

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Structural Discrimination between Activating and Repressing Response Elements in the p53 Protein Family

Permalink

<https://escholarship.org/uc/item/1jc2k8z1>

Author

Ramos, Ana

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Structural Discrimination between Activating and Repressing Response Elements in the
p53 Protein Family

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

in

Chemistry

by

Ana Ramos

Committee in charge:

Professor Héctor Viadiu, Chair
Professor Russell Doolittle, Co-Chair
Professor Alexander Hoffmann
Professor J. Andrew McCammon
Professor Stanley Opella
Professor Gentry Patrick

2015

Copyright

Ana Ramos, 2015

All rights reserved

The Dissertation of Ana Ramos is approved, and it is acceptable in quality
and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2015

DEDICATION

The dissertation is dedicated to my family, especially to my parents for their support and unconditional love.

EPIGRAPH

Every one who is seriously involved in the pursuit of science becomes convinced that a spirit is manifest in the laws of the Universe a spirit vastly superior to that of man, and one in the face of which we with our modest powers must feel humble. In this way the pursuit of science leads to a religious feeling of a special sort, which indeed is quite different from the religiosity of someone more naïve

-Albert Einstein

TABLE OF CONTENTS

Signature page.....	iii
Dedication.....	iv
Epigraph.....	v
Table of Contents.....	vi
List of Figures.....	x
List of Tables.....	xv
Acknowledgements.....	xviii
Vita.....	xx
Abstract of the Dissertation.....	xxi
CHAPTER 1 INTRODUCTION.....	1
Transcription Factors Binding to Recognition Sites.....	2
Transcription Factors Key Regulators of Transcription.....	2
Specificity of Transcription Factors to the DNA Binding Site	5
The p53 Transcription Factor.....	6
Structural Organization of p53 Transcription Factor	8
Binding Specificity of p53 Transcription Factor to its Response Elements.....	10
Cooperativity in p53 Transcription Factor	12
Activation, Function and Degradation of p53 Transcription Factor	12
The p53 Response Elements Role in Activating or Repressing p53 Target Genes.....	16
The p73 Transcription Factor.....	18
Cooperative Binding Mechanism of p73.....	20
Focus of Study	20

CHAPTER 2 DNA BINDING AFFINITY OF THE p73 DNA BINDING DOMAIN TO HALF-SITE RESPONSE ELEMENTS.....	22
Introduction.....	23
Structure and stability of the DNA binding domain.....	23
Binding of p73 DBD to DNA.....	24
Materials and Methods.....	25
Subcloning and Protein Expression.....	25
Protein Purification.....	25
Measuring DNA-binding Affinity by Fluorescence Polarization.....	26
Analysis of DNA Binding Data.....	29
Results.....	30
p73 DNA-Binding Domain Purification.....	30
Binding of p73DBD to half-site response elements	33
Discussion	42
CHAPTER 3 RESPONSE ELEMENT SPECIFICITY OF THE Δ Np73 δ ISOFORM.....	44
Introduction.....	45
Materials and Methods.....	47
Subcloning of Δ Np73 δ	47
Protein Expression and Purification	47
Sedimentation Velocity Experiments	48
Sedimentation Velocity Analysis	49
DNA-Binding Fluorescence Polarization Assays.....	50
DNA-Binding Fluorescence Polarization Analysis.....	53

Results.....	54
Δ Np73 δ purification	54
Oligomerization State of Δ Np73 δ Isoform.....	58
Discussion	69
Oligomerization State of the p73 DBD and the Δ Np73 δ Isoform.....	69
Specificity of Response Element Binding in the Transcription Factor p73	71
Cooperativity of Response Element Binding in the Transcription Factor p73.....	74
CHAPTER 4 A HOOGSTEEEN BASE CONFORMATION IN p53 CORRELATES WITH ACTIVE RESPONSE ELEMENTS.....	79
Introduction.....	80
Materials and Methods.....	82
Subcloning of the p53DBD Protein.....	82
Expression of the p53DBD Protein	82
Subcloning of the AcLys120-p53DBD Gene.....	82
Expression of the AcLys120-p53DBD Protein	83
Purification of p53DBD and AcK120-p53DBD Proteins	84
Mass Spectrometry Analysis of the AcLys120-p53DBD Protein	85
Measuring p53DBD and AcK120-p53DBD Proteins Binding Affinity to Half-site Response Elements by Fluorescence Polarization.....	87
DNA Binding Data Analysis	88
Crystallization of p53DBD with Activating and Repressing Response Elements	88
Diffraction Data Collection and Structure Determination.....	92
Results.....	93
Genetically incorporation of acetyl-lysine into p53DBDTAG.....	93

Expression and Purification of the p53DBD and AcLys120-p53DBD Proteins.....	93
DNA Binding Affinity of the p53DBD and AcLys120-p53DBD Proteins.....	98
Data collection.....	103
Molecular Replacement and Refinement	103
Crystal packing.....	107
Structure Interpretation	110
Overall Structural Analysis of the AGGCA and GGACA Activating Response Element and the TTTCA Repressing Response Element.....	110
Protein- Protein interactions	116
DNA analysis.....	120
Protein-DNA interactions	130
Discussion	146
The p53DBD dimers bound to activating response elements show similar conformation compared with p53DBD dimers bound to repressing element.....	146
The Adenines in the Center of the Activating Half-Site Response Elements Flip to a Hoogsteen Conformation.....	147
DNA Binding Affinity by p53DBD and AcLys120-p53DBD Remains Unaltered upon Changing the Nucleotides in the Flanking Triplet Sequence	151
CHAPTER 5 CONCLUSIONS AND FINAL MODEL.....	153
Conclusions.....	154
Bibliography	159

LIST OF FIGURES

Figure 1.1.	The three major groups of transcription factors.....	4
Figure 1.2.	Sequence identity between different domains of the p53 family.....	7
Figure 1.3.	Structural organization of the transcription factor p53.....	7
Figure 1.4.	The p53 DNA binding domain structure.....	9
Figure 1.5.	The p53 pathway in normal cells.....	14
Figure 1.6.	The p53 pathway in stressed cells.....	15
Figure 1.7.	Activating and repressing sequences.....	17
Figure 1.8.	The p73 isoforms.....	19
Figure 2.1.	Purification of the p73DBD protein.....	31
Figure 2.2.	Elution profile of p73DBD protein.....	32
Figure 2.3.	Binding affinity of p73DBD bound to half-site of the consensus sequence.....	35
Figure 2.4.	Binding affinity graphs of p73DBD bound to a half site consensus sequence with different nucleotides at position 1.....	36
Figure 2.5.	Binding affinity graphs of p73DBD bound to a half site consensus sequence with different nucleotides at position 2.....	37
Figure 2.6.	Binding affinity graphs of p73DBD bound to a half site consensus sequence with different nucleotides at position 3.....	38
Figure 2.7.	Binding affinity graphs of p73DBD bound to a half site consensus sequence with different nucleotides at position 4.....	39

Figure 2.8.	Binding affinity graphs of p73DBD bound to a half site consensus sequence with different nucleotides at position 5.....	40
Figure 3.1.	The organizational domain of Δ Np73 δ	55
Figure 3.2.	Purification of the Δ Np73 δ isoform.....	56
Figure 3.3.	Elution profile of Δ Np73 δ in a size-exclusion column.....	57
Figure 3.4.	Sedimentation coefficient profile of Δ Np73 δ in solution unbound and bound to half-site and full-site response elements.....	58
Figure 3.5.	Binding affinity graph of Δ Np73 δ bound to the reference half-site response element.....	61
Figure 3.6.	Binding affinity graphs of Δ Np73 δ bound to a half-site response element with different nucleotides at position 1.....	62
Figure 3.7.	Binding affinity graphs of Δ Np73 δ bound to a half-site response element with different nucleotides at position 2.....	63
Figure 3.8.	Binding affinity graphs of Δ Np73 δ bound to a half-site response element with different nucleotides at position 3.....	64
Figure 3.9.	Binding affinity graphs of Δ Np73 δ bound to a half-site response element with different nucleotides at position 4.....	65
Figure 3.10.	Binding affinity graphs of Δ Np73 δ bound to a half-site response element with different nucleotides at position 5.....	66
Figure 3.11.	Binding affinity graphs of Δ Np73 δ to full-sites compared to half-site response elements.....	67

Figure 3.12.	Sedimentation coefficients of the p73 DBD bound and unbound to half and full-site response elements.....	71
Figure 3.13.	Binding profile of p73DBD and Δ Np73 δ	77
Figure 3.14.	p73DBD and Δ Np73 δ specificity.....	78
Figure 4.1.	Crystals of p53DBD in Complex with half-site Response Elements.....	91
Figure 4.2.	Mass spectrometry analysis of the p53 derived peptides from trypsin-digested AcLys120-p53DBD.....	95
Figure 4.3.	Purification of p53DBD protein.....	96
Figure 4.4.	Purification of acetylated AcLys120-p53DBD.....	96
Figure 4.5.	Elution profile of the p53DBD protein.....	97
Figure 4.6.	Elution profile of AcLys120-p53DBD.....	97
Figure 4.7.	DNA binding affinity graphs of p53DBD bound to half-site response elements with modifications in the first three nucleotides of each quarter-site response element.....	99
Figure 4.8.	DNA binding affinity graphs of AcLys120-p53DBD bound to half-site response elements with modifications in the first three nucleotides of each quarter-site response element.....	100
Figure 4.9.	Summary of the binding experiments of p53DBD and AcLys120-p53DBD bound to half-site response elements.....	101
Figure 4.10.	Diffraction Pattern of the p53DBD-TTCA Complex at 1.92Å Resolution.....	105
Figure 4.11.	Unit Cell Packing Ribbon model of the dimers packing inside a unit cell.....	108
Figure 4.12.	Structures of the p53DBD dimers bound to half-site response element.....	109

Figure 4.13.	Alignment of the three crystal structures.....	111
Figure 4.14.	Alignment of AGGCA with TTTCA showing differences in protein loops and helixH1.....	112
Figure 4.15.	Overall structure analysis.....	113
Figure 4.16.	Distances and angles between the COMs of the 2 monomers in p53DBD structures.....	115
Figure 4.17.	Monomer-monomer interaction.....	117
Figure 4.18.	Electron density map of the DNA.....	122
Figure 4.19.	Comparison of Watson-Crick base pair with the AGGCA, GGACA and TTTCA structures at their central base pair A-T.....	123
Figure 4.20.	Hoogsteen and Watson and Crick Geometry.....	124
Figure 4.21.	DNA in the AGGCA complex. a. DNA of AGGCA complex minor groove aligned with DNA minor groove containing Watson and Crick pairing.....	125
Figure 4.22.	Parameters describing the orientation and position of a base-pair.....	129
Figure 4.23.	Interactions of the protein and DNA in the minor groove of the DNA half-site response element.....	132
Figure 4.24.	Interactions of p53DBD residues with the DNA minor groove.....	133
Figure 4.25.	Overall p53DBD interactions with the mayor groove of the DNA half-site response element.....	134
Figure 4.26.	Interactions in the major groove.....	135
Figure 4.27.	Lysine conformation is different in all complexes.....	136

Figure 4.28. Binding affinity graphs of p53DBD bound to a half-site response element with different nucleotides in the flanking triplet sequence.....	144
Figure 4.29. C1'-C1' distances for the two central A-T base pairs of p53 structures.....	145
Figure 5.1. The molecular mechanism used for p53 to recognize activating response elements.....	156-158

LIST OF TABLES

Table 1.1.	DNA-binding transcription factor distribution in five organism.....	3
Table 1.2.	Crystal structures of p53 bound to DNA sequences.....	11
Table 2.1.	List of double stranded DNA sequences with fluorescein attached to the 5' of the DNA used for the binding experiments.....	27
Table 2.2.	Polarizers position used for the fluorescence anisotropy experiments.....	28
Table 2.3.	Dissociation constants of p73DBD bound to the different sequences tested.....	41
Table 3.1.	List of the 5'-fluorescein-labeled (FAM) double-stranded-DNA dodecamers used for the fluorescence anisotropy experiments.....	51
Table 3.2.	List of the double-stranded-DNA 5'-fluorescein-labeled (FAM) 22 mers used for the fluorescence anisotropy experiments.....	52
Table 3.3.	Dissociation constants of $\Delta Np73\delta$ bound to the studied half-site response elements.....	68
Table 3.4.	Dissociation constants of $\Delta Np73\delta$ bound to half-site and full-site response elements.....	68
Table 4.1.	List of 5'-fluorescein oligonucleotides used for the binding experiments.....	87
Table 4.2.	The palindromic double stranded DNA sequences containing two binding sites used to crystallize the p53DBD-DNA complex.....	90
Table 4.3.	Binding affinity of p53DBD and AcLys120-p53DBD bound to half-site response element.....	102
Table 4.4.	Data collection and refinement statistics of the three-complexes.....	106

Table 4.5.	Root mean square deviation of comparing the monomers in AGGCA, GGACA and TTTCA structures.....	113
Table 4.6.	Monomer-monomer interactions in the AGGCA complex.....	118
Table 4.7.	Monomer-monomer interactions in the GGACA complex.....	119
Table 4.8.	Monomer-monomer interactions in the TTTCA complex.....	120
Table 4.9.	Geometrical parameters of the AGGCA Half-Site DNA and B-DNA.....	126
Table 4.10.	Geometrical DNA Parameters of the GGACA Half-Site DNA and B-DNA.....	127
Table 4.11.	Geometrical DNA Parameters of the TTTCA Half-Site.....	128
Table 4.12.	Opening parameter increased in Hoogsteen base pair geometry.....	129
Table 4.13.	Interactions of monomer A with DNA observed in all the three complexes.....	137
Table 4.14.	Cysteine 277 contacts with different nucleotides in the same position.....	137
Table 4.15.	Contacts of Arg 283 and Lys 120 with different nucleotides in AGGCA and GGACA complexes.....	138
Table 4.16.	Contacts sharing by only two complexes.....	138
Table 4.17.	Contacts observed only in GGACA complex.....	138
Table 4.18.	Contacts observed only in AGGCA complex.....	139
Table 4.19.	Contacts observed only in AGGCA complex.....	140
Table 4.20.	Interactions of monomer B with DNA observed in all the three complexes.....	141
Table 4.21.	Cysteine 277 contacts with different nucleotides in the same position.....	141
Table 4.22.	Contacts of Lys120 with nucleotides in AGGCA and GGACA complexes.....	142
Table 4.23.	Contacts sharing by only two complexes.....	142

Table 4.24.	Contacts observed only in GGACA complex.....	142
Table 4.25.	Contacts observed only in AGGCA complex.....	143
Table 4.26.	Contacts observed only in TTTCA complex.....	143

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Héctor Viadiu for providing the opportunity to be part of the Viadiu Lab, for his mentorship, his patience and for being an inspiration in my career as a scientist.

To my parents, I am thankful for their support during these five years, for all the pieces of advice, and for being there for me in my moments of sadness and distress. Also, for giving me hope when I felt hopeless and for teaching me that as a scientist, I am rediscovering God's creation; therefore, I should be patience and persistent. To Julio, my fiancé, for his patience, love and understanding during long hours of work, and for making me laugh and caring for me. To my siblings, Moisés, thank you for our long chats on the phone where we talked about life and how we are what we think about; to Emmanuel, for his pieces of advice regarding life and relations with people, for his willingness to help and his caring. To my lovely Gema, my sister, for listening to me and helping me with your continuous message of love. To my brother and sister in law, Fer and Selene, for their caring and for listening to me when I was trying to explain my research project. To my nephews, Ludwig, Yaref, Emiliano, Joaquín and Alfonso, for making me realize that life is an enjoyable path. To my aunt and uncle, Alma and Alvaro, for their caring during the last five years; and to my cousins José and Carlos, for keeping in touch with me during this time.

To the Viadiu Lab, especially to Hidayath, the lovely posdoc whom I admire. I appreciate his help during the development of my project; thank you for all your teachings and even after you left lab, thank you for replying to my emails regarding my project. To my dearest friend Michelle, for your unconditional help and support during the five years, for making me feel close to home since you became like sister to me. To my lovely friend Kate, for your caring and messages full of hope and faith. To my dear Aki, Nikki, Ha and Vivien, for the moments we share

and the chats we had about our home countries. To George, for his help during the last years of my PhD. I have no other words to express my gratitude to him. To Kevin and Jessie, my mentees, for their patience, understanding, and for increasing my English vocabulary. To Thien and Frank, for our conversations regarding science, for your friendship and support. To Jill, for being in the lab with me until the last day and for your friendship. To my good friend, Jorge, for our conversations about science and life. To my posdoc friends, Ana, Estér, Alicia and Ismael, for making me feel be part of the Mexican-postdoc group.

Finally, I would like to acknowledge all the people that I met and contributed to this work either by cheering me up or by sharing their knowledge with me.

Chapter 2, in full, is contained in the manuscript named Structural Determinants of DNA Binding Specificity in the Transcription Factor p73: a Hierarchical Model of Response Element Recognition. Ramos, Ana and Viadiu, Hector. The dissertation/thesis author is the primary investigator and author of this paper.

VITA

- 2005-2009 Bachelor of Science in Chemistry
UNAM, Mexico
- 2010-2011 Teaching Assistant, Department of Chemistry and Biochemistry
University of California, San Diego
- 2010-2012 Masters of Science, Chemistry and Biochemistry
University of California, San Diego
- 2010-2015 Doctor of Philosophy, Chemistry and Biochemistry
University of California, San Diego

PUBLICATIONS

- Y. Ciribilli, P. Monti, A. Bisio, H. T. Nguyen, A. S. Ethayathulla, **A. Ramos**, G. Foggetti, P. Menichini, D. Menendez, M. a Resnick, H. Viadiu, G. Fronza, and A. Inga, "Transactivation specificity is conserved among p53 family proteins and depends on a response element sequence code.," *Nucleic Acids Res.*, vol. 41, no. 18, pp. 8637–53, Oct. 2013.

FIELDS OF STUDY

Major Field: Biochemistry

Studies in Biochemistry and Biophysics, Professor Héctor Viadiu

HONORS AND AWARDS

- 2012-2014 UC-Mexus-CONACYT Fellow
- 2010-2012 CONACYT Fellow
- 2009 B.S. Graduated with honors, Chemistry, UNAM, Mexico

ABSTRACT OF THE DISSERTATION

Structural Discrimination between Activating and Repressing Response Elements in the
p53 Protein Family

by

Ana Ramos

Doctor of Philosophy in Chemistry

University of California, San Diego, 2015

Professor Héctor Viadiu, Chair

The differential expression of genes is a fundamental hallmark of cell development. Transcription factors are key proteins that trigger gene transcription. Traditionally, it is considered that a transcription factor binding to a response element is sufficient to trigger transactivation of the regulated gene. Nonetheless, a view is emerging that considers that sole binding of the transcription site to a response element does not warrant gene activation. In my thesis work, I study by structural and biochemical techniques how the members of the p53 transcription family distinguish between response elements that will trigger gene activation or gene repression.

The p53 transcription family that comprises p53, p63 and p73 are some of the most studied proteins, particularly, because mutations in the p53 gene are present in more than 50% of cancers. Together, the three transcription factors regulate hundreds of genes in development and stress pathways. To trigger transcription, the members of the p53 family bind in a specific manner to a response element, a DNA sequence composed of two 10 bp half-sites response elements with the 5'-PuPuPuCA/TGPyPyPy-3' consensus sequence. Functional p53 and p73 require the formation of oligomers to bind with high affinity and specificity to its response element. Recently, based on functional studies, response elements have been divided into activating and repressing. Although, there is vast amount of information about p53 and p73 function and structure, little is known about their molecular mechanism to discriminate between activating and repressing response elements.

By using DNA binding assays with fluorescence polarization and crystal structure determination of protein-DNA complexes, I studied biochemically and structurally how p53 and p73 are able to distinguish between activating and repressing response elements. The analysis of my data allowed me to conclude that: 1) the different nucleotides in the p53-p73 half-site response element have each a distinctive role in discriminating whether the DNA-binding domain will distinguish between strongly or weakly bound response elements; 2) the DNA binding domain dimer changes conformation to discriminate between activating and repressing response elements in a mechanism triggered by a lysine in loop L1 and that involves the switch of the central adenine to a Hoogsteen conformation; and 3) the oligomerization domain increases cooperativity and specificity, while the acetylation of the lysine in loop L1 does not have an effect on the binding affinity of the DNA binding domain to its response element.

My work represents the clearest molecular view to date on how a transcription factor binding to its response element is not sufficient to acquire a conformation that promotes transactivation.

Chapter 1

Introduction

Transcription Factors Binding to Recognition Sites

Recently, studies have revealed that recognition sites of transcription factors are active players in the process that enables transcription factors to activate or repress target genes. Experimental data have suggested that upon binding to their recognition site, transcription factors assume different conformations, enabling them to interact with particular co-factors to either activate transcription to varying degrees, to have no effect, or to inhibit transcription. Therefore, if recognition sites are able to modify the transcription factors' activating or repressing role, it is crucial to understand how transcription factors are able to recognize them among a myriad of binding sites that would lead to different outcomes. The mechanism that transcription factors follow to recognize activating or repressing binding sites is still uncertain and a description of the molecular mechanism explaining the recognition pattern is needed to achieve a better understanding of the biological process that leads to gene expression.

Transcription Factors Key Regulators of Transcription

Transcription factors are cell components that regulate biological processes like cell cycle, intracellular metabolism, physiological balance, differentiation and development of cells [1] [2] [3]. Transcription factors are found in all living organism due to their paramount function as gene regulators.

Analyses of the genome sequence and prediction of the Gene Ontology (GO) data base showed the presence of 1,000 to 3,000 sequence-specific DNA binding domains (DBD) in humans corresponding to 8% of the total encoded proteins. Those DNA binding domains are part of transcription factors [4] (Table 1.1.). From the transcription factors coded by these genes, only 62 have been experimentally demonstrated to have DNA binding properties and regulatory functions [5].

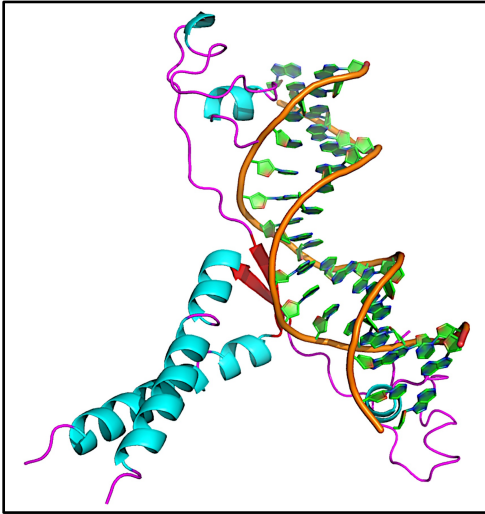
Table 1.1. DNA-binding transcription factor distribution in five organisms

Number of DNA-binding transcription factors in five organisms			
Organism	Number of transcripts	Number of proteins DNA-binding domains	Percentage of transcripts containing DNA-binding
<i>E. coli</i>	4280	267	6.2
<i>S. cerevisiae</i>	6357	245	3.9
<i>C. elegans</i>	31 677	1463	4.6
<i>H. sapiens</i>	32 036	2604	8.1
<i>A. thaliana</i>	28 787	1667	5.7

Source: Babu et al. (2004). Structure and evolution of transcriptional regulation networks, *Curr. Opinion in Structural biology*, 14(3), 283-91.

In general, the transcription factors (TFs) are classified based upon the structure of their DNA binding domains, which play an important role in DNA binding specificity [6]. Among the several conserved groups of transcription factors, 3 are predominantly observed that account for 80% of the total regulatory proteins in the human genome: the Zinc-coordinating (675 TFs), the homeodomain (257 TFs) and the helix-loop-helix (87 TFs) groups [5] (Figure 1.1.).

a. Zinc-coordinating group: GAL4 domain



b. Homedomain: Engrailed home



c. Helix loop helix group: Interferon regulatory factor 2

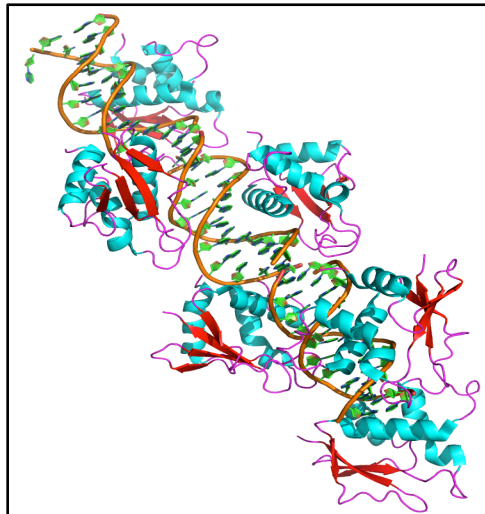


Figure 1.1. The three major groups of transcription factors. a. GAL4 (PID 1ZME) represents the Zinc-coordinating group. b. Engrailed home domain (PID 2HDD) is a member of the Homedomain group. c. Interferon regulatory factor (PID 2IRF) is part of the Helix-loop-helix group.

Specificity of Transcription Factors to the DNA Binding Site

In order for the TF to bind in a specific manner, it needs to bind tighter to the target site compared with a non-specific DNA sequence. In general, the equilibrium constant of a protein binding to a non specific DNA is 10^3 - 10^6 M⁻¹ with a net free energy change (ΔG) of -4 to -7 kcal/mole per site, whereas the equilibrium constant of the protein binding to a specific site is 1000 to $> 10^7$ times tighter, but it must not exceed 10^{12} since the binding needs to be reversible. The overall energy change for specific binding is in the range of -11 to -15 kcal/mole per site [7] [8].

Transcription factors recognize their specific DNA sequences by two mechanisms: base readout and shape readout. In the base readout mechanism, specificity is achieved by direct or indirect interactions. Direct contacts are formed between protein side chains and nucleotide bases located mainly in the major groove via hydrogen bonds or hydrophobic interactions [9] [10], whereas indirect contacts comprise water-mediated interactions of the protein side chains with the DNA bases in the two DNA grooves. In the major groove, the functional groups of the bases are more accessible to protein side chains than in the minor groove; thus, a highest potential for base readout exist in the major groove compared with the minor groove [11]. Finally, the sequence specific interactions determine the specificity by base readout. In the shape readout mechanism, the DNA conformation/structure plays a critical role in protein recognition. The DNA deformation from the B-DNA could be local (kink, minor and major groove) or global (bend, A- or Z-DNA) [12]. It is worth mentioning that DNA deformations influence the base readout, making the bases in the major and minor groove more or less accessible. An example where both mechanisms, base and shape readout, are used to achieve specificity is given by the *Escherichia coli* Trp Operator. The structural information from free-DNA and the DNA-protein complex showed that specificity is achieved by 3 main components: bases readout, structural changes driven by the DNA sequence, and water-mediated interactions. From 10 water molecules

conserved in the free-DNA and DNA-protein complex, only three of them were mediating contacts between the protein and the nitrogen atoms of purines. Also, the DNA from the DNA-protein complex was bent by 15° as a consequence the minor groove got wider and the major groove narrowed; however, the overall shape of the bound DNA still resembled the B-DNA form observed in the free-DNA structure [13] [14].

The p53 Transcription Factor

The p53 protein is a member of the loop-sheet-helix group, along with p73 and p63, which altogether form the p53 family. All three members of the p53 family share high sequence homology especially in its DNA binding domain, DBD (Figure 1.2.). This family is also part of the zinc-coordinating group of transcription factors. This group of transcription factors is characterized by the formation of a tetrahedral coordination to one or two zinc ions with cysteine and histidine residues in the DNA binding domain. This coordination net confers increasing stability to the core structure of the protein[15].

The p53 transcription factor consists of 393 amino acids and it is organized in three major functional domains: the N-terminus, the DNA binding domain and the C-terminus (Figure 1.3.). The X-ray structure of the full-length p53 has not been elucidated yet due to flexible regions in the N and C-termini of the protein [16]; nonetheless, the structure of full-length p53 in the absence of DNA was reconstructed using single-particle electron microscopy [17]. Based on the reconstructed 3D structure, full-length p53 is a tetramer formed by two dimers. The tetramer adopts C2 symmetry and has a relaxed conformation that upon binding to DNA changes to a tight conformation. In the relaxed structure, unbound to DNA, the DNA binding domains of the two dimers do not interact with one another since the distance between each DBD increases to 45 Å and 71 Å in each dimer. Therefore, a dramatic conformational change has to occur when bound to DNA in order to get the DBDs of each dimer to contact one another as observed in the low-pass filtered crystal structure of p53 DBD bound to DNA [18].

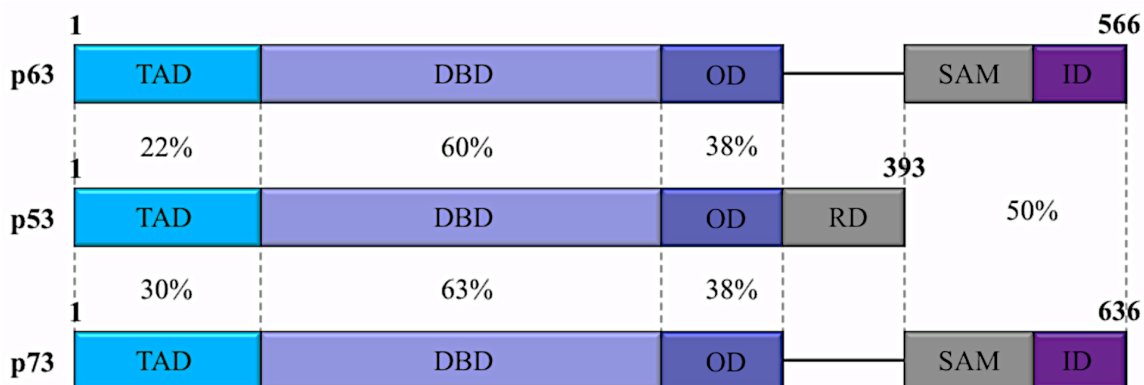


Figure 1.2. Sequence identity between different domains of the p53 family. The p53 family is conserved in transactivation domain (TAD), DNA binding domain (DBD) and oligomerization domain (OD). Besides, p73 and p63 have an additional two domains in the C-terminus: the sterile alpha domain (SAM) and the inhibitory domain (ID). The percentage of similarity between p53/p63 (top) and p53/p73 (bottom) is depicted.



Figure 1.3. Structural organization of transcription factor p53. The p53 protein consists of 393 amino acids and the DNA binding domains comprises over 50% of the total protein residues content.

Structural Organization of p53 Transcription Factor

The three domains of p53 play specific roles in order to fulfill p53's role as a tumor suppressor. The N-terminal of p53, called the transactivation domain (TAD), comprises residues 1-92 and it induces the transcriptional activity of the p53-regulated genes. This region has two different patches of residues, one of which contains a high content of aspartic and glutamic acid residues; the second patch is rich in proline residues. Apart from these defined regions; the transactivation domain is further divided into transcriptional activation domain 1 (TAD1, residues 1-42) and transcription activation domain 2 (TAD2, residues 43-92) based on the genes that are activated. For instance, TAD1 primes transcription of cell arrest genes and TAD2 drives the activation of apoptotic genes [19]. The TAD is an intrinsically disordered region, a hallmark in transcription factors because it facilitates binding to diverse target proteins [20]; nevertheless, upon binding to its target protein, the TAD acquires a secondary structure.

The domain that confers binding specificity to the p53 transcription factor is the DNA binding domain (DBD), it is conformed by residues 93 to 293 and it is the most structured domain of p53 (Figure 1.3. and 1.4.). Around 80% of mutations presenting in human tumors occur in this domain [21]. The structure of the DBD could be described as an immunoglobulin-like anti-parallel β -sandwich, comprised of two antiparallel β sheets of four (S1, S3, S8 and S5) and five (S10, S9, S4, S7 and S6) strands; these 2 sheets form the hydrophobic inner core. Apart from the 2 sheets, the DNA binding domain contains 2 helices (H1 and H2) and 3 loops (L1, L2, L3) that emerge from the side of the hydrophobic inner core. Finally, a tetrahedrally coordinated Zn with Cys176, His179, Cys238 and Cys242 is found close to H1, which confers stability to the hydrophobic core. (Figure 1.4.) [22] [23].

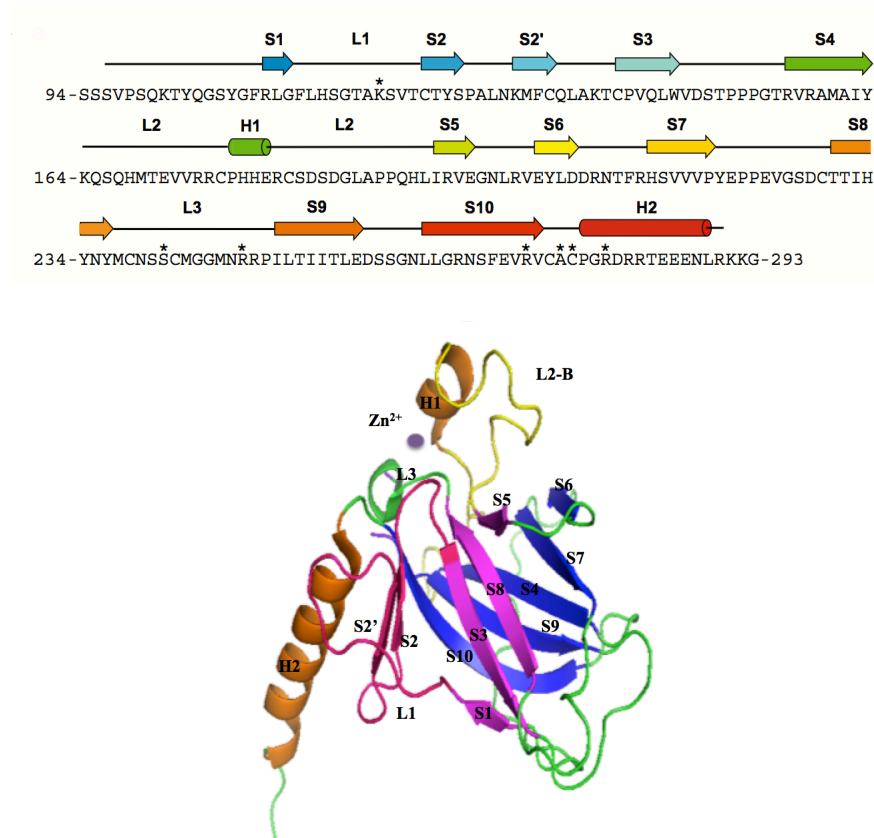


Figure 1.4. The p53 DNA binding domain is the most structured domain in the p53 transcription factor (PDB ID 1TSR).

The oligomerization domain (OD) is responsible for dimer and tetramer formation in solution in the absence of DNA. It is a segment comprising residues from 325 to 360 [24]. The OD is linked to the DBD by a short domain called the nuclear localization signal (NLS, residues 305-321) that promotes p53 transport to the nucleus [25]. The oligomerization domain is located in the C-terminal region of p53. Although it is rarely mutated in cancer [26], it is important for the protein-protein interactions within the two dimers of the functional p53 tetramer [27] [28].

The p53 regulatory domain (RD) is another unstructured region and consists of 30 amino acids in the C-terminal region of the protein comprising residues 362 to 393. Like the TAD, upon binding to its protein partners, the RD gains some secondary structure [16]. For instance, in complex with deacetylase Sir2Tm, p53 RD adopts a helical structure [29]. The basic character of this domain, which derives from the presence of lysines and arginines, makes it bind non-specifically to DNA [30]. Also, the RD domain is target of posttranslational modifications such as acetylation, methylation, phosphorylation and ubiquitination that enables p53 degradation by the proteasome [31][32] [33].

Binding Specificity of p53 Transcription Factor to its Response Elements

Transcription factors bind to a specific site that is recognized by the DNA binding domain. The p53 protein binds to the recognition site or response elements usually located before the promoter of the target gene. Most p53 response elements follow a consensus sequence consisting of two decameric repeats containing the sequence -PuPuPuCA/TGPyPyPy- where Pu= guanine or adenine, Py= thymine and cytosine [34]; moreover, half sites of biological significance are separated by 0-3 nucleotides [35].

Structures of DNA binding domain (DBD) and the DNA binding domain together with the oligomerization domain (DBD-OD) of p53 bound to consensus sequence have been solved by X-ray crystallography (Table 1.2.). Based on this structural information, p53 DNA binding domain makes direct and indirect (water mediated) contacts with the DNA through five common residues in the major groove and two residues in the minor groove; although contacts may vary depending on the DNA sequence. In the major groove, Lys120 from L1, Cys277, which is located between H2 and S10, and Arg280 in H2 make direct contacts with the nucleotide bases in the DNA. Finally, two residues, Ala276 that is located between H2 and S10, and Arg273 in S10, interact with DNA through the phosphate backbone. In the minor groove, Ser241 from L3 contacts the phosphate backbone and Arg248 from L3 adopts different conformations depending

on the DNA sequence in which it may contact the DNA backbone or the nucleotide bases mediated by water molecules [22] [36] [37] [38].

Table 1.2. Crystal structures of p53 bound to DNA sequences

Quarter-site	Description	Resolution (Å)	PDB ID	Reference
5'-GGGCA-3'	First p53 core domain complexed with DNA	2.2	1TSR	Cho et al., 1994
5'-GGGCA-3'	p53 tetramer with two half-sites	1.8	2AC0	Kitayner et al., 2006
5'-AGGCA-3'	p53 tetramer with two half-sites	2.2	2ATA	Kitayner et al., 2006
5'-GGACA-3'	p53 tetramer with two half-sites	1.85	2AHI	Kitayner et al., 2006
5'-GGACA-3'	p53 dimer with one half-site	2.5	2ADY	Kitayner et al., 2006
5'-GAGCA-3'	p53 dimer crosslinked to 16bp	2.3	2GEQ	Ho et al., 2006
5'-GGGCA-3'	p53 tetramer bound to a full site	1.91	3KZ8	Kitayner et al., 2010
5'-AGGCA-3'	p53 tetramer bound to a full site	2.14	3KMD	Chen et al., 2010
5'-GGGCA-3' 5'-AGACA-3'	p53 DBDOD mutated bound to full site as a dimer	2.4	3Q05	Petit et al., 2011
5'-GGGCA-3' 5'-AGACA-3'	p53 DBDOD mutated bound to p21 response element	2.8	3TS8	Emamzadah et al., 2011
5'-GGGCA-3' 5'-AGACA-3'	p53 DBDOD mutated bound to full site as a dimer with S121F and V122G DBD mutations	2.9	4MZR	Emamzadah et al., 2014

Source: Viadiu, H. (2008). Molecular architecture of tumor suppressor p53. *Current Topics in Medicinal Chemistry*, 8, 1327–1334, adapted by Ana Ramos

Cooperativity in p53 Transcription Factor

In order for p53 to transactivate genes, it needs to oligomerize. p53 forms dimers at low concentrations, whereas higher concentrations of p53 result in tetramer formation, which is reflected in the dimer to tetramer equilibrium constant of 3 μM , indicating the dependency of concentration in the formation of p53 tetramers [39]. Also, upon DNA binding to a full site, p53 needs to tetramerize in order to better transactivate the target genes. The p53CT truncation containing the DNA binding domain and the oligomerization domain binds to its recognition site in the nM interval; whereas, the p53 DBD by itself binds in a μM range [39]. The difference in binding affinity could be explained by an increase in protein-protein contacts, protein-DNA interactions and DNA flexibility allowing a highly cooperative binding process, which is a hallmark in almost all transcription factors. Apart from an increase in affinity, cooperativity plays a pivotal role in p53 as a switch between cell cycle and apoptosis gene activation since recent studies have shown that high cooperativity increases p53 apoptotic functions, whereas weak cooperativity favors activation of cell cycle genes. Consequently, cooperativity contributes to the wide-ranging tumor suppressor activity of p53 [40].

Activation, Function and Degradation of p53 Transcription Factor

The tumor suppressor p53 was discovered in 1979 and it was the first tumor suppressor to be identified [41] [42] [43]. It is the most commonly mutated gene, appearing in 50% of human cancers including liver, ovary, esophagus, colorectal, lung, and brain [44]. In unstressed cells, p53 is present in low concentrations, and its short half-life only ranges from 6-20 minutes from the moment it is translated in the cytoplasm. MDM-2 is an ubiquitin ligase that confers the short half-life to p53 by directly interacting with the tumor suppressor protein and catalyzing its mono-ubiquitination [45] [46]. Furthermore, CBP/p300 (CREB binding protein transcriptional coactivator/p300) adds ubiquitin molecules to the mono-ubiquitinated p53 to be finally degraded by the proteasome in the cytoplasm [47] (Figure 1.5.).

Upon stress signals such as UV-light, hypoxia, DNA damage and heat shock, p53 is activated by different posttranslational modifications such as acetylation, phosphorylation, methylation, which avoids p53 degradation by MDM-2, its negative regulator [48] [49]. Another mechanism is the activation of positive p53 regulators such as p14ARF and c-Abl which bind to MDM-2 and inhibit its interaction with p53; thus, keeping high p53 concentrations in the damaged cells [50][51]. Then, p53 is exported to the nucleus where it binds specifically to its response element, usually located in the promoter region of the target genes, to activate or repress the target genes [52] (Figure 1.6.).

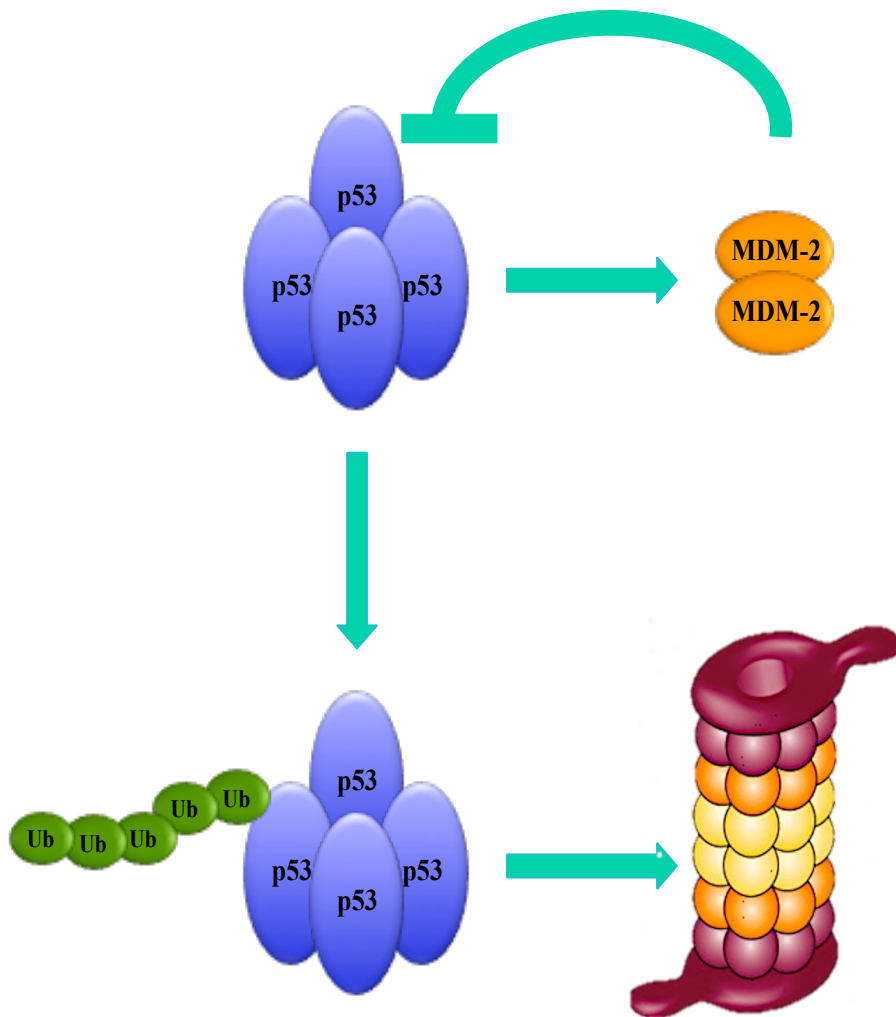


Figure 1.5. The p53 pathway in normal cells.

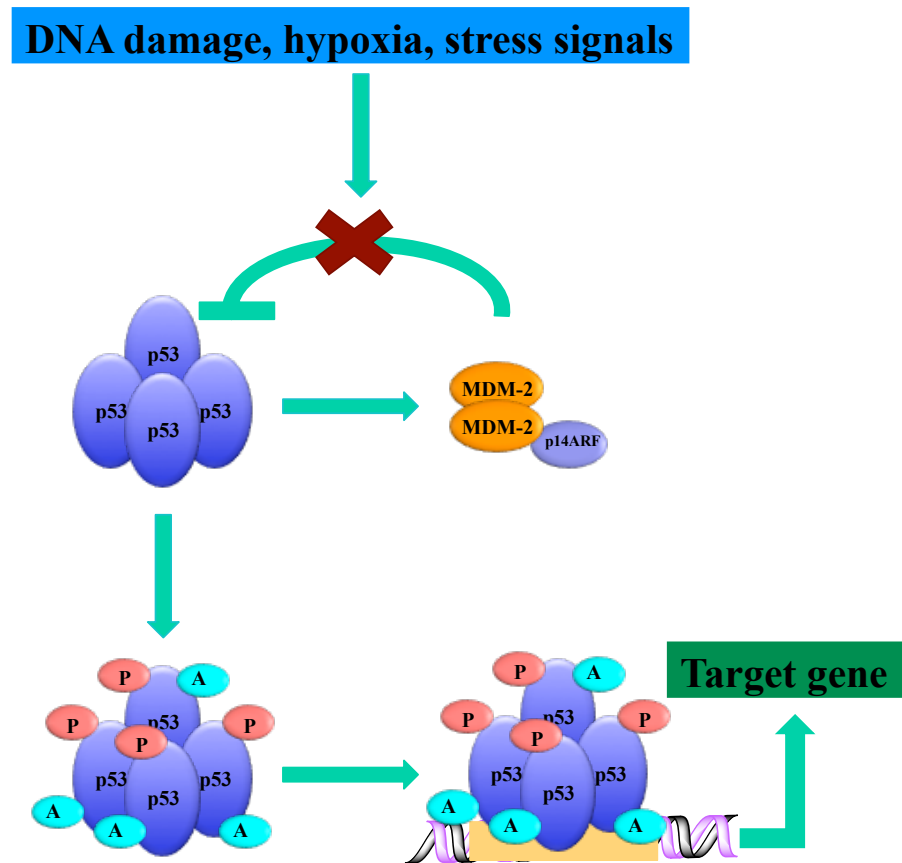


Figure 1.6. The p53 pathway in stressed cells.

The p53 Response Elements Role in Activating or Repressing p53 Target Genes

The p53 transcription factor regulates around 129 genes and binds to 160 response elements (p53 regulated genes can contain more than one response element) [35]. The response element (RE) comprises two decamers, each with a consensus sequence of PuPuPuCA/TGPyPyPy [34] although variations in the consensus sequence are frequently observed. The conserved cytosine and its pair guanine -CA/TG- are crucial for p53 binding specificity. By performing luciferase reported assays to test activation and repression of p53 target genes upon p53 binding to different response elements, Wang et al. defined rules describing the function of the other nucleotides in the p53 RE. The core CWWG is the major determinant in the activation and repression function where the doublets pair in the central core (CWWG) AT, TT and AA have activating properties, while the TC, GA, TG, CA, GC, GG, CC, CG show repressing features. The second determinant is the triplet flanking sequence, which exerts a modulating effect (Figure 1.7.) and finally, the adjacent nucleotide to the core XCWWGY shows the strongest positional effect [53]. These three rules might serve as a starting point to understand the complex regulatory networks of p53 transcription factor in gene regulation.

