 32, 3867-3876.

Vanhoof, G., Goossens, F., De Meester, I., Hendriks, D., and Scharpe, S. (1995). Proline motifs in peptides and their biological processing. FASEB J 9, 736-744.

Walker, J. R., Roth, J. R., and Altman, E. (2001). An *in vivo* study of novel bioactive peptides that inhibit the growth of *Escherichia coli*. J Peptide Res 58, 380-388.

Walter, R., Simmons, W. H., and Yoshimoto, T. (1980). Proline specific endo- and exopeptidases. Mol Cell Biochem 30, 111-127.

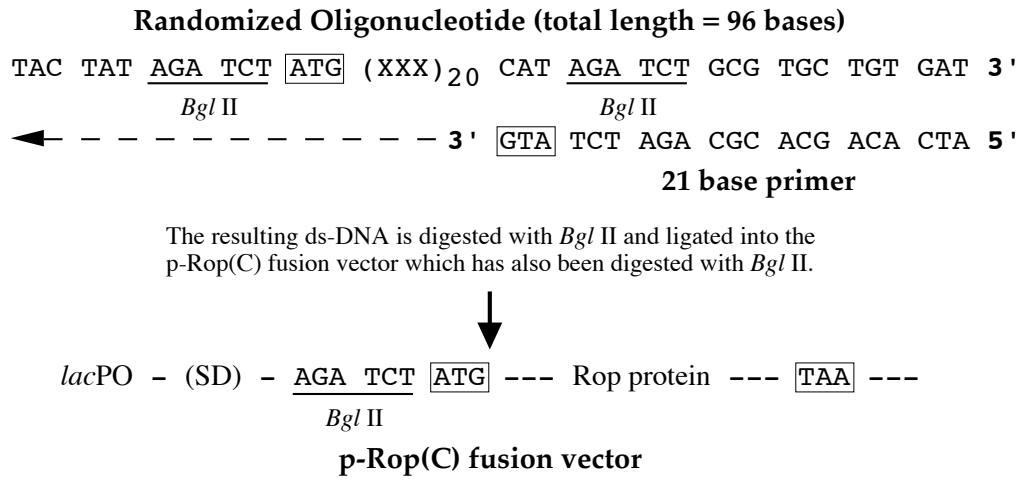
Warren, J. W., Walker, J. R., Roth, J. R., and Altman, E. (2000). Construction of a highly regulable expression vector, pLAC11, and its multipurpose derivatives, pLAC22 and pLAC33. Plasmid 44, 138-151.

Woodley, J. F. (1994). Enzymatic barriers for GI peptide and protein delivery. Crit Rev Ther Drug Carrier Sys 11, 61-95.

Yao, J., Feher, V. A., Espejo, F., Reymond, M. T., Wright, P. E., and Dyson, H. J. (1994). Stabilization of a type VI turn in a family of linear peptides in water solution. J Mol Biol 243, 736-753.

Yaron, A. and Naider, F. (1993). Proline-dependent structural and biological properties of peptides and proteins. Crit Rev Biochem Mol Biol 28, 31-81.

A. p-Rop(C) Inhibitors



B. p(N)Rop- Inhibitors

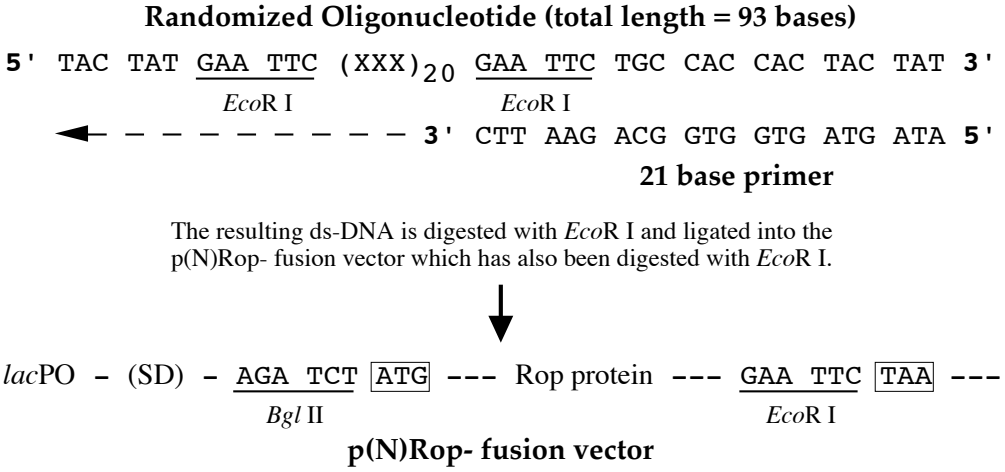


Figure 3.1 Scheme for generating p-Rop(C) and p(N)Rop- inhibitor peptide libraries. Wherever an X occurs in the oligonucleotide, a random mixture of all four of the nucleotides, A, C, G, and T, was used. (A) p-Rop(C) peptide library where the carboxyl terminus of the peptides are fused to the amino terminus of the Rop protein. Because of the way the oligonucleotide library has been engineered, either orientation of the incoming digested dsDNA fragment results in a fusion product. (B) p(N)Rop- peptide library where the amino terminus of the peptides are fused to the carboxyl terminus of the Rop protein. Because of the way the oligonucleotide library has been engineered, either orientation of the incoming digested dsDNA fragment results in a fusion product.

Table 3.1 Inhibitory effects of the p-Rop(C) and p(N)Rop- peptide inhibitors

Inhibitor	% Inhibition of the growth rate
p-Rop(C)1	87
p-Rop(C)2	99
p-Rop(C)3	85
p-Rop(C)4	98
p-Rop(C)5	95
p-Rop(C)6	99
p-Rop(C)7	91
p-Rop(C)8	86
p-Rop(C)9	93
p-Rop(C)10	91
p(N)Rop-1	81
p(N)Rop-2	96
p(N)Rop-3	95
p(N)Rop-4	92
p(N)Rop-5	99
p(N)Rop-6	93
p(N)Rop-7	87
p(N)Rop-8	91
p(N)Rop-9	95
p(N)Rop-10	96

p-Rop(C) inhibitors are stabilized by fusing the carboxyl terminus of the peptide to the amino terminus of the Rop protein. p(N)Rop- inhibitors are stabilized by fusing the amino terminus of the peptide to the carboxyl terminus of the Rop protein. The inhibitory effects were determined as described in the text using p-Rop(C) or p(N)Rop- as a control. The data is the average of duplicate experiments.

Table 3.2 Summary of the frequency and potency at which the different types of inhibitor peptides can be isolated

Type of inhibitor peptide	Frequency at which a potent 2-day inhibitor peptide can be isolated	Average inhibition (%)	Average potency (%)[a]
anchorless	1 in 20,000	25	65
protected at the C-terminus via Rop	1 in 625	92	98
protected at the N-terminus via Rop	1 in 429	93	98

[a]The average potency was calculated as previously described by Walker *et al*., 2001.

CHAPTER 4

USE OF CHARGED AND HELICAL MOTIFS TO STABILIZE BIOACTIVE INHIBITOR PEPTIDES *IN VIVO*

Abstract

Using an *in vivo* genetic approach which we developed to screen for novel bioactive peptides that can inhibit the growth of *Escherichia coli*, we found that protecting randomized peptides via a terminal opposite charge motif, or the incorporation of hydrophilic α-helix-forming amino acids, increased the frequency at which potent bioactive inhibitor peptides could be isolated. Secondary structure predictive algorithms and CD spectroscopy were used to examine the structural nature of the putative α-helical inhibitor peptides.

Introduction

The completion of the human genome project has revitalized interest in the utilization of polypeptides for the treatment of previously incurable diseases. The discovery of bioactive peptides implicated in various biological functions has also increased interest in the potential therapeutic use of these 2 to 50-amino acid molecules. Peptides have a broad range of potential clinical benefits, with possible applications in some of the most prevalent public health categories such as bacterial, fungal or viral infections (Lawless *et al*., 1996; Bulet *et al*., 2004), cardiovascular disease (Boerrigter and Burnett, 2004), neurological disorders (Permanne *et al*., 2002; Adessi *et al*., 2003), and cancer (Hetian *et al*., 2002; Liu *et al*., 2004). Indeed, peptides of various size and derivation are already in wide use as drugs and a number of promising novel candidates for therapeutic use are currently in development for diseases such as AIDS, Alzheimer's, malaria, and various forms of cancer (Lien and Lowman, 2003).

Over the past two decades, novel synthetic bioactive peptides have been primarily engineered by two *in vitro* approaches. In the first approach a randomized library of 6-10 amino acid peptides is prepared by one of several combinatorial synthetic techniques (Eichler *et al.*, 1995; Schultz and Schultz, 1996; Al-Obeidi *et al*., 1998). The second approach involves the cloning of a randomized oligonucleotide library into a coat protein gene of a filamentous phage that allows peptides 6-38 amino acids in length to be expressed on the surface of the phage particles (Cwirla *et al*., 1990; Devlin *et al*., 1990; Smith and Scott, 1993; Sidhu *et al*., 2000). The peptide libraries obtained via these methods are typically mixed with a matrix-bound protein target, and bound peptides are subsequently eluted and sequenced.

An *in vivo* genetic system for screening and isolating novel bioactive peptides that inhibit the growth of *Escherichia coli* has been developed. Two different motifs were characterized

which seem to offer protection from enzymatic degradation by endogenous proteases and peptidases, as evidenced by the fact that peptides with these motifs incorporated into their structure are isolated much more frequently from the *in vivo* screen than those with no stabilizing motif (Walker *et al*., 2001; Walker *et al*., 2003). The high degree of success observed when the highly stable helix-turn-helix Rop protein was used as a protective anchor for inhibitor peptides, led to the question: Are there other peptides that exhibit stable secondary structures that might be adapted to serve as protective motifs? Upon examination of the structures of the hundreds of naturally occurring bioactive peptides, it was determined that there were some common motifs that could be utilized in the *in vivo* system. Antimicrobial peptides are found in many different organisms across the taxonomic kingdoms of life, but one thing most of these molecules have in common is a cationic amphipathic α-helical structure. Peptides such as maganin, mastoparan and melittin exhibit this secondary structure and are potent antibacterial compounds (Bechinger *et al*., 1993; Cachia *et al*., 1986; Terwilliger and Eisenberg, 1982). A number of peptide hormones utilized therapeutically also possess observed α-helical structure, including glucagon, endorphin and secretin (Bedarkar *et al*., 1977; Blanc *et al*., 1983; Bodanszky and Bodansky, 1986). Another intriguing structure is the terminal opposite charge motif which is utilized in nature by the conantokin subfamily of the conotoxins which exhibits a looped structure stabilized by amino acid residues of opposite charge at their termini (Olivera *et al*., 1991).

It was decided to attempt to incorporate a randomized α-helical structure into the *in vivo* genetic screen through the use of a specific subset of amino acids that have helix-forming tendencies and are hydrophilic to allow for high solubility in aqueous solutions. An attempt was also made to mimic the terminal opposite charge motif in our genetic system by deliberately

designing oligonucleotides that would encode for negative amino acids at the amino terminus, and positive residues at the carboxyl terminus of an otherwise randomized peptide (Figure 4.1).

Materials and Methods

Media for the in vivo studies

Rich Luria-Bertani and minimal M9 media used in this study were prepared as described by Miller (1972). Ampicillin was used in rich media at a final concentration of 100 μg/mL and in minimal media at a final concentration of 50 μg/mL. Isopropyl β-D-thiogalactoside (IPTG) was added to media at a final concentration of 1 mM.

Bacterial strains and plasmids

ALS225 which is MC1061/F'*lacI*^{Q1}*Z+Y+A+* was the *E. coli* strain used in this study. The genotype for MC1061 is *araD139* Δ*(araABOIC-leu)7679* Δ*(lac)X74 galU galK rpsL hsr-hsm+* (Casadaban and Cohen, 1980). pLAC11 is the highly regulable expression vector that was used in this work (Warren *et al*., 2000).

Generating the randomized opposite charge peptide library

To construct randomized 24-amino acid peptides based on the opposite charge motif, the oligonucleotides 5'-TAC TAT AGA TCT ATG GAA GAC GAA GAC $(XXX)_{16}$ CGT AAA CGT AAA TAA TAA GAA TTC GTA CAT-3' and 5'-GCA TTT GCA TTT ATT ATT CTT AAG CAT GTA-3' were used. After extension the resulting dsDNA was digested with *Bgl*II and *Eco*RI and ligated into the pLAC11 expression vector that had been digested with the same two restriction enzymes.

Generating the randomized hydrophilic α-helical peptide library

To construct randomized 18-amino acid peptide libraries based on the α-helical motif, the oligonucleotides 5'-TAC TAT AGA TCT ATG $(VAN)_{17}$ TAA TAA GAA TTC TGC CAG CAC

TAT-3' and 5'-ATA GTG CTG GCA GAA TTC TTA TTA-3' were used. After extension the resulting dsDNA was digested with *Bgl*II and *Eco*RI and ligated into the pLAC11 expression vector that had been digested with the same two restriction enzymes. In the randomized oligonucleotides an N denotes that an equimolar mixture of the nucleotides A, C, G, or T was used, while a V denotes that an equimolar mixture of the nucleotides A, C, or G was used.

Sequencing the coding regions of the inhibitor peptide clones

The forward primer 5'-TCA TTA ATG CAG CTG GCA CG-3' and the reverse primer 5'-TTC ATA CAC GGT GCC TGA CT-3' were used to sequence both strands of the inhibitor peptide clones from plasmid DNA isolated via midiprep. If an error-free consensus sequence could not be obtained from these two sequencing runs, both strands of the inhibitor peptide clone in question were resequenced using the forward primer 5'-TAG CTC ACT CAT TAG GCA CC-3' and the reverse primer 5'-GAT GAC GAT GAG CGC ATT GT-3'. The second set of primers were designed to anneal downstream of the first set in the pLAC11 vector.

Secondary structure prediction

The propensity of the putative helical inhibitory peptides to form α-helices was initially analyzed using prediction algorithms including those developed by Chou and Fasman and Garnier, Osguthorpe, and Robson. Peptide sequences were also analyzed using the multiple algorithms of the NPS@ Consensus Secondary Structure Prediction website (Combet *et al*., 2000).

Circular dichroism

CD measurements were made using a Jasco-710 spectropolarimeter interfaced and controlled by a computer. The measurements were made using a circular quartz cell of 0.1-cm path length. All peptide samples were at a concentration of 100 μM in 20 mM sodium phosphate

buffer (pH 6.0) with 2,2,2-trifluoroethanol (TFE) added to 50%. Measurements were made over a range of 250-190 nm, a scan speed of 20 nm/min with a sensitivity of 50 millidegrees, a resolution of 0.2 nm, and a band width of 1.0 nm. Five scans were accumulated and averaged, and all dichroic spectra were corrected by subtraction of the background and then smoothed via an internal algorithm in the Jasco software package, J-710 for Windows. The CD spectra are reported as the mean residue molar ellipticity ([θ]) in degrees•cm^2•$dmol^{-1}$. Secondary structure was estimated by utilizing the MRE value at 222 nm. Taking peptide length into account, percent α-helix = 100[MRE]/39,500(1 – 2.57/*n*), where *n* is the number of amino acid residues (Scholtz *et al*., 1991).

Synthetic peptides

Synthetic peptides (2B and 115B) were obtained from Sigma-Genosys (The Woodlands, TX, USA) based on the primary sequences determined from the five putative helical clones and the data from the secondary structure prediction.

