Iowa Research Online
The University of Iowa's Institutional Repository

University of Iowa
Iowa Research Online

Theses and Dissertations

2007

DNA sequence selectivity and kinetic properties of de novo designed metalloprotein dimers

Siu Wah Wong-Deyrup
University of Iowa

Copyright 2007 Siu Wah Wong-Deyrup

This dissertation is available at Iowa Research Online: http://ir.uiowa.edu/etd/185

Recommended Citation
Wong-Deyrup, Siu Wah. "DNA sequence selectivity and kinetic properties of de novo designed metalloprotein dimers." PhD (Doctor of Philosophy) thesis, University of Iowa, 2007.
http://ir.uiowa.edu/etd/185.

Follow this and additional works at: http://ir.uiowa.edu/etd

Part of the Chemistry Commons

DNA SEQUENCE SELECTIVITY AND KINETIC PROPERTIES OF *DE NOVO* DESIGNED METALLOPROTEIN DIMERS

by

Siu Wah Wong-Deyrup

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Chemistry
in the Graduate College of
The University of Iowa

May 2007

Thesis Supervisor: Associate Professor Sonya J. Franklin

ABSTRACT

In our efforts to engineer a DNA binding and cleaving protein with greater sequence discrimination, we have designed dimeric proteins derived from *engrailed* homeodomain and calmodulin. Previous research by our group has shown that a hydrolytically active lanthanide binding site can be incorporated into a DNA binding motif. To understand protein-DNA interaction and improve the sequence selectivity of the chimeric complex, two lanthanide-binding homodimers were designed and expressed.

One of the dimers, F2, is coupled together by a flexible polypeptide linker and the other, R7C, is a disulfide cross-linked cysteine mutant at the N-terminus. Studies of fluorescence of tryptophan residues document that the overall affinity for lanthanide and calcium is similar to traditional EF-hand peptides (1-10 μM). Metal titrations monitored by circular dichroism (CD) revealed that the secondary structures of the dimers contained a lower degree of α-helicity than the designed monomeric protein due to additional modifications, but because of their flexibility and their two active-site domain, hydrolytic activity was several folds faster than our previously designed proteins and peptides. Unlike earlier reports on our chimeras, F2 also demonstrated the capability to hydrolyze DNA in the presence of some biological relevant metal ions suggesting different cleavage mechanisms were carried out. Extensive DNA sequencing studies on cleavage patterns with oligonucleotide duplexes confirmed the unique sequence selectivity and kinetic properties of F2. Two *engrailed* homeodomain target sites, TAATTA, were favored for hydrolytic activity corresponding to one domain acting as a DNA anchor on the first target site while the other was an "opportunist" at recognizing the second site. Nonetheless, the hydrolytic behavior at the phosphodiester bond on a specific dsDNA sequence is in good agreement with the behavior of restriction endonucleases. Unlike restriction enzymes, metallated F2 has not only demonstrated the ability to cleave DNA plasmid, but it also excises the entire nucleotide on a selected sequence. This homodimer

is the first example of an active and selective hydrolytic artificial nuclease based on the modular turn substitution design approach that can be a potential template for genomic modification.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

DNA SEQUENCE SELECTIVITY AND KINETIC PROPERTIES OF *DE NOVO* DESIGNED METALLOPROTEIN DIMERS

by

Siu Wah Wong-Deyrup

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Chemistry
in the Graduate College of
The University of Iowa

May 2007

Thesis Supervisor: Associate Professor Sonya J. Franklin

Copyright by

SIU WAH WONG-DEYRUP

2007

All Rights Reserved

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Siu Wah Wong-Deyrup

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Chemistry at the May 2007 graduation.

Thesis Committee:_____
Sonya J. Franklin, Thesis Supervisor

_____
Donald M. Cannon

_____
Amnon Kohen

_____
Sarah C. Larsen

_____
Subramanian Ramaswamy

To my parents Kin Sang Lau and Cheuk Yut Wong

Strength does not come from physical capacity. It comes from its indomitable will.

Mahatma Gandhi

If we knew what we are doing it wouldn't be research.

Albert Einstein

ACKNOWLEDGMENTS

It has been a great journey in Iowa City for the last six years. During my graduate studies at the University of Iowa, my intellectual and emotional development has accelerated tremendously. Not only did the graduate experience enable me to understand and fulfill my passion in life, but I was also very lucky that I got to share this experience with my dearest friends and family. Special thanks to Dr. Sonya Franklin's caring and teaching that contributed the most to my growth as a person professionally and spiritually. She is the superhero who is a scientist, advisor, mother, and wife to whom I will always look up.

Having a highly diverse research group has given me an opportunity to be involved in teamwork and learn to love people as they are; I especially thank my cubicle-mates, Samantha Soebbing and Kinesha Harris. The three of us are so different, yet worked perfectly together due to our strength in different areas as it is true in the saying, "Three minds are better than one." I have learned as much of American culture as science from my amazing group members. The staff in the Chemistry department, who I cannot name all, Sharon Robertson, Janet Kugley, Amber Seaton, and Tim Koon, were utmost helpful and cheerful in many ways. Their warm smiles always brighten my day.

Assisting general chemistry lab for Dr. Russell Larsen was a wonderful teaching experience. He gracefully showed his students and reminded me of how exciting and fun chemistry is. Nonetheless, I am thankful to my committee, professors Cannon, Kohen, Larsen, and Ramaswamy, for their helpful research instruction and advice.

I am very gratified that I got to meet many awesome friends in Iowa City whose friendship kept my insanity intact throughout all these years. My "hot pot" buddies made me feel like lowa was a home away from home. My deepest appreciation goes to my host family in Florida, Tommy, Caroline, and Alida Gertz-Robert who took me in and loved me as part of their family for the entire six years when I could not return home.

Caroline and Alida showed me how to be “kick-ass” as a woman both literally and theoretically. They have become my second family whom I will always cherish.

At home, I would like to thank you for the encouragement and perseverance of my aunt, Auntie Goat, who is no doubt the inspiration for my success nowadays. Nonetheless, no one else can compare to my devoting and loving parents, Kin Sang Lau and Cheuk Yut Wong. I would not be the person I am now without them as they have given up so much in order to provide me the best quality of life. They have showed me to be strong through tough times and celebrate the happiness in life.

Last but not the least, I feel absolutely lucky that I have met my husband, Stephen Deyrup, who is my best friend and my companion in life. We have gone through many huddles together throughout all these years and his unconditional love, support, and trust for me remain untainted. His smile is my sunshine when the sky is grey; his arms are my shelter; and his presence in my life shows me that the glass is more than half full.

ABSTRACT

In our efforts to engineer a DNA binding and cleaving protein with greater sequence discrimination, we have designed dimeric proteins derived from *engrailed* homeodomain and calmodulin. Previous research by our group has shown that a hydrolytically active lanthanide binding site can be incorporated into a DNA binding motif. To understand protein-DNA interaction and improve the sequence selectivity of the chimeric complex, two lanthanide-binding homodimers were designed and expressed.

One of the dimers, F2, is coupled together by a flexible polypeptide linker and the other, R7C, is a disulfide cross-linked cysteine mutant at the N-terminus. Studies of fluorescence of tryptophan residues document that the overall affinity for lanthanide and calcium is similar to traditional EF-hand peptides (1-10 μM). Metal titrations monitored by circular dichroism (CD) revealed that the secondary structures of the dimers contained a lower degree of α-helicity than the designed monomeric protein due to additional modifications, but because of their flexibility and their two active-site domain, hydrolytic activity was several folds faster than our previously designed proteins and peptides. Unlike earlier reports on our chimeras, F2 also demonstrated the capability to hydrolyze DNA in the presence of some biological relevant metal ions suggesting different cleavage mechanisms were carried out. Extensive DNA sequencing studies on cleavage patterns with oligonucleotide duplexes confirmed the unique sequence selectivity and kinetic properties of F2. Two *engrailed* homeodomain target sites, TAATTA, were favored for hydrolytic activity corresponding to one domain acting as a DNA anchor on the first target site while the other was an "opportunist" at recognizing the second site. Nonetheless, the hydrolytic behavior at the phosphodiester bond on a specific dsDNA sequence is in good agreement with the behavior of restriction endonucleases. Unlike restriction enzymes, metallated F2 has not only demonstrated the ability to cleave DNA plasmid, but it also excises the entire nucleotide on a selected sequence. This homodimer

is the first example of an active and selective hydrolytic artificial nuclease based on the modular turn substitution design approach that can be a potential template for genomic modification.

TABLE OF CONTENTS

LIST OF TABLES xi

LIST OF FIGURES xii

LIST OF ABBREVIATIONS xx

CHAPTER

1 INTRODUCTION 1

1.1 Background 1
1.2 Calcium Binding Sites and EF-hand motif 3
1.3 Substitution of Calcium with Lanthanides 4
1.4 DNA interaction with HTH Protein 5
1.5 DNA Hydrolytic Activity for Enzyme Design 7
1.6 The Benefit of Dimers 8
1.7 *De Novo* Protein Design 9
1.8 Description of the Thesis Contents 10

2 PHYSICAL CHARACTERIZATION OF A CHIMERIC DIMER, F2 18

2.1 Introduction 18
2.2 Experimental Procedures 20
2.2.1 Chimeric Design 20
2.2.2 DNA Cloning 20
2.2.3 Protein Purification 21
2.2.4 Tryptophan Fluorescence Titrations 22
2.2.5 Circular Dichorism Titrations 23
2.2.6 Thermal Denaturation 23
2.2.7 DNA Gel Shift Assays 24
2.3 Results and Discussion 25
2.3.1 Protein Design and Purification 25
2.3.2 Tryptophan Fluorescence Titrations 25
2.3.3 Secondary Structure Analysis by Circular Dichroism 26
2.3.4 DNA Binding Studies 28
2.4 Conclusions 28

3 NUCLEASE ACTIVITIES OF THE CHIMERIC DIMER, F2 39

3.1 Introduction 39
3.2 Experimental Procedures 40
3.2.1 Reagents and Proteins 40
3.2.2 DNA Plasmid Cleavage Assay 41
3.2.3 pH-dependent DNA Cleavage Assay 41
3.2.4 Time-dependent DNA Cleavage Assay 42
3.2.5 Metal-dependent DNA Cleavage Assay 43
3.2.6 BNPP Kinetic Studies 43
3.2 7 Phosphate Inhibition Studies 44
3.2.6.1 Investigation with BNPP 44
3.2.6.2 Investigation with DNA Plasmid 45

3.3 Results and Discussion 46
3.3.1 DNA Plasmid Cleavage Assay 46
3.3.2 pH-dependent DNA Cleavage Assay 47
3.3.3 Time-dependent DNA Cleavage Assay 48
3.3.4 Metal-dependent DNA Cleavage Assay 49
3.3.5 BNPP Kinetic Studies 50
3.3.6 Phosphate Inhibition Studies 51
3.4 Conclusions 53

4 F2-DNA INTERACTIONS AND THE EXTENT OF SEQUENCE SELECTIVITY 65

4.1 Introduction 65
4.2 Experimental Procedures 67
4.2.1 Electrophoresis Mobility Shift Assay (EMSA) 67
4.2.1.1 PCR-coupled EMSA 67
4.2.1.2 EMSA for Binding Site Selection 67
4.2.1.3 CNBr Sepharose DNA Selection 69
4.2.2 ^{32}P-labeled Sequencing Gels 70
4.2.2.1 Sequencing DNA by Maxam-Gilbert Technique 70
4.2.2.2 Mechanism Comparison via DNase I Footprinting 70
4.2.2.3 Cleavage Assay with ^{32}P-labeled DNA Duplexes 71
4.2.3 Circular Dichroism of F2 in the Presence of DNA Duplexes 72
4.2.4 ^{32}P-labeled Kinetics Assay 72
4.2.5 ^{32}P-labeled at 3′- and 5′-end Site Matching Assay 73
4.3 Results and Discussion 74
4.3.1 Electrophoresis Mobility Shift Assay of C2 and F2 74
4.3.1.1 EMSA for Binding Site Selection with F2 74
4.3.1.2 CNBr Sepharose DNA Selection with F2 & C2 75
4.3.2 ^{32}P-labeled Sequencing Gels 77
4.3.2.1 Cleavage assay with ^{32}P-labeled DNA Duplexes 77
4.3.2.2 Hydrolytic versus Oxidative Cleavage Mechanism 80
4.3.3 Circular Dichroism of F2 in the Presence of DNA Duplexes 81
4.3.4 ^{32}P-labeled Kinetics Assay 82
4.3.5 ^{32}P-labeled at 3′- and 5′-end Site Matching Assay 83
4.4 Conclusions 84

5 DNA SEQUENCE PREFERENCE OF P3W AND P5B 101

5.1 Introduction 101
5.2 Experimental Procedures 103
5.2.1 Peptides Design and Synthesis 103
5.2.2 DNA and Other Reagents 104
5.2.3 Circular Dichroism in the Presence/Absence of DNA 104
5.2.4 PCR-coupled EMSA 105
5.2.5 EMSA for Binding Site Selection 105
5.3 Results and Discussion 107
5.3.1 Design and Characterization of Chimeric HTH-EF-hand Peptides 107
5.3.2 Selection of High Affinity DNA Sites 108
5.3.3 Consensus Target Sequences 109
5.3.4 Peptide structure as a Function of DNA Binding 111

5.4 Conclusions 113

6 PHYSICAL CHARACTERIZATION OF A CHIMERIC DIMER, R7C ... 124

6.1 Introduction 124
6.2 Experimental Procedures 125
6.2.1 Chimeric Design 125
6.2.2 DNA Cloning 125
6.2.3 Protein Purification 126
6.2.4 Protein Dimerization 127
6.2.5 Tryptophan Fluorescence Titrations 127
6.2.6 Circular Dichroism Titrations 127
6.2.7 DNA Plasmid Cleavage Assay 128
6.3 Results and Discussion 128
6.3.1 Chimeric Design and DNA Cloning 128
6.3.2 Protein Purification and Dimerization 129
6.3.3 Tryptophan Fluorescence Titrations 130
6.3.4 Circular Dichroism Titrations 131
6.3.5 DNA Plasmid Cleavage Assay 131
6.4 Conclusions 132

7 CONCLUSIONS 142

APPENDIX: DNA CLONING AND PROTEIN PURIFICATION OF CROM 144

REFERENCES 161

LIST OF TABLES

Table

1.1 Natural occurrences of oligomeric proteins in *Escherichia coli*. This table is adapted from reference 69 16

3.1 Kinetic rates of BNPP cleavage for metalloprotein/peptides 59

4.1 DNA sequences for ^{32}P-labeled sequencing gel. Each group has identical potential cleavage sites represented in blue with different base pair gaps between two sites as shown underlined. The change of sequence pattern was indicated in orange 93

4.2 The rate of reaction of Eu_2F2 with different DNA sequences observed from ^{32}P-labeled polyacrylamide gel electrophoresis. The red letters indicated the changes between sequences 97

5.1 Alpha-helical content of peptides in the presence and absence of DNA, based on molar ellipticity at 222 nm. 100% helical content was assumed to be 30,000 (deg•cm^2)/dmol that is adapted from reference 132. Note that these values are necessarily estimates, since small changes in DNA structure could also contribute to negative intensity in the UV 123

6.1 A summarized table of the α-helicity of our chimeric designs, a standard (myoglobin), and wt *engrailed* homeodomain. Since the CD measurement of myoglobin was less helical than that was projected from its PDB structure, the α-helicity was normalized based on the known α-helicity of myoglobin at pH 7.0 139

LIST OF FIGURES

Figure

1.1 (A) X-ray structure of the calmodulin (1CLL). The Ca^{2+} ions are shown in white spheres as the calcium-binding loops are represented in light blue. (B) The EF-hand proteins are based on a motif characterized by a helix-loop-helix structure. The motif's name is derived from its thumb and forefinger pair on a hand. These figures are adapted from reference 20..... 12

1.2 Coordinating residues of an EF-hand site from parvalbumin. The coordinating position of a typical EF-hand site is illustrated as the residues at loop positions 1, 3, 5, 7, 9, and 12 are coordinated to the bound metal ions. Residue 9 is bound through water as the "gateway position". This figure is adapted from reference 28..... 13

1.3 The crystal structure of *engrailed* homeodomain (2HDD.pdb). Helix 1 is in red while helix 2 is shown in orange. Helix 3 is the DNA-binding domain which is represented in green..... 14

1.4 Diagram showing major groove contacts by *engrailed* homeodomain mostly from helix 3. DNA phosphates are represented by circles connected by lines. DNA base contacts are indicated by a thick line. Hydrogen bonds are depicted by broken lines. Water is shown as W in a circle, a filled circle, or a circle and square. DNA contacts with the N-terminal residues, R3 and R5, are not shown. This figure is adapted from reference 51..... 15

1.5 The EF-hand loop (1OSA) and *engrailed* homeodomain (2HDD) are shown in blue/green and red/yellow respectively. The overlaying of two motives created a chimeric monomer model, C2 (in the middle). A chimeric dimer protein model, F2 (in green), is designed to be covalently connected through a flexible residues linker (in black and not helical) and binds to the major groove of DNA (in grey). Both structures of C2 and F2 models are only determined by Siwss PDB viewer..... 17

2.1 Three-dimensional view of the chimeric monomer design, C2. The DNA-binding HTH motif (blue) and the Ca-binding loop EF-hand motif (pink) with calcium ion (black) are structurally similar at α-α corner. Therefore, the turn sequence of the HTH motif can be modularly replaced by an EF-hand motif. 30

2.2 (A) The amino acid sequence of chimeric F2. Parent protein sequence is indicated in bold for the homeodomain and shaded for the EF-hand Ca-binding loop. The 7-residue linker between two monomers is boxed. (B) Schematic representation of our chimeric dimer design with two approaches..... 31

2.3 A schematic map of F2 DNA cloning. The cloning involved three stages: (1) Amplification of C2 with a *Xho*I site at both termini. (2) Ligation of the amplified C2 into the original vector which already contains one C2. (3) Insertion of the linker and the deletion of a start and a stop codons that are adjacent to the linker via multi-site mutagenesis technique. *Nde*I and *Xho*I are unique restriction enzyme sites for pET21a(+). N and C indicate the direction of the termini. M represents the start condon, methionine, while • represents the stop condon 32

2.4 Oligonucleotide sequences were ordered and used as primers during PCR amplification and multi-site mutagenesis. The sites for restriction enzyme digestion are underlined. The DNA encoding for the 7-residue linker, GGTGGAG in amino acids, is shaded. 32

2.5 SDS-PAGE images of F2 purification steps. Lane M is prestained broad range marker. Lane 1 shows non-induced protein while lane 2 is induced one. The French pressed pellet was kept lane 3 and extracted by CB-1000 is lane 4. The protein was further purified by PEI precipitation (lane 5) and Heparin column (lane 6). F2 was finally purified through size exclusion column to 99% purity (lane 7) 33

2.6 The MALDI-TOF spectrum of purified F2. The calculated molecular weight of F2 is 15820 *m/z* via DNASTAR software and the measured molecular weight of F2 is $(F2+H)^+$ at 15800 *m/z*, $(F2+H)^{2+}$ at 7911*m/z*, and $(F2+H)^{3+}$ at 5274 *m/z* which is within the accuracy range 33

2.7 Fluorescence titrations of F2 with metal ions. The fluorescence intensity of Trp was followed at 350 nm (280 nm excitation) as a function of added metal (5 μM F2/10 mM HEPES pH 7.0 and 50 mM NaCl). The solid lines represent a linear non-least square algorithm fitting to a 1:1 association model, from which K_d was obtained. (*Inset*) Emission spectra showed throughout one titration, from 0 μM $LuCl_3$, upper most line to 100 μM $LuCl_3$, bottom most line. 34

2.8 CD spectra of apoF2 (black), Eu_2F2 (blue), Ca_2F2 (green), and Mg_2F2 (red). The sample contained 15 μM F2 in 10 mM HEPES pH 7.0 and 50 mM NaCl 35

2.9 The thermal denaturation of apoF2 (black), Eu_2F2 (blue), Ca_2F2 (green) observed at A_{222}. The melting temperatures for apoF2, Eu_2F2, and Ca_2F2 are 45, 54, and 33 °C respectively fitted with a four-parameter algorithm which equation is adapted from reference 73 36

2.10 Agarose gel electrophoresis of supercoiled pTYB1 DNA bound with (A)ApoF2, (B)Ca_2F2, (C)Mg_2F2, and (D)Eu_2F2 in a buffer of 10 mM HEPES pH 7.0 and 50 mM NaCl. Lane M: Lambda Hind III fragment marker (Novitrogen). The increasing number lanes are indicated as shown with increasing concentrations of protein binding to DNA. 37

2.11 Agarose gel of DNA binding assay in the presence of apoF2. Lane M is a λ-Hind III fragment marker for size measurement. Lane 0 contained DNA plasmid only while lanes 4-64 represent the concentration of apoF2 in μM added to DNA plasmid. 38

2.12 A plot of % gel mobility as a function of the concentration of metallo-/apo-F2. The gray arrow indicates the estimated overall binding affinity for F2 which is 900 nM. 38

3.1 Agarose gel electrophoresis of supercoiled pTYB1 DNA cleaved by increasing concentrations of apoF2 or metallated F2 (10 mM HEPES pH 7.0 and 50 mM NaCl as indicated). Lane M: Lambda Hind III fragment marker (Novitrogen). Lane 0: Supercoiled pTYB1 DNA in the absence of protein. The remaining lanes are indicated as shown with increasing concentration of protein which has interacted with DNA. 54

3.2 (A) The structural comparison of projected model and NMR structure of LaC2 demonstrated that the orientation of helix 3 was flipped to the opposite side of helix 1. The metal binding loop and the metal ion are represented in blue and gray respectively. (B) A 90° turn view of (A) 55

3.3 Agarose gel electrophoresis of supercoiled pTYB1 DNA treated with 1.50 μM metallated F2 that was incubated at different pHs (10 mM HEPES pH 7.0 and 50 mM NaCl). Lane C: Supercoiled DNA without protein. Lane 4-13 indicates the pH condition of DNA-protein complex. The lanes with cleavage activity were indicated in orange letters 56

3.4 Agarose gel electrophoresis of supercoiled pTYB1 DNA treated with 643 nM Ca_2F2 that was incubated for different lengths of time (10 mM HEPES pH 7.0 and 50 mM NaCl). Lane C: Supercoiled DNA before treatment with the Geneclean kit. Lane L: Linear pTYB1 DNA digested with *EcoR*I. Lane 0-7 indicate the amount of incubation time in hours. 57

3.5 A graph plotting percentage of agarose gel band intensity from figure 3.3 as a function of time indicating the appearance of a DNA cleaved by Ca_2F2 over time 58

3.6 A plot of the pseudo-first-order rate (k_{obs}) as a function of Ca_2F2 concentration observed via agarose gel electrophoresis. The data was fitted with Michaelis-Menten kinetics equation (blue dotted line) to obtain V_{max}' and K_M. The error bar on the y-axis is the standard deviation over 3 trials. 58

3.7 Agarose gel electrophoresis of supercoiled pTYB1 DNA treated with 1.50 μM metallated F2 incubated at 37 °C. Lane M is the λ Hind III DNA fragment marker and lane DNA is the supercoiled DNA without protein. 60

3.8 A time course experiment of BNPP cleavage at various concentrations of metalloprotein. Absorbance is monitored at 400 nm coinciding with the product of cleavage activity, nitrophenol 61

3.9 The hydrolytic activities of BNPP with metalloproteins and metal ions were compared 62

3.10 The % hydrolytic activity of Eu_2F2 was inhibited by the addition to phosphate in the BNPP solution 63

3.11 Agarose gel electrophoresis of pTYB1 plasmid with Eu_2F2 to determine the level of inhibition of cleavage activity in the presence of phosphate or europium. 64

4.1 (A) DNA library with eleven randomized base pair (N_{11})(green letters) and PCR primers for electromobility gel shift assay (EMSA). The TAATTA site is located on the left of N_{11}. The primers (arrows) were ^{32}P-radiolabeled at the 5′-terminus prior to PCR. The restriction enzyme sites, *BamHI* and *HindIII*, are shown in gray boxes. The other restriction enzyme that is shown in blue box is used for the insertion verification of the DNA library that was inserted into pUC19 plasmid. (B) DNA library with eight randomized base pair (N_8)(green letters) and PCR primers for electromobility gel shift assay (EMSA). The primers (arrows) were ^{32}P-radiolabeled at the 5′-terminus prior to PCR. The restriction enzyme sites, *BamHI* and *EcoRI*, are shown in gray boxes 87

4.2 Polyacrylamide gel shift assay of DNA binding library with increasing amounts of Ca_2F2. The box containing peptide-bound DNA was excised from the gel, and the DNA purified and amplified for the next round. 88

4.3 (A) The alignment of enriched DNA library binding to Ca_2F2. A total of 51 sequences were selected from 2 trials. The frequencies of each nucleotide for the alignment were summarized in (B) and the consensus sequence is 5′-T(C/G)(C/G)(C/G)(C/G)-3′ 89

4.4 (A) The alignment of enriched DNA library binding to CaC2. A total of 31 sequences were selected from 2 trials. The frequencies of each nucleotide for the alignment were summarized in (B) and the consensus sequence is 5′-GGC-3′ 90

4.5 (A) The alignment of enriched DNA library binding to Ca_2F2. A total of 24 sequences were selected from 2 trials. The frequencies of each nucleotide for the alignment were summarized in (B) and the consensus sequence is 5′-(A/T)CC(C/G)(C/T)-3′ 91

4.6 ^{32}P-labeled polyacrylamide gel of DNA sequences, Non specific, Wong 8, and Wong 5TA in the presence of Eu_2F2, Ca_2F2 and apoF2 represented as lane Eu, Ca, and Apo respectively. The DNA nucleotides were sequenced using Maxam-Gilbert sequencing technique, as indicated on lane C and G+A. The cleavage sites were represented in orange letters 92

4.7 The cleavage sites of Eu_2F2 and Ca_2F2 with different DNA sequence patterns. The arrows indicated the cleavages (black: sequence as shown, grey: the complimentary sequence). The length of the arrow represented the intensity of the cleavage band on ^{32}P-labeled acrylamide gels. The base pair gap between two TAAT(NN) sites was underlined. The change of sequence pattern from the previous group was shown in orange letters 94

4.8 A three-dimensional view of the cleavage sites of group A to D with various base pair gaps (indicated on top of the picture) between two TAATNN sites. The cleavage intensity was related to the warm color of the nucleotide. The warmer the color is; the more intensive cleavage band was on the ^{32}P-labeling polyarcylamide gel 95

4.9 A polyacrylamide gel electrophoresis of Wong 5Rev incubated with Eu_2F2 at different time intervals. Lane Eu^{3+} represented the incubation of Wong 5Rev with $EuCl_3$ only. Lane DNA is the incubation of Wong 5Rev only. Lane 0-60: time of incubation in the presence of Eu_2F2. Lane G+A and C+T showed the DNA underwent Maxam-Gilbert sequencing reaction to be sequenced as the markers 96

4.10 (A) Polyacrylamide gel electrophoresis of cleavage products of a 5′-end-labeled 21 bp DNA duplex, Wong22F. Lane 1, 2, 5, and 6: Maxam-Gilbert sequencing lanes producing phosphate termini (p = 3′OPO_3 termini, blue arrows). Lane 3 and 7: DNase I treated DNA producing hydroxyl termini (h = 3′-OH termini, red arrows). The chimera Ca_2F2 cleaved DNA, producing only 3′-OH termini for both complimentary strands of Wong3Rev. (B) The arrows (blue for Maxam-Gilbert sequencing and red for DNase I treated DNA) represented the cleavage sites from the view of a DNA molecule structure 98

4.11 CD spectra of apoF2 in the presence and absence of DNA duplex 99

4.12 The subtraction of apoF2 in the presence of different DNA duplex sequences, Non-specific, cs 1, and Wong 3Rev, with the CD spectra of DNA duplexes..... 99

4.13 (A) The schematic model of the cleavage site observed from 3′-end- and 5′-end labeling on the same sequence of a DNA duplex, Wong3Rev. The dashed lines indicated the cleavage sites observed from both labeled ends polyacrylamide gel electrophoresis. (B) The colors that were coordinated with (A) were indicated in a three-dimensional view on a DNA duplex 100

5.1 (A) Schematic depiction of the chimeric peptide design. The DNA-binding helix-turn-helix motif (purple) and the Ca-binding EF-hand motif (green) are structurally analogous alpha-alpha corners, and thus can be combined by modularly replacing turn sequences. (B) Sequences of designed chimeric peptides P3W (33 aa, based on *engrailed*) and P5b (32 aa, based on *antennapedia*). The EF-hand metal-binding loop (designated +1 through +12) is shaded in each case..... 117

5.2 DNA library with eight randomized base pairs (N_8) and PCR primers for EMSA binding analysis. The primers (arrows) were ^{32}-P-radiolabeled at the 5′-terminus prior to PCR. The restriction enzyme sites, *BamH1* and *EcoR1* are shown in gray boxes 118

5.3 Polyacrylamide gel shift assay of DNA binding library with increasing amounts of P5b, in the presence of Ca^{2+}. The dotted box containing peptide-bound DNA was excised from the gel, and the DNA purified and amplified for the next round. CaP5b/DNA binding conditions: 10 mM Tris, pH 7.4, 500 μM $CaCl_2$, 10% glycerol 119

5.4 Aligned binding sites for CaP3W obtained by EMSA selection. Two families of target sites were identified, with the consensus sequences tabulated below. Two trials of 5 rounds of selection were performed, and the 44 sequences analyzed together (trial number indicated on the right). Asterisks indicate sequences identical in 7 of 8 positions. Four sequences were unaligned 120

5.5 Circular dichroism spectra of metallopeptides in the presence and absence of equimolar duplex DNA (E1: non-specific target; E2: target binding sequence). Data are presented as molar ellipticity based on peptide concentration. (A) Spectra of 15 μM EuP3W, EuP5b, and free E2 duplex, 50 mM Tris, 50 mM NaCl, pH 7.8. (B) Eu-peptide spectra and difference spectra ((Eu-peptide + DNA)-free DNA signal) with binding target E2. (C) Eu-peptide spectra and difference spectra ((Eu-peptide + DNA)-free DNA signal) with non-specific target E1 121

5.6 Circular dichrosim spectra of P3W and P5b with and without Eu^{3+} (pH = 7.80 at 25 °C, 50 mM Tris and 50 mM NaCl). EuP5b is poorly folded in aqueous solution, but has similar secondary structure to EuP3W when in 50% trifluorethanol solvent (dashed gray arrow). 122

6.1 A chimeric dimer protein model, R7C (in green), is designed to form dimer with disulfide bond on the flexible N-terminal of C2 by mutation from Arg7 to Cys7 (in yellow and arrowed). The dimer is proposed to interact with the major groove of DNA (in grey) using Swiss PDB Viewer software 134

6.2 (A) The amino acid sequence of chimeric F2. Parent protein sequence is indicated in bold for the homeodomain and shaded for the EF-hand Ca-binding loop. The 7-residue linker between two monomers is boxed. (B) Schematic representation of our chimeric dimer design with two approaches 135

6.3 The primers, Wong05F and Wong05R, were used for the site-directed point mutagenesis from Arg7 to Cys7 as underlined in order to create a site for dimerization through disulfide bond 136

6.4 A 15% SDS-PAGE of R7C purification steps. Lane M is prestained broad range marker. Lane 1 shows induced protein. Lane 2 is the French pressed supernatant. The protein was further purified by PEI precipitation and remained in supernatant as shown in lane 3. The supernatant after dialyzed against low salt solution is shown in lane 4 and was eluted by Heparin column as shown in lane 5. Lane 6 showed the dimerization of R7C using 2,2-dipyridyl disulfide where lane 5 and 6 were run with SDS-PAGE without 2-mercaptoethanol 136

6.5 The MALDI-TOF spectra of purified R7C. The calculated molecular weight of R7C is 7702 Da via DNAstar software and the measured molecular weight of F2 is $R7C^{+}$ at 7703 *m/z* and dimerized $R7C^{+}$ at 15410 *m/z* which is within the accuracy range 137

6.6 Fluorescence titrations of R7C with metal ions. The fluorescence intensity of Trp was followed at 350 nm (280 nm excitation) as a function of added metal. The solid lines represent a a linear non-least square algorithm fitting to a 1:1 association model, from which K_d was obtained. (*Inset*) Emission spectra showed throughout one titration, from 0 μM $EuCl_3$ (*Top*) to 40 μM $EuCl_3$ (*Bottom*) 138

6.7 CD spectra of apoR7C (black) titrated with increasing amount of $EuCl_3$. Sample contained 15 μM F2 in 10 mM HEPES pH 7.0 and 50 mM NaCl 139

6.8 The mean residue ellipticity of EuR7C at 222 nm with the titration of increasing equivalent of $EuCl_3$ 140

6.9 Agarose gel electrophoresis pTYB1 cleavage assay of EuR7C at different conditions. All R7C samples consisted of 10 mM HEPES pH 7.0 and 50 mM NaCl 141

A.1 A three-dimensional view of the λ Cro-DNA interaction was presented. The third helix (in red) is bound to the major groove and bends the DNA duplex as a dimer. The third β-strand interacts to each other to form a dimer. The figure is adapted from reference 154 154

A.2 The α-α turn (in blue) between helix 2 and 3 of wt Cro was replaced by the EF-hand Ca-binding loop (in red) of calbindin. A silent point mutation (in purple) on Aln29 (GCG to GCT) was preformed to create a restriction enzyme site of *AfeI* to aid the insertion of Ca-binding loop. The bold red letters show where calcium coordination occurs 155

A.3 (A) Primers used for the site-mutagenesis from Q37 to P37. Wong03F and Wong03R are complimentary to each other. The bases significant in the nucleotide change from CAA to CCA are represented in bold green letters. (B) The different on amino acid sequence of CroM and Q37P is indicated in red 155

A.4 (A) Amino acid sequence of the chimera, CroM. The parent protein sequence is indicated in bold (lambda Cro) and the inserted EF-hand Ca-binding loop from calbindin is underlined. (B) The EF-hand loop of calbindin (2BCA) and lambda Cro (5CRO) are shown in green/grey and blue/grey respectively. The overlay of the two motifs and the replacement of the α–α turn (in grey) on λ Cro created a chimera that could dimerize spontaneously, CroM 156

A.5 A 15% SDS-PAGE of wt Cro's induction conditions. Lane M is pre-stained broad range marker. Lane N is the non-induced protein. The protein was induced at different OD_{600} levels. Lane 1-3 and 6 indicates the amount of induction time in hours 157

A.6 A 15% SDS-PAGE of CroM's induction conditions. Lane M is pre-stained broad range marker. Lane N is the non-induced protein. The protein was induced at different OD_{600} levels. Lane 1-3 indicates the amount of induction time in hours 158

A.7 A 15% SDS-PAGE of pUQP37's and pUCroM's induction conditions. Lane M is pre-stained broad range marker. Lane N is the non-induced protein. The protein was induced at OD_{600} level of 1.00. Lane 1-3 indicates the amount of induction time in hours 159

A.8 A 15% SDS-PAGE of pUCroM's purification steps. Lane M is pre-stained broad range marker. Lane N is the non-induced protein while lane I is the protein induced for 2.5 hours. The cell culture was French pressed to yield supernatant (FPS) and pellet (FPP). The FPS fraction was further purified with PEI precipitation as shown in lane PEIP for pellet and PEIS for supernatant. The supernatant was then concentrated and loaded to a Sephadex G-50 fine (S-G50)..... 160

LIST OF ABBREVIATIONS

°C	Celsius degree
Apo-	Non-metal bound
Arg	Arginine
APS	Ammonium persulfate
BNPP	Bis(4-nitrophenyl) phosphate
BSA	Bovine serum albumin
CD	Circular dichorism
CIAP	Calf intestinal alkaline phosphatase
Cys	Cysteine
DNA	Deoxyribonucleic acid
dNTP	2′-deoxynucleoside 5′-triphosphate mix
dsDNA	Double-stranded deoxyribonucleic acid
DTT	1,4-dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoresis mobility shift assay
ESI-MS	Electrospray ionization-mass spectrometry
HEPES	N-Cyclohexyl-2-aminoethanesulfonic acid
HTH	Helix-turn-helix
IDT	Integrated DNA technologies
IPTG	Isopropyl 1-thio-β-D-galactopyranoside
K	Second order rate
k_{cat}	Turnover number
K_d	Dissociation constant
Da	Dalton
K_M	Michaelis-Menton constant

k_{obs}	Pseudo first order rate
MALDI-TOF	Matrix-assisted desorption/ionization time-of-flight
MES	2-(N-morpholino)ethanesulfonic acid
NPP	4-nitrophenyl phosphate
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PEI	Poly(ethlyenimine)
PMSF	Phenylmethylsulfonyl fluoride
Poly(dI-dC)	Synthetic polynucleotide polymer
SDS	Sodium dodecyl sulfate
ssDNA	Single-stranded deoxyribonucleic acid
TAE	Tris-acetate EDTA
TB	Tris-borate
TEMED	Tetramethylethylenediamine
Tris	Trishydroxymethylaminomethane
Trp	Tryptophan
UV-Vis	Ultraviolet-visible
V_{max}	Maximum velocity
wt	Wild type

CHAPTER 1

INTRODUCTION

1.1 Background

Many proteins are known to be very specific and unique in function due to being shaped by millions of years of natural selection. They are essential in biology starting from crucial cell regulations to cell death. Proteins play an integral part in regulating reactions when folded properly, but display severely negative consequences when not.[1] Extensive numbers of native protein structures and functions have been investigated with the help of the blossoming technology for both NMR and crystallographic structural determination over the past few decades. Yet, there is much work to be done in order to understand the key principles for proper protein folding.

One way to study protein folding and function is to design novel protein structures from scratch. Besides creating a structure from an unnatural sequence of amino acids, an alternative and equally effective approach is to design novel proteins by modifying native scaffolds. The advantages of utilizing a native protein scaffold are the stable and unique properties that are benefited from the native protein. Thus, successful protein design is possible through incorporation of an additional functionality into a well-characterized protein based on their common sequence motifs. *De novo* protein design is an innovative tool for building and improving these highly regulated machineries that have been shaped by nature allowing us to investigate key interactions of proteins.

Metal ions are ubiquitous in biological systems having essential roles in oxygen transport, proper neurological function, and transcription pathways. Metalloenzymes comprise the majority of the enzyme family and can catalyze some of the most difficult biological functions from structural rearrangement to catalysis to transfer of information.[2,3] With this in mind, it is not surprising that metalloproteins represent at least 1/3 of all structurally characterized proteins and approximately contribute to half of

proteins known from nature.[4] In addition, DNA and RNA molecules have been shown to be involved in catalyzing important biological reactions in the presence of metal ions over the past 20 years.[5-7] The introduction of metal ions into designed proteins also provides rich spectroscopic properties for the investigation of their function and structure. Not only do we gain insights into the structure and function of proteins, we can create more selective and stable alternatives for biochemical and medicinal applications. Recently, the creation of small designed proteins, such as a zinc-finger-based enzyme with catalytic and HIV antiviral behaviors have shown promise as therapeutic treatments.[8,9] As a result, *de novo* protein design incorporating metal binding sites is an intriguing technique for protein design.

Our research is based on an idea of teaching an old dog a new trick. Previous studies have given new functionalities to engineered helix-bundle and miniature peptides, which in both cases monomeric domains interacted with DNA.[10-12] Work in this thesis took a step forward by engineering a metal-binding loop to a DNA-binding protein dimer so that we can further understand the significant contributions of protein dimerization upon DNA interactions. This project could provide a potential prospect for DNA sequence specific cancer drugs that would be potent, yet have fewer side-effects than the current treatments.

With the increasing recognition of a large number of homodimers found in living systems, one cannot help to wonder about the functional benefits of dimer formation. The importance of dimerization to catalytic/structural behavior is particularly pronounced in small DNA-binding protein units (<100kDa) as they have been shown to behave like "hinges" to hold DNA and/or interface with "active" sites.[13] Owing to the enhanced structural stability and extreme specificity, the homodimer template is a challenging target for protein engineering. There are many examples of restriction endonucleases that are homodimers with great specificity for defined palindromic sequences and DNA cleavage activities.[14,15] They utilize hard Lewis acids such as Ca^{2+}, Mg^{2+}, and Zn^{2+} to

activate water as a nucleophile to cleave P-O bond. Those homodimers recognize specific base pair sequences and catalyze the hydrolysis of the phosphodiester bond on both strands of DNA. These unique properties of homodimers encourage the design of dimeric artificial endonucleases as alternate tools for bioengineering. Therefore, we hope to engineer metallochimeric dimers that give greater sequence discrimination and stronger DNA affinity, leading to more robust hydrolytic activities in the presence of metal ions like Ca^{2+} and Ln^{3+}.

1.2 Calcium Binding Sites and EF-hand motif

Calcium plays a vital role in the physiology and biochemistry of organisms, especially in signal transduction pathways. Calcium ions are the key players as secondary messengers in signal transduction in eukaryotic cells. They enter into the cytoplasm through the cell membrane via calcium-binding proteins.[16,17] Several different calcium-modulated proteins, such as troponin-C and calmodulin, bind to calcium with high selectivity even in the presence of up to 10^5-fold higher concentration of Na^+, K^+, and Mg^{2+}.[18] This is particularly important for regulatory sites in Ca^{2+} signal proteins including calmodulin and its homologues. Because of its tolerance to engineered sequence changes and its simple substrate that requires relatively few geometrical variables, Ca^{2+} sites are well suited for the protein engineering approach.

Calmodulin is the protein of interest in our lab as it belongs to a family with highly conserved helix-loop-helix calcium binding domain known as EF-hand motif (Figure 1.1A).[19,20] The term "EF-hand" was first coined as a pictorial representation of the orthogonal "thumb and forefinger" geometry was described by Kretsinger and Nockolds 27 years ago (Figure 1.1B). The EF-hand motif consists of two α helices grasping a Ca^{2+} atom in between. Most EF-hand proteins undergo a conformational rearrangement upon binding to Ca^{2+}. Nonetheless, the calcium binding in calmodulin usually occurs at physiological concentrations. In fact, the affinity of calcium to the EF-

hand loops has been shown to vary by six orders of magnitude ($K_d = 10^{-4}$ to 10^{-9} M) depending upon the protein environment, the binding geometry, and the number of charged ligand residues.[21-25] By investigating the rules governing Ca^{2+}-binding affinity, it can facilitate our understanding of Ca^{2+}-mediated protein stability.

The calcium binding loop in the EF-hand motif consists of a 12-residue highly conserved sequence.[26] Calcium ions are coordinated by carbonyl groups on the side chain of residues on the loop. The calcium ion is wrapped around by the loop and coordinated by seven oxygens in a pentagonal bipyramidal arrangement. The occupation of ±x, ±y, and ±z positions in the Cartesian framework assume an almost octahedral arrangement and is denoted as 1(x), 3(y), 5(z), 7(-y), 9(-x), and 12(-z; bidentate) (Figure 1.2).[27,28] With one vertex, –z, shared by the two carboxyl oxygen atoms of a glutamate side chain that binds in a bidentate manner to Ca^{2+}, position –y occupied by a conserved backbone carbonyl group and –x being generally a water molecule, this demonstrates a very similar coordination to that of lanthanide ions which is a good candidate for calcium substitution. In addition, the second shell ligand at –x position that stabilize the water is also referred to as the "gateway position" which has the most variations on metal to loop contact.[29,30] This position allows for the coordination of solvent molecules that are in the closest proximity to the metal. The changes of these key metal-binding residues provide a fundamental tool to tune the binding and kinetics of the metal center.

1.3 Substitution of Calcium with Lanthanides

Lanthanides, Ln^{3+}, have long been established as spectroscopic probes for studying the refined structures of calcium binding sites due to their striking similarity in properties.[31] In particular, the ionic radius of Ln^{3+} and Ca^{2+} are virtually identical and the preference of Ln^{3+} for hard oxygen donors make it a good candidate for substitution at the calcium binding site. The negligible crystal field splitting allows for flexible and similar coordination at the EF-hand loop. Because of the relatively small crystal field

stabilization energy, the coordination number and geometry are influenced by a balance between the metal-ligand interaction and ligand-ligand repulsion. The lanthanides typically have eight to nine coordination species while calcium forms seven to eight. Therefore, lanthanides can coordinate with hydroxyl with its extra ligand resulting in hydrolysis. Most lanthanides prefer to adopt a trivalent oxidation state, due to facile removal of the valence $6s^2$ and $5d^1$ electrons which is the only oxidation state that is easily accessible under biological condition.[32] Its higher charge also results in greater binding affinity at the calcium-binding site and stronger Lewis acid that can promote hydrolytic cleavage of DNA and RNA at the phosphodiester bonds.[33]

The paramagnetic properties of lanthanides arise from unpaired f orbital electrons (for all but La and Lu) are particularly useful for NMR structural studies and luminescence studies in some of the lanthanides, such as Eu^{3+} and Tb^{3+}. They can be utilized as a luminescent probe to investigate the chemical environment at the binding sites. Indeed, there have been numerous studies on Ln-coordination complexes which have been utilized as chiral shift agents and probes of biological structure.[34,35] Ln^{3+}-substituted calcium protein techniques have been extensively used for defining both local and long-range structure in signal proteins.[36-38] All in all, not only do we gain a good deal of structural information and stability of the binding site by replacing calcium with lanthanides, but we also can utilize lanthanide's striking catalytic activity to create a metalloenzyme with DNA hydrolytic activity.

1.4 DNA interaction with HTH Protein

A homeodomain parental scaffold was chosen as the DNA-binding motif of dimeric designs based on its well-characterized structure and DNA interactions. Homeodomains play important roles in regulating transcription and development in eukaryotes.[39,40] They are a family of highly conserved proteins that specifically bind to promoter DNA sequence. They consist of 60 amino acids that form a three helical

bundle. The three helices are connected by a loop and a turn. Helices 1 and 2 are connected by a loop where as helices 2 and 3, the DNA recognition region, are linked together by a turn (Figure 1.3).[41-44] Thus, the essential DNA recognition region was defined as the helix-turn-helix (HTH) motif. Only a limited number of highly conservative DNA binding motifs have been classified.[45-47] These domains include the homeodomain, zinc-finger, basic helix-loop-helix, and leucine zipper proteins. Among these, the α-helix is one of the common features for DNA groove-binding properties and it could provide the key to understand in the principles for protein-DNA interactions.

The homeodomain family recognizes the 5′-TAATNN-3′ sequence (with the last two sites as variables) and *engrailed* homeodomain in particular recognizes the 5′-TAATTA-3′ sequence specifically through both hydrophobic and hydrogen bonding interactions between the side chains of amino acids and the DNA phosphate backbone.[48-50] Since the DNA-binding helix of homeodomains are basic and relatively bulky in nature, they have an affinity towards the negatively charged DNA molecules and are well-suited to associate with the major groove of DNA while the arm of the N-terminus contacts at the adjacent minor groove of DNA (Figure 1.4).[51,52] *Engrailed* homeodomain which regulates segmentation in *drosophila* larvae was the first transcription factor to be co-crystallized with DNA and has been well characterized. With this knowledge, a chimera based on the small, well characterized *engrailed* homeodomain can serve as a primary model for protein-DNA interaction.

