THE SOX9 DIMERIZATION DOMAIN AND ITS ROLE IN COOPERATIVE DNA BINDING

SARAH NEETA RAMSOOK

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

GRADUATE PROGRAM, DEPARTMENT OF BIOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

CANADA

AUGUST 2015

ABSTRACT

Nucleic acid binding proteins are key regulators in developmental processes and controlling the onset of diseases. Human SOX9 is a group E member of the SOX transcription family of proteins characterized by a conserved amino acid region first identified as the high mobility group (HMG) with high affinity for DNA. In parallel, a 38 amino acid region preceding the HMG domain described as the dimerization domain supports the formation of cooperative dimers on promoters with inverted sites, by binding to the HMG domain. Cooperative binding occurs when the binding at a first site increases the rate at which binding occurs at a second site between identical or non-identical proteins. Using mutagenesis and electrophoretic mobility shift assays (EMSA) the trans-dimerization model of dimeric SOX9 binding to DNA was studied. The molecular model of the SOX9 dimeric complex is characterized as a dimer by way of the amphipathic helix of the dimerization domain of one SOX9 protein docks on the HMG domain of an identical SOX9 protein in the DNA-bound state. Cooperative binding of one protein to the consensus-binding site may facilitate binding to the non-consensus site. This model provides insight to the characterized mutations associated with Campomelic dysplasia, an embryonic malformation of the skeletal and reproductive system.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation for my mentor Dr. Logan Donaldson whom taught me the spirit of research and always gave me the benefit of doubt. He continues to believe in the work of his students and their results and pushes them to think creatively in all aspects of their life. He embodies a quote that I once read about having the attitude and the substance of a genius. Without the opportunity and mentoring provided by Dr. Logan Donaldson this dissertation would not be possible. I would like to thank my committee members Dr. Yi Sheng, Dr. Dorota Crawford, and Dr. Thilo Womelsdorf for your expertise and direction. I am grateful for all of your insight throughout my Masters and all that you have done to help support my career and future endeavors. Thank you to my collaborators Dr. Mark Bayfield, Sam Sharifi and Ana Vakiloroayaeu who have been apart of my research.

Alongside the support of my wonderful lab I gained the scientific intuition that transforms a student into a scientist. I would like to extend my sincerest gratitude to Dr. Jamie Kwan and Ekaterina Smirinova. Your strength, teachings, and patience have made this journey unforgettable. I will forever be in debt to your time, love, and what I have achieved as a result of you both. To my entire lab and Life Science Building friends, past and present, thank you for your care and company. You have all made this achievable and less stressful in your own way with great laughs, walks, conversations, and memories both inside and outside of LSB. Graduate school has graciously built life long friendships and love that will forever be irreplaceable.

Finally, to my friends and family that has made this entire journey even possible. You kept me striving, enabled me to relax, believed in me throughout my career choices and my nature to be a perpetual student at heart. You have been by my side for at least half of my life and are a true testament to my strength with every experience that has crossed my path. To my loving Mom and Dad, thank you for being the two kindest and compassionate people in my life. You have an unconditional love and support for my dreams that is insurmountable. You enable me to experience any and everything that may come my way and only worked your hardest in order for me to seize every opportunity. You are my personal councilors on all aspects of life. I will forever be indebt to you and I only hope to reciprocate all the hard work, love, laughs, and warm meals I have come home to on a daily basis. I love you both endlessly.

"To laugh often and love much; to win the respect of intelligent persons and the affection of children; to earn the approbation of honest citizens and endure the betrayal of false friends; to appreciate beauty; to find the best in others; to give of one's self; to leave the world a bit better; to have played and laughed with enthusiasm and sung with exultation; to know even one life has breathed easier because you have lived
—this is to have succeeded."
- Bessie Anderson Stanley

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

A/Ala	Alanine
AA	Amino Acid
ACAN	Aggrecan
AEBSF	4-(2-Aminoethyl)-benzenesulfonyl fluoride HCL
AGC-1	Aggrecan 1
AMH	Anti-Mullerian Hormone
BMP	Bone Morphogenetic Protein
bp	base pair
C	Carbon
CC36	36-base pair palindrome DNA
C/C'	Tandemly inverted promoters with unequal binding sites
CD	Campomelic Dysplasia
CD	Circular Dichrosim
CD-Rap	Cartilage-Derived Retinoic Acid-sensitive Protein
Col2a1	Collagen, type II, alpha 1
Col2a2	Collagen, type II, alpha 2
Col9a1	Collagen, type IX, alpha 1
Col9a2	Collagen, type IX, alpha 2
Col10a1	Collagen, type X, alpha 1
Col11a2	Collagen, type XI, alpha 2
CTNNB1	β-Catenin, Cadherin-Associated Protein, Beta 1
D	Aspartic Acid
DAX1	Dosage-sensitive sex reversal, Adrenal hypoplasia critical region, on chromosome X, gene 1
DBD	DNA Binding Domain
DNA	Deoxyribonucleic Acid
DPPA4	Developmental Pluripotency Associated 4
E	Glutamic Acid
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Assay
ETS1	E26 Transformation-Specific family 1
F	Phenylalanine

FBXO15	F-Box
FGF4	Fibroblast Growth Factor 4
FGF9	Fibroblast Growth Factor 9
FITC	Fluorescein Isothiocyanate
FOXL2	Forkhead box L2
FPLC	Fast Protein Liquid Chromatography
G	Glycine
GLI2/3	GLI Zinc Finger 2,3
H	Hydrogen
HCl	Hydrochloric acid
His_6	6 Histidine-tagged
hLa	human La
HMG	High Mobility Group
HSQC	Heteronuclear Single Quantum Coherence
I	Isoleucine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRES	Internal Ribosome Entry Site
K	Lysine
kb	kilobases
K_d	Dissociation constant
kDa	kilo Daltons
L/Leu	Leucine
LAM	La Motif
LARP7	La Ribonucleoprotein Domain Family, Member 7
LB	Luria-Bertani
EMSA	Electrophoretic Mobility Shift Assay
EDTA	Ethylene-di-amine-tetra-acetic acid
FITC	Fluorescein Isothiocyanate
M	Molar
MaSC	Mammary Stem Cell
MATN1	Matrilin-1, cartilage matrix protein
MEF2C	Myocyte-specific Enhancer Factor 2C
MHz	Megahertz
mM	milimolar

Mmp13	Matrix metallopeptidase 13
MWCO	Molecular Weight Cut Off
MYB	Myeloblastosis proto-oncogene transcription factor family
N	Nitrogen
NaCl	Sodium Chloride
NF1A	Nuclear Factor 1A
NHP6A	Non-histone chromosomal protein 6A
Ni-NTA	Nickel-Nitrotriacetic Acid
NLS	Nuclear Localization Signal
nm	Nanometer
NMR	Nuclear Magnetic Resonance spectroscopy
NR5A1	Nuclear Receptor family 5, group A, member 1
NRE	Nuclear Retention Element
NSS	Non-sequence Specific
nt	Nucleotide
OD	Optical Density
P32	Phosphorus-32
Pax1	Paired Box 1
PDB	Protein Data Bank (Accessed via pdb.org)
PGDS	Prostaglandin D Synthase
pH	potential Hydrogen
PKA	Protein Kinase A
PMSF	Phenylmethylsulphonylfluoride
POU5F1	POU Domain, class 5, transcription factor 1
PQA	Proline Glutamine Alanine
PQS	Proline Glutamine Serine
Q	Glutamine
R	Arginine
RAC1	Ras-Related C3 Botulinum Toxin Substrate 1, Rho-Family, Small GTP Binding Protein Rac1
RANKL	Receptor Activator of Nuclear factor Kappa-B Ligand
RNA	Ribonucleic Acid
RPC	Retinal Progenitor Cell
RRM	RNA Recognition Motif
RSPO1	R-spondin-1

S9WT	Single site DNA
SBM	Short Basic Motif
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SF1	Steroidogenic factor-1
SNAIL2	Snail Family Zinc Finger 2
SOX	SRY related high mobility group box
SRY	Sex Determining Region Chromosome Y
SS	Sequence Specific
T300	20mM Tris-HCl, 300mM NaCl
T/Thr	Threonine
TA	Transcriptional Activation Domain
TBE	Tris, Boric Acid, EDTA
TDF	Testies Determining Factor
TER	Telomerase
UTF1	Undifferentiated Embryonic cell Transcription Factor 1
UV	Ultraviolet
V	Valine
Vegfa	Vascular endothelial growth factor A
WNT4	Wingless-type MMTV integration site family, member 4
XX	Female
XY	Male
Y	Tyrosine
YAP1	Yes-Associated Protein 1
α	alpha helix
°C	degrees Celsius
μL	microlitre
μM	micromolar

CHAPTER I

LITERATURE REVIEW

1.1 The High Mobility Group (HMG) Family of Proteins

High mobility group (HMG) proteins originated from their high electrophoretic mobility on polyacrylamide gels (Bianchi and Agresti, 2005). They are the second most abundant chromatin binding proteins after histones (Bianchi and Agresti, 2005). HMG proteins contribute to the fine-tuning of transcription (Bianchi and Agresti, 2005). This group exerts inclusive genomic functions via the establishment of inactive and active chromatin domains and ultimately regulating genome accessibility (Bianchi and Agresti, 2005). The HMG domain consists of three alpha helices separated by loops that have a high affinity for bent, distorted, or linear DNA (Thomas, 2001). HMG proteins can be divided into three derived families based on their function **(Figure 1)** *(i)* the HMG-I(Y) or AT-hook family, *(ii)* the HMG-14/17 or nucleosome binding family, and *(iii)* the HMG1/2 or HMG-box family (Bianchi and Agresti, 2005). The HMG-box family can further be sub-characterized by the protein either binding to bent DNA described as the canonical HMG domain, while a small subset retain the propensity to bind to a specific sequence and induce a bend in the DNA, defined as the non-canonical HMG domain.

1.2 The Sox Family of Proteins: A Non-Canonical DNA-Binding Protein Family

The Sex-Determining Region Y (SRY) of the Y chromosome and the Sry-related HMG box (SOX) are transcription factor regulators that retain dynamic roles in

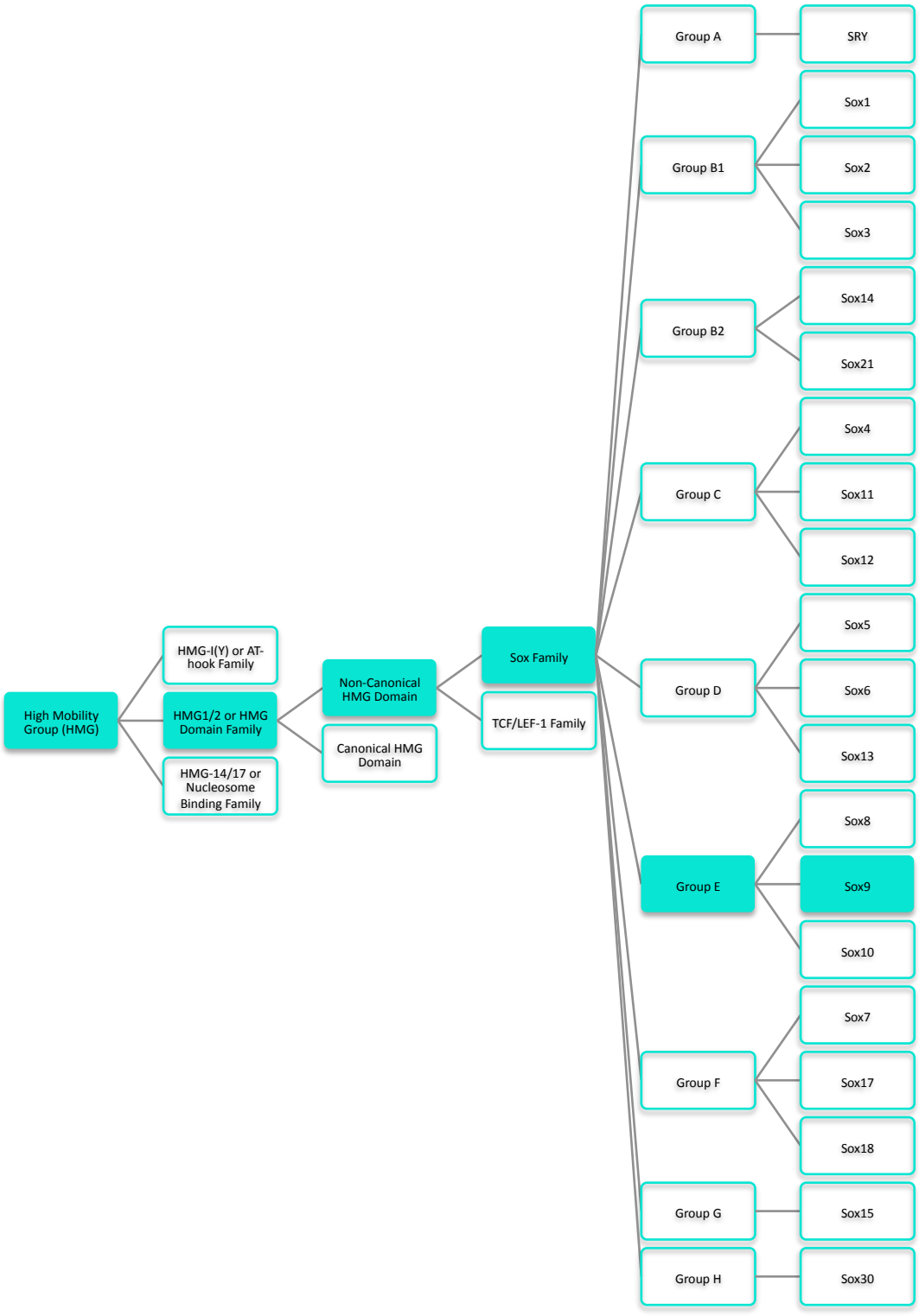


Figure 1: Schematic representation of the High Mobility Group (HMG) family. Sox9 is a transcription factor that is part of the Group E family of Sox proteins that binds DNA via a non-canonical HMG domain. (*Recreated from* Shokofeh Shahangian)

vertebrate developmental processes. SOX; expressed in Homo sapiens, or Sox; expressed in lower organisms, are a family of DNA-binding proteins. Sox proteins bind to ATTGTT sequence motifs via their HMG domain (Kamachi and Kondoh, 2013). It is the binding of the HMG domain to the minor groove of DNA that elicits bending in the DNA towards its major groove (Kamachi and Kondoh, 2013). The HMG domain in mice, adapted from Kamachi and Kondoh (2013), can be illustrated as a phylogenic tree **(Figure 2A)** defining classifications of SOX proteins from groups A to H as a result of genome duplication (Kamachi and Kondoh, 2013). Such groups differ in the amino acid sequence of the HMG domain **(Figure 2B),** however within a single group conservation of amino acid sequence homology and protein domain organization is maintained by sharing greater than 80% amino acid identity (Sarkar and Hochedlinger, 2013; Wegner, 1999). Structures outside of the HMG box diverge between mammalian species while the HMG in itself remains highly conserved (Kamachi and Kondoh, 2013). Redundancy within the Sox gene groups enables alterations in the amino acid sequences and changes to their functional role over evolutionary time (Kamachi and Kondoh, 2013). Sox family of proteins require binding of a transcription factor partner **(Figure 3A)** to an adjacent site on the DNA as a means to induce transcriptional activation or repression and to establish Sox-DNA complexes, increasing stability and regulating transcriptional activity (Kamachi and Kondoh, 2013).

Although Sox Group A is characteristic of mammalian Y-chromosomes and contains only Sry, groups B to F are distinguishable in invertebrates **(Figure 2C)** (Kamachi and Kondoh, 2013). Group B is sub-divided into B1 and B2 that differ in neural and sensory tissue targets, specification of the vertebrate brain, and are involved in

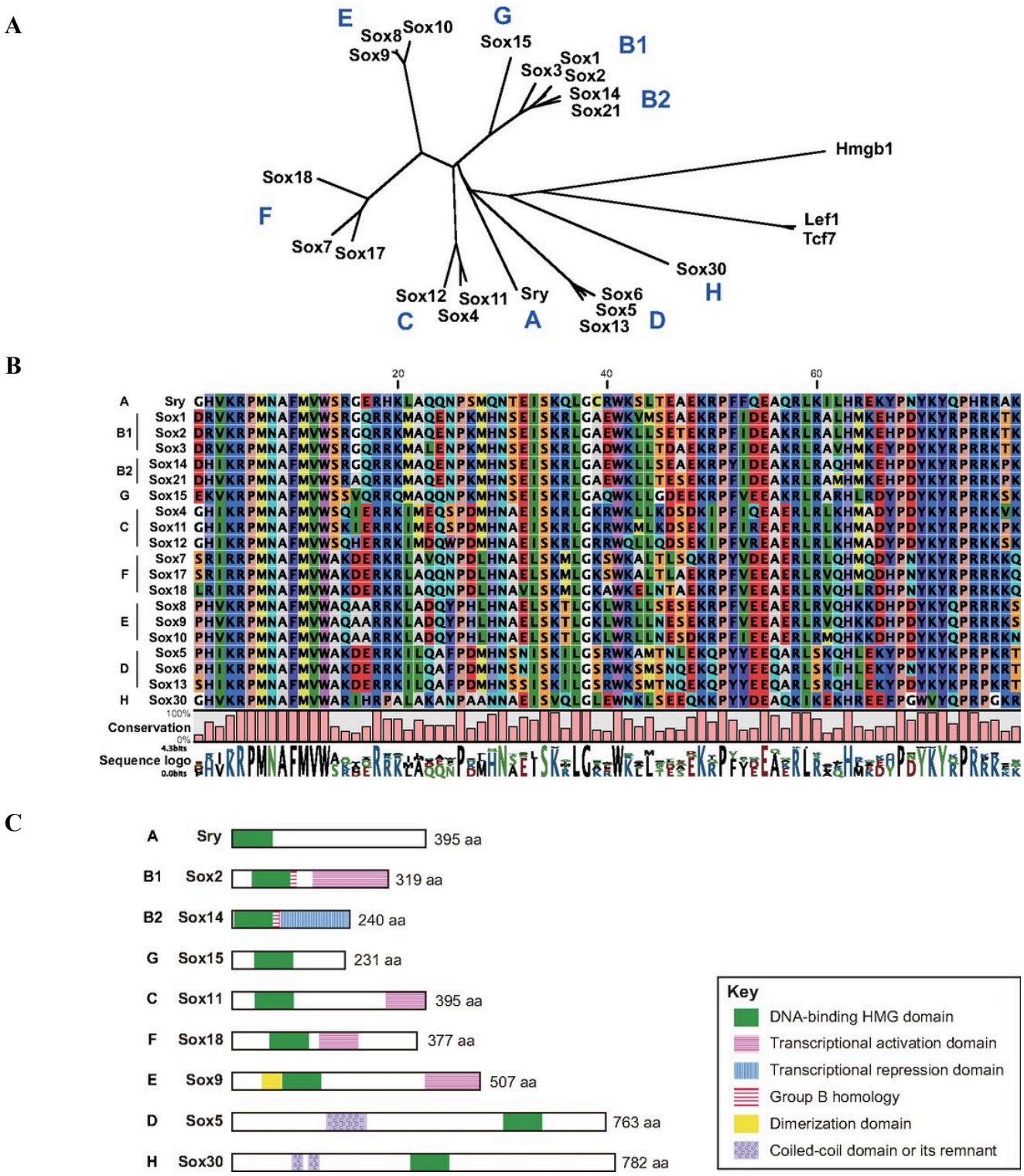


Figure 2: (**A**) Phylogenic tree of the high mobility group (HMG) domains of the mouse including the members within multiple evolutionary Sox groups. (**B**) Alignment of the Sox HMG domains. The sequences of 20 mouse Sox proteins arranged by the relationships of their sequence containing 79 amino acids. Conservation of amino acid sequences are defined by their logos. (**C**) The diversity of mouse Sox proteins organized by their domain structures demonstrating the various Sox groups. The basic structures remain conserved amongst members within each group. (*Adapted from* Kamachi, Y., and Kondoh, H. (2013). Sox proteins: regulators of cell fate specification and differentiation. Development 140, 4129–4144.)

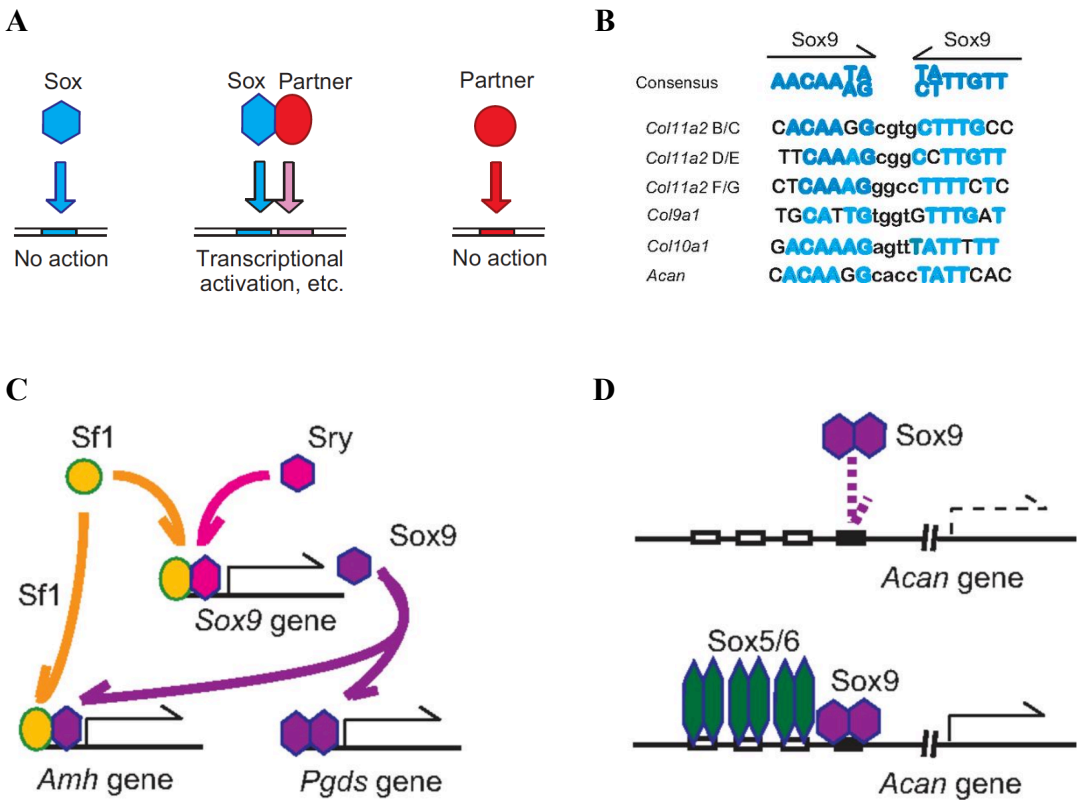


Figure 3: (**A**) Representation of Sox-partner complexes demonstrating DNA-binding and subsequent function. (**B**) Sox9 homodimer-binding sequences that are demonstrated to bind to DNA enhancers of collagen and related genes. Sox9 homodimer-binding sequences differ from the Sox consensus sequence at both sites. (**C**) Sox9-partner complexes that mediate developmental processes. Specifically, Sox9 induction is activated via Sry-Sf1 during male gonad development. (**D**) Sox9 dimerization with an identical Sox9 protein to employ a homo-dimer and initiate the dimerization of the SoxD group proteins; Sox5/Sox6, to bind to a nearby DNA site during chondrocyte development. Acan, aggrecan; Col9a1, collagen, type IX, α1; Col10a1, collagen, type X, α1; Col11a2, collagen, type XI, α2; Dppa4, developmental pluripotency associated 4; Fbxo15, F-box protein 15; Fgf4, fibroblast growth factor 4; Pou5f1, POU domain, class 5, transcription factor 1; Utf1, undifferentiated embryonic cell transcription factor 1; Amh, anti-Mullerian hormone; Pgds, prostaglandin D synthase; Sf1, steroidogenic factor 1; Sox, SRY-related HMG-box. (*Adapted from* Kamachi, Y., and Kondoh, H. (2013). Sox proteins: regulators of cell fate specification and differentiation. Development 140, 4129–4144.)

neural plate development (Okuda et al., 2006; Uchikawa et al., 2011). B2 also differs in that it contains a transcriptional repression domain unlike B1 that exhibits an activation domain in the C-terminal (Kamachi and Kondoh, 2013). Sox21 of the Sox B2 Group represses expression to promote neurogenesis (Matsuda et al., 2012). Sox Group C is involved in governing reproduction and sexual development and along with Sox Group E and F are similar in organization to Group B1 (Kamachi and Kondoh, 2013). Sox Group D contains a long N-terminal sequence with a Leucine-zipper motif containing repetitive Leucine residues (Landschulz et al., 1988). It is followed by a coiled-coil domain as a result of a glutamine-rich region as a means to dimerize with other Sox D proteins mediating both homo or hetero dimerization (Kamachi and Kondoh, 2013). This differs from Sox Group E that contains a unique defined self-dimerization domain important during cartilage formation and sex determination in early embryonic development (Kamachi and Kondoh, 2013). Group F has been associated with vascular and hematopoietic development and contains an activation domain significantly closer than Group C and E to the N-terminus (Chung et al., 2011). Both Sox Group G and Sox Group H are specific to mammals with Group G possessing the relatively smallest construct and Group H (Sox 30) sustaining the largest Sox construct (Meeson et al., 2007; Osaki et al., 1999).

1.3 The Sox Group E Family of SOX Proteins: The SOX9 Transcription Factor

The Sox Group E family of human SOX proteins including SOX8, SOX9, and SOX10 have been shown to bind to DNA as both a monomer and dimer (Bernard et al., 2003). These proteins are identified by their capability to cooperatively dimerize in a fashion that is DNA-dependent on a consensus site; (A/T)(A/T)CAA(A/T)G (Harley et

al., 1994). Such dimerization is reliant on approximately 40 amino acids N-terminal to the HMG domain. This is in stark contrast to SOX5, SOX6, and SOX11 that are part of the Group D class of proteins whereby dimerization dependency relies on a Leucine zipper motif also in a DNA-dependent manner (Takamatsu et al., 1995). More importantly, the differentiation between monomeric or dimeric binding to a variety of promoters can specialize the function of SOX9 during development (Bernard et al., 2003).

SOX9 is an approximately 60kDa transcription factor protein containing 507 amino acids. Its structure **(Figure 8A)** contains an N-terminal dimerization domain followed by an HMG domain flanked by two nuclear localization signals (NLS). The C-terminal is comprised of a PQA rich region followed by a transcriptional activation domain. The crystal structure of the HMG domain of SOX9 complexed with DNA (SOX9-DNA) at a single DNA site demonstrates the affinity for SOX9 to bind and bend DNA **(Figure 4).** SOX9 is a transcription factor that recognizes the sequence CCTTGAG. It can form either a monomeric or dimeric structure to determine its subsequent function by binding to specific partners including an identical SOX9 protein forming a homodimer, or an alternative binding partner forming a heterodimer, and activating targeted genes **(Figure 3C & 3D)** (Kamachi and Kondoh, 2013). The alterations between monomeric and dimeric binding states and alternative SOX9 binding partners assume specific confirmations depending on the complex formed that elicit specific transcriptional regulation (Kamachi and Kondoh, 2013). As a result, the binding sequences differ from the consensus-binding site between a single factor and SOX9-partner complexes **(Figure 3B)**.

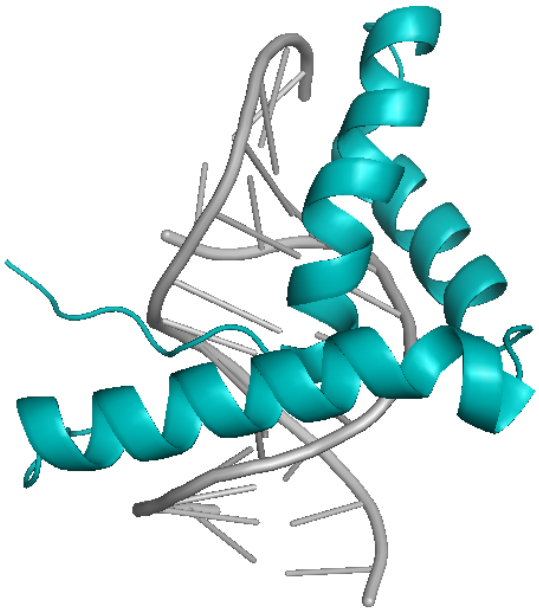


Figure 4: Crystal structure of the HMG domain containing three alpha helices of the transcription factor SOX9 bound to DNA (SOX9-DNA) of Homo sapiens at 2.77 A resolution (*Recreated from* PDB ID: **4EUW**).

1.4 A Monomeric SOX9: Sex Determination

The role of SOX9 in sex determination was first revealed post-discovery of the sex-determining region Y (SRY) protein responsible for the initiation of male sex determination in humans. SOX9 was identified via genetic analysis of patients with human Campomelic dysplasia (CD) whereby XY individuals with heterozygous SOX9 mutations demonstrated male-to-female sex reversals (Kamachi and Kondoh, 2013). A knockout of SOX9 leads to male gonad defects and XY gonad dysgenesis (Sekido and Lovell-Badge, 2008; Wilhelm et al., 2007). In mice, Sry is transiently expressed in supporting somatic cells of the gonads that have yet to be determined between embryonic day 10.5 to 12.5 (Kamachi and Kondoh, 2013; Sekido and Lovell-Badge, 2008). During this period of development, Sry expression will trigger the differentiation of these somatic cells into Sertoli cells **(Figure 5A)** (Kamachi and Kondoh, 2013). Using the testis-specific enhancer (Tes) of Sox9, synergistically Sox9 and Steroidogenic factor 1 (Sf1) are directly activated as a result of *Sry* (Kamachi and Kondoh, 2013). Once expression of Sox9 is activated passed a critical threshold level, the expression levels are maintained in the Sertoli cells with a positive-feedback loop (Kamachi and Kondoh, 2013). The auto-regulated positive feedback mechanism of Sox9 is maintained via Sox9 activating the Tes enhancer in a cooperative manner with Sf1 (Eggers et al., 2014; Kamachi and Kondoh, 2013). Sox9 as a monomer can then activate other genes responsible for testis development including prostaglandin D synthase (Pgds) (Kamachi and Kondoh, 2013).

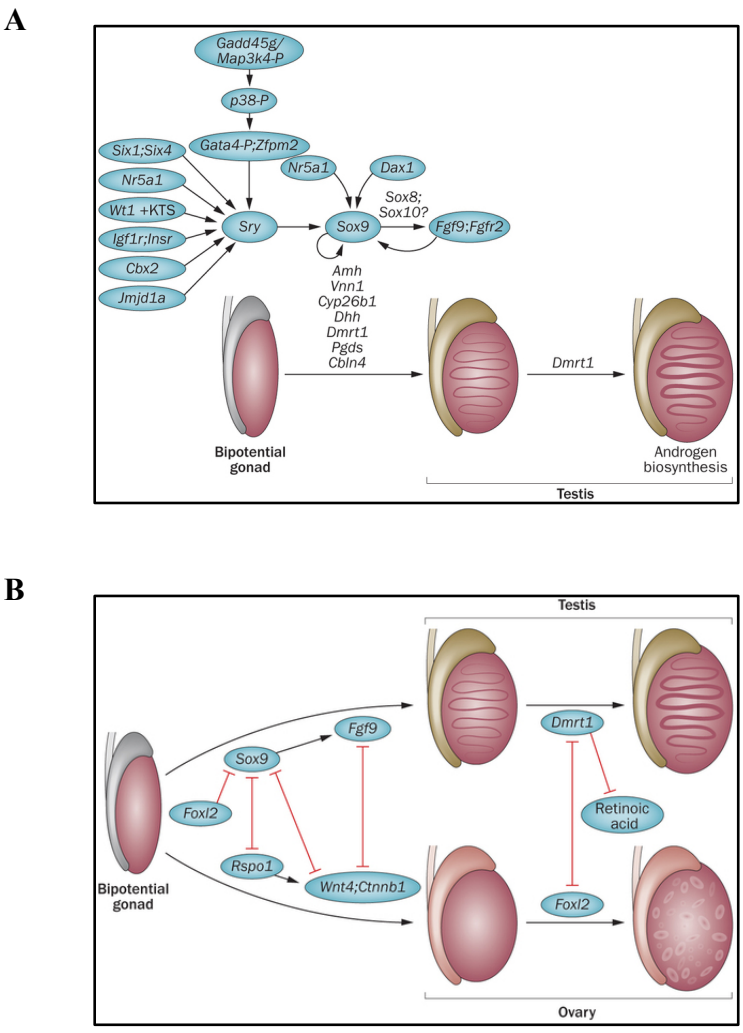


Figure 5: (**A**) The pathway regulating testes development and differentiation in part requires that Sry is transiently expressed in the bipotential gonad with the highest level of expression during development at embryonic day 12.5. This in turn increases the expression of Sox9 in the developing testes. Sox9 maintains a positive feedback loop by way of auto-regulation and under the influence of co-factors *Sry, Nr5a1, Dax1, Fgf9*, and many other genes. (**B**) The two pathways leading to either testes or ovary development is partially dependant on the regulation of Sox9 along with an array of genes and proteins. *Foxl2, Rspo1, Wnt4,* and *Ctnnb1*(β-catenin) genes contribute to inhibition of Sox9 and Fgf1 that in turn inhibit the development of ovaries by then suppressing the expression of *Rspo1* and *Wnt4-Ctnnb1*. This process is maintained by other factors throughout the adult life. (*Adapted from* Eggers, S., Ohnesorg, T., and Sinclair, A. (2014). Genetic regulation of mammalian gonad development. Nat Rev Endocrinol *10, 673–683.)*

SOX9 plays a pivotal role in male sexual development and male sexual differentiation due to the regulation of the anti-Müllerian hormone (AMH) activated by the cooperative action of SOX9 and Sf1 (Kamachi and Kondoh, 2013). Cooperative binding occurs when the binding at one site increases the rate at which binding occurs at a second site between identical or non-identical proteins. AMH is strictly regulated in male-sex determination during embryonic development via reversion of the Müllerian duct. Although lost in males, this duct is associated with the uterine tubes, uterus, cervix, and vagina in females (De Santa Barbara et al., 1998). It is Sf1, located in the AMH proximal promoter that is essential for AMH gene activation (De Santa Barbara et al., 1998). SOX9 can produce AMH in Sertoli cells to inhibit the development of the female reproductive system **(Figure 5B)** (De Santa Barbara et al., 1998). Rather, it promotes the development of male sexual organs (De Santa Barbara et al., 1998). This process is initiated when the Testis determining factor encoded by SRY activates SOX9 by binding to the enhancer sequence upstream of the gene (De Santa Barbara et al., 1998). SOX9 activates a glia-activating factor identified as Fibroblast growth factor 9 (Fgf9) to enable development of the testis cord and increase Sertoli cell count (Kamachi and Kondoh, 2013). In addition, SOX8 is expressed when SOX9 is up-regulated and plays a role in the maintenance of testis function during later stages of development (Kamachi and Kondoh, 2013).

