

Thesis
2
2002

This is to certify that the

dissertation entitled

THE X-RAY CRYSTALLOGRAPHIC STRUCTURES OF THE
MSX-1 HD/DNA COMPLEX AND THE OCT-1 POU/U1
OCTAMER/SNAP190 TERNARY COMPLEX

presented by

Stacy L. Hovde

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Analytical Chemistry

Date

8/20/02

James H. Deiger
Major professor

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

**THE X – RAY CRYSTALLOGRAPHIC STRUCTURES OF THE MSX – 1 HD/DNA
COMPLEX AND THE OCT – 1 POU/U1 OCTAMER/SNAP 190 TERNARY
COMPLEX**

By

Stacy L. Hovde

A DISSERTATION

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

DOCTOR OF PHILOSOPHY

Department of Chemistry

2002

ABSTRACT

The X – RAY CRYSTALLOGRAPHIC STRUCTURES OF THE MSX-1 HD/DNA COMPLEX AND THE OCT-1 POU/U1 OCTAMER/SNAP 190 TERNARY COMPLEX

By

Stacy L. Hovde

X-ray structures of two homeodomain complexes have been determined at high resolution. The Msx-1 HD/DNA complex was solved to 2.2 Å and the Oct-1 POU/U1 DNA/SNAP 190 peptide complex was solved to 2.4 Å. Though both structures contain homeodomains they have completely different functions but their mode of binding to DNA is conserved. The high resolution of both structures allow us to get a clearer picture of the important interactions between the domains and allows us to look at water mediated interactions. These structures were solved using the molecular replacement method.

The Msx-1 homeodomain protein plays a crucial role in craniofacial, limb and nervous system development. Homeodomain DNA-binding domains are comprised of 60 amino acids that show a high degree of evolutionary conservation. We have determined the structure of the Msx-1 homeodomain complexed to DNA at 2.2 Å resolution. The structure has an unusually well ordered n-terminal arm with a unique trajectory across the minor groove of the DNA. DNA specificity conferred by bases flanking the core TAAT sequence is explained by well ordered water-mediated interactions at Q50. Most interactions seen at the TAAT sequence are typical of the interactions seen in other homeodomain structures. Comparison of the Msx-1/HD structure to all other high

resolution HD/DNA complex structures indicate a remarkably well conserved sphere of hydration between the DNA and protein in these complexes.

Oct-1 is a ubiquitously expressed protein that interacts with SNAP 190, the largest subunit of SNAPc. We have studied the interaction of Oct-1 POU/SNAP 190 peptide to the U1 octamer element. Transcriptional activation of the highly expressed human U1 snRNA genes is dependent upon an octamer element contained within an upstream enhancer. This octamer element recruits the transcriptional activator protein Oct-1 that activates U1 transcription by direct protein contacts between Oct-1 and the general transcription factor SNAP190. Surprisingly, given the highly expressed nature of the U1 genes, the U1 octamer only weakly recruits the Oct-1 POU DNA binding domain but recruitment is stimulated by a peptide containing the region of SNAP190 previously identified as the target for Oct-1. Structural analysis of a co-complex of Oct-1 POU domain on the U1 octamer with a peptide from SNAP190 revealed that SNAP190 makes extensive contacts with the Oct-1 POU-specific domain. Interestingly, SNAP190 also makes DNA contacts within the enhancer. Together, this data suggests that the general transcription machinery can assist activator recruitment to weak enhancers. Moreover, SNAP190 occupies a similar trajectory between the Oct-1 POU-specific domain and POU-homeo domain as that observed for the B-cell specific co-regulator protein Oca-B when complexed with the Oct-1 POU domain on a high affinity octamer element. Thus, SNAP190 and Oca-B interactions with Oct-1 are likely mutually exclusive. Weak enhancer recognition by Oct-1 at the U1 promoter may be important for preventing tissue specific transcriptional squelching by inappropriate recruitment of co-regulatory proteins.

For romance readers everywhere

ACKNOWLEDGEMENTS

I want to thank my advisor Jim Geiger for guiding me along and being a great boss. I also want to thank our collaborator Dr. R. William Henry for providing a lot of insight into our joint project. The Geiger group started out working closely with Dr. Al Tulinsky's group. His advice and anecdotes about the old days were invaluable. Dr. Cory Abate – Shen supplied the protein and the plasmid for the Msx-1 protein. Otto Sorenson provided sound advice on Oct-1 purification. Special thanks to Dr. Craig Hinkley for doing the Oct-1 biochemical experiments. Certainly I could not have done without the counsel of Dr. Jorge Ríos, who really helped me out in the beginning with crystallographic software. How he put up with my never-ending questions I will never know. Thanks a lot Jorge.

The Geiger lab has always been a fun place to be, with the place virtually being overrun with women. I must say that it has been like a big family with group outings and parties. In the beginning there were three of us; Michelle, Tyra, and I. I remember those days fondly and I want to thank Michelle for her patience in teaching me the basics. Since then we have expanded to a rather large group but the presence of undergraduates has always been a bonus. The people I want to mention include virtually the whole lab: Michelle, Tyra, Marta, Sara, Erika, Xiangshu, Aimee, Katie, Laura, Elena, Mike, Chris, Paul, Keith, and Adam. Tyra, Elena, Aimee, Katie, and Laura have worked directly on the SNAP project and it has been a great help. A huge thanks goes out to Aimee, Katie, and Laura for setting up hundreds of boxes in the last year. Finally I want to acknowledge Marta Abad for all of our great times together in the lab and outside of the

lab. The place wouldn't have been the same without you "Smarta". Let me just say that Women Rule! I would also like to thank my husband Dan for putting up with my late nights and odd hours and for moving to Michigan to be with me.

I will always have fond memories of Michigan State but the ones that will stick with me are priceless. Jim Geiger dancing at my wedding. Michigan State University ice hockey games with the rest of the chemistry geeks. Marta's ice-skating attempt. The croquet match at Jim's house with various group members cheating. The Christmas party at Jim's. Erika's bowling technique comes to mind. Injury prone Sara. High speed Mike and Keith. Xiangshu's ticklishness. The knitting circle was great too – Aimee, Erika, Marta, Gwynne, Sara, and Keith. Additionally I will never forget Andrej crashing his tricycle at high speed into a pole at the synchrotron. I want to mention Emily Brown and thank her for a wonderful partnership in Science Theatre. It was a big part of my life and I will never forget you. Finally I must thank Gwynne Osaki for being a great friend throughout the years here at State and being so very helpful! I really relied on your advice in the early days and I am glad that you are moving on to better things. Good luck in your future endeavors. Last but not least I want to encourage Xiangshu because she is the hardest worker I ever met and has had the worst luck. Things will get better because you deserve it and you work for it. I will miss you so much when you graduate so please keep in touch and I wish you the very best in life.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	xi
ABBREVIATIONS	xvii
Chapter I: INTRODUCTION	
1.1 Transcriptional Regulation	1
1.2 Homeotic Genes	8
1.3 Msx –1	13
1.4 Oct –1	18
1.5 SNAPc	24
1.6 References	29
Chapter II: X-RAY CRYSTAL STRUCTURE DETERMINATION	
2.1 Msx-1/DNA Complex	35
2.1.1 Crystallization	36
2.1.2 Structure Determination	39
2.1.3 Molecular Replacement and Structure Refinement	44
2.1.4. Materials and Methods	47
2.2 Oct-1/SNAP 190/DNA Complex	54
2.2.1 Crystallization and data collection	54
2.2.2 Molecular Replacement and Structure Refinement	55
2.2.3 Materials and Methods	59
2.3 References	65
Chapter III: THE THREE DIMENSIONAL STRUCTURE OF THE MSX-1 HD / DNA COMPLEX.	
3.1 Overall structure of the Msx-1/DNA Complex	67
3.2 Protein/DNA recognition	72
3.2.1 Residue Q50	76
3.2.2 Residue A54	78
3.3 DNA minor groove and n-terminal arm interactions.	79
3.4 Hydration of the HD/DNA interface	81
3.5 Structure of the DNA	84
3.6 Msx-1 HD protein interactions	90
3.7 Conclusions	92
3.8 References	94
Chapter IV: THE THREE DIMENSIONAL STRUCTURE OF THE OCT-1 POU/ U1 OCTAMER/ SNAP190 TERNARY COMPLEX.	
4.1 Overall structure of the Oct-1/U1/SNAP190 ternary complex	96

4.2 SNAP190 assists Oct-1 binding to the U1 octamer	98
4.3 Structure of the Oct-1/ U1 octamer / SNAP190 peptide	101
4.4 Comparison of Oct-1 POU to other HDs and POU proteins	107
4.5 The OCA-B co-activator and SNAP190 general factor target Oct-1 similarly.	113
4.6 Cooperative promoter recognition and activation of human U1 transcription.	120
4.7 Conclusions	124
4.8 References	126

APPENDIX

Appendix 2.1 Protein Purification Buffers	128
Appendix 3.1 Msx-1 – DNA contacts compared to other monomer HD structures.	130
Appendix 4.1 Protein – Protein interactions between SNAP190 and Oct-1 POU.	134

LIST OF TABLES

CHAPTER I: INTRODUCTION

Table 1.1	Known homeodomain structures.	12
Table 1.2	DSE sequences found in a variety of snRNA promoters.	21

CHAPTER II: X-RAY CRYSTAL STRUCTURE DETERMINATION

Table 2.1	Crystal parameters for the Msx-1 HD/DNA Complex	36
Table 2.2	Statistics for the Msx-1 HD/DNA data sets	38
Table 2.3	Refinement statistics for Msx-1 HD/DNA complex	46
Table 2.4	DNA sequences used in Msx-1 Crystallization Trials.	50
Table 2.5	Iodinated DNAs used in crystallization trials.	52
Table 2.6	Crystal parameters for the Oct-1/U1/SNAP 190 peptide crystal	55
Table 2.7	Statistics for the ternary complex data collection	55
Table 2.8	Refinement statistics for Oct-1/U1 DSE/SNAP 190 (884-910)	57
Table 2.9	DNA Sequences used in ternary complex crystallization.	61
Table 2.10	Iodinated DNA Sequences used in ternary complex crystallization.	61
Table 2.11	Peptide Sequences used in Crystallization attempts.	63
Table 2.12	Ternary complexes that have been set up.	63

CHAPTER III: THE THREE DIMENSIONAL STRUCTURE OF THE MSX-1 HD/DNA COMPLEX

Table 3.1	Salt Bridges in Msx-1HD	71
Table 3.2	Conserved water table	82

CHAPTER IV: THE THREE DIMENSIONAL STRUCTURE OF THE OCT-1 POU/U1 OCTAMER/SNAP 190 TERNARY COMPLEX

Table 4.1	Summary of base specific contacts among POU proteins	111
Table 4.2	Conserved water comparison	113

LIST OF FIGURES

Images in this dissertation are presented in color.

CHAPTER I: INTRODUCTION

Figure 1.1	A. Schematic of a eukaryotic gene (top). B. Transcription initiation via Pol II only (bottom).	3
Figure 1.2	An example of a zinc finger DNA binding domain. The Zif268 protein – DNA complex at 1.6 Å (13). Zinc atoms are shown in green.	5
Figure 1.3	Leucine zipper element. GCN4 basic region leucine zipper / DNA complex at 2.9 Å.	6
Figure 1.4	Antennapedia HD/DNA structure at 2.4 Å looking down the recognition helix.	7
Figure 1.5	Consensus HD sequence from a compilation of 346 HD sequences. The bold residues are conserved in 80% of the sequences. The underlined residues denote every 10 th residue.	9
Figure 1.6	Repression regions of the full length Msx-1 protein.	17
Figure 1.7	Oct-1 POU / H2B DNA Complex. DNA is shown in green. The HD is shown in blue (left – 3 helices). The POU specific domain is shown in red (right – 4 helices).	20
Figure 1.8	DNA sequences of the Oct-1 H2B and the Pit-1 Prl-1P binding sites. There is a 4bp spacing between the two domains in the Pit –1 in addition to the radically different DNA sequence. Arrows indicate NH ₂ – terminal to COOH – terminal orientation of each domain. The broken lines show the disordered linker. DNA sequences are shown 5' to 3' on the top strand.	22
Figure 1.9	SNAPc – dependent Pol III transcription.	25
Figure 1.10	Pol II versus Pol III SNAPc – dependent transcription.	25
Figure 1.11	Schematic representation of the SNAP190 amino acid sequence showing functionally relevant domains.	26

Figure 1.12 Oct-1 mediated SNAPc transcription.	26
---	----

CHAPTER II: X-RAY CRYSTAL STRUCTURE DETERMINATION

Figure 2.1 Hanging drop vapor diffusion method. The reservoir contains precipitating agents that cause crystals to form.	37
Figure 2.2 Orthorhombic Crystal of Msx-1 HD/DNA4. The crystal has dimensions of $0.4 \times 0.2 \times 0.2 \text{ mm}^3$.	37
Figure 2.3 Ramachandran Plot of the Msx – 1 Homeodomain Residues.	46
Figure 2.4 Msx-1 Gel. M.W. Standards: Purple 42,000, Orange 32,000, Red 17,900, and Blue 7,200.	48
Figure 2.5 Crystals of the Msx-1 HD/DNA Complexes. A. DNA3 complex crystals $0.2 \times 0.2 \times 0.1 \text{ mm}^3$. B. DNA8 complex crystals $0.1 \times 0.1 \times 0.05 \text{ mm}^3$.	51
Figure 2.6 Oct-1/U1/SNAP 190 (884 – 910) with dimensions of $0.7 \times 0.05 \times 0.025 \text{ mm}^3$.	54
Figure 2.7 Ramachandran Plot for the Oct-1 POU and SNAP 190 residues. There are no residues in disallowed regions. The red areas indicate the most favorable regions and the yellow areas indicate additional allowed regions. The triangles represent glycines.	58
Figure 2.8 SDS – PAGE Gel of Oct-1 POU bound to Glutathione Beads. M.W. of purple is 42,000.	60
Figure 2.9 A. Oct-1/U1/SNAP 190 (52mer) grown in 20% PEG 6000 and 0.1 M Sodium Acetate pH 5.5 ($0.05 \times 0.1 \times 0.02 \text{ mm}^3$). B. Oct-1/U6/SNAP 190 (52 mer) grown in 7% PEG 6000 and 0.14 M Sodium Acetate pH 5.5 ($0.3 \times 0.2 \times 0.02 \text{ mm}^3$).	62

CHAPTER III: THE THREE DIMENSIONAL STRUCTURE OF THE MSX-1 HD / DNA COMPLEX

Figure 3.1 The three dimensional structure of the Msx-1 HD/DNA Complex. The view is looking down the recognition helix.	68
Figure 3.2 The hydrophobic core residues that are integral to protein stability.	69

Figure 3.3	The salt bridges that connect the three helices.	70
Figure 3.4	Overlay of the Msx-1 HD (cyan), Antennapedia (dk.blue), engrailed (yellow), and even-skipped (red). The DNA is from the Msx-1 structure and is present to provide a reference point.	73
Figure 3.5	Sequence Alignment of Homeodomains. In the majority sequence every 10 th residue is underlined. The Msx-1 residues involved in DNA recognition are denoted by an asterisk (*), and those involved in HD core stabilization are marked with a carat (^).	74
Figure 3.6	The contacts between the DNA and the protein in the complex. The major and minor groove DNA contacts are shown by squares and circles respectively. Dotted lines indicate hydrogen bonding, solid lines indicate hydrophobic interactions.	75
Figure 3.7	Simulated annealing omit map of the Q50 – water – DNA interaction, contoured at 1.5 σ . Picture was made with Setor.	77
Figure 3.8	The conserved water ring that surrounds A54 and fills the cavity present between the protein and the DNA backbone in this region.	80
Figure 3.9	Stereoview of the trajectory of the N – terminal arm of Msx-1. Hydrogen bonds are represented by dotted lines.	80
Figure 3.10	Conserved water network present in the homeodomain – DNA complexes studied. The waters are shown in gold.	83
Figure 3.11	A. Plot of DNA base roll as a function of DNA sequence. Horizontal dotted or dashed lines indicate average values for B-DNA. The sequence of the top strand only is shown. All parameters were calculated with the program Curves. B. Plot of DNA helical twist (B) as a function of HoxB1 DNA sequence. Horizontal dotted or dashed lines indicate average values for B-DNA. The sequence of the top strand only is shown.	85 86
Figure 3.12	Overall triple helix interaction for the stacking DNA. Strand 1 is in red and yellow while strand 2 is in purple and blue.	88
Figure 3.13	Triple helix schematic for the unusual helical interaction of the stacked DNAs.	89
Figure 3.14	The triple helix trio of Gua16:Cyt18 and Ade31, a previously unseen interaction in triple helix combinations.	89

CHAPTER IV: THE THREE DIMENSIONAL STRUCTURE OF THE OCT-1 POU/U1 OCTAMER/SNAP 190 TERNARY COMPLEX

- Figure 4.1 The three dimensional structure of Oct-1 POU/ U1 octamer/ SNAP190 peptide at 2.4 Å. The DNA is in silver, the Oct-1 POU protein in gold, and the SNAP190 is in green. This picture was created with the Ribbons program. 97
- Figure 4.2 DNA sequence used in crystallization (U1) – top strand shown only, compared to the H2B octamer. Base changes are shown in bold. The octamer sequence is numbered 1-8, with the equivalent base on the opposite strand indicated by a prime (A4' corresponds to the base paired to T4 for example). 99
- Figure 4.3 Electrophoretic mobility shift assays were performed using 0.1 ng (lanes 2 and 5) or 1 ng (lanes 3 and 6) of human Oct-1 POU-domain protein with DNA probes containing a human histone H2B (lanes 1-3) or U1 snRNA (lanes 4-6) octamer element. Lanes 1 and 4 contain the probes alone. The position of the POU complex is indicated (DNA/Oct-1). 99
- Figure 4.4 Electrophoretic mobility shift assays were performed using DNA probes containing a human histone H2B (lanes 1 and 2) or U1 snRNA octamer element (lanes 3-8) with 1 ng (lane 2) or 30 ng (lanes 4-6) of human Oct-1 POU-domain protein alone (lanes 2 and 4), with 10 µg SNAP190 peptide (lane 5) or with an equimolar amount of a control peptide (lane 6). Lanes 7 and 8 contain the SNAP190 or control peptides alone, respectively. Lanes 1 and 3 contain probe DNA alone. The position of the POU complex is indicated(DNA/Oct-1). 101
- Figure 4.5 SNAP190 Peptide sequence with the portion used in crystallization indicated. The residues in bold show sequence identity to the OCA-B peptide. 101
- Figure 4.6 Schematic representation of the protein/DNA contacts within the SNAP190/Oct-1/U1 octamer complex. The red contacts and arrows are the same in all three structures. Orange represents those found in SNAP190 and Oct-1 only. Pink represents those found in SNAP190 and OCA-B only. Blue contacts are unique to our structure. Those residues in black are the SNAP190 peptide contacts to the DNA. 103
- Figure 4.7 Hydrophobic interactions dominate the interaction between SNAP190 and Oct-1. A stereo view of the POU₅ interaction 104

with the SNAP190 C-terminal helix with the POU domain (gold) and SNAP190 (green) is shown. The view is looking down the SNAP190 helix. There are several hydrophobic interactions and two hydrogen bonds shown with dotted lines.

- Figure 4.8 A key determinant of transcriptional specificity within Oct-1 is well positioned for hydrogen bonding with SNAP190. Shown is a simulated annealing omit electron density map contoured at 1.8σ around SNAP190 K900, Oct-1 POU_s E7, and SNAP190 E904. SNAP190 E904 buttresses K900, accurately positioning it to make a critical salt bridge with POU_s E7. All of the protein is shown in dark blue. This figure was made using Setor. 106
- Figure 4.9 Sequence alignment of a few POU containing proteins. Oct-1 and Oct-2 are nearly identical with both being human proteins. Oct-4 and Oct-6 are mouse proteins. Pit-1 is a rat protein while Unc-86 is from *Caenorhabditis elegans*. Every 10th residue is marked with a (.) and alpha helices are denoted. Differences in sequence are indicated. 108
- Figure 4.10 Sequence alignment of homologous regions of SNAP190 and OCA-B. The region surrounding the SNAP190 peptide used in the crystallization is indicated. The homology between the two sequences is demoted with bold text. H indicates the helical region that is common to both structures. The green circles above (SNAP190) and below the (OCA-B) sequence alignment denote contacts made to DNA and the red circles donate contacts made to the Oct-1 POU domain. 115
- Figure 4.11 Overlay of the Oct-1/U1 octamer/SNAP190 (gold), Oct-1/H2B octamer (red), and the Oct-1/H2B octamer/OCA-B (dark blue) complex structures. The peptides have been colored independently with the SNAP190 peptide in green and the OCA-B in dark gray. Additional contacts between OCA-B and the Oct-1 POU_{HD} that are not observed in the SNAP190 structure rotate the POU_{HD} DNA recognition helix relative to its position in the other two structures (enlarged in next figure 4.12). 118
- Figure 4.12 Enlargement of the recognition helix region in which the OCA-B (blue) helix shifts by more than 3.5Å, interacting with the N terminal of the OCA-B peptide (gray). The two residues shown are one example of an interaction between the OCA-B HD and the OCA-B peptide. 119
- Figure 4.13 The arginine 49 interaction to the different base pair at position 4. A. The POU domain is in gold while the DNA 122

is in silver. The R49 moves down to make a closer contact to the G3 oxygen while it has a longer contact to A4'. This is the direct opposite of what is seen in the Oct-1/H2B (panel B) and Oct-1/H2B/OCA-B structures.

Figure 4.14 R102 collision in the case of OCA-B protein and U1 octamer 123
with position 6 altered. The U1 DNA is in silver and the OCA-B homeodomain is shown in blue. The Oct-1/H2B/OCA-B and Oct-1/U1/SNAP190 structures were overlaid and the result is that R102 can not make its normal contact with base A6 – in fact it is repelled by the G6 NH₂ group. In fact the whole R102 side chain is pushed out of the DNA groove and interacts with the protein and the DNA backbone.

ABBREVIATIONS

A - alanine
A/ADE – adenine
ant - antennapedia
APS – Advanced Photon Source

bp – base pair
BMP – Bone Morphogenic Proteins

C – cysteine
C/CYT – Cytosine
C terminal – carboxy terminal
C α - the alpha carbon in the peptide bond
CC – correlation coefficient
CCD – charge-coupled-device

D – aspartic acid
DNA – Deoxyribonucleic Acid
DNase – Deoxyribonucleosidase
DEAE – diethylaminoethyl cellulose
DTT – dithiothreitol
DSE – distal sequence element

E – glutamic acid
eve – even-skipped
EMSA – electrophoretic mobility shift assays
en – engrailed
Exd – Extradenticle

F – phenylalanine
F_{hkl} – structure factor

G – glycine
G/GUA – Guanine
GH-1 – growth hormone
GST – Glutathione S-Transferase

H – histidine
HTH – helix turn helix
HD – homeodomain
hox – homeobox gene in mammal
Hepes – N-[2-hydroxyethyl] piperazine-N'-[ethane sulfonic acid]

I – isoleucine
IPTG – Isopropyl-B-D-Thiogalactopyranoside
Isl – insulin gene enhancer protein
Iodo – iodine
ID – identification

K – lysine

L – leucine

M – methionine
msh – muscle specific homeobox (invertebrates)
Msx – Muscle specific (x denotes vertebrates)
MAT – Mating type
MCM – MADS-box transcription factor
MAD – multiple wavelength anomalous dispersion
 μg – microgram
 mm - millimeter
M.W. – molecular weight
MORE – more PORE
MPD – 2-methyl-2,4-pentanediol

N – asparagine
 ng - nanogram
 n-terminal (N-terminal) – amino terminal
ND – not deposited

Oct - octamer

P – proline
POL – polymerase
PDB – Protein Data Bank
 prd - paired
POU – comes from Pit Oct Unc proteins
POU_{HD} – POU homeodomain
POU_s – POU specific domain
Pbx – vertebrate ortholog of extradenticle
Pit - pituitary
Prl-1P – prolactin
PEG – polyethylene glycol
PORE – palindromic octamer factor Recognition element
PMSF – phenyl methyl sulfonyl fluoride
Pax – also of the paired class
PSE – proximal sequence element

Q – glutamine

R – arginine
RNA – Ribonucleic Acid
RNAP – Ribonucleic Acid Polymerase
Rmsd – root mean square deviation

S – serine
SNAP – small nuclear RNA activating protein
SNAPc - small nuclear RNA activating protein complex
snRNA – small nuclear RNA
SIR – single isomorphous replacement
SDS-PAGE – sodium dodecyl sulfate – poly acrylamide gel electrophoresis
SBC – Structural Biology Center

T – threonine
T/THY – Thymine
TBP – TATA Binding Protein
TFII(A) – Transcription Factor II (x)
Tris – 2-Amino-2-(hydroxymethyl)-1,3-propanediol
TDB – thrombin digestion buffer

U – Iodinated Uracil
Ubx - Ultrabithorax
URL – uniform resource locator (web address)

V – valine
VND/NK-2 – ventrolateral neurogenic anlage

W – tryptophan
Wat – water

Y – tyrosine

CHAPTER 1: INTRODUCTION

1.1 Transcriptional Regulation

Eukaryotic transcriptional regulation is a complex process that involves complex sets of regulatory elements and three different RNA polymerases, I, II, and III. The RNA Pol I system transcribes ribosomal RNA; the RNAP II system transcribes all messenger RNAs and some small nuclear RNAs (snRNAs); the RNAP III system transcribes all transfer RNAs and the rest of the snRNAs. Upstream of the initiation site there may be different combinations of specific DNA sequences, each of which is recognized by a corresponding site – specific DNA – binding protein. The upstream regulatory elements can be divided into two main categories, promoter and enhancer elements. A schematic of this is shown in Figure 1.1A. Within the promoter element is a common region for binding general transcription factors (GTFs) seen in Figure 1.1B.

