

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Studying Response Element Specificity in the Transcription Factor p73 by Molecular Dynamics

Permalink

<https://escholarship.org/uc/item/75k0p9rv>

Author

Murano, George Anthony

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Studying Response Element Specificity in the Transcription Factor p73
by Molecular Dynamics

A Thesis submitted in partial satisfaction of the requirements for the degree
Master of Science

in

Chemistry

by

George Anthony Murano

Committee in charge:

Professor Hector Viadiu, Chair
Professor Rommie Amaro, Co-Chair
Professor Stanley Opella

2014

Copyright

George Anthony Murano, 2014

All rights reserved.

The Thesis of George Anthony Murano is approved and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2014

DEDICATION

This work is dedicated to

I. A. Bergmann

EPIGRAPH

I made this letter very long, because I did not have the leisure to make it shorter.

Blaise Pascal

TABLE OF CONTENTS

Signature Page.....	iii
Dedication.....	iv
Epigraph.....	v
Table of Contents.....	vi
List of Abbreviations.....	viii
List of Figures.....	ix
List of Tables.....	xiii
Acknowledgements.....	xiv
Vita.....	xv
Abstract of the Thesis.....	xvi
Chapter 1.....	1
Introduction	
A. Significance.....	2
B. The Role of the p53-Family in Transcription Regulation.....	3
C. TP73 and the p53 Gene Family.....	6
D. p73 Structure.....	7
E. The Recognition Element.....	9
F. Motivation of Study.....	11
References.....	14
Chapter 2	21
Methods	
A. Molecular Dynamics (MD) Simulation.....	22
B. System Preparation.....	25
C. MD Simulation Protocol.....	26
D. Analysis.....	28
E. Initial Crystal Structure Dimer Comparison.....	29
F. p53 Response Element Study.....	32
References.....	40

Chapter 3.....	43
Results: Response Element Contact Analysis	
A. Lys-138.....	44
B. Ser-139, Cys-297, and Arg-268.....	46
C. Additional DNA-Binding Residues.....	48
D. Arg-300 and Lys-138.....	49
E. Loop L1 analysis.....	51
References.....	72
Chapter 4.....	73
Results: Quaternary Structure Analysis of the p73 dimer-RE complex	
A. Dimer-Response Element Complex Conformation.....	74
B. GAACA Dimer Conformational Change.....	76
C. Sequence-Independent Dimer Motion.....	77
D. Dimer Interface and Conformation.....	79
E. Mechanism of Conformational Change.....	80
Chapter 5.....	93
Discussion	
A. Consequences of Guanine-Rich Sequences.....	94
B. Model for p73 Response Element Recognition.....	97
References.....	111
Appendix.....	112

LIST OF ABBREVIATIONS

A	adenine
AUC	analytical ultracentrifugation
bp	base pair
C	cytosine
CTD	C-terminal domain
K _d	dissociation constant
DBD	DNA-binding domain
G	guanine
GTFs	general transcription factors
OD	oligomerization domain
Pu	purine
Py	pyrimidine
RE	response element
RMSD	root mean square deviation
RNAPII	RNA polymerase II
SAM	sterile alpha motif
SAS	solvent assessable surface
T	thymine
TA	transactivation domain
TSS	transcription start site
W	adenine or thymine

LIST OF FIGURES

Figure 2.1. Diagram summarizing MD protocol for simulation.....	28
Figure 2.2. p53 Response Element Analysis.....	39
Figure 3.1. Summary of Lys-138 Hydrogen Bond Contacts to the Response Element.	55
Figure 3.2. Snapshot of Lys-138 in Contact with Response Element Positions 1 – 3...	56
Figure 3.3. Lys-138 Contacts to DNA Monitored by Distance (Part 1).....	57
Figure 3.4. Lys-138 Contacts to DNA Monitored by Distance (Part 2).....	58
Figure 3.5. Snapshot of Ser-139 in Contact with Response Element Positions 1 – 3...	59
Figure 3.6. Ser-139 Contacts to DNA Monitored by Distance (Part 1).....	60
Figure 3.7. Ser-139 Contacts to DNA Monitored by Distance (Part 2).....	61
Figure 3.8. Distance between Cys-297 and position-3 nucleotide.....	62
Figure 3.9. Snapshot of Arg-268 in Contact with Response Element Positions 1 – 3...	63
Figure 3.10. Arg-268 Contacts to DNA Monitored by Distance (Part 1).....	64
Figure 3.11. Arg-268 Contacts to DNA Monitored by Distance (Part 2).....	65
Figure 3.12. Summary of DNA-Binding Residue hydrogen bond percentage.....	67
Figure 3.13. Distance between the carbon chains of Lys-138 and Arg-300.....	68
Figure 3.14. Percentage of time that the loop L1 and helix H2 hydrogen bonds occur during the trajectory.....	69
Figure 4.1. Surface Image of Dimer-Response Element Complex.....	84
Figure 4.2. Backbone Dimer-Response Element Complex Image.....	85
Figure 4.3. Dimer-Axis Rotational States.....	87
Figure 4.4. Intra-Dimer Rotation.....	88
Figure 4.5. Transition of GGGCA_4g82_swap from open to closed Dimer-DNA conformation.....	90

Figure 4.6. Transition of GAACA_3vd1_swap to open Dimer-DNA complex.....	91
Figure 4.7. GAACA_4g82 Transition to relaxed state.....	92
Figure 5.1. Conformational arrangement of p73 tetramer with and without guanine-rich quarter site.....	105
Figure 5.2. Dimer and Helix H2 Rotation between Specifically and Nonspecifically bound dimers.....	106
Figure 5.3. Distances between helix H2s of a crystal structure of p73-DBD Tetramer Bound to a 0-Spacer Full-Site Response Element.....	107
Figure 5.4. Possible p73 arrangements of OD in unbound and bound states.....	108
Figure 5.5. Possible Mechanism for p73 Recognition of Nonspecific DNA and Response Element.....	109
Figure 5.6. P73 RE Recognition Models Based on OD Connectivity.....	110
Figure A.1. Distance between Ala-296 and position-4 phosphate.....	113
Figure A.2. Distance between Asn-259 and position-4 phosphate.....	114
Figure A.3. Distance between Ser-261 and position-4 phosphate.....	115
Figure A.4. Distance between Arg-293 and position-5 phosphate of GGGCA_3vd1 and GGGCA_4g82_swap systems.....	116
Figure A.5. Distance between Arg-293 and position-5 phosphate of GAACA_4g82 and GAACA_3vd1_swap systems.....	117
Figure A.6. Distance between Arg-300 and guanine complimentary to position-four cystosine for GGGCA_3vd1 system.....	118
Figure A.7. Distance between Arg-300 and guanine complimentary to position four cystosine for GGGCA_4g82_swap system.....	119
Figure A.8. Distance between Arg-300 and guanine complimentary to position four cystosine for GAACA_4g82 system.....	120
Figure A.9. Distance between Arg-300 and guanine complimentary to position four cystosine for GAACA_3vd1_swap system.....	121

Figure A.10. Distance between hydrogen bond partners of contacts concerning intraloop L1(top), and between loop L1 and helix H2(bottom) for GGGCA_3vd1 system.....	122
Figure A.11. Distance between hydrogen bond partners of contacts concerning intraloop L1(top), and between loop L1 and helix H2(bottom) for GGGCA_4g82_swap system.....	123
Figure A.12. Distance between hydrogen bond partners of contacts concerning intraloop L1(top), and between loop L1 and helix H2(bottom) for GAACA_4g82 system.....	124
Figure A.13. Distance between hydrogen bond partners of contacts concerning intraloop L1(top), and between loop L1 and helix H2(bottom) for GAACA_3vd1_swap system.....	125
Figure A.14. RMSD of loop L1 with respect GAACA(top) and GGGCA(bottom) crystal structures for the GGGCA_3vd1 system.....	126
Figure A.15. RMSD of loop L1 with respect GAACA(top) and GGGCA(bottom) crystal structures for the GGGCA_4g82_swap system.....	127
Figure A.16. RMSD of loop L1 with respect GAACA(top) and GGGCA(bottom) crystal structures for the GAACA_4g82 system.....	128
Figure A.17. RMSD of loop L1 with respect GAACA(top) and GGGCA(bottom) crystal structures for the GAACA_3vd1_swap system.....	129
Figure A.18. B-factor of monomers A and B of GGGCA(3vd1) crystal structure.....	130
Figure A.19. B-factor of monomers A and B of GAACA(4g82) crystal structure.....	131
Figure A.20. Principal axis angle between dimer and DNA for all systems.....	132
Figure A.21. Dimer Rotation Through Dimer Principal Axis.....	133
Figure A.22. Dimer Rotation with respect to plane of dimer.....	134
Figure A.23. Monitoring residue distances across monomers representing van der Waal interactions in the dimer interface for the GGGCA systems.....	135
Figure A.24. Monitoring residue distances across monomers representing van der Waal interactions in the dimer interface for the GAACA systems.....	136

Figure A.25. Monitoring of potential hydrogen bond partners in the dimer interface, Part 1.....	137
Figure A.26. Monitoring of potential hydrogen bond partners in the dimer interface, Part 2.....	138
Figure A. 27. Monitoring of potential hydrogen bond partners in the dimer interface, Part 3.....	139

LIST OF TABLES

Table 2.1. MD Simulation Systems.....	26
Table 2.2. RMSD Comparison of 0 bp and 2 bp spacer dimers with AGGCA response element as well as 1 bp and 2 bp spacer dimers with GGGCA response element.....	34
Table 2.3. RMSD Comparison of AGGCA and GAACA 0 bp spacer dimers aswell as AGGCA and GGGCA 2 bp spacer dimers.....	35
Table 2.4. RMSD Comparison of 0 and 1 bp spacer dimers regardless of response element sequence.....	36
Table 2.5. RMSD Comparison between dimers belonging to the same tetramer.....	37
Table 2.6. RMSD comparison of selected monomers.....	38
Table 3.1. Hydrogen bond percentage of Ser-139, Cys-297, and Arg-268 to DNA...	66
Table 3.2. Number of clusters and percentage of the four most representative clusters for loop L1 generated from RMSD-cutoff of 1.36 Å.....	70
Table 3.3. RMSD Comparison between loop L1 of cluster 1 representative structure and GAACA, and GGGCA crystal structures.....	71
Table 4.1. Average Angle between Dimer and DNA Principal Axes from 60 – 130 ns of simulation.....	86
Table 4.2. Buried SAS Area of Dimer Interface.....	89

ACKNOWLEDGEMENTS

This study is the brainchild of a man obsessed (in the good sense ☺, an attribute worthy of any great scientist) with trying to understand the p53-family. I have only been fortunate to be a pupil of his for the last few years. I am eternally grateful in more ways than he can imagine for his guidance and support in every respect. Thank you Prof. Viadiu.

The work presented here would not have been possible without the help and support of Prof. Rommie Amaro and Dr. Ozlem Demir. Rommie exposed me to the basic principles of molecular dynamics as a student in her computational chemistry class. She has been incredibly gracious and generous throughout this project, particularly in regards to the computational resources, which allotted me the opportunity to do this study. Ozlem has been instrumental in helping me in the critical day-to-day problems of a molecular dynamics project. I want to thank you so much for your patience and constant support that you have given me.

Lastly, and certainly not least, I would like to thank the incredible environment that was fostered in the Viadiu lab due to the magnificent people that composed our group. In particular, I wish to thank Ana Ramos, Thien Nguyen, and Jill Tee-Sy who have withstood the longest tenure with me there, and all of the former members that have made the lab an incredibly warm and comforting place to be.

VITA

- 2012 Bachelor of Science with Majors in Bioengineering: Pre-Med and
Biochemistry/Chemistry
University of California, San Diego
- 2012 – 2014 Teaching Assistant, Department of Chemistry and Biochemistry
University of California, San Diego
- 2014 Master of Science in Chemistry
University of California, San Diego

FIELDS OF STUDY

Major Field: Biochemistry

Studies in Structural Biology and Biophysics
Professor Hector Viadiu

Studies in Computational Chemistry
Professor Rommie Amaro

ABSTRACT OF THE THESIS

Studying Response Element Specificity in the Transcription Factor p73
by Molecular Dynamics

by

George Anthony Murano

Master of Science in Chemistry

University of California, San Diego, 2014

Professor Hector Viadiu, Chair
Professor Rommie Amaro, Co-Chair

p73 is responsible for the regulation of many of the cell's critical life processes along with the entire p53-family. Additionally, due to its transcriptional cross-over present among the p53-family members, great promise is held for p73 to substitute for the tumor suppression function absent in mutant-p53, which is present in 50 % of all cancers.

Critically important for the regulation of transcription is the response element (RE) sequence recognition by the p53-family. Although some is known about how the response element affects the level of transcription, little is understood, in particular concerning the mechanism for how transcription is regulated for p53-family genes. In this work, molecular dynamics (MD) is used to study the recognition of the response element flanking-site by the p73-DBD dimer. In particular, two response elements sequences are examined: an apoptotically relevant, guanine-rich RE, and a cell cycle arrest/DNA repair relevant, adenine-rich RE. The results highlight a conformational change of the p73-DBD/RE complex as a result of the guanine-rich sequences, suggesting a distinct mechanism of transcription regulation compared to RE's that lack a guanine-rich flanking-site. In addition, simulations showed conformational states of a partially unbound p73-DBD/RE complex that serve to propose a model for p73 recognition of nonspecific DNA and REs.

CHAPTER 1

INTRODUCTION

A. Significance

The p53-family forms an incredibly important set of transcription factors, responsible for the regulation of many of the cell's critical life processes. Individually, the p53 family proteins regulates genes related to apoptosis, cell cycle repair, senescence, autophagy, chromosome stability, neurogenesis, inflammation, development, morphogenesis, and many more (Crum & McKeon, 2010; Kaghad et al., 1997; Mills et al., 1999; Murray-Zmijewski, Slee, & Lu, 2008; Pozniak et al., 2000; Vousden & Prives, 2009; Yang et al., 1999). And, together, they potentially regulate a collection of thousands of genes, in a manner that looks as diverse as the number of genes they control (Smeenk et al., 2008). Since the discovery of p53 in 1979, the p53-family has flooded the field with over 60,000 publications, which have contributed to the rapid expansion of the field, despite many publications not withstanding the test of time (Lane & Levine, 2010). An important contribution to this voluminous and large interest in the p53-family was its link to cancer. Importantly, a p53 mutation has been found to be present in at least 50% of all cancers (from the International Agency for Cancer Research TP53 Mutation Database and the TP53 Website). This, in addition to many other manifestations of p53's connection to matters of cell fate, led to p53's classification as a tumor suppressor, which subsequently has also been extended to include p73 (Tomasini et al., 2008). Unlike p53, the p73 mutations in human tumors are relatively rare, representing only 0.5% (Melino, De Laurenzi, & Vousden, 2002). Interestingly, p73 along with p63 have shown to be able to activate many p53-dependent genes (Lin et al., 2009). The facts that p73 shows little evidence of mutation in cancers and that its function also overlaps with the activation of

p53 pathways related to apoptosis show promise for p73 to substitute the tumor suppressor function of p53 when the functionality of p53 is compromised due to mutation (Dotsch, Bernassola, Coutandin, Candi, & Melino, 2010; Yoon, Ha, Lee, & Chi, 2014). Additionally, structural studies of mutant p73 have served to inspire methods of reactivating p53 (Viadiu, Fronza, & Inga, 2014). The p53-family shows a remarkable similarity across the family members, evidenced by 63% sequence identity between the p73 and p53 DNA-binding domains (DBD) (p63 has a similar percentage with p53) (Dotsch et al., 2010). This similarity consequently explains, in part, p53-family's ability to cross-over and activate several of the other family member's associated genes. In addition to p73's hopeful promise of rescuing the tumor suppression function of a p53-debilitated cell, p73 serves in many respects to expand the understanding of the p53-family regulation of transcription.

In relation to the work presented here, the subsequent sections of this chapter will present the regulatory context of the p53-family, specifically the basic transcription machinery and p53-family's role in transcription activation (Section B), describe the TP73 gene (Section C), highlight the known structure characteristics of p73 (Section D), define the p53-family recognition site (Section E), and lastly, describe the motivation for this study (Section F).

B. The Role of the p53-Family in Transcription Regulation

The general role of the p53-family, like most transcription factors, is to assist in the regulation of the gene transcription process. The predominant manner that the p53-

family regulates transcription is by binding to promoter sites or response elements. These response elements lie predominantly upstream, but also downstream of the transcription start site (TSS), however less commonly in introns and exons (Riley, Sontag, Chen, & Levine, 2008). Not only is the p53-family crucial for activating transcription, but also is responsible for the repression of many genes. Repression seems to take on several forms which can include occlusion of a transactivating protein binding site due to binding site-overlap (Budhram-Mahadeo et al., 1999; K. C. Lee, Crowe, & Barton, 1999; Ori et al., 1998), inhibiting proper interaction of transactivating proteins through protein-protein contact (Innocente & Lee, 2005a, 2005b; Kanaya et al., 2000; Ohlsson, Kley, Werner, & LeRoith, 1998), as well as inducing histone modifications in chromatin (Murphy et al., 1999). Activation seems to be more closely associated with direct binding and p53, in particular, activates transcription by recruiting general transcription factors to assemble the basal transcription complex (Farmer, Colgan, Nakatani, Manley, & Prives, 1996; Thut, Chen, Klemm, & Tjian, 1995) as well as enzymes that are important to chromatin remodeling through histone modifications (Gu & Roeder, 1997; Gu, Shi, & Roeder, 1997). In addition to binding to the appropriate response element, the placement of the p53-family recognition site has been generally linked to transcription control such that response elements closer to the TSS increase the transactivation potential (Riley et al., 2008). The interaction of the p53-family member with the transcription initiation complex is further nuanced by the fact that DNA looping can occur, similar as in the case of transcription factor-binding proteins, through several distally bound p53 complexes stacking together to bring separated DNA regions closer (Stenger et al., 1994).

Due to the p53-family's important role in transcription initiation, the basal transcriptional complex, minimally responsible for the initiation of transcription, is briefly described here. The catalytic actor in transcribing the mammalian genetic code, RNA polymerase II (RNAPII), forms a pre-initiation complex (PIC) along with a set of general transcription factors (GTFs) that include TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, and TFIIH (Butler & Kadonaga, 2002). This complex is assembled on the core promoter, which provides the appropriate set of DNA regulatory motifs that bind subunits of the aforementioned GTFs and stretches roughly ~40 bps upstream and downstream of the transcription start site (TSS or +1 position) (Smale, 2001). The assembly of the PIC onto the core promoter is initiated by the TFIID, which is critical to the nucleation of the basal transcription complex by binding to promoters that specify the +1 position (Albright & Tjian, 2000; Burley & Roeder, 1996). These promoters can include the TATA box, the initiator element (Inr), the downstream promoter element (DPE), as well as the downstream core element (CDC), which all can be bound by distinct subunits of the TFIID. The binding of TFIID is subsequently followed by TFIIB, which can recognize the TFIIB recognition element (BRE) directly upstream and downstream of the TATA box (Deng & Roberts, 2005, 2007). The assembly of the PIC is completed with the binding of TFIIF, TFIIIE, and TFIIH in complex with unphosphorylated RNAPII (Orphanides, Lagrange, & Reinberg, 1996; Woychik & Hampsey, 2002). In relation to the p53-family, p53 is specifically able to help recruit several PIC components to the core promoter such as TBP, a TFIID subunit (X. Chen, Farmer, Zhu, Prywes, & Prives, 1993; Liu, Miller, Koeffler, & Berk, 1993; Seto et al., 1992), TFIIA and TFIIH (Ko & Prives, 1996; Xing, Sheppard, Corneillie, & Liu, 2001).

C. TP73 and the p53 Gene Family

Part of the diversity exhibited by the p53-family in transcription regulation is due to the diverse set of transcribed isoforms, particularly in p63 and p73. Common to all members of the p53 family, their genes encode for a N-terminal transactivation domain (TA) largely used for recruitment of other transcription factors, a central DNA-binding domain (DBD) responsible for recognition of its response element, as well as an oligomerization domain (OD) involved in the formation of the quaternary structure (Murray-Zmijewski, Lane, & Bourdon, 2006). Both p63 and p73 additionally have a sterile alpha motif (SAM) domain that is involved in protein-protein interactions important for transactivation whereas p53 contains a C-terminal domain (CTD) that undergoes many posttranslational modifications important to transactivation (Murray-Zmijewski et al., 2006). Also unique to TP63 and TP73 genes are the two transcription promoter regions, P1 and P2, located upstream of exon 1 and within intron 3, respectively, giving rise to splicing variants that have a N-terminal transactivation (TA) domain similar to p53, called TAp63 and TAp73, as well as splicing isoforms without the TA domain designated as Δ Np63 and Δ Np73 (Kaghad et al., 1997; Yang et al., 1998).

The TP73 gene comprises 15 exons of which due to alternative splicing can make 29 theoretical isoforms, although only 14 of them have been described (Murray-Zmijewski et al., 2006). All possible isoforms come about from at least seven alternatively spliced C-terminal isoforms (Melino, Lu, Gasco, Crook, & Knight, 2003; Moll & Slade, 2004) and at least four alternatively spliced N-terminal isoforms (Stiewe,

Zimmermann, Frilling, Esche, & Putzer, 2002). These two groups that largely comprise the TAp73 isoforms and the Δ Np73 isoforms have been identified to carry out distinct transcriptionally relevant functions. The TAp73 isoform has a close structural similarity to TAp53 and has been identified to transcriptionally activate p53 target genes, consequently being viewed predominantly as a tumor suppressor (C. W. Lee & La Thangue, 1999; Strano et al., 2005; Tomasini et al., 2008). In contrast, Δ Np73 lacks the TA domain associated with much of the transcription regulation activity in many p53-dependent regulations but also p53-independent transcription. Instead, Δ Np73 has been characterized largely to have a repressive role many times interfering with p53 and TAp73 binding to respective response elements (Grob et al., 2001; Ishimoto et al., 2002; Nakagawa et al., 2003; Nakagawa et al., 2002). Consequently, Δ Np73 has been largely attributed to have oncogenic functions.

D. p73 Structure

The response element recognition of p73 is done by the DBD, which has a structure common across the p53-family. The following account of the DBD structure is largely informed by the first published papers of the p73 tetramers recognizing a full-site response element, (Ethayathulla et al., 2012) and (Ethayathulla, Nguyen, & Viadiu, 2013). It is comprised of a twisted β -sandwich made up of two anti-parallel β -sheets. The loops and turns that emerge from the β -strand core of the DBD, help to make up, in one end of the β -sandwich, the dimerization surface as well as the DNA recognition sites of the monomer while the set of loops and turns at the other end of the β -sandwich, largely

serve to provide the connectivity between β -strands facing the solvent surface. The former set of loops combine with other secondary structure to form the larger assembly of the p73 structure bound to DNA. In particular, helices H1 and 3₁₀, which reside in loops L2 and L3, respectively, form the dimer interface. In addition, a loop-sheet-helix motif forms the majority of the DNA recognition element which is comprised by loop L3 that binds by nonspecific and specific residues to the DNA minor groove, followed by two anti-parallel β -sheets that bring back the protein to the DNA recognizing face and allows residues prior to the formation of helix H2 as well as ones constituting the helix H2 to recognize the DNA. The loosely structured region made up of loops 2 and 3 is maintained by a tetrahedral zinc coordination-site, which is made up of residues Cys-194 and His-197 from helix H1 in loop L2 and Cys-258-and Cys-262 from helix 3₁₀ in loop L3.

The full-site recognition of p73 to the response element is made, as in the case of p53 and p63, by a dimer of dimers conformation, such that one dimer recognizes one half-site of the DNA by individual monomer recognition of the response element quarter-site. The dimerization interface is formed by van der Waal interactions and hydrogen bonds from the mentioned loops L2 and L3 of each monomer that sit in the middle of the half-site response element. Two equally structured, dimer-DNA half-sites make the full-site tetramer complex, largely maintained by van der Waal interactions between the adjacent dimers that form the tetramerization interface.

High-resolution structures of any of the full-length unmodified p53-family members or described isoform in tetramer form have yet to be determined. As a result, knowledge of the exact domain assembly of the p53-family of proteins and its context is largely unknown. Some low-resolution electron microscopy reconstructions of full-length

p53 have been solved, but show distinct domain arrangements particularly of the OD with respect to the DBD (Melero et al., 2011; Okorokov et al., 2006; Pham, Lucumi, Cheung, & Viadiu, 2012; Tidow et al., 2007). This diversity in full-length p53 arrangement might suggest a great flexibility of the tetramer depending on the context to adopt different quaternary arrangements. A crystal structure of modified-p53 as a tetramer bound to DNA was solved and included the DBD and the OD, but without the NLS-linker region that connects the DBD and OD regions. This structure showed OD connectivity between monomers adjacent to one another, but from different dimers (Emamzadah, Tropia, & Halazonetis, 2011; Petty et al., 2011). Extrapolating any of these possible conformations to the other p53-family members is difficult because of the different isoforms encountered in p73 and p63 and, in addition, is further complicated by the distinct CTD encountered in p73 and p63 compared to their p53 analog.

E. The Recognition Element

The response element sequence that binds p53-family members is characterized by two consensus half-sites (5'-Pu1-Pu2-Pu3-C4-W5-W5'-G4'-Py3'-Py2'-Py1'-3'; where Pu = purine, C = cytosine, W = adenine or thymine, and Py = pyrimidine). Functional half-site response elements, can be separated by a spacer of 0 to 3 bps in length (Riley et al., 2008). Interestingly, the ratio of response elements that have no spacer to ones with spacer is roughly 50:50. In the case of p73, the tetramer complex bound to a full-site response element is maintained, likely, up to a 2 bps spacer (Ethayathulla et al., 2012). This is due to the loss of contact that form the tetramerization

interface between adjacent dimers as a result of the half-site separation and the added DNA rotation after every added spacer. This suggests that p53-family member structural organization, in the case of response elements with medium to long spacers, form dimers rather than tetramers, although their role is still uncertain.

The response element half-site can be thought of as having a very loose palindromic-like organization highlighted by three components. First, the most critical element is, the position-4 cytosine that is almost always present twice in each half-site in a palindromic arrangement across the middle of each half-site. Second, the fifth position is represented by an adenine or thymine, although a palindromic adenine arrangement across quarter-sites is very common. Lastly, the flanking regions represented by positions 1 – 3 are the most variable, but most often contain purines. Remarkably, the vast majority of the native response elements that have been validated have mismatches from the consensus sequence, notably 95% of them (Menendez, Inga, & Resnick, 2009). This percentage speaks for the fact that there is likely an element of redundancy among the sequences. Nonetheless, the response element is constituted essentially from the mentioned three key parts that make up the necessary and sufficient conditions for appropriate transcription control. Of greatest priority, the highly conserved fourth position is critical for the binding of the response element by p73. This is due to the critical recognition of the guanine, complimentary to the fourth position cytosine, which is contacted by way of Arg-300 in p73 (analogously Arg-280 and Arg-311 in p53 and p63, respectively) (C. Chen, Gorlatova, Kelman, & Herzberg, 2011; Ethayathulla et al., 2012; Kitayner et al., 2006). Therefore, a complete half-site would be able to properly recognize a p53-family member dimer by way of both guanines with help from the

nonspecific DNA backbone contacts. The fifth position also plays an important role in recognition by p53-family of proteins through Arg-268 (Arg-248 and Arg-279 in p53 and p63, respectively), in addition to adopting a favorable DNA conformation at the center of the half-site believed to help proper complex formation (Ethayathulla et al., 2013; Kitayner et al., 2006). In the case of p53, the presence of an AT sequence at the center of the half-site seems to favor a Hoogsteen base pairing, allowing a more favorable DNA conformation (Kitayner et al., 2010). Finally, the flanking sequence is recognized by several residues, however their contacts are highly dependent on the sequence present. Most importantly, Lys-138 (Lys-120 and Lys-149 in p53 and p63, respectively) forms an important contact with the flanking-site, most notably when in presence of a guanine-rich sequence (Ethayathulla et al., 2013; Kitayner et al., 2006).

F. Motivation of Study

Response element sequences of the p53-family show distinct sequence profiles depending on the corresponding gene function. Interestingly, the comparison between p73 response elements for genes of cell cycle arrest and DNA repair show an increased frequency of adenines compared to apoptotic response elements that show a higher presence of guanines in position 1 – 3 (Ethayathulla et al., 2013). In support of this correlation, the author of this work conducted a non-exhaustive analysis of experimentally verified response element sequences for p53 taken from a compiled table in (Wang, Xiao, & Ren, 2009), showing a significant increased presence of guanines in the flanking sequence of apoptotic response elements compared to a higher presence of

adenines in cell cycle arrest and DNA repair response elements (shown in chapter2, section F). Structures of p53 and p73 (Ethayathulla et al., 2013; Kitayner et al., 2006) have highlighted the importance of Lys-120 in p53, and analogously Lys-138 in p73, to recognize the flanking quarter-site sequence. In the case of p73, structural studies of p73-DBD tetramer bound to guanine-rich and adenine-rich response elements show a distinct conformational change of loop L1 (Ethayathulla et al., 2013; Ethayathulla et al., 2012). Specifically, loop L1 conformation is extended further into the major groove enabled by Lys-138 binding to position-2 and -3 when in the presence of a guanine-rich sequences. In contrast, loop L1 recesses away from the major groove due to Lys-138's lack of contact with the response element in an adenine-rich environment.

Functional studies, although not exhaustive, have shown the flanking sequences of response element to play an important role in modulating the transcriptional behavior of p53 (Wang et al., 2009). Interestingly, separate studies of p53 have shown that acetylated p53 at Lys-120 is necessary for apoptotic gene transcription, specifically in *puma* and *bax* (response elements with multiple guanine-rich flanking sites) to occur (Sykes et al., 2006; Tang, Luo, Zhang, & Gu, 2006). Despite knowing some aspects of the flanking-site including Lys-120/138's importance in recognition of the flanking sequence as seen in p53 and p73, the distinct flanking sequence profiles in relation to apoptotic and nonapoptotic sequences, as well as its functional role in directing transactivation of apoptotic genes, little is known of the mechanism for how flanking site recognition can cause conformational change in the protein and structurally how it may lead to modulating transcription control.

The work presented here tries to understand the p73 recognition of the response element flanking-site believed to be important in directing transcription regulation. In doing so, a reductionist approach is taken to try to understand response element recognition using the basic biochemically and likely biologically-relevant unit of p73, the p73DBD-dimer/half-site complex. Specifically, using molecular dynamics (MD), this study examines complexes with two distinct response element sequences, guanine-rich (apoptotically relevant) and adenine-rich (Cell Cycle Arrest/DNA Repair relevant) flanking sequences. In addition, the transition going from the guanine-rich to the adenine-rich complex, and vice versa, will be examined to try to understand the mechanism of the conformational change. The subsequent chapters describe in detail the system set-up and the running of the MD calculations (chapter 2), the conformational analysis of the simulated systems as well as the comparative differences with respect to the starting crystal structures (chapters 3 and 4), and finally potential consequences from the results obtained as well as a model for response element recognition that takes into account the relevancy of the flanking sequence (chapter 5).

REFERENCES

- Albright, S. R., & Tjian, R. (2000). TAFs revisited: more data reveal new twists and confirm old ideas. *Gene*, 242(1-2), 1-13.
- Budhram-Mahadeo, V., Morris, P. J., Smith, M. D., Midgley, C. A., Boxer, L. M., & Latchman, D. S. (1999). p53 suppresses the activation of the Bcl-2 promoter by the Brn-3a POU family transcription factor. *J Biol Chem*, 274(21), 15237-15244.
- Burley, S. K., & Roeder, R. G. (1996). Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem*, 65, 769-799. doi: 10.1146/annurev.bi.65.070196.004005
- Butler, J. E., & Kadonaga, J. T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev*, 16(20), 2583-2592. doi: 10.1101/gad.1026202
- Chen, C., Gorlatova, N., Kelman, Z., & Herzberg, O. (2011). Structures of p63 DNA binding domain in complexes with half-site and with spacer-containing full response elements. *Proc Natl Acad Sci U S A*, 108(16), 6456-6461. doi: 10.1073/pnas.1013657108
- Chen, X., Farmer, G., Zhu, H., Prywes, R., & Prives, C. (1993). Cooperative DNA binding of p53 with TFIID (TBP): a possible mechanism for transcriptional activation. *Genes Dev*, 7(10), 1837-1849.
- Crum, C. P., & McKeon, F. D. (2010). p63 in epithelial survival, germ cell surveillance, and neoplasia. *Annu Rev Pathol*, 5, 349-371. doi: 10.1146/annurev-pathol-121808-102117
- Deng, W., & Roberts, S. G. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev*, 19(20), 2418-2423. doi: 10.1101/gad.342405
- Deng, W., & Roberts, S. G. (2007). TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma*, 116(5), 417-429. doi: 10.1007/s00412-007-0113-9
- Dotsch, V., Bernassola, F., Coutandin, D., Candi, E., & Melino, G. (2010). p63 and p73, the ancestors of p53. *Cold Spring Harb Perspect Biol*, 2(9), a004887. doi: 10.1101/cshperspect.a004887
- Emamzadah, S., Tropia, L., & Halazonetis, T. D. (2011). Crystal structure of a multidomain human p53 tetramer bound to the natural CDKN1A (p21) p53-

- response element. *Mol Cancer Res*, 9(11), 1493-1499. doi: 10.1158/1541-7786.MCR-11-0351
- Ethayathulla, A. S., Nguyen, H. T., & Viadiu, H. (2013). 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*, 288(7), 4744-4754. doi: 10.1074/jbc.M112.408039
- Ethayathulla, A. S., Tse, P. W., Monti, P., Nguyen, S., Inga, A., Fronza, G., & Viadiu, H. (2012). Structure of p73 DNA-binding domain tetramer modulates p73 transactivation. *Proc Natl Acad Sci U S A*, 109(16), 6066-6071. doi: 10.1073/pnas.1115463109
- Farmer, G., Colgan, J., Nakatani, Y., Manley, J. L., & Prives, C. (1996). Functional interaction between p53, the TATA-binding protein (TBP), and TBP-associated factors in vivo. *Mol Cell Biol*, 16(8), 4295-4304.
- Grob, T. J., Novak, U., Maisse, C., Barcaroli, D., Luthi, A. U., Pirnia, F., . . . Tobler, A. (2001). Human delta Np73 regulates a dominant negative feedback loop for TAp73 and p53. *Cell Death Differ*, 8(12), 1213-1223. doi: 10.1038/sj.cdd.4400962
- Gu, W., & Roeder, R. G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*, 90(4), 595-606.
- Gu, W., Shi, X. L., & Roeder, R. G. (1997). Synergistic activation of transcription by CBP and p53. *Nature*, 387(6635), 819-823. doi: 10.1038/42972
- Innocente, S. A., & Lee, J. M. (2005a). p53 is a NF-Y- and p21-independent, Sp1-dependent repressor of cyclin B1 transcription. *FEBS Lett*, 579(5), 1001-1007. doi: 10.1016/j.febslet.2004.12.073
- Innocente, S. A., & Lee, J. M. (2005b). p73 is a p53-independent, Sp1-dependent repressor of cyclin B1 transcription. *Biochem Biophys Res Commun*, 329(2), 713-718. doi: 10.1016/j.bbrc.2005.02.028
- Ishimoto, O., Kawahara, C., Enjo, K., Obinata, M., Nukiwa, T., & Ikawa, S. (2002). Possible oncogenic potential of DeltaNp73: a newly identified isoform of human p73. *Cancer Res*, 62(3), 636-641.
- Kaghad, M., Bonnet, H., Yang, A., Creancier, L., Biscan, J. C., Valent, A., . . . Caput, D. (1997). Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell*, 90(4), 809-819.

- Kanaya, T., Kyo, S., Hamada, K., Takakura, M., Kitagawa, Y., Harada, H., & Inoue, M. (2000). Adenoviral expression of p53 represses telomerase activity through down-regulation of human telomerase reverse transcriptase transcription. *Clin Cancer Res*, 6(4), 1239-1247.
- Kitayner, M., Rozenberg, H., Kessler, N., Rabinovich, D., Shaulov, L., Haran, T. E., & Shakked, Z. (2006). Structural basis of DNA recognition by p53 tetramers. *Mol Cell*, 22(6), 741-753. doi: 10.1016/j.molcel.2006.05.015
- Kitayner, M., Rozenberg, H., Rohs, R., Suad, O., Rabinovich, D., Honig, B., & Shakked, Z. (2010). Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat Struct Mol Biol*, 17(4), 423-429. doi: 10.1038/nsmb.1800
- Ko, L. J., & Prives, C. (1996). p53: puzzle and paradigm. *Genes Dev*, 10(9), 1054-1072.
- Lane, D., & Levine, A. (2010). p53 Research: the past thirty years and the next thirty years. *Cold Spring Harb Perspect Biol*, 2(12), a000893. doi: 10.1101/cshperspect.a000893
- Lee, C. W., & La Thangue, N. B. (1999). Promoter specificity and stability control of the p53-related protein p73. *Oncogene*, 18(29), 4171-4181. doi: 10.1038/sj.onc.1202793
- Lee, K. C., Crowe, A. J., & Barton, M. C. (1999). p53-mediated repression of alpha-fetoprotein gene expression by specific DNA binding. *Mol Cell Biol*, 19(2), 1279-1288.
- Lin, Y. L., Sengupta, S., Gurdziel, K., Bell, G. W., Jacks, T., & Flores, E. R. (2009). p63 and p73 transcriptionally regulate genes involved in DNA repair. *PLoS Genet*, 5(10), e1000680. doi: 10.1371/journal.pgen.1000680
- Liu, X., Miller, C. W., Koeffler, P. H., & Berk, A. J. (1993). The p53 activation domain binds the TATA box-binding polypeptide in Holo-TFIID, and a neighboring p53 domain inhibits transcription. *Mol Cell Biol*, 13(6), 3291-3300.
- Melero, R., Rajagopalan, S., Lazaro, M., Joerger, A. C., Brandt, T., Veprintsev, D. B., . . . Valle, M. (2011). Electron microscopy studies on the quaternary structure of p53 reveal different binding modes for p53 tetramers in complex with DNA. *Proc Natl Acad Sci U S A*, 108(2), 557-562. doi: 10.1073/pnas.1015520107
- Melino, G., De Laurenzi, V., & Vousden, K. H. (2002). p73: Friend or foe in tumorigenesis. *Nat Rev Cancer*, 2(8), 605-615. doi: 10.1038/nrc861

- Melino, G., Lu, X., Gasco, M., Crook, T., & Knight, R. A. (2003). Functional regulation of p73 and p63: development and cancer. *Trends Biochem Sci*, 28(12), 663-670. doi: 10.1016/j.tibs.2003.10.004
- Menendez, D., Inga, A., & Resnick, M. A. (2009). The expanding universe of p53 targets. *Nat Rev Cancer*, 9(10), 724-737. doi: 10.1038/nrc2730
- Mills, A. A., Zheng, B., Wang, X. J., Vogel, H., Roop, D. R., & Bradley, A. (1999). p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature*, 398(6729), 708-713. doi: 10.1038/19531
- Moll, U. M., & Slade, N. (2004). p63 and p73: roles in development and tumor formation. *Mol Cancer Res*, 2(7), 371-386.
- Murphy, M., Ahn, J., Walker, K. K., Hoffman, W. H., Evans, R. M., Levine, A. J., & George, D. L. (1999). Transcriptional repression by wild-type p53 utilizes histone deacetylases, mediated by interaction with mSin3a. *Genes Dev*, 13(19), 2490-2501.
- Murray-Zmijewski, F., Lane, D. P., & Bourdon, J. C. (2006). p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. *Cell Death Differ*, 13(6), 962-972. doi: 10.1038/sj.cdd.4401914
- Murray-Zmijewski, F., Slee, E. A., & Lu, X. (2008). A complex barcode underlies the heterogeneous response of p53 to stress. *Nat Rev Mol Cell Biol*, 9(9), 702-712. doi: 10.1038/nrm2451
- Nakagawa, T., Takahashi, M., Ozaki, T., Watanabe, K., Hayashi, S., Hosoda, M., . . . Nakagawara, A. (2003). Negative autoregulation of p73 and p53 by DeltaNp73 in regulating differentiation and survival of human neuroblastoma cells. *Cancer Lett*, 197(1-2), 105-109.
- Nakagawa, T., Takahashi, M., Ozaki, T., Watanabe Ki, K., Todo, S., Mizuguchi, H., . . . Nakagawara, A. (2002). Autoinhibitory regulation of p73 by Delta Np73 to modulate cell survival and death through a p73-specific target element within the Delta Np73 promoter. *Mol Cell Biol*, 22(8), 2575-2585.
- Ohlsson, C., Kley, N., Werner, H., & LeRoith, D. (1998). p53 regulates insulin-like growth factor-I (IGF-I) receptor expression and IGF-I-induced tyrosine phosphorylation in an osteosarcoma cell line: interaction between p53 and Sp1. *Endocrinology*, 139(3), 1101-1107. doi: 10.1210/endo.139.3.5832
- Okorokov, A. L., Sherman, M. B., Plisson, C., Grinkevich, V., Sigmundsson, K., Selivanova, G., . . . Orlova, E. V. (2006). The structure of p53 tumour suppressor

- protein reveals the basis for its functional plasticity. *EMBO J*, 25(21), 5191-5200. doi: 10.1038/sj.emboj.7601382
- Ori, A., Zauberman, A., Doitsh, G., Paran, N., Oren, M., & Shaul, Y. (1998). p53 binds and represses the HBV enhancer: an adjacent enhancer element can reverse the transcription effect of p53. *EMBO J*, 17(2), 544-553. doi: 10.1093/emboj/17.2.544
- Orphanides, G., Lagrange, T., & Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev*, 10(21), 2657-2683.
- Petty, T. J., Emamzadah, S., Costantino, L., Petkova, I., Stavridi, E. S., Saven, J. G., . . . Halazonetis, T. D. (2011). An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity. *EMBO J*, 30(11), 2167-2176. doi: 10.1038/emboj.2011.127
- Pham, N., Lucumi, A., Cheung, N., & Viadiu, H. (2012). The tetramer of p53 in the absence of DNA forms a relaxed quaternary state. *Biochemistry*, 51(41), 8053-8055. doi: 10.1021/bi301193k
- Pozniak, C. D., Radinovic, S., Yang, A., McKeon, F., Kaplan, D. R., & Miller, F. D. (2000). An anti-apoptotic role for the p53 family member, p73, during developmental neuron death. *Science*, 289(5477), 304-306.
- Riley, T., Sontag, E., Chen, P., & Levine, A. (2008). Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol*, 9(5), 402-412. doi: 10.1038/nrm2395
- Seto, E., Usheva, A., Zambetti, G. P., Momand, J., Horikoshi, N., Weinmann, R., . . . Shenk, T. (1992). Wild-type p53 binds to the TATA-binding protein and represses transcription. *Proc Natl Acad Sci U S A*, 89(24), 12028-12032.
- Smale, S. T. (2001). Core promoters: active contributors to combinatorial gene regulation. *Genes Dev*, 15(19), 2503-2508. doi: 10.1101/gad.937701
- Smeenk, L., van Heeringen, S. J., Koeppel, M., van Driel, M. A., Bartels, S. J., Akkers, R. C., . . . Lohrum, M. (2008). Characterization of genome-wide p53-binding sites upon stress response. *Nucleic Acids Res*, 36(11), 3639-3654. doi: 10.1093/nar/gkn232
- Stenger, J. E., Tegtmeyer, P., Mayr, G. A., Reed, M., Wang, Y., Wang, P., . . . Mastrangelo, I. A. (1994). p53 oligomerization and DNA looping are linked with transcriptional activation. *EMBO J*, 13(24), 6011-6020.
- Stiewe, T., Zimmermann, S., Frilling, A., Esche, H., & Putzer, B. M. (2002). Transactivation-deficient DeltaTA-p73 acts as an oncogene. *Cancer Res*, 62(13), 3598-3602.

