UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Molecular Mechanisms Behind Receptor Signaling

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biology

by

Katherine Yoshie Blain

Committee in charge:

Professor Senyon Choe, Chair
Professor Tony Hunter, Co-Chair
Professor Partho Ghosh
Professor Joan Heller Brown
Professor Kit Pogliano

2010

The Dissertation of Katherine Yoshie Blain is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2010

DEDICATION

To my parents,

Alvin and Lorna Blain,

Who made all of this possible with their endless love, encouragement, and patience.
Thank you for making the sacrifices that enabled me to obtain a higher education.

And also to my family, friends, and loved ones,
Who have been there to support me throughout my life, never leaving my side.

EPIGRAPH

We know nothing in reality; for truth lies in an abyss.

***Democritus**, Greek Philosopher (c. 420 BCE)*

The important thing in science is not so much to obtain new facts
as to discover new ways of thinking about them.

***Sir William Bragg**, British physicist (1862 - 1942)*

There is no area of the world that should not be investigated by scientists.
There will always remain some questions that have not been answered.
In general, these are the questions that have not yet been posed.

***Linus Pauling**, American Chemist (1901-1994)*

Let me tell you the secret that has led me to my goal.
My strength lies solely in my tenacity.

***Louis Pasteur**, French biologist & bacteriologist (1822 - 1895)*

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
LMPG	1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-(1-glycerol)]
MES	2-(N-morpholino) ethanesulfonic acid
ATP	Adenosine-5'-triphosphate
[γ-32P] ATP	Adenosine-5'-triphosphate gamma phosphorus-32
E4	Adenovirus early region 4
ALS	Advanced Light Source
ABD	Agonist binding domain
Ala, A	Alanine
AIDS	Aquired Immune Deficiency Syndrome
A. fulgidus	*Archeoglobus fulgidus*
Arg, R	Arginine
Asn, N	Asparagine
Asp, D	Aspartic acid
k_{on}	association rate constant
B. subtilis	*Bacillus subtilis*
βMe	beta mercaptoethanol
β-gal	beta-galactosidase
Ca^{2+}	Calcium
$CaCl_2$	Calcium chloride
CA	catalytic and ATP-binding

CCL	CC Chemokine Ligand
CCR5	CC chemokine receptor
CF	cell-free
CSMP	Center for Structures of Membrane Proteins
CD	cluster of differentiation
CDL	combinatorial dual-labeling
CTD	COOH-terminal domain
Cys, C	Cysteine
Da	dalton
DNA	deoxyribonucleic acid
DHp	dimerization and histidine phosphotransfer
K_d	dissociation constant
k_{off}	dissociation rate constant
DTT	dithiothreitol
Endo H	endoglycosidase H
EAM	energy absorbing molecules
EC	equilibrium concentration
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ECD	extracellular domain
FPLC	fast protein liquid chromatography
FLIPR	fluoroumetric imaging plate reader

FC-12	FOS-choline-12
GPCR	G protein-coupled receptor
Glu, E	Glutamic acid
Gln, Q	Glutamine
Gly, G	Glycine
g	gram
HBSS	Hank's Balanced Salt Solution
His, H	Histidine
HK	histidine kinase
HAMP	Histidine kinase, Adenyl cyclase, Methyl-accepting protein and Phosphatase
h	hour
HIV-1	human immunodeficiency virus type 1
hIMP3	human integral membrane protein 3
IC	inhibitory concentration
IMP	integral membrane protein
I	iodine
iGluR	ionotropic glutamate receptor
Ile, I	Isoleucine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	kilo dalton
Leu, L	Leucine

LBD	ligand binding domain
L	Liter
LLG	log-likelihood gain
Lys, K	Lysine
M-tropic	macrophage-tropic
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulfate
MS	mass spectrometry
MPDB	Membrane Protein Data Bank
Mistic	Membrane-Integrating Sequence for Translating IMP Constructs
Met, M	Methionine
MCP	methyl-accepting chemotaxis protein
μL	microliter
μM	micromolar
mg	milligram
min	minute
M	Molar; mol L^{-1}
MR	molecular replacement
MWCO	molecular weight cutoff
MAD	multiple wavelength anomalous dispersion
NMDA	*N*-methyl-D-aspartate
nM	nanomolar

NTD	NH_2-terminal domain
Ni-NTA	nickel-nitrilotriacetic acid
N_2	nitrogen
NCS	non-crystallographic symmetry
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect
NOESY	nuclear overhauser enhancement spectroscopy
ONPG	*o*-nitrophenyl β-D-galactopyranoside
ORF3	open reading frame 3
OD	optical density
Redox	oxidation-reduction
PRE	paramagnetic relaxation enhancement
PBMC	peripheral blood mononuclear cells
Phe, F	Phenylalanine
PMSF	phenylmethanesulfonyl fluoride
PEG	polyethylene glycol
PCR	polymerase chain reaction
KCl	potassium chloride
p-CF	precipitating CF
Pro, P	Proline
PDB	Protein Data Bank
PDBID	Protein Data Bank identification code

PDBTM	Protein Data Bank of Transmembrane Proteins
P. aeruginosa	*Pseudomonas aeruginosa*
RR	response regulator
rpm	revolutions per minute
RNA	ribonucleic acid
RNP	ribonucleoprotein domain
RRM	RNA recognition motif
SSM	secondary-structure matching
SeMet	selenomethionine
Ser, S	Serine
SPA	sinapinic acid
SAD	single anomalous dispersion
SIRAS	single isomorphous replacement with anomalous scattering
sRNA	small ribonucleic acid
Na_2CO_3	sodium carbonate
NaCl	Sodium Chloride
SDS	sodium dodecyl sulfate
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
NaI	sodium iodide
Na_2P0_4	sodium phosphate
SSRL	Stanford Synchrotron Radiation Laboratory
SELDI-TOF	surface enhanced laser desorption/ionization time of flight

TCL-1	T cell line-tropic
Thr, T	Threonine
x g	times gravity
TGF-β	Transforming Growth Factor beta
TM, TMD	transmembrane, transmembrane domain
TROSY	transverse relaxation optimized spectroscopy
Tris	tris (hydroxymethyl) aminomethane
Trp, W	Tryptophan
TCS	Two-component system
Tyr, Y	Tyrosine
Val, V	Valine
v	version
KvPae	voltage gated K^+ Channel-like protein from *P. aeruginosa*

LIST OF SYMBOLS

α	alpha
$Å^3$	Angstrom cubed
$Å^2$	Angstrom squared
Å	Angstrom; 1 x 10^{-10} m
β	beta
°	degree
°C	degree celcius
Δ	delta
γ	gamma
$\rightarrow$	gene activation
$\dashv$	gene repression
V_m	Matthews Coefficient, $Å^3$/Da
μ	micro
$\bar{v}$	partial specific volume
%	percent
Φ	phi
σ	sigma

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost I would like to thank my family and friends for their continued support throughout my education. In particular, I would like to thank my parents Alvin and Lorna, and brother Brent, for always lending an ear and guiding me along this journey. I would also like to thank my boyfriend, Jay, whose support, patience, and reassurance has helped me get to this point.

I thank my mentor Senyon Choe, for taking me into his lab, for believing in my abilities to achieve higher goals, and encouraging me to settle for nothing but success. He has created an environment where a scientist is free to think outside the box and is provided with the opportunities to grow. It is always refreshing to see his enthusiasm and interest in my projects and I will never forget our bets on scientific theories/experimental outcomes over the wager of ice cream. I would also like to thank Witek Kwiatowski, who has been there for me since day one. He has taught me the tools to understand and ask questions, has always made time for discussions and encourages me to not only challenge myself, but also believe in myself. I could probably go on for pages but in short his unwavering support has helped me in an immeasurable way.

I also thank all of the past and current Structural Biology Laboratory members who have supported and helped me throughout the years. In particular, Innokentiy Maslennikov and Christian Klammt, who have both taught me and guided me along the way, they have incredible patience and took the time to teach me NMR. Last but not least I would like to thank Luis Esquivies, for having to sit next to me throughout the past few years and listen to me jabber on about my experiments, thanks for lending me your ears.

(Oh yea, and educating me about the Spanish language, Indie music and trivia, a graduate education wouldn't be complete without that.) All in all, I could not have asked for a better place to learn.

Moreover, I would like to thank The American Heart Association and the H. A. Mary K. Chapman Charitable Trust and The Mary K. Chapman Foundation whom have both generously provided support for me during my graduate career. I would also like to thank the support and help from the members of the Center for Structures of Membrane Proteins (CSMP). The knowledge, encouragement, and positive environment from this center have made learning more about membrane proteins fun and enjoyable.

Chapter 1 and Chapter 2, in part, is a reprint of the material as it appears in Biochemistry 2007. "Blain, K. Y., Kwiatkowski, W., Zhao, Q., La Fleur, D., Naik, C., Chun, T.W., Tsareva, T., Kanakaraj, P., Laird, M. W., Shah, R., George, L., Sanyal, I., Moore, P. A., Demeler, B., and Choe, S. Structural and functional characterization of CC Chemokine CCL14. *Biochemistry*. 2007; 46(35): 10008-15." The dissertation author was the primary investigator and author of this paper.

Chapter 1 and Chapter 3, in part, is a reprint of the material as it appears in *Protein & Peptide Letters* 2010: "Xie, W., Blain, K. Y., Kuo, M. M., Choe, S. "Protein engineering of bacterial histidine kinase receptor systems." *Protein Pept Lett.* 2010; 17 (7): 867-873" The dissertation author was one of the primary investigators and authors of this paper.

Chapter 1 and Chapter 4, in part, is a reprint of the material as it appears in *Biochemistry* 2010. "Blain, K.Y., Kwiatkowski, W., Choe, S. "The functionally active Mistic-fused histidine kinase receptor, EnvZ." *Biochemistry*. 2010; 49 (42): 9089-9095. The dissertation author was the primary investigator and author of this paper.

Chapter 1 and Chapter 5, in part, will be prepared for submission for publication of the material and may appear as: "Klammt, C., Maslennikov, I., Blain, K.Y., and Choe, S. Solution NMR structure of human integral membrane proteins. 2010. *In preparation.*" The dissertation author was one of the primary investigators of this work.

Chapter 1 and Chapter 5, in part, has in part been submitted for publication: "Farina, A. N., Blain, K. Y., Maruo, T., Kwiatkowski, W., Choe, S., and Nakagawa, T. 2010 Separation of domain contacts is required for heterotetrameric assembly of functional NMDA receptors. 2010. *Submitted.*" The dissertation author was one of the primary investigators and authors of this paper.

Chapter 1 and Chapter 5, in part, is currently being prepared for submission for publication of the material and may appear as: "Ou, H., Blain, K. Y., Kwiatkowski, W., Choe, S., and O'Shea, C. Structural Assembly of E4-ORF3. 2010. *In preparation.*" The dissertation author was one of the primary investigators of this work.

VITA

EDUCATION

2004 Bachelor of Science, Molecular Biology and Biochemistry, University of California, Irvine

2010 Doctor of Philosophy, Biology, University of California, San Diego

RESEARCH FELLOWSHIPS, AWARDS AND EXPERIENCES

2005-2010 Member of the Center for Structures of Membrane Proteins.

2005-2010 H. A. Mary K. Chapman Charitable Trust and The Mary K. Chapman Foundation

2006-2008 American Heart Association Predoctoral Fellowship

2006 Experimenter. *RapiData: Data Collection and Structure Solving at the NSLS, A Practical Course in Macromolecular X-ray Diffraction Measurement.* National Synchrotron Light Source, Brookhaven National Laboratory. Upton, New York.

2008 Student. *CCP4 school: from data processing to structure refinement and beyond.* Advanced Photon Source, Argonne National Laboratory. Argonne, Illinois.

2008 TA Excellence in Teaching Award, University of California, San Diego, Division of Biological Sciences

2008 Annual Meeting Presentation. Center for Structures of Membrane Proteins. University of California, San Francisco.

2009 Annual Meeting Presentation. Center for Structures of Membrane Proteins. The Salk Institute for Biological Studies.

PUBLICATIONS

1) **Blain KY**, Kwiatkowski W, Zhao Q, La Fleur D, Naik C, Chun TW, Tsareva T, Kanakaraj P, Laird MW, Shah R, George L, Sanyal I, Moore PA, Demeler B, Choe S. "Structural and functional characterization of CC chemokine CCL14." *Biochemistry.* 2007; 46(35):10008-15.

2) Maslennikov I, Krupa M, Dickson C, Esquivies L, **Blain K**, Kefala G, Choe S, Kwiatkowski W. "Characterization of protein detergent complexes by NMR, light scattering, and analytical ultracentrifugation." *J Struct Funct Genomics.* 2009; 10(1):25-35.

3) Xie W, **Blain KY**, Kuo MM, Choe S. "Protein engineering of bacterial histidine kinase receptor systems." *Protein Pept Lett.* 2010; 17 (7): 867-873

4) **Blain KY**, Kwiatkowski W, Choe S. "The functionally active Mistic-fused histidine kinase receptor, Envz." *Biochemistry*. 2010; 49 (42): 9089-9095.

5) Farina AN, **Blain KY**, Maruo T, Kwiatkowski W, Choe S, Nakagawa T. "Separation of domain contacts is required for heterotetrameric assembly of functional NMDA receptors." *Submitted.* 2010.

ABSTRACT OF THE DISSERTATION

The Molecular Mechanisms Behind Receptor Signaling

by

Katherine Yoshie Blain

Doctor of Philosophy in Biology

University of California, San Diego, 2010

Professor Senyon Choe, Chair
Professor Tony Hunter, Co-Chair

Cell membranes are a crucial component to the life of a cell. The membrane defines boundaries, provides structural elements, and contains proteins that serve as sensor receptors to transmit external cues across the phospholipid bilayer, either from the environment to the cell's interior or from the cytosol to a particular sub-cellular compartment. One type of proteins found spanning these lipid bilayers are known as the

integral membrane proteins (IMPs). Since the dysfunction of this class of proteins has been associated with multiple disease states, a large initiative has developed to target them for new pharmaceutical therapeutics.

Unfortunately, many obstacles have slowed the progression of IMP research. Some of these deterrents include low IMP over-expression levels, detergent selection, and detergent extraction. The drawbacks to studying IMPs can account for the small percentage of IMP structures that have been deposited in the PDB.

Here I describe the structural and functional studies of multiple receptors from different classes. The first study includes analysis of the chemokine CCL14 and the G protein-coupled receptor CCR5 and provides new information for the potential to use chemokines in CCR5 receptor targeted HIV-1 therapeutics. Other studies presented here include the structure determination of hIMP3, the NH_2-terminal domain of a glutamate receptor, and an adenoviral protein and can aid in the development of drugs related to human neurological diseases and cancer.

To speed the progression of IMP studies, there is a current need for the improvement of existing methods and for the creation of new ones. Previous studies have been compiled here to summarize the creation of histidine kinase receptor chimeras and the benefits of engineering new receptor signal transduction circuits. Moreover, I discuss the Mistic-fusion system for use in boosting IMP expression levels while maintaining functionality of the fused receptor EnvZ. Lastly, I illustrate that by combining cell-free expression, the CDL-strategy, and the use of fast NMR analysis, the structure of hIMP3 was solved within a few months. By using new techniques and determining the structures

of IMPs and their signaling partners, we can gain a better understanding of the molecular mechanisms behind receptor signaling.

CHAPTER 1:

Introduction

1.1 Introduction

Constituting about 30% of both prokaryotic and eukaryotic proteomes, are the integral membrane proteins. These proteins play a vital role in mediating a cell's interaction with its surrounding environment by first sensing extracellular signals then transmitting them into the cell to create a specific response. Alterations to their natural function can be correlated to multiple disease states and thus a large interest has developed in targeting these proteins for pharmaceutical therapeutics (*1-6*). Unfortunately, many bottlenecks hamper the progression of research in this field, such as integral membrane protein over-expression, detergent extraction, and detergent selection for structural and functional analysis. Such obstacles have accounted for the fact that only 261 unique membrane protein structures have been determined (October 2010,

http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html) as per the Stephen White laboratory, while about 1288 structures of transmembrane proteins and their fragments were deposited to the PDB according to PDBTM (October 2010, http://pdbtm.enzim.hu/.) (*7, 8*), and 1066 structures solved according to Membrane Protein Data Bank (MPDB, http://www.mpdb.tcd.ie/) (*9*).

Due to this demand for a better understanding of integral membrane proteins, structural and functional studies were completed and presented here on multiple receptors from different classes. One of the first systems I studied, covered in Chapter 2, was that of the human G protein-coupled receptor CCR5 and the CC Chemokine CCL14. Determining the structure and analysis of the function of CCL14, has yielded insights in oligomerization and has provided a basis for the use of this chemokine in CCR5 targeted HIV-1 therapeutics. Moreover, the information provided from the NMR and crystallographic membrane protein structures presented here of hIMP3 and GluN1 NTD contain structural insights valuable to the scientific community for the design and creation of new drugs used in the treatment of many widespread human diseases. I also illustrate that protein crystallography can be a powerful tool in regards to other types of proteins such as the adenoviral protein E4-ORF3. The crystal structure has provided us with a means to visually recognize 3D structural motifs that cannot be determined from the sequence alone, yielding insight into the mode of higher oligomerization and function of this protein. Collectively, these results have provided us with a look at atomic resolution into these different protein systems, offering us a new perspective into the molecular mechanisms behind receptor signaling and protein function.

Since there are many obstacles that hinder the study of membrane proteins there is a current need for the improvement of existing methods and the creation of new methods to aid in membrane protein structure determination and in the study of membrane protein function. In Chapter 3, I compiled some of the current studies that have been completed on protein engineering on histidine kinase receptors to create new signal transduction systems within cells. These studies cover the methodology of creating chimeras to investigate the signal transduction circuits of two-component systems and explore the possibilities of creating new unique synthetic systems that can be utilized for medical and industrial use.

In addition to the current methods used to study protein function and engineer new pathways, we have employed new techniques to improve the efficiency of structure determination of membrane proteins. One of these techniques is the use of the *Bacillus subtilis* protein, Mistic, as a fusion partner of membrane proteins during expression. Although previous studies have shown that Mistic's ability to boost eukaryotic and prokaryotic membrane protein expression levels (*10, 11*), the functional state of the fused membrane protein remained unknown. In Chapter 4 we address this question by utilizing the prokaryotic two-component system involving the histidine kinase receptor EnvZ and its cognate response regulator OmpR. This study illustrates that in the case of EnvZ, the Mistic-fusion system can be used to increase expression levels of a receptor and to maintain functionality both *in vivo* and *in vitro* of the cargo protein, providing an alternate means to address the problem of low expression yields. Another newly developed technique that I have utilized is coupling cell-free expression, the CDL-

strategy, and the use of paramagnetic probes for acquiring long distance information, to determine the structure of a human integral membrane protein. Utilizing this new technique I show in Chapter 5 that the hIMP3 solution structure was determined within just a few months time. Thus illustrating the benefits of exploring new methods to improve the efficiency and speed of membrane protein structure determination. Lastly, in order to improve the use of current techniques in structural biology it is important to address problems that are encountered during structure determination. In Chapter 5, I address the obstacles that were faced during our ineffective attempts at solving the crystal structure of GluN1 NTD using molecular replacement, and provide possible answers as to why this method yielded unsuccessful results. By utilizing new techniques and solving the structures of membrane proteins and their signaling components, we can gain a better mechanistic understanding of signaling systems that are not well understood and thus can be a very important contribution to structure based-drug design.

1.2 The G-protein coupled receptor CCR5 and the CC Chemokine CCL14

Chemokines are small secreted proteins that activate seven transmembrane G protein-coupled receptors (GPCR) in order to play an important role in the immune and inflammatory response systems (*12, 13*). These small 8-10 kDa proteins have been associated with immunological disease states such as asthma (*14*), arteriosclerosis (*15*), and rheumatoid arthritis (*16, 17*). The role of chemokines have also been associated with Acquired Immune Deficiency Syndrome (AIDS), since both chemokines and Human Immunodeficiency Virus Type -1 (HIV-1) utilize the same GPCR CC Chemokine

Receptor 5 (CCR5) for entry into cells (*18-23*). Certain chemokines such as MIP-1α (CCL3), MIP-1β (CCL4), MCP-2 (CCL8), and RANTES (CCL5) have been previously shown to inhibit HIV-1 entry into cells (*24-26*). Thus illustrating the potential for targeting CCR5 with chemokines for HIV-1 therapeutics.

CC Chemokine Ligand 14, CCL14, is a human CC chemokine that is of recent interest because of its natural ability, upon proteolytic processing of the first eight NH_2-terminal residues, to bind to and signal through HIV-1 co-receptor CCR5. We report X-ray crystallographic structures of both full-length CCL14 and signaling-active, truncated CCL14 [9-74] determined at 2.23 Å and 1.8 Å, respectively. Although CCL14 and CCL14 [9-74] differ in their ability to bind CCR5 for biological signaling, we find that the NH_2-terminal eight amino acids [residues 1 through 8] are completely disordered in CCL14 and both show an identical mode of dimeric assembly characteristic of the CC type chemokine structures. However, analytical ultracentrifugation studies reveal that the CCL14 is stable as a dimer at a concentration as low as 100 nM, whereas CCL14 [9-74] is fully monomeric at the same concentration. By the same method, the equilibrium between monomers of CCL14 [9-74] and higher-order oligomers is estimated to be of $EC_{1,4}$ = 4.98 μM for monomer-tetramer conversion. The relative instability of CCL14 [9-74] oligomers as compared to CCL14 is also reflected in the K_d's that are estimated by the surface plasmon resonance method to be ~9.84 nM and 667 nM for CCL14 and CCL14 [9-74], respectively. This ~60-fold difference in stability at a physiologically relevant concentration can potentially account for their different signaling ability. Functional data from the activity assays by intracellular calcium flux and inhibition of CCR5-mediated HIV-1 entry show that only CCL14 [9-74] is fully active at these near-

physiological concentrations where CCL14 [9-74] is monomeric and CCL14 is dimeric. These results together suggest that the ability of CCL14 [9-74] to monomerize can play a role for cellular activation.

1.3 Protein engineering on histidine kinase two-component systems and the use of chimeras to study new receptor signaling mechanisms

Two-component systems involving the His-Asp phosphotransfer are commonly utilized for signal transduction in prokaryotes in which the two essential components are a sensor histidine kinase receptor and a response regulator. The histidine kinase receptor usually receives a signal and autophosphorylates on a conserved histidine residue and subsequently transfers this phosphoryl group to a conserve aspartate on the response regulator. The response regulator generally acts as a transcription factor, where upon phosphorylation turns on or off the transcription of genes.

Histidine kinase receptors contain a conserved topology consisting of two transmembrane domains separated by an extracellular sensor domain, and a COOH-terminal cytoplasmic domain. The cytoplasmic domain can further be divided into two domains: (A) an NH_2-terminal dimerization and histidine phosphotransfer (DHp) domain and (B) a COOH-terminal catalytic and ATP-binding (CA) domain. Typically, a conserved α-helical linker, HAMP (Histidine kinase, Adenyl cyclase, Methyl-accepting protein and Phosphatase) domain, is found connecting the second transmembrane domain to the cytoplasmic domain.

Despite great efforts in structural and functional characterization of signal perception mechanisms, the exact signaling mechanisms remain elusive for many two-component systems. Mimicking the natural two-component system signaling pathways, chimeric receptor kinases and response regulators have been constructed through the process of swapping modular domains of related two-component systems. To design chimeras with new signaling pathways, domains from different proteins that have little relationship at the primary structural level but carrying desirable functional properties can be conjoined to engineer novel two-component systems. These chimeras maintain the ability to respond to environmental stimulants by regulating protein phosphorylation to produce downstream output signals. Depending on the nature of external signals, chimeric two-component systems can serve as a novel tool not only to examine the natural signaling mechanisms in two-component systems, but also for industrial and clinical applications.

1.4 Utilizing the benefits of Mistic while maintaining functionality of the histidine kinase receptor EnvZ

Integral membrane proteins are required for major cellular functions and account for about 30 % of all proteins in both prokaryotic and eukaryotic organisms and are thus important targets for therapeutics (*1-4*). However, one major obstacle to overcome when working with this family of proteins is their tendency to express at low levels. To overcome this problem alternate means of expression are often utilized. One such method that has been shown to boost both prokaryotic and eukaryotic membrane protein

expression levels is the use of Mistic-fusion system (*10, 11*). Mistic is a small *Bacillus subtilis* protein that is of current interest to the field of structural biology and biochemistry because of its unique ability to increase integral membrane protein yields in *Escherichia coli* expression. Although Mistic's ability to increase membrane protein yields has been shown, its ability to preserve functionality of the fused membrane protein remained unknown. I chose to address this question by studying the prokaryotic two-component system, EnvZ-OmpR.

Using the osmo-sensing histidine kinase receptor, EnvZ, an *E. coli* two-component system, and its cytoplasmic cognate response regulator, OmpR, we provide the first evidence that a Mistic-fused integral membrane protein maintains functionality both *in vitro* and *in vivo*. When the purified and detergent-solubilized receptor EnvZ is fused to Mistic, it maintains the ability to autophosphorylate on residue His_{243} and phosphotransfers to residue Asp_{55} located on OmpR. Functionality was also observed *in vivo* by means of a β-galactosidase assay in which RU1012 [Φ(*ompC-lacZ*)*10-15*, Δ*envZ::*Km^r] cells transformed with Mistic-fused EnvZ led to an increase in downstream signal transduction events detected by the activation of *ompC* gene expression. These findings illustrate that Mistic preserves the functionality of the Mistic-fused cargo protein and thus provides a beneficial alternate approach to study integral membrane proteins by not only improving expression levels but also for direct use in functional characterization.

1.5 Rapid determination of the solution NMR structure of hIMP3

The histidine kinase receptor EnvZ was my initial target in trying to solve the solution structure by NMR. I first tried to truncate it into fragments that would be more

likely to contain uniform structural properties needed to acquire good quality NMR spectra. After locating certain domains of EnvZ I decided to create constructs that included the first two transmembrane domains and the periplasmic sensor domain. After optimizing parameters such as detergent type and concentration, salt content, temperature and pH, EnvZ 1-190 appeared to yield the best quality spectra; however, it was not sufficient enough to begin assignment. The EnvZ (1-190) spectra contained two different types of signals, sharp and round (most likely the periplasmic domain) and broad and unresolved (most likely the transmembrane region). This inconsistency in signal-type confirms non-uniform dynamics behavior of the periplasmic domain and the transmembrane regions of EnvZ (1-190), such that further improvement of the parameters will need to be tested to improve the quality of the spectra.

With the creation of a newly developed technique that combines cell-free expression and rapid NMR analysis (*27*) to determine the structures of transmembrane proteins, we decided to solve the structure of a two transmembrane human integral membrane protein, which due to our preparation for publication, I will refer to as human integral membrane protein 3 (hIMP3). This newly developed technique overcomes many of the difficulties found in NMR membrane protein structure determination by utilizing a cell-free combinatorial dual-labeling (CDL) strategy. Within a day after 6 hIMP3 CDL samples were expressed using the p-CF system and [^{1}H-^{15}N]-TROSY-HSQC and ^{1}H-^{15}N plane of HNCO spectrum were collected, 26 % of the protein backbone assignment was completed for unique amino acid pairs in the sequence. By analyzing 3D-HNCA, -HNCOCA and -HNCACB spectra using the initial unambiguous assignments as anchor

points for further sequential assignment, we currently have 71 % of the protein assigned. Within just a few months the paramagnetic relaxation enhancement (PRE) effect (*28, 29*) was utilized to determine distances and determine the structure of hIMP3. The speed and efficiency of the CDL strategy combined with cell free expression and fast NMR analysis provides a new technique that can be used to quickly determine structures of integral membrane proteins. Here I have shown that a human integral membrane protein can be solved in a short period of time and thus this technique can be an important tool that will provide new structural insights in receptor signaling and therapeutic design and development.

1.6 Crystallographic structure and analysis of the NMDA receptor GluN1 NTD

N-methyl-D-aspartate (NMDA) receptors are one of the three subfamilies that comprise the ionotropic glutamate receptors (iGluRs). These receptors are found in the mammalian brain and are involved in the development and function of the brain (*30, 31*). Impaired function of the NMDA receptors has been related to many neurological and psychiatric disease states, such as epilepsy, neuropathic pain, schizophrenia, seizure, stroke, Alzheimer's and Parkinson's disease and are thus important therapeutic targets (*32, 33*). Therefore it is of current interest to solve the structure of these receptors for a better mechanistic understanding behind NMDA receptor signaling, which could lead to insights into structure based therapeutics.

In addition to the benefits obtained in the medical community by determining a new protein structure through the method of X-ray crystallography, there are also benefits to the scientific structural biology community, in gaining a better understanding of the

problems encountered during the steps taken to determining the structure. Here we address a problem during molecular replacement and find ways to explain the cause and identify the commonality of this occurrence in previous glutamate receptors solved and in the structures deposited in the Protein Data Bank.

We have determined the X-ray crystallographic structure of the NMDA receptor GluN1 NH_2-terminal domain (NTD) to 3.4 Å resolution. Initial attempts to first solve the phases using molecular replacement yielded unsuccessful results and led us to solve the structure using a combination of SIRAS and SAD with iodine signal. The final solution contains 3 molecules in the asymmetric unit which are arranged in the primitive hexagonal crystal system, space group $P3_121$ with unit cell dimensions $a = b = 164.7$, $c = 147.3$, $\alpha = \beta = 90.0$, $\gamma = 120.0$, and a high solvent content of 71 %. The monomeric GluN1 NTD structure contains a clamshell-like architecture that is similar to the structures of other glutamate receptor domains and is further divided into two sub-domains, R1 and R2 that are connected by three loops. Four GluN1 NTD monomers create two unique dimers that together form a tetrameric arrangement, the biological relevance of this structure will still need to be investigated.

After solving the structure we found that the solution contains 3 molecules in the asymmetric unit, where molecular replacement solutions predicted 4-5 molecules, and could account for the failed attempts at molecular replacement. Molecular replacement programs utilize the Matthews coefficient to predict the number of molecules in the asymmetric unit based on solvent content, unit cell dimensions and the molecular weight of the molecules within that unit cell. After completing a search of the Protein Data Bank

(http://www.rcsb.org/pdb/) for high solvent structures determined, and finding other glutamate NTD structures solved with high solvent content (*34, 35*), we determined that this was not such an uncommon occurrence. Moreover, another contributing factor to the unsuccessful endeavor at molecular replacement could have been due to the differences between the search models that we chose and GluN1 NTD. Although the overall clam-shell like motif appears similar to other glutamate receptor search models that we used, when overlaid, the major differences are visualized in not only the loop regions, but also in the Cα backbone trace.

Lastly, the number of *N*-linked glycosylation sites was determined through SELDI TOF mass spectrometry. Through this analysis we found that there are a total of 6 glycosylation sites. In the GluN1 NTD sample that was used to set up crystal trays we had a mixture of one population of GluN1 NTD which contained 5 sites of glycosylation where all 5 sites were de-glycosylated after treatment with endo-H, and another population where we had 6 sites of glycosylation where 5 of the 6 sites were de-glycosylated after treatment with endo-H. This corresponds well to our crystal structure where we see 5 sites of glycosylation at Asn_{61}, Asn_{203}, Asn_{239}, Asn_{276}, and Asn_{300}.

1.7 Utilizing protein crystallography to understand oligomerization in the DNA tumor viral protein, E4-ORF3

The tumor-suppressor p53 is a protein that is constitutively expressed in normal cells, where activation signals like DNA damage can cause stabilization of p53 and a rise

in p53 levels which eventually lead to an inhibition of cell growth (*36, 37*). It has been found that in many types of cancer p53 is inhibited either through point mutations, gene losses or rearrangements (*38-40*). Similarly, tumor mutations and DNA tumor virus proteins deregulate the same targets and mechanisms that subsequently cause abnormal cell proliferation (*41*). Thus viral proteins are of current interest for finding new targets that are deregulated in cancer and for designing new viral cancer therapies (*42*).

One such viral protein is E4 open reading frame 3 (E4-ORF3), which is one of many proteins encoded by adenovirus early region 4 and is responsible for regulating the viral lytic program along with other cellular processes (*43, 44*). Current studies performed by the O'Shea laboratory have shown that E4-ORF3 forms a novel and dominant epigenetic structure that silences p53-transcription (*45*). They have shown that E4-ORF3 assembles into a high-order oligomer creating a nuclear scaffold that prevents p53 from binding to DNA at p53 target promoters through the formation of H3K9me3 heterochromatin formation at p53 target promoters.

I have solved the crystallographic structure of E4-ORF3 to 2.06 Å resolution in four different space groups and analyzed its architecture to gain a better understanding of function from a structural perspective. Comparison of E4-ORF3 to other 3 dimensional structures of proteins contained in the Protein Data Bank (http://www.rcsb.org/pdb/) using the program DALI (*46*), revealed that E4-ORF3 contains similar conserved features found in the RNA recognition motif (RRM). Although sequence similarity was low, this search revealed a similar RRM architecture to that of IS*Hp608* TnpA transposase. Through analysis of the crystal structure I propose a possible mode of fiber formation

where the dimer composes the basic building block. I also suggest mutations based on the structure that can be created in future studies to disrupt and recover the hydrophobic core of E4-ORF3. Collectively these results will shed light on the mystery behind E4-ORF3's ability to form a high order nuclear scaffold that prevents p53 from binding to target promoters.

1.8 Acknowledgements:

Chapter 1, in part, is a reprint of the material as it appears in *Biochemistry* 2007. "Blain, K. Y., Kwiatkowski, W., Zhao, Q., La Fleur, D., Naik, C., Chun, T.W., Tsareva, T., Kanakaraj, P., Laird, M. W., Shah, R., George, L., Sanyal, I., Moore, P. A., Demeler, B., and Choe, S. Structural and functional characterization of CC Chemokine CCL14. *Biochemistry*. 2007 46 (35): 10008-15." The dissertation author was the primary investigator and author of this paper.

Chapter 1, in part, is a reprint of the material as it appears in *Protein & Peptide Letters* 2010: "Xie, W., Blain, K. Y., Kuo, M. M., Choe, S. "Protein engineering of bacterial histidine kinase receptor systems." *Protein Pept Lett.* 2010; 17 (7): 867-873" The dissertation author was one of the primary investigators and authors of this paper.

Chapter 1, in part, is a reprint of the material as it appears in *Biochemistry* 2010. "Blain, K. Y., Kwiatkowski, W., Choe, S. "The functionally active Mistic-fused histidine kinase receptor, Envz." *Biochemistry*. 2010; 49 (42): 9089-9095. The dissertation author was the primary investigator and author of this paper.

Chapter 1, in part, has been submitted for publication of the material: "Farina, A. N., Blain, K. Y., Maruo, T., Kwiatkowski, W., Choe, S., and Nakagawa, T. Separation of

domain contacts is required for heterotetrameric assembly of functional NMDA receptors. 2010. *Submitted.*" The dissertation author was one of the primary investigators and authors of this paper.

Chapter 1, in part, will be prepared for submission for publication of the material and may appear as: "Klammt, C., Maslennikov, I., Blain, K.Y., and Choe, S. Solution NMR structure of human integral membrane proteins. 2010. *In preparation.*" and "Ou, H., Blain, K. Y., Kwiatkowski, W., Choe, S., and O'Shea, C. Structural Assembly of E4-ORF3. 2010. *In preparation.*" The dissertation author was one of the primary investigators of this work

1.9 References

1. Arkin, I. T., Brunger, A. T., and Engelman, D. M. (1997) Are there dominant membrane protein families with a given number of helices?, *Proteins 28*, 465-466.

2. Wallin, E., and von Heijne, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms, *Protein Sci 7*, 1029-1038.

3. White, S. H., and Wimley, W. C. (1999) Membrane protein folding and stability: physical principles, *Annu Rev Biophys Biomol Struct 28*, 319-365.

4. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, R. D., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J. F., Dougherty, B. A., Bott, K. F., Hu, P. C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison, C. A., 3rd, and Venter, J. C. (1995) The minimal gene complement of Mycoplasma genitalium, *Science 270*, 397-403.

5. Frishman, D., and Mewes, H. W. (1997) Protein structural classes in five complete genomes, *Nat Struct Biol 4*, 626-628.

6. Kihara, D., and Kanehisa, M. (2000) Tandem clusters of membrane proteins in complete genome sequences, *Genome Res 10*, 731-743.

7. Tusnady, G. E., Dosztanyi, Z., and Simon, I. (2004) Transmembrane proteins in the Protein Data Bank: identification and classification, *Bioinformatics 20*, 2964-2972.

8. Tusnady, G. E., Dosztanyi, Z., and Simon, I. (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank, *Nucleic Acids Res 33*, D275-278.

9. Raman, P., Cherezov, V., and Caffrey, M. (2006) The Membrane Protein Data Bank, *Cell Mol Life Sci 63*, 36-51.

10. Roosild, T. P., Vega, M., Castronovo, S., and Choe, S. (2006) Characterization of the family of Mistic homologues, *BMC Struct Biol 6*, 10.

11. Kefala, G., Kwiatkowski, W., Esquivies, L., Maslennikov, I., and Choe, S. (2007) Application of Mistic to improving the expression and membrane integration of

histidine kinase receptors from Escherichia coli, *J Struct Funct Genomics 8*, 167-172.

12. Murphy, P. M., Baggiolini, M., Charo, I. F., Hebert, C. A., Horuk, R., Matsushima, K., Miller, L. H., Oppenheim, J. J., and Power, C. A. (2000) International union of pharmacology. XXII. Nomenclature for chemokine receptors, *Pharmacol Rev 52*, 145-176.

13. Murphy, P. M. (1996) Chemokine receptors: structure, function and role in microbial pathogenesis, *Cytokine Growth Factor Rev 7*, 47-64.

14. Smit, J. J., and Lukacs, N. W. (2006) A closer look at chemokines and their role in asthmatic responses, *Eur J Pharmacol 533*, 277-288.

15. Nelken, N. A., Coughlin, S. R., Gordon, D., and Wilcox, J. N. (1991) Monocyte chemoattractant protein-1 in human atheromatous plaques, *J Clin Invest 88*, 1121-1127.

16. Koch, A. E., Kunkel, S. L., Harlow, L. A., Mazarakis, D. D., Haines, G. K., Burdick, M. D., Pope, R. M., Walz, A., and Strieter, R. M. (1994) Epithelial neutrophil activating peptide-78: a novel chemotactic cytokine for neutrophils in arthritis, *J Clin Invest 94*, 1012-1018.

17. Robinson, E., Keystone, E. C., Schall, T. J., Gillett, N., and Fish, E. N. (1995) Chemokine expression in rheumatoid arthritis (RA): evidence of RANTES and macrophage inflammatory protein (MIP)-1 beta production by synovial T cells, *Clin Exp Immunol 101*, 398-407.

18. Berger, E. A., Murphy, P. M., and Farber, J. M. (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease, *Annu Rev Immunol 17*, 657-700.

19. Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M., and Berger, E. A. (1996) CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1, *Science 272*, 1955-1958.

20. Doranz, B. J., Rucker, J., Yi, Y., Smyth, R. J., Samson, M., Peiper, S. C., Parmentier, M., Collman, R. G., and Doms, R. W. (1996) A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors, *Cell 85*, 1149-1158.

