UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Assessing a potential correlation between protein stability and function of the HMG domain of Lymphoid Enhancer Binding Factor-1

Permalink

https://escholarship.org/uc/item/5xj829hb

Author

Ly, Youly

Publication Date

2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

SAN DIEGO STATE UNIVERSITY

Assessing a potential correlation between protein stability and function of the HMG domain of Lymphoid Enhancer Binding Factor-1

A Dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Chemistry

by

Youly Ly

Committee in Charge:

University of California, San Diego

Professor John Crowell
Professor Elizabeth Komives

San Diego State University

Professor John Love, Chair
Professor Terry Frey
Professor William Stumph

2012

This Dissertation of Youly Ly is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

San Diego State University

2012

This dissertation is dedicated to

My parents, Sivly, and Michael

TABLE OF CONTENTS

Signature Page…… iii

Dedication…….. iv

Table of Contents……... v

List of Figures…… vii

List of Tables……. x

Acknowledgements…… xi

Vita……. xii

Abstract…….. xiii

I. Chapter 1 - Introduction…….. 1

1.1 Introduction…….. 1

1.2 The HMG Family of Proteins……. 3

1.2.1 HMG-14/-17 (old nomenclature) – HMGN (new nomenclature)…….. 5

1.2.2 HMG-I/Y (old nomenclature) – HMGA (new nomenclature)…….. 7

1.2.3 HMG-1/-2 (old nomenclature) – HMGB (new nomenclature)…….. 10

1.3 Lymphoid enhancer binding factor 1 (LEF-1)……... 15

II. Chapter 2 – Homology Analysis of LEF-1 and the Design of Stable LEF-1 Variants…… 27

2.1 Introduction – Sequence similarity of LEF-1 in vertebrates….. 28

2.1.1 Materials and methods…… 30

2.1.2 Results and Discussion…….. 31

2.2 Introduction – Sequence similarity of the HMG box within the HMGB family……. 40

2.2.1 Identification and sequence alignment of the HMG domain of human HMGB family members……. 42

2.2.2 Rational and computational design of LEF-1 Variants…… 46

2.2.3 Construction of the LEF-1 Variants…… 49

2.2.4 Protein expression and purification……. 59

2.2.5 Results and Discussion…… 60

III. Chapter 3 – The stability of LEF-1 and its variants

3.1 Introduction……... 67

3.2 Materials and methods…….. 69

3.3 Results……... 73

3.4 Discussion……. 85

IV. Chapter 4 – Functional Studies

4.1 Introduction…… 96

4.2 Materials and Methods……... 98

4.3 Results and Discussion…….. 102

4.4 Future Direction…….. 111

References……. 114

LIST OF FIGURES

Figure 1.1 Structural features of the high mobility group N (HMGN) proteins……..... 5

Figure 1.2. Amino acid alignment of human HMGA proteins…….. 9

Figure 1.3. NMR structure of the HMG box region of the B-domain of rat HMGB1……. 11

Figure 1.4. Regions of interactions…… 20

Figure 2.1 Multiple alignment of LEF-1 protein sequences…….. 33

Figure 2.2. Multiple sequence alignment of the HMG-box domain from members of the HMGB family……. 43

Figure 2.3. Selected amino acid positions for computational mutagenesis…. 48

Figure 2.4. DNA and amino acid sequence of the HMG domain of LEF-1… 49

Figure 2.5. Site directed mutagenesis primers for variants 2 and 4……... 50

Figure 2.6. Sequences of the six overlapping oligonucleotides used to assemble the gene for variant 5…… 52

Figure 2.7. Sequences of the six overlapping oligonucleotides used to assemble the LEF-1 gene (variant 6)…… 53

Figure 2.8. Sequences of the six overlapping oligonucleotides used to assemble the LEF-1 gene (variant 7)…… 54

Figure 2.9. Sequences of the six overlapping oligonucleotides used to assemble the LEF-1 gene (variant 8)…… 5

Figure 2.10. Sequences of the six overlapping oligonucleotides used to assemble the LEF-1 gene (variant 9)……. 56

Figure 2.11. Turn 1 region of the HMG domain of LEF-1…… 62

Figure 3.1. Protein stability analyzed by limited proteolysis……. 72

Figure 3.2. Far UV circular dichroism scans and thermal unfolding curves for wild type and the nine variants…… 74

Figure 3.3. Spectra of Wild Type and Variant 1…… 76

Figure 3.4. Spectra of Variant 2 and Variant 3…….. 77

Figure 3.5. Spectra of Variant 4 and Variant 5…….. 78

Figure 3.6. Spectra of Variant 6 and Variant 7…….. 79

Figure 3.7. Spectra of Variant 8 and Variant 9…….. 80

Figure 3.8. Limited proteolysis of LEF-1 and the variants…… 82

Figure 3.9. Correlation between limited proteolysis and protein thermal stability……. 83

Figure 3.10. Major hydrophobic core region of hUBF HMG Box-5 (PDB: 1L8Y). ……. 90

Figure 4.1. Analysis of the equilibrium binding of LEF-1 and the variants to a oligonucleotide containing the LEF-1 consensus recognition sequence in a electrophoretic mobility shift assay. 99

Figure 4.2. Structure of the 440 bp DNA probes used in the circulation permutation analysis…… 100

Figure 4.3. List of primers used in generating the 440bp DNA probes for circular permutation analysis…… 101

Figure 4.4. Circular permutation analysis of DNA bending induced by binding of wild type and the variants…… 104

Figure 4.5. Schematic representation of Groucho/TLE…… 109

LIST OF TABLES

Table 1.1. Revised nomenclature for the HMG family……... 4

Table 2.1. Percent homology between regions of hLEF-1 compared to mouse, frog, fish, and chicken…….. 32

Table 2.2. Abbreviation and protein name for sequence specific HMGB family members……. 44

Table 2.3. Abbreviation and protein names for non-sequence specific HMGB family members…… 45

Table 2.4: Variants of LEF-1 made from rational and computational design. 65

Table 3.1. Amino acid sequence of LEF-1 and the nine variants with their corresponding melting temperature…….. 75

Table 3.2 Amino acid sequence of LEF-1 and the nine variants with their corresponding melting temperature and time required for 50% degradation by limited proteolysis…… 84

Table 4.1 Summary of the Kd values of WT and the nine variants…….. 103

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor John Love for all the guidance and time he has provided me in my graduate career. I would also like to thank Professor Elizabeth Komives, Professor Crowell, Professor Terry Frey, and Professor William Stumph for your time and understanding through the exam, defense and the scheduling process. I would also like to thank Dr. Elizabeth Komives, Dr. William Stumph, Dr. Tom Huxford, Dr. Aileen Knowles, and Dr. Peter van der Geer for allowing me to use your lab equipment. I would also like to thank my colleagues Nora Barakat, Nermeen Barakat, Nesreen Barakat, Lisa Carmody, Sarah Hamilton, Mackenzie Patterson, Aditi Apte, and Mario Navarro for all of your insightful advice on experiments and presentations. I would also like to thank my parents, brother and sister for your support during my graduate career. In addition, I would like to thank the California Metabolic Research Foundation for the Arne N. Wick Pre-doctoral Research fellowship.

VITA

2002	B.S., Biochemistry, University of California, Riverside
2002-2003	Laboratory Assistant, University of California, Riverside
2005-2009	Teaching Associate, Chemistry Department San Diego State University
2009	Recipient of an Arne N. Wick Pre-doctoral Research Fellowship from the California Metabolic Research Foundation.
2012	Ph.D., Chemistry, University of California, San Diego and San Diego State University (Joint Doctoral Program)

PUBLICATIONS

Lan Ge, Youly Ly, Morley D. Hollenberg and Kathryn A. DeFea, B-arrestin-dependent sequestration of MAPK to pseudopodia is involved in PAR-2 mediated chemotaxis, *J. Biol. Chem*, 2003, 278 (36): 34418-26.

Youly Ly and John J. Love, Assessing the potential correlation between protein stability and function of the HMG domain of Lymphoid Enhancer Binding Factor-1, in preparation.

FIELDS OF STUDY

Major Field: Biochemistry and Molecular Biology

Studies in protein stability and function
Professor John Love

ABSTRACT OF THE DISSERTATION

Assessing a potential correlation between protein stability and function of the HMG domain of Lymphoid Enhancer Binding Factor-1

by

Youly Ly

Doctor of Philosophy in Chemistry

University of California, San Diego, 2012

San Diego State University, 2012

Professor John Love, Chair

Lymphoid enhancer binding factor-1 (LEF-1) is an architectural transcription factor that is a member of the HMGB family. The NMR spectra of the LEF-1 DNA binding domain (HMG domain) complexed with DNA exhibits the characteristic NMR spectra of a folded globular protein. In contrast, NMR analysis of the free LEF1 HMG domain revealed the presence of significant structural instability and conformational heterogeneity. Its disorder-to-order transition upon binding to DNA renders LEF-1 an excellent candidate for exploring the potential role stability plays in the function of the HMG domain. To gain insight into the possible role of LEF-1

disorder, or lack of stability prior to binding to DNA, we undertook the challenge of introducing stability into the HMG domain of LEF-1. This work will first demonstrate the use of homology modeling, structural analysis, and protein design algorithms (ORBIT) in an attempt to stabilize the HMG domain of LEF-1. Nine variants of the HMG domain of LEF-1 were created and structurally analyzed using NMR, CD, and limited proteolysis. Five variants were reported to be more stable than wild type. Lastly, we assessed the degree by which the nine mutant variants function in binding and bending cognate DNA. The results indicate that increased stability does not affect the degree by which the LEF-1 domain bends DNA but does have a small yet consistent affect on the binding affinity.

Chapter 1

Introduction

1.1 Introduction

Multicellular organisms depend on the cells ability to store, retrieve, process, and translate the genetic information that is stored in DNA (Cleynen and Van de Ven,Wim J M, 2008). Large amounts of DNA are packaged into the nucleus of eukaryotic cells by coiling around histones to form a subunit referred to as the nucleosome (Stros, 2010). Nucleosomes are then packaged together to form a chromatin fiber which is condensed into a chromosome. The cell has the ability to duplicate the chromosome and distribute the duplicated DNA to daughter cells to allow for genetic information to be passed from one generation to the next (Cleynen and Van de Ven,Wim J M, 2008).

The chromatin fiber not only functions as a packaging tool but also it controls the access of regulatory factors to their cognate DNA binding sites and allows interactions between distant regulatory elements. This is achieved by changes in chromatin folding and the remodeling of chromatin as a consequence of DNA bending and unwinding (Stros, 2010) (Cleynen and Van de Ven,Wim J M, 2008). The Histone H1 family and a large and diverse family of HMG (High Mobility Group) proteins function to mediate structural changes in chromatin. (Stros, 2010). The HMG family is subdivided into three families named HMGA (the HMG-AT hook family), HMGN (the HMG-nucleosome binding family), and HMGB (the HMG-box family) (Bustin, 2010). Each HMG family has a distinct functional motif (*i.e.* HMGA contains a

named the "AT-hook", in HMGB it is the HMG-box, and in HMGN it is the nucleosomal binding domain) (Stros, 2010). Interestingly, in addition to these structural changes in chromatin, there are also structural changes, on a more minor scale, in some of the HMG family members upon binding DNA (*e.g.,* the Lymphoid Enhancer Binding Factor-1, LEF-1, transitions from a disordered to ordered structure upon binding DNA) (Love *et al*., 2004). To gain insight into the potential roles of LEF-1's disorder, or lack of stability prior to binding to DNA, we undertook the challenge of introducing stability into the HMG domain of LEF-1. The remainder of this chapter is dedicated to illustrating how a diverse group of proteins (the HMG family) with different structures and DNA-binding motifs function to modulate chromatin structure.

1.2 The HMG Family of Proteins

Histones and the high mobility group (HMG) proteins play an important role in the assembly and remodeling of chromatin (Zhang and Wang, 2008) (Cleynen and Van de Ven, 2008). Second to histones, the HMG proteins are highly abundant in the nucleus (Zhang and Wang, 2008) (McCauley *et al*., 2007) and not only remodel the assembly of chromatin but also regulate gene transcription in eukaryotic cells by distorting, bending, or modifying the structure of the DNA (Zhang and Wang, 2008). These proteins were identified more than 30 years ago by Goodwin and coworkers (Goodwin *et al*., 1973) in an insoluble fraction that was obtained during the isolation of pure histones. The amino acid sequences of the HMG proteins were difficult to elucidate due to their tendency to aggregate after dissociation from their chromatin complex.

Goodwin and coworkers demonstrated that non-histone proteins can be removed from calf thymus chromatin by extracting with 0.35 M sodium chloride (Goodwin *et al*., 1973). The extracted proteins contained high and low mobility proteins as determined by polyacrylamide gel electrophoresis. Trichloroacetic acid was added to the sodium chloride extraction to precipitate the low mobility proteins. The isolation of the high mobility proteins has allowed scientists to further characterize these proteins and discover their roles in a variety of cellular processes, including progression of cancer, DNA repair, and infectious/inflammatory disorders (Zhang and Wang, 2008). HMG proteins are architectural transcription factors that modify the structure of their binding partners to regulate DNA-dependent activities

that include transcription, replication, recombination, and repair (Cleynen and Van de Ven, 2008).

Today the HMG superfamily is subdivided into three families that are defined by unique functional motifs. The families are HMGA which contains the AT-hook motif; HMGB which contain the HMG-box motif; and HMGN which contains the nucleosomal binding motif (Bustin, 2010). It was not until 2001 that a systematic method was introduced to name the three families (Bustin, 2001). Previously, the HMGA family was known as HMG-I/HMG-Y/HMG-C; the HMGB family was known as HMG-1/HMG-2; and the HMGN family was known as HMG-14/HMG-17 (Bustin, 2001). A summary of the revised nomenclature is shown in Table 1.1 followed by descriptions of each protein family.

Table 1.1 Revised nomenclature for the HMG family.
Abbreviations: HMG, high mobility group; NBD, nucleosome binding domain; ATH, AT-hook

New Name	Old Name	Functional Motif
HMGB	HMG-1/HMG-2	HMG-box
HMGN	HMG-12/HMG-17	NBD
HMGA	HMG-I/HMG-Y /HMG-C	ATH

1.2.1 HMG-14/-17 (old nomenclature) – HMGN (new nomenclature)

There are 6 protein members of the HMGN family, HMGN1, HMGN2, HMGN3a, HMGN3b, HMGN4, and HMGN5 which are only found in vertebrates (Reeves, 2010)(Gerlitz *et al*., 2009) (Hock *et al*., 2007). HMGN1-4 consists of approximately 90 amino acids while HMGN5 consists of approximately 400 amino acids. All HMGN proteins contain three functional regions: 1) two nuclear localization signal (NLS) regions, 2) a nucleosomal binding domain (NBD), and 3) an acidic tail called the chromatin-unfolding domain (CHUD) (Figure 1.1) (Reeves, 2010) (Gerlitz *et al*., 2009).



Figure 1.1 Structural features of the high mobility group N (HMGN) proteins. Functional domains of the HMGN proteins include two nuclear localization signal domains (NLS), a nucleosomal binding domain (NBD), and an acidic tail referred to as chromatin unfolding domain (CHUD).

Unlike other DNA binding domains found in HMGA or HMGB, the three dimensional structure of the NBD has yet to be determined. The NBD is made up of 30 amino acids that contains an eight amino acid sequence (RRSARLSA) that is found to be conserved in all proteins found in the HMGN family (Gerlitz *et al*., 2009) . Of the eight amino acids found to be conserved in the family, the sequence R-S-RL is found to be important for non-sequence specific binding to the nucleosome (Gerlitz *et al*., 2009). HMGNs are the only non-histone proteins that bind specifically to the 147 base pair nucleosome core particle to enhance transcription, replication, and DNA repair

(Reeves, 2010). In addition the C-terminus has been found to be important for the binding affinity of chromatin and transcriptional activation.

1.2.2 HMG-I/Y (old nomenclature) – HMGA (new nomenclature)

In mammals, HMGA proteins are transcribed from two genes, HMGA1 and HMGA2 (Reeves, 2010)(Pfannkuche *et al*., 2009). There are three splice variants of HMGA1 -HMGA1a, HMGA1b, and HMGA1c. HMGA2 is transcribed to full length HMGA2 protein (Reeves, 2010)(Cleynen and Van de Ven, 2008). All HMGA proteins contain a AT-hook region that is 8 or 9 amino acids long with the following amino acid sequence (Arg-Gly-Arg-Pro) flanked by additional amino acids on each side (*e.g.* Lys-Arg-Gly-Arg-Gly-Arg-Pro-Arg-Lys) (Pfannkuche *et al*., 2009) (Cleynen and Van de Ven, 2008). The arginine residues found in this region are responsible for the electrostatic interaction involved in the non-sequence specific binding between the AT-hook region and the minor groove of the B-form DNA (Cleynen and Van de Ven, 2008). NMR studies have shown the disorder to order transition of the AT-hook region upon binding to DNA (Cleynen and Van de Ven, 2008). Three AT-hook regions can be found in proteins belonging to the HMGA family (Cleynen and Van de Ven, 2008). The first two AT-hook regions found in HMGA1 and HMGA2 are identical. In the third AT-hook region, the region differs in two positions (Figure 1.2). It is not necessary for all three AT-hook regions to bind the same DNA segment. There are some instances where one or two may be bound to one segment while the other one is bound to another DNA segment. HMGA proteins utilize the AT-hook region to regulate gene expression by changing the DNA conformation upon binding to the AT-hook region which allows interaction among other transcription factors that flank the AT-hook region (Cleynen and Van de Ven, 2008). This allows the HMGA

proteins to take part in a variety of biological processes that include cell growth, proliferation, differentiation and cell death (Cleynen and Van de Ven, 2008). The AT-hook region is followed by several negative amino acids in the C-terminus region of which the function is still unknown

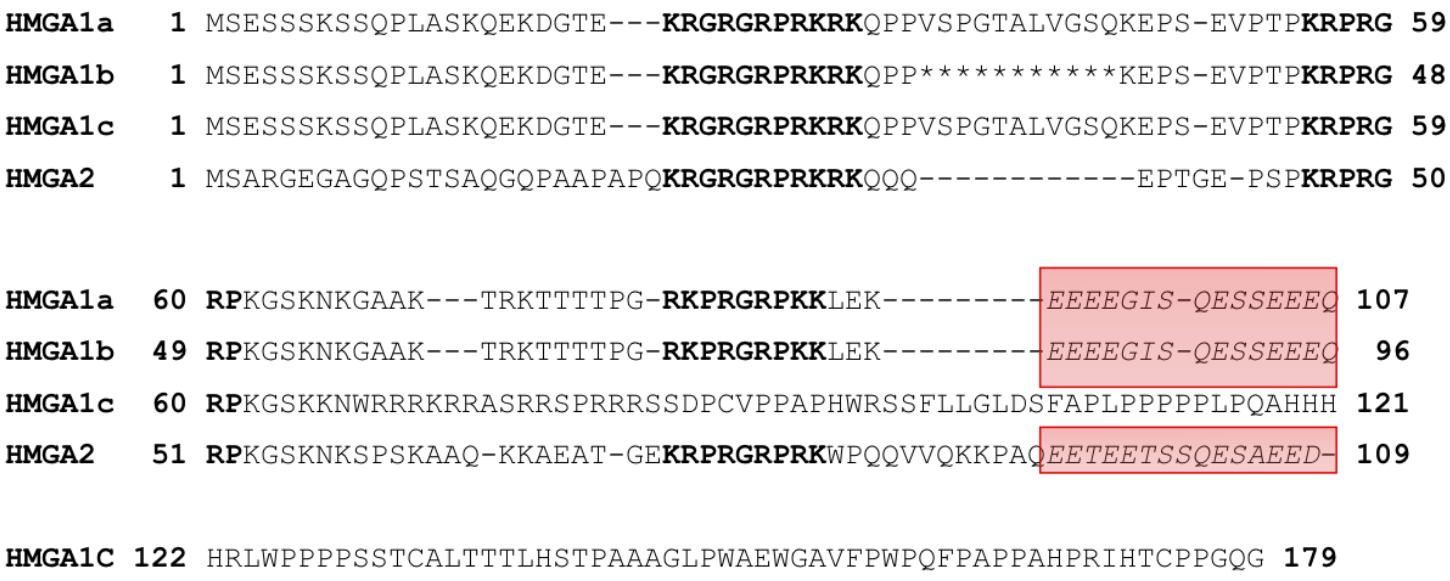


Figure 1.2. Amino acid alignment of human HMGA proteins. AT-hooks (bold); C-terminal acid tail (red box;italics). The amino acid residues deleted in HMG1b compared to HMG1a (asterisks). Figure adapted from (Cleynen and Van de Ven, 2008).

1.2.3 HMG-1/-2 (old nomenclature) – HMGB (new nomenclature)

The HMGB family is the most abundant family of the HMG proteins and is characterized by a DNA binding domain, the HMG-box (the human HMGB family is listed in Figure 2.5) (Gerlitz *et al*., 2009). The LEF-1 domain, which is the primary focus of this thesis, is a member of this family. The HMG boxes are made up ~75 amino acids that fold to form an L-shaped structure consisting of three α-helices (Figure 1.3) (Stros *et al*., 2007). The long arm of the L comprises the N-terminal strand and helix III while helices I and II make up the short arm of the L-shaped structure. The L-shaped structure is partially held together by a hydrophobic core (referred to herein as the major hydrophobic core) and one minor core (Travers, 2000). It should be noted that the three-dimensional structure of the HMG box is highly conserved within this family whereas the amino acid sequences of the HMG boxes are fairly diverse (Stros *et al*., 2007).

Members of the HMGB family are known to regulate transcription and replication. In addition, it has been shown that members of the HMGB family play a role in the determination of cell phenotype and also play important roles as extracellular signaling molecules (Gerlitz *et al*., 2009). This family can be divided into two groups (Stros *et al*., 2007). The first group consists of HMG-Box proteins that are non-sequence specific, usually contain more than one HMG domain, and contain an acidic C-terminal tail (Stros *et al*., 2007). The second group is made up of HMG-Box proteins that do not have an acidic C-terminal tail, are sequence specific, and normally contain one HMG domain (Stros *et al*., 2007).

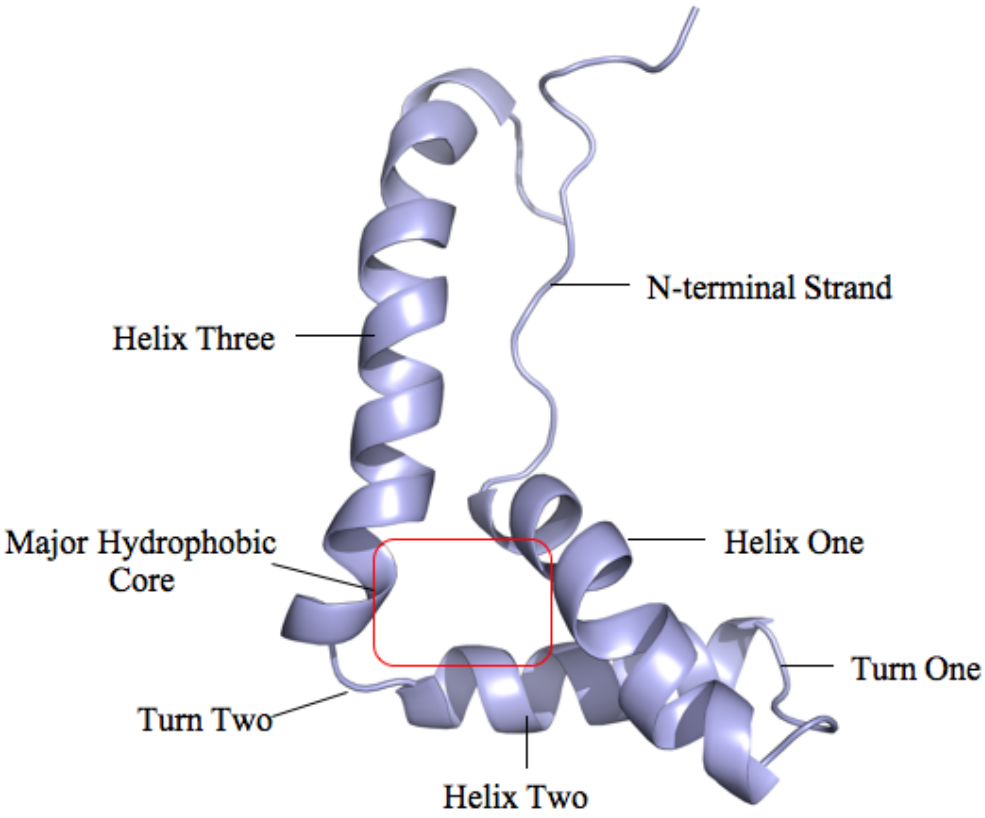


Figure 1.3. NMR structure of the HMG box region of the B-domain of rat HMGB1. The characteristic L-shaped fold of the HMG box is comprised of an extended N-terminal strand and 3 α-helices. The boxed region indicates the conserved major hydrophobic core. (Weir *et al.,* 1993, Protein Data Bank ID code – 1HME)

Interestingly, the HMG box of the non-sequence specific subclass (*e.g.* HMGB1 and HMG-D) exhibit conformational stability in the absence of DNA whereas the sequence specific family members are less ordered (*e.g.* LEF-1) (Love *et al*., 2004). The NMR spectra of the HMG domains from the non-sequence specific subclass are relatively high in quality and, due to their L shape structure, these domains tumble anisotropically as rigid ellipsoids. It has been theorized that the structural stability observed for the HMG domains of non-sequence specific family members, for example human upstream binding factor (hUBF), may be due to the presence of a greater number of hydrophobic residues in the hydrophobic core compared to other HMG domains (Liu *et al*., 2009). Masse *et al.* performed a sequence comparison of 33 HMGB proteins from a variety of organisms ranging from yeast to human and discovered the amino acid residues that make up the hydrophobic core are well conserved in both sequence specific family members and non-sequence specific family members (Masse *et al*., 2002). Interestingly, Masse and coworkers report a negatively charged amino acid (*i.e.* Glu57) within the major hydrophobic core that is well conserved among non-sequence specific family members. The negatively charged amino acid is commonly surrounded by positive amino acids to stabilize the fold of helix 2. In addition to Glu57, the major hydrophobic core is stabilized by aromatic residues. For example, three tightly stacked tryptophan residues occupy the hydrophobic core of hUBF (W14, W41, W52). The corresponding residues in HMGB1 are F17, W45, and Y56 and are packed in an edge-to-face manner.

For members of the sequence-specific group of HMGB1 proteins, the NMR spectra are usually incomplete (*e.g.,* Sox-4 and LEF-1) and is likely due to the mobile and dynamic movement of various regions of the proteins as for example the helix termini. The rmsd values increase significantly at the N- and C- termini of Sox-4 solution structures (Love *et al*., 2004). The structural dynamics in the N- and C-terminal strand is also seen in Sox-5 and LEF-1. The $^{3}J_{NH_{\alpha}}$ coupling constant values for residues found from the midpoint of helix III to the C-terminal increase in value and is indicative of the loss of α-helical structure towards the C-terminus. Interestingly, structural order is partially induced at the C-terminal end of the Sox-5 protein upon reducing the temperature from 25°C to 12°C. On the other hand, lowering the temperature of LEF-1 from 30°C to 22, 12, 10, or 5°C did not improve the NMR spectra and thus indicates that LEF-1 is likely more disordered than Sox-5 (Love *et al*., 2004).

In addition to the potential structural features described above, the lack of stable structure in sequence specific family members may also be attributed to increased dynamics in the major hydrophobic core (Figure 1.3). For example, the HMG box 2 of human mitochondrial transcription factor A (TFAM) contains two residues Y162 and Y200 that exhibit multiple conformations and potentially contribute to the instability of the HMG domain. The melting temperature of HMG box 2 (T_m is 23°C) reflects the absence of stability in the major hydrophobic core which is also reported with other sequence specific family members in the absence of DNA (*e.g.* Sox-5 and SRY) (Wong *et al*., 2009). Wong and coworkers were able to

impart stability in to the HMG domain by extending the N-terminal end (addition of 11 residues) of the HMG box which resulted in a higher T_m of 30°C. The observed increase in stability can be attributed to the fact that the N-terminal region packs against the C-terminal helix 3.

1.3 Lymphoid enhancer binding factor 1 (LEF-1)

Lymphoid enhancer binding factor 1 (LEF-1) was first identified in pre-B and T lymphocytes where it was shown to bind in the T cell receptor (TCR) α enhancer region (Travis et al., 1991). About five years later, LEF-1 and the closely related T cell factor (TCF) proteins (TCF-1, TCF-3, and TCF-4) were identified as nuclear binding partners of β-catenin that regulate the transcription of Wnt/β-catenin target genes (Robertson *et al*., 2011) (Vacik and Lemke, 2011)(Behrens *et al*., 1996) (Molenaar *et al*., 1996) (van de Wetering *et al*., 1997). Many of these target genes are involved in cell fate decisions, cell proliferation and differentiation, and changes in cell shape (Arce *et al*., 2006) (Wisniewska *et al*., 2010).