1.5 A Dimeric SOX9: Chondrogenesis

The role of SOX9 is not limited to sex determination but rather plays a significant role in early development, programmed cell death, as well as maintenance of the cell cycle. Its function in early cartilage and skeletal development was first reported as a result of heterozygous human SOX9 mutations as a cause for the skeletal disorder; CD (Coustry et al., 2010; Kamachi and Kondoh, 2013). Such heterozygous mice suffer perinatal mortality due to abnormal skeletal development (Bi et al., 2001; Kamachi and Kondoh, 2013). Specifically, during chondrogenesis; the process by which cartilage is formed, SOX9 expression levels are high and maintained throughout the chondrocyte lineage (Kamachi and Kondoh, 2013; Ng et al., 1997; Zuscik et al., 2008). SOX9 is critical for multi-step chondrocyte differentiation, function, and has been deemed as a master regulator in controlling the secretion of cartilage matrix proteins (Hino *et al,* 2014). In addition, the expression of SOX9 is present and essential throughout multiple stages of cartilage formation **(Figure 6A)** including the commitment process from an undifferentiated mesenchymal cells to osteochondroprogenitor cells, and further downstream to condensed mesenchymal cells, differentiated chondrocytes, and proliferating chondrocytes (Akiyama et al., 2002) (Zuscik et al., 2008). This eventually leads to the formation of hypertrophic chondrocytes and finally calcified cartilage (Akiyama et al., 2002). SOX9 is present within distinct cell types, time points, and located in the radial cellular zone during the process of chondrogenesis **(Figure 6B & 6C)**. However, SOX9 inhibits the transition from proliferating to hypertrophy cells (Akiyama et al., 2002). Sox9 knockdowns lead to chondrogenesis defects including congenital diseases like CD as a result of either activation or repression of targeted genes such as *Col10a1* (Dy et al., 2012; Leung et al., 2011).

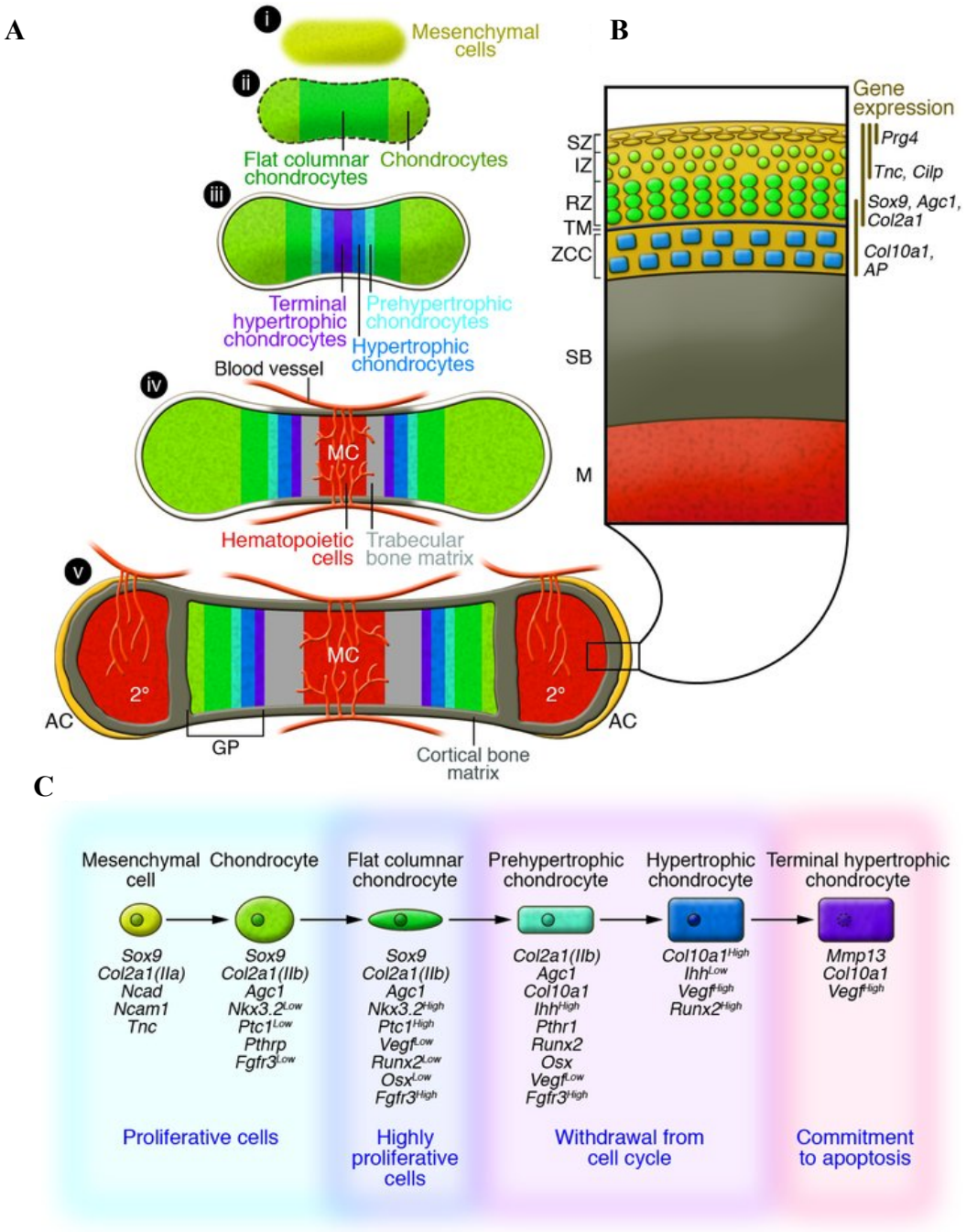


Figure 6: (**A**) Model for endochondral bone development. SOX9 is expressed and functions as a homo-dimer during multiple stages of the chondrocyte differentiation pathway. This includes the participation of an array of co-factors. (**B**) Distinct cellular zones whereby Sox9 is present in the radial zone (RZ). (**C**) Model for the process of chondrogenesis and the differentiation of chondrocytes. (*Adapted from* Zuscik, M.J., Hilton, M.J., Zhang, X., Chen, D., and O'Keefe, R.J. (2008). Regulation of chondrogenesis and chondrocyte differentiation by stress. J. Clin. Invest. *118, 429–438.)*

The DNA binding site sequences for Sox9-partner complexes differ from the consensus sequences for single factors binding to DNA due to a specific confirmation by the complex to elicit transcriptional regulation. During the process of chondrogenesis, the Sox9 homodimer binding sequences differ from the Sox9 consensus sequence at both sites **(Figure 3B)** (Kamachi and Kondoh, 2013). Its expression profile is similar to that of type II collagen gene *Col2a1* that encodes the major cartilage matrix protein (Lefebvre et al., 1997). The binding of SOX9 is a potent activator to this particular intronic enhancer and is essential for the activation of the *Col2a1* gene (Kamachi and Kondoh, 2013; Lefebvre et al., 1997). More importantly, SOX9 directly regulates an array of collagen genes (Lefebvre et al., 1997). Furthermore, Sox9 and Col2a1 are expressed in cells during chondrogenesis (Zhao et al., 1997). Other SOX proteins including SOX5 and SOX6 of the Sox D family also mediate the entire process (Akiyama et al., 2002). SOX9 is essential for the expression of SOX5 and SOX6 (Akiyama et al., 2002). Strong activation of the minimal enhancer *Col2a1* only occurs when SOX9 and proteins SOX5 and SOX6 bind to several sites in the enhancer (Han and Lefebvre, 2008; Kamachi and Kondoh, 2013). This has been defined in literature as a "Sox trio" that corresponds to a SOX9 homodimer with multiple SOX5/6 dimers binding to an adjacent DNA site (Kamachi and Kondoh, 2013). In the presence of normal SOX9 levels, a double knockout of SOX5/6 leads to a lack of chondrogenesis, while an individual knockout of either SOX5 or SOX6 leads to mild skeletal abnormalities (Kamachi and Kondoh, 2013; Smits et al., 2001). This suggests that there lies some form of cooperation between SoxD-SoxE families that enables an important molecular mechanism regulating chondrogenesis (Kamachi and Kondoh, 2013). More importantly, the SOX9 dimer-binding sequences

present in enhancers of genes that regulate chondrocyte development differ from SOX consensus binding sequences in both sites (Kamachi and Kondoh, 2013).

1.6 The Importance of the Dimerization Domain

The dimerization domain is a region of amino acids defined by its affinity to homodimerize or heterodimerize depending on the nature of the subunits to ultimately control gene expression and cell function (Feuerstein et al., 1994). Studies have demonstrated that patients who harbor mutations including substitutions and in-frame deletions preceding the HMG domain, i.e. the dimerization region, instigate a direct onset of semi lethal skeletal anomalies characterized in CD, male-female sex reversal, and developmental defects in cartilage formation (Lefebvre et al., 1997; Sock et al., 2003). One of the first mutations outside of the HMG domain associated with XY male patients that were not sex-reversed but diagnosed with CD was the A76E mutation. It is the structural conformational change due to dimerization that enables activation of promoters through dimeric binding sites, however this is lost in two mutant SOX9 patient studies (Sock, 2003). During cartilage formation cartilage specific promoter sequences like pro-α1 (II) collagen gene *Col2a1* contain HMG domain binding sites that are inverted (Lefebvre et al., 1997). These sites are spaced a couple of base pairs apart and maintain an affinity for two closely adjacent SOX9 proteins (Lefebvre et al., 1997).

The dimerization domain of SOX9 is suggested to play a critical role in modifying the Col2a1 chromatin to initiate transcription (Lefebvre et al., 1997). DNA binding differs between wildtype SOX9 and a SOX9 dimerization mutant such that the wild-type SOX9 dimeric complex increases at a slower rate with a sigmoidal progression (Lefebvre et al., 1997). This differs from a mutated dimerization domain that increases

linearly (Lefebvre et al., 1997). Stability of DNA-SOX9 complexes is not affected by the presence of an active dimerization domain and transcription is activated however, chromatin remodeling is hindered in the dimerization domain mutant (Lefebvre et al., 1997). Ultimately, the dimerization domain enables a stronger activation *Col2a1* promoter/enhancer-luciferase than the mutant SOX9 but interactions with the chromatin are unaffected (Lefebvre et al., 1997).

1.7 HMG Proteins Bind and Bend DNA

DNA bending by the HMG-box of HMG proteins like SOX9 are established on the basis of HMG-box-DNA structures. The important characteristics of the HMG-box-DNA complex is that bulky hydrophobic amino acids present in the HMG-box intercalates between successive base-pairs within the minor groove of DNA (Stros, 2010). Together with partial unwinding and winding of the minor groove elicits bending toward the major groove of DNA (Stros, 2010). The contact with the minor groove of the DNA double helix forces the DNA to bend to and L-shape angle varying from 30° to 110° **(Figure 7)** (Lefebvre et al., 2007). The inserted hydrophobic residues of the HMG-box flanked by basic residues are conserved and bind to the phosphodiester bond of DNA to create stability in the protein-DNA complex (Stros, 2010). Within the HMGB1/2 there contains two boxes defined as A and B. HMG-box B has a greater responsibility for bending DNA than A, as A lacks a primary intercalating residue (Stros, 2010). The residues generally responsible are Phenylalanine and Isoleucine. DNA bending by non-sequence-specific proteins differ from sequence-specific proteins such that non-specific proteins are associated with multiple hydrophobic residues at two sites in DNA unlike sequence-specific proteins whereby only one intercalation residue is observed (Stros, 2010).

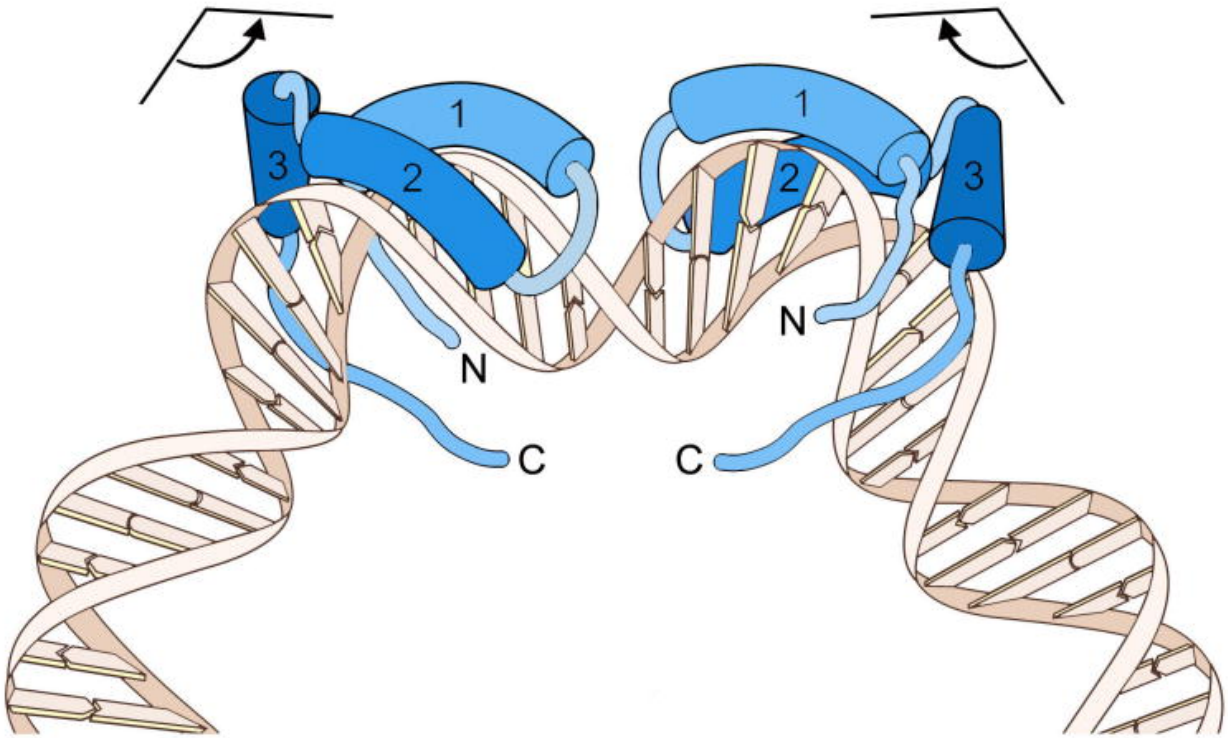


Figure 7: The SOX HMG Box domain bound to a DNA double helix and inducing a L-shaped bend in the DNA that ranges from 30° to 110°. (*Adapted from* Lefebvre, V., Dumitriu, B., Penzo-Méndez, A., Han, Y., & Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. The International Journal of Biochemistry & Cell Biology, 39(12), 2195–2214. http://doi.org/10.1016/j.biocel.2007.05.019)

For example, NHP6A is a chromatin-associated protein in *Saccharomyces cerevisiae* from the HMG1/2 family of non-specific DNA binding proteins (Allain et al., 1999). The NMR-based model of the NHP6A-DNA complex structure illustrates the HMG domain of the protein positioned onto the minor groove of the DNA by which the distortion site contains two kinks in the DNA where Methionine and Phenylalanine are inserted and deemed critical for bending (Allain et al., 1999). A similar crystal structure is confirmed with a non-histone HMG chromosomal protein from *Drosophila melanogaster* whereby HMG-D binds a DNA duplex in a non-sequence-specific manner (Murphy et al., 1999).

1.8 The Functional Significance of DNA Bending

Relevant to functionality, the importance of DNA bending is observed by virtue of *in vitro* assays (Stros, 2010). Mutations of these critical intercalating residues demonstrate a decrease in mobility in these residues and in HMGB1 living cells preventing binding to chromatin (Stros, 2010). Furthermore, these residues are required for stimulation of progesterone receptor transcriptional activity as well as the expression from the human topoisomerase IIα gene promoter *in vivo* due to the impairment of DNA bending (Stros, 2010). It enables looping as well as flexibility in DNA by transiently even binding to DNA (Stros, 2010). This has been confirmed by electron microscopy and dual laser optical tweezer experiments (Stros, 2010). Aside from specific binding, random binding to DNA for a short period of time enables the DNA to bend with more fluidity and local flexibility at specific binding sites. Such binding could contribute to an increase in cooperativity. Combined, bending and looping can provide a mechanism for proteins to promote activity of gene promoters by the enhancement of binding of transcription

factors as well as spatially guiding regulatory sequences into close proximity (Stros, 2010).

Alternatively, the anticancer activity of *cis*-diamminedichloroplatinum(II) (cisplatin) is a product of its ability to damage DNA by bending and unwinding the duplex (Ohndorf et al., 1999). The structural change exposes sites preferable for HMG domain proteins to bind to the cisplatin-modified DNA to mediate antitumor properties of the drug (Ohndorf et al., 1999). X-ray studies revel that the HMG-domain of the protein binds to the widened minor groove of the DNA and a specific Phenylalanine residue at position 37 intercalates into a hydrophobic notch (Ohndorf et al., 1999). In addition, binding is significantly reduced in a substitution mutation with Alanine at position 37 (Ohndorf et al., 1999). The HMG-box domain in human SRY (hSRY) male-sex determining factor, binds DNA in a sequence-specific manner of the minor groove of DNA initiating a substantial bend (Murphy et al., 2001). Most of the point mutations that result in 46XY sex reversals are within the HMG domain (Murphy et al., 2001).The clinical mutation M64I corresponds to M9I in the HMG of hSRY and reduces the extent of bending of DNA from 54° in the wildtype to 41° in the mutant (Murphy et al., 2001). This is due to changes in the roll and tilt angles in the base pair interactions with residue 9 and only partial intercalation of Isoleucine 13 as per the 3D model studied via NMR spectroscopy (Murphy et al., 2001). The bend in the wildtype is a result of steric repulsion in the sugar of the second adenine combined with the bulkiness of Methionine 9 (Murphy et al., 2001). The position of Methionine 9 is fixed by a hydrogen bond with the guanidino group of Arg17 (Murphy et al., 2001). When Methionine is replaced by Isoleucine, the hydrogen bond is lost due to the presence of a less bulky methyl group in

isoleucine and braces the sugar moieties thereby decreasing the roll and tilt angle (Murphy et al., 2001).

1.9 SOX Proteins Bend DNA to Elicit Function

The activity of SOX proteins via DNA bending is an essential factor for the regulation of gene expression. Studies have well characterized the ability of SOX proteins to first bind and subsequently elicit a bend in the DNA changing the structure of the DNA and the protein complex. Mouse Sox2 participates in the transactivation of the fibroblast growth factor (Fgf4) gene in the inner cell mass of the blastocyst (Scaffidi and Bianchi, 2001). Mutations to either the HMG box of Sox2 or base-mutations to the Fgf4 enhancer resulted in less bending of the DNA from 80° to 42° or 0° (Scaffidi and Bianchi, 2001). Reduced bending resulted in a decrease in transcription activity for all mutants except for one Sox2 HMG mutant whereby reduced bending resulted in a more powerful transactivation in comparison to the wildtype protein (Scaffidi and Bianchi, 2001).

The dynamic properties of protein-DNA binding via the HMG box are in part due to the nature of the sequence-specific (SS) or non-sequence-specific (NSS) HMG box proteins (Dragan et al., 2004). DNA recognition sequences differ in stability. SS proteins are identified as partially unfolded at room temperature however when in complex with DNA stabilizes and results in cooperation (Dragan et al., 2004). The difference between SS and NSS are that SS are formed with an enthalpy close to zero and a negative heat capacity unlike NSS that forms with a high positive enthalpy and positive heat capacity(Dragan et al., 2004). The complex formed between the SS HMG box-DNA complexes are a more tightly compact interface via extensive van der Waals contact

points between apolar groups than NSS HMG box-DNA complexes (Dragan et al., 2004). The bending of DNA via the NSS-DNA interaction is highly dependant on the electrostatic component of the binding energy (Dragan et al., 2004). Unlike DNA bending by SS-DNA complexes that are highly driven by the non-electrostatic component (Dragan et al., 2004).

The existence of the HMG box enables Sox5 to induce a large bend in double helix DNA to regulate the function of gene expression during the post-meiotic phase of spermatogenesis (Connor et al., 1994). DNA bending by Sox5 is demonstrated with circular permutation assays (Connor et al., 1994). Annealed synthetic nucleotides containing the Sox5 binding site were constructed via an electromobility shifts assay (EMSA) and the center of DNA bending to the Sox5 binding site was mapped to visualize the quantification of R_{bound}/R_{free} against flexure displacement (Connor et al., 1994). Flexure displacement was defined as the distance of the centre of Sox5 *in vitro* binding motif from the 5' end of the probe divided by the total length of the probe (Connor et al., 1994). As expected, the assay revealed that the minimum centre of bending was located to the AACAAT motif (Connor et al., 1994). The bending of DNA via Sox proteins may play a critical role in the assembly of transcriptional enhanceosomes (Lefebvre et al., 2007). These are defined as the higher-order protein complexes assembled at the enhancer that regulate target gene expression (Lefebvre et al., 2007).

Circular permutation assays using circular permutated probes have been at the forefront of demonstrating that the Sox Group E family of proteins bind and bend DNA. The *Col2a1* enhancer has demonstrated to bend by the specific binding of SOX9 by mapping the flexure displacements (Lefebvre et al., 1997). There is a correlation between the expression of *Col2a1* and SOX9 in chondrogenic cells that is important for cartilage formation during embryonic development (Lefebvre et al., 1997). Yeast-two-hybrid screens have demonstrated that Sox8 and Sox10 proteins of the Group E family of Sox proteins posses DNA-binding and DNA-bending for the architectural function of Sox proteins on their target gene promoter (Wissmüller et al., 2006). The stabilization of the Sox protein when bound to DNA is a product of the Sox protein initiating varying degrees of bending in the DNA as a result of a flexible angular surface of the L-shaped HMG Box domain that becomes fixed upon the binding to DNA (Lefebvre et al., 2007). The dual nature of the protein-DNA complex is complimentary and important as SOX instructs the DNA how to bend, while the DNA instructs the SOX protein on how to complete its tertiary fold (Lefebvre et al., 2007). As a result, there is an enhanced stability formed by the complex and novel functionality (Lefebvre et al., 2007). Cooperative binding of Sox10 to a dimeric binding site C/C′ is dependent on 40 amino acids preceding the HMG (Peirano and Wegner, 2000). Sox10 wildtype and spacing mutants were studied to determine the effect on cooperative binding. Spacing variants of sequences inserted between site C and site C′ demonstrated that increasing the sequence length increases the bending angle by up to 20° (Schlierf et al., 2002). Consequently, an increase in sequence length decreases cooperativity and quantity of bound dimer (Schlierf et al., 2002). All three SOX E proteins have overlapping expression patterns during

embryogenesis that could confer to functional redundancy but mainly control seemingly unrelated developmental processes in different tissues (Wissmüller et al., 2006).

1.10 Clinical Role of SOX9

SOX9 expression is a contributing factor in several well-studied developmental malformations and diseases **(Table 1)**. The absence of SOX9 has been associated with CD, a form of skeletal dysplasia. CD has distinctive dysmorphological characteristics including cleft palate, bowing of long bones, short stature, instability of the cervical spine leading to cord compression, scoliosis, and club feet (Mansour et al., 1995). Infants are known to die as a neonate (Mansour et al., 1995). With only a single functional copy of the *SOX9* gene defective cartilage primordia and premature skeletal mineralization is prevalent (Bi et al., 2001). The *SOX9* gene is located at 17q24 on chromosome 17 consisting of approximately 5400 bases in size. SOX9 has also been identified to play a role in sex reversal in humans during early embryonic development (Foster, 1996). Studies confirm that such reversal is a result of mutations to the SRY gene in a XY individual affecting SOX9 expression levels and results in a sex-reversed female phenotype (Foster, 1996). Consequently, mutation studies have confirmed de novo mutations in the *SOX9* gene in patients with both autosomal sex reversal and CD (Foster, 1996; Kwok et al., 1995). SOX9 participates in a multitude of developmental processes including the growth plate of endochondral bones, hair bulging via hair follicle stem cells and early skin morphogenesis required for formation of hair follicles and sebaceous glands (Fantauzzo

Table 1: The array of tissues and purposes of activation of SOX9 with associated binding partners as well as monomeric or dimeric binding sites of SOX9 in order to activate known targeted genes. SOX9 knockouts lead to changes in phenotypes and associated congenital human diseases during embryonic development.

Tissues/ Purpose of Activation	Known Target Genes	Available SOX9 Binding Sites	Known Partner Factors	Phenotypes of Knockout	Associated Congenital Human Disease
Chondrocytes/ Chondrogenesis	*CD-Rap*	Monomer			
Chondrocytes/ Chondrogenesis	*Agc-1*	Monomer			
Chondrocytes/ Chondrogenesis	*Col2a1*	Monomer/Dimer	Sox5/6		
Chondrocytes/ Chondrogenesis	*Col2a2*	Dimer	Sox5/6		
Chondrocytes/ Chondrogenesis	*Col9a2*	Dimer	Sox5/6		
Chondrocytes/ Chondrogenesis	*Col10a1* (activation)		Mef2c	Chondrogenesis defects	Campomelic dysplasia
Chondrocytes/ Chondrogenesis	*Col10a1* (repression)		Gli2/3	Chondrogenesis defects	Campomelic dysplasia
Chondrocytes/ Chondrogenesis	*Col11a2*	Dimer	Sox5/6		
Chondrocytes/ Chondrogenesis	*Acan*	Dimer	Sox9, Sox5/6	Chondrogenesis defects	Campomelic dysplasia
Chondrocytes/ Chondrogenesis	*Matrilin-1*	Dimer	Sox5/6		
Sex-determination	*SF-1*	Monomer			
Sex-determination	*AMH*	Monomer			
Gonad	*Sox9, Amh, Pgds*	Monomer/Dimer	Sf1, Sox9	Male gonad defect	XY gonad dysgenesis
Hair Bulge				Alopecia	Hypertrichosis
Otic Placode				Otic placode invagination defect	
Neural Stem Cells				Failure to maintain neural stem	
Glial Progenitor Cells			Nf1a	Imparied gliogenesis	
Oligodendrocyte Progenitors	Supressor of fused myelin basic protein			Loss of oligodendrocytes	
Neural Crest	*Snail2*		Snail2	Loss of neural crest delamination	
Neural Crest	*Sox10*	Dimer	Sox9, Ets1, Myb		
Retinal Progenitor Cells				Loss of stem cell differentiation in Müller glial cells	
Pancreatic Progenitors				Loss of exocrine pancreatic duct cells	
Mammary Stem Cells		Cooperativity with Slug		Loss of stem cell maintenance	
Intestinal Stem/ Progenitor Cells				Depletes stem cells and Paneth cells	
Liver Duct Cells				Lineage labels hepatocytes post injury	

et al., 2012; Nowak et al., 2008; Vidal et al., 2005; 2001). Knockouts of SOX9 lead to alopecia or hair loss from the scalp and hypertrichosis; a SOX9 dysregulation (Nowak et al., 2008).

SOX9 is a significant player in the regulation of the state of stem cells. It is identified as a key contributor to liver regeneration, fibrosis, formation of liver duct cells, and tracing of hepatocytes post injury (Ein et al., 2014; Furuyama et al., 2011). This diverse protein contributes to distal tip lung cell formation (Rawlins, 2011). Pancreatic progenitors lead to the development of exocrine pancreatic duct cells with Sox9 genetic deletions conferring to a loss of pancreatic progenitors, intestinal stem, and Paneth cells (Furuyama et al., 2011; Kopp et al., 2011; Sato et al., 2011; Seymour et al., 2007). SOX9 expression is associated with the development of retinal progenitor cells (RPCs) with genetic deletions facilitating the loss of the potential for stem cell differentiation in Müller glial cells (Poché et al., 2008). Mammary stem cell (MaSC) activity is blocked upon the inhibition of either Sox9 or Slug in mammary epithelial cells (Guo et al., 2012). Conversely, transient co-expression of either from an exogenous source enables the conversion of differentiated luminal cells to MaSCs such that Sox9 and Slug cooperatively determine the state of MaSCs (Guo et al., 2012).

The expression of SOX9 is essential during neuronal development. Normal otic placode development can become impacted by otic placode invagination defects when Sox9 is knocked out (Barrionuevo et al., 2008). Such tissue specific knockdowns also lead to failure of neural stem maintenance from the development of neural stem cells including the determination of glial fate in the developing spinal cord (Cheng et al., 2009; Scott et al., 2010; Stolt et al., 2003). This includes impaired gliogenesis during defective glial

progenitor development (Kang et al., 2012; Stolt et al., 2003). Sox9 is expressed in the developing CNS and maintains the multipotency of stem cells (Cheng et al., 2009) (Fantauzzo et al., 2012). From the Group E family, both Sox9 and Sox10 play a major role in the formation of oligodendrocyte progenitors (Pozniak et al., 2010; Stolt et al., 2003; 2002; 2006). Oligodendrocytes modulated by Sox D proteins are known to target the suppressor of the fused myelin basic protein however deficits in Sox9 lead to a loss of oligodendrocytes (Pozniak et al., 2010; Stolt et al., 2002; 2003; 2006). In the cranial neural crest, there is a cooperation between Sox9 with Ets1 and cMyb to activate the *Sox10* gene in a Sox9-dependant manner (Agarwal et al., 2011; Betancur et al., 2010). When phosphorylated, Sox9 also plays a critical role in neural crest formation however a deficiency or lack of Sox9 contributes to the loss of neural crest delamination via downstream regulation by bone morphogenetic protein (BMP) and canonical Wnt (Betancur et al., 2010; Cheung and Briscoe, 2003; Liu et al., 2013). This includes the cooperative action of Sox9, Snail2, and Protein Kinase A (PKA) (Sakai et al., 2006).

The role of SOX9 in cancer is extremely diverse including its up-regulation in tumors associated with pancreatic cancer and high expression in gastric cancers. SOX9 regulates miR-32 in osteosarcoma and can aid as a marker in its potential progression and prognosis (Camaj et al., 2014; Choi et al., 2014; Xu et al., 2014; Zhu et al., 2013). In esophageal cancer, the ability of stem cell properties to be acquired has demonstrated to be dependant on the expression of SOX9 in a YAP-1 driven manner (Song et al., 2014). SOX9 expression mediated by Rac1 has been demonstrated in inner meniscus cells (Kanazawa et al., 2014). More importantly, meniscus cells have been studied as a donor for tissue engineering (Nakata et al., 2001).

1.11 SOX9 Therapeutic Role

The expression of SOX9 is altered in tumorigenesis such as osteosarcoma. Upregulation of SOX9 was evident in 72% of aggressive human osteosarcoma tissues and expression further increases during advanced clinical stages indicating its fundamental role in the progression of bone cancer (Zhu et al., 2013). Such advanced stages include distant metastasis and a reduced response to chemotherapy (Zhu et al., 2013). This suggests that SOX9 can be used as a tool for early detection and prognosis of osteosarcoma and a target for optimal clinical treatment (Zhu et al., 2013). Small molecules could be developed to target the binding of SOX9 or compete for binding on DNA and prevent subsequent clinical stages of cancer progression. Similarly, adenovirus with down regulated Sox9 can be injected into tumors to prevent tumor growth. Likewise, meniscus cells have been described as a candidate for tissue-specific engineering (Nakata et al., 2001). Meniscus cells are critical for fibrocartilaginous tissue in the knee joint as well as other joint stability mechanisms such as load distribution (Adesida et al., 2007). When abolished, there is a higher incidence of osteoarthritis (Adesida et al., 2007). SOX9 expression has been associated with an increase in oxygen-tension that is correlated to the suppression of cartilage-like matrix formation (Adesida et al., 2007). The high incidence of SOX9 expression could be used as a marker for the onset of osteoarthritis as well as maintaining the viability of meniscus cells for tissue engineering as these cells could be cultured, propagated, and seeded onto a collagen scaffold (Nakata et al., 2001).

1.12 Thesis Overview and Objective

This study features a model for the interaction of the dimerization domain with the HMG domain of two identical SOX9 [71/184] constructs. The two wildtype constructs featured in this study are SOX9 [71/184] containing both the dimerization domain and HMG domain and SOX9 [101/184] that contains only the HMG domain **(Figure 8B).** This study utilizes two different DNA constructs; a single site S9WT and CC36; a 36 bp double site palindrome **(Figure 8C & 8D).** CC36 differs from a canonical sequence at the first position and is similar to the first C binding site in the tandem C/C' region of the P_o gene promoter that binds to SOX10 (Peirano and Wegner, 2000). SOX9 proteins bind to CC36 via a primary consensus and secondary non consensus site that are inverted palindromes comparable to the *Col2a1* gene. SOX9 will play a role in cooperative DNA binding via its dimerization domain.

Upon examining the structural interaction between DNA and the DNA binding protein SOX9, the hydrophobic platform and formation of a helix bundle can be defined. This study can provide insight into the sequential process of DNA binding, the relationship between the dimerization domain and HMG domain, and if dimerization affects the affinity of DNA-binding. This also includes whether the stability of SOX9 is affected by the presence of the dimerization domain. In addition, four mutations of SOX9 [71/184] at A118E and A119E within the first alpha helix (α_1) of the HMG domain and L142Q and L145E within the second and linker region of the second (α_2) and third (α_3) alpha helix **(Figure 9A)** can test the role of the conserved di-alanine site and conserved leucines characteristic of the Sox E family of proteins **(Figure 9B)** can be determined via mutagenesis and biological assays. Comprehensively, conducting such experiments can

assist in determining if co-operative binding can play a role such that the binding of one SOX9 [71/184] protein to one DNA site facilitates binding of a second SOX9 [71/184] to a second adjacent site. This can establish a binding mechanism by characterizing the affinity for the dimerization domain with the HMG domain and develop a novel binding structural model between SOX9 with a double-site DNA.