Despite the highly conserved sequence and the simplicity of the motif, DNA specificity is substantially different from one homeodomain to another.[53,54] Even though the *engrailed* homeodomain makes extensive contacts with DNA phosphate backbones directly or through water, such as R3 and R5 in the N-terminals and L47, Q50, N51, and M54 in helix 3, the mutation of *engrailed* homeodomain at Q50K alters this DNA-binding specificity from TAATTA to TAATCC.[55,56] This sensitivity of proteins to subtle

changes of the residue sequence and structure can make modulating the target sequence possible and is particularly interesting in novel protein-DNA design.

Homeodomains are also known to be internalized by cells in cultures and thus bind to DNA in the nucleus.[57,58] This interesting property could be applicable for delivering a small engineered protein with specific DNA hydrolytic function into the nucleus, thus providing a new method of drug delivery.

1.5 DNA Hydrolytic Activity for Enzyme Design

The hydrolysis of phosphodiester bonds is critical to proper function of biological systems, including DNA repair and excision, integration, and signal transduction. It is also important to mediate DNA cleavage by hydrolytic pathways since radical species from oxidative cleavage can result in non-specific reactions and thus cytotoxicity. Non-redox active metalloenzymes (Zn^{2+}, Ca^{2+}, and Mg^{2+}) have long been known for their DNA hydrolytic ability and are abundant in living systems.[59,60]

Because of the increasing interests in applications of biomimetic nucleases, numerous small synthetic metallonuclease-like complexes have been designed.[61,62] However, most of the metal complexes lack specificity, have unacceptable enzymatic efficiency, and are unable to cleave double-stranded DNA. Not until recently, with the incorporation of lanthanide ions and lanthanide complexes has catalytic hydrolysis of double-stranded DNA been made possible.[63,64] One noteworthy effort to enhance sequence-specificity and cleavage efficiency was the creation of a Ce^{4+} iminodiacetate antisense oligonucleotide by Komiyama's group.[65] Lanthanides are good candidates when compared to transition metals due to their strong Lewis acidity, lack of redox chemistry (except for Ce^{4+}), rich spectroscopic and labile chemical properties which allow for direct measurements and rate enhancement in DNA hydrolysis.

Greater promise for metal-promoted DNA hydrolysis may be found in the development of metalloprotein design. Since DNA-binding proteins are already

equipped with the ability to recognize DNA sequences specifically, lanthanide-incorporated protein design has great potential for DNA hydrolytic chemistry and thus may provide important insights for the creation of artificial restriction enzymes. In our study, Ln^{3+} is substituted into the calcium-binding loop that is incorporated in a structurally homologous DNA-binding domain.[66-68] An "artificial" metalloprotein with hydrolytic activity towards specific DNA patterns is thereby created.

1.6 The Benefit of Dimers

As previously mentioned, we are particularly interested in how dimerization can impact activity and selectivity of a designed metalloprotein. A glance at globular proteins that carry out complex regulatory or catalytic functions reveals that their structures parallel their complexity. In fact, many of the globular proteins are composed of multiple subunits. A striking feature among globular proteins is revealed when a survey on the oligomeric state of proteins is performed on the Protein Data Bank (PDB). Over half of the oligomeric proteins are homodimers.[13] A survey of *E. coli* proteins in the SWISS-PROT annonated protein sequence also suggests the predominance of dimeric proteins (Table 1.1).[69] Restriction endonucleases, one of the common bioengineering tools currently used for DNA sequence manipulations is also often dimeric. These restriction endonucleases have great specificity at defined palindromic sequences and hydrolyze double-stranded DNA activated in the presence of metal ions.[15,70,71]

As demonstrated by the high natural occurrence of dimers, there are many physical advantages of being dimeric in addition to the biological advantages. The homodimeric arrangement usually maximizes stability and avoids unwanted aggregation.[69] With the reduction of surface area in dimers, the diffusion of substrates to enzyme active sites is improved, resulting in greater activity.[72] Dimer formation accounts for a broad range of allosteric effects in many enzymes, leading to cooperative binding, which assists the binding affinity to substrate.[73] The dimerization of p53 tumor

suppressor protein stabilizes DNA binding ,[74] and recent research on a novel RNase in which a dimer exhibits a 2-fold increase in dsRNA and ssRNA enzymatic activity over a monomer on a per monomer basis are great examples of the benefits of dimerization.

Recently, dinuclear metal complexes have been shown to generate greater DNA hydrolytic activity. In one example, double Lewis acids, Co^{3+}, and bridging oxide activation gives 11 orders of magnitude rate increase in phosphate hydrolysis.[75] Several dinuclear lanthanide macrocyclic complexes have been shown to perform double-stranded DNA hydrolysis as well.[76] With the consideration of these advantages, the DNA specificity of a protein, the hydrolytic property of lanthanide metals, and a promising novel dimeric design approach suggesting a true artificial enzyme can be achieved by protein design.

1.7 *De Novo* Protein Design

De novo peptide and protein design was first introduced about two decades ago. This approach starts with a postulated or known flexible three-dimensional protein and aims to characterize the fold of a protein. The goal is to understand how to associate a given protein sequence with its own unique shape. In recent years, *de novo* design has evolved into a structure-based protein design approach to create predictable three-dimensional shapes.[10,77,78] Moreover, the studies on DNA hydrolytic activity with lanthanide complex,[79,80] have led us to pursue peptides/proteins design that can incorporate metal-binding functionality into new scaffolds. Our research has shown that the chimeras we built have exhibited robust structure as well as some degree of sequence selectivity and hydrolytic activity.[10,11,81] Utilizing protein design to create enzymatic metalloproteins is a critical building block towards developing selective chemotherapy agents that can target sequences of choice.

The combination of activity and selectivity for our new protein was formulated by combining geometrically similar regions from proteins with different functionalities. The

consensus Ca-binding EF-hand motif of calmodulin and the helix-turn-helix (HTH) motif of *engrailed* homeodomain, a DNA-binding domain, were carefully chosen due to their structural similarity at turns.[11,66] The α–α corner between α2 and α3 of homeodomain was modularly substituted by the metal binding loop of calmodulin (Figure 1.5). With the advantage of the Ln^{3+} Lewis acids mentioned above, the incorporation of various Ln^{3+} at the metal binding loop is also a useful probe for structural and functional studies. By incorporating both a metal-binding site and a DNA binding domain in the nuclease design, the artificial endonuclease is able to deliver hydrolytic metals to the DNA backbone utilizing protein-DNA interaction.

We have also demonstrated robust secondary structure and DNA-sequence discrimination in cleavage with a cerium-bound 33-mer peptide (P3W) and a metallated monomeric protein (C2).[81,82] The DNA recognition was shown to be stronger and more specific in C2 (full domain) than that in P3W (half domain) (data not shown). Many studies have shown that homo/heterodimers are abundant in the biological systems and improve functions as summarized by Finazzi-Agrò et.al.[13] Short synthetic DNA binding peptides have also been demonstrated to bind to multiple direct-repeat sequences with increased DNA sequence selectivity.[83,84] These findings shine light on building chimeric dimers based on our successful scaffolds to further increase the hydrolytic activity and enable the recognition of target sequences that are long enough to ensure specificity.[85,86] Furthermore, a chimeric dimer with two active sites that can catalyze double strand DNA, and could therefore lead to a new paradigm for restriction enzymes.

1.8 Description of the Thesis Contents

Chapter II details the design of the metallochimera root from the DNA-binding domain, *engrailed* homeodomain, and the calcium-binding EF-hand loop from calmodulin. The idea of dimeric design and our approach will be discussed in detail.

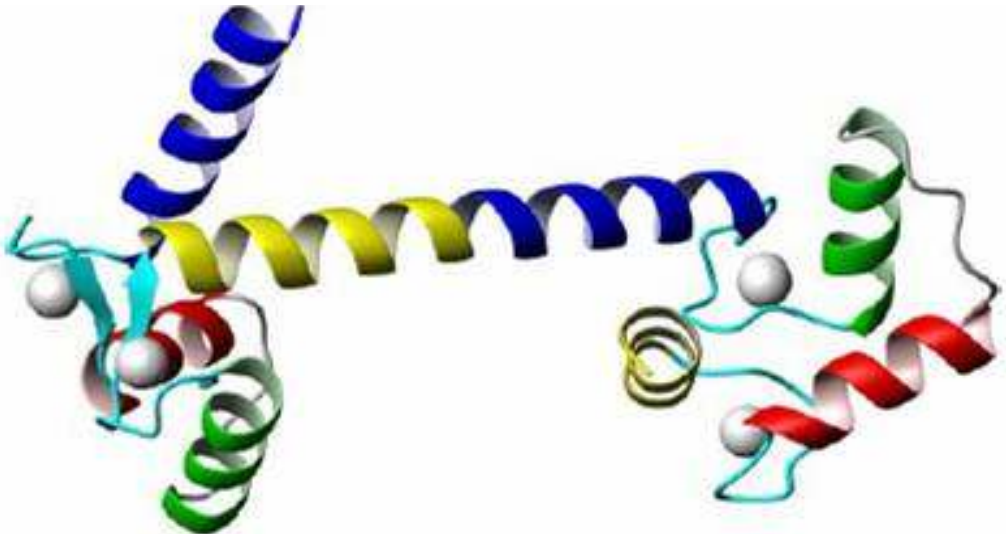
Protein purification and physical characterization of the first generation of chimeric dimer, F2, will be described.

The nuclease activities of F2 over various conditions in the presence of DNA plasmid and a small molecule DNA analog will be illustrated in Chapter III. The DNA hydrolytic behavior of F2 will also be compared with other nucleases and chimeras in this chapter.

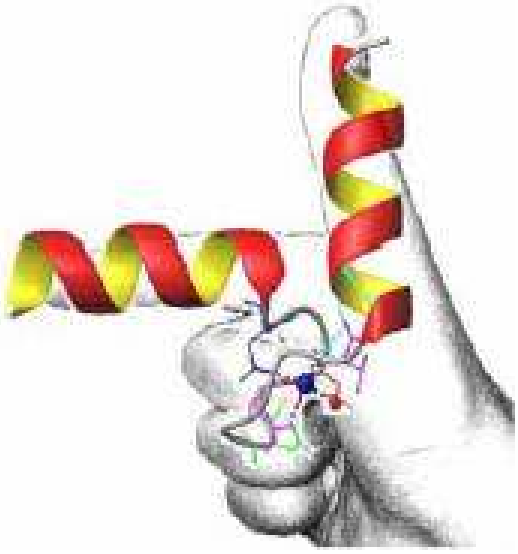
In Chapter IV, the interaction and kinetic activity of F2 with selected DNA sequences will be carefully investigated. A discussion of DNA cleavage and binding mechanism will also be included.

Chapter V mainly describes the DNA sequence selectivity properties of two chimeric peptides, P3w and P5b, precursors to the protein work described in previous chapters. This work has been published in the Journal of Biological Inorganic Chemistry.[87]

The purification and physical characterization of a second dimeric design, R7C will be discussed in Chapter VI. Due to the difficulties of disulfide bond dimerization of R7C, more reliable results have yet to be elucidated.



(A)



(B)

Figure 1.1 (A) X-ray structure of the calmodulin (1CLL). The Ca^{2+} ions are shown in white spheres as the calcium-binding loops are represented in light blue. (B) The EF-hand proteins are based on a motif characterized by a helix-loop-helix structure. The motif's name is derived from its thumb and forefinger pair on a hand. These figures are adapted from reference 20.

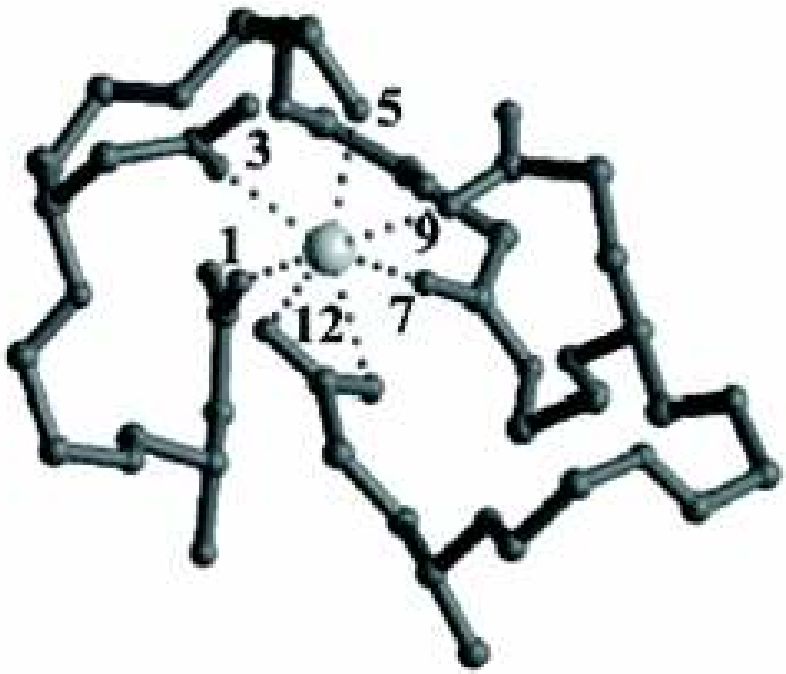


Figure 1.2 Coordinating residues of an EF-hand site from parvalbumin. The coordinating position of a typical EF-hand site is illustrated as the residues at loop positions 1, 3, 5, 7, 9, and 12 are coordinated to the bound metal ions. Residue 9 is bound through water as the "gateway position". This figure is adapted from reference 28.

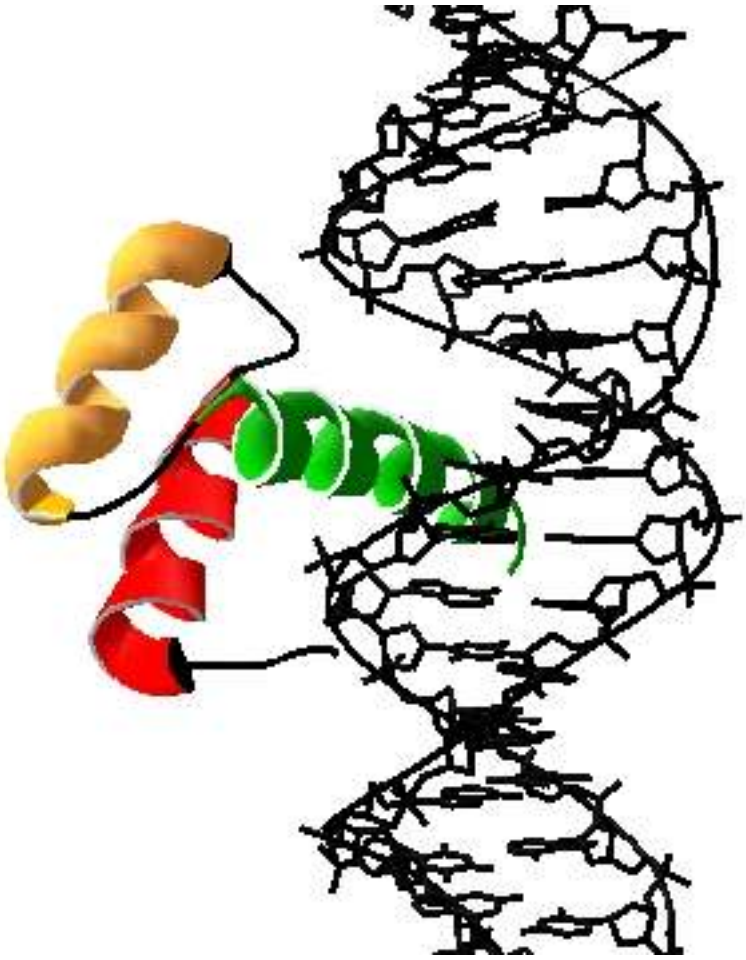


Figure 1.3 The crystal structure of *engrailed* homeodomain (2HDD.pdb). Helix 1 is in red while helix 2 is shown in orange. Helix 3 is the DNA-binding domain which is represented in green.

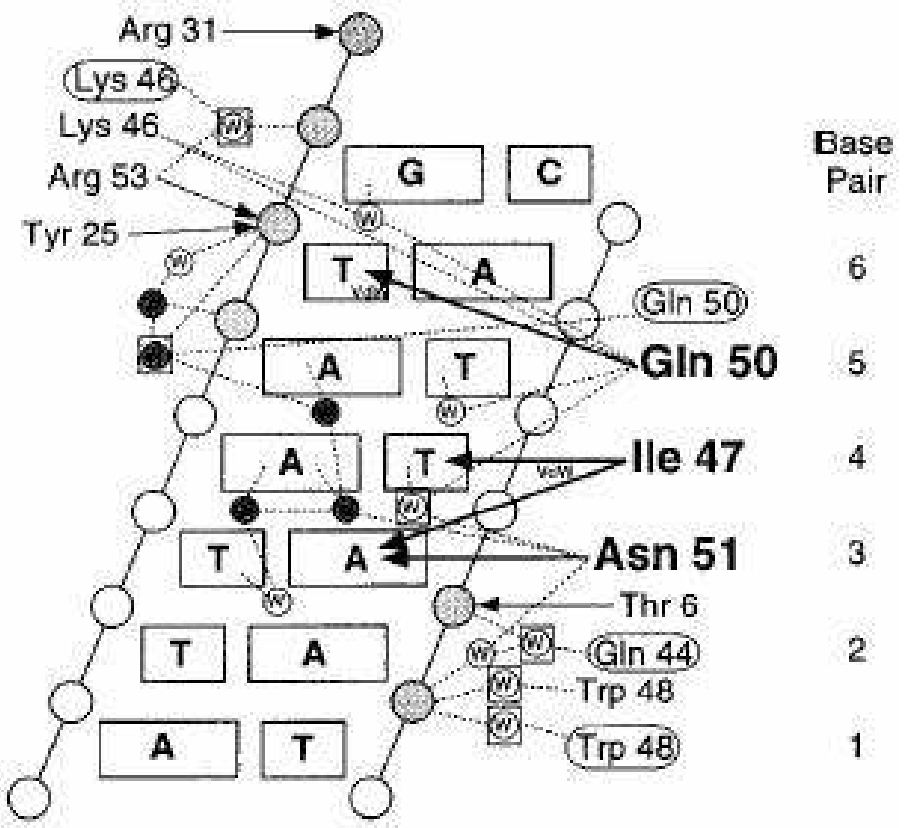


Figure 1.4 Diagram showing major groove contacts by *engrailed* homeodomain mostly from helix 3. DNA phosphates are represented by circles connected by lines. DNA base contacts are indicated by a thick line. Hydrogen bonds are depicted by broken lines. Water is shown as W in a circle, a filled circle, or a circle and square. DNA contacts with the N-terminal residues, R3 and R5, are not shown. This figure is adapted from reference 51.

Table 1.1 Natural occurrences of oligomeric proteins in *Escherichia coli*. This table is adapted from reference 69.

Oligomeric state	Number of homooligomers	Number of heterooligomers	Percent
Monomer	72	---	19.4
Dimer	115	27	38.2
Trimer	15	5	5.4
Tetramer	62	16	21.0
Pentamer	1	1	0.1
Hexamer	20	1	5.6
Heptamer	1	1	0.1
Octamer	3	6	2.4
Nonamer	0	0	0.0
Decamer	1	0	0.0
Undecamer	0	1	0.0
Dodecamer	4	2	1.6
Higer oligomers	8	---	2.2
Polymers	10	---	2.7

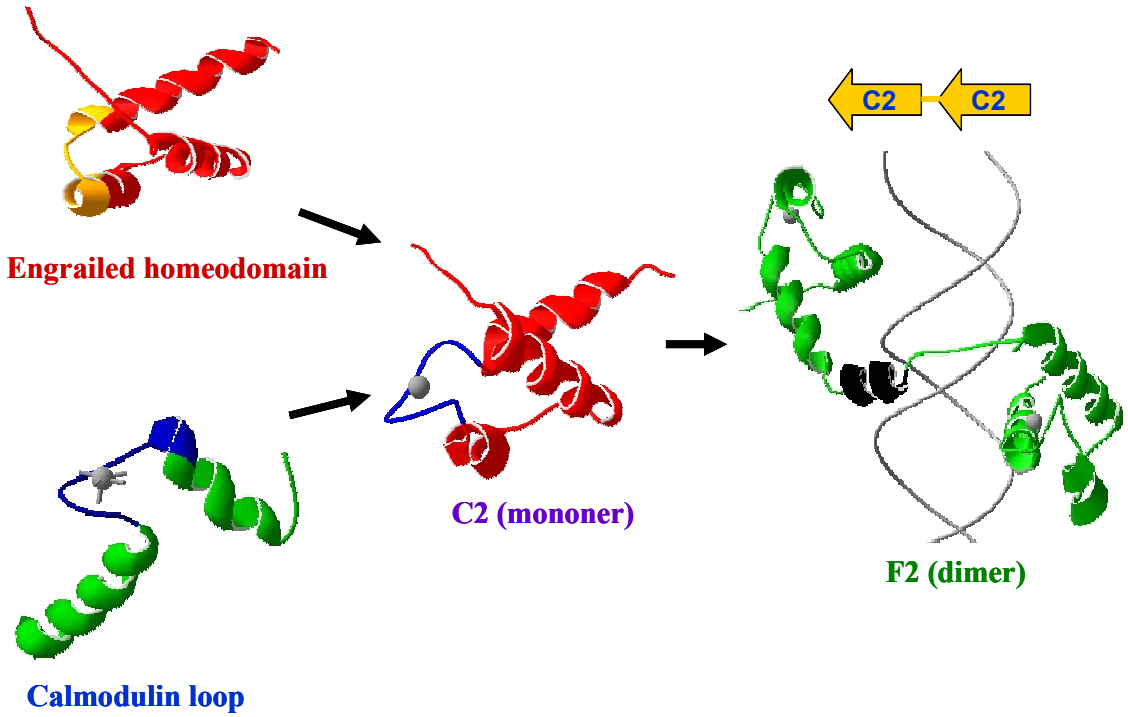


Figure 1.5 The EF-hand loop (1OSA) and *engrailed* homeodomain (2HDD) are shown in blue/green and red/yellow respectively. The overlaying of two motives created a chimeric monomer model, C2 (in the middle). A chimeric dimer protein model, F2 (in green), is designed to be covalently connected through a flexible residues linker (in black and not helical) and binds to the major groove of DNA (in grey). Both structures of C2 and F2 models are only determined by Siwss PDB viewer.

CHAPTER 2

PHYSICAL CHARACTERIZATION OF THE CHIMERIC DIMER F2

2.1 Introduction

One of the current ongoing goals of pharmaceutical science is to develop tools to improve cancer treatment through high specificity and selectivity in biological systems and relatively low toxicity to patients. The ability to only target a small collection of overexpressed genes in tumorous tissues is one of the great drug development interests and one to which protein design is well-suited. Peptides/proteins incorporating metal-binding functionality into new scaffolds have therefore been extensively studies as they exhibited robust structure as well as some degree of sequence selectivity and hydrolytic activity towards DNA.[10,11,81]

We have demonstrated robust secondary structure, modest metal binding ability, and DNA-sequence discrimination in cleavage with a cerium-bound 33-mer peptide (P3W) and a metallated monomeric protein (C2)(Figure 2.1), though their hydrolytic activity towards DNA were not as readily as expected.[81,82] Nonetheless, the recognition of DNA sequence in well-folded C2 (full domain) was stronger and more specific than that in P3W (half domain) (data not shown) illustrated how structure affects its functionality. These findings shine light on building a chimeric homodimer based on our successful scaffolds to further increase the hydrolytic activity and enable the recognition of target sequences that are long enough to ensure specificity.[85,86] In addition, homodimeric proteins, such as lambda Cro[88-90] and paired homeodomain,[91] exhibit greater stability and stronger DNA binding affinity than their monomer/mutant forms. Short synthetic DNA binding peptides have also been demonstrated to bind to multiple direct-repeat sequences with increased DNA sequence selectivity.[83,84] Therefore, chimeric dimers with two active sites that can catalyze double stranded DNA can act as a paradigm for artificial restriction enzymes.

Many natural proteins are constructed of repeated units. Therefore, it is common to observe the linkage of independent folding units into a single polypeptide chain thorough evolution. Single-chain antibodies are an example of the *in vitro* fusion of separated genes encoding interacting subunits as a way to obtain recombinant proteins with desirable properties.[92] Nucleic acid binding proteins have also been constructed into single-chain dimers to study fundamental issues in macromolecular recognition, protein folding, and stability.[93-95]

To obtain a covalent dimer which maintains the intrinsic DNA affinity of the homeodomain/monomeric chimera, a linkage should not perturb the structure or flexibility of the homeodomain/monomer. With the observation that the N-terminal arm of C2 is flexible and the close proximity of both C- and N-termini of C2 to the major groove, a flexible linkage between the C-terminus of one domain and the N-terminus of the next may be compatible with the structure and function of the chimeric dimer, F2.

Herein we report the characterization of the dimer's secondary structure, stability, metal affinity, and DNA-binding affinity as its DNA hydrolytic and selectivity properties will be described in detail in Chapter 3. In this chapter, the metal binding loop demonstrated that it was structurally and functionally well preserved within the modified domain. The stability of the europium-bound homodimer was greater than that of the monomer yet gave slightly less secondary structure due to the addition of the linker. DNA-binding ability was preserved in the dimer from the parental homeodomain. As a result, the dimeric approach is the first example of a designed nuclease that can exploit the exquisite specificity achieved by DNA binding proteins to deliver a hydrolytic metal to a given sequence for selective cleavage with greater nuclease activity and sequence discrimination than any previously designed metalloprotein.

2.2 Experimental Procedures

2.2.1 Chimeric Design

The known crystal structures of *engrailed* homeodomain (2HDD)[96] and the Ca-binding EF-hand motif of calmodulin (1EXR)[97] were oriented and aligned manually using Swiss PDB Viewer.[98] The design of the chimeric monomer was based on the overlaying of the HTH motif of *engrailed* homeodomain at the α–α turn and calmodulin EF-hand metal-binding loop due to their similar topology which was described previously (Figure 2.1).[66,68] The homodimer was constructed by covalently linking a series of seven flexible residues between the C-terminus of one monomer and the N-terminus of the other monomer to keep the dimer intact yet flexible enough for both domains to have DNA associations (Figure 1.5).[99] The complete residue sequence of the designed dimer contains 132 amino acids and the schematic representation of the chimeric dimer design is shown in Figure 2.2. The characterization of F2 design (head to tail connection) is the focus of this chapter while the R7C design (tail to tail connection) will be discussed further in Chapter 6.

2.2.2 DNA Cloning

F2 cloning was carried out using a vector, pET21a(+)(Novagen, 5369bp), which has a T7 promoter without an affinity tag for purification. It contains the DNA encoding for monomeric chimera with *Nde*I and *Xho*I sites at each side of DNA_{C2}, as shown in Figure 2.3, template 1 (provided by Sunghyut Lim, University of Iowa). Primers, Wong 09F and 07R, were ordered from IDT (Figure 2.4) to replicate DNA_{C2} that had *Xho*I at both termini by polymerase chain reaction (PCR). The PCR conditions were 95 °C for 30 seconds, 55 °C for 1 minute, and 68 °C for 1 minute over 25 cycles. The PCR product and DNA_{C2} were then digested by enzyme *Xho*I at 37 °C for 2 hours. The digested template 1 (Figure 2.3) was de-phosphorylated by 1 uL of calf intestinal alkaline phosphatase (CIAP) in 1 x CIAP buffer at 37 °C for 2 hours to prevent self-ligation at

*Xho*I sites. The digested template and de-phosphorylated PCR product were purified using a PCR purification kit (Qiagen) so that there would be no interference by other buffer constituents and unwanted DNA pieces during ligation. Purified PCR product was ligated into DNA_{C2} at *Xho*I sites. The reaction mixture was incubated at 16 °C for 6 hours and the ligated DNA was then transformed into DH5α competent cells. After the orientation and size of the inserted PCR product was verified, Primers that included the 7-residue linker and has the deletion of start or stop codon on both ends of linker, were designed and ordered from IDT(Figure 2.4) to carry out multi-site mutagenesis between two DNA_{C2}'s (Figure 2.3, template 2). The multi-site mutagenesis procedure was described by Stratagene. The final vector product was isolated from cells via a DNA purification kit (Qiagen; Figure 2.3, template 3). The DNA construct encoding the target gene was confirmed by University of Iowa DNA facility.

2.2.3 Protein Purification

The DNA plasmid encoding the F2 sequence was transformed into BL21(DE3)star competent cells. The colony was then grown in 100 μg/mL Ampicillin LB broth at 37 °C to OD_{600} of 1.0. The culture was expressed for 2.5 hours at 28 °C in the presence of 1 mM of isopropyl 1-thio-β-D-galactopyranoside (IPTG). The cell pellet was re-suspended in lysis buffer consisting of 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 5 mM 1,4-dithiothreitol (DTT), 1 mM EDTA, 1mM phenylmethylsulfonyl fluoride (PMSF), and 300 μg/ml of freshly prepared lysozyme. The suspended cell solution was incubated for 30 minutes at 4 °C. The pellet and supernatant were separated via centrifugation and the supernatant was then discarded. The pellet was re-suspended in the lysis buffer before being French pressed at 16,000 psi. The inclusion body was re-dissolved in CB-1000 buffer (10 mM N-Cyclohexyl-2-aminoethanesulfonic acid (HEPES) pH 8.0, 1 mM EDTA, 1 M urea, 1 M NaCl, and 1% Triton-X*)*. A total concentration of 0.1% Poly(ethlyenimine) (PEI) was added to the supernatant to

precipitate negatively charged molecules and the target protein was then mostly precipitated by slowly stirring in ammonium sulfate powder until 25% saturated. The precipitate was again dissolved in the CB-1000 buffer and the solution was dialyzed gradually from 1000 to 50 mM NaCl with 10 mM HEPES pH 8.0, 1 M urea, and 1 mM $CaCl_2$. The desalted target protein was bound to Heparin Sepharose 6 Fast Flow column (Amersham) and eluted at 500 to 700 mM NaCl without urea. The resulting 95% pure concentrated protein was further purified with a Sephadex G-50 Fine column (Amersham) in 10mM HEPES pH 7.0 and 50 mM NaCl buffer to 99% purity and dialyzed against metal-free solution consisting of chelex-100 (Bio-Rad). The purity and molecular weight of apoF2 was confirmed by silver stained SDS-PAGE and matrix-assisted desorption/ionization time-of-flight mass spectrometry (MADLI-TOF MS) respectively. The concentration of filtered apoF2 was determined from A_{260} via UV-Vis spectrometer. The extinction coefficient and amino acid composition of F2 were calculated by amino acid analysis (University of Iowa Molecular Analysis facility).

2.2.4 Tryptophan Fluorescence Titrations

The metal-binding affinity of F2 was studied by monitoring the fluorescence intensity of the single tryptophan (Trp) residue in each domain (W_{52} & W_{120}) as a function of metal concentration. The spectra were collected on an Aminco-Bowman Series 2 flourimeter. Each sample of 5 μM protein in 10mM HEPES pH 7.0 and 50 mM NaCl was excited at 280 nm and the Trp emission was monitored at 350 nm as described previously.[100] The spectra were scanned from 400 to 300 nm in 1.0 nm intervals at 25 °C. Metal was added to the protein solution as chlorides and equilibrated for 5 minutes. Stoichiometric amounts of metal ions, including Ca^{2+}, Mg^{2+}, Fe^{2+}, Fe^{3+}, Mn^{2+}, Zn^{2+}, La^{3+}, Eu^{3+}, and Lu^{3+}, were added to the protein as the chloride salt, freshly prepared from 10 mM stock. The intensity as a function of metal concentration was described as a binding curve. Since no co-operative metal binding was observed from the binding curve, the

data was well-fit by a single exponential 1:1 association model per binding site with a linear non-least square algorithm.[100,101]

2.2.5 Circular Dichroism Titrations

Circular dichroism (CD) spectra of F2 titrating with Eu^{3+}, Ca^{2+}, and Mg^{2+} were detected using an Olis Cary-17 DS Conversion spectrophotometer at 20 ˚C under N_2 atmosphere in a 0.1-cm cell. Myoglobin containing a high degree of alpha-helices and loops was used as a control to correct measured helicities at certain buffer condition by normalizing measured values with the myoglobin standard. Spectra were collected from 260 to 190 nm at 1.0 nm resolution and were smoothed with Olis software (25 boxcar smoothing function). The spectra were corrected using a background scan of the buffer as a reference. The sample contained 15 μM of protein in 10 mM HEPES pH 7.0 and 50 mM NaCl titrating with increasing stoichiometric quantities of metal ion ranging from 0 to 4. Stoichiometric amounts of metal ions were added to protein as the chloride salts, from 10 mM stock freshly prepared by weight.

2.2.6 Thermal Denaturation

CD spectra of 15 μM apoF2, Ca_2F2, and Eu_2F2 in 10 mM HEPES pH 7.0 and 50 mM NaCl were measured from 5-80 °C at 5 °C intervals using an Olis Cary-17 DS Conversion spectrophotometer at 20 ˚C under N_2 atmosphere in a 0.1-cm cell with stopper. The sample was equilibrated for 2 minutes in a 1.0 x 0.1-cm cuvette in a thermostated cell chamber using a Peltier-type temperature controller. Spectra were collected from 260 to 210 nm at 1.0 nm resolution and smoothed with Olis software (11 boxcar smoothing function). The percentage of unfolded conformation as a function of temperature was plotted to describe an unfolding transition as temperature increases. The data was interpolated with a four-parameter algorithm by using a non-linear least square equation to obtain the melting temperature.[102] The protein fraction present in the

unfolded conformation (F_u), equilibrium constant (K_{eq}), and Gibbs free energy were calculated using the following equations:[103]

Equation 1.1: $F_u = (y_F\text{-}y)/(y_F\text{-}y_U)$

Equation 1.2: $K_{eq} = [U]/[N] = F_u/(1\text{-}F_u)$

Equation 1.3: $\Delta G° = -RT \ln K_{eq} = -RT \ln[(y_F\text{-}y)/(y\text{-}y_U)]$

where y_F and y_U represent the percentage of y in the folded and unfolded states, respectively.

2.2.7 DNA Gel Shift Assays

The affinity of apoF2, Ca_2F2, and Eu_2F2 for supercoiled plasmid DNA was examined by agarose gel electrophoresis. Samples of 100 ng of plasmid (pTYB1, 7477 bp, New England Biolabs) with increasing concentrations of each ligand ranging 0-720 nM were incubated for 5 minutes at room temperature in a buffer of 10 mM HEPES pH 7.0 and 50 mM NaCl. The plasmid was chosen because it has several occurrences of the sequence TAAT, which is the common subset of the 6 base pair binding site for homeodomains. A 1% agarose gel was run for approximately 1 hour in 1 x TA buffer (Tris-acetate) at 120 V. The gels were visualized under UV-light and photographed with a Kodak digital imager. The supercoiled plasmid's migration on each lane was analyzed by a Kodak image analyzer. Control lanes contained only plasmid, buffer, and loading dye which were treated in the same manner as other samples.

2.3 Results and Discussion

2.3.1 Protein Design and Purification

The purification steps and the purity of F2 were followed by 15% SDS-PAGE (Figure 2.5). The size of apoF2 was confirmed by MADLI-TOF with $(F2+H)^{+}$ at (calcd for apoF2, 15820 *m/z*) 15800 *m/z*, $(F2+H)^{2+}$ at 7911 *m/z*, and $(F2+H)^{3+}$ at 5274 *m/z* (Figure 2.6). Since the mass accuracy for MADLI-TOF is ±0.1% for weights greater than 10 kDa, the calculated molecular weight of F2 is within the error range. The number of residues (132 amino acids) and the extinction coefficient for F2 (13,314.6 $M^{-1}cm^{-1}$) were calculated by amino acid analysis which further confirmed the homogeneity and size of the chimera.

2.3.2 Tryptophan Fluorescence Titrations

The metal binding affinity of F2 was determined by fluorescence titration following the Trp fluorescence intensity change in the presence of metal ions. The emission at 350 nm was observed after excitation at 280 nm (Figure 2.7, inset). The fluorescence intensity decreased systematically and gave a blue shift (~4 nm) of λ_{max} as the concentration of metal ions in solution increased, suggesting the tryptophan moved into a more hydrophobic environment caused by metal binding. Typically, the less solvent exposed environment increases the Trp emission intensity, but a quenched intensity was observed in the folding study of homeodomains due to its unique water-indole interaction upon folding.[104] With two metal binding sites in the chimeric dimer, it binds one to two metal equivalents with binding affinity of typical EF-hand-derived peptides.[105-107] Even though isolated EF-hand loops have been shown to be highly cooperative upon dimerization,[27,108] the two EF-hands in F2 were well isolated from each other and no cooperative binding was observed. Therefore, a linear non-least square algorithm fitting of 1:1 metal:protein associative model was applied to calculate the overall dissociation constant (K_d) of metal binding to the dimer. The conditional

dissociation constants for La^{3+}, Eu^{3+}, and Lu^{3+} were 1.4 ± 0.2, 1.6 ± 0.2, and 2.8 ± 0.9 μM respectively, which were very similar to the designed peptides and monomeric protein described previously.[68,82] Metallated F2 seemed to have no significant size discrimination among Ln^{3+} ions (Figure 2.7). In contrast, the K_d for Ca^{2+} and Mg^{2+} were 3.5 ± 0.1 μM and 2.5 ± 0.5 μM respectively. The K_d of Ca^{2+} is approximately 10-100 fold stronger than observed in other chimeras despite its lower valence charge than lanthanide ions. In general, the EF-hand loop should have no affinity towards smaller ions, such as Mg^{2+} due to the EF-hand loop's important role in discriminating Mg^{2+} from Ca^{2+} since there is 1000 fold more Mg^{2+} in cells than Ca^{2+}. It is important to note that Mg^{2+} has much slower water exchange than that of Ca^{2+} and Ln^{3+}. Therefore, its bigger sized, hydrated Mg^{2+} can give an outer-sphere coordination with the loop. To understand the possible Mg^{2+} binding mechanism, we utilized a complex with similar hydrated radius and inert ligand exchange, $Co(NH_6)_3$. Its similarity in binding affinity to Mg^{2+} would support an outer-sphere binding mode to hydrated Mg^{2+}. The binding affinity of $Co(NH_6)_3$, 2.0 ± 0.6 μM, was very similar to that of Mg^{2+}. The metal titration with increased salt concentration, 150 mM NaCl resulted in no change of metal binding affinity indicating that the measurements were not caused by collision interactions. Unlike Mg^{2+}, no hydrolytic activity was observed with $Co(NH_6)_3$ on an agarose gel electrophoresis (data not shown) further confirmed that hydrolysis was carried out through the water on Mg^{2+} rather than electrostatic interaction. Moreover, very weak (>mM range) metal binding affinity was observed with Fe^{2+}, Fe^{3+}, Mn^{2+}, Co^{2+}, and Zn^{2+} ions showing that the proteins had metal selectivity though the binding mechanism was significantly different from Mg^{2+}.

2.3.3 Secondary Structure Analysis by Circular Dichroism

CD spectroscopy was applied to investigate the global protein formation upon metal binding. Previous CD and NMR studies have shown that the monomeric chimera,

C2, exhibited a modest amount of secondary structure in the absence and presence of lanthanides.[82] Yet the % α-helicity in apoC2 and EuC2 is less folded than in the wild-type homeodomain [49] as shown by the disruption of the short α2 helix by severe residue overlap at α2 in the NMR spectra and of the 12-residue Ca-binding loop insertion nearby.[109] An estimated α-helical secondary structure based on molar ellipticity [θ_{222}] from CD data [110] indicates that the apoF2 was 16% α-helical while increasing to 22% upon Eu^{3+} binding (Figure 2.8). F2 was less folded when compared to the 25% and 32% α-helicity observed on apoC2 and EuC2 respectively, yet their shape and the induced pattern were similar. The reduction of helicity could be due to the insertion of a flexible linker which causes stress on the C- and N-termini upon folding. Correspondingly, the α-helicity of myoglobin measured by CD at pH 7.0 is only 30% compared to the 39% calculated α-helicity via Swiss PDB viewer. This indicated that the secondary structure of the chimera could be higher than its CD measurement after normalized with myoglobin. There was no change of secondary structure in the presence of Ca^{2+} and Mg^{2+} ions demonstrating that the native structure of the protein was not altered upon binding because the divalent charge of these metal ions gives less rigidity to the metal-binding loop than that of trivalent lanthanides. The interactions between helices were less rigid resulting in less overall secondary structure of F2 in the presence of divalent ions. The secondary structure proposed for F2 is supported by the data from temperature denaturation experiments. The melting temperature (Tm) of apoF2, Ca_2F2, and Eu_2F2 are 45, 33, and 54 °C respectively (Figure 2.9). The standard Gibbs free energy that was calculated from Equation 3 are 3.60, 2.76, and 11.06 kJ/mol for apoF2, Ca_2F2, and Eu_2F2 respectively suggesting that Eu_2F2 had overall better structural stability than Ca_2F2 and apoF2 had. The sigmoid shape of F2 was more pronounced than that of C2 indicating that F2 has a greater conformational transitional energy barrier than that of C2. The difference in observations may be due to the assistance of cooperative folding of one domain to another. Regardless of the lesser % helicity in the dimer, the overall protein

interaction and structural stability is stronger than the monomer showing that the dimer model improved the overall structural stability and thus may behave more like the parental DNA-binding properties.

2.3.4 DNA Binding Studies

An agarose gel retardation assay was performed to investigate DNA association of F2 (Figure 2.10). The choice of pTYB1 plasmid was used because it contains a single *engrailed* homeodomain site, TAATTA, and several TAAT sites (the core sequence for the homeodomain family). The percentage of the gel retardation enhanced as the concentration of apoF2 and metallo-F2 increased. Specific DNA interactions were observed at lower protein concentrations (up to 300 fold greater DNA) yet non-specific binding was also observed at higher protein concentration as the protein-DNA complex had no gel mobility due to the abundance of protein masking the negatively charged DNA plasmid (Figure 2.11). Although the consensus sequence for the dimer has yet to be elucidated, the DNA binding affinity displayed in this assay here was estimated to be starting at 900 nM (Figure 2.12). The overall binding affinity of apoF2 and Eu_2F2 is slightly stronger than that of Mg_2F2 and Ca_2F2 results that are consistent with the secondary structure studies. The more robust the secondary structure, the better that the protein fit to the DNA major groove of the recognition sequence. Further investigation on nuclease activity and DNA binding selectivity will be discussed in chapter 3 and 4 respectively.

2.4 Conclusions

A homodimer with a combination of structural and DNA binding properties was successfully designed by incorporating an EF-hand metal binding loop into a DNA binding *engrailed* homeodomain. The unique metal binding loop environment of the dimer is very intriguing because it is the first chimera to display such versatile metal selectivity which allows for smaller ions to settle in the active site and such properties

can be compared with some restriction endonucleases. Its secondary structure was slightly less folded yet more stable to denaturation than the monomeric form. These properties could allow more elasticity on reorganization upon DNA binding and therefore may give greater DNA hydrolytic activity. Regardless, this design provides a building block on understanding the importance of dimers and their protein-DNA interaction. The investigations on hydrolytic activity of the metal-bound dimer can be understood through kinetic assays and its sequence selectivity was determined by ^{32}P DNA sequencing gels, which will be further discussed in the next two chapters.

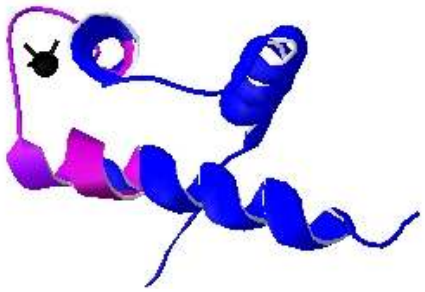


Figure 2.1 Three-dimensional view of the chimeric monomer design, C2. The DNA-binding HTH motif (blue) and the Ca-binding loop EF-hand motif (pink) with calcium ion (black) are structurally similar at α-α corner. Therefore, the turn sequence of the HTH motif can be modularly replaced by an EF-hand motif.

(A)

```
          1         10        20        30        40        50
H2N-MDEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLDKDGDGTIDEREIK
    IWFQNKRAKIKKSGGTGGAGDEKRPRTAFSSEQLARLKREFNENRYLTER
    RRQQLDKDGDGTIDEREIKIWFQNKRAKIKKS-COOH
```

(B)

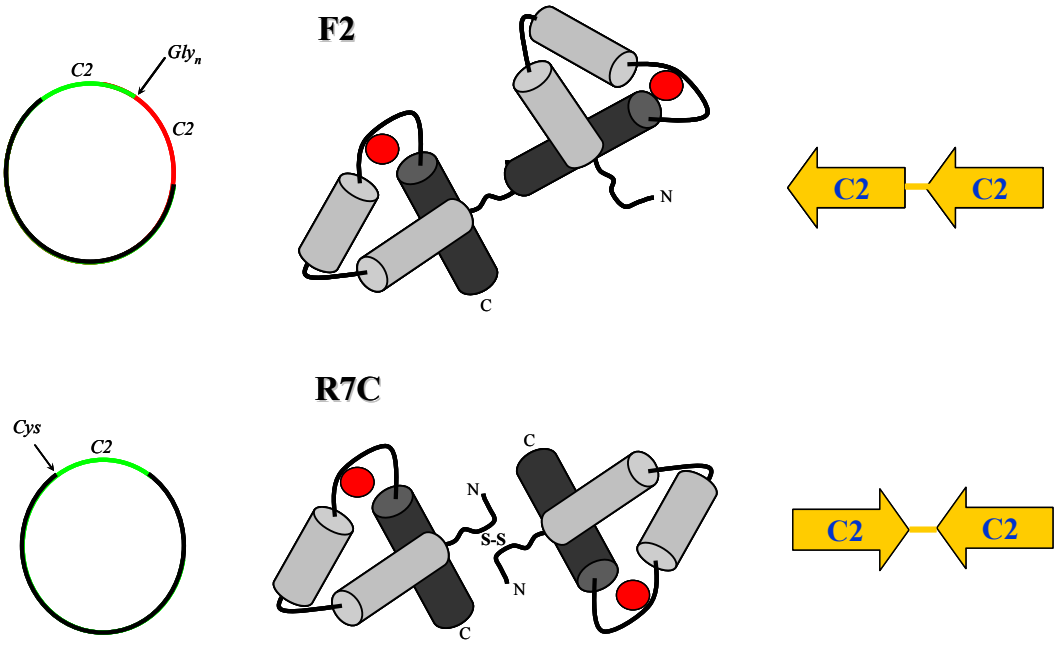


Figure 2.2 (A) The amino acid sequence of chimeric F2. Parent protein sequence is indicated in bold for the homeodomain and shaded for the EF-hand Ca-binding loop. The 7-residue linker between two monomers is boxed. (B) Schematic representation of our chimeric dimer design with two approaches.