Chemicals and reagents

Extension reactions were carried out using Klenow polymerase from New England Biolabs (Beverly, MA, USA) while ligation reactions were performed using T4 DNA ligase from Invitrogen (Carlsbad, CA, USA). Isopropyl β-D-thiogalactoside (IPTG) was obtained from Diagnostic Chemicals Limited (Prince Edward Island, Canada). 2,2,2-trifluoroethanol (TFE) was obtained from Sigma-Aldrich Company (St. Louis, MO, USA).

Results

Isolation and characterization of inhibitor peptides that are protected by a terminal opposite charge motif

In order to isolate potential inhibitor peptides with an opposite charge motif consisting of four negatively charged amino acid residues at the amino terminus and four positively charged residues at the carboxyl terminus, a randomized oligonucleotide library that encoded up to 24-amino acid peptides was cloned into the pLAC11 vector as shown in Figure 4.2A and transformed into *E. coli* under repressed conditions. Since most of the routinely employed expression vectors produce significant amounts of protein from cloned genes even when grown under repressed conditions, there was concern that it might not be possible to clone potent inhibitor peptides unless they could be isolated under completely repressed conditions. For this reason, the highly regulable pLAC11 expression vector was used in these studies (Warren *et al*., 2000). Since encoded peptides that are expressed using pLAC11 are under the control of the wild-type *lac* operon, their expression can be turned on or off by the presence or absence of the gratuitous inducer IPTG. Using the grid-patching technique described in Walker *et al*., 2001, in which the clones are patched onto both rich repressing plates and minimal inducer plates, 20,000 potential candidates were screened. Six potent IPTG-dependent inhibitors were isolated in this screen. The inhibitors were verified by preparing plasmid DNA from the clones and retransforming into ALS225 to verify that they still demonstrated inhibition. The potency of the inhibitors was then determined via growth rate analysis as described in Walker *et al*., 2001. As indicated in Table 4.1, the terminal opposite charge peptides inhibited the bacterial growth rate at an average of 50% with an average potency of 88% (Table 4.7).

In order to continue analysis of the inhibitors, we next wanted to verify that all of the inhibitor candidates contained 78-bp inserts as expected (Figure 4.2A). Although, five out of the six did, one of the most potent clones was found to have been terminated before the double TAATAA termination site that was incorporated into the oligonucleotide (Table 4.2).

Interestingly, this truncated peptide is negatively charged over the first half of its length and contains two positively charged amino acids at its carboxyl terminus, and is the second best inhibitor of the group.

Isolation and characterization of inhibitor peptides that are protected by an overall α-helical motif

The α-helical structure is used by many of the naturally occurring bioactive peptides in order to achieve a stable structure. It was rationalized that because of the nature of the genetic code and the arrangement of its codons, it should be possible to create a motif that would yield randomized 18-amino acid hydrophilic peptide libraries with a propensity to form α-helices. Since in most α-helices there are 3.6 amino acid residues per complete turn, the 18-amino acid length was chosen in order to generate α-helical peptides which contained five complete turns (Creighton, 1993). The use of hydrophilic amino acids would yield peptides that should be soluble in the cellular cytosol. As shown in Table 4.3, the use of a [(CAG)A(TCAG)] codon mixture yields the hydrophilic amino acids histidine, glutamine, lysine, aspartic acid, and glutamic acid which are α-helix formers, as well as asparagine which is a weak α-helix breaker. If this codon mixture were used to randomly build an α-helical peptide, asparagine would be expected to occur no more than every sixth position which is acceptable in an α-helical structure according to either Chou-Fasman or Garnier-Osguthorpe-Robson prediction rules. Additionally, several well-characterized proteins have been observed to contain up to three weak α-helical breaker amino acids within a similarly sized α-helical region of the protein (Creighton, 1993).

In order to isolate potential inhibitor peptides with an overall α-helical motif, a randomized oligonucleotide library that encoded up to 18-amino acid peptides was cloned into the pLAC11 vector as shown in Figure 4.2B and transformed into *E. coli* under repressed

conditions. Using the grid-patching technique, 12,000 potential candidates were screened. Five potent IPTG-dependent inhibitors were isolated in this screen and verified as described above. As indicated in Table 4.4, the peptides inhibited bacterial growth at levels that averaged 42% and had an overall potency of 83% (Table 4.7). Sequence analysis of the coding regions for the five inhibitors is shown in Table 4.5.

Secondary structure prediction and analysis of putative α-helical inhibitor peptides

Predictive information on the structure and function of a polypeptide of known sequence can be obtained through use of appropriate algorithms. We employed several of the most commonly used secondary structure prediction algorithms to examine the sequences of our five putative α-helical peptides as well as sequences of seven representative peptides whose 3D structures have been solved via X-ray crystallography or NMR (Table 4.6). Based on the consensus predictions of our five putative helical peptides, we chose to examine two of them further via the use of circular dichroism spectroscopy in an attempt to determine the α-helical content relative to the algorithmic predictions. Peptides 2B and 115B were synthesized and CD spectra was obtained as described in Materials and Methods (Figure 4.4). Although consensus algorithms predicted helical content of 50% and 40.9% for peptides 2B and 115B, respectively, the CD spectra of these peptides demonstrated just over 1% and 2% helical content, respectively. Interestingly, the peptide (2B) with the highest predicted helical content was also the one that demonstrated the greatest inhibition of bacterial growth.

Discussion

In previous studies of novel bioactive peptides that inhibited the growth of *E. coli*, two highly protective motifs were discovered that dramatically increased the frequency at which inhibitor peptides were isolated versus the "anchorless" approach; a proline dipeptide cap at

either the amino or carboxyl terminus of the peptide, or a fusion of the highly stable, small protein Rop at either terminus (Walker *et al*., 2003). In this study, the implementation of two new motifs in the *in vivo* genetic screen was investigated, a terminal opposite charge motif consisting of four negatively-charged amino terminal residues and four positively-charged carboxyl terminal residues bracketing a randomized 16-amino acid region, and an overall hydrophilic α-helical structure based on the random incorporation of His, Gln, Asn, Lys, Asp, and Glu residues into an 18-amino acid peptide. As shown in Table 4.7, both of these strategies demonstrated positive results by allowing the isolation of inhibitor peptides at a frequency 6 to 8 times higher than that of the unprotected peptide screen. And while the frequency at which inhibitors were isolated was not nearly as great as when the Rop and proline motifs were used in the screen, the inhibitors that were isolated had a potency that was nearly as effective as the proline protected peptides.

Despite the fact that the putative α-helical peptide inhibitors did not demonstrate significant helical content as determined by CD analysis, there must be some mechanism intrinsic to this particular motif that helped to stabilize peptides or inhibitors of bacterial growth would not have been isolated at frequencies so far above those of the anchorless peptides. It may be beneficial to alter the codon selection of the α-helical motif strategy to include T at the second position, as in [(CAG)(AT)(CAGT)], as this would have the effect of incorporating the strong helix-forming amino acid, leucine, into the randomized peptides as well as lending them some hydrophobic character, which may in fact be preferable as the majority of the α-helical antimicrobial peptides are in fact cationic amphipathic helices (Epand and Vogel, 1999).

References

Adessi, C., Frossard, M.J., Boissard, C., Fraga, S., Bieler, S., Ruckle, T., Vilbois, F., Robinson, S.M., Mutter, M., Banks, W.A., and Soto, C. (2003). Pharmacological profiles of peptide drug candidates for the treatment of Alzheimer's disease. J Biol Chem 278, 13905-13911.

Al-Obeidi, F., Hruby, V.J., and Sawyer, T.K. (1998). Peptide and peptidomimetic libraries. Molecular diversity and drug design. Mol Biotechnol 9, 205-223.

Bechinger, B., Zasloff, M., and Opella, S.J. (1993). Structure and orientation of the antibiotic peptide magainin in membranes by solid-state nuclear magnetic resonance spectroscopy. Protein Sci 2, 2077-2084.

Boerrigter, G. and Burnett, J.C. Jr. (2004). Recent advances in natriuretic peptides in congestive heart failure. Expert Opin Investig Drugs 13, 643-652.

Bozou, J.C., Amar, S., Vincent, J.P., and Kitabgi, P. (1986). Neurotensin-mediated inhibition of cyclic AMP formation in neuroblastoma N1E115 cells: involvement of the inhibitory GTP-binding component of adenylate cyclase. Mol Pharmacol 29, 489-496.

Bulet, P., Stocklin, R., and Menin, L. (2004). Anti-microbial peptides: from invertebrates to vertebrates. Immunol Rev 198, 169-184.

Cachia, P.J., Van Eyk, J., Ingraham, R.H., McCubbin, W.D., Kay, C.M., and Hodges, R.S. (1986). Calmodulin and troponin C: a comparative study of the interaction of mastoparan and troponin I inhibitory peptide [104-115]. Biochemistry 25, 3553-3562.

Casadaban, M.J. and Cohen, S.N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J Mol Biol 138, 179-207.

Chou, P.Y. (1990). Prediction of protein structural classes from amino acid compostitions. In Prediction of protein structure and the principles of protein conformation. Ed. Fasman, G.D. (Plenum Press: New York), pp. 549-586

Chou, P.Y. and Fasman, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol 47, 45-148.

Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000). NPS@: network protein sequence analysis. Trends Biochem Sci 25, 147-150.

Creighton, T.E. (1993). Conformational properties of polypeptide chains. In Proteins: structures and molecular properties. (W.H. Freeman and Company: New York), pp. 182-186.

Cwirla, S.E., Peters, E.A., Barrett, R.W., and Dower, W.J. (1990). Peptides on phage: a vast library of peptides for identifying ligands. Proc Natl Acad Sci USA 87, 6378-6382.

Devlin, J.J., Panganiban, L.C., and Devlin, P.E. (1990). Random peptide libraries: a source of specific protein binding molecules. Science 249, 404-406.

Eichler, J., Appel, J.R., Blondelle, S.E., Dooley, C.T., Dorner, B., Ostresh, J.M., Perez-Paya, E., Pinilla, C., and Houghten, R.A. (1995). Peptide, peptidomimetic, and organic synthetic combinatorial libraries. Med Res Rev 15, 481-496.

Epand, R.M. and Vogel, H.J. (1999). Diversity of antimicrobial peptides and their mechanisms of action. Biochim Biophys Acta 1462, 11-28.

Garnier, J., Osguthorpe, D.J., and Robson, B. (1978). Analysis of the accuracy and implications of simple method for predicting the secondary structure of globular proteins. J Mol Biol 120, 97-120.

Hetian, L., Ping, A., Shumei, S., Xiaoying, L., Luowen, H., Jian, W., Lin, M., Meisheng, L., Junshan, Y., and Chengchao, S. (2002). A novel peptide isolated from a phage display library

inhibits tumor growth and metastasis by blocking the binding of vascular endothelial growth factor to its kinase domain receptor. J Biol Chem 277, 43137-43142.

Lawless, M.K., Barney, S., Guthrie, K.I., Bucy, T.B., Petteway, S.R. Jr., and Merutka, G. (1996). HIV-1 membrane fusion mechanism: structural studies of the interactions between biologically-active peptides from gp41. Biochemistry 35, 13697-13708.

Le-Nguyen, D., Heitz, A., Chiche, L., Castro, B., Boigegrain, R.A., Favel, A., and Coletti-Previero, M.A. (1990). Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. Biochemie 72, 431-435.

Lien, S. and Lowman, H.B. (2003). Therapeutic peptides. Trends Biotechnol 21, 556-562.

Liu, R., Hsieh, C.Y., and Lam, K.S. (2004). New approaches in identifying drugs to inactivate oncogene products. Semin Cancer Biol 14, 13-21.

Miller, S., Janin, J., Lesk, A.M., and Chothia, C. (1987). Interior and surface of monomeric proteins. J Mol Biol 196, 641-656.

Olivera, B.M., Rivier, J., Scott, J.K., Hillyard, D.R., and Cruz, L.J. (1991). Conotoxins. J Biol Chem 266, 22067-22070.

O'Neill, K.T. and DeGrado, W.F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. Science 250, 646-651.

Permanne, B., Adessi, C., Saborio, G.P., Fraga, S., Frossard, M.J., Van Dorpe, J., Dewachter, I., Banks, W.A., Van Leuven, F., and Soto, C. (2002). Reduction of amyloid load and cerebral damage in a transgenic mouse model of Alzheimer's disease by treatment with a β-sheet breaker peptide. FASEB J 16, 860-862.

Roseman, M.A. (1988). Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. J Mol Biol 200, 513-522.

Scholtz, J.M., Qian, H., York, E.J., Stewart, J.M., and Baldwin, R.L. (1991). Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. Biopolymers 31, 1463-1470.

Schultz, J.S. and Schultz, J.S. (1996). The combinatorial library: a multifunctional resource. Biotechnol Prog 12, 729-743.

Siddle, K. and Hutton, J.C. (1990). Peptide Hormone Action. (IRL Press: Oxford).

Sidhu, S.S., Lowman, H.B., Cunningham, B.C., and Wells, J.A. (2000). Phage display for selection of novel binding peptides. Methods Enzymol 328, 333-363.

Smith, G.P. and Scott, J.K. (1993). Libraries of peptides and proteins displayed on filamentous phage. Methods Enzymol 217, 228-257.

Terwilliger, T.C. and Eisenberg, D. (1982). The structure of melittin. I. Structure determination and partial refinement. J Biol Chem 257, 6010-6015.

Walker, J.R., Altman, R.K., Warren, J.W., and Altman, E. (2003). Using protein-based motifs to stabilize peptides. J Peptide Res 62, 214-226.

Walker, J.R., Roth, J.R., and Altman, E. (2001). An *in vivo* study of novel bioactive peptides that inhibit the growth of *Escherichia coli*. J Peptide Res 58, 380-388.

Warren, J.W., Walker, J.R., Roth, J.R., and Altman, E. (2000). Construction of a highly regulable expression vector, pLAC11, and its multipurpose derivatives, pLAC22 and pLAC33. Plasmid 44, 138-151.

Wolfenden, R., Andersson, L., Cullis, P.M., and Southgate, C.C. (1981). Affinities of amino acid side chains for solvent water. Biochemistry 20, 849-855.

