A STRUCTURAL INVESTIGATION OF THE CHEMOKINE DIMER INTERFACE

A Senior Honors Thesis

by

Garret Lance Hayes

Submitted to the Office of Honors Programs
& Academic Scholarships
Texas A&M University
in partial fulfillment of the requirements of the

UNIVERSITY UNDERGRADUATE
RESEARCH FELLOWS

April 2003

Group: Life Sciences 1

A STRUCTURAL INVESTIGATION OF THE CHEMOKINE DIMER INTERFACE

A Senior Honors Thesis

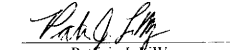
by

Garret Lance Hayes

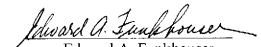
Submitted to the Office of Honors Programs
& Academic Scholarships
Texas A&M University
in partial fulfillment of the requirements of the

UNIVERSITY UNDERGRADUATE
RESEARCH FELLOWS

Approved as to style and content by:



Patricia J. LiWang
(Fellows Adviser)



Edward A. Funkhouser
(Executive Director)

April 2003

Group: Life Sciences 1

ABSTRACT

A Structural Investigation of the Chemokine Dimer

Interface. (April 2003)

Garret Lance Hayes
Department of Biochemistry
Texas A&M University

Fellows Adviser: Dr. Patricia LiWang
Department of Biochemistry

Macrophage Inflammatory Protein(MIP)-1β is a dimeric protein of the CC chemokine subfamily. Interest in both the structure and function of MIP-1β has increased rapidly over the past several years, resulting primarily from the discovery that MIP-1β has the ability to block infection of white blood cells by HIV. This discovery has instigated further inquiry into exactly how the structure of MIP-1β correlates to its function and, likewise, to its anti-HIV capacity.

Of particular interest to us is the contrast between the structure of MIP-1β and a similar dimeric protein from the CXC chemokine subfamily, Interleukin-8 (IL-8). A comparison of the monomeric forms of IL-8 and MIP-1β shows a striking similarity between the two proteins. By examining the monomeric subunits of these two proteins, one would expect that similarity in their quaternary structure would follow suit. However, it has been shown that dimers of MIP-1β and IL-8 differ greatly, with one interacting along the first β-strand and C-terminal helix (IL-8) and the other along the N-

terminus (MIP-1β). Such contrasting quaternary structures for essentially identical monomeric subunits is quite rare.

In order to learn more about the protein-protein interactions that lead to dimerization and to more fully understand the chemokine dimmer interface, we attempted to alter the sequence of MIP-1β such that it will dimerize in like fashion to IL-8. Previous rational mutations have been successful in producing folded monomers but led us to believe that a strategy involving random mutagenesis was required for formation of the alternate dimer. Randomization coupled with a phage λ selection system was used to search for MIP-1β variants that dimerize like IL-8. This process has thus far produced several variants currently under investigation by NMR. Preliminary data suggests that these proteins may exhibit some type of folded, alternate structure.

This work is dedicated to the teachers of Vidor High School. Thank you for showing me the thrills of learning.

ACKNOWLEDGEMENTS

I would like to thank my thesis adviser, Dr. Patricia LiWang, for her continued support and encouragement. Her overflowing enthusiasm for science has continually inspired me to seek out answers to the difficult questions. What inspired her to take a chance on an inexperienced freshman the world may never know, but I think, I hope, the journey has been exciting for the both of us.

I would also like to express my gratitude to all members of the LiWang laboratory, especially Craig Cassidy, Ioannis Vakonakis, Melissa McCornack, Erik Meyer, and Chris Jao. You have been there to answer questions, provide support, lend advice, and, when needed, dole out criticism. I am confident that my undergraduate experience has been a richer one because of your involvement. Melissa and Erik deserve a special thanks for all their help on this thesis.

I am also indebted to the lab of Dr. James Hu of Texas A&M University for their continued assistance. A special thanks is due to Leonardo Marino-Ramirez for his guidance and seemingly endless patience.

Finally, I would like to thank my parents for their absolute support in most everything I do. Their belief in my abilities has always been a comfort and given me the confidence to persist in my ambitions. I can only hope to have been a good steward of their trust.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

A_N	Absorbance at N nm
AIDS	Acquired Immune Deficiency Syndrome
Amp	Ampicillin
AU	Absorbance units
BSA	Bovine Serum Albumin
CCR	CC Receptor
CXCR	CXC Receptor
dd	distillied deionized
DNA	Deoxyribonucleic acid
EDTA	Ethylene diaminetetraacetic acid
EtBr	Ethidium bromide
GAG	Glycosaminoglycan
h; min; s	hour(s); minute(s); second(s)
HIV	Human Immunodeficiency Virus
HV1	Helix Variant 1
HV2	Helix Variant 2
HV3	Helix Variant 3
HV4	Helix Variant 4
HPLC	High Performance Liquid Chromatography

HSQC	Heteronuclear Single Quantum Coherence
IL-8	Interleukin 8
IPTG	Isopropyl-β-D-thiogalactopyranoside
Kan	Kanamycin
L; ml; ml	Liters, milliliters, microliters
LB	Luria Broth
M; mM	Molar, millimolar
LV1	Loop Variant 1
LV2	Loop Variant 2
LV3	Loop Variant 3
MCP-1	Monocyte Chemoattractant Protein 1
MGSA	Melanoma Growth-Stimulatory Activity
MIP-1α	Macrophage Inflammatory Protein 1α
MIP-1β	Macrophage Inflammatory Protein 1β
nm	nanometers
NMR	Nuclear Magnetic Resonance
OD_N	Optical density at N nm
PCR	Polymerase chain reaction
PF-4	Platelet Factor 4
pfu	Plaque forming units
RANTES	Regulated upon Activation of Normal T-cell Expressed and Secreted

RLV1	Rational Loop Variant 1
RLV2	Rational Loop Variant 2
RNA	Ribonucleic acid
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
TM	Tris-Magnesium
TFA	Trifluoroacetic acid
WT	Wild-type

CHAPTER I

INTRODUCTION

An Historical Perspective on Protein Structure

Over the past half century, structure determination has proven to be an incredibly powerful tool allowing deep insight into the biological minutiae of the living world. The strides taken in the capacity for structure elucidation are nothing less than phenomenal and highlight the rise of structural biology to the forefront of modern science (Figure 1.1) (Smith, 1999; Wuthrich, 2001). With the increased ability to determine structure quickly and accurately, a massive library of structures is hurriedly accumulating that serves as a guide to drug design, mechanism elucidation, and protein function.

However, the ability of scientists to *predict* structure is much less tangible and underlines the need for even more inquiries into the fundamental properties that lead to protein folding (Heringa, 2000). The search for an all-encompassing structural viewpoint is as fervent as ever. Closely linked to this general investigation of the structural paradigm is the study of protein dimerization. While much is known about what *influences* dimer association and dissociation, the process of *predicting* how proteins dimerize is still extremely challenging (Cherfils and Janin, 1993; Janin, 1995; Shoichet and Kuntz, 1996; Sternberg *et al.*, 1998).

This thesis follows the style and format of the *Journal of Biomolecular NMR*.

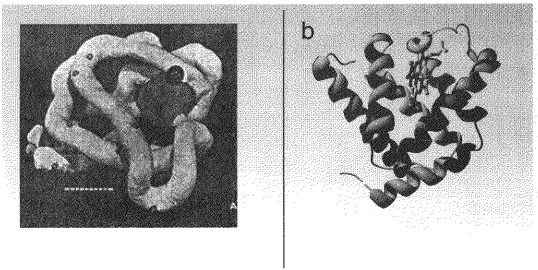


Figure 1.1 The evolution of structure. Shown above is the first X-ray crystallography structure determined, myoglobin (Kendrew *et al.*, 1958) (**a**). Adjacent (**b**) is a more recent high resolution structure of myoglobin (Takano, 1984). Reprint permission of Figure (**a**) granted by the Nature Publishing Group.

Conveniently then, the study of dimerization is one satisfying method to investigate protein folding. By studying the processes and interactions that lead to dimerization, a firmer grasp of the generic rubric that governs protein folding can be realized; such an investigation is the general intent of this thesis. Further elucidation of the general protein-protein interactions that lead to dimerization was accomplished using a class of secreted proteins, the chemokines, as models of study.

Chemokines

Chemokines are small (~8-10kDa), secreted proteins involved in the human immune response. Derived from the phrase *chemo*tactic cyto*kines*, "chemokine" aptly describes their role in the recruitment of white blood cells to sites of infection via chemotaxis. In addition to their part in immune system response, chemokines are also involved in several immunological diseases, including rheumatoid arthritis, ateriosclerosis, and most notably AIDS (Baggiolini *et al.*, 1997; Burke-Gaffney *et al.*, 2002; Dragic *et al.*, 1996; Mellado *et al.*, 2001; Patel *et al.*, 2001; Volin *et al.*, 1998).

The chemokine family consists of over 50 proteins grouped into two broad major subfamilies (CC and CXC) and two minor subfamilies (C and CX_3C) according to the arrangement of conserved cysteine residues near the N-terminus. The CC major subfamily contains those chemokines with contiguous cysteine residues while the CXC major subfamily consists of chemokines with an interceding residue between the conserved cysteines. As will be demonstrated, chemokines within a subfamily share

characteristic structural properties (Clore and Gronenborn, 1995; Fernandez and Lolis, 2002; Rollins, 1997).

Some well known chemokines include RANTES, MIP-1β, and MIP-1α (CC subfamily) as well as IL-8 and PF-4 (CXC subfamily). Figure 1.2 details some of the roles and properties of common chemokines.

Chemokine Structure

Chemokine structure has been well documented over the past decade, leading to a wide breadth of knowledge about the structural characteristics that are conserved between various chemokines and those unique properties indicative of one particular chemokine or chemokine subfamily. Elucidation of the high resolution structure of over 20 chemokines has permitted extensive characterization of the conserved chemokine structure (Clore *et al.*, 1990; Chung *et al.*, 1995; Crump *et al.*, 1997; Crump *et al.*, 1998; Handel and Domaille, 1996; Kim *et al.*, 1994; LiWang *et al.*, 1999; Lodi *et al.*, 1994; Mayer and Stone, 2000; Meunier *et al.*, 1997; Mizoue *et al.*, 1999; Qian *et al.*, 1999; Rajarathnam *et al.*, 2001; St. Charles *et al.*, 1989; Sticht *et al.*, 1999; Young *et al.*, 1999). These structures demonstrate that chemokines exhibit a conserved monomeric fold consisting of three main portions:

a.) a highly disordered N-terminal region followed by

b.) a β-sheet comprised of three anti-parallel β-strands and

c.) a C-terminal α-helix

	Subfamily	Chromosomal location	Target Cells	HIV inhibition?
RANTES	CC	17q	Monocytes, T lymphocytes, Basophils, Eosinophils, NK cells, Dendritic cells, Mast cells	YES
MIP-1β	CC	17q	Monocytes, T lymphocytes, Hematopoietic precursor cells, Basophils	YES
MIP-1α	CC	17q	T lymphocytes, Basophils, Hematopoietic precursor cells, Monocytes, Eosinophils, Neutrophils, Mast cells, NK cells, Epidermal keratinocytes, Astrocytes, Spermatogonia, Dendritic cells, Osteoclasts	YES
PF-4	CXC	4q	Fibroblasts, Platelets, Adrenal microvasculas pericytes, Mast cells, Basophils,Megakaryocytes	NO
IL-8	CXC	4q	Neutrophils, T-lymphocytes, Basophils, Eosinophils, Keratinocytes, HUVECs	NO
MCP-1	CC	17q	Monocytes, Hematopoeitic precursors, T lymphocytes, Basophils, Eosinophils, Mast cells, NK cells, Cardiac myocytes, Osteoclasts, Microglial cells, Dendritic cells	NO
MGSA	CXC	4q	Neutrophils, Lymphocytes, Monocytes, Epidermal melanocytes	NO

Figure 1.2 Descriptions of some well-studied chemokines and their properties (Vaddi *et al.*, 1997).

An example of this characteristic fold is given in Figure 1.3.

The quaternary structure of chemokines is more variable than the well-conserved monomeric fold. Although many chemokines do exhibit some type of quaternary structure (for monomeric exceptions see (Crump *et al.*, 1997; LiWang *et al.*, 1999; Mizoue *et al.*, 1999; Rajarathnam *et al.*, 2001; Shao *et al.*, 1998; Sticht *et al.*, 1999)), dimerization among monomeric subunits is segregated according to subfamily membership. Those chemokines belonging to the CC subfamily form dimer contacts primarily along the N-terminus, thus forming interactions between the amino end and the opposite subunit. In contrast, CXC chemokine dimers are constructed primarily by interactions along the first β-strands of each unit and between the C-terminal helix and the flexible loop region between β-strand 1 and 2 (Baggiolini, 1997; Fernandez and Lolis, 2002). These two dimer types are illustrated in Figure 1.4.

The biological relevance of the chemokine dimer is not clearly known (Laurence *et al.*, 2000; Paavola *et al.*, 1998). The inability to demonstrate receptor cross-binding by chemokines of different subfamilies does suggest that dimeric structure may play a role in receptor recognition (Baggiolini, 1997; LiWang *et al.*, 1999). In addition, the importance of the N-terminus in receptor activation seems to imply that the N-terminal oriented dimers of CC chemokines might have some effect on biological function (Baggiolini, 1997; Clark-Lewis *et al.*, 1995; Crump *et al.*, 1997; Gong and Clark-Lewis, 1995; Gong *et al.*, 1996; Pakianathan *et al.*, 1997; Proudfoot *et al.*, 1996; Simmons *et al.*, 1997; Steitz *et al.*, 1998; Struyf *et al.*, 1997; Struyf *et al.*, 1999; Weber *et al.*, 1996; Zhang *et al.*, 1994).

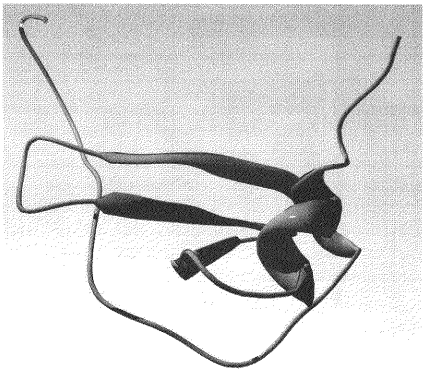


Figure 1.3 The structure of the CC chemokine RANTES. This view demonstrates the typical "chemokine" fold: a disordered N-terminus, three antiparallel β-strands, and a C-terminal α-helix (Skelton *et al.*, 1995).

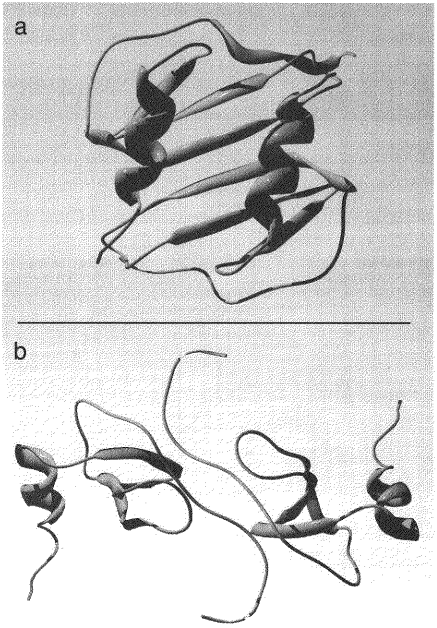


Figure 1.4 Views of the two most common chemokine dimers. At top is the CXC chemokine PF-4 (**a**). At bottom is the CC chemokine RANTES (**b**). These dimer forms are well conserved throughout the chemokine subfamilies (Skelton *et al.*, 1995; St. Charles *et al.*, 1989)

Biological Role of Chemokines

The chemokines play an active role in the human immune system, mainly as recruiters of white blood cells to sites of infection (Fernandez and Lolis, 2002). All chemokines bind to seven transmembrane domain G-protein coupled receptors linked to the initiation of chemotaxis (Baggiolini *et al.*, 1994; Murphy, 1994). Generally, these receptors are designated depending on the ligand subfamily (e.g. CCR or CXCR). The exact nature of chemokine receptor binding is currently unclear. While some evidence does point toward a "surface binding" model, the specific interactions that occur between the chemokine and receptor are still unclear (Crump *et al.*, 1997; Meyer and Stone, 2001; Pakianathan *et al.*, 1997). Furthermore, the role that chemokine quaternary structure plays in receptor binding is equally nebulous. Recent studies have shown that with some CC chemokines, monomeric variants can bind and activate the CCR receptor (Laurence *et al.*, 2000).

In addition to the significant role that chemokines play in the immunological response, the binding of chemokines to seven transmembrane domain receptors also has a dramatic impact on HIV infectious progression. In order to gain entrance to white blood cells and proceed with viral infection, HIV recognizes a set of extracellular receptors that often include chemokine receptors. Therefore, inhibition of HIV binding can be accomplished by binding of chemokines to their native receptors. For example, the chemokine MIP-1β recognizes the CCR5 extracellular receptor. By binding to this receptor, MIP-1β can exclude the viral particle from gaining cellular access (Figure 1.5) (Baggiolini *et al.*, 1997).

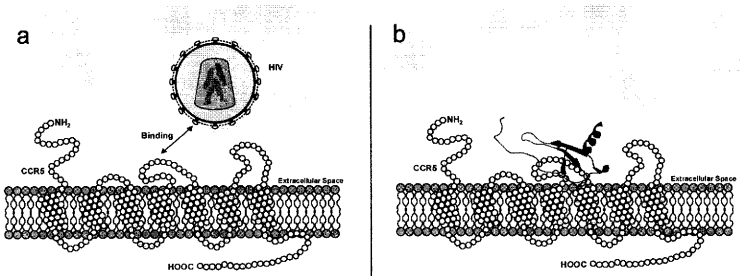


Figure 1.5 The exclusion of HIV from the CCR5 receptor. This figure demonstrates how the CC chemokine MIP-1β excludes a viral particle from gaining cellular access. Because the CCR5 receptor is required for HIV entrance into the host, binding of MIP-1β to its native receptor obstructs HIV from achieving proper CCR5 access. This type of HIV interference occurs in many chemokines.

Chemokines have also been shown to exhibit binding to glycosaminoglycans (GAGs), a process which is intricately linked to the chemotaxis of white blood cells (Tanaka *et al.*, 1993; Springer, 1994). The GAG heparin is uniquely interesting since it has been shown that its presence is required for the full anti-HIV capacity of some chemokines (Burns *et al.*, 1999; Wagner *et al.*, 1998). The GAG-chemokine interaction is thought to be mainly a function of electrostatic attraction between the negatively charged GAG particle and positively charged residues on the chemokine surface (Amara *et al.*, 1999; Kuschert *et al.*, 1999; Koopman and Krangel, 1997; Laurence *et al.*, 2001; Lortat-Jacob *et al.*, 2002). Recent investigations on MIP-1β have further characterized this interaction and produced the first NMR spectra at neutral pH of an anti-HIV CC chemokine interacting with GAGs (McCornack *et al.*, 2003).

Using Chemokines to Study Structure

In what way do can chemokines be used to study the general problem of structure determination and more precisely, the processes that lead to dimerization? The investigation detailed in this thesis will focus primarily on two chemokines: Macrophage Inflammatory Protein-1β (MIP-1β) and Interleukin-8 (IL8). These two peptides will be used to investigate the nature of the chemokine dimer interface and, more broadly, the general interactions that generate dimers. As will be described, the structural characteristics of these two proteins provide an interesting means to examine at the dimerization paradigm.