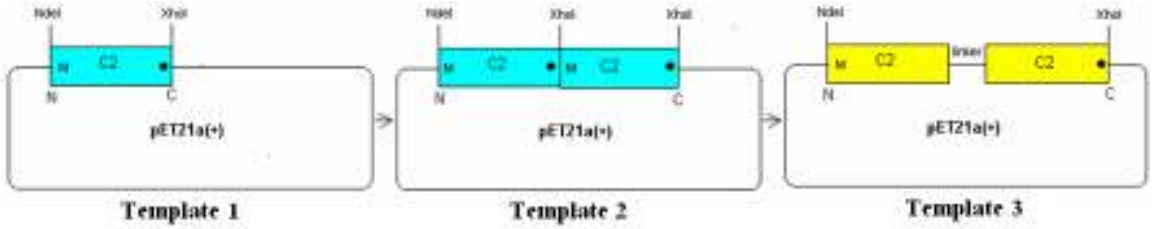


Figure 2.3 A schematic map of F2 DNA cloning. The cloning involved three stages: (1) Amplification of C2 with a *Xho*I site at both termini. (2) Ligation of the amplified C2 into the original vector which already contains one C2. (3) Insertion of the linker and the deletion of a start and a stop codons that are adjacent to the linker via multi-site mutagenesis technique. *Nde*I and *Xho*I are unique restriction enzyme sites for pET21a(+). N and C indicate the direction of the termini. M represents the start condon, methionine, while • represents the stop condon.

```
                                  XhoI
Wong 09F: 5'GAA GGA GAT ATA CTC GAG GAC GAC AAG CGC C 3'
                                  XhoI
Wong 07R: 5'AAG AAG TCG TAG CTC GAG CAC CAC CAC 3'

Wong 10F: 5'C AAG AAG TCG GGT GGT ACT GGT GGA GCC GGT GAC GAG AAG C 3'
Wong 10R: 3'G TTC TTC AGC CCA CCA TGA CCA CCT CGG CCA CTG CTC TTC G 5'
```

Figure 2.4 Oligonucleotide sequences were ordered and used as primers during PCR amplification and multi-site mutagenesis. The sites for restriction enzyme digestion are underlined. The DNA encoding for the 7-residue linker, GGTGGAG in amino acids, is shaded.

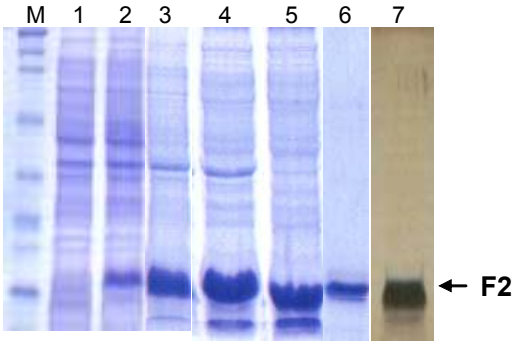


Figure 2.5 SDS-PAGE images of F2 purification steps. Lane M is prestained broad range marker. Lane 1 shows non-induced protein while lane 2 is induced one. The French pressed pellet was kept lane 3 and extracted by CB-1000 is lane 4. The protein was further purified by PEI precipitation (lane 5) and Heparin column (lane 6). F2 was finally purified through size exclusion column to 99% purity (lane 7).

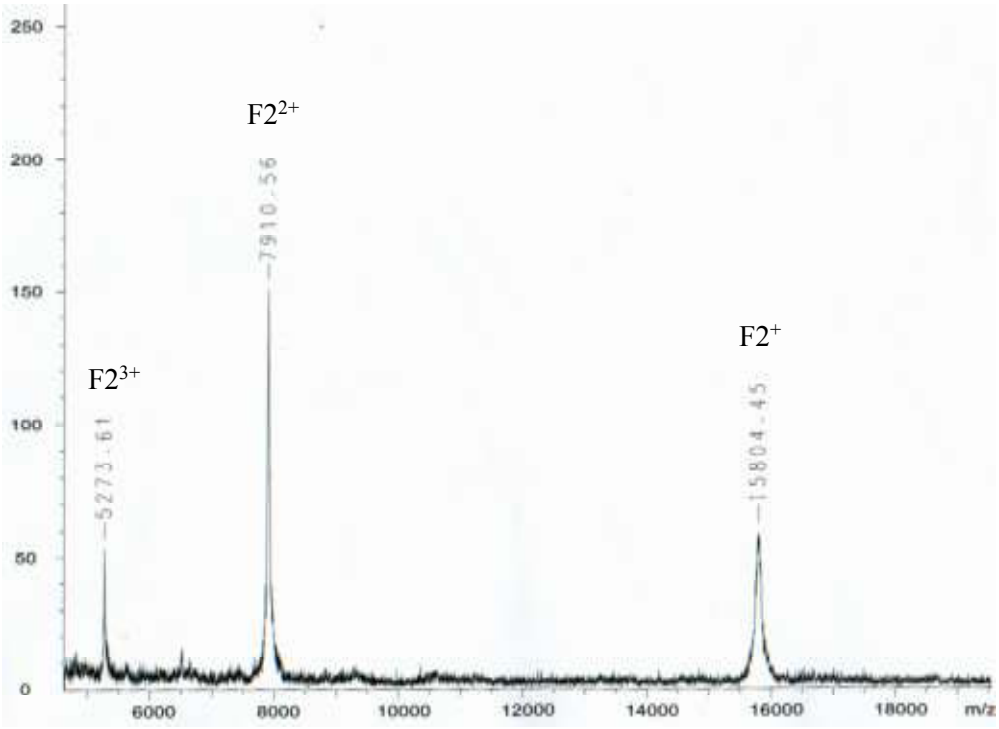


Figure 2.6 The MALDI-TOF spectrum of purified F2. The calculated molecular weight of F2 is 15820 *m/z* via DNASTAR software and the measured molecular weight of F2 is $(F2+H)^{+}$ at 15810 *m/z*, $(F2+H)^{2+}$ at 7911 *m/z*, and $(F2+H)^{3+}$ at 5274 *m/z* which is within the accuracy range.

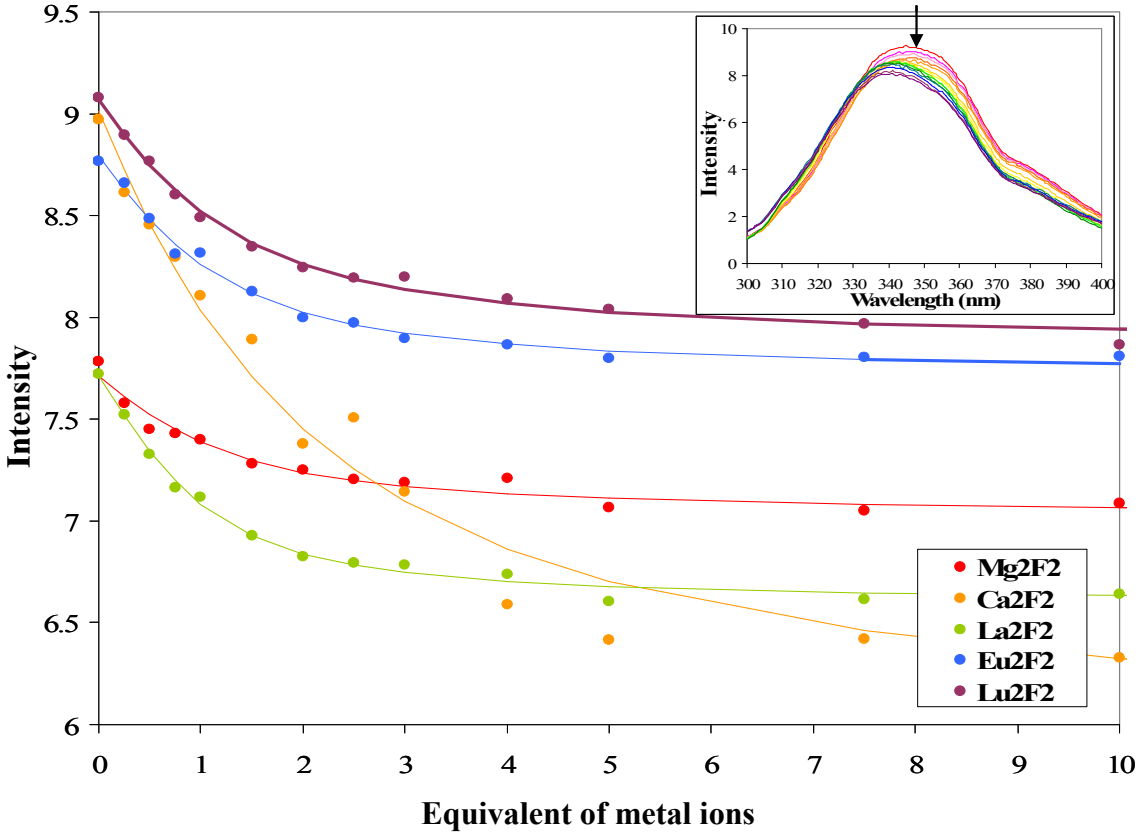


Figure 2.7 Fluorescence titrations of F2 with metal ions. The fluorescence intensity of Trp was followed at 350 nm (280 nm excitation) as a function of added metal (5 μM F2/10 mM HEPES pH 7.0 and 50 mM NaCl). The solid lines represent a linear non-least square algorithm fitting to a 1:1 association model, from which K_d was obtained. (*Inset*) Emission spectra showed throughout one titration, from 0 μM $LuCl_3$, upper most line to 100 μM $LuCl_3$, bottom most line.

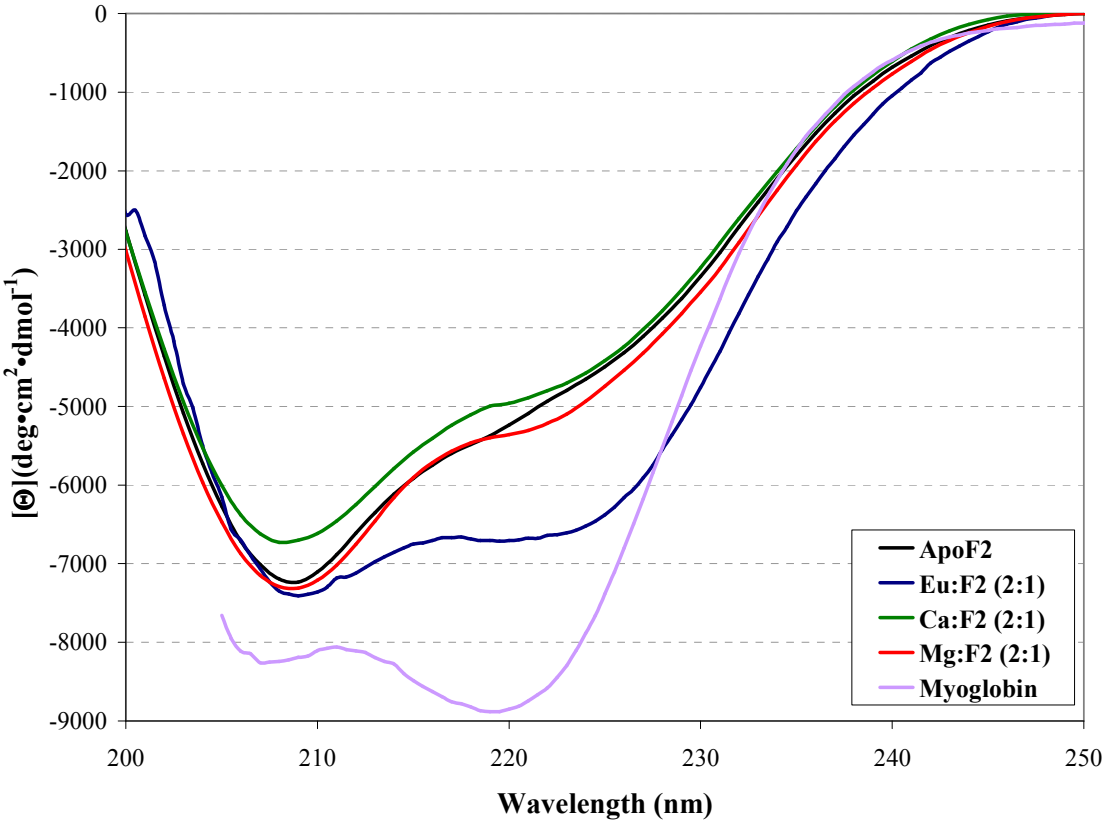


Figure 2.8 CD spectra of apoF2 (black), Eu_2F2 (blue), Ca_2F2 (green), and Mg_2F2 (red). The sample contained 15 μM F2 in 10 mM HEPES pH 7.0 and 50 mM NaCl.

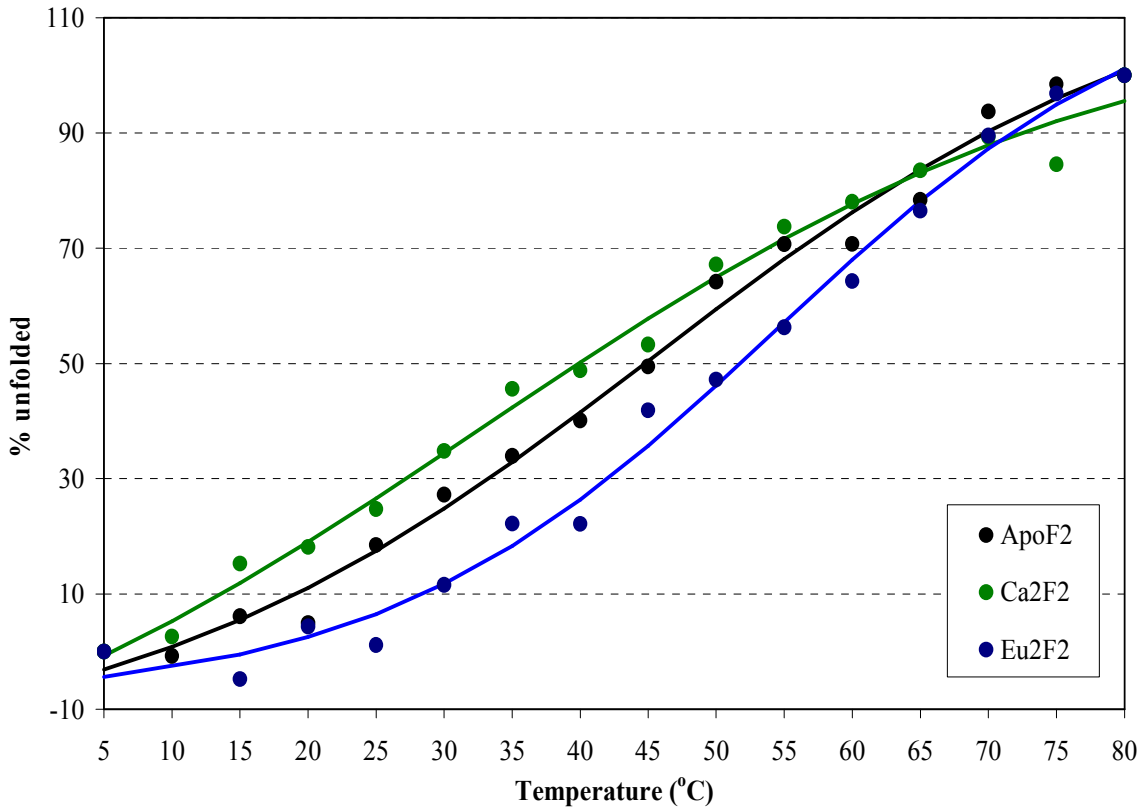


Figure 2.9 The thermal denaturation of apoF2 (black), Eu_2F2 (blue), Ca_2F2 (green) observed at A_{222}. The melting temperatures for apoF2, Eu_2F2, and Ca_2F2 are 45, 54, and 33 °C respectively fitted with a four-parameter algorithm which equation is adapted from reference 73.

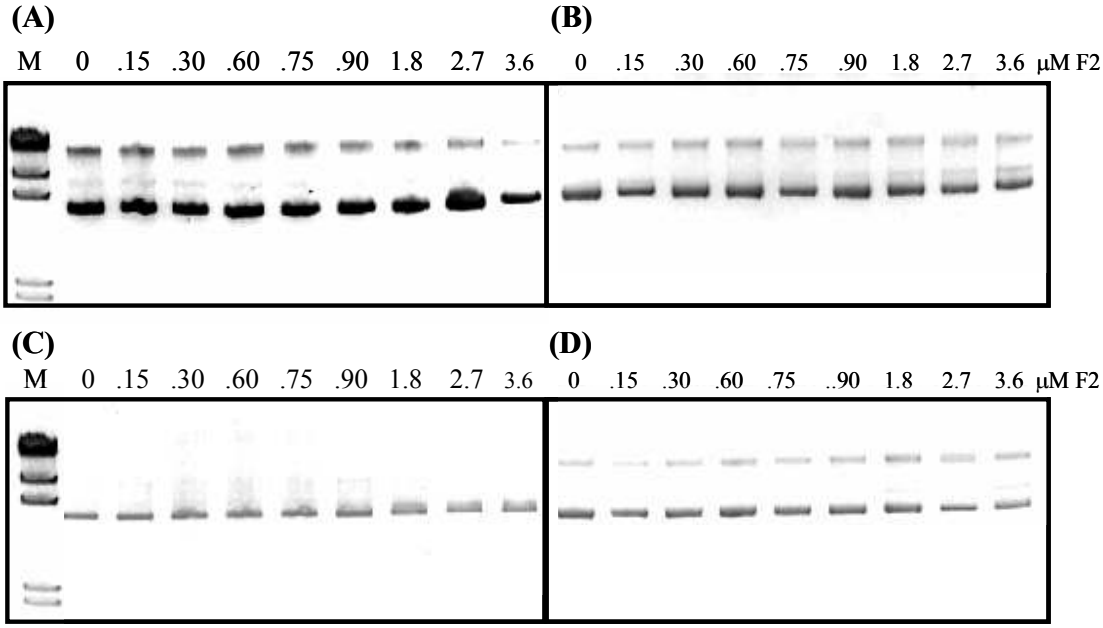


Figure 2.10 Agarose gel electrophoresis of supercoiled pTYB1 DNA bound with (A)ApoF2, (B)Ca_2F2, (C)Mg_2F2, and (D)Eu_2F2 in a buffer of 10 mM HEPES pH 7.0 and 50 mM NaCl. Lane M: Lambda Hind III fragment marker (Novitrogen). The increasing number lanes are indicated as shown with increasing concentrations of protein binding to DNA.

Figure 2.11 Agarose gel of DNA binding assay in the presence of apoF2. Lane M is a λ-Hind III fragment marker for size measurement. Lane 0 contained DNA plasmid only while lanes 4-64 represent the concentration of apoF2 in μM added to DNA plasmid.

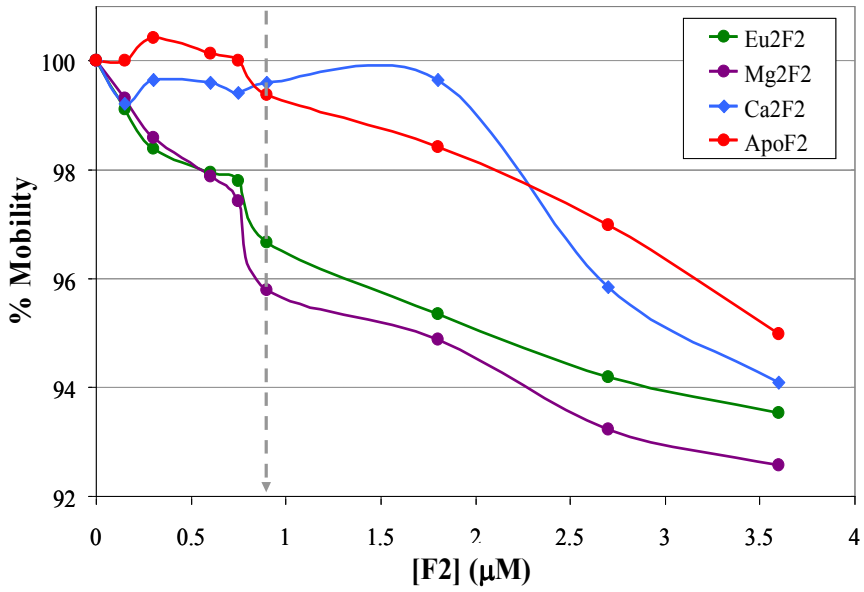


Figure 2.12 A plot of % gel mobility as a function of the concentration of metallo-/apo-F2. The gray arrow indicates the estimated overall binding affinity for F2 which is 900 nM.

CHAPTER 3

NUCLEASE ACTIVITIES OF THE CHIMERIC DIMER, F2

3.1 Introduction

One of the important characters in our chimeric design is the ability to incorporate new reactivity and specificity into a novel protein. The redesign of a structurally well-characterized and robust protein is a powerful approach to achieve new function abilities with the inclusion of metal binding sites. This designed metalloprotein can potentially give rise to new enzymatic activities that are unprecedented in biology and therefore is especially intriguing. In particular, the dimeric design is the next stepping stone from miniature protein/peptides design to an advanced biologically relevant approach, as a dimer has approximately doubled the DNA target sites length resulting in greater reactivity and specificity. Several groups engineered novel metalloproteins by incorporating metal sites into an unrelated scaffold in order to pinpoint reactivity of the metal sites or to avoid biological hazards sourced of the native protein.[10,111,112] Common examples of these applications are to incorporate a heme group in helix bundles or the studies of isolated copper binding sites from prion proteins. Our *de novo* design of a metalloprotein nuclease is based on this premise of the addition of a new metal site into a new context, leading to novel functionality.

The structural similarity of the HTH motif of a DNA binding protein and the calcium binding EF-hand motif has led us to consider the incorporation of a metal site into the α–α turn of a DNA-binding protein. Our previously designed miniature peptides, P5b and P3W, exhibited the abilities to both target and catalyze cleavage of DNA sequences, but their reactivity with dsDNA has yet to be refined and improved. To further enhance both selectivity and reactivity of the chimeras by taking advantage of being dimeric in biological systems, a structural-based dimer, F2, was designed as described in previous chapters.

The trivalent lanthanide ions, Ln^{3+}, have long been known to be an useful probe for studying the structure and function of Ca-binding proteins due to Ln^{3+}'s rich spectroscopic properties, their DNA hydrolytic activity, and their similar size, coordination environment, and ligand preferences to Ca^{2+}.[35] The Ca-binding EF-hand has never been reported to have enzymatic activity in its native context within physiological systems. But because Ln^{3+} have a higher charge to size ratio and a higher number of coordination ligands than Ca^{2+}, Ln^{3+} are a strong Lewis acid and thus water bound to the metals are better nucleophiles which promotes hydrolytic cleavage of dsDNA. By incorporating Ca-binding and DNA binding domains in the design of an artificial endonuclease, the catalytic metal is capable of performing specific DNA backbone cleavage through protein-DNA interactions.

The DNA cleavage and inhibition activity of metallo-F2 was investigated and the results are presented in this chapter. Activities were also compared between a DNA analog and a supercoiled DNA plasmid in order to gain an insight on the cleavage preference of F2. The reactivity of the chimeric dimer has been shown to be reactive not only in the presence of Ln^{3+}, but also with Ca^{2+} and other smaller metal ions. Contrasting DNA reactivity results between the monomer and the dimer were puzzling, yet theses results can be comparable with restriction enzymes as they are abundant in homodimers. The kinetic studies of F2 open the door to understanding the functions of dimers in biological systems and lay a path towards multi-domain protein design, thereby providing new avenues for future metalloprotein designs.

3.2 Experimental Procedures

3.2.1. Reagents and Proteins

Buffers, agarose and other reagents were obtained from Fisher Scientific. Protein was prepared and purified as described in Chapter 2. The stock solutions of metal chlorides, $LnCl_3 \cdot (H_2O)_6$, $FeCl_3 \cdot (H_2O)_6$, $FeCl_2 \cdot (H_2O)_4$, $CaCl_2$, $CoCl_2 \cdot (H_2O)_6$,

$MnCl_2 \cdot (H_2O)_4$, $MgCl_2 \cdot (H_2O)_6$, and $ZnCl_2$, were made fresh by weight prior to use. The stock solution of $FeCl_2$ was prepared freshly and the experiment was completed within an hour. All solutions were prepared in deionized distilled water (MilliQ 18 mΩ).

3.2.2 DNA Plasmid Cleavage Assay

To demonstrate the catalytic activity of an exposed EF-hand, the cleavage of supercoiled plasmid DNA complexes with metallated F2 was observed via agarose gel electrophoresis. Various concentrations of Ca_2F2 or Eu_2F2 solutions (0-10 μM) were freshly prepared by adding stoichometric amounts (2:1) of metal ions before adding to 350 ng of plasmid (pTYB1, New England Biolabs) in 10 mM HEPES pH 7.0 and 50 mM NaCl. The complex was incubated for 10 minutes at 37 °C prior to the extraction of DNA using a Geneclean kit (Bio-rad). The samples with loading dye were loaded to a 1% agarose gel that was run for approximately 1 hour in 1 x TAE buffer (Tris-acetate-EDTA) at 120 V. The gels were visualized under UV-light and photographed with a Kodak digital imager.

3.2.3 pH-dependent DNA Cleavage Assay

The optimal pH for cleavage of supercoiled DNA plasmid in the presence of Eu_2F2 and Ca_2F2 was determined by agarose gel electrophoresis. A buffer was prepared with a combination of sterile acetic acid, HEPES, MES, and ethanolamine to provide a large pH range and to control material consistency. The pH was adjusted to desired values using NaOH and HCl, and followed with a pH meter. Samples of 220 ng pTYB1in the10 mM buffer and 50 mM NaCl at different pHs were mixed and incubated with 1.50 μM metallated F2 at 37 °C for 10 minutes. The reaction was quenched by the addition of 2.5 mM EDTA. The pH was adjusted back to a narrow pH range (5.0-7.6) with 500 mM HEPES pH 7.0 to preserve consistency. The samples were loaded onto a 1% agarose gel and run for approximately 1 hour in 1 x TAE buffer at 120 V. The gels were visualized under UV-light and photographed with a Kodak digital imager.

3.2.4 Time-dependent DNA Cleavage Assay

To address the cleavage rate of supercoiled DNA plasmid with a hydrolytically active metallo-F2, cleavage product formation was investigated at different incubation times (from 0 to 10 hours) via agarose gel electrophoresis. Samples of 100 ng of pTYB1 were mixed and incubated with 643 nM Eu_2F2 at 37 °C in a buffer of 10 mM HEPES pH 7.0 and 50 mM NaCl. The reaction was quenched by 100 mM EDTA at different incubation time intervals. The DNA of each sample was isolated by a Geneclean kit and was loaded onto a 1% agarose gel that was then run for approximately 1 hour in 1 x TAE buffer at 120 V. A control sample at 0 time was run as well to ensure the quenching and Geneclean purification did not impact cleavage results. The gel was visualized under UV-light and photographed with a Kodak digital imager. The band intensity of cleaved and uncleaved DNA was measured with a Kodak digital image analyzer. The intensity of the supercoiled band was corrected by a factor of 1.47 as a result of its lower staining capacity by ethidium bromide.[113] A graph with the percentage of the agarose gel DNA band intensity was plotted as a function of time to follow the appearance of cleaved DNA over time.

By measuring the degree of cleavage activity observed from agarose gels at various F2 concentrations at 10 minutes time intervals, the pseudo-first-order rate and the rate of reaction was thus calculated using the following equations.

Equation 3.4: $y = y_0 \exp(-k_{obs}x)$

Equation 3.5: $k_{obs}' = V_{max}'[F2]/(K_M+[F2])$

The pseudo-first-order rate was obtained from equation 3.1 which was then simplified into $\ln(y/y_0)=-k_{obs}x$, where y is the total percentage of the cleaved DNA at a given time, y_0 is the total percentage of the uncleaved DNA including nicked and linear

bands at time 0, and x is the incubation time for the hydrolytic rate, k_{obs} was then determined from the slope of the $\ln(y/y_0)$ in the function of incubation time (x). V_{max}' is the maximal reaction velocity.

The K_M and V_{max}' was calculated using equation 3.2.[62] DNA cleavage assays with constant DNA concentration (350 ng) and various Ca_2F2 concentrations (μM) were performed, and band intensities were plotted against substrate concentration in terms of equation 3.2. The equation 3.2 was then fitted with Michaelis-Menten kinetics and the V_{max}' and K_M were then calculated using SigmaPlot 8.0 software.

3.2.5 Metal-dependent DNA Cleavage Assay

Since F2 has been shown to have binding affinity towards the smaller sized ion, Mg^{2+}, in chapter 2, it is important to understand its metal versatility by investigating its nucleolitic ability in the presence of various metal ions other than the standard metals for an EF-hand loop (Ca^{2+} and Ln^{3+}). Metal ions, such as Mg^{2+}, Co^{2+}, Mn^{2+}, Zn^{2+}, Fe^{2+}, and Fe^{3+}, are abundant in biological system and were therefore selected for this assay. A concentration of 643 nM apoF2 was incubated with two equivalent amounts of metal ions at 37 °C for 5 minutes. The metallated-F2 was then mixed and incubated with 100 ng of DNA plasmid at 37 °C for 60 minutes in a buffer consisting of 10 mM HEPES pH 7.0 and 50 mM NaCl. Control samples of DNA plasmid with various metal ions were run at the same time to determine DNA cleavage induced by metal ion itself. Samples were quenched with EDTA and loaded onto a 1% agarose gel. The gel was then run for approximately 1 hour in 1 x TAE buffer at 120 V and visualized under UV-light and photographed with a Kodak digital imager. The assay was repeated three times and averaged to confirm the consistency of the results.

3.2.6 BNPP Kinetic Studies

In order to determine the phosphadiester cleavage rate of metallo-F2, a DNA analog, Bis(4-nitrophenyl) phosphate (BNPP) was used as substrate, since the hydrolysis

product is bright yellow. Therefore, a time-course observation at 400 nm via UV-Vis spectrophotometer (Cary 300 Bio) was used to follow the reaction. The UV-Vis cuvettes, the holder, and the metallo-protein were pre-incubated at 37 °C for at least 10 minutes to ensure stable temperature from the beginning of the experiment. Various concentrations of Ca_2F2 or Eu_2F2 in a buffer of 10 mM HEPES pH 7.0 and 50 mM NaCl were mixed immediately with freshly prepared BNPP to the final concentration of 500 μM. The control samples in the reference cells contained 500 μM of final concentration of BNPP in the same buffer. Measurements were recorded from 350 to 450 nm every 10 minutes for the first hour and every 30 minutes for the remaining 9 hours of incubation. The absorbance at 400 nm was plotted as a function of time. The apparent first order rate, k_{obs}, was obtained by fitting the first 90-minute slope from a graph of an absorbance at 400 nm versus time. The second order rate, K, was obtained from the slope of k_{obs} as a function of protein concentrations. The experiment was repeated twice to obtain a standard deviation. The K values for $CaCl_2$ and $EuCl_2$ were measured for the basal control relative to metallo-F2.

3.2.7 Phosphate Inhibition Studies

3.2.7.1 Investigation with BNPP

Previously kinetic studies of chimeric peptides with BNPP have revealed hydrolytic activity inhibition by the product of BNPP/NPP cleavage, phosphate ion. Therefore, the degree of phosphate inhibition on BNPP cleavage by metallo-F2 was determined by adding increasing stoichometric amounts of phosphate. The phosphodiester cleavage of BNPP by Eu_2F2 was monitored at 400 nm via UV-Vis spectrophotometer with different equivalents of sodium phosphate pH 7.0. The cuvettes and the holder were pre-incubated at 37 °C for at least 10 minutes. Eu_2F2 (10 μM) was dissolved in a buffer of 10 mM HEPES pH 7.0 and 50 mM NaCl with different equivalents of sodium phosphate pH 7.0, and these were incubated at 37 °C for 5

minutes. The samples were then mixed with freshly prepared BNPP to the final concentration of 500 μM immediately before scanning from 350 to 450 nM. The control samples in the reference cells contained 500 μM BNPP in the same buffer. The measurement was scanned every 3 minutes for the first 30 minutes and every 20 minutes for the remaining of the 9.5 hours of incubation. The absorbance at 400 nm was plotted against time to calculate the first order rate constant. The percentage of activity was calculated based on assigned 100% activity to a non-sodium phosphate sample. A graph of % activity was plotted as a function of the equivalents of sodium phosphate to determine the actual inhibition factor for metallo-F2 by phosphate ions.

3.2.7.2 Investigation with DNA Plasmid

The cleavage inhibition of M_2F2-DNA complex in the presence of phosphate can potentially be utilized as a "switch" for hydrolytic activity in biological systems. The degree of cleavage activity toward DNA plasmid substrate was observed using agarose gel electrophoresis. All the samples consisted of a buffer of 10 mM HEPES pH 7.0 and 50 mM NaCl with different equivalents of sodium phosphate at pH 7.0, which was then pre-incubated with 2 μM Eu_2F2 at 37 °C for 10 minutes. A concentration of 0.02 μM of pTYB1 was then added to the sample and the mixtures were incubated at 37 °C for 12 hours.

In order to determine whether the inhibition was caused by the association of Eu^{3+} in F2 with free PO_4^{3-} in solution resulting $EuPO_4$-protein precipitate, a set of samples was prepared to contain excess amounts of $EuCl_3$ at different equivalents of phosphate. Different levels of sodium phosphate was mixed and incubated with 60 μM $EuCl_3$ at 37 °C for 10 minutes to form $EuPO_4$ in order to eliminate the ability to $EuPO_4$-protein precipitation. The samples were finally incubated at 37 °C for 12 hours after adding 0.02 μM of pTYB1 and 0.2 μM Eu_2F2 to the mixture. Samples with loading dye were charged onto a 1% agarose gel and were then run for approximately 1 hour in 1 x TAE buffer at

120 V. The gels were visualized under UV-light and photographed with a Kodak digital imager. The assay was repeated twice to confirm the consistency of the results.

3.3 Results and Discussion

3.3.1 DNA Plasmid Cleavage Assay

The DNA plasmid cleavage assay was performed with apoF2, Ca_2F2, and Eu_2F2 in order to investigate the hydrolytic activity of F2 in the presence and absence of metal ions (Figure 3.1). The DNA plasmid pTYB1 was chosen because it contains a single *engrailed* homeodomain site, TAATTA, and several TAAT sites (the core sequence for the homeodomain family). The cleavage activity was followed as supercoiled (uncleaved) DNA converted into nicked/linear DNA with increasing concentration of metallo-F2 (Figure 3.1), and then ultimately to smaller fragments. The samples with apoF2 are the controls for the assay since F2 requires metal ions to exhibit reactivity. Not only was Eu_2F2 observed to carry-out dsDNA cleavage, but Ca_2F2 also displayed a similar degree of activity. Yet, the monomeric calcium-bound peptides and protein had been shown to have no hydrolytic activity at neutral pH in previous studies even they are active with Eu^{3+}.[11,81,87,114] Even though Eu^{3+} itself hydrolyzed phosphodiester bonds at k_{cat}/K_M of 1.7,[63] its rate (lane Eu^{3+} in Figure 3.1) was much slower than that with Eu_2F2 indicating that activity was primarily contributed by the metallodimer. The degree of hydrolytic activity was especially surprising since our previously designed monomers have limited reactivity with DNA plasmids.

In fact, NMR structure of LaC2 has shown to have a different globe fold than *engrailed* causing the EF-hand loop to be more buried than predicted.[109] F2 may shape more closely to the *engrailed* which allows for more solvent accessibility on the metal sites and lowers the activation barrier for bond breakage with the assistance of two active sites. Perhaps, the overall structure of F2 was in a different configuration from C2 since the third helix of C2 was observed to be flipped 180° away from the predicted model

even though the major interaction between helix 1 and 3 remained intact (Figure 3.2). As a result, the energy barrier between the predicted model and observed structure could be small. The structure of the dimer could, therefore, easily interconvert with the predicted model form which in turn could promote metal availability on the loop region. The dimer may have an enhanced ability to hinge to the major grooves with the help of an extra binding domain and thus gave greater hydrolytic activity. As the concentration of metallated F2 was increased, the DNA was further cleaved into smaller fragments suggesting that the sequence recognition of F2 is not as specific as wild type homeodomain. According to the DNA-binding assay in chapter 2 (Figure 2.11), F2 has some degree of sequence specificity at a concentration of <4 μM indicating that the threshold between non-specific and specific recognition/cleavage could be quite narrow suggesting that several conformations of F2 could give different degrees of contacts on the major grooves. Nonetheless, F2 has been shown to have sequence selectivity which will be discussed in detail in chapter 4.

3.3.2 pH-dependent DNA Cleavage Assay

The pH-dependent DNA cleavage assay was performed to determine the optimal working conditions for the metalloprotein. Eu_2F2 had a wider pH working range (6-11) with DNA plasmid than that of Ca_2F2 (7-10) as indicated in Figure 3.3. The more robust secondary structure and stability of Eu_2F2 than Ca_2F2 may result in its broader activity range in pH. Therefore, Eu_2F2 showed greater tolerance in harsh environments as Eu^{3+} is relatively well-complexed in the loop.

The cleavage activity of both metalloproteins at basic pH condition is a strong indication of metal's accessibility to substrate. Presumably, the residues could be deprotonated by water or hydroxide in the basic solution, negatively charged side chains at the metal-binding loop became better nucleophiles for metal coordination. Therefore,

the hard metal ion, Eu^{3+}, was well-coordinated in the loop which contributed to the reactivity of metallo-F2.

3.3.3 Time-dependent DNA Cleavage Assay

To address the rate of DNA cleavage activity with metallo-F2, the protein-DNA complex was incubated at various reaction times ranging from 0 to 10 hours, and was analyzed using agarose gel electrophoresis (Figure 3.4). With the addition of Ca_2F2, the intensity of the uncleaved (supercoiled) DNA band decreased while the cleaved (nicked and linear) DNA bands brightened simultaneously over time. The first appearance of the nicked DNA band followed by the appearance of the linear DNA band indicating that dsDNA was cleaved through a two-step mechanism. The band intensity of cleaved and uncleaved DNA were then quantified and plotted against time (Figure 3.5). The cleavage rate started to slow down after 5 hours of incubation demonstrating the effective reactivity time of Ca_2F2 at a constant concentration.

The pusedo-first-order kinetics was calculated using Equation 3.1 at different concentrations of Ca_2F2 and averaged over 3 trials (Figure 3.6). The K_M and V_{max}' were thus determined as 2.3 ± 0.6 μM and $17.2 \pm 1.8 \times 10^{-4}$ s^{-1} respectively by fitting with the Michealis-Menton kinetics equation (Equation 3.2). The k_{cat} was calculated as $V_{max}/[E_0]$, where $[E_0]$ is the concentration of Ca_2F2 (0.643 μM). As a result, the substrate specificity (k_{cat}/K_M) of Ca_2F2 is 7.50×10^2 $M^{-1}s^{-1}$. When comparing the k_{cat}/K_M of Ca_2F2 with that of Eu^{3+} [63] and restriction enzymes, BamHI [115] and EcoRV [116], the specificity of the restriction enzyme is three magnitudes greater than that of Ca_2F2 (Table 3.1). Nonetheless, Ca_2F2 is the first small (<100 kDa) metallochimeric dimer with substantial dsDNA cleavage activity and the studies on sequence preference of F2 will be discussed further in Chapter 4.

3.3.4 Metal-dependent DNA Cleavage Assay

Since the Trp fluorescence and CD titration studies of F2 have demonstrated that F2 has the capacity to bind to metal ions other than Ca^{2+} and Ln^{3+}, the reactivity with other biologically relevant metal ions should be investigated to understand the metal binding versatility of F2. We, therefore, may gain insights on the uniqueness of the metal binding mechanism in our designed dimer (Figure 3.7). Even though apoF2 appeared to be not completely metal-free on the gel, the same experiment has been performed four times with various batches of purified apoF2. Therefore, the consistency of the results led us to believe that the observations of the metal cleavage activity were reasonable. A control DNA cleavage assay in the presence of metal ions alone was performed to show that the hydrolytic activity on DNA was predominately induced by metallo-F2 not metal ion alone. It was very intriguing that not only did Ca^{2+}- and Ln^{3+}-bound F2 show hydrolytic activities, but in the company of other smaller metal ions, Mg^{2+}, Co^{2+}, and Mn^{2+}, F2 displayed an even greater DNA hydrolytic activity.

The cleavage mechanism of Mg^{2+} might be carried out through a pathway similar to that of restriction enzymes.[15] In these examples, the P-O bond of DNA backbone coordinates to the metal site became a better leaving group/electrophile and thus caused the breakage of phosphodiester bonds with the assistance of Mg^{2+}. Mn^{2+} has been reported to enhance the hydrolytic activity of an enzyme when Mn^{2+} was surrogated to a Mg-binding protein.[117,118] Nonetheless, the weak metal binding affinity of F2 with Co^{2+} and Mn^{2+} strongly indicates that both metal ions may appear to be just electrostatic association and do not have a well defined binding interaction at the metal-binding loop of F2 which in turn implied that the DNA cleavage mechanisms could be completely different from that of Mg^{2+}, Ca^{2+}, and Ln^{3+}.

In contrast, the monomeric protein, C2, hardly demonstrated any DNA hydrolytic activity and was shown to bind to Ca^{2+} and Ln^{3+} selectively yet its binding affinity with

Ca^{2+} was 10 times weaker than that of F2. These differences in metal ion versatility between F2 and C2 suggest that the metal-binding loop of the dimer is shaped to be more accessible than that of LaC2 since its NMR structure has different global fold than *engrailed* causing the EF-hand loop to be more buried than predicted.[109] However, the actual metal binding and cleaving mechanisms of metallated F2 have yet to be elucidated due to the lack of F2's crystal structure.

3.3.5 BNPP Kinetic Studies

Ln^{3+} complexes have been known to cleave phosphate esters through a hydrolytic rather than a redox mechanism. Many of the publications on small lanthanide complexes utilized a DNA analog to study their rate constants which is on the order of $k_{obs} = 10^{-2}$ - $10^{-6} s^{-1}$.[11,119-121] BNPP was chosen for this work since its phosphodiester cleavage can easily be monitored by UV-Vis spectrometry. As the nitrophenolate group is released from BNPP by the addition of water causing ester bond cleavage from the phosphate, the nitrophenolate group absorbs at 400 nm making the solution yellow in color. The pseudo-first order rate constant of the metalloprotein was then determined by monitoring the release of nitrophenolate group at 400 nm over time with excess amount of substrate (500μM BNPP) as less than 10% product conversion in 10 hours of reaction. A decrease in rate was observed after 90 minutes of incubation for all concentrations of M_2F2 implicating that decomposition of protein at 37 °C could contribute to the gradual rate reduction. The data was therefore fitted with linear regression for the initial 90-minute of reaction and the rate constant (k_{obs}) was calculated for different concentrations of the metalloprotein (Figure 3.8).

It is important to note that the cleavage rate of BNPP with the metalated dimer was not significantly faster than with the metalated monomer while there were visible cleavage rate differences with DNA plasmid. The slower calculated BNPP hydrolysis rate with the dimer would be affected by the DNA specificity of *engrailed* homeodomain.

Homeodomains are known to have very specific residue contact with the DNA major groove while BNPP is a small, negatively charged molecule that requires little regioselectivity for hydrolytic activity. The visible, cleaved DNA observed in the DNA cleavage gel is a strong indication of selective dimer-DNA interactions. Therefore, the reduced hydrolytic activity observed on BNPP could be affected by molecule-protein collisions and blockage of active sites with charged BNPP.

We have shown that Eu_2F2 and Ca_2F2 have much more robust DNA cleavage activities than previously designed chimeras in Figure 3.1. BNPP was thus chosen for direct comparison of hydrolytic rate among chimeras. The rate constant of BNPP in aqueous solution alone was estimated at $6 \times 10^{-11} s^{-1}$ at 37°C at pH 7.[122] Eu_2F2 and Ca_2F2 carried out BNPP hydrolysis with second order rate constant (K) of 1.36×10^{-1} and 8.14×10^{-2} $M^{-1}s^{-1}$, respectively, which is about 3 to 6 times faster than those of other chimeric designs (Table 3.1). The rate constant of Eu^{3+} was similar to that of chimeric monomers, while no cleavage activity of BNPP was observed in the presence of Ca^{2+}. The hydrolytic activity exhibited by monomers could be primarily caused by the presence of free Eu^{3+} in the protein solution. Meanwhile, the rate constant of F2 was about two orders of magnitude faster than Eu^{3+}. The data from the BNPP studies along with the observations from the DNA cleavage assay indicated that F2 is indeed the first truly structural-based artificial nuclease with significant DNA hydrolytic activity.

3.3.6 Phosphate Inhibition Studies

Our previously studied chimeric metallopeptides have shown that excess phosphate could inhibit the hydrolytic activity of chimeras. The rationale behind phosphate's role on our chimeras' reactivity with BNPP is not well understood. Here we examine the activity of Eu2F2 in the presence of excess phosphate with BNPP and DNA plasmid with the intention to utilize phosphate as a "switch" for DNA hydrolytic activity. A graph of % hydrolytic activity of Eu_2F2 was plotted as a function of sodium phosphate

equivalents in a 500 μM BNPP solution agreed with the inhibition results of our chimeric peptides (Figure 3.9). The activity reduced dramatically to <10% at two equivalents of phosphate which is equal amount to the Eu-binding sites in the dimer suggesting that a $EuPO_4$-protein precipitate or a PO_4^{3-} bound Eu_2F2 could be the cause of inhibition. However, an unexpected result of the phosphate inhibition study with a DNA plasmid using agarose gel electrophoresis has suggested otherwise (Figure 3.10). The activity of the metalloprotein remained the same with up to three equivalents of phosphate while the DNA band intensity was affected in the presence of excess Eu^{3+}. These two results seemed to contradict each other, but in fact they provided us with evidence on how phosphate influences the catalytic activity of Eu_2F2.

The rules of charge repulsion and attraction play an important role in our hypothesis. The negatively charged phosphate might coordinate to Eu^{3+} at the binding sites and thus inhibit the interaction of the active sites with the relatively small BNPP molecules. On the other hand, the large, highly negatively charged DNA plasmid repulsed the negatively charged phosphates which allowed room for the Eu_2F2-DNA interactions. The excess free positively charged Eu^{3+}, however, occupied the negatively charge DNA backbone and thus inhibited Eu_2F2 cleavage. The effects from the charged Eu^{3+} could be observed with lane Eu^{3+} where supercoiled DNA band migrated slower than lanes without excess Eu^{3+}, such as lane DNA (Figure 3.10). The charge of the DNA plasmid was partially neutralized by Eu^{3+} and thus was less affected by the current resulting less band migration. The DNA cleavage activity in the presence of excess phosphate also showed that the Eu^{3+} was well situated in the binding loop where the precipitation of $EuPO_4$ was discouraged. Phosophate ions indeed inhibited hydrolysis with BNPP, but a basic yet important insight of what facilitates the inhibition was revealed with the DNA plasmid inhibition study, thereby presenting a plausible role of free phosphate inhibition of the nuclease.