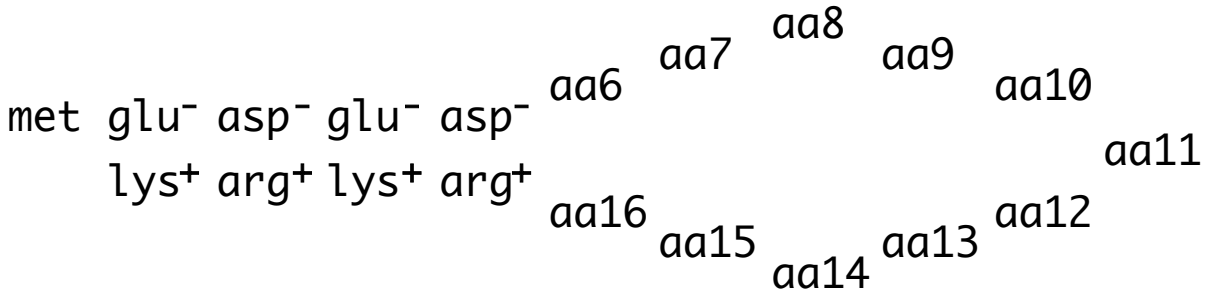


Figure 4.1 Theoretical structure for terminal opposite charge motif inhibitor peptides

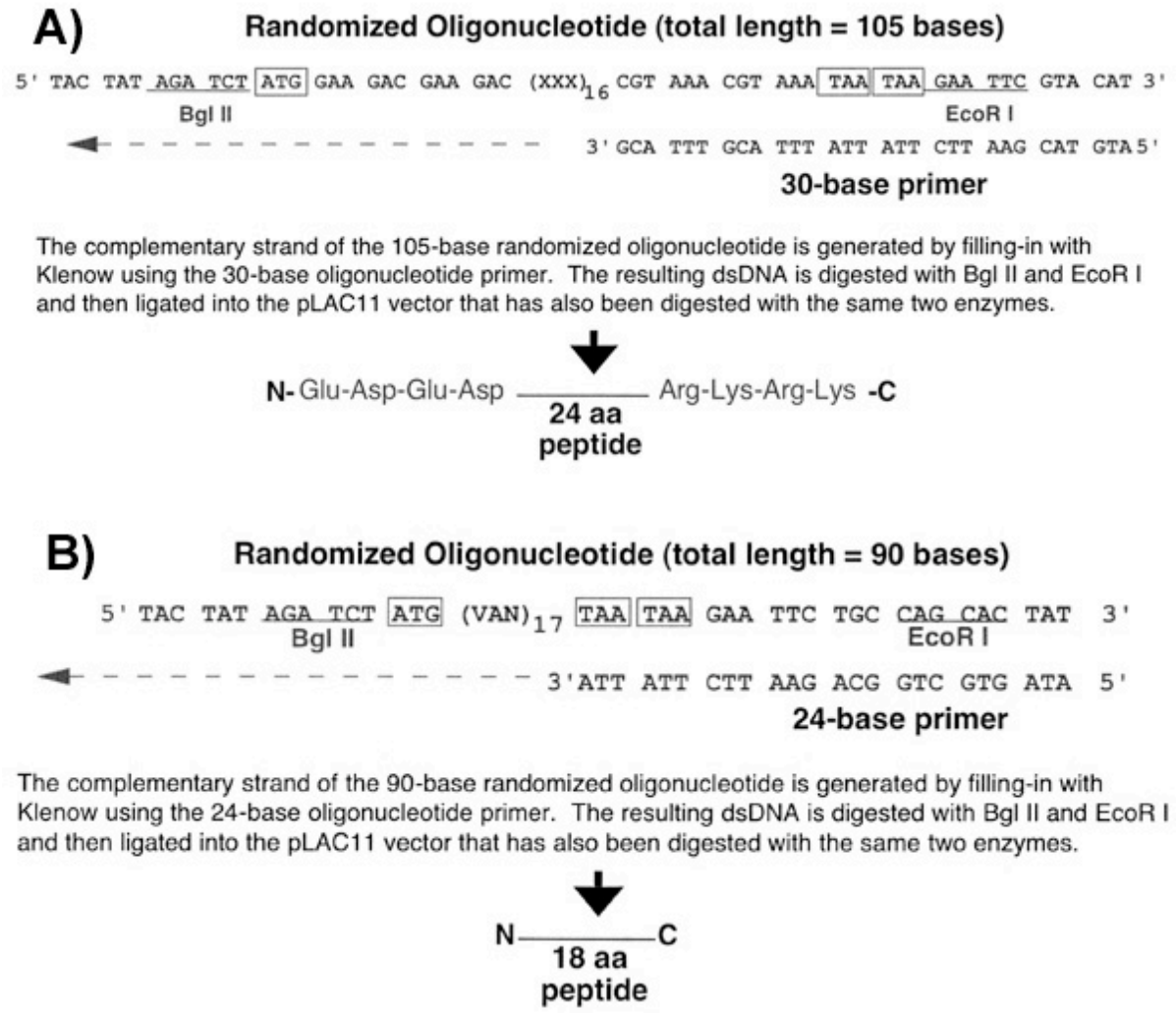


Figure 4.2 Scheme for generating p-Rop(C) and p(N)Rop- inhibitor peptide libraries.

Wherever an X or N occurs in the oligonucleotide, a random mixture of all four of the nucleotides, A, C, G, and T, was used.

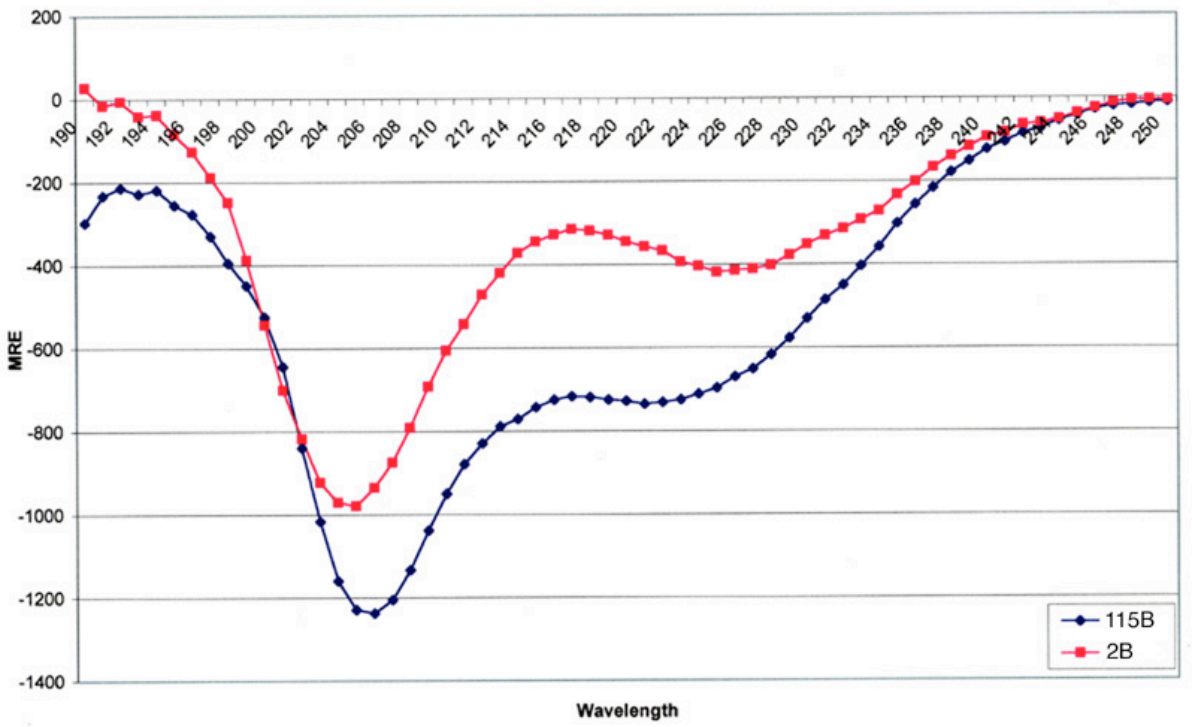


Figure 4.3 CD spectra of top two putative α-helical inhibitor peptides. Peptide concentration 100 μM in 20 mM sodium phosphate buffer (pH 6.0)/50% TFE.

Table 4.1 Growth curve results for inhibitor peptides with opposite charge motif

Inhibitor peptide	% inhibition in minimal media
18B	29.6
23	36.8
24	43.1
36A	58.1
45A	51.1
55A	81.8

Growth rates for cells containing the induced inhibitors were determined as described in the text and percent inhibition was calculated by comparing these values to the growth rate of cells that contained the induced pLAC11 vector. The averaged values of three independent determinations are shown.

Table 4.2 Sequence analysis of the insert region from the opposite charge motif inhibitor clones and the peptides that they are predicted to encode

18B - 25 aa

```
          -   -   -   -   -                                   -                                   +   +   +   +
AGATCT ATG GAA GAC GAA GAC GAG GGT GCG TCA GCG TGG GGA GCA GAA CTT TGG TCG TGG CAG TCG GTG CGT AAA CGT AAA TAA TAA
       M   E   D   E   D   E   G   A   S   A   W   G   A   E   L   W   S   W   Q   S   V   R   K   R   K   *   *
GAATTC
```

23 - 25 aa

```
          -   -   -   -                                   +                                   +   +   +   +
AGATCT ATG GAA GAC GAA GAC GGT CTA GGC ATG GGG GGT GGG TTG GTC AGG CTC ACT TTA TTA TTC TTC CGT AAA CGT AAA TAA TAA
       M   E   D   E   D   G   L   G   M   G   G   G   L   V   R   L   T   L   L   F   F   R   K   R   K   *   *
GAATTC
```

24 - 25 aa

```
          -   -   -   -       -   +                   +                               +   +   +   +   +
AGATCT ATG GAA GAC GAA GAC GGG GAG AGG ATC CAG GGG GCC CGC TGT CCA GTA GCG CTG GTA GAT AGA CGT AAA CGT AAA TAA TAA
       M   E   D   E   D   G   E   R   I   Q   G   A   R   C   P   V   A   L   V   D   R   R   K   R   K   *   *
GAATTC
```

36A - 11 aa

```
          -   -   -   -   -   +       +       +
AGATCT ATG GAA GAC GAA GAC GAC AGG GGG CGT GGG CGG TAG CTT TAA GTT GCG CTA AGT TGC GAG ATA CGT AAA CGT AAA TAA TAA
       M   E   D   E   D   D   R   G   R   G   R   *
GAATTC
```

45A - 25 aa

```
          -   -   -   -                   +   +                                       -   +   +   +   +
AGATCT ATG GAA GAC GAA GAC GGG GGG GCC GGG AGG AGG GCC TGT CTT TGT TCC GCG CTT GTT GGG GAA CGT AAA CGT AAA TAA TAA
       M   E   D   E   D   G   G   A   G   R   R   A   C   L   C   S   A   L   V   G   E   R   K   R   K   *   *
GAATTC
```

55A - 25aa

```
          -   -   -   -   +   +   +   -   +           +       +                               +   +   +   +
AGATCT ATG GAA GAC GAA GAC AAG CGT CGC GAG AGG AGT GCA AAA GGG CGT CAT GTC GGT CGG TCG ATG CGT AAA CGT AAA TAA GACTGT
       M   E   D   E   D   K   R   R   E   R   S   A   K   G   R   H   V   G   R   S   M   R   K   R   K   *
```

The landmark *Bgl*II and *Eco*RI restriction sites for the insert region are underlined. Amino acid charges are noted above the associated codons.

Table 4.3 Codon selection for randomized α-helical peptide inhibitors

TTT	**phe**	h_a	TCT	ser		TAT	tyr	b_a	TGT	cys	
TTC	**phe**	h_a	TCC	ser		TAC	tyr	b_a	TGC	cys	
TTA	**leu**	H_a	TCA	ser		TAA	OCH		TGA	OPA	
TTG	**leu**	H_a	TCG	ser		TAG	AMB		TGG	trp	
CTT	**leu**	H_a	CCT	pro	B_a	CAT	his	h_a	CGT	arg	
CTC	**leu**	H_a	CCC	pro	B_a	CAC	his	h_a	CGC	arg	
CTA	**leu**	H_a	CCA	pro	B_a	CAA	gln	h_a	CGA	arg	
CTG	**leu**	H_a	CCG	pro	B_a	CAG	gln	h_a	CGG	arg	
ATT	**ile**	h_a	ACT	thr		AAT	asn	b_a	AGT	ser	
ATC	**ile**	h_a	ACC	thr		AAC	asn	b_a	AGC	ser	
ATA	**ile**	h_a	ACA	thr		AAA	lys	h_a	AGA	arg	
ATG	**met**	H_a	ACG	thr		AAG	lys	h_a	AGG	arg	
GTT	**val**	h_a	GCT	**ala**	H_a	GAT	asp	h_a	GGT	gly	B_a
GTC	**val**	h_a	GCC	**ala**	H_a	GAC	asp	h_a	GGC	gly	B_a
GTA	**val**	h_a	GCA	**ala**	H_a	GAA	glu	H_a	GGA	gly	B_a
GTG	**val**	h_a	GCG	**ala**	H_a	GAG	glu	H_a	GGG	gly	B_a

Boldface amino acids are hydrophobic, while outlined amino acids are hydrophilic.

H_a designates residues that are strong helix-formers, h_a designates residues that are weak helix-formers.

B_a designates residues that are strong helix-breakers, b_a designates residues that are weak helix-formers.

Table 4.4 Growth curve results for inhibitor peptides with α-helical motif

Inhibitor peptide	% inhibition in minimal media
2B	56.6
18	23.1
115B	44.2
148	44.8
173	42.6

Growth rates for cells containing the induced inhibitors were determined as described in the text and percent inhibition was calculated by comparing these values to the growth rate of cells that contained the induced pLAC11 vector. The averaged values of three independent determinations are shown.

Table 4.5 Sequence analysis of the insert region from the α-helical motif inhibitor clones and the peptides that they are predicted to encode

2B - 18 aa

```
AGATCT ATG CAT GAC GAA CAA GAG GAG GAG CAC AAT AAA AAG GAT AAC GAA AAA GAA CAC TAA TAA GAATTC
       M   H   D   E   Q   E   E   E   H   N   K   K   D   N   E   K   E   H   *   *
```

115B - 22 aa

```
AGATCT ATG CAG CAG GAG CAC GAG CAA GGC AGG ATG AGC AAG AGG ATG AAG AAT AAT AAG AAT TCT CAT GTT TGA
       M   Q   Q   E   H   E   Q   G   R   M   S   K   R   M   K   N   N   K   N   S   H   V   *
```

148 - 22 aa

```
AGATCT ATG AAC CAT CAT AAT GAG GCC ATG ATC AAC ACA ATG AAA ACG AGG AAT AAT AAG AAT TCT CAT GTT TGA
       M   N   H   H   N   E   A   M   I   N   T   M   K   T   R   N   N   K   N   S   H   V   *
```

173 - 18 aa

```
AGATCT ATG AAC GAC GAC AAT CAG CAA GAG GAT AAT CAT GAT CAG CAT AAG GAT AAC AAA TAA TAA GAATTC
       M   N   D   D   N   Q   Q   E   D   N   H   D   Q   H   K   D   N   K   *   *
```

18 - 18 aa

```
AGATCT ATG CAC AAC CAG GAT AAC GAA CAA GAC GAG GAG GAT AAC GAG GAA CAG GAG GAG TAA TAA GAATTC
       M   H   N   Q   D   N   E   Q   D   E   E   D   N   E   E   Q   E   E   *   *
```

The landmark *Bgl*II and *Eco*RI restriction sites for the insert region are underlined.

Table 4.6 Percent α-helix content of peptides based on predictive algorithms

	Published	DPM	DSC	GOR1	GOR2	GOR4	HNNC	PHD	SOPM	SOPMA	Consensus
PKI	30	30	0	0	0	0	15	15	30	20	10
Endorphin	40	25.8	51.6	41.9	48.4	41.9	48.4	48.4	29.0	19.4	45.2
Motilin	50	40.9	45.5	59.1	50	40.9	40.9	40.9	45.5	50	40.9
Dermaseptin	70	64.7	94.1	79.4	88.2	61.8	85.3	85.3	82.4	82.4	85.3
Mastoparan	78	71.4	100	100	100	35.7	71.4	92.9	0	0	78.6
Melittin	80	30.8	34.6	50	57.7	26.9	53.9	69.2	57.7	57.7	53.9
Maganin	>90	34.8	73.9	100	69.6	17.4	39.1	78.3	95.7	95.7	78.3
2B		22.2	0	100	88.9	33.3	27.8	33.3	83.3	83.3	50
18		11.1	0	77.8	55.6	22.2	0	0	55.6	55.6	33.3
115B		0	50	68.2	59.1	36.4	45.5	31.8	50.0	50.0	40.9
148		0	40.9	68.2	59.1	31.8	36.4	45.5	45.5	45.5	40.9
173		0	0	72.2	55.6	22.2	5.6	0	22.2	22.2	11.1

Highlighted boxes indicate best prediction method based on NPS@ Consensus Secondary Structure Prediction website: http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html (Combet *et al*., 2000)

Table 4.7 Summary of the frequency and potency at which the different types of inhibitor peptides can be isolated

Type of inhibitor peptide	Frequency at which a potent 2-day inhibitor peptide can be isolated	Average inhibition (%)	Average potency (%)
anchorless[a]	1 in 20,000	25	65
protected at the C-terminus via Rop[b]	1 in 625	92	98
protected at the N-terminus via Rop[b]	1 in 429	93	98
protected at the C-terminus and/or N-terminus via proline dipeptide residues[b]	1 in 625	53	89
protected with a terminal opposite charge motif	1 in 3,333	50	88
protected with an α-helical structural motif	1 in 2,400	42	83

[a]in Walker *et al*., 2001.