The Structures of MIP-1β and IL8

A cursory comparison of the individual, monomeric structures of CXC chemokine IL-8 and CC chemokine MIP-1β shows a striking similarity between the two proteins; each contains a disordered N-terminal region followed by three anti-parallel β-strands and a C-terminal α-helix (Figure 1.6) (Baldwin *et al.*, 1991; Clore *et al.*, 1990; Lodi *et al.*, 1994). The sequence identity between MIP-1β and IL-8 in 59 selected residues (Figure 1.7) is 20%; the root mean square difference between Cα atoms in these regions is only 1.6Å (Lodi *et al.*, 1994). Based on this cursory examination, it is not unreasonable to further conclude that similarity in the intermolecular, quaternary structure would also be demonstrated. However, the functional dimers of MIP-1β and IL-8 differ greatly in structure: one dimerizes along the first β-strand (IL-8) while the other forms contacts along the N-terminus (MIP-1β) (Figure 1.8). The type of quaternary structure exhibited is indicative of the parent subfamily of each chemokine: IL-8 shows a dimerization typical of the CXC subfamily and MIP-1β association is indicative of the CC chemokines. However, such contrasting quaternary structures for essentially identical monomeric subunits is quite rare; very few instances exist where two structures containing greater than 15% sequence similarity dimerize so differently (Lodi *et al.*, 1994; Royer *et al.*, 1985; Shaanan, *et al.*, 1991).

The contrast between MIP-1β and IL8 dimers immediately proposes a tempting question of protein engineering: given the very similar monomeric structures of the two proteins, can one of the chemokines be altered to form a dimer resembling the other?

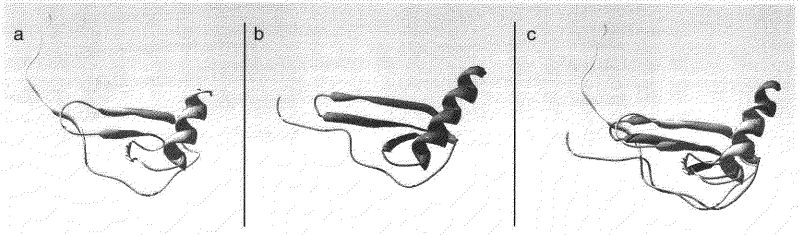


Figure 1.6 MIP-1β and IL-8 compared. This figure demonstrates the monomeric similarity between MIP-1β (**a**) and IL-8 (**b**). Both demonstrate the standard chemokine fold and when overlayed (**c**) present a fair amount of structural consistency (Clore *et al.*, 1990; Lodi *et al.*, 1994).

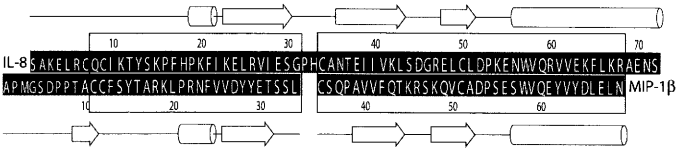


Figure 1.7 Sequence comparison of MIP-1β and IL-8. The amino acids sequences and secondary structure of MIP-1β and IL-8 are compared in this figure. The boxed regions indicate areas where sequence identity is 20%; sequence similarity is approximately 35%. Regions not included in the boxed areas are the variable N-termini, the end of the IL-8 helix, and a portion of the IL-8 flexible loop connecting strands β1 and β2. These regions are the areas of greatest variation between the two molecules (Lodi *et al.*, 1994).

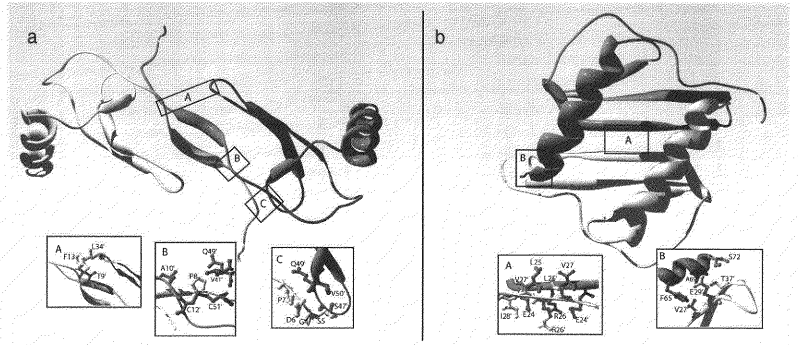


Figure 1.8 MIP-1β and IL-8 dimers. Here is shown a comparison between the MIP-1β (**a**) and IL-8 (**b**) quaternary structures. Despite having very similar monomeric forms, these two molecules dimerize very differently. MIP-1β forms dimer contacts primarily along the N-terminus while IL-8 relies on interactions of the first β-strands as well as the helix-loop contacts. Inset boxes more closely describe the pertinent residues in dimer formation (Clore *et al.*, 1990; Lodi *et al.*, 1994).

Such an endeavor would undoubtedly shed much light on the nature of the chemokine dimer interface and provide information on the role that the chemokine dimer plays in biological function. In addition, the task of switching the chemokine dimer will help to garner general knowledge of what is necessary to form an alternate quaternary structure.

Previous Strategies

Attempts have previously been made to alter the MIP-1β dimer from its native form to an arrangement that more closely resembles the IL-8 dimer. Rational protein design was used to logically manipulate the MIP-1β sequence in order to promote the formation of an alternate dimer. Since most CC chemokines require an intact N-terminus to accomplish dimerization, all rational designs lacked the first eight amino acids of the MIP-1β molecule; this ensured that no wild-type dimerization could occur. Various other mutations were then attempted to force the assumption of an IL-8 dimer. These attempts and their relative successes are summarized in Figure 1.9. Particular notice should be taken of the variant labeled MIPSTART as it will factor greatly into the future investigations of this project. The inability of rational mutation to successfully produce a folded alternate dimer led to the conclusion that a different approach was needed to realize formation of the alternate dimer.

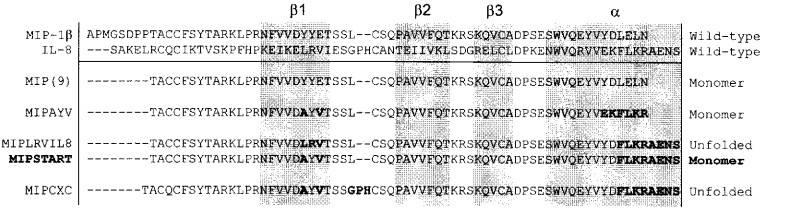


Figure 1.9 Summary of previous rational mutations. This figure depicts the previous attempts at MIP-1β rational mutation to form an alternate IL-8 type dimer. Listed at the top are the wild-type sequences of MIP-1β and IL-8. The mutants created and their names are listed below the wild type sequences. To the right are listed the dimerization results of the mutations. All mutants were lacking the first eight amino acids of the N-terminus. This was done to prevent a return to the wild-type MIP-1β dimer. The mutant listed as MIPSTART will be the template for future randomization schemes.

The Specific Purpose of this Study

As explained, previous rational mutations of MIP-1β have been successful in producing a set of folded monomers but have failed to produce a stable, alternate IL-8 type dimer. Therefore, random mutagenesis was proposed as a method to produce variants that can be assessed for IL-8 type dimerization. Such an assessment was accomplished by implementing a system that easily selects for dimerization. The previously designed variant MIPSTART was chosen as the beginning template for randomization. The impetus for selecting this variant as the randomization template was as follows:

a.) MIPSTART is fully monomeric and lacks the first eight residues of the N-terminus. Randomization is unlikely to cause reversion back to a wild-type MIP-1β dimer as this would require the addition of N-terminal residues. Therefore, any dimers produced by randomization must be either IL-8 in fashion or an equally intriguing alternate arrangement.

b.) MIPSTART already contains some elements of the IL-8 monomer structure that in theory, help promote the formation of the IL-8 type dimer (Figure 1.10).

Using MIPSTART as a template for random mutagenesis in combination with a dimerization selection system permitted insight as to why two proteins with such comparable monomeric structure form drastically different dimeric arrangements. Nuclear magnetic resonance (NMR) spectroscopy was used to analyze any dimeric MIPSTART variants produced by the randomization/selection process. Accomplishing

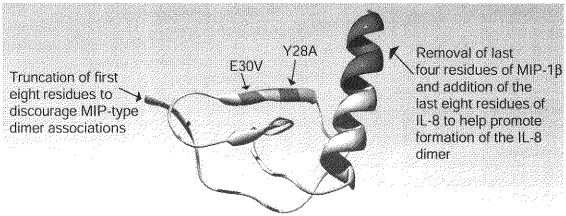


Figure 1.10 Structure of MIPSTART. Graphical representation of MIPSTART illustrating the salient structural features. Regions altered to encourage IL-8 type dimerization are noted. The two residue changes on β1 (E30V and Y28V) are an attempt to produce a more hydrophobic area that might assist subunit interaction. The hybrid nature of this molecule was the reason for its selection as the randomization template.

this goal allowed further understanding into what exactly is required to switch the quaternary chemokine arrangement and gave insight into the general properties of the chemokine dimer interface.

CHAPTER II

THE PHAGE λ SYSTEM AND RANDOMIZATION

Introduction

The previous chapter detailed that this project will focus on the randomization of a monomeric MIP-1β variant (MIPSTART) with the intent of forcing an IL-8 type dimer. The randomization of MIPSTART obviously necessitates the implementation of some selection system that can expedite the process of mutant analysis. There is a high probability that of the thousands of mutants produced by the randomization process, only a few will exhibit dimerization. Since direct structural analysis of thousands of mutants by NMR would be nearly impossible, a system that can dependably analyze mutants is needed. Ideally, this system would possess three main characteristics:

a.) Ability to distinguish between the monomeric and homodimeric forms of a MIP-1β variant

b.) Employment of a selection process through which quaternary form can be easily, quickly, and confidently detected

c.) Exhibit high-throughput capability that permits efficient selection of thousands of variants within a feasible time table

Several different systems were analyzed in order to find one best tailored to the project scope; a selection system based on the bacteriophage λ was chosen as the most appropriate system for the goal at hand. The phage λ selection system, developed by Dr.

Jim Hu of Texas A&M University and Dr. Robert Sauer of the Massachusetts Institute of Technology, is of great relevance to this endeavor because of its ability to quickly and confidently detect the presence of homodimers. At the time of the project inception, the λ system had been used to study dimerization of defined mutants (Zhang *et al.*, 1999) and to study interactions among genomic proteins (Marino-Ramirez and Hu, 2002), but had never been applied to a high-throughput analysis of random mutant libraries. The implementation of the λ system as well as a solution to the high-throughput problem will now be addressed.

Life Cycle of Phage λ

The bacteriophage λ is an infectious agent of *E.coli* that exploits and manipulates the host machinery to promote reproduction and proliferation. The life cycle of phage λ is indicative of most viral-type infections: the cycle is a continual repetition of the attack, usurp, and reproduce mantra. Upon adherence of the bacteriophage to the host cell exterior, infection is begun by injection of linear phage DNA into the cell lumen. This linear DNA circularizes and commences usurpation of host machinery. In the early stages of phage DNA transcription, several key decisions are made about life cycle progression; these decisions determine the fate of the phage life cycle for the immediate future. Essentially, there are two options the phage may pursue after initial infection: a lytic pathway and a lysogenic pathway. Should environmental conditions exist that are conducive to rapid phage proliferation (e.g. a fresh and plentiful nutrient supply), the phage will execute the lytic pathway. Such a decision results in immediate production of

phage DNA and proteins; new phage particles are then constructed and accumulate until lysis of the host occurs. This lysis causes dispersion of particles that continue the infectious scheme. Thus, the immediate result of a lytic pathway is destruction of the host cell. In contrast, the lysogenic pathway does not result in immediate phage reproduction. Environmental conditions that are not permissible to rapid phage proliferation will cause the phage system to initiate a series of events that lead to incorporation of phage DNA into the host genome. The infectious DNA will remain incorporated until environmental conditions are permissive, then dissociate from host DNA and commence the lytic pathway. Therefore, the immediate effect of lysogeny is continued host vitality. To summarize, the key aspects of phage reproduction are that the lytic phase of the phage life cycle leads to cell death (lysis) while the lysogenic phase results in cell vitality (Prescott *et al.*, 1999). This life cycle is illustrated in Figure 2.1. Attention shall now be turned to one of the prominent regulators of the lytic/lysogenic decision: the λ repressor.

The Structure and Role of λ Repressor

The λ repressor protein is a key regulator of the lysis/lysogeny decision in the phage life cycle. The repressor protein, along with several other early transcription products, serves to determine the fate of the phage and thus, indirectly, the host cell. Specifically, the λ repressor is responsible for promoting the establishment of lysogeny and halting lytic progress. By binding to specific sites on newly circularized phage DNA, λ repressor halts transcription of genes pertinent to maintaining a lytic pathway

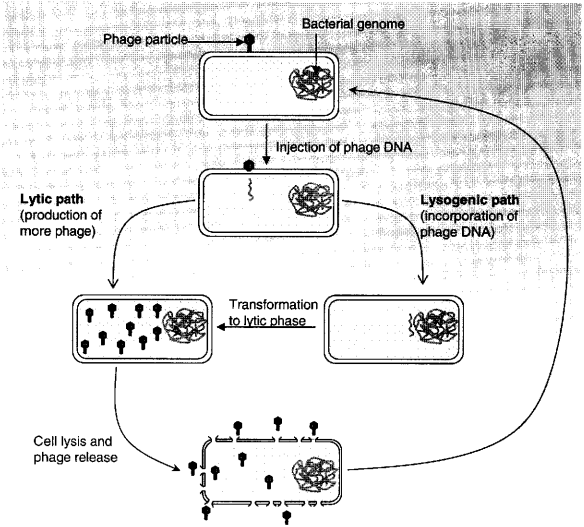


Figure 2.1 The phage λ life cycle. Shown here is the process of phage λ infection. The two pathway, lysis and lysogeny, are noted in the figure. The indicated transformation from lysogeny to lysis occurs due to environmental cues. However, in an abundance of functional λ repressor, the phage is typically unable to escape lysogeny and the host cell persists.

and allows for promotion of the lysogenic option. Essentially, an abundance of λ repressor will force the phage to proceed through a lysogenic life cycle.

The method by which λ repressor accomplishes such a task is heavily correlated to its structure. The repressor is 236 amino acids in length and consists of two domains, amino and carboxyl, connected by a flexible linker bridge. Dimerization of the repressor occurs via protein-protein contacts in the carboxyl domains; binding of DNA is exclusively accomplished through a tandum amino domain arrangement (Figure 2.2). This structure allows for the binding of phage DNA and results in the obstruction of RNA polymerase transcription of phage genes vital to a lytic life cycle. Thus, λ repressor effectively halts the phage processes that would otherwise lead to host destruction.

In summary, two key points concerning the repressor structure are as follows: dimerization of the repressor is an absolute requisite for binding of the associated amino domains to the phage DNA and this dimerization is facilitated by contacts through the carboxyl domains (Lewin, 2000; Russell, 1996; Snustad and Simmons, 2000). This unique structure-function relationship of the λ repressor and its role in the lysis/lysogeny decision immediately suggest a method by which the manipulation of the repressor structure can provide an easy means for dimer selection.

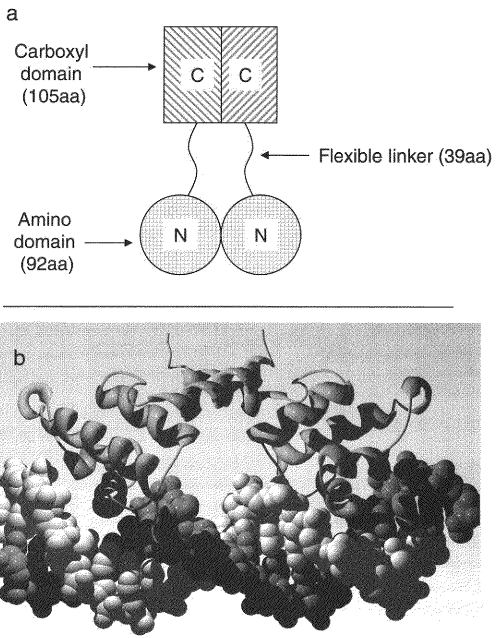


Figure 2.2 The λ repressor. **a)** Diagram showing the general structure of λ repressor in its dimeric form. **b)** Structure of dimeric λ repressor N-terminal domains binding to the phage operator (Beamer and Pabo, 1992).

The Selection System Rationale

The structural characteristics of λ repressor discussed in the previous section suggested to Hu and coworkers an ingenious way by which the life cycle of phage λ can be exploited to test for dimerization. By replacing the C-terminal domain of λ repressor with a selected peptide, allowing this fusion repressor to propagate within the host cell, and then infecting with phage λ, a simple life-death assay can be achieved for dimerization. Should the new fusion repressor maintain the ability to dimerize, the infecting phage will be forced into lysogeny and the host will remain robust. In contrast, should the fusion repressor fail to dimerize, infecting phage will be able to enter the lytic cycle and thus destroy the host. Therefore, this is a simple life-death selection that easily and quickly permits analysis of quaternary structure (Hu *et al.*, 1990; Hu *et al.*, 1993; Hu, 1995; Kim and Hu, 1995; Reidhaar-Olson *et al.*, 1991; Zeng, Zhu *et al.*, 1997; Zeng, Herndon *et al.*, 1997). This system is diagrammed in Figure 2.3. It must be noted that the success of the system is dependent on obtaining a phage stock that is deleted for the cI gene, the region of the phage DNA that codes for native λ repressor. Such a deletion is required because the fusion repressor will be introduced to the host cell via an independent plasmid. Production of native λ repressor from a phage stock would interfere with the life-death selection process.

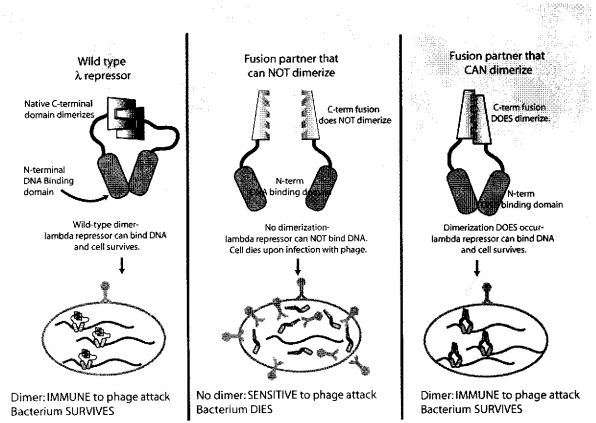


Figure 2.3 The phage λ selection system. The figure illustrates that phage λ can be used to select for dimerization by replacing the C-terminal end of the λ repressor with a novel partner. The presence or absence of protein dimerization can be determined by a simple life-death observation.