3.4 Conclusions

A painting of the fearsome creature in Greek mythology, chimera, has finally been completed. The fire-breathing head of a goat (Ca-binding EF-hand loop), the body of a lion (DNA-binding *engrailed* homeodomain), and the tail of a dragon (new functionality with Ln^{3+} substitution) are packed into this chimeric dimer. Its power is indeed shown to be larger than its original components combined. The structural-based chimeric design is a powerful tool for gene modification and targeting as well as building a foundation for dimeric design which could be used for potential pharmaceutical enhancement. We have shown that there are advantages and disadvantage in designing a homodimer. The hydrolytic activity of the dimer with DNA plasmid and BNPP was several magnitudes greater than of the monomer. Not only did the dimer cleave double stranded DNA, but it also exhibited a potential enzymatic behavior with biological relevant metal ions other than the well-known hydrolytic active Ln^{3+}. However, it was very difficult to obtain structural information of the flexible and DNA cleaving dimer which eliminates the opportunity of co-crystallization with natural DNA duplex. Nonetheless, the overall information we have gained from this dimeric design has outweighed the difficulties on its structural determination.

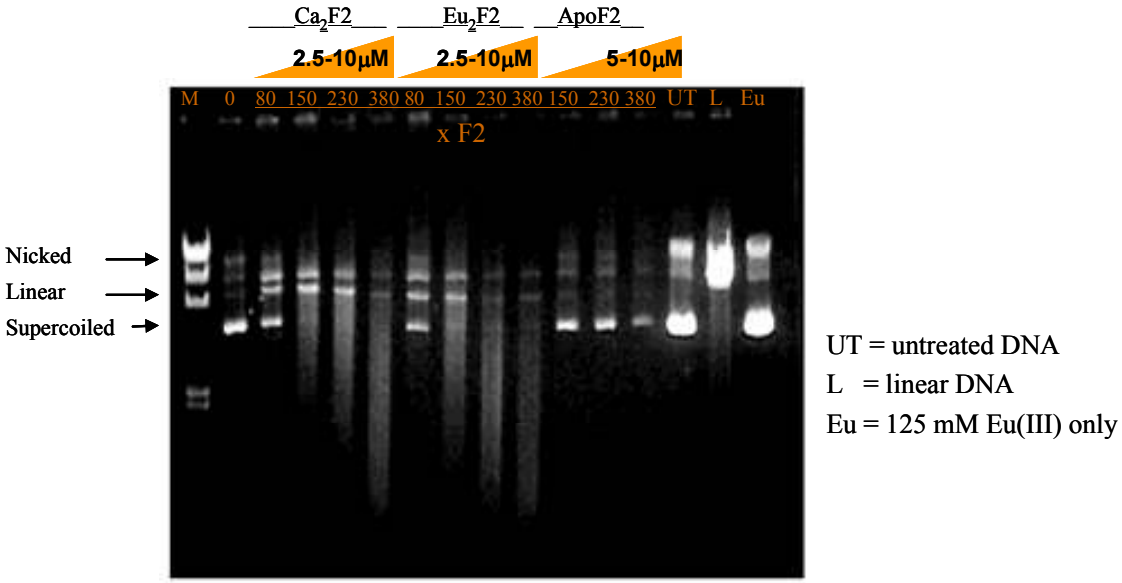
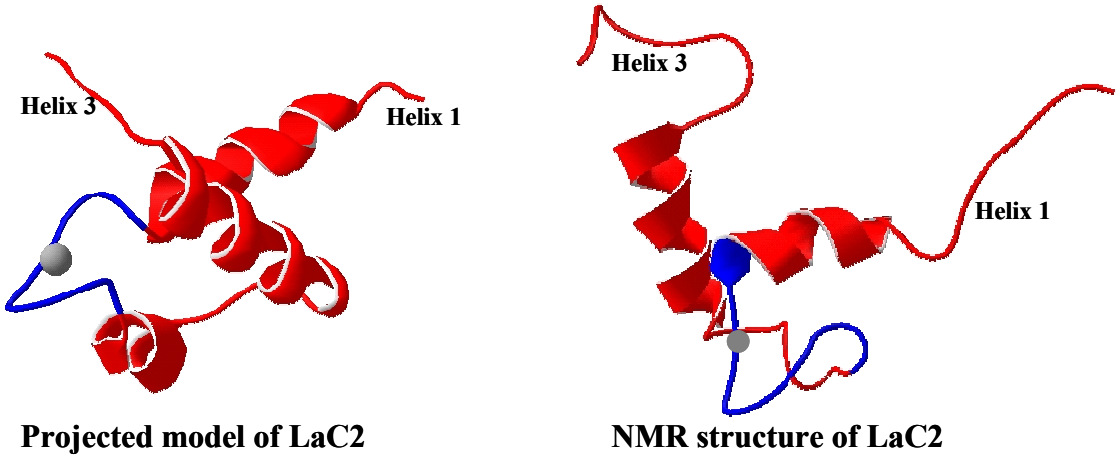


Figure 3.1 Agarose gel electrophoresis of supercoiled pTYB1 DNA cleaved by increasing concentrations of apoF2 or metallated F2 (10 mM HEPES pH 7.0 and 50 mM NaCl as indicated). Lane M: Lambda Hind III fragment marker (Novitrogen). Lane 0: Supercoiled pTYB1 DNA in the absence of protein. The remaining lanes are indicated as shown with increasing concentration of protein which has interacted with DNA.

(A)



(B)

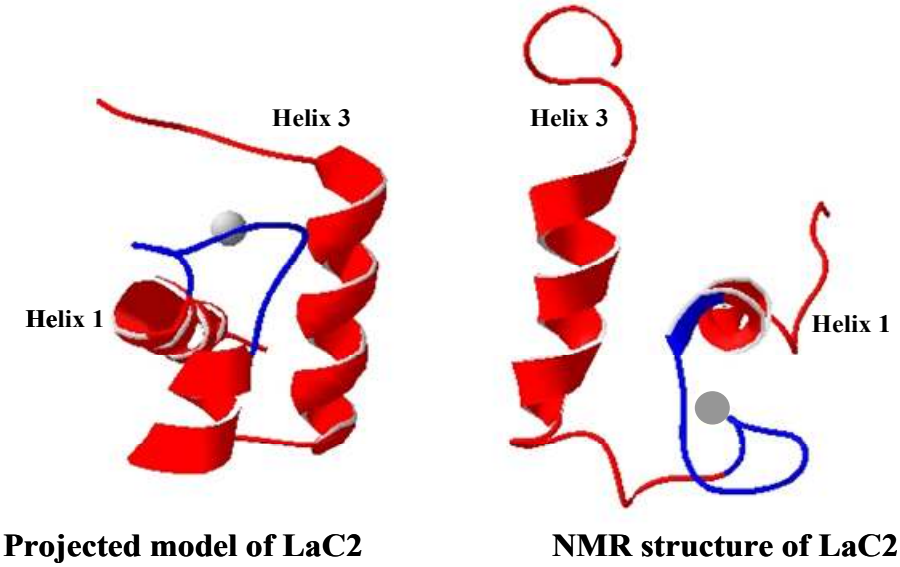


Figure 3.2 (A) The structural comparison of projected model and NMR structure of LaC2 demonstrated that the orientation of helix 3 was flipped to the opposite side of helix 1. The metal binding loop and the metal ion are represented in blue and gray respectively. (B) A 90° turn view of (A).

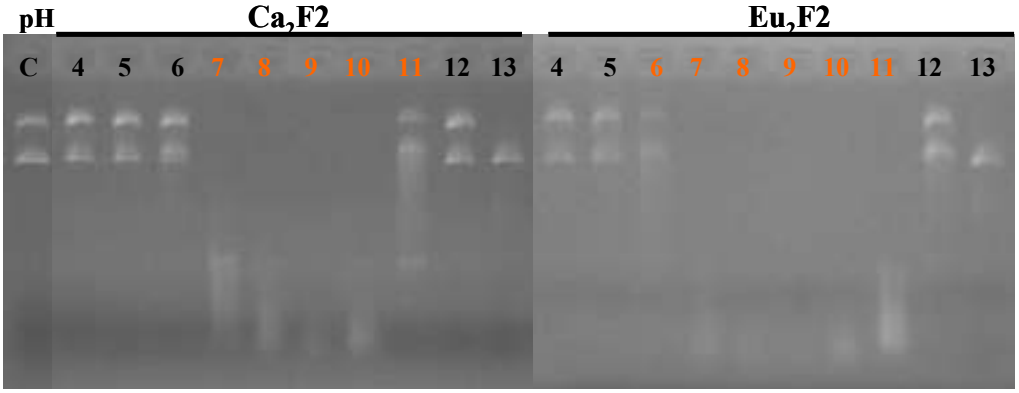


Figure 3.3 Agarose gel electrophoresis of supercoiled pTYB1 DNA treated with 1.50 μM metallated F2 that was incubated at different pHs (10 mM HEPES pH 7.0 and 50 mM NaCl). Lane C: Supercoiled DNA without protein. Lane 4-13 indicates the pH condition of DNA-protein complex. The lanes with cleavage activity were indicated in orange letters.

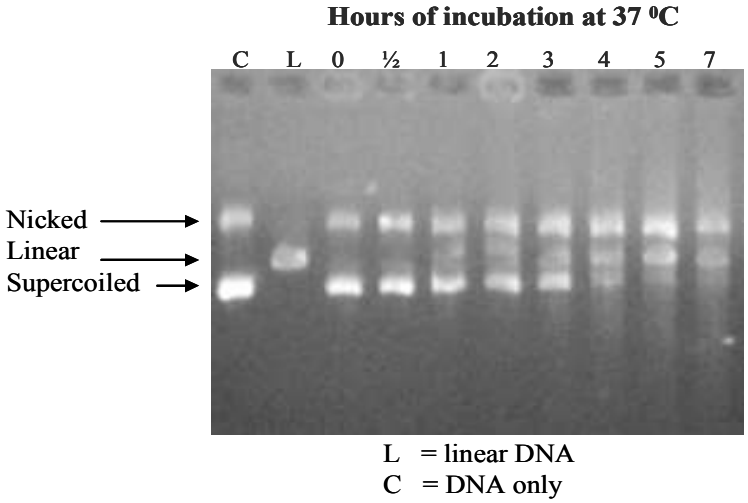


Figure 3.4 Agarose gel electrophoresis of supercoiled pTYB1 DNA treated with 643 nM Ca_2F2 that was incubated for different lengths of time (10 mM HEPES pH 7.0 and 50 mM NaCl). Lane C: Supercoiled DNA before treatment with the Geneclean kit. Lane L: Linear pTYB1 DNA digested with *EcoR*I. Lane 0-7 indicate the amount of incubation time in hours.

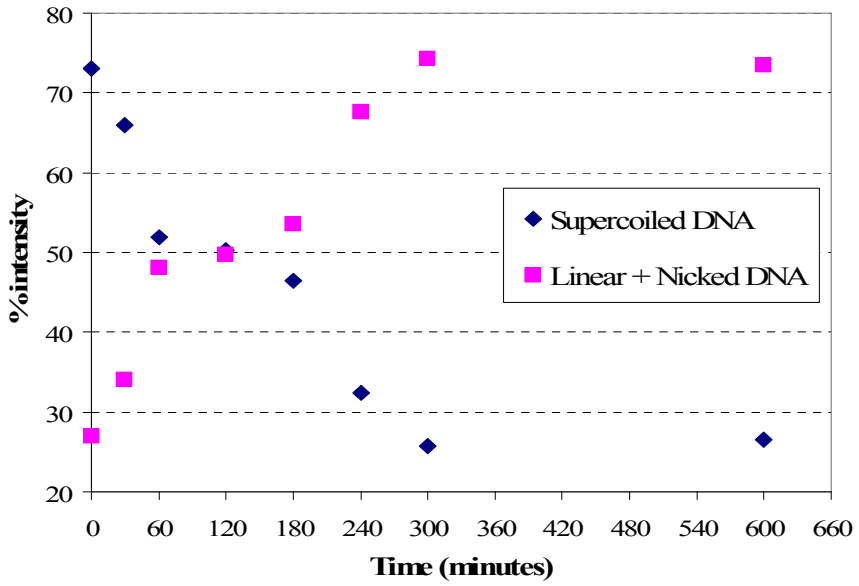


Figure 3.5 A graph plotting percentage of agarose gel band intensity from figure 3.3 as a function of time indicating the appearance of a DNA cleaved by Ca_2F2 over time.

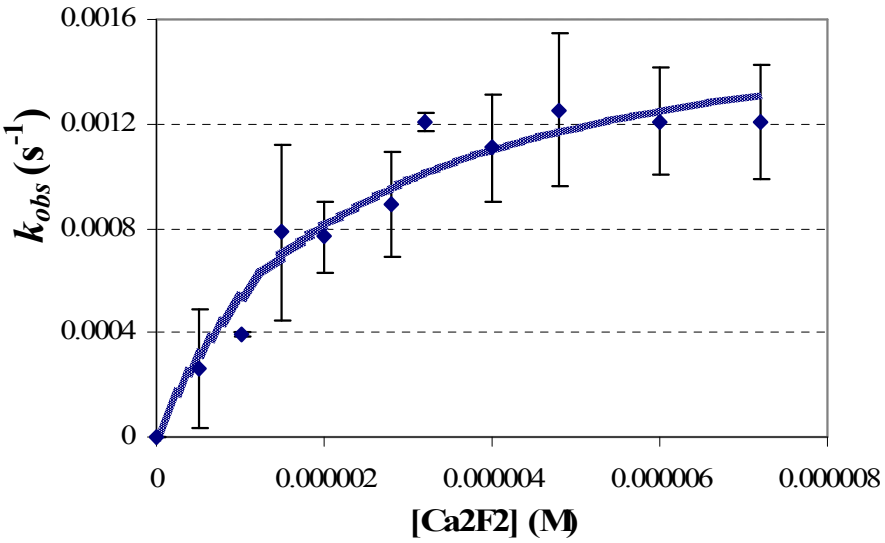


Figure 3.6 A plot of the pseudo-first-order rate (k_{obs}) as a function of Ca_2F2 concentration observed via agarose gel electrophoresis. The data was fitted with Michaelis-Menten kinetics equation (blue dotted line) to obtain V_{max}' and K_M. The error bar on the y-axis is the standard deviation over 3 trials.

Table 3.1 Kinetic rates of BNPP cleavage for metalloprotein/peptides.

Complex	K ($M^{-1}s^{-1}$)		Substrate
Eu^{3+}	2.1×10^{-2}		BNPP
EuP3A[a]	1.2×10^{-2}		BNPP
EuP3E[a]	4.7×10^{-3}		BNPP
EuP3W[a]	3.1×10^{-2}		BNPP
EuC2[b]	2.5×10^{-2}		NPP
Eu_2F2	1.4×10^{-1}		BNPP
Ca_2F2	8.2×10^{-2}		NPP
Complex	**K_M(M)**	**k_{cat}/K_M ($M^{-1}s^{-1}$)**	**Substrate**
Ca_2F2	2.3×10^{-6}	7.5×10^{2}	DNA Plasmid
Eu^{3+}[c]	3.9×10^{-5}	1.7×10^{0}	DNA Plasmid
BamHI[d]	2.6×10^{-6}	2.6×10^{5}	Oligonucleotide
Mg-EcoRV[e]	5.5×10^{-7}	1.4×10^{6}	Oligonucleotide

Source: [a]Data is collected by Kinesha Harris. [b]Data is collected by Sunghyuk Lim. [c]Data is adapted from reference 66. [d]Data is adapted from reference 117. [e]Data is adapted from reference 118.

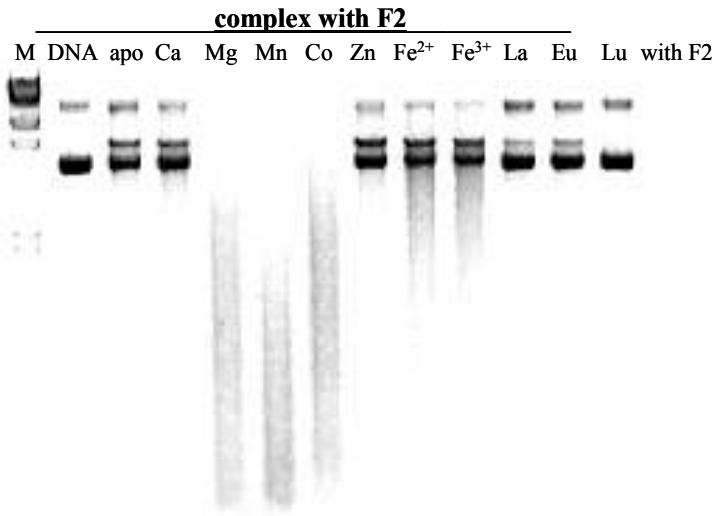


Figure 3.7 Agarose gel electrophoresis of supercoiled pTYB1 DNA treated with 1.50 μM metallated F2 incubated at 37 °C. Lane M is the λ Hind III DNA fragment marker and lane DNA is the supercoiled DNA without protein.

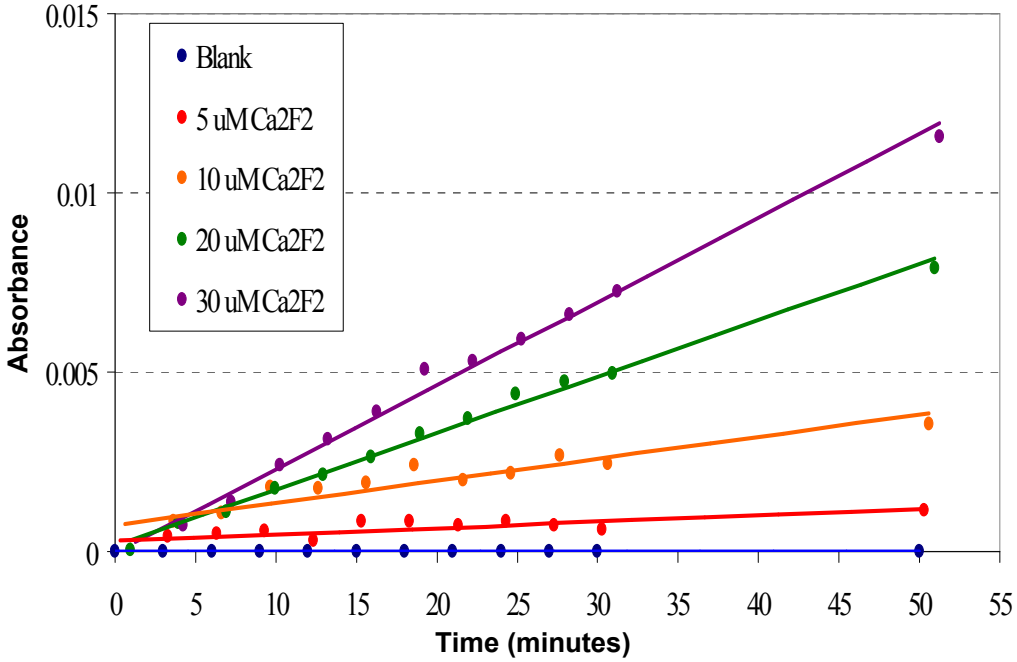


Figure 3.8 A time course experiment of BNPP cleavage at various concentrations of metalloprotein. Absorbance is monitored at 400 nm coinciding with the product of cleavage activity, nitrophenol.

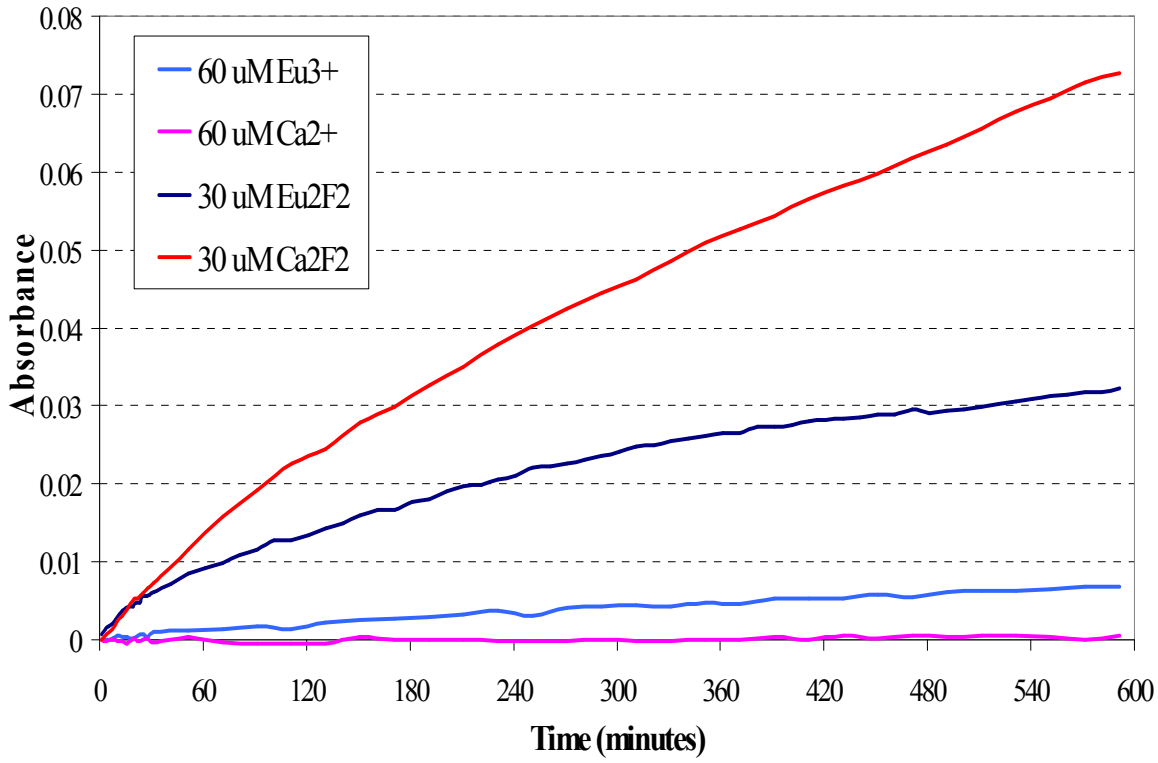


Figure 3.9 The hydrolytic activities of BNPP with metalloproteins and metal ions were compared.

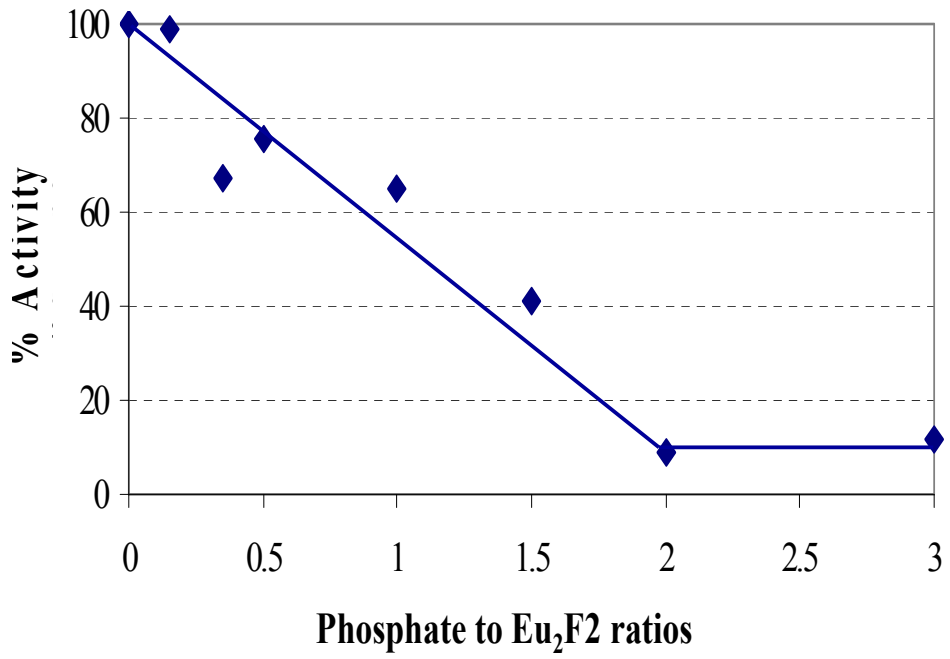


Figure 3.10 The % hydrolytic activity of Eu_2F2 was inhibited by the addition to phosphate in the BNPP solution.

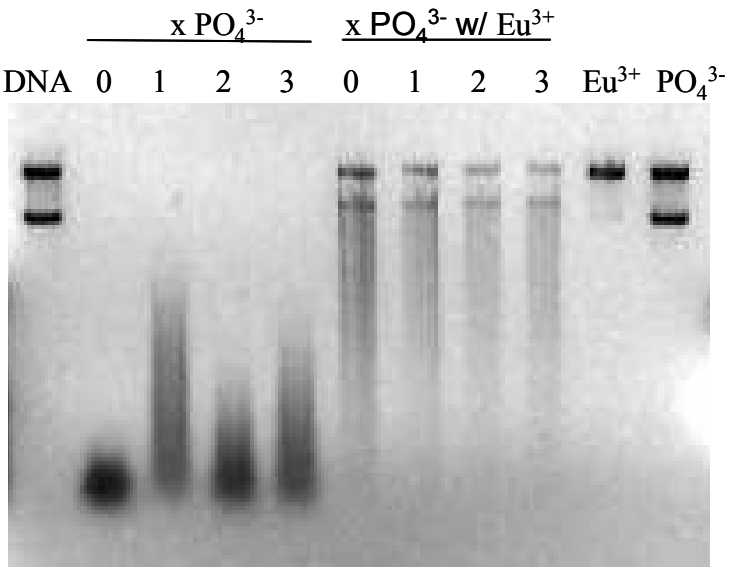


Figure 3.11 Agarose gel electrophoresis of pTYB1 plasmid with Eu_2F2 to determine the level of inhibition of cleavage activity in the presence of phosphate or europium.

CHAPTER 4

F2-DNA INTERACTIONS AND THE EXTENT OF SEQUENCE SELECTIVITY

4.1 Introduction

Transcription factors play important roles in regulation of gene expression.[39,123] The regulation controlled by DNA-binding transcription factors are highly complex and occurs in nearly every metabolic pathway. Among the DNA-binding proteins, homeodomains are our focus of interest because it is an extremely simple model system for understanding protein-DNA interactions. The family consists of proteins with a domain of 60 amino acids in length. This domain has much of the DNA-binding specificity and affinity of the larger proteins.[124] Several DNA-homeodomain complexes have been well-characterized via crystallography and NMR.[42-44] The family of homeodomains is highly conservative in structure even though their targeted DNA sequence recognition sequences were considerably diverse, suggesting that homeodomains are the ideal template for understanding DNA recognition.[49,54,125,126] In addition to its appealing DNA recognition and structural properties, it has the ability to internalize into cells[57,58] which can be a conventional application for drug delivery.

Engrailed homedomain is the prototypical and widely studies homeodomain. It consists of 3 α-helices and it recognizes a consensus sequence of 5′-TAATTA-3′. The third helix makes sequence-specific contacts in the major groove while the flexible N-terminal arm makes contact in the adjacent minor groove. The base-specific contacts are made primarily by residues I47, Q50, and N51 in the major groove. The R3 and R5 of the N-terminal arm are essential for the base contacts within the TAAT core.[50,51] Studies have shown that omitting the N-terminal arm reduced the specificity and affinity of *engrailed* homeodomain towards DNA.[53] *De novo* protein design with the conservation

of the N-terminal arm and helix 3 is essential for maintaining these protein-DNA interactions.

We have shown that designed chimeras obtain parental properties by utilizing those advantages of the DNA-binding domain of *engrailed* homeodomain and the substitution of a robust Ca-binding loop at the α-α turn.[68,127] Since a long, flexible polypeptide linker provides large DNA association freedom for the two subunits and greatly increases the local concentration of subunits for dimeric interactions, a structure-based chimeric homodimer was designed with subunit connection by a linker to potentially improve DNA selectivity and induce dsDNA cleavage.

The DNA selectivity and kinetic activity toward specific sequences were investigated in this chapter. The consensus binding site sequences of monomeric and dimeric chimeras were determined and compared. The consensus sequence for the monomer is much more apparent than that of the dimer. However, the hydrolytic activity with DNA plasmid was hardly measurable in the metallomonomer while it was observed with metallodimer within one minute of DNA association. This suggested that strong sequence binding affinity was not parallel with the accessibility of active sites to the DNA backbone. The ^{32}P-labeled experiments have shown that the DNA cleavage activity was carried through hydrolytic mechanism on both strands with nucleotide excision from the sequence in the process. This homodimer is the first example of an active and selective hydrolytic artificial dimeric nuclease based on the modular turn substitution design approach.

4.2 Experimental Procedures

4.2.1 Electrophoresis Mobility Shift Assay (EMSA)

4.2.1.1 PCR-coupled EMSA

The 5′-end of two primers (Figure 4.1) were radiolabeled with [α-^{32}P]dATP and T4 polynucleotide kinase for an hour at 37 ˚C and the reaction quenched by adding 20 mM EDTA to a final concentration of 10 mM. Unincorporated nucleotides were removed by micro Bio-Spin 6 chromatography column (Bio-Rad), which has a 5-base pair or 6000 dalton exclusion limit. The randomized library (N_8 or N_{11}-oligonucleotide) was then amplified and radiolabeled by PCR. The 100 µL reaction mixture contained 1X *pfu* enzyme buffer (100 mM KCl, 100 mM $(NH_4)_2SO_4$, 200 mM Tris-HCl, pH = 8.75, 20 mM $MgSO_4$, 1% Triton X-100, and 1 mg/mL BSA), 10 mM *dNTP*s (Invitrogene), 400 ng/µL of radiolabeled primers (primer 1 and 2 for C2 and 3 and 4 for F2), 66 nM of the annealed oligonucleotide, and 2 µL (1U) of *pfu* DNA polymerase (Stratagene). The mixture was amplified with 25 cycles of 95 °C for 30 s, 55 °C for 60 s, and 65 °C for 30 s (Mastercycler PCR, Eppendorf). A final extension of 72 °C for 10 minutes was included in generating the initial labeled libraries prior to EMSA. The PCR products were purified by extracting with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with cold ethanol. Ethanol was removed from the DNA pellet by vacuum centrifugation (Labconco).

4.2.1.2 EMSA for Binding Site Selection

The ^{32}P-radiolabeled DNA containing the randomized N_{11} core was incubated with increasing concentrations of chimeric protein in 25 µL binding buffer (10 mM HEPES pH 7.0, 50 mM NaCl, 5% glycerol, 150 mM NaCl, 4 mM Ca^{2+}, 0.5 mg BSA and 0.5 mg poly(dI-dC)). The mixtures with F2 were incubated for 5 minutes at room temperature. A 6% 1X TB (Tris-Borate) polyacrylamide gel was prepared for each

assay, and pre-run at 300 V for 30 minutes prior to loading to eliminate interference from residual ammonium persulfate (APS) and tetramethylethylenediamine (TEMED). An aliquot of 12.5 μL from each reaction mixture was quickly loaded onto the wells and the gel was run at 4 ˚C and 155 V for 1.5 hours. Tracking dye was not added to the mixture but was run in an adjacent lane to track gel migration during the electrophoresis. After electrophoresis, the gel was exposed to a Molecular Dynamics PhosphorImager Screen (Amersham Biosciences) for either an hour at room temperature or overnight at -20 ˚C, scanned and digitized (50 μm resolution; Storm PhosphorImager), then analyzed with Molecular Dynamics ImageQuant software. In each round of selection, a "bound band" section was excised from the gel at the lowest concentration of protein producing an observable gel shift. A section was chosen from just above the "unbound" band, as the "bound" band was somewhat diffuse. This is presumably due to rapid protein-DNA dissociation kinetics during electrophoresis. The selected DNA was extracted by soaking the crushed gel slice in DNA elution buffer (100 mM NaCl, 10 mM Tris-HCl pH = 6.91, 2 mM EDTA) for 2 hours at 37 ˚C, and gel materials removed by Micro Bio-Spin Chromatography (Bio-Rad). The selected DNA was further purified from residual protein by phenol extraction (phenol:chloroform:isoamyl alcohol (25:24:1)) and ethanol precipitation. The purified DNA was again amplified by PCR using radiolabeled primers as described above. Residual polymerase was removed from the amplified DNA by phenol extraction and ethanol precipitation, and then used directly for the next round of binding site selection.

After three rounds of selection and amplification, the pool of DNA enriched in higher affinity sites was amplified a final time with non-labeled primers. This final pool of DNA was digested with restriction enzymes *BamHI* and *EcoRI* and subcloned into pUC19 plasmid. The plasmid was transformed in DH5α cells, plated, and DNA isolated from individual colonies was sequenced by the University of Iowa DNA facility.

4.2.1.3 CNBr Sepharose DNA Selection

A weight of 0.25 g of dry cyanogen-bromide activated sepharose 4 FF resin (Amersham Biosciences) was soak in deionized water for one hour. The coupling solution consists of 100 mM HEPES pH 7.0, 500 mM NaCl, and 500 mM $CaCl_2$ and cold wash solution consists of 1 mM HCl were prepared. The wet resin was packed into a micro bio-spin column (Bio-rad). A ratio of 0.5:1 of coupling solution to medium (CaC2 or Ca_2F2) volume was mixed into the resin. The resin was prepared and activated for coupling by following the product manual. The coupling reaction was preformed at 4 °C overnight in order for the exposed amine residues, lysine (K) and arginine (R), of F2 to attach covalently to the resin. The absorbance at 280 nm was monitored prior and after the coupling reaction to check the efficiency of the coupling. Before introducing 25 uL of purified ^{32}P-labeled DNA duplex with random N_8/N_{11}-nucleotide library (Figure 4.1A & B), the protein-bound resin was equilibrated with 50 mM HEPES pH 7.0, 100 mM NaCl, 50 mM Ca_2Cl, 10 μg poly(dI-dC), and 20 μg tRNA. The labeled DNA was incubated with the protein-bound resin for 2 minutes. The unbound nucleotides were then washed out and the bound nucleotides were then eluted from the column. The bound nucleotides were purified via phenol extraction (phenol:chloroform:isoamyl alcohol (25:24:1)) and ethanol precipitation. The purified DNA was amplified by PCR using radiolabeled primers as decribed above. Residual polymerase was removed from the amplified DNA by phenol extraction and ethanol precipitation, and then used directly for the next round of binding site selection. After three rounds of selection and amplification, the pool of DNA enriched in higher affinity sites was amplified a final time with non-labeled primers. This final pool of DNA was digested with restriction enzymes *BamHI* and *EcoRI* and subcloned into pUC19 plasmid. The plasmid was transformed in DH5α cells, plated, and DNA isolated from individual colonies was sequenced by the University of Iowa DNA facility.

4.2.2 ^{32}P-labeled Sequencing Gels

4.2.2.1 Sequencing DNA by Maxam-Gilbert Technique

The oligonucleotides were synthesized and desalted by IDT (Coralville, IA) (Table 4.1). The 5′-end of 50 μM designed oligonucleotide was radiolabeled with [α-^{32}P]dATP and T4 polynucleotide kinase for an hour at 37 ˚C and the reaction quenched by adding 20 mM EDTA to a final concentration of 10 mM. Unincorporated nucleotides were removed by micro Bio-Spin 6 chromatography column (Bio-Rad), which has a 5-base pair or 6000 dalton exclusion limit. A standard protocol for sequencing single strand ^{32}P-radiolabeled DNA by the Maxam-Gilbert technique was followed.[128]

The sequences for the studies are showed in Table 4.1. The sequences are separated into 6 groups (A-E) depending on the orientation of both potential cleavage sites, such as TAAT and the sequences designed for group E consist of no TAAT site which are used the control for non-TAAT site cleavage(Table 4.1). Group A to D contain two TAATNN sites that are orientated at different directions and there are different gap distances between two potential cleavage sites within each group.

4.2.2.2 Mechanism Comparison via DNase I Footprinting

The 5′-end of 50 μM designed oligonucleotide was radiolabeled with [α-^{32}P]dATP and T4 polynucleotide kinase for an hour at 37 ˚C and the reaction was quenched by adding 20 mM EDTA to a final concentration of 10 mM. Unincorporated nucleotides were removed by a micro Bio-Spin 6 chromatography column (Bio-Rad), which has a 5-base pair or 6000 dalton exclusion limit. The 5′-end labeled oligonucleotide was annealed with the non-labeled complimentary strand at 95 °C for 5 minutes in a buffer of 10 mM HEPES pH 7.0 and 50 mM NaCl. The heated DNA strands were annealed by cooling down to room temperature slowly. A standard protocol for sequencing single strand ^{32}P-radiolabeled DNA via the DNase I footprinting technique was performed.[129] Since the length of oligonucleotide is less than 20 base pairs, the

optimal footprinting condition for the DNA duplex was 12 μl of 10 μg/ml cold DNase I incubated at room temperature for 30 s. The DNase I reaction was quenched with 20 mM EDTA and the DNA duplex was further purified from DNase I by phenol extraction (phenol:chloroform:isoamyl alcohol (25:24:1)) and ethanol precipitation before loading onto a 20% urea-polyacrylamide gel.

4.2.2.3 Cleavage assay with ^{32}P-labeled DNA duplexes

The 5′-end labeled oligonucleotide was annealed with the non-labeled complimentary strand at 95 °C for 5 minutes in the buffer of 10 mM HEPES pH 7.0 and 50 mM NaCl. The heated strands were allowed to anneal by cooling down to room temperature slowly. Two equivalent of Eu^{3+} or Ca^{2+} were added to apoF2 as the chloride salt and the mixture was equilibrated for 15 minutes at room temperature. A concentration of 8 μM metalloprotein was mixed with 8 nmole of DNA duplexes and incubated at 37 °C overnight. The DNA was further purified from residual protein and salt by phenol extraction (phenol:chloroform:isoamyl alcohol (25:24:1)) and ethanol precipitation. A 20% urea-polyacrylamide gel was prepared for each assay, and pre-run at 60 watts for 30 minutes prior to loading in order to eliminate interference from residual APS and TEMED. An aliquot of 12.5 μL with 1 x loading buffer from each reaction mixture was quickly loaded into the wells. The gel was run at 300 watts for the first 5 minutes or until the temperature reached 50 °C in order to prevent re-annealing and hairpin structure formation. The wattage was then adjusted to 60 watts and run for 1.5 hours. After electrophoresis, the gel was exposed to a Molecular Dynamics PhosphorImager Screen (Amersham Biosciences) for either an hour at room temperature or overnight at -20 ˚C, scanned and digitized (50 μm resolution; Storm PhosphorImager, then analyzed with Molecular Dynamics ImageQuant software).

4.2.3 Circular Dichroism of F2 in the Presence of DNA Duplexes

CD spectra of ApoF2 and Eu_2F2 were recorded in the presence or absence of DNA duplexes on an Olis Cary-17 DS Conversion spectrophotometer at 20 ˚C under N_2 atmosphere in a 0.1-cm cell. Spectra were collected from 320 to 200 nm at 0.5 nm resolution and smoothed with Olis software (21 boxcar smoothing function). Samples of 15 μM apoF2 containing 10 mM HEPES pH 7.0 and 50 mM NaCl added to equivalent Eu^{3+} as the chloride salt to form Eu_2F2 and the mixture was equilibrated at 20˚C for 15 minutes. Stock solutions of duplex DNA were prepared in 10 mM HEPES pH 7.0 and 50 mM NaCl and annealed by heating to 95 ˚C for 5 minutes followed by cooling to room temperature slowly. Three DNA duplex were tested: Non-specific contains no TAATTA site: 5′-GAGCGGAGGCGCAGCAGCTTAGATAAGG-3′, cs 1 with one TAATTA site (underlined): 5′-GACGTAATTACCCCCAAGCTG-3′, and Wong 3Rev with two TAATTA sites (underlined): 5′-GACGTAATTAGTAATTAGC-3′. The spectrum of 1:1 duplex-protein mixture was subtracted with that of the DNA duplex sample to obtain the protein spectrum in the presence of DNA. The molar ellipticity of the three subtracted protein spectrum was plotted against the wavelength.

4.2.4 ^{32}P-labeled Kinetics Assay

In order to determine the hydrolytic rate of the metallo-F2 at selected cleavage sites and determine if there is any sequence-dependent cleavage activity, we looked at cleavage under single-site cleavage conditions. Four DNA sequences, Wong 3Rev, Wong 4Rev, and Wong5Rev (Group D), and cs 1, were used for this experiment since group D sequences gave the most efficient cleavage model for metallo-F2 (Figures 4.7 and 4.8). Sequence cs 1 consists of only one TAATTA site which can be used for site specific activity comparison. One strand of each oligonucleotide was radiolabeled at 5′-end with [α-^{32}P]dATP and T4 polynucleotide kinase for an hour at 37 ˚C and the reaction

quenched by adding 20 mM EDTA to a final concentration of 10 mM. Unincorporated nucleotides were removed by a micro Bio-Spin 6 chromatography column (Bio-Rad), which has a 5-base pair or 6000 dalton exclusion limit. The 5′-end labeled oligonucleotide was mixed with the non-labeled complimentary strand at 95 °C for 5 minutes in a buffer of 10 mM HEPES pH 7.0 and 50 mM NaCl, then annealed by cooling down to room temperature slowly. The annealed 0.8 nmole DNA duplex was mixed and incubated with a constant concentration of Eu_2F2 at various incubation time intervals from 0 to 60 minutes. The reaction was quenched with 20 mM EDTA at each time interval and DNA was isolated by phenol extraction and ethanol precipitation prior loading to a 20% urea-acrylamide gel. The gel was run and developed as describe in section 4.2.2.3. The percentage of band intensity relative to the entire lane was analyzed and calculated for each lane using Molecular Dynamics ImageQuant software. The percentage of cleaved DNA was plotted against time to obtain the pseudo first order rate, k_{obs}, at 0.6 μM Eu_2F2 averaged from two trials. The second order rate was measured by calculating k_{obs} at various Eu_2F2 concentrations.

4.2.5 ^{32}P-labeled at 3′- and 5′-end Site Matching Assay

Since sequence Wong 3Rev was the most efficient cleavage model for metallo-F2, it was used for this assay to determine its cleavage pattern or preference. By labeling the sequence at both termini and observing the cleavage pattern from both directions, the cleavage site can be matched and its cleavage pattern can therefore be understood. The 5′-end labeling is described in section 4.2.2.2 and the 3′-end labeling is performed using a standard protocol.[130] The markers for the sequence were obtained via DNase I footprinting (section 4.2.2.2) and Maxam-Gilbert (section 4.2.2.1) techniques. A concentration of 2 μM Eu_2F2 and Ca_2F2 was mixed and incubated with labeled duplex at 37 °C for 2 hours. The reaction was quenched with 20 mM EDTA and the residual metalloprotein was removed from the cleaved DNA by phenol extraction and ethanol

precipitation prior loading to a 20% urea-acrylamide gel. The gel was run and developed as describe in section 4.2.2.3. The cleavage site was matched according to the sequenced markers.

4.3 Results and Discussion

4.3.1 Electrophoresis Mobility Shift Assay of C2 and F2

4.3.1.1 EMSA for Binding Site Selection with F2

EMSA is a useful technique to collect a consensus sequence or family of DNA binding site(s) of our chimeric metalloprotein. A library of randomized sequences was incorporated as a cassette into a larger oligonucleotide (Figure 4.1B). With the TAATTA site included in the designed randomized sequences, it may provide an anchor for dimer-DNA recognition, we determined that an EMSA experiment without TAATTA site had too many variants to consider, such as base pair gap between two sites and the length for consensus sequence determination (data not shown). The randomized library was amplified and labeled by PCR using ^{32}P-radiolabeled primers complementary to the sequences to either side of the N_{11}-site. The labeled library was then subjected to an EMSA with increasing concentrations of Ca_2F2 (Figure 4.2). The shifted band was then excised from the gel, and then the resulting DNA enriched in preferred binding sites was again amplified by PCR. This method is an adaptation of published methods.[87]

In the initial EMSA screen, a gel shift was observed at 0.4 μM Ca_2F2. After 3 rounds of selection, the target DNA was enriched in higher affinity sites. After three selections, the midpoint of the gel shift had decreased to 0.24 μM Ca_2F2, corresponding to about 2-fold enhancement in binding affinity. Unlike the smearing observed in the protein-DNA complex band with our previously designed peptides (described in chapter 5), there was a distinct shift band with F2-DNA complex suggesting the protein-DNA interaction is much stronger or less transient as a full domain rather than a “half-domain”

peptides. This observation was also supported by the DNA binding assay in chapter 2 where F2 had DNA binding affinity at the low micromolar range.

After three rounds of EMSA selection, the DNA library was amplified with unlabeled primers and subcloned into the multicloning site of pUC19 by the ligation of the restriction enzyme sites on the oligonucleotides. The plasmid was transformed into DH5α cells to grow individual colonies for sequencing. The resulting sequences were aligned manually to identify consensus binding targets, as shown in Figure 4.3. Initial alignments took into consideration the most commonly occurring 3-mer and 4-mer sequences. Both forward and reverse directions (relative to primer) were considered, but each occurrence was included in the alignment only once. The consensus sequence for F2 was thus summarized as 5′-T(C/G)(C/G)(C/G)(C/G)-3′ with a range of 71-90% nucleotide identity among 51 selected in two trials of DNA sequences. There was a high population (~75%) of three consecutive sites of G/C-rich nucleotides located next to the TAATTA site indicating that the binding site for one of the F2 domain was adjacent to the TAATTA site.

4.3.1.2 CNBr Sepharose DNA Selection with F2 & C2

CNBr-activitated sepharose DNA selection was also performed to shorten the protein-DNA interaction time to 2 minutes as Ca_2F2 has been shown to be hydrolytically active in the presence of DNA duplexes. The same selection was also performed with CaC2 to determine the differences in DNA recognition site between a monomer and a dimer.

After three round of DNA selection, the DNA libraries of CaC2 and Ca_2F2 were collected and sequenced. The apparent binding site for C2 is distinct with a consensus sequence of 5′-GGC-3′ with over 88% nucleotide identity among the selected 31 DNA sequence library from two trials (Figure 4.4). Even though the consensus sequence of CaC2 was very different from the parental one (5′-TAATTA-3′), it is quite similar to one

of the consensus sequences obtained from a "half-domain" peptide, CaP3W, which is 5′-(A/T)GGG(C/T)-3′ with a core of 5′-GGG-3′.[87]

Due to the weaker binding affinity and the DNA reactivity of Ca_2F2, its binding site apparently was not as distinct as that of CaC2. The consensus sequence of Ca_2F2 was aligned as 5′-(A/T)CC(C/G)(C/T)-3′ with the range of 71-92% nucleotide identity among the selected 24-sequence library from two trials (Figure 4.5). When comparing the consensus sequences of Ca_2F2 with both selection methods, both are resulted in G/C rich 3- or 4-mer sequences implying that both methods were valid for CaC2 and Ca_2F2 consensus sequence determination. In addition, the binding selectivity studies on our designed peptides have shown that their recognition sites were not similar to that of the wild-type[87] suggesting that it was not surprising for both metalloproteins to have a high frequency of G/C rich 3- or 4-mer in their DNA libraries.