21. Dragic, T., Litwin, V., Allaway, 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) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5, *Nature 381*, 667-673.

22. Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P. D., Wu, L., Mackay, C. R., LaRosa, G., Newman, W., Gerard, N., Gerard, C., and Sodroski, J. (1996) The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates, *Cell 85*, 1135-1148.

23. Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R., and Landau, N. R. (1996) Identification of a major co-receptor for primary isolates of HIV-1, *Nature 381*, 661-666.

24. Gong, W., Howard, O. M., Turpin, J. A., Grimm, M. C., Ueda, H., Gray, P. W., Raport, C. J., Oppenheim, J. J., and Wang, J. M. (1998) Monocyte chemotactic protein-2 activates CCR5 and blocks CD4/CCR5-mediated HIV-1 entry/replication, *J Biol Chem 273*, 4289-4292.

25. Cocchi, F., DeVico, A. L., Garzino-Demo, A., Arya, S. K., Gallo, R. C., and Lusso, P. (1995) Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells, *Science 270*, 1811-1815.

26. Greco, G., Mackewicz, C., and Levy, J. A. (1999) Sensitivity of human immunodeficiency virus infection to various alpha, beta and gamma chemokines, *J Gen Virol 80 (Pt 9)*, 2369-2373.

27. Maslennikov, I., Klammt, C., Hwang, E., Kefala, G., Okamura, M., Esquivies, L., Mors, K., Glaubitz, C., Kwiatkowski, W., Jeon, Y. H., and Choe, S. (2010) Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis, *Proc Natl Acad Sci U S A 107*, 10902-10907.

28. Battiste, J. L., and Wagner, G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data, *Biochemistry 39*, 5355-5365.

29. Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R., and Choe, S. (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression, *Science 307*, 1317-1321.

30. Cowan, W. M., Sudhof, T. C., and Stevens, C. F. (2001) *Synapses*, The Johns Hopkins Univ. Press, Baltimore, MD.

31. Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (1995) *Essentials of Neural Science and Behavior*, Appelton & Lange, East Norwalk, CT.

32. Cull-Candy, S., Brickley, S., and Farrant, M. (2001) NMDA receptor subunits: diversity, development and disease, *Curr Opin Neurobiol 11*, 327-335.

33. Kemp, J. A., and McKernan, R. M. (2002) NMDA receptor pathways as drug targets, *Nat Neurosci 5 Suppl*, 1039-1042.

34. Karakas, E., Simorowski, N., and Furukawa, H. (2009) Structure of the zinc-bound amino-terminal domain of the NMDA receptor NR2B subunit, *EMBO J 28*, 3910-3920.

35. Kumar, J., Schuck, P., Jin, R., and Mayer, M. L. (2009) The N-terminal domain of GluR6-subtype glutamate receptor ion channels, *Nat Struct Mol Biol 16*, 631-638.

36. Kubbutat, M. H., Jones, S. N., and Vousden, K. H. (1997) Regulation of p53 stability by Mdm2, *Nature 387*, 299-303.

37. Bates, S., and Vousden, K. H. (1996) p53 in signaling checkpoint arrest or apoptosis, *Curr Opin Genet Dev 6*, 12-18.

38. Hollstein, M., Sidransky, D., Vogelstein, B., and Harris, C. C. (1991) p53 mutations in human cancers, *Science 253*, 49-53.

39. Levine, A. J. (1990) The p53 protein and its interactions with the oncogene products of the small DNA tumor viruses, *Virology 177*, 419-426.

40. Michalovitz, D., Halevy, O., and Oren, M. (1991) p53 mutations: gains or losses?, *J Cell Biochem 45*, 22-29.

41. Levine, A. J. (2009) The common mechanisms of transformation by the small DNA tumor viruses: The inactivation of tumor suppressor gene products: p53, *Virology 384*, 285-293.

42. O'Shea, C. C. (2005) DNA tumor viruses -- the spies who lyse us, *Curr Opin Genet Dev 15*, 18-26.

43. Bridge, E., and Ketner, G. (1989) Redundant control of adenovirus late gene expression by early region 4, *J Virol 63*, 631-638.

44. Huang, M. M., and Hearing, P. (1989) Adenovirus early region 4 encodes two gene products with redundant effects in lytic infection, *J Virol 63*, 2605-2615.

45. Soria, C., Estermann, F. E., Espantman, K. C., and O'Shea, C. C. (2010) Heterochromatin silencing of p53 target genes by a small viral protein, *Nature 466*, 1076-1081.

46. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res 38 Suppl*, W545-549.

CHAPTER 2:

Structural and Functional Characterization of CC Chemokine CCL14

2.1 Introduction

Chemokines (chemotactic cytokines) are small secreted proteins of 8 to 10 kDa that are widely known for their participation in the immune and inflammatory response systems. Their ability to recruit and activate leukocytes to sites of injury or infection is made possible through the activation of specific seven transmembrane G protein-coupled receptors (GPCR) (*1, 2*). Chemokines are generally grouped into the following three subfamilies based on function: 'homeostatic', 'inflammatory', and 'dual function'. Homeostatic chemokines are constitutively produced and secreted, whereas inflammatory chemokines are produced when there is a stimulus or during an infection and they cause the migration of leukocytes to the injured or infected site. The subfamily 'dual function' refers to the chemokines which have dual abilities of the first two subfamilies (*3*). Structurally, these chemokines are classified into the following four categories based on the position of the first two cysteine residues: CC, CXC, C and CX3C (*4, 5*). Currently 28 mammalian CC chemokines have been discovered, with 25 of these found in humans, making the CC chemokines the largest subfamily (*6*). The CXC chemokines Platelet Factor 4 (CXCL4) (*7*) and Interleukin 8 (CXCL8) (*8-10*) were the first chemokines for which a three-dimensional structure was solved. Currently, the structures of several

chemokines have been elucidated through the methods of both NMR and X-ray crystallography. These studies revealed that chemokines share a common structural scaffold consisting of three anti-parallel β-strands and a COOH-terminal α-helix (*11*), but they also showed dissimilarities in the mode of assembly into dimers among different subfamilies. Within the same subfamily, the mode of dimeric assembly is identical.

Due to their role in the immune response, chemokines have been linked to the regulation of many different immunological disorders such as asthma (*12*), arteriosclerosis (*13*), and rheumatoid arthritis (*14, 15*). Pertaining further medically is the role of certain chemokines to Acquired Immune Deficiency Syndrome (AIDS). HIV-1 initiates infection of target cells by fusion of both the viral and target cell membranes. This action is mediated by the binding of the viral envelope glycoprotein (Env) and a CD4 receptor on human target cells. Critical to the efficient infection is the presence of a co-receptor. The importance of the co-receptor provided the explanation underlying HIV-1 tropism by which certain individuals are not readily infected by the virus. In the body two chemokine receptors, CCR5 and CXCR4, are known as the co-receptors for effective HIV entry (*16*). Macrophage-tropic (M-tropic) strains of HIV primarily utilize CCR5 as their co-receptor for entry into cells (*17-21*), where as T cell line-tropic (TCL-tropic) strains are selective for CXCR4 (*22*). The specific chemokines that have been shown to interact with CCR5 are MIP-1α (CCL3), MIP-1β (CCL4), MCP-2 (CCL8), and RANTES (CCL5). All of these chemokines have been shown to inhibit HIV-1 entry into cells (*23-25*). In addition, small-molecule CCR5 antagonists such as vicriviroc (SCH-D, Schering-Plough) and maraviroc (UK-427,857, Pfizer) have been shown to significantly

reduce viral load in HIV-1 patients, demonstrating the therapeutic utility of targeting CCR5 as a means to treat HIV-1/AIDS patients (*26*).

Originally isolated from the hemofiltrate of chronic renal failure patients, plasmatic hemofiltrate CC chemokine 1 or CC Chemokine Ligand 14 (CCL14) is a prototypical CC-type chemokine. CCL14 is found in human plasma at high concentrations (1.5 – 10 nM) and is constitutively expressed in many different tissues (*27*). CCL14 utilizes CCR1, CCR3, and CCR5 for chemotaxis and is involved in the activation of eosinophils, monocytes, and T lymphoblasts. The full-length CCL14 of 74 amino acids is naturally turned into a high-affinity ligand for CCR5 through proteolytic processing by urokinase plasminogen activator and plasmin (*28*). This cleaved form, CCL14 [9-74], has been shown to block HIV entry and replication (*29, 30*), whereas the CCL14 itself has not. Likewise, the CCL14 [9-74] is able to signal by inducing Ca^{2+} flux and migration of different leukocytes (*29*), whereas CCL14 is not, consistent with the difference in their CCR5 binding affinity.

In our study, we report the high-resolution crystal structures of CCL14 and CCL14 [9-74]. Combined with data from analytical ultracentrifuge and BIAcore studies, we demonstrate that CCL14 [9-74] can convert to a monomer more readily than CCL14 at physiological concentrations. *In vitro* assays confirm that CCL14 [9-74], not CCL14, is able to induce calcium flux and to inhibit HIV entry at a concentration where CCL14 [9-74] is a monomer and CCL14 is a dimer.

2.2 Crystal Structures of CCL14 and CCL14 [9-74]

The structures of CCL14 and CCL14 [9-74] have been solved by molecular replacement using the program Phaser (*31-33*), at a resolution of 2.2 Å for CCL14 in space group C2, and 1.8 Å for CCL14 [9-74] in space group $P2_12_12_1$ (Table 2.1). The monomeric subunits of these structures are characterized by the general CC chemokine fold, consisting of a 3-stranded anti-parallel β-sheet in a Greek key motif, connected by loops and followed by a C-terminal α-helix (Figure 2.1A). These structures also contain some common chemokine features such as a NH_2-terminal β-strand (βN_{term}), from Ser^{14} to Cys^{17}, followed by a short $\alpha 3_{10}$ helix, Arg^{27} to Arg^{29}, which directly precedes the first β-strand of the 3-stranded anti-parallel β-sheet. The three β-strands that create the anti-parallel β-sheet are comprised of residues Ile^{30} to Glu^{35} (β1), Ile^{45} to Thr^{49} (β2), and Gly^{52} to Thr^{57} (β3). The C-terminal α-helix is formed by residues Lys^{62} to Met^{71} (Figure 2.1A). Two disulfide bridges are formed between the following four cysteine residues: Cys^{16}-Cys^{40} and Cys^{17}-Cys^{56}. These two disulfide bridges link the NH_2-terminal region with β3 and with the hairpin loop between β1 and β2, stabilizing the monomeric structure. Interactions between two antiparallel β-strands (βN_{term}- βN_{term}) stabilize the dimer without a covalent bond between them (Figure 2.1B). The two homodimers (A-B and C-D of CCL14 or E-H and F-G of CCL14 [9-74]) assume the same intermolecular interaction resulting in the same dimeric configuration. At this high resolution it is very clear that the first nine amino acids on the NH_2-terminus are not present in the electron density of any monomers in the CCL14 structure, indicating high mobility in this region.

2.3 Crystal Structures reveal higher order oligomers for CCL14 and CCL14 [9-74]

Although both CCL14 and CCL14 [9-74] are similar in monomeric and dimeric structures, they differ slightly among those found in the asymmetric unit. CCL14 forms a tetramer in the asymmetric unit and by two-fold symmetry creates an octamer. In all four monomers (A, B, C, D) in CCL14, the first nine residues, Thr^{1} to Gly^{9}, were not present in the electron density maps. Lys^{72} to Asn^{74} in monomer A and Asn^{74} in monomer D are also disordered (Figure 2.2A). CCL14 [9-74] also packs as a tetramer (monomers E, F, G, H) in the asymmetric unit, but in contrast to the octameric quaternary state of CCL14, it does not form a higher order oligomer with any symmetry mates. In monomers F, G, and H of CCL14 [9-74], the COOH-terminal residues Lys^{72} to Asn^{74} are disordered (Figure 2.2B).

To test if the difference in quaternary structure plays a role in the activity of the chemokine, we designed two point mutants: CCL14 E15A and CCL14 [9-74] E15A. The CCL14 [9-74] E15A mutation was designed to destabilize the tetramer assembly by placing a point mutation in the dimer interface at residue Glu^{15} (Figure 2.2), and CCL14 E15A was created as a control. Interestingly, the point mutation did alter the tetrameric arrangement of CCL14 [9-74] to that of CCL14 forming the octameric packing, and as illustrated below, CCL14 E15A remained inactive, but CCL14 [9-74] E15A increased HIV inhibition by 10- and 50-fold (IC50 and IC90) compared to that of wild-type.

2.4 Concentration Dependent Inhibition of HIV-1 Entry and Ca^{2+} Flux Activity

The ability of the CCL14 proteins to inhibit entry of HIV-1 was performed by testing their efficacy to block entry of a CCR5 tropic recombinant HIV-1 virus containing

the luciferase reporter gene into activated CD4 T-cells. While both CCL14 [9-74] and CCL14 [9-74] E15A mediated greater than 95% inhibition of HIV-1 entry, neither CCL14 nor CCL14 E15A demonstrated any dose-dependent inhibition of HIV-1 entry as shown for a representative donor in Figure 2.3A. The IC50 and IC90 concentrations for CCL14 [9-74] were 146 nM and 6.3 μM respectively, while for CCL14 [9-74] E15A they were 14.2 nM and 260 nM. Hence introduction of the E15A mutation reduces the IC50 and IC90 concentrations by 10- and 50-fold.

Ca^{2+} flux assays were completed in parallel to test the ability to signal through CCR5 by inducing an increase in intracellular Ca^{2+} concentrations. Correlating with the data from the HIV-1 infectivity assay, the truncated CCL14 [9-74] and CCL14 [9-74] E15A were able to induce high levels of intracellular Ca^{2+}, demonstrating their ability to bind to CCR5 and signal. In contrast, mobilization of intracellular Ca^{2+} was not detected by CCL14 and CCL14 E15A (Figure 2.3B). We also note that there is an insignificant difference between CCL14 [9-74] and CCL14 [9-74] E15A in the calcium flux assay. Although this contrasts to the CCL14 [9-74] E15A data showing an increased inhibition of HIV-infectivity, as reflected in the ~10-fold reduction of the IC-50 concentration compared to CCL14 [9-74], alteration of the tetrameric assembly of CCL14 [9-74] by the introduction of E15A mutation is an observation potentially useful in developing anti-HIV therapeutics.

2.5 Monomer Dissociation Rates Determined by Analytical Ultracentrifugation

Crystal structures of CCL14 and CCL14 [9-74] dimers do not show conformational difference between the two to account for the difference in functional

binding. Since CCL14 is found in human plasma at a concentration around 10 nM (*27*), sedimentation equilibrium absorbance analytical ultracentrifugation experiments were used to measure the oligomeric state at low concentrations. By measuring absorbance at 220 nm in 10 mM phosphate buffer, CCL14 and CCL14 E15A remain homogeneously as a dimer in solution at a concentration as low as 100 nM, indicating that the K_d for the monomer-dimer equilibrium must be far less than 100 nM. This analysis is insufficient to show if they will form monomeric states (Table 2.2), when diluted to *in vivo* concentration. However, data for CCL14 [9-74] and CCL14 [9-74] E15A were fit very well to a monomer-dimer-tetramer model using the program UltraScan (*34*), and the equilibrium concentration (EC) of the dissociation of tetramers to monomers was found to be in the low micromolar concentrations, with $EC_{1,4}$'s of 4.98 μM and 2.36 μM for CCL14 [9-74] and CCL14 [9-74] E15A, respectively (Table 2.2, Figure 2.4).

Since absorbance-based analytical ultracentrifugation is not suitable to derive the monomer-dimer equilibrium constants smaller than 100 nM for full-length CCL14 (Table 2.2), surface plasmon resonance (BIAcore) analysis was performed as an alternative method to estimate the binding affinities of the chemokines CCL14 and CCL14 [9-74]. We assumed that at very low concentration, the non-covalent CCL14 dimers must dissociate into monomers. In these experiments, each chemokine was immobilized at a low concentration where it was extensively diluted to a monomeric state on the BIAcore chip surface. A blank surface in flow cell 1 was used as a control. At these low concentrations, therefore, protein binding of CCL14 and CCL14 [9-74] in the eluent can simulate dimerization as measured as their ability to bind to their corresponding immobilized partner. The equilibrium dissociation constants, K_d*'s,* for CCL14 and

CCL14 [9-74] are 9.84 nM and 667 nM, respectively, and the K_d*'s* for CCL14 E15A and CCL14 [9-74] E15A are 15.3 nM and 664 nM, respectively. These results show that the dimerization stability of CCL14 is about 60-fold higher than that of CCL14 [9-74], and CCL14 E15A is about 40-fold higher than that of CCL14 [9-74] E15A.

2.6 Discussion

Proteolytic processing of the first eight amino acids activates CCL14 prior to its binding to CCR5 (*29, 30, 35*). X-ray crystallographic analysis shows that CCL14 and CCL14 [9-74] do not have conformational differences in their monomeric fold and dimeric assemblies, thus also revealing similarities to other CC chemokine structures such as CCL5 (*36, 37*), CCL4 (*38*) and CCL2 (*39*). A study on NH_2-terminal processing through the creation of CCL14 variants with lengths ranging from CCL14 [6-74] to CCL14 [12-74] has shown that only CCL14 [9-74] and CCL14 [10-74] block HIV infectivity (*30*). Furthermore, it is well known that certain residues in the NH_2-terminal region of many other CC chemokines are involved in controlling receptor binding and activation (*11, 40-44*). Many have also revealed that NH_2-terminal modifications to chemokines can result in antagonistic, agonistic or super-agonistic properties. A recent study on the NH_2-terminally modified CCL14 analog n-Nananoyl-CC Chemokine Ligand 14, shows the induction of the internalization of CCR3 and desensitization of CCR3-mediated calcium release (*45*). The amino acids directly adjacent to the first eight residues in CCL14 may also be important for identification by the receptor such that the first eight amino acids need to be removed in order for binding and correct recognition to occur. However, in light of our structural analysis it is unclear how the flexibility of the

first eight or nine amino acids may be an important property that affects the ability of CCL14 to bind to CCR5.

Here we suggest that CCL14's ability to monomerize affects its activity. Using analytical ultracentrifugation studies, performed at low concentrations which were closer to those found *in vivo,* the dissociation of dimers to monomers of CCL14 [9-74] happens at near physiological concentration. If we assume that the monomerization of CCL14 is a step in activation, it is unclear how the flexible NH_2-terminal eight residues contribute to the stability of the dimer. The possibility that the monomer is the active form of chemokines was first proposed when the functional unit of a chemically synthesized analog of the CXC chemokine IL-8, was shown to be a monomer (*46*). Since then many studies have been carried out on different CC chemokines including CCL1 (*47*), CCL2 (*40, 47*), CCL3 (*48, 49*), CCL4 (*41, 42*), CCL7 (*50*), and CCL8 (*50*), all suggesting that the monomer is important for binding and activation of chemokine receptors. Some analogues of other CC chemokines obtained by modification to their NH_2-terminus still bind to their receptors but do not signal and thus act as antagonists. Therefore, additional mechanisms of activation caused by NH_2-terminal alterations may also play a role (*51*).

The chemokine activity assays, analytical ultracentrifugation data, and BIAcore data presented here show that inactive CCL14 forms an approximately 60-fold more stable dimer than that of active CCL14 [9-74]. Together, these results suggest how the CCL14 dimer may result in the active CCL14 [9-74] monomer at physiological concentration. Although CCL14 is found circulating in the blood at 10 nM (*27*), it is well known that chemokines form gradients when they home to sites of infection. With this in mind, CCL14 is probably at concentrations that are higher than 10 nM at the time of

activation and are near the concentration where CCL14 is dimeric and CCL14 [9-74] is monomeric and thus may account for their difference in activity.

2.7 Materials and Methods

Crystallization and Data Collection.

Crystals of CCL14 were grown at room temperature by hanging drop vapor diffusion. Each drop contained equal volumes of protein solution and reservoir solution. CCL14 crystals were grown by mixing equal volumes (2 μl each) of protein (20 mg/ml, in 0.2 M NaCl, 10mM MES, pH 6.0) and reservoir solution. CCL14 crystals grew overnight in solution No. 21 from the Hampton PEG/Ion screen containing 0.2 M Sodium Formate, 20% w/v Polyethylene glycol 3,350 (PEG 3350). CCL14 [9-74] crystals were grown by mixing equal volumes (2 μl each) of protein (14 mg/ml, in 0.25 M NaCl, 20 mM MES, pH 6.2) and reservoir solution. CCL14 [9-74] crystals grew after two days in solution No. 18 from the Hampton Screen 1 containing 0.2 M Magnesium acetate tetrahydrate, 0.1 M Sodium cacodylate trihydrate, pH 6.5 and 20% PEG 8000. For X-ray diffraction experiments, crystals were taken directly from the droplets with a fiber loop and flash frozen in liquid N_2 with no cryo-protectant. The diffraction data for CCL14 were collected at beamline 9.1 at the Stanford Synchrotron Radiation Laboratory and the CCL14 [9-74] diffraction data were collected at the Advanced Light Source on beamline 8.2.1. X-ray data sets were integrated and scaled using the program HKL-2000 version 0.98.692i (*52*). The space group of CCL14 was found to be C2 with unit cell dimensions a = 94.35, b = 78.07, c = 59.64, $\alpha = 90$, $\beta = 121.65$, $\gamma = 90°$. The space group

of CCL14 [9-74] was found to be $P2_12_12_1$ with unit cell dimensions a = 56.14, b = 57.55, c = 85.44, $\alpha = \beta = \gamma = 90°$.

Crystal Structure Determination and Refinement.

The crystal structures of CCL14 and CCL14 [9-74] were solved by molecular replacement with the program Phaser (*31-33*) in the CCP4 program suite (*53*). The search model for CCL14 was the AOP RANTES dimer (PDB code: 1B3A) (*54*). A rotation and translation search using the dimer of AOP RANTES identified two dimers in the asymmetric unit, which by two-fold crystallographic symmetry form an octameric packing assembly. The molecular replacement search for CCL14 [9-74] had a solution for two dimers in the asymmetric unit, which did not generate the same octameric assembly of full-length CCL14. The protein chain models were built using the programs ARP/wARP version 6.1 (*55, 56*), O (*57*) and Coot version 0.2 (*58, 59*). The program Refmac5 version (*53, 60*) was used to refine both structures.

The final crystallographic *R* factor and free *R* factor for CCL14 are 18.7% and 23.4%, respectively, for reflections within 56-2.2 Å. Moreover, the final crystallographic *R* factor and free *R* factor for CCL14 [9-74] are 17.4% and 22.8%, respectively, for reflections within 46.9-1.8 Å. The root mean squared deviations in bond lengths and bond angles for CCL14 are 0.021 Å and 1.69 Å, respectively. The root mean squared deviations in bond lengths and bond angles for CCL14 [9-74] are 0.013 Å and 1.24 Å, respectively. The Ramachandran plot calculated for the final CCL14 model using the program PROCHECK (*61*) shows the conformations of 96.9% of residues are located within the most favored regions and 3.1% of residues are located within the additionally

allowed regions. The Ramachandran plot calculated for the final CCL14 [9-74] model shows the conformations of 95.9% of residues are located within the most favored regions and 4.1% of residues are located within the additionally allowed regions. All water molecules have densities of 1 σ or greater in the final $2F_o$-F_c map.

Sedimentation Equilibrium Ultracentrifugation

All protein samples were dialyzed in 10 mM Na_2P0_4, 150 mM NaCl at pH 7.5, Sedimentation equilibrium analytical ultracentrifugation experiments were performed using a Beckman Optima XL-I analytical ultracentrifuge (Beckman, CA) at 20 °C and six-channel centerpieces. Three pairs of blanks and protein samples at three different concentrations were loaded at a volume of 120 ul each and were centrifuged against 130 ul of the equivalent buffer blank. Centrifugation was performed at rotor speeds and durations of 20,000 rpm for 19 h, 27,500 rpm for 18 h, 35,000 rpm for 17 h, 42,500 rpm for 15 h, and 50,000 rpm for 13 h. At each speed the cells were scanned across their radius at 280 nm or 220 nm. To confirm equilibrium, a second scan was taken one hour later and compared to the first scan. Since low concentrations are required to measure an accurate K_d, absorbances were measured at 220 nm for all samples. The protein loading concentrations used were 2.4 μM, 3.2 μM, 6.4 μM for all samples.

Analysis was performed by a global non-linear least squares method using the program UltraScan version 9.1 (*34*). For CCL14 and CCL14 E15A, data was fitted to a one-component ideal model, and for CCL14 [9-74] and CCL14 [9-74] E15A, the data was fit to a monomer-dimer-tetramer equilibrium model. 95% confidence intervals were estimated by Monte Carlo analysis. CCL14 E15A and CCL14 [9-74] E15A are single-

point mutants replacing Glu15 to Ala. The mutation was predicted to disrupt the tetrameric assembly interface where Glu15 is found. The molecular weights that were calculated from the fit indicated the oligomeric state of the protein under the specific solution condition. The density was calculated based on the composition of the buffer used, via the buffer calculation module within UltraScan (*34*). The partial specific volume ($\bar{v}$) was calculated based on its primary amino acid sequence and the temperature (20 °C).

Surface Plasmon Resonance Analysis (BIAcore)

CCL14, CCL14 [9-74], CCL14 E15A and CCL14 [9-74] E15A binding analyses were performed on a BIACore 1000 instrument (BIAcore). Primary amine coupling was used to immobilize CCL14 on a CM4 and CM5 sensor chip (BIAcore) in flow cell 2 and to immobilize CCL14 [9-74] (BIAcore) in flow cell 3. Flow cell 1 was used as a reference surface. The chemokines were immobilized at a concentration of 5 μg/ml, in filtered 10 mM Na acetate buffer at pH 5.0. The proteins were immobilized by pulsing the injection of chemokine over the chip surface at a flow rate of 10μl per min until the desired number of response units (500 RU) was reached. The binding assays were performed in filtered 10 mM HEPES pH 7.5, 150 mM NaCl, 3 mM EDTA, and 0.005% Tween 20 at a flow rate of 30 μl per min for a total of 2 minutes. Seven different concentrations, including a zero concentration, of each chemokine were run. Global fits to binding curves using a 1:1 binding with mass transfer model and separate fits of the association (k_{on}) and dissociation (k_{off}) rate constants were used, both yielding similar

results. The overall K_d was then calculated using the BIAEVALUATION Ver. 4.1 software (Biacore).

Calcium Flux Assay

CCR5 transfected 293 T were detached from T-75 cell culture flask using cell dissociation buffer and after washing two times with HBSS buffer, cells were resuspended at $2.5X10^6$ cells/ml. The cells were then labeled with Fluo-4AM (Molecular Probe # P-3000MP) at a final concentration of 4 μg/ml for 60 minutes at RT in dark. The labeled cells were washed twice in HBSS buffer and resuspended at $2.5X10^6$ cells/ml. 150 μl of labeled cell suspension was transferred to a polyD lysine coated 96-well plate (Biocoat # 359109) and centrifuged at 500 rpm for 1 minute. 50 μl of various concentrations of CCL14 and CCL14 mutants were added and the increase in intracellular calcium was determined using FLIPR.

HIV infectivity Assay

PBMCs were isolate from buffycoats from HIV-seronegative donors. Cells were stimulated with anti-CD3 antibody and IL-2 for 48 h followed by depletion of CD8+ T cells. Then, 150,000 cells were seeded per well in 96-well flat bottom plates. Cells were pre-incubated with CCL14 protein in triplicate at a range of concentrations up to 1600 nM. After 2 h pre-incubation, cells were infected with luciferase HIV pseudotyped with the CCR5 tropic JRFL envelope protein that is capable of a single round of viral entry and infection (*62*). Following 48 h of incubation, cells were lysed and assayed for luciferase activity in a luminometer. % Entry is calculated for each dose tested by the

following formula: (level of luciferase activity observed in the absence of CCL14 protein/level of luciferase observed in presence of CCL14 protein) x 100%. The level of inhibition is plotted against concentration and an IC50 (amount of protein required to block 50% entry) and IC90 (amount of protein required to block 90% entry) calculated.

2.8 Figures

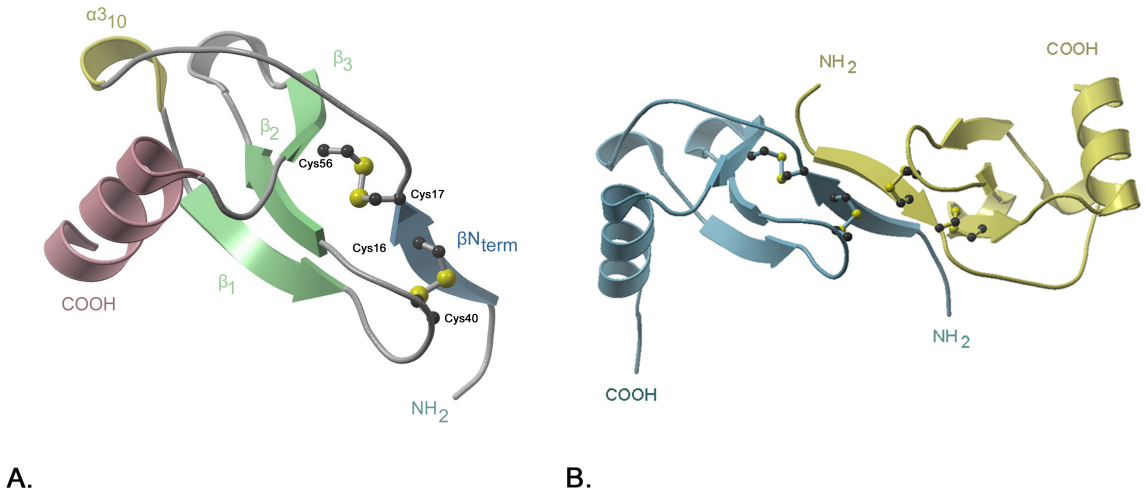

Figure 2.1: Secondary Structure of the CCL14 Monomer and Dimer.
(A) The CCL14 monomer shows similarities to the conserved CC-type chemokine fold characterized by three anti-parallel β-sheets followed by a c-terminal α-helix. Where β-strands are illustrated in blue and green, and helices are shown in yellow and pink, cysteines which form the two disulfide bonds are shown in yellow. (B) The dimer interface is composed of amino acid contacts between the NH_2-terminal strands βN_{term} of each monomer. Two disulfide bonds formed by the four highly conserved cysteine residues, which help stabilize the overall structure, are oriented in a similar fashion between CCL14 and CCL5. As seen here the three stranded anti-parallel β-sheet and COOH-terminal α-helix are the core structural scaffold of the CC chemokine family. Individual monomers are presented in blue and yellow. Figures were prepared and rendered using Molscript Version 2.1.2 (*63*) and POV-Ray (*64*).

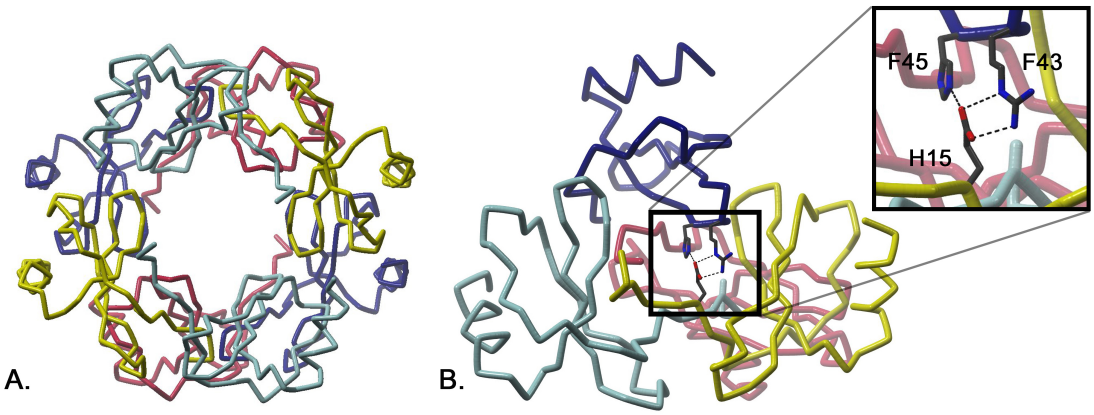

Figure 2.2: CCL14 and CCL14 [9-74] Crystal Structures.
(A) The CCL14 crystal structure is illustrated where each color represents monomers A, B, C, and D in the asymmetric unit. The asymmetric unit consists of a tetramer which is built by two dimers such that dimer 1 is composed of the monomers A-B (red-blue) and dimer 2 is composed of monomers C-D (yellow-teal). By a two-fold symmetry operation, the two tetramers form an octameric assembly. (B) The CCL14 [9-74] crystal structure is illustrated where each color represents monomers E, F, G, and H in the asymmetric unit. This tetrameric assembly differs from that of the CCL14 structure shown in (A), There can be no octamer assembly that can be formed by any symmetry mates of CCL14 [9-74] in the crystal. Two dimers are shown, where dimer 1 is composed of monomers E-H (teal-yellow) and dimer 2 is composed of monomers F-G (blue-red). The figure on the right of the tetramer zooms into the core of the tetrameric assembly where residue Glu15 resides. The numbers in the zoomed-in window represent the monomer (F or H) followed by the residue number. Figures were prepared and rendered using Molscript Version 2.1.2 (*63*), and POV-Ray (*64*).

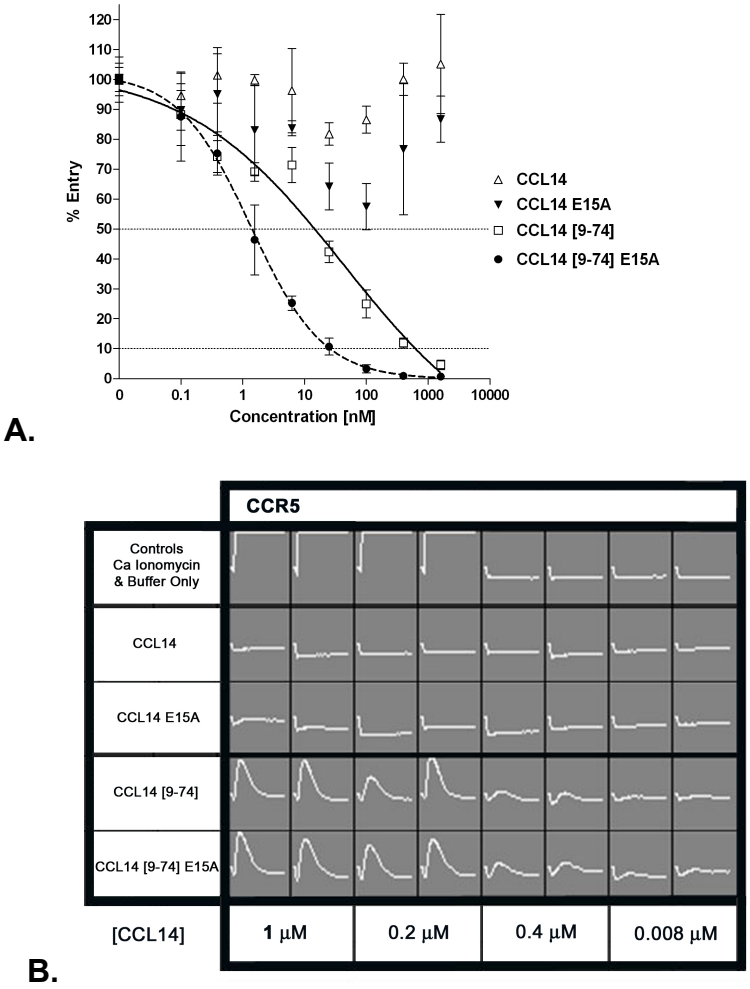

Figure 2.3: HIV Entry Inhibition and Calcium Mobilization by CCL14 proteins.
(A) Inhibition of CCR5-tropic HIV-1 virus infection of activated human PBMCs by CCL14 proteins was determined. CCL14 [9-74] and CCL14 [9-74] E15A blocked greater than 90% of HIV infection while full-length CCL14 proteins had no dose-dependent effect. CCL14 [9-74] demonstrated an IC50 of 146 nM and an IC90 of 6.3 uM while CCL14 E15A [9-74] demonstrates IC50 and IC90 of 14.2 nM and 260 nM. (B) Intracellular Ca^{2+} levels were measured in the presence of CCL14 and mutants at four different chemokine concentrations (bottom row). Ca Ionomycin and buffer only, were used as positive and negative controls, respectively. Ca Ionomycin controls are illustrated in the left four boxes in the first row and the buffer only control is shown in the right four boxes in the first row. Full-length CCL14 and CCL14 E15A do not produce an increase in intracellular Ca^{2+} levels (top two rows under controls). Truncated CCL14 [9-74] and CCL14 [9-74] E15A show an increase in intracellular Ca^{2+} levels at concentrations from 0.2 µM to 1 µM (bottom two rows). Their effective concentrations are approximately similar to each other.

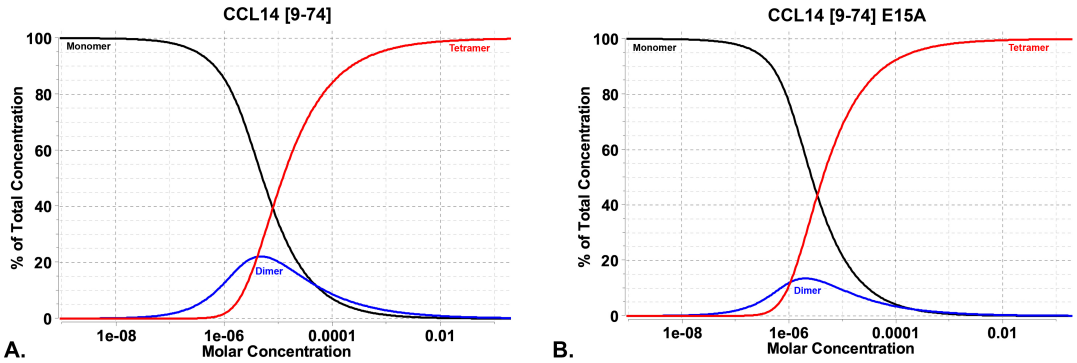

Figure 2.4: Self-Association Profiles for CCL14 [9-74] and CCL14 [9-74] E15A. Self-association curve-fit profiles were generated for (A) CCL14 [9-74] and (B) CCL14 [9-74] E15A from the analytical ultracentrifuge data using the program UltraScan (*34*). See methods for analytical details. The lines represent the different oligomeric states present in the samples as follows: black (monomer), blue (dimer), red (tetramer).

2.9 Tables

Table 2.1: Data collection and refinement statistics.