Group	Half-site 1		Half-site 2		Luciferase assay*
	RRR	YYY	RRR	YYY	
p21	GAA	TCC	CAA	TTG	550.7 ± 21.4
Activating	G ₃	T ₃	A ₃	C ₃	643.4 ± 16.0
	A ₃	C ₃	A ₃	C ₃	556.5 ± 28.2
	G ₃	T ₃	G ₃	T ₃	446.7 ± 36.3
	A ₃	C ₃	G ₃	T ₃	431.5 ± 12.4
	G ₃	T ₃	G ₃	A ₃	573.9 ± 64.0
	A ₃	C ₃	T ₃	C ₃	448.2 ± 47.9
	G ₃	A ₃	G ₃	T ₃	435.5 ± 22.4
	T ₃	C ₃	A ₃	C ₃	433.7 ± 31.3
	T ₃	C ₃	T ₃	C ₃	303.5 ± 35.2
	T ₃	G ₃	A ₃	C ₃	98.8 ± 7.8
Repressing	A ₃	G ₃	A ₃	G ₃	63.9 ± 13.5
	T ₃	C ₃	A ₃	G ₃	55.7 ± 3.9
	A ₃	G ₃	T ₃	C ₃	37.7 ± 2.8
	A ₃	C ₃	T ₃	G ₃	36.1 ± 1.2
	C ₃	A ₃	G ₃	A ₃	68.8 ± 4.8
	T ₃	G ₃	T ₃	C ₃	47.8 ± 1.0
	T ₃	C ₃	T ₃	G ₃	37.8 ± 1.0
	G ₃	A ₃	C ₃	A ₃	30.4 ± 0.9
	C ₃	A ₃	C ₃	A ₃	67.8 ± 3.2
	C ₃	A ₃	T ₃	G ₃	52.2 ± 1.9
	T ₃	G ₃	C ₃	A ₃	40.4 ± 4.4
	T ₃	G ₃	T ₃	G ₃	35.6 ± 0.9

Figure 1.7. The normalized luciferase activity of 20 luciferase constructs with different combinations of the triplet flanking sequence using p21 response element as background. G₃= GGG, A₃= AAA, C₃= CCC, T₃= TTT. Source: Wang, B., Xiao, Z. & Ren, E. C. Redefining the p53 response element. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14373–8 (2009).

The p73 Transcription Factor

In contrast to p53 that is mutated in 50% of human cancers, p73 is mutated in less than 0.5% of cases [54]. Additionally, p73 can activate p53 target genes leading to apoptosis and DNA repair [55] by recognizing and binding to p53 response elements in a specific manner. The p73 specificity for p53 response elements is derived from the high degree of sequence homology between p53 and p73 DNA binding domains (Figure 1.4.). Furthermore, protein-DNA contacts are conserved in both p53 (Lys120, Cys277 and Arg280) and p73 (Lys138, Cys297 and Arg300) [56]. Therefore, p73 could theoretically also act as a tumor suppressor and it could become a therapeutic target to replace mutated p53 [57]. Apart from its tumor suppressor activity, p73 is also involved in pheromonal sensory, chromosome stability, inflammation, neurogenesis and osteoblastic differentiation pathways [58] [59]. Like p53, p73 is constituted by three major domains: N-terminus, the DNA binding and the C-terminus domain (Figure 1.2.) where the TAD, the DBD and OD share sequence identity with p53 domains. On the other hand, p73 C-terminus contains an extra domain named the sterile- α -motif domain (SAM) that is involved in protein-protein interactions [60].

Unlike the TPp53 gene, the TP73 gene gives rise to different mRNAs. The different mRNAs are the result of alternate splicing and the use of two different promoters. The p73 isoforms regulated by two different promoters are named TAp73 and Δ Np73. The TAp73 contains the transactivation domain encoded by exons 2 and 3, whereas the Δ Np73 lacks the transactivation domain. Moreover, the alternative splicing in the C-terminus produces 7 different isoforms (α - η) [58] [61][62] (Figure 1.8.). The TA and Δ N isoforms have opposite functions. Whereas TA isoforms are activators of p53 responsive genes regulating growth arrest and cell death as tumor suppressors, the Δ N isoforms block the transactivation activity of TAp73 and p53 transcription factors and they are incapable of inducing gene expression [63] [64] [65] [66].

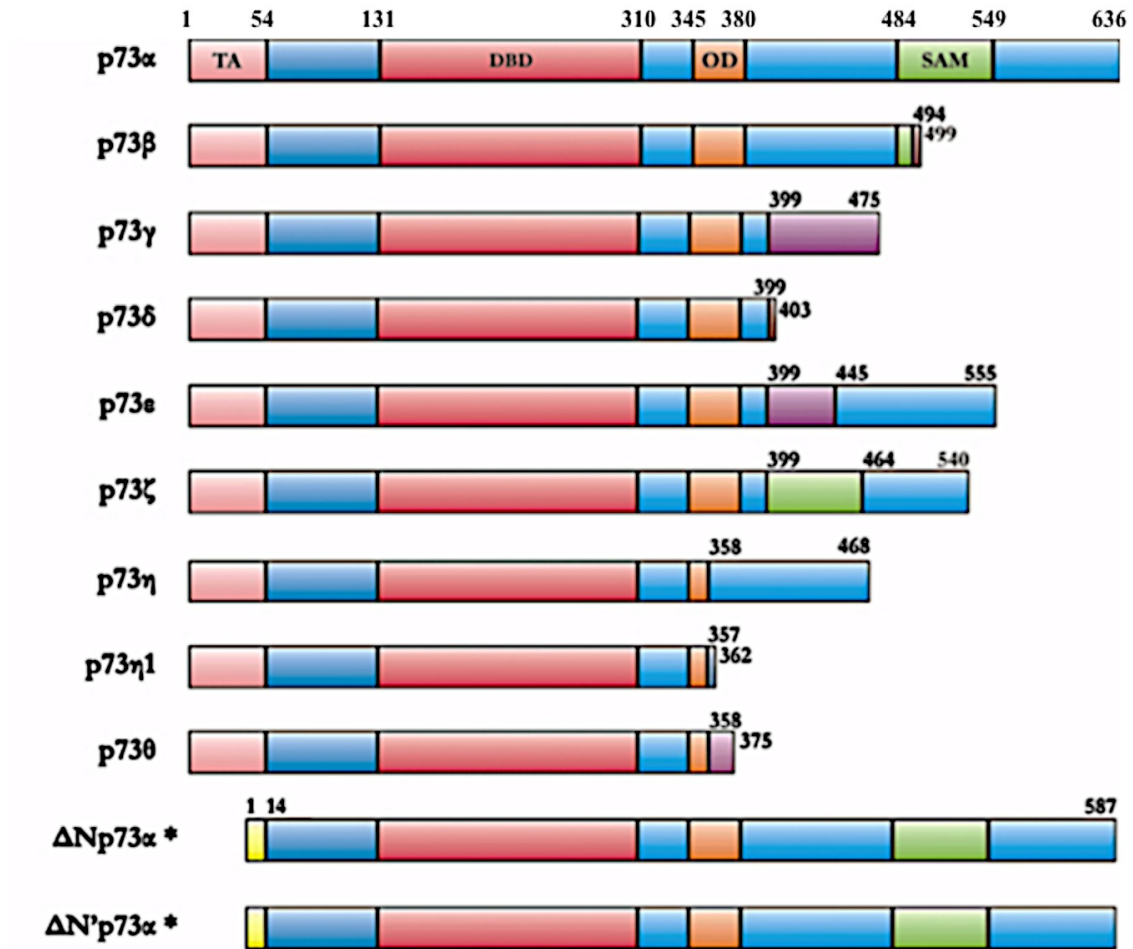


Figure 1.8. Isoforms of p73 from two different promoter regions and alternative splicing: TAp73 containing the transactivation domain and $\Delta Np73$ without the transactivation domain. Source: Pluta, A., Nyman, U., Joseph, B., Robak, T., Zhivotovsky, B., & Smolewski, P. (2006). The role of p73 in hematological malignancies. *Leukemia*, 20(5), 757–66

Cooperative Binding Mechanism of p73

As in p53 transcription factor, p73 binds to its recognition site in a cooperative manner [67] [68]. The p73 DNA binding domain in solution behaves as a monomer that upon DNA binding forms dimers or tetramers depending on the DNA length. The p73 DNA binding domain (DBD) binds to the consensus sequence in a μM range, which is similar to p53 DBD binding affinity [56] [39]. However, the binding affinity of p73 increases when the oligomerization domain is present. For instance, the binding affinity of the isoform $\Delta\text{Np73}\delta$ (containing the DBD and OD) is 10 times greater than the p73 DBD bound to full and half-site consensus sequence [56] [68]. Thus, both p53 and p73 transcription factors bind to its recognition site in a similar way where oligomerization is the first step to bind in a cooperative way with increased specificity and affinity.

Focus of Study

Based on the fact that the oligomerization domain in the naturally occurring isoform $\Delta\text{Np73}\delta$ confers an increase in affinity compared with the DNA binding domain of p73, and that the oligomerization domain triggers a cooperative behavior in $\Delta\text{Np73}\delta$ upon binding to the DNA [68] [56], I hypothesized that the increase in binding affinity and the observed cooperative behavior might also affect the specificity of the protein to the response elements (REs) since transcription factors' cooperative behavior results in an increase of binding affinity and enhancement of specificity [9] [10]. Therefore, in the first part of the thesis, I aim to understand the specificity requirements of both the p73 DBD and the biological relevant $\Delta\text{Np73}\delta$ isoform, which will enable us to comprehend better the role of the oligomerization domain in p73 binding specificity.

Based on the rules governing the roles of the nucleotides in the RE established by Wang et al., p53 could activate or repress a target gene depending on the response element sequence it

binds to; however, the molecular mechanism that p53 uses to distinguish between those two pathways is poorly understood. Structural information shows that the conformation of Lys120 in p53 changes depending on the RE sequence p53 binds [36]; besides, the central doublets (WW), AT, TT and AA in the response elements –PuPuPuCWWGPyPyPy- adopt a Hoogsteen geometry upon p53 binding[69] [38]. Taking together these information, I asked how the change in lysine conformation and the observed Hoogsteen geometry in the central doublets upon p53 binding to the DNA relate to p53's ability to recognize between a repressing and activating response element (RE). Therefore, in the second part of my thesis, I aim to give an answer to that question by investigating at a molecular level the recognition pathway p53 employs to differentiate between activating and repressing RE in order to initiate or abrogate transcription of the target gene.

Chapter 2

DNA Binding Affinity of the p73

DNA Binding Domain to Half-Site

Response Elements

Introduction

The p73 protein is a homologue of the p53 tumor suppressor. Both proteins contain an N-terminus domain, the DNA binding domain and a C-terminus domain. Among the three domains, the DNA binding domain shares the highest sequence identity of 58% between p73 and p53 that makes p73 an attractive candidate to study due to its high homology to p53, its capacity to bind some of the p53 response elements [70] and its minimum occurrence of mutations related to cancer [54].

Structure and stability of the DNA binding domain

The DNA binding domain of p73 comprises 30% of the entire p73 protein with 197 amino acids. In solution, the CD secondary analysis of p73DBD protein indicated that 34.6% is β -sheet and 7.4% is alpha helix at 20°C [67]. This data is totally supported by the crystal structure of the p73 DBD that showed an immunoglobulin-like fold with two β sheets built from antiparallel strands each. One of the β sheets has four β strands- S1, S3, S5 and S8- and the second one has five β strands – S4, S6, S7, S9, S10. Three loops emerging from the core immunoglobulin fold serve as strand-strand connectors. For instance, loop 1 links β strands S1 and S3, it also contains two small β strands, S2 and S2'; loop 2 is the longest of the 3; therefore, it is divided into L2A and L2B and it connects H1 and S5. Finally, loop 3 brings together S8 and S9. An important feature of p73 DNA binding domain relevant for dimerization and DNA binding is the coordination of a Zn^{2+} ion with Cys194 and His197 from H1 and Cys258 and Cys262 from L3.

Unlike p53 DBD, p73 DBD is thermodynamically stable. Based on spectroscopic data, p73 DBD possesses a T_m of 51.2°C compare with 42°C of p53 DBD [67]. The stability factor could be useful when designing a cancer therapy using p73 where p53 is mutated.

Binding of p73 DBD to DNA

The p73 DBD behaves as a monomer in solution; upon DNA binding, it can form a dimer or a tetramer depending on the length of the DNA [56]. Like p53 DBD, p73 DBD can bind to half and full-site of sequences following the consensus rule [34]. The DNA binding domain of p73 binds to half site (5'-GGGCATGCCC-3') and to a full site (5'-GGGCATGCCCCGGGCATGCCC-3') consensus sequence with an affinity of 2.9 μ M and 2.4 μ M, respectively [56]. Based on the crystal structures of p73 DBD bound to consensus sequences, the residues involved in the interaction with DNA are located in Loop1, Strand10, Helix2 and L3. Some of these residues contact the bases or the backbone of the DNA. For instance, Lys138 (L1) contacts guanine or adenine in the second position; cysteine 297 (H2) contacts cytosine at position 4th, and Arg300 (H2) contacts the complementary base of cytosine, guanine, in the 4th position, all of these contacts happen in the major groove of the DNA. Other contacts with the phosphate backbone of the DNA and side chains of Ser261, Arg268 (L3) and Arg296 (S10) occur to stabilize the complex. Although, the residues of p73 DBD that contact the specific bases and phosphates in the backbone of the DNA are known, little is understood on how the binding affinity of the protein is affected when contacting bases in the DNA are replaced. It is unclear how deviations from half site of the consensus sequence –5'- PuPuPuCA/TA/TGPyPyPy -3' could affect its binding affinity. Therefore, in this chapter, I showed the effect on the binding affinity of p73 DBD upon replacing each of the bases in the consensus sequence shedding light on the importance of the consensus nucleotides in the mechanism of DNA binding of p73.

Materials and Methods

Subcloning and Protein Expression

Residues 115-312 that codify for the human p73 DNA-binding domain (DBD) were cloned into the pET28a over-expression vector with a His-tag at the N-terminus. Throughout this thesis, I will refer to the gene product of this over-expression construct as the p73DBD protein. The *EcoRI* and *HindIII* restriction sites in the polylinker were used for cloning. The full-length human p73 gene served as a template for the PCR amplification of the DBD domain with the selected cloning flanking sites. The vector carrying the p73DBD construct was used to transform BL21/DE-3 *E. coli* competent cells that were selected with the antibiotic kanamycin. Cells were grown in LB medium with 30 µg/ml of kanamycin at 37 °C. Upon reaching an absorbance at 600 nm of 0.6 AU, cells were induced with 0.5 mM IPTG and grown overnight at 25°C.

Protein Purification

After induction, cells were harvested by centrifugation at $2,846 \times g$ for 30 min at 5°C. The supernatant was discarded and the cells in the pellet were resuspended in lysis buffer (0.5 M sodium chloride, 20 mM sodium citrate (pH 6.1) and 10 µM zinc chloride). Cells were lysed using a microfluidizer and centrifugated at $104,444 \times g$ to remove the cell debris from the cytoplasmic fraction. The protein was purified using affinity chromatography by incubating overnight the soluble fraction containing the p73DBD protein with 1 ml of nickel-nitrilotriacetic acid (Ni-NTA) resin from Roche at 5°C. To remove the contaminant proteins bound non-specifically to the affinity resin, the resin was washed with 150 ml of lysis buffer and 50 ml of lysis buffer supplemented with 10 mM imidazole. Finally, the protein was eluted in lysis buffer supplemented with 200 mM imidazole. Afterward, the protein was concentrated to 1.7 mM using a centrifugal filter concentrator. To further purify the protein, one milliliter was loaded into a Superdex 200 gel filtration column equilibrated with binding buffer (100 mM sodium chloride, 10

mM sodium citrate (pH 6.1) 5 mM DTT and 5 μ M zinc chloride). The degree of purity was determined by running a 15% SDS-PAGE and the identity of the protein was confirmed by a western blot using an anti His-tag antibody from Roche.

Measuring DNA-binding Affinity by Fluorescence Polarization

The p53 family binds to a consensus sequence or response element consisting of a two repeated decamer separated by 0-13 nucleotides; the decamer -5'-PuPuPuC(A/T)GPyPyPy-3' contains a half site of the consensus sequence. To study the affinity of the p73DBD protein for different half-site response element sequences, a set of 5'-fluorescein-labeled 12 base-pair dsDNAs were acquired from Integrated DNA Technologies (IDT). Table 2.1. lists the 16 oligonucleotides used. The sequence GGGCATGCCC was used as a reference for the known high affinity of the p73DBD protein for this sequence [56]. Based on this reference sequence, individual single mutations were introduced in each quarter-site to explore the contribution of each position for binding. In this work, to refer to each position of the quarter-site where the p73DBD binds, I numbered them from the outside to the center of the half-site sequence. For example, for the sequence 5'-GGGCATGCCC-3', the first G is referred as position 1, the second G as position 2, the third as position 3, the C as position 4 and the next position as position 5. The positions of the adjacent inverted repeat are numbered with the equivalent numbers from the center 5, 4, 3, 2 and 1 (See Table 2.1.). I added an extra nucleotide at the 5' of the sequence to covalently link the Fluorescein to the response element in order to ensure a proper annealing of each sequence used.

Table 2.1. List of double stranded DNA sequences with fluorescein attached to the 5' of the DNA used for the binding experiments. The numbers represent the positions of the nucleotides in each sequence.

Name-ID	Sequence
12345 GGGCA	1234554321 5'-FAM-aGGGCATGCCct-3' 3'-tCCCCGTACGGGa-FAM-5'
12345 AGGCA	1234554321 5'-FAM-aAGGCATGCCt-3' 3'-tTCCGTACGGa-FAM-5'
12345 CGGCA	1234554321 5'-FAM-aCGGCATGCCt-3' 3'-tGCCGTACGGc-FAM-5'
12345 TGGCA	1234554321 5'-FAM-aTGGCATGCCa-3' 3'-tACCGTACGGT-FAM-5'
12345 GAGCA	1234554321 5'-FAM-aGAGCATGCt-3' 3'-tCTCGTACGAGa-FAM-5'
12345 GCGCA	1234554321 5'-FAM-aGCGCATGCCt-3' 3'-tCGCGTACGCGa-FAM-5'
12345 GTGCA	1234554321 5'-FAM-aGTGCATGCAct-3' 3'-tCAGTACGTGa-FAM-5'
12345 GGACA	1234554321 5'-FAM-aGGAACATGTCCt-3' 3'-tCCGTGTACAGGa-FAM-5'
12345 GGCCA	1234554321 5'-FAM-aGGCCATGGCCt-3' 3'-tCCGGTACCGGa-FAM-5'
12345 GGTCA	1234554321 5'-FAM-aGGTCATGACc-3' 3'-tCCAGTACTGGa-FAM-5'
12345 GGGAA	1234554321 5'-FAM-aGGGAATTCc-3' 3'-tCCCTTAAGGa-FAM-5'
12345 GGGGA	1234554321 5'-FAM-aGGGGATCCc-3' 3'-tCCCCTAGGGa-FAM-5'
12345 GGGTA	1234554321 5'-FAM-aGGGTATACc-3' 3'-tCCCCATATGGa-FAM-5'
12345 GGGCC	1234554321 5'-FAM-aGGGCCGCCc-3' 3'-tCCCCGCCGGa-FAM-5'
12345 GGGCG	1234554321 5'-FAM-aGGGCGGCCc-3' 3'-tCCCCGCGGGa-FAM-5'
12345 GGGCT	1234554321 5'-FAM-aGGGCTAGCCc-3' 3'-tCCCCATCGGa-FAM-5'

To measure the DNA binding constant of the p73DBD protein for each selected half-site response element, I prepared 15 to 17 serial protein dilutions from 100 μM to 2 nM. Each protein concentration was mixed with 50 nM 5'-fluorescein-labeled 12-mer dsDNA in a 350 μL final volume of binding buffer (100 mM sodium chloride, 10 mM sodium citrate (pH 6.1) 5 mM DTT and 5 μM zinc chloride). Tubes were incubated for 30 minutes at 5°C before making the fluorescence polarization measurements. To determine the fluorescence polarization, four fluorescence intensity measurements with different filter arrangements at excitation and emission wavelengths of 494 nm and 521 nm were done using a Hitachi F-2000 fluorescence spectrometer (see Table 2.2.). For each half-site response element studied, the binding experiment was repeated three times.

Table 2.2. Polarizers' position used for the fluorescence anisotropy experiments

Measure Polarization	Position Excitation Filter	Position Emission Filter
i/	90°	0°
i//	90°	90°
I/	0°	0°
I//	0°	90°

Analysis of DNA Binding Data

Total polarization was calculated using the general equation:

$P(\lambda) = (I_{//} - G \times I_{\perp}) / (I_{//} + G \times I_{\perp})$, where $P(\lambda)$ = polarization at certain emission or excitation wavelength; $I_{//}$ = excitation filter at 0° and emission filter at 90°; $G = (i_{//}) / (i_{\perp})$; I_{\perp} = excitation and emission filter at 0°; $i_{//}$ = excitation and emission filter at 90°; i_{\perp} = excitation filter at 90° and emission filter at 0° [71].

To fit the experimentally obtained polarization values from the above equation, the Graph Pad Prism software was used. To obtain the EC₅₀ or K_d values that correspond to the protein concentration at which half of the existing DNA sites are occupied by protein, I used non-linear analysis using the sigmoidal dose-response and variable-slope sigmoidal dose-response equations. The second equation evaluates the existence of cooperativity.

Results

p73 DNA-Binding Domain Purification

I purified residues 115 to 312 of the recombinant human p73 DNA-binding domain (p73DBD) by affinity and size exclusion chromatographies. Taking advantage of the poly-histidine attached to the N-terminus of the protein, I purified p73DBD with Ni-NTA affinity resin yielding milligram amounts of the target protein.

To remove any impurities of high molecular weight present after eluting it from the affinity resin with 200 mM imidazole, I concentrated the sample with centrifugal filter concentrators to a final volume of 1 ml and I injected 1 ml into a size-exclusion chromatography column. The elution profile of this column showed a peak with a voltage of 2000 mAU; it spiked at 37 ml leveling off for the following 3 ml (Figure 2.2.). I collected and concentrated the gel-filtration fractions from volumes 36 to 41 milliliters.

To assess the purity of the p73DBD sample, I ran a 15% SDS-PAGE with samples at every step of the purification protocol. Figure 2.1 depicts each of the purification steps. The final lane in the gel of Figure 2.1. corresponds to the final purified and concentrated sample of p73DBD after the last size exclusion chromatography step. The p73DBD sample migrates at the expected molecular weight of around 26 kDa. The Coomassie-stained gel of the final purification step shows a highly pure p73DBD protein sample. In order to verify the identity of the p73DBD protein and the His-tag, I carried out a western blot using an anti-His-tag antibody. As also seen in Figure 2.1., I confirmed the presence of the His-tagged target protein by the presence of identical bands at the expected molecular weight in the Coomassie-stained PAGE and the Western-blot (Figure 2.1.). Finally, I quantified the amount of protein obtained by Bradford assay, which typically corresponded to 20 to 25 mg of pure p73DBD per each 2 liters of over-expressed bacterial culture.

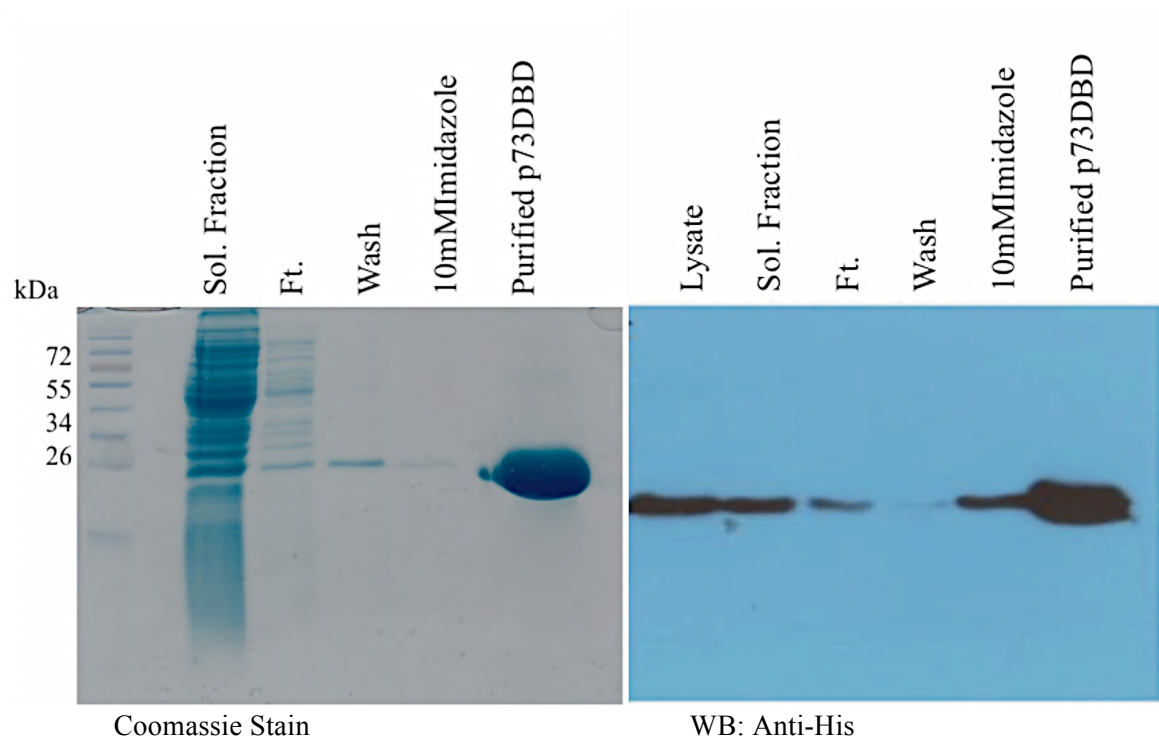


Figure 2.1. Purification of recombinant human His-tagged p73DBD using affinity and size-exclusion chromatography. Coomassie-stained SDS PAGE and western blot, showed that p73DBD migrated at 26 kDa in a 15% SDS-PAGE.

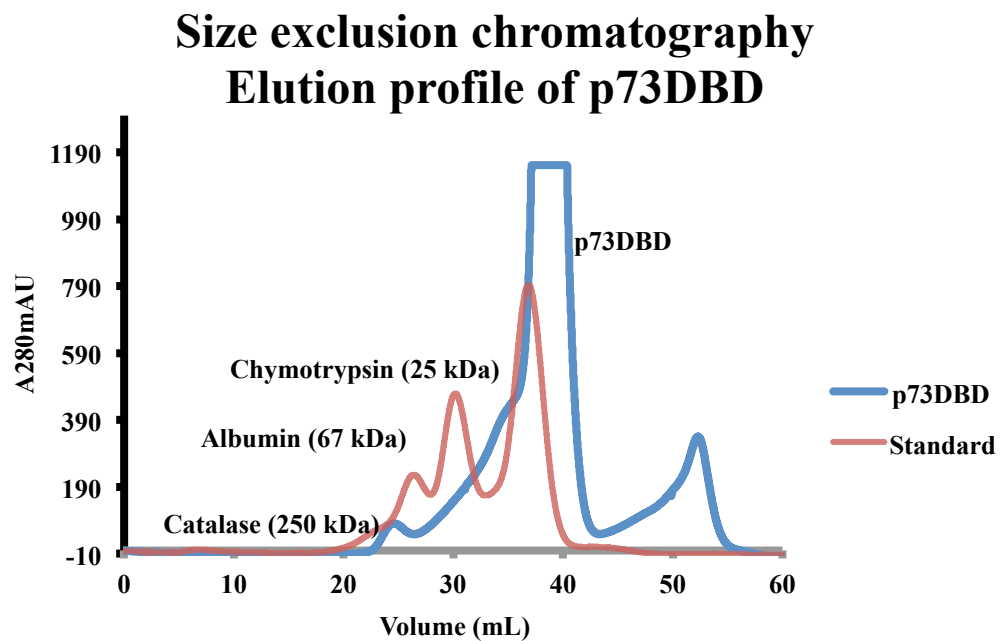


Figure 2.2. Elution profile of p73DBD. The chromatogram showed that p73DBD eluted at 37mL that is comparable to the elution profile of chymotrypsin with a molecular weight of 25 kDa.

Binding of p73DBD to half-site response elements

In order to investigate the contribution to specificity of each nucleotide in the quarter-site response element of p73, I measured, by fluorescence anisotropy experiments, the binding affinity of p73DBD to half-site response elements with variations in each position. All the dodecamer oligonucleotides used had a fluorescein molecule attached to the 5' of the oligonucleotide that allowed me to measure the changes in fluorescence polarization that occur once the protein binds the DNA.

I used the software GrapPad Prism, to calculate the affinity constants (K_d) of p73DBD for each of the studied half-site response elements. Non-linear analyses were performed to compare two different models. The first one considered a no-cooperative hill slope of 1 (dose-response equation), and the second one a cooperative hill slope different than 1 (variable-slop dose-response equation). Based on the extra-sum of squares F test method, the software chose the model that fit better the experimental data providing the best value of K_d with or without hill slope for each experimental data set. The extra-sum of squares F test method compares the improvement of the SS (sum of squares) with the more complicated model (more parameters: variable-slope sigmoidal dose- response equation) against loss of degrees of freedom. Figures 2.3. to 2.8. show the fitted p73DBD titration curves for each of the studied half-site response elements and list their corresponding DNA dissociation constant after the analysis with the GraphPad Prism software package.

The dissociation constant (K_d) for the consensus half-site response element used as a reference, quarter-site sequence: 5'-GGGCA-3', is 3.54 μ M (Figure 2.3.). I took this sequence as the reference because I knew it had the highest affinity for DNA that I had measured. By introducing individual variations to this basic response element (Table 2.1.), I was able to analyze the contribution of each position to the p73DBD DNA binding.

By changing guanine in the first, second and third position to adenine, I observed that the dissociation constant remained comparable to the value measured for 5'-GGGCA-3': 3.12, 4.5 and 4.41 μM , respectively (Figure 2.4.a., 2.5.a. and 2.6.a.). On the other hand, when I exchanged purines for pyrimidines in the first position, the binding affinity of the protein bound to the half site consensus sequence decreased by three fold: 12.82 and 13.35 μM respectively (Figure 2.4.b. and c.). A similar behavior is observed when cytosine or thymine replaced purines in the second position where the binding affinity of p73DBD to the half site sequences decreased around two fold: 8.08 and 6.45 μM , respectively (Figure 2.5.b. and c.). Moreover, in the third position, the dissociation constant increased two and three fold when pyrimidines replaced purines, 12.43 and 8.55 μM (Figure 2.6.). Overall, mutating purines to pyrimidines in the triplet flanking of the consensus sequence decreased the binding affinity of the protein to the DNA two and three fold.

The most affected position was the fourth one where replacing cytosine to adenine or guanine produced an increment of the dissociation constant of seven fold: 22.86 and 21.51 μM , (Figure 2.7.a. and 2.7.b.). Nonetheless, when thymine replaced cytosine, the binding affinity slightly decreased where the increment of the dissociation constant was two fold, approximately (Figure 2.7.c.). In the fifth position, the calculated K_d was 10.09, 11.53 and 13.55 μM upon changing adenine to cytosine, guanine and thymine, respectively (Figure 2.8.).

In summary, the p73DBD protein binds to a half site of the consensus sequence with an affinity of 3.49 μM . Changing the nucleotides at fourth position produces the highest effect in binding where the affinity of the protein to the DNA decreased about seven-fold; then, modifications in the fifth position moderately decreases the affinity as well as the mutations performed in the triplet flanking sequence where pyrimidines replaced purines. Finally, the binding affinity of p73DBD protein to a half site sequence was comparable to the reference sequence when adenine replaces guanine in the first, second and third position.

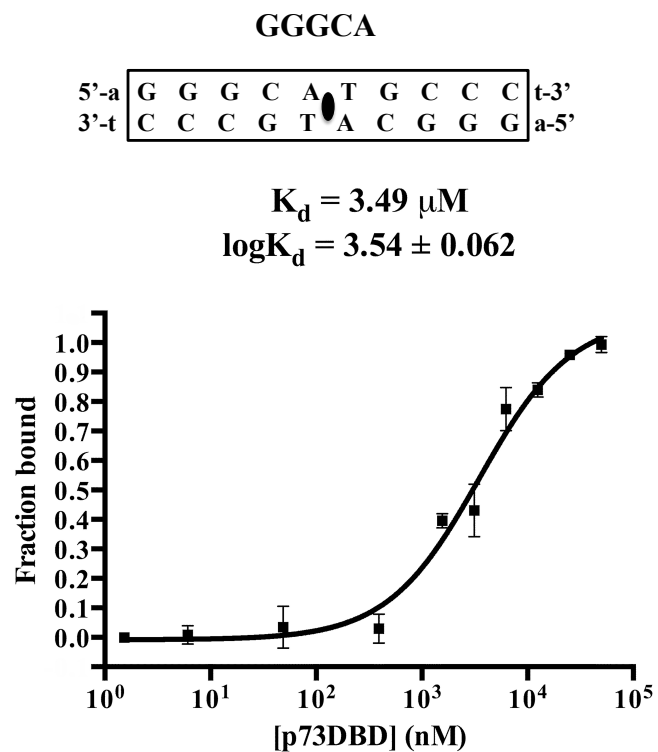


Figure 2.3. Binding affinity graph of p73DBD protein bound to half site of the consensus sequence.

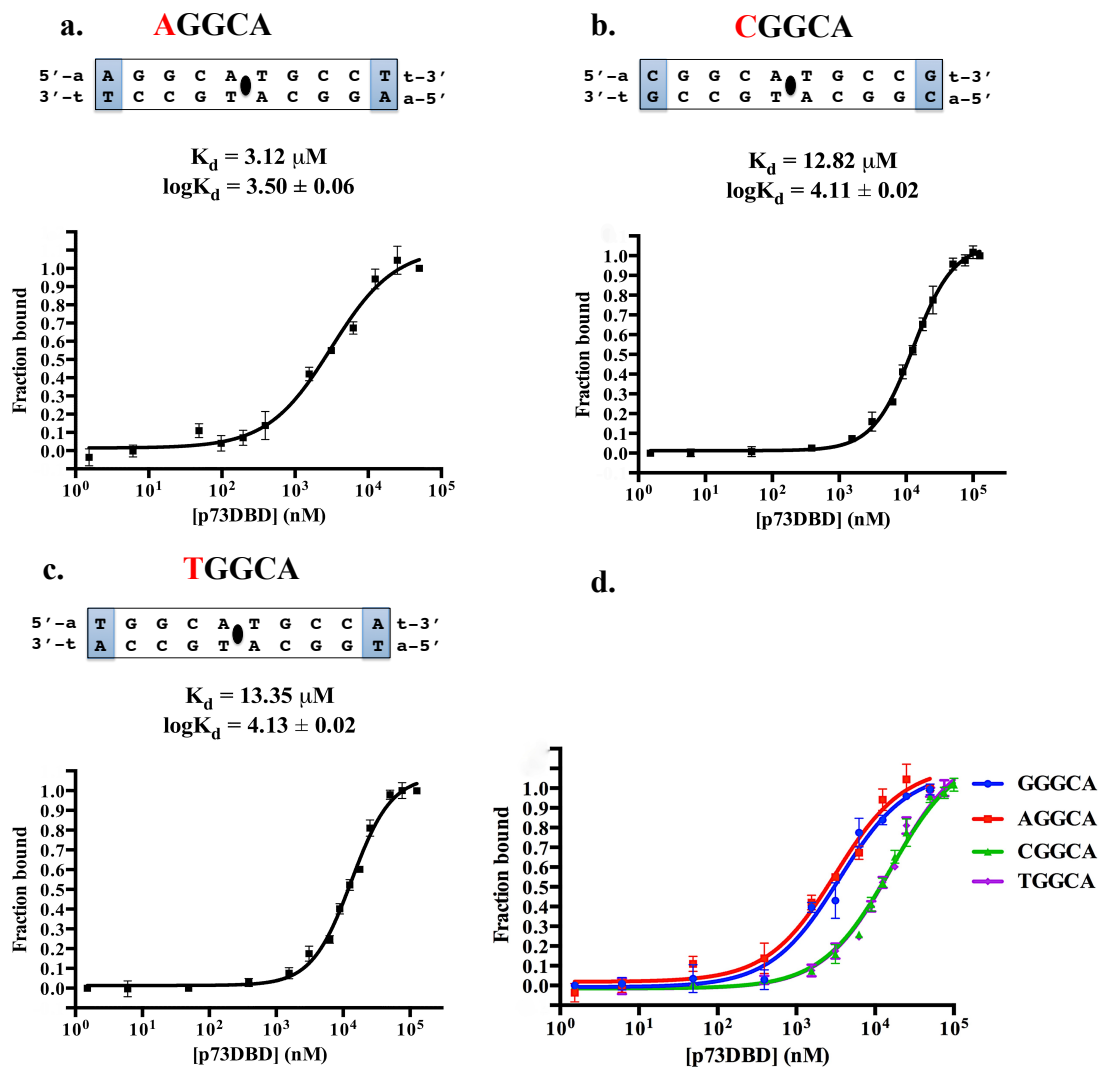


Figure 2.4. Binding affinity graphs of p73DBD bound to a half site consensus sequence with different nucleotides at position 1. Graphs a., b., and c. showed the binding affinity of p73DBD to sequences with A, C and T in the first position. d. depicted a summary of the three graphs compared with the consensus sequence (dark blue).

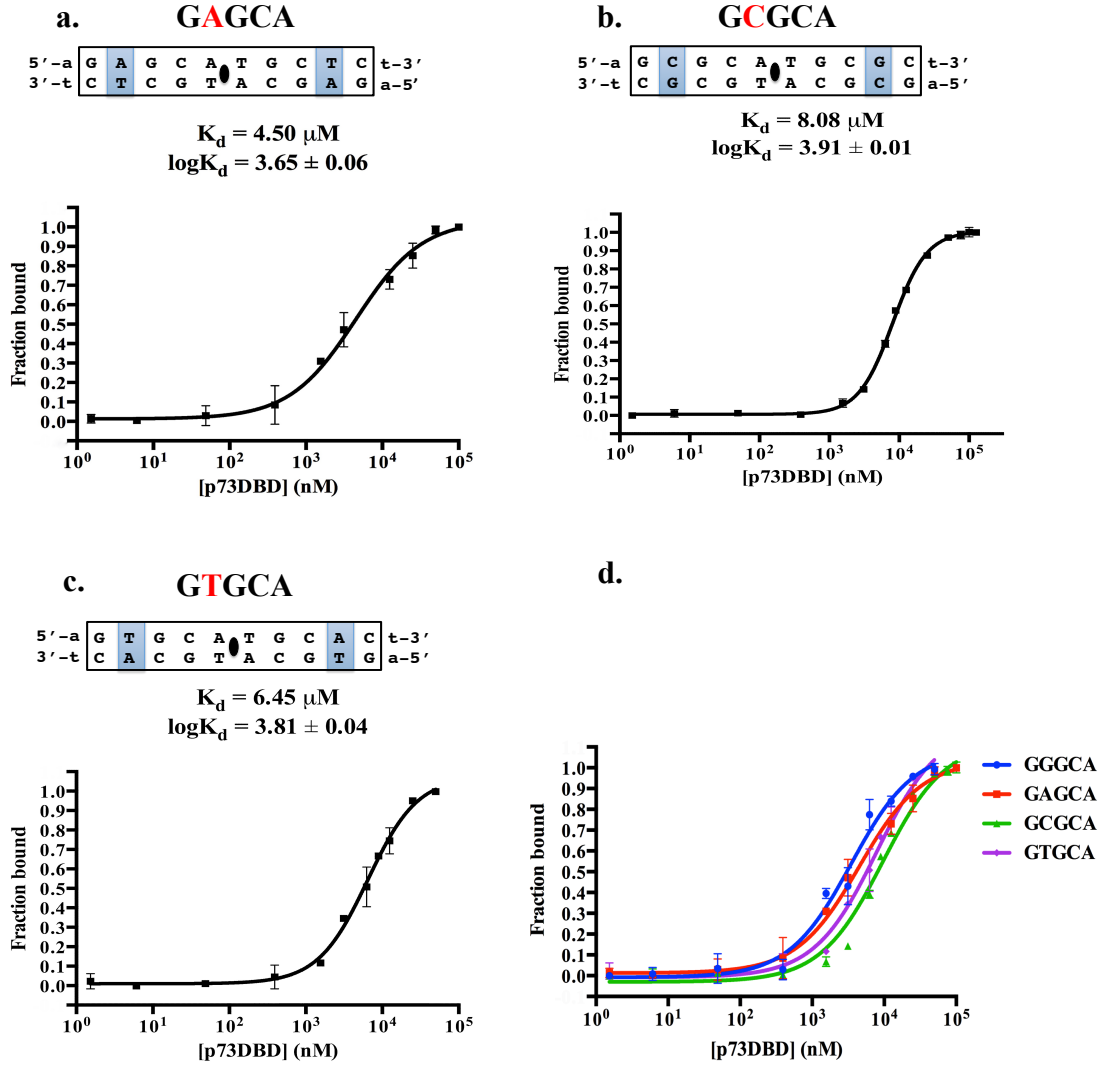


Figure 2.5. Binding affinity graphs of p73DBD bound to a half site consensus sequence with different nucleotides at position 2. a., b., and c. showed the binding affinity of p73DBD to sequences with A, C and T in the second position. d. depicted a summary of the three graphs compared with the consensus sequence (dark blue).

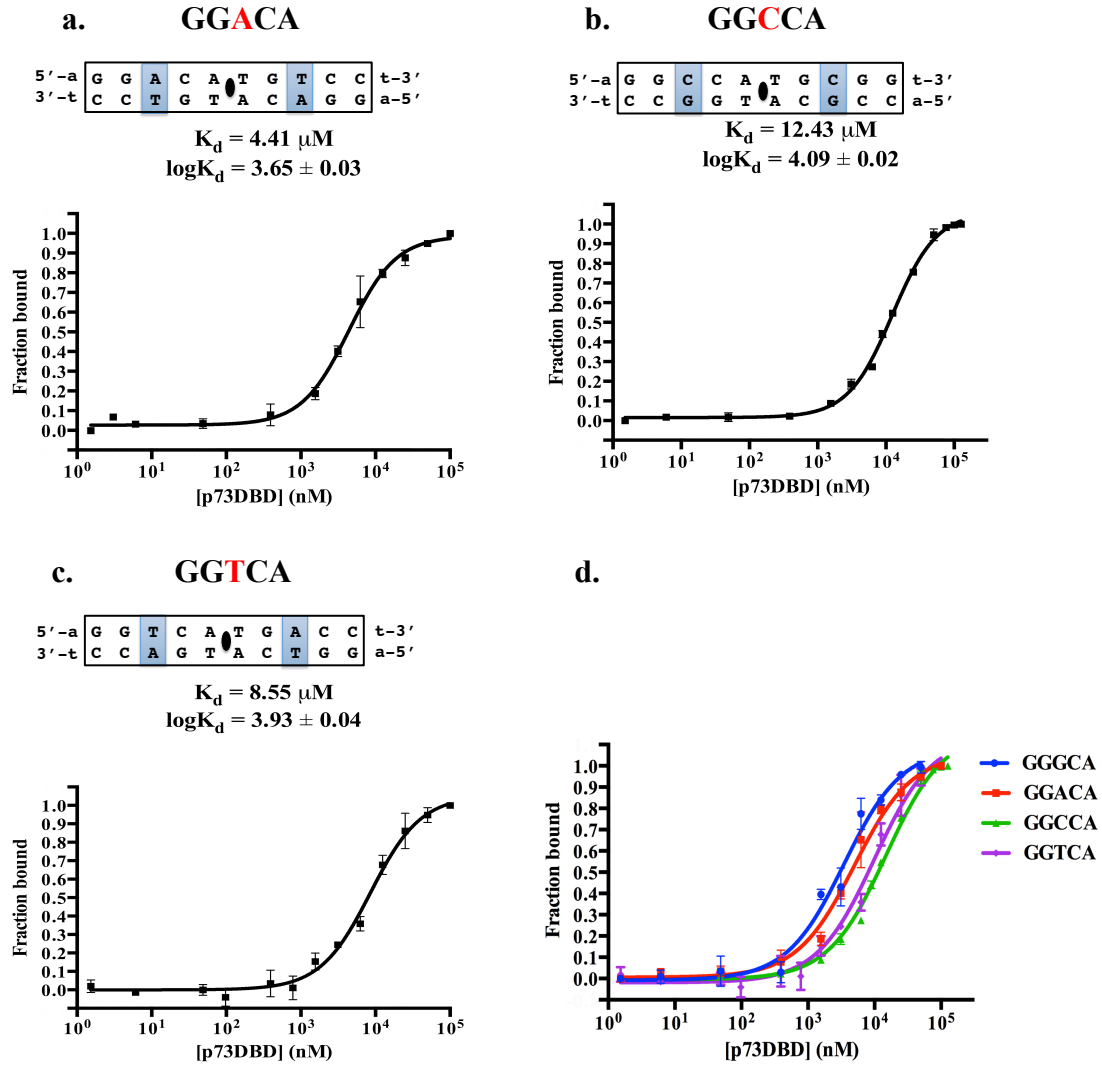


Figure 2.6. Binding affinity graphs of p73DBD bound to a half site consensus sequence with different nucleotides at position 3. a., b., and c. showed the binding affinity of p73DBD to sequences with A, C and T in the third position. d. depicted a summary of the three graphs compared with the consensus sequence (dark blue).

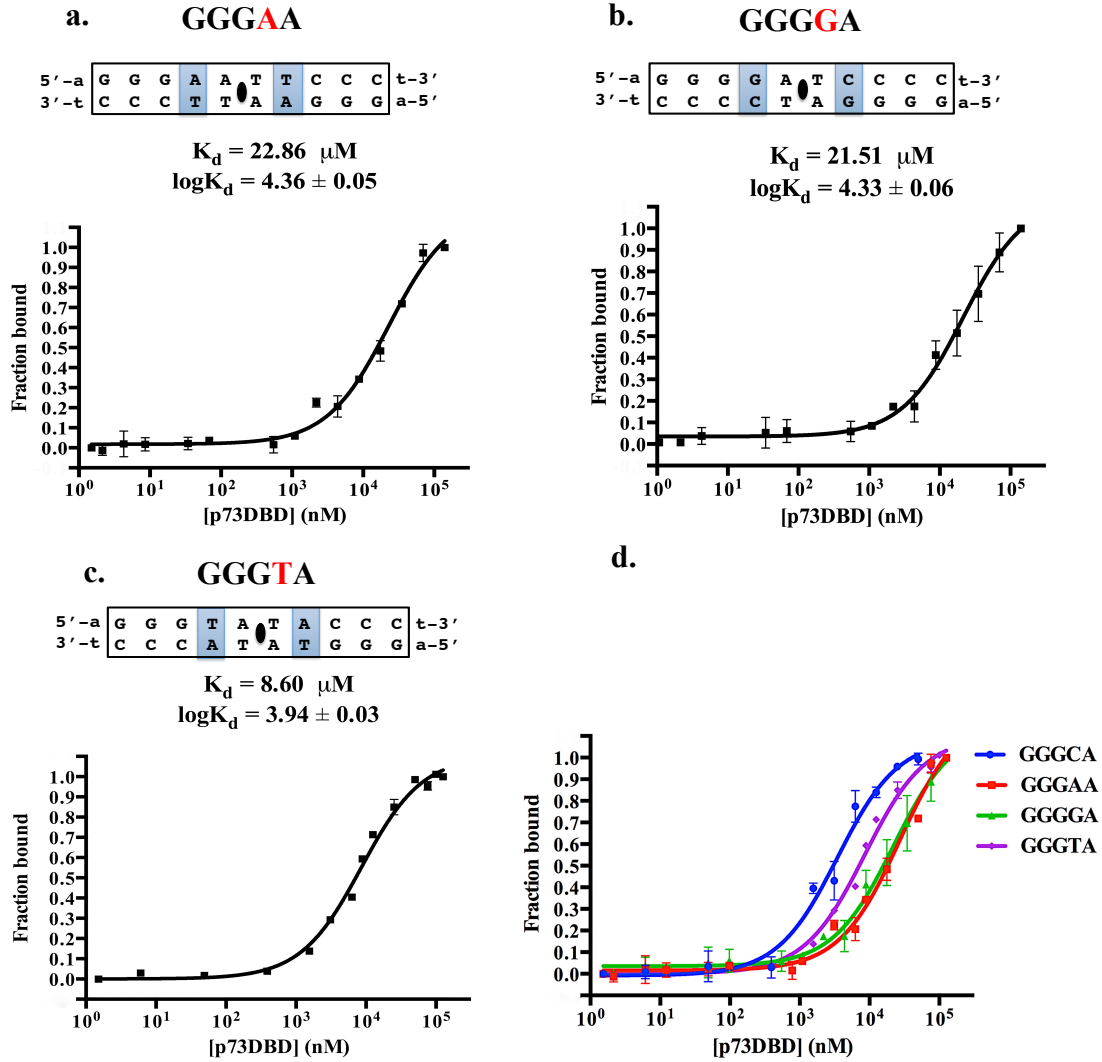


Figure 2.7. Binding affinity graphs of p73DBD bound to a half site consensus sequence with different nucleotides at position 4. a., b., and c. showed the binding affinity of p73DBD to sequences with A, G and T in the fourth position. d. depicted a summary of the three graphs compared with the consensus sequence (dark blue).