- Strano, S., Monti, O., Pediconi, N., Baccarini, A., Fontemaggi, G., Lapi, E., . . . Blandino, G. (2005). The transcriptional coactivator Yes-associated protein drives p73 gene-target specificity in response to DNA Damage. *Mol Cell*, 18(4), 447-459. doi: 10.1016/j.molcel.2005.04.008
- Sykes, S. M., Mellert, H. S., Holbert, M. A., Li, K., Marmorstein, R., Lane, W. S., & McMahon, S. B. (2006). Acetylation of the p53 DNA-binding domain regulates apoptosis induction. *Mol Cell*, 24(6), 841-851. doi: 10.1016/j.molcel.2006.11.026
- Tang, Y., Luo, J., Zhang, W., & Gu, W. (2006). Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. *Mol Cell*, 24(6), 827-839. doi: 10.1016/j.molcel.2006.11.021
- Thut, C. J., Chen, J. L., Klemm, R., & Tjian, R. (1995). p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. *Science*, 267(5194), 100-104.
- Tidow, H., Melero, R., Mylonas, E., Freund, S. M., Grossmann, J. G., Carazo, J. M., . . . Fersht, A. R. (2007). Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc Natl Acad Sci U S A*, 104(30), 12324-12329. doi: 10.1073/pnas.0705069104
- Tomasini, R., Tsuchihara, K., Wilhelm, M., Fujitani, M., Rufini, A., Cheung, C. C., . . . Mak, T. W. (2008). TAp73 knockout shows genomic instability with infertility and tumor suppressor functions. *Genes Dev*, 22(19), 2677-2691. doi: 10.1101/gad.1695308
- Viadiu, H., Fronza, G., & Inga, A. (2014). Structural Studies on Mechanisms to Activate Mutant p53. *Subcell Biochem*, 85, 119-132. doi: 10.1007/978-94-017-9211-0_7
- Vousden, K. H., & Prives, C. (2009). Blinded by the Light: The Growing Complexity of p53. *Cell*, 137(3), 413-431. doi: 10.1016/j.cell.2009.04.037
- Wang, B., Xiao, Z., & Ren, E. C. (2009). Redefining the p53 response element. *Proc Natl Acad Sci U S A*, 106(34), 14373-14378. doi: 10.1073/pnas.0903284106
- Woychik, N. A., & Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. *Cell*, 108(4), 453-463.
- Xing, J., Sheppard, H. M., Corneillie, S. I., & Liu, X. (2001). p53 Stimulates TFIID-TFIIA-promoter complex assembly, and p53-T antigen complex inhibits TATA binding protein-TATA interaction. *Mol Cell Biol*, 21(11), 3652-3661. doi: 10.1128/MCB.21.11.3652-3661.2001

- Yang, A., Kaghad, M., Wang, Y., Gillett, E., Fleming, M. D., Dotsch, V., . . . McKeon, F. (1998). p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol Cell*, 2(3), 305-316.
- Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R. T., . . . McKeon, F. (1999). p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature*, 398(6729), 714-718. doi: 10.1038/19539
- Yoon, M. K., Ha, J. H., Lee, M. S., & Chi, S. W. (2014). Structure and Apoptotic Function of p73. *BMB Rep.*

CHAPTER 2

METHODS

A. Molecular Dynamics (MD) Simulation

Molecular dynamics simulation is a computational technique that utilizes Newtonian laws of motion to simulate a molecular system of interest. In order to model molecular motions and obtain the position of atoms at discrete moments in time, numerical integration techniques are employed, specifically the Verlet algorithm. The Verlet algorithm is a Taylor series expansion of the position functions, $r_i(t + h)$ and $r_i(t - h)$, where the subscript i refers to, in our case, a particular atom within an N-body system. A Taylor series is a representation of a function as an infinite series many times used, as employed here, as a way to approximate a function by using just a few terms. This, in computational terms, helps to limit the computational cost of solving the complete equation. The functions $r_i(t + h)$ and $r_i(t - h)$ can therefore be interpreted as approximating the position at two moments in time. The Taylor series expansion of the two functions give the following relations,

$$r_i(t + h) \approx r_i(t) + \dot{r}_i(t)h + \frac{\ddot{r}_i(t)h^2}{2}$$

$$r_i(t - h) \approx r_i(t) - \dot{r}_i(t)h + \frac{\ddot{r}_i(t)h^2}{2}.$$

In order to relate the two position functions, both are added together. The summation of the above relations gives the following equation,

$$r_i(t + h) + r_i(t - h) \approx 2r_i(t) + \ddot{r}_i(t)h^2.$$

In our system, like many other physical systems, the energy potential is approximated to be some known relation (see below). Since one can relate the potential energy to the force acting on an object (force is the negative of the position derivative of energy in one-dimensional terms), one can indirectly relate the potential to the position of the atom in question by introducing Newton's equation of motion, $F_i = m_i \ddot{r}_i$, into the relation. The Verlet algorithm is represented as the following,

$$r_i(t + h) \approx 2r_i(t) - r_i(t - h) + \frac{F_i}{m_i}h^2.$$

As discussed above, one can relate the force to the potential energy and specifically in three dimensions, the force acting on any given atom is the gradient of the potential energy, $F_i = -\nabla_i V(r^N)$. The potential energy of a molecular system is given by the molecular force field, which models intra- and inter-molecular interactions with a relatively simple four-component expression,

$$\begin{aligned} V(r^N) = & \sum_{bonds} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 + \sum_{dihedrals} \frac{V_n}{2} (1 + \cos[n\omega - \gamma]) \\ & + \sum_{i=1}^N \sum_{\substack{j=i+1 \\ nonbonded}}^N \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right). \end{aligned}$$

The potential energy is a function of the position of N atoms. The first two terms use harmonic potentials to model bond and angle energies. First, to model each bond's energy, the harmonic potential provides a reasonable approximation to the Morse potential (models the potential energy of a diatomic molecule) in the area close to the equilibrium bond position. Although vibrational effects of molecules at ground state are not taken into effect, the model is sufficiently accurate for the systems that one studied. The bonding term is characterized by a reference bond length, $l_{i,0}$, which represents the length when all other force field terms are zero, an actual bond length, l_i , and a force constant, k_i , which like $l_{i,0}$, is determined by factors such as bond hybridization, bond order, and atom types. In the second term, the angle energy refers to the angle formed by three adjacent atoms with two bonds that forms each valence angle. The energy is identically modeled as the bonding energy and its parameters can be, likewise, analogously interpreted.

The third term in the potential energy equation represents the energy contribution for each dihedral angle, ω . The dihedral energy is characterized by V_n , which largely represents the height of the energetic barrier, n , the number of energy minima through a 360° rotation, and the phase, γ , indicating the angle at which the energy minimum is located. Several force fields can also include improper dihedrals and cross terms.

The last term comprises the non-bonded interactions, which are composed of van der Waals treated with a Lennard-Jones potential, which models the potential energy representing the interaction between a pair of neutral atoms, and electrostatic interactions which are described by a Coulombic potential (models the potential energy between

charged particles, taking into account the magnitude of the charge as well as the distance between the two atoms).

B. System Preparation

The goal here is to modify the PDB structures so that they are ready for simulation. Apart from the original GGGCA and GAACA systems (minimally modified protein and DNA structures), two swapped structures were created where GGGCA and GAACA structures will swap with DNA structure of the other. Additionally, protonation states were assigned to residues to make sure their configuration is appropriate. Finally, the conditions that it would be in if it were in solution are “mimicked” by adding solvent as well as the appropriate salt conditions.

The starting coordinates for GGGCA and GAACA systems were obtained from the dimer formed by chains A and B including their corresponding response element half site of PDB id's 3vd1 (Ethayathulla et al., 2012) and 4g82 (Ethayathulla, Nguyen, & Viadiu, 2013), respectively. Table 2.1 summarizes the four systems simulated. To generate the swapped structures, both crystal structures were superimposed and the nucleic structures were swapped. Additionally, two A-T bp were added to each end of the response element for the four systems, making the sequence length 14 bps total. Monomer chains were modified to obtain chains with equal length. This was obtained by aligning all monomers with respect to chain A of structure 3vd1 and adding residues to the n-terminal for the other monomers such that all chains went from residue 111 to 313.

Table 2.1. MD Simulation Systems

System	DNA Sequence	Protein Structure(PDB id)	DNA(PDB id)	Copies
GGGCA_3vd1	5'-atGGGCATGCCCat-3'	3vd1	3vd1	1&2
GGGCA_4g82_swap	5'-atGGGCATGCCCat-3'	4g82	3vd1	1&2
GAACA_4g82	5'-atGAACATGTTCat-3'	4g82	4g82	1&2
GAACA_3vd1_swap	5'-atGAACATGTTCat-3'	3vd1	4g82	1&2

Histidine protonation states were determined by using PROPKA (Bas, Rogers, & Jensen, 2008; Li, Robertson, & Jensen, 2005; Olsson, Sondergaard, Rostkowski, & Jensen, 2011; Sondergaard, Olsson, Rostkowski, & Jensen, 2011). Molprobit (Chen et al., 2010) was used to determine mis-fit during structural characterization and flip histidine, asparagine, and glutamate side chains 180°. The coordinated zinc residues were modeled by the tetrahedrally coordinated cationic dummy method (Pang, 1999). Crystallographic waters were retained and systems were solvated with TIP3P water box (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983). Periodic boundary conditions were used such that the distance between the solute and the edge of the water box was at least 15 Å. All system topology and coordinate files were prepared using the amber ff99SB force field (Hornak et al., 2006) for protein residues and the parmbsc0 force field (Perez et al., 2007) for nucleic acids. System charges were neutralized with chlorine ions. To account for salt crystallization conditions, 100 mM of sodium and chloride ions were additionally added. System sizes for all four systems were around 70,000 atoms.

C. MD Simulation Protocol

Once the systems are setup, the systems can begin to acclimate by two series of steps before simulation. First, minimizations allow for the system reach its energetic minimum. Although starting structures, especially crystal structures, are likely to be very close to the energetic minimum, factors like crystal packing can account for deviations from the minimum. Additionally, modifications made to the crystal structure most likely need time to adjust to their lowest energetic state. Solvating the starting structure by the introduction of the water molecules and salt ions can also cause some steric clashes between atoms, which are alleviated by the minimization step. Subsequently, equilibration serves to gently raise the temperature of the system, most likely, to 310 K. This step avoids unfavorable conformations due to abrupt temperature changes. Lastly, the system is ready for unrestrained simulation.

Minimization was conducted in a series of four discrete relaxation steps for a total of 45,000 time-steps. Initially, hydrogens and the additional 4 A-T bps that were modeled into each system were relaxed for the first 5,000 time-steps. Subsequently, another 5,000 time-steps were used to relax hydrogens, waters, ions and the 4 A-T bps. In addition, 10,000 time-steps were used to relax hydrogens, waters, ions, the 4 A-T bps, and residue side chains. Finally, 25,000 time-steps were used to relax the entire system. Minimization was followed by a process of equilibration for a total of 1 ns of restrained MD simulation at 310 K. A series of four 250,000 time-steps were used each with a decreasing harmonic restrain coefficient on backbone atoms.

Unrestrained MD using NAMD 2.7 (Phillips et al., 2005) was performed for 130 ns with a 1 femto-second time-step on all systems. Temperature was maintained at a constant 310 K using Langevin dynamics with a dampening frequency of 5 ps^{-1} . Pressure

was maintained at 1 atm using the Nose Hoover-Langevin piston (Feller, Zhang, Pastor, & Brooks, 1995; Martyna, Tobias, & Klein, 1994) with decay and period times of 50 and 100 femtoseconds, respectively. The Particle Mesh Ewald method was used for long-range electrostatics with a non-bonding cutoff of 14 Å. Figure 2.1 summarizes the sequence of steps for MD simulation.

Minimization, equilibration, as well as the first 50 ns of simulation were performed at the San Diego Supercomputer Center. The latter part of the unrestrained MD, 50 - 130 ns, was performed at the Texas Advanced Computing Center.

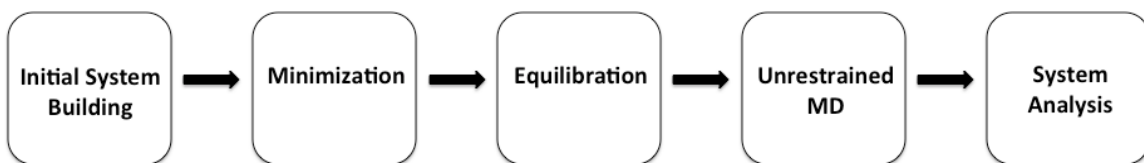


Figure 2.1. Diagram summarizing MD protocol for simulation.

D. Analysis

Analysis of simulation systems was done using both ptraj and cpptraj (Roe & Cheatham, 2013) from AmberTools13 (Case, D.A. et al., 2012) for measurement of distances, angles, and RMSD. GROMOS algorithm (Hess, Kutzner, van der Spoel, & Lindahl, 2008) with GROMACS4.0.5 (Van Der Spoel et al., 2005) analysis software was used for loop L1 analysis of all systems. Starting trajectories were aligned with respect to protein backbone and trajectory was clustered according to loop L1 atoms. Final results shown were done using an RMSD cut-off of 1.36 Å. Visual analysis of trajectories,

individual frames, PDB structures, as well as image rendering, were done using VMD (Humphrey, Dalke, & Schulten, 1996) and Chimera (Pettersen et al., 2004). Hydrogen bonds were characterized using distance and angle calculations, which were done with cpptraj, and post-processed using Linux scripts to filter for instances with cut-off criteria for distance and angle values of <3.5 Å and $\leq 120^\circ$ (angle between electron-donor, hydrogen and, electron-acceptor), respectively.

E. Initial Crystal Structure Dimer Comparison

Starting coordinates for the simulated systems were altered, as previously described in the methods chapter, from two crystallographic structures of p73 DNA-binding domain (DBD) in complex with two half-site response elements. The starting crystal structures of our study were p73 DBD dimer in complex with the half-site 5'-GGGCATGCCC-3' (PDB id: 3vd1) and the half-site 5'-GAACATGTTC-3' (PDB id: 4g82). To simplify our molecular system to the smaller complex of biochemical significance, instead of tetramers, we studied only dimers in complex with half-sites because velocity sedimentation experiments have shown that in solution the p73 DBD dimer is the only molecular specie bound to 10-14 bp oligonucleotides that only contain half-site response elements. An added complication that we avoided was that the original crystal structures had nucleotides spacing the distance between half-sites. Our study is based on the comparison between the p73 DBD dimer bound to two half-sites, labeled as GGGCA and GAACA systems (See Table 2.2).

Although the simulated systems are dimers and therefore contain no spacers, the difference in initial structures due to spacer length is still worth considering. One of the structures in question is referenced in (Ethayathulla et al., 2012), which compares the RMSD of dimers from tetramer structures with spacers of 0-, 1-, 2-, and 4-bps from the following PDB ids: 3vd0, 3vd1, and 3vd2. This comparison largely concluded that the dimer structure is not altered between tetramers with 0 or 1 spacers while one dimer of the 2-bp spacer structures changes greatly as a consequence of a significant translation and rotation that the dimers must overcome in order to maintain its tetramer structure. The structure with a 4-bp spacer seems to resemble more closely the altered 2-bp spacer dimer rather than the 0 and 1-bp spacer structures. In addition, the GAACA structure was compared against the aforementioned structures (AGGCA, GGGCA, and GGACA) by obtaining the RMSD with respect to the tetramers and monomers while aligned with respect to the tetramer and monomer, respectively (Ethayathulla et al., 2013). RMSD values around 2 Å or greater show that there are notable differences across these structures regardless of sequence or spacer length. Due to the significant comparative nature of this study, a further uniform comparison across all these structures was done to more precisely define these differences.

As an alternative method to measure the RMSD, rather than aligning to the entire dimer, structures were aligned to one monomer and the RMSD value of the second monomer in the dimer was recorded. In order to examine exclusively the effects of the spacer, given that there are four distinct response element sequences, Table 2.2 shows the comparison for structures with the same sequence but different spacer. The data shows large differences between 0- and 2-bp spacer dimers as well as 1- and 2-bp spacer dimers

with the exception of regions that are responsible for DNA-binding and forming the dimer interface such as helix H3₁₀, loop L3, and especially helix H1.

A comparison between structures with the same spacer but different sequences was also done as shown in Table 2.3. The sequence comparison with 2 bp spacer AGGCA and GGGCA dimers showed similar structures for only two out of the four shown in Table 2.3. The two distinct conformations are attributed to the strained tetramer formed by the 2-bp spacer.

When comparing AGGCA and GAACA dimers with no spacers, RMSD values of more than 3.5 Å are observed (Table 2.3). Furthermore, comparing the structures of the simulated systems, 1-bp spacer GGGCA dimer and 0-bp spacer GAACA dimer, there are similarly large RMSD values around 3.5 Å. Interestingly, a comparison between the 0-bp AGGCA dimers and the 1-bp spacer GGGCA dimers, which was the structure simulated, found in Table 2.4 shows these two structures to be very similar with RMSD values between 0.5 and 1 Å. These values are comparable to RMSD values of 0- and 1-bp spacer dimers of the same tetramer, found in Table 2.5, which for the purpose of these comparisons can be interpreted as largely noise. As a consequence, a 1-bp spacer does not seem to alter the dimer conformation.

In addition to comparing dimers, it is important to establish whether the monomers of the starting coordinates of the simulation were different. A comparison of the 1bp GGGCA dimers and 0-bp GAACA dimers, shows loop L1 to be significantly higher than the rest of the secondary structures with an average RMSD value 2.036 ± 0.121 Å (Table 2.6). Other regions with moderate RMSD values of around 1 Å are unstructured loop/turns, which largely lie between the antiparallel β strands. When

comparing the 1-bp GGGCA dimers to a 0-bp spacer dimer with a similar sequence, AGGCA, very little difference is observed represented by RMSD values ranging from around 0.2 to 1 Å and significantly, the average loop L1 is 0.771 ± 0.168 Å (Table 2.6). As a reference, the RMSD comparison of different 0-bp AGGCA dimers is also shown in Table 2.6.

F. p53 Response Element Study

In support of the surprising correlation of relatively high presence in guanines in the flanking sites of apoptotic p73 response elements compared to a higher presence of adenine in the same flanking sites of nonapoptotic response elements (Ethayathulla et al., 2013), an analysis using a compiled list of p53 response elements (Wang, Xiao, & Ren, 2009) was done to see if this same observation is seen in verified p53 response elements. The analysis was done using a simple Java program developed by the author which looked at the matching frequency per position across half-sites of p53 response elements in a palindromic fashion. In essence, the algorithm is designed to palindromically check each position to see if the nucleotides match (i.e. positions 1 and 20, positions 2 and 19, and so on...) and if so, then that particular position would increment by one for that response element. Response elements were analyzed according to their associated gene function. The 125 response elements analyzed were divided into 14 gene categories. The categories of most relevance due to the sample size (≥ 11) were apoptosis, cell cycle arrest, and DNA repair are shown in Figure 2.2. The motivation for picking this manner of characterizing the response elements was to see which nucleotides really mattered in

each position. To match nucleotides across half-site would increase the weight of that particular nucleotide according to position as opposed to counting the number of times each nucleotide appears in each position.

Even more affirmatively than the p73 response element analysis, apoptotic response elements showed a higher prevalence of guanine in positions 1 – 3, compared to cell cycle arrest and DNA repair response elements that have higher adenines in those same flanking sites. This provides further evidence of a possible link between apoptotic response elements and the p73 and p53 transcription control. Although not all response elements have a majority of guanines in flanking regions, the prevalence of guanines in these regions may highlight an important evolutionary mechanism relevant to transcription of apoptotic genes.

Table 2.2. RMSD Comparison of 0 bp and 2 bp spacer dimers with AGGCA response element as well as 1 bp and 2 bp spacer dimers with GGGCA response element. RMSD values are broken down by protein region. First row represents the spacer of the dimers compared, second row is the response element sequence (palindromic) of the dimers compared, and the third row shows the dimers identified by the pdbID and the letter of each monomer composing the dimer.

	Spacer 0 & Spacer 2	Spacer 0 & Spacer 2	Spacer 0 & Spacer 2	Spacer 0 & Spacer 2	Spacer 1 & Spacer 2	Spacer 1 & Spacer 2	Spacer 1 & Spacer 2	Spacer 1 & Spacer 2
Sec. Structure Reg.	AGGCA 3vd0_AB & 3vd0_IJ	AGGCA 3vd0_AB & 3vd0_LK	AGGCA 3vd0_DC & 3vd0_IJ	AGGCA 3vd0_DC & 3vd0_LK	GGGCA 3vd1_AB & 3vd1_IJ	GGGCA 3vd1_AB & 3vd1_LK	GGGCA 3vd1_DC & 3vd1_IJ	GGGCA 3vd1_DC & 3vd1_LK
nter	6.2	4.2	6.0	4.4	6.2	4.0	5.8	4.6
S1	5.6	4.2	5.4	4.3	6.9	3.7	6.0	4.0
L1	5.4	3.5	5.7	3.6	6.0	3.6	5.9	3.8
S2	4.9	3.0	5.1	2.9	5.7	2.9	5.2	3.0
bS2S2p	6.1	3.2	6.0	3.2	6.6	2.6	6.0	2.7
S2p	4.1	2.7	4.4	3.1	4.9	2.3	4.8	2.6
bS2pS3	2.8	2.4	3.2	2.8	3.4	2.6	3.6	2.7
S3	5.2	3.7	5.2	4.0	6.2	3.7	5.6	4.2
bS3S4	6.2	6.0	6.0	5.7	7.3	5.6	6.1	6.1
S4	3.4	3.6	3.5	3.5	4.2	3.3	3.9	3.7
L2A	3.5	2.7	3.3	2.9	4.2	2.2	3.9	2.6
H1	1.0	1.3	0.8	1.6	1.1	1.5	0.8	1.5
L2B	2.1	1.8	2.3	2.4	2.1	2.3	2.2	2.7
S5	3.0	3.2	3.2	3.4	3.8	3.4	3.7	3.7
bS5S6	3.0	3.6	3.3	3.8	3.2	4.4	3.6	4.4
S6	1.9	3.3	1.9	3.4	1.8	3.5	2.0	3.7
bS6S7	2.4	3.3	2.2	3.6	3.5	3.3	2.6	4.1
S7	3.0	3.6	3.1	3.8	3.5	3.6	3.4	3.9
bS7S8	6.9	5.4	6.7	5.4	7.7	5.5	7.1	5.9
S8	4.3	3.6	4.4	3.8	4.9	3.7	4.6	4.0
bS8H310	1.9	1.8	2.2	2.1	2.4	1.8	2.8	2.1
H310	1.8	1.5	1.7	1.9	2.4	1.4	2.3	1.7
L3	2.6	1.7	2.4	2.2	3.2	1.3	3.0	1.6
S9	3.9	3.7	4.0	3.8	5.0	3.5	4.7	3.7
bS9S10	3.9	5.1	4.4	5.4	5.2	4.9	4.5	5.7
S10	4.4	3.7	4.3	4.1	5.3	3.4	4.9	4.1
bS10H2	3.0	2.1	3.6	2.6	3.7	2.1	4.0	2.6
H2	5.2	2.6	5.4	2.7	5.8	2.2	5.5	2.2
Average	3.8	3.2	3.9	3.4	4.5	3.2	4.2	3.5

Table 2.3. RMSD Comparison of AGGCA and GAACA 0 bp spacer dimers aswell as AGGCA and GGGCA 2 bp spacer dimers. RMSD values are broken down by protein region. First row represents the spacer of the dimers compared, second row is the response element sequence(palindromic) of the dimers compared, and the third row shows the dimers are identified by the pdbID and the letter of each monomer composing the dimer.

Sec. Structure Reg.	Spacer 0	Spacer 0	Spacer 2	Spacer 2	Spacer 2	Spacer 2
	AGGCA & GAACA	AGGCA & GAACA	AGGCA & GGGCA	AGGCA & GGGCA	AGGCA & GGGCA	AGGCA & GGGCA
	3vd0_AB & 4g82_AB	3vd0_DC & 4g82_AB	3vd0_IJ & 3vd1_IJ	3vd0_IJ & 3vd1_LK	3vd0_LK & 3vd1_IJ	3vd0_LK & 3vd1_LK
nter	6.0	5.8	1.5	9.3	9.9	0.8
S1	4.4	4.1	0.9	8.1	9.5	0.7
L1	3.4	3.3	1.1	7.3	8.1	0.9
S2	2.5	2.6	0.6	6.2	7.2	0.5
bS2S2p	3.7	3.6	0.8	7.2	8.5	1.1
S2p	2.3	2.6	0.7	5.4	6.6	0.8
bS2pS3	2.0	2.2	0.7	4.3	4.7	0.3
S3	4.4	4.3	0.9	7.7	8.7	0.3
bS3S4	6.6	6.6	1.5	10.4	11.2	0.5
S4	4.2	4.4	1.0	6.3	7.4	0.6
L2A	3.5	3.4	0.9	5.5	6.7	0.9
H1	0.8	1.0	0.6	1.5	1.8	0.7
L2B	2.6	2.4	0.8	3.1	3.2	0.8
S5	3.4	3.4	0.9	5.2	5.9	0.4
bS5S6	4.8	4.9	0.8	5.9	6.2	0.6
S6	4.0	4.1	0.7	4.8	5.1	0.6
bS6S7	4.4	4.4	1.3	5.3	6.7	1.5
S7	4.5	4.6	0.8	5.9	6.8	0.6
bS7S8	5.0	4.9	1.1	10.2	11.4	0.5
S8	3.6	3.7	0.8	6.6	7.4	0.3
bS8H310	1.2	1.4	0.7	2.9	3.5	0.3
H310	1.6	1.3	0.7	2.8	3.0	0.3
L3	2.1	2.2	0.7	3.6	4.5	0.9
S9	4.5	4.6	1.2	6.8	8.2	0.5
bS9S10	6.7	7.3	1.8	8.3	10.0	0.8
S10	4.7	4.6	1.1	7.3	8.4	0.5
bS10H2	0.8	1.2	0.6	3.7	4.6	0.6
H2	2.1	1.9	0.6	5.9	6.9	0.8
Average	3.6	3.6	0.9	6.0	6.9	0.7

Table 2.4. RMSD Comparison of 0 and 1 bp spacer dimers regardless of response element sequence. RMSD values are broken down by protein region. First row represents the spacer of the dimers compared, second row is the response element sequence(palindromic) of the dimers compared, and the third row shows the dimers are identified by the pdb id and the letter of each monomer.

	Spacer 0 & Spacer 1	Spacer 0 & Spacer 1	Spacer 0 & Spacer 1	Spacer 0 & Spacer 1	Spacer 0 & Spacer 1	Spacer 0 & Spacer 1
	AGGCA & GGGCA	AGGCA & GGGCA	AGGCA & GGGCA	AGGCA & GGGCA	GAACA & GGGCA	GAACA & GGGCA
Sec. Structure Reg.	3vd0_AB & 3vd1_AB	3vd0_AB & 3vd1_DC	3vd0_DC & 3vd1_AB	3vd0_DC & 3vd1_DC	4g82_AB & 3vd1_AB	4g82_AB & 3vd1_DC
nter	0.7	1.5	0.6	1.2	5.6	5.4
S1	0.5	0.7	1.0	0.4	4.8	3.8
L1	0.6	0.8	0.8	0.8	3.3	3.1
S2	0.4	0.5	0.6	0.6	2.7	2.2
bS2S2p	0.7	1.0	0.8	0.7	3.8	3.0
S2p	0.6	0.7	0.8	0.4	2.5	2.4
bS2pS3	0.4	0.5	0.6	0.4	2.0	2.0
S3	0.5	0.9	0.4	0.6	4.5	3.9
bS3S4	0.7	1.7	1.1	1.2	6.8	5.8
S4	0.4	0.5	0.6	0.6	4.2	4.0
L2A	0.5	0.7	0.8	0.5	3.5	3.3
H1	0.5	0.8	1.0	0.3	0.9	1.0
L2B	0.5	1.2	0.9	0.8	2.4	2.3
S5	0.3	0.7	0.2	0.4	3.3	3.1
bS5S6	0.5	0.9	0.6	0.7	4.6	4.8
S6	0.6	0.6	0.6	0.4	3.6	3.8
bS6S7	0.5	1.5	0.7	1.2	4.1	4.0
S7	0.5	0.7	0.5	0.4	4.4	4.3
bS7S8	0.5	0.9	0.7	0.8	4.9	4.2
S8	0.4	0.7	0.5	0.5	3.5	3.2
bS8H310	0.3	0.6	0.7	0.3	1.1	1.4
H310	0.4	0.7	1.0	0.5	1.9	1.1
L3	0.6	0.9	0.9	0.7	2.4	2.1
S9	0.4	0.5	0.6	0.5	4.6	4.2
bS9S10	0.6	1.4	1.2	1.7	6.7	6.0
S10	0.4	1.1	0.8	0.6	4.6	4.3
bS10H2	0.5	0.8	0.7	0.4	1.0	1.2
H2	0.6	0.9	1.2	0.5	2.5	1.8
Average	0.5	0.9	0.8	0.6	3.6	3.3

Table 2.5. RMSD Comparison between dimers belonging to the same tetramer. RMSD values are broken down by protein region. First row represents the spacer of the dimers compared, second row is the response element sequence (palindromic) of the dimers compared, and the third row shows the dimers are identified by the pdbID and the letter of each monomer composing the dimer.

	Spacer 0	Spacer 2	Spacer 1	Spacer 2	Spacer 4
	AGGCA	AGGCA	GGGCA	GGGCA	GGACA
Sec. Structure Reg.	3vd0_AB & 3vd0_DC	3vd0_IJ & 3vd0_LK	3vd1_AB & 3vd1_DC	3vd1_IJ & 3vd1_LK	3vd2_AB & 3vd2_DC
nter	0.7	9.7	1.3	9.6	0.6
S1	0.6	8.7	1.0	9.0	0.4
L1	0.7	7.7	0.8	7.7	0.7
S2	0.4	6.7	0.6	6.7	0.3
bS2S2p	0.6	8.1	1.0	7.6	0.6
S2p	0.7	6.1	0.7	6.0	0.6
bS2pS3	0.7	4.3	0.3	4.7	0.6
S3	0.5	8.0	0.8	8.4	0.5
bS3S4	1.1	10.6	1.6	11.0	0.6
S4	0.3	6.6	0.6	7.2	0.4
L2A	0.6	6.0	0.9	6.2	0.6
H1	0.7	1.7	1.0	1.7	0.6
L2B	1.0	2.9	1.1	3.4	1.4
S5	0.5	5.3	0.5	5.9	0.3
bS5S6	0.6	5.6	0.8	6.5	0.6
S6	0.5	4.7	0.6	5.3	0.7
bS6S7	0.7	5.6	1.3	6.3	0.7
S7	0.4	6.1	0.6	6.7	0.5
bS7S8	0.4	10.5	0.8	11.2	1.9
S8	0.4	6.7	0.6	7.3	0.6
bS8H310	0.5	3.0	0.8	3.5	0.6
H310	0.7	2.8	1.0	3.1	0.5
L3	0.8	4.1	1.1	4.1	0.5
S9	0.4	7.1	0.7	8.0	0.5
bS9S10	1.3	8.7	1.2	9.6	0.8
S10	0.8	7.6	1.0	8.1	0.5
bS10H2	0.8	4.1	0.7	4.1	0.8
H2	1.0	6.6	1.0	6.2	2.9
Average	0.7	6.3	0.9	6.6	0.7

Table 2.6. RMSD comparison of selected monomers. The number of comparisons for each average RMSD value were 6, 16, and 8, for the following comparisons AGGCA, AGGCA & GGGCA, and GGGCA & GAACA, respectively.

	Spacer 0		Spacer 0 & Spacer 1		Spacer 1 & Spacer 2	
	AGGCA		AGGCA & GGGCA		GGGCA & GAACA	
Sec. Structure Reg	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
nter	0.7	0.1	0.8	0.3	1.1	0.3
S1	0.4	0.1	0.4	0.2	0.6	0.2
L1	0.8	0.3	0.8	0.2	2.0	0.1
S2	0.4	0.1	0.4	0.1	0.5	0.2
bS2S2p	0.8	0.3	0.6	0.2	0.7	0.2
S2p	0.4	0.1	0.4	0.1	0.5	0.2
bS2pS3	0.4	0.1	0.4	0.1	0.5	0.1
S3	0.4	0.1	0.4	0.0	0.5	0.1
bS3S4	1.0	0.2	0.9	0.3	1.2	0.3
S4	0.3	0.1	0.3	0.1	0.3	0.1
L2A	0.5	0.1	0.5	0.1	0.5	0.1
H1	0.5	0.1	0.5	0.1	0.5	0.1
L2B	1.0	0.2	0.9	0.2	0.9	0.2
S5	0.4	0.1	0.3	0.1	0.3	0.1
bS5S6	0.7	0.1	0.6	0.2	0.8	0.2
S6	0.6	0.3	0.5	0.2	0.5	0.1
bS6S7	0.6	0.2	0.6	0.3	0.6	0.3
S7	0.3	0.2	0.3	0.1	0.3	0.1
bS7S8	0.7	0.2	0.7	0.2	1.2	0.6
S8	0.3	0.1	0.4	0.1	0.6	0.2
bS8H310	0.2	0.0	0.3	0.1	0.3	0.1
H310	0.4	0.1	0.4	0.1	0.7	0.3
L3	0.5	0.1	0.5	0.1	0.6	0.1
S9	0.4	0.1	0.4	0.1	0.4	0.1
bS9S10	1.3	0.3	1.0	0.3	0.9	0.2
S10	0.5	0.1	0.5	0.1	0.5	0.1
bS10H2	0.4	0.1	0.4	0.1	0.6	0.1
H2	0.7	0.1	0.6	0.2	0.7	0.2

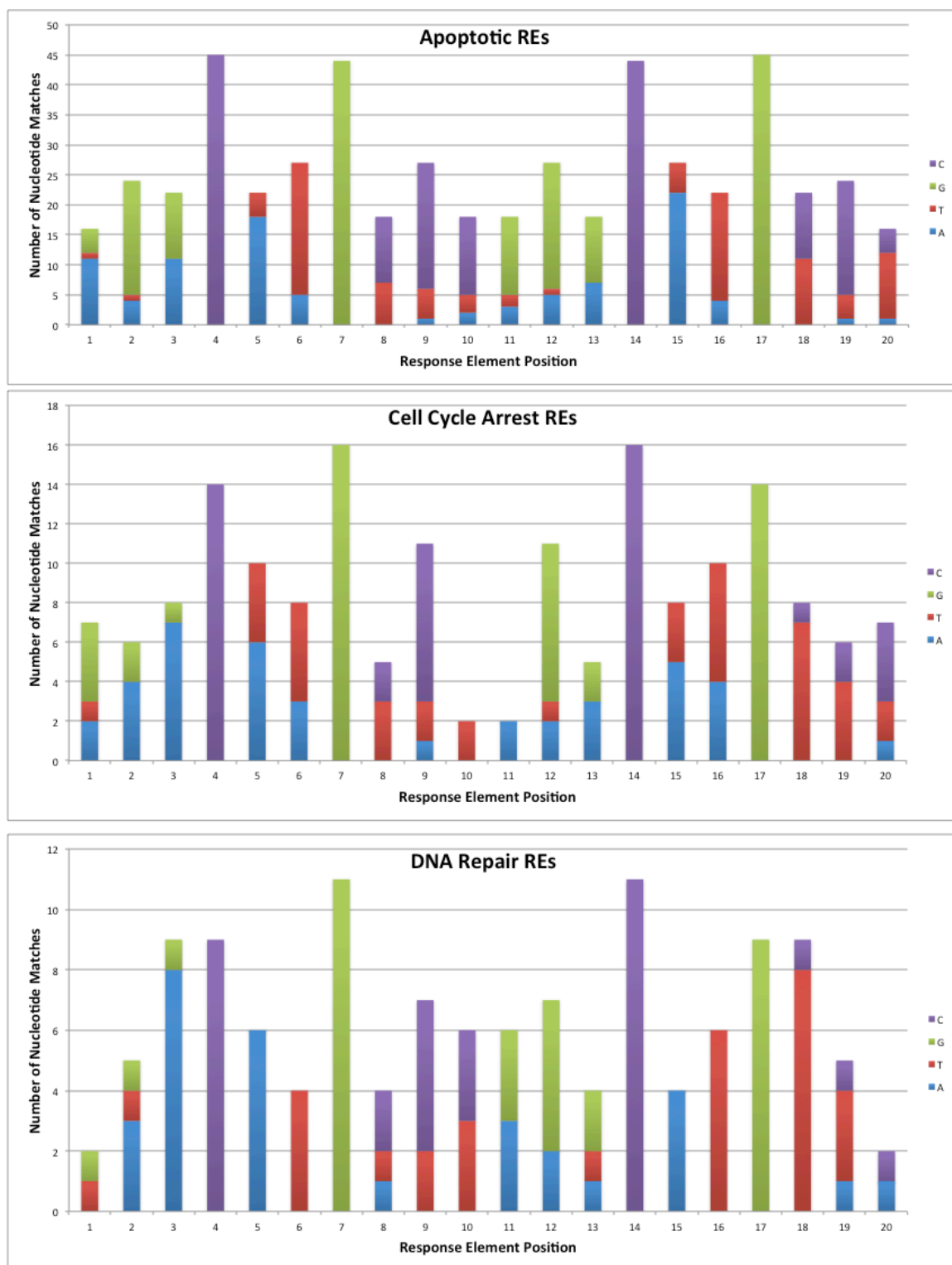


Figure 2.2. p53 Response Element Analysis. Response elements pertaining to three of the most prevalent p53 controlled gene functions are shown above. The above graphs are redundant across half-sites.

REFERENCES

- Bas, D. C., Rogers, D. M., & Jensen, J. H. (2008). Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins*, 73(3), 765-783. doi: 10.1002/prot.22102
- Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., . . . Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 1), 12-21. doi: 10.1107/S0907444909042073
- Ethayathulla, A. S., Nguyen, H. T., & Viadiu, H. (2013). 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*, 288(7), 4744-4754. doi: 10.1074/jbc.M112.408039
- Ethayathulla, A. S., Tse, P. W., Monti, P., Nguyen, S., Inga, A., Fronza, G., & Viadiu, H. (2012). Structure of p73 DNA-binding domain tetramer modulates p73 transactivation. *Proc Natl Acad Sci U S A*, 109(16), 6066-6071. doi: 10.1073/pnas.1115463109
- Feller, S. E., Zhang, Y. H., Pastor, R. W., & Brooks, B. R. (1995). Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *Journal of Chemical Physics*, 103(11), 4613-4621. doi: Doi 10.1063/1.470648
- Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3), 435-447. doi: Doi 10.1021/Ct700301q
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3), 712-725. doi: 10.1002/prot.21123
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling*, 14(1), 33-38. doi: Doi 10.1016/0263-7855(96)00018-5
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics*, 79(2), 926-935.