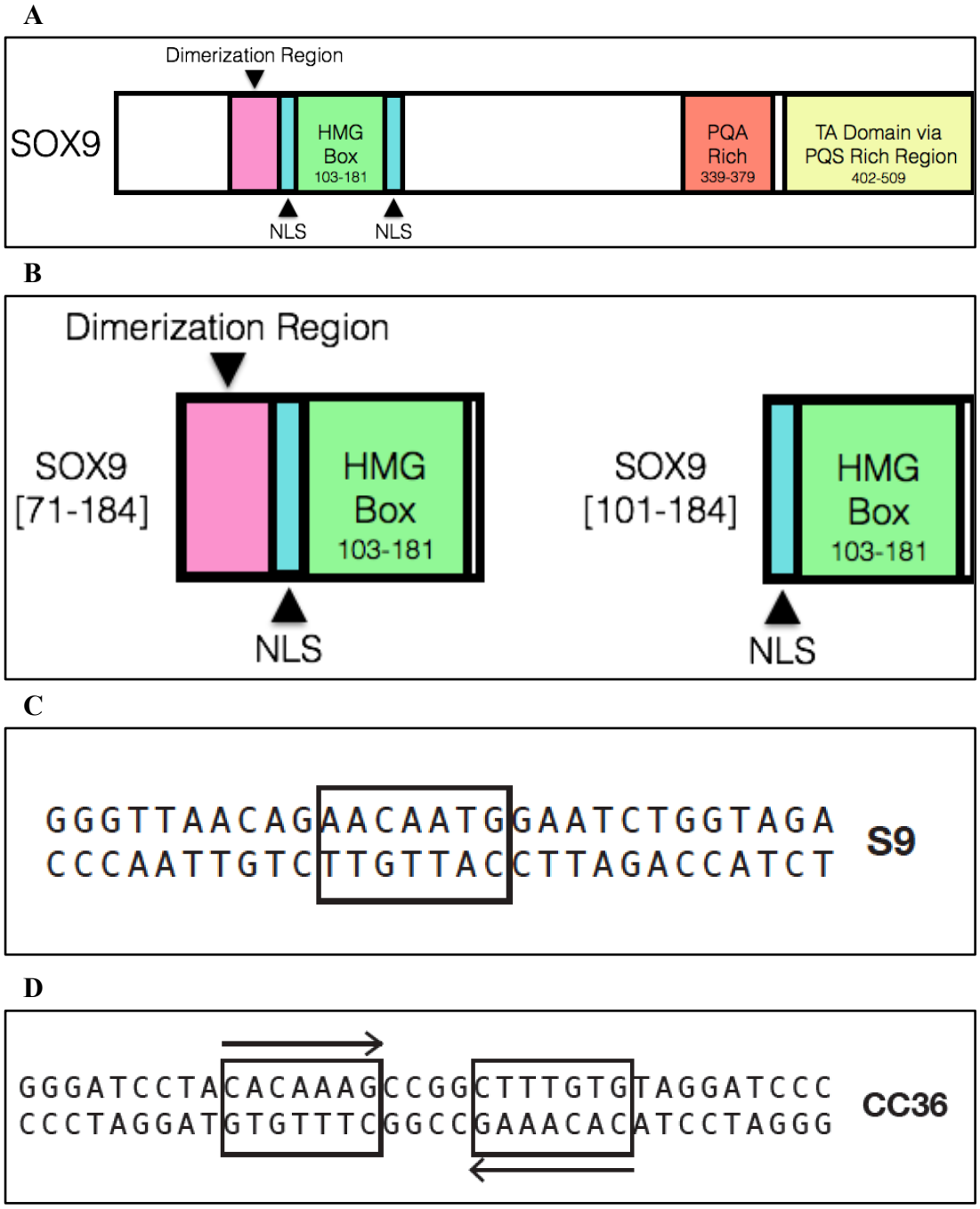
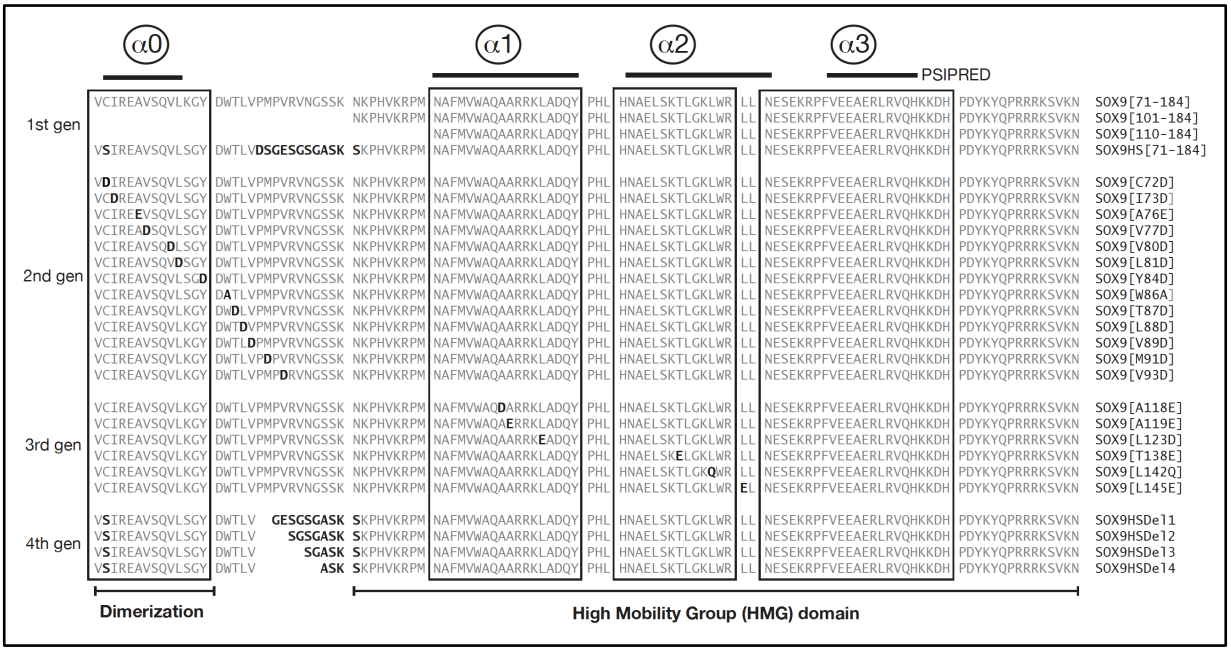


Figure 8: (**A**) Domain organization of SOX9. (**B**) Wildtype SOX9 [71/184] and [101/184] constructs used to conduct this study. (**C**) Single site DNA S9WT baring one canonical binding site. (**D**) A 36 bp palindromic oligonucleotide; CC36, used with the inverted heptameric binding sites indicated within the boxes (*Courtesy of* Dr. Logan Donaldson).

A



B

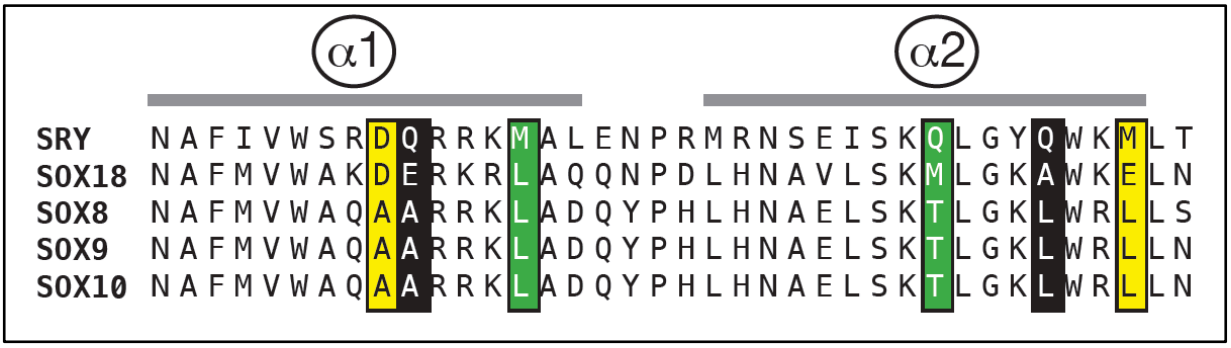


Figure 9: (**A**) Four generations of Human SOX9 mutants. Boxes indicate the sequence boundaries of the dimerization domain (α_0) as well as the HMG domain (α_1, α_2, α_3) that was determined from the crystal structure of the SOX9:DNA complex (PDB: 4EUW). The PSIPRED secondary structure prediction is located above the sequence. (**B**) The sequence alignment of the HMG domains of SOX8, SOX9, and SOX10 that are within the Sox E Group of proteins as well as two non-dimerizing proteins; SRY and SOX18. The black highlights the regions whereby substitutions prevent dimerization while yellow highlighting denotes the regions that substitutions have no affect on dimerization. Green highlighting is indicative of substitutions that affect protein expression, stability, or folding and directly could inhibit the conducting of further experimental assays. The di-alanine residiues in the first alpha helix and two leucines in the second alpha helix are characteristic of the Sox E family of proteins. (*Courtesy of* Dr. Logan Donaldson).

CHAPTER II

EXPERIMENTAL METHODS

2.1 Plasmid Construction

A custom His_6-SOX9 [71/184]gs DNA insert containing a NdeI/XbaI site was developed and purchased from GenScript Corporations. The insert was sub-cloned by a previous lab member into pUC57 and transformed via electroporation into *E.coli* BL21:DE3 (Novagen) to produce His_6-SOX9 [71/184]gs. The cells were plated bearing Kanamycin resistance and grown overnight at 37°C. Individual colonies were sequenced, tested for expression, and prepared as individual glycerol stocks stored in the -80°C freezer for long term use.

2.2 Protein Expression and Purification of His_6-SOX9 [71/184]gs, His_6-MBP-SOX9 [101/184], His_6-SOX9gs [71/184] A118E, His_6-SOX9gs [71/184] A119E, His_6-SOX9gs [71/184] L142Q, and His_6-SOX9gs [71/184] A145E

E.coli BL21 (DE3) cells expressing either His_6-SOX9 [71/184]gs, His_6-SOX9 [101-184], or His_6-SOX9gs [71/184] mutants A118E, A119E, L142Q, and L145E were grown from individual 150mL overnight cultures in 500mL flasks of sterile LB supplemented with 50 μg/mL of Kanamycin and a fraction of the glycerol stock of the clones. The day cultures were grown at 37°C shaking at 200rpm in 1.5 liters of Luria-Bertani (LB) broth supplemented with 50 μg/mL of Kanamycin in a 2.8 liter Allen flask. Day cultured cells were grown to an exponential phase reaching an optical density (OD_{600}) of ~0.8-1.0. Cells were induced with 1.0mM IPTG (Wisent Inc) and the temperature was reduced to 25°C at 200rpm for a five hour post-induction period for

soluble protein production. Following the post-induction period, cell pellets were collected by centrifugation at 7000rpm for 15 minutes at 4°C and placed in the -20°C freezer for a minimum of one hour. Cells were re-suspended over ice in 25mL of 6M urea supplemented with T300 (10mM Tris-HCl ph 8.0, 300mM NaCl). The re-suspended cells were lysed via three passes of compression at 15000psi using the Thermo Spectronic French® Pressure Cell Press. The cell lysate was kept over ice until centrifugation at 17000rpm for 30 minutes at 4°C using a Beckman centrifuge. The soluble extract was then increased to a pH of 7.0 using Tris-HCl pH 7.5 and purified over a 10mL slurry of NI-NTA resin (Qiagen) column with an automated pump system (Amersham Biosciences Pump P-50). The column was primed in T300 at 2ml/minute and upon adding the lysate it was followed by sequential 50mL washes of 6M, 4M, 2M, and 1M urea supplemented with T300 at 3ml/minute. The column was further washed with 150mL of 15mM imidazole supplemented with T300 at 3ml/minute. The his-tagged SOX9 proteins were eluted with 20mM EDTA enriched with T300 at a rate of 2mL/minute following priming of the column with 20mM EDTA for 20 seconds. After collecting approximately 80mL of elution, 1mM PMSF was added. The elution was concentrated down via ultrafiltration using a Sartorius Stedim Biotech Vivaspin Turbo 15 5000 MWCO filter to 15mL of elution with gentile pipetting between each hour of spin at 5600rpm (Eppendorf Centrifuge 5430 R) at 4°C.

The purified and concentrated His_6-MBP-SOX9 [101/184] was placed in 1.5mL eppendorf tubes whereby 3-4 units of Sigma-Aldrich Human Thrombin T1063 was added to each tube containing 1mL of concentrated protein. Tubes were placed on a nutator overnight at 4°C as a means of liberating the SOX9 construct from the his-MBP tag. 4-

(2-Aminoethyl)-benzenesulfonyl fluoride HCL (AEBSF) (Bioshop) was added after 16-18 hours to inactivate thrombin and prevent further cutting. For all SOX9 constructs the 15mL of elution was spun down in 1.5mL eppendorf tubes at 4°C in a benchtop centrifuge (Thermo Scientific Sorvall Legend Micro 17 R Centrifuge) at 13300rpm for 10 minutes to remove any precipitate. 5mL of the SOX9 construct elution was further purified at a time by fast protein liquid chromatography (FPLC) (Amersham pharmacia biotech AKTA FPLC) gel filtration using a Sephacryl S100 16/60 column (GE Biosciences) equilibrated with SOX9 buffer (10mM NaP, 100mM NaCl, 5mM EDTA, 0.2% NaN_3, and 1mM PMSF) ph 6.0 in order to prevent tryptophan proteases from attacking the protein. Fractions pertaining to the molecular weight of the protein of interest were then confirmed at the correct molecular weight on a 12% SDS gel using a protein ladder (Biorad Precision Plus Protein™ Dual Color Standards) and visualized with Fast SeeBand Single Step Protein staining solution (FroggaBio) on the Alpha Imager HP system (Alpha Innotech). Verified fractions were pooled and concentrated by ultrafiltration using a Pall Corporation Macrosep® Advance Centrifugal Device 3000 MWCO filter for His_6-MBP-SOX9 [101/184] and Sartorius MWCO 5000 for all other SOX9 constructs.

Alternatively, if concentration of SOX9 proteins were being lost by the FPLC due to precipitation or affinity to the membrane of the column, dialysis provided a suitable alternative for buffer exchanging. SOX9 was concentrated down to 12mL and injected into a Thermo Scientific Slide-A-Lyzer ® Dialysis Cassette 3500 MWCO using as syringe as per protocol. The cassette was place in a 1L beaker filled with SOX9 buffer and a magnetic stirrer kept the cassette spinning for 1 hour at 4°C. This step was repeated

with fresh buffer after one hour for one final hour and then exchanged again with fresh buffer to spin overnight at 4°C. The following day, the buffer exchanged SOX9 was then removed from the cassette with a syringe and ready for testing. The molar concentrations for all SOX9 constructs were calculated from the A_{280} (Spectronic Unicam Genesys 10uv) (using the extinction coefficient calculated using the ExPASy ProtPram tool (http://web.expasy.org/protparam/) with respect to the confirmed amino acid sequences.

2.3 Non-Radioactive Electro-mobility Shift Assays (EMSA)

Complexes were titrated together between SOX9 protein constructs with a fixed concentration of DNA of either a single SOX9 binding site; S9WT, or a double SOX9 binding site; 36 bp oligonucleotide CC36. The complementary lengths of DNA were designed in 36 base pair length (IDT). The DNA-protein binding reactions were performed with increasing protein concentrations from 0 to 4.0μM in increments of 0.5μM titrated with a constant final DNA concentration of 1μM per reaction in a SOX-DNA binding buffer (10mM NaP, 100mM NaCl, 5mM EDTA, 1mM PMSF, 0.2% NaN_3, pH 6.0, and 10% glycerol). The final reaction volume was 40μL per sample using the SOX9 binding buffer supplemented with 10% glycerol. Complexes were incubated over ice for 30-45 minutes and resolved on 10% TBE gels in the cold room at 4°C with an added 1kB DNA ladder. The gels were stained with 1X SYBR Green I solution (Invitrogen) in a 1:10000 dilution for 40 minutes and visualized with an Alpha Imager HP system (Alpha Innotech). Radioactive EMSA's were performed in a similar manner in the nanomolar range, with the exception of P-32 radioactive tagging (Perkin Elmer) of the DNA and complex formation performed by Ana Vakiloroayaei. Radioactive detection was conducted on a Typhoon 9400 Multi-Format Imager.

2.4 Fluorescent-Peptide Electro-mobility Shift Assays (EMSA)

A shift assay was used to detect the binding of a FITC-peptide **(GVSIREAVSQVLSGYD)** sequence of the SOX9 dimerization domain that was bound to a complex with either wildtype or mutant SOX9 proteins to the main oligonucleotide CC36 with the double SOX9 binding sites as well as the single SOX9 binding site S9WT. Complexes were constructed and incubated over ice for 30 minutes. Protein-DNA complexes were titrated together with a fixed final concentration of 1μM DNA of either S9WT or CC36 with a fixed concentration of 2μM of SOX9 protein constructs SOX9 [71/184], SOX9 [101/184], or SOX9 [71/184] A118E, A119E, L142Q, and L145E to form 1:2 DNA:SOX9 protein binding reactions. FITC labeled peptide was added to the complex with an increasing concentration from 0 to 40μM (0μM, 2μM, 3μM, 6μM, 8μM, 10μM, 15μM, 20μM, and 40μM). The final reaction volume was 40μL per sample using the SOX9 binding buffer supplemented with 10% glycerol. Complexes were incubated over ice for 30-45 minutes in the dark to prevent further light exposure and resolved on 10% TBE gels with an added 1kb DNA ladder in the cold room at 4°C. The gels were visualized using a Carestream 4000 MM Pro Image Station using a multi-wavelength illumination source with corresponding filters for FITC detection at 495nm excitation and 520 nm emission that is characteristic to a FITC tag (Carestream Health). Imaging and band intensities were measured using the Carestream Molecular Imaging Software.

2.5 Circular Dichroism (CD)

The far UV spectra of His_6-SOX9 [71/184] and His_6-MBP-SOX9 [101/184] were acquired with a Jasco J-810 CD Spectrometer in conjunction with Spectra Manager 1.54.03 and Spectra Analysis 1.54.04 software (Jasco 815CD) at 20°C using a cell with a constant 0.1 cm path length for all measurements. Spectra for SOX9 protein constructs alone, either CC36 or S9WT alone, as well as the DNA-protein complex spectras were recorded over a wavelength range from 190-300 nm. All spectra samples were acquired in a CD binding buffer (10mM NaH_2PO_4 and 100mM KF). A blank scan of buffer alone was also performed to use as a baseline for data processing. All proteins and DNA had a final concentration of 30μM in a final volume of 200μL per reaction. There was a total of three scans performed for each sample and the data was normalized to mean residue ellipticity using the molar protein concentration and number of amino acids for each of the individual SOX9 protein constructs. The final CD spectrums for the SOX9 proteins of interest were determined by subtracting the free CC36 or S9WT DNA spectra (30μM) from the complex (both at 30μM each). Thermal denaturation experiments were also conducted for both SOX9 protein alone and in complex with either DNA and monitored with the molar ellipticity at 208nm and 222nm at 2°C/minute from 4°C to 90°C. Spectra deconvolution was achieved using the online software analysis program CONTINLL (Dicroweb) and K2D2 (http://k2d2.ogic.ca).

2.6 Fluorescence Anisotropy

In this study, fluorescence anisotropy was used to determine the binding constants and kinetics of the reaction between the single site FITC-S9WT and an increasing concentration of either SOX9 [71/184] or SOX9 [101/184]. 100 μL of pure SOX9 [71/184] or [101/184] protein starting at 400μM was serially diluted to concentrations ranging from 200μM down to 1μM with a constant concentration of 1μM of FITC-S9WT in a final reaction volume of 40μL per sample using the SOX9 binding buffer. Reactions were placed on ice for 30-45 minutes. 10μL reactions were then added to a black background 384 well plate in an order of increasing protein concentration and repeated in four rows. In addition, four rows each of the buffer alone, protein in buffer alone, and DNA in buffer alone were also subjected to polarization tests in two separate readings. Foil was placed over the well plate to prevent the exposure of light and placed in a centrifuge for 5 minutes at 1000rpm in order to settle the samples to the bottom of each well. The plate was read using a Synergy H4 Hybrid reader using polarization intensities with an excitation at 495nm and emission at 520nm that is characteristic to a FITC tag. Each reading contains two readings for parallel and two readings for perpendicular whereby a final polarization is calculated using the Gen5 software corresponding to the Synergy apparatus. The final four readings from each of the rows were averaged for a final polarization value for each concentration and then repeated three times for a final average.

2.7 ^{1}H-^{15}N HSQC Nuclear Magnetic Resonance (NMR) Spectroscopy of SOX9 [101/184]

^{15}N-isotopically labeled SOX9 [101/184] complexed with the single site S9WT DNA was produced by growing 3L of *E.coli* BL21:DE3 in M9 minimal media supplemented with 1 g/L [^{15}N] ammonium chloride as a means of being the sole nitrogen source. Singly labeled proteins were generated for ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) investigations. Previous growth and purification protocols were implemented and following gel filtration, pure protein was concentrated to 0.1-0.15M using a Pall Corporation MWCO 3000 filter overnight at 4°C and supplemented with 10% D_2O. A 1:1 concentration of DNA:SOX9 were complexed together and left over ice for 30-45 minutes and HSQC spectras were acquired on a 700 MHz Bruker Ascend NMR spectrometer.

CHAPTER III

RESULTS

3.1 Stability of SOX9 Constructs

The expression of SOX9 [71/184], detagged-SOX9 [101/184], SOX9 [71/184] A118E, SOX9 [71/184] A119E, SOX9 [71/184] L142Q, and SOX9 [71/184] A145E **(Figure 10)** were all confirmed with their correct molecular weights by SDS-PAGE. SOX9 is sensitive to temperature, pH, and salt conditions. Detagged-SOX9 [101/184] expression was more limited in 15N media than LB, which required an increase in culturing with lower and longer post-induction temperatures in order to compensate for a decrease in expression. All six expressed proteins remained stable in a sodium phosphate buffer pH 6-8 with salt concentrations between 100-500mM. Aggregation became an issue for the constructs over 100mM in concentration as precipitate formed and facilitated a loss of more than 50% of protein by the final stages of purification and storage. When kept on ice, loss of protein due to degradation and aggregation were greatly reduced. In addition, protease inhibitors were added to elutions and final buffers to reduce serine proteases as well as buffers were reduced to pH 6 to reduce tryptophan proteases. Both constructs generated ~90% of recombinant proteins were produced as insoluble aggregates. This was resolved by solubilizing the inclusion bodies with 6 M urea prior to conducting affinity purifications. The shelf life for the SOX9 proteins were approximately 5-10 days however with the use of 30% glycerol and storage in the -20°C freezer, the shelf life could be extended for a couple of months.

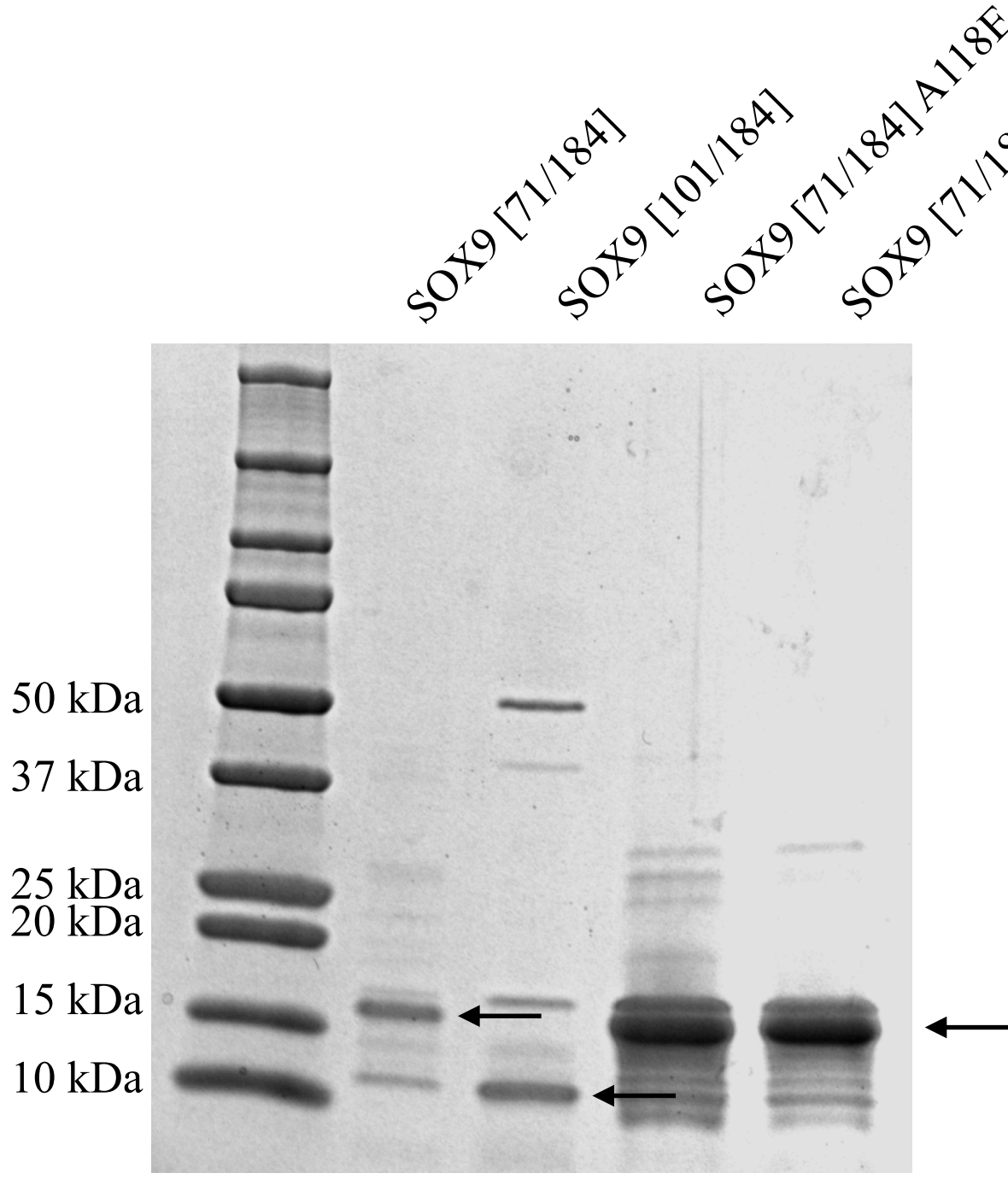
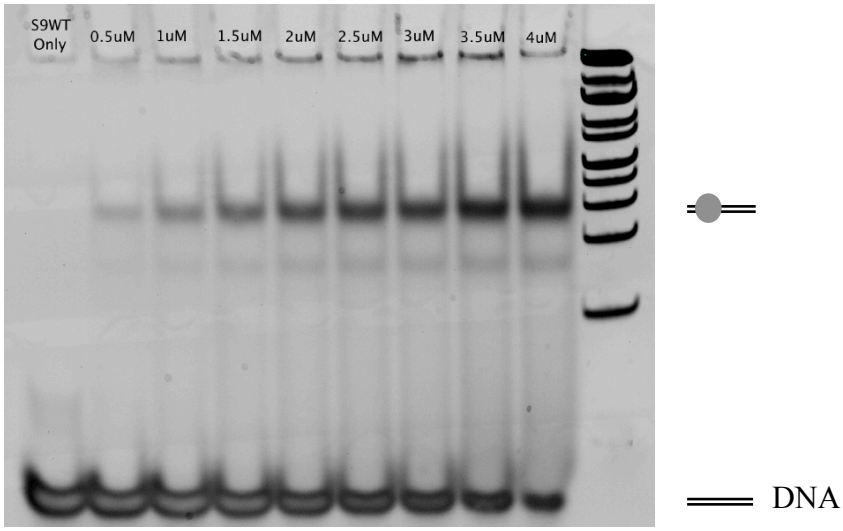


Figure 10: Confirmation of SOX9 constructs SOX9 [71/184], detagged-SOX9 [101/184], SOX9 [71/184] A118E, and SOX9 [71/184] A119E with their correct molecular weights at 15kDa, 10kDa, 15kDa, and 15kDa respectively. Visualized on a 12% SDS gel using a protein ladder (Biorad Precision Plus Protein™ Dual Color Standards) and visualized with Fast SeeBand Single Step Protein staining solution (FroggaBio) on the Alpha Imager HP system (Alpha Innotech). Arrows indicate the band with the correct molecular weight.

3.2 DNA Binding of Wildtype SOX9 [71/184] and [101/184] and SOX9 Mutants

Non-radioactive electrophoretic mobility shifts assays (EMSAs) of S9WT complexed with SOX9 [71/184] **(Figure 11A)** or monomeric SOX9 [101/184] **(Figure 11B)** with S9WT were used to visualize and confirm that SOX9 does bind to DNA and demonstrating that the minimal domain required for binding is the HMG domain. This is demonstrated by the binding between SOX9 [101/184] and S9WT or CC36. CC36 complexed with either SOX9 [71/184] **(Figure 12A)** or SOX9 [101/184] **(Figure 12B)** demonstrates the role of cooperativity when the dimerization domain is present. With just the HMG present SOX9 [101/184] initially binds to one site until the concentration of protein is about 2μM at which point it binds to and saturates both sites on CC36. While SOX9 [71/184] containing both the dimerization domain and the HMG bound to both sites almost simultaneously, revealing that even at low concentrations immediate patterns of occupying both sites on the DNA. When compared to SOX9 [71/184] HMG domain mutants A118E and A119E the A118E mutant **(Figure 15A)** was still able to demonstrate dimerization with a similar pattern the wildtype SOX9 [71/184]. However A119E **(Figure 15B)** demonstrated a loss of dimerization. Although, both mutant constructs retained the ability to bind to DNA. Similarly, the SOX9 [71/184] L142Q mutant bound to CC36 **(Figure 16A)** demonstrates a loss of dimerization comparable to A119E. Likewise, L145E mutant **(Figure 16B)** maintains its ability to dimerize and exhibit cooperativity much like the A118E mutant.

A



B

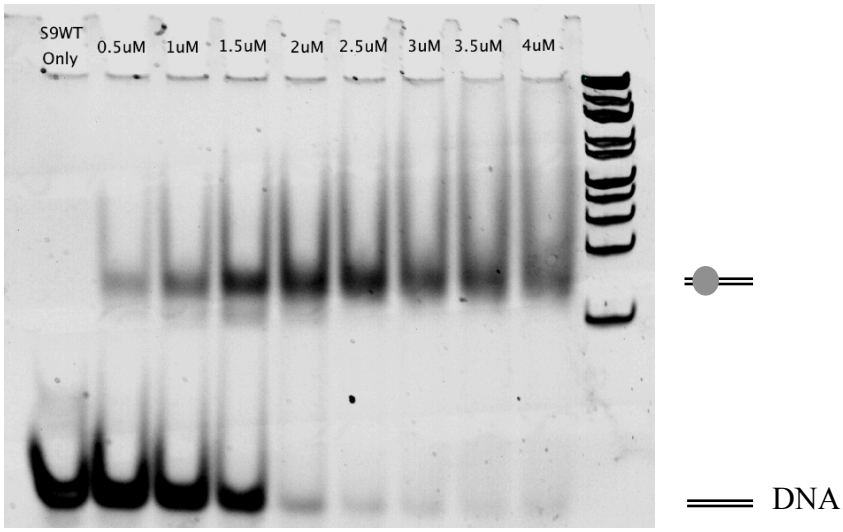
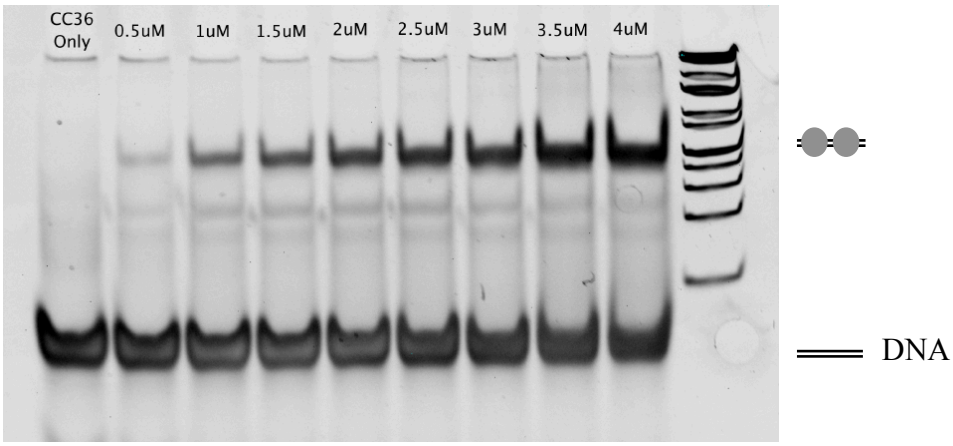


Figure 11: Non-radioactive electrophoretic mobility shift assays of wildtype SOX9 constructs with the single-site S9WT DNA. (**A**) EMSA shift of wildtype SOX9 [71/184] and (**B**) of wildtype SOX9 [101/184] with an increasing concentration (0μM, 0.5μM, 1μM, 1.5μM, 2μM, 2.5μM, 3μM, 3.5μM, 4μM) in complex with 1μM of S9WT. Complex was resolved on a 10% TBE gel and stained with 1:10000 dilution of SYBR Green. The final lane contains a 1kb DNA ladder.