RNA Polymerase I and III transcribe only a limited set of genes while there is a huge variety of Pol II transcribed genes that encode proteins. However the activities of Pol I and III dominate cellular transcription, combining for 80% of RNA synthesis in growing cells (1). Pol I is the most specific polymerase in that it exhibits stringent species specificity. The promoters from different species vary widely in sequence but have a similar layout and function. Pol I works with the required upstream binding factor (UBF) and selectivity factor (SL1) for basal expression (2). SL1 is a complex consisting of TATA – binding protein (TBP) and TBP – associated Pol I specific factors (TAF-1). Pol II and Pol III promoters are more homologous and the factor determining polymerase specificity is the presence of a TATA box. Usually the Pol II system recognizes the

TATA box while the Pol III system does not, in the case of snRNA genes this is exactly the opposite. Sometimes Pol specificity is determined by exact spacing between certain elements in the gene or the exact sequence of the promoter (3). Pol II and III both require TBP and the GTFs to activate transcription. Pol I, II, and III may also utilize different sets of enhancer elements that the GTFs or other proteins recognize.

The enhancer elements contain specific sequences recognized by transcription factors. These enhancers can be as far away as 20,000 base pairs (bp) from the promoter they are controlling. Therefore there is a great need for many proteins to work together over long stretches of DNA for efficient gene expression. Each gene has the same DNA control sequences, but not every cell has the same set of DNA binding proteins. Cell – specific gene expression depends on the complement of transcription factors present at any one time in the cell. There are two classes of transcription factors, those required for the expression of all structural genes transcribed by a given polymerase and specific ones that are found in a restricted range of cell types and are responsible for the expression of cell – type specific proteins. The GTFs include TBP and assorted transcription factors TFIIA – F (Figure 1.1B). RNA polymerases can not recognize and bind to promoters on their own so these general factors recruit it to the promoter (4-9).

Different transcription factors show specificity for various DNA sequences. Over 80% of the known transcription factors have one of three distinct structural motifs. These motifs provide a three - dimensional scaffold that dictates proper positioning of the protein against the DNA. The three main motifs are zinc fingers, leucine zippers, and helix turn helix domains. Classic zinc fingers involve two histidines and two cysteines

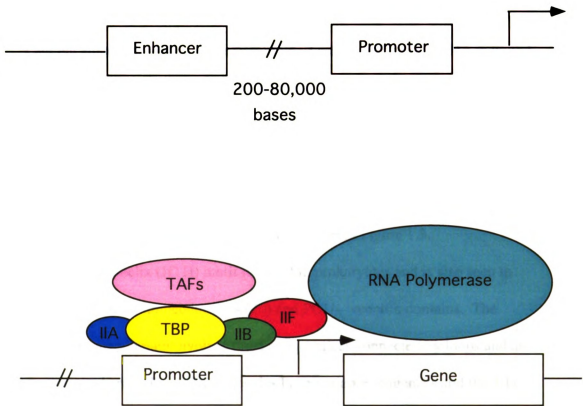


Figure 1.1. A. Schematic of a eukaryotic gene (top). B. Transcription initiation via Pol II only (bottom).

coordinated to a zinc atom that creates a loop between the residues and forms a DNA binding region. This finger motif is repeated in tandem to recognize DNA sequences of different lengths. An example of this is shown in Figure 1.2.

Leucine zippers provide a dimerization interface through which DNA is bound. In the leucine zipper there are leucine residues every 7th position and these leucines come from two alpha helical monomers. These two helices come together in a Y shape and interact with DNA as a dimer. The leucine zipper contains a 4-3 heptad repeat of hydrophobic and non-polar residues that pack together. The two alpha helices grip the major groove of the DNA like forceps. This can be seen in Figure 1.3.

The helix turn helix (HTH) motif is found in prokaryotes and is also seen in eukaryotes in the form of homeodomains (10) and POU – specific domains. The structure of the homeodomain involves three alpha helices connected by loops and an extended n-terminal arm. The HTH molecules have variable sequences and the 3-D globular folds are variable. The recognition helix is shorter, the mode of docking on the DNA is not conserved, and binding often occurs as homodimers (11). Most homeodomains recognize a 5' –ATTA – 3' binding site, and have four invariant residues in the recognition helix. The recognition helix makes base specific contacts in the major groove and is involved in fixing the reading head in the major groove. The flexible n – terminal arm tracks across the minor groove of the DNA (12). An example of a HD/DNA structure is shown in Figure 1.4. Two homeodomain containing proteins will be the focus of this thesis.

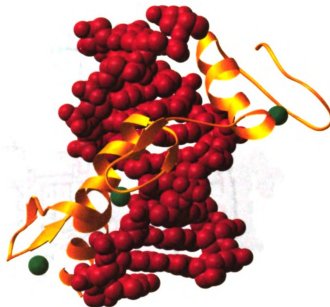


Figure 1.2 An example of a zinc finger DNA binding domain. The Zif268 protein – DNA complex at 1.6 Å (13). Zinc atoms are shown in green.



Figure 1.3 Leucine zipper element. GCN4 basic region leucine zipper / DNA complex at 2.9 Å (14).

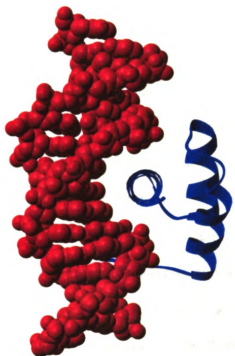


Figure 1.4 Antennapedia HD/DNA structure at 2.4 Å looking down the recognition helix (15).

1.2. Homeotic Genes

Homeotic gene families have been extensively studied because of their fundamental role in development. These genes specify the body plan, pattern formation, cell fate, and are involved in genetic control of development. They are arranged in close complexes and are expressed in the same order as they are arranged on the chromosome (colinearity rule). Homeotic genes are master control genes that share a common 180 base pair (bp) sequence referred to as the homeobox which encodes a 60 amino acid region called the homeodomain. There is a high degree of evolutionary conservation among homeodomains in eukaryotes. The homeodomain represents the DNA-binding domain of larger homeodomain proteins and allows sequence - specific recognition of sets of target genes by the homeodomain proteins.

Homeodomains are highly conserved across many species. The consensus sequence derived from a compilation of 347 homeodomains is given in Figure 1.5. There are seven positions that are occupied by the same amino acid in more than 95% of the sequences, and 10 others are conserved in more than 80% of the sequences, while in 12 additional positions only two different amino acids are found in more than 80% of the sequences. These highly conserved amino acids define the homeodomain (16).

Homeodomain sequences can be subdivided into classes on the basis of several criteria: sequence identity, sequence similarity in flanking regions, organization into gene clusters, association with other sequence motifs, and positions of introns. Even - skipped (eve) functions as a segmentation gene and is located on a different chromosome from the

RRRKRTAYTRYQLLELEKEFHFNRYLTRRRRIELA
HSLNLTERQVKIWFQNRMRMKWKKEN

Figure 1.5 Consensus HD sequence from a compilation of 346 HD sequences (16). The bold residues are conserved in 80% of the sequences. The underlined residues denote every 10th residue.

homeotic complexes. The engrailed class (en) has 4 highly conserved protein segments outside of the homeodomain. The paired class (prd) has the homeodomain associated with a second DNA - binding domain. All of the prd homeodomains have a serine at position 50. The paired like class (prd-like) has a sequence similarity to prd. The POU proteins were first isolated as transcription factors but later were found to contain homeodomains. The known POU genes have a cysteine at position 50. There are other classes classified in the same manner (16).

Most of the DNA sequences that interact favorably with homeodomains contain a tetranucleotide (ATTA or on the other strand 5' - TAAT -3'), which is the core motif. Mutational analysis confirmed that each conserved base pair in the ATTA core contributed to high binding affinity (17). There is also a preference for certain bases preceding the ATTA core depending upon the amino acid at position 50. The amino acid at position 50 has been shown to be involved in discriminative recognition of distinct classes of DNA sequences (16). In the *Ant* complex, the methionine at position 54 contacts bp 3 preceding the core motif. Depending upon the length and identity of the side chain at position 54 it should contact bp 3 or 4 and contribute to the homeodomain's sequence preference (12). Little is known about the role of the other particular amino acid side chains in the role of discrimination of binding sequences.

The high affinity binding sites contain the ATTA motif. The medium affinity binding sites have strong sequence conservation but often do not have the ATTA core motif. The precise DNA sequence of these sites leads to functional specificity of the homeodomain proteins *in vivo*. Differences found in N-terminal sequences between various homeodomains translate into slightly different binding preferences, which then

contribute to different biological functions. The N-terminal arm is involved in selective protein-protein interactions and this does contribute to the functional specificity of homeotic proteins (16). DNA binding specificity combined with the association with other transcription factors could account for functional specificity.

There are many HD structures that have been solved by X-Ray Crystallography and NMR (12, 15, 18-44). Table 1.1 lists all of the current structures of homeodomains. The consensus mode of binding is that of the recognition helix to the major groove and the n – terminal arm to the minor groove. In quite a few cases there is a second domain that is attached to the homeodomain and acts like a clamp on the opposite side of the DNA.

The structures listed in Table 1.1 came from the protein data bank found at the following URL: <http://www.rcsb.org/pdb>. The PDB accession number gives access to structural data for the protein or complex listed. ND refers to coordinates that were not deposited so I have listed the reference for the paper that discusses these structures. There may be other structures that have not been deposited but to my knowledge this is a complete list of the homeodomains that have been solved to date. It is a requirement of most publications that coordinates be deposited before a paper is published.

Table 1.1 Known homeodomain structures.

	Protein	Resolution/ NMR	Reference/ PDB Accession ID
HD Monomers	Rat Insulin Gene Enhancer Protein Isl-1	NMR	(28)1BW5
	Pbx	NMR	(45)1DU6
	Engrailed	NMR	(33)1ENH
	MATa1	NMR	(46)1F43
	Thyroid transcription factor (TTF-1)	NMR	(47)1FTT
	Fushi Tarazu	NMR	(31)1FTZ
	Oct-2 POU	NMR	(38)1HDP
	Rat Liver LFB1/HNF1	NMR	(29)2LFB
	Oct-3 POU	NMR	(48)1OCP
	VND/NK2	NMR	(49)1QRY
HD/DNA complexes	Antennapedia/DNA	NMR 2.4	(50)1AHD (15) 9ANT
	MAT α 2/DNA	2.7	(12)1APL
	MATa1/ MAT α 2/DNA	2.5	(20)1AKH (19)1YRN
	MAT α 2/MCM1/DNA	2.25	(34)1MNM
	MATa1/MAT α 2-3A/DNA	2.3	(in press)1LE8
	Pbx1/HoxB1/DNA	2.35	(24)1B72
	Ubx/Exd/DNA	2.4	(23)1B8I
	Engrailed	2.8	(25)1HDD
		2.2	(22)3HDD
	Engrailed mutant	2.0	(51)1DUO
	Paired/DNA	2.0	(18)1FJL
	Even-skipped/DNA	2.0	(21)1JGG
	Msx-1/DNA	2.2	(52)1IG7
	VND/NK-2/DNA	NMR	(53)1NK2
	Oct-1 POU/DNA	3.0	(36)1OCT
	Oct-1 POU/Oca-B/DNA	3.2	(35)1CQT
	Oct-1 POU/DNA(MORE)	1.9	(42)1E3O
	Oct-1 POU/DNA(PORE)	2.7	(42)1HFO
	Pit-1/DNA	2.3	(43)1AU7
	Pit-1/GH-1 DNA	3.0	(44)ND
	Pit-1/Prl-1P DNA	3.05	(44)ND

1.3 Msx – 1

The study of the molecular processes that regulate mouse embryonic development led to the identification of numerous genes whose protein products control gene expression during embryogenesis. These are transcriptional regulatory proteins that establish and maintain the appropriate patterns of spatial and temporal gene expression (54). These genes share the conserved homeobox which encodes the homeodomain. The *hox* gene family was studied because of its similarity to the *Drosophila* homeotic gene, *antennapedia*. The *hox* genes are expressed early in the developing mouse embryo, from 7.5 to 8.5 days in overlapping patterns. The colinearity of chromosome organization mentioned earlier has been conserved from *Drosophila* to *Homo sapiens* and might be a molecular code that provides positional information during development. The murine gene *hox 7.1* (Msx-1) is a member of a small gene family and its sequence is more closely related to the *Drosophila* gene *msh* (muscle-specific homeobox) than to the *antennapedia* gene. The *hox 7.1* gene is expressed throughout the developing neural tube and the anterior boundaries of expression extend to the presumptive midbrain (54).

The DNA binding specificity of the murine homeodomain protein *hox 7.1* was determined and many of the selected sites were flanked by Gua or Cyt nucleotides (54). The consensus binding site ACTAATTG was identified for *hox 7.1*. The nucleotides on the 5' end of the TAAT did not affect binding greatly when altered, but when the nucleotides on the 3' end were altered there was a significant reduction in the binding activity of the protein. Substitutions within the TAAT core abolished the activity for the protein (54). The amino acid side chain at position 50 distinguished binding to sites that had either a C or TG flanking the TAAT on the 3' end.

The role of *Msx* genes in tissues has been examined. Three subclasses have been determined; the vertebrate *Msx-1*, *Msx-2*, *Msx-3*, and the invertebrate *msh* genes. The early expression of these genes in the differentiation of diverse organs suggests they have a fundamental role in development (55, 56). Determining the functionality of these subclasses will help in understanding the molecular processes that distinguish the development of individual tissues. *Msx-1* in the mouse has shown expression in the uterus, cervix, vagina, uterine wall, and other reproductive organs. It has also shown activity in the lateral mesoderm, dorsal ectoderm, neural plate, dorsal region of the brain, cranial neural crest cells, facial processes, tooth germs, eye, ear, nose, and many other tissues (57). Mutations provide the most direct evidence concerning the function of the *Msx* genes. Mice that are homozygous for a targeted insertion in *Msx-1* fail to form teeth and have craniofacial abnormalities. They also have a cleft palate which could provide a model for some human cleft palate syndromes (57).

Other mutations in *Msx-1* affect the development of bones in the head. It appears that only some bones require *Msx* activity for their formation. The *Msx* genes might function in determining the characteristic shapes of specific bones. The control of bone formation also involves bone morphogenic proteins (BMP's), but this process is poorly understood. The human *Msx* proteins have been isolated and their positions on the chromosome have been identified. The mouse studies could be extended to the human *Msx-1* and its role in congenital malformations could be established (57). The human, chicken, and mouse *Msx-1* homeodomains are identical. Seventeen out of the 293 amino acids for the human full length *Msx-1* protein are different from the mouse *Msx-1*.

The role of Msx-1 in transcriptional repression was investigated (58). The common feature among transcriptional regulatory proteins is that they function to transduce cellular signaling events to changes in gene expression. Many transcription factors were identified based on their connection to abnormal cellular processes. The functionality of the homeodomain *in vivo* is thought to be as a scaffold for protein-protein interactions. Msx-1 has an N-terminal region that contains a high percentage of alanines, glycines, and prolines. These residues are frequently associated with transcriptional regulatory domains. The C terminal region is also rich in alanines and both the N and C termini have a high percentage of hydrophobic residues.

Though very few genes have been identified that are regulated by the Msx proteins in organisms, both *in vivo* and *in vitro* studies show that these proteins act as potent transcriptional repressors (55, 58-62). Surprisingly, this transcriptional repression activity is independent of Msx DNA binding sites. Instead, repression activity appears to be transduced via protein-protein interactions with basal machinery factors and with other homeodomain proteins. In fact Msx-1 binds tightly and specifically to TBP while Msx-2 interacts specifically with TFIIF (60, 62). Mutations that abrogate these interactions correlate with loss of repression activity and are localized to the n-terminal arm of the homeodomain. Msx-1 also directly interacts with members of three other homeodomain protein families, Dlx (Dlx 2 and Dlx 5), LIM(Lhx2), and Pax (Pax3) (61, 63, 64). In each case, the interactions are localized to the homeodomain and in each case the interaction abrogates DNA binding of both proteins, and neutralizes the transcriptional effects of each. It thus appears that these interactions may lead to functional antagonism *in vivo* in tissues where there is coexpression of Msx-1 and these other proteins. In

Figure 1.6 the regions involved in repression are shown along with the homeodomain (58).

It turns out that all domains of Msx-1 are required for maximal repressor function. The observation that the repression occurs in the absence of the homeodomain DNA binding sites does not rule out the possibility that DNA binding contributes to its function as a transcriptional regulator. Sequestering the homeodomain by DNA binding could preclude its transcriptional function (58).

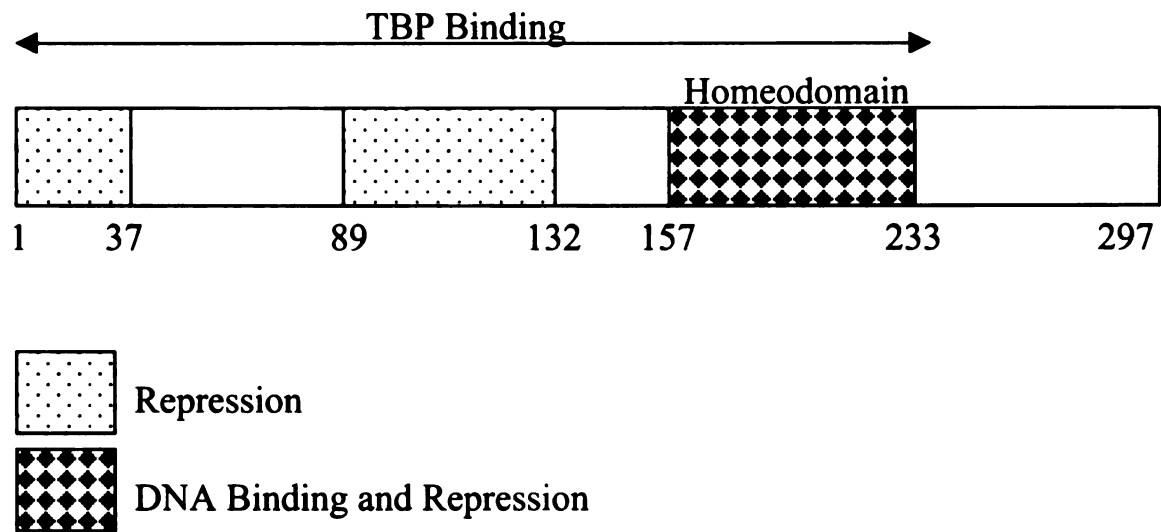


Figure 1.6 Repression regions of the full length Msx-1 protein.

1.4 Oct –1

Oct-1 is a ubiquitous protein that interacts directly with the basal transcriptional machinery. It is a member of the POU domain – containing transcriptional activators involved in the regulation of a variety of genes. The POU domain is the DNA binding domain and consists of two domains, a canonical homeodomain (POU_H) and a helix-turn-helix like POU-specific domain (POU_S)(65). These two domains are connected by a flexible linker that varies in length. Crystal structures of several POU domain/DNA complexes, including Pit-1(43, 66) and Oct-1 (67-69) have been determined and confirm this model. The POU domain proteins typically bind either a TAATGARAT (R=purine) motif or an eight base pair motif called an octamer sequence. The canonical octamer sequence is utilized in several promoters and has the sequence ATGCAAAT. While the POU_S domain binds the ATGC sequence, the POU_H domain interacts with the AAAT sequence. Different POU domain proteins can bind to octamer sequences that vary in relative orientation, sequence and spacing between the two basepair half sites (65). The Oct-1 POU domain bound to the H2B octamer site has been solved to a resolution of 3.0Å (Figure 1.7).

The DNA binding specificity of the Oct-1 POU protein has been studied in great detail. The Oct-1 protein binds to an octamer site present in the distal sequence element (DSE) which lies upstream of the transcription start site. The Oct-1 POU protein binds optimally to the site ATGCAAAT with no spacing between the half sites. A one bp insertion between the two half sites decreases the binding constant 10-fold while a 2 – 3 bp insertion causes a 100 fold decrease in the binding affinity (67). The Oct-1 POU

protein can bind to sequences that have position 4 altered (C to T) because this does not interrupt the interaction of Arg 49 (POU_S) to that base (70). Arg 49 also does not seem to mind the mutation of G3/C4 to T3/T4 and seems capable of adapting to base changes in this area (71). The base at position 5 can either be an A or a T with no difference in the binding affinity between the two because the HD binding site is TAAT. Oct-1 POU also recognized oligonucleotides containing the homeodomain binding site (TAAT) with higher affinity than the free POU_{HD} suggesting involvement of POU_S in determining binding affinity to these sites. Altering the flanking bases of the octamer showed no adverse affects on binding constants.

The DNA octamer sites found in a variety of snRNA promoters are very similar. The most common one is the H2B site ATGCAAAT which binds with high affinity to Oct-1. Table 1.2 lists the octamer sequences from some DSEs found in snRNA promoters. There are not too many changes from the standard H2B sequence. There are virtually no cases of extra basepairs between the two half sites or with the orientation of the two half sites differing. There are cases where the orientation of the octamer sequence relative to the start site of transcription is variable (U6). Some of the octamers contain significant base changes that are detrimental to Oct-1 binding (70, 72, 73). Interestingly there has been some experimentation with different DNA sequences that contain totally different binding sites for the POU proteins. In these cases the orientation of the POU specific domain relative to the homeodomain is flipped and its position and spacing are altered (65). This is mainly due to the differences in DNA sequences, the dimerization interface, and the length of the linker which in the case of Pit-1 is a short 15 amino acids. This is shown in a schematic in Figure 1.8.

protein can bind to sequences that have position 4 altered (C to T) because this does not interrupt the interaction of Arg 49 (POU_s) to that base (70). Arg 49 also does not seem to mind the mutation of G3/C4 to T3/T4 and seems capable of adapting to base changes in this area (71). The base at position 5 can either be an A or a T with no difference in the binding affinity between the two because the HD binding site is TAAT. Oct-1 POU also recognized oligonucleotides containing the homeodomain binding site (TAAT) with higher affinity than the free POU_{HD} suggesting involvement of POU_s in determining binding affinity to these sites. Altering the flanking bases of the octamer showed no adverse effects on binding constants.

The DNA octamer sites found in a variety of snRNA promoters are very similar. The most common one is the H2B site ATGCAAAT which binds with high affinity to Oct-1. Table 1.2 lists the octamer sequences from some DSEs found in snRNA promoters. There are not too many changes from the standard H2B sequence. There are virtually no cases of extra basepairs between the two half sites or with the orientation of the two half sites differing. There are cases where the orientation of the octamer sequence relative to the start site of transcription is variable (U6). Some of the octamers contain significant base changes that are detrimental to Oct-1 binding (70, 72, 73).

Interestingly there has been some experimentation with different DNA sequences that contain totally different binding sites for the POU proteins. In these cases the orientation of the POU specific domain relative to the homeodomain is flipped and its position and spacing are altered (65). This is mainly due to the differences in DNA sequences, the dimerization interface, and the length of the linker which in the case of Pit-1 is a short 15 amino acids. This is shown in a schematic in Figure 1.8.

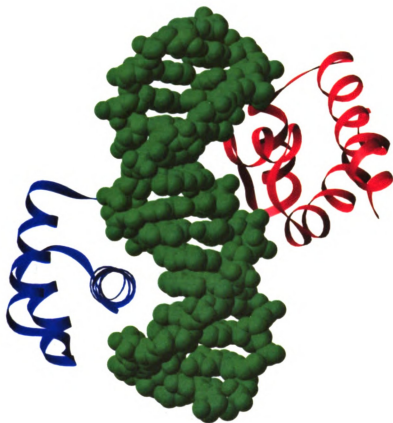
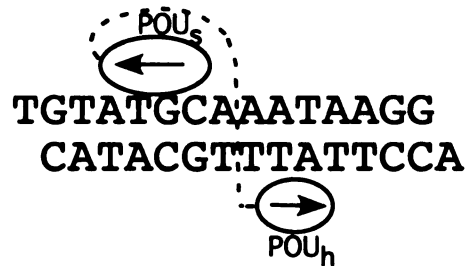


Figure 1.7 Oct-1 POU / H2B DNA Complex (36). DNA is shown in green. The HD is shown in blue (left – 3 helices). The POU specific domain is shown in red (right – 4 helices).

Table 1.2 DSE sequences found in a variety of snRNA promoters.

Promoters	Octamer sequence
H2B Octamer	ATGCAAAT
Human Genes	
U1	ATGTAGAT
U2	ATGCAAAT
U3	ATGCTAAT
U4B	ATTAGCAT
U4C	ATTTGCAT
U11	ATTTGCAT
U6	ATTTGCAT
7SK	ATTTAGCAT
	TTTAGCAT
	ATTTGCTAT
H1	ATGGAATT
	ATTTGCAT
MRP/Th	ATTTGCAT
HY3	ATGCAAAT
Herpesvirus saimiri	
HSUR1	ATTTGCAT
HSUR2	ATTTGAAT
HSUR3	ATTTGAGT
HSUR4	ATGCAAAT
	ATTTGAAT
HSUR5	ATTTGAAT
Xenopus genes	
U1a (major)	ATGTAAAC
U1b (major)	ATGCAAAT
U2	ATGCAAAT
U5	ATTTGCAT
U6	ATTTGCAT
Chicken genes	
U1 52A	ATGCAAAT
U1 52B	ATGCAGAT
U1 52C	ATGCAAAT
U1 2.5	ATGCAAAT
U2	ATGCAAAT
U4B	CTTTGCAT
U4X	ATTACCAT

Oct-1 H2B DNA



Pit-1 Prl-1P DNA

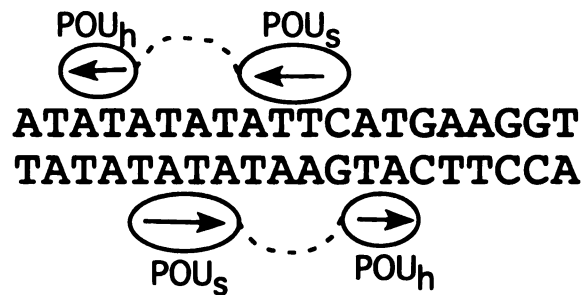


Figure 1.8 DNA sequences of the Oct-1 H2B and the Pit-1 Prl-1P binding sites. There is a 4bp spacing between the two domains in the Pit -1 in addition to the radically different DNA sequence. Arrows indicate NH₂ – terminal to COOH – terminal orientation of each domain. The broken lines show the disordered linker. DNA sequences are shown 5' to 3' on the top strand.