Materials and Methods

Production of Phage Stocks

Two 50µl phage stocks were obtained from the lab of Dr. James Hu: phage λ KH54 and phage λ h80. Both of these phage stocks are deleted for the cI gene and differ only in the recognition of the receptors needed to gain initial access to the host cell. The significance of this difference will be noted later (see p.XX of this thesis). These phage stocks were used to produce new phage for use in setting up the selection system. The protocol for this phage production was also obtained from the Hu laboratory. Serial dilutions of existing phage stocks were made to 10^{-2}, 10^{-4}, 10^{-5}, 10^{-6}, and 10^{-7}. 100µl of the 10^{-5}, 10^{-6}, and 10^{-7} dilutions were preadsorbed to MC1061 *E.coli* cells from an overnight culture for thirty minutes. 3ml of tryptone top agar was added to each bacterial-phage mixture and then transferred to freshly poured tryptone agar plates. After solidification of top agar, plates were placed into a 37°C oven overnight. Subsequently, a plate with well-defined plaques was selected and a single plaque was transferred by Pasteur pipette to 1ml of Tris-Magnesium (TM). After vortexing, four separate mixtures of 200µl of TM-phage and 50µl MC1061 cells were prepared. 3ml of tryptone top agar was added and then transferred to freshly poured tryptone agar plates. Incubation of the four plates was carried out at 37°C overnight. 5ml of TM was then added to the surface of each phage plate and then sealed and stored at 4°C for 48 hours. After this time, the phage-TM mixture was decanted, placed in a 15ml Falcon tube with 2-3 drops of choloroform, labeled as "stock" and stored at 4°C. Titring of the phage was

carried out by creating serial dilutions of stock solutions which were used to infect MC1061 cells. Plaques were counted and used to calculated the titre in plaque forming units (pfu)/ml. In order to achieve optimal infection rates, all stocks were assured to be between 10^{12} and 10^{14} pfu/ml.

Construction of MIPSTART and IL8 Fusion Repressors

The first step in establishment of the λ system was to ensure that the life-death selection works accurately using known controls. To test this, two fusion repressors from MIPSTART and wild-type IL-8 were created. As previously mentioned, the MIPSTART protein will be the beginning point for random mutagenesis and is a monomer. Wild type IL-8, of course, is a dimer. Dr. Jim Hu kindly provided the JH391 fusion repressor plasmid. This plasmid is diagrammed in Figure 2.4. JH391 contains the N-terminal domain and flexible linker bridge of the λ repressor followed by a lacZ stuffer region. This construct is under the control of a lac promoter. The lacZ stuffer can be removed by digestion of the plasmid with the Sal1 and BamH1 enzymes. Subsequently, a gene of interest can be inserted adjacent to the flexible linker bridge region. When expressed in *E.coli* this results in production of the fusion repressor. Oligonucleotide primers were used to amplify the MIPSTART and IL-8 genes from stock minipreps; a 3' BamH1 site and 5' Sal1 and Nco1 restriction sites (the Nco1 site is for rapid shuttling of the insert from the JH391 plasmid to the pET-32a(+) expression plasmid) were added. PCR products were purified by gel electrophoresis followed by cleanup with a Qiagen Gel Extraction kit. Purified PCR products were then digested in a

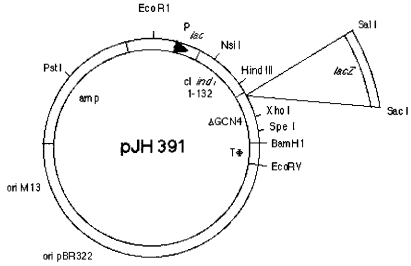


Figure 2.4 The JH391 plasmid. The JH391 plasmid shown here is used for constructing fusion λ repressors. The N-terminal domain (amino acids 1-132) is marked on the plasmid as cI. Before insertion of genes, the lacZ stuffer region is removed by digestion with SalI and BamH1 restriction enzymes. Genes of interest are then cloned into the vector between the SalI and BamH1 restriction sites. Production of the fusion repressor is controlled by a lac promoter (dark arrow at 1:00) This figure is courtesy of Dr. Jim Hu of Texas A&M University..

20µl mixture containing 14µl purified DNA, 2µl New England Biolabs BamH1 Buffer, 2µl New England Biolabs 10X BSA buffer, 1µl New England Biolabs Sal1 restriction enzyme, and 1µl New England Biolabs BamH1 restriction enzyme for three hours. Restriction digestions were purified using the Qiagen PCR purification kit and eluted in 40µl ddH_2O. Concurrently, the JH391 vector was digested in an analogous manner with Sal1 and BamH1 restriction enzymes and then purified using gel electrophoresis followed by clean up with the Qiagen Gel Extraction Kit (elution in 40µl ddH_2O). The digested plasmid and inserts were then ligated together in 20µl mixtures containing 10µl insert, 4µl vector, 2µl New England Biolabs T4 DNA Ligase Buffer, 3.5µl water, and 0.5µl New England Biolabs T4 DNA Ligase. Ligation was allowed to proceed for twenty four hours at 16°C. 5µl of the ligation mixtures were transformed into Stratagene XL-1 Blue $CaCl_2$ competent cells by heat shock at 42°C and plated onto LB agar plates containing ampicillin. After 24 hours of growth, individual colonies were picked, grown in culture, and DNA was harvested using the Qiagen MINIPREP Kit. The JH391MIPSTART and JH391IL8 constructs were then stored at -20°C in 40µl of ddH_2O.

Analysis of MIPSTART and IL8 Performance

The JH391MIPSTART and JH391IL8 constructs were next tested for performance in the λ selection system. JH391MIPSTART and JH391IL8 were transformed into the AG1688 *E.coli* cell line obtained from Dr. Jim Hu and allowed to

grow overnight. A "cross-streak" test was then performed to examine the success or failure of the selection system in analyzing these two proteins. Phage λ KH54 was first streaked horizontally across an LB agar plate. Next, streaks of cells containing constructs to be tested were drawn vertically across the phage line. The vitality of the cells at the interface of the two streaks indicates the presence or absence of a functional λ repressor. Figure 2.5 shows the results of this experiment for MIPSTART and IL-8 constructs. As can be seen, the two constructs perform as expected: MIPSTART, a monomer, shows clear signs of necrosis at the phage-bacteria interface while IL-8, a dimer, shows complete vitality. This experiment verifies that the controls, the "starting" and "ending" points of the structural study both perform as expected *in vivo*.

Creation of a High-Throughput Method

The success of MIPSTART and IL-8 to perform as expected in the λ system demonstrates that the selection system is indeed adequate for the purposes of this investigation. However, as previously mentioned and as Figure 2.5 demonstrates, the cross-streak method of dimerization analysis is not conducive to high-throughput analysis. The prospect of streaking thousands, or millions, of colonies is simply not feasible. Therefore, an expansion of the system to a level where high-throughput is possible was conducted. It seemed that the most straightforward method of accomplishing such an expansion was to simply transform MIPSTART and IL-8 into AG1688 *E.coli* cells and then plate a limited amount of the transformation mixture onto

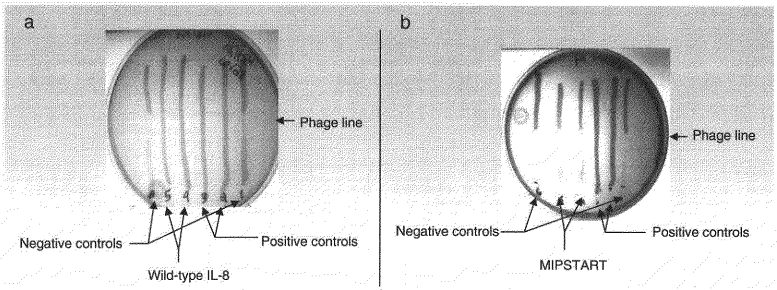


Figure 2.5 Cross-streak of MIPSTART and IL-8. The performance of IL-8 (**a**) and MIPSTART (**b**) in the phage λ selection system is shown. Positive controls are JH391 constructs containing the entire wild-type λ repressor. Negative controls are JH391 constructs containing only the N-terminal region of the repressor (lacks a dimerization domain). Resistance or sensitivity to phage lysis can be determined by examination of the bacteria-phage interface. This test demonstrates that both MIPSTART and IL-8 perform as expected: IL-8 survives because of dimerization while MIPSTART, a monomer, does not convey phage immunity.

plates already replete with a phage solution. To test this, 150µl of the KH54 phage λ was plated onto LB agar plates followed by 100µl of AG1688 heat shock transformation mixtures of MIPSTART, IL-8, and appropriate controls on top of the phage. Early experiments in this manner were somewhat successful but led to the appearance of many colonies on plates that should not have exhibited any dimerization. It was decided that the appearance of these colonies was the result of a rather obvious culprit: natural selection. By only using one type of phage in the experiments, selection was occurring for those cells that might have phage resistance from means other than λ repressor. Therefore, the conclusion was reached that cells should be exposed to more than one type of phage. By doing this, the chances of an abnormally resistant bacterium from surfacing are significantly reduced.

The procedure was modified to account for this by plating 150µl of KH54 and 150µl of h80 phage λ onto a plate before addition of transformed cells. The results of these experiments are shown in Figure 2.6. As can be seen from this figure, the high-throughput method of analysis was achieved, with many thousands of colonies effectively screened by the phage system. Some false positive colonies do occur in both the negative control and MIPSTART plates in Figure 2.6, but the number is so low (1 and 4 colonies, respectively) to be almost inconsequential. It can be imagined that such a process could be applied to a randomized pool by transformation of the pool into AG1688 cells that would then be exposed, as above, to phage.

However, a significant difference in the application of this system to randomized libraries is the need to obtain an accurate count of how many possible colonies were

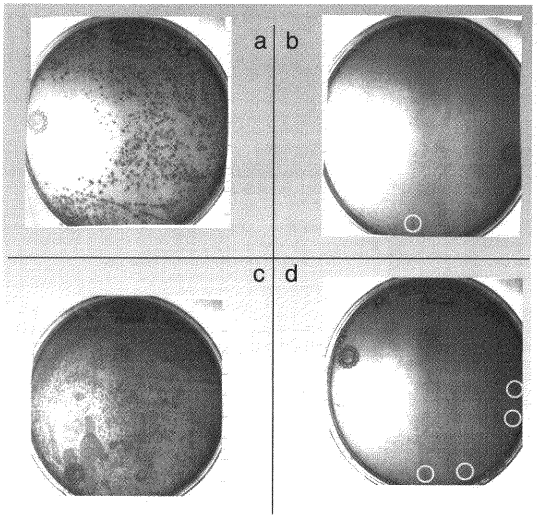


Figure 2.6 High-throughput selection of MIPSTART and IL-8. Plates coated with phage were exposed to cells transformed with JH391 constructs and observed for colony growth. Four separate tests were performed: positive control (**a**), negative control (**b**), IL-8 wild-type dimer (**c**), and MIPSTART monomer (**d**). Positive and negative controls were the same as those described in Figure 2.5. Observed cell growth in (**b**) and (**d**) is marked with a light circle. Although these colonies do indicate false positives, overall the system effectively selects against thousands of colonies. Only a small portion persist as false positives. The results shown here demonstrate that the high-throughput system does indeed accurately select against fusion repressors incapable of dimerization.

analyzed by the selection process. This need necessitates a slight modification of the process. The transformation of randomized libraries into AG1688 *E.coli* cells was accomplished by electroporation in order to maximize the amount of mutants that can be screened by one selection iteration. After successful electroporation of a randomized library, a small portion (1%) of the transformation mixture is separated and plated on LB agar plates that do not contain phage. By counting the number of colonies that appear on the non-infected plate, extrapolation of the number of colonies screened on the phage-exposed plate is possible. In this manner, the number of different variants from a randomized pool that have been screened can be estimated. Refinement of this high-throughput system has been achieved and it is possible to screen up to 180,000 random variants on one plate. On average, approximately 50,000 mutants can be screened on one phage-coated plate. The appearance of false positives does occur at a regular rate, but never more than 20 colonies during a single selection. Such a small number is easily analyzed further by a low-throughput method such as the cross-streak.

Randomization

With the successful set up of a phage λ selection system, the next logical step was to develop a process by which MIPSTART could be successfully randomized. Several possibilities existed that could potentially serve as a randomization protocol. The use of a "sloppy" polymerase, such as Taq, was considered and even attempted, but failed to provide the level of randomization desired. This is not unreasonable considering that Taq only introduces approximately one base pair change for every

20,000 base pairs replicated, a rate that is fairly useless when working with proteins whose genetic sequences are only 200 base pairs long. Other procedures were only cursorily examined as they did not give the type and depth of randomization desired. The solution that seemed most appropriate was the use of randomized oligonucleotide primers.

Randomized oligonucleotide primers were ordered directly from Integrated DNA Technologies (IDT). When ordering such primers, bases to be randomized were designated as "N" in the sequence. The primers produced by IDT contained a mixture of adenine, cytosine, thymine, and guanine at locations so designated. IDT indicated that there was a slight weighting of bases towards adenine and thymine at these locations, but that the primer stock provided did contain every possible combination of base arrangements. By performing randomization in this manner, not only was an incredible range of mutants obtained, but mutagenesis but mutagenesis was directed to regions hypothesized to be important for forming the alternate dimer. In this manner, the versatility of random mutagenesis was combined with rational design to create a very powerful mutagenesis scheme.

Randomization using these primers was accomplished using polymerase chain reactions. For randomization of regions within approximately twenty base pairs of the 5' or 3' termini, a traditional PCR could be performed to create a library of mutants. For randomization of regions distant from the termini, a different approach was taken. This procedure was developed specifically to facilitate randomization of regions in the bulk of the gene. The general scheme of the procedure is outlined in Figure 2.7. This process

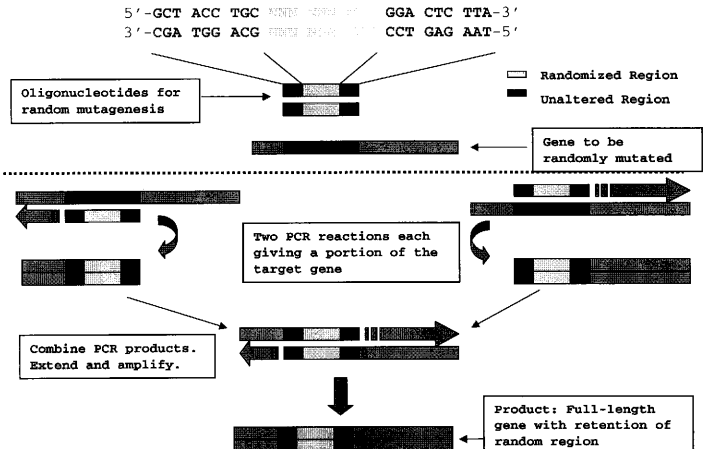


Figure 2.7 Random PCR mutagenesis. The scheme for randomizing regions of the MIPSTART distant from the 5' and 3' termini. Random oligonucleotide primers obtained from Integrated DNA Technologies are used to introduce base pair substitutions through a series of polymerase chain reactions.

uses two separate PCRs to accomplish randomization. Initially, two separate but complementary oligonucleotides were created for the region to be randomized. These primers were used with terminal oligonucleotide primers to create "half-gene products" that were then combined to create a "full-gene product" randomized only in the desired region. This randomized gene library was then be analyzed using the λ selection system.

The System in Total

The two main sections of the process used to search for alternate dimers have been outlined: the λ selection system and the randomization process. It is the coupling of these two processes that allow for a complete methodology by which to change MIPSTART from a monomer to an IL-8 type dimer. Once randomization has been completed, the randomized gene product and the JH391 plasmid are subjected to restriction digestion by SalI and BamH1 restriction enzymes. The randomized inserts and JH391 plasmid are then ligated together (the ligation process used for randomized products will be described in Chapter III) and transformation of the ligated product into AG1688 *E.coli* cells by electroporation is performed. As mentioned previously, a small portion of the transformation mixture is removed and plated under permissive conditions to allow accurate counting of the mutants screened. After the screening is complete, colonies that thrive on the phage coated plate are harvested and grown in culture overnight. Subsequently, the mutant DNA is harvested using a Qiagen Miniprep Kit. Retransformation of the DNA (by $CaCl_2$ heat shock) into AG1688 cells is then

performed and a cross streak test executed. This second phage test is to ensure that all colonies harvested from the initial high-throughput plate are in fact resistant to phage based on the dimerization of λ repressor and not on artifacts of natural selection. Mutant DNA that continually demonstrates the ability to convey phage resistance is then shuttled into an expression vector and produced for examination by NMR.

CHAPTER III

ANALYSIS OF RANDOM VARIANTS

Introduction

The establishment of a selection system and refinement of a random mutagenesis technique described in the last chapter was the staging ground for a concerted mutagenesis of the MIPSTART protein. With the means necessary to randomize MIPSTART and select for dimerization in place, the next step was to actually implement the randomization/selection scheme.

There are three areas of MIPSTART that seem to be promising targets for randomization: the first β-strand, the loop region between β-strands 1 and 2, and the C-terminal α-helix. Figure 3.1 demonstrates the location of these three regions. A comparison with Figure 1.8 immediately demonstrates that these regions play pertinent roles in formation of the IL-8 quaternary structure. In the IL-8 dimer, interactions between the first β-strands of opposite subunits help to stabilize the quaternary structure through hydrogen bonding and hydrophobic side chain contacts. Mutation of residues in the first β-strand has been shown to decrease the ability of IL-8 to self-associate (Lowman *et al.*, 1997; Rajarathnam *et al.*, 1994).

Also playing a key role in maintenance of the IL-8 dimer structure are the interactions between the C-terminal α-helix and the β1-β2 loop region. Hydrogen bonding between these regions is thought to contribute significantly to IL-8

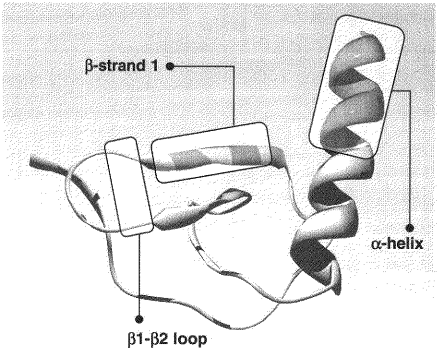


Figure 3.1 Important MIPSTART regions. Areas of MIPSTART considered prime candidates for randomization are indicated above. Comparison of these regions with Figure 1.8 indicates that all three highlighted areas participate in IL-8 dimer formation. It is assumed that only β1-β2 loop residues near the β-strands participate in dimer formation as these residues are most likely to interact with the C-terminal helix.

dimerization; removal of the last six residues of the C-terminal IL-8 helix results in a completely monomeric variant (Lowman *et al.*, 1997; Rajarathnam *et al.*, 1997).

There are three areas, therefore, that seem promising avenues for random mutagenesis: the first β-strand, the loop region between β1 and β2, and the C-terminal α-helix. Randomization of these three areas followed by selection with the phage λ system was carried out in an attempt to locate possible variants that demonstrate formation of the alternate dimer.

Materials and Methods

Randomization of MIPSTART Template

The JH391MIPSTART template constructed in Chapter II was used as a randomization template. Complementary oligonucleotide primers were used to effect randomization in the first β-strand as described in Chapter II. Randomization of the β1-β2 loop region was also carried out using the procedure outlined in Chapter II. Terminal oligonucleotide primers used were complementary to the MIPSTART 5' and 3' ends. The terminal 5' oligonucleotide primer incorporated SalI and NcoI restriction sites before the 5' end of the MIPSTART gene. The terminal 3' primer added a BamHI restriction site. Helix randomization was accomplished by traditional PCR using a terminal 3' oligonucleotide primer with randomized positions.