A



B

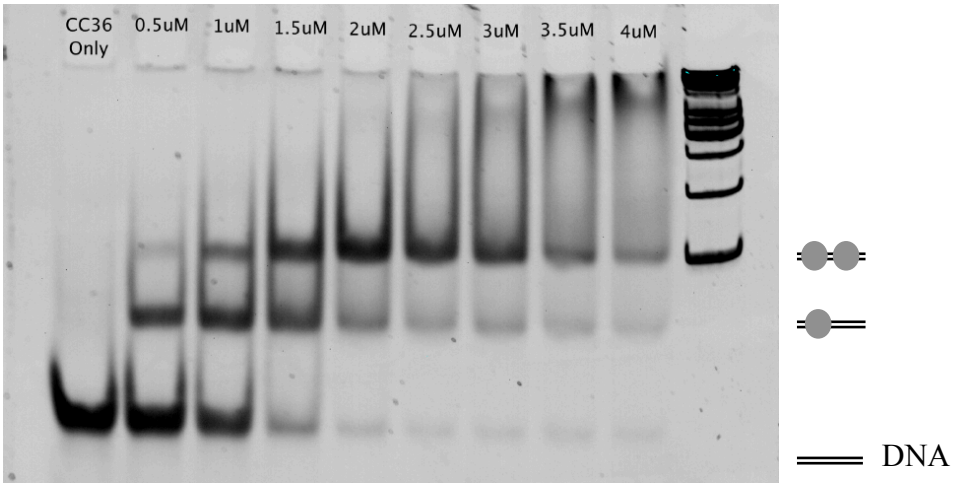


Figure 12: Non-radioactive electrophoretic mobility shift assays of wildtype SOX9 constructs with the double-site 36bp CC36 DNA. (**A**) EMSA shift of wildtype SOX9 [71/184] demonstrating cooperativity and (**B**) of wildtype SOX9 [101/184] occupying the one of two sites on the DNA first with an increasing concentration (0μM, 0.5μM, 1μM, 1.5μM, 2μM, 2.5μM, 3μM, 3.5μM, 4μM) in complex with 1μM of CC36. Complex was resolved on a 10% TBE gel and stained with 1:10000 dilution of SYBR Green. The final lane contains a 1kb DNA ladder.

Radioactive EMSAs provided a means to determine the K_d for SOX9 [71/184] with S9WT **(Figure 13A)**, SOX9 [101/184] with S9WT **(Figure 13B)**, SOX9 [71/184] with CC36 **(Figure 13C)**, and SOX9 [101/184] with CC36 **(Figure 13D)**. Binding shifts were demonstrated under 20nM of SOX9 [101/184] with 9nM of S9WT while other complexes demonstrated visible shifts under 50nM. The plotted percent of SOX9 protein unbound to either DNA **(Figure 13E)** revealed a sigmoidal binding curve. However, the algorithmic calculations using Prism (GraphPad Software) of the K_d were approximately 35nM for SOX9 [101/184] with S9WT and approximately 100nM for the other constructs. SOX9 [71/184] with S9WT bound to 83% completion and as a result will need to be repeated.

3.3 The Surface on the Sox Group E HMG Domain also Participates in Binding with the Dimerization Domain

FITC-peptide of the dimerization domain sequence (α_0) of SOX9 enabled florescence electrophoretic mobility shift assays to be conducted using a 2:1 preformed complex of SOX9 [71/184] or SOX9 [101/184] with DNA. An increase in the intensity was present with an increase in peptide concentration for SOX9 [71/184] bound to S9WT **(Figure 14A)** and when bound to CC36 **(Figure 14C)**. A similar intensity increase was detected when SOX9 [101/184] was bound to S9WT **(Figure 14B)** or CC36 **(Figure 14D).** Peptide titrated with DNA alone demonstrated no binding with any construct. However, the mutants SOX9 [71/184] A118E and A119E demonstrate very different affinities for the FITC-peptide. SOX9 [71/184] A118E bound to CC36 **(Figure 15C)** reveals a similar increase in intensity in FITC-peptide as the concentration increases, as did the wildtype SOX9 constructs. However, SOX9 [71/184] A119E bound to CC36 **(Figure 15D)** reveals no detection of the peptide binding to the preformed complex. The

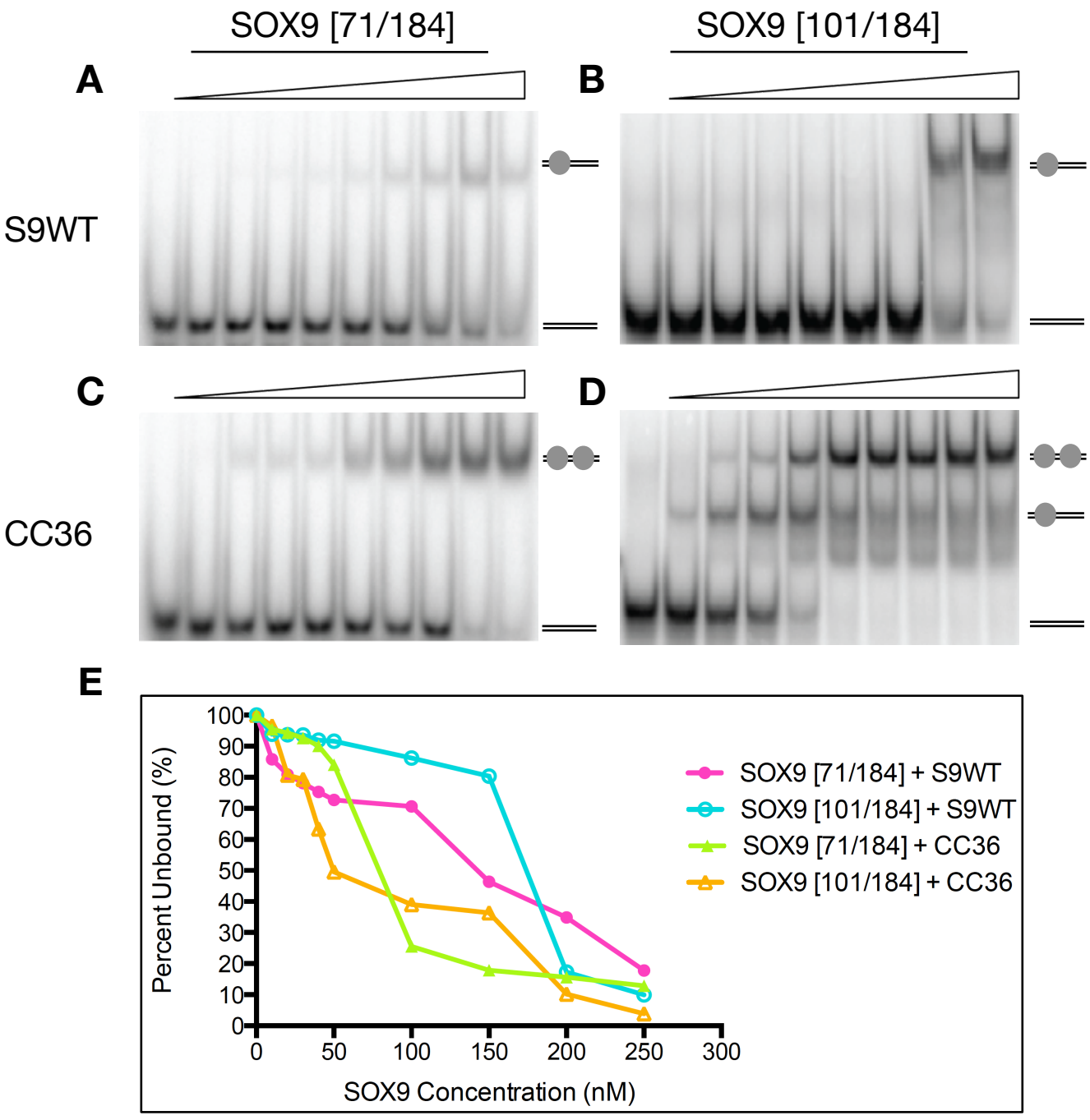


Figure 13: Radioactive electrophoretic mobility shift assays of S9WT with (**A**) SOX9 [71/184] and (**B**) SOX9 [101/184] and using CC36 with (**C**) SOX9 [71/184] and (**D**) SOX9 [101/184] with an increasing concentration of SOX9 proteins (0nM, 10nM, 20nM, 30nM, 40nM, 50nM, 100nM, 150nM, 200mM, and 250nM). (**E**) Sigmodial plot of the unbound percentages of the SOX9 proteins bound to DNA.

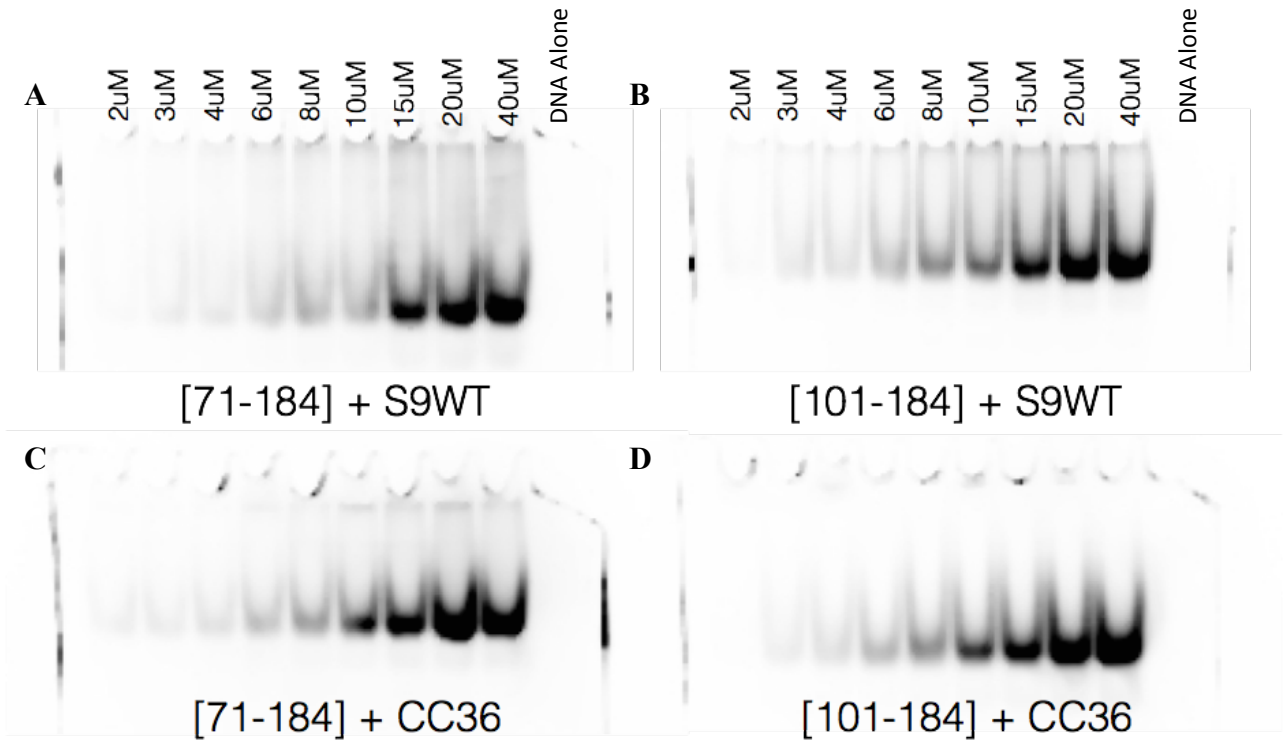


Figure 14: FITC-peptide of SOX9 dimerization domain EMSA with 2µM of wildtype (**A**) SOX9 [71/184] with 1µM S9WT; (**B**) SOX9 [101/184] with 1µM S9WT; (**C**) SOX9[71/184] with 1µM CC36; (**D**) SOX9 [101/184] with 1µM CC36. FITC-peptide in an increasing concentration (2µM, 3µM, 4µM, 6µM, 8µM, 10µM, 15µM, 20µM, 40µM, and 5µM peptide with 1µM DNA alone). Complex was resolved on a 10% TBE gel and detected with UV-light on a CareStream Imager (Carestream Health). There was no detection of FITC-peptide binding to DNA alone.

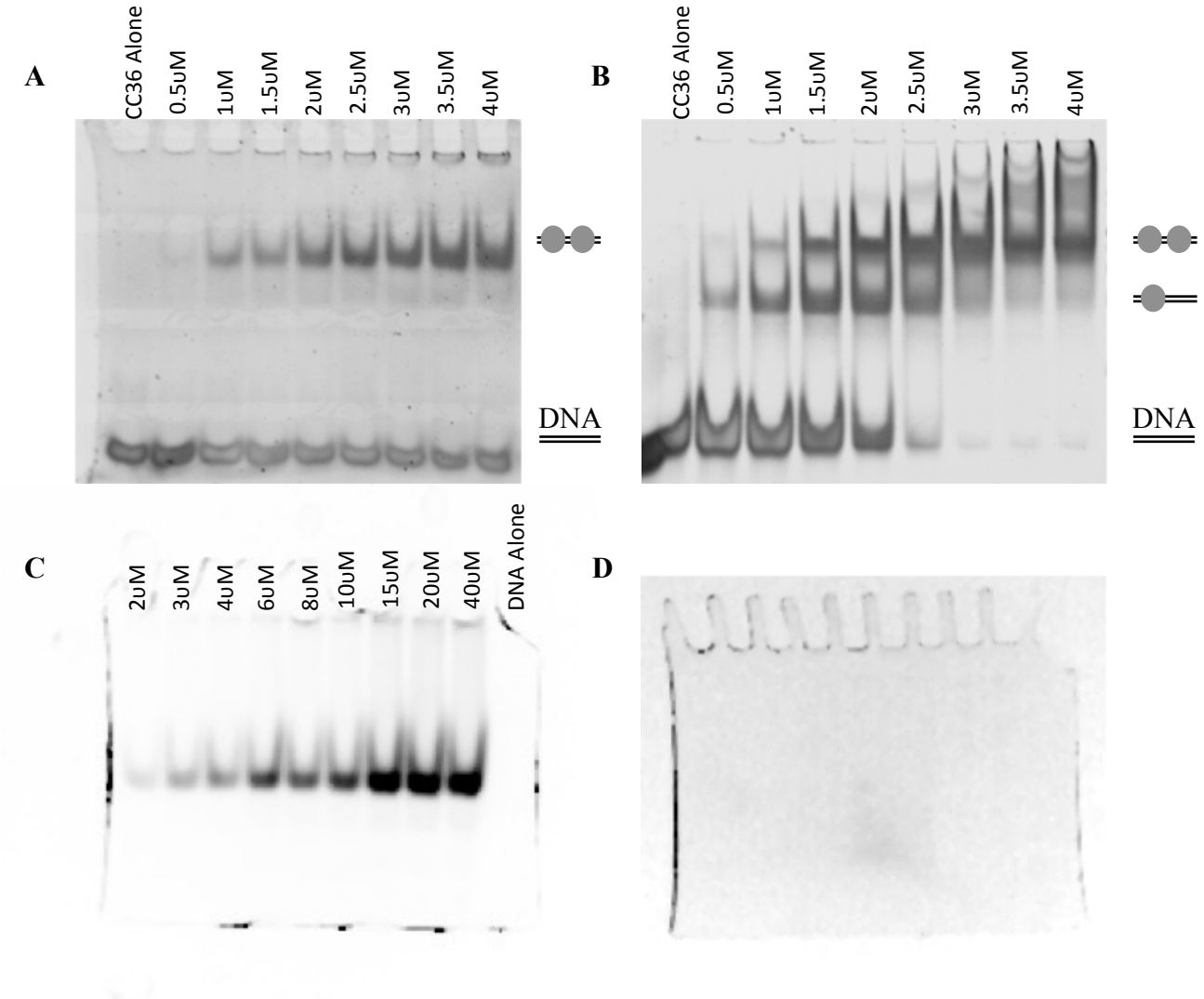


Figure 15: Non-radioactive electrophoretic mobility shift assays of mutant SOX9 constructs with the double-site 36bp CC36 DNA. (**A**) EMSA shift of mutant SOX9 [71/184] A118E and (**B**) of mutant SOX9 [71/184] A119E with an increasing concentration (0μM, 0.5μM, 1μM, 1.5μM, 2μM, 2.5μM, 3μM, 3.5μM, 4μM) in complex with 1μM of CC36. Complex was resolved on a 10% TBE gel and stained with 1:10000 dilution of SYBR Green. The final lane contains a 1kb DNA ladder. (**C**) FITC-peptide of SOX9 dimerization domain EMSA with 2μM SOX9 [71/184] A118E with 1μM of CC36 DNA (**D**) or 2μM SOX9 [71/184] A119E and 1μM CC36 with 1μM of CC36 DNA. Increasing concentration of FITC-peptide (2μM, 3μM, 4μM, 6μM, 8μM, 10μM, 15μM, 20μM, 40μM, and 5μM peptide with 1μM CC36 DNA alone). Complex was resolved on a 10% TBE gel and detected with UV-light on a CareStream Imager (Carestream Health). There was no binding detection of peptide with the mutant SOX9 [71/184] A119E or detection of FITC-peptide binding to DNA alone.

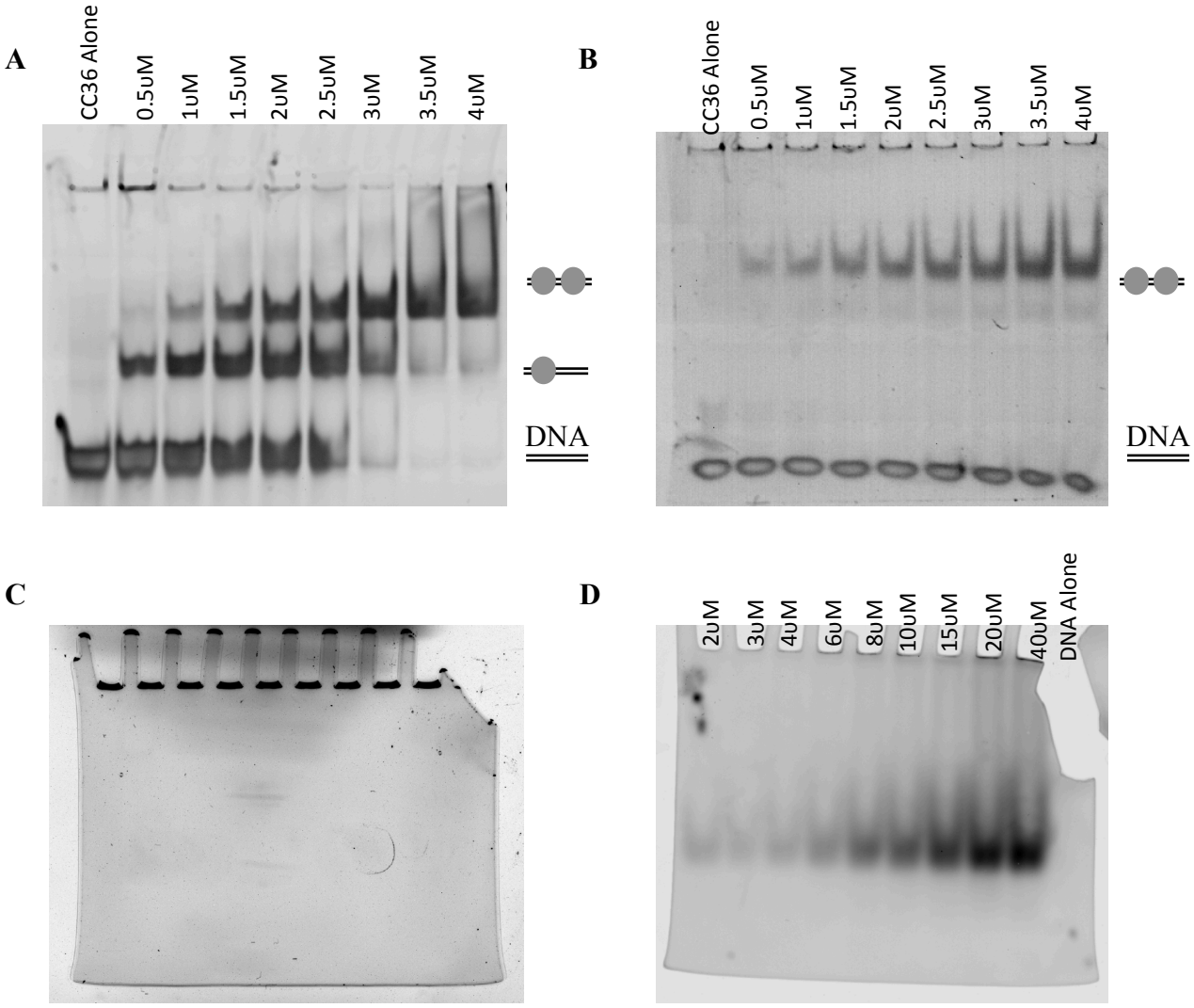


Figure 16: Non-radioactive electrophoretic mobility shift assays of mutant SOX9 constructs with the double-site 36bp CC36 DNA. (**A**) EMSA shift of mutant SOX9 [71/184] L142Q and (**B**) of mutant SOX9 [71/184] L145E with an increasing concentration (0μM, 0.5μM, 1μM, 1.5μM, 2μM, 2.5μM, 3μM, 3.5μM, 4μM) in complex with 1μM of CC36. Complex was resolved on a 10% TBE gel and stained with 1:10000 dilution of SYBR Green. The final lane contains a 1kb DNA ladder. (**C**) FITC-peptide of SOX9 dimerization domain EMSA with 2μM SOX9 [71/184] L142Q with 1μM of CC36 DNA (**D**) or 2μM SOX9 [71/184] L145E and 1μM CC36 with 1μM of CC36 DNA. Increasing concentration of FITC-peptide (2μM, 3μM, 4μM, 6μM, 8μM, 10μM, 15μM, 20μM, 40μM, and 5μM peptide with 1μM CC36 DNA alone). Complex was resolved on a 10% TBE gel and detected with UV-light on a CareStream Imager (Carestream Health). There was no binding detection of peptide with the mutant SOX9 [71/184] L142Q or detection of FITC-peptide binding to DNA alone.

K_d and Hill coefficient values associated with the peptide binding to any of the SOX9-DNA complexes **(Table 2)** are fairly similar; ~8-9 K_d with a value of 3 for the Hill coefficient, and conform to a similar sigmoidal curve when normalized **(Figure 17).** This is with the exception of SOX9 [71/184] A119E whereby there were no band intensities to measure. Similarly, the SOX9 [71/184] L142Q mutant bound to CC36 **(Figure 16C)** was unable to bind with the peptide and no bands were detected. Unlike, mutant L145E **(Figure 16D)** that retained an affinity for the peptide of the dimerization domain and exhibited an increase in fluorescence intensities as the peptide concentration increased.

3.4 Stability or DNA Binding Affinity of the Wildtype SOX9-DNA Complex is Not Enhanced by the Dimerization Domain

The qualitative circular dichroism spectroscopy scans for SOX9 [71/184] demonstrated a shift in the alpha helix profile when bound to S9WT and CC36 compared to the protein alone as demonstrated by the raw data confirming the presence of the complex **(Figure 18A & 18B)** and the actual shifts when the DNA was subtracted **(Figure 18C)**. Similarly, a shift in the profile was present in the scans for SOX9 [101/184] bound to S9WT and CC36. This is confirmed by the raw data complex formed between the SOX9 protein and either S9WT or CC36 **(Figure 19A & 19B)** as well as when DNA is subtracted from the profiles **(Figure 19C)**. Spectra deconvolution was achieved using the online software analysis program CONTINLL (Dicroweb) and K2D2 (http://k2d2.ogic.ca) to determine 91% alpha helix protein structure for SOX9 [71/184] alone and when complexed to CC36 and 88% when complexed to S9WT. Similarly, SOX9 [101/184] alone or when bound to either CC36 or S9WT had a 94% alpha helix structure with a 0.01% likelihood of a beta sheet present. The secondary structure of both SOX9 protein constructs alone were also confirmed using the secondary structure

Table 2: Kd and Hill values for His_6-SOX9 wildtype and mutant constructs with S9WT or CC36 DNA

	SOX9 [71/184] + CC36	SOX9 [101/184] + CC36	SOX9 [71/184] A118E + CC36	SOX9 [71/184] + S9WT	SOX9 [101/184] + S9WT
Kd	8.35±0.27	7.95±0.21	6.90±2.16	10.45±0.51	7.02±0.21
Hill	3.170	2.900	2.160	3.260	2.830

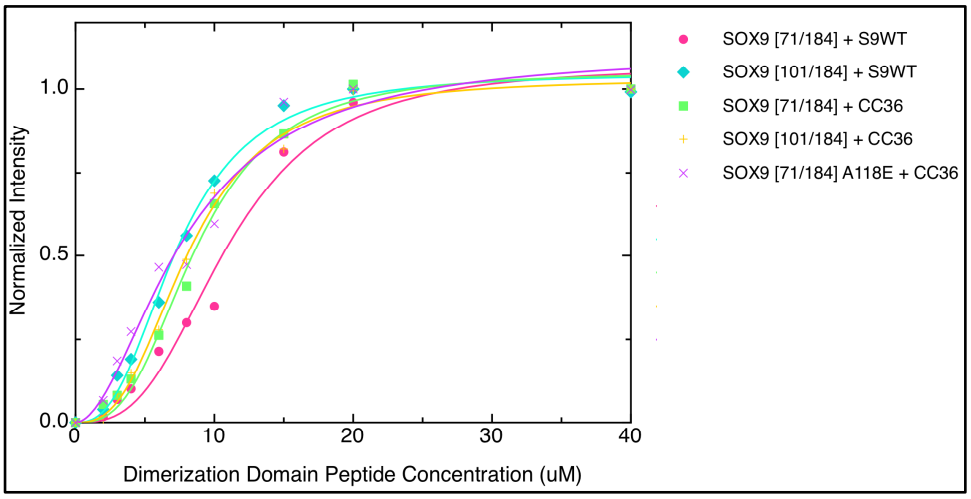
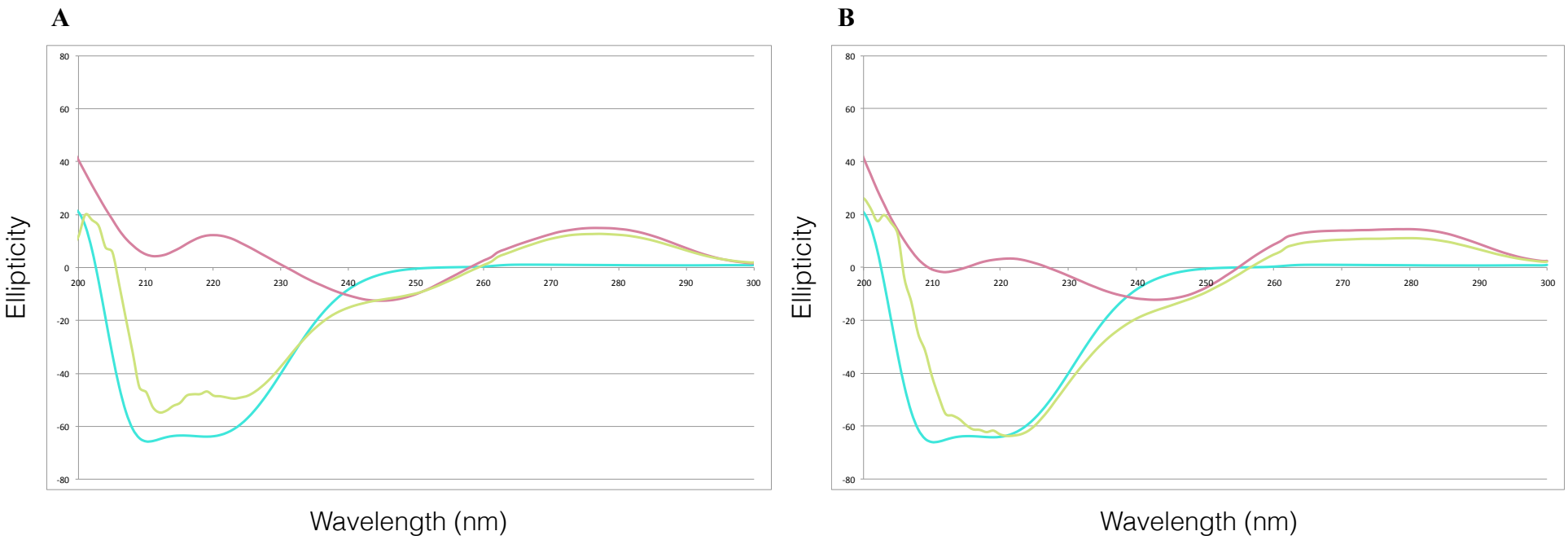


Figure 17: Normalized intensities for His_6-SOX9 wildtype and mutant constructs against FITC-peptide concentration (μM) of the SOX9 dimerization domain sequence.



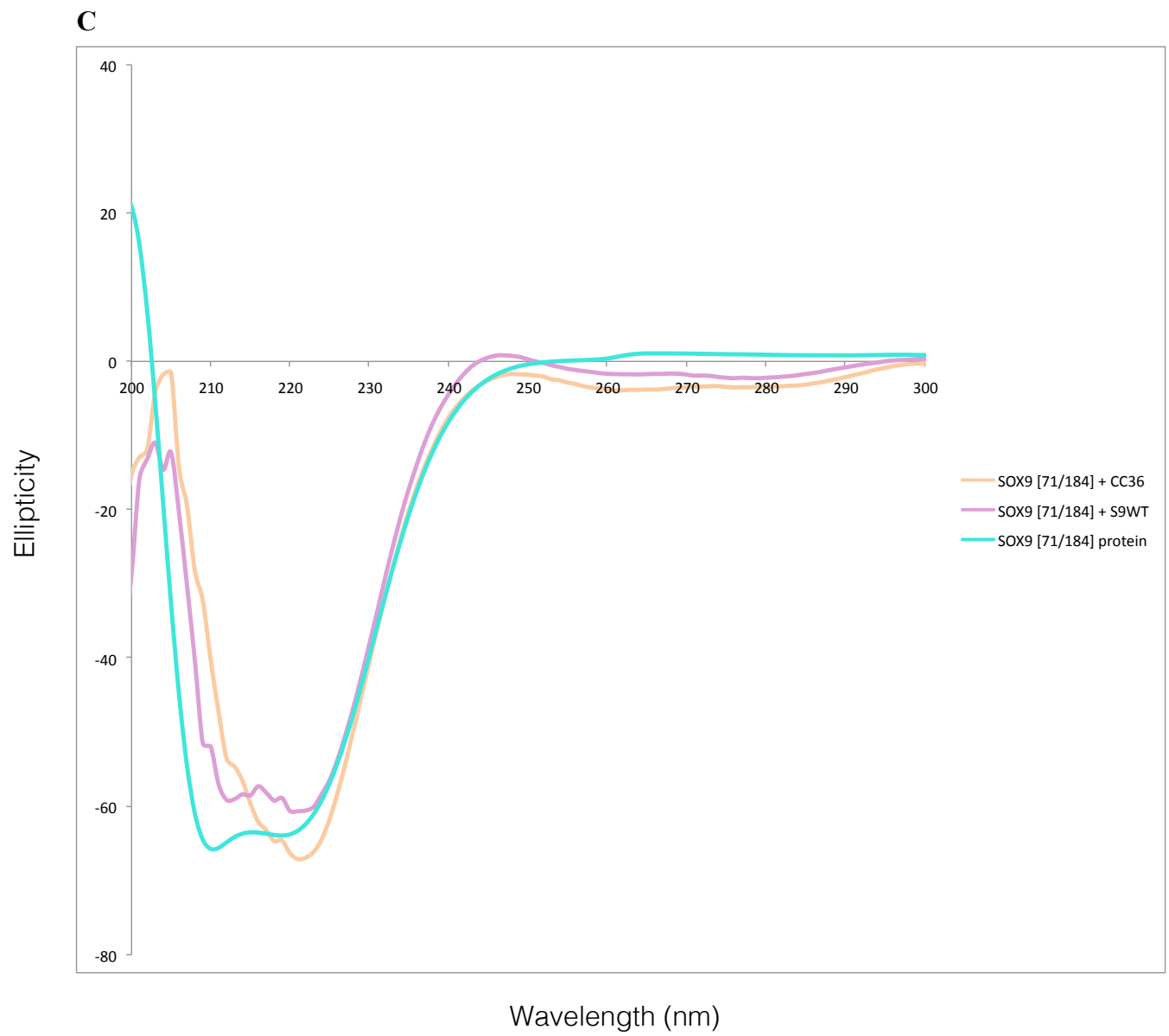
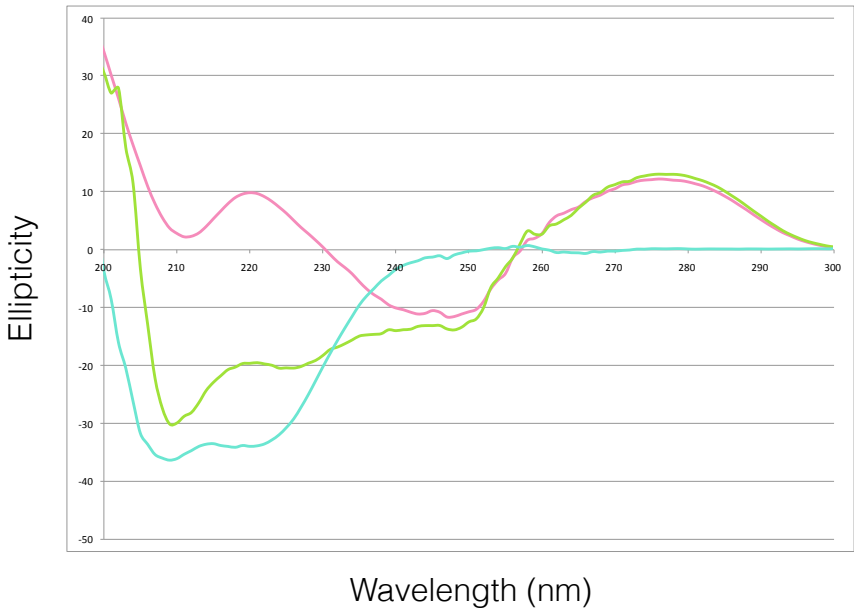
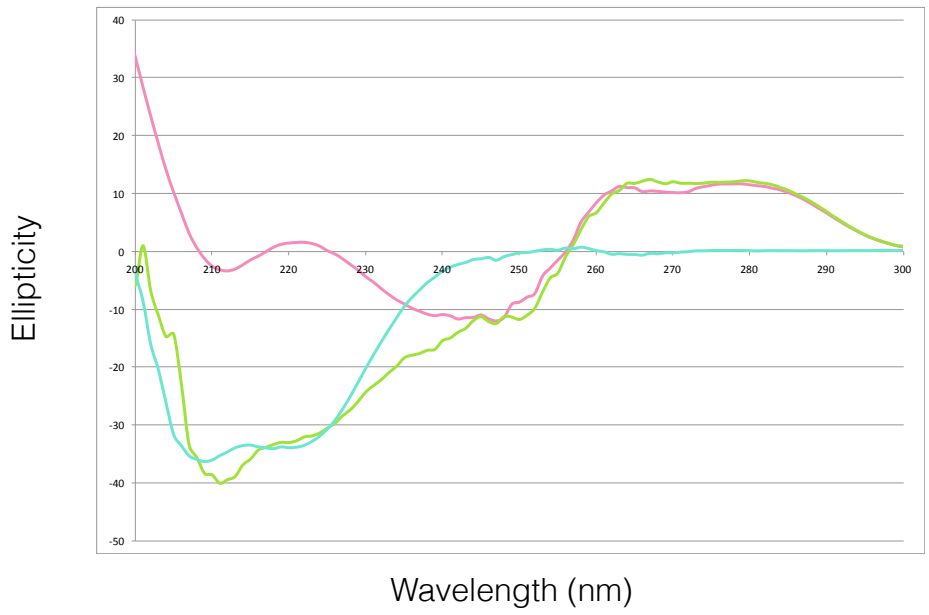


Figure 18: Raw data of free (blue) and bound (green) His_6-SOX9 [71/184] at 20°C with (**A**) S9WT DNA (pink) (**B**) CC36 DNA (pink). CD spectra of the protein and DNA region (200-300nm). (**C**) Normalized CD spectra of His_6-SOX9 [71/184] with either S9WT (purple) or CC36 (orange) compared to His_6-SOX9 [71/184] protein alone (blue) demonstrates and altered alpha helix profile when bound to DNA.

