University of Alabama in Huntsville

LOUIS

Dissertations | UAH Electronic Theses and Dissertations

2012

Preferred DNA binding orientation of the hyperthermophile protein Sac7d

Gokul Krishna Turaga

Follow this and additional works at: https://louis.uah.edu/uah-dissertations

Recommended Citation

Turaga, Gokul Krishna, "Preferred DNA binding orientation of the hyperthermophile protein Sac7d" (2012). *Dissertations*. 301.
https://louis.uah.edu/uah-dissertations/301

This Dissertation is brought to you for free and open access by the UAH Electronic Theses and Dissertations at LOUIS. It has been accepted for inclusion in Dissertations by an authorized administrator of LOUIS.

PREFERRED DNA BINDING ORIENTATION OF THE HYPERTHERMOPHILE PROTEIN Sac7d

by

GOKUL KRISHNA TURAGA

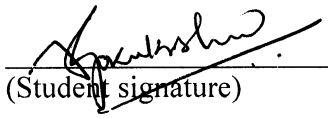
A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
The Biotechnology Science & Engineering Program
to
The School of Graduate Studies
of
The University of Alabama in Huntsville

HUNTSVILLE, ALABAMA

2012

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree from The University of Alabama in Huntsville, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by my advisor or, in his/her absence, by the Director of the Program or the Dean of the School of Graduate Studies. It is also understood that due recognition shall be given to me and to The University of Alabama in Huntsville in any scholarly use which may be made of any material in this dissertation.



(Student signature)

08/20/2012

(date)

DISSERTATION APPROVAL FORM

Submitted by Gokul Krishna Turaga in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Biotechnology Science & Engineering Program and accepted on behalf of the Faculty of the School of Graduate Studies by the dissertation committee.

We, the undersigned members of the Graduate Faculty of The University of Alabama in Huntsville, certify that we have advised and/or supervised the candidate on the work described in this dissertation. We further certify that we have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Biotechnology Science & Engineering Program.

20 AUG 2012
Dr. John Shriver (Date) Committee Chair

Dr. Stephen Edmondson

Dr. Robert McFeeters

Dr. Pamela Twigg

Dr. Roy Magnuson

Dr. Joseph D. Ng Program Director

Dr. John J. Fix College Dean

9/18/12
Dr. Rhonda K. Gaede Graduate Dean

ABSTRACT
The School of Graduate Studies
The University of Alabama in Huntsville

Degree Doctor of Philosophy Program Biotechnology, Science and Engineering

Name of Candidate Gokul Krishna Turaga

Title Preferred DNA binding orientation of the hyperthermophile protein Sac7d

Sac7d is a 7.6 kDa hyperthermophile DNA binding protein that binds to the minor groove of the DNA with no known sequence specificity. Binding introduces a sharp kink (~70°) and the kink is stabilized by the intercalation of the side chains of V26 and M29. No evidence of cooperativity was observed in reverse titrations (DNA into protein) followed by measuring the quenching of intrinsic tryptophan fluorescence. Forward titrations (protein into DNA) followed by monitoring the change in circular dichroism of the DNA indicate a cooperative unwinding in the DNA with increasing protein-binding density. An explanation for this apparent synergistic effect has not been presented.

The focus of the present study is the identification of possible preferences in binding orientation for adjacent Sac7d molecules and the potential influence of DNA deformation (bending). The binding of adjacent Sac7d on the DNA has been studied by characterizing the DNA binding of Sac7d dimers in head-to-head, tail-to-tail and head-to-tail orientations, and the binding of non-crosslinked Sac7d at adjacent sites on DNA using paramagnetic relaxation enhancement (PRE). Further, packaging multiple Sac7d molecules on elongated DNA molecules was studied by fitting the small angle X-ray scattering (SAXS) data considering multiple binding orientations for adjacent Sac7d

molecules. The analysis of DNA binding of Sac7d dimers indicates that K9C dimer (tail-to-tail dimer) shows a 1000-fold increase in DNA binding affinity compared to the Sac7d monomer. Cross-linking Sac7d monomers on DNA also indicates a clear preference for dimers in the tail-to-tail orientation. The PRE assay using non-crosslinked Sac7d also indicates a strong preference for tail-to-tail orientation. The previous experimental SAXS data for Sac7d-32-mer [d(GC)] complex was shown to be consistent with tail-to-tail, head-to-head orientation for adjacent Sac7d molecules suggested by the present work. All of the data are consistent with a preference for tail-to-tail orientation of packaging Sac7d onto DNA. The results are discussed relative to the previous work on Sac7d binding to DNA, including NMR and x-ray crystal structures.

Abstract Approval: Committee Chair _____
Dr. John Shriver

Department Chair _____
Dr. Joseph D. Ng

Graduate Dean Rhonda Kay Gaede 9/18/12
Dr. Rhonda K. Gaede

ACKNOWLEDGMENTS

I sincerely thank my advisor, Dr. John Shriver for suggesting this research project and for providing assistantship throughout my graduate studies. I am grateful for his guidance, support, encouragement and patience. I thank Dr. Stephen Edmondson for his guidance, teaching and patience. I am extremely grateful for the opportunity to learn from him. I thank my other committee members Dr. Pamela Twigg, Dr. Robert McFeeters and Dr. Roy Magnuson for their valuable suggestions and continuous cooperation.

I thank all the present and the past members in Shriver lab for their support. Specially, I thanks to Kelley Smith for making my life in the lab easier. I would like to thank all the lab rats in the material science building for their friendship over the years and for helping me out on various occasions. I thank everyone in the laboratory of structural biology and in McFeeters' lab for their help. I would like to thank the Podila lab for providing access to their PCR machine and nanodrop.

I would like to thank Dr. William Setzer and the department of chemistry for supporting me with a graduate teaching assistantship for a part of my graduate studies. I thank Sharon Blair for all her assistance.

I appreciate the support and encouragement of all my friends. I cannot thank my family enough for their love, support and constant encouragement. I appreciate my wife for everything she is.

Finally, I thank the National Institutes of Health for the funds.

TABLE OF CONTENTS

Page

LIST OF FIGURES XIV

LIST OF TABLES XXII

Chapter Page

1. INTRODUCTION 1

1.1 General background 1

1.2 Protein-DNA interactions 3

1.3 DNA recognition 4

1.3.1 Direct Recognition 5

1.3.2 Indirect Recognition 7

1.4 DNA bending 8

1.4.1 Sequence dependent DNA bending 8

1.4.2 Protein binding-induced DNA bending 10

1.5 Significance of DNA bending and indirect readout 13

1.5.1 Local sequence dependent minor groove shape recognition 15

1.5.2 Deformability of the DNA to be able to form a snug fit with the proteins .. 17

1.5.3 Refolding or restructuring of proteins upon binding 18

1.5.4 Layers of specificity 18

1.6. Origin of specificity from non-specific binding and DNA bending 19

1.7 Sac7d-DNA binding system 20

1.7.1 General background 20

1.7.2 DNA binding and bending 20

1.7.3. Rationale for the present work 25

2. ENGINEERED SAC7D DIMERS – CROSSLINKED IN DIFFERENT ORIENTATIONS 27

2.1 Introduction 27

2.2 Model building and mutagenesis 30

2.3 Preparation of the mutant proteins 34

2.4 Characterization of the Sac7d mutants 34

2.5 Engineering Sac7d dimers 39

2.5.1 Oxidation 39

2.5.2 Chemical cross-linking 39

2.6 Cross-linking Sac7d in head-to-tail orientation 40

2.6.1 Cloning C-terminal His-tag 43

2.6.2.Expression and purification of His-tag K9C (C-terminal His-tag tail cysteine mutant) 45

2.6.3 Head-to-tail cross-linking and purification 45

2.7 Characterization of Sac7d dimers 49

2.7.1 Electrophoresis 49

2.7.2 Circular dichroism 49

2.7.3 Thermal stability 54

2.7.4 Nuclear magnetic resonance 56

2.8 Structural studies of Sac7d dimers using residual dipolar coupling analysis . 61

2.8.1 Data collection 64

2.8.2 Results 65

2.8.3 RDCs for head-to-tail mutant 69

2.9 Discussion 69

3. DNA BINDING OF SAC7D DIMERS 71

3.1 Introduction 71

3.2 Fluorescence spectroscopy 72

3.2.1 Fluorescence data analysis 74

3.2.2 Binding parameters for Sac7d dimers 77

3.2.2.1. Binding parameters for dimers cross-linked in head-to-head orientation 77

3.2.2.2. Binding parameters for tail-to-tail orientation 80

3.2.2.3. Binding parameters for head-to-tail orientation 82

3.3. Sac7d-DNA binding monitored by the change in the circular dichroism of DNA 84

3.4 DNA affinity Chromatography 93

3.4.1 Preparation of DNA affinity columns 94

3.4.2 Preparation of Sac7d dimers cross linked in different orientations 94

3.4.3 Separation of Sac7d dimers cross linked in different orientations 95

3.5 Summary 100

4. IDENTIFICATION OF NEAREST NEIGHBOR ORIENTATION BY CROSSLINKING SAC7D ON DNA 103

4.1 Introduction 103

4.2 Materials and methods 104

4.3 Results and discussion 105

4.4 Conclusions 108

5. PARAMAGNETIC RELAXATION ENHANCEMENT ASSAY 110

5.1 Introduction 110

5.2 Paramagnetic Relaxation Enhancement 113

5.2.1 Application of PRE for biomolecules 114

5.3 Materials and methods 115

5.3.1 Oligonucleotides and introduction of the paramagnetic probe 115

5.3.2 Protein sample preparation and NMR titrations 116

5.4 Results and discussion 117

5.5 Conclusions 133

6. ANALYSIS OF SMALL ANGLE X-RAY SCATTERING OF SAC7D-DNA COMPLEX 134

6.1 Introduction 134

6.2 Model building 135

6.3 Methodology for the calculation and comparison of scattering profiles 146

6.4 Results and discussion 147

6.5 Conclusion 150

7. DISCUSSION 152

7.1 Sac7d-DNA binding and DNA unwinding 152

7.2 DNA binding of Sac7d dimers 154

7.2.1 Sac7d dimers 154

7.2.2 DNA binding of Sac7d dimers 155

7.3 DNA binding of non-cross linked Sac7d 159

7.4 Preferred binding orientation 160

7.5 Role of intercalation and DNA bending in preferential orientation and indirect readout 163

7.6 Packaging multiple Sac7d on DNA and potential role of Sac7d *in vivo* 164

8. CONCLUSION AND FUTURE WORK 166

8.1 Conclusions 166

8.2 Future work 168

APPENDICES 169

APPENDIX A: BIOCHEMISTRY AND MOLECULAR BIOLOGY TECHNIQUES AND DATA COLLECTION 170

APPENDIX B: STANDARD LABORATORY EQUIPMENT, MATERIALS, SOLUTIONS, BUFFERS, ENZYMES, ORGANISMS, MEDIA AND ANTIBIOTICS 181

APPENDIX C: SCRIPTS USED IN THE STUDY FOR DATA ANALYSIS 183

APPENDIX D: CHEMIACL CROSS-LINKERS, CHEMICAL MODIFICATIONS AND PROTEIN SEQUENCES 196

REFERENCES 198

LIST OF FIGURES

Figure page

1.1 Crystal structure of Sac7d (grey) bound to 8 base pair DNA (green) (5'-GCGATCGC) (1azp). 22

1.2 Regular head-to-tail packaging of Sac7d molecules on to 32-mer [d(GC)] modeled from the 1azp coordinates 23

1.3 NMR solution structure depicting two Sso7d molecules bound in preferential head-to-head orientation on a 12-base pair DNA (5'-CTAGCGCGCTAG-3')(1bbx). 24

2.1 Sac7d structure depicting the 'head' (blue) and 'tail' (orange) regions 29

2.2 A head-to-head model showing two Sac7d molecules (grey ribbon models) placed in head-to-head orientation on B-DNA (green ball and stick model). The head region residues are labeled. 31

2.3 A tail-to-tail model showing two Sac7d molecules (grey ribbon models) placed in tail-to-tail (tail residues in light grey) orientation on B-DNA (green ball and stick model). The tail region residues are labeled. 32

2.4 A head-to-tail model showing two Sac7d molecules (grey ribbon models) placed in head-to-tail orientation on B-DNA (green ball and stick model). The tail region of the protein on the top points to the head region of the protein at the bottom. The tail and head regions residues are labeled. 33

2.5 Cation-exchange chromatography elution profile showing two distinct peaks corresponding to monomer and dimer Protein flow is indicated in black. Red line indicates 0.0M to 1.0 M NaCl gradient. The small peak between the monomer and dimer was consistently observed in all the preparation and it ran as monomer Sac7d on SDS-PAGE. The cysteine mutation may be responsible for change in affinity to the cation-exchange column and salt dependence of affinity due to differential exposure of amino acid side chains responsible for binding..... 35

2.6 SDS-PAGE image showing monomer and dimer mixture obtained from cation-exchange column for mutant protein preparation. The gel contents are: Lane 1: molecular weight ladder, lane 2: recombinant Sac7d, lane 3: Sac7d cysteine mutant preparation with DTT, lane 4: Sac7d cysteine mutant preparation from the cation-exchange column without DTT. Sac7d runs at 10kDa in a 16.0% Tris-Tricine SDS-PAGE 36

2.7 CD curves for Sac7d (black bold curve), K9C dimer (dotted red curve) K9C with DTT (blue curve) show good agreement indicating that the mutations and dimerization do not affect the secondary structure of the proteins. 37

2.8 Over lay of binding curves for Sac7d monomer and K9C monomer (tail cysteine mutant with DTT) in 10 mM KH_2PO_4, pH=7.2 using 100-mer [d(GC)]. The binding parameters for the curves are shown in the table below (table 2.1). Quenching is defined as (F_i-F_0/F_0). DNA/protein ratio is the number of DNA nucleotides (mM) per protein (mM of monomer)..... 38

2.9 SDS-PAGE of Sac7d dimers (K9C dimers). Panel-A: Diamide induced oxidation: Lane 1 shows molecular weight ladder, lane 2 shows the monomer dimer mixture obtained from cation-exchange column, lane 3 shows complete dimerization obtained from diamide treatment. Panel-B: Bi-functional cross-linker induced cross-linking: Lane1 shows molecular weight ladder, Lanes 2-4 show dimer fractions that are separated (on cation-exchange column Figure 2.11) from residual monomer followed by treatment with bi-functional cross-linking agent. Lane 5 shows monomer Sac7d. 41

2.10 Separation of monomer and dimer fractions resulting from chemical cross-linking 42

2.11 Scheme for the separation Sac7d dimers with different cross-linking orientations by the introduction of a C-terminal His-tag to the tail cysteine mutant. Panel 1 shows the formation of a mixture of dimers with three possible orientations. Panel 2 shows the dimers with different number of His-tags that can be separated on a Ni-column using imidazole gradient..... 44

2.12 SDS-PAGE image showing the three types of dimers. Lane1: Molecular weight ladder, lanes 2,3,5,6 different dimer fractions for control purpose, Lane 4: intentionally left blank, Lane7: the mixture of three types of dimers as three distinct bands (that are loaded on to the Ni-Column) 46

2.13 SDS-PAGE image of fractions separated through Ni-column confirming the separation of head-to-tail dimer from the other two orientations. The difference in the molecular weights for the dimers in three different orientations is due to the extra leader segment containing the His-tag and the spacer region. Tail-to-tail dimer (K9CHis-tag dimer) with two sets of extra segments is the heaviest. Head-to-head dimer (K28C dimer) with no His-tags is the lightest. Head-to-tail dimer (K9CHis-tag-K28C) with one His-tag has intermediate weight. Lane 1: molecular weight marker, lane 2: K9CHis-tag dimer, lane 3: K9CHis-tag-K28C (head-to-tail) dimer, lanes 4 and 5: K28C dimer (with no His-tags) obtained in the Ni-column flow through, lane 7: K9C His-tag preparation showing K9C His-tag monomer and K9C His-tag dimer and lane 8: K28C preparation showing K28C monomer and dimer bands and lanes 6 is intentionally left blank..... 48

2.14 Far UV CD spectra of Sac7d monomer and K9C dimer (tail-to-tail)..... 51

2.15 Far UV CD spectra of Sac7d monomer and K9C-K28C dimer (tail-to-head) 52

2.16 Far UV CD spectra of Sac7d monomer and K28C dimer (head-to-head)..... 53

2.17 Comparison of the DCS curves for Sac7d and K9C (tail cysteine mutant). The blue curve represents Sac7d monomer, red curve represents K9C dimer and the green curve represents a mixture K9C monomer and dimer. The corresponding denaturation parameters are displayed. The solution conditions were 10mM KH_2PO_4, pH=7.2. The protein concentration was ~1mg/ml..... 56

2.18 Comparison of ^{15}N-HSQC of ^{15}N-Sac7d (black peaks) and ^{15}N-K9C dimer (red peaks). The assignments of Sac7d are shown..... 58

2.19 Comparison of ^{15}N-HSQC of ^{15}N-Sac7d (black peaks) and ^{15}N-K28C dimer (red peaks). The assignments of Sac7d are shown..... 59

2.20 Comparison of ^{15}N-HSQC of ^{15}N-Sac7d (black peaks) and ^{15}N-K9CHis-tag-K28C dimer (head-to-tail) (red peaks). The assignments of Sac7d are shown. The extra unassigned (red) peaks represent the residues in His-tag and the spacer between the His-tag and the spacer..... 60

2.21 Comparison of correlation chemical shifts of K28C-Sac7d and native Sac7d K28C mutation is labeled. Two other extra peaks are observed in the ^{15}N-HSQC of K28C that were not assigned..... 61

2.22 Comparison of correlation chemical shifts of K9C-Sac7d and native Sac7d. K9C mutation is labeled..... 62

2.23 Comparison of experimental and calculated N-H RDCs for K9C dimer (tail-to-tail dimer)..... 65

2.24 Figure 2.21 Comparison of experimental and calculated N-H RDCs for K28C dimer (head-to-head dimer)..... 66

2.24 Superimposition of K9C (blue) structure obtained from XPLOR on Sac7d (1sap) (Green) solution structure 67

2.25 Superimposition of K28C (blue) structure obtained from XPLOR on Sac7d (1sap) (Green) solution structure..... 68

3.1 Reverse titration of Sac7d with 8mer [d(GC)] followed by fluorescence quenching (open circles). (Conditions were 0.01M phosphate buffer, pH=7.2 at 20 ºC). The bold line is the fit obtained that gives the values of Ka (binding constant), n (site size) and %Q (percentage quenching). The concentration of the protein was 1.5

μM. Five μl injections of stock concentrations of DNA (0.3 mM) were added to the protein solution until the quenching remained constant. X-axis represents the ratio of DNA to protein concentration and Y-axis represents fluorescence quenching. DNA/protein ratio is the number of DNA nucleotides (mM) per protein (mM of monomer). 73

3.2 Comparison of fluorescence quenching data for the dimers cross-linked in different orientations in a reverse titration with 100-mer [d(GC)]. DNA/protein ratio is the number of DNA nucleotides (mM) per protein (mM of monomer). .. 84

3.3 Forward titration of Sac7d with dGdC 100-mer followed by circular dichroism. The initial CD spectrum for the DNA in the absence of protein is shown bold and labeled ([protein]:[DNA]=0). Spectra for progressively increasing g ratios of protein:DNA (approximately 0.01, 0.02, 0.03, 0.04, 0.05, 0.07, 0.11, 0.16, 0.21 and 0.26) in 0.01 mM KH_2PO_4 (pH=7.2) are shown. The initial concentration of 100-mer [d(GC)] was ~65μM (McAfee, Edmondson et al. 1996). 85

3.4 Plot of fractional change in Δε values against the ratio of protein to DNA. The markers represent the data (triangles = polyd(AG).poly d(CT), circles = poly d(AC)-poly d(GT), squares = poly d(GC) and diamonds = poly d(IC)) and the smooth are simulated using McGhee Von Hippel model for polynucleotide binding. The figure has been extracted from (McAfee, Edmondson et al. 1996). 86

3.5 Forward titration of dGdC 100-mer with K9C dimer followed by circular dichroism. The initial CD spectrum for the DNA in the absence of protein is shown bold and labeled ([protein]:[DNA]=0). Spectra for progressively increasing g ratios of protein:DNA (approximately 0.02, 0.06, 0.08, 0.11, 0.14, 0.17, 0.2 and 0.25) in 0.01 mM KH_2PO_4 (pH=7.2) are shown. The initial concentration of 100-mer [d(GC)] was ~65 μM. 88

3.6 Forward titration of dGdC 100-mer with K28C dimer followed by circular dichroism. The initial CD spectrum for the DNA in the absence of protein is shown bold and labeled ([protein]:[DNA]=0). Spectra for progressively increasing g ratios of protein:DNA (approximately 0.02, 0.03, 0.05, 0.06, 0.08, 0.11, 0.13, 0.16, 0.19, 0.24 and 0.32) in 0.01 mM KH_2PO_4 (pH=7.2) are shown. The initial concentration of 100-mer [d(GC)] was ~65 μM. 89

3.7 Forward titration of dGdC 100-mer with K9CHis-tag-K28C dimer followed by circular dichroism. The initial CD spectrum for the DNA in the absence of protein is shown bold and labeled ([protein]:[DNA]=0). Spectra for progressively increasing g ratios of protein:DNA (approximately 0.02, 0.03, 0.05, 0.07, 0.09, 0.1, 0.12, 0.14, 0.16, 0.18, 0.21 and 0.24) in 0.01 mM KH_2PO_4 (pH=7.2) are shown. The initial concentration of 100-mer [d(GC)] was ~65 μM. 90

3.8 Forward titration of dGdC 100-mer with K9CHis-tag-G27C dimer followed by circular dichroism. The initial CD spectrum for the DNA in the absence of protein

is shown bold and labeled ([protein]:[DNA]=0). Spectra for progressively increasing g ratios of protein:DNA (approximately 0.02, 0.05, 0.07, 0.09, 0.09, 0.12, 0.14, 0.16, 0.23, 0.30 and 0.38) in 0.01 mM KH_2PO_4 (pH=7.2) are shown. The initial concentration of 100-mer [d(GC)] was ~65μM. 91

3.9 Fractional change in Δε285 for 100-mer [d(GC)] with increasing concentration of Sac7d and Sac7d dimers cross-linked in head-to-head, tail-to-tail and head-to-tail orientations: Sac7d (open circles), K28C dimer (open triangles), K9C dimer (open squares) and K9C His-tag-K28C dimer (open diamonds) in 10mM KH_2PO_4 (pH=7.2). Data points are connected with solid lines to demonstrate progressive change in CD signal. 93

3.10 DNA affinity chromatogram obtained by injecting a dimer mixture obtained from subjecting equimolar amounts of monomer K28C and monomer K9C (head and tail mutants). The elution profile was obtained with 1M NaCl gradient. 96

3.12 The DNA affinity chromatogram was obtained by injecting the dimer mixture described in the figure 3.6 along with a pure preparation of K28C dimer (head-to-head). The elution profile was obtained with 1M NaCl gradient 98

3.13 The DNA affinity chromatogram was obtained by injecting the dimer mixture described in the figure 3.6 along with a pure preparation of K9C dimer (tail-to-tail). The elution profile was obtained with 1M NaCl gradient. 99

4.1 Cross-linking adjacent Sac7d on DNA with 1bbx sequence. 105

4.2 Cross-linking adjacent Sac7d on DNA with 24mer [d(IC)] sequence..... 106

4.3 Cross-linking adjacent Sac7d on DNA with 100-mer [d(GC)] sequence..... 106

5.1 Scheme depicting the use of PRE to study the preferential binding orientation of adjacent Sac7d molecules on DNA. The top panel displays the protein molecule (with its head and tail regions clearly pointed out) and DNA with terminal EDTAMn^{2+} modifications (paramagnetic centers). Lower panel displays the three possible binding orientations for adjacent Sac7d molecules: 1. Proteins bound in head-to-head orientation resulting in placing the tail regions close to the paramagnetic label 2. Proteins bound in tail-to-tail orientation resulting placing the head residues close to the paramagnetic label. 3. Proteins bound in head-to-tail (or random orientation on a longer oligonucleotide). Analysis of the ^{15}N-HSQC spectrum of ^{15}N labeled protein and EDTA-Mn^{2+}labled DNA should indicate the preferential binding orientation based on the set of peaks that broaden out..... 111

5.2 ^{15}N-HSQC of ^{15}N-Sac7d without DNA. ~2.5mg of protein in 10 mM KH_2PO_4 pH=5.0 with 10% D_2O. The peaks labeled showed selective drop in intensity with the addition of [d(GC)]EDTA-Mn^{2+} (as shown in figure 5.3). The peaks that are

broaded only with the addition of Mn^{2+} are labeled. Certain peaks like K9, N37 that disappear with the addition of DNA with Ca^{2+} are not labeled. 117

5.3 ^{15}N-HSQC of ^{15}N-Sac7d without DNA. ~2.5mg of protein and [d(GC)]EDTA-Mn^{2+} in 10 mM KH_2PO_4 pH=5.0 with 10% D_2O. The ratio of concentrations of DNA to protein was 0.25. The peaks labeled showed selective drop in intensity with the addition of [d(GC)]EDTA-Mn^{2+}. 118

5.4 Two Sac7d molecules (grey ribbon) bound in tail-to-tail orientation on a 11-mer dGdC (green wire). The EDTA modification (multicolor wire) is shown at the two ends of the DNA. Labeled in blue are the amino acid residues that are in proximity to the EDTA-Mn^{2+} paramagnetic center for which the peak intensity decreased with the addition of [d(GC)]EDTA-Mn^{2+}. 119

5.5 Two Sac7d molecules (grey ribbon) bound in head-to-head orientation on a 11-mer dGdC (green wire). The EDTA modification (multicolor wire) is shown at the two ends of the DNA. Labeled in blue are the amino acid residues that are in proximity to the EDTA-Mn^{2+} paramagnetic center for which the peak intensity should decrease with the addition of [d(GC)]EDTA-Mn^{2+}, in case of a preferential head-to-head orientation for the adjacent Sac7d molecules. 120

5.6 The decrease in intensity for the peaks in the head (blue and red curves) and the tail (black curves) regions of Sac7d with the addition of [d(GC)]EDTA-Mn^{2+}. The concentration of DNA is expressed in mM nucleotides. The intensity of the peaks is plotted on the Y-axis and the intensity values are normalized to start all the curves at 1.0 122

5.7 A Plot of percentage decrease in the peak intensity for the residues in head and tail regions with the addition of [d(GC)]EDTA-Mn^{2+}. The ratio of concentration of DNA to protein is ~0.3 (mM nucleotides to mM Protein). 123

5.8 The decrease in intensity for the peaks in the head (blue and red curves) and the tail (black curves) regions of Sac7d with the addition of [d(GC)]EDTA-Ca2+. The concentration of DNA is expressed in mM nucleotides. The intensity of the peaks is plotted on the Y-axis and the intensity values are normalized to start all the curves at 1.0 124

5.9 A Plot of percentage decrease in the peak intensity for the residues in head and tail regions with the addition of [d(GC)]EDTA-Ca^{2+}. The ratio of concentration of DNA to protein is ~0.3 (mM nucleotides to mM Protein). 125

5.10 Comparison of percentage decrease in peak intensity for Sac7d and VMSac7d with addition of [d(GC)]EDTA-Mn^{2+}. The ratio of concentration of DNA to protein is 0.3 127

5.11 Model showing the residues that broaden out with the addition of [d(GC)] EDTA-Mn^{2+} to VMSac7d. Two Sac7d molecules (grey ribbon) bound in tail-to-tail orientation on a 11mer [d(GC)] (green wire). The EDTA modification (multicolor wire) is shown at the two ends of the DNA. Residues labeled in blue represent the residues that broaden out both in case of Sac7d and VMSac7d. Residues labeled in red represent those that specifically broaden out with VMSac7d only. 128

5.12 The decrease in intensity for the peaks in the head (blue and red curves) and the tail (black curves) regions of Sac7d with the addition of 1bbx EDTA-Mn2+. The concentration of DNA is expressed in mM nucleotides. 130

5.13 A Plot of percentage decrease in the peak intensity for the residues in head and tail regions with the addition of [d(GC)]EDTA-Ca^{2+}. The ratio of concentration of DNA to protein is ~0.4 (mM nucleotides to mM Protein)..... 131

5.14 Comparison of percentage decrease in peak intensity for Sac7d and VMSac7d with addition of [d(GC)]EDTA-Mn2+. The ratio of concentration of DNA to protein is ~0.3. 132

6.1 Three different Sac7d-32-mer [d(GC)] complexes created for analyzing the SAXS data considering sequential tail-to-tail, head-to-head orientation for adjacent Sac7d on DNA. The kinks and the number of base pairs separating them are pointed out. 136

6.2 4-4-4 building unit shows two Sac7d molecules bound in tail-to-tail orientation. The two kinks are modeled between three 4 base pair B-DNA segments. The building unit being defined here consists of two 4bp B-DNA segments and two Sac7d molecules. The third B-DNA segment shown is a part of the next 4-4- unit. 138

6.3 3-5-3 unit shows two Sac7d molecules bound in tail-to-tail orientation. The two kinks are modeled between 3bp, 5bp and 3bp B-DNA segments. The 3-5-3 binding unit consists of a 3bp and a 5bp segment with two Sac7d molecules bound in tail-to-tail orientation. 139

6.4 2-6-2 unit shows two Sac7d molecules bound in tail-to-tail orientation. The two kinks are modeled between two base pair, six base pair and two base pair B-DNA segments. The 2-6-2 binding unit consists of a 2bp and a 6bp segment with two Sac7d molecules bound in tail-to-tail orientation. 140

6.5 4-4-4 in head-to-tail orientation shows two Sac7d molecules bound in head-to-tail orientation. The two kinks are modeled between three 4 base pair B-DNA segments..... 141

6.6 4-4-4 complex with adjacent Sac7d molecules placed on a 32-mer [d(GC)] B-DNA in sequential tail-to-tail, head-to-head orientation with a separation of 4 base pairs between successive kinks. 142

6.7 3-5-3 complex with adjacent Sac7d molecules placed on a 32-mer [d(GC)] B-DNA in sequential tail-to-tail, head-to-head orientation with a separation of three, five and three base pairs between successive kinks. 143

6.8 2-6-2 complex with adjacent Sac7d molecules placed on a 32-mer [d(GC)] B-DNA in sequential tail-to-tail, head-to-head orientation with a separation of two, six and two base pairs between successive kinks. 144

6.9 4-4-4 head-to-tail complex with adjacent Sac7d molecules placed on a 32-mer [d(GC)] B-DNA in sequential head-to-tail orientation with a separation of 4 base pairs between successive kinks. 145

6.10 Comparison of the scattering profiles calculated for the 4-4-4, 3-5-3 and 2-6-2 models with the experimental data for Sac7d-32-mer [d(GC)]. The experimental data was scaled to match the calculated curves for the models. The black curves (bold=4-4-4, dashed=3-5-3 and dotted=2-6-2) represent the calculated scattering profiles for the constructed models and red dots represent the experimental data. 148

7.1 Solution structure of Sso7d-12mer DNA (1bbx). The DNA is show in green (ball and stick) and the protein in dark grey (ribbon). The anomalies with respect to the Sac7-DNA crystal structure (1azp) are pointed out and labeled (red). 161

LIST OF TABLES

Table page

2.1 Comparison of binding parameters for Sac7d and K9C with DTT. Both the titrations were conducted in 10 mM KH_2PO_4, pH=7.2 using 100-mer [d(GC)]. Q_{max} is the maximum fluorescence quenching. The Reverse titrations were conducted using the general procedure described in appendix A.2.5 38

2.2 List of bi-functional cross linker used in the study. Appendix D.1 lists the full names of the cross-linkers mentioned here. 40

2.3 Calculated secondary structure fractions for Sac7d and Sac7d dimers in different cross-linking orientations. Structural fractions were calculated from a CD spectrum for each protein. CD spectra were collected in 0.01 M phosphate buffer, pH=7.2 at 20 ºC. Error measurements are ±1 standard deviation. *Secondary structure fractions determined from NMR solution structure of Sac7d (1sap) (Edmondson, Qiu et al. 1995). 53

3.1 Binding parameters for K28C dimer 78

3.2 Binding parameters for G27C dimer (head-to-head orientation) 79

3.3 DNA binding parameter of K9C dimer (Tail-to-Tail) 80

3.4 DNA binding parameters of G10 dimer (tail-to-tail) 81

3.5 Binding parameters for head-to-tail dimers: K9CHis-tag-K28C and K9CHis-tag-G27C 82

6.1 Structural parameters for experimental scattering data and the model data obtained by computing the parameters based on an all atom weighted approach. Rc was determined by linear regression analysis of the I(Q).Q analysis of the scattering data after correction for the slit beam geometry of the scattering instrument. The diameters were calculated from the relation D= 2Rc√2. 149

To my beloved pinni-babai.

CHAPTER 1

INTRODUCTION

1.1 General background

The central dogma of molecular biology states that DNA is the information core of all the cellular processes with the exception of RNA in viruses. DNA is a linear polymer that needs to be packaged and concomitantly accessed for gene expression. Packaging and accessing DNA are the two mutually inhibiting tasks that are of utmost importance among all cellular processes, accomplished by a variety of DNA binding proteins that either bind to or interact with DNA either to package or to transcribe.

All the three major domains of life are challenged by the daunting task of packaging very long DNA to fit in the cells that are smaller by several orders of magnitude. Bacteria have an average genome size of 4.6 Mb with a DNA molecule having a contour length of 1.3 mm that is packaged in a cell size of 1-2 microns (Dame 2005; Luijsterburg, Noom et al. 2006; Macvanin and Adhya 2012). Eukaryotic cells face even tougher task of packaging 2 m long DNA into a nucleus of 10 μm diameter (White and Bell 2002).

The three domains of life accomplish the task of packaging DNA by different mechanisms. Bacteria employ a variety of nucleoid-associated proteins, whereas eukaryotes employ two copies of four histones. Achaea have similarities to both bacteria and eukaryotes in their DNA compacting mechanisms. They have histone homologues in euryacheaota and nucleoid associated proteins in crenarcheota (Travers and Muskhelishvili 2005).

The level of complexity of DNA compaction is different in bacteria and eukaryotes. In eukaryotes, histones package DNA into higher order structures whereas in bacteria and crenarchaeota several factors are shown to influence DNA packaging including nucleoid-associated proteins (Luijsterburg, Noom et al. 2006; Zimmerman 2006; Marenduzzo, Micheletti et al. 2010). However, the commonality across the domains is that the DNA packaging proteins bind to DNA and cause structural distortion. The process of packaging DNA by proteins involves both non-specific and specific interactions between DNA and proteins. Both histones and the histone-like architectural proteins are typically basic and bind without sequence specificity to the phosphate backbone of the DNA through electrostatic interactions. But there are several reports that point to a certain degree of specificity resulting from the mechanisms of DNA packaging. For example, there is a strategic distribution of AA, TT dinucleotide steps in the 180 mer nucleosome sequences in order to facilitate convenient winding around the histone core (Olson and Zhurkin 2011). Structural features of such genomic sequences can contribute to specificity and some times cooperativity in the process of protein-DNA binding and in genome compaction. There are evidences for the DNA structure induced specificity and cooperativity in the genome compaction in the case of eukaryotes and bacteria but such

information in archaea is scanty. In order to understand the mechanism of DNA packaging by proteins it is important to better understand DNA bending, the origin of specificity and factors inducing specificity. In this work, the DNA binding of a hyperthermophile protein Sac7d and associated DNA bending has been studied to identify a potential influence of protein induced DNA bending on binding specificity.

1.2 Protein-DNA interactions

Protein-DNA interactions are of critical importance to all the cellular processes. A wide array of proteins likes the polymerases, transcriptional factors and architectural proteins including the histones interact with DNA. The interactions are either sequence specific or non-sequence specific depending on the function of the proteins. Sequence specific interactions typically involve interactions between nitrogen bases of the DNA in the major groove and the amino acid side chains of the proteins. Functional proteins like replication initiation factors, restriction enzymes, transcriptional factors bind to specific sequences associated with their function. DNA replication is initiated at the origin of replication (ORC) sequence that is composed of AT rich 13-mer repeats and four 9mer repeats. Several copies of *dnaA* bind specifically to the 9-mer repeats to initiate DNA replications (Okumura, Yoshimura et al. 2012). TATA box binding (TBP) protein binds specifically to the TATA box (5'-TATAAA-3') a sequence located upstream of the transcription site (Kim, Geiger et al. 1993).

Non-sequence specific binding typically involves electrostatic interactions between the phosphate backbone of the DNA and the positively charged amino acids of

the proteins. Architectural proteins that are typically rich in basic amino acids bind DNA in a sequence independent manner. Some architectural proteins induce DNA bending serving their ultimate goal of DNA compaction and genome packaging. However, many proteins that have a cognate sequence associated with their function, bind to random DNA sequences away from their sequence of interest in a nonspecific manner in order to expedite the search process. They scan through a long range of non-specific sequences before they can recognize and establish specific interactions with their cognate sequences. The shift from non-specific binding mode to specific binding mode and recognition of the specific DNA sequence involves optimization of the interactions in order to achieve the required specificity and increased affinity (Rohs, Jin et al. 2010).

1.3 DNA recognition

Process of recognizing specific DNA sequences by proteins typically involves the change in the binding mode from nonspecific to specific mode by optimizing the interactions between multiple proteins. This optimization may involve deformation of DNA or partial refolding of the protein, or both, that result either in establishing additional contacts between DNA and protein molecules, or change in the interacting partner residues due to the changes in protein-DNA orientation.

Specificity in protein-DNA interactions involves not only the recognition of specific DNA sequences but also the recognition of structural features of DNA. Specificity arising from the recognition of specific nucleotide base pair sequences by the proteins is called direct readout or direct recognition. Specificity arising from the

recognition of specific structural characteristics of DNA is called indirect readout or indirect recognition.

1.3.1 Direct Recognition

Direct recognition commonly involves the specific interactions between the nucleotide bases in the major groove of the DNA and the amino acid side chains in the DNA binding domain of the proteins. This kind of recognition typically involves hydrogen bonding between the nucleotide bases and the amino acid chains. Base read out may also involve hydrophobic interactions. Typically, hydrogen bonding is specific for purines and hydrophobic contacts are specific for pyrimidines (Rohs, Jin et al. 2010). Hydrophobic contacts have been reported in several bacteriophage DNA binding proteins such as 434 repressor and P22 C2 repressor in complexes with their respective operator sequences and in Cro protein to its operator sequence (Aggarwal, Rodgers et al. 1988; Harrison and Aggarwal 1990).

Binding specificity due to hydrogen bonding is due to the hydrogen bond pattern or geometry and number of contacts between the base pairs and the amino acid side chains. Compared to hydrogen bonds with shared partners and single H-bond, those involving multiple donor and acceptor atoms confer the highest degree of specificity. Hydrogen bonds can also be mediated by intervening water molecules. Trp repressor has water-mediated hydrogen bonds (Otwinowski, Schevitz et al. 1988; Harrison and Aggarwal 1990). Water-mediated hydrogen bonds play a significant role in specific base readout in RXR, retinoid associated receptor (RAR) and several other DNA-protein complexes that primarily involve lysines and arginines mediated binding (Rastinejad,

Wagner et al. 2000). In the case of Lac repressor (LacI), water mediated hydrogen bonding is observed in binding to non-specific sequences while specific binding to the cognate sequence does not involve water. The four possible types of nucleotide base pairs have different hydrogen bonding patterns. Binding specificity due to hydrogen bonding is primarily achieved by unique binding patterns adopted by various protein folds. For example, different proteins in a family may share a common recognition mechanism and pattern of hydrogen bonds, such as the homeodomain proteins, repressors proteins like Trp repressor, lambda repressor etc.,. Hydrogen bonding may be supplemented with other recognition modes or interactions in order to achieve the required specificity, affinity and differentiation among different members of the same family. Certain zinc finger domain containing proteins like Cys_2-Cys_2-GATA proteins form hydrogen bonds in the major groove and simultaneously bind in the minor groove. HMG box family proteins and some TATA box binding proteins form hydrogen bonds and also utilize the local DNA helical architecture (Kim, Nikolov et al. 1993; Kim, Geiger et al. 1993). Within a given family of DNA-binding proteins further level of specificity is achieved by other modes of recognition involving shape, groove architecture and flexibility of DNA that are recognized by specific amino acids that are unique for a given protein. Multiple recognition modes, other than the base readout mechanism may be employed to achieve the required binding affinity and specificity. Such non-base pair readout recognition mechanisms constitute indirect readout of DNA.

1.3.2 Indirect Recognition

Indirect recognition involves the specific recognition of structural features of the DNA by the DNA binding domains of the proteins. Differences in electronic configuration and hydrogen bonding pattern results in unique major or minor grove structure, electrostatic potential and groove hydration that could potentially serve as recognition elements and confer specificity in protein binding.

The term "indirect readout" was introduced by Otwinowski *et. al.,* in the description of tryptophan repressor-operator binding (Otwinowski, Schevitz et al. 1988). The tryptophan repressor-operator binding surface is wide with 24 direct interactions and 6 solvent-mediated hydrogen bonds to the phosphate groups. Operator sequence recognition is achieved neither through direct hydrogen bonds nor through non-polar contacts with nucleotide bases, but through the geometry of the phosphate backbone that facilitates the formation of a stable interface.

DNA binding proteins recognize the structural information in the DNA duplex by a variety of mechanisms. In general, the formation of a protein-DNA complex that involves DNA bending depends on the capability of that particular sequence to adapt to a specific structure in the complex. The DNA helix alters its typical B-DNA conformation to resemble other forms of DNA, and bend or undergo large structural changes resulting in strand unwinding and groove opening in order to optimize the protein-DNA binding interface (Travers 1989; Hegde, Grossman et al. 1992; Kim, Nikolov et al. 1993; Kim, Geiger et al. 1993; Shakked, Guzikevich-Guerstein et al. 1994; Lu, Shakked et al. 2000).

1.4 DNA bending

DNA bending can be defined as deviation from the ideal structural conformation of the duplex due to the influence of constraining factors. DNA bending primarily arises from a relative roll between otherwise parallel adjacent base pair planes resulting in a bend in the long axis of the molecule. The general determinants of DNA curvature are torsional flexibility that involves twist between adjacent bases and bending flexibility that involves the deviations in the long axis of the molecule. DNA bending alters groove width that results in the variation of electrostatic potential and degree of hydration associated with the grooves. DNA bending can result either from sequence dependent intrinsic curvature or from protein binding (Nadeau and Crothers 1989; Olson, Gorin et al. 1998)

1.4.1 Sequence dependent DNA bending

DNA molecules behave as elastic rods with a persistence length ~150 bp or 50 nm and are described using worm-like chain model (Hagerman 1988; Taylor and Hagerman 1990; Geggier and Vologodskii 2010). Longer DNA molecules exhibit varying degree of intrinsic curvature that is dependent on their sequence and base composition (Olson, Gorin et al. 1998; Geggier and Vologodskii 2010). The difference in the flexibilities of different sequences and base pair steps is attributed to their differences in stacking energies.

Stacking interactions stabilizes pyrimidine-R steps in a DNA duplex and any differences in these interactions result in DNA bending. The sequence RGCY (R= A/G and Y=C/U) readily bends at the central GC step due to the preferential stacking of GC

on top of each other. The junctions of GC and AT steps have a bend of ~10°-20° (Dickerson, Goodsell et al. 1994). The crystal structure of d 5'-CATGGCATG-3' has a kink at the central GC (Dickerson, Goodsell et al. 1994; Allemann and Egli 1997). However, stacking differences at base pair steps may not always create bends. The crystal structure of duplex 5'-CGCGAATTCGCG-3' has an 18° bend at GC/AT of one end and no bend at the other end (Wing, Drew et al. 1980; Drew, Wing et al. 1981).

The deviation from ideal double helical pattern of the adjacent base steps can result from roll, helical twist, or propeller twist (Travers 2004). Roll is defined as the rotation of adjacent base pair planes with respect to the base-pairing axis. Helical twist is defined as the relative rotation of adjacent base pair planes with respect to the helical axis. Propeller twist is defined as the relative rotation between two bases within a base pair along the base-pairing axis. The three kinds of deviations arise from differences in the participating nucleotides (electronic/functional groups/modifications etc.) that impose variations in the structural aspects of DNA resulting in sequence dependent structural signatures.

In addition to specific local kinks and deformation, global flexibility and groove structure of DNA can be affected by nucleotide composition and sequence. In poly(dA).poly(dT) (poly A tracts), repeating AT/TA base pairs form inter base hydrogen bonds in the major groove resulting in global rigidity due to increased propeller twisting (Nelson, Finch et al. 1987). Global rigidity of the AT tracts is also attributed to the stearic hindrance caused by the methyl groups of T that are in proximity to each other and the sugar phosphate backbone (Aggarwal, Rodgers et al. 1988). In ApA, TpT and ApT steps, propeller-twisting results in bifurcated hydrogen bonds that are accompanied by a

negative roll and result in enhanced minor groove narrowing (Rohs, Sklenar et al. 2005; Haran and Mohanty 2009). Among all the possible combination of adjacent base steps TpA is the most flexible and referred to as hinge joint (Shakked and Rabinovich 1986; Mack, Chiu et al. 2001). In the ApT base step the stacking of methyl groups with the adjacent sugar phosphate backbone prevents the roll of base pair planes whereas in case of TpA step the methyl group projects into the major groove causing no stearic hindrance. In addition, stearic clash of the diagonal adenines from the cross strands leads to local widening of the minor groove creating a flexible hinge in the otherwise rigid A tracts (Koo, Wu et al. 1986). TG/CA steps are also known to be flexible for similar reasons. A correlation between the experimentally calculated stacking energies and the base step flexibility is observed. GC steps have highest stacking energies and hence have very low conformational variability suggesting the rigidity of the base steps. TG/CA and TA steps have the lowest stacking energies rendering highest flexibility. AA/TT and AT steps have intermediate stacking energies (Protozanova, Yakovchuk et al. 2004; Geggier and Vologodskii 2010).

Thus, nucleotide composition and sequence influence both local and global structure of the DNA. As a result, certain sequences are easier to bend compared to others and they may carry a physiological significance by serving as amenable targets for certain specific DNA bending proteins.

1.4.2 Protein binding-induced DNA bending

Many DNA binding proteins induce varying degree of structural changes in DNA. Bending DNA is not limited only to the DNA packaging proteins but also certain

functional proteins induce structural deformation in DNA. DNA bending may serve as a binding determinant or recognition element in protein-DNA complexes with bent DNA.

Individual protein components can bind DNA at regions distant from one another in order to bring them together. For example, for the subunits of the nucleoprotein complex to assemble it is necessary to bring distant regions of the DNA into proximity . In addition, experimental evidence indicates that protein binding induced bending is structural switch for the transcriptional activation and can regulate transcriptional potency(Pérez-Martín and Espinosa 1993; Parekh and Hatfield 1996; Ukiyama, Jancso-Radek et al. 2001).

Proteins binding to DNA can either induce asymmetric charge neutralization in the phosphate backbone resulting in DNA bending or insert amino acid side chains inbetween the base pairs resulting in sharp kinks (Travers 2004; Privalov, Dragan et al. 2009). Protein-DNA salt bridge interactions on one helical face of DNA neutralize residual phosphate charge on that face resulting in unbalanced set of inter-phosphate repulsions along the helix axis. As a result, the DNA helix tends to relax towards that neutralized face. This kind of protein induced DNA bending was originally discussed by Rich and co-workers (Crothers 1994). In this mode of bending, DNA assumes a bent conformation along the protein surface that is typically concave. Catabolite Activator Protein (CAP) binding to its binding site with a two-fold symmetry is a typical example of this binding mode. CAP binds to a palindromic binding site and introduces two kinks, one each in two halves of the binding site (5'-AAATGTGATCT-3'). Primary kinks are facilitated by the flexibility of the 6 and 7 base pair step. The primary kinks along with two smaller kinks result in an over all curvature of 90° (von Hippel and Berg 1986;

Gartenberg and Crothers 1988; Lesser, Kurpiewski et al. 1990; Pabo and Sauer 1992; Chen, Vojtechovsky et al. 2001).

DNA bending by intercalation involves the insertion of hydrophobic amino acid side chains inbetween stacked base pairs of DNA resulting in a sharp kink in the helix (Werner, Gronenborn et al. 1996; Ren, Jenkins et al. 2000). The weakened base stacking interactions are supplemented by the intercalation of hydrophobic amino acid side chains. A group of hydrophobic DNA binding proteins that specifically bind in the minor groove of the DNA have been described that insert one or more hydrophobic amino acid side chains inbetween stacked base pairs leading to partial unstacking and severe kinking. High-resolution structures of several of these proteins in complex with DNA have been solved. They include TATA box binding protein (Kim, Nikolov et al. 1993; Kim, Geiger et al. 1993; Juo, Chiu et al. 1996), repressors like the purine repressor (PuR) (Schumacher, Choi et al. 1994) and the lactose repressor (LacI) (Lewis, Chang et al. 1996; Bell and Lewis 2000), the high mobility group (HMG) of mouse lymphoid enhancement factor (LEF-1) (Love, Li et al. 1995), HMG domain of human sex determining region Y (SRY) (Werner, Huth et al. 1995), HMG I, integration host factor (IHF) (Rice, Yang et al. 1996), Sac7d (Robinson, Gao et al. 1998) and Sso7d (Gao, Su et al. 1998). In the case of Lac repressor and Purine repressor, binding to their respective operator sequences results in the local widening of the minor groove that is stabilized the partial intercalation of the side chains of two leucines (Kalodimos, Biris et al. 2004; Kalodimos, Boelens et al. 2004). TBP introduces a significant kink of ~45° by the intercalation of two phenylalanine residues resulting in the unwinding of the TATA box over seven base pairs by 105° that is compensated by one-third of a turn of positive

supercoiling (Chasman, Flaherty et al. 1993; Kim, Geiger et al. 1993; Kosa, Ghosh et al. 1997; Chastain and Sinden 1998). HMG box binding proteins also introduce DNA kinking by amino acid side chain intercalation. SRY and LEF-1 bend DNA by the insertion of isoleucone and methionine respectively into the minor grove resulting in ~54° and ~120° kinks (Love, Li et al. 1995; Werner, Huth et al. 1995). SOX-17, another HMG domain containing protein introduces a sharp kink in DNA at an ApA step through the intercalation of phenylalanine and methionine. IHF introduces a 160 degree 'U' turn in DNA reversing the direction of the double helix (Swinger and Rice 2004). Two proline residues present at the tips of opposing arms of the protein induce bends that are nine base steps apart. Hyperthermophile proteins Sac7d and Sso7d also introduce severe kinking. Crystal structure of single Sac7d molecule bound to a 8 base pair DNA shows a kink of ~70°. Both Sac7d and Sso7d are known to bend DNA in a similar manner through the intercalation of the side chains of methionine and valine residues between adjacent base steps (Peters, Edmondson et al. 2004; Peters, Edmondson et al. 2005). Insertion of side chains of two amino acids at the same base pair step is also observed in TBP-DNA complex (Chasman, Flaherty et al. 1993; Kim, Nikolov et al. 1993; Kim, Geiger et al. 1993; Kosa, Ghosh et al. 1997).

1.5 Significance of DNA bending and indirect readout

DNA bending is one way to allow indirect readout. The protein binding affinity depends on the ease with which the DNA can be deformed into its structure in the complex (Halford and Marko 2004; Travers 2004; Rohs, West et al. 2009; Tullius 2009). For example, a negative correlation is observed between the binding free energy and the

DNA deformation energy in case of binding studies of bacterial integration host factor (IHF) with its known DNA binding sites (Aeling, Opel et al. 2006). Cyclization studies with E2, a human papilloma viral protein (HPV) using varying DNA sequences, indicate that DNA curvature and flexibility are the main parameters responsible for the affinity changes (Hegde 2002; Watkins, Mohan et al. 2010). In the case of nucleosomal core DNA, there is a periodic placement of AT rich sequences along one helical phase and short GC rich regions in the opposite helical phase facilitating the required bend and alternating narrow and wide groove widths to constitute the binding sites to wrap around the histone octamer (Travers 2004). Zinc finger domain-containing proteins recognize the flexibility of the ATA sequence for specific binding to their cognate sequences (Fairall, Schwabe et al. 1993). Restriction endonuclease EcoRV recognizes its target sequences due to the deformability of the hinge TpA step (Zahran, Daidone et al. 2010). The binding site of γδ resolvase depends on the flexibility of TpA (Yang and Steitz 1995). TBP, serum response factor (SRF), E*co*RI, E*co*RV and E*co*RI DNA methyl transferase (M.E*co*RI) bind their target sequences with flexible AT steps that confer specificity. Thus, certain DNA sequences deviate from the ideal B-DNA structure to optimize protein-DNA binding, and they serve as recognition elements for indirect readout.