Rational Mutation of MIPSTART

The rational mutation of MIPSTART loop regions was achieved by oligonucleotide primers ordered from Integrated DNA Technologies. Because the rational mutations attempted in Chapter III occur in internal regions of the MIPSTART gene, the PCR protocol from Chapter II was adapted as a means to incorporate these rational mutations into the MIPSTART gene. Complementary oligonucleotides containing the desired mutations were ordered and used to carry out a PCR procedure as described in Chapter II.

Ligation of Randomized Variants into JH391 Plasmid

After purification of the randomized DNA library, both the JH391 plasmid and the randomized DNA library were subjected to restriction digestion as follows: 14µl of plasmid/library DNA, 2µl New England BioLabs 10X BSA, 2µl New England BioLabs BamH1 buffer, 1µl New England BioLabs Sal1 Restriction Enzyme, and 1µl New England BioLabs BamH1 restriction enzyme were combined in a 1.5ml microcentrifuge tube and allowed to incubate overnight at 37°C. Both insert and plasmid DNA were run on a low melt agarose (1.7% and 1.0%, respectively) gel containing EtBr for purification. Appropriate bands were excised, purified using the Qiagen Gel Extraction Kit, and eluted in 40µl ddH_2O. Quantification of DNA was then performed using a Beckman DU640 Spectrophotometer. Concentration of DNA was determined using the equation (Sambrook *et al.*, 1989):

$$\text{Concentration (ng/µl)} = OD_{260}(\text{Dilution})(50\text{ng/µl}/OD_{260})$$

After shaking, 1µl of cells were plated onto an LB-Kan/Amp agar plate for determination of colony count and 900µl of cells were plated onto an LB-Kan/Amp agar plate pre-treated with 150µl of phage λ KH54 and 150µl phage λ h80. Plates were allowed to incubate overnight at 37°C. Subsequently, the plates were removed and the 1µl non-phage plate used to determine colony count. Any surviving colonies on the phage-treated plate were transferred to 2XYT broth and cultured overnight at 37°C. The plasmid DNA was then harvested using a Qiagen Miniprep Kit.

Retransformation and Low-Throughput Phage Screening of Randomized Variants

After miniprepping of high-throughput survivors, retransformation of harvested DNA into AG1688 cells was accomplished using a standard $CaCl_2$ transformation procedure. 1µl of DNA was added to 100µl of $CaCl_2$ AG1688 cells and allowed to incubate on ice for 1 hour. Cells were then transferred to a 42°C hot water bath and incubated for 45s. This was followed by incubation on ice for another 2 minutes. 400µl LB was added and cells were placed in a shaker at 200rpm and 37° for 45 minutes. Cells were then spun at 3000rpm for five minutes in a microcentrifuge, the supernatant discarded, and resuspension in 100µl LB achieved. This entire 100µl solution was then plated onto an LB-Kan/Amp plate and allowed to incubate overnight at 37°C. A colony was picked and used in a cross-streak test to verify resistance to phage lysis. Those variants that showed phage sensitivity after retransformation were discarded.

Phage System Efficiency

The procedure for phage selection just outlined has been characterized to a great extent throughout the course of this work. Several key factors should be noted when working with this system. As detailed in Chapter II, a maximum of 180,000 colonies have been successfully screened in one selection iteration. However, the *average* number of colonies that can be successfully screened in one iteration is approximately 50,000. Typically, 2-25 colonies demonstrate phage immunity after a selection. Out of the colonies that survive the selection process, most (~95%) were false positives detected upon retransformation and cross-streaking. The remaining 10% were true positives that will be discussed in the next section. This efficiency is well within an acceptable range. Because of the appearance of false positives and the statistical nature of phage selection, it is assumed that one must screen a number of colonies that is at least two times the number of possible mutants to claim complete analysis of the random library. For example, if one is working with a pool of 2,000 mutants, selection on at least 4,000 colonies must be performed.

Sequencing of Surviving Variants

Those variants that demonstrated phage resistance after restransformation were sequenced using the Texas A&M University Gene Technologies Laboratory protocol. 10µl of miniprep DNA was dried down in a 0.2ml PCR tube using a Savant SpeedVac Concentrator and resuspended in 3µl ddH_2O. To this were added 2µl of 5µM T7 terminator oligonucleotide primer and 2µl of BIGDYE (obtained from the GeneTech

Lab). A sequencing program was then executed on an MJ Research thermocycler that allowed for DNA sequencing. The amplified DNA was purified from free nucleotides by passage over a column containing Sephadex G50 resin (for removal of free nucleotides). This solution was then dried using a Savant SpeedVac Concentrator and shipped to Gene Technologies for separation and analysis.

Shuttling of Surviving Variants to pET-32a(+) Expression Vector

In order to express interesting variants, transfer of DNA between the JH391 selection vector to the T7-containing pET-32a(+) expression vector was necessary (Figure 3.2). Amplification of the variant from the JH391 plasmid was accomplished using PCR with a 5' MIPSTART complementary oligonucleotide primer and a 3' T7 terminator oligonucleotide primer. This PCR product was purified using gel electrophoresis and a Qiagen Gel Extraction Kit. Digestion of the product was achieved using the NcoI and BamH1 restriction enzymes in a fashion identical to that already described. Ligation into a digested and purified pET32a(+) plasmid was performed as follows: 10µl insert, 4µl plasmid, 2µl T4 DNA Ligase Buffer, 3.5µl ddH_2O, and 0.5µl T4 DNA ligase were combined and incubated overnight at 16°C. 5µl of the raw ligation mixture was then added to 100µl of Stratagene's XL-1 Blue $CaCl_2$ competent cells and allowed to incubate on ice for 1 hour. The mixture was then transferred to a 42°C hot water bath and incubated for 45s. This was followed by an immediate 2 minute incubation on ice after which 400µl of LB was added and the cells were placed in a shaker at 200rpm and 37°C for 45 minutes. The cells were then spun at 3000rpm for five

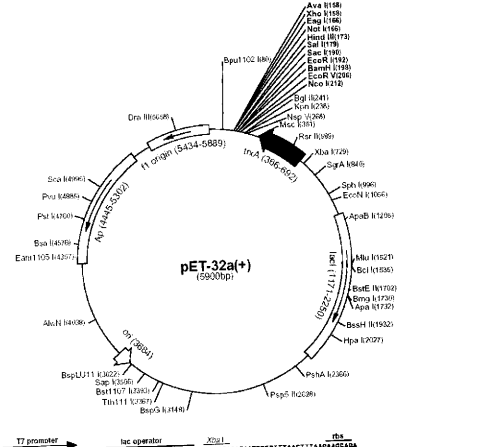


Figure 3.2 The pET-32a(+) plasmid. Shown above is the vector map for the pET-32a(+) expression plasmid . Genes were cloned into the pET vector either between Nco1 and BamH1 (for protein production with fusion tags) or between Nde1 and BamH1 (for protein production without fusion tags). Cleavage of fusion tags was accomplished by enterokinase digestion at the site indicated on the map. Protelolytic digestion leaves two residues, alanine and methionine, as artifacts of the cleavage process. This can be avoided by producing protein without fusion tags (reprint permission granted by EMD Biosciences, Inc., Novagen).

minutes, the supernatant decanted, and the pellet resuspended in 100µl LB. The entire mixture was plated onto LB-Amp plates and allowed to incubate overnight at 37°C. Colonies were selected and placed in LB broth overnight and plasmid DNA was harvested using the Qiagen Miniprep Kit.

Expression of Variants

Expression of variants was accomplished using the Novagen BL21(DE3) *E.coli* strain. Transformation of variant DNA in the pET-32a(+) plasmid into $CaCl_2$ competent BL21 cells was accomplished by heat shock transformation as already described. A colony was then picked and grown in 2ml LB culture at 37°C to an OD_{550} of 0.7-0.9, an 800µl sample removed for glycerol preparation (15% glycerol, 85% cell culture), and then induced by addition of 1mM IPTG. Growth was continued for another 2 hours and then a 200µl sample was removed and prepared for analysis by SDS-PAGE. Large scale production and isotopic labeling of protein was achieved by addition of 200µl of the previously prepared glycerol freeze to a 100ml minimal media solution containing ^{15}N ammonium chloride as the exclusive source of nitrogen. This culture was allowed to shake overnight at 37°C and 200rpm. 35ml of the 100ml culture was then added to a 1L minimal media preparation containing ^{15}N ammonium chloride as the nitrogen source and allowed to grow to an OD_{550} of 0.8. Samples were taken for SDS-PAGE analysis and glycerol stock preparation and then 1mM IPTG added. Growth was allowed to continue for another 4 hours and then another sample for SDS-PAGE analysis was

removed. Cells were then spun down in a Sorvall RC-5B centrifuge using a GS3 rotor at 6000rpm for thirty minutes and then placed in a -20° freezer for storage.

SDS-PAGE Analysis

SDS-PAGE was used to ensure proper protein expression. SDS-PAGE samples were prepared by addition of 2X loading buffer to samples followed by boiling for five minutes. 10µl of each sample was loaded into a 17% polyacrylamide tricine gel with a 4% stacking layer. Electrophoresis was carried out at 60V for 20 minutes followed by a 120V segment for 1 hour. Gels were stained in 50% methanol, 10% acetic acid, and 40% Coomasie Blue Stock. Removal of excess dye was achieved by rinsing with this same solution lacking the Coomasie Blue stain.

Isolation of Protein

Cells stored at -20° were resuspended in 20 ml of a high salt cracking buffer containing 500mM NaCl, 20mM Tris (pH 8), 1mM EDTA, 5mM β-mercaptoethanol, and 5mM benzamidine. Cells were then lysed by French pressing at 16000psi two times followed by centrifugation at 19,000g for 1 hour. Supernatants were decanted and the pellet resuspended in 10ml of denaturing buffer containing 5M guanidinium-HCl, 50mM NaCl, 20mL Tris(pH 8), 1mM EDTA, and 5mM β-mercaptoethanol and allowed to sit at room temperature for 2 hours. The solution was centrifuged at 15,000g for 30 minutes and the supernatant refolded by introduction into a refolding buffer of same composition minus the denaturing agent. The refolding was accomplished by drop-wise addition over

the course of thirty minutes. The protein was then allowed to continue refolding at room temperature for another 2 hours. The solution was centrifuged at 10,000g for thirty minutes and the supernatant dialyzed against a 2L solution of 20mM Tris (pH 8) at 4°C. Buffer was switched twice at six hour intervals and upon completion of dialysis the solution was centrifuged at 10,000g for 20 minutes. The supernatant was removed and stored at 4°C. For some variants, purification was also attempted by isolating protein from the lysis supernatant after French Pressing. In this case, the supernatant isolated after cell lysis by French Pressing was not refolded but immediately purified using a Ni^{2+} saturated HiTrap chelating column.

Purification of Variants

Proteins inserted into the pET-32a(+) plasmid using the Nco1-BamH1 cutting scheme are expressed with a variety of fusion tags attached to the N-terminus (Figure 3.2). These tags serve to assist in the purification process. After initial purification is complete, these tags may be removed by proteolytic cleavage using the protease enterokinase. Cleavage in this manner adds an additional two residues to the end of the target protein (alanine and methionine, in that order).

Most variants were initially purified by pellet refolding and then subjection to reverse phase high performance liquid chromatography on a Vydac C4 column. Refolded pellets after dialysis were adjusted to a pH of 2.5 using trifluoroacetic acid (~0.1% vol/vol TFA) and adjusted to a composition of 10% acetonitrile. The column was rinsed at 5ml/min with 40ml of 10% acetonitrile, 90% ddH_2O, both at 0.1% TFA.

Protein was loaded onto the column at a rate of 2-4ml/min and then the column was rinsed with an additional 20ml of 10% acetonitrile, 90% ddH_2O (0.1%TFA). A gradient was applied from 10-100% acetonitrile (0.1% TFA). Elution of protein was monitored with absorbance at 280nm. Isolated proteins were then collected, lyophilized, and stored at room temperature in preparation for proteolytic cleavage.

In some instances, initial purification was achieved by nickel affinity chromatography using the poly-histidine fusion tag. Passage of proteins attached to such a tag over a chelating resin containing Ni^{2+} results in separation of the target protein from most other cellular components. This type of purification scheme is used when isolation of protein from the cell lysis supernatant (after French Pressing) is desired. Cell lysis supernatants with added NaCl (200mM) were passed over one or more Pharmacia Hi-Trap Chelating columns charged with $NiSO_4$. An Amersham Pharmacia system was used to load supernatants onto the charged chelating columns at a rate of 2ml/min (an exception to this is the purification of Loop Variant 3-this was run at less than 1ml/min). Columns were washed with 50mM Tris (pH 8), 500mM NaCl, and 5mM imidazole, then a 50mM imidazole "bump" to remove any indiscriminately bound proteins. A gradient from 50-500mM imidazole was then applied to remove any protein that was bound in significant proportions to the column. Absorbance at 280nm was used to monitor elution of protein. Fractions containing the desired protein were collected and dialyzed against 20mM Tris (pH 8) at 4°C in preparation for proteolytic cleavage.

After proteolytic cleavage, the digested protein was isolated from fusion tags by reverse phase high performance liquid chromatography on a Vydac C4 column. Cleaved

proteins were adjusted to a pH of 2.5 using trifluoroacetic acid (~0.1% TFA) and adjusted to a composition of 10% acetonitrile. The column was rinsed at 5ml/min with 40ml of 10% acetonitrile, 90% ddH_2O, both at 0.1% TFA. Protein was loaded onto the column at a rate of 2-4ml/min and then the column was rinsed with an additional 20ml of 10% acetonitrile, 90% ddH_2O (0.1%TFA). A gradient was applied from 10-100% acetonitrile (0.1% TFA). Again, elution of proteins was monitored with absorbance at 280nm. Isolated proteins were then lyophilized and stored at room temperature.

Proteolytic Cleavage

Proteins containing an enterokinase cut site were digested in cutting buffer (20mM tris[pH 8], 50mM NaCl, 2mM $CaCl_2$). Protein concentration was kept as high as possible, with the cutting solution maintaining an absorbance at 280nm near 0.1-0.3. Approximately 0.2 units of enterokinase per ml was added to the cutting solution and allowed to digest over several days. Progress of the cleavage was monitored by SDS-PAGE.

NMR Spectroscopy

All spectra were recorded at 25°C on a Varian Inova 500 or 600 MHz spectrometer. These spectrometers are equipped with xyz-shielded gradient triple resonance probes. Sample tubes were obtained from Shigemi, Inc. HSQC spectra were acquired with 512* points in the ^1H (direct), and 128* points in the nitrogen dimension.

Spectral width for ^{1}H was approximately 6000-8000Hz and for ^{15}N was 1500-1700Hz. Referencing was relative to DSS, using the method described by Wishart *et al.*

Results and Analysis

Results of βl Randomization

Randomization of the first β-strand was targeted to promote interaction between subunits in an IL-8 type dimer. Because of the apparently salient role the first β-strand played in forming the IL-8 type dimer, it was assumed this would be an ideal starting point for implementing the randomization/selection process previously developed. As this was the first attempt at randomization using the specialized PCR scheme, it was unclear the extent to which random bases could be incorporated into the oligonucleotide primer before annealing became too nonspecific. To assess this, oligonucleotide primer pairs with increasing amounts of randomized bases were ordered. Specifically, primers with 3, 4, 9, and 12 randomized positions centered on the first β-strand were obtained. The ability of the primers to successfully produce PCR products using the randomization scheme was then assessed. All primers successfully produced amplified DNA when subjected to the PCR protocol. The conclusion was reached that there was no hindrance to using primers with a highly randomized content, at least up to 12 positions. To date, the efficiency of primers with a randomized base content higher than 12 has not been tested. Sequence analysis of randomized products ligated into plasmid DNA verified that the desired product had been produced and that positions expected to demonstrate randomization did indeed present such variation (Figure 3.3).

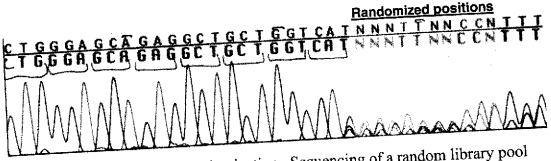


Figure 3.3 Verification of randomization. Sequencing of a random library pool demonstrated that regions altered via PCR random mutagenesis exhibit low sequence cohesion. The highlighted area marks positions randomized using the procedure described in Chapter II. These positions are clearly ambiguous and show that the PCR procedure works as predicted.

Randomization and selection of first β-strand variants proceeded as described in Materials and Methods. Figure 3.4 demonstrates the positioning of the randomized bases and the number of possible variants that could be produced from such randomization. Based on colony counts performed during the selection process, an estimated 400,000 colonies were screened. This should have been sufficient to effectively screen all 185,220 possible variants. From this selection process, two variants demonstrated the consistent ability to convey resistance to phage λ. After sequencing of these variants, it was discovered that both contain stop codons engineered in β1 positions. The sequences of these variants are indicated in Figure 3.5. The presence of stop codons in the first β-strand was perplexing since it is not entirely clear why a short (~20 amino acids) polypeptide chain consistently allows cell vitality when exposed to phage. It is possible that the dimerization of λ repressor is brought about by simple nonspecific association of the peptide. Formation of disulfide bridges between the cysteine residues of each peptide could also result in non-specific dimerization. Although the reducing environment of *E.coli* would seem to preclude this, it is not entirely unreasonable as previous studies using this system demonstrate that short peptides rich in cysteine do in fact cause dimerization of the λ repressor amino domains (Zhang *et al.*, 1999).

Currently, the exact relevance of these peptides is unclear and further investigations into their ability to promote repressor dimerization is intended. One possible avenue of inquiry would be to express and purify the entire repressor amino domain with the peptide attached and investigate the structure with NMR.

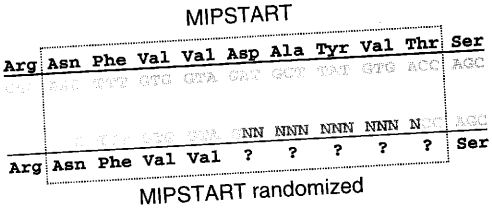


Figure 3.4 β1 randomization of MIPSTART. Shown above is the randomization of the first β-strand using twelve randomized positions. The boxed region is the β1 boundary. The randomized positions allow for a total of 185,220 different peptide combinations (including stop codons).

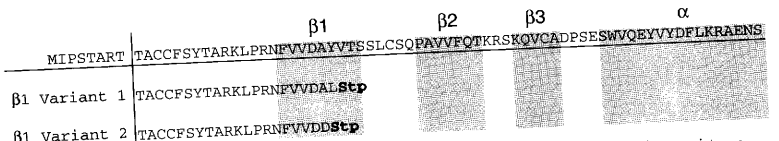


Figure 3.5 Results of β1 randomization of MIPSTART. Two variants consistently demonstrated phage resistance after randomization of the first β-strand. The sequences of these two variants are diagrammed above.