In addition to β-catenin, ALY (Ally of AML-1(acute myeloid leukemia factor) and LEF-1) (Bruhn *et al*., 1997), Notch (Ross and Kadesch, 2001), TGFβ_3 (transforming growth factor-β,Nawshad and Hay, 2003), and SMADS were identified as coactivators of LEF-1. The overexpression of ALY increases the activity of the TCRα enhancer by binding to LEF-1 which brings other DNA bound factors into close proximity of one another for transcription to occur (Bruhn *et al*., 1997). The Notch mediated coactivation of LEF-1 involves the intracellular domain of Notch (NICD) (Ross and Kadesch, 2001). High levels of NICD have been shown to bind to LEF-1 which increases the activity of the TCRα enhancer. TGFβ_3 is a secreted cytokine that has many biological functions that include pivotal roles in embryonic development (Nawshad and Hay, 2003). TGFβ_3 upregulates LEF-1 mRNA, and associated transcriptional activity of LEF-1, and has been demonstrated to induce

mouse palatal epithelial mesenchymal transformation. TGFβ_3 induces phosphorylation of Smad2 which is known to increase the transcriptional activity of LEF-1. LEF-1 also functions in a variety of roles that include organogenesis, T and B cell development, and a collection of a number of diseases (Arce *et al*., 2006) (Hoppler and Kavanagh, 2007).

T cell development. The complex differentiation process of T lymphocytes involves multiple steps that are defined by the expression of specific cell surface antigens (*e.g.* CD44 and CD25) and intracellular proteins. The means by which transcription factors (*e.g.* LEF-1 and TCF-1) regulate differentiation has been reviewed by Leiden and Thompson (Leiden and Thompson, 1994) and Clevers and Grosschedl (Clevers and Grosschedl, 1996). Okamura and colleagues report impairment to T cell differentiation and TCRα expression only when mice carry mutations in both the LEF-1 and TCF-1 gene. T cell differentiation is arrested at the immature CD8+ single-positive stage and TCRα gene transcription is severely reduced (Okamura *et al*., 1998).

Organogenesis. In vertebrates, organogenesis primarily involves two types of tissues, the epithelium and mesenchyme. The interactions between these two tissues controls the initiation of organ development, followed by morphogenesis, and terminal cytodifferentiation (Kratochwil *et al*., 1996). The initiation of organ development involves the thickening of the epithelium which either branches within the mesenchyme or folds around the mesenchyme to form a papillary structure. The final step (terminal cytodifferentiation) involves the differentiation of the epithlium and

mesenchyme cells that regulates the synthesis of the gene products specific for organ function (Kratochwil *et al*., 1996). LEF-1 has been shown to be involved in the epithelial-mesenchymal tissue interaction during teeth and hair development (Sasaki *et al*., 2005)(Kratochwil *et al*., 1996). LEF-1 regulates the fibroblast growth factor (FGF)-4 gene which then governs the expression of FGF-3 in the dental mesenchyme (Sasaki *et al*., 2005). LEF-1 deficient mice lack teeth, mammary glands, whiskers, and body hair (van Genderen *et al*., 1994). In addition to teeth and hair development, LEF-1 has been shown to be expressed in the neural crest, branchial arches, limb buds, thymus, lung, kidney, tooth germs, brain, and inner ear (van Genderen *et al*., 1994).

In development, one of the earliest and most important steps of all organisms is the establishment of the embryonic axes (*e.g.* in a bilateral animal: anterior-posterior axis, dorsal-ventral axis, left and right axis) (Croce and McClay, 2006). This establishes the foundation needed for the development of tissues and organs. The role of the Wnt signaling pathway in axis specification has been well studied in *Xenopus* development. Wnt signaling has been shown to function on the dorsal side of the early *Xenopus* embryo. When LEF-1 is artificially introduced into the ventral side of the embryo, it induces abnormal dorsal development.

Cancers and disorders. The level of LEF-1 expression has been shown to potentially be correlated to cancers or disorders. For example, the continuous activation of the Wnt signaling pathway leads to colon cancer (Jeong *et al*., 2007). As a result, there is an accumulation of beta-catenin in the nucleus which interacts with LEF-1 and other TCF proteins (TCF-1, TCF-3, and TCF-4) to activate target genes

(*e.g.* MYC, CCND1, MMP7, and TCF1). LEF-1 is abnormally expressed in the colon to accommodate the accumulation of beta-catenin in the nucleus (Hovanes *et al.*, 2001). On the other hand, low levels of LEF-1 can lead to disorders such as congenital neutropenia (CN). CN is characterized by incomplete maturation in granulopoiesis in the bone marrow, absence of mature granulocytes in peripheral blood, and bacterial infections (Skokowa and Welte, 2007). More than 20% of patients with CN progress to acute myelogenous leukemia (AML) (Skokowa *et al.*, 2006). LEF-1 is among several transcription factors (*e.g.* RUNX, PU.1, C/EBP-α) that regulate granulopoiesis. Patients with CN have low expression levels of LEF-1 and in some cases no expression at all. The restoration of LEF-1 in CN individuals restores granulopoiesis and displays the critical role LEF-1 plays in granulopoiesis (Skokowa and Welte, 2007).

In addition to the increased levels of LEF-1 expression found in certain cancerous tissue, domain deletions of LEF-1 can also lead to diseases such as T-cell acute lymphoblastic leukemia (T-ALL). T-ALL is characterized by aberrant thymocyte differentiation (Gutierrez *et al.*, 2010) (Ferrando et al., 2002). Thymocyte differentiation is indicated by the expression of two receptors, CD4 and CD8 (Dik *et al.*, 2005). In T-ALL patients, the developmental arrest occurs at the transition from a CD8 immature single positive (ISP) stage to the double-positive stage (Gutierrez *et al.*, 2010).The disruption in thymocyte differentiation is due to allelic deletions in exon 3 of chromosome 4. As a result, the context dependent activation domain and the HMG domain of LEF-1 are not translated. In sabaceous tumors, there is a mutation or deletion of the NH_2 terminus of LEF-1 that prevents β-catenin binding (Niemann *et al.*, 2002). The double nucleotide substitutions in exon 1 of the LEF-1 gene results in the following

amino acid substitutions E45K and S61P which is inhibiting β-catenin binding to LEF-1. Researchers have shown by expressing a LEF-1 variant with the deleted β-catenin binding domain causes sebaceous tumors to develop.

LEF-1 is also known to be involved in the development of pancreatic cancer. In pancreatic cancer, LEF-1 lacks exon VI which has been shown to enhance cell migration of pancreatic cancer cells (Jesse *et al*., 2010). This isoform of LEF-1 lacks parts of the context-dependent regulatory domain (CRD) and performs different regulatory activities than full length LEF-1. LEF-1 (without exon VI) inhibits cell to cell adhesion and as a result enhances cell migration. The 84 base pairs missing in LEF-1 (without exon VI) may function as a binding site for its interaction partners (*e.g*. Smad2 and Smad4). The interaction between LEF-1 and Smad2/Smad4 regulate the E-cadherin promoter activity (Jesse *et al*., 2010).

Structure. Full length LEF-1 is 397 amino acids long and has a calculated molar mass of ~44 kD (Travis *et al*., 1991). Like many transcription factors, LEF-1 has multiple domains. LEF-1 is made up of two transactivation domains, a beta-catenin domain (βBD) (residues 1-99) and a context-dependent activation domain (CAD) (residues 99-217). The CAD domain is followed by a DNA binding domain (residues 296-380) referred to as the high mobility group domain that places it in the HMGB family (Figure 1.4) (Daniels and Weis, 2005).

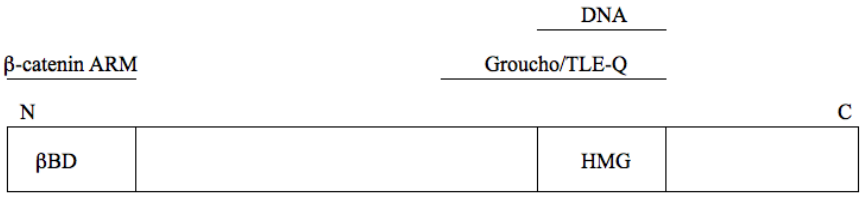


Figure 1.4. Regions of interactions. A diagram of the primary structure of LEF-1 showing the β-catenin binding domain (βBD), the Groucho/TLE-Q binding domain and the HMG DNA binding domain. (ARM, β-catenin armadillo repeat domain)

The NMR spectra of the LEF-1 DNA binding domain complexed with DNA exhibits the characteristic NMR spectra of a folded globular protein (Lee and Vasmatzis, 1997). In contrast, NMR analysis of the free LEF1 DNA binding domain revealed the presence of significant structural instability and conformational heterogeneity (Lee and Vasmatzis, 1997). The two dimensional [^{15}N, ^{1}H] HSQC spectrum of the free protein is poorly dispersed, lacks the appropriate number of peaks (~63 out of a potential 83), and exhibits quite heterogeneous peak shape and intensity. Instability in the region encompassing the first and second helices renders all of the signals from helix one and most of the signals from helix two non-observable on the NMR time scale.

These observations in the NMR analysis give rise to certain questions regarding the LEF-1 structure such as, why does this domain exhibit significant disorder and structural heterogeneity prior to binding DNA? Could the flexibility be cooperative with the gross structural perturbations that occur within the DNA upon bending? Furthermore, is this behavior functionally essential or simply an artifact of

studying an excised domain? If this is not a thermodynamic necessity, then has it evolved to modulate the kinetics of binding? What impact would the introduction of stability have on the role LEF-1 plays as an architectural transcription factor, and for that matter, could stability be introduced?

The unique disorder-to-order transition of the HMG domain of LEF-1 upon binding DNA places LEF-1 among a growing number of proteins that are identified as intrinsically disordered proteins (IDP). Some IDPs are completely disordered and only form a tertiary structure upon binding to their target (*e.g.* protein or DNA) while others contain motifs within the disordered protein that fold upon binding to their target (Wright and Dyson, 2009). Many of these IDPs participate in cellular control mechanisms, signaling and are at the center of protein interaction networks (Wright and Dyson, 2009). The intrinsically disordered regions can be greater than 50 residues and are commonly found in eukaryotes (Minezaki *et al*., 2006) (Dyson and Wright, 2005). These regions provide a large binding surface for ligands and permits a protein to bind to multiple targets (Miller, 2009). IDPs can bind to multiple targets due to its structural plasticity which allows for IDPs to adopt several conformations (Wright and Dyson, 2009).

Liu and coworkers have reported that 82% of eukaryotic transcription factors contain intrinsically disordered regions (Liu *et al*., 2006). The intrinsic disorder may have evolved to overcome the thermodynamic and kinetic challenges in sequence specific binding of DNA (Dyson and Wright, 2005). The flexibility found in transcription factors is utilized during transcription activation where complexes of

many different proteins need to be formed and changes in the chromatin structure occurs. The family of proteins that contain leucine zippers also exhibit structurally disordered basic regions that take part in important cellular functions, developmental regulation, metabolism, and responses to environmental changes (Miller, 2009). Members of the leucine zipper family are made up of a positively charged DNA binding domain, a basic region, and a sequence of heptad leucine repeats known as the leucine zipper. Biophysical studies have shown that there is a lack of structure outside the DNA binding domain where a coil to α-helix transition in the leucine zipper region occurs upon homodimerization (Miller, 2009). Leucine zipper proteins homodimerize to bind to their target DNA site where the DNA stabilizes the α-helix which induces helical formation of the basic region. The unfolding and refolding of the protein allows it to change binding partners which is a way to regulate gene expression.

LEF-1 is also a transcription factor that regulates gene expression. Its disorder-to-order transition upon binding to DNA renders LEF-1 an excellent candidate for exploring the role stability plays in the function of the HMG domain. LEF-1, a DNA sequence specific transcription factor shares the characteristic L-shaped fold of non-sequence specific HMGBs but only when bound to DNA. LEF-1 may have evolved from non-sequence specific family members and thus sequence homology analysis could be utilized for the goal of imparting stability on the HMG domain of LEF-1. To gain insight into the role of LEF-1's significant disorder prior to binding to DNA we undertook the challenge of introducing stability into the HMG domain of LEF-1.

For many years protein engineers have attempted and succeeded in thermostabilizing proteins that are present in nature as well as proteins used for

industrial processes (Kazlauskas and Bornscheuer, 2009) (Joo *et al*., 2010). Although success has been achieved in a number of cases, the thermostabilization process for proteins still remains an important issue and a significant challenge. In general, thermostabilization entails three potential steps: 1) choosing residues to mutate in the protein sequence based on the 3-dimensional structure (which involves rational design or possibly the randomization of the codon for particular amino acid positions), 2) choosing a method that will incorporate the desired changes (site-directed mutagenesis or total gene synthesis), and 3) identifying which mutant variants show improved properties (which normally entails the use of a screen or selection process) (Kazlauskas and Bornscheuer, 2009). For example, a goal of many protein engineering projects is to increase the thermostability of important enzymes. To achieve this a structure-based approach may be utilized with the assumption that a more rigid structure will result in a more stable enzyme at higher temperatures (Kazlauskas and Bornscheuer, 2009). A high quality 3-dimensional structure is critical for this process as it is needed to determine whether the structure can be stabilized upon introduction of either a disulfide bond, a salt bridge, a proline residue to potentially stabilize a turn, or hydrophobic residues in a flexible region in the protein to create a stabilizing hydrophobic core in the protein (Kazlauskas and Bornscheuer, 2009). Another approach one may take to stabilize a protein is directed evolution combined with the use of bioinformatics tools. This involves the assumption that well conserved amino acids contribute to the overall stability of the protein (Kazlauskas and Bornscheuer, 2009). The homology modeling approach involves comparing related sequences to

that of the target sequence and mutating residues to amino acids that occur most frequently at that particular position.

Another approach to stabilize a protein is achieved through computational protein design. This involves identifying amino acid sequences that are compatible with a given three dimensional structure. Protein design algorithms perform energy assessments on the amino acid sequences and with the use of optimization methods such as dead-end elimination, Monte Carlo simulated annealing, genetic algorithms, or optimization theory approaches low energy sequences are identified (Saven, 2011). For many years there has been great success in stabilizing proteins through computational desgin (Kim *et al*., 2010)(Ko *et al*., 2007) (Korkegian *et al*., 2005). For example, Kim and colleagues redesigned *Coprinus cinereus* heme peroxidase computationally to improve its thermostability for commercial use (Kim *et al*., 2010). They used a RosettaDesign (Kuhlman and Baker, 2000) method that utilizes an energy function to evaluate the fitness of a given sequence for a given three dimensional structure and a monte carlo search algorithm for sampling sequence space (Kim *et al*., 2010). Eight mutants of *Coprinus cinereus* heme peroxidase were produced from RosettaDesign and two of the eight mutants were determined experimentally more stable than wild type. Kim *et. al.* were able to design a more thermostable enzyme without affecting its function (Kim *et al*., 2010).

Over the years, studies have been published that explore the correlation between protein stability and function (Teilum *et al*., 2011). The published studies have shown weak, positive, and negative correlation betweeen protein stability and

function. For example, single site mutations were performed on 26 amino acid sites of acyl coenzyme A to test for stability and function. The stability of the 28 variants of acyl coenzyme A binding protein (ACBP) were measured with chemical denaturation followed by intrinsic tryptophan and tyrosine fluorescence and its binding affinity for dodecanoyl-CoA was measured with isothermal titration calorimetry (Kragelund *et al*., 1999). The results of the studies demonstrated the following: some variants were stabilized with greater affinity for dodecanoyl-CoA, some variants were destabilized with a decrease in affinity for dodecanoyl-CoA, and some variants showed no correlation at all between protein stability and function (Kragelund *et al*., 1999). This study is a good example of a weak correlation between protein stability and function. Another study demonstrates a strong correlation between protein stability and function. Krageland and colleagues mutated residues (*e.g*. Phe 10, Leu 23, Leu 28, Val 61, and Phe 66) in the hydrophobic core of calbindin D_{9k} to either alanine or glycine and showed a decrease in protein stability and a decrease in the affinity for Ca^{2+} (Kragelund *et al*., 1998). The L23G mutation demonstrated the largest decrease in the affinity for Ca^{2+}. Krageland and coworkers suggest the backbone of Leu is involved in the binding loop of Ca^{2+} and may participate in the stabilizing interactions of the adjacent Ca^{2+} binding site. Thus, the replacement of a bulky hydrophobic residue with a glycine decreases its stability and affinity for Ca^{2+}. The last study demonstrates no correlation between protein stability and function. Torrado and colleagues explored the role of a pair of amino acids (residues 19 and 30) in the stability of the homeodomain protein. They performed the following mutation (V19R and I30D) in an

attempt to form a salt bridge between residues 19 and 30. Their results demonstrated the destabilization of the homeodomain protein (*e.g.* wild type T_m = 56.3°C and variant T_m = 36.2) and showed destabilization had no affect on the binding affinity for the DNA. In addition, we present our study of the potential correlation between protein stability and function of the HMG domain of LEF-1.

This work will address the following: our attempts to rationally design stable variants of LEF-1 (Chapter 2), the structural analysis and characterization of nine mutant variants of LEF-1 (Chapter 3), and lastly assess the degree by which the nine mutant variants function in binding and bending cognate DNA (Chapter 4).

Chapter 2

Homology Analysis of LEF-1 and the Design of Stable LEF-1 Variants

In this chapter a sequence homology analysis is described and it is divided into two categories, 1) sequence comparison of LEF-1 to other vertebrate sequences and 2) sequence comparison of the LEF-1 HMG domain against other HMGB family members. The amino acid sequence of the human LEF-1 variant was used (*i.e.* 'blasted') to find homologues in other vertebrates and an alignment of LEF-1 among these different vertebrates showed functional domains to be well conserved. The parameters of the blast search will follow the introduction and a discussion of the strong correlation between amino acid conservation and protein function will conclude the first analysis. An introduction to the homology modeling approach we used to thermally stabilize LEF-1 will follow, which proved to be successful and resulted in four variants (*i.e.* variants 1-4) with increased thermal stability (Chapter 3). A more traditional structure-based approach is described next to design four additional variants (*i.e.* variants 5-8). Lastly, a suite of protein design algorithms developed by the Mayo laboratory was utilized in an attempt to create a more thermostable variant (Dahiyat and Mayo, 1996). The description of the parameters of the blast search, computational design, construction of the variants, and protein expression/purification will follow the introduction and conclude with a discussion of the amino acid positions that participate in the stability of the HMG domain of LEF-1.

2.1 Introduction – Sequence similarity of LEF-1 in vertebrates

The alignment of multiple protein sequences has provided vital information for protein classification and analysis. Multiple sequence alignments make it possible to determine conserved residues that may participate in protein function or structure. Residues involved in protein function or structure are often conserved whereas residues found on the surface of a protein vary more widely throughout evolution (Schueler-Furman and Baker, 2003). The conserved residues involved in protein function or structure are often grouped together and given a domain name to identify the significance of the amino acids in the primary sequence. This information has been used to develop bioinformatic programs to aid in the search for related sequences in the protein sequence space (Gribskov *et al*., 1990) (Birney *et al*., 1996). In addition, multiple sequences alignment databases have been developed to analyze data from the large genome sequencing projects (Schultz *et al*., 1998) (Bateman *et al*., 1999).

The goal of this section of my thesis was, in part, to determine residue positions and amino acid substitutions that could possibly lead to an increase in thermal stability for LEF-1. In addition, we also wanted to determine if any other proteins had sequences similar to the sequences that conferred greater stability to the LEF-1 HMG domain. We also desired to ascertain correlations between LEF-1 sequences and its different functions. To shed light on whether a conserved function is correlated with a conserved sequence, the following section is provided and describes the results of using the LEF-1 sequence (*i.e.,* 'blasted') to possibly find homologues in other organisms. LEF-1 has been shown to be important in the development of

eukaryotic organisms by participating in the Wnt signaling pathway responsible for regulating development from early axis specification to organogenesis (Robertson *et al*., 2011). We have utilized the program Blast (Altschul *et al*., 1997) to report the amino acid similarities in four vertebrates.

2.1.1 Materials and Methods

The amino acid sequence of the human variant of LEF-1 (NCBI Reference Sequence: NP_057353.1) was used as a query to identify LEF-1 variants in mouse (taxid: 10090), frog (taxid: 8355), fish (taxid: 7898), and chicken (taxid: 9030) using the program blastp (http://www.expasy.org/tools/blast/) [query against the UniProt Knowledgebase (Swiss-Prot + TrEMBL)]. The auto-select comparison matrix option was utilized and the threshold parameter was set at 10 for the blast search. An output of a list of potentially matching sequences is provided with the most relevant sequences listed first. The sequences were then aligned with ClustalW2 (Larkin et al., 2007) using the following parameters: pairwise alignment options (alignment type slow; protein weight matrix gonnet; gap open 10; gap extension 0.1) and multiple sequence alignment options (protein weight matrix gonnet; gap open 10; gapextension 0.2; gap distances 5; no end gaps; iteration none; numiter 1; clustering NJ).

2.1.2 Results and Discussion

Sequence of LEF-1 compared to vertebrates

Alignment of the LEF-1 protein among different vertebrates reveals a remarkable conservation in the primary amino acid sequence (Figure 2.1, Table 2.1). The LEF-1 protein sequences from mouse (Milatovich *et al*., 1991), frog (Pukrop *et al*., 2001), zebrafish (Gamba *et al*., 2010), and chicken (Doi *et al*., 2011) have greater than 80% homology compared to the full-length human LEF-1 protein sequence. The sequence similarity amongst these species for the β-catenin domain and the CAD domain is greater than 76% compared to hLEF-1 while the HMG domain displays greater than 98% compared to hLEF-1. There is only a single amino acid difference in the mouse HMG domain compared to the human variant. This strong evolutionary conservation may, in part, be associated with Wnt signaling as described below.

Wnt signaling is an important and well conserved pathway in metazoan animals (Hoppler and Kavanagh, 2007). LEF-1 is a member of the TCF family of transcription factors that participates in the Wnt signaling pathway which regulates development from early axis specification to organogenesis (Robertson *et al*., 2011) (Atcha *et al*., 2007). LEF-1 needs to be conserved to function properly in the Wnt signaling pathway for an organism to develop and survive. The Wnt signaling pathway begins with the Wnt ligand binding to the Frizzled receptor and tranducing intracellular signals through either a β-catenin dependent pathway or a β-catenin independent pathway. In the Wnt/β-catenin dependent pathway, the intracellular signal disassembles the β-catenin degradation complex. This complex is comprised of Axin,

APC (adenomatous polyposis coli), GSK3 (glycogen synthase kinase-3), GBP (GSK3 binding protein), CKs (casein kinases) 1α/2 and PP2A (protein phosphatase 2A) and allows β-catenin to enter the nucleus to bind to TCF/LEF transcription factors. This process initiates the transcritional activity responsible for cellular fate such as stem cell determination, proliferation, and differentiation (Sethi and Vidal-Puig, 2010). A diversity of isoforms of TCF and LEF-1 exist to meet the demands of the vertebrate system.

Table 2.1. Percent homology between regions of hLEF-1 compared to mouse, frog, fish, and chicken.

	Percent Homology			
	Full Length LEF-1	Beta-catenin domain	CAD domain	HMG domain
Mouse vs Human	97.7	93.9	98.3	100
Frog vs Human	80.2	81.8	88.1	98.8
Fish vs Human	86	76.8	78.8	100
Chicken vs Human	94.6	85.9	96.6	100

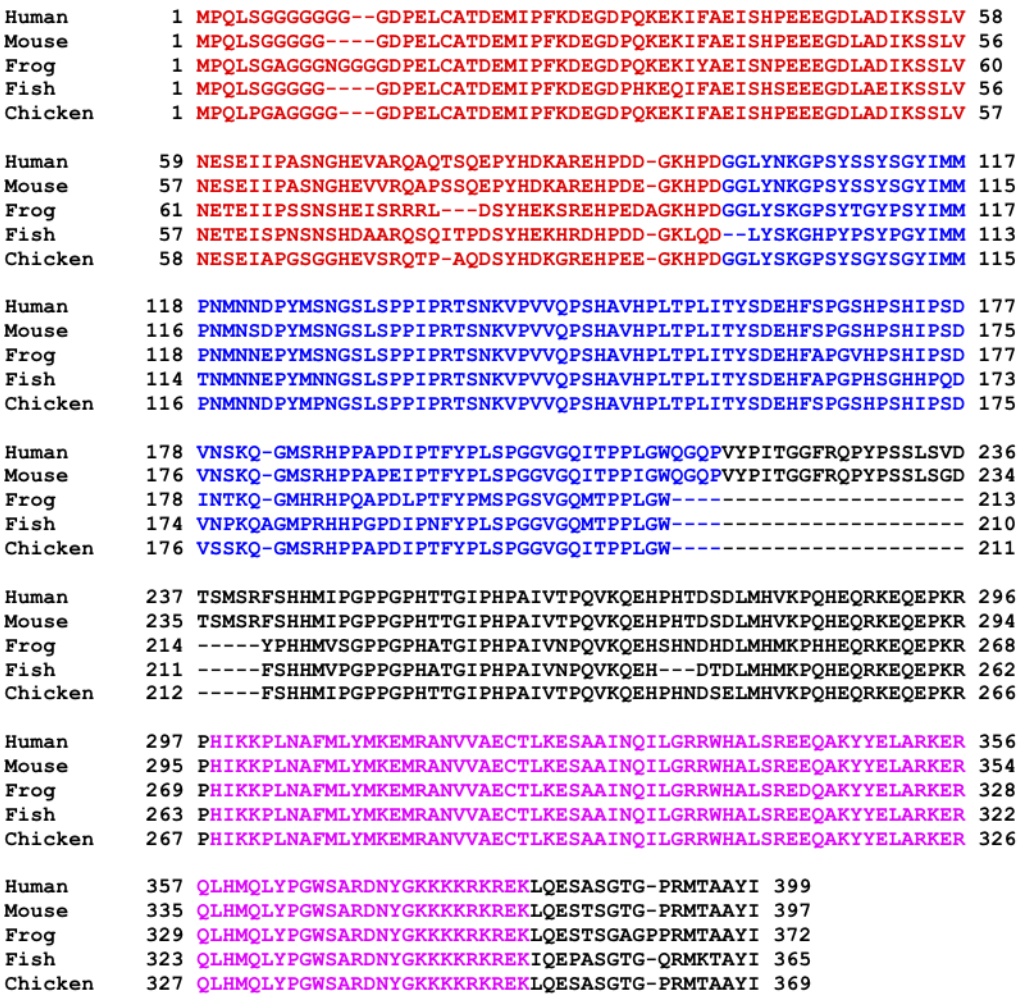


Figure 2.1 Multiple alignment of LEF-1 protein sequences. The protein sequences of the human (Milatovich *et al*., 1991), mouse (Milatovich *et al*., 1991), frog (Pukrop *et al*., 2001), fish (Gamba *et al*., 2010), and chicken (Doi *et al*., 2011) LEF-1s were aligned. Percent homology of each LEF-1 compared to human LEF-1 are as follows: mouse vs human, 97.7%; frog vs human, 80.2%; fish vs human, 86%; chicken vs human, 94.6%. Percent homology of each HMG domain (purple) of LEF-1 compared to human LEF-1 are as follows: mouse vs human, 100%; frog vs human, 98.8%, fish vs human, 100%; chicken vs human, 100%. β-catenin domain is denoted in red and context-dependent activation domain (CAD) in blue.

Vertebrates have three TCF genes (*e.g.,* TCF-1, TCF-3, and TCF-4) and one LEF-1 gene (Robertson *et al*., 2011) (Vacik and Lemke, 2011)(Macdonald *et al*., 2009). Invertebrates have one TCF gene and no alternative transcripts (Robertson *et al*., 2011)(Hoppler and Kavanagh, 2007). The TCF gene found in invertebrates consists of four domains: an N-terminal β-catenin domain, a central domain, an HMG domain, and a C-terminal tail which provides the framework for the vertebrate TCF gene (Atcha *et al*., 2007). TCF genes found in vertebrates also include isoforms of TCF that have lost parts of these domains and/or include new peptide motifs. In vertebrates, alternative promoters are used for translating TCF-1 or LEF-1. As a result, shorter isoforms of TCF-1 or LEF-1 is produced without the N-terminal β-catenin domain. These shorter isoforms allow the cells to repress the Wnt signaling pathway while allowing others to proceed (Hoppler and Kavanagh, 2007). In addition to alternative promoters translating shorter isoforms of TCF-1 and LEF-1, alternative splicing in the central domain creates isoforms of TCF-1 and LEF-1 genes that lack the central domain exon. The central domain exon contributes to the body axis development of mammals. As the embryo forms, its overall body pattern is determined by the establishment of three clear axes: the anterior-posterior axis (head-tail), the dorsal-ventral axis (back-belly), and the left-right axis. The central domain is crucial for the formation of the body axis. The introduction of the central domain exon increased the activity of Xenopus LEF-1 during secondary axis induction (Gradl *et al*., 2002). Lastly, alternative splicing at the C-terminus gives rise to isoforms with a variety of C-terminal tails expressed with or without three conserved regions: CRARF

motif (a highly conserved 5 amino acid sequence in the C-terminus), the RKKKCIRY motif (a highly conserved 8 amino acid sequence in the C-terminus), and the binding sites for CtBP cofactor. The CRARF motif and the RKKCIRY motif of TCF-1 or TCF-4 can activate the promoter responsible for expressing LEF-1. In colon cancer, cancer cells take advantage of the aberrant expression of LEF-1 to proliferate (Najdi *et al*., 2011).