Proteins not only recognize the intrinsic curvature of DNA but also recognize the bending introduced by other neighboring proteins. IHF recognizes pre-formed conformational characteristics of consensus DNA sequences at high affinity binding sites (indirect readout) whereas at other sites the relative binding affinity depends on the deformation energy required by the nearest neighbor to force the DNA into a conformation that is required for the complex formation (Aeling, Opel et al. 2006).

Thus, recognition of specific DNA sequences may be achieved by a combination of recognition events. They may operate individually or in unison to bring about the required specificity. The following is a description of some indirect readout mechanisms and determinants of indirect specificity that have been characterized.

1.5.1 Local sequence dependent minor groove shape recognition

Narrow minor grooves locally enhance the electrostatic potential through electrostatic focusing (Rohs, West et al. 2009). Studies using superoxide dismutase revealed that the most negative electrostatic potentials of the DNA phosphates are observed in the grooves due to electrostatic focusing. The narrow active site focuses the field lines away from the protein and into the dielectric solvent. A quantitate treatment of this has been obtained by studies on the MogR-binding site system (Jayaram, Sharp et al. 1989; Honig and Nicholls 1995). Differential electrostatic potential was also shown to be important in codon-anticodon recognition in tRNA and in shaping electrical potential in diverse mRNAs, and in shifting pKas in RNA catalytic sites (Sharp, Honig et al. 1990; Tang, Alexov et al. 2007).

Minor groove shape is essential for achieving functional specificity of Hox proteins. Sequence dependence of minor groove width and the corresponding variation in the electrostatic potential are essential in differentiating minor differences in nucleotide sequences in various members of the *Drosophila* Hox protein *s*ex *c*omb *r*educed (SCR) (Chastain and Sinden 1998).

The Arg repressor in Mycobacterium uses 4 stranded β sheets that lie above the groove to detect narrow minor groove shape via the phosphate backbone. This

recognition does not involve the insertion of side chains of amino acids which is unusual for the mechanism of recognition of narrow minor groove (Cherney, Cherney et al. 2009). Usually, the specific detection of a narrow minor groove involving proteins with arginines is associated with the insertion of the guanidine group of arginine into the narrow minor groove. LEAFY gene regulator binds as homodimer in which the arginines present in the N-terminal arm of the individual monomers bind to distant minor groove regions (Hamès, Ptchelkine et al. 2008). There are several examples of arginines being employed to bring about increased DNA binding affinity and specificity. The structure of DNA binding domain of several proteins is designed such that more than one arginine is brought together to constitute the DNA binding domain. Arginine mediated specific recognition of the minor groove facilitates hetero-dimerization of Oct-1 and MogR and homo-dimerization of Scr-Exd. Specificity rendered by arginine insertion to the minor groove may be due to the stabilization of a bend like in IHF or due to the stabilization of linear DNA by preventing deformation as in the case of HMG-I (Y) (Bewley, Gronenborn et al. 1998).

Lysines are also known to recognize a narrow minor groove. IFN-β - enhansosome structure has several lysines stabilizing the minor groove. However, due to the higher free energy associated with the removal of lysines from water, they are less favored in comparison to arginines (Escalante, Nistal-Villán et al. 2007; Panne, Maniatis et al. 2007; Panne 2008). Histidines are also known to bind to narrow minor grooves in Scr-Exd fкh 250 (Joshi, Passner et al. 2007).

1.5.2 Deformability of the DNA to be able to form a snug fit with the proteins

DNA can deviate from the ideal B-form to change its structure in such a way that the binding surface is optimized (Travers 1989; Shakked, Guzikevich-Guerstein et al. 1994; Lu, Shakked et al. 2000; Garvie and Wolberger 2001; Rohs, West et al. 2009). HPV-E2 binds to its target sequence causing a severe bend by binding to successive major grooves of the sequence. Specificity is achieved by multiple contacts between amino acids and bases in the interface that result in optimizing the binding interface (Brukner, Dlakic et al. 1993; Travers 2004). Groove opening is observed in the case of TBP in order to make the binding tight (Brukner, Sánchez et al. 1995; Buttinelli, Minnock et al. 1998; Travers 2004). Hinc II endonuclease recognizes its cognate sequence GTYRAC due to the deformability of the central YpR step with CpG being the most favored (Sierk, Zhao et al. 2001).

Indirect readout primarily depends on the propensity of a duplex to assume a conformation that is observed in the protein-DNA complex. However, the DNA sequences that cause the required structural alteration may not be in contact with the protein. Certain linker sequences can offer the required flexibility to bring two halves of a binding site together so that they bind to the binding sites on the protein, while the actual flexible linkers is not bound. RevErb nuclear hormone receptor specifically binds to a site containing two TpA steps. Neither of the TpA steps make direct contact with the protein, but specificity in RevErb-DNA binding is achieved by the DNA distortion allowed by the TpA steps (Fairall, Schwabe et al. 1993; Horton, Dorner et al. 2001; Sierk, Zhao et al. 2001).

1.5.3 Refolding or restructuring of proteins upon binding

'Induced fit' like binding may contribute to indirect readout. Some DNA binding proteins transition from unfolded or unstructured state into a structured state upon binding to DNA. Transcriptional factors such as *Fos* and *Ju*n bind DNA as helical hetero dimers that are otherwise unstructured (Ellenberger, Brandl et al. 1992; Travers 1998). Structural transformation is observed in Lymphoid Enhancement Factor (LEF-1) upon binding to DNA (Love, Li et al. 1995). Zinc finger domain of the retinoid x receptor (RXR) which is otherwise disordered, undergoes proper folding upon binding to DNA. The structural modification of binding partners is aimed at optimizing the binding surface in order to achieve desired function. Sequence specific DNA binding of Cys_2 His_2 zinc finger domains facilitates the proper binding of the next zinc finger domain to the major groove by stabilizing the preceding helix by folding and capping of the linker domain (Laity, Dyson et al. 2000). Glucocorticoid receptor (GR) undergoes conformational changes up on DNA binding resulting the exposure of transcription activation surfaces. The complex of GR with its cognate DNA sequence results in the distortion of the DNA in order to optimize the binding surface and to facilitate transcription activation (Lefstin and Yamamoto 1998).

1.5.4 Layers of specificity

Homeodomain protein family is a very good example for describing the presence of several layers of specificity and diversity within a given family of proteins. Among all the homoedomain family members, N51 is found to be uniformly critical for recognizing the AT rich sequences in the TATA box. A finer line of specificity is achieved with Q50

in order to differentiate among the various proteins in the homeodomain family. In case of Q50, it binds to TAATTG or TAATTA. In case of K50, the sequence recognized is TAATCC (Hanes and Brent 1989; Treisman, Gönczy et al. 1989; Tucker-Kellogg, Rould et al. 1997; Grant, Rould et al. 2000).

1.6. Origin of specificity from non-specific binding and DNA bending

From the previous discussion of inter-dependence of binding affinity and the flexibility of DNA it can be inferred that non-sequence dependent protein-induced DNA bending and associated potential indirect recognition mechanisms are interconnected. Non-specific DNA binding and DNA bending can be described as a coupled reaction where the relatively unfavorable DNA bending is accomplished by coupling it to the more favorable protein-DNA binding (Travers 2004; Aeling, Opel et al. 2006; Tullius 2009; West, Rohs et al. 2010). Specificity may arise even from the information-poor sugar phosphate backbone in the minor groove, as in the case of wrapping of genomic DNA around the histone core. Sugar phosphate backbone contacts may play a role in specificity through the positioning of protein recognition elements in orientations that facilitate the formation of more specific contacts (Luscombe and Thornton 2002; Siggers, Silkov et al. 2005). A clear understanding of the non-sequence specific DNA binding induced bending and its relation to a potential indirect readout mechanism requires a comprehensive characterization of non-sequence specific DNA binding. The present study presents the characterization of the non-specific DNA binding of Sac7d and the investigation of potential indirect specificity arising from protein binding induced DNA bending.

1.7 Sac7d-DNA binding system

Sac7d-DNA binding system has been used for the present study due to the non-specific protein-DNA binding and protein induced DNA binding properties. The details of Sac7d-DNA binding are described below.

1.7.1 General background

Sulfolobus, a hyperthermophile crenarcheote expresses a number of small DNA binding proteins that are considered to be associated with DNA packaging and gene regulation (Lurz, Grote et al. 1986; Choli, Wittmann-Liebold et al. 1988; Edmondson and Shriver 2001). Based on the molecular weight, the DNA binding proteins of *Sulfolobus* are categorized into 7 kDa, 8 kDa and 10 kDa proteins. The 7kDa proteins of *Sulfolobus* are collectively referred to as Sul7 proteins (White and Bell 2002). Sac7d (from *S.acidocaldarious*) and Sso7d (from *S.solfactaricus*) are the most extensively studied among the Sul7 members. They are stable over a wide range of conditions of temperature, pH and salt concentration, making them advantageous systems for biophysical studies.

1.7.2 DNA binding and bending

The structure and thermodynamics of Sac7d and its DNA binding has been characterized. It is a 7.6 kDa protein with high thermal stability. It binds without sequences specificity to the minor groove of the DNA with an average binding site size of four base pairs. Sac7d-DNA crystal structure (1azp) (figure 1.1) indicates that Sac7d binding introduces a significant kink in the DNA backbone due to the insertion of the

side chains of two amino acids, viz. Val 26 and Met 29, into the DNA. The significant kinking and unwinding of DNA was shown to contribute to the ΔCp of binding. The binding is entropy driven and the unfavorable enthalpy of binding is attributed to the ionic interaction associated with the binding and the cost of DNA distortion (Shriver, Peters et al. 2001; Peters, Edmondson et al. 2004; Peters, Edmondson et al. 2005).

Based on the high expression level and significant DNA bending and distortion caused by Sac7d, it is believed to serve a DNA packaging function in cells. However, there is no clear understanding about the packaging of multiple Sac7d molecules onto extended duplex DNA. Small angle X-ray scattering (SAXS) analysis of Sac7d-DNA complex (multiple Sac7d molecules bound to 32 base pair DNA) (figure 1.2) indicated a head-to-tail packaging of adjacent proteins (Krueger, McCrary et al. 1999) . The structure indicates that the adjacent Sac7d molecules bend the DNA in opposite directions resulting in an over all zigzag topology with negligible packaging. However, the NMR solution structure of Sso7d-DNA complex indicates that two molecules of Sso7d are bound to a GC rich 12mer DNA in a two-fold symmetry with a preferential head to head orientation (Agback, Baumann et al. 1998). Notably unlike the Sac7d-DNA crystal structure, the NMR structure of Sso7d-DNA complex does not show significant DNA bending. It is not clear if this ambiguity is either due to the head-to-head packaging or due to limitations of NMR (e.g. fast averaging of constraints).

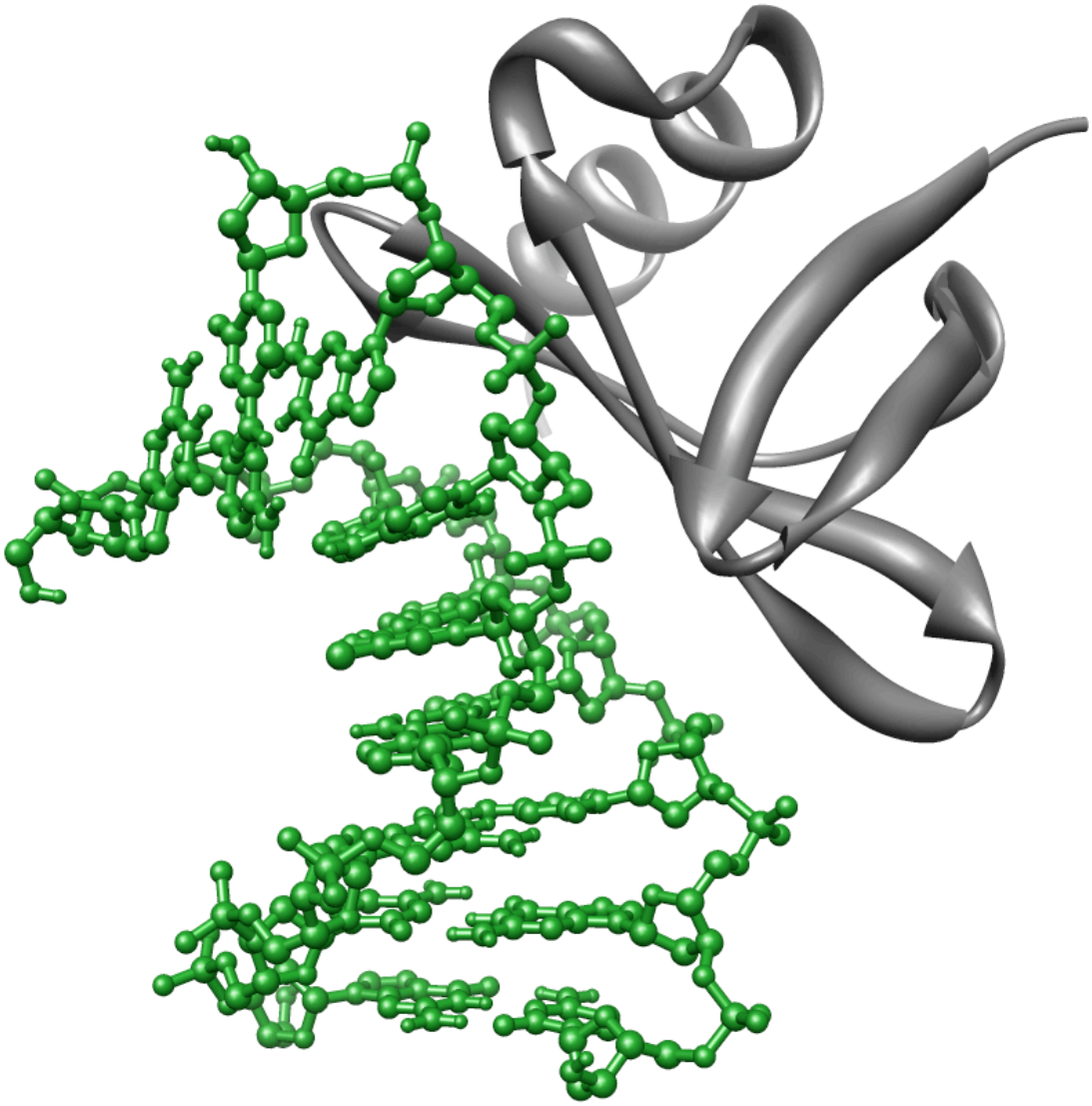


Figure 1.1 Crystal structure of Sac7d (grey) bound to 8 base pair DNA (green) (5'-GCGATCGC) (1azp).

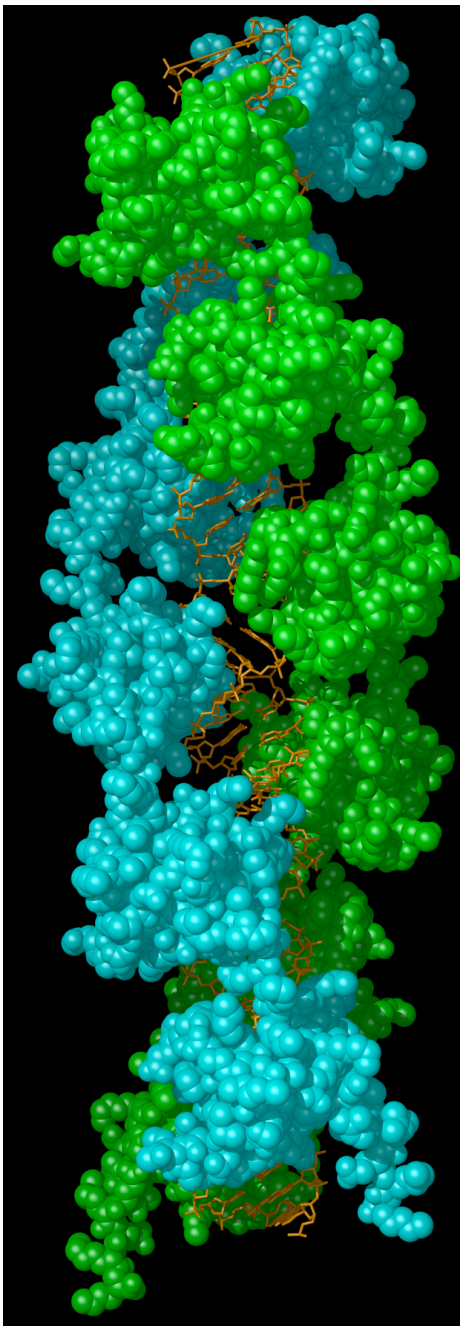


Figure 1.2 Regular head-to-tail packaging of Sac7d molecules on to 32-mer [d(GC)] modeled from the 1azp coordinates

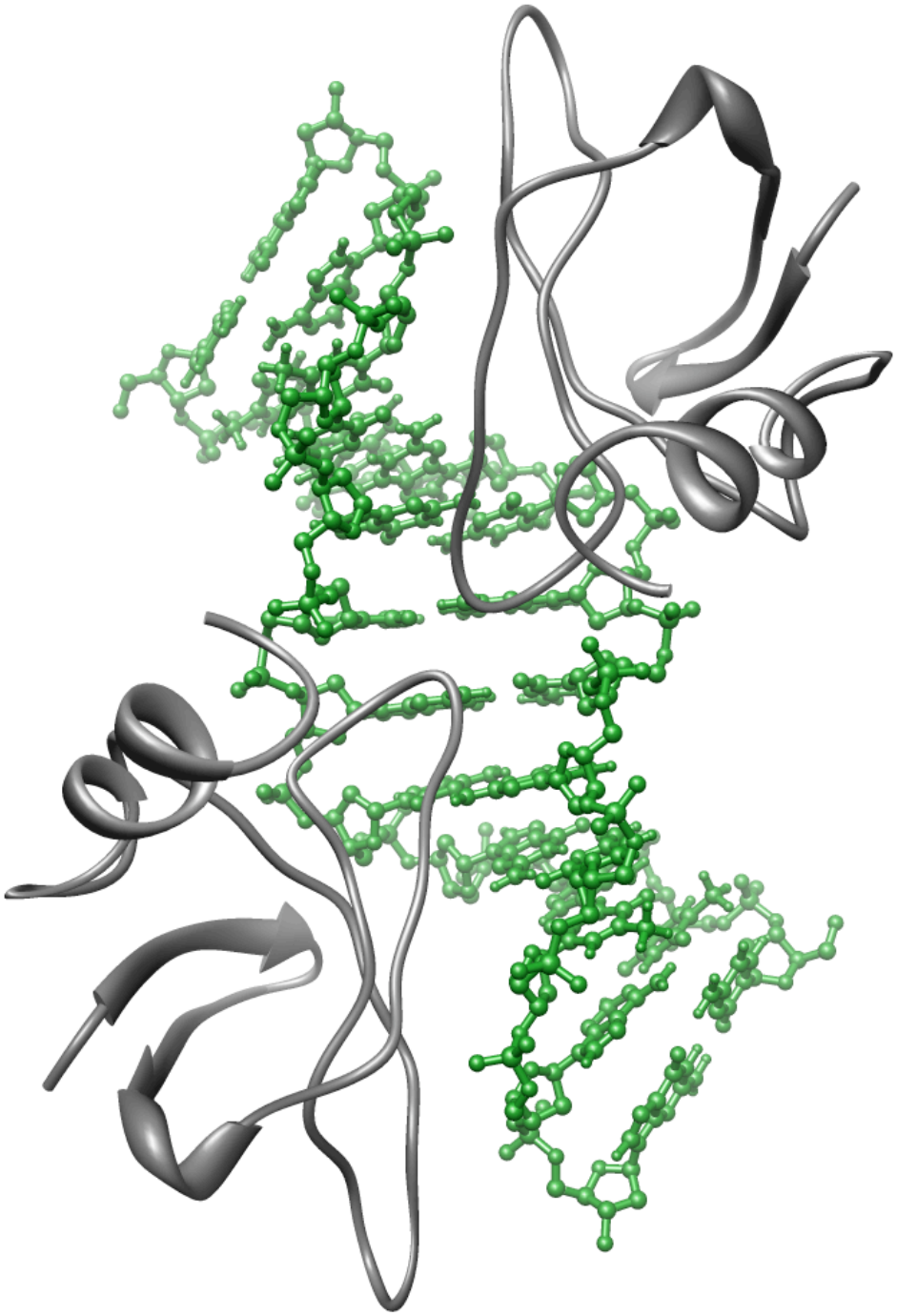


Figure 1.3 NMR solution structure depicting two Sso7d molecules bound in preferential head-to-head orientation on a 12-base pair DNA (5’-CTAGCGCGCTAG-3’)(1bbx).

1.7.3. Rationale for the present work

The purpose of this study is to carefully investigate if there is a preferred packing orientation of Sac7d on DNA. It could be expected that Sac7d binding induced DNA distortion could influence binding orientation of subsequent Sac7d molecule that binds at the adjacent binding site on the DNA. We are interested in knowing if there is potential role of DNA bending in influencing the relative binding orientation of two adjacent Sac7d molecules on the DNA.

The change in the binding orientation of the adjacent Sac7d molecules varies the distance between the successive kinks in the DNA thereby affecting the overall DNA packaging. It is not known if the DNA distortion introduced by the binding of a Sac7d molecule favors a set orientation for the next adjacent protein molecule resulting in DNA induced preferred binding orientation that translates to indirect read out mechanisms. Reverse titrations (DNA into protein) of Sac7d with DNA followed by W24 fluorescence quenching showed no evidence for cooperative interactions between proteins. However, forward titrations (protein into DNA) followed by circular dichroism of the DNA demonstrated a cooperative unwinding of the DNA with increasing binding density. Initially, at low binding density, the CD showed negligible unwinding, but with increasing binding density there was a sigmoidal increase in unwinding. This indicates that the energetics of structural changes induced in the DNA were influenced by the proximity of bound species, presumably due to reinforcement of unfavorable unwinding by neighboring species (McAfee, Edmondson et al. 1996). Neither a quantitative nor a structural description of this apparent cooperative effect is available. It is unclear if such

an effect might promote a preferred orientation for packing Sac7d onto DNA. Also, the information about the binding orientation of adjacent Sac7d molecules on DNA from SAXS and NMR structure do not agree.

In order to more thoroughly investigate this problem of DNA bending induced cooperativity and packing of Sac7d onto DNA, it is necessary to study of the binding of multiple Sac7d molecules onto DNA and investigate the possibility of preferential binding orientation between adjacent Sac7d molecules that could potentially stem from protein induced DNA bending. In this work, we present the investigation of the potential preferential binding orientation of two Sac7d molecules that bind at adjacent sites on the DNA using multiple approaches. We present the DNA binding analysis of Sac7d dimers cross-linked in different orientations and non-cross linked Sac7d molecules to understand the determinants of potential specificity in the binding orientation.

CHAPTER 2

ENGINEERED Sac7d DIMERS – CROSSLINKED IN DIFFERENT ORIENTATIONS

2.1 Introduction

Many DNA binding proteins bind to duplex DNA as homo or heterodimers (Nadassy, Wodak et al. 1999; Luscombe, Austin et al. 2000; Jones, Montcouquiol et al. 2006). Examples include the lac repressor, λ repressor, IHF, HU, CAP, GCN4, Myc, Max and Mad. The increased binding site size that results from dimerization leads to both increased affinity and specificity. Increased affinity results from the decreased entropic penalty required for the binding of the second site once the first interaction has been made. Dimerization has been promoted by design to increase the affinity of the LacI and λ repressor fragments (Falcon and Matthews 2001). In both cases, site-directed mutagenesis was used to introduce cysteine residues at the known dimer interface, which were expected to stabilize the dimer and remove the entropic penalty. λ repressor showed a 10-fold enhancement in affinity as a result of covalent crosslinking, and the LacI repressor dimer showed a10^4-fold higher affinity for DNA.

In order to study the potential preference in binding orientation for adjacent Sac7d on DNA it is necessary to study the binding of two adjacent proteins on DNA simultaneously. Appropriate cross-linking of two Sac7d molecules provides a convenient method to enforce proximity in defined orientations and study Sac7d-DNA binding at adjacent binding sites on DNA. Cross-linking in different possible orientations should help in identifying any potential preference in orientation between the two proteins that bind at adjacent sites on DNA. The dimer that is cross-linked in the potential preferred orientation (should one exist) should exhibit increased binding affinity compared the Sac7d monomer. Increased affinity results from the decreased entropic penalty required for the binding of the second site once the first interaction has been made. Native Sac7d contains no cysteines, so introduction of cysteines at specific sites through mutagenesis should provide a convenient means to crosslink monomers in a controlled manner using oxidation or bi-functional crosslinking reagents. This technique facilitates engineering dimers of Sac7d that are connected in defined orientations.

Sac7d contains the conserved chromo domain fold that belongs to the OB (*O*ligonucleotide/*O*ligosaccharide *B*inding)-fold class found in many DNA binding proteins both in prokaryotes and in eukaryotes (Murzin 1993). The chromo domain is a conserved fold with approximately 50 amino acids and is characterized by three β-strands packed against an α-helix. NMR solution structure of Sac7d (1sap) (Edmondson, Qiu et al. 1995) and the crystal structure of Sac7d-DNA complex (1azp) (Robinson, Gao et al. 1998) indicate that K9 in the β-turn between the first and the second strands of the N-terminal β-ribbon appears as a "tail" extending away from the globular body of the protein, with the head being the opposite end containing intercalating V26 and V29

residues. The ‘head’ and the ‘tail’ region are clearly pointed out in Figure 2.1. Selective mutation of single amino acid residues in the “head" and “tail” regions to cysteines provide cross-linking sites to form dimers with defined orientation of the monomer.

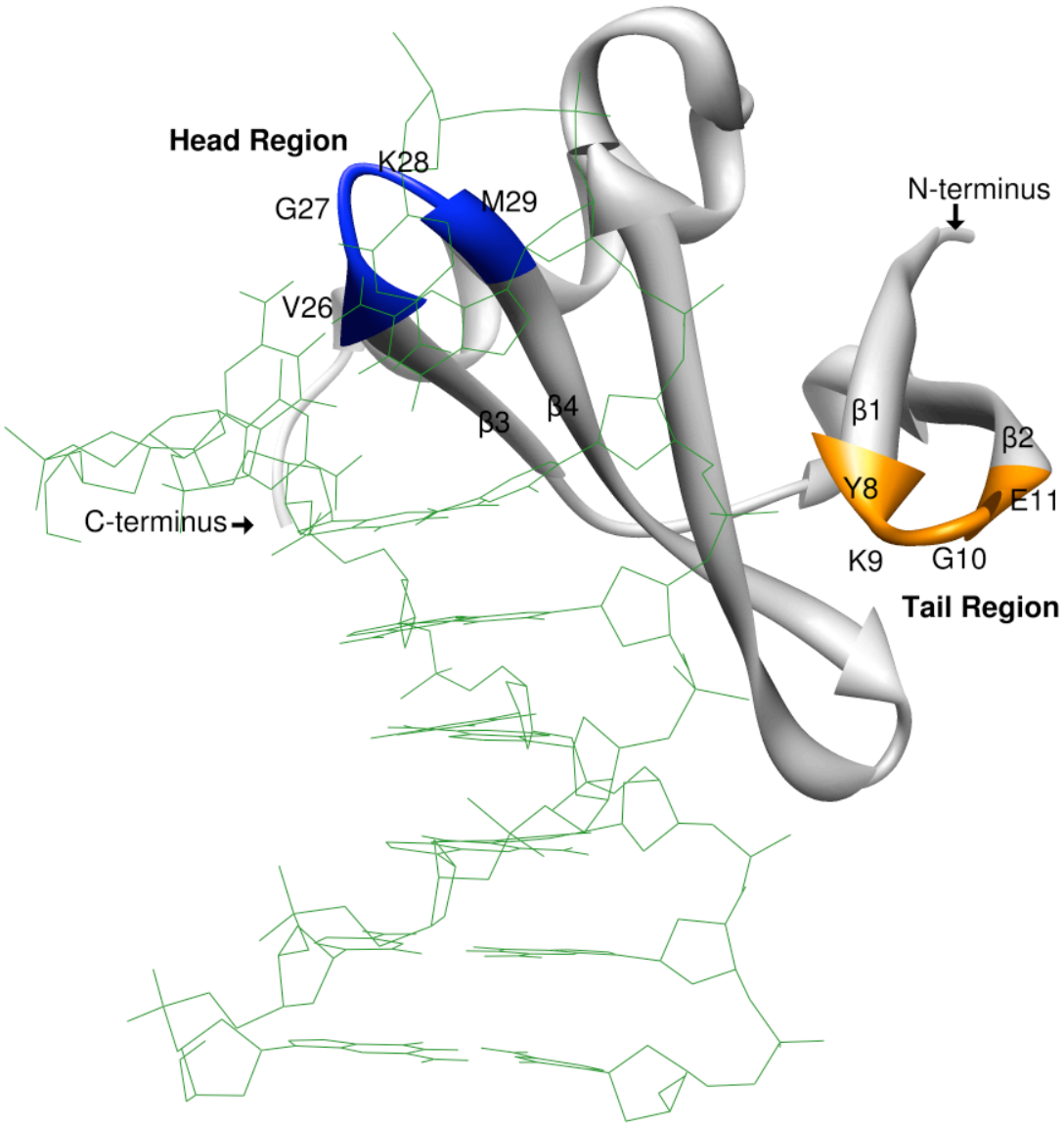


Figure 2.1 Sac7d structure depicting the ‘head’ (blue) and ‘tail’ (orange) regions

The dimers with monomers arranged in head-to-head, tail-to-tail and head-to-tail can be used to study and compare the change in DNA binding properties with change in the binding orientation of adjacent protein molecules on DNA.

2.2 Model building and mutagenesis

Models of B-DNA with two Sac7d molecules bound in different orientations were used to select the amino acid residues for cysteine mutations. B-DNA segments and crystal structure coordinates of Sac7d-DNA complex (1azp) (Robinson, Gao et al. 1998) were used to construct the models in Chimera (Sanner, Olson et al. 1996; Pettersen, Goddard et al. 2004). Adjacent Sac7d molecules were placed on B-DNA according to the coordinates from the crystal structure of Sac7d-DNA complex (1azp). Kinks in the DNA caused by dimer binding were also modeled from constraints from 1azp. Three different models were constructed with adjacent Sac7d molecules oriented in head-to-head (Figure 2.2), tail-to-tail (Figure 2.3) and head-to-tail (Figure 2.4) on DNA. Based on the models, the amino acid residues that are in closest proximity in different orientations were selected for site directed mutagenesis. Figure 2.2 with the two adjacent Sac7d molecules in head-to-head orientation shows K28 and G27 in the head region of the two adjacent proteins in proximity. Similarly, Figure 2.3 with tail-to-tail orientation shows K9 and G10 of the two adjacent proteins in proximity. K28 and G27 in the head region and K9 and G10 in the tail region have been selectively and separately mutated to cysteine to create single cysteine mutants of Sac7d. Site-directed mutagenesis was performed as described in Appendix A.1.9 and the mutations were confirmed using sequence analysis.

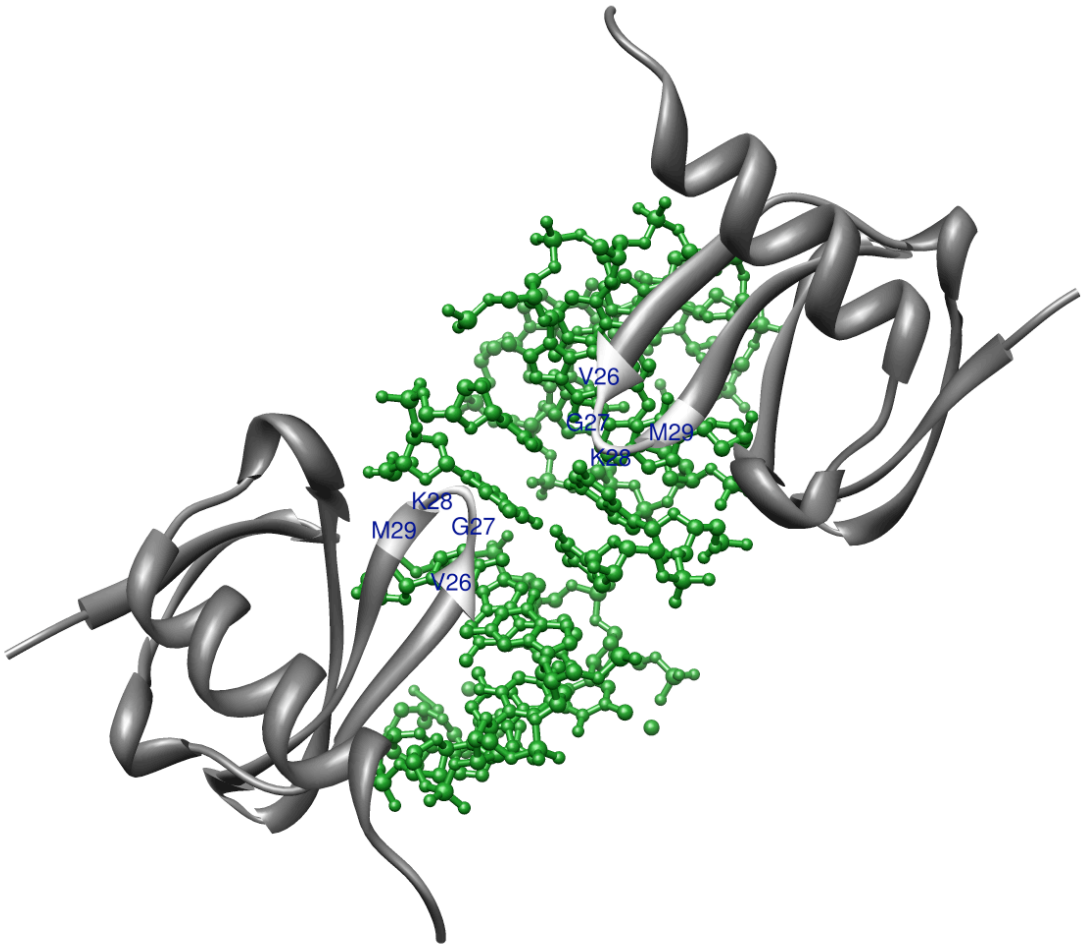


Figure 2.2 A head-to-head model showing two Sac7d molecules (grey ribbon models) placed in head-to-head orientation on B-DNA (green ball and stick model). The head region residues are labeled.

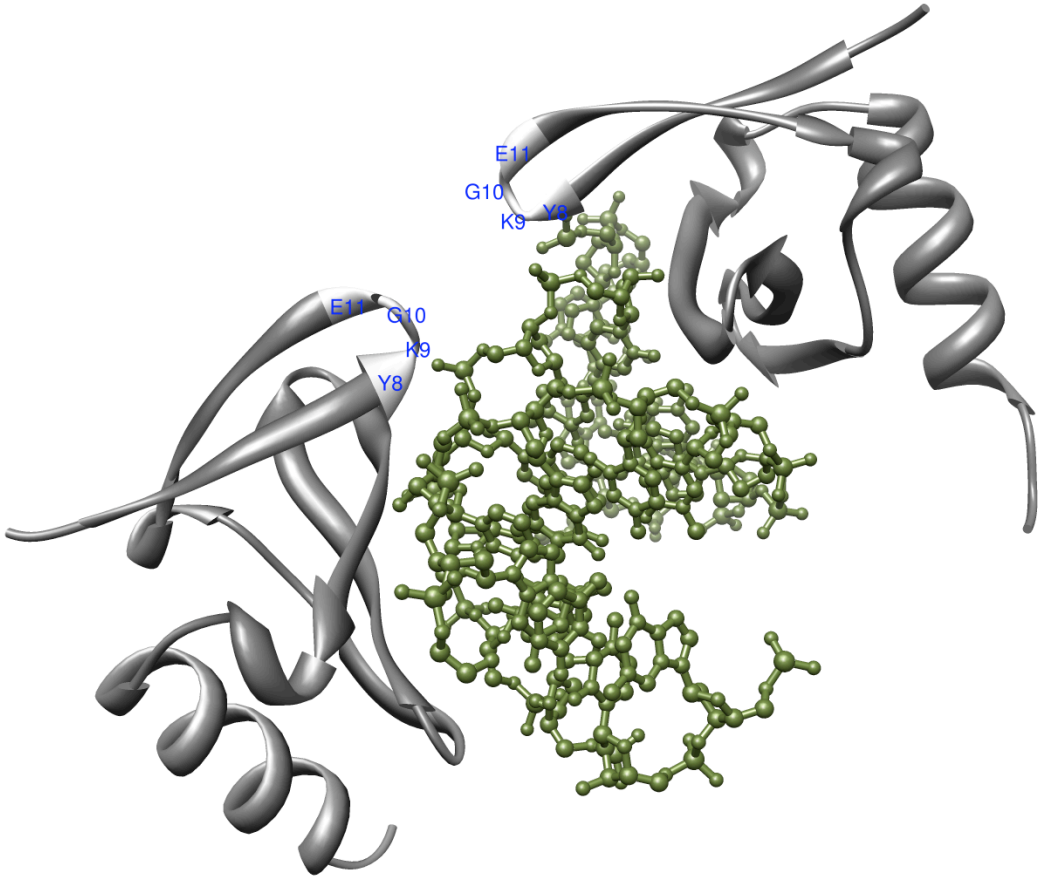


Figure 2.3 A tail-to-tail model showing two Sac7d molecules (grey ribbon models) placed in tail-to-tail (tail residues in light grey) orientation on B-DNA (green ball and stick model). The tail region residues are labeled.

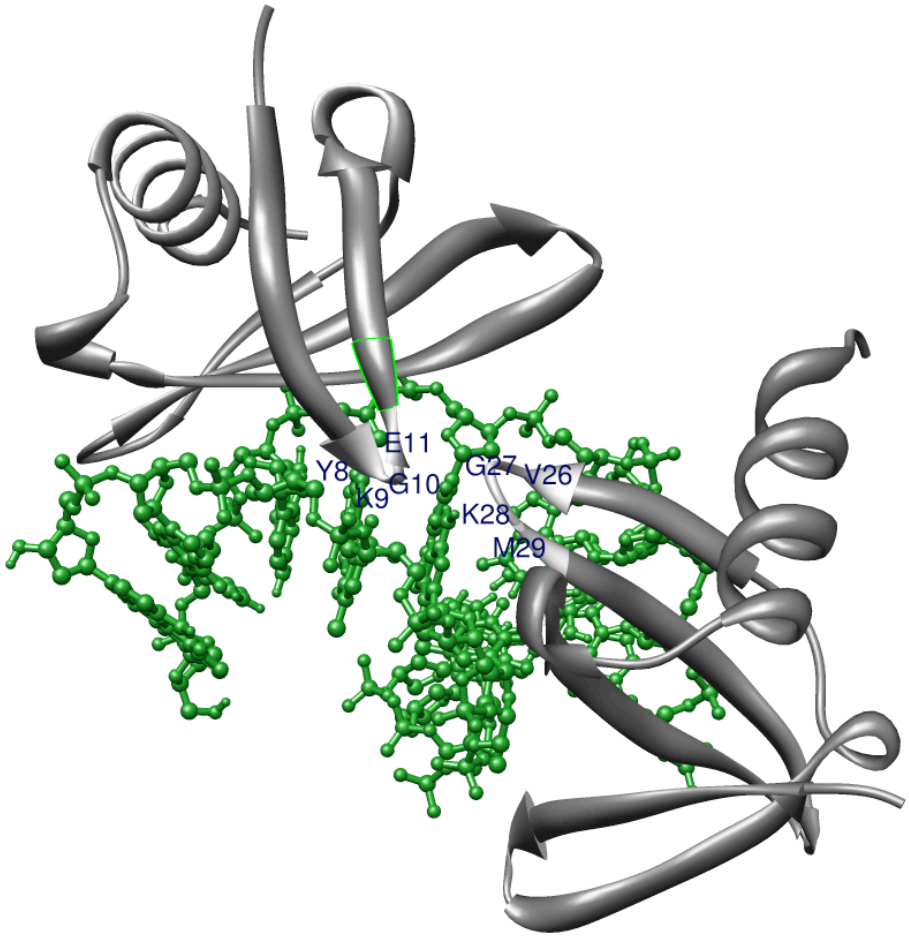


Figure 2.4 A head-to-tail model showing two Sac7d molecules (grey ribbon models) placed in head-to-tail orientation on B-DNA (green ball and stick model). The tail region of the protein on the top points to the head region of the protein at the bottom. The tail and head regions residues are labeled.

2.3 Preparation of the mutant proteins

Sac7d mutant proteins have been expressed and purified according to the method previously described for recombinant Sac7d purification (appendix A.1.2). The final protein output obtained from cation-exchange chromatography using 1 M NaCl gradient showed two well resolved peaks (Figure 2.5) as opposed to a single peak observed in the preparation of recombinant Sac7d. Based on sodium dodecyl sulfate-polyacrylamide electrophoresis (SDS-PAGE) (Figure 2.6), the two peaks were identified as monomer mutant Sac7d and disulfide cross-linked dimer respectively. Similar chromatography profiles were observed with all the mutants under study, with the dimer always eluting at higher salt concentration, i.e., more tightly bound to the ion exchange column.

2.4 Characterization of the Sac7d mutants

Biochemical, biophysical and structural characterization of the Sac7d mutants was conducted to study the effect of mutations on the secondary structure and the DNA binding properties of the protein. The effect of mutation on the secondary structure of the protein has been investigated using circular dichroism (CD) (Figure 2.7). ^{15}N-HSQC of the mutant proteins did not show any major differences in chemical shifts when compared to those of Sac7d. Chemical shift values of the peaks around the specific mutation were altered and negligible difference was observed in the remaining region. Mutations also did not alter the DNA binding properties of the protein (monomer), studied by reverse titrations (DNA into protein) in the presence of reducing agent (20 mM DTT) followed by monitoring the quenching of W24 using fluorescence spectroscopy (Figure 2.8).

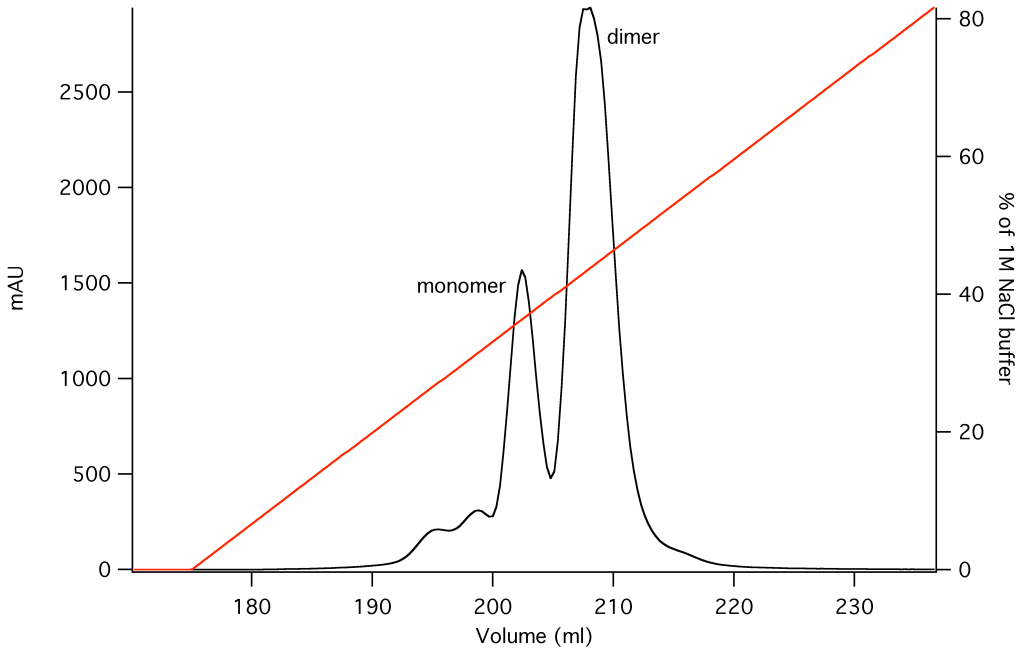


Figure 2.5 Cation-exchange chromatography elution profile showing two distinct peaks corresponding to monomer and dimer Protein flow is indicated in black. Red line indicates 0.0M to 1.0 M NaCl gradient. The small peak between the monomer and dimer was consistently observed in all the preparation and it ran as monomer Sac7d on SDS-PAGE. The cysteine mutation may be responsible for change in affinity to the cation-exchange column and salt dependence of affinity due to differential exposure of amino acid side chains responsible for binding.

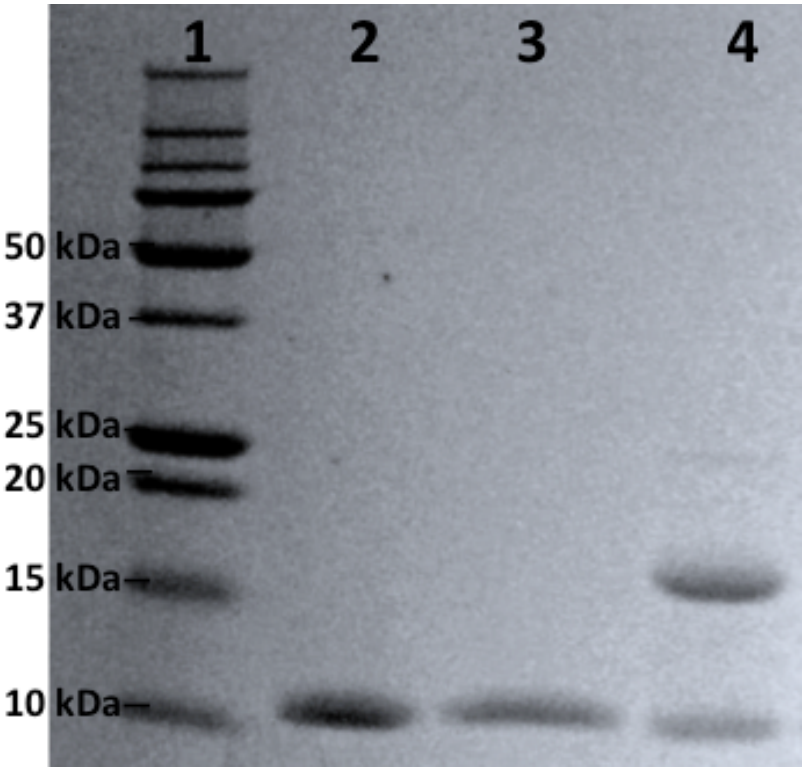


Figure 2.6 SDS-PAGE image showing monomer and dimer mixture obtained from cation-exchange column for mutant protein preparation. The gel contents are: Lane 1: molecular weight ladder, lane 2: recombinant Sac7d, lane 3: Sac7d cysteine mutant preparation with DTT, lane 4: Sac7d cysteine mutant preparation from the cation-exchange column without DTT. Sac7d runs at 10kDa in a 16.0% Tris-Tricine SDS-PAGE

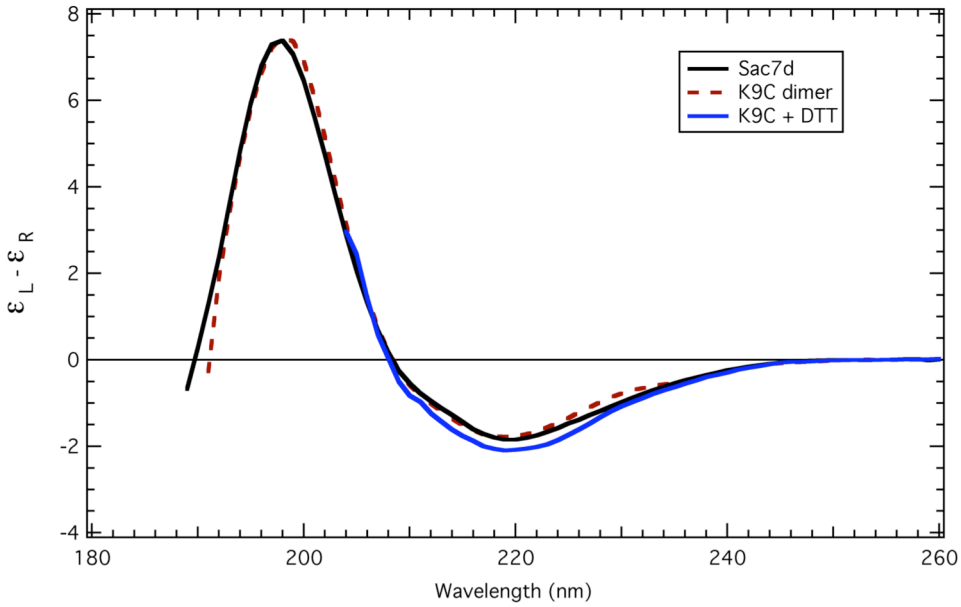


Figure 2.7 CD curves for Sac7d (black bold curve), K9C dimer (dotted red curve) K9C with DTT (blue curve) show good agreement indicating that the mutations and dimerization do not affect the secondary structure of the proteins.

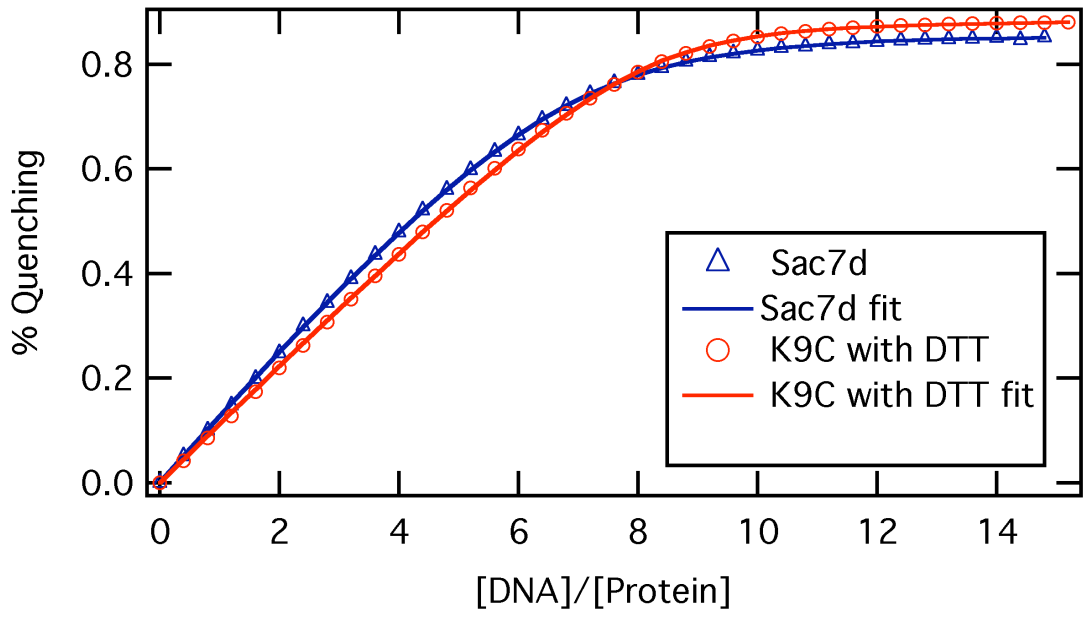


Figure 2.8 Over lay of binding curves for Sac7d monomer and K9C monomer (tail cysteine mutant with DTT) in 10 mM KH_2PO_4, pH=7.2 using 100-mer [d(GC)]. The binding parameters for the curves are shown in the table below (table 2.1). Quenching is defined as (F_i-F_0/F_0). DNA/protein ratio is the number of DNA nucleotides (mM) per protein (mM of monomer).

	Binding constant $Kx10^{-7}$ (M^{-1})	Site size (bp)	Qmax
Sac7d	1.227±0.056	2.933±0.0062	0.9047±0.007
K9C with DTT	3.43±0.23	3.513±0.013	0.8876±0.0012

Table 2.1 Comparison of binding parameters for Sac7d and K9C with DTT. Both the titrations were conducted in 10 mM KH_2PO_4, pH=7.2 using 100-mer [d(GC)]. Q_{max} is the maximum fluorescence quenching. The Reverse titrations were conducted using the general procedure described in appendix A.2.5

2.5 Engineering Sac7d dimers

The Sac7d cysteine mutants were subjected to cross-linking conditions either by oxidation using diamide or by using bi-functional cross-linking agents specific for the sulfhydryl group.

2.5.1 Oxidation

Dimerization of cysteine mutants of Sac7d was achieved by treatment with diamide (Appendix D.1). Diamide is an oxidizing agent that has been shown to cause disulfide bond formation in proteins with free cysteine groups. It was first reported for glutathione oxidation (Kosower, Kosower et al. 1969). Optimum reagent concentration and reaction conditions were achieved by conducting a series of trails by varying diamide concentration and reaction time. Optimum reaction conditions and the method are described in appendix A.1.11. A 0.1 mM final concentration of diamide resulted in almost complete dimerization. Figure 2.9 shows the SDS-PAGE image of dimers obtained by diamide-induced oxidation.