Results of Loop Randomization

The flexible loop region following the first β-strand and preceding β-strand 2 also seemed an ideal place to target randomization. As already discussed, interactions between the C-terminal α-helix and this loop region are absolutely necessary for maintenance of the IL-8 dimer. Therefore, changes in the loop region might promote interaction with the helix. The randomization procedure was conducted as previously described in Materials and Methods. The location of randomized bases was chosen so that residues thought important for helix interaction were altered while other residues which already play a dominant role in maintaining *monomeric* structure (such as the nearby cysteine) were not changed (Figure 3.6). After randomization and selection, calculations indicated that a total of at least 500,000 different colonies had been screened; this was judged to be sufficient to cover most of the 194,481 possible variants. From this selection process, three variants showed the ability to consistently convey resistance to phage. Out of these three, two contained randomly integrated stop codons. The third contained codons for amino acids in each randomized position. The sequences of these three variants are shown in Figure 3.7. As with the β-strand, variants containing stop codons have not yet been further characterized. However, Loop Variant 3 (LV3) was immediately shuttled into the pET-32a(+) plasmid for expression and purification.

Unfortunately, purification of LV3 proved to be very cumbersome. Several attempts were made to successfully purify the protein. After shuttling to the pET-32a(+) expression vector, the variant was expressed with the thioredoxin, poly-Histidine, and S tags of the pET-32a(+) vector (Figure 3.2). Initially, the cell lysis pellet from French

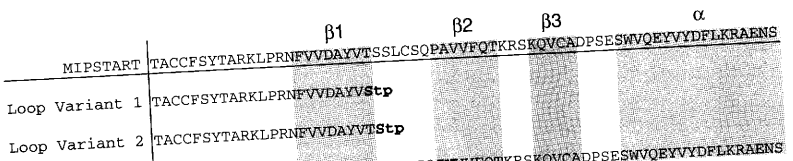


Figure 3.7 Results of β1-β2 loop randomization of MIPSTART. Three variants consistently demonstrated phage resistance after randomization of the loop region. The sequences of these three variants are diagrammed above.

pressing was refolded and purified over a reverse phase C4 column as indicated in Materials and Methods. The trace produced by the Amersham Pharmacia system is shown in Figure 3.8. As can be seen, very little if any protein was eluted from the column. The fraction containing the desired protein was collected, lyophilized, and digested with enterokinase to remove fusion tags. This cleavage reaction was again run over a reverse phase C4 column to separate the protein product from fusion tags. As Figure 3.8 shows, the amount of protein produced was again extremely low and extraordinarily suspect. This protein was collected, lyophilized, and stored for later NMR study.

In an attempt to find an alternate route to purification of LV3, purification from the cell lysis supernatant was attempted. A large, 3L cell growth was used to express the LV3 protein; the supernatant after French Pressing was passed over a HiTrap chelating column saturated with Ni^{2+}. The UV trace for this purification is shown in Figure 3.9. In this case, only a single, low, broad peak eluted from the column. This peak was collected, dialyzed, and then subjected to proteolytic digestion by enterokinase. The cleavage reaction was then passed over a C4 reverse phase column for purification. Again, only a small amount of protein was obtained. This protein was collected, lyophilized and stored for later NMR study.

The protein isolated from both purification schemes was used to obtain NMR HSQC spectra. These HSQC spectra for LV3 are shown in Figure 3.10. The poor dispersion of peaks and clustering towards the center of the spectra indicates an unfolded protein. It is not yet clear why the LV3 protein was so difficult to purify or

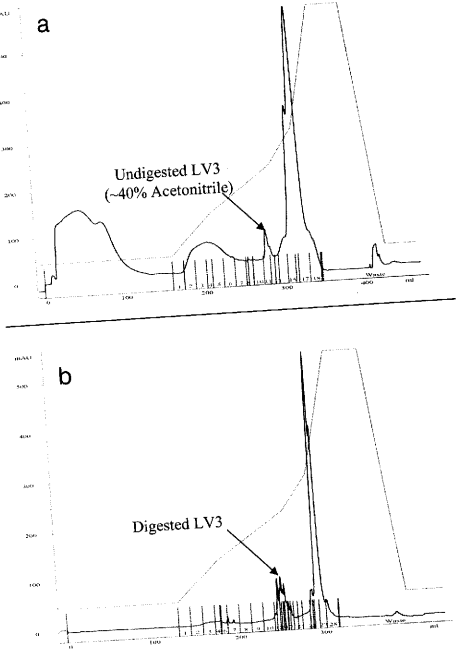


Figure 3.8 LV3 Pellet Purification. Shown above are UV traces at 280nm of the purification of LV3. Purification before enterokinase digestion (**a**) produced a small peak in the region expected for MIP-1β. Purification after enterokinase digestion (**b**) produced several small peaks near 40% These peaks were pooled together and lyophilized..

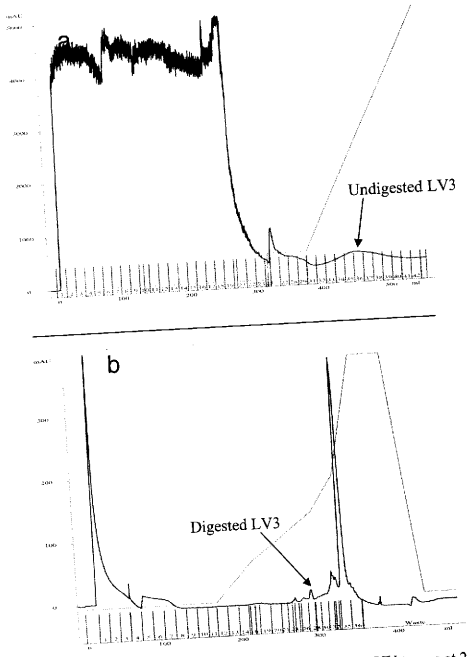


Figure 3.9 LV3 Supernatant Purification. Shown above are UV traces at 280nm of the purification of LV3 supernatant. Purification before enterokinase digestion (**a**) was performed over a HiTrap chelating column saturate with Ni^{2+} and produced a broad, low peak that was isolated and dialyzed. Purification after enterokinase digestion (**b**) was performed on a C4 reverse phase column and produced a single small peak. This peak was isolated and lyophilized for NMR analysis.

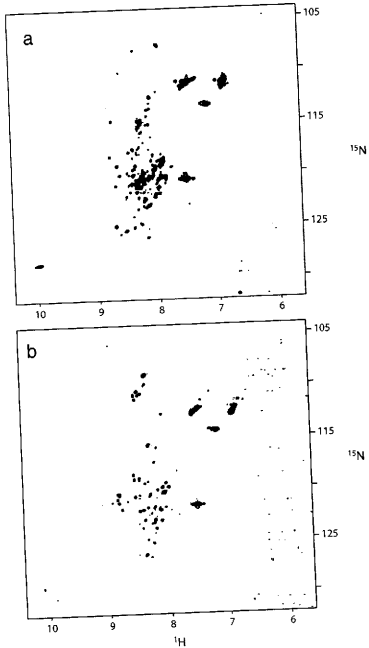


Figure 3.10 HSQC spectra of Loop Variant 3. HSQC spectra of Loop Variant 3 are shown above. (**a**) LV3 purified by pellet refolding followed by enterokinase digestion and purification. (**b**) LV3 purified from cell lysis supernatant by a Ni^{2+} column followed by enterokinase digestion and purification. These spectra contain many peaks clustered at the center, thus indicating that the protein is unfolded *in vitro*. It is unclear why this protein consistently passes the λ selection system.

why a protein that produces seemingly positive results *in vivo* appears unfolded once purified.

An Attempt at Rational Mutation

Because the loop region seemed to be extremely important to forming the IL-8 type dimer, a rational approach was attempted at this area. Rational Loop Variant 1 (RLV1) and Rational Loop Variant 2 (RLV2) were designed to change the MIPSTART loop region to resemble more closely the analogous IL-8 area. Figure 3.11 shows the changes made in RLV1 and RLV2 in order to imitate the IL-8 loop. These rational mutations were carried out as described in Materials and Methods and expressed in the pET-32a(+) plasmid. These proteins were purified as described in Materials and Methods by refolding of the cell lysis pellet followed by proteolytic cleavage. An HSQC spectrum of both proteins was obtained; these spectra are shown in Figure 3.12. The clustering of peaks in the center indicates that both proteins are unfolded. The unsuccessful mutagenesis of the loop region by both random and rational means seems to indicate that alteration of this area is highly detrimental to structure integrity. A cysteine residue involved in disulfide bond formation present in this area may make this region highly sensitive to changes. It is possible that any mutation made in the loop region will not be beneficial to overall structural coherence.

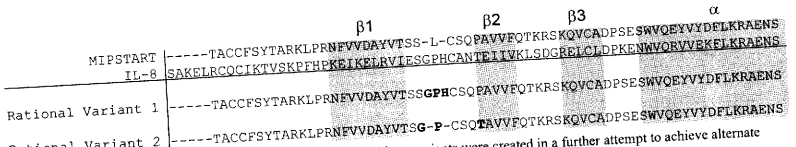


Figure 3.11 Rational loop variants of MIPSTART. Two rational loop variants were created in a further attempt to achieve alternate dimerization. The sequences of these two variants as compared to MIPSTART and IL-8 are demonstrated above.

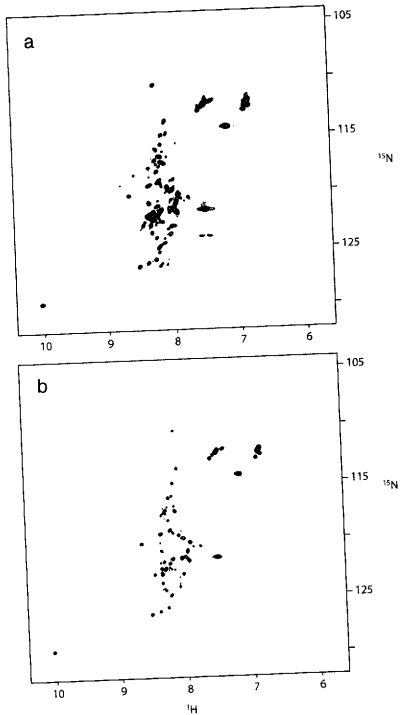


Figure 3.12 HSQC spectra of Rational Loop Variants. HSQC of Rational Loop Variants 1 (**a**) and 2 (**b**) are shown. The large clustering of peaks towards the center of both spectra indicates both proteins are unfolded *in vitro*.

Continuation of Loop Randomization

Although the loop region seems less than ideal for mutation, it was believed that continued random mutagenesis might be beneficial. Because the continued appearance of stop codons in the randomization process is not desired, new oligonucleotide primers for the loop region were ordered that would partially reduce the possibility of stop codon appearance. These primers were ordered with the specification that at the third position of a randomized codon every base except adenine could be inserted. Because adenine is the third base in two out of the three stop codons, this is expected to alleviate most of the problem of stop codon appearance. However, elimination of adenine from the third position, while not *completely* eliminating any amino acid from incorporation, does reduce the *likelihood* of incorporating some amino acids. The benefit of eliminating stop codons from passing the screening test was judged to outweigh this problem. The randomized portion of the loop region is outlined in Figure 3.13. Randomization and selection using this new primer system was continued until it was felt that all 194,481 possible variants had been examined (~480,000 colonies were screened). No variants that consistently passed the selection system were detected. This lack of positive variants is most likely attributed to the elimination of the stop codons.

Results of Helix Randomization

The C-terminal α-helix was the last region where randomization was believed to be necessary. As shown in Figure 1.8, interactions between the helix and the loop region are likely to be necessary for IL-8 type dimer formation. Based on an analysis of

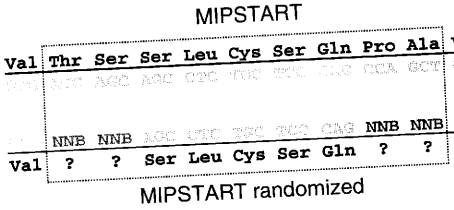


Figure 3.13 β1-β2 loop randomization without stop codons. Shown above is the randomization of the β1-β2 loop region using twelve randomized positions. The boxed region is the loop region. The letter "B" indicates substitution of every nitrogenous base except adenine. The randomized positions allow for a total of 194,481 different peptide combinations (including stop codons).

the structures of IL-8 and MIP-1β, the entire MIPSTART helix was not randomized; only residues that could form potential interactions with the loop region were included in the randomization process. Figure 3.14 demonstrates this concept and diagrams the randomized portion of the helix. Selection of the randomized helix library was carried out as previously described. One of the key advantages of helix randomization was the reduced role stop codons played in the selection process. Because the helix is at the C-terminal end of the protein, a stop codon in this region will still allow formation of a near-intact protein. Also, since deletion of helix residues is increasingly detrimental to dimer formation, it is unlikely that incorporation of stop codons will promote dimerization. This is in contrast to randomization of the loop and β1 regions where stop codons produced a short polypeptide that did, in theory, promote association by some non-specific means.

The success of the helix randomization was astounding in comparison to previous randomization attempts. According to colony counts approximately 6000 colonies have been screened, an amount that includes a significant portion of the 9,261 possible mutants produced by randomization. Currently, the process of producing helix mutants and subjecting them to the selection process is still ongoing. However, four variants that demonstrate phage immunity have been obtained. The sequences of these four variants are outlined in Figure 3.15. Although there does not seem to be any immediate correlation between the amino acid substitutions present in these four variants, there does seem to be a disproportionate number of positively charged residues.

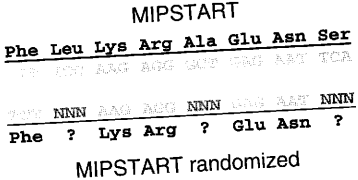


Figure 3.14 Helix randomization of MIPSTART. Shown above is the randomization of the helix region using nine randomized positions. All randomized residues are all on the face of the C-terminal helix that interacts with the loop region. The randomized positions allow for a total of 9,261 different peptide combinations (including stop codons).

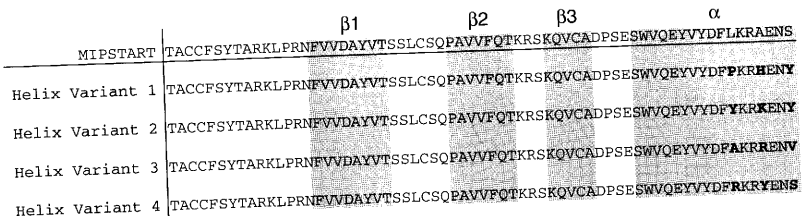


Figure 3.15 Results of helix randomization of MIPSTART. Four variants consistently demonstrated phage resistance after randomization of the helix region. The sequences of these four variants are diagrammed above.

Helix Variant 1 (HV1) was shuttled to the pET-32a(+) plasmid and inserted between the Nco1 and BamH1 restriction sites. HV1 was then expressed and purified by Dr. Erik Meyer of the LiWang laboratory. Because HV1 was inserted between Nco1 and BamH1, fusion tags were removed by proteolytic cleavage. The purified HV1 was then analyzed by NMR and an HSQC spectrum obtained. This spectrum is shown in Figure 3.16.

The HV1 HSQC spectrum demonstrates several peculiar properties. Initially, it was believed that there were too few peaks present in the spectrum to account for all amino acids of the protein. However, expression of HV1 was analyzed by SDS-PAGE and showed that the expression product migrates at the expected molecular weight. It is possible that the protein may be slightly unfolded and thus some peaks may be clustered near the center.

Subsequent to HV1 expression, production of the HV2 protein was begun. The HV2 gene was shuttled from the JH391 plasmid to the pET-32a(+) plasmid as described in Materials and Methods. As opposed to previous constructions using Nco1 and BamH1 in pET-32a(+), HV2 was inserted between the Nde1 and BamH1 restriction sites of the pET-32a(+) vector. By doing this, no fusion tags were added and a step was removed from the purification process (no proteolytic digestion). HV2 was simply refolded from a cell lysis pellet and purified using a single pass on a reverse phase C4 column. In addition, production of the HV2 protein without fusion tags eliminated the N-terminal methionine and alanine residues left as a proteolytic artifact. Recent

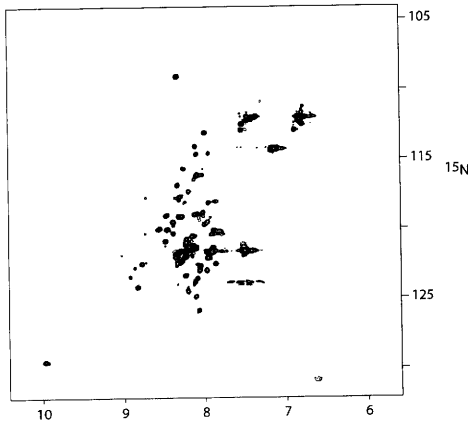


Figure 3.16 HSQC spectrum of HV1. Above is shown the HSQC of Helix Variant 1. The spectrum appears to be lacking enough peaks to account for all residues of the HV1 protein. However, some unfolding may have resulted in clustering of peaks near the spectrum center.

experiments in the LiWang laboratory suggested that these residues might be detrimental to an analysis of the HV2 NMR data. After successful purification of the HV2 protein product, HV3 was also expressed and purified (both proteins were expressed and purified by of Dr. Erik Meyer). An HSQC spectrum of each protein was obtained (Figure 3.17). The spectra of HV2 and HV3 demonstrate significant peak shifting away from the MIPSTART HSQC spectrum. This shifting indicates a definite change in protein conformation, and the implications of this will now be discussed.

Helix Variant Significance

The helix variants HV2 and HV3 presented are most worthy of note because they appear to differ significantly from the original MIPSTART template when analyzed with NMR (Figure 3.16). This difference suggests many possibilities, several of which are most intriguing to the work at hand. Shifts in peaks on an HSQC spectrum indicate a change in local environmental conditions for the N-H pair affected. This change can be the result of a nearby residue substitution that slightly alters the surrounding environs, or it can result from a much larger difference, such as newly formed dimer contacts. The amount of shifting seen in HV2 and HV3 is significant, but it is unclear if the change is large enough to warrant assumption of large, intermolecular alterations. However, it is not unreasonable to suggest that the shifting observed is the result of some type of weak dimeric contact affecting certain residues. It is equally likely that the shifting simply results from the alteration of the helix and surrounding local environment.

Determination of which of these possibilities is more accurate requires further structural investigation. As peak assignments for these molecules are not yet available, it is impossible to know which residues have shifted and the significance that shift may have. Peak assignment will require more time and effort but will be shortly underway.

In addition, analytical ultracentrifugation analysis of these mutants and the MIPSTART template will begin shortly. Hopefully, this analysis will indicate quickly if there is at least some dimeric tendency in the HV2 and HV3 proteins. Should these variants present indications of dimerization, a full structural analysis will begin, including determination of a high-resolution structure.

CHAPTER IV

SUMMARY AND CONCLUSION

Summary

The past three chapters of this thesis have presented and detailed an approach to studying the dimer interface of chemokines. Chapter I outlined the fundamental basis of the project, including its relevance to chemokine biochemistry and structural biology in general. The impetus rests largely on a desire to more fully understand the type of protein-protein interactions that lead to chemokine dimerization and, more broadly, quaternary structure formation in general. As a means of studying this process, a methodology was devised that subjected a monomeric MIP-1β variant titled MIPSTART to random mutagenesis and subsequent selection with a unique phage λ screening system. Chapter II described the steps taken to set up and expand the phage λ screening system so that it would best suit the needs of this project. All relevant controls, including MIPSTART and the IL-8 wild-type dimer performed as expected in the system. A high-throughput method of screening was also devised and a process for MIPSTART randomization created.