[b]in Walker *et al*., 2003.

CHAPTER 5

EXAMINATION OF ROP-LIKE AND OTHER HIGHLY STABLE HELICES AS POTENTIAL PEPTIDE-STABILIZING MOTIFS

Abstract

The highly stable, largely helical Rop protein has been demonstrated to be an effective stabilizing anchor for preventing the degradation of small bioactive peptides. In this study, a group of naturally occurring α-helices based on Rop and other helical proteins, as well as potential helical peptides designed *de novo* were investigated to determine if they could serve as potential stabilizing groups. Using a combined approach of secondary structure prediction, CD spectroscopic calculation of helical content, and *in vitro* rat serum degradation assays, two highly stable helices were found that had half-lives which were significantly higher than other peptides of similar size, and similar to those of other small stable proteins.

Introduction

Peptides of varying size and derivation have found widespread use as therapeutic and diagnostic agents. A brief survey of these agents currently being marketed or in clinical trials shows that the majority has been derived from natural products (Latham, 1999; Lien and Lowman, 2003). However, combinatorial library screening now plays an increasingly important role in the lead discovery process. In many cases, peptides isolated by combinatorial techniques have been incredibly promising during *in vitro* studies, but have failed during *in vivo* trials due mainly to the rapid degradation of peptides by endogenous host peptidases and proteases (Kelley, 1996). For example, the biological stability of glucagon-like peptide 1 (GLP-1), involved in treatment of type 2 diabetes, initially showed a half-life of only minutes during *in vivo* studies (Gallwitz *et al*., 2000). Likewise, thymopentin, the synthetic pentapeptide that reproduces the biological effect of the 49 amino acid hormone thymopoietin, exhibits a half-life in plasma of approximately 30 seconds (Heavner *et al*., 1986).

To avoid this rapid degradation of peptides, researchers have employed methods such as conformational constraint via cyclization, use of D-amino acids, or use of unusual amino acids to stabilize potential therapeutic agents (Hruby and Bonner, 1994). Another method of protection involves the administration of peptidase/protease inhibitors in concert with the peptide of interest (Bai and Amidon, 1992). For example, researchers found that the presence of boroleucine prolonged the half-lives of leucine enkephalins 4.0 to 6.4-fold in rat plasma (Hussain *et al*., 1990). The addition of protective motifs such as methyl and amide groups to the ends of peptides have yielded longer efficacy for peptides such as GLP-1 and anti-HIV peptide T140 (Gallwitz *et al*., 2000; Tamamura *et al*., 2001). However, even though this type of terminal modification may offer protection from degradation of peptides, often the increased stability is

accompanied by a reduction in bioactivity. When thymopentin is modified by either amino-terminal acetylation or carboxyl-terminal amidation, a complete loss of biological activity is observed (Heavner *et al*., 1986). For this reason, the technique known as PEGylation has become increasingly utilized in an effort to protect peptides against degradation without negatively impacting bioactivity. Covalently linking polyethylene glycol (PEG) moieties to amino groups of polypeptides is a protective method that has been used for over 30 years, and is currently undergoing a resurgence of popularity as researchers use the strategy to increase the *in vivo* stability of numerous peptide and protein drug candidates, such as growth hormone-releasing hormone (GRF), interleukin-2 (IL-2), and human interferon-α-2b (Campbell *et al*., 1997; Katre *et al*., 1987; Glue *et al*., 2000).

Many naturally occurring peptides are able to maintain secondary structure or contain protective groups that are thought to play a role in resisting degradation by peptidases, and certain peptide therapeutics demonstrate such a stabilized structure. For example, gramicidin S, polymixin B and bacitracin are antibiotic peptides that have a cyclized structure (Hancock and Chapple, 1999). Other natural bioactive peptides contain certain structural groups such as α-helix, β-sheet and disulfide bond-stabilized loops which may provide peptidase protection (Epand and Vogel, 1999; Powers and Hancock, 2003). Glucagon and endorphin are examples of peptides that exhibit α-helical structure, while vasopressin and oxytocin are among those stabilized by disulfide bonds (Bedarkar *et al*., 1977; Blanc *et al*., 1983; Fong *et al*., 1964; Urry *et al*., 1968).

Walker *et al.* have shown that peptide stability can be dramatically affected by fusing the small 63-amino acid Rop protein to one terminus of a bioactive peptide. Presumably, the highly stable helix-turn-helix protein enhances the stability of the peptide that it is linked to by

protecting it from degradation by exopeptidases (Walker *et al*., 2003). The possibility of the Rop protein acting in a protective nature to fused peptides is not surprising in view of the very stable α-helical secondary structure of this small protein (Banner *et al*., 1987; Castagnoli *et al*., 1989). Rop has already been engineered for use as a carrier molecule for *in vivo* and *in vitro* presentation studies for bioactive molecules by altering the loop region to incorporate short peptide sequences (Kresse *et al*., 2001). Since the highly stable structure of Rop is primarily due to its a-helical composition, the possibility of using small α-helical peptides and protein domains as protective anchors was investigated in this study.

Materials and Methods

Synthetic peptides

Peptides used in this study were custom synthesized by Sigma-Genosys (The Woodlands, TX, USA) and the primary sequences are listed in Table 5.1.

Secondary structure prediction

The propensity of the putative helical inhibitory peptides to form α-helices was initially analyzed using prediction algorithms including those developed by Chou and Fasman and Garnier, Osguthorpe, and Robson. Peptide sequences were also analyzed using the multiple algorithms of the NPS@ Consensus Secondary Structure Prediction website (Combet *et al*., 2000).

Circular dichroism

CD measurements were made using a Jasco-710 spectropolarimeter interfaced and controlled by a computer. The measurements were made using a circular quartz cell of 0.1-cm path length. All peptide samples were in 20 mM sodium phosphate buffer (pH 6.0) at a concentration of 100 μM. Measurements were made over a range of 250-190 nm, a scan speed

of 20 nm/min with a sensitivity of 50 millidegrees, a resolution of 0.2 nm, and a band width of 1.0 nm. Five scans were accumulated and averaged, and all dichroic spectra were corrected by subtraction of the background and then smoothed via an internal algorithm in the Jasco software package, J-710 for Windows. The CD spectra are reported as the mean residue molar ellipticity ([θ]) in degrees•cm^2•$dmol^{-1}$. Secondary structure was estimated by utilizing the MRE value at 222 nm. Taking peptide length into account, percent α-helix = 100[MRE]/39,500(1 – 2.57/*n*), where *n* is the number of amino acid residues (Scholtz *et al*., 1991).

In vitro degradation studies

810 μL of rat serum was mixed with 90 μL of peptide at a concentration of 1 mg/mL in 10 mM Tris (pH 8.0) and incubated at 37°C. 100 μL aliquots were removed at 1, 2, 4, 8, and 24 hour intervals, 200 μL of methanol was added and the samples were vortexed, placed on ice for 5 minutes, and centrifuged for 5 minutes. 250 uL of supernatant was removed and stored at –70°C until analysis via LC/MS conducted by Bay Bioanalytical Laboratory (Hercules, CA, USA). A Shimadzu Advp LC system coupled to a PE SCIEX API3000 Triple Quadrupole MS was used for LC/MS analysis. The peptides were separated on a 5μm particle size Thermo-Hypersil Keystone BetaBasic C_{18} 2 x 10 mm Guard Column using a 5 to 95% acetonitrile gradient that contained 0.1% trifluoroacetic acid. The most abundant +3 or +4 charge state of each peptide was analyzed.

To avoid potential discrepancies among different lots of rat serum, all serum used in these studies was normalized using the randomized and XPP peptides described in previous work (Walker *et al*., 2003). Only sera which were within 10% of the published values were used in these experiments.

Results

Selection of peptides for the study

Nine peptides with potentially high helical content were selected for this study (Table 5.1). Peptide 1 was designed with six repeating units of three hydrophilic helix-forming amino acids, glutamic acid, lysine and glutamine (EKQ). Peptide 2 was derived from one of the most highly stable α-helical protein fragments studied to date, discovered after researchers elucidated the structure of the ribosomal protein L9 and found this particular helix which connects the two globular domains of the protein (Kuhlman *et al*., 1997). Peptide 3 is an α-helical fragment of the Rop-like antiparallel four-helix bundle α-catenin adhesion modulation domain (Yang *et al*., 2001; Pokutta *et al*., 2002). Peptide 4 is the S-alpha fragment of the C-type lectin mannose binding protein (Sheriff *et al*., 1994). Peptides 5 and 6 are two proline-rich peptides consisting of ten repeating tripeptide units (PPG and PPA, respectively) synthesized in an attempt to develop a stable protective anchor that mimics a unique secondary structure called a polyproline II helix, which is a preferred solution conformer occurring when four or more proline residues are found in a row. It has been suggested that this structure is responsible for stabilizing a group of proline-rich antibacterial peptides including the bacternecins and indolicidin (Gennaro *et al*., 1989; Falla and Hancock, 1997; Ladokhin *et al*., 1997). Peptide 7 is a 24-amino acid version of the EKQ peptide, synthesized after it was postulated that the 18-amino acid length of Peptide 1 was too short to exhibit significant helical character. Peptides 8 and 9 are derived from the carboxyl and amino terminals of the most stable derivative (RLP-3) of a series of small helix-turn-helix proteins designed *de novo* and intended to mimic Rop by dimerizing into antiparallel four-helix bundles (Betz *et al*., 1996; Betz *et al*., 1997).

Algorithmic prediction of secondary structure

The helical peptides were analyzed using the secondary structure prediction algorithms from the NPS@ website and all were predicted to contain high percentages of helical content (72-92%) with the exception of Pep5 and Pep6 (Table 5.2). These algorithms were designed to predict the basic polypeptide secondary structures α-helix, β-sheet, β-turn and random coil, but do not account for the polyproline II helix. Therefore, it was not unexpected that Pep 5 and Pep 6 were predicted to have zero percent helical content when analyzed, due to their repeating PPG and PPA primary sequence.

Calculated helical content via CD spectroscopy

The nine peptides were subjected to CD spectroscopy in order to calculate their actual helical content in solution. Following initial analysis in sodium phosphate buffer, only four peptides, Pep2, Pep7, Pep8, and Pep9, had spectra consistent with demonstrable α-helical content, but calculations showed the percentages to be quite low compared to the algorithmic predictions (Table 5.3). TFE was added to the peptide solutions to a final concentration of 50% to determine if this helix-stabilizing chemical would increase the observed helical content. Increases in calculated helical content was observed in all seven peptides examined, with the differences demonstrated by the four peptides that showed measurable helicity in aqueous solution ranging from 1.4-fold higher for the Ribosomal L9 fragment to a 9.2-fold increase for the $(EKQ)_8$ peptide (Table 5.3).

Half-life determination via in vitro degradation assay

Based on the above results, seven peptides were subjected to an *in vitro* rat plasma degradation assay and half-lives were determined following analysis via LC/MS. Peptides 2, 3, 4, and 9 had short half-lives ranging from 30 minutes to just over 2 hours, while Peptides 6, 7,

and 8 demonstrated a high degree of stability which greatly surpassed that of control peptides and more closely approached that of small protein controls, ranging from 14 hours to 4 days (Table 5.4 and 5.5). Peptide 1 was the shorter 18-amino acid version of the EKQ peptide which failed to exhibit helicity during examination by CD spectroscopy and was therefore excluded from the degradation assay and LC/MS analysis. Peptide 5 was the $(PPG)_{10}$ putative polyproline II helix, but CD spectra failed to demonstrate the curve characteristic of this secondary structure, and it was also excluded from half-life determination.

Discussion

In an examination of nine helical peptides selected for their potential ability to act as stabilizing motifs for bioactive peptides, two promising candidates were identified by this study. One of these was designed with a repeating sequence of three hydrophilic, helix-forming amino acids (Glu-Lys-Gln). This result was somewhat surprising considering the limited success that other researchers have observed during attempts to generate designed α-helices using limited sets of amino acids. A model 51 amino acid peptide consisting of tandem repeats of two helix-forming amino acids (Lys-Lys-Leu-Leu) and containing a turn sequence (Asn-Pro-Gly) showed no significant conformation by CD under conditions of neutral pH (Goto and Aimoto, 1991). The other peptide that demonstrated high stability in this study was designed to adopt the secondary structure known as a polyproline II (PPII) helix. This result was less surprising as this left-handed helical conformation is similar to the basic structural type observed in the highly stable collagen triple helix, although the character of its stabilization is different (Fraser *et al*., 1979). Aside from proline, the other four amino acids with Chou-Fasman parameters higher for PPII helices greater than four residues than for other secondary structures are Ala, Arg, Gln and Ser (Adzhubei and Sternberg, 1993). In contrast, glycine is rare, with the lowest frequency for

PPII helices four residues or longer among all types of secondary structure. It has been suggested that due to the lack of a regular network of hydrogen bonds in this conformation, the presence of Gly in longer PPII helices is more destabilizing than in other secondary structures (Sreerama and Woody, 1992). This may account for the apparent lack of secondary structure exhibited by the $(PPG)_{10}$ peptide in these experiments.

It was observed in this study that the predicted α-helical percentage of peptides by various secondary structure algorithms was not an accurate guide to the actual helical content of these peptides in solution. These algorithms were designed to predict the secondary structure of proteins and some are in fact based on subsets of the existing protein structure databases such as the Protein Data Bank (PDB) and the Cambridge Structural Database (CSD) (Berman *et al*., 2000). However, neither the PDB nor the CSD are specifically aimed at peptides, such that the peptide structural data they contain are only a small subset of the databases. Furthermore, it would appear as though the calculation of helical content in peptides via CD spectroscopy is affected to some degree by the small size of these molecules and their dynamic nature in solution. While peptides do show similarities with proteins, the higher order structural features of the larger molecules certainly influence the conformational characteristics of the amino acid residues. An α-helix found in the interior of a protein has amino and carboxyl termini that are bound by the protein chain on either end, limiting the conformational flexibility of the structure. Helices found inside proteins are further stabilized by interactions with other parts of the protein. Conversely, helical peptides are not stabilized by any of these interactions. More importantly, their amino and carboxyl termini are free, allowing for the helix to wind and unwind depending upon various solution characteristics such as pH, ionic strength and temperature. Finally, it is perhaps not unexpected that the α-helical peptides demonstrated percentages of helix content far

below that predicted by secondary structure algorithms when one considers the findings of researchers from the Baldwin lab. Their observations led to the conclusion that "the helix propensities of most amino acids oppose folding; consequently, the majority of isolated helices derived from proteins are unstable, unless specific side-chain interactions stabilize them" (Chakrabartty *et al*., 1994).