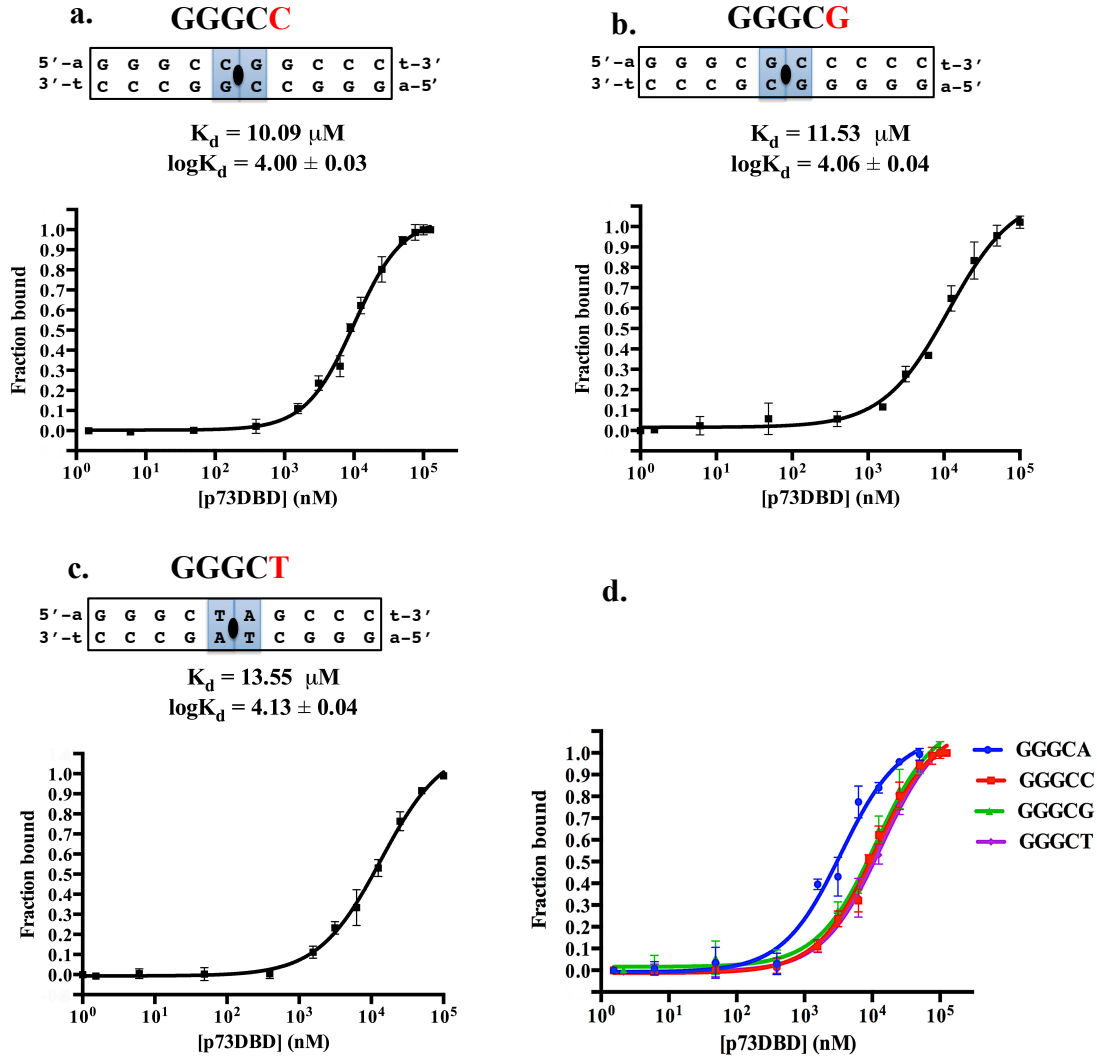


Fig 2.8. Binding affinity graphs of p73DBD bound to a half site consensus sequence with different nucleotides at position 5. a., b., and c. showed the binding affinity of p73DBD to sequences with C, G and T in the fifth position. d. depicted a summary of the three graphs compared with the consensus sequence (dark blue).

Table 2.3 Dissociation constants of p73DBD bound to the different sequences tested

Sequence	Kd (μ M)
GGGCA	3.49
First position	
A GGCA	3.12
C GGCA	12.82
T GGCA	13.35
Second position	
G A GCA	4.5
G C GCA	8.08
G T GCA	6.45
Third position	
GG A CA	4.41
GG C CA	12.43
GG T CA	8.55
Fourth position	
GGG A A	22.86
GGG G A	21.51
GGG T A	7.74
Fifth position	
GGGC C	10.09
GGGC G	11.53
GGGC T	13.55

Discussion

The protein p73DBD eluted from the gel-filtration column as a monomer as observed from the size exclusion profile. This confirmed our previous hydrodynamic experiments using analytical ultracentrifugation that showed that p73DBD is a monomer in solution [72].

From the fluorescence anisotropy experiments, pure p73DBD binds to half-site of the consensus sequence 5'-GGGCATGCCC-3' with a micromolar affinity of 3.54 μ M, which is similar to the one measured in previous studies [56].

The binding affinity of p73DBD decreased by two and three fold as compared with the 5'-GGGCATGCCC-3' consensus sequence (reference sequence) when the triplet flanking sequence was replaced by pyrimidines. The K_d remained comparable to the reference sequence when I measured the affinity of the protein to DNA sequences where adenine replaced guanine at the first, second and third position. These results are in agreement with the consensus rule – PuPuPuC(A/T)(T/A)PyPyPy- where I was expecting a lower binding of the protein to DNA when pyrimidines occupied the first three positions of the consensus sequence quarter site (Figure 2.4., 2.5. and 2.6.).

The crystal structure of p73 complexed with DNA shows that the pairing base at position 4, guanine, makes two hydrogen bonds with arginine 300 of p73; consequently, the absence of cytosine at this position would disrupt the two hydrogen bonds made by its complementary base, guanine [56]. Therefore, as expected I observed the greatest decrease in affinity or the largest K_d value when guanine or adenine substituted cytosine at the fourth position [34]. However, our data showed that thymine in the fourth position decreased the binding affinity of the protein to the DNA by only two fold (Figure 2.7.). I speculated that the presence of thymine in the fourth position could play a more noticeable role in the decrease in affinity of the protein to the DNA only when more domains of p73 are present. Furthermore, the fifth position was the second most affected with the nucleotide mutations where a three -fold decrease in affinity occurred compared

with the reference sequence (Figure 2.8.). Interestingly, I was expecting a similar binding when thymine occupied the fifth position according to the consensus rule; nevertheless, the drop in affinity upon having thymine was comparable to the one observed when cytosine was present in this position, 13.55 and 10.09 μM , respectively. This behavior could be explained by the lack of interaction of the side chain of Arg268 of p73DBD to the DNA backbone phosphate when adenine is not present in the fifth position as seen in the crystal structure of p73 DNA binding domain bound to a response element with cytosine replacing adenine at this position [73].

From these results, I conclude that this systematic approach where mutations were performed at each position of the consensus sequence half-site enables to distinguish that the p73DBD binding affinity to that sequence is determined in a hierarchical manner where the fourth position plays the most important role in the binding affinity, followed by the fifth position and lastly, the triplet flanking sequence as seen in Table 2.3.

Chapter 2, in full, is contained in the manuscript named Structural Determinants of DNA Binding Specificity in the Transcription Factor p73: a Hierarchical Model of Response Element Recognition. Ramos, Ana and Viadiu, Hector. The dissertation/thesis author is the primary investigator and author of this paper.

Chapter 3

Response Element Specificity

of the $\Delta Np73\delta$ Isoform

Introduction

The p53 tumor suppressor protein is an oligomer in the absence of DNA. The domain responsible for its oligomerization in solution was determined to be in the C-terminus comprising residues 311 to 363 [74][75][76].

In p53, the oligomerization domain plays an important role as an enhancer of the interaction with DNA by increasing the binding affinity of the protein to its response elements by 10 to 100 times in comparison with the p53 DNA binding domain by itself [39]. Another function of the oligomerization domain in p53 is its ability to promote conformational changes in the p53 oligomers as well as to increase DNA bending [69] [77]. Finally, the p53 oligomerization domain might serve as a scaffold for protein-protein interactions [78]

The oligomerization domain of p53 shares a sequence identity of 40% with its paralog p73 [79]. Moreover, p73 oligomerization domain forms a tetramer described as a dimer of dimers as in p53 oligomerization domain [80]. Despite the sequence similarity with p53, p73 oligomerization domain contains an extra alpha-helix comprising residues 383 to 395. The extra alpha-helix is important for tetramers formation and stability and for cell-cycle arrest transcriptional activity of p73. Moreover, the extra helix allows the p73 oligomerization domain to form hetero-oligomers with p63 oligomerization domain that also has a conserved extra alpha-helix [81] [80].

Similarly to p53, p73 requires the oligomerization domain to function as a transcription factor [82], and as stabilizer of the overall p73 tetramer [80]

Although plenty information is known about the structure and function of the oligomerization domain of p73, little is known about its effect on DNA binding specificity. The only data available show that p73 isoform, $\Delta Np73\delta$, can distinguish between two different full-site consensus sequences (GGGCA or GAACA in the pentamer repeat) by a change in the cooperative behavior and in the binding affinity [68]. Hence, in this chapter, I will study the

binding affinity of p73 DNA binding domain together with the oligomerization domain using the biological relevant isoform $\Delta Np73\delta$ to elucidate the function of the oligomerization domain in the context of the DNA-binding domain.

Materials and Methods

Subcloning of Δ Np73 δ

A DNA fragment codifying for residues 115 to 402 of the human p73 gene was cloned, using *Eco*RI and *Hind*III restriction sites, into the protein over-expression vector pET28 with an N-terminal His-tag. This gene fragment codifies for the Δ Np73 δ isoform of the human p73 gene. The accuracy of the DNA sequence in the modified vector was verified by sequencing the vector region carrying the inserted gene. Throughout this thesis, I will refer to the gene product of such protein over-expression construct as the Δ Np73 δ protein. The pET28 vector containing the Δ Np73 δ gene was used to transform BL21/DE-3 *E. coli* competent cells using the kanamycin-resistance gene codified in the pET28 vector as a selection marker. To have a permanent frozen stock of cells able to over-express the Δ Np73 δ protein, cells were grown in LB medium with 30 μ g/mL of kanamycin at 37°C and 1 ml stocks were frozen at -80°C with a 50% final glycerol concentration.

Protein Expression and Purification

To over-express the Δ Np73 δ protein, a 1 μ l aliquot of frozen cells was grown overnight in 5 ml LB medium with 30 μ g/ml of kanamycin at 37°C. After the culture grew overnight to saturation, it was used to start a 1 lt LB culture. The regular overexpression protocol was carried out with 4 lts of LB media. Upon reaching an absorbance of 0.6 AU at 600 nm, cells were induced with a final 0.5 mM IPTG concentration and grown at 25°C for 6 to 8 hours. After induction, cells were harvested by centrifugation at $2,845.9 \times g$ for 30 min at 5°C. The supernatant was discarded and the cells in the pellet were resuspended in lysis buffer with 1.5 M sodium chloride, 50 mM Tris (pH 8.0). Cells were lysed using a microfluidizer and centrifuged at $104,443.6 \times g$ to remove the cell debris from the cytoplasmic fraction. The Δ Np73 δ protein was purified using affinity chromatography by incubating overnight the soluble fraction from the

centrifugation step with 1 ml of nickel-nitrilotriacetic acid (Ni-NTA) resin and 20 mM of imidazole at 5°C.

To remove the non-specifically-bound contaminant proteins to the nickel resin, the resin was washed with 100 ml of lysis buffer and 150 ml of the same lysis buffer supplemented with 60 mM imidazole. Finally, the protein was eluted with lysis buffer supplemented with 500 mM imidazole. Then, the protein was concentrated to a volume of 2 ml using a centrifugal filter concentrator. To further purify the protein, one milliliter was loaded into a Superdex 200 gel filtration column equilibrated with binding buffer (100 mM sodium chloride, 10 mM sodium citrate (pH 6.1), 5 mM DTT and 5 μ M zinc chloride). The degree of purity was determined by running a 10% SDS-PAGE and the identity of the protein was confirmed by a western blot using an anti His-tag antibody from Roche.

Sedimentation Velocity Experiments

Sedimentation velocity experiments were performed in binding buffer (100 mM sodium chloride, 10 mM sodium citrate (pH 6.1), 5 mM DTT and 5 μ M zinc chloride) using a Beckman XL-1 analytical ultracentrifuge with an AnTi60 rotor. The experiment used 400 μ l of buffer containing Δ Np73 δ protein at a concentration of 64.5 μ M and 400 μ l of buffer without the protein as a reference. Sample and buffer were loaded into separate compartments of a double sector centerpiece. All the experiments performed were carried out at 20°C and at a speed of 201240 \times g. To investigate the oligomerization state of the protein in solution without DNA, radial scans were collected at 280 nm. On the other hand, oligonucleotides with a fluorescein molecule attached to their 5' allowed us to follow the sedimentation of the protein-DNA complex where the concentration of the protein was 64.5 μ M and the DNA 3.4 μ M. To assess the oligomerization of Δ Np73 δ bound to half and full-site response elements, radial scans measuring absorption at the fluorescein absorbance maximum of 488 nm were collected.

Sedimentation Velocity Analysis

To obtain the sedimentation coefficients of the species present in the sample, the data were analyzed using a continuous $c(s)$ distribution model with the SEDFIT software [83]. The analysis requires the partial specific volume of the protein, the buffer viscosity, and the buffer density, which were calculated in SEDNTRERP [84]. The molecular weight of the species in the sample with unique sedimentation and diffusion coefficients was estimated in SEDFIT using the equation:

$$M = \frac{sRT}{D(s)(1-\rho\tilde{v})} \quad (2)$$

where M = molecular weight in Da; s = sedimentation coefficient; R = gas constant; T = temperature; D = diffusion coefficient; ρ = density of the buffer; and \tilde{v} = partial specific volume calculated based on the protein sequence. Two assumptions are made when applying this formula. The first one states that the diffusion coefficient, D , is a function of the sedimentation coefficient, s , and the second one relates to the fact that all species in the sample have a unique weight-average frictional ratio, $(f/f_0)_w$, [83]. The weight-average frictional ratio is iteratively determined during the continuous distribution analysis by non-linear regression; and the calculated value is used to estimate the diffusion coefficient D :

$$D(s) = \frac{\sqrt{2}}{18\pi} kT s^{-1/2} \left(\eta \left(\frac{f}{f_0} \right)_w \right)^{-3/2} \left(\frac{(1-\rho\tilde{v})}{\tilde{v}} \right)^{1/2} \quad (3)$$

where k = Boltzman constant; T = temperature; S = sedimentation coefficient; η = viscosity; $(f/f_0)_w$ = weight-average frictional ratio; ρ = the density of the buffer and \tilde{v} = the partial specific volume.

Therefore, the oligomeric state of the protein can be estimated by comparing the molecular weight of the monomer (calculated from the sequence) with the molecular weight obtained from equation 2, which SEDFIT calculates automatically.

DNA-Binding Fluorescence Polarization Assays

To study the $\Delta Np73\delta$ protein affinity towards different half-site response elements, a set of 12 and 22 bp 5'-fluorescein-labeled oligonucleotides were acquired from Integrated DNA Technologies (IDT). Table 3.1 lists the seventeen 12 bp half-site oligonucleotides used, and Table 3.2 lists the two 22 bp full-site oligonucleotides used. To study the binding of $\Delta Np73DBD\delta$ to half-site response elements, I used the same experimental approach as in Chapter 2 to study the DNA specificity of the p73 DBD. Moreover, I used the same sequences where the consensus 5'-GGGCATGCCC-3' half-site was used as a reference and single mutations were introduced in each quarter-site to explore the contribution to binding of each nucleotide position. I numbered each position of the quarter site from the 5' end to the center of the half-site sequence, as I did for the p73DBD binding experiments (Table 3.1.).

Table 3.1. List of the 5'-fluorescein-labeled (FAM) double-stranded-DNA dodecamers used for the fluorescence anisotropy experiments

Name-ID	Sequence
12345 GGGCA	1234554321 5'-FAM-aGGGCATGCCct-3' 3'-tCCCGTACGGGa-FAM-5'
12345 AGGCA	1234554321 5'-FAM-aAGGCATGCCtT-3' 3'-tTCCGTACGGaA-FAM-5'
12345 CGGCA	1234554321 5'-FAM-aCGGCATGCCGt-3' 3'-tGCCGTACGGCa-FAM-5'
12345 TGGCA	1234554321 5'-FAM-aTGGCATGCCAt-3' 3'-tACCGTACGGTa-FAM-5'
12345 GAGCA	1234554321 5'-FAM-aGAGCATGCTCt-3' 3'-tCTCGTACGAGa-FAM-5'
12345 GCGCA	1234554321 5'-FAM-aGCGCATGCGCt-3' 3'-tCGCGTACGCGa-FAM-5'
12345 GTGCA	1234554321 5'-FAM-aGTGCATGCAct-3' 3'-tCAGCTACGTGa-FAM-5'
12345 GGACA	1234554321 5'-FAM-aGGAACATGTCct-3' 3'-tCCGTGTACAGGa-FAM-5'
12345 GGCCA	1234554321 5'-FAM-aGGCCATGGCct-3' 3'-tCCGGTACCGGa-FAM-5'
12345 GGTCA	1234554321 5'-FAM-aGGTCATGACct-3' 3'-tCCAGTACTGGa-FAM-5'
12345 GGGA	1234554321 5'-FAM-aGGGAATTCct-3' 3'-tCCCTTAAGGGa-FAM-5'
12345 GGGGA	1234554321 5'-FAM-aGGGGATCCCCt-3' 3'-tCCCCTAGGGGa-FAM-5'
12345 GGGTA	1234554321 5'-FAM-aGGGTATACCCt-3' 3'-tCCCAATATGGGa-FAM-5'
12345 GGGC	1234554321 5'-FAM-aGGGCGGCCct-3' 3'-tCCCGCGCGGGa-FAM-5'
12345 GGGCG	1234554321 5'-FAM-aGGGCGCGGCCct-3' 3'-tCCCGCGCGGGa-FAM-5'
12345 GGGCT	1234554321 5'-FAM-aGGGCTAGCCCt-3' 3'-tCCCGATCGGGa-FAM-5'

Table 3.2. List of the double-stranded-DNA 5'-fluorescein-labeled (FAM) 22 mers used for the fluorescence anisotropy experiments

Name-ID	Sequence
12345	5' -FAM-aGGGCATGCCCCGGGCATGCCCt-3'
GGGCA-20	3' -tCCCGTACGGGCCCGTACGGGa-FAM-5'
12345	5' -FAM-aGAA CATGTTTCGAA CATGTTTct-3'
GAA CA-20	3' -tCTTGTACAAGCTTGTACAAGa-FAM-5'

To measure the DNA binding constant of the $\Delta Np73\delta$ protein for each half-site response element, I prepared 15 to 17 samples of the $\Delta Np73\delta$ protein going from 40 μM to 2 nM. Each sample with a unique protein concentration was mixed with 50 nM of each 5'-fluorescein-labeled dsDNA, either a 12-mer or a 22-mer. The final volume for each assay was adjusted to 350 μL with binding buffer (100 mM sodium chloride, 20 mM sodium citrate (pH 6.1), 5 mM DTT, and 5 μM zinc chloride). The same binding buffer was used to prepare the protein stock and its dilutions, thus I had identical ionic conditions in each assay.

Before measuring the binding constants by fluorescence polarization, all the mixed protein-DNA final solutions were incubated for at least 4 hours at room temperature to reach its equilibrium binding. The fluorescence polarization measurements were performed in the same manner as I described in Chapter 2. Briefly, for each protein concentration, four fluorescence intensity measurements were done, each with a unique filter arrangement. The maxima excitation and emission wavelengths of the fluorescein molecule were used, 494 and 521 nm, respectively. Measurements were done with a Hitachi F-2000 fluorescence spectrometer. For each half-site response element studied, the binding experiment was repeated three times.

DNA-Binding Fluorescence Polarization Analysis

The GraphPad Prism software was used to fit the binding data to two non-linear equations, one without cooperativity and one accounting cooperativity. Both, a sigmoidal-dose-response and a variable-slope sigmoidal-dose-response equations were used to estimate the EC_{50} or K_d values that correspond to the protein concentration at which half of the protein present were bound to DNA. A more detailed explanation of the binding data analysis used can be found in Chapter 2.

Results

Δ Np73 δ purification

The Δ Np73 δ isoform construct that I purified derives from residues 115 to 402 of the full-length human p73 transcription factor. This construct has the DNA-binding domain (DBD), the nuclear localization signal (NLS), the oligomerization domain (OD), and five extra residues that result from the alternative splicing of the C-terminus in the delta (δ) isoform (P-T-W-G-P) (Figure 3.1.).

After overexpression, cell harvesting and cell rupture, I proceeded to purify the protein as mentioned in the Methods section. The purification protocol consisted of two chromatographic steps: affinity and size exclusion. The His-tag in the N-terminus of the Δ Np73 δ construct permitted the use of affinity purification with Ni-nitrilotriacetic acid (Ni-NTA) resin. After the loading and washing steps, the protein was eluted from the column with 500 mM of imidazole. Using filter concentrators, the sample was concentrated to a volume of 2 ml. Then, successive size exclusion chromatography runs with injections of 1 ml were performed. In each run, the Δ Np73 δ protein peak eluted in the fractions between 26 and 31 milliliters, which were collected and concentrated (Figure 3.3.).

To assess the purity of Δ Np73DBD δ sample, I ran a 10% SDS-PAGE with the samples of every step in the purification protocol (Figure 3.2.). The last lane in the gel of Figure 3.2. corresponds to the final purified and concentrated Δ Np73 δ protein after the final size exclusion chromatography step. The Δ Np73 δ sample migrated at the expected molecular weight of 34 kDa.

I further verified the identity of the purified His-tagged Δ Np73 δ protein by carrying out a western blot with an anti-His-tag antibody. As seen in Figure 3.2, I confirmed the presence of the His-tagged Δ Np73 δ protein by the presence of a band at an identical position that the Coomassie-stained SDS-PAGE band.

Finally, protein concentration was quantified by the Bradford assay. Typically, 2 liters of overexpressed bacterial culture yielded between 10 to 13 milligrams of pure Δ Np73 δ protein.

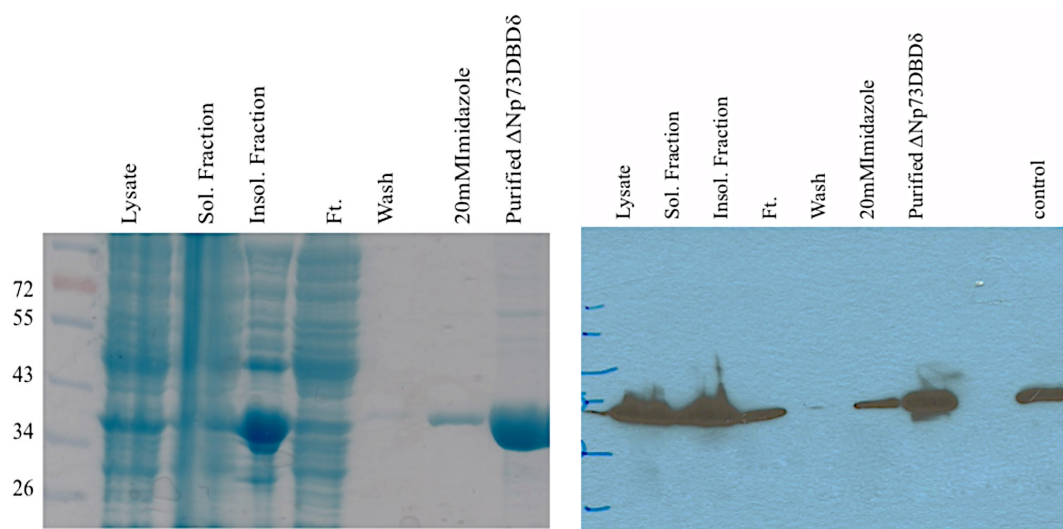
```

115  I P S N T D Y P G P H H F E V T F Q Q S S T A K S A T W T Y S P L L K K
151  L Y C Q I A K T C P I Q I K V S T P P P G T A I R A M P V Y K K A E H V
188  T D V V K R C P N H E L G R D F N E G Q S A P A S H L I R V E G N N L S
224  Q Y V D D P V T G R Q S V V V P Y E P P Q V G T E F T T I L Y N F M C N
260  S S C V G G M N R R P I L I I T L E M R D G Q V L G R R S F E G R I C A
297  C P G R D R K A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K
332  Q S P P A V P A L G A G V K K R R H G D E D T Y Y L Q V R G R E N F E I
368  L M K L K E S L E L M E L V P Q P L V D S Y R Q Q Q Q L L Q R P T W G P

```

DNA binding domain (DBD) residues 115-312	Nuclear localization signal (NLS) residues 327-347	Oligomerization domain (OD) residues 351-399	Delta end (δ) residues 400-404
---	--	--	---

Figure 3.1. The Δ Np73 δ isoform containing the DNA binding domain, the nuclear localization signal, the oligomerization domain and the last five residues in the splicing of the C-terminus of the delta (δ) isoform (P-T-W-G-P) used in the experiments



Coomassie stain

WB: antihis

Figure 3.2. Purification of recombinant human His-tagged Δ Np73 δ using affinity and size exclusion chromatography. SDS PAGE analysis, Coomassie stained SDS PAGE (left) and western blot (right), showed that Δ Np73 δ migrated at 34 kDa in a 10% SDS PAGE.

Size exclusion chromatography Elution profile Δ Np73 δ

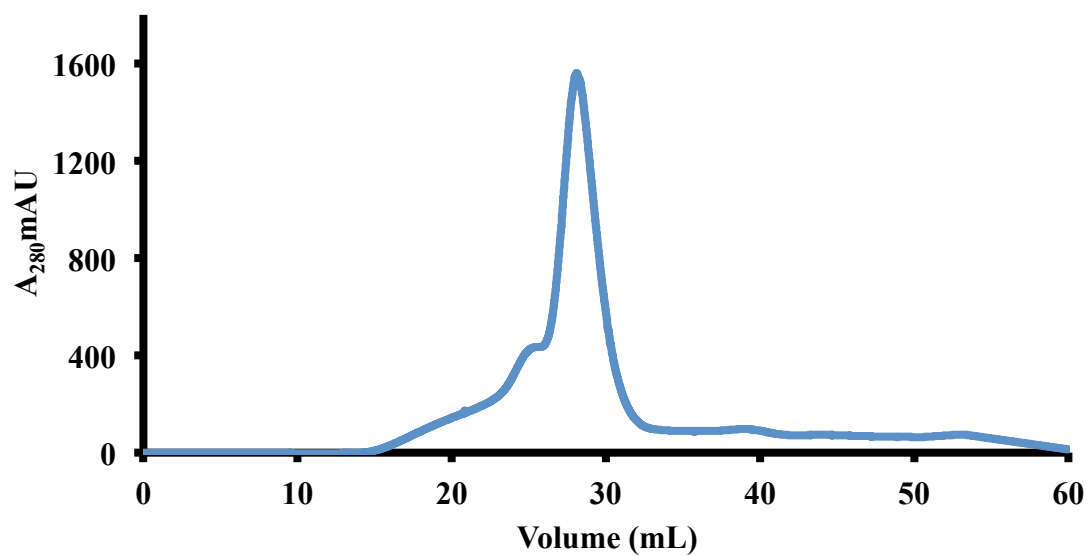


Figure 3.3. Elution profile of Δ Np73 δ in a size-exclusion column. The chromatogram showed the retention volume of Δ Np73 δ to be 26 ml.

Oligomerization State of $\Delta\text{Np73}\delta$ Isoform

To investigate the oligomerization state of the $\Delta\text{Np73}\delta$ isoform, I measured its sedimentation coefficient in sedimentation velocity experiments in the absence and presence of DNA. According to the analysis of the hydrodynamic experiments, when the $\Delta\text{Np73}\delta$ protein is in solution in the absence of DNA, its sedimentation coefficient is 4.94 S. The sedimentation and diffusion coefficients calculated, using the software SEDFIT, as solutions to the Lamm equation suggest that the $\Delta\text{Np73}\delta$ protein in the absence of DNA in solution is a dimer. When I ran identical sedimentation velocity experiments with the $\Delta\text{Np73}\delta$ protein in the presence of a 10 bp half-site response element, the determined sedimentation coefficient increased to 6.06 S. Thus, a dimer of $\Delta\text{Np73}\delta$ binds to a half-site response element. Finally, when the experiments were carried out with a 20 bp full-site response element, the sedimentation coefficient was 7.87 S. The solution of the Lamm equation, using the estimated sedimentation and diffusion coefficients, indicates that the $\Delta\text{Np73}\delta$ protein is a tetramer when bound to a full-site response element (Figure 3.4.).

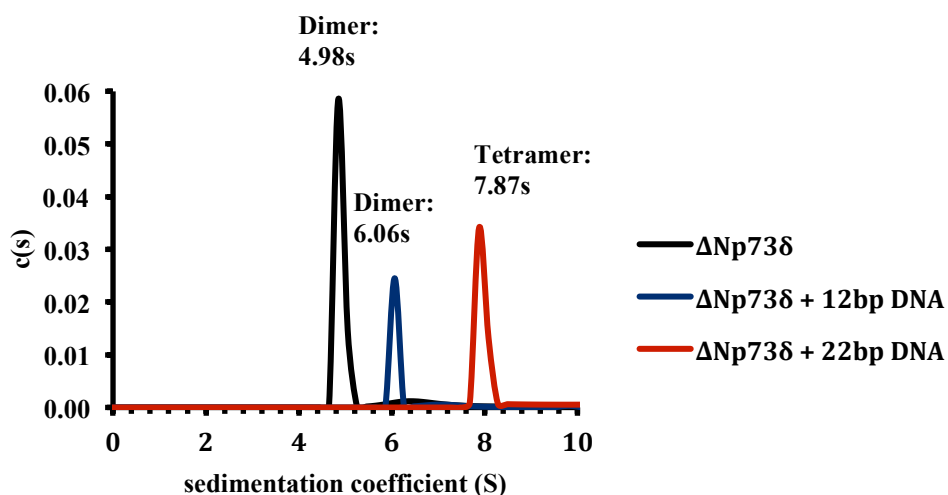


Figure 3.4. Sedimentation coefficient profile of $\Delta\text{Np73}\delta$ in solution unbound and bound to half-site and full-site response elements.

Binding of Δ Np73 δ to Half-Site Response Elements

The goal of our work with the Δ Np73 δ isoform has been to determine the role of the C-terminus of this isoform, particularly its oligomerization domain, in response element recognition. As I had done for p73 DBD, I carried out equivalent binding affinity assays and oligomerization state experiments for the Δ Np73 δ isoform. In Chapter 2, I reported how p73 DBD DNA binding changes when bound to half-site response elements with variations at different positions. The experiments reported in this chapter for the Δ Np73 δ isoform follow the same logic than the ones reported in the previous chapter for the p73 DBD construct.

In the binding affinity studies, I used the oligonucleotide 5'-tGGGCATGCCCa-3' as a reference sequence. The p73 DBD had shown a higher affinity for the reference sequence than for other sequences [56]. I named the reference sequence as GGGCA to highlight the two inverted-repeat quarter-sites that form the half-site response element contained in the oligonucleotide. I followed the same nomenclature for the rest of half-sites studied (Table 3.1.). The K_d of the Δ Np73 δ isoform towards the reference GGGCA half-site was 171.7 nM, which was the highest affinity obtained in all the experiments that I carried out with the Δ Np73 δ isoform (Figure 3.5.).

As I had done for the p73 DBD protein, I systematically modified every nucleotide position in the reference half-site response element. My aim was to understand how dependent Δ Np73 δ recognition is to sequence changes of the response element. I report the binding data for half-sites with modifications in each of the five positions of the quarter-sites. For modifications in the first position, the dissociation constants of the AGGCA, CGGCA and TGGCA half-sites were 199.4 nM, 1060 nM and 1149 nM, respectively (Figure 3.6.). When the first guanine in each quarter-site was replaced for an adenine, the binding affinity of the Δ Np73 δ protein was similar to the reference sequence. Instead, when the nucleotide was changed to either cytosine or thymine, the DNA affinity of the Δ Np73 δ isoform dropped almost ten times. For response elements with modifications in the second position of each quarter-site, changing guanine to adenine slightly

increased the K_d from 171.7 nM to 875.9 nM. On the other hand, when a cytosine or a thymine replaced the second guanine, the DNA binding affinity of the $\Delta Np73\delta$ isoform decreased by 70 and 20 times to have dissociation constants of 11972 and 2990 nM, respectively (Figure 3.7.). The third position was the least affected by the substitution of guanine in the reference sequence to adenine, cytosine or thymine. When guanine was substituted by an adenine the K_d was 581.4 nM, 379.3 nM when modified to a cytosine, and 365.8 nM when changed to a thymine (Figure 3.8.). Changes in the fourth position showed the largest changes in DNA affinity, together with changes in position two. When the cytosine in the reference sequence was changed by guanine or thymine, the K_d values were 5285 and 12243 nM, respectively. Nonetheless, when adenine substituted cytosine, the 171.7 to 890.8 nM decrease in DNA binding affinity of the $\Delta Np73\delta$ isoform was less drastic than for the guanine and thymine substitutions (Figure 3.9.). Lastly, in the position five of the quarter-site, all the substitutions showed a decrement in the DNA affinity of $\Delta Np73\delta$. The dissociation constants when cytosine, guanine or thymine replaced adenine were 2022 nM, 1288 nM and 1701 nM, respectively (Figure 3.10.). The overall binding profile of $\Delta Np73\delta$ obtained from this set of experiments is shown in Figure 3.14.

To explore the effect of having a full-site response element, instead of the minimal half-site, I measured the DNA binding affinity of $\Delta Np73\delta$ for two full-site response elements with a sequence similar to our half-site reference sequence. I called such full-site response elements GGGCA-20 and GAACA-20 (Table 3.2.). The binding affinities of $\Delta Np73\delta$ to the GGGCA-20 and GAACA-20 full-sites were 163.6 nM and 297.8 nM, respectively. These values were roughly comparable to the affinities measured for the GGGCA and GAACA half-sites of 171.7 and 180.6 nM (Figure 3.11.).

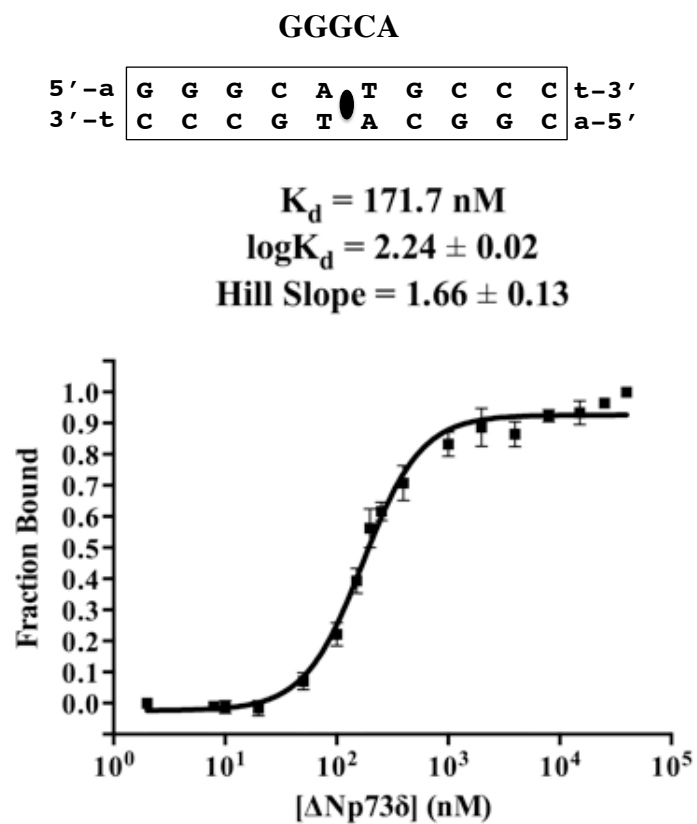


Figure 3.5. Binding affinity graph of $\Delta\text{Np73}\delta$ bound to the reference half-site response element.

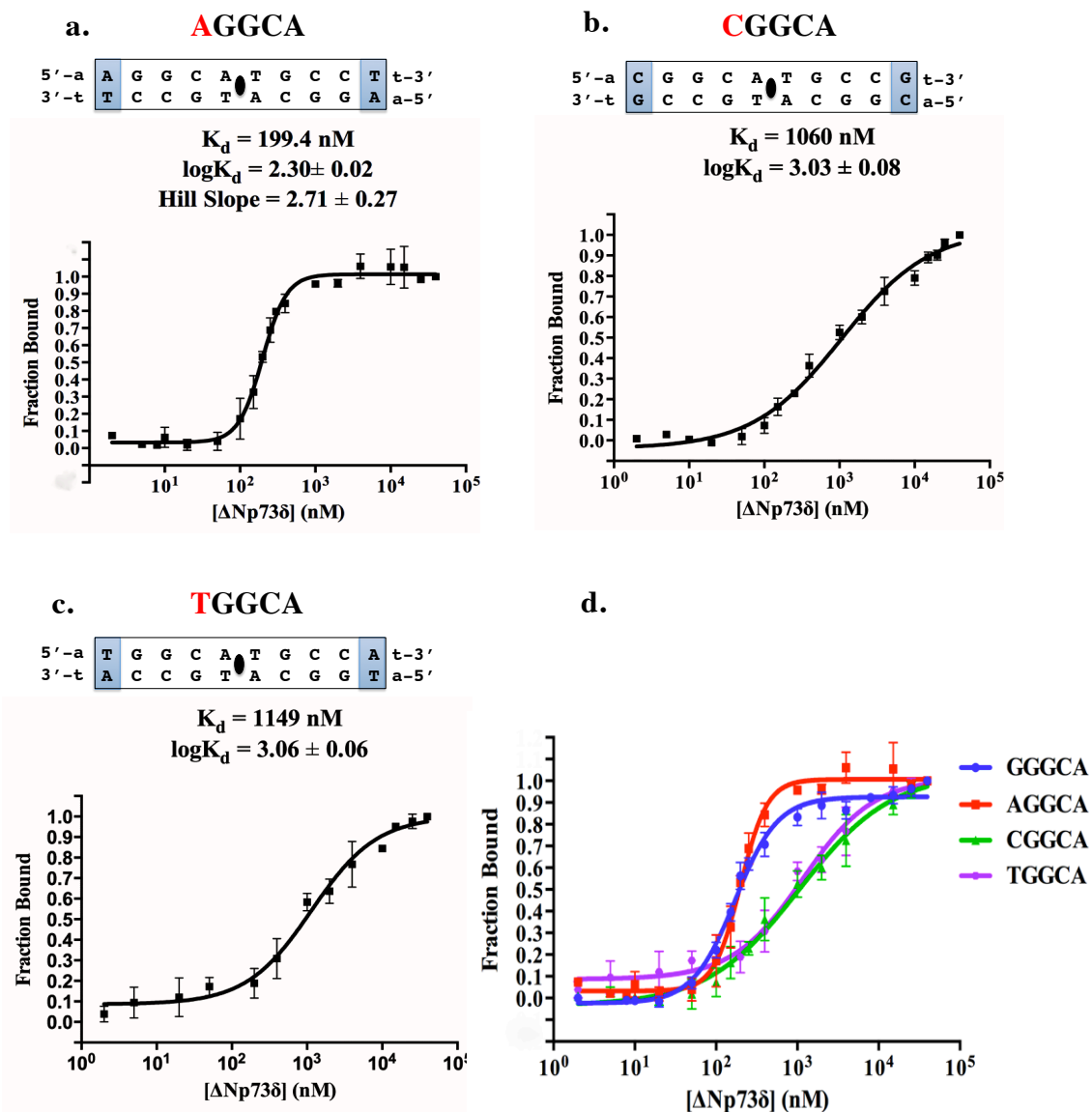


Figure 3.6. Binding affinity graphs of $\Delta Np73\delta$ bound to a half-site response element with different nucleotides at position 1. Graphs a., b., and c. showed the binding affinity of $\Delta Np73\delta$ to sequences with A, C and T in the first position. Graph d. depicted a summary of the three graphs compared with the reference half-site sequence (dark blue).

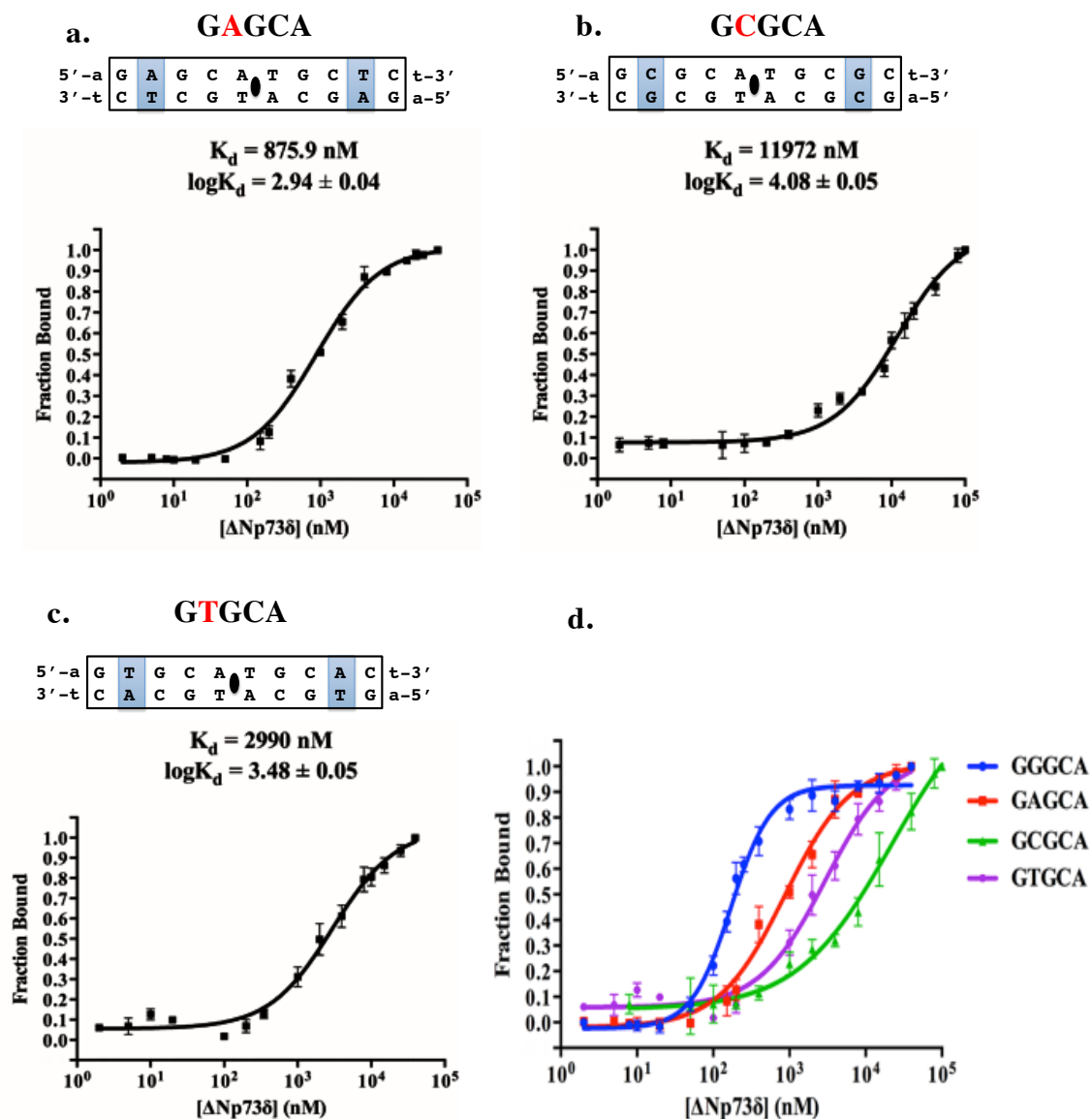


Figure 3.7. Binding affinity graphs of $\Delta\text{Np73}\delta$ bound to a half-site response element with different nucleotides at position 2. Graphs a., b., and c. showed the binding affinity of $\Delta\text{Np73}\delta$ to sequences with A, C and T in the second position. Graph d. depicted a summary of the three graphs compared with the reference half-site sequence (dark blue).