A



B



C

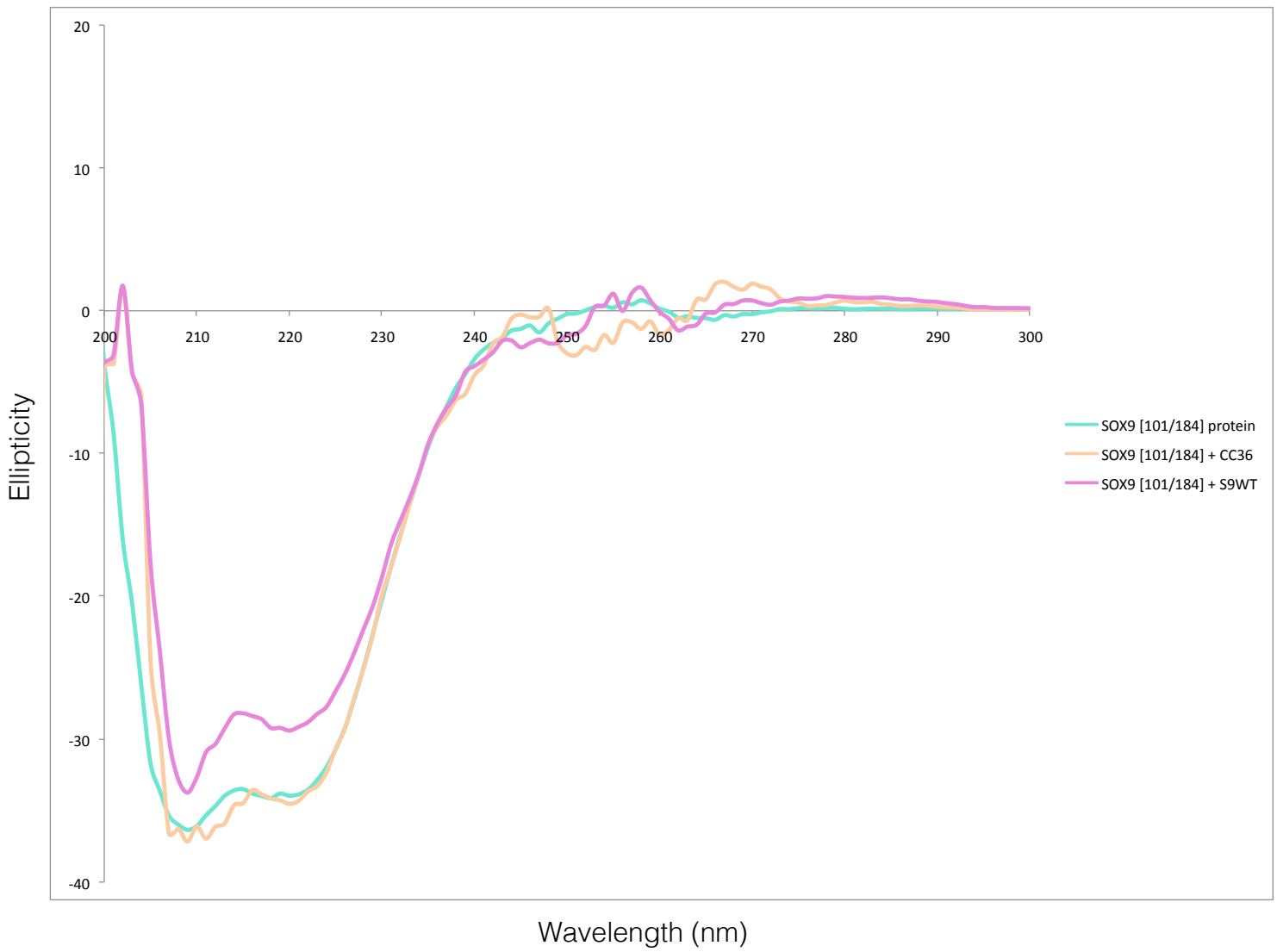


Figure 19: Raw data of free (blue) and bound (green) His_6-SOX9 [101/184] at 20°C with (**A**) S9WT DNA (pink) (**B**) CC36 DNA (pink). CD spectra of the protein and DNA region (200-300nm). (**C**) Normalized CD spectra of His_6-SOX9 [101/184] with either S9WT (purple) or CC36 (orange) compared to His_6-SOX9 [101/184] protein alone (blue) demonstrates and altered alpha helix profile when bound to DNA.

prediction program PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) **(Figure 20)**. As forecasted, the program predicted that SOX9 [71/184] contains one alpha helix within the dimerization domain and three alpha helices within the HMG domain while SOX9 [101/184] only contains the three alpha helices within the HMG domain.

The dimeric SOX9 [71/184] and monomeric SOX9 [101/184] proteins revealed a thermal denaturation midpoint (T_m) at 45 ˚C for either protein constructs alone. This indicates that the HMG domain remains largely unfolded at room temperature. Given that the melting temperatures of both are the same, it suggests that the dimerization domain does not play a significant role in the stabilization of the HMG domain. When the studies were performed with the proteins in complex with the 36 bp DNA duplex, CC36; a palindrome containing two inverted CACAAAG binding sites, as well as a single site DNA; S9WT the Tm increased respectively for SOX9 [71/184] **(Figure 21A)** when complexed with S9WT to 68°C and CC36 to 69°C. Likewise, the values for SOX9 [101/184] **(Figure 21B)** increased when complexed to S9WT to 65°C and with CC36 to 65°C. Furthermore, melting of the DNA alone was approximately 60°C. This proposes that stabilization of the DNA-bound protein complex is not hindered or affected by any protein-protein contacts as a result of the dimerization region being present.

The reaction between the single site FITC-S9WT and an increasing concentration of either SOX9 [71/184] or SOX9 [101/184] enabled a change in rotational time of the molecules such that if the SOX9 protein bound to the FITC-DNA it would slow down the rotation of the DNA and enable an increase in detection of polarization. The fluorophore bound to S9WT demonstrated a difference in polarization between the unbound and bound state. SOX9 [71/184] bound to FITC-S9WT demonstrated a smaller spike in

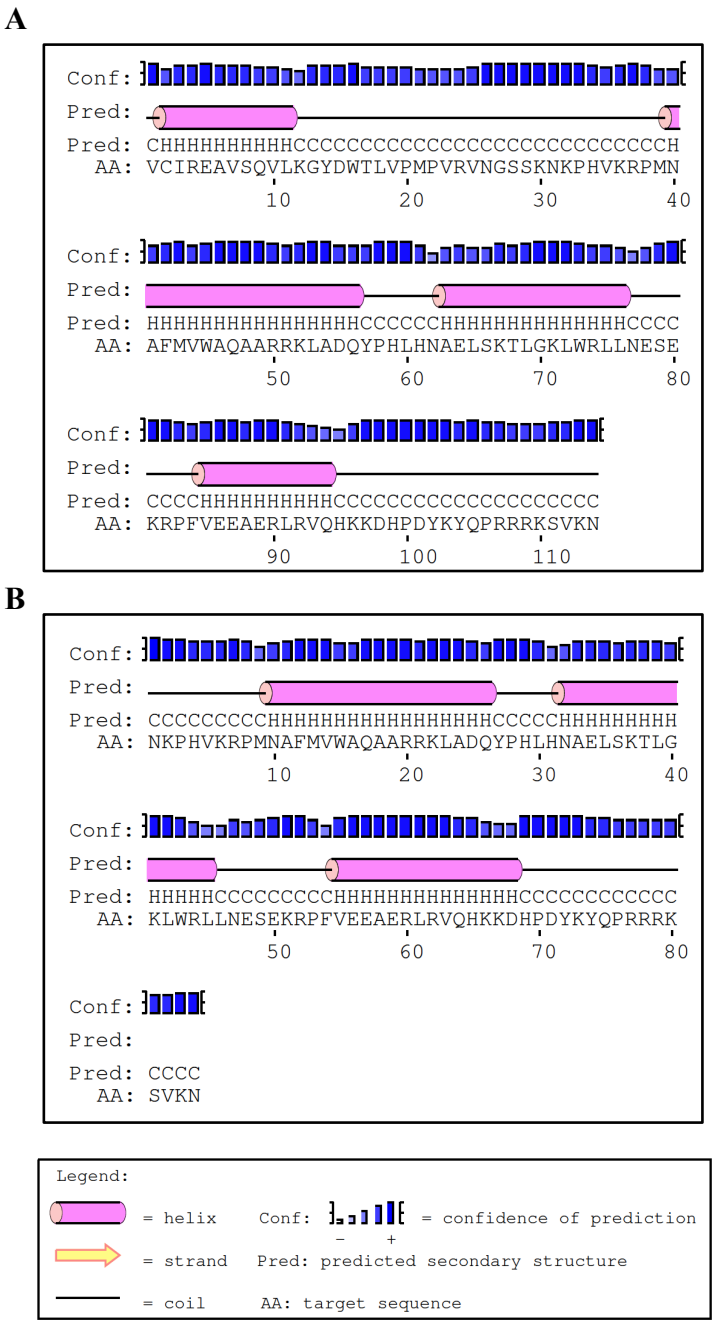


Figure 20: Secondary structure prediction of SOX9 constructs. The alpha helices prediction of **(A)** SOX9 [71/184] **(B)** SOX9 [101/184] using prediction program PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/).

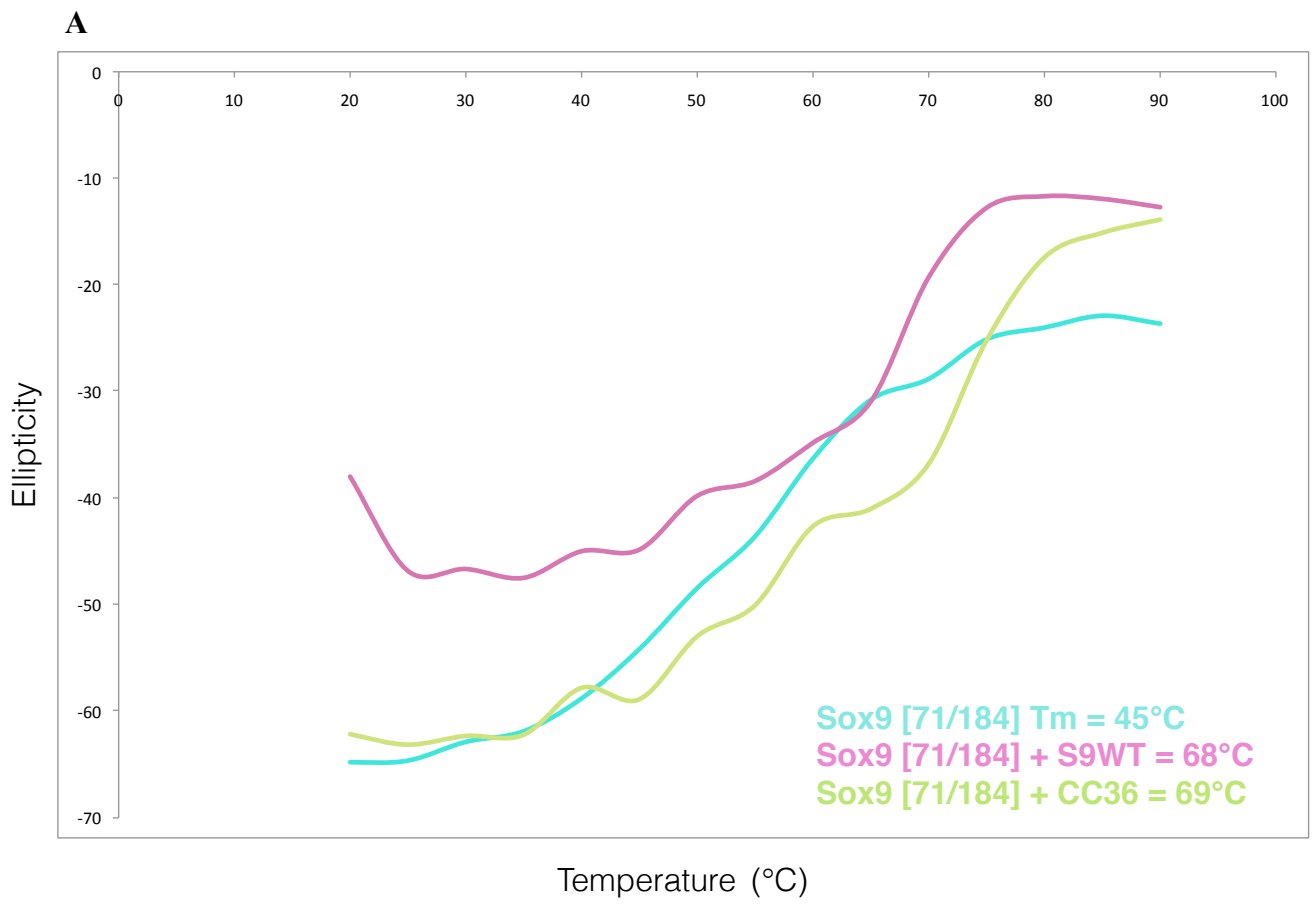


Figure 21: Melting of His_6-SOX9 wild type constructs free and bound to S9WT or CC36. (**A**) His_6-SOX9 [71/184] alone (blue) melted at 45°C and bound to S9WT (purple) melted at 68°C or bound to CC36 (green) melted at 69°C. (**B**) His_6-SOX9 [101/184] alone (blue) melted at 45°C and bound to S9WT (purple) melted at 65°C or bound to CC36 (green) melted at 65°C. Melting of both His_6-SOX9 wild type constructs at very similar temperatures is indicative that the thermal stability is not altered with the addition of the dimerization domain in the His_6-SOX9 [71/184] construct.

polarization at 10nM, 20nM, 50nM, and a larger spike at 100nM. This is similar to SOX9 [101/184] bound to FITC-S9WT whereby there was an initial spike at 10nM, 15nM, 40nM, and a greater spike at 100nM. However, the data was not consistent or accurate enough to determine a saturation curve and K_d values from.

3.5 ^{1}H-^{15}N-^{13}C HSQC Nuclear Magnetic Resonance (NMR) Spectroscopy of His_6-SOX9 [101/184]

HSQC data collected from the 700 MHz Bruker magnet for SOX9 [101/184] complexed with the single site S9WT required a program known as NMR Analysis in order to asses the peak data. As a result, approximately 56 of the 83 peaks were present. An imino test was performed to confirm the presence of DNA with 10 defined peaks as predicted. However, protein degradation hindered a complete spectrum from being constructed and assigned.

CHAPTER IV

DISCUSSION

4.1 The Binding Mechanism of SOX9

Cellular processes such as chondrogenesis and sex determination are regulated by the structural confirmations and complex formations of SOX9. Oligomer formation is one of the unique mechanisms employed by SOX9 proteins that differentiate its mode of specificity and subsequent function. Binding of both SOX9 [71/184] and SOX9 [101/184] with S9WT or CC36 provided an indication that the HMG domain of SOX9 is the only platform required to form a SOX9-DNA complex and retain the affinity for DNA. Using radioactive and non-radioactive electrophoretic mobility shift assays with wildtype SOX9 [71/184] and SOX9 [101/184] it is established that the presence of an intact dimerization domain enables the binding to DNA CC36 at both sites directly and subsequently demonstrating cooperative dimerization. This suggests that a stable complex is formed in the presence of an intact dimerization domain. More so, the reciprocal interaction of one dimerization region with another HMG domain on a tandem promoter is a fundamental attribute of a novel binding mechanism.

Furthermore, disruption of the two conserved alanine residues in the Sox E Family of proteins demonstrated that when substituted, the first of two Alanines in the HMG domain at A118E had no impact on the ability of the SOX9 [71/184] mutant to form a stable confirmation of binding to both sites on CC36 in a similar cooperative manner as the wildtype SOX9 [71/184]. SOX9 [71/184] A119E loss the ability to dimerize and cooperatively bind to the double-sited oligonucleotide CC36.

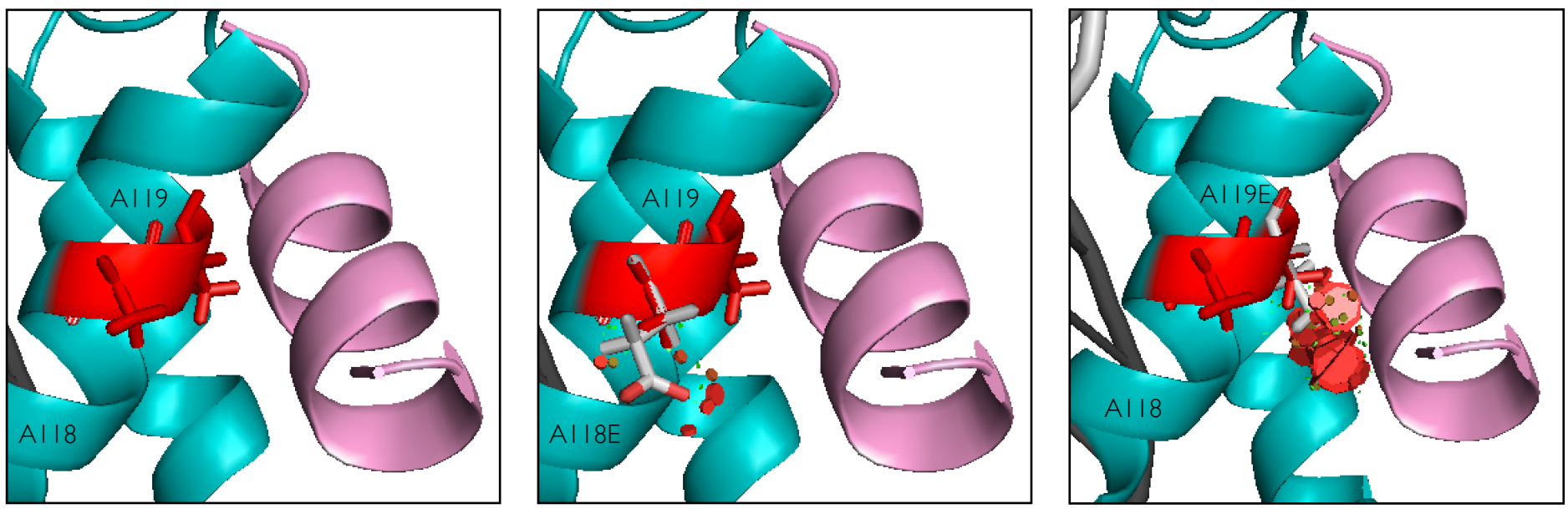
In addition, there is a loss of affinity between the mutant HMG domain and the dimerization domain. This suggests that the second Alanine residue plays a more critical role in exposing a greater binding surface related with the complex formed between two identical SOX9 [71/184] on a double site DNA. Both mutants still retain a surface that permits the HMG domain to bind with DNA. With respect to the two Leucines unique to the Sox E Group of proteins, the mutants L142Q and L145E, the SOX9 [71/184] L142Q mutant is present within the second alpha helix of the three that make up the HMG domain. Upon mutating the second alpha helix (α_2) with a single amino acid substitution SOX9 [71/184] L142Q the hydrophilic replacement reduces surface exposure losing the ability to form a dimer in a cooperative manner and affinity to bind to the dimerization domain. Unlike the SOX9 [71/184] L145E mutant that is present in the two amino acid linker region between the second and third alpha helices of the HMG domain, L145E does not affect the proteins ability to bind cooperatively or bind to the peptide of the dimerization domain. This suggests that the third alpha helix (α_3) of the HMG domain may have a less significant role and potentially not dock onto the alpha helix of the dimerization domain (α_0) unlike the first and second alpha helices that dock directly onto α_0 to form a stable complex.

A118 and L145 of the HMG domain are oriented away from the dimerization domain-docking groove such that the A118E or L145E substitution does not interfere with the binding surface. Replacing either Alanine or Leucine with a Glutamic acid would only propel the longer hydrophilic side chain to orient further out into the external environment as opposed to burying itself into the inner hydrophobic core.

Although A119 and L142 are also on the surface, its stable conformational state is aligned toward the docking groove. A mutation at A119E or L142Q could contribute a negative charge that generates steric hindrance between the HMG and dimerization domain surfaces that would otherwise make contact **(Figure 22)**.

Previous mutant studies conducted by Shokofeh Shahangian demonstrated that particular amino acids I73, A76, V77, V80, L81, and Y84 were critical for mediating dimerization which supports the data that the dimerization domain folds into an amphipathic helix post its commitment to binding (Bernard et al., 2003). This can also contribute to the notion that the thermal stability of the SOX9-DNA complex is not affected by the presence of the dimerization domain. When bound to CC36, SOX9 [71/184] with an intact dimerization domain exhibits cooperativity.

A



B

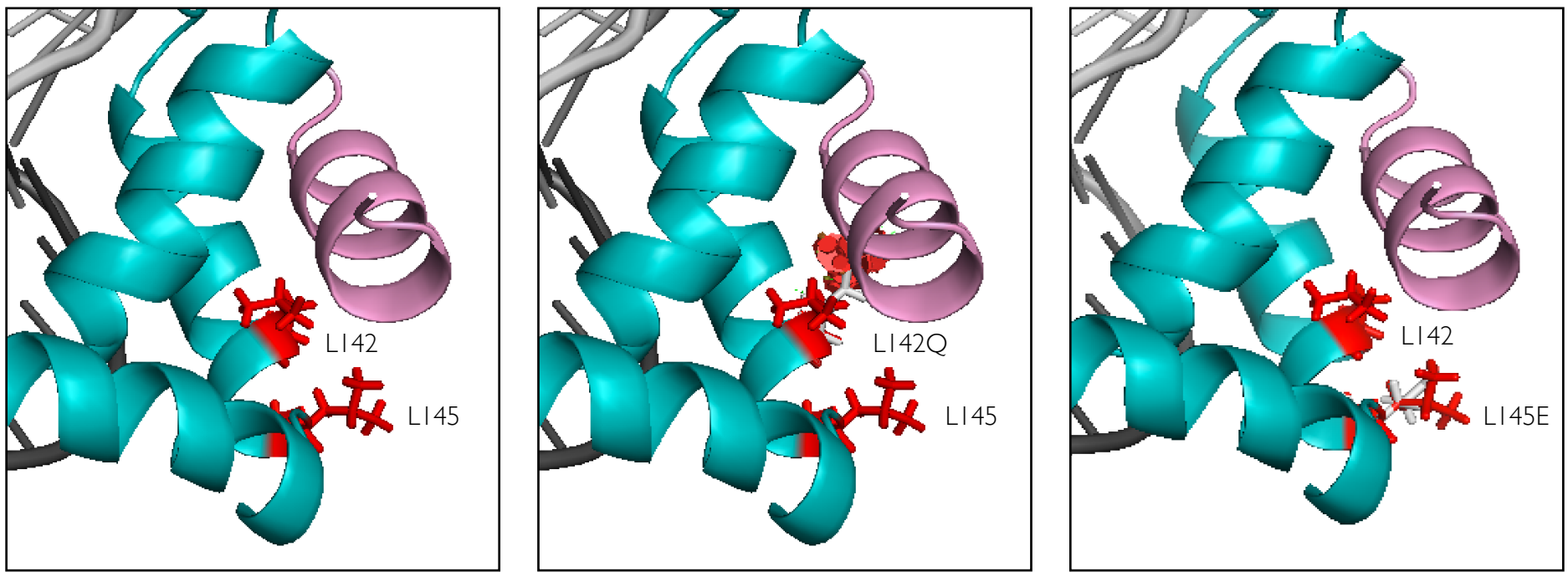


Figure 22: Point mutations via Pymol to the HMG domain of SOX9. **(A)** Representation of single point mutations at A118E and A119E in the first alpha helix of the HMG domain. A119E demonstrates greater steric hindrance than A118E **(B)** Single point mutations at L142Q and L145E in the second alpha helix of the HMG domain. L142Q demonstrates greater steric hindrance than L145E.

4.2 The Effect of the Dimerization Domain

SOX9 [71/184] complexed to S9WT or CC36 shows that the peptide can interfere with a preformed complex. The hydrophobic surface of the peptide assists in stabilizing the complex. The complex between SOX9 [101/184] with S9WT or CC36 demonstrates how the peptide i.e., the dimerization domain, can specifically recognize the HMG domain-DNA interface to generate another plane of structural and functional specificity **(Figure 23**. With the dimerization in trans, the HMG domain is the platform the peptide requires for binding. Amongst the Sox E family of proteins, the amphipathic helix of the dimerization domain can bind across the hydrophobic cleft of the HMG domain. The dimerization domain does not affect DNA binding given that electrophoretic mobility shifts are still detected even in the absence of the dimerization domain with SOX9 [101/184]. Given that there was no florescent signal of the peptide with DNA alone it suggests that the dimerization domain does not posses an affinity for either the single or double site DNA.

Given that A118, A119, L142, and L145 are largely exposed surfaces and hydrophobic as per the SOX9-DNA crystal structure, mutations to the HMG domain that substitute the two conserved alanines or the unique Leucines with a hydrophilic amino acid native to SRY or SOX18 does not hinder the ability of SOX9 to bind to DNA as demonstrated by the shifts stained with SYBR Green. Fluorescence shifts with the dimerization peptide revealed no indication of the peptide binding to the SOX9 A119E-DNA complex as well as the L142Q-DNA complex. This is consistent with the loss of dimerization demonstrated by the A119E and L142Q shift. Although the HMG retains the ability to bind to DNA and form a complex, the initial

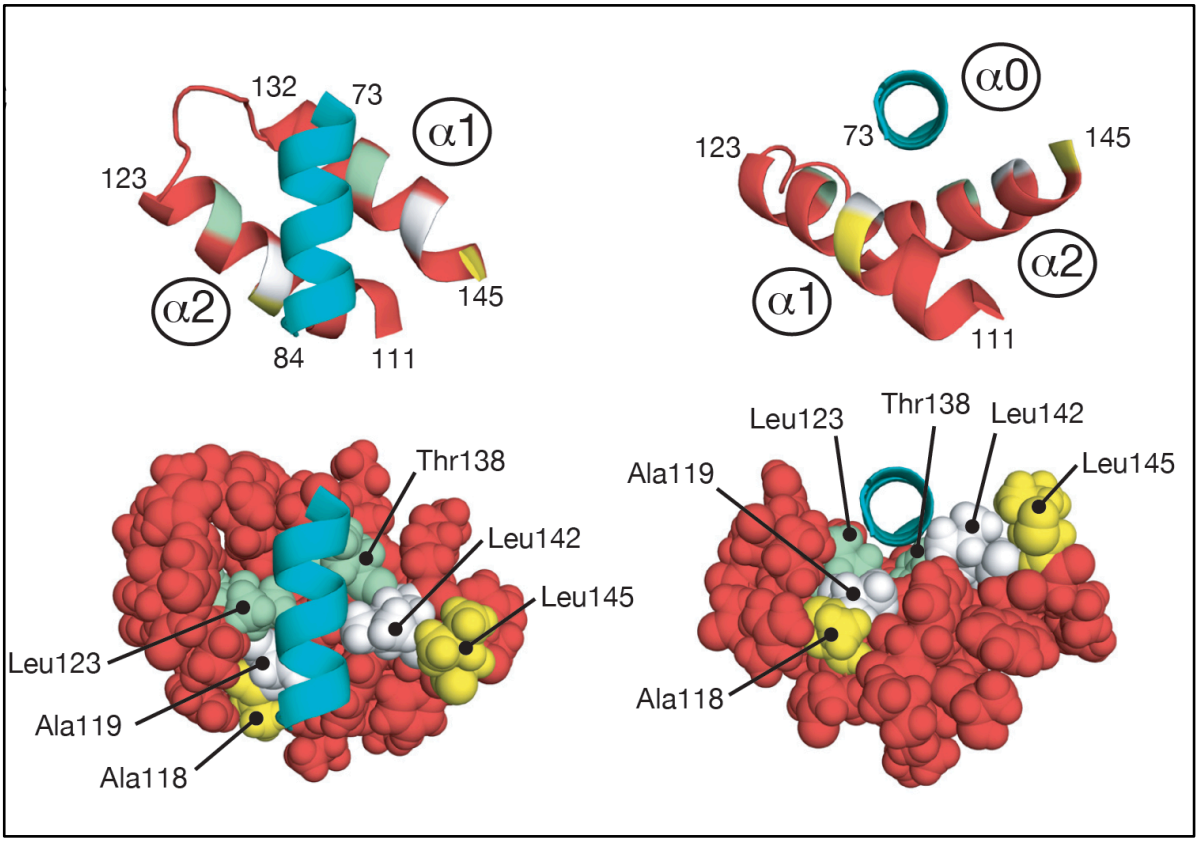


Figure 23: A docking simulation of how the alpha helix (α_0) of the dimerization domain of SOX9 would dock onto the first and second alpha helix (α_1, α_2) of its own HMG domain. The surface exposure of specific amino acids are also revealed including the di-alanine sites at A118 and A119 and two leucines at L142 and L145 that are unique to the Sox E Group of proteins (*Courtesy of* Dr. Logan Donaldson).

exposed surface that becomes mutated prevents the peptide from interfering and binding with the preformed complex. A118E and L145E maintain their ability to dimerize and as a result the peptide can obstruct and bind the preformed complex enabling detection of the peptide.

4.3 Characterization of the Dimerization and HMG Domain of SOX9 [71/184]

Peptide disassociation constants SOX9 [71/184], SOX9 [101/184], and SOX9 mutant A118E were calculated using pro Fit by Quantum Soft with the assistance of Dr. Logan Donaldson. The binding of the peptide on the complex of SOX9 [71/184] containing the dimerization domain with CC36 DNA employs a more sigmoidal-type progression unlike the SOX [101/184]-CC36 complex that demonstrates a more linear binding slope. This exact pattern is visible with the P-32 radioactive EMSAs between SOX9 [71/184] and SOX9 [101/184] complexed with CC36. As Coustry and colleagues demonstrated, the slower more sigmoidal progression of SOX9 [71/184] with a functional dimerization domain is indicative that binding of SOX9 to its recognition sites of the DNA is a more cooperative process (Coustry et al., 2010). This provides insight into the role of the dimerization domain in chromatin remodeling to initiate transcription and how the dimerization domain may elicit a stronger activation of the promoter or enhancer by running at a higher efficiency and having a greater output of transcripts (Coustry et al., 2010). ChiP assays have previously confirmed that an intact dimerization domain of SOX9 is essential for chromatin remodeling via Mnase digestion assays in order to activate transcription of the *Col2a1* enhancer containing a double-site

consensus and non-consensus sequence. Despite the non-functional dimerization domain deletion mutant interacting with the chromatin, in specific instances, the SOX9 mutant was unable to activate transcription like the intact dimerization domain wildtype SOX9 (Coustry et al., 2010).

Without the presence of a functional dimerization domain such as the SOX9 mutant or in the case of this study SOX9 [101/184], these particular constructs are unable efficiently disrupt the chromatin necessary for transcription once confronted by the nucleosomes on the template (Coustry et al., 2010). It is possible that the dimerization domain upstream of HMG domain on SOX9 recruits co-factors that link to RNA polymerase II and assist in unraveling the chromatin for the HMG domain to dock onto the DNA and initiate transcription. However, in the absence of the domain, interactions with chromatin are maintained and limited transcription of some reporters containing promoters that are targets of SOX9 can still be initiated (Coustry et al., 2010). The binding affinity of SOX9 [71/184], SOX9 [101/184], and SOX9 mutant A118E with either DNA did not differ in the order of magnitude for the K_d or the Hill coefficient for peptide binding with the respective plotting of the normalized intensities.