Data collection and model refinement statistics				
	CCL14		CCL14 [9-74]	
space group	*C*2		*P*212121	
unit cell parameters (Å, deg)	*a* = 94.3		*a* = 56.1	
	b = 78.1		*b* = 57.6	
	c = 59.6		*c* = 85.4	
	β = 121.7		*β* = 90.0	
data collection statistics				
Beamline	9.1 SSRL		8.2.1 ALS	
wavelength (Å)	1.000		1.000	
resolution range (Å)	50-2.23 (2.31-2.23)		50-1.82 (1.89-1.82)	
observed reflections	74369		203161	
unique reflections	17814		25453	
Redundancy	4.2	(4.2)[b]	8.0	(7.3)[b]
R_{merge} (%)	3.9	(10.0)[b]	5.4	(17.0)[b]
Mean I/σ (I)	30.98	(14.3) [b]	35.32	(11.2) [b]
completeness (%)	98.9	(99.2)[b]	99.6	(96.3)[b]
refinement statistics				
R_{cryst} (%)	18.7		17.35	
R_{free} (%)[a]	23.4		22.786	
average *B* factor (Å^2)	37.739		14.096	
root mean square deviation				
bond lengths (Å)	0.021		0.013	
bond angles (Å)	1.685		1.243	
number of atoms total:	2314		2467	
Protein	2130		2102	
Water	184		365	
number of missing residues	40		9	
ramachandran plot (%):				
most favored	96.9		95.9	
additional allowed	3.1		4.1	
generously allowed	0.0		0.0	

[a] calculated from 5% of data not used in refinement. [b] Numbers in parentheses correspond to highest resolution shell (Å).

Table 2.2: Equilibrium sedimentation analytical ultracentrifugation analysis.

Equilibrium Sedimentation Analytical ultracentrifugation analysis of CCL14 and CCL14 [9-74]				
Sample	MW	$K_{d1,2}$	$K_{d1,4}$	best fit model
CCL14	1.628 x 10^4 Da (+85/-67)[a]	----------	----------	ideal one component
CCL14 E15A	1.677 x 10^4 Da (+68/-71)[a]	----------	----------	ideal one component
CCL14 [9-74]	8284 Da (+665/-496)[a]	12.32 μM (+52.34/-9.33)[a]	12.35 x 10^{-17} M^3 (+9.94 x 10^{-17}/-6.56 x 10^{-17})[a] ($EC_{1,4}$= 4.98 μM)[b]	monomer-dimer-tetramer
CCL14 [9-74] E15A	8336 Da (+879/-1036)[a]	10.00 μM (+14.60/-3.90)[a]	1.319 x 10^{-17} M^3 (+1.25 x 10^{-17}/-0.678 x 10^{-17})[a] ($EC_{1,4}$ = 2.36 μM)[b]	monomer-dimer-tetramer

[a]The numbers in parenthesis represent the 95% confidence intervals for the determined parameters.
[b]EC represents the equilibrium concentration.

2.10 Acknowledgements

This work was supported by grants from the National Institutes of Health HD013527 (S.C.) and RRR022200 (B.D.) and the National Science Foundation DBI-9974819 (B.D.). I would like to thank the fellowship support from the American Heart Association and from the H. A. Mary K. Chapman Charitable Trust and The Mary K. Chapman Foundation. I would like acknowledge the following co-authors that helped contribute to the work that was completed: Witek Kwiatkowski, Qinghai Zhao, David La Fleur, Chethana Naik, Tae-Wook Chun, Tatiana Tsareva, Palanisamy Kanakaraj, Michael W. Laird, Rutul Shah, Lisa George, Indra Sanyal, Paul A. Moore, Borries Demeler, and Senyon Choe. In addition, I would like to thank Andy Gacia, Jingli Cong, Jason Bock and Kimberly Florence for their help with protein production and bioassay.

Chapter 2, in part, is a reprint of the material as it appears in Biochemistry 2007. "Blain, K. Y., Kwiatkowski, W., Zhao, Q., La Fleur, D., Naik, C., Chun, T.W., Tsareva, T., Kanakaraj, P., Laird, M. W., Shah, R., George, L., Sanyal, I., Moore, P. A., Demeler, B., and Choe, S. Structural and functional characterization of CC Chemokine CCL14. *Biochemistry*. 2007; 46(35): 10008-15." The dissertation author was the primary investigator and author of this paper.

2.11 References

1. Murphy, P. M., Baggiolini, M., Charo, I. F., Hebert, C. A., Horuk, R., Matsushima, K., Miller, L. H., Oppenheim, J. J., and Power, C. A. (2000) International union of pharmacology. XXII. Nomenclature for chemokine receptors, *Pharmacol Rev 52*, 145-176.

2. Murphy, P. M. (1996) Chemokine receptors: structure, function and role in microbial pathogenesis, *Cytokine Growth Factor Rev 7*, 47-64.

3. Moser, B., Wolf, M., Walz, A., and Loetscher, P. (2004) Chemokines: multiple levels of leukocyte migration control, *Trends Immunol 25*, 75-84.

4. Baggiolini, M., Dewald, B., and Moser, B. (1997) Human chemokines: an update, *Annu Rev Immunol 15*, 675-705.

5. Proudfoot, A. E. (1998) The chemokine family. Potential therapeutic targets from allergy to HIV infection, *Eur J Dermatol 8*, 147-157.

6. Bacon, K., Baggiolini, M., Broxmeyer, H., Horuk, R., Lindley, I., Mantovani, A., Maysushima, K., Murphy, P., Nomiyama, H., Oppenheim, J., Rot, A., Schall, T., Tsang, M., Thorpe, R., Van Damme, J., Wadhwa, M., Yoshie, O., Zlotnik, A., and Zoon, K. (2003) Chemokine/chemokine receptor nomenclature, *Cytokine 21*, 48-49.

7. St Charles, R., Walz, D. A., and Edwards, B. F. (1989) The three-dimensional structure of bovine platelet factor 4 at 3.0-A resolution, *J Biol Chem 264*, 2092-2099.

8. Clore, G. M., Appella, E., Yamada, M., Matsushima, K., and Gronenborn, A. M. (1990) Three-dimensional structure of interleukin 8 in solution, *Biochemistry 29*, 1689-1696.

9. 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 et al. (1991) Crystal structure of interleukin 8: symbiosis of NMR and crystallography, *Proc Natl Acad Sci U S A 88*, 502-506.

10. Clore, G. M., and Gronenborn, A. M. (1995) Three-dimensional structures of alpha and beta chemokines, *Faseb J 9*, 57-62.

11. Clark-Lewis, I., Kim, K. S., Rajarathnam, K., Gong, J. H., Dewald, B., Moser, B., Baggiolini, M., and Sykes, B. D. (1995) Structure-activity relationships of chemokines, *J Leukoc Biol 57*, 703-711.

12. Smit, J. J., and Lukacs, N. W. (2006) A closer look at chemokines and their role in asthmatic responses, *Eur J Pharmacol 533*, 277-288.

13. Nelken, N. A., Coughlin, S. R., Gordon, D., and Wilcox, J. N. (1991) Monocyte chemoattractant protein-1 in human atheromatous plaques, *J Clin Invest 88*, 1121-1127.

14. Koch, A. E., Kunkel, S. L., Harlow, L. A., Mazarakis, D. D., Haines, G. K., Burdick, M. D., Pope, R. M., Walz, A., and Strieter, R. M. (1994) Epithelial neutrophil activating peptide-78: a novel chemotactic cytokine for neutrophils in arthritis, *J Clin Invest 94*, 1012-1018.

15. Robinson, E., Keystone, E. C., Schall, T. J., Gillett, N., and Fish, E. N. (1995) Chemokine expression in rheumatoid arthritis (RA): evidence of RANTES and macrophage inflammatory protein (MIP)-1 beta production by synovial T cells, *Clin Exp Immunol 101*, 398-407.

16. Berger, E. A., Murphy, P. M., and Farber, J. M. (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease, *Annu Rev Immunol 17*, 657-700.

17. Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M., and Berger, E. A. (1996) CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1, *Science 272*, 1955-1958.

18. Doranz, B. J., Rucker, J., Yi, Y., Smyth, R. J., Samson, M., Peiper, S. C., Parmentier, M., Collman, R. G., and Doms, R. W. (1996) A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors, *Cell 85*, 1149-1158.

19. Dragic, T., Litwin, V., Allaway, 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) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5, *Nature 381*, 667-673.

20. Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P. D., Wu, L., Mackay, C. R., LaRosa, G., Newman, W., Gerard, N., Gerard, C., and Sodroski, J. (1996) The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates, *Cell 85*, 1135-1148.

21. Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R., and Landau, N. R. (1996) Identification of a major co-receptor for primary isolates of HIV-1, *Nature 381*, 661-666.

22. Feng, Y., Broder, C. C., Kennedy, P. E., and Berger, E. A. (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor, *Science 272*, 872-877.

23. Gong, W., Howard, O. M., Turpin, J. A., Grimm, M. C., Ueda, H., Gray, P. W., Raport, C. J., Oppenheim, J. J., and Wang, J. M. (1998) Monocyte chemotactic protein-2 activates CCR5 and blocks CD4/CCR5-mediated HIV-1 entry/replication, *J Biol Chem 273*, 4289-4292.

24. Cocchi, F., DeVico, A. L., Garzino-Demo, A., Arya, S. K., Gallo, R. C., and Lusso, P. (1995) Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells, *Science 270*, 1811-1815.

25. Greco, G., Mackewicz, C., and Levy, J. A. (1999) Sensitivity of human immunodeficiency virus infection to various alpha, beta and gamma chemokines, *J Gen Virol 80 (Pt 9)*, 2369-2373.

26. Westby, M., and van der Ryst, E. (2005) CCR5 antagonists: host-targeted antivirals for the treatment of HIV infection, *Antivir Chem Chemother 16*, 339-354.

27. Schulz-Knappe, P., Magert, H. J., Dewald, B., Meyer, M., Cetin, Y., Kubbies, M., Tomeczkowski, J., Kirchhoff, K., Raida, M., Adermann, K., Kist, A., Reinecke, M., Sillard, R., Pardigol, A., Uguccioni, M., Baggiolini, M., and Forssmann, W. G. (1996) HCC-1, a novel chemokine from human plasma, *J Exp Med 183*, 295-299.

28. Vakili, J., Standker, L., Detheux, M., Vassart, G., Forssmann, W. G., and Parmentier, M. (2001) Urokinase plasminogen activator and plasmin efficiently convert hemofiltrate CC chemokine 1 into its active, *J Immunol 167*, 3406-3413.

29. Detheux, M., Standker, L., Vakili, J., Munch, J., Forssmann, U., Adermann, K., Pohlmann, S., Vassart, G., Kirchhoff, F., Parmentier, M., and Forssmann, W. G. (2000) Natural proteolytic processing of hemofiltrate CC chemokine 1 generates a potent CC chemokine receptor (CCR)1 and CCR5 agonist with anti-HIV properties, *J Exp Med 192*, 1501-1508.

30. Munch, J., Standker, L., Pohlmann, S., Baribaud, F., Papkalla, A., Rosorius, O., Stauber, R., Sass, G., Heveker, N., Adermann, K., Escher, S., Kluver, E., Doms, R. W., Forssmann, W. G., and Kirchhoff, F. (2002) Hemofiltrate CC chemokine 1[9-74] causes effective internalization of CCR5 and is a potent inhibitor of R5-tropic human immunodeficiency virus type 1 strains in primary T cells and macrophages, *Antimicrob Agents Chemother 46*, 982-990.

31. Read, R. J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood, *Acta Crystallogr D Biol Crystallogr 57*, 1373-1382.

32. Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) Likelihood-enhanced fast rotation functions, *Acta Crystallogr D Biol Crystallogr 60*, 432-438.

33. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Likelihood-enhanced fast translation functions, *Acta Crystallogr D Biol Crystallogr 61*, 458-464.

34. Demeler, B. (2005) *UltraScan A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments. Modern Analytical Ultracentrifugation: Techniques and Methods.*, D.J. Scott, S.E. Harding and A.J. Rowe. Eds. Royal Society of Chemistry (UK), 210-229.

35. Tsou, C. L., Gladue, R. P., Carroll, L. A., Paradis, T., Boyd, J. G., Nelson, R. T., Neote, K., and Charo, I. F. (1998) Identification of C-C chemokine receptor 1 (CCR1) as the monocyte hemofiltrate C-C chemokine (HCC)-1 receptor, *J Exp Med 188*, 603-608.

36. Chung, C. W., Cooke, R. M., Proudfoot, A. E., and Wells, T. N. (1995) The three-dimensional solution structure of RANTES, *Biochemistry 34*, 9307-9314.

37. Skelton, N. J., Aspiras, F., Ogez, J., and Schall, T. J. (1995) Proton NMR assignments and solution conformation of RANTES, a chemokine of the C-C type, *Biochemistry 34*, 5329-5342.

38. Lodi, P. J., Garrett, D. S., Kuszewski, J., Tsang, M. L., Weatherbee, J. A., Leonard, W. J., Gronenborn, A. M., and Clore, G. M. (1994) High-resolution solution structure of the beta chemokine hMIP-1 beta by multidimensional NMR, *Science 263*, 1762-1767.

39. Handel, T. M., and Domaille, P. J. (1996) Heteronuclear (1H, 13C, 15N) NMR assignments and solution structure of the monocyte chemoattractant protein-1 (MCP-1) dimer, *Biochemistry 35*, 6569-6584.

40. 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) Monomeric monocyte chemoattractant protein-1 (MCP-1) binds and activates the MCP-1 receptor CCR2B, *J Biol Chem 273*, 33157-33165.

41. Laurence, J. S., Blanpain, C., Burgner, J. W., Parmentier, M., and LiWang, P. J. (2000) CC chemokine MIP-1 beta can function as a monomer and depends on Phe13 for receptor binding, *Biochemistry 39*, 3401-3409.

42. Kim, S., Jao, S., Laurence, J. S., and LiWang, P. J. (2001) Structural comparison of monomeric variants of the chemokine MIP-1beta having differing ability to bind the receptor CCR5, *Biochemistry 40*, 10782-10791.

43. Pakianathan, D. R., Kuta, E. G., Artis, D. R., Skelton, N. J., and Hebert, C. A. (1997) Distinct but overlapping epitopes for the interaction of a CC-chemokine with CCR1, CCR3 and CCR5, *Biochemistry 36*, 9642-9648.

44. Wells, T. N., Power, C. A., Lusti-Narasimhan, M., Hoogewerf, A. J., Cooke, R. M., Chung, C. W., Peitsch, M. C., and Proudfoot, A. E. (1996) Selectivity and antagonism of chemokine receptors, *J Leukoc Biol 59*, 53-60.

45. Forssmann, U., Hartung, I., Balder, R., Fuchs, B., Escher, S. E., Spodsberg, N., Dulkys, Y., Walden, M., Heitland, A., Braun, A., Forssmann, W. G., and Elsner, J. (2004) n-Nonanoyl-CC chemokine ligand 14, a potent CC chemokine ligand 14 analogue that prevents the recruitment of eosinophils in allergic airway inflammation, *J Immunol 173*, 3456-3466.

46. Rajarathnam, K., Sykes, B. D., Kay, C. M., Dewald, B., Geiser, T., Baggiolini, M., and Clark-Lewis, I. (1994) Neutrophil activation by monomeric interleukin-8, *Science 264*, 90-92.

47. Paolini, J. F., Willard, D., Consler, T., Luther, M., and Krangel, M. S. (1994) The chemokines IL-8, monocyte chemoattractant protein-1, and I-309 are monomers at physiologically relevant concentrations, *J Immunol 153*, 2704-2717.

48. Mantel, C., Kim, Y. J., Cooper, S., Kwon, B., and Broxmeyer, H. E. (1993) Polymerization of murine macrophage inflammatory protein 1 alpha inactivates its myelosuppressive effects in vitro: the active form is a monomer, *Proc Natl Acad Sci U S A 90*, 2232-2236.

49. Avalos, B. R., Bartynski, K. J., Elder, P. J., Kotur, M. S., Burton, W. G., and Wilkie, N. M. (1994) The active monomeric form of macrophage inflammatory protein-1 alpha interacts with high- and low-affinity classes of receptors on human hematopoietic cells, *Blood 84*, 1790-1801.

50. Kim, K. S., Rajarathnam, K., Clark-Lewis, I., and Sykes, B. D. (1996) Structural characterization of a monomeric chemokine: monocyte chemoattractant protein-3, *FEBS Lett 395*, 277-282.

51. Baggiolini, M. (1998) Chemokines and leukocyte traffic, *Nature 392*, 565-568.

52. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods in Enzymology 276*, 307-326.

53. Collaborative, Computational, Project, Number, and 4. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr D Biol Crystallogr 50*, 760-763.

54. Wilken, J., Hoover, D., Thompson, D. A., Barlow, P. N., McSparron, H., Picard, L., Wlodawer, A., Lubkowski, J., and Kent, S. B. (1999) Total chemical synthesis and high-resolution crystal structure of the potent anti-HIV protein AOP-RANTES, *Chem Biol 6*, 43-51.

55. Lamzin, V. S., and Wilson, K. S. (1993) Automated refinement of protein models, *Acta Crystallogr D Biol Crystallogr 49*, 129-147.

56. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat Struct Biol 6*, 458-463.

57. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr A 47 (Pt 2)*, 110-119.

58. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.

59. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr D Biol Crystallogr 60*, 2256-2268.

60. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr D Biol Crystallogr 53*, 240-255.

61. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *Journal of Applied Crystallography 26*, 283-291.

62. Moir, S., Lapointe, R., Malaspina, A., Ostrowski, M., Cole, C. E., Chun, T. W., Adelsberger, J., Baseler, M., Hwu, P., and Fauci, A. S. (1999) CD40-Mediated induction of CD4 and CXCR4 on B lymphocytes correlates with restricted susceptibility to human immunodeficiency virus type 1 infection: potential role of B lymphocytes as a viral reservoir, *J Virol 73*, 7972-7980.

63. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *J Appl Cryst 24*, 946-950.

64. POV-Team. (1997) Persistence of Vision Ray Tracer v3.02. World Wide Web: http://www.povray.org.

CHAPTER 3:

Protein Engineering of Bacterial Histidine Kinase Receptor Systems

3.1 Introduction

A TCS is typically comprised of a sensor histidine kinase (HK) receptor and a response regulator (RR). The HK functions as the signal-sensing receptor and auto-phosphorylates the conserved histidine residue on itself upon signal recognition and subsequently transfers this phosphoryl group to a conserved aspartate residue on the RR. In general, the RR is a transcription factor and once phosphorylated, it binds to DNA to activate or repress the transcription of corresponding genes. HK/RRs are probably the most commonly used bacterial sensory systems in nature. The topology of a HK begins with a short cytoplasmic fragment followed by a membrane-spanning helix (TM1) and an extracellular (or periplasmic) sensor domain that is connected to a C-terminal cytosolic portion via a second transmembrane helix (TM2) (Figure 3.1 A) (*1*). While the periplasmic domain varies in sequence and length, reflecting the need to respond to different environmental stimuli, the cytoplasmic portion of the HK typically consists of an array of more conserved modular structures, which in turn folds into two distinct domains: (A) a N-terminal dimerization and histidine phosphotransfer (DHp) domain and (B) a C-terminal catalytic and an ATP-binding (CA) domain. A conserved alpha-helical region known as the HAMP (Histidine kinase, Adenyl cyclase, Methyl-accepting protein and Phosphatase) linker domain connects the TM2 helix to the cytosolic domain and its linkage has been considered to play a central role in transmitting signals across the

membrane (*2, 3*). The DHp domain is responsible for the dimerization of HKs and contains the critical phosphor-accepting histidine residue whereas the CA and DHp domain is responsible for molecular recognition of its cognate RR as well as the hydrolysis of ATP.

The similar modular architectures of HKs and RRs with such a broad variety of signal recognition capabilities allow for the coupling of a wide variety of input signals to appropriate output responses through a conserved phosphotransfer pathway. A handful of chimeric TCSs have been used to explore their functions. The successes of these examples may be generalized to study the molecular events of signal transduction across membranes. Due to the limited scope of this mini-review, it is impossible to provide a comprehensive report, yet we will recapitulate the representative work that has provided substantial understanding of the structure-function relationship of these chimeric systems.

Past work completed on the engineering of receptor kinases mainly employed a domain-swapping strategy where a receptor protein and a HK each contributed a functional module. Among these chimeric systems, the fusion between a chemoreceptor and a HK is the most common. The ability to move toward favorable environmental conditions, called chemotaxis, is common among motile bacteria. There are five bacterial methyl-accepting chemotaxis proteins (MCPs): Tsr, Tar, Tap, Trg and Aer (*4, 5*). Each of these chemoreceptors recognizes different chemical attractants or repellents and responds to a change in the concentration of these periplasmic chemoeffectors. With the exception of Aer, the other four are transmemberane receptors that have a similar structure to many HKs like EnvZ, which makes domain-swapping relatively easy.

3.2 Tar-EnvZ Chimeras: Taz1 and Tez1

The EnvZ-OmpR pair is a well-studied phosphorelay system in *E. coli* (*6-8*). It involves an inner membrane histidine kinase receptor, EnvZ, that responds to changes in medium osmolarity through a series of autophosphorylation, phosphotransfer and phosphatase activities in conjunction with its cognate cytoplasmic response regulator, OmpR. Together this TCS transduces a yet unknown signal through the inner membrane and ultimately regulates the expression of the two major outer membrane porin genes, *ompF* and *ompC* (*9-14*). Another well-known *E. coli* inner membrane protein, Tar, is classified in the family of methyl-accepting chemotaxis proteins, MCPs. The chemoreceptor Tar mediates cell chemotaxis by responding to a variety of different ligands including the amino acid aspartate. Although lacking in primary sequence similarity, Tar has a structural organization comparable to that of EnvZ where both consist of three domains: the periplasmic receptor domain, two transmembrane domains, and the cytoplasmic domain (*12, 15*). Since the specific ligand for EnvZ is unknown, two chimeras have been created, Taz1 and Tez1, where particular domains of EnvZ and Tar have been fused differently to characterize the signal transduction mechanism in response to aspartate.

The chimera Taz1 is created by fusing the periplasmic sensor domain, transmembrane domains, and the cytoplasmic HAMP domain of Tar to the cytoplasmic DHp, and CA signaling domains of EnvZ (Figure 3.1 B) (*16*). Instead of responding to changes in osmolarity, Taz1 senses the presence of aspartate and in turn activates the transcription of the *ompC* gene. An *ompC-lacZ* reporter gene assay indicates that the binding of one Asp molecule to the Tar receptor in Taz1 is sufficient to trigger signal

transduction when the concentration of Asp molecule is in the range of 10^{-3} M, a much less sensitive detection limit than the Tar receptor alone (~ 10^{-6} M). The production of a functional fusion protein suggests a common mechanism of transmembrane signaling shared by both Tar and EnvZ. The Tar receptor forms a dimer where the two Asp-binding pockets are formed by the interfaces of two subunits. Mutations are introduced in the Asp-binding pocket of Tar (to disrupt the Asp binding capacity) and in the DHp or G1, G2 domains of EnvZ (to eliminate autophosphorylation or the ATP binding capacity). Through different combinations of these mutations, interesting phenotypes have been observed and analyzed (*17*). These experiments demonstrate that one Asp-binding pocket and one EnvZ active site is sufficient to regulate the cellular concentration of OmpR-P upon Asp binding. However, in two other mutant constructs Asp fails to elicit an induction of *ompC-lacZ* (*17*). These results may have stemmed from a conformational change in helix 4 in the Tar receptor upon Asp binding (*18, 19*). Helix 4 forms part of the Asp-binding pocket and transduces the conformational change through a "swinging-piston" motion to TM2 (*18, 19*), which propagates to the EnvZ cytoplasmic domain through the HAMP domain. This Asp-dependent regulation is detected only when the functional His243- residue and the functional ATP-binding pocket are both present. These studies on the Taz1 chimera contributed a better understanding of how signal transduction across the membrane occurs, by illustrating how specific subunits within the different domains in a receptor dimer function in response to a ligand.

Similar to the structural composition of Taz1, the chimera Tez1 is also a fusion between Tar and EnvZ, except that the cytoplasmic HAMP domain of EnvZ replaces that

of Tar (Figure 3.1 C). Although the structural organization is similar to that of Taz1, Tez1 does not respond to aspartate (*20*), illustrating that the HAMP domain is critical for proper transduction. The NMR structure of a dimeric HAMP domain from the hyperthermophilic bacterium *A. fulgidus* (*21*) indicates that it forms a parallel four-helical coiled-coil fold with each subunit contributing two helices. The HAMP domain of Tar does not share a high sequence identity with that of EnvZ, but both structures are predicted to form the conserved helix–turn–helix motif. Interestingly, the Tar HAMP domain possesses only one extra residue than that in EnvZ. Since the initial chimeric protein has a nonregulatable phenotype, extensive mutational analysis such as deletion, addition, or random mutagenesis in the EnvZ HAMP domain of Tez1 has been carried out. It is found that a single Ala insertion at the beginning of the HAMP domain rescues the Asp-nonregulatable phenotype. Moreover, the position and the nature of the residue type are important since the Ala residue is necessary required to fully restore the phenotype. Missense mutations TezP185Q or TezR184Q identified by the error-prone PCR method also enable the chimera to respond to Asp. As observed in the Ala insertion experiment, the substitution by Gln at these two positions is quite specific (*20*). Therefore the region leading to helix I in the EnvZ HAMP domain is crucial for its function. These studies illustrate that the HAMP domains share a similar molecular mechanism of signal transduction. More importantly, it suggests that the HAMP domains are finely tuned and slight changes in configuration can influence their functions.

3.3 Trg-EnvZ Chimera: Trz1

Analogous to Taz1 and Tez1, another EnvZ chimera known as Trz1, is also created through the fusion of the two transmembrane domains, the ribose- and galactose-binding protein-sensing domain, and the cytoplasmic HAMP domain of the chemosensor Trg to the cytoplasmic DHp and CA domains of EnvZ (*22*). This hybrid receptor has been shown to be functional, where upon recognition of the ligand-occupied ribose-binding protein by the sensor domain causes the activation of the *ompC* promoter. Using an *ompC-lacZ* reporter gene assay, Baumgartner and colleagues demonstrate that in the presence of ribose there is concentration dependent β-galactosidase activity that is more than 20-fold higher than that in the absence of ribose (*22*). Furthermore, alignment of the primary sequences of EnvZ, Tar, and Trg reveal a few common residues of interest located in the TM1 and HAMP domains, which again highlights the importance of these regions.

3.4 Tar-ArcB Chimera-Tab

In addition to the discoveries made by the creation of a variety of different chimeras using the EnvZ kinase, functional characterization using other HKs has also been completed. The ArcB/A TCS is in charge of anoxic redox control and through a sophisticated phosphorelay process can regulate the expression of more than 30 operons depending on the redox conditions of growth (*23-25*). ArcB bears a complex cytosolic structure in that it consists of three catalytic domains: a primary transmitter domain with a conserved His residue at position 292, a receiver domain with a conserved Asp at position 576, and a secondary transmitter domain with a conserved His at position 717 (Figure 3.2) (*26, 27*). Under reducing conditions, ArcB undergoes autophosphorylation at

His292. The phosphoryl group is then transferred to ArcA via a His292/Asp576/His717/Asp54 phospho-relay process (*28-30*). Under oxidizing conditions, the reverse reaction will occur and the enzyme catalyzes the dephosphorylation of ArcA-P (*31*). Another unorthodox structural feature of ArcB is the lacking of an apparent periplasmic domain, consisting of only 16 amino acid residues (*32*). An initial construct of Tar-ArcB or Tab1 is created by an in-frame fusion of the N-terminus of Tar (residues 1–212 ending immediately after TM2) and the cytosolic domains of ArcB (residues 78–778) and the hybrid kinase is highly active as an autokinase even under aerobic conditions (*33*). In order to evaluate the significance of the relative orientations of the two ArcB segments set by the dimeric membrane anchor, a series of hybrids are constructed in which the TM2 of Tar is extended with progressively longer C-termini: Tar1–213-ArcB78–778 (Tab-2), Tar1–214-ArcB78–778 (Tab-3), and Tar1–215-ArcB78–778 (Tab-4). Each extra residue of Tar added twists to the helical wheel of the ArcB kinase by an angle of 100° due to the rigid helical nature of TM2 that protrudes into the cytoplasm. As a result, each construct represents a different spatial relationship between the Tar anchor and the cytosolic ArcB portion and different phenotypes are observed. Tab-2, with one more residue of Tar at the hybrid junction than Tab-1, is constitutively inactive under both aerobic and anaerobic growth conditions, as shown by the β-galactosidase activity assay. Tab-3, with one additional residue than Tab-2, behaves like the wildtype ArcB, displaying a wildtype phenotype. Tab-4, possessing another additional residue than Tab-3, has a relative relationship between the two ArcB moieties similar to that of Tab-1. This dimer once again is active as a kinase, being constitutively turned “on”.

Based on the completely different catalytic properties of various Tab constructs, it is proposed that the signal transduction process of ArcB occurs via a rotation-dependent mechanism caused by its cytosolic portion. This mechanism is in disagreement with the piston-like displacement employed by Tar, whose vertical displacements between the two transmembrane segments within the same subunit is believed to regulate the catalytic activity of the cytosolic domain upon ligand binding.

3.5 Cph1-EnvZ Chimera: Cph8

Chemoreceptors are not the only proteins that have been fused to HK receptors to create functional chimeras. A previous study has shown that the light sensing phytochrome Cph1 from cyanobacterium can be fused to the cytoplasmic domain of EnvZ to create the chimera, Cph8, and still maintain functionality (*34*). Like histidine kinase TCSs, a phytochrome is composed of an extracellular sensor domain and an intracellular response-regulator (*35*), but in contrast, the phytochrome response regulator cannot regulate gene expression directly. However, when fused to EnvZ and transformed, along with specific phycocyanobilin genes that are required for proper functioning of the photoreceptor domain, into a ΔEnvZ *E. coli* strain with an *ompC-lacZ* gene fusion, Cph8 demonstrates a strong response to light as detected by light repressed gene expression (*34*). In the presence of red light, Cph8 remains in a state where autophosphorylation is inhibited and gene expression is repressed. In the absence of light Cph8 is activated and can be visualized through the addition of the β-gal substrate S-gal, which upon catalysis yields a black precipitate. In this example, the TCS chimera Cph8 possesses an image-processing role with its protein-engineered circuit by which bacteria

can now function as biological films producing high-definition images. The research on this engineered HK photoreceptor allows for a better understanding of signal transduction within TCS systems and provides possibilities for potential application in such fields as biological materials and bacterial microlithography.

3.6 TGF-β receptor-EnvZ Chimera: TB-EnvZ

A human TGF-β receptor protein consists of a ligand-sensing extracellular domain (ECD), a transmembrane helix and a cytoplasmic kinase domain. They can be expressed in *E. coli* using an N-terminal Mistic fusion (*36*). This technique would enable us to create novel chimeric TCSs in *E. coli* that are able to sense an array of human TGF-beta ligands and then trigger downstream responses in a ligand-restricted manner. Although such a TGF-β-sensible *E. coli* system has not been created, one can conceivably adopt the native EnvZ/OmpR signaling system as the starting point for engineering. As an example, an *E. coli* strain carrying an antibiotic resistance gene can be created as a template strain for *in vivo* genetic selection. In this modified strain, the chromosomal *ompc* gene downstream to the EnvZ/OmpR signaling pathway (normally expressed at high osmolarity) is replaced by an antibiotic resistant gene. Once the engineered cell is confirmed to show antibiotic resistance in high osmolarity conditions, it could be further engineered to delete the chromosomal EnvZ gene to prevent the signaling through the native EnvZ/OmpR pathway. For a TGF-β-EnvZ chimera, TB-EnvZ, the N-terminal TM1 and extracellular sensor domain of native EnvZ is to be replaced with the ECD of a TGF-β receptor. A library of TB-EnvZ constucts can be

created and transformed into the template *E. coli* strain for genetic selection using the antibiotics. Once functional TB-EnvZ chimeras are isolated, other genes of interest can replace the antibiotics resistance gene. To create novel specificity, new amino acids can be introduced by random mutagenesis to create *E. coli* that can respond to other ligands, and expand to introduce into other organisms to "synthesize" organisms with novel biological functionalities.

3.7 Engineering on Response Regulators

In addition to the manipulation of HK receptors, engineering on response regulators has also been carried out in some mechanistic studies of TCSs. The RR is usually comprised of two domains: the highly conserved receiver domain that interacts with HKs and accepts the incoming phosphoryl group, and the effector domain that outputs the signal, through regulation of gene transcription. Like their kinase counterparts, engineering on RRs is mainly accomplished through a 'domain-swapping' method based on the modular nature of RR (*37*). Specifically, it involves the fusion of the receiver domain from one response regulator with an output (DNA-binding) domain from another.

3.7.1 OmpR – PhoB chimera

The PhoR–PhoB two-component regulatory system is required to allow *E. coli* to sense environmental inorganic phosphate levels (*38*). Under limiting phosphate conditions, the PhoRB TCS activates the transcription of the phosphate (Pho) regulon,

which is responsible for the uptake of phosphorus-containing compounds. The RR, PhoB is highly homologous to OmpR - both in sequence and structure. OmpR and PhoB belong to the winged-helix-turn-helix family of DNA binding proteins (*39, 40*) (Figure 3.3 A) and their C-terminal structures superimpose well (*40*). However, the process by which OmpR and PhoB regulate their effector domains is quite different. Phosphorylation of the N terminus in OmpR boosts the DNA binding affinity of the C terminus, whereas phosphorylation of the N-terminus in PhoB relieves the inhibition of the C terminus so that it can bind to cognate DNA more easily. In either case, the interdomain linker appears to play an indispensable role. To better understand the activation and regulation mechanisms, chimeras between OmpR and PhoB containing either interdomain linker are constructed (*41*). When the $PhoB_N$-$OmpR_C$ chimeras are expressed in the *ompR* null reporter strains containing *ompF-lacZ* or *ompC-lacZ* operon fusions (*42*), both chimeras are impaired in activating the transcription of OmpR-dependent genes, although the level of phosphorylation (by PhoR) of the chimeras is comparable to that of PhoB alone. Consequently, the lack of transcriptional activation is not a result of under-phosphorylation. Interestingly, the $OmpR_N$-$PhoB_C$ chimeras are constitutively active in an *envZ* null strain, as demonstrated by their abilities to activate the transcription of the PhoB-specific reporter gene *phoA*. Hence, $OmpR_N$ does not exert an inhibitory effect on the binding of $PhoB_C$ to DNA. In addition, in the $envZ^+$ background, the RRB chimera (the construct containing the OmpR linker) is able to induce a threefold increase of *phoA* transcription while that for RBB chimera (the construct containing the PhoB linker) remains the same. It therefore can be concluded that the activation of $PhoB_C$ requires the OmpR linker. Further studies show that a functional chimera requires a linker of defined

length and composition, because specific substitutions for certain amino acids lead to an *ompF*(Con) *ompC* null phenotype, while changes in the linker length have a large influence on the activation efficiency. Altogether, this work indicates that regulation of effector domain by either N terminus requires its cognate interdomain linker (*41*).

3.7.2 PhoP'-'YycF chimera

Like the PhoR–PhoB TCS, PhoR-PhoP system activates or represses Pho regulon genes to overcome a phosphate deficiency. YycG-YycF is a highly conserved TCS in many Gram-positive bacteria and is essential for the viability of *B. subtilis* (*43*). However the characteristics of the YycF regulon and the exact functions of YycG-YycF system are yet to be identified. In this example, the PhoR/PhoP TCS is used to investigate the unknown functions of the YycG-YycF TCS and the hybrid gene PhoP'-'YycF fusing the receiver domain from *phoP* with the linker and the output domain from *yycF* is created (*44*). Bacteria bearing the hybrid regulator gene are grown under phosphate-limiting conditions to induce expression of the PhoP'-'YycF chimera. In response to phosphate limitation, the PhoR kinase phosphorylates the receiver domain of PhoP in the hybrid, and then YycF in the hybrid regulator binds to its specific DNA promoter. To examine the transcriptome contents, total RNA of the *B. subtilis* cells is extracted, reverse-transcribed and analyzed by a macroarray system. The result demonstrates that, during phosphate-limiting growth, the PhoP'-'YycF hybrid is capable of eliciting a response and activates the expression of *yocH,* which encodes a potential autolysin (*45*). The binding and affinity of YycF for *yocH* have been characterized by gel mobility shift and DNase I

footprinting assays. Subsequent Northern blotting analysis shows that this activation only occurs in the presence of PhoR kinase, indicating that PhoR/PhoP'-'YycF is a functional hybrid TCS.

Using the same strategy, Howell and colleagues also construct YycF'-'PhoP, fusing the receiver domain of YycF and the output domain as well as the linker of PhoP (*44*). This reverse hybrid response regulator is also active, suggesting the domains between PhoP and YycF are inter-changeable. This study represents a new approach to systematically engineer RRs to create TCSs of novel functions.

3.8 Discussion

In summary, HK and RR proteins are two major protein components that are excellent candidates for modular rearrangements. The fusion systems reviewed here allow us to gain insights into TCS signaling pathways. As more crystal structures of HK-RR or RR-DNA promoter complexes are reported (*46-50*) (Figure 3.3), systematic approaches in engineering TCSs begin to emerge. In a recent crystal structure of a HK-RR complex, the molecular mechanism of the phosphorelay from the transmitting domain of HK853 and the receiver domain of its cognate RR468 is described (*46*). With the detailed knowledge of these interactions, the reprogramming of a TCS pathway is possible through rational design of the cytoplasmic domain of a HK to change its specificity for the response regulator. For instance, the substrate specificity of EnvZ has been changed through mutational studies guided by a computational approach (*51*). Additionally, one can even swap the effector domain of a RR with that from a paralog within or across species barriers to create transcriptional regulatory machinery using an

evolutionarily unrelated set of genes and ultimately monitor or control gene expression. These synthetic organisms equipped with the ‘rewired’ signaling pathways will find wide use in future research and in clinical and industrial applications.