2.5.2 Chemical cross-linking

Sac7d cysteine mutants were also cross-linked with bi-functional cross-linking agents with different cross-linker lengths that are specific for sulfhydryl groups. Table 2.2 shows the different cross linkers used in the study and their cross-linking lengths. The purpose of engineering Sac7d dimers using cross-linkers was to provide increased distance and flexibility between the two monomers when binding at adjacent binding

sites on DNA. A flexible cross-linker also eliminates any possible constraints that might otherwise be imposed due to S-S cross-linking.

Sac7d dimers were engineered according to the procedure described in appendix A.1.12. The chemical cross-linking reaction did not result in complete dimerization. Residual monomer was separated from the desired dimer fraction by cation-exchange chromatography using a 1 M NaCl gradient. The chromatogram representing the separation of desired dimer fraction for the monomer is shown the in Figure 2.10. Dimer formation is conformed using SDS- PAGE electrophoresis as shown in Figure 2.9.

Cross-linker	Cross-linker length
BMBD	10.2 Å
BMB	10.9Å
pPDM	12.2 Å
BMH	13.0 Å
BM(PEG)2	14.7 Å

Table 2.2 List of bi-functional cross linker used in the study. Appendix D.1 lists the full names of the cross-linkers mentioned here.

2.6 Cross-linking Sac7d in head-to-tail orientation

Unlike the creation of head-to-head or tail-to-tail Sac7d dimers, creation of Sac7d dimers cross-linked in a head-to-tail orientation was not straightforward. Subjecting an equimolar mixture of head-cysteine mutant monomer and tail-cysteine mutant monomer to cross-linking conditions results in a mixture of dimers with head-to-head, tail-to-tail and head-to-tail orientations. Preliminary experiments showed that the dimers in three different orientations could not be separated by ion exchange or affinity chromatography on a preparative scale. In order to selectively purify the mutant dimers with three different orientations, an additional differentiating parameter was necessary. Introduction

of a terminal His-tag to one of the mutant monomers provides a key additional tool to separate the Sac7d dimers cross-linked in different orientations.

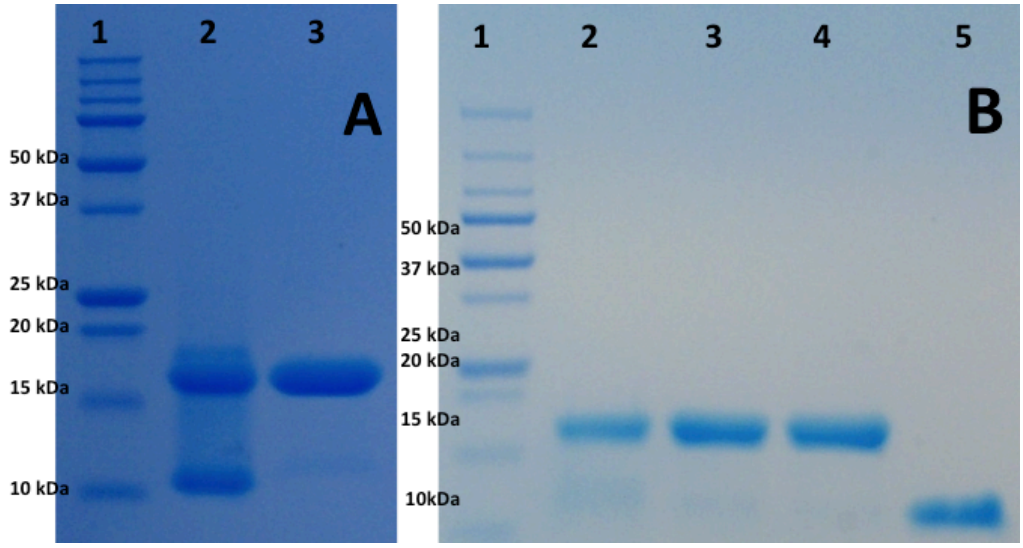


Figure 2.9 SDS-PAGE of Sac7d dimers (K9C dimers). Panel-A: Diamide induced oxidation: Lane 1 shows molecular weight ladder, lane 2 shows the monomer dimer mixture obtained from cation-exchange column, lane 3 shows complete dimerization obtained from diamide treatment. Panel-B: Bi-functional cross-linker induced cross-linking: Lane1 shows molecular weight ladder, Lanes 2-4 show dimer fractions that are separated (on cation-exchange column Figure 2.11) from residual monomer followed by treatment with bi-functional cross-linking agent. Lane 5 shows monomer Sac7d.

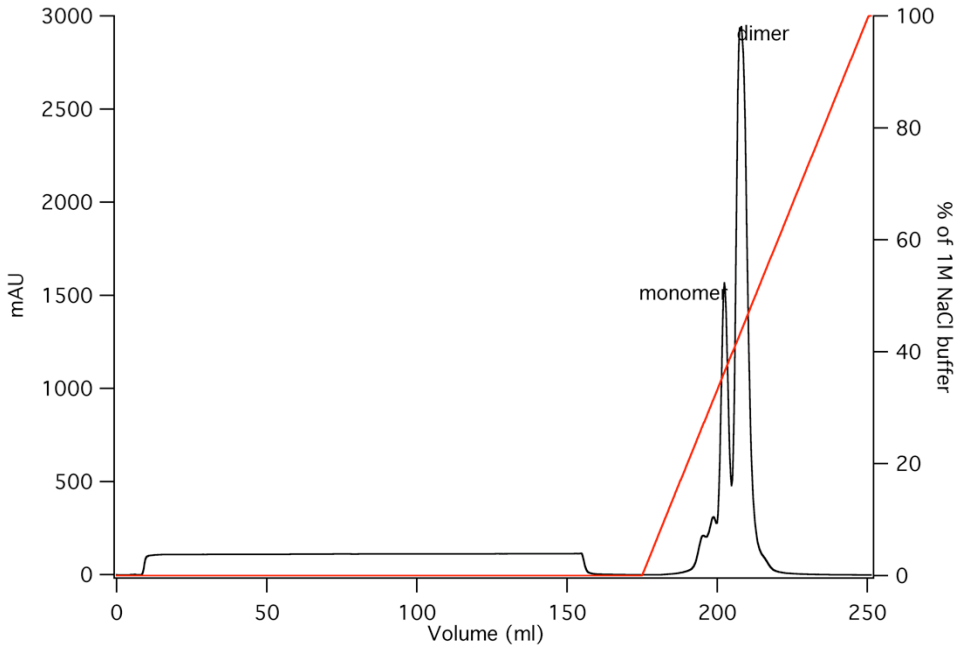


Figure 2.10 Separation of monomer and dimer fractions resulting from chemical cross-linking

The scheme presented in the Figure 2.11 describes the problem associated (panel 1) with cross-linking Sac7d in head-to-tail orientation and its solution (panel 2) by the introduction of a His-tag. By introducing a C-terminal His-tag to the tail-cysteine monomer and subjecting an equimolar mixture of monomeric head-cysteine mutant and monomeric His-tag tail-cysteine mutant results in a mixture of dimers as described previously, i.e., a mixture of head-to-head, tail-to-tail and head-to-tail. But, due to the presence of a terminal His-tag on the tail cysteine-mutant, the dimers in the three different orientations differ in the number of His-tags per dimer. Head-to-head dimer has no His-tags, tail-to-tail dimer has two His-tags and the head-to-tail dimer has one His-tag. Due to the presence of different number of His-tags, the dimers cross-linked in different orientations bind with different affinities to a Ni-column and can be separated using imidazole gradient.

2.6.1 Cloning C-terminal His-tag

C-terminal His-tag was introduced in the K9C sequence using pET-28(b) expression system. The sequence for K9C was extracted from pET-3(b) expression system using restriction enzymes *Eco*RI *and Bam*HI and introduced into pET-28(b) between the same restriction sites. The His-tag and the protein coding sequence are separated by spacer region consisting of 20 amino acids that is built into the pET-28(b) vector. The purpose of the spacer region is to provide flexibility. Cloning results were confirmed using sequence analysis.

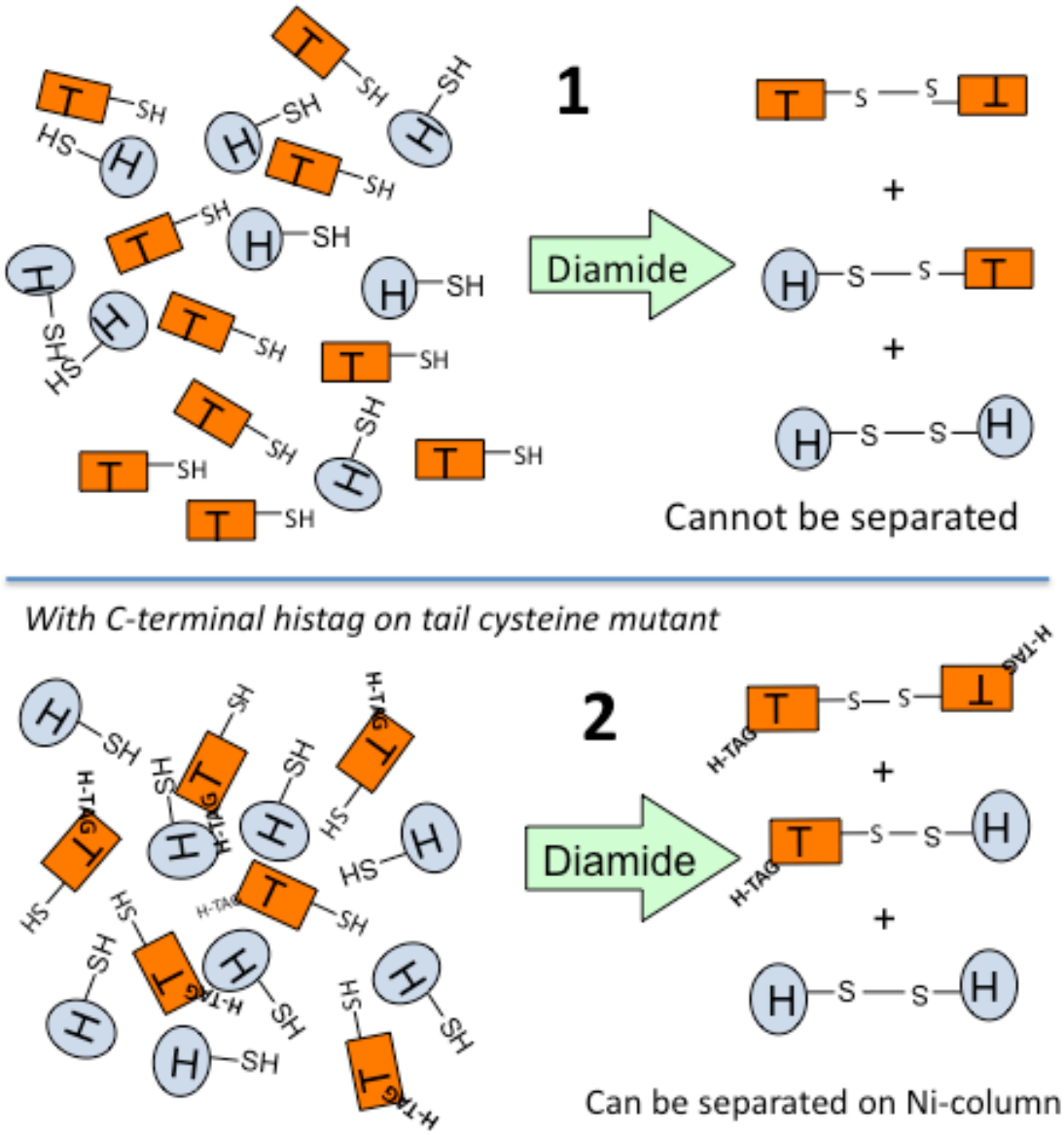


Figure 2.11 Scheme for the separation Sac7d dimers with different cross-linking orientations by the introduction of a C-terminal His-tag to the tail cysteine mutant. Panel 1 shows the formation of a mixture of dimers with three possible orientations. Panel 2 shows the dimers with different number of His-tags that can be separated on a Ni-column using imidazole gradient.

2.6.2 Expression and purification of His-tag K9C (C-terminal His-tag tail cysteine mutant)

His-tag K9C carrying pET-28(b) vector system was transformed into *E.coli* according to the transformation procedure described in the appendix A.1.10. The protein production was induced with 1 mM final concentration of IPTG. The His-tag protein was localized in the inclusion bodies and hence the protein was purified according to the general protocol for His-tag-protein purification from inclusion bodies.

2.6.3 Head-to-tail cross-linking and purification

Head-to-tail cross-linking of Sac7d was achieved by subjecting an equimolar mixture of K9C-His-tag monomer (tail mutant) and K28C monomer (head mutant) to cross-linking conditions. The resultant dimers consisted of K9C His-tag dimer (tail-to-tail) with two His-tags per dimer, K28C dimer (head-to-head) with zero His-tags per dimer and K28C-K9C His-tag dimer (head-to-tail) with one His-tag per dimer. The mixture of dimers obtained from the cross-linking reaction was separated by purification through Ni-affinity column chromatography using imidazole gradient. K28C dimer with no His-tags did not bind to the Ni-matrix and was eluted in the flow through while the sample was being loaded. The bound protein was eluted with a linear gradient from 0.0 M to 0.5 M imidazole gradient. With increasing imidazole gradient, head-to-tail and tail-to-tail dimers with one and two His-tags per dimer, respectively, were eluted as two well-resolved peaks. The fractions were confirmed using SDS-PAGE. Figure 2.12 to 2.14 represent the chromatogram profiles and SDS-PAGE pictures depicting the separation of the dimers in three different orientations. Figure 2.12 shows the SDS-PAGE image

having the mixture of three dimers. Imidazole elution profile obtained by the injection of the mixture of dimers in three different orientations on to Ni-column is depicted in Figure 2.13. Figure 2.14 shows the SDS-PAGE image for the conformation of the three dimer fractions separated using Ni-column chromatography.

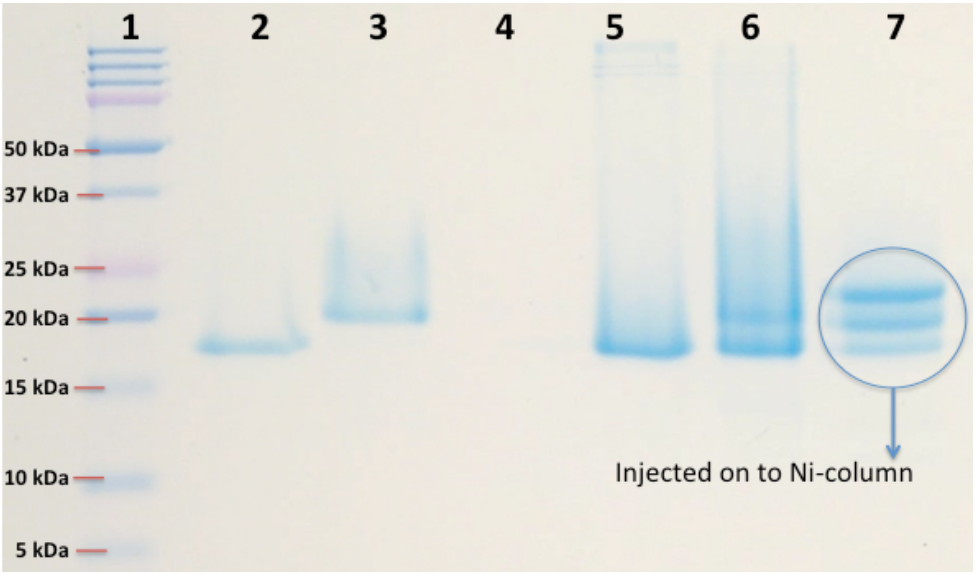


Figure 2.12 SDS-PAGE image showing the three types of dimers. Lane1: Molecular weight ladder, lanes 2,3,5,6 different dimer fractions for control purpose, Lane 4: intentionally left blank, Lane7: the mixture of three types of dimers as three distinct bands (that are loaded on to the Ni-Column)

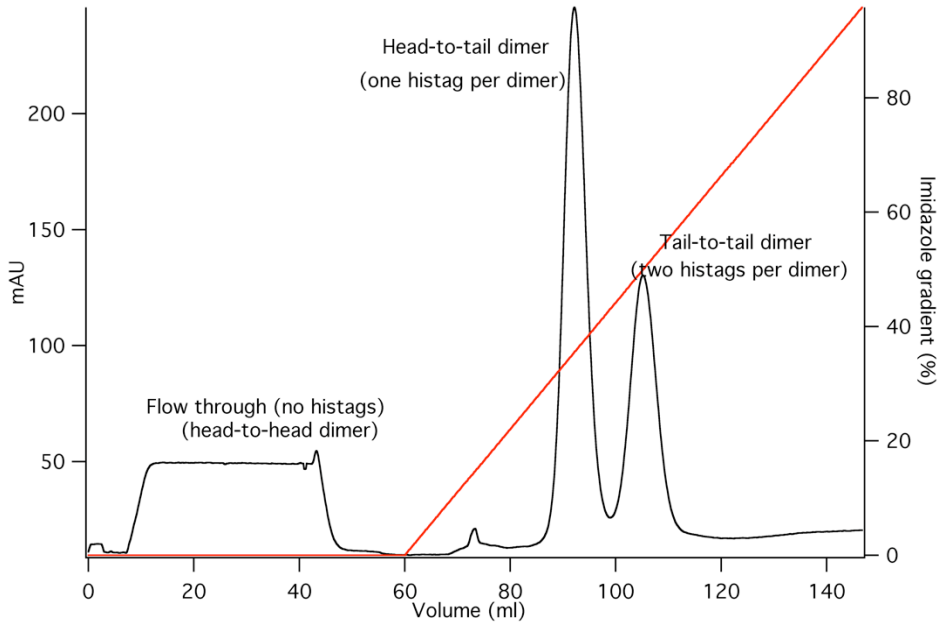


Figure 2.12 Ni-column chromatogram profile showing the separation of the K28C dimer (head-to-head) in the flow through, K9C-K28C dimer (head-to-tail) and K9C dimer (tail-to-tail) eluting the in the same order with increasing imidazole gradient (red line).

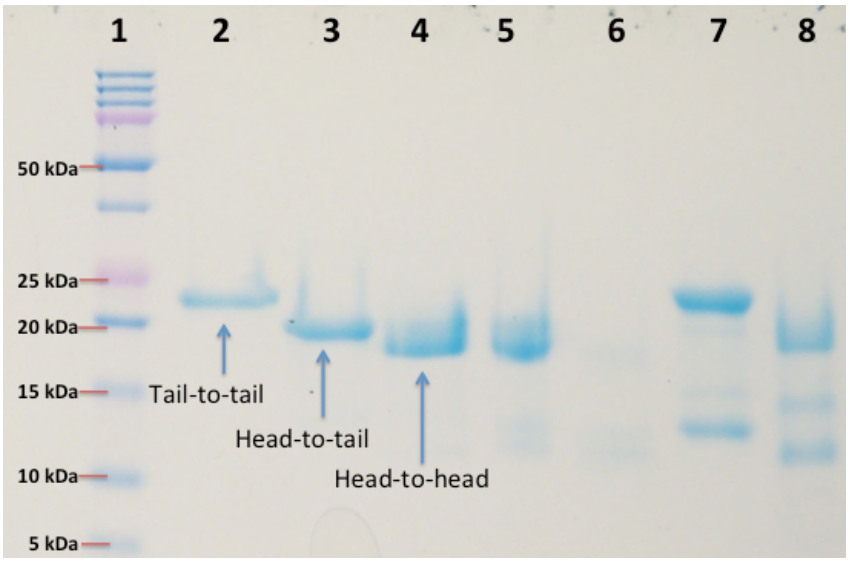


Figure 2.13 SDS-PAGE image of fractions separated through Ni-column confirming the separation of head-to-tail dimer from the other two orientations. The difference in the molecular weights for the dimers in three different orientations is due to the extra leader segment containing the His-tag and the spacer region. Tail-to-tail dimer (K9CHis-tag dimer) with two sets of extra segments is the heaviest. Head-to-head dimer (K28C dimer) with no His-tags is the lightest. Head-to-tail dimer (K9CHis-tag-K28C) with one His-tag has intermediate weight. Lane 1: molecular weight marker, lane 2: K9CHis-tag dimer, lane 3: K9CHis-tag-K28C (head-to-tail) dimer, lanes 4 and 5: K28C dimer (with no His-tags) obtained in the Ni-column flow through, lane 7: K9C His-tag preparation showing K9C His-tag monomer and K9C His-tag dimer and lane 8: K28C preparation showing K28C monomer and dimer bands and lanes 6 is intentionally left blank.

2.7 Characterization of Sac7d dimers

Sac7d dimers were characterized using electrophoresis, circular dichroism, differential scanning calorimetry and NMR spectroscopy as described below.

2.7.1 Electrophoresis

SDS-PAGE was used after cation-exchange chromatography as a preliminary confirmation for the formation of Sac7d dimers. The engineered dimer fractions were analyzed on denaturing polyacrylamide gels to confirm the formation of dimers. Under reducing conditions of electrophoresis, the dimers made with oxidation using diamide were reduced to monomers, while the dimers made with chemical cross-linkers did not. Sac7d dimers created by both by diamide-induced oxidation and chemical cross-linking were stable for months at 4 °C (previous Figures 2.5, 2.6, 2.9, 2.10, 2.11, 2.12 and 2.13).

2.7.2 Circular dichroism

Circular dichroism (CD) is sensitive to the secondary structure of proteins and peptides. CD spectroscopy is a form of light spectroscopy that measures the difference in the absorption of left and right circularly polarized light by a substance. Alpha-helices, parallel and anti-parallel beta sheets, turns and random coil have signature patterns of CD spectra between 260 nm and 180 nm (far UV region). There are many examples of the use of CD to analyze the secondary structure of proteins and there are many reviews describing the technique and applications of CD (Johnson 1988; Bloemendal and Johnson 1995; Woody, Sugeta et al. 1996).

CD spectra for Sac7d and the mutant proteins have been collected and analyzed for α-helix, β-strands, β-turns and other secondary structural content using the CDPro software package (Sreerama, Venyaminov et al. 1999). The CDPro software package consists of three programs (SELCON3, CDSSTR and CONTIN) for analyzing protein CD spectra. The analysis programs use set of reference proteins (CD and X-ray diffraction data) to calculate a CD spectrum for comparison to the experimental spectrum. Secondary structure fraction predictions are given as output once a reasonable fit is obtained between the calculated and the experimental data. The three analysis programs allow independent calculation of standard deviation between calculated spectra. Figures 2.14 through 2.16 display the wavelength scan of the three dimer orientations studied compared to Sac7d.

The CD spectra of the mutant monomers and dimers of K9C (tail mutant), K28C (head-to-head) dimer and K9CHis-tag and K28C dimer (head-to-tail) showed no significant differences when compared to that of the recombinant Sac7d monomer. This showed that cysteine mutations and the resultant dimerization did not affect the folding and the secondary structure of Sac7d.

Table 2.2 list the secondary structure fractions predicted from the CD spectrum of each dimer using CDPro software package. Values are averages of three independent data sets obtained from SELCON3, CDSSTR and CONTINLL. Error measurements are ±1 standard deviation determined from the three calculated values for each parameter.

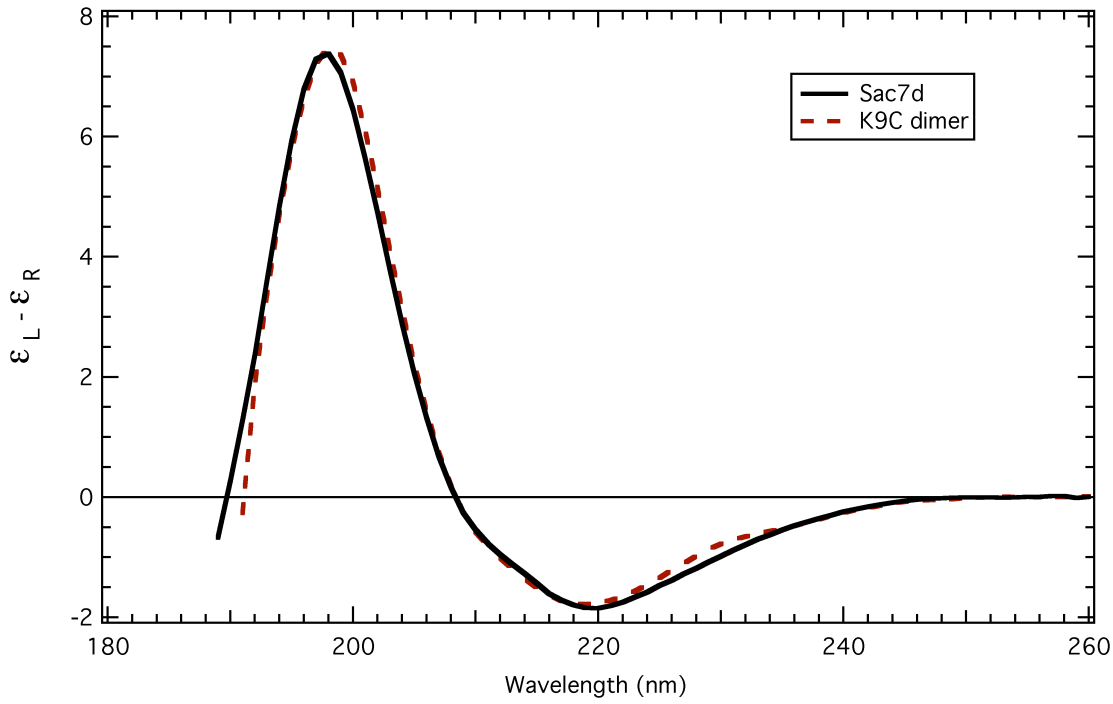


Figure 2.14 Far UV CD spectra of Sac7d monomer and K9C dimer (tail-to-tail)

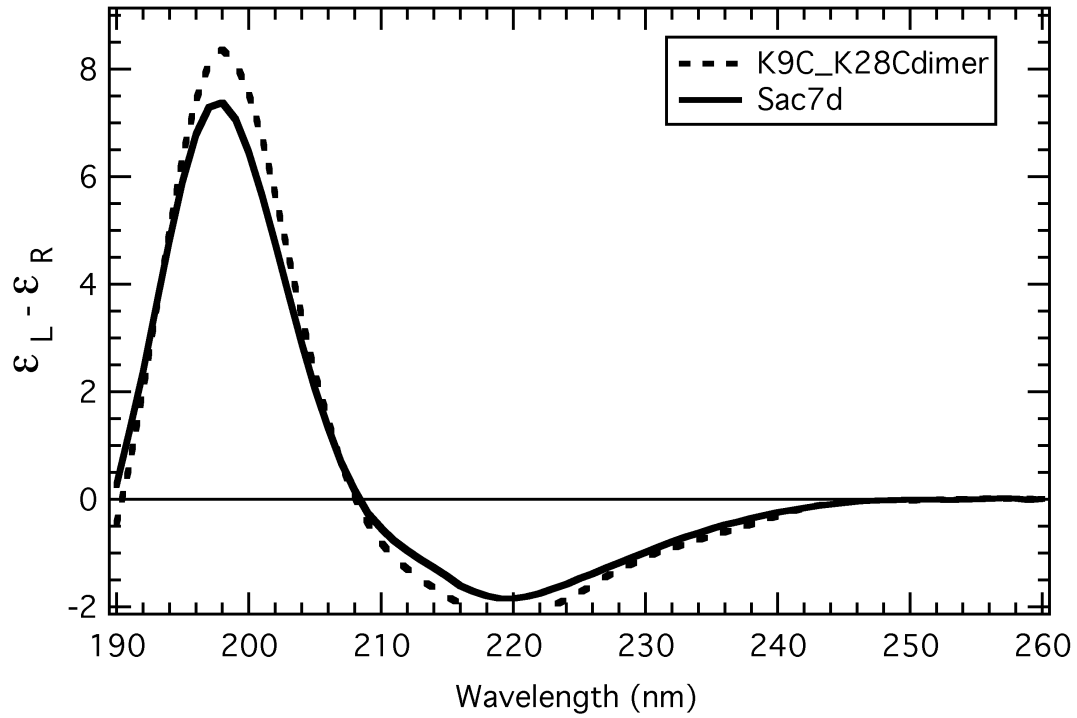


Figure 2.15 Far UV CD spectra of Sac7d monomer and K9C-K28C dimer (tail-to-head)

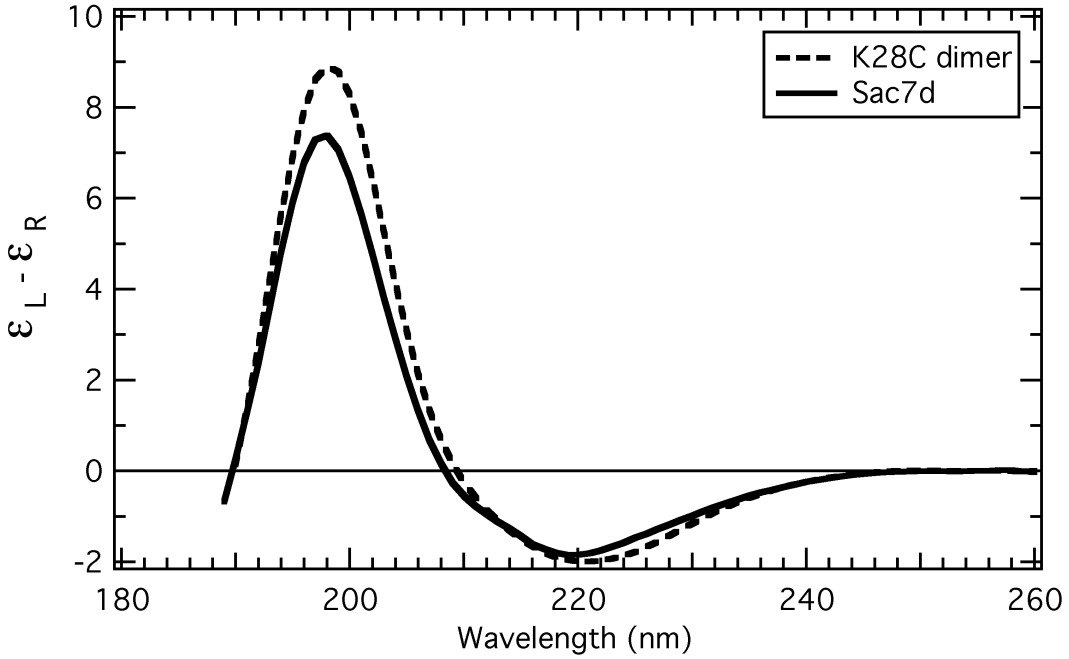


Figure 2.16 Far UV CD spectra of Sac7d monomer and K28C dimer (head-to-head)

Protein	α-helix	β-strand	β-turn	Others
Sac7d	24.2±1.9	33±1.4	16.2±1.5	26.5±4.1
K9C dimer	31.5±1.4	35.5±1.11	13.2±1.18	19.4±1.52
K28C dimer	25.05±1.3	34.7±1.18	16.15±1.53	23.725±1.41
K9ChtK28C	28.25±1.56	26.65±1.18	21.125±1.14	24.025±1.3
Sac7d*	34	36	12	18

Table 2.3 Calculated secondary structure fractions for Sac7d and Sac7d dimers in different cross-linking orientations. Structural fractions were calculated from a CD spectrum for each protein. CD spectra were collected in 0.01 M phosphate buffer, pH=7.2 at 20 ºC. Error measurements are ±1 standard deviation. *Secondary structure fractions determined from NMR solution structure of Sac7d (1sap) (Edmondson, Qiu et al. 1995).

The structure of Sac7d in solution was determined by NMR to be composed of 36% of β-strand, 12% β-turns, 23% α-helix and 11% 3_{10} helix. The CDPro analysis of the CD spectrum for Sac7d calculates 35.1±1.4% β-sheet, 22.1±1.9% α-helix, 15.9±1.5% β-turn and 26.8±4.1% other secondary structure. Values for α-helix and β-strand composition correspond well with the NMR solution structure of Sac7d (Johnson 1988). However, the software does not distinguish 3_{10}-helix from the β-turn and other secondary structure. As a result the two structural elements contribute to the calculated β-turn and other structural values.

2.7.3 Thermal stability

Differential scanning calorimetry (DSC) scans were conducted to study the thermal stability of cysteine mutants of Sac7d. DSC most readily determines the thermodynamic parameters for heat-induced unfolding of protein. For hyperthermophile proteins, DSC is particularly useful because of the ability to collect data up to 130 °C. DSC measures the heat capacity of the sample as a function of temperature. As the temperature of the sample enters a range where an unfolding transition occurs, the sample takes on an excess heat capacity because a portion of the heat transferred to the sample cell is used in unfolding the protein. The post-transition baseline of the unfolding transition does not necessarily return to the pre transition baseline level since the heat capacity of the unfolded state is higher than the heat capacity of the folded state. A typical DSC scan is characterized by a pre-transition baseline, a peak centered at the mid point temperature of the transition (T_m), and a displaced post-transitional baseline. The integrated area under the transition curve is equal to the heat ($\Delta H°$) of the unfolding

process. For a monomer, the difference in the pre and the post-transition base lines when extrapolated at the T_m represents the change in the heat capacity (ΔCp) of the unfolding.

For a more accurate measurement of ΔCp, extensive data collection involves the measurement of $\Delta H°$ values at varying conditions (pH, buffer conditions etc.). The method of obtaining ΔCp from the slope of a plot of $\Delta H°$ versus Tm at different pH conditions is called Kirchoff's analysis.

The calorimetric enthalpy is the total area under the DSC endotherm. It represents the total excess heat needed to unfold the protein. ΔH is function of the heat capacity and the temperature and standard calorimetric enthalpy is defined at the mid-point temperature. Equation 2.1 represents the expression for calorimetric enthalpy.

$$\Delta H\,(T) = \Delta H\,(T_m) - \Delta \mathrm{Cp}\,(T_m - T) \quad . \quad (2.1)$$

DSC data can be integrated to obtain a progress curve from which the temperature dependence of the equilibrium constant (K_{un}) can be obtained. The enthalpy change obtained from such a progress curve is known as van't Hoff enthalpy (ΔH_{vh}).

$$\Delta H_{vh} = -Rd\,(\ln K_{un})/d(1/T) \quad . \quad (2.2)$$

The ratio of $\Delta H_{cal} / \Delta H_{vh}$ can be used to indicate if the unfolding process is two-state unfolding or not.

Figure 2.17 shows the comparison of DSC scans for Sac7d and K9C mutant. The values for T_m and ΔH for the cysteine mutants were comparable to those obtained for

Sac7d monomer, indicating that the cysteine mutation and dimerization did not effect the stability of the protein.

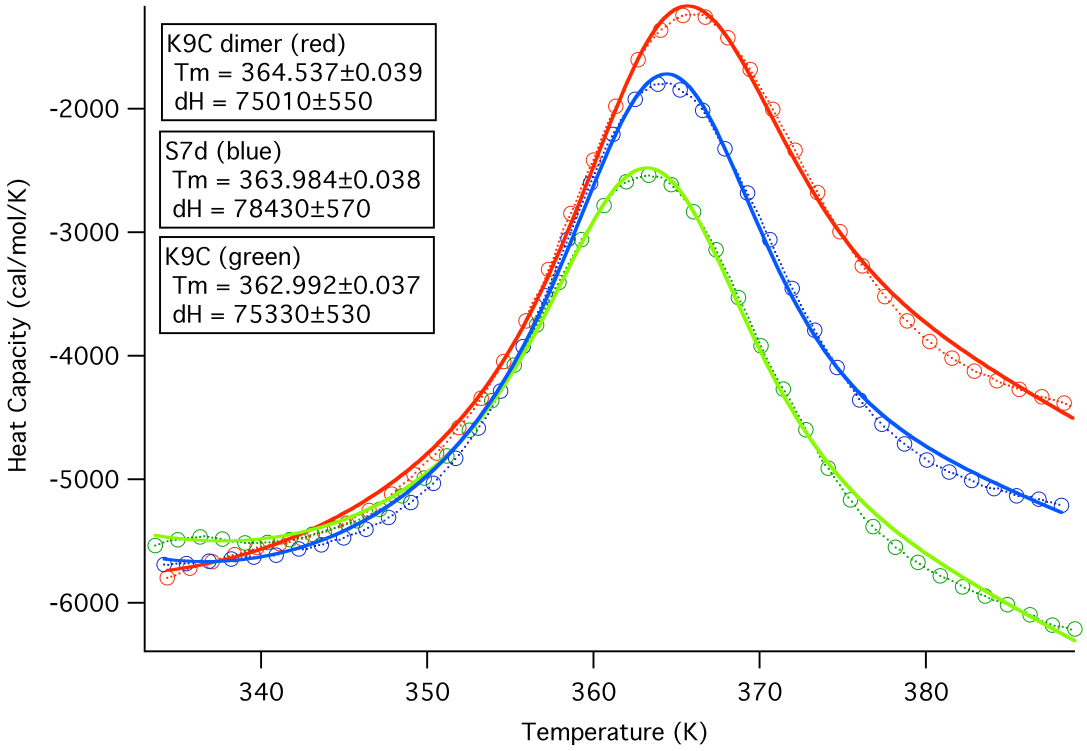


Figure 2.17 Comparison of the DCS curves for Sac7d and K9C (tail cysteine mutant). The blue curve represents Sac7d monomer, red curve represents K9C dimer and the green curve represents a mixture K9C monomer and dimer. The corresponding denaturation parameters are displayed. The solution conditions were 10mM KH_2PO_4, pH=7.2. The protein concentration was ~1mg/ml.

2.7.4 Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) spectroscopy was employed to study the structural differences in Sac7d mutants and cross-linked Sac7d dimers. Two-dimensional ^{15}N-HSQC spectra were collected for all the mutant proteins and dimers. The ^{15}N-HSQC spectrum is unique for a given protein structure and may be described as a protein

"finger-print". Resonances observed in ^{15}N-HSQC spectra arise from labeled nitrogen (^{15}N) and hydrogen (^{1}H) atoms of each amino acid. Assignments were made using NMR ViewJ (java) software package. Sample preparation, data collection and processing were done according to the procedure described in the appendices A.1.4 and A.2.4. Two-dimensional ^{15}N-HSQC spectra of K9C dimer were collected on Varian INOVA 800MHz spectrometer at 25 ºC. The data was processed using NMR Pipe and VNMRJ, and analyzed using NMRVIEW (Delaglio, Grzesiek et al. 1995; Johnson 2004). The scripts used for data processing are shown in the appendix C.

Chemical shifts for each amino acid in the mutant proteins were compared to the chemical shifts for Sac7d. Major differences in the chemical shift values were not observed with the introduction of cysteine mutations and dimerization. Figures 2.18 through 2.20 show the ^{15}N-HSQC of K9C dimer, K28C dimer and K9CHis-tag-K28C dimer. In the three spectra, black peaks correspond to the Sac7d and red peaks correspond to the respective mutant dimers. Comparison of the ^{15}N-HSQC spectra of Sac7d and the mutant dimers indicate very little difference in the chemical shifts, indicating little influence of cysteine mutations and dimerization on the structure of Sac7d. Extra peaks corresponding to the His-tag regions are observed in the ^{15}N-HSQC of K9CHis-tag-K28C dimer. Fi

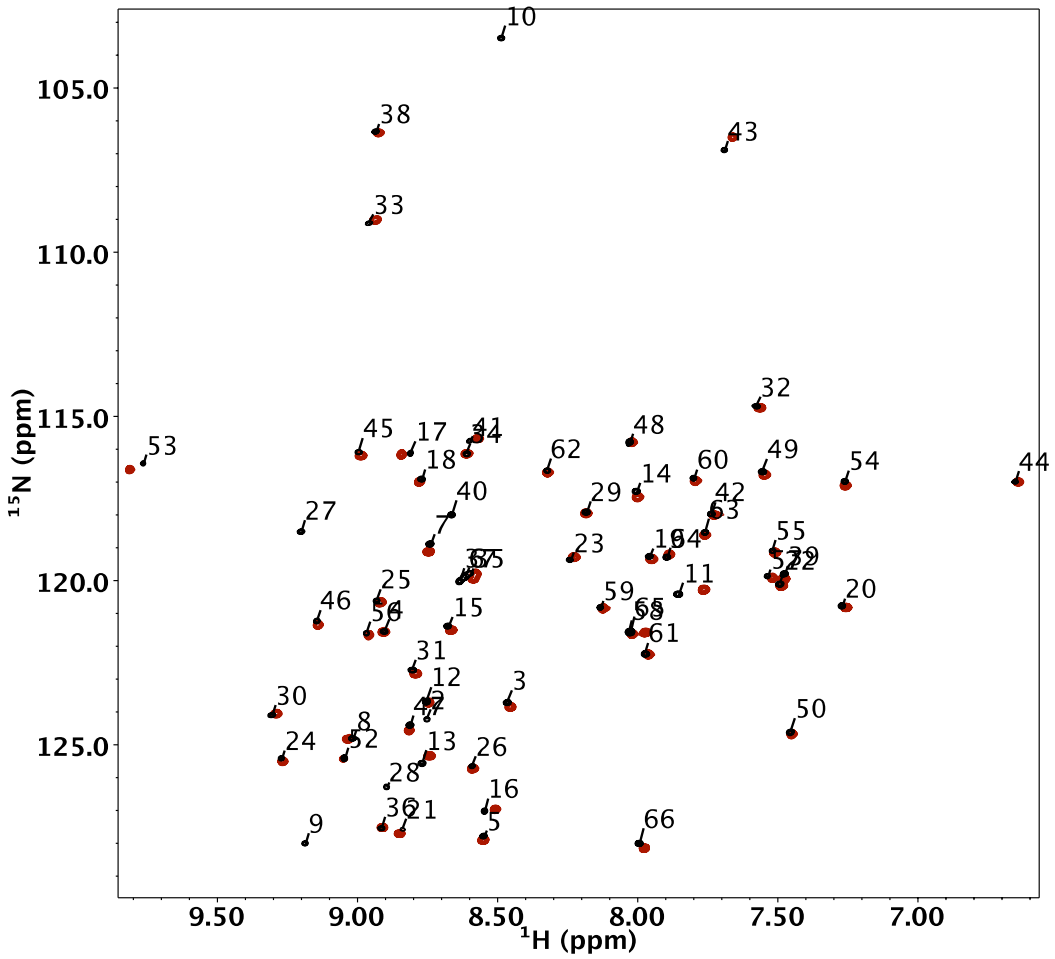


Figure 2.18 Comparison of ^{15}N-HSQC of ^{15}N-Sac7d (black peaks) and ^{15}N-K9C dimer (red peaks). The assignments of Sac7d are shown.

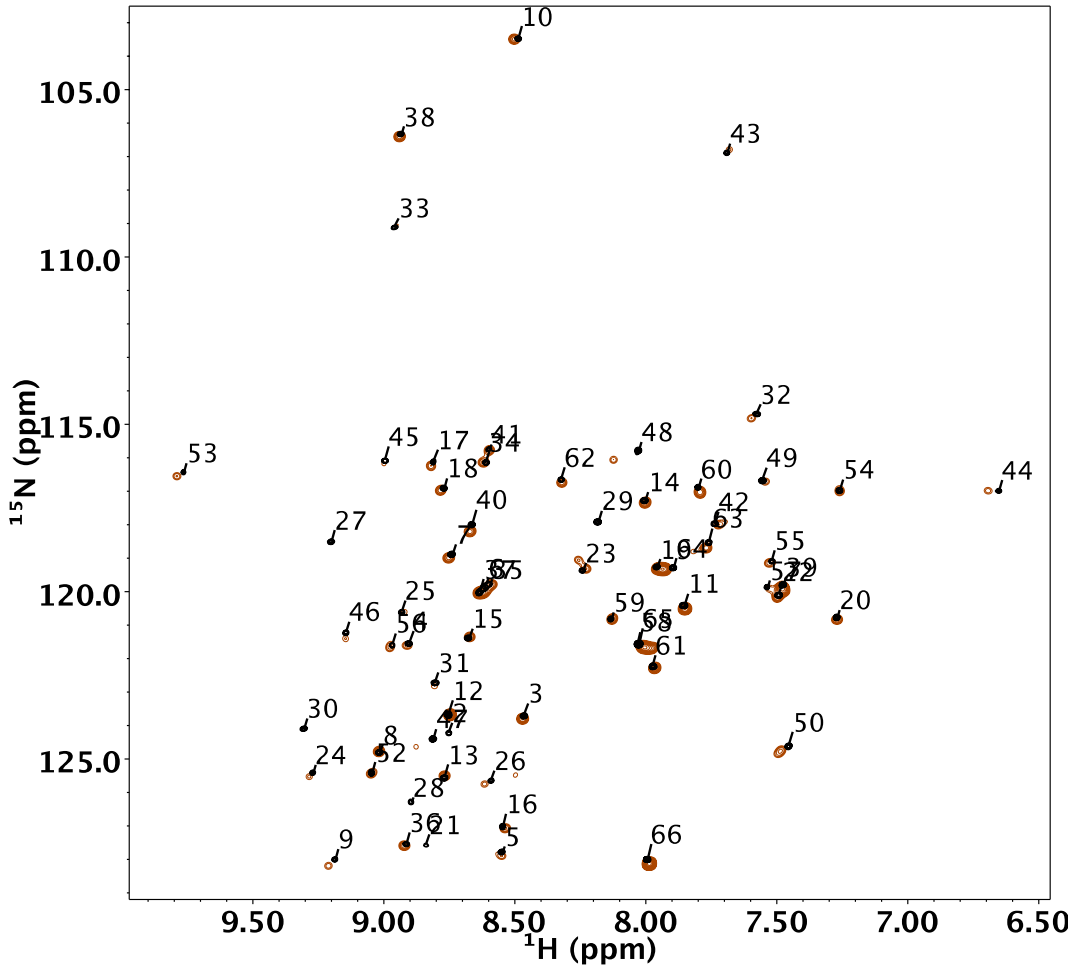


Figure 2.19 Comparison of ^{15}N-HSQC of ^{15}N-Sac7d (black peaks) and ^{15}N-K28C dimer (red peaks). The assignments of Sac7d are shown.

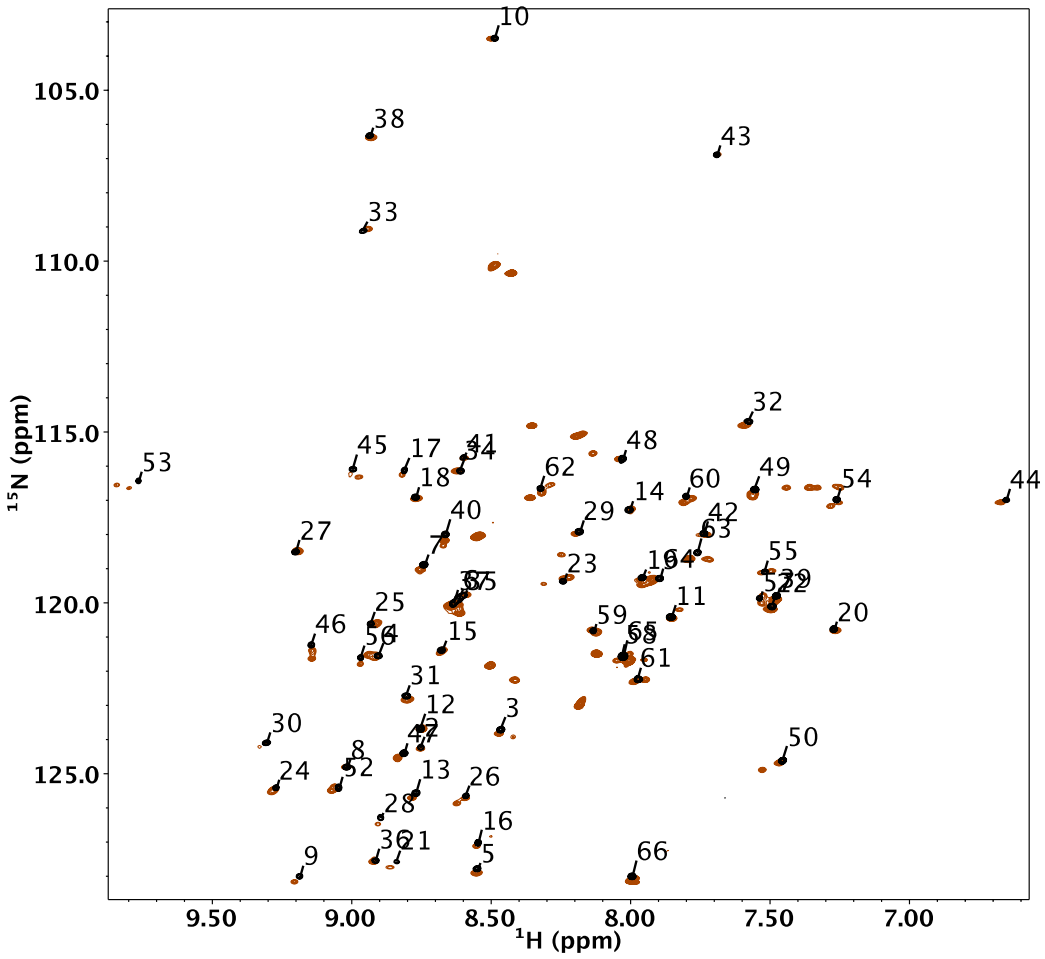


Figure 2.20 Comparison of ^{15}N-HSQC of ^{15}N-Sac7d (black peaks) and ^{15}N-K9CHis-tag-K28C dimer (head-to-tail) (red peaks). The assignments of Sac7d are shown. The extra unassigned (red) peaks represent the residues in His-tag and the spacer between the His-tag and the spacer.

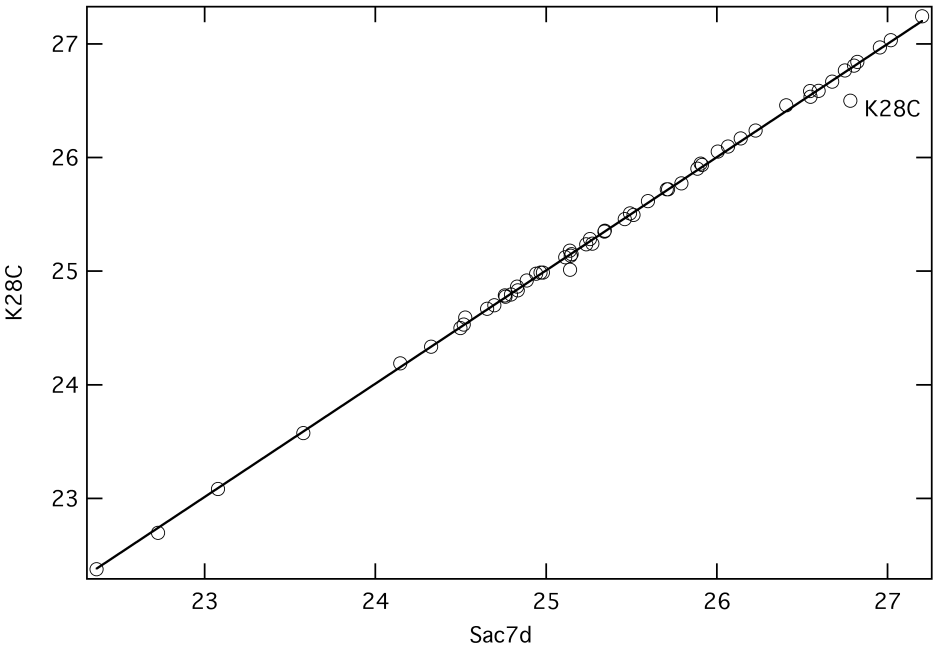


Figure 2.21 Comparison of correlation chemical shifts of K28C-Sac7d and native Sac7d K28C mutation is labeled. Two other extra peaks are observed in the ^{15}N-HSQC of K28C that were not assigned.