The variety of TCF/LEF-1 isoforms provides the vertebrate system a set of molecular tools to perform various roles in development, stem cell maintenance, and immune function. The important role TCF/LEF-1 plays in development, stem cell maintenance, and immune function (as described below) may be the reason why TCF/LEF-1 is so strongly conserved in vertebrates.

Role of TCF/LEF-1 in development

In Xenopus development, Wnt signalling has been shown to regulate dorsal development, mesoderm induction, and midbrain patterning (Hoppler and Kavanagh, 2007). All TCF proteins and LEF-1 have been shown to either repress or activate development on the dorsal side of the embryo. These proteins are activators of development if they do not contain the following amino acid sequence, LVPQ in the central domain. TCF-1 does not have the LVPQ sequence in the central domain and has been shown to activate Wnt signaling on the dorsal side of the embryo and activate mesoderm induction by assembling the transcription activating complex (Standley *et al*., 2006). TCF-3 contains the LVPQ amino acid sequence and has been shown to

repress Wnt signaling in dorsal development and mesoderm induction. TCF-4 and LEF-1 also represses Wnt signaling if it contains the amino acid sequence LVPQ in the central domain (Hoppler and Kavanagh, 2007). In midbrain patterning, TCF-4 is required for midbrain patterning. The presence of the LVPQ and SFLSS (sequence also located in the central domain) amino acid sequence in TCF-4 enables it to activate midbrain patterning.

Role of TCF/LEF-1 in stem cell maintenance

TCF/LEF-1 proteins are also responsible for stem cell maintenance in the small intestine, large intestine, skin, and hair. TCF-4 activates stem cell proliferation in the crypts of the small intestine and the large intestine. The crypt is the part of the epithelium that lies beneath the villi (Reya and Clevers, 2005). Colorectal cancer cells take advantage of the TCF-4 protein to generate more cancerous cells in the small intestine and large intestine. In addition to utilizing TCF-4, cancerous cells utilize LEF-1 to generate more Wnt signalling target genes. Lastly, TCF-1 has been shown to repress the stem cell proliferation in the small intestine and the large intestine (Hoppler and Kavanagh, 2007).

Role of TCF/LEF-1 in skin and hair

LEF-1 and TCF-3 are both expressed in the skin. In mammals, skin and hair are produced in the embryo and maintained in the adult populations of multipotent stem cells (Alonso and Fuchs, 2006). The structure and tissues that make up the hair

include: the epidermis, the sebaceous gland, the outer-root sheath, the inner-root sheath, and the hair (Alonso and Fuchs, 2006). LEF-1 is present in the cell lineage that gives rise to the inner-root sheath and it is also found in hair. TCF-3 is present in stem cells and the lineage that produces the outer-root sheath. Van Genderen and colleagues have reported LEF-1 knockout mice to have very few hair follicles while Merrill and colleagues had no success in studying TCF-3 knockout mice due to the lethality of not having TCF-3 (van Genderen *et al*., 1994) (Merrill *et al*., 2001).

Role of TCF/LEF-1 in the immune system

The TCF/LEF proteins were first identified in the immune system (Wetering *et al*., 1991). TCF-1 isoforms are expressed at various stages of T-cell development in the thymus while LEF-1 is expressed in developing T-cells in the thymus and in developing B cells in the bone marrow (Wetering *et al*., 1996) (Travis *et al*., 1991). TCF-1 deficient mice show early impairment in T-cell differentiation (Clevers and Grosschedl, 1996). The differentiation stalls at the immature single positive (ISP) stage, a T-cell which expresses the pre-TCR with the TCR β-chain but has not yet rearranged the TCRα gene. ISP cells are the most actively proliferating thymocytes but in TCF-1 deficient mice they are small and noncycling (Clevers and Grosschedl, 1996). In LEF-1 knockout mice, T-cell differentiation is not affected but mice lack hair, mammary glands, and teeth (van Genderen *et al*., 1994).

The conservation in the β-catenin domain

The conserved amino acids in the β-catenin binding domain are required for the protein/protein interaction that takes place between β-catenin/LEF-1. The basic structure of β-catenin consists of an unstructured N-terminal tail, an armadillo repeat domain, and a C-terminal acidic tail (Sun and Weis, 2011). The armadillo repeat domain consists of 12 armadillo repeats that form a superhelix of α-helices that bind to the β-catenin domain of TCF/LEF-1, APC, or cadherins. Sun and colleagues have reported the following amino acids necessary for the interaction between LEF-1 and β-catenin: Asp19 of LEF-1 forms a salt bridge with Lys435 and Arg474 and nonpolar contacts with Cys429 and His470 of β-catenin, Met21 of LEF-1 packs against Asp390, Thr393, Asn426, and Cys429 of β-catenin, Ile22 of LEF-1 packs against Asn426, and Phe24 of LEF-1 packs against Arg386 and Gly422 (Sun and Weis, 2011). The conserved residues Asp19, Met21, Ile22, and Phe24 are also reported to participate in the binding event between other TCFs/β-catenin interactions (Graham et al., 2000) (Poy *et al*., 2001). Mutation of Asp19 of LEF-1 impairs the interaction between β-catenin and LEF-1 (Hsu *et al*., 1998).

The conservation in the HMG domain

The conserved amino acids in the HMG domain of LEF-1 are required for the DNA/protein interactions needed for regulating embryogenesis in each species. The HMG domain of LEF-1 recognizes the following sequence CCTTTGAAG that lies in the Wnt response elements of target genes (Arce *et al*., 2006). Several amino acids in

the HMG domain of LEF-1 have been shown to participate in the DNA/protein interaction. Hydrophobic residues on the concave surface formed by helices 1 and 2 (*i.e.* Ala 9, Phe 10, Met 11, Met 14, Val 22, Ala 31, and Gly 38), N-terminal residues (*i.e.* Lys 4, Lys 5, Pro 6, Leu 7, and Asn 8), and C-terminal residues (*i.e.* Ala 72, Arg 73, Tyr 76, Lys 79, Lys 81, and Arg 82) come into contact with the DNA to induce a bend of ~100° which facilitates the interactions between transcription factors bound at sites flanking the LEF-1 DNA binding site (Love *et al*., 1995). Mutation of residues 4, 5, 7, 11, 14, 38, 41, 52, 80, and 81 impairs or abrogates DNA binding (Giese *et al*., 1991).

2.2 Introduction – Sequence similarity of the HMG box within the HMGB family

A single amino acid sequence does not reveal any evolutionary information regarding conservation but when compared to homologous sequences a wealth of information may be disclosed. Homologous sequences often reveal well conserved amino acid residues that participate in protein stability or function. As a result, many of the conserved residues are located in hydrophobic cores (residues involved in protein stability) and/or in enzyme active sites or protein-protein or protein-ligand binding sites (residues involved in function)(Schueler-Furman and Baker, 2003). In addition, a family of homologous sequences often share a common three dimensional fold (Shih *et al*., 2012).

LEF-1 is a member of the well established HMGB family that is divided into two subclasses. The non-sequence specific subclass exhibits conformational stability in the absence of DNA whereas the sequence specific family members are usually less ordered. A blastp search was performed to identify the HMGB family members and an alignment of each family member with LEF-1 was performed. In designing more stable LEF-1 variants we mutated particular residue positions to amino acids that occurred most frequently at particular position within the HMG domain of LEF-1. As a result, four variants displayed an increase in thermal stability (*i.e.* variants 1-4) (Chapter 3). The success of the previous approach led to the design of four additional variants (*i.e.,* variants 5-8) from a more traditional structure-based approach. In addition, a suite of protein design algorithms (ORBIT) was utilized to calculate a best

fit amino acid sequence for the three dimensional structure of LEF-1(Dahiyat and Mayo, 1996).

2.2.1 Identification and sequence alignment of the HMG domain of human HMGB family members

The amino acid sequence of the human variant of LEF-1 (residues 296-380, NCBI Reference Sequence: NP_057353.1) was used as a query to identify HMGB proteins in the human taxonomic group (taxid: 9606) (http://www.expasy.org/tools/blast/) using the program blastp (query against the UniProt Knowledgebase (Swiss-Prot + TrEMBL)). The auto-select comparison matrix option was utilized and the threshold parameter was set at 10 for the blast search. The blast results (the 100 best scoring sequences) also provided an alignment of the HMG domain of each protein with respect to LEF-1 (Figure 2.2). Each protein was categorized as a sequence specific member or non-sequence specific member based on the references listed in Tables 2.2 and 2.3.

A

```
                10        20        30        40        50        60        70
LEF1      2 HIKKPLNAFMLYMKEMRANVVAESTLKESAAINQILGRRWHALSREEQAKYYELARKERQLHMQLYPGWSARDNYGK 78
TCF1    269  IKKPLNAFMLYMKEMRAKVIAECTLKESAAINQILGRRWHALSREEQAKYYELARKERQLHMQLYPGWSARDNYG 343
TCF3    345 HVKKPLNAFMLYMKEMRAKVVAECTLKESAAINQILGRKWHNLSREEQAKYYELARKERQLHSQLYPTWSARDNYG 420
TCF4    349 HIKKPLNAFMLYMKEMRAKVVAECTLKESAAINQILGRRWHALSREEQAKYYELARKERQLHMQLYPGWSARDNYG 424
SRY      60  VKRPMNAFIVWSRDQRRKMALENPRMRNSEISKQLGYQWKMLTEAEKWPFFQEAQKLQAMHREKYPNYKYR     130
CIC     199 HIRRPMNAFMIFSKRHRALVHQRHPNQDNRTVSKILGEWWYALGPKEKQKYHDLAFQVKEAHFKAHPDW        267
Sox1     51  VKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKVMSEAEKRPFIDEAKRLRALHMKEHPDYKYR     121
Sox2     41  VKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYR     111
Sox3    139  VKRPMNAFMVWSRGQRRKMALENPKMHNSEISKRLGADWKLLTDAEKRPFIDEAKRLRAVHMKEYPDYKYR     209
Sox4     58 HIKRPMNAFMVWSQIERRKIMEQSPDMHNAEISKRLGKRWKLLKDSDKIPFIREAERLRLKHMADYPDYKYR     129
Sox5    555 HIKRPMNAFMVWAKDERRKILQAFPDMHNSNISKILGSRWKAMTNLEKQPYYEEQARLSKQHLEKYPDY        623
Sox6    620 HIKRPMNAFMVWAKDERRKILQAFPDMHNSNISKILGSRWKSMSNQEKQPYYEEQARLSKIHLEKYPNY        688
Sox7     45  IRRPMNAFMVWAKDERKRLAVQNPDLHNAELSKMLGKSWKALTLSQKRPYVDEAERLRLQHMQDYPNYKYR     115
Sox8    101 HVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLSESEKRPFVEEAERLRVQHKKDHPDY        169
Sox9    104 HVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNESEKRPFVEEAERLRVQHKKDHPDY        172
Sox10   103 HVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNESDKRPFIEEAERLRMQHKKDHPDY        171
Sox11    48 HIKRPMNAFMVWSKIERRKIMEQSPDMHNAEISKRLGKRWKMLKDSEKIPFIREAERLRLKHMADYPDYKYR     119
Sox12    39 HIKRPMNAFMVWSQHERRKIMDQWPDMHNAEISKRLGRRWQLLQDSEKIPFVREAERLRLKHMADYPDYKYR     110
Sox13   405 HIKRPMNAFMVWAKDERRKILQAFPDMHNSSISKILGSRWKSMTNQEKQPYYEEQARLSRQHLEKYPDY        473
Sox14     7 HIKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSEAEKRPYIDEAKRLRAQHMKEHPDYKYR      78
Sox15    49  VKRPMNAFMVWSSAQRRQMAQQNPKMHNSEISKRLGAQWKLLDEDEKRPFVEEAKRLRARHLRDYPDYKYR     119
Sox17    68  IRRPMNAFMVWAKDERKRLAQQNPDLHNAELSKMLGKSWKALTLAEKRPFVEEAERLRVQHMQDHPNYKYR     138
Sox18    85  IRRPMNAFMVWAKDERKRLAQQNPDLHNAVLSKMLGKAWKELNAAEKRPFVEEAERLRVQHLRDHPNYKYR     155
Sox21     7 HVKRPMNAFMVWSRAQRRKMAQENPKMHNSEISKRLGAEWKLLTESEKRPFIDEAKRLRAMHMKEHPDYKYR      78
Sox30   336 HVKRPMNAFMVWARIHRPALAKANPAANNAEISVQLGLEWNKLSEEQKKPYYDEAQKIKEKHREEFPGW        404
HBP1    435   KRPMNAFMLFAKKYRVEYTQMYPGKDNRAISVILGDRWKKMKNEERRMYTLEAKALAEEQ                494
BBX      81   RRPMNAFLLFCKRHRSLVRQEHPRLDNRGATKILADWWAVLDPKEKQKYTDMAKEYKDAFMKANPGY        147
TFAM     51   KKPVSSYLRFSKEQLPIFKAQNPDAKTTELIRRIAQRWRELPDSKKKIYQDAYRAEWQVYKE              112
```

B

```
HMGB1    96   KRPPSAFFLFCSEYRPKIKGEHPGLSIGDVAKKLGEMWNNTAADDKQPYEKKAAKLKE                  153
HMGB2    96   KRPPSAFFLFCSEHRPKIKSEHPGLSIGDTAKKLGEMWSEQSAKDKQPYEQKAAKLKE                  153
HMGB3    94   KRPPSGFFLFCSEFRPKIKSTNPGISIGDVAKKLGEMWNNLNDSEKQPYITKAAKLKE                  151
HMGB4    94   RRPPSSFLLFCQDHYAQLKRENPNWSVVQVAKATGKMWSTATDLEKHPYEQRVALLRAKYFEELELY         160
HMG20A  104   KSPLTGYVRFMNERREQLRAKRPEVPFPEITRMLGNEWSKLPPEEKQRYLDEADRDKERYMK               165
HMG20B   71   KAPVTGYVRFLNERREQIRTRHPDLPFPEITKMLGAEWSKLQPTEKQRYLDEAEREKQQYMKELRAYQQSEAY  143
UBF1    113   KKPLTPYFRFFMEKRAKYAKLHPEMSNLDLTKILSKKYKELPEKKKMKYIQDFQREKQ                  170
SSRP1   548   KRPMSAYMLWLNASREKIKSDHPGISITDLSKKAGEIWKGMSKEKKEEWDRKAEDARRDYEKAMKEYEG       616
UBTFL1  101   KRPLTAYNRFFKESWPQYSQMYPGMRSQELTKILSKKYRELPEQMKQKYIQDFRKEKQ                  158
Tox     262   QKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDGLGEEQKQVYKKKTEAAK                   318
Tox2    256   QKPVSAYALFFRDTQAAIKGQNPSATFGDVSKIVASMWDSLGEEQK                              301
Tox3    256   QKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDSLGEEQKQVYKRKTEAAKKE                 314
Tox4    224   QKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDSLGEEQKQVYKRKTEAAKKE                 282
```

Figure 2.2. Multiple sequence alignment of the HMG-box domain from members of the HMGB family. Sequence alignment of members of the sequence specific subclass (A) and members of the non-sequence specific subclass (B) of HMGB proteins. The protein name for each abbreviation is listed in Tables 2.1 and 2.2. The numbering scheme at the top of the figure refers to the amino acid positions within the HMG domain of LEF-1.

Table 2.2: Abbreviation and protein name for sequence specific HMGB family members.

Sequence specific HMGB family members

Abbreviation	Protein Name	Accession Number	Reference
LEF-1	Lymphoid enhancer-binding factor 1	Q9UJU2	Love *et al.*, 1995
TCF-1	T-cell specific transcription factor 1	P36402	van de Wetering *et al*., 1991
TCF-3	T-cell specific transcription factor 3	Q9HCS4	Brantjes *et al.*, 2002
TCF-4	T-cell specific transcription factor 4	Q9NQB0	Brantjes *et al*., 2002
SRY	Testis determining factor	Q05066	King *et al.*, 1993
CIC	Protein capicua homolog	Q96RK0	Malki *et al*., 2010
Sox1	Transcription factor Sox1	O00570	Malki *et al*., 2010
Sox2	Transcription factor Sox2	P48431	Malki *et al*., 2010
Sox3	Transcription factor Sox3	P41225	Malki *et al*., 2010
Sox4	Transcription factor Sox4	Q06945	Malki *et al*., 2010
Sox5	Transcription factor Sox5	P35711	Malki *et al*., 2010
Sox6	Transcription factor Sox6	P35712	Malki *et al*., 2010
Sox7	Transcription factor Sox7	Q9BT81	Malki *et al*., 2010
Sox8	Transcription factor Sox8	P57073	Malki *et al*., 2010
Sox9	Transcription factor Sox9	P48436	Malki *et al*., 2010
Sox10	Transcription factor Sox10	P56693	Malki *et al*., 2010
Sox11	Transcription factor Sox11	P35716	Malki *et al*., 2010
Sox12	Transcription factor Sox12	O15370	Malki *et al*., 2010
Sox13	Transcription factor Sox13	Q9UN79	Malki *et al*., 2010
Sox14	Transcription factor Sox14	O95416	Malki *et al*., 2010
Sox15	Transcription factor Sox15	O60248	Malki *et al*., 2010
Sox17	Transcription factor Sox17	Q9H6I2	Malki *et al*., 2010
Sox18	Transcription factor Sox18	P35713	Malki *et al*., 2010
Sox21	Transcription factor Sox21	Q9Y651	Malki *et al*., 2010
Sox30	Transcription factor Sox30	O94993	Malki *et al*., 2010
HBP1	HMG box-containing protein 1	O60381	Sampson *et al*., 2001
BBX	Bobby sox homolog	Q8WY36	Sanchez Diaz *et al*., 2001
TFAM	Transcription factor A, mitochondrial	Q00059	Wong *et al.*, 2009

Table 2.3: Abbreviation and protein names for non-sequence specific HMGB family members.

Non-sequence specific HMGB family members

Abbreviation	Protein Name	Accession Number	Reference
HMGB1	High mobility group protein B1	P09429	Stros, 2010
HMGB2	High mobility group protein B2	P26583	Stros, 2010
HMGB3	High mobility group protein B3	O15347	Stros, 2010
HMGB4	High mobility group protein B4	Q8WW32	Stros, 2010
HMG20A	High mobility group protein 20A	Q9NP66	Sumoy *et al*., 2000
HMG20B	High mobility group protein 20B	Q9P0W2	Sumoy *et al*., 2000
SSRP1	FACT complex subunit SSRP1	Q08945	Stros *et al*., 2007
UBF1	Upstream binding factor 1	P17480	Liu *et al*., 2009
UBTFL1	Putative upstream-binding factor 1-like protein 1	P0CB47	Yamada *et al.*, 2010
Tox	Thymus high mobility group box protein tox	O94900	O'Flaherty and Kaye 2003
Tox2	TOX high mobility group box family member 2	Q96NM4	O'Flaherty and Kaye 2003
Tox3	TOX high mobility group box family member 3	O15405	O'Flaherty and Kaye 2003
Tox4	TOX high mobility group box family member 4	O94842	O'Flaherty and Kaye 2003

2.2.2 Rational and computational design of LEF-1 Variants

Four variants were created using homology modeling with sequence data from the HMGB family of proteins (described in further detail in Section 2.2.5) and four variants were created using structural comparisons to relatively stable variants from the HMGB family (described in further detail in Section 2.2.5). No mutations were made to amino acid residues that came into contact with the DNA (*i.e.* amino acid residues within 4 angstrom of the DNA). The amino acid residues within 4 angstrom of the DNA in the 12 NMR structures of the HMG domain of LEF-1 (PDB code 2LEF) were determined with the Pymol software package. The final variant, variant 9, was designed using the output from the Optimization of Rotamers Based on Iterative Techniques (ORBIT) suite of protein design algorithms developed by the Mayo laboratory (Dahiyat and Mayo, 1996). The average of twelve NMR structures of the HMG domain of LEF-1 (without the coordinates of the bound DNA) (PDB code 2LEF) were used as the starting template for computational calculations. The conformations of amino acid substitutions at five positions (*i.e.,* 13, 17, 37, 44, and 52, Figure 2.3A) were represented with a set of backbone independent side chain rotamers derived from the rotamer library of Dunbrack and Cohen (Dunbrack and Cohen, 1997). Positions F10, E16, R40, W41, E48, Q49, K51, and Y53 (Figure 2.3B) were floated (*i.e.,* the amino acid conformation but not the amino acid identity was allowed to vary to allow for conformational adjustments). The energies of all possible combinations of the considered amino acids (A, V, L, I, F, Y, and W) at the chosen positions (13, 17, 37, 44, and 52) were calculated using a force field containing terms

describing van der Waals, solvation, electrostatic, and hydrogen bond interactions. The resulting energies were used as an input for a search algorithm referred to as the dead end elimination (DEE) algorithm (Maeyer *et al*., 2000). DEE utilizes the results of a molecular mechanics scoring function to identify one optimal sequence and conformation with the lowest energy called the optimal (ground state). The optimal sequence returned from ORBIT is the following: F13, Y17, L37, L44, F52. ORBIT returned wild type sequences at position 37 and 44.

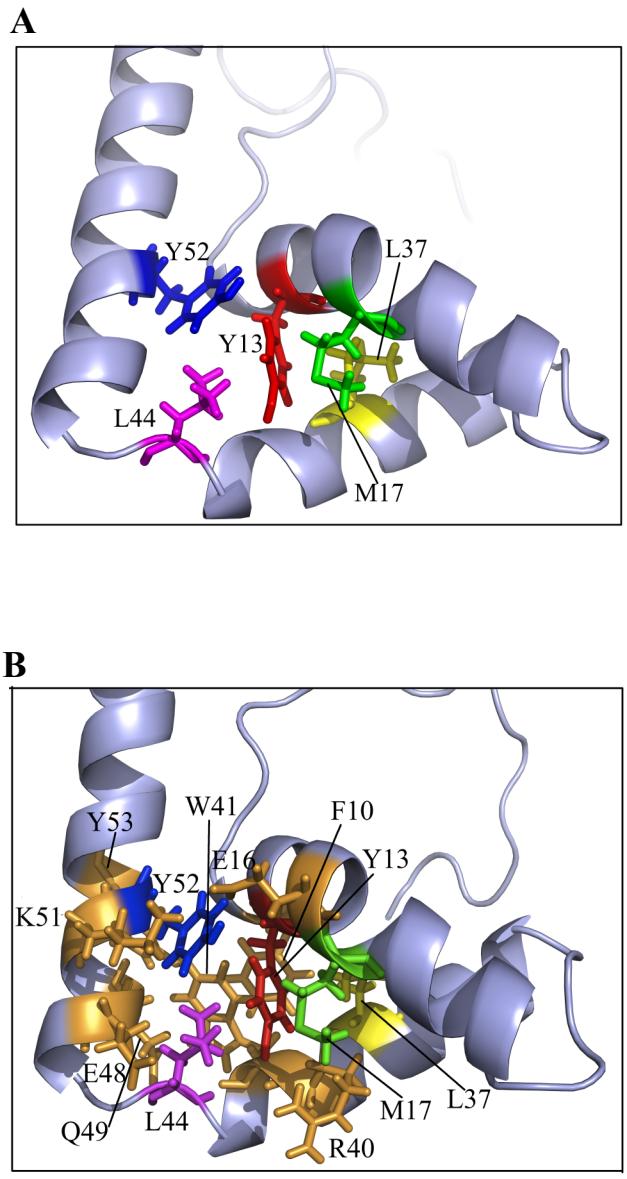


Figure 2.3 Selected amino acid positions for computational mutagenesis. A, computational calculations were performed with ORBIT on the five amino acid positions shown above: Y13 (red), M17 (green), L37 (yellow), L44 (magenta), and Y52 (blue). **B,** the conformations of the following amino acids: F10, E16, R40, W41, E48, Q49, K51, and Y53 (orange) were allowed to vary for computational calculations.

2.2.3 Construction of the LEF-1 Variants

The gene for the wild-type version of the LEF-1 HMGB domain was obtained from Dr. John Love and is contained in a pET21a(+) plasmid (Novagen). The restriction enzyme sites that flank the HMG-domain of LEF-1 are Nde I (5' end) and EcoRI (3' end) (Figure 2.4).

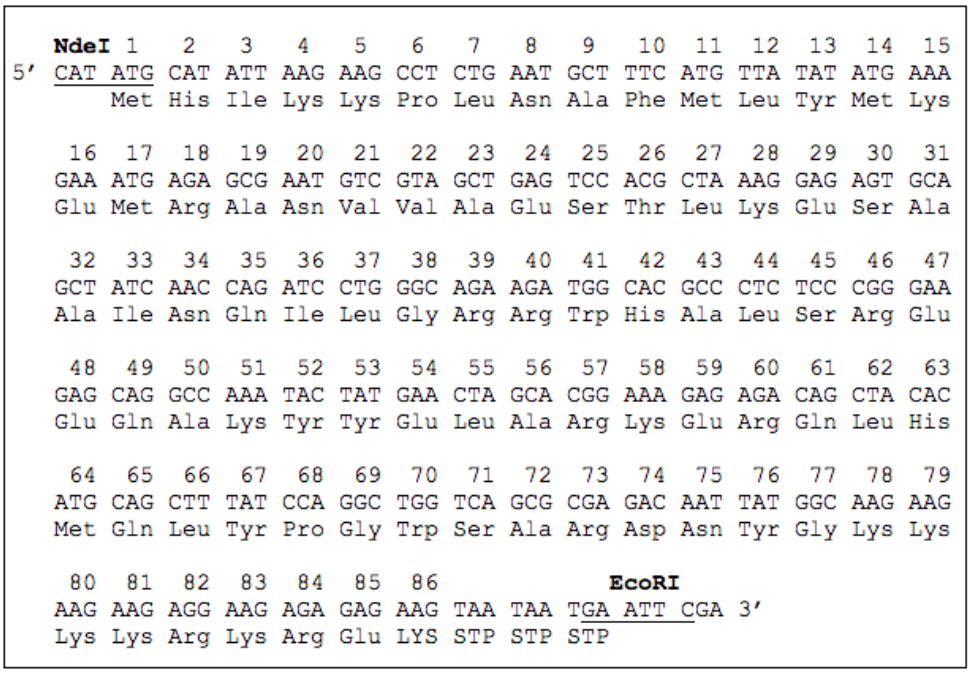


Figure 2.4. DNA and amino acid sequence of the HMG domain of LEF-1. The DNA sequence of the HMG domain of LEF-1 was cloned into the pET21a(+) vector. NdeI and EcoRI sites are underlined.

All DNA oligonucleotides that were used for site-directed mutagenesis and total-gene synthesis were obtained from Integrated DNA technologies, Inc. and purified by electrophoresis through a denaturing polyacrylamide gel. Oligonucleotides were visualized by UV and gel slices were crushed and soaked in 1X

SSC (0.15 M Sodium Chloride and 0.015 M sodium citrate pH 7) for 20 hours. Eluted oligonucleotides were ethanol precipitated, air dried, and resuspended in water.

Variants 1-4 were created using site directed mutagenesis. Site directed mutagenesis involves the design of two complimentary primers containing the mutation of interest. An example of two primers used to create variants 2 and 4 is shown below in Figure 2.5.

Variant 2 and 4 complimentary primers

5′CTC TCC CGG GAA GAG AAG GCC AAA TAC TAT G 3′

5′CAT AGT ATT TGG CCT TCT CTT CCC GGG AGA G 3′

Figure 2.5 Site directed mutagenesis primers for variants 2 and 4. The nucleotide highlighted in red is the mutation incorporated into wild type or variant 3 to create the lysine at position 49 during PCR.

The two complimentary primers, plus DNA parental template, 50 μM dNTPs, and PfuTurbo polymerase were incubated in an appropriate buffer. The mixture was subjected to the following cycling parameters to generate the DNA template with the desired mutation: 1 cycle of 30 seconds at 95°C, 16 cycles of 30 seconds at 95°C, 1 minute at 55°C, 5 minutes at 68°C, 1 cycle of 5 minutes at 4°C. To digest the methylated parental strand 10 U of DpnI enzyme was added to the mixture and incubated at 37°C overnight. 1 uL of the DpnI treated mixture was transformed into XL1-Blue electrocompetent cells and plated on 100 mg/ml AMP resistant agar plates. Colonies were then selected for DNA amplification and DNA sequencing was

performed to confirm the presence of the desired mutation and that the gene was in the correct reading frame.

Variants 5-9 were created using total gene synthesis. Six overlapping oligonucleotides were designed for each variant (Figures 2.6-2.10) and PCR was used to generate the entire gene.