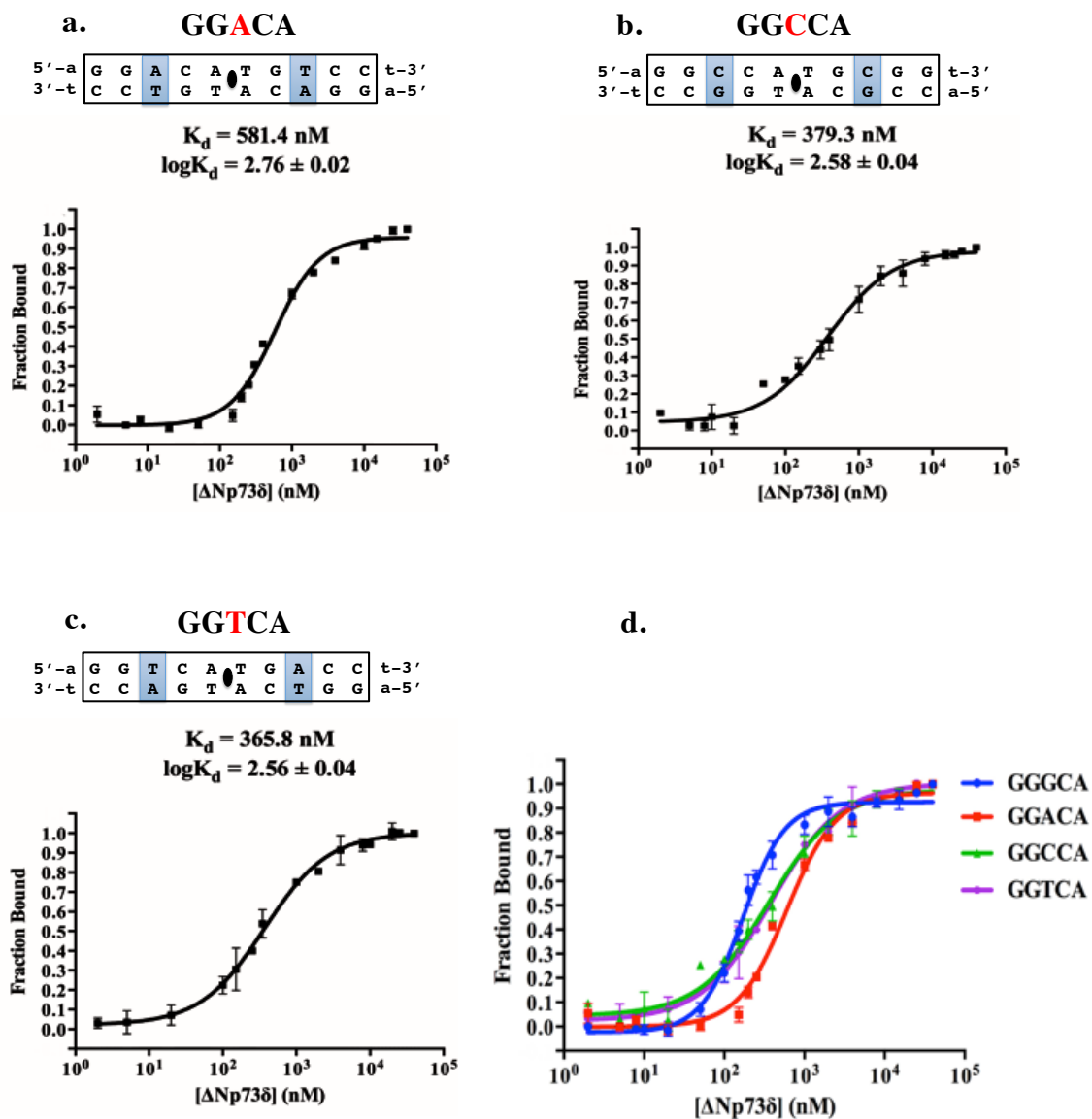


Figure 3.8. Binding affinity graphs of $\Delta Np73\delta$ bound to a half-site response element with different nucleotides at position 3. Graphs a., b., and c. showed the binding affinity of $\Delta Np73\delta$ to sequences with A, C and T in the third position. Graph d. depicted a summary of the three graphs compared with the reference half-site sequence (dark blue).

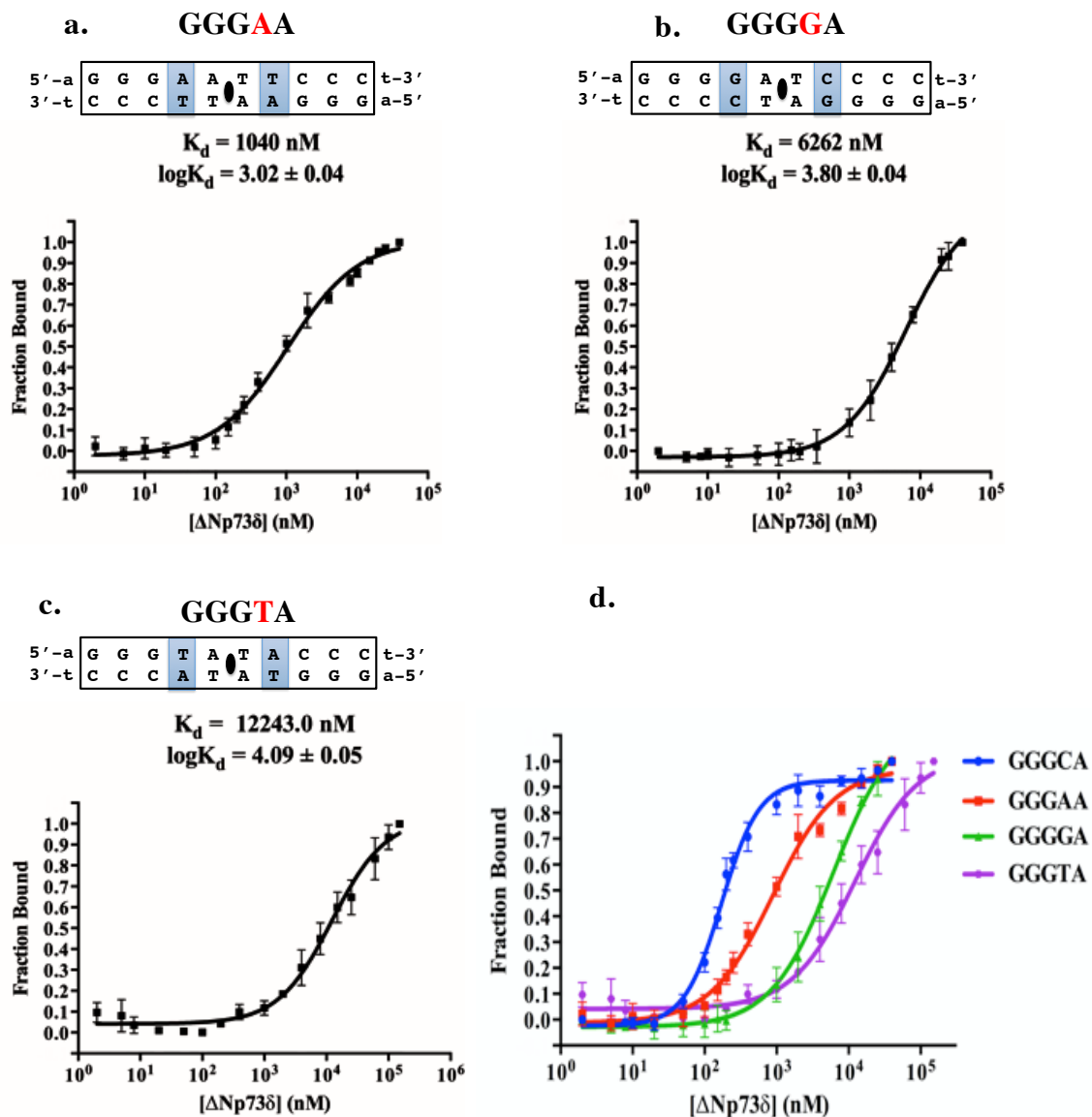


Figure 3.9. Binding affinity graphs of $\Delta Np73\delta$ bound to a half-site response element with different nucleotides at position 4. Graphs a., b., and c. showed the binding affinity of $\Delta Np73\delta$ to sequences with A, G and T in the fourth position. Graph d. depicted a summary of the three graphs compared with the reference half-site sequence (dark blue).

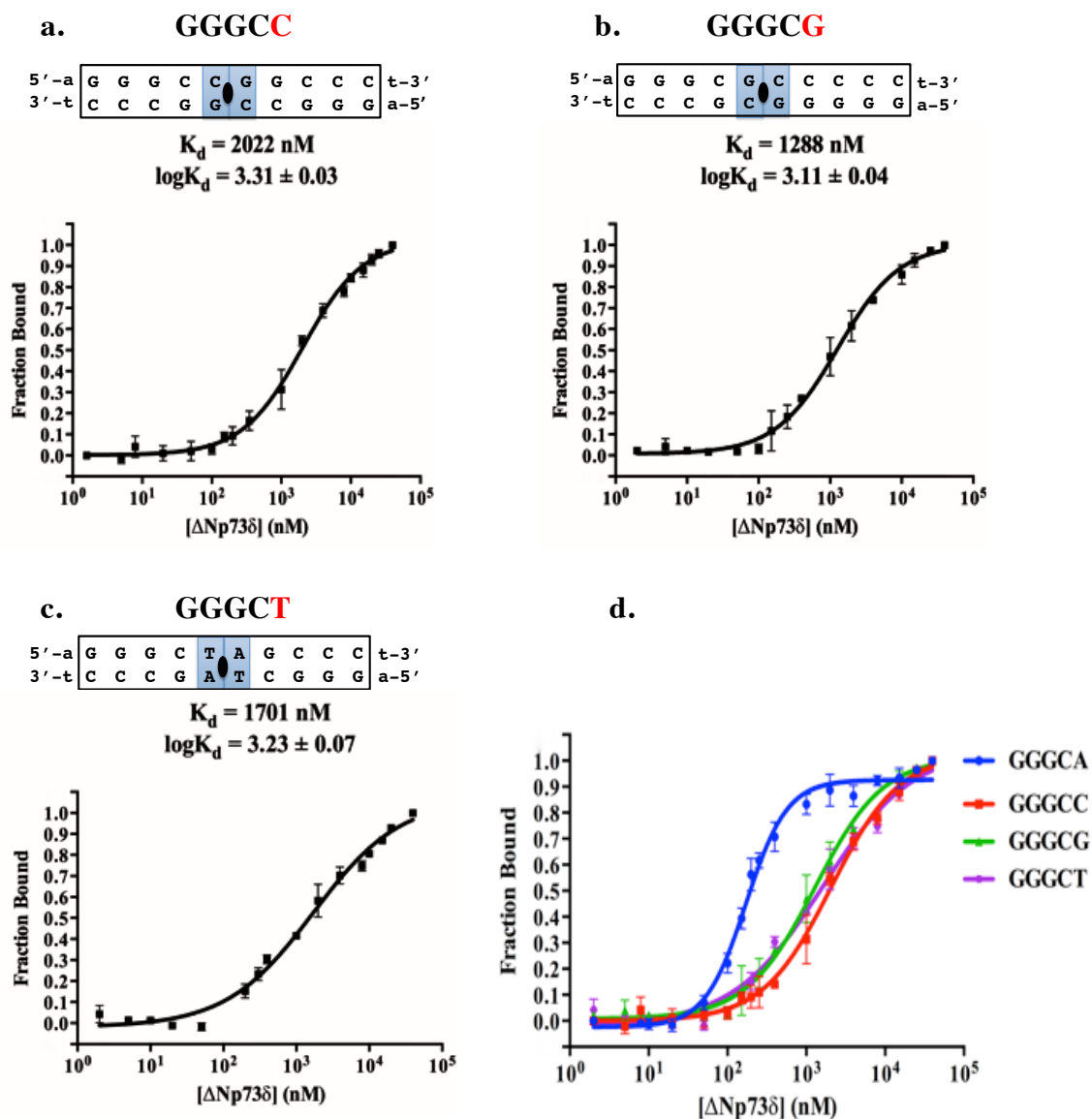


Figure 3.10. Binding affinity graphs of $\Delta Np73\delta$ bound to half-site response elements with different nucleotides at position 5. Graphs a., b., and c. showed the binding affinity of $\Delta Np73\delta$ to sequences with C, G and T in the fifth position. Graph d. depicted a summary of the three graphs compared with the reference half-site sequence (dark blue).

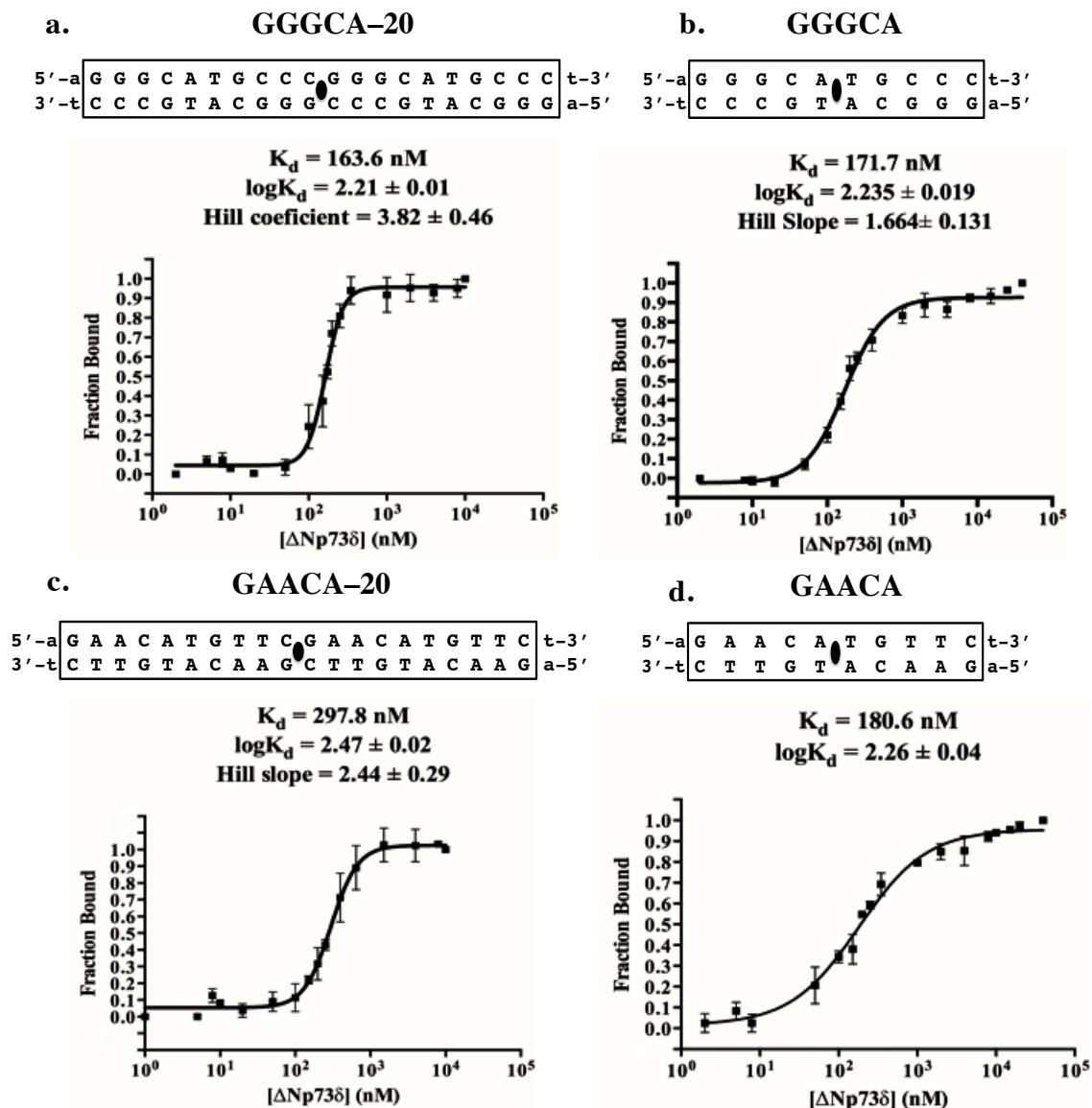


Figure 3.11. Binding affinity graphs of $\Delta\text{Np73}\delta$ to full-sites compared to half-site response elements. Graphs a. and c. showed the binding affinity of $\Delta\text{Np73}\delta$ to full-site response elements, GGGCA-20 and GAACA-20. Graph b. and d. depicted the binding affinity of $\Delta\text{Np73}\delta$ bound to half-sites response elements, GGGCA and GAACA.

Table 3.3. Dissociation constants of $\Delta Np73\delta$ bound to the studied half-site response elements

Sequence	Kd (μ M)	Hill coefficient
GGGCA	0.17	1.66 ± 0.13
First position		
A GGCA	0.19	2.71 ± 0.27
C GGCA	1.06	1
T GGCA	1.15	1
Second position		
G A GCA	0.88	1
G C GCA	11.97	1
G T GCA	2.99	1
Third position		
GG A CA	0.58	1
GG C CA	0.38	1
GG T CA	0.37	1
Fourth position		
GGG A A	0.89	1
GGG G A	6.26	1
GGG T A	12.24	1
Fifth position		
GGG C	2.02	1
GGG G	1.29	1
GGG T	1.70	1

Table 3.4. Dissociation constants of $\Delta Np73\delta$ bound to half-site and full-site response elements

Sequence	Kd (μ M)	Hill coefficient
GGGCA-20	0.16	3.82 ± 0.46
GGGCA	0.17	1.66 ± 0.13
G A CA-20	0.29	2.44 ± 0.29
G A CA	0.18	1

Discussion

In order to understand p73 response element specificity and to separate the contribution of the DBD from the one from the OD, I studied two simplified constructs that were amenable of biochemical characterization: the p73 DBD and the Δ Np73 δ isoform. I used purified proteins to study its oligomerization, its DNA specificity and its cooperativity of binding.

Oligomerization State of the p73 DBD and the Δ Np73 δ Isoform

Hydrodynamic experiments with the purified p73 DBD construct and Δ Np73 δ isoform in the absence of DNA showed differences in their ability to form oligomers. I mentioned in Chapter 2 that pure p73 DBD remains as a monomer in solution with a sedimentation coefficient of 2.10 S, and it is unable to dimerize in the absence of DNA [72]. Instead, the Δ Np73 δ isoform in the absence of DNA is already a dimer with a sedimentation coefficient of 4.94 S and an estimated molecular weight of 71 kDa – close to the theoretical 68 kDa dimer molecular weight. The main difference between the purified p73 DBD construct and Δ Np73 δ isoform is the presence of the oligomerization domain in the Δ Np73 δ isoform. The oligomerization domain leads to the dimerization of the Δ Np73 δ isoform, even in the absence of DNA [80]. The fact that Δ Np73 δ is a dimer in solution is due to the ability of the oligomerization domain (OD) to form a stable interface [81]

Hydrodynamic experiments with the purified p73 DBD construct and Δ Np73 δ isoform in the presence of DNA showed a similar tendency to dimerize for proteins. In the case of the p73 DBD, the absence of the oligomerization domain makes the protein to remain as a monomer and it only dimerizes in the presence of DNA (Figure 3.12.) [72]. In the case of the Δ Np73 δ isoform, upon binding to a half-site response element the sedimentation coefficient increased to 6.03 S. The sedimentation coefficient value increment indicates that the Δ Np73 δ isoform remains as a dimer upon binding to a half-site response element and that a change in the shape of the protein upon binding the DNA to a more spherical shape occurred since the protein sedimented faster

with the DNA. Finally, once the $\Delta\text{Np73}\delta$ isoform can bind to a full-site response element, the s value of $\Delta\text{Np73}\delta$ -full-site complex increases to 7.93 S, which can be explained due to the formation of a tetramer according to the calculated molecular weight (Figure 3.4.).

In summary, the hydrodynamic experiments showed that the $\Delta\text{Np73}\delta$ isoform and the p73 DBD oligomerize into dimers or tetramers when bound to DNA depending on the DNA length. However, in the absence of DNA, their behavior differs. The p73 DBD is a monomer and $\Delta\text{Np73}\delta$ is a dimer underlying the importance of the oligomerization domain in promoting the association of monomers, which is key in DNA binding specificity as observed in p53 [85] [86]. Our sedimentation velocity experiments did not allow quantifying the constant of dissociation of the monomer to dimer formation of p73 DBD and the $\Delta\text{Np73}\delta$ isoform, but I can safely conclude that the constant of dimerization for the $\Delta\text{Np73}\delta$ isoform is much lower than for the p73 DBD alone. As I carried out all the sedimentation velocity experiments in the sub-millimolar range (64.5 μM), I can conclude that the constant of dimer formation for the p73 DBD is in the millimolar or higher range, while for the $\Delta\text{Np73}\delta$ isoform is in the micromolar or lower range. In brief, for all the p73 DBD experiments, I had a predominantly monomer specie, while for the $\Delta\text{Np73}\delta$ isoform, I had a predominantly dimer specie.

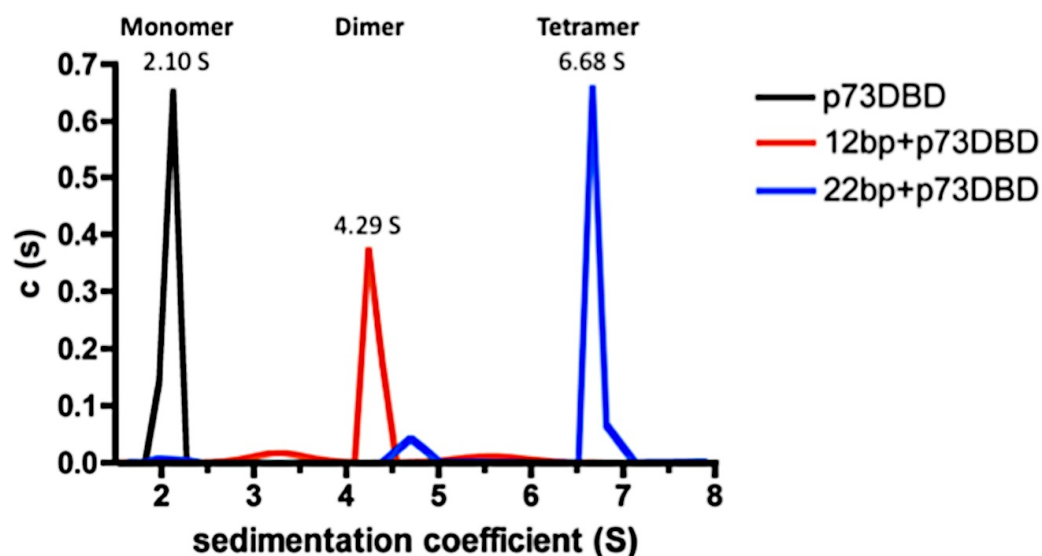


Figure 3.12. Sedimentation coefficients of the p73 DBD bound and unbound to half and full-site response elements [72].

Specificity of Response Element Binding in the Transcription Factor p73

To study response element specificity, I used a reference half-site response element sequence that p73 DBD and the $\Delta Np73\delta$ isoform showed a higher affinity for it in binding and crystallization experiments: the 5'-GGGCATGCCC-3' half-site response element. I observed that the K_d of $\Delta Np73\delta$ bound to the 5'-GGGCATGCCC-3' half-site reference sequence is 20 times lower than the one for the p73 DBD - 171.7 and 3490 nM, respectively. As the presence of the oligomerization domain in the $\Delta Np73\delta$ isoform is the main difference between both proteins, the binding experiments indicate that the oligomerization domain plays a role in the increased binding affinity observed for the $\Delta Np73\delta$ isoform. For all the response elements studied, I always observed the same tendency of the $\Delta Np73\delta$ to have a higher binding than the p73 DBD to the same DNA sequence.

Similarly as was shown for the p73 DBD, fluorescence polarization binding experiments with the $\Delta Np73\delta$ isoform showed that some nucleotide positions within the response element are more important than others to determine DNA binding affinity (Figure 3.13.). I divided the

effects in three categories: those effects in the first three nucleotides of each quarter-site, the effects in the fourth position of the quarter-site, and, finally in the fifth positions of the quarter-sites.

First, I observed that replacing the guanines in the first, second, and third positions of the reference 5'-GGGCATGCCC-3' half-site sequence by pyrimidines reduced the DNA binding affinity of the $\Delta Np73\delta$ isoform. This effect was not observed when the guanine was substituted by the purine adenine. This outcome was not surprising since it has been described that the consensus sequence of response elements that control p53 activated genes favors purines and not pyrimidines at those positions: 5'-PuPuPuC(A/T)(T/A)GPyPyPy-3' [34]. The switch of a purine by a pyrimidine in the second position of each quarter-site has the greatest effect of any of the three external positions in the studied half-site sequence (5'-**PuPuPu**C(A/T)(T/A)G**PyPyPy**-3'). Specifically, changing guanine to cytosine or thymine at positions 2 of each quarter-site increases the DNA dissociation constant more than changes at positions one or three. Such increased DNA dissociation constant can be explained by the disruption of the contact from lysine 138 in loop L1 of the p73 DBD with the purine at the second position [73]. By changing the purine to pyrimidine, the hydrogen bond made between the O6 keto oxygen of the guanine with the amino group of lysine is broken, and consequently, the dissociation constant increases [73]. These results show the importance of a purine in the second position of each quarter-site to enhance p73 affinity for DNA. In brief, when mutations are performed in the triplet nucleotides flanking the half-site response element sequence, the effect in the DNA binding affinity of the p73 DBD and the $\Delta Np73\delta$ isoform is similar.

Second, when the cytosine in the fourth position was swapped to adenine, guanine or thymine, I observed a considerable rise of the dissociation constant for both the p73 DBD and the $\Delta Np73\delta$ isoform. Sequences having adenine, guanine or thymine in the fourth position of each quarter-site showed a substantially increment in the dissociation constant. As the cytosine at the

fourth position is the most conserved among the p53 canonical sequence, the decrease in affinity was expected since the hydrogen bonds between Arg300 in the p73 DBD and the guanine that is complementary to the cytosine in the fourth position are completely disrupted.

Third, also similarly to the p73 DBD, the DNA affinity for the $\Delta Np73\delta$ isoform decreased when cytosine, guanine or thymine occupied the fifth position, instead of adenine. Since the canonical p53 response sequence has an adenine or a thymine in the fifth position of the quarter-site, I expected to have a similar dissociation constant upon replacing adenine with thymine, but I found also a decrease in affinity as I had found for p73 DBD. As mention in chapter 2, the decrease in affinity due to variation in this position can be explained by the lack of interaction of Arg268 with the distorted phosphate backbone that has been observed in all the p73-DNA crystal structures [73].

Both, the p73 DBD and the $\Delta Np73\delta$ isoform share similar DNA binding affinity profiles (Figure 3.13.). The differences were more a matter of degree in the selectivity of nucleotides at certain positions. For example, although the binding affinity of the p73 DBD and the $\Delta Np73\delta$ isoform dropped upon changing the guanine in the second position for a pyrimidine, the affinity of the $\Delta Np73\delta$ isoform decreased 17 to 70-fold; in comparison the DNA affinity of p73 DBD that decrease only 2.3-fold (Figure 3.13.). The lower affinity showed by the $\Delta Np73\delta$ isoform for sequences containing cytosine or thymine in the second position might be derived from the restriction of the oligomerization domain to the selectivity of the p73 DBD for consensus nucleotides. Also, there are scale differences in the increase of the dissociation constant upon changing cytosine in the fourth position: 6 to 7-fold for the p73 DBD and 5 to 72-fold for the $\Delta Np73\delta$ isoform (Figure 3.13.).

In a manner related to the structure of loop L1, the oligomerization domain seems to regulate DNA specificity since I observed in our published data that $\Delta Np73\delta$ isoform is able to distinguish between sequences containing GGGCA and GAACA in the pentamer repeats by

showing a greater affinity for GGGCA compared with GAACA sequence [68]. In an important conclusion to understand the differential expression of genes controlled by different response elements, clearly the oligomerization domain is able to communicate with the DBD to enhance binding to GGGCA-containing response elements over GAACA ones.

Cooperativity of Response Element Binding in the Transcription Factor p73

As the members of the p53 transcription family act as tetramers to activate transcription, it is important to understand the possibility that the monomers that recognize the quarter-site response elements communicate with each other once they bind the full-site response element to create a cooperative effect that might result in a tighter response element binding at lower transcription factor concentration. Disregarding the effect that chromatin structure might have on the transcriptional activity, the properties of DNA binding by the specific transcription factors, like p53 and p73, to the response elements of the enhancer are likely to have profound effects on the transcriptional activity of the genes that are controlled by these factors. For this reason, I analyzed the cooperativity effect on DNA binding in the two studied p73 constructs, the p73 DBD and the Δ Np73 δ proteins.

I reported an increased DNA affinity of Δ Np73 δ in comparison with the p73 DBD for all the studied half-site response elements. Apart from the difference in DNA binding affinity, the Δ Np73 δ isoform showed a cooperative DNA binding to the DNA (Figure 3.5.); while the p73 DBD did not show cooperativity (Figure 2.3.). The analysis for cooperativity was performed for both proteins where after measuring the experiments, the raw data was fit into two non-linear equations, one accounting for cooperativity and the other not. For p73 DBD, all the analyzed data fit better the equation with a Hill coefficient equals to one meaning a lack of cooperativity in the binding of p73 DBD to the sequences tested.

When a protein shows cooperativity, the binding of one ligand increases the binding to the second site. In our case, the binding of one Δ Np73 δ monomer to one quarter-site response

element increases the affinity of another $\Delta\text{Np73}\delta$ monomer to the adjacent quarter-site response element. As our analysis of cooperativity in the two p73 proteins shows, the DNA binding affinity of the $\Delta\text{Np73}\delta$ isoform is higher than the one for the p73 DBD for all the half-site response elements that I studied. As the only difference between the two proteins is the presence of the NLS sequence and the oligomerization domain that are present in the $\Delta\text{Np73}\delta$ isoform and absent in the p73 DBD protein, I postulate that the oligomerization domain is involved in the DNA binding mechanism. The simplest explanation for a mechanism that awaits structural support is that, once the first DBD has bound to the response element, the oligomerization domain, that forms a stable dimer and brings together two DNA binding domains, multiplies the affinity by radically increasing the chance of the second DBD to bind the DNA.

For both proteins, the p73 DBD and the $\Delta\text{Np73}\delta$ isoform, when purines were replaced by pyrimidines in the first three position of each quarter-site or any mutation was introduced in the fourth or fifth position, the K_d of for DNA increased (lower affinity). But, only for $\Delta\text{Np73}\delta$ isoform, the slope of the binding curve became less steep with a Hill coefficient close to 1 that reflected a loss of the cooperativity (Material and methods Chapter 2). The only change that continued to show cooperativity was when adenine replaced guanine in the first position (Figure 3.6.a.). In brief, $\Delta\text{Np73}\delta$ isoform binds in a cooperative way only when GGGCA and AGGCA are present in the half-site response element (Figure 3.5. and 3.6.a., Table 3.3.).

The $\Delta\text{Np73}\delta$ isoform also binds to a full-site response element GGGCA-20 (3.11.). To determine the binding affinity of $\Delta\text{Np73}\delta$ binds to a full-site response element, I measured the dissociation constant for two full-sites called GGGCA-20 and GAACA-20: the K_d were 163.6 nM and 297.8 nM, respectively. The binding affinity for the GGGCA half-site was virtually identical (171.7 nM) to the GGGCA-20 full-site (163.6 nM) (Figure 3.11.). Also, the cooperative behavior was present in both with a Hill coefficient of 3.83 and 1.66 for a full and half-site, respectively (Table 3.3. and 3.4.). Moreover, when I measured the binding affinity of $\Delta\text{Np73}\delta$ to

the GAACA half-site and the GAACA-20 full-site, the binding affinity continued to be comparable, 180.6 versus 297.8 nM. Although in this case, the cooperativity was lost in the GAACA half-site. Based on these results, I speculate that guanine in the positions two and three of the quarter-site enhances cooperativity of the $\Delta Np73\delta$ isoform for the response element (Figure 3.11.). Cooperativity is also observed for a p53 protein construct that contains the DNA binding domain and the oligomerization domain bound to p21 and MDM2-controlling response elements [39]. By comparing the results of the two sets of experiments, I showed that the presence of the oligomerization domain of p73 increases the DNA binding affinity and its selectivity towards certain nucleotides at positions one, two and three of the quarter-site response elements; clearly the oligomerization domain is able to communicate with the DBD to enhance binding to GGGCA-containing response elements over GAACA ones.

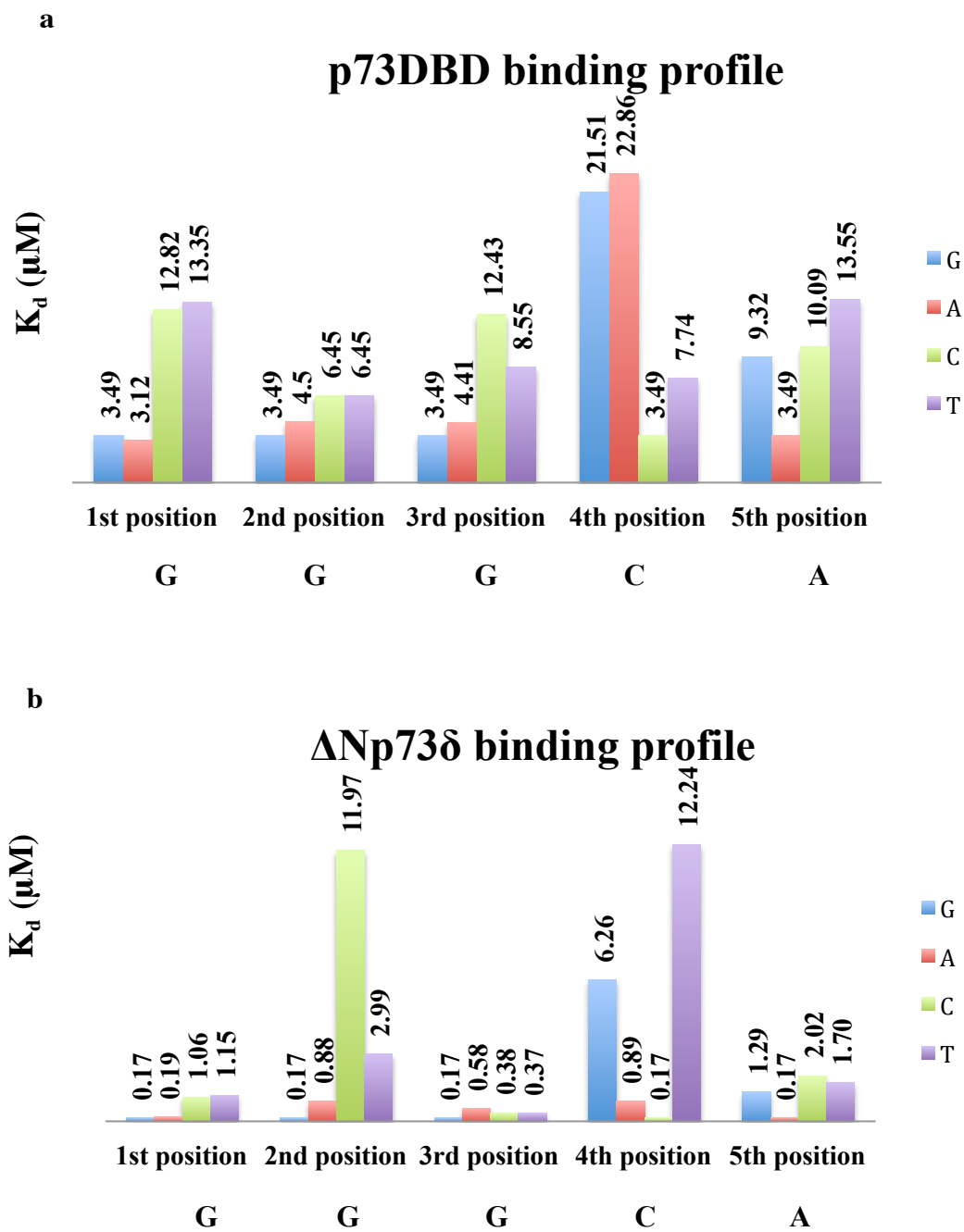


Figure 3.13. Binding profile of p73DBD and $\Delta\text{Np73}\delta$. The K_d units are in μM units.

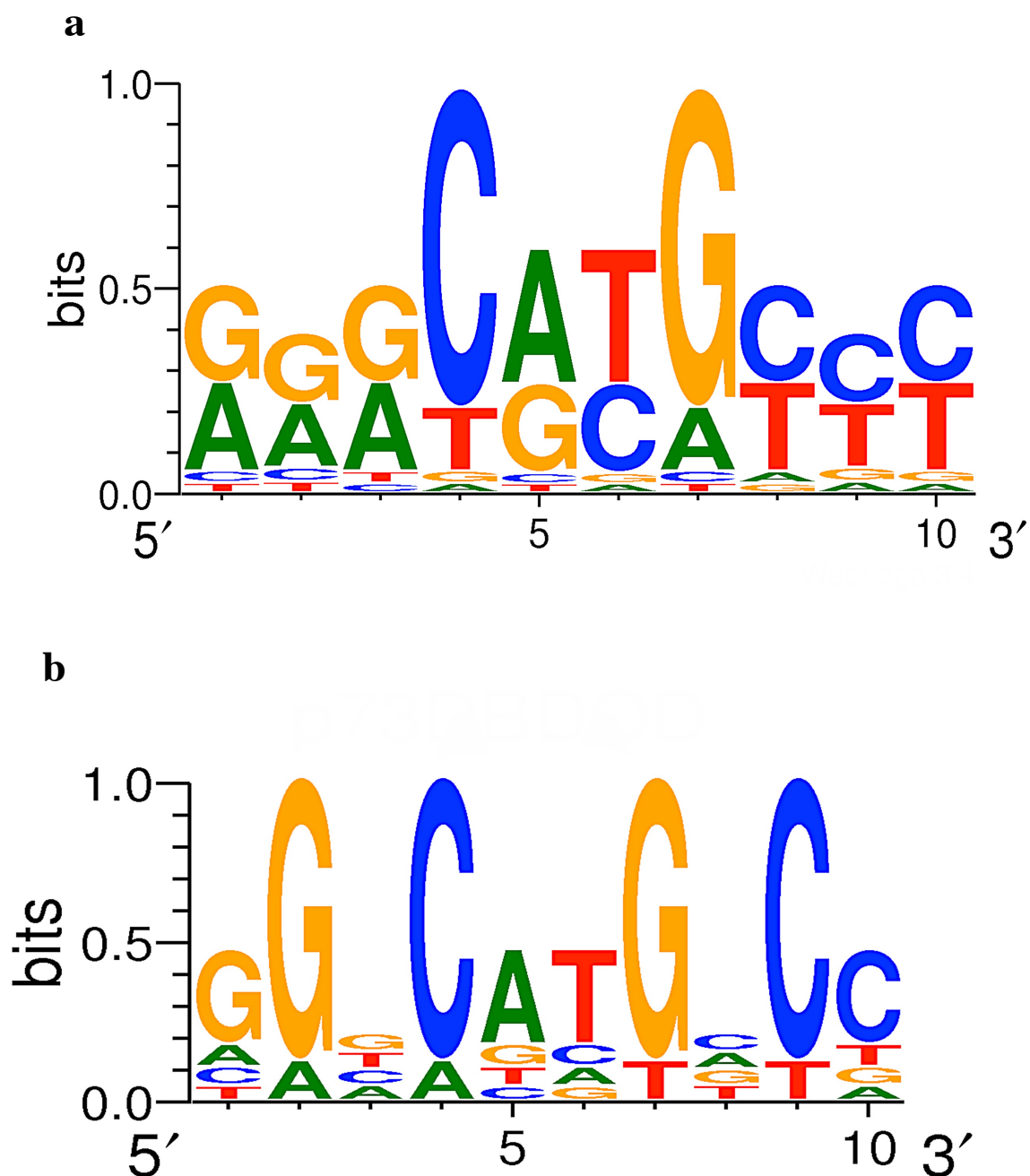


Figure 3.14. p73DBD and $\Delta Np73\delta$ specificity. The WebLogo displays a summary of the half-site response element specificity of p73DBD and $\Delta Np73\delta$. The closest to the unity in the y axis meant the high selectivity of p73 to a certain nucleotide in a position of the half-site consensus sequence.

Chapter 4

A Hoogsteen Base Conformation

in p53 Correlates with

Active Response Elements

Introduction

Transcription factors are key molecules that regulate the use of genetic information. They can activate or repress the expression of genes depending on the cell's needs. A key step on gene activation is the recognition of the response element sequence by the transcription factor. For example, the Oct1 transcription factor acts as a weak transactivator when bound to the ATGCAAAT recognition site, whereas when bound to the TAATGA(Pu)AT Oct1 recognition site, transcription is strongly activated [87]. Pit1 transcription factor follows a similar activation/repression mechanism where the binding site is a key regulator of the structure adopted by the transcription factor and its activity [88]. The effect of the recognition site on the transcription factor function can vary from strong to low or null level of transcription of the target genes. Then, if transcription factors are dependent on the sequence of the response element to activate or repress a gene, how do transcription factors 'read' the response element sequence to either activate or repress the target gene?

In the case of our model, the p53 family of transcription proteins, considering the traditional 5'-PuPuPuCA/TGPyPyPy-3' consensus sequence for half-site p53 response elements, Wang et al. investigated deeper and classified response elements as activating and repressing by measuring transactivation level of 162 verified p53 response elements by Luciferase gene reporter assays[53]. They proposed and tested three rules to predict if a response element sequence would activate or repress. First, the core CWWG in the half-site p53 response elements mostly determines the activation and repression function; second, the three nucleotides flanking the core XXXCWWGYYY in each side exerts a modulating effect on transactivation; and, third, the adjacent nucleotide to the core XCWWGY has the strongest requirement to conform to the consensus in order to maintain activation [53]. Although this information provides insights on the response element sequences that make p53 to activate or to repress a gene, still the molecular

mechanism used by p53 to discriminate the activating from the repressing response element is not established.

Our goal is to describe the molecular mechanism that allows p53 to discriminate between activating and repressing response elements. Structural studies have shown that the p53 DBD forces the two central nucleotides in each half-site (CWWG) to adopt a consistently deformed DNA conformation with respect to the expected standard B-DNA geometry. More recent p53 DBD crystal structures in complex with half and full-site response elements to significantly higher resolution with AT, TT, and AA in the central two nucleotides have observed that adenines flips to a Hoogsteen base conformation, instead of the canonical Watson and Crick conformation [38] [69]. Interestingly, this structural observation matches the transactivation activity assays that determine that the two central nucleotides make p53 to activate or repress the downstream gene [53]. Moreover, p53 structures with AGGCA and GGACA in each quarter-site response element show that Lys120 adopts different conformations. In the AGGCA sequence, Lys120 is contacting the two consecutive guanines, whereas, in the GGACA one, Lys120 interacts with the second guanine and the thymine of the opposite strand [36]. Based on these functional and structural data, I investigated the relation of p53 activating and repressing response elements with Hoogsteen geometry and the different conformations of Lysine120 observed in the p53DBD structures bound to the consensus sequence. In this chapter, to study at the molecular level the recognition pattern of p53 to the different sequences. I describe the structures of p53 DBD bound to two activating response elements (AGGCA and GGACA) and one repressing sequence (TTTCA). Furthermore, I performed binding experiments to study the effect of activating and repressing response elements in p53DBD and AcLys120-p53DBD affinity to the DNA.

Materials and Methods

Subcloning of the p53DBD Protein

Residues 92 to 292 encoding for the DNA binding domain of human p53 with a His-tag at the N-terminus were subcloned, using *Eco*R1 and *Hind*III restriction sites, into the overexpression vector pET28. The sequence of the vector was verified by sequencing the inserted gene. Throughout this thesis, I will refer to the overexpressed gene product of the residues 92 to 292 of human p53 as the p53DBD protein. BL21/DE-3 *E. coli* competent cells were transformed with the pET28 vector containing the p53 DBD construct and were selected with the antibiotic kanamycin. The BL21/DE-3 *E. coli* cells with pET28 containing the inserted p53 DBD construct were stored in 50% glycerol at -80°C until further use.

Expression of the p53DBD Protein

Cells were grown in LB medium with 30 µg/mL of kanamycin at 37°C. Upon reaching an absorbance of 0.6 AU at 600 nm, cells were induced with 0.5 mM IPTG and grown overnight at 25°C.

Subcloning of the AcLys120-p53DBD Gene

The human p53 DNA binding domain (p53 DBD), comprising residues 92 to 292, was acetylated at position of Lys120 by incorporating N-acetyl-lysine at the TAG amber stop codon in a technique developed by Jason Chin at MRC [89]. The Lys120 codon was mutated to the TAG codon in the human p53 DNA binding domain gene by site-directed mutagenesis with the following primers: 5'-acagcggcaccgcgTAGtctgtccctgcac-3' and the antisense 5'-gtgcaggtgacagaCTAcgcggtgccgctgt-3'. The gene fragment codifying for residues 92 to 292 of the p53 DBD with the TAG stop codon at position Lys120 was cloned into the pCDFduet expression vector with *Bam*H1 and *Sal*I restriction sites. The expression vector pCDFduet with the DNA sequence codifying for the His-tagged p53DBD with the TAG codon in the Lys120 position was

co-transformed with pBK-ACKRS-3 containing the codifying sequence for the synthetase ACKRS-3. Both vectors, pCDFduet and pBk-ACKRS-3, were co-transformed in BL21/DE-3 *E. coli* competent cells that were selected for streptomycin and kanamycin antibiotics [90].

Expression of the AcLys120-p53DBD Protein

Cells containing the N-terminal His-tagged p53DBD gene with the TAG codon in the Lys120 position and the ACKRS-3/tRNA^{Acua} pair were grown in 2xYT medium with 30 µg/mL of kanamycin and 50 µg/mL of streptomycin at 37°C. To make the ACKRS-3/tRNA^{Acua} pair incorporate the acetyl-lysine into the Lys120 position, I added 5 mM acetyl-lysine to the 2xYT medium. Before induction, at an absorbance of 0.4 at 600 nm, 50 mM of nicotinamide was added to the medium in order to inhibit the deacetylases activity. Upon reaching an absorbance of 0.6 at 600 nm, cells were induced with 0.5 mM IPTG and grown overnight [91]. During the induction process, the orthogonal pair incorporates the acetyl-lysine into the stop codon of the p53DBD gene with the TAG codon in the Lys120 position. I will refer to the acetylated protein as AcLys120-p53DBD.

Purification of p53DBD and AcK120-p53DBD Proteins

The purification preparations of the p53DBD and AcK120-p53DBD proteins were carried out with the same protocol. After induction, cells were harvested by centrifugation at $2,846 \times g$ for 30 min at 5 °C. The supernatant was discarded and the pelleted cells were resuspended in lysis buffer (0.5 M sodium chloride, 20 mM sodium citrate (pH 6.1), and 10 μ M zinc chloride). Cells were lysed in a microfluidizer and centrifuged at $104,444 \times g$ to remove the cell debris from the cytoplasmic fraction. The proteins were purified by affinity chromatography. The cytoplasmic fraction containing the p53DBD protein was incubated overnight with 1 ml of nickel-nitrilotriacetic acid agarose resin (Ni-NTA) at 5°C. To remove the non-specifically-bound protein contaminants from the resin, the resin was washed with 50 ml of lysis buffer and 150 ml of lysis buffer supplemented with 20 mM imidazole. For the AcLys120-p53DBD protein, an extra step was added to wash the resin with 50 ml of lysis buffer with 50 mM imidazole. Finally, the protein was eluted with lysis buffer supplemented with 350 mM imidazole. Subsequently, proteins were concentrated using the centrifugal filter concentrators to a volume of 3-4 ml.

To further purify the protein, I injected, one milliliter at a time, the 3 or 4 milliliters into a Superdex 200 size-exclusion column equilibrated with binding buffer (100 mM sodium chloride, 10 mM sodium citrate (pH 6.1), 10 mM DTT, and 10 μ M zinc chloride). The degree of purification was determined by running a 15% SDS-PAGE and the identity of the protein was confirmed by a western blot using an anti His-tag antibody.

Mass Spectrometry Analysis of the AcLys120-p53DBD Protein

To verify the site-specific incorporation of acetyl-lysine at position Lys120 of the p53 DBD, trypsin in-gel digestion and LC-MS/MS analysis were performed. The small gel fragments (about 1 mm long) containing the protein of interest were destained with 3 to 4 washes of acetonitrile and ammonium bicarbonate at room temperature with an incubation of 15 minutes between each wash. Then, the sample was alkylated by adding 250 μ L of 10 mM DTT/100 mM of ammonium bicarbonate with 45 minutes incubation at 56°C; immediately after, 250 μ L of 55 mM of iodoacetamide in 100 mM of ammonium bicarbonate was added and the samples were incubated at room temperature for 30 minutes in the dark. After the alkylation step, trypsin digestion was performed where 50 μ L of a solution containing 20 mM calcium chloride, 100 mM of ammonium bicarbonate and 1.25 μ L of 1 μ g/ μ L of trypsin was added to each sample. The samples were incubated for 30 minutes at 4°C, and they were transferred to a 37°C bath for overnight incubation. After digestion, the gel fragments were treated with gel extraction buffer (1% formic acid and 2% acetonitrile) and the supernate was removed. This step was repeated twice. The pooled samples were dried in a speed vacuum concentrator. Before the LC-MS/MS analysis, 10 μ L of a solution containing 2% formic acid and 5% of acetonitrile were added to the dried sample.