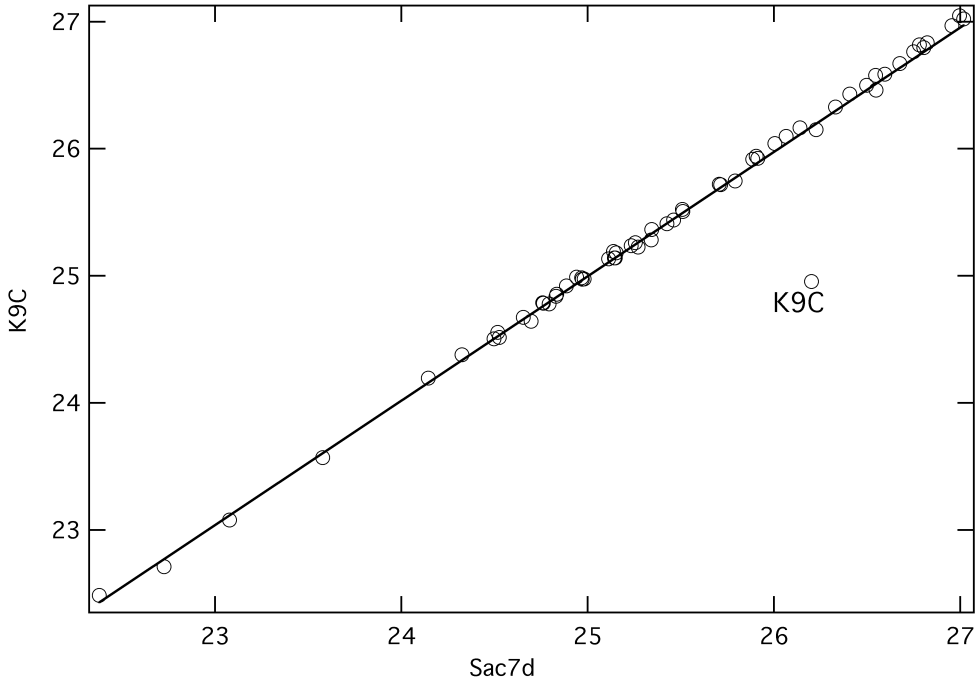


Figure 2.22 Comparison of correlation chemical shifts of K9C-Sac7d and native Sac7d. K9C mutation is labeled.

2.8 Structural studies of Sac7d dimers using residual dipolar coupling analysis

Residual dipolar coupling (RDC) has been increasingly important in bio-molecular structural studies. RDCs are observed when two magnetic nuclei that are proximate to each other are partially aligned. The dipolar interactions, unlike scalar coupling that takes place through the electrons intervening the two nuclei, occurs through space. When a molecule that is otherwise isotropic is partially ordered, it shows residual dipolar coupling for the interacting nuclei. The RDC between the two nuclei is dependent on the distance between the two nuclei, gyromagnetic ratio and the angle between the magnetic field and the bond vector.

Mathematically, it can be represented as,

$$D_{ij} = \frac{\mu_o \gamma_i \gamma_j h}{(2\pi r)^3} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle \quad . \qquad (2.3)$$

Where, r is the distance between the two nuclei under consideration, γi and γj are the gyromagnetic ratios of the nuclei *i and j,* μ_0 is the permittivity of space, θ is the angle between the intermolecular vector and the magnetic field and h is Planck's constant. The residual dipolar coupling (D_{ij}) is given in hertz when the SI units are used for all parameters in the equation. In an isotropic situation, where all the possible orientations are equally sampled, the term in the parentheses is reduced to zero and thus the RDC observed. It is thus necessary to partially align the molecules to observe RDCs. In a partially aligned sample, the splitting observed is the sum of the scalar coupling and the RDC. The scalar coupling is first measured without the partial alignment, so that it can be subtracted from the splitting obtained in the presence of alignment media to obtain the

residual dipolar coupling. This can be done either before the addition of the alignment media or in some media by changing the temperature in the presence of the media. The need for alignment introduces more parameters, in addition to r and θ, to obtain the RDCs. The direction of the ordering frame with respect to the coordinate system fixed by the molecule, the extent of ordering and also the degree of asymmetry need to be represented.

2.8.1 Data collection

RDC data was collected at 18.7 T using partially aligned ^{15}N- labeled protein in a liquid crystalline medium consisting of *n*-alkyl-poly(ethylene glycol) and hexanol (Rückert and Otting 2000; Biyani, Kahsai et al. 2005). Spectra of the unaligned samples were collected and spectra in partially aligned state have been collected on the same sample after the addition of the liquid crystallization-promoting medium. Single bond ^{1}H-^{15}N RDCs were obtained using ^{1}H,^{15}N HSQC-IPAP pulse sequence (obtained from Dr. L. Kay at University of Toronto). The raw data sets have been processed with linear prediction and zero filling. RDCs were extracted using in-house NMRView scripts that are presented in appendix C. RDC values have been analyzed using MODULE (Dosset, Hus et al. 2001). Structural analysis using the RDC information was performed using XPLOR (Schwieters, Kuszewski et al. 2003; Clore and Schwieters 2006).

2.8.2 Results

Figures 2.23 and 2.24 show the comparison of experimental and fitted N-H RDCs for the head-to-head dimer and the tail-to-tail dimer. Figures 2.24 and 2.25 show the structure solved based on the RDCs for K28C dimer and K9C dimer.

The comparison of the experimental and the fitted N-H RDCs indicate that the two monomer subunits are properly folded and maintain an overall secondary structure that is comparable to that of the monomer Sac7d. The structure of the tail cysteine mutant (K9C) and the head cysteine mutant (K28C) line up very well with that of Sac7d monomer. The overall secondary structure and folding are well conserved.

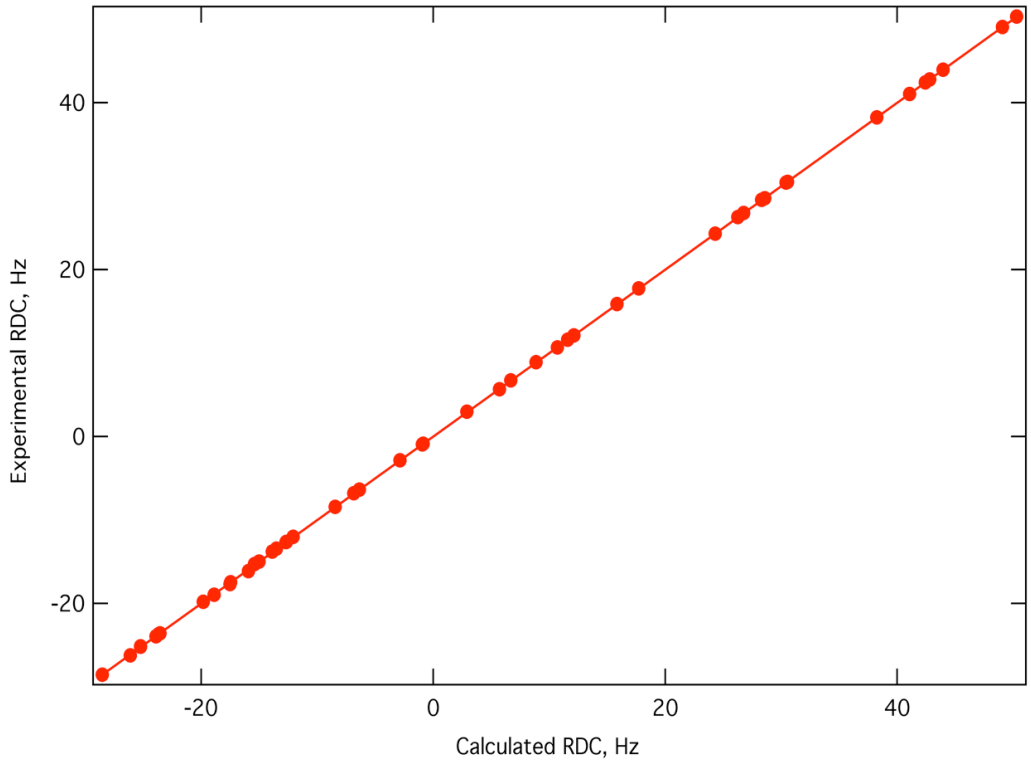


Figure 2.23 Comparison of experimental and calculated N-H RDCs for K9C dimer (tail-to-tail dimer).

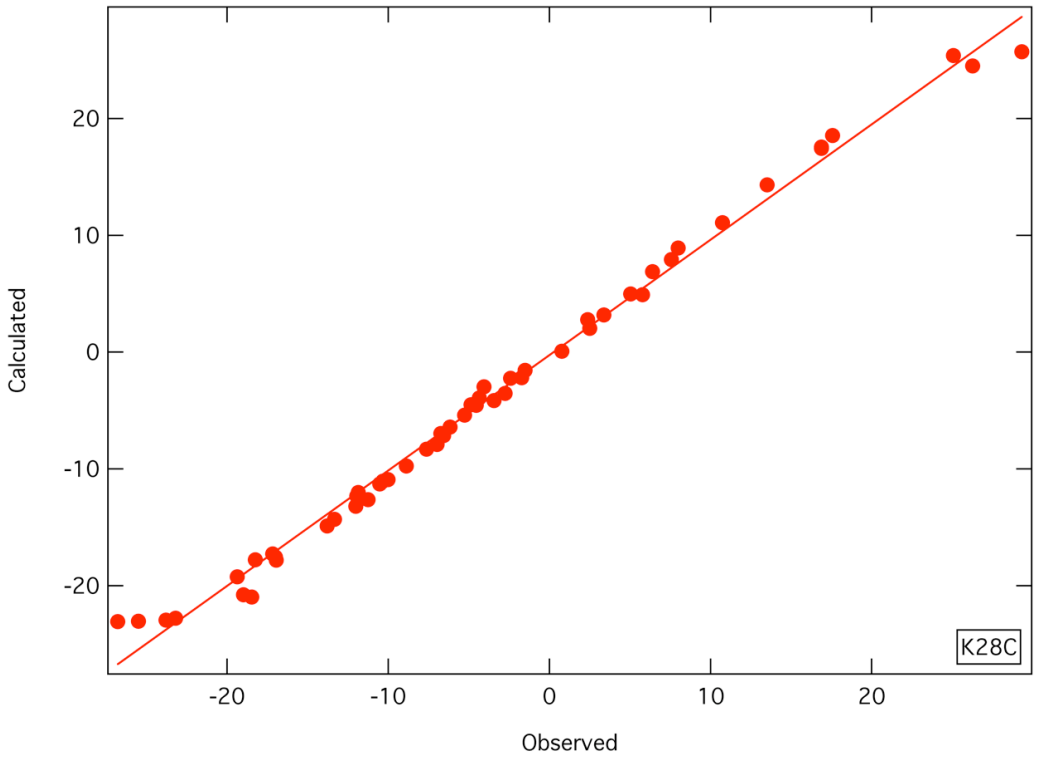


Figure 2.24 Figure 2.21 Comparison of experimental and calculated N-H RDCs for K28C dimer (head-to-head dimer).

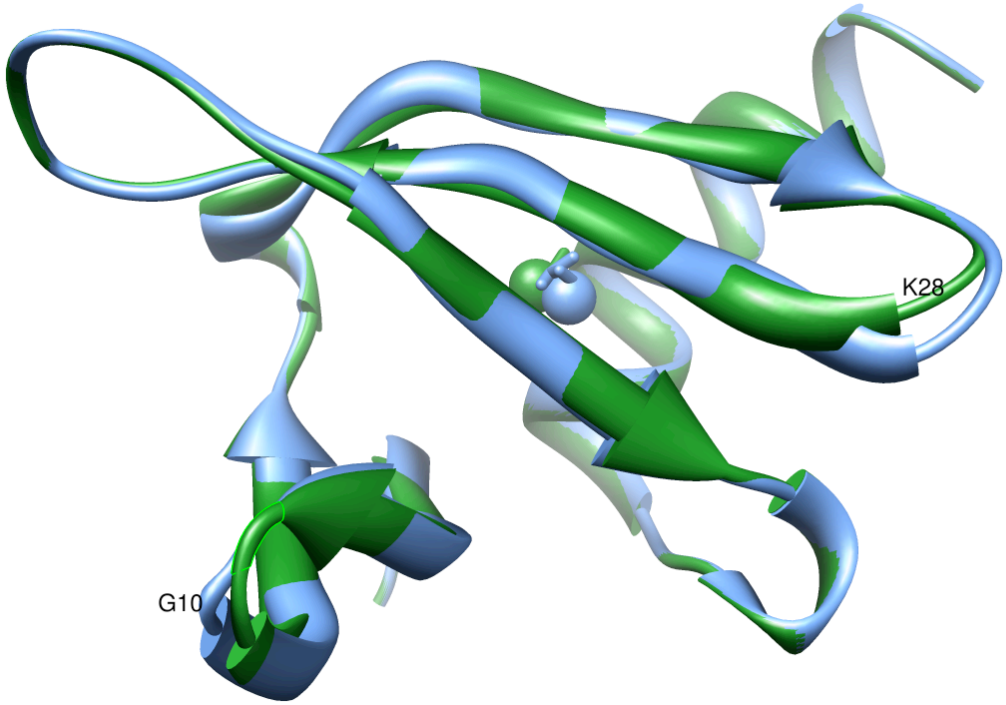


Figure 2.24 Superimposition of K9C (blue) structure obtained from XPLOR on Sac7d (1sap) (Green) solution structure

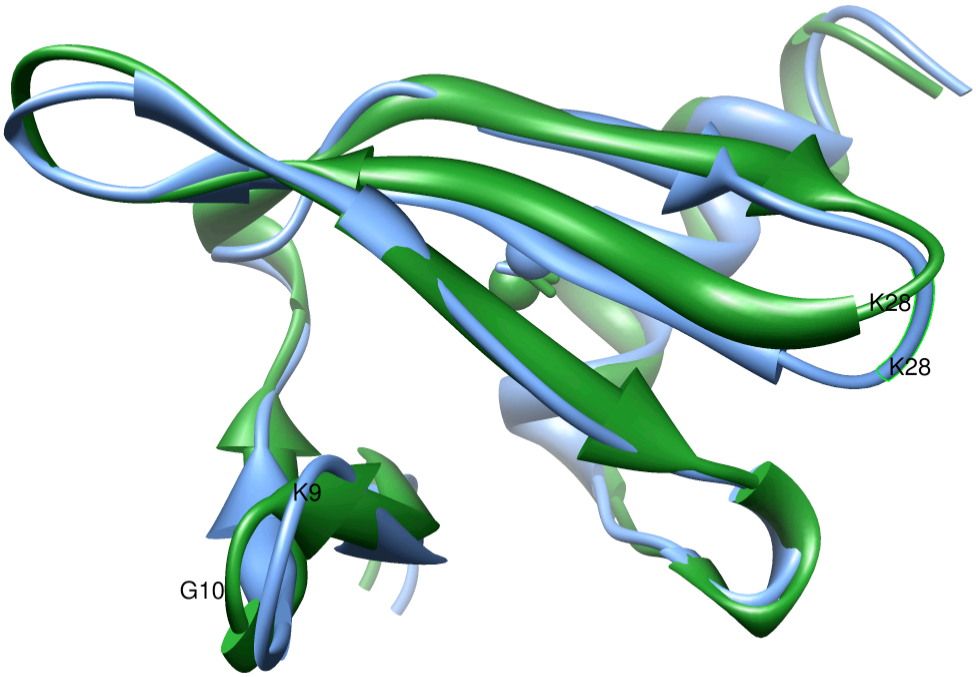


Figure 2.25 Superimposition of K28C (blue) structure obtained from XPLOR on Sac7d (1sap) (Green) solution structure.

2.8.3 RDCs for head-to-tail mutant

It was not possible to calculate the RDCs for the head-to-tail mutant due to the presence of two dissimilar monomer units in the dimer. Due to the presence of his-tag K9C as one of the subunits in the dimer, the ^{15}N-HSQC of the dimers showed more than one peak for each residue. It was difficult to correctly assign the peaks to exact residues in IPAP data. Therefore an RDC analysis of the head-to-tail dimer could not be performed.

2.9 Discussion

In order to investigate the potential preference in the binding orientation for adjacent Sac7d molecules on DNA, Sac7d dimers have been engineered by cross-linking two Sac7d monomers in different orientations. Dimerization was achieved by introducing site-directed single cysteine mutations in selected positions and by cross-linking the sulfhydryl groups by oxidation using diamide or by chemical cross-linking using bi-functional cross-linking agents that are specific for sulfhydryl groups. Engineering head-to-head and tail-to-tail dimers was straightforward. But, the separation of head-to-tail dimers from the mixture of dimers in three possible orientations required an additional differentiating parameter. Subjecting a mixture of monomer cysteine mutants in the head and tail regions results in a mixture of dimers in head-to-head, head-to-tail and tail-to-tail orientations. C-terminal His-tag was cloned to the tail cysteine mutant to separate head-to-tail dimers from the dimer mixture. The difference in the cross-linking orientation translates to difference in the number of His-tags per dimer that in turn translates to

different affinities to Ni-column. Three dimer orientations were separated using Ni-column affinity chromatography.

Primary characterization by ion exchange chromatography and SDS-PAGE confirmed the formation of Sac7d dimers. Biophysical characterization of the dimers prepared both by oxidation and by chemical cross-linking indicated that the dimers were stable and the introduction of cysteine mutations and dimerization did not affect the secondary structure, thermal stability and DNA binding properties of Sac7d. The presence of extra leader sequence in the C-terminal His-tag of the tail mutant provided a convenient method for the clear identification of dimers with different cross-linking orientations as distinct bands on SDS-PAGE gels. Collection of CD spectra for the cysteine mutants and the dimers in the far UV region followed secondary structure analysis using CDPro software package indicated that the cysteine mutation and dimerization did not affect the protein folding and the secondary structure. DSC studies on the mutants indicated that the thermal stability of cysteine mutants was comparable to that of Sac7d while the dimers had independent unfolding for each monomer unit. ^{15}N-HSQC spectra of the mutant dimers indicate that the chemical shifts of the residues around the mutation were affected but the remaining residues showed chemical shits that matched closely with Sac7d. RDC analysis for head-to-head and tail-to-tail dimers was conducted to study the relative orientation of the two monomers. RDC analysis for the head-to-tail dimer was difficult due to the appearance of separate peaks for the same amino acid residue corresponding to the two monomers.

CHAPTER 3

DNA BINDING OF Sac7d DIMERS

3.1 Introduction

DNA binding of Sac7d dimers cross-linked in different orientations is studied to understand the effect of difference in cross-linking orientation of adjacent proteins on DNA binding. A preferred orientation for two Sac7d molecules binding at adjacent binding sites on DNA should result in increased binding affinity for a Sac7d dimer that is cross-linked in that orientation because the binding of one half of the dimer should facilitate binding of the second monomer unit due to decreased entropic penalty. The effect of cross-linking orientation on Sac7d-DNA binding has been studied in three ways: 1. Reverse titrations (DNA into protein) followed by monitoring the fluorescence quenching of W24 in Sac7d upon DNA binding, 2. Forward titrations (protein into DNA) followed by monitoring the change in the circular dichroism of the DNA with protein binding and 3. DNA affinity chromatography. The DNA binding energetics of all the mutant dimers described in chapter two are described here with different DNA sequences and lengths.

3.2 Fluorescence spectroscopy

Aromatic amino acids contribute to the intrinsic fluorescence of the proteins. Contribution from tryptophan and tyrosine is primary while, contribution from phenylalanine is significant only in the absence of the former two amino acids. The fluorescence of tryptophan is usually much stronger than that of tyrosine and phenylalanine. The average values of absorptivity and quantum yields for tryptophan, tyrosine and phenylalanine are 5600/0.2, 1400/0.14 and 257/0.04 (molar absorptivity: L mol^{-1} cm^{-1}) respectively. UV spectroscopic studies indicate that tryptophan and tyrosine have a strong absorption band at ~280 nm, and excitation near this wavelength results in characteristic emission profiles. The maximum fluorescence of tryptophan is centered around 350 nm when excited at 295 nm and the maximum fluorescence of tyrosine is around 303 nm when excited at 275 nm. The difference in the emission wavelengths of the two amino acids aids in the isolation of respective individual intrinsic fluorescence contributions from them. The intensity and the wavelength maximum fluorescence emission of tryptophan depend on the surrounding environment. This feature in complement with the sensitivity of fluorescence can be utilized to study the changes in the local environment of the tryptophan side chains by employing a minimum amount of protein.

Sac7d sequence contains a single tryptophan, two phenylalanines and two tyrosines (appendix D.3). W24 is present in the DNA binding site and forms hydrogen bonds with DNA in the Sac7d-DNA complex (Robinson, Gao et al. 1998). Interaction of W24 with DNA in Sac7d-DNA binding results in the quenching of the intrinsic fluorescence approximately by 90%(McAfee, Edmondson et al. 1996).

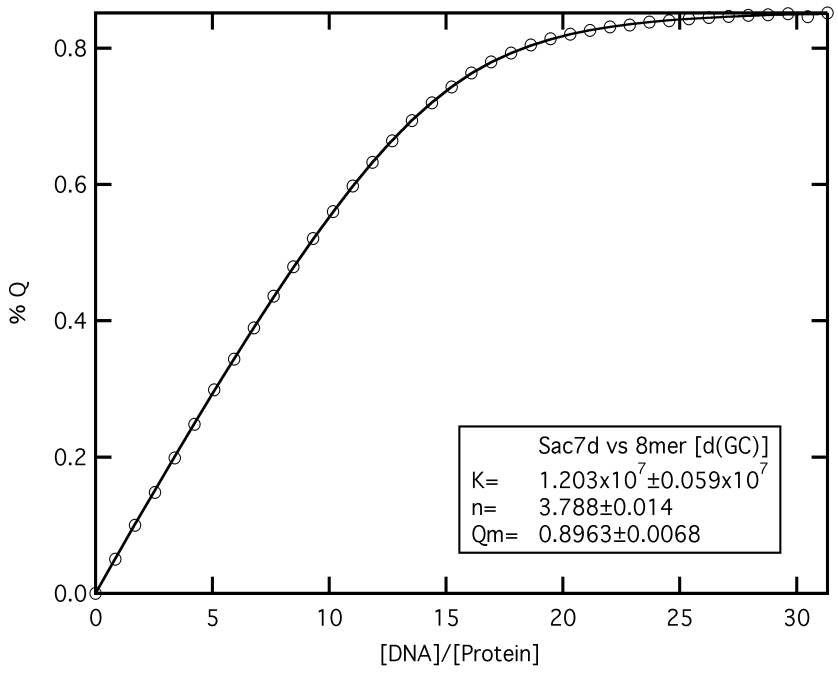


Figure 3.1 Reverse titration of Sac7d with 8mer [d(GC)] followed by fluorescence quenching (open circles). (Conditions were 0.01M phosphate buffer, pH=7.2 at 20 ºC). The bold line is the fit obtained that gives the values of Ka (binding constant), n (site size) and %Q (percentage quenching). The concentration of the protein was 1.5 µM. Five µl injections of stock concentrations of DNA (0.3 mM) were added to the protein solution until the quenching remained constant. X-axis represents the ratio of DNA to protein concentration and Y-axis represents fluorescence quenching. DNA/protein ratio is the number of DNA nucleotides (mM) per protein (mM of monomer).

It has been shown that the comparison of emission spectra for free Sac7d and poly (dGdC) saturated with Sac7d illustrates the quenching of internal fluorescence due to DNA binding (McAfee, Edmondson et al. 1996). It has also been shown previously by the model independent ligand binding density function method of Bujalowski and Lohman (Bujalowski and Lohman 1987) that the fractional change in fluorescence quenching of single tryptophan in Sac7d observed upon DNA binding is equal to the fraction of protein bound (McAfee, Edmondson et al. 1996). Therefore, binding isotherms can be constructed and the fluorescence data is able to be fit using nonlinear

regression to obtain intrinsic binding affinities, site sizes and maximal fluorescence quenching. Figure 3.1 displays titration data for Sac7d with an 8 mer [d(GC)]. Intrinsic fluorescence quenching of tryptophan has been used to measure the binding of Sac7d mutant dimers to DNA with variable sequence and length. The binding parameters obtained from DNA-Sac7d dimer binding were compared to those of DNA-Sac7d monomer binding in order to study the effect of dimerization and binding orientation of adjacent Sac7d molecules on DNA.

3.2.1 Fluorescence data analysis

The dependence of the observed intrinsic tryptophan fluorescence quenching, Q_{obs}, on the total DNA concentration for each titration was fit using non-cooperative, non-specific DNA binding model developed by McGhee and von Hippel (McGhee and von Hippel 1974) , as modified by Lohman and Mascotti by linear regression (Lohman and Mascotti 1992). This model allows the analysis of binding data for non-sequence-specific binding in terms of observed fluorescence quenching (Q_{obs}), observed affinity (K_{obs}), site-size (n), ligand binding density (mole of ligand bound per mole of lattice residue, ν) and free ligand (Lf). Q_{obs} is defined by $(F-F_{obs})/F$, where F and F_{obs} are the fluorescence intensity of protein in the absence and the presence of DNA respectively. Root mean square errors in the optimization were evaluated using an iterative computer algorithm, which numerically searches for the value of the binding density, ν(i), which corresponds to the i^{th} experimental total DNA concentration, $D_{t\,exp}(i)$. An initial value for the concentration of the protein bound to DNA, $L_b(i)$, is obtained from

$$L_b(i) = [Q_{obs}(i)/Q_{max}]L_t(i) \quad . \quad (3.1)$$

Since $L_b/L_t = Q_{obs}/Q_{max}$ (proven by ligand binding density function analysis (McAfee, Edmondson et al. 1996)), where Q_{obs} (i) is initially taken to bee the experimentally observed value. L_b (i) and L_t (i) are the bound and total concentrations of protein for data point I, respectively and Q_{max} is the maximal fluorescence quenching. Using the definition of the binding density, ν(i), an initial value for ν(i) is obtained.

$$\nu(i) = L_b(i)/ D_{t\,exp} \quad . \quad (3.2)$$

The binding density is then substituted into the McGhee and Von Hippel model (McGhee and von Hippel 1974), as modified by Lohman and Mascotti to include both cooperative and non-cooperative binding (Lohman and Mascotti 1992), to obtain the concentration of free protein, $L_f(i)$.

$$L_f(i) = \nu(i)/[\, K_{obs}[1\text{-}n\nu(i)]\{2\omega[1\text{-}n\,\nu(i)]/[(2\omega\text{-}1) \times [1\text{-}n\nu(i)] + \nu(i) + R]\}^{n-1}\{[1\text{-}(n+1)\,\nu(i)+R]/[2[1\text{-}n\nu(i)]]\}^2] \quad , \quad (3.3)$$

where,

$$R= \{[1\text{-}(n+1)\,\nu(i)]^2 + 4\omega\nu(i)[1\text{-}n\,\nu(i)]\}^{1/2} \quad , \quad (3.4)$$

where n is the site size and ω is the cooperativity parameter and K_{obs} is the intrinsic binding constant observed in the specified conditions. A new value for $L_b(i)$ is calculated from the total protein concentration,

$$Lb(i) = Lt(i) – Lf(i) \quad . \quad (3.5)$$

and the corresponding value for the total DNA concentration $D_{t\ calc}$ is calculated using the definition of the binding density

$$D_{t\ calc}(i) = L_b(i)/ \nu(i) \quad . \quad (3.6)$$

$D_{t\ calc}(i)$ is compared to the experimental value $D_{t\ exp}(i)$, and ν(i) is iteratively incremented until the difference between the calculated and the experimental D_t values is acceptably small (typically less than 0.01% error). The value of $Q_{obs}(i)$, which corresponds to the final $D_{t\ calc}$ is calculated by rearrangement of Equation (4.1):

$$Q_{obs}(i) = Q_{max}\ L_b(i)/L_t(i) \quad . \quad (3.7)$$

Thus, $Q_{obs}(i)$ has been calculated for a given value of the $D_{t\ exp}(i)$. Unless mentioned otherwise, the data for the titration reported here was fit assuming non-cooperative binding, i.e., constraining the value of ω to 1.

3.2.2 Binding parameters for Sac7d dimers

DNA binding of the mutant dimers of Sac7d was studied by titrating small aliquots of DNA into the protein sample. The standard reverse titration conditions were 10 mM KH_2PO_4, pH=7.2 at 20 °C and the procedure is described in appendix A.2.5. Reverse titrations for Sac7d dimers in head-to-head, tail-to-tail and head-to-tail orientations have been conducted. K9C and G10C were selected for cross-linking in the tail region to create K9C homodimer and G10C homodimer. Similarly, K28C and G27C were selected for cross-linking in the head region to create K28C homodimer and G27C homodimer. K9C-G10C and K28C-G27C heterodimers were not created. K9C His-tag-K28C dimer and K9C His-tag-G27C dimer are the two head-to-tail dimers studied.

3.2.2.1. Binding parameters for dimers cross-linked in head-to-head orientation

The DNA binding of K28C dimer and G27C dimer were studied by reverse titrations using fluorescence. Tables 3.1and 3.2 list the DNA binding parameters for K28C dimer and G27C dimer, respectively. The DNA binding of disulfide linked K28C dimer was studied with multiple DNA sequences. The DNA binding parameters for K28C dimer with multiple DNA sequences studied were comparable to those of Sac7d monomer. The binding parameters were comparable both in the case of disulfide cross-linking and chemical cross-linking. Exceptional binding parameters were obtained only in the case of K28C disulfide dimer binding to 12-mer dGdC and 1bbx sequence (5'-CTAGCGCGCTAG-3'). The DNA binding affinity of K28C disulfide dimer with 12mer dGdC and 1bbx sequence was ten-fold greater than that for Sac7d monomer. Maximum quenching in case of binding with 1bbx sequence was ~80%, indicating native like

protein-DNA interaction. With 12-mer dGdC the binding constant was ten-fold greater when compared to that of Sac7d monomer. There was no increase in binding affinity either with 12-mer dGdC or with 1bbx sequence by the introduction of a flexible cross-linker between the two monomers of the K28C dimer.

	$Kx10^{-7}$ (M^{-1})	Site size (n)	Qmax
K28C_Xlinkers with 100-mer dGdC			
K28C_DMB	7.990±0.022	1.8186±0.011	0.7489±0.0012
K28C_BMH	2.240±0.22	2.044±0.017	0.7351±0.0014
K28C_BM$(PEG)_2$	2.050±0.14	1.890±0.012	0.68863±0.0010
K28C_BMDB	3.410±0.11	1.9123±0.0045	0.66560±0.00033
K28C_diamide	1.280±0.26	3.28±0.0062	0.8262±0.00465
Sac7d	1.07±0.1	3.88±0.01	0.885±0.001
K28C_Xlinkers with 12 mer NMR structure sequence			
K28C_DMB	0.185±0.017	2.000±0.044	0.9128±0.0067
K28C_BMH	0.272±0.014	2.336±0.023	0.9080±0.0030
K28C_BM$(PEG)_2$	0.380±0.02	1.880±0.018	0.8307±0.0023
K28C_BMDB	0.435±0.033	2.016±0.028	0.8149±0.0029
K28C_pPDM	0.812±0.062	1.307±0.018	0.7283±0.0017
K28C_diamide	19.88±2.70	2.5333±0.0095	0.7715±0.0028
Sac7d	1.38±0.0850	3.594±0.019	0.6215±0.0010
K28C diamide with variable seq. length			
5mer dGdC	2.07±0.32	2.316±0.031	0.7459±0.0027
6mer dGdC	5.3±1.2	2.911±0.043	0.5039±0.0016
8mer dGdC	4.12±0.33	2.740±0.014	0.61496±0.00065
10mer dGdC	0.8±0.11	1.890±0.037	0.6172±0.0029
12mer dGdC	19±42	1.556±0.022	0.5807±0.0041
1bbx (12mer)	29.2±4.5	2.702±0.010	0.8025±0.0033
100-mer dGdC	1.28±0.26	3.28±0.0062	0.8262±0.00465
100-mer dAdT	8.1±1.4	3.984±0.024	0.6846±0.0019

Table 3.1 Binding parameters for K28C dimer

DNA binding parameters for G27C dimer (table 3.2) with disulfide bonds and with chemical cross-linkers with multiple DNA sequences did not show any differences when compared to those of Sac7d monomer. Unlike the K28C disulfide dimer, the G27C dimer (linked by a disulfide bond) did not show increased affinity either for 12mer dGdC or the 1bbx sequence.

	$Kx10^{-7}$ (M^{-1})	Site size (n)	Qmax
G27C_Xlinkers with 100-mer dGdC			
G27C_DMB	0.813±0.022	2.0643±0.021	0.71795±0.0021
G27C_BMH	3.22±0.19	2.044±0.017	0.7251±0.0012
G27C_BM$(PEG)_2$	2.14±0.12	2.190±0.012	0.67963±0.00087
G27C_BMDB	4.014±0.11	2.0123±0.0045	0.68520±0.00029
G27C_diamide	$1.46x10^{7}$±0.41	3.46±0.0056	0.7862±0.00346
Sac7d	1.07±0.1	3.88±0.01	0.885±0.001
G27C_Xlinkers with 12mer NMR structure sequence			
G27C_DMB	2.04±0.012	2.042±0.029	0.8428±0.0072
G27C_BMH	1.89±0.0141	2.336±0.023	0.892±0.0029
G27C_BM(PEG)2	0.402±0.02	1.880±0.018	0.8307±0.0034
G27C_BMDB	0.482±0.029	2.016±0.028	0.902±0.0019
G27C_pPDM	0.796±0.081	1.307±0.018	0.7846±0.0026
G27C_diamide	1.01±0.24	3.581±0.076	0.5828±0.0052
Sac7d	1.38±0.085	3.594±0.019	0.6215±0.0010
G27C diamide with variable seq. length			
5mer dGdC	0.3269±0.045	3.371±0.048	0.5687±0.0074
6mer dGdC	0.0476±0.0044	0.989±0.074	0.873±0.012
8mer dGdC	0.0624±0.0034	1.123±0.04	0.783±0.0055
10mer dGdC	0.139±0.026	1.78±0.043	0.812±0.044
12mer dGdC	1.28±0.4	3.46±0.032	0.8148±0.006
12mer (1bbx)	1.66±0.4	3.547±0.062	0.8297±0.0058
100-mer dGdC	$1.01x10^{7}$±0.24	3.58±0.076	0.583±0.0052
100-mer dAdT	0.635±0.057	2.835±0.024	0.752±0.0036

Table 3.2 Binding parameters for G27C dimer (head-to-head orientation)

3.2.2.2. Binding parameters for tail-to-tail orientation

The DNA binding of K9C dimer and G10C dimer was studied using fluorescence. Table 3.3 lists the binding parameters for K9C dimer. The DNA binding of K9C disulfide dimer and K9C dimers cross-linked with different flexible cross-linkers has been studied using multiple DNA sequences.

K9C_diamide	Kx10^{-9} (M^{-1})	Site size (n)	Qmax
100dGdC	80±82	3.175±0.027	0.7255±0.0023
12 dGdC	0.0321±0.0083	2.355±0.029	0.681±0.012
12mer (1bbx)	0.0194±0.0032	2.46±0.046	0.678±0.02
8 dGdC	3.7±1.1	3.156±0.012	0.7639±0.0029
8mer (1azp)	0.92±0.26	3.252±0.017	0.7992±0.0052
6 dGdC	0.133±0.016	2.0719±0.0026	0.7289±.0026
5dGdC	0.065±0.033	2.337±0.048	0.549±0.014
with 100-mer [d(GC)]			
K9C_DMB	0.0221±0.0017	3.1303±0.0092	0.7617±0.0092
K9C_BMH	0.0491±0.0036	3.463±0.012	0.8709±0.001
K9C_BM$(PEG)_2$	0.0268±0.0011	3.7557±0.0082	0.89026±0.00089
K9C_BMDB	0.0407±0.0003	3.743±0.011	0.9064±0.001
K9C_pPDM	0.0338±0.0017	1.7232±0.0062	0.80285±0.00059
K9C_diamide	80±82	3.175±0.027	0.7255±0.0023

Table 3.3 DNA binding parameter of K9C dimer (Tail-to-Tail)

The binding constant for K9C disulfide dimer with 100-mer [d(GC)] was 1000 times greater than that for Sac7d monomer. Binding constants for K9C disulfide dimer with 8 mer [d(GC)], 1azp sequence, and with 10-mer [d(GC)] was 100-fold greater and ten-fold greater with 6mer [d(GC)] than that for Sac7d monomer. With 12-mer [d(GC)] and 5-mer [d(GC)]s, the binding affinity dropped and the values were comparable to those of Sac7d monomer. Considering a binding site size of ~4 base pairs per Sac7d monomer, [d(GC)] 12-mer is four base pairs longer than the required eight base pairs for the dimer. Binding of K9C dimer leaves ~1/3rd of oligonucleotide vacant and it is not

usable for another dimer. There is possibility for a monomer subunit (of a dimer) to bind to the remaining length with its other monomer dangling off. In case of 5-mer [d(GC)] the DNA is not long enough to accommodate the dimer. Introduction of flexible cross-linkers with variable length between the two monomer-units did not result in increased affinity as compared to that of Sac7d monomer.

G10C_diamide	$Kx10^{-7}$ (M^{-1})	Site size (n)	Qmax
100dGdC	7.9±1.4	2.652±0.016	0.8469±0.0069
12 dGdC	1.23 ±2.4	2.89±0.023	0.864±0.046
12mer (1bbx)	1.04±0.085	2.6357±0.0074	0.882±0.0042
10 dGdC	2.12±0.62	2.604±0.012	0.807±0.0026
8 dGdC	1.02±0.14	2.596±0.015	0.853±0.051
8mer (1azp)	1.02± 3.1	1.268±0.010	0.761±0.0014
6 dGdC	3.2±1.4	2.98±0.012	0.8132±0.0062
5dGdC	1.862.1	1.96±0.028	0.6014±0.054
with 100-mer [d(GC)]			
G10C_DMB	1.609±0.085	2.6657±0.0071	0.9112±0.0048
G10C_BMH	2.34±0.22	2.604±0.012	0.8907±0.0066
G10C_BM$(PEG)_2$	1.21±0.13	2.639±0.014	0.911±0.011
G10C_BMDB	1.02±0.1	2.596±0.015	0.879±0.011
G10C_pPDM	0.704± 0.031	1.268±0.010	0.7241±0.0013
G10C_diamide	$7.9\pm1.4x10^{7}$	2.652±0.016	0.8469±0.0069

Table 3.4 DNA binding parameters of G10 dimer (tail-to-tail)

Table 3.4 lists the binding parameters for G10C dimer. The binding affinity of G10C disulfide dimer with 100-mer [d(GC)] was greater than that of Sac7d monomer but did not approach that of K9C disulfide dimer. However, no increase in binding affinity was observed with any other sequences or with the use of flexible cross-linkers between the two G10C monomers.

3.2.2.3. Binding parameters for head-to-tail orientation

Head-to-tail dimers were created by cross-linking K9C in the tail region either with K28C or with G27C in the head region. As described in Chapter 2, K9C mutant with a C-terminal His-tag (K9C His-tag) was used to create head-to-tail dimers in order to separate the dimers with different cross-linking orientation. The conservation of increased binding affinity observed with His-tag K9C dimer indicated that the C-terminal His-tag did not interfere with the DNA binding properties of Sac7d. Binding parameters for the head-to-tail dimers are presented in table 3.5.

K9CHis-tag-K28C			
	$Kx10^{-7}$ (M^{-1})	Site size (n)	Qmax
5 dGdC	0.86±0.61	2.486±0.043	0.4716±0.0046
6 dGdC	2.46±2.6	2.43±0.014	0.3657±0.003
8 dGdC	2.27±47	3.52±0.024	0.45939±0.00088
8mer (1azp)	1.35±13.0	3.23±0.032	0.56432±0.00074
10 dGdC	0.601±1.6	3.296±0.05	0.875±0.066
12 dGdC	0.238±0.41	2.363±0.032	0.875±0.039
12mer (1bbx)	0.181±3.9	2.636±0.032	0.639±0.011
100dGdC	2.30±1.6	2.262±0.013	0.59960±0.00075
100dAdT	2.93±4.4	3.438±0.029	0.8023±0.0027
K9CHis-tag-G27C			
5dGdC	0.24±0.61	3.02±0.039	0.5016±0.0062
6 dGdC	1.75±14	3.03±0.032	0.58252±0.00072
8 dGdC	0.782±0.29	2.635±0.010	0.48417±0.00084
8mer (1azp)	0.684±1.6	3.092±0.08	0.846±0.082
10 dGdC	0.221±0.41	3.063±0.032	0.825±0.068
12 dGdC	0.186±0.41	2.84±0.038	0.856±0.048
12mer (1bbx)	2.34±1.8	3.2503±0.0092	0.7163±0.0072
100dGdC	1.71±1.2	2.6981±0.0098	0.7900±0.0055
100dAdT	2.52±2.4	3.132±0.026	0.7963±0.0026

Table 3.5 Binding parameters for head-to-tail dimers: K9CHis-tag-K28C and K9CHis-tag-G27C

DNA binding properties of both the head-to-tail dimers, K9C-K28C and K9C-G27C were studied with multiple DNA sequences. Similar to the dimers in head-to-head and tail-to-tail orientations, the binding affinity of head-to-tail dimers with 5-mer [d(GC)] was lower than that with the other sequence lengths due to space constraints. In contrast to head-to-head and tail-to-tail dimers, the binding affinities of head-to-tail dimers for 8-mer, 10-mer and 12-mer were 10 fold lower than those in case of Sac7d monomer. No increased binding affinity is observed for both K9C-K28C and K9C-G27C with change in the DNA sequence, except for a 10-fold greater binding affinity for the binding of K9C-G27C dimer with 8-mer [d(GC)], when compared to that of Sac7d monomer. The maximum quenching values for binding with 5-mer, 6-mer and 8-mer [d(GC)] sequences were lower than those with monomer Sac7d.

Figure 3.2 displays a comparison of the binding data obtained from reverse titrations of dimers in different cross-linking orientations with 100-mer [d(GC)].

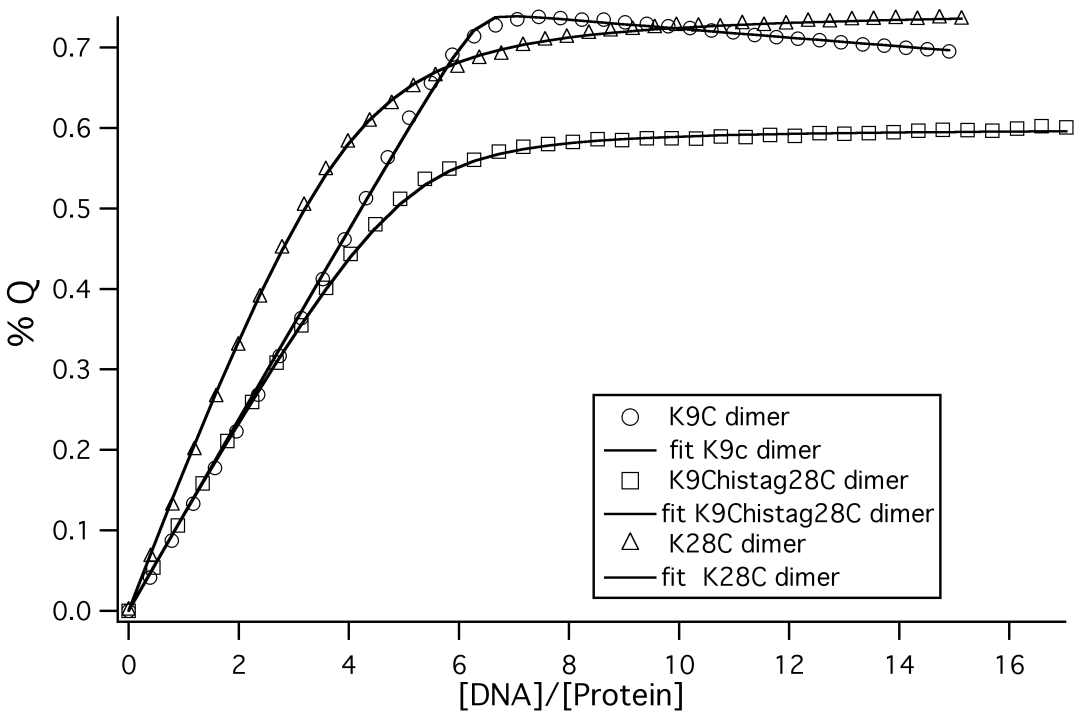


Figure 3.2 Comparison of fluorescence quenching data for the dimers cross-linked in different orientations in a reverse titration with 100-mer [d(GC)]. DNA/protein ratio is the number of DNA nucleotides (mM) per protein (mM of monomer).

3.3. Sac7d-DNA binding monitored by the change in the circular dichroism of DNA

Binding of Sac7d to DNA followed by monitoring the change in the CD of DNA has shown that there is a cooperative structural change in the DNA with increase in protein binding density (McAfee, Edmondson et al. 1996). A significant increase in the CD spectrum poly [d(GC)] centered around 285 nm was observed with the addition of increasing amounts of protein.

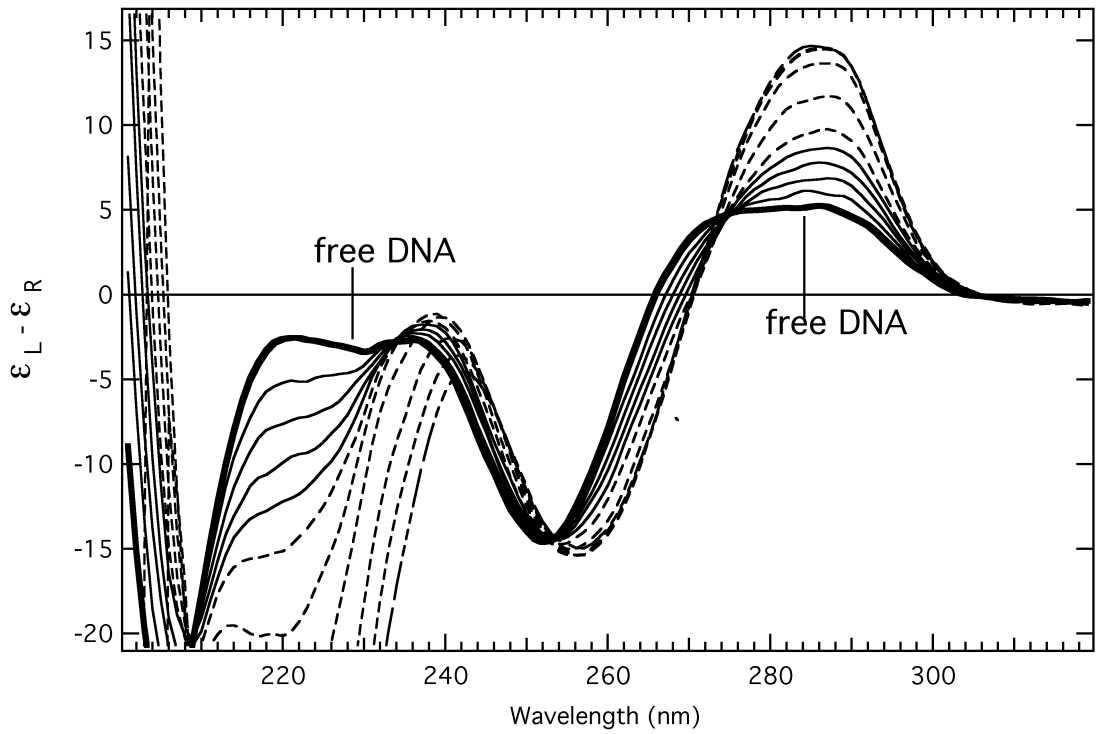


Figure 3.3 Forward titration of Sac7d with dGdC 100-mer followed by circular dichroism. The initial CD spectrum for the DNA in the absence of protein is shown bold and labeled ([protein]:[DNA]=0). Spectra for progressively increasing g ratios of protein:DNA (approximately 0.01, 0.02, 0.03, 0.04, 0.05, 0.07, 0.11, 0.16, 0.21 and 0.26) in 0.01 mM KH_2PO_4 (pH=7.2) are shown. The initial concentration of 100-mer [d(GC)] was ~65μM (McAfee, Edmondson et al. 1996).

The CD spectra are nearly isodichroic at 250 nm and 270 nm suggesting the presence of two predominant species. With addition of increasing amount of protein, the CD band around 220 nm wavelength becomes increasingly negative. At wavelengths above 250 nm, CD band depicts contributions from the DNA. It has been shown that the CD changes in the long-wavelength region represent Sac7d binding induced conformational changes in DNA (McAfee, Edmondson et al. 1996). Figure 3.3 displays the titration of poly [d(GC)] with Sac7d. The study was conducted using multiple sequences (poly [d(GC)], poly [d(IC), poly [d(AG)]-poly [d(CT)] and poly [d(AC)]-poly

[d(GT)]) and it showed that the DNA sequence influenced the cooperativity associated with the protein binding induced structural change in DNA. Figure 3.4 displays the plot of fractional change in Δε values with increasing amount of protein.

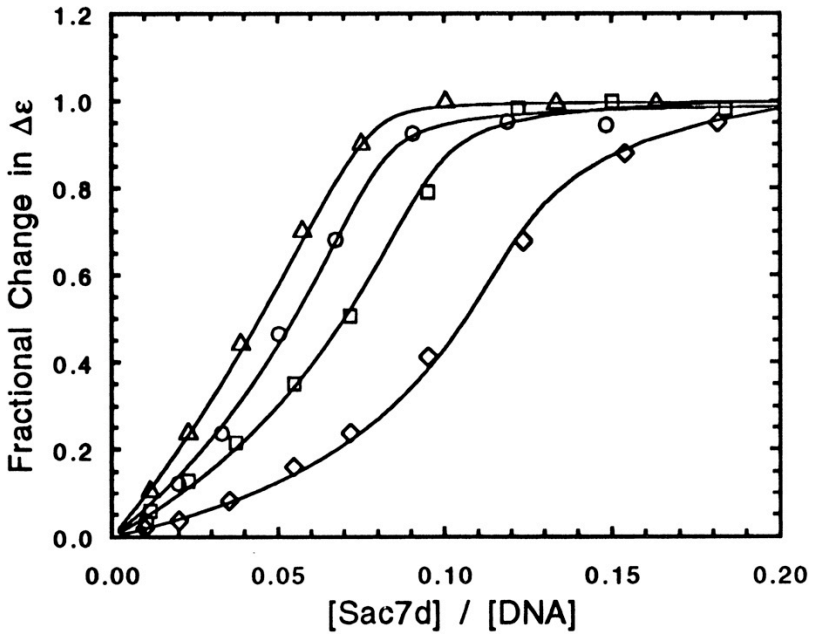


Figure 3.4 Plot of fractional change in Δε values against the ratio of protein to DNA. The markers represent the data (triangles = polyd(AG).poly d(CT), circles = poly d(AC)-poly d(GT), squares = poly d(GC) and diamonds = poly d(IC)) and the smooth are simulated using McGhee Von Hippel model for polynucleotide binding. The figure has been extracted from (McAfee, Edmondson et al. 1996).

A sequence-dependent sigmoidicity in unwinding was observed in fractional change of Δε values with increase in [Sac7d] to [DNA] ratio. It was suggested that the sigmoidicity in unwinding was dependent on the inherent nature of the sequence to undergo structural change. Poly [d(IC)] does not show a structural transition until near saturation, unlike poly [d(GC)] that shows a structural change at a much lower protein concentration ([protein]:[DNA] ~ 0.05) (McAfee, Edmondson et al. 1996). A correlation

between inherent DNA flexibility and sigmoidicity of DNA unwinding was noted (McAfee, Edmondson et al. 1996).

In order to study the influence of cross-linking orientation of adjacent proteins on the structural changes in DNA, a forward titrations were conducted with poly [d(GC)] using Sac7d dimers cross-linked in different orientations (head-to-head, tail-to-tail and head-to-tail). Separate forward titrations were conducted by titrating small aliquots of dimers cross-linked in different orientations into 100-mer [d(GC)] followed by monitoring the change in the CD of the DNA. The structural change in DNA imposed by protein binding was studied with increasing protein concentration. Similar to the structural change imposed by Sac7d binding, the binding of Sac7d dimers also induced increase in the CD band around 285 nm with increasing protein concentration. But the change in the CD spectrum of poly [d(GC)] varied with cross-linking orientation of the dimer. Figures 3.5 through 3.8 display the titrations of poly [d(GC)] with K9C disulfide dimer (tail-to-tail), K28C disulfide dimer (head-to-head), K9C-K28C disulfide dimer (head-to-tail) and K9C-G27C disulfide dimer (head-to-tail).