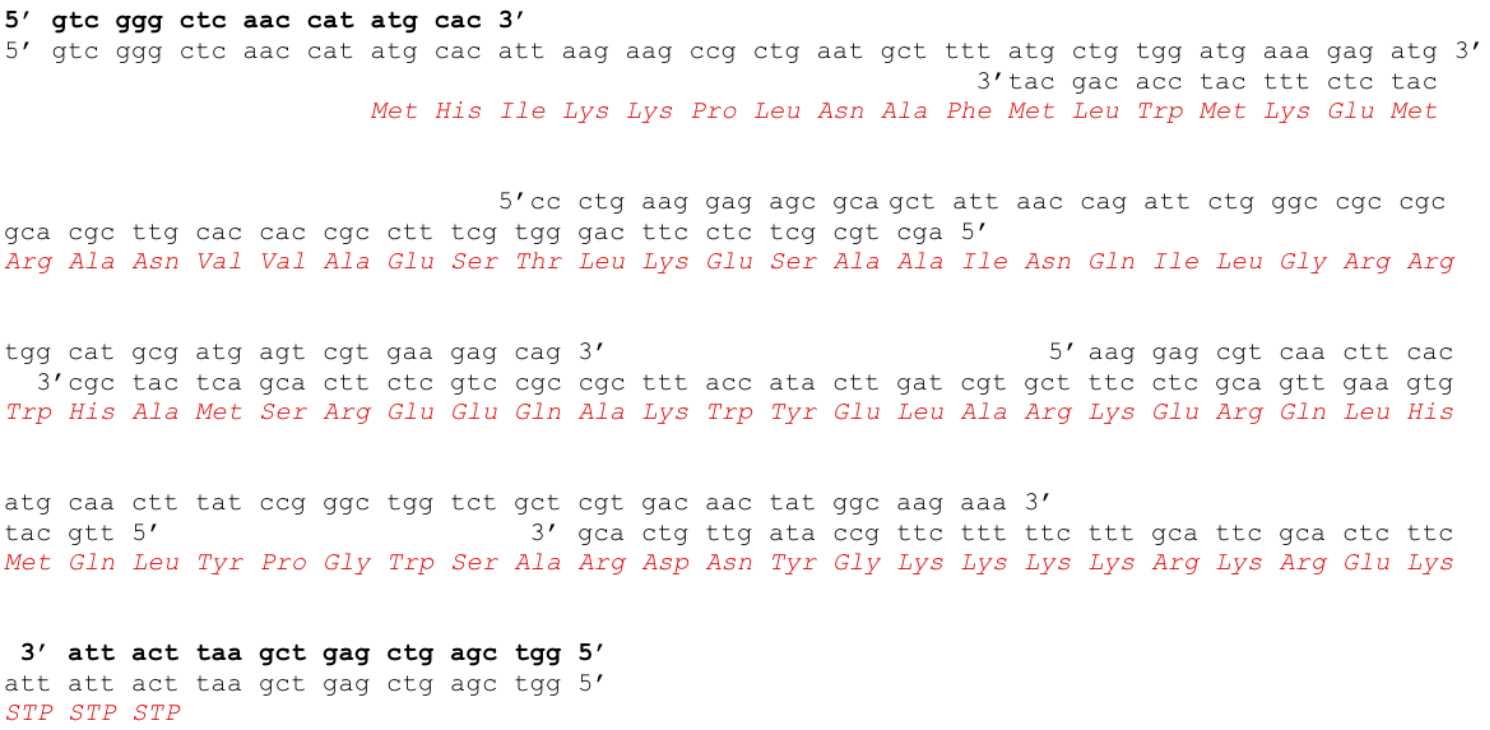


Figure 2.6 Sequences of the six overlapping oligonucleotides used to assemble the gene for variant 5. The three oligonucleotides in the upper row correspond to the coding strand and the oligonucleotides in the lower row correspond to the non-coding strand. The amino acid sequence of LEF-1 is shown in red and the short sequences (bold text) at the termini are primers used in the second PCR.

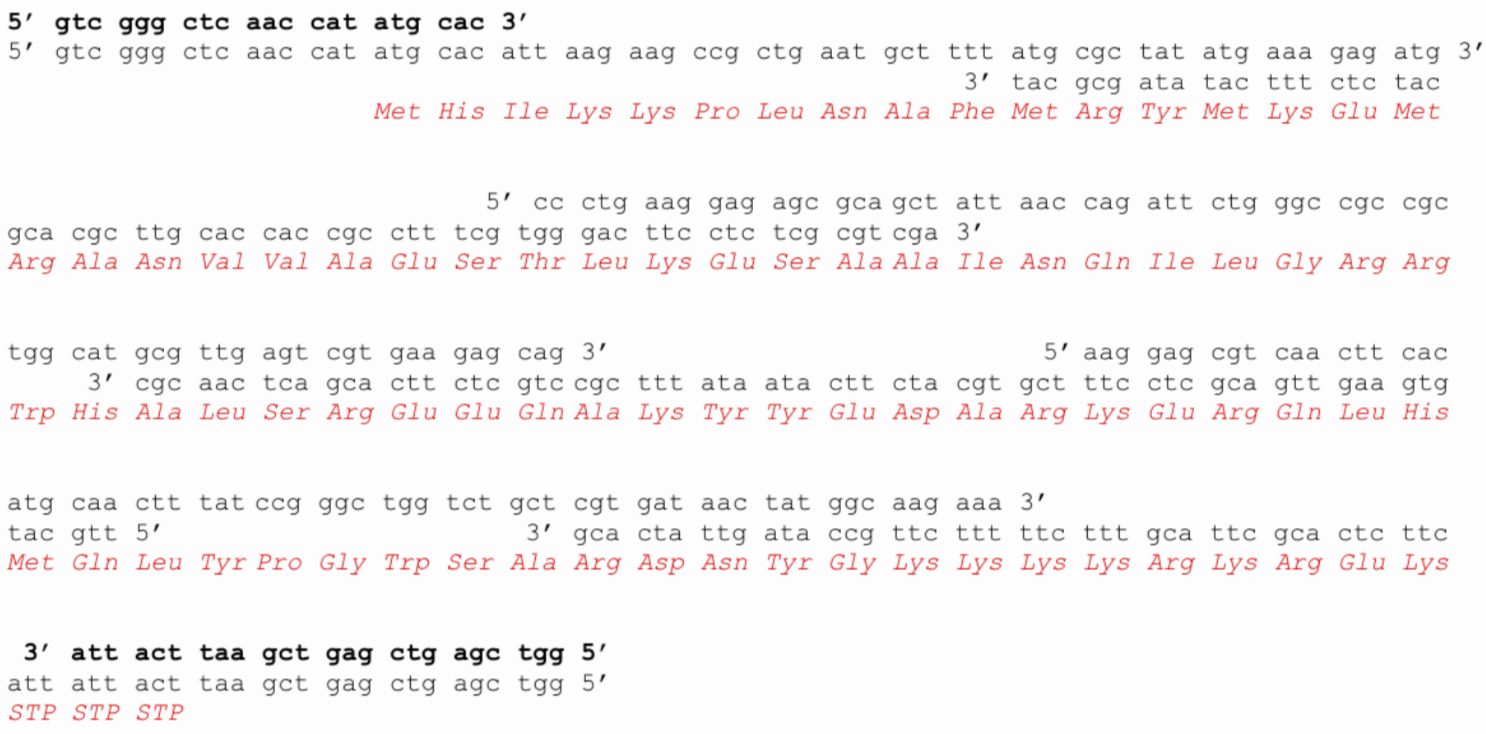


Figure 2.7 Sequences of the six overlapping oligonucleotides used to assemble the LEF-1 gene (variant 6). The three oligonucleotides in the upper row correspond to the coding strand and the oligonucleotides in the lower row correspond to the non-coding strand. The amino acid sequence of LEF-1 is shown in red and the short sequences (bold text) at the termini are primers used in the second PCR.

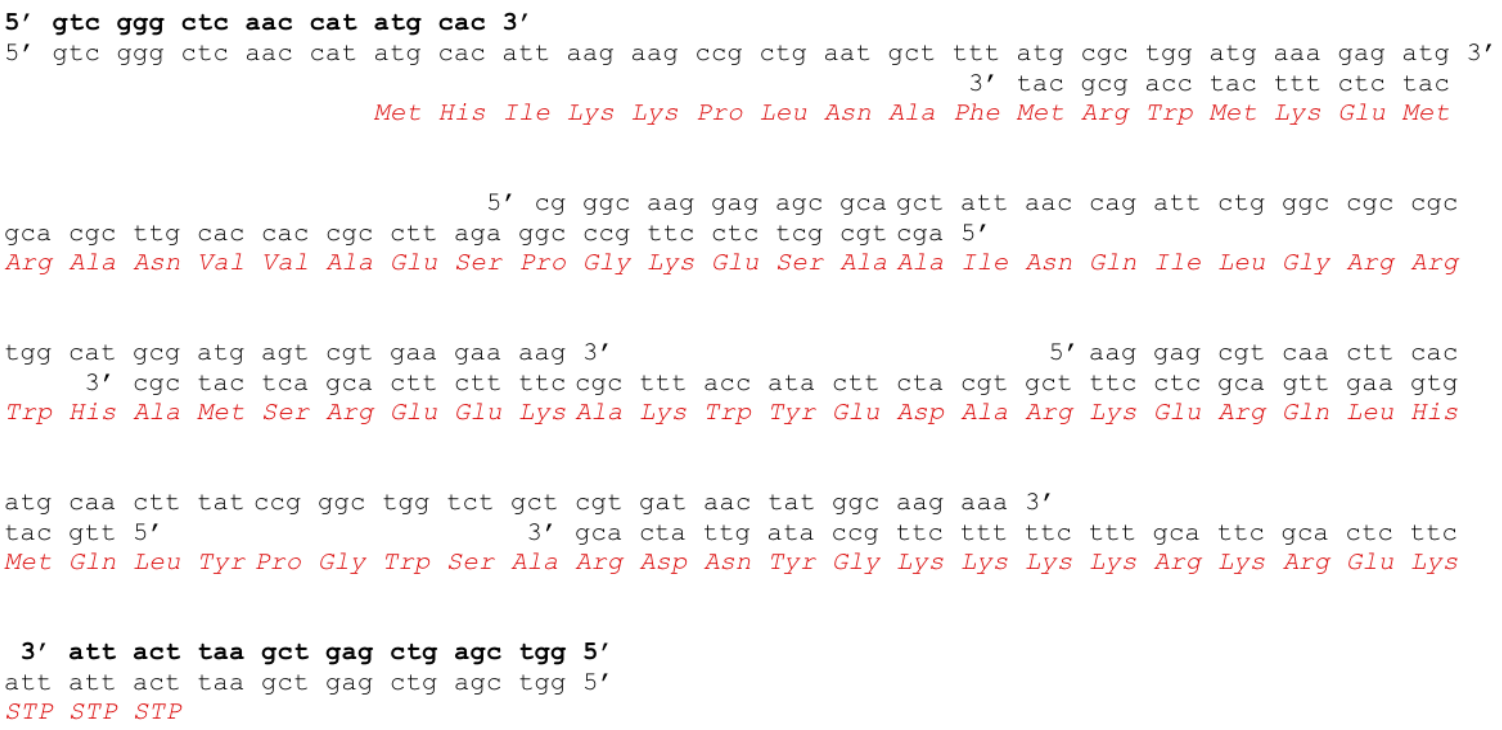


Figure 2.8 Sequences of the six overlapping oligonucleotides used to assemble the LEF-1 gene (variant 7). The three oligonucleotides in the upper row correspond to the coding strand and the oligonucleotides in the lower row correspond to the non-coding strand. The amino acid sequence of LEF-1 is shown in red and the short sequences (bold text) at the termini are primers used in the second PCR.

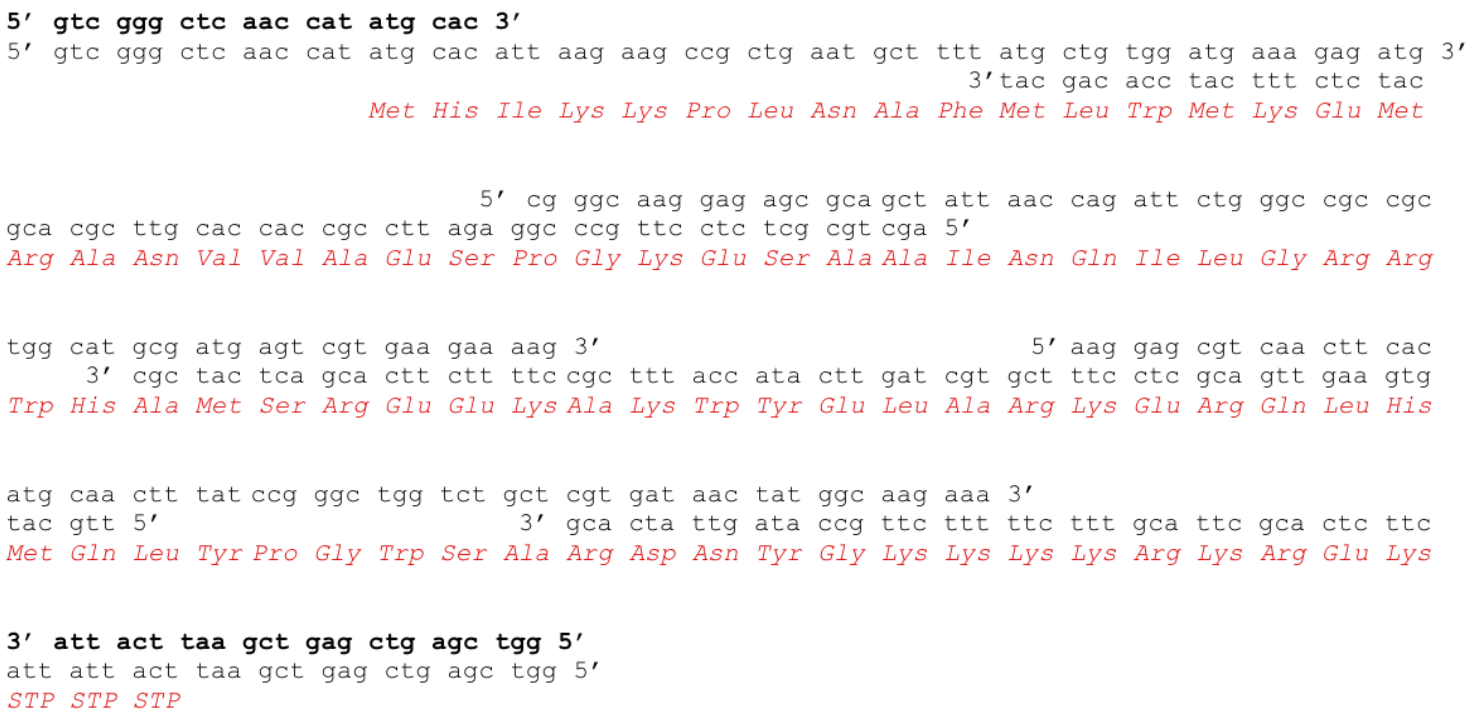


Figure 2.9 Sequences of the six overlapping oligonucleotides used to assemble the LEF-1 gene (variant 8). The three oligonucleotides in the upper row correspond to the coding strand and the oligonucleotides in the lower row correspond to the non-coding strand. The amino acid sequence of LEF-1 is shown in red and the short sequences (bold text) at the termini are primers used in the second PCR.

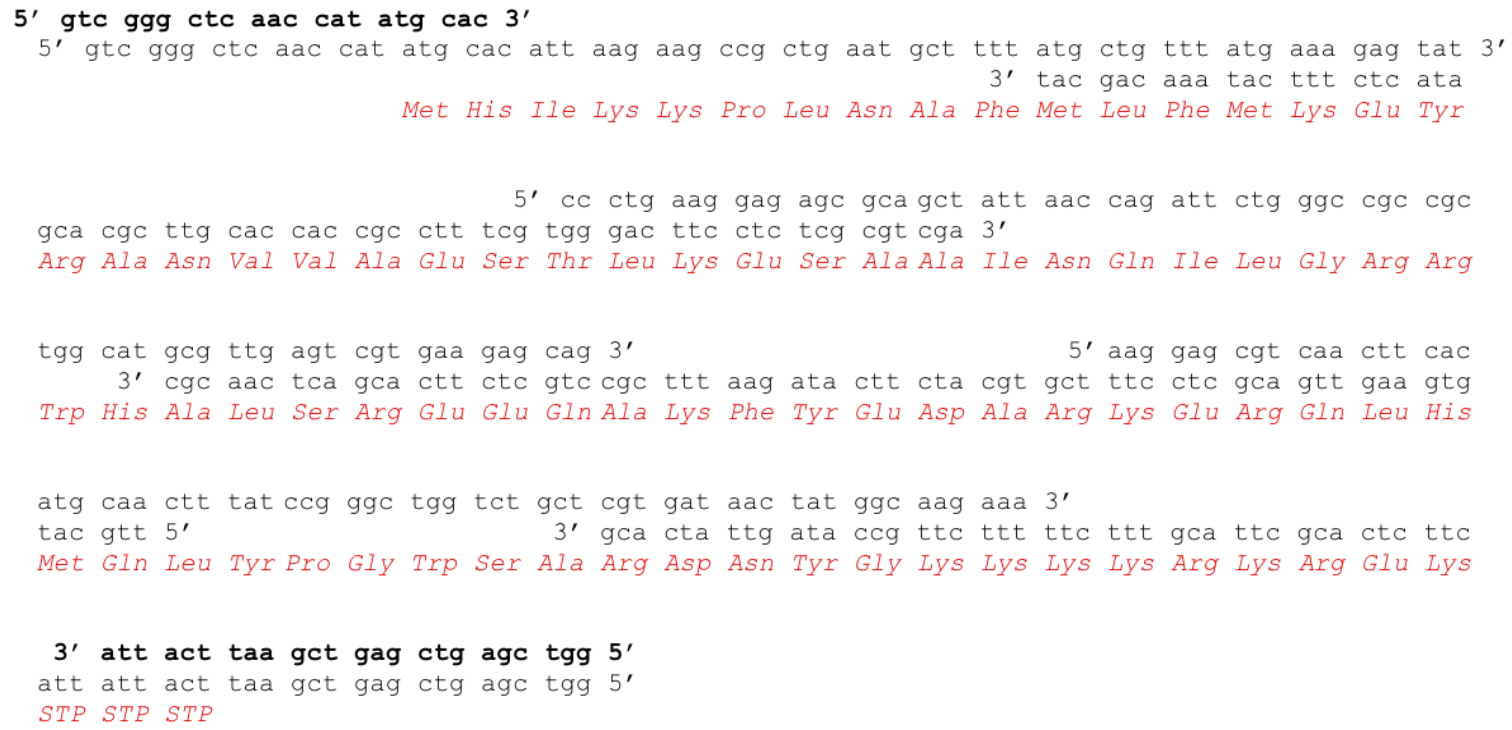


Figure 2.10 Sequences of the six overlapping oligonucleotides used to assemble the LEF-1 gene (variant 9). The three oligonucleotides in the upper row correspond to the coding strand and the oligonucleotides in the lower row correspond to the non-coding strand. The amino acid sequence of LEF-1 is shown in red and the short sequences (bold text) at the termini are primers used in the second PCR.

It should be noted that, when total gene synthesis was utilized to create a LEF-1 variant, that the rare arginine codons found in the wild-type LEF-1 gene (*i.e.,* AGG and AGA) were exchanged for one of the four higher frequency arginine codons utilized by bacteria (*i.e.,* CGX where X corresponds to all four bases). For total-gene synthesis two rounds of PCR were required for generating the full length gene. In the first round all six overlapping oligomers were incubated with 50 uM dNTPs, PfuTurbo Buffer (Strategene), and PfuTurbo Polymerase. The mixture was subjected to the following temperature cycling parameters to generate the appropriate variant: 25 cycles of 30 seconds at 95°C, 30 seconds at 59°C, 2 minutes at 72°C, 1 cycle of 10 minutes at 72°C, 1 cycle of 10 minutes at 4°C. In the second round of PCR, 10 uL of PCR product from round one was incubated with primers listed in Figures 2.3-2.7, and 50 μM dNTPs, PfuTurbo Buffer (Strategene), and PfuTurbo Polymerase. This mixture was subjected to the following cycling parameters: 30 cycles of 1 minute at 95°C, 1 minute at 52°C, 2 minutes at 72°C, 1 cycle of 10 minutes at 72°C, and 1 cycle of 10 minutes at 4°C. PCR products from the second round of amplification were isolated and purified via agarose gel electrophoresis (Sambrook et al, 1989). The resolved bands were excised and extracted from the 1% agarose gel by utilizing the gel extraction kit developed by Qiagen. The synthesized genes and the pET21a(+) vector were incubated with NdeI and EcoRI at 37° C for six hours and underwent another round of agarose gel electrophoresis isolation, purification, and extraction, as mentioned above. The digested genes and vector were ligated with 5U of T4 Ligase (Invitrogen) and 1 uL of that mixture was electroporated into XL1-Blue cells. The

transformed cells were plated on agar that contained 100 mg/mLampicillin. . Colonies were selected for DNA amplification and DNA sequencing was performed to confirm the correct gene sequence and to verify the correct reading frame.

2.2.4 Protein Expression and Purification

The pET21a(+) expression system was used for the expression of LEF-1 and its variants in *Escherichia coli* BL21 (DE3) cells. It was previously demonstrated that, due to the presence of six rare arginine codons in the LEF-1 sequence, that protein expression was unsuccessful in *E. coli.* To address this issue it was necessary to cotransform the bacteria with the plasmid pUBS520 (dnaY), which constitutively expresses the rare $tRNA_{AGG/AGA}$. This plasmid contains the *E. coli dnaY* gene which codes for the minor arginine codons AGA and AGG (Garcia *et al*., 1986; Brinkman *et al*., 1989). The cells were grown in Luria-Bertani medium at 37°C and induced with 1.0 mM IPTG at an OD of 1.4. The cells were harvested and subjected to a freeze/thaw cycle, sonication, and centrifugation at 6,000 rpm. Upon PAGE analysis the LEF-1 variants were found to be located in the insoluble pellet after sonication and centrifugation (*i.e.,* inclusion bodies). The LEF-1 variants in the pellet were solubilized with 6M guanidine hydrochloride. Impurities were then separated from the resolubilized LEF-1 variants upon centrifugation at 6,000 rpm. The resulting supernatant was dialyzed against 50 mM potassium phosphate (pH 6.5) and after centrifugation at 6,000 rpm the supernatant containing the protein of interest was purified by HPLC and lyophilized. The purity of the protein was confirmed on SDS-PAGE gel and the molecular weight of the protein was confirmed with mass spectrometry.

2.2.5 Results and Discussion

Sequence comparison of the protein sequences obtained via a blast search against the wild-type LEF-1 sequence revealed that LEF-1 is one of the only members of the family that does not have a proline residue at position 26 (Figure 2.2). Other members of the HMGB family contain a 25H-26P-27G or 25H-26P-27A sequence in this region. These sequences are located in the turn 1 region of the structure of LEF-1, just beyond the C-terminus of helix one and may function to break helix I and stabilize the subsequent turn (Figure 2.11). Based on this information, two variants were created using site directed mutagenesis (*i.e.,* T26P and TL26,27PG). Initially, two dimensional [^{15}N, ^{1}H] HSQC spectra were collected on these variants (Figure 4.3, Chapter 4) and while the spectrum for the single mutant, T26P showed no appreciable improvement relative to the wild-type spectrum the double mutant, TL26,27PG did reveal an increased number of peaks and an improvement in signal quality and resolution. These findings indicate that the introduction of the 'helix breaking' Pro-Gly di-peptide into this region possibly results in a more compact fold and improved relaxation properties.

The success obtained for the double mutant, TL26,27PG led us to pursue the possibility of enhancing the overall stability of the free LEF-1 domain by redesigning certain residue positions such that they contained well conserved amino acids from the HMGB family. For example, the blast search also revealed that the LEF-1 sequence is one of the only sequences that does not contain a Lys residue at position 49 (for the

WT LEF-1 sequence a glutamine residue is found at position 49). This mutation is incorporated into variants 2, 4, 7 and 8.

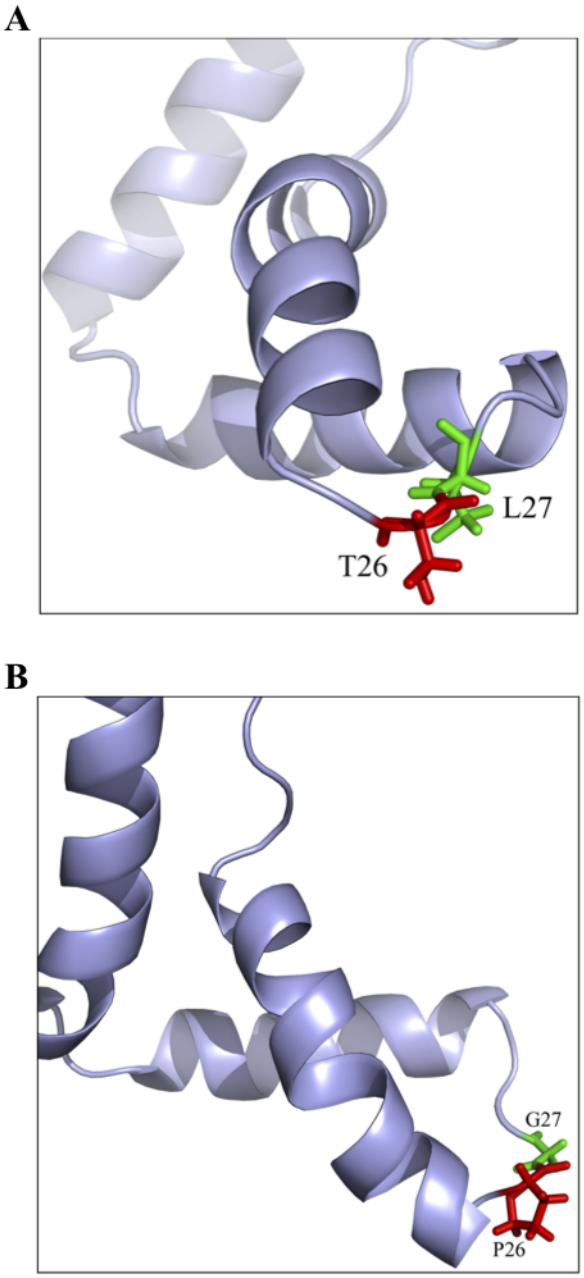


Figure 2.11 Turn 1 region of the HMG domain of LEF-1. Wild type LEF-1 (PDB: 2LEF) has a Thr 26 and Leu 27 in the turn 1 region of the HMG domain (**A**). Other HMGB family members (*i.e.* HMGB1, PDB: 1HME) have a Pro 26 and Gly 27 or Pro 26 and Ala 27 in the first turn region of the HMG domain (**B**).

Structural analysis combined with sequence homology comparisons of the HMGB family indicate that four residue positions in the central hydrophobic core may contribute to the overall stability of the HMG domain. These positions include Tyr-13, Trp-41, Leu-44, and Tyr-52. In an attempt to increase the overall stability of the LEF-1 domain the amino acids at these four positions from more stable variants (*e.g.,* SSRP1, HMGD) were introduced into the LEF-1 sequence (*i.e.,* Y13W, L44M, Y52W; position 41 in the LEF-1 sequence is already a tryptophan).

Structural comparisons of HMGB domains, for which structures are available, also reveal that a salt bridge between helix 1 and 3 is found in some members of the HMGB family. The following mutation in LEF-1, L12R and L55D was introduced into the LEF-1 sequence to potentially form a salt bridge between helix 1 and 3.

The last variant, variant 9, was created using the suite of protein design algorithms collectively referred to as ORBIT. The five positions that were chosen for computational mutagenesis and energy assessment are positions 13, 17, 37, 44, and 52. These positions contribute to the formation of the major hydrophobic core of LEF-1. The wild type amino acids at the 5 selected positions are Y13, M17, L37, L44, and Y52. The ORBIT calculation on the DNA free LEF-1 structure returned the following results F13, Y17, L37, L44, and F52 with a calculated energy of -156 kcal/mol. The calculated results reflect amino acids found in members of HMGB family. Alignment of the HMGB family (Figure 2.1) reveals that at positions 13, 17, and 52 aromatic residues are conservatively substituted (*i.e.,* residues interchange between F, Y, and W at position 13; residues interchange between F, Y, and H at position 17; residues

interchange between F and Y at position 52). Positions 13, 17, and 52 play an important role in maintaining the integrity of the protein fold (i.e. L-shaped fold) and are putatively conserved for structural reasons. Interestingly, aromatic residues (i.e. residues interchange between F, Y, and H) are only found at position 17 in non-sequence specific members, HMGB1, 2, 3, and 4. The introduction of F13 and Y17 creates a π-π stacking interaction between the aromatic rings of the amino acid residue that stabilizes the protein fold.

At this point it is important to note that, in total, 9 mutant variants were created that incorporated different sets of mutations. In some cases single point mutations were incorporated into the LEF-1 sequence and in other cases multiple mutations were introduced. The sequence changes for all nine mutant variants, plus the corresponding WT residues, are listed below in Table 2.4.

Table 2.4: Variants of LEF-1 made from rational and computational design.