The digested peptides were analyzed by ultra high-pressure liquid chromatography in tandem with nanospray-ionization mass spectroscopy (LC-MS/MS). The nanospray ionization experiments were performed in a nanoscale reverse phase column followed by TripleTof 5600 hybrid mass spectrometry. Peptides were eluted from the C18 column into the mass spectrometer using a 5–80% linear gradient of acetonitrile (ACN) at a flow rate of 250 μ L/min for 1 hr. Two buffers were used to create the ACN gradient: buffer A (98% H₂O, 2% ACN, 0.1% formic acid, and 0.005% TFA) and buffer B (100% ACN, 0.1% formic acid, and 0.005% TFA). The MS/MS data were acquired in a data-dependent manner. The MS1 data with an m/z ratio of 400 to 1250

were acquired for 250 ms and the MS2 data with m/z ratio values of 50 to 2,000 for 48 milliseconds. Finally, the collected data were analyzed using Protein Pilot 4.5 (ABSCIEX) for peptide identifications.

Measuring p53DBD and AcK120-p53DBD Proteins Binding Affinity to Half-site

Response Elements by Fluorescence Polarization

To measure the binding affinity of the p53DBD and the AcK120-p53DBD proteins for different half-site response elements, a set of 5'-fluorescein-labeled oligonucleotides were acquired from Integrated DNA Technologies (IDT). Table 4.1 lists the 4 sequences used. Mutations in the first three positions of each quarter-site were introduced to explore their binding contribution.

Table 4.1. List of 5'-fluorescein oligonucleotides used for the binding experiments

Function	Name	Sequence
Activating	GGGCA	5'-FAM-aGGGCATGCCctataGGGCATGCCct-3'
Activating	AAACA	5'-FAM-aAAACATGTTTtataAAACATGTTTt-3'
Repressing	CCCA	5'-FAM-aCCCATGGGgtataCCCATGGGgt-3'
Repressing	TTTCA	5'-FAM-aTTTCATGAAAtataTTTCATGAAAt-3'

To measure the DNA binding constant of the p53DBD and AcK120-p53DBD for the four half-site response elements, I prepared 13 to 15 serial dilutions with protein concentrations from 20 μ M to 2 nM. Each protein concentration was mixed with 50 nM 5'-fluorescein-labeled 26-mer dsDNA in a 350 μ l final volume of binding buffer (100 mM sodium chloride, 10 mM sodium citrate (pH 6.1), 5 mM DTT, and 5 μ M zinc chloride). Before making the fluorescence polarization measurements, tubes were incubated for 30 minutes at 5°C.

Fluorescence polarization was determined as described in Chapter 2. Briefly, four fluorescence intensity measurements with four different filter arrangements were measured at excitation and emission wavelengths of 494 and 521 nm, respectively, in a Hitachi F-2000 fluorescence spectrometer. For each half-site response element studied, the binding experiment was repeated three times.

DNA Binding Data Analysis

The Graph-Pad Prism software was used to analyze the data. Non-linear analyses were performed to fit the data and obtain EC_{50} values. Two equations were used: a sigmoidal dose-response and a sigmoidal dose-response with variable slope. The EC_{50} or K_d value corresponds to the protein concentration at which half of the DNA sites are occupied. For a more detailed explanation about the analysis, please refer to the Materials and Methods section in Chapter 2.

Crystallization of p53DBD with Activating and Repressing Response Elements

In order to understand the structural basis of p53 activation, I crystallized p53DBD in complex with response elements that activate or repress p53-controlled genes. The pure p53DBD protein in a buffer of 100 mM sodium chloride, 10 mM sodium citrate (pH 6.1), 5 mM DTT, and 10 μ M zinc chloride was concentrated to 5 to 10 mg/ml. Three palindromic 26-mer oligonucleotides that form a double-stranded DNA by bending a TATA-linker to create a half-site response element were used to assemble the protein-DNA complex (Table 4.2). To crystallize the complexes, I used the hanging-drop vapor diffusion method with VDX multi-well plates at room temperature. In a final volume of 30 μ l, 26.1 μ l of pure p53DBD protein at 0.6 mM was mixed with 3.9 μ l of DNA at 1 mM concentration. The crystallization drops had a protein:DNA ratio of 4:1, a two-fold DNA excess from the expected ratio for the complex of 2:1. After mixing the protein and DNA solutions, the complex was incubated at room temperature for at least one hour and any precipitation was removed before setting the tray by centrifugation. The hanging-drop had 1 μ l of soluble complex and 1 μ l of reservoir condition. Optimal crystallization conditions varied slightly between the three complexes.

Crystals of p53DBD bound to the AGGCA activating oligonucleotide grew in 22-30% PEG-3350, 0.26-0.32 mM sodium chloride, and 0.1 M bis-tris methane (pH 6.8). The crystals had a V shaped form and grew in a period of 5 days (Figure 4.1.a.). Crystals also grew in potassium chloride, instead of sodium chloride, but not of the same quality. The condition where I obtained

the largest and sharped-edged crystals was in PEG-3350 26%, 0.3 M sodium chloride, and 0.1 M bis-tris methane (pH 6.8).

Similarly, crystals with p53DBD bound to the GGACA response element were observed at conditions with 20-27% PEG-3350, 0.25-0.4 mM sodium chloride, and 0.1 M bis-tris methane (pH 6.5-6.6). The crystals were small, amorphous and with not well defined edges (Figure 4.1b) and they 5-7 days to grow. Crystals also grew in potassium chloride; however they were smaller than with sodium chloride. The condition where I obtained the crystals that diffracted was 24% PEG-3350, 0.3 M sodium chloride, and bis-tris methane (pH 6.6).

Crystals with p53DBD bound to the TTTCA response element grew at a slightly higher salt concentration compared with the AGGCA and GGACA complexes. The cross-shaped, large, and well-defined edged crystals were obtained in 30% PEG-3350, 0.4 M sodium chloride, and 0.1 M bis-tris methane (pH 6.8) (Figure 4.1c). The crystals took 5-7 days to grow. Prior to data collection, I tore the clustered crystal into single crystals, and I collected the diffraction data in one part of the cross-shaped crystals. Prior to data collection, all crystals were frozen in a cryoprotectant solution using the same reservoir condition where they grew, but increasing the PEG-3350 concentration to 35%.

Table 4.2. The palindromic double stranded DNA sequences containing two binding sites used to crystallize the p53DBD-DNA complex.

Function	Name	Sequence
Activating	AGGCA	5' -FAM-aAGGCATGCCTtataAGGCATGCCTt-3'
	GGACA	5' -FAM-aGGACATGTCctataGGACATGTCct-3'
Repressing	TTTCA	5' -FAM-aTTTCATGAAAtataTTTCATGAAAt-3'

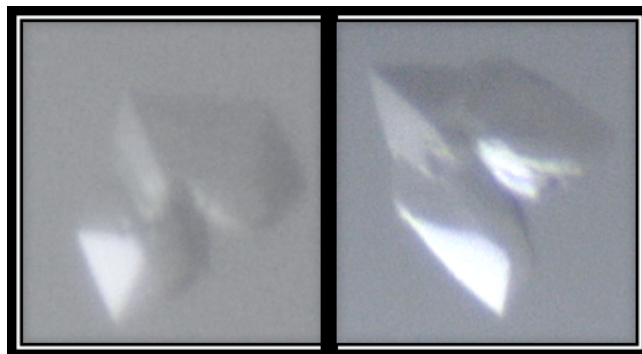
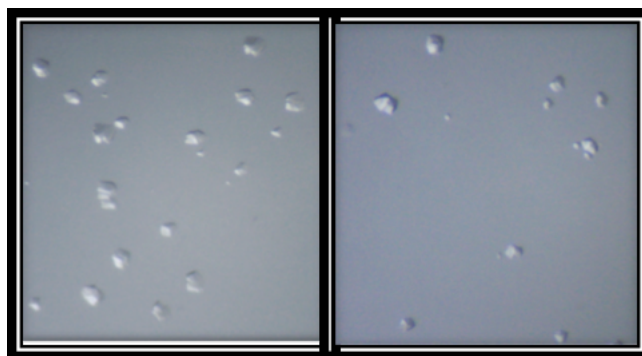
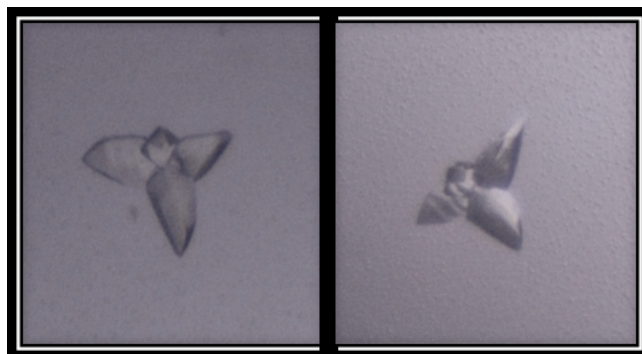
a. AGGCA complex**b. GGACA complex****c. TTTCA complex**

Figure 4.1. Crystals of p53DBD in Complex with Half-site Response Elements. a. V-shape of the crystals obtained when p53DBD binds to AGGCA. b. Small amorphous crystals formed when p53DBD binds to GGACA. c. Three-leaf crystals grown when p53DBD binds TTTCA.

Diffraction Data Collection and Structure Determination

Crystals were diffracted at beamline BL7-1 of the Stanford Synchrotron Radiation Lightsource at a wavelength of 1.127 Å. The diffraction data was collected on an ADSC-Q315 CCD detector. The collected data was indexed, integrated and scaled in HKL2000 [92] and the obtained reflection file from HKL2000 was further used to solve the structure. The phase information for each data set was found by molecular replacement using Phaser [93]. The search model used in molecular replacement was a p53DBD dimer bound to half-site response element (PDB ID: 4GTG). The atomic model refinement was performed in Phenix [94] using rigid body, simulated annealing, positional and B-factor refinement. Weight optimization of the phases was used at the end of the refinement to establish better phase probability distributions [95]. Steps of positional and b-factor refinement in Phenix were intercalated with manual refinement steps in Coot [96]. To verify the model and to avoid model bias, composite omits maps were calculated in Phenix and used for manual refinement in Coot. Finally, structure validation was carried out with the Phenix comprehensive validation tool. During refinement, I monitored in Phenix and manually corrected in Coot regions with disfavored structural features; for example: bonds, angles, side-chain rotamers and Ramachandran plot outliers [97]. Besides, validation in Phenix, also the program PROCHECK was further use to obtain an extensive geometric analysis of the structure [98].

Results

Genetically incorporation of acetyl-lysine into p53DBDTAG

Acetyl-lysine was first specifically incorporated into a protein using N-acetyl-lysyl-tRNA and synthetase/tRNA^{Acua} from *M. barkeri* to acetylate myoglobin [89]. The system was further optimized to increase the efficiency of acetyl-lysine incorporation into the targeted residue by optimizing the AcKRS-3 tRNA-synthetase used in this work [90]. From the MS/MS analysis, I identified peptides containing acetylated Lys120 with 99% confidence, thus confirming the efficiency and fidelity of the AcKRS-3 tRNA-synthetase and the tRNA^{Acua} to work in BL21/DE-3 *E. coli* cells (Figure 4.2.).

Expression and Purification of the p53DBD and AcLys120-p53DBD Proteins

I purified the p53DBD and AcLys120-p53DBD proteins with a two-step purification protocol (Figures 4.3. and 4.4.). The first step was affinity chromatography using the poly-histidine tag in the N-terminus of the protein construct. The second step was a size-exclusion chromatography step.

In order to remove high molecular impurities present after eluting the proteins from the affinity resin with 350 mM imidazole, I concentrated the eluted sample to 1 ml volume with centrifugal filter concentrators and injected it into the size-exclusion column. As observed from the elution profile of these columns, the p53DBD and AcLys120-p53DBD proteins started to elute at a retention volume of 36 ml with the peak at 40 ml (Figures 4.5 and 4.6). For each protein, I collected and concentrated the gel-filtration fractions from volumes 36 to 42 ml.

To determine the degree of purity of the p53DBD and AcLys120-p53DBD preparations, I ran a 15% SDS-PAGE with samples from every step of the purification protocol. I observed intense bands for the p53DBD protein (23 kDa theoretical Mw) and for the AcLys120-p53DBD protein (23.55 kDa theoretical Mw) below the 26 kDa reference protein ladder band. No visible

higher or lower molecular weight impurities were observed in the final preparation (last lane in Figures 4.3. and 4.4.).

To verify the identity of the p53DBD and AcLys120-p53DBD proteins, I carried out a Western blot using an anti-His-tag antibody. I confirmed the presence of the His-tagged target proteins by observing bands in the Western blot that matched the expected molecular weight and the bands observed in the Coomassie-stained SDS-PAGE (Figures 4.3. and 4.4.). The final pure preparations were concentrated using centrifugal filter concentrators. To corroborate the presence of the acetyl group at the Lys120 position, I performed an in-gel digestion followed by a mass spectrometry analysis. From the analysis, I was able to observe peptides with Lys120 acetylated (Figure 4.2.). Finally, I quantified the protein concentration of the purified samples using the Bradford assay. Following this protocol, I obtained a yield 7-9 mg of pure p53DBD and 3-4 mg of pure AcLys120-p53DBD per liter of bacterial culture.

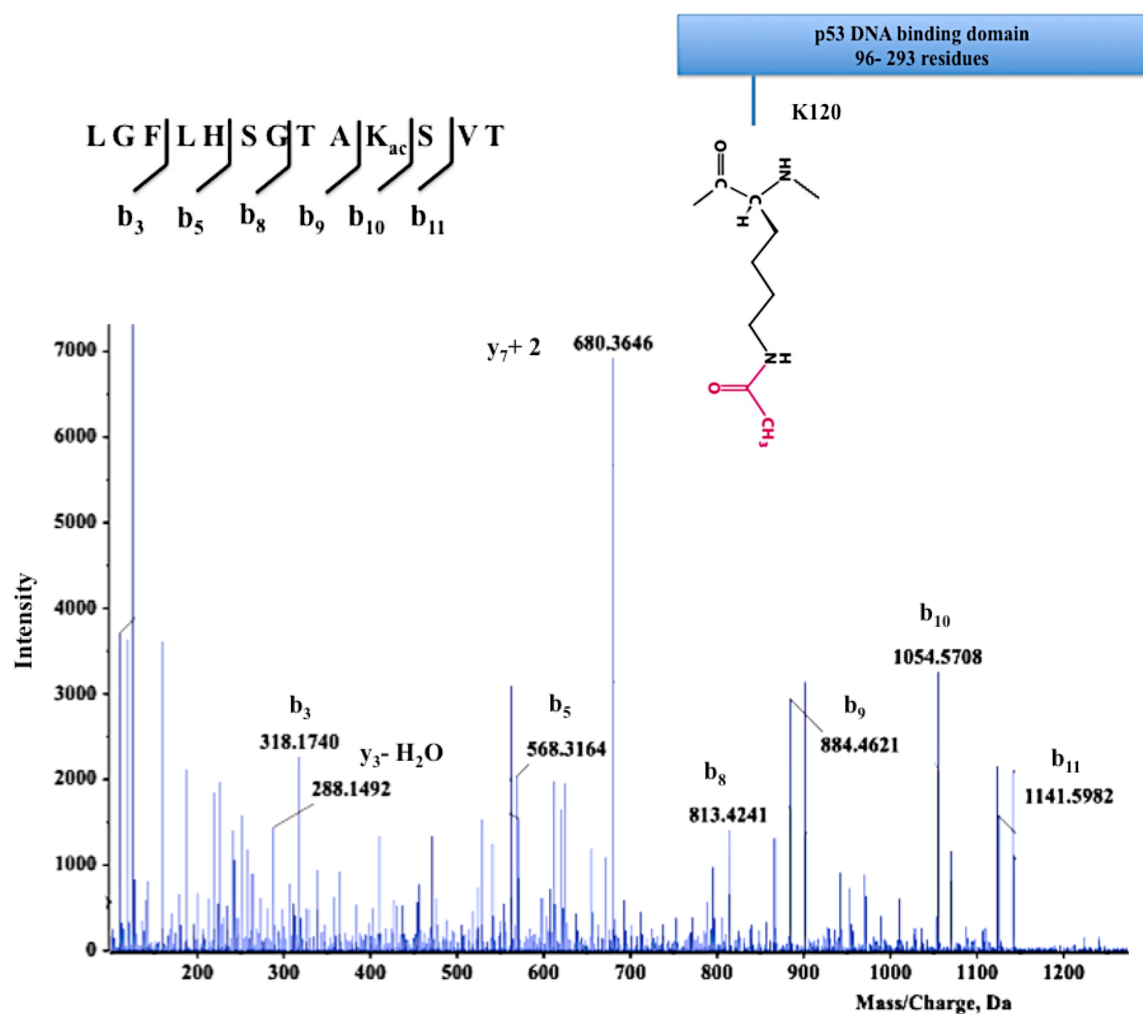
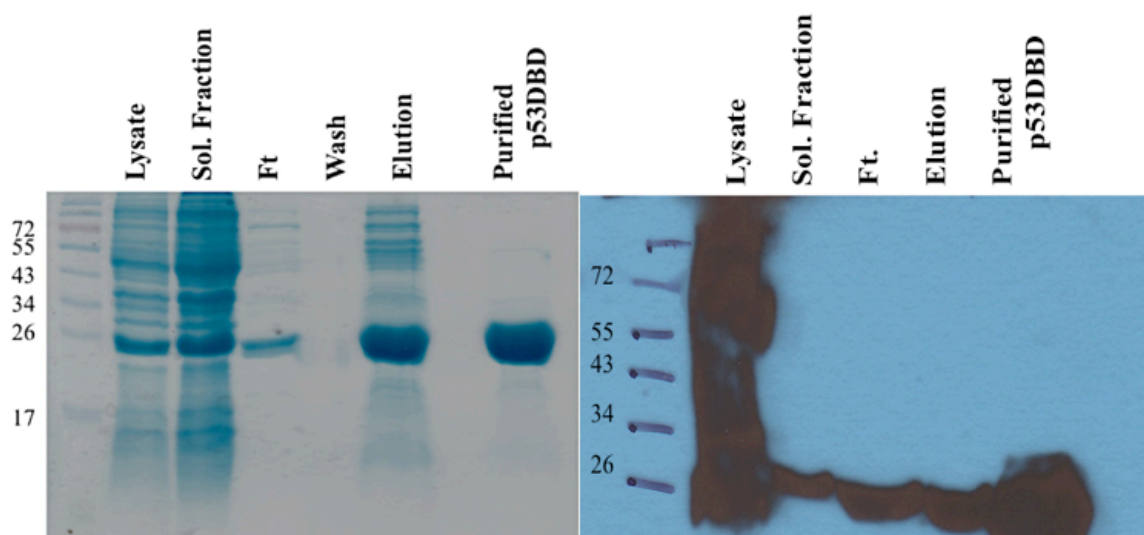


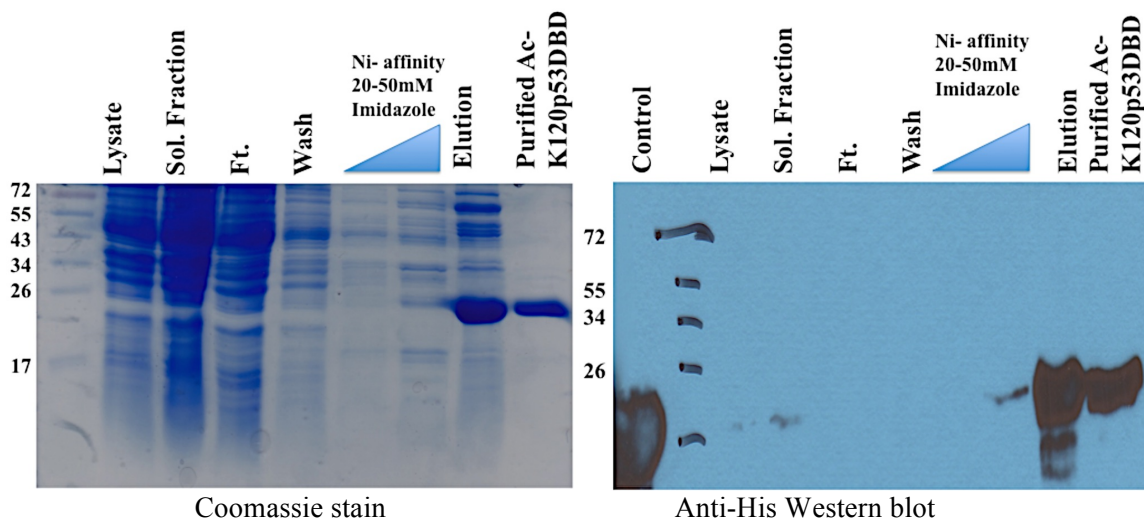
Figure 4.2. Mass spectrometry analysis of the p53-derived peptides from trypsin-digested AcLys120-p53DBD.



Coomassie stain

WB: anti His

Figure 4.3. Purification of recombinant human His-tagged p53DBD using affinity and size exclusion chromatography. SDS-PAGE analysis, Coomassie stained SDS-PAGE and western blot, showed that p53DBD migrated below 26 kDa in a 15% SDS-PAGE.



Coomassie stain

Anti-His Western blot

Figure 4.4. Purification of recombinant human His-tagged acetylated AcLys120-p53DBD using affinity and size-exclusion chromatography. SDS-PAGE analysis, Coomassie-stained SDS-PAGE and Western blot, showed that AcLys120-p53DBD migrated below the 26 kDa in a 15% SDS-PAGE as in p53DBD.

Size exclusion chromatography Elution profile p53DBD

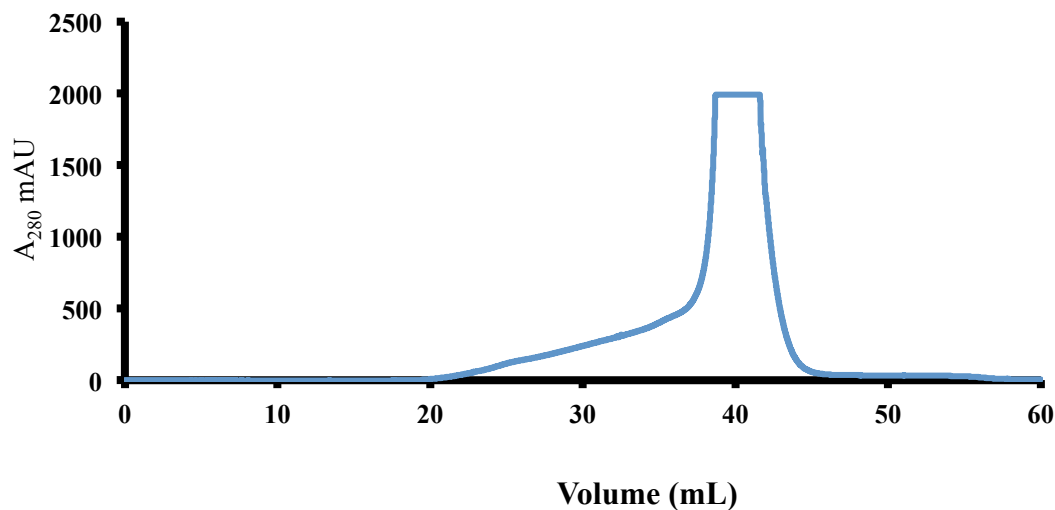


Figure 4.5. Elution profile of the p53DBD protein. The chromatogram showed that p53DBD eluted at 36 ml and peaked at 40 ml.

Size exclusion chromatography Elution profile AcLys120-p53DBD

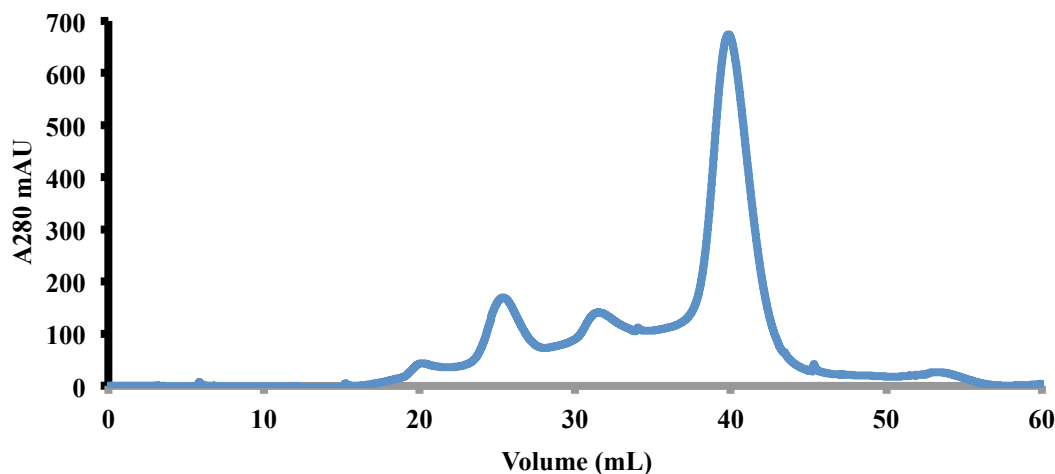


Figure 4.6. Elution profile of AcLys120-p53DBD. The chromatogram showed that AcLys120-p53DBD eluted at 36 ml and peaked at 40 ml. As seen from Figure 4.5 the elution profile from the affinity column of p53DBD is similar to the one I observed for the acetylated protein. The behavior of both proteins in the size-exclusion column is identical

DNA Binding Affinity of the p53DBD and AcLys120-p53DBD Proteins

The acetylation of Lys120 in p53 has been postulated as a mechanism to regulate the activation of apoptotic genes [99] [100]. As Lys120 in loop L1 of p53 often establishes hydrogen bonds with the nucleotides in the position 2 and/or 3 of each quarter-site response element [36], I decided to investigate the role of Lys120 acetylation in binding to half-site response elements with modifications in the first three nucleotides of each quarter-site response element. Specifically, I measured DNA binding affinity of the p53DBD, acetylated and not acetylated, to half-site response elements by changes in the fluorescence polarization. The sequences of the half-site response elements used were: 5'-GGGCATGCCC-3', 5'-AAACATGTTT-3', 5'-CCCCATGGGG-3', and 5'-TTTCATGAAA-3' (Table 4.1. shows the name and sequence of the used sequences). I analyzed the experimental data in the GraphPad Prism software, as already described in Chapter 2.

The DNA binding affinity of p53DBD towards the studied response elements was not affected by changes in the first three nucleotides of each quarter-site response elements (Figure 4.7.). When p53DBD was not acetylated, the dissociation constants were almost the same for all of them: 1.10 μM for GGGCA, 1.07 μM for AAACA, 0.95 μM for CCCCA, and 1.04 μM for TTTCA. On the other hand, the cooperativity of binding as measured by the Hill coefficient decreased when the triplets AAA, CCC, and TTT replaced GGG. For the AcLys120-p53DBD, the dissociation constants were 2.31 μM for GGGCA, 2.67 μM for AAACA 1.78 μM for CCCCA, and 1.56 μM for TTTCA (Figure 4.8.).

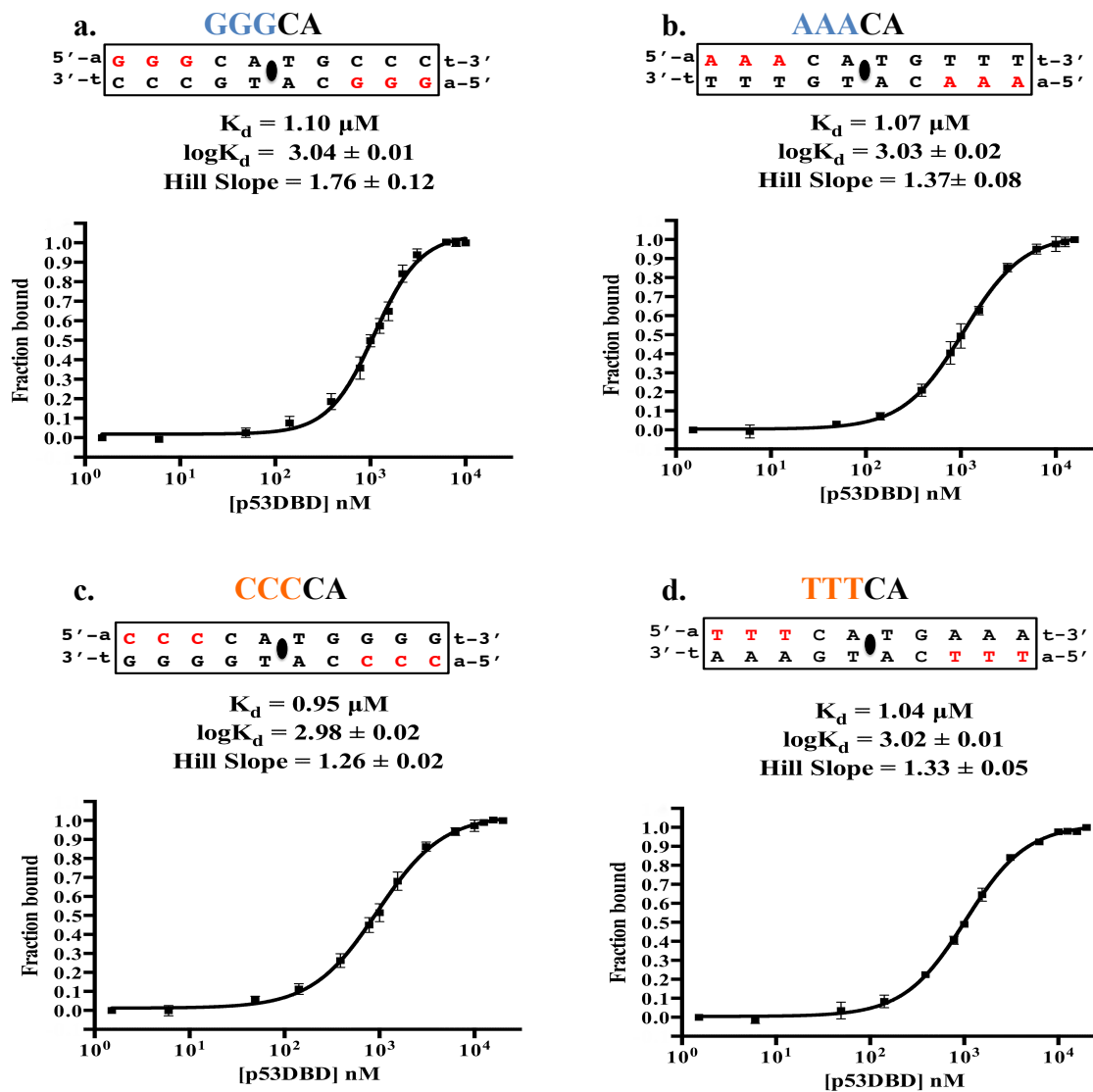


Figure 4.7. DNA binding affinity graphs of p53DBD bound to half-site response elements with modifications in the first three nucleotides of each quarter-site response element.

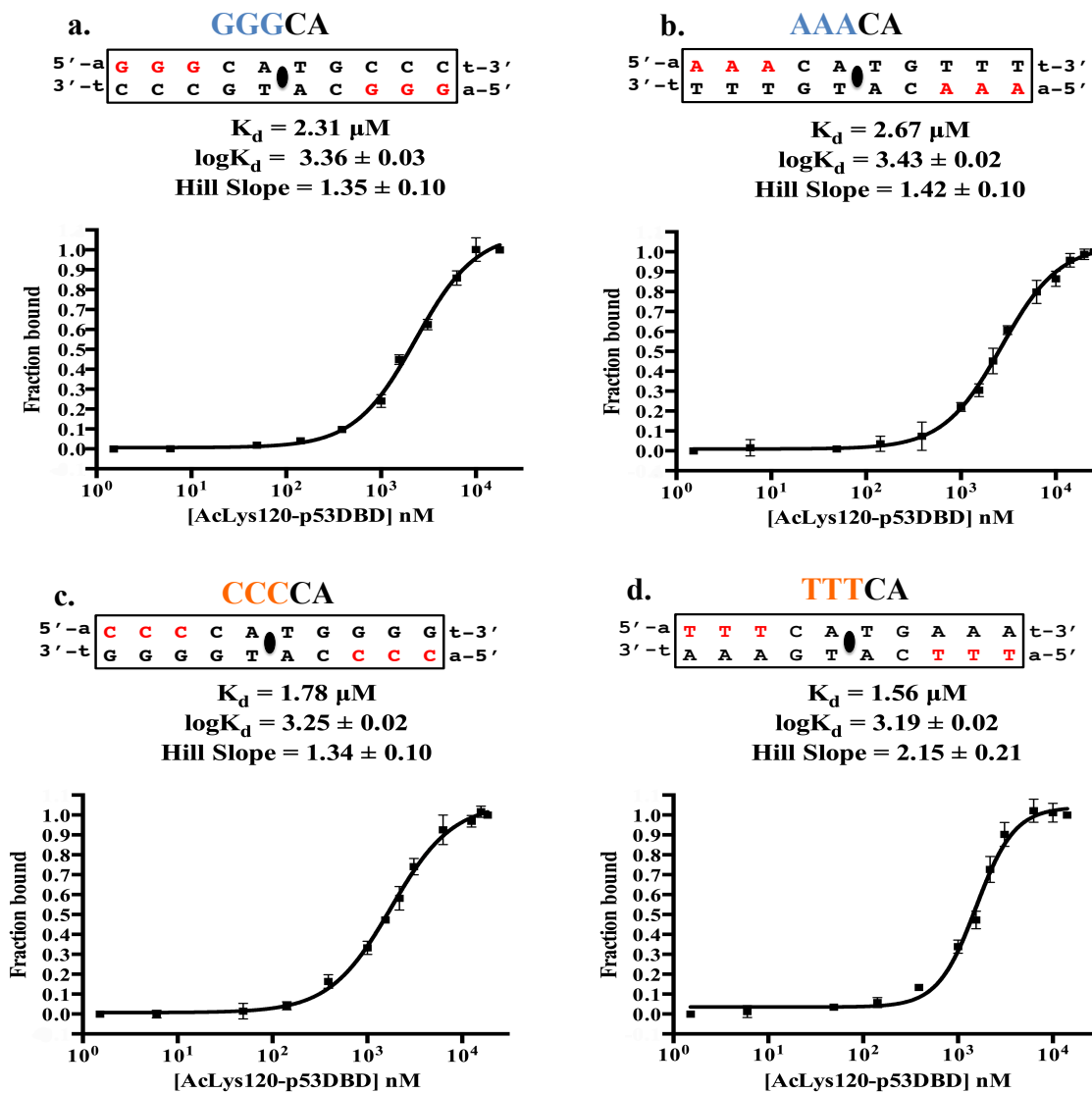
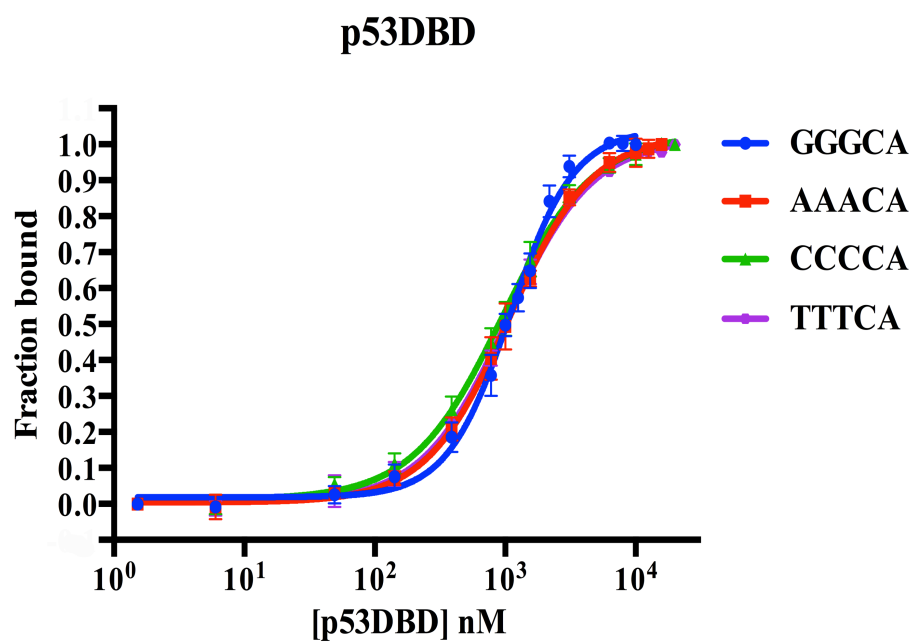


Figure 4.8. DNA binding affinity graphs of AcLys120-p53DBD bound to half-site response elements with modifications in the first three nucleotides of each quarter-site response element.

a.



b.

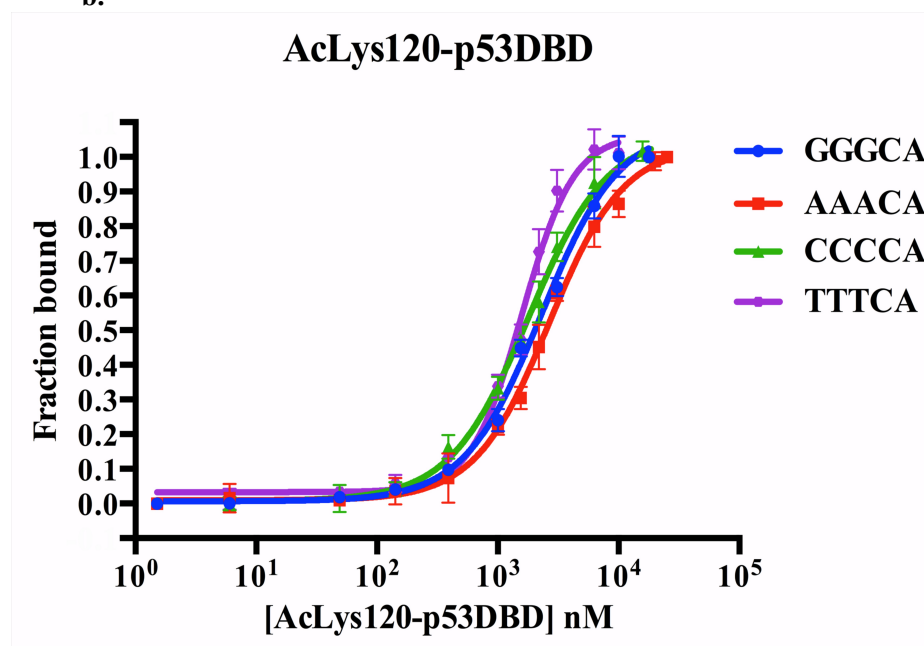


Figure 4.9. Summary of the binding experiments of p53DBD and AcLys120-p53DBD bound to half-site response elements. a. Summary of p53DBD bound to four different sequences. b. Summary of AcLys120-p53DBD bound to four different sequences.

Table 4.3. Binding affinity of p53DBD and AcLys120-p53DBD bound to half-site response element

Sequence	p53DBD		Ac-K120p53DBD	
	K_d (μ M)	Hill coefficient	K_d (μ M)	Hill coefficient
GGGCA	1.10	1.76 ± 0.12	2.31	1.35 ± 0.10
AAACA	1.07	1.37 ± 0.08	2.67	1.42 ± 0.10
CCC CA	0.95	1.26 ± 0.02	1.78	1.34 ± 0.10
TTT CA	1.04	1.33 ± 0.05	1.56	2.15 ± 0.21

Data collection

As mentioned before, crystals were diffracted at beamline BL7-1 of the Stanford Synchrotron Radiation Lightsource at a wavelength of 1.127 Å (Figure 4.10.). The diffraction data were collected on an ADSC-Q315 CCD detector. I solved the structures of three p53DBD-DNA complexes with three different half-site response elements. As shown in Table 4.2, I named the structures using the quarter-site response element that it is unique to each crystal structure: AGGCA, GGACA and TTTCA, respectively. The three crystals belong to the same $P2_1P2_1P2_1$ space group in an orthorhombic unit cell with unique unit cell lengths and three identical 90° angles ($a \neq b \neq c$ and $\alpha = \beta = \gamma = 90^\circ$). The AGGCA, GGACA and TTTCA crystals diffracted at resolutions of 1.95 Å, 2.22 Å, and 1.92 Å, respectively. The three data sets had an I/σ ratio greater than two in the last high-resolution shell (Table 4.4.). The data completeness for each data set was very high: 100% for the TTTCA crystals and almost 100% for the AGGCA and GGACA ones. Furthermore, the spread of measurement in equivalent reflections or R-merge value of the last data range, for the three structures fell in the acceptable range according to standard criteria (30-40% in the highest resolution shell). The data redundancy or multiplicity, which is the average number of measurements per unique reflection, for the three data sets was higher than 4 in all the resolution shells. Overall, the data collection statistics show high quality diffraction data sets for the three crystals.

Molecular Replacement and Refinement

The three structures were solved by molecular replacement using a dimer of p53DBD bound to a half-site response element (PDBID: 4GTG) as a search model. The initial electron density map of each structure was calculated using the phases from the molecular replacement solution and the experimentally observed structure factors. The number of measured reflections used during refinement was: 36,720 for the AGGCA complex, 24,154 for the GGACA one, and 37,838 for the TTTCA one.

After several cycles of automatic refinement in Phenix and manual refinement in Coot, I obtained an Rwork (a measure of the discrepancy between the model and the experimental data) of 17.6% for AGGCA, 17.6% for GGACA, and 18% for TTTCA. Similarly, the final Rfree (the disagreement of the model with 5% of the experimental data not included in the refinement) were 21.8% for AGGCA, 22.2% for GGACA, and 22.1% for TTTCA. The overall B-factor, a parameter reflecting the mobility of the atom or the precision of the atom's position, for the three structures solved was 24.5 Å² for AGGCA, 26.5 Å² for GGACA and 27.7 Å² for TTTCA. It is interesting to note that the TTTCA structure, in spite of having data at higher resolution, it has a higher B-factor than the AGGCA and GGACA structures. This is probably consequence of having a different conformation that results in different crystal packing than the AGGCA and GGACA molecules. All the polypeptide backbone torsion angles in the three structures are in allowed regions of the Ramachandran plot, which is an important indicator of the quality of the final model [101]. Table 4.4 showed the data-collection and refinement statistics of the three structures solved.

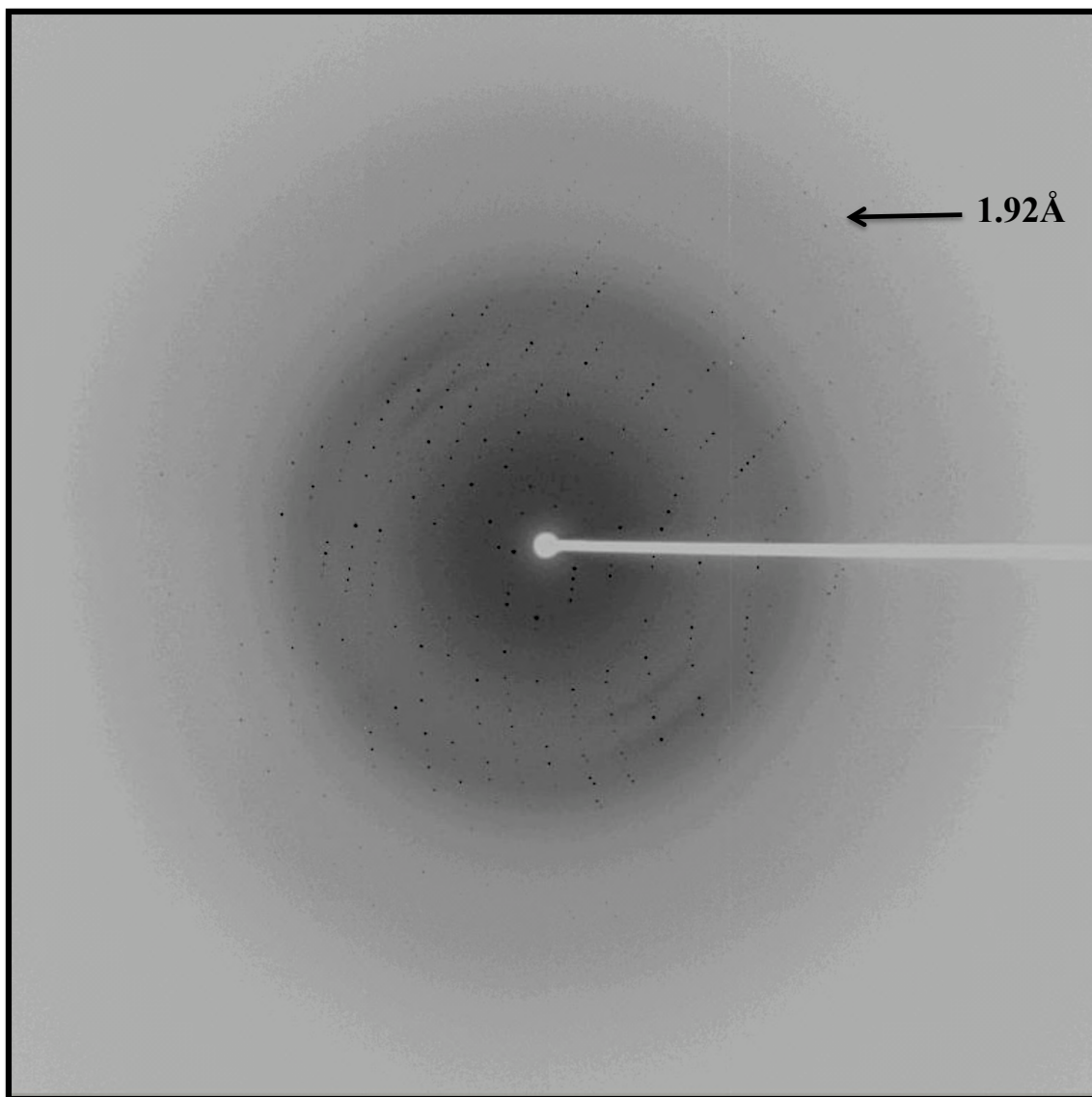


Figure 4.10. Diffraction Pattern of the p53DBD-TTTCA Complex at 1.92Å Resolution. The data were collected in the beamline 7-1 of the Stanford Synchrotron Radiation Lightsource.