3.9 Figures

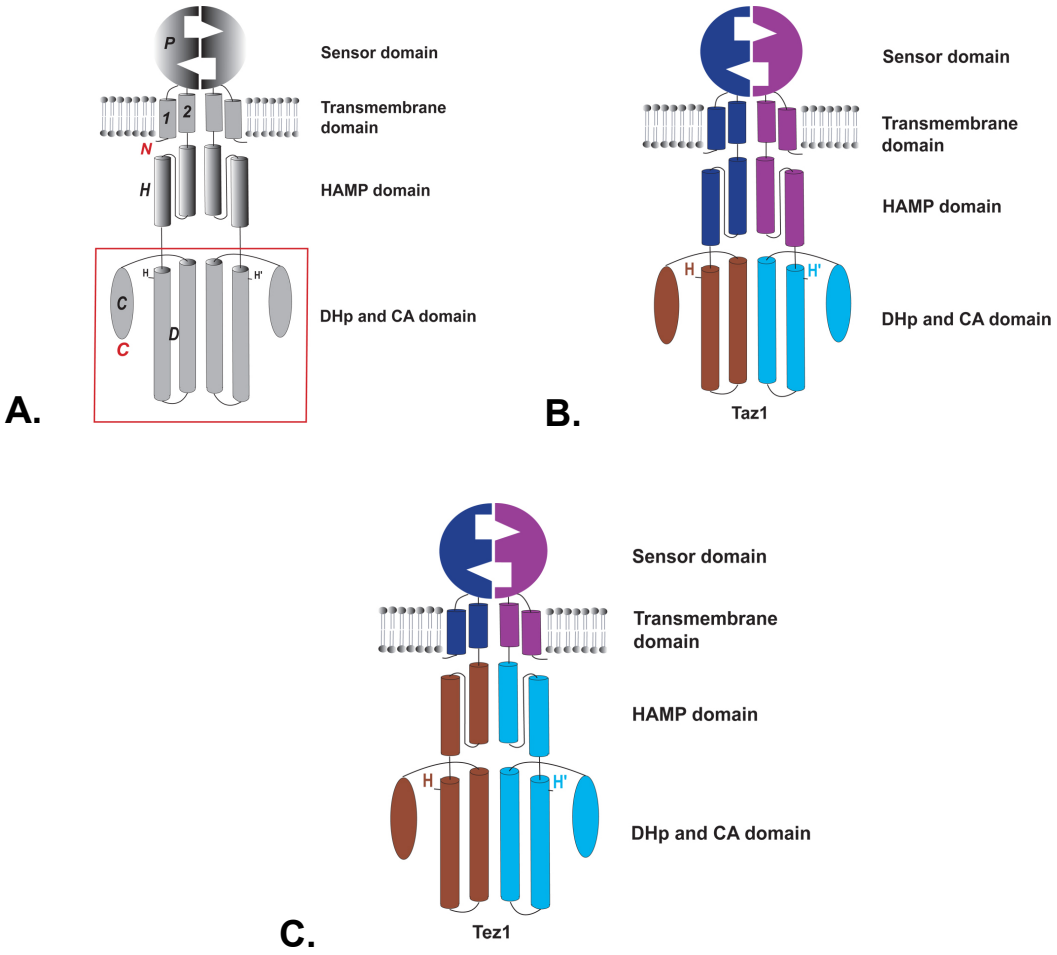

Figure 3.1: The diagram of Tar-EnvZ chimeras.
(A) The illustration of EnvZ dimeric structure. The domains are represented as follows by bold italic letters and numbers: *1* for TM1, *2* for TM2, *P* for periplasmic domain, *H* for HAMP domain, *D* for DHp, *C* for cytosolic kinase domain. The N and C termini are represented by red bold italic letters *N* and *C* respectively while the phosphor-accepting histidine is shown by regular bold letter H. (B) The Taz1 construct is composed of the DHp and CA domains from EnvZ and a periplasmic domain, transmembrane domain, and the HAMP domain from Tar. H denotes for His243 in DHp domain. (C) The Tez1 construct with the HAMP domain of Taz1 replaced with EnvZ HAMP domain. In (b) and (c), domains from Tar are shown in blue and purple while domains from EnvZ are shown in brown and cyan (52).

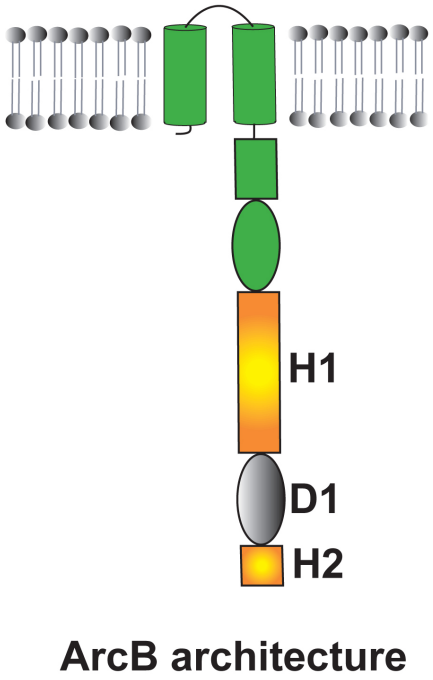

Figure 3.2: The modular domains of the HK ArcB.
ArcB is composed of a short periplasmic segment between the 2 TM helices and a complex cytosolic portion where it consists of three catalytic domains following the PAS domain: a primary transmitter domain H1, a receiver domain D1, and a secondary transmitter domain H2 (52).

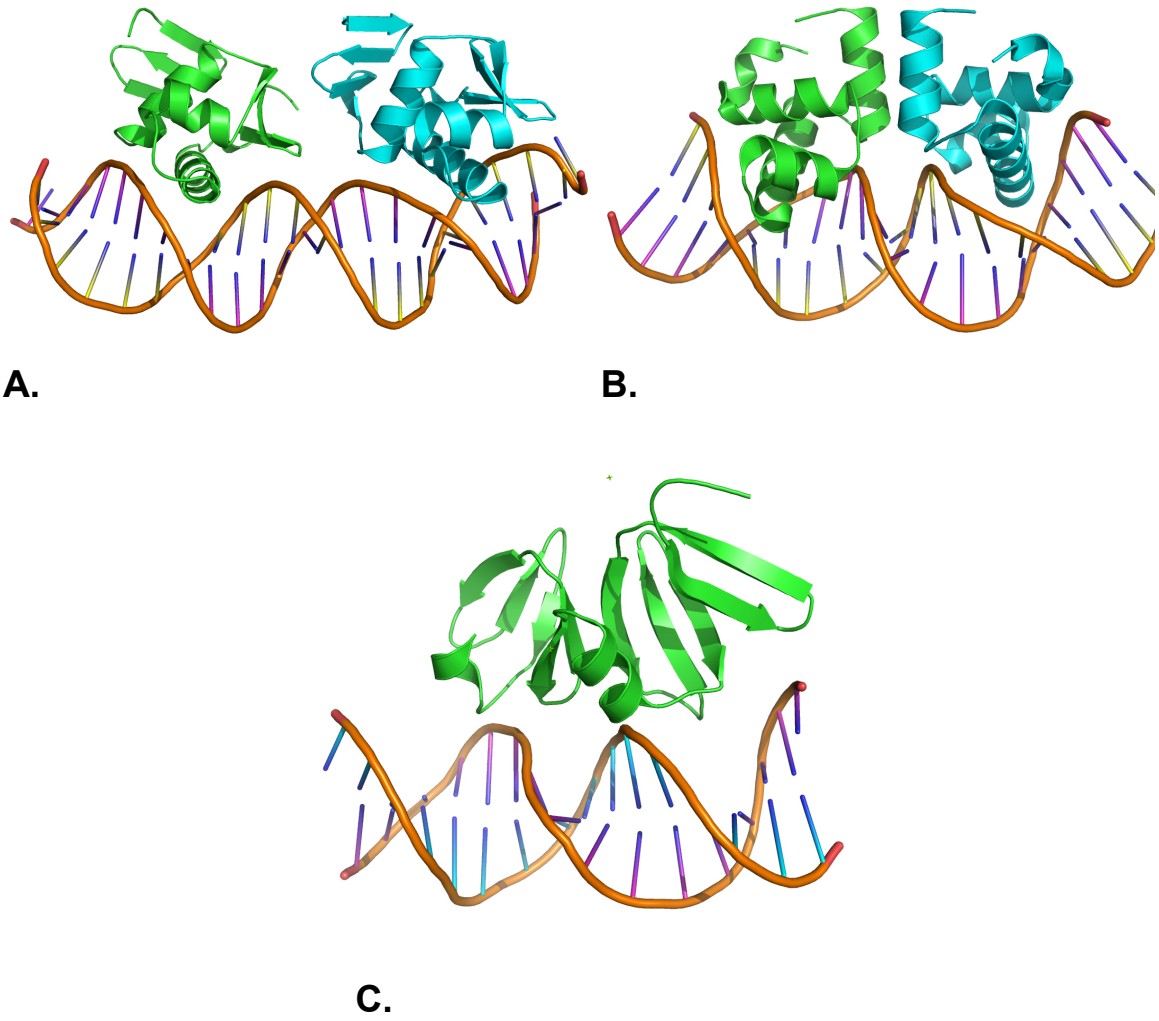

Figure 3.3: The ribbon representations of different DNA-binding modes illustrated by crystal structures of RR-DNA complexes.
(A) The OmpR/PhoB winged-helix domain. (B) The NarL/FixJ four-helix helix-turn-helix domain. (C) The β fold of AgrA LytTR domain (52).

3.10 Acknowledgements

This work was supported by grants from the National Institutes of Health, IFEZ and WCU-Korea (S.C.). I would like to thank the fellowship support from the H. A. Mary K. Chapman Charitable Trust and The Mary K. Chapman Foundation. I would also like to acknowledge the following co-authors that helped contribute to the work that was completed: Wei Xie, Mario Meng-Chiang Kuo and Senyon Choe. Chapter 3, in part, is a reprint of the material as it appears in *Protein & Peptide Letters* 2010: "Xie, W., Blain, K. Y., Kuo, M. M., Choe, S. "Protein engineering of bacterial histidine kinase receptor systems." *Protein Pept Lett.* 2010; 17 (7): 867-873" The dissertation author was one of the primary investigators and authors of this paper.

3.11 References

1. Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000) Two-component signal transduction, *Annu Rev Biochem 69*, 183-215.

2. Fabret, C., Feher, V. A., and Hoch, J. A. (1999) Two-component signal transduction in Bacillus subtilis: how one organism sees its world, *J Bacteriol 181*, 1975-1983.

3. Williams, S. B., and Stewart, V. (1999) Functional similarities among two-component sensors and methyl-accepting chemotaxis proteins suggest a role for linker region amphipathic helices in transmembrane signal transduction, *Mol Microbiol 33*, 1093-1102.

4. Stock, J. B., Surette, M.G. (1996) Chemotaxis. In: Escherichia coli and Salmonella: cellular and molecular biology, (Neidhardt, F. C., Curtiss, R. III, Ingraham, J. L., Lin, E. C., Low, K. B. Jr., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umbarger, H. E., Ed.) 2nd ed., pp 1193-1129, ASM Press, Washington, D.C.

5. Bibikov, S. I., Biran, R., Rudd, K. E., and Parkinson, J. S. (1997) A signal transducer for aerotaxis in Escherichia coli, *J Bacteriol 179*, 4075-4079.

6. Egger, L. A., Park, H., and Inouye, M. (1997) Signal transduction via the histidyl-aspartyl phosphorelay, *Genes Cells 2*, 167-184.

7. Forst, S. A., and Roberts, D. L. (1994) Signal transduction by the EnvZ-OmpR phosphotransfer system in bacteria, *Res Microbiol 145*, 363-373.

8. Pratt, L. A., and Silhavy, T. J. (1995) Identification of base pairs important for OmpR-DNA interaction, *Mol Microbiol 17*, 565-573.

9. Aiba, H., Nakasai, F., Mizushima, S., and Mizuno, T. (1989) Phosphorylation of a bacterial activator protein, OmpR, by a protein kinase, EnvZ, results in stimulation of its DNA-binding ability, *J Biochem 106*, 5-7.

10. Aiba, H., Nakasai, F., Mizushima, S., and Mizuno, T. (1989) Evidence for the physiological importance of the phosphotransfer between the two regulatory components, EnvZ and OmpR, in osmoregulation in Escherichia coli, *J Biol Chem 264*, 14090-14094.

11. Forst, S., and Inouye, M. (1988) Environmentally regulated gene expression for membrane proteins in Escherichia coli, *Annu Rev Cell Biol 4*, 21-42.

12. Forst, S., Comeau, D., Norioka, S., and Inouye, M. (1987) Localization and membrane topology of EnvZ, a protein involved in osmoregulation of OmpF and OmpC in Escherichia coli, *J Biol Chem 262*, 16433-16438.

13. Forst, S., Delgado, J., and Inouye, M. (1989) Phosphorylation of OmpR by the osmosensor EnvZ modulates expression of the ompF and ompC genes in Escherichia coli, *Proc Natl Acad Sci U S A 86*, 6052-6056.

14. Igo, M. M., and Silhavy, T. J. (1988) EnvZ, a transmembrane environmental sensor of Escherichia coli K-12, is phosphorylated in vitro, *J Bacteriol 170*, 5971-5973.

15. Krikos, A., Mutoh, N., Boyd, A., and Simon, M. I. (1983) Sensory transducers of E. coli are composed of discrete structural and functional domains, *Cell 33*, 615-622.

16. Utsumi, R., Brissette, R. E., Rampersaud, A., Forst, S. A., Oosawa, K., and Inouye, M. (1989) Activation of bacterial porin gene expression by a chimeric signal transducer in response to aspartate, *Science 245*, 1246-1249.

17. Yang, Y., Park, H., and Inouye, M. (1993) Ligand binding induces an asymmetrical transmembrane signal through a receptor dimer, *J Mol Biol 232*, 493-498.

18. Chervitz, S. A., and Falke, J. J. (1996) Molecular mechanism of transmembrane signaling by the aspartate receptor: a model, *Proc Natl Acad Sci U S A 93*, 2545-2550.

19. Milburn, M. V., Prive, G. G., Milligan, D. L., Scott, W. G., Yeh, J., Jancarik, J., Koshland, D. E., Jr., and Kim, S. H. (1991) Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand, *Science 254*, 1342-1347.

20. Zhu, Y., and Inouye, M. (2003) Analysis of the role of the EnvZ linker region in signal transduction using a chimeric Tar/EnvZ receptor protein, Tez1, *J Biol Chem 278*, 22812-22819.

21. Hulko, M., Berndt, F., Gruber, M., Linder, J. U., Truffault, V., Schultz, A., Martin, J., Schultz, J. E., Lupas, A. N., and Coles, M. (2006) The HAMP domain structure implies helix rotation in transmembrane signaling, *Cell 126*, 929-940.

22. Baumgartner, J. W., Kim, C., Brissette, R. E., Inouye, M., Park, C., and Hazelbauer, G. L. (1994) Transmembrane signalling by a hybrid protein: communication from the domain of chemoreceptor Trg that recognizes sugar-binding proteins to the kinase/phosphatase domain of osmosensor EnvZ, *J Bacteriol 176*, 1157-1163.

23. Iuchi, S., and Lin, E. C. (1988) arcA (dye), a global regulatory gene in Escherichia coli mediating repression of enzymes in aerobic pathways, *Proc Natl Acad Sci U S A 85*, 1888-1892.

24. Lynch, A. S., Lin, E. C. C. In: Regulation of Gene Expression in Escherichia coli, (Lin, E. C. C., Lynch, A.S. , Ed.), Landes, Georgetown, TX.

25. Lynch, A. S., Lin, E. C. C. . In: Escherichia coli and Salmonella typhimurium: Cellular and Molecular biology, (Neidhardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M. Schaechter, M., Umbarger, H. E., Ed.), pp 1526-1538, American Society for Microbiology, Washington, D. C.

26. Iuchi, S., Matsuda, Z., Fujiwara, T., and Lin, E. C. (1990) The arcB gene of Escherichia coli encodes a sensor-regulator protein for anaerobic repression of the arc modulon, *Mol Microbiol 4*, 715-727.

27. Ishige, K., Nagasawa, S., Tokishita, S., and Mizuno, T. (1994) A novel device of bacterial signal transducers, *EMBO J 13*, 5195-5202.

28. Iuchi, S., and Lin, E. C. (1992) Purification and phosphorylation of the Arc regulatory components of Escherichia coli, *J Bacteriol 174*, 5617-5623.

29. Georgellis, D., Lynch, A. S., and Lin, E. C. (1997) In vitro phosphorylation study of the arc two-component signal transduction system of Escherichia coli, *J Bacteriol 179*, 5429-5435.

30. Kwon, O., Georgellis, D., and Lin, E. C. (2000) Phosphorelay as the sole physiological route of signal transmission by the arc two-component system of Escherichia coli, *J Bacteriol 182*, 3858-3862.

31. Georgellis, D., Kwon, O., De Wulf, P., and Lin, E. C. (1998) Signal decay through a reverse phosphorelay in the Arc two-component signal transduction system, *J.Biol.Chem. 273*, 32864-32869.

32. Kwon, O., Georgellis, D., Lynch, A. S., Boyd, D., and Lin, E. C. (2000) The ArcB sensor kinase of Escherichia coli: genetic exploration of the transmembrane region, *J Bacteriol 182*, 2960-2966.

33. Kwon, O., Georgellis, D., and Lin, E. (2003) Rotational on-off switching of a hybrid membrane sensor kinase Tar-ArcB in Escherichia coli., *J Biol Chem 278*, 13192-13195.

34. Levskaya, A., Chevalier, A., Tabor, J., Simpson, Z., Lavery, L., Levy, M., Davidson, E., Scouras, A., Ellington, A., Marcotte, E., and Voigt, C. (2005) Synthetic biology: engineering Escherichia coli to see light., *Nature 438*, 441-442.

35. Yeh, K., Wu, S., Murphy, J., and Lagarias, J. (1997) A cyanobacterial phytochrome two-component light sensory system., *Science 277*, 1505-1508.

36. Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R., and Choe, S. (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression, *Science 307*, 1317-1321.

37. Parkinson, J. S. (1995) In *Genetic approaches for signaling pathways and proteins* (Hock, J. A., Silhavy, T. J., Ed.), pp 9-23, American Society for Microbiolog Press, Washington, D. C.

38. Wanner, L. B. (1996) Escherichia coli and Salmonella typhimurium: Cellular and Molecular biology, (Neidhardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M. Schaechter, M., Umbarger, H. E., Ed.), pp 1357-1381, American Society for Microbiology, Washington, D. C.

39. Kenney, L. (2002) Structure/function relationships in OmpR and other winged-helix transcription factors., *Curr Opin Microbiol 5*, 135-141.

40. Itou, H., and Tanaka, I. (2001) The OmpR-family of proteins: insight into the tertiary structure and functions of two-component regulator proteins., *J Biochem 129*, 343-350.

41. Walthers, D., Tran, V. K., and Kenney, L. J. (2003) Interdomain linkers of homologous response regulators determine their mechanism of action, *J Bacteriol 185*, 317-324.

42. Hall, M. N., and Silhavy, T. J. (1981) The ompB locus and the regulation of the major outer membrane porin proteins of Escherichia coli K12, *J Mol Biol 146*, 23-43.

43. Fabret, C., and Hoch, J. A. (1998) A two-component signal transduction system essential for growth of Bacillus subtilis: implications for anti-infective therapy, *J Bacteriol 180*, 6375-6383.

44. Howell, A., Dubrac, S., Andersen, K. K., Noone, D., Fert, J., Msadek, T., and Devine, K. (2003) Genes controlled by the essential YycG/YycF two-component system of Bacillus subtilis revealed through a novel hybrid regulator approach, *Mol Microbiol 49*, 1639-1655.

45. Smith, T. J., Blackman, S. A., and Foster, S. J. (2000) Autolysins of Bacillus subtilis: multiple enzymes with multiple functions, *Microbiology 146 (Pt 2)*, 249-262.

46. Casino, P., Rubio, V., and Marina, A. (2009) Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction, *Cell 139*, 325-336.

47. Blanco, A. G., Sola, M., Gomis-Ruth, F. X., and Coll, M. (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator, *Structure 10*, 701-713.

48. Maris, A. E., Sawaya, M. R., Kaczor-Grzeskowiak, M., Jarvis, M. R., Bearson, S. M., Kopka, M. L., Schroder, I., Gunsalus, R. P., and Dickerson, R. E. (2002) Dimerization allows DNA target site recognition by the NarL response regulator, *Nat Struct Biol 9*, 771-778.

49. Sidote, D. J., Barbieri, C. M., Wu, T., and Stock, A. M. (2008) Structure of the Staphylococcus aureus AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding, *Structure 16*, 727-735.

50. Zapf, J., Sen, U., Madhusudan, Hoch, J. A., and Varughese, K. I. (2000) A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction, *Structure 8*, 851-862.

51. Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M., and Laub, M. T. (2008) Rewiring the specificity of two-component signal transduction systems, *Cell 133*, 1043-1054.

52. Xie, W., Blain, K. Y., Kuo, M. M., and Choe, S. (2010) Protein engineering of bacterial histidine kinase receptor systems, *Protein Pept Lett 17*, 867-873.

CHAPTER 4:

The Functionally Active Mistic-Fused Histidine Kinase Receptor, EnvZ

4.1 Introduction

Accounting for approximately 30% of all proteins in both prokaryotic and eukaryotic organisms are the integral membrane proteins. They are required for major cellular functions and are thus important pharmaceutical targets (*1-4*). Unfortunately, structural and biochemical studies of integral membrane proteins are hampered in part by low levels of expression. Therefore, a heterologous expression system is often employed to overcome this setback. Mistic is a 13 kDa, 110-amino acid *Bacillus subtilis* protein that has unique structural and functional properties. The NMR structure of Mistic has illustrated that it consists of a four α-helical bundle with a hydrophilic surface (*5*). Mistic differs from other membrane-integrated proteins in that it appears to interact with the lipid bilayer and can bypass the traditional cellular translocon machinery for membrane integration. Previous studies illustrate that both prokaryotic and eukaryotic membrane protein expression levels were boosted when target proteins are fused to Mistic (*6, 7*). Despite the utility of this Mistic-fusion system in improving expression levels of membrane proteins, the critical question still remains whether the overexpressed cargo protein remains functional as a fusion partner to Mistic. In this study, we chose to analyze the prokaryotic two-component signal transduction system EnvZ-OmpR to address this question.

Prokaryotic organisms utilize two-component signal transduction systems as their principal mode for adapting to various environmental stresses (*8*). One of the most

widely studied and best characterized two-component systems involves the interaction between the osmo-sensing histidine kinase receptor, EnvZ and its cytoplasmic cognate response regulator, OmpR (*8-10*). EnvZ is a 450-amino acid inner membrane protein consisting of an NH_2-terminal cytoplasmic tail, periplasmic sensor domain, two transmembrane domains, and a COOH-terminal cytoplasmic domain. The cytoplasmic domain is further divided into a HAMP-linker (histidine kinases, adenylyl cyclases, methyl-accepting chemotaxis proteins, and phosphatases), DHp (dimerization and histidine phosphotransfer) domain and the CA (catalytic and ATP-binding) domain (*11-14*) (Figure 4.1 A). Upon activation EnvZ will autophosphorlyate on His_{243} (*15*) and phophotransfer the phosphoryl group to residue Asp_{55} on OmpR. Phosphorylated OmpR, OmpR-P, functions as a transcription factor and subsequently controls the expression of the genes for the two major outer membrane porins OmpF and OmpC (*16-18*). Like the majority of other histidine kinases, EnvZ has two major functions, possessing not only kinase activity but also the ability to act as a phosphatase when complexed with OmpR (*16, 18*), where it dephosphorylates OmpR-P and in turn regulates the concentration of OmpR-P in the cytoplasm (*19*).

In addition to its use in studies involving two-component phosphorelays, EnvZ has also been exploited for various protein engineering purposes (*20*). In the past, EnvZ has been utilized to create many different chimeras which involve domain swapping with different chemoreceptors such as Tar (*21, 22*) and Trg (*23*) in order to study the signaling mechanisms behind two-component systems. More recently, EnvZ has been fused to the cyanobacterium light sensing phytochrome, Cph1, to create a chimera which functions in a unique system with an image processing role, thus permitting bacteria to exhibit

properties like that of film (*24*). In this study we have used EnvZ to test if the Mistic fusion affects its catalytic and signaling capabilities (Figure 4.1 B). Here we provide the first report which illustrates that the *E. coli* histidine kinase receptor, EnvZ, maintains its transmembrane signaling abilities when fused to Mistic, based on the data from both *in vitro* assays, through autophosphorylation and phosphotransfer to OmpR, and from *in vivo* assays through activation of *ompC-lacZ* gene expression, with Mistic-fused EnvZ.

4.2 Expression and Purification of EnvZ, Misticated EnvZ and OmpR

One major challenge of studying membrane proteins is the ability to overexpress and isolate a pure homogeneous sample. With the use of a Mistic fusion (Figure 4.1 B), approximately 19 mg of pure homogenous Misticated EnvZ was obtained by Ni-NTA affinity and size exclusion chromatography from a liter of cultured media. In contrast, EnvZ expressed without Mistic fusion yielded approximately 7 mg of pure homogeneous protein per liter of culture. The soluble response regulator OmpR also expressed to large quantities as well, yielding approximately 7 mg of pure homogenous protein by Ni-NTA affinity and size exclusion chromatography (Figure 4.2).

4.3 Autophosphorylation and Phosphotransfer of EnvZ and Misticated EnvZ in vitro.

[γ-32P] ATP kinase assay was used to test EnvZ and Misticated EnvZ's ability to autophosphorylate *in vitro*. Here purified EnvZ and Misticated EnvZ samples were incubated in the presence of Mg^{2+} and [γ-32P] ATP. Autoradiography was performed after running samples on a 10 % acrylamide gel. The soluble cytoplasmic domain

demonstrated the ability to autophosphorylate in the absence of the sensor and transmembrane domains and confirmed the location of autophosphorylation as previously shown (*17, 25*), (Figure 4.3, lane 1).

Figure 4.3, lane 3 and 9, illustrate that both purified full-length EnvZ and Misticated EnvZ, when solubilized in FC-12, are also able to autophosphorylate. To demonstrate that this autophosphorylation event occurs on the predicted site of phosphorylation, His_{243}, the point mutant H243V was created which has been previously shown to knock out kinase activity (*17*). Figure 4.3 (lanes 6 and 12) shows that no autophosphorylation takes place for EnvZ H243V, indicating that the autophosphorylation is dependent on residue His_{243}.

EnvZ's ability to phosphotransfer to its cognate response regulator OmpR can also be detected using this method. When solubilized in FC-12, both EnvZ and Misticated EnvZ were incubated in the presence of OmpR and [γ-32P] ATP and exhibited the ability to autophosphorylate and phosphotransfer to OmpR (Figure 4.3, lanes 3-4, 9-10). When EnvZ H243V and Misticated EnvZ H243V were incubated in the presence of OmpR, no phosphotransfer (Figure 4.3, lanes 7 and 13) took place, demonstrating the dependence of these events on the initial autophosphorylation of EnvZ His_{243}. To determine if phosphorylation of OmpR is dependent on EnvZ, we repeated the assay in the absence of EnvZ and Misticated EnvZ. No phosphorylation was dectected, thus illustrating its dependence on the histidine kinase (Figure 4.3, lane 2). To confirm the phosphorylation site, Asp_{55} of OmpR, we created the point mutant D55Q which was previously described to knock out phosphorylation (*26*). Figure 4.3 (lanes 5 and 11) shows that EnvZ and Misticated EnvZ are not able to phosphotransfer to OmpR D55Q,

confirming its residue specificity for residue Asp_{55}. The reaction mixture alone, in the absence of EnvZ, does not phosphorylate OmpR D55Q non-specifically (Figure 4.3, lane 8).

In this kinase assay there appears around a 2-fold difference in intensity between equivalent bands in lanes 4 and 10 (Figure 4.3), but the relative ratio between EnvZ and OmpR bands in each lane remains largely similar. To explain the intensity difference between equivalent bands between lanes, there can be simply a minor calibration error due to the varying amounts of protein loaded on the gel. More importantly, however, this may be due to continuous phosphorylation on the kinase and phospho-transfer to OmpR, then concurrent dephosphorylation, which eventually depletes phosphorus source in the reaction mixture. Dephosphorylation occurs only in the EnvZ-OmpR complex, not on EnvZ alone, thus autophosphorylation by EnvZ in the absence of OmpR can accumulate its labeling. So there is a certain equilibrium that must be considered to trap its maximal labeling. The reaction time was optimized to show the most phospho-labeling for the phosphotransfer experiments. The Misticated EnvZ and OmpR phosphotransfer experiment as well as the wild-type EnvZ and OmpR experiment were ran for the same lengths of time.

4.4 β-galactosidase Assay Illustrates EnvZ and Misticated EnvZ Signaling in vivo

The ability of EnvZ and Misticated EnvZ to autophosphorylate and phosphotransfer *in vitro* suggests that the CA domain and DHp domain are properly oriented allowing for such activities to take place. However, this does not give insight into the functionality of the periplasmic sensor domain, transmembrane domains, or the

activity of all domains of EnvZ and Misticated EnvZ as a whole. In order to analyze the activity of all the domains of EnvZ and to look at EnvZ's ability to signal downstream to the level of inducing porin expression, an *in vivo* β-galactosidase assay was performed. Various EnvZ constructs were transformed into two different *E. coli* strains: RU1012 [Φ(*ompC-lacZ*)*10-15*, Δ *envZ*::Km^r] (*21*) and MC4100 (*lac*$^-$). In this assay we electroporated cells with full-length EnvZ, Misticated EnvZ, EnvZ H243V, Misticated EnvZ H243V, and Misticated KvPae, a voltage gated K^+ Channel-like protein from *Pseudomonas aeruginosa*, as a negative control. In addition, cells were tested in the absence of vector, as an additional negative control. Cells were grown to mid-log phase at 37°C. 10 ml aliquots were harvested before the cells reached mid-log phase, and 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 20 h post mid-log phase. The samples were centrifuged, frozen, and the Miller Assay was completed on all samples.

The MC4100 (*lac*$^-$) *E. coli* cell strain was used as a control to ensure that any activity seen in the Miller Assay was due to the production of β-gal. When the Miller Assay was completed there was no significant β-gal activity, measured in Miller Units, from any of the six samples (Figure 4.4 A). The use of the RU1012 [Φ(*ompC-lacZ*)*10-15*, Δ *envZ*::Km^r] (*21*) strain allows for the measurement of *ompC* gene expression as the consequence of the downstream signal transduction events of EnvZ and Misticated EnvZ. EnvZ- and Misticated EnvZ-transformed cells showed activity before mid-log phase with a rise in activity up to 3 h post mid-log phase and continuing up to 20 h post mid-log phase. The negative control which lacked a transformed vector exhibited a low level of β-gal activity starting at 2 h post mid-log phase in comparison to EnvZ and Misticated

EnvZ samples. Null mutants EnvZ H243V and Misticated EnvZ H243V exhibited the lowest amount of activity than any other samples (Figure 4.4 B).

4.5 Discussion

In this study, we provide direct evidence that Mistic-fused full-length EnvZ is active both *in vitro* and *in vivo*. By means of the *in vitro* [γ-32P] ATP kinase assay, we illustrate that EnvZ and Misticated EnvZ autophosphorylate residue His_{243}, the conserved site of phosphorylation. In addition, we show that EnvZ and Misticated EnvZ are both able to phosphotransfer the phosphoryl group from EnvZ His_{243} to OmpR $Asp_{55,}$ in a site specific manner. Results from these assays demonstrate that both full-length constructs solubilized in the presence of the detergent FC-12 have a properly folded and functional cytoplasmic domain.

To further address the functionality of the whole receptor as a transmembrane signaling molecule, we performed an *in vivo* β-galactosidase assay using RU1012 *E. coli* cells, where EnvZ and Misticated EnvZ were tested to determine the activation of *ompC-lacZ* gene expression through the binding of OmpR to the *ompC* promoter. The results from this experiment illustrate that both EnvZ and Misticated EnvZ are active in their natural cell environment and are able to transduce a downstream signal such that the *ompC* promoter becomes activated. The lowest β-gal activity was found in the H243V samples and might be caused by the physical presence of the null EnvZ receptor. The expression of the non-functional EnvZ could possibly interfere or turn off alternative pathways subsequently inhibiting the otherwise recoverable *ompC* gene expression. When the negative control was tested in the absence of vector, a small rise in activity was

seen 2 h post mid-log phase, which could be explained by the complexity of the gene regulation system of the major *E. coli* outer membrane porins.

Since OmpC is one of the major *E. coli* outer membrane porins under complex gene regulation (*27*), the slight rise in β-gal activity seen in our negative control of the *in vivo* β-gal assay, could be due to the interference of other pathways attempting to compensate for the absence of EnvZ. Both major *E. coli* outer membrane porins, OmpC and OmpF, are under the control of a very intricate regulatory system comprised of many components within the cell including sRNAs such as MicF (*28-31*), MicC (*32*), RseX (*33, 34*), RybB (*35, 36*), and Ipex (*34, 37, 38*) which function by forming base pairs with their target mRNAs in the translation start site region and thus prevent translation. There are also numerous indirect/direct protein regulators some of which include: Rob, SoxS, MarA, CpxR, Lrp, HU, IHF, and H-NS (Figure 4.5). Belonging to the AraC/XylS family of transcriptional regulators are Rob, SoxS, and MarA, and they function by repressing OmpF expression through activation of *micF* transcription (*30, 39-48*). The histidine kinase receptor CpxA responds to different stimuli some of which include misfolded proteins and alkaline pH and functions in conjunction with the response regulator CpxR to positively and negatively regulate *ompC* and *ompF*, respectively (*49-51*). The activity of the regulator Lrp increases when the cell is exposed to conditions of limited accessibility to nutrients like that of minimal medium, where this protein negatively regulates *ompC* and positively regulates *ompF* through repression of *micF* (*52, 53*). Histone-like proteins such as HU, IHF, and H-NS are also involved in the regulation of OmpC and OmpF. The nucleoid protein HU partakes in porin regulation through its involvement in a pathway which decreases OmpF levels through regulation of *micF*

expression (*54*). IHF is a DNA-binding protein that functions by not only negatively regulating the *ompR-envZ* operon but also by negatively regulating both *ompC* and *ompF* by binding near their promoter region (*55-58*). The histone-like protein H-NS plays a role through repression of *ompC* and affects OmpF expression through the regulation of *micF* (*59, 60*). Due to the complexity of porin regulation it is possible that one of these other pathways may take over the regulation of the *ompC* gene when EnvZ is absent.

Since there is an increasing demand to overcome the difficulties facing the structural studies of integral membrane proteins, biochemists and structural biologists have looked into alternative modes to not only increase the expression level of integral membrane proteins, but also to ensure that these proteins are functionally active. We describe in this study that the Mistic-fusion system provided one such alternative, by not only increasing the expression level of EnvZ but also preserving its functional activity both *in vitro* and *in vivo*.

4.6 Materials and Methods

Strains and Plasmids:

All vector construction was Gateway (Invitrogen)- adapted. For the [γ-32P] ATP Kinase Assay, Mistic was fused to the NH_2–terminus of all targets (referred to as 'Misticated'), following an NH_2-terminal octahistidine-tag in the Gateway-adapted vector, pMis3.0E (*5*). The genes that were not Misticated were placed in frame in a Gateway-adapted NH_2-terminal nonylhistidine-tagged vector modified from pET28, (referred to as 'pHis9', non-Misticated). A thrombin cleavage site is present between the histidine tag and the target protein on each construct. The Gateway vector pDEST17 (Invitrogen) was used for expression of non-Misticated targets in the β-galactosidase assay to keep the antibiotic resistance (Amp) consistent between Misticated and non-Misticated samples.

E. coli BL21 (DE3) cells (Invitrogen) were used for expression of all samples. Experimental *E. coli* RU1012 [Φ(*ompC-lacZ)10-15*, Δ*envZ::*Km^r] cells (courtesy of Dr. Masayori Inouye) and control *E. coli* MC4100 cells (courtesy of Dr. Kit Pogliano) were used for the β-galactosidase assay.

Expression and Membrane Isolation

Recombinant vectors were used to transform *E. coli* BL21 (DE3) cells (Invitrogen). A 5 ml overnight culture was used to inoculate 1 L of Terrific Broth (EMD) at a 1:1000 ratio. Cells were grown at 37°C to OD_{600}= 1. Temperature was decreased to 18°C and 0.5 mM IPTG was added to induce expression of EnvZ constructs. For the soluble proteins OmpR and EnvZ cytoplasmic domain, cells were grown at 37°C to

OD_{600}= 0.4. 1 mM IPTG was added to induce expression, the temperature was kept at 37°C and cells were harvested 3 h later.

The cell pellet was weighed and the lysis buffer was added at 4X the weight of the cell pellet (20 mM Tris pH 8.0, 200 mM NaCl, 10 mM EDTA, 5 mM PMSF). The pellet was resuspended, then 5 mM β-Me and 1 mg/ml lysozyme was added and the sample was stirred at 4°C for 30 m. Cells were further lysed by sonication 3X on ice for a total of 1 m, pulses at 1 s on and 2 s off, in volumes of no more than 40 ml at a time. The sample was then centrifuged at 100,000 x g for 2 h. The pellet was resuspended in lysis buffer and centrifuged at 10,000 x g for 20 m. The supernatant was collected and centrifuged at 100,000 x g for 2 h. The membrane pellets were then re-suspended in cold salt wash buffer (20 mM Tris pH 8.0, 500 mM NaCl, 5 mM βMe, 5 mM PMSF, 10 mM EDTA) and stirred overnight at 4°C. The next day, the salt-washed membranes were centrifuged at 100,000 x g for 2 h, the pellet was then re-suspended in cold storage buffer (20mM Tris pH 8.0, 0.1 M NaCl, 5 mM βMe, 20 % v/v glycerol & protease inhibitor cocktail tablets (Roche)). The homogenous membrane mixture was then aliquoted and frozen at -80°C.

EnvZ Purification

Membranes were solubilized in solubilization buffer (20 mM Tris pH 8.0, 20 mM FC-12, 0.3 M NaCl, 1 mM $MgCl_2$, 5 mM βMe) and stirred overnight at 4°C. Solubilized membranes were centrifuged 100,000 x g 2 h. The protein was purified on a Ni-NTA column (Qiagen) and the detergent was exchanged by washing with Wash buffer (20 mM Tris pH 8.0, 0.2 M NaCl, 4 mM FC-12, 10 mM imidazole, 3 mM βMe). The protein was

eluted with Elution buffer (20 mM Tris pH 8.0, 0.2 M NaCl, 4 mM FC-12, 0.3 M imidazole, 3 mM βMe) and concentrated to 2 ml using a Vivaspin concentrator and injected on a S200 16/60 size exclusion column (Pharmacia) with FPLC buffer (20 mM Tris pH 8.0, 0.2 M NaCl, 1 mM DTT, 1 mM EDTA, 2 mM FC-12). To digest with thrombin, a 1:200 molar ratio of thrombin:protein was added and the sample was dialyzed overnight at 4°C, post Ni-NTA purification. Following cleavage on the next day, the thrombin and uncleaved protein were removed by purification on a benzamidine and Ni-NTA column, the protein was concentrated to 2 ml and purified via size exclusion chromatography as stated above.

OmpR and EnvZ Cytoplasmic Domain Purification

Cells were lysed as stated above using Lysis buffer (20 mM Tris pH 8.8, 0.3 M NaCl, 1 mM imidazole, 5 mM PMSF, 5 mM βMe). Protein was purified on a Ni-NTA column (Qiagen), washed with Wash buffer (20 mM Tris pH 8.8, 0.3 M NaCl, 20 mM imdazole), and eluted with Elution buffer (20 mM Tris pH 8.8, 0.3 M NaCl, 0.25 M imidazole, 2 mM $CaCl_2$). Thrombin (Sigma) was added with a 1:2000 dilution, and dialyzed overnight 4°C, 3500 MWCO tubing in dialysis buffer (20 mM Tris pH 8.8, 0.3 M NaCl, 2.5 mM $CaCl_2$). After thrombin cleavage, the protein was purified on Ni-NTA and benzamidine resin to remove uncleaved protein and thrombin. The sample was concentrated in a Vivaspin concentrator and injected on a S200 16/60 size exclusion column (Pharmacia) with FPLC buffer (20 mM Tris pH 8.8, 0.2 M NaCl, 5 mM DTT, 1 mM EDTA).