The positive cooperativity was approximately 3 for all constructs with the exception of SOX9 [71/184] A119E. This compares to the strong cooperativity between oxygen and hemoglobin (Stefan and Le Novère, 2013). Furthermore, given that the peptide was unable to bind to the A119E mutant, a disassociation constant was unable to

be calculated. The K_d for the peptide binding to the various SOX9-DNA preformed complexes were all within the same boundaries of 10μM. The A118E mutant has a greater negative hydrophilic side-chain oriented away from the hydrophobic core that could alter the binding surface for the peptide to bind to the preformed SOX9-DNA complex. Likewise, the lack of binding of the dimerization domain peptide to the mutated HMG domain in SOX9 [71/184] A119E suggests that the original Alanine at position 119 of the HMG domain is an essential residue of the HMG binding pocket and plays a significant role in creating an affinity interface between the dimerization domain and the HMG domain **(Figure 23)**. The peptide did not have any platform within the HMG domain of the A119E mutant to bind to the protein. Steric hindrance could alter the reactivity pattern of the SOX9-DNA whereby the HMG domain binding pocket becomes less surface-exposed and prevent a docking site for the peptide. Both SOX9 [71/184] and SOX9 [101/184] enable peptide binding providing that SOX9 [71/184] with the intact dimerization domain and HMG domain can create a trans-dimerization model that enables the peptide to intervene with the pre-formed complex while SOX9 [101/184] containing just the HMG domain provides a surface exposed groove for the peptide to dock onto.

The HMG domain of SOX family is structured and maintains conserved folding. Conserved amino acids within the Sox E group family include Q117, A118, A119, T138, and L142 (Schlierf et al., 2002). The structure of the HMG domain consists of three alpha helices that are oriented in an L-shaped arrangement. The mutation study of A117V of SOX10 that is equivalent to A118 of SOX9, disrupts cooperative binding (Schlierf et al., 2002). This study is indicative of an interaction between the

helices of the HMG domain and the dimerization region potentially due to the conservation of amino acids (Goji et al., 1998; Schlierf et al., 2002). The mutational study of A119E demonstrated a monomeric band shift with SOX [71/184] on DNA as opposed to the wildtype band shifts. In addition, the HMG is defined as having two wings; the major wing consisting of alpha helices 1 and 2 and the minor wing defined as helix 3 (Cary et al., 2001). Both wings are thought to work independent of each other and although both wings bind to DNA it is the major wing that is flexible enough enabling it to have the potential to be the working wing associated with protein interactions (Cary et al., 2001). Likewise, SOX9 [71/184] may have similar interaction fate between helices 1 and 2 as SOX10.

This study puts forth that stable binding of the peptide in the ternary complex involves peptide-protein contacts with the HMG domain but not the dimerization domain of the SOX9 protein. Such binding does not involve peptide-DNA contacts given that there is no detection of peptide with DNA alone (Chasman et al., 1999). Binding of the dimerization domain peptide can stimulate cooperative trans-dimerization with SOX9 [101/184] whereby the complementary interaction of one dimerization region with another HMG domain on a tandem promoter is a fundamental attribute of the binding mechanism. This could be the model required for processes like chondrogenesis providing a point of control for tissue-specific regulation of gene expression by a SOX transcription factor. The dimerization domain may confer additional stimulatory effects on transcription by interfering with the transactivation of enhancers such as *Col9a2* and *Col11a2* involved in chondrocyte differentiation (Bernard et al., 2003).

4.4 Predicted SOX9 [71/184] Model and Validation

The long linker located between the C-terminus of the dimerization domain and N-terminus of the HMG domain facilitates the orientation between both domains. This is suggestive of an alternative novel trans-dimerization model **(Figure 24)** of SOX9 [71/184] homodimers whereby the dimerization domain of one protein docks onto the HMG domain of the second protein. The docking of the hydrophobic face of the dimerization domain helix to the HMG domain is optimal because it provides a favorable position for the hydrophobic amino acids to orient themselves to preferably minimize surface exposure. A119E mutant is known to contribute to Campomelic Dysplasia and although it retains the ability to bind to DNA, it does not maintain cooperativity in parallel to the wildtype SOX9 [71/184]. This is suggestive that the dimerization domain helix may also shield the residues exclusive to the Sox E group of proteins. This model could be applied to SOX10 such that there is an increase in affinity for the non-consensus binding site after the binding of the first SOX protein to the consensus sequence (Bernard et al., 2003; Sock, 2003). Non-consensus binding of SOX10 is initiated by the binding of a single SOX10 docking on the consensus binding site which triggers the folding of the dimerization such that it will then bind to the HMG domain of a second identical SOX10 protein (Schlierf et al., 2002). The second SOX protein would have a higher affinity for binding to a non-consensus sequence that will be followed by the folding of the dimerization domain of the second protein and binding to the HMG domain of the first protein to establish a stable complex in trans increasing the stability of the proteins as a complex (Schlierf et al., 2002). The trans-dimerization model is only applicable when the two binding sites are not equal.

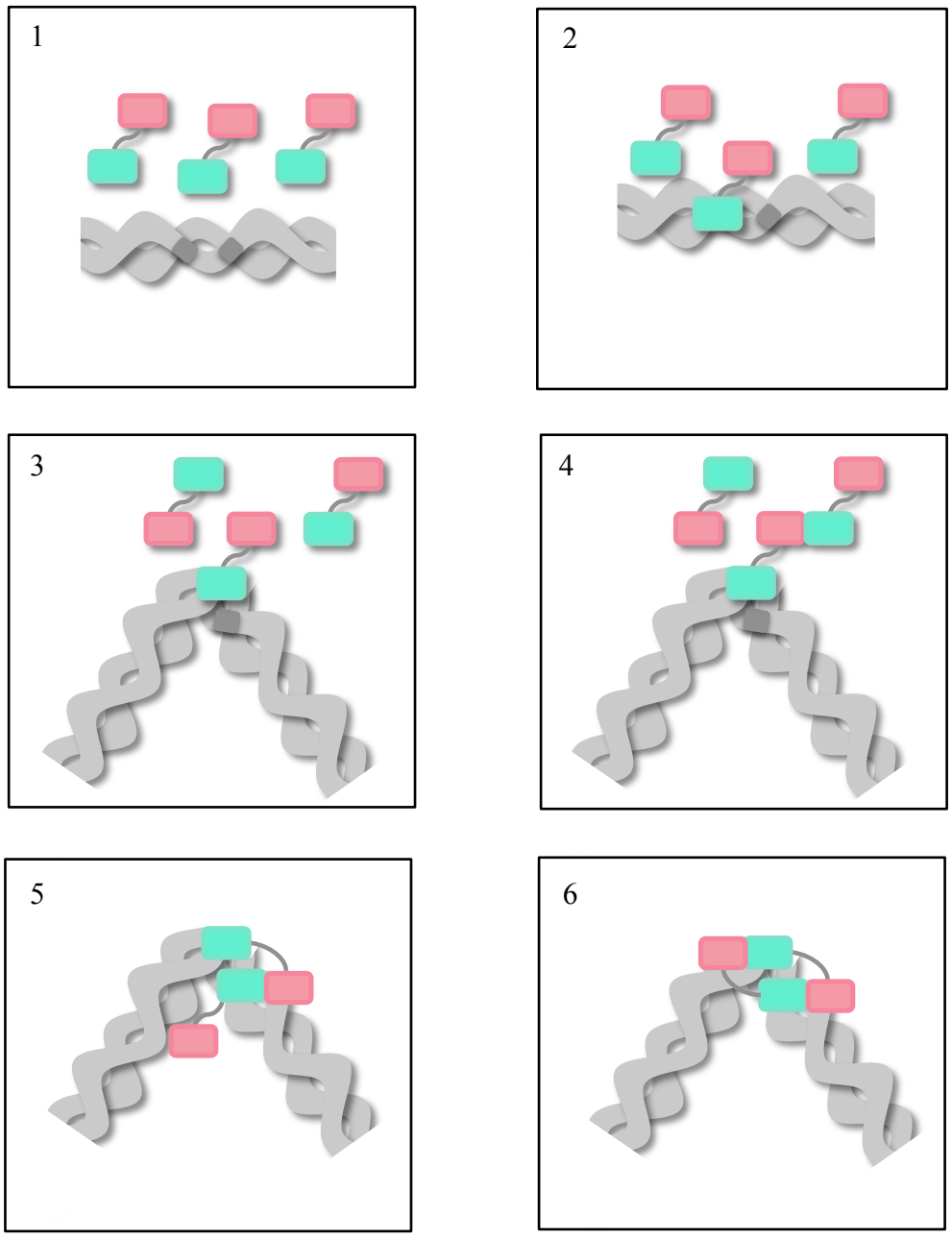


Figure 24: The trans-dimerization model. SOX9 [71/184] is represented by its HMG domain (green) and dimerization domain (pink) attached by a linker region with a double site (dark grey) DNA (light grey). The model infers that the HMG domain of the first SOX9 [71/184] binds to the consensus DNA site and induces a bend in the DNA. The HMG domain of a second identical SOX9 [71/184] docks onto the dimerization domain of the first SOX9 [71/184] protein which is followed by the HMG domain of the second SOX9 [71/184] binding to the second non-consensus site on the DNA. This enables the dimerization domain of the second SOX9 [71/184] to bind to the HMG domain of the first SOX9 [71/184].

For example, tandemly inverted promoters that demonstrate unequal binding sites due to a difference in affinity similar to the Po promoter binding to SOX10 (Peirano and Wegner, 2000). Furthermore, the difference in sequences of SOX9 binding sites for example between *Col11a2* and *Col2a1* can affect concentration-dependent binding and as a result influence binding affinities between DNA, SOX9, and the recruitment of multiple transcription factors that SOX9 supports and in turn supports SOX9 affecting the accessibility of chromatin sites (Coustry et al., 2010).

The dimerization domain helix of SOX9 has a hydrophobic face. Those hydrophobic amino acids are oriented such that surface exposure is minimized. As per the structural model, the positioning of the dimerization domain over the HMG domain will decrease the surface exposure as well as mask the hydrophobic residues **(Figure 25A & 25B)**. Upon binding SOX9 will bind to the minor groove of DNA and cause it to bend 108° toward the major groove. The bending of DNA may contribute to the regulatory functions of SOX9 **(Figure 25C)**.

4.5 The Importance of Trans Binding of SOX proteins to DNA

The trans-dimerization hypothesis emerged from a study that switched helices α 1/α 2/α 3 from the HMG domain of SOX10 with those from SOX11, a Group C family member that does not dimerize (Schlierf et al., 2002). From this study, it appeared the α 1/α 2 contained an important determinant of dimerization as they could not be swapped with their corresponding helices from SOX11. The study suggests that the dimerization domain and the Sox10-specific HMG domain have to be present on the same Sox10

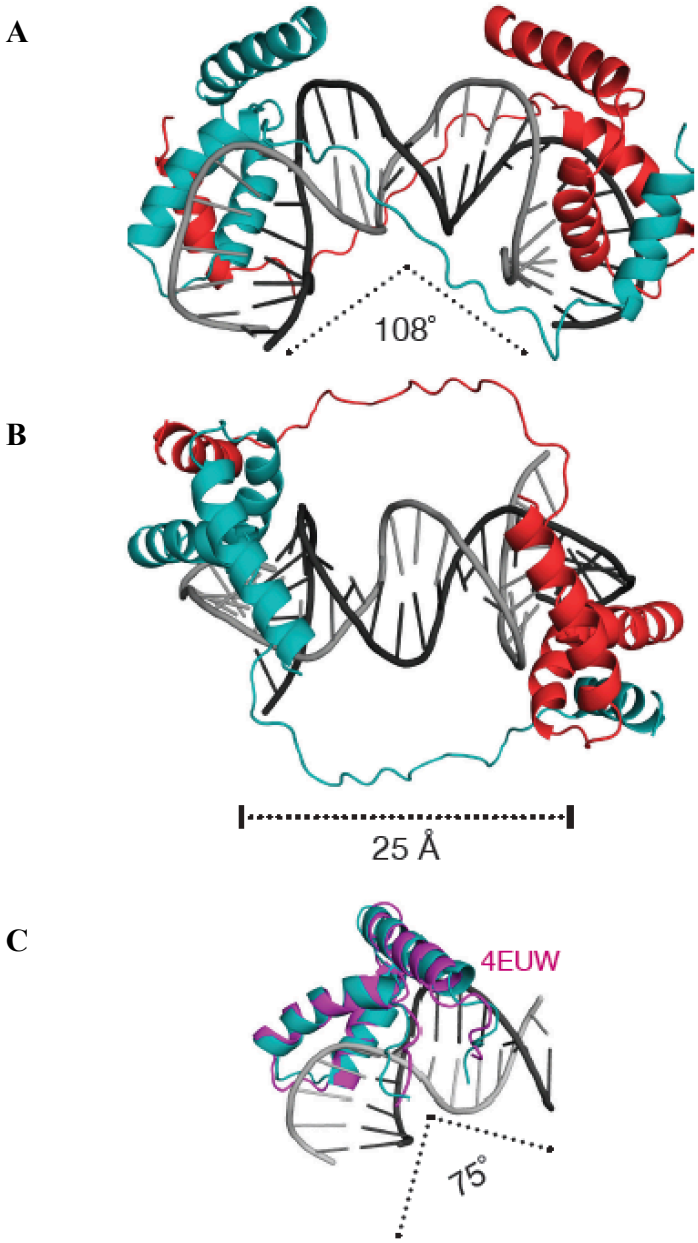


Figure 25: (**A**, **B**) A molecular model rotated at two angles of a dimerized SOX9 protein complexed with a double-site DNA. The HMG bound to DNA bends the DNA such that the dimerization domain of one SOX9 protein binds to the HMG domain of a second identical protein. The HMG domain of the second identical SOX9 is also bound to a second site on DNA. The two sites on the DNA are palindromic sequences that are characteristic of the high affinity sites on DNA for the SOX family of proteins. (**C**) The model demonstrates how the HMG domain of SOX9 is capable of bending DNA at a single site (PDB: **4EUW**). (*Courtesy of* Dr. Logan Donaldson).

molecule for intramolecular function of both domains with the potential for two HMG domains and two dimerization domains to interact with each other in a trans-binding manner between two Sox10 molecules (Schlierf et al., 2002). The mixture of binding complexes could be a direct result of the difference in endogenous sources of Sox9 under different physiological conditions such as incubation time. Binding to DNA differs between cis and trans depending on the incubation period (Lippard and Hoeschele, 1979). Additionally, orientation of the transcription factor can contribute to complex formation. Both the cis and trans complex formation can contribute to marked changes in the expression of Sox9 and isoforms within tissues over the course of development in the embryo. Each complex structure can play a role in the response patterns and trans-activating specific response elements that may have a distinct function or distinct gene products associated with its activation and subsequent function.

4.6 The Clinical Affect of Improper or Lack of SOX9-DNA Complex Formation

The lack of binding of SOX9 to DNA or improper formation of the protein-DNA complex leads to drastic affects on early embryonic development in terms of both sex determination and cartilage formation. During testis development SOX9 increases in expression in order to maintain a positive feedback loop to influence to activation of co-factors including growth factors as well as many other genes. In addition, SOX9

auto-regulates itself to maintain a threshold expression level to inhibit factors such as R-spondin-1 that is necessary to inhibit ovary formation. Without the binding of SOX9 to a sex-determination specific promoter or enhancer, the downstream cascade of events necessary to maintain the male sex determination cannot be activated and the propensity for a female-sex reversal phenotype becomes highly probable.

Improper or the lack of SOX9-DNA complex formation can lead to defects in chondrogenesis. Specifically SOX9 is a regulator of the type II collagen gene. A mutation to the dimerization domain deeming it nonfunctional can reduce or inhibit the activation of chondrocyte-specific reporters that are vital to initiate the process of chondrogenesis during early embryonic development. Misexpression of SOX9 results in ectopic cartilage formation of the limbs as well as alter the aggregation properties of limb mesenchymal cells such that cell condensation is reduced or lost ultimately affecting early skeletogenesis (Healy et al., 1999). Additionally, SOX9 plays a critical role in the cartilage differentiation programme in dermomyotomal cells that generate axial musculature and dermis (Healy et al., 1999). Such misimpression of SOX9 can inhibit the process of such cells committing to its normal cell fate by affecting the regulation of Pax1, a marker for ventral development downstream (Healy et al., 1999).

Alternatively, formation of the SOX9-DNA complex at an unsuitable time can have negative implications on development. SOX9 is a negative regulator of bone marrow formation, cartilage vascularization and endochondral ossification (Hattori et al., 2010). Complex formation can affect the ability of proliferating chondrocytes to further

differentiate into hypertrophic chondrocytes and eventually calcified cartilage (Akiyama et al., 2002). Its down regulation may be necessary to initiate cartilage-to-bone transition in the growth plate (Hattori et al., 2010). Under the control of the *BAC-Col10a1* promoter, transgenic newborn mice demonstrate vascular invasion into hypertrophic cartilage and impaired cartilage resorption delaying the formation of endochondral bone and resulting in reduced bone growth (Hattori et al., 2010). The inhibition of terminal differentiation is a result of Vegfa, Mmp13, RANKL, and osteopontin expression. Suppression is a result of SOX9 binding to SRY sites in the Vegfa gene (Hattori et al., 2010). Overall, SOX9 acts upon several target genes. Enhancers with varying affinities for SOX9 contribute to differential regulation of target genes (Mertin et al., 1999).

CHAPTER V

SUMMARY

5.1 Conclusion

Following a trans-dimerization DNA-SOX9 model, the hydrophobic platform defined by the alpha helices α_1 and α_2 of the HMG domain provides a favorable binding surface for the acceptance of the single alpha helix of the dimerization domain (α_0) forming a three-helix bundle. Cooperative binding is mediated such that the binding of one SOX9 protein to the consensus-binding site facilitates binding to the non-consensus site on DNA. In addition, the DNA binding affinity is not affected by mutations to the unique di-alanine cluster or two leucines present in the HMG domain. The thermal stability of the HMG domain as well as DNA binding is not improved or diminished by the presence of a dimerization domain. Dimerization at tandemly inverted promoters with unequal binding sites such as C/C' utilizes the trans-dimerization recruitment to serve as a cogent method of binding. Dimerization is probable once the stronger consensus site is bound to a SOX9 protein and can then recruit a second SOX9 to the weaker non-consensus site. This differs from two equal binding sites whereby dimerization can equally proceed via either site (Peirano and Wegner, 2000). This suggests that one dimerization region with another HMG domain on a tandem promoter is a fundamental attribute of the binding mechanism**.**

5.2 Future Works

NMR of SOX9 [101/184] with S9WT would provide visualization of the minimal binding of the HMG domain to a single site DNA providing the most resolved spectra. This should differ from SOX9 [71/184] with S9WT given the affinity for its own dimerization domain to bind to its own HMG domain. If an NMR sample becomes hindered by degradation a more suitable alternative would be to crystallize the complex under various conditions and determine the structure via X-ray diffraction analysis. Crystallization was previously attempted with SOX9 [71/164] and CC36 however, the crystal did not diffract. Alternatively, a SOX9-SOX5/6 co-complex on a bigger DNA may be more stable for crystallization. In addition, to assist with determining the ratio of cis and trans complexes within the mixture further cell work would need to be conducted. This includes incubation assays to determine the dependency of the type of complex formation on temperature and time of incubation. *In vitro*, this study could be affected by incubation time and temperatures such that the on-rate of binding of SOX9 almost exclusively as a dimer affecting cooperation and the stability of the SOX9-DNA complex can reach a maximum that varies between one minute and two hours (Coustry et al., 2010). Fluorescence anisotropy could be used to study the dynamics of folding between the unbound, partially bound, and bound states given that the fluorophore is bound to a relatively large molecule such as the SOX9 molecule itself. Concurrently, kinetics and sedimentation studies would reveal the reaction mixture (Lippard and Hoeschele, 1979).

APPENDIX A:
A STRUCTURAL MODEL OF THE RNA BINDING SITE IN THE C-TERMINAL OF THE HUMAN LA PROTEIN

6.1 ABSTRACT

The human La protein (hLa) is a RNA-binding protein with diverse roles in viral and cellular RNA metabolism. The RNA chaperone activity of hLa assists with the folding of RNA into secondary structures. Folding assists in functional structures favorable for internal ribosome entry site (IRES)-mediated translation initiation by assisting the correct start site usage. The C-terminal of hLa is highly unstructured and upon binding to RNA and initiating a conformational change is predicted to form an alpha helix in the short basic motif (SBM) of hLa. Transitions in secondary structure are studied using nuclear magnetic resonance (NMR) titrations between the unbound protein and its bound state to a 12-nucletoide trailer RNA; a small subsection of pre-tRNA that hLa is known to bind to. In addition, numerous NMR titrations were conducted to determine the extent to which hLa binds to an array of RNAs.

6.2 THESIS OBJECTIVE & OVERVIEW

Given the structural similarities between Larp7 and hLa it begs the question whether bound RNA can induce a confirmation change and formation of an extended third alpha helix in hLa in a similar manner to the complex formed between Larp7 and RNA **(Figure 26B)** (Singh et al., 2013). The array of RNAs includes a 12-nucleotide trailer, A20, 5'OH double stranded and single stranded RNA, and triphosphate RNA. The 12-nucleotide trailer RNA is part of a larger pre-tRNA that hLa is known to bind to.

More importantly, there could be two essential binding sites in the C-terminal constructs of hLa working in a cooperative manner. This includes the initial recognition and binding affinity of the RNA recognition motif 2 (RRM2) for RNA as well as the presence of the SBM to then lock in and form a final low energy conformational state of the RNA-hLa complex. The affinity of the SBM for RNA is only increased once the RRM2 has recognized the RNA first. The approach is to use Nuclear Magnetic Resonance (NMR) spectroscopy to titrate a stable hLa construct with the 12-nucleotide trailer. This is a result of previous circular dichroism data that suggests that complexing with this particular RNA yields the lowest energy structure and potentially most stable complex. The objective is to determine if the SBM becomes structured and forms an extended third alpha helix upon binding to RNA. It is predicted that the RNA binds into a groove made between the third alpha helix and the rest of the RRM2. The structure of the hLa [225/334] construct has been predetermined by Jacks *et al 2003* and the re-assignments can assist in the determining novel assignment post-C-terminal 334 on the CTD of hLa (Jacks et al., 2003; Sanfelice et al., 2004). The constructs for the hLa **(Figure 27B)** in this study begin at amino acid 225 and end at various amino acids including 332, 339, 344, 363, 375, and 408 at the C-terminal. It has been demonstrated in preliminary data that the SBM is important for binding.

6.3 LITERATURE REVIEW: Human La (hLa): A RNA-Binding Protein

La was first identified as an autoantigen in Sjorgens Syndrome B and Lupus Erthymatosus (Kuehnert et al., 2015). Human La (hLa) exhibits numerous roles in viral and cellular RNA metabolism (Kuehnert et al., 2015). These processes include stabilizing mRNA, mRNA translation via interaction between its RNA-binding surfaces and mRNA,

processing of RNA polymerase III transcripts, and micro RNA processing (Dong et al., 2004; Kuehnert et al., 2015; Teplova et al., 2006). It is a ubiquitous RNA-binding protein that has chaperone activity in eukaryotic cells (Phizicky and Hopper, 2010; Wolin and Cedervall, 2002). It is an essential protein in mice but can be knocked out in yeast. It binds to a variety of RNA transcripts such as pre-tRNAs and RNA Polymerase II transcripts (Kotik-Kogan et al., 2008). As a RNA chaperone, La assists with non-covalent folding or unfolding of RNA (Herschlag, 1995). hLa binds and assists RNA into folding into a functional secondary structure (Kuehnert et al., 2015). This is favorable for translation initiation assisting with the correct start site usage as well as IRES-mediated translation given that mutants lacking the C-terminal region are unable to promote poliovirus IRES-mediated translation (Svitkin et al., 1994). It contains an atypical RNA Recognition Motif (RRM) consisting of five beta sheets and three alpha helices that directly recognizes RNA (Singh et al., 2013). In addition, the highly disordered C-terminal of hLa enables it to assist in the biogenesis of RNA precursors (Kucera et al., 2011). In parallel, the C-terminal of hLa becomes structurally organized upon binding to RNA as a result of its basic amino acids being involved in RNA interactions (Goodier et al., 1997). It prevents newly synthesized RNA from forming non-functional structures and prevents exonuclease digestion (Alfano et al., 2004; Wolin and Cedervall, 2002). It is ATP independent and maintains a high degree of disorder even in its lowest energy structural confirmation.

The study of hLa as an RNA-binding protein is based on the entropy transfer model that describes the role of structural disorder in chaperone function **(Figure 26A)**. A disordered chaperone binds to a partially mis-folded RNA or protein substrate non-specifically (Tompa and Csermely, 2004). The recognition motif anchors the chaperone

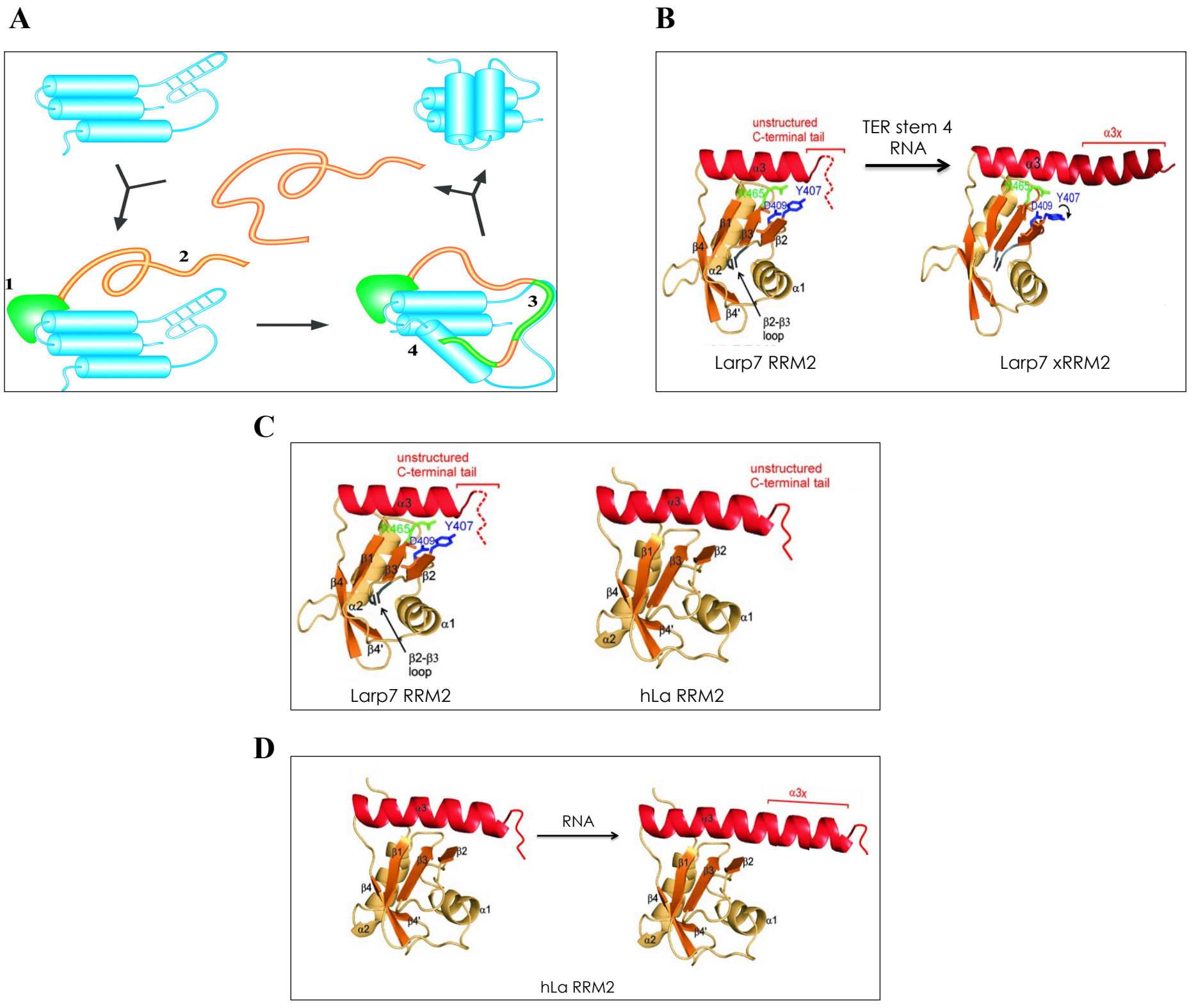


Figure 26: (**A**) The entropy transfer model describes the role of structural disorder in chaperone function. First, a disordered chaperone (orange) binds to a partially mis-folded RNA or protein substrate (blue) nonspecifically. The recognition motif (green) anchors the chaperone to the substrate. Secondly, a disordered region of the chaperone projects outward and eventually makes contact with the disordered region of the substrate. The chaperone becomes ordered while the substrate unfolds due to entropy transfer. The unfolded substrate creates flexibility in itself as a means to find a native or lower energy confirmation by maintaining close proximity to the folded chaperone. It is possible that the chaperone has multiple contact and release in rapid time points until it finally releases a folded substrate and can resume the entire catalytic processes again. (*Adapted from* Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. Faseb J. *18*, 1169–1175.) (**B**) Structure of Larp7 RRM2 with an unstructured C-terminal tail prior to the binding of a TER stem 4 RNA to the RRM2. Upon binding a confirmation change in the structure of the protein-RNA complex leads to the formation of an alpha helix C-terminal to the third alpha helix. (**C**) The structural similarities between Larp7's RRM2 and hLA's RRM2 include the four beta sheets and three alpha helices as well as the unstructured C-terminal tail at the end of the third alpha helix. (**D**) The change in structural confirmation of hLA's RRM2 is a result of RNA binding. The unstructured C-terminal tail becomes an alpha helix due to entropy transfer. (*Adapted from* Singh, M., Choi, C.P., and Feigon, J. (2013). xRRM: a new class of RRM found in the telomerase La family protein p65. RNA Biol *10*, 353–359.)

to the substrate (Naeeni et al., 2012). Secondly, a disordered region of the chaperone projects outward and eventually makes contact with the disordered region of the substrate (Tompa and Csermely, 2004). The chaperone becomes ordered while the substrate unfolds due to entropy transfer (Tompa and Csermely, 2004). The unfolded substrate creates flexibility in itself as a means to find a native or lower energy confirmation by maintaining close proximity to the folded chaperone (Tompa and Csermely, 2004). It is possible that the chaperone has multiple contact and release in rapid time points until it finally releases a folded substrate and can resume the entire catalytic processes again (Tompa and Csermely, 2004).

The relationship between p65; a member of the Larp7 family, and hLa are similar in their CTD structures but with less conserved sequences in their CTD than NTD enabling the structures to differ between the two proteins. In the CTD, they both contain five beta sheets and three alpha helices in a βαββαβ'βα fold including the unstructured C-terminal tail at the end of the third alpha helix **(Figure 26C)** (Singh et al., 2013). This differs from a canonical RRM that consists of four beta sheets and two alpha helices in a βαββαβ fold (Singh et al., 2013). An unstructured C-terminal tail in Larp7 forms a structured alpha helix C-terminal to the third alpha helix upon binding of a telomerase RNA (TER) Stem 4 RNA to the xRRM2 (Singh et al., 2013). The xRRM2 is an atypical motif with one extra alpha helix and beta sheet in comparison to the standard RRM (Singh et al., 2013). Analysis of p65 revealed a novel-binding site in the xRRM2 of hLa that has affinity for RNA. The RRM2 of p65 is atypical with a disordered tail. Upon binding to RNA, p65 becomes ordered for a hierarchy of assembly between of telomerase RNA (Singh et al., 2012).

The crystal structure of p65-C with Stem IV of TER illustrates that when bound to RNA, the C-terminal tail is now structured with an alpha helix extension into the third alpha helix (α3) (Singh et al., 2012). This extension binds to the major groove of the S4 of TER opening two C-G base pairs on the either side of G121 and A122 that results in a 105° bend of the RNA. The GA bulge protrudes out and interacts with residues on the RRM (Singh et al., 2012). The high-affinity binding requires both the atypical RRM and the C-terminal tail (Singh et al., 2012). This provides insight about the importance of the loop prior to the last alpha helix in the C-terminal and if it behaves in a similar manner to the shift in p65 (Singh et al., 2013). Sam Sharifi from Dr. Mark Bayfield's lab conducted radioactive electro mobility shift assays demonstrating that increasing the length of the C-terminal of hLa constructs enables shifting to a RNA bound state at lower hLa concentrations (i.e. a lower K_d). This suggests that the binding constant will decrease as the hLa constructs increase in their C-terminal amino acids. In addition, K316A and K317A mutations to the alpha helix demonstrate a loss of function. This provides insight into the importance of this final alpha helix to make contact and how mutations to this specific region may abolish the structure of the alpha helix.