	Residue Position								
	12	13	17	26	27	44	49	52	55
Variant									
WT	L	Y	M	T	L	L	Q	Y	L
1				P					
2							K		
3				P	G				
4				P	G		K		
5		W				M		W	
6	R								D
7	R	W		P	G	M	K	W	D
8		W		P	G	M	K	W	
9		F	Y					F	

Summary. Sequence homology analysis of LEF-1 has provided insight into the conserved residues that may participate in the function and structure of LEF-1. The sequence alignment of the human LEF-1 variant against four different vertebrates showed two functional domains to be well conserved. The conserved amino acids in the β-catenin domain and HMG domain are required for two binding events to occur in the Wnt signaling pathway that is responsible for regulating development from early axis specification to organogenesis.

The sequence comparison of the LEF-1 HMG domain against other HMGB family members revealed conserved residues that play a role in the stability of the HMG domain. Three residues located in the turn regions of LEF-1 were found to be highly conserved in the HMGB family. These residues were introduced into four variants and all demonstrated an increase in thermal stability (Chapter 3). The traditional structure-based approach was used next to design four additional variants. In addition, a suite of protein design algorithms were used to create a more thermostable variant.

Chapter 3

The stability of LEF-1 and its variants

3.1 Introduction

The HMG domain of LEF-1 exhibits a disorder-to-order transition upon binding to DNA (Love, *et al*., 2004). The structural instability of the HMG domain prior to binding DNA provides an excellent opportunity to introduce stability into the 86 amino acid domain. When bound to DNA, LEF-1 shares the characteristic L-shaped fold of non-sequence specific HMGB proteins. LEF-1 may have evolved from non-sequence specific family members and thus sequence homology analysis was utilized as a means to impart stability in the HMG domain of LEF-1 (Chapter 2).

In this chapter we describe the structural stability of the HMG domain of LEF-1 and its variants as determined using circular dichroism (CD), NMR, and *in vitro* proteolysis. Thermodynamic characterization of wild type and the nine variants were assessed with CD by monitoring the thermal denaturation at $\lambda_{222\ nm}$. We then show the difference in chemical shift dispersion in the two dimensional [^{1}H, ^{15}N] HSQC spectrum of LEF-1 and the variants. This analysis is followed by the results of *in vitro* proteolysis, where the intrinsic stabilities of LEF-1 and the variants are further characterized.

This introductory section is followed by the materials and methods used to study LEF-1 and the variants, followed by CD results, a comparison of the two dimensional [^{1}H, ^{15}N] HSQC spectra of LEF-1 and the different variants, and concluding remarks on the *in vitro* proteolysis analysis. This chapter concludes with a

discussion on the components of the HMG domain that contribute to the overall stability of the HMG domain of LEF-1.

3.2 Materials and Methods

Construction and Expression of LEF-1 variants— The construction and expression of LEF-1 and its variants used in the circular dichroism experiments are as described in Chapter 2. Proteins used in the NMR experiments are described as follows. Calcium chloride chemically competent BL21(DE3) cells were first transformed with the plasmid pUBS520. The transformed cells were plated onto agar plates containing 50 ug/ml of kanamycin and were subjected to a second round of calcium chloride treatment (Sambrook *et al*., 1989) for transformation of pET-21A containing LEF-1 or the another variants. The transformed cells were plated on 100 ug/ml ampicillin and 50 ug/ml kanamycin minimal media agar plates. Minimal media consisted of standard M9 salts (Sambrook *et al*., 1989) with 2 grams of $^{15}NH_4SO_4$ and 6 grams of ^{12}C glucose per liter. The M9 minimal media was supplemented with Basal Medium Eagle Vitamin Solution (GibcoBRL). A colony was inoculated into 10 ml of minimal media overnight with no shaking at 21°C. The culture was then incubated at 37°C until it reached an OD_{600} ~ 0.6 which then was inoculated into 1 liter of minimal media. The cells were grown at 37°C to an OD_{600} ~1.3 to 1.5 and induced with 1 mM IPTG for 4 hours. The cells were harvested, subjected to a freeze/thaw cycle, sonication, and centrifugation at 6,000 rpm to produce a pellet that contained the protein of interest and the insoluble fraction of the lysed cells (inclusion bodies). The pellet was solubilized with 6M guanidine hydrochloride and impurities were isolated by centrifugation at 6,000 rpm. The remaining supernatant was dialyzed against 50 mM potassium phosphate (pH 6.5) and after centrifugation at 6,000 rpm the supernatant

containing the protein of interest was purified by HPLC and lyophilized. The purity of the protein was confirmed on SDS-PAGE gel and the molecular weight of the protein was confirmed with liquid chromatography mass spectrometry.

Circular Dichroism

The CD data was collected on a Jasco-810 Spectrometer (located at the University of San Diego) and equipped with a thermoelectric unit and using a 0.1 mm path length cell. Protein sample concentrations were 50 μM in 10 mM potassium phosphate and 50 mM potassium chloride at pH 7. Thermal melts were monitored at 222 nm. Data was collected every 1°C with an equilibration time of 2 minutes. For the thermal denaturation curves, the data was normalized by first shifting the points linearly such that the $[\theta]_{222nm}$ value at 4°C was zero. A scaling factor was obtained for each set by dividing the maximum $[\theta]_{222nm}$ value for all sets at 95°C by the $[\theta]_{222nm}$ value at 95°C for each set. All the data points for each set were then scaled by the unique scaling factor calculated for each set.

Nuclear Magnetic Resonance

NMR spectra were acquired on Varian spectrometers operating at 600 MHz. Protein concentrations were usually ~1 mM in 90% H_2O/10% ^2H_2O mixture containing 50 mM potassium phosphate pH 7.0, 25 mM potassium chloride, 50 μM NaN_3, and 10 μM EDTA. Standard 2D [^1H, ^{15}N] HSQC spectra were collected at 20°C for all variants. The programs NMRDraw and NMRPipe were used to process the NMR data, and the

program NMRView (One Moon Scientific, Inc.) was used to generate and analyze the spectra.

In vitro proteolysis analysis

A stock concentration of 20 mg/ml of trypsin (ICN Biomedicals, IncTM) in 20mM Tris, pH 7.5, 50mM NaCl, and 2 mM $MgCl_2$ was made for this assay which was further diluted to a final concentration of 6.667 X 10^{-3} mg/mL per reaction. The reactions were carried out at 20°C with 6.667 X 10^{-3} mg/mL trypsin and ~15 uM of protein. The reactions were stopped by adding 5X SDS loading buffer followed by boiling for 3 minutes. The reactions were loaded on a 20% SDS polyacrylamide gel and stained with coomassie for quantification with ImageJ. The 50% degradation time-point was estimated from the disappearance of the HMG domain of LEF-1/or variants (86 amino acids) using densitometry (Figure 3.1).

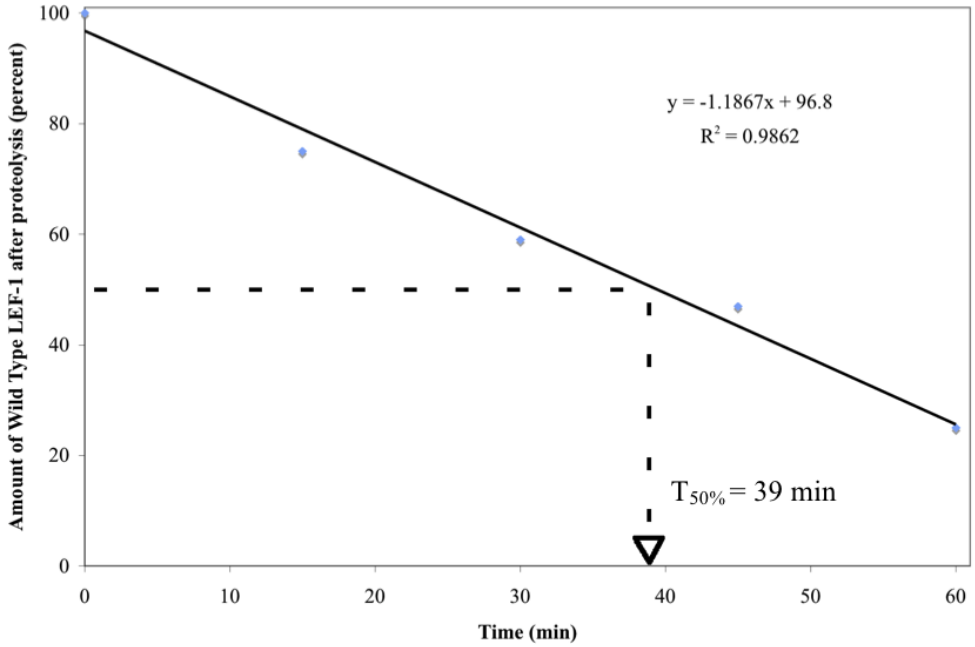


Figure 3.1. Protein stability analyzed by limited proteolysis. An example of how the amount of time was determined for 50% of the protein (wild type LEF-1, HMG domain) to be proteolyzed by trypsin. The values of the optical density in the unproteolyzed and proteolyzed bands were used to plot the percentage of protein remaining after proteolysis as a function of time. This was performed for all variants and all three trials.

3.3 Results

Circular dichroism

CD spectra were collected for LEF-1 HMG box (WT) and the nine variants in the far UV region. In the far UV region minima were observed at 208 and 222 nm indicating a helical structure present in both WT and the nine variants (Figure 3.2A). The thermal denaturation of wild type and the nine variants were monitored at 222 nm over a temperature range from 4˚C to 95˚C (Figure 3.2B). The reported T_m for WT and the nine variants are reported in Table 3.1. Six variants displayed melting temperatures higher than wild type (*i.e.* variant 1 T_m = 41ºC, variant 2 T_m = 44ºC, variant 3 T_m = 43ºC, variant 4 T_m = 46ºC, variant 5 T_m = 46ºC, and variant 8 T_m = 53ºC). In addition, three variants are reported to have melting temperatures lower than wild type (*i.e.* variant 6 T_m = 25ºC, variant 7 T_m = 35ºC, and variant 9 T_m = 34ºC). The slopes of the melting curve are fairly similar and indicate cooperative unfolding and thus provide further evidence that the variants likely maintain the general wild type fold.

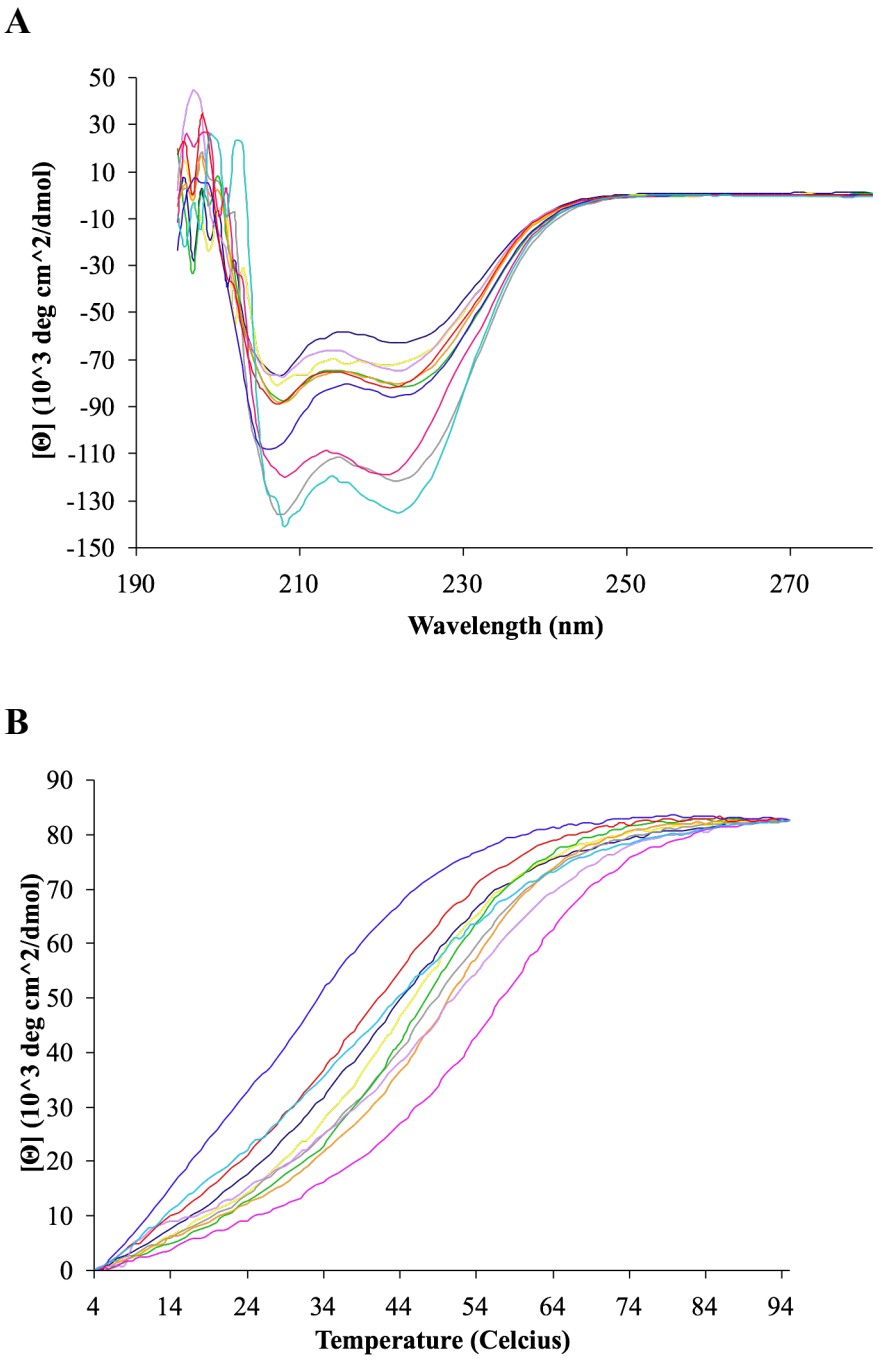


Figure 3.2. Far UV circular dichroism scans and thermal unfolding curves for wild type and the nine variants. (a) The CD spectra of the nine variants are nearly identical to LEF-1 HMG box. ■Wild type, ■Variant 1, ■Variant 2, ■Variant 3, ■Variant 4, ■Variant 5, ■Variant 6, ■Variant 7, ■Variant 8, ■Variant 9 (data not normalized). (b) Thermal denaturation of all variants monitored by CD at 222 nm (normalized as described in Materials and Methods).

Table 3.1. **Amino acid sequence of LEF-1 and the nine variants with their corresponding melting temperature.** Melting temperatures were calculated from the thermal denaturation curves generated by circular dichroism.

	Residue Position									
	12	13	17	26	27	44	49	52	55	
Variant										Tm (°C)
WT	L	Y	M	T	L	L	Q	Y	L	38
1				P						41
2							K			44
3				P	G					43
4				P	G		K			46
5		W				M		W		46
6	R								D	25
7	R	W		P	G	M	K	W	D	35
8		W		P	G	M	K	W		53
9		F	Y					F		34

[^{1}H, ^{15}N] HSQC Spectra

Two-dimensional [^{1}H,^{15}N] HSQC spectra were collected for ten variants (Figures 3.3-3.7). The spectrum of wild type is relatively poorly dispersed, contains ~63 out of a potential 83 backbone amide peaks, and exhibits heterogeneous line width and intensity, all characteristic of a flexible and dynamic molecule. The spectra of variants 1, 2, 5, 6, 7, and 9 exhibited no improvement relative to wild type. While the spectra of variants 3, 4, and 8 contained more peaks, a higher degree of dispersion and more homogeneous peak shape and intensity indicative of a more compact fold that tumbles more isotropically (*i.e.,* exhibiting more homogeneous T2 relaxation).

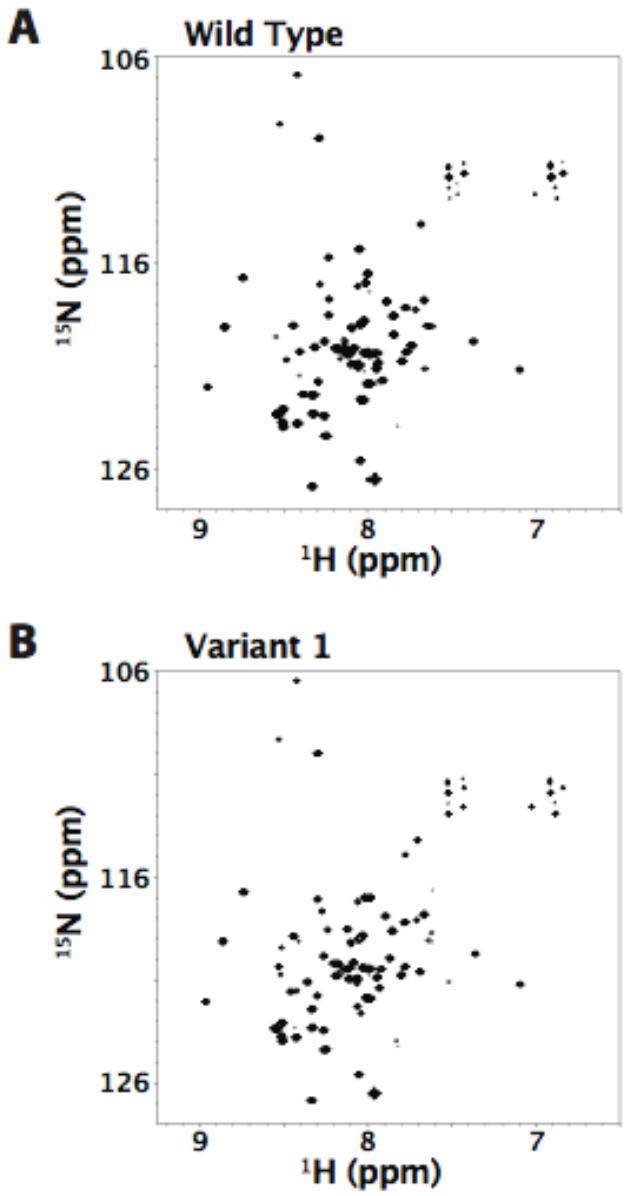


Figure 3.3 Spectra of Wild Type and Variant 1. [^{1}H, ^{15}N] HSQC Spectra of ^{15}N-labeled wild type (LEF-1 HMG box), **A** and Variant 1, **B**.

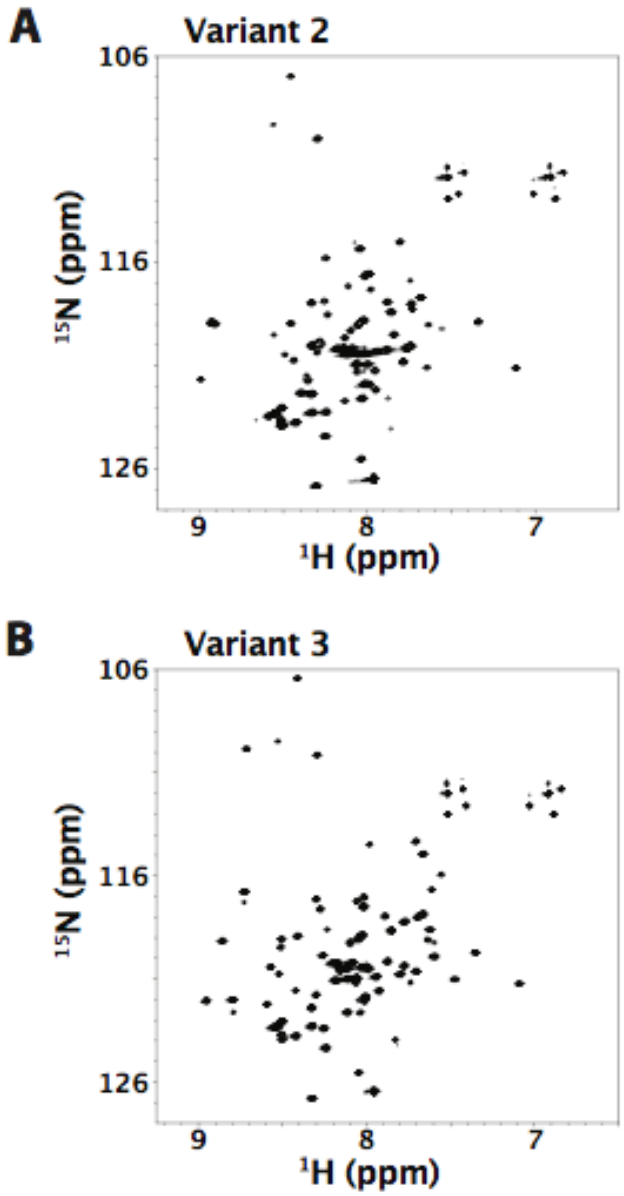


Figure 3.4 Spectra of Variant 2 and Variant 3. [^{1}H, ^{15}N] HSQC Spectra of ^{15}N-labeled variant 2, **A** and Variant 3, **B**.

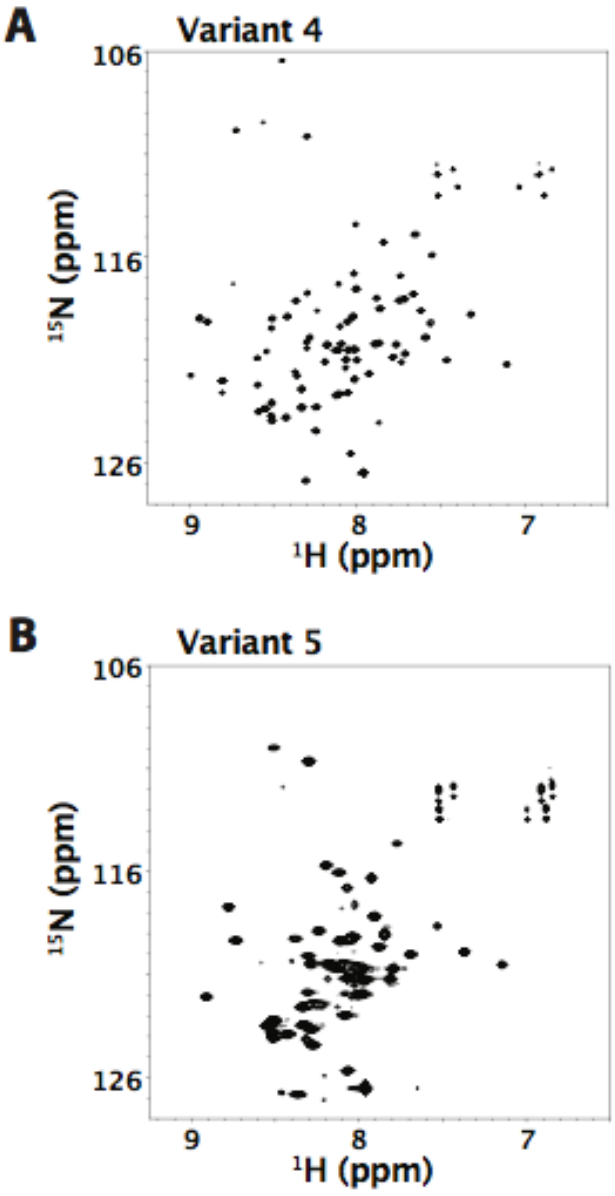


Figure 3.5 Spectra of Variant 4 and Variant 5. [^{1}H, ^{15}N] HSQC Spectra of ^{15}N-labeled variant 4, **A** and Variant 5, **B**.

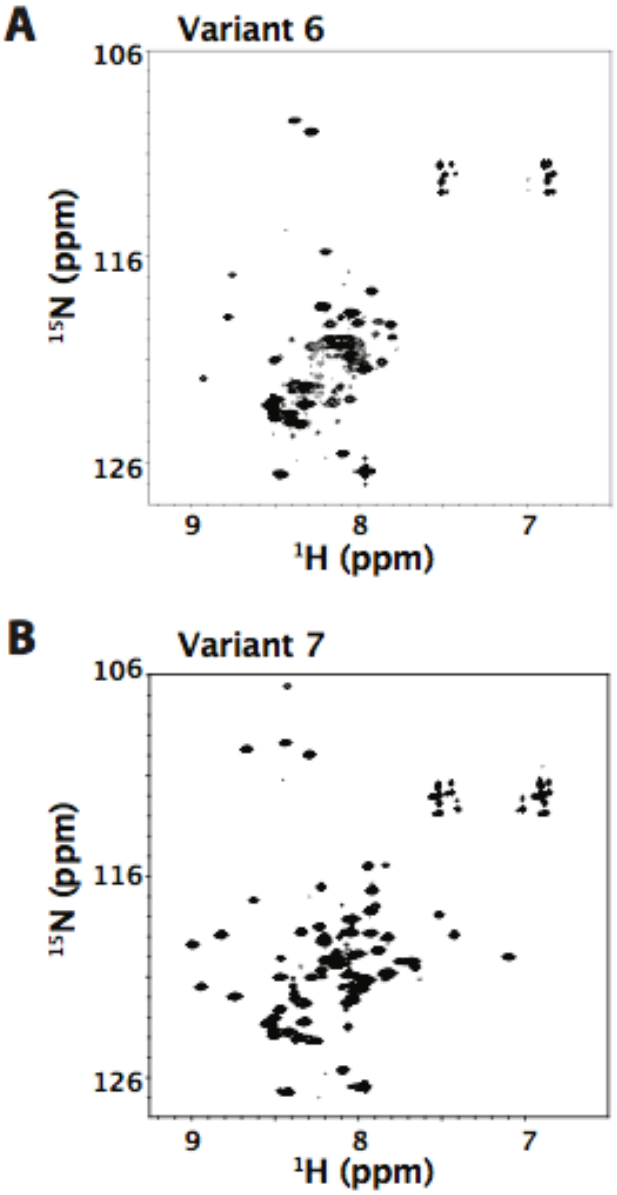


Figure 3.6 Spectra of variant 6 and Variant 7. [^{1}H, ^{15}N] HSQC Spectra of ^{15}N-labeled variant 6, **A** and Variant 7, **B**.

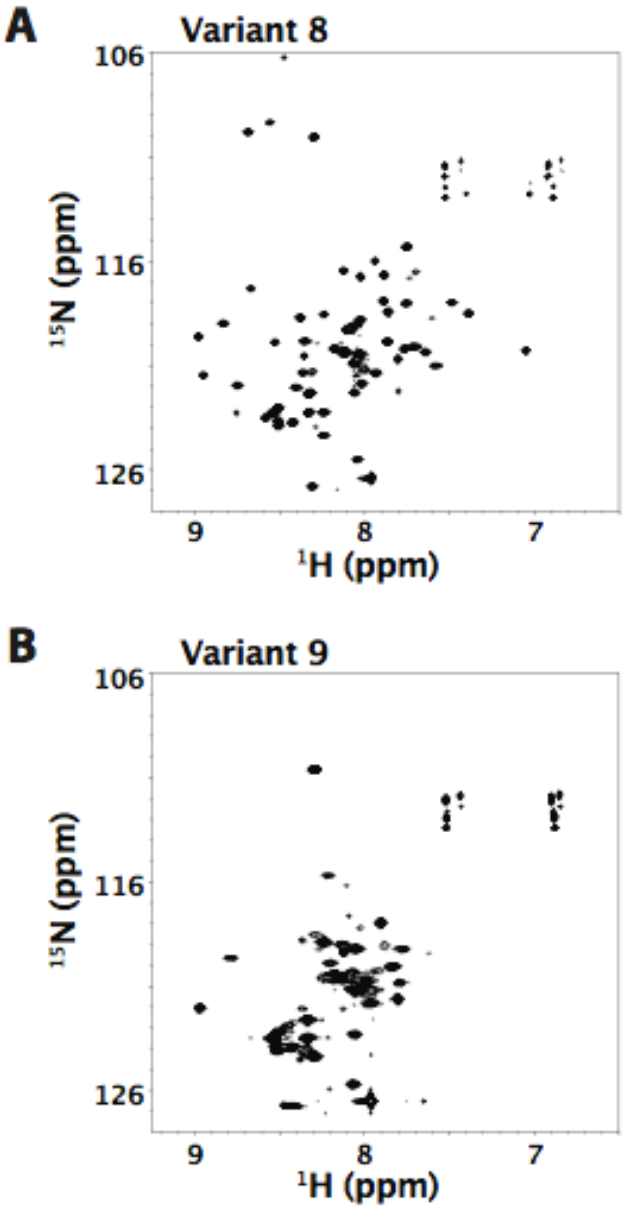


Figure 3.7 Spectra of Variant 8 and Variant 9. [^{1}H, ^{15}N] HSQC Spectra of ^{15}N-labeled variant 8, **A** and Variant 9, **B**.