Future studies will likely include an examination of longer variants of the EKQ peptide to determine if an increase in peptide length leads to higher α-helical content and greater overall stability. Also planned is an examination of the two individual α-helices that comprise the highly stable Rop protein monomer, in a manner similar to that conducted with the two helical halves of the RLP-3 protein in this study. Finally, in accordance with having the highest propensities toward forming PPII helices based on Chou-Fasman conformational parameters, peptides consisting of repeating PPQ and PPR units may be studied.

References

Adzhubei, A.A. and Sternberg, M.J.E. (1993). Left-handed polyproline II helices commonly occur in globular proteins. J Mol Biol 229, 472-493.

Bai, J.P. and Amidon, G.L. (1992). Structural specificity of mucosal-cell transport and metabolism of peptide drugs: implication for oral peptide drug delivery. Pharm Res 9, 969-978.

Banner, D.W., Kokkinidis, M., and Tsernoglou, D. (1987). Structure of the ColE1 Rop protein at 1.7 Å resolution. J Mol Biol 196, 657-675.

Bedarkar, S., Blundell, T.L., Dockerill, S., Tickle, I.J., and Wood, S.P. (1977). Polypeptide hormone-receptor interactions: the structure and receptor binding of insulin and glucagon. Ciba Found Symp 60, 105-121.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res 28, 235-242.

Betz, S.F., Bryson, J.W., Passador, M.C., Brown, R.J., O'Neil, K.T., and DeGrado, W.F. (1996). Expression of *de novo* designed alpha-helical bundles. Acta Chem Scand 50, 688-696.

Betz, S.F., Liebman, P.A., and DeGrado W.F. (1997). *De novo* design of native proteins: characterization of proteins intended to fold into antiparallel, Rop-like, four-helix bundles. Biochemistry 36, 2450-2458.

Blanc, J.P., Taylor, J.W., Miller, R.J., and Kaiser, E.T. (1983). Examination of the requirement for an amphiphilic helical structure in beta-endorphin through the design, synthesis, and study of model peptides. J Biol Chem 258, 8277-8284.

Brownlees, J. and Williams, C.H. (1993). Peptidases, peptides, and the mammalian blood-brain barrier. J Neurochem 60, 793-803.

Campbell, R.M., Heimer, E.P., Ahmad, M., Eisenbeis, H.G., Lambros, T.J., Lee, Y., Miller, R.W., Stricker, P.R., and Felix, A.M. (1997). Pegylated peptides. V. Carboxy-terminal PEGylated analogs of growth hormone-releasing factor (GRF) display enhanced duration of biological activity *in vivo*. J Pept Res 49, 527-537.

Castagnoli, L., Scarpa, M., Kokkinidis, M., Banner, D.W., Tsernoglou, D., and Cesareni, G. (1989). Genetic and structural analysis of the ColE1 Rop (Rom) protein. EMBO J 8, 621-629.

Chakrabartty, A., Kortemme, T., and Baldwin, R.L. (1994). Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Protein Sci 3, 843-852.

Chou, P.Y. and Fasman, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol 47, 45-148.

Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000). NPS@: network protein sequence analysis. Trends Biochem Sci 25, 147-150.

Epand, R.M. and Vogel, H.J. (1999). Diversity of antimicrobial peptides and their mechanisms of action. Biochim Biophys Acta 1462, 11-28.

Falla, T.J. and Hancock, R.E. (1997). Improved activity of a synthetic indolicidin analog. Antimicrob Agents Chemother 41, 771-775.

Fong, C.T., Silver, L., and Louie, D.D. (1964). Necessity of the disulfide bond of vasopressin for antidiuretic activity. Biochem Biophys Res Commun 14, 302-306.

Fraser, R.D.B., MacRae, T.P., and Suzuki, E. (1979). Chain conformation in the collagen molecule. J Mol Biol 129, 463-481.

Gallwitz, B., Ropeter, T., Morys-Wortmann, C., Mentlein, R., Siegel, E.G., and Schmidt, W.E. (2000). GLP-1-analogues resistant to degradation by dipeptidyl-peptidase IV *in vitro*. Reg Pep 86, 103-111.

Garnier, J., Osguthorpe, D.J., and Robson, B. (1978). Analysis of the accuracy and implications of simple method for predicting the secondary structure of globular proteins. J Mol Biol 120, 97-120.

Gennaro, R., Skerlavaj, B., and Romeo, D. (1989). Purification, composition, and activity of two bactenecins, antibacterial peptides of bovine neutrophils. Infect Immun 57, 3142-3146.

Glue, P., Fang, J.W., Rouzier-Panis, R., Raffanel, C., Sabo, R., Gupta, S.K., Salfi, M., and Jacobs, S. (2000). Pegylated interferon-α2b: pharmacokinetics, pharmacodynamics, safety, and preliminary efficacy data. Hepatitis C Intervention Therapy Group. Clin Pharmacol Ther 68, 556-567.

Goto, Y. and Aimoto, S. (1991). Anion and pH-dependent conformational transition of an amphiphilic polypeptide. J Mol Biol 218, 387-396.

Hancock, R.E.W. and Chapple, D.S. (1999). Peptide antibiotics. Anti Agents Chem 43, 1317-1323.

Heavner, G.A., Kroon, D.J., Audhya, T., and Goldstein, G. (1986). Biologically active analogs of thymopentin with enhanced enzymatic stability. Peptides 7, 1015-1019.

Hruby, V.J. and Bonner, G.G. (1994). Design of novel synthetic peptides including cyclic conformationally and topographically constrained analogs. Methods Mol Biol 35, 201-240.

Hussain, M.A., Rowe, S.M., Shenvi, A.B., and Aungst, B.J. (1990). Inhibition of leucine enkephalin metabolism in rat blood, plasma and tissues *in vitro* by an aminoboronic acid derivative. Drug Metab Dispos 18, 288-291.

Katre, N.V., Knauf, M.J., and Laird, W.J. (1987). Chemical modification of recombinant interleukin 2 by polyethylene glycol increases its potency in the murine Meth A sarcoma model. Proc Natl Acad Sci USA 84, 1487-1491.

Kelley, W.S. (1996). Therapeutic peptides: the devil is in the details. Biotechnology (N.Y.) 14, 28-31.

Kresse, H.P., Czubayko, M., Nyakatura, G., Vriend, G., Sander, C., and Bloecker, H. (2001). Four-helix bundle topology re-engineered: monomeric Rop protein variants with different loop arrangements. Pro Eng 14, 897-901.

Kuhlman, B., Yang, H.Y., Boice, J.A., Fairman, R., and Raleigh, D.P. (1997). An exceptionally stable helix from the ribosomal protein L9: implications for protein folding and stability. J Mol Biol 270, 640-647.

Ladokhin, A.S., Selsted, M.E., and White, S.H. (1997). Bilayer interactions of indolicidin, a small antimicrobial peptide rich in tryptophan, proline, and basic amino acids. Biophys J 72, 794-805.

Latham, P.W. (1999). Therapeutic peptides revisited. Nat Biotechnol 17, 755-757.

Lien, S. and Lowman, H.B. (2003). Therapeutic peptides. Trends Biotechnol 21, 556-562.

Pokutta, S., Drees, F., Takai, Y., Nelson, W.J., and Weis, W.I. (2002). Biochemical and structural definition of the l-afadin- and actin-binding sites of α-catenin. J Biol Chem 277, 18868-18874.

Powers, J.S. and Hancock, R.E. (2003). The relationship between peptide structure and antibacterial activity. Peptides 24, 1681-1691.

Rawlings, N.D. and Barrett, A.J. (1993). Evolutionary families of peptidases. Biochem J 290, 205-218.

Scholtz, J.M., Qian, H., York, E.J., Stewart, J.M., and Baldwin, R.L. (1991). Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. Biopolymers 31, 1463-1470.

Sheriff, S., Chang, C.Y., and Ezekowitz, R.A. (1994). Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil. Nat Struct Biol 3, 789-794.

Sreerama, N. and Woody, R.W. (1994). Poly(pro)II helices in globular proteins: identification and circular dichroic analysis. Biochemistry 33, 10022-10025.

Tamamura, H., Omagari, A., Hiramatsu, K., Gotoh, K., Kanamoto, T., Xu, Y., Kodama, E., Matsuoka, M., Hattori, T., Yamamoto, N., Nakashima, H., Otaka, A., and Fujii, N. (2001). Development of specific CXCR4 inhibitors possessing high selectivity indexes as well as complete stability in serum based on an anti-HIV peptide T140. Bioorg Med Chem Lett 11, 1897-1902.

Urry, D.W., Quadrifoglio, F., Walter, R., and Schwartz, I.L. (1968). Conformational studies on neurohypophyseal hormones: the disulfide bridge of oxytocin. Proc Natl Acad Sci USA 60, 967-974.

Walker, J.R., Altman, R.K., Warren, J.W., and Altman, E. (2003). Using protein-based motifs to stabilize peptides. J Peptide Res 62, 214-226.

Walker, J. R., Roth, J. R., and Altman, E. (2001). An *in vivo* study of novel bioactive peptides that inhibit the growth of *Escherichia coli*. J Peptide Res 58, 380-388.

Warren, J.W. and Altman, E. Peptide stabilization with terminal opposite charge and α-helical motifs. Manuscript in preparation.

Warren, J. W., Walker, J. R., Roth, J. R., and Altman, E. (2000). Construction of a highly regulable expression vector, pLAC11, and its multipurpose derivatives, pLAC22 and pLAC33. Plasmid 44, 138-151.

Yang, J., Dokurno, P., Tonks, N.K., and Barford, D. (2001). Crystal structure of the M-fragment of α-catenin: implications for modulation of cell adhesion. EMBO J 20, 3645-3656.

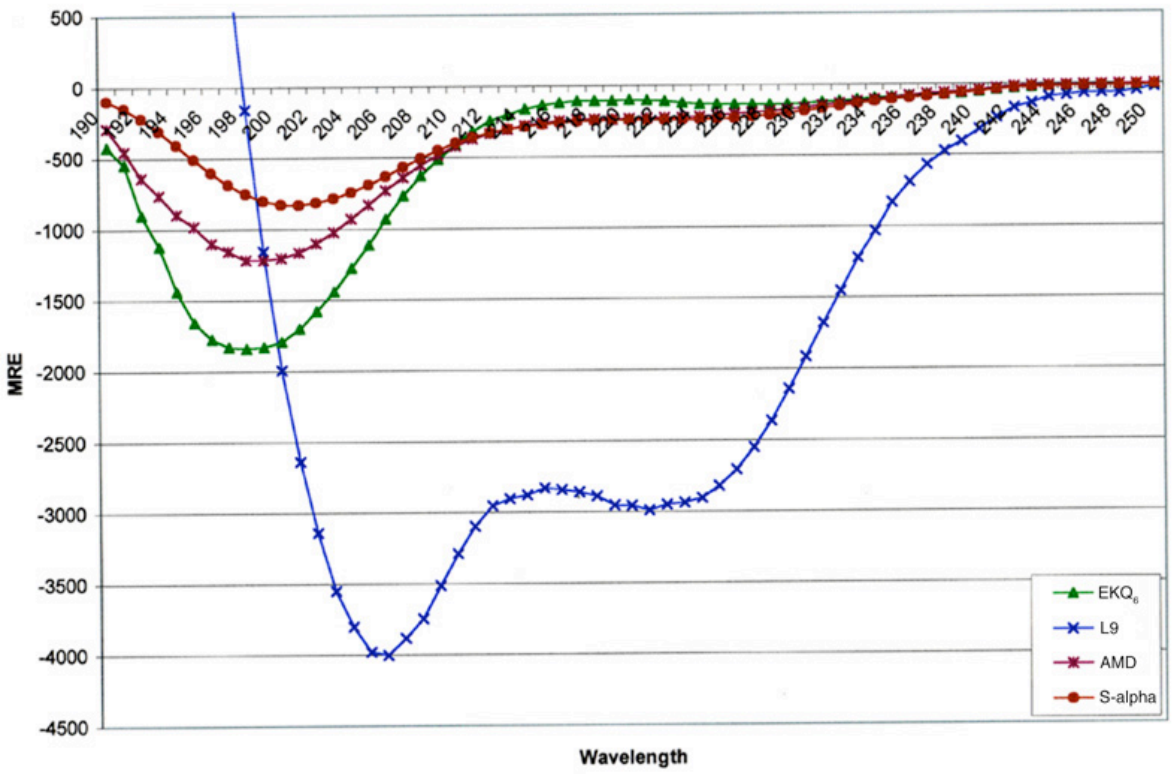


Figure 5.1 CD spectra of putative helical peptides 1-4. Peptide concentration 100 mM in 20 mM sodium phosphate buffer (pH 6.0).

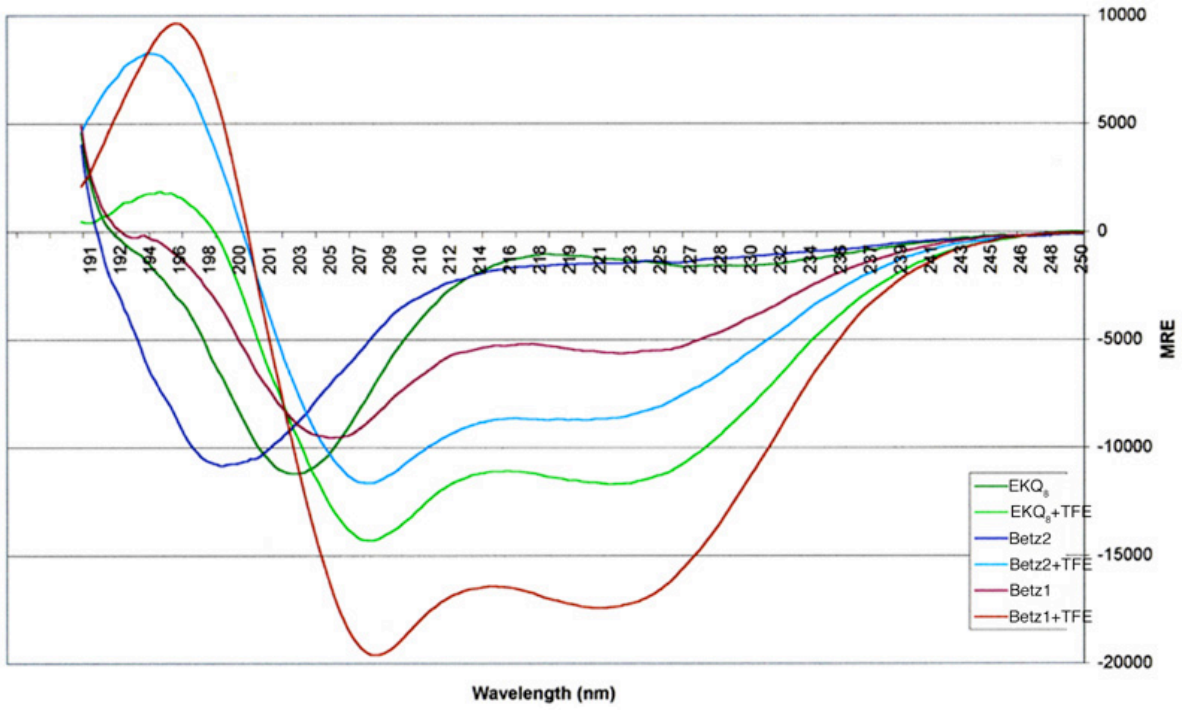


Figure 5.2 CD spectra of putative helical peptides 7-9. Peptide concentration 100 mM in 20 mM sodium phosphate buffer (pH 6.0) with and without 50% TFE.