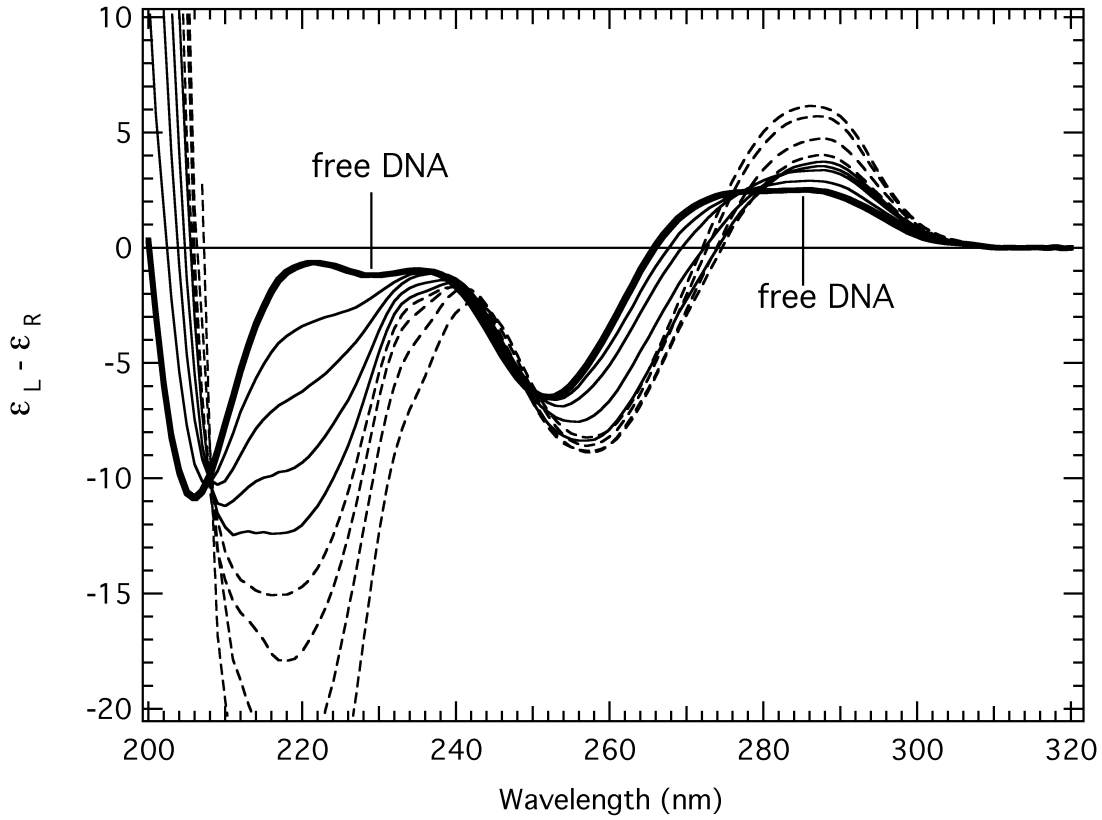


Figure 3.5 Forward titration of dGdC 100-mer with K9C dimer followed by circular dichroism. The initial CD spectrum for the DNA in the absence of protein is shown bold and labeled ([protein]:[DNA]=0). Spectra for progressively increasing g ratios of protein:DNA (approximately 0.02, 0.06, 0.08, 0.11, 0.14, 0.17, 0.2 and 0.25) in 0.01 mM KH_2PO_4 (pH=7.2) are shown. The initial concentration of 100-mer [d(GC)] was ~65 μM.

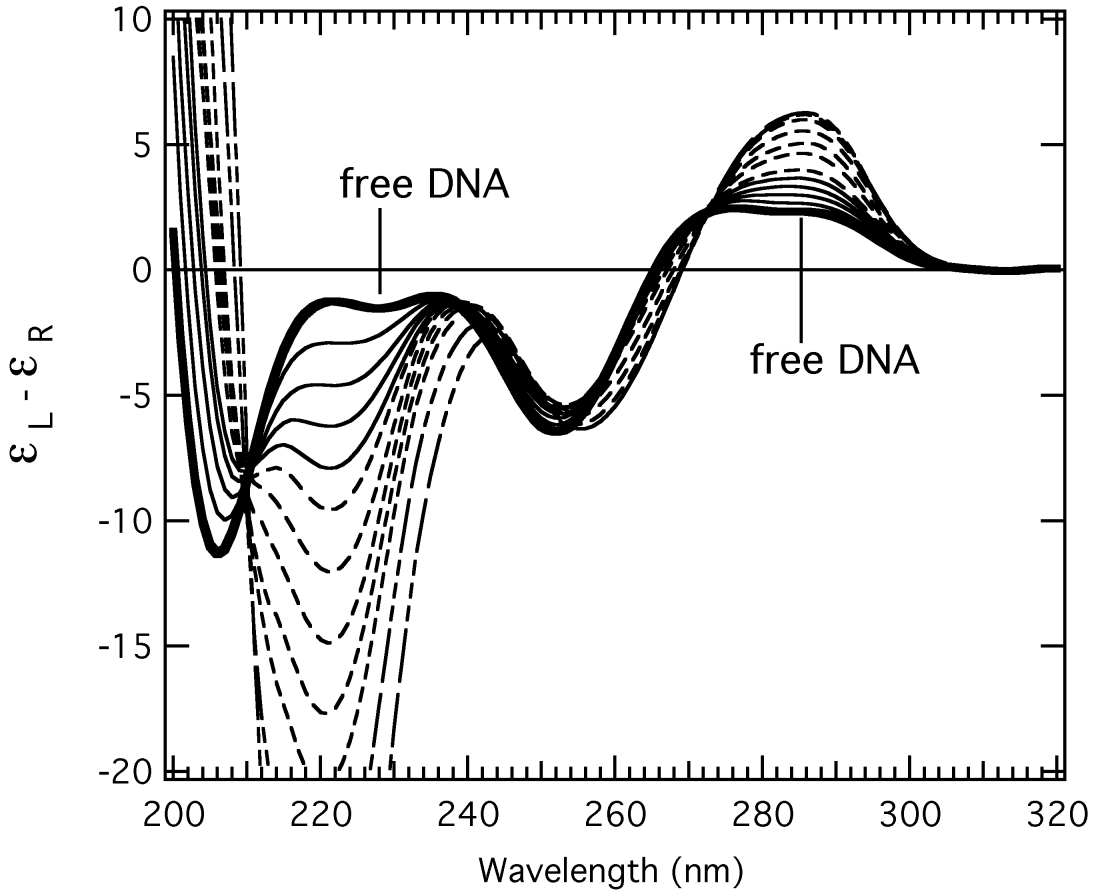


Figure 3.6 Forward titration of dGdC 100-mer with K28C dimer followed by circular dichroism. The initial CD spectrum for the DNA in the absence of protein is shown bold and labeled ([protein]:[DNA]=0). Spectra for progressively increasing g ratios of protein:DNA (approximately 0.02, 0.03, 0.05, 0.06, 0.08, 0.11, 0.13, 0.16, 0.19, 0.24 and 0.32) in 0.01 mM KH_2PO_4 (pH=7.2) are shown. The initial concentration of 100-mer [d(GC)] was ~65 μM.

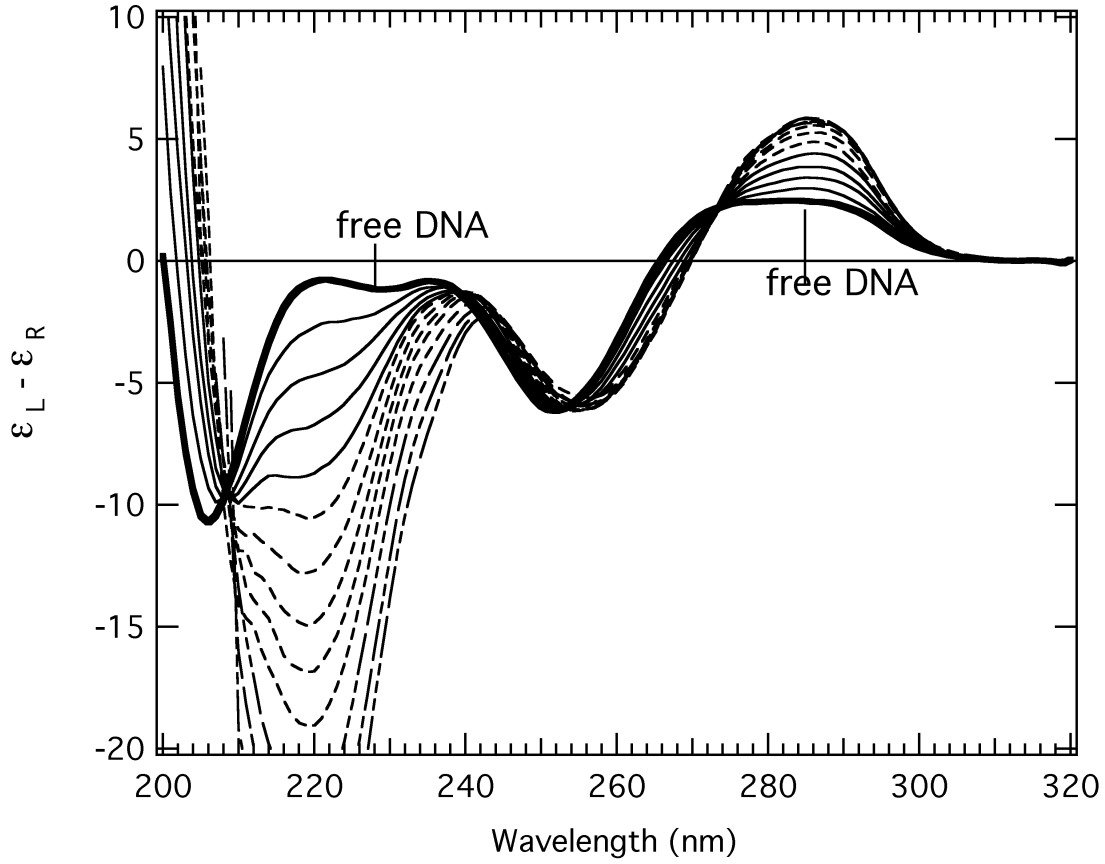


Figure 3.7 Forward titration of dGdC 100-mer with K9CHis-tag-K28C dimer followed by circular dichroism. The initial CD spectrum for the DNA in the absence of protein is shown bold and labeled ([protein]:[DNA]=0). Spectra for progressively increasing g ratios of protein:DNA (approximately 0.02, 0.03, 0.05, 0.07, 0.09, 0.1, 0.12, 0.14, 0.16, 0.18, 0.21 and 0.24) in 0.01 mM KH_2PO_4 (pH=7.2) are shown. The initial concentration of 100-mer [d(GC)] was ~65 μM.

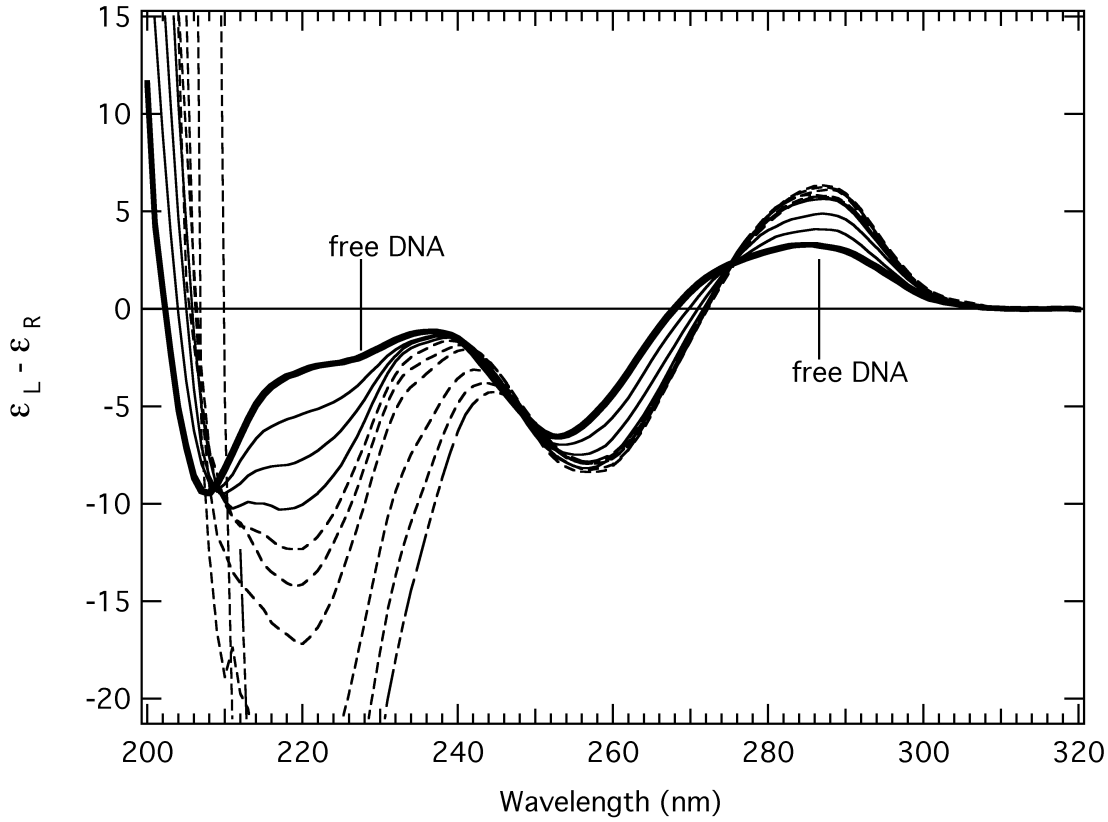


Figure 3.8 Forward titration of dGdC 100-mer with K9CHis-tag-G27C dimer followed by circular dichroism. The initial CD spectrum for the DNA in the absence of protein is shown bold and labeled ([protein]:[DNA]=0). Spectra for progressively increasing g ratios of protein:DNA (approximately 0.02, 0.05, 0.07, 0.09, 0.09, 0.12, 0.14, 0.16, 0.23, 0.30 and 0.38) in 0.01 mM KH_2PO_4 (pH=7.2) are shown. The initial concentration of 100-mer [d(GC)] was ~65μM.

The change in the CD of DNA for the forward titration of K9C disulfide dimer into 100-mer [d(GC)] was different compared that of Sac7d 100-mer [d(GC)]. The isodichroic point at 250nm was consistent with the Sac7d-100-mer [d(GC)] titration, but there was no isodichroic point at longer wavelength region. The CD spectra for the first two aliquots of protein ([protein]:[DNA] ~0.02 and ~0.06) were isodichroic with that

of100-mer [d(GC)] at 275 nm. The CD spectra for the subsequent aliquots of protein do not pass through the isodichroic point at 275 nm. In case of forward titration of 100-mer [d(GC)] with K28C disulfide dimer the isodichroic point at the long wavelength ~275 nm is consistent with that in case of Sac7d into 100-mer [d(GC)]. For the head-to-tail dimers the isodichroic point in the long wavelength region is consistent with that in case of Sac7d into 100-mer [d(GC)] but differ with the Sac7d titration in the short wavelength region.

The fractional change $\Delta\varepsilon$ at 285nm from each titration have been plotted as a function of the ratio of protein to DNA concentration (Figure 3.9). The fractional change in the $\Delta\varepsilon_{285}$ for K9CHis-tag-K28C (in head-to-tail orientation) is comparable to that of Sac7d. The structural change introduced by K28C dimer (in head-to-head orientation) is comparable to that of Sac7d monomer but degree of unwinding is less compared to that of Sac7d. The fractional change in $\Delta\varepsilon_{285}$ for the binding of K9C dimer (in tail-to-tail orientation) indicates cooperative unwinding in DNA. The curve is comparable to that obtained with a titration of Sac7d monomer into poly [d(IC)] (figure 3.4) (McAfee, Edmondson et al. 1996).

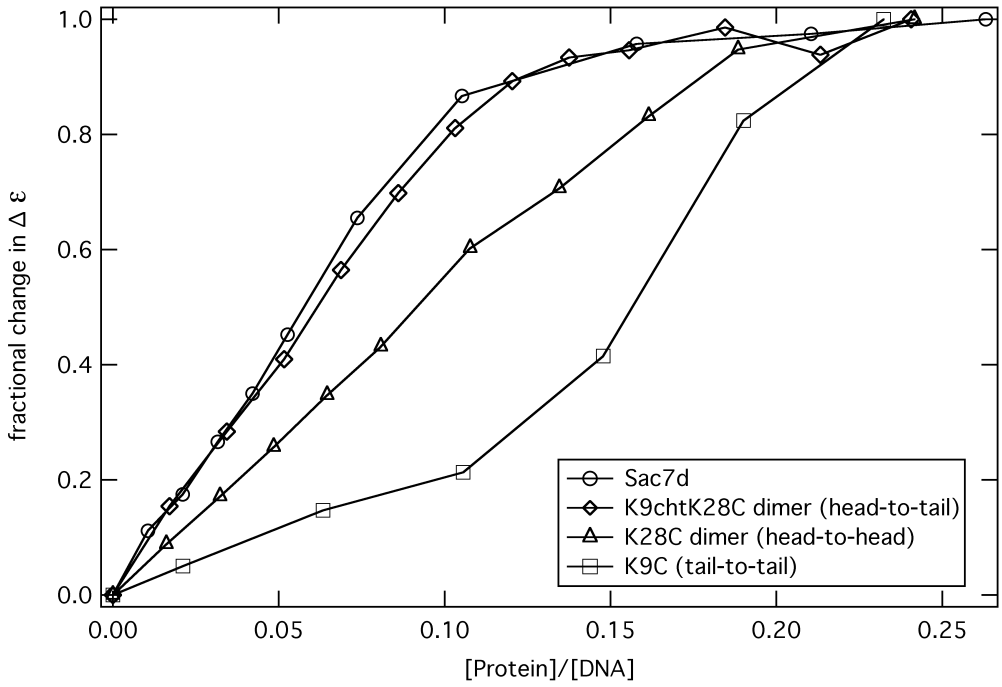


Figure 3.9 Fractional change in Δε285 for 100-mer [d(GC)] with increasing concentration of Sac7d and Sac7d dimers cross-linked in head-to-head, tail-to-tail and head-to-tail orientations: Sac7d (open circles), K28C dimer (open triangles), K9C dimer (open squares) and K9C His-tag-K28C dimer (open diamonds) in 10mM KH_2PO_4 (pH=7.2). Data points are connected with solid lines to demonstrate progressive change in CD signal.

3.4 DNA affinity Chromatography

Reverse titrations monitoring the tryptophan quenching with increasing DNA concentrations suggested that the K9C dimer with tail-to-tail orientation has a higher DNA affinity when compared to that of Sac7d monomer and dimers in other orientations. The DNA binding affinity of Sac7d dimers cross-linked in different orientations was also studied by DNA affinity chromatography. DNA affinity chromatography provides a means to study the DNA binding affinity for all the three-dimer orientations in a single experiment.

Relative differences in DNA binding affinity for the dimers cross-linked in different orientations were identified based on their respective affinities for DNA affinity matrix by using gradient elution.

3.4.1 Preparation of DNA affinity columns

DNA affinity columns were prepared by immobilizing double stranded DNA 80 mer [d(GC)] on N-hydroxy succinamide (NHS) activated HP columns (purchased from GE healthcare). The matrix material is made up of polypropylene, which is biocompatible and non-interactive with biomolecules. The matrix is functionalized with NHS that specifically react with free anime groups ($-NH_2$). Double stranded oligonucleotides with amino group modification on five prime end (purchased from Integrated DNA technologies [IDT]) were immobilized on the matrix according to the protocol provided with the column (Appendix A.2.7).

3.4.2 Preparation of Sac7d dimers cross linked in different orientations

The sample applied on the DNA affinity columns was a mixture of Sac7d dimers cross-linked in different orientations. Equimolar amounts of head cysteine mutant and tail cysteine mutant were pooled and reduced to monomers with 10 mM DTT. After carrying out the reduction process for 20 minutes at room temperature, DTT was removed by gel exclusion using PD-10 desalting columns (GE healthcare). The elutate obtained from the PD-10 columns was treated with 0.1 mM dimaide (final concentration) to induce dimerization. Dimerization of equimolar mixture of head-cysteine mutant monomers and tail-cysteine mutant monomers dimerization should result in three dimer populations,

viz., head-to-head, tail-to-tail and head-to-tail. Some amount of monomer was also present as a result of incomplete dimerization. The mixture of dimers is applied onto the DNA affinity column to separate the constituents based on their relative DNA binding affinities under 1 M NaCl gradient.

3.4.3 Separation of Sac7d dimers cross linked in different orientations

Dimers cross-linked in different orientations described in the previous section were loaded onto the DNA affinity column connected to an AKTA HPLC system. The column was equilibrated with 10 mM KH_2PO_4 buffer (pH=7.2). Protein samples were eluted with a 0-1 M NaCl gradient using10 mM KH_2PO_4 buffer (pH=7.2). The elution profile of the run performed using the dimer sample alone resulted in four poorly resolved peaks. In order to identify the peak composition, subsequent runs were performed by adding pure preparation of a dimer cross-linked in a specific orientation to the dimer sample used in the initial run. The purpose of keeping the initial injection sample constant and adding one species in pure preparation at a time is to specifically identify the peak corresponding to the species added (in pure preparation). With the addition of a particular dimer, the volume of the peak corresponding to that particular dimer should increase proportionally. Figures 3.10 through 3.13 represent the elution profiles obtained by injecting only the dimers sample, dimer sample with Sac7d monomer, dimer sample with head-to-head dimer and dimer sample with tail-to-tail dimer. The first three peaks (in the order of their elution with increasing salt gradient) were identified as Sac7d monomer, head-to-head (K28C) dimer and tail-to-tail (K9C) dimer.

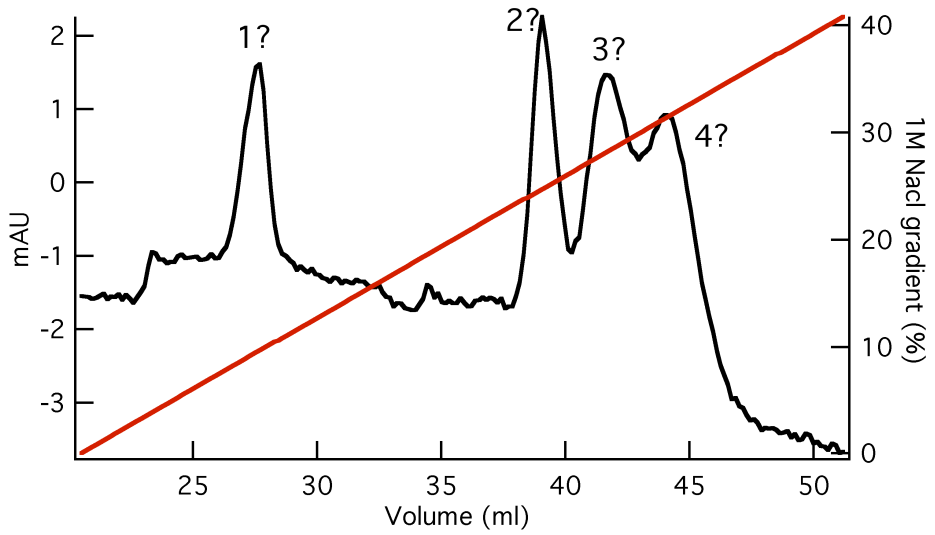


Figure 3.10 DNA affinity chromatogram obtained by injecting a dimer mixture obtained from subjecting equimolar amounts of monomer K28C and monomer K9C (head and tail mutants). The elution profile was obtained with 1M NaCl gradient.

Peaks were identified by selectively adding a particular species to the original dimer injection mixture (that was maintained constant) in separate runs. Addition of Sac7d monomer to the original dimer mixture, the intensity of the first peak eluting at lowest salt gradient increased indicating that it corresponds to Sac7d (figure 3.11). Subsequently, in separate chromatography runs, the addition of K28C dimer (head-to-head) (figure 3.12) and K9C dimer (tail-to-tail) (figure 3.13) to the original dimer mixture resulted in the increase in the intensity of peaks corresponding to K28C dimers and K9C dimer facilitating their identification. The intensity of the last peak that was eluted at the highest salt concentration did not increase either with the addition of monomer Sac7d or with the addition of the other two dimers (both K9C and K28C) indicating that it should be the dimer cross-linked in head-to-tail orientation. Due to the poor resolution of the peaks this technique was not useful for obtaining pure preparation of the head-to-tail dimer.

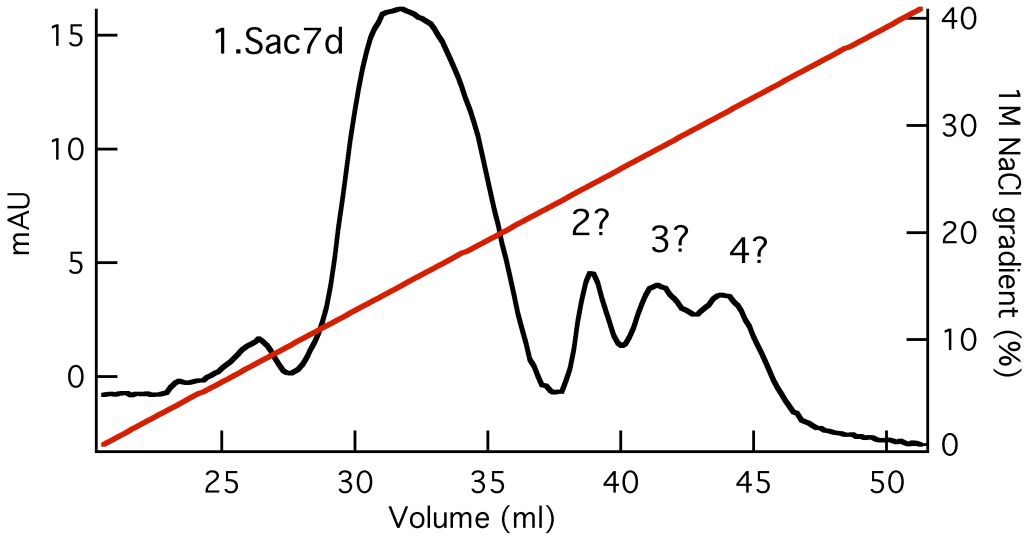


Figure 3.11 The DNA affinity chromatogram was obtained by injecting the dimer mixture described in the figure 3.6 along with a pure preparation of Sac7d monomer. The elution profile was obtained with 1 M NaCl gradient

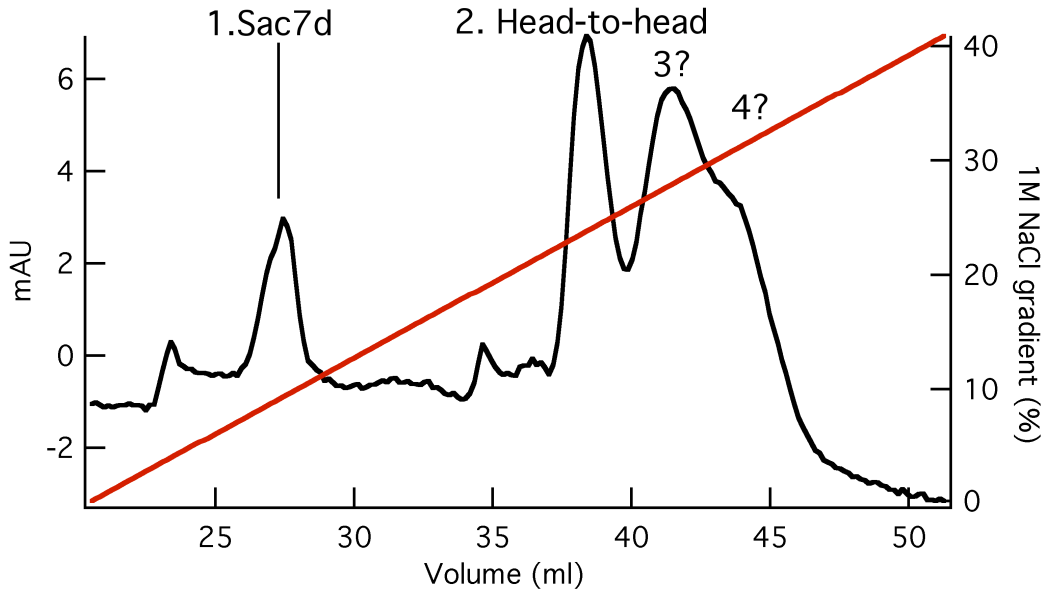


Figure 3.12 The DNA affinity chromatogram was obtained by injecting the dimer mixture described in the figure 3.6 along with a pure preparation of K28C dimer (head-to-head). The elution profile was obtained with 1M NaCl gradient

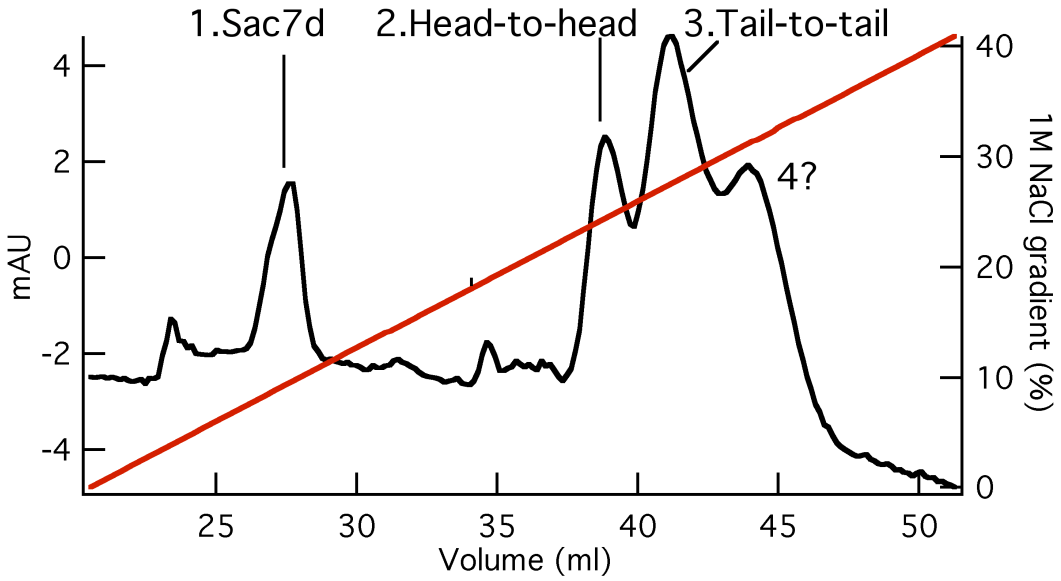


Figure 3.13 The DNA affinity chromatogram was obtained by injecting the dimer mixture described in the figure 3.6 along with a pure preparation of K9C dimer (tail-to-tail). The elution profile was obtained with 1M NaCl gradient.

3.5 Summary

DNA binding of Sac7d dimers cross-linked in the three orientations (head-to-head, tail-to-tail and head-to-tail) was studied with both reverse and forward titrations. Reverse titrations using fluorescence spectroscopy pointed out that K9C disulfide dimer in tail-to-tail orientation has 100-1000-fold greater binding affinity than that of Sac7d. Increased binding affinity of K9C disulfide dimer was independent of the DNA sequence indicating negligible role of sequence dependent DNA structure recognition by the tail-to-tail orientation by two Sac7d molecules binding at adjacent binding sites on DNA. Head-to-head orientation did not show any increased DNA binding affinity except for the 10-fold increase in binding affinity of K28C disulfide dimer that is specific to 1bbx sequence. The specific increase may be attributed to the specific interaction of the side chains of R42 resides with the narrow minor groove regions at the AT steps flanking the GC triplet in the 1bbx sequence. There are several reports of DNA binding proteins that employ arginine side chains to interact with narrow regions of minor groove, in specific AT rich regions. The interactions between the guanidine group of arginine and the narrow minor groove impose specificity and result in increased affinity (Rohs, Jin et al. 2010).

The plot of fractional change in $\Delta\varepsilon$ versus the ratio of protein to DNA concentration clearly points out the difference in the change of CD spectrum of the DNA with the change in the cross-linking orientation of the Sac7d molecules. The forward titrations of K9C disulfide dimer with 100-mer [d(GC)] indicates cooperative unwinding of DNA with the increase in protein concentration. The forward titrations of head-to-head and head-to-tail dimers are comparable to that of Sac7d in having two isodichroic points and the DNA structural change leveling off at saturating protein concentrations. The

fractional change in the Δε curve for K9C dimer is comparable to that obtained with the forward titration of Sac7d into dIdC that shows a cooperative unwinding in DNA with increasing Sac7d concentration (McAfee, Edmondson et al. 1996). This indicates that at low protein concentration, it is difficult to unwind the DNA and it takes a synergistic effect from multiple proteins to bring about cooperative unwinding.

The structural change in the DNA observed in the forward titrations with Sac7d dimers cross-linked in different orientations indicates that tail-to-tail orientation promotes cooperative unwinding in DNA. This indicates that the proteins must be within a set distance to cause structural change in the DNA. As a result, in a forward titration no structural change in DNA is noticed until a specific protein to DNA ratio is reached. But in case of reverse titrations (DNA into protein), with the addition of first few aliquots of DNA into the protein, all the protein is bound. The DNA is saturated and the protein dimers are bound in close proximity to influence the binding orientation of each other mediated by the DNA deformation. This indicates that the observations from forward and the reverse titrations agree with each other.

The potential preference in the binding orientation for adjacent proteins imposed by the DNA deformation caused by one half of the dimer on the second half. Hence, in case of a binding of Sac7d dimers cross-linked in all the possible orientations, the dimer cross-linked in potentially preferred orientation should out compete the dimers cross-linked in alternate orientations. DNA affinity chromatography was performed to test this hypothesis and to separate the dimers based in the differences in their DNA binding affinities. Based on NaCl gradient elution and peak identification, Sac7d dimer cross-linked in head-to-tail orientation was inferred to be binding to the DNA columns with

maximum affinity. It was not possible to separate the fractions of different dimers due to poor resolution of the peaks. Poor resolution may be attributed to the inadequate sensitivity in separation and also indicative of minute difference in the binding affinities between the different fractions.

It is difficult to identify the dimer with preferential binding orientation based on the DNA chromatography data. The order of elution may not reflect the binding constant at a single lower salt concentration, as the effect on salt concentration on the affinities of the dimers may be different. The difference in affinities could potentially result from the non-uniform deformations caused by the simultaneous binding of dimers in all the three cross-linking orientations. As a result the structural change in the DNA imposed by simultaneous binding of all the dimer orientations may not depict the structural change imposed due to dynamic non-sequence specific binding of Sac7d monomer. Constraints due to cross-linking may also interfere with otherwise dynamic non-specific binding mode of Sac7d. Cross-linking may also influence the preferential binding orientation and the increased affinity resulting from it. In order to confirm the preference in binding orientation, the binding of two adjacent Sac7d molecules on DNA should be also studied without constrains from cross-linking as well.

CHAPTER 4

IDENTIFICATION OF NEAREST NEIGHBOR ORIENTATION BY CROSSLINKING Sac7d ON DNA

4.1 Introduction

Cross-linking Sac7d molecules in different possible orientations provides a convenient means to probe for any potential preference in DNA binding orientation for adjacent Sac7d molecules, by searching for increased affinity resulting from difference in cross-linking orientation. But dimer binding may also be influenced by crosslinking constraints that are otherwise absent in dynamic non-specific binding of monomer Sac7d. In order to study the binding orientation of adjacent Sac7d molecules on DNA without the constraints from cross-linking, the process of cross-linking can be delayed until the binding is done. Proteins bound to DNA are subject to cross-linking conditions in order to cross-link the amino acid residues that are in proximity as a result of DNA binding. The predominant dimer species formed due to the influence from DNA binding is identified by comparison with a cross-linking reaction in the absence of DNA, that serves as control. The method is extended to study the influence of DNA sequence in dictating

potential preference in binding orientation for two Sac7d molecules that bind at adjacent sites. The cross-linking reactions are analyzed using gel electrophoresis to analyze the dimers species formed as a result cross-linking in the presence of DNA. Potential preference in the binding orientation for adjacent Sac7d molecules resulting from DNA binding was studied by qualitative analysis of the dimers using SDS-PAGE.

Equimolar amounts of single cysteine mutants of Sac7d in head (K28C) and tail (K9C His-tag) regions are bound to DNA as monomers in the presence of DTT. DTT is removed by desalting and the protein-DNA complex is subject to cross-linking conditions. The dimers obtained are analyzed using SDS-PAGE and mass-spectroscopy.

4.2 Materials and methods

The concentration of the protein and the DNA for the binding reactions were selected in such away that the DNA was saturated. Saturation of the DNA with the Sac7d single cysteine mutants results in placing the adjacent protein in adequate proximity to facilitate cross-linking. Multiple DNA sequences and sequence lengths (described below) are used in this study.

Equimolar amounts of head cysteine mutant (K28C) and tail cysteine mutant (K9CHis-tag) at final concentration of 21 μM were mixed with double stranded DNA at a final concentration of 15 μM (4bp binding sites). Multiple DNA sequences were tested in separate experiments, with similar concentrations. The protein-DNA complex was incubated with 25 mM DTT in a total volume of 2.5 ml for 30 minutes at room temperature. DTT is removed by size exclusion chromatography on PD-10 columns. The flow through was treated with appropriate cross-linker as described in appendix A.1.12.

Three sets of control reactions were conducted in the same manner: a) with DNA-without cross-linker, b) without DNA-with cross-linker and c) without both DNA and cross-linker.

4.3 Results and discussion

Figures 4.1 through 4.3 display the cross-linking reactions with 1bbx sequence (12mer: 5'-CTAGCGCGCTAG-3') 24 mer [d(IC)] and 100-mer [d(GC)]. In all the three figures lanes 3-4, 5-6 and 7-8 are pairs of control reaction and experimental reaction conditions, control being without DNA and experiment being in the presence of DNA. Lanes 3-4 are without any cross-linker, lanes 5-6 are with diamide (for cross-linking by oxidation) and lanes 7-8 are with p-PDM (for chemical cross-linking).

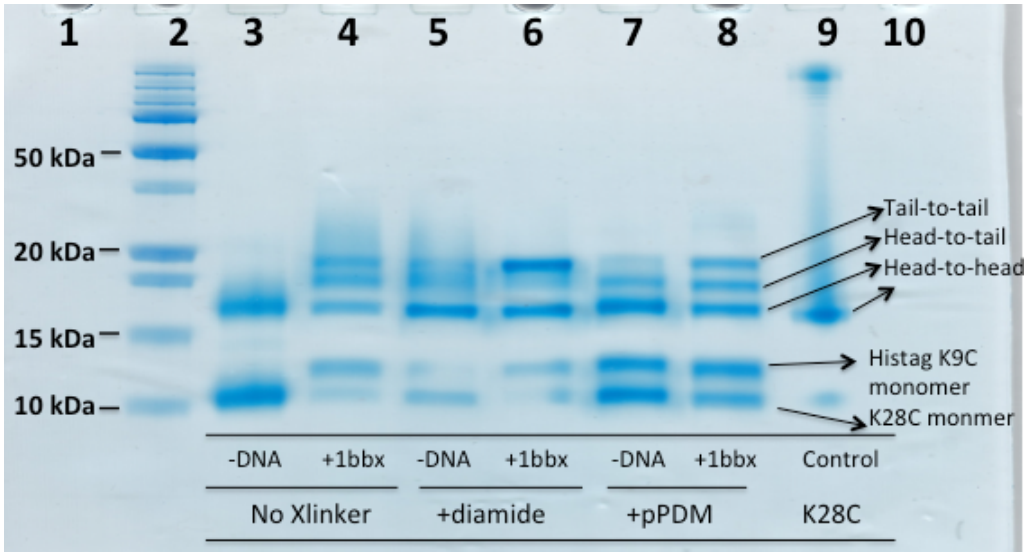


Figure 4.1 Cross-linking adjacent Sac7d on DNA with 1bbx sequence.

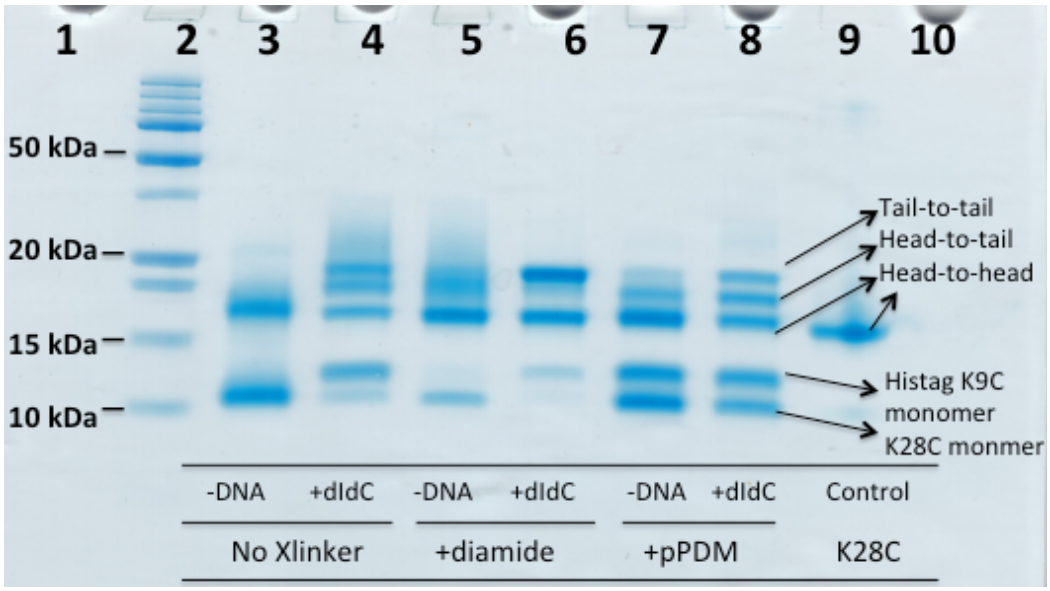


Figure 4.2 Cross-linking adjacent Sac7d on DNA with 24mer [d(IC)] sequence.

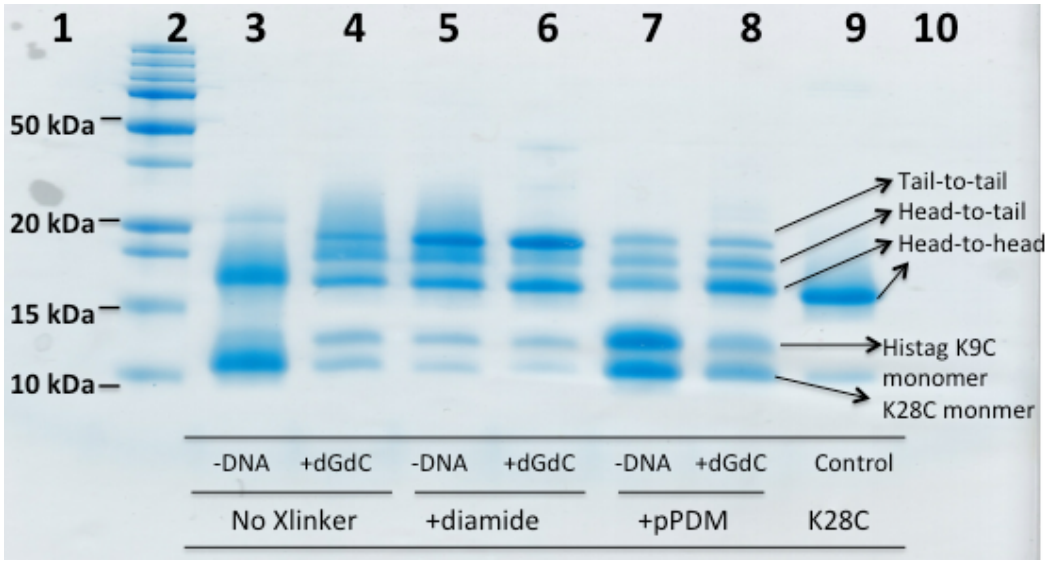


Figure 4.3 Cross-linking adjacent Sac7d on DNA with 100-mer [d(GC)] sequence

Lane 9 displays head-to-head dimers (with out any His-tags) and serves a control. In all the three gels, lane 3 is a control reaction containing only K28C and K9CHis-tag without DTT and cross-linking agent. Lane 4 is similar to the contents of lane 3 but with appropriate DNA. Lane 4 displays well resolved dimer bands even in the absence of a cross-linking agent.

In lanes 5 and 6 the control and the experimental reaction were treated with 0.1 mM final concentration of diamide. DNA was added to the experimental reaction and appropriate volume of 10mM phosphate buffer was added to the control reaction. In case of addition of 1bbx, [d(IC)] tail-to-tail dimer band (K9CHis-tag dimer) is intense compared to the control reaction. A clear difference in intensity for different dimer population is not apparent in case of [d(GC)].

In case of lanes 7 and 8, control and the experimental reaction were treated with 40 μM of pPDM. DNA is added to the experimental reaction and appropriate volume of 10mM phosphate buffer was added to the control reaction. In case of 1bbx and [d(IC)] the intensity of the band corresponding to the tail-to-tail dimer is greater in the experimental in comparison to the control lane. In case of 100-mer [d(GC)] it is difficult to distinguish the intensity of the dimers bands in the control and experimental conditions. But there is a clear reduction in the intensity of the monomer bands in the experimental condition when compared to the control reaction without [d(GC)].

The difference in the intensities of the bands corresponding to the three dimer orientations in the control reactions (without the DNA) may be attributed to the differences in the accessibility of the cross-linking groups in the different mutants (the head and the tail mutants).

4.4 Conclusions

Addition of DNA to the Sac7d cysteine mutants resulted in the formation of dimers even in the absence of cross-linking conditions. This indicates that Sac7d-DNA binding results in positioning the adjacent proteins close enough for cross-linking of sulfhydryl groups in the cysteine mutants. In the presence of cross-linking conditions there is a clear preference for tail-to-tail binding orientation evident from the stronger intensity of the band corresponding to K9CHis-tag dimer in case of 1bbx and [d(IC)]. Thus the results obtained from the cross-linking on DNA clearly indicate that K9C His-tag dimer (tail-to-tail) orientation is the predominant dimer species obtained by cross-linking monomer cysteine Sac7d mutants that are bound to DNA. This indicates that a majority of the protein molecules bound with their respective tail regions pointing towards each other.

Binding of proteins with tails pointing at each other should result in a head-to-head orientation with the neighbor on the other side. However, cross-linking may not be possible between all the ends pointing towards each other due to two reasons: 1. Lack of reactive groups and 2. Lack of adequate proximity between the reactive groups. Lack of reactive groups can be rectified by experimental modification. Lack of proximity is dictated by protein binding-induced DNA deformation. Introducing reactive sites in both head and tail region of the same molecule to obtain dimers (in both head-to-head and tail-to-tail orientations) and oligomers raises a new question. Which orientation out of the two obtained was formed as result of protein binding induced DNA deformation? In order to identify the preferred orientation imposed by protein binding induced DNA deformation on the protein that binds at the adjacent site, the binding of two Sac7d molecules should

be studied on oligonucleotide that can accommodate no more than two Sac7d molecules. 1bbx sequence used in the present cross-linking study has been shown to bind two-Sso7d (Sac7d homologue from *Sulfolobus sulfactaricus*) (Agback, Baumann et al. 1998).

Cross-linking on 1bbx showed the formation of tail-to-tail dimers along with a band corresponding to head-to-head dimers. Hence, in order to obtain a clear indication of preferred orientation between adjacent Sac7d molecules on DNA arising from protein binding induced DNA deformation, the constraints imposed by cross-linking should be eliminated. A confirmation for preferred orientation between adjacent Sac7d molecules on DNA should be obtained by studying the protein-DNA binding without cross-linking the adjacent molecules to allow the dynamic non-specific binding.

CHAPTER 5

PARAMAGNETIC RELAXATION ENHANCEMENT ASSAY

5.1 Introduction

The indication of preferred DNA binding orientation for adjacent Sac7d molecules obtained from DNA binding analysis of Sac7d dimers should be confirmed by studying the binding of Sac7d to adjacent sites on DNA in the absence of cross-linking constraints. The binding orientation of adjacent Sac7d molecules on the DNA should be investigated by allowing for dynamic non-sequence specific binding mode of Sac7d that might otherwise be influenced by constraints from cross-linking. An ideal test system to study the preferential binding orientation imposed by Sac7d binding on the adjacent proteins should consist of two Sac7d molecules and DNA with two DNA binding sites so that the potential influence of first protein binding event on the second protein binding can be clearly tested. Also, the system should enable clear identification of the relative orientation in which the adjacent proteins are bound on the DNA.

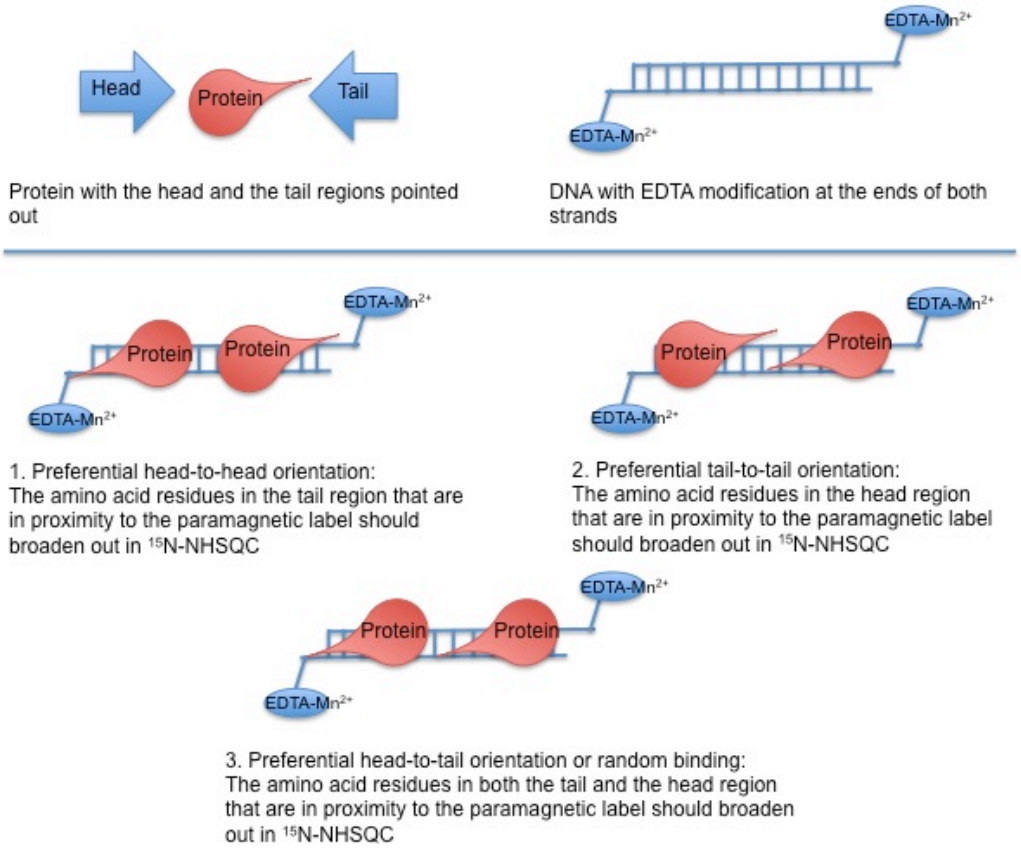


Figure 5.1 Scheme depicting the use of PRE to study the preferential binding orientation of adjacent Sac7d molecules on DNA. The top panel displays the protein molecule (with its head and tail regions clearly pointed out) and DNA with terminal EDTAMn^{2+} modifications (paramagnetic centers). Lower panel displays the three possible binding orientations for adjacent Sac7d molecules: 1. Proteins bound in head-to-head orientation resulting in placing the tail regions close to the paramagnetic label 2. Proteins bound in tail-to-tail orientation resulting placing the head residues close to the paramagnetic label. 3. Proteins bound in head-to-tail (or random orientation on a longer oligonucleotide). Analysis of the ^{15}N-HSQC spectrum of ^{15}N labeled protein and EDTA-Mn^{2+}labled DNA should indicate the preferential binding orientation based on the set of peaks that broaden out.

Paramagnetic relaxation enhancement (PRE) assay meets the above requirements for the analysis of potential preference in DNA binding orientation for adjacent Sac7d on DNA. PRE provides a convenient method to identify the nuclei (with a magnetic moment) that are in proximity to a paramagnetic center (~20Å). In a ^{1}H, ^{15}N-HSQC,

introduction of a paramagnetic center results in the broadening of the peaks corresponding to nuclei that are close to the paramagnetic center. Therefore, the proximity of amino acid residues in labled protein samples to a paramagnetic center can be identified using the PRE assay. Introduction of one paramagnetic center each at the two ends of a B-DNA that can accommodate only two Sac7d molecules serves as the ideal test system mentioned above. Scheme presented in figure 5.1 depicts the system presented above and the discussion to follow.