- Li, H., Robertson, A. D., & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, 61(4), 704-721. doi: 10.1002/prot.20660
- Martyna, G. J., Tobias, D. J., & Klein, M. L. (1994). Constant-Pressure Molecular-Dynamics Algorithms. *Journal of Chemical Physics*, 101(5), 4177-4189. doi: Doi 10.1063/1.467468
- Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M., & Jensen, J. H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. *Journal of Chemical Theory and Computation*, 7(2), 525-537. doi: Doi 10.1021/Ct100578z
- Pang, Y. P. (1999). Novel zinc protein molecular dynamics simulations: Steps toward antiangiogenesis for cancer treatment. *Journal of Molecular Modeling*, 5(10), 196-202.
- Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A., & Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophysical Journal*, 92(11), 3817-3829. doi: Doi 10.1529/Biophysj.106.097782
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605-1612. doi: 10.1002/jcc.20084
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., . . . Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16), 1781-1802. doi: Doi 10.1002/Jcc.20289
- Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, 9(7), 3084-3095. doi: Doi 10.1021/Ct400341p
- Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M., & Jensen, J. H. (2011). Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values. *Journal of Chemical Theory and Computation*, 7(7), 2284-2295. doi: Doi 10.1021/Ct200133y
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, 26(16), 1701-1718. doi: 10.1002/jcc.20291

Wang, B., Xiao, Z., & Ren, E. C. (2009). Redefining the p53 response element. *Proc Natl Acad Sci U S A*, 106(34), 14373-14378. doi: 10.1073/pnas.0903284106

CHAPTER 3

RESULTS: Response Element Contact Analysis

A. Lys-138

This section, like much of this chapter, largely characterizes the behavior of the most important DNA-binding residues. The aim is to describe their behavior in a comparative manner between guanine- and adenine-rich systems as well as describe whether their contacts are similar to those seen in the starting crystal structures. In addition, with respect to the swapped structures, it is important to characterize their behavior in relation to the unmodified response element systems (GGGCA_3vd1 and GAACA_4g82). This section will characterize the most critical residue for the recognition of the flanking-site sequence, Lys-138.

Differences in dimer and monomer conformation between GGGCA and GAACA structures can be largely attributed to Lys-138 in loop L1, which contacts the DNA depending on the nucleotide sequence present (Ethayathulla, Nguyen, & Viadiu, 2013). The GGGCA crystal structure shows that Lys-138 contacts positions 2 and 3 via hydrogen bonds to O6 oxygen of both guanines. Surprisingly, Lys-138 of the GAACA structure makes no side chain contact to DNA. In order to characterize the behavior of Lys-138 in the MD analysis, the hydrogen bond percentage between Lys-138 and DNA is summarized in Figure 3.1. In the GGGCA_3vd1 systems, Figure 3.1 highlights the strong contacts of Lys-138 with guanines from positions 2 and 3. This is the case for monomers A and B of both copies. The strength of the contact is due the favorable electron-donor environment when there is consecutive guanines (Figure 3.2). From Figure 3.3, the distance of Lys-138 to the first three positions with respect to time shows the steady contacts formed to nucleotides 2 and 3 as a consequence of its hydrogen bonds to O6

oxygen and N7 nitrogen from both guanines. For the GGGCA_4g82_swap system, Lys-138 hydrogen bond analysis of the system shows a partial transition whereby in copy 1, Lys-138 of monomer B shows strong contact to nucleotides 1' and 2' while in monomer A there is little contact to phosphates in positions 0 and 1. Similarly in copy 2, Lys-138 of monomer B strongly contacts nucleotides 2' and 3' while, in monomer A, there is little to no contact to phosphates 0 and 1 (Figure 3.1). An analysis of the distance between Lys-138 and positions 1-3 of the GGGCA_4g82_swap system correlates to the hydrogen bond data. Figure 3.3 shows Lys-138 of monomer A contacting phosphates in both copies 1 and 2, while Lys-138 of monomer B shows contacts to nucleotides 1', 2', and 3'. An important distinction between monomer B of copy 1 and 2 is that while copy 1 maintains contact with nucleotides 1' and 2', copy 2 transitions to make contacts exclusively to nucleotides 2 and 3, as the GGGCA_3vd1 system.

In the GAACA_4g82 system, hydrogen bond contacts of Lys-138 to nucleotides are a more diverse set of binding partners where contacts are to several nucleotides and to phosphates of position 0, 1, and 2 (Figure 3.1). Monitoring the distance of Lys-138 to nucleotides, one sees that for monomers B in copy 1 and 2 as well as monomer A in copy 1, Lys-138 goes from contacting phosphates to nucleotides 1 and 2 (Figure 3.4). In the GAACA_3vd1_swap system, hydrogen bond contacts of Lys-138 show a similar diversity of nucleotide and phosphate contacts from positions 0, 1, and 2 (Figure 3.1). The distance of Lys-138 to nucleotides, shown in Figure 3.4, show partial contacts to both nucleotides and phosphates 1 and 2 to varying degrees for monomers A and B of copy 2 and monomer A of copy 1 while in monomer B of copy 1 almost complete contact to nucleotides 1 and 2 occurs.

B. S139, C297, and R268

Similar to Lys-138, Ser-139, Cys-297, and Arg-268 also display behavior that changes based on nucleotide sequence. Ser-139, which resides next to Lys-138, is largely dependent on the relative position of loop L1 to determine the response element point of contact. Analysis of the GGGCA_3vd1 system showed no direct contact by Ser-139 to DNA, however this residue makes a very stable contact water-mediated contact with guanine in position 1 (Figure 3.5). The stability between Ser-139 and the position-1 guanine is seen in the time-lapsed distance monitoring shown in Figure 3.6. GAACA_4g82 and GAACA_3vd1_swap systems showed Ser-139 make nonspecific contacts to position-0 and -1 phosphates. This is confirmed in the distance monitoring of Ser-139 where most of the time is spent between both phosphates either by direct or water-mediated contact (Figure 3.7). In the GGGCA_4g82_swap system, Ser-139 did not contact nucleotide positions 1 and 2 but rather phosphates 0 and 1, except for monomer B of copy 2 GGGCA_4g82_swap, which showed a nucleotide contact to position-0' thymine (Figure 3.6).

Cys-297 displays a distinct behavioral contrast depending on the nucleotide environment. In the GGGCA systems, the SG sulfur of Cys-139 acts as the electron donor to the N4 nitrogen of the cytosine complimentary to position-3 guanine (Figure 3.8). This contact is present in all the copies of the guanine rich systems except for copy1 of GGGCA_4g82_swap. The stability of the contact between Cys-297 to position-3 is increased because Cys-297 forms an additional hydrogen bond with Ser-139, which

contacts guanines in positions 1 and 2. The GAACA systems however, show a less stable binding to the DNA, largely through water-mediated contacts. Not only does Cys-297 show a variety of binding partners, mainly nucleotides in position 1-3, it also seems to have no DNA contact at some instances of the trajectory (Figure 3.8, bottom). To recapitulate, taking into account the three residues discussed so far, the guanine-rich environment allows Lys-138, Ser-139, and Cys-297 to form a very stable and cohesive ensemble of contacts in the major groove of the response element as compared to a more transient and diverse group of interactions in the adenine-rich environment.

Of all the DNA-binding residues, Arg-268 contacts the most diverse set of nucleotides and phosphates. Nonetheless, a subtle but distinct contact pattern is discerned between the guanine- and adenine-rich systems. In the GGGCA_3vd1 system, Arg-268 mainly contacts the position-5 phosphates (Figure 3.9, top) as seen from the distance monitoring of all monomers and copies in Figure 3.10. Additionally, one notices that loop L3 where Arg-268 resides is slightly further away from the minor groove in comparison to the adenine-rich systems which show loop L3 protrude closer toward the minor groove. In the GAACA_3vd1 and GAACA_4g82_swap systems, Arg-268 more readily enter the minor groove and contact nucleotides in positions 4 and 5 (Figures 3.11). Although Arg-268 is able to make direct contacts with position-4 and -5 nucleotides, a large part of the time it does so through a water-mediated contact. The water-mediated contacts allow Arg-268 to interact many times with two nucleotides simultaneously. From the DNA-binding residue analysis, most monomers of the GGGCA_4g82_swap system did not converge from the contacts observed in the adenine-rich systems to the contacts observed in the GGGCA_3vd1 system, except for monomer B of copy 2. Here,

Arg-268 adopts the position-5 phosphate contact while the other monomers of the GGGCA_4g82_swap system mostly prefer the position-5 nucleotide (Figure 3.10). Table 3.1 summarizes the percentage of time that hydrogen bonds of Ser-139, C297, and Arg-300 exist. Most of the contact percentages reported, especially the Arg-268 contacts to nucleotide position 4 and 5 as well as the Ser-139 contacts to guanines 1 and 2, are underestimated (or not seen) since water-mediated contacts do not contribute to the hydrogen bond percentage reported.

C. Additional DNA-Binding Residues

Although Lys-138, Ser-139, Cys-297, and Arg-268 seem to respond to the nucleotide changes in positions 2 and 3, an important set of DNA-binding residues bind to DNA nonspecifically. Of this group, several DNA-binding residues were monitored (Figure 3.12). Several of the contacts shown are maintained for almost all of the simulations irrespective of system. The residues involved are the following: Asn-259, Ser-261, Arg-293, Ala-296, and Cys-297. Additionally, Arg-300 is included in this group since it also retains its contact with the response element independent of systems simulated, even though it binds specifically to the guanine complementary to position-4 cytosine. Cys-297, which was already mentioned as playing an important role in binding to the cytosine complementary to position-3 guanine, is also involved in the nonspecific binding of the response element.

The binding of all of these residues is heavily concentrated on the position-4 phosphate. One set of residues, which includes Ala-296, Cys-297, and Arg-300, combine

to form a hydrogen bond network, which binds to the major groove face of the fourth position phosphate (the O2P oxygen). Arg-300 which binds to the guanine complementary to position-4 cytosine also forms a water-mediated contact with position-4 phosphate and the backbone nitrogen, N, of Cys-297. Ala-296, which sits on the loop preceding helix H2 along with Cys-297, contacts the fourth position phosphate directly (Figure 3.12). The other set of residues, Asn-259, Ser-261, and Arg-293, directly contact fourth and fifth position phosphates via the minor groove face. While Ala-296 contacts the O2P oxygen, residues Asn-259 and Ser-261, which reside in loop L3, contact the O1P oxygen from the fourth position phosphate. Arg-293, which resides on sheet S10 and sits above Asn-259 and Ser-261, contacts the fifth position phosphate (Figure 3.12). The exception to all the described contacts is found in copy 1 of the GGGCA_4g82_swap system where Asn-259 of monomer B now contacts the O2P oxygen of the position-4' phosphate, Arg-293 contacts O1P oxygen of the position-4' phosphate, and Arg-300 of monomers A and B contact nucleotide 2 and position-5 phosphate, respectively. Additionally, residues Ala-296 of monomer A and Asn-259 of monomer A completely lose contact with the response element (Appendix Figures A.1 and A.2). Given such an extensive change in the contact profile of nonspecific binding residues as well as in residues that bind to the first three nucleotides, the swapped system dimer is likely to be undergoing some element of unbinding from the response element. Appendix Figures A.1 – 5 show the detailed distance analysis of residues making direct contact with positions 4 and 5 phosphates.

D. R300 and K138

Although Arg-300 maintains contact at the fourth position for the most part of all trajectories, there are some differences between systems as mentioned in the GGGCA_4g82_swap case. Furthermore, it is interesting to note that the GAACA_4g82 system shows a slightly lower percentage of time contacting with position 4 as compared to the GGGCA_3vd1 system (Figure 3.12). Particularly, monomer A of copy 1 is contacted only 10% of the time. One can postulate that a less stable Lys-138 might influence the instability of Arg-300. This correlation was also previously described (Pan & Nussinov, 2010). Additionally, GGGCA_3vd1 is the only system that exclusively touches the fourth position. A visual inspection of trajectories show that, although most other systems contact the fourth position, they also can contact other nucleotide positions simultaneously, specifically the third-position nucleotide. Further analysis of Arg-300 was done by monitoring the distance between Arg-300 and guanine that is complementary to the cytosine in the fourth position (Appendix Figures A.6 – 9). Of the 16 monomers simulated, composing a total of 8 dimers, three showed Arg-300 contacts to the fourth position 10% of the time or less. Additionally, two of those same three monomers, which belong to GAACA_4g82 and GGGCA_4g82_swap systems, had no nucleotide-specific contact with Lys-138.

Both Arg-300 and Lys-138 bind better when the guanine-rich sequence is present, but why? Arg-300 and Lys-138 contact their respective binding partners (Figure 3.13). The GGGCA_3vd1 system is distinct with respect to the rest. Not only do Arg-300 and Lys-138 bind to the same nucleotides across both monomers and copies, but also both residues reach into the major groove as a single aliphatic chain bound by van der Waals

interactions, rather than to two smaller distinct side-chains in a sea of polar solvent. More quantitatively, the graphs in Figure 3.13 show the difference across all four systems of the distance between the CG carbon of Lys-138 and the CB carbon of Arg-300. Apart from the GGGCA_3vd1 system, two other monomers also manage to retain this proximity for the entire trajectory, them being monomer B of the copy 2 GGGCA_4g82 system as well as monomer B of copy 1 GAACA_4g82 system.

In addition to the aliphatic interaction the line of communication between Arg-300 and Lys-138 contains a second component. In the guanine-rich systems, the backbone of Lys-138 can contact the main-chain of helix H2 by Arg-300 and Gly-299. In most cases, the carbonyl oxygen can simultaneously hydrogen bond with the backbone nitrogens from both Arg-300 and Gly-299 (Figures 3.14). Again, the GGGCA_3vd1 system is unique in that it consistently maintains this contact (Appendix Figures A.10 – 13). In systems GAACA_4g82 and GAACA_3vd1_swap, contact between the two secondary structures is initially present, however is lost towards the end of the trajectory. Interestingly, monomer B of copy 2 GGGCA_4g82_swap system makes this contact within 5 ns of the end of the trajectory. A summary of the hydrogen bond percentage between Lys-138, Arg-300, and Gly-299 is shown in Figure 3.14.

E. Loop L1 analysis

A comparison of the loop L1 from the GAACA and GGGCA crystal structures shows marked differences in the loop L1 conformation. In addition to the unique coupling effect between Lys-138 and Arg-300, which is in part transmitted by the loop

L1 main-chain, loop L1 establishes a hydrogen bond between the backbone carbonyl oxygen of Ala-137 and the backbone secondary amine nitrogen of Ala-140. This contact acts as a bridge providing support in loop L1 between each segment of the loop before and after the turn (Figure 3.14). The intra-loop hydrogen bond was more prevalent in systems with the GGGCA response element. The percentage of contact duration for the GGGCA_3vd1 systems was more than three quarters of the simulated time compared to roughly around half of the time for the GAACA_4g82 systems. The swapped systems showed some effects from their new response elements by showing percentages largely between the values reported from the original systems. The distance between Ala-137 and Ala-140 of the intra-loop contact is shown in Figures 3.10 - 13 (top); importantly, a transition for the swapped systems can be observed, especially in the GAACA_3vd1_swap system.

Due to the flexibility of loop L1, RMSD based clustering was used to characterized the diversity of the loop L1 conformations within a trajectory as well as across systems. Using a RMSD cutoff of 1.36 Å, the number of clusters for all loop L1 per monomer is summarized in Table 3.2. Between the two original systems, one sees that GGGCA_3vd1 adopts a fewer amount of distinct conformations, represented by a slightly lower cluster count. Additionally, the most representative clusters are reported Table 3.2 with their associated percentage, representing the how much of the trajectory the respective cluster represent. A look at the cluster 1 shows that monomer A for both copies had significantly higher representative percentages than the GAACA response elements sequences. The swapped systems show comparable cluster counts except for two monomers in the GGGCA_4g82_swap system and one in the GAACA_3vd1_swap

system, where the number of clusters has significantly increased. Monomers with significantly high cluster counts correlate with monomers with low or no nucleotide or phosphate hydrogen bond contacts from Lys-138 (Figure 3.1). This correlation suggests that Lys-138 helps to stabilize loop L1 such that its conformation is less variable.

What does loop L1 conformation look like? In order to give an idea of what kind of conformation loop L1 adopts during the simulation, a comparison with the starting crystal structures was performed. Using a representative frame from cluster 1, the RMSD between these two structures was measured with respect to the loop L1 crystal structures. Most all comparisons showed large differences with the exception of GGGCA_3vd1 cluster 1 and the GGGCA crystal structure, which showed very small RMSD values - well under 1 Å (Table 3.3). Additionally, the RMSD of loop L1 was measured for the entire trajectory with respect to the GAACA and GGGCA crystal structures (Appendix Figures 3.14 – 17). The GGGCA_3vd1 systems clearly show a close similarity with the GGGCA crystal structure while the GAACA_4g82 systems were distinct than both crystal structures. A comparison of the B-factor values of the GGGCA and GAACA crystal structures show that the GAACA structure has significantly higher B-factor values for loop L1 compared to the GGGCA structure (Appendix Figures A.18 and A.19). A high B-factor would indicate a high degree of disorder, which largely explains the GAACA_4g82 system loop L1 not resembling the GAACA crystal structure. The swapped systems seem to undergo transitions as observed from slight increases in RMSD at the beginning of the trajectories. Interestingly, despite large deviations with respect to the GGGCA crystal structure for most of the trajectory, monomer B of copy 2, and to a

lesser extent monomer B of copy1, seems to converge with the crystal structure loop L1 during the last 5 ns of the trajectory (Appendix Figure A.15, bottom).

In summary, one observes distinct characteristics between systems with GGGCA and GAACA response elements. In the adenine-rich environment, Loop L1 adopts subtle, but nonetheless, diverse conformations whereas the guanine-rich environment promotes a uniquely favorable conformation as seen from loop L1's stability as well as its ability to transition to the same state adopted in the GGGCA crystal structure.

Percentage of HB Duration between K138 and DNA

MD cp1

18 23

			88	81									
P0	P1	P2	P3	P4	P5	P5'	P4'	P3'	P2'	P1'	P0'	P-1'	
t0	G1	G2	G3	C4	A5	T5'	G4'	C3'	C2'	C1'	a0'		
a0	C1	C2	C3	G4	T5	A5'	C4'	G3'	G2'	G1'	t0'		
P-1	P0	P1	P2	P3	P4	P5	P5'	P4'	P3'	P2'	P1'	P0'	
									78	60			

MD cp2

GGGCA_3vd1
GGGCA_4g82_swap

14 22 7

			95	80									
P0	P1	P2	P3	P4	P5	P5'	P4'	P3'	P2'	P1'	P0'	P-1'	
t0	G1	G2	G3	C4	A5	T5'	G4'	C3'	C2'	C1'	a0'		
a0	C1	C2	C3	G4	T5	A5'	C4'	G3'	G2'	G1'	t0'		
P-1	P0	P1	P2	P3	P4	P5	P5'	P4'	P3'	P2'	P1'	P0'	P-1'
									93	78			
									59	83			

MD cp1

21 16 45 17
16 51 72

P0	P1	P2	P3	P4	P5	P5'	P4'	P3'	P2'	P1'	P0'	P-1'	
t0	G1	A2	A3	C4	A5	T5'	G4'	T3'	T2'	C1'	a0'		
a0	C1	T2	T3	G4	T5	A5'	C4'	A3'	A2'	G1'	t0'		
P-1	P0	P1	P2	P3	P4	P5	P5'	P4'	P3'	P2'	P1'	P0'	P-1'
									34	38	23		
									69	62			

MD cp2

GAACA_4g82
GAACA_3vd1_swap

49 10 44 18
9 39 22 37

P0	P1	P2	P3	P4	P5	P5'	P4'	P3'	P2'	P1'	P0'	P-1'	
t0	G1	A2	A3	C4	A5	T5'	G4'	T3'	T2'	C1'	a0'		
a0	C1	T2	T3	G4	T5	A5'	C4'	A3'	A2'	G1'	t0'		
P-1	P0	P1	P2	P3	P4	P5	P5'	P4'	P3'	P2'	P1'	P0'	P-1'
									57				
									22	13	35		

Figure 3.1. Summary of Lys-138 Hydrogen Bond Contacts to the Response Element. Hydrogen bonds that occurred between multiple electron donors of a particular nucleotide or phosphate during the simulation were counted toward the percentage total of the trajectory.

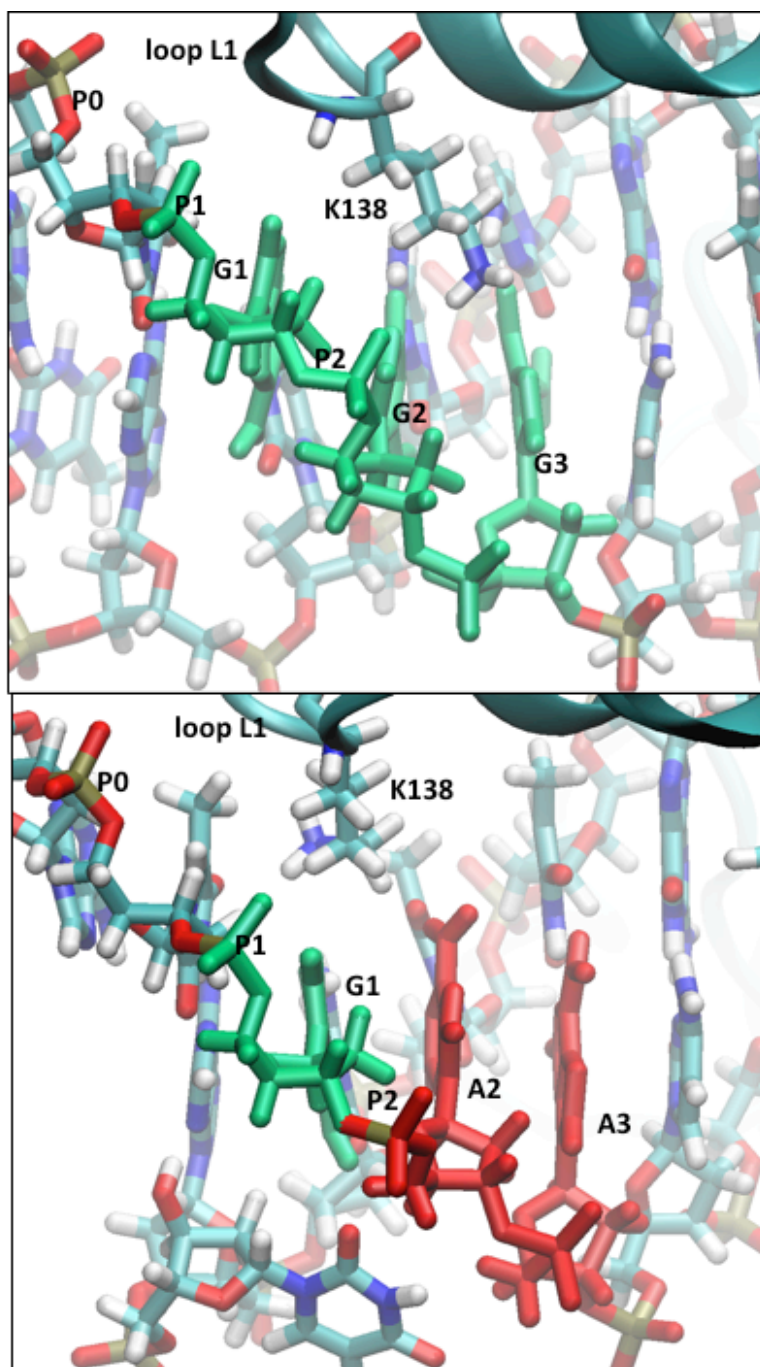


Figure 3.2. Snapshot of Lys-138 in Contact with Response Element Positions 1 – 3. Top, Lys-138 is shown when in presence of a GGGCA sequence. Bottom, Lys-138 is shown when in presence of a GAACA sequence.

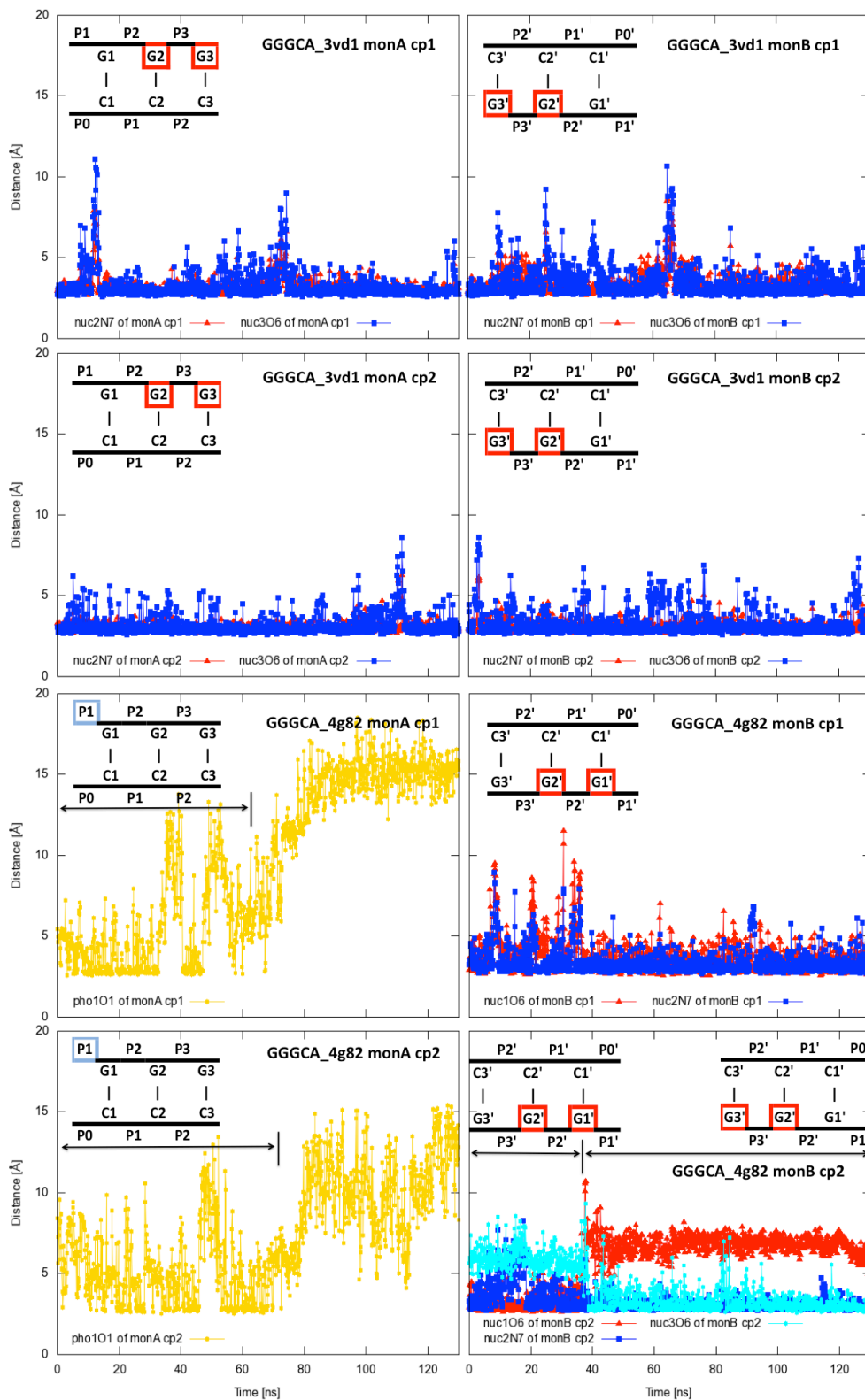


Figure 3.3. Lys-138 Contacts to DNA Monitored by Distance. Top four graphs show Lys-138 contacts in the GGGCA_3vd1 system. Bottom four graphs Lys-138 contacts in the GGGCA_4g82_swap system. Dashed boxes represent water-mediated contacts as opposed to direct contacts.

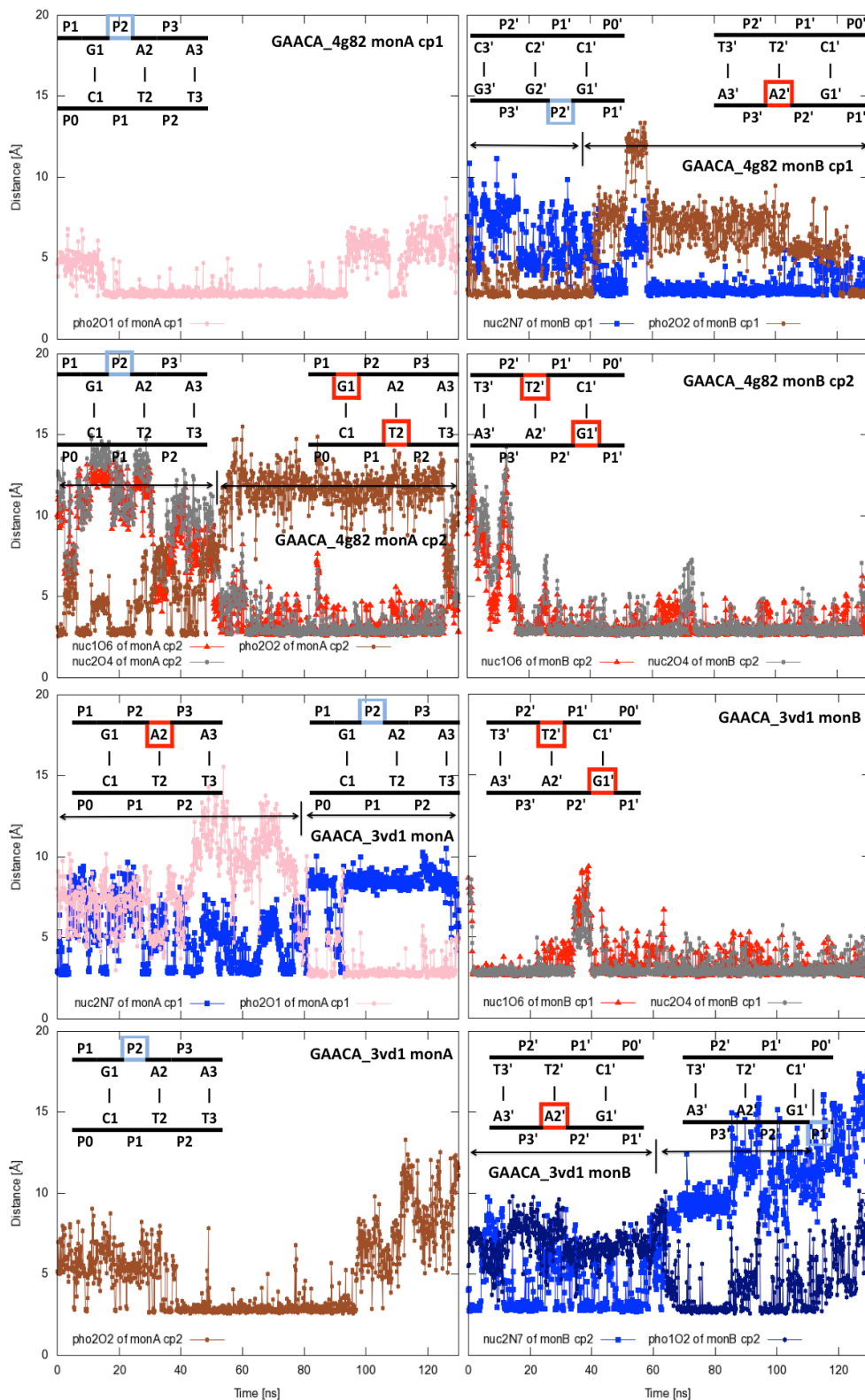


Figure 3.4. Lys-138 Contacts to DNA Monitored by Distance. Top four graphs show Lys-138 contacts in the GAACA_4g82 system. Bottom four graphs Lys-138 contacts in the GAACA_3vd1_swap system. Dashed boxes represent water-mediated contacts as opposed to direct contacts.

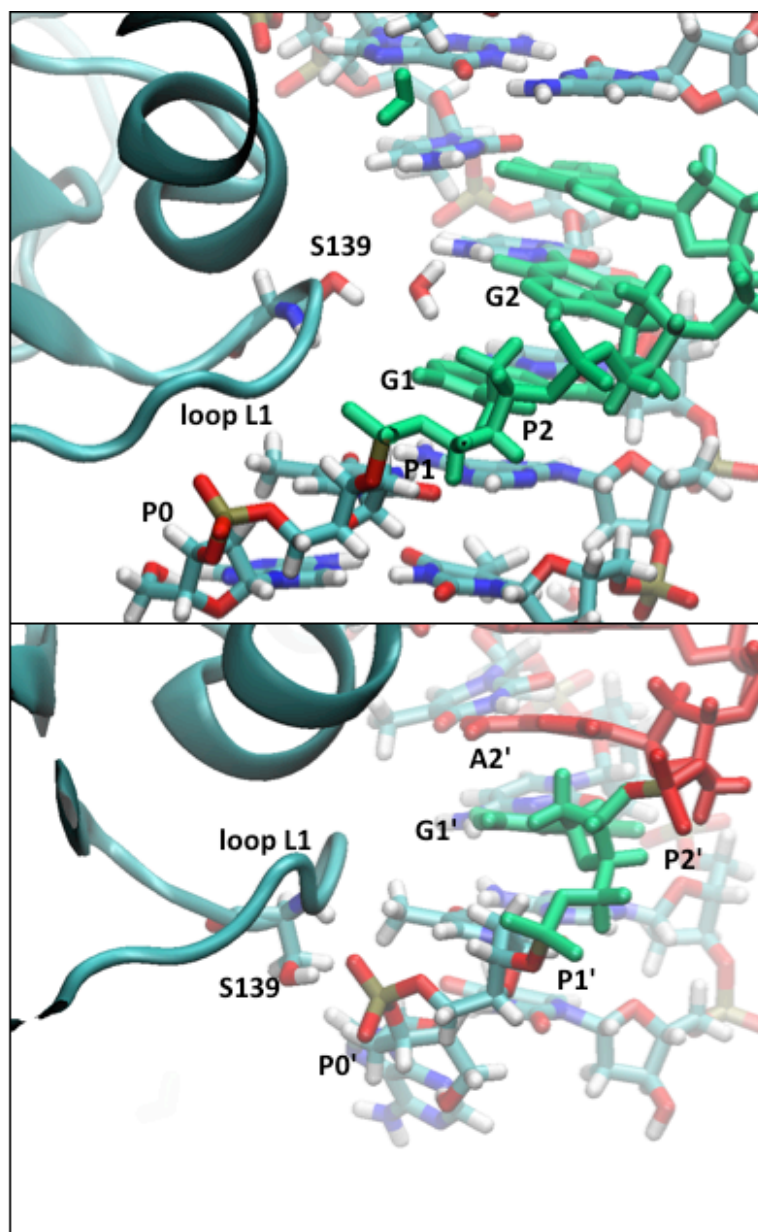


Figure 3.5. Snapshot of Ser-139 in Contact with Response Element Positions 1 – 3. Top, Ser-139 is shown when in presence of a GGGCA sequence. Bottom, Ser-139 is shown when in presence of a GAACA sequence.

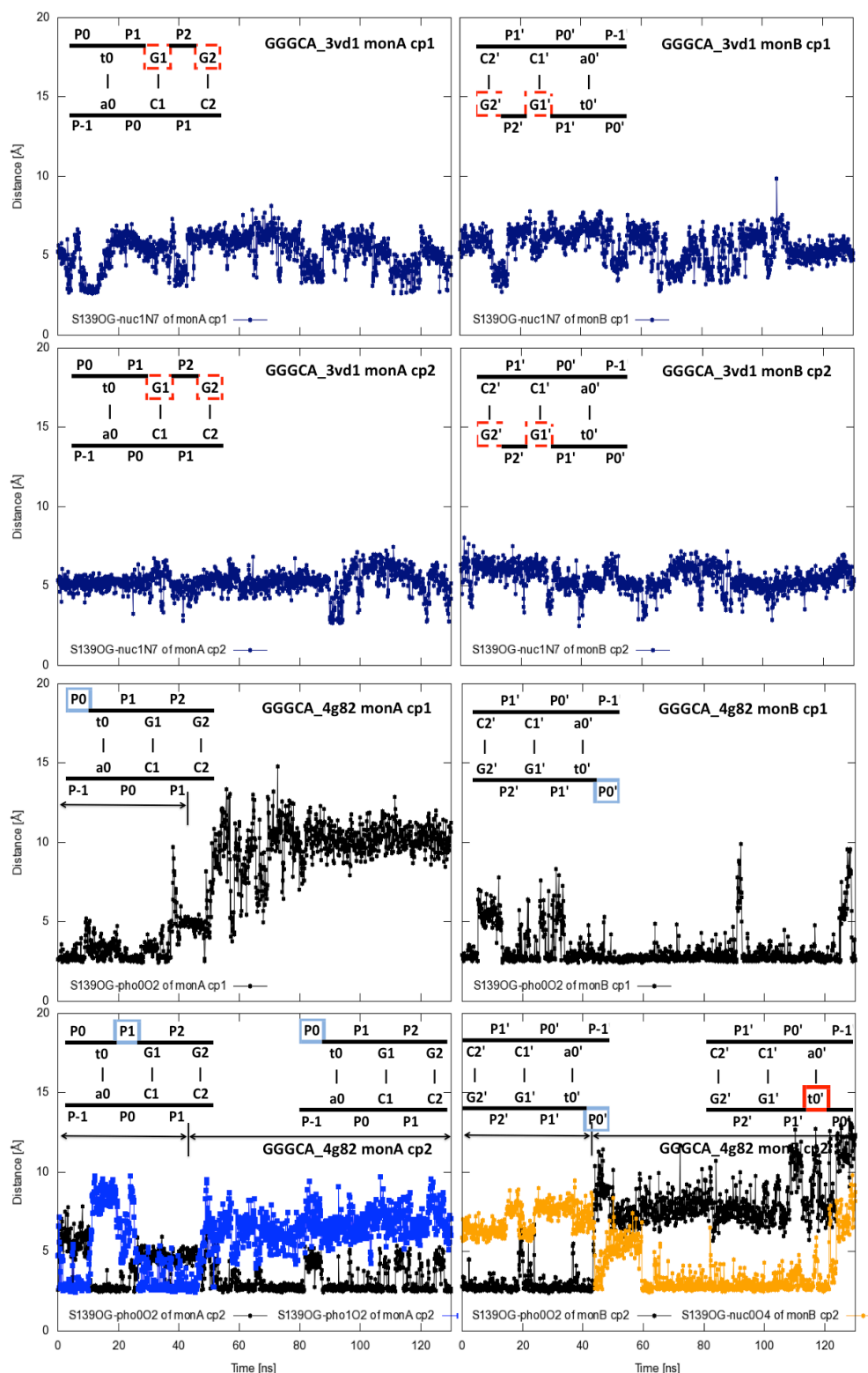


Figure 3.6. Ser-139 Contacts to DNA Monitored by Distance. Top four graphs show Ser-139 contacts in the GGGCA_3vd1 system. Bottom four graphs Ser-139 contacts in the GGGCA_4g82_swap system. Dashed boxes represent water-mediated contacts as opposed to direct contacts.

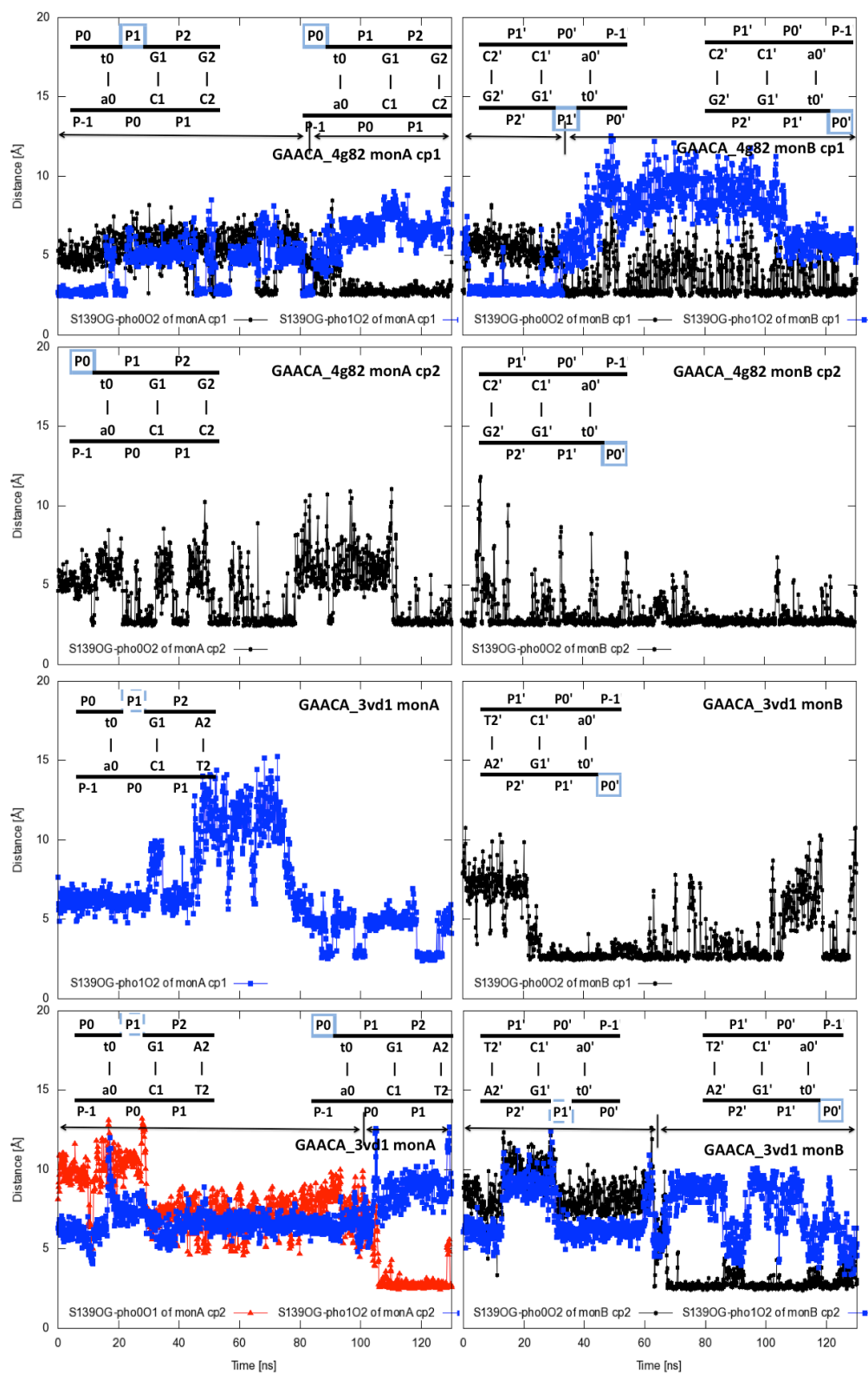


Figure 3.7. Ser-139 Contacts to DNA Monitored by Distance. Top four graphs show Ser-139 contacts in the GAACA_4g82 system. Bottom four graphs Ser-139 contacts in the GAACA_3vd1_swap system. Dashed boxes represent water-mediated contacts as opposed to direct contacts.