[γ-32P] ATP Kinase Assay

1 μM of protein was added to the reaction mixture (100 mM Tris pH 8.0, 100 mM KCl, 10 mM $MgCl_2$, 10% glycerol) along with 20 μM cold ATP, 10 μCi [γ ^{32}P] ATP in a total volume of 20 μl, and incubated at room temperature for 15 m. For EnvZ/OmpR phosphotransfer, the complex was incubated at room temperature for 10 m before and after addition of ATP. 10 μl of 2X SDS PAGE sample buffer was added to stop the reaction. Samples were heated in a 95°C water bath for 2 m and loaded on a 10% SDS polyacrylamide gel along with the Biorad precision plus pre-stained molecular weight marker. The gel was incubated with amberlite cation/anion exchange resin (Polysciences, Inc., polylite MB-3) to absorb free [γ ^{32}P] ATP, dried, then exposed to KODAK BioMax XAR film for analysis.

β-galactosidase Assay

RU1012 cells or MC4100 cells were electroporated with the recombinant plasmid of choice and spread. The next day colonies were picked and a 5 ml overnight culture was made. Cells were diluted 1:100 into 150 ml Loria Broth (EMD). Cells were then grown to mid-log phase and induced with 0.5 mM IPTG. 10 ml aliquots were taken out before induction and 0.5 h, 1 h, 1.5 h, 2 h, 2.5 h, 3 h, and 20 h post induction, harvested and frozen in -80°C.

Cells were thawed and resuspended in chilled Z buffer (0.06 M $Na_2P0_4{\cdot}7H_2O$, 0.04 M $NaH_2PO_4{\cdot}H_20$, 0.01 M KCl, 0.001 M $MgSO_4$, 0.05 M βMe, pH 7.0) and normalized to 0.2 OD_{600}. A fraction of cells were taken out and diluted to about 1:2 with Z buffer for a total of 1 ml, if this ratio did not yield a sufficient yellow color, the ratio

was changed by the addition of more cells (or less cells if sample turned yellow too fast). The cells were permeabilized by the addition of 100 µl chloroform and 50 µl 0.1% SDS, then vortexed and equilibrated for 5 m in a 28°C water bath. The reaction was started by the addition of 0.2 ml 4 mg/ml ONPG, followed by incubation at 28°C. The reaction was terminated by the addition of 0.5 ml 1 M Na_2CO_3 once a sufficient yellow color developed. Cells were centrifuged at 17,000 x g for 5 m to remove chloroform and cell debris. OD_{420} and OD_{550} was recorded for all samples to calculate the units of activity (*61*).

4.7 Figures

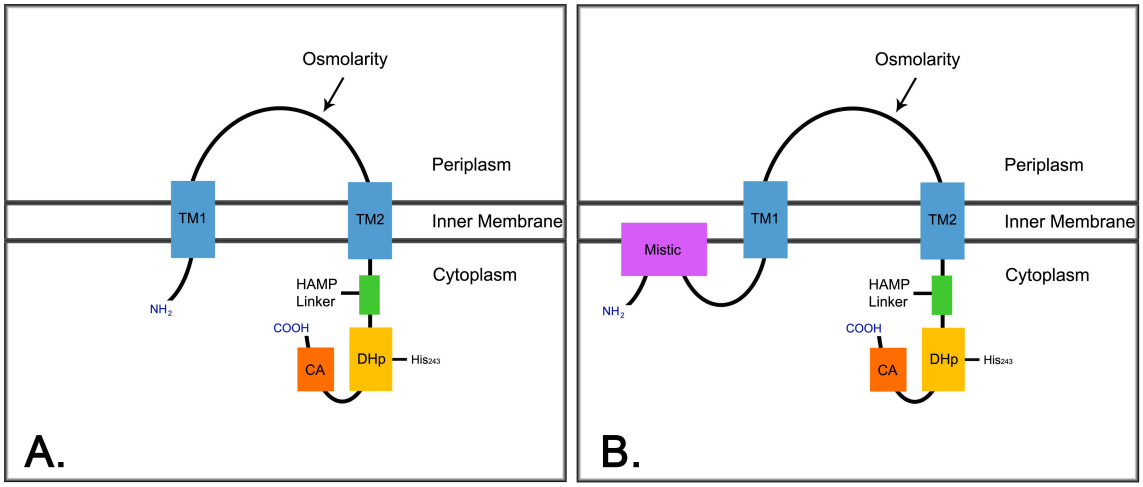

Figure 4.1: Schematic Representation of EnvZ's Domain Organization.
(A) EnvZ is composed of two-transmembrane domains (blue), a periplasmic sensor domain, and a cytoplasmic domain composed of a HAMP Linker (green), DHp (gold), and CA domain (orange). (B) EnvZ domain organization when Mistic is NH2-terminally fused.

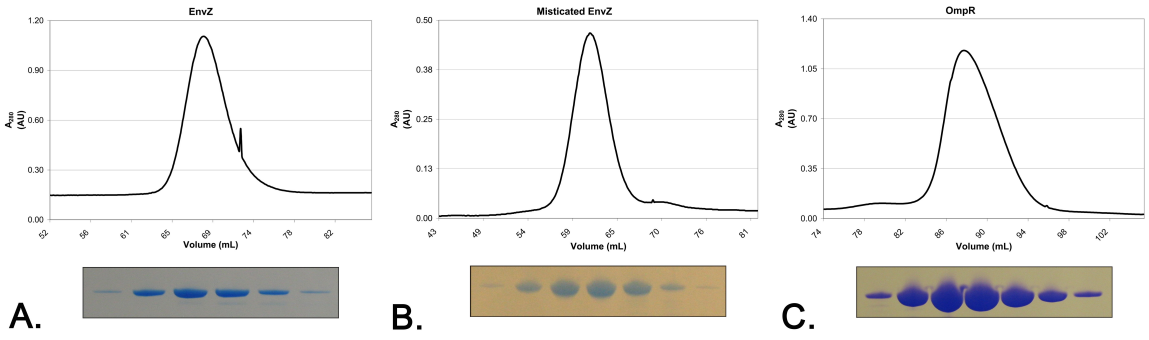

Figure 4.2: Purification of EnvZ, Misticated-EnvZ, and OmpR.
Size exclusion chromatogram profiles of EnvZ (A), Misticated EnvZ (B), and OmpR (C) from a FPLC run on a S200 16/60 column, post Ni-NTA affinity chromatography. Coomassie stained SDS-PAGE gels of the peak fractions are shown below each chromatogram.

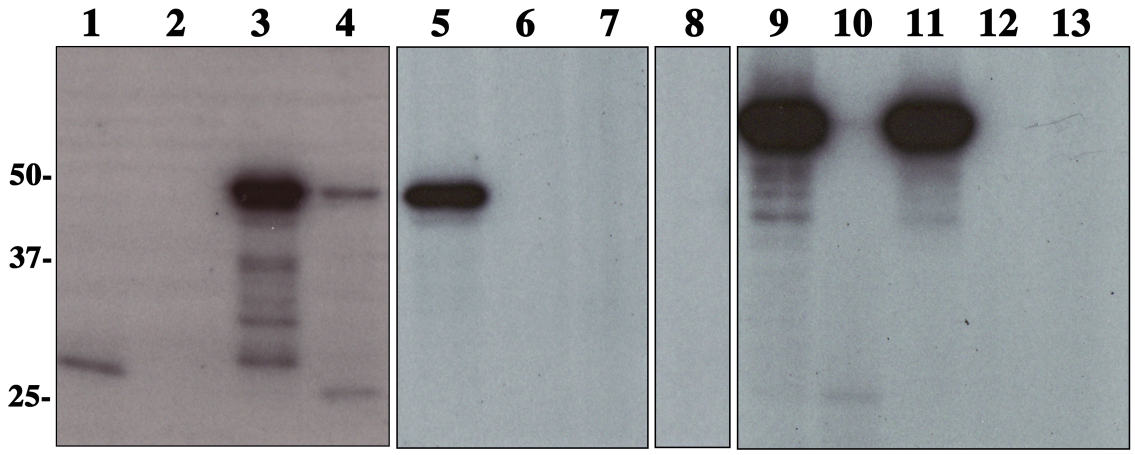

Figure 4.3: [γ-32P] ATP Kinase Assay Detecting Autophosphorylation and Phosphotransfer of EnvZ and Misticated EnvZ.
Autoradiogram of samples which were incubated in the presence of [γ-32P] ATP as described previously. Lane 1: EnvZ cytoplasmic domain; Lane 2: OmpR; Lane 3: EnvZ; Lane 4: EnvZ (upper band) and OmpR (lower band); Lane 5: EnvZ and OmpR D55Q; Lane 6: EnvZ H243V; Lane 7: EnvZ H243V and OmpR; Lane 8: OmpR D55Q; Lane 9: Misticated EnvZ; Lane 10: Misticated EnvZ (upper band) and OmpR (lower band); Lane 11: Misticated EnvZ and OmpR D55Q; Lane 12: Misticated EnvZ H243V; Lane 13: Misticated EnvZ H243V and OmpR.

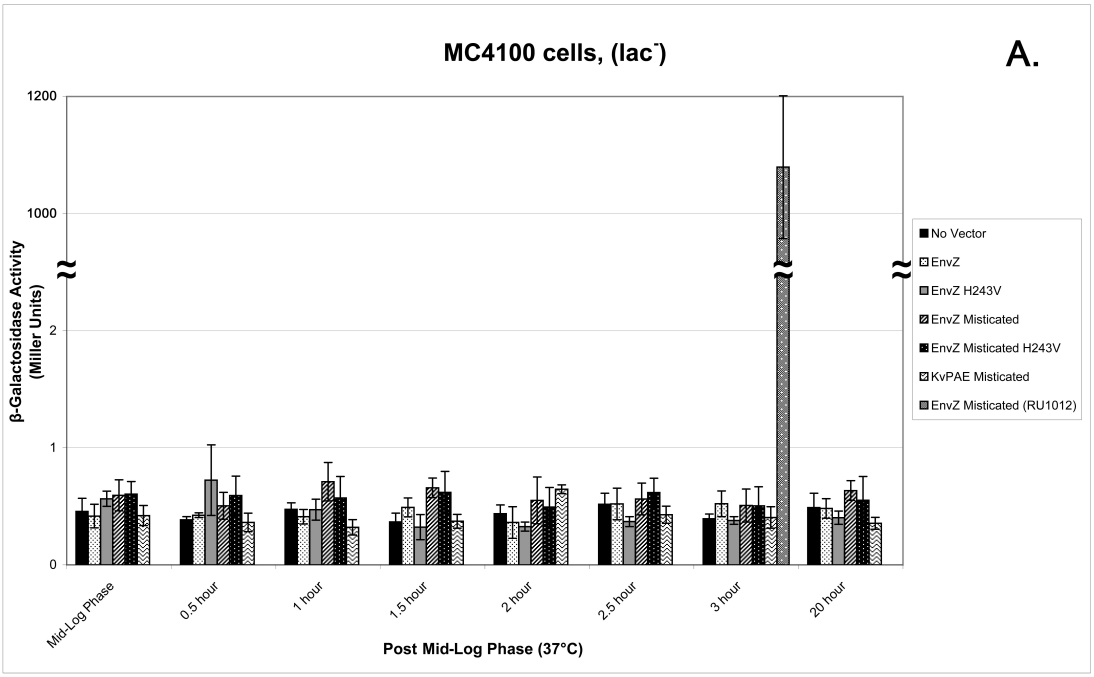

Figure 4.4: β-Galactosidase Assay using the MC4100 and RU1012 *E. coli* strains illustrating EnvZ and Misticated EnvZ Activity.
(A) The MC4100 *E. coli* strain illustrating that any β-galactosidase activity (in Miller Units) is a result of the production of β-galactosidase produced by the lac-Z reporter gene. The following samples were tested: no vector control, EnvZ, EnvZ H243V mutant, Mistcated EnvZ, Misticated EnvZ H243V, and Misticated KvPAE. 10 ml aliquots of cells were harvested and frozen 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 20 h post mid log phase. The Miller Assay was completed when all aliquots were collected.

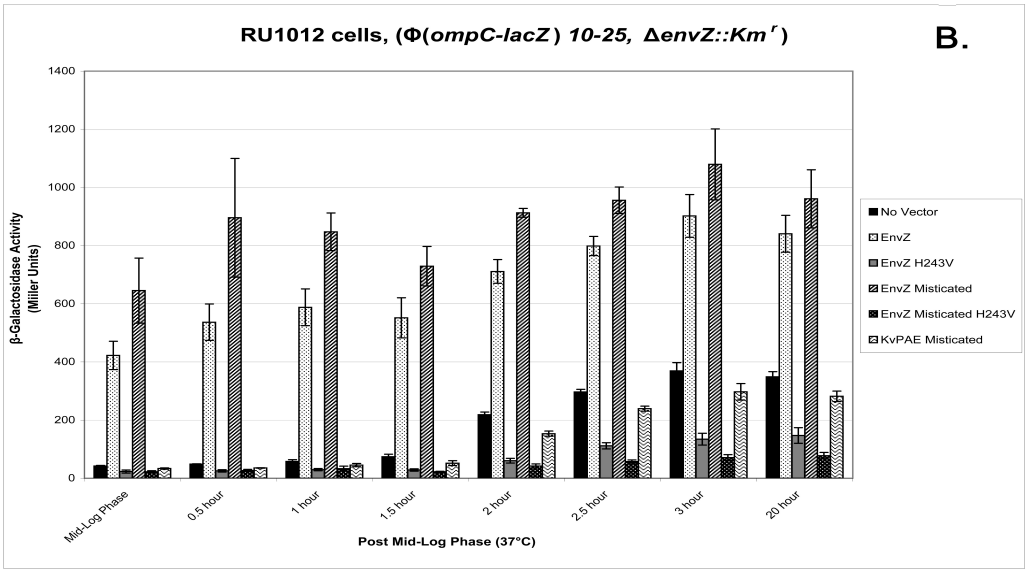

Figure 4.4: β-Galactosidase Assay using the MC4100 and RU1012 *E. coli* strains illustrating EnvZ and Misticated EnvZ Activity.
(B) The experimental RU1012 [Φ(*ompC-lacZ*)*10-15*, Δ *envZ*::Km^r] *E. coli* strain was tested to measure downstream signaling of EnvZ and Misticated-EnvZ as a result of β-galactosidase activity (in Miller Units). The following samples were tested: no vector control, EnvZ, EnvZ H243V mutant, Mistcated EnvZ, Misticated EnvZ H243V, and Misticated KvPAE. 10 ml aliquots of cells were harvested and frozen 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 20 h post mid log phase. The Miller Assay was completed when all aliquots were collected.

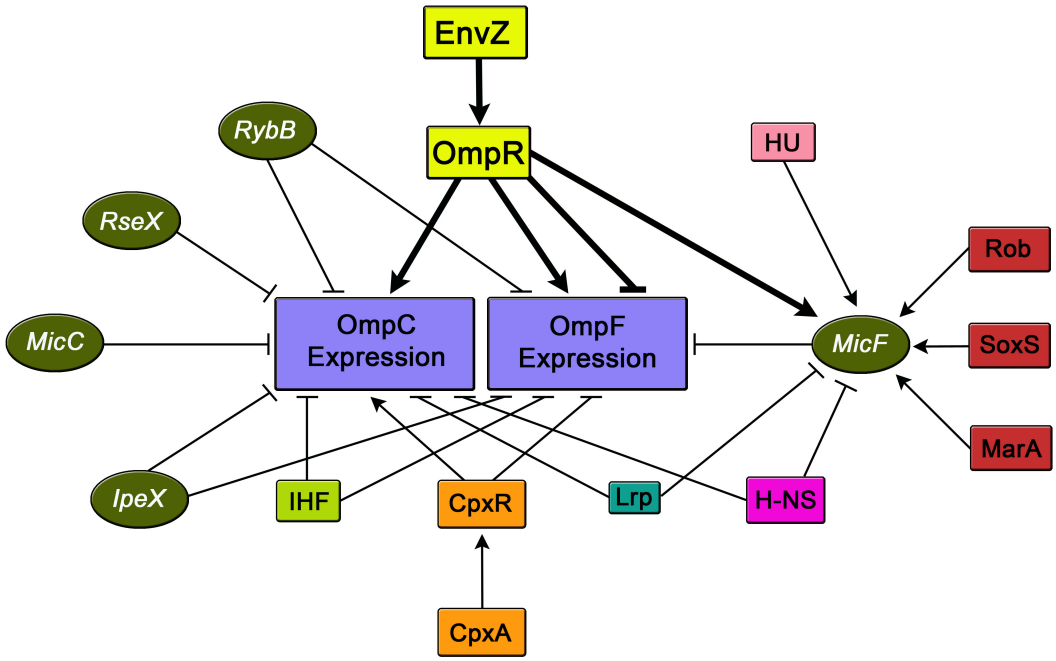

Figure 4.5: The Complexity of OmpC and OmpF Porin Expression Regulation. These are a number of the several different factors within the cell which are involved in the regulation of OmpC and OmpF porin expression. Various constituents illustrated include regulatory sRNAs (green circles), indirect and direct protein regulators (multi-colored boxes). The → symbol represents gene activation and ⊣ represents gene repression.

4.8 Acknowledgements

This work was supported by grants from the National Institutes of Health GM74929 (S.C.). K.Y.B. acknowledges the fellowship support from the American Heart Association and from the H. A. and Mary K. Chapman Charitable Trust and the Mary K. Chapman Foundation. We would like to thank Tony Hunter and Jill Meisenhelder from the Salk Institute for providing the facilities for, and the help with the [γ-32P] ATP kinase assay. We would like to thank Innokentiy Maslennikov, Georgia Kefala, Mizuki Okamura, Luis Esquivies and Chris Dickson for their help during discussions. We thank M. Inouye for the gift of the RU1012 strain, and K. Pogliano for the gift of the MC4100 strain. Chapter 4, in part, is a reprint of the material as it appears in *Biochemistry* 2010. Blain, K.Y., Kwiatkowski, W., Choe, S. "The functionally active Mistic-fused histidine kinase receptor, EnvZ." *Biochemistry*. 2010; 49 (42): 9089-9095. The dissertation author was the primary investigator and author of this paper.

4.9 References

1. Arkin, I. T., Brunger, A. T., and Engelman, D. M. (1997) Are there dominant membrane protein families with a given number of helices?, *Proteins 28*, 465-466.

2. Wallin, E., and von Heijne, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms, *Protein Sci 7*, 1029-1038.

3. White, S. H., and Wimley, W. C. (1999) Membrane protein folding and stability: physical principles, *Annu Rev Biophys Biomol Struct 28*, 319-365.

4. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, R. D., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J. F., Dougherty, B. A., Bott, K. F., Hu, P. C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison, C. A., 3rd, and Venter, J. C. (1995) The minimal gene complement of Mycoplasma genitalium, *Science 270*, 397-403.

5. Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R., and Choe, S. (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression, *Science 307*, 1317-1321.

6. Roosild, T. P., Vega, M., Castronovo, S., and Choe, S. (2006) Characterization of the family of Mistic homologues, *BMC Struct Biol 6*, 10.

7. Kefala, G., Kwiatkowski, W., Esquivies, L., Maslennikov, I., and Choe, S. (2007) Application of Mistic to improving the expression and membrane integration of histidine kinase receptors from Escherichia coli, *J Struct Funct Genomics 8*, 167-172.

8. Hoch, J. A., and Silhavy, T. J. (1995) *Two-Component Signal Transduction*, ASM Press, Washington, DC.

9. Egger, L. A., Park, H., and Inouye, M. (1997) Signal transduction via the histidyl-aspartyl phosphorelay, *Genes Cells 2*, 167-184.

10. Forst, S. A., and Roberts, D. L. (1994) Signal transduction by the EnvZ-OmpR phosphotransfer system in bacteria, *Res Microbiol 145*, 363-373.

11. Forst, S., Comeau, D., Norioka, S., and Inouye, M. (1987) Localization and membrane topology of EnvZ, a protein involved in osmoregulation of OmpF and OmpC in Escherichia coli, *J Biol Chem 262*, 16433-16438.

12. Park, H., and Inouye, M. (1997) Mutational analysis of the linker region of EnvZ, an osmosensor in Escherichia coli, *J Bacteriol 179*, 4382-4390.

13. Park, H., Saha, S. K., and Inouye, M. (1998) Two-domain reconstitution of a functional protein histidine kinase, *Proc Natl Acad Sci U S A 95*, 6728-6732.

14. Dutta, R., Qin, L., and Inouye, M. (1999) Histidine kinases: diversity of domain organization, *Mol Microbiol 34*, 633-640.

15. Roberts, D. L., Bennett, D. W., and Forst, S. A. (1994) Identification of the site of phosphorylation on the osmosensor, EnvZ, of Escherichia coli, *J Biol Chem 269*, 8728-8733.

16. Igo, M. M., Ninfa, A. J., Stock, J. B., and Silhavy, T. J. (1989) Phosphorylation and dephosphorylation of a bacterial transcriptional activator by a transmembrane receptor, *Genes Dev 3*, 1725-1734.

17. Forst, S., Delgado, J., and Inouye, M. (1989) Phosphorylation of OmpR by the osmosensor EnvZ modulates expression of the ompF and ompC genes in Escherichia coli, *Proc Natl Acad Sci U S A 86*, 6052-6056.

18. Aiba, H., Mizuno, T., and Mizushima, S. (1989) Transfer of phosphoryl group between two regulatory proteins involved in osmoregulatory expression of the ompF and ompC genes in Escherichia coli, *J Biol Chem 264*, 8563-8567.

19. Tokishita, S., Yamada, H., Aiba, H., and Mizuno, T. (1990) Transmembrane signal transduction and osmoregulation in Escherichia coli: II. The osmotic sensor, EnvZ, located in the isolated cytoplasmic membrane displays its phosphorylation and dephosphorylation abilities as to the activator protein, OmpR, *J Biochem 108*, 488-493.

20. Xie, W., Blain, K. Y., Kuo, M. M., and Choe, S. (2010) Protein Engineering of Bacterial Histidine Kinase Receptor Systems, *Protein Pept Lett 17*, 867-873.

21. Utsumi, R., Brissette, R. E., Rampersaud, A., Forst, S. A., Oosawa, K., and Inouye, M. (1989) Activation of bacterial porin gene expression by a chimeric signal transducer in response to aspartate, *Science 245*, 1246-1249.

22. Zhu, Y., and Inouye, M. (2003) Analysis of the role of the EnvZ linker region in signal transduction using a chimeric Tar/EnvZ receptor protein, Tez1, *J Biol Chem 278*, 22812-22819.

23. Baumgartner, J. W., Kim, C., Brissette, R. E., Inouye, M., Park, C., and Hazelbauer, G. L. (1994) Transmembrane signalling by a hybrid protein:

communication from the domain of chemoreceptor Trg that recognizes sugar-binding proteins to the kinase/phosphatase domain of osmosensor EnvZ, *J Bacteriol 176*, 1157-1163.

24. Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy, M., Davidson, E. A., Scouras, A., Ellington, A. D., Marcotte, E. M., and Voigt, C. A. (2005) Synthetic biology: engineering Escherichia coli to see light, *Nature 438*, 441-442.

25. Igo, M. M., and Silhavy, T. J. (1988) EnvZ, a transmembrane environmental sensor of Escherichia coli K-12, is phosphorylated in vitro, *J Bacteriol 170*, 5971-5973.

26. Kanamaru, K., Aiba, H., and Mizuno, T. (1990) Transmembrane signal transduction and osmoregulation in Escherichia coli: I. Analysis by site-directed mutagenesis of the amino acid residues involved in phosphotransfer between the two regulatory components, EnvZ and OmpR, *J Biochem 108*, 483-487.

27. Batchelor, E., and Goulian, M. (2006) Imaging OmpR localization in Escherichia coli, *Mol Microbiol 59*, 1767-1778.

28. Mizuno, T., Chou, M. Y., and Inouye, M. (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA), *Proc Natl Acad Sci U S A 81*, 1966-1970.

29. Schmidt, M., Zheng, P., and Delihas, N. (1995) Secondary structures of Escherichia coli antisense micF RNA, the 5'-end of the target ompF mRNA, and the RNA/RNA duplex, *Biochemistry 34*, 3621-3631.

30. Delihas, N., and Forst, S. (2001) MicF: an antisense RNA gene involved in response of Escherichia coli to global stress factors, *J Mol Biol 313*, 1-12.

31. Coyer, J., Andersen, J., Forst, S. A., Inouye, M., and Delihas, N. (1990) micF RNA in ompB mutants of Escherichia coli: different pathways regulate micF RNA levels in response to osmolarity and temperature change, *J Bacteriol 172*, 4143-4150.

32. Chen, S., Zhang, A., Blyn, L. B., and Storz, G. (2004) MicC, a second small-RNA regulator of Omp protein expression in Escherichia coli, *J Bacteriol 186*, 6689-6697.

33. Fernandez-Mora, M., Oropeza, R., Puente, J. L., and Calva, E. (1995) Isolation and characterization of ompS1, a novel Salmonella typhi outer membrane protein-encoding gene, *Gene 158*, 67-72.

34. Douchin, V., Bohn, C., and Bouloc, P. (2006) Down-regulation of porins by a small RNA bypasses the essentiality of the regulated intramembrane proteolysis protease RseP in Escherichia coli, *J Biol Chem 281*, 12253-12259.

35. Papenfort, K., Pfeiffer, V., Mika, F., Lucchini, S., Hinton, J. C., and Vogel, J. (2006) SigmaE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay, *Mol Microbiol 62*, 1674-1688.

36. Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001) Identification of novel small RNAs using comparative genomics and microarrays, *Genes Dev 15*, 1637-1651.

37. Pugsley, A. P., and Schnaitman, C. A. (1978) Identification of three genes controlling production of new outer membrane pore proteins in Escherichia coli K-12, *J Bacteriol 135*, 1118-1129.

38. Castillo-Keller, M., Vuong, P., and Misra, R. (2006) Novel mechanism of Escherichia coli porin regulation, *J Bacteriol 188*, 576-586.

39. Rosenberg, E. Y., Bertenthal, D., Nilles, M. L., Bertrand, K. P., and Nikaido, H. (2003) Bile salts and fatty acids induce the expression of Escherichia coli AcrAB multidrug efflux pump through their interaction with Rob regulatory protein, *Mol Microbiol 48*, 1609-1619.

40. Bennik, M. H., Pomposiello, P. J., Thorne, D. F., and Demple, B. (2000) Defining a rob regulon in Escherichia coli by using transposon mutagenesis, *J Bacteriol 182*, 3794-3801.

41. Rosner, J. L., Dangi, B., Gronenborn, A. M., and Martin, R. G. (2002) Posttranscriptional activation of the transcriptional activator Rob by dipyridyl in Escherichia coli, *J Bacteriol 184*, 1407-1416.

42. Li, Z., and Demple, B. (1994) SoxS, an activator of superoxide stress genes in Escherichia coli. Purification and interaction with DNA, *J Biol Chem 269*, 18371-18377.

43. Gil, F., Hernandez-Lucas, I., Polanco, R., Pacheco, N., Collao, B., Villarreal, J. M., Nardocci, G., Calva, E., and Saavedra, C. P. (2009) SoxS regulates the expression of the Salmonella enterica serovar Typhimurium ompW gene, *Microbiology 155*, 2490-2497.

44. Gallegos, M. T., Schleif, R., Bairoch, A., Hofmann, K., and Ramos, J. L. (1997) Arac/XylS family of transcriptional regulators, *Microbiol Mol Biol Rev 61*, 393-410.

45. Miller, P. F., and Sulavik, M. C. (1996) Overlaps and parallels in the regulation of intrinsic multiple-antibiotic resistance in Escherichia coli, *Mol Microbiol 21*, 441-448.

46. Balague, C., and Vescovi, E. G. (2001) Activation of multiple antibiotic resistance in uropathogenic Escherichia coli strains by aryloxoalcanoic acid compounds, *Antimicrob Agents Chemother 45*, 1815-1822.

47. Cohen, S. P., Hachler, H., and Levy, S. B. (1993) Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in Escherichia coli, *J Bacteriol 175*, 1484-1492.

48. Hachler, H., Cohen, S. P., and Levy, S. B. (1991) marA, a regulated locus which controls expression of chromosomal multiple antibiotic resistance in Escherichia coli, *J Bacteriol 173*, 5532-5538.

49. Batchelor, E., Walthers, D., Kenney, L. J., and Goulian, M. (2005) The Escherichia coli CpxA-CpxR envelope stress response system regulates expression of the porins ompF and ompC, *J Bacteriol 187*, 5723-5731.

50. Dorel, C., Lejeune, P., and Rodrigue, A. (2006) The Cpx system of Escherichia coli, a strategic signaling pathway for confronting adverse conditions and for settling biofilm communities?, *Res Microbiol 157*, 306-314.

51. Raivio, T. L. (2005) Envelope stress responses and Gram-negative bacterial pathogenesis, *Mol Microbiol 56*, 1119-1128.

52. Ferrario, M., Ernsting, B. R., Borst, D. W., Wiese, D. E., 2nd, Blumenthal, R. M., and Matthews, R. G. (1995) The leucine-responsive regulatory protein of Escherichia coli negatively regulates transcription of ompC and micF and positively regulates translation of ompF, *J Bacteriol 177*, 103-113.

53. Calvo, J. M., and Matthews, R. G. (1994) The leucine-responsive regulatory protein, a global regulator of metabolism in Escherichia coli, *Microbiol Rev 58*, 466-490.

54. Painbeni, E., Caroff, M., and Rouviere-Yaniv, J. (1997) Alterations of the outer membrane composition in Escherichia coli lacking the histone-like protein HU, *Proc Natl Acad Sci U S A 94*, 6712-6717.

55. Tsui, P., Helu, V., and Freundlich, M. (1988) Altered osmoregulation of ompF in integration host factor mutants of Escherichia coli, *J Bacteriol 170*, 4950-4953.

56. Huang, L., Tsui, P., and Freundlich, M. (1990) Integration host factor is a negative effector of in vivo and in vitro expression of ompC in Escherichia coli, *J Bacteriol 172*, 5293-5298.

57. Ramani, N., Huang, L., and Freundlich, M. (1992) In vitro interactions of integration host factor with the ompF promoter-regulatory region of Escherichia coli, *Mol Gen Genet 231*, 248-255.

58. Tsui, P., Huang, L., and Freundlich, M. (1991) Integration host factor binds specifically to multiple sites in the ompB promoter of Escherichia coli and inhibits transcription, *J Bacteriol 173*, 5800-5807.

59. Deighan, P., Free, A., and Dorman, C. J. (2000) A role for the Escherichia coli H-NS-like protein StpA in OmpF porin expression through modulation of micF RNA stability, *Mol Microbiol 38*, 126-139.

60. Suzuki, T., Ueguchi, C., and Mizuno, T. (1996) H-NS regulates OmpF expression through micF antisense RNA in Escherichia coli, *J Bacteriol 178*, 3650-3653.

61. Miller, J. H. (1972) *Experiments in Molecular Genetics*, [Cold Spring Harbor, N.Y.] : Cold Spring Harbor Laboratory, 352-355.

CHAPTER 5:

Additional Structures Solved

hIMP3, GluN1, E4-ORF3

5.1 The NMR Solution Structure of Human Integral Membrane Protein 3

5.1.1 Introduction

The progression of structure determination of integral membrane proteins has been hindered by the complexities associated with membrane protein overexpression, solubilization, and purification. The ability to obtain quality X-ray crystal diffraction has proven difficult due to the disordered nature of the membrane protein crystals. Moreover, NMR spectra analysis has been hindered by signal broadening caused by intrinsic internal mobility of the helical membrane proteins and low overall tumbling rate of large protein detergent complexes, because of the need for detergent for protein solubilization. The recent progress of structure determination of membrane proteins by methods such as X-ray crystallography, NMR, electron diffraction and cryo-electron microscopy has been slow. Only 1288 transmembrane protein structures has been determined according to PDBTM (October 2010, http://pdbtm.enzim.hu/.) (*1, 2*), and 1066 membrane protein structures solved according to Membrane Protein Data Bank (MPDB, http://www.mpdb.tcd.ie/) (*3*). The numbers reported by those databases include whole membrane proteins, membrane protein fragments, and non-unique structures deposited to the PDB. These numbers correspond to less than 3% of all protein structures deposited into the PDB, and represent less than 1% of known structures of unique proteins. Where as an even smaller number of 261 unique whole membrane protein structures have been determined according to the Stephen White laboratory database (October 2010, http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html). Although the number reported by the Stephen White laboratory includes mostly crystal structures and excludes some NMR structures this number is still rather small in comparison to the total number

of protein structures that have been deposited into the Protein Data Bank, which is about 68,998 (November 2010). Due to these obstacles and the current need for new unique membrane protein structures, for a better biological and biochemical understanding of function and for structure based drug design, there is a large effort to find new methods to overcome these bottlenecks.

One such newly developed technique that can be utilized to overcome these difficulties is the cell-free combinatorial dual-labeling (CDL) strategy. The CDL strategy combines of the efficient and fast cell-free (CF) expression and rapid NMR analysis to quickly determine the structures of integral membrane proteins. This strategy has been previously used to determine the backbone structures, within 8 months, of the transmembrane domains of three classes of *E. coli* histidine kinase receptors, ArcB, QseC, and KdpD (*4*). To gain a better understanding of this technique and because my attempts at solving the solution structure of the 2 transmembrane domain and extracellular domain of the histidine kinase receptor EnvZ were unsuccessful, I decided to try and solve the structure of a human integral membrane protein, which due to our preparation for publication, I will refer to as human integral membrane protein 3 (hIMP3).

5.1.2 EnvZ's large periplasmic sensor domain and requirement for detergent provided obstacles in obtaining quality NMR spectra

The histidine kinase receptor EnvZ is a 450 amino acid integral membrane protein which was discussed in detail in chapters 3 and 4, and was my initial target in trying to

solve the solution structure by NMR. Since EnvZ in whole is about 50 kDa in size, we decided to truncate the protein and separate it into fragments that would be more likely to possess uniform structural properties and dynamics behavior, necessary for success at obtaining good quality NMR spectra. Since the NH_2-terminal portion of EnvZ is composed of an NH_2-terminal cytoplasmic tail (AA1-15), two transmembrane domains, TM1 (AA16-47) and TM2 (AA163-179), surrounding a periplasmic sensor domain (AA48-162) we decided to include these secondary structural elements into our construct. We thus created the boundary cut off at the charged residue of Glu190, which will result in a construct containing two transmembrane domains and the periplasmic sensor domain at a size of 22 kDa.

Multiple parameters were tested to obtain the highest quality NMR spectra possible. Detergent type and concentration were two of the first parameters to be tested and we found that EnvZ (1-190) behaved best in 50 mM FC-12 with a protein:detergent molar ratio of about 1:500 to 1:1000. Salt concentration was also adjusted and we found that removing the salt in the sample increases sensitivity and thus spectra quality. After screening temperature and pH we found that the lower pH of 4.8 and a temperature of 45°C also aided in obtaining a better spectra. We have also created multiple truncated constructs by varying the position of the COOH-terminal residue cut off such as creating EnvZ (1-180) and EnvZ (1-158), however these samples did not show an improvement.

Although we optimized the conditions for EnvZ (1-190) by testing multiple parameters to improve sample behavior and spectral data collection we were not able to obtain spectra of sufficient quality to begin assignment. If we increased the detergent

concentration, the peaks within the α-helical and transmembrane regions improved, but caused the peaks within the β-sheet region to disappear (Figure 5.1.1). However, if an insufficient amount of detergent is added to the sample, some peaks in the β-sheet region would appear, but the transmembrane domains appear to destabilize causing signal broadening within the α-helical region. The high detergent concentration could cause the soluble periplasmic domain to unfold, since we do not see the expected β-sheet cross-peak pattern as seen in previous NMR studies (*5*) on the periplasmic domain of EnvZ (region highlighted in blue Figure 5.1.1). Usually the dispersion and deviation from random coil values (about ~8.1-8.5 ppm range for ^{1}H) for β-sheets is much better, than for α-helices; however, the dispersion in our EnvZ (1-190) spectra is not enough to expect β-sheet structure. The EnvZ (1-190) spectra contain two types of signals, one being sharp and rounded (most likely the periplasmic domain) and one being broad and unresolved (most likely the transmembrane regions), which confirm non-uniform dynamics behavior of the periplasmic domain and transmembrane regions of EnvZ (1-190). Further improvement of the NMR parameters will need to be addressed to increase the quality of the spectra before assignment can be initiated.

5.1.3 Expression, Data Collection, and Assignment of hIMP3

hIMP3 was synthesized by utilizing the precipitating CF (p-CF) expression mode (*6*), where it was created as a precipitate that is subsequently solubilized by the lipid-like detergent LMPG. LMPG is a detergent of choice because (i) it successfully solubilizes most of the CF-expressed proteins from precipitant and (ii) it allows for the mobility of

the protein inside the micelle thus sharpening the resonance lines (*6, 7*). The CF expression system is very useful because it enables preparative scale expression of hIMPs in combination with unique isotopical labeling possibilities. Further, it does not require purification or the addition of affinity tags, and NMR studies can be completed directly after expression and solublization. One batch yielded about 3 mg after an overnight expression in 3 ml.

In order to speed up the NMR spectra assignment process that is completed by sequential assignment of backbone resonances we exploited the CDL strategy that was recently created for membrane proteins. With the use of the CDL strategy we designed a combinatorial selective ^{15}N- and 1-^{13}C-labeling scheme individually optimized for hIMP3 amino acid sequence. The scheme contains 6 samples with different sets of ^{15}N- and 13C-labeled amino acids. The combinatorial approach reduces the number of samples that are required to assign the second ^{15}N-labeled amino acid in every pair in a sequence, and optimization of the scheme for particular protein sequence minimizes the spectral complexity of every sample in the scheme. After 6 CDL samples were expressed using the p-CF system, [^{1}H-^{15}N]-TROSY-HSQC and ^{1}H-^{15}N plane of HNCO spectra were collected for each CDL sample. The spectra were then analyzed, and about 26 % of the protein backbone assignment was completed within a day for a majority of unique amino acid pairs in the sequence. These unambiguous assignments were used as valuable anchor points for further sequential assignment of the remaining resonances by analyzing 3D-HNCA, -HNCOCA and -HNCACB spectra.

For some strongly overlapping and thus remaining more challenging cross-peaks in the transmembrane regions we used single amino acid labeling in combination with point-directed assignment which follows the same idea as CDL strategy by just using one sample. Again, this is based on single amino-acid type selective dual-isotope (^{15}N-^{13}C) labeling of two types of amino acids that are located consecutively in a protein sequence (*8*). Cross-peaks in both the [^{15}N-^{1}H]-HSQC and the HNCO spectra will appear for every pair of residues in which the first amino acid is labeled with ^{13}CO and the second is labeled with ^{15}NH (Figure 5.1.2). A 3D NOESY spectrum with a mixing time of 120 ms was measured to facilitate sequential assignment and to obtain assignment of side chain Hα, Hβ and Hγ protons. Currently, we have approximately 71 % of the protein assigned (Figure 5.1.3).

5.1.4 Structure Determination of hIMP3

A common method for deriving interatomic distance information is through use of the Nuclear Overhauser effect (NOE) spectroscopy (NOESY), where the volumes of cross-peaks relate to the distances between spatially close spins. However, there are several limitations of NOE-based approach in membrane protein NMR: (i) only distances below 6 Å could be detected without special labeling (like partial deuteration of the protein in order to reduce very fast relaxation of ^{1}H magnetization), which is much less than an average distance from backbone HN group (usually used for direct detection) and most of the ^{1}H atoms on neighboring TM helix; (ii) therefore the sidechain-sidechain NOEs are very important but they are very difficult to obtain because the strong and

sharp ^{1}H resonances from detergent interfere with analysis of these NOEs; (iii) internal mobility of membrane protein TM domains in millisecond timescale (comparable with mixing times) broadens resonances and averages NOE values. We therefore turned to an alternate method known as the PRE (paramagnetic relaxation enhancement) effect to determine distances.