Chapter III detailed the progress of this project so far, including the successes and failures of the randomization/selection procedure to produce an alternate MIPSTART dimer. Randomization of the first β-strand and loop regions of the MIPSTART protein yielded somewhat disappointing results; four short polypeptides and

one full length but unfolded protein were the result of this portion of the process. This region of the protein seems recalcitrant to modification. However, the next area of randomization, the C-terminal α-helix, proved much more fruitful.

Randomization and selection of the C-terminal helix region of MIPSTART provided several promising protein variants that may possibly prove worthwhile. HV2 and HV3 have been expressed and purified and show considerable deviation from MIPSTART when analyzed using an HSQC NMR experiment. This deviance is promising, yet must be approached with some skepticism. As discussed in Chapter II, more examination of HV2 and HV3 is about to begin in order to better characterize the structural characteristics of each protein.

In addition, the HV1 protein was expressed and purified but gave unusual results when analyzed with NMR. An apparent lack of N-H peaks on the HSQC spectrum suggests that this protein is behaving oddly. Further investigation will be needed.

These results are promising and indicate at least a partial success of the project goal. A discussion of the implications these experiments follows in the conclusion section.

Conclusion

An exploration of the options for studying the chemokine dimer interface and, as a part of that, the universal properties that lead to dimerization in general has only just begun. The lack of knowledge concerning chemokine dimerization presents a wide array of experimental possibilities with which to investigate the process of monomer

association. The research presented here is a significant first step in understanding more about what leads to that dimerization and how to manipulate it, but much work remains to be done. The data gained from this project has already influenced the future direction of this research and what steps should be taken to best achieve the alternate dimer. Therefore, a discussion of the important conclusions and ideas to be garnered from the information presented follows. This discussion will focus on the significance of pertinent results as well as future steps that should be taken.

β-strand and Loop Region Variants

The random variants produced from randomization of the first β-strand and loop regions of the MIPSTART template were less than successful in forming alternate dimers. The persistent appearance of truncated variants produced by incorporation of a stop codon suggests that the small polypeptides are capable of promoting repressor dimerization. How this can be or what significance it may have for the phage system is nebulous at best. The possibility of further study of these small peptides by NMR analysis has not been ruled out.

The failure of the single full-length loop variant (LV3) to demonstrate stable structure *in vitro* raises interesting questions about the sensibility of mutating near the β1 and loop regions. Because MIPSTART is a small protein with a small hydrophobic core, it is possible that modification of residues near the center of the protein may simply cause too much structural strain and compromise integrity. The presence of a cysteine involved in disulfide bridge formation within the loop region may also present problems.

Any nearby residue substitutions that affect loop orientation may adversely affect disulfide formation. As the two disulfide bridges present in the MIPSTART structure are extremely important for structural integrity, this conclusion is not entirely unreasonable. However, this does not mean randomization of this region should be abandoned. Predicting the behavior of the cysteine in every possible circumstance is not feasible and there may be conditions such that substitution of nearby residue does not impair disulfide formation. Randomization of this area is not futile, but it should be approached with appreciation for the unique environmental circumstances.

Randomization for the Future

The randomization performed throughout the scope of this project has been fairly extensive, but further randomization is needed to fully exhaust the wide array of possibilities that might help in formation of the alternate dimer. As mentioned in Chapter III, randomization of the helix is nearing completion. However, as previously described, this current strategy of helix randomization is directed so that only residues thought to interact with the loop region are mutated. The next logical step in helix randomization, and one that will shortly begin, is mutation of the entire helix region rather than selected residues. This will provide a more general and broader approach to the idea of helix randomization. Furthermore, isolation of randomization to only single specific areas may be somewhat provincial in outlook. Randomization of multiple sites at once would give an even greater pool of variation from which an alternate dimer might arise. For example, randomization of the helix and loop regions simultaneously

would vastly increase the types and numbers of potential interactions that can form. Similarly, randomization of the β-strand, loop region, and α-helix concurrently would give a vast array of variants that could form many different intermolecular associations. While expansion of the randomization scope might serve to create a variant that no longer remotely resembles MIPSTART, the importance of expanding the randomization scope in the near future cannot be overemphasized.

Vice Versa

The intense scrutiny conducted in this attempt to force MIPSTART to dimerize in like fashion to IL-8 raises the interesting query of whether a similar approach can be taken for IL-8. In a reversal of the work here, can an IL-8 monomer be forced to dimerize like MIP-1β? While the answer to this is unknown, the process to achieve such a goal is already underway in the LiWang laboratory. Current investigations that are modeled closely off of the work done in this project are designed to explore the possibility of forming a MIP-type dimer from IL-8 monomers. Obviously, these projects are closely intertwined; this interconnection assists in looking at the projects from several viewpoints and hopefully will continue to promote ideas for new avenues of attack. It will be fascinating to see what answers come from both of these projects and how closely the alternate dimers of each species resemble the wild-type structure they are attempting to mimic.

Exploring the Heterodimer

While the main intent of this project has been the formation of a MIPSTART homodimer that resembles IL-8, the possibility of investigating heterodimer formation must not be excluded. Formation of a heterodimer between MIP-1β and IL-8 would be a potential springboard for ideas on how to achieve the goal of an alternate homodimer. In addition, the achievement of a MIP-IL8 heterodimer in and of itself would be a significant success. There are several approaches that could be taken to produce a heterodimer. In analogy to the process employed here, a yeast 2-hybrid system could be used to select for heterodimer formation between MIP-1β and IL-8. Also, use of rational design to promote association between the two species would no doubt be fruitful. Heterodimers could be sought between wild-type MIP-1β and IL-8, between MIPSTART and wild-type IL-8, between a monomeric IL-8 (IL8START) and wild-type MIP-1β, or, most challenging, a heterodimer between MIPSTART and IL8START. Regardless of the approach taken, whether it be rational or random, and whether it employs the use of biological screening systems or not, the point remains that the search for a heterodimer could be just as intriguing and informative as the search currently underway in this project.

Sine Die

In conclusion, presented here is a biochemical work on the chemokine dimer interface. In an attempt to investigate protein-protein interactions that generate the characteristic chemokine dimer, random mutagenesis coupled with a phage λ selection

system was used to alter MIP-1β so that it dimerizes in similar fashion to IL-8. While a stable alternate dimer of MIPSTART that resembles IL-8 has not yet been discovered, there is little doubt that such an achievement is rapidly approaching. Recently produced variants show great promise and will be intensely studied in the near future. An outline of future steps that seem necessary to completely satisfy this study on the chemokine dimer interface has also been presented. The work conducted on this project was most graciously funded by the National Science Foundation and Pfizer, Inc..

REFERENCES

REFERENCES

Amara, A. Lorthioir, O., Valenzuela, A., Margerus, A., Thelen, M., Montes, M., Virelizier, J.L., Delepierre, M., Baleux, F., Lortat-Jacob, H., and Arenzana-Seisdedos, F. (1999) *J. Biol. Chem.* **274**, 23916-23925.

Baggiolini, M., Dewald, B., and Moser, B. (1994) *Adv. Immunol.* **55**, 97-179.

Baggiolini, M., Dewald, B., and Moser, B. (1997) *Annu. Rev. Immunol.* **15**, 675-705.

Baldwin, E.T., Weber, I.T., St. Charles, R., Xuan, J.C., Appella, E., Yamada, M., Matsushima, K., Edwards, B.F., Clore, G.M., Gronenborn, A.M., and Wlodawer, A. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 502-506.

Beamer, L.J. and Pabo, C.O. (1992) *J. Mol. Biol.* **227**, 177-196.

Burke-Gaffney, A., Brooks, A., and Bogle, R.G. (2002) *Vasc. Pharmacol.* **38**, 283-292.

Burns, J.M., Lewis, G.K., and DeVico, A.L. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14499-14504.

Cherfils, J. and Janin, J. (1993) *Curr. Opin. Struct. Biol.* **3**, 265-269.

Chung, C.-W., Cooke, R.M., Proudfoot, A.E.I., and Wells, T.N.C. (1995) *Biochemistry* **34**, 9307-9314.

Clark-Lewis, I., Kim, K.-S., Rajarathnam, K., Gong, J.-H., Dewald, B., Moser, B., Baggiolini, M., and Sykes, B.D. (1995) *J. Leuokoc. Biol.* **57**, 703-711.

Clore, G.M., Appella, E., Yamada, M., Matushima, K., and Gronenborn, A.M. (1990) *Biochemistry* **29**, 1689-1696.

Clore, G.M. and Gronenborn, A.M. (1995) *FASEB J.* **9**, 57-62.

Crump, M.P., Gong, J.-H., Loetscher, P., Rajarathnam, K., Amara, A., Arenzana-Seisdedos, F., Virelizier, J.-L., Baggiolini, M., Sykes, B.D., and Clark-Lewis, I. (1997) *EMBO J.* **16**, 6996-7007.

Crump, M.P., Rajarathnam, K., Kim, K.-S., Clark-Lewis, I., and Sykes, B.D. (1998) *J. Biol. Chem.* **273**, 22471-22479.

Dragic, T., Litwin, V., Allaways, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P., and Paxton, W.A. (1996) *Nature* **381**, 667-673.

Fernandez, E.J. and Lolis, E. (2002) *Annu. Rev. Pharmacol. Toxicol.* **42**, 469-499.

Gong, J.-H., Uguccioni, M., Dewald, B., Baggiolini, M., and Clark-Lewis, I. (1996) *J. Biol. Chem.* **271**, 10521-10527.

Gong, J.-H. and Clark-Lewis, I. (1995) *J. Exp. Med.* **181**, 631-640.

Handel, T.M., and Domaille, P.J. (1996) *Biochemistry* **35**, 6569-6584.

Heringa, J. (2000) *Curr. Protein Pept. Sci.* **1**, 273-301.

Hu, J., O'Shea, E.K., Kim, P.S., and Sauer, R.T. (1990) *Science* **250**, 1400-1403.

Hu, J.C., Newell, N.E., Tidor, B., and Sauer, R.T. (1993) *Protein Sci.* **2**, 1072-1084.

Hu, J.C. (1995) *Structure* **3**, 431-433.

Janin, J. (1995) *Prog. Biophys. Mol. Biol.* **64**, 145-166.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958) *Nature* **181**, 662-666.

Kim, K.S., Clark-Lewis, I., and Sykes, B.D. (1994) *J. Biol. Chem.* **269**, 32909-32915.

Kim, Y.-I., and Hu, J.C. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7510-7514.

Koopman, W., and Krangel, M.S. (1997) *J. Biol. Chem.* **272**, 10103-10109.

Kuschert, G.S., Coulin, F., Power, C.A., Proudfoot, A.E., Hubbard, R.E., Hoogewerf, A.J., and Wells, T.N. (1999) *Biochemistry* **38**, 12959-12968.

Laurence, J.S., Blanpain, C., Parmentier, M., Burgner, J.W., and LiWang, P.J. (2000) *Biochemistry* **39**, 3401-3409.

Laurence, J.S., Blanpain, C., DeLeener, A., Parmentier, M., and LiWang, P.J. (2001) *Biochemistry* **40**, 4990-4999.

Lewin, B. (2000) *Genes* **7th ed.**, 325-345.

LiWang, A.C., Wang, Z.-X., Sun, Y., Peiper, S.C., and LiWang, P.J. (1999) *Protein Sci.* **8**, 2270-2280.

Lodi, P.J., Garrett, D.S., Kuszewski, J., Tsand, M., Weatherbee, J.A., Leonard, W.J., Gronenborn, A.M., and Clore, G.M. (1994) *Science* **263**, 1762-1767.

Lortat-Jacob, H., Grosdidier, A., and Imberty, A. (2002) *Proc. Natl. acad. Sci. U.S.A.* **99**, 1229-1234.

Lowman, H.B., Fairbrother, W.J., Slagle, P.H., Kabakoff, R., Liu, J., Shire, S., and Hebert C.A. (1997) *Protein Sci.* **6**, 598-608.

Marino-Ramirez, L. and Hu, J.C. (2002) *Yeast* **19**, 641-650.

Mayer, K.L., and Stone, M.J. (2000) *Biochemistry* **39**, 8382-8395.

Mayer, K.L., and Stone, M.J. (2001) *J. Biol. Chem.* **13**, 911-916.

McCornack, M.A., Cassidy, C.K., and LiWang, P.J. (2003) *J. Biol. Chem.* **278**, 1946-1956.

Mellado, M., Rodriguez-Frade, J.M., Manes, S., Martinez, A.C. (2001) *Annu. Rev. Immunol.* **19**, 397-421.

Meunier, S., Bernassau, J.M., Guillemot, J.C., Ferrara, P., and Darbon, H. (1997) *Biochemistry* **36**, 4412-4422.

Mizoue, L.S., Bazan, J.F., Johnson, E.C., and Handel, T.M. (1999) *Biochemistry* **38**, 1402-1414.

Murphy, P.M. (1994) *Annu. Rev. Immunol.* **12**, 593-633.

Paavola, C.D., Hemmerich, S. Grunberger, D., Polsky, I., Bloom, A., Freedman, R., Mulkins, M., Bhakta, S., McCarley, D., Wiesent, L., Wong, B., Jarnagin, K., and Handel, T.M. (1998) *J. Biol. Chem.* **273**, 33157-33165.

Pakianathan, D.R., Kuta, E.G., Artis, D.R., Skelton, N.J., and Hebert, C.A. (1997) *Biochemistry* **36**, 9642-9648.

Patel, D.D., Zachariah, J.P., Whichard, L.P. (2001) *Clin. Immunol.* **98**, 39-45.

Prescott, L.M., Harley, J.P., and Klein, D.A. (1999) *Microbiology* **4th ed.**, 365-369.

Proudfoot, A.E., Power, C.A., Hoogewerf, A.J., Montjovent, M.O., Borlat, F., Offord, R.E., and Wells, T.N. (1996) *J. Biol. Chem.* **271**, 2599-2603.

Qian, Y.Q., Johanson, K.O., and McDevitt, P. (1999) *J. Mol. Biol.* **294**, 1065-1072.

Rajarathnam, K., Sykes, B.D., Kay, C.M., Dewald, B., Geiser, T., Baggiolini, M., and Clark-Lews, I. (1994) *Science* **264**, 90-92.

Rajarathnam, K., Kay, C.M., Clark-Lewis, I., and Sykes, B.D. (1997) *Methods Enzymol.* **287**, 89-105

Rajarthnam, K., Li, Y., Rohrer, T., and Gentz, R. (2001) *J. Biol. Chem.* **276**, 4909-4916.

Reidhaar-Olson, J.F., Bowie, J.U., Breyer, R.M., Hu, J.C., Knight, K.L., Lim, W.A., *et al.* (1991) *Methods Enzymol* **208**, 564-586.

Rollins, B.J. (1997) *Blood* **90**, 909-928.

Royer, W.E., Love, W.E., and Fenderson, F.F. (1985) *Nature* **316**, 277-280.

Russel, P.J. (1996) *Genetics* **4th ed.**, 529-534.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* **2nd ed.**, E.4.

Shaanan, B., Lis, H., and Sharon, N. (1991) *Science* **254**, 862-866.

Shao, W., Jerva, F., West, J., Lolis, E., and Schweitzer, B.I. (1998) *Biochemistry* **37**, 8303-8313.

Shoichet, B.K., Kuntz, I. (1996) *Chem. Biol.* **3**, 151-156.

Simmons, G., Clapham, P.R., Picard, L., Offord, R.E., Rosenkilde, M.M., Schwartz, T.W., Buser, R., Wells, T.N., and Proudfoot, A.E. (1997) *Science* **276**, 276-279.

Skelton, N.J., Aspiras, F., Ogez, J., Schall, T.J. (1995) *Biochemistry* **34**, 5329-5342.

Smith, T. (1999) *Nat. Struc. Biol.* **6**, 411.

Snustad, D.P. and Simmons, M.J. (2000) *Principles of Genetics* **2nd ed.**, 630-637.

Springer, T.A. (1994) *Cell* **76**, 301-314.

St. Charles, R., Walz, D.A., and Edwards, B.F.P. (1989) *J. Biol. Chem.* **264**, 2092-2099.

Takano, T. (1984) *Methods and Applications in Crystallographic Computing*, 262.

Sternberg, M., Gabb, H., and Jackson, R. (1998) *Curr. Opin. Struct. Biol.* **8**, 250-256.

Steitz, S.A., Hasegawa, K., Chiang, S.L., Cobb, R.R., Castro, M.A., Lobl, T.J., Yamada, M., Lazarides, E., Cardarelli, P.M. (1998) *FEBS Lett.* **430**, 158-164.

Sticht, H., Escher, S.E., Schweimer, K., Forssman, W.-G., Rosch, P., and Aderman, K. (1999) *Biochemistry* **38**, 5995-6002.

Struyf, S., DeMeester, I., Scharpe, S., Lenaerts, J.P., Menten, P., Wang, J.M., Proost, P., and VanDamme, J. (1998) *Euro. J. Immunol.* **28**, 1262-1271.

Struyf, S., Proost, P., Schols, D., De Clercq, E., Opdenakker, G., Lenaerts, JP., Detheux, M., Parmentier, M., De Meester, I., Scharpe, S., Van Damme, J. (1999) *J. Immunol.* **162**, 4903-4909.

Tanaka, Y., Adams, D.H., Hubscher, S., Hirano, H., Siebenlist, U., and shaw, S. (1993) *Nature* **361**, 79-82.

Vaddi, K., Heller, M., and Newton, R.C. (1997) *The Chemokine Facts Book*, 23,38,46,86,107,118.

Volin, M.V., Shah, M.R., Tokuhira, M., and Haines, G.K., Woods, J.M., and Koch, A.E. (1998) *Clin. Immunol. Immunopathol.* **89**, 44-53.

Wagner, L., Yang, O.O., Garcia-Zepeda, E.A., Ge, Y., Kalams, S.A., Walker, B.D., Pasternack, M.S., and Luster, A.D. (1998) *Nature* **391**, 908-911.

Weber, M., Uguccioni, M., Baggiolini, M., Clark-Lewis, I., and Dahinden, C.A. (1996) *J. Exp. Med.* **183**, 681-685.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., and sykes, B.D. (1995) *J. Biomol. NMR* **6**, 135-140.

Wuthrich, K. (2001) *Nat. Struct. Biol.* **8**, 923-925.

Young, H., Roongta, V., Daly, T.J., and Mayo, K.H. (1999) *Biochem. J.* **338**, 591-598.

Zeng, X., Zhu, H., Lashuel, H.A., and Hu, J.C. (1997) *Protein Sci* **6**, 2218-2226.

Zeng, X., Herndon, A.M., and Hu, J.C. (1997) *Proc Natl Acad Sci U.S.A.* **94**, 3673-3678.

Zhang, Y.J., Rutledge, B.J., and Rollins, B.J. (1994) *J. Biol. Chem.* **269**, 15918-15924.

Zhang, Z., Murphy, A., Hu, J.C., and Kodadek, T. (1999) *Curr. Biol.* **9**, 417-420.