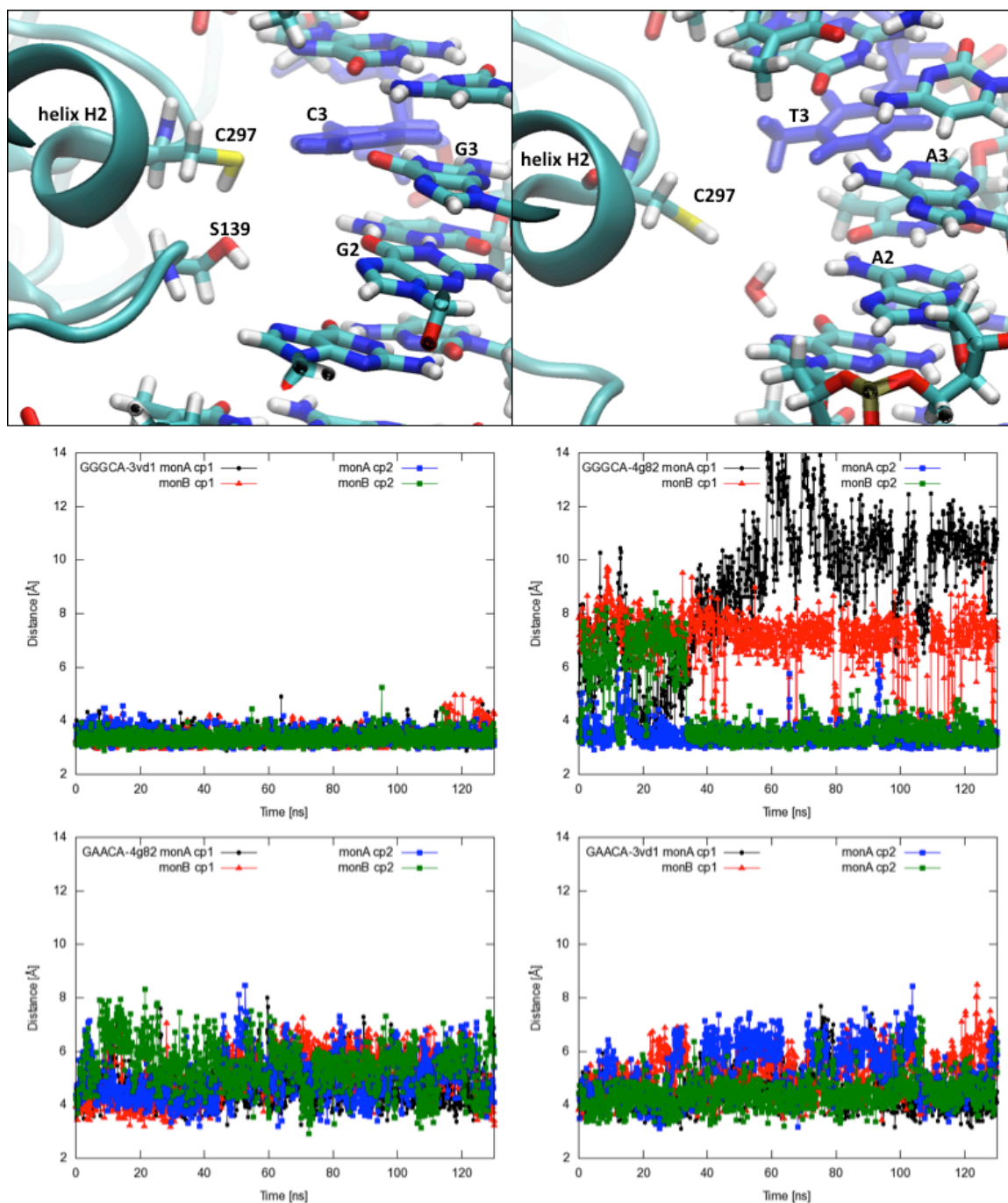


Figure 3.8. Distance between Cys-297 and position 3 nucleotide. Top, Arg-268 is shown when in presence of a GGGCA(left) and GAACA(right) sequence. Bottom, The relative distance is measured from the SG sulfur of Cys-297 and N4 nitrogen for GGGCA systems and O4 oxygen for GAACA systems. Systems are shown clock-wise starting from the top left: GGGCA_3vd1, GGGCA_4g82_swap, GAACA_3vd1_swap, GAACA_3vd1.

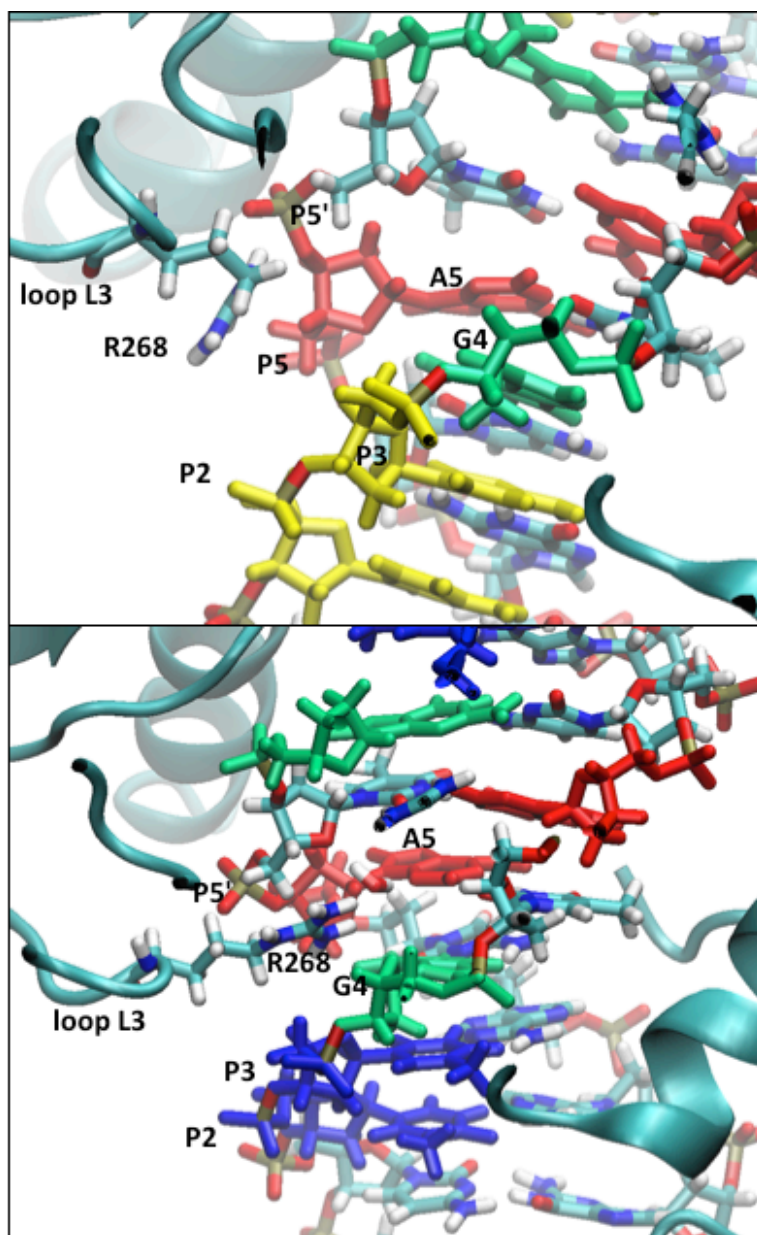


Figure 3.9. Snapshot of Arg-268 in Contact with Response Element Positions 1 – 3. Top, Arg-268 is shown when in presence of a GGGCA sequence. Bottom, Arg-268 is shown when in presence of a GAACA sequence.

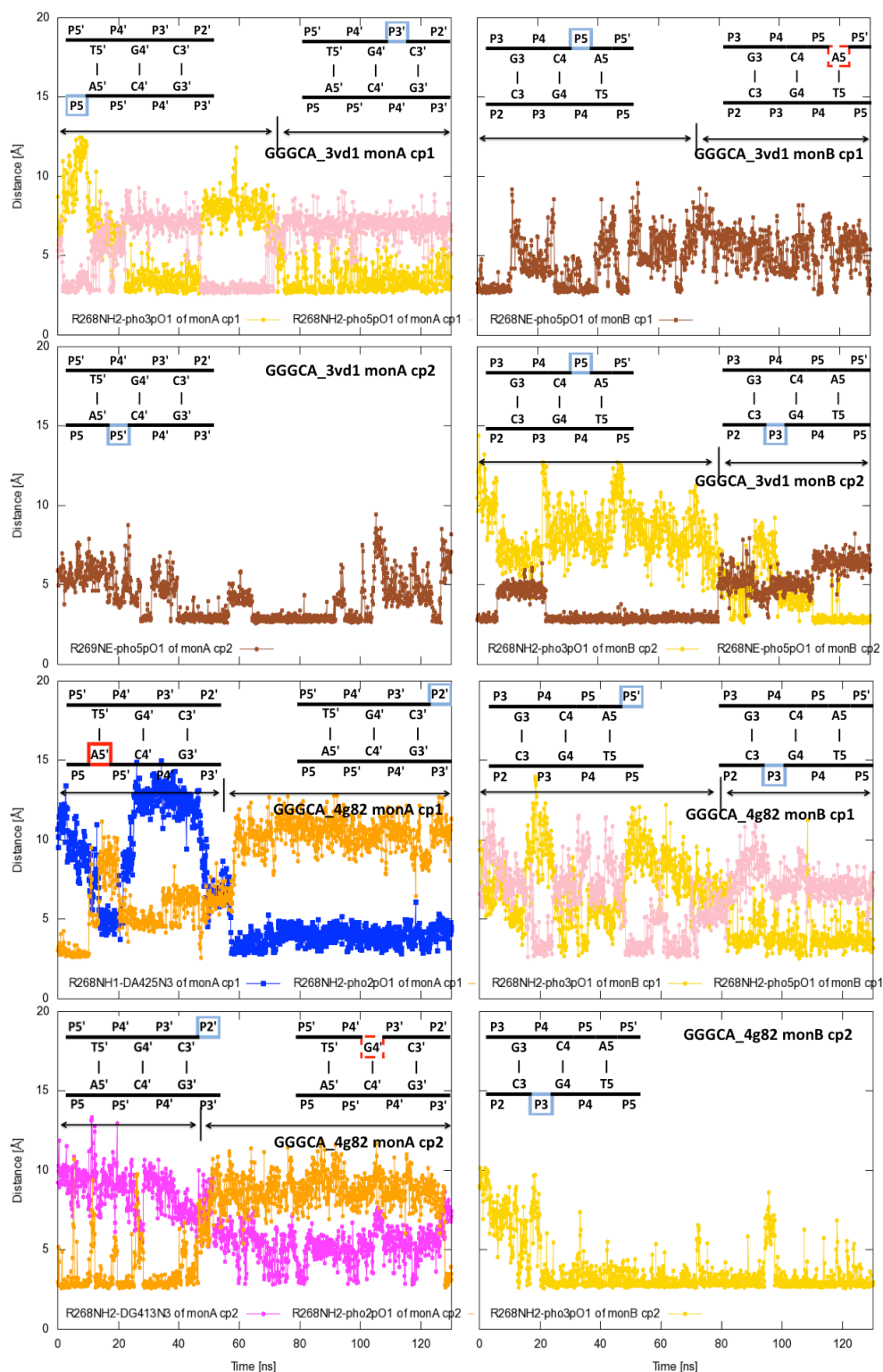


Figure 3.10. Arg-268 Contacts to DNA Monitored by Distance. Top four graphs show Arg-268 contacts in the GGGCA_3vd1 system. Bottom four graphs Arg-268 contacts in the GGGCA_4g82_swap system. Dashed boxes represent water-mediated contacts as opposed to direct contacts.

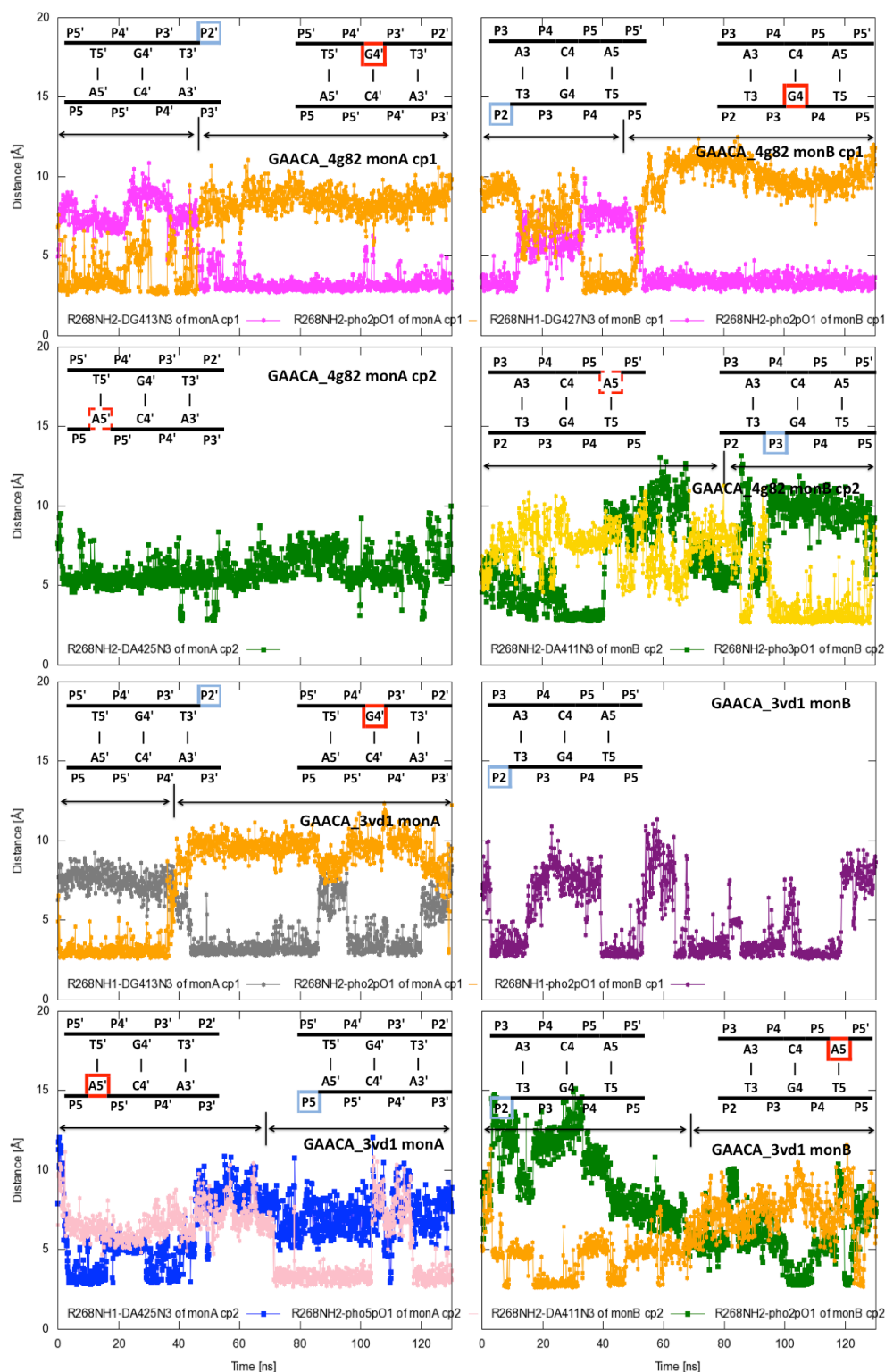
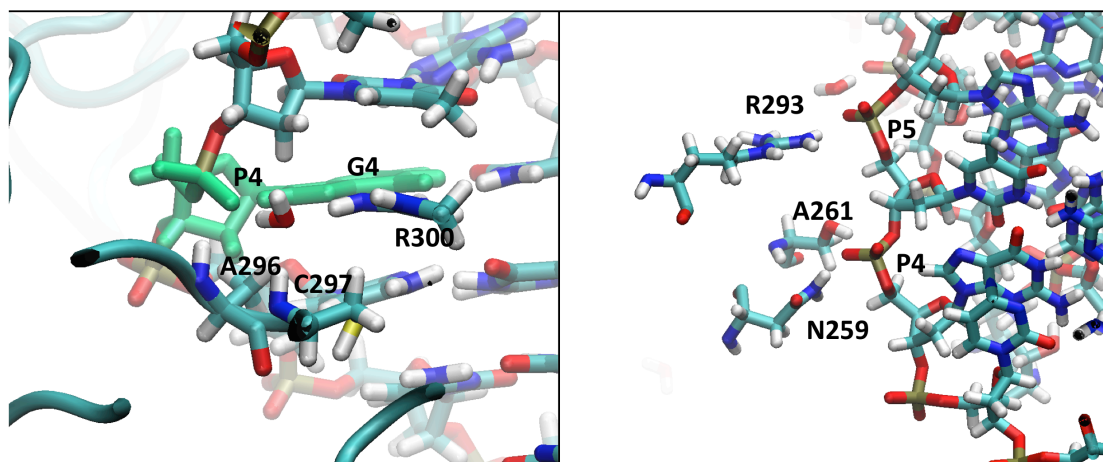


Figure 3.11. Arg-268 Contacts to DNA Monitored by Distance. Top four graphs show Arg-268 contacts in the GAACA_4g82 system. Bottom four graphs Arg-268 contacts in the GAACA_3vd1_swap system. Dashed boxes represent water-mediated contacts as opposed to direct contacts.

Table 3.1. Hydrogen bond percentage of Ser-139, Cys-297, and Arg-268 to DNA. This table represents the most observed contacts and the percentage of each contact present with respect to each monomer. Water-mediated contacts not included which represent a significant portion of contact time especially in the Arg-258 contacts to position 5 nucleotides of GAACA systems.

		S139OG_pho0	S139OG_pho1	S139N_pho0	C297_nuc3pN4
GGGCA_3vd1	monA_cp1	0	0	0	75
	monB_cp1	0	0	0	81
	monA_cp2	0	0	0	71
	monB_cp2	0	0	0	78
GGGCA_4g82_swap	monA_cp1	30	0	22	1
	monB_cp1	77	0	26	2
	monA_cp2	58	18	82	68
	monB_cp2	29	0	23	45
GAACA_4g82	monA_cp1	37	23	27	0
	monB_cp1	47	22	1	0
	monA_cp2	41	8	5	0
	monB_cp2	81	0	13	0
GAACA_3vd1_swap	monA_cp1	5	11	0	0
	monB_cp1	58	0	24	0
	monA_cp2	18	0	0	0
	monB_cp2	44	0	46	0

		R268_pho2pO1	R268_pho5pO1	R268_pho3pO1	R268_DA425N3	R268_DG413N3
GGGCA_3vd1	monA_cp1	16	25	37	0	0
	monB_cp1	14	33	0	0	0
	monA_cp2	21	44	1	1	0
	monB_cp2	12	49	8	0	0
GGGCA_4g82_swap	monA_cp1	13	0	0	13	1
	monB_cp1	15	15	20	0	1
	monA_cp2	31	8	0	3	0
	monB_cp2	2	0	48	1	0
GAACA_4g82	monA_cp1	23	16	0	2	35
	monB_cp1	13	1	0	16	20
	monA_cp2	0	9	0	2	0
	monB_cp2	21	23	2	12	4
GAACA_3vd1_swap	monA_cp1	25	0	9	2	29
	monB_cp1	53	5	8	2	0
	monA_cp2	19	33	0	7	1
	monB_cp2	38	19	4	5	2



System		R300_nuc4p	A196_pho4p	R293_pho5p	N259_pho4p	S261_pho4p
GGGCA_3vd1	monA_cp1	99	99	81	95	50
	monB_cp1	100	98	74	88	68
	monA_cp2	95	96	87	76	91
	monB_cp2	100	100	52	68	99
GGGCA_4g82_swap	monA_cp1	1	15	98	28	91
	monB_cp1	0	97	20	20	92
	monA_cp2	93	97	70	65	91
	monB_cp2	78	99	92	99	86
GAACA_4g82	monA_cp1	10	99	96	98	90
	monB_cp1	98	100	98	88	99
	monA_cp2	84	85	82	70	71
	monB_cp2	80	86	95	79	58
GAACA_3vd1_swap	monA_cp1	100	100	76	97	91
	monB_cp1	89	94	90	71	92
	monA_cp2	67	97	87	92	89
	monB_cp2	99	97	76	96	85

Figure 3.12. Summary of DNA-Binding Residue direct contacts hydrogen bond percentage. Top, snapshot of residue sets bound position 4 and 5 of response element. Bottom, This table represents the most observed contacts and reported is the percentage of each contact present with respect to each monomer.

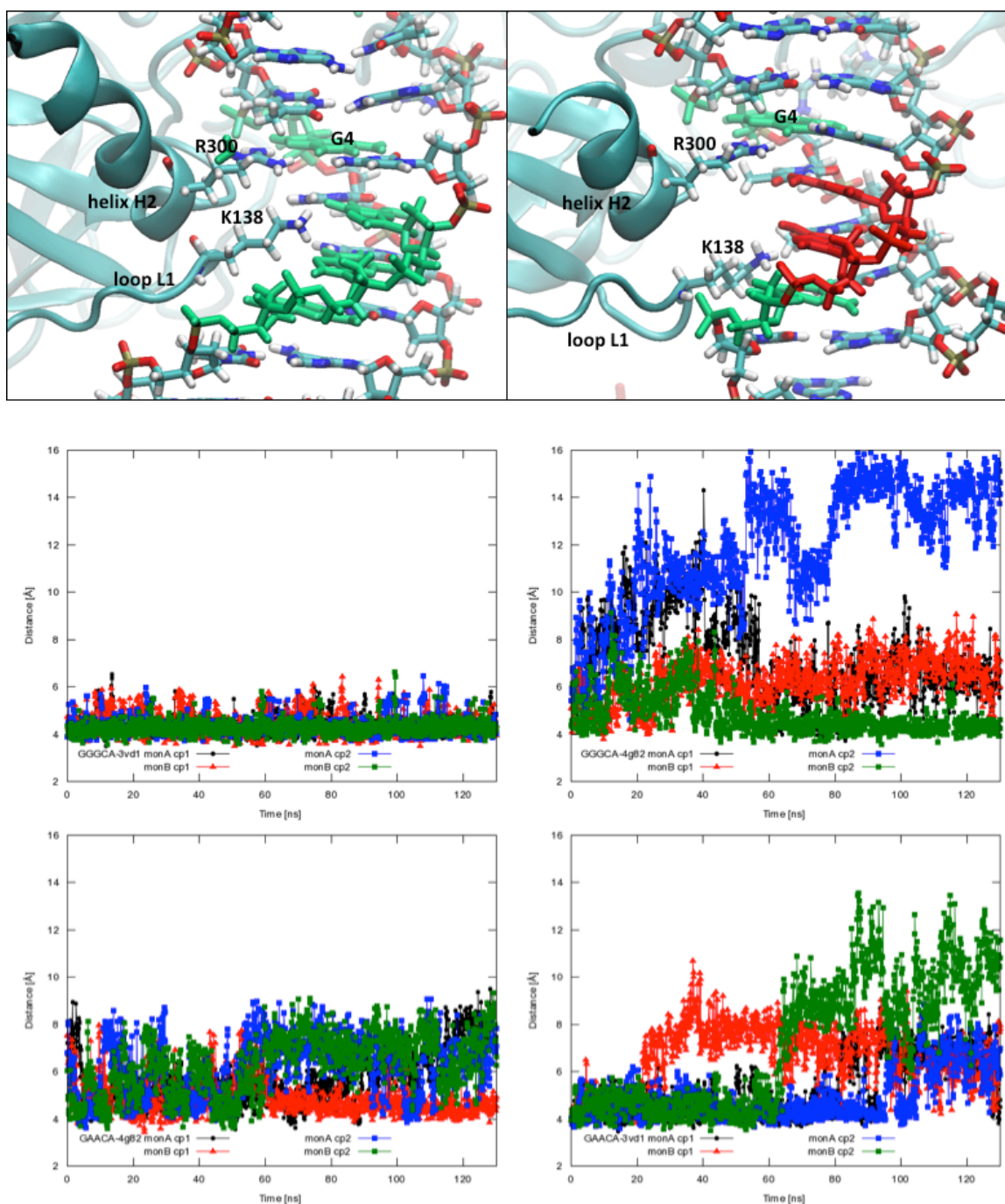


Figure 3.13. Distance between the carbon chains of Lys-138 and Arg-300. Top, snapshot of Arg-300 with Lys-138 in GGGCA and GAACA systems. Bottom, The distance between the carbon chains of the two residues was measured between the CG carbon on Lys-138 and CB carbon on Arg-300. Clockwise starting at the top left represent the systems in the following order: GGGCA_3vd1, GGGCA_4g82_swap, GAACA_3vd1, GAACA_3vd1_swap.

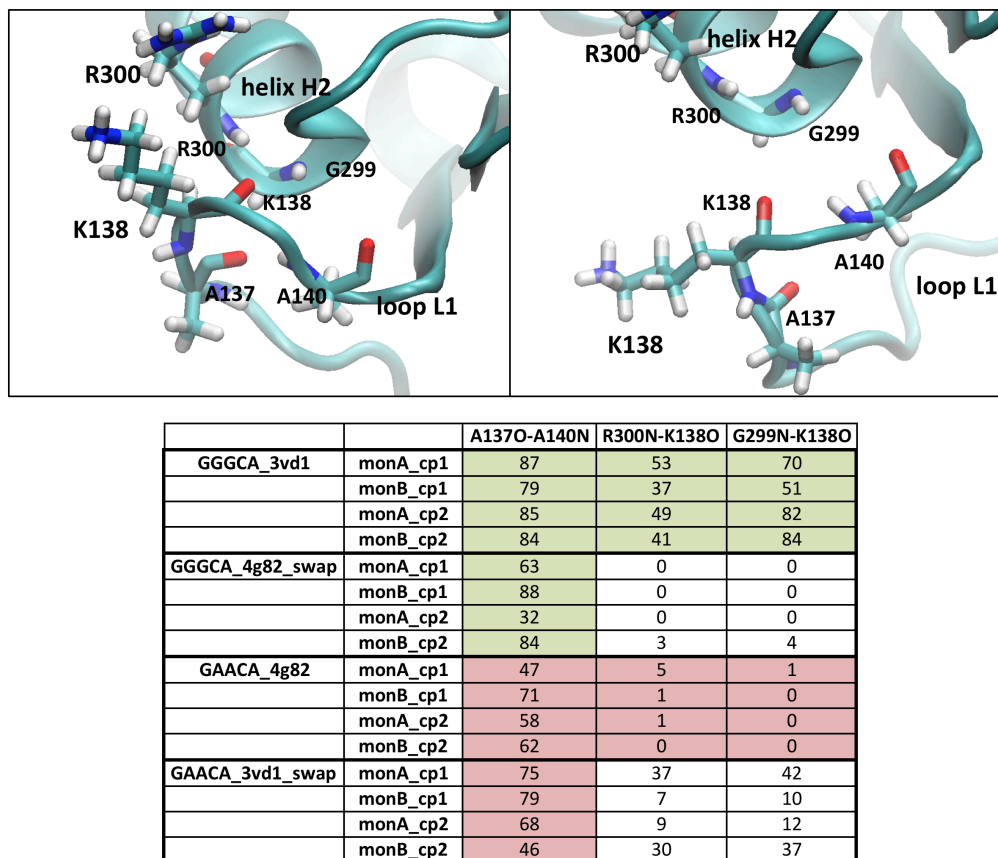


Figure 3.14. Percentage of time that the loop L1 and helix H2 hydrogen bonds occur during the trajectory. Top, snapshot of loop L1 and helix H2 contact atoms for GGGCA(left) and GAACA(right) systems. Bottom, table which summarizes the hydrogen bond percentage where each row represents a monomer of a system. Three hydrogen bond percentage values are reported including the intraloop L1 contact composed of oxygen, O, of Ala-137 and nitrogen, N, of Ala-140, and two contacts between loop L1 and helix H2 which are N of Arg-300 and O of Lys-138, and N of Gly-299 and O of Lys-138.

Table 3.2. Number of clusters and percentage of the four most representative clusters for loop L1 generated from RMSD-cutoff of 1.36 Å. Top, table summarizes the number of cluster for each monomer of system. Bottom, table summarizes percentage of clusters, going from left to right, cluster 1 – 4.

	monA_cp1	monB_cp1	monA_cp2	monB_cp2
GGGCA_3vd1	9	21	10	19
GGGCA_4g82_swap	38	14	87	21
GAACA_4g82	28	17	22	17
GAACA_3vd1_swap	17	14	22	72

	monA_cp1	monB_cp1	monA_cp2	monB_cp2
GGGCA_3vd1	85.7/5.1/4.6/1.4	57.5/15.9/10.9/3.7	85.5/7.0/2.8/2.3	39.0/19.3/12.7/6.6
GGGCA_4g82_swap	46.9/23.1/4.7/3.8	80.3/7.6/4.7/3.4	19.7/7.4/7.2/5.7	61.9/11.1/7.0/6.4
GAACA_4g82	67.9/6.7/6.0/2.9	65.5/20.1/5.0/2.7	48.1/18.9/8.7/4.9	63.1/10.5/7.6/7.1
GAACA_3vd1_swap	64.1/17.9/4.4/3.5	68.5/13.7/6.1/5.7	66.3/9.6/7.4/3.5	37.9/11.7/8.8/6.9

Table 3.3. RMSD Comparison between loop L1 of cluster 1 representative structure and GAACA, and GGGCA crystal structures. Loop L1 structures were superimposed and RMSD measured.

	GAACA_4g82_L1A	GAACA_4g82_L1B	GGGCA_3vd1_L1A	GGGCA_3vd1_L1B
GGGCA_3vd1_L1A_cp1	1.3	1.2	0.7	0.7
GGGCA_3vd1_L1B_cp1	1.1	1.1	0.6	0.7
GGGCA_3vd1_L1A_cp3	1.3	1.2	0.5	0.5
GGGCA_3vd1_L1B_cp3	1.5	1.4	1.0	0.8
GGGCA_4g82_L1A_cp1	1.4	1.2	1.5	1.5
GGGCA_4g82_L1B_cp1	1.4	1.2	1.3	1.3
GGGCA_4g82_L1A_cp3	3.1	3.0	3.0	3.0
GGGCA_4g82_L1B_cp3	1.3	1.2	1.2	1.2
GAACA_4g82_L1A_cp1	1.0	1.0	0.6	0.7
GAACA_4g82_L1B_cp1	1.4	1.1	1.4	1.3
GAACA_4g82_L1A_cp2	1.5	1.3	1.5	1.5
GAACA_4g82_L1B_cp2	1.5	1.3	1.7	1.7
GAACA_3vd1_L1A_cp1	1.1	1.1	0.4	0.6
GAACA_3vd1_L1B_cp1	1.3	1.1	1.3	1.3
GAACA_3vd1_L1A_cp2	1.4	1.1	1.4	1.4
GAACA_3vd1_L1B_cp2	1.2	1.1	0.6	0.7

REFERENCES

- Ethayathulla, A. S., Nguyen, H. T., & Viadiu, H. (2013). 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*, 288(7), 4744-4754. doi: 10.1074/jbc.M112.408039
- Pan, Y., & Nussinov, R. (2010). Lysine120 interactions with p53 response elements can allosterically direct p53 organization. *PLoS Comput Biol*, 6(8). doi: 10.1371/journal.pcbi.1000878

CHAPTER 4

RESULTS : Quaternary Structure Analysis of the p73 dimer-RE Complex

A. Dimer-Response Element Complex Conformation

The most significant of all measured structural parameters in our analysis is the relation between dimer and response element. With a few exceptions discussed in subsequent sections, dimer arrangement is relatively constant irrespective of system. As a consequence, to evaluate the structural relation of the dimer to the response element, the dimer itself can be treated as a rigid object. The results obtained reveal that nucleotide sequence directs the dimer-DNA complex to different quaternary arrangements. The nature of the relationship between the dimer and the DNA can be thought of as a scissor mechanism whereby the dimer and DNA act as two rigid bodies that pivot by way of a hinge formed by nonspecific interactions between the DNA-binding residues and the DNA backbone atoms along the minor groove. The difference in nucleotide sequence naturally dictates the opening and closing of the complex. As a consequence of the Lys-138 interaction with positions 1 – 3 of the response element (discussed in detail in Chapter 3), the GGGCA sequence is able to draw the monomer closer to the DNA. Contrastingly, the GAACA sequence allows for far less interaction with Lys-138, therefore drawing the monomer away from the DNA. An analysis of the dimer-DNA complex conformation reveals the described difference (Figures 4.1 and 4.2). In order to measure the degree of opening and closing of the complex, the angle between the principal axes of the dimer and DNA was measured (average values reported are for the last 70 ns of simulation). The GGGCA_3vd1 system showed similar results for both copies and on average showed an angle of 119°. The GAACA_4g82 system adopted two distinct conformational states resulting in angle averages of 102° and 110°.

With respect to the swapped systems, most copies adopted one of the three distinct conformational states. Both copies of the GAACA_3vd1_swap system adopted the 110° angle. In the GGGCA_4g82_swap system, copy 1 seemed to adopt a conformation in between 102° and 110° while copy 2 adopted the closed conformation of 119°. An explanation for copy 1 of the GGGCA_4g82_swap system to adopt an angle that is in between the two open conformations is likely due to several factors. Arg-300's contact to DNA at position 4, which is critical for specific binding to p73-response elements, is absent in this copy. Additionally, Lys-138 manages to make a very stable contact with nucleotides in positions 1 and 2, which to a small degree helps compensate for the loss of contact at position 4 from Arg-300. A summary of the average angles per copy is shown in Table 4.1. Graphs of the angle for all systems through the entire trajectory are shown in Appendix Figure A.20.

In order to compare the simulation results with the p73 crystallographic structures, angle measurements of the 3vd0, 3vd1, 3vd2, and 4g82 crystal structures were obtained. Unlike the dimer conformation which changes in the 2- and 4-spacer structures, dimer-DNA angle remains relatively constant for the AGGCA and GGGCA dimer structures with 0-, 1-, and 2- spacers with an average value of 106.2° +/- 1.7°. Contrastingly, the GAACA structure has a significantly lower angle of 94°. Comparing the average value for the guanine-rich structures and the GAACA structure leads to a similar angle difference as the difference between GGGCA and GAACA simulated dimer systems. The increased angle in the crystallographic structures compared to the simulated systems is likely due to the interactions in the tetramer interfaces, which increase stability

and consequently lower the necessity for residues to interact with DNA components to maintain complex stability.

Importantly, angle measurements for the GGACA 4-spacer structures, which likely resemble a dimer structure since no tetramer interface is present, show average values of $113 \pm 1.1^\circ$, which are within the closed and open angles observed in the simulated systems. A likely explanation to why the angle lies between the two conformations could be due to Lys-138 contacting positions 1 and 2 instead of positions 2 and 3 in the GGGCA system.

B. GAACA Dimer Conformational Change

Apart from the two unique dimer-DNA conformations observed in the GAACA_4g82 system, further analysis shows that the relaxed conformation also undergoes a dimer conformational change. Analogous to a dumbbell with free rotating ends, the monomers rotate relative to one another where the axis of rotation is roughly through the dimer principal axis. In order to measure the difference in rotational conformation a dihedral angle was defined whereby the axis of rotation goes through the dimer. The angle is defined by four alpha-carbons, two in each monomer, all of them residing in beta-strands. The atoms in order of connectivity are: Ser-289 and Asn-145 from monomer A, and Asn-145 and Ser-289 from monomer B. The defined dihedral angle and the two distinct dimer conformations are shown in Figure 5.3. The rotation is observed after 50 ns of simulation and largely stays at this conformation for the remainder of the trajectory. In order to compare angle values for both conformations, the

reported average values are from 60 – 130 ns of simulation. The conformation shown on top represents the conformations adopted by GGGCA_3vd1 and GAACA_4g82 copy 2 dimers. On average both GGGCA_3vd1 copies have a dihedral angle of 161° . Similarly the GAACA_4g82 copy 2, which adopts the more acute angle relative to the response element, shows a average angle of 155° . The conformation of the more relaxed complex of the GAACA_4g82 system has an average angle of about 141° , which results in a 20° difference between states.

Most swapped systems tend to adopt the larger angle conformation although to slightly different degrees. GGGCA_4g82_swap systems adopt both conformations where copy 1 shows a smaller degree of rotation at 144° while copy 2 shows an average of 158° . GAACA_3vd1_swap systems behave similarly, adopting the larger angle conformation (average 154°). Appendix Figure A.21 shows graphs of the dihedral angle for the entire trajectories of all systems.

C. Sequence-Independent Dimer Motion

Starting crystal structure RMSD comparison showed a noticeable difference between the dimer of the GGGCA and GAACA crystal structures. The RMSD measurement of one monomer while superimposing the other revealed differences in all secondary structure regions except for three regions concerning helix H1, responsible for forming the dimer interface, and two DNA-binding regions, helix H3₁₀ and loop L3. This result suggests the three regions mentioned act as a pivot for the monomers to move

relative to one another. The resulting change between the two dimer structures can be characterized as a rotation between monomers within the plane of the dimer.

A similar motion was identified in system dynamics but to an even greater extent that was observed in the starting crystal structures. In order to characterize and compare this motion observed in the dynamics to the starting structures, a dihedral angle was defined to measure the angular rotation through the dimer plane. The dihedral angle was defined such that its principal axis of rotation is perpendicular to the dimer plane. The angle chosen is defined by alpha-carbons residing on beta-strands of the following residues listed in order of connectivity: Thr-143, Glu-291, Ser-235, which all reside in monomer A, and Arg-176 from monomer B. An image of the defined dihedral angle as well as the range of the rotation is shown in Figure 5.4. A comparison across all systems shows no significant difference between guanine-rich and adenine-rich systems (Appendix Figure A.22). The average of each system average over the trajectory is calculated to be $85.7^{\circ} \pm 3.1^{\circ}$. Although on average these systems show no difference, there are large variations within a trajectory with values which in extreme cases can range from about 75° - 100° .

The same dihedral measurement on the starting crystal structures reveal angles of 91° and 86° for the GGGCA and GAACA structures, respectively. The difference in results between the crystal structures and the simulations allow us to infer that tetramer interfaces provide additional support such that the rotation within the dimer plane is a natural consequence to compensate for the added quaternary interactions. It must be stressed though, that this dimer angle difference is relatively small in magnitude and

more importantly is inconsequential to the dimer-DNA conformation difference, which is response element dependent.

D. Dimer Interface and Dimer Conformation

Despite no sequence-dependent dimer plane rotation was observed in the dynamics, the fluctuation of this angle throughout the trajectory resulted in large conformational motions for all the systems. Considering the three regions that largely act as a pivot, helix H3₁₀ shows little change largely due to DNA- and zinc-binding residues, which are unaltered in all systems. Additionally, loop L3 shows some displacement, but only as a result of nucleotide sequence change as oppose to constant fluctuation. In contrast, analysis of the dimer interface showed a diverse set of weak interactions largely resulting in unstable monomer-monomer contacts.

The dimer interface is comprised of an important set of van der Waals interactions as well as hydrogen bonds. Irrespective of the simulation system, Pro-195 and Leu199 to different degrees comprise the van der Waal component of the dimer interface. The interaction partners are the following: Leu-199 (mon. A) with Pro-198 (mon. B), Pro-195 (mon. A) with Pro-195 (mon. B), and Pro-185 (mon. A) with Leu-199 (mon. B). Of the three interactions, the proline-proline interaction is by far the most stable and acts as the best point of contact between monomers (Appendix Figures A.23 and A.24). Hydrogen bond interactions occur sparingly for all systems with two exceptions: GGGCA_3vd1 copy 1 and GAACA_4g82 copy 2 (Appendix Figures A.25 - 27). These two systems manage to maintain a hydrogen bond network involving contacts between Asn-196 and

Val-263, and Asn-196 between both monomers. Comparatively, starting crystal structures show this same hydrogen bond arrangement. In addition, Val-263 also contributes to the interface through van der Waals interactions resulting in a more stable interface.

In order to characterize further the dimer interface, the buried solvent-accessible surface (SAS) area was calculated at discrete times from 50 – 130 ns every 20 ns (Table 4.2). Similar to the van der Waals and hydrogen bonds contacts, SAS values were similar across most systems. The exceptions were copy 1 of GGCA_3vd1 and copy 2 of GAACA_3vd1, which showed values slightly lower than the rest. These results correlate with the loss of van der Waals interactions between Leu-199 and Pro-185 for those two copies (Appendix Figures A.23 and A.24). Also in agreement with contact analysis of the dimer interface as well as the relative monomer movement is the relatively large variability of the buried SAS surface area within a trajectory.