B-DNA with 10 to 12 base pairs that can accommodate two Sac7d molecules are used. A paramagnetic center is introduced at the two ends of the DNA molecule by chelating Mn^{2+} ions to EDTA that is attached to deoxythymidine (dT). In a ^{1}H, 15N-HSQC spectrum collected with appropriate ratio of ^{1}H, ^{15}N-Sac7d and DNA with EDTA-Mn^{2+} so that the DNA is saturated in order to allow for adjacent binding of proteins, the peaks corresponding to the amino acids that are in proximity to the paramagnetic label should selectively broaden out due to paramagnetic relaxation. The relative binding orientation is understood by identifying the ends of the two proteins that are close to the paramagnetic center. In case of preferential head-to-head orientation between the two proteins, the tail regions point away from each other and hence they are close to the paramagnetic center at the 5'-end of both the DNA strands. As a result peaks corresponding to the tail region amino acids broaden out. In case of a preferential tail-to-tail orientation, the head regions pointing away from each other are in proximity to the paramagnetic center resulting their broadening out. In case of a preferential head-to-tail orientation or random binding orientation, the peaks corresponding to the amino acids in both the head and the tail regions should broaden out (figure 5.1).

5.2 Paramagnetic Relaxation Enhancement

PRE arises from magnetic dipolar interactions between a nucleus (^{1}H and ^{15}N) and unpaired electrons (paramagnetic center) and results in an increase in the relaxation rate of the nuclear magnetization (Bloembergen and Morgan 1961; Clore, Tang et al. 2007; Iwahara, Tang et al. 2007). The magnitude of PRE is proportional to r^{-6} where, r is the distance between the paramagnetic electron and the nucleus. As the magnetic moment of the unpaired electrons is large, so are the PRE effects and hence can provide long-range distance information (Clore, Tang et al. 2007). As the effect of PRE is distance dependent, the intensity of the peaks of the corresponding amino acid residues varies with their proximity to the paramagnetic center. The proximity of a given domain in a biomolecule such as protein or DNA can be estimated with the increase or decrease in the intensity of the peaks of the corresponding nuclei. Introduction of paramagnetic metal ions such as Mn^{2+} serves as a paramagnetic center. Chemical attachment of EDTA at the desired position serves as a convenient method to chelate metal ions. Chelating diamagnetic ions like Ca^{2+} to EDTA serve as diamagnetic controls to study the effect of paramagnetic relaxation. There may be a decrease in peak intensity in experiments with EDTA-Ca^{2+} also. The decrease in the intensity of the peaks in the experiments with EDTA-Ca^{2+} may be attributed to the exchange broadening with the addition of macromolecule with the label. In case of protein-DNA complexes, the decrease in peak intensity may also be due to the increase in the size of molecules. The size of the molecules in solution also affects the spin relaxation. The rotational motion of the molecules is primarily characterized by correlation time (τ_c) which is proportional to the relaxation times. With the increase in the size of the complexes due to protein DNA

binding, the correlation time increases. In order to maintain constant integrated peak intensities (when there is no change in the number of nuclei) peak height decreases. Comparison of the decrease in the peak intensity from the paramagnetic and diamagnetic experiments clearly indicates the effect of PRE and hence the proximity of nuclei of interest to the paramagnetic center (Wüthrich 1986; Cavanagh, Fairbrother et al. 1995).

5.2.1 Application of PRE for biomolecules

PRE has been extensively used in the study of metallo-proteins due to the presence of intrinsic paramagnetic centers. Introduction of extrinsic paramagnetic centers in biomolecules like proteins and DNA has also been carried out. PRE arising from unpaired electrons with an isotropic g-tensor as in the case of a nitric oxide (NO) spin label or EDTA-Mn^{2+} has extensive applications in the study of biomolecules. The advantage of an isotropic system over the anisotropic ones is absence of pseudo-contact shifts and Curie-spin relaxation that could potentially exhibit significant cross-correlation with other relaxation mechanisms (Clore and Gronenborn 1991; Clore, Tang et al. 2007). PRE arising from unpaired electrons with an isotropic g-tensor have been extensively used for studying the structures and interactions between the biological macromolecules (Kosen 1989; Battiste and Wagner 2000; Gaponenko, Howarth et al. 2000; Varani, Gunderson et al. 2000; Donaldson, Skrynnikov et al. 2001; Dvoretsky, Gaponenko et al. 2002; Hansen, Hass et al. 2003; Iwahara, Anderson et al. 2003; Iwahara, Schwieters et al. 2004; Iwahara, Schwieters et al. 2004; Iwahara, Akagi et al. 2004; Roosild, Greenwald et al. 2005; Iwahara and Clore 2006; Iwahara, Zweckstetter et al. 2006; Liang, Bushweller et al. 2006; Tang, Iwahara et al. 2006). PRE has also been used for studying the non-

specific protein-DNA interactions (Johnson, Brun et al. 1999; Card, Erbel et al. 2005). Intermolecular PRE can provide a powerful probe in the fast exchange regime for the detection and the characterization of transient and lowly populated in macromolecular binding (Ramos and Varani 1998; Jain, Venot et al. 2001).

Non-specific protein-DNA involving fast exchange has been studied using PRE (Iwahara, Schwieters et al. 2004; Iwahara, Zweckstetter et al. 2006). The Sac7d-DNA binding system involves non-sequence specific binding and operates with fast exchange. Use of paramagnetic relaxation enhancement to analyze the potential preference in the binding orientation for adjacent Sac7d on DNA is presented here.

5.3 Materials and methods

5.3.1 Oligonucleotides and introduction of the paramagnetic probe

Oligonucleotides with an EDTA attachment at the penultimate dT were purchased from Midland certified reagent company (Midland, Texas). The EDTA modification reaction is customized for the thiamine ring and hence the DNA sequence should essentially contain a dT at the position where EDTA modification is required. Structure of EDTA attachment to the thiamine ring is shown in the appendix D.1. EDTA serves as a chelating agent for the introduction of metal ions. Paramagnetic ions like Mn^{2+} chelated to EDTA serve as the paramagnetic center to study PRE. Diamagnetic ions like Ca^{2+} are chelated for control spectra.

Single stranded EDTA modified oligonucleotides in 10 mM KH_2PO_4 pH 7.2 were annealed to obtain double stranded DNA. Appropriate amounts of $MnCl_2$ were added to the DNA solution and incubated to allow chelation of all the EDTA groups with Mn^{2+}.

The DNA samples were purified by anion exchange chromatography (using Resource Q column) using 10 mM KH_2PO_4 followed by elution with1M NaCl gradient. The fractions obtained were desalted and concentrated using centrifugal filter unit (Millipore) concentrators for use in NMR titrations.

5.3.2 Protein sample preparation and NMR titrations

^{15}N-labeled Sac7d was prepared and purified according to the standard procedure described in the appendix A.1.4. 500 µl of NMR sample was prepared at a final concentration of 0.3 mM of ^{15}N-Sac7d in 10 mM KH_2PO_4, with 10% D_2O and pH=5.0. A small volume of 4,4 dimethyl, 4-silapentane sulfonic acid (DSS, ~5 µl of 14 mM) was added for reference. Aliquots of DNA with the paramagnetic label were added in increments so that the DNA to protein concentration ratios were approximately in the order of 0, 0.05, 0.1, 0.15, 0.2 and 0.3. ^{15}N-HSQC spectra were collected at each increment. The data was processed using NMRView Java software to obtain the intensities of the peaks corresponding to all the amino acids in Sac7d. The change in the intensity of the peaks corresponding to amino acids in different regions of the protein (head and tail) with the addition of DNA containing paramagnetic label was compared. The degree of change in intensities of peaks corresponding to amino acids in particular regions of Sac7d (with increase in the DNA with paramagnetic label) translates to their respective proximities to the paramagnetic center. The peaks corresponding to the amino acids that are close to the paramagnetic label broaden out more than those that are far away from the label.

5.4 Results and discussion

Figure 5.2 and 5.3 show the ^{15}N-HSQC images for ^{15}N-Sac7d without any DNA and with 12-mer [d(GC)]EDTA-Mn^{2+} labeled DNA respectively. The intensity of peaks for G27, K28, and V30 in the turn between β3 and β4 (head-region) decreased with addition of EDTA-Mn^{2+} labeled DNA. The intensity of G43 and V45 in β5 also decreased as the later half of β5 is close to the defined head region.

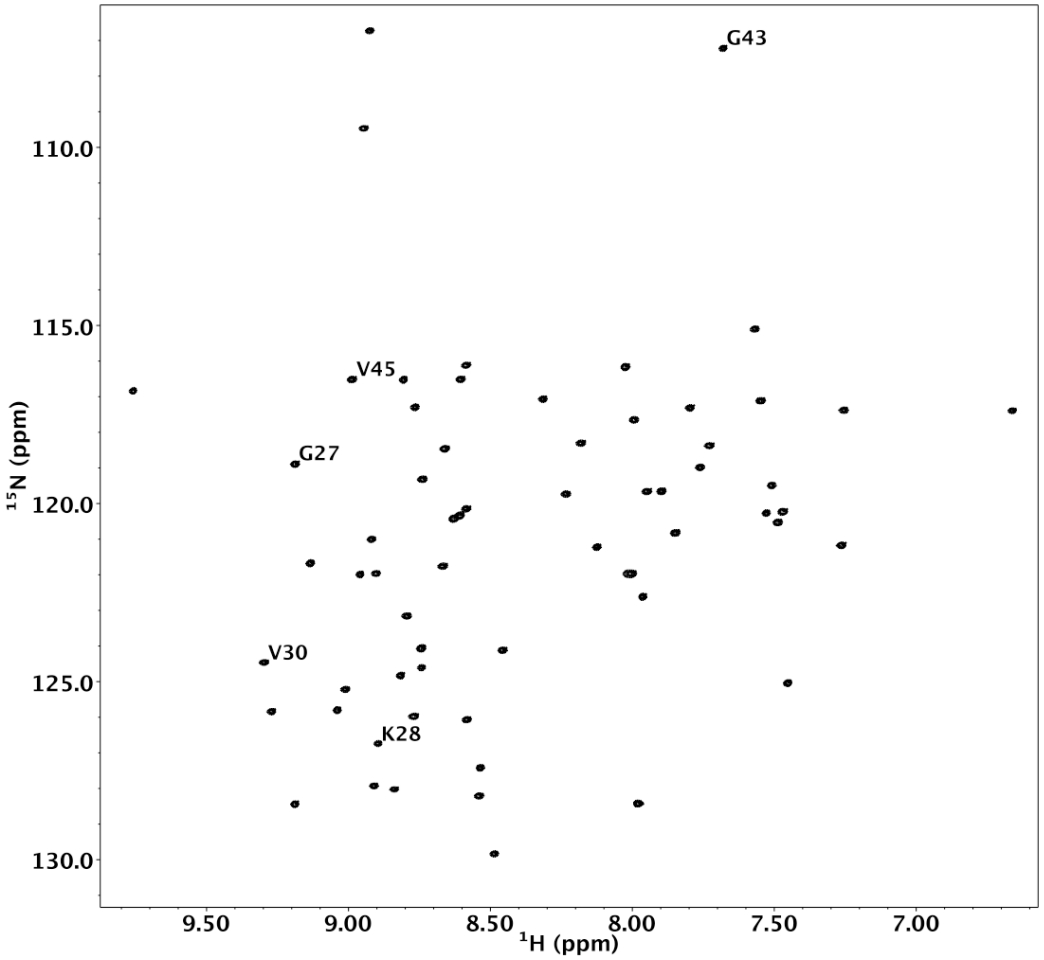


Figure 5.2 ^{15}N-HSQC of ^{15}N-Sac7d without DNA. ~2.5mg of protein in 10 mM KH_2PO_4 pH=5.0 with 10% D_2O. The peaks labeled showed selective drop in intensity with the addition of [d(GC)]EDTA-Mn^{2+} (as shown in figure 5.3). The peaks that are broaded only with the addition of Mn^{2+} are labeled. Certain peaks like K9, N37 that disappear with the addition of DNA with Ca^{2+} are not labeled.

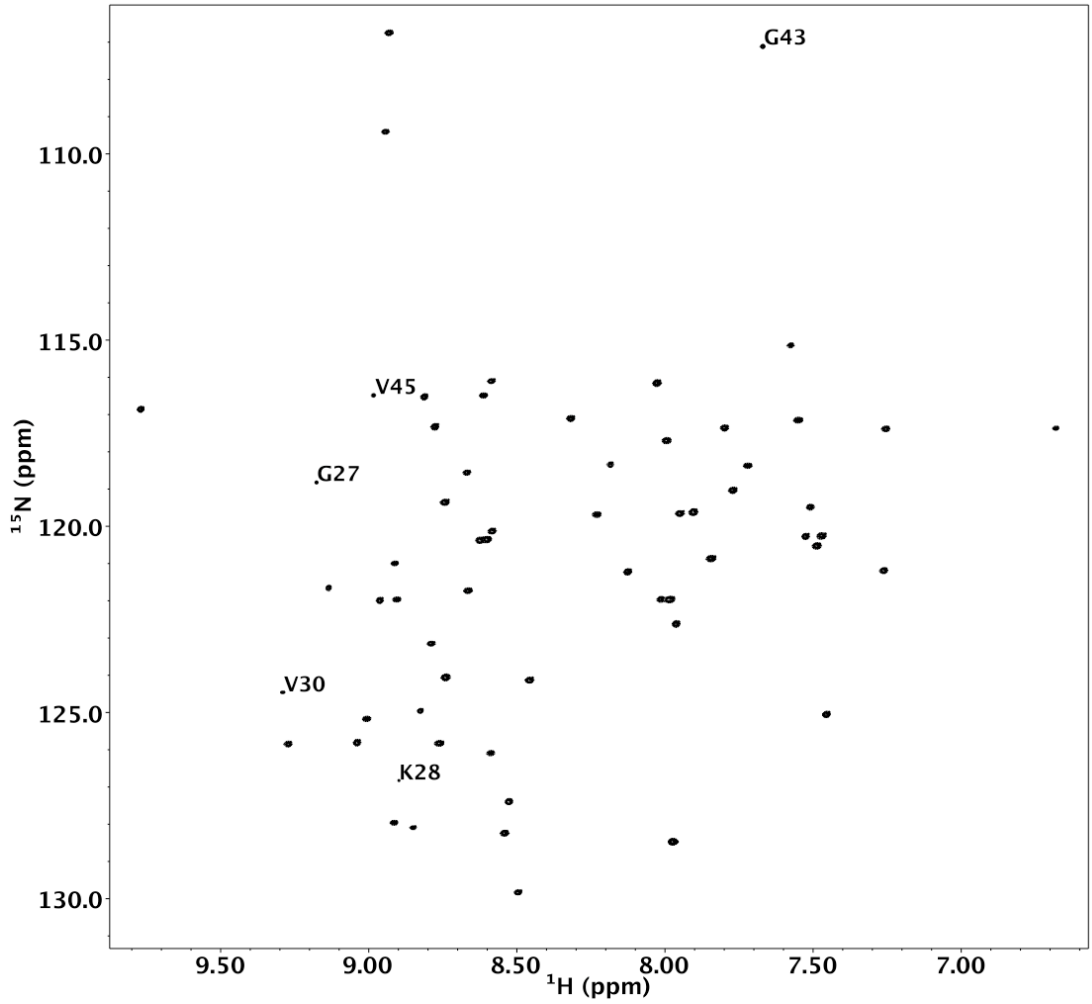


Figure 5.3 ^{15}N-HSQC of ^{15}N-Sac7d without DNA. ~2.5mg of protein and [d(GC)]EDTA-Mn^{2+} in 10 mM KH_2PO_4 pH=5.0 with 10% D_2O. The ratio of concentrations of DNA to protein was 0.25. The peaks labeled showed selective drop in intensity with the addition of [d(GC)]EDTA-Mn^{2+}.

Figure 5.4 displays a model with two Sac7d molecules in tail-to-tail orientation on a B-DNA (11 base-pairs) with EDTA modification. The residues that showed decrease in the peak intensity with the addition of DNA with paramagnetic label are shown in blue. The model indicates that the decrease in the intensity for G27, K28, V30, G43 and V45 can be explained if the two Sac7d molecules are bound in a preferential tail-to-tail

orientation. Figure 5.5 displays two Sac7d molecules bound in head-to-head orientation on a B-DNA (11 base-pairs) with EDTA modification for comparison. Based on the model in figure 5.5, binding of two proteins in a head-to-head orientations should result in the broadening of peaks corresponding to the amino acids in the tail region (Y8, K9 and G10), in the turn between β4 and β5 and the first half of β5.

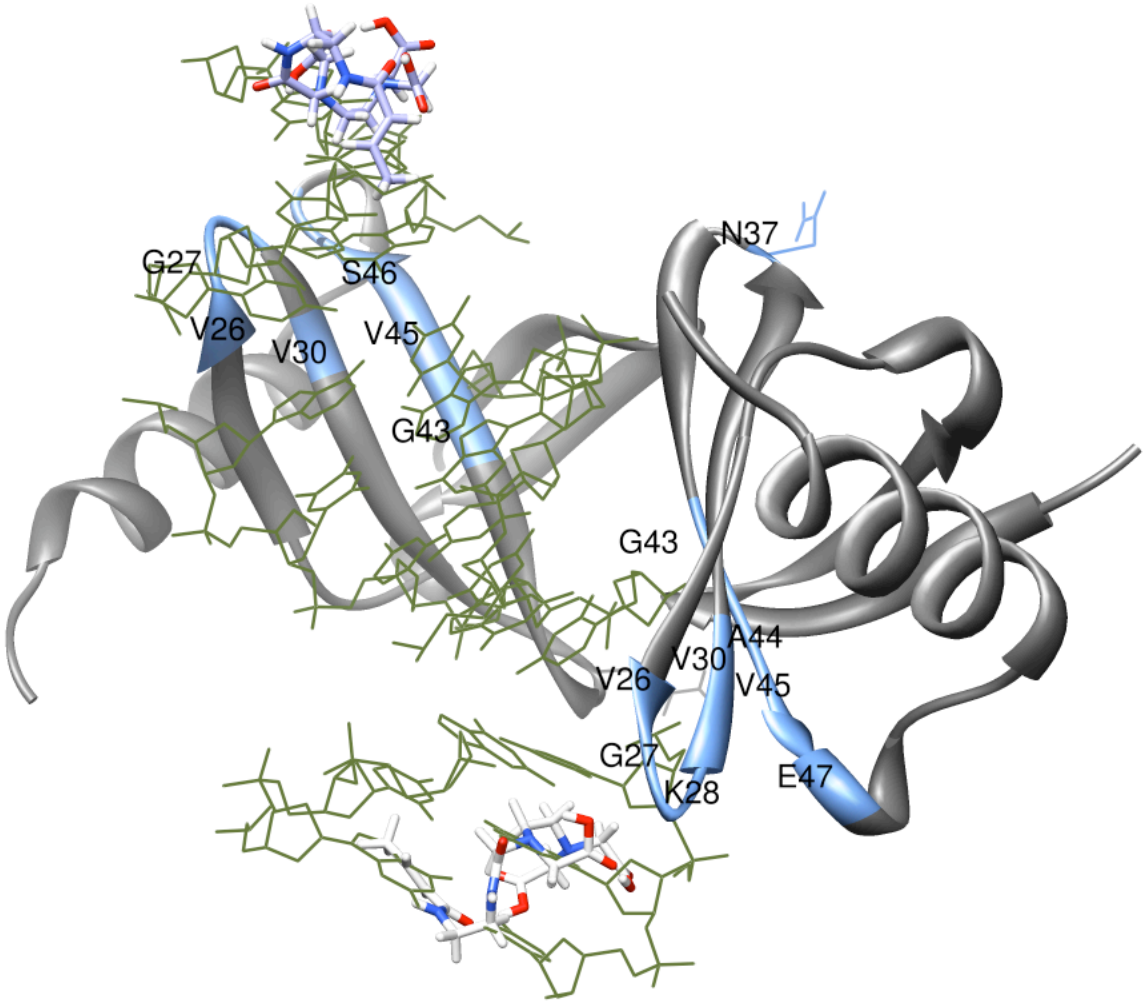


Figure 5.4 Two Sac7d molecules (grey ribbon) bound in tail-to-tail orientation on a 11-mer dGdC (green wire). The EDTA modification (multicolor wire) is shown at the two ends of the DNA. Labeled in blue are the amino acid residues that are in proximity to the EDTA-Mn^{2+} paramagnetic center for which the peak intensity decreased with the addition of [d(GC)]EDTA-Mn^{2+}.

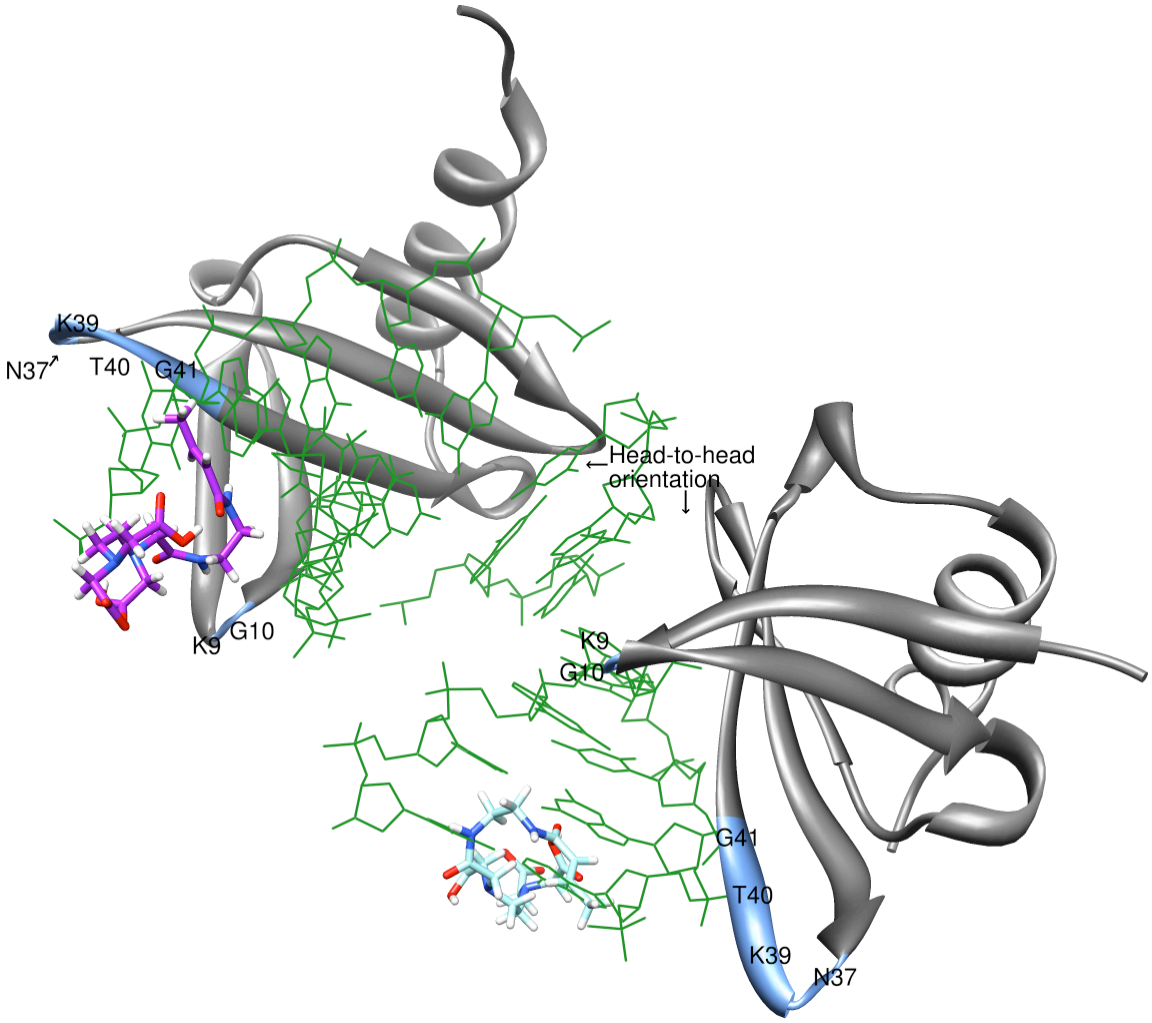


Figure 5.5 Two Sac7d molecules (grey ribbon) bound in head-to-head orientation on a 11-mer dGdC (green wire). The EDTA modification (multicolor wire) is shown at the two ends of the DNA. Labeled in blue are the amino acid residues that are in proximity to the EDTA-Mn^{2+} paramagnetic center for which the peak intensity should decrease with the addition of [d(GC)]EDTA-Mn^{2+}, in case of a preferential head-to-head orientation for the adjacent Sac7d molecules.

Figure 5.6 displays the drop in peak intensities for amino acids in different regions plotted against the ratio of DNA to protein concentration. The intensity drop profiles indicate that with addition of increments of DNA with paramagnetic label the intensity of the peaks in the head region (described in the model 5.4) decreased to a greater extent in comparison to those in the tail region. Figure 5.7 displays the percentage decrease in the peak intensities in the form of histograms for clear understanding. The data indicate that there is a degree of decrease in intensity for the residues in the head region greater than those in the tail region. K9 in the tail region is an exception. The intensity of K9 decreased (almost disappeared) with the addition of DNA with paramagnetic label.

Figures 5.8 and 5.9 display the data for control experiments conducted with EDTA-Ca^{2+} DNA. Replacing Mn^{2+} with Ca^{2+} eliminates the paramagnetic relaxation effects and serves as a control. Titrations using DNA without paramagnetic label indicates that the intensity of K9 and G27 decreased with the addition of DNA. Hence decrease in the intensity of K9 in the tail region may not due to paramagnetic relaxation. The degree of decrease in the intensities of the remaining peaks in the head-region was less with Ca^{2+} than that with Mn^{2+}. The decrease in the peak intensities with the addition of DNA with EDTA-Ca^{2+} may be due to exchange broadening due to increase in molecular weight. Comparing the decrease in the peak intensities with the addition of DNA with EDTA-Mn^{2+} and EDTA-Ca^{2+} clearly indicates the effect of paramagnetic relaxation enhancement. The decrease in the peak intensity due to PRE represents the proximity of the respective nuclei to the EDTA-Mn^{2+} label.

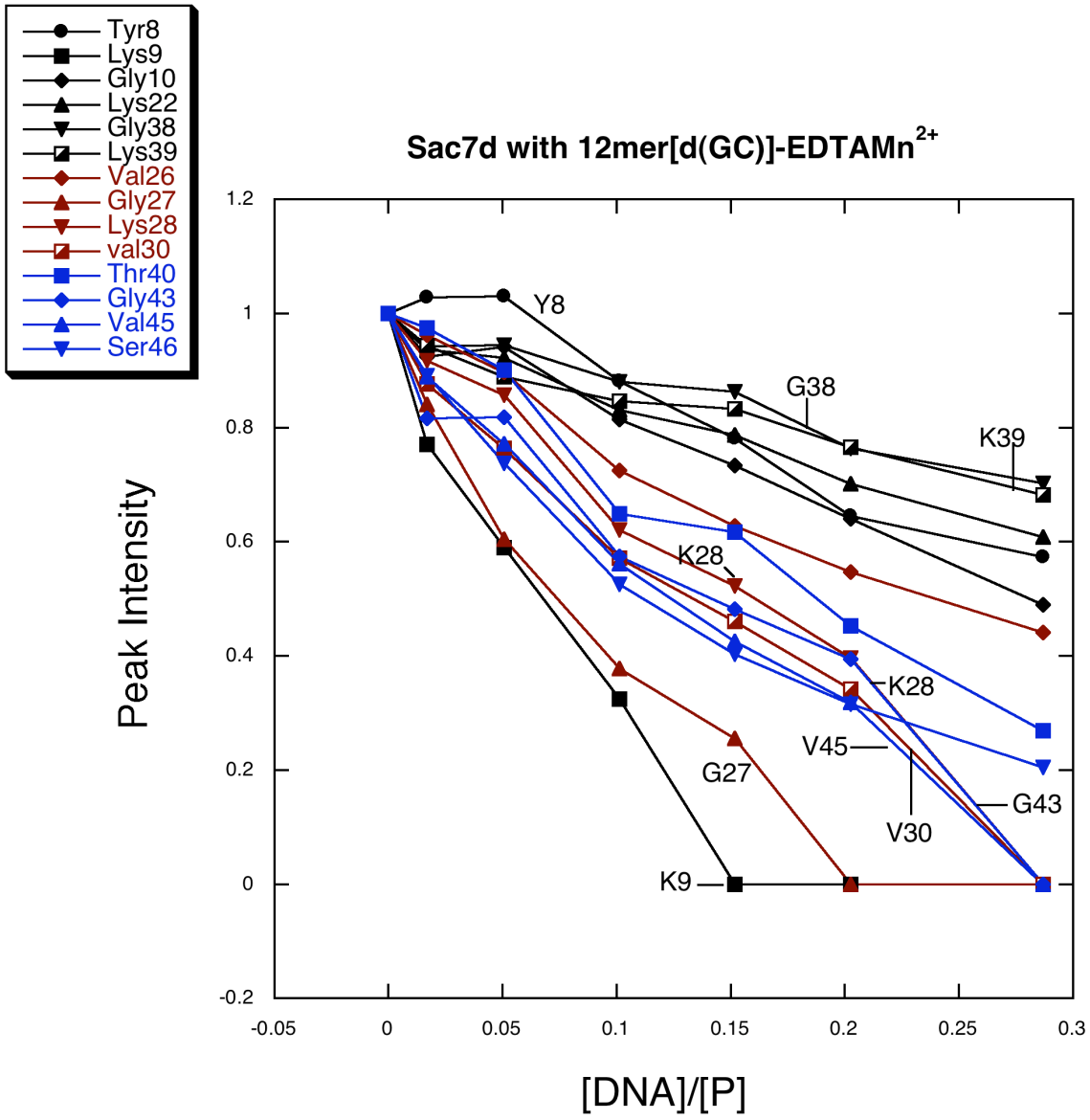


Figure 5.6 The decrease in intensity for the peaks in the head (blue and red curves) and the tail (black curves) regions of Sac7d with the addition of [d(GC)]EDTA-Mn^{2+}. The concentration of DNA is expressed in mM nucleotides. The intensity of the peaks is plotted on the Y-axis and the intensity values are normalized to start all the curves at 1.0

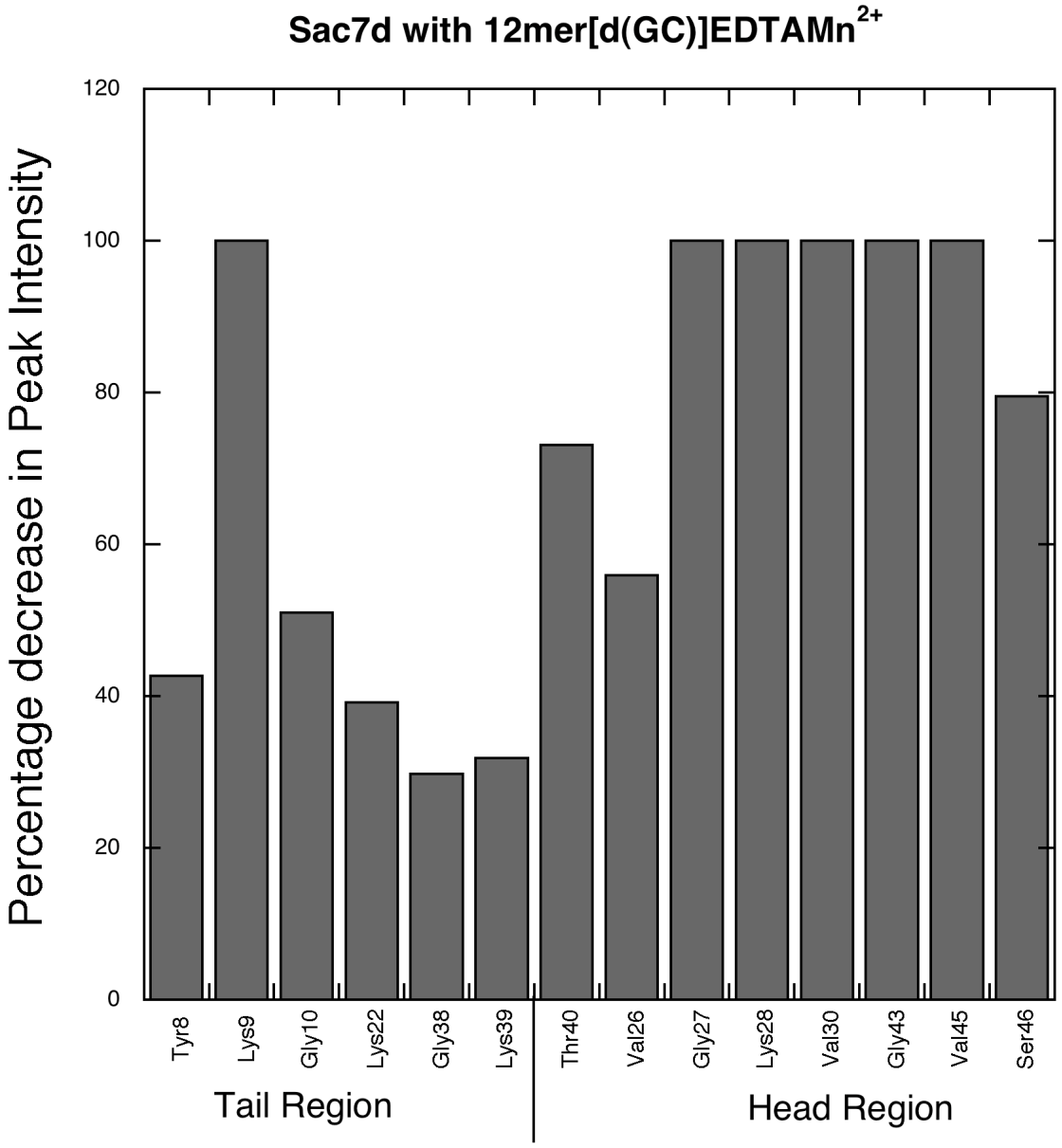


Figure 5.7 A Plot of percentage decrease in the peak intensity for the residues in head and tail regions with the addition of [d(GC)]EDTA-Mn^{2+}. The ratio of concentration of DNA to protein is ~0.3 (mM nucleotides to mM Protein).

Comparison of the PRE data using DNA with Mn^{2+} and Ca^{2+} indicates a preference in tail-to-tail orientation between adjacent Sac7d molecules on 12mer[d(GC)]. In order to understand the influence of DNA bending imposed by Sac7d on the preferred orientation, the PRE analysis was conducted using V26A, M29A double mutant of Sac7d with [d(GC)]EDTA-Mn^{2+}.

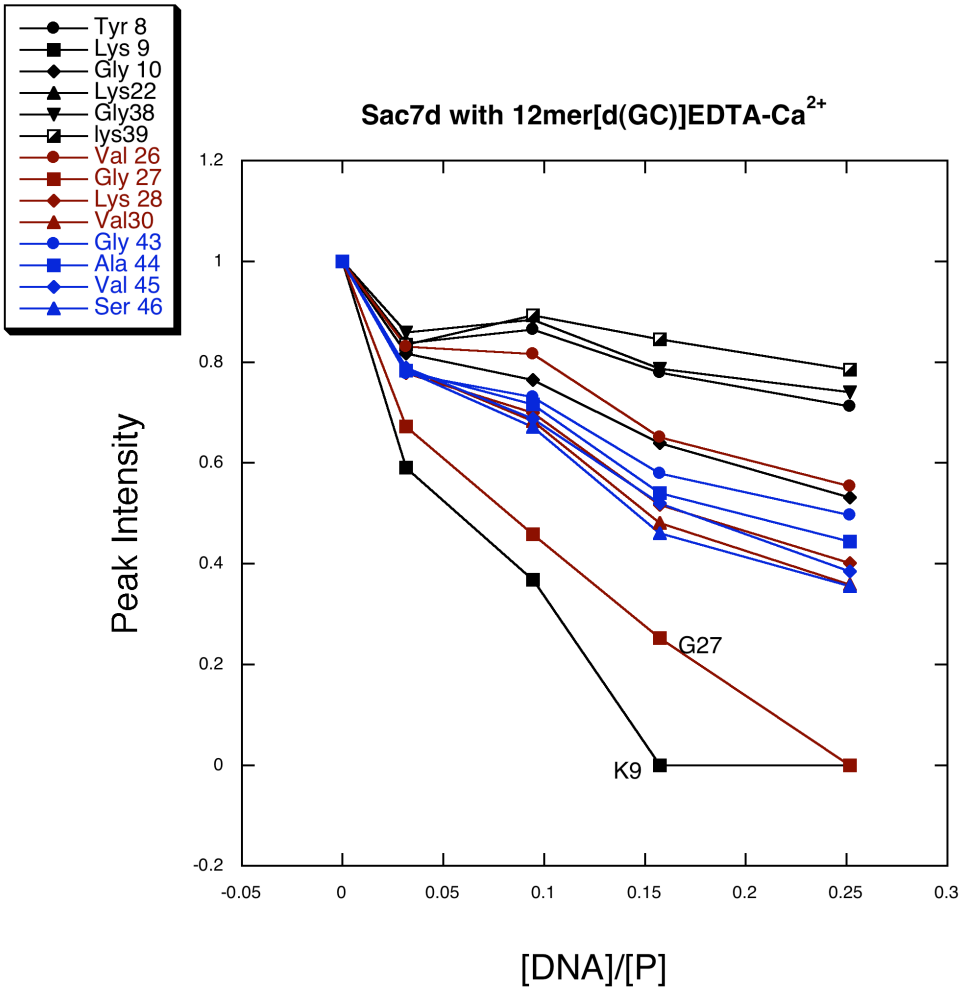


Figure 5.8 The decrease in intensity for the peaks in the head (blue and red curves) and the tail (black curves) regions of Sac7d with the addition of [d(GC)]EDTA-Ca2+. The concentration of DNA is expressed in mM nucleotides. The intensity of the peaks is plotted on the Y-axis and the intensity values are normalized to start all the curves at 1.0

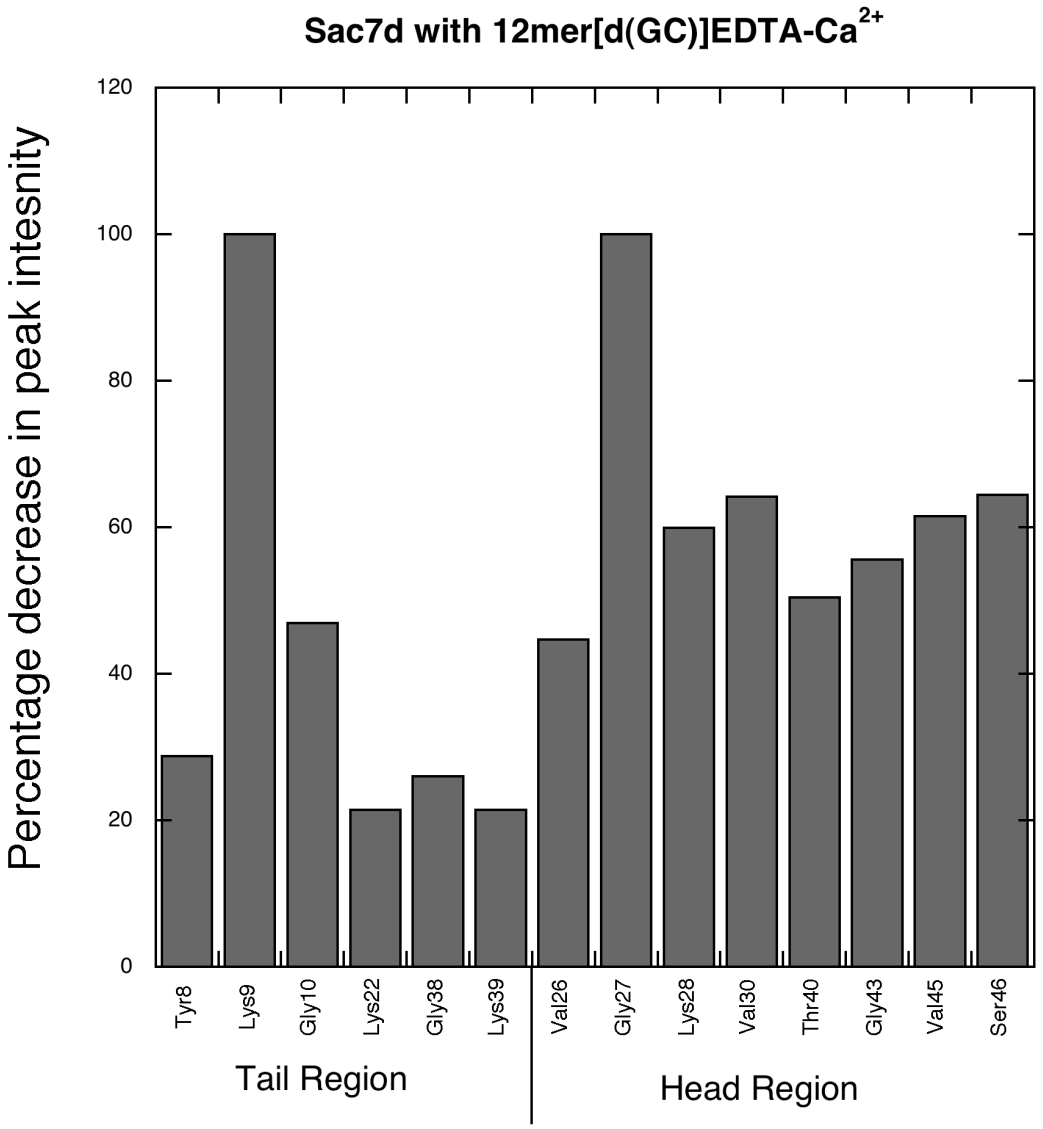


Figure 5.9 A Plot of percentage decrease in the peak intensity for the residues in head and tail regions with the addition of [d(GC)]EDTA-Ca^{2+}. The ratio of concentration of DNA to protein is ~0.3 (mM nucleotides to mM Protein).

The binding of VMSac7d to DNA results in a lesser degree of bending (~50°) due to shorter intercalating side chains (Chen, Ko et al. 2005). Also, forward titration of VMSac7d into poly [d(GC)] followed by monitoring the change in circular dichroism of the DNA does show structural change in DNA (Peters, Edmondson et al. 2005). Hence,

comparison of decrease in peak intensity profiles with the addition of [d(GC)]EDTA-Mn^{2+} to Sac7d and to VMSac7d in separate titrations should indicate the influence of protein induced DNA distortion on preferential binding orientation for the adjacent proteins. Figure 5.10 displays the comparison of the decrease in peak intensities for Sac7d and VMSac7d with the addition of [d(GC)]EDTA-Mn^{2+}. The model in figure 5.11 shows the specific residues that broadened out for Sac7d and VMSac7d (blue-for both Sac7d and VMSac7d, red-only for VMSac7d). A clear difference in the intensity drop profile is observed between Sac7d and VMSac7d. The intensity of K39, T40, G41 in β5, G10 in the turn between β1 and β2 and K22, V23 in the beginning of β3 (later part of β3 is regarded as head region) decreased in case of VMSac7d binding to [d(GC)]EDTA-Mn^{2+}which is contrary to Sac7d-[d(GC)]EDTA-Mn^{2+} binding. In addition to the residues listed above the intensities of the peaks corresponding to residues in the head region also decreased which is comparable to Sac7d-[d(GC)]EDTA-Mn^{2+} binding. Hence, the absence of DNA bending due to the introduction of V26A and M29A double mutation eliminated the preferential binding orientation between the adjacent Sac7d molecules that was otherwise present.

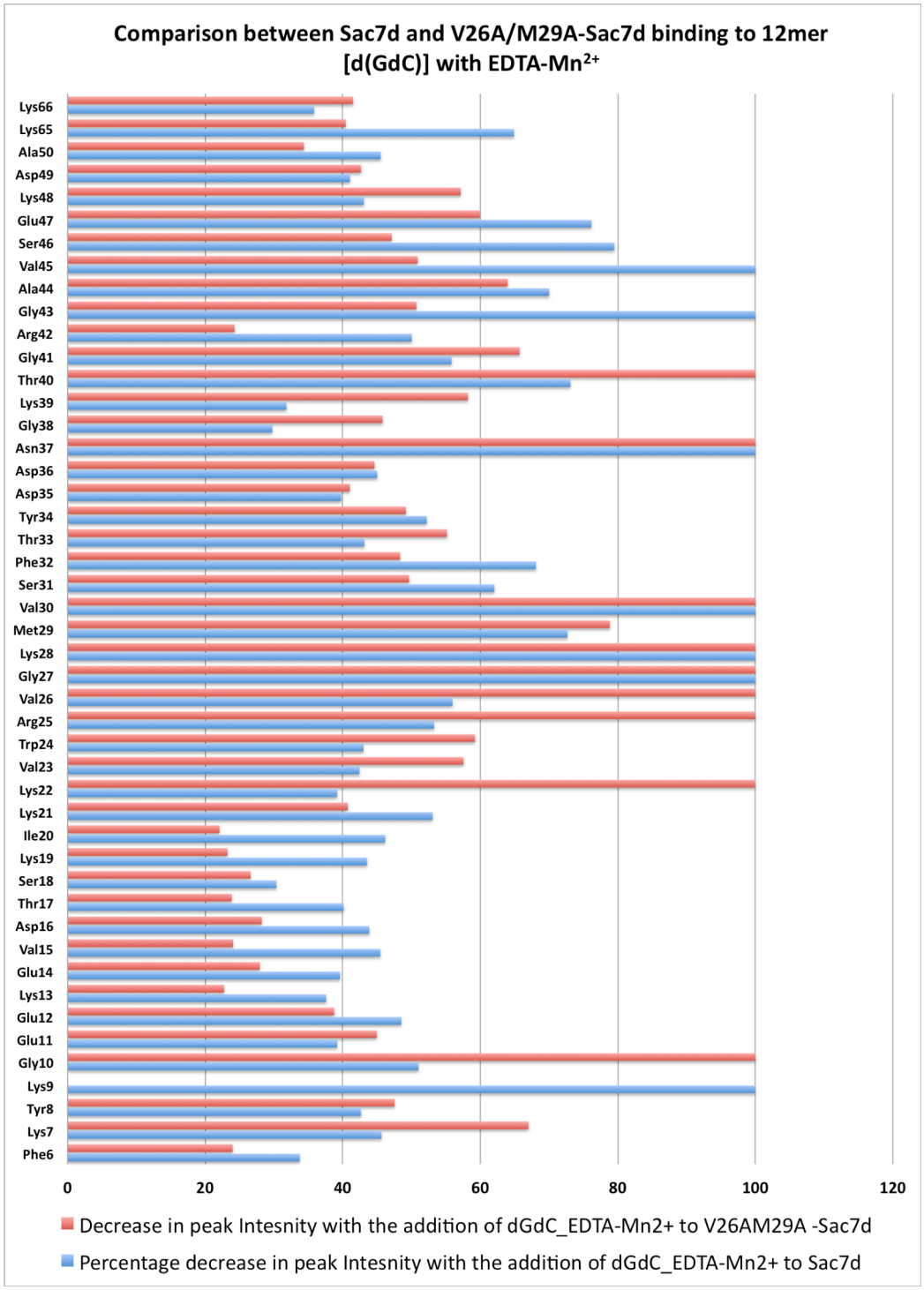


Figure 5.10 Comparison of percentage decrease in peak intensity for Sac7d and VMSac7d with addition of [d(GC)]EDTA-Mn^{2+}. The ratio of concentration of DNA to protein is 0.3.

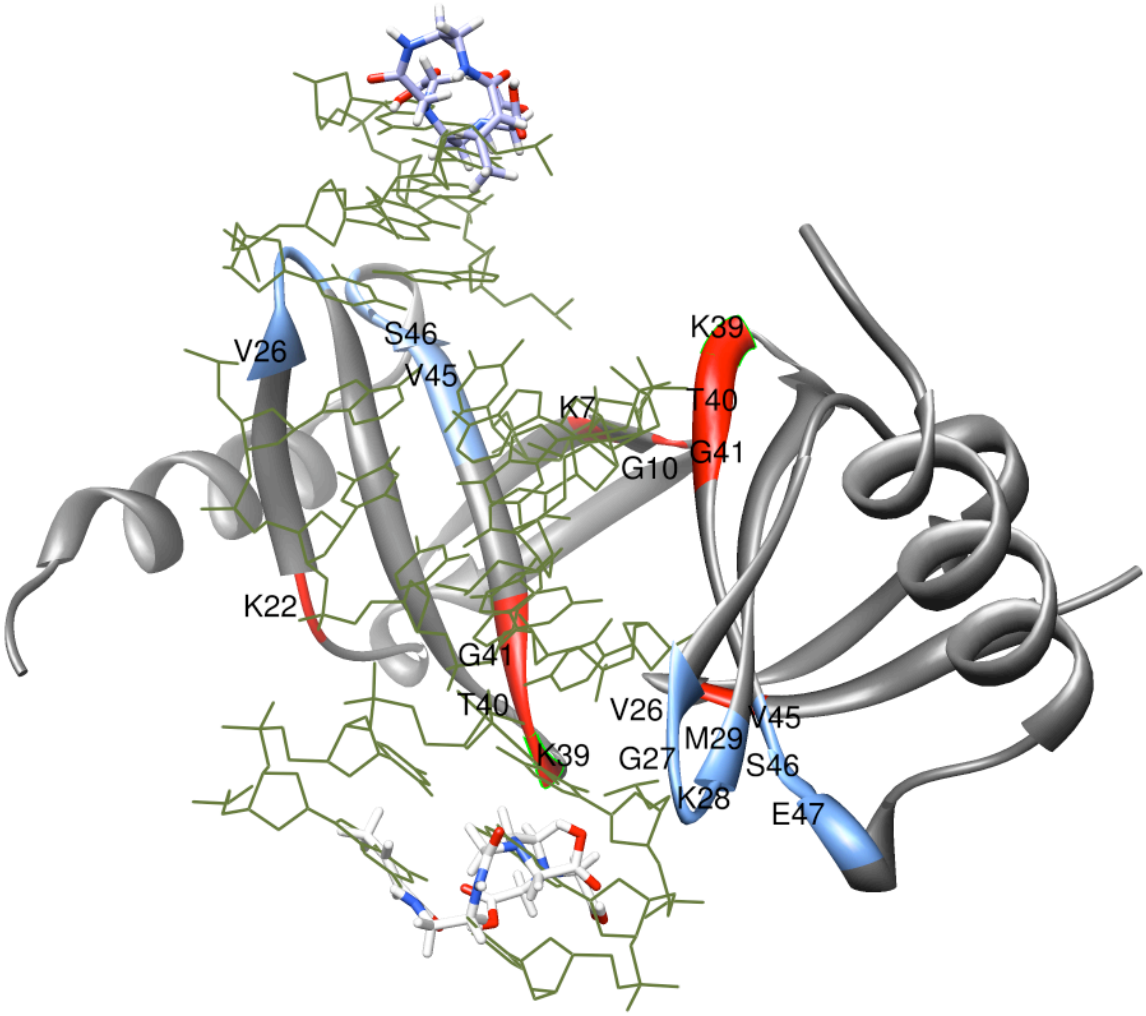


Figure 5.11 Model showing the residues that broaden out with the addition of [d(GC)] EDTA-Mn^{2+} to VMSac7d. Two Sac7d molecules (grey ribbon) bound in tail-to-tail orientation on a 11mer [d(GC)] (green wire). The EDTA modification (multicolor wire) is shown at the two ends of the DNA. Residues labeled in blue represent the residues that broaden out both in case of Sac7d and VMSac7d. Residues labeled in red represent those that specifically broaden out with VMSac7d only.

Comparison of the PRE data from Sac7d-[d(GC)]EDTA-Mn^{2+} titrations, Sac7d-[d(GC)]EDTA-Ca^{2+} titrations and VMSac7d-[d(GC)]EDTA-Mn^{2+}titrations indicates a clear tail-to-tail preference in the binding of adjacent Sac7d molecules on the DNA. Also,

there is a clear indication for the influence of protein binding induced DNA bending on the binding orientation.