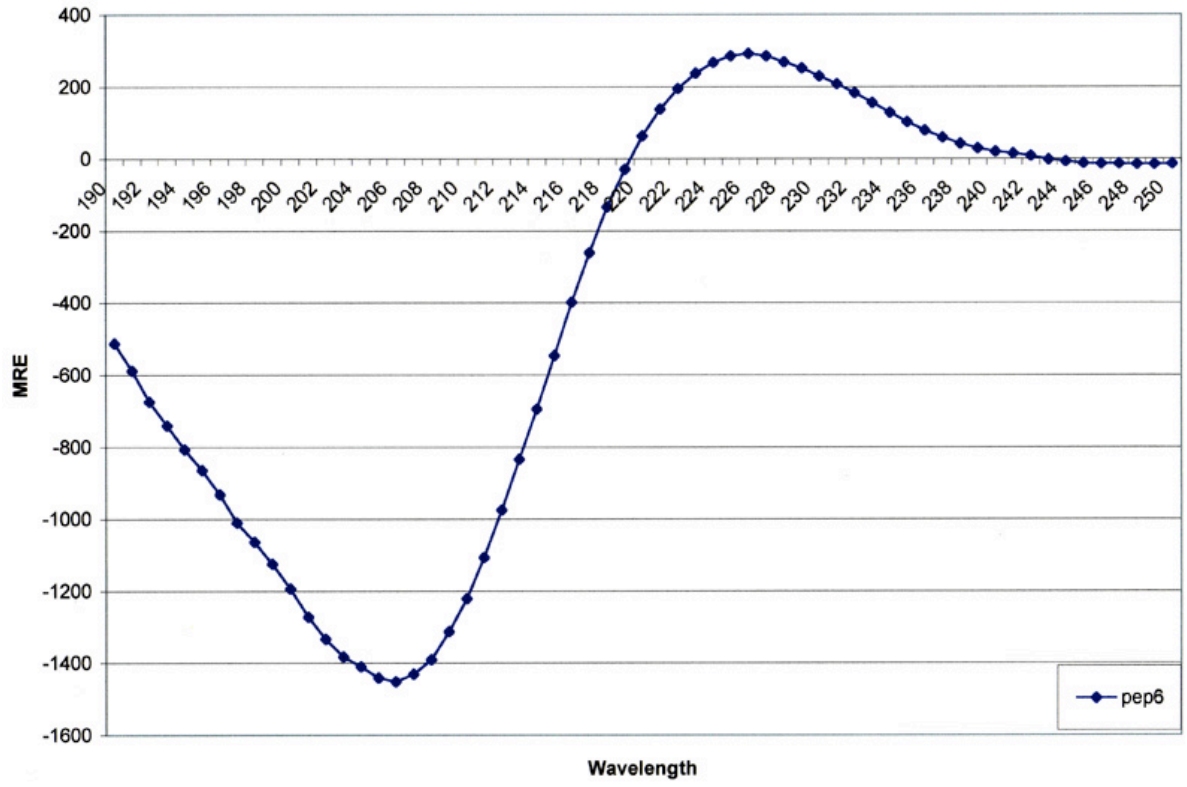


Figure 5.3 CD spectra of putative helical peptide 6. Peptide concentration 100 mM in 20 mM sodium phosphate buffer (pH 6.0).

Table 5.1 Sequences of peptides

Pep1:	EKQEKQEKQEKQEKQEKQ (18aa) repeating unit of hydrophilic helix-forming amino acids
Pep2:	PANLKALEAQKQKEQRQAAEELANAKKLKEQLEK (34aa) Ribosomal Protein L9 peptide fragment
Pep3:	LAVSAAHIAEDVNKCVIALQEKDVDGLDRRTAGAIRGRAARVI (42aa) two-helical bundle α-catenin AMD peptide
Pep4:	AASERKALQTEMARIKKALTA (21aa) S-alpha peptide fragment of C-type lectin MBP
Pep5:	PPGPPGPPGPPGPPGPPGPPGPPGPPGPPG (30aa) putative polyproline II helix
Pep6:	PPAPPAPPAPPAPPAPPAPPAPPAPPA (30aa) putative polyproline II helix
Pep7:	EKQEKQEKQEKQEKQEKQEKQEKQ (24aa) repeating unit of hydrophilic helix-forming amino acids
Pep8:	ELLKEVEELEKKVDKLYKIVEH (22aa) C-terminal helix of Betz RLP-3
Pep9:	SAQELLKIARRLRKEAKELLKRAEH (25aa) N-terminal helix of Betz RLP-3

Table 5.2 Predicted % α-helix content of potential stable helical peptide anchors compared to the two helices of the Rop protein

	DPM	DSC	GOR1	GOR2	GOR4	HNNC	PHD	SOPM	SOPMA	Consensus
Pep1	78	0	100	100	100	44	72	94	94	78
Pep2	85	82	100	100	74	71	91	100	100	91
Pep3	60	42	81	84	53	47	74	77	70	72
Pep4	81	86	100	100	52	76	90	95	95	90
Pep5	0	0	0	0	0	0	0	0	0	0
Pep6	0	0	0	0	0	0	0	0	0	0
Pep7	83	0	100	100	63	38	79	96	96	88
Pep8	59	86	100	100	55	55	91	100	100	91
Pep9	84	80	100	100	64	88	92	96	96	92
Rop1	57	82	96	100	64	79	82	89	89	86
Rop2	85	54	96	81	46	85	88	88	88	88

Prediction method based on NPS@ Consensus Secondary Structure Prediction website: http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html (Combet *et al*., 2000)

Table 5.3 Calculated % α-helical content of peptides from CD spectra

	20mM Na phos	20 mM Na phos + 50% TFE
Pep1	0.3	ND
Pep2	8.1	11.0
Pep3	0.6	18.7
Pep4	0.7	26.2
Pep5	0.03	ND
Pep6	0.5	3.8
Pep7	3.6	33.1
Pep8	4.2	24.7
Pep9	15.8	49.0

ND = not determined

Table 5.4 ***In vitro*** **rat plasma degradation assays of control peptides and small proteins**

Peptide	Plasma half-life (min)
GLP-1	13.3
Galanin	70.5
VIP	155
PTH	182
Small protein	
Rop	1420
Gst	>2880
Trx	>2880

Table 5.5 Comparison of predicted and calculated α-helicity with peptide half-life from *in vitro* rat plasma degradation studies

	Predicted α-helicity (%)	CD calculated α-helicity (%)	Plasma half-life (min)
Pep2	91	11.0	30
Pep3	72	18.7	30
Pep4	90	26.2	90
Pep6	ND	ND	5220
Pep7	88	33.1	5760
Pep8	91	24.7	840
Pep9	92	49.0	144

Pep6 is a putative polyproline II helix and as such was not amenable to secondary structure algorithmic prediction or calculation of helicity by CD.

CHAPTER 6

SELF-DIRECTED INHIBITION

Abstract

Secondary structural domains such as α-helices and β-pleated sheets have been documented to be key nucleation factors that contribute to protein folding and the achievement of a final tertiary structure. Based on this fact it was proposed that if key structural domains of a protein were overexpressed *in vivo* they would interfere with the proper folding of the protein into its correct tertiary active form. This study examined this possibility using two different target proteins, LacZ (β-galactosidase) of *Escherichia coli* and XylE (catechol-2,3-oxygenase) of *Pseudomonas putida*, as well as results obtained from similar studies conducted by other researchers.

Introduction

Selective inhibition of specific gene targets represents a general approach to the analysis of gene function and a possible method of treatment for various genetically-linked diseases. Inhibition for these purposes can be achieved by either disruption of the gene of interest or by introduction of genetic elements that interfere with the expression of the target gene or the function of gene products. One such element is antisense RNA which involves the production of RNA sequences complementary to the mRNA of the target gene. The inhibitory effect is caused by the binding of antisense RNA to mRNA which blocks translation and prevents protein synthesis (Takayama and Inouye, 1990). Though the theory is sound, many antisense RNA constructs that have shown promise during *in vitro* studies, such as those designed to inhibit influenza or SV40 viral replication, have demonstrated little or no biological effect *in vivo* (Hasan *et al*., 1988; Kerr *et al*., 1988; Leiter *et al*., 1989). Another genetic element that has been examined as an approach to targeted gene inhibition involves the use of mutated proteins that interfere with the function of the native protein when expressed in a wild-type background. These dominant-negative mutants may be derived from normal proteins by truncation of domains involved in specific functional interactions such as oligomerization (Herskowitz, 1987). However, unless the domain structure of the target protein is well characterized it becomes problematic to predict which, if any, segments of the protein act as dominant negative mutants. Furthermore, *in vivo* studies have revealed that the dominant negative form of some proteins fail to inhibit their native targets in the manner exhibited during *in vitro* experiments due to rapid degradation, possibly related to instability of the tertiary structure (Ramalho *et al*., 2002).

In response to the shortcomings of the antisense and dominant negative protein approaches, the Roninson lab developed a technique in which both short sense and antisense

gene elements which exert a phenotypic effect are generated by random fragmentation of the DNA of a target gene or multigene complex (Roninson *et al*., 1995). They used the term genetic suppressor element (GSE) to describe these inhibitors of gene function and first demonstrated the concept by showing that randomly fragmented bacteriophage lambda DNA increased survival of *E. coli* cells transformed with the fragment expression library upon challenge with phage λ (Holtzmayer *et al*., 1992). Other studies by this group of researchers further demonstrated the technique through the isolation of GSEs which inhibited the function of human p53 and human topoisomerase II and helped to elucidate the biological effects of the various domains of these proteins (Gudkov *et al*., 1993; Ossovskaya *et al*., 1996). Another group of scientists has applied a similar approach to the analysis of the pheromone response pathway in *Saccharomyces cerevisiae* and collectively referred to the random short peptides and larger protein fragments that inhibited the pathway as "perturbagens" (Caponigro *et al*., 1998).

In an effort to make targeted protein inhibition a realistic tool for researchers it was decided to undertake an investigation of a method similar to those described above based on what is known about the mechanics of protein folding. It has become increasingly clear that as a protein folds from its primary amino acid sequence key domains of the protein interact with each other and facilitate the folding of the protein into its final tertiary structure (Branden and Tooze, 1991). Interestingly, the size of these key domains, which are typically stable secondary structures such as α-helices and β-sheets, appears to be 20-30 amino acids in length. It was proposed that random fragments from a particular protein could be expressed in a cell as novel bioactive peptides, and might interfere with the folding of the intact parental protein, thereby rendering it nonfunctional (Figure 6.1). It was decided to test this hypothesis using the LacZ (β-galactosidase) protein of *Escherichia coli* and XylE (catechol-2,3-oxygenase) protein of

Pseudomonas putida as targets, since both proteins have well-characterized crystal structures and both phenotypic and colorimetric assays can be employed to easily screen for LacZ^- and $XylE^-$ colonies and measure enzymatic activity of the native proteins respectively.

Materials and Methods

Media

Rich Luria-Bertani media used in this study was prepared as described by Miller (1972). Ampicillin was used in rich media at a final concentration of 100 μg/mL, chloramphenicol was used in rich media at a final concentration of 20 μg/mL. Isopropyl β-D-thiogalactoside (IPTG) was added to media at a final concentration of 1 mM.

Bacterial strains and plasmids

The bacterial strains and plasmids that were used in this study are listed in Table 6.1.

Construction of the pSTOP expression vector

To construct the pSTOP expression vector (Figure 6.2) the 5' phosphorylated oligonucleotides 5'-GAT CCT AAT TAA TTA ACA GCT AGC A-3' and 5'-AGC TTG CTA GCT GTT AAT TAA TTA G-3' were used. After the oligos were annealed, the resulting dsDNA was ligated into the pUC8 expression vector that had been digested with *Bam*HI and *Hin*dIII.

Construction of the pUC8.lacZ vector

To construct the pUC8.*lacZ* vector, midiprep DNA of the pTrc99A.*lacZ* plasmid was restricted with *Bam*HI and *Hin*dIII and gel isolated to yield a ~3.1 Kb fragment containing the *lacZ* gene. This fragment was then ligated to the ~2.6 Kb fragment from the plasmid pUC8 that had been digested with the same two restriction enzymes.

Construction of the pTrc99A.xylE vector

To construct the pTrc99A.*xylE* vector, primers 5'-ATC AGA CTG CAG GAG GTA ACA GCT ATG AAC AAA GGT GTA ATG CGA CC-3' and 5'-TAG CAG TGG CAG CTC TGA AAG CTT TGC ACA ATC TCT GCA ATA AGT CG-3' were used to PCR amplify a 972 bp fragment from the plasmid pXE60, which contains the wild-type *Pseudomonas putida xylE* gene isolated from the TOL pWWO plasmid (Greated *et al*., 2002). The resulting fragment was gel isolated, digested with *Pst*I and *Hin*dIII, and then ligated into the 4164 bp fragment from the plasmid pTrc99A that had been digested with the same two restriction enzymes.

Construction of the pUC18.xylE vector

To construct the pUC18.*xylE* vector, midiprep DNA of the pTrc99A.*xylE* was restricted with *Pst*I and *Hin*dIII and gel isolated to yield a ~950 bp fragment containing the *xylE* gene. This fragment was then ligated to the ~2.6 Kb fragment from the plasmid pUC18 that had been digested with the same two restriction enzymes. Ligation mixtures were transformed into ALS225 and plated on LB + IPTG plates. Transformants were sprayed with 100 mM catechol and yellow $XylE^+$ colonies were selected and checked by gel restriction analysis.

Construction of the pACYC184.xylE vector

To construct the pACYC184.xylE vector, midiprep DNA of the pTrc99A.*xylE* was cut with *Nar*I and *Hin*dIII, gel isolated to yield a ~950bp fragment containing the *xylE* gene, and blunt-ended (polished) with Klenow. This fragment was then ligated to the ~3.2 Kb fragment from the plasmid pACYC184 that had been digested with *Hin*cII. Ligation mixtures were transformed into ALS225 and plated on LB+Cam+IPTG plates. Transformants were sprayed with 100 mM catechol and yellow $XylE^+$ colonies were selected and checked by gel restriction analysis.

SDI cloning scheme for LacZ

The pUC8.*lacZ* construct was digested for 30 minutes with DNAseI, and 10% of the digestion was analyzed via gel electrophoresis to assess the efficiency of the fragmentation, while the remainder was run through a DNA clean-up kit and then polished with Klenow to produce blunt ends for ligation into the pSTOP expression vector that was restricted with *Sma*I, dephosphorylated, and gel isolated. Fragments of *lacZ* were blunt-end ligated with T4 ligase into the 2.67 Kb pSTOP gel isolate. Ligation products were transformed into electrocompetent ALS269 and plated on LB+Amp+IPTG plates. Ten clones were selected at random and miniprep plasmid DNA was isolated and digested with *Hae*II to verify that the library incorporated *lacZ* gene fragments at a frequency of at least 50%. *Hae*II digestion of pSTOP yields three fragments, a 1871 bp large fragment, a 370 bp small fragment, and a 429 bp variable fragment that contains the cloning site for the vector. Upon successful ligation with an average-sized fragment from a DNaseI digested gene library, the variable fragment should show an increase in size of ~200 bp, which is easily ascertained by gel electrophoresis (Figure 6.3).