E. Mechanism of Conformational Change

Analysis of the swapped systems' behavior describes the transition between the open and closed dimer-DNA conformations seen in the simulations from the unaltered starting structures. It is important to note that minimization and equilibration steps altered the initial crystal dimer-DNA complex structures from open to more closed conformations. It is observed that at the start of the simulation, initial dimer-DNA complex resembles the closed conformation of 110°. The conformational change during the steps prior to free dynamics can be largely accounted for by the absence of the tetramer interface.

Copy 2 of the GGGCA_4g82_swap system illustrates the transition from around 112° to 122° dimer-DNA angle change (Figure 4.5). This transition starts by the stabilization of the Arg-300 DNA contact from both monomers. Arg-300 of monomer A is initially contacting the guanine complementary to position-4 cytosine with only the NH2 nitrogen of its guanidium group. Before 20 ns of simulation, Arg-300 rearranges such that both NH1 and NH2 nitrogens contact the O6 and N7 electron donors of position-4 guanine. Subsequently, Arg-268 of monomer B slips out of the minor groove pocket by loosing contact with position-5' adenine and contacts the position-3' phosphate. This rearrangement allows Arg-300 of monomer B to regain contact with the position-4' guanine which was immediately lost after the start of the simulation. The anchoring of both Arg-300s to position-4 guanines allows the other components of the major groove response element recognition to bind to their respective partners. Cys-297 of monomer B is now able to stably contact cytosine complementary to position-3' guanine. Additionally, Lys-138 of monomer B, which had been making a stable contact to positions 1 and 2 guanines from the start of the simulation, subsequently moves to contact position-2' and -3' guanines. Further stabilization of the monomer B response element recognition happens when loop L1 forms a stable contact with helix H2 via the backbone contacts of Lys-138 (loop L1), and Gly-299 and Arg-300. Interestingly, the transition from the open to the closed dimer-DNA conformation just described, only requires one monomer of the dimer making stable contacts to position- 1 – 3 nucleotides, importantly Lys-138 to position-2 and -3 guanines.

The transition to the open conformation is seen in copy 1 of the GAACA_3vd1 system (Figure 4.6). In contrast to the open to closed conformational change described

above, the closed to open transition is far subtler since the starting dimer-DNA angle closely resembles the angle after the transition. Nonetheless, important dimer-DNA contact differences exist which mark the particularity of each conformation.

The transition begins by loss of contact between loop L1 and helix H2 of monomer B. After 40 ns, Arg-268 of monomer A moves from contacting position-2 phosphate to position-4 guanine in the minor groove. Lys-138 of monomer A which has lost nucleotide contact prior to the start of simulation, briefly returns by contacting position-1 guanine between ~50 – 70 ns of simulation. Soon after Lys-138 loses contact to position 1, loop L1 and helix H2 of the same monomer also lose contact. Additionally, Cys-297 of monomer B, which started contacting position-3 thymine, has slowly progressed to lose any contact with the nucleotide. As a consequence to these residue changes, the nucleotide-specific recognition of the response element of the dimer has lost important components such that monomer A recognizes only position-4 guanine and position-3 thymine via Arg-300 and Cys-297, respectively, while monomer B only recognizes position-4' guanine and position-1' guanine via Arg-300 and Lys-138, respectively. In addition, loop L1 and helix H2 contacts are lost in both monomers, which further contribute to the instability the nucleotide-specific recognition contrasting greatly with the contact profile in the guanine-rich systems.

Unlike the transitions discussion so far, the open conformation adopted by copy 1 of the GAACA_4g82 system not only represents the most open complex observed, almost forming a perpendicular Dimer-DNA conformation, but also undergoes a monomer-monomer rotation through the dimer principal axis (Figure 4.7). Like copy 1 of

the GGGCA_4g82_swap system, which does a similar transition, it represents the most unique conformational change of all the systems.

Not long after the start of the simulation, the most important nucleotide-specific contacting residue, Arg-300 of monomer A, loses complete contact with the position-4 guanine. After 30ns, contact between loop L1 and helix H2 in monomer B is lost most likely due to Lys-138 rearrangement in the same monomer. Consequentially, Cys-297 also of monomer B loses contact to position-3' thymine. The weakening of the major groove contacts toward the ends of the response element, most critically Arg-300, allow the repositioning of Arg-268 from both monomers such that the residues move from contacting phosphates to position-4' and -5' nucleotides in the minor groove. Like the other two aforementioned transitions, the loss of contacts by components composing the binding at the major groove allow for the dimer-DNA angle to decrease. Unique to this system is the dimer rotation associated with the change in dimer-DNA angle. As previously emphasized, the importance of Arg-300 seems to be indispensable in the proper binding of to the response element. The analysis of Arg-300 contacts to position-4 guanine shows only two system losing this contact, copy 1 of GAACA_4g82 and copy 2 of GGGCA_4g82_swap (see Chapter 3). Consequently, both copies undergo dimer rotations to similar degrees.

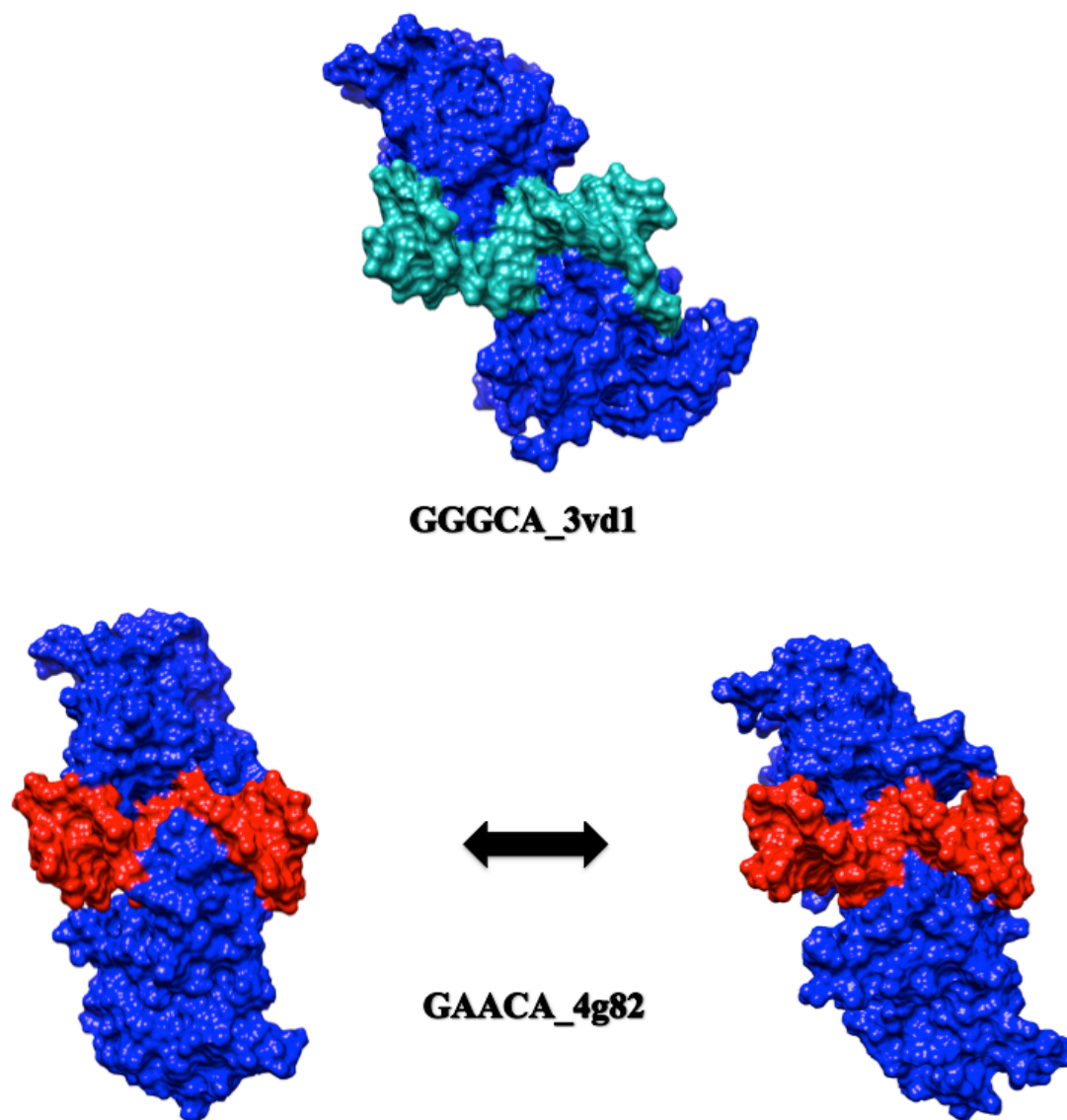


Figure 4.1. Surface Image of Dimer-Response Element Complex. Images show a representative frame of the conformations adopted in their respective trajectories. Top, GGGCA_3vd1 show a dimer conformation that swivels closer to the ends of the response element. Bottom, GAACA_4g82 adopts two conformations that are to different degrees closer to perpendicular arrangements with respect to the response element.

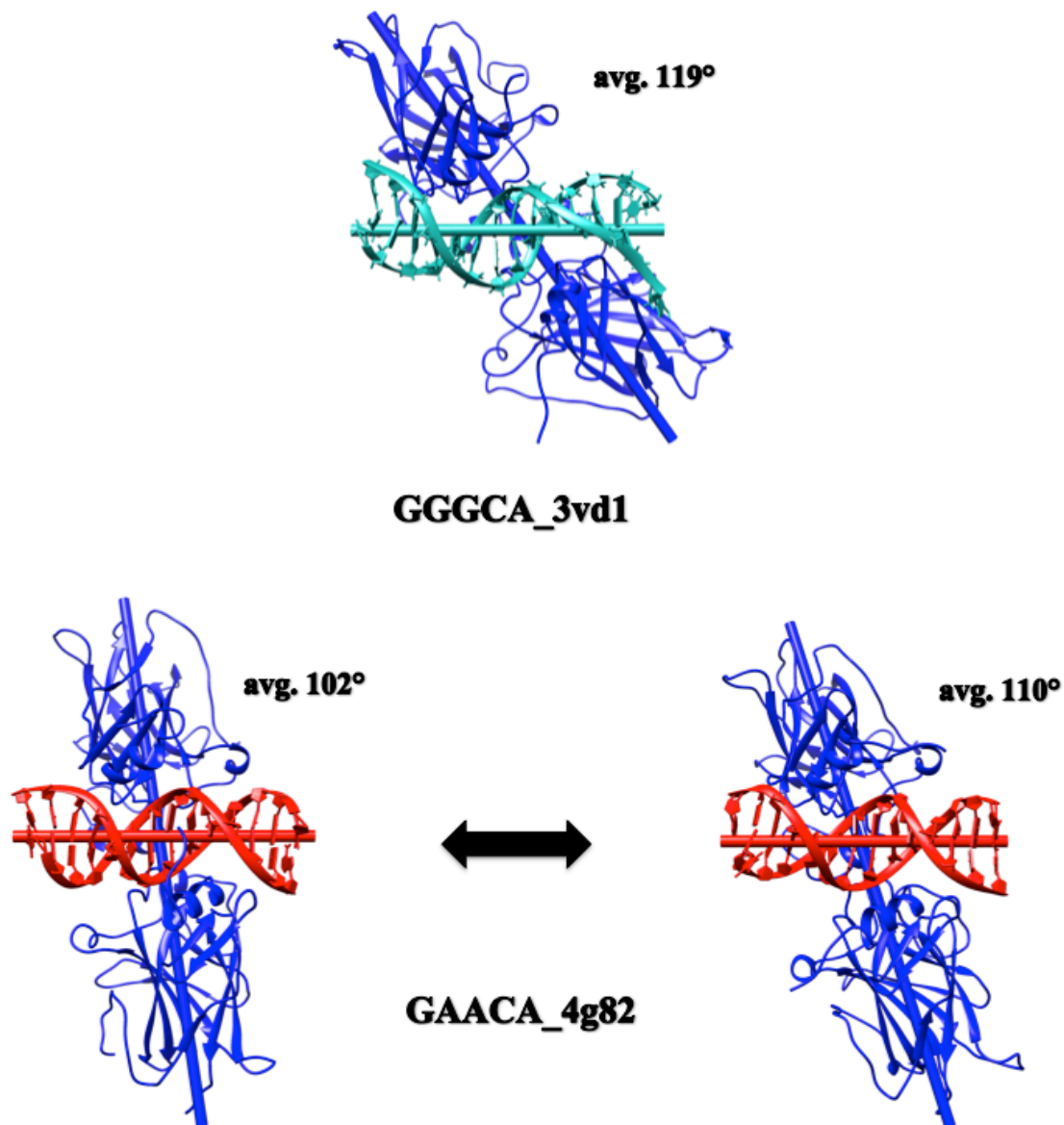


Figure 4.2. Backbone Dimer-Response Element Complex Image. Representative complex images show principal axes of the dimer and DNA represented in blue as well as green and red, respectively. Angle between the dimer and DNA principal axes is shown adjacent to the corresponding angle. GAACA_4g82 systems adopt two distinct dimer conformations with respect to DNA which both tend toward more perpendicular conformations compared to the GGGCA_3vd1 system. Averages reported are from 60 – 130 ns of simulation. The GGGCA_3vd1 system average is comprised of both simulated copies. The two distinct states reported for the GAACA_4g82 system correspond to copy 1, left, and copy 2, right.

Table 4.1. Average Angle between Dimer and DNA Principal Axes from 60 – 130 ns of simulation.

GGGCA_3vd1	cp 1	121
GGGCA_3vd1	cp 2	116
GGGCA_4g82_swap	cp 1	107
GGGCA_4g82_swap	cp 2	121
GAACA_4g82	cp 1	102
GAACA_4g82	cp 2	110
GAACA_3vd1_swap	cp 1	109
GAACA_3vd1_swap	cp 2	112

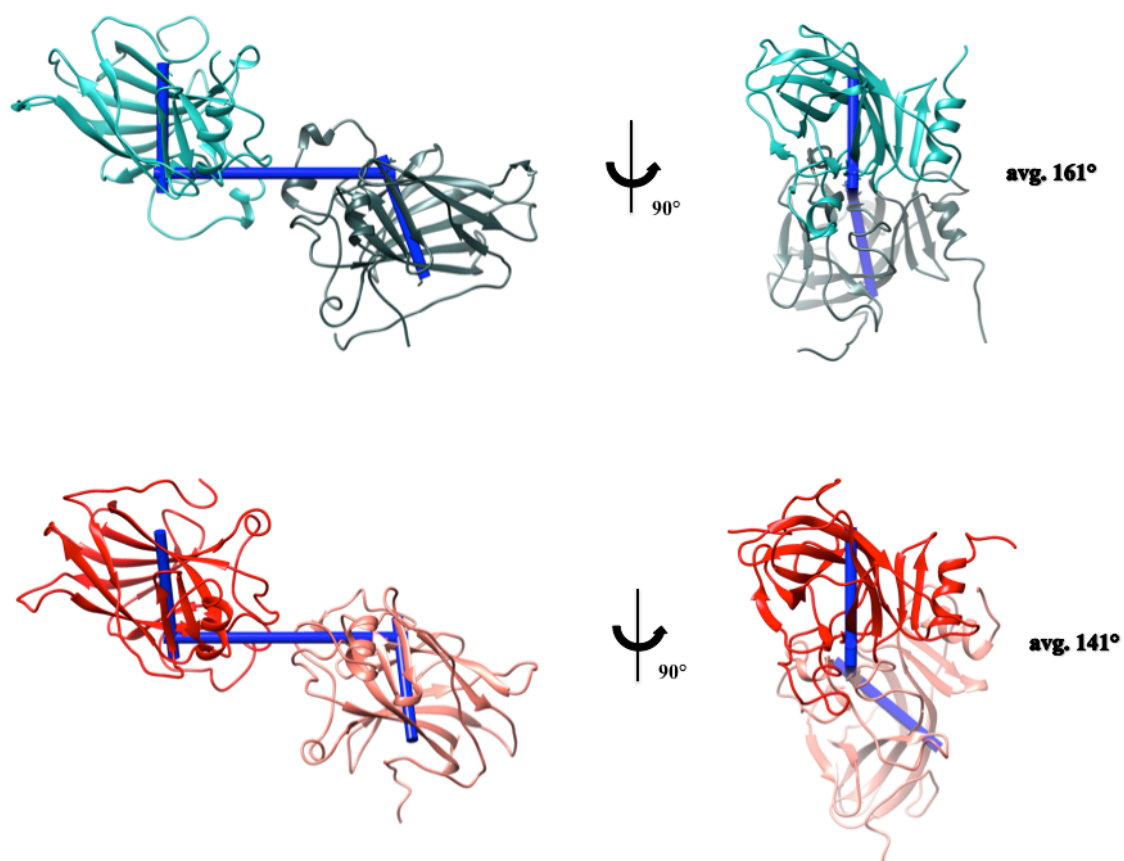


Figure 4.3. Dimer-Axis Rotational States. Two Rotamer-like conformational states through the dimer principal axis are observed. Systems that adopt the 110° or 119° dimer-DNA angle conformations adopt dihedral angle in the anti-like configuration shown in top row. Systems that adopt the more relaxed 102° dimer-DNA angle conformation also have a different rotamer-like conformation which is about 20° less.

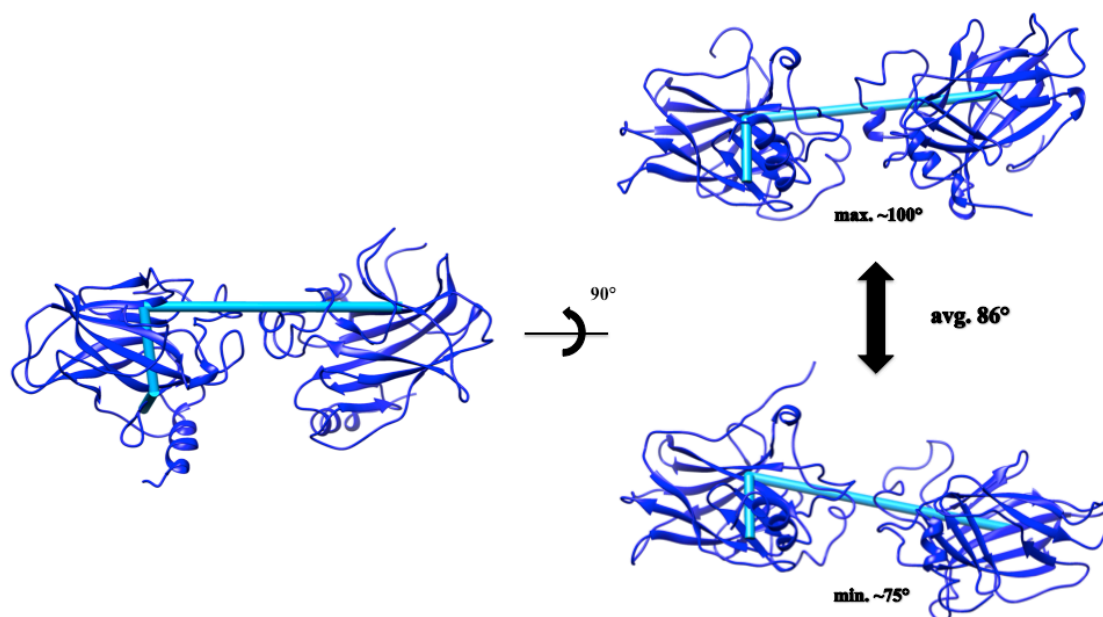


Figure 4.4. Intra-Dimer Rotation. Dimer rotation perpendicular to is dimer principal axis averages about 86° and change is independent of response element sequence.

Table 4.2. Buried SAS Area of Dimer Interface.

System Copy	50ns	70ns	90ns	110ns	130ns	Average	St. Dev.
GGGCA_3vd1_cp1	354	250	297	268	320	298	41
GGGCA_3vd1_cp2	196	214	189	230	237	213	21
GGGCA_4g82_cp1	245	272	257	245	260	256	11
GGGCA_4g82_cp2	263	293	252	258	222	258	25
GAACA_4g82_cp1	219	285	289	208	287	258	40
GAACA_4g82_cp2	181	275	259	259	202	235	41
GAACA_3vd1_cp1	253	291	332	289	181	269	57
GAACA_3vd1_cp2	197	165	109	121	133	145	36

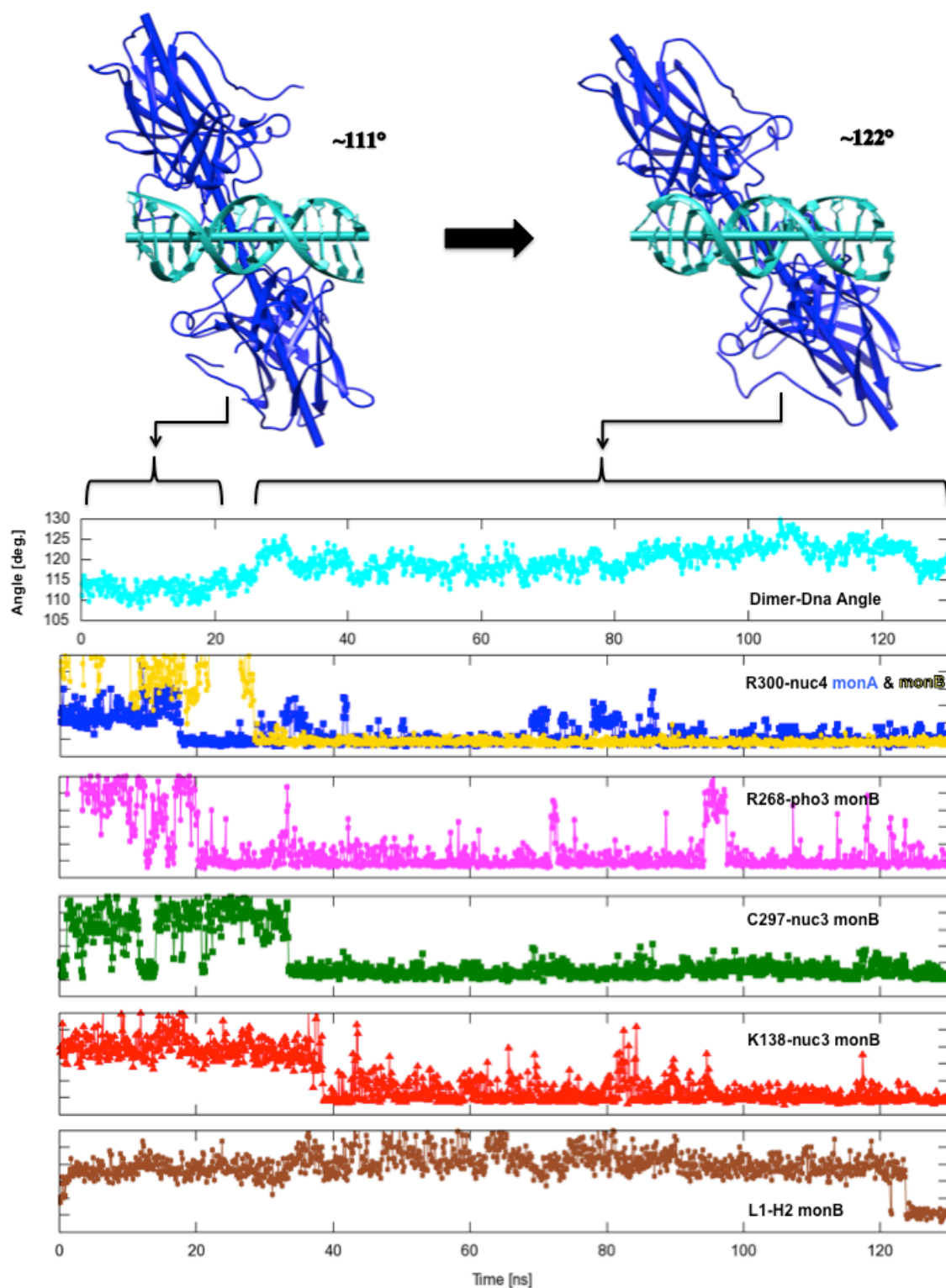


Figure 4.5. Transition of GGGCA_4g82_swap from open to closed Dimer-DNA conformation.

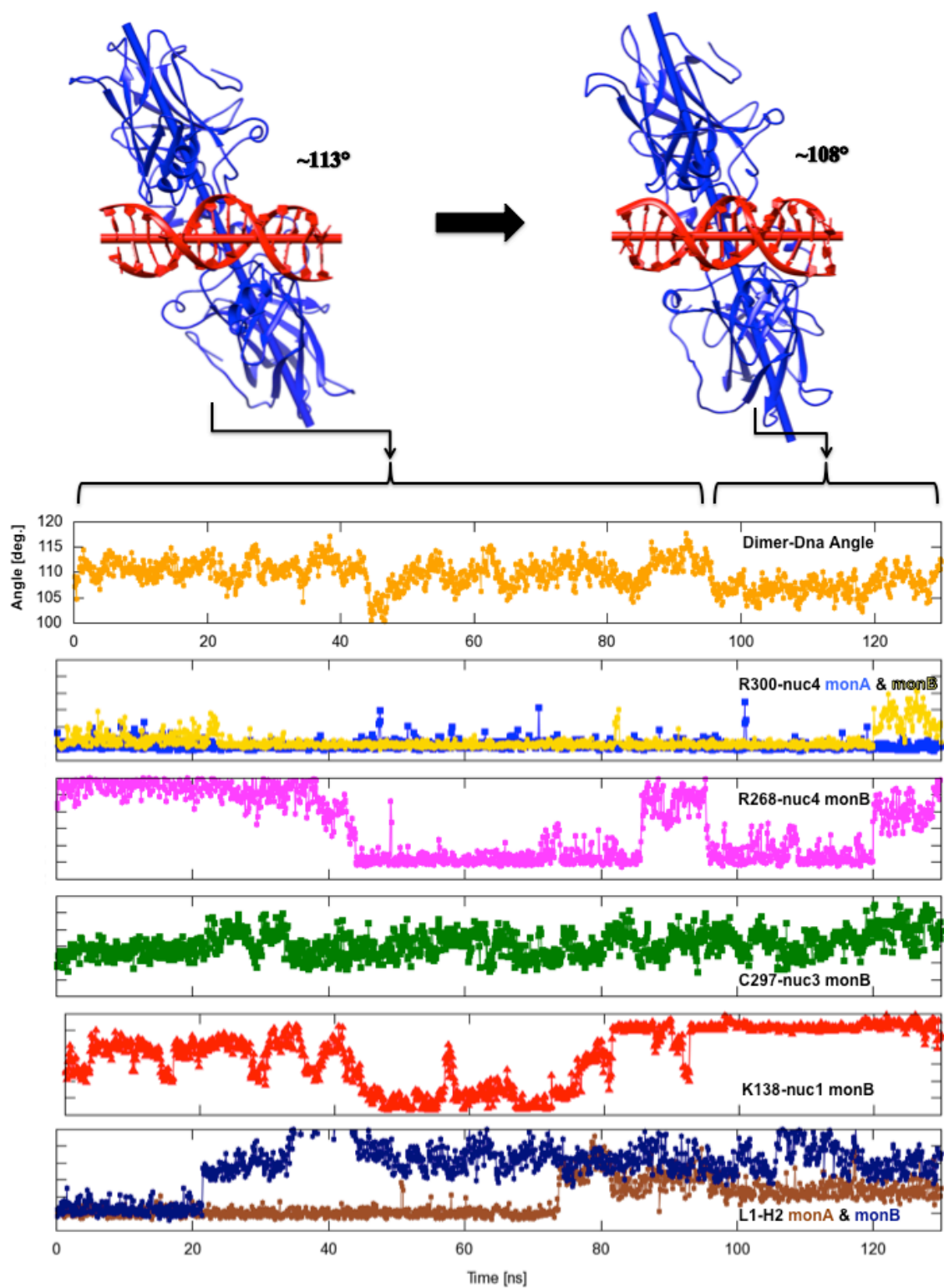


Figure 4.6. Transition of GAACA_3vd1_swap to open Dimer-DNA complex.

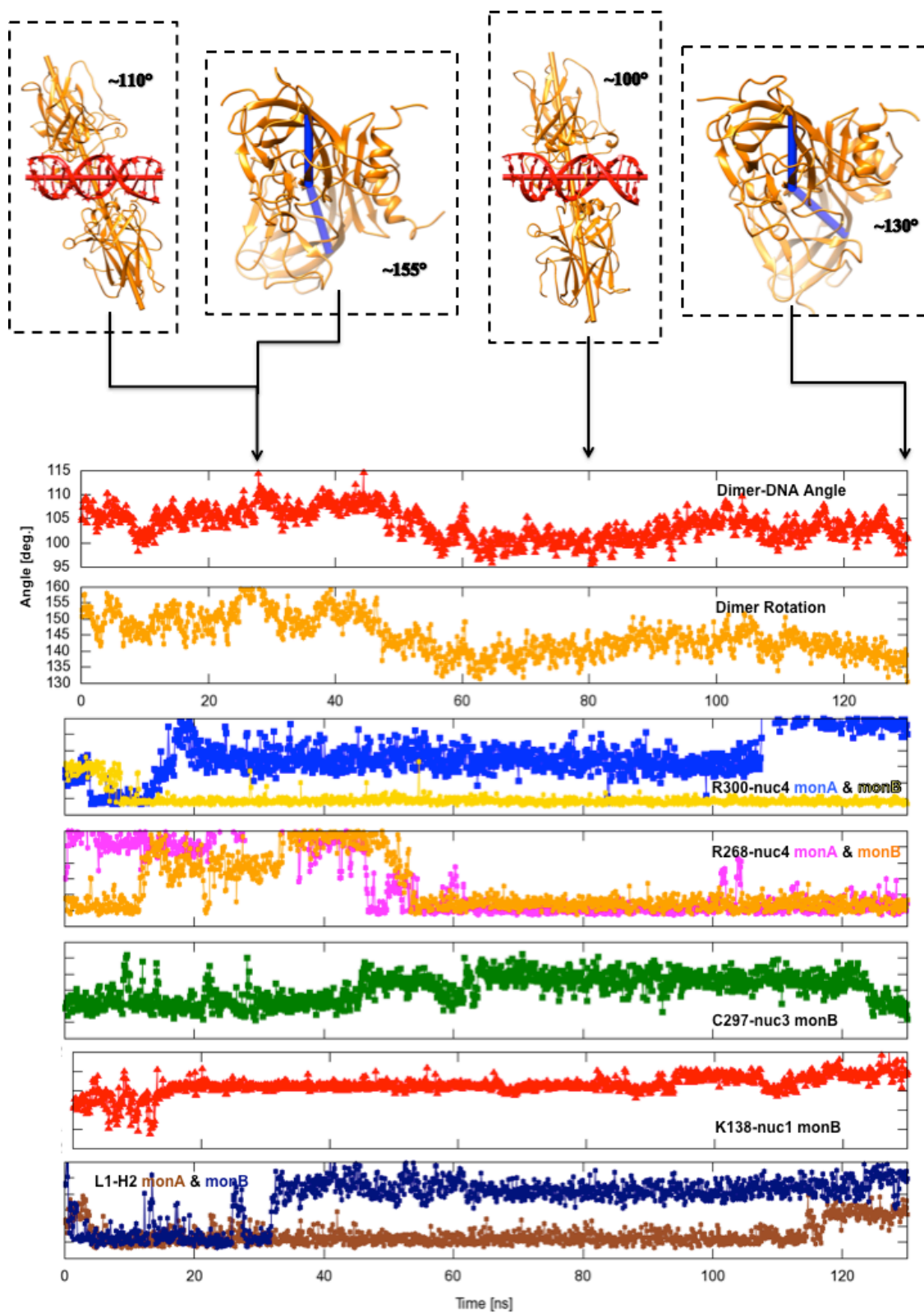


Figure 4.7. GAACA_4g82 Transition to relaxed state.

CHAPTER 5

DISCUSSION

A. Consequences of Guanine-Rich Sequences

The comparison between the guanine- and adenine-rich systems, show an important conformational change in relation to p73-DBD and the response element. Interestingly, several systems displayed heterogeneous binding profiles in relation to both monomers of the dimer as seen in several adenine-rich systems and most importantly, copy 2 of GGGCA_4g82_swap. In the guanine-rich swapped system, monomer A while maintaining contact with to position-4 guanine by Arg-300, it showed no important nucleotide contacts to position 2 and 3 guanines by Lys-138, which is crucial for the closed dimer-DNA conformation. Instead, Lys-138 of monomer B contacts to position 2 and 3 guanine guanines sufficed to form the closed conformation. As such, copy 2 of the GGGCA_4g82_swap system illustrates the transitions from closed to open conformation with Lys-138 contacting position 2 and 3 guanine to only one quarter-site. This case poses the question of what is sufficient for proper p73 response element recognition? Native p53-family response elements represent an immense combination of sequences particularly in positions 1 – 3. Naturally, deciphering first, why are there so many nucleotide arrangements, second, what can they all mean in terms of the p73 conformation, and third, what does this mean in terms of directing p73 transactivation? From binding studies one knows that purines in positions 1 – 3 show the best binding affinity to p73-DBD and even better to the Δ Np73 δ isoform (p73-DBD with oligomerization domain (OD)), especially guanines (Ramos A. G., 2014). In addition, guanine-rich sequences, particularly those containing at least two consecutive guanines in positions 1 – 3, show cooperativity in the Δ Np73 δ isoform (Ramos A. G., 2014). Both

crystallographic structures of p73-DBD tetramer complexes and MD simulations of the dimer complex relatively show open and closed conformations of the GGGCA and GAACA systems, respectively. Furthermore, the simulation studies show that although the GGGCA swapped system with Lys-138 contacting guanines in positions 2 and 3 is able to transition from the open to closed conformation, a similar contact profile in the GAACA unaltered and swapped systems where Lys-138 in one or both quarter sites contacts nucleotides from positions 1 or 2 is unable to change to the closed conformation and furthermore may be prone to unbinding at least as a dimer without the OD. Based on the communicative ability of p73 either as DBD-dimer as seen in the simulations or from binding studies of the dimer with OD, one postulates that conformational change of p73-dimer is likely to occur with only one guanine-rich quarter-site. One can extend this to say, given evidence of communication among monomers of a tetramer in the presence of guanines in position 1 - 3, that one guanine-rich quarter-site would also be sufficient for p73-tetramer conformational change.

There have been some studies that try to rationalize directing transactivation to particular cellular pathways by response element sequence patterns (Wang, Xiao, & Ren, 2009). They suggest certain patterns or rules that can explain many of the response elements belonging to particular functional pathways but none that can explain all. Based on the dynamics study of the p73 dimer, as well as the tetramer structures and experimental binding studies, one proposes that the flanking-site sequence can also direct the transactivation to a particular pathway. In relation to the guanine-rich sequence, p73 function might be more similar to a binary switch, such that a guanine-rich quarter-site sequence may serve to propel the functional consequence appropriate to that

conformation. This model would be independent of any particular cellular transactivation pathway and instead may be used in many functionally distinct response elements (apoptotic, cell-cycle repair, etc.), although this mechanism may be important and therefore evolutionarily conserved in many apoptotic response elements.

Studies of response element sequences have characterized proapoptotic response elements to contain an increased number of guanines in the flanking 1 – 3 quarter-site positions suggesting the importance of these nucleotides in initiation of a cell death response as mentioned in the chapter 1. Additionally, post-translational modifications of p53 seem critical for triggering apoptosis in proapoptotic genes, specifically the acetylation of Lys-120 (analogously Lys-138 in p73). This has been shown to occur in the proapoptotic targets *puma* and *bax*. Interestingly, the p53 response elements for these two genes contain more than one quarter-site with two or more consecutive guanines. One can imagine that, acetylating Lys-120 would liberate the interaction between loop L1 and the flanking region of the response element, subsequently allowing the protein-DNA conformation to transition from a closed to open conformation and allowing the transactivation of the gene. Like previously stated, although many contain an increased presence of guanines, several proapoptotic response elements also contain no guanine-rich sequences. Therefore, this type of mechanism is not necessarily exclusive to apoptotic transactivation mechanism but to several genes independent of function. One can further highlight that response elements with a guanine-rich sequence may serve to increase the regulation possibilities of the p53-family-dependent genes, as well as act as an additional checkpoint for important pathway events such that binding alone is not

sufficient for carrying out the message for transactivation but also appropriate conformational change of the protein-DNA complex (Figure 5.1).

B. Model for p73 Response Element Recognition

Conformational states that include an open dimer-DNA complex as well as an axial rotation of the dimer present in copies 1 of GAACA_4g82 and GGGCA_4g82_swap, present an interesting case for response element recognition. To summarize, dimer-DNA complex form a perpendicular-like conformation largely due to lack of stable Lys-138 interaction with position-2 and -3 guanines as well as Arg-300 contact lose to position-4 guanine. The perpendicular conformation has important similarities to electron microscopy models of full length p53 tetramer in the absence of DNA where the tetramer conformation forms a perpendicular arrangement with respect to all four monomers (Pham, Lucumi, Cheung, & Viadiu, 2012). In relation to the dimer simulation, should a tetramer complex be formed, a perpendicular-like arrangement would be necessary rather than an angled complex in order to form the dimer interface. Additionally, the loss of Arg-300 interaction with position-4 guanine causes a dimer rotation through its principal axis. The loss of important nucleotide-specific interactions including Arg-300, Lys-138, Cys-297, and Ser-139, make interaction to the DNA-backbone the predominant form of contact. These conditions suggest a possible model for p73 bound to nonspecific DNA.

Using the nonspecifically bound dimer-DNA complex just described, a model for p73 response element recognition can be suggested. To consider a model one must take

into account the OD, which is involved in all of the p73 isoforms except one. Importantly, analytical ultracentrifugation experiments reveal that p73-DBD with OD, specifically the $\Delta\text{Np73}\delta$ isoform, in the absence of DNA form a dimer in solution and not a tetramer, (Ramos A. G., 2014). This naturally leads to presume that p73 DNA recognition occurs as a dimer. Additionally, binding of p73 with or without the OD occurs only as a dimer in the presence of a half-site or a tetramer in the presence of a full-site response element.

Fluorescence Polarization experiments show considerable differences in binding affinity such that p73-DBD bound to half- or full- site response elements (dimer- or tetramer- DNA complex formed) have K_d values that vary between 2 – 3 μM while $\Delta\text{Np73}\delta$ bound to full-site response element show values which can vary between tens and a few hundred of nM in value (Ethayathulla et al., 2012)(Ramos A. G., 2004). Additionally, simulation results showing relative unbinding of the dimer in copies 1 of GAACA_4g82 and GGGCA_4g82_swap is congruent with experimental values of binding affinity, which shows poor affinity of p73-DBD for its consensus sequence. Further experimental evidence shows that cooperativity exists in the $\Delta\text{Np73}\delta$ isoform when it binds to the half-site (as a dimer) as well as the full-site (as a tetramer) response element, however no cooperativity is seen in the p73-DBD binding to either half-site or full-site response elements, (Ramos A. G., 2014). Given that $\Delta\text{Np73}\delta$ is a dimer in the absence of DNA in solution, one can assume that the OD provides a strong enough interaction between both monomers to provide stability for the dimer, which is absent in the p73-DBD. Furthermore, the poor binding affinity of the p73-DBD as well as the

degree of unbinding which occurs in some simulation systems, show evidence that the p73-DBD dimer-DNA complex is relatively unstable.

Structural analysis of the nonspecifically bound dimer-DNA complex seen in copies 1 of GAACA_4g82 and GGGCA_4g82_swap systems, show important structural consequences of the dimer and tetramer arrangement. Since the OD comes soon after helix H2 in the DBD, it is interesting to evaluate the change in position by helix H2. The dimer rotation caused by nonspecific binding to the response element shows an interesting rotation (by the same angle as the dimers) of the helix H2s. Figure 5.2 shows the dimer and helix H2 rotation in representative frames of copies 1 of GGGCA_3vd1 and GAACA_4g82 systems. The relative dimer rotation shown is defined by the same dihedral discussed in the chapter 4. Importantly the distance between helix H2 of monomers A and B, defined by the carbon-alpha of Glu-306 of both monomers, come significantly closer together such that the nonspecifically bound dimer shows a H2 helix-helix distance of 32 Å instead of 40 Å. Due to this rotation, one can assume that in the case of p73 dimers with added ODs, these domains would be also in effect closer to one another, possibly in a similar dimer conformation as it would be in the absence of DNA.

Analysis of the H2 helix-helix distance within the dimer of the crystal structures of p73-DBD bound to a full-site response element shows a similar distance as found in the specifically bound dimer conformation shown in Figure 5.3. Interestingly, the same measurement across dimers on the same side of the response element reveals a much closer distance that is roughly the same as the distance between helix H2s in the nonspecifically bound dimer conformation (shown in Figure 5.2). This is the case for 0-

and 1-spacer tetramer structures of AGGCA, GAACA (both 0-spacer), and GGGCA (1-spacer) sequences, respectively.

Assuming that OD's can bind to other OD's one at a time, using the given structural parameters, one postulates that in going from a DNA-free to DNA-bound state on a full-site response element, p73 changes the connectivity of OD's between the monomer of its own dimer to the monomer forming the tetramer interface on the same side of the DNA. OD contact would ideally provide a stronger communicative link between dimers than just the tetramer interface, which largely acts as a support framework. This structural change might provide an explanation to the binding studies that show cooperativity as a tetramer as well as a dimer. For dimers that bind to half-site response elements, since a communicative effect between monomers of the p73-DBD dimer from the simulations is observed, the role of the ODs in these cases would be to enhance the stability of the complex that lacks the support of the tetramer interface. Figure 5.4 shows H2 helix-helix measurement, as well as other parameters, for the p73 dimer, and the two possible OD connected tetramers.