The linker between the POU_S and POU_{HD} is variable in length for many POU proteins. Pit-1 (rat) for example has a short 15 amino acid linker while the Oct-1 human protein has a 24 amino acid linker. Binding to the natural octamer site by Oct-1 requires a minimal linker of 10 – 14 amino acids (74). Varying the length of the linker from 2 to 37 amino acids showed that the smaller linkers (<23 aa) had a lower affinity for the octamer site. However, lengthening the linker does not compensate for the low affinity that the Oct-1 protein has for DNA with extra spacing between the half sites (ATGC—AAAT). Klemm and Pabo showed that the isolated POU_S and POU_{HD} could bind cooperatively even in the absence of the linker even though the two parts of the protein do not contact each other (75). Overlapping DNA contacts near the center of the octamer site may mediate the cooperativity and explain why the non-spaced octamer is the preferred site. The linkers among the POU proteins do not show any sequence homology except for one glutamate that when mutated to a lysine led to a 2.5 fold reduction in affinity for the octamer binding site (74).

1.5 SNAPc

SNAPc is the small nuclear RNA activating protein complex. The SNAPc complex has five polypeptides, SNAP19, SNAP43, SNAP45, SNAP50, and SNAP190 (76-80). SNAPc dependent genes all contain the PSE (proximal sequence element) that is specifically recognized by the SNAPc complex. The most interesting aspect of SNAPc mediated transcription is that either RNAP II or RNAP III can be recruited depending upon promoter context. A schematic of the SNAPc dependent Pol III transcription is shown in Figure 1.9 (U6 for example). Promoters that are recognized by Pol III have both a TATA box and a PSE while promoters transcribed by Pol II only have the PSE (U1) (Figure 1.10). Nevertheless, TBP is required for transcription of both classes of genes.

TBP interacts with the N-terminal part of SNAP190, the largest part of the SNAP_c complex. A schematic of the SNAP190 protein is shown in Figure 1.11. TBP and SNAP_c have built in mechanisms that prevent them from binding to DNA efficiently on their own. The N- terminus of TBP inhibits its binding to TATA boxes while the C-terminal portion of SNAP190 inhibits its binding to the DNA (81, 82). These two proteins dissociate slowly from DNA so perhaps this prevents binding to inappropriate sites. Interestingly, the two inhibitory portions of TBP and SNAP190 are required for cooperative binding of TBP with SNAP_c and SNAP_c with Oct-1.

Oct -1 interacts directly with SNAP190, the largest part of the SNAPc complex and activates transcription (Figure 1.12). The Oct-1 interacting region is also depicted in the schematic in Figure 1.11. The interaction between Oct-1 and SNAP190 has been

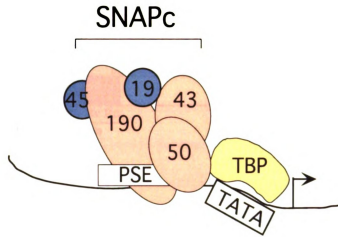


Figure 1.9 SNAPc – dependent Pol III transcription.

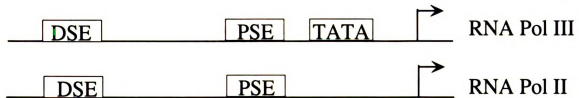


Figure 1.10 Pol II versus Pol III SNAPc – dependent transcription.

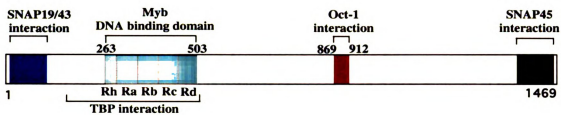


Figure 1.11 Schematic representation of the SNAP190 amino acid sequence showing functionally relevant domains.

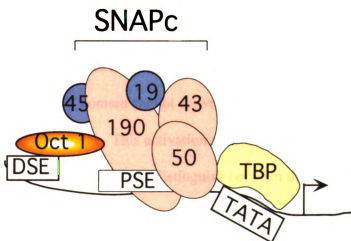


Figure 1.12 Oct-1 mediated SNAPc transcription.

localized to a small region of the protein (800-930). A smaller part of this region (869-912) is homologous to the first 63 amino acids of Oca-B (OBF-1 or Bob-1), a B-cell specific coactivator that associates with Oct-1 bound to octamer motifs and increases transcription from immunoglobulin promoters (83). Binding of OBF-1 to the Oct-1/DNA complex is sensitive to changes in octamer sequence as an Ade is required at both positions 5 and 6 (84, 85). There is no data regarding DNA sequence specificity of SNAP190 binding to the Oct-1/DNA complex. This suggests that some remarkable processes are conserved between Pol II and Pol III transcriptional activation. Apparently direct interaction between a basal initiation factor (SNAPc) and an activator (Oct-1) bypasses the need for a coactivator (OBF-1) in some contexts. This is also perhaps the best characterized, functionally critical direct interaction between a basal initiation factor and a transcriptional activator.

Although capable of activating immunoglobulin genes in lymphoid cells, Oct-1 preferentially activates transcription of snRNA genes (86). In addition to the POU domain, Oct-1 also contains an activation domain, but surprisingly, the Oct-1 POU DNA binding domain is sufficient to maintain robust activation of snRNA gene transcription. In contrast, the Pit-1 POU domain cannot activate snRNA transcription even though these two POU domains are similar. This activation specificity served as the platform to identify molecular discriminators that distinguish between the two POU domains. For example, mutational analysis of the Oct-1 and Pit-1 POU domains revealed that a single amino acid at position 7 of helix 1 within the POU₅ domain contributes to this activator specificity. The corresponding positions within Oct-1 and Pit-1 encode a glutamic acid and an arginine, respectively, and changing this amino acid within the Oct-1 POU₅

domain from the glutamic acid to an arginine (Oct-1 POUE7R) abolished the ability of the Oct-1 POU domain to activate snRNA transcription (83).

Both basal and activated SNAP-dependent transcription can be reconstituted from snRNA promoters *in vitro*. However, Oct-1 dependent activated transcription requires the DSE to be moved within a few base pairs of the PSE on naked DNA templates (87). The native U6 promoter sequence can also be used in activated transcription, but only after being reconstituted with nucleosomes. This reconstitution results in the positioning of a nucleosome between the PSE and DSE on this promoter. A similarly positioned nucleosome can also be seen on the U1 promoter. This positioned nucleosome can also be detected *in vivo*. From these data it appears that a positioned nucleosome is required to fold the promoter DNA such that the DSE and PSE sequences are close enough for Oct-1 and SNAP190 to interact and for activated transcription to occur (88).

1.6 References

1. Paule, M. R., and White, R. J. (2000) *Nuc. Acids Res.* 28, 1283-98.
2. Jacob, S. T., and Ghosh, A. K. (1999) *J. Cell Biochem. Suppl.* 41-50.
3. Hernandez, N. (2001) *J. Biol. Chem.* 276, 26733-6.
4. Reinberg, D., Orphanides, G., Ebright, R., Akoulitchev, S., Carcamo, J., Cho, H., Cortes, P., Drapkin, R., Flores, O., Ha, I., Inostroza, J. A., Kim, S., Kim, T. K., Kumar, P., Lagrange, T., LeRoy, G., Lu, H., Ma, D. M., Maldonado, E., Merino, A., Mermelstein, F., Olav. (1998) *Cold Spring Harb. Symp. Quant. Biol* 63, 83-103.
5. Paule, M. R., and White, R. J. (2000) *Nuc. Acids Res.* 28, 1283-1298.
6. Chedin, S., Ferri, M. L., Peyroche, G., Andrau, J. C., Jourdain, S., Lefebvre, O., Werner, M., Carles, C., Sentenac, A. (1998) *Cold Spring Harbor Symp. Quant. Biol.* 63, 381-389.
7. Hampsey, M., and Reinberg, D. (1999) *Curr. Opin. Gen. Develop.* 9.
8. Bell, S. D., and Jackson, S. P. (1998) *Cold Spring Harbor Symp. Quant. Biol.* 63, 41-51.
9. Reeder, R. H. (1999) *Prog. Nucleic Acid Res. Mol. Biol.* 62, 293-327.
10. Desplan, C., Theis, J., and O'Farrell, P. H. (1988) *Cell* 54, 1081-90.
11. Gehring, W. J., Quan, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994) *Cell* 78, 211-223.
12. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) *Cell.* 67, 517-28.
13. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) *Structure* 4, 1171-80.
14. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell* 71, 1223-37.
15. Fraenkel, E., and Pabo, C. O. (1998) *Nat. Struct. Biol.* 5, 692-7.

16. Gehring, W. J., Affolter, M., and Burglin, T. (1994) *Annu. Rev. Biochem.* 63, 487-526.
17. Laughon, A. (1991) *Biochemistry* 30, 11357-67.
18. Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) *Cell*. 82, 709-19.
19. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) *Science*. 270, 262-9.
20. Li, T., Jin, Y., Vershon, A. K., and Wolberger, C. (1998) *Nuc. Acids Res.* 26, 5707-18.
21. Hirsch, J. A., and Aggarwal, A. K. (1995) *EMBO J.* 14, 6280-6291.
22. Fraenkel, E., Rould, M. A., Chambers, K. A., and Pabo, C. O. (1998) *J. Mol. Biol.* 284, 351-61.
23. Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999) *Nature* 397, 714-9.
24. Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., and Wolberger, C. (1999) *Cell* 96, 587-97.
25. Kissinger, C. R., Liu, B. S., Martin Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) *Cell*. 63, 579-90.
26. Gruschus, J. M., Tsao, D. H., Wang, L. H., Nirenberg, M., and Ferretti, J. A. (1999) *J. Mol. Biol.* 289, 529-45.
27. Guntert, P., Qian, Y. Q., Otting, G., Muller, M., Gehring, W., and Wuthrich, K. (1991) *J. Mol. Biol.* 217, 531-40.
28. Ippel, H., Larsson, G., Behravan, G., Zdunek, J., Lundqvist, M., Schleucher, J., Lycksell, P. O., and Wijmenga, S. (1999) *J. Mol. Biol.* 288, 689-703.
29. Schott, O., Billeter, M., Leiting, B., Wider, G., and Wuthrich, K. (1997) *J. Mol. Biol.* 267, 673-83.
30. Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., and Wuthrich, K. (1989) *Cell*. 59, 573-80.
31. Qian, Y. Q., Furukubo Tokunaga, K., Resendez Perez, D., Muller, M., Gehring, W. J., and Wuthrich, K. (1994) *J. Mol. Biol.* 238, 333-45.

32. Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T., and Pabo, C. O. (1997) *Structure* 5, 1047-54.
33. Clarke, N. D., Kissinger, C. R., Desjarlais, J., Gilliland, G. L., and Pabo, C. O. (1994) *Protein Science* 3, 1779-1787.
34. Tan, S., and Richmond, T. J. (1998) *Nature* 391, 660-6.
35. Chasman, D., Cepek, K., Sharp, P. A., and Pabo, C. O. (1999) *Genes Dev.* 13, 2650-7.
36. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) *Cell*. 77, 21-32.
37. Morita, E. H., Shirakawa, M., Hayashi, F., Imagawa, M., and Kyogoku, Y. (1995) *Protein Sci.* 4, 729-39.
38. Sivaraja, M., Botfield, M. C., Mueller, M., Jancso, A., and Weiss, M. A. (1994) *Biochemistry* 33, 9845-55.
39. Ceska, T. A., Lamers, M., Monaci, P., Nicosia, A., Cortese, R., and Suck, D. (1993) *Embo J.* 12, 1805-10.
40. Leiting, B., De Francesco, R., Tomei, L., Cortese, R., Otting, G., and Wuthrich, K. (1993) *Embo J.* 12, 1797-803.
41. Viglino, P., Fogolari, F., Formisano, S., Bortolotti, N., Damante, G., Di Lauro, R., and Esposito, G. (1993) *FEBS Lett.* 336, 397-402.
42. Remenyi, A., Tomilin, A., Pohl, E., Lins, K., Philippsen, A., Reinbold, R., Scholer, H. R., and Wilmanns, M. (2001) *Mol. Cell* 8, 569-80.
43. Jacobson, E. M., Li, P., Leon-del-Rio, A., Rosenfeld, M. G., and Aggarwal, A. K. (1997) *Genes Dev.* 11, 198-212.
44. Scully, K. M., Jacobson, E. M., Jepsen, K., Lunyak, V., Viadiu, H., Carriere, C., Rose, D. W., Hooshmand, F., Aggarwal, A. K., and Rosenfeld, M. G. (2000) *Science* 290, 1127-31.
45. Sprules, T., Green, N., Featherstone, M., and Gehring, K. (2000) *Biochemistry* 39, 9943-50.
46. Anderson, J. S., Forman, M. D., Modleski, S., Dahlquist, F. W., and Baxter, S. M. (2000) *Biochemistry* 39, 10045-54.

47. Esposito, G., Fogolari, F., Damante, G., Formisano, S., Tell, G., Leonardi, A., Di Lauro, R., and Viglino, P. (1996) *Eur. J. Biochem.* **241**, 101-13.
48. Morita, E. H., Shirakawa, M., Hayashi, F., Imagawa, M., and Kyogoku, Y. (1995) *Protein Sci.* **4**, 729-39.
49. Gruschus, J. M., Tsao, D. H., Wang, L. H., Nirenberg, M., and Ferretti, J. A. (1999) *J. Mol. Biol.* **289**, 529-45.
50. Billeter, M., Qian, Y. Q., Otting, G., Muller, M., Gehring, W., and Wuthrich, K. (1993) *J. Mol. Biol.* **234**, 1084-93.
51. Grant, R. A., Rould, M. A., Klemm, J. D., and Pabo, C. O. (2000) *Biochemistry* **39**, 8187-92.
52. Hovde, S., Abate-Shen, C., and Geiger, J. H. (2001) *Biochemistry* **40**, 12013-21.
53. Gruschus, J. M., Tsao, D. H., Wang, L. H., Nirenberg, M., and Ferretti, J. A. (1997) *Biochemistry* **36**, 5372-80.
54. Catron, K. M., Iler, N., and Abate, C. (1993) *Mol. Cell. Biol.* **13**, 2354-65.
55. Catron, K. M., Wang, H., Hu, G., Shen, M. M., and Abate Shen, C. (1996) *Mech. Dev.* **55**, 185-99.
56. Shimeld, S. M., McKay, I. J., and Sharpe, P. T. (1996) *Mech. Dev.* **55**, 201-10.
57. Davidson, D. (1995) *Trends in Genetics* **11**.
58. Catron, K. M., Zhang, H., Marshall, S. C., Inostroza, J. A., Wilson, J. M., and Abate, C. (1995) *Mol. Cell. Biol.* **15**, 861-71.
59. Semenza, G. L., Wang, G. L., and Kundu, R. (1995) *Biochem. Biophys. Res. Commun.* **209**, 257-62.
60. Zhang, H., Catron, K. M., and Abate Shen, C. (1996) *Proc. Natl. Acad. Sci. U S A.* **93**, 1764-9.
61. Zhang, H., Hu, G., Wang, H., Sciavolino, P., Iler, N., Shen, M. M., and Abate-Shen, C. (1997) *Mol. Cell. Biol.* **17**, 2920-2932.
62. Newberry, E. P., Latifi, T., Battaile, J. T., and Towler, D. A. (1997) *Biochemistry* **36**, 10451-62.
63. Bendall, A. J., Rincon Limas, D. E., Botas, J., and Abate Shen, C. (1998) *Differentiation* **63**, 151-7.

64. Bendall, A. J., Ding, J., Hu, G., Shen, M. M., and Abate-Shen, C. (1999) *Development* 126, 4965-76.
65. Herr, W., and Cleary, M. A. (1995) *Genes Dev.* 9, 1679-1693.
66. Scully, K. M., Jacobson, E. M., Jepsen, K., Lunyak, V., Viadiu, H., Carriere, C., Rose, D. W., Hooshmand, F., Aggarwal, A. K., and Rosenfeld, M. G. (2000) *Science* 290, 1127-1131.
67. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) *Cell* 77, 21-32.
68. Tomilin, A., Remenyi, A., Lins, K., Bak, H., Leidel, S., Vriend, G., Wilmanns, M., and Scholer, H. R. (2000) *Cell* 103, 853-64.
69. Chasman, D., Cepek, K., Sharp, P. A., and Pabo, C. O. (1999) *Genes Dev.* 13, 2650-7.
70. Stepchenko, A. G., and Polyanovskii, O. L. (1996) *Molecular Biology* 30, 296-302.
71. Cleary, M. A., and Herr, W. (1995) *Mol. Cell. Biol.* 15, 2090-100.
72. Bendall, A. J., Sturm, R. A., Danoy, P. A., and Molloy, P. L. (1993) *Eur. J. Biochem.* 217, 799-811.
73. Verrijzer, C. P., Alkema, M. J., van Weperen, W. W., Van Leeuwen, H. C., Strating, M. J., and van der Vliet, P. C. (1992) *Embo J.* 11, 4993-5003.
74. van Leeuwen, H. C., Strating, M. J., Rensen, M., de Laat, W., and van der Vliet, P. C. (1997) *Embo J.* 16, 2043-53.
75. Klemm, J. D., and Pabo, C. O. (1996) *Genes Dev.* 10, 27-36.
76. Henry, R. W., Sadowski, C. L., Kobayashi, R., and Hernandez, N. (1995) *Nature* 374, 653-6.
77. Henry, R. W., Ma, B., Sadowski, C. L., Kobayashi, R., and Hernandez, N. (1996) *Embo J.* 15, 7129-36.
78. Henry, R. W., Mittal, V., Ma, B., Kobayashi, R., and Hernandez, N. (1998) *Genes Dev.* 12, 2664-72.
79. Sadowski, C. L., Henry, R. W., Kobayashi, R., and Hernandez, N. (1996) *Proc. Natl. Acad. Sci. U S A* 93, 4289-93.

80. Wong, M. W., Henry, R. W., Ma, B., Kobayashi, R., Klages, N., Matthias, P., Strubin, M., and Hernandez, N. (1998) *Mol. Cell. Biol.* **18**, 368-77.
81. Mittal, V., and Hernandez, N. (1997) *Science* **275**, 1136-40.
82. Mittal, V., Ma, B., and Hernandez, N. (1999) *Genes Dev.* **13**, 1807-21.
83. Ford, E., Strubin, M., and Hernandez, N. (1998) *Genes Dev.* **12**, 3528-40.
84. Cepek, K. L., Chasman, D. I., and Sharp, P. A. (1996) *Genes Dev.* **10**, 2079-88.
85. Gstaiger, M., Georgiev, O., van Leeuwen, H., van der Vliet, P., and Schaffner, W. (1996) *Embo J.* **15**, 2781-90.
86. Herr, W. (1992) in *Transcriptional Regulation* (Yamamoto, S. M. a. K., Ed.) pp 1103-1135, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
87. Mittal, V., Cleary, M. A., Herr, W., and Hernandez, N. (1996) *Mol. Cell. Biol.* **16**, 1955-65.
88. Zhao, X., Pendergrast, P. S., and Hernandez, N. (2001) *Mol. Cell* **7**, 539-49.

CHAPTER II : X-RAY CRYSTAL STRUCTURE DETERMINATION

2.1 Msx-1/DNA Complex

The initial step in structure determination is the production of single and well diffracting crystals. It is essential to have large quantities of purified protein and DNA. The complex is set up in a specific ratio for crystallization using sparse matrices of precipitating solutions that have proven success in crystallizing various proteins. One of the most common methods for crystallizing macromolecules is the hanging drop vapor diffusion method (Figure 2.1). In this method, a drop containing a mixture of the complex and precipitating solution is equilibrated against a reservoir containing the precipitant. The complex slowly precipitates and the molecules adopt identical orientations that form an orderly three-dimensional array of molecules held together by non-covalent interactions. The crystallization process involves screening of DNA to obtain the best complex and setting up thousands of drops. These drops have to be monitored for precipitation and solubility behavior. Based on the results from the initial sparse screens, new screens can be developed to optimize initial crystals. Eventually the iterative process leads to crystals suitable for X-ray diffraction. This strategy was followed in the crystallization of both the Msx-1/DNA complex and the Oct-1/SNAP 190/DNA complexes.

2.1.1 Crystallization

The murine Msx-1 HD protein (166 - 225) was over - expressed in E-coli and purified by Nickel affinity chromatography. The native protein was screened for crystallization using various DNA sequences. The Msx-1 HD/DNA complex was crystallized (Figure 2.2) and data was collected at cryogenic temperatures (123K) to prevent crystal decay.

The Msx-1 HD/DNA crystals belong to the $P2_12_12_1$ space group with unit cell dimensions $a=33.66 \text{ \AA}$, $b=60.96 \text{ \AA}$, $c=83.37 \text{ \AA}$. There is one molecule per asymmetric that includes 55% solvent. This value is within the observed range for protein crystals (1). The crystal parameters for the Msx-1 HD/DNA complex are listed in Table 2.1. Data was 99.8% complete for 3,778 unique reflections derived from a total of 11,914 reflections. Detailed data collection statistics are found in Table 2.2 for the Msx-1 HD/DNA crystals.

Table 2.1 Crystal parameters for the Msx-1 HD/DNA Complex

Crystal Form	Orthorhombic
Space Group	$P2_12_12_1$
Unit Cell	$a=33.66 \text{ \AA}$ $b=60.96 \text{ \AA}$ $c=83.37 \text{ \AA}$
Solvent Content	55%
Molecules per asymmetric unit	1

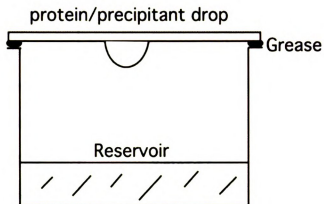


Figure 2.1 Hanging drop vapor diffusion method. The reservoir contains precipitating agents that cause crystals to form.

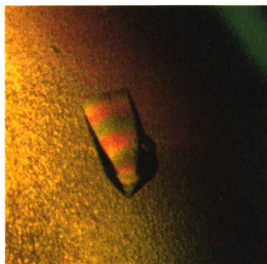


Figure 2.2 Orthorhombic Crystal of Msx-1 HD/DNA4. The crystal has dimensions of $0.4 \times 0.2 \times 0.2 \text{ mm}^3$.

Table 2.2 Statistics for the Msx-1 HD/DNA data sets

	Native	Iodo – DNA Complex
Resolution Range (Å)	40.0 – 2.15	40.0 – 2.50
(last resolution shell)	2.25 – 2.15	2.82 – 2.69
Cell parameters (Å)	a=33.66 b=60.96 c=83.37 Å	a=33.6 b=59.89 c=83.64 Å
Completeness (%)	98.9(99.8)	99.2(100.0)
R_{merge} (I) [*](%)	6.2(29.8)	7.7(24.6)
<I>/<σ_I>	17.2(2.9)	17.6(7.1)

1. Both data sets were collected at the Michigan State University Macromolecular X-ray Facility home source.

[†]Values in parenthesis refer to the last resolution shell.

^{*} $R_{\text{merge}} = \frac{\sum \sum |I_i - \langle I \rangle|}{\sum \sum |I_i|}$, where I_i is an individual intensity measurement and $\langle I \rangle$ is the average intensity for this reflection, with summation over all data.

2.1.2 Structure Determination

After high resolution data have been obtained, an electron density map is calculated from which the structure of the protein is determined. This map includes density for all of the molecules that form the lattice. This includes protein main and side chain atoms as well as DNA and solvent molecules. This map may include any ions or substrates that also make up the crystalline lattice. Electron density is represented by the following equation,

$$\rho(x,y,z)=\frac{1}{V}\sum_h\sum_k\sum_l F_{hkl}e^{-2\pi(hx+ky+lz)}$$

The structure factor (F_{hkl}) for each reflection labeled hkl , is a complete description of all of the atoms that contribute to that reflection. F_{hkl} is a wave function with frequency, amplitude and phase. The frequency is that of the X-ray source and the amplitude is proportional to the square root of the measured intensity of the reflection. The one thing left that we need in order to calculate our map is the phase. This is known in crystallography as the phase problem.

This phase problem can be solved using several methods, molecular replacement, isomorphous replacement, or multiple wavelength anomalous dispersion (MAD). The molecular replacement method is most commonly used when the protein is homologous to another protein for which the structure has been solved. This is the method that was used to solve the Msx-1 HD/DNA structure. Several structures of homeodomains had been previously solved so there were many models from which to start.

Molecular replacement uses phases from a known protein structure as an initial estimate of the phases for your protein and works backwards to calculate structure factors (F_{calc}). The ability to work backwards relies on the Patterson function. It is the Fourier transform using coefficients $F(hkl)^2$. This function will contain sets of peaks representing intra - atomic vectors (2). The rotation function is calculated by transforming the crystallographic system into a Cartesian coordinate system and then converting this into a system of Eulerian angles (α, β, γ) which defines any orientation of the molecule. The density of crystal 2 is converted into that of crystal 1 by rotating crystal 2 through γ about Z, then through β about Y, and then through α about Z relative to the fixed axes. The rotation function uses Patterson maps to determine the correct orientation of the model. This is performed by the automated Patterson search routine of the program AmoRe (3).