In vitro proteolysis

The cleavage of peptide bonds by proteases takes place when the proper stereospecific structure is created to fit into the active site of the enzyme. Proteins of lower overall stability are degraded faster than proteins of higher stability due to a more highly malleable nature. To further characterize the overall stability of wild type and the nine variants we subjected all variants to *in vitro* proteolysis analysis using the serine protease, trypsin (Figure 3.8). A summary of the results of the *in vitro* proteolysis analysis are shown in Figure 3.9 and Table 3.2, where it shows the amount of time to degrade 50% of the protein. In general, proteins with a higher melting temperature than wild type displayed a degradation time similar to wild type (*i.e.* wild type T_m = 38ºC, $T_{50\%}$ = 41.0 ± 1.5 min; variant 1 T_m = 41ºC, $T_{50\%}$ = 40.7 ± 1.7 min; variant 2 T_m = 44ºC, $T_{50\%}$ = 44.2 ± 3.4 min; variant 3 T_m = 43ºC, $T_{50\%}$ = 40.9 ± 2.2 min; variant 4 T_m = 46ºC, $T_{50\%}$ = 42.8 ± 0.9 min, variant 5, T_m = 46ºC, $T_{50\%}$ = 43.3 ± 2.6 min; variant 8, T_m = 53ºC, $T_{50\%}$ = 48.4 ± 5.1 min). Proteins with a lower melting temperatures than wild type exhibit a degradation time less than 40 minutes (*i.e.* variant 6 T_m = 25ºC, $T_{50\%}$ = 26.2 ± 3.2 min; variant 7 T_m = 35ºC, $T_{50\%}$ = 38.9 ± 1.3 min).

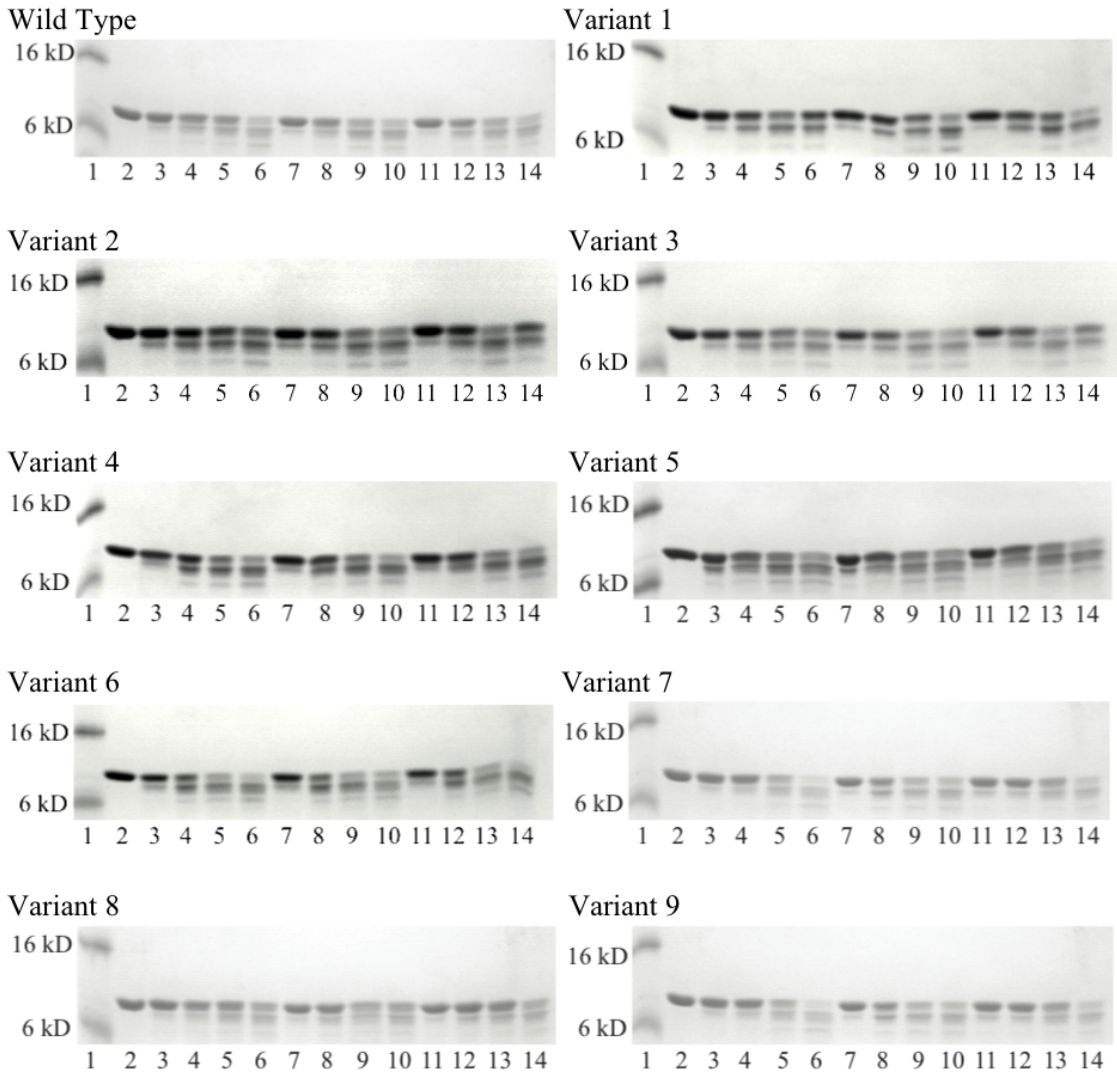


Figure 3.8 Limited proteolysis of LEF-1 and the variants. SDS-PAGE analysis of the products of trypsin (4 μg/μL) digestion for the following time points 0 minutes (lane 2), 15 minutes (lanes 3, 7, and 11), 30 minutes (lanes 4, 8, and 12), 45 minutes (lanes 5, 9, and 13), and 60 minutes (lanes 6, 10, and 14).

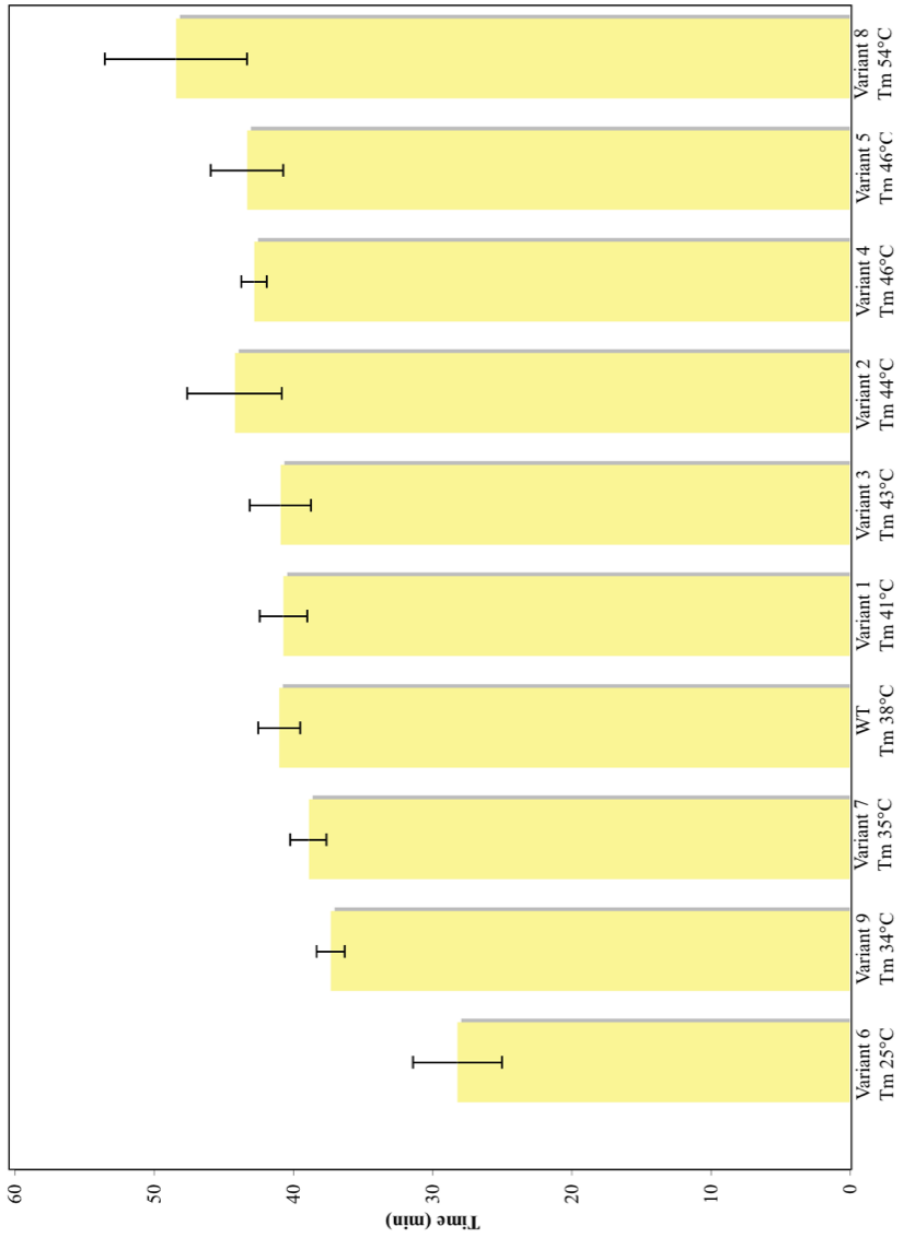


Figure 3.8 Correlation between limited proteolysis and protein thermal stability. The amount of time required to degrade 50% of LEF-1 and the variants.

Table 3.2 Amino acid sequence of LEF-1 and the nine variants with their corresponding melting temperature and time required for 50% degradation by limited proteolysis. The 50% degradation time-point was estimated from the disappearance of the HMG domain of LEF-1/or variants (86 amino acids) using densitometry. Results are mean ± S.D. of multiple experiments (n=3).

	Residue Position										
	12	13	17	26	27	44	49	52	55		
Variant										Tm (°C)	Time 50% (min)
WT	L	Y	M	T	L	L	Q	Y	L	38	41.0 ± 1.5
1				P						41	40.7 ± 1.7
2							K			44	44.2 ± 3.4
3				P	G					43	40.9 ± 2.2
4				P	G		K			46	42.8 ± 0.9
5		W				M		W		46	43.3 ± 2.6
6	R								D	25	28.2 ± 3.2
7	R	W		P	G	M	K	W	D	35	38.9 ± 1.3
8		W		P	G	M	K	W		53	48.4 ± 5.1
9		F	Y					F		34	37.3 ± 1.0

3.4 Discussion

Local protein folding upon DNA binding has been recognized since 1994 (Spolar and Record, 1994). Proteins that fold upon binding to their targets have been found to participate in cell cycle control, transcription, and translational regulation (Wright and Dyson, 1999). The sequence specific members of the HMGB family is a part of the growing list of proteins that fold upon binding to DNA. All of the current high resolution structures of the HMG domain of the HMGB family consists of three α helices arranged in a characteristic L-shaped fold (Figure 1.3) (Churchill, *et al*., 2010) (Liu, *et al*., 2009) (Kasai, *et al*., 2005) (Masse, *et al*., 2002) (Xu, *et al*., 2002) (Cary, *et al*., 2001) (Murphy, *et al*., 1999) (Hardman, *et al*., 1995) (Love, *et al*., 1995) (Werner, *et al*., 1995) (Read, *et al*., 1993). These high resolution structures were determined either in the presence (*i.e.* usually HMG domains of sequence specific family members) or absence of DNA (*i.e.* usually HMG domains of non-sequence specific family members). The long arm of the L comprises the N-terminal strand and helix III while helices I and II make up the short arm of the L-shaped structure. The L-shaped structure is partially held together by the major hydrophobic core and one minor core (Travers, 2000). Mutagenesis and structural analysis of the L-shaped fold has helped us identify three areas of the HMG domain that contribute to the overall stability of the domain.

Turn 1 region between helix 1 and 2

Sequence comparison of the HMG domain of LEF-1 reveals that LEF-1 is one of the few members of the HMGB family that does not have a proline residue at position 26 (Figure 2.2). In other members of the family there is often a His-Pro-Gly or His-Pro-Asp sequence in this region. The introduction of a proline at position 26 modestly increased the thermal stability of variant 1 (wild type T_m = 38ºC; variant 1, T_m = 41ºC) however showed no improvement in the HSQC spectra (Figure 3.3), and a half life similar to wild type (Table 3.2, Figure 3.9). On the other hand, the introduction of a proline and glycine at positions 26 and 27 (variant 3) increased the thermostability of LEF-1 and an improved HSQC spectra with more peaks, a higher degree of dispersion and more homogeneous peak shape and intensity (wild Type T_m = 38°C; variant 3 T_m = 43°C; Table 3.1, Figure 3.4). However, the half life of variant 3 is similar to wild type (wild type $T_{50\%}$ = 41.0 ± 1.5 min, variant 3 $T_{50\%}$ = 40.9 ± 2.2 min). The discrepancy between the protease assay results, CD and NMR may be due to the limitations of the proteolysis assay. Hubbard reported the conformational changes needed for limited proteolysis require proteins to unfold up to 12 amino acids long in a chain segment (Hubbard, 1998). Thus, our proteolysis studies of wild type, variant 1 and 3 showed similar half lives because there are likely no large conformational changes in the polypeptide to observe a difference in proteolysis.

The proline residue at position 26 may function to break helix 1 and stabilize the subsequent turn and bring amino acids from helix 1 and 2 into close proximity of one another to form the major hydrophobic core. The absence of a helix breaking

'active turn' in LEF-1 results in the subsequent loss of rigidity which contributes to the instability observed for this region (Figure 3.3). Liu *et al*., Phillips *et al*., and Weiss *et al*. report the formation of the helix 1-turn-helix 2 motif prior to the packing of aromatic side chains in the major hydrophobic core (discussed below) of the HMG domain of hUBF Box-5 (Liu, *et al*., 2009), SRY (Phillips, *et al.,* 2011), and Sox-5(Weiss, 2001). The formation of the helix-turn-helix motif is also reported in other helix bundle proteins (*e.g.* Engrailed homeodomain (EnHD) (Religa, *et al*., 2007) and KIX domain (Schanda, *et al*., 2008)). Three helix proteins fold via the formation of the helix 2-turn-helix 3 motif or helix 2-helix 3 hairpin (Liu, *et al*., 2009). Monte Carlo folding simulations of EnHD, B domain of protein A, and the Villin headpiece (three helix bundle proteins) show all three proteins undergo the formation of the helix-turn-helix motif prior to the third helix formation (Yang, *et al*., 2008).

Major hydrophobic core

The major hydrophobic core is believed to be the first region of the HMG domain to form due to hydrophobic collapse. Native state hydrogen deuterium experiments have shown the hydrophobic core of the HMG domain of NHP6A and hUBF to be the first portion of the protein to adopt its secondary and tertiary structure (Masse, *et al*., 2002) (Liu, *et al*., 2009). In the free HMG domain of LEF-1, the major hydrophobic core is partially formed between helices two and three whereas in other HMGB proteins such as UBF, SRY, Sox-4, and Sox-5 the major hydrophobic core is partially formed between helices one and two (Liu, *et al*., 2009) (Phillips, *et al*., 2011).

The four residue positions in the central hydrophobic core that contribute significantly to the overall stability of the HMG domain are Tyr-13, Trp-41, Leu-44, and Tyr-52. Mutations made at those positions (variant 5, Y13W, L44M, Y52W; position 41 in the LEF-1 sequence is already a tryptophan) increased the thermostability of the HMG domain of LEF-1 (wild type T_m = 38°C, variant 5 T_m = 46°C, Table 3.1 and Figure 3.5) and displayed a [^{1}H,^{15}N] HSQC spectra similar to wild type. In addition, the half life is similar to wild type (wild type $T_{50\%}$ = 41.0 ± 1.5 min, variant 5 $T_{50\%}$ = 43.3 ± 2.6 min). We predict the addition of the tryptophan residues in the major hydrophobic core in variant 5 may have created a nonpolar environment where hydrophobic interactions can form between nonpolar residues (Trp-13, Met-44, Trp-41, and Trp-52). These interactions may function to stabilize the tertiary structure of the protein which was demonstrated to be the case for a non-sequence specific family member, hUBF (Wang, *et al*., 2007). Wang and coworkers report the tryptophans in box 5 of hUBF to be stacked one on top of the other as shown in Figure 3.10. The hydrogen exchange kinetics reported by Wang and coworkers show slow H/D exchange of amide protons that correspond to residues Trp-13, Trp-41, and Trp-52. Residues Trp-13, Trp-41, and Trp-52 exhibiting slow H/D exchange rates are most likely taking part in hydrogen bonds and other interactions in the protein structure. These interactions include the aromatic rings of Trp-13 and Trp-52 participating in polar electrostatic interactions that involve π-electrons circulating above and below the faces of the aromatic rings that function to stabilize the hydrophobic core. Although variant 5 has a melting temperature that is 8 degrees

higher than wild-type the [^{1}H,^{15}N] HSQC spectra reveals no structural improvement relative to wild type. This is likely due to the fact that the three mutations that give rise to variant 5 (*i.e.,* Y13W, L44M, Y52W) result in larger side-chains which generally affords greater hydrophobic contacts but no apparent increase in structurally specificity. The HSQC spectra indicate that the structure for variant 5 lacks specificity which may be due to inefficient packing plus the conformational fluctuations of helix one and two due to the absence of the turn stabilizing Pro 26 and Gly 27 mutations. In addition, no difference is reported in the half life of variant 5 due to the lack of conformational change in the polypeptide.

As discussed in the previous section, the presence of a proline residue at position 26 and a glycine at position 27, enables the amino acids in helices one and two to pack more efficiently and thus possibly contribute to a more favorable formation of the major hydrophobic core. This may be, in part, why variant 8 is more thermally stable than variant 5 (variant 5 Tm = 46°C; variant 8 Tm = 53°C). Higher conformational stability is exhibited in the [^{1}H,^{15}N] HSQC spectra of variant 8 in agreement with the higher T_m as measured by circular dichroism. The [^{1}H,^{15}N] HSQC spectra of variant 8 contains more peaks, a higher degree of dispersion and more homogeneous peak shape and intensity than variant 5 (Figures 3.4 and 3.6). In addition, the proteolysis results appear to agree with the CD and NMR data as variant 8 exhibited the highest degree of resistance to proteolysis (*i.e.,* has the longest half-life).

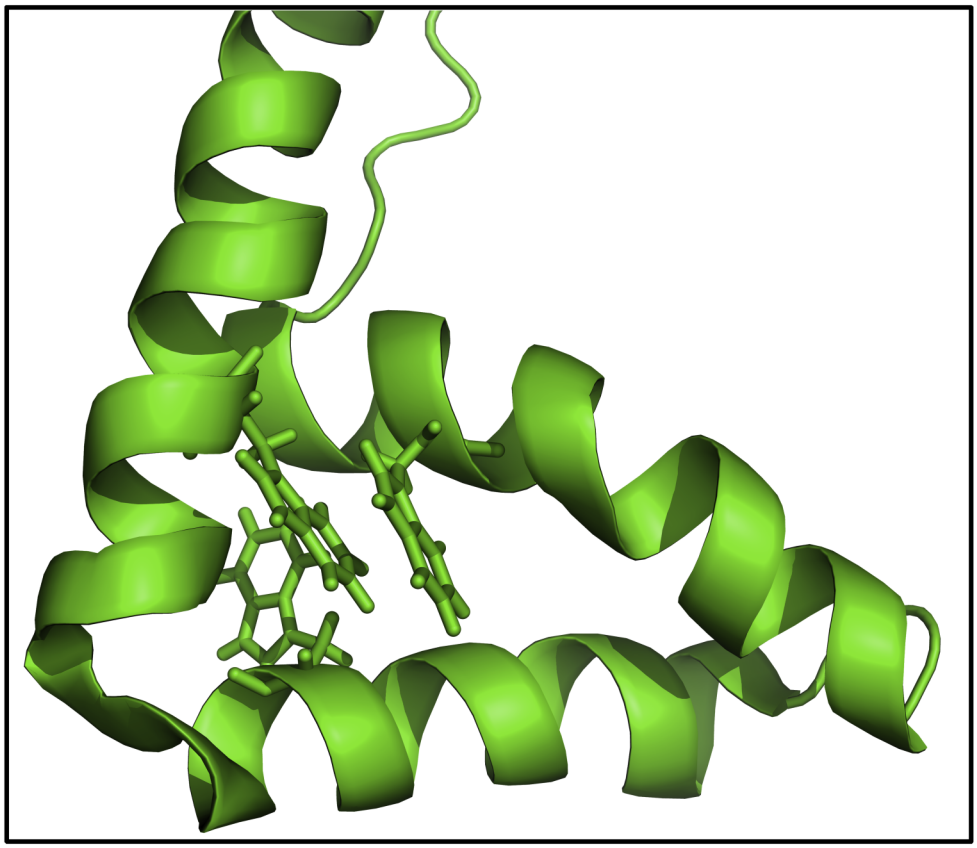


Figure 3.10 Major hydrophobic core region of hUBF HMG Box-5 (PDB: 1L8Y). The side-chains of three tryptophans (W14, W41, and W52) stack one on top of the other to stabilize the HMG domain of UBF.

Turn 2 region between helix 2 and 3

Sequence comparison of the HMG domain of LEF-1 reveals that LEF-1 is the only member of the HMGB family that does not have a lysine residue at position 49 (Figure 2.1). Our studies have shown the introduction of a lysine at position 49 results in an increase in the thermostability of the HMG domain of LEF-1 (variant 2 T_m = 44°C, variant 4 T_m = 46°C, variant 8, T_m = 53°C, Table 3.1) and a significant improvement in the [^{1}H,^{15}N] HSQC spectra (variant 4, Figure 3.4) suggesting lysine is structurally important for the formation of the hydrophobic core and the turn 2 region of the HMG domain. The single mutation, Q49K showed a slight improvement in [^{1}H,^{15}N] HSQC spectra compared to wild type with more peaks and a slight improvement in peak dispersion. When the single mutation is combined with the T26P and L27G mutations a distinct change in the [^{1}H,^{15}N] HSQC spectra was observed. The [^{1}H,^{15}N] HSQC spectra of variant 4 contains more peaks, a higher degree of dispersion and more homogeneous peak shape and intensity, all characteristics of a more well folded protein. As discussed earlier, the proline residue at position 26 may function to break helix 1 and stabilize the subsequent turn and bring amino acids from helix 1 and 2 into close proximity of one another while also possibly enabling better packing in the major hydrophobic core. The addition of the following mutation, Q49K to T26P and L27G possibly contributed to the formation of the hydrophobic core by stabilizing turn 2 (turn region between helix 2 and 3). Several studies have been published that indicate Lys-41 is important for stabilizing turn 2 of the HMG domain. The first study published by Read and coworkers showed the methyl group of Thr-44

interacting with the γ and ε methylenes of Lys-49 of HMG1which functions to anchor turn 2 onto the hydrophobic core (Read, *et al*., 1993). The side chains of Lys-49 have also been reported to pack along the face of the indole ring of Trp-41 with its ε-amino group positioned to hydrogen bond with the carbonyl group of Trp-41. As a result, a positive charge is placed at a suitable location to interact with the dipole of helix 2 (Read, *et al*., 1993). In a second study, Jones and colleagues reported the protons of Lys-49 of HMGD to be shifted upfield in the NMR spectrum. Based on these NMR findings, they postulated Lys-49 to be interacting with aromatic residues in the hydrophobic core (Jones, *et al*., 1994). In a third study, Werner and coworkers created a SRY variant with the following mutation K49I which resulted in a disruption of the three dimensional fold by introducing two bulky methyl groups (Werner, *et al*., 1995). They suggested the methyl groups of isoleucine may have displaced Phe-55 from the hydrophobic core and partially displacing Leu-46 thereby destabilizing the packing between helix 2 and helix 3 (Werner, *et al*., 1995). In another study Weir and colleagues showed an interaction between Lys-53 and Trp-45 in HMGB1 (Weir, *et al*., 1993). Lys-53 and Trp-45 join Phe-14, Phe-17, and Trp-56 to form the hydrophobic cluster between the two arms. In the last study, Kasai and colleagues reported one lysine residue stabilizing turn 1(between helix 1 and 2) and another lysine residue stabilizing turn 2(between helix 2 and 3) of SSRP1 (structure specific recognition protein 1) (Kasai, *et al.,* 2005).

Destabilized variants

Three variants are reported to have lower melting temperatures and [^{1}H,^{15}N] HSQC spectra that exhibits worse properties than wild type (*i.e.* variant 6 T_m = 25ºC, variant 7 T_m = 35ºC, and variant 9 T_m = 34ºC, Table 3.1, Figures 3.5 and 3.6). The mutations introduced in variant 6 were meant to create a salt bridge between helix 1 and 3 using residue positions 12 and 55. The introduction of an arginine and glutamate at position 12 and 55 respectively ultimately resulted in the lowering of the melting temperature for variant 6 by 13 degrees and a [^{1}H,^{15}N] HSQC spectrum that has fewer peaks, heterogeneous peak shape and intensity relative to wild type. In addition, variant 6 was proteolyzed more rapidly than wild type (Table 3.2 and Figure 3.7) (wild type $T_{50\%}$ = 41.0 ± 1.5 min, variant 6 $T_{50\%}$ = 28.2 ± 3.2 min), suggesting variant 6 is less stable than wild type. The decrease in stability may be due to the hydrophilic nature of Arg 12 and Glu 55 which may result in a more favorable interaction with ions in the solvent instead of the desired salt bridge between helix 1 and 3.

Variant 7 is another variant which displayed a small decrease in stability compared to wild type. We incorporated all the mutations in our study in variant 7 (excluding the mutations made in variant 9). This resulted in lowering the melting temperature from 38ºC to 35ºC, an [^{1}H,^{15}N] HSQC spectrum similar to wild type, and a half life similar to wild type (wild type $T_{50\%}$ = 41.0 ± 1.5 min, variant 7 $T_{50\%}$ = 38.9 ± 1.3 min). These results suggest that the mutations modestly destabilized the HMG domain of LEF-1. We hypothesize the destabilizing effects associated with the Arg 12 and Glu 55 mutations is the likely cause for the low stability observed for this variants.

It appears to be fairly clear that substitution of leucine residues at positions 12 and 55 for the introduced 'salt-bridge' residues (*i.e.,* Arg and Glu) eliminated favorable (and non-directional) van der Waals contacts between the leucine residues themselves as well as with other hydrophobic portions of the major hydrophobic core. Any favorable electrostatic energy afforded by the Arg 12 and Glu 55 mutations obviously could not compensate for the loss of the hydrophobic interactions.

The last variant created with the aid of a protein design algorithm, ORBIT resulted in a variant with a small decrease in the melting temperature (from 38ºC to 35ºC), an lower quality [^{1}H,^{15}N] HSQC spectrum that exhibited fewer peaks, a lower degree of dispersion, heterogenous peak shape and intensity, and a small decrease in half life ($T_{50\%}$ = 37.3 ± 1.0). These results indicate the mutations modestly destabilized the HMG domain of LEF-1. We predict the decrease in stability is due to the loss in hydrogen bonding between Tyr 13 and Met 17 when these residues were mutated to Phe 13 and Tyr 17. In addition, we predict there is a loss in hydrogen bonding between Tyr 52 and Glu 16 when Tyr 52 was mutated to Phe 52. The combined effects of these mutations, although calculated to be stabilizing, clearly resulted in a more destabilized variant. This finding is a fairly strong indicator that the prediction of mutations that are more stabilizing is a complicated matter and a greater understanding (and more accurate mathematical emulation) of the forces that stabilize tertiary structure is needed.

Summary. Mutagenesis and structural analysis of the HMG domain of LEF-1 has helped us identify three areas of the domain that contribute to the overall stability. The presence of a proline and glycine in the first turn region (between helix 1 and 2) increased the thermal stability of the HMG domain of LEF-1. The proline residue at position 26 may function to break helix 1 and stabilize the subsequent turn and bring amino acids from helix 1 and 2 into close proximity of one another to form the hydrophobic core. The introduction of a lysine in the second turn region (between helix 2 and 3) also increased the thermal stability of the HMG domain of LEF-1 suggesting Lys 49 to be structurally important for the formation of the hydrophobic core and the turn 2 region of the HMG domain. Lastly, the replacement of polar residues with nonpolar residues in the hydrophobic core increased the stability of the HMG domain of LEF-1.

Chapter 4

Functional Studies

4.1 Introduction

LEF-1 is a transcription factor that functions as a nuclear mediator of the canonical Wnt/β-catenin signaling pathway (Yee *et al*., 2005). The HMG domain of LEF-1 has been shown to bind to the following sequence CCTTTGAA within the T-cell receptor (TCR)-α gene enhancer region (Neuweiler *et al*., 2009) (Lee and Vasmatzis, 1997). The HMG domain of LEF-1 not only recognizes a specific sequence but it also induces a bend in the DNA ~100° (Love *et al*., 2004) to coordinate the binding of other transcription factors. Due to the ability of LEF-1 to bend the DNA and inability to activate transcription in reporter assays, LEF-1 is categorized as an architectural transcription factor (Zhang and Wang, 2008).

In this chapter we report how the electrophoretic mobility shift assay (EMSA) and the circular permutation assay were used to characterize the ability of the LEF-1 mutant variants to bind and bend cognate DNA. The electrophoretic mobility shift assay was used to assess the binding affinity of the DNA while the circular permutation assay was used to compare the DNA bending ability of LEF-1 and variants. The combined use of these two assays allowed us to gain insight into any effects stability had on the HMG domain of LEF-1 to bind and bend the DNA.