The PRE approach relies on relaxation enhancement of the ^{1}H atoms of a protein caused by external or internal, covalently bound to the protein, paramagnetic label. The most efficient approach, described (*9, 10*) uses paramagnetic spin labels attached to the cysteine residues introduced into the protein of interest by mutagenesis. The PRE of ^{1}H atoms located within 14-25 Å from the label can be measured and converted to the distance restraints similar to the NOE data.

Six cysteine mutants were created, p-CF expressed and [^{15}N-^{1}H]-TROSY-HSQC spectra were taken for hIMP3 samples before labeling, with paramagnetic spin label, and with sterically similar diamagnetic spin label. The integral intensities of the cross-peaks were measured in these spectra and used to derive the long-range distance constraints as described in (*10*).

The structure was calculated by using the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift deviations from the "random coil" values, which defined the torsion angle restraints for helical regions (*11, 12*), and the long-range distance constraints derived by analysis of the PRE data. The backbone spatial structures of hIMP3 was calculated through an interactive procedure that included structure calculation by the CYANA program (*13*) followed by the distance constraints refinement. The 20 conformers from the last CYANA calculation

cycle with the lowest target function were analyzed (illustrated in stereo in Figure 5.1.4). The stable regions spanning residues 14-87 of the best 20 structures are shown superimposed by the coordinates of the backbone atoms of hIMP3 helical regions 17-22, 30-46, and 65-82. The NMR solution structure of hIMP3 consists of two transmembrane domains (residues 30-46 and 65-82) and an N-terminal cytosolic α-helix (residues 17-22) (Figure 5.1.5).

5.1.5 Discussion

Although initial attempts at solving the solution NMR structure of EnvZ (1-190) proved unsuccessful, I was able to rapidly assign NMR resonances for the 2-TM integral membrane protein hIMP3 using the CDL strategy. This assignment was completed within just a few weeks allowing us to build a preliminary 2 TM structure of hIMP3. The development of the CDL strategy (*4*) has created a technique that can be used for the rapid production of membrane proteins through CF expression and NMR analysis. With only 20 human integral membrane protein structures known so far (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html), this technique allows for a potential increase in the number of membrane protein structures available. As I have illustrated here, the structure of a human integral membrane protein can be solved using this method within a only a few months time and can thus be a powerful tool for structure based drug design.

5.1.6 Materials and Methods

NMR experiments

High-resolution NMR spectra were recorded at 37° C on a Bruker 700 MHz spectrometer, equipped with five radio-frequency channels and a triple-resonance Cryo-probe with a shielded z-gradient coil. [^{15}N, ^{1}H]-TROSY-HSQC and TROSY-based (*14*) HNCO experiments were measured for each selectively [^{15}N, ^{13}C]-labeled sample for combinatorial assignment. TROSY-based experiments HNCA, HNCO (*15*), HNCACB, and HNCOCA (*16*), as well as 3D ^{15}N-resolved TROSY-[^{1}H, ^{1}H]-NOESY (mixing time 120 ms) were also used for sequential assignment of backbone ^{1}H, ^{15}N, ^{13}C, and side chain ^{1}H resonances.

The CDL strategy

The CDL strategy is composed of four steps: (i) designing the combinatorial labeling scheme; (ii) parallel expression of CDL samples (p-CF expression system) (3.0 ml of reaction mixture for each CDL sample of hIMP3) and solubilization in the same buffer which eliminates any differences in cross peak positions; (iii) short measurement of [^{1}H-^{15}N]-TROSY-HSQC and 2D-HNCO spectra (about ½ -1½ hour per spectrum) for each CDL sample, all the samples for the combinatorial assignment for hIMP3 were measured in only 1-2 days; (iv) analysis of the spectra and assignment of ^{1}H-^{15}N cross peaks using the CARA program (*17*).

Point-directed assignment

The point-directed assignment is based on selective dual isotope (^{15}N and ^{13}C) labeling of two types of amino acids, located consecutively in a protein sequence. For every pair of residues in which the first amino acid is labeled with $^{13}C^{O}$ and the second amino acid is labeled with $^{15}N^{H}$, cross peaks in both the [^{15}N-^{1}H]-HSQC and the HNCO spectra arise (*8*). If the second amino acid in a pair is labeled with $^{15}N^{H}$ and the first amino acid is not labeled with $^{13}C^{O}$, there will only be a cross peak in the HSQC spectrum. For pairs in which the second amino acid is not labeled with ^{15}N, there will be no cross peaks in either spectrum. By analyzing the presence and absence of cross peaks in the [^{15}N-^{1}H]-HSQC and the HNCO spectra recorded from a selectively labeled sample one can define the type of amino acids for backbone resonances participating in these cross peaks, and assign them to a particular amino acid pair. If a pair is unique in the sequence, an exact assignment of the $^{1}H^{N}$, $^{15}N^{H}$, and $^{13}C^{O}$ resonances to the residues that comprise that pair is instantly made.

Paramagnetic Relaxation Enhancement

Measurement of the paramagnetic relaxation enhancement *(PRE) effect* was performed as described (*9, 10*). [^{15}N, ^{1}H] TROSY spectra were measured consequently with all cysteine mutants before spin labelling, after the labelling with MTSL (oxidized form) or DML (equivalent of reduced state). All the spectra were transformed identically, and their integral intensities were calibrated against the intensities in the spectra of the DML-labeled samples using 8-12 cross peaks with the minimal relative signal decrease.

Distance constraints were derived from the measured PRE effect according to the procedure described in reference (*10*).

Solution NMR analysis of H-D exchange

The stability of the secondary structure of the hIMP3 TM domains was studied by exchange of backbone labile protons to solvent deuterons. The ^{15}N-labeled protein was expressed in the p-CF mode in 100% H_2O and solubilized by 2% LMPG in 100% D_2O. The [^{15}N, ^{1}H]-TROSY-HSQC spectra were collected starting at 10 minutes after solubilization. Analysis of localization of the HN protons, demonstrating slow exchange to solvent deuterons, showed that the majority of backbone amide hydrogen's located in the TM helices participated in stable hydrogen bonds.

Structure calculation and analysis

Torsion angle restraints were defined from the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift deviations from the "random coil" values (*11, 12*). The long-range distance constraints were derived by analysis of the PRE effect as described above. An interactive procedure, which included structure calculation by the CYANA program (*13*) followed by the distance constraints refinement, was used to calculate the backbone spatial structures of hIMP3. The 20 conformers with the lowest target function of the last CYANA calculation cycle were analyzed and presented in Figure 5.1.4.

5.1.7 Figures

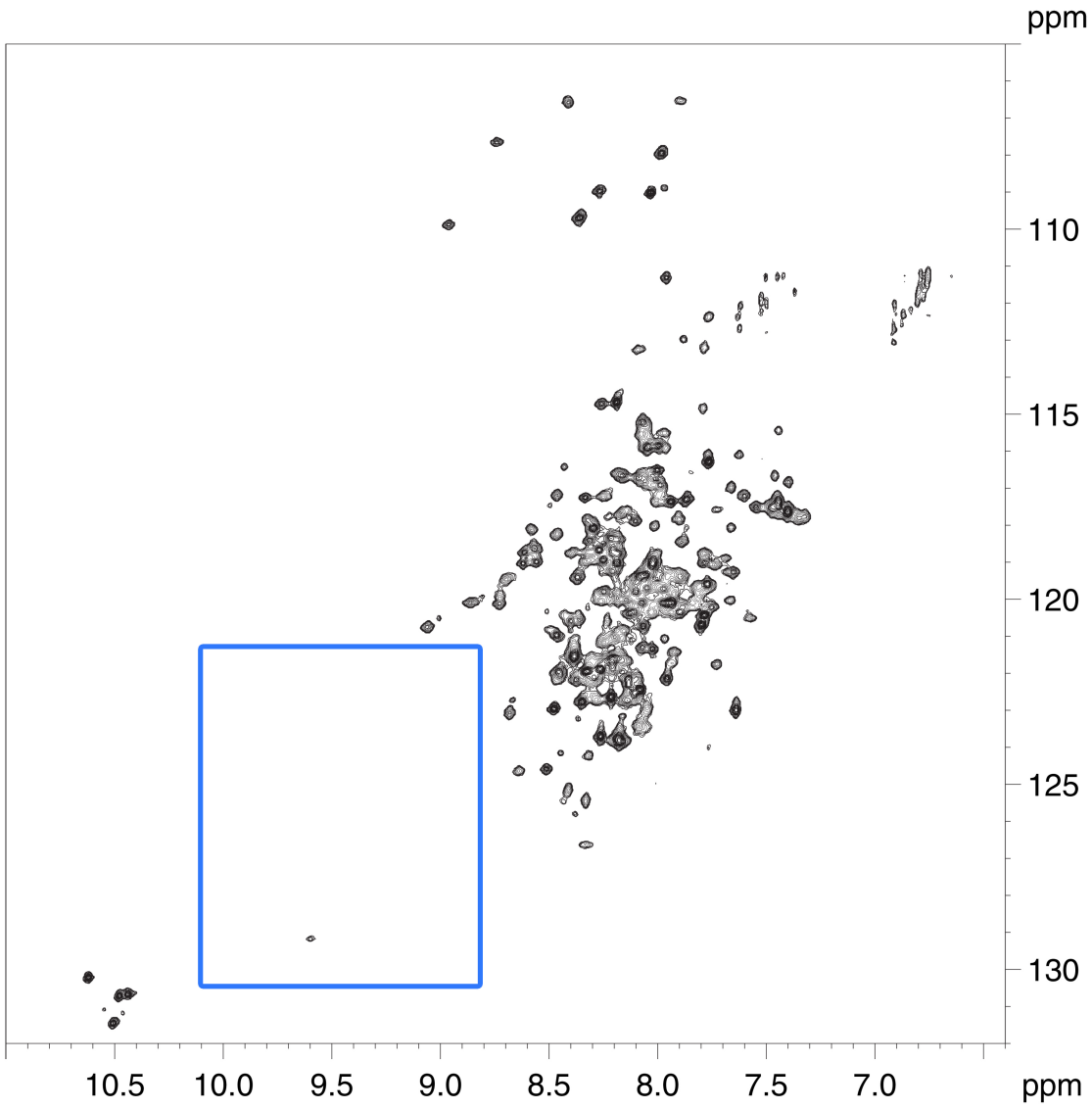

Figure 5.1.1: EnvZ (1-190) [^{15}N]-TROSY-HSQC NMR spectra in high detergent concentration.
EnvZ (1-190) [15N]-TROSY-HSQC NMR spectra in 50 mM FC-12, 20 mM MES pH 4.8, at 45°C (318K) illustrates missing peaks in the expected β-sheet region (blue).

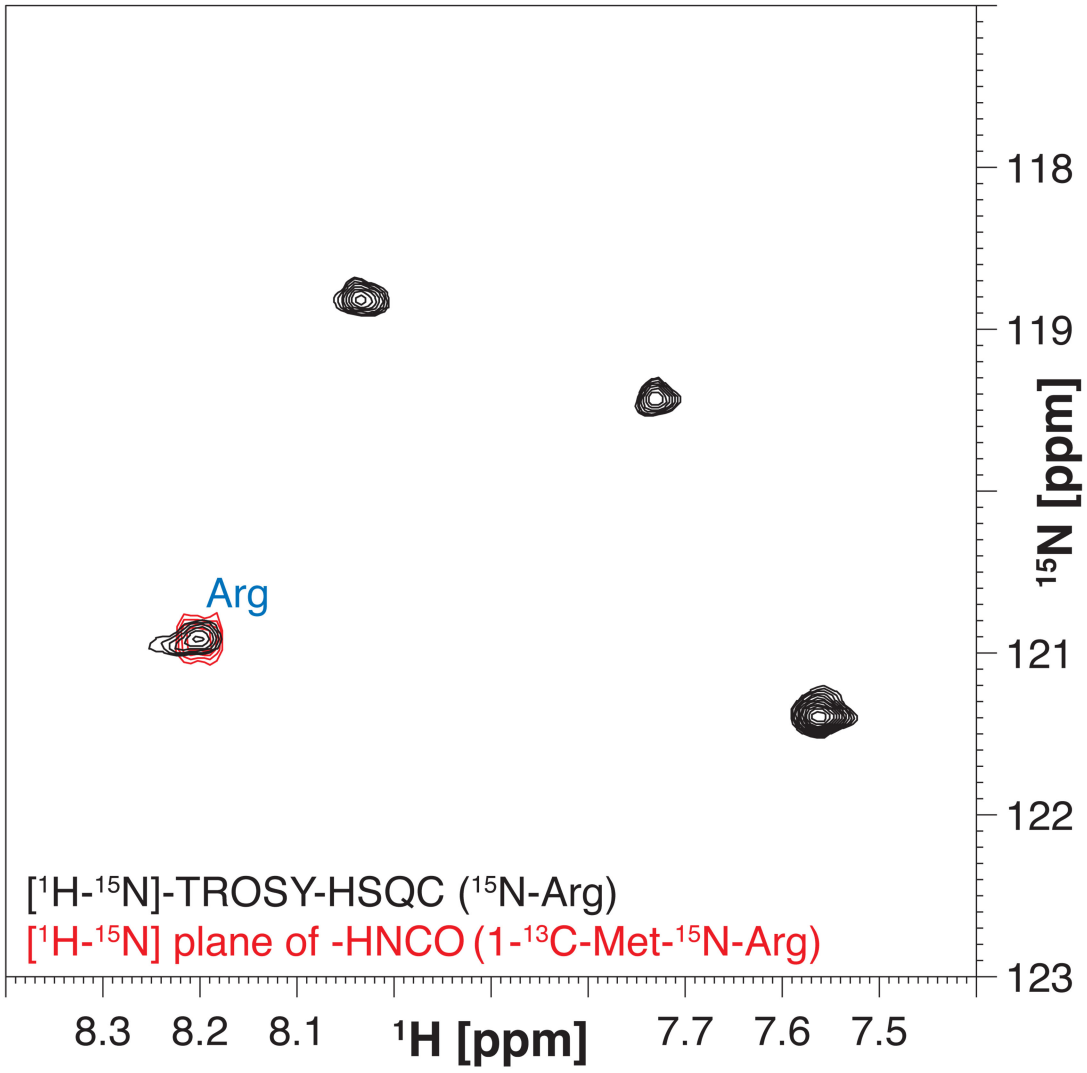

Figure 5.1.2: Overlaid [^{1}H-^{15}N]-TROSY-HSQC NMR spectra of ^{15}N-Arg, 1-^{13}C Met labeled hIMP3.
The [^{1}H-^{15}N]-TROSY-HSQC spectra (black) overlaid with the [^{1}H-^{15}N]-TROSY-HSQC plane of an –HNCO experiment (red).

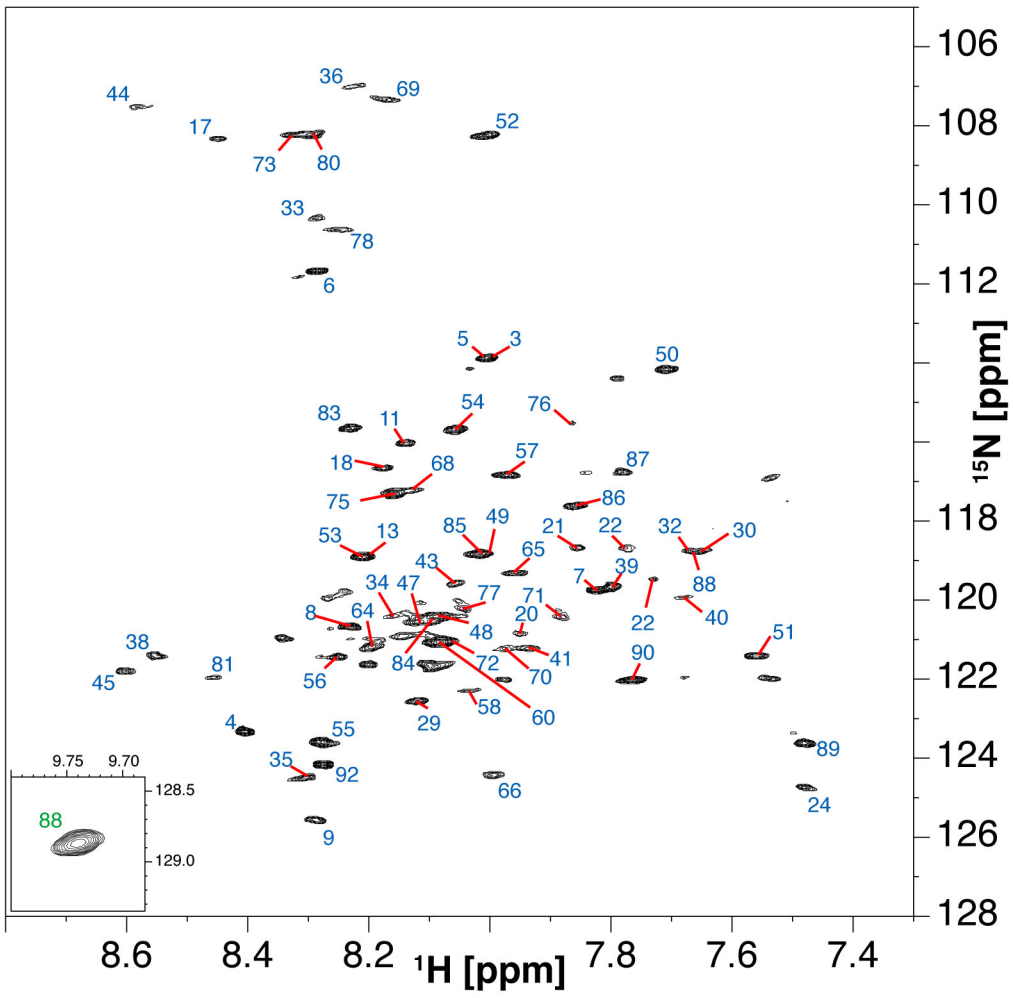

Figure 5.1.3: Assignment of hIMP3 [^{1}H-^{15}N]-TROSY-HSQC spectra.
The assignment of the HN cross-peaks to the specific residue in the hIMP3 sequence is shown by blue numbers.

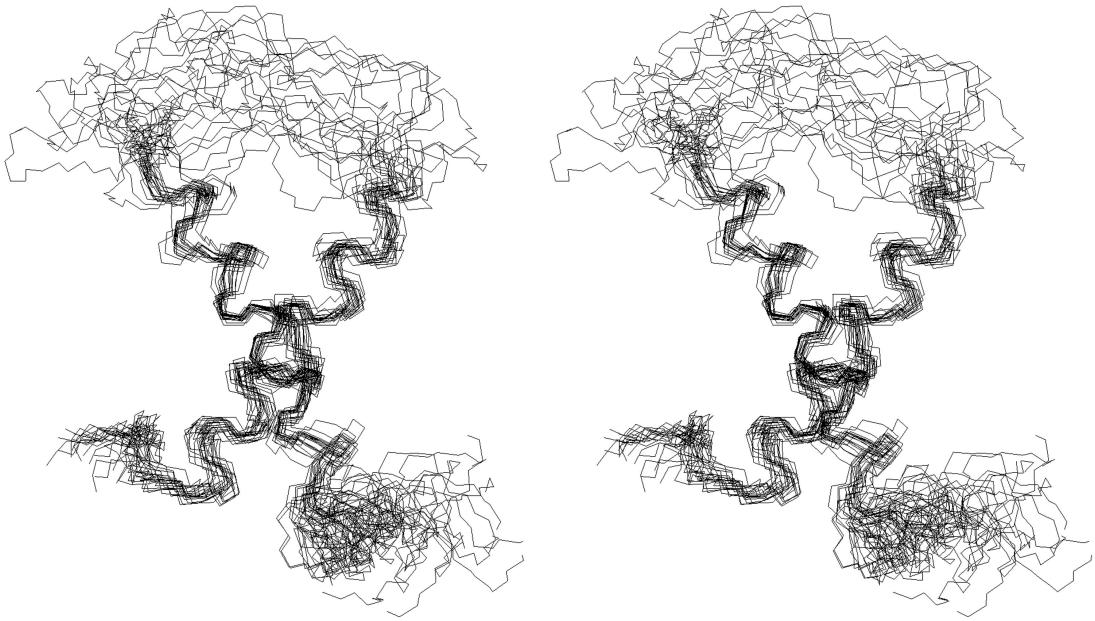

Figure 5.1.4: The stereoview of the top 20 hIMP3 solutions.
The best 20 structures are shown superimposed by the coordinates of backbone atoms of helical regions 17-22, 30-46, and 65-82.

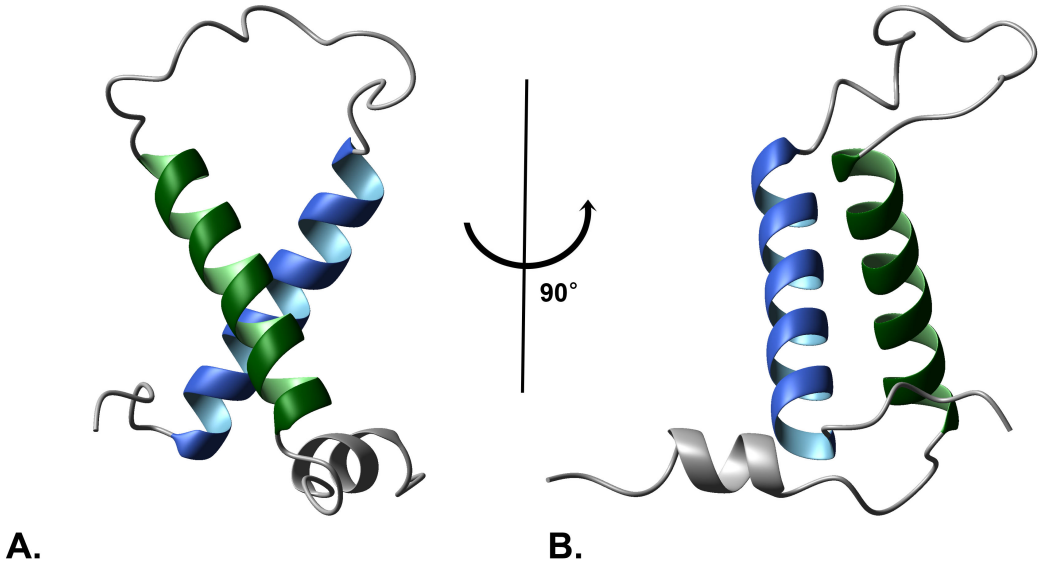

Figure 5.1.5: NMR solution structure of hIMP3.
hIMP3 consists of two transmembrane domains, illustrated in green and blue (residues 30-46, and 65-82), and one NH_2-terminal cytosolic α-helix (residues 17-22) shown in silver. (A) Depicts the front view of hIMP3 and (B) is rotated 90° around the *y* axis.

5.2 The crystallographic structure of NH_2-terminal domain of the NMDA receptor GluN1

The following experiments were done collaboratively between Choe and Nakagawa laboratories.

5.2.1 Introduction

The major excitatory neurotransmitter, L-glutamate is found in the mammalian brain and plays an important role in brain development and function. When L-glutamate is released from the presynaptic nerve terminals, it diffuses across the synaptic cleft and activates a class of ligand-gated ion channels, known as the ionotropic glutamate receptors (iGluRs), located in the post-synaptic membrane (*18, 19*). Three subfamilies constitute the iGluR family and are classified based on their selective agonist binding properties and sequence similarity of the receptor subunits (*20-22*). These subfamilies are the α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) (GluA1-GluA4), kainate (GluK1-GluK5), and *N*-methyl-D-aspartate (NMDA) (GluN1, GluN2A-GluN2D, GluN3A-GluN3B)(*23-27*).

In order to activate members of the NMDA receptor subfamily, both L-glutamate and glycine agonists have to bind and membrane depolarization has to take place to relieve a Mg^{2+} block (*28, 29*). Signal transduction cascades that control synaptic strength related to neuronal development and memory formation are activated when the NMDA receptor opens causing the entry of Ca^{2+} into the postsynaptic space (*30-32*). These receptor channels are the trigger for processes that lead to synaptic plasticity, synaptogenesis and excitotoxicity (*33-35*). The dysfunction of these receptors has been correlated with many different neurological and psychiatric disease states, such as

epilepsy, neuropathic pain, schizophrenia, seizure, stroke, Alzheimer's and Parkinson's disease, and are thus important therapeutic targets (*36, 37*).

The NMDA receptors are composed of a similar topology consisting of four domains. The large extracellular domain is divided into two sub-domains, an NH_2-terminal (NTD) and a ligand binding domain (LBD). The LBD is responsible for glutamate and glycine binding and initiates the opening of the channel. The channel pore is composed of three transmembrane domains (TMD) containing a re-entrant pore loop (*38*). The last domain is the cytoplasmic COOH-terminal domain (CTD), which is responsible for interactions with cytosolic proteins and signal transduction cascades (*39*).

NMDA receptors are assembled as obligate heterotetramers (*27*) and most of these receptors are composed of 2 GluN1 subunits coupled with 2 GluN2 subunits, or 1 GluN2 and 1 GluN3 subunit (*27, 40, 41*). Here we report the crystallographic structure of the GluN1 NTD, which was expressed and crystallized by Anthony N. Farina from Terunaga Nakagawa laboratory. The GluN1 NTD structure was determined to 3.4 Å resolution limit and refined to R=21.6% and R_{free}=26.7%. We investigate why molecular replacement was unsuccessful, and why a combination of SIRAS and SAD using iodine signal had to be used. The GluN1 NTD crystals were also examined to explain why they contained a high solvent content of 71% and if this is an outlier of the Matthews coefficient distribution. The GluN1 NTD is an important part of NMDA receptors, and its structure can lead to insights regarding biological function of glutamate receptors.

5.2.2 Determining the phases using a combination of SIRAS and SAD

After collecting the native data set with gluN1-NTD crystals diffracting to 3.4 Å on our Rigaku MicroMax-007 home source X-ray, we first attempted to determine the phases using molecular replacement. These initial attempts were to select search models based on previous glutamate receptor NTDs, because they seem to be composed of a conserved 2 sub-domain clam-shell motif. Searching the Protein Data Bank (http://www.rcsb.org/pdb/) for previously solved glutamate receptors, the following models were selected: the kainate receptor GluR6 NTD (PDBID-3H6G) (*42*), the AMPA-receptor GluA2 NTD (PDBID-3KG2) (*43*), and the NMDA receptor NR2B NTD (PDBID-3JPW) (*44*). However the molecular replacement program Phaser v. 2.1.4 (*45, 46*) output incorrect solutions that would not refine.

Since molecular replacement did not yield successful solutions, alternative methods for determining the structure were investigated. By growing the crystals in the presence and in the absence of NaI, iodine incorporation into the crystal grown with NaI was confirmed during the data processing step by comparing the χ^2 values of equivalent-reflections-merged data with Friedel pairs included and not included during the merging process. The benefit of using iodine is that this heavy atom has a high anomalous scattering signal (~7 electrons) at our home X-ray source (CuKα anode, wavelength 1.5418 Å), in comparison many other heavy atoms. Bromine, for example, has as a low anomalous scattering signal (~1.25 electrons) at CuKα wavelength and requires the use of a tunable synchrotron X-ray source where you can tune the wavelength to the absorption edge of bromine. Even after tuning the X-ray to the absorption edge of

bromine at the synchrotron, bromine still emits almost 2X less anomalous scattering signal (~3.7 electrons) when compared to iodine anomalous scattering signal on our home X-ray source Figure 5.2.1. This high anomalous scattering signal allowed enabled the single wavelength anomalous dispersion (SAD) phasing method. Additionally, by collecting a data set from native crystals (crystals grown without NaI), a combination of SAD and single isomorphous replacement with anomalous scattering (SIRAS) methods of phasing could be used. The additional benefit of this procedure was that it allowed us to be able to determine the structure without shipping the crystals to the synchrotron, or subjecting them to radiation damage often caused by synchrotron radiation.

Using the autoSHARP (*47, 48*) program from the CCP4 suite (*46*), which combined SIRAS and SAD with iodine signal, we were able to locate the iodine atom sites and improve phases through solvent flattening, which yielded electron density maps that permitted the manual building of the protein chains in COOT version 0.6.2-pre-1 (*49*). The final solution contains three molecules in the asymmetric unit, which are arranged in space group $P3_121$ with unit cell dimensions $a = b = 164.7$, $c = 147.3$. The native data set was used for final refinement, which was completed after multiple cycles of refinement using REFMAC V 5.5.0109 (*50*) utilizing non-crystallographic symmetry (NCS) restraints between chains and converged to R=21.6% and R_{free}=26.7% (Table 5.2.1).

5.2.3 Crystal structure of NH_2 terminal domain of GluN1

The monomeric GluN1 NTD possesses a clamshell-like architecture similar to the structures of other glutamate receptor domains and is further divided into two sub-domains, R1 and R2, that are connected by three loops (Figure 5.2.2 A). The NH_2-terminus begins with R1 and the COOH-terminus ends in the R2 domain where the loop regions that compose domains R1 and R2 create three interdomain crossings. GluN1 NTD contains six potential sites of *N*-linked glycosylation, in which five of these six *N*-acetyl-glucosamine sites (Asn_{61}, Asn_{203}, Asn_{239}, Asn_{276}, and Asn_{300}) are seen in the electron density.

Four GluN1 NTD monomers make up a tetrameric assembly in which each monomer is listed as mA, mB, mC, mD (Figure 5.2.2 B). Two distinct dimers compose the GluN1 NTD tetramer. Dimer 1 is composed of monomers mA-mB and mD-mC (Figure 5.2.3 A). In this dimer, the dimeric interface is composed of residues found within the R1 sub-domains of each monomer. Dimer 2 is composed of monomers mB-mC and mD-mA (Figure 5.2.3 B). Dimer 2 differs from dimer 1 in that the dimeric interface is composed of residues found in the R1 and R2 sub-domains of one monomer and the R2 sub-domain of the second monomer. The biological relevance of these different oligomeric assemblies will still need to be investigated.

The packing arrangement of GluN1 NTD contains a basic repeating tetramer with a pseudo two-fold axis of symmetry separated by large gaps of solvent in between (Figure 5.2.4 A-B). Interestingly, by looking at the packing arrangement of the tetramers

one can visualize the high solvent content of 71% contained within the GluN1 NTD crystals. The high solvent content is addressed in a later section.

5.2.4 Why molecular replacement did not work

The root-mean-square deviation of the models from our solved structure, which was calculated after we solved it by the combination of SIRAS and SAD, turned out to be 3.77 Å for GluR6 NTD (*42*), 2.72 Å for NR2B NTD (*44*), and 3.64 Å for GluR2 NTD (*43*), as determined by the program Superpose (*51*) using structural alignment based on secondary structure matching in the CCP4 suite (*46*). When these three structures were superimposed using SSM in COOT version 0.6.2-pre-1 (*49*), we found that although the overall clam-shell like motif is similar, there are major differences in not only the loop regions, but also in the Cα backbone trace (Figure 5.2.5). These differences could be one of the factors that caused our attempts at finding phases using molecular replacement to fail. We attempted many different combinations of molecular replacement by creating search models using poly-alanine chains, homology models, and searching for truncated domains without loop and linker regions.

In addition to the problems with structural similarities between the structure and the models used in molecular replacement, the high solvent content contributed to the difficulties of obtaining the correct molecular replacement solution. Most of the molecular replacement programs predicted 4-5 molecules in the asymmetric unit based on the Matthews coefficient and the solutions did not have any clashes when symmetry equivalent molecules were generated. However, such solutions would refine with a high *R*-factor ~50%.

5.2.5 Is having such a high solvent content an outlier of the Matthews coefficient distribution?

In X-ray crystallography one of the first steps in structure determination is estimating the number of molecules in the asymmetric unit. Solvent occupies a large percentage of the volume of a protein crystal. In 1968 and 1976, Matthews noticed that in the majority of structures the solvent occupied, on average, about 43% of the crystal volume (*52, 53*). He found that the distribution of the Matthews Coefficient (V_m) was useful in the analysis of protein crystals to estimate the number of molecules per asymmetric unit, especially for crystals in the molecular weight range under 70 kDa. Although decades have passed since Matthew's original studies of solvent content in protein crystals, the Matthews Coefficient is still used today for analysis of protein and solvent content of the asymmetric unit. The solvent content and V_m, of a crystal can be calculated by two factors, the unit cell dimensions and the molecular weight of the molecules within that unit cell (*52*). Solvent content is also important in SAD experiments, in which it is needed in order to use the technique of solvent flattening (*54-56*). Solvent flattening resolves phase ambiguity and improves the quality of the phases and thus the electron density as well.

In 1968 Matthews observed that the average solvent content of protein crystals was 43% (*52*) and a study completed in 2003 by Kantardjieff and Rupp comparing V_m and resolution of diffraction data suggests that crystals with a lower solvent content tend to diffract to a higher resolution (*57*). However, our GluN1 NTD crystals contained a relatively high solvent content, 71%, in comparison to Matthew's observed average of

43%, and diffracted to 3.4 Å on our home source and 3.0 Å at the synchrotron. Initial inspection of the Matthews Coefficient for our GluN1 NTD crystals based on Matthews V_m distribution from 1968, one would assume that there were about 5 molecules per asymmetric unit, yielding a Matthews Coefficient and solvent content of V_m =2.55 Å^3/Da, 51.73%. However, upon determining the structure for GluN1 NTD we discovered that there were three molecules per asymmetric unit yielding a Matthews Coefficient of V_m = 4.24 Å^3/Da and 71.04% solvent. Besides the discrepancy between the models we used and the final solution, the number of molecules in the asymmetric unit based on the Matthews coefficient was another factor contributing to our unsuccessful attempts at molecular replacement.

To determine how common this occurrence is in the Protein Data Bank (http://www.rcsb.org/pdb/), we decided to perform a search for unique protein structures (with a sequence identity cut off set at 90 %) solved with crystals containing a high solvent content. After searching the PDB for structures with a solvent content above 71% between the X-ray diffraction resolution of 1 Å to 5 Å resolution limit, we found that there are over 1000 structures deposited (November 2010). The majority of these protein structures lie between the X-ray resolution limit range of 2.5 Å to 3.0 Å, about the resolution range of our GluN1 NTD crystals (Figure 5.2.6). Moreover, a large portion of these structures, about ~ 36 %, diffract to 2.5 Å limit or better. Interestingly, five of the structures lie in the resolution limit range between 1.0 Å to 1.5 Å and include the following PDBID's: 3EOJ, 2FGQ, 1YBK, 1XUX, and 1LLM. These outlier structures consist of proteins, peptides, protein-DNA complexes, and RNA.

In addition to the many structures that have been deposited in the PDB with a high solvent content, a previous study has shown a correlation between the solvent content of a crystal and its symmetry. When 9081 protein crystal structures deposited in the PDB were examined for solvent content, Matthews coefficient, and oligomeric state, it was found that higher symmetry is related to higher solvent content (*58*). In this study it was found that those crystallizing in higher crystal systems, such as the hexagonal and cubic systems, the mean Matthews coefficient and solvent contents were V_m = 3.08 Å^3/Da, 57 % solvent content, and V_m = 3.44 Å^3/Da, 61 % solvent content, respectively. When this study compared and ranked the Matthews coefficient and solvent content as a function of the 28 most frequent space groups, $P3_121$ was ranked number 10, with an average Matthews coefficient of V_m = 2.96 Å^3/Da, 56% solvent content, and space group $P2_12_12_1$ ranked at number 1. Space groups $I4$, $I422$, $I4122$ and $P6_322$ had the highest solvent content ranging from 57 % to 76 %. Similar to GluN1 NTD, other glutamate receptor NH_2 terminal domain structures like NR2B NTD (*44*) and GluR6 NTD (*42*), have also been solved in the high symmetry primitive hexagonal Bravais lattice and both also contain a high solvent content of 77 %. The findings of this previous study correlates not only with other glutamate structures but also with our GluN1 NTD structure crystallizing in the higher symmetry hexagonal crystal system with the space group $P3_121$ and provides insight into the high solvent content in terms of symmetry.

5.2.6 GluN1-NTD *N*-linked glycosylation sites determined by SELDI

Since GluN1-NTD was expressed and treated with endoglycosidase H (Endo H) before crystallization, we decided to determine the exact number of glycosylation sites in the sample that was being crystallized. To do this we utilized surface-enhanced laser desorption/ionization time of flight (SELDI-TOF) mass spectrometry to evaluate the masses of purified GluN1-NTD before and after treatment with EndoH. These experiments yielded two peaks which represent two populations of GluN1-NTD glycosylated at a different number of sites. The first peak (labeled in red) illustrates a mass shift from 53,000.3 Da to 46, 482.9 Da, and corresponds to the cleavage by Endo H of 5 out of 6 sites of glycosylation. The second peak (labeled in blue) contains a mass shift from 51,811.1 Da to 45,206.2 Da and correlates with the cleavage by Endo H of 5 out of 5 sites of glycosylation (Figure 5.2.7). The remaining peaks in the spectra belong to GluN1-NTD protein degradation products. This data agrees with the crystal structure that we solved in that 5 of these cleaved glycosylation sites are resolved in the electron density (Asn_{61}, Asn_{203}, Asn_{239}, Asn_{276}, and Asn_{300}).

5.2.7 Discussion

Here we have determined the crystallographic structure of GluN1 NTD and examined its architecture along with its tetrameric and dimeric oligomerization states. We have also illustrated that SELDI-TOF mass spectrometry can be used to identify the number of *N*-linked glycosylations on a purified sample which can be of potential use to those attempting to crystallize proteins expressed in eukaryotic systems. Accurately

assessing the homogeneity of a glycosylated sample could increase the likelihood of obtaining crystals.

After solving the GluN1 NTD structure through a combination of SIRAS and SAD, we examined the problems that were encountered with molecular replacement. We found that the high solvent content of the crystals, most likely due to crystal symmetry, created a high Matthews coefficient. Since molecular replacement programs estimate the number of molecules in the asymmetric unit based on the Matthews coefficient, they output solutions with four to five molecules per asymmetric unit, when actually the solution contains three. This information is important because it may lead to faster structure determination in cases in which there is a high solvent content.

5.2.8 Materials and Methods

Crystal Structure Data Collection, Determination, and Refinement.

The diffraction data were collected for iodine and native crystals at the Salk Institute, SBL-C, on Rigaku MicroMax-007 at a resolution of 3.4 Å. The iodine and native X-ray data sets comprises 330111 and 192346 reflections, respectively and were integrated and scaled using the program HKL-2000 version 0.98.692i (*59*). The space group of GluN1-NTD was found to be $P3_121$ with unit cell dimensions $a, b = 164.7$, $c = 147.0$, $\alpha, \beta = 90$, $\gamma = 120$. The Matthews coefficient for GluN1 was determined using the program MATTHEWS_COEF (*52, 57*) within the CCP4 suite by inputting the mass as determined by SELDI-TOF.