Figure 5.5 shows a model in a sequence of diagrams representing the progression of p73 DNA recognition. P73, which acts as a dimer connecting the two monomers by interaction between ODs, first recognizes nonspecific DNA through regions loop L3, helix 3₁₀, and sheet S10, which recognizes the DNA backbone by residues Asn-259, Ser-261, and Arg-293, respectively. Rather than facing the DNA backbone via the minor groove, these nonspecific interactions would first approach the DNA strand from the major groove. Additionally, this conformation would still allow Arg-300 to sense the nucleotide environment since both monomers now face the DNA major groove. The

dimer, therefore, will be anchored between the ODs and be able to scan along the DNA tracks until it can recognize its response element.

Recognition of the response element is triggered when a sequence of the form 5'-XXXCWWGXXX-3' is sensed, where W is A or T and X can be any nucleotide. Arg-300 would be first to recognize the characteristic fourth position, which consequently will pull both monomers closer and around toward the minor groove and allow addition nonspecific contacts by Ala-296 and Cys-297, both right below Arg-300, to contact the position-4 phosphate. This would also allow Cys-297, via its side chain, to contact the position-3 nucleotide as well as to form the dimer interface between helices H1 and 3₁₀. The final arrangement of the dimer-response element complex (open or closed conformation) is fine-tuned by the nucleotide environment in positions 1 – 3 that is sensed by Lys-138 and consequently the arrangement the Arg-268 adopts in relation to the fifth position. At this stage, the ODs most likely act as a support for the dimer-DNA complex, which is needed as it lacks the stability of the tetramer interface.

After recruitment of a second p73 dimer to the adjacent half-site whereby stabilization of the tetramer complex occurs, ODs which had tethered the monomers to form the dimer, now are able to rearrange and bridge across dimers. Monomers, which bind to the same side of the response element, will in addition to their tetramer interface also interact through their respective OD. In the case of a response element with a guanine-rich quarter site, the communicative ability of the tetramer complex will be enhanced through the major-groove nucleotide-specific residue network of contacts, which not only is improved by the stability of both Arg-300 and Lys-138 but of the whole major groove recognition component. The dimer that recognizes the guanine-rich

sequences and adopts the more closed conformation would also induce the conformational change in the other dimer through the interaction of the ODs.

Further consideration for the aforementioned model of DNA recognition can be further deliberated by a thought-experiment on the two possible models. From binding studies, one knows the binding affinity is greatly increased with the presence of the OD. One also knows from the simulations that p73-DBD dimer is quite flexible at adapting its conformation to the nucleotide environment, which most likely is helped by a lower binding affinity compared to the DBD-OD construct. From this one gathers that there is an important consequence as a result of OD rearrangement in the tetramer form, if that were the case. The change in OD connectivity might enhance the communication between dimers such that the first dimer bound to the response element does not restrict the potential tetramer conformation from an informed full-site recognition and hence not dictate the final tetramer conformation. One senses that a structurally robust dimer-response element complex, the intra-dimer OD connected model, would have little incentive to accommodate for the dimer conformation adopted in the second half-site as a result of the nucleotide environment. Therefore it is likely that without OD rearrangement, the dimer to bind to the empty half-site would have to adopt a similar conformation from the already bound dimer-RE half-site complex since adopting an appropriate tetramer interface requires both dimers to sit next to each other in a parallel manner. In contrast, an already bound dimer whose ODs loose contact between monomers is likely to be nimble enough to provide more flexibility for the waiting half-site complex. One might expect that a free OD might help recruit the second dimer to the response element as well as coordinate to adopt the appropriate tetramer-RE

conformation by sensing the complete response element. The consequences of these models are illustrated using the case of a response element with one guanine-rich flanking site where the first dimer to bind does so at the non-guanine-rich half-site. Under the static model, one proposes that the dimer-DNA conformation, adopted from the sequence of the bound half-site, will remain unchanged regardless of potential dimer conformation that might be adopted as a result of the sequence of the second half-site. In contrast, the dynamic model presents a conformation of the initially bound dimer that is more flexible, due to the loss of monomer-monomer OD interaction, whereby after the second dimer binds to the half-site, the newly formed tetramer can sense the guanine-rich sequence of the second half-site, due to the added dimer-dimer OD interactions, and consequently conform to the more closed conformation. Figure 5.6 illustrates the two models in the case of a guanine-rich flanking sequence in the full-site response element.

In conclusion, the work laid out here highlights two findings of the characteristics of the p73 dimer-RE complex. The first outlines a significant conformational change in the dimer-DNA complex as a consequence of the flanking nucleotide sequence, i.e. position 1 – 3. Furthermore, guanine-rich sequences seem to provide the necessary conditions for communication of the p73 dimer without the OD. This result naturally extends to p73 tetramer conformation, such that full-site response elements would be able to dictate p73 conformations based on the flanking quarter-site sequences; importantly, only one guanine-rich flanking sequence in the full-site response element would be necessary to direct the p73-RE complex to the closed conformation. The second finding shows a unique dimer conformation that was adopted by the weaker binding sequence, namely GAACA, which gave some insight into the nature of an unbound p73 dimer and

to the possible connectivity of the OD. Using this conformation, a plausible model for p73 response element and nonspecific DNA recognition was developed. Both findings have helped, in concert with the known experimental data, to deduce attributes of the p73 protein that can be extended to other little known domains of p73 and the p53-family area of research and hopefully aid in an effort for better understanding.

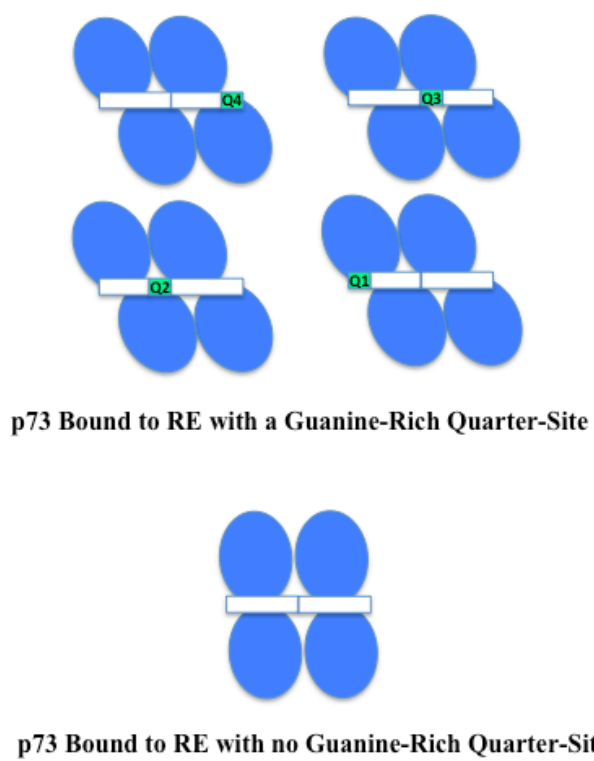
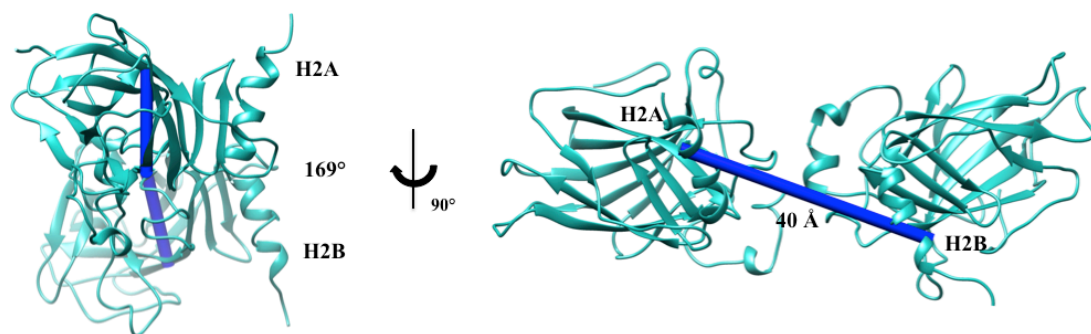
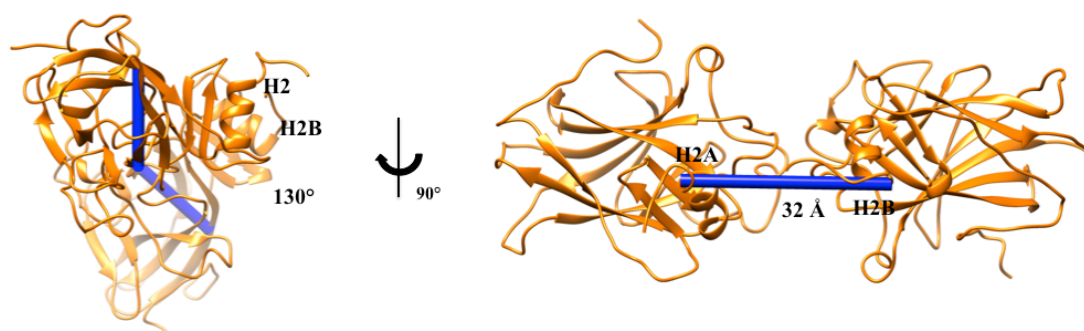


Figure 5.1. Conformational arrangement of p73 tetramer with and without guanine-rich quarter site.



Specifically Bound Dimer Conformation



Nonspecifically Bound Dimer Conformation

Figure 5.2. Dimer and Helix H2 Rotation between Specifically and Nonspecifically bound Dimers. Shown left, are the dimer rotation angle for each respective conformation. Shown right, are the distances between helix H2 of monomers A and B. H2 helix-helix distances are measured using Glu-306 from both monomers.

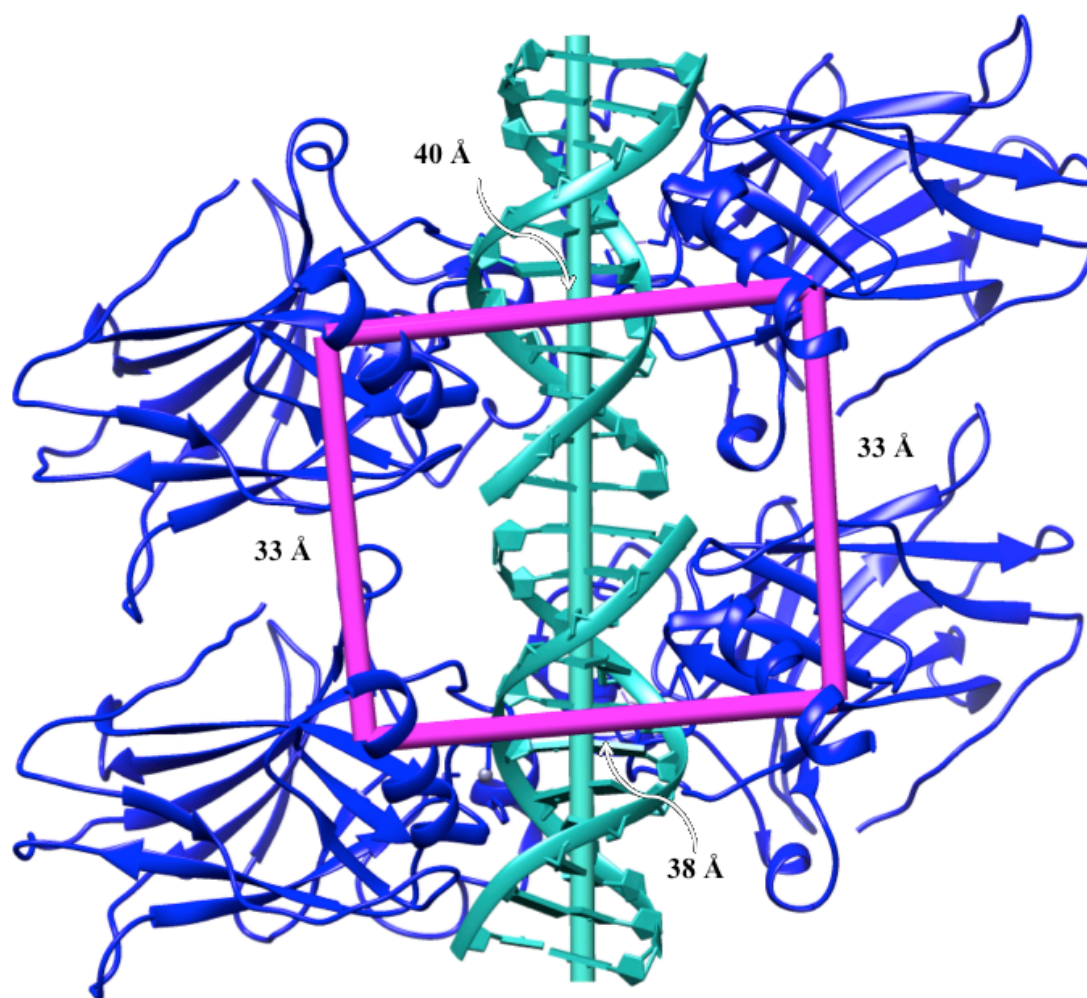
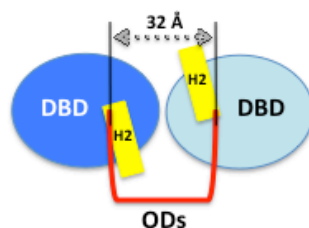
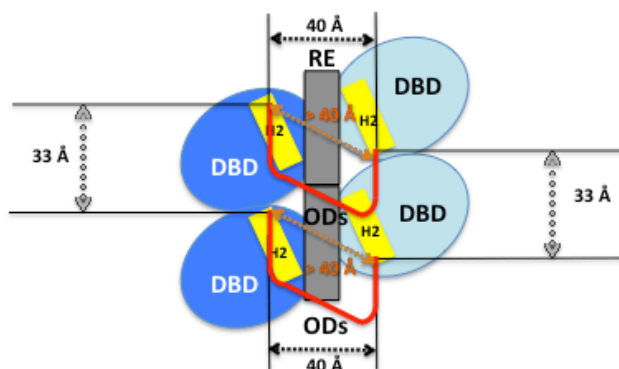


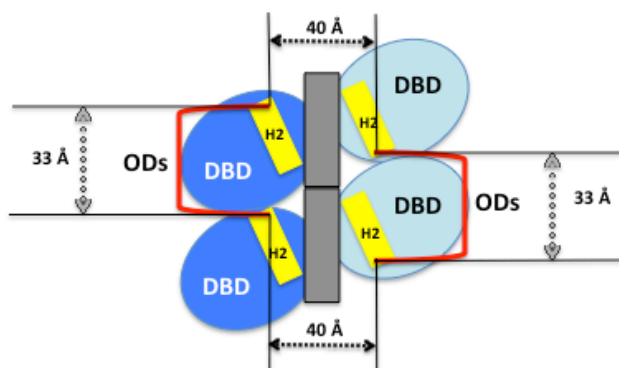
Figure 5.3. Distances between helix H2s of a crystal structure of p73-DBD Tetramer Bound to a 0-Spacer Full-Site Response Element. H2 helix-helix distances are measured using Glu-306 from both monomers.



p73 Dimer in Absence of RE



p73 Tetramer Conformation with intra-Dimer OD Connectivity



p73 Tetramer Conformation with Dimer-Dimer OD Connectivity

Figure 5.4. Possible p73 arrangements of OD in unbound and bound states. Top, OD connects from helix H2 of one monomer to helix H2 of other monomer (value shown is the distance measured between Glu-306 of helix H2 monomers A and B taken from a snapshot of the copy 1 GAACA_4g82 system). Middle, p73 tetramer conformation is shown with intra-dimer OD connectivity. Bottom, p73 tetramer conformation is shown with dimer-dimer OD connectivity. Distance values shown for tetramer arrangements are from p73 tetramer crystal structure(0-spacer PDB id: 3vd0) measurements between Glu-306 of helix H2.

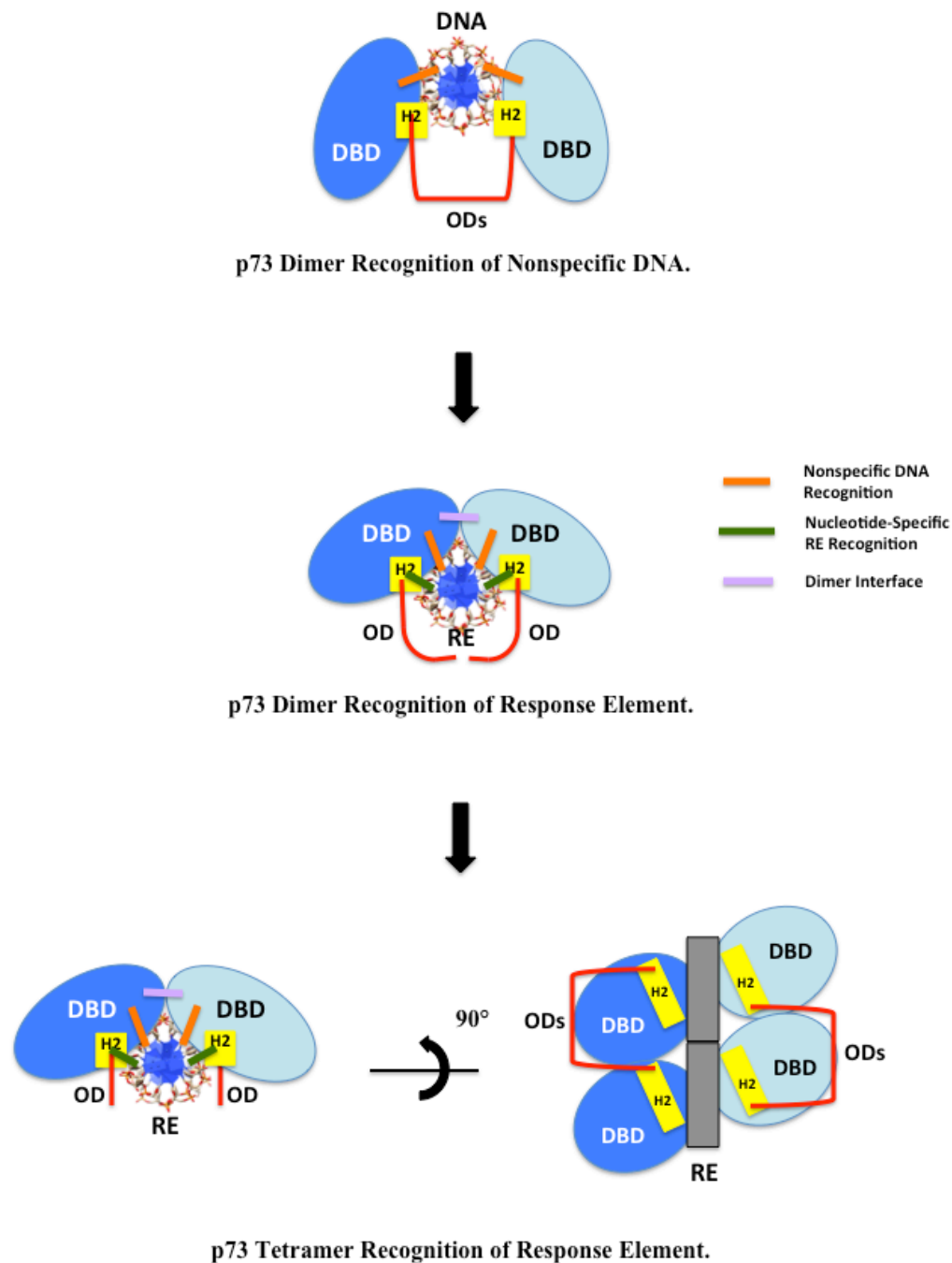
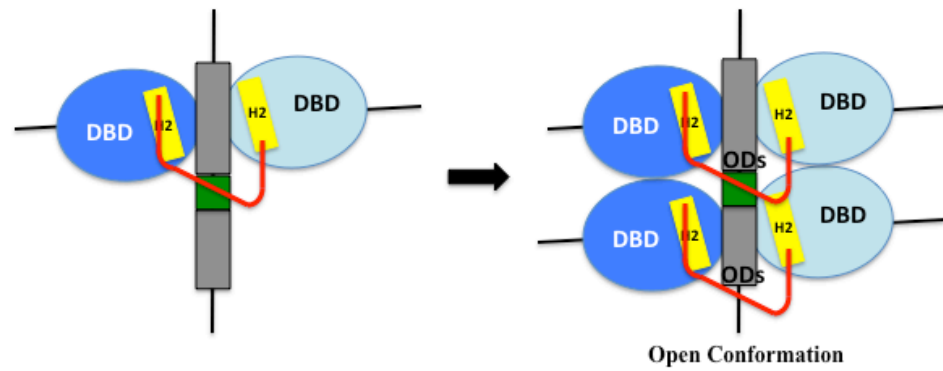
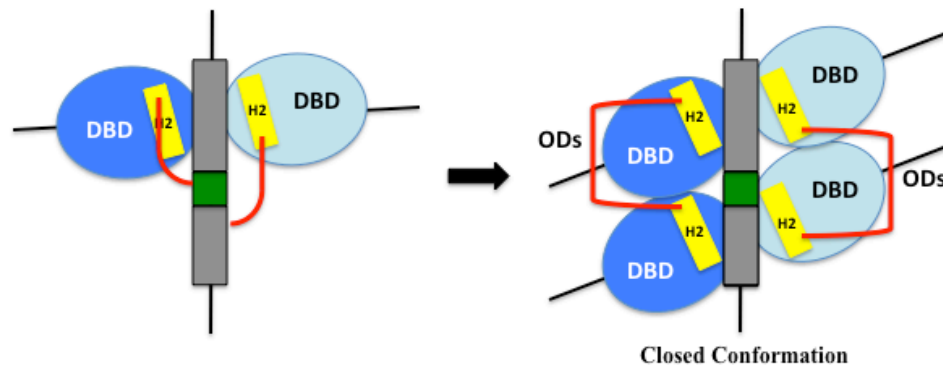


Figure 5.5. Possible Mechanism for p73 Recognition of Nonspecific DNA and Response Element. Top, p73 dimer recognizes nonspecific DNA by phosphate backbone. Middle, p73 dimer recognizes response element by nonspecific and nucleotide-specific interactions. It is also able to contact between monomers by forming a dimer interface. Bottom, p73 tetramer recognizes response element similarly to the dimer. Additionally, OD connectivity changes from contacting within a dimer to across dimers of the adjacent side to response element.



Static Model of p73 RE Recognition



Dynamic Model of p73 RE Recognition

Figure 5.6. P73 RE Recognition Models Based on OD Connectivity. Top, static model retains monomer-monomer OD interaction of the first dimer RE half-site recognition. Dimer adopts conformation sensed by nucleotide sequence. Empty half-site is recognized by second dimer that adopts the same conformation as the previously bound dimer. Bottom, dynamic model loses OD interaction within the dimer after RE recognition. Conformation of dimer-half-site complex is transient. Binding of the second dimer allows OD interaction across dimer for adjacent monomers on the same side of response element. Tetramer conformation is informed by the sequence of all four quarter-sites.

REFERENCES

- Ethayathulla, A. S., Tse, P. W., Monti, P., Nguyen, S., Inga, A., Fronza, G., & Viadiu, H. (2012). Structure of p73 DNA-binding domain tetramer modulates p73 transactivation. *Proc Natl Acad Sci U S A*, *109*(16), 6066-6071. doi: 10.1073/pnas.1115463109
- Pham, N., Lucumi, A., Cheung, N., & Viadiu, H. (2012). The tetramer of p53 in the absence of DNA forms a relaxed quaternary state. *Biochemistry*, *51*(41), 8053-8055. doi: 10.1021/bi301193k
- Ramos, A. G. (2014). PhD Dissertation, University of California-San Diego.
- Wang, B., Xiao, Z., & Ren, E. C. (2009). Redefining the p53 response element. *Proc Natl Acad Sci U S A*, *106*(34), 14373-14378. doi: 10.1073/pnas.0903284106

APPENDIX

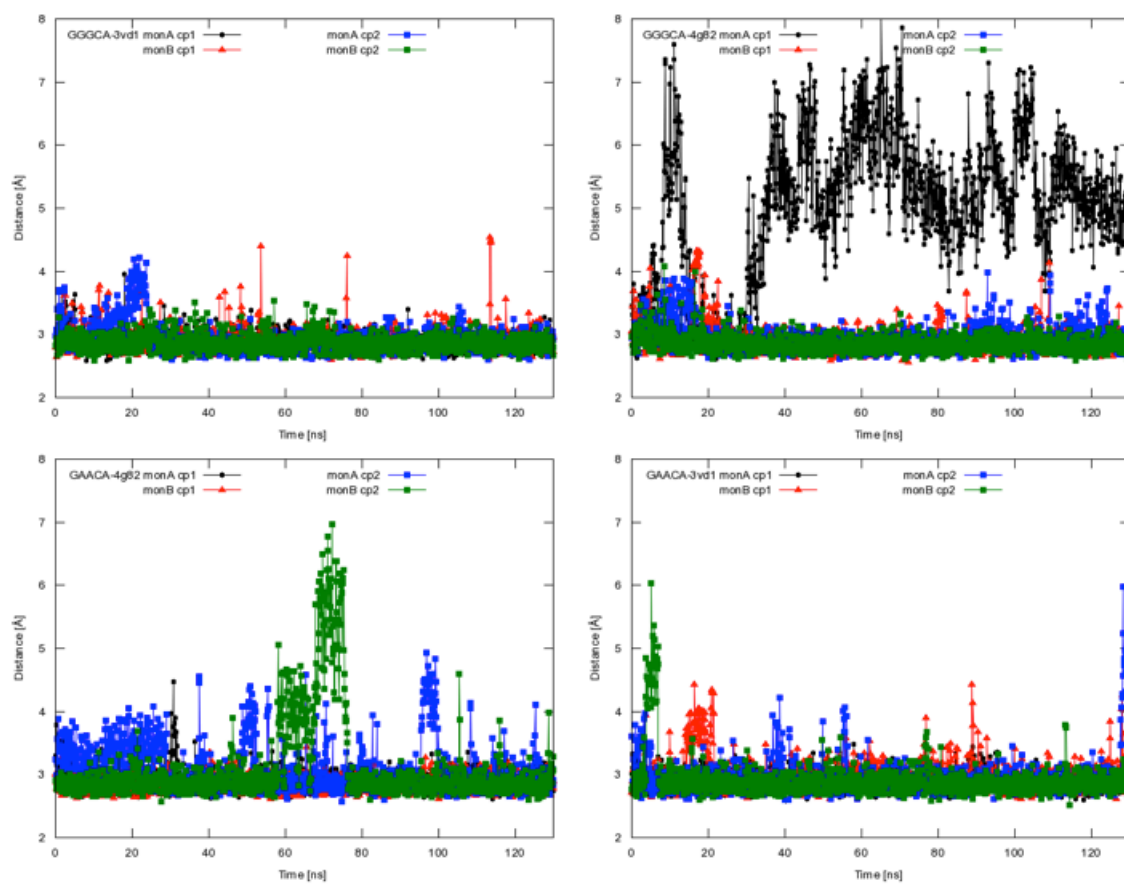


Figure A.1. Distance between Ala-296 and position-4 phosphate. Distance monitored from backbone nitrogen, N, of Ala-296 to oxygen, O2P, of phosphate 4.

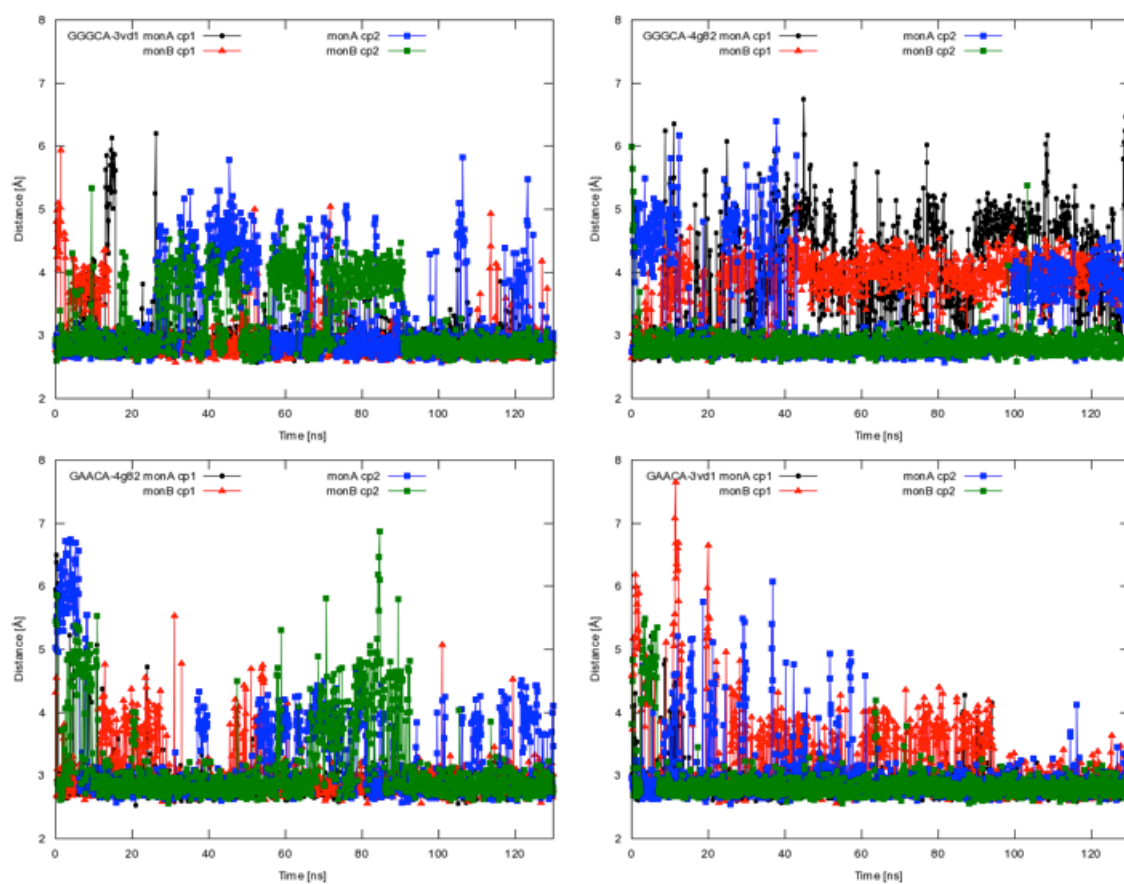


Figure A.2. Distance between Asn-259 and position-4 phosphate. Distance monitored from nitrogen, ND2, of Asn-259 to oxygen, O1P, of phosphate 4.

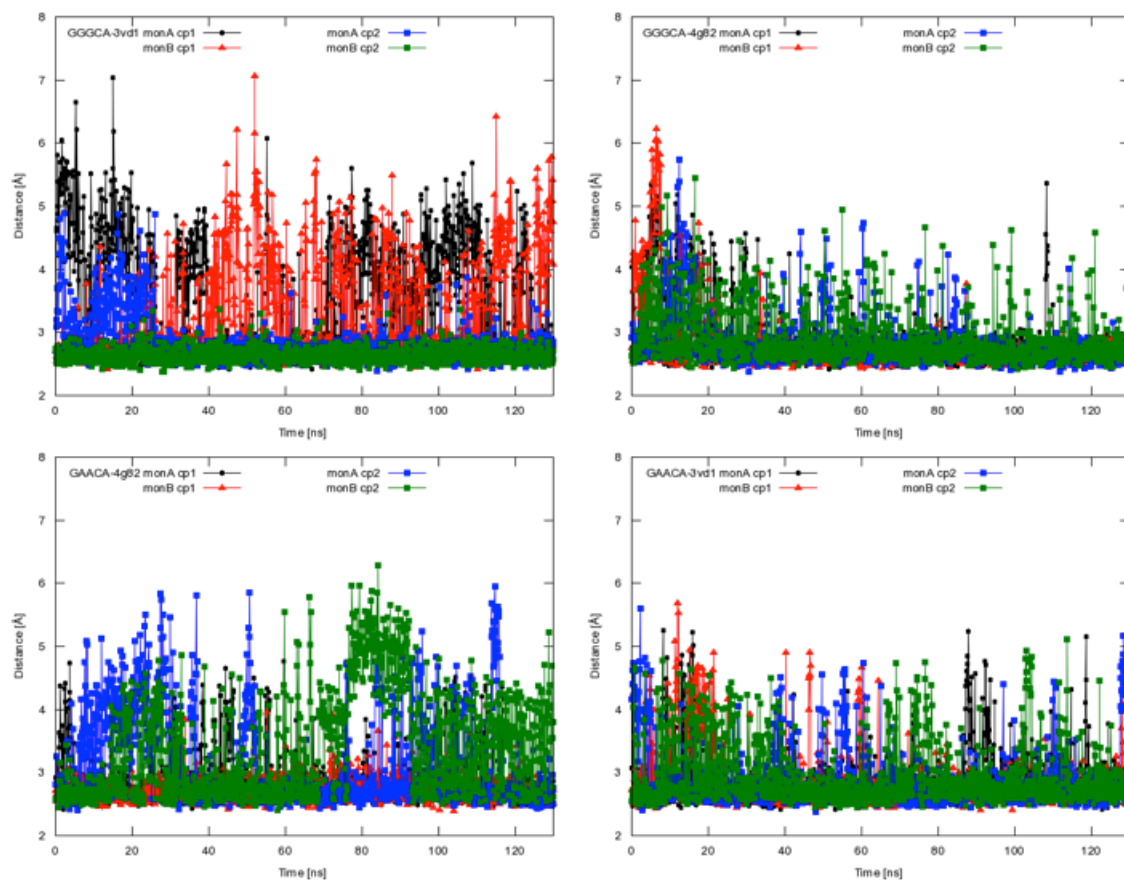


Figure A.3. Distance between Ser-261 and position-4 phosphate. Distance monitored from oxygen, OG, of Ser-162 to oxygen, O1P, of phosphate 4.

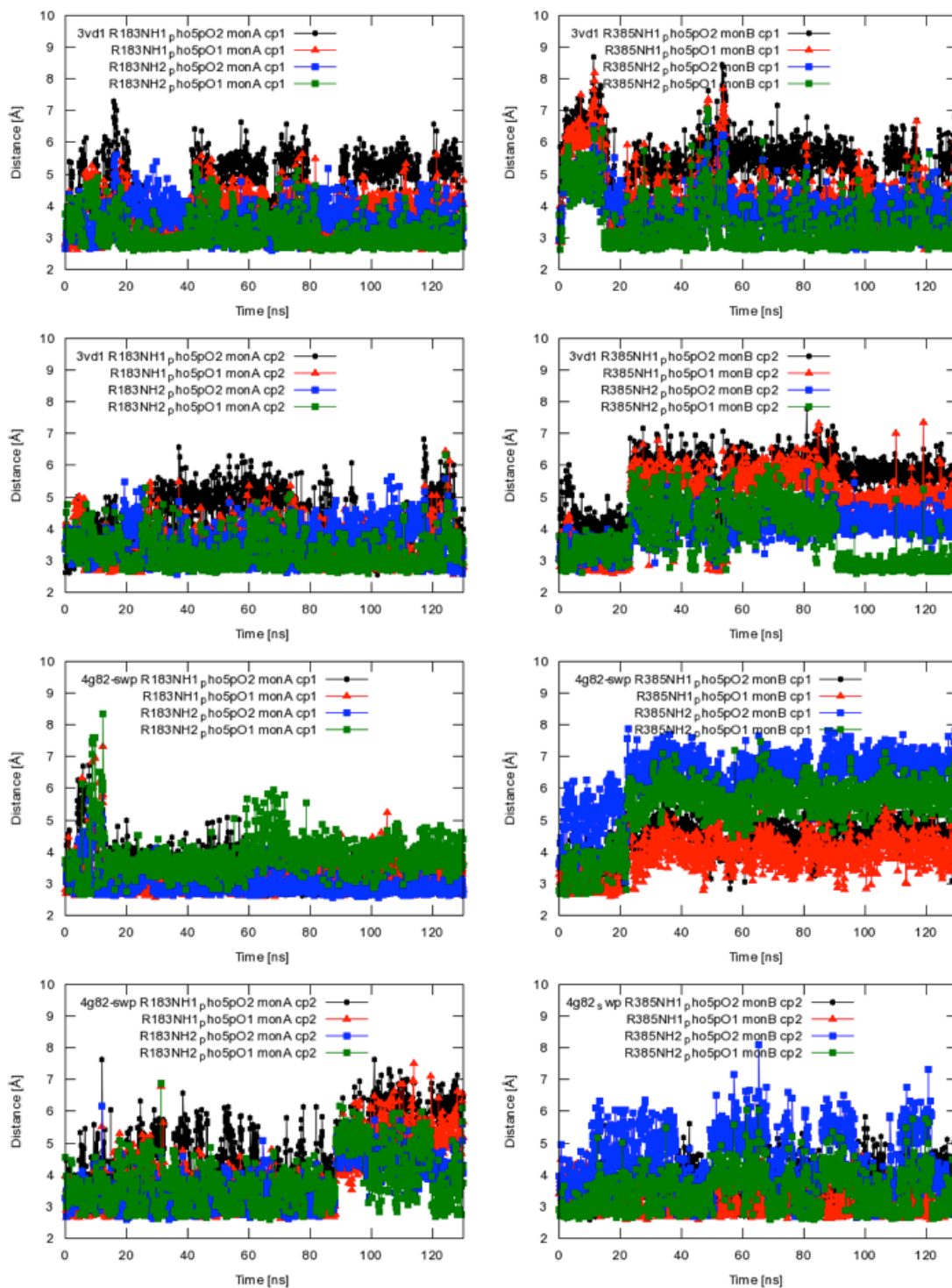


Figure A.4. Distance between Arg-293 and position-5 phosphate of GGGCA_3vd1 and GGGCA_4g82_swap systems. Distance monitored of four contacts: NH1-O2, NH1-O1, NH2-O2, NH2-O1.

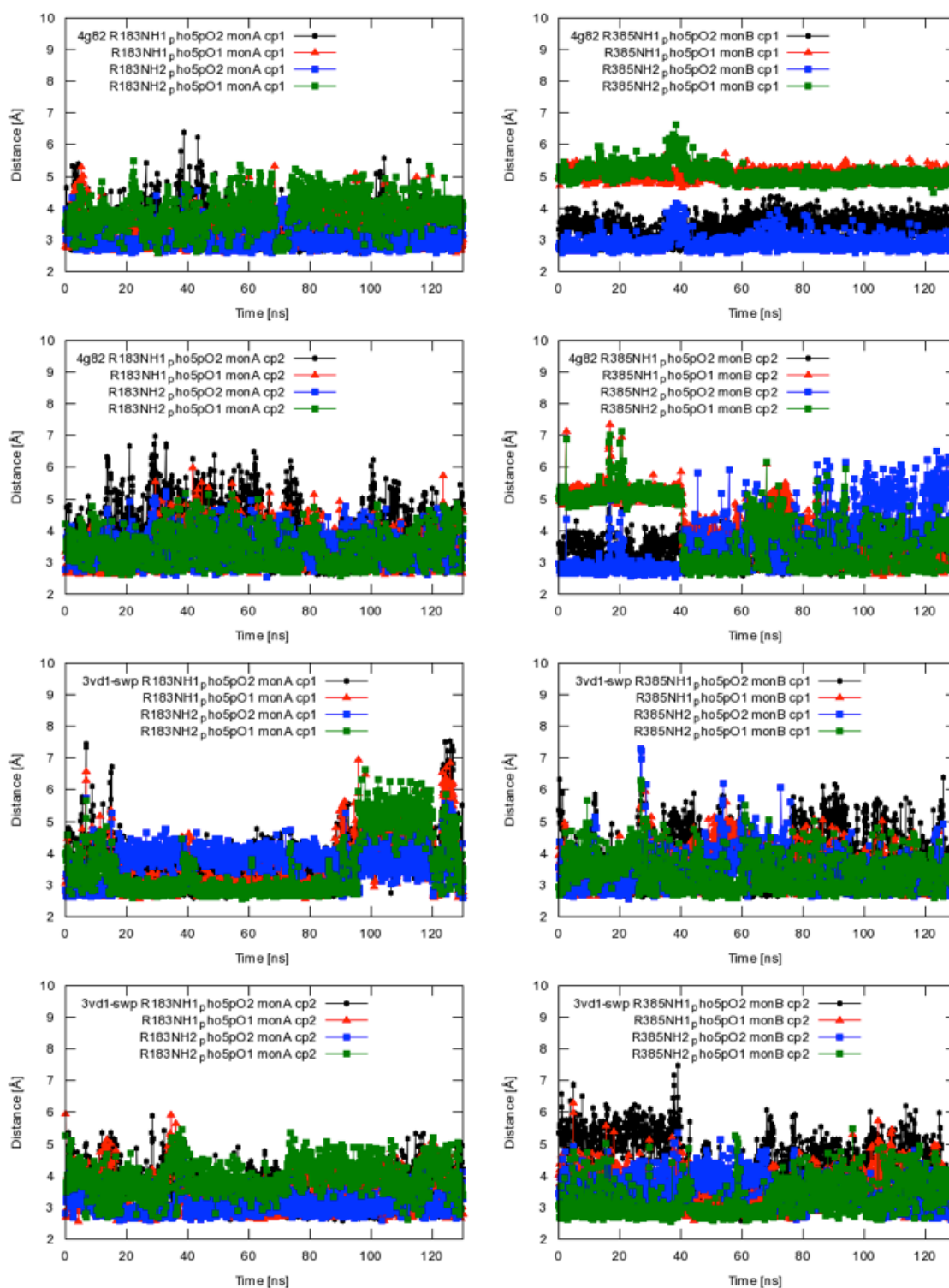


Figure A.5. Distance between Arg-293 and position-5 phosphate of GAACA_4g82 and GAACA_3vd1_swap systems. Distance monitored of four contacts: NH1-O2, NH1-O1, NH2-O2, NH2-O1.