Once the correct orientation of the model has been found the correct position can be determined through the translation function. The intermolecular vectors (cross vectors) are used to find the translation between molecules after their orientations have been found. The translation vector for each atom is split into two vectors. One (t) in the direction of the rotation axis and the other (s) perpendicular to the axis direction. The vector t is the same for all atoms while vector s depends on the distance of each atom from the axis of rotation. The superimposition of the self vector set on the cross vectors would give a number of positions where some agreement between the vector peaks is obtained (4).

The results are evaluated based on the agreement of the calculated structure factors with the observed ones from the data. The structure factors of a properly positioned model are then calculated and a correlation coefficient (CC) and an R-factor

value will determine if a correct solution has been obtained. A good solution will yield a high CC and a low R-factor. These two values are defined below, where F_{obs} and F_{calc} are the observed and calculated structure factors.

$$R = \frac{\sum_{hkl} \|F_{obs} - F_{calc}\|}{\sum_{hkl} |F_{obs}|} \times 100$$

$$C.C. = \frac{\sum_{hkl} \left(|F_{obs}|^2 - \overline{|F_{obs}|^2} \right) \left(|F_{calc}|^2 - \overline{|F_{calc}|^2} \right)}{\left[\sum_{hkl} \left(|F_{obs}|^2 - \overline{|F_{obs}|^2} \right)^2 \sum_{hkl} \left(|F_{calc}|^2 - \overline{|F_{calc}|^2} \right)^2 \right]^{1/2}}$$

In the case of the Msx-1HD/DNA complex there were problems fitting the DNA so a complex was crystallized that had iodo-uracil substituted for thymine. This incorporates isomorphous replacement into the equation. In isomorphous replacement, differences in the diffraction pattern are measured after the introduction of a heavy atom, in this case iodine. For isomorphous replacement to work you need an atom or group of atoms with many electrons, you also need the protein molecules within the crystal to be identically bound by the heavy atom with no change in the native crystal lattice. For

DNA/protein complexes it is easy to chemically modify the DNA and insure that the second criterion above is met.

The coordinates and the diffraction pattern of the heavy atom alone can be determined by calculating the difference between the diffraction patterns of the native crystal and the heavy atom crystal. With the diffraction pattern of the heavy atom, there are a small number of atoms and an easier structure to be determined. In the case of the Msx-1 HD/DNA structure, the complex is rather small in size so to provide additional phase information only one heavy atom substitution was required. The iodo-uracil crystals were grown in a similar condition to that of the native structure. The heavy atom search routine was utilized from mlphare (5).

In the case of the Iodine DNA derivative there is an anomalous scattering component present in the data and these differences can sometimes be large enough for phasing. Anomalous scattering comes from the heavy atom having an absorption edge near the wavelength used to collect data. Iodine can have an anomalous difference for CuK α radiation (1.54 Å) of about 5% (for small proteins) (6). An anomalous difference Patterson is then calculated and then compared to the isomorphous difference Patterson. A peak that is in both is most likely to be the heavy atom.

The Patterson (P_{uvw}) search is performed by the calculation of a vector map known as the Patterson function. This Patterson map is calculated based on the product of the electron density at two points separated by a distance u ,

$$P_{uvw} = \int_V \rho(x,y,z)\rho(x+u,y+v,z+w)dV$$

$$P_{uvw} = \frac{1}{V} \sum_h \sum_k \sum_l |F^2_{hkl}| e^{-2\pi(hu + kv + lw)}$$

The Patterson will have peaks that have varying heights. The peak height relationship depends somewhat on overlapping peaks, which can be compensated for by peak sharpening and removal of the origin peak. The peak height also depends on the number of electrons in the atoms, a vector between two Hg atoms will have a much higher peak in the Patterson than a C – Hg vector. The result obtained from the Patterson function is peaks located at positions defined by the inter-atomic vectors between heavy atoms. Symmetry within the unit cell is used to translate these inter-atomic vectors into crystallographic coordinates. Unfortunately our choice of thymine for derivitization was disordered in the structure thereby rendering our heavy atom data useless.

2.1.3 Molecular Replacement and Structure Refinement

There were many available options for the model in solving our structure. We used the even-skipped HD/DNA structure (PDB ID# 1JGG)(7). The surface side chains that were not identical to Msx-1 were changed to Ala. The three central base pairs (AAT) of the DNA in the binding site were used in conjunction with the pared down protein as the search model. Though the model contained only one third of the total asymmetric unit, a weak but correct molecular replacement solution was obtained with a C.C. of 19.3% and an R-factor of 46.5%.

The next step in solving the structure involves calculating phases from the model and using these in conjunction with the observed experimental amplitudes to compute an electron density map. This should resemble the real molecule more closely than the original model. Refinement involves slightly adjusting the position of the atoms in the molecule to satisfy a number of requirements of the program, including realistic bond lengths and angles, realistic energies, all the while comparing what you have built to the data. The overall goal is to bring F_{calc} and F_{obs} into agreement. This agreement is represented by a numerical value, the R-factor which is represented by the following equation:

$$R_{factor} = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|} \times 100$$

The correlation coefficient is another indication of this. It is almost independent of scaling between F_{calc} and F_{obs} and is a better indication of progress when the R-factor is high.

The Msx-1/HD structure was refined using the simulated annealing method (8, 9). Molecular dynamics is used to simulate the various parameters in the conformational space in the molecule. The annealing process involves heating of the particles until they are in the liquid phase. Then the molecule is slowly cooled so that these particles will go into the lowest energy state. The target function consists of an empirical potential energy which is described by the stereochemistry and non-bonding interactions in the macromolecule.

During refinement resolution is extended to the high resolution data cutoff and solvent molecules are added to the structure. Waters are only seen in cases where the data is better than 3.0Å. In the case of the Msx-1/HD complex the final resolution was 2.2Å and there were a total of 153 water molecules. The final refinement parameters are listed in Table 2.3. The final model consists of 1,219 non-hydrogen atoms with 2 amino acids and 3 DNA bases disordered. Additionally the side chains of residues R21, Q22, and R57 were disordered.

Figure 2.3 represents the Ramachandran plot of the Msx-1 HD/DNA structure. A Ramachandran diagram is a plot of ϕ (angle between N and $C\alpha$) versus ψ (angle between C and $C\alpha$). Based on geometry and considering steric restrictions the ϕ and ψ angles must be within certain values which are marked by the different shaded areas in Figure 2.3. All residues that have a side chain must lie within these allowed regions. In the Msx-1 HD/DNA complex there are no residues that lie outside the allowed regions.

Table 2.3 Refinement statistics for Msx-1 HD/DNA complex

<i>R</i> -factor	19.8 %
<i>R</i> _{free}	26.8%
Resolution	8.0 – 2.2Å
rmsd Bonds	0.0192 Å
rmsd Angles	2.174°

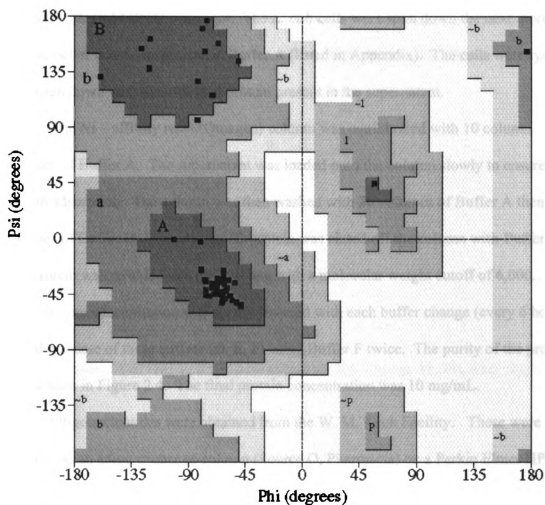


Figure 2.3. Ramachandran Plot of the Msx – 1 Homeodomain Residues.

2.1.4. Materials and Methods

2.1.4.1. Over-expression and Purification

The Msx-1 HD (166 - 225) was overexpressed in *E. coli* and purified to homogeneity by nickel-affinity chromatography. Our collaborator Dr. Cory Abate – Shen sent us some glycerol stocks of the Msx-1 HD. These were plated on ampicillin (0.1 mg/mL) and kanomycin (0.025 mg/mL) resistant agar plates. They were then grown overnight in 50mL flasks of LB containing ampicillin and kanomycin. They were transferred into 1L flasks containing the antibiotics and grown to an O.D. of 0.5, then induced with 1mM IPTG overnight. The *E. coli* cells were spun down the next morning and the pellet was re-suspended in Buffer A (listed in Appendix). The cells were lysed, then spun down with the unfolded protein present in the supernatant.

The Ni – affinity resin (Quiagen) column was equilibrated with 10 column volumes of Buffer A. The supernatant was loaded onto the column slowly to ensure protein adsorption. The column was then washed with 20 volumes of Buffer A then 10 volumes of Buffer B. The Msx-1 HD protein was eluted off the column with Buffer C. The protein was refolded in a dialysis bag with a molecular weight cutoff of 6,000. The Guanidinium concentration was slowly lowered with each buffer change (every 6 hours) over the course of three buffers (D, E, F) using Buffer F twice. The purity of the protein can be seen in Figure 2.4. The final protein concentration was 10 mg/mL.

Oligonucleotides were obtained from the W. M. Keck Facility. These were then purified on an anion exchange column (Source Q, Pharmacia) on a Perkin Elmer HPLC with UV detection at 260 nm. One μ mole of DNA was loaded onto the column with buffer A (10 mM NaOH and 0.2 M NaCl) and eluted with a shallow gradient of buffer B

(10mM NaOH and 1.0 M NaCl). Collected fractions were neutralized with 1 M Tris at pH 7.5. For concentration, the sample was diluted with four volumes of 10 mM Tris and loaded on a 1 ml DEAE cellulose column and the DNA was then eluted with a 1M NaCl elution buffer. The DNA was further concentrated in a centricon-3 (Amicon) concentrator. DNA strands were annealed in equimolar amounts. The final DNA concentrations were approximately 2mM. The derivitized DNA was purified in the same manner but was shielded from light to slow degradation of the iodine. Table 2.4 lists all of the oligonucleotides utilized in crystallization trials as well as their results. Table 2.5 lists all of the heavy atom DNA strands utilized and their results. The Msx-1 HD was complexed to all of the DNAs in a 1:1.2 molar ratio and buffer exchanged into Buffer G.



M.W. Msx-1

Figure 2.4 Msx-1 Gel. M.W. Standards: Purple 42,000, Orange 32,000, Red 17,900, and Blue 7,200.

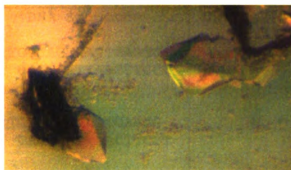
2.1.4.2 Crystallization

All complexes were screened for crystallization using the hanging drop vapor diffusion method (Figure 2.1). The reservoir contained 300 μ L of the precipitating solution and a 3 μ L hanging drop consisting of a 1:1 protein/DNA complex to precipitating solution ratio. The search for initial crystallization conditions was performed through sparse matrix sampling by using different crystallization screens at room temperature (10, 11). Crystals formed at 298K from a solution containing 12% PEG 4000 and 0.1 M sodium acetate, pH 4.6. The best crystals grew from DNA 4 (Tables 2.4 & 2.5). The crystals grew in 4 weeks to maximum dimensions of 0.4 \times 0.2 \times 0.2 mm³ (Figure 2.2). Other crystals of the Msx-1 HD complexed to different DNA are in Figure 2.5.

Table 2.4. DNA sequences used in Msx-1 Crystallization Trials.

Name	DNA Sequence	Crystallized	Diffacted to
DNA1	TTCACTAATTGA AGTGATTAACTA	Yes	-
DNA2	TGTCACTAATTGAA CAGTGATTAACTTA	Yes	8.0 Å
DNA3	TGTCACTAATTGAAGG CAGTGATTAACCTCCA	Yes	3.2 Å
DNA4	TGTCACTAATTGAAGG CAGTGATTAACCTCCT	Yes	2.1 Å
DNA5	TTCACTAATTGA AGTGATTAACCT	No	-
DNA6	TTCACTAATTGAA AGTGATTAACTTA	Yes	-
DNA7	TAACCGATATGTGG TTGGCTATACACCA	No	-
DNA8	TGCATAATCACCCGGG CGTATTAGTGGGCCCT	Yes	8.0 Å
DNA9	TAGTGATTTCCGCC TCACTAAAGGCGGA	Yes	-
DNA10	TGTCACTGATTGAAGG CAGTGACTAACTTCCT	No	-
DNA11	TGTCACTAATTAAAGG CAGTGATTAATTCCT	No	-
DNA12	TGTCACTAATTTAAGG CAGTGATTAATTCCT	No	-

A.



B.

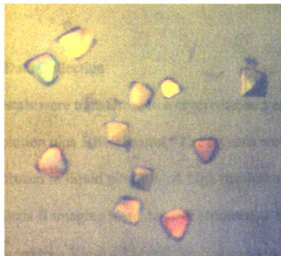


Figure 2.5. Crystals of the Msx-1 HD/DNA Complexes. A. DNA3 complex crystals $0.2 \times 0.2 \times 0.1 \text{ mm}^3$. B. DNA8 complex crystals $0.1 \times 0.1 \times 0.05 \text{ mm}^3$.

Table 2.5. Iodinated DNAs used in crystallization trials.

Name	DNA Sequence	Crystallized	Diffacted to
DNA4A1	TGUCACTAATTGAAGG CAGTGATTAACCTTCCT	Yes	2.5Å
DNA4A2	TGTCACUAATTGAAGG CAGTGATTAACCTTCCT	No	-
DNA4A3	TGTCACTAATUGAAGG CAGTGATTAACCTTCCT	No	-

2.1.4.3 Native Data Collection

The crystals were transferred to a cryoprotectant containing the original precipitating solution plus 30% glycerol. The crystals were then mounted in nylon cryo-loop and flash frozen in liquid nitrogen. A high resolution data set was collected at home using a MSC Raxis II imaging plate detector (Molecular Structure Corp., TX). CuK α X-rays were produced by a Rigaku RU 200 rotating anode source operating at 5 kW (50 kV \times 100 mA). Data was collected to a resolution of 2.1 Å. The crystal to detector distance was 100mm, and 120° worth of data was collected with an oscillation angle of 1.0°. Diffraction data were processed with DENZO and scaled with SCALEPACK (12).

2.1.4.4 SIR data collection

An SIR experiment was also carried out, in which chemically modified DNA was crystallized with Msx-1 HD. The best crystals came from DNA 4A1 (Table 2.4). The crystals were cryoprotected and frozen as previously described. Data were collected over

226° with oscillations of 2.0°. A total of 94,663 reflections were measured at our home source. We attempted to find the one iodine present in the DNA with the programs xtalview (13) and mlphare (5). But due to the unfortunate luck of the substituted DNA base being disordered; the molecular replacement method was the only method used to solve the structure. All model building was done using TURBO FRODO and refinements were carried out using CNS (14). A lot of the crystallographic software is part of the CCP4 suite of programs (15).

2.2 Oct-1/SNAP 190/DNA Complex

2.2.1 Crystallization and data collection

The Oct-1/U1 octamer/SNAP 190 (884-910) complex was crystallized (Figure 2.6) and a complete data set to a resolution of 2.3Å was collected. The crystals are triclinic (P1) with unit cell dimensions of $a=36.43$, $b=54.97$, $c=77.61$ Å, $\alpha=94.93$, $\beta=99.59$, $\gamma=109.25^\circ$. We have two molecules of the complex in the asymmetric unit which corresponds to 53% solvent in the crystal. This value is in the range for protein crystals (1). The crystal parameters for the complex crystal are listed in Table 2.6. A synchrotron X-ray diffraction data set was collected on a single crystal to a resolution of 2.3Å. Detailed data statistics are listed in Table 2.7.

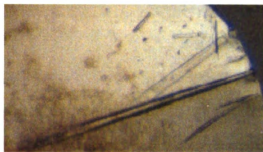


Figure 2.6. Oct-1/U1/SNAP 190 (884 – 910) with dimensions of $0.7 \times 0.05 \times 0.025$ mm³.

Table 2.6 Crystal parameters for the Oct-1/U1 octamer/SNAP190 peptide crystal

Crystal Form	Triclinic
Space Group	P1
Unit Cell	a=36.43, b=54.97, c=77.61Å $\alpha=94.93, \beta=99.59, \gamma=109.25^\circ$
Solvent Content	53%
Molecules per asymmetric unit	2

Table 2.7 Statistics for the ternary complex data collection

Wavelength (Å)	0.97942
Resolution Range (Å)	50.0 – 2.4 (2.48 – 2.38)
Completeness, %	94.1(92.5)
I/σ	12.7(3.7)
R_{merge}, %	9.6(30.9)
Unique Reflections	23,432
Measured Reflections	110,313

2.2.2 Molecular Replacement and Structure Refinement

The structure was solved by molecular replacement using the structure of the Oct-1/DNA structure (PDB ID 1OCT) (16). The first molecule was found using the Oct-1 POU specific and HD domains. Its correlation coefficient was 22.1% while the R-factor was 53.5. Attempts to find the second molecule using this model were unsuccessful. The model was cut into its two separate domains and the second molecule was searched for

with each one individually while fixing the position of molecule one. The POU specific domain was successful in finding itself in the second molecule of the asymmetric unit. The first POU domain was overlaid onto the second one giving us two mostly complete molecules to finish building and refinement. The correlation coefficient for molecule one with 10bp of DNA and the second POU specific domain was 50.7 with an R-factor of 46.1.

Multiple rounds of structure refinement using the simulated annealing method followed by the addition of waters and extension of the resolution yielded the final refinement parameters listed in Table 2.8. The final model includes all of the protein except for the first three residues in the POU domains. The homeodomain is missing the last three residues in one molecule and the first four in the other. SNAP190 is missing the first three residues (884-886) in both molecules. The linker was more ordered than seen in previous structures. Additionally, we see seven residues that form one additional turn of helix at the end of the POU domain and then a loopy region. The two molecules are a little different in terms of side chain motion (maybe 5% difference overall). Overlaying the two molecules results in a 0.0 rmsd for the pair. The Ramachandran plot is shown in Figure 2.7 and it shows that there are no residues in the disallowed regions.

Table 2.8 Refinement statistics for Oct-1/U1 octamer/SNAP 190 (884-910)

<i>R-factor</i>	22.8
<i>R_{free}</i>	29.4
Resolution	2.4Å
rmsd Bonds	0.02
rmsd Angles	2.6

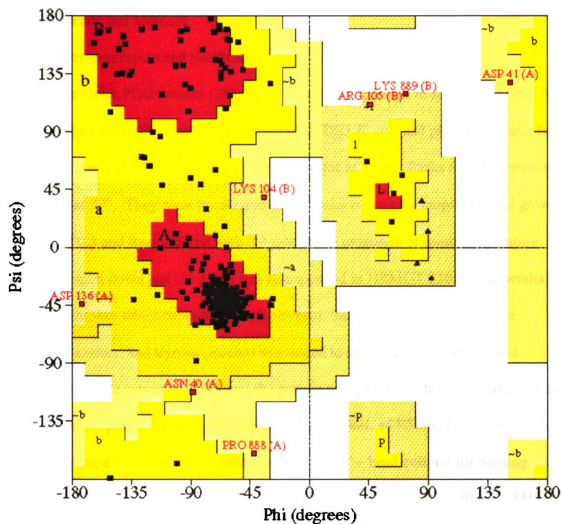


Figure 2.7 Ramachandran Plot for the Oct-1 POU and SNAP 190 residues. There are no residues in disallowed regions. The red areas indicate the most favorable regions and the yellow areas indicate additional allowed regions. The triangles represent glycines.

2.2.3 Materials and Methods

2.2.3.1 Over-expression and purification

The Oct-1 POU protein (284 - 439) was over-expressed in *E. coli* as a GST fusion protein. The plasmid was transformed into BL21DE3 *E. coli* and plated on ampicillin resistant plates. The colonies were grown overnight in 50 mL flasks of LB containing ampicillin at 37°. They were transferred to 1L flasks containing ampicillin and grown to an O.D. of 0.5 and then induced with 0.4 mM IPTG overnight at room temperature. The cells were spun down and the pellet was resuspended in HEMGT 250 (see appendix). A protease inhibitor tablet containing chymotrypsin, thermolysin, papain, pronase, pancreatic extract, and trypsin (Roche) was added to the mixture and then lysed.

The lysate was spun down and the protein was present in the supernatant. The GST – bead (Sigma) column was equilibrated with 50mL of HEMGT 250 in the cold room. The supernatant was then loaded onto the GST – bead column for binding overnight. The beads were then washed with 50 mL of HEMGT 250, 30 mL of HEMGT 100, 20mL 1x TDB, and finally 10 mL of 1x TDB – DTT. The beads were checked for protein binding, and then 20 units of thrombin (Sigma) were added to the beads with overnight shaking. The thrombin was inhibited with PMSF (final concentration of 0.5 mM) and the protein was eluted with TDB – DTT. The SDS PAGE gel is shown in Figure 2.8. The final protein concentration was ~5mg/mL.

The DNA was purified as described for the Msx-1 DNA strands. All of the DNA sequences used in attempted crystallizations are listed in Table 2.9 along with the results. We also tried derivatized DNAs and those are listed in Table 2.10 along with the results. A few different crystal forms are shown in Figure 2.9.

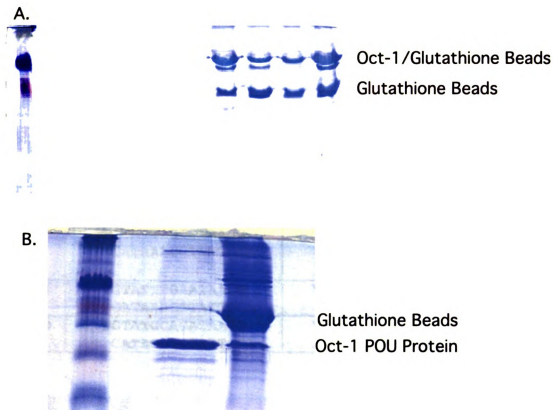


Figure 2.8. SDS – PAGE Gel of Oct-1 POU bound to Glutathione Beads. M.W. of purple is 42,000.

Table 2.9. DNA Sequences used in ternary complex crystallization.

	DNA Sequences (5'-3')	Crystals (27 mer)	Crystals (52 mer)	Diffraction Limit (Å)
H2B	TGTATGCAAATAAGG CATACGTTTATTCCA	Yes	Yes	3.2/not tested
TLF1	ATGTATGCAAATAAGG CATACGTTTATTCCAT	No	-	-
TLF2	TGTATGCAAATAAGG CATACGTTTATTTCCT	No	-	-
TLF3	GTGTATGCAAATAAGG CATACGTTTATTCCCA	Yes	-	None at synchrotron
U1	TGTATGTAGATAAGG CATACATCTATTCCA	Yes	Yes	2.3 at synchrotron/ 2.3 at synchrotron
U6	TGTATTTGCATAAGG CATAAACGTATTCCA	Yes	Yes	5.0 at home and synchrotron/ 2.6 at synchrotron
HSUR1	TGTATTTGAATAAGG CATAAACTTATTCCA	Yes	-	9.0 at synchrotron
U152B	TGTATGCAGATAAGG CATACGTCTATTCCA	Yes	-	None at home

Table 2.10 Iodinated DNA Sequences used in ternary complex crystallization.

Name	DNA Sequence	Crystals	Diffraction to
U1AI	TGUATGTAGATAAGG CATACATCTATTCCA	Yes	-
U1AI2	TGTATGTAGAUAAAGG CATACATCTATTCCA	Yes	-
U1BI	TGTATGTAGATAAGG CATACATCTAUTCCA	Yes	-
U1BI2	TGTATGTAGATAAGG CATACATCUATTCCA	Yes	-

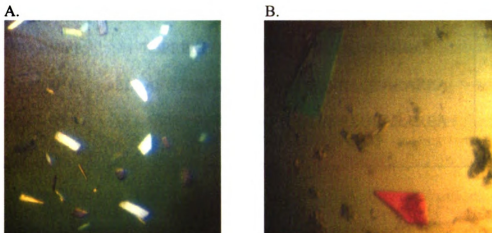


Figure 2.9. A. Oct-1/U1/SNAP 190 (52mer) grown in 20% PEG 6000 and 0.1 M Sodium Acetate pH 5.5 ($0.05 \times 0.1 \times 0.02 \text{ mm}^3$). B. Oct-1/U6/SNAP 190 (52 mer) grown in 7% PEG 6000 and 0.14 M Sodium Acetate pH 5.5 ($0.3 \times 0.2 \times 0.02 \text{ mm}^3$).

SNAP 190 peptides were ordered from the Keck Facility. The four peptides we tried are listed in Table 2.11. They were all purified by reverse phase chromatography on a C18 column (Vydax) using an acetonitrile gradient. Peptides were lyophilized and reweighed and set up in complexes with DNA and Oct-1. The three components were set up in a ratio of 1:1.2:3 with Oct-1:DNA:SNAP 190 peptide. The complexes that have been screened so far are listed in Table 2.12. The complexes were buffer exchanged into 0.1 M Hepes pH 7.9, 10 mM DTT. Crystals did not form in the absence of DTT.

Table 2.11 Peptide Sequences used in Crystallization attempts.

Name	Peptide Sequence	Crystals
18mer (888 – 903)	PKPKTVSELLQEKRLQ	No
27mer (884 – 910)	TGPRPKPKTVSELLQEKRLQEARAREA	Yes
32mer (879 – 910)	SLLASTGPRPKPKTVSELLQEKRLQEARAREA	Yes
52mer (879 – 930)	SLLASTGPRPKPKTVSELLQEKRLQEARAREA TRGPVVLPSQLLVSSSVILQ	Yes

Table 2.12 Ternary complexes that have been set up.