This introductory section is followed by a description of the materials and methods used to study the DNA binding and bending by LEF-1 and the variants (section 4.2). The results of EMSA (section 4.3) and the circular permutation assay

(section 4.4) are described, and finally a discussion and conclusion section is provided that also describes possible future experiments that might be accomplished to gain further insight in the effect of stability on its ability to bind and bend the DNA.

4.2 Materials and Methods

Electrophoretic Mobility Shift Assay

The probe used for the assay is a 5' end biotin label 34 bp duplex (5'-TCGACGTAGGGCACCCTTTGAAGCTCTCCCTCA-3'). The probe was added to the appropriate protein at 25°C in the following buffer (20mM Tris-HCl pH 7.5, 50mM NaCl, 5mM $MgCl_2$, 10% glycerol, 50 ug/mL BSA). The binding reaction was electrophoresed in a 6% non-denaturing polyacrylamide gel and transferred onto a nylon membrane. Free and complexed DNA were detected by chemiluminescence (Figure 4.1) and the quantification of the bands were done using the ImageJ software package. The values obtained from ImageJ for bound and free DNA from each gel were used to plot the fraction of bound DNA as a function of the concentration of LEF-1. A best fit curve was obtained for the plot using IgorPro 4.0 software. The equation generated from the best fit curve was used to estimate the concentration required for half-maximal binding (taken as an estimate for the K_d). The raw results from the EMSA analysis are provided in Figure 4.1.

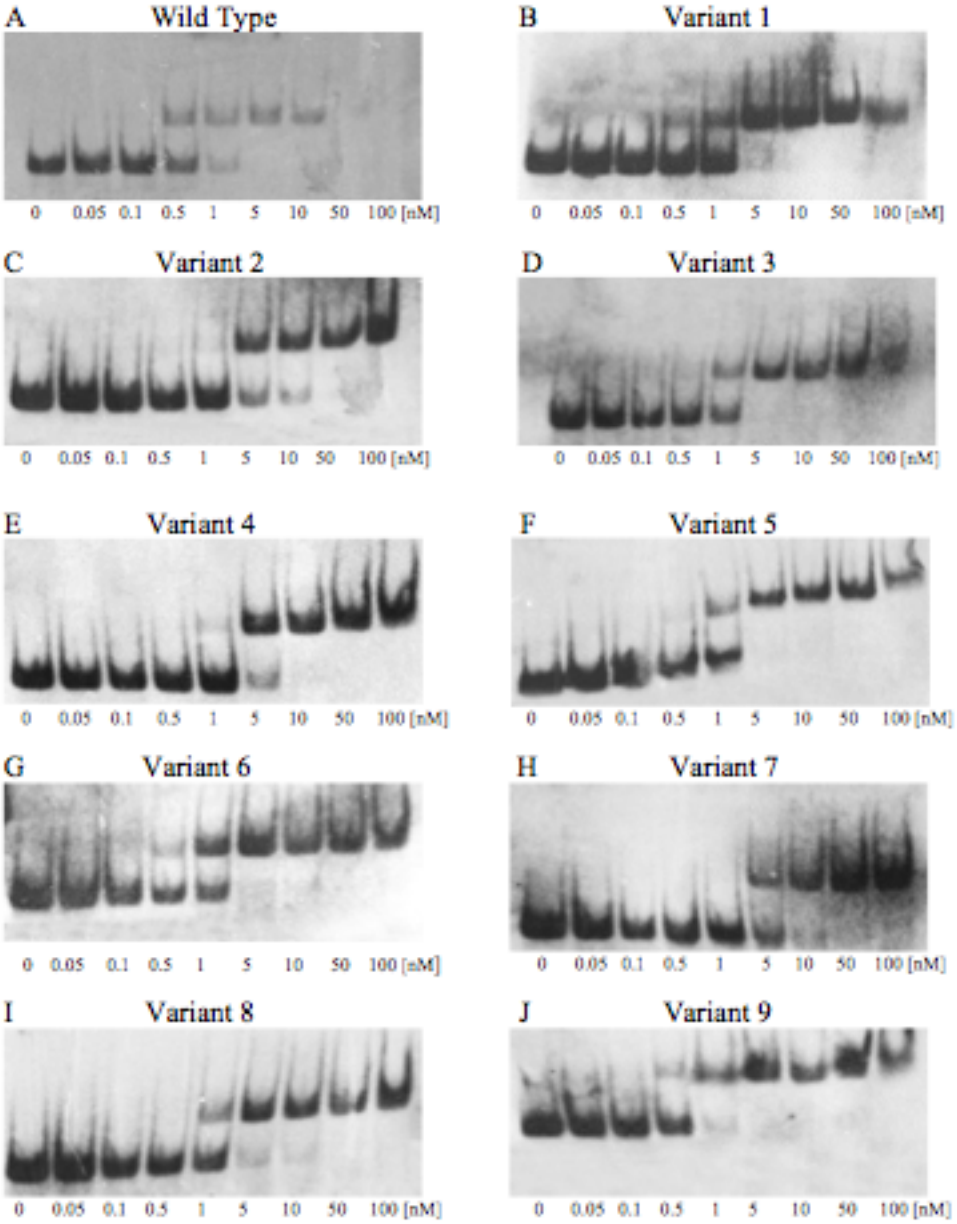


Figure 4.1. Analysis of the equilibrium binding of LEF-1 and the variants to a oligonucleotide containing the LEF-1 consensus recognition sequence in a electrophoretic mobility shift assay. The values from the band intensity of bound and free DNA from a-b were used to plot the fraction of bound DNA as a function of the log of protein. (A) Wild type; (B) Variant 1; (C) Variant 2; (D) Variant 3; (E) Variant 4; (F) Variant 5; (G) Variant 6; (H) Variant 7; (I) Variant 8; (J) Variant 9. Equilibrium binding constant was calculated using Igor Pro.

Circular Permutation Assay

The 440 bp probe containing the LEF-1 binding site was obtained from a circular permutated plasmid using the following restriction enzymes EcoRI, HindIII, BstNI, RsaI, EcoRV, EaeI, and BamHI (Figure 4.2). The 440bp probes were further amplified by PCR with primers listed in Figure 4.3. The contents of the binding reaction buffer are 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM $MgCl_2$, 10% glycerol, and 50 ug/mL BSA. The 440bp probes were then biotinylated with terminal deoxynucleotidyl transferase (Fermentas). Protein and DNA were electrophoresed on a 6% non-denaturing polyacrylamide gel and visualized by chemiluminesence.

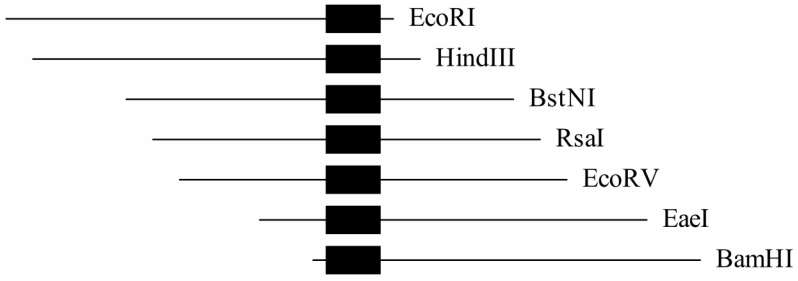


Figure 4.2. Structure of the 440 bp DNA probes used in the circulation permutation analysis. The probes were generated by PCR and contains the DNA binding sequence for wild type (indicated by the black box). Each DNA is designated by the origin of these fragments from restriction digests of the original circular plasmid.

Primer Name	Primer Sequence
EcoRI Forward	5′ CTCATGTTTGACAGCTTATCAT 3′
EcoRI Reverse	5′ CCCAGATCTAGACCCCCTCGAG 3′
HindIII Forward	5′ AGCTTTAATGCGGTAGTTTATCAC 3′
HindIII Reverse	5′ TATCGATGATAAGCTGTCAAACAT 3′
BstNI Forward	5′ TGGATGCTGTAGGCATAGGC 3′
BstNI Reverse	5′ GGGTGACGGTGCCGAGGATG 3′
RsaI Forward	5′ ACTGCCGGGCCTCTTGCGGGAT 3′
RsaI Reverse	5′ ACCGGCATAACCAAGCCTATG 3′
EcoRV Forward	5′ ATCGTCCATTCCGACAGCATCGCC 3′
EcoRV Reverse	5′ ATCCCGCAAGAGGCCCGGCAGTAC 3′
EaeI Forward	5′ GCCGCCCAGTCCTGCTCGCTTC 3′
EaeI Reverse	5′ AAAGCGGTCGGACAGTGCTCCGAG 3′
BamHI Forward	5′ CCCGAGCTCCCCCTCGACGTAG 3′
BamHI Reverse	5′ CACAGGACGGGTGTGGTCGCCATG 3′

Figure 4.3. List of primers used in generating the 440bp DNA probes for circular permutation analysis.

4.3 Results and Discussion

In Chapter Three we described the design and creation of LEF-1 variants that were more stable than wild type and some variants that were less stable than wild type. Wild type LEF-1 has a melting temperature of 38°C and an equilibrium binding constant of 1.03 ± 0.05 nM estimated by EMSA (Table 4.1, Figure 4.1). Two variants, variant 7 (T_m = 35ºC) and variant 8 (T_m = 53°C) exhibit small yet consistent decrease in binding affinity (variant 7, K_d = 8.61 ± 0.93 nM; variant 8, K_d = 2.36 ± 0.51 nM) (Table 4.1, Figure 4.1). These variants (as with all variants) exhibit the same relative mobilities in the polyacrylamide gel as those observed with wild type in the circular permutation assay (Figure 4.4). These results suggest the mutations introduced into variants 7 and 8 likely interfere with some aspect of binding interactions and whatever these effects are they do not have any apparent effects on bending DNA. For these variants the addition and increased packing of the tryptophans (Trp 13 and Trp 52) in the hydrophobic core may have increased the overall volume of the major hydrophobic core and possibly changed the surface complementarity between the protein and DNA. This may have lead to weaker protein-DNA interactions and the associated increase for the binding constants for these two variants. No effects in the DNA bend were reported and this may be due to the fact that no mutations were made in the primary side chains (Met 10 and Tyr 75) that participate in DNA bending. Met 10 has been shown in the LEF-1/DNA interaction to be partially inserted into the DNA which partially offsets the energy penalty for disruption of favorable base stacking which contributes to DNA bending. Tyr 75 also contributes to DNA bending

through its insertion into the minor groove near the 5' end of the TCR- α enhancer region (Love *et al*., 1995).

Two variants with the same melting temperature and different [^{1}H,^{15}N] HSQC spectra (Chapter 3)(variant 4 and variant 5, T_m = 46°C) show different binding affinities for cognate DNA. Variant 4 (T26P, L27G, Q49K) (T_m = 46°C), which exhibits high conformational specificity displayed a slight increase in binding affinity for DNA (K_d = 0.6 ± 0.2 nM) (Table 4.1, Figure 4.1) compared to wild type (K_d = 1.03 ± 0.05 nM)

Table 4.1 Summary of the K_d values of WT and the nine variants.

	Residue Position										
	12	13	17	26	27	44	49	52	55		
Variant										Tm (°C)	Kd (nM)
WT	L	Y	M	T	L	L	Q	Y	L	38	1.03 ± 0.05
1				P						41	1.16 ± 0.55
2							K			44	1.11 ± 0.41
3				P	G					43	1.08 ± 0.38
4				P	G		K			46	0.60 ± 0.02
5		W				M		W		46	1.30 ± 0.30
6	R								D	25	1.20 ± 0.20
7	R	W		P	G	M	K	W	D	35	8.61 ± 0.93
8		W		P	G	M	K	W		53	2.36 ± 0.51
9		F	Y					F		34	0.74 ± 0.05

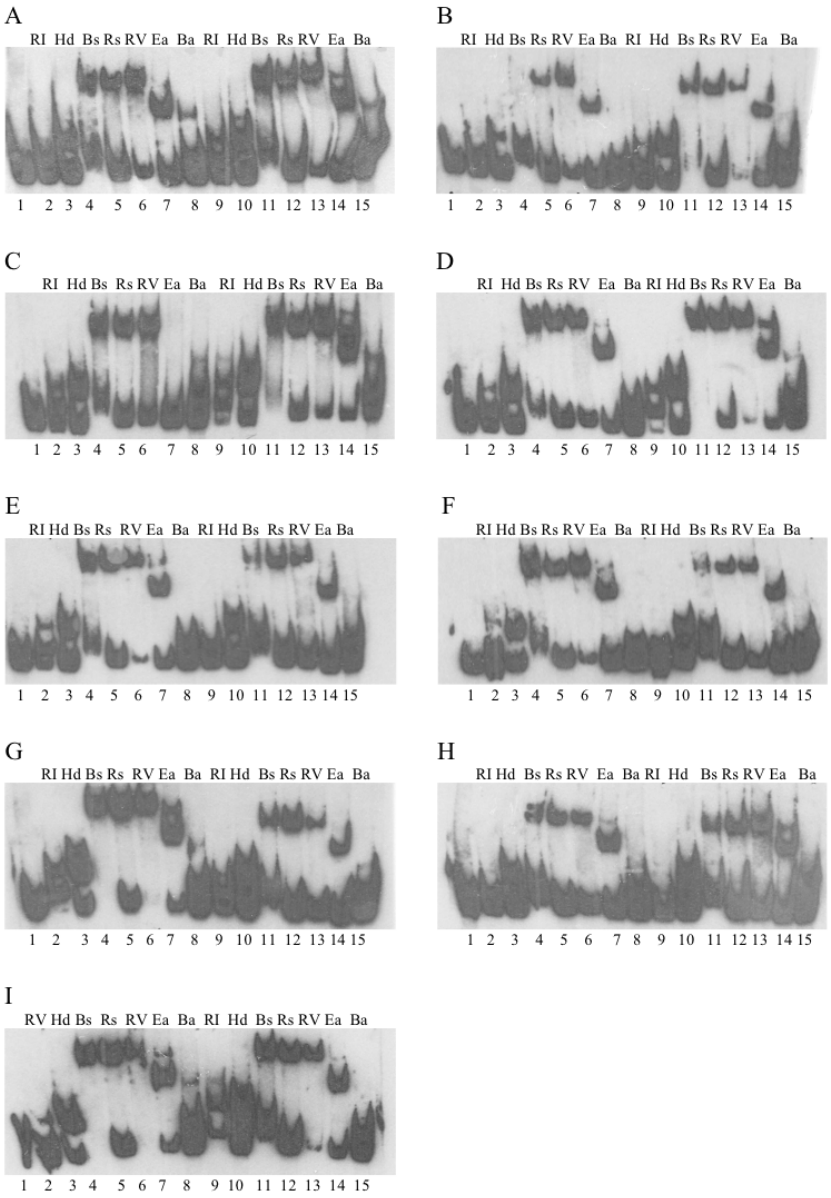


Figure 4.4. Circular permutation analysis of DNA bending induced by binding of wild type and the variants. The probes were generated by PCR and contains the DNA binding sequence for LEF-1. Each DNA is designated by the origin of these fragments from Giese et al. Electrophoretic mobility shift assay of wild type (A-I) in lanes 2-8 and the following variants in lanes 9-15: (A) Variant 1, (B) Variant 2, (C) Variant 3, (D) Variant 4, (E) Variant 5, (F) Variant 6, (G) Variant 7, (H) Variant 8, (I) Variant 9 bound to DNA. DNA probe is indicated above the lane of interest and abbreviations correspond to the labeled DNA in Figure 5.2. Lane 1 (a-i) contains the free DNA probe without protein.

(Table 4.1, Figure 4.1),whereas variant 5 showed an equilibrium binding constant (1.30 ± 0.30 nM) similar to wild type. As stated above both variants exhibited no differences in DNA bending relative to each other as well as to the wild type domain. The results suggest the mutations that gave rise to variant 4 had a modest impact on the protein-DNA interaction as its binding constant implies a slightly higher affinity. The amino acid substitutions in variant 4 were made in the turn regions of the HMG domain of LEF-1 whereas the mutations made in variant 5 (Y13W, L44M, Y52W) were in the hydrophobic core. Thus in variant 4 the packing of the hydrophobic core and the surface complementarity between the HMG domain and the DNA may be similar to or slightly better than wild type. It is possible that since tighter folding appears to occur prior to DNA binding with this variant, that there is less of an entropic penalty paid for the conformational ordering that the wild-type variant must pay upon DNA binding. Since there is presumably no gross changes to the overall structure for variant 4 (as determined by NMR analysis) then important residues, such as Met 10, are readily available for binding DNA,which may be the reason for the slight increase in binding affinity. Of particular interest, variant 5 has the same melting temperature as that of variant 4 yet exhibits measurably greater conformational variability as determined by NMR. Although the binding constants are possibly too close to definitely assign possible reasons it is tempting to speculate that the reason for the reduced binding constant is related to the lower conformational specificity exhibited by variant 5 as well as variant 7. Variant 7 not only exhibits low conformational specificity by NMR, its structure is likely also perturbed enough such

that important contact residues are likely not positioned for proper DNA binding, hence the slight reduction in binding affinity. On the other hand variant 8 exhibits the highest thermal stability, relatively high conformational specificity, yet also reduced binding affinity. This is likely due to structural changes that take place in the major hydrophobic core that likely result in the rearrangement of residues that may not be optimal for as efficient binding affinity. Of particular interest is the variant designed using the ORBIT suite of protein design algorithms. This variant has almost the lowest thermal stability (T_m = 34°C), displays high dynamics as exhibited by NMR analysis, yet has a measured binding constant that is indicative of affinity that is on par with that of the wild type variant. Interestingly, the sequence returned from the ORBIT design (*i.e.,* Y13F, M17Y, L37L, L44L, Y52F) is highly similar to the structurally stable HMGB1 domain (*i.e.,* F13, Y17, L37, T44, Y51). The results suggest phenylalanine may pack more favorably in the hydrophobic core than tryptophan when this domain is bound to DNA. It appears that the structure of the major hydrophobic core for this variant, although more mobile, is not as perturbed as the other variants and thus gives rise to equal if not slightly stronger affinity.

Lastly, four variants with thermal stabilites greater than wild type (variant 1 T_m = 41ºC, variant 2 T_m = 44ºC, variant 3 T_m = 43 ºC, variant 5 T_m = 46 ºC) and one variant less stable than wild type (variant 6 T_m = 25ºC) exhibit equilibrium binding constants similar to wild type (*i.e.* wild type, 1.03 ± 0.05 nM; variant 1, 1.16 ± 0.55 nM; variant 2, 1.11 ± 0.41 nM; variant 3, 1.08 ± 0.38 nM; variant 5, 1.30 ± 0.30 nM). In addition, all variant-DNA complexes showed the same alterations in their relative

mobilities in the polyacrylamide gel as those observed with wild type in the circular permutation assay (Figure 4.4). These results suggest that mutating residues that do not contact DNA have little to no effect on binding and bending DNA *in vitro*. Thus far, studies have only shown the mutation of amino acid residues that come into direct contact with the DNA to significantly alter the DNA binding affinity and bend angle (Giese *et al*., 1991) (Carlsson *et al*., 1993) (Read *et al*., 1994) (Churchill *et al*., 2010) (Murphy *et al*., 2001). Thus, our results compliment these studies along with the findings of the HMGB1b stability and function studies performed by Bell *et. al.* (Bell *et al*., 2002). Bell and coworkers stabilized HMGB1b by engineering His residues into the HMG domain of HMGB1b. They show metal binding to a pair of histidines which stabilized HMGB1b but had no effect on its binding affinity. The authors report HMGB1b as a marginally stable protein and any attempt to stabilize the protein increased its helicity and had no effect on its binding affinity. In addition, destabilizing the HMG domain of LEF-1 or HMGB1 of *Chironomus* or rat did not significantly alter the protein's binding affinity or bending capabilities (Wisniewski *et al*., 1997) (Falciola *et al*., 1994) (Teo *et al*., 1995). Wisniewski and colleagues mutated a well conserved tryptophan at position 14 to a phenylalanine or alanine which unfolded the protein. The binding affinity was measured by EMSA and showed no change in either of the mutants when compared to wild type (Wisniewski *et al*., 1997). Thus the function of the conformational flexibility in the HMG domain of HMGB protein family members may be to facilitate its fitting and binding to various forms of DNA or binding partners.

Several studies have shown the HMG domain binding to various forms of DNA and proteins. In one study, Wisniewski and colleagues reported the HMG domain (box A) of HMGB1 binding to various forms of DNA, such as single stranded, double stranded, cruciform, or other unusual DNA conformations (Wisniewski *et al*., 1997). Thus the flexibility in the HMG domain allows the protein to find the appropriate fit for its binding partner. In addition to binding DNA, the HMG domain of HMGB family members also bind proteins. The conformational flexibility in the HMG domain of mouse HMGB2 allows the protein to bind to the POU domain of the octamer transcription factor, Oct2 in addition to binding DNA (Zwilling *et al*., 1995). Thus, the conformational disorder in the HMG domain of LEF-1 may allow the protein to adapt to different binding partners.

LEF-1 has been shown to bind to β-catenin, Groucho/Transducin-like-Enhancer (TLE1-4), and DNA (Arce *et al*., 2009) (Daniels and Weis, 2005). The interaction between LEF-1 and Groucho/TLE occurs in the context dependent activation domain (residues 242-247) and HMG domain (residues 296-396). Drosophila Groucho and its human orthologous forms, TLE1-4, function as transcription co-repressors within the Wnt signaling pathway (Levanon *et al*., 1998). Groucho and TLE are nuclear proteins that are globally expressed and do not bind DNA. They repress by binding to other proteins (bHLH, Runx, Pax5/BSAP, Six, and LEF/TCFs) (Arce *et al*., 2009). Groucho and TLE's contain five domains (*e.g.* Q, GP, CcN, SP, and WD) (as shown below Figure 4.5).

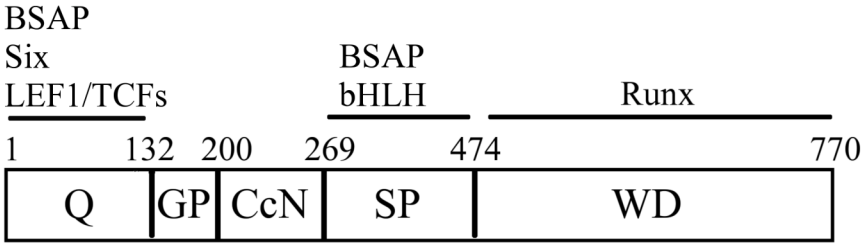


Figure 4.5. Schematic representation of Groucho/TLE. A diagram of the primary structure of Groucho/TLE. The proteins that bind to each domain are listed above each domain (BSAP, B-cell specific activator protein; bHLH, basic helix-loop-helix protein; Runx, runt domain transcription factor).

Groucho/TLE represses the LEF-1 transcription cascade until a Wnt protein binds a Frizzled receptor and the lipoprotein receptor-related protein 5/6 (LRP 5/6). The protein Disheveled is then recruited to the receptors which then recruits Axin and GSK3. This disrupts the protein degradation complex and there is an accumulation of β-catenin in the cytosol. β-catenin then translocates to the nucleus where the corepressor Groucho/TLE is displaced from its binding site on the LEF-1/TCF protein by β-catenin and transcription is then initiated (Arce *et al*., 2009). β-catenin has been shown to bind to residues 1-143 (β-catenin binding domain) and 265-397 (includes parts of the central domain and the HMG domain) of LEF-1 (Daniels and Weis, 2005). Thus the disorder may be to accomodate the binding of β-catenin, TLE/Groucho, or DNA to the HMG domain.

Summary. The electrophorectic mobility shift assay (EMSA) provided insight into the binding affinity of the DNA while the circular permutation assay was used to compare the DNA bending ability of LEF-1 and the variants. The results indicate the change in stability of the HMG domain of LEF-1 does not affect the degree by which the LEF-1 domain bends DNA but does have a small yet consistent affect on the binding affinity. Overall, the packing of amino acid residues in the major hydrophobic core is predicted to dictate how well the HMG domain of LEF-1 binds DNA. For those variants that contain the addition of tryptophan residues in the major hydrophobic core, there is a small decrease in binding affinity. The tryptophans in the hydrophobic core may have increased the overall volume of the major hydrophobic core and possibly change the surface complementarity between the protein and DNA. Some variants showed a small increase in the binding affintiy for cognate DNA. Mutations made in these variants did not include tryptophans in the major hydrophobic core. Thus the packing of the hydrophobic core and the surface complementarity between the HMG domain and the DNA may be similar to or slightly better than wild type. Lastly, some variants with greater stability than wild type and one variant less stable exhibit equilibrium binding constants similar to wild type. These reults suggest that mutating residues that do not contact DNA have little to no effect on binding and bending DNA *in vitro*. Thus far, studies have only shown the mutation of amino acid residues that come into direct contact with the DNA to significantly alter the DNA binding affinity and bend angle.

4.4 Future Direction

Future Study #1

LEF-1 is a mediator of the Wnt signaling pathway and is activated by the recruitment of β-catenin (Sethi and Vidal-Puig, 2010). LEF-1 is an architectural transcription factor that initiates transcription by bending its cognate DNA to facilitate interaction between transcription factors (e.g. Ets-1, PEBP2α, and ATF-CREB) (Love *et al*., 2004). The stability of the HMG domain of LEF-1 may have a greater impact on transcriptional activity *in vivo* than *in vitro*. To study Wnt/β-catenin signaling, two commonly used assays involves transcription based reporter systems (*e.g.* TOPFlash, SuperTOPFlash, and β-catenin activated reporter system(BAR)), and a measure of transcription profile of Wnt/β-catenin target genes (Biechele and Moon, 2008).

Transcription based reporter systems used to study the Wnt/β-catenin signaling pathway involve the placement of multiple (*e.g.* three, eight, or twelve) recognition sequences for TCF/LEF-1 (*i.e.* LEF-1, CCTTTGAAG; TCF-1, CCTTTGATC) upstream of a basal promoter (*e.g. c-fos*, thymidine kinase, Clonetech's TA promoter, or Promega's minP promoter) followed by the luciferase reporter gene in a plasmid that is transiently transfected into human cell lines (*e.g.* Hela or HEK293). The observed luciferase activity in cultured cells is an artificial readout of the Wnt/β-catenin signaling pathway. To compliment the reporter studies, the transcription of *Axin2*, a Wnt/β-catenin signaling pathway target gene, should be monitored by real time PCR. Thus, transiently transfecting wild type or the variants with the transcription based reporter system in cultured cells will allow one to monitor

the LEF-1/variants transcriptional activity through the Wnt/β-catenin signaling pathway.

Future Study #2

The intrinsic disorder in the HMG domain of LEF-1 may allow the protein to adapt to different binding partners (discussed in the previous section). To determine the effects of stability on LEF-1's ability to bind to TLE, one can measure the binding affinity (binding constant) of the HMG domain for TLE. SPR (surface plasmon resonance) can be used to measure the binding constant of wild type and the variants for TLE. TLE will be immobilized on a gold plated sensor chip. Wild type LEF-1 will be placed in a dielectric medium (buffered solution) that will flow over the immobilized TLE sensor chip. As wild type LEF-1 binds to TLE it will change the reflected polarized light from the sensor chip and an increase in the SPR signal (RU, response unit) will occur. The addition of LEF-1 will then cease and the dissociation of LEF-1 from TLE will occur. The SPR signal (RU, response unit) will decrease. The association rate (k_a) and the dissociation rate (k_d) can be calculated from the SPR signal and the binding constant (K_D) can be calculated by dividing k_d / k_a .

Future Study #3

The two-dimensional [^{1}H,^{15}N] HSQC spectra of variant 4 (T26P, L27G, Q49K, Figure 3.4) and variant 8 (Y13W, T26P, L27G, L44M, Q49K, Y52W, Figure 3.6) contained more peaks, a higher degree of dispersion and more homogeneous peak

shape and intensity indicative of a more compact fold that tumbles more isotropically (*i.e.,* exhibiting more homogeneous T2 relaxation) than wild type LEF-1 (Figure 3.2). These features render variant 4 and variant 8 ideal candidates for NMR structure elucidation. The NMR structure of variant 4 and 8 will provide insight to the intermolecular interactions that stabilize the HMG domain of LEF-1.

REFERENCES

Chapter 1

Arce, L., Yokoyama, N.N., Waterman, M.L. (2006) Diversity of LEF/TCF action in development and disease. Oncogene 25:7492-7504.

Behrens, J., von Kries, J.P., Kühl, M., Bruhn, L., Wedlich, D., Grosschedl, R., Birchmeier, W. (1996) Functional interaction of beta-catenin with the transcription factor LEF-1. Nature 382:638-642.

Bruhn, L., Munnerlyn, A., Grosschedl, R. (1997) ALY, a context-dependent coactivator of LEF-1 and AML-1, is required for TCRalpha enhancer function. Genes & Development 11:640-653.

Bustin, M. (2001) Revised nomenclature for high mobility group (HMG) chromosomal proteins. Trends Biochem Sci 26:152-153.

Bustin, M. (2010) High mobility group proteins. BBA - Gene Regulatory Mechanisms 1799:1-2.

Clevers, H.C., Grosschedl, R. (1996) Transcriptional control of lymphoid development: Lessons from gene targeting. Immunol Today 17:336-343.