The crystal structure of GluN1-NTD was solved by a combination of SIRAS single isomorphous replacement with anomalous scattering and SAD –single wavelength anomalous dispersion using iodine signal. Phase calculation was completed using the autoSHARP (*47, 48*) software in the CCP4 program suite 6.1.13 (*46*). The iodine atom sites were located with the program SHELXD (*60*) and solvent flattening was performed in SOLOMON (*56*) yielding electron density maps that permitted the manual building of the protein chains using the program COOT version 0.6.2-pre-1 (*49*). The final solution contains 3 molecules in the asymmetric unit each containing 5 ASN linked NAG (N-acetyl glucosamine) molecules. Using the native data set, refinement of the structure was completed after multiple cycles of refinement using REFMAC V 5.5.0109 (*50*) utilizing non-crystallographic symmetry (NCS) restraints and converged to R and R_{free} of 21.6% and 26.7%, respectively.

The final data processing and refinement statistics are listed in Table 5.2.1. The root mean squared deviations in bond lengths and bond angles are 0.011 Å and 1.522 deg, respectively. The Ramachandran plot calculated for the final model using the PROCHECK software (*61*) illustrates that the final conformations for 80% of the residues are located in the most favored region and 14.7 % of the residues are located in the additionally allowed regions. All water molecules have a density of 1 σ or greater in the $2F_o$-F_c map.

SELDI-TOF MS - protein profiling and analysis

Each sample was analyzed on a reversed-phase hydrophobic H50 chip array. All spots on the H50 chips were activated with 5 μl of 50% acetonitrile and allowed to dry. 5 μl of H50 buffer (Ciphergen Biosystems, Inc.) was added to each spot and incubated at room temperature for 2 min., this was repeated once. Purified GluN1-NTD samples were added after removal of the H50 buffer and the chip was incubated at room temperature in a humid chamber for 30 min. with shaking. The protein sample was removed and the chip was air dried for 5-10 min. Once dry, 1 μl of sinapinic acid (SPA) energy absorbing molecules (EAM) (Bio-Rad Laboratories, Inc.) was applied to each spot twice. The chips were allowed to air-dry and then placed into the Protein Biological System II-C mass spectrometer reader (Ciphergen Biosystems, Inc.) and TOF spectra were collected by averaging 200 laser shots with an intensity of 237 and a detector sensitivity of 8. The optimization range was from 20,000 to 40,000 Da and the high mass was set to 180,000 Da.

5.2.9 Figures and Tables

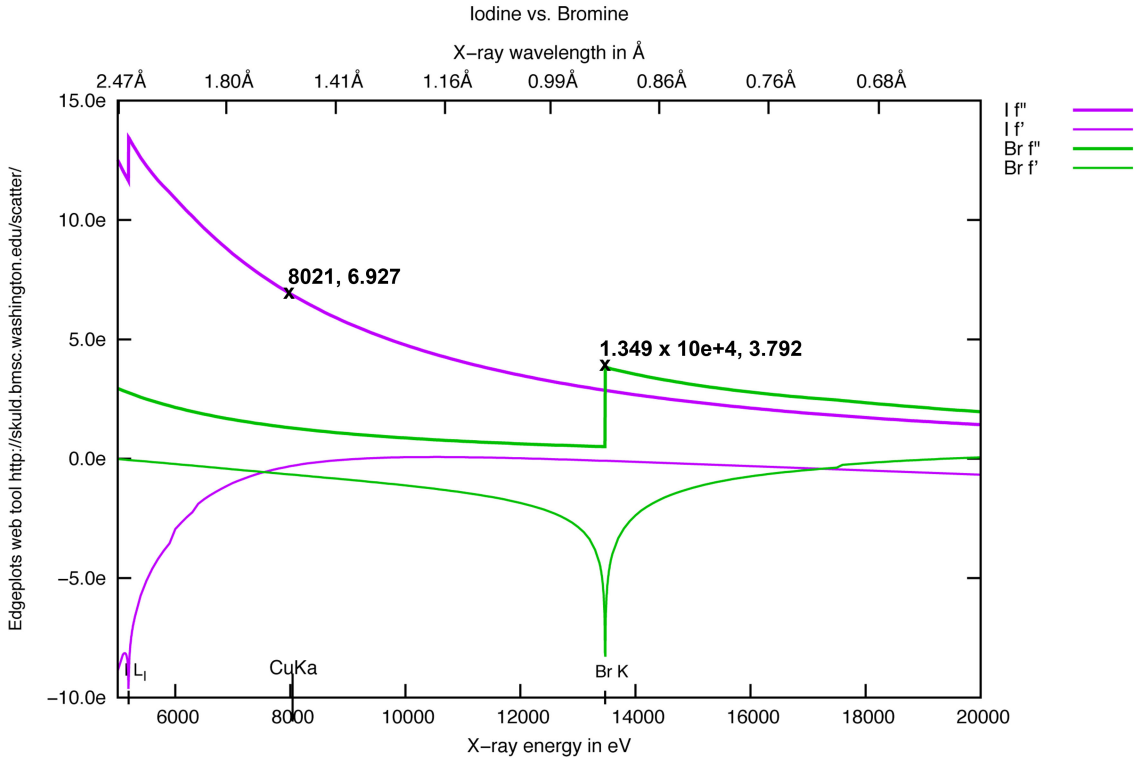

Figure 5.2.1: Iodine vs. bromine absorption edge plot comparing f' and f" over X-ray energy.
Iodine (purple) f' and f" is compared to bromine f' and f" (green). The anomalous signal of iodine at the CuKα X-Ray energy is noted on the purple plot and the the absorption edge of bromine is noted on the green plot. Figure created by http://skuld.bmsc.washington.edu/scatter/ (*62-65*).

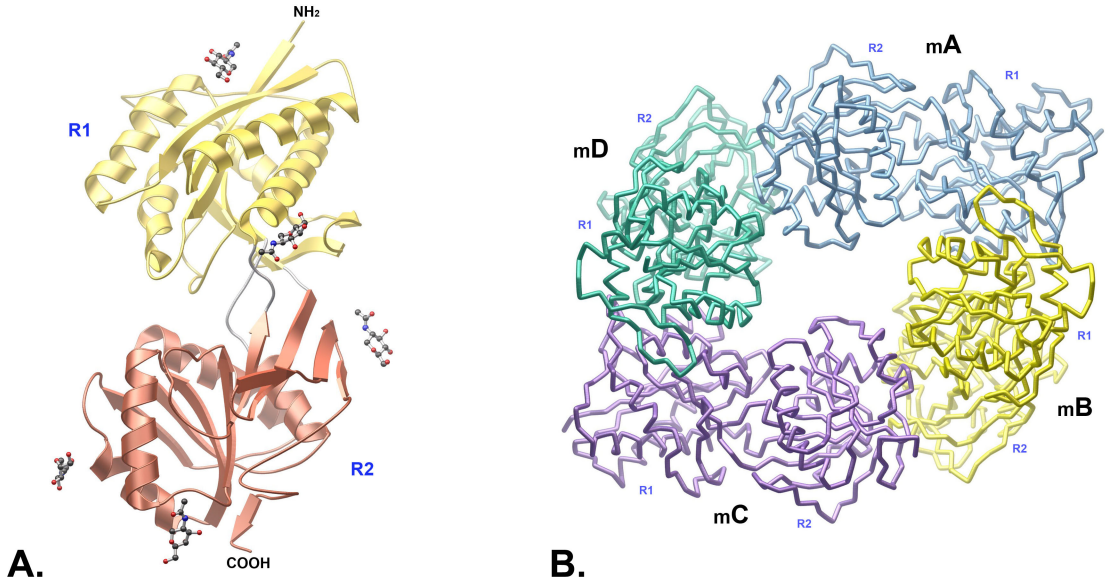

Figure 5.2.2: GluN1 NTD crystal structure.
(A) Monomeric GluN1 NTD composed of R1 (gold) and R2 (rust) sub-domains connected by 3 interdomain loops (gray). 5 *N*-acetyl-glucosamines are modeled illustrated in gray. (B) Tetrameric GluN1 NTD composed of 4 monomers labeled mA (blue), mB (yellow), mC (purple), and mD (green). R1 and R2 sub-domains are labeled in blue. Figure was prepared and rendered using Molscript Version 2.1.2 (*66*) and POV-Ray (*67*).

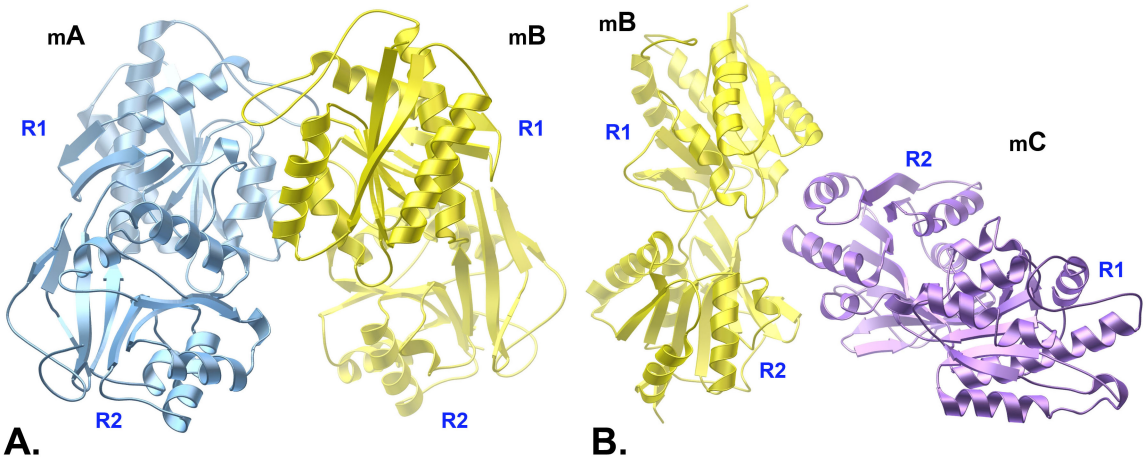

Figure 5.2.3: GluN1 NTD dimer structures.
(A) GluN1 NTD dimer 1, where the dimeric interface is composed of the R1 sub-domains from monomer mA (blue) and monomer mB (gold). (B) GluN1 NTD dimer 2, where the dimeric interface is composed of the R1 and R2 sub-domains of monomer mB (gold) and R2 of mC (purple). R1 and R2 sub-domains are labeled in blue. Figure was prepared and rendered using Molscript Version 2.1.2 (*66*) and POV-Ray (*67*).

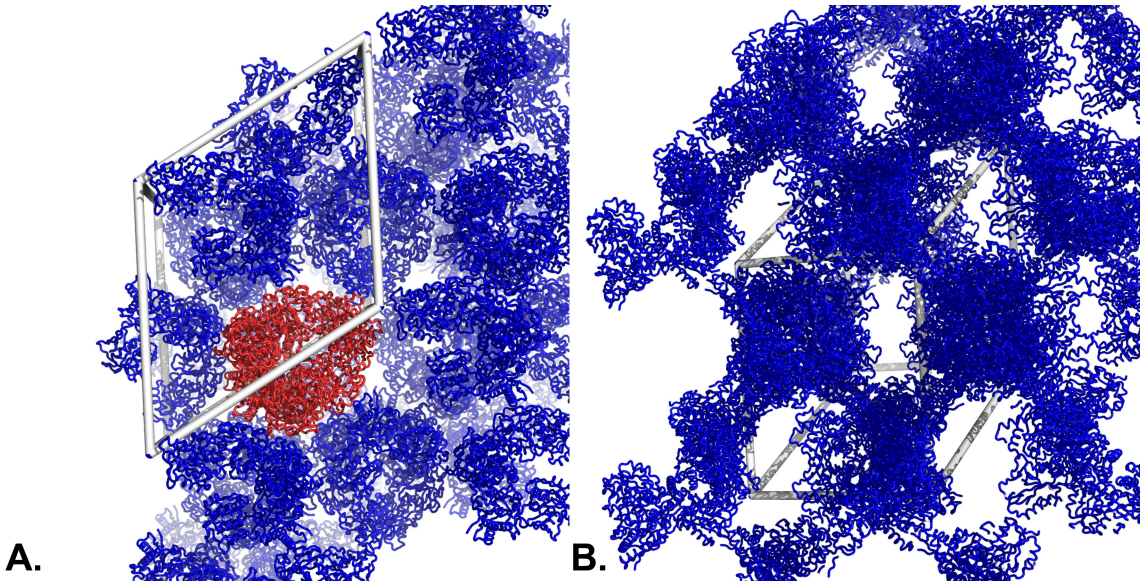

Figure 5.2.4: The packing arrangement of the GluN1 NTD crystals arranged in the primitive hexagonal Bravais lattice.
(A) The basic repeating unit is the GluN1 NTD tetramer (represented in red). Gaps filled with solvent can be seen in between the tetramers. (B) Alternate view of GluN1 NTD packing arrangement. Figure was prepared and rendered using using PyMOL v. 1.3.x (Schrödinger LLC).

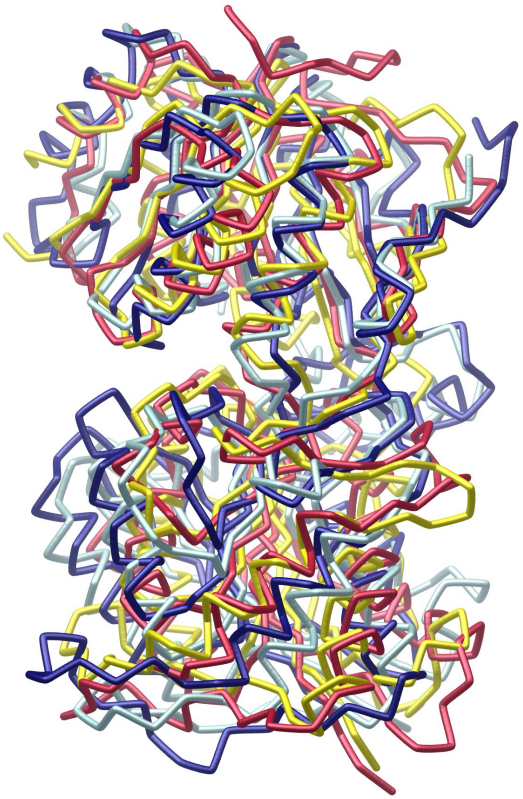

Figure 5.2.5: Glutamate receptor search models superimposed on GluN1 NTD. GluR6 NTD (cyan), NR2B NTD (yellow), and GluR2 NTD (navy) are overlaid onto GluN1 NTD (red). Figure was created using SSM in COOT version 0.6.2-pre-1 (*49*) and the image was prepared and rendered using Molscript Version 2.1.2 (*66*) and POV-Ray (*67*).

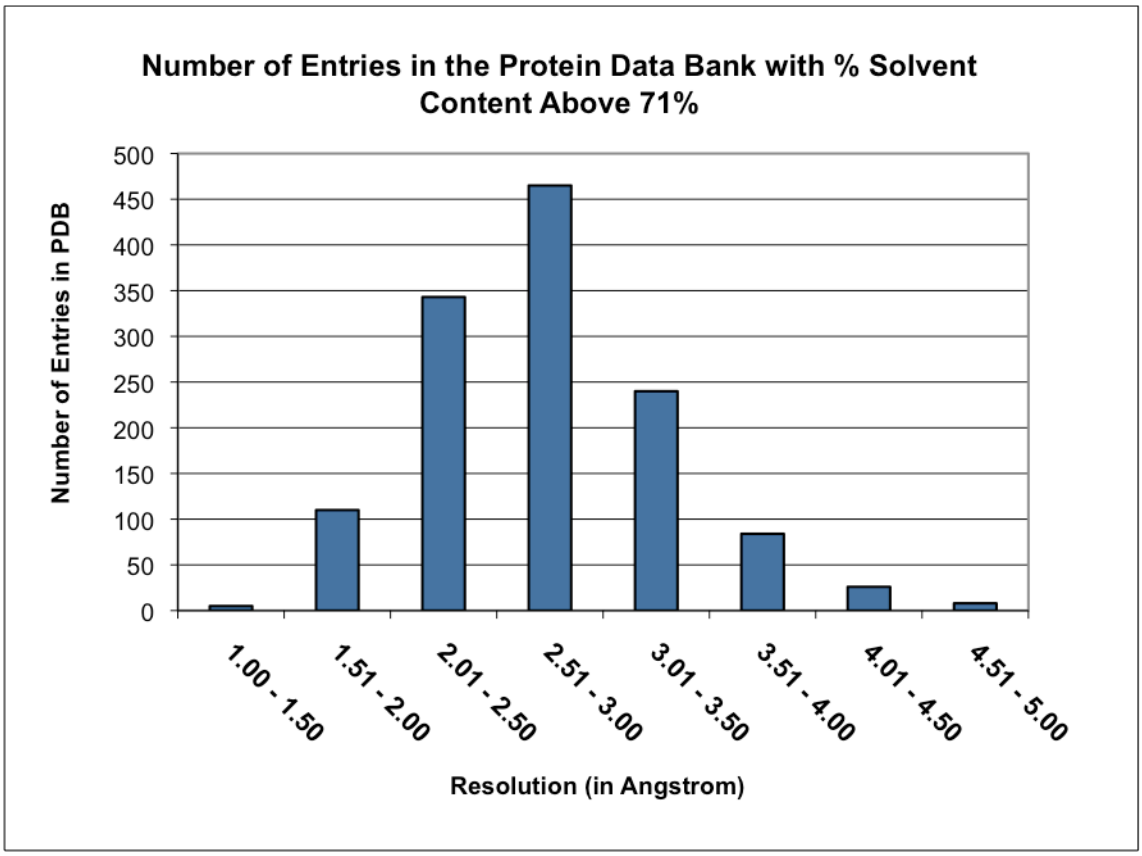

Figure 5.2.6: Histogram of the number of protein structure entries in the Protein Data Bank with a solvent content above 71 % between the resolution ranges of 1 Å and 5 Å.

X-ray resolution is listed on the X-axis and the number of entries in the PDB is listed on the Y-axis. Multiple PDB entries with a sequence similarity of 90% are represented by a single structure.

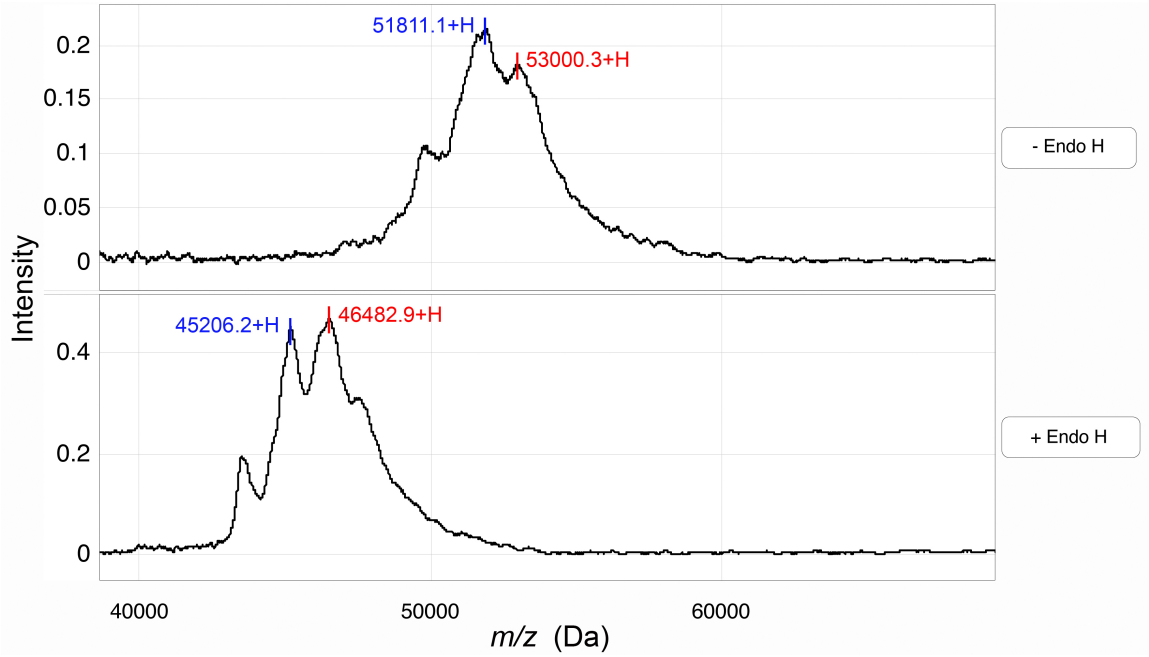

Figure 5.2.7: GluN1-NTD *N*-linked glycosylation sites determined by SELDI. Purified GluN1-NTD samples were analyzed on a H50 protein chip array using SELDI-TOF MS to compare masses of GluN1-NTD before (top, - Endo H) and after (bottom, + Endo H) treatment with Endo H. Masses labeled in red correspond to GluN1-NTD glycosylated at 6 sites, where as those labeled in blue correspond to GluN1-NTD glycosylated at 5 sites. The mass of the red peak shifts from 53,000.3 Da to 46,482.9 Da, which correlates to the cleavage of 5 of the 6 sites of glycosylation post Endo H treatment. The mass of the blue peak shifts from 51811.1 Da to 45206.2 Da, which correlates with the cleavage of all 5 sites of glycosylation post Endo H treatment. All other peaks in the spectra are protein degradation products.

Table 5.2.1: GluN1 NTD data collection and refinement statistics for iodine and native data sets.

Data collection and model refinement statistics		
	GluN1-NTD (iodine)	GluN1-NTD (native)
space group	$P3_121$	$P3_121$
unit cell parameters (Å, deg)	a, b = 164.7	a, b = 164.7
	c = 147.0	c = 147.3
	α, β = 90.0	α, β = 90.0
	γ = 120.0	γ = 120.0
data collection statistics		
Beamline	Rigaku MicroMax-007 (The Salk Institute, SBL-C)	Rigaku MicroMax-007 (The Salk Institute, SBL-C)
wavelength (Å)	1.5418	1.5418
resolution range (Å)	100-3.40 (3.46-3.40)	100-3.40 (3.46-3.40)
observed reflections	330111	192346
unique reflections	32227	32023
Redundancy	10.2 (10.1)[b]	6.0 (5.8)[b]
R_{merge} (%)	17.3 (68.5)[b]	14.6 (63.1)[b]
Mean I/σ (I)	15.0 (3.4)[b]	12.9 (2.8)[b]
completeness (%)	99.9 (100.0)[b]	99.9 (100.0)[b]
refinement statistics		
R_{cryst} (%)	21.7	21.6
R_{free} (%)[a]	28.3	26.7
average B factor (Å^2)	76.284	74.395
root mean square deviation		
bond lengths (Å)	0.010	0.011
bond angles (deg)	1.484	1.522
number of atoms total:	8983	8968
Protein	8725	8725
Water	22	25
Chlorine	7	8
Iodine	19	0
N-Acetyl Glucosamine	210	210
ramachandran plot (%):		
most favored	79.8	80.0
additional allowed	15.2	14.7
generously allowed	2.4	2.8
disallowed allowed	2.7	2.6

[a] calculated from 5% of data not used in refinement. [b] Numbers in parentheses correspond to highest resolution shell (Å).

5.3 Crystallographic structure of Adenovirus E4-ORF

This work is a collaboration between Horng Der Ou from Clodagh O'Shea's laboratory and the Choe laboratory.

5.3.1 Introduction

Both tumor mutations and DNA tumor virus proteins deregulate the same cellular targets and mechanisms, that drive abnormal cell proliferation (*68*). These small viral proteins function by interacting with key cellular proteins that in turn alter normal protein signaling pathways. Therefore viral proteins are of interest for identifying new targets that are deregulated in cancer and for the design of new therapies based on viral and cellular protein interactions (*69*).

The tumor-suppressor p53 is a protein that is constitutively expressed in normal cells where degradation leads to its regulation (*70*). Activation signals such as DNA damage causes a stabilization of p53 and a rise in p53 levels, leading to an inhibition of cell growth (*71*). In order to achieve normal cell growth and development, tight regulation of p53 is required (*72-74*). Inactivation of p53 through rearrangements, gene losses, or point mutations has been found in many types of cancer (*75-77*).

Adenovirus early region 4 (E4) encodes many proteins responsible for regulating the viral lytic program and other cellular processes, one such protein is known as E4 open reading frame 3 (ORF3) (*78, 79*). This 116 amino acid protein has been recently found by the O'Shea laboratory to form a novel and dominant epigenetic structure that silences p53-transcription, regardless of p53 stabilization and phosphorylation (*80*). E4-ORF3 assembles into a high-order oligomer to create a nuclear scaffold that specifies

heterochromatin formation at p53 target promoters, and prevents p53 from binding to DNA. From these findings the oligomerization of E4-ORF3, which creates this nuclear scaffold, appears to be critical for its function (*80*).

Here I present the crystallographic structure of E4-ORF3 that I have solved to 2.06 Å resolution. Upon analysis of the crystal structure it appears that E4-ORF3 shares a similar architecture to IS*Hp608* TnpA Transposase, where both contain a motif similar to that of the RNA recognition motif. Based on the crystal structure we present a possible mode of higher order oligomerization and fiber formation where the dimer composes the basic building block. I also propose structure-based point mutations that could be created to disrupt and recover the strong hydrophobic core found within E4-ORF3. Together these findings will help us better understand how this viral protein assembles into these higher-ordered oligomers that compose such an important nuclear scaffold.

5.3.2 Crystallographic structure of E4-ORF3

In order to block the fiber formation and increase solubility, mutations were introduced in E4-ORF3. Crystals of the mutated protein were relatively easy to obtain. The E4-ORF3 protein was expressed, purified and crystallized by Horng Der Ou from Clodagh O'Shea's laboratory. Based on the sequence similarity we could not identify a model for the molecular replacement in the PDB so direct phasing methods had to be used. Many heavy metal soaks including K_2PtCl_4, $HgCl_2$, and samarium acetate hydrate Sm (OOCCH2) x3 H_2O, however, all yielded unsuccessful results. However crystals of Se-Met protein grew and many of them were screened at SSRL to obtain good quality

diffraction. Subsequently, using Se-Met Multiple –wavelength anomalous dispersion (MAD) method, the crystal structure of E4-ORF3 mutant was solved in the space group $P2_12_12_1$. The protein crystallized in 3 other space groups, $P2_1$, $P6_122$, and $C222_1$. These 3 crystal structures were all solved by molecular replacement. Structures in space groups $P2_12_12_1$, $P2_1$, and $P6_122$ contained two molecules per asymmetric unit forming a conserved dimer. The structure solved in space group $C222_1$ contained one molecule in the asymmetric unit. However, through crystallographic symmetry it also formed the same dimer as seen in all other solved mutant forms. Data was collected and the structure was interpreted at 2.06 Å ($P2_12_12_1$), 2.30 Å ($P2_1$), 2.84 Å ($P6_122$), and 2.20 Å ($C222_1$) resolution (Table 5.3.1).

The monomeric subunit of mutant E4-ORF3 is characterized by a $\beta_1\alpha_1\alpha_2\beta_2\beta_3\alpha_3\beta_4$ topology, where the fold is composed of one 3-stranded anti-parallel β-sheet and three α-helices packed against the β-sheet forming a hydrophobic core (Figure 5.3.1). Residues 98-116 form a COOH-terminal 18 amino acid arm-like coil region containing β4 from amino acids Phe_{108} to Leu_{111}. When the sequence was compared to that of E4-ORF3 identified in 13 other human adenoviral serotypes using CLUSTALW (*81, 82*), 35% of the sequence was fully conserved where the majority of the conserved residues were located in β_1, α_1, and α_3 (Figure 5.3.2).

From the crystallographic structures, we are able to see that the basic subunit of E4-ORF3 is a conserved dimer that is present in all 4 crystal forms. The dimerization interface is composed of β_1-β_3 from each monomer that together create a hydrophobic core with multiple hydrogen bonds. The COOH-terminal 18 amino acid coil region

containing β_4, forms an arm that crosses over the opposing open ends of the β_1-β_3 dimerization interface extending from β_1 of the first monomer to β_2 of the second monomer (Figure 5.3.1).

5.3.3 Similarity of E4-ORF3 to IS*Hp608* TnpA Transposase

Results from a DALI search (*83*), in which E4-ORF3 was compared to the three-dimensional structure of proteins in the Protein Data Bank (http://www.rcsb.org/pdb/), indicated that the E4-ORF3 monomer also contains conserved features found in the RNA recognition motif (RRM). The RRM, also known as the RNA-binding domain (RBD) or ribonucleoprotein domain (RNP), is one of the most abundant protein domains in eukaryotes, but is also found throughout all life kingdoms (*84, 85*). Similar to E4-ORF3, the RRM is composed of a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology as illustrated by the first structure containing the RNA recognition motif, the NH_2-terminal RRM of U1 small nuclear ribonuceloprotein A (PDBID- 1OIA) (Figure 5.3.3) (*86*). However, E4-ORF3 contains an additional α helix, α_2 located between α_1 and β_2, and it contains a 3-stranded anti-parallel β-sheet, instead of a 4-stranded β-sheet. This conserved motif containing a hydrophobic core may yield mechanistic insights in fiber formation.

One of the top search hits that was obtained from the DALI (*83*) 3D protein structure search was the *Helicobacter pylori* protein DNA transposase IS*Hp608* TnpA (PDBID-2A6M) with a Z-score of 9.2, and RMSD of 2.9 (*87*). Like that of E4-ORF3, IS*Hp608* TnpA also shares the similar conserved RRM motif. When E4-ORF3 was superimposed on IS*Hp608* TnpA the RRM fold was apparent where E4-ORF3 β_1, α_2, β_2,

β_3, and α_3 secondary structures overlapped (Figure 5.3.4 A). When the sequences were compared for secondary structure similarity based on the known coordinates using DSSP (*88*) and DALI (*83*), only about 7.8% of residues were conserved between the two proteins (Figure 5.3.4 B), 50% of these conserved residues are also conserved in other E4 ORF3 proteins found in 13 other human adenoviral serotypes as seen in the alignment shown in Figure 5.3.2. The majority of these residues are hydrophobic and are concentrated in α_2, β_2, and α_3.

Although the E4-ORF3 monomer is similar to that of IS*Hp608* TnpA, the mode of dimerization is quite different. The IS*Hp608* TnpA dimer is arranged in a head-to-tail orientation, whereas the E4-ORF3 is oriented head-to-head. In the E4-ORF3 dimer, the anti-parallel β-sheets from each monomer compose the dimeric interface, where hydrogen bonds and multiple hydrophobic residues create a stable dimer that is seen in all crystal structures solved. Since the dimeric interface is composed of such strong interactions we believe that this dimer could be the basic building block of the fiber scaffold. If we assume that this dimer is the building block of fiber formation, then one possible explanation of how E4-ORF3 forms a fiber within the nucleus would be to create a continuous hydrophobic core, in which multiple dimer subunits would dock to either end. Contacts between dimers would be formed through continuation of the anti-parallel β-sheet, through β_2 and β_4 of each dimer. For this to occur, the COOH terminal arm composed of the last 18 amino acids would have to rotate enough to allow the exposure of the hydrophobic core of the dimeric interface allowing for the β_2-β_4/β_4-β_2 interaction between two different dimer subunits (Figure 5.3.5). According to this theory, the

repeating unit in the hydrophobic anti-parallel β-sheet between two dimers would be $\beta_4\beta_1\beta_3\beta_2$:$\beta_4\beta_1\beta_3\beta_2$ creating one side of the anti-parallel β-sheet and $\beta_2\beta_3\beta_1\beta_4$:$\beta_2\beta_3\beta_1\beta_4$ composing the opposing side of the sheet, where ":" represents the interface between two dimers.

5.3.4 Discussion

Here the crystal structure of mutant human adenovirus E4-ORF3 has revealed a stable dimer which we assume to be the basic building block of fiber formation in the nuclear scaffold that is seen in previous studies using high resolution confocal microscopy (*80*). We have proposed one possible mechanism of fiber formation in which there is a continuation of the hydrophobic core formed by the dimeric interface, by the rotation of the COOH terminal arm such that another dimer can dock forming β_2-β_4 interactions, thus elongating the anti-parallel β-sheet on each side of the dimeric interface.

To test the role of the internal hydrophobic core in fiber formation, mutations can be made along the α_2 and α_3 helices where hydrophobic interactions are created between helices within the monomer. Residues that contribute to the interactions made between these helices are listed in (Figure 5.3.6). Since residue Ile_{31} on α_2 is a highly conserved hydrophobic Ile residue both in other Adenovirus serotypes (Figure 5.3.2) and in IS*Hp608* TnpA (Figure 5.3.4) mutating it may disrupt the hydrophobic core. Changing Ile_{31} to either an Asp or a Glu would introduce both bulk and negative charge to the hydrophobic core (Figure 5.3.7 A). The deletion of this conserved hydrophobic residue

and the introduction of both charge and size of the side chain, may be enough to disrupt the hydrophobic interactions. In order to recover this disruption of the two helices a complementary mutation could be made on the opposing helix α_3. Residue Ile_{84} is another conserved hydrophobic residue among Adenovirus serotypes (Figure 5.3.2); however, it is a Ser in IS*Hp608* TnpA (Figure 5.3.4). If Ile_{84} was mutated to Ser_{84}, the decrease in size of the side chain along with the introduction of positive charge, may complement the I31D and I31E point mutations, thus stabilizing the two helices and hydrophobic core (Figure 5.3.7 B, C). The I84S mutations when modeled places the side chain at a bonding distance at around ~3 Å away from D_{31} or E_{31}. Asp_{31} and Glu_{31} may also make contact with Ser_{87} thus providing additional stability to the two helices. Therefore if the I31D or I31E mutation was made to α_2 then either mutation alone should disrupt the hydrophobic core yet when coupled with the α_3 mutation I84S it should regain stability of the core through the compensating charge interactions made between Glu/Asp and Ser. In that experiment one could potentially turn on and off fiber formation through introduction of two point mutations.

5.3.5 Materials and Methods

Data Collection.

The diffraction data were collected for SeMet and native crystals at Stanford Synchrotron Radiation Lightsource beamline 9-2 to 2.06 Å ($P2_12_12_1$), 2.30 Å ($P2_1$), 2.84 Å ($P6_122$), and 2.20 Å ($C222_1$) resolution. The datasets were integrated and scaled using the program HKL-2000 version 0.98.692i (1).

Crystal Structure Determination and Refinement.

The crystal structure of E4-ORF3 was solved by MAD –multiple wavelength anomalous dispersion using selenium signal. Phase calculation was completed using the autoSHARP (2-3) software in the CCP4 program suite 6.1.13 (4). The selenium atom sites were located with the program SHELXD (5) and solvent flattening was performed in SOLOMON (6) yielding electron density maps that permitted the building of the protein chains using the programs ARP/wARP v. 7.1.0 (*50, 89-92*) and COOT version 0.6.2-pre-1 (7). The final solution contains 2 molecules in the asymmetric unit except for the dataset collected in the $C222_1$ space group which had only 1 molecule per asymmetric unit. Refinement of the structure was completed after multiple cycles of refinement using REFMAC V 5.5.0109 (8) utilizing various weighting terms and NCS (non-crystallographic symmetry) restraints and converged to R and R_{free} of 0.21 and 0.27 ($P2_12_12_1$), 0.22 and 0.28 ($P2_1$), 0.24 and 0.31 ($P6_122$), and 0.27 and 0.35 ($C222_1$) respectively.

The current data processing and refinement statistics are listed in Table 4.1.1. The root mean squared deviations in bond lengths and bond angles are 0.025 Å and 2.126 deg ($P2_12_12_1$), 0.022 Å and 2.001 deg ($P2_1$), 0.015 Å and 1.847 deg ($P6_122$), and 0.019 Å and 2.141 deg ($C222_1$), respectively. All water molecules have a density of 1 σ or greater in the $2F_o$-F_c map.

5.3.6 Figures and Tables

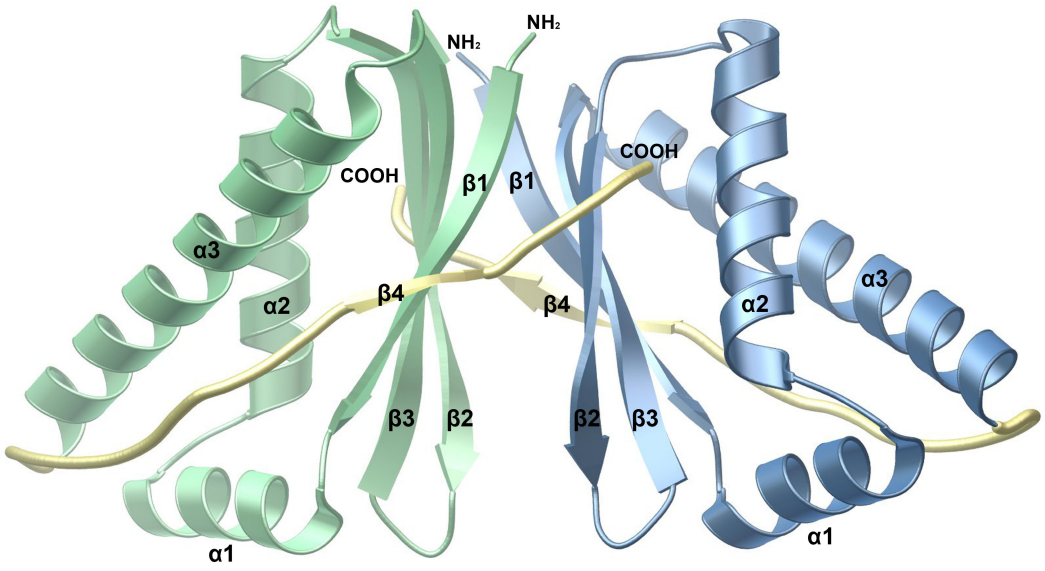

Figure 5.3.1: Crytallographic structure of E4 ORF3 dimer.
The E4 ORF dimer is composed of two monomeric subunits represented in green and blue. The COOH terminal arm containing β_4 is represented in yellow. Figures were prepared and rendered using Molscript Version 2.1.2 (*66*) and POV-Ray (*67*).