The length of the target site was not known *a priori*, since the binding behavior of these chimeric constructs had not been previously investigated. It was expected that the targets would be no longer than that of the parental *engrailed* or *antennapedia* homeodomains (6 b.p.) and likely somewhat shorter (4-5 b.p), since the apparent DNA binding affinity of the chimeras are at least 3-5 magnitudes weaker than the wild-type *engrailed* homeodomain.[49]

Here we have created chimeras that had parental DNA binding ability, yet gave different sequence preferences. The binding affinity and the EMSA results of C2 and F2 were consistent in which C2 has 2-fold stronger binding affinity with DNA plasmid and thus showed stronger evidence on obtaining a consensus sequence. However, as we will see in the next section, the binding targets resulted in this assay may not represent the targets F2 that cleave most selectively.

4.3.2 ^{32}P-labeled Sequencing Gels

4.3.2.1 Cleavage Assay with ^{32}P labeled DNA Duplexes

Our previous studied chimeric peptides, P3W and P5b, have been shown to target a distinct family of consensus sequences[87] and P3W had modest DNA sequence selectivity[81] even though the peptides were essentially only a "half-domain" chimera with the absence of the first helix. With the results of DNA hydrolytic activity and the consensus sequence of F2 which has two full domains, we can determine its sequence preference via ^{32}P-labeled sequencing gel assay with three different sequences (Figure 4.6). The sequences are described as shown below:

Non specific (N.S.): 5′-CCAGAACAAGCGGTGCAAGATCAAGAAGT-3′

Wong8: 5′-GAGGTGGTCATGG-3′

Wong 5TA: 5′-GACGTAATGTAGCATTATAGC-3′

N.S. is a GC- and AA-rich sequence in the absence of a homeodomain recognition site, TAAT. Wong8 has one TCAT site (underlined) which is one site different from TAAT while Wong5TA is designed with two parlindromic TAAT sites (boxed) for metallo-F2. Both sequencing gels of metallo-F2 with Wong4 and Wong5TA resulted in the appearance of bands suggesting a selective cleavage occurred at certain nucleotide sites. However, the intensity of those cleavage bands were several magnitudes weaker with Wong4 compared to that of Wong5TA even with a time-course sequencing gel assay (data not shown) indicating that sequence Wong5TA is more cleavage preference by F2. However, F2 did not have sequence preference with Wong8 which has a similar target site showing that there is some degree of selectivity against non-target sequences, and very little cleavage was observed within the N.S. sequence.

Since sequence containing target sites of homeodomain gave the greatest number and most intense cleavage bands, we designed five groups of sequences with different orientations of both target sites (TAAT or ATTA) and different lengths of gap between

the two sites within each group (Table 4.1) in order to understand the cleavage patterns and preference of metallo-F2. Did the orientations of two target sites affect the cleavage ability of F2? Since F2 is a single-chain dimer linked by a polypeptide, did the gap between two target sites influence the cleavage pattern or selectivity of F2? Those are the questions we could determine from the five-group sequencing gel assay (Figure 4.7). The arrows represented a visible cleavage band on a certain nucleotide site. The height of the arrow represented the intensity of the band. Both strands of the sequence were 5′-end labeled and run to determine cleavage sites as a duplex. A three-dimensional view of the cleavage patterns of group A to D was summarized in Figure 4.8.

The cleavage patterns of the sequences in group A were very similar except for Wong 4 with the 4 base pairs gap. It had the most cleavage bands in a distinct cleavage pattern, presumably because its first site was in fact TAATTA instead of TAAT. As a result, group B was designed with TAATTA at the first site with the same base pair gap as group A to understand whether or not the addition of TA played a factor to the different cleavage pattern in Wong 4. The cleavage pattern of Wong 3AT and 5AT were very similar to that of Wong 3 and 5 suggesting the addition of TA at the first site did not play a major role on the unique cleavage pattern. However, the migration of cleavage band down the duplex as a function of the base pair gap was observed in group A and B. A cleavage band before ATTA site remained persistent even with different gap length, but no nucleotide preference, G or C, was observed on that band. This observation suggested that there was cleavage selectivity one nucleotide before the second target site and was independent of the gap lengths between two target sites. The design of group C was thus focused on the second target site. The second site was changed from ATTA<u>TA</u> to ATTA<u>AT</u>, a complimentary sequence of *engrailed* homeodomain recognition site, TAATTA. This was designed to allow us to understand the cleavage influence of the second site. Not only did the cleavage pattern changed dramatically with increased number of cleavage bands on the complimentary strand in group C, but cleavage bands

on the second site that were observed in group A and B had also disappeared. When the direction of the second site has changed into TAATTA in group D, some cleavage bands on the second site were regained yet the effect of the gap was different. Instead of a consistence cleavage before the ATTA site independent of the gap length, cleavage remained on the same nucleotide within the second site in group D. Moreover, the cleavage on T nucleotide on the 5th of the 6 base pair *engrailed* target at the first site was steady thorough all groups. This further verified that F2 has sequence selectivity on the T nucleotide within the *engrailed* site. The cleavage pattern was not gap-dependent yet was greatly influenced by the pattern on the second site. The lesser amount of sequence discrimination behavior on the first site could be related its function as an anchor for DNA recognition, the same role for all sequences tested. This property was further confirmed with a time-course sequence gel assay in this chapter (Figure 4.9). A cleavage band first appeared on T nucleotide at the first site as bands started to appear on the second site after 10 minutes of incubation.

In group E, only one TAATTA site was designed followed by two potential F2 consensus binding sequences, GGGGGAA and CCCCCAA, as the consensus sequence of F2 have shown to be G/C rich by EMSA assay. The frequency of G and C nucleotides were over 75% following the TAATTA site in the EMSA assay. So, this set tested the relationship between the binding and cleavage targets. Many of the cleavage bands disappeared in group E even though the group contained the potential consensus sequences. It is important to note that the consensus sequence for less hydrolytically active C2 is also G/C rich as 5′-GGC-3′ suggesting the strong DNA binding affinity is not an indicator for strong hydrolytic activity. These non-linear observations could be influenced by the orientation "face" of our chimeras. The active site is located at the beginning of the third helix, the DNA contact domain. Therefore, where the domain contacted with DNA did not always lead to DNA cleavage as the active site is outside of the DNA contact region. As a result, the DNA binding orientation with GGC and

TAATTA sequences could be different which led to the alteration of DNA backbone accessibility to the active sites. Nonetheless, homeodomains have shown to diffuse and hopping down the DNA duplex like beads along a string upon recognizing TAAT sites.[131] Therefore, it is not surprising that our chimeras might have the same behavior upon DNA recognition where F2 was orientated for hydrolytic cleavage in the association of TAATTA sites.

4.3.2.2 Hydrolytic versus Oxidative Cleavage Mechanism

The catalysis of DNA strands can be carried out by various mechanisms. Two of the major DNA cleavage mechanisms in biological systems are the hydrolytic reaction that involves breaking a phosphodiester bond with the introduction of water and the oxidative reaction that free radicals reduced an aromatic ring of DNA by extracting a H˙. During the oxidation reaction, the sugar ring is reduced by free radicals resulting in hydroxyl ends on both termini. On the other hand, the hydrolysis of phosphodiester bond by endonuclease produces 5′-phosphate and 3′-hydroxyl end products on a DNA duplex. In order to determine the cleavage mechanism of metallo-F2, a DNA sequence is treated with chemical or enzymatic reaction to give end products reflecting a different mechanism and observable by different gel migrations for a given nucleotide (Figure 4.10A). The DNA duplex that was subjected to Maxam-Gilbert sequencing technique conditions produced 5′- and 3′-hydroxyl termini, while DNase I treatment of DNA gave 5′-phosphate and 3′-hydroxyl termini nucleotides (Figure 4.10B). Since both techniques produced different 5′-end products (hydroxyl and phosphate), we can thus determine the cleavage mechanism by treating 5′-end labeled DNA, Wong 3Rev, with both techniques and comparing these migration standards to F2. The cleavage bands on both DNA strands in the presence of metallo-F2 matched the DNA end products from the DNase I treatment instead of the Maxam-Gilbert technique indicating that F2 cleaved DNA through hydrolysis to give 3′-hydroxyl and 5′-phosphate termini. This mechanism is in

agreement with that of endonuclease and further confirmed the behavior of this chimera followed the artificial endonuclease design.

4.3.3 Circular Dichroism of F2 in the Presence of DNA Duplexes

To observe the influence of F2 on the structure of the dimer (and vice versa), the change in ellipticity of spectra was followed by CD spectroscopy. The molar ellipticity of apoF2 was observed in the presence and absence of three DNA duplexes. Sequence Non-specific (N.S.) contained no parental *engrailed* target sequence 5′-TAATTA-3′. Sequence Cs 1 contained one F2 target binding and one parental *engrailed* target sequences and sequence Wong 3Rev had two parental *engrailed* target sequences. As discussed previously, only Wong 3Rev had significant specific cleavage. The CD spectrum of annealed DNA oligonucleotides alone had CD spectra characteristic of stacked, B-form DNA (Figure 4.11).[132] The addition of apoF2 to an equimolar amount of each DNA duplex resulted in an increase in negative ellipticity below 260 nm. The difference CD spectrum of the DNA-bound F2 was obtained by subtracting the free DNA signal from that of F2-DNA (Figure 4.12). ApoF2 showed changes at 280 and 250 nm in the F2-Wong 3Rev complex indicative of DNA conformational change upon association. In the presence of duplexes, apoF2 appeared to maintain the most α-helicity with Wong 3Rev (2 parental sites), followed by cs 1 (1 parental site), and then N.S. (0 parental site) observed from the peak minima at 222 nm. However, the α-helicity of F2 was much smaller in the association of all three DNA duplexes suggesting that the secondary structure of F2 underwent a significant conformational change upon DNA binding which could be caused by stress of linker between two domains and the unwinding of termini upon binding to DNA. However, this may not be a very quantitative tool, as α-helicity of standard protein, myoglobin, was greatly underrepresented as well.

4.3.4 ^{32}P-labeled Kinetics Assay

The rate of cleavage activity on certain sites can be specifically determined via ^{32}P-labeled cleavage gel assay with selected sequences. The oligonucleotides of Wong 3Rev, Wong 4Rev, Wong 5Rev, and cs 1 were chosen for rate measurement and comparison. The sequences of Wong 3Rev, Wong 4Rev, and Wong 5Rev have two TAATTA sites while cs 1 has only one TAATTA site adjacent to a proposed consensus binding sequence of F2. As section 4.3.2.1 mentioned above, the rate of cleavage with a non-TAAT site sequence was several fold slower than with a TAAT sequence. The rate comparison among these sequences could allow us to understand the rate-dependent sequence discrimination due to recognition of these TAATTA site(s).

An example of a ^{32}P-labeled cleavage gel as a function of time with Wong 5Rev is illustrated in Figure 4.9. A band at T_1 started to appear after one minute of incubation. Three minutes later, a second band at T_2 started to be visible and its band intensity became as strong as at T_1 after 10 minutes of incubation and the second TAATTA site began to show cleavage bands at the 30-minute incubation time interval. As both sites' cleavage did not take place simultaneously, the recognition property at both TAATTA sites was distinctly different. It also suggests that T_1 site, a universal cleavage site for all sequence groups we investigated, could act as an anchor for site recognition. Apparently, the first site is more preferable for DNA cleavage by metallo-F2 while the second site was behaved as an "opportunist" site as more time was required for its recognition and cleavage. This property was also supported by the first order and second order rate for the overall sequences (Table 4.2). Sequence Wong 4Rev, Wong 5Rev, and cs 1 have very similar rates at K~ 40 $M^{-1}s^{-1}$ while Wong 3Rev was 3 times faster. This is likely because an extra cleavage bands was observed with Wong 3Rev, which could increase the overall cleavage percentage of the sequence and thus could be relevant to rate differences. When comparing the rate at T_1 nucleotide among these four sequences, the rates were relatively similar (data not shown) indicating that the selection on the second

TAATNN site was not rate-limiting for initial recognition and any cleavage. Nonetheless, there was still modest sequence cleavage selectivity between the consensus and TAATTA sequences at the second target site as much fewer cleavage bands were observed with sequences in the absence of second TAATNN site (cs1 and 2). These sequence selectivity results of F2 are consistent with that of the CD duplexes and EMSA experiments.

4.3.5 ^{32}P-labeled at 3′- and 5′-end Site Matching Assay

After determining the cleavage sites and pattern of metallo-F2, it is important to follow the cleavage pattern starting at different ends of a DNA sequence in order to further understand the mechanism contributing to this cleavage pattern. Like many restriction enzymes that are homodimers, we would like to ask whether or not it produced double stranded cleavage, and if so, whether that cleavage leaves a blunt or sticky end product. Sequence Wong 3Rev was utilized for this site matching experiment because it has the most intensive and distinct cleavage pattern. Each strand (both complimentary strands) of Wong 3Rev was labeled at 3′- and 5′-end with Klenow fragment exo$^-$ and T4 polynucleotide kinase respectively. The location of cleavage sites were followed by markers sequencing by the Maxam-Gilbert and the DNase I footprinting techniques. The gel results of 3′- and 5′-end labeling on both strands were summarized on Figure 4.13. The colored lines demonstrate the length of sequence from each terminus observed on the gel. The sequence lengths from both termini labelings were matched in same color of lines resulting four best matched cleavage sites on Wong 3Rev in the presence metallo-F2 (Figure 4.13A). It is interesting that not only did Ca_2F2 cleave dsDNA, but Ca_2F2 also could excise an entire nucleotide on a sequence. This result yield a very intriguing aspect in understanding the protein-DNA interaction of an engineered dimeric protein, because it seems to excise whole bases, leaving sticky ends. Thus, it is particularly

appealing for the purpose of gene therapy, and could be potentially useful for gene disruption or repair in cancerous cells.

A three-dimensional view of the cleavage site was described in Figure 4.13B. The cleavage sites are in close proximity to each other at the major grooves which is an optimal DNA binding region for a bulky α-helix. The cleavage at the second site is also only one nucleotide away from the first target site suggesting the two active sites of the dimer are in close proximity to each other which may assist the hydrolytic activity of DNA by two-metal ion mechanisms proposed for restriction endonuclease.[15,70,133] Nonetheless, this first image is a model constructed to demonstrate the cleavage sites of a dimeric chimera while the crystal structure of F2 has yet to be determined.

4.4 Conclusions

The F2-DNA interaction in the aspect of the chimera's specificity and affinity toward target sequences, as well as conformational changes in the presence of DNA were addressed in detail in this chapter. The EMSA technique was utilized to identify and compare the consensus binding sequences for F2 and C2, although they did not appear to be the consensus cleavage target sites. The structural changes of DNA sequences upon F2 binding were monitored via CD spectroscopy. The sequence cleavage specificity at two adjacent target sites was carefully investigated via ^{32}P-labeling polyacryamide gel assay. Finally, the cleavage mechanisms and patterns of F2 were determined with various radio-labeling techniques.

The degree of sequence specificity and catalytic activity of this dimeric chimera has further demonstrated that it is much more intriguing and active than our previously designed monomeric peptides/protein. The recognition binding site of the dimer consists of a T followed by G/C rich 5-mer which was at least two times longer than that of the monomer's binding site (5′-GGC-3′) with a binding affinity in the low micromolar range.

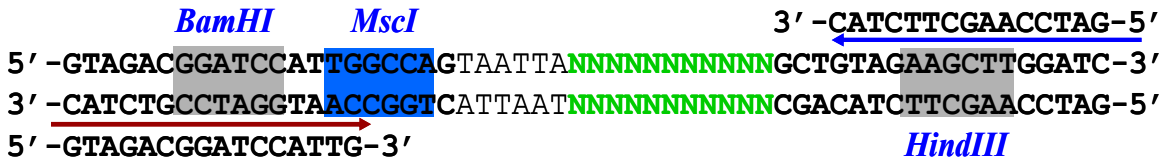
The ^{32}P-labeling sequence experiment with different groups of sequence designs and duplex studies via CD were performed in order to further understand the cleavage patterns and sequence preference in different sequences. It is clear that the structures of DNA and apoF2 responded distinctly with different sequences. A sequence containing two TAATTA sites resulting in DNA conformational change and a better secondary structure of F2 over sequences with one or none TAATTA site.

F2 cleavage via ^{32}P-labeled experiment showed that cleavage at the first target site was universally favorable and constant among sequences while the cleavage pattern was influenced by the alteration on the second site. This suggests that there are two recognition mechanisms where the first site acts as an anchor for aiding the recognition of the second site. The ^{32}P-labeled kinetic studies further verified the site recognition behavior of F2 that was based on rate measurements of the sequences.

By determining the way DNA was cleaved by metallo-F2, we can further understand the cleavage mechanism and pattern of this artificial endonuclease. The cleavage produces were determined via various ^{32}P-labeling techniques as the band of a nucleotide migrated differently depending on the end products yielded from various sequencing techniques. With the band comparison from DNase I footprinting (hydrolytic cleavage) and Maxam-Gilbert sequencing (leaving basic phosphate ends) sequencing techniques, F2 was shown to cleave DNA backbone through hydrolysis leaving complementary hydroxyl and phosphate ends which was parallel to the nucleolytic property of most endonuclease. The experiment of the 3′- and 5′-end labeled site matching assay showed that not only did the metallo-dimer cleave dsDNA, but it also excised an entire nucleotide at certain sites demonstrating the benefit of dimeric design. With the help of two active sites in close proximity, the catalytic activity of the nuclease was increased dramatically and the excision of an entire nucleotide was made possible. The first three-dimensional image demonstrating the cleavage pattern of this structure-based chimeric dimer was thus constructed.

The ability of F2 to bind and cleave dsDNA in the presence metal ions was demonstrated in previous chapters. As a result, the DNA recognition and cleavage mechanisms of metallo-F2 could be further addressed in this chapter. We here demonstrated the first structure-based dimeric design with enhanced recognition and cleavage properties. Providing the advantage of a longer recognition site with a metallo-dimer with enhanced hydrolytic activity, the recognition/cleavage patterns of F2 were also strongly depending on the design of both target sites. This precise yet flexible site recognition property can be potentially useful for customized gene therapy in cancerous cells, especially with the ability of excise certain nucleotides and perhaps even for gene repair in the future. We hope to obtain a crystal structure of metallo-F2 in the near future in order to further understand the cleavage mechanism especially on the metal sites. The unique F2-DNA interaction is also an intriguing topic to investigate as it is one more step closer to understand the complicated protein-DNA interactions as well as our goal of designing potent and specific drug applications.

(A)



(B)

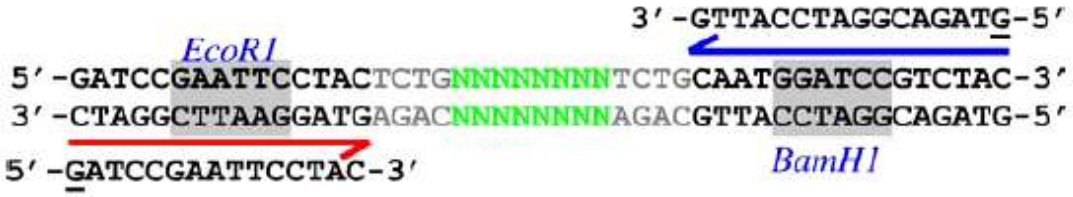


Figure 4.1 (A) DNA library with eleven randomized base pair (N_{11})(green letters) and PCR primers for electromobility gel shift assay (EMSA). The TAATTA site is located on the left of N_{11}. The primers (arrows) were ^{32}P-radiolabeled at the 5′-terminus prior to PCR. The restriction enzyme sites, *BamHI* and *HindIII*, are shown in gray boxes. The other restriction enzyme that is shown in blue box is used for the insertion verification of the DNA library that was inserted into pUC19 plasmid. (B) DNA library with eight randomized base pair (N_8)(green letters) and PCR primers for electromobility gel shift assay (EMSA). The primers (arrows) were ^{32}P-radiolabeled at the 5′-terminus prior to PCR. The restriction enzyme sites, *BamHI* and *EcoRI*, are shown in gray boxes.

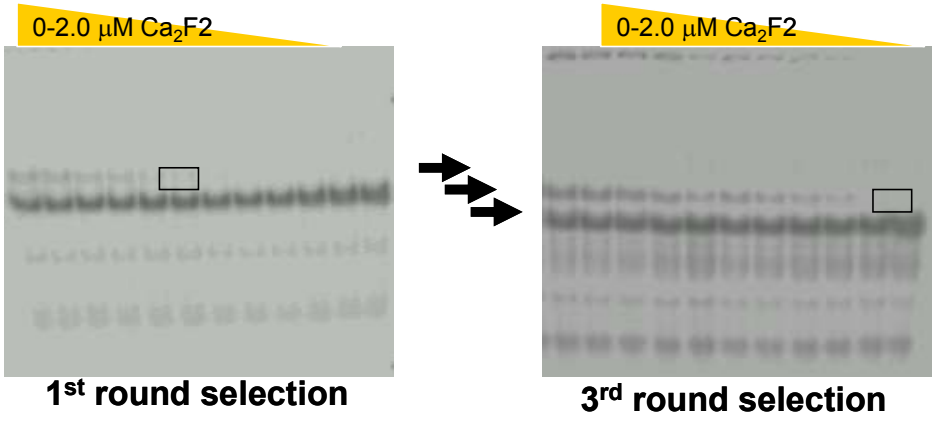
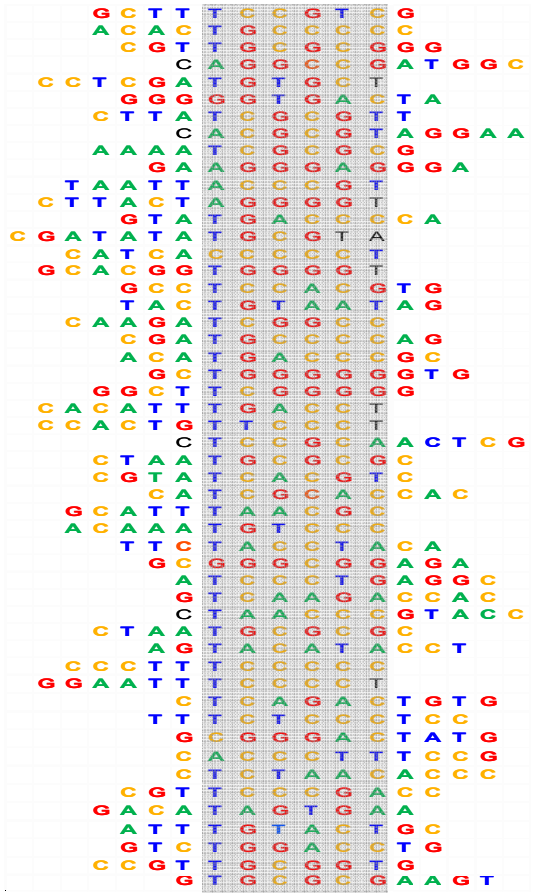


Figure 4.2 Polyacrylamide gel shift assay of DNA binding library with increasing amounts of Ca_2F2. The box containing protein-bound DNA was excised from the gel, and the DNA purified and amplified for the next round.

(A)



(B)

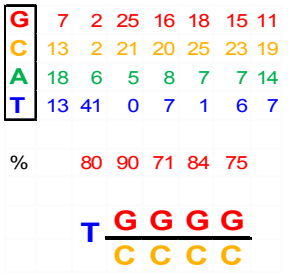
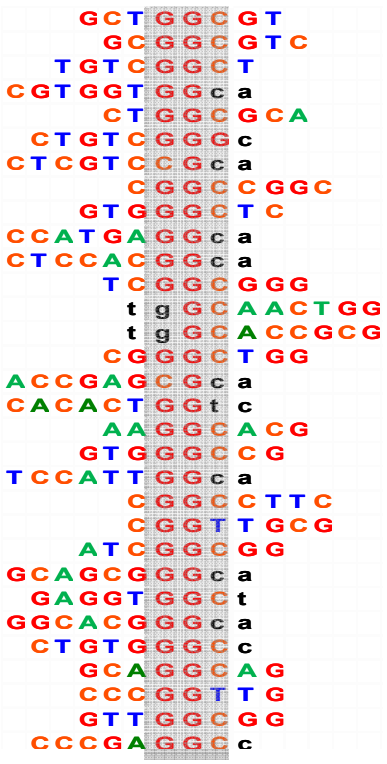


Figure 4.3 (A) The alignment of enriched DNA library binding to Ca_2F2. A total of 51 sequences were selected from 2 trials. The frequencies of each nucleotide for the alignment were summarized in (B) and the consensus sequence is 5′-T(C/G)(C/G)(C/G)(C/G)-3′.

(A)



(B)

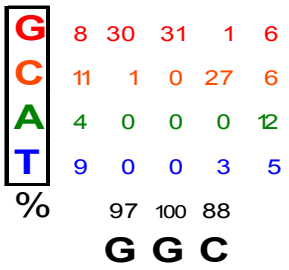
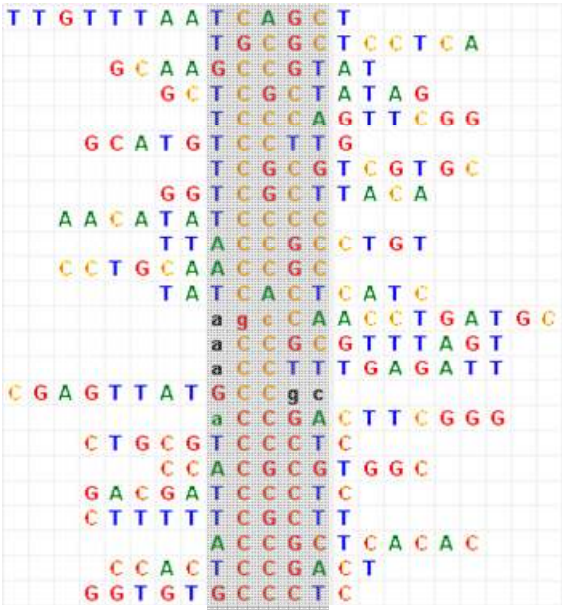


Figure 4.4 (A) The alignment of enriched DNA library binding to CaC2. A total of 31 sequences were selected from 2 trials. The frequencies of each nucleotide for the alignment were summarized in (B) and the consensus sequence is 5′-GGC-3′.

(A)



(B)

Figure 4.5 (A) The alignment of enriched DNA library binding to Ca_2F2. A total of 24 sequences were selected from 2 trials. The frequencies of each nucleotide for the alignment were summarized in (B) and the consensus sequence is 5′-(A/T)CC(C/G)(C/T)-3′.

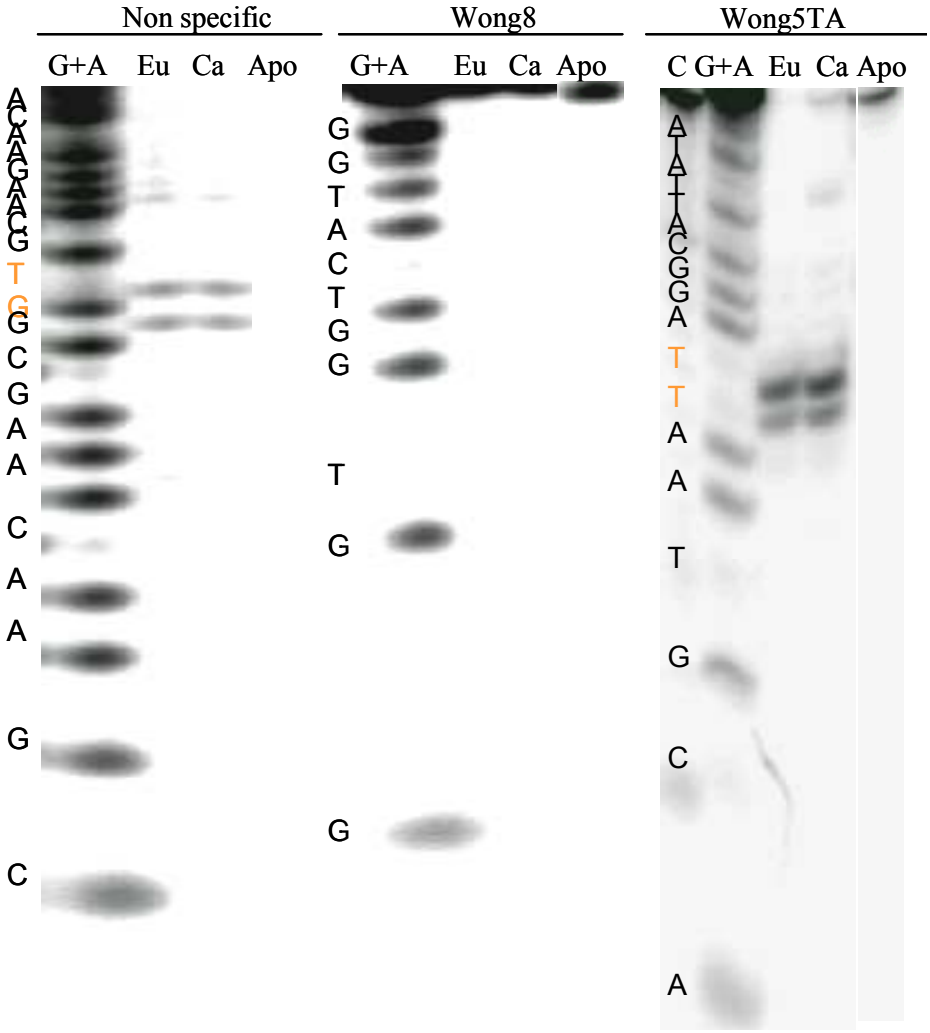


Figure 4.6 ^{32}P-labeled polyacrylamide gel of DNA sequences, Non specific, Wong 8, and Wong 5TA in the presence of Eu_2F2, Ca_2F2 and apoF2 represented as lane Eu, Ca, and Apo respectively. The DNA nucleotides were sequenced using Maxam-Gilbert sequencing technique, as indicated on lane C and G+A. The cleavage sites were represented in orange letters.

Table 4.1 DNA sequences for ^{32}P-labeled sequencing gel. Each group has identical potential cleavage sites represented in blue with different base pair gaps between two sites as shown underlined. The change of sequence pattern was indicated in orange.

Group	Name	Sequence (5′ →3′)
A	Wong 3	GACGTAATAGCATTATAGC
	Wong 4	GACGTAATTAGCATTATAGC
	Wong 5	GACGTAATGTAGCATTATAGC
B	Wong 3TA	GACGTAATTAGATTATAGC
	Wong 4	GACGTAATTAGCATTATAGC
	Wong 5TA	GACGTAATTAGGCATTATAGC
C	Wong 3AT	GACGTAATTAGATTAATGC
	Wong 4AT	GACGTAATTAGCATTAATGC
	Wong 5AT	GACGTAATTAGGCATTAATGC
D	Wong 3Rev	GACGTAATTAGTAATTAGC
	Wong 4Rev	GACGTAATTAGCTAATTAGC
	Wong 5Rev	GACGTAATTAGGCTAATTAGC
E	cs 1	GACGTAATTACCCCCAAGCTG
	cs 2	GACGTAATTAGGGGGAAGCTG

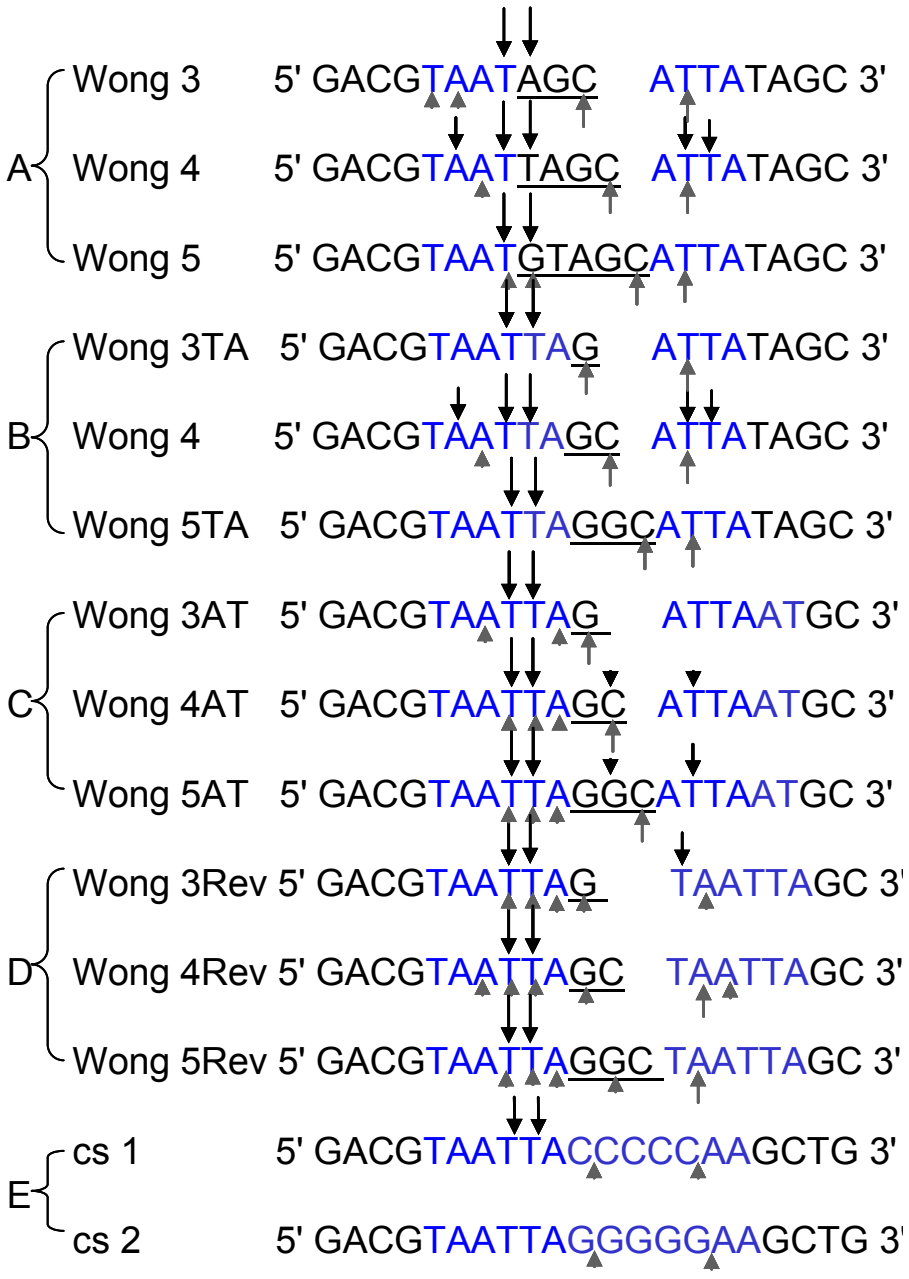


Figure 4.7 The cleavage sites of Eu_2F2 and Ca_2F2 with different DNA sequence patterns. The arrows indicated the cleavages (black: sequence as shown, grey: the complimentary sequence). The length of the arrow represented the intensity of the cleavage band on ^{32}P-labeled acrylamide gels. The base pair gap between two TAAT(NN) sites was underlined. The change of sequence pattern from the previous group was shown in orange letters.

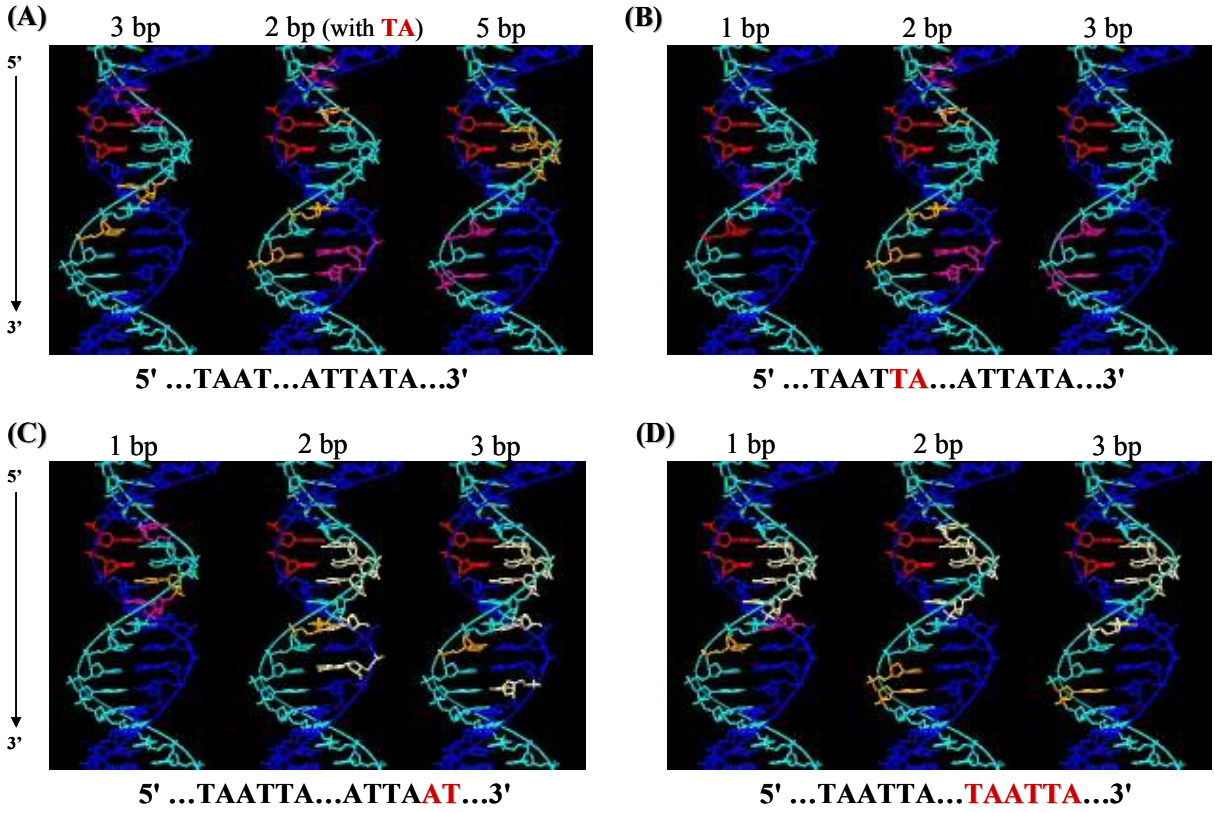


Figure 4.8 A three-dimensional view of the cleavage sites of group A to D with various base pair gaps (indicated on top of the picture) between two TAATNN sites. The cleavage intensity was related to the warm color of the nucleotide. The warmer the color is; the more intensive cleavage band was on the ^{32}P-labeling polyarcylamide gel.

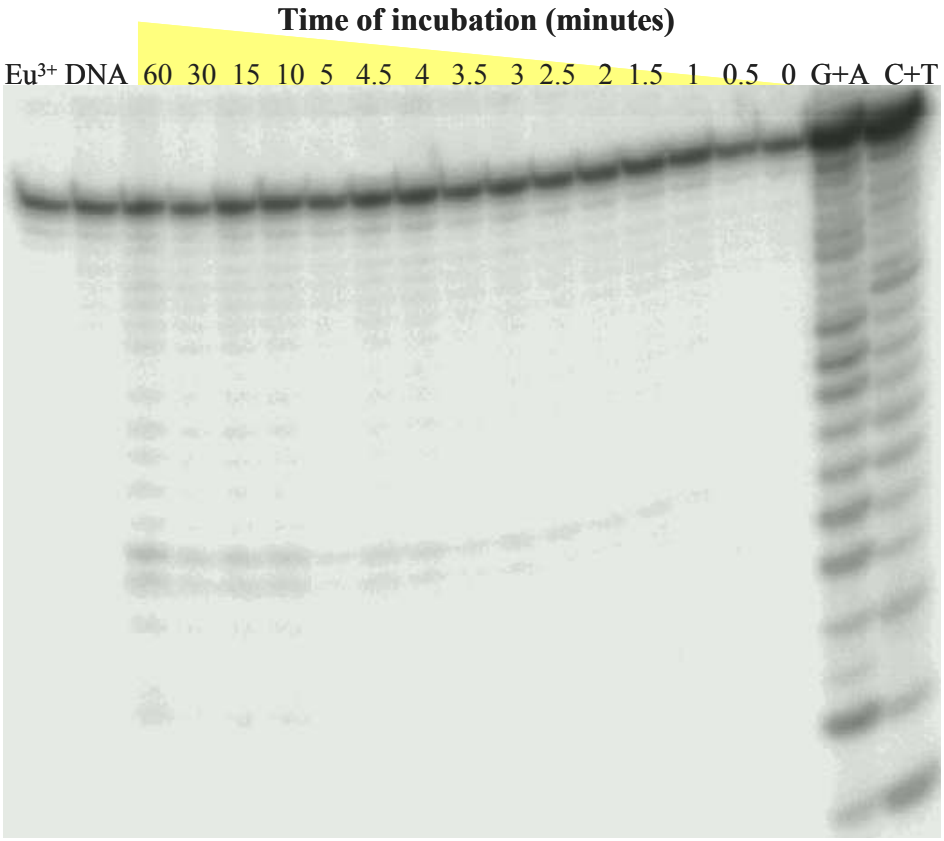


Figure 4.9 A polyacrylamide gel electrophoresis of Wong 5Rev incubated with Eu_2F2 at different time intervals. Lane Eu^{3+} represented the incubation of Wong 5Rev with $EuCl_3$ for 60 minutes. Lane DNA is the incubation of Wong 5Rev only. Lane 0-60: time of incubation in the presence of Eu_2F2. Lane G+A and C+T showed the DNA underwent Maxam-Gilbert sequencing reaction to be sequenced as the markers.

Table 4.2 The rate of reaction of Eu_2F2 with different DNA sequences observed from ^{32}P-labeled polyacrylamide gel electrophoresis. The red letters indicated the changes between sequences.

Name	Sequence	k_{obs} at 610 nM	Std. Dev.	K ($M^{-1}s^{-1}$)
Wong 3Rev	TAATTAG TAATTA	11.3 x 10^{-5} s^{-1}	± 4.7	120.2
Wong 4Rev	TAATTAGC TAATTA	4.2 x 10^{-5} s^{-1}	± 3.9	34.2
Wong 5Rev	TAATTAGGCTAATTA	2.3 x 10^{-5} s^{-1}	± 0.2	45.2
cs 1	TAATTACCCCCAA	3.6 x 10^{-5} s^{-1}	± 2.9	36.4

(A) **(B)**

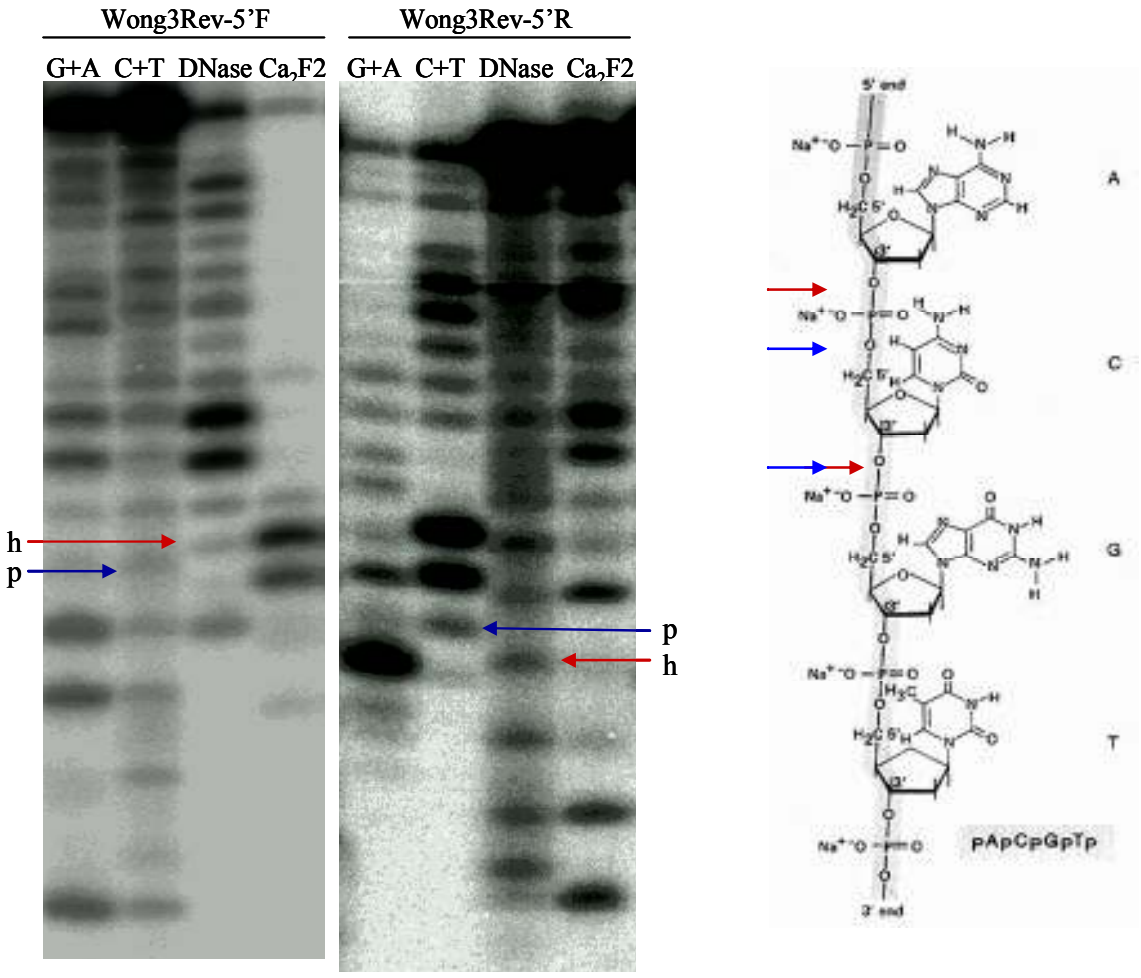


Figure 4.10 (A) Polyacrylamide gel electrophoresis of cleavage products of a 5′-end-labeled 21 bp DNA duplex, Wong22F. Lane 1, 2, 5,and 6: Maxam-Gilbert sequencing lanes producing phosphate termini (p = 3′OPO_3 termini, blue arrows). Lane 3 and 7: DNase I treated DNA producing hydroxyl termini (h = 3′-OH termini, red arrows). The chimera Ca_2F2 cleaved DNA, producing only 3′-OH termini for both complimentary strands of Wong3Rev. (B) The arrows (blue for Maxam-Gilbert sequencing and red for DNase I treated DNA) represented the cleavage sites from the view of a DNA molecule structure.