Cleynen, I., Van de Ven,W.J.M. (2008) The HMGA proteins: A myriad of functions (review). Int J Oncol 32:289-305.

Croce, J.C., and McClay, D.R. (2006). The canonical Wnt pathway in embryonic axis polarity. Semin. Cell Dev. Biol. 17, 168-174.

Daniels, D.L., and Weis, W.I. (2005). Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. Nat Struct Mol Biol 12, 364-371.

Dik, W.A., Pike-Overzet, K., Weerkamp, F., de Ridder, D., F.E. de Haas, E., Baert, M.R.M., van der Spek,P., Koster, E.E.L., Reinders, M.J.T., van Dongen, J.J.M., Langerak, A.W., and Staal, F.J.T. (2005). New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. J. Exp. Med. 201, 1715-1723.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nature Reviews Molecular Cell Biology 6, 197-208.

Ferrando, A.A., Neuberg, D.S., Staunton, J., Loh, M.L., Huard, C., Raimondi, S.C., Behm, F.G., Pui, C.H., Downing, J.R., Gilliland, D.G.*, et al.* (2002). Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell 1, 75-87.

Gerlitz, G., Hock, R., Ueda, T., and Bustin, M. (2009). The dynamics of HMG protein-chromatin interactions in living cells. Biochem. Cell Biol. 87, 127-137.

Goodwin, G.H., Sanders, C., and Johns, E.W. (1973). A new group of chromatin-associated proteins with a high content of acidic and basic amino acids. Eur. J. Biochem. 38, 14-19.

Gutierrez, A., Sanda, T., Ma, W., Zhang, J., Grebliunaite, R., Dahlberg, S., Neuberg, D., Protopopov, A., Winter, S.S., Larson, R.S.*, et al.* (2010). Inactivation of LEF1 in T-cell acute lymphoblastic leukemia. Blood 115, 2845-2851.

Hock, R., Furusawa, T., Ueda, T., and Bustin, M. (2007). HMG chromosomal proteins in development and disease. Trends Cell Biol. 17, 72-79.

Hoppler, S., and Kavanagh, C.L. (2007). Wnt signalling: variety at the core. J. Cell. Sci. 120, 385-393.

Hovanes, K., Li, T.W., Munguia, J.E., Truong, T., Milovanovic, T., Marsh, J.L., Holcombe, R.F., and Waterman, M.L. (2001). Beta-catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. Nat. Genet. 28, 53-57.

Jeong, E.G., Lee, S.H., Yoo, N.J., and Lee, S.H. (2007). Mutational analysis of Wnt pathway gene LEF1 in common human carcinomas. Dig. Liver Dis. 39, 287-288.

Jesse, S., Koenig, A., Ellenrieder, V., and Menke, A. (2010). Lef-1 isoforms regulate different target genes and reduce cellular adhesion. Int. J. Cancer 126, 1109-1120.

Joo, J.C., Pohkrel, S., Pack, S.P., and Yoo, Y.J. (2010). Thermostabilization of Bacillus circulans xylanase via computational design of a flexible surface cavity. J. Biotechnol. 146, 31-39.

Kazlauskas, R.J., and Bornscheuer, U.T. (2009). Finding better protein engineering strategies. Nat Chem Biol 5, 526-529.

Kim, S.J., Lee, J.A., Joo, J.C., Yoo, Y.J., Kim, Y.H., and Song, B.K. (2010). The development of a thermostable CiP (Coprinus cinereus peroxidase) through in silico design. Biotechnol Progress 26, 1038-1046.

Ko, M., Bowen, L.D., Narasimhan, D., Berlin, A.A., Lukacs, N.W., Sunahara, R.K., Cooper, Z.D., and Woods, J.H. (2007). Cocaine esterase: interactions with cocaine and immune responses in mice. J. Pharmacol. Exp. Ther. 320, 926-933.

Korkegian, A., Black, M.E., Baker, D., and Stoddard, B.L. (2005). Computational thermostabilization of an enzyme. Science 308, 857-860.

Kragelund, B.B., Jönsson, M., Bifulco, G., Chazin, W.J., Nilsson, H., Finn, B.E., and Linse, S. (1998). Hydrophobic core substitutions in calbindin D9k: effects on Ca2+ binding and dissociation. Biochemistry (N. Y.) 37, 8926-8937.

Kragelund, B.B., Poulsen, K., Andersen, K.V., Baldursson, T., Krøll, J.B., Neergåard, T.B., Jepsen, J., Roepstorff, P., Kristiansen, K., Poulsen, F.M., and Knudsen, J. (1999). Conserved residues and their role in the structure, function, and stability of acyl-coenzyme A binding protein. Biochemistry (N. Y.) 38, 2386-2394.

Kratochwil, K., Dull, M., Farinas, I., Galceran, J., and Grosschedl, R. (1996). Lef1 expression is activated by BMP-4 and regulates inductive tissue interactions in tooth and hair development. Genes Dev. 10, 1382-1394.

Kuhlman, B., and Baker, D. (2000). Native protein sequences are close to optimal for their structures. Proceedings of the National Academy of Sciences 97, 10383-10388.

Lee, B., and Vasmatzis, G. (1997). Stabilization of protein structures. Curr. Opin. Biotechnol. 8, 423-428.

Leiden, J.M., and Thompson, C.B. (1994). Transcriptional regulation of T-cell genes during T-cell development. Curr. Opin. Immunol. 6, 231-237.

Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic disorder in transcription factors. Biochemistry 45, 6873-6888.

Liu, Z., Zhang, J., Wang, X., Ding, Y., Wu, J., and Shi, Y. (2009). Temperature-induced partially unfolded state of hUBF HMG Box-5: conformational and dynamic investigations of the Box-5 thermal intermediate ensemble. Proteins 77, 432-447.

Love, J.J., Li, X., Chung, J., Dyson, H.J., and Wright, P.E. (2004). The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. Biochemistry (N. Y.) 43, 8725-8734.

Masse, J.E., Wong, B., Yen, Y., Allain, F.H.T., Johnson, R.C., and Feigon, J. (2002). The S. cerevisiae architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding. J. Mol. Biol. 323, 263-284.

McCauley, M.J., Zimmerman, J., Maher, L.J., and Williams, M.C. (2007). HMGB binding to DNA: single and double box motifs. J. Mol. Biol. 374, 993-1004.

Miller, M. (2009). The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. Curr Protein Pept Sci 10, 244-269.

Minezaki, Y., Homma, K., Kinjo, A.R., and Nishikawa, K. (2006). Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. J. Mol. Biol. 359, 1137-1149.

Molenaar, M., Wetering, M.v.D., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destrée, O., and Clevers, H. (1996). XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. Cell 86, 391-399.

Nawshad, A., and Hay, E.D. (2003). TGFbeta3 signaling activates transcription of the LEF1 gene to induce epithelial mesenchymal transformation during mouse palate development. J. Cell Biol. 163, 1291-1301.

Niemann, C., Owens, D.M., Hülsken, J., Birchmeier, W., and Watt, F.M. (2002). Expression of DeltaN Lef1 in mouse epidermis results in differentiation of hair follicles into squamous epidermal cysts and formation of skin tumours. Development 129, 95-109.

Okamura, R.M., Sigvardsson, M., Galceran, J., Verbeek, S., Clevers, H., and Grosschedl, R. (1998). Redundant regulation of T cell differentiation and TCRalpha gene expression by the transcription factors LEF-1 and TCF-1. Immunity 8, 11-20.

Pfannkuche, K., Summer, H., Li, O., Hescheler, J., and Dröge, P. (2009). The High Mobility Group Protein HMGA2: A Co-Regulator of Chromatin Structure and Pluripotency in Stem Cells? Stem Cell Rev and Rep 5, 224-230.

Reeves, R. (2010). Nuclear functions of the HMG proteins. Biochim. Biophys. Acta 1799, 3-14.

Robertson, S.M., Lo, M., Odom, R., Yang, X., Medina, J., Huang, S., and Lin, R. (2011). Functional analyses of vertebrate TCF proteins in C. elegans embryos. Dev. Biol. 355, 115-123.

Ross, D.A., and Kadesch, T. (2001). The notch intracellular domain can function as a coactivator for LEF-1. Mol. Cell. Biol. 21, 7537-7544.

Sasaki, T., Ito, Y., Xu, X., Han, J., Bringas, P., Maeda, T., Slavkin, H.C., Grosschedl, R., and Chai, Y. (2005). LEF1 is a critical epithelial survival factor during tooth morphogenesis. Dev. Biol. 278, 130-143.

Saven, J.G. (2011). Computational protein design: engineering molecular diversity, nonnatural enzymes, nonbiological cofactor complexes, and membrane proteins. Curr. Opin. Chem. Biol. 15, 452-457.

Skokowa, J., Cario, G., Uenalan, M., Schambach, A., Germeshausen, M., Battmer, K., Zeidler, C., Lehmann, U., Eder, M., Baum, C.*, et al.* (2006). LEF-1 is crucial for neutrophil granulocytopoiesis and its expression is severely reduced in congenital neutropenia. Nat. Med. 12, 1191-1197.

Skokowa, J., and Welte, K. (2007). LEF-1 is a decisive transcription factor in neutrophil granulopoiesis. Ann. N. Y. Acad. Sci. 1106, 143-151.

Stros, M., Launholt, D., and Grasser, K.D. (2007). The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cell Mol. Life Sci. 64, 2590-2606.

Stros, M. (2010). HMGB proteins: Interactions with DNA and chromatin. Biochim. Biophys. Acta 1799, 101-113.

Teilum, K., Olsen, J.G., and Kragelund, B.B. (2011). Protein stability, flexibility and function. Biochim. Biophys. Acta 1814, 969-976.

Travers, A. (2000). Recognition of distorted DNA structures by HMG domains. Curr. Opin. Struct. Biol. 10, 102-109.

Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. (1991). LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function. Genes & Development 5, 880-894.

Vacik, T., and Lemke, G. (2011). Dominant-negative isoforms of Tcf/Lef proteins in development and disease. Cell Cycle 10, 4199-4200.

van de Wetering, M., Cavallo, R., Dooijes, D., van Beest, M., van Es, J., Loureiro, J., Ypma, A., Hursh, D., Jones, T., Bejsovec, A.*, et al.* (1997). Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. Cell 88, 789-799.

van Genderen, C., Okamura, R., Farinas, I., Quo, R., Parslow, T., Bruhn, L., and Grosschedl, R. (1994). Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. Genes Dev. 8, 2691-2703.
Wisniewska, M.B., Misztal, K., Michowski, W., Szczot, M., Purta, E., Lesniak, W., Klejman, M.E., Dabrowski, M., Filipkowski, R.K., Nagalski, A., Mozrzymas, J.W., and Kuznicki, J. (2010). LEF1/beta-catenin complex regulates transcription of the Cav3.1

calcium channel gene (Cacna1g) in thalamic neurons of the adult brain. Journal of Neuroscience 30, 4957-4969.

Wong, T.S., Rajagopalan, S., Freund, S.M., Rutherford, T.J., Andreeva, A., Townsley, F.M., Petrovich, M., and Fersht, A.R. (2009). Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53. Nucleic Acids Res. 37, 6765-6783.

Wright, P.E., and Dyson, H.J. (2009). Linking folding and binding. Curr. Opin. Struct. Biol. 19, 31-38.

Zhang, Q., and Wang, Y. (2008). High mobility group proteins and their post-translational modifications. Biochim. Biophys. Acta 1784, 1159-1166.

Chapter 2

Alonso, L., and Fuchs, E. (2006). The hair cycle. J. Cell. Sci. 119, 391-393.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

Arce, L., Yokoyama, N.N., and Waterman, M.L. (2006). Diversity of LEF/TCF action in development and disease. Oncogene 25, 7492-7504.

Atcha, F.A., Syed, A., Wu, B., Hoverter, N.P., Yokoyama, N.N., Ting, J.T., Munguia, J.E., Mangalam, H.J., Marsh, J.L., and Waterman, M.L. (2007). A unique DNA binding domain converts T-cell factors into strong Wnt effectors. Mol. Cell. Biol. 27, 8352-8363.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Finn, R.D., and Sonnhammer, E.L. (1999). Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. Nucleic Acids Res. 27, 260-262.

Birney, E., Thompson, J.D., and Gibson, T.J. (1996). PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. Nucleic Acids Res. 24, 2730-2739.

Clevers, H.C., and Grosschedl, R. (1996). Transcriptional control of lymphoid development: lessons from gene targeting. Immunol. Today 17, 336-343.

Dahiyat, B.I., and Mayo, S.L. (1996). Protein design automation. Protein Sci. 5, 895-903.

Doi, T., Puri, P., Bannigan, J., and Thompson, J. (2011). Altered PITX2 and LEF1 gene expression in the cadmium-induced omphalocele in the chick model. Pediatr. Surg. Int. 27, 495-499.

Dunbrack, R.L., and Cohen, F.E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci. 6, 1661-1681.

Gamba, L., Cubedo, N., Lutfalla, G., Ghysen, A., and Dambly-Chaudiere, C. (2010). Lef1 controls patterning and proliferation in the posterior lateral line system of zebrafish. Dev. Dyn. 239, 3163-3171.

Giese, K., Amsterdam, A., and Grosschedl, R. (1991). DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. Genes {\&} Development 5, 2567-2578.

Gradl, D., König, A., and Wedlich, D. (2002). Functional diversity of Xenopus lymphoid enhancer factor/T-cell factor transcription factors relies on combinations of activating and repressing elements. J. Biol. Chem. 277, 14159-14171.

Graham, T.A., Weaver, C., Mao, F., Kimelman, D., and Xu, W. (2000). Crystal structure of a beta-catenin/Tcf complex. Cell 103, 885-896.

Gribskov, M., Lüthy, R., and Eisenberg, D. (1990). Profile analysis. Meth Enzymol 183, 146-159.

Hoppler, S., and Kavanagh, C.L. (2007). Wnt signalling: variety at the core. J. Cell. Sci. 120, 385-393.

Hsu, S.C., Galceran, J., and Grosschedl, R. (1998). Modulation of transcriptional regulation by LEF-1 in response to Wnt-1 signaling and association with beta-catenin. Mol. Cell. Biol. 18, 4807-4818.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., Mcwilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R.*, et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R., and Wright, P.E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature 376, 791-795.

Macdonald, B.T., Tamai, K., and He, X. (2009). Wnt/ß-Catenin Signaling: Components, Mechanisms, and Diseases. Developmental Cell 17, 9-26.

Maeyer, M.D., Desmet, J., and Lasters, I. (2000). The dead-end elimination theorem: mathematical aspects, implementation, optimizations, evaluation, and performance. Methods Mol. Biol. 143, 265-304.

Merrill, B.J., Gat, U., DasGupta, R., and Fuchs, E. (2001). Tcf3 and Lef1 regulate lineage differentiation of multipotent stem cells in skin. Genes & Development 15, 1688-1705.

Milatovich, A., Travis, A., Grosschedl, R., and Francke, U. (1991). Gene for lymphoid enhancer-binding factor 1 (LEF1) mapped to human chromosome 4 (q23-q25) and mouse chromosome 3 near Egf. Genomics 11, 1040-1048.

Najdi, R., Holcombe, R.F., and Waterman, M.L. (2011). Wnt signaling and colon carcinogenesis: beyond APC. J Carcinog 10, 5-22.

Poy, F., Lepourcelet, M., Shivdasani, R.A., and Eck, M.J. (2001). Structure of a human Tcf4-beta-catenin complex. Nat. Struct. Biol. 8, 1053-1057.

Pukrop, T., Gradl, D., Henningfeld, K.A., Knochel, W., Wedlich, D., and Kuhl, M. (2001). Identification of two regulatory elements within the high mobility group box transcription factor XTCF-4. J. Biol. Chem. 276, 8968-8978.

Reya, T., and Clevers, H. (2005). Wnt signalling in stem cells and cancer. Nature 434, 843-850.

Robertson, S.M., Lo, M., Odom, R., Yang, X., Medina, J., Huang, S., and Lin, R. (2011). Functional analyses of vertebrate TCF proteins in C. elegans embryos. Dev. Biol. 355, 115-123.

Schueler-Furman, O., and Baker, D. (2003). Conserved residue clustering and protein structure prediction. Proteins 52, 225-235.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA 95, 5857-5864.

Sethi, J.K., and Vidal-Puig, A. (2010). Wnt signalling and the control of cellular metabolism. Biochem. J. 427, 1-17.

Shih, C., Chang, C., Lin, Y., Lo, W., and Hwang, J. (2012). Evolutionary information hidden in a single protein structure. Proteins 80, 1647-1657.

Standley, H.J., Destrée, O., Kofron, M., Wylie, C., and Heasman, J. (2006). Maternal XTcf1 and XTcf4 have distinct roles in regulating Wnt target genes. Dev. Biol. 289, 318-328.

Sun, J., and Weis, W.I. (2011). Biochemical and structural characterization of ß-catenin interactions with nonphosphorylated and CK2-phosphorylated Lef-1. J. Mol. Biol. 405, 519-530.

Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. (1991). LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function. Genes & Development 5, 880-894.

Vacik, T., and Lemke, G. (2011). Dominant-negative isoforms of Tcf/Lef proteins in development and disease. Cell Cycle 10, 4199-4200.

van de Wetering, M., Castrop, J., Korinek, V., and Clevers, H. (1996). Extensive alternative splicing and dual promoter usage generate Tcf-1 protein isoforms with differential transcription control properties. Mol. Cell. Biol. 16, 745-752.

van de Wetering, M., Oosterwegel, M., Dooijes, D., and Clevers, H. (1991). Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. EMBO J. 10, 123-132.

van Genderen, C., Okamura, R., Farinas, I., Quo, R., Parslow, T., Bruhn, L., and Grosschedl, R. (1994). Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. Genes Dev. 8, 2691-2703.

Chapter 3

Cary, P.D., Read, C.M., Davis, B., Driscoll, P.C., and Crane-Robinson, C. (2001). Solution structure and backbone dynamics of the DNA-binding domain of mouse Sox-5. Protein Sci. 10, 83-98.

Chen, J., and Stites, W.E. (2001). Packing is a key selection factor in the evolution of protein hydrophobic cores. Biochemistry (N. Y.) 40, 15280-15289.

Churchill, M.E.A., Klass, J., and Zoetewey, D.L. (2010). Structural Analysis of HMGD--DNA Complexes Reveals Influence of Intercalation on Sequence Selectivity and DNA Bending. J. Mol. Biol. 403, 88-102.

Dill, K.A., Ozkan, S.B., Shell, M.S., and Weikl, T.R. (2008). The protein folding problem. Annu Rev Biophys 37, 289-316.

Dobson, C.M. (2003). Protein folding and misfolding. Nature 426, 884-890.

Dunker, A.K., Oldfield, C.J., Meng, J., Romero, P., Yang, J.Y., Chen, J.W., Vacic, V., Obradovic, Z., and Uversky, V.N. (2008). The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics 9 Suppl 2, S1.

Dyson, H.J., and Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. Curr. Opin. Struct. Biol. 12, 54-60.

Hardman, C.H., Broadhurst, R.W., Raine, A.R., Grasser, K.D., Thomas, J.O., and Laue, E.D. (1995). Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. Biochemistry (N. Y.) 34, 16596-16607.

Hazy, E., and Tompa, P. (2009). Limitations of induced folding in molecular recognition by intrinsically disordered proteins. Chemphyschem 10, 1415-1419.

Jones, D.N., Searles, M.A., Shaw, G.L., Churchill, M.E., Ner, S.S., Keeler, J., Travers, A.A., and Neuhaus, D. (1994). The solution structure and dynamics of the DNA-binding domain of HMG-D from Drosophila melanogaster. Structure 2, 609-627.

Kasai, N., Tsunaka, Y., Ohki, I., Hirose, S., Morikawa, K., and Tate, S. (2005). Solution structure of the HMG-box domain in the SSRP1 subunit of FACT. J. Biomol. NMR 32, 83-88.

Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic disorder in transcription factors. Biochemistry 45, 6873-6888.

Liu, Z., Zhang, J., Wang, X., Ding, Y., Wu, J., and Shi, Y. (2009). Temperature-induced partially unfolded state of hUBF HMG Box-5: conformational and dynamic investigations of the Box-5 thermal intermediate ensemble. Proteins 77, 432-447.

Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R., and Wright, P.E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature 376, 791-795.

Love, J.J., Li, X., Chung, J., Dyson, H.J., and Wright, P.E. (2004). The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. Biochemistry (N. Y.) 43, 8725-8734.

Masse, J.E., Wong, B., Yen, Y., Allain, F.H.T., Johnson, R.C., and Feigon, J. (2002). The S. cerevisiae architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding. J. Mol. Biol. 323, 263-284.

Maxwell, K.L., and Davidson, A.R. (1998). Mutagenesis of a buried polar interaction in an SH3 domain: sequence conservation provides the best prediction of stability effects. Biochemistry (N. Y.) 37, 16172-16182.

Metallo, S.J. (2010). Intrinsically disordered proteins are potential drug targets. Curr. Opin. Chem. Biol. 14, 481-488.

Miller, M. (2009). The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. Curr Protein Pept Sci 10, 244-269.

Murphy, F.V., Sweet, R.M., and Churchill, M.E. (1999). The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. EMBO J. 18, 6610-6618.

Phillips, N.B., Racca, J., Chen, Y., Singh, R., Jancso-Radek, A., Radek, J.T., Wickramasinghe, N.P., Haas, E., and Weiss, M.A. (2011). Mammalian testis-determining factor SRY and the enigma of inherited human sex reversal: frustrated induced fit in a bent protein-DNA complex. J. Biol. Chem. 286, 36787-36807.

Read, C.M., Cary, P.D., Crane-Robinson, C., Driscoll, P.C., and Norman, D.G. (1993). Solution structure of a DNA-binding domain from HMG1. Nucleic Acids Res. 21, 3427-3436.

Religa, T.L., Johnson, C.M., Vu, D.M., Brewer, S.H., Dyer, R.B., and Fersht, A.R. (2007). The helix-turn-helix motif as an ultrafast independently folding domain: the pathway of folding of Engrailed homeodomain. Proc. Natl. Acad. Sci. USA 104, 9272-9277.

Schanda, P., Brutscher, B., Konrat, R., and Tollinger, M. (2008). Folding of the KIX domain: characterization of the equilibrium analog of a folding intermediate using 15N/13C relaxation dispersion and fast 1H/2H amide exchange NMR spectroscopy. J. Mol. Biol. 380, 726-741.

Sigalov, A.B. (2010). Protein intrinsic disorder and oligomericity in cell signaling. Mol.BioSyst. 6, 451-461.

Spolar, R.S., and Record, M.T. (1994). Coupling of local folding to site-specific binding of proteins to DNA. Science 263, 777-784.

Sugase, K., Dyson, H.J., and Wright, P.E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447, 1021-1025.

Travaglini-Allocatelli, C., Ivarsson, Y., Jemth, P., and Gianni, S. (2009). Folding and stability of globular proteins and implications for function. Curr. Opin. Struct. Biol. 19, 3-7.

Travers, A. (2000). Recognition of distorted DNA structures by HMG domains. Curr. Opin. Struct. Biol. 10, 102-109.

Uversky, V.N., and Dunker, A.K. (2010). Understanding protein non-folding. Biochim. Biophys. Acta 1804, 1231-1264.

Vuzman, D., and Levy, Y. (2012). Intrinsically disordered regions as affinity tuners in protein-DNA interactions. Mol.BioSyst. 8, 47-57.

Wang, D., Zhang, J., Jin, X., Wu, J., and Shi, Y. (2007). Investigation of the structural stability of hUBF HMG box 5 by native-state hydrogen exchange. Biochemistry (N. Y.) 46, 1293-1302.

Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R., Laue, E.D., and Thomas, J.O. (1993). Structure of the HMG box motif in the B-domain of HMG1. EMBO J. 12, 1311-1319.

Weiss, M.A. (2001). Floppy SOX: mutual induced fit in hmg (high-mobility group) box-DNA recognition. Mol. Endocrinol. 15, 353-362.

Werner, M.H., Huth, J.R., Gronenborn, A.M., and Clore, G.M. (1995). Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. Cell 81, 705-714.

Wright, P., and Dyson, H. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J. Mol. Biol. 293, 321-331.

Wright, P.E., and Dyson, H.J. (2009). Linking folding and binding. Curr. Opin. Struct. Biol. 19, 31-38.

Xu, Y., Yang, W., Wu, J., and Shi, Y. (2002). Solution structure of the first HMG box domain in human upstream binding factor. Biochemistry (N. Y.) 41, 5415-5420.

Yang, J.S., Wallin, S., and Shakhnovich, E.I. (2008). Universality and diversity of folding mechanics for three-helix bundle proteins. Proc. Natl. Acad. Sci. USA 105, 895-900.

Chapter 4

Arce, L., Pate, K.T., and Waterman, M.L. (2009). Groucho binds two conserved regions of LEF-1 for HDAC-dependent repression. BMC Cancer 9, 159.

Bell, A.J., Xin, H., Taudte, S., Shi, Z., and Kallenbach, N.R. (2002). Metal-dependent stabilization of an active HMG protein. Protein Eng. 15, 817-825.

Biechele, T.L., and Moon, R.T. (2008). Assaying beta-catenin/TCF transcription with beta-catenin/TCF transcription-based reporter constructs. Methods Mol. Biol. 468, 99-110.

Boehr, D.D., Nussinov, R., and Wright, P.E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol 5, 789-796.

Carlsson, P., Waterman, M.L., and Jones, K.A. (1993). The hLEF/TCF-1 alpha HMG protein contains a context-dependent transcriptional activation domain that induces the TCR alpha enhancer in T cells. Genes & Development 7, 2418-2430.

Churchill, M.E.A., Klass, J., and Zoetewey, D.L. (2010). Structural Analysis of HMGD--DNA Complexes Reveals Influence of Intercalation on Sequence Selectivity and DNA Bending. J. Mol. Biol. 403, 88-102.

Daniels, D.L., and Weis, W.I. (2005). Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. Nat Struct Mol Biol 12, 364-371.

Falciola, L., Murchie, A.I., Lilley, D.M., and Bianchi, M. (1994). Mutational analysis of the DNA binding domain A of chromosomal protein HMG1. Nucleic Acids Res. 22, 285-292.

Garza, A.S., Ahmad, N., and Kumar, R. (2010). Role of intrinsically disordered protein regions/domains in transcriptional regulation. Life Sci. 84, 189-193.

Giese, K., Amsterdam, A., and Grosschedl, R. (1991). DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. Genes & Development 5, 2567-2578.

Hippel, P.H.v., and Berg, O.G. (1989). Facilitated target location in biological systems. J. Biol. Chem. 264, 675-678.

Lee, B., and Vasmatzis, G. (1997). Stabilization of protein structures. Curr. Opin. Biotechnol. 8, 423-428.

Levanon, D., Goldstein, R.E., Bernstein, Y., Tang, H., Goldenberg, D., Stifani, S., Paroush, Z., and Groner, Y. (1998). Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. Proc. Natl. Acad. Sci. USA 95, 11590-11595.

Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R., and Wright, P.E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature 376, 791-795.

Love, J.J., Li, X., Chung, J., Dyson, H.J., and Wright, P.E. (2004). The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. Biochemistry (N. Y.) 43, 8725-8734.

Miller, M. (2009). The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. Curr Protein Pept Sci 10, 244-269.

Murphy, E.C., Zhurkin, V.B., Louis, J.M., Cornilescu, G., and Clore, G.M. (2001). Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. J. Mol. Biol. 312, 481-499.

Neuweiler, H., Sharpe, T.D., Rutherford, T.J., Johnson, C.M., Allen, M.D., Ferguson, N., and Fersht, A.R. (2009). The folding mechanism of BBL: Plasticity of transition-state structure observed within an ultrafast folding protein family. J. Mol. Biol. 390, 1060-1073.

Read, C.M., Cary, P.D., Preston, N.S., Lnenicek-Allen, M., and Crane-Robinson, C. (1994). The DNA sequence specificity of HMG boxes lies in the minor wing of the structure. EMBO J. 13, 5639-5646.

Sethi, J.K., and Vidal-Puig, A. (2010). Wnt signalling and the control of cellular metabolism. Biochem. J. 427, 1-17.

Teo, S.H., Grasser, K.D., Hardman, C.H., Broadhurst, R.W., Laue, E.D., and Thomas, J.O. (1995). Two mutations in the HMG-box with very different structural consequences provide insights into the nature of binding to four-way junction DNA. EMBO J. 14, 3844-3853.

Wisniewski, J.R., Hessler, K., Claus, P., and Zechel, K. (1997). Structural and functional consequences of mutations within the hydrophobic cores of the HMG1-box domain of the Chironomus high-mobility-group protein 1a. Eur. J. Biochem. 243, 151-159.

Yee, A., Savchenko, A., Ignachenko, A., Lukin, J., and Xu, X. (2005). NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. J. Am. Chem. Soc. 127, 16512-16517.

Zhang, Q., and Wang, Y. (2008). High mobility group proteins and their post-translational modifications. Biochim. Biophys. Acta 1784, 1159-1166.

Zwilling, S., König, H., and Wirth, T. (1995). High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors. EMBO J. 14, 1198-1208.