	18mer	27mer	32mer	52mer
H2B	X	X		X
TLF1	X	X		
TLF2	X	X		
TLF3	X	X		
U1	X	X	X	X
U6	X	X	X	X
HSUR1	X	X		
U152B	X	X		

2.2.3.2 Crystallization and data collection

All complexes were extensively screened for crystallization using the hanging drop vapor diffusion method (Figure 2.1). All of the complexes were screened using several sparse crystallization screens at 298 K. Previous Oct-1 complexes produced crystals at room temperature so we focused our efforts there. A 2 μ L hanging drop in a 1:1 complex to precipitant solution ratio was equilibrated against 300 μ L of precipitant solution. The best crystals came with the Oct-1/U1/SNAP 190 (884-910) complex in 20% isopropanol, 20% PEG 4000, and 0.1 M Sodium Citrate pH 5.6. The crystals appeared in about 2 to 3 weeks and were approximately 0.7 \times 0.05 \times 0.025 mm³ (Figure 2.6).

The crystals were transferred to a cryoprotectant solution containing 30% MPD in addition to the precipitant condition and flash frozen in liquid nitrogen. X-Ray diffraction data to 2.3 Å was collected at the SBC beamline at APS (Argonne, IL). Data was collected using a custom built 3 \times 3 array (3072 \times 3072 pixels) CCD area detector. The crystal to detector distance was 200mm and two data sets were collected on this crystal. The two were scaled together and a total of 430° of data were collected. Diffraction data were processed using HKL2000 (17). This structure was solved by molecular replacement using the program AmoRe (3). All model building was done with TURBO FRODO and refinements were done with CNS (14).

2.3 References

1. Matthews, B. W. (1968) *J. Mol. Biol.* 33, 491-7.
2. Stout, G. H., and Jensen, L. H. (1989) *X-ray structure determination : a practical guide*, 2nd ed., Wiley, New York.
3. Navaza, J. (1994) *Acta. Cryst. A50*, 157-163.
4. Blundell, T. L., and Johnson, L. N. (1976) *Protein crystallography*, Academic Press, New York.
5. Otwinowski, Z. (1991) (Wolf, W., Evans, P. R., & Leslie, A. G. W. (eds), Ed.), CCP4 Daresbury Study Weekend, nos. DL/SCI/R32 Warrington WA4 4AD, UK: Daresbury Laboratory, for Daresbury Laboratory.
6. McRee, D. E., and David, P. R. (1999) *Practical protein crystallography*, 2nd ed., Academic Press, San Diego, Calif.
7. Hirsch, J. A., and Aggarwal, A. K. (1995) *EMBO J.* 14, 6280-6291.
8. Brunger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science* 235, 458-460.
9. Kirkpatrick, S., Gelatt, C., and Vecchi, M. (1983) *Science* 220, 671-680.
10. Cudney, R., Patel, S., Wesisgraber, K., Newhouse, Y., and McPherson, A. (1994) *Acta. Cryst. D50*, 414-423.
11. Jancarik, J., and Kim, S.-H. (1991) *J. Appl. Cryst.* 24, 409-411.
12. Otwinowski, Z. (1993) in *Data Collection and Processing* (Sawyer, L., Issacs, N., and Bailey, S., Eds.) pp 56-62, SERC Daresbury Laboratory, Daresbury, U.K.
13. McRee, D. E. (1999) *J. Struct. Biol.* 125, 156-65.
14. Brunger, A. T. (1992) *X-PLOR, version 3.1, a System for X-ray Crystallography and NMR.*, Yale University Press, New Haven, CT.
15. CCP4. *Acta. Cryst. D50*, 760-763.
16. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) *Cell* 77, 21-32.

17. Otwinowski, Z. M., W. (1997) *Methods in Enzymology* 276, 307-325.

CHAPTER III: THE THREE DIMENSIONAL STRUCTURE OF THE MSX-1 HD / DNA COMPLEX

3.1 Overall structure of the Msx-1/DNA Complex

The three dimensional structure of the Msx-1 HD/DNA complex provides some additional insight into homeodomain – DNA binding. This is one of several homeodomain – DNA structures that have been solved. We have done a thorough analysis of all of the known crystal structures of these complexes. There are many similarities but some subtle differences. These include several conserved waters and subtle differences in DNA binding. The complete structure consists of 58 of the 60 amino acids in the HD, 29 of the 32 bases of DNA, and 153 water molecules.

The overall structure of the complex is a 3 helix protein which includes a HTH domain and an extended flexible N – terminal arm. The recognition helix of the HD binds to the major groove of the DNA and the N – terminal arm tracks across the minor groove of the DNA (Figure 3.1). Hydrophobic core residues and three key salt bridges hold the protein together (Figure 3.2 & 3.3). Mutation of these residues severely compromises both the transcriptional repression activity as well as the DNA-binding affinity of the Msx-1 HD, indicating the importance of these residues for structural integrity (*1*). The HD and DNA interaction involves several base specific contacts, water mediated contacts, and non – specific interactions.

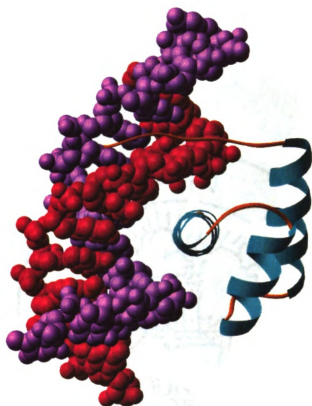


Figure 3.1 The three dimensional structure of the Msx-1 HD/DNA Complex. The view is looking down the recognition helix.

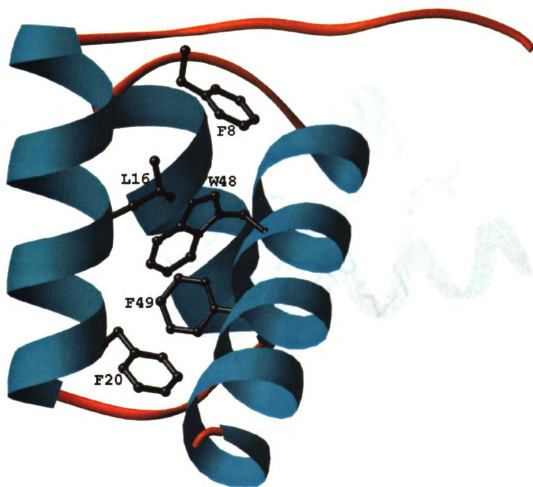


Figure 3.2 The hydrophobic core residues that are integral to protein stability.

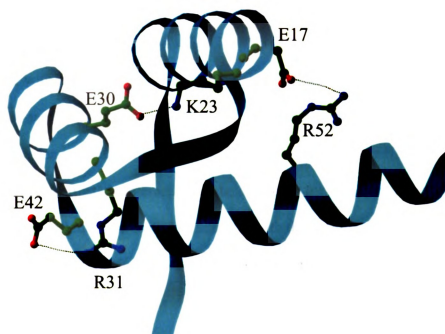


Figure 3.3 The salt bridges that connect the three helices.

The most significant structural differences between HDs are seen in the n-terminal arm conformation. In fact much of this region is disordered in many of the known HD structures. The Msx-1 HD n-terminal arm on the other hand, is unusually well ordered with only the 1st residue of the HD not accounted for in the final model. The Msx-1 HD is additionally stabilized by three salt bridges (Figure 3.3), which serve as “electrostatic crosslinks” between each of the three helices (Table 3.1). The salt bridge between E30 and K23 links helix one with helix two; the salt bridge between E42 and R31 links helix two with helix three; the salt bridge between E17 and R52 links helix three with helix one. Similar, though not identical salt bridges are observed in most of the other HD structures, indicating these electrostatic interactions to be important for domain stability. For example, in most HD’s residue 30 is basic and interacts with the highly conserved E19 residue, which still represents an electrostatic interaction between helix one and two.

Table 3.1 Salt Bridges in Msx-1HD

Anion	Cation	Distance
Glu 30 COO ⁻	Lys 23 NH ₃	2.82 Å
Glu 42 COO ⁻	Arg 31 NH ₂	4.57 Å
Glu 17 COO ⁻	Arg 52 NH ₂	3.69 Å

The Msx-1 HD is homologous to several other homeodomains in their structures and their sequences. A superposition of the Msx-1 HD and the other crystallized HDs results in root mean square deviations between 0.54 and 0.85 Å. A few of the HDs are overlaid in Figure 3.4 to illustrate this point. In Figure 3.5 a sequence alignment of the homeodomains is shown. The overall fold of the homeodomain is very similar in these structures, as is the way in which they interact with the DNA. Though most homeodomains bind DNA exclusively as monomers, some also bind as heterodimers, the MATa1/MATα2 heterodimer (2, 3) and the HoxB1/Pbx-1 heterodimer (4) being examples. Other families of homeodomain proteins contain a second DNA binding domain within their sequence. These include the *POU* and *Paired* classes of homeobox genes.

3.2 Protein/DNA recognition

Most of the protein/DNA interactions between the HD recognition helix and the DNA major groove are common among other HD/DNA complexes whose structures are known, especially those HD's that recognize the canonical TAAT sequence. The I47 in Msx-1 makes a hydrophobic contact to Ade9 and Thy10 but this contact is seen no matter what residue is there, with valine and asparagine also occurring at position 47. Figure 3.6 schematically summarizes all of the interactions seen in the Msx-1 HD/DNA structure.

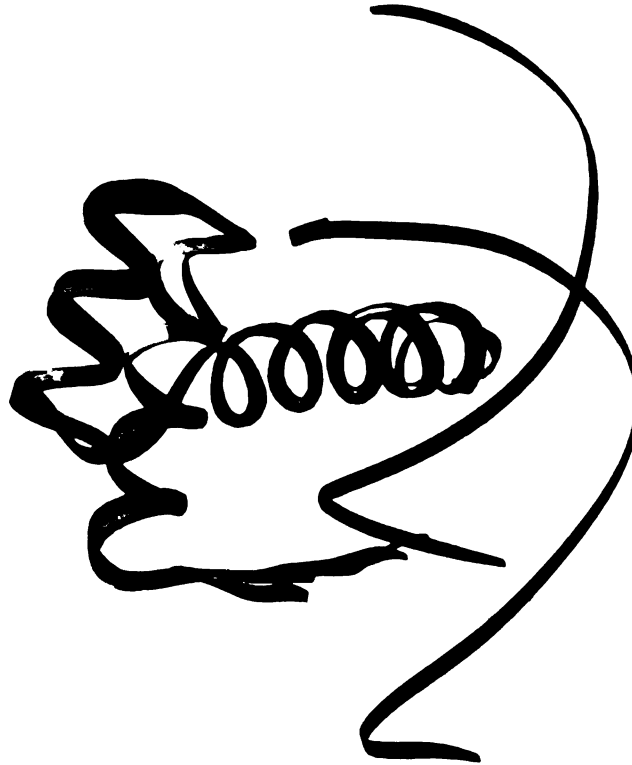


Figure 3.4 Overlay of the Msx-1 HD (cyan), Antennapedia (dk.blue), engrailed (yellow), and even-skipped (red). The DNA is from the Msx-1 structure and is present to provide a reference point.

	* * ^ ^ ^ ^ *
Msx-1	NRKPRTPTTAAQLL <u>ALER</u> KFRQKQ---YLSIAERAEF
MATa1	SPKCKSSISPQARAFLEQVF--RRKQSLNSKE-KEEV
MATalpha2	KPYRGHRFTKENVRILES W FAKNIENPYLDTKGLENL
pbx	ARR <u>KRRN</u> FNKQATEILNEYFYSHLSNPYPSEEAKEEL
hoxb1	PSGLRTNFTTRQLELEKEFHFNK---YLSRARRVEI
ubx	<u>RRRGR</u> QTYTRYQLELEKEFHNTNH---YLTRRRRIEM
exd	ARR <u>KRRNF</u> SKQASEILNEYFYSHLSNPYPSEEAKEEL
engrailed	EKR <u>PRTAF</u> SSEQLARLKREFNENR---YLTERRRQQL
antennapedia	<u>RKRGR</u> QTYTRYQLELEKEFHFNH---YLTRRRRIEI
even-skipped	VRRYRTAF <u>TRD</u> QLGRLEKEFYKEN---YVSRPRRCCL
paired	QRRSRTTF <u>SASQ</u> LDELE <u>RAF</u> ERTQ---YDPDIYTREEL
Majority	ARRGRTTFTK <u>QQLLELEKEFH</u> SNR---YLSRERREEL
	^ * ^ ^ * ^ *
Msx-1	SSSLSLTETQ <u>VKIWFQ</u> NRRRAKAKRLQ
MATa1	AKKCGITPLQVRVWFINKRMRSK
MATalpha2	MKNTSLSRIQIKNWVSNNRRKEKTTIT
pbx	AKKCGITVSQVSNWFGNKRIRYKKNI
hoxb1	AATLELNETQ <u>VKIWFQ</u> NRRMKQKKRE
ubx	AHALCLTERQIKIWFQNNRMKLKKEI
exd	ARKCGITVSQVSNWFGNKRIRYKKNI
engrailed	SSELGLNEAQIKIWFQNKRAKIKKST
antennapedia	AHALCLTERQIKIWFQNNRMKWKKEN
even-skipped	AAQLNLPESTIKVWFQNNRMKDQRQR
paired	AQRTNLTEARIQVWFQNNRRARLRKQH
Majority	AKKGLLTESQIKIWFQNNRMKLKKEI

Figure 3.5. Sequence Alignment of Homeodomains. In the majority sequence every 10th residue is underlined. The Msx-1 residues involved in DNA recognition are denoted by an asterisk (*), and those involved in HD core stabilization are marked with a carat (^).

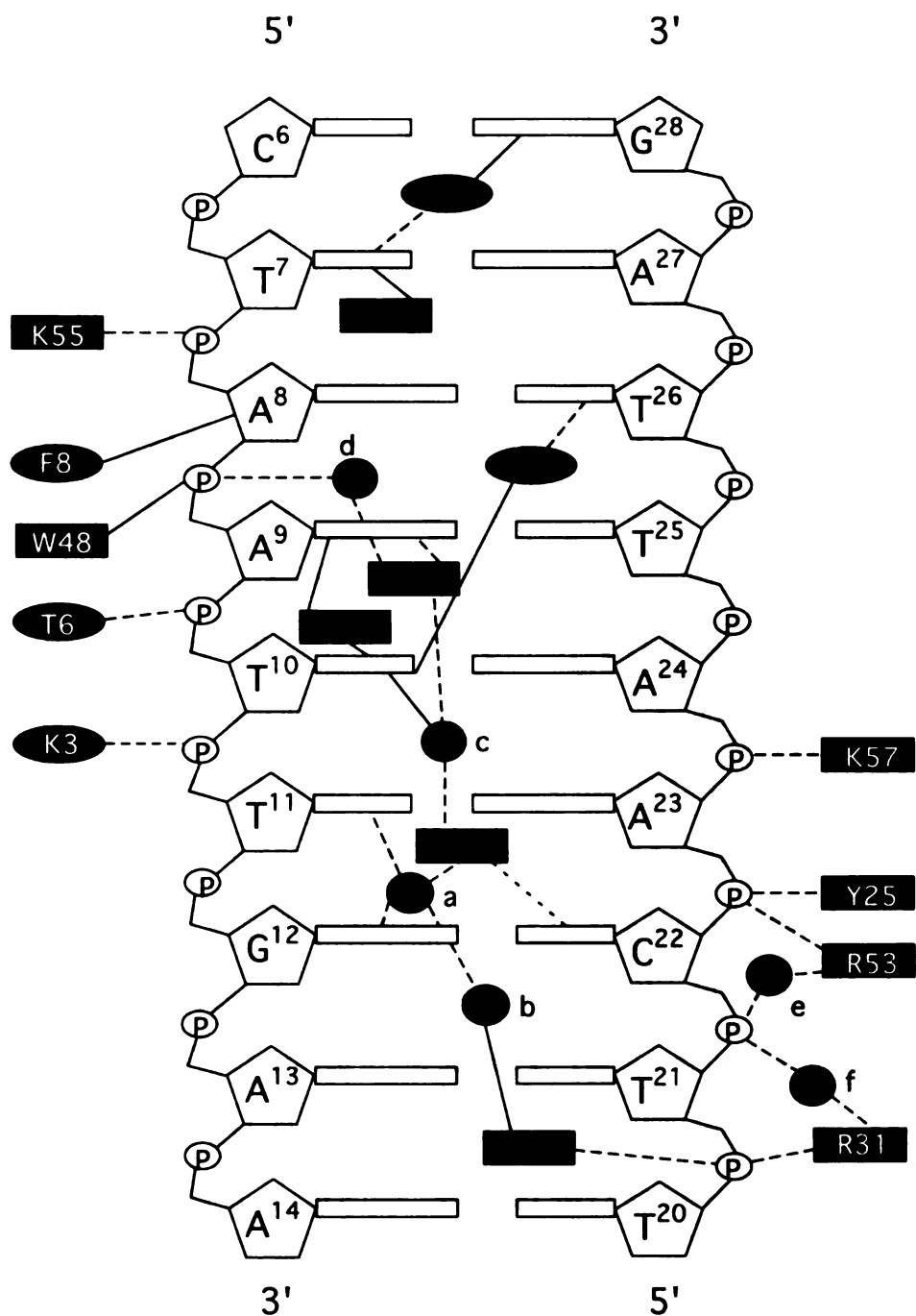


Figure 3.6 The contacts between the DNA and the protein in the complex. The major and minor groove DNA contacts are shown by squares and circles respectively. Dotted lines indicate hydrogen bonding, solid lines indicate hydrophobic interactions.

3.2.1 Residue Q50

Residue 50 deserves special mention because it is one of the critical residues involved in discriminative DNA recognition for distinct classes of homeodomains. When residue 50 is K, there is a clear preference for 5'-TAATCC-3' sequences over other sequences such as GG and TA. The structure of the Q50K mutant *engrailed* homeodomain bound to a GG containing DNA sequence clearly explains this preference by showing a direct hydrogen bond between K50 and the GG of the DNA (5). What is more difficult to explain, however, is why Q50 variants *disfavor* the GG sequence relative to other sequences such as TG, which is part of the core binding sequence for the Msx-1 homeodomain, or TA, which is the preferred sequence for the wild type *engrailed* homeodomain. We believe that GG sequences are disfavored both by the Msx-1 and *engrailed* homeodomains because of the water mediated interaction between Q50 and DNA. In the Msx-1 homeodomain/DNA complex there is a water-mediated interaction between Q50, Thy11, Gua12 and Wat a (Figure 3.7). The 26.9 B-factor of Wat a is low compared to the average B-factors for all of the waters in the structure (43.6), indicating this water to be quite well ordered in our structure. The coordination sphere of Wat a interacts with Q50, Thy11, Gua 12 and Wat b serving to completely define its hydrogen bonding pattern. Substitution of either Thy11 or Gua12 with a Cyt nucleotide would cause a hydrogen bonding clash with Wat a necessitating some structural reorganization in this region. An almost identical duplication of the water-mediated interaction seen in the Msx-1/DNA complex can be seen in the *engrailed*/DNA complex structure (6). Binding studies for both the Msx-1 and engrailed homeodomain/DNA complexes have

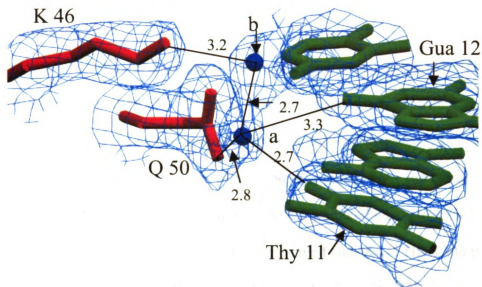


Figure 3.7 Simulated annealing omit map of the Q50 – water – DNA interaction, contoured at 1.5σ . Picture was made with Setor (7).

shown that the presence of a Cyt nucleotide at either of these positions causes a significant loss of binding affinity for each of these complexes (5, 8). This interaction interface defines most of the Msx-1 base specificity flanking the core TAAT sequence. The conclusion to be drawn from this is that the fully coordinated Wat a is *the* critical component in the Thy11/Gua12 base specificity for Msx-1 and other homeodomains which favor this sequence.

An example of a structural reorganization in this region can be seen in the HoxB1-Pbx1/DNA complex structure (4). Though a water is present very near to the location of Wat a, the presence of a Cyt nucleotide has forced Q50 to move completely out of the region. Q50 actually makes interactions with the phosphate backbone in this structure. The paired HD structure also shows an important water mediated interaction for Q50 but the DNA in this region is a bit different from the engrailed and Msx-1 HD's and therefore we do not see this exact pattern in the paired structure (9). The paired HD also has a mutation of the DNA from TAATCA to TAACGA which encompasses part of the core sequence. There is a reduction in binding from this mutation but how much is due to the flanking sequence is hard to infer (9).

3.2.2 Residue A54

In most homeodomains, residue 54 is a reasonably large hydrophobic residue (I or M) that fills a cavity that exists between the recognition helix and the surface of the DNA, often making direct interactions with the DNA (2, 4, 10-12). In contrast, the Msx-1, *engrailed*, and paired homeodomains have alanines at this position (6, 9). In Msx-1 and *engrailed* the lack of a large side chain produces an identical ring of ordered water

molecules that surround the residue, some of them making interactions with the DNA (Figure 3.8). These water-mediated interactions serve to replace the direct hydrophobic interaction that was lost. The paired HD has four of these waters but also has an R at position 57, which swings into this area to fill the gap. In the other two structures with A54, K57 has a different conformation that does not interfere with the water ring.

3.3 DNA minor groove and n-terminal arm interactions.

In addition to the interface between the HD recognition helix and the DNA major groove, which is reminiscent of bacterial helix-turn-helix/DNA interactions, homeodomains have additionally an n-terminal arm region lacking secondary structure that snakes across the minor groove, making critical interactions with the DNA. The Msx-1 HD n-terminal arm makes three direct interactions with DNA (Figure 3.9). T6 makes a hydrogen bond to the DNA phosphate backbone, which is commonly found in other HD structures containing this residue (4, 6, 13). R5 makes a direct base contact to Thy7, which is virtually identical to the interaction seen in all other HD/DNA structures that have an R at position 5. In contrast to most other HD/DNA structures, however, the n-terminal arm of Msx-1 is well ordered all the way to R2. In fact R2 forms a tight hydrogen bonding/salt bridge interaction with base Thy26, this can also be seen in the paired structure (9). The average B factor of the n-terminal arm residues in Msx-1 is 40 Å² which compares well with the average B factor of the entire protein (32.4 Å²), indicating the n-terminal arm to be well ordered relative to the rest of the structure. We believe that the structural integrity of the Msx-1 n-terminal arm is partially preserved

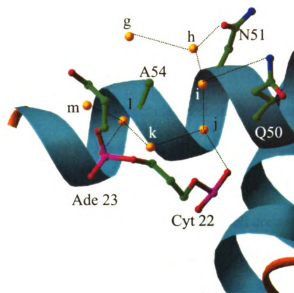


Figure 3.8 The conserved water ring that surrounds A54 and fills the cavity present between the protein and the DNA backbone in this region.

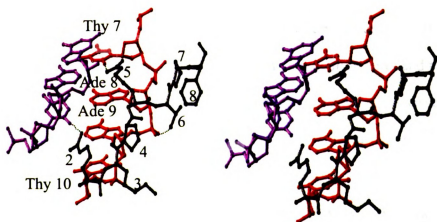


Figure 3.9 Stereoview of the trajectory of the N – terminal arm of Msx-1. Hydrogen bonds are represented by dotted lines.

by the presence of two unconserved prolines in the sequence. The paired structure does not contain these prolines but is involved in a dimerization interaction with helix II and helix III of its second HD.

3.4 Hydration of the HD/DNA interface

The protein-DNA interface not only includes protein-DNA interactions, but protein-water and DNA-water interactions as well. Though the complex buries 1832 \AA^2 of surface area between the protein and DNA, there is also significant hydration within the interface. It turns out that this hydration is remarkably well conserved between the homeodomain/DNA structures. By overlaying all known HD/DNA complex structures and looking for conserved waters, i.e. waters in different structures that were no more than 1.5 \AA away from corresponding Msx-1 waters, we observed a remarkable conservation of the water structure in these complexes (Table 3.2). Table 3.2 contains those HD complexes that diffract to better than 2.5 \AA . Figure 3.10 depicts the Msx-1/DNA structure with all 16 of the most conserved waters. Note how virtually all of these waters create a hydration sphere *between* the recognition helix and the major groove of the DNA. 130 of the possible 180 conserved waters are observed (72%). 23 of the 50 missing waters are in the three structures whose water structure is least conserved (MAT α 1, MAT α 2 and *Antenpedia* HD/DNA complexes). Of the 130 conserved waters identified in this analysis only sixteen are more than 1.2 \AA away from the corresponding waters of the Msx-1 structure, and most are within 1 \AA of an Msx-1 water. Thirteen of the sixteen waters make bridging interactions between the protein and DNA while three make contacts only to the DNA. The overwhelming majority of these waters are making

similar interactions in each of the HD/DNA structures. All of these interactions and distances are tabulated in the appendix. It appears that the strong conservation in HD/DNA interactions extends to the sphere of hydration of these complexes. Interestingly, no conserved waters are seen outside the protein/DNA interface. These waters have a low average B-factor of 30 Å² when compared to the average water B factor in the Msx-1/DNA complex which is 43.6 Å².