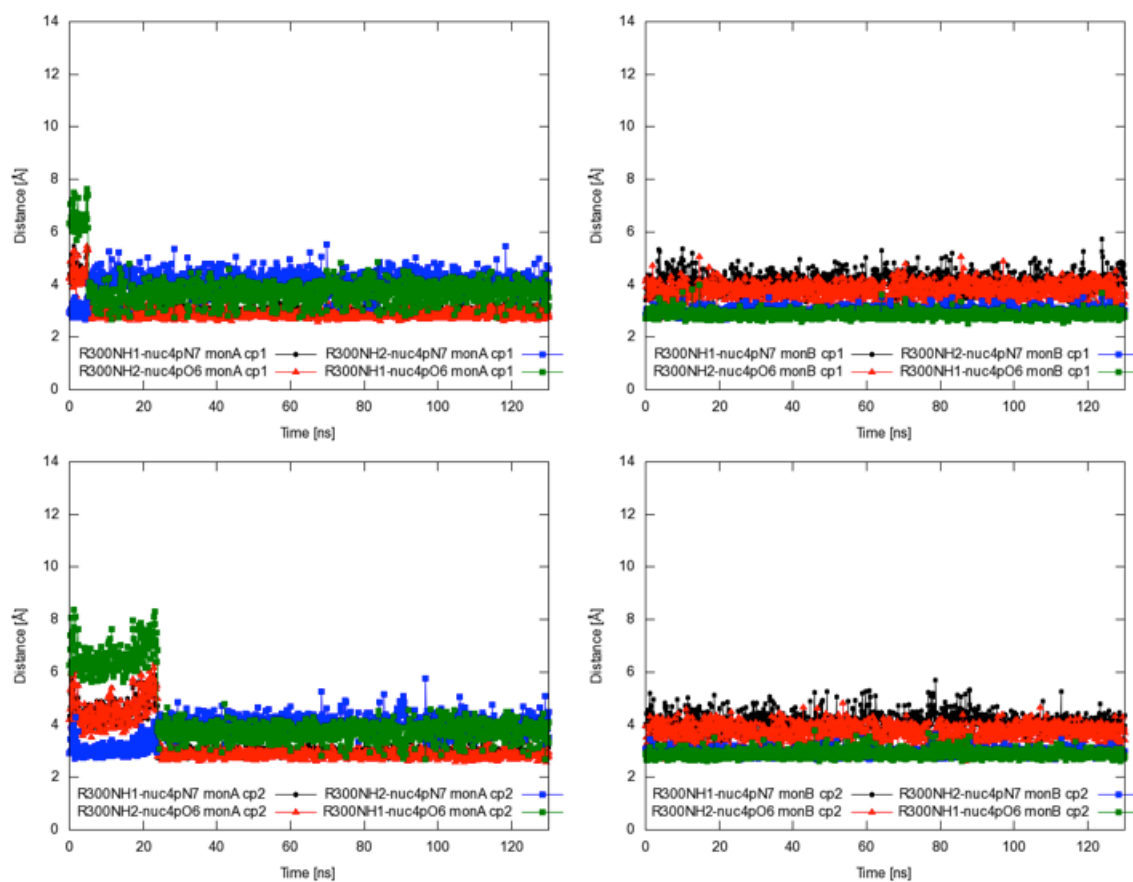


Figure A.6. Distance between Arg-300 and guanine complimentary to position four cystosine for GGGCA_3vd1 system. Each contact is identified by position and type of contact, the electron donor atom, the monomer, and copy (i.e. R300NH1-nuc4pN7 of monA cp1).

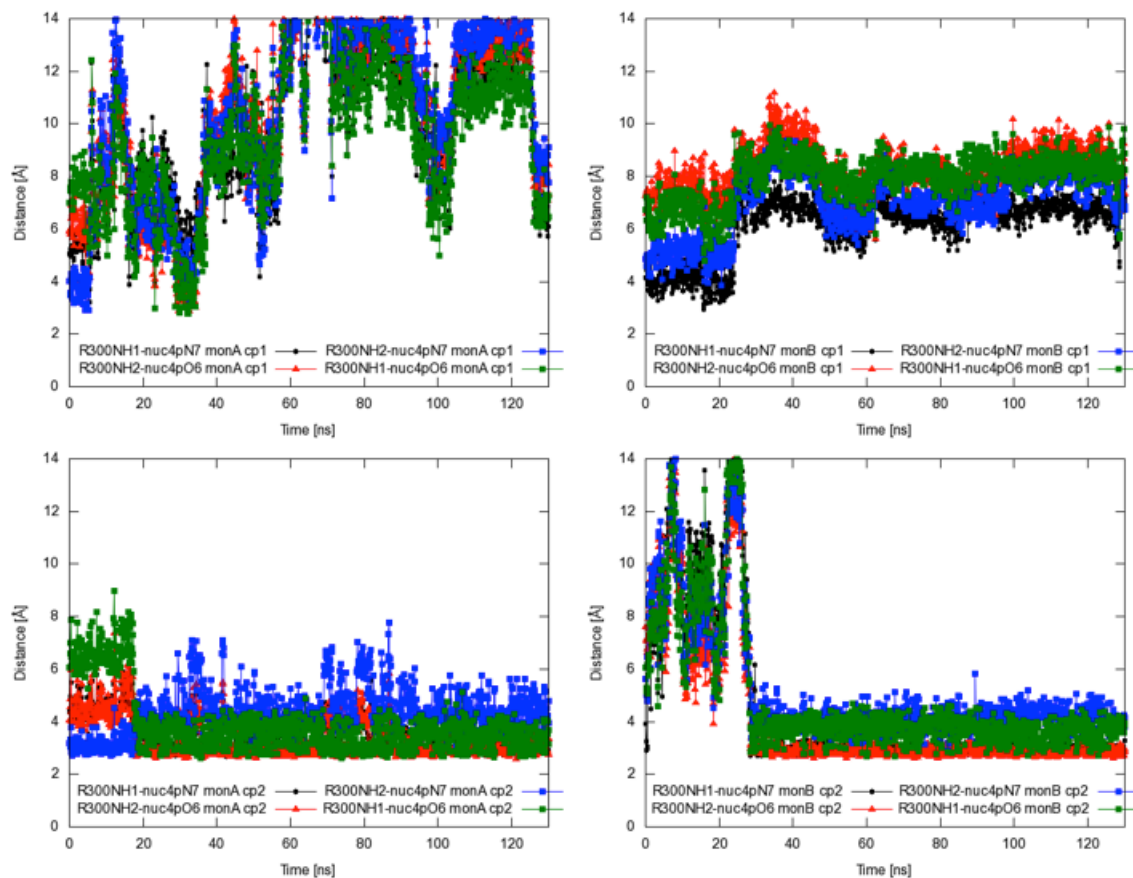


Figure A.7. Distance between Arg-300 and guanine complimentary to position four cystosine for GGGCA_4g82_swap system. Each contact is identified by position and type of contact, the electron donor atom, the monomer, and copy (i.e. R300NH1-nuc4pN7 of monA cp1).

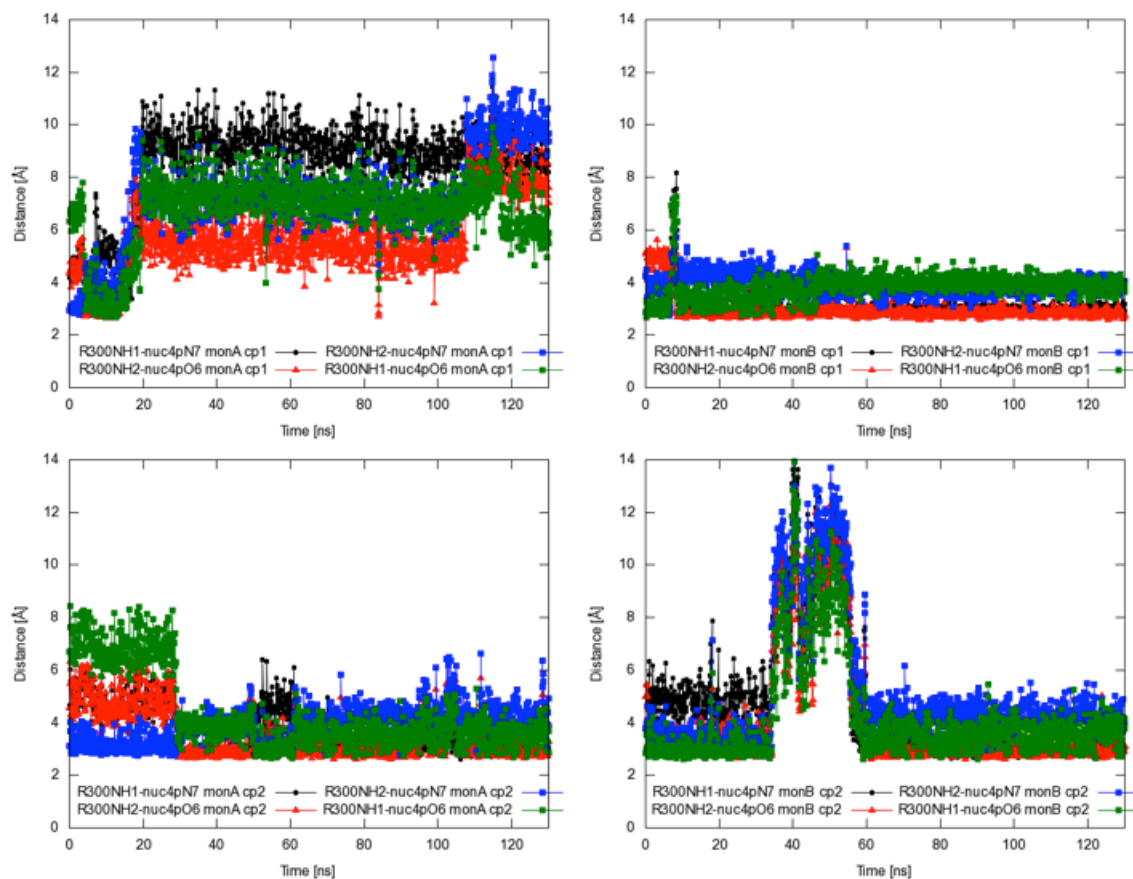


Figure A.8. Distance between Arg-300 and guanine complimentary to position four cystosine for GAACA_4g82 system. Each contact is identified by position and type of contact, the electron donor atom, the monomer, and copy(i.e. R300NH1-nuc4pN7 of monA cp1).

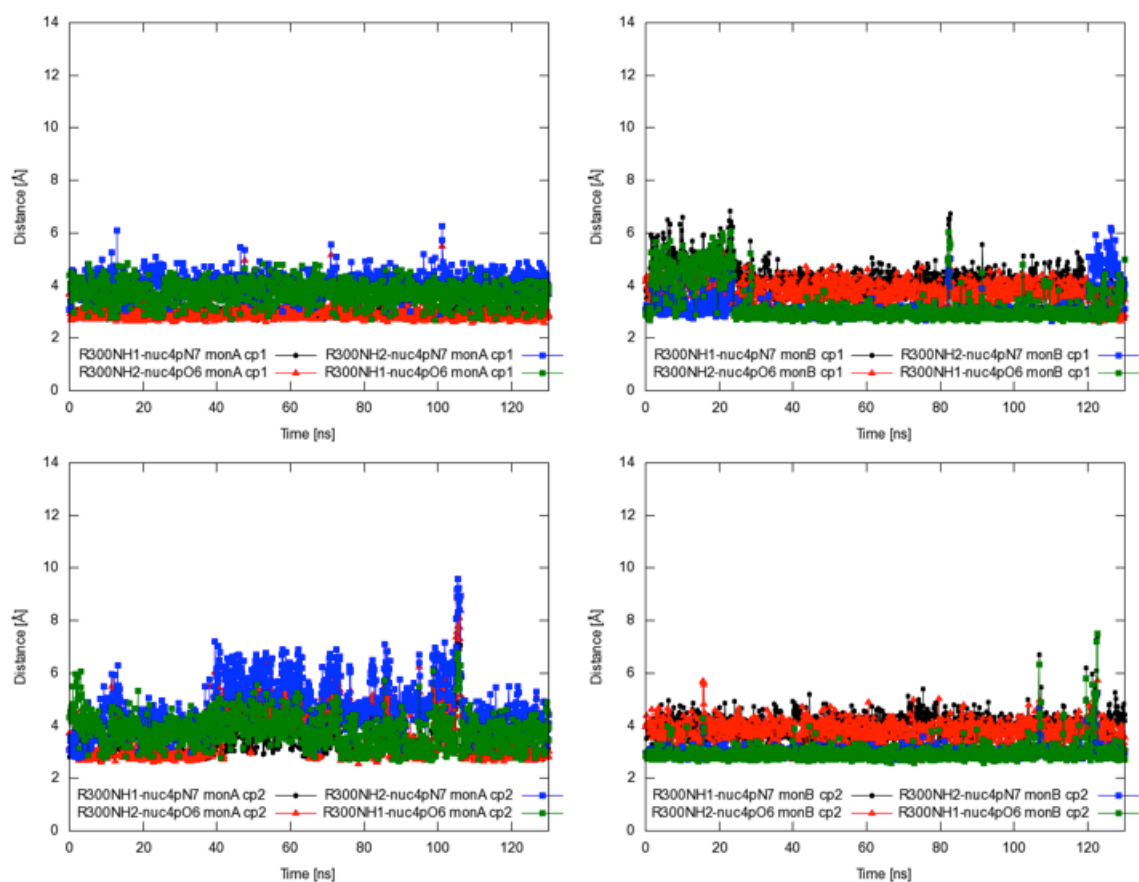


Figure A.9. Distance between Arg-300 and guanine complimentary to position four cystosine for GAACA_3vd1_swap system. Each contact is identified by position and type of contact, the electron donor atom, the monomer, and copy (i.e. R300NH1-nuc4pN7 of monA cp1).

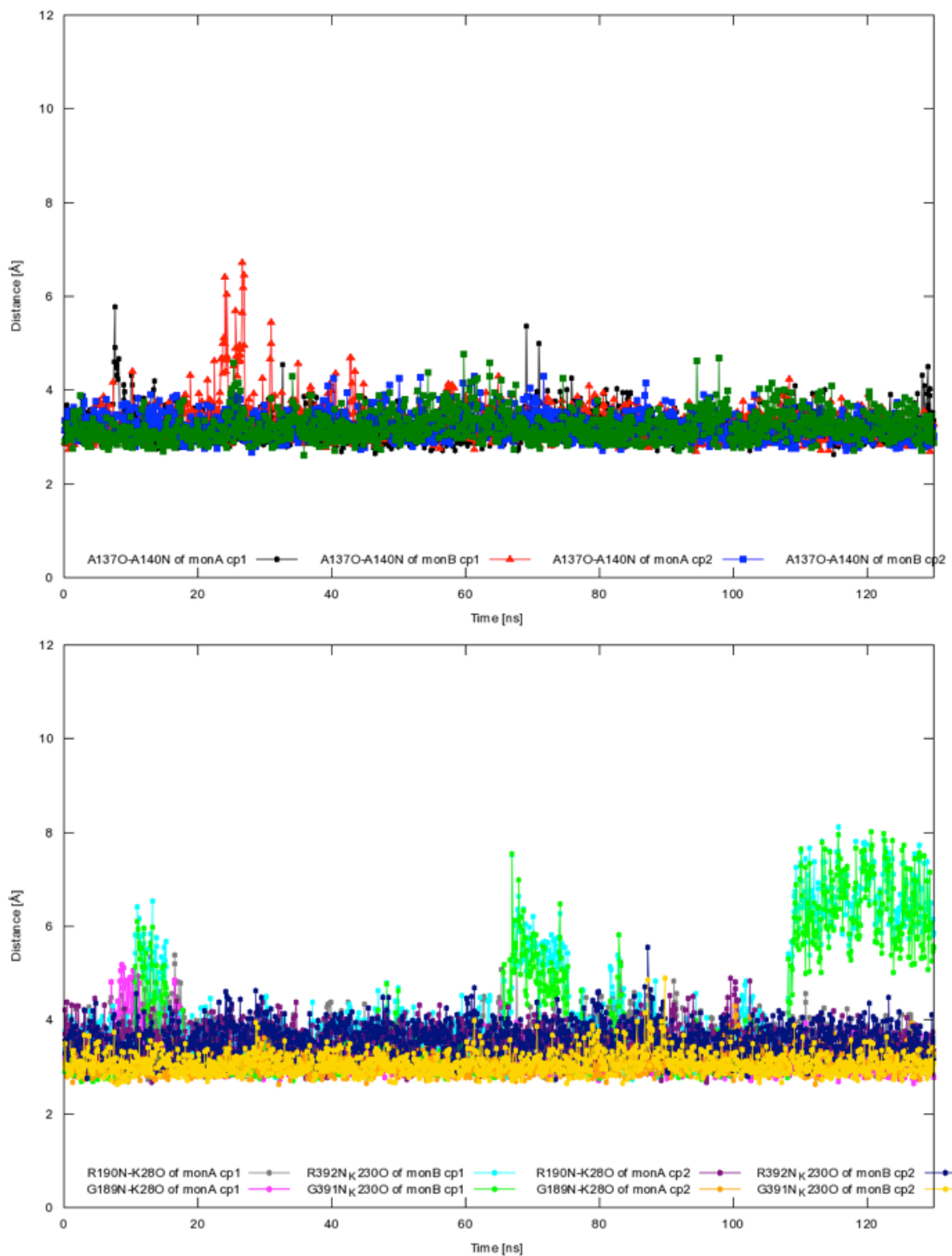


Figure A.10. Distance between hydrogen bond partners of contacts concerning intraloop L1(top), and between loop L1 and helix H2(bottom) for GGGCA_3vd1 system. The intraloop L1 contact is monitored for the four monomers corresponding from both simulation copies. The contacts between loop L1 and helix H2 are monitored for the two contacts(Arg-300 and Lys-138, and Gly-299 and Lys-138) per monomer(all four monomers reported).

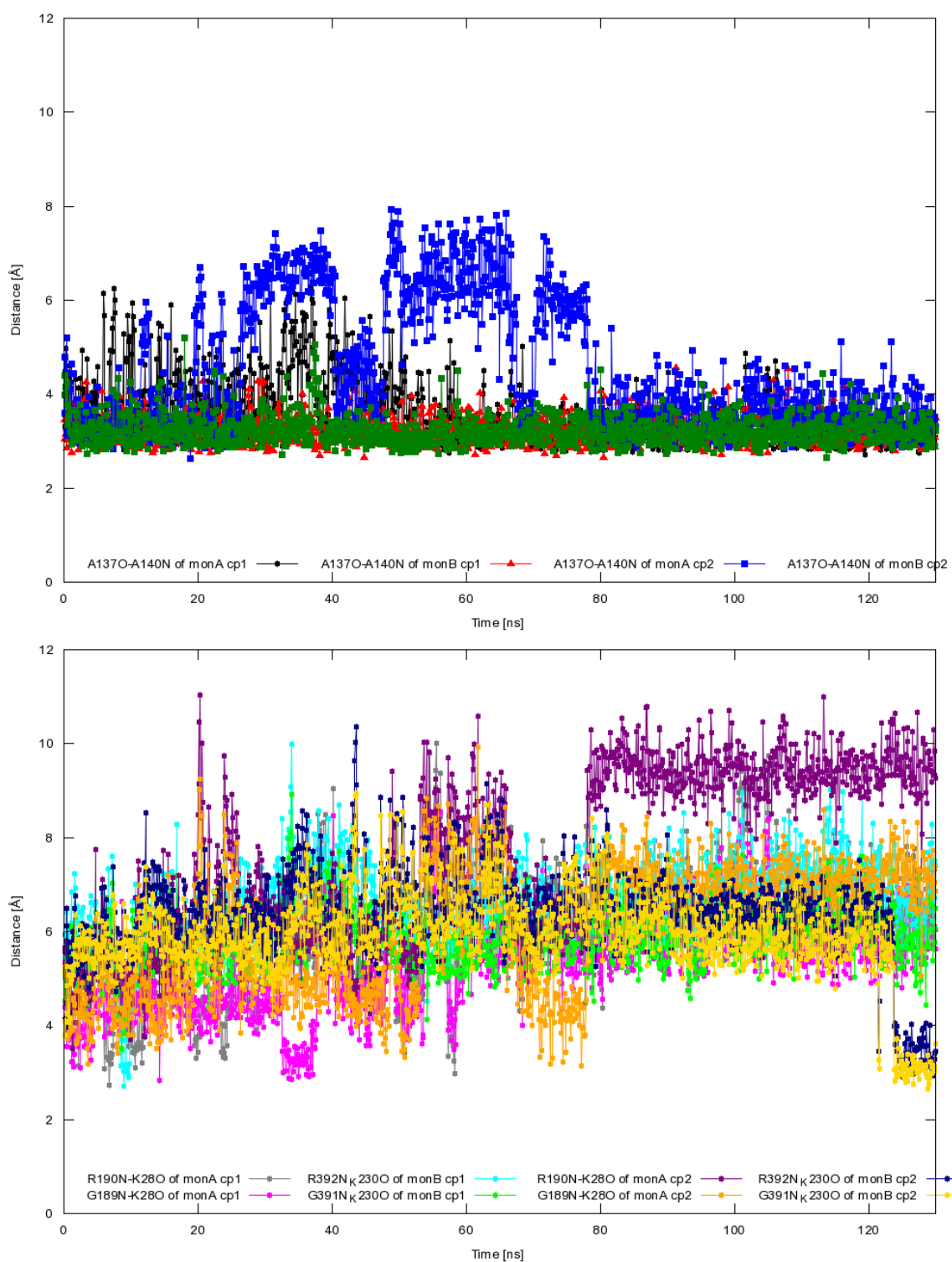


Figure A.11. Distance between hydrogen bond partners of contacts concerning intraloop L1(top), and between loop L1 and helix H2(bottom) for GGGCA_4g82_swap system.

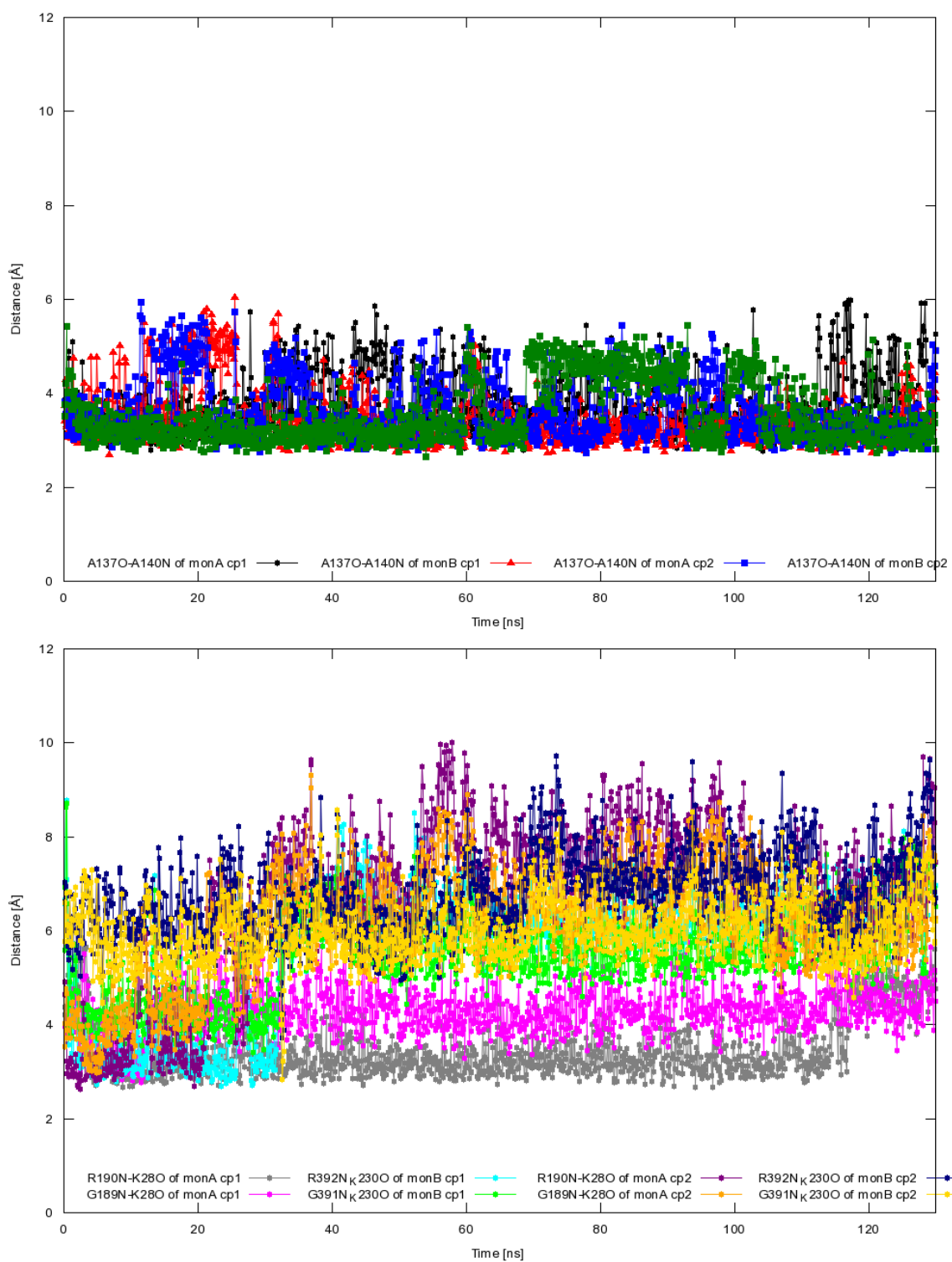


Figure A.12. Distance between hydrogen bond partners of contacts concerning intraloop L1(top), and between loop L1 and helix H2(bottom) for GAACA_4g82 system.

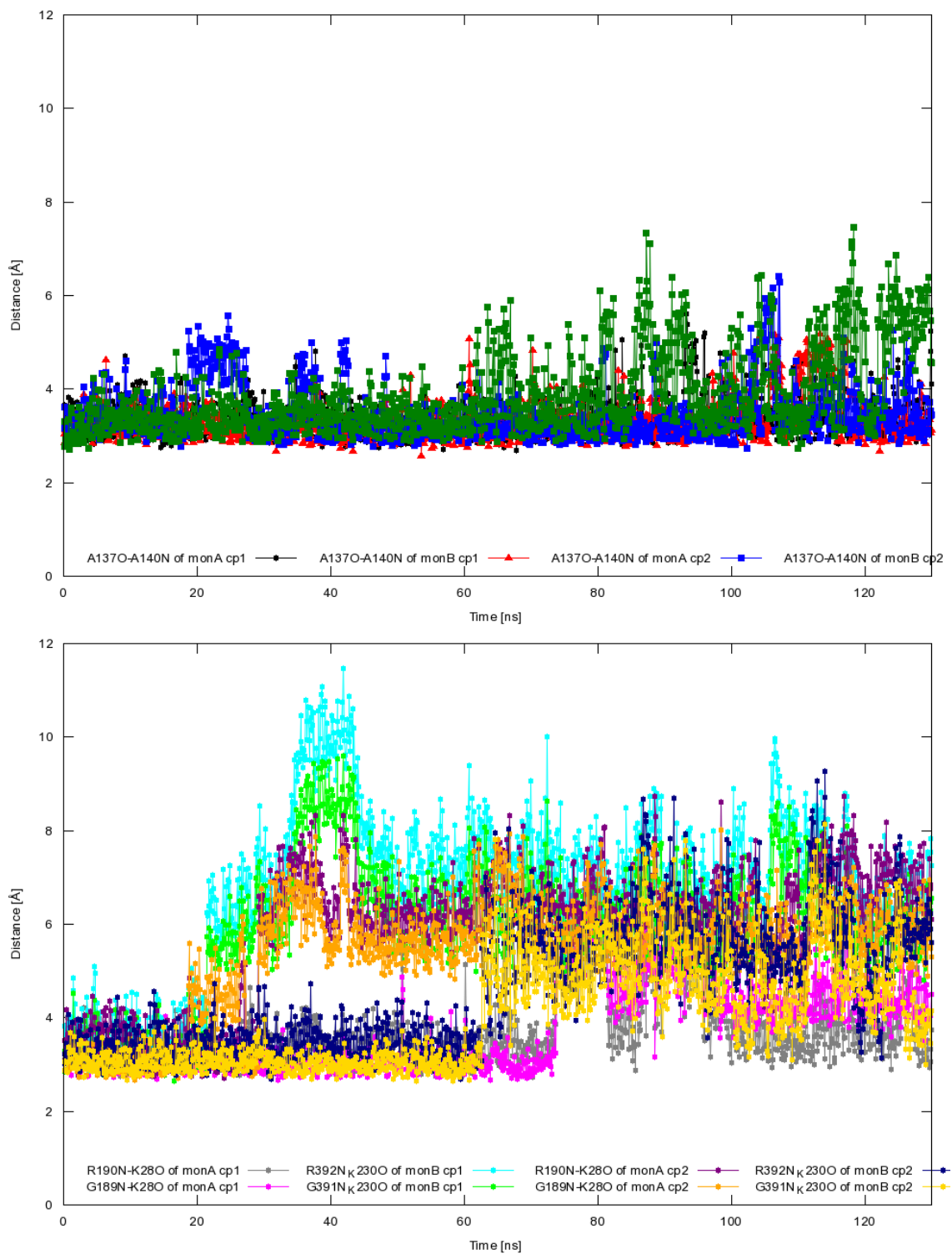


Figure A.13. Distance between hydrogen bond partners of contacts concerning intraloop L1(top), and between loop L1 and helix H2(bottom) for GAACA_3vd1_swap system.

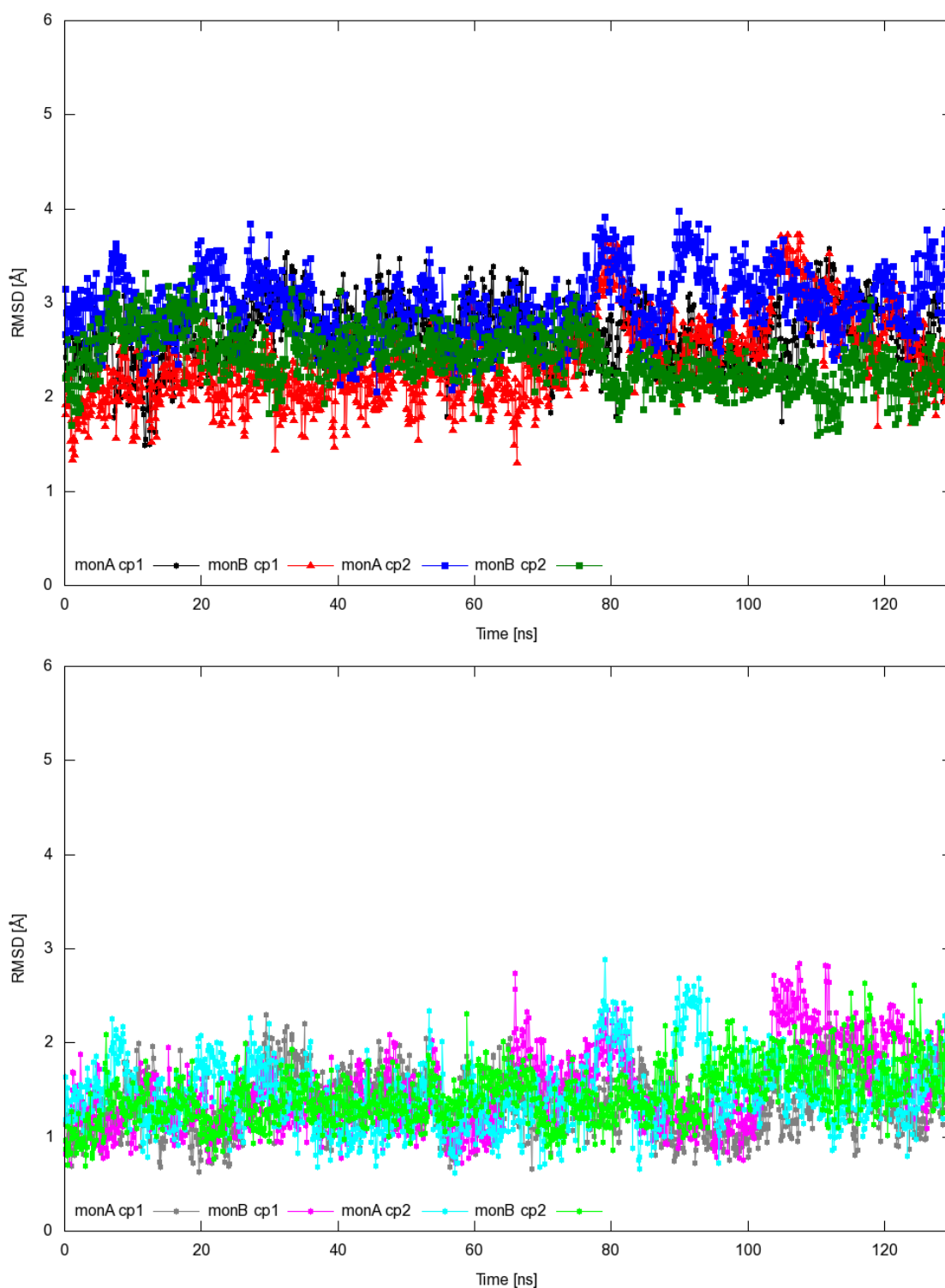


Figure A.14. RMSD of loop L1 with respect GAACA(top) and GGGCA(bottom) crystal structures for the GGGCA_3vd1 system. Corresponding monomers were aligned and RMSD of loop L1 was measured.

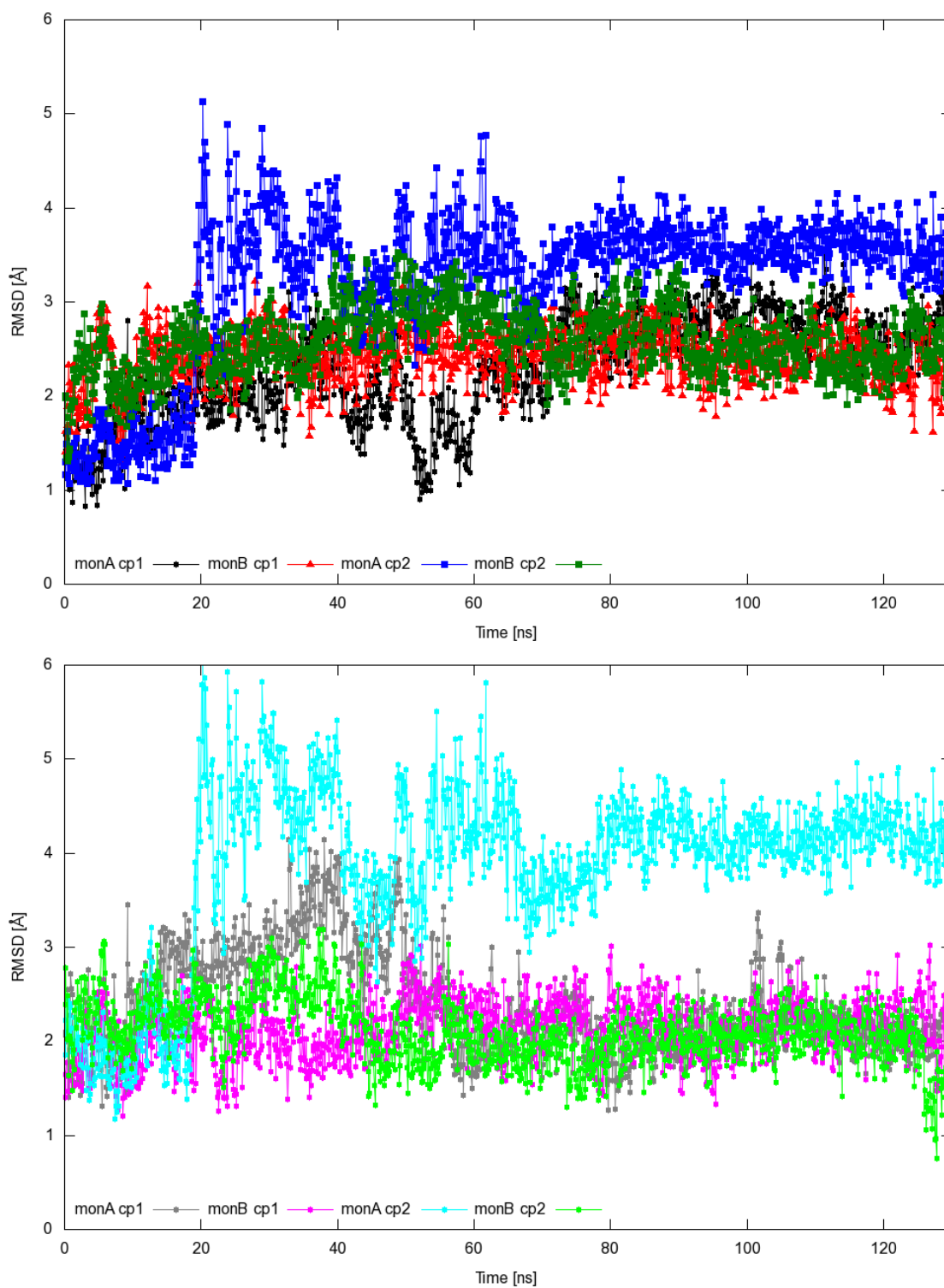


Figure A.15. RMSD of loop L1 with respect GAACA(top) and GGGCA(bottom) crystal structures for the GGGCA_3vd1 system.

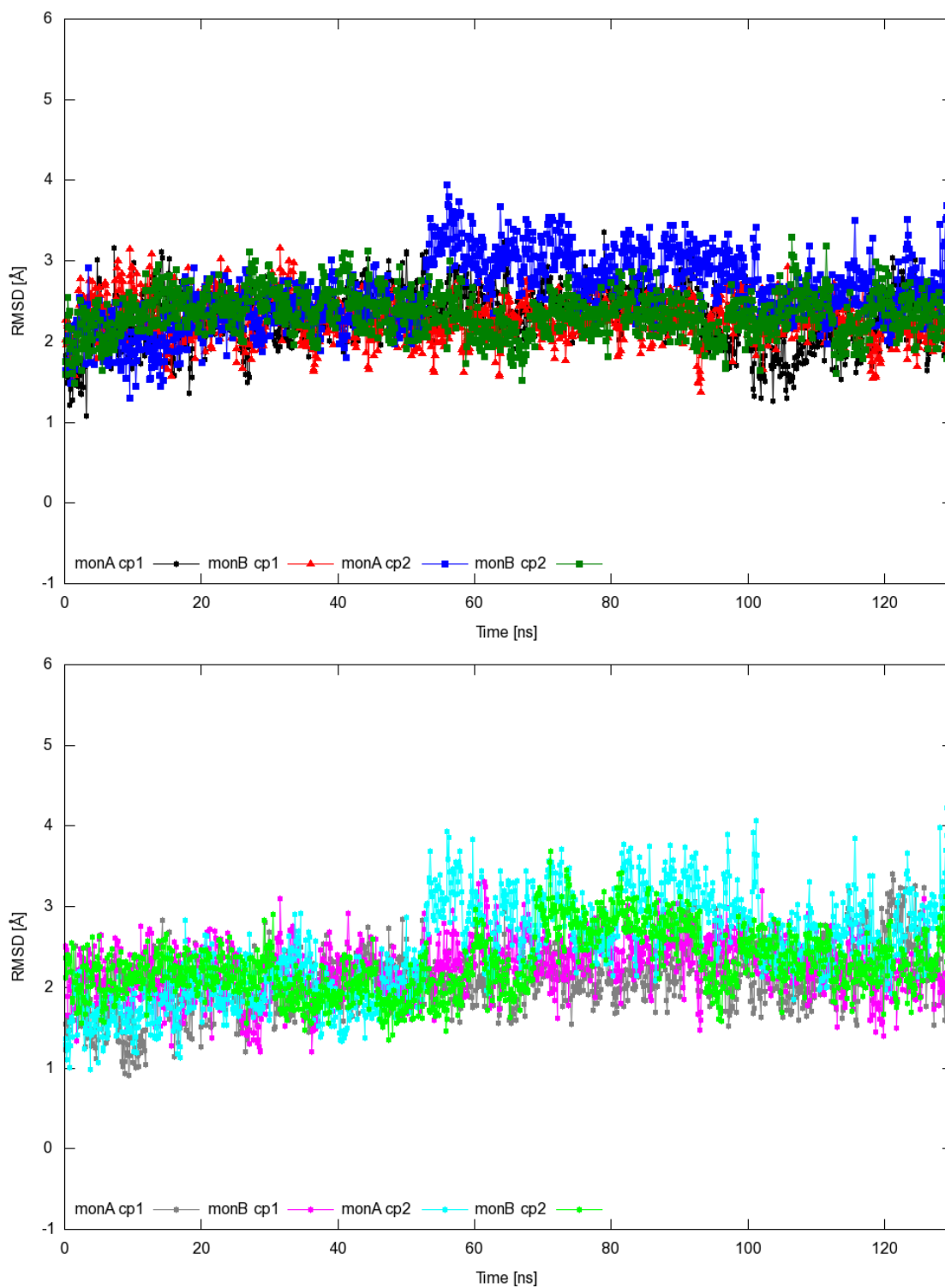


Figure A.16. RMSD of loop L1 with respect GAACA(top) and GGGCA(bottom) crystal structures for the GAACA_4g82 system.