SDI cloning scheme for XylE

The pUC18.*xylE* construct was digested for 30 minutes with DNaseI, and 10% of the digestion was analyzed via gel electrophoresis to assess the efficiency of the fragmentation, while the remainder was run through a DNA clean-up kit and then polished with Klenow to produce blunt ends for ligation into the pSTOP expression vector that was restricted with *Sma*I, dephosphorylated, and gel isolated. Fragments of *xylE* were blunt-end ligated with T4 ligase into the 2.67 Kb pSTOP gel isolate. Ligation products were transformed into electrocompetent ALS748 and plated on LB+Amp+Cam+IPTG plates. Ten clones were selected at random and

miniprep plasmid DNA was isolated and digested with *Hae*II to verify that the library incorporated *xylE* gene fragments at a frequency of at least 50% (Figure 6.4).

Chemicals and reagents

Blunt-ending reactions were carried out using Klenow polymerase from New England Biolabs (Beverly, MA, USA) while ligation reactions were performed using T4 DNA ligase from Invitrogen (Carlsbad, CA, USA). Random gene fragmentation was performed using DNaseI from Promega (Madison, WI, USA). IPTG was obtained from Diagnostic Chemicals Limited (Prince Edward Island, Canada). Catechol was obtained from Sigma Chemicals (St. Louis, MO, USA).

Results

Screen for Self-Directed Inhibitors of LacZ

The library of randomized ~200 bp fragments of the lacZ gene encoding ~66-amino acid peptide fragments of β-galactosidase was transformed into ALS269 in bulk and 20,000 clones were screened for white $LacZ^-$ colonies among the blue $LacZ^+$ background. No $LacZ^-$ colonies were isolated in this screen.

Screen for Self-Directed Inhibitors of XylE

The library of randomized ~200 bp fragments of the xylE gene encoding ~66-amino acid peptide fragments of catechol-2,3-oxygenase was transformed into ALS748 in bulk and 20,000 clones were screened for white $XylE^-$ colonies among the yellow $XylE^+$ background after the plates had been sprayed with 100 mM catechol. No $XylE^-$ colonies were isolated in this screen.

Discussion

A principal goal in protein folding is to determine the mechanism of folding and the properties of folding intermediates. Protein folding is postulated to begin with the formation of

individual secondary structures (microdomains) such as α-helices and β-sheets (Karplus and Weaver, 1976; Kim and Baldwin 1982). These transient structures are envisioned to be stabilized by packing against each other, later steps continuing in a hierarchical manner until the native tertiary structure is achieved (Baldwin, 1982; Branden and Tooze, 1991). Interestingly, the size of these key microdomains is typically on the scale of 20-30 amino acids in length. It was from this standpoint that the hypothesis of Self-Directed Inhibition (SDI) was developed, as an attempt to approach gene inactivation from the aspect of protein folding. It was proposed that randomized fragments from a gene of interest could be expressed *in vivo* as novel bioactive peptides which might interfere with the correct folding of the native parental protein, thereby rendering it nonfunctional (Figure 6.1). Furthermore, SDI predicts that inhibitory sense-oriented fragments from a gene product of interest would be clustered around regions of the protein critical to folding such as the microdomains described above.

Other scientists have utilized a similar approach to generate inhibitors of various proteins, but have not postulated disruption of protein folding as a causative agent of inhibition. Instead, these researchers indicate that the isolated fragments which have an inhibitory effect on target proteins are found around the active site or demonstrate homology to domains of the protein responsible for oligomerization, DNA binding, transactivation, etc. (Gudkov *et al*., 1993; Ossovskaya *et al*., 1996; Gallagher *et al*., 1997). An examination of some of the gene products for which GSEs were successfully isolated was conducted by comparing the regions where inhibitory fragments clustered with the secondary structures of these proteins. According to the SDI hypothesis, one would expect to see a correlation between GSEs and secondary structures such as α-helices and β-sheets that serve as nucleation sites during early stages of protein folding. Seven sense-oriented fragments of human topoisomerase IIα were isolated from a

20,000-clone library (1 in 2857) and identified as GSEs, and five of these inhibitory peptides were clustered in a region purportedly involved with dimerization and protein-DNA interaction (Gudkov *et al*., 1993). No significant secondary structure is apparent in the region of this protein where GSEs cluster as would be predicted by the SDI model (Figure 6.5). Three sense-oriented fragments of HIV-1 reverse transcriptase were isolated from a library of 80,000 clones (1 in 26,667) and identified as GSEs (Dunn *et al*., 1999). These inhibitory peptides all clustered in a 28 amino acid region lacking secondary structure near the amino terminus of the protein (Figure 6.5). Ten sense-oriented fragments of human p53 tumor suppressor protein were isolated from a 100,000-clone library (1 in 10,000) and identified as GSEs, and five of these inhibitory peptides clustered in a region thought to be responsible for oligomerization (Ossovskaya *et al*., 1996; Gallagher *et al*., 1997). There is an α-helix present in this region consistent with the SDI theory (Figure 6.5). In two of the three cases studied the clustering of GSEs were not found to overlap with regions of the proteins exhibiting secondary structure, thus seeming to disprove the SDI hypothesis. Therefore, it would appear that techniques of this type will probably not contribute to the elucidation of protein folding for gene products of interest, but seem to be better suited to aid in the identification of key functional domains and generation of protein inhibitors. Consistent with this idea is the isolation of truncated loss-of-function mutants of the Wilms tumor suppressor (WT1) that are observed to oligomerize more efficiently or form more stable complexes with full-length WT1 (Moffett *et al*., 1995). Furthermore, the inability to isolate any inhibitory fragments in the screening of two 20,000-clone libraries for two bacterial proteins which spanned a broad range of size may indicate that only a small percentage of protein targets are amenable to these techniques in general.

References

Amann, E., Ochs, B., and Abel, K-J. (1988). Tightly regulated *tac* vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene 69, 301-315.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H., and Falkow, S. (1977). Construction and characterization of new cloning vehicles II. A multipurpose cloning system. Gene 2, 95-113.

Branden, C. and Tooze, J. (1991). Introduction to protein structure. (Garland Publishing, Inc.: New York), pp. 247-267.

Caponigro, G., Abedi, M.R., Hurlburt, A.P., Maxfield, A., Judd, W., and Kamb, A. (1998). Transdominant genetic analysis of a growth control pathway. Proc Natl Acad Sci USA 95, 7508-7513.

Casadaban, M. and Cohen, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J Mol Biol 138, 179-207.

Dunn, S.J., Park, S.W., Sharma, V., Raghu, G., Simone, J.M., Tavassoli, R., Young, L.M., Ortega, M.A., Pan, C.H., Alegre, G.J., Roninson, I.B., Lipkina, G., Dayn, A., and Holzmayer, T.A. (1999). Isolation of efficient antivirals: genetic suppressor elements against HIV-1. Gene Ther 6, 130-137.

Gallagher, W.M., Cairney, M., Schott, B., Roninson, I.B., and Brown, R. (1997). Identification of p53 genetic suppressor elements which confer resistance to cisplatin. Oncogene 14, 185-193.

Greated, A., Lambertsen, L., Williams, P.A., and Thomas, C.M. (2002). Complete sequence of the IncP-9 TOL plasmid pWW0 from *Pseudomonas putida*. Environ Microbiol 4, 856-871.

Gudkov, A.V., Zelnick, C.R., Kazarov, A.R., Thimmapaya, R., Suttle, D.P., Beck, W.T., and Roninson, I.B. (1993). Isolation of genetic suppressor elements, inducing resistance to

topoisomerase II-interactive cytotoxic drugs, from human topoisomerase II cDNA. Proc Natl Acad Sci USA 90, 3231-3235.

Hasan, N., Somasekhar, G., and Szybalski, W. (1988). Antisense RNA does not significantly affect expression of the *galK* gene of *Escherichia coli* or the *N* gene of coliphage lambda. Gene 72, 247-252.

Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. Nature 329, 219-222.

Holzmayer, T.A., Pestov, D.G., and Roninson, I.B. (1992). Isolation of dominant negative mutants and inhibitory antisense RNA sequences by expression selection of random DNA fragments. Nucleic Acids Res 20, 711-717.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F., and Boyer, H.W. (1977). Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science 198, 1056-1063.

Kerr, S.M., Stark, G.R., and Kerr, I.M. (1988). Excess antisense RNA from infectious recombinant SV40 fails to inhibit expression of a transfected, interferon-inducible gene. Eur J Biochem 175, 65-73.

Leiter, J.M., Krystal, M., and Palese, P. (1989). Expression of antisense RNA fails to inhibit influenza virus replication. Virus Res 14, 141-160.

Miller, J.H. (1972). "Experiments in molecular genetics." Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Moffett, P., Bruening, W., Nakagama, H., Bardeesy, N., Housman, D., Housman, D.E., and Pelletier, J. (1995). Antagonism of WT1 activity by protein self-association. Proc Natl Acad USA 92, 11105-11109.

Ossovskaya, V.S., Mazo, I.A., Chernov, M.V., Chernova, O.B., Strezoska, Z., Kondratov, R., Stark, G.R., Chumakov, P.M., and Gudkov, A.V. (1996). Use of genetic suppressor elements to dissect distinct biological effects of separate p53 domains. Proc Natl Acad Sci USA 93, 10309-10314.

Ramalho, J.S., Anders, R., Jaissle, G.B., Seeliger, M.W., Huxley, C., and Seabra, M.C. (2002). Rapid degradation of dominant-negative Rab27 proteins *in vivo* precludes their use in transgenic mouse models. BMC Cell Biol 3, 26.

Roninson, I.B., Gudkov, A.V., Holzmayer, T.A., Kirschling, D.J., Kazarov, A.R., Zelnick, C.R., Mazo, I.A., Axenovich, S., and Thimmapaya, R. (1995). Genetic suppressor elements: new tools for molecular oncology--Thirteenth Cornelius P. Rhoads Memorial Award Lecture. Cancer Res 55, 4023-4028.

Takayama, K.M. and Inouye, M. (1990). Antisense RNA. Crit Rev Biochem Mol Biol 25, 155-184.

Vieira, J. and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19, 259-268.

Warren, J. W., Walker, J. R., Roth, J. R., and Altman, E. (2000). Construction of a highly regulable expression vector, pLAC11, and its multipurpose derivatives, pLAC22 and pLAC33. Plasmid 44, 138-151.

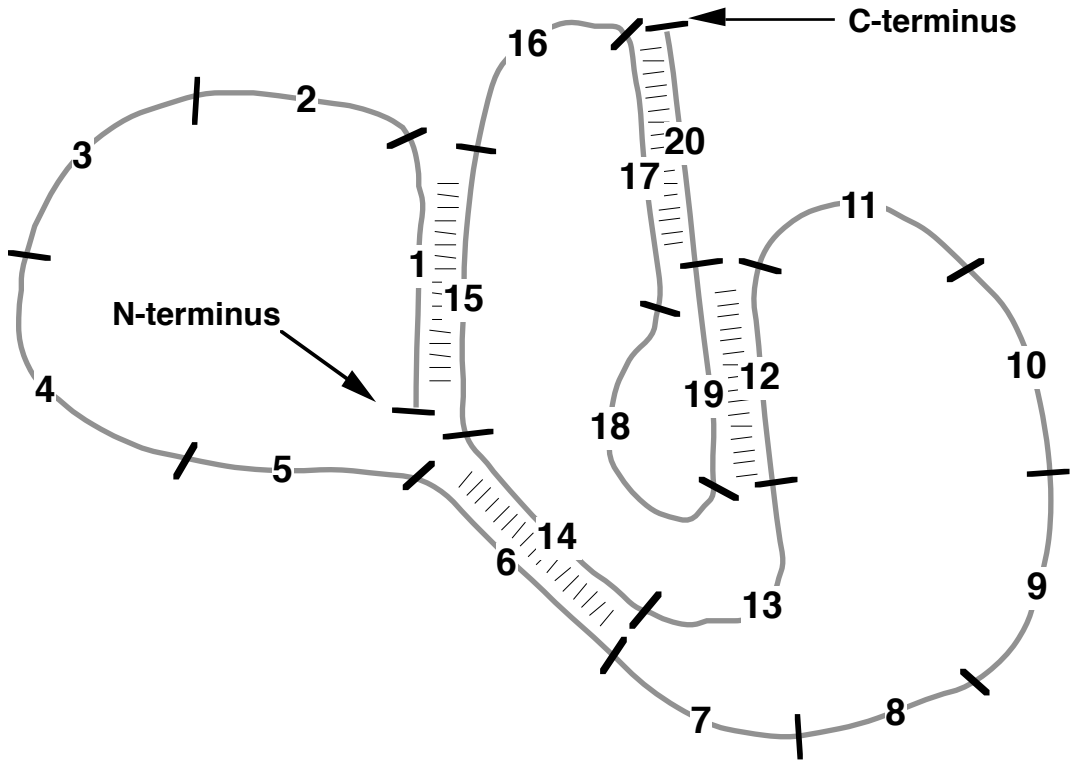


Figure 6.1 Self-Directed Inhibition hypothesis. A hypothetical 600-amino acid protein is shown divided into twenty equal intervals. In this protein, intervals 1 and 15, 6 and 14, 12 and 19, and 17 and 20 interact with each other as depicted in the figure as the protein folds into a final active tertiary form. According to the SDI hypothesis, if any one of these 30-amino acid intervals was expressed independently in a wild-type background, it might have an adverse effect on the folding of the native protein and result in inhibition of biological activity.

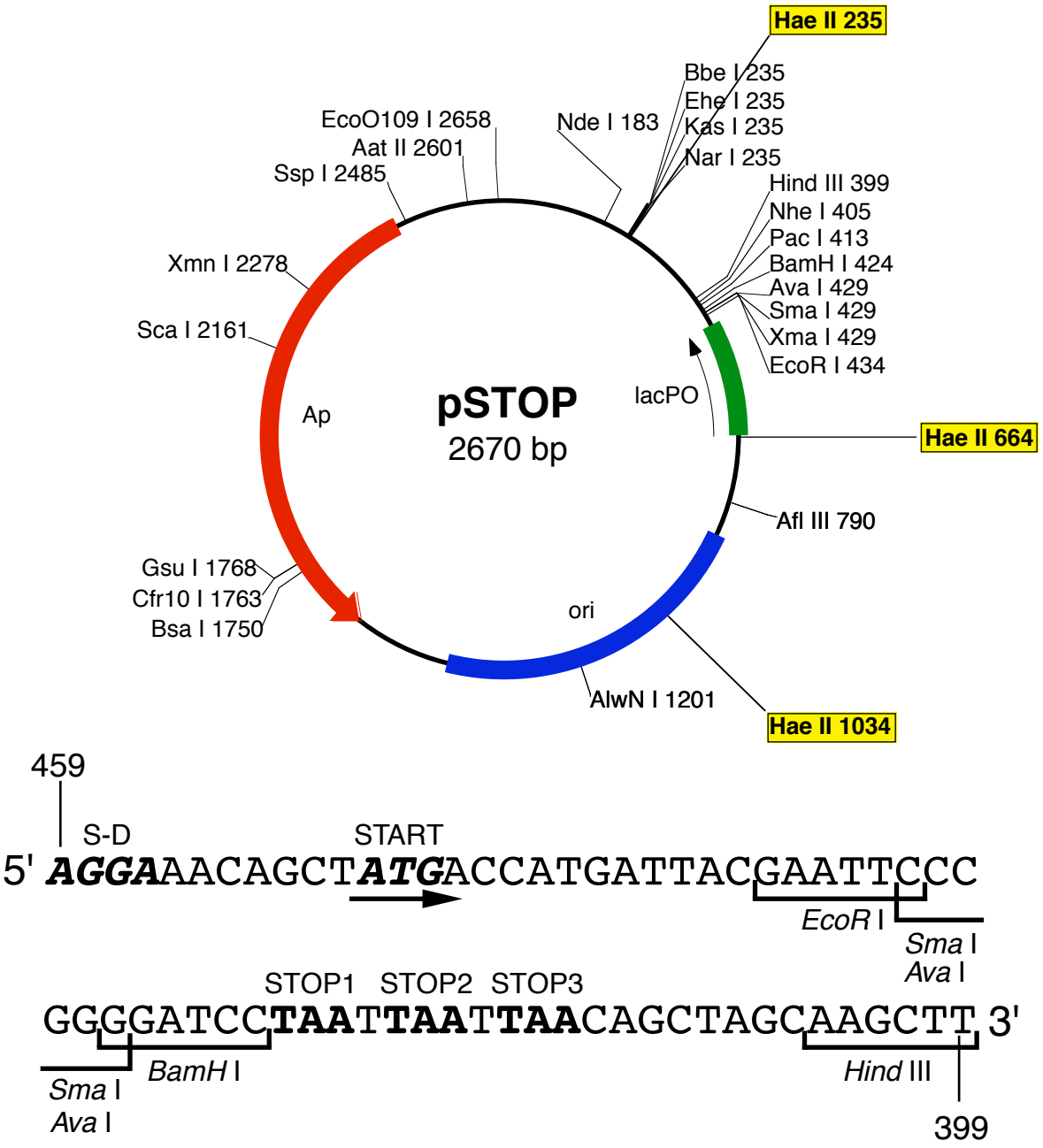


Figure 6.2 **pSTOP vector.** Cloning region designed with stop codons in all three reading frames to express gene fragments of random length in the proper sense.