Table 3.2 Conserved water table

HD/DNA Complex	Res. Limit (Å)	# Conserved waters	Data Collection Condition	Total # waters	PDB ID#
Antennapedia	2.4	8	room temp	38	9ANT
MATa1/MATα2	2.5	α1-7, α2-10	frozen	58	1AKH
Even-skipped	2.0	12	frozen	68	1JGG
HoxB1/Pbx1	2.35	hox-10,pbx-13	frozen	61	1B72
Ubx/Exd	2.4	ubx-15,exd-12	frozen	110	1B8I
Engrailed	2.2	13	room temp	53	3HDD
Engrailed Mutant	1.9	14	frozen	183	2HDD
Paired	2.0	12	frozen	242	1FJL
Msx-1	2.2	16	frozen	153	1IG7

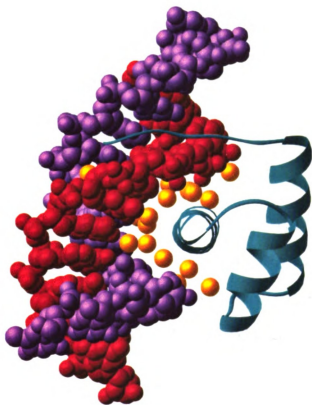


Figure 3.10 Conserved water network present in the homeodomain – DNA complexes studied. The waters are shown in gold.

pr

60

to

Th

gr

inc

the

to

the

be

gle

Di

HI

th

rel

so

tw

sig

for

the

3.5 Structure of the DNA

Most of the DNA/HD complexes have a modest bend in the DNA caused by protein binding. In the MAT α 1/MAT α 2/DNA heterodimer complex there is a significant 60° bend in the DNA (2) while in the majority there is a modest bend of about 10-13° toward the major groove (2, 4-6, 11-13). The paired HD complex shows a 21° bend (9). This bend appears to be caused by the protein pulling the DNA toward it on the major groove side. This deformation is very common in major groove binding proteins including helix turn helix proteins and other DNA binding proteins (14-17). The DNA in the Msx-1 HD/DNA complex, however, exhibits a much more severe bend of 28° relative to other monomeric HD/DNA complexes. The very close structural homology between the Msx-1 HD/DNA complex and other HD/DNA complexes indicated that the extra bending was not due to protein/DNA interactions. Further evidence of this can be gleaned from Figure 3.11 where the base roll angle is plotted for each sequence in the DNA. The roll angle is shown for both the Msx-1 HD/DNA complex and for the HoxB1 HD in the HoxB1-Pbx1/DNA complex, a structure that is highly homologous to that of the Msx-1 HD/DNA complex (4). The base roll parameter can give an indication of relative bend per base for a DNA sequence, since a DNA bend must be accommodated by some base roll. As shown in Figure 3.11, the base roll per residue is very similar for the two structures in the core binding region, but the Msx-1 base roll parameters deviate significantly at the end of the sequence where the crystal packing-induced triple helix is formed. It is apparently the formation of this triple helix that causes the larger bend in the DNA of the Msx-1 HD/DNA complex.

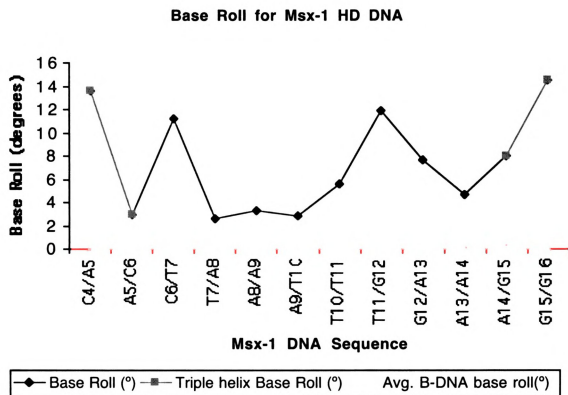


Figure 3.11 A. Plot of DNA base roll as a function of DNA sequence. Horizontal dotted or dashed lines indicate average values for B-DNA. The sequence of the top strand only is shown. All parameters were calculated with the program Curves (18).

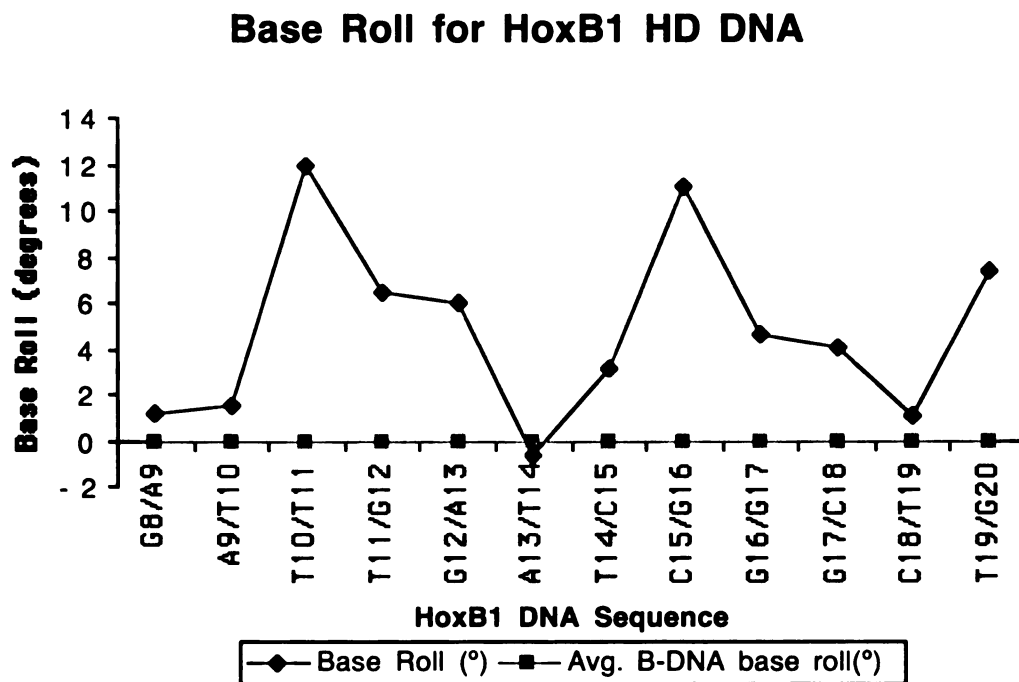


Figure 3.11 B. Plot of DNA helical twist (B) as a function of HoxB1 DNA sequence. Horizontal dotted or dashed lines indicate average values for B-DNA. The sequence of the top strand only is shown.

This triple helix formed when the first two bases of the double stranded portion of the DNA melted (Gua2 and Thy3). One of the resulting single stranded segments (Ade31 and Cyt32) then formed a triple helix interaction with the end of a neighboring DNA in the crystal, while the other single stranded segment (Thy1, Gua2 and Thy3) was not seen in the structure and is likely disordered. The result is that a triple helix is formed between Gua15-Cyt19 (Cyt32) and Gua16-Cyt18 (Ade31) with the base in parentheses being the third base forming the triple helix between the Watson and Crick base pairs. In addition, a third triple helix base step is formed by the interaction of an overhanging Thy17 with Cyt4-Gua30, which is butted against Gua16-Cyt18 in the crystalline lattice. The complete triple helix is depicted in Figures 3.12 and 3.13. There are a few possible explanations for this phenomenon. Some DNase I footprinting experiments have shown that there is a propensity for triple helices to form in GA and GT rich oligonucleotides (19) such as ours. The Msx-1 DNA has an AGG sequence in the region where the triple helix forms. Also, in all three steps of triple helix seen in this structure base protonation is observed. We believe this protonation to be a result of the comparatively low pH (pH = 4.6 in the crystallization). The protonation values for the nucleotides appear to support this theory with the pKa's of the ring nitrogens being around pH 4 (20). This low pH served to both destabilize the double helix and to stabilize the triple helix by base protonation. Figure 3.14 shows the Gua16-Cyt18 (Ade31) triple helix step, a structure not seen previously in DNA or protein/DNA structures.

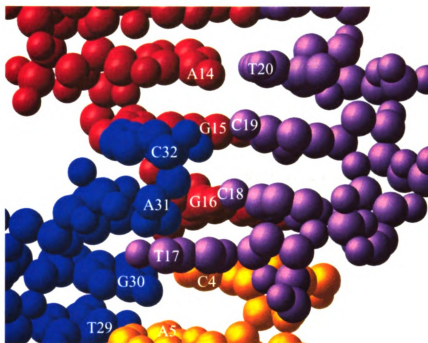


Figure 3.12 Overall triple helix interaction for the stacking DNA. Strand 1 is in red and yellow while strand 2 is in purple and blue.

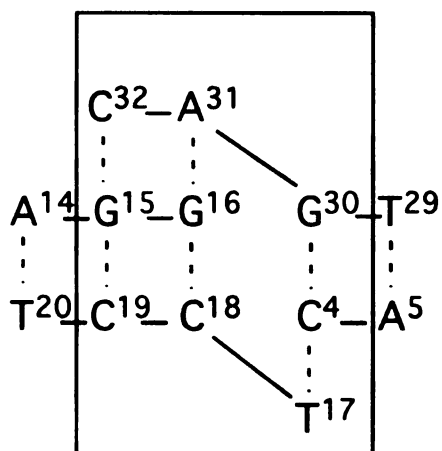


Figure 3.13 Triple helix schematic for the unusual helical interaction of the stacked DNAs.

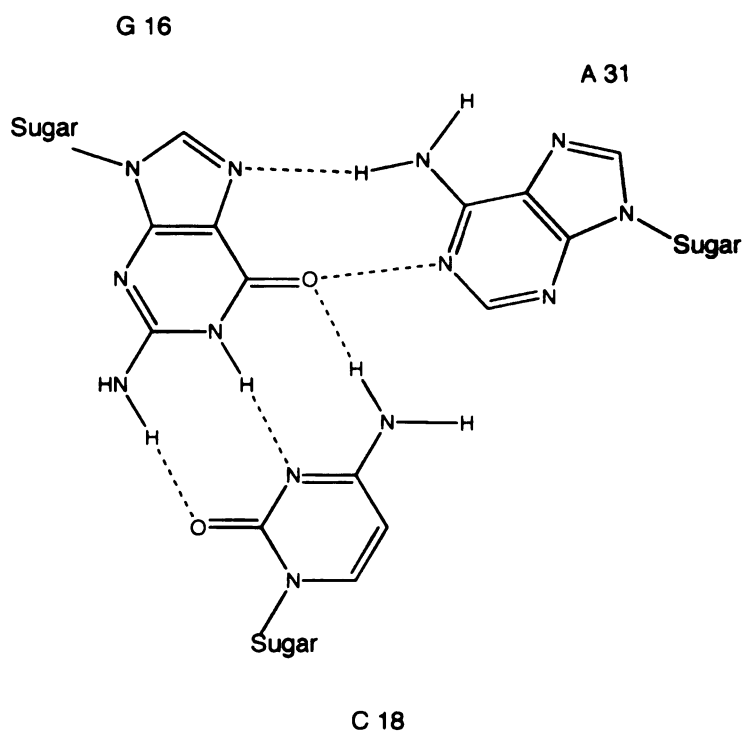


Figure 3.14 The triple helix trio of Gua16:Cyt18 and Ade31, a previously unseen interaction in triple helix combinations.

The source of the 15° bend in the region of the triple helix is more difficult to explain, since triple helix formation has not previously been correlated with DNA bending. Clearly, however, in this unusual case of triple helix formation, modest DNA bending is a result. There are examples in the literature of a triple helix forming at the ends of double stranded DNA with overhangs (21-23) and these have been utilized in designing DNAs for crystallization purposes. The triple helix in this case is not a pseudo continuous helix but one that has packed against their opposing strands. So it is a rare occurrence in itself but the way that it has formed makes it even more unusual.

3.6 Msx-1 HD protein interactions.

In addition to its interaction with DNA, the Msx-1 HD makes specific and functionally important interactions with other proteins as well. Specifically GST pulldown experiments and gel shift analyses have shown that Msx-1 binds TBP and the TBP/TATA box DNA complex. While residues in the n-terminal arm of the homeodomain are required for this interaction (1), residues in other parts of the HD had little effect on this interaction.

Specifically the F8A, R5A, K3A triple mutation completely abrogated Msx-1 HD/TBP interaction. All three of these residues reside in the n-terminal arm. In contrast, both the double mutant, L16A, F20A; and the triple mutant I47A, E50A, N51A bind TBP with near wild type affinity. While L16 and F20 lie in helix one, I47, E50 and N51 are all located in the DNA binding interface of helix 3. All of these mutants have lost measurable DNA binding affinity, indicating that the interaction with TBP is independent of the protein's DNA binding interface. Both the helix one mutants and the n-terminal

arm mutants were inactive toward transcriptional repression in the context of the full length protein. Mutants in helix three, on the other hand, were still capable of transcriptional repression, again indicating that the DNA binding interface is not required for this activity. The two residues in helix I are part of the conserved hydrophobic core, and when mutated to Alanine probably cause important changes in the structure of the HD. Our structure indicates that the structure of the Msx-1 n-terminal arm is somewhat unique relative to many other HDs and also indicate it to be comparatively less flexible. It is possible that this increased structural rigidity is important in defining a TBP binding interface.

In addition to its interaction with TBP, a member of the transcriptional basal machinery, Msx-1 appears to make interactions with several other members of the homeobox gene family. These interactions appear to play important regulatory roles in transcription. The HD protein Dlx, a transcriptional activator, is one of these proteins. Heterodimerization between Msx-1 and Dlx causes mutual loss of transcriptional activity (24). Mutations both in the n-terminal arm and in the recognition helix of Msx-1 severely compromise this interaction, indicating the interaction to encompass the entire DNA binding surface of the protein. Presumably, a similar interaction surface is used in the interactions between Msx-1 and other homeodomain proteins such as Pax3 and Dlx2 (25, 26).

The same mutations studied above for TBP binding were also studied for Dlx binding. The n-terminal arm residues again proved important as the Msx-1/Dlx interaction was demolished when residues in this region were mutated. Helix I and helix III mutations also reduced dimerization, indicating that the interaction is at least partially

mediated via the recognition helix. The dimerization of the Msx and Dlx proteins seems to differ from the interaction of Msx-1 with TBP. Gel retardation assays indicate that Msx and Dlx proteins bind to homeodomain sites as monomers. Dimerization excludes DNA binding and mutual repression of their transcriptional activities is the result.

3.7 Conclusions

Comparison of our 2.2 Å Msx-1 homeodomain/DNA complex structure with all other homeodomain/DNA complexes has illustrated some enlightening aspects of homeodomain structure and DNA recognition. Most notably comparison of ordered water molecules between HD/DNA complex structures has identified a well conserved hydration interface between the recognition helix and major groove. More insight has been gained concerning how residue Q50 determines HD binding site preferences. The coordination sphere of Wat a (a very conserved water) dictates the interaction of Q50 with the DNA flanking sequence. Mutations of this flanking sequence led to a significant loss in binding affinity of the complex, mainly due to the disruption of the critical water coordination sphere of Wat a. Our structure has reaffirmed the importance of water-mediated interactions for HD/DNA binding.

There are also important differences between our structure and those of other structures. The n-terminal arm of Msx-1 is unusually well ordered, exhibiting clear electron density to the second residue. We believe that two proline residues not usually seen in HD n-terminal arms stabilize the structure. A combination of HD/DNA interactions and the formation of an unusual triple helix packing interaction leads to a significant, 28° bend in the DNA, which is significantly larger than that seen in other

monomeric HD/DNA complex structures. Though an artifact of crystal packing interactions, the triple helix formed between stacked DNA helices provides an interesting example of triple helix structure. It contains an unusual GC (A) step not previously seen in triple helices and leads to a significant DNA bend not typically seen in triple helix structures.

3.8 References

1. Zhang, H., Catron, K. M., and Abate Shen, C. (1996) *Proc. Natl. Acad. Sci. U S A.* 93, 1764-9.
2. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) *Science.* 270, 262-9.
3. Tan, S., and Richmond, T. J. (1998) *Nature* 391, 660-6.
4. Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., and Wolberger, C. (1999) *Cell* 96, 587-97.
5. Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T., and Pabo, C. O. (1997) *Structure* 5, 1047-54.
6. Fraenkel, E., Rould, M. A., Chambers, K. A., and Pabo, C. O. (1998) *J. Mol. Biol.* 284, 351-61.
7. Evans, S. V. (1993) *J Mol. Graph* 11, 134-8, 127-8.
8. Catron, K. M., Iler, N., and Abate, C. (1993) *Mol. Cell. Biol.* 13, 2354-65.
9. Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) *Cell.* 82, 709-19.
10. Hirsch, J. A., and Aggarwal, A. K. (1995) *Embo J.* 14, 6280-91.
11. Fraenkel, E., and Pabo, C. O. (1998) *Nat. Struct. Biol.* 5, 692-7.
12. Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999) *Nature* 397, 714-9.
13. Hirsch, J. A., and Aggarwaal, A. K. (1995) *EMBO J.* 14, 6280-6291.
14. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) *Cell.* 77, 21-32.
15. Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993) *Nature* 364, 412-20.
16. Konig, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996) *Cell* 85, 125-36.
17. Mondragon, A., and Harrison, S. C. (1991) *J. Mol. Biol.* 219, 321-34.
18. Lavery, R., and Sklenar, H. (1988) *J. Biomol. Struct. Dyn.* 6, 63-91.

19. Chandler, S. P., and Fox, K. R. (1996) *Biochemistry* 35, 15038-48.
20. Ts'o, P. O. P., and Eisinger, J. (1974) *Basic principles in nucleic acid chemistry*, Academic Press, New York,.
21. Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., and Sigler, P. B. (1991) *Nature* 352, 497-505.
22. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) *Science* 253, 1001-7.
23. Van Meervelt, L., Vlieghe, D., Dautant, A., Gallois, B., Precigoux, G., and Kennard, O. (1995) *Nature* 374, 742-4.
24. Zhang, H., Hu, G., Wang, H., Sciavolino, P., Iler, N., Shen, M. M., and Abate-Shen, C. (1997) *Mol. Cell. Biol.* 17, 2920-2932.
25. Bendall, A. J., Rincon Limas, D. E., Botas, J., and Abate Shen, C. (1998) *Differentiation* 63, 151-7.
26. Bendall, A. J., Ding, J., Hu, G., Shen, M. M., and Abate-Shen, C. (1999) *Development* 126, 4965-76.

CHAPTER IV: The Three Dimensional Structure of the Oct-1 POU/ U1 octamer/ SNAP190 ternary complex.

4.1 Overall structure of the Oct-1/U1/SNAP190 ternary complex

The three dimensional structure of the Oct-1 POU/U1 octamer/SNAP190 ternary complex represents the first structure of a transcriptional activator in a complex with its general transcription factor target. The structure of the Oct-1 POU protein has been studied bound to DNA and additionally DNA and OCA-B (1, 2). It has two distinct domains a POU specific (POU_s) and a POU homeodomain (POU_{HD}) connected by a flexible 24 amino acid linker. The Oct-1 POU protein has been solved in complex with the optimal binding site from the H2B promoter (ATGCAAAT) to 3.0 Å. The Oct-1/DNA/OCA-B structure also utilized the H2B octamer site while including a B-cell specific co-activator OCA-B to a resolution of 3.2 Å. The Oct-1/U1 octamer /SNAP190 complex has been solved to a resolution of 2.4 Å and includes all but 10 residues of the linker, the base overhangs of the DNA and 3 residues of the SNAP190 peptide. There are 244 water molecules in both complexes in the asymmetric unit.

The POU_s domain consists of 4 alpha helices while the POU_{HD} domain consists of 3 alpha helices and a flexible N terminal arm that is the same as all of the other homeodomains previously discussed (Figure 4.1). There are three major parts of the protein involved in DNA binding, helix 3 of the POU_s, the n-terminal arm of the POU_{HD}, and the recognition helix of the POU_{HD}. The SNAP190 peptide has an alpha helix that interacts with POU_s and a flexible portion that interacts with the DNA. The DNA used in

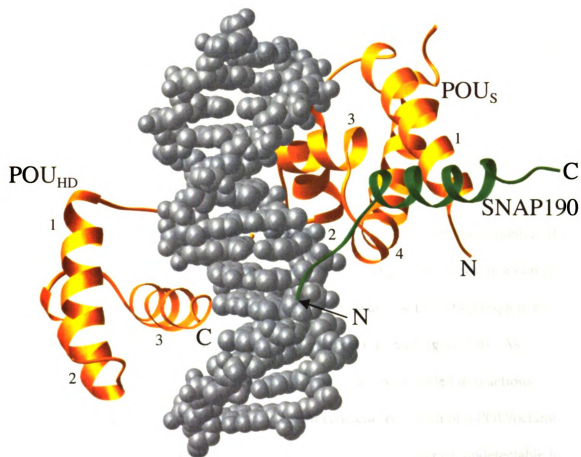


Figure 4.1. The three dimensional structure of Oct-1 POU/ U1 octamer/ SNAP190 peptide at 2.4 Å. The DNA is in silver, the Oct-1 POU protein in gold, and the SNAP190 is in green. This picture was created with the Ribbons program (3).

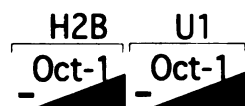
this ternary complex contains the U1 octamer binding site but uses the flanking sequences found in the H2B structures. This structure gives us insight into the interaction of Oct-1 and SNAP190 with the U1 octamer in particular and leads us to rethink how activators work.

4.2 SNAP190 assists Oct-1 binding to the U1 octamer

As shown in Figure 4.2, the U1 octamer element differs from the canonical H2B octamer sequence at position 4, where a T replaces C, and position 6, where a G replaces A. To assay the relative affinity of the Oct-1 POU domain for these two sequences, electrophoretic mobility shift assays (EMSA) were performed (Figure 4.3). As increasing amounts of Oct-1 POU domain (0.1 and 1 ng) were added to reactions containing a high affinity H2B octamer element efficient formation of a POU/octamer complex was observed (compare lanes 2 and 3 to lane 1). In contrast, undetectable levels of complex formation were observed for reactions containing 0.1 or 1 ng of the Oct-1 POU domain and the human U1 octamer element probe (lanes 5 and 6, respectively). Overall, the relative affinity of the Oct-1 POU domain for the human U1 octamer is several orders of magnitude lower than for the H2B octamer element (Figure 4.3). The weak affinity of the Oct-1 POU domain for the U1 octamer is surprising, given that the human U1 snRNA genes are highly expressed. As discussed, the SNAP190 subunit of SNAP_C is one target for Oct-1 POU domain activation of human snRNA genes. To determine whether SNAP190 can stimulate DNA binding by Oct-1 electrophoretic mobility shift assays were performed with the Oct-1 POU domain and a region of

		-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12			
U1	5'	-	T	G	T	A	T	G	T	A	G	A	T	A	A	G	G	-	3'
H2B	5'	-	T	G	T	A	T	G	C	A	A	A	T	A	A	G	G	-	3'

Figure 4.2 DNA sequence used in crystallization (U1) – top strand shown only, compared to the H2B octamer. Base changes are shown in bold. . The octamer sequence is numbered 1-8, with the equivalent base on the opposite strand indicated by a prime (A4' corresponds to the base paired to T4 for example).



-DNA/Oct-1



Figure 4.3 Electrophoretic mobility shift assays were performed using 0.1 ng (lanes 2 and 5) or 1 ng (lanes 3 and 6) of human Oct-1 POU-domain protein with DNA probes containing a human histone H2B (lanes 1-3) or U1 snRNA (lanes 4-6) octamer element. Lanes 1 and 4 contain the probes alone. The position of the POU complex is indicated (DNA/Oct-1). Experiment done by Dr. Craig Hinkley.

SNAP190 containing amino acids 884-910, which was previously shown to contact Oct-1 POU domain during transcriptional activation by Oct-1 (4). As shown in Figure 4.4, efficient complex formation is observed when 1 ng of the Oct-1 POU domain is incubated with the high affinity H2B octamer element (lane 2). In contrast, weak complex formation was observed in reactions containing 30 ng of the Oct-1 POU domain and the U1 octamer element, as expected (lane 4). Interestingly, when the SNAP190 (884-910) peptide (sequence in Figure 4.5) was also included in these reactions significant enhancement of DNA binding by the Oct-1 POU domain was now observed (lane 5). The effect of the peptide was specific because comparable amounts of an irrelevant peptide had no effect on DNA binding by Oct-1 POU (lane 6). In addition, neither of the peptides could bind DNA in the absence of the Oct-1 POU domain (lanes 7 and 8). Therefore, SNAP190 stimulates Oct-1 POU domain binding to the weak U1 octamer element.

4.3 Structure of the Oct-1/ U1 octamer / SNAP190 peptide.

Recognition of regulatory elements by transcriptional activator proteins is presumed to be a prerequisite for subsequent recruitment of the general transcription machinery to gene promoters. Yet, the Oct-1 POU domain only weakly recognizes the human U1 octamer element though binding is stimulated by SNAP190, suggesting that synergistic promoter recognition by these factors contributes to the activation of these genes. In order to understand the mechanism for SNAP190-mediated enhancement of

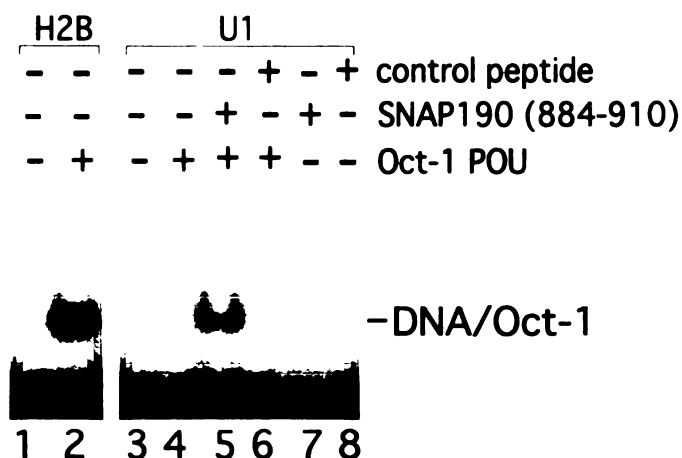


Figure 4.4 Electrophoretic mobility shift assays were performed using DNA probes containing a human histone H2B (lanes 1 and 2) or U1 snRNA octamer element (lanes 3-8) with 1 ng (lane 2) or 30 ng (lanes 4-6) of human Oct-1 POU-domain protein alone (lanes 2 and 4), with 10 μ g SNAP190 peptide (lane 5) or with an equimolar amount of a control peptide (lane 6). Lanes 7 and 8 contain the SNAP190 or control peptides alone, respectively. Lanes 1 and 3 contain probe DNA alone. The position of the POU complex is indicated (DNA/Oct-1). Experiment done by Dr. Craig Hinkley.