APPENDIX A

SELECTED PROTOCOLS

PCR Protocol for Random Mutagenesis

(for randomization near the 3' terminus)

Combine in 0.2ml PCR tube:

1µl	Template DNA
4µl	dNTPs
5µl	Thermopol Buffer
3µl	5' Sense Oligo Complementary to Template
3µl	3' Antisense Oligo with randomized positions
33µl	ddH_2O
1µl	DeepVent DNA Polymerase

Use the following thermocycle to amplify DNA with randomization:

Step 1:	95°C	30s
Step 2:	55°C	30s
Step 3:	72°C	30s

Repeat 30X

Amplified DNA is then purified by gel electrophoresis

PCR Protocol for Random Mutagenesis

(internal regions of a gene)

Prepare two simultaneous PCR mixtures in two different 0.2ml PCR tubes:

Reaction #1

- 1µl Template DNA
- 4µl dNTPs
- 5µl Thermopol Buffer
- 3µl 5' Sense oligo (0.1mg/ml) complementary to internal region of gene (with randomized bases)
- 3µl 3' Antisense oligo complementary to 3' terminus of gene
- 33µl ddH_2O
- 1µl DeepVent DNA Polymerase

Reaction #2

- 1µl Template DNA
- 4µl dNTPs
- 5µl Thermopol Buffer
- 3µl 5' Antisense oligo (0.1mg/ml) complementary to internal region of gene (with randomized bases)
- 3µl 3' Sense oligo complementary to 5' terminus of gene

33µl ddH_2O

1µl DeepVent DNA Polymerase

Amplify both reactions with the following thermocycle:

Step 1:	95°C	30s
Step 2:	55°C	30s
Step 3:	72°C	30s

Repeat 30X

Both "half-genes" are purified using gel electrophoresis then the following PCR mixture is prepared in a 0.2ml PCR tube:

4µl dNTPs

5µl Thermopol Buffer

4µl 5' Purified "half-gene" (Reaction #1)

4µl 3' Purified "half-gene" (Reaction#2)

26µl ddH_2O

1µl DeepVent DNA Polymerase

The following preamplification thermocycle is then run:

Step 1:	95°C	30s
Step 2:	55°C	30s
Step 3:	72°C	30s

Repeat 10X

After preamplification, the following is added to the PCR mixture:

3µl 5' Terminal Sense Oligo (0.1mg/ml) Complementary to Template

3µl 3' Terminal Antisense Oligo (0.1mg/ml) Complementary to Template

The following thermocycle is then run:

Step 1:	95°C	30s
Step 2:	55°C	30s
Step 3:	72°C	30s

Repeat 30X

After amplification, the randomized gene is isolated and purified using gel electrophoresis.

Preparation of $CaCl_2$ Competent Cells

1. Start overnight with 2-3ml of LB
2. Add 1-2ml of overnight into 100ml LB
3. Incubate and shake at 200 rpm
4. Grow to an OD_{550} of about 1
5. Aliquot culture into two separate (sterile) 50ml tubes
6. Let stand on ice for 20 minutes
7. Make 0.1 M $CaCl_2$ and put on ice
8. Centrifuge in swing bucket at 3500 rpm, 10 min, 4oC
9. Pour off supernatant
10. Resuspend pellet in 16ml of 0.1 M $CaCl_2$ per 50 ml culture
11. Ice for 30-60 min
12. Centrifuge at 3500 rpm, 10 min, 4^oC
13. Pour off supernatant
14. Resuspend pellet in 2-4 ml ice cold $CaCl_2$ per tube.
15. Repeat step 12-14 two or three times On the final repeat only add 1-2ml of $CaCl_2$
16. (Optional) let age 24hrs
17. If making glycerol freeze add 1ml 80% glycerol for every 4ml $CaCl_2$ suspended cells. Pre-chill tubes.

Electrocompetent Cell Preparation

Autoclave beforehand:

1.5L 10% glycerol (use deionized water!)
500ml 2XYT
Two medium size centrifuge bottles

Procedure:

1. From a freshly streaked plate of cells, pick colonies and inoculate 3 Falcon tubes with 2ml of 2XYT each (total of 6ml)

2. Shake cells overnight at 37°C

3. Inoculate 500ml autoclaved 2XYT with 6ml overnight cultures and shake at 37°C

4. Grow until A_{600} of .5-.7 (if this absorbance is not reached within 2.5-3 hours, then cells will NOT be competent)

5. Prechill the centrifuge rotor to ~2°C. Chill 10% glycerol on ice.

6. Centrifuge cells in medium size centrifuge bottles in a cold rotor at 4000rpm for 15 minutes

7. Remove as much of the supernatant as possible. It is better to lose a few cells than to leave any supernatant behind.

8. Gently resuspend the pellets in a total of 500ml ice-cold 10% glycerol taking care NOT to lyse them.

9. Centrifuge as in step 6.

10. Resuspend in 250ml of ice-cold 10% glycerol. Centrifuge as in step 6.

11. Resuspend in 125ml of ice-cold 10% glycerol. Centrifuge as in step 6.

12. Resuspend in a final volume of ~1ml ice-cold 10% glycerol. The cell concentration should be between $1x10^{10}$ and $3x10^{10}$ cells/ml.

13. Make 200µl aliquots of this suspension in 1.5ml tubes and store at -80°C.

1X ^{15}N Minimal Media

To 900ml autoclaved water add:

10X ^{15}N Minimal Salts	100ml
Trace Elements Solution	5ml
1M $MgCl_2 \cdot 6H_2O$	5ml
Thiamine(20mg/ml)	2ml
Ampicillin(100mg/ml)	2ml
20% Glucose	40ml

10X ^{15}N Minimal Salts
(g per 100ml of 10X)

KH_2PO_4	13.0g
K_2HPO_4	10.0g
Na_2HPO_4	9.0g
K_2SO_4	2.4g
$^{15}NH_4Cl$	1.1g

Adjust the pH of the solution to 7.20-7.4 using KOH. Autoclave.

Trace Elements Solution
(g/100ml)

Add each ingredient and wait several minutes before adding the next ingredient.

Rule of thumb: wait at least 10 minutes between adding each ingredient.

$CaCl_2 2H_2O$	0.6g
$FeSO_4 7H_2O$	0.6g
$MnCl_2 4H_2O$	0.115g
$CoCl_2 6H_2O$	0.08g
$ZnSO_4 7H_2O$	0.07g
$CuCl_2 2H_2O$	0.03g
H_3BO_3	0.002g
$(NH_4)_6Mo_7O_{24} 4H_2O$	0.025g
EDTA (disodium!)	0.50g

APPENDIX B

TABLE OF OLIGONUCLEOTIDE PRIMERS

IDT Reference Number	Oligonucleotide Name	Primer Sequence (5'-3')	Description
1112634	IL8_BamH1_Anti	GCG CGC GGA TCC TTA TGA ATT CTC AGC CC	Complementary to 3' end of IL-8 and MIPSTART; adds a BamH1 site to the 3' end
1157806	MIP(9)_Sal1_Nco1_S	CGC GCG TCG ACC ATG GCT ACC GCC TGC TGC TTT TC	Complementary to 5' end of MIPSTART; adds a Sal1 and Nco1 site to the 5' end
1371219	MIPAYVAENS_A28-Y29_Rdm_S	CGC AAC TTT GTG GTA GAT GNN NAT GTG ACC AGC AGC CTC TGC	Used in combination with its antisense partner to introduce 3 randomized positions into the β1 strand
1371220	MIPAYVAENS_A28-Y29_Rdm_A	GCA GAG GCT GCT GGT CAC ATN NNC ATG TAC CAC AAA GTT GCG	Used in combination with its sense partner to introduce 3 randomized positions into the β1 strand
1394681	MIPAYV_Y29_V30_Rdm_S	GTG GTA GAT GCT TNN NNG ACC AGC AGC CTC	Used incombination with its antisense partner to introduce 4 randomized positions into the β1 strand
1394682	MIPAYV_Y29_V30_Rdm_A	GAG GCT GCT	Used in

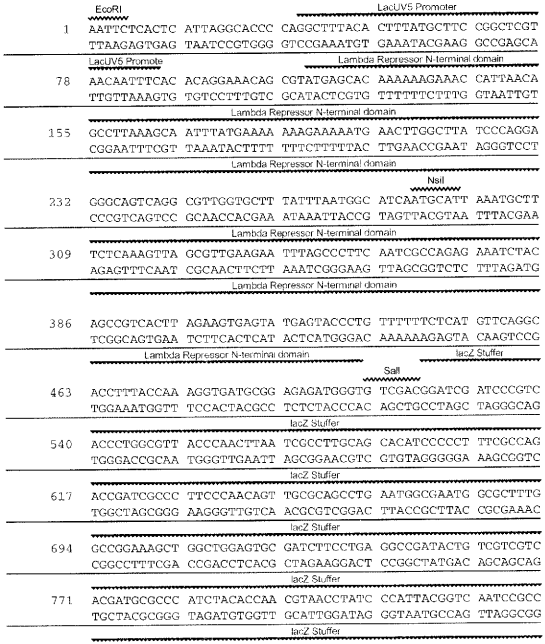
		GGT CNN NNA AGC ATC TAC CAC	combination with its sense partner to introduce 4 randomized positions into the β1 strand
1421679	MIP9AYV_D27-V30_Rdm_S	CGC AAC TTT GTG GTA GAT GNN NNN NNN NTG ACC AGC AGC CTC TGC	Used in combination with its antisense partner to introduce 9 randomized positions into the β1 strand
1421680	MIP9AYV_D27-V30_Rdm_A	GCA GAG GCT GCT GGT CAN NNN NNN NNC ATC TAC CAC AAA GTT GCG	Used in combination with its sense partner to introduce 9 randomized positions into the β1 strand
1704416	MIPAYV_D27-T31_Rdm_S	CCT CGC AAC TTT GTG GTA GNN NNN NNN NNN NCC AGC AGC CTC TGC TCC	Used in combination with its antisense partner to introduce 12 randomized positions into the β1 strand
1704417	MIPAYV_D27-T31_Rdm_A	GGA GCA GAG GCT GCT GGN NNN NNN NNN NNC TAC CAC AAA GTT GCG AGG AAG	Used in combination with its sense partner to introduce 12 randomized positions into the β1 strand
2752449*	MIP_Loop_Rdm_S	GTA GAT TAC TAT GAG NNN NNN AGC CTC TGC TCC CAG	Used in combination with its antisense partner to introduce 12

		NNN NNN GTG GTA TTC CAA ACC	randomized positions into the β1-β2 loop
2752450*	MIP_Loop_Rdm_A	GGT TTG GAA TAC CAC NNN NNN CTG GGA GCA GAG GCT NNN NNN CTC ATA GTA ATC TAC	Used in combination with its sense partner to introduce 12 randomized positions into the β1-β2 loop
4032370	MIP9_Loop_Rdm_NOA_S	GTA GAT GCT AGA GTG NNB NNB AGC CTC TGC TCC CAG NNB NNB GTG GTA TTC CAA ACC	Used in combination with its antisense partner to introduce 12 randomized positions into the β1-β2 loop; reduces stop codon appearance
4232861	MIP9_Loop_Rdm_NOT_A	GGT TTG GAA TAC CAC VNN VNN CTG GGA GCA GAG GCT VNN VNN CAC TCT AGC ATC TAC	Used in combination with its sense partner to introduce 12 randomized positions into the β1-β2 loop; reduces stop codon appearance
5011087	MIP(9)_Sal1_Nco1_Ext_S	CGC GCG TCG ACC ATG GCT ACC GCC TGC TGC TTT TCT TAC ACC	Complementary to 5' end of MIPSTART but longer than MIP(9)_Sal1_Nco1_S; adds a Sal1 and Nco1 site to the 5' end
5011088	MIPSTART_S33G_L34P_P38T_S	GTG ACC AGC GGC CCT TGC	Used in combination with

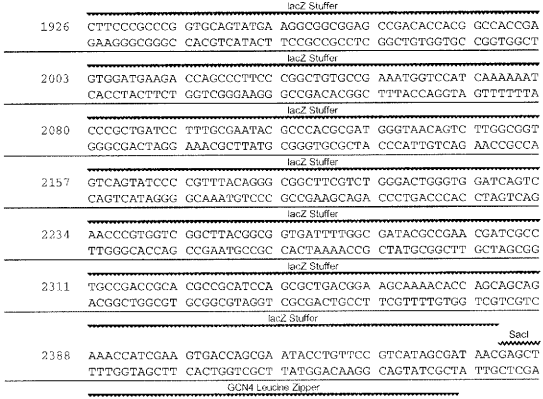
		TCC CAG ACC GCT GTG GT	its antisense partner to introduce internal mutations to create Rational Loop Variant XX
5011089	MIPSTART_S33G_L34P_P38T_A	ACC ACA GCG GTC TGG GAG CAA GGG CCG CTG GTC AC	Used in combination with its sense partner to introduce internal mutations to create Rational Loop Variant XX
5419027	MIPSTART_HelRdm_BamH1_A	GCG CGC GGA TCC TTA NNN ATT CTC NNN CCT CTT NNN AAA GTC ATA CAC GTA CTC CTG GAC	Complementary to the 3' end of MIPSTART; introduces randomization into selected residues; adds a BamH1 site
6540683	MIPSTART_HelTerm_BamH1_A	GCG CGC GGA TCC TTA TGN NNN NNN NNN NNT CTT CAA AAA GTC ATA CAC GTA CTC	Complementary to 3' end of MIPSTART; randomizes contiguous bases of the helix; adds a BamH1 site to the 3' end

APPENDIX C

COMPLETE SEQUENCE OF PLASMID JH391


EcoRI
LacUV5 Promoter
1
AATTCTCACTC ATTAGGCACCC CAGGCTTTACA CTTTATGCTTC CGGCTCGT
TTAAGAGTGAG TAATCCGTGGG GTCCGAAATGT GAAATACGAAG GCCGAGCA
LacUV5 Promote
Lambda Repressor N-terminal domain
78
AACAATTTCAC ACAGGAAACAG CGTATGAGCAC AAAAAAGAAAC CATTAACA
TTGTTAAAGTG TGTCCTTTGTC GCATACTCGTG TTTTTTCTTTG GTAATTGT
Lambda Repressor N-terminal domain
155
GCCTTAAAGCA ATTTATGAAAA AAAGAAAAATG AACTTGGCTTA TCCCAGGA
CGGAATTTCGT TAAATACTTTT TTTCTTTTTAC TTGAACCGAAT AGGGTCCT
Lambda Repressor N-terminal domain
NsiI
232
GGGCAGTCAGG CGTTGGTGCTT TATTTAATGGC ATCAATGCATT AAATGCTT
CCCGTCAGTCC GCAACCACGAA ATAAATTACCG TAGTTACGTAA TTTACGAA
Lambda Repressor N-terminal domain
309
TCTCAAAGTTA GCGTTGAAGAA TTTAGCCCTTC AATCGCCAGAG AAATCTAC
AGAGTTTCAAT CGCAACTTCTT AAATCGGGAAG TTAGCGGTCTC TTTAGATG
Lambda Repressor N-terminal domain
386
AGCCGTCACTT AGAAGTGAGTA TGAGTACCCTG TTTTTTCTCAT GTTCAGGC
TCGGCAGTGAA TCTTCACTCAT ACTCATGGGAC AAAAAAGAGTA CAAGTCCG
Lambda Repressor N-terminal domain
lacZ Stuffer
SalI
463
ACCTTTACCAA AGGTGATGCGG AGAGATGGGTG TCGACGGATCG ATCCCGTC
TGGAAATGGTT TCCACTACGCC TCTCTACCCAC AGCTGCCTAGC TAGGGCAG
lacZ Stuffer
540
ACCCTGGCGTT ACCCAACTTAA TCGCCTTGCAG CACATCCCCCT TTCGCCAG
TGGGACCGCAA TGGGTTGAATT AGCGGAACGTC GTGTAGGGGGA AAGCGGTC
lacZ Stuffer
617
ACCGATCGCCC TTCCCAACAGT TGCGCAGCCTG AATGGCGAATG GCGCTTTG
TGGCTAGCGGG AAGGGTTGTCA ACGCGTCGGAC TTACCGCTTAC CGCGAAAC
lacZ Stuffer
694
GCCGGAAAGCT GGCTGGAGTGC GATCTTCCTGA GGCCGATACTG TCGTCGTC
CGGCCTTTCGA CCGACCTCACG CTAGAAGGACT CCGGCTATGAC AGCAGCAG
lacZ Stuffer
771
ACGATGCGCCC ATCTACACCAA CGTAACCTATC CCATTACGGTC AATCCGCC
TGCTACGCGGG TAGATGTGGTT GCATTGGATAG GGTAATGCCAG TTAGGCGG
lacZ Stuffer
848
GGTTGTTACTC GCTCACATTTA ATGTTGATGAA AGCTGGCTACA GGAAGGCC
CCAACAATGAG CGAGTGTAAAT TACAACTACTT TCGACCGATGT CCTTCCGG


lacZ Stuffer
925
TAACTCGGCGT TTCATCTGTGG TGCAACGGGCG CTGGGTCGGTT ACGGCCAG
ATTGAGCCGCA AAGTAGACACC ACGTTGCCCGC GACCCAGCCAA TGCCGGTC
lacZ Stuffer
1002
ACCTGAGCGCA TTTTTACGCGC CGGAGAAAACC GCCTCGCGGTG ATGGTGCT
TGGACTCGCGT AAAAATGCGCG GCCTCTTTTGG CGGAGCGCCAC TACCACGA
lacZ Stuffer
1079
GAAGATCAGGA TATGTGGCGGA TGAGCGGCATT TTCCGTGACGT CTCGTTGC
CTTCTAGTCCT ATACACCGCCT ACTCGCCGTAA AAGGCACTGCA GAGCAACG
lacZ Stuffer
1156
CGATTTCCATG TTGCCACTCGC TTTAATGATGA TTTCAGCCGCG CTGTACTG
GCTAAAGGTAC AACGGTGAGCG AAATTACTACT AAAGTCGGCGC GACATGAC
lacZ Stuffer
1233
AGTTGCGTGAC TACCTACGGGT AACAGTTTCTT TATGGCAGGGT GAAACGCA
TCAACGCACTG ATGGATGCCCA TTGTCAAAGAA ATACCGTCCCA CTTTGCGT
lacZ Stuffer
1310
GGCGGTGAAAT TATCGATGAGC GTGGTGGTTAT GCCGATCGCGT CACACTAC
CCGCCACTTTA ATAGCTACTCG CACCACCAATA CGGCTAGCGCA CTGTGATG
lacZ Stuffer
1387
GTGGAGCGCCG AAATCCCGAAT CTCTATCGTGC GGTGGTTGAAC TGCACACC
CACCTCGCGGC TTTAGGGCTTA GAGATAGCACG CCACCAACTTG ACGTGTGG
lacZ Stuffer
1464
AAGCCTGCGAT GTCGGTTTCCG CGAGGTGCGGA TTGAAAATGGT CTGCTGCT
TTCGGACGCTA CAGCCAAAGGC GCTCCACGCCT AACTTTTACCA GACGACGA
lacZ Stuffer
1541
CGAGGCGTTAA CCGTCACGAGC ATCATCCTCTG CATGGTCAGGT CATGGATG
GCTCCGCAATT GGCAGTGCTCG TAGTAGGAGAC GTACCAGTCCA GTACCTAC
lacZ Stuffer
1618
GCTGATGAAGC AGAACAACTTT AACGCCGTGCG CTGTTCGCATT ATCCGAAC
CGACTACTTCG TCTTGTTGAAA TTGCGGCACGC GACAAGCGTAA TAGGCTTG
lacZ Stuffer
1695
ACCGCTACGGC CTGTATGTGGT GGATGAAGCCA ATATTGAAACC CACGGCAT
TGGCGATGCCG GACATACACCA CCTACTTCGGT TATAACTTTGG GTGCCGTA
lacZ Stuffer
1772
GATCCGCGCTG GCTACCGGCGA TGAGCGAACGC GTAACGCGAAT GGTGCAGC
CTAGGCGCGAC CGATGGCCGCT ACTCGCTTGCG CATTGCGCTTA CCACGTCG
lacZ Stuffer
1849
CATCTGGTCGC TGGGGAATGAA TCAGGCCACGG CGCTAATCACG ACGCGCTG
GTAGACCAGCG ACCCCTTACTT AGTCCGGTGCC GCGATTAGTGC TGCGCGAC