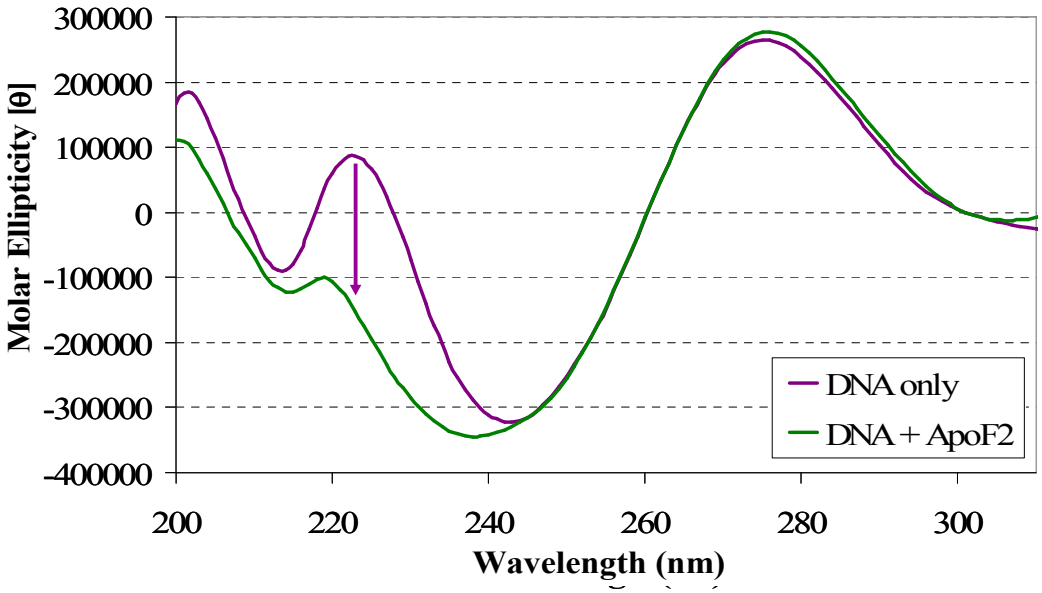


Figure 4.11 CD spectra of apoF2 in the presence and absence of DNA duplex.

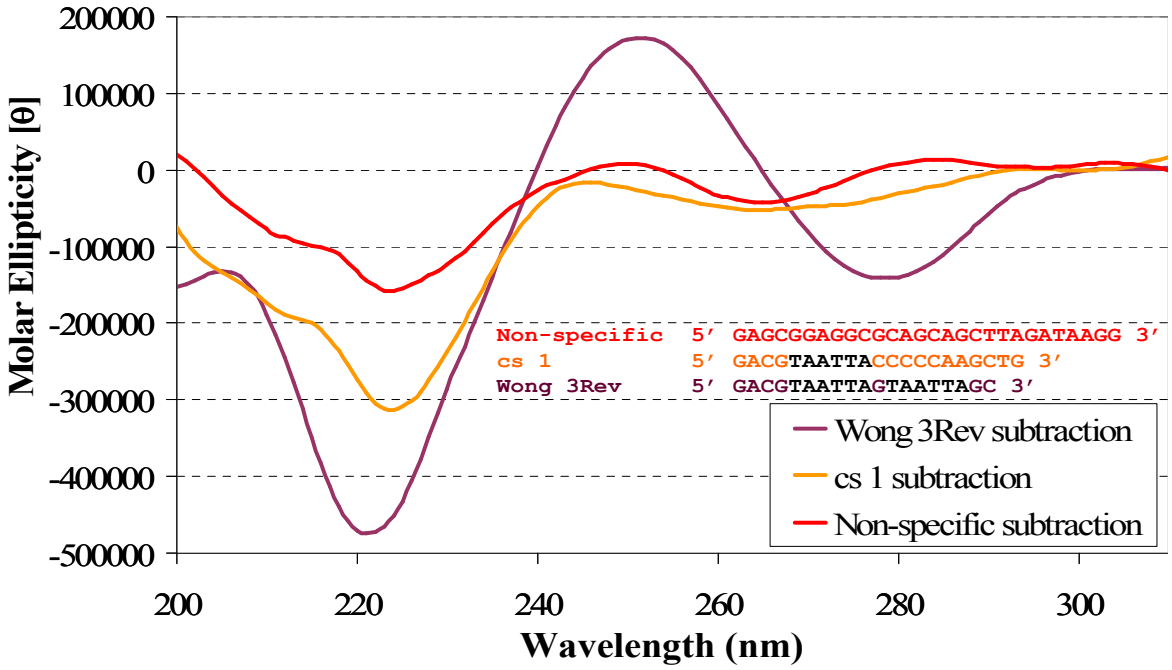
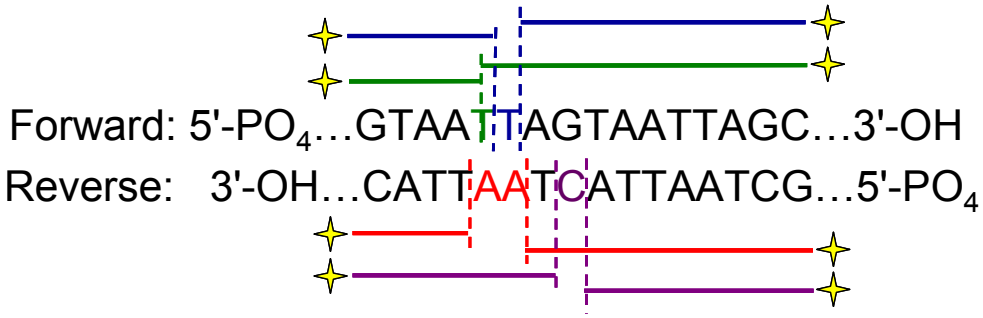


Figure 4.12 The subtraction of apoF2 in the presence of different DNA duplex sequences, Non-specific, cs 1, and Wong 3Rev, with the CD spectrum of DNA duplexes.

(A)



(B)

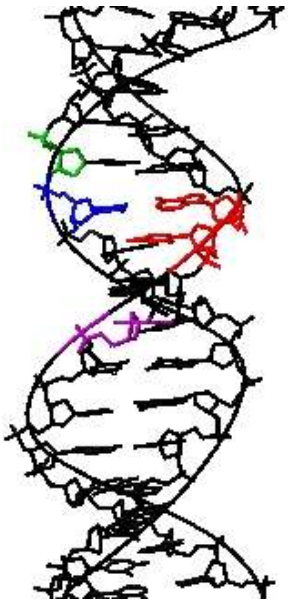


Figure 4.13 (A) The schematic model of the cleavage site observed from 3′-end- and 5′-end labeling on the same sequence of a DNA duplex, Wong3Rev. The dashed lines indicated the cleavage sites observed from both labeled ends polyacrylamide gel electrophoresis. (B) The colors that were coordinated with (A) were indicated in a three-dimensional view on a DNA duplex.

CHAPTER 5

DNA SEQUENCE PREFERENCE OF P3W AND P5B

5.1 Introduction

This chapter has been published in the Journal of Biological Inorganic Chemistry.[87] The EMSA experiment of P5b was performed by Dr. Youngbae Kim. The writing of this paper is a collaboration consult among the authors.

Among the most fundamental partnerships in biology is the transcription factor protein domain associating with its target DNA recognition element to regulate biological cascades. The selective protein-DNA binding involved encompasses many subtle, weak interactions. It is this ensemble of structural stabilization and flexibility, folding and dehydration, and specific hydrophobic and hydrogen-bonding contacts that direct the remarkable discrimination of targets from the extensive background of nearly identical sites. One of the goals of protein design is to engineer functional and selective artificial enzymes, which includes engineering such protein-DNA interactions. We recently demonstrated DNA-sequence discrimination in cleavage by cerium bound to a small 33-mer peptide (P3W) designed to retain a DNA-binding fold as well as incorporate a metal-binding site for reactivity.[68,81] In the current study, we sought to determine whether the observed non-random cleavage pattern produced by Ce(IV)P3W was a result of DNA sequence recognition by our chimeric metallopeptides, and if so, to identify the target DNA-binding site(s).

Designing an agent to target sequences of choice is a challenge that tests our understanding of protein folding, metal-binding, tuning of enzymatic activity, and intermolecular interactions between proteins and nucleic acids. Recognition of specific promoter regions *in vivo* generally entails direct sequence recognition by a small DNA-binding domain, which is then augmented by the recruiting and assembly of more

complex multi-protein structures. However, the initial direct recognition of the DNA-target is often achieved with surprisingly small DNA-binding motifs of less than 10 kDa, such as the homeodomains, leucine zippers, or zinc fingers. This suggests that even smaller peptides (less than 5 kDa) may be able to bind DNA with significant affinity and selectivity if they have defined structure, and several notable examples of such peptide motifs have been designed and described.[134-136]

We have approached the design of functional DNA-binding peptides by capitalizing on the structural similarity of two unrelated protein motifs. The EF-hand Ca-binding motif and the helix–turn–helix (HTH) DNA-binding motif are analogous supersecondary structures, each comprising two helices flanking an approximately 90° turn. In one case, the turn is a short β-turn, and in the other, the turn is a longer, 12-residue metal-binding loop. We postulated that these two turns were modular and thus could be substituted into the opposite motif with retention of overall structure. Therefore, we designed hybrid peptides that included the DNA-binding helices of the HTH motif and the Ca-binding loop of the EF-hand, as shown schematically in figure 5.1. It is notable that these minimalist designs incorporate only the C-terminal helices of the homeodomain (α2 and α3, omitting the N-terminal tail and α1), so in essence our peptides represent an incomplete "half-domain." However, despite their small size, we find that the chimeric Ln^{3+} metallopeptides indeed fold to an organized HTH structure and hydrolytically cleave DNA with some sequence preference.[135,137]

The ability to recognize a consensus site or family of sites is an important goal in the design of artificial enzymes. Our earlier work using sequencing gels showed Ce(IV)P3W generated nonrandom cleavage patterns within plasmid DNA fragments,[81] which suggests the folded peptide was delivering a metal to certain sites based on protein–DNA interactions. However, we could not determine the true extent of binding and sequence targeting by scanning a mere fraction of all possible sequences with that method. Here we utilized an iterative selection approach, enriching preferred binding

sites from a library of sequences. We find that in the presence of Ca^{2+}, designed peptides P5b and P3W (Figure 5.1) preferentially bind to a family of TG-rich DNA sequences similar to the site(s) previously shown to be cleaved by Ce(IV)P3W. Furthermore, by testing two similar metallopeptides that differ in their 2° structure content in the absence of DNA, we can compare a system where recognition must involve structural changes (induced-fit mechanism) to one that may or may not fold differently upon binding (simple docking mechanism).

5.2 Experimental Procedures

5.2.1 Peptides Design and Synthesis

The design of peptide P3W based on a helix-turn-helix motif (*engrailed* homeodomain) and the concensus EF-hand metal-binding loop has been described in detail previously.[68,81] P5b was designed similarly, with manual Protein Data Bank structural overlays of calmodulin (1OSA, first EF-hand)[137] and the related homeodomain, antennapedia (9ANT),[126] rather than *engrailed*.

Peptides were synthesized by standard *N*-α-(9-fluorenylmethyloxycarbonyl)-amino acid coupling chemistry, cleaved from resin, and purified by high-performance liquid chromatography to more than 95% purity (P3W, New England Peptide, Fitchburg, MA, USA; P5b, Caltech Peptide Synthesis Facility, Pasadena, CA, USA). The concentrations of the peptides were determined by absorption spectroscopy from the measured extinction coefficients (P3W ε_{280}=7,289 M^{-1} cm^{-1}; P5b ε_{280}=12,678 M^{-1} cm^{-1}), calculated by peptide digestion and quantification (University of Iowa Peptide Synthesis Facility). Eu^{3+} and Ca^{2+} binding affinities of P5b were determined by Trp fluorescence titrations, as described previously.[68]

5.2.2 DNA and Other Reagents

DNA oligonucleotides were synthesized by Integrated DNA Technologies, Coralville, IA. A synthetic 47-mer DNA oligonucleotide and complement were designed with PCR primer sequences at each end, and a library of randomized octamer sequences (65496 sequences) in the center (Figure 5.2). The randomized region was incorporated by coupling with a mixed base cocktail at these steps in the automated synthesis (phosphoramidite percentage weighted by base coupling reactivity). Each PCR primer sequence included a unique restriction site (*BamH1* and *EcoR1,* respectively), so that the final library could be cloned into a vector and sequenced. Buffers, acrylamide, and other reagents were obtained from Fisher Scientific. $EuCl_3 \cdot (H_2O)_6$ (99.99%) was from Aldrich, and stock solutions were made fresh by weight prior to use. Solutions were prepared in deionized distilled water (MilliQ 18 mΩ)

5.2.3 Circular Dichroism in the Presence/Absence of DNA

CD spectra of metallopeptides EuP3W and EuP5b were recorded with and without DNA on an Olis Cary-17 DS Conversion spectrophotometer at 20 ˚C under N_2 atmosphere in a 0.1-cm cell. Spectra were collected from 320 to 200 nm at 0.5 nm resolution and smoothed with Olis software (21 boxcar smoothing function). Samples contained 15 μM peptide in 50 mM tris(hydroxymethyl)aminomethane buffer (Tris pH = 7.8 at 25˚C) and 50 mM NaCl, with 15 μM DNA duplex as indicated. Stoichiometric Eu^{3+} was added to peptides as the chloride salt, from 10 mM stock freshly prepared by weight. Two DNA oligonucleotides were tested: E1, a self-complementary 14-mer oligonucleotide containing the parental (*engrailed*) target sequence (5′-GAGCTAATTAGCTC-3′), and E2, containing one P3W (shaded) and two P5b target sequences (underlined and double underlined) (5′-CATAGGGTGGCG-3′ and its complement 5′-CGCCACCCTATG-3′). Stock solutions of duplex DNA (1 mM duplex)

were prepared in 10 mM Tris-HCl (pH = 7.8), 50 mM NaCl and annealed by heating to 95 ˚C for 5 minutes followed by slow cooling to room temperature.

5.2.4 PCR-coupled EMSA

The 5′-end of two primers were radiolabeled with [α-^{32}P]*dATP* and T4 polynucleotide kinase for an hour at 37 ˚C and the reaction quenched by adding 20 mM EDTA to a final concentration of 10 mM. Unincorporated nucleotides were removed by micro Bio-Spin 6 chromatography column (Bio-Rad), which has a 5-base pair or 6000 dalton exclusion limit. The randomized library (N_8-oligonucleotide) was then amplified and radiolabeled by PCR. The 100 µL reaction mixture contained 1X *pfu* enzyme buffer (100 mM KCl, 100 mM $(NH_4)_2SO_4$, 200 mM Tris-HCl, pH = 8.75, 20 mM $MgSO_4$, 1% Triton X-100, and 1 mg/mL BSA), 10 mM *dNTP*s (Invitrogene), 400 ng/µL of radiolabeled primers (primer 1 and 2), 66 nM of double stranded N_8-oligonucleotide, and 2 µL (1U) of *pfu* DNA polymerase (Stratagene). The mixture was amplified with 25 cycles of 95 °C X 30 s, 55 °C X 60 s, and 65 °C X 30 s (Mastercycler PCR, Eppendorf). A final extension of 72 °C X 10 minutes was included in generating the initial labeled libraries prior to EMSA. The PCR products were purified by extracting with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with cold ethanol. Ethanol was removed from the DNA pellet by vacuum centrifugation (Labconco).

5.2.5 EMSA for Binding Site Selection

The ^{32}P-radiolabeled DNA containing the randomized N_8 core was incubated with increasing concentrations of chimeric peptide in 25 µL binding buffer (10mM Tris-HCl (pH = 7.4), 250 µM Ca^{2+}, and 7.5% glycerol for P5b; 10mM Tris-HCl (pH = 6.8), 300 µM Ca^{2+}, and 6% glycerol for P3W). The mixtures were incubated for 3 hours at room temperature. A 6% 1X TB polyacrylamide gel was prepared for each assay, and pre-run at 300 V for 30 minutes prior to loading to eliminate interference from residual ammonium persulfate and TEMED. An aliquot (12.5 µL) from each reaction mixture

was quickly loaded into the wells and the gel run at 4 ˚C and 155 V for 1.5 hours. Tracking dye was not added to the mixture but was run in an adjacent lane to track gel migration during electrophoresis. After electrophoresis, the gel was exposed to a Molecular Dynamics PhosphorImager Screen (Amersham Biosciences) for either an hour at room temperature or overnight at -20 ˚C, scanned and digitized (50 μm resolution; Storm PhosphorImager), then analyzed with Molecular Dynamics ImageQuant software. In each round of selection, a "bound band" section was excised from the gel at the lowest concentration of peptide producing an observable gel shift. A section was chosen from just above the "unbound" band, as the "bound" band was somewhat diffuse. This is presumably due to rapid peptide-DNA dissociation kinetics during electrophoresis. The selected DNA was extracted by soaking the crushed gel slice in DNA elution buffer (100 mM NaCl, 10 mM Tris pH = 6.91, 2 mM EDTA) for 2-4 hours at 37 ˚C, and gel materials removed by Micro Bio-Spin Chromatography (Bio-Rad). The selected DNA was further purified from residual peptide and salt by phenol extraction (phenol:chloroform:isoamyl alcohol (25:24:1)) and ethanol precipitation. The purified DNA was amplified by PCR using radiolabeled primers as described above. Residual polymerase was removed from the amplified DNA by phenol extraction and ethanol precipitation, and then used directly for the next round of binding site selection.

After five rounds of selection and amplification, the pool of DNA enriched in higher affinity sites was amplified a final time with non-labeled primers. This final pool of DNA was digested with restriction enzymes *BamHI* and *EcoRI* and subcloned into pUC19 plasmid. The plasmid was transformed in DH5α cells, plated, and DNA isolated from individual colonies was sequenced by the University of Iowa DNA facility. For each peptide, two trials of 5 rounds of amplification each were preformed. There was no significant difference in sequence patterns between the two trials, and thus the sequences were analyzed together. P5b: 39 total sequences (16 trial 1, 23 trial 2); P3W: 44 total sequences (25 trial 1, 19 trial 2).

5.3 Results and Discussion

5.3.1 Design and Characterization of Chimeric HTH-EF-hand Peptides

Peptides P3W and P5b were designed based on the structural similarity of the helix-turn-helix and the EF-hand, two physiologically unrelated motifs (Figure 5.1). The chimeric peptides comprise helices from DNA-binding homeodomains (*engrailed* and *antennapedia*) and one Ca-binding loop from *calmodulin*. The synthesis and characterization of P3W has been described previously.[68] P5b was designed with similar modular substitution of the HTH by an EF-hand Ca-binding loop. Each 32- and 33-mer peptide incorporates a single metal-binding site within the DNA-binding motif. As a result, the chimeras have the ability to bind to a metal ion (calcium or trivalent lanthanides) as well as DNA.

Residues F_6 to E_{21} of P5b are derived from the *calmodulin* sequence, which in the native structure comprises the Ca-loop and includes a full α-turn of N-terminal helices. Notably, several important hydrophobic interactions on either side of the metal-binding loop that are important for stabilizing the Ca-binding pocket are conserved in EF-hands (residues -1/+13/+16, relative to the metal-binding loop, which is designated +1 through +12; see Figure 5.1).[138] P3W retains these contacts (I_8, I_{21}, W_{24}) and is well-folded as evidenced by our previous NMR solution structure of LaP3W.[68] P5b in contrast retains a hydrophobic contact at position -1/+13 (F_9 and I_{22}) yet the important hydrophobic pair at positions -1/+16 was omitted to test whether these designs also flexibly accommodate hydrophobic contacts at -1/+14 (F_9/W_{23}) or -1/+15 (F_9/F_{24}).

Europium titrations of both designed peptides were followed by circular dichroism spectroscopy to determine the affect of metal on secondary structure (Figure 5.7).[68] Although some HTH contacts are retained in P5b, the disruption of -1/+16 contact resulted in a substantially less robust fold that has little secondary structure outside the

metal-binding domain in the absence of a DNA-target. P3W becomes more helical upon binding equimolar Eu^{3+}, as indicated by increased negative ellipticity at 222 nm, while the secondary structure of P5b remained nearly unchanged even with excess Eu^{3+}. In 50% trifluorethanol, an organic solvent that enhances α-helix secondary structure, EuP5b approaches the secondary structure content of EuP3W, suggesting EuP5b can access a more complex fold in certain contexts. Similar effects are seen for Ca^{2+}, though the alpha-helical content is somewhat lower than with the trivalent lanthanides.

Trp fluorescence titrations were used to measure the metal binding affinity based on fluorescence quenching as a function of metal concentration.[66,100] Both peptides contain a tryptophan residue that provides a spectroscopic handle for subtle local structural changes upon metal complexation. The dissociation constants of EuP3W and CaP3W are 5.9 ± 0.3 μM and 35 ± 1.4 μM respectively,[68] while the K_d for EuP5b was found to be 12.6 ± 1.9 μM and 70 ± 8 μM respectively. Although both peptides had stronger affinity towards lanthanide ions than calcium ions, the chimeric EF-hand derivatives bind both types of ions, and do so with similar relative affinities. Although metallated P5b has less secondary structure than P3W, the peptides' binding affinities towards Ca^{2+} and Eu^{3+} are similar.

5.3.2 Selection of High Affinity DNA Sites

In order to determine the consensus nucleotide binding site(s) of our chimeric metallopeptides, we adopted an iterative approach to finding the target sequence. A library of randomized 8-mer sequences was incorporated into a 47-mer oligonucleotide of the sequence shown in Figure 5.2. This library was amplified by PCR using ^{32}P-radiolabeled primers complementary to the sequences to either side of the N_8-site. The labeled library was then subjected to an electromobility gel shift assay (EMSA) with increasing concentrations of P3W and P5b peptide (Figure 5.3). The shifted band was then excised from the gel, and the resulting DNA enriched in preferred binding sites was

again amplified by PCR. This method is an adaptation of published methods,[49,139] but includes 300 μM calcium in the samples and in the gel to promote peptide folding and DNA binding. Importantly, Ca^{2+} does not promote cleavage; we have found that unlike the lanthanide-loaded peptides, Ca-peptides are virtually inactive as hydrolytic nucleases at neutral pH (data not shown).

Although the Ca-peptides are similar in design and overall charge (CaP3W pI = 9.8; CaP5b pI = 11.5), CaP5b has a background (non-specific) binding affinity for DNA approximately an order of magnitude stronger than CaP3W. In the initial screen, a gel shift was observed at 1.5-2.0 μM CaP5b, and at 15 μM CaP3W. After 5 rounds of selection, the target DNA for both systems was enriched in higher affinity sites. After selection, the midpoint of the gel shift had decreased to ~200 nM CaP5b, and 1.0 μM CaP3W, corresponding to an enhancement in binding affinity of 8-10-fold for P5b, and 15-fold for P3W. The fact that bands due to peptide-DNA adducts are smeared rather than distinct indicates that the DNA-binding interactions are relatively labile, consistent with our inability to observe DNA-footprinting effects in sequencing gels (data not shown).

5.3.3 Consensus Target Sequences

After five rounds of selection, the DNA library was amplified with unlabeled primers and subcloned into the *BamH1/EcoR1* site of pUC19. The plasmid was transformed into DH5α cells, plated, and individual colonies sequenced. The resulting sequences were aligned to identify consensus binding targets, as shown in Figures 5.4 and 5.5. Initial alignments took into consideration the most commonly occurring 4-mer and 5-mer sequences, and those sequences that were identical (P5b: two examples of 5′-GCCATAGC-3′) or nearly so (P3W: 5′-(A)AGGGCAA(G)-3′; 5′-(T/G)ATGTGCA-3′). Both forward and reverse directions (relative to primer) were considered, but each occurrence was included in the alignment only once. The consensus alignments were

further confirmed by motif analyses with search algorithms Motif Sampler[140] and MDscan.[141]

The apparent binding site(s) for each peptide are distinct, yet related. For both CaP5b and CaP3W, the selected sequences were rich in 5′-TG-3′ steps. For CaP5b, 5′-TG-3′ (17%), 5′-AG-3′ (15%), and 5′-GT-3′ (13%) comprised nearly half of all base steps, and for CaP3W, 5′-TG-3′ (20%) and 5′-GG-3′ (15%) comprised a third of the steps.

The length of the target site was not known *a priori*, since the binding behavior of these chimeric constructs had not been previously investigated. It was expected that the targets would be no longer than that of the parental *engrailed* or *antennapedia* homeodomains (6 b.p.), and likely somewhat shorter (4-5 b.p.), since the chimeras in essence represent "half-domains" relative to the larger parental homeodomains. CaP5b recognized a single family of sites 4-6 b.p. in length (5′-pur-T-pur-G-(G/C)-T-3′). There is a strong consensus in the first four positions, and a moderate consensus in the next two positions (Figure 5.4). There is a notable preference for A (51% A, 33% G) in the first purine position, though no strong preference is observed between purines at position three. The fifth position has more ambiguity, but a modest preference for G/C (G + C = 67%). The sixth position is predominantly T (44%).

CaP3W, in contrast, recognizes two distinct families of 5 b.p. sites (Figure 5.5). Although these sequences may be considered as two overlapping subsets of 5′-X-T-G-(G/T)-G-X-3′ with a common 4 b.p. core, the distinction in the third position of the core is significant (63% T in site A, 90% G in site B), and is the basis of considering these target families separately. The occurrence of two target sites suggests unique peptide-base contacts in each case, presumably arising from different binding modes or orientations.

The manual alignments were confirmed with the results of search algorithms MDscan and Motif Sampler,[140,141] which consider statistical analyses of most commonly

occurring sequences. For CaP5b, the 4 b.p. sequences with the highest frequency are 5′-GTAG-3′, 5′-TAGG-3′, and 5′-TAGC-3′, representing 9% of all occurrences. Alignment of these 4-mers at the common TAG triad suggests a binding preference for 5′-(G)-T-A-G-(G/C)-3′. The Motif Sampler algorithm identified two 5 b.p. sequences (5′-(G/A)-T-A-G-(G/C)-3′ and 5′-X-(T/A)-(T/A)-G-(C/G)-3′). MDscan, a similar algorithm that statistically scores multiple local sequence alignments, found overlapping 6 b.p. sequences with a 5′-TAGGA-3′ core (i.e., 5′-GTAGGA-3′ and 5′-TAGGAC-3′). Thus, these analyses are consistent with the manual alignment of the CaP5b target given in Figure 5.4.

For CaP3W, the four most frequently occurring 4-mer sequences are 5′-GTGC-3′, 5′-GGTG-3′, 5′-TGAG-3′, and 5′-TGGG-3′, representing 11% of all 4 b.p. sequences within the EMSA results. Alignments of these sequences at the TG diad result in a 6 b.p. sequence (5′-GGTGXG-3′), likely longer than is truly recognized by the small peptide. Motif Sampler identified one 5 b.p. sequence like manually aligned Site A (5′-(G/T)-T-T-G-(G/T)-3′) and a less clear second target family similar to Site B (5′-(G/C)-A-(G/C)-(G/A)-(G/T)-3′). MDscan identified only one target sequence (5′-TGTGC-3′), related to Site A. When searching for a 6 b.p. consensus, both algorithms included an adenine base to the 5′-side of the sequences, which is consistent with the manual alignments as well.

5.3.4 Peptide Structure as a Function of DNA Binding

To observe the influence of DNA-binding on metallopeptide structure, the change in peptide secondary structure was followed by difference CD spectroscopy. The molar ellipticity of each metallopeptide was observed in the presence and absence of duplex DNA. Two 14-mer nucleotide sequences were tested. One (E1) contained the parental *engrailed* target sequence 5′-TAATTA-3′, which was found not to be a binding target for either CaP3W or CaP5b. The second (E2) contained target sequences identified from both peptides. The CD spectrum of annealed DNA oligonucleotides alone had CD

spectra characteristic of stacked, B-form DNA.[132] The addition of metallopeptide to an equimolar amount of either duplex E1 or E2 resulted in an increase in negative ellipticity below 260 nm, while the positive, nucleotide-derived band at 272 nm remained essentially unchanged (Figure 5.6A). The difference CD spectrum of the DNA-bound metallopeptide was obtained by subtracting the free DNA signal from that of EuP3W-DNA or EuP5b-DNA (Figures 5.6B and 5.6C). For metallopeptides binding to E1 (no target site), neither metallopeptide showed large change in the relative intensities of the 204 and 222 nm bands, which report on peptide secondary structure. Additionally, nucleotide E1 appeared to have no change in duplex structure in the presence of metallopeptide, based on the positive band at 272 nm. However, for metallopeptides binding to E2 (containing target sites), greater differences in peptide structure and α-helicity were observed (Figure 5.6B). EuP3W, which already adopts a folded structure in the absence of DNA, remained well-folded when DNA-bound. The absolute intensity of the EuP3W bands increased in the presence of DNA target, although they changed little in the relative intensities (blue arrow in Figure 5.6B). This suggests that binding target DNA stabilizes, but likely does not significantly change the helix-turn-helix structure of EuP3W. Estimates of the percentage helicity based on ellipticity at 222nm [142] confirm this trend (Table 5.1). However, the absolute numbers are likely less informative in this case, as small changes in DNA structure could contribute to the negative intensity in the UV.

In contrast, EuP5b showed significant changes in both the intensities and the ratio of the peptide-based bands, suggesting this less-organized metallopeptide adopts significantly greater helical content upon binding its target (red arrow). Additionally, EuP5b appears to impose subtle structural changes upon the target DNA, particularly noticeable with an increase in the negative band at 250 nm. However, neither target DNA nor random DNA was significantly altered by peptide binding, although the

secondary structure of the chimeras in the presence of DNA was dependent on nucleotide sequence.

5.4 Conclusion

An important question addressed by this work is whether cleavage specificity and binding specificity are related, the goal of structure-based metalloenzyme design. Although neither cleavage assays nor binding assays identified *unique* sequence targets for the small peptides, the EMSA analysis shows that both CaP3W and CaP5b recognize distinct families of target sequences.

Ce(IV)P3W was found previously to cleave DNA with modest sequence preference,[81] though the DNA fragments tested were necessarily a small subset of all possible targets, and did not encompass the consensus targets identified here. However, among the sequences found to be preferentially cleaved were the thymine and adenine (underlined) within the sequence context 5′-T<u>T</u>C<u>A</u>CCT-3′ (complement 5′-AGG<u>T</u>G<u>A</u>A-3′), which is very similar to the consensus target Site A (compare Site A: 5′-ATGTG-3′ to cleavage target complement 5′-AGG<u>T</u>G<u>A</u>A-3′). Note also that 5′-GGTG-3′ is among the most frequently targeted 4-mer sequences by CaP3W. Thus, the Ce(IV) cleavage results are consistent with oriented metallopeptide binding to DNA, and cleavage to one side of the binding site.

The recognition targets for CaP3W and CaP5b are related, yet distinct, which raises the question of what factors dictate the difference in behavior. Both chimeric peptides were designed by modular loop substitution of a Ca-EF-hand binding site into the turn of a DNA-binding HTH motif. However, the peptides differ in two important ways. First, they derive from different parental HTH motifs (*engrailed vs. antennapedia*), so the recognition helices are unique. Second, the extent of secondary structure in solution differs. P3W in the presence of Ca^{2+} or Ln^{3+} folds significantly, as demonstrated by CD spectroscopy and the NMR solution structure of LaP3W.[68] In

contrast, P5b is poorly folded even with excess metal ions (Figure 5.7). In the presence of trifluoroethanol, EuP5b can achieve similar 2° content to folded EuP3W, indicating P5b can accommodate greater helical structure. More importantly, EuP5b has dramatically increased helical content in the presence of DNA (Figure 5.6 and Table 5.1).

Even though EuP5b was not as well-folded as EuP3W in aqueous solution, the secondary structure of EuP5b was induced more significantly than that of EuP3W upon binding DNA that contains recognition sites for both peptides. This affect was only observed in DNA with target sequences and not with non-specific binding sites, suggesting that the induction of secondary structure was sequence selective and DNA-dependent. Moreover, the stacking conformations of DNA were altered to a greater extent within recognition sites, showing that there are structural adjustment on both DNA and peptides upon binding to the target sequences. As a result, rigid peptide secondary structure is not a prerequisite for the metallopeptides to recognize a target site. In fact this is consistent with the behavior of a number of native homeodomains themselves, including *engrailed, NK-2*, and *Pbx1* some of which are known to adopt a canonical extended recognition helix (α3) or additional C-terminal helices only upon binding their target DNA.[51,143-146]

EuP3W, unlike EuP5b, has super secondary structure independent of DNA binding, although this structure also appears to be enhanced and stabilized by specific nucleic acid interactions. A pre-organized structure could potentially approach the DNA groove in several ways. The HTH motif in different protein contexts interacts with DNA in a variety of orientations, from nearly parallel to nearly perpendicular to the DNA helical axis.[147] Since P3W has at least 2 distinct families of sites, it appears that two predominant binding orientations occur. Assuming the peptides bind in the major groove as do the parental motifs (likely, based on the small perturbations in DNA structure evidenced by CD and the native helix-turn-helix precedence), the two binding modes must differ in their contacts to the central guanidine of 5′-TG<u>G</u>G-3′ (H-bonding) vs. the

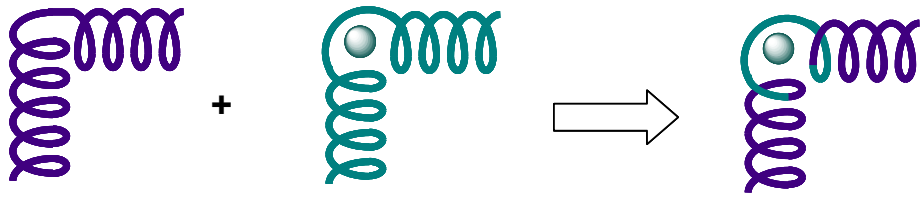
thymidine of 5′-TGTG-3′ (van der Waals contact to the T-methyl group). As discussed previously,[81] the omission of the N-terminal tail would be expected to preclude binding to the first two bases of the consensus homeodomain target (5′-TAATXX-3′), and might also allow much greater flexibility in the relative orientation of the recognition helix in the groove.

Intriguingly, the very G-rich sequence recognized by CaP3W (5′-AGGGT-3′) is more similar to the telomer repeat (5′-…TTAGGGTTAGGGTTAGGG…-3′) than the parental *engrailed* sequence (5′-TAATTA-3′). The telomer repeat is selectively recognized by the HTH-containing protein RAP1, whose recognition helix has little sequence homology with *engrailed* or P3W.[148] However, an overlay of the HTH domains of RAP1 and *engrailed* shows that the region of interactions is shifted by approximately one α-helical turn (3-4 residues, and ~30 degrees) between the two. Thus, if the *engrailed* HTH motif were rotated in the major groove to contact the G-rich sequence in a manner similar to RAP1's binding mode, the important pair of RAP1 Arg residues making specific DNA contacts to the G base triplet (i, i+4 residues R_{542}, R_{546} of RAP1)[148] could be mimicked by a pair of *engrailed* Lys residues (i, i+4 residues K_{46}, K_{50} of *engrailed*). While certainly this postulate to explain the recognition of the RAP1 target by the chimeric P3W is speculative without further structural studies in the presence of DNA, the ability of an *engrailed*-derived "half-domain" to target and binding a telomer-like sequence is consistent with contacts through an altered face of the folded recognition helix.

It is clear that although the designed metallopeptides do not exhibit the level of specificity of the parental motifs, they behave as structured recognition units. This confirms our earlier conclusion that the selectivity in cleavage demonstrated by Ce(IV)P3W is due to binding selectivity,[81] and thus these peptides are true artificial nucleases in that they recognize and cleave a specific DNA sequence target.

The DNA-binding targets of two designed chimeric metallopeptides were determined by an EMSA selection assay. Although these peptides represent essentially a "half-domain" relative to the parental homeodomains from which they are derived, both have a nonrandom sequence preference for distinct families of DNA sites. Metallated P3W, which has significant secondary structure in the absence of DNA, recognizes two similar yet unique sequence families. This suggests that P3W contacts the DNA through two distinct binding "faces." In contrast, metallated P5b, which folds significantly only in the presence of DNA, recognizes one DNA sequence family. This induced-fit mechanism allows the flexible peptide to bind more tightly to its target DNA than peptide P3W. These results confirm our goal of structure-based peptide design to create a DNA-binding artificial nuclease with true sequence discrimination.

(A)



(B)

	-1 +1	+12	+16
P3W	TERRRQQLDKDGDGTIDEREIKIWFQNKRAKIK		
P5b	TRRRRFSLFDKDGDGTITTKEIWFQNRRMKWK		

Figure 5.1 (A) Schematic depiction of the chimeric peptide design. The DNA-binding helix-turn-helix motif (purple) and the Ca-binding EF-hand motif (green) are structurally analogous alpha-alpha corners, and thus can be combined by modularly replacing turn sequences. (B) Sequences of designed chimeric peptides P3W (33 aa, based on *engrailed*) and P5b (32 aa, based on *antennapedia*). The EF-hand metal-binding loop (designated +1 through +12) is shaded in each case.

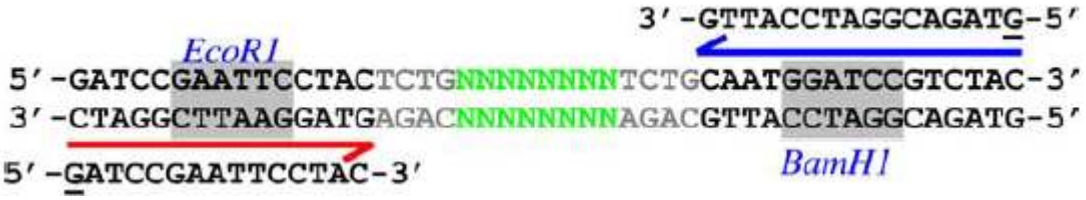


Figure 5.2 DNA library with eight randomized base pairs (N_8) and PCR primers for EMSA binding analysis. The primers (arrows) were ^{32}-P-radiolabeled at the 5′-terminus prior to PCR. The restriction enzyme sites, *BamH1* and *EcoR1* are shown in gray boxes.

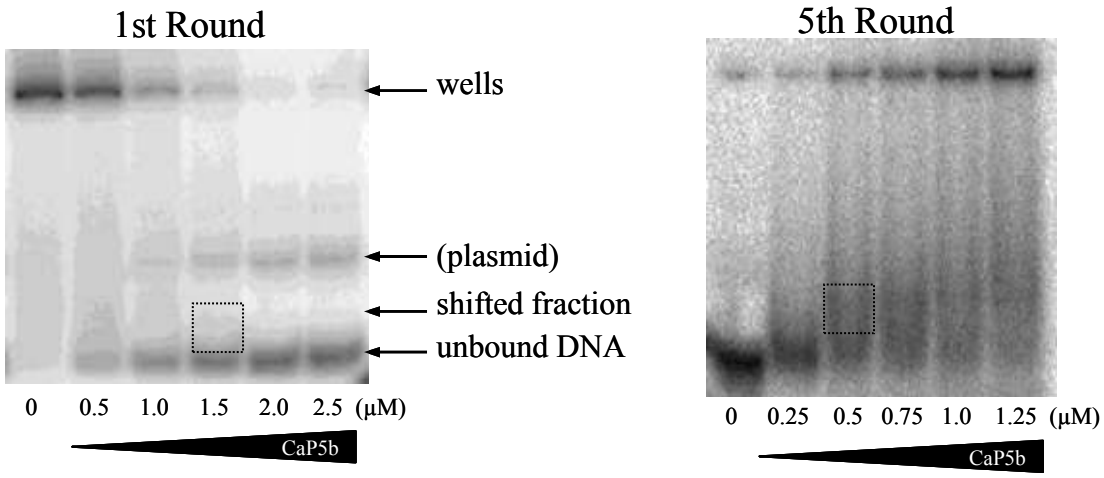
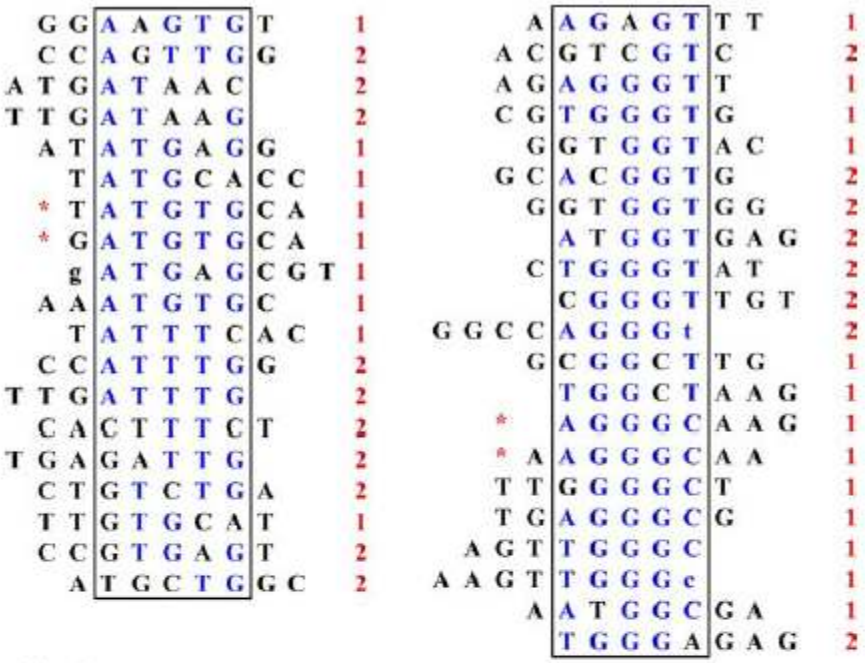


Figure 5.3 Polyacrylamide gel shift assay of DNA binding library with increasing amounts of P5b, in the presence of Ca^{2+}. The dotted box containing peptide-bound DNA was excised from the gel, and the DNA purified and amplified for the next round. CaP5b/DNA binding conditions: 10 mM Tris, pH 7.4, 500 μM $CaCl_2$, 10% glycerol.



Site A

	1	2	3	4	5
a	13	2	2	5	2
t	1	15	6	12	0
c	1	0	2	2	3
g	4	2	9	0	14
	A	**T**	**G/T**	**T**	**G**
%	**68**	**79**	**79**	**63**	**74**

Site B

	1	2	3	4	5
a	9	0	1	0	1
t	6	5	0	0	13
c	2	1	1	2	7
g	4	15	19	19	0
	A/T	**G**	**G**	**G**	**T/C**
%	**71**	**71**	**90**	**90**	**95**

Unaligned:

C A T A G T G C
C T G T A G T T
G T T C G A G A
G C T G C A G C

Figure 5.4 Aligned binding sites for CaP3W obtained by EMSA selection. Two families of target sites were identified, with the consensus sequences tabulated below. Two trials of 5 rounds of selection were performed, and the 44 sequences analyzed together (trial number indicated on the right). Asterisks indicate sequences identical in 7 of 8 positions. Four sequences were unaligned.

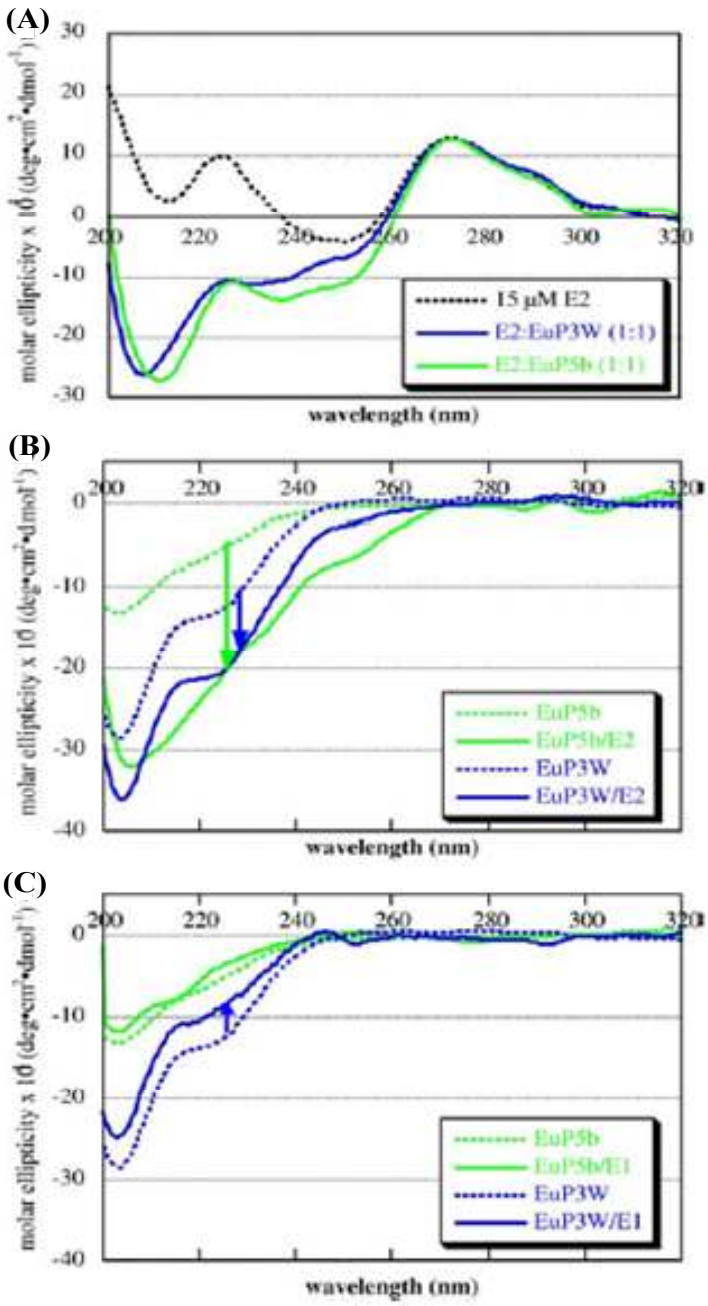


Figure 5.5 Circular dichroism spectra of metallopeptides in the presence and absence of equimolar duplex DNA (E1: non-specific target; E2: target binding sequence). Data are presented as molar ellipticity based on peptide concentration. (A) Spectra of 15 µM EuP3W, EuP5b, and free E2 duplex, 50 mM Tris, 50 mM NaCl, pH 7.8. (B) Eu-peptide spectra and difference spectra ((Eu-peptide + DNA)-free DNA signal) with binding target E2. (C) Eu-peptide spectra and difference spectra ((Eu-peptide + DNA)-free DNA signal) with non-specific target E1.