The tail-to-tail preference in the binding orientation obtained does not agree with the NMR solution structure of Sso7d (1bbx) determined by Agback *et.al.,* that shows two Sso7d molecules bound in a preferential head-to-head orientation to a 12 mer GC rich sequence (Agback, Baumann et al. 1998). PRE analysis was conducted with 1bbx sequence to test the preference in binding orientation for 1bbx. Figure 5.12 shows the decrease in the intensity of the peaks for the residues in the head region with the addition of 1bbx-EDTA-Mn^{2+}. Figure 5.13 shows the percentage decrease in the peak intensity with the addition of 1bbx-EDTA-Mn^{2+}. The degree of decrease in the intensities for head region residues was more than those in the tail region. The intensity decrease profile for Sac7d-[d(GC)]EDTA-Mn^{2+} and that for Sac7d-1bbxEDTA-Mn^{2+} were comparable indicating a preference for tail-to-tail orientation for adjacent Sac7d on 1bbx sequence but not for head-to-head orientation as previously reported by Agback *et.al.,* (Agback, Baumann et al. 1998). Figure 5.14 displays the comparison of decrease in the intensities of amino acids for PRE titrations of Sac7d-[d(GC)]EDTA-Mn^{2+} and Sac7d-1bbxEDTA-Mn^{2+}. The percentage decrease in the intensity values for the two experiments were obtained at a DNA to protein concentration of ~0.3. The intensity decrease profile for 1bbx-Sac7d binding matches with that of Sac7d-[d(GC)] binding with a few exceptions. The degree of decrease in intensity for R25 is greater for 1bbx-Sac7d than for Sac7d-[d(GC)] binding. Conversely, the degree of decrease in intensity for V30 is greater for 1bbx-Sac7d binding than for Sac7d-[d(GC)] binding. N37 also does not broaden out in case 1bbx-Sac7d binding. A search for any possible evidence for a preferential head-to-

head orientation on 1bbx indicated no evidence. No evidence of relative decrease in intensity for the residues in the tail region (which is required for head-to-head orientation) was found in Sac7d-1bbx binding. The differences in intensity changes of R25 and N37 for 1bbx and [d(GC)] indicate a probable difference between Sac7d-[d(GC)] and Sac7d-1bbx binding.

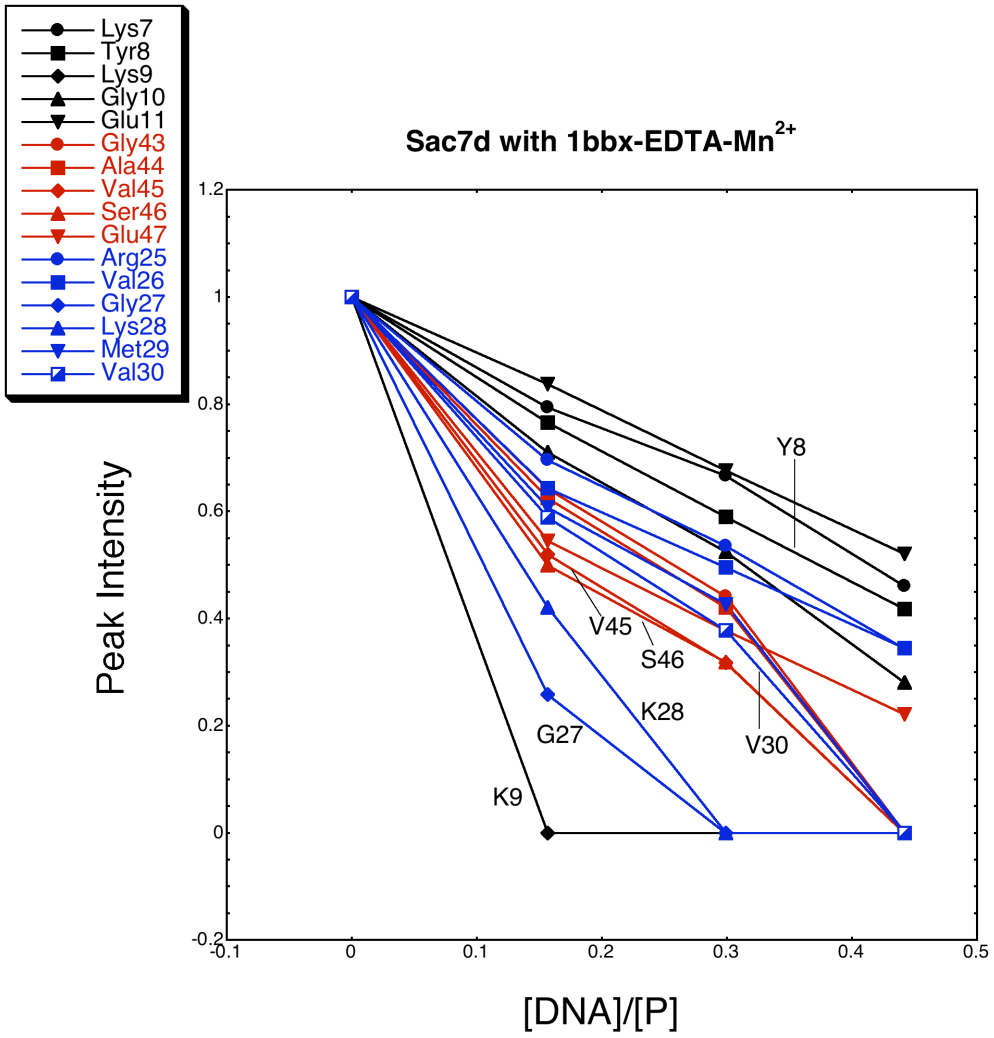


Figure 5.12 The decrease in intensity for the peaks in the head (blue and red curves) and the tail (black curves) regions of Sac7d with the addition of 1bbx EDTA-Mn2+. The concentration of DNA is expressed in mM nucleotides.

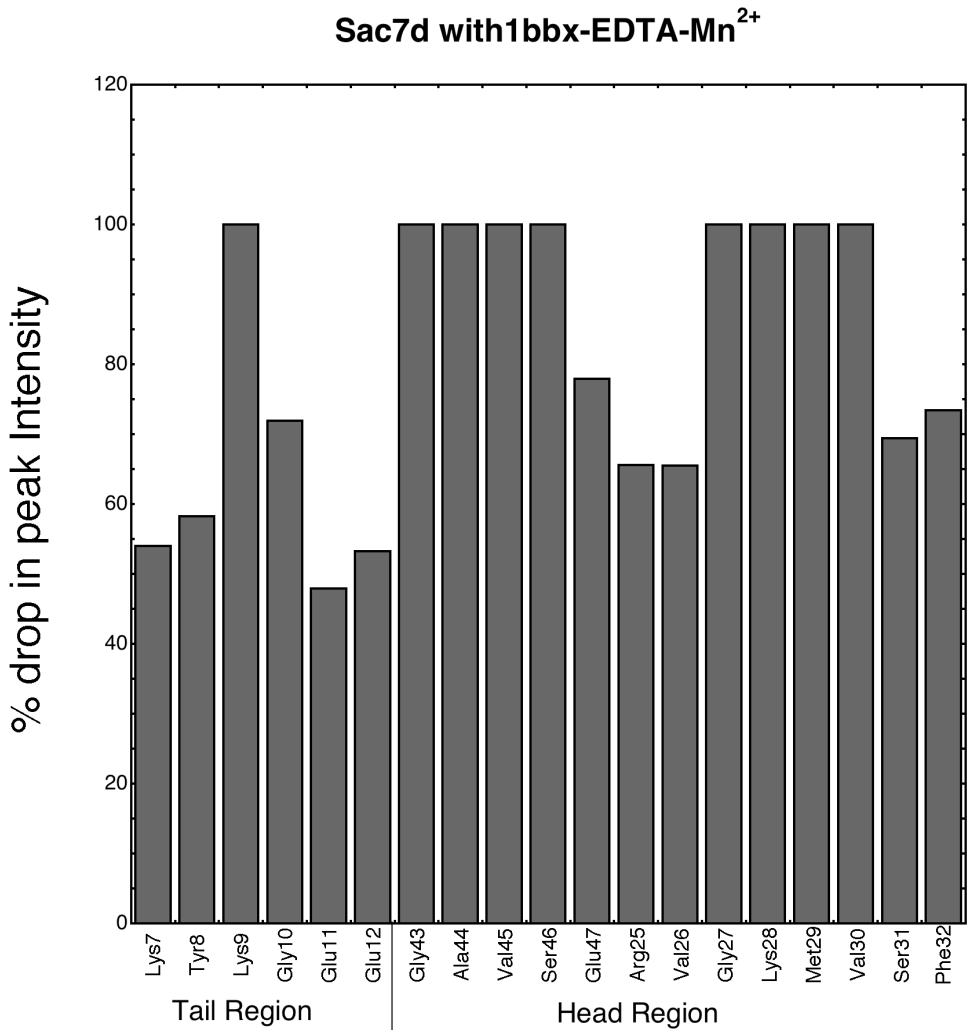


Figure 5.13 A Plot of percentage decrease in the peak intensity for the residues in head and tail regions with the addition of [d(GC)]EDTA-Ca^{2+}. The ratio of concentration of DNA to protein is ~0.4 (mM nucleotides to mM Protein).

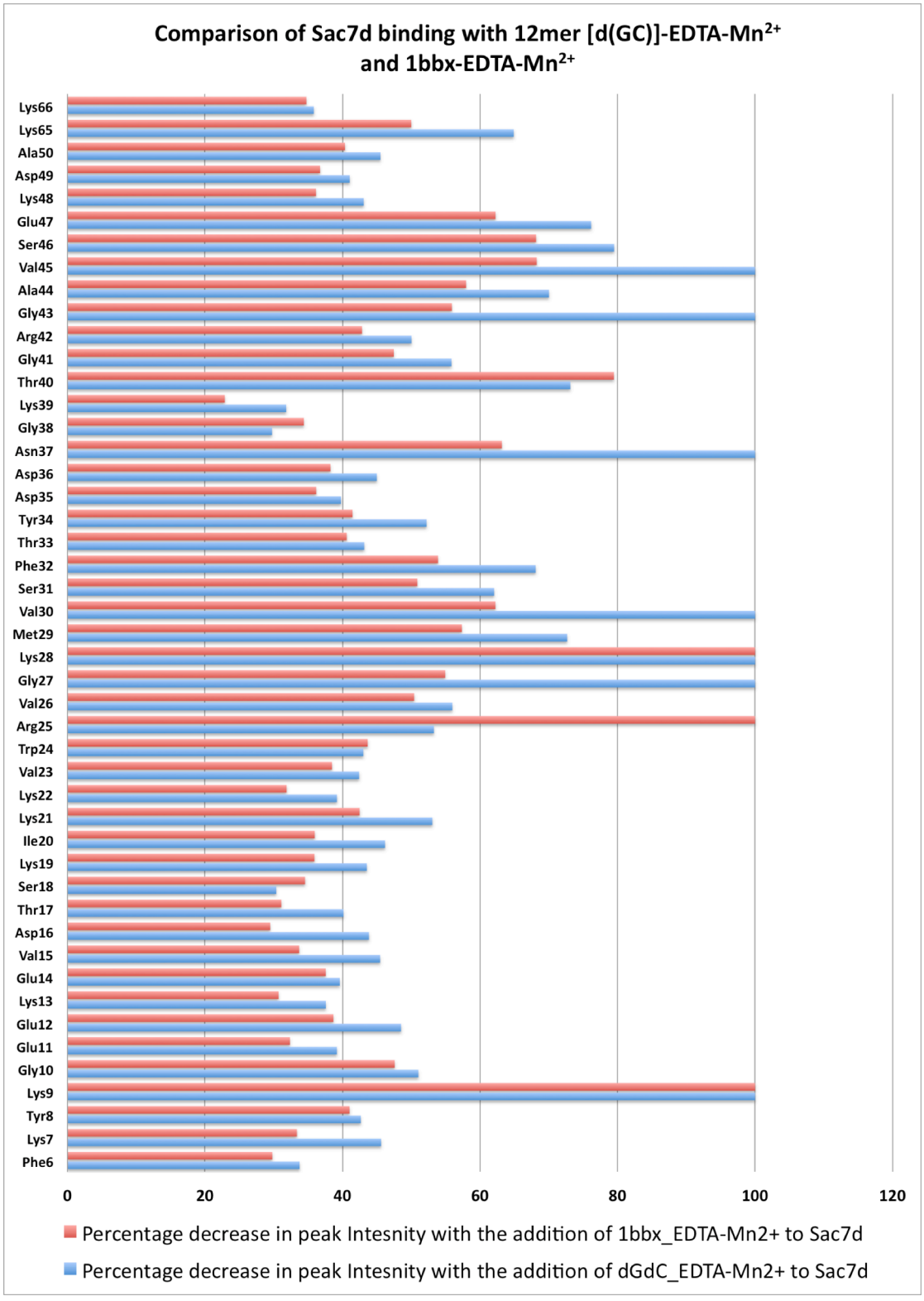


Figure 5.14 Comparison of percentage decrease in peak intensity for Sac7d and VMSac7d with addition of [d(GC)]EDTA-Mn2+. The ratio of concentration of DNA to protein is ~0.3.

5.5 Conclusions

PRE analysis facilitated the investigation of preferential binding orientation for adjacent Sac7d molecules on DNA by allowing unconstrained Sac7d-DNA binding. The comparison of the decrease in the intensities for the amino acid peaks in ^{15}N-HSQC of Sac7d obtained with the addition of increments of DNA with EDTA-Mn^{2+} indicated a clear preference for tail-to-tail orientation of adjacent proteins on DNA. The decrease in intensity due to paramagnetic relaxation was confirmed by conducting control experiments using diamagnetic Ca^{2+} labeled DNA. Comparison of PRE data for Sac7d and VMSac7d with [d(GC)]EDTA-Mn^{2+} provided indication for influence of protein binding induced DNA bending on binding orientation of adjacent Sac7d molecules on DNA. PRE analysis with 1bbx sequence using ^{15}N-labled Sac7d showed no indication of preferential head-to-head orientation for adjacent Sac7d on 1bbx as indicated by the NMR solution structure for Sso7d as determined by Agback *et.al.,* (Agback, Baumann et al. 1998), instead a preference for tail-to-tail orientation was observed.

Due to the high intensity of the paramagnetic signal, the amount of DNA with paramagnetic label used for the titrations was very low. At DNA to protein concentration ratios exceeding ~0.45, it was difficult to observe a clear distinction between the changes in the intensities for different regions of the protein. A part of it could also be attributed to the fast exchange associated with the Sac7d-DNA binding. For example, the intensity of K9, G27 and N37 decreased with [d(GC)]EDTA-Mn^{2+}–^{15}N-Sac7d, VMSac7d–[d(GC)]EDTA-Mn^{2+}bindng and [d(GC)]EDTA-Ca^{2+}–^{15}N-Sac7d, indicating that the decrease in intensity may be attributed to the fast exchange associated with Sac7d–DNA binding.

CHAPTER 6

ANALYSIS OF SMALL ANGLE X-RAY SCATTERING OF Sac7d-DNA COMPLEX

6.1 Introduction

DNA binding analysis of Sac7d dimers and the paramagnetic relaxation enhancement (PRE) assay indicate a preference in tail-to-tail orientation for adjacent Sac7d on DNA. A preferential tail-to-tail orientation implies that Sac7d molecules on elongated DNA can bind in a sequential tail-to-tail, head-to-head fashion. Previous study of small angle X-ray scattering analysis (SAXS) for Sac7d-32-mer [d(GC)] complex considered only a sequential head-to-tail orientations for adjacent Sac7d molecules (Krueger, McCrary et al. 1999). SAXS is sensitive to the size and shape of macromolecules in solution and is useful for obtaining information about the conformations of higher order structures of protein and DNA complexes (Bak, Hanna et al. 1995; Olah, Gray et al. 1995). This chapter describes the attempt to fit the previous SAXS data for Sac7d-32 mer [d(GC)] complex according to the preferential tail-to-tail orientation suggested by the DNA binding analysis of Sac7d dimers and PRE analysis.

In the present study scattering profiles calculated for Sac7d-DNA model complexes constructed using Chimera a molecular modeling software (Pettersen, Goddard et al. 2004) have been compared with the experimental scattering data for Sac7d-32 mer [d(GC)] complex collected previously (Krueger, McCrary et al. 1999). The model protein-DNA complexes have been constructed by placing adjacent Sac7d molecules on a 32 mer [d(GC)] in sequential head-to-head, tail-to-tail fashion. The sequential tail-to-tail and head-to-head placement of adjacent proteins was based on the evidence for preferential tail-to-tail orientation for adjacent Sac7d molecules obtained from DNA binding analysis of Sac7d dimer and PRE analysis.

6.2 Model building

The DNA-protein complex models were constructed in Chimera (Pettersen, Goddard et al. 2004; Goddard, Huang et al. 2005; Couch, Hendrix et al. 2006; Meng, Pettersen et al. 2006; Goddard, Huang et al. 2007; Morris, Huang et al. 2007; Pintilie, Zhang et al. 2010) using B-DNA segments and the coordinates of Sac7d-DNA crystal structure (1azp). The kinks in the DNA were also modeled according to the 1azp crystal structure (Robinson, Gao et al. 1998). Eight Sac7d molecules were placed on 32-mer [d(GC)] to create seven kinks. The 32-mer [d(GC)] was saturated by placing adjacent Sac7d molecules in sequential tail-to-tail, head-to-head orientation and by varying the distance between the successive kinks. Three different models of Sac7d-32-mer [d(GC)] were created with adjacent Sac7d molecules placed in such a way that the distances between the successive kinks is different in each model. The reason for considering three different models with different distances between the kinks is to avoid stearic clashes

between the adjacent proteins that might otherwise be in place when the adjacent proteins are placed with the kinks every four base pairs.

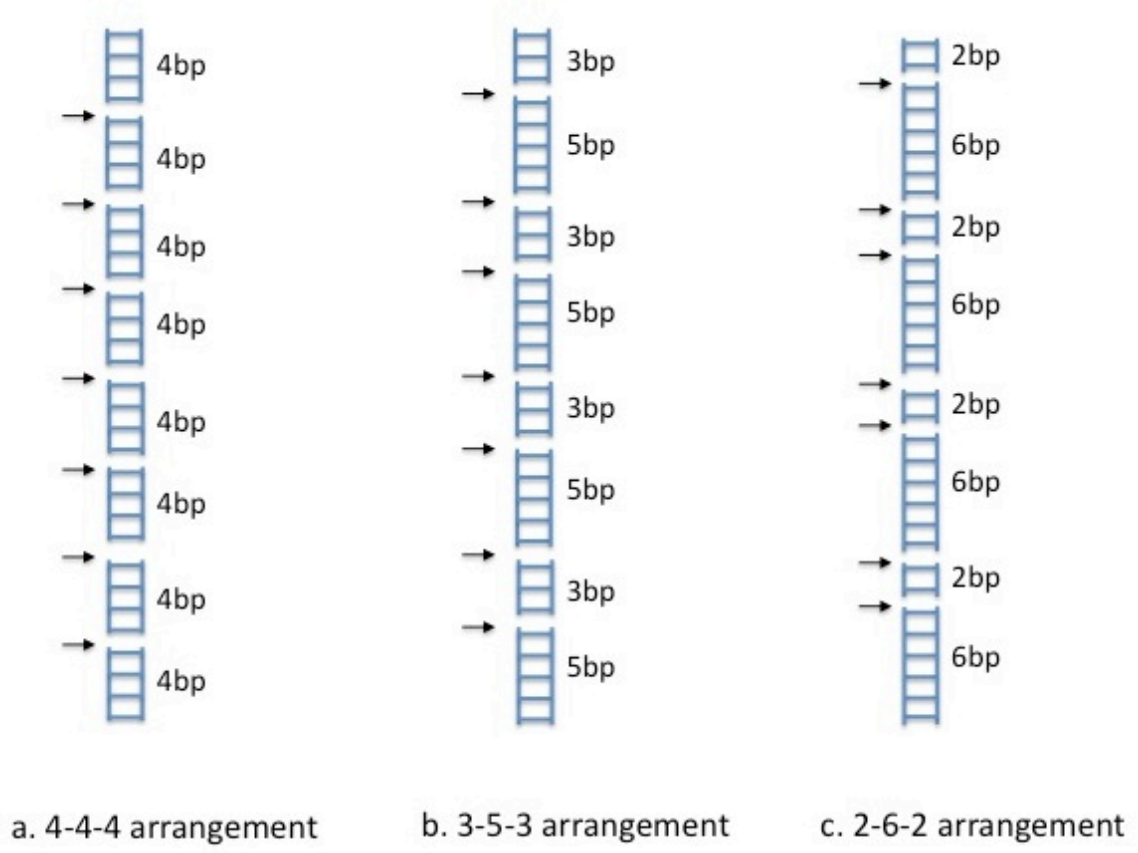


Figure 6.1 Three different Sac7d-32-mer [d(GC)] complexes created for analyzing the SAXS data considering sequential tail-to-tail, head-to-head orientation for adjacent Sac7d on DNA. The kinks and the number of base pairs separating them are pointed out.

The different models under study differ in the spacing between successive kinks. Considering the binding site of four base pairs, two Sac7d molecules can bind every eight base pairs and the kinks created by them can be separated by four and four base pairs (4+4=8) or three and five base pairs (3+5=8) or two and six base pairs (2+6=8). Therefore on a 32-mer [d(GC)], eight Sac7d molecules can bind creating seven kinks. The binding can be in such away that the kinks may be 4-4-4- base pairs apart or 3-5-3- base pairs apart or 2-6-2- base pairs apart. These three possible arrangements were

considered to create three model complexes of eight Sac7d molecules on a 32-mer [d(GC)] and they were named based on the distance between the successive kinks as 444 arrangement, 353 arrangement and 262 arrangement. (A possible combination of these arrangements such as 4-4-3-5 or 4-4-2-6 etc., were not considered for simplicity). Figure 6.1 displays the three different types of complexes described above.

Figures 6.2 through 6.4 depict the building units for the three models described above. These models display the arrangement of protein molecules and the kinks using B-DNA segments. These building units were used to construct Sac7d-32-mer B-DNA complexes for which the scattering profiles were calculated. The building units are being shown before the 32-mer complex for clarity. The kink in the models is built using the constraints from the crystal structure of Sac7d-DNA complex (Robinson, Gao et al. 1998). Phosphodiester bonds were not modeled between the DNA segments to distinguish the regions and hence the gaps seen in the duplex do not represent breaks in the DNA.

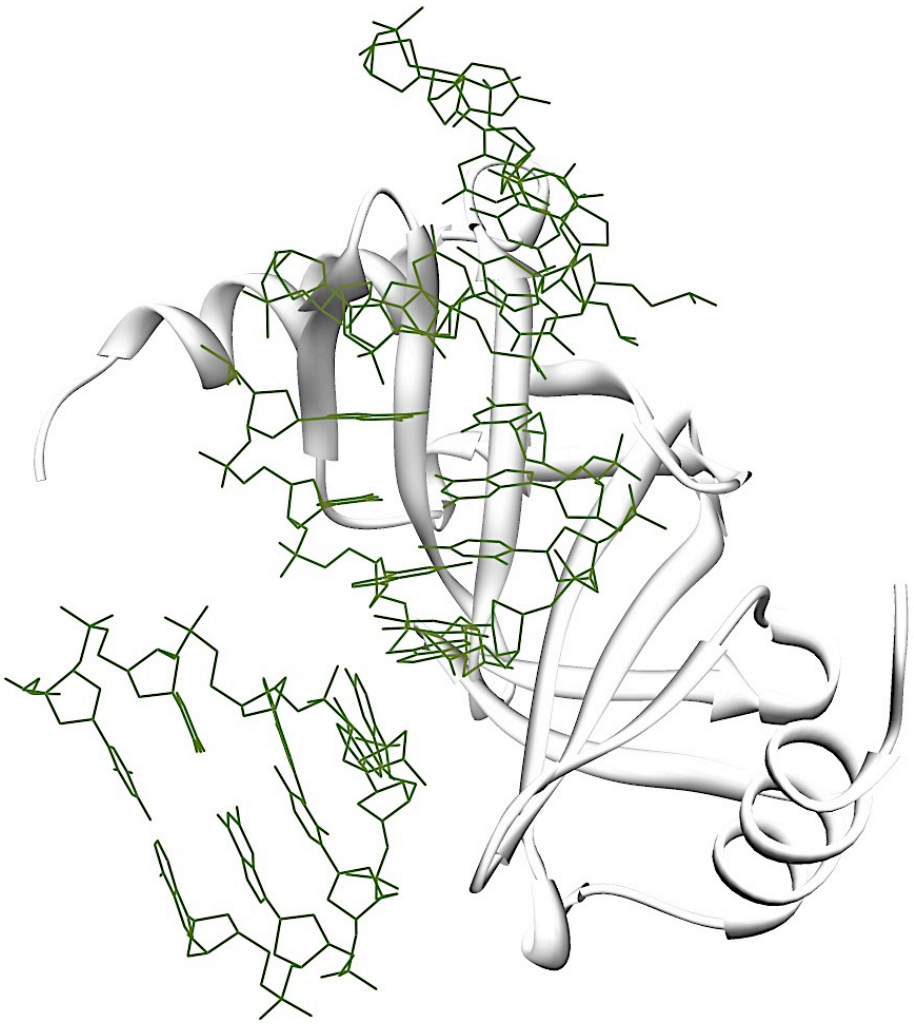


Figure 6.2 4-4-4 building unit shows two Sac7d molecules bound in tail-to-tail orientation. The two kinks are modeled between three 4 base pair B-DNA segments. The building unit being defined here consists of two 4bp B-DNA segments and two Sac7d molecules. The third B-DNA segment shown is a part of the next 4-4- unit.

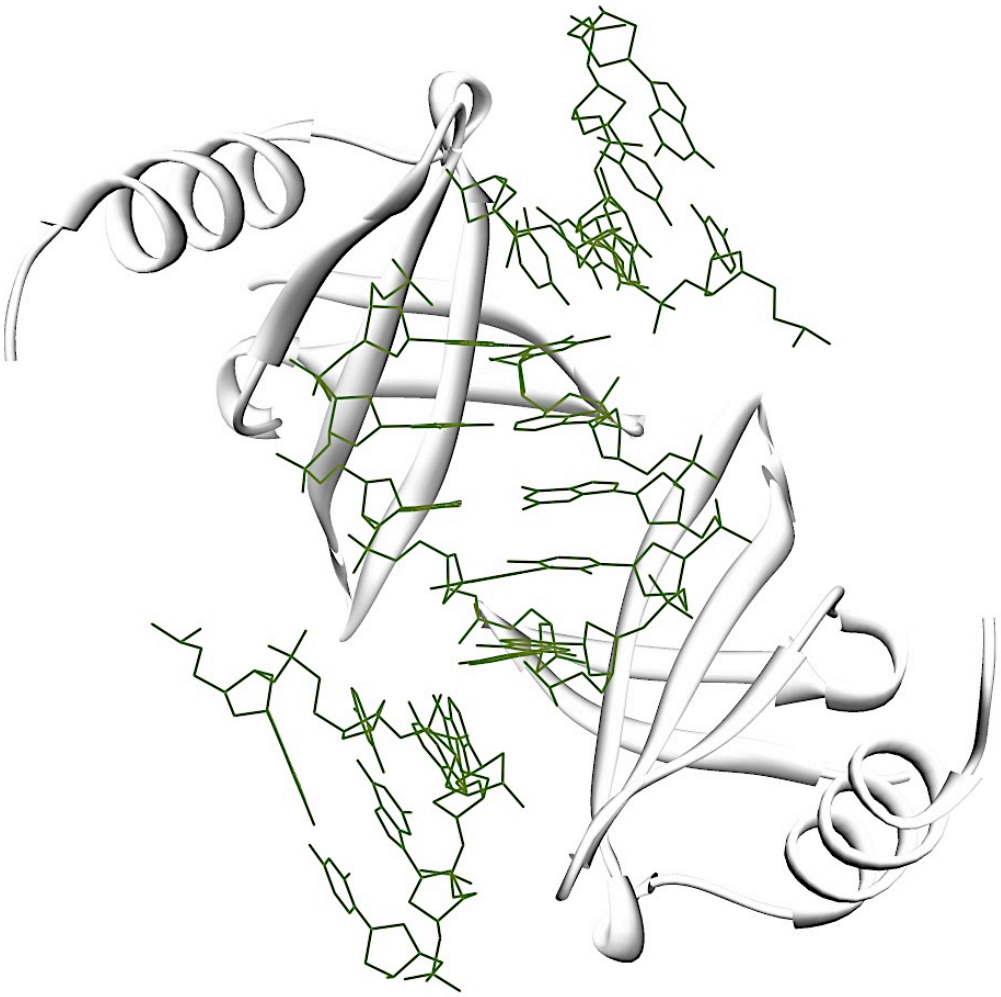


Figure 6.3 3-5-3 unit shows two Sac7d molecules bound in tail-to-tail orientation. The two kinks are modeled between 3bp, 5bp and 3bp B-DNA segments. The 3-5-3 binding unit consists of a 3bp and a 5bp segment with two Sac7d molecules bound in tail-to-tail orientation.

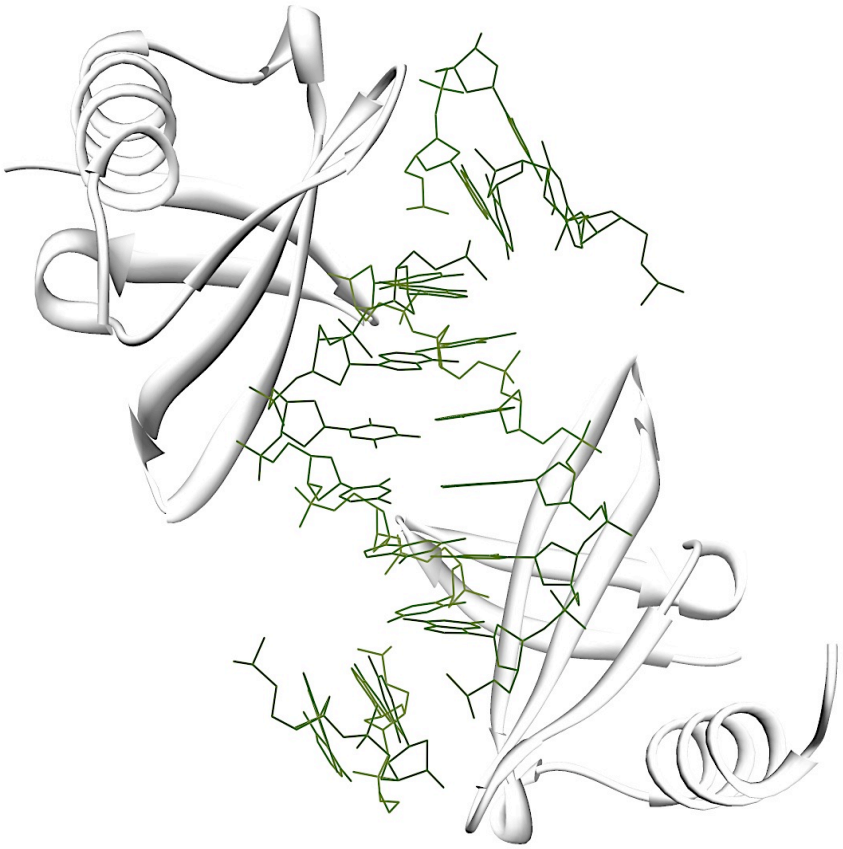


Figure 6.4 2-6-2 unit shows two Sac7d molecules bound in tail-to-tail orientation. The two kinks are modeled between two base pair, six base pair and two base pair B-DNA segments. The 2-6-2 binding unit consists of a 2bp and a 6bp segment with two Sac7d molecules bound in tail-to-tail orientation.

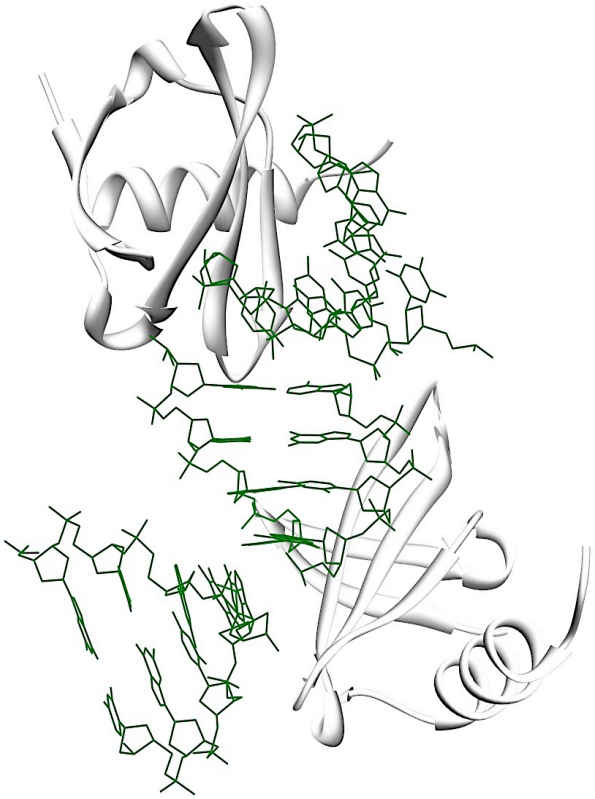


Figure 6.5 4-4-4 in head-to-tail orientation shows two Sac7d molecules bound in head-to-tail orientation. The two kinks are modeled between three 4 base pair B-DNA segments.

In all the above models, Sac7d molecules are bound to the minor groove of the DNA with a sequential tail-to tail, head-to-head orientation except Figure 6.5 in which the orientation is sequential head-to-tail orientation. Figure 6.5 represents the model previously used for Sac7d-DNA SAXS analysis that considered only a sequential head-to-tail packaging for Sac7d on DNA (Krueger, McCrary et al. 1999). In the present study, along with the models with sequential tail-to-tail, head-to-head orientation the previous model was also studied for comparison. Figures 6.6 to 6.9 represent the Sac7d-32 base pair [d(GC)] B-DNA complexes in arrangements described above. The following

complexes were obtained by placing the building units described above to create a 32-mer B-DNA.

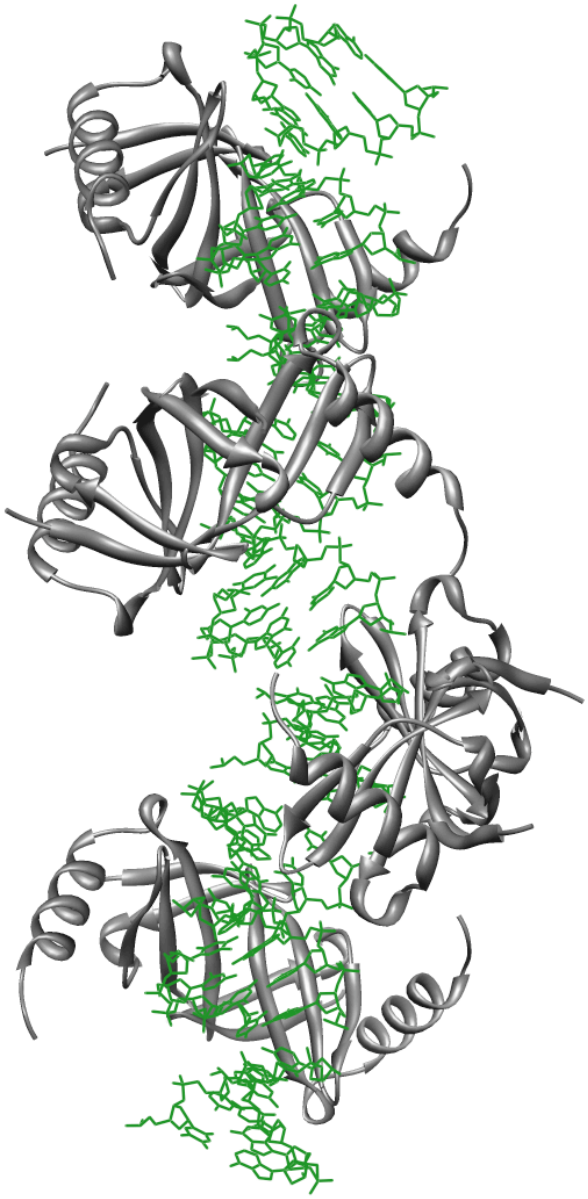


Figure 6.6 4-4-4 complex with adjacent Sac7d molecules placed on a 32-mer [d(GC)] B-DNA in sequential tail-to-tail, head-to-head orientation with a separation of 4 base pairs between successive kinks.

Figure 6.7 3-5-3 complex with adjacent Sac7d molecules placed on a 32-mer [d(GC)] B-DNA in sequential tail-to-tail, head-to-head orientation with a separation of three, five and three base pairs between successive kinks.

Figure 6.8 2-6-2 complex with adjacent Sac7d molecules placed on a 32-mer [d(GC)] B-DNA in sequential tail-to-tail, head-to-head orientation with a separation of two, six and two base pairs between successive kinks.

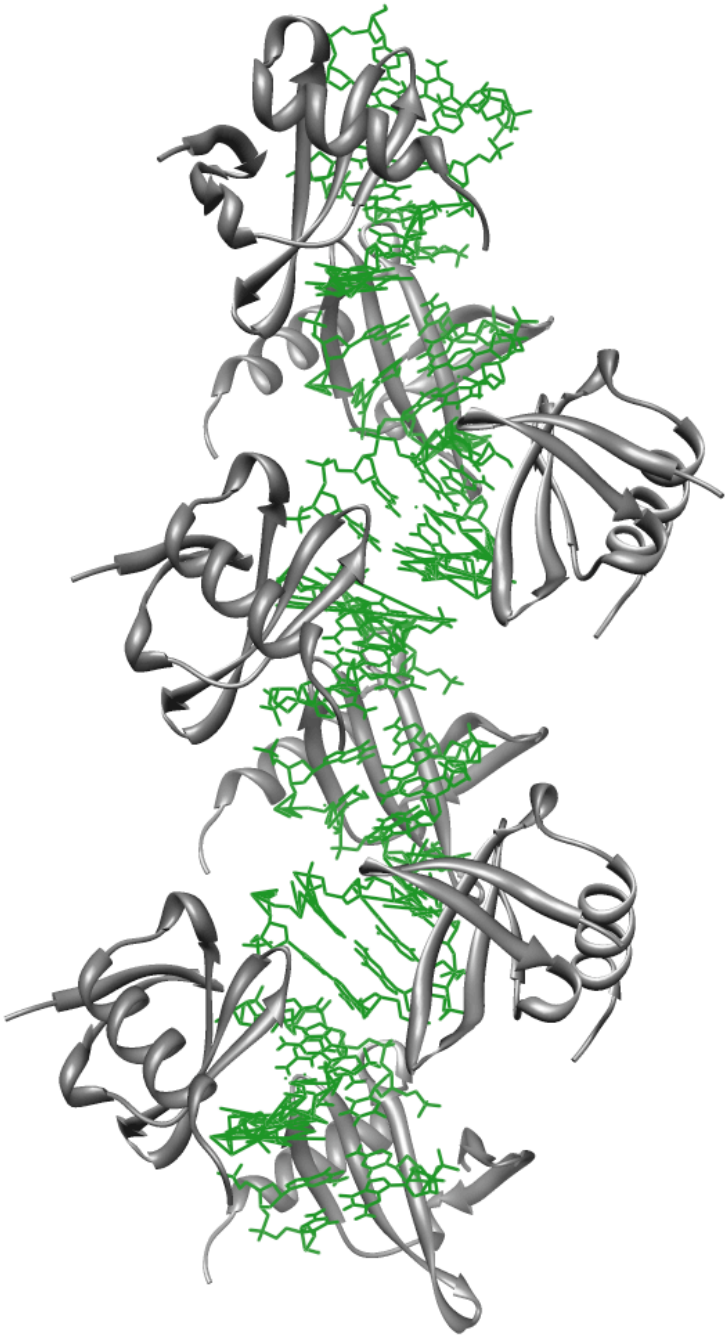


Figure 6.9 4-4-4 head-to-tail complex with adjacent Sac7d molecules placed on a 32-mer [d(GC)] B-DNA in sequential head-to-tail orientation with a separation of 4 base pairs between successive kinks.

6.3 Methodology for the calculation and comparison of scattering profiles

The SAXS data for Sac7d-32-mer [d(GC)] complex were collected at Los Alamos National Laboratory by Krueger *et.al.,* (Krueger, McCrary et al. 1999). Scattering data were reduced to *I(Q)* versus *Q* and analyzed. *I(Q)* is the scattering X-ray intensity per unit solid angle. *Q* is the amplitude of the scattering vector and it is given by $4\pi(\sin\theta)/\lambda$, where 2θ is the scattering angle and λ is the wavelength of the scattered X-rays (1.54 Å). Scattering data were collected for *Q* values from 0.08 to 0.25 $Å^{-1}$. Scattering data were collected at several concentrations of the complex ranging from 5 to 20mg/ml.

Scattering data were corrected for the slit geometry of the scattering instrument that has the effect of smearing the experimental scattering data. The GNOM analysis of the Svergun and colleagues was used to calculate the *P(r)* functions for the complexes from the scattering data (Serdyuk, Tsalkova et al. 1987). *P(r)* is the frequency of the vector length connecting small-volume elements within the entire volume of the scattering particle, and goes to zero at the maximum dimension of the particle, d_{max}. The radius of gyration of cross section, *Rc,* was calculated using the Guinier approximations (Bevington and Robinson 2003) to the lowest *Q* data for each complex. *Rc* is the weighted, root-mean-square distance of all the elemental areas from the center of the cross-sectional area of the rod-shaped particle.

An all-atom procedure was use to calculate *P(r)* function for the models of the protein-DNA complexes. Points were randomly selected within the Van der Waals volume of pairs of the randomly selected atoms and the *P(r)* distribution function was calculated by summing the distances between the selected points weighted by their

atomic scattering power. Intensity curves, *I(Q),* were calculated from the model *P(r)* functions using the relation:

$$I(Q) = 4\pi \int P(r) \sin(Q*r)/(Q*r)\, dr$$

The programs used for the calculations are presented in appendix C.2. To compare the scattering profiles calculated from the models with the experimental data, the model data was convoluted with the instrumental slit smearing function (scripts in appendix C.2). The experimental scattering data were scaled to the scattering profiles calculated for each model over a Q range of 0.01-0.2 $Å^{-1}$ using least-square linear program based on singular value decomposition. The fit between model scattering profiles and the experimental data was evaluated using reduced chi-square (χ^2) values.

6.4 Results and discussion

Figure 6.10 displays the comparison of the experimental and model scattering profiles. The data for three types of models described above are compared with the experimental data available for Sac7d-32mr [d(GC)] DNA. A reasonable fit (χ^2=1.48) was obtained for the scattering data of 353-complex and the experimental data. Plot of the experimental data and the scattering profile for 353-complex also indicate good fit. The overall errors in the fits to the 444-complex and 262-complex may be attributed to not optimizing the geometric parameters involved in the kink (bend angle, rotation angle and horizontal displacements). Optimization of these geometric parameters may further improve the quality of fit for 353-complex.

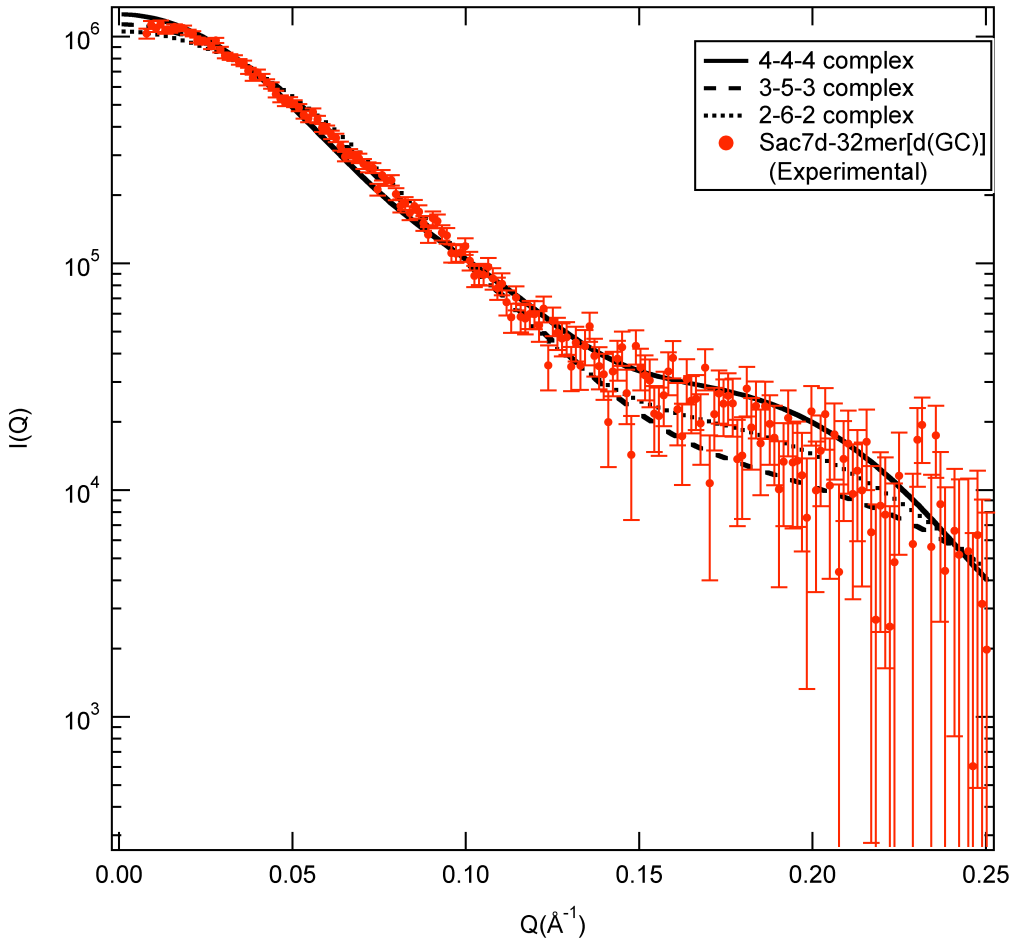


Figure 6.10 Comparison of the scattering profiles calculated for the 4-4-4, 3-5-3 and 2-6-2 models with the experimental data for Sac7d-32-mer [d(GC)]. The experimental data was scaled to match the calculated curves for the models. The black curves (bold=4-4-4, dashed=3-5-3 and dotted=2-6-2) represent the calculated scattering profiles for the constructed models and red dots represent the experimental data.

Model	Rc (Å)	Rg (Å)	D(Å)	d_{max}(Å)
444 head-to-tail	17.347	34.41	49.06472533	120
444	17.715	34.34	50.10558651	120
353	16.612	30.38	46.985 8314	120
262	26.157	28.14	73.9831683	120
Sac7d/32-mer				
0:32	7.4±0.1	31.0±0.3	20.9±0.3	100
1:32	8.7±0.1	31.0±0.3	24.6±0.3	100
7:32	16.1±0.2	34.3±0.2	45.5±0.5	120
Sac7d/poly[d(GC)]				
1:4	17.6±0.2	129	49.8±0.6	450
Sac7d/*E.coli* DNA				
1:4	17.9±0.5	131	50.6±1.4	450

Table 6.1 Structural parameters for experimental scattering data and the model data obtained by computing the parameters based on an all atom weighted approach. Rc was determined by linear regression analysis of the I(Q).Q analysis of the scattering data after correction for the slit beam geometry of the scattering instrument. The diameters were calculated from the relation D= 2Rc√2.

Table 6.1 lists the structural parameters for each complex obtained from the pdb files of the complexes using in-house scripts that follow an all atom weighted approach. The script used for computing the structural parameters from pdb files is show in appendix C.2. Comparison of the structural parameters of the three different models under study with the experimental parameters indicates that the structural parameters for the 353-complex with the adjacent Sac7d molecules in sequential tail-to-tail, head-to-head orientation are comparable to those of the experimental data of Sac7d-32-mer [d(GC)]. The experimental value of radius of gyration (R_g) 34.3±0.2 Å is comparable to 34.41Å for the complex in 444-orientation with the adjacent Sac7d molecules in sequential tail-to-tail, head-to-head orientation. The values of cross sectional radii of

gyration (R_c) and the diameter of the complex with 353-orientation compare closely with the experimental data. Based on the comparison of the structural parameters for the model complexes with the adjacent Sac7 molecules in head-to-tail orientation (previous SAXS data) and the current models under study with the adjacent Sac7d molecules in head-to-head, tail-to-tail orientation indicates that the parameters of the models with head-to-head/tail-to-tail compare well with the experimental data for Sac7d-32-mer [d(GC)].

6.5 Conclusion

X-ray scattering for three types of complexes of Sac7d-32-mer [d(GC)] with adjacent Sac7d in tail-to-tail, head-to-head orientation were calculated. The three complexes differ in the distance between the successive kinks. The calculated scattering data was compared with that of the experimental data collected previously for Sac7d-32-mer [d(GC)] complex to test the preference for tail-to-tail orientation obtained from the DNA binding aSac7d dimer and PRE analysis.

The scattering data and geometric parameters for the model complexes were computed using in-house script that follow an all atom weighted approach. The comparison of the scattering profiles for the model data with the experimental data indicated that the model in 353-orientation with adjacent Sac7d molecules in sequential tail-to-tail, head-to-head orientation was in good agreement with the experimental data. The over all quality of the fits for 444 complexes and 262 complexes was not good because the different geometrical parameters involved in the modeling the kink structure were not optimized. Therefore optimization of bend angle, rotation angle and horizontal

displacements may further improve the quality of fits. Comparison of geometric parameters of the model data with the experimental data also indicates that the models with adjacent Sac7d molecules in sequential in tail-to-tail, head-to-head orientation were comparable to those obtained for experimental data.

The results indicate that the experimental data for Sac7d 32-mer [d(GC)] can be explained considering a sequential tail-to-tail, head-to-head orientation for adjacent molecules which is in agreement with the preferential tail-to-tail orientation obtained from DNA binding analysis of Sac7d dimers and PRE analysis.

CHAPTER 7

DISCUSSION

7.1 Sac7d-DNA binding and DNA unwinding

Sac7d binds along the minor grove of DNA and bends DNA into the major grove. The bend is stabilized by the insertion of the side chains of V26 and M29 (Robinson, Gao et al. 1998; Peters, Edmondson et al. 2005). The crystal structure of Sac7d bound to an eight base pair duplex DNA (1azp) shows a sharp kink in the DNA between the second and the third base pair steps with a clear unwinding of the helix (Robinson, Gao et al. 1998). Sac7d binding induced DNA unwinding is clearly indicated by forward titrations (protein into DNA) followed by monitoring the change in circular dichroism of the DNA. With an increasing protein concentration, a cooperative unwinding of the DNA was observed (McAfee, Edmondson et al. 1996). With increase in protein concentration the CD spectrum of the DNA showed a gradual increase in intensity of the CD band at long wavelength (around 280 nm). An increase in the peak at 280 nm is characteristic of A-DNA and represents helix widening resulting from DNA unwinding. The DNA

unwinding with an increase in protein concentration was cooperative and the degree of cooperativity was sequence dependent(McAfee, Edmondson et al. 1996). Poly [d(IC)] showed the highest cooperativity followed by poly [d(GC)], while poly [d(AG)]-poly[d(CT)] and poly [d(AC)]-poly[d(GT)] showed negligible cooperativity. There was no evidence of cooperativity in forward titrations (DNA into protein) of poly [d(GC)] with Sac7d monitored by fluorescence (McAfee, Edmondson et al. 1996). Reverse titrations (DNA into protein) followed by monitoring the fluorescence quenching of W24 in Sac7d also did not indicate cooperative binding mode (Peters, Edmondson et al. 2005).

Cooperative DNA unwinding with increase in protein binding density implies that the unwinding imposed by the protein molecules operates through adjacent proteins. It can be expected new protein molecules bind in a set orientation in relation to their adjacent, already bound proteins so that the unwinding occurs in a synergistic manner. A recognition mechanism is required for the incoming proteins to bind in a set orientation adjacent to the already bound proteins so that the over all DNA unwinding is in phase. In-coming Sac7d molecules binding adjacent to the already bound Sac7d molecules encounter a DNA binding site that is different in structure than that encountered by the Sac7d molecules that are initially bound. The difference in the structure of the binding site arises from DNA distortion caused by the bound proteins. For the cooperative mode of unwinding to propagate, the incoming Sac7d molecules should indirectly recognize (structure based recognition) the modified DNA structure or the modified binding site and bind to it in a defined orientation. Hence, a potential preference in binding orientation for adjacent Sac7d molecules may arise from DNA unwinding in a cooperative manner.

In addition, the inherent sequence dependent structural features of the DNA may also impose potential preference in orientation for adjacent Sac7d on DNA.

The suspected correlation between DNA unwinding and preferred orientation for adjacent Sac7d molecules has been the focus of the present work. In order to investigate the potential preference in binding orientation for two adjacent Sac7d molecules on DNA, the binding of Sac7d to adjacent sites on DNA has been studied using three different approaches. First, DNA binding of Sac7d-dimers cross-linked in different possible binding orientations has been studied. Second, DNA binding of Sac7d molecules without cross-linking was studied by the paramagnetic relaxation (PRE) assay to allow the dynamic non-specific mode of Sac7d-DNA without cross-linking constraints. Third, as an intermediate between the cross-linked dimers and Sac7d molecules without cross-linking, Sac7d molecules bound to the DNA were cross-linked on DNA to identify the preference in binding orientation for adjacent proteins on DNA. Also, packaging multiple Sac7d molecules on elongated DNA has been studied by fitting the previous small angle X-ray scattering (SAXS) data (Krueger, McCrary et al. 1999) based on the data obtained from the three approaches stated above.