The structure of the human La (hLa) protein contains 408 amino acids in length and weighs 47 kDa. hLa contains three RNA binding surfaces **(Figure 27A)** including the La motif (LAM), RNA recognition motifs (RRM) 1 (RRM1) and a non-canonical RRM2 (Kuehnert et al., 2015). It consists of the N-terminal Domain (NTD) containing the LAM and RRM1 making up the La Module. The C-terminal Domain (CTD) contains the RRM2, nuclear retention signal (NRE), short basic motif (SBM), and nuclear localization signal (NLS) (Horke et al., 2004; Wolin and Cedervall, 2002). All three RNA-binding

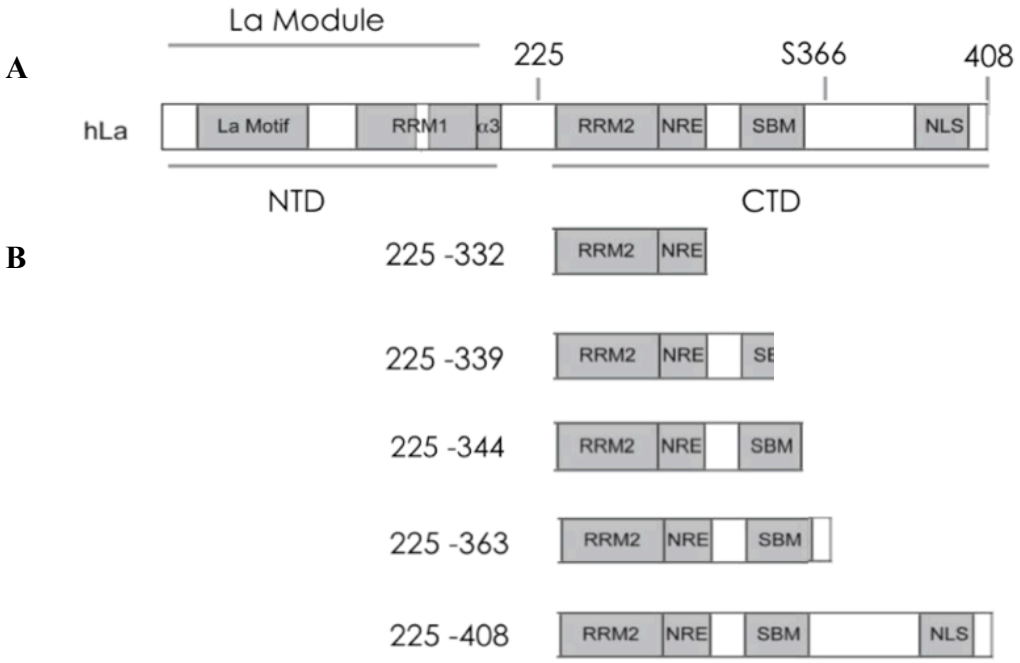


Figure 27: (**A**) Structure of the human La (hLa) protein. The N-terminal domain (NTD) consists of the La module that harbors the La Motif and the RNA Recognition Motif 1 (RRM1). The C-terminal domain of interest beginning at amino acid 225 consists of RRM2, the Nuclear Retention signal (NRE), the Short Basic Motif (SBM), and Nuclear Localization Signal (NLS). In the CTD, Serine 366 is a site for phosphorylation. (**B**) The various CTD constructs of hLa that are tested as part of the study, along with their conserved domains. (*Courtesy of* Sam Sharifi)

surfaces of hLa interact with the Hepatitis C virus and is required for the virus to translate via an internal ribosome entry site within independent regions (Ali et al., 2000; Martino et al., 2012; Pudi et al., 2003). However, the RRM1 and RRM2 cooperate to internally recognize the RNA sequences derived from the Hepatitis B virus (Alfano et al., 2004; Horke et al., 2002). In the CTD, Serine 366 **(Figure 27A)** is a site for phosphorylation (Fan et al., 1997; Schwartz et al., 2004). When hLa is present in the nucleus it gets phosphorylated by Casein Kinase-2 (CK2), however this phosphorylation does not occur in the cytoplasm (Schwartz et al., 2004). In hLa, Serine 366 lies in a region with basic amino acids N-terminal to it and acidic amino C-terminal to 366 (Kuehnert et al., 2015).

6.4 EXPERIMENTAL METHODS: Protein Expression and Purification of His_6-hLa [225/332], His_6-hLa [225/339], His_6-hLa [225/344], His_6-hLa [225/363], and His_6-hLa [225/408] for NMR

E.coli BL21 (DE3) cells expressing all His_6-hLa constructs were grown at 37°C shaking at 200rpm in 3 Ls of M9 minimal media supplemented with 1 g/L [^{15}N] ammonium chloride as a means of being the sole nitrogen source. The media was supplemented with 50 μg/mL of Kanamycin in a 2.8 liter Allen flask. Cells were grown to an exponential phase reaching an optical density (OD_{600}) of ~0.8-1.0. Cells were induced with 1.0mM IPTG (Wisnet Inc) and the temperature was reduced to 16°C at 200rpm for a 16-18 hour post-induction period for soluble protein production. Following the post-induction period, cell pellets were collected by centrifugation at 7000rpm for 15 minutes at 4°C and placed in the -20°C freezer for a minimum of one hour. Cells were re-suspended over ice in 25mL of T300 supplemented with 100μL of Thermo Scientific Halt™ Protease Inhibitor Cocktail (100X). The re-suspended cells were lysed via three

passes of compression at 15000psi using the French Cell Press. The cell lysate was kept over ice until centrifugation at 17000rpm for 30 minutes at 4°C using a Beckman centrifuge. The soluble extract was then increased to pH 7.0 using Tris-HCl pH 7.5 and purified over a 10mL slurry of NI-NTA resin (Qiagen) column primed in T300 at 2ml/minute. This was followed by a 100mL wash of 2M urea supplemented with T300 at 3ml/minute provided that the lysate was added over the beads. The column was further washed with 150mL of 10mM and 40mM imidazole supplemented with T300 at 3ml/minute. The his-tagged hLa proteins were eluted with 250 mM imidazole enriched with T300 at a rate of 2mL/minute following priming the column with 250 mM imidazole for approximately 20 seconds. After collecting approximately 80mL of elution, 1mM PMSF was added.

The elution was concentrated down via ultrafiltration using a Sartorius 5000 MWCO filter to 5mL spinning 5600rpm (Eppendorf Centrifuge 5430 R) at 4° C. The elution was further purified using the AKTA FPLC gel filtration using a Sephacryl S100 16/60 column (GE Biosciences) equilibrated with T300 supplemented with NP-40 and 0.2% NaN_3. Fractions pertaining to the correct molecular weight of the protein of interest were then confirmed at the correct molecular weight on a 12% SDS gel using a protein ladder (Biorad) and visualized with Fast SeeBand Single Step Protein staining solution (FroggaBio) on the Alpha Imager HP system (Alpha Innotech). Verified fractions were pooled and concentrated by ultrafiltration using a Sartorius 5000 MWCO. The molar concentrations for all SOX9 constructs were calculated from the A_{280} using the extinction coefficient calculated using the ExPASy ProtPram tool (http://web.expasy.org/protparam/) using

the confirmed amino acid sequence. Pure protein was concentrated to 0.2-0.3M using a MWCO 5000 and supplemented with 10% D_2O for unbound protein. In addition, a 1:1 ratio at 0.6 mM of a 12-nucleotide RNA trailer to hLa protein was constructed to collect bound HSQC data.

6.5 RESULTS: Stability of hLa Constructs

The expressions of hLa [225/332], hLa [225/339], hLa [225/344], hLa[225/363], hLa [225/408] were all confirmed with their correct molecular weight at 13.2kDa, 14kDa, 14.5kDa, 15.4kDa, and 20.6kDa. Expressions of all hLa constructs **(Figure 28)** were more limited in N15 media than LB. All constructs were subjected to a profuse amount of degradation and aggregation. However this issue was controlled for by culturing a higher quantity of cells, lower post-induction temperatures, protease cocktail inhibitors, PMSF, urea washes, and keeping the lysates, protein fractions, and samples on ice and frozen during overnight periods at -20°C with a minimum of 10% glycerol.

^{1}H-^{15}N-^{13}C HSQC Nuclear Magnetic Resonance (NMR) Spectroscopy of His_6-hLa [225/332], His_6-hLa [225/363], and His_6-hLa [225/363]

HSQC data were collected on a 700 MHz Bruker magnet at a 1:1 concentration with various RNAs and hLa constructs. Shifts in peaks from the spectra of unbound hLa verses bound RNA-hLa demonstrated the affinity of hLa for particular RNAs. When His_6-hLa [225/332] was titrated with 5'OH double-stranded RNA **(Figure 29A)** there was peak shifting. This was a similar case when His_6-hLa [225/408] was titrated with 5'OH double-stranded RNA **(Figure 29B)**. Likewise, hLa[225/363] peaks shifted when bound to 5'OH double-stranded RNA **(Figure 30A)** as well as poly-A20 RNA **(Figure 30B)**.

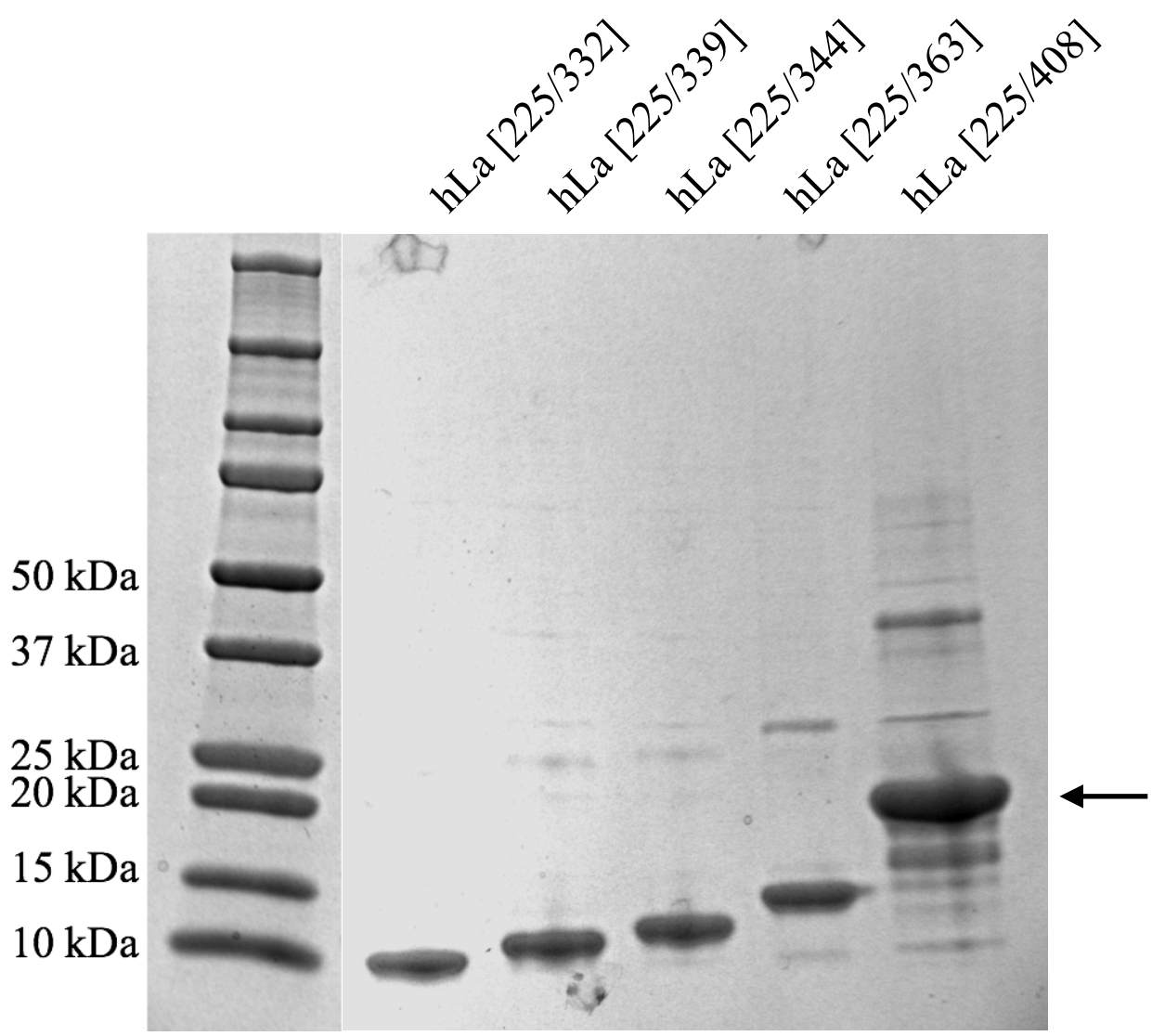


Figure 28: Confirmation of hLa constructs; hLa [225/332], hLa [225/339], hLa[225/344], hLa[225/363], hLa [225/408] are also confirmed at the correct molecular weights at 13.2kDa, 14kDa, 14.5kDa, 15.4kDa, and 20.6kDa. Visualized on a 12% SDS gel using a protein ladder (Biorad Precision Plus Protein™ Dual Color Standards) and visualized with Fast SeeBand Single Step Protein staining solution (FroggaBio) on the Alpha Imager HP system (Alpha Innotech). Arrows indicate the band with the correct molecular weight.

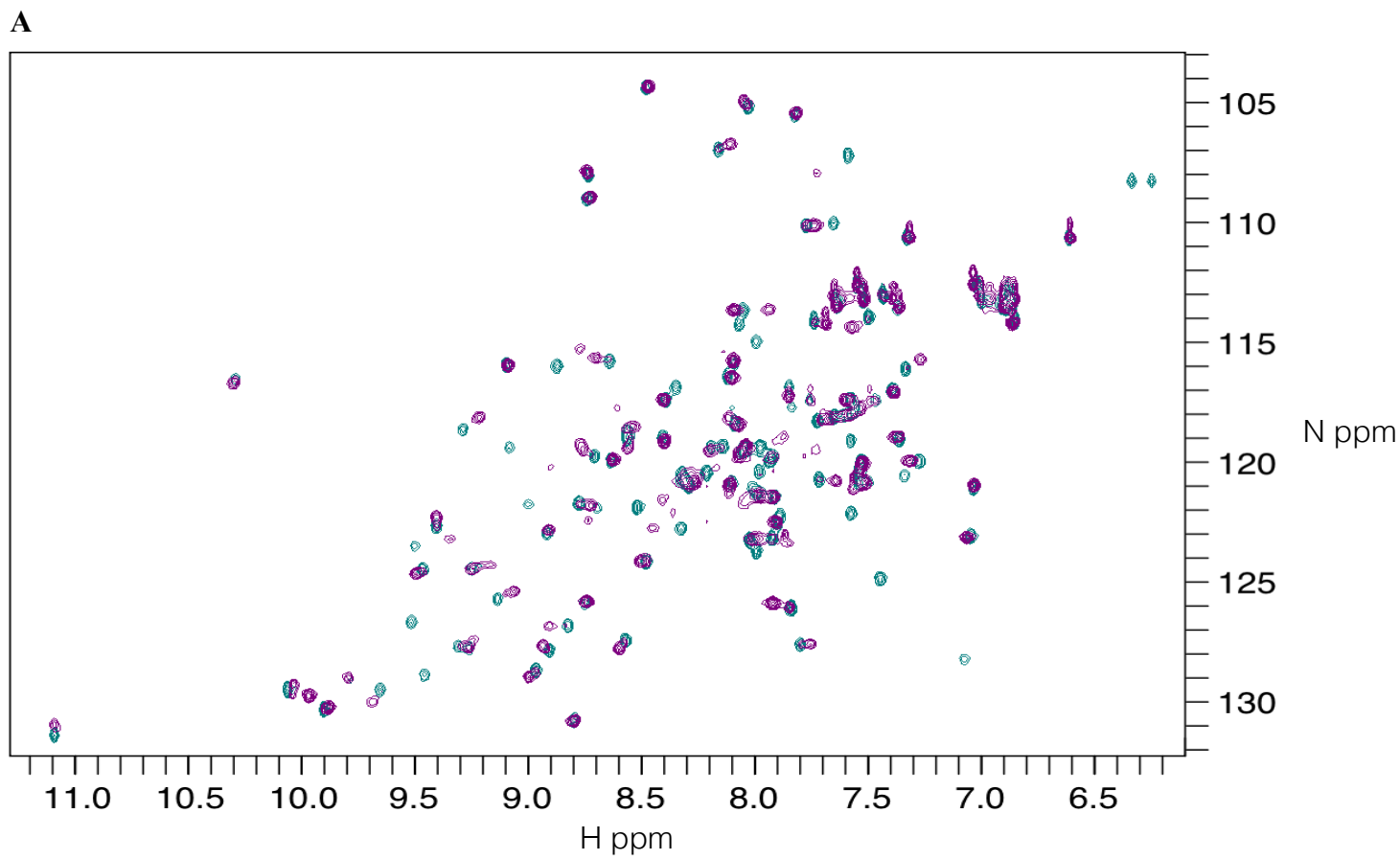


Figure 29: ^{1}H-^{15}N HSQC spectrum of hLa constructs unbound and bound with 5'OH RNA. (**A**) ^{1}H-^{15}N HSQC spectra of hLa [225/332] unbound (purple) and bound to 5'OH RNA (green). (**B**) ^{1}H-^{15}N HSQC spectra of hLa [225/408] unbound (red) and bound to 5'OH RNA (blue). The shifts in peaks of both constructs unbound compared to when bound to 5'OH RNA are indicative of binding and a conformational change in the structure upon formation of a RNA-hLa complex.

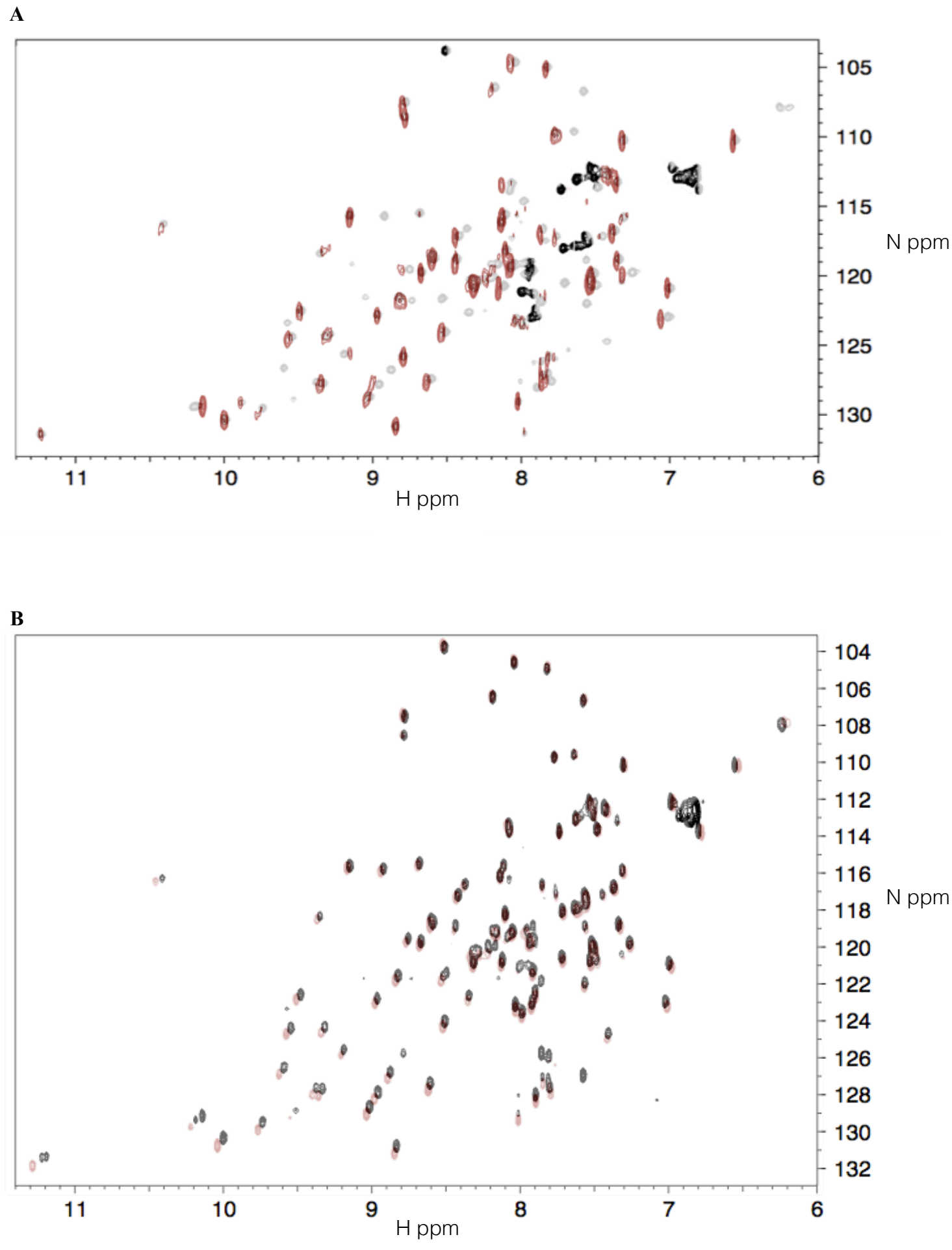


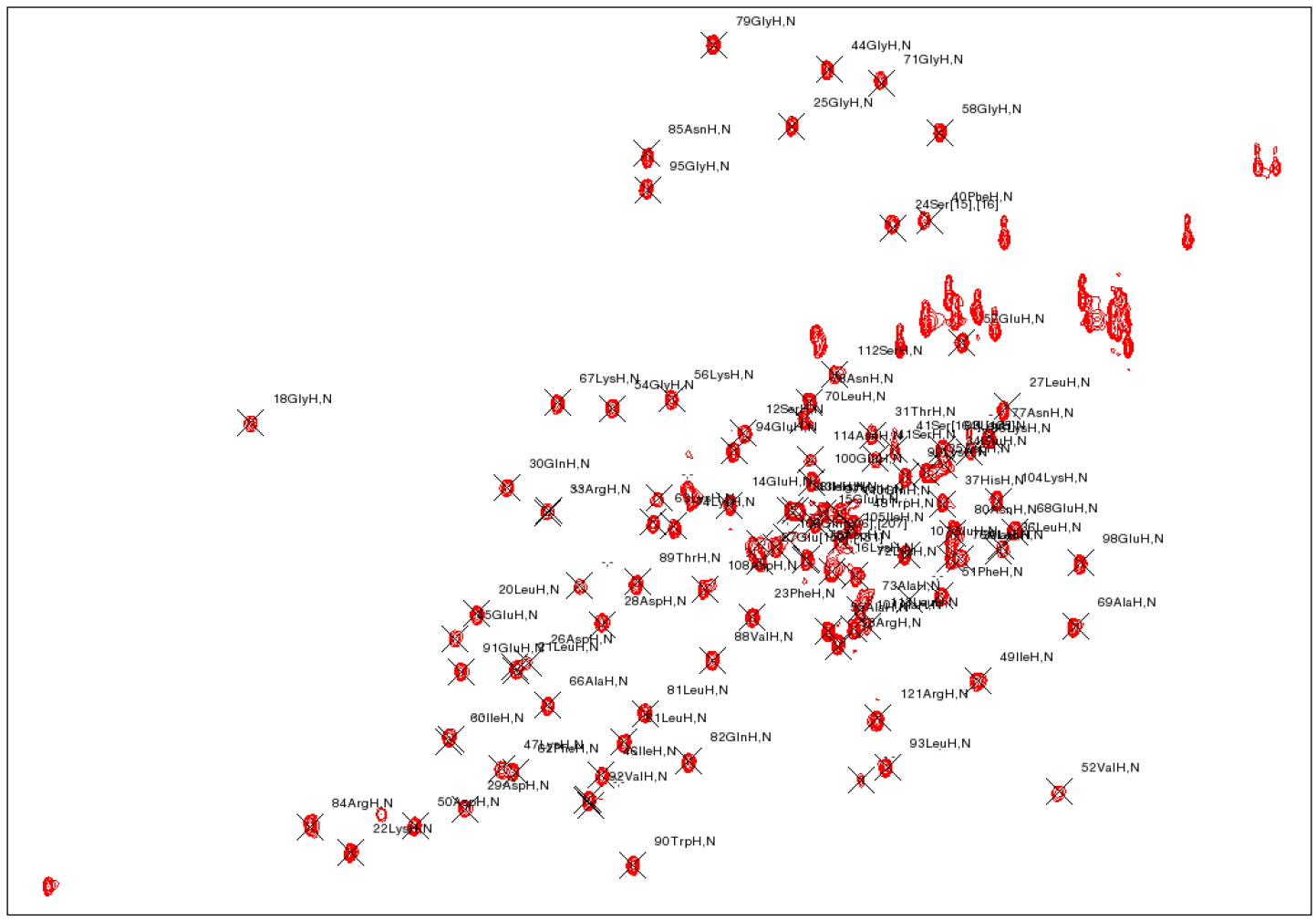
Figure 30: ^{1}H-^{15}N HSQC spectrum of hLa [225/363] unbound and bound to RNA. (**A**) ^{1}H-^{15}N HSQC spectra of hLa [225/363] unbound (brown) and bound to single stranded 5'OH (black). (**B**) ^{1}H-^{15}N HSQC spectra of hLa [225/363] unbound (brown) and bound to A20 RNA (black). The shifts in peaks of hLA [225/363] unbound compared to when bound to either RNA are indicative of binding and a conformational change in the structure upon the formation of a RNA-hLa complex.

HSQC spectra of hLa [225/332] unbound **(Figure 31A)** and bound to a 12 nucleotide RNA trailer **(Figure 31B)** in a 1:1 concentration of RNA to protein were reassigned using the original assignments assigned by Jacks *et al* 2003 (Jacks et al., 2003). The overlay **(Figure 32A)** of both the unbound and bound state differed in their spectra and significant shifts were present. Based on the calculated difference in the HSQC peak shifts via Dr. Logan Donaldson's UNIX Perl Script **(Figure 32B)** between the unbound and bound spectra it is apparent that there is binding between the 12-nucleotide trailer RNA and the hLa [225/332] construct. Several shifts that occurred after binding included F23, L30, L36, Q45, F51, G54, G58, I60, L61, R84, and I105. There was prominent peak broadening post binding of E87, I106, Q109, and E111 in one region of the spectra as well as E34, D35, L83, K99, K103 within another region of the spectra. K103 only became present post-binding. In addition, peaks for F51 and V52 disappeared in the complexed spectra.

6.6 DISCUSSION: Predicted hLa [225/332] Model and Validation

The disappearance of peaks between the unbound and bound state could reflect transient binding consistent with protein dynamics at the milisecond-to-microsecond time scale. Plotting the largest peak shifts of hLa [225/332] bound to the 12-nucleotide trailer RNA on the PDB structure of hLa [225/334] demonstrated that there is an increase in binding toward the end of the C-terminal of the construct **(Figure 33A & 33B)**. There is an assumption that the strongest shifts (yellow<orange<red) and surface exposed amino acids **(Figure 33C & 33D)** are involved in RNA binding and provides a

A



B

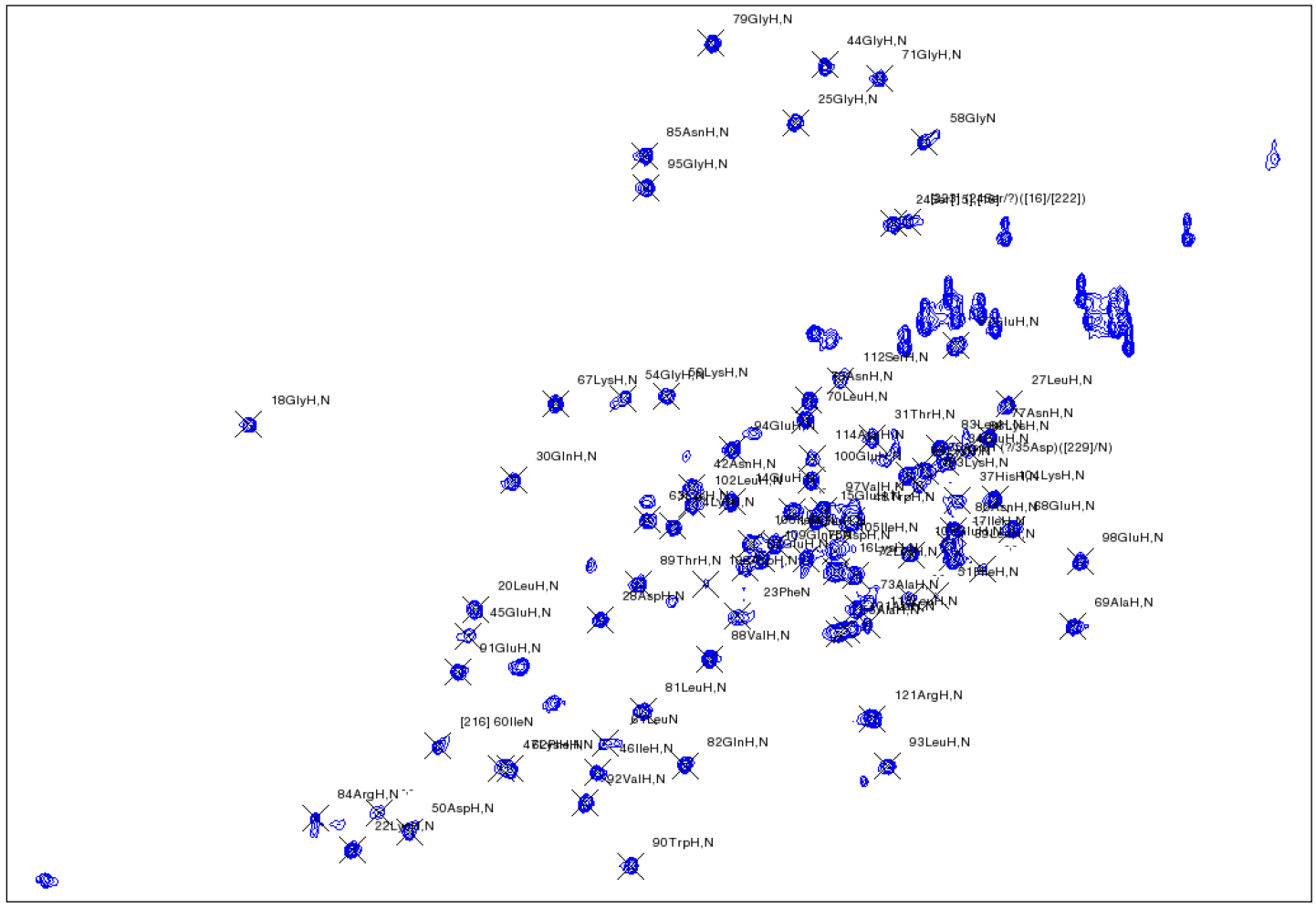
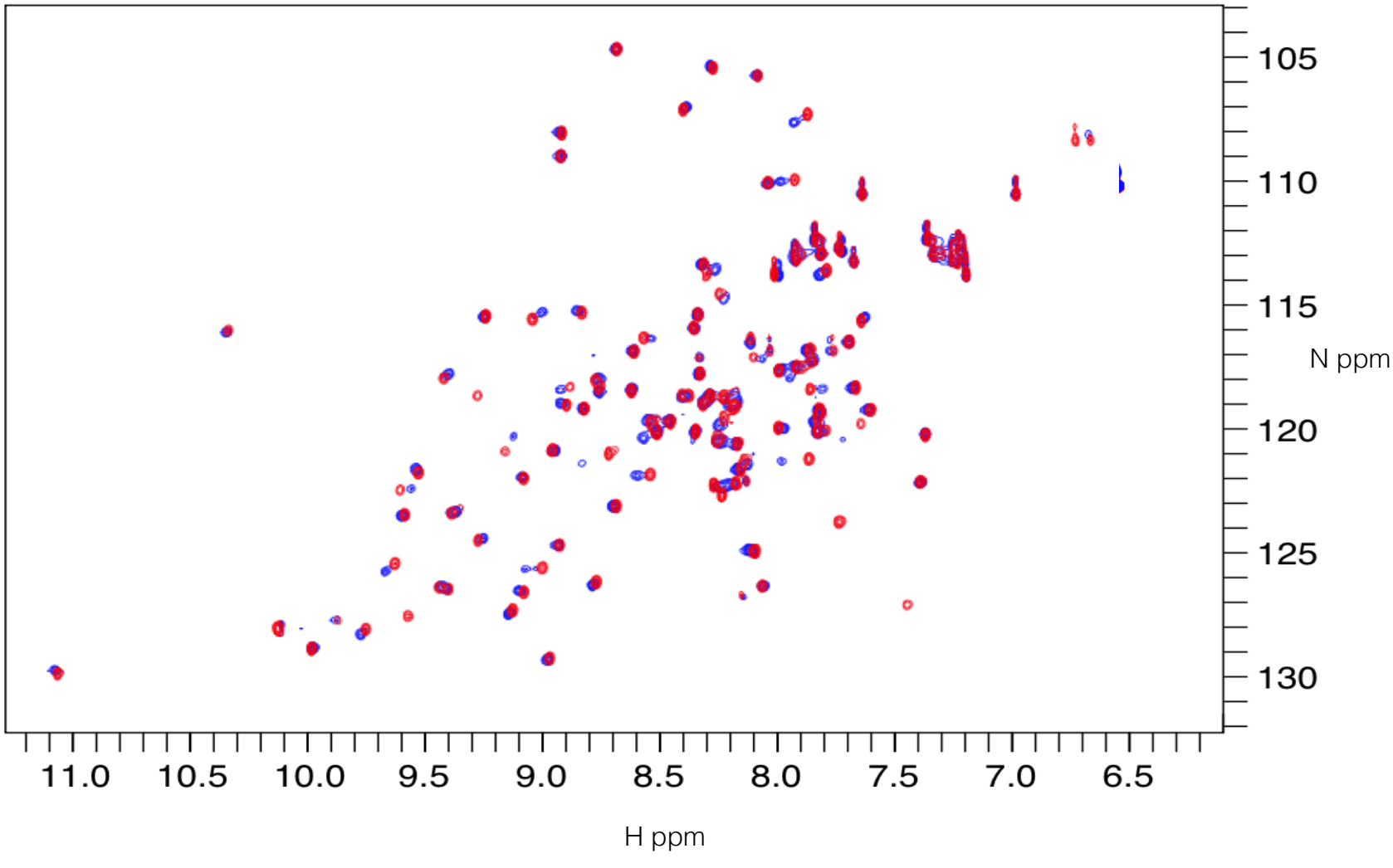


Figure 31: ^{1}H-^{15}N HSQC spectrum of hLa [222/332] (**A**) unbound (red) (**B**) bound to a 12 nucleotide RNA trailer (blue). The peaks are re-assigned based on Jacks *et al* (2003) assignments using the software NMR Analysis.