Table 4.4. Data collection and refinement statistics of the three-complexes

	AGGCA complex 26 mer	GGACA complex 26 mer	TTTCA complex 26 mer
Data collection			
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
a (Å)	44.9	45	44.1
b (Å)	91.2	91.0	94.2
c (Å)	120.4	119.1	119.7
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50- 1.95 (1.98- 1.95)	50- 2.22 (2.26- 2.22)	50- 1.92 (1.95- 1.92)
R _{sym} or R _{merge}	9.6 (30.6)	9.7 (35.6)	7.6 (38.8)
I/ σ	25.1 (5.9)	27.3 (6.5)	36.5 (6)
Completeness (%)	99.5 (99.7)	99.9 (99.5)	100 (100)
Redundancy	4.7 (4.7)	7.0 (6.8)	7.2 (7.2)
Refinement			
Resolution (Å)	36.74- 1.95	42.51-2.22	43.82- 1.92
No. reflections	36720	24154	37838
R _{work} /R _{free}	17.6/21.8	17.6/22.2	18/22.1
Molecules in the asymmetric unit			
Protein/dsDNA (26bp)	2.0/1.0	2.0/1.0	2.0/1.0
No. of atoms	4201	3963	4168
Protein and DNA	3700	3660	3687
Zn ²⁺ ion	2	2	2
Water	499	301	479
Average B-factors	24.5	26.5	27.7
Protein and DNA	24.2	26.3	27.4
Metal ions	17.1	22.3	24.9
Water	26.5	28.4	30.5
R.m.s. deviations			
Bond lengths (Å)	0.007	0.008	0.008
Bond angles	1.07	1.07	1.09
Ramachandran Plot (%)			
Resd. in most favored region	91.5	93.2	89.2
Resd. in additional allowed region	8.2	6.5	10.5
Resd. in generously allowed region	0.3	0.3	0.3
Resd. in disallowed region	0	0	0
Molprobrity score	1.14	1.12	1.26
Molprobrity percentile	100	100	99

Crystal packing

The unit cell of each of the three crystal forms studied contains four asymmetric units, since four equivalent positions exist in the $P2_1P2_1P2_1$ space group. According to the Matthews coefficient estimation of $V_m = 2.3 \text{ \AA}^3/\text{Da}$, each asymmetric unit has a p53DBD dimer bound to half-site response element (Figure 4.11.). Each protein-DNA complex has a 26-mer oligonucleotide with the 5'-end annealed with its 3'-end to form a double-stranded DNA with a TATA single-stranded loop at one end. The resulting double-stranded DNA forms a half-site response element with two quarter-sites that each can bind a p53DBD monomer. The loop in one end of the half-site makes impossible the formation of a tetramer. Consequently, our results explain the dimeric protein-DNA interactions, without tetrameric interactions (Figure 4.12.).

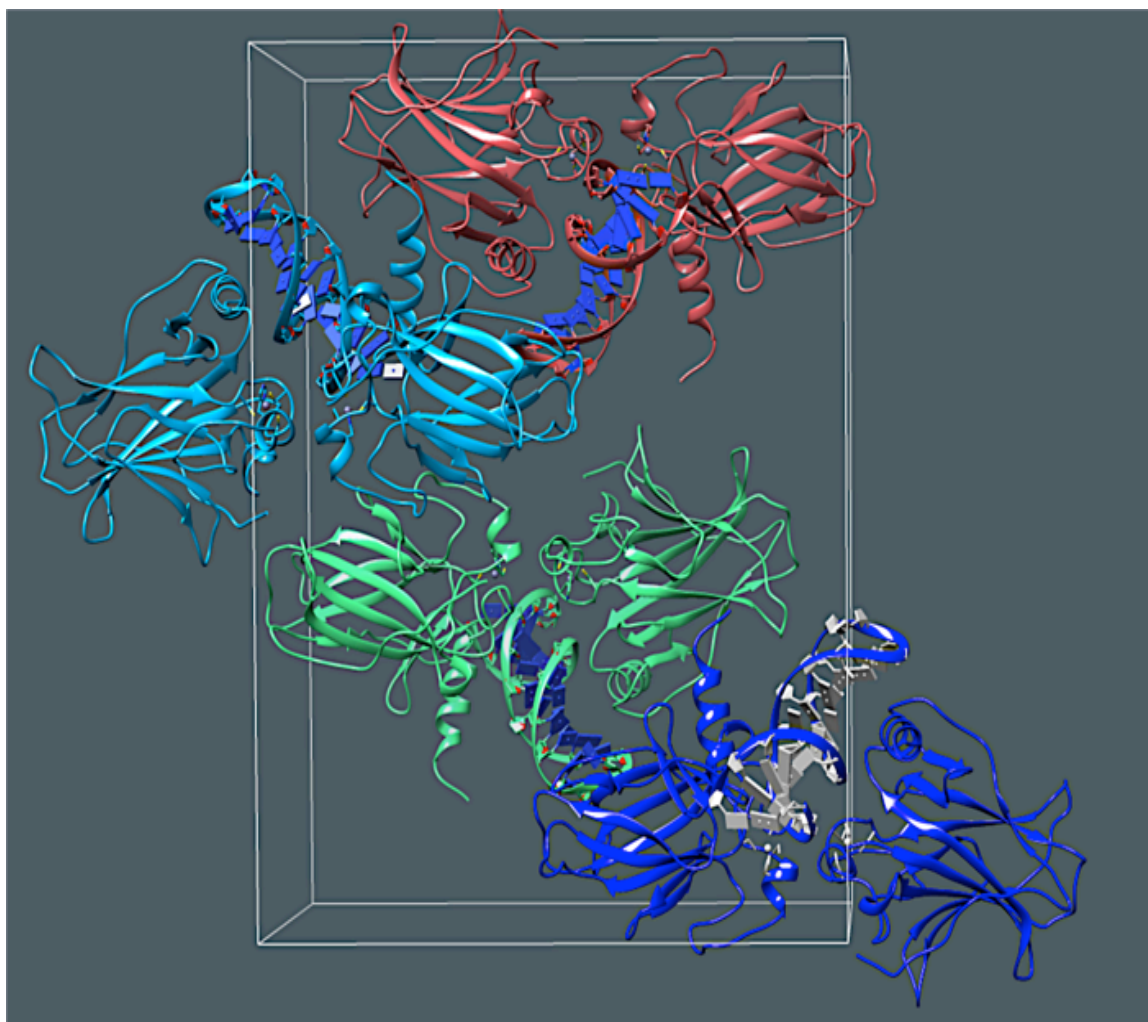


Figure 4.11. Unit Cell Packing. Ribbon model of the dimers packing inside a unit cell.

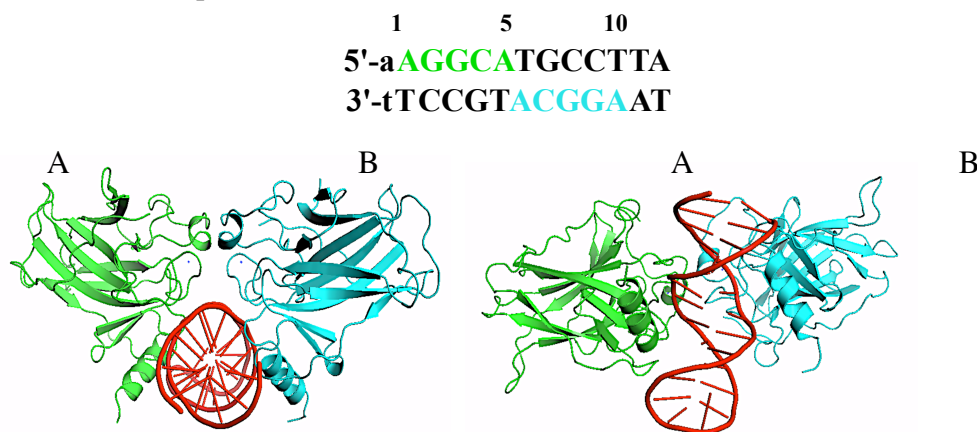
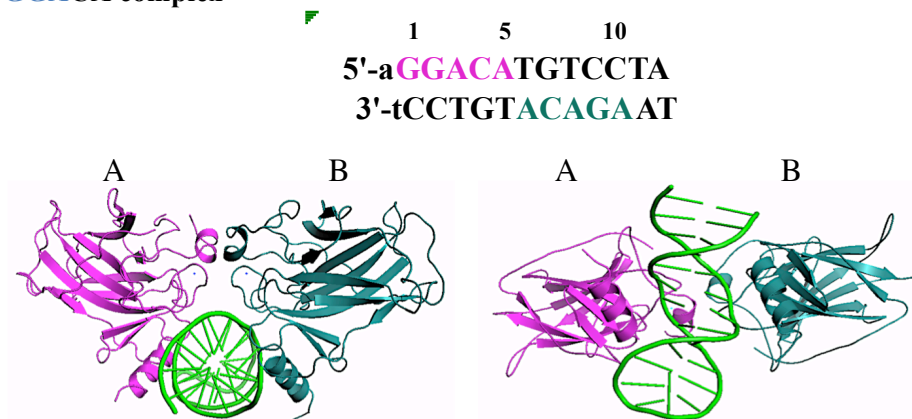
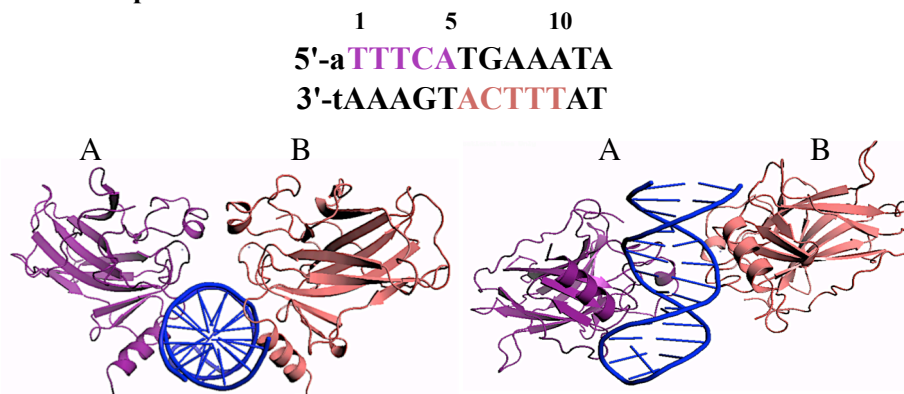
a. **AGGCA complex**b. **GGACA complex**c. **TTTCA complex**

Figure 4.12. Structures of the p53DBD dimers bound to half-site response element. a. Complex AGGCA. b. Complex GGACA. c. Complex TTTCA. Molecules are shown in two different DNA views; the two monomers are indicated as A and B in the three structures

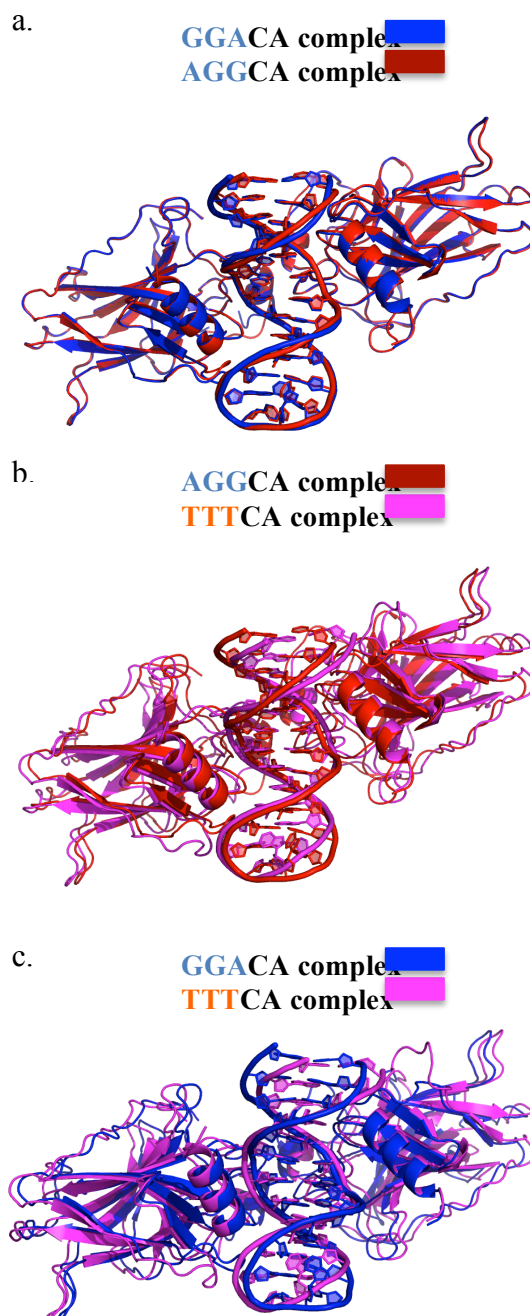
Structure Interpretation

Overall Structural Analysis of the AGGCA and GGACA Activating Response

Element and the TTTCA Repressing Response Element

In order to get information regarding the structural similarities and differences among the three structures, I performed α -Carbons alignment using Pymol (Figure 4.13.). The structures with the activating response elements (AGGCA and GGACA) closely align with an root mean square (rms) deviation of 0.196 Å, whereas alignment of the structure containing the repressing sequence (TTTCA) with any of the activating ones has a significantly large rms deviation of 1.69 Å and 1.75 Å with the AGGCA and GGACA structures, respectively. The main structural differences between activating and repressing structures are located in the protein loops, helix H1 and the DNA (Figure 4.14.).

As observed from Figure 4.15., the p53DBD monomer structure is maintained in all the structures where I observed the tetrahedral coordinated Zn by Cys176 in loop L2, His179 in helix H1, Cys238 and Cys242 in loop L3, which contributes to the overall stability of the p53 DNA binding domain (Figure 4.15.). On the other hand, after the alignment of TTTCA monomers with either AGGCA or GGACA monomers, the root mean square (rms) deviation values slightly increased compare with the rms alignment values of the monomers among the AGGCA and GGACA structures (Table 4.5.).



4.13. Alignment of the three crystal structures. a. AGGCA complex (red) alignment with TTTCA (magenta). b. GGACA complex (blue) alignment with TTTCA complex (magenta). c. AGGCA complex (red) alignment with GGACA (blue).

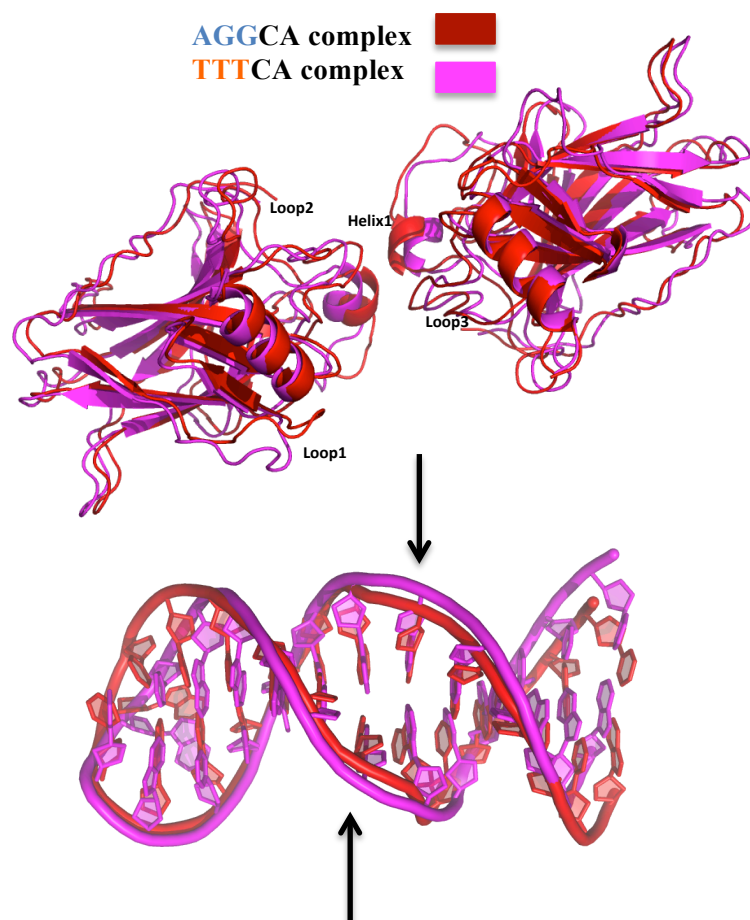


Figure 4.14. Alignment of AGGCA with TTTCA showing main differences in protein loops and helix H1. In the DNA from the AGGCA structure, distortions in the B-form are observed compared with the DNA in the TTTCA structure.

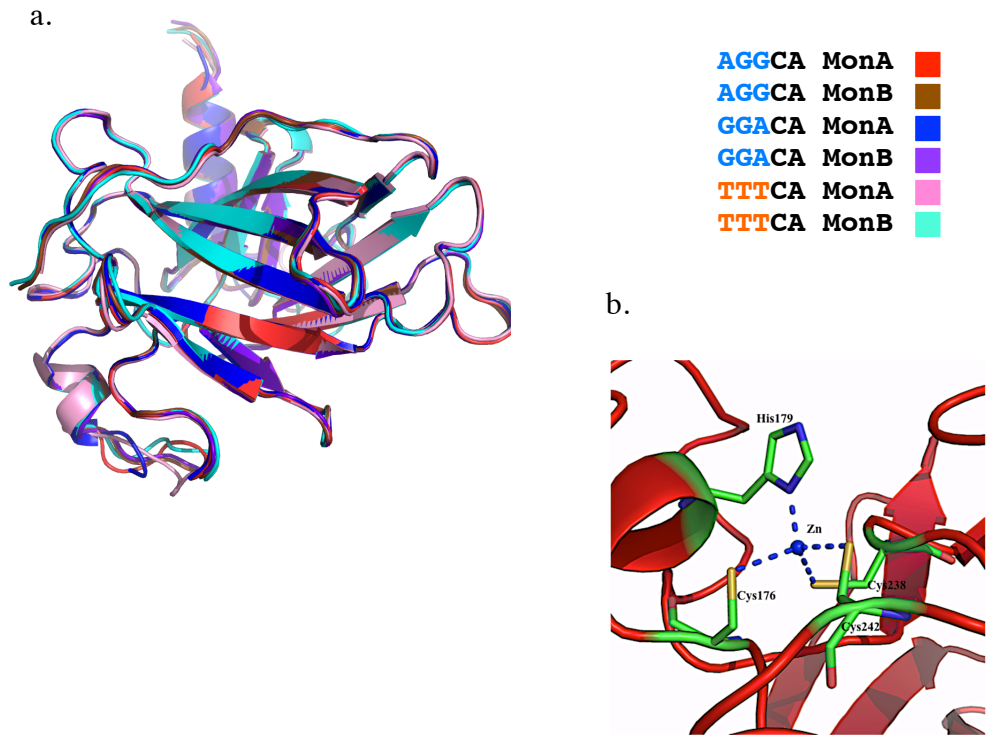


Figure 4.15. Overall Structure Analysis. a. Alignment of all the monomers in the different structures. b. Coordination of Zn ion with residues from L2 (Cys176), L3 (Cys238, Cys242) and H1 (His179) observed in all structures.

Table 4.5. Root mean square deviation of comparing the monomers in AGGCA, GGACA and TTTCA structures

rms deviations	AGGCA		GGACA		TTTCA	
	MonA	MonB	MonA	MonB	MonA	MonB
AGGCA MonA		0.1	0.1	0.2	0.2	0.2
AGGCA MonB	0.1		0.2	0.1	0.3	0.2
GGACA MonA	0.1	0.2		0.2	0.3	0.2
GGACA MonB	0.2	0.1	0.2		0.3	0.2
TTTCA MonA	0.2	0.3	0.3	0.3		0.3
TTTCA MonB	0.2	0.2	0.2	0.2	0.3	

To further characterize the discrepancies observed in the alignment, I calculated, for each structure, the distance between the center of mass (COM) of the monomers using the CALCOM software [102]. The monomers bound to a repressing response element are separated by 45 Å, while the ones bound to activating sequences are separated by only 42 Å (Figure 4.16.). Moreover, if one looks along the axis formed by the length of the DNA molecule, the angle formed between the two monomers in each dimer is 101.7° in the AGGCA structure, 102° in the GGACA structure and 110.3° in the TTTCA structure (Figure 4.16.). Together, these comparisons indicate that the monomers in the TTTCA structure are separated 3 Å and 8° more with respect to the monomer arrangement found in the AGGCA and the GGACA structures. Such rearrangement already explains the larger unit b axis in the unit cell of the TTTCA structure.

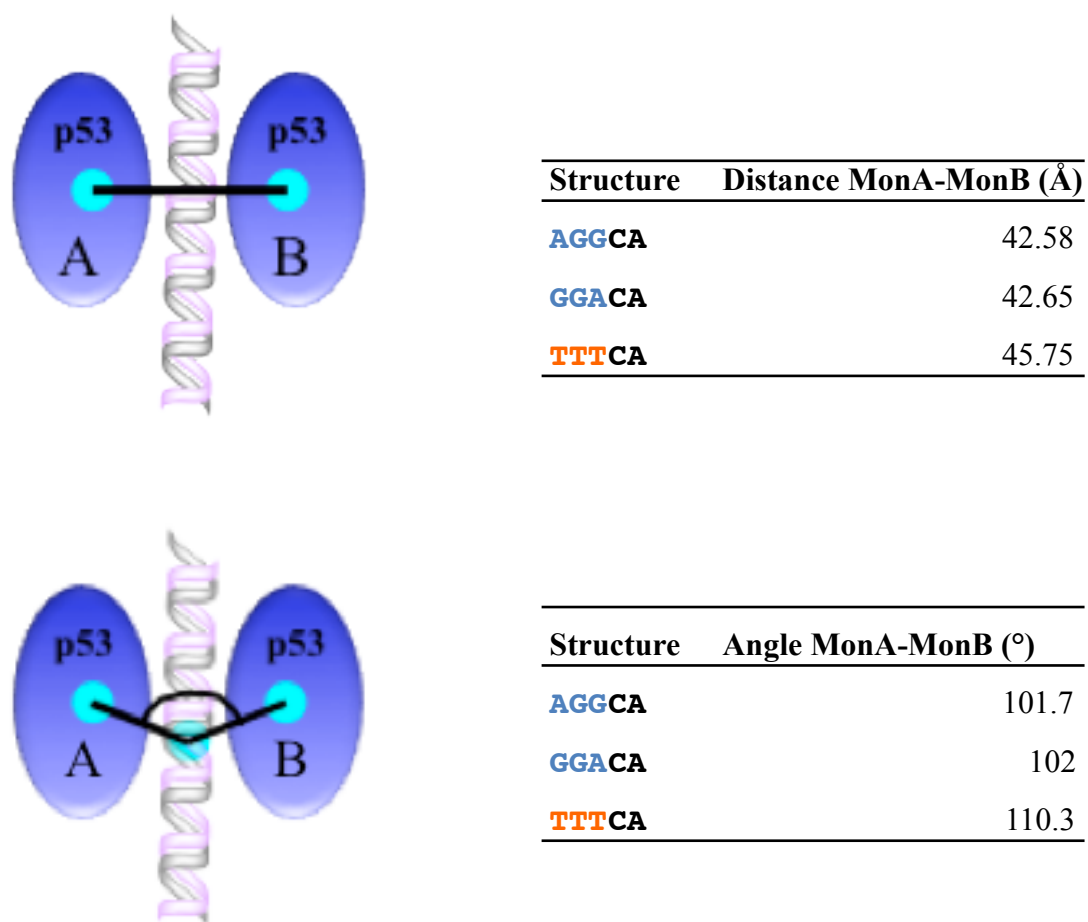


Figure 4.16. Distances and angles between the centers of mass (COM) of the two monomers in the p53DBD structures.

Protein- Protein interactions

The p53 DBD monomer dimerizes upon DNA binding and the interactions between both monomers are located in loop L3 and helix H1 of each p53 DBD. These monomer-monomer interactions are Van der Waals contacts and hydrogen bonds (Figure 4.17.).

Although the interacting secondary elements are the same for the three complexes, the number of residues forming the dimeric interface is quite different in each structure (Figure 4.17.). In the AGGCA complex, the residues involved in the dimeric interactions were Pro177, His178, His179 and Glu180 from helix H1, Met243 and Gly244 from loop L3, and Arg181 from loop L2 (Table 4.4.). In the GGACA complex, neither His179 nor Glu180 in helix H1 contribute to the monomer-monomer interface (Table 4.6.). Instead, in the TTTCA complex only His178 from helix H1, Arg181 from loop L2, and Met243 from loop L3 form the dimeric interface (Table 4.6.). To get more information about the interface of p53DBD monomers, I calculated the solvent accessible surface area (SASA) in each structure using Areaimol [103][104], for AGGCA and GGACA structures, the area accessible to solvent was pretty similar of 22,855.9 Å² and 23,143.1 Å², respectively, whereas the SASA for TTTCA dramatically decreased to 16,099.8 Å², which explains the reduce number of residues interacting in the intra-face of the dimer. The change in the number of residues forming the monomer-monomer interface result in two dimer conformations: a closed one for the GGACA and AGGCA structures and an opened one for the TTTCA structure (Figure 4.16.). In brief, the residues involved in the dimeric interface are similar among the three complexes; however, the number of contacts between these residues is greatly reduced in the TTTCA dimer in comparison with the AGGCA and GGACA dimers.

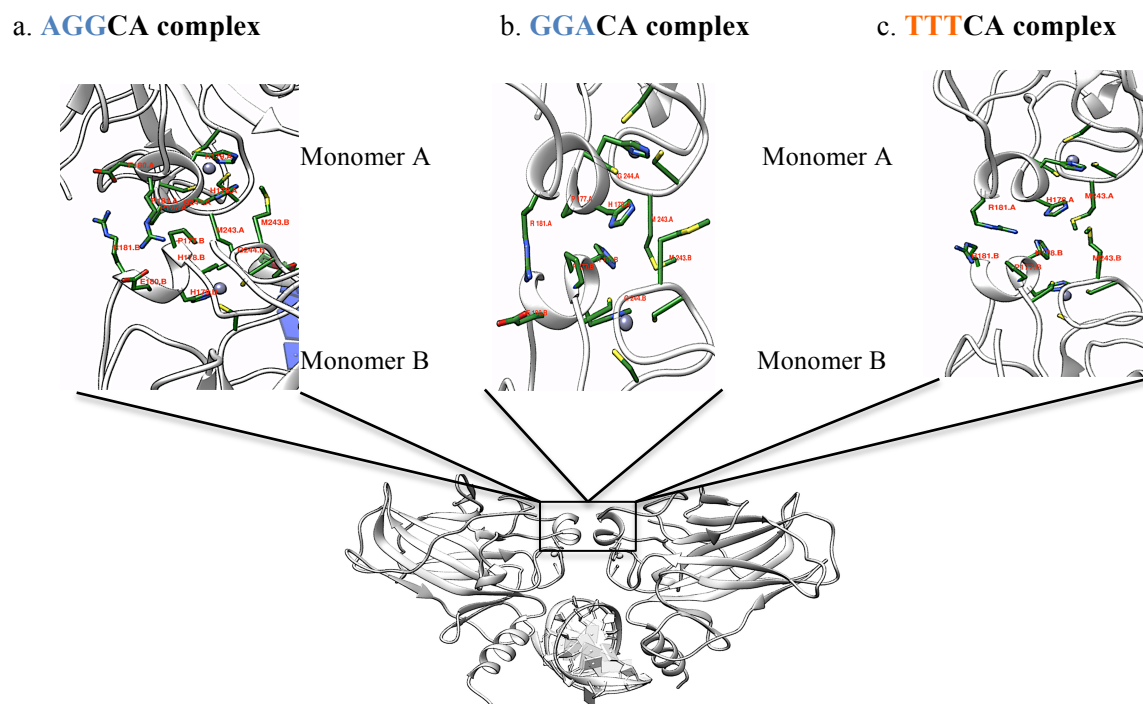


Figure 4.17. Monomer-monomer interactions. a. AGGCA complex interactions of the two monomers. b. GGACA complex dimer interface contains less atom-atom contacts than AGGCA. c. TTTCA complex has the least number of interactions between the monomer

Table 4.6. Monomer-monomer interactions in the AGGCA complex

Residue	Chain A number	Atom	Residue	Chain B number	Atom	Distance
ARG	181	NH1	GLU	180	OE1	2.4
ARG	181	NH1	GLU	180	CD	3.2
GLU	180	OE1	ARG	181	NH1	3.3
HIS	178	CD2	MET	243	CE	3.4
MET	243	CE	HIS	178	CD2	3.4
ARG	181	CZ	GLU	180	OE1	3.4
HIS	178	NE2	MET	243	CE	3.5
MET	243	CE	HIS	178	NE2	3.5
GLY	244	CA	HIS	178	ND1	3.5
HIS	178	ND1	GLY	244	CA	3.6
GLY	244	CA	HIS	178	CE1	3.7
HIS	178	CE1	GLY	244	CA	3.7
ARG	181	CZ	PRO	177	CB	3.7
ARG	181	NH2	PRO	177	CG	3.8
MET	243	CE	HIS	179	CE1	3.8
PRO	177	CG	HIS	178	CA	3.8
ARG	181	NH1	PRO	177	CA	3.8
HIS	178	CA	PRO	177	CG	3.8
MET	243	CE	HIS	179	NE2	3.8
PRO	177	CB	ARG	181	CD	3.8
ARG	181	CZ	PRO	177	CG	3.9
ARG	181	NE	PRO	177	CB	3.9
HIS	179	CE1	MET	243	CE	3.9
PRO	177	CB	HIS	178	CA	4.0
PRO	177	CB	PRO	177	CB	4.0
HIS	178	CD2	MET	243	CB	4.0
HIS	178	CA	PRO	177	CB	4.0
PRO	177	CG	HIS	178	CB	4.1
HIS	178	CB	PRO	177	CG	4.1

Table 4.7. Monomer-monomer interactions in the GGACA complex

Residue	Chain A number	Atom	Residue	Chain B number	Atom	Distance
ARG	181	NH2	GLU	180	OE1	3.1
HIS	178	ND1	GLY	244	CA	3.3
MET	243	CG	HIS	178	ND1	3.4
MET	243	CG	HIS	178	CE1	3.4
GLY	244	CA	HIS	178	CD2	3.5
HIS	178	CE1	GLY	244	CA	3.5
HIS	178	CE1	GLY	244	N	3.6
HIS	178	ND1	GLY	244	N	3.6
GLY	244	CA	HIS	178	NE2	3.7
PRO	177	CG	HIS	178	CA	3.8
HIS	178	CA	PRO	177	CG	3.8
ARG	181	CD	PRO	177	CB	3.8
HIS	178	N	PRO	177	CB	3.8
PRO	177	CB	PRO	177	C	3.9
HIS	178	CA	PRO	177	CB	3.9
PRO	177	CB	HIS	178	N	3.9
PRO	177	CB	HIS	178	CA	3.9
HIS	178	CD2	MET	243	CB	4.0
PRO	177	CB	PRO	177	CB	4.0

Table 4.8. Monomer-monomer interactions in the TTTCA complex

Residue	Chain A number	Atom	Residue	Chain B number	Atom	Distance (Å)
ARG	181	CZ	PRO	177	CB	3.3
HIS	178	NE2	MET	243	CE	3.5
HIS	178	CD2	MET	243	SD	3.5
HIS	178	CE1	MET	243	CE	3.6
ARG	181	CD	ARG	181	NH2	3.6
HIS	178	NE2	MET	243	SD	3.7
ARG	181	NH1	PRO	177	CB	3.8
MET	243	CE	HIS	178	CE1	4.0

DNA analysis

The DNA molecule in the three structures has a well-defined electron density map where I observed that the 14th base flipped away from the helical axis in all three complexes (Figure 4.18.). The 14th base is part of the loop formed when the 5' and 3' ends anneal to form the double stranded DNA.

A hallmark in the DNA of AGGCA and GGACA complexes is the central base pair nucleotide, A-T, adopting a Hoogsteen base pairing geometry, whereas the DNA in the TTTCA complex showed the canonical Watson and Crick base pairing conformation. Figure 4.19. showed the different geometry adopted by the two central base pair, A-T, in the three different complexes. In Hoogsteen geometry, N7 of the adenine makes a hydrogen bond with N3 of the thymine, this interaction is different in Watson and Crick geometry where N1 of the adenine is the one making a hydrogen bond with N3 of thymine as observed in Figure 4.20. Consequently, in the Hoogsteen pairing of A-T, the glycosyl bond conformation changed to syn in the adenine compare with anti conformation in a Watson and Crick geometry.

The Hoogsteen geometry has been observed in p53 DBD complexes [69] [38] where one of the consequences of having this geometry is the narrowing of the minor groove (Figure 4.21.) that is supported by our results as well. I observed that the distance between carbon1 of the ribose from adenine to the carbon1 of the ribose from thymine of the two base pairs in the dodecamer sequence was 8.4 and 8.4 for the AGGCA complex, and 8.1 and 8.2 for the GGACA complex which agreed with the distance showed in a Hoogsteen geometry. In contrast, in the TTTCA complex, the distance between carbon1 of the ribose from adenine to the carbon1 of the ribose from thymine, 10.2 Å and 10.1Å, was similar to the distance of these two same atoms in a Watson and Crick conformation of the central AT pairs (Figure 4.20.). Furthermore, another difference found between the DNA of AGGCA and GGACA with the one of TTTCA complexes was the position and direction values of the central base pair duplets (A-T) relative to its neighbor

pair. As expected, AGGCA and GGACA roll, tilt and twist values were similar, but differ greatly from the B-DNA form (Table 4.9., 4.10.). On the other hand, the three angular values for TTTCA structure indicated a slightly change of location and direction compare to the B-DNA parameters (Table 4.11.). Finally, the slide and opening parameter of AGGCA and GGACA differed by almost 4 and 60 units, respectively, with the TTTCA value showing that AGGCA and GGACA helical axis were shifted (Table 4.9., 4.10., 4.11. and 4.12.). These differences in the parameters describing the location and position of the bases relative to its neighbor show the deformation of the DNA observed in the AGGCA and GGACA structures.

In brief, Hoogsteen base pairing geometry present in the central A-T duplets of the activating-sequence AGGCA and GGACA structures narrowed the minor groove; increased the angular parameters; changed the glycosyl bond conformation of the adenine to syn and shifted the helical axes resulting in a distorted B-DNA form.

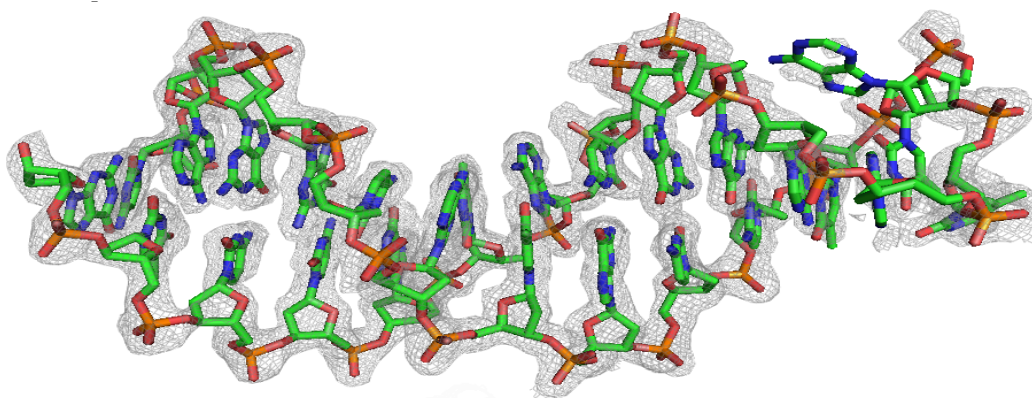
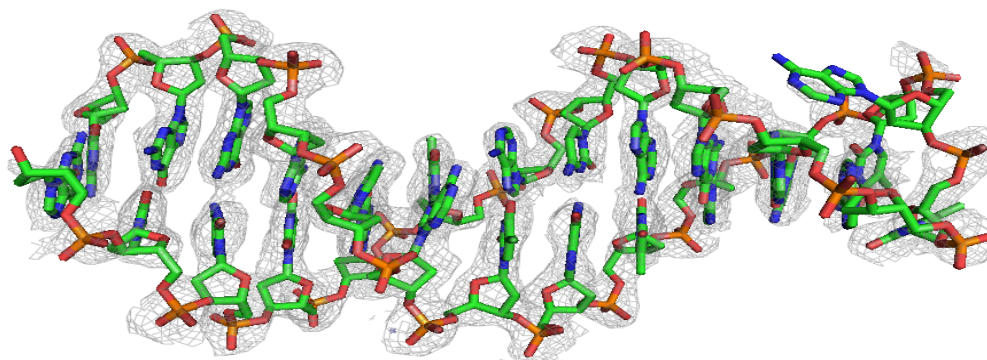
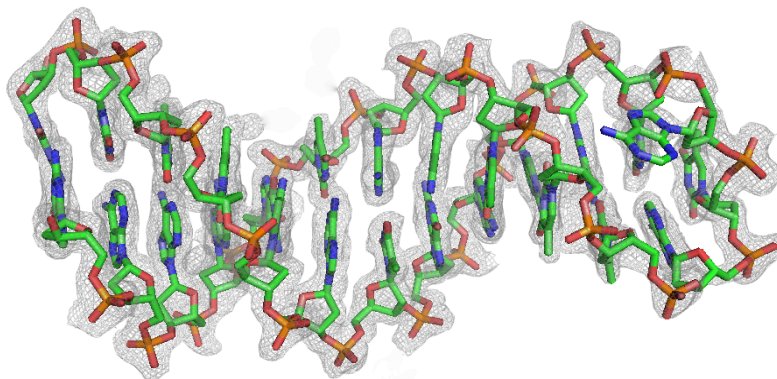
a. **AGGCA**b. **GGACA**c. **TTTCA complex**

Figure 4.18. Electron density map of the DNA. a. Composite omit map $2F_o - F_c$ of the DNA with sequence AGGCA. The electron density is contour at 1σ cutoff. b. Composite omit map $2F_o - F_c$ of the DNA with sequence AGGCA. The 14TH base flipped away from the helical axis as in the other two complexes. c. Composite omit map $2F_o - F_c$ of the DNA with sequence TTTCA.

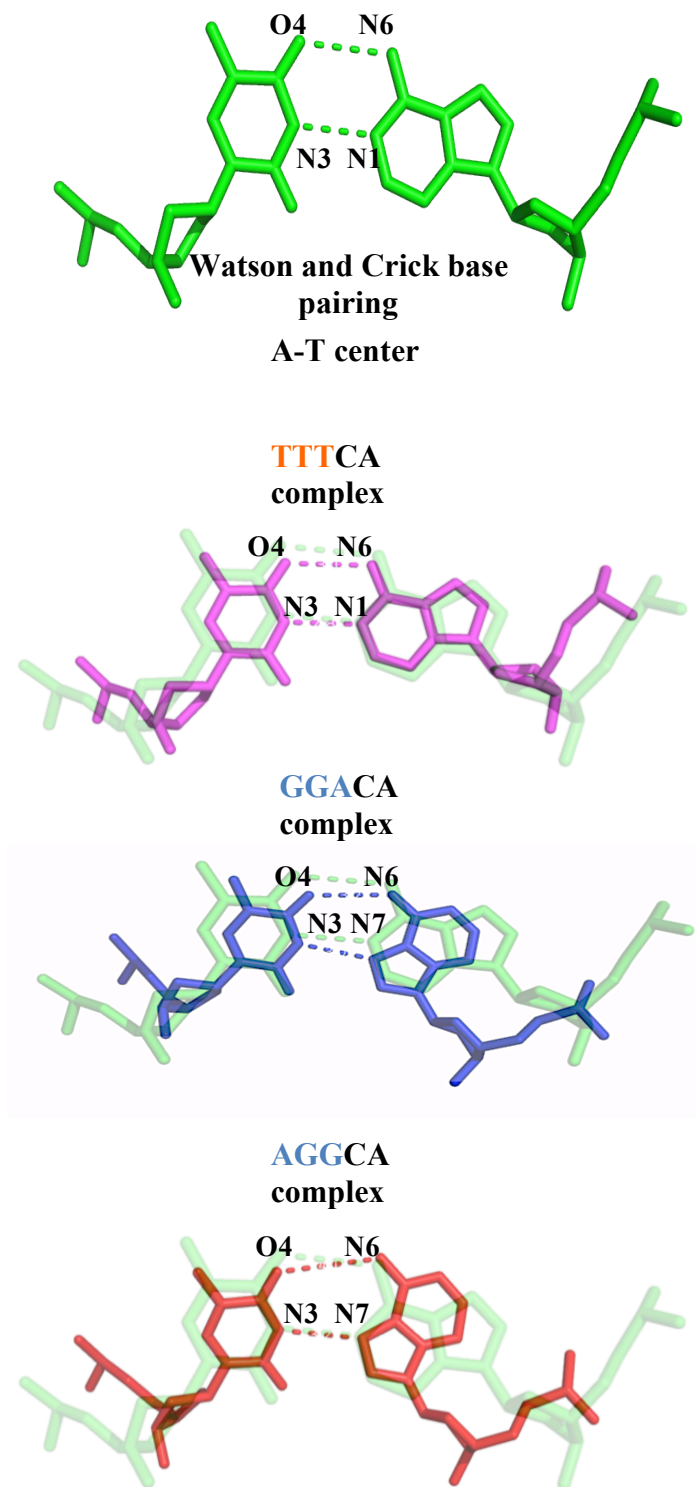


Figure 4.19. TTTCA complex retains the Watson-Crick conformation. In the AGGCA and GGACA complexes, a Hoogsteen conformation of central pair A-T is observed. For simplicity, only one pair in the center is shown.

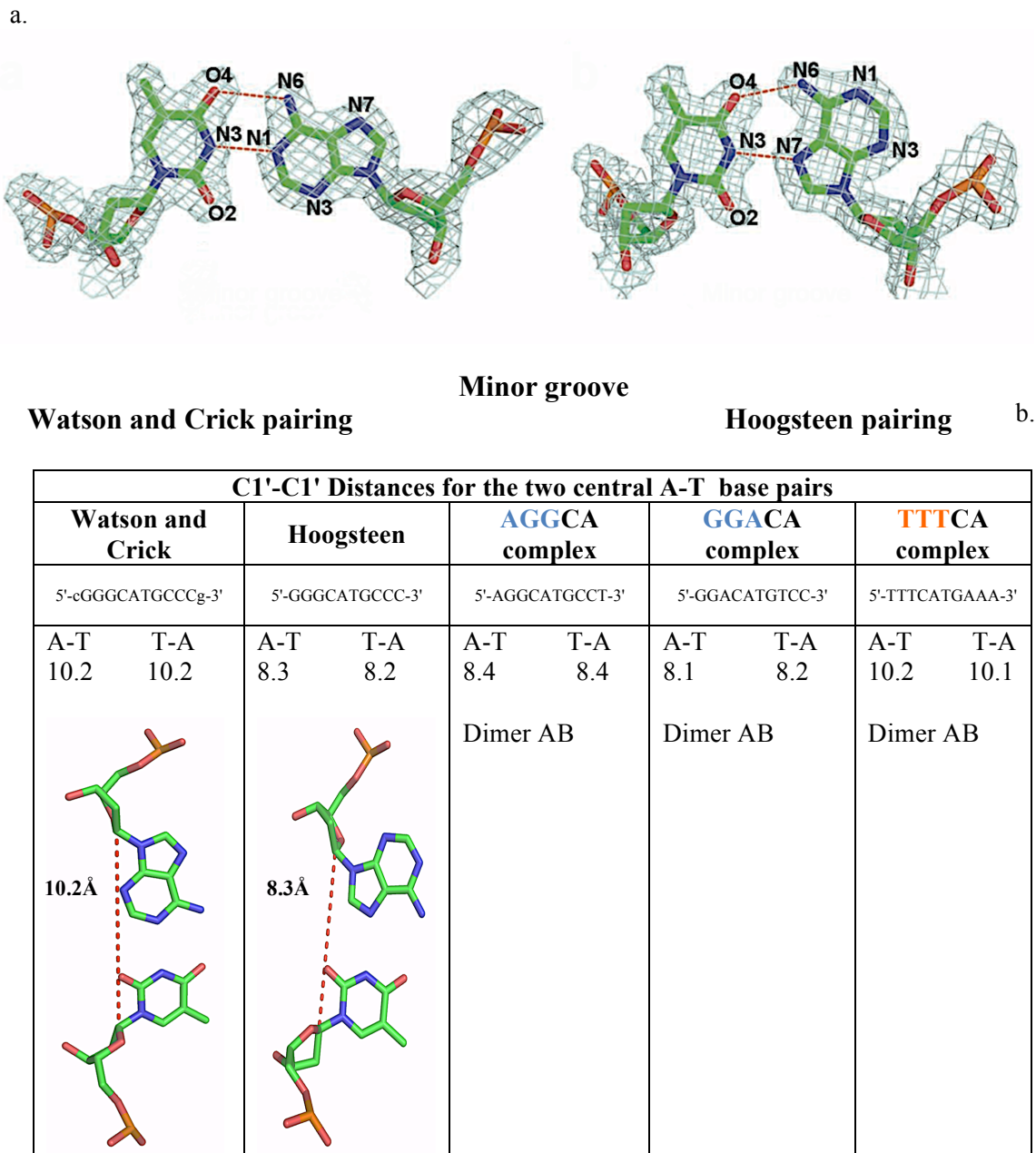
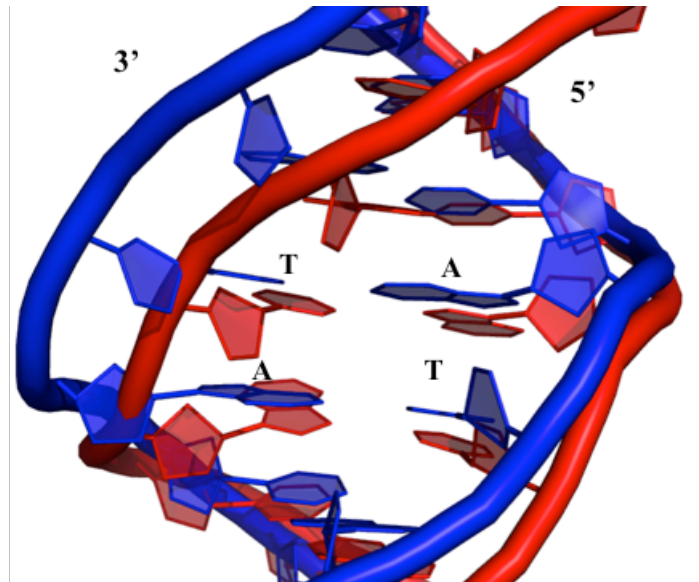
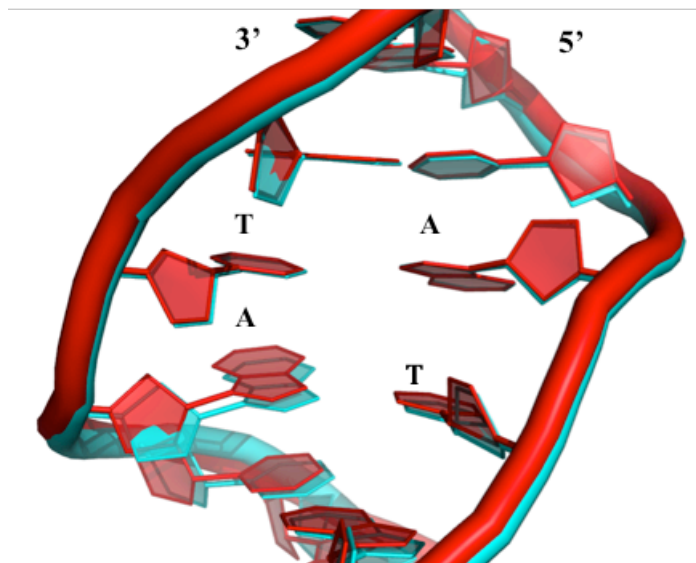


Figure 4.20. Hoogsteen and Watson and Crick Geometry a. Watson and crick base pair A-T from PDB2AC0 and Hoogsteen base pair A-T from PDB3IGK, both are shown with 2Fo-Fc electron density maps at a 1σ cutoff taken from Kitayner et al., 2010. b. Distances of the interstrand carbon 1 of the ribose from the adenine to the carbon 1 of the ribose from the thymine of the two base pairs in the dodecamer sequence. The Watson and Crick base pairing was taken from PDB2AC0 and the Hoogsteen from PDBD4GTG. AGGCA and GGACA complex retain a Hoogsteen conformation, while TTTCA complex presents a Watson and Crick conformation in the two central A-T pair.

a.



b.



Watson and Crick model ■ Hoogsteen model ■ AGGCA complex ■

Figure 4.21. DNA in the AGGCA complex. a. DNA of AGGCA complex minor groove aligned with DNA minor groove containing Watson and Crick pairing (PDB2AC0). b. DNA of AGGCA complex minor groove aligned with minor groove with Hoogsteen pairing (PDB4GTG). The effect of having Hoogsteen base pairing is reflected by a narrow minor groove compare with a Watson and Crick base pairing.