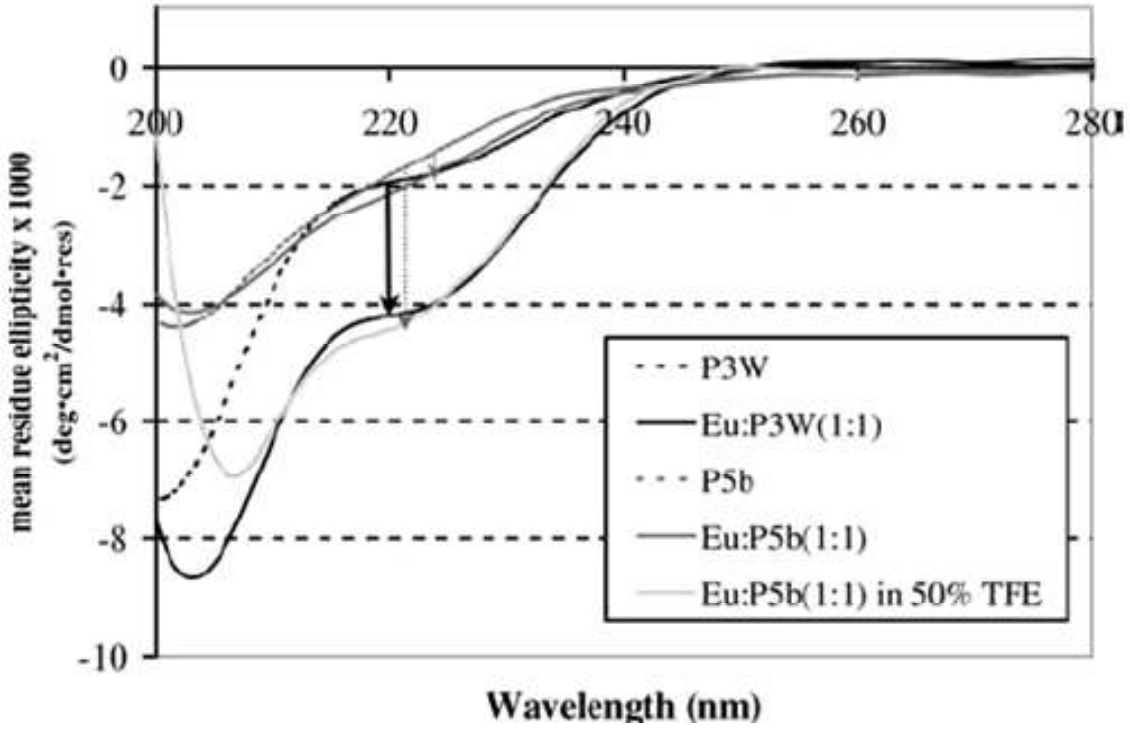


Figure 5.6 Circular dichrosim spectra of P3W and P5b with and without Eu^{3+} (pH = 7.80 at 25 °C, 50 mM Tris and 50 mM NaCl). EuP5b is poorly folded in aqueous solution, but has similar secondary structure to EuP3W when in 50% trifluorethanol solvent (dashed gray arrow).

Table 5.1 Alpha-helical content of peptides in the presence and absence of DNA, based on molar ellipticity at 222 nm. 100% helical content was assumed to be 30,000 (deg•cm^2)/dmol that is adapted from reference 132. Note that these values are necessarily estimates, since small changes in DNA structure could also contribute to negative intensity in the UV.

Mean Residue Ellipticity	**[222](deg•cm^2•dmol^{-1})**	**Total % α-helicity**
EuP5b	-1988	7
EuP5b/E1	-1447	5
EuP5b/E2	-7094	24
EuP3W	-3970	13
EuP3W/E1	-2988	10
EuP3W/E2	-6424	21

CHAPTER 6

PHYSICAL CHARACTERIZATION OF A CHIMERIC DIMER, R7C

6.1 Introduction

In our efforts to design metalloprotein dimers, various methods of designing artificial dimers based on our successful monomer model were utilized. One of the design approaches is dimerization through a polypeptide linker connecting the head (N-terminus) to the tail (C-terminus) of each monomer. This dimer was described thoroughly in Chapter 2. Another approach is to connect each monomer at the same terminus, head to head or tail to tail (Figure 1.5 and 6.1).

The dimer design with head to tail and head to head connections were postulated as feasible models (Figure 6.2). Both models have large degree of freedom on DNA interaction given by the flexible N-terminus or the flexible polypeptide linker. The model of the tail to tail dimer was not favored because of the rigid DNA interaction between the DNA binding domain of the third helix and the major groove. The termini of both third helices point to an opposite direction demonstrating that a long linker may be needed for the tail to tail dimer design and the rigidity of the C-terminus might not be compatible for structural modification.

Our studies of the head to tail dimer, F2, showed it to be a hydrolytically active DNA selective artificial endonuclease as described in Chapters 2 through 4. In order to understand the key contributions on the kinetic and selective properties of F2, it was important to compare its properties with a similar dimer model (Figure 6.1). The head to head dimer design, R7C, was therefore a great candidate for investigating the metallo-dimer system that enhanced hydrolysis and selectivity mechanisms. Since DNA hydrolytic activities between a monomer and a dimer are dramatically different as described in Chapter 3, the dimerization of a dimer through disulfide bonding can be a key for activity control. The oxidation and reduction of the disulfide bond on R7C can

thus be utilized as a "switch" for hydrolytic activity. A model with non-covalently linked domains can also allow us to understand the importance of being dimeric. Does dimer formation of a structural-based metalloprotein contribute to the enhancement of DNA hydrolysis and selectivity? Is the behavior of F2 dependant solely on the direction of its design? This simple and short model allows us to focus on what the advantages of dimerization may bring and thus shine light on further dimeric protein engineering that encourage particular activity of a native protein.

6.2 Experimental Procedures

6.2.1 Chimeric Design

The known crystal structures of *engrailed* homeodomain (2HDD)[96] and the Ca-binding EF-hand motif of calmodulin (1EXR)[97] were oriented and aligned manually using Swiss PDB Viewer.[98] The design of the chimeric monomer was based on the overlaying of the HTH motif of engrailed homeodomain at the α–α turn and calmodulin EF-hand metal-binding loop because of their similar topology.[66,68]

The dimer design with head to head connection was considered due to its close proximity and flexibility at the N-terminus of each domain while both DNA-binding domains were well-situated at the major grooves (Figure 6.1). Since the chimera did not contain a cysteine residue, the homodimer was designed to dimerize through disulfide bond by performing site mutation from Arg7 to Cys7 at the flexible N-terminus of the monomer, C2.

6.2.2 DNA Cloning

R7C mutagenesis was carried out using a vector, pET21a(+)(Novagen, 5369bp), containing the DNA encoding for the monomeric chimera (provided by Sunghyut Lim, University of Iowa), which has a T7 promoter without an affinity tag for purification. Primers, wong05F & R, were ordered from IDT for point mutation (Figure 6.3). The

procedure for the mutagenesis was performed using the site-directed point mutagenesis kit and procedure by Stratgene. The PCR conditions were 95 °C for 30 seconds, 55 °C for 30 minutes, and 68 °C for 8 minutes over 18 cycles. The final product was isolated from cells via a DNA purification kit (Qiagen). The DNA construct encoding the target protein was confirmed by the University of Iowa DNA facility.

6.2.3 Protein Purification

The DNA plasmid encoding for R7C was transformed into BL21(DE3) star competent cells. The colony was then grown in 100 μg/mL ampicillin LB broth at 37 °C to an OD_{600} level of 1.0. The culture was expressed for 2.5 hours at 28 °C in the presence of 1 mM IPTG. The cell pellet was re-suspended in lysis buffer consisting 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 50 mM DTT, 1 mM EDTA, and 1mM PMSF. After being French pressed at 16,000 psi, the inclusion bodies were re-dissolved in CB-1000 buffer. A total concentration of 0.1% PEI was added to the supernatant to precipitate negatively charged molecules and the target protein was then mostly precipitated by slowly stirring in ammonium sulfate powder until 25% saturated. The pellet was again dissolved in CB-1000 buffer and the solution was dialyzed gradually from 1000 to 50 mM NaCl with 10mM HEPES pH 8.0, 50 mM DTT, 1 M urea, and 1 mM $CaCl_2$. The desalted target protein was bound to a Heparin Sepharose 6 Fast Flow column (Amersham) and was eluted at 400 to 600 mM NaCl without urea. The 95% pure concentrated protein was further purified with a Sephadex G-50 Fine column (Amersham) in 10mM HEPES pH 7.0 and 50 mM NaCl buffer to 98% purity and dialyzed to metal-free condition. The concentration of filtered apoR7C was determined from A_{260} via UV-Vis spectrometer (ε_{280} = 7090 $M^{-1}cm^{-1}$). The purity and molecular weight of apoF2 was confirmed by silver stained SDS-PAGE and MADLI-TOF respectively. Myoglobin was used for the MADLI-TOF calibration.

6.2.4 Protein Dimerization

Since R7C was shown to be mostly monomeric in an oxygenated environment, chemically induced dimerization was performed. A ratio of 1:2 RC7 to 2,2-dipyridyl disulfide (PDS), a disulfide formation reagent, was prepared freshly and dissolved in a buffer consisting of 10 mM HEPES pH 8.0 and 50 mM NaCl.[149] The PDS-protein mixture was then incubated at 4 °C overnight. Native gels in both acidic and basic conditions were attempted with poor resolution. As a result, the mixture was subsequently run by SDS-PAGE in the absence 2-mercaptoethanol to determine the degree of dimerization with the introduction of PDS.

6.2.5 Tryptophan Fluorescence Titrations

The metal-binding affinity of R7C was studied by monitoring the fluorescence intensity of a single tryptophan (Trp) residue in each domain (W_{52}) as a function of metal concentration. The spectra were collected on an Aminco-Bowman Series 2 flourimeter. Each sample of 5 μM protein in 10mM HEPES pH 7.0, 10 mM DTT, and 50 mM NaCl was excited at 280 nm and the Trp emission was monitored at 350 nm as described previously.[100] R7C was presumably a monomer in the presence of DTT. The spectra were scanned from 400 to 300 nm in 1.0 nm interval at 25 °C. Metal was added to the protein solution and equilibrated for 5 minutes. The intensity at 350 nm as a function of metal concentration was plotted and described as a binding curve. Since no co-operative metal binding as a monomer was observed from the binding curve, the data was well-fit by a single exponential 1:1 association model with a linear non-least square algorithm [100,101].

6.2.6 Circular Dichroism Titrations

CD spectra of R7C titrating with Eu^{3+} were recorded using an Olis Cary-17 DS Conversion spectrophotometer at 20 ˚C under N_2 atmosphere in a 0.1-cm cell. Data were collected from 260 to 190 nm at 1.0 nm resolution and smoothed with Olis software (25

boxcar smoothing function). The spectra were corrected using a background scan of the buffer as a reference. The sample contained 25 μM of protein in 10 mM HEPES pH 7.0 and 50 mM NaCl titrating with increasing stoichiometric quantities of Eu^{3+} ranging from 0 to 4. Stoichiometric amounts of metal ions were added to protein as the chloride salt, from a 10 mM freshly prepared stock. A summary of the α-helicity of chimeras and myoglobin (standard) was compared and normalized based on the factor difference between the CD and PDB measurements of myoglobin.

6.2.7 DNA Plasmid Cleavage Assay

To demonstrate the catalytic activity of exposed EF-hands, the cleavage of supercoiled plasmid DNA complexes with Ca_2R7C was performed and observed via agarose gel electrophoresis. Samples of a 3.5 μM of Ca_2R7C_2 only, Ca_2R7C_2 with 100 mM DTT, Ca_2R7C_2 with 50 mM EDTA, and Ca_2R7C_2 with 8 M urea were freshly prepared before being added to 9 nM of pTYB1 plasmid in a 10 mM HEPES pH 7.0 and 50 mM NaCl buffer. The mixtures were incubated for 2 hours at 37 °C prior to extraction of DNA using a Geneclean kit (Bio-rad). The samples with loading dye were loaded onto a 1% agarose gel that was run for approximately 1 hour in 1 x TAE buffer at 120 V. The gels were visualized under UV-light and photographed with a Kodak imager.

6.3 Results and Discussion

6.3.1 Chimeric Design and DNA Cloning

The contact orientation of engrailed homeodomains with DNA major grooves was considered for the dimeric design by utilizing the known co-crystal structure of two homeodomains docking individually on a DNA helix (2HDD.pdb). In the head to head design, the DNA contact orientation of the second domain would be opposite to that of the head to tail design. As a result, the orientation of the protein contact to DNA major

groove is in reversed direction where the flexible N-terminals are at close proximity (Figure 6.1). A point mutation from Arg7 to Cys7 was considered due to the structural similarity of Arg and Cys residues which would therefore minimize structural alteration and their close proximity of the N-terminal arms in the minor groove as a dimer is essential for maintaining important DNA contacts. A head to head orientated dimer was thus formed through a disulfide bond.

Point mutagenesis was successfully performed to create cysteine residue near the N-terminus. The DNA construct was confirmed by the University of Iowa DNA facility and DNASTAR software following sequencing.

6.3.2 Protein Purification and Dimerization

The purity of R7C was followed by SDS-PAGE throughout the purification steps. The products of several purification steps including French pressed, PEI, dialysis, and cation-exchange Heparin column supernatant/eluent are shown in Figure 6.4. In order to determine the ratio of R7C dimers in solution, air-oxidized and chemically oxidized purified R7C was run on lane 5 and 6 in the absence of 2-mercaptoethanol. A very low percentage of R7C was dimeric in an oxygenated environment as indicated by the light band at ~15 kDa in lane 5 while about 30% of R7C was dimerized with the help of PDS overnight as shown in lane 6. However, the degree of dimerization was still not high enough for accurate activity measurements as a dimer.

Poor ability to dimerize is a common problem encountered in recombinant proteins especially as a non-specific "surface" residue. In biological systems, formation of disulfide bonds between proteins or enzymes are often induced by chaperones or other enzymes.[150] Several research groups have applied a 1:2 protein to PDS ratio to form inter- or intra-disulfide bond(s) in proteins.[149] Even though the percentage of dimerized R7C increased, the amount of dimers remained less than 50% of the total population after the introduction of PDS.

Nonetheless, the size of R7C was confirmed by MADLI-TOF as a monomer $(R7C+H)^+$ at 7703 *m/z* and a dimer of $(2R7C+H)^+$ at 15410 *m/z* (Figure 6.5). Because of the mass accuracy for MADLI-TOF is ± 0.1% for masses greater than 10 kDa, the calculated molecular weight of R7C for the monomer and dimer which is 7702 and 15403 g/mol respectively are within the error range.

6.3.3 Tryptophan Fluorescence Titrations

Trp fluorescence titrations were performed to determine the metal binding affinity of R7C by following the change of intensity of Trp fluorescence in the presence of metal ions. Since there is only one tryptophan, loaded at the beginning of the third helix turn of homeodomain, Trp fluorescence was able to be utilized for measuring the metal binding affinity on the Ca-binding loop that was inserted between helices 2 and 3. The emission at 350 nm was observed after excitation at 280 nm (Figure 6.6, inset). The fluorescence intensity decreased systematically as the concentration of metal ions in solution increased, suggesting that the Trp moved into a more hydrophobic environment upon metal addition. Typically, a less solvent exposed environment increases the Trp emission intensity, but a quenched intensity was observed in the folding study of homeodomains due to its unique water-indole interaction upon folding.[104] The fluorescence intensity observed at 350 nm was plotted against the equivalent of Eu^{3+} added to the solution (Figure 6.6). A linear non-least square algorithm fitting of 1:1 metal to protein per site associative model was then applied to calculate the overall dissociation constant (K_d) of metal binding to the dimer. The average K_d and R^2 value of R7C with Eu^{3+} were calculated from four trials which are 0.5 ± 0.2 μM and 1.0 ± 0.0 respectively. This K_d value is approximately 5 to 10 times stronger than that of C2 and F2 suggesting that there was an influence on metal binding affinity even with less than 50% dimeric presence in the solution.

Isolated EF-hand loops have shown to be highly cooperative and increase in calcium binding affinity upon dimerization.[27,108] It is possible that the freedom of rotation on the disulfide bond could bring both EF-hand loops in close proximity resulting an augmented metal binding affinity, or better folding of the domains due to packing against one another.

6.3.4 Circular Dichroism Titrations

CD spectroscopy was applied to investigate the global protein formation upon metal binding. An estimated α-helical secondary structure based on normalized molar ellipticity [θ_{222}] from CD data [110] showed that the apoR7C was 60% α-helical and the % helicity increased to 77% in the presence of excess Eu^{3+} (Figure 6.7). The α-helicity and shape of R7C are very similar to that of the chimeric monomer, C2, suggesting that the point mutation at the N-terminus and the disulfide bond dimerization did not alter the overall secondary structure of the chimera. A table of the α-helicity of our chimeric designs, a standard (highly helical myoglobin from equine heart) is summarized in Table 6.1. The helicity of the chimeric designs and control were measured under the same conditions while the result of the wild type homeodomain is obtained from a Sauer's paper[49] suggesting that the secondary structure of the chimeras could be higher than the calculation from CD data.

The CD spectra plotted against equivalents of Eu^{3+} in Figure 6.8 indicates that R7C requires two equivalents of ion to completely induce the secondary structure of R7C. This is consistent with the presence of two metal binding sites in a dimer and the measured metal binding affinity of R7C and is indicative of structural influence of dimerization.

6.3.5 DNA Plasmid Cleavage Assay

To investigate the DNA hydrolytic activity of a disulfide-formed dimer in comparison of a single-chain dimer and a monomer, a DNA plasmid cleavage assay of

monomeric/dimeric R7C was performed (Figure 6.9). The presence of a denaturing reagent, 8 M urea, is utilized as a control to ensure the activity is not caused by small molecules in the solution. The addition of EDTA to R7C showed that the reactivity was due to the presence of metal ions at the active sites of the metalloprotein. The introduction of a reducing reagent, DTT, forces R7C to be monomeric by reducing the disulfide bond. The partially dimerized R7C cleaved double stranded DNA similarly to the single-chain dimer, F2, while the activity of monomeric R7C behaved like the monomer, C2. This suggests that the dramatic change in cleavage activity is indeed due to the dimeric nature of the chimera and it is indicative of the advantages of dimeric design even with different linking orientations of the subunits. The large degree of design freedom and enhanced enzymatic activities based on dimeric designs set a milestone for understanding the benefits for basic protein-protein interactions.

6.4 Conclusions

Dimeric designs inspired by a structural-based chimeric monomer has been successfully designed, expressed, purified, and characterized. Not only did the metallodimers obtain both native proteins' unique functionalities (the ability to bind to DNA and to metal ions), but they also enhanced the double stand DNA hydrolytic activity dramatically. Comparison of two different dimeric designs, head to tail and head to head connection models, clearly demonstrates that both models have distinctly increased reactivities suggesting dimeric design is not limited to naturally occurring dimers and thus the possibilities for dimeric design can be numerous, for example, recombinant antibodies. Dimerization may also be used as an activation switch for a particular functionality may also easily be controlled as a dimer. Future studies on structural comparison of R7C and F2 in the association of DNA would be useful to understand whether different binding modes are presented and therefore be able to modified cleavage pattern accordingly.

These studies on chimeric dimers allow us to compare and understand the principles and potential of structural-based protein design. Moreover, these dimeric templates are good candidates for potential genome modification, such as DNA sequence targeting and gene disruption, with great specificity.

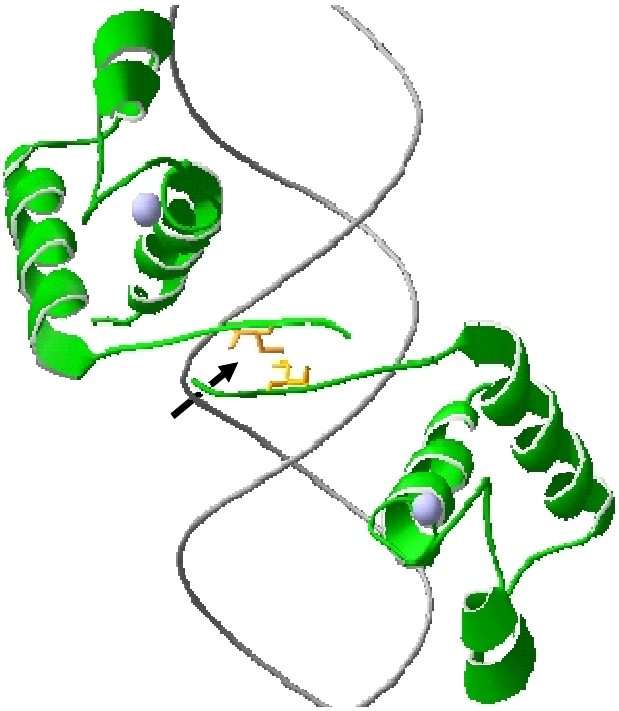


Figure 6.1 A chimeric dimer protein model, R7C (in green), is designed to form dimer with disulfide bond on the flexible N-terminal of C2 by mutation from Arg7 to Cys7 (in yellow and arrowed). The dimer is proposed to interact with the major groove of DNA (in grey) using Swiss PDB Viewer software.

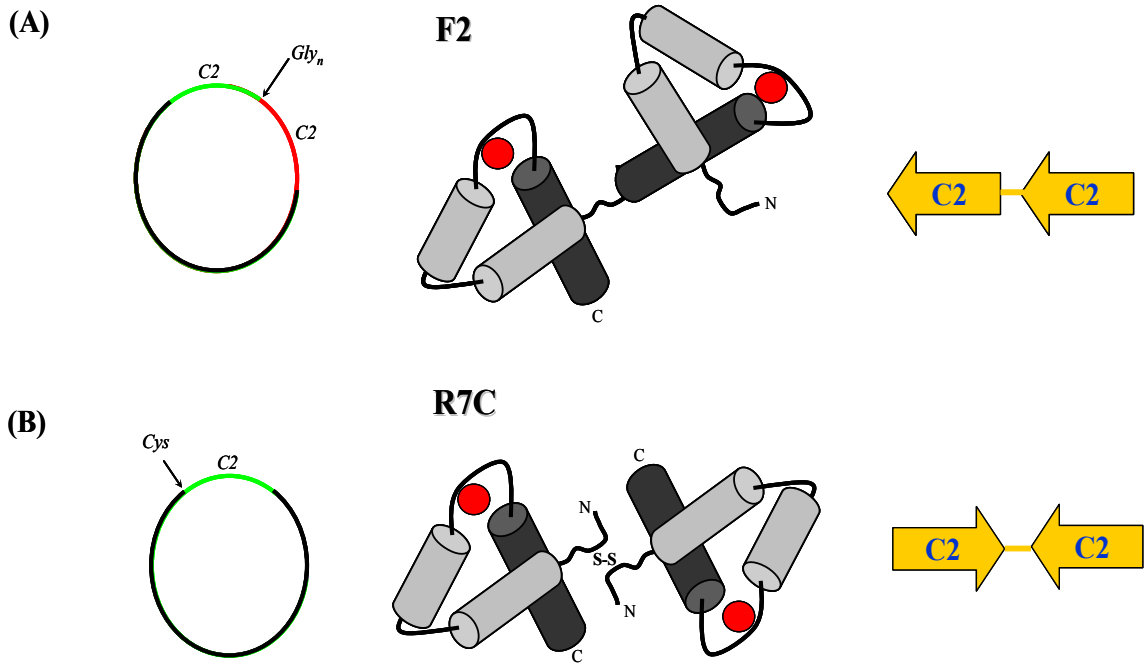


Figure 6.2 Schematic representation of our chimeric dimer design with two approaches. (A) Dimer was formed with a polypeptide linker between N- and C-termini. (B) R7C is dimerized at N-termini through disulfide bond at Cys7.

Wong05F: 5′-GAC GAG AAG CGT CCA TGC ACC GCG TTC TCC AGC-3′

Wong05R: 3′-CTG CTC TTC GCA GGT ACG TGG CGC AAG AGG TCG-5′

Figure 6.3 The primers, Wong05F and Wong 05R, were used for the site-directed point mutagenesis from Arg7 to Cys7 as underlined in order to create a site for dimerization through disulfide bond.

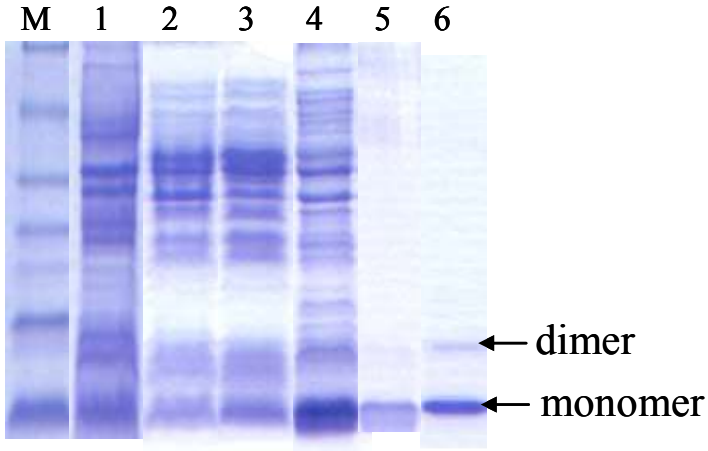


Figure 6.4 A 15% SDS-PAGE of R7C purification steps. Lane M is prestained broad range marker. Lane 1 shows induced protein. Lane 2 is the French pressed supernatant. The protein was further purified by PEI precipitation and remained in supernatant as shown in lane 3. The supernatant after dialyzed against low salt solution is shown in lane 4 and was eluted by Heparin column as shown in lane 5. Lane 6 showed the dimerization of R7C using 2,2-dipyridyl disulfide where lane 5 and 6 were run with SDS-PAGE without 2-mercaptoethanol.

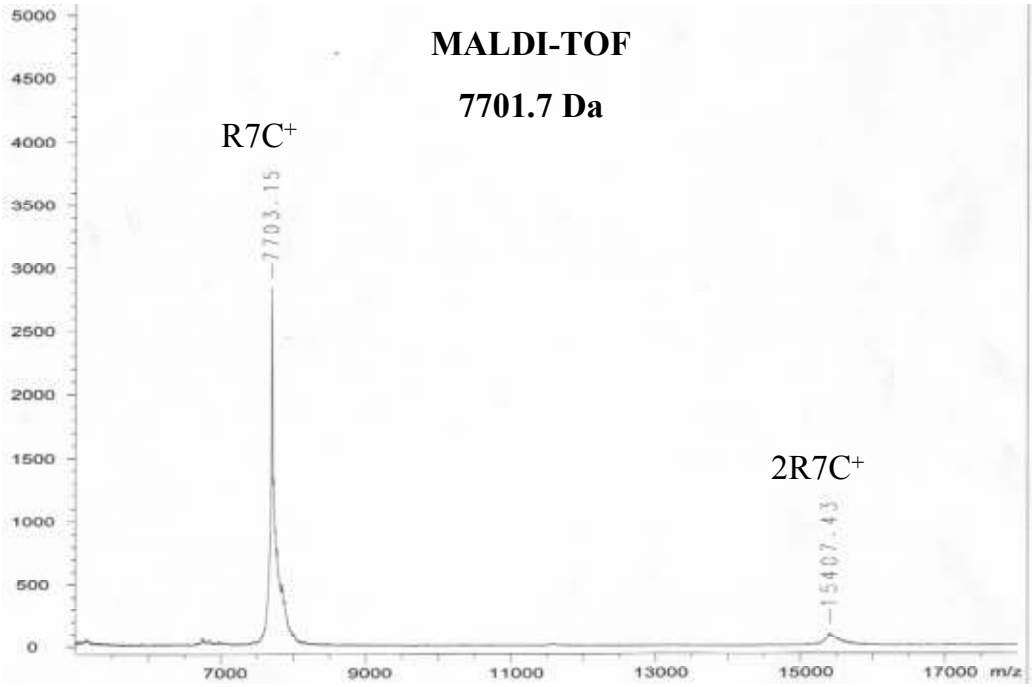


Figure 6.5 The MALDI-TOF spectra of purified R7C. The calculated molecular weight of R7C is 7702 Da via DNAstar software and the measured molecular weight of F2 is $R7C^+$ at 7703 *m/z* and dimerized $R7C^+$ at 15410 *m/z* which is within the accuracy range.

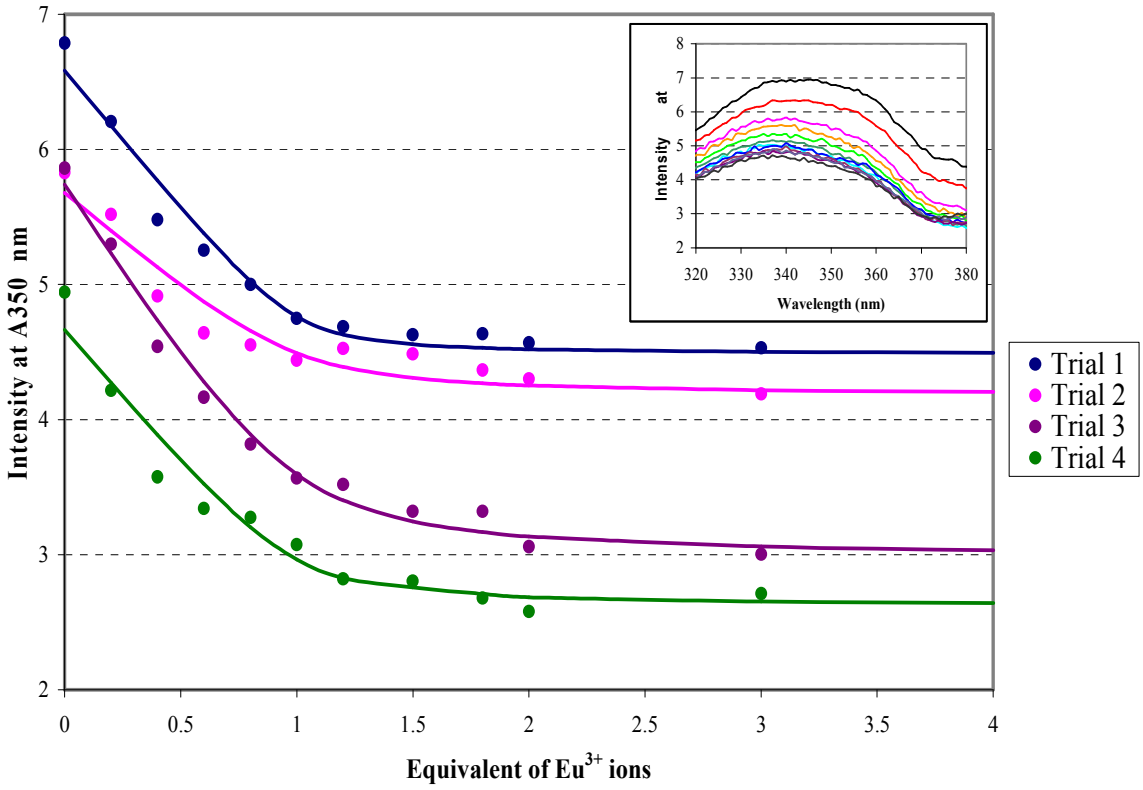


Figure 6.6 Fluorescence titrations of R7C with metal ions. The fluorescence intensity of Trp was followed at 350 nm (280 nm excitation) as a function of added metal. The solid lines represent a a linear non-least square algorithm fitting to a 1:1 association model, from which K_d was obtained. (*Inset*) Emission spectra showed throughout one titration, from 0 µM $EuCl_3$ (*Top*) to 40 µM $EuCl_3$ (*Bottom*).

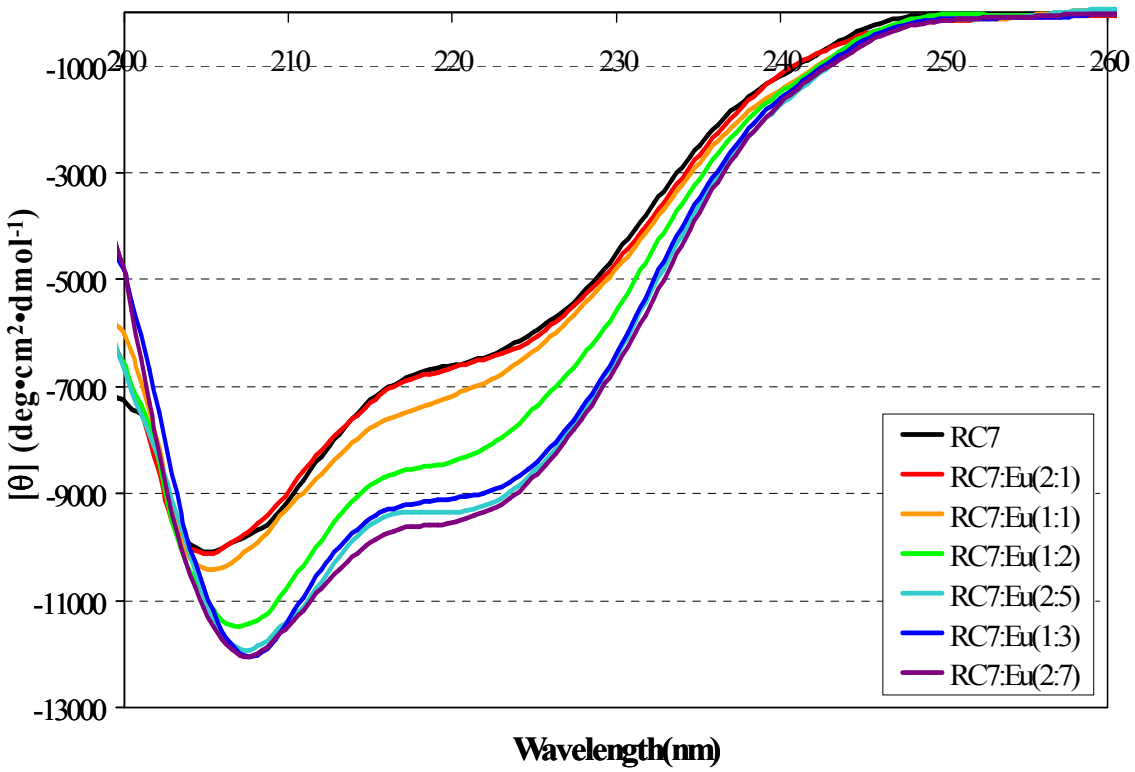


Figure 6.7 CD spectra of apoR7C (black) titrated with increasing amount of $EuCl_3$. Sample contained 15 μM F2 in 10 mM HEPES pH 7.0 and 50 mM NaCl.

Table 6.1 A summarized table of the α-helicity of our chimeric designs, a standard (myoglobin), and wt *engrailed* homeodomain. Since the CD measurement of myoglobin was less helical than that was projected from its PDB structure, the α-helicity was normalized based on the known α-helicity of myoglobin at pH 7.0.

α-helicity measured by	**wt HD**	**C2**	**R7C**	**F2**	**Myoglobin at pH 7.0 (standard)**
PDB/model	64	48	48	45	39
Normalized CD (apo form)	66	36	32	24	55
Normalized CD (Eu added)	---	45	43	31	---

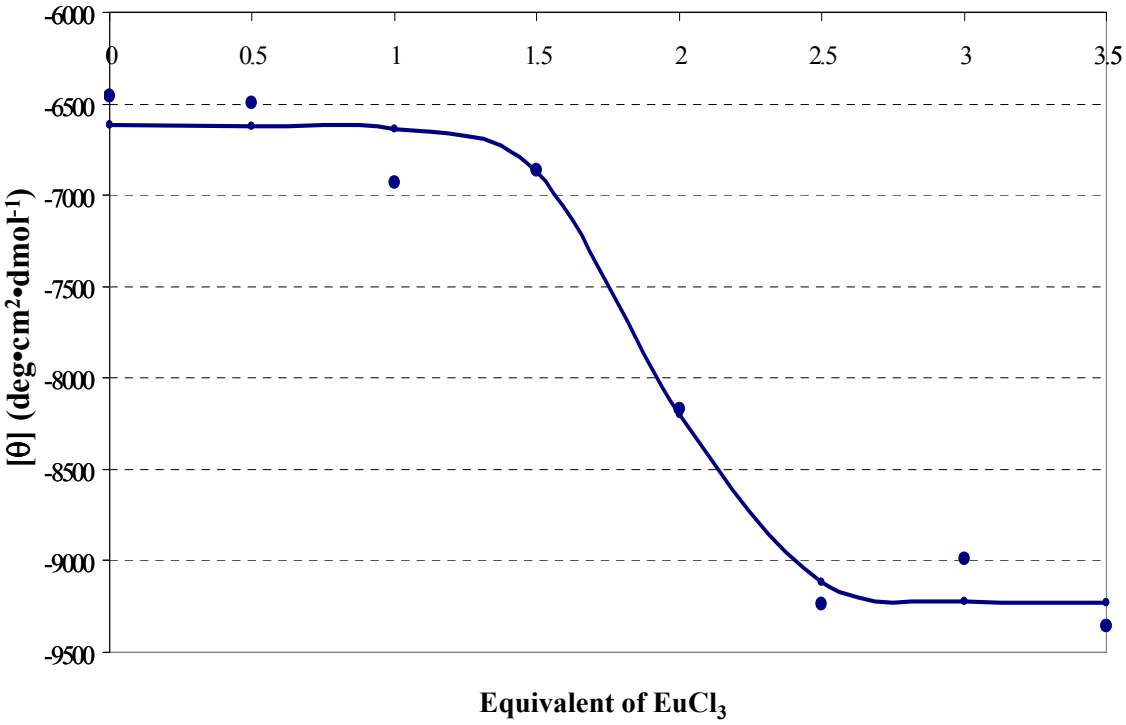


Figure 6.8 The mean residue ellipticity of EuR7C at 222 nm with the titration of increasing equivalent of $EuCl_3$.

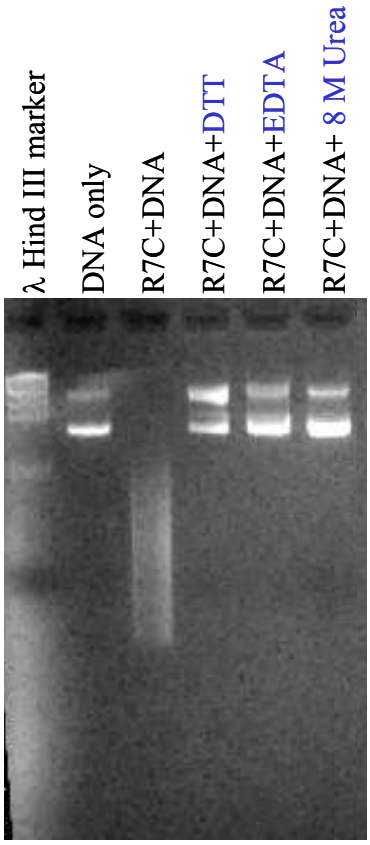


Figure 6.9 Agarose gel electrophoresis pTYB1 cleavage assay of EuR7C at different conditions. All R7C samples consisted of 10 mM HEPES pH 7.0 and 50 mM NaCl.

CHAPTER 7

CONCLUSIONS

A painting of the fearsome creature in Greek mythology, chimera, has finally been completed. The fire-breathing head of a goat (Ca-binding EF-hand loop), the body of a lion (DNA-binding *engrailed* homeodomain), and the tail of a dragon (new functionality with Ln^{3+} substitution) are packed into this chimeric dimer. Its power as a dimer has indeed been shown to be larger than its original components and its monomer combined.

The purity and the characterizations of F2 confirmed that the improved effect was solely due to the designed, folded dimer protein as increased purity enhanced its cleavage activity. The observations on significant metal binding affinity and cleavage activity of F2 in the presence of metal ions other than Ln^{3+}, such as Mg^{2+} and Ca^{2+}, were unexpected. Yet, these open a door for the understanding of the unique dimeric properties, as how protein-protein interactions may alter the metal selectivity on active sites. Even though the dissociation constant of metal ions was measured to be in the micromolar range, the persistent apoF2 cleavage results on the metal-dependent cleavage gel suggest that it is possible that metal ions could already be associated strongly with F2. Electrospray ionization mass spectrometry (ESI MS) would be a valid technique to determine the presence of metals in F2 in future studies.

The studies on F2-DNA interactions using EMSA and extensive numbers of ^{32}P-labeled DNA sequences provided us insights into DNA cleavage and binding selectivity of the dimer. Both DNA binding and cleavage consensus target sequence of F2 were notably different suggesting the interface for DNA binding does not promote hydrolytic activity on DNA backbone. These observations corresponded with those of C2 (monomer) as it rarely had any cleavage activity even when it selectively bound to a consensus DNA sequence. The consistent gap- and sequence-dependent cleavage

patterns of metallo-F2 demonstrated that the interactions were carried out by a dimer as a whole, as one domain behaved as an anchor to DNA on the first target site while the other recognized the second target site like an "opportunist". In addition, F2's hydrolytic behavior at the phosphodiester bond on a specific dsDNA sequence was shown to be in good agreement with the behavior of restriction endonucleases even though it also excised the entire nucleotide on a selected sequence. However, the structural information of the flexible and DNA cleaving dimer was very difficult to obtain after a number of attempts. This is at least in part due to its robust hydrolytic activity with metal ions, which eliminates the opportunity of co-crystallization with a natural DNA duplex. As a result, an attempt to understand such an intriguing cleavage mechanism using Swiss PDB Viewer, computational simulations, or co-crystallization with unnatural oligonucleotides is feasible to explore. In addition, a DNA sensor designed with fluorescence probes from Yi Lu's group could be modified and applied for kinetic studies of F2 on selective sequences.[151]

We have shown that there are advantages and disadvantage in designing a homodimer. Nonetheless, the overall information we have gained from this dimeric design has outweighed the difficulties on its structural determination, as this homodimer is the first example of an active and selective hydrolytic artificial nuclease based on the modular turn substitution design approach.

Both dimeric models, F2 and R7C, have shown to distinctly increase reactivities suggesting dimeric design is not limited to naturally occurring dimers. Thus, the possibility for dimeric design can be numerous, such as recombinant antibodies. These dimeric templates are also good candidates for potential genome modification, such as DNA sequence targeting and gene disruption, with great specificity which could be used for potential pharmaceutical enhancement.

APPENDIX

DNA CLONING AND PROTEIN PURIFICATION OF CROM

Introduction

Many DNA binding proteins present as oligomers perform various tasks in complicated biological pathways. One approach to structure-function studies of large complex proteins is to adopt a strategy of "divide and conquer".[152] However, the studies of individual subunits are often less meaningful as the protein-protein interaction between subunits improve functionality and stability of the whole complex.

The Cro repressor of bacteriophage lambda is among the simplest DNA binding proteins (Figure A.1). It is an important transcription regulatory protein that binds to operator DNA as a dimer consisting of two identical 66 residue subunits.[90] The dimeric interface involves a short antiparallel β-ribbon at the third β-strand. Its structure was one of the first sequence-specific DNA-binding proteins to be determined in three dimensions.[153] The DNA recognition site of λ Cro consists of a 17-base pair consensus operator which undergoes a large conformational change upon protein-DNA interaction.[154] Because of the simplicity of the rigid α-helical DNA recognition element and its well-characterized structure as a homodimer, it has become the prototypical helix-turn-helix DNA-binding protein for numerous spectroscopic and thermodynamic studies.[155-157]

In solution, λ Cro adopts predominately the monomic form in the absence of DNA, suggesting that the dimerization of λ Cro is weak with the equilibrium dissociation constant ranging from 0.3-3 μM.[90,158] However, Cro dimers are only required for half-saturation of operator DNA at as low as nanomolar concentrations.[90] Since the formation of dimer is at a nanomolar level, the binding affinity between the Cro dimer and DNA must be high. This range has been proposed to be on the order of 10^{-9}-10^{3} M.[158]

Our previously designed chimeric miniature peptides and monomeric protein have shown the ability to maintain their robust and selective metal-binding site with modest DNA binding selectivity.[11,66-68,87,159] Therefore, the selectivity properties of the designed chimera can be enhanced, as the length of the DNA recognition region becomes longer as a helix-turn-helix DNA binding homodimer. Lambda Cro is of interest to our dimeric design because of its nanomolar protein concentration requirement for dimerization in the presence of DNA and its small size which facilitates site-specific modification. We intend to utilize this well-studied dimer template to further understand and improve our *de novo* structural-based chimeric design.

Experimental Procedures

Chimeric Design

The design of the 76-residue chimeric protein, CroM, was based on overlays of the known crystal structures of DNA-binding λ Cro (5CRO)[160] and a Ca-binding EF-hand motif, which was oriented using Swiss PDB Viewer software. The HTH and EF-hand motifs consist of two helices at approximately right angles to each other allowing for modular substitution of the HTH motif of λ Cro at the α–α turn with an EF-hand metal binding loop. Several crystal coordinates of EF-hand proteins were downloaded from the Protein Data Bank (PDB): calmodulin (1OSA),[161,162]) parvalbumin (5PAL),[162] calbindin (1BCA)[163], and α-lactalbumin (1ALC)[164] to determine the best structurally aligned chimera. These four EF-hand metal binding loops were superimposed with λ Cro to determine the most suitable substitution. The EF-hand loop of calbindin gave the best structural fit with the α–α turn of λ Cro and was therefore the first model chosen for chimeric dimer studies.

DNA Cloning

Insertion of the metal binding loop into λ Cro

The λ Cro sequence in the pUC19 vector was prepared by Dr. Tim Kovacic. Before the insertion of the metal binding loop of calbindin, a silent mutation was made on Aln29 to create a restriction site of *AfeI* (Figure A.2) because the *AfeI* site could be utilized for restriction site insertion of the metal binding loop nearby. The silent mutation was performed via the Quickchange site-directed mutagenesis kit (Stratagene). The desalted primers for the mutation were ordered from DNA facility. The vector was amplified via PCR (Mastercycler PCR, Eppendorf) with 12 cycles of 95 °C for 30 s, 55 °C for 60 s, and 68 °C for 6 minutes. The quality of the PCR product was verified by 1% agarose gel. The mutated vector was purified using a midi-prep purification kit (Promega). The concentration of the mutated vector was determined by UV-Vis spectrometry observing at 260 nm. The mutated vector was sequenced by the University of Iowa DNA facility via universal primers, Marcy and Ben. The mutated vector was then cleaved at *BglII* and *AfeI* sites in order to allow the insertion of the metal binding loop of calbindin. The metal binding loop was ligated into the sites by incubating with T4 ligase at 12 °C overnight, and the ligated DNA was transformed into DH5α competent cells. The replicated DNA was then extracted from cells via mini-prep kit (QAlgene). The modified DNA sequence was confirmed by the University of Iowa DNA facility. Since the pET21a(+) vector (Invitrogen) has the T7 promoter for protein expression, the DNA sequence encoding the chimera was inserted between *EcoRI* and *HindIII* sites in the multi-cloning site of pET21a(+). The sequence of the ligated DNA was then analyzed by the University of Iowa DNA facility and was verified by the DNASTAR software. The verified DNA plasmid was transformed into a competent cell strain, BL21(DE3)star. The cell culture was then induced with 1 mM IPTG at various

cell optical densities (OD_{600}) and induction times in order to determine the optimal protein expression level.