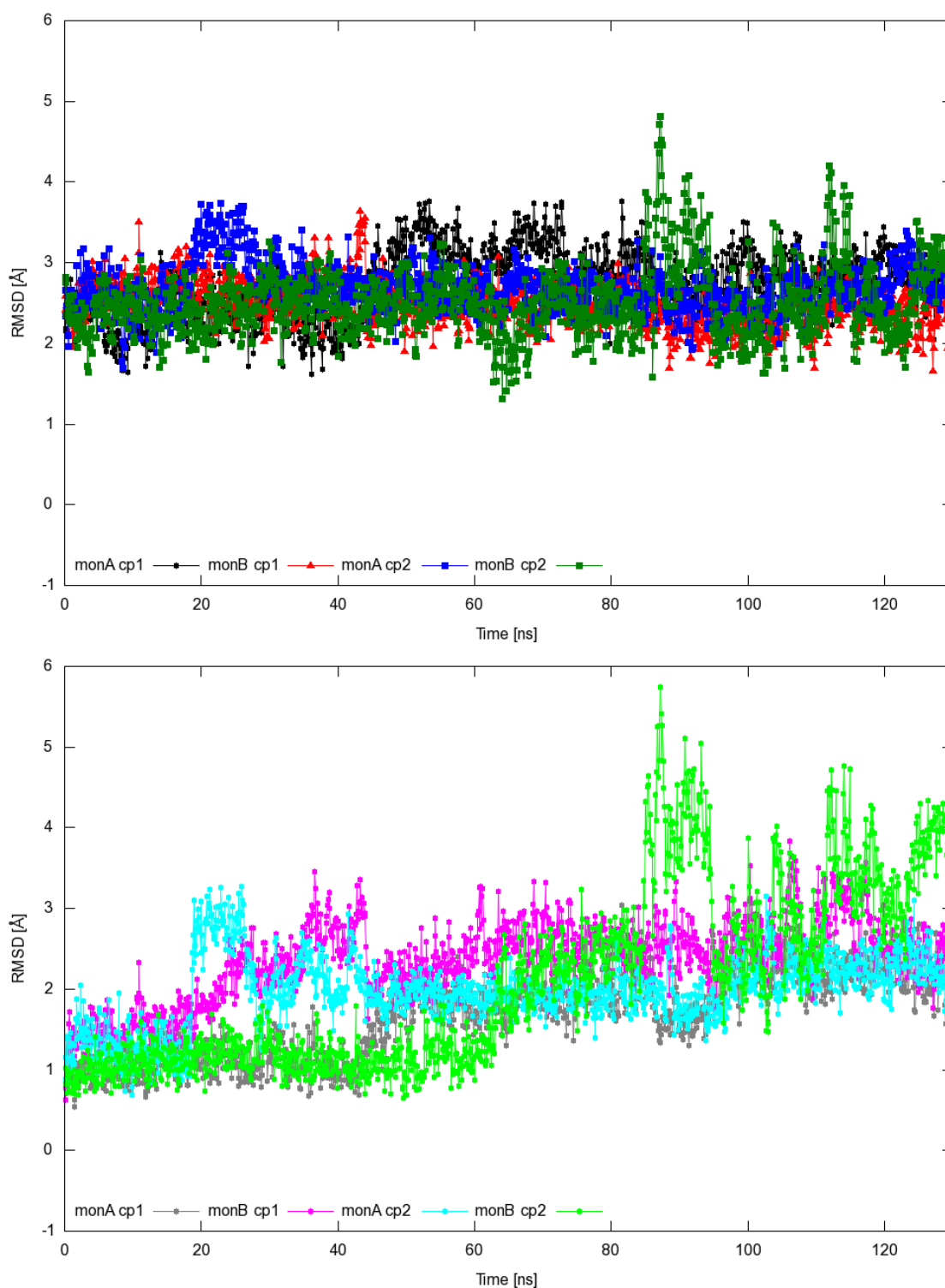


Figure A.17. RMSD of loop L1 with respect GAACA(top) and GGGCA(bottom) crystal structures for the GAACA_3vd1_swap system.

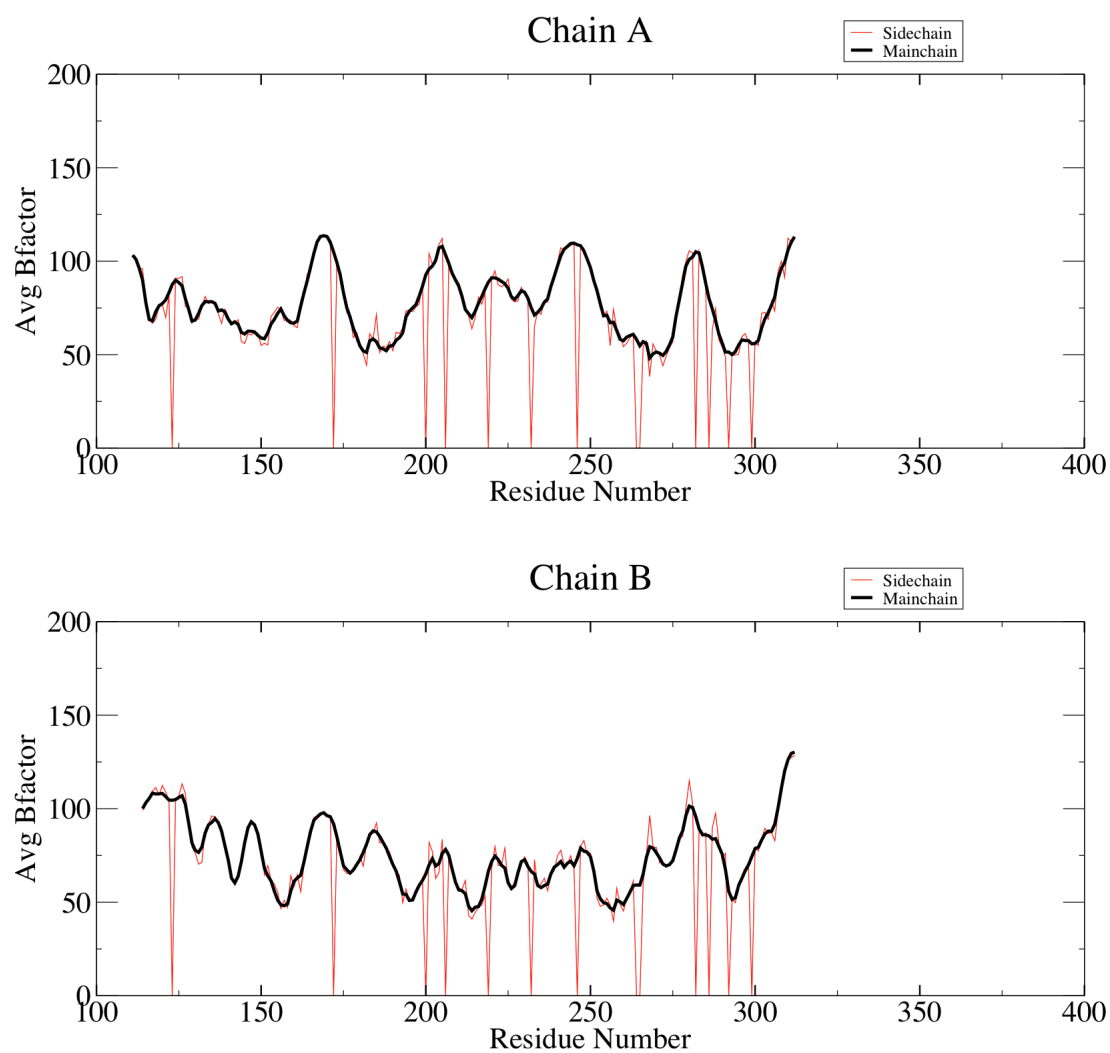


Figure A.18. B-factor of monomers A and B of GGGCA(3vd1) crystal structure.

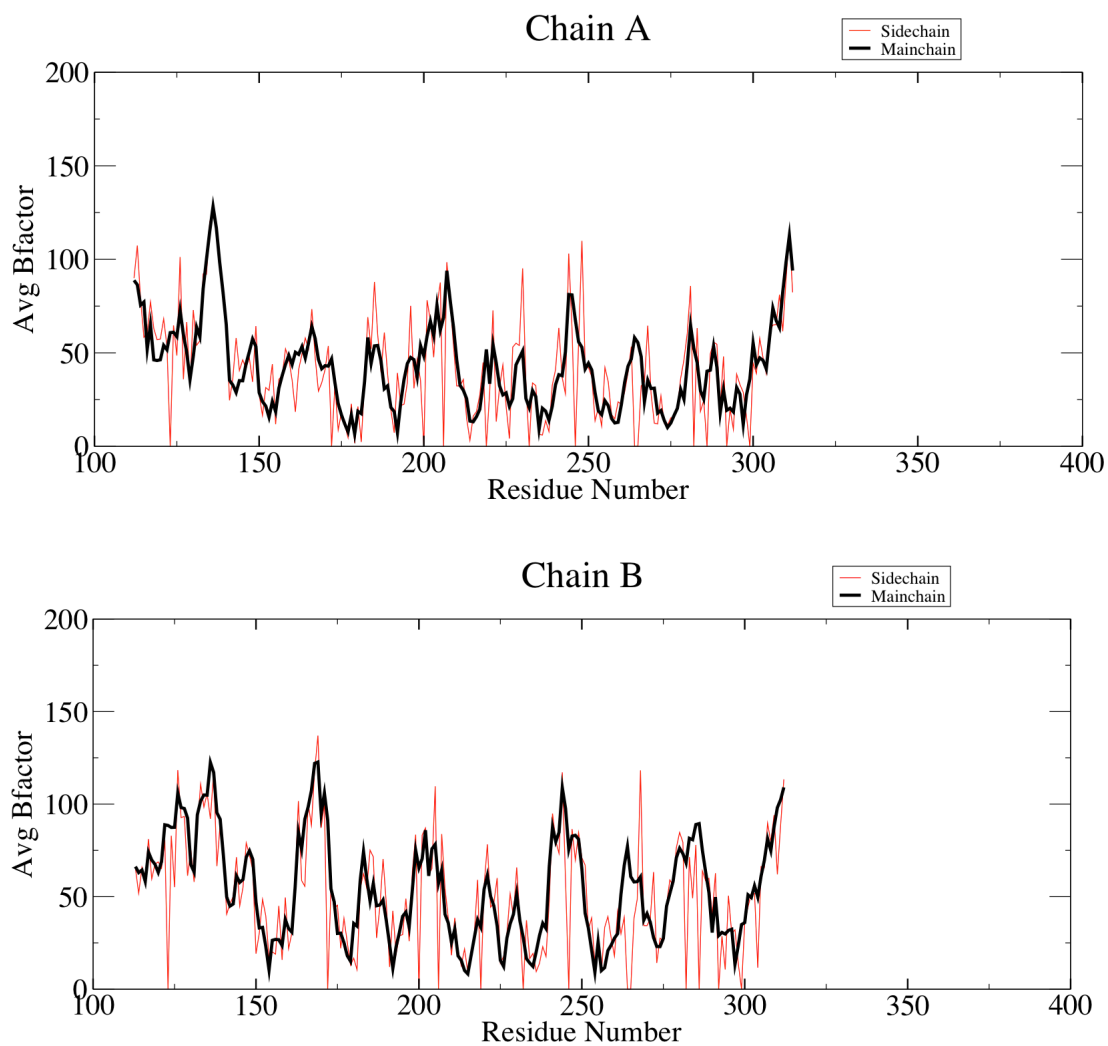


Figure A.19. B-factor of monomers A and B of GAACA(4g82) crystal structure.

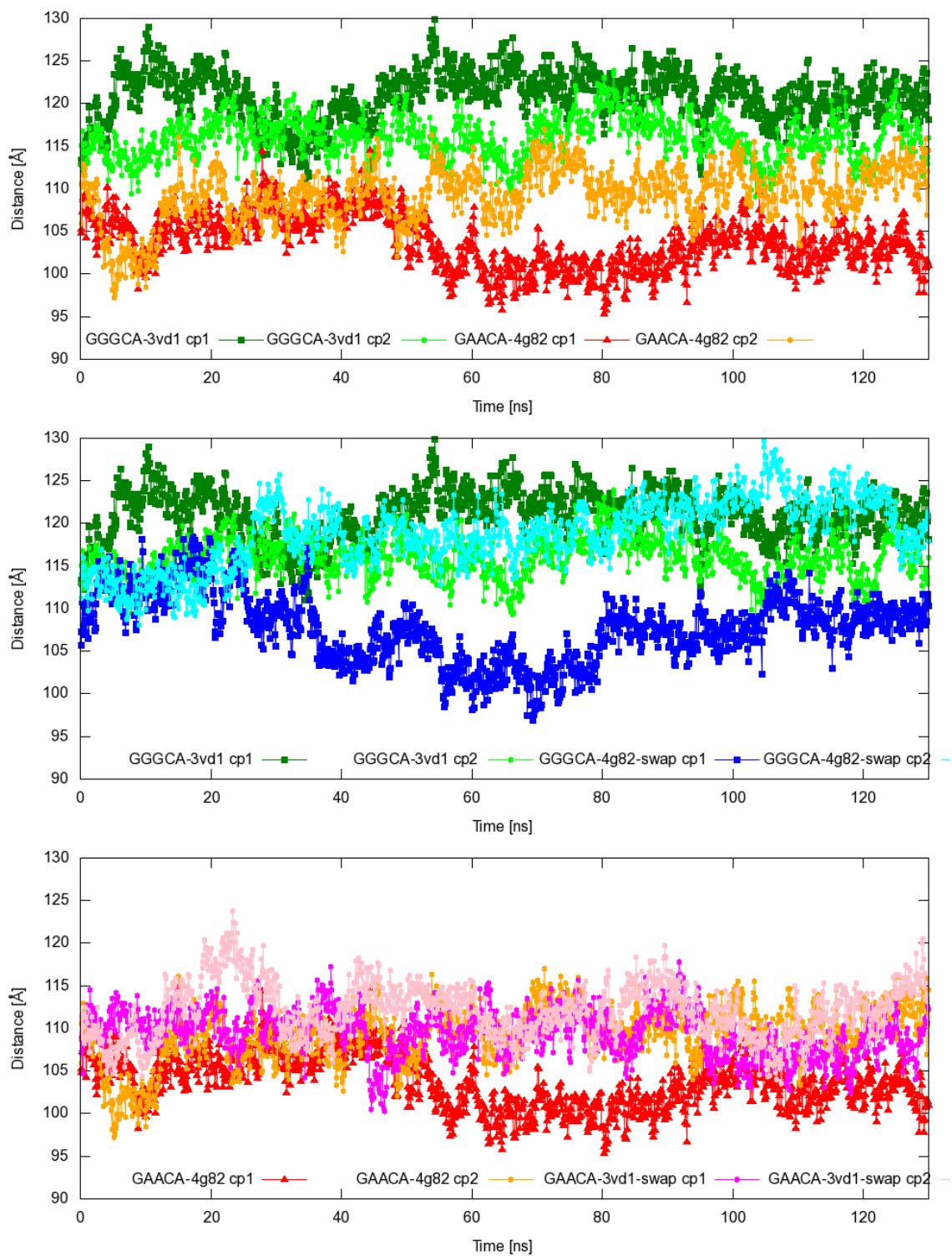


Figure A.20. Principal axis angle between dimer and DNA for all systems.

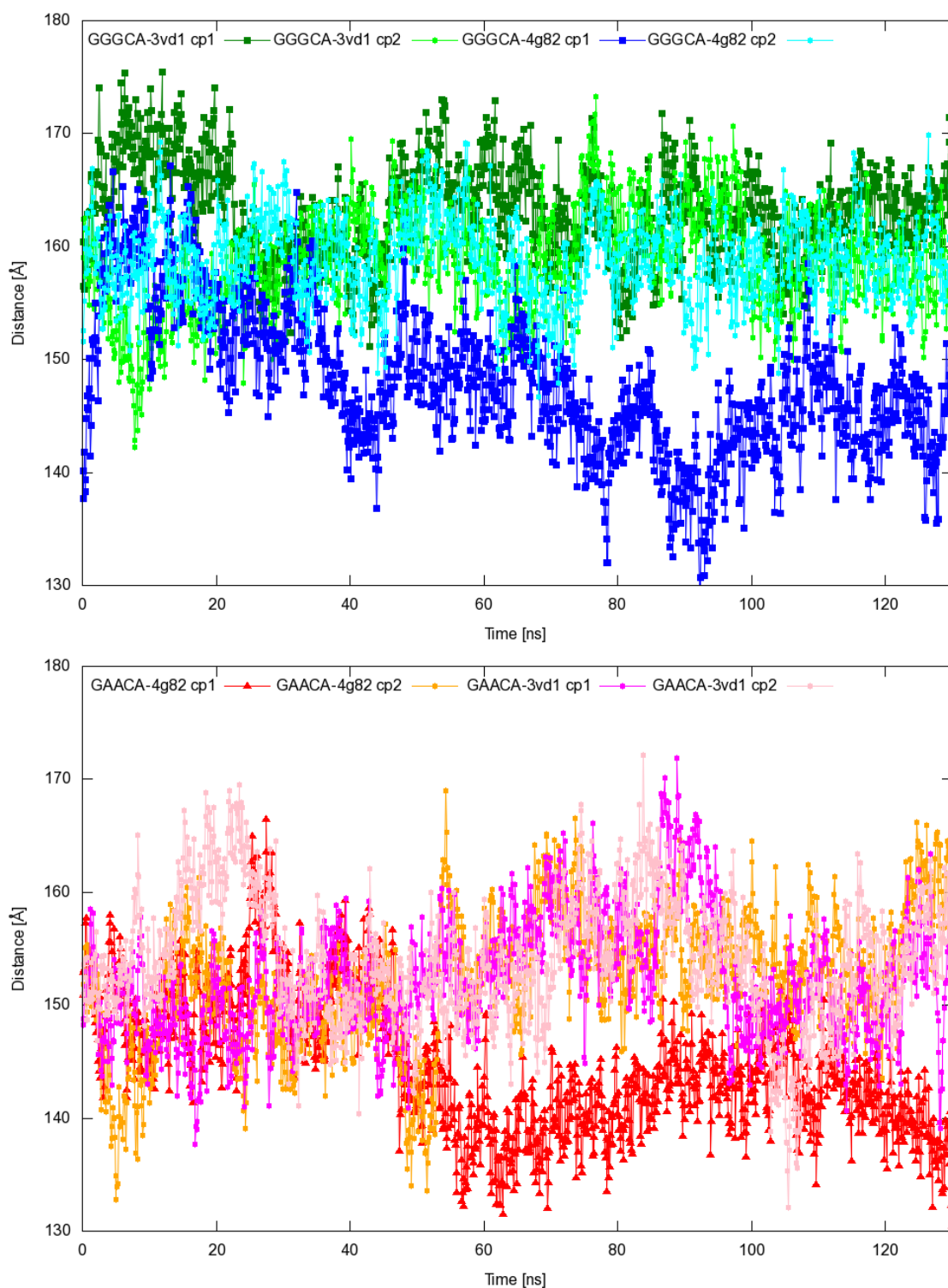


Figure A.21. Dimer Rotation Through Dimer Principal Axis.

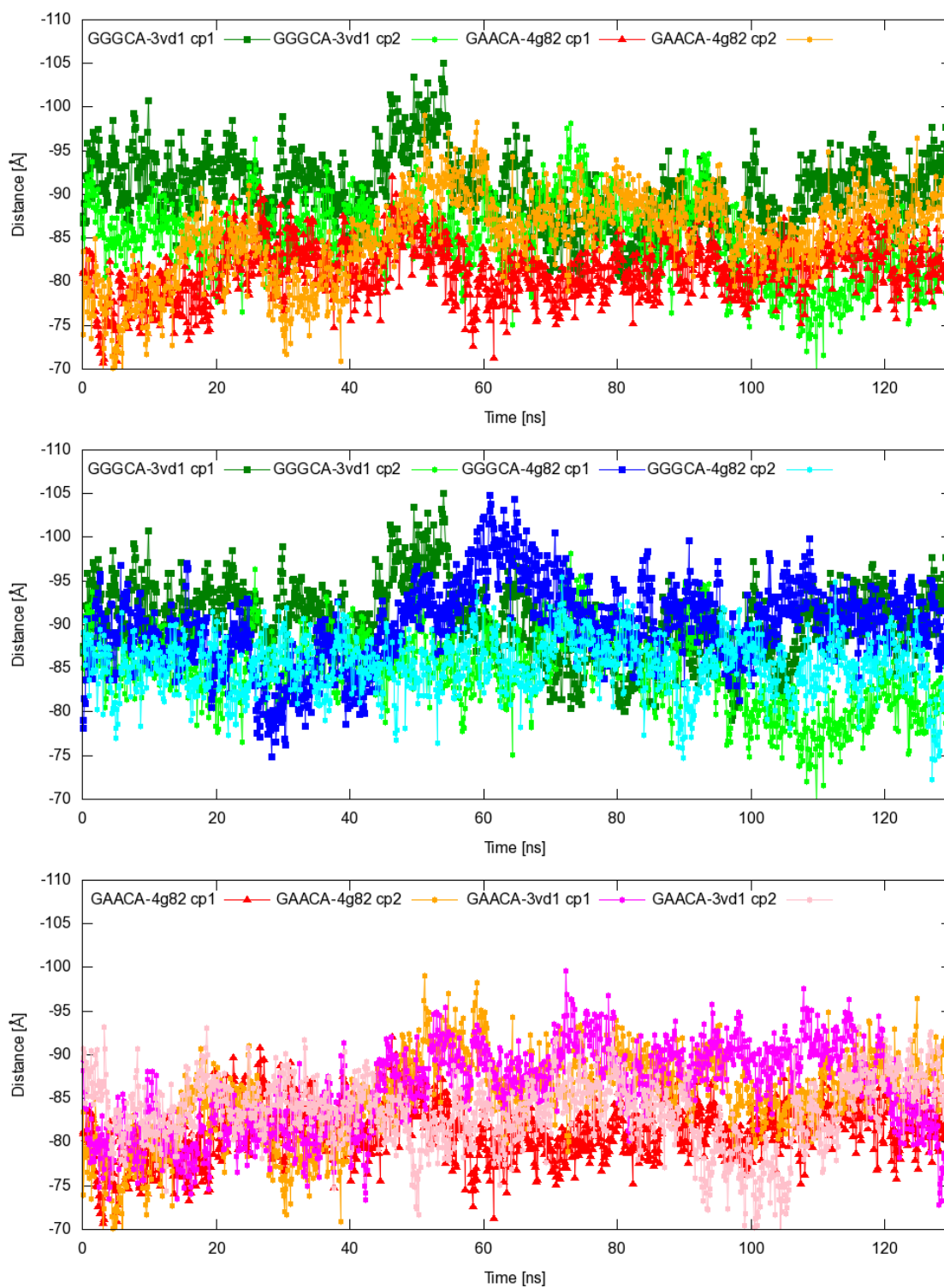


Figure A.22. Dimer Rotation with respect to plane of dimer.

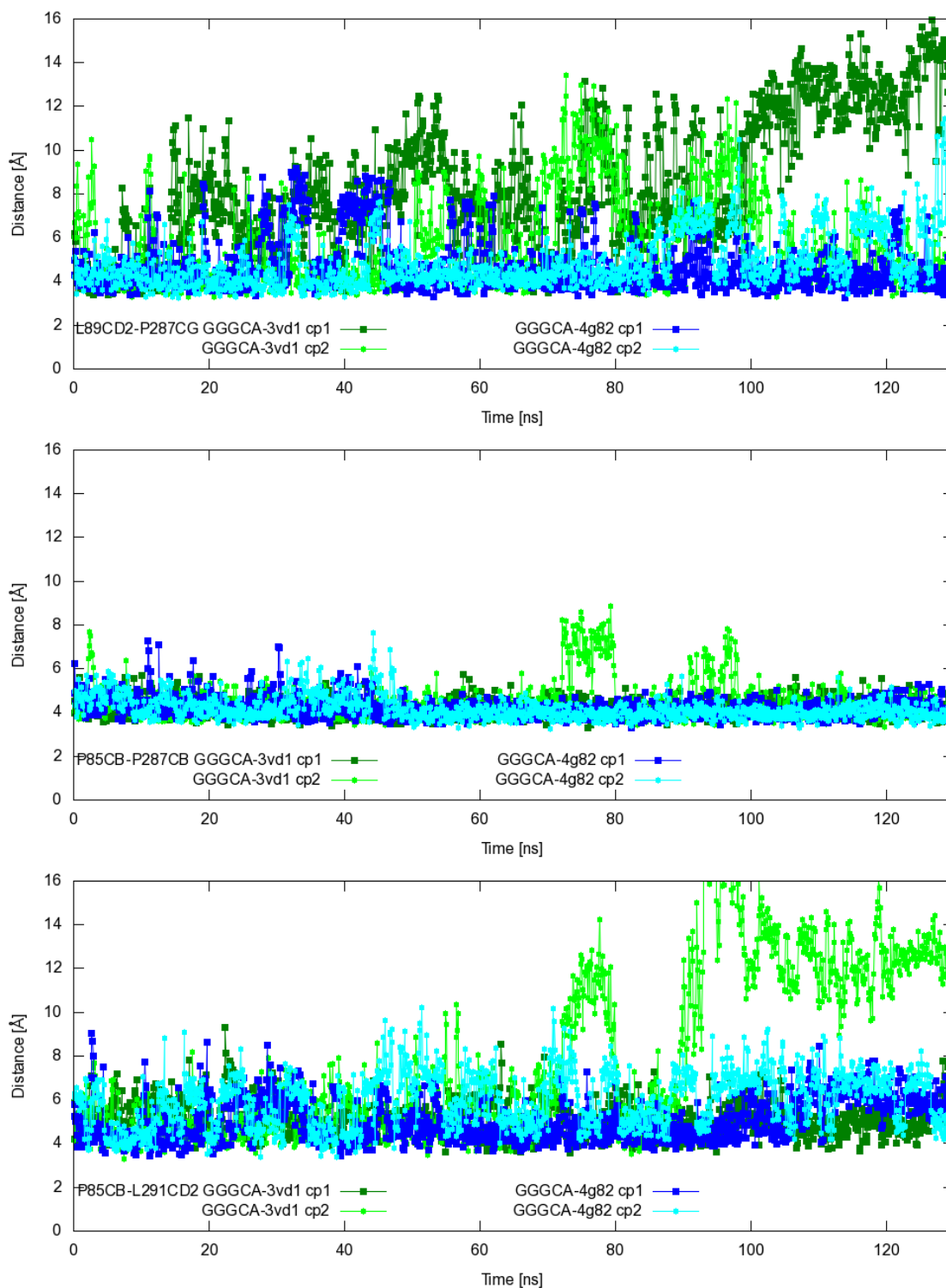


Figure A.23. Monitoring residue distances across monomers representing van der Waal interactions in the dimer interface for the GGGCA systems.

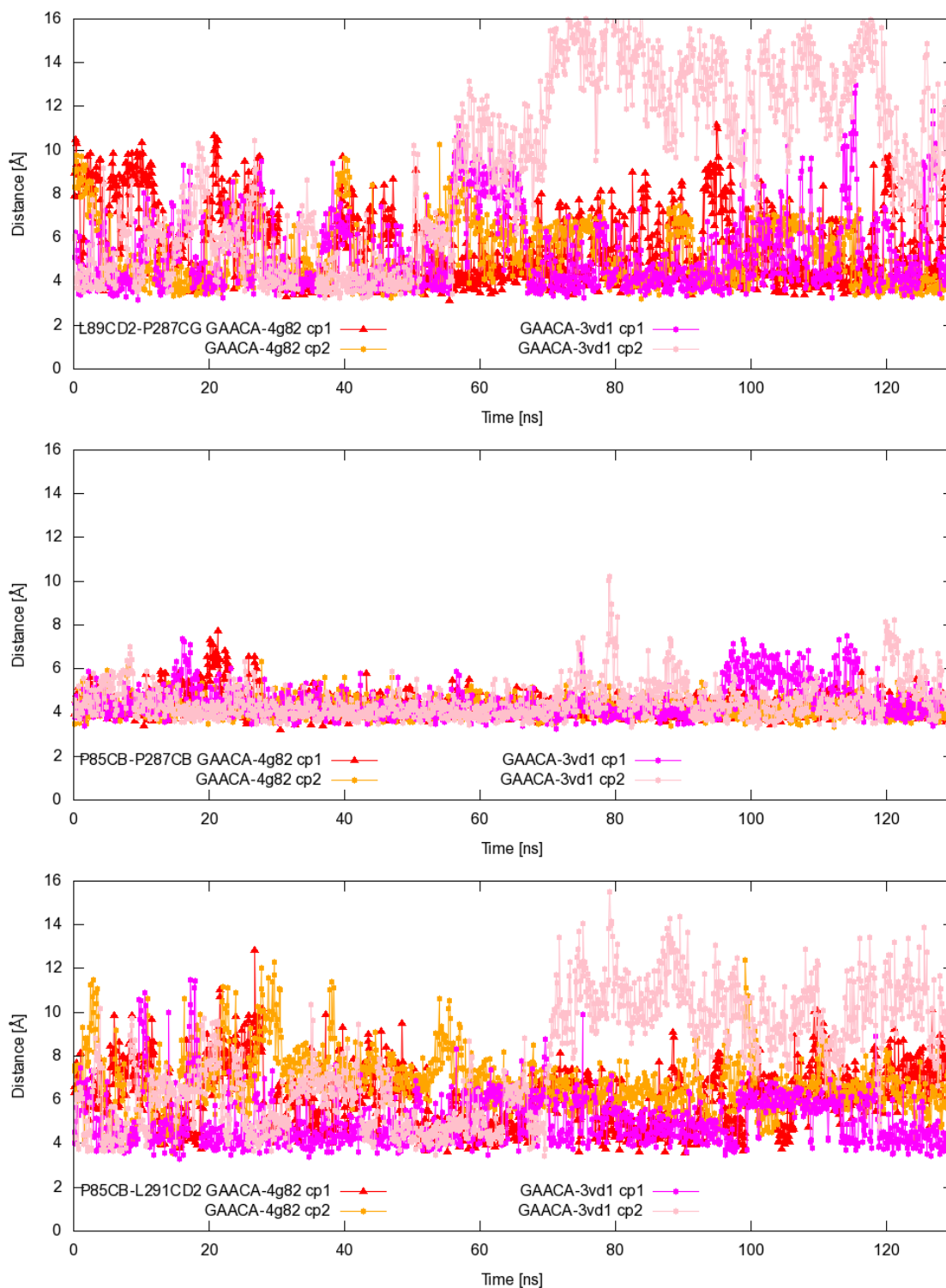


Figure A.24. Monitoring residue distances across monomers representing van der Waal interactions in the dimer interface for the GAACA systems.

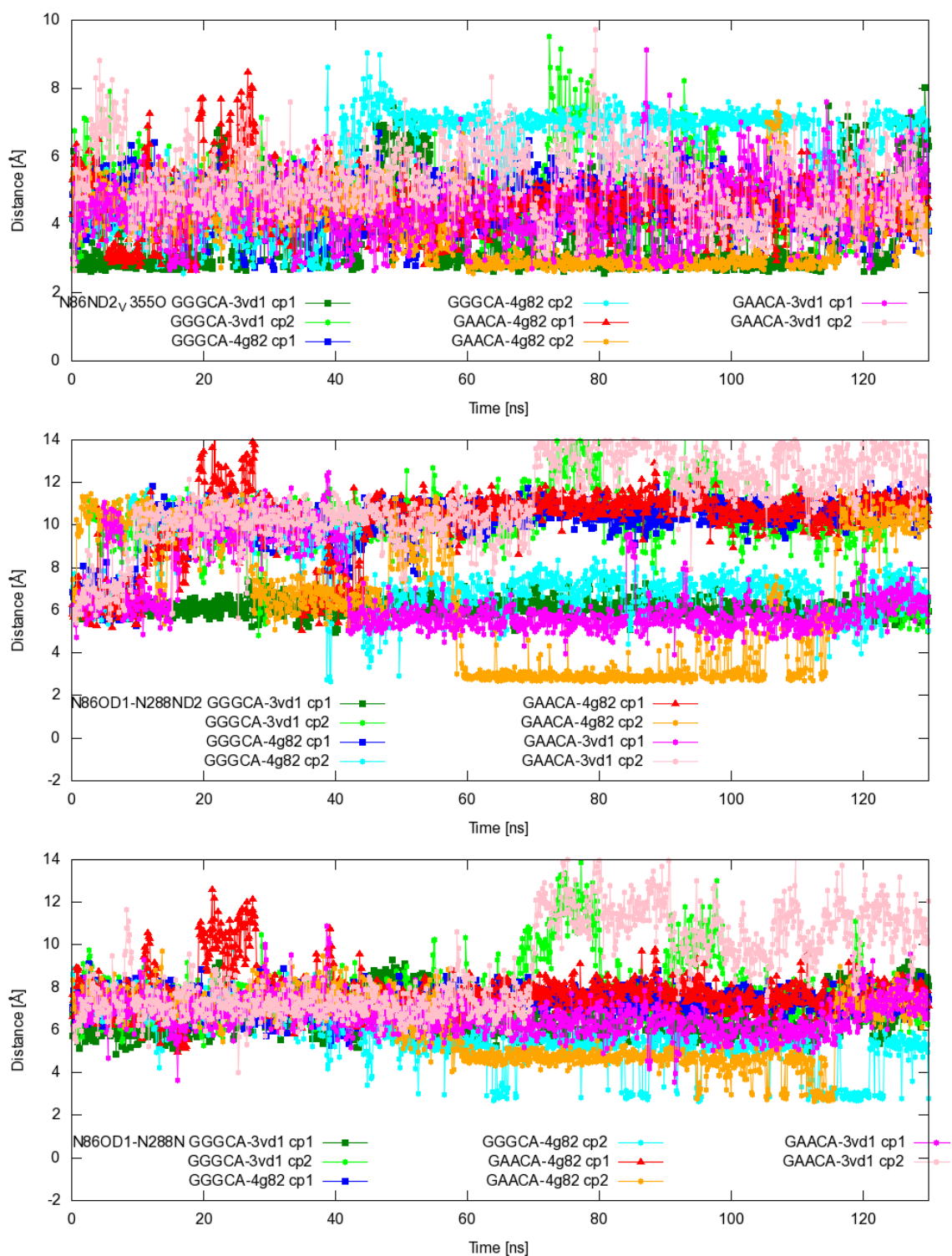


Figure A.25. Monitoring of potential hydrogen bond partners in the dimer interface, Part 1.

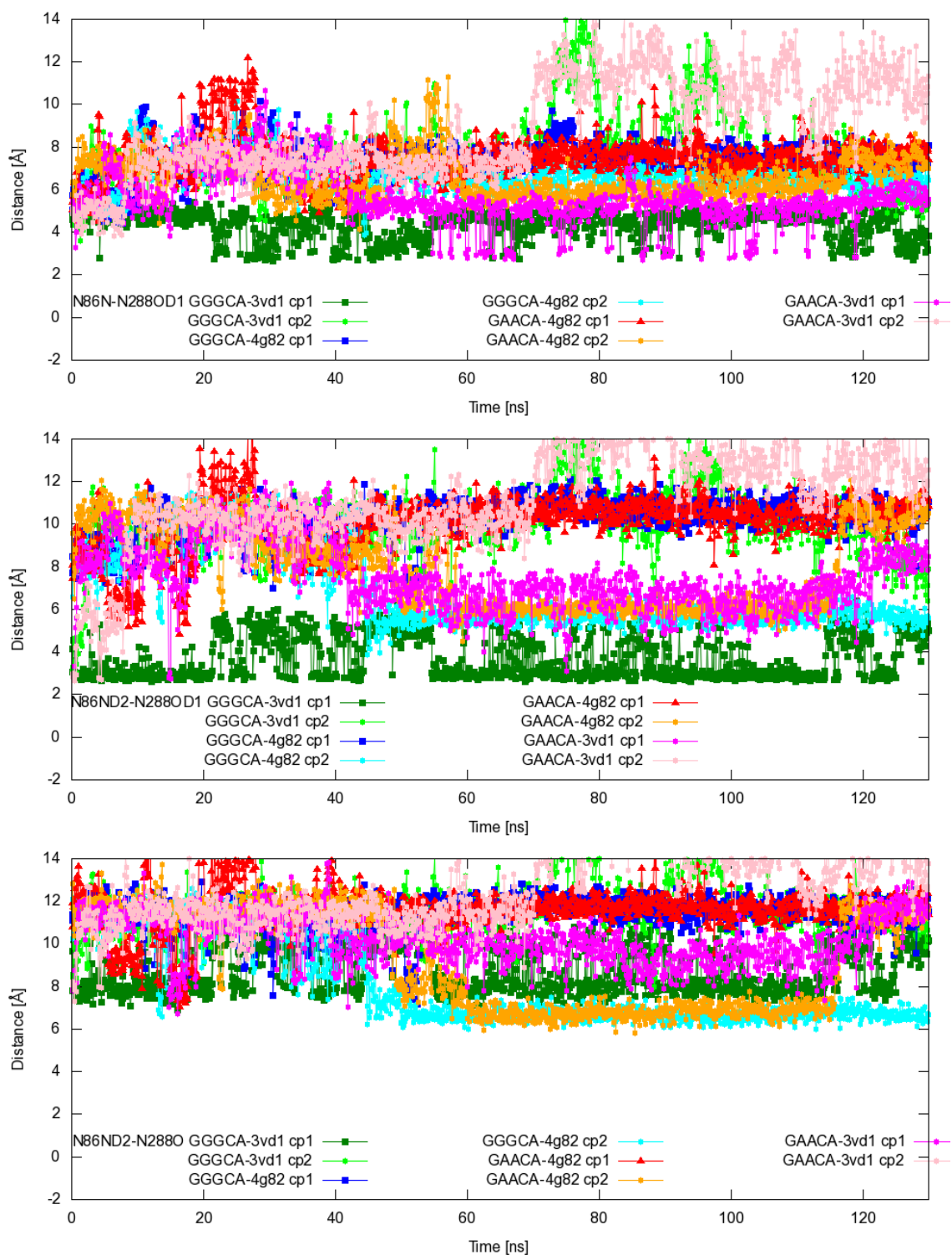


Figure A.26. Monitoring of potential hydrogen bond partners in the dimer interface, Part 2.

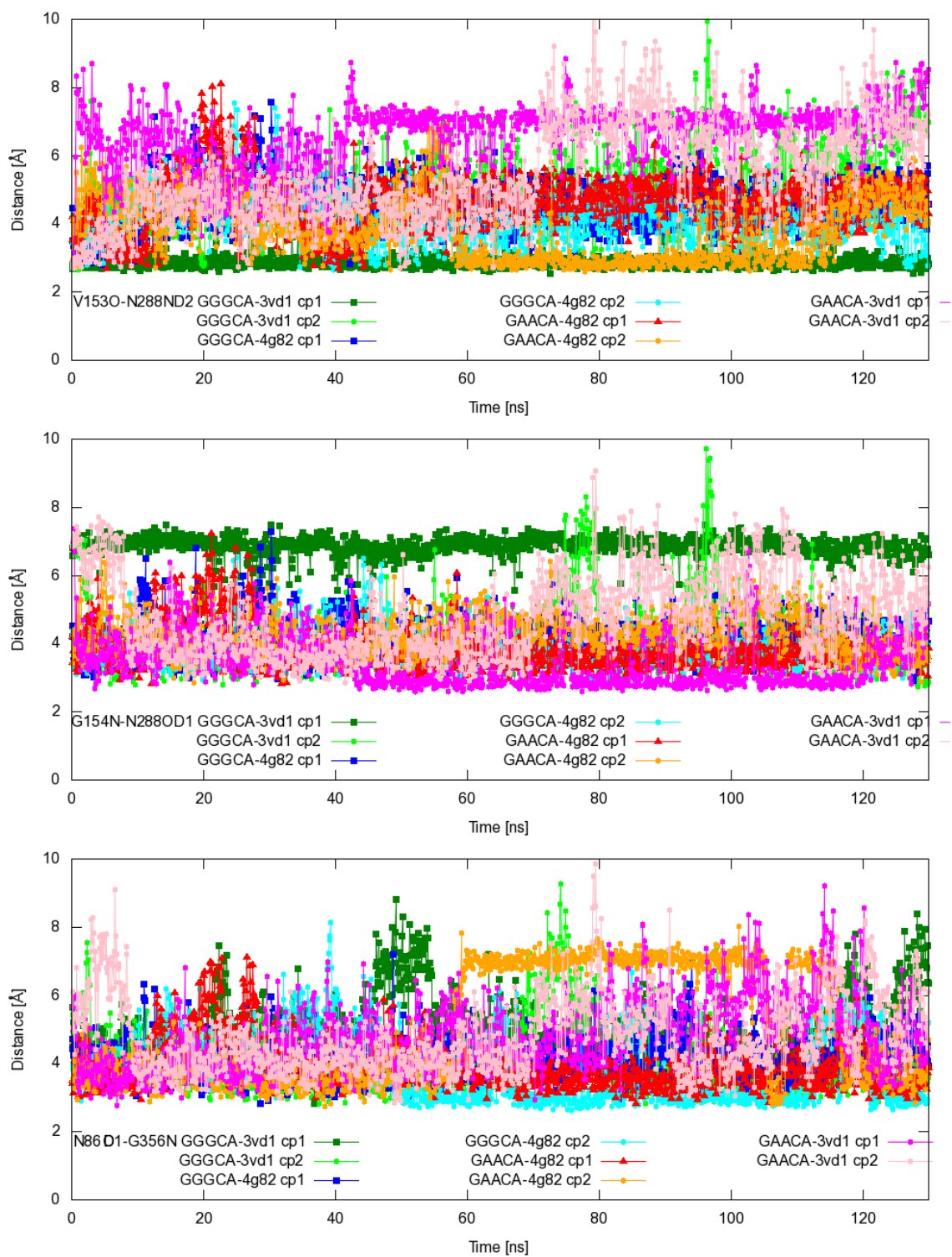


Figure A.27. Monitoring of potential hydrogen bond partners in the dimer interface, Part 3.