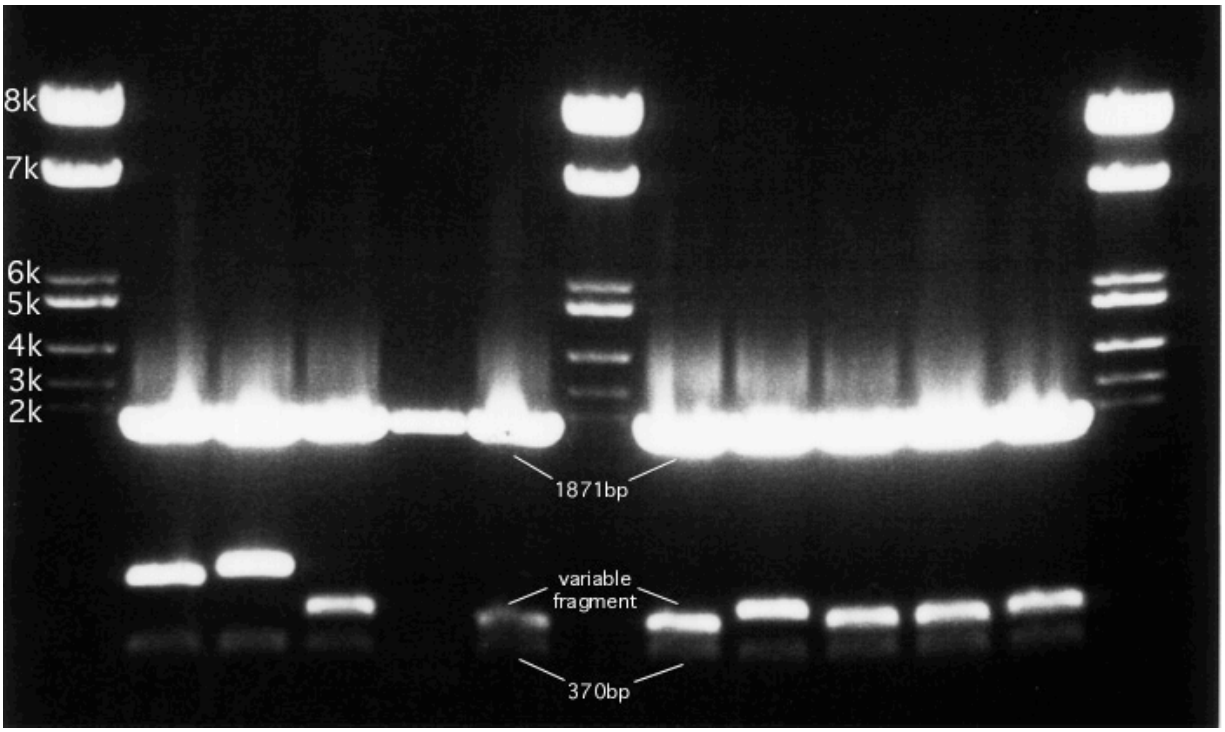


Figure 6.3 Restriction analysis of 10 random clones from LacZ SDI screen

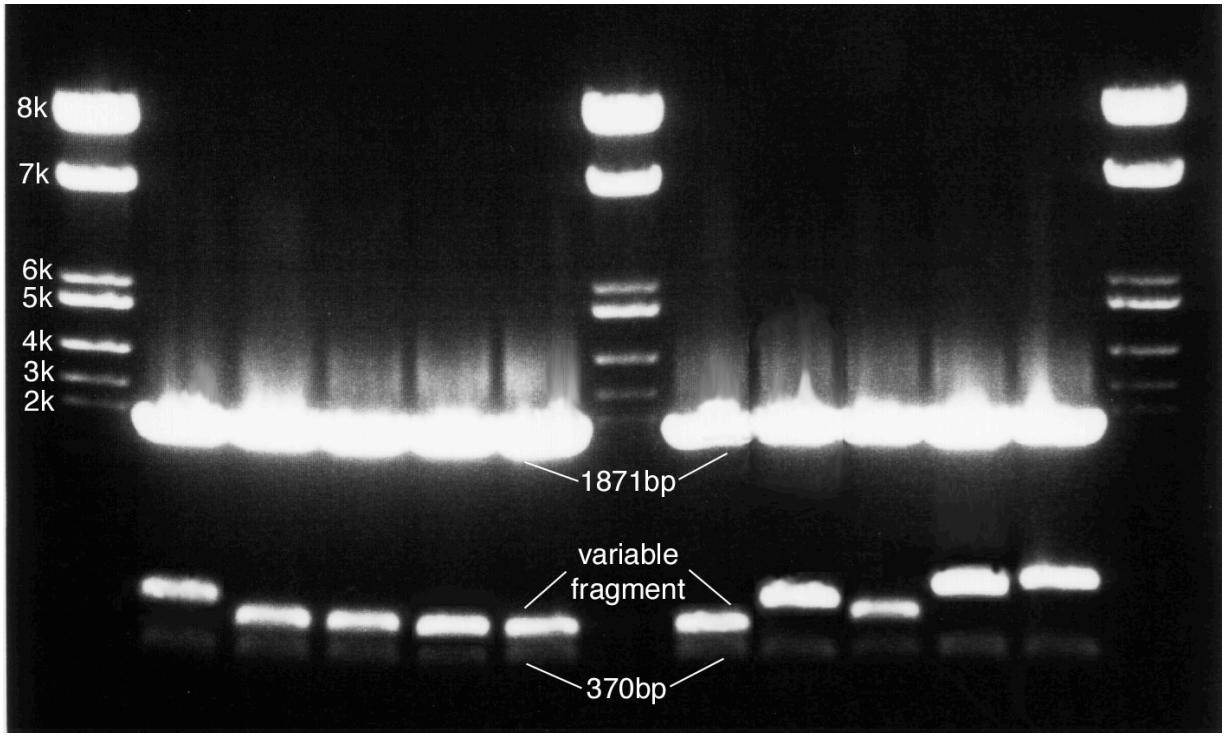


Figure 6.4 Restriction analysis of 10 random clones from XylE SDI screen

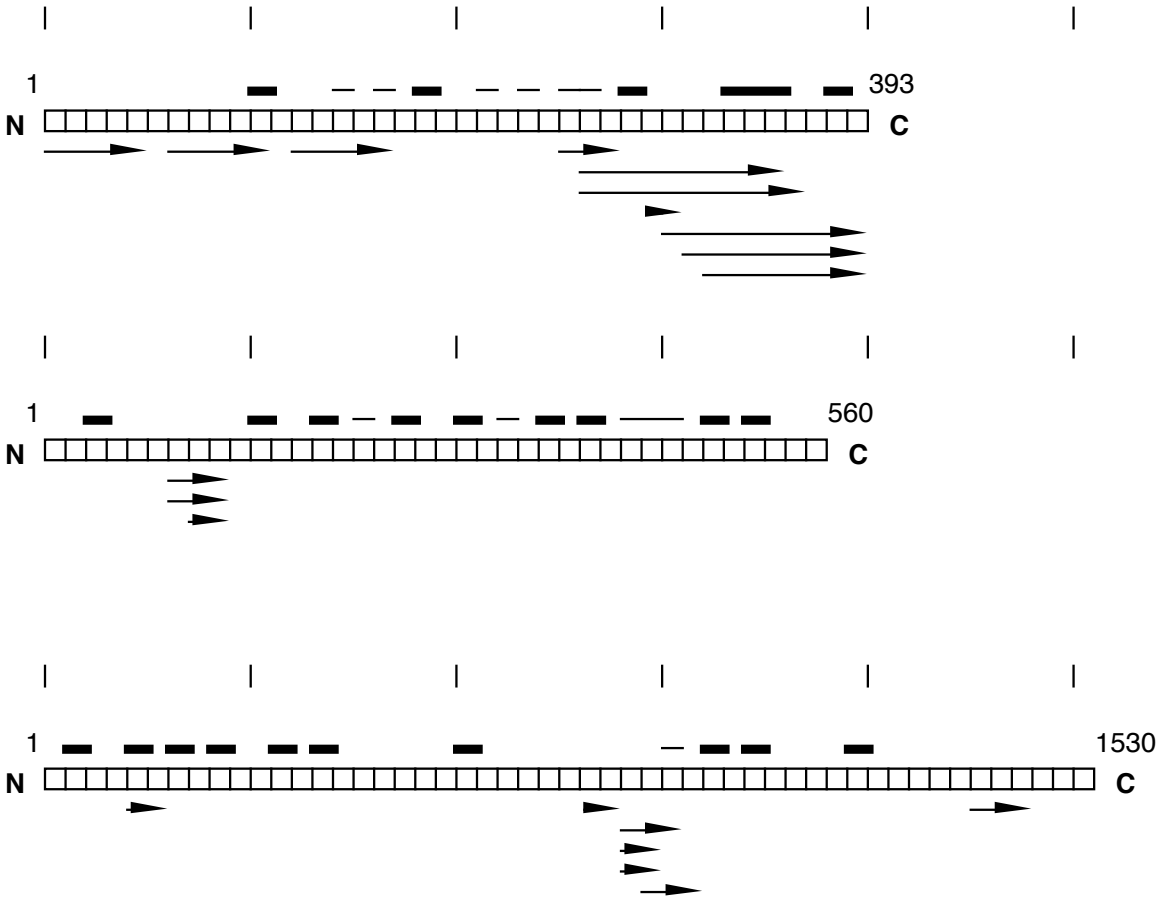


Figure 6.5 GSEs isolated for human p53 tumor suppressor protein (top), HIV-1 reverse transcriptase (center), and human topoisomerase IIα (bottom). Sense-oriented inhibitors are represented by arrows below the protein diagram, α-helices by thick lines above, and β-sheets by thin lines above. Note that protein diagrams are not to scale; p53 is segmented into 10-amino acid blocks, HIV-1 RT into 15-amino acid blocks, and topo IIa into 30-amino acid blocks. Data described in Gudkov *et al*., Ossovskaya *et al*., Gallagher *et al*., and Dunn *et al*.

Table 6.1 Bacterial strains and plasmids used in this work

Bacterial strains (*E. coli*)			
Laboratory name	Original name	Genotype	Reference or source
ALS224	MC1061	*araD139* Δ*(araABOIC-leu)7679* Δ*(lac)X74 galU galK rpsL hsr- hsm+*	Casadaban and Cohen, 1980
ALS225		MC1061 / F'$lacI^{Q1}$ Z^+ Y^+ A^+	Warren *et al*., 2000
ALS269	CSH27	F- *trpA33 thi tyr*	Miller, 1972
ALS528		ALS224 pXE60	This work
ALS748		ALS225 pACYC184-xylE	This work
ALS524	XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* / F' *proAB* $lacI^Q$ Δ*(lacZ)M15* Tn10	Stratagene

Plasmids		
Plasmid name	Relevant characteristics	Reference or source
pACYC184	wild-type *lac* promoter / operator, Cam^R, Tet^R, ColE1 replicon	Itakura *et al*., 1977
pBR322	Amp^R, Tet^R, ColE1 replicon	Bolivar *et al*., 1977
pTrc99A	*trc* promoter / operator, $lacI^Q$, Amp^R, ColE1 replicon	Amann *et al*., 1988
pUC8	*lac* promoter / operator, Amp^R, ColE1 replicon	Vieira and Messing, 1982
pUC18	*lac* promoter / operator, Amp^R, ColE1 replicon	W. Gilbert
pXE60	wild-type TOL pWWO *xylE* gene, Amp^R	J. Westpheling

CHAPTER 7

CONCLUSIONS

The development of bioactive peptides as clinically useful therapeutics has been greatly hindered by their poor metabolic stability *in vivo*. The aim of this thesis was to investigate protein-based motifs and their potential for protecting peptides against degradation. A novel, highly regulable expression vector, pLAC11, was designed and constructed to allow potent growth inhibitory peptides to be isolated from *E. coli* bacteria utilized in a genetic system devised to screen randomized peptides. In comparative studies with other commonly used expression vectors, pLAC11 successfully demonstrated the regulable nature necessary to express inhibitory clones which might otherwise kill the cells producing the bioactive peptides if the expression could not be turned off completely.

During early studies, a deletion event occurred in the plasmid that caused randomized peptides to be fused to the highly stable 63-amino acid Rop protein, producing two extremely potent inhibitors. To investigate the deliberate implementation of this potentially protective motif, a vector was designed to fuse Rop to either the amino or carboxyl terminus of a randomized 20-amino acid peptide. Screening combined libraries totaling 16,000 clones it was found that the fusion of Rop to either end of randomized peptides resulted in the isolation of potent inhibitors at an averaged frequency of 1 in 527, a 38-fold increase compared to the 1 in 20,000 found when peptides are left unprotected. This seemed to indicate that Rop might aid in stabilizing peptides against enzymatic degradation *in vivo*.

It was postulated that the helix-turn-helix structure which lends Rop its highly stable nature might have been the key to its apparent ability to protect bioactive peptides, and it was decided to investigate other stable secondary structures which might have similar effects. Oligonucleotides were designed for use in the *in vivo* genetic screen to yield peptides with a terminal opposite charge motif flanking a randomized core of residues and peptides composed of randomized hydrophilic α-helix-forming amino acids, and these strategies increased the frequency at which potent inhibitors were isolated, by 6-fold and 8-fold respectively.

In a search for more stable protective motifs, a group of naturally-occurring α-helices based on Rop and other helical proteins as well as potential helical peptides designed *de novo* were investigated through the use of predictive secondary structure algorithms, CD spectroscopic analysis and *in vitro* rat plasma degradation assays. Two highly stable helical peptides were identified which exhibited half-lives 52-fold greater than other representative peptides and similar to that of small stable proteins. These peptides could potentially be adapted to serve as protective protein-based anchors to help stabilize bioactive peptides.