A



B

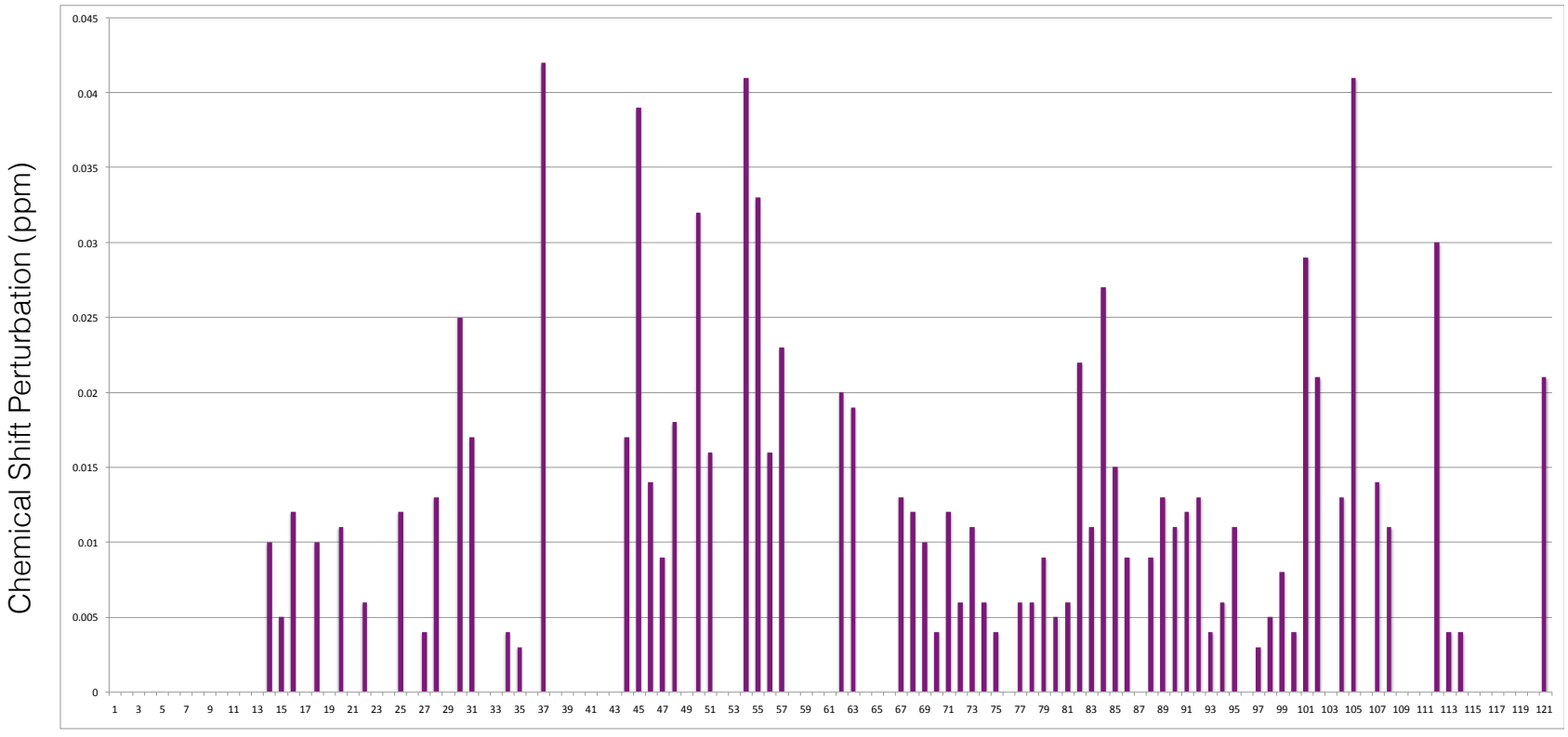


Figure 32: (**A**) ^{1}H-^{15}N HSQC spectrum of hLa [222/332] unbound (red) and bound to a 12 nucleotide RNA trailer (blue). The shifts in peaks of hLA [225/332] unbound compared to when bound to the 12-nt trailer RNA are indicative of binding and a conformational change in the structure from unbound to bound. (**B**) Plotted difference in the shift coordinates for each assigned peak between unbound hLa [225/332] and bound hLa [225/332] with a 12 nucleotide trailer RNA.

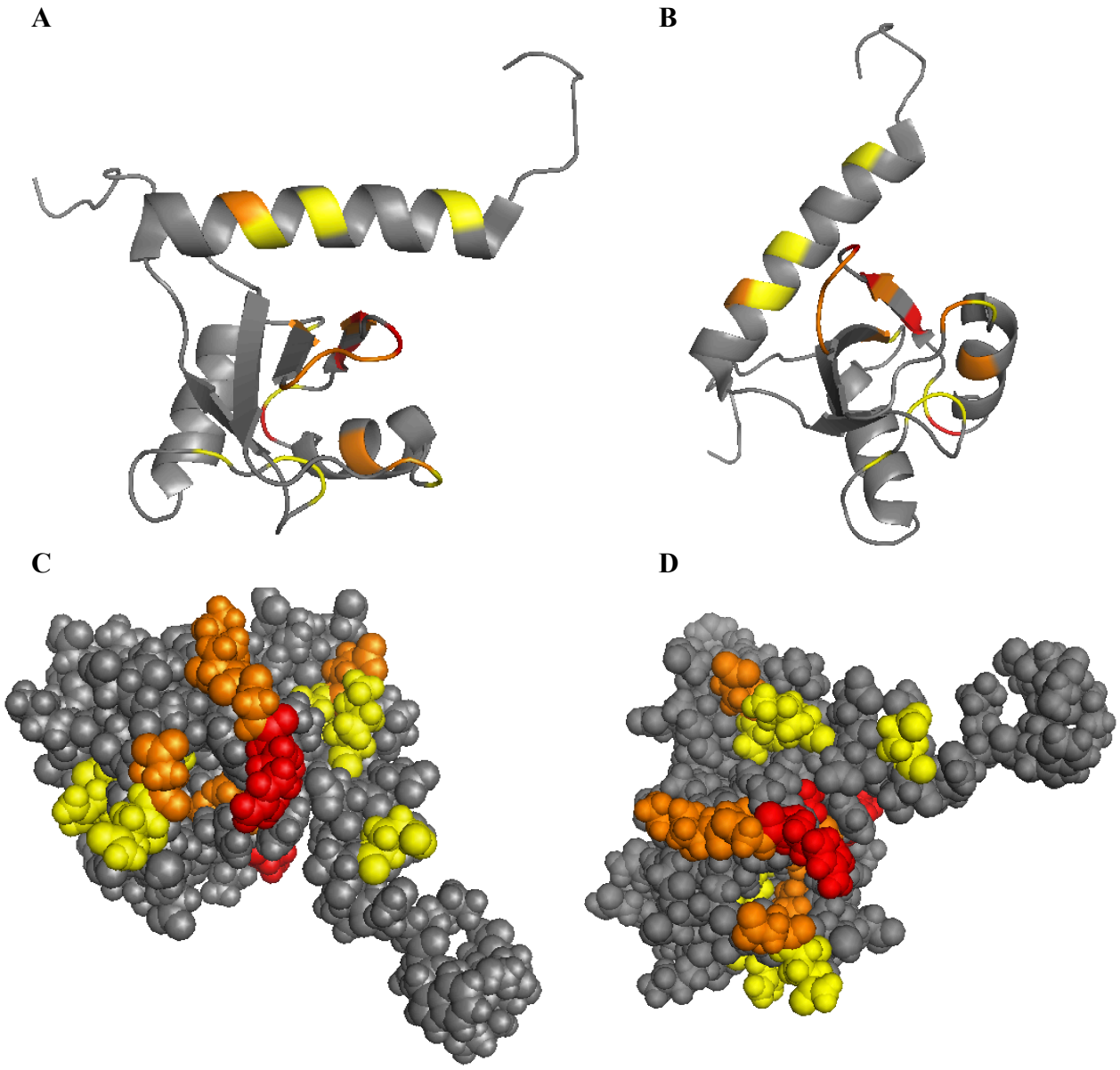


Figure 33: NMR structure (PDB ID: **1OWX**) of hLa [225/334] RRM2 with significant peak shifts between unbound hLa [225/332] and bound hLa[225/332] with a 12-nucleotide trailer RNA. Significance determined as weak (yellow); 0.015-0.019, medium (orange); 0.02-0.029, and strong (red) ≥ 0.03. (**A**, **B**) Cartoon figures rotated at two different angles. (**C**, **D**) Reveal the potential binding groove for RNA to dock into hLa.

docking site for RNA to bind to hLa and engage in electrostatic interactions. There is also significant shifting in loop three similar to the loop three that shifts in p65. This provides some insight as to there being an affinity for RNA along the C-terminal and the potential for a change in the structural confirmation of a longer hLa C-terminal construct. The ability for the RRM2 to recognize RNA may demonstrate cooperative binding such that it increases the affinity for the SBM to become engaged and clasp over RNA. This could be the mechanism to which the C-terminal of hLa changes confirmation from an unstructured tail to an extended structured alpha helix upon binding to RNA.

6.7 CONCLUSION

hLa demonstrates a high affinity for various RNAs. The affinity between the RRM2 for RNA may demonstrate cooperative binding such that it increases the affinity for the SBM to become engaged with RNA and endure a change in confirmation from an unstructured tail to a structured alpha helix. More importantly, the structural confirmation is altered in the presence of RNA and affects its functional state much like the commitment of SOX9 when bound to DNA.

6.8 FUTURE WORKS

Using the most stable C-terminal hLa construct, a ^{15}N-^{13}C NMR sample will be produced and titrated with the 12-nucleotide trailer RNA to compare the bound and unbound shifts in order to assign novel peaks to the potentially structured C-terminal when bound to RNA and determine if an extension of the third alpha helix is conformed post-RNA binding. Testing the effect of phosphorylation of hLa could enhance the understanding of hLa-RNA binding. In the event of hLa serine 366 phosphorylation, the

acidic region is exposed and could be decreasing the affinity of hLa to bind to RNA. If Serine 366 is not phosphorylated, the acidic amino acids are less exposed enabling the basic amino acids to bind to RNA.

CHAPTER VII

REFERENCES

Adesida, A.B., Grady, L.M., Khan, W.S., Millward-Sadler, S.J., Salter, D.M., and Hardingham, T.E. (2007). Human meniscus cells express hypoxia inducible factor-1alpha and increased SOX9 in response to low oxygen tension in cell aggregate culture. Arthritis Res. Ther. *9*, R69.

Agarwal, P., Verzi, M.P., Nguyen, T., Hu, J., Ehlers, M.L., McCulley, D.J., Xu, S.-M., Dodou, E., Anderson, J.P., Wei, M.L., et al. (2011). The MADS box transcription factor MEF2C regulates melanocyte development and is a direct transcriptional target and partner of SOX10. Development *138*, 2555–2565.

Akiyama, H., Chaboissier, M.-C., Martin, J.F., Schedl, A., and de Crombrugghe, B. (2002). The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. Genes Dev. *16*, 2813–2828.

Alfano, C., Sanfelice, D., Babon, J., Kelly, G., Jacks, A., Curry, S., and Conte, M.R. (2004). Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. Nat. Struct. Mol. Biol. *11*, 323–329.

Ali, N., Pruijn, G.J., Kenan, D.J., Keene, J.D., and Siddiqui, A. (2000). Human La antigen is required for the hepatitis C virus internal ribosome entry site-mediated translation. J. Biol. Chem. *275*, 27531–27540.

Allain, F.H., Yen, Y.M., Masse, J.E., Schultze, P., Dieckmann, T., Johnson, R.C., and Feigon, J. (1999). Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding. EMBO J. *18*, 2563–2579.

Barrionuevo, F., Naumann, A., Bagheri-Fam, S., Speth, V., Taketo, M.M., Scherer, G., and Neubüser, A. (2008). Sox9 is required for invagination of the otic placode in mice. Dev. Biol. *317*, 213–224.

Bernard, P., Tang, P., Liu, S., Dewing, P., Harley, V.R., and Vilain, E. (2003). Dimerization of SOX9 is required for chondrogenesis, but not for sex determination. Hum. Mol. Gen. *12*, 1755–1765.

Betancur, P., Bronner-Fraser, M., and Sauka-Spengler, T. (2010). Genomic code for Sox10 activation reveals a key regulatory enhancer for cranial neural crest. Proc. Natl. Acad. Sci. U.S.A. *107*, 3570–3575.

Bi, W., Huang, W., Whitworth, D.J., Deng, J.M., Zhang, Z., Behringer, R.R., and de Crombrugghe, B. (2001). Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. Proc. Natl. Acad. Sci. U.S.A. *98*, 6698–6703.

Bianchi, M.E., and Agresti, A. (2005). HMG proteins: dynamic players in gene regulation and differentiation. Curr. Opin. Genetics Dev. *15*, 496–506.

Camaj, P., Jäckel, C., Krebs, S., De Toni, E.N., Blum, H., Jauch, K.-W., Nelson, P.J., and Bruns, C.J. (2014). Hypoxia-independent gene expression mediated by SOX9 promotes aggressive pancreatic tumor biology. Mol. Cancer Res. *12*, 421–432.

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.

Chasman, D., Cepek, K., Sharp, P.A., and Pabo, C.O. (1999). Crystal structure of an OCA-B peptide bound to an Oct-1 POU domain/octamer DNA complex: specific recognition of a protein-DNA interface. Genes Dev. *13*, 2650–2657.

Cheng, L.-C., Pastrana, E., Tavazoie, M., and Doetsch, F. (2009). miR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. Nat. Neurosci. *12*, 399–408.

Cheung, M., and Briscoe, J. (2003). Neural crest development is regulated by the transcription factor Sox9. Development *130*, 5681–5693.

Choi, Y.J., Song, J.H., Yoon, J.H., Choi, W.S., Nam, S.W., Lee, J.Y., and Park, W.S. (2014). Aberrant expression of SOX9 is associated with gastrokine 1 inactivation in gastric cancers. Gastric Cancer *17*, 247–254.

Chung, M.I.S., Ma, A.C.H., Fung, T.-K., and Leung, A.Y.H. (2011). Characterization of Sry-related HMG box group F genes in zebrafish hematopoiesis. Exp. Hematol. *39*, 986–998.e5.

Connor, F., Cary, P.D., Read, C.M., Preston, N.S., Driscoll, P.C., Denny, P., Crane-Robinson, C., and Ashworth, A. (1994). DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. Nucleic Acids Res. *22*, 3339–3346.

Coustry, F., Oh, C.-D., Hattori, T., Maity, S.N., de Crombrugghe, B., and Yasuda, H. (2010). The dimerization domain of SOX9 is required for transcription activation of a chondrocyte-specific chromatin DNA template. Nucleic Acids Res. *38*, 6018–6028.

De Santa Barbara, P., Bonneaud, N., Boizet, B., Desclozeaux, M., Moniot, B., Sudbeck, P., Scherer, G., Poulat, F., and Berta, P. (1998). Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Müllerian hormone gene. Mol. Cell. Biol. *18*, 6653–6665.

Dong, G., Chakshusmathi, G., Wolin, S.L., and Reinisch, K.M. (2004). Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch. EMBO J. *23*, 1000–1007.

Dragan, A.I., Read, C.M., Makeyeva, E.N., Milgotina, E.I., Churchill, M.E.A., Crane-Robinson, C., and Privalov, P.L. (2004). DNA binding and bending by HMG boxes: energetic determinants of specificity. J. Mol. Biol. *343*, 371–393.

Dy, P., Wang, W., Bhattaram, P., Wang, Q., Wang, L., Ballock, R.T., and Lefebvre, V. (2012). Sox9 directs hypertrophic maturation and blocks osteoblast differentiation of growth plate chondrocytes. Dev. Cell *22*, 597–609.

Eggers, S., Ohnesorg, T., and Sinclair, A. (2014). Genetic regulation of mammalian gonad development. Nat. Rev. Endocrinol. *10*, 673–683.

Ein, S.H., Masiakos, P.T., and Ein, A. (2014). The ins and outs of pyloromyotomy: what we have learned in 35 years. Pediatr. Surg. Int. *30*, 467–480.

Fan, H., Sakulich, A.L., Goodier, J.L., Zhang, X., Qin, J., and Maraia, R.J. (1997). Phosphorylation of the human La antigen on serine 366 can regulate recycling of RNA polymerase III transcription complexes. Cell *88*, 707–715.

Fantauzzo, K.A., Kurban, M., Levy, B., and Christiano, A.M. (2012). Trps1 and its target gene Sox9 regulate epithelial proliferation in the developing hair follicle and are associated with hypertrichosis. PLoS Genet. *8*, e1003002.

Feuerstein, R., Wang, X., Song, D., Cooke, N.E., and Liebhaber, S.A. (1994). The LIM/ double zinc-finger motif functions as a protein dimerization domain. Proc. Natl. Acad. Sci. U.S.A. *91*, 10655–10659.

Foster, J.W. (1996). Mutations in SOX9 cause both autosomal sex reversal and campomelic dysplasia. Acta. Paediatr. Jpn. *38*, 405–411.

Furuyama, K., Kawaguchi, Y., Akiyama, H., Horiguchi, M., Kodama, S., Kuhara, T., Hosokawa, S., Elbahrawy, A., Soeda, T., Koizumi, M., et al. (2011). Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. Nat. Genet. *43*, 34–41.

Goji, K., Nishijima, E., Tsugawa, C., Nishio, H., Pokharel, R.K., and Matsuo, M. (1998). Novel missense mutation in the HMG box of SOX9 gene in a Japanese XY male resulted in campomelic dysplasia and severe defect in masculinization. Hum. Mutat. *Suppl 1*, S114–S116.

Goodier, J.L., Fan, H., and Maraia, R.J. (1997). A carboxy-terminal basic region controls RNA polymerase III transcription factor activity of human La protein. Mol. Cell. Biol. *17*, 5823–5832.

Guo, W., Keckesova, Z., Donaher, J.L., Shibue, T., Tischler, V., Reinhardt, F., Itzkovitz, S., Noske, A., Zürrer-Härdi, U., Bell, G., et al. (2012). Slug and Sox9 cooperatively determine the mammary stem cell state. Cell *148*, 1015–1028.

Han, Y., and Lefebvre, V. (2008). L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer. Mol. Cell. Biol. *28*, 4999–5013.

Harley, V.R., Lovell-Badge, R., and Goodfellow, P.N. (1994). Definition of a consensus DNA binding site for SRY. Nucleic Acids Res. *22*, 1500–1501.

Hattori, T., Müller, C., Gebhard, S., Bauer, E., Pausch, F., Schlund, B., Bösl, M.R., Hess, A., Surmann-Schmitt, C., Mark, von der, H., et al. (2010). SOX9 is a major negative regulator of cartilage vascularization, bone marrow formation and endochondral ossification. Development *137*, 901–911.

Healy, C., Uwanogho, D., and Sharpe, P.T. (1999). Regulation and role of Sox9 in cartilage formation. Dev. Dyn. *215*, 69–78.

Herschlag, D. (1995). RNA chaperones and the RNA folding problem. J. Biol. Chem. *270*, 20871–20874.

Horke, S., Reumann, K., Rang, A., and Heise, T. (2002). Molecular characterization of the human La protein.hepatitis B virus RNA.B interaction in vitro. J. Biol. Chem. *277*, 34949–34958.

Horke, S., Reumann, K., Schweizer, M., Will, H., and Heise, T. (2004). Nuclear trafficking of La protein depends on a newly identified nucleolar localization signal and the ability to bind RNA. J. Biol. Chem. *279*, 26563–26570.

Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P.D., Curry, S., and Conte, M.R. (2003). Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. Structure *11*, 833–843.

Kamachi, Y., and Kondoh, H. (2013). Sox proteins: regulators of cell fate specification and differentiation. Development *140*, 4129–4144.

Kanazawa, T., Furumatsu, T., Matsumoto-Ogawa, E., Maehara, A., and Ozaki, T. (2014). Role of Rho small GTPases in meniscus cells. J. Orthop. Res. *32*, 1479–1486.

Kang, P., Lee, H.K., Glasgow, S.M., Finley, M., Donti, T., Gaber, Z.B., Graham, B.H., Foster, A.E., Novitch, B.G., Gronostajski, R.M., et al. (2012). Sox9 and NFIA coordinate a transcriptional regulatory cascade during the initiation of gliogenesis. Neuron *74*, 79–94.

Kopp, J.L., Dubois, C.L., Schaffer, A.E., Hao, E., Shih, H.P., Seymour, P.A., Ma, J., and Sander, M. (2011). Sox9+ ductal cells are multipotent progenitors throughout development but do not produce new endocrine cells in the normal or injured adult pancreas. Development *138*, 653–665.

Kotik-Kogan, O., Valentine, E.R., Sanfelice, D., Conte, M.R., and Curry, S. (2008). Structural analysis reveals conformational plasticity in the recognition of RNA 3' ends by the human La protein. Structure *16*, 852–862.

Kucera, N.J., Hodsdon, M.E., and Wolin, S.L. (2011). An intrinsically disordered C terminus allows the La protein to assist the biogenesis of diverse noncoding RNA precursors. Proc. Natl. Acad. Sci. U.S.A. *108*, 1308–1313.

Kuehnert, J., Sommer, G., Zierk, A.W., Fedarovich, A., Brock, A., Fedarovich, D., and Heise, T. (2015). Novel RNA chaperone domain of RNA-binding protein La is regulated by AKT phosphorylation. Nucleic Acids Res. *43*, 581–594.

Kwok, C., Weller, P.A., Guioli, S., Foster, J.W., Mansour, S., Zuffardi, O., Punnett, H.H., Dominguez-Steglich, M.A., Brook, J.D., and Young, I.D. (1995). Mutations in SOX9, the gene responsible for Campomelic dysplasia and autosomal sex reversal. Am. J. Hum. Genet. *57*, 1028–1036.

Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science *240*, 1759–1764.

Lefebvre, V., Huang, W., Harley, V.R., Goodfellow, P.N., and de Crombrugghe, B. (1997). SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. Mol. Cell. Biol. *17*, 2336–2346.

Lefebvre, V., Dumitriu, B., Penzo-Méndez, A., Han, Y., and Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int. J. Biochem. Cell Biol. *39*, 2195–2214.

Leung, V.Y.L., Gao, B., Leung, K.K.H., Melhado, I.G., Wynn, S.L., Au, T.Y.K., Dung, N.W.F., Lau, J.Y.B., Mak, A.C.Y., Chan, D., et al. (2011). SOX9 governs differentiation stage-specific gene expression in growth plate chondrocytes via direct concomitant transactivation and repression. PLoS Genet. *7*, e1002356.

Lippard, S.J., and Hoeschele, J.D. (1979). Binding of cis- and trans-dichlorodiammineplatinum(II) to the nucleosome core. Proc. Natl. Acad. Sci. U.S.A. *76*, 6091–6095.

Liu, J.A.J., Wu, M.-H., Yan, C.H., Chau, B.K.H., So, H., Ng, A., Chan, A., Cheah, K.S.E., Briscoe, J., and Cheung, M. (2013). Phosphorylation of Sox9 is required for neural crest delamination and is regulated downstream of BMP and canonical Wnt signaling. Proc. Natl. Acad. Sci. U.S.A. *110*, 2882–2887.

Mansour, S., Hall, C.M., Pembrey, M.E., and Young, I.D. (1995). A clinical and genetic study of campomelic dysplasia. J. Med. Genet. *32*, 415–420.

Martino, L., Pennell, S., Kelly, G., Bui, T.T.T., Kotik-Kogan, O., Smerdon, S.J., Drake, A.F., Curry, S., and Conte, M.R. (2012). Analysis of the interaction with the hepatitis C virus mRNA reveals an alternative mode of RNA recognition by the human La protein. Nucleic Acids Res. *40*, 1381–1394.

Matsuda, S., Kuwako, K.-I., Okano, H.J., Tsutsumi, S., Aburatani, H., Saga, Y., Matsuzaki, Y., Akaike, A., Sugimoto, H., and Okano, H. (2012). Sox21 promotes hippocampal adult neurogenesis via the transcriptional repression of the Hes5 gene. J. Neurosci. *32*, 12543–12557.

Meeson, A.P., Shi, X., Alexander, M.S., Williams, R.S., Allen, R.E., Jiang, N., Adham, I.M., Goetsch, S.C., Hammer, R.E., and Garry, D.J. (2007). Sox15 and Fhl3 transcriptionally coactivate Foxk1 and regulate myogenic progenitor cells. EMBO J. *26*, 1902–1912.

Mertin, S., McDowall, S.G., and Harley, V.R. (1999). The DNA-binding specificity of SOX9 and other SOX proteins. Nucleic Acids Res. *27*, 1359–1364.

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.

Murphy, F.V., Sehy, J.V., Dow, L.K., Gao, Y.G., and Churchill, M.E. (1999). Co-crystallization and preliminary crystallographic analysis of the high mobility group domain of HMG-D bound to DNA. Acta Crystallogr. D *55*, 1594–1597.

Naeeni, A.R., Conte, M.R., and Bayfield, M.A. (2012). RNA chaperone activity of human La protein is mediated by variant RNA recognition motif. J. Biol. Chem. *287*, 5472–5482.

Nakata, K., Shino, K., Hamada, M., Mae, T., Miyama, T., Shinjo, H., Horibe, S., Tada, K., Ochi, T., and Yoshikawa, H. (2001). Human meniscus cell: characterization of the primary culture and use for tissue engineering. Clin. Orthop. Relat. Res. S208–S218.

Ng, L.J., Wheatley, S., Muscat, G.E., Conway-Campbell, J., Bowles, J., Wright, E., Bell, D.M., Tam, P.P., Cheah, K.S., and Koopman, P. (1997). SOX9 binds DNA, activates transcription, and coexpresses with type II collagen during chondrogenesis in the mouse. Dev. Biol. *183*, 108–121.

Nowak, J.A., Polak, L., Pasolli, H.A., and Fuchs, E. (2008). Hair follicle stem cells are specified and function in early skin morphogenesis. Cell Stem Cell *3*, 33–43.

Ohndorf, U.M., Rould, M.A., He, Q., Pabo, C.O., and Lippard, S.J. (1999). Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. Nature *399*, 708–712.

Okuda, Y., Yoda, H., Uchikawa, M., Furutani-Seiki, M., Takeda, H., Kondoh, H., and Kamachi, Y. (2006). Comparative genomic and expression analysis of group B1 sox genes in zebrafish indicates their diversification during vertebrate evolution. Dev. Dyn. *235*, 811–825.

Osaki, E., Nishina, Y., Inazawa, J., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Ohsugi, M., Tezuka, T., Yoshida, M., and Semba, K. (1999). Identification of a novel Sry-related gene and its germ cell-specific expression. Nucleic Acids Res. *27*, 2503–2510.

Peirano, R.I., and Wegner, M. (2000). The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences. Nucleic Acids Res. *28*, 3047–3055.

Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. Genes Dev. *24*, 1832–1860.

Poché, R.A., Furuta, Y., Chaboissier, M.-C., Schedl, A., and Behringer, R.R. (2008). Sox9 is expressed in mouse multipotent retinal progenitor cells and functions in Müller glial cell development. J. Comp. Neurol. *510*, 237–250.

Pozniak, C.D., Langseth, A.J., Dijkgraaf, G.J.P., Choe, Y., Werb, Z., and Pleasure, S.J. (2010). Sox10 directs neural stem cells toward the oligodendrocyte lineage by decreasing Suppressor of Fused expression. Proc. Natl. Acad. Sci. U.S.A. *107*, 21795–21800.

Pudi, R., Abhiman, S., Srinivasan, N., and Das, S. (2003). Hepatitis C virus internal ribosome entry site-mediated translation is stimulated by specific interaction of independent regions of human La autoantigen. J. Biol. Chem. *278*, 12231–12240.

Rawlins, E.L. (2011). The building blocks of mammalian lung development. Dev. Dyn. *240*, 463–476.

Sakai, D., Suzuki, T., Osumi, N., and Wakamatsu, Y. (2006). Cooperative action of Sox9, Snail2 and PKA signaling in early neural crest development. Development *133*, 1323–1333.

Sanfelice, D., Babon, J., Kelly, G., Curry, S., and Conte, M.R. (2004). Resonance assignment and secondary structure of the La motif. J. Biomol. NMR *29*, 449–450.

Sarkar, A., and Hochedlinger, K. (2013). The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell *12*, 15–30.

Sato, T., van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., van den Born, M., Barker, N., Shroyer, N.F., van de Wetering, M., and Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature *469*, 415–418.

Scaffidi, P., and Bianchi, M.E. (2001). Spatially precise DNA bending is an essential activity of the sox2 transcription factor. J. Biol. Chem. *276*, 47296–47302.

Schlierf, B., Ludwig, A., Klenovsek, K., and Wegner, M. (2002). Cooperative binding of Sox10 to DNA: requirements and consequences. Nucleic Acids Res. *30*, 5509–5516.

Schwartz, E.I., Intine, R.V., and Maraia, R.J. (2004). CK2 is responsible for phosphorylation of human La protein serine-366 and can modulate rpL37 5'-terminal oligopyrimidine mRNA metabolism. Mol. Cell. Biol. *24*, 9580–9591.

Scott, C.E., Wynn, S.L., Sesay, A., Cruz, C., Cheung, M., Gomez Gaviro, M.-V., Booth, S., Gao, B., Cheah, K.S.E., Lovell-Badge, R., et al. (2010). SOX9 induces and maintains neural stem cells. Nat. Neurosci *13*, 1181–1189.

Sekido, R., and Lovell-Badge, R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. Nature *453*, 930–934.

Seymour, P.A., Freude, K.K., Tran, M.N., Mayes, E.E., Jensen, J., Kist, R., Scherer, G., and Sander, M. (2007). SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proc. Natl. Acad. Sci. U.S.A. *104*, 1865–1870.

Singh, M., Choi, C.P., and Feigon, J. (2013). xRRM: a new class of RRM found in the telomerase La family protein p65. RNA Biol. *10*, 353–359.

Singh, M., Wang, Z., Koo, B.-K., Patel, A., Cascio, D., Collins, K., and Feigon, J. (2012). Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. Mol. Cell *47*, 16–26.

Smits, P., Li, P., Mandel, J., Zhang, Z., Deng, J.M., Behringer, R.R., de Crombrugghe, B., and Lefebvre, V. (2001). The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. Dev. Cell *1*, 277–290.

Sock, E. (2003). Loss of DNA-dependent dimerization of the transcription factor SOX9 as a cause for campomelic dysplasia. Hum. Mol. Gen. *12*, 1439–1447.

Sock, E., Pagon, R.A., Keymolen, K., Lissens, W., Wegner, M., and Scherer, G. (2003). Loss of DNA-dependent dimerization of the transcription factor SOX9 as a cause for campomelic dysplasia. Hum. Mol. Gen. *12*, 1439–1447.

Song, S., Ajani, J.A., Honjo, S., Maru, D.M., Chen, Q., Scott, A.W., Heallen, T.R., Xiao, L., Hofstetter, W.L., Weston, B., et al. (2014). Hippo coactivator YAP1 upregulates SOX9 and endows esophageal cancer cells with stem-like properties. Cancer Res. *74*, 4170–4182.

Stefan, M.I., and Le Novère, N. (2013). Cooperative binding. PLoS Comput. Biol. *9*, e1003106.

Stolt, C.C., Lommes, P., Sock, E., Chaboissier, M.-C., Schedl, A., and Wegner, M. (2003). The Sox9 transcription factor determines glial fate choice in the developing spinal cord. Genes Dev. *17*, 1677–1689.

Stolt, C.C., Rehberg, S., Ader, M., Lommes, P., Riethmacher, D., Schachner, M., Bartsch, U., and Wegner, M. (2002). Terminal differentiation of myelin-forming oligodendrocytes depends on the transcription factor Sox10. Genes Dev. *16*, 165–170.

Stolt, C.C., Schlierf, A., Lommes, P., Hillgärtner, S., Werner, T., Kosian, T., Sock, E., Kessaris, N., Richardson, W.D., Lefebvre, V., et al. (2006). SoxD proteins influence multiple stages of oligodendrocyte development and modulate SoxE protein function. Dev. Cell *11*, 697–709.

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

Svitkin, Y.V., Meerovitch, K., Lee, H.S., Dholakia, J.N., Kenan, D.J., Agol, V.I., and Sonenberg, N. (1994). Internal translation initiation on poliovirus RNA: further characterization of La function in poliovirus translation in vitro. J. Virol. *68*, 1544–1550.

Takamatsu, N., Kanda, H., Tsuchiya, I., Yamada, S., Ito, M., Kabeno, S., Shiba, T., and Yamashita, S. (1995). A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein. Mol. Cell. Biol. *15*, 3759–3766.

Teplova, M., Yuan, Y.-R., Phan, A.T., Malinina, L., Ilin, S., Teplov, A., and Patel, D.J. (2006). Structural basis for recognition and sequestration of UUU(OH) 3' temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. Mol. Cell *21*, 75–85.

Thomas, J.O. (2001). HMG1 and 2: architectural DNA-binding proteins. Biochem. Soc. Trans. *29*, 395–401.

Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J. *18*, 1169–1175.

Uchikawa, M., Yoshida, M., Iwafuchi-Doi, M., Matsuda, K., Ishida, Y., Takemoto, T., and Kondoh, H. (2011). B1 and B2 Sox gene expression during neural plate development in chicken and mouse embryos: universal versus species-dependent features. Dev. Growth Differ. *53*, 761–771.

Vidal, V.P.I., Chaboissier, M.-C., Lützkendorf, S., Cotsarelis, G., Mill, P., Hui, C.-C., Ortonne, N., Ortonne, J.-P., and Schedl, A. (2005). Sox9 is essential for outer root sheath differentiation and the formation of the hair stem cell compartment. Curr. Biol. *15*, 1340–1351.

Wegner, M. (1999). From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res. *27*, 1409–1420.

Wilhelm, D., Hiramatsu, R., Mizusaki, H., Widjaja, L., Combes, A.N., Kanai, Y., and Koopman, P. (2007). SOX9 regulates prostaglandin D synthase gene transcription in vivo to ensure testis development. J. Biol. Chem. *282*, 10553–10560.

Wissmüller, S., Kosian, T., Wolf, M., Finzsch, M., and Wegner, M. (2006). The high-mobility-group domain of Sox proteins interacts with DNA-binding domains of many transcription factors. Nucleic Acids Res. *34*, 1735–1744.

Wolin, S.L., and Cedervall, T. (2003). THE LA PROTEIN. Http://Dx.Doi.org/10.1146/Annurev.Biochem.71.090501.150003 71, 375–403.

Xu, J.-Q., Zhang, W.-B., Wan, R., and Yang, Y.-Q. (2014). MicroRNA-32 inhibits osteosarcoma cell proliferation and invasion by targeting Sox9. Tumour Biol. *35*, 9847–9853.

Zhao, Q., Eberspaecher, H., Lefebvre, V., and de Crombrugghe, B. (1997). Parallel expression of Sox9 and Col2a1 in cells undergoing chondrogenesis. Dev. Dyn. *209*, 377–386.

Zhu, H., Tang, J., Tang, M., and Cai, H. (2013). Upregulation of SOX9 in osteosarcoma and its association with tumor progression and patients' prognosis. Diagn. Pathol. *8*, 183.

Zuscik, M.J., Hilton, M.J., Zhang, X., Chen, D., and O'Keefe, R.J. (2008). Regulation of chondrogenesis and chondrocyte differentiation by stress. J. Clin. Invest. *118*, 429–438.