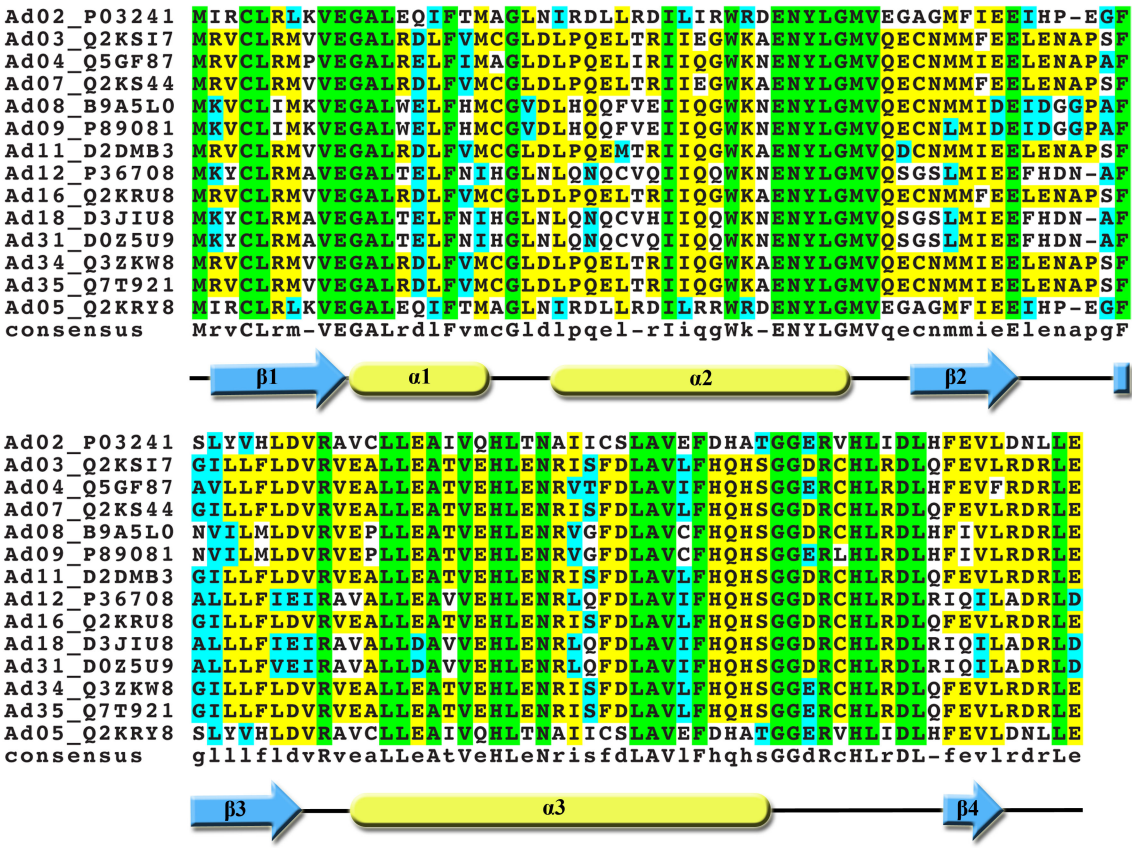

Figure 5.3.2: E4-ORF3 Sequence Alignment.
Sequence alignment to E4-ORF3 from 13 other human adenoviral serotypes. E4-ORF3 from human adenoviral serotype 5 is listed on the last line above the consensus. Fully conserved residues are represented green, identical residues are yellow, and similar residues are cyan. Uniprot IDs (*93, 94*) are listed to the right of the serotype, and the secondary structure of E4-ORF3 from Ad05 is list at the bottom. This figure was generated using CLUSTALW(*81, 82*) and BOXSHADE v. 3.3.1 (K. Hofman and M.D. Baron).

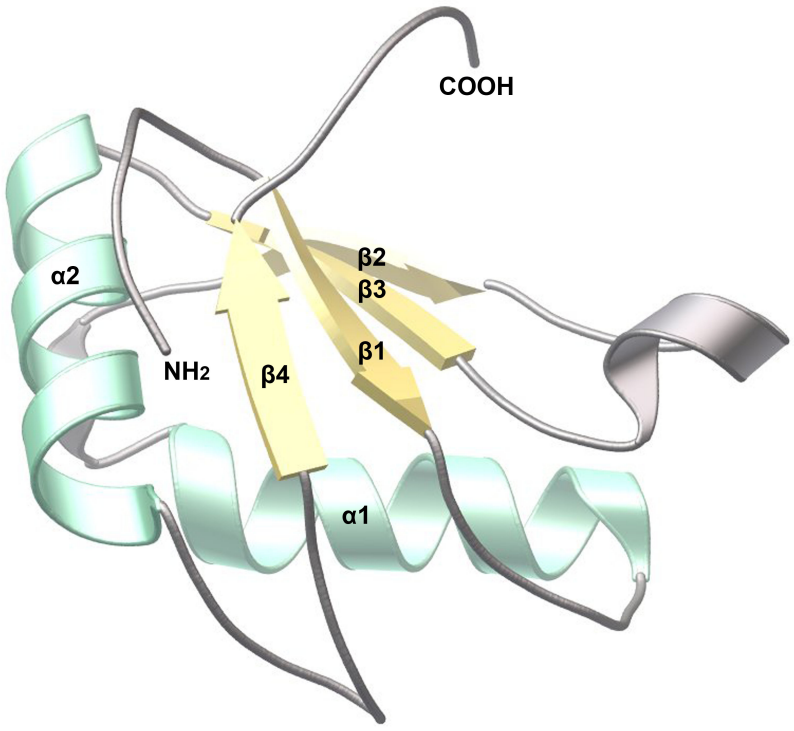

Figure 5.3.3: The RNA recognition motif as illustrated by the NH_2-terminal RRM of U1 small nuclear ribonucleoprotein A.
The RRM motif is composed of a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology as illustrated by the first structure containing the RNA recognition motif, the NH_2-terminal RRM of U1 small nuclear ribonucleoprotein A (PDBID- 1OIA) (*86*). Figures were prepared and rendered using Molscript Version 2.1.2 (*66*) and POV-Ray (*67*).

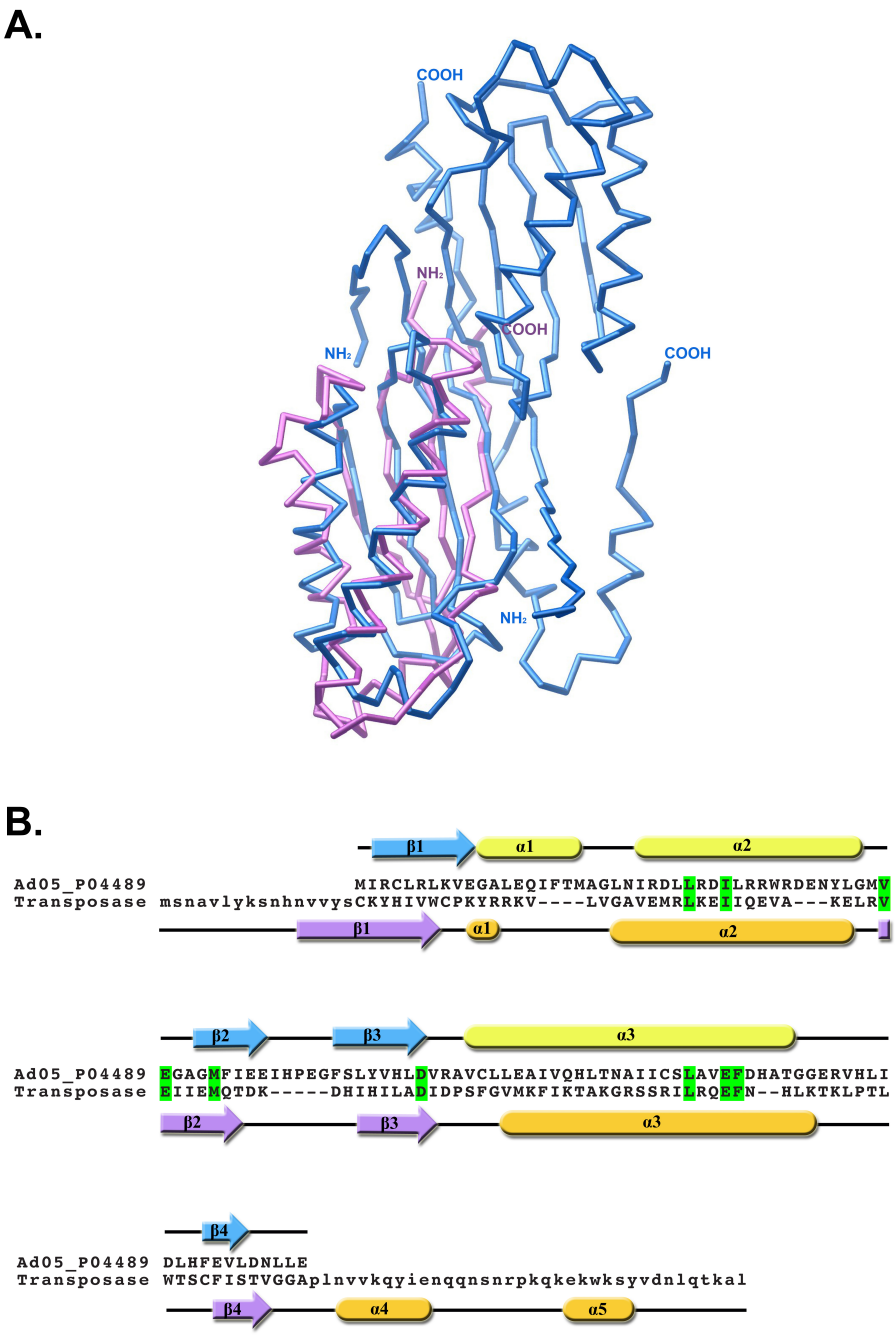

Figure 5.3.4: E4-ORF3 shares similar motif as DNA transposase IS*Hp608* TnpA. (A) E4-ORF3, represented in purple, superimposed on IS*Hp608* TnpA, represented in blue (PDBID: 2A6M) (*87*). (B) Sequence alignment of E4-ORF3 (top) and IS*Hp608* TnpA (bottom) based on secondary structural elements determined from their coordinates, where conserved residues are highlighted in green. Alignment created using the programs DSSP (*88*) and DALI (*83*).

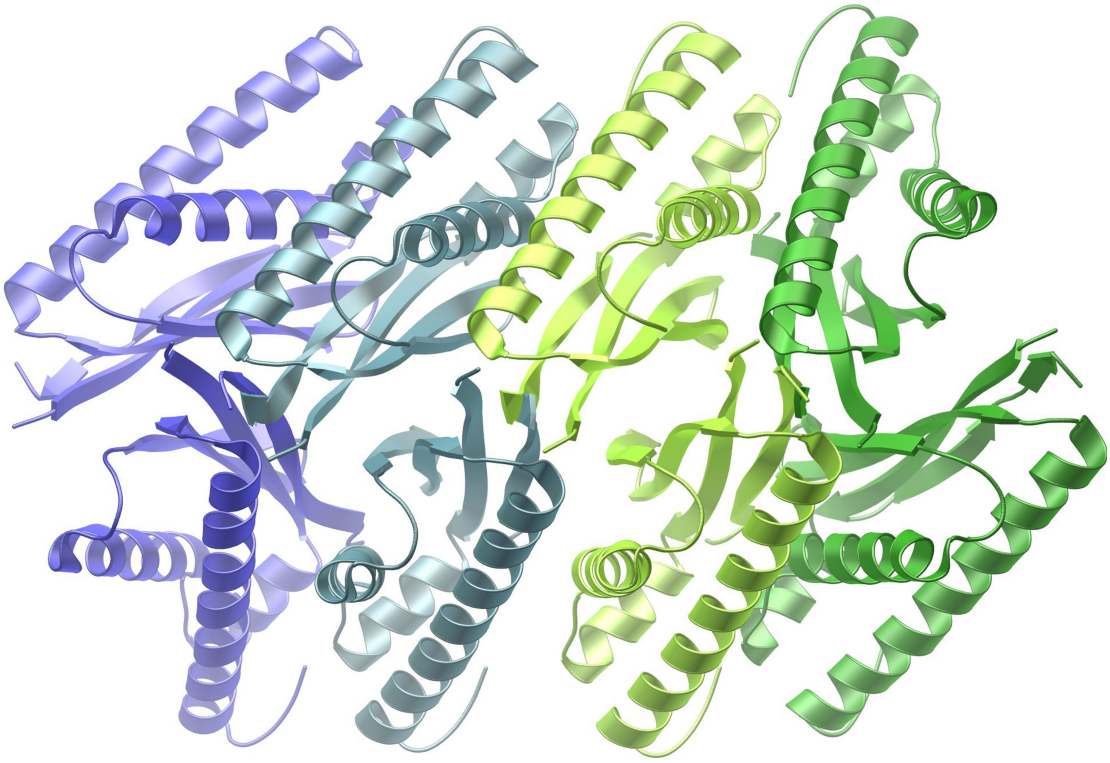

Figure 5.3.5: E4-ORF3 proposed mechanism of fiber formation and dimer subunit assembly.
Here an E4-ORF3 fiber is depicted through the assembly of 4 dimeric subunits oriented end-to-end such that the COOH terminal arm is rotated allowing for the β_2 of one dimer to make contact with the β_4 of the next dimer, resulting in the continuation of the anti-parallel β-sheet and hydrophobic core. All four dimers are represented in a different color: blue, cyan, light green, and green. This model was created using COOT version 0.6.2-pre-1 (*49*) and figures were prepared and rendered using Molscript Version 2.1.2 (*66*) and POV-Ray (*67*).

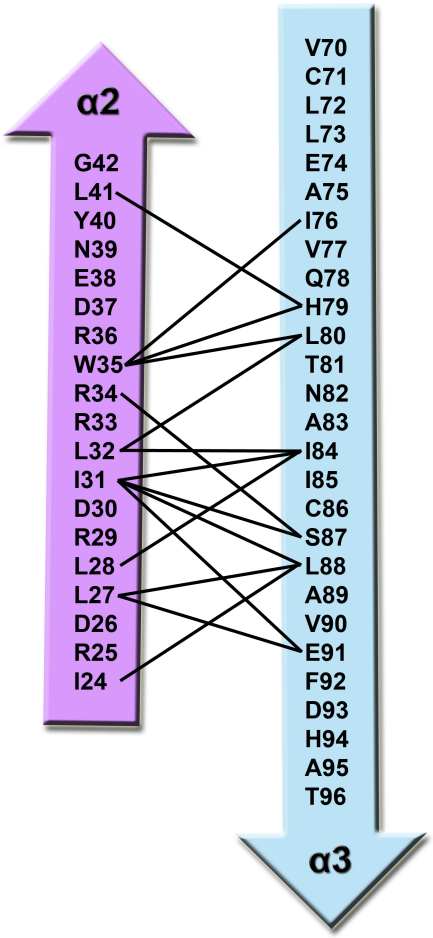

Figure 5.3.6: Contacts made between residues in E4-ORF3 helices α_2 and α_3.
α_2, represented in purple and α_3, represented as blue, are shown with contacts illustrated as lines between residues in each helix.

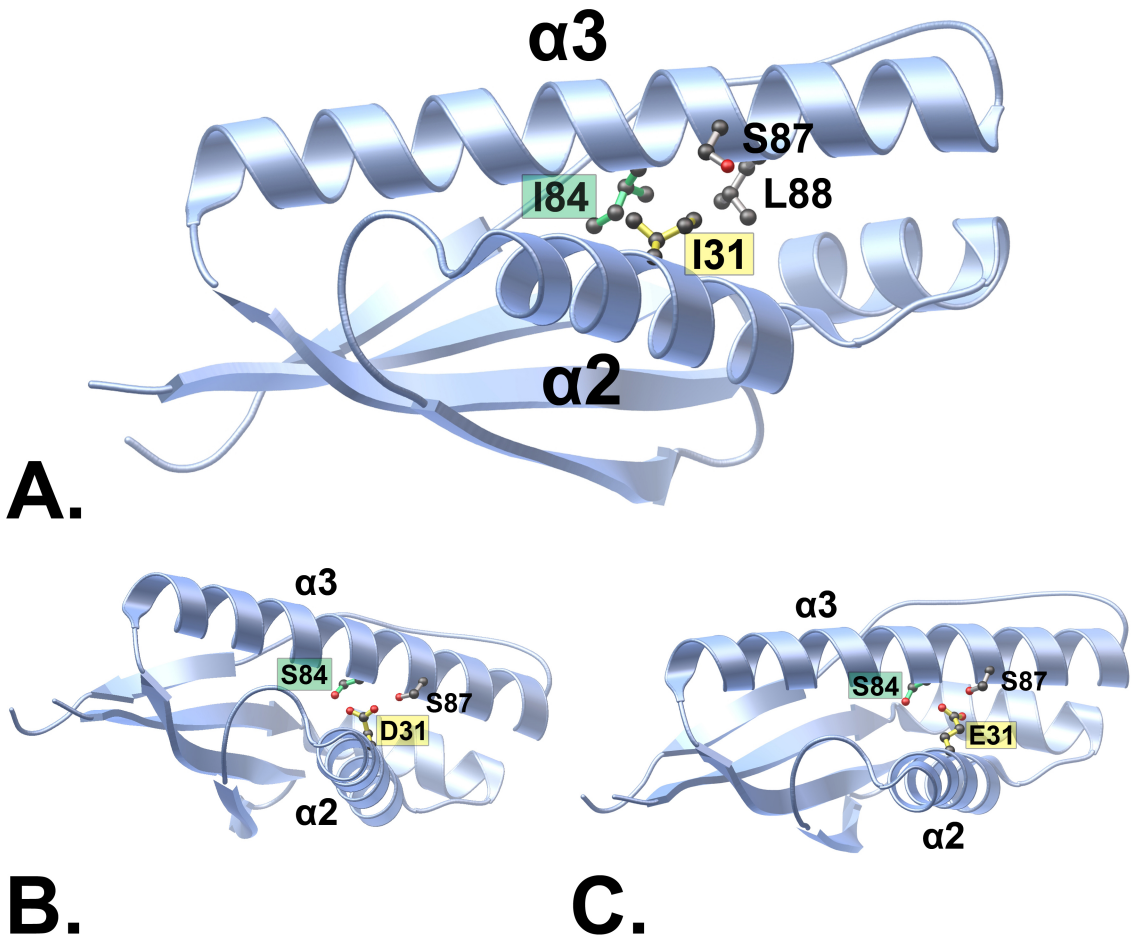

Figure 5.3.7: Point directed mutagenesis to manipulate E4-ORF3 hydrophobic core. (A) Monomeric E4-ORF3 where Ile31 (yellow) located on α2 and Ile84 (green), Ser87 and Leu88 side chains are represented as ball-and-stick. (B) Models of the E4-ORF3 I31D I84S double mutant and the (C) I31E I84S double mutant that illustrate possible contacts made within the hydrophobic core. Figures were prepared and rendered using Molscript Version 2.1.2 (66) and POV-Ray (67).

Table 5.3.1: E4-ORF3 Data collection and refinement statistics.

Data collection and model refinement statistics								
	E4-ORF3 Mutant (SeMet)		E4-ORF3 Mutant (native)		E4-ORF3 Mutant (native)		E4-ORF3 Mutant (native)	
space group	$P2_12_12_1$		$P2_1$		$P6_122$		$C222_1$	
unit cell parameters (Å, deg)	a = 56.43		a = 43.6		a, b = 81.267		a = 50.68	
	b = 67.22		b = 60.0		c = 181.187		b = 66.46	
	c = 87.34		c = 58.2		α, β = 90.0		c = *89.88*	
	α, β, γ = 90.0		α=γ = 90.0		γ =120.0		α, β, γ = 90.0	
			β = 96.0					
data collection statistics								
Beamline	SSRL BL9-2		SSRL BL9-2		SSRL BL9-2		SSRL BL9-2	
wavelength (Å)	0.979		0.979		0.979		0.979	
resolution range (Å)	50.00-2.06 (2.10-2.06)		100.0-2.30 (2.38-2.30)		50.00-2.84 (2.94-2.84)		100.00-2.20 (2.28-2.20)	
observed reflections	276893		43830		100511		27633	
unique reflections	21373		12223		8987		6310	
Redundancy	13.0	(8.2)[b]	3.6	(3.0)[b]	11.2	(10.9)[b]	4.4	(3.0)[b]
R_{merge} (%)	14.1	(57.8)[b]	6.9	(28.7)[b]	6.7	(31.4)[b]	8.1	(40.4)[b]
Mean I/σ (I)	20.4	(2.9) [b]	17.6	(3.0) [b]	37.8	(7.6) [b]	15.8	(2.23) [b]
completeness (%)	99.6	(93.9)[b]	89.8	(67.1)[b]	99.9	(99.8)[b]	79.0	(49.1)[b]
refinement statistics								
R_{cryst} (%)	20.94		21.8		24.0		26.84	
R_{free} (%)[a]	26.45		27.6		31.05		34.65	
average B factor (Å^2)	31.86		33.67		39.13		50.02	
root mean square deviation								
bond lengths (Å)	0.025		0.022		0.015		0.019	
bond angles (deg)	2.126		2.001		1.847		2.141	

[a] calculated from 5% of data not used in refinement. [b] Numbers in parentheses correspond to highest resolution shell (Å).

5.4 Acknowledgements

I would like to thank the fellowship support from the H. A. and Mary K. Chapman Charitable Trust and the Mary K. Chapman Foundation.

I acknowledge the following co-authors that have helped contribute to the work that was completed on hIMP3: Christian Klammt, Innokentiy Maslennikov, and Senyon Choe. Chapter 5, in part, will be prepared for submission for publication of the material and may appear as: "Klammt, C., Maslennikov, I., Blain, K.Y., and Choe, S. Solution NMR structure of human integral membrane proteins. 2010. *In preparation.*" The dissertation author was one of the primary investigators of this work.

I also acknowledge the following co-authors that have helped contribute to the work that was completed on GluN1 NTD: Anthony N. Farina, Tomohiko Maruo, Witek Kwiatkowski, Senyon Choe, and Terunaga Nakagawa. Chapter 5, in part, has in part been submitted for publication: "Farina, A. N., Blain, K. Y., Maruo, T., Kwiatkowski, W., Choe, S., and Nakagawa, T. 2010 Separation of domain contacts is required for heterotetrameric assembly of functional NMDA receptors. 2010. *Submitted.*" The dissertation author was one of the primary investigators and author of this paper.

In addition, I acknowledge the following co-authors that helped contribute to the work that was completed on E4-ORF3: Witek Kwiatkowski, Senyon Choe, Horng Der Ou and Clodagh O'Shea. The E4-ORF3 protein was expressed, purified and crystallized by Horng Der Ou from Clodagh O'Shea's laboratory. Chapter 5, in part, is currently being prepared for submission for publication of the material and may appear as: "Ou, H., Blain, K. Y., Kwiatkowski, W., Choe, S., and O'Shea, C. Structural Assembly of E4-

ORF3. 2010. *In preparation.*" The dissertation author was one of the primary investigators of this work.

5.5 References

1. Tusnady, G. E., Dosztanyi, Z., and Simon, I. (2004) Transmembrane proteins in the Protein Data Bank: identification and classification, *Bioinformatics 20*, 2964-2972.

2. Tusnady, G. E., Dosztanyi, Z., and Simon, I. (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank, *Nucleic Acids Res 33*, D275-278.

3. Raman, P., Cherezov, V., and Caffrey, M. (2006) The Membrane Protein Data Bank, *Cell Mol Life Sci 63*, 36-51.

4. Maslennikov, I., Klammt, C., Hwang, E., Kefala, G., Okamura, M., Esquivies, L., Mors, K., Glaubitz, C., Kwiatkowski, W., Jeon, Y. H., and Choe, S. (2010) Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis, *Proc Natl Acad Sci U S A 107*, 10902-10907.

5. Khorchid, A., Inouye, M., and Ikura, M. (2005) Structural characterization of Escherichia coli sensor histidine kinase EnvZ: the periplasmic C-terminal core domain is critical for homodimerization, *Biochem J 385*, 255-264.

6. Klammt, C., Lohr, F., Schafer, B., Haase, W., Dotsch, V., Ruterjans, H., Glaubitz, C., and Bernhard, F. (2004) High level cell-free expression and specific labeling of integral membrane proteins, *Eur J Biochem 271*, 568-580.

7. Krueger-Koplin, R. D., Sorgen, P. L., Krueger-Koplin, S. T., Rivera-Torres, I. O., Cahill, S. M., Hicks, D. B., Grinius, L., Krulwich, T. A., and Girvin, M. E. (2004) An evaluation of detergents for NMR structural studies of membrane proteins, *J Biomol NMR 28*, 43-57.

8. Kainosho, M., and Tsuji, T. (1982) Assignment of the three methionyl carbonyl carbon resonances in Streptomyces subtilisin inhibitor by a carbon-13 and nitrogen-15 double-labeling technique. A new strategy for structural studies of proteins in solution, *Biochemistry 21*, 6273-6279.

9. Battiste, J. L., and Wagner, G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data, *Biochemistry 39*, 5355-5365.

10. Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R., and Choe, S. (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression, *Science 307*, 1317-1321.

11. Wishart, D. S., and Sykes, B. D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data, *J Biomol NMR 4*, 171-180.

12. Luginbuhl, P., Szyperski, T., and Wuthrich, K. (1995) Statistical basis for the use of 13Calpha chemical shifts in protein structure determination. , *J. Magn. Reson. B., 109*, 5355-5365.

13. Guntert, P. (2004) Automated NMR structure calculation with CYANA, *Methods Mol Biol 278*, 353-378.

14. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, *Proc Natl Acad Sci U S A 94*, 12366-12371.

15. Salzmann, M., Pervushin, K., Wider, G., Senn, H., and Wuthrich, K. (1998) TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins, *Proc Natl Acad Sci U S A 95*, 13585-13590.

16. Salzmann, M., Wider, G., Pervushin, K., Senn, H., and W√ºthrich, K. (1999) TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins, *Journal of the American Chemical Society 121*, 844-848.

17. Keller, R. (2004) The Computer Aided Resonance Assignment Tutorial (Cantina Verlag, Goldau, Switzerland).

18. Cowan, W. M., Sudhof, T. C., and Stevens, C. F. (2001) *Synapses*, The Johns Hopkins Univ. Press, Baltimore, MD.

19. Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (1995) *Essentials of Neural Science and Behavior*, Appelton & Lange, East Norwalk, CT.

20. Hollmann, M., and Heinemann, S. (1994) Cloned glutamate receptors, *Annu Rev Neurosci 17*, 31-108.

21. Dingledine, R., Borges, K., Bowie, D., and Traynelis, S. F. (1999) The glutamate receptor ion channels, *Pharmacol Rev 51*, 7-61.

22. Brauner-Osborne, H., Egebjerg, J., Nielsen, E. O., Madsen, U., and Krogsgaard-Larsen, P. (2000) Ligands for glutamate receptors: design and therapeutic prospects, *J Med Chem 43*, 2609-2645.

23. Boulter, J., Hollmann, M., O'Shea-Greenfield, A., Hartley, M., Deneris, E., Maron, C., and Heinemann, S. (1990) Molecular cloning and functional expression of glutamate receptor subunit genes, *Science 249*, 1033-1037.

24. Keinanen, K., Wisden, W., Sommer, B., Werner, P., Herb, A., Verdoorn, T. A., Sakmann, B., and Seeburg, P. H. (1990) A family of AMPA-selective glutamate receptors, *Science 249*, 556-560.

25. Sommer, B., Burnashev, N., Verdoorn, T. A., Keinanen, K., Sakmann, B., and Seeburg, P. H. (1992) A glutamate receptor channel with high affinity for domoate and kainate, *EMBO J 11*, 1651-1656.

26. Moriyoshi, K., Masu, M., Ishii, T., Shigemoto, R., Mizuno, N., and Nakanishi, S. (1991) Molecular cloning and characterization of the rat NMDA receptor, *Nature 354*, 31-37.

27. Monyer, H., Sprengel, R., Schoepfer, R., Herb, A., Higuchi, M., Lomeli, H., Burnashev, N., Sakmann, B., and Seeburg, P. H. (1992) Heteromeric NMDA receptors: molecular and functional distinction of subtypes, *Science 256*, 1217-1221.

28. Mayer, M. L., Westbrook, G. L., and Guthrie, P. B. (1984) Voltage-dependent block by Mg2+ of NMDA responses in spinal cord neurones, *Nature 309*, 261-263.

29. Johnson, J. W., and Ascher, P. (1987) Glycine potentiates the NMDA response in cultured mouse brain neurons, *Nature 325*, 529-531.

30. MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J., and Barker, J. L. (1986) NMDA-receptor activation increases cytoplasmic calcium concentration in cultured spinal cord neurones, *Nature 321*, 519-522.

31. Kerchner, G. A., and Nicoll, R. A. (2008) Silent synapses and the emergence of a postsynaptic mechanism for LTP, *Nat Rev Neurosci 9*, 813-825.

32. Sudhof, T. C., and Malenka, R. C. (2008) Understanding synapses: past, present, and future, *Neuron 60*, 469-476.

33. Bliss, T. V., and Collingridge, G. L. (1993) A synaptic model of memory: long-term potentiation in the hippocampus, *Nature 361*, 31-39.

34. Choi, D. W., and Rothman, S. M. (1990) The role of glutamate neurotoxicity in hypoxic-ischemic neuronal death, *Annu Rev Neurosci 13*, 171-182.

35. Constantine-Paton, M., Cline, H. T., and Debski, E. (1990) Patterned activity, synaptic convergence, and the NMDA receptor in developing visual pathways, *Annu Rev Neurosci 13*, 129-154.

36. Cull-Candy, S., Brickley, S., and Farrant, M. (2001) NMDA receptor subunits: diversity, development and disease, *Curr Opin Neurobiol 11*, 327-335.

37. Kemp, J. A., and McKernan, R. M. (2002) NMDA receptor pathways as drug targets, *Nat Neurosci 5 Suppl*, 1039-1042.

38. Hollmann, M., Maron, C., and Heinemann, S. (1994) N-glycosylation site tagging suggests a three transmembrane domain topology for the glutamate receptor GluR1, *Neuron 13*, 1331-1343.

39. Scannevin, R. H., and Huganir, R. L. (2000) Postsynaptic organization and regulation of excitatory synapses, *Nat Rev Neurosci 1*, 133-141.

40. Schorge, S., and Colquhoun, D. (2003) Studies of NMDA receptor function and stoichiometry with truncated and tandem subunits, *J Neurosci 23*, 1151-1158.

41. Ulbrich, M. H., and Isacoff, E. Y. (2007) Subunit counting in membrane-bound proteins, *Nat Methods 4*, 319-321.

42. Kumar, J., Schuck, P., Jin, R., and Mayer, M. L. (2009) The N-terminal domain of GluR6-subtype glutamate receptor ion channels, *Nat Struct Mol Biol 16*, 631-638.

43. Sobolevsky, A. I., Rosconi, M. P., and Gouaux, E. (2009) X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor, *Nature 462*, 745-756.

44. Karakas, E., Simorowski, N., and Furukawa, H. (2009) Structure of the zinc-bound amino-terminal domain of the NMDA receptor NR2B subunit, *EMBO J 28*, 3910-3920.

45. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.

46. Collaborative, Computational, Project, Number, and 4. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr D Biol Crystallogr 50*, 760-763.

47. de La Fortelle, E., and Bricogne, G. (1997) Maximum-Likelihood Heavy-Atom Parameter Refinement for the Multiple Isomorphous Replacement and

Multiwavelength Anomalous Diffraction Methods, In *Methods in Enzymology*, pp 472-494, Academic Press, New York.

48. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP, *Methods Mol Biol 364*, 215-230.

49. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.

50. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr D Biol Crystallogr 53*, 240-255.

51. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr D Biol Crystallogr 60*, 2256-2268.

52. Matthews, B. W. (1968) Solvent content of protein crystals, *J Mol Biol 33*, 491-497.

53. Matthews, B. W. (1976) X-ray Crystallographic Studies of Proteins, In *Annual Review of Physical Chemistry*, pp 493 -493.

54. Wang, B. C. (1985) Resolution of phase ambiguity in macromolecular crystallography, *Methods Enzymol 115*, 90-112.

55. Leslie, A. (1987) A reciprocal-space method for calculating a molecular envelope using the algorithm of B.C. Wang, *Acta Crystallographica Section A 43*, 134-136.

56. Abrahams, J. P., and Leslie, A. G. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase, *Acta Crystallogr D Biol Crystallogr 52*, 30-42.

57. Kantardjieff, K. A., and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals, *Protein Sci 12*, 1865-1871.

58. Chruszcz, M., Potrzebowski, W., Zimmerman, M. D., Grabowski, M., Zheng, H., Lasota, P., and Minor, W. (2008) Analysis of solvent content and oligomeric states in protein crystals--does symmetry matter?, *Protein Sci 17*, 623-632.

59. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, In *Methods in Enzymology* (Carter Jr., C. W., and Sweet, R. M., Ed.), pp 307-326, Academic Press, New York.

60. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD, *Acta Crystallogr D Biol Crystallogr 58*, 1772-1779.

61. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *Journal of Applied Crystallography 26*, 283-291.

62. Cromer, D. T. (1965) Anomalous dispersion corrections computed from self-consistent field relativistic Dirac-Slater wave functions, *Acta Crystallogr 18*, 17-23.

63. Cromer, D. T., and Liberman, D. (1970) Relativistic calculation of anomalous scattering factors for X Rays, *J Chem Phys 53*, 1891-1898.

64. Cromer, D. T., and Liberman, D. A. (1981) Anomalous dispersion calculations near to and on the long-wavelength side of an absorption edge, *Acta Crystallographica Section A 37*, 267-268.

65. Brennan, S., and Cowan, P. L. (1992) A suite of programs for calculating x-ray absorption, reflection and diffraction performance for a variety of materials at arbitrary wavelengths, *Rev. Sci. Instrum. 63*, 850 - 853.

66. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *J Appl Cryst 24*, 946-950.

67. POV-Team. (1997) Persistence of Vision Ray Tracer v3.02. World Wide Web: http://www.povray.org.

68. Levine, A. J. (2009) The common mechanisms of transformation by the small DNA tumor viruses: The inactivation of tumor suppressor gene products: p53, *Virology 384*, 285-293.

69. O'Shea, C. C. (2005) DNA tumor viruses -- the spies who lyse us, *Curr Opin Genet Dev 15*, 18-26.

70. Kubbutat, M. H., Jones, S. N., and Vousden, K. H. (1997) Regulation of p53 stability by Mdm2, *Nature 387*, 299-303.

71. Bates, S., and Vousden, K. H. (1996) p53 in signaling checkpoint arrest or apoptosis, *Curr Opin Genet Dev 6*, 12-18.

72. Barak, Y., and Oren, M. (1992) Enhanced binding of a 95 kDa protein to p53 in cells undergoing p53-mediated growth arrest, *EMBO J 11*, 2115-2121.

73. Momand, J., Zambetti, G. P., Olson, D. C., George, D., and Levine, A. J. (1992) The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation, *Cell 69*, 1237-1245.

74. Chen, J., Marechal, V., and Levine, A. J. (1993) Mapping of the p53 and mdm-2 interaction domains, *Mol Cell Biol 13*, 4107-4114.

75. Hollstein, M., Sidransky, D., Vogelstein, B., and Harris, C. C. (1991) p53 mutations in human cancers, *Science 253*, 49-53.

76. Levine, A. J. (1990) The p53 protein and its interactions with the oncogene products of the small DNA tumor viruses, *Virology 177*, 419-426.

77. Michalovitz, D., Halevy, O., and Oren, M. (1991) p53 mutations: gains or losses?, *J Cell Biochem 45*, 22-29.

78. Bridge, E., and Ketner, G. (1989) Redundant control of adenovirus late gene expression by early region 4, *J Virol 63*, 631-638.

79. Huang, M. M., and Hearing, P. (1989) Adenovirus early region 4 encodes two gene products with redundant effects in lytic infection, *J Virol 63*, 2605-2615.

80. Soria, C., Estermann, F. E., Espantman, K. C., and O'Shea, C. C. (2010) Heterochromatin silencing of p53 target genes by a small viral protein, *Nature 466*, 1076-1081.

81. Higgins, D. G., Bleasby, A. J., and Fuchs, R. (1992) CLUSTAL V: improved software for multiple sequence alignment, *Comput Appl Biosci 8*, 189-191.

82. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res 22*, 4673-4680.

83. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res 38 Suppl*, W545-549.

84. Maruyama, K., Sato, N., and Ohta, N. (1999) Conservation of structure and cold-regulation of RNA-binding proteins in cyanobacteria: probable convergent evolution with eukaryotic glycine-rich RNA-binding proteins, *Nucleic Acids Res 27*, 2029-2036.

85. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) The Pfam protein families database, *Nucleic Acids Res 30*, 276-280.

86. Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A, *Nature 348*, 515-520.

87. Ronning, D. R., Guynet, C., Ton-Hoang, B., Perez, Z. N., Ghirlando, R., Chandler, M., and Dyda, F. (2005) Active site sharing and subterminal hairpin recognition in a new class of DNA transposases, *Mol Cell 20*, 143-154.

88. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577-2637.

89. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat Struct Biol 6*, 458-463.

90. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2002) ARP/wARP's model-building algorithms. I. The main chain, *Acta Crystallogr D Biol Crystallogr 58*, 968-975.

91. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003) ARP/wARP and automatic interpretation of protein electron density maps, *Methods Enzymol 374*, 229-244.

92. Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J., and Perrakis, A. (2004) Towards complete validated models in the next generation of ARP/wARP, *Acta Crystallogr D Biol Crystallogr 60*, 2222-2229.

93. (2010) The Universal Protein Resource (UniProt) in 2010, *Nucleic Acids Res 38*, D142-148.

94. Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B. E., Martin, M. J., McGarvey, P., and Gasteiger, E. (2009) Infrastructure for the life sciences: design and implementation of the UniProt website, *BMC Bioinformatics 10*, 136.

CHAPTER 6:

Discussion and Future Direction

Since the majority of drugs today target membrane proteins, it is clear that this family of proteins is of vital importance to our medical and scientific communities. Therefore it is to our best interest to gain an in depth understanding of how they function so that we cannot only improve the pharmaceutical therapeutics that we currently use, but also develop new ones.

Unfortunately, the progress of research in this particular field has been slow due to integral membrane protein complexity and the many obstacles associated with working with them. Although the three-dimensional structure of a protein is an essential component for the understanding of their function, not many have been determined due to the bottlenecks hampering these studies. However, here I have utilized protein crystallography and NMR to determine structures such as CCL14, hIMP3, and Glu1 NTD from different receptor signaling systems to gain a better understanding of biological function and aid in the progression of research in this field. In addition, I utilized crystallography to determine the adenoviral protein E4-ORF3 structure and propose a potential mechanism of oligomerization.

To increase the efficiency of integral membrane protein research it is crucial to improve the methods that we currently have and to exploit and develop new ways to study them. Here I have gathered and reviewed past research that has been completed on the protein engineering of bacterial histidine kinases to create chimeras through a domain-swapping strategy. These studies have illustrated that we have the ability to create new synthetic signal transduction circuits that can be useful in both medical and clinical applications. In addition to the use of chimeras to study integral membrane

proteins, I exploit the Mistic-fusion system to improve the levels of overexpression of the histidine kinase receptor EnvZ. Moreover, I demonstrate for the first time that Mistic not only boosts EnvZ expression levels, but maintains functionality both *in vitro* and *in vivo*. These findings illustrate that the Mistic-fusion system can be used as beneficial method to improve overexpression levels without diminishing the activity of the cargo protein. In addition to the Mistic-fusion system I employed the newly developed method of coupling cell-free expression, the CDL strategy, and fast NMR analysis to solve the structure of hIMP3 in a short period of time. Despite the past efforts to determine the structures of membrane proteins, only 20 human integral membrane protein structures are known so far (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html), thus illustrating the power of utilizing new methods.

Gaining a better understanding of the current techniques used in today's research is also important so that we can increase the efficiency in the future. Here I examined the problems that were encountered in molecular replacement during structure determination of GluN1 and found that the high solvent content of the crystals could have been the cause. Through the employment of new techniques and addressing the problems with current techniques we can potentially increase the number of integral membrane protein structures solved. This is of imperative need since the three-dimensional structures of membrane proteins and their signaling partners can yield detailed information, at atomic resolution, on the molecular mechanisms behind receptor signaling.