Table 4.9. Geometrical parameters of the **AGGCA** Half-Site DNA and B-DNA

Local base-pair step parameters of AGGCA DNA						
Step	Shift	Slide	Rise	Tilt	Roll	Twist
AG/CT	0.2	-0.3	3.3	2.5	0.2	31.6
GG/CC	1.0	-0.1	3.5	5.0	2.6	32.0
GC/GC	-0.4	-0.2	3.0	1.2	-0.9	36.2
CA/TG	3.3	-3.7	-0.8	-173.0	2.2	175.9
AT/AT	-0.5	-3.8	-0.2	97.0	-135.9	177.7
TG/CA	-0.9	3.3	3.7	-2.3	7.6	-1.0
GC/GC	0.3	-0.6	3.3	0.6	2.1	32.7
CC/GG	-0.4	-0.5	3.4	-3.6	4.0	35.7
CT/AG	0.6	0.0	3.0	1.8	0.3	30.1

Local base-pair step parameters of B-DNA (PDB ID: 1BNA)						
Step	Shift	Slide	Rise	Tilt	Roll	Twist
CG/CG	-0.4	0.2	3.5	-3.4	6.4	40.3
GC/GC	0.5	0.2	3.5	0.8	-4.7	38.2
CG/CG	-0.3	0.7	3.0	3.6	8.0	24.5
GA/TC	0.0	0.1	3.4	-2.7	3.2	40.9
AA/TT	0.1	-0.3	3.3	-0.7	1.0	35.4
AT/AT	0.3	-0.6	3.3	1.8	-2.8	34.8
TT/AA	-0.3	-0.2	3.3	3.0	0.7	35.4
TC/GA	0.0	0.0	3.4	0.3	-0.1	39.3
CG/CG	0.4	0.9	3.2	-3.3	3.9	29.4
GC/GC	-1.3	0.4	3.7	-4.7	-12.2	40.8
CG/CG	0.8	0.1	3.2	3.1	-3.1	32.6

Table 4.10. Geometrical DNA Parameters of the **GGACA** Half-Site DNA and B-DNA

Local base-pair step parameters of GGACA DNA						
Step	Shift	Slide	Rise	Tilt	Roll	Twist
GG/CC	-0.61	0.63	3.29	-2.39	-0.99	35.01
GA/TC	0.46	-0.31	3.48	3.16	-4.94	42.58
AC/GT	0.59	-0.76	3.14	2.63	-2.2	33.33
CA/TG	3.06	-3.77	-0.65	-168.77	4.41	165.19
AT/AT	-0.42	-3.98	-0.92	99.31	-134.87	154.45
TG/CA	-1.03	3.13	3.67	-1.23	10.13	-0.5
GT/AC	-0.7	-0.88	3.09	-2.54	0.98	32.15
TC/GA	-0.1	-0.61	3.3	-1.9	1.55	33.92
CC/GG	0.17	-0.57	3.16	-0.86	6.29	29.8

Local base-pair step parameters of B-DNA (PDB ID: 1BNA)						
Step	Shift	Slide	Rise	Tilt	Roll	Twist
CG/CG	-0.4	0.2	3.5	-3.4	6.4	40.3
GC/GC	0.5	0.2	3.5	0.8	-4.7	38.2
CG/CG	-0.3	0.7	3.0	3.6	8.0	24.5
GA/TC	0.0	0.1	3.4	-2.7	3.2	40.9
AA/TT	0.1	-0.3	3.3	-0.7	1.0	35.4
AT/AT	0.3	-0.6	3.3	1.8	-2.8	34.8
TT/AA	-0.3	-0.2	3.3	3.0	0.7	35.4
TC/GA	0.0	0.0	3.4	0.3	-0.1	39.3
CG/CG	0.4	0.9	3.2	-3.3	3.9	29.4
GC/GC	-1.3	0.4	3.7	-4.7	-12.2	40.8
CG/CG	0.8	0.1	3.2	3.1	-3.1	32.6

Table 4.11. Geometrical DNA Parameters of the **TTT**CA Half-Site

Local base-pair step parameters of TTTCA DNA						
Step	Shift	Slide	Rise	Tilt	Roll	Twist
CT/AG	-0.5	-1.0	3.1	4.7	5.8	33.4
TT/AA	-0.4	-0.6	3.3	1.0	-4.8	36.7
TT/AA	-0.3	-0.4	3.2	-0.2	-5.3	36.0
TC/GA	-0.3	-0.4	3.4	-2.3	-3.0	39.9
CA/TG	0.6	0.7	3.2	2.7	3.5	40.0
AT/AT	0.2	-0.1	2.8	-0.1	6.8	21.0
TG/CA	-0.7	0.8	3.2	-4.0	1.2	41.0
GA/TC	0.3	-0.3	3.5	3.5	-2.4	38.6
AA/TT	0.0	-0.3	3.2	-0.8	-2.1	38.6
AA/TT	0.4	-0.6	3.4	4.1	-8.4	35.6
AT/AT	0.9	-0.7	2.9	-1.5	2.2	22.7

Local base-pair step parameters of B-DNA (PDB ID: 1BNA)						
Step	Shift	Slide	Rise	Tilt	Roll	Twist
CG/CG	-0.4	0.2	3.5	-3.4	6.4	40.3
GC/GC	0.5	0.2	3.5	0.8	-4.7	38.2
CG/CG	-0.3	0.7	3.0	3.6	8.0	24.5
GA/TC	0.0	0.1	3.4	-2.7	3.2	40.9
AA/TT	0.1	-0.3	3.3	-0.7	1.0	35.4
AT/AT	0.3	-0.6	3.3	1.8	-2.8	34.8
TT/AA	-0.3	-0.2	3.3	3.0	0.7	35.4
TC/GA	0.0	0.0	3.4	0.3	-0.1	39.3
CG/CG	0.4	0.9	3.2	-3.3	3.9	29.4
GC/GC	-1.3	0.4	3.7	-4.7	-12.2	40.8
CG/CG	0.8	0.1	3.2	3.1	-3.1	32.6

Table 4.12. Opening parameter increased in Hoogsteen base pair geometry

B-DNA		TTTCA		GGACA		AGGCA	
Base-pair	Opening	Base-pair	Opening	Base-pair	Opening	Base-pair	Opening
C-G	-3.67	T-A	3.12	G-C	2.34	A-T	-3.15
G-C	-4.02	T-A	1.53	G-C	-0.15	G-C	-2.08
C-G	-2.35	T-A	1.19	A-T	3.02	G-C	0.83
G-C	-1.3	C-G	1.65	C-G	2.63	C-G	2.34
A-T	1.84	A-T	5.68	A+T	65.35	A+T	66.27
A-T	5.56	T-A	9.88	T+A	-65.41	T+A	-68.07
T-A	7.93	G-C	1.68	G-C	4.66	G-C	1.54
T-A	0.83	A-T	2.66	T-A	0.75	C-G	1.76
C-G	-0.87	A-T	2.16	C-G	1.92	C-G	-1.22
G-C	-1.13	A-T	0.16	C-G	1.87	T-A	-0.44

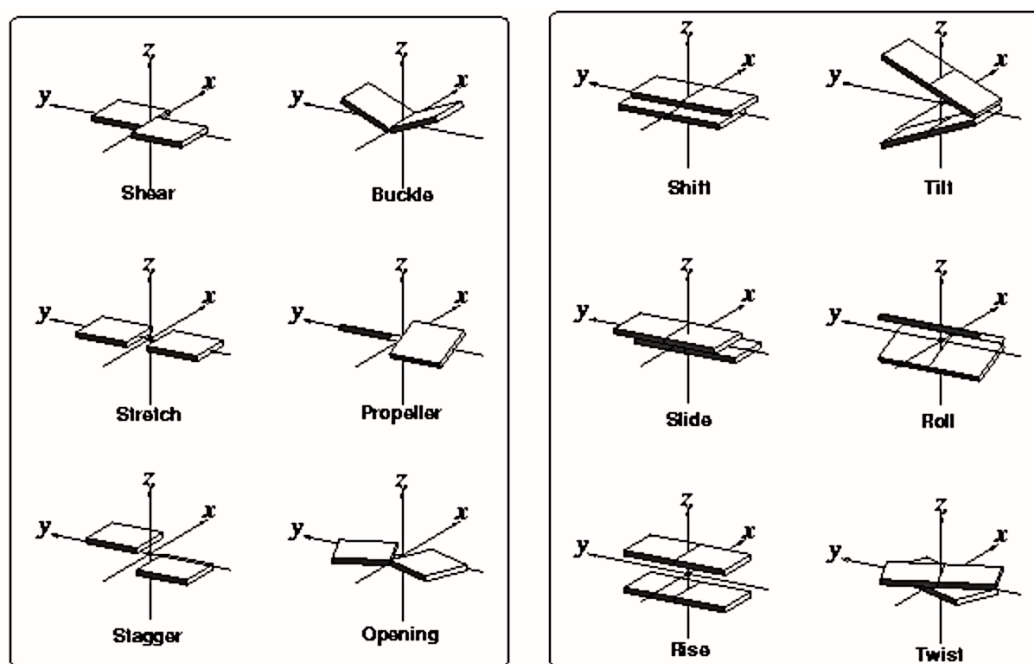


Figure 4.22. Parameters describing the orientation and position of one base-pair relative to another.

Protein-DNA interactions

The protein-DNA contact analysis of the three structures shows that each monomer in the dimer of the three structures binds in a perfectly symmetrical manner to each quarter-site response element. Therefore, I will only describe the interactions between residues in monomer A with DNA.

In the minor groove, Ser241 and Arg248 from loop L3 interact with the DNA backbone and the nucleotide bases in the three structures (Figure 4.23.). Ser241 interacts with the DNA backbone and C5' of Gua7 (Figure 4.24.) (Table 4.13. and 4.20.) Arg248 contacts C5' of Thy6 and the phosphate backbone of that same nucleotide, it also contacts the O4 and N3 of Gua21 and via a water molecule the O2 of Thy6 in the AGGCA and GGACA structures (Figure 4.24.) (Table 4.16. and 4.23.). This conformation of Arg248 was previously observed in our laboratory where the structure of p53DBD bound to half-site consensus sequence was solved at 1.2 Å resolution [38]. On the other hand, Arg248 interacts with O3' of Ade22 and with the phosphate backbone of Thy6 in the TTTCA structure (Figure 4.24) (Table 4.19. and 4.26.).

Protein-DNA interactions are observed in both the minor and the major groove. In the major groove, five residues contact the DNA backbone and bases in the three structures: Arg273, Cys275, Ala276, Cys277 and Arg280 from β -strand S10 and helix H2 (Figure 4.25.). Arg273, Ala276 and Cys275 contact the phosphate backbone of Thy6 and Gua7, while Arg280 interacts directly with the N7 and O6 of Gua7 (Figure 4.26.) (Table 4.13. and 4.20.). Moreover, Cys277 contacts N4 of Cyt8 in AGGCA complex, C7 of Thy8 in GGACA complex and N6 of Ade8 in TTTCA complex (Figure 4.26.) (Table 4.14. and 4.21.).

The major difference in the protein-DNA interactions observed in all three sequences is the position of Lys120 located in Loop1. In the three structures, Lys120 contacts different nucleotides (AGGCA and GGACA structure) or it is totally flipped away from the DNA (TTTCA structure) (Figure 4.27.). In the AGGCA structure, Lys120 interacts mainly with C8, C5 and N7

of Ade15, as well as with O6 and C5 of Gua16 and finally with the O6 and N7 of Gua17 (Table 4.15. and 4.22.). On the other hand, Lys120 contacts C8 and N7 of Gua15, and C8 and N7 of Gua16 in the GGACA structure (Table 4.17. and 4.24.). In contrast, no interactions are observed between Lys120 and the DNA in the TTTCA structure where the residue is completely flipped away from the half-site response element sequence (Figure 4.27.).

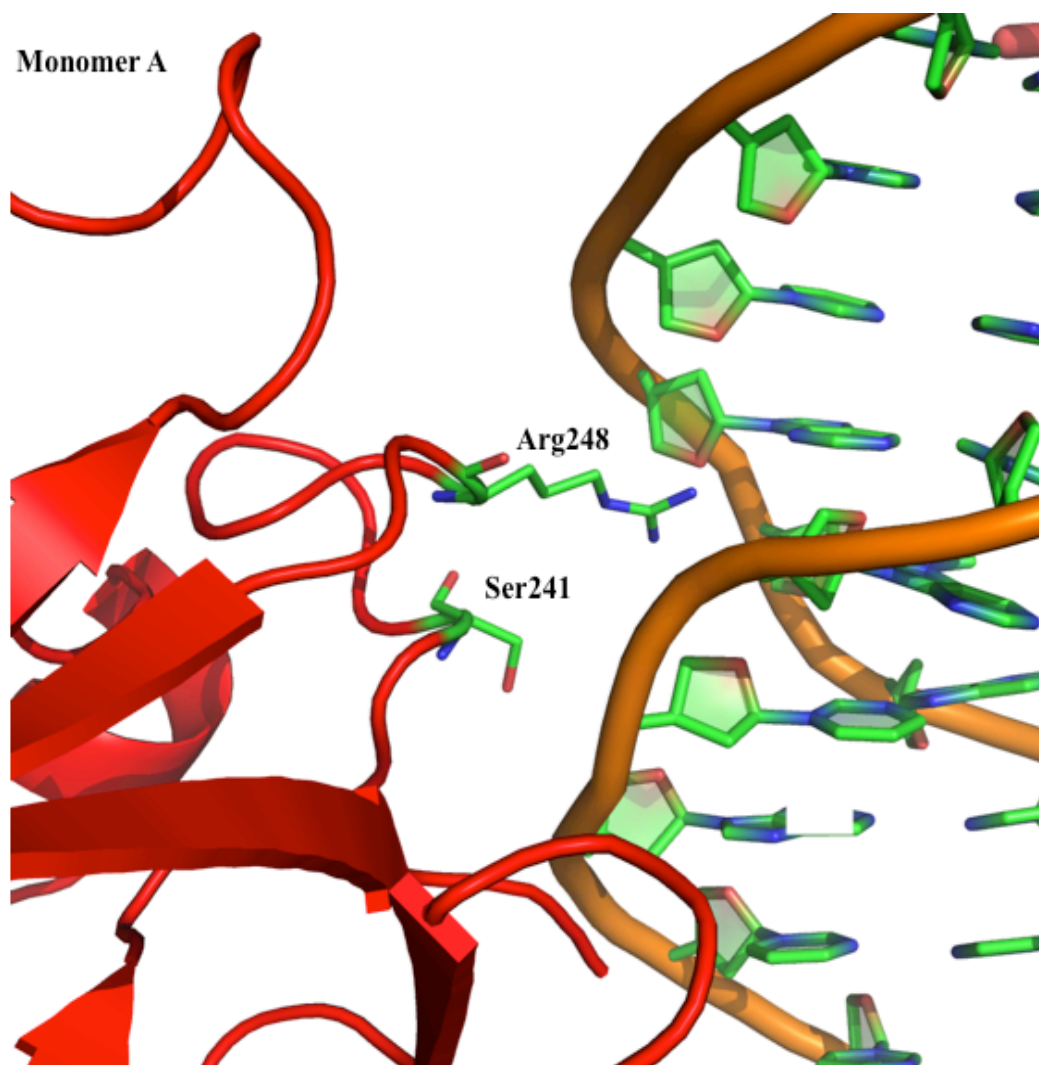


Figure 4.23. Interactions of the protein and DNA in the minor groove of the DNA half-site response element. The monomer A of AGGCA complex shows the two major contacts in the minor groove observed in the three structures.

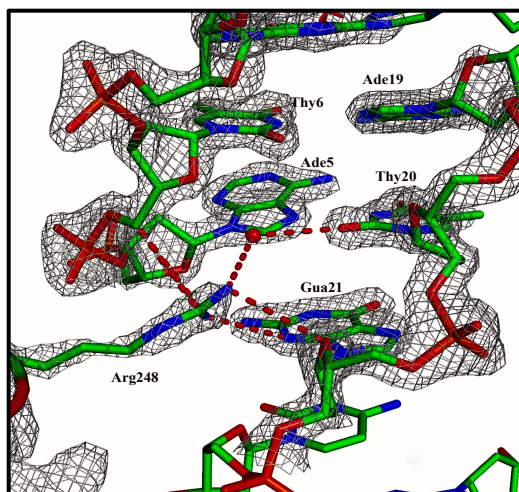
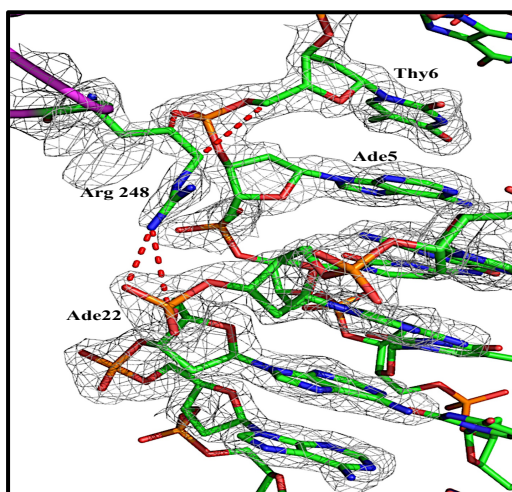
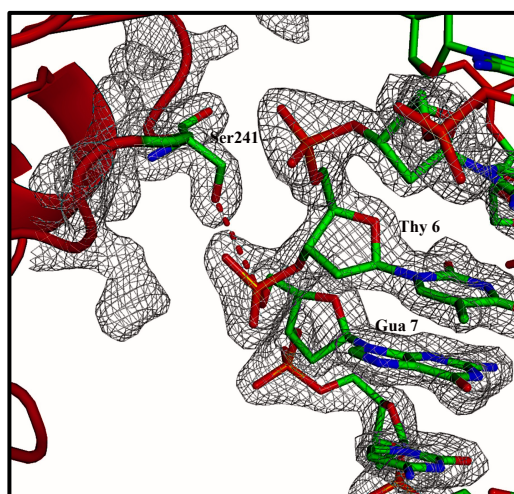
a. **AGGCA** and **GGACA** structuresb. **TTTCA** structurec. **AGGCA**, **GGACA** and **TTTCA** structures

Figure 4.24. Interactions of p53DBD residues with the DNA minor groove. a. Arg248 interaction of AGGCA and GGACA complex with the minor groove does not change. b. However, in TTTCA complex, Arg248 possess a different conformation, thus its interaction with the minor groove of the DNA is different from the other two complexes. c. Ser241 interaction with the minor groove is the same in the three complexes.

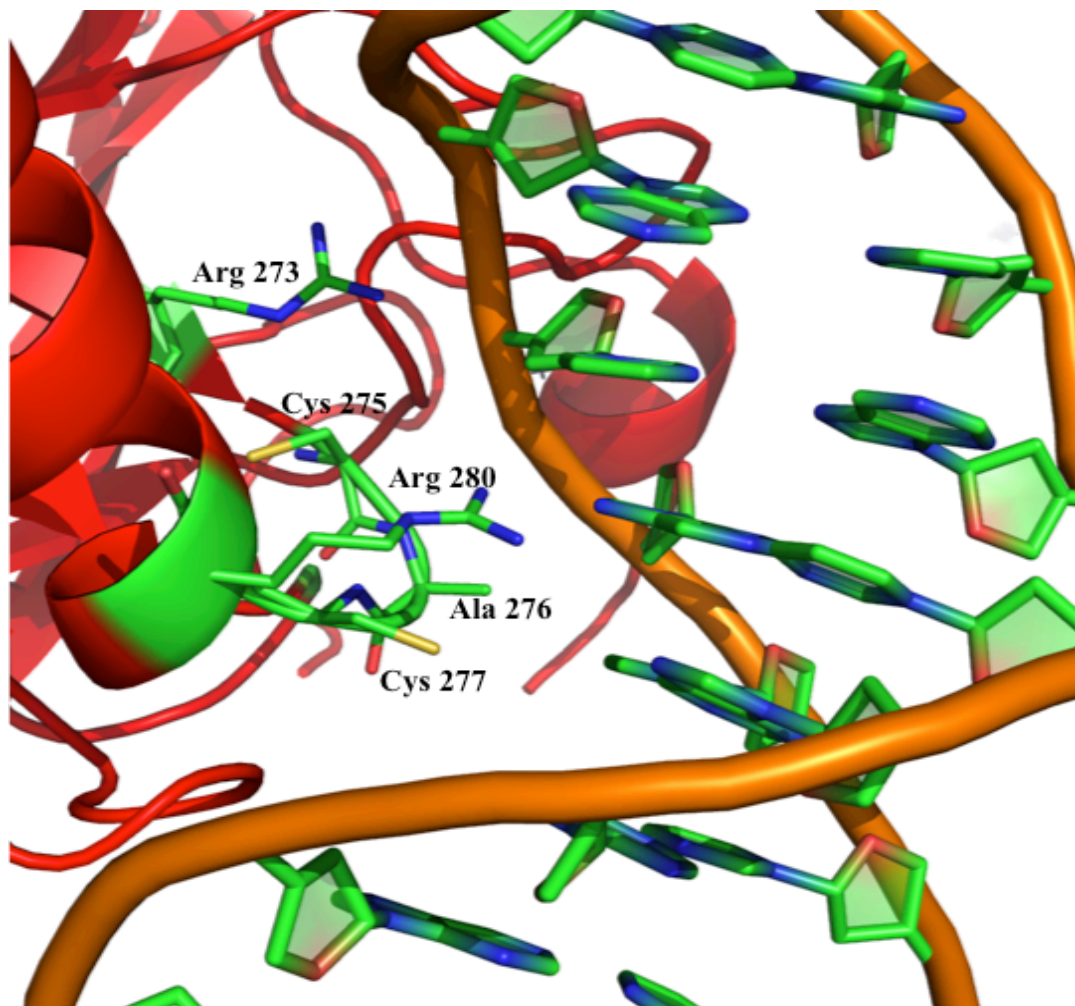


Figure 4.25. Overall p53DBD interactions with the mayor groove of the DNA half-site response element. The three complexes show the same five residues depicted in here making contact with the DNA.

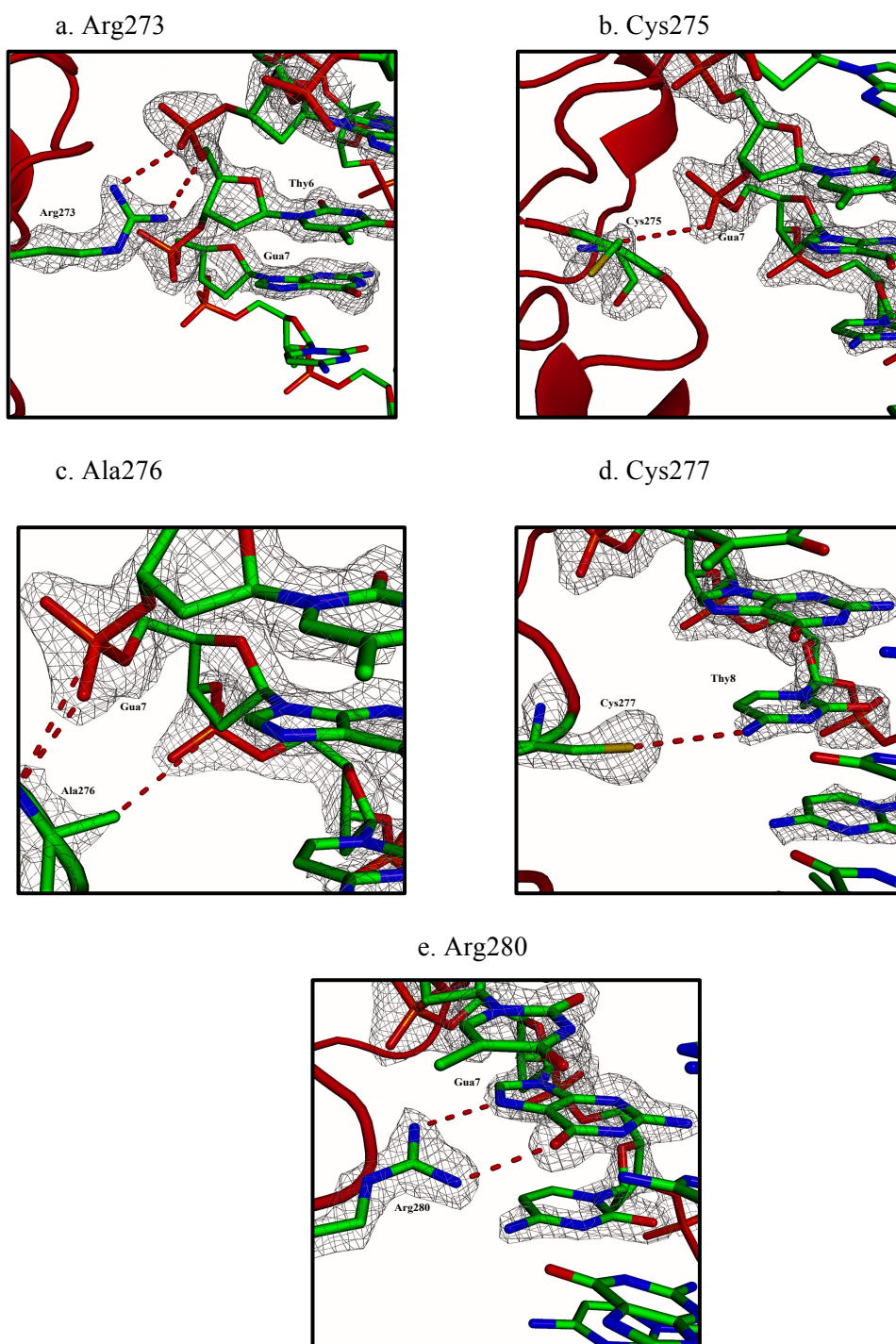


Figure 4.26. Interactions in the major groove. Interactions of Arg273 (a), Cys275 (b) and Ala276 (c), (d) Cys277 and (e) Arg280 with the major groove of the half-site response element in the AGGCA, GGACA and TTTCA complexes.

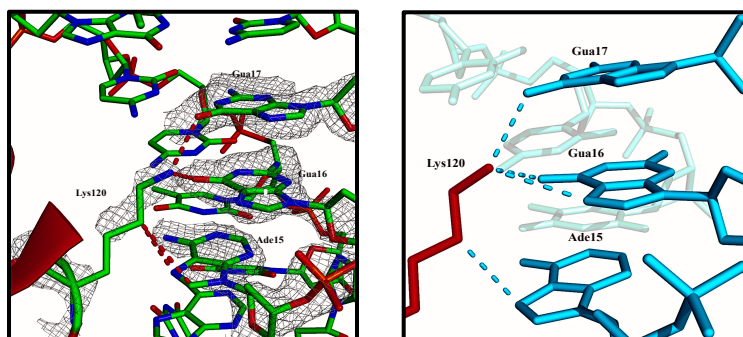
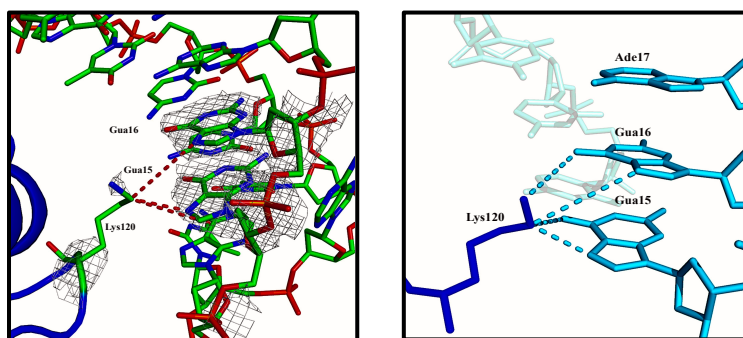
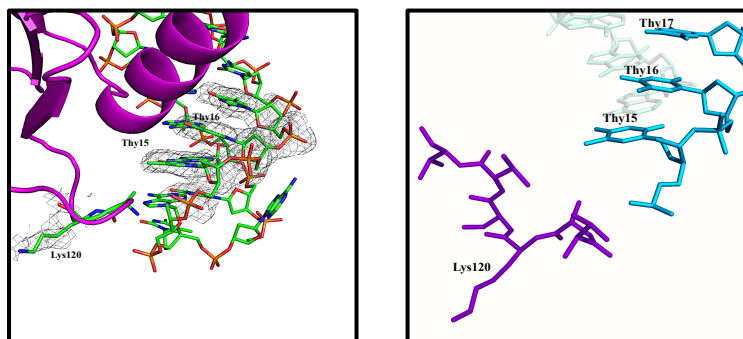
a. **AGGCA** complexb. **GGACA** complexc. **TTTCA** complex

Figure 4.27. Lysine conformation is different in all complexes. a. Lysine interacts with Ade15, Gua16 and Gua17 in the AGGCA complex. b. In GGACA complex, Lys120 contacts Gua15 and Gua16. c. Lys120 is completely away from the half-site response element sequence in TTTCA complex. Both monomers display the same conformation of Lys120 interacting with DNA, monomer A is shown as an example.

Table 4.13. Interactions of monomer A with DNA observed in all the three complexes

Residue	Monomer A number	Atom	Nucleotide	Number	Atom	Distance(Å)
SER	241	OG	Gua	7	OP1	2.6
ALA	276	N	Gua	7	OP2	2.8
ARG	280	NH1	Gua	7	O6	2.9
ARG	273	NH1	Thy	6	OP2	2.9
ARG	273	NH2	Thy	6	OP2	2.9
ARG	280	NH2	Gua	7	N7	3.0
ARG	283	NH1	Ade	15	OP2	3.2
ARG	273	NH2	Thy	6	O5'	3.2
ARG	273	CZ	Thy	6	OP2	3.2
SER	241	CB	Gua	7	OP1	3.3
CYS	277	SG	Gua	8	N4	3.5
SER	241	OG	Gua	7	C5'	3.5
CYS	275	CB	Gua	7	OP2	3.5
ARG	273	NH2	Thy	6	P	3.6
SER	241	OG	Gua	7	P	3.8
ARG	273	NH2	Thy	6	C2'	3.8
ARG	273	NH2	Thy	6	C3'	3.8
ARG	273	NH1	Thy	6	P	3.9
ALA	276	CB	Gua	7	C2'	3.9
ALA	276	N	Gua	7	P	3.9

Table 4.14. Cysteine 277 contacts with different nucleotides in the same position

Residue	Monomer A number	Atom	Nucleotide	Number	Atom	Distance(Å)	Complex
CYS	277	SG	Cyt	8	N4	3.5	AGGCA
CYS	277	SG	Thy	8	C7	3.7	GGACA
CYS	277	SG	Ade	8	N6	3.5	TTTCA

Table 4.15. Contacts of Arg 283 and Lys 120 with different nucleotides in AGGCA and GGACA complexes

Residue	Monomer A number	Atom	Nucleotide	Number	Atom	Distance(Å)	Complex
ARG	283	NH1	Ade	15	OP2	3.2	AGGCA
ARG	283	NH1	Gua	15	OP2	3.2	GGACA
LYS	120	CD	Ade	15	C8	3.6	AGGCA
LYS	120	CD	Gua	15	N7	3.8	GGACA
LYS	120	CE	Gua	15	C8	3.8	GGACA
LYS	120	CD	Ade	15	C5	3.8	AGGCA

Table 4.16. Contacts sharing by only two complexes

Residue	Monomer A number	Atom	Nucleotide	Number	Atom	Distance(Å)	Complex
ARG	248	NH1	Thy	6	C5'	3.3	AGGCA, GGACA
ARG	248	NH2	Gua	21	O4'	3.3	AGGCA, GGACA
ALA	276	CB	Gua	7	OP2	3.4	AGGCA, TTTCA
ARG	248	NH1	Gua	21	N3	3.5	AGGCA, GGACA
ARG	248	CZ	Thy	6	C5'	3.6	AGGCA, GGACA
CYS	275	CA	Gua	7	OP2	3.6	AGGCA, TTTCA
ARG	248	CD	Thy	22	C5'	3.7	AGGCA, GGACA
CYS	275	CA	Gua	7	P	4.3	GGACA, TTTCA

Table 4.17. Contacts observed only in GGACA complex

Residue	Monomer A number	Atom	Nucleotide	Number	Atom	Distance(Å)
LYS	120	CE	Gua	15	N7	3.0
ARG	248	CB	Thy	6	OP1	3.6
ARG	248	CG	Thy	6	OP1	3.7
ARG	280	NH1	Gua	7	N7	3.7
ALA	276	CB	Thy	8	C7	3.7
ARG	248	NH2	Gua	21	C1'	3.8
ARG	248	CD	Thy	22	C4'	4.1
CYS	275	CB	Gua	7	P	4.3

Table 4.18. Contacts observed only in AGGCA complex

Residue	Monomer A number	Atom	Nucleotide	Number	Atom	Distance(Å)
ARG	283	NH2	Ade	15	OP2	2.9
LYS	120	NZ	Gua	16	O6	3.0
LYS	120	NZ	Gua	17	O6	3.0
LYS	120	CD	Ade	15	N7	3.2
LYS	120	NZ	Gua	16	N7	3.2
ARG	280	NH1	Cyt	8	N4	3.3
LYS	120	CE	Gua	16	O6	3.3
ARG	283	CZ	Ade	15	OP2	3.4
LYS	120	CB	Ade	15	N7	3.4
ARG	280	NH1	Gua	17	O6	3.4
ARG	280	NH2	Thy	6	C6	3.6
LYS	120	NZ	Gua	16	C6	3.6
ALA	276	CA	Gua	7	OP2	3.7
LYS	120	CB	Ade	15	C8	3.7
ARG	283	NH2	Ade	15	P	3.7
ARG	283	NH1	Ade	15	P	3.7
LYS	120	CE	Cyt	8	N4	3.8
LYS	120	CG	Ade	15	N7	3.8
ARG	280	NH2	Thy	6	C2'	3.8
ALA	276	CB	Gua	7	P	4.3

Table 4.19. Contacts observed only in AGGCA complex

Residue	Monomer A number	Atom	Nucleotide	Number	Atom	Distance(Å)
ARG	283	NH2	Thy	15	OP2	2.7
ARG	248	NH2	Ade	23	OP1	2.8
ARG	248	NH2	Ade	22	O3'	3.0
ARG	280	NH1	Thy	17	O4	3.0
ALA	119	N	Thy	15	OP2	3.3
ALA	119	CB	Thy	15	OP2	3.4
ALA	119	CB	Ade	14	C2'	3.4
ARG	248	NH2	Ade	23	P	3.5
ARG	248	CG	Thy	6	OP1	3.5
ARG	248	CB	Thy	6	OP1	3.5
ARG	280	NH1	Ade	8	N6	3.6
ARG	283	NH2	Thy	15	P	3.6
ARG	248	CD	Thy	6	OP1	3.6
ARG	280	NH2	Thy	6	C2'	3.8
ALA	119	N	Ade	14	C3'	3.8
ARG	280	CB	Thy	16	C7	3.9
ALA	119	CB	Ade	14	C3'	3.9
ARG	248	NH2	Ade	23	C5'	3.9
ARG	248	CD	Thy	6	C5'	4.0
ARG	280	CD	Thy	16	C7	4.0
ARG	280	CG	Thy	16	C7	4.2
SER	241	CB	Gua	7	P	4.4

Table 4.20. Interactions of monomer B with DNA observed in all the three complexes

Residue	Monomer B number	Atom	Nucleotide	Position	Atom	Distance(Å)
SER	241	OG	Gua	21	OP1	2.7
ALA	276	N	Gua	21	OP2	2.8
ARG	280	NH1	Gua	21	O6	2.9
ARG	273	NH1	Thy	20	OP2	2.9
ARG	273	NH2	Thy	20	OP2	3.0
ARG	280	NH2	Gua	21	N7	3.0
ARG	273	NH2	Thy	20	O5'	3.3
ARG	273	CZ	Thy	20	OP2	3.3
SER	241	CB	Gua	21	OP1	3.4
ARG	280	NH2	Thy	20	C6	3.6
CYS	275	CB	Gua	21	OP2	3.6
ARG	248	CG	Thy	20	OP1	3.6
SER	241	OG	Gua	21	C5'	3.6
ARG	273	NH2	Thy	20	P	3.7
SER	241	OG	Gua	21	P	3.8
ALA	276	N	Gua	21	P	3.9
ARG	273	NH1	Thy	20	P	3.9
ALA	276	CB	Gua	21	C2'	4.0
CYS	275	CA	Gua	21	P	4.4

Table 4.21. Cysteine 277 contacts with different nucleotides in the same position

Residue	Monomer B number	Atom	Nucleotide	Number	Atom	Distance(Å)	Complex
CYS	277	SG	Cyt	22	N4	3.4	AGGCA
CYS	277	SG	Ade	22	N6	3.5	TTTCA
CYS	277	SG	Thy	22	C7	3.8	GGACA

Table 4.22. Contacts of Lys 120 with different nucleotides in AGGCA and GGACA complexes

Residue	Monomer A number	Atom	Nucleotide	Number	Atom	Distance(Å)	Complex
LYS	120	NZ	Gua	3	O6	2.6	AGGCA
LYS	120	NZ	Gua	2	O6	3.2	GGACA
LYS	120	CE	Gua	2	O6	3.3	GGACA
LYS	120	CE	Gua	3	O6	3.4	AGGCA
LYS	120	CE	Gua	1	N7	3.6	GGACA
LYS	120	CE	Gua	2	N7	3.6	AGGCA

Table 4.23. Contacts sharing by only two complexes

Residue	Monomer B number	Atom	Nucleotide	Number	Atom	Distance(Å)	Complex
ALA	276	CB	Gua	21	OP2	3.4	AGGCA, TTTCA
ALA	276	CB	Gua	21	OP2	3.4	AGGCA, TTTCA
ALA	276	CB	Gua	21	OP2	3.4	AGGCA, TTTCA
ARG	248	CB	Thy	20	OP1	3.6	AGGCA, GGGCA
ARG	248	NH1	Thy	20	C5'	3.7	AGGCA, GGGCA
ARG	248	CZ	Thy	20	C5'	3.7	AGGCA, GGGCA
ARG	280	NH1	Thy	20	C7	3.8	AGGCA, GGGCA
ARG	280	CZ	Thy	20	C7	3.9	AGGCA, GGGCA

Table 4.24. Contacts observed only in GGACA complex

Residue	Monomer B number	Atom	Nucleotide	Position	Atom	Distance(Å)
LYS	120	NZ	Thy	22	O4	3.3
MET	243	CE	Gua	7	O3'	3.3
ARG	248	NH2	Gua	7	O4'	3.4
MET	243	CE	Thy	8	OP1	3.5
LYS	120	CE	Gua	1	C8	3.5
LYS	120	CE	Gua	2	N7	3.5
ALA	276	CB	Thy	22	C7	3.7
ARG	280	NE	Thy	20	C7	3.8
MET	243	CE	Gua	7	C4'	4.0
ARG	248	CD	Thy	8	C5'	4.0
MET	243	CE	Thy	8	P	4.0
MET	243	CE	Gua	7	C3'	4.0

Table 4.25. Contacts observed only in AGGCA complex

Residue	Monomer B number	Atom	Nucleotide	Position	Atom	Distance(Å)
LYS	120	N	Ade	1	OP2	2.8
LYS	120	NZ	Gua	2	N7	2.8
SER	121	N	Cyt	0	O5'	2.9
ARG	248	NH1	Gua	7	N3	3.3
LYS	120	NZ	Gua	3	C6	3.4
SER	121	C	Cyt	0	O5'	3.4
SER	121	CA	Cyt	0	O5'	3.5
ALA	119	CA	Ade	1	OP2	3.5
LYS	120	CB	Ade	1	OP2	3.6
LYS	120	NZ	Gua	3	N7	3.6
LYS	120	NZ	Gua	2	C5	3.6
LYS	120	CE	Gua	2	O6	3.6
ALA	276	CA	Gua	21	OP2	3.7
LYS	120	CD	Gua	2	N7	3.7
ALA	119	CB	Cyt	0	O5'	3.7
SER	121	CB	Cyt	0	O5'	3.7
LYS	120	NZ	Gua	2	C8	3.8
ARG	273	NH2	Thy	20	C2'	3.9
ARG	273	NH2	Thy	20	C3'	3.9
LYS	120	CD	Ade	1	N7	3.9
ALA	276	CB	Gua	21	C8	4.0
ALA	276	CB	Gua	21	P	4.3
ALA	119	CB	Ade	1	P	4.4
ALA	119	CA	Ade	1	P	4.4

Table 4.26. Contacts observed only in TTTCA complex

Residue	Monomer B number	Atom	Nucleotide	Position	Atom	Distance(Å)
ARG	283	NH1	Thy	1	OP2	3.1
VAL	122	CB	Cyt	0	O5'	3.3
ARG	280	NH1	Ade	22	N6	3.4
ARG	280	NH2	Thy	20	C2'	3.8
VAL	122	CG2	Cyt	0	C5'	4.0
VAL	122	CB	Cyt	0	C5'	4.0
ARG	280	CB	Thy	1	C7	4.1
ARG	248	CG	Thy	21	OP1	3.3
ARG	248	CD	Thy	21	OP1	3.4

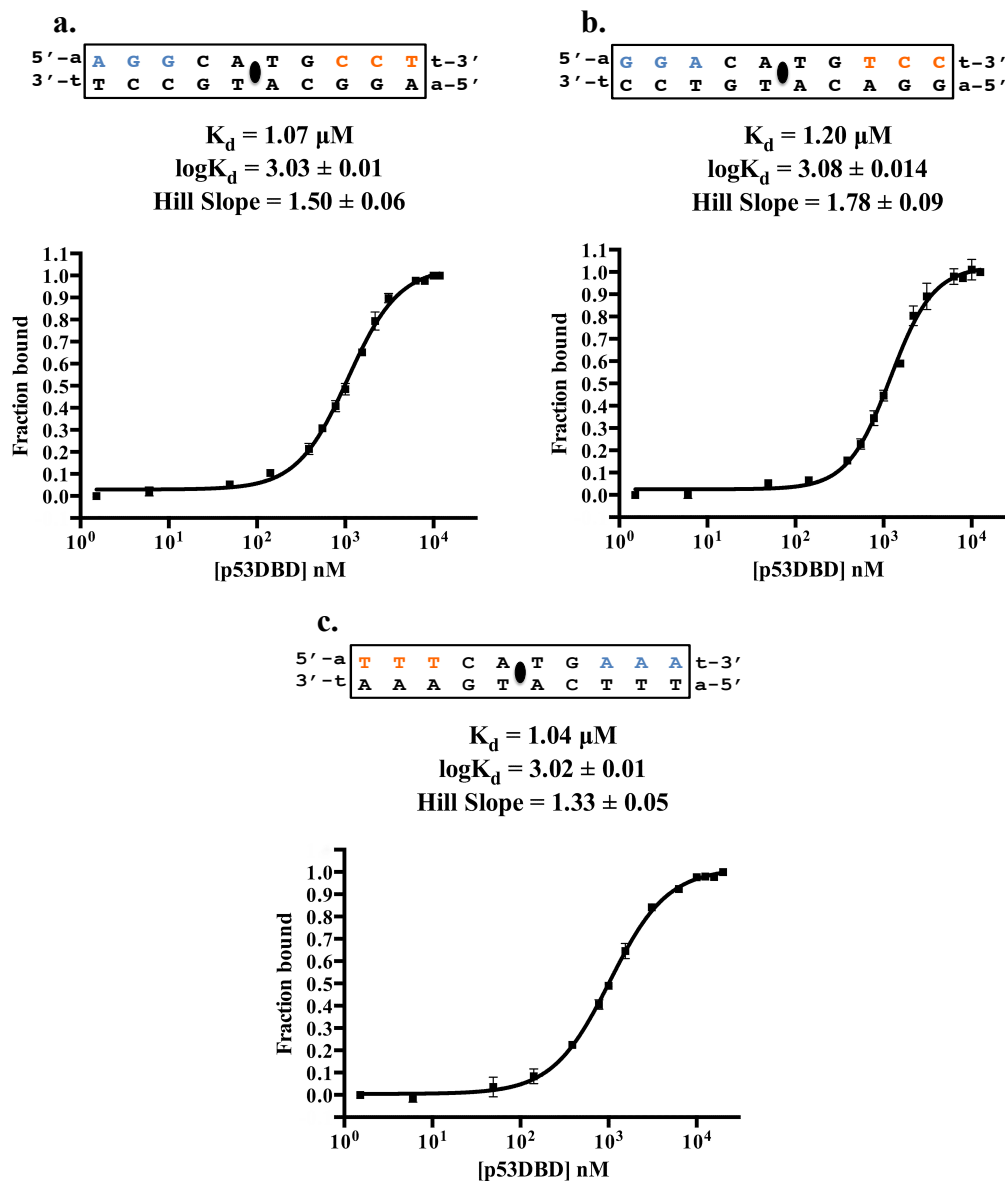


Figure 4.28. Binding affinity graphs of p53DBD bound to a half-site response element with different nucleotides in the flanking triplet sequence.

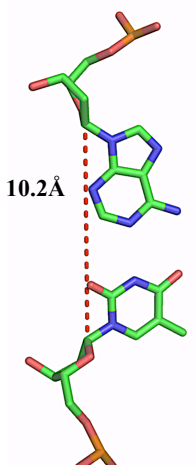
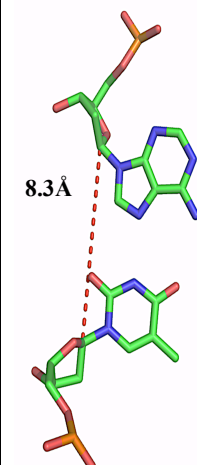
C1'-C1' Distances for the two central A-T base pairs									
Watson and Crick		Hoogsteen		Petty et al.		Emamzadah et al.		Emamzadah et al.	
5'-cGGGCATGCCCg-3'		5'-GGGCATGCCC-3'		PID = 3Q05		PID= 4MZR Tetramer		PID= 3TS8 Tetramer	
A-T	T-A	A-T	T-A	A-T	T-A	T-A	T-A	A-T	T-A
10.2	10.2	8.3	8.2	9.7	9.5	10.1	9.7	9.6	9.7
				9.2	9.6	9.8	9.8	9.2	9.6
				Tetramer Contiguous full-site Consensus sequence		Tetramer Contiguous full-site Consensus sequence		Tetramer Contiguous full-site p21 response element	

Figure 4.29. C1'-C1' distances for the two central A-T base pairs of p53 structures containing the mutated DNA binding domain and the modified oligomerization domain bound to a contiguous full-site.

Discussion

The p53 response elements were defined based on a consensus rule; although this approach has served to understand the basis of p53 binding specificity, now it is evident that response elements' sequences can give us further information about p53 activity on activating or repressing a target gene. Recently, Wang et al., showed that p53 response elements could be divided according to their ability to transactivate the luciferase gene in an in vitro assay into active or repressive. Nevertheless, the final outcome (activating or repressing) is a sum of several factors, such as p53 mutations, transcriptional cofactors, variable spacer length, etc. The approach taken by Wang et al., although simplistic, helps to shed light on the complicated process of how p53 activates or represses transcription. The primary determinant of this activity might lie in the p53 response element. Furthermore, the molecular mechanism used by p53 to distinguish between the activating and repressing elements is still unknown. Therefore, in this chapter, I have taken a biochemical and structural approach to explain such mechanism.

The p53DBD dimers bound to activating response elements show similar conformation compared with p53DBD dimers bound to repressing element.