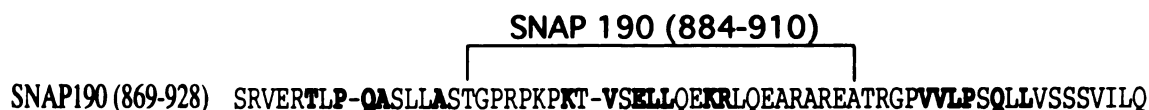


Figure 4.5 SNAP190 Peptide sequence with the portion used in crystallization indicated. The residues in bold show sequence identity to the OCA-B peptide.

DNA binding by the Oct-1 POU domain, X-ray structural analysis of a ternary complex of SNAP190, Oct-1 POU domain, and the human U1 octamer element was pursued. The ternary complex was formed using the 27 residue SNAP190 peptide encompassing amino acids 884-910, the Oct-1 POU domain containing both the POU_S and POU_{HD} DNA binding modules, and a 14mer DNA oligonucleotide based on the non-canonical octamer sequence found in the U1 DSE. This arrangement buries 3124Å² of surface area compared to 3858Å² for the OCA-B structure. The relative orientation of the POU_{HD} and POU_S domains is similar to that seen in the structures of the Oct-1 POU/H2B octamer complex and the Oct-1/H2B/OCA-B complex, with the two DNA binding modules binding to the major groove on opposite sides of the DNA (Figure 4.1). While the POU_S interacts with the first four basepairs in the sequence (ATGT), the POU_{HD} interacts with the last four basepairs (AGAT). Thus, interactions between SNAP190 and the Oct-1 POU domain barely perturb the overall structure of the POU domain. A more detailed characterization of the atomic interactions is described in Figures 4.6 and 4.7. Figure 4.6 shows the protein interaction with the DNA while figure 4.7 details the POU_S and SNAP190 interactions. The complete list of SNAP190/Oct-1 POU interactions are listed in the appendix.

In this structure, the ordered portion of the SNAP190 peptide begins at residue R887, which makes a salt bridge to the phosphate backbone of the DNA at position A8'. An additional contact between SNAP190 P890 and the phosphate backbone at position T7' is also observed. Thus, DNA contacts by SNAP190 within the octamer element contribute to stable DNA binding by this complex. The SNAP190 chain then traverses the phosphate backbone and ends in a 4-turn helix containing residues 892-906. This

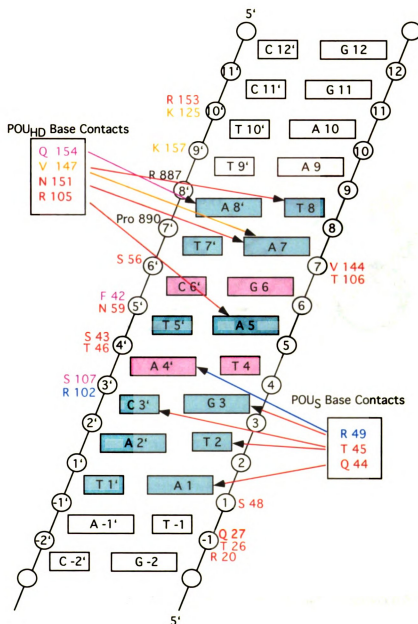


Figure 4.6 Schematic representation of the protein/DNA contacts within the SNAP190/Oct-1/U1 octamer complex. The red contacts and arrows are the same in all three structures. Orange represents those found in SNAP190 and Oct-1 only. Pink represents those found in SNAP190 and OCA-B only. Blue contacts are unique to our structure. Those residues in black are the SNAP190 peptide contacts to the DNA.

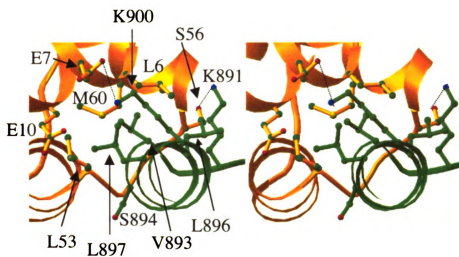


Figure 4.7. Hydrophobic interactions dominate the interaction between SNAP190 and Oct-1. A stereo view of the POU_s interaction with the SNAP190 C-terminal helix with the POU domain (gold) and SNAP190 (green) is shown. The view is looking down the SNAP190 helix. There are several hydrophobic interactions and two hydrogen bonds shown with dotted lines.

helix packs snugly against a surface of the POU₅ domain, making extensive contacts with POU₅ domain helix 1 and the loop connecting helices 2 and 3. The interactions between SNAP190 and Oct-1 are largely hydrophobic, with the core of the interaction defined by T892, V893, S894, L896 and L897 of SNAP190. The side chains of these residues make several hydrophobic contacts with the side-chain residues in the POU₅ domain, including L6, E7, E10, L53 and M60. Three hydrogen bonds are also made involving the main-chains of SNAP190 and OCT-1 POU₅: the SNAP190 V893 main-chain amide nitrogen to the POU₅ L55 main-chain carbonyl, the SNAP190 S894 main-chain amide nitrogen to the POU₅ L53 main-chain carbonyl, and the side-chain oxygen of T392 to the main-chain oxygen and nitrogen of L55. A perfect complement of shape is revealed in this interaction where the side-chains of the two peptides direct the main-chain atoms to make these hydrogen bonds. It is this shape complement that renders specificity to the interaction in spite of the dominance of main-chain hydrogen bonds in the interface. Notably, there are two side-chain hydrogen bonds seen in our structure, one is between K900 of SNAP190 and E7 of the POU₅ domain (Figure 4.8) and the other is seen between S56 and K891. The side chain of K891 has two conformations in both molecules one of which binds to S56 at 3.2Å and the other interacts with E899 of the SNAP190 peptide at 2.7Å. Interestingly, the K900 – E7 interaction is sufficient to dictate activator specific regulation of human snRNA gene transcription by POU domain proteins and is critical for transcriptional activation of human snRNA genes by the Oct-1 POU domain (4-6). This interaction is buttressed by a salt bridge between SNAP190 K900 and E904 that correctly positions K900 for this critical interaction with the Oct-1 activator. Importantly, this well – coordinated interaction helps to maintain the alignment

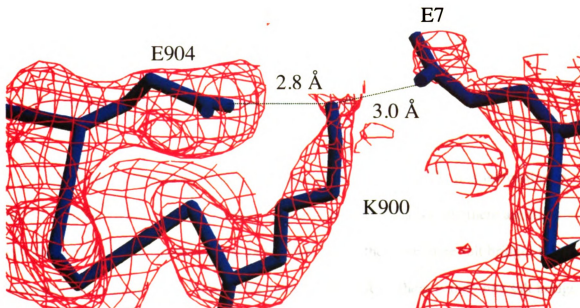


Figure 4.8 A key determinant of transcriptional specificity within Oct-1 is well positioned for hydrogen bonding with SNAP190. Shown is a simulated annealing omit electron density map contoured at 1.8σ around SNAP190 K900, Oct-1 POU_s E7, and SNAP190 E904. SNAP190 E904 buttresses K900, accurately positioning it to make a critical salt bridge with POU_s E7. All of the protein is shown in dark blue. This figure was made using Setor (7).

of the critical hydrophobic interface between the Oct-1 POU domain and SNAP190 and explains, in part, the role of this single salt bridge in defining transcriptional specificity by POU domain proteins.

4.4 Comparison of Oct-1 POU to other HDs and POU proteins

The Oct-1 homeodomain can be compared to the Msx-1 HD. It is interesting to note that although the sequence identity among homeodomains is high, there are some key contacts that are disrupted in the Oct-1 HD. While there are three salt bridges that were thought to stabilize the Msx-1 HD, in the case of Oct-1 there is only one salt bridge between helix one and helix three (3.73 Å). This salt bridge is between Glu 117 and Arg 152 which is a very common interaction in the HDs. The other two key salt bridges are not present in the Oct-1 HD because of the difference in the sequence. This difference can be seen in Figure 4.9 which shows an alignment of many POU proteins. It appears that the additional salt bridges are not as important when there is the POU₅ domain there to stabilize things. The POU domains bind to widely different DNA sequences, but there are a few critical contacts that monomer HDs make to the DNA that are also found in POU HDs. The invariant N151 always makes a contact to an Ade in the major groove of the DNA and this is also seen in the POU proteins. The conserved I147 in HDs (see Figure 1.5) is a V in the POU proteins, but still makes contact with a thymine in the major groove. Q154 is also conserved in the POU family unlike the rest of the HDs which have a variable residue in this position. In all of the structures of POU domains this Q154 interacts with an Ade in the major groove. These three contacts come from the

1 75
 Oct-1 EEPDLEELEQFAKTFKQRRIKLGFQTQGDVGLAMGKLY---GNDFSQTTISRFEALNLSFKNMCKLKP LLEKWLNDAE
 Oct-2R.....
 Oct-6 .DAPSSDD.....Q.....A.....L.T.....V.....C.....Q.....N...EETD
 Oct-4 DMKALQK.....LL.K.T.Y.A.....TL.V.F---KV.....C.....Q.L.....R.....VEE.D
 Pit-1 MDSPEIR.....NE.V.....Y.TN.E.LAAVH---SE.....C.N.Q.....A..AI.S...EE..
 Unc-86 DMDT.PRQ..T..EH.....V.A...K.LAH.KMPGV.S-L..S.C..S.T..HN.VA...I.HS..EK..

α1 α2 α3 α4

Oct-1 NLSSDSSLSSPSALNSPGIEGLSR
Oct-2 TMSVDSLSPNQLSSPSLGFDGLPGR
Oct-6 SSSGSTNLDKIAAQGR
Oct-4 NNENLQEICKSETLVQA
Pit-1 QVGALYNEKVGANER
Unc-86 EAMKQKDTIGDINGILPNTD

1 60

Oct-1 RRKKRTSIETNIRVALEKSFLENQKPTSEEITMIADQLNMEKEVIRVWFNRRQKEKRIN
Oct-2V.F.....A.....LL.E.H.....
Oct-6 K.....VGVGK...SH..KCP..SAH...GL..S..QL.....V.....MT
Oct-4 .KR.....NRV.WS..TM..KCP..SLQQ..H..N..GL..D.V.....G..SS
Pit-1 K..R..T.SIAAKD...RH.G.HS..S.Q..MRM.EE..L...V.....R..VK
Unc-86 KKR.....AAPEKRE..QF.KQQPR.SG.R.AS...R.DLK.N.V.....Q...Q..DF

$\alpha 1$ $\alpha 2$ $\alpha 3$

108

recognition helix of the homeodomain and contact the major groove of the DNA. The Q154 contact is found only in the POU family but it has not been mutated to determine its specific effect on DNA binding.

Another interesting fact about the POU proteins is the conserved residue Cysteine 50 in the HD. In the structures of Oct-1 POU proteins, the residue at position 50 does not interact with DNA, however in the Pit-1 POU structures the C50 has a van der Waals contact with a Thy that lies in the flanking sequence 5' to the octamer (8). Stepchenko mutated the cysteine at position 50 to all other amino acids and tested their binding to the natural octamer ATGCAAAT and the homeodomain binding site TAAT to determine the role of this conserved amino acid in the POU protein family (9). For those proteins that bound to the TAAT site the strength of the binding was determined in part by the 3' flanking sequences (TAATNN) which we found to be true in the case of the Msx-1 HD binding to its cognate DNA binding site. Based on their mutational data it seems as if no residue at position 50 except cysteine can give the POU proteins the capability of highly selective recognition of their specific targets and binding to them independently of the 3' – flanking nucleotides. The role of Cys 50 is still unclear.

The N-terminal arm of the homeodomain for the Oct-1/SNAP190/U1 octamer is very similar to the other Oct-1 structures that have been solved. The N-terminal arm is well ordered out to residue R101 which is one residue longer than the usual start of the homeodomain (residue 102) and while there appears to be density for part of the linker, it is not well ordered enough to trace. The R102 makes contacts to the minor groove of the DNA and helps to anchor the N-terminal arm. The average B-factor for the N-terminal arm is anywhere from 3 to 10 Å² higher than the average B-factor for the rest of the

protein. The linker connecting R101 to S82 is a disordered flexible region whose function is not yet clear.

The differences in the linker region are the major differences in the POU protein family. The residues that contact the DNA in the Oct-1 protein and the Pit-1 protein are exactly the same but Pit-1 proteins bind to widely different sequences. Pit-1 does not bind a highly conserved consensus sequence like the Oct-1 and Oct-2 proteins (10). Table 1 summarizes the similar DNA base specific contacts among the POU domain structures solved. The table indicates the specific base involved in the contacts and includes both molecules if there is more than one in the asymmetric unit of the structure. The Pit-1 POU protein is in the Pit-1, Prl-1P, and the GH-1 structures. The others are structures of the Oct-1 POU bound to different DNA sequences. The POU proteins were overlaid either with the POU specific or the POU HD depending upon the part of the structure being studied. The base specific contacts seem unaltered so the specificity for a certain POU protein for a certain DNA sequence must lie in the arrangement of the domains on the DNA and perhaps whether it forms a monomer or a dimer.

The water network that was observed in the monomer HDs was compared with the POU proteins. Table 2 summarizes the number of conserved waters (from the Msx-1 structure) that are found in the POU structures. The SNAP190 structure has the lowest number of conserved waters (8/32) for both molecules. The real anomaly is the one water found in the Oct-1/MORE structure which has the unusual POU_s and POU_{HD} arrangement. The numbers for the Prl-1P and GH-1 structures are not included because pdb information was not available. The Oct-1/H2B and Oct-1/H2B/OCA-B

Table 4.1 Summary of base specific contacts among POU proteins.

	SNAP 190	OCA-B	Oct-1	MORE	PORE	Pit-1	Prl-1P	GH-1
S43						T5':T5'		
Q44	A1:A1	A1:A1	A1	A1	A1',A2': T1	A3:A3	A3:A3	A4:A4
T45	T2,C3': T2,C3'	T2,C3': G4',C3'	T2, C3'	T2,C3'	G2,C3': T2,A3'	T4,T5': T4,T5'	T4,A5': T4,T5'	T5,A6': T5,C6'
R49	G3,G4': G3	G3,G4': G4',T5'	G3, G4'	G3,G8'	G4':G4'	G6': G6'	G6': A6'	G7': A7'
R102			T6'		T6':T6'			
R105	A5,A4': A5,A4	A5,G4': G4'	A5, G4'		A5:A5	A5,T4': A5,A4'	T5':T4	
V147	A7:A7	T8	T8			T8:T8	T8:T8	T9:T9
C150						T10': T10'	T11': T10'	T10': T10'
N151	A7	A7	A7	A5	A7:A7	A7:A7	A7:A7	A8:A8
Q154	A8',T9': A8',T9'	A8'.T9': A8',T9'		A6'	A8':A8'	A8': A8'	A8': A8'	A9': A9'

structures do not have waters due to their resolutions (3.0 & 3.2 Å). The main reason the SNAP190 structure does not have as many waters as say the Pit-1 structure is unclear. The SNAP190 ternary complex was crystallized in iso-propanol and the other were not, so maybe this has had a small effect. Even the Pit-1 and PORE structures have about half of the “conserved” waters seen in Msx-1. The POU proteins have two domains bound to the DNA while most of the structures are dimers, so the need for water mediated interactions between the HD and the DNA are perhaps not as necessary. The majority of the absences in the table are due to DNA base changes and/or residue differences. The ones that we see being conserved among the monomer HDs and the POU protein HDs are the ones independent of a specific sequence (such as phosphate backbone and protein backbone interactions). The number of conserved waters can not be correlated to resolution or total number of waters. The average B-factors for the “conserved” waters compared to all of the waters in that particular structure is also included in the table.

Table 4.2 Conserved water comparison.

HD/DNA Complex	Resolution Limit (Å)	# Conserved Waters	Data Collection Condition	Total # Waters	B-factor Ratio (Å²)	Ref/PDB ID
Msx-1	2.2	16	frozen	153	30/43.6	(11), 1IG7
SNAP190	2.4	A-5, B-3	frozen	244	17.6/24.6	-
Pit-1	2.3	A-9, B-8	frozen	176	34.9/41.1	(8),1AU7
MORE	1.9	1	frozen	138	-	(12), 1E3O
PORE	2.7	A-9, B-7	frozen	119	41.6/45.9	(12), 1HFO

4.5 The OCA-B co-activator and SNAP190 general factor target Oct-1 similarly.

As shown in Figure 4.10, the Oct-1 interacting regions of SNAP190 and OCA-B exhibit significant homology. Many of the conserved residues in the C-terminal helix region of SNAP190 and OCA-B make important interactions with the POU_s domain. Virtually all of the hydrophobic contacts and both main-chain hydrogen bonds are conserved between the two structures. In spite of this conservation, there are also

differences in the two interfaces. The most notable difference in the two structures is the extension of the C-terminal helix by one turn in our SNAP190 structure. None of the residues in this C-terminal helix region are conserved between SNAP190 and OCA-B and these residues were disordered in the Oct-1/OCA-B/H2B octamer structure. This extra turn of the SNAP190 helix contains E904 which makes a buttressing salt bridge to K900, stabilizing its orientation for interaction with the Oct-1 POU₅ E7 side chain. This electrostatic interaction is critical for POU domain transcription specificity (4-6). In our structure these side chains make a tight interaction (3.0 Å) as compared to comparable positions in the OCA-B structure wherein the interactions are more distant (5 Å) (see also Figure 4.8).

The two structures are even more significantly divergent in their N-terminal regions. The OCA-B peptide tracks across the minor groove of the DNA, making several hydrophobic and main-chain hydrogen bonding interactions with the T/A basepair at position 5 in the H2B octamer sequence. These interactions are predominantly mediated by V22 and V24 of OCA-B. In fact, OCA-B confers additional DNA binding specificity to the complex because it will only bind to Oct-1/DNA complexes that have A/T base pairs at positions 5 and 6 (13). Since the U1 octamer lacks the A/T base pair at position 6, it is not expected to have affinity for OCA-B. In contrast, those amino acids within OCA-B that make base specific contacts are not conserved within SNAP190. Instead, SNAP190 tracks the phosphate backbone on only one side of the DNA making only two direct contacts with the DNA phosphate backbone (R887 and P890). In contrast to OCA-B, SNAP190 would likely not confer DNA specificity to the complex since its DNA interactions are with the phosphate backbone. Importantly, this critical base at position 6

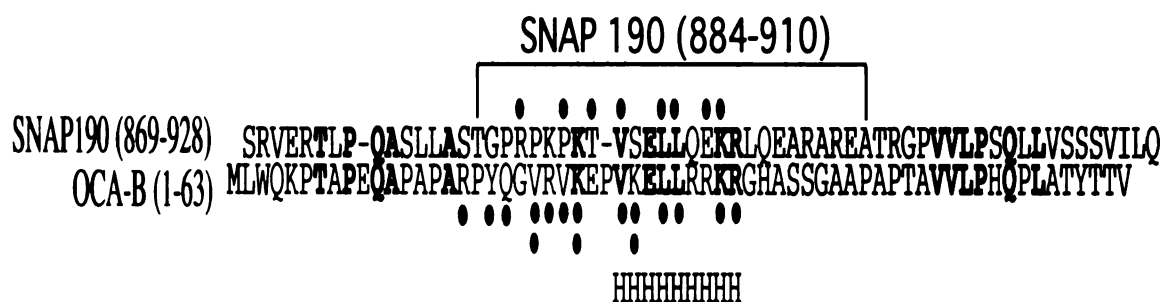


Figure 4.10 Sequence alignment of homologous regions of SNAP190 and OCA-B. The region surrounding the SNAP190 peptide used in the crystallization is indicated. The homology between the two sequences is demoted with bold text. H indicates the helical region that is common to both structures. The green circles above (SNAP190) and below the (OCA-B) sequence alignment denote contacts made to DNA and the red circles donate contacts made to the Oct-1 POU domain.

in the U1 octamer reduces the affinity of OCA-B for DNA binding without compromising the ability of SNAP190 to assist in Oct-1 activator binding to the U1 octamer element.

The far N-terminus of the OCA-B peptide also makes several contacts with the POU_{HD} domain, resulting in the DNA being surrounded by the Oct-1 POU/OCA-B peptide complex. No similar interactions are observed between SNAP190 and POU_{HD} in the SNAP190/Oct-1 POU/U1 octamer complex structure. These differences are also not surprising given that none of the residues in OCA-B that interact with the POU_{HD} domain are conserved in the SNAP190 sequence (Figure 4.10). An interesting consequence of the interaction between OCA-B and the POU_{HD} can be seen in Figures 4.11 and 4.12. Here we have aligned the POU_S domains from the Oct-1/H2B octamer, Oct-1/H2B octamer/OCA-B and the Oct-1/U1 octamer/SNAP190 structures (Figure 4.11). While the Oct-1/H2B octamer and Oct-1/U1 octamer/SNAP190 structures are very similar, with both the POU_S, octamer octamer and POU_{HD} well aligned, there is a significant change in the relative position of POU_{HD} in the Oct-1/H2B octamer/OCA-B structure. In fact the second helix of the POU_{HD} has moved by more than 3.5 Å relative to the positions of either our structure or the Oct-1/H2B octamer structure (Figure 4.12). The motion combines a translation and rotation centered in the middle of the POU_{HD} DNA recognition helix. Both motions serve to pull the POU_{HD} recognition helix closer to the OCA-B peptide, allowing the interactions between OCA-B and helix 3 of the POU_{HD} to occur. This motion occurs without significant compensatory movement of the half-site DNA bound by POU_{HD}. Nevertheless, most of the interactions between the POU_{HD} and

DNA are preserved. In contrast, SNAP190 does not interact with POU_{HD} and the relative orientations of POU_S, DNA and POU_{HD} are very similar to that seen in the original Oct-1/H2B octamer binary complex, in spite of the change in octamer DNA sequence. This indicates that the motion of POU_{HD} is due to its direct interaction with OCA-B and not with other more indirect interactions. These interactions alter the trajectory of flanking DNA on the POU_{HD} side of the octamer. Here all three complexes exhibit significant structural differences, sometimes resulting in altered protein/DNA interaction (Figure 4.6). In the case of the OCA-B ternary complex, the flanking DNA on the POU_{HD} side has been pulled toward the domain as it has moved toward OCA-B. The consequences of flanking sequence movement for transcriptional activation are not known, but may influence the relative orientation of downstream components of the pre-initiation complex at these promoters.



Figure 4.11 Overlay of the Oct-1/U1 octamer/SNAP190 (gold), Oct-1/H2B octamer (red), and the Oct-1/H2B octamer/OCA-B (dark blue) complex structures. The peptides have been colored independently with the SNAP190 peptide in green and the OCA-B in dark gray. Additional contacts between OCA-B and the Oct-1 POU_{HD} that are not observed in the SNAP190 structure rotate the POU_{HD} DNA recognition helix relative to its position in the other two structures (enlarged in next figure 4.12).

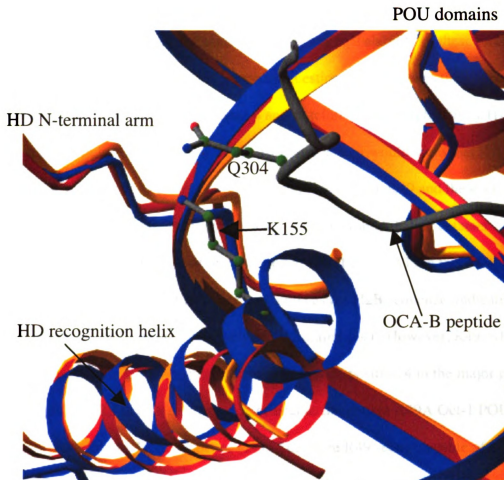


Figure 4.12 Enlargement of the recognition helix region in which the OCA-B (blue) helix shifts by more than 3.5\AA , interacting with the N terminal of the OCA-B peptide (gray). The two residues shown are one example of an interaction between the OCA-B HD and the OCA-B peptide.

4.6 Cooperative promoter recognition and activation of human U1 transcription

Promoter recognition by activator proteins is key to the ability of these factors to modulate transcriptional activity. One important question is what makes the U1 octamer, which contains two base changes (positions 4 and 6) relative to the high affinity H2B octamer, a poor binding site for Oct-1 POU. At first glance it would seem that R49 from the Oct-1 POU_s could play a critical role, as the contact between R49 and the 4' G (now a 4' A in the U1 octamer) is lost. However, Cleary and Herr found that an octamer sequence with both the 3 and 4 positions changed to T (ATTTAAAT) in the H2B octamer sequence has similar affinity for Oct-1 POU as the H2B sequence, indicating that changing position 4 to T does not adversely affect binding (14). However, R49, which is the only residue that makes base-specific interactions with position 4 in the major groove of the Oct-1 POU/H2B octamer structure, is critical as the mutant R49A Oct-1 POU has very little affinity for any DNA sequence (14). Therefore R49 seems capable of making critical but flexible interactions with DNA in this complex. We see evidence for this flexibility in our structure. Although R49 has lost its interaction to position 4, it compensates by moving significantly to make a much tighter hydrogen bond with the O6 of G at position 3 than it does in the Oct-1/H2B octamer complex structure (2.8 Å in our structure versus 3.5 Å in the Oct-1 POU/H2B octamer structure). When T's replaces both positions 3 and 4 as in the ATTTAAAT sequence, R49 could potentially make a tight hydrogen bond with the O4 carbonyl of T at position 3. On the other hand, we predict that modification of position 3 to C and position 4 to T (ATCTAAAT) would remove both possibilities for R49 hydrogen bonding and result in a sequence that has

much reduced affinity for Oct-1 POU. Similar flexibility is seen in the Pit-1 structures. In all the Oct-1/DNA complexes solved to date, R49 makes base specific hydrogen bonds with the base at position 3 or 4', depending on the sequence, which suggests that its possible interactions and motions are restricted to these bases (14). Figure 4.13 shows the interaction of R49 with the DNA bases. The U1 octamer also contains a base change at position 6 from A to G, which also causes changes to the Oct-1 POU/DNA interface. In both the Oct – 1 POU/OCA-B and H2B octamer structures, Oct-1 R102 makes contacts with both the sugar and bases of the 5', 6', 7 and 8 positions. In our Oct-1/U1 octamer/SNAP190 structure the trajectory of R102 is altered due to steric collision with the amine of G6, preventing these interactions from occurring (Figure 4.14). Instead, R102 points away from the DNA. Given the above arguments, the base change in the 6 position may have the more devastating effect on Oct-1 POU/DNA binding, although both base changes likely contribute to the reduced affinity for the U1 octamer.