Mutation of CroM

Primers, wong03F and wong03R, were designed for the point mutation from nucleotide CAA to CCA in order to change the encoded residue from Gln37 to Pro37 in the vector pET21a(+)(Figure A.3). This idea was inspired by the prolonged intracellular half-life (>3 hours) observed in wt Cro with the same mutation.[165] The introduction of proline results in enriched secondary structure and thus slows the protein degradation rate. With slower degradation, we hope to obtain more target proteins from the protein induction. A Quickchange site-directed mutagenesis kit (Stratagene) was used to perform the site mutation. The sequence of the mutated CroM, Q37P, was analyzed by the University of Iowa DNA facility and verified by DNASTAR software. The verified DNA plasmid was transformed into BL21(DE3)star competent cells. The cell culture was then induced with 1 mM IPTG at various cell densities and induction times in order to determine the optimal protein expression level.

Construction of CroM and Q37P in the pUCroRS vector

Since wt Cro has shown to be induced in large quantity with the pUCroRS vector, we can insert CroM and Q37P into pUCroRS in the hope to improve their expression levels. Both DNA encoding chimeras were inserted at the restriction enzyme sites of *HindIII* and *NcoI*. The CroM and Q37P that were inserted in pUCroRS were referred to as pUCroM and pUQ37P respectively. The insertion was analyzed by the University of Iowa DNA facility and verified by DNASTAR software. The verified DNA plasmid was transformed into a compatible cell strain, X90. The cell culture was then induced with 1 mM IPTG at various cell densities and induction times in order to determine the optimal protein expression level.

Protein Purification

The DNA plasmid encoding pUQ37P and pUCroM sequences were individually transformed into X90 competent cells. The colonies were then grown in 100 μg/mL Ampicillin LB broth at 37 °C to OD_{600} of 1.0. The culture was expressed for 2.5 hours at 28 °C in the presence of 1 mM of IPTG. The cell pellet was re-suspended in lysis buffer consisting of 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 5 mM DTT, 1 mM EDTA, and 1mM PMSF. After being French pressed at 16,000 psi, a total concentration of 0.1% PEI was added to the supernatant. After the PEI assay, the solution was then mostly precipitated by slowly stirring in 25% saturated ammonium sulfate powder. The pellet was re-dissolved in 6 M urea. The dissolved proteins were loaded on a Sephadex G-50 Fine column (Amersham) in 10 mM HEPES pH 7.0 and 50 mM NaCl buffer. However, the eluted column fractions containing the target protein were no longer visible on 15% SDS-PAGE. Several column fractions that seemed to contain low μM amounts of protein were sent to analyze the size of the protein by electrospray ionization-mass spectrometry (ESI-MS).

Results and Discussion

Chimeric Design

Several EF-hand proteins: calmodulin, parvalbumin, calbindin, and α-lactalbumin were superimposed onto the α–α turn of the HTH motif in a DNA binding protein, λ Cro. The crystal structures of proteins were obtained from the Swiss Protein Data Bank and a visual overlaying of the metal-binding loop of the EF-hand proteins with λ Cro was performed using Swiss PDB viewer software. The modular substitution with each EF-hand motif was overlaid with λ Cro to obtain the best possible fit. The calcium binding loop of calbindin gave the best structural fit with λ Cro at the α–α turn (Figure A.4). The amino acid modification on wt Cro is indicated in Figure A.4A. A three-dimensional view of the proposed chimeric design is shown in Figure A.4B as the metal-binding loop

of calbindin has been inserted between helix 2 and 3 of λ Cro with the minimum disruption on both parental structures resulting in a 76-residue chimera that can potentially form a dimer through antiparallel β-strands.

DNA cloning

Insertion of the metal binding loop into λ Cro

Before inserting the DNA sequence encoding the calbindin metal binding loop, a restriction enzyme site of *AfeI* was created by performing a silent mutation to aid the loop insertion into the λ Cro sequence. The 76-residue chimera with the combination of metal DNA binding, DNA binding, and dimerization domains is called CroM. The insertion was carried out in pUC19 which is a vector mainly used for DNA replication. Because of a lack of robust protein expression by pUC19 in the absence of a T7 promotor, the DNA sequence encoding CroM was inserted into the pET21a(+) plasmid which is commonly used for gene expression. The inserted sequence was analyzed and confirmed by the DNA facility and DNASTAR software respectively.

The pET21a(+) vector carries the *lac I* encoding for the *lac* repressor. The *lac* repressor operates by binding at the *lac* operon region which prevents the production of lactose as a carbon source in bacteria. The induction of IPTG (an analog of lactose) causes the repressor to change shape thus making it unable to bind to the *lac* operon. This enables the gene expression downstream of the T7 promoter and the target protein was thus expressed in a large quantity.

The wt Cro and CroM in the pET21a(+) vector were both induced by IPTG. The cell cultures were induced at various OD_{600} levels and lengths of time. The expressed protein was visualized via 15% SDS-PAGE which is a common method of protein mass determination (Figures A.5 and A.6). A protein (broad range) marker was used to evaluate the size of proteins. The last band of the marker is known to be 6.4 kDa which is close to the size of the target proteins (7 kDa). After an hour of induction with a vector

containing wt Cro, a band at ~7 kDa appeared suggesting the protein production of wt Cro (Figure A.5). In contrast, there was no visible band of CroM (~7 kDa) even after 3 hours of induction at various levels of OD_{600} (Figure A.6). Different induction temperatures, 28 °C and 37°C, were also applied but the expression of CroM remained elusive. This suggested that the modification done to wt Cro resulted in unstable protein.

Mutation of CroM to Q37P

A large literature search on the stability of mutant Cro revealed that approximately half of the 66 residue positions of Cro with mutations resulted in dramatically reduced protein expression. The destabilization of the fold in mutants generally results in extreme sensitivity to intracellular proteolysis. As a result, the corresponding mutants are severely degraded in cells.[166] Nonetheless, Saucer's group later demonstrated that the mutation of five different sites in λ Cro reduced the proteolytic hypersensitivity of the mutant via an antibody screen.[165] One of the mutations (Gln37 to Pro37) that was located at the N-terminal end of the third helix displayed no change in intracellular half-life for more than 3 hours. This position is especially important for our chimera design since the metal-binding loop was inserted at the turn between helices 2 and 3 of λ Cro. The rigidity provided by proline at that position may stabilize the fold of Cro and thus reduce its susceptibility for proteolysis. A Quickchange site-directed mutagenesis was therefore used to perform the site mutation changing from Gln37 to Pro37. The sequence of that mutant, Q37P, in the pET21a(+) vector was confirmed by the DNA facility.

The cell culture of Q37P was induced by IPTG at various levels of cell densities, lengths of expression time, and induction temperatures. Its protein expression level was followed by 15% SDS-PAGE. The Q37P mutant was expressed at the same level as CroM. (data not shown).

Construction of CroM and Q37P in pUCroRS vector

Dr. Mossing, who has published numerous studies on λ Cro suggested the use of a pUCroRS vector with a *tac* (hybrid *trp-lac*) promoter and a compatible cell strain, JM109, W3110 iQ or X90, with overexpression of *lac* repressor (*lac iQ*) for the Cro mutants expression [167] because λ Cro's over-expression is lethal to the host cells. The vector, pUCroRS, has a *tac* promoter that is repressed by the *lac* repressor. When there are more plasmids than repressors, the host RNA polymerase binds to the *tac* promoter and produces Cro mRNA. Therefore, cell strains that produce enough *lac* repressor to keep the *tac* promoter repressed are essential for protein expression of λ Cro.

To investigate the protein expression effectiveness of CroM and Q37P in pUCroRS, we inserted each mutant into the vector via the *HindIII* and *NcoI* restriction enzyme sites. When CroM and Q37P were individually inserted into pUCroRS, the resulting vectors were named pUCroM and pUQ37P respectively. Both insertions were verified by the University of Iowa DNA facility and DNASTAR software. Those DNA plasmids were then transformed into a cell strain containing *lac iQ*, X90.

The cell cultures of pUCroM and pUQ37P were induced with IPTG at an OD_{600} of 1.0. The protein expression was analyzed at various lengths of incubation time via SDS-PAGE (Figure A.7). A band at ~7kDa became visible for pUCroM and pUQ37P after 3 and 2 hours of induction, respectively, suggesting the vector and cell strain that produce excess *lac* repressors were essential for healthy expression of our Cro mutants.

Protein Purification

The cells of induced pUCroM and pUQ37P were collected after 3 hours of incubation. Since the vector did not have an affinity tag, several steps of purifications were required sequentially to completely purify the target protein. The target proteins were tracked with SDS-PAGE (Figure A.8). Proteins were released from the cell membrane by lysing with French Press (Lane FPS). The negatively charged proteins and

DNA were then precipitated with PEI while the target proteins remained in solution (Lane PEIS). The target proteins were salted out from smaller and membrane proteins with 25% saturated ammonium sulfate powder (data not shown). The pellets were re-dissolved with 6 M urea and then were loaded onto a size-exclusion column, Sephadex G-50 fine. Even though the large proteins eluted from the column, none of the eluted fractions had a visible band at ~7 kDa which is indicative of pUCroM's and pUQ37P's protein even in the case with silver straining, a more sensitive tool for protein detection. Several column fractions that seemed to contain low μM amounts of target proteins were sent to for analysis by ESI-MS, but no promising results were obtained (data not shown). Several purification modifications were made, such as ultracfiltration, cation (Sepharouse SF-fine), and anion exchange (Resource Q) columns, to improve the purity and reduce purification steps before their rapid degradation, though this effort was hampered by the stringent purification steps. Nonetheless, the protein was finally purified, although, the protein may be susceptible to degradation over time which will produce inconsistent results.

Conclusion

This is our first attempt to design and express a chimeric dimer based on a native dimeric HTH domain. The advantages for using Cro as the template are its structural simplicity, its extensively studied characterizations, and its high affinity with a long consensus sequence. It also only requires an antiparallel β-ribbon interface for dimerization. However, native proteins have evolved to achieve an optimal balance between stability and activity. Mutations can result in a decrease or increase in stability or activity because the structure-property relationship of proteins is yet to be understood.

The three-dimensional structure of the designed chimera was analyzed using Swiss PDB Viewer. The metal binding loop of calbindin was determined to be the best model among four EF-hand motif, though the loop modification on λ Cro disrupted its

fold stability which led to cell proteolysis. This indicated that the secondary structure of λ Cro dominated over its native function which is in favor for gene regulations. Since λ Cro is produced from the very first transcription step of λ early gene expression, this unique property can prevent errors from passing through the pathway.[168]

Our observations on the expression of Cro mutants shed light on the key principles of structural-based protein design. In order to engineer a promising protein, not only do its functionalities and modular similarity need to be carefully considered, but its structural stability and its susceptibility to modification are also important for the success of *de novo* design. But once a protein has been proven to be a functional and stable candidate for protein design, such as small zinc finger proteins, its application for medical innovations could be infinite.

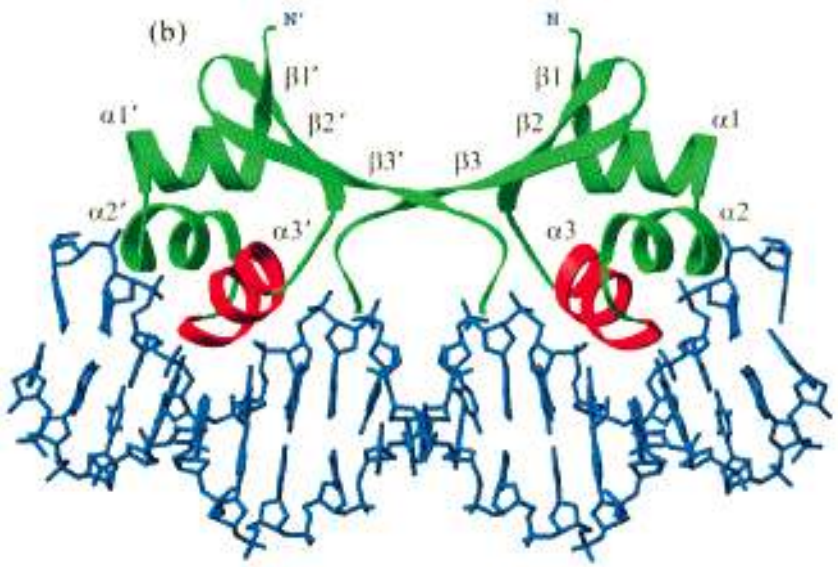


Figure A.1 A three-dimensional view of the λ Cro-DNA interaction. The third helix (in red) is bound to the major groove and bends the DNA duplex as a dimer. The third β-strand interacts to each other to form a dimer. The figure is adapted from reference 154.

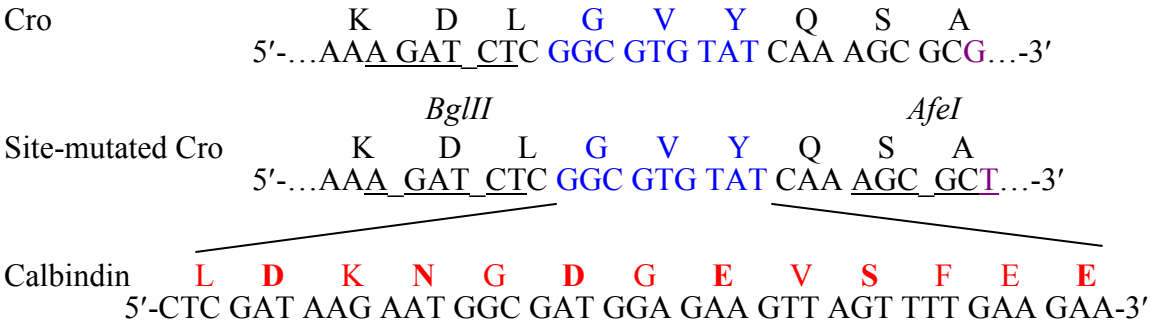


Figure A.2 The α-α turn (in blue) between helix 2 and 3 of wt Cro was replaced by the EF-hand Ca-binding loop (in red) of calbindin. A silent point mutation (in purple) on Aln29 (GCG to GCT) was preformed to create a restriction enzyme site of *AfeI* to aid the insertion of Ca-binding loop. The bold red letters show where calcium coordination occurs.

(A) Wong03F: 5′-GTT AGT TTT GAA GAA C**C**A AGC GCT ATC AAC AAG-3′

Wong03R: 3′-CAA TCA AAA CTT CTT G**G**T TCG CGA TAG TTG TTC-5′

(B) CroM: NH_3…TKTAKDLLDKNGDGEVSFEEQSAINKAI…COOH

Q37P: NH_3…TKTAKDLLDKNGDGEVSFEEPSAINKAI…COOH

Figure A.3 (A) Primers used for the site-mutagenesis from Q37 to P37. Wong03F and Wong03R are complimentary to each other. The bases significant in the nucleotide change from CAA to CCA are represented in bold green letters. (B) The different on amino acid sequence of CroM and Q37P is indicated in red.

(A)

1 10 20

NH$_3$-**MEQRITLKDYAMRFGQTKTA**

KDLLDKNGDGEVSFEE**QSAI**

NKAIHAGRKIFLTINADGSV

YAEEVKPFPSNKKTTA-*COOH*

(B)

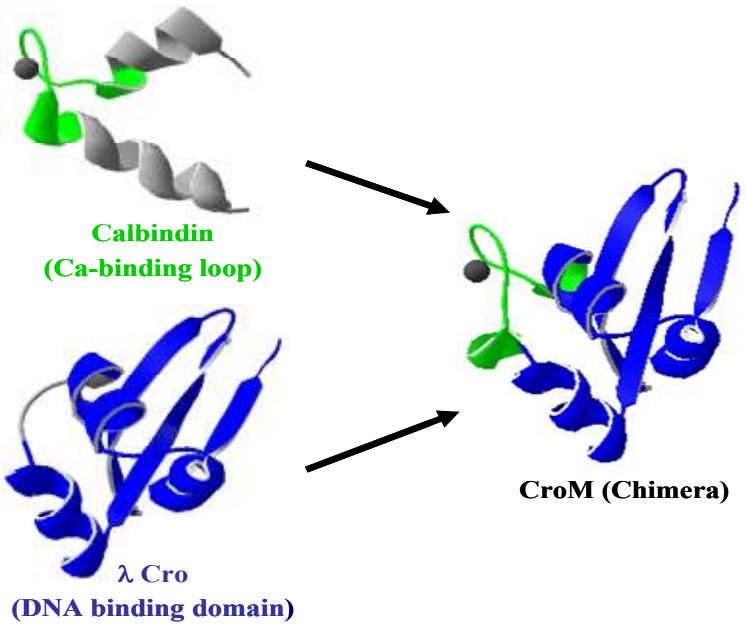


Figure A.4 (A) Amino acid sequence of the chimera, CroM. The parent protein sequence is indicated in bold (lambda Cro) and the inserted EF-hand Ca-binding loop from calbindin is underlined. (B) The EF-hand loop of calbindin (2BCA) and lambda Cro (5CRO) are shown in green/grey and blue/grey respectively. The overlay of the two motifs and the replacement of the α–α turn (in grey) on λ Cro created a chimera that could dimerize spontaneously, CroM.

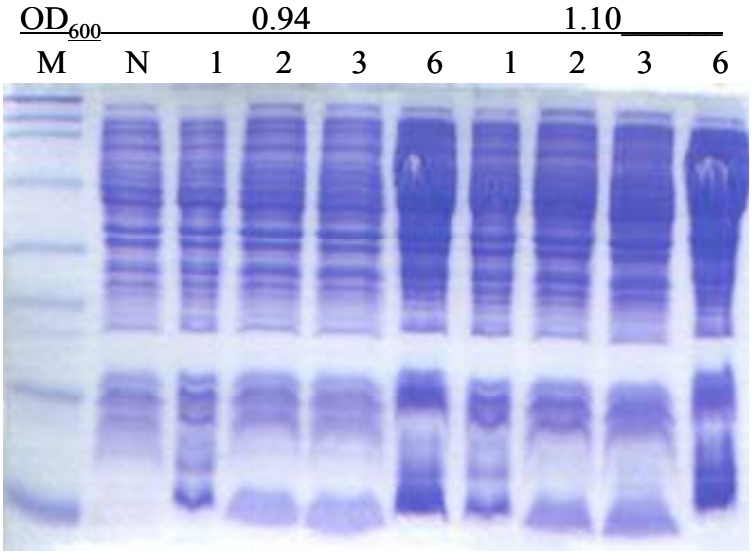


Figure A.5 A 15% SDS-PAGE of wt Cro's induction conditions. Lane M is pre-stained broad range marker. Lane N is the non-induced protein. The protein was induced at different OD_{600} levels. Lane 1-3 and 6 indicates the amount of induction time in hours.

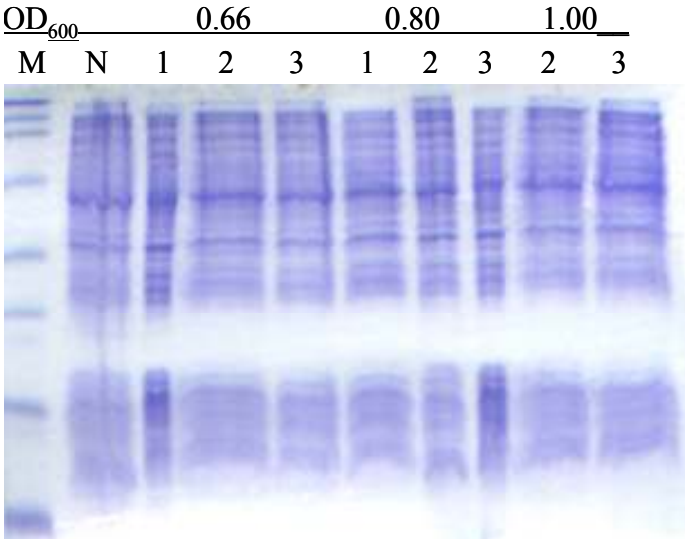


Figure A.6 A 15% SDS-PAGE of CroM's induction conditions. Lane M is pre-stained broad range marker. Lane N is the non-induced protein. The protein was induced at different OD_{600} levels. Lane 1-3 indicates the amount of induction time in hours.

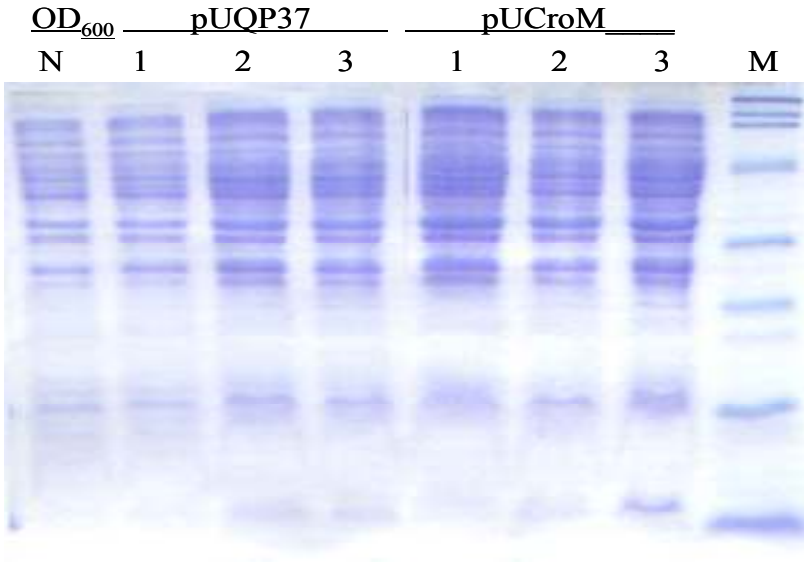


Figure A.7 A 15% SDS-PAGE of pUQP37's and pUCroM's induction conditions. Lane M is pre-stained broad range marker. Lane N is the non-induced protein. The protein was induced at OD_{600} level of 1.00. Lane 1-3 indicates the amount of induction time in hours.

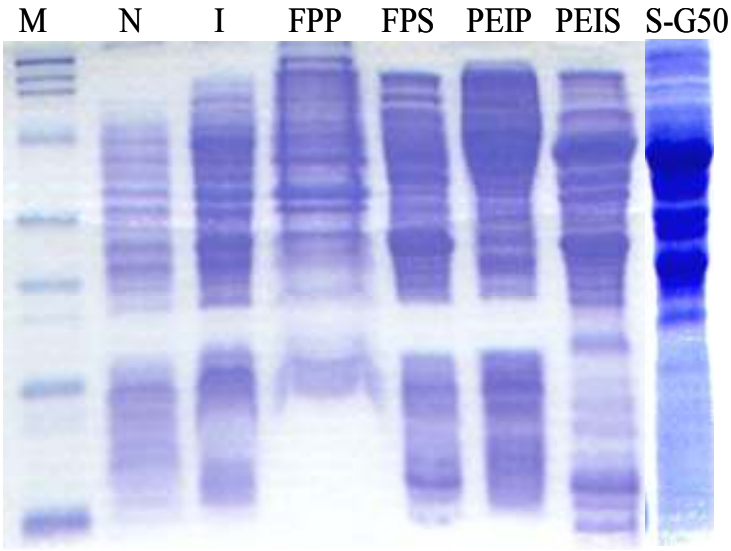


Figure A.8 A 15% SDS-PAGE of pUCroM's purification steps. Lane M is pre-stained broad range marker. Lane N is the non-induced protein while lane I is the protein induced for 2.5 hours. The cell culture was French pressed to yield supernatant (FPS) and pellet (FPP). The FPS fraction was further purified with PEI precipitation as shown in lane PEIP for pellet and PEIS for supernatant. The supernatant was then concentrated and loaded to a Sephadex G-50 fine (S-G50).

REFERENCES

(1) Carrell, R. W.; Gooptu, B. *Curr. Opin. Struct. Biol.* **1998**, *8*, 799-809.

(2) Kaim, W.; Shcewderski, B. *Bioinorganic Chemistry: Inorganic Elements in Chemistry of Life*; Wiley: New York, 1994.

(3) Lippard, S. J.; Berg, J. M. *Principles of Bioinorganic Chemistry*; University Science Books: Mill Valley, 1994.

(4) Thomson, A. J.; Gray, H. B. *Curr. Opin. Chem. Biol.* **1998**, *2*, 155-158.

(5) Doudna, J. A.; Cesh, T. R. *Nature (London)* **2002**, *418*, 222-228.

(6) Joyce, G. F. *Annu. Rev. Biochem.* **2004**, *73*, 791-836.

(7) Lu, Y. *Chem. Eur. J.* **2002**, *8*, 4588-4596.

(8) Pabo, C. O.; Peisach, E.; Grant, R. A. *Annu. Rev. Biochem.* **2001**, *70*, 313-340.

(9) Root, M. J.; Kay, M. S.; Kim, P. S. *Science* **2001**, *291*, 884-888.

(10) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.* **1999**, *68*, 779-819.

(11) Welch, J. T.; Sirish, M.; Lindstrom, K. M.; Franklin, S. J. *Inorg. Chem.* **2001**, *40*, 1982-1984.

(12) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364-1368.

(13) Mei, G.; Venere, A. D.; Rosato, N.; Finazzi-Agro, A. *FEBS Journal* **2005**, *272*, 16-27.

(14) Jeltsch, A.; Wenz, C.; Wende, W.; Selent, U.; Pingoud, A. *Trends in Biotech.* **1996**, *14*, 235-238.

(15) Cowan, J. A. *Nucl. Acids Mol. Biol.* **2004**, *14*, 339-358.

(16) Strynadka, N. C.; James, M. N. G. *Encyclopedia of Inorganic Chemistry*; 1st ed.; John Wiley and Sons, Ltd.: New York, 1993.

(17) McPhalen, C. A.; Strynadka, N. C.; James, M. N. *Adv. Protein Chem.* **1991**, *42*, 77-144.

(18) Falke, J. J.; Synder, E. E.; Thatcher, K. C.; Voertler, C. S. *Biochemistry* **1991**, *30*, 8690-8697.

(19) Kretsinger, R. H.; Nockolds, C. E. *J. Biol. Chem.* **1973**, *248*, 3313-3326.

(20) Capozzi, F.; Casadei, F.; Luchinat, C. *J. Biol. Inorg. Chem.* **2006**, *11*, 949-962.

(21) Linse, S.; Brodin, P.; Johansson, C.; Thulin, E.; Grundstrom, T.; Forsen, S. *Nature* **1988**, *335*, 651-652.

(22) Linse, S.; Forsen, S. *Adv. Second Messenger Phosphoprotein Res.* **1995**, *30*, 89-151.

(23) Linse, S.; Helmersson, A.; Forsen, S. *J. Biol. Chem.* **1991**, *266*, 8050-8054.

(24) Wu, X.; Reid, R. E. *Biochemistry* **1997**, *36*, 8649-8656.

(25) Wu, X.; Reid, R. E. *Biochemistry* **1997**, *36*, 3608-3616.

(26) Godzik, A.; Boguta, G. *Studia Biophys.* **1989**, *129*, 241-250.

(27) Wojcik, J.; Goral, J.; Pawlowski, K.; Bierzynski, A. *Biochemistry* **1997**, *36*, 680-687.

(28) Cates, M. S.; Teodoro, M. L.; Phillips, J., George N. *Biophys. J* **2002**, *82*, 1133-1146.

(29) Drake, S. K.; Falke, J. J. *Biochemistry* **1996**, *35*, 1753-1760.

(30) Drake, S. K.; Lee, K. L.; Falke, J. J. *Biochemistry* **1996**, *35*, 6697-6705.

(31) Lim, S.; Franklin, S. J. *Cell. Mol. Life Sci.* **2004**, *61*, 2184-2188.

(32) Huheey, J. E.; Keiter, E. A.; Keiter, R. L. *Inorganic Chemistry*; 4th ed.; HarperCollins College Publishers: New York, 1993.

(33) Morrow, J. R.; Buttrey, L. A.; Shelton, V. M.; Berback, K. A. *J. Am. Chem. Soc.* **1992**, *114*, 1903-1905.

(34) Bari, L. D.; Salvadori, P. *Coord. Chem. Rev.* **2005**, *249*, 2854-2879.

(35) Kremer, C.; Torres, J.; Dominguez, S.; Mederos, A. *Coord. Chem. Rev.* **2005**, *249*, 567-590.

(36) Maniccia, A. W.; Yang, W.; Li, S.-y.; Johnson, J. A.; Yang, J. J. *Biochemistry* **2006**, *45*, 5848-5856.

(37) Snyder, E. E.; Buoscio, B. W.; Falke, J. J. *Biochemistry* **1990**, *29*, 3937-3943.

(38) Bruno, J.; Horrocks Jr., W. D.; Zauhar, R. J. *Biochemistry* **1992**, *31*, 7016-7026.

(39) Scott, M. P.; Tamkun, J. W.; Hartzell, G. W. *Biochim. Biophys. Acta* **1989**, *989*, 25-48.

(40) Gehring, W. J.; Affolter, M.; Burglin, T. *Annu. Rev. Biochem.* **1994**, *63*, 487-526.

(41) Laughon, A. *Biochemistry* **1991**, *30*, 11357-11367.

(42) Qian, Y.; Billeter, M.; Otting, G.; Muller, M.; Gehring, W. J.; Wuthrich, K. *Cell* **1989**, *59*, 573-580.

(43) Kissinger, C.; Liu, B.; Martin-Blanco, E.; Kornberg, T.; Pabo, C. O. *Cell* **1990**, *63*, 579-590.

(44) Wolberger, C.; Vershon, A.; Liu, B.; Johnson, J. A.; Pabo, C. O. *Cell* **1991**, *67*, 517-528.

(45) Freemont, P. S.; Lane, A. N.; Sanderson, M. R. *Biochem. J.* **1991**, *278*, 1-23.

(46) Brennan, R. G.; Matthew, B. W. *Trends Biochem. Sci.* **1989**, *14*, 286-290.

(47) Patikoglou, G.; Burley, S. K. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 289-325.

(48) Gehring, W. J.; Qian, Y. Q.; Billeter, M.; Furukubotokunaga, K.; Schier, A. F.; Resendezperez, D.; Affolter, M.; Otting, G.; Wuthrich, K. *Cell* **1994**, *78*, 211-223.

(49) Ades, S. E.; Sauer, R. T. *Biochemistry* **1994**, *33*, 9187-9194.

(50) Draganescu, A.; Tullium, T. D. *J. Mol. Biol.* **1998**, *276*, 529-536.

(51) Fraenkel, E.; Rould, M. A.; Chambers, K. A.; Pabo, C. O. *J. Mol. Biol.* **1998**, *284*, 351-361.

(52) Garcia-Bellido, A.; Santamaria, P. *Genetics* **1972**, *72*, 87-104.

(53) Dragan, A. I.; Li, Z.; Makeyeva, E. N.; Milgotina, E. I.; Liu, Y.; Crane-Robinson, C.; Privalov, P. L. *Biochemistry* **2006**, *45*, 141-151.

(54) Wang, L.-H.; Chmelik, R.; Nirenberg, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12721-12726.

(55) Treisman, J.; Gonczy, P.; Vashishtha, M.; Harris, M.; Desplan, C. *Cell* **1989**, *59*, 553-562.

(56) Hanes, S. D.; Brent, R. *Science* **1991**, *251*, 426-430.

(57) Daniele, D.; Gerard, C.; Alain, P. *Trends Cell Biol.* **1998**, *8*, 84-87.

(58) Cosgaya, J. M.; Aranda, A.; Cruces, J.; Martin-Blanco, E. *J. Cell Sci.* **1998**, *11 (Pt 16)*, 2377-2384.

(59) Cowan, J. A. *Chem. Rev.* **1998**, *98*, 1067-1088.

(60) Wilcox, D. E. *Chem. Rev.* **1996**, *96*, 2435-2458.

(61) Sreedhara, A.; Cowan, J. A. *J. Biol. Inorg. Chem.* **2001**, *6*, 337-347.

(62) Sreedhara, A.; Freed, J. D.; Cowan, J. A. *J. Am. Chem. Soc.* **2000**, *122*, 8814-8824.

(63) Rammo, J.; Hettich, R.; Roigk, A.; Schneider, H.-J. *Chem. Commun.* **1996**, 105-107.

(64) Branum, M. E.; Que, L. *J. Biol. Inorg. Chem.* **1999**, *4*, 593-600.

(65) Komiyama, M. J. *J. Biochem. (Tokyo)* **1995**, *118*, 665-670.

(66) Kim, Y.; Welch, J. T.; Lindstrom, K. M.; Franklin, S. J. *J. Biol. Inorg. Chem.* **2001**, *6*, 173-181.

(67) Franklin, S. J.; Lim, S. *Cell Mol. Life Sci.* **2004**, *61*, 2184-2188.

(68) Welch, J. T.; Kearney, W. R.; Franklin, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3725-3730.

(69) Goodsell, D. S.; Olson, A. J. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105-153.

(70) Cowan, J. A. *Chem. Rev.(Washington, DC)* **1998**, *98*, 1067-1087.

(71) Chevalier, B.; Sussman, D.; Otis, C.; Noel, A.-J.; Turmel, M.; Lemieux, C.; Stephens, K.; Monnat, R. J.; Stoddard, B. L. *Biochemistry* **2004**, *43*, 14015-14026.

(72) Sharp, K.; Fine, R.; Honig, B. *Science* **1987**, *236*, 1460-1464.

(73) Perutz, M. *Rev. Biophys.* **1989**, *22*, 139-237.

(74) McLure, K. G.; Lee, P. W. K. *EMBO J.* **1998**, *17*, 3342-3350.

(75) Williams, N. H.; Cheung, W.; Chin, J. *J. Am. Chem. Soc.* **1998**, *120*, 8079-1451.

(76) Zhu, B.; Zhao, D.-Q.; Ni, J.-N.; Zeng, Q.-H.; Huang, B.-Q.; Wang, Z.-L. *Inorg. Chem. Commun.* **1999**, *2*, 351-353.

(77) Lombardi, A.; Summa, C. M.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6298-6305.

(78) Walsh, S. T. R.; Cheng, H.; Bryson, J. W.; Roder, H.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5486-5491.

(79) Komiyama, M. J. *Biochemistry* **1995**, *118*, 665-670.

(80) Takenaka, S.; Ihara, T.; Tagaki, M. J. *J. Mol. Recognit.* **1990**, *3*, 156-162.

(81) Kovacic, R. T.; Welch, J. T.; Franklin, S. J. *J. Am. Chem. Soc.* **2003**, *125*, 6656-6662.

(82) Lim, S.; Franklin, S. J. *Protein Sci* **2006**, *15*, 2159-2165.

(83) Morii, T.; Yaname, J.; Aizawa, Y.; Makino, K.; Sugiura, Y. *J. Am. Chem. Soc.* **1996**, *118*, 10011-10017.

(84) Talanian, R. V.; McKnight, C. J.; kim, P. S. *Science* **1990**, *249*, 769-771.

(85) Wilson, D. S.; Guenther, B.; Desplan, C.; Kuriyan, J. *Cell* **1995**, *82*, 709-719.

(86) Xu, W.; Rould, M. A.; Jun, S.; Desplan, C.; Pabo, C. O. *Cell* **1995**, *80*, 639-650.

(87) Wong-Deyrup, S. W.; Kim, Y.; Franklin, S. J. *J. Biol. Inorg. Chem.* **2006**, *11*, 17-25.

(88) Gitehson, G. I.; Griko, Y. V.; Kurochkin, A. V.; Rogov, V. V.; Kutyshenko, V. P.; Kirpichnikov, M. P.; Provalov, P. L. *FEBS Lett.* **1991**, *289*, 201-204.

(89) Takeda, Y.; Flkmanis, A.; Echols, H. *J. Biol. Chem.* **1997**, *252*, 6177-6183.

(90) Jana, R.; Hazbun, T. R.; Mollah, A. K. M. M.; Mossing, M. C. *J. Mol. Biol.* **1997**, *273*, 402-416.

(91) Czerny, T.; Schaffner, G.; Busslinger, M. *Gene & Dev.* **1993**, *7*, 2048-2061.

(92) Huston, J. S.; Levinson, D.; Mudgett-Hunter, M.; Tai, M. S.; Novotny, J.; Margolies, M. N.; Ridge, R. J.; Bruccoleri, R. E.; Haber, E.; Crea, R.; Oppermann, H. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5879-5883.

(93) Simoncsits, A.; Chen, J.; Percipalle, P.; Wang, S.; Toro, I.; Pongor, S. *J. Mol. Biol.* **1997**, *267*, 118-131.

(94) Robinson, C. R.; Sauer, R. T. *Biochemistry* **1996**, *35*, 109-116.

(95) Robinson, C. R.; Sauer, R. T. *Biochemistry* **1996**, *35*, 13878-13884.

(96) Tucker-Kellogg, L.; Rould, M. A.; Chambers, K. A.; Ades, S. E.; Sauer, R. T.; Pabo, C. O. *Structure* **1997**, *5*, 1047-1054.

(97) Wilson, M. A.; Brunger, A. T. *J. Mol. Biol.* **2000**, *301*, 1237-1256.

(98) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714-2723.

(99) Jana, R.; Hazbun, T. R.; Fields, J. D.; Mossing, M. C. *Biochemistry* **1998**, *37*, 6446-6455.

(100) Sirish, M.; Schneider, H.-J. *Chem. Commun.* **1999**, 907-908.

(101) Wilcox, C. S. *Frontiers in Supramolecular Organic Chemistry and Photochemistry*; VCH: Weinheim, Germany, 1991.

(102) David Peretz; Scott, M. R.; R., D. G.; Williamson, A.; Burton, D. R.; Cohen, F. E.; Prusiner, S. B. *Protein Sci* **2001**, *10*, 854-863.

(103) Teles, R. C.; Calderon, L. d. A.; Medrano, F. J.; Barbosa, J. A. R. G.; Guimaraes, B. G.; Santoro, M. M.; de Freitas, S. M. *Biophys. J* **2005**, *88*, 3509-3517.

(104) Stollar, E. J.; Gelpi, J. L.; Velankar, S.; Golovin, A.; Orozco, M.; Luisi, B. F. *Protein* **2004**, *57*, 1-8.

(105) Marsden, B. J.; Shaw, G. S.; Sykes, B. D. *Biochem. Cell. Biol.* **1990**, *68*, 587-601.

(106) Procyshyn, R. M.; Reid, R. E. *J. Biol. Chem.* **1994**, *269*, 1641-1647.

(107) Gariepy, J.; Sykes, B. D.; Hodges, R. S. *Biochemistry* **1983**, *22*, 1765-1772.

(108) Siedlecka, M.; Goch, G.; Ejchart, A.; Sticht, H.; Bierzynski, A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 903-908.

(109) Lim, S. *Ph.D. Thesis* **2006**.

(110) Chen, Y.-H.; Yang, J. T.; Martinez, H. M. *Biochemistry* **1972**, *11*, 4120-4131.

(111) Peretz, D.; Scott, M. R.; R., D. G.; Williamson, A.; Burton, D. R.; Cohen, F. E.; Prusiner, S. B. *Protein Sci.* **2001**, *10*, 854-863.

(112) Sheilds, S. B.; Franklin, S. J. *Biochemistry* **2004**, *43*, 16986-16091.

(113) Bernadou, J.; Pratviel, G.; Bennis, F.; Girardet, M.; Meumier, B. *Biochemistry* **1989**, *28*, 7268.

(114) Kim, Y.; Franklin, S. J. *Inorganica Chimica Acta* **2002**, *341*, 107-112.

(115) Acharya, A. S.; Roy, K. B. *Biochem. Biophys. Res. Commun.* **2001**, *287*, 153-159.

(116) Erskine, S. G.; Baldwin, G. S.; Halford, S. E. *Biochemistry* **1997**, *36*, 7567-7576.

(117) Buy, C.; Girault, G.; Zimmermann, J. L. *Biochemistry* **1996**, *35*, 9880-9891.

(118) Maguire, M. E.; Cowan, J. A. *BioMetals* **2002**, *15*, 203-210.

(119) Komiyama, M. J. *Chem. Commun.* **1999**, *16*, 1443-1444.

(120) Ott, R.; Kramer, R. *Appl. Microbiol. Biotechnol.* **1999**, *52*, 761-767.

(121) Franklin, S. J. *Curr. Opinion Chem. Biol.* **2001**, *5*, 201-208.

(122) Chin, J.; Banaszcayk, M.; Jubian, V.; Zou, X. *J. Am. Chem. Soc.* **1989**, *111*, 186-190.

(123) Phillips, K.; Luisi, B. F. *J. Mol. Biol.* **2000**, *302*, 1023-1039.

(124) Kornberg, T. *J. Biol. Chem.* **1993**, *268*, 26813-26816.

(125) Simon, M. D.; Sato, K.; Weiss, G. A.; Shokat, K. M. *Nucl. Acids Res.* **2004**, *32*, 3623-3631.

(126) Fraenkel, E.; Pabo, C. O. *Natl. Struct. Biol.* **1998**, *5*, 692-697.

(127) Lim, S.; Franklin, S. J. *Protein Sci.* **2006**, *15*, 2159-2165.

(128) Maxam, A. M.; Gilbert, W. *65* **1980**.

(129) Sambrook, J.; Russell, D. W. In *Molecular Cloning: a laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: New York, 2000; Vol. 3, pp 17.11-17.12.

(130) Sambrook, J.; Russell, D. W. In *Molecular Cloning: a laboratory manual*; Cold Spring Harbor Laboratory Press: New York, 2000; Vol. 2, pp 12.40-12.44.

(131) Iwahara, J.; Zweckstetter, M.; Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15062-15067.

(132) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. *Biopolymers* **1970**, *9*, 1059-1077.

(133) Kostrewa, D.; Winkler, K. F. *Biochemistry* **1995**, *34*, 683-696.

(134) Chin, J. W.; Schepartz, A. *J. Am. Chem. Soc.* **2001**, *123*, 2929-2930.

(135) Friesen, W. J.; Darby, M. K. *J. Biol. Chem.* **2001**, *276*, 1968-1973.

(136) Zondlo, N. J.; Schepartz, A. *J. Am. Chem. Soc.* **1999**, *121*, 6938-6939.

(137) Rao, S. T.; Wu, S.; Satyshur, K. A.; Ling, K.-Y.; Kung, C.; Sundaralingam, M. *Protein Sci* **1993**, *2*, 436.

(138) Falke, J. J.; Drake, S. K.; Hazard, A. L.; Peersen, O. B. *Q Rev. Biophys.* **1994**, *27*, 219-290.

(139) Morris, J. F.; Hromas, R.; Rauscher, F. J. I. *Mol. Cell Biol.* **1994**, *14*, 1786-1795.

(140) Thijs, G.; Lescot, M.; Marchal, K.; Rombauts, S.; Moor, B. D.; Rouze, P.; Moreau, Y. *Bioinformatics* **2001**, *17*, 1113-1122.

(141) Liu, X. S.; Brutlag, D. L.; Liu, J. S. *Nat. Biotechnol.* **2002**, *20*, 835-839.

(142) Kelly, S. M.; Jess, T. J.; Price, N. C. *Biochem. Biophys. Acta* **2005**, *1751*, 119-139.

(143) Ledneva, R. K.; Alexeevskii, A. V.; Vasilev, S. A.; Spirin, S. A.; Karyagina, A. S. *Mol. Biol.* **2001**, *35*, 647-659.

(144) Tsao, D. H. H.; Gruschus, J. M.; Wang, L.-H.; Nirenberg, M.; Ferretti, J. A. *Biochemistry* **1994**, *33*, 15053-15060.

(145) Stollar, E. J.; Mayor, U.; Lovell, S. C.; Federici, L.; Freund, S. M. V.; Fersht, A. R.; Luisi, B. F. *J. Biol. Chem.* **2003**, *278*, 43699-43708.

(146) Jabet, C.; Gitti, R.; Summers, M. F.; Wolberger, C. *J. Mol. Biol.* **1999**, *291*, 521-530.

(147) Luscombe, N. M.; Austin, S. E.; Berman, H. M.; Thornton, J. M. *Genome Biol.* **2000**, *1*, 1-37.

(148) Konig, P.; Giraldo, R.; Chapman, L.; Rhodes, D. *Cell* **1996**, *85*, 125-136.

(149) Korodi, I.; Asboth, B.; Polgar, L. *Biochemistry* **1986**, *25*, 6895-6900.

(150) Lauber, T.; Schulz, A.; Rosch, P.; Marx, U. C. *Biochemistry* **2004**, *43*, 10050-10057.

(151) Liu, J.; Brown, A. K.; Meng, X.; Cropek, D. M.; Istok, J. D.; Watson, D. B.; Lu, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2056-2061.

(152) Albright, R. A.; Mossing, M. C.; Matthew, B. W. *biochemistry* **1996**, *35*, 735-742.

(153) Anderson, W. F.; Ohlendorf, D. H.; Takeda, Y.; Matthew, B. W. *Nature* **1981**, *290*, 754-758.

(154) Albright, R. A.; Matthew, B. W. *J. Mol. Biol.* **1998**, *280*, 137-151.

(155) Anderson, W. F.; Takeda, Y.; Ohlendorf, D. H.; Matthew, B. W. *J. Mol. Biol.* **1982**, *159*, 745-751.

(156) Brennan, R. G.; Matthew, B. W. *J. Biol. Chem.* **1989**, *264*, 1903-1906.

(157) Harrison, R. *Nature* **1991**, *353*, 715-719.

(158) Darling, P. J.; Holt, J. M.; Ackers, G. K. *Biochemistry* **2000**, *39*, 11500-11507.

(159) Franklin, S. J.; Welch, J. T. *Comment Inorg. Chem.* **2005**, *26*, 127-164.

(160) Ohlendorf, D. H.; Tronrud, D. E.; Matthew, B. W. *J. Mol. Biol.* **1998**, *280*, 129-136.

(161) Chattpadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. *J. Mol. Biol.* **1992**, *228*, 1177-1192.

(162) Roquet, F.; Declercq, J. P.; Tinant, B.; Rambaud, J.; Parello, J. *J. Mol. Biol.* **1992**, *223*, 705-720.

(163) Kordel, J.; Pearlman, D. A.; Chazin, W. J. *J. Biomol. NMR* **1997**, *10*, 231-243.

(164) Acharya, K. R.; Stuart, D. I.; Walker, N. P.; Lewis, M.; Phillips, D. C. *J. Mol. Biol.* **1989**, *208*, 99-127.

(165) Pakula, A. A.; Sauer, R. T. *Proteins: Structure, Function, and Genetics* **1989**, *5*, 202-210.

(166) Pakula, A. A.; Young, V. B.; Sauer, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8829-8833.

(167) Mossing, M. C.; Sauer, R. T. *Science* **1990**, *250*, 1712-1715.

(168) Satumba, J. W.; Mossing, M. C. *Biochemistry* **2002**, *41*, 14216-14224.