The conformation of the p53DBD dimer in the AGGCA and GGACA structures bound to activating response elements is similar, as measured by their structural alignment (r.m.s. dev. = 0.2), their monomer-monomer centers of mass distances (43 Å for both) and their monomer-monomer angles with respect to the main DNA axis (101.7° and 102°) (Figure 4.16.). On the other hand, the same values differ for the p53DBD bound to the TTTCA repressing sequence; the structural alignments have a larger r.m.s. dev. when the TTTCA structure is compared to the AGGCA (1.7) and the GGACA (1.8) structures, the monomer-monomer centers of mass distance is 46 Å and the dimer angle with respect to the DNA is 110.3°. These data demonstrate that the

dimer conformation has a more open conformation when bound to repressing sequences than when bound to activating sequences.

The Adenines in the Center of the Activating Half-Site Response Elements Flip to a Hoogsteen Conformation

Previous results on the crystal structures of the p53 DBD bound to full and half-site response elements show that the adenine nucleotides in the central base pair A/T doublets can flip 180° acquiring a Hoogsteen conformation [69] [38]. I also found the central adenine in a Hoogsteen conformation in the two AGGCA and GGACA crystal structures that are bound to the activating response elements. In contrast, in the TTTC structure bound to a repressive response element, all the base pairs in the half-site response element are in the standard Watson-Crick conformation.

The conformational switching of the central adenines in the middle of the half-site response element from a Watson-Crick to a Hoogsteen conformation was the most striking difference between the activating and the repressing sequences. Nonetheless, it was not the only difference. To further understand the similarities between the AGGCA and GGACA structures on one hand, and the dissimilarities between the TTTC and the AGGCA or the GGACA structures on the other hand, I analyzed the dimer interface and the protein-DNA interactions in the three structures.

Regarding the dimeric interface, the interacting residues are located in helix H1 and loop L3, as previously seen in all the previously solved p53 structures [22] [105] [36]. However, the number of interactions between monomers in AGGCA structure was greater than that for GGACA; for instance, in AGGCA there are 29 interactions between the monomers, whereas in GGACA structure, 19 interactions are observed. Moreover, TTTC structure possessed only 8 protein-protein interactions within the dimer. Also, looking at the vertical axis of the DNA, the angle formed by the two monomers in AGGCA and GGACA structures are pretty similar, 101.7°

and 102° compared with TTTCA with an angle of 110.3° explaining the fact that reduced number of interactions are observed in TTTCA structure due to a more opened dimer (Figure 4.16.). Finally, the accessible surface area in TTTCA structure is smaller, $16,099.8 \text{ \AA}^2$, than the one calculated for AGGCA and GGACA structures, $22,855.9 \text{ \AA}^2$ and 23143.1 \AA^2 . Taking together all these data, I can conclude that the dimer interface is greatly affected by monomer conformations where in the repressing TTTCA sequence, the monomers seem tighter or in a closed-pack form leading to a reduced accessible solvent area, less number of interactions between monomers and a opened dimer compared with the activating sequence.

By analyzing the DNA structure and the DNA-protein interactions, I was able to explain the reason why an opened-dimer exists in the TTTCA structure while a closed dimer is present in the AGGCA and GGACA structures. The best analogy to explain the complex conformational changes is to think that the dimerization interface acts as a movable hinge that adjust according to the sequence of the half-site response element. In order to transition from the opened conformation (TTTCA structure) to the closed conformation (GGACA and AGGCA structures), there are three main points of interactions. First, as our binding studies suggest, the recognition by Arg280 of guanine in the fourth conserved base pair is the first and most important determinant of binding. Second, Arg248 reads the narrowness of the minor groove and, if it corresponds to a narrow A-T step, Arg248 forms hydrogen bonds with both phosphate backbones, enters deep into the minor groove, by further narrowing the minor groove forces the switching of the central adenine to Hoogsteen conformation and makes a water-mediated hydrogen-bond with the O2 of the central thymine. And, third, the Lys120 reads the external nucleotides; in case of finding guanines, it forms hydrogen bonds and brings the entire loop L1 closer to the DNA. The regulatory movement of Lys120 is coupled to the one from Cys277, which is also reading the composition of the third base pair. In brief, Lys120 reads the presence of purines in one strand,

while Cys277 reads the presence of pyrimidines in the other strand; if both residues contact the bases, the dimer forms a closed conformation.

Furthermore, Arg248 adopted a similar conformation in AGGCA and GGACA structures as the one indicated in the GGGCA structure solved by Nguyen and Kitayner et al. where the residue made a direct contact with N7 of Gua7 and the O of Thy6 via a water molecule. In contrast, Arg248 adopted a different geometry in the TTTCA structure where it interacted only with the phosphate backbone of Thymine6.

On the other hand, Lys120 directly contacted the guanine at position 2 and 3, there are interactions with Adenine1, but they are mostly Van der Waals (Figure 4.27.a.). Like in the AGGCA structure, Lys120 contacted also the guanines but at positions 1 and 2 in the GGACA structure (Figure 4.27.b.). Finally, in the TTTCA structure, Lys120 flipped away from the DNA and no interaction was made (Figure 4.27.c.). Also, MD studies on p73 tetramers bound to GGGCA and GAACA response elements suggested that when Lys120 interacts with the Guanines in the triplet flanking sequence, there is an effect in which the monomers of each dimer come close, whereas in the GAACA, this effect is much more diffuse [106]. Therefore, the effect of Lys120 interaction with Guanines in the triplet flanking sequence could enhance the formation of Hoogsteen base pairing by inducing a closer conformation that facilitates Arg248 interaction within the minor groove resulting in a change of conformation from Watson and Crick to Hoogsteen in the central base pair doublets A/T. It is important to point out that binding affinity was not affected by changes in the flanking sequence of the solved structures (Figure 4.28.) indicating the null correlation between the formation of Hoogsteen geometry and the binding affinity of the protein to DNA in this study.

Thus, our results show that the presence or absence of the non-canonical Hoogsteen base pairing in a dimer of p53 DBD bound to half-site response element is dependent on not only the direct interaction of Arg248 in the minor groove (Nguyen, 2012), but also on the conformation of

Lysine120 favoring the interaction with guanines. These findings suggest a new DNA recognition mode for p53. Also, our results rule out the crucial role of inter-dimer interactions in the formation of Hoogsteen base pairing [69] since I analyzed one high resolution dimer structure (AGGCA) containing the non-canonical geometry of Hoogsteen. To further confirm our findings where I hypothesize the important role of Lys120 conformation and Arg248 contacting the base pairs in the minor groove, mutants of p53 DBD at both position could be performed and the structures solved to verify our hypothesis.

Although our results are supported by high-resolution structures of the p53DBD bound to AGGCA and TTTCA, I do not dismiss the possibility that the N or C-terminus might play a role in this new DNA recognition mode by p53. Structures containing the DNA binding and the oligomerization domain of p53 bound to contiguous full site consensus sequence show the canonical Watson and Crick base pairing geometry in the central nucleotides A-T doublets instead of Hoogsteen conformation [107] [108] [109]. However, the distance between C1 of the ribose in adenine and the C1 of the ribose in thymine are smaller than expected for the canonical Watson and Crick base pairing conformation (Figure 4.29.). Besides, these structures contained several mutations in the DNA binding domain and in the linker between the DNA binding domain and the oligomerization domain that might affect the conformation of the DNA upon protein binding.

Finally, our study sheds light on the way p53 might recognize its activating response elements via the formation of Hoogsteen base pair in the central A-T doublets produced by Arg248 and Lys120 conformations leading to an overall closer conformation of the p53DBD monomers. Further structural studies with p53 DBD and other p53 domains (not mutated) bound to contiguous and non-contiguous full-site response elements might be needed to confirm the novel DNA recognition mechanism of p53.

DNA Binding Affinity by p53DBD and AcLys120-p53DBD Remains Unaltered upon Changing the Nucleotides in the Flanking Triplet Sequence

The results from the binding experiments showed that the DNA binding affinity of the p53DBD protein is unaffected by changes in the first three nucleotides in each quarter-site response element (Figure 4.7.). On the other hand, the Hill coefficient decreased when AAACA, CCCCA and TTTCA substituted GGGCA from 1.76 to 1.0. Therefore, a lost in a cooperative behavior is observed when guanines were replaced by adenines, cytosines or thymines, even when the dissociation constants remain unchanged.

The binding affinity of AcLys120-p53DBD to half-site response element was similar to the binding affinity of AcLys120-p53DBD to half-site sequences with the triplet flanking mutated to pyrimidines (Table 4.3.). As observed from our data, the binding affinity of AcLys120-p53DBD is comparable to all the sequences tested even though a lost in the interaction of Lys120, which is acetylated and lacks the positive charged, with the guanines in the triplet flanking sequence made us expected a significant dropped in affinity when the protein was bound to GGGCA half-site consensus sequence. The calculated Hill coefficient for the binding of AcLys120-p53DBD to GGGCA, AAACA, and CCCCA was closed to the unity resulting in a non-cooperative binding process. However, for the binding of AcLys120-p53DBD to TTTCA, the process was cooperative since the calculated Hill coefficient was 2.15; moreover the affinity was the highest among the sequences tested.

Finally, the binding affinity of AcLys120-p53DBD to the half site sequences with a quarter site containing GGGCA or AAACA or CCCCA or TTTCA is similar compared with the binding affinity of p53 DBD to the same sequences (Table 4.3.). Therefore, the acetylation of Lys120 in the p53 DBD does not disturb the binding affinity of p53DBD to half site consensus sequence. The ability for p53DBD to bind half-site activating or repressing sequences remained unchangeable upon the acetylation of Lys120. The only noticeable difference relied on the fact

that p53 DBD binds to GGGCA in a cooperative way, whereas AcLys120-p53DBD does not. On the other hand, AcLys120-p53DBD showed a cooperative behavior when bound to TTTCA while in p53 DBD the cooperative behavior was lost when bound to TTTCA.

Chapter 5

Conclusions and Final Model

Conclusions

A fundamental biological problem that remains unsolved is to know how the cell decides when to activate or inactivate genes. Due to the importance of p53 mutations in the emergence of cancer cells, I decided to use the p53 protein family as a model to understand response element recognition by transcription factors. In spite of the great amount of structural and functional information about p53 and its family members, a molecular mechanism on how p53 differentiate between a myriad of potential activating and repressing response elements in the genome remains uncovered. My biochemical and structural work suggest a mechanism on how members of the p53 protein family distinguish between activating and repressing response elements.

The DNA binding experiments suggest a hierarchical response element recognition where some nucleotides are more important than others to determine the binding of p73 and p53 to its response element. The p53 family of transcription factors binds to 20 bp response elements that have two adjacent 5'-PuPuPuCATGPyPyPy-3' half-sites. My binding data suggests that: 1) the conserved cytosine in the fourth position of each palindromic quarter-site is the most important residue for binding; 2) there is a preference for AT over TA, GC or CG as the bases in the two central nucleotides of the half-site response element; and 3) the flanking three nucleotides in each side of the half-site response element are not very important for determining DNA binding affinity.

The structural experiments with sequences that have been reported as activating and inactivating suggest a sequential step-wise response element recognition by p53. I defined two basic conformations, an opened one bound to repressing response elements and a closed one bound to activating response elements. The proposed steps on the p53 mechanism of response element recognition are: 1) The p53 DBD binds as a dimer to any DNA. 2) The p53 DBD slides to bind to a 10 bp half-site defined by a central 5'-CNNG-3' sequence. 3) Depending on the sequence of the flanking nucleotides Lys120 will contact the bases or not. If the response element

is activating, Lys120 contacts the second base; instead if the response element is inactivating, Lys120 does not recognize the bases and the dimer remains in an opened conformation. 4) In active response elements that contain a central CATG, CAAG or CTTG sequence, Arg248 approaches the narrow minor groove, narrows it even further by contacting both phosphate backbones, such DNA conformational change promotes the breakage of the Watson-Crick base pair, the adenine flips to a Hoogsteen conformation and Arg248 makes a water-mediated contact with the central thymine that stabilizes the dimer in a closed conformation (Figure 5.1).

Besides these findings, I started to study two other factors that influence response element recognition. First, the presence of the oligomerization domain, instead of only the DNA-binding domain, makes p73 to increase specificity and cooperativity. On the other hand, I studied the effect of Lys120 acetylation that is known to regulate p53 activation of apoptotic genes, but I did not discover a direct effect on DNA binding.

Finally, the p53 model of response element recognition explains in molecular terms how the members of the p53 family proteins can discriminate between activating and repressing response elements. Further studies should focus on how the conformational changes that I found transmit the activation or repression signal to the full-length p53 family proteins and to the general transcription factors responsible of recruiting the transcriptional machinery.

1. p53 binds as a dimer to any DNA

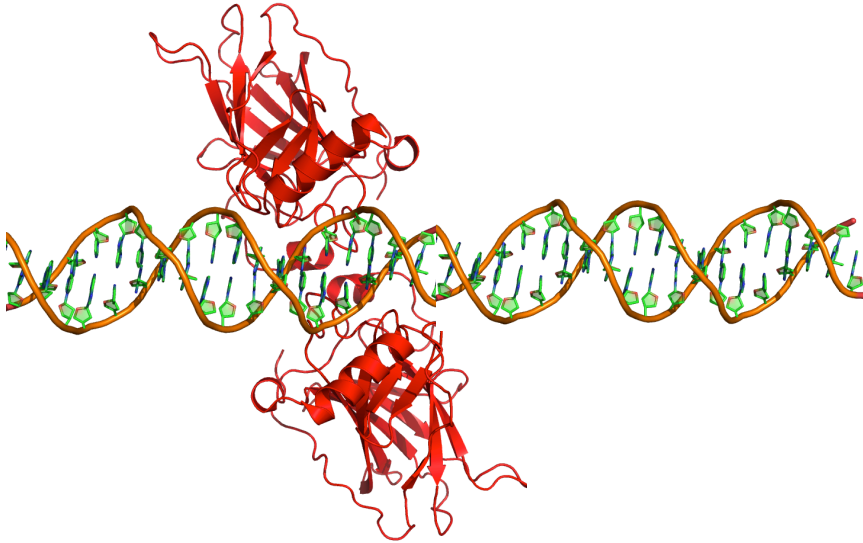


Figure 5.1.a. The molecular mechanism used for p53 to recognize activating response elements.

2. p53 slides to bind a 10bp half-site defined by 5'-CNNG-3'

5'-XXX**CNNG**YYY-3'

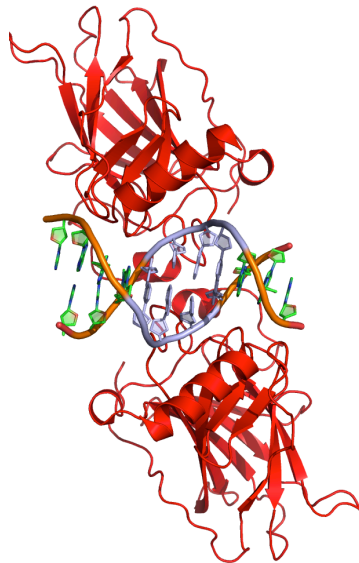
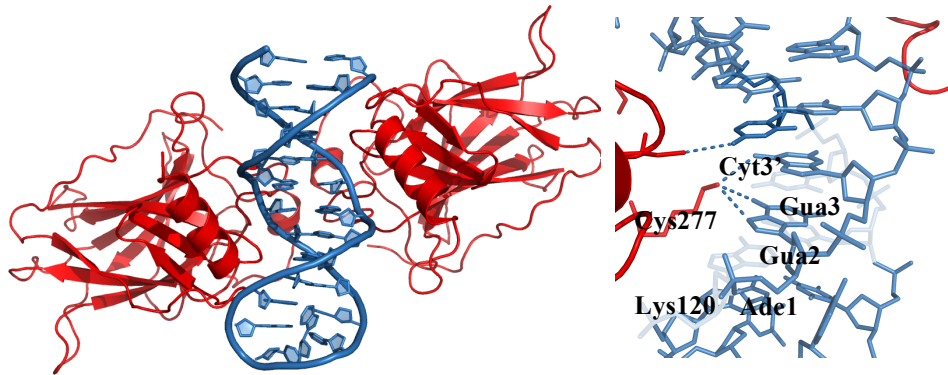


Figure 5.1.b. The molecular mechanism used for p53 to recognize activating response elements.

3. Lys120 contacts DNA in activating response elements and Lys120 does not contact DNA in repressing response elements

AGGCA



TTTCA

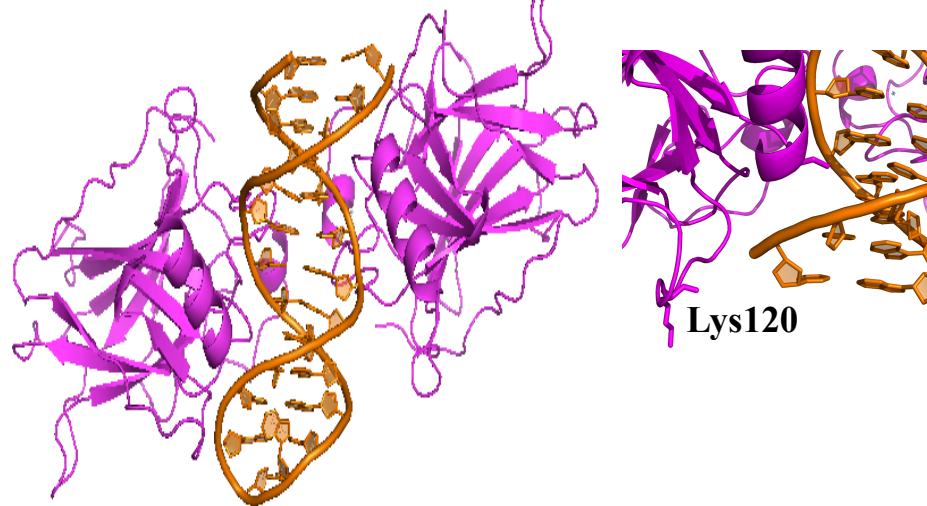


Figure 5.1.c. The molecular mechanism used for p53 to recognize activating response elements.

4. In active response elements containing a central CATG, CAAG and CTTG, Arg248 approaches the minor groove and promotes that Ade flips to Hoosteen geometry.

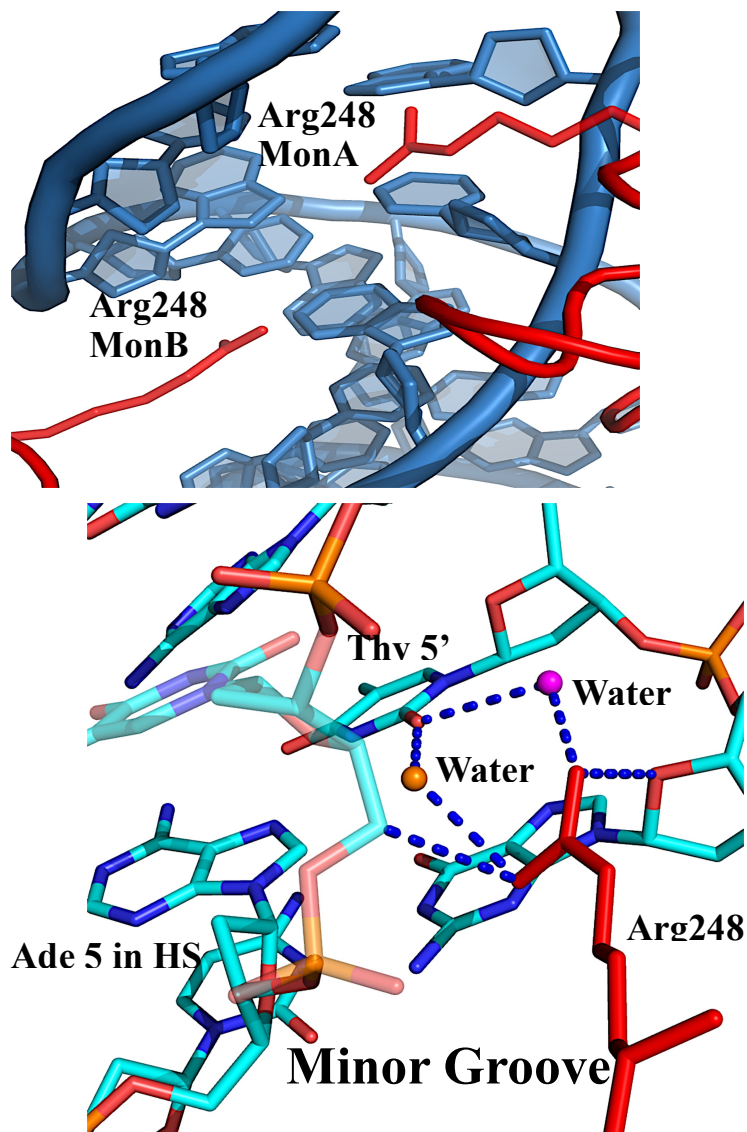


Figure 5.1.d. The molecular mechanism used for p53 to recognize activating response elements.

Bibliography

- [1] G. Bain, E. C. Maandag, D. J. Izon, D. Amsen, A. M. Kruisbeek, B. C. Weintraub, I. Krop, M. S. Schlissel, A. J. Feeney, and M. van Roon, "E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements.," *Cell*, vol. 79, no. 5, pp. 885–92, Dec. 1994.
- [2] D. Accili and K. C. Arden, "FoxOs at the crossroads of cellular metabolism, differentiation, and transformation.," *Cell*, vol. 117, no. 4, pp. 421–6, May 2004.
- [3] I. Simon, J. Barnett, N. Hannett, C. T. Harbison, N. J. Rinaldi, T. L. Volkert, J. J. Wyrick, J. Zeitlinger, D. K. Gifford, T. S. Jaakkola, and R. A. Young, "Serial regulation of transcriptional regulators in the yeast cell cycle.," *Cell*, vol. 106, no. 6, pp. 697–708, Sep. 2001.
- [4] M. M. Babu, N. M. Luscombe, L. Aravind, M. Gerstein, and S. A. Teichmann, "Structure and evolution of transcriptional regulatory networks.," *Curr. Opin. Struct. Biol.*, vol. 14, no. 3, pp. 283–91, Jun. 2004.
- [5] J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe, "A census of human transcription factors: function, expression and evolution.," *Nat. Rev. Genet.*, vol. 10, no. 4, pp. 252–63, Apr. 2009.
- [6] A. Jolma, J. Yan, T. Whittington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J. M. Vaquerizas, R. Vincentelli, N. M. Luscombe, T. R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, and J. Taipale, "DNA-binding specificities of human transcription factors.," *Cell*, vol. 152, no. 1–2, pp. 327–39, Jan. 2013.
- [7] P. H. von Hippel and O. G. Berg, "On the specificity of DNA-protein interactions.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 83, no. 6, pp. 1608–12, Mar. 1986.
- [8] L. Jen-Jacobson, "Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state.," *Biopolymers*, vol. 44, no. 2, pp. 153–80, Jan. 1997.
- [9] M. Suzuki, S. E. Brenner, M. Gerstein, and N. Yagi, "DNA recognition code of transcription factors.," *Protein Eng.*, vol. 8, no. 4, pp. 319–28, Apr. 1995.
- [10] R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann, and B. Honig, "The role of DNA shape in protein-DNA recognition.," *Nature*, vol. 461, no. 7268, pp. 1248–53, Oct. 2009.
- [11] N. C. Seeman, J. M. Rosenberg, and A. Rich, "Sequence-specific recognition of double helical nucleic acids by proteins.," *Proc. Natl. Acad. Sci.*, vol. 73, no. 3, pp. 804–808, Mar. 1976.

- [12] R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, and R. S. Mann, "Origins of specificity in protein-DNA recognition.," *Annu. Rev. Biochem.*, vol. 79, pp. 233–69, Jan. 2010.
- [13] Z. Shakked, G. Guzikovich-Guerstein, F. Frolow, D. Rabinovich, A. Joachimiak, and P. B. Sigler, "Determinants of repressor/operator recognition from the structure of the trp operator binding site.," *Nature*, vol. 368, no. 6470, pp. 469–73, Mar. 1994.
- [14] Z. Otwinowski, R. W. Schevitz, R. G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, and P. B. Sigler, "Crystal structure of trp repressor/operator complex at atomic resolution.," *Nature*, vol. 335, no. 6188, pp. 321–9, Sep. 1988.
- [15] N. M. Luscombe, S. E. Austin, H. M. Berman, and J. M. Thornton, "An overview of the structures of protein-DNA complexes.," *Genome Biol.*, vol. 1, no. 1, p. REVIEWS001, Jan. 2000.
- [16] S. Bell, C. Klein, L. Müller, S. Hansen, and J. Buchner, "p53 contains large unstructured regions in its native state.," *J. Mol. Biol.*, vol. 322, no. 5, pp. 917–27, Oct. 2002.
- [17] N. Pham, A. Lucumi, N. Cheung, and H. Viadiu, "The tetramer of p53 in the absence of DNA forms a relaxed quaternary state.," *Biochemistry*, vol. 51, no. 41, pp. 8053–5, Oct. 2012.
- [18] R. Aramayo, M. B. Sherman, K. Brownless, R. Lurz, A. L. Okorokov, and E. V Orlova, "Quaternary structure of the specific p53-DNA complex reveals the mechanism of p53 mutant dominance.," *Nucleic Acids Res.*, vol. 39, no. 20, pp. 8960–71, Nov. 2011.
- [19] K. L. Harms and X. Chen, "The functional domains in p53 family proteins exhibit both common and distinct properties.," *Cell Death Differ.*, vol. 13, no. 6, pp. 890–7, Jun. 2006.
- [20] R. Dawson, L. Müller, A. Dehner, C. Klein, H. Kessler, and J. Buchner, "The N-terminal Domain of p53 is Natively Unfolded.," *J. Mol. Biol.*, vol. 332, no. 5, pp. 1131–1141, Oct. 2003.
- [21] O. M. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, "Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database," *Hum Mutat*, vol. 28, no. 6, pp. 622– 629, 2013.
- [22] Y. Cho, S. Gorina, P. Jeffrey, and N. Pavletich, "Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations," *Science (80-.).*, vol. 265, no. 5170, pp. 346–355, Jul. 1994.
- [23] K. Zhao, X. Chai, K. Johnston, A. Clements, and R. Marmorstein, "Crystal structure of the mouse p53 core DNA-binding domain at 2.7 Å resolution.," *J. Biol. Chem.*, vol. 276, no. 15, pp. 12120–7, Apr. 2001.

- [24] K. Sakaguchi, H. Sakamoto, D. Xie, J. W. Erickson, M. S. Lewis, C. W. Anderson, and E. Appella, "Effect of phosphorylation on tetramerization of the tumor suppressor protein p53.," *J. Protein Chem.*, vol. 16, no. 5, pp. 553–6, Jul. 1997.
- [25] T. Inoue, J. Stuart, R. Leno, and C. G. Maki, "Nuclear import and export signals in control of the p53-related protein p73.," *J. Biol. Chem.*, vol. 277, no. 17, pp. 15053–60, Apr. 2002.
- [26] T. Soussi, K. Dehouche, and C. Bérout, "p53 website and analysis of p53 gene mutations in human cancer: forging a link between epidemiology and carcinogenesis.," *Hum. Mutat.*, vol. 15, no. 1, pp. 105–13, Jan. 2000.
- [27] C. Götz, P. Scholtes, A. Prowald, N. Schuster, W. Nastainczyk, and M. Montenarh, "Protein kinase CK2 interacts with a multi-protein binding domain of p53.," *Mol. Cell. Biochem.*, vol. 191, no. 1–2, pp. 111–20, Jan. 1999.
- [28] C. Delphin, K. P. Huang, C. Scotto, A. Chapel, M. Vincon, E. Chambaz, J. Garin, and J. Baudier, "The in vitro phosphorylation of p53 by calcium-dependent protein kinase C--characterization of a protein-kinase-C-binding site on p53.," *Eur. J. Biochem.*, vol. 245, no. 3, pp. 684–92, May 1997.
- [29] M. S. Cosgrove, K. Bever, J. L. Avalos, S. Muhammad, X. Zhang, and C. Wolberger, "The structural basis of sirtuin substrate affinity.," *Biochemistry*, vol. 45, no. 24, pp. 7511–21, Jun. 2006.
- [30] K. McKinney, M. Mattia, V. Gottifredi, and C. Prives, "p53 linear diffusion along DNA requires its C terminus.," *Mol. Cell*, vol. 16, no. 3, pp. 413–24, Nov. 2004.
- [31] M. H. Kubbutat, R. L. Ludwig, M. Ashcroft, and K. H. Vousden, "Regulation of Mdm2-directed degradation by the C terminus of p53.," *Mol. Cell. Biol.*, vol. 18, no. 10, pp. 5690–8, Oct. 1998.
- [32] M. S. Rodriguez, J. M. Desterro, S. Lain, D. P. Lane, and R. T. Hay, "Multiple C-terminal lysine residues target p53 for ubiquitin-proteasome-mediated degradation.," *Mol. Cell. Biol.*, vol. 20, no. 22, pp. 8458–67, Nov. 2000.
- [33] C. Dai and W. Gu, "p53 post-translational modification: deregulated in tumorigenesis.," *Trends Mol. Med.*, vol. 16, no. 11, pp. 528–36, Nov. 2010.
- [34] W. S. el-Deiry, S. E. Kern, J. A. Pietenpol, K. W. Kinzler, and B. Vogelstein, "Definition of a consensus binding site for p53.," *Nat. Genet.*, vol. 1, no. 1, pp. 45–9, Apr. 1992.
- [35] T. Riley, E. Sontag, P. Chen, and A. Levine, "Transcriptional control of human p53-regulated genes.," *Nat. Rev. Mol. Cell Biol.*, vol. 9, no. 5, pp. 402–12, May 2008.
- [36] M. Kitayner, H. Rozenberg, N. Kessler, D. Rabinovich, L. Shaulov, T. E. Haran, and Z. Shakked, "Structural basis of DNA recognition by p53 tetramers.," *Mol. Cell*, vol. 22, no. 6, pp. 741–53, Jun. 2006.

- [37] H. Viadiu, "Molecular architecture of tumor suppressor p53.," *Curr. Top. Med. Chem.*, vol. 8, pp. 1327–1334, 2008.
- [38] S. Nguyen, "Structural Basis of Hoogsteen Base Pair Recognition by the DNA-Binding Domain of the Transcription Factor p53," University of California, San Diego, 2012.
- [39] R. L. Weinberg, D. B. Veprintsev, and A. R. Fersht, "Cooperative binding of tetrameric p53 to DNA," *J. Mol. Biol.*, vol. 341, no. 5, pp. 1145–1159, Aug. 2004.
- [40] K. Schlereth, R. Beinoraviciute-Kellner, M. K. Zeitlinger, A. C. Bretz, M. Sauer, J. P. Charles, F. Vogiatzi, E. Leich, B. Samans, M. Eilers, C. Kisker, A. Rosenwald, and T. Stiewe, "DNA binding cooperativity of p53 modulates the decision between cell-cycle arrest and apoptosis.," *Mol. Cell*, vol. 38, no. 3, pp. 356–68, May 2010.
- [41] D. P. Lane and L. V. Crawford, "T antigen is bound to a host protein in SY40-transformed cells," *Nature*, vol. 278, no. 5701, pp. 261–263, Mar. 1979.
- [42] D. I. H. Linzer and A. J. Levine, "Characterization of a 54K Dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells," *Cell*, vol. 17, no. 1, pp. 43–52, May 1979.
- [43] C. A. Finlay, P. W. Hinds, and A. J. Levine, "The p53 proto-oncogene can act as a suppressor of transformation.," *Cell*, vol. 57, no. 7, pp. 1083–93, Jun. 1989.
- [44] M. Olivier, M. Hollstein, and P. Hainaut, "TP53 mutations in human cancers: origins, consequences, and clinical use.," *Cold Spring Harb. Perspect. Biol.*, vol. 2, no. 1, p. a001008, Jan. 2010.
- [45] P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, and N. P. Pavletich, "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain.," *Science*, vol. 274, no. 5289, pp. 948–53, Nov. 1996.
- [46] O. Schon, A. Friedler, M. Bycroft, S. M. V Freund, and A. R. Fersht, "Molecular mechanism of the interaction between MDM2 and p53.," *J. Mol. Biol.*, vol. 323, no. 3, pp. 491–501, Oct. 2002.
- [47] S. R. Grossman, M. E. Deato, C. Brignone, H. M. Chan, A. L. Kung, H. Tagami, Y. Nakatani, and D. M. Livingston, "Polyubiquitination of p53 by a ubiquitin ligase activity of p300.," *Science*, vol. 300, no. 5617, pp. 342–4, Apr. 2003.
- [48] C. L. Brooks and W. Gu, "Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation.," *Curr. Opin. Cell Biol.*, vol. 15, no. 2, pp. 164–71, Apr. 2003.
- [49] Y. Tang, W. Zhao, Y. Chen, Y. Zhao, and W. Gu, "Acetylation is indispensable for p53 activation.," *Cell*, vol. 133, no. 4, pp. 612–26, May 2008.

- [50] Y. Zhang, Y. Xiong, and W. G. Yarbrough, "ARF promotes MDM2 degradation and stabilizes p53: ARF-INK4a locus deletion impairs both the Rb and p53 tumor suppression pathways.," *Cell*, vol. 92, no. 6, pp. 725–34, Mar. 1998.
- [51] Z. Goldberg, R. Vogt Sionov, M. Berger, Y. Zwang, R. Perets, R. A. Van Etten, M. Oren, Y. Taya, and Y. Haupt, "Tyrosine phosphorylation of Mdm2 by c-Abl: implications for p53 regulation.," *EMBO J.*, vol. 21, no. 14, pp. 3715–27, Jul. 2002.
- [52] E. Appella and C. W. Anderson, "Post-translational modifications and activation of p53 by genotoxic stresses.," *Eur. J. Biochem.*, vol. 268, no. 10, pp. 2764–72, May 2001.
- [53] B. Wang, Z. Xiao, and E. C. Ren, "Redefining the p53 response element.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 34, pp. 14373–8, Aug. 2009.
- [54] G. Melino, V. De Laurenzi, and K. H. Vousden, "p73: Friend or foe in tumorigenesis.," *Nat. Rev. Cancer*, vol. 2, no. 8, pp. 605–15, Aug. 2002.
- [55] Y.-L. Lin, S. Sengupta, K. Gurdziel, G. W. Bell, T. Jacks, and E. R. Flores, "p63 and p73 transcriptionally regulate genes involved in DNA repair.," *PLoS Genet.*, vol. 5, no. 10, p. e1000680, Oct. 2009.
- [56] A. S. Ethayathulla, P.-W. Tse, P. Monti, S. Nguyen, A. Inga, G. Fronza, and H. Viadiu, "Structure of p73 DNA-binding domain tetramer modulates p73 transactivation.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 16, pp. 6066–71, Apr. 2012.
- [57] H. S. Bell and K. M. Ryan, "Targeting the p53 Family for Cancer Therapy: 'Big Brother' Joins the Fight," *Cell Cycle*, vol. 6, no. 16, pp. 1995–2000, Oct. 2014.
- [58] M. Kaghad, H. Bonnet, A. Yang, L. Creancier, J.-C. Biscan, A. Valent, A. Minty, P. Chalon, J.-M. Lelias, X. Dumont, P. Ferrara, F. McKeon, and D. Caput, "Monoallelically Expressed Gene Related to p53 at 1p36, a Region Frequently Deleted in Neuroblastoma and Other Human Cancers," *Cell*, vol. 90, no. 4, pp. 809–819, Aug. 1997.
- [59] C. D. Pozniak, S. Radinovic, A. Yang, F. McKeon, D. R. Kaplan, and F. D. Miller, "An anti-apoptotic role for the p53 family member, p73, during developmental neuron death.," *Science*, vol. 289, no. 5477, pp. 304–6, Jul. 2000.
- [60] M. S. Irwin and W. G. Kaelin, "p53 family update: p73 and p63 develop their own identities.," *Cell Growth Differ.*, vol. 12, no. 7, pp. 337–49, Jul. 2001.
- [61] V. De Laurenzi, A. Costanzo, D. Barcaroli, A. Terrinoni, M. Falco, M. Annicchiarico-Petruzzelli, M. Levrero, and G. Melino, "Two new p73 splice variants, gamma and delta, with different transcriptional activity.," *J. Exp. Med.*, vol. 188, no. 9, pp. 1763–8, Nov. 1998.
- [62] V. D. De Laurenzi, M. V. Catani, A. Terrinoni, M. Corazzari, G. Melino, A. Costanzo, M. Levrero, and R. A. Knight, "Additional complexity in p73: induction by mitogens in

- lymphoid cells and identification of two new splicing variants epsilon and zeta.," *Cell Death Differ.*, vol. 6, no. 5, pp. 389–90, May 1999.
- [63] Y. Ueda, M. Hijikata, S. Takagi, T. Chiba, and K. Shimotohno, "Transcriptional activities of p73 splicing variants are regulated by inter-variant association.," *Biochem. J.*, vol. 356, no. Pt 3, pp. 859–66, Jun. 2001.
 - [64] C. A. Jost, M. C. Marin, and W. G. Kaelin, "p73 is a simian [correction of human] p53-related protein that can induce apoptosis.," *Nature*, vol. 389, no. 6647, pp. 191–4, Sep. 1997.
 - [65] O. Ishimoto, C. Kawahara, K. Enjo, M. Obinata, T. Nukiwa, and S. Ikawa, "Possible oncogenic potential of DeltaNp73: a newly identified isoform of human p73.," *Cancer Res.*, vol. 62, no. 3, pp. 636–41, Feb. 2002.
 - [66] T. J. Grob, U. Novak, C. Maisse, D. Barcaroli, A. U. Lüthi, F. Pirnia, B. Hügli, H. U. Graber, V. De Laurenzi, M. F. Fey, G. Melino, and A. Tobler, "Human delta Np73 regulates a dominant negative feedback loop for TAp73 and p53.," *Cell Death Differ.*, vol. 8, no. 12, pp. 1213–23, Dec. 2001.
 - [67] S. Patel, T. T. T. Bui, A. F. Drake, F. Fraternali, and P. V Nikolova, "The p73 DNA binding domain displays enhanced stability relative to its homologue, the tumor suppressor p53, and exhibits cooperative DNA binding.," *Biochemistry*, vol. 47, no. 10, pp. 3235–44, Mar. 2008.
 - [68] Y. Ciribilli, P. Monti, A. Bisio, H. T. Nguyen, A. S. Ethayathulla, A. Ramos, G. Foggetti, P. Menichini, D. Menendez, M. a Resnick, H. Viadiu, G. Fronza, and A. Inga, "Transactivation specificity is conserved among p53 family proteins and depends on a response element sequence code.," *Nucleic Acids Res.*, vol. 41, no. 18, pp. 8637–53, Oct. 2013.
 - [69] M. Kitayner, H. Rozenberg, R. Rohs, O. Suad, D. Rabinovich, B. Honig, and Z. Shakked, "Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs.," *Nat. Struct. Mol. Biol.*, vol. 17, no. 4, pp. 423–9, Apr. 2010.
 - [70] K. Harms, S. Nozell, and X. Chen, "The common and distinct target genes of the p53 family transcription factors.," *Cell. Mol. Life Sci.*, vol. 61, no. 7–8, pp. 822–42, Apr. 2004.
 - [71] J. R. Lakowics, *Principles of Fluorescence Spectroscopy*, 3rd Editio. 2006, pp. 553–582.
 - [72] A. Tse, "DNA binding specificity of p73 DNA binding domain," 2012.
 - [73] A. S. Ethayathulla, H. T. Nguyen, and H. Viadiu, "Crystal structures of the DNA-binding domain tetramer of the p53 tumor suppressor family member p73 bound to different full-site response elements.," *J. Biol. Chem.*, vol. 288, no. 7, pp. 4744–54, Feb. 2013.

- [74] Y. Wang, M. Reed, P. Wang, J. E. Stenger, G. Mayr, M. E. Anderson, J. F. Schwedes, and P. Tegtmeier, "p53 domains: identification and characterization of two autonomous DNA-binding regions.," *Genes Dev.*, vol. 7, no. 12B, pp. 2575–86, Dec. 1993.
- [75] N. P. Pavletich, K. A. Chambers, and C. O. Pabo, "The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots.," *Genes Dev.*, vol. 7, no. 12B, pp. 2556–64, Dec. 1993.
- [76] K. Iwabuchi, B. Li, P. Bartel, and S. Fields, "Use of the two-hybrid system to identify the domain of p53 involved in oligomerization.," *Oncogene*, vol. 8, no. 6, pp. 1693–6, Jun. 1993.
- [77] Y. Pan and R. Nussinov, "Structural basis for p53 binding-induced DNA bending.," *J. Biol. Chem.*, vol. 282, no. 1, pp. 691–9, Jan. 2007.
- [78] T. Dobner, N. Horikoshi, S. Rubenwolf, and T. Shenk, "Blockage by adenovirus E4orf6 of transcriptional activation by the p53 tumor suppressor.," *Science*, vol. 272, no. 5267, pp. 1470–3, Jun. 1996.
- [79] M. Kaghad, H. Bonnet, A. Yang, L. Creancier, J. C. Biscan, A. Valent, A. Minty, P. Chalon, J. M. Lelias, X. Dumont, P. Ferrara, F. McKeon, and D. Caput, "Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers.," *Cell*, vol. 90, no. 4, pp. 809–19, Aug. 1997.
- [80] A. C. Joerger, S. Rajagopalan, E. Natan, D. B. Veprintsev, C. V Robinson, and A. R. Fersht, "Structural evolution of p53, p63, and p73: implication for heterotetramer formation.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 42, pp. 17705–10, Oct. 2009.
- [81] D. Coutandin, F. Löhr, F. H. Niesen, T. Ikeya, T. A. Weber, B. Schäfer, E. M. Zielonka, A. N. Bullock, A. Yang, P. Güntert, S. Knapp, F. McKeon, H. D. Ou, and V. Dötsch, "Conformational stability and activity of p73 require a second helix in the tetramerization domain.," *Cell Death Differ.*, vol. 16, no. 12, pp. 1582–9, Dec. 2009.
- [82] S. Nozell, Y. Wu, K. McNaughton, G. Liu, A. Willis, J. C. Paik, and X. Chen, "Characterization of p73 functional domains necessary for transactivation and growth suppression.," *Oncogene*, vol. 22, no. 28, pp. 4333–47, Jul. 2003.
- [83] P. Schuck, "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling.," *Biophys. J.*, vol. 78, no. 3, pp. 1606–19, Mar. 2000.
- [84] R. T. M. and P. S. L. Laue T.M., Shah B.D., *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Royal Society of Chemistry, 1992, pp. 90–125.
- [85] E. Shaulian, A. Zauberman, J. Milner, E. A. Davies, and M. Oren, "Tight DNA binding and oligomerization are dispensable for the ability of p53 to transactivate target genes and suppress transformation.," *EMBO J.*, vol. 12, no. 7, pp. 2789–97, Jul. 1993.

- [86] P. Chène, “The role of tetramerization in p53 function.,” *Oncogene*, vol. 393, 2001.
- [87] S. Walker, S. Hayes, and P. O’Hare, “Site-specific conformational alteration of the Oct-1 POU domain-DNA complex as the basis for differential recognition by Vmw65 (VP16).,” *Cell*, vol. 79, no. 5, pp. 841–52, Dec. 1994.
- [88] K. M. Scully, E. M. Jacobson, K. Jepsen, V. Lunyak, H. Viadiu, C. Carrière, D. W. Rose, F. Hooshmand, A. K. Aggarwal, and M. G. Rosenfeld, “Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification.,” *Science*, vol. 290, no. 5494, pp. 1127–31, Nov. 2000.
- [89] H. Neumann, S. Y. Peak-Chew, and J. W. Chin, “Genetically encoding N(epsilon)-acetyllysine in recombinant proteins.,” *Nat. Chem. Biol.*, vol. 4, no. 4, pp. 232–4, Apr. 2008.
- [90] H. Neumann, S. M. Hancock, R. Buning, A. Routh, L. Chapman, J. Somers, T. Owen-Hughes, J. van Noort, D. Rhodes, and J. W. Chin, “A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation.,” *Mol. Cell*, vol. 36, no. 1, pp. 153–63, Oct. 2009.
- [91] E. Arbely, E. Natan, T. Brandt, M. D. Allen, D. B. Veprintsev, C. V Robinson, J. W. Chin, A. C. Joerger, and A. R. Fersht, “Acetylation of lysine 120 of p53 endows DNA-binding specificity at effective physiological salt concentration.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 20, pp. 8251–6, May 2011.
- [92] Z. Otwinowski and W. Minor, “Processing of X-ray Diffraction Data Collected in Oscillation Mode,” *Methods Enzymol.*, vol. 276, 1997.
- [93] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, “Phaser crystallographic software.,” *J. Appl. Crystallogr.*, vol. 40, no. Pt 4, pp. 658–674, Aug. 2007.
- [94] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart, “PHENIX: A comprehensive Python-based system for macromolecular structure solution,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 213–221, 2010.
- [95] B. Rupp, *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. Garland Science and Taylor & Francis Group, LLC, 2010, pp. 1–809.
- [96] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, “Features and development of Coot,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 4, pp. 486–501, 2010.
- [97] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, “MolProbity: all-atom structure validation for macromolecular crystallography.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 1, pp. 12–21, Jan. 2010.

- [98] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, "PROCHECK: a program to check the stereochemical quality of protein structures," *J. Appl. Crystallogr.*, vol. 26, no. 2, pp. 283–291, Apr. 1993.
- [99] Y. Tang, J. Luo, W. Zhang, and W. Gu, "Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis.," *Mol. Cell*, vol. 24, no. 6, pp. 827–39, Dec. 2006.
- [100] S. M. Sykes, H. S. Mellert, M. a Holbert, K. Li, R. Marmorstein, W. S. Lane, and S. B. McMahon, "Acetylation of the p53 DNA-binding domain regulates apoptosis induction.," *Mol. Cell*, vol. 24, no. 6, pp. 841–51, Dec. 2006.
- [101] G. N. Ramachandran and V. Sasisekharan, "Conformation of polypeptides and proteins.," *Adv. Protein Chem.*, vol. 23, pp. 283–438, Jan. 1968.
- [102] S. Costantini, A. Paladino, and A. M. Facchiano, "CALCOM: a software for calculating the center of mass of proteins.," *Bioinformation*, vol. 2, no. 7, pp. 271–2, Jan. 2008.
- [103] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson, "Overview of the CCP4 suite and current developments.," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 67, no. Pt 4, pp. 235–42, Apr. 2011.
- [104] E. B. Saff and A. B. J. Kuijlaars, "Distributing many points on a sphere," *Math. Intell.*, vol. 19, no. 1, pp. 5–11, Dec. 1997.
- [105] W. C. Ho, M. X. Fitzgerald, and R. Marmorstein, "Structure of the p53 core domain dimer bound to DNA.," *J. Biol. Chem.*, vol. 281, no. 29, pp. 20494–502, Jul. 2006.
- [106] G. Murano, "Molecular Dynamics of p73," University of California San Diego, 2014.
- [107] T. J. Petty, S. Emamzadah, L. Costantino, I. Petkova, E. S. Stavridi, J. G. Saven, E. Vauthey, and T. D. Halazonetis, "An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity.," *EMBO J.*, vol. 30, no. 11, pp. 2167–76, Jun. 2011.
- [108] S. Emamzadah, L. Tropaia, and T. D. Halazonetis, "Crystal Structure of a Multidomain Human p53 Tetramer Bound to the Natural CDKN1A (p21) p53-Response Element.," *Molecular Cancer Research*, vol. 9, no. 11, pp. 1493–1499, 2011.
- [109] S. Emamzadah, L. Tropaia, I. Vincenti, B. Falquet, and T. D. Halazonetis, "Reversal of the DNA-binding-induced loop L1 conformational switch in an engineered human p53 protein," *J. Mol. Biol.*, vol. 426, no. 4, pp. 936–944, 2014.