Previous reports have shown that inhibition of DNA binding plays a role in the transcriptional specificity for snRNA genes. It has already been established that full length human TBP binds DNA alone very poorly and that the inhibition to TBP DNA binding can be relieved by the SNAP_C complex on U6 promoters (4, 15). It appears that the inhibition is mediated by the N-terminus of TBP and that SNAP_C interaction probably causes a conformational change between the N- and C- termini of TBP. Further the DNA binding affinity of SNAP190 is also inhibited, this time by the C-terminus of the protein. This inhibition is relieved by interactions between the C-terminus of SNAP190 and the

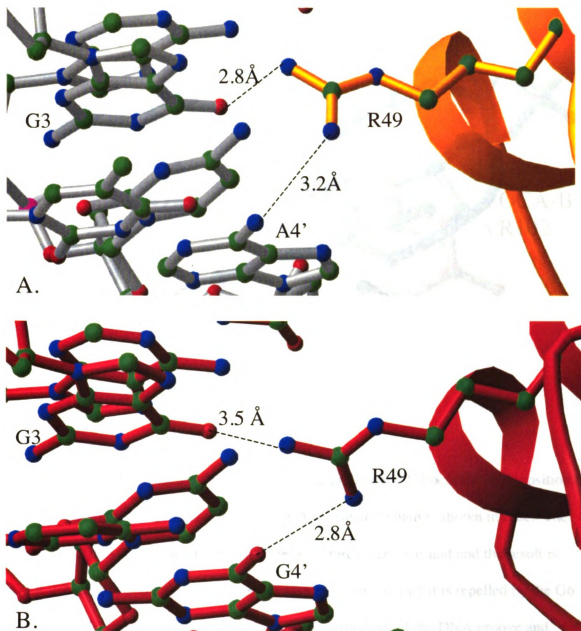


Figure 4.13 The arginine 49 interaction to the different base pair at position 4. A. The POU domain is in gold while the DNA is in silver. The R49 moves down to make a closer contact to the G3 oxygen while it has a longer contact to A4'. This is the direct opposite of what is seen in the Oct-1/H2B (panel B) and Oct-1/H2B/OCA-B structures.

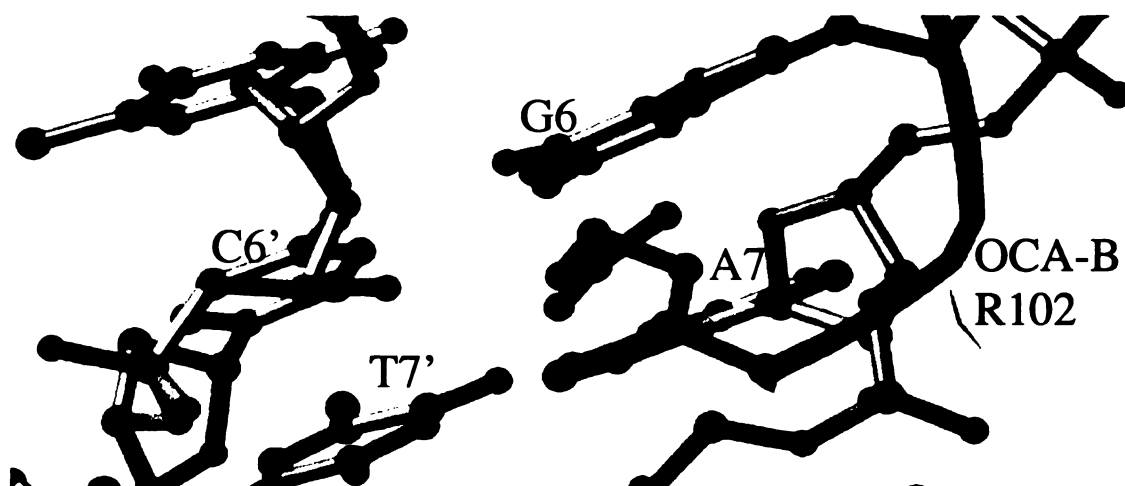


Figure 4.14 R102 collision in the case of OCA-B protein and U1 octamer with position 6 altered. The U1 DNA is in silver and the OCA-B homeodomain is shown in blue. The Oct-1/H2B/OCA-B and Oct-1/U1/SNAP190 structures were overlaid and the result is that R102 can not make its normal contact with base A6 – in fact it is repelled by the G6 NH_2 group. In fact the whole R102 side chain is pushed out of the DNA groove and interacts with the protein and the DNA backbone.

Oct-1/DNA complex. In a further twist to this theme, we show here that Oct-1 DNA binding is inhibited at the U1 octamer and that this inhibition can be relieved by interactions between Oct-1 and SNAP190. DNA binding cooperativity thus becomes a specificity tool, preventing these proteins from interacting at this or other promoters or DNA sequences in the absence of their functional partners. This is one way of preventing genome-wide squelching. Given that many SNAP-dependent genes are highly expressed, it may be important to keep SNAP_C from binding at functionless DNA sites. It now appears that at least at the U1 promoter, all proteins that interact directly with DNA are inhibited for this binding in the absence of at least one of the other proteins or protein complexes that interact with DNA. Binding of the full complex requires a series of synergistic interactions between all three factors.

4.7 Conclusions

This structure highlights subtle differences between the Oct-1 POU structures that have been solved. Essentially the mode of recognition remains the same with virtually identical protein/DNA contacts in all of the structures. Differences come into play in the flanking DNA regions where there is more play in the structures. The main differences are due to the actual identity of the DNA bases when you compare the H2B sequence with the U1 sequence. With the change of two base pairs you get a drastic reduction in the binding of Oct-1 to the DNA. This is most likely due to the loss of direct base contacts to the new bases. However since U1 is so widely expressed there is no reason to think that Oct-1 would not bind to U1 *in vivo*. Clearly the stepwise mechanism for the activation of transcription can not explain the Oct-1/U1 octamer/SNAP190 binding

assays. Oct-1 binds weakly to the U1 octamer site despite the fact the U1 genes are transcribed at a high rate. So it appears all three factors need to be present for decent binding to the U1 octamer site and eventually transcription of the U1 genes.

Obviously since OCA-B and SNAP190 bind in the same place they would compete with each other in the cell. There is no data on the binding of the OCA-B peptide to the U1 octamer, almost all experiments are done with the H2B octamer due to the high affinity binding of this site to Oct-1 and other Oct proteins (Oct-2 for example). So we can only suggest the base pair change at position 6 would disrupt the critical contact between OCA-B and DNA and thus cause the binding constant to decrease significantly.

4.8 References

1. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) *Cell*. 77, 21-32.
2. Chasman, D., Cepek, K., Sharp, P. A., and Pabo, C. O. (1999) *Genes Dev.* 13, 2650-7.
3. Carson, M. (1991) *J. Appl. Crystallogr.* 24, 958-961.
4. Ford, E., Strubin, M., and Hernandez, N. (1998) *Genes Dev.* 12, 3528-40.
5. Mittal, V., Cleary, M. A., Herr, W., and Hernandez, N. (1996) *Mol. Cell. Biol.* 16, 1955-65.
6. Murphy, S. (1997) *Nucleic Acids Res.* 25, 2068-76.
7. Evans, S. V. (1993) *J Mol. Graph* 11, 134-8, 127-8.
8. Jacobson, E. M., Li, P., Leon-del-Rio, A., Rosenfeld, M. G., and Aggarwal, A. K. (1997) *Genes Dev.* 11, 198-212.
9. Stepchenko, A. G., Luchina, N. N., and Pankratova, E. V. (1997) *Nucleic Acids Res.* 25, 2847-2853.
10. Herr, W., and Cleary, M. A. (1995) *Genes Dev.* 9, 1679-1693.
11. Hovde, S., Abate-Shen, C., and Geiger, J. H. (2001) *Biochemistry* 40, 12013-21.
12. Remenyi, A., Tomilin, A., Pohl, E., Lins, K., Philippsen, A., Reinbold, R., Scholer, H. R., and Wilmanns, M. (2001) *Mol. Cell* 8, 569-80.
13. Babb, R., Cleary, M. A., and Herr, W. (1997) *Mol. Cell. Biol.* 17, 7295-305.
14. Cleary, M. A., and Herr, W. (1995) *Mol. Cell. Biol.* 15, 2090-100.
15. Mittal, V., Ma, B., and Hernandez, N. (1999) *Genes Dev.* 13, 1807-21.

APPENDIX

Appendix 2.1 Protein purification Buffers.

Msx- 1 Buffers

Buffer A:

6M Guanidine HCl

25mM Sodium Phosphate pH 8.0

Buffer B:

6M Guanidine HCl

25mM Sodium Phosphate pH 6.0

Buffer C:

6M Guanidine HCl

25mM Sodium Phosphate pH 5.0

Buffer D:

1.0 M Guanidine HCl

25mM Sodium Phosphate pH 7.4

10% glycerol

50 mM KCl

5 mM MgCl₂

10 mM DTT

Buffer E:

0.1 M Guanidine HCl

25mM Sodium Phosphate pH 7.4

10% glycerol

50 mM KCl

5 mM MgCl₂

5 mM DTT

Buffer F:

25mM Sodium Phosphate pH 7.4

10% glycerol

50 mM KCl

5 mM MgCl₂

1 mM DTT

Buffer G:

5mM Tris

10% Glycerol

50mM KCl

5mM β-mercaptoethanol.

Oct-1 Buffers:

HEMG1 250:

25mM Hepes pH 7.9

2 mM EDTA

12.5 mM MgCl₂

10% Glycerol

0.1% Tween-20

250 mM KCl

3mM DTT

HEMG1 100:

25mM Hepes pH 7.9

2 mM EDTA

12.5 mM MgCl₂

10% Glycerol

0.1% Tween-20

100 mM KCl

3mM DTT

TDB (Thrombin Digestion Buffer) (10x):

200mM Tris HCl pH 8.4

1.5 M NaCl

25 mM CaCl₂

TDB – DTT (10x):

200mM Tris HCl pH 8.4

1.5 M NaCl

25 mM CaCl₂

3 mM DTT

Appendix 3.1 Msx-1 – DNA contacts compared to other monomer HD structures.

The structures were overlaid using the alpha carbon atoms in the three alpha helices. The r.m.s.d. was calculated with the alpha carbons only. Abbreviations are as follows: paired (prd), engrailed at 2.2Å(neweng), engrailed mutant (engmut), even-skipped (eve), Antennapedia X-ray (antx), MATa1 (a1), MATalpha2 (a2).

Three structures are heterodimers (a1,a2), (pbx, hoxb1), and (ubx,exd). Each homeodomain in the pair was overlaid on Msx-1 to compare the water structure. In the case of antx, the two HDs are the same so molecule A waters were inspected first and any waters not found were looked for in molecule B. Molecule B waters are indicated by a (*). The same goes for eve, but in this case the molecules in the pdb are labeled C and D. In the case of the heterodimer (a1, a2), as there are so few waters we inspected two structures of the heterodimeric complex. They were bound to somewhat different DNA sequences. The second structure was published in 1998. In each case the two domains were aligned separately with msx-1. The (*) entries in the table indicate these waters came from the later structure. The interactions are listed in Table 2. When the structures were overlaid we renumbered all of the other structures to match the msx-1 numbering scheme for an easier direct comparison. All protein, DNA, and waters have been renumbered. The interactions noted are within 3.5Å of each other. The water interactions listed are specific (numbered) for other conserved waters in the table and general for non-conserved waters. A letter next to the water number denotes the water's appearance in a figure in the paper.

Table 1. Part one of direct water comparisons.

Msx-1 water #	prd	neweng	engmut	eve	pbx	hoxb1
(1IG7)	(1FJL)	(3HDD)	(2HDD)	(1JGG)	(1B72)	(1B72)
161	0.51	0.67	0.27	0.97*	X	0.5
162 (D)	0.14	0.1	0.29	0.32	0.67	0.39
163 (E)	X	0.76	0.61	0.59	1.02	0.69
168 (A)	1	0.39	1.33	0.78*	0.14	1.18
169 (H)	1.43	0.68	0.29	X	1.53	0.72
170	X	X	X	0.53	X	X
175	1.1	X	1.58	X	1.02	1.56
178	1.53	X	1.12	0.61	1.51	X
186 (C)	0.58	1.2	X	1.33	0.88	0.26
187	0.23	0.46	0.27	0.34	0.69	0.48
189	X	0.53	0.26	X	1.43	X
202	0.16	0.4	0.26	0.46	0.63	X
203	0.56	0.54	0.46	0.37	1.53	0.32
219 (J)	0.77	0.31	0.55	0.36	X	X
244 (G)	0.92	0.44	0.41	X	1.47	X
330 (I)	X	0.2	0.55	0.66	1.07	0.84
Avg	0.74	0.51	0.6	0.61	1.05	0.69
r.m.s.d.	0.67	0.59	0.64	0.65,0.59	0.75	0.59

Table 1 Cont. Part two of the direct water comparisons.

Msx-1 water #	ubx	exd	a1	a2	antx	avg
(1IG7)	(1B8I)	(1B8I)	(1YRN)	(1AKH)	(9ANT)	
161	1.04	X	X	X	1.19*	0.74
162 (D)	0.81	0.88	X	0.5	0.65	0.48
163 (E)	0.53	X	0.55	0.73*	0.42*	0.65
168 (A)	0.94	0.77	1.45	X	X	0.89
169 (H)	1.45	0.78	1.39	X	X	1.03
170	0.97	1.48	X	1.14	1.24	1.07
175	0.42	0.49	1.42	1.58	X	1.15
178	0.94	X	0.81	1.49	0.58	1.07
186 (C)	0.91	1.21	0.84	X	0.13	0.82
187	0.84	0.52	0.63*	0.21	0.39	0.46
189	0.7	1.18	X	0.68*	X	0.79
202	0.67	1.05	X	0.81	0.44*	0.54
203	1.17	X	X	0.95	X	0.74
219 (J)	0.7	0.52	X	0.6	X	0.54
244 (G)	1.02	1.01	X	X	X	0.88
330 (I)	X	0.77	X	X	X	0.68
Avg	0.87	0.89	1.01	0.87	0.63	
r.m.s.d.	0.55	0.86	0.66,0.62	0.69,0.68	0.57,0.59	

Table 2. Protein – DNA contact comparison among HDs.

Water #	Msx-1	prd	New Engrailed	Engrailed Mutant	Even Skipped	Ptx
161	Ade 8	Ade 8	Ade 8	Ade 8	Ade 8, Ade 9 M54	X
162 (D)	Ade 8 Phosphate N51 Wat 187	Ade 8 Phosphate N51 Wat 187	Ade 8 Phosphate N51 Wat 187	Ade 8 Phosphate N51 Wat 187	Ade 8 Phosphate N51 Wat 187	Ade 8 Phosphate N51 Wat 187
163 (E)	Thy 21 Phosphate R53, K46 backbone Wat 178	X	Gua 21 Phosphate R53, K46 backbone	Gua 21 Phosphate R53, K46 backbone	R53, K46 backbone Wat 178	Thy 21 Phosphate R53, S46 backbone Wat 178
168 (A)	Thy 11 Q50 Wat	Cyt 11 water	Thy 11 Q50	Cyt 11	Thy 11, Cyt 22	Thy 11, Cyt 22 water
169 (H)	Ade 24 N51 Wat 330	Ade 24 N51 Wat 244	Ade 24 N51 Wat 244,330	Ade 24 N51 Wat 244,330	X	Ade 24 N51 Wat 244
170	Ade 9 Phosphate T43 water	X	X	X	Ade 9 Phosphate	X
175	Thy 10, Thy 25 R2	Thy 10, Thy 25 R2	X	Thy 10, Thy 25 water	X	Thy 10, Thy 25 water
178	Thy 21 Phosphate R53, L26 backbone wat 163	Cyt 21 Phosphate Q46 water	X	Gua 21 Phosphate R53, L26 backbone	R53, V26 backbone Wat 163	Thy 21 Phosphate R53 Wat 163, water
186 (C)	Ade 9, Thy 10 Q50, N51	Thy 10 Q50, N51 water	Thy 10 Q50, N51	X	Thy 10 Q50, N51	Thy 10 N51
187	Ade 9 Phosphate Q44 backbone Wat 162	Ade 9 Phosphate R44 backbone Wat 162	Ade 9 Phosphate Q44 backbone Wat 162	Ade 9 Phosphate Q44 backbone Wat 162	Ade 9 Phosphate T44 backbone Wat 162	Ade 9 Phosphate Q44 backbone Wat 162
189	Cyt22 Phosphate water	X	Thy 22 Phosphate water	Gua 22 Phosphate water	X	Cyt 22 Phosphate Y25, K57
202	Ade 8 Phosphate W48 backbone water	Ade 8 Phosphate W48 backbone water	Ade 8 Phosphate W48 backbone	Ade 8 Phosphate W48 backbone water	Ade 8 Phosphate W48, R52 backbone water	Gua 8 Phosphate W48 backbone Wat 203
203	Ade 8 Phosphate W48	Ade 8 Phosphate W48 water	Ade 8 Phosphate W48, L13	Ade 8 Phosphate W48 water	Ade 8 Phosphate W48, L13 water	W48 Wat 202, water
219 (J)	Cyt 22 Phosphate Q50 backbone Wat 330, water	Thy 22 Phosphate R57, Q50 backbone	Thy 22 Phosphate Q50 backbone Wat 330	Gua 22 Phosphate K50 backbone Wat 330	Q50 backbone Wat 330	X
244 (G)	Ade 24	Ade 24 Wat 169, water	Ade 24 Ala 54 Wat 169	Ade 24 Ala 54 Wat 169	X	Ade 24 Wat 169
330 (I)	Ade 23 Q50 Backbone Wat 169	X	Ade 23 Q50 Backbone Wat 169, 219	Gua 23 K50 Wat 169, 219	Ade 23 M54 Wat 219	Ade 23 G50 Backbone water

Table 2. Cont. Protein – DNA contact comparisons continued.

Water #	Max-1	Hox B1	Ubx	Exd	MATa1	MATa2	Antennapedia (x-ray)
161	Ade 8	Gua 8	Thy 8	X	X	X	Ade 8 water
162 (D)	Ade 8 Phosphate N51 Wat 187	Gua 8 Phosphate N51 Wat 187	N51 Wat 187	Ade 8 Phosphate N51 Wat 187	Ade 8 Phosphate N51	Ade 8 Phosphate N51 Wat 187	Ade 8 Phosphate N51 Wat 187
163 (E)	Thy 21 Phosphate R53, K46 backbone Wat 178	Cyt 21 Phosphate R53, K46 backbone	R53, S46 backbone Wat 178	X	Thy 21 Phosphate R53, R46 backbone	Cyt 21 Phosphate R53, K46 backbone Wat 178	Gua 21 Phosphate R53, K46 backbone Wat 178
168 (A)	Thy 11 Q50 Wat	Cyt 22 Wat	Gua 11, Gua 12, Cyt 22	Thy 11 Wat	Thy 12, Gua 11 water	X	X
169 (H)	Ade 24 N51 Wat 330	Ade 24 N51 Wat 330	Ade 24 N51 Wat 244	Ade 24 N51 Wat 244,330	Ade 24 N51	X	X
170	Ade 9 Phosphate T43 water	X	Thy 10 Phosphate water	Thy 10 Phosphate N47 water	X	Ade 9 Phosphate N47	Thy 10 Phosphate I47, R43
175	Thy 10, Thy 25 R2	Thy 25	Thy 10, Thy 25 water	Thy 10, Thy 25 waters	Thy 10	Thy 25 water	X
178	Thy 21 Phosphate R53, L26 backbone wat 163	X	R53, L26 backbone Wat 163	X	Thy 21 Phosphate R53, L26 backbone	Cyt 21 Phosphate R53, L26 backbone	Gua 21 Phosphate R53, L26 backbone
186 (C)	Ade 9, Thy 10 Q50, N51	Thy 10 N51	Thy 10, Cyt 23 Q50, N51	Thy 10	Thy 10, Cyt 23 I50	X	Thy 10, Cyt 23 Q50, N51
187	Ade 9 Phosphate Q44 backbone Wat 162	Ade 9 Phosphate Q44 backbone Wat 162	Y8, Q44 backbone Wat 162, water	Ade 9 Phosphate Q44 backbone Wat 162	Ade 9 Phosphate Q44 and W48 backbones	Ade 9 Phosphate Q44 backbone Wat 162	Y8, Q44 backbone Wat 162
189	Cyt22 Phosphate water	X	Cyt 22 Phosphate M54 water	Ade 22 Phosphate I54 water	X	Ade 22 Phosphate K57 water	X
202	Ade 8 Phosphate W48 backbone water	X	W48 Wat 203, waters	W48, K52 backbone water	X	W48 water	Ade 8 Phosphate W48, W48 backbone
203	Ade 8 Phosphate W48	Gua 8 Phosphate W48, L13 water	Wat 202, water	X	X	Thy 8 Phosphate W48	X
219 (J)	Cyt 22 Phosphate Q50 backbone Wat 330, water	X	Cyt 22 Phosphate M54, Q50 backbone	Ade 22 Phosphate I54, G50 backbone Wat 330	X	Ade 22 Phosphate S50 backbone	X
244 (G)	Ade 24	X	Ade 24 Wat 169	Ade 24 Wat 169	X	X	X
330 (I)	Ade 23 Q50 Backbone Wat 169	Gua 23 Q50 Wat 169	X	Ade 23 G50 Backbone Wat 169, 219	X	X	X

Appendix 4.1 Protein – Protein contacts found between the SNAP190 peptide and the Oct – 1 POU protein. The interactions listed are within a 4.0 Å cutoff.

SNAP190			Oct-1 POU			distance (Å)
#	name	atom	#	name	atom	
890	A PRO	CB	56	A SER	CB	3.88
891	A LYS	CG	56	A SER	CA	3.86
891	A LYS	CG	56	A SER	CB	3.98
891	A LYS	CG	57	A PHE	N	3.73
891	A LYS	CD	56	A SER	CB	3.04
891	A LYS	CD	56	A SER	C	3.03
891	A LYS	CD	57	A PHE	CA	3.36
891	A LYS	CD	57	A PHE	CB	3.37
891	A LYS	CE	56	A SER	CA	3.78
891	A LYS	CE	56	A SER	CB	3.41
891	A LYS	CE	56	A SER	C	3.70
891	A LYS	CE	57	A PHE	CA	3.42
891	A LYS	CE	57	A PHE	CB	3.07
891	A LYS	CE	58	A LYS	N	3.85
891	A LYS	NZ	56	A SER	CA	3.84
891	A LYS	NZ	56	A SER	CB	3.44
891	A LYS	NZ	56	A SER	OG	3.70
891	A LYS	NZ	56	A SER	C	3.31
891	A LYS	NZ	57	A PHE	CG	3.69
891	A LYS	NZ	57	A PHE	C	2.93
891	A LYS	NZ	58	A LYS	CA	3.67
891	A LYS	NZ	58	A LYS	CB	3.84
891	A LYS	NZ	58	A LYS	CG	3.78
892	A THR	N	55	A LEU	O	3.95
892	A THR	CA	53	A LEU	O	3.93
892	A THR	CA	55	A LEU	C	3.90
892	A THR	CA	55	A LEU	O	3.20
892	A THR	CB	53	A LEU	O	3.52
892	A THR	C	53	A LEU	O	3.87
892	A THR	C	55	A LEU	O	3.75
893	A VAL	N	53	A LEU	O	3.66
893	A VAL	N	55	A LEU	O	3.22
893	A VAL	CB	60	A MET	SD	3.65
893	A VAL	CG2	55	A LEU	O	3.81
893	A VAL	CG2	60	A MET	CB	3.99
893	A VAL	CG2	60	A MET	CG	3.40
893	A VAL	CG2	60	A MET	SD	3.09
894	A SER	N	53	A LEU	CB	3.82

894	A	SER	N	53	A	LEU	O	3.04
894	A	SER	CA	53	A	LEU	O	3.68
894	A	SER	CB	53	A	LEU	CB	3.96
894	A	SER	CB	53	A	LEU	O	3.18
896	A	LEU	CB	6	A	LEU	CD1	3.97
896	A	LEU	CB	6	A	LEU	CD2	3.78
896	A	LEU	CD1	6	A	LEU	CD2	3.97
896	A	LEU	O	6	A	LEU	CD1	3.87
897	A	LEU	CD1	10	A	GLU	CB	3.80
897	A	LEU	CD1	10	A	GLU	CD	3.47
897	A	LEU	CD1	10	A	GLU	OE1	3.17
897	A	LEU	CD1	10	A	GLU	OE2	3.88
897	A	LEU	CD1	53	A	LEU	CD1	3.72
897	A	LEU	CD2	6	A	LEU	CB	3.56
897	A	LEU	CD2	10	A	GLU	OE1	3.78
900	A	LYS	CE	7	A	GLU	CD	3.49
900	A	LYS	CE	7	A	GLU	OE1	3.22
900	A	LYS	CE	7	A	GLU	OE2	3.85
900	A	LYS	NZ	7	A	GLU	CG	3.76
900	A	LYS	NZ	7	A	GLU	CD	3.04
900	A	LYS	NZ	7	A	GLU	OE1	3.27
900	A	LYS	NZ	7	A	GLU	OE2	2.96

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02331 8193