lacZ Stuffer
1926
CTTCCCGCCCG GTGCAGTATGA AGGCGGCGGAG CCGACACCACG GCCACCGA
GAAGGGCGGGC CACGTCATACT TCCGCCGCCTC GGCTGTGGTGC CGGTGGCT
lacZ Stuffer
2003
GTGGATGAAGA CCAGCCCTTCC CGGCTGTGCCG AAATGGTCCAT CAAAAAAT
CACCTACTTCT GGTCGGGAAGG GCCGACACGGC TTTACCAGGTA GTTTTTTA
lacZ Stuffer
2080
CCCGCTGATCC TTTGCGAATAC GCCCACGCGAT GGGTAACAGTC TTGGCGGT
GGGCGACTAGG AAACGCTTATG CGGGTGCGCTA CCCATTGTCAG AACCGCCA
lacZ Stuffer
2157
GTCAGTATCCC CGTTTACAGGG CGGCTTCGTCT GGGACTGGGTG GATCAGTC
CAGTCATAGGG GCAAATGTCCC GCCGAAGCAGA CCCTGACCCAC CTAGTCAG
lacZ Stuffer
2234
AACCCGTGGTC GGCTTACGGCG GTGATTTTGGC GATACGCCGAA CGATCGCC
TTGGGCACCAG CCGAATGCCGC CACTAAAACCG CTATGCGGCTT GCTAGCGG
lacZ Stuffer
2311
TGCCGACCGCA CGCCGCATCCA GCGCTGACGGA AGCAAAACACC AGCAGCAG
ACGGCTGGCGT GCGGCGTAGGT CGCGACTGCCT TCGTTTTGTGG TCGTCGTC
lacZ Stuffer
SacI
2388
AAACCATCGAA GTGACCAGCGA ATACCTGTTCC GTCATAGCGAT AACGAGCT
TTTGGTAGCTT CACTGGTCGCT TATGGACAAGG CAGTATCGCTA TTGCTCGA
GCN4 Leucine Zipper
SpeI
BamHI
2465
AACGAAGTTGC GCGCCTGAAAA AACTAGTTGGT GAACGTTGAGG ATCCGGCG
TTGCTTCAACG CGCGGACTTTT TTGATCAACCA CTTGCAACTCC TAGGCCGC
T7 terminator
2542
TGAGTTGGCTG CTGCCACCGCT GAGCAATAACT AGCATAACCCC TTGGGGCC
ACTCAACCGAC GACGGTGGCGA CTCGTTATTGA TCGTATTGGGG AACCCCGG
T7 terminator
EcoRV
2619
TGCTGAAAGGA GGAACTATATG CCCGGATATCG TCCATTCCGAC AGCATCGC
ACGACTTTCCT CCTTGATATAC GGGCCTATAGC AGGTAAGGCTG TCGTAGCG
2696
CTATATGCGTT GATGCAATTTC TATGCGCACCC GTTCTCGGAGC ACTGTCCG
GATATACGCAA CTACGTTAAAG ATACGCGTGGG CAAGAGCCTCG TGACAGGC
2773
GCTCGCTTCGC TACTTGGAGCC ACTATCGACTA CGCGATCATGG CGACCACA
CGAGCGAAGCG ATGAACCTCGG TGATAGCTGAT GCGCTAGTACC GCTGGTGT
2850
CCCTTGAGAGC CTTCAACCCAG TCAGCTCCTTC CGGTGGGCGCG GGGCATGA
GGGAACTCTCG GAAGTTGGGTC AGTCGAGGAAG GCCACCCGCGC CCCGTACT
2927
CTTCTTTATCA TGCAACTCGTA GGACAGGTGCC GGCAGCGCTCT GGGTCATT
GAAGAAATAGT ACGTTGAGCAT CCTGTCCACGG CCGTCGCGAGA CCCAGTAA

3004	GCGCGACGATG	ATCGGCCTGTC	GCTTGCGGTAT	TCGGAATCTTG	CACGCCCT
	CGCGCTGCTAC	TAGCCGGACAG	CGAACGCCATA	AGCCTTAGAAC	GTGCCGGA
3081	GCCACCAAACG	TTTCGGCCAGA	AGCAGGCCATT	ATCGCCGGCAT	GGCGGCCG
	CGGTGGTTTGC	AAAGCCGCTCT	TCGTCCGGTAA	TAGCGGCCGTA	CCGCCGGC
3158	GTTCGCGACGC	GAGGCTGGATG	GCCTTCCCCAT	TATGATTCTTC	TCGCTTCC
	CAAGCGCTGCG	CTCCGACCTAC	CGGAAGGGGTA	ATACTAAGAAG	AGCGAAGG
3235	AGGCCATGCTG	TCCAGGCAGGT	AGATGACGACC	ATCAGGGACAG	CTTCAAGG
	TCCGGTACGAC	AGGTCCGTCCA	TCTACTGCTGG	TAGTCCCTGTC	GAAGTTCC
3312	ACTTCGATCAC	TGGACCGCTGA	TCGTCACGGCG	ATTTATGCCGC	CTCGGCGA
	TGAAGCTAGTG	ACCTGGCGACT	AGCAGTGCCGC	TAAATACGGCG	GAGCCGCT
3389	TGTAGGCCCCG	CCCTATACCTT	GTCTGCCTCCC	CGCGTTGCGTC	GCGGTGCA
	ACATCCGCGGC	GGGATATGGAA	CAGACGGAGGG	GCGCAACGCAG	CGCCACGT
3466	TGGAAGCCGGC	GGCACCTCGCT	AACGGATTCAC	CACTCCAAGAA	TTGGAGCC
	ACCTTCGGCCG	CCGTGGAGCGA	TTGCCTAAGTG	GTGAGGTTCTT	AACCTCGG
3543	ATGCGCAAACC	AACCCTTGGCA	GAACATATCCA	TCGCGTCCGCC	ATCTCCAG
	TACGCGTTTGG	TTGGGAACCGT	CTTGTATAGGT	AGCGCAGGCGG	TAGAGGTC
3620	AGCGTTGGGTC	CTGGCCACGGG	TGCGCATGATC	GTGCTCCTGTC	GTTGAGGA
	TCGCAACCCAG	GACCGGTGCCC	ACGCGTACTAG	CACGAGGACAG	CAACTCCT
3697	TACTGGTTAGC	AGAATGAATCA	CCGATACGCGA	GCGAACGTGAA	GCGACTGC
	ATGACCAATCG	TCTTACTTAGT	GGCTATGCGCT	CGCTTGCACTT	CGCTGACG
3774	CAACAACATGA	ATGGTCTTCGG	TTTCCGTGTTT	CGTAAAGTCTG	GAAACGCG
	GTTGTTGTACT	TACCAGAAGCC	AAAGGCACAAA	GCATTTCAGAC	CTTTGCGC
3851	TTCCGGATCTG	CATCGCAGGAT	GCTGCTGGCTA	CCCTGTGGAAC	ACCTACAT
	AAGGCCTAGAC	GTAGCGTCCTA	CGACGACCGAT	GGGACACCTTG	TGGATGTA
3928	ACCCTGAGTGA	TTTTTCTCTGG	TCCCGCCGCAT	CCATACCGCCA	GTTGTTTA
	TGGGACTCACT	AAAAAGAGACC	AGGGCGGCGTA	GGTATGGCGGT	CAACAAAT
4005	CATGTTCATCA	TCAGTAACCCG	TATCGTGAGCA	TCCTCTCTCGT	TTCATCGG
	GTACAAGTAGT	AGTCATTGGGC	ATAGCACTCGT	AGGAGAGAGCA	AAGTAGCC
4082	TCCCCCTTACA	CGGAGGCATCA	AGTGACCAAAC	AGGAAAAAACC	GCCCTTAA
	AGGGGGAATGT	GCCTCCGTAGT	TCACTGGTTTG	TCCTTTTTTGG	CGGGAATT
4159	ACATTAACGCT	TCTGGAGAAAC	TCAACGAGCTG	GACGCGGATGA	ACAGGCAG
	TGTAATTGCGA	AGACCTCTTTG	AGTTGCTCGAC	CTGCGCCTACT	TGTCCGTC
4236	CGCTCATGAGC	TTTACCGCAGC	TGCCTCGCGCG	TTTCGGTGATG	ACGGTGAA
	GCGACTACTCG	AAATGGCGTCG	ACGGAGCGCGC	AAAGCCACTAC	TGCCACTT
4313	AGACGGTCACA	GCTTGTCTGTA	AGCGGATGCCG	GGAGCAGACAA	GCCCGTCA
	TCTGCCAGTGT	CGAACAGACAT	TCGCCTACGGC	CCTCGTCTGTT	CGGGCAGT
4390	TGTCGGGGCGC	AGCCATGACCC	AGTCACGTAGC	GATAGCGGAGT	GTATACTG
	ACAGCCCCGCG	TCGGTACTGGG	TCAGTGCATCG	CTATCGCCTCA	CATATGAC
4467	ATTGTACTGAG	AGTGCACCATA	TGCGGTGTGAA	ATACCGCACAG	ATGCGTAA
	TAACATGACTC	TCACGTGGTAT	ACGCCACACTT	TATGGCGTGTC	TACGCATT

```
4544  TTCCGCTTCCT CGCTCACTGAC TCGCTGCGCTC GGTCGTTCGGC TGCGGCGA
      AAGGCGAAGGA GCGAGTGACTG AGCGACGCGAG CCAGCAAGCCG ACGCCGCT

4621  TAATACGGTTA TCCACAGAATC AGGGGATAACG CAGGAAAGAAC ATGTGAGC
      ATTATGCCAAT AGGTGTCTTAG TCCCCTATTGC GTCCTTTCTTG TACACTCG

4698  CGTAAAAAGGC CGCGTTGCTGG CGTTTTTCCAT AGGCTCCGCCC CCCTGACG
      GCATTTTTCCG GCGCAACGACC GCAAAAAGGTA TCCGAGGCGGG GGGACTGC

4775  TCAGAGGTGGC GAAACCCGACA GGACTATAAAG ATACCAGGCGT TTCCCCCT
      AGTCTCCACCG CTTTGGGCTGT CCTGATATTTC TATGGTCCGCA AAGGGGGA

4852  TTCCGACCCTG CCGCTTACCGG ATACCTGTCCG CCTTTCTCCCT TCGGGAAG
      AAGGCTGGGAC GGCGAATGGGC TATGGACAGGC GGAAAGAGGGA AGCCCTTC

4929  TGTAGGTATCT CAGTTCGGTGT AGGTCGTTCGC TCCAAGCTGGG CTGTGTGC
      ACATCCATAGA GTCAAGCCACA TCCAGCAAGCG AGGTTCGACCC GACACACG

5006  CTGCGCCTTAT CCGGTAACTAT CGTCTTGAGTC CAACCCGGTAA GACACGAC
      GACGCGGAATA GGCCATTGATA GCAGAACTCAG GTTGGGCCATT CTGTGCTG

5083  GTAACAGGATT AGCAGAGCGAG GTATGTAGGCG GTGCTACAGAG TTCTTGAA
      CATTGTCCTAA TCGTCTCGCTC CATACATCCGC CACGATGTCTC AAGAACTT

5160  AGAAGGACAGT ATTTGGTATCT GCGCTCTGCTG AAGCCAGTTAC CTTCGGAA
      TCTTCCTGTCA TAAACCATAGA CGCGAGACGAC TTCGGTCAATG GAAGCCTT

5237  CAAACAAACCA CCGCTGGTAGC GGTGGTTTTTT TGTTTGCAAGC AGCAGATT
      GTTTGTTTGGT GGCGACCATCG CCACCAAAAAA ACAAACGTTCG TCGTCTAA

5314  AAGATCCTTTG ATCTTTTCTAC GGGGTCTGACG CTCAGTGGAAC GAAAACTC
      TTCTAGGAAAC TAGAAAAGATG CCCCAGACTGC GAGTCACCTTG CTTTTGAG

5391  TTATCAAAAAG GATCTTCACCT AGATCCTTTTA CGCGCCCTGTA GCGGCGCA
      AATAGTTTTTC CTAGAAGTCGA TCTAGGAAAAT GCGCGGGACAT CGCCGCGT

5468  CGCGCAGCGTG ACCGCTACACT TGCCAGCGCCC TAGCGCCCGCT CCTTTCGC
      GCGCGTCGCAC TGGCGATGTGA ACGGTCGCGGG ATCGCGGGCGA GGAAAGCG

5545  TTCGCCGGCTT TCCCCGTCAAG CTCTAAATCGG GGGCTCCCTTT AGGGTTCC
      AAGCGGCCGAA AGGGGCAGTTC GAGATTTAGCC CCGGAGGGAAA TCCCAAGG

5622  CCCCAAAAAAC TTGATTTGGGT GATGGTTCACG TAGTGGGCCAT CGCCCTGA
      GGGGTTTTTTG AACTAAACCCA CTACCAAGTGC ATCACCCGGTA GCGGGACT

5699  TGGAGTCCACG TTCTTTAATAG TGGACTCTTGT TCCAAACTGGA ACAACACT
      ACCTCAGGTGC AAGAAATTATC ACCTGAGAACA AGGTTTGACCT TGTTGTGA

5776  GATTTATAAGG GATTTTGCCGA TTTCGGCCTAT TGGTTAAAAAA TGAGCTGA
      CTAAATATTCC CTAAAACGGCT AAAGCCGGATA ACCAATTTTTT ACTCGACT

5853  TAACAAAATAT TAACGTTTACA ATTTAAATATT TGCTTATACAA TCTTCCTG
      ATTGTTTTATA ATTGCAAATGT TAAATTTATAA ACGAATATGTT AGAAGGAC
                                                    Ampicillin

5930  CCGGGGTAAAT CAATCTAAAGT ATATATGAGTA AACTTGGTCTG ACAGTTAC
      GGCCCCATTTA GTTAGATTTCA TATATACTCAT TTGAACCAGAC TGTCAATG
```


Ampicillin Resistance
6007
TCTCAGCGATC TGTCTATTTCG TTCATCCATAG TTGCCTGACTC CCCGTCGT
AGAGTCGCTAG ACAGATAAAGC AAGTAGGTATC AACGGACTGAG GGGCAGCA
Ampicillin Resistance
6084
TTACCATCTGG CCCCAGTGCTG CAATGATACCG CGAGACCCACG CTCACCGG
AATGGTAGACC GGGGTCACGAC GTTACTATGGC GCTCTGGGTGC GAGTGGCC
Ampicillin Resistance
6161
GCCAGCCGGAA GGGCCGAGCGC AGAAGTGGTCC TGCAACTTTAT CCGCCTCC
CGGTCGGCCTT CCCGGCTCGCG TCTTCACCAGG ACGTTGAAATA GGCGGAGG
Ampicillin Resistance
6238
AAGCTAGAGTA AGTAGTTCGCC AGTTAATAGTT TGCGCAACGTT GTTGCCAT
TTCGATCTCAT TCATCAAGCGG TCAATTATCAA ACGCGTTGCAA CAACGGTA
Ampicillin Resistance
6315
TCGTCGTTTGG TATGGCTTCAT TCAGCTCCGGT TCCCAACGATC AAGGCGAG
AGCAGCAAACC ATACCGAAGTA AGTCGAGGCCA AGGGTTGCTAG TTCCGCTC
Ampicillin Resistance
6392
AAAAGCGGTTA GCTCCTTCGGT CCTCCGATCGT TGTCAGAAGTA AGTTGGCC
TTTTCGCCAAT CGAGGAAGCCA GGAGGCTAGCA ACAGTCTTCAT TCAACCGG
Ampicillin Resistance
6469
CAGCACTGCAT AATTCTCTTAC TGTCATGCCAT CCGTAAGATGC TTTTCTGT
GTCGTGACGTA TTAAGAGAATG ACAGTACGGTA GGCATTCTACG AAAAGACA
Ampicillin Resistance
6546
TTCTGAGAATA GTGTATGCGGC GACCGAGTTGC TCTTGCCCGGC GTCAACAC
AAGACTCTTAT CACATACGCCG CTGGCTCAACG AGAACGGGCCG CAGTTGTG
Ampicillin Resistance
6623
AACTTTAAAAG TGCTCATCATT GGAAAACGTTC TTCGGGGCGAA AACTCTCA
TTGAAATTTTC ACGAGTAGTAA CCTTTTGCAAG AAGCCCCGCTT TTGAGAGT
Ampicillin Resistance
6700
GTTCGATGTAA CCCACTCGTGC ACCCAACTGAT CTTCAGCATCT TTTACTTT
CAAGCTACATT GGGTGAGCACG TGGGTTGACTA GAAGTCGTAGA AAATGAAA
Ampicillin Resistance
6777
ACAGGAAGGCA AAATGCCGCAA AAAAGGGAATA AGGGCGACACG GAAATGTT
TGTCCTTCCGT TTTACGGCGTT TTTTCCCTTAT TCCCGCTGTGC CTTTACAA
6854
ATATTATTGAA GCATTTATCAG GGTTATTGTCT CATGAGCGGAT ACATATTT
TATAATAACTT CGTAAATAGTC CCAATAACAGA GTACTCGCCTA TGTATAAA
6931
TAGGGGTTCCG CGCACATTTCC CCGAAAAGTGC CACCTGACGTC TAAGAAAC
ATCCCCAAGGC GCGTGTAAAGG GGCTTTTCACG GTGGACTGCAG ATTCTTTG

```
                                                        EcoRI
                                                          v
7008   AAAAATAGGCG TATCACGAGGC CCTTTCGTCTT CAAG
       TTTTTATCCGC ATAGTGCTCCG GGAAAGCAGAA GTTC
```

VITA

Garret Lance Hayes was born to Willie and Lawanda Hayes in Pt. Arthur, Texas on January 2, 1981. He attended Vidor High School in Vidor, TX where he was an active member of the Vidor High School Band and Mock Trial Team. During his high school career, Garret obtained the level of second degree black belt, was an active instructor of TaeKwonDo, regularly competed in state piano competition, and worked as a weekend disc jockey for KQXY 94.1. Garret graduated from Vidor High School in 1999 as valedictorian with National Merit Scholar designation and continued on to Texas A&M University where he majored in biochemistry. While a freshman at Texas A&M, Garret met Professor Patrcia LiWang and began working in her research laboratory as an undergraduate assistant and later as an undergraduate researcher. The research he conducted while working with Professor LiWang allowed him to receive a Goldwater Research Fellowship, a Pfizer Summer Undergraduate Research Fellowship, and an NSF Research Experience for Undergraduates Fellowship. Garret was also involved in the Texas A&M Young Democrats as Vice President, the University Scholars program, the Research Fellows program, and several honor societies. He currently maintains a 4.0 grade point average. Garret will be pursuing graduate studies in biochemistry in the fall of 2003.