7.2 DNA binding of Sac7d dimers

7.2.1 Sac7d dimers

Sac7d dimers have been synthesized in three different orientations head-to-head, tail-to-tail and head-to-tail orientations. Native Sac7d sequence does not contain cysteines, so introduction of single cysteine mutations at desired positions allowed the

creation of convenient cross-linking sites. CD, DSC, NMR and DNA binding studies monitoring the fluorescence quenching of W24 in Sac7d indicated that the effect of cysteine mutations was negligible on protein structure and DNA binding.

Creating Sac7d dimers cross-linked in head-to-tail orientation was not straightforward. Introduction of a C-terminal His-tag to the tail cysteine mutant (K9C) facilitated the separation of head-to-tail dimer from the dimers with other orientations. The extra His-tag region did not interfere with the protein folding and DNA binding properties. The extra amino acid residues in the His-tag region and the spacer region between the His-tag and the protein sequence provided a convenient method for clearly distinguishing dimers with different cross-linking orientations using gel electrophoresis. The increase in the molecular weight due to the introduction of the His-tag was responsible for the resolution of His-tag-Sac7d as a separate band (separated from Sac7d without His-tag). The dimers cross-linked in different orientations using His-tag K9C with head cysteine mutants (K28C or G27C) differed in the number of His-tag regions per dimer resulting in the difference of their molecular weights and hence the distinction of dimers with different cross-linking orientations on gels.

7.2.2 DNA binding of Sac7d dimers

The DNA binding properties of the Sac7d dimers cross-linked in different orientations were studied using fluorescence titrations, circular dichroism and DNA affinity chromatography.

Reverse titrations (DNA into protein) followed by monitoring the fluorescence quenching of W24 were used to study the DNA binding properties of Sac7d dimers.

DNA binding properties showed the most change for the K9C dimer (tail-to-tail dimer). When compared to Sac7d monomer, K9C dimer bound to 100-mer [d(GC)] with 1000-fold higher affinity ($K=\sim10^{10}\,M^{-1}$).

Dimerization was previously shown to increase the binding affinity for several DNA binding proteins. Increased DNA binding affinity was reported in engineered dimers of both Lac and λ repressors. λ repressor showed a 10-fold enhancement in affinity as a result of covalent crosslinking, and the LacI repressor dimer showed a 10^4-fold higher affinity for DNA (Falcon and Matthews 2001; Kalodimos, Folkers et al. 2001).

The increase in the affinity of dimers can be attributed to a decreased entropic penalty for the second binding event after the first protein is bound. However, the increase in affinity is seen only if the binding of the second protein unit is facilitated by a favorable binding site. Dimerization in head-to-head and head-to-tail did not result in increased DNA binding affinity. The increase in the DNA binding affinity for K9C dimer indicates that first protein favored the binding of the second protein via a favorable DNA distortion. This claim is supported by the decrease in binding affinity of K9C dimer with the introduction of V26A and M29A double mutation in K9C dimer. The crystal structure of V26A, M29A Sac7d (VMSac7d)- DNA complex shows a roll angle at kink site of 26°, a reduction from 66° observed in the Sac7d-DNA complex. The overall bending is reduced to ~50° with the introduction of the double mutation. The binding decreased to $\sim10^7\,M^{-1}$ from $\sim10^{10}\,M^{-1}$ for K9C dimer with the introduction of V26A and M29A double mutation in titrations with constant with 100-mer [d(GC)].

Further evidence for DNA deformation induced preference in protein binding orientation was obtained from monitoring the DNA unwinding followed by the binding of K9C dimer. Forward titrations (protein into DNA) followed by monitoring the circular dichroism of the DNA indicated cooperative unwinding of 100-mer [d(GC)] with the increase in the binding density of K9C dimer. The CD band at the long wavelength (~280 nm) region for the 100-mer [d(GC)] increased in a cooperative manner with increasing protein. Previous studies indicate that the degree of cooperativity of DNA unwinding with increase in Sac7d binding density was dependent on DNA sequence (McAfee, Edmondson et al. 1996). Highest sigmoidicity in unwinding was observed with poly [d(IC)] followed by poly [d(GC)]. Negligible cooperativity was observed with poly [d(AG)]-poly[d(CT)] and poly[d(AC)]-ploy[d(GT)] (McAfee, Edmondson et al. 1996). The cooperative unwinding of 100-mer [d(GC)] with increase in the binding density K9C dimer observed in the present study is comparable to that observed with poly[d(IC)]-Sac7d binding previously reported. The increase in the degree of cooperative unwinding of [d(GC)] with the addition of K9C dimer indicates that the unwinding induced by the two monomer units of the K9C dimer operates in synergistic manner.

K9C dimer binds DNA with enhanced binding affinity (from reverse titrations using fluorescence) and introduces cooperative unwinding in DNA (forward titrations using CD). Increase in affinity translates to increased ease of binding of second protein after the first protein is bound. Cooperative unwinding indicates that the unwinding induced by the two adjacent proteins is synergistic. Therefore the increase in binding affinity observed with reverse titrations and the cooperative unwinding of the DNA observed in the forward titrations appear to be correlated. However, cooperative

unwinding in DNA does not indicate cooperative protein-DNA binding. Previous studies indicate no evidence of cooperativity for Sac7d-DNA binding in forward titrations (protein into DNA) conducted by fluorescence (McAfee, Edmondson et al. 1996).

The change in the long wavelength CD band represents conformational changes in DNA. It has been reported that the CD band at ~275nm is indicative of helix winding angle and base pair propellant twist (Baase and Johnson 1979; Johnson, Dahl et al. 1981). Binding of Sac7d to multiple DNA sequences (section 7.1) increased the CD band of DNA at long wavelength (~275 nm-290 nm) but the degree of increase in intensity was sequence dependent. Sac7d binding does not promote strand separation, but rather stabilizes DNA by increasing the T_m by ~40 °C (McAfee, Edmondson et al. 1995; McAfee, Edmondson et al. 1996). It was suggested that the susceptibility of the DNA to the protein induced conformational change is not related to binding affinity but reflects conformational preferences of the DNA (McAfee, Edmondson et al. 1996). Poly [d(IC)] does not undergo structural deformation until the protein concentration is close to saturation ([protein]:[DNA]=~0.2) unlike poly[d(GC)] that undergoes structural change with little protein ([protein]:[DNA]=~<0.1). Hence, DNA sequence dependent flexibility and structure play a role in deciding the degree of sigmoidicity imposed by Sac7d binding and cooperativity.

Sigmoidicity may also be imposed by protein binding as observed in the case of K9C dimer-100-mer [d(GC)] binding. Sigmoidal unwinding of DNA with protein binding requires synergistic action of two proteins. It is not possible to induce cooperative unwinding by non-associated protein binding i.e., the two Sac7d molecules should bind within a minimum distance so that the unwinding imposed by the two

proteins are reinforced in a synergistic manner. That means the unwinding introduced by a protein should influence the binding of its nearest neighbor or vice versa.

Present results indicate that in addition to the inherent DNA structure and bendability, the protein binding induced DNA bending also influences the binding of nearest neighbor Sac7d molecules. These deductions were validated by investigating the influence of protein binding induced DNA deformation on the binding of nearest neighbor without cross-linking constraints.

7.3 DNA binding of non-cross linked Sac7d

Single cysteine mutants of Sac7d were bound to DNA as monomers and cross-linked on DNA. Control reactions were conducted by cross-linking the cysteine mutants in the absence of DNA. A predominance of tail-to-tail dimers was observed in the cross-linking reactions conducted in the presence of DNA when compared to the control reactions in the absence of DNA. This indicates that there is preference in binding orientation for adjacent proteins on DNA. Dimer formation in the presence of DNA also indicates that the proteins bind in adequate proximity to facilitate cross-linking, supporting a previous claim that two adjacent Sac7d molecules should bind within a set distance to induce cooperative unwinding in DNA. Therefore, a DNA imposed preferred binding orientation is observed for two Sac7d molecules that bind at adjacent binding sites.

There are reports of DNA mediated assembly of proteins. For example, DNA mediated dimerization was reported for phage 434-repressor and yeast GCN4 DNA-binding domain (Guarnaccia, Raman et al. 2004). Dimers of phage 434-repressor and

yeast GCN4 DNA-binding domain were assembled on their respective cognitive sequences. The *in vitro* binding studies with DNA binding domains of 434-repressor and yeast GCN4 DNA without their dimerization domains resulted in a preferential dimerization on their respective cognitive sequences. Further, high specificity was observed for 434-repressor-cognitive site binding by competing with related engineered mutants of 434-repressor (Guarnaccia, Raman et al. 2004). DNA binding proteins not only recognize the inherent structural signature resulting from base composition and sequence but also recognize the bending introduced by other neighboring proteins. IHF recognizes pre-formed conformational characteristics of consensus DNA sequences at high affinity binding sites where as at other sites the relative binding affinity depends on the deformation energy required for the complex formation (Aeling, Opel et al. 2006).

Further, preference in DNA binding orientation for adjacent Sac7d on DNA was obtained by studying the Sac7d-DNA binding with out cross-linking using paramagnetic relaxation enhancement (PRE). The data indicated a clear preference in tail-to-tail orientation for adjacent Sac7d molecules on 12mer [d(GC)]. The preference in orientation was lost with the introduction of V26A, M29A double mutation.

7.4 Preferred binding orientation

DNA binding data of dimers, cross-linking Sac7d cysteine mutant on DNA and paramagnetic relaxation enhancement clearly point to a preferred tail-to-tail orientation for adjacent Sac7d molecules on DNA. This is in disagreement with the previous reports of Sso7d-DNA binding (Sac7d homologue from *Sulfolobus solfataricus*).

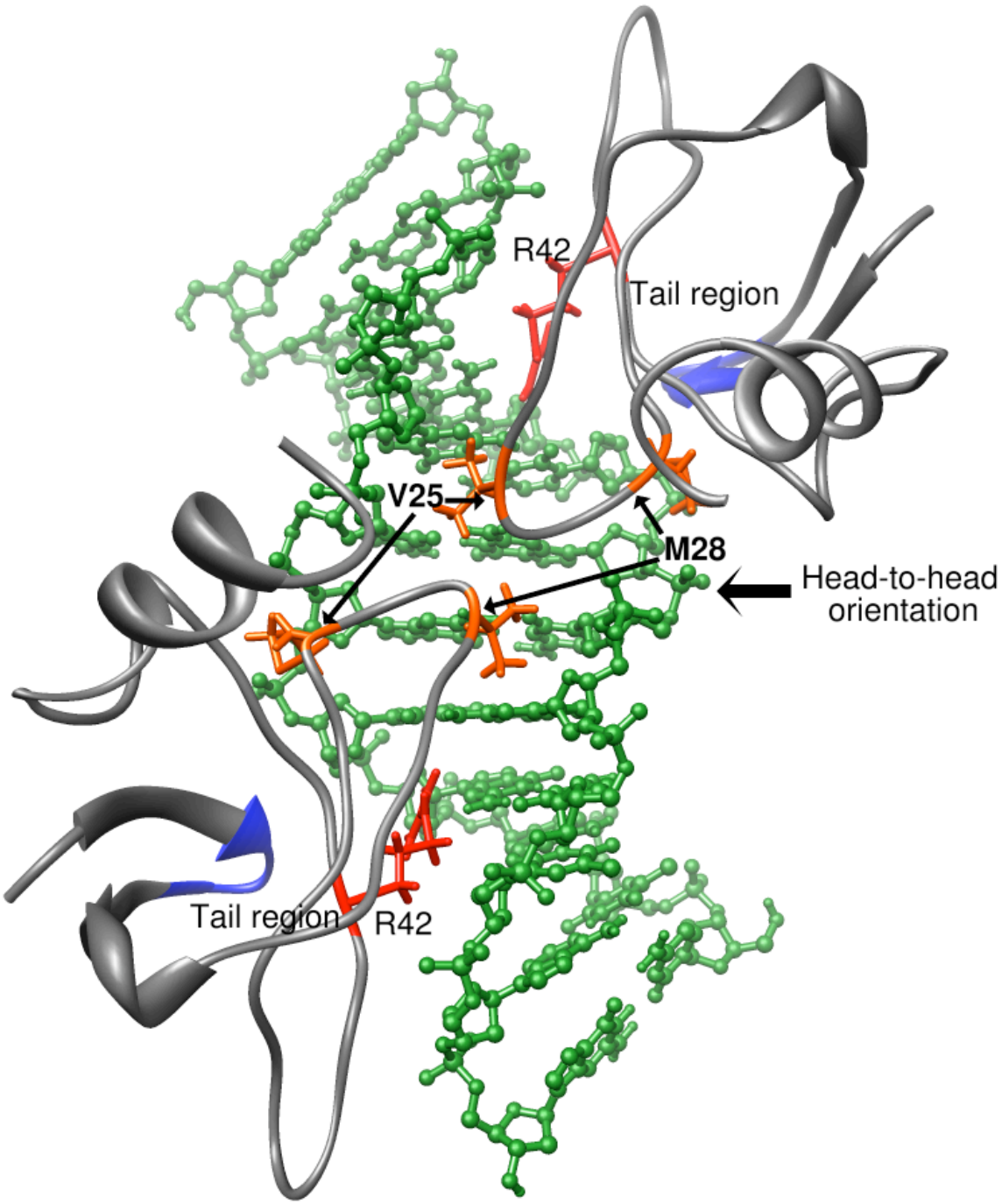


Figure 7.1 Solution structure of Sso7d-12mer DNA (1bbx). The DNA is show in green (ball and stick) and the protein in dark grey (ribbon). The anomalies with respect to the Sac7-DNA crystal structure (1azp) are pointed out and labeled (red).

The solution structure of Sso7d shows two Sso7d molecules in a preferential head-to-head orientation with two-fold symmetry on a 12mer GC rich sequence (1bbx) (Agback, Baumann et al. 1998). The Sso7d-1bbx structure is shown in Figure 7.1 and the deviations from the crystal structure of Sac7d are noted. The DNA is not bent in the NMR structure which is not in agreement with the Sac7d-DNA crystal structure (1azp) (Robinson, Gao et al. 1998). Therefore the head-to-head orientation for adjacent Sso7d molecules reported by 1bbx appears to be the result of poor quality NMR data.

The DNA binding surface of Sac7d or Sso7d that constitutes the β-barrel containing three β-sheets (β3,β4 and β5) is not clearly folded in the Sso7d-DNA complex. Also, the intercalating residues, V26 and M29 (V25 and M28 in Sso7d; numbering in Sso7d structure starts at second residue) point away from the DNA and are not involved in intercalation. In addition, the interaction of other amino acids observed in the crystal structure of Sac7d are not observed in the Sso7d-DNA complex. Therefore, the preference in head-to-head orientation for the adjacent proteins presented in 1bbx is not valid without the characteristic interactions of the protein and DNA that are observed in crystal structure. The preference for tail-to-tail orientation was tested using the 1bbx sequence. Crosslinking cysteine mutants of Sac7d bound to the 1bbx sequence resulted in a clear predominance in the tail-to-tail dimers in comparison to control reactions conducted without DNA. A PRE analysis using the 1bbx sequence with EDTA-Mn^{2+} label did not show any indication of preference for head-to-head orientation for adjacent Sac7d on 1bbx. The peak intensity profile was comparable to that observed in case of Sac7d-[d(GC)]EDTA-Mn^{2+} PRE analysis indicating a tail-to-tail orientation.

7.5 Role of intercalation and DNA bending in preferential orientation and indirect readout

Mutation of the intercalating residues in the K9C dimer results in a decrease of binding affinity. The crystal structure of V26A/M29A Sac7d-DNA complex has reduced bending (~50°) due to a decrease in the length of interacting side chains (Chen, Ko et al. 2005). V26A/M29A double mutation in K9C dimer resulted in a decrease of DNA binding affinity. The PRE assay with V26A/M29A Sac7d also did not indicate preferred tail-to-tail orientation that was otherwise observed with recombinant Sac7d. This indicates intercalating residues may be responsible for conferring preference in binding orientation for adjacent Sac7d molecules on DNA.

Intercalating residues are known to confer specificity in protein-DNA interactions involving non-sequence specific binding. Several high mobility group (HMG) chromosomal proteins that are common to all eukaryotes are known to bind to DNA by inserting one more amino acid side chains into DNA structure (Klass, Murphy et al. 2003). The crystal structure of HMG domain of *Drosophila melanogaster* protein HMG-D bound to DNA shows no-sequence specific model of DNA recognition mediated by intercalation of amino acid chains (Murphy, Sweet et al. 1999). V32 and T33 partially intercalate the DNA and serve as elements of specificity along with water mediated hydrogen bonds formed by S10. The specificity conferred by the intercalation and water-mediated hydrogen bonds serve as non-sequence specific recognition equivalents for HMG-D protein, when compared to its sequence specific counter part lymphoid enhancement factor LEF-1 (Murphy, Sweet et al. 1999). A putative transcription factor encoded by the sex-determining region of the human Y chromosome (SRY), a HMG

domain containing protein follows indirect recognition of DNA by partial intercalation of nonpolar side chains in the DNA minor groove (King and Weiss 1993). Based on the multiple examples, a clear correlation between DNA binding affinity and DNA bending propensity has been suggested.

Present data indicates a correlation between intercalation induced DNA bending and preference in binding orientation for adjacent Sac7d. A correlation between intercalation and sigmoidicity cannot be confirmed due to the lack of DNA unwinding with the binding of V26A/M29A Sac7d but a clear bend (~50°) in the crystal structure of VMSac7d-DNA complex. A correlation between binding affinity and intercalation cannot be confirmed because increased binding affinity of Sac7d monomer with poly [d(IC)] was not observed (McAfee, Edmondson et al. 1996).

7.6 Packaging multiple Sac7d on DNA and potential role of Sac7d *in vivo*

Sac7d is known to stabilize DNA by increasing the T_m by ~40 °C (McAfee, Edmondson et al. 1995). Forward titrations (protein into DNA) followed by monitoring the CD of DNA also indicate a gradual increase in the CD band at the long wavelength region indicating a transition from B-DNA to A-DNA. DNA unwinding observed by the change in circular dichroism of the DNA with increase in protein binding density cannot be correlated to DNA packaging functions of Sac7d. DNA packaging in nucleosome formation and in bacteriophage DNA packaging the CD band in the long wavelength decreases (Maestre, Gray et al. 1971; Cowman and Fasman 1978; McAfee, Edmondson et al. 1996). Previous studies also indicate that increasing ionic strength and decreasing temperature increase the helical twist of the DNA in solution and result in a decrease of

CD band at long wavelength (~275nm) (Gennis and Cantor 1972; Zimmer and Luck 1974; McAfee, Edmondson et al. 1996).

Small angle X-ray scattering analysis (SAXS) of multiple Sac7d molecules bound to a 32-mer [d(GC)] indicated that adjacent Sac7d binding in a sequential head-to-tail fashion along the minor groove of the DNA. The structure shows the adjacent Sac7d molecules bound on opposite sides of the DNA and the resulting kinks are in opposite directions leading to an over all zigzag structure of DNA. SAXS analysis indicates that there is essentially no DNA compaction. Hence, a potential regulatory role for Sac7d was suggested (Krueger, McCrary et al. 1999). Sac7d may also work with other DNA binding proteins to bring about DNA packaging. The preferential head-to-tail orientation for adjacent Sac7d on DNA suggested by the SAXS data does not agree with the preferential head-to-head orientation for two Sso7d molecules bound in a on a12mer GC-rich sequence (1bbx) (Agback, Baumann et al. 1998).

Based on the tail-to-tail preference indicated by other data, models of Sac7d-DNA complexes with Sac7d packed on to 32-mer[d(GC)] were created and their scattering profiles were computed and compared with the experimental scattering data from Sac7d-32-mer[d(GC)] complex. The comparison indicates that sequential head-to-head/tail-to-tail may be possible for multiple Sac7d on DNA. The binding of Sac7d in sequential tail-to-tail and head-to-head also do not seem to induce significant packing of DNA.

CHAPTER 8

CONCLUSION AND FUTURE WORK

8.1 Conclusions

A potential preference in the DNA binding orientation for adjacent Sac7d molecules on DNA has been investigated. Sac7d binding at two adjacent sites on DNA has been studied in three ways: 1. Characterization of DNA binding of Sac7d dimers cross-linked in three different orientations which are possible for the two Sac7d molecules binding at adjacent sites on DNA. 2. Characterization of the binding orientation of adjacent Sac7d molecules on DNA without cross-linking constraints, using paramagnetic relaxation enhancement (PRE). 3. Cross-linking the nearest neighbor Sac7d that are bound to adjacent sites on the DNA. Also, the packing of multiple Sac7d molecules on elongated DNA was tested using small angle X-ray scattering.

K9C dimer (tail-to-tail dimer) introduced cooperative DNA unwinding with increase in protein binding density, indicating a potential preference for tail-to-tail orientation of adjacent Sac7d. Sac7d dimers cross-linked in tail-to-tail orientation (K9C dimer) showed 100-1000-fold increase in binding affinity indicating potential preference.

The correlation between DNA bending and the potential preference in orientation was confirmed by studying the DNA binding affinity using V26A/M29A K9C dimer. Introduction of double mutation of V26A and M29A resulted in decrease of DNA binding affinity. VMSac7d does not introduce significant DNA deformation compared to native Sac7d. Therefore the loss of preference in binding orientation with the introduction of VM double mutation indicates the influence of protein induced DNA deformation on the preferential binding orientation for adjacent Sac7d molecules.

The PRE analysis of Sac7d with [d(GC)] suggested a clear preference for tail-to-tail orientation for adjacent Sac7d molecules. The correlation of DNA bending with the preference in binding orientation has been tested by conducted the PRE analysis using V26A/M29A Sac7d. The potential preference in tail-to-tail orientation observed with Sac7d was lost with the introduction of V26A and M29A double mutation. The PRE analysis with 1bbx sequence-Sac7d indicated no preference for head-to-head orientation previously suggested by Sso7d-1bbx solution structure (Agback, Baumann et al. 1998).

Cross-linking cysteine mutants of Sac7d that are bound at adjacent sites on DNA indicated a predominance of dimer population with tail-to-tail orientation. Cross-linking Sac7d cysteine mutants on 1bbx sequence formed dimers with preferential tail-to-tail orientation, which in clear disagreement with the preferential head-to-head orientation previously suggested by the NMR structure (Agback, Baumann et al. 1998).

All the three approaches indicate a clear preference for tail-to-tail orientation of adjacent Sac7d molecules. The data also indicates the influence of protein induced DNA deformation on the preferential binding orientation for adjacent Sac7d molecules. Present

data suggests an indication of correlation between the increase in DNA binding affinity and DNA bending caused due to intercalation.

8.2 Future work

Structural characterization of the tight binding dimer in tail-to-tail orientation (K9C dimer) is required for better understanding of the correlation between the increase in DNA binding affinity and preference in DNA binding orientation. The influence of length of intercalation should also be studied to establish the correlation between intercalation and preferential binding orientation.

Stability, non-sequence specific DNA binding and bending properties of Sac7d have been exploited for protein engineering (Yang and Wang 2004). The DNA binding, local unwinding and strand stabilization can be applied for protein engineering to enhance efficiency and processivity for DNA binding regulatory proteins.

APPENDICES

APPENDIX A

BIOCHEMISTRY AND MOLECULAR BIOLOGY TECHNIQUES AND DATA COLLECTION

A.1. Molecular techniques

A.1.1 Long-term Storage of *E. coli*

E. coli strains were maintained at –70 °C and stored for long periods. Cell cultures are grown to an optical density at 600 nm (OD_{600}) between 0.5 and 0.7. In a sterile 1.5ml microcentrifuge tube, 800 µl of cell culture is vortexed with 200 µl of sterile 50% glycerol solution. The samples were frozen in a liquid nitrogen bath and kept at -–70 °C for long-term storage.

A.1.2 Expression of Recombinant Protein

Two *E. coli* cell lines, BL21(DE3)pLysS and Rosetta(DE3)pLysS, were used in conjunction with pET3b plasmids to express recombinant protein. The Rosetta(DE3) pLysS *E. coli* cell strain contains extra copies of the rare *E. coli* argU, ileY, and leuW tRNA genes and was used to increase expression yield. The protein expression system

was developed by William F. Studier and co-workers, at Brookhaven National Laboratory (Studier, Rosenberg et al. 1990). Previous efforts inserted the Sac7d gene into pET3b to express recombinant 7 kDa Protein (McAfee, Edmondson et al. 1995). The protein was expressed with no additional residues Other than the initiating methionine (lacking the blocking formyl group indicated by sequencing and mass spectrometry). All Sac7d mutant proteins were produced using the materials and methods described below.

A seed culture was prepared by inoculating 20 ml of sterile LB media containing 60 μg/m1 ampicillin with *E. coli* Rosetta(DE3)pLysS cells containing the appropriate protein gene. The culture was incubated overnight at 37 °C with 250 rpm shaking. An intermediate culture was prepared by inoculating 4 ml of the overnight culture into 200 ml of sterile LB media containing 60 μg/m1 ampicillin. The intermediate culture was incubated at 37 °C with 225 rpm shaking or until the culture reached an OD= ~1.0 (approximately two hours). Aliquots of 50 ml of the intermediate culture were inoculated into 2 L of sterile LB media containing 60 μg/m1 ampicillin. The 2 L cultures were pre-warmed to 37 °C for inoculation. Final cultures were incubated at 37 °C with 250 rpm shaking until the OD reached ~1.0 (approximately 3 hours). Expression of the gene of interest was induced by addition of 3.85 ml of a stock solution of IPTG (final concentration 100 μg/m1). Protein expression was allowed to proceed for one hour at 37 °C with 225 rpm shaking. Cells were then harvested by centrifugation at 3750 rpm for 15 minutes at 4 °C. The supernatant was discarded and the cell pellets were immediately stored at –70 °C.

A.1.3 Expression of recombinant protein in minimal media

A seed culture was prepared by inoculating 7.75 ml of sterile water containing 60 μg/m1 ampicillin, 100μl of 0.1 g/ml NH_4C1, 2 ml of M9 salts, 20 μL 1 M MgSO4, 200μl 20% glucose, 2μl of thiamine hydrochloride and 2 μl 1 M $CaC1_2$ with *E. coli* Rosetta(DE3)pLysS cells containing the appropriate protein gene. The culture was incubated overnight at 37 °C with 250 rpm shaking. An intermediate culture was prepared by inoculating the 10ml overnight culture into 153 ml of sterile water containing 60 μg/ml ampicillin, 2 ml of 0.1 g/ml NH_4C1, 240 ml of M9 salts, 400 μl 1 M $MgSO_4$, 4 ml 20% glucose, 20μl of thiamine hydrochloride and 20 μl 1 M $CaC1_2$. The intermediate culture was incubated at 37 °C with 250 rpm shaking until the culture reached an $0D_{600}$= ~1.0. Aliquots of 50 ml of the intermediate culture were inoculated into 1.53 L of sterile water containing 60 μg/m1 ampicillin, 20 ml of 0.1 g/m1 NH_4C1, 400 ml of M9 salts, 4 ml 1 M $MgSO_4$, 40 ml 20% glucose, 200μl of thiamine hydrochloride and 200 μL 1 M $CaC1_2$. The 2 L cultures were pre-warmed to 37 °C for inoculation. Final cultures were incubated at 37 °C with 250 rpm shaking until the $0D_{600}$ reached ~1.0. Expression of the gene of interest was induced by addition of 3.85 ml of a stock solution of IPTG (final concentration 100 μg/m1). Protein expression was allowed to proceed for one hour at 37 °C with 225 rpm shaking. Cells were then harvested by centrifugation at 3750 rpm for 15 minutes at 4 °C. The supernatant was discarded and the cell pellets were immediately stored at –70 °C.

A.1.4 ^{15}N labeling of protein samples

Proteins for NMR experiments were labeled with ^{15}N by expression in minimal media supplemented with $^{15}NH_4C1$ in the growth medium. Growth and expression of the protein was performed as described in Section A.1.3.

A.1.5 Purification of Recombinant Sac7d from *E. coli*

E. coli cells were taken from –70 °C storage, placed on ice, and allowed to thaw. Thawed cell pellets were re-suspended in 100 ml of 10 mM Tris-HC1 (pH 8.0), 1 mM EDTA (pH 8.0), 0.1 % Triton X-100, and 0.5 mM PMSF. The cells were lysed on ice by sonication with output control set to 8, duty cycle set to 75%, and timer set to 2.3 minutes. Sonication was repeated three times with three minutes of cooling time allowed between each event. Between the second and the third sonication pulse, the cell-lysate is treated with 25 mg of DNase I and incubated at 37 °C for 5 minutes. The suspension was centrifuged at 50,000 rpm at 4 °C for 1 hour T1250 rotor in a Sorvall Discovery 90SE ultracentrifuge. The supernatant was dialyzed for 12 hours at 4 °C in 1000 molecular weight cut-off (MWCO) dialysis tubing against pre-chilled 0.2 M H_2SO_4. The precipitate was removed by centrifugation at 40,000 rpm for 30 minutes at 4 °C. The supernatant was then dialyzed 4 times against 10mM KH_2PO_4 (pH 7.2). Following dialysis, the solution was centrifuged at 40,000 rpm for 30 minutes at 4 °C. The final supernatant was filtered and applied to a 5 ml SP sepharose cation-exchange column on an Amersham Pharmacia Biotech HPLC AKTA purifier equilibrated with 10mM KH_2PO_4 (pH 7.2). The protein was eluted with a 75-mL linear gradient of NaC1 (0 to 1.0 M) in 10mM KH_2PO_4 (pH 7.2). The Sac7d and cysteine mutants of Sac7d eluted ~0.3 M NaCl. Protein

concentrations were determined using an ε_{280} of 1.1 mg ml^{-1} cm^{-1} for Sac7d and cysteine mutants of Sac7d.

A.1.6 Protein Storage and Handling

Eluted protein was stored at 4 °C in elution buffer in 10mM KH_2PO_4 (pH 7.2) with a small amount of sodium azide. Eluted protein solution was dialyzed extensively against the appropriate buffer or buffer exchanged using PD-10 desalting columns (GE Health care) for each particular experiment. Desired protein concentration was reached using Centrifugal filtration units (Millipore) (3000 MWCO).

A.1.7 Plasmid DNA Isolation

Double stranded plasmids (pET3b), were isolated and purified using Qiaprep Spin mini Kit obtained from Qiagen. The kit included buffers P1, P2, N3 PB, PE and EB. An *E.coli* culture was inoculated into 5ml of LB media containing 60µb/ml of ampicillin. The culture was incubated at 37 °C overnight at 250 rpm shaking. Cells were collected into sterile 1.5 ml microcentrifuge tubes by centrifugation at 14000. The supernatant was discarded after each spin. Cell pellet was resuspended in 200 µl P1 buffer to which RNase was already added before being stored. The cell suspension is then lysed by the addition 200 µl of buffer P2. The tubes were gently mixed to ensure proper mixing of the lysis solution. To each tube 350 µl of N3 solution was added and the contents were gently mixed by inverting the tubes several times. Tubes were not vortexed at this point. Mixing with N3 resulting the precipitation of cell debris that was removed by centrifugation for 20 minutes at 14,000 rpm. The supernatant was applied on to the columns provided with

the kit and centrifuged for 30 seconds inorder to bind the plasmid DNA on to column filter. After binding the DNA, the column filter was washed with 500µl of PB buffer by centrifuging for 30 seconds at 14,000 rpm. The column was then washed with 750µl of PE buffer to which 24ml of ethanol was already added. This was followed by another round of centrifugation to remove any residual wash buffer. Next, 30µl of deionized waster that was pre-heated to ~65 °C was applied to the center of each column. The column was allowed to stand for 1 minute. Double-stranded plasmid DNA was eluted into a sterile 1.5 ml microcentrifuge tube by 1 minute of centrifugation at 14,000 rpm for 30 seconds.

A.1.8 Competent *E. coli* Cells

Commercial competent cells obtained from Agilent along with the lightning mutagenesis were used for transforming all the mutant plasmid into *E. coli* cells.

A.1.19 Site-directed Mutagenesis

Site directed mutagenesis was conducted using Quick-change mutagenesis kit (obtained from Stratagene) or Lightning Mutagenesis kit (obtained from Agilent). The kit included dNTPs, 5x reaction buffer, Pfu ultra polymerase enzyme. Control reaction primer and Dnp restriction enzyme. To introduced mutation correct primers were designed using the primer design program available in Stratagene's website. Primers were obtained from IDT. To introduce mutations, PCR reaction was conducted according to the parameters recommended in the kit. The PCR mixture was treated with 2µl of Dnp restriction digestion enzyme for 5minutes at 37 °C. The Dnp treated reaction mixture was

used for transforming the competent cells that are available with the kit (section A.1.10). Plasmids were isolated from few of the colonies obtained to confirm the mutations by sequence analysis.

A.1.10 Chemical Transformation

A water bath was pre-heated to 42 °C, 5 ml of LB media was pre-heated to 37 °C, and an appropriate number of 15 ml sterile culture tubes were incubated on ice for 20 minutes. Five 1 of each plasmid to be transformed was mixed with 100 µ1 of competent cells in the pre-cooled 15 ml sterile culture tubes. The mixture was kept on ice for 20 minutes. The culture tubes were held in the 42 °C water bath for 1 minute and immediately place on ice. One ml of pre-warmed LB media was then added and the culture was incubated at 37 °C with 250 rpm shaking for 1 hour. Four ml of LB medium with 60 µg/m1 ampicillin was added to each culture tube. The culture was incubated at 37 °C with 250 rpm shaking for 12-16 hours. Cell stocks were prepared as described above.

A.1.11 Oxidation of cysteine mutants of Sac7d using diamide

The monomer and dimer protein fractions of cysteine mutants of Sac7d obtained from cation-exchange chromatography were pooled and treated with 0.1mM final concentration of diamide for ~ 1 hour at room temperature. The reaction mixture was diluted and loaded on to cation-exchange column for the separation of monomer and dimer fractions.

A.1.12 Chemical cross-linking of cysteine mutants of Sac7d using bifunctional cross-linking agents

The monomer and dimer protein fractions of cysteine mutants of Sac7d obtained from cation-exchange chromatography were pooled and treated with 10mM of DTT for ~30 minutes to create monomers. The DTT reaction was desalted using PD-10 desalting columns (GE Healthcare). The eluate from the PD-10 columns was treated with appropriate cross-linker (with a stock concentration of 20mM) such that the ratio of protein to cross-linker is approximately 1:5. The reactions were incubated at room temperature in dark for ~ 1 hour. The cross-linking reactions were then applied on to cation-exchange column to separate the monomer and the dimer fractions.

A.2 Data collection

A.2.1 Circular Dichroism (CD)

All the CD experiments were conducted using Olis 1500 instrument. Experimental temperature was controlled by Peltier thermopile temperature control unit. Data was collected in both near UV (320 nm to 200 nm) and far UV (260 nm – 180 nm) regions. Spectra were collected with 1 nm increments with an averaging time of 10 seconds. CD wavelength spectra were smoothed by procedures described by Savitsky and Golay (Savitzky and Golay 1964). The CD instrument was calibrated using *d*-camphor-10-silfonic acid. The ratio De_{190}/De_{290} was 2.0.

1 cm path length cells were used for far UV CD measurements that were typically used for obtained the CD scans of proteins. A protein concentration 0.15mg/ml was used. Protein-DNA binding experiments were conducted in 1 cm cells. The titration were

started with ~3ml of DNA ($O.D_{260}$=~0.5) solution in 10mM KH_2PO_4 (pH=7.2). Data was exported as ascii files and processed using igor.

A.2.2 Differential Scanning calorimetry

Typical DSC scans were done using 1 mg/ml protein concentration in 10mM KH_2PO_4 pH=7.2. The scans were conducted using VP-DSC instrument. The thermal denaturation data were fit using in-house software to obtain the melting parameters.

A.2.3 Small angle X-ray scattering

Present study employed programs to calculate the structural parameters from the pdb file using an all atom-weighted approach. The programs utilize GNOM analysis that follows alternate linear regularization method to calculate *P(r)* functions of the protein-DNA complexes. The scripts used in the analysis are displayed in the appendix C.

A.2.4 Nuclear Magnetic Resonance (NMR) Spectroscopy Sample Preparation

Sac7d and mutant proteins were overexpressed in E. coli in minimal media and purified as described in Section A.1.3 and A.1.4. Purified protein was concentrated in centrifugal filter units (Millipore). 15 ml volumes of protein solution were concentrated to approximately 5 ml. Then the total volume was brought to 15 ml with 10mM KH_2PO_4 (pH=7.2). This process was repeated several times to ensure that the protein sample was free of salt and buffer. In final centrifugation step the sample was concentrated down to ~ 1 ml. Samples for data collection were prepared to a final volume of ~600µl. With 10% D_2O and ~5µl of DSS. Quantity of the protein was ~5 mg/ml. The pH of each sample was

adjusted to 5.0 or 7.0 using dilute HCl and KOH and was measured with a Radiometer glass electrode. All NMR spectra were collected on a Varian 800 MHz or Varian 500 MHz NMR spectrometer. Data were processed using NMR pipe, NMRView and VNMR-J softwares using the scripts shown in appendix C.

A.2.5 Fluorescence Spectroscopy

Fluorescence titration measurements were performed on a Fluoromax-3 spectrofluorometer with 5 nm excitation and 5 nm emission slit widths. Binding titrations were performed with excitation at 295 nm and emission monitored at 355 nm. Reverse titrations (DNA into protein) were performed by adding aliquots of concentrated nucleotide solution to a protein sample in a 4 ml quartz fluorescence cell with stirring using a circular magnetic stirrer within the cell. Protein solutions were made gravimetrically from stock protein solutions of concentrations determined spectroscopically. Nucleotide concentrations were determined spectrophotometrically in a 1 cm cuvette using the appropriate extinction coefficient. Protein concentrations were determined spectrophotometrically in a 1 cm cuvette using the previously determined extinction coefficient of 1.1 for Sac7d (McAfee, Edmondson et al. 1995). Emission spectra and apparent intensities of solutions with high optical densities may be distorted due to absorption of emitted photons by the experimental sample. Attenuation of emission, inner filter effect, is most pronounced for fluorophores that have significant overlap of the absorption and emission spectra. However, because there was no attenuation of fluorescence signal observed during titrations, correction was not made for the inner filter effect during analysis. Binding parameters were obtained by using the

noncooperative McGhee-von Hippel model for non-specific binding (McGhee and von Hippel 1974).

A.2.6 Preparation of DNA affinity columns

NHS-activated sepharose 1ml columns were obtained from GE Healthcare. DNA oligonucleotides with a 5'-NH_2 modification were obtained from IDT (Integrated DNA technologies). The ligand coupling was performed using a 2 ml syringe for HiTrap 1 ml according to the following protocol. The isopropanol was washed out with 1 mM HCl, ice-cold. (3 × 2 ml)Care was taken to exceeding the flow rates of 1 ml/min (1/2 drop/sec) at this stage. Immediately 1 ml ligand solution was injected onto the column. The column was sealed and allowed to stand for 15–30 minutes at 25°C (or 4 hours at 4°C).

Any excess active groups that have not coupled to the ligand were deactivated and the non-specifically bound ligands were washed out using the following the procedure:

Buffer A: 0.5 M ethanolamine, 0.5 M NaCl, pH 8.3

Buffer B: 0.1 M sodium acetate, 0.5 M NaCl, pH 4

1. Buffer A of 3 × 2 ml was injected.
2. Buffer B of 3 × 2 ml was injected.
3. Buffer A of 3 × 2 ml was injected.
4. The column was left for 15–30 min in room temp (or ~4 hours in 4°C).
5. Buffer B of 3 × 2 ml was injected.
6. Buffer A of 3 × 2 ml was injected.
7. Buffer B of 3 × 2 ml was injected.
8. Finally, 2 ml of binding buffer with neutral pH was injected to adjust the pH.

APPENDIX B

STANDARD LABORATORY EQUIPMENT, MATERIALS, SOLUTIONS, BUFFERS, ENZYMES, ORGANISMS, MEDIA AND ANTIBIOTICS

Unless otherwise stated, all chemicals used were analytical grade and all solutions were freshly prepared as necessary. Unless otherwise stated all chemicals were obtained from Sigma Chemical Co., St. Louis, MO 63178. Unless otherwise stated all enzymes were supplied by New England Biolabs. Radioactivity was supplied by New England Nuclear (NEN), a DuPont corporation. Water used in this work was purified by a Reverse Osmosis system in line with a Barnstead-Thermolyne NanoPURE ultrapure water system. Protein dialysis was done using 1000 MWCO tubing from Fisher Biotech, Spectrum, Inc. UV spectrophotometric measurements were made using a Varian model UV/Vis Cary 300 Bio spectrophotometer.

All routine pH measurements were done with a Radiometer glass electrode connected to ThermoFisher digital reader.

Sonication of cells was done using a Branson Sonic Power Company, Danbury, CT 06810, Sonifier model 250. Preparative centrifugation was done using a Beckman

Instruments, Inc., optima L series, preparative ultracentrifuge, Palo Alto, CA 94304 or a Kendro Laboratory Equipment Sorvall, Discovery 90SE model, preparative ultracentrifuge, Newtown, CT 06470-2337. A DuPont Instruments, Sorvall RC-5B Super-speed refrigerated centrifuge was used for routine centrifugation. A J6B Beckman refrigerated centrifuge or a Avanti J-25 Beckman refrigerated centrifuge was used to harvest saturated cultures of *E. coli.* For experiments requiring use of polymerase chain reactions, Eppendorf Mastercycler Gradient thermal cycler was used.

Competent Cells: TSS: 85% LB medium, 10% PEG (wt/vol), MW 8000, Sigma), 5% DMSO (vol/vol), and 50 mM $MgCl_2$ (pH 6.5). Solution was autoclaved to sterilize and stored at 4 °C for < 2 weeks.

Plasmid isolation; Bacterial media and antibiotics: LB: 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaC1, adjust solution to pH 7.4, was sterilized by autoclaving.

LB agar plates: LB media made including 15 g/L agar before autoclaving. Desired antibiotic(s) was added after solution had cooled to 60 °C, and then plates were poured.

M9 salt: 64 g $Na_2HPO_4 \cdot 7H_20$, 15 g KH_2PO_4, 2.5 g NaC1 dissolved to 1L H_20 then autoclaved.

Ampicillin: 100 mg/ml (1000 x) stock solution in H_20. Filter sterilize and stored at 4 °C for up to one month.

APPENDIX C

SCRIPTS USED IN THE STUDY FOR DATA ANALYSIS

C.1 Scripts used for processing NMR data

C.1.1 A template used for processing 2D NMR data

```
#!/bin/csh
nmrPipe -in test.fid \
| nmrPipe -fn SOL -fl 16 \
#| nmrPipe -fn POLY -time \
#| nmrPipe  -fn LP -before -pred 3 \
#| nmrPipe -fn SP -off 0.45 -end 0.98 -pow 2 -c 0.5 \
# SP parameters for 1H \
| nmrPipe -fn SP -off 0.45 -end 0.98 -pow 2 -c 0.5 \
# GM parameters for 1H \
#| nmrPipe -fn GM -g1 20 -g2 35 -g3 .0 -c 0.5 \
| nmrPipe -fn ZF -size 4096 \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 0 -p1 0.0 -di -verb \
| nmrPipe -fn EXT -left -sw \
#| nmrPipe  -fn EXT -x1 9ppm -xn 6ppm  -sw \
| nmrPipe -fn TP \
| nmrPipe -fn LP \
# GM parameters for 13C \
#| nmrPipe  -fn GM -g1 30 -g2 110 -g3 .01 -c 0.5 \
# GM parameters for 15N \
#| nmrPipe -fn GM -g1 10 -g2 25 -g3 .0 -c 0.5 \
# SP parameters \
| nmrPipe -fn SP -off 0.45 -end 0.98 -pow 2 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
# | nmrPipe  -fn FT -neg -verb \
```

```
| nmrPipe -fn PS -p0 0 -p1 0 -di -verb \
#| nmrPipe -fn PS -p0 -90 -p1 180 -di -verb \
#| nmrPipe  -fn CS -ls 2.0ppm -sw \
#| nmrPipe  -fn BASE -nw 3 -nl 0% 5% 95% 100% \
| nmrPipe -fn TP \
#| nmrPipe -fn POLY -auto \
| nmrPipe -ov -out f1f3.dat
#nmrPipe -in f1f3.dat | pipe2xyz -nv -ov -out f1f3.nv
```

C.1.2 Script used assigning 2D IPAP data

```
# Use for assigning ipap xpeaks
# Change names of spectra and peaklists in code below
# to match your names.

# Spectra and associated xpeak lists are
# hsqc, adif, asum, udif, and usum
# i - expand about hsqc cursor
# I - more expand about Nhsqc cursor
# o - zoom out to predefined axis limits
# a - get assignemnt of closest xpeak of Nhsqc
#     and assign to last xpeak of
#     adiflist, asumlist, udiflist, and usumlist

bind .win_hsqc.0 <Key-a> {assign}
proc assign {} {
set pid [nv_peak closest k9cNhList]
echo $pid
nv_peak panel $pid
set lbln [nv_peak elem N15.L $pid]
set lblh [nv_peak elem H1.L $pid]
set pknum [nv_peak n adiflist]
set pkn [expr $pknum-1]
nv_peak elem N15.L adiflist.$pkn $lbln
nv_peak elem H1.L adiflist.$pkn $lblh
set pknum [nv_peak n asumlist]
set pkn [expr $pknum-1]
nv_peak elem N15.L asumlist.$pkn $lbln
nv_peak elem H1.L asumlist.$pkn $lblh
set pknum [nv_peak n udiflist]
set pkn [expr $pknum-1]
nv_peak elem N15.L udiflist.$pkn $lbln
nv_peak elem H1.L udiflist.$pkn $lblh
set pknum [nv_peak n usumlist]
set pkn [expr $pknum-1]
nv_peak elem N15.L usumlist.$pkn $lbln
nv_peak elem H1.L usumlist.$pkn $lblh
nv_peak panel $pid
}
```

```
bind .win_hsqc.0 <Key-n> {next}
proc next {} {
set xrng [expr 0.3]
set yrng [expr 2.75]
set pid [nv_peak closest k9cNhList]
set pknum [nv_peak idnum $pid]
set newpid [expr $pknum+1]
echo now using hsqc peak number $newpid
set px [nv_peak elem H1.P k9cNhList.$newpid]
set py [nv_peak elem N15.P k9cNhList.$newpid]
set x1 [expr $px+$xrng]
set x2 [expr $px-$xrng]
set y1 [expr $py+$yrng]
set y2 [expr $py-$yrng]
nv_win act .win_hsqc.0
nv_win ppm x $x1 $x2
nv_win ppm y $y1 $y2
nv_win draw
nv_win act .win_adif.0
nv_win ppm x $x1 $x2
nv_win ppm y $y1 $y2
nv_win draw
nv_win act .win_asum.0
nv_win ppm x $x1 $x2
nv_win ppm y $y1 $y2
nv_win draw
nv_win act .win_udif.0
nv_win ppm x $x1 $x2
nv_win ppm y $y1 $y2
nv_win draw
nv_win act .win_usum.0
nv_win ppm x $x1 $x2
nv_win ppm y $y1 $y2
nv_win draw
nv_win act .win_hsqc.0
}

bind .win_hsqc.0 <Key-I> {expand}
proc expand {} {
set xrng [expr 0.6]
set yrng [expr 5.5]
nv_win act .win_hsqc.0
set px [nv_win cross1x]
set py [nv_win cross1y]
set xl [expr $px-$xrng]
set xh [expr $px+$xrng]
set yl [expr $py-$yrng]
set yh [expr $py+$yrng]
nv_win ppm x $xl $xh
nv_win ppm y $yl $yh
nv_win draw
nv_win act .win_adif.0
nv_win ppm x $xl $xh
```

```
nv_win ppm y $yl $yh
nv_win draw
nv_win act .win_asum.0
nv_win ppm x $xl $xh
nv_win ppm y $yl $yh
nv_win draw
nv_win act .win_udif.0
nv_win ppm x $xl $xh
nv_win ppm y $yl $yh
nv_win draw
nv_win act .win_usum.0
nv_win ppm x $xl $xh
nv_win ppm y $yl $yh
nv_win draw
}

bind .win_hsqc.0 <Key-i> {expand2}
proc expand2 {} {
set xrng [expr 0.3]
set yrng [expr 2.75]
nv_win act .win_hsqc.0
set px [nv_win cross1x]
set py [nv_win cross1y]
set xl [expr $px-$xrng]
set xh [expr $px+$xrng]
set yl [expr $py-$yrng]
set yh [expr $py+$yrng]
nv_win ppm x $xl $xh
nv_win ppm y $yl $yh
nv_win draw
nv_win act .win_adif.0
nv_win ppm x $xl $xh
nv_win ppm y $yl $yh
nv_win draw
nv_win act .win_asum.0
nv_win ppm x $xl $xh
nv_win ppm y $yl $yh
nv_win draw
nv_win act .win_udif.0
nv_win ppm x $xl $xh
nv_win ppm y $yl $yh
nv_win draw
nv_win act .win_usum.0
nv_win ppm x $xl $xh
nv_win ppm y $yl $yh
nv_win draw
}

bind .win_hsqc.0 <Key-o> {compress}
proc compress {} {
set x1 [expr 9.77]
set x2 [expr 6.67]
set y1 [expr 131.4]
```

```
set y2 [expr 102.1]
nv_win act .win_hsqc.0
nv_win ppm x $x1 $x2
nv_win ppm y $y1 $y2
nv_win draw
nv_win act .win_adif.0
nv_win ppm x $x1 $x2
nv_win ppm y $y1 $y2
nv_win draw
nv_win act .win_asum.0
nv_win ppm x $x1 $x2
nv_win ppm y $y1 $y2
nv_win draw
nv_win act .win_udif.0
nv_win ppm x $x1 $x2
nv_win ppm y $y1 $y2
nv_win draw
nv_win act .win_usum.0
nv_win ppm x $x1 $x2
nv_win ppm y $y1 $y2
nv_win draw
nv_win act .win_hsqc.0
}
```

C.1.3 Script used for residual dipolar coupling from cross peak files

```
#!/usr/bin/perl

# read in 4 xpk files (dif and sum of aligned and unaligned
sample)
# constructed from IPAP data (2D spectra and xpk files)
# and calculate RDC for assigned peaks

# calculates aligned minus isotropic.
# if one of the atoms is N then the sign is reversed.
# so the sign of the RDC should be correct.

# option for normalizing DC to NH vector

# CAUTION - the sign of hnco DC may need to be obtained
elsewhere.
# Do not rely on this script for hnco.

# RDC scaling factors relative to N-H are
#    N-H      1.0
#    C-H     -2.1
#    Co-Ca   -1/5.05 = -0.198
#    HN-Co   -1/3.2  = -0.312
#    N-Co     1/8.33 =  0.120

# added type #5 for NhsqcS3 sequence 2007
```

```
#print "unaligned diff xpk: ";
#chop($u_dif = <STDIN>);
#print "unaligned sum xpk: ";
#chop($u_sum = <STDIN>);
#print "aligned diff xpk: ";
#chop($a_dif = <STDIN>);
#print "aligned sum xpk: ";
#chop($a_sum = <STDIN>);
$u_dif="udiflist.xpk";
$u_sum="usumlist.xpk";
$a_dif="adiflist.xpk";
$a_sum="asumlist.xpk";

print "1. N-Hn ipap (2D or 3D)\n";
print "2. N-Co from hnco\n";
print "3. Hn-Co from hnco\n";
print "4. Co-Ca from hnco or 2D-S3\n";
print "5. N-Hn from NhsqcS3 (coupling in H dimension)\n";
print "Select type of experiment: ";
chop($type = <STDIN>);
if($type == 2) {
    print "enter value of Jf: ";
    chop($Jf = <STDIN>);
    }
if($type == 4) {
    print "enter lambda scaling for 2D-S3, if any: ";
    chop($lambda = <STDIN>);
    if($lambda == 0) {
        $lambda=1; }
    }
$norm = 0;
if ($type >= 2) {
     print "normalize to NH ?: ";
     chop($_ = <STDIN>);
     if(/y/i) {
          $norm=1; }
     }

print "output RDC file name: ";
chop($output = <STDIN>);
$udi=0;
$usi=0;
$adi=0;
$asi=0;
open(IN,$u_dif) || die "open unaligned diff failed";
for ($i=0; $i<4; $i++) {
    <IN>;
    }
$_=<IN>;
s/\{//g;
s/\}//g;
@w=split;
```

```
if($type == 3 || $type==5) {
    $freq = $w[0];
    } else {
            $freq = $w[1];
            }
print "frequency is $freq \n";
$_=<IN>;
while (<IN>) {
    chop;
    @w = split;
    if ($w[1] ne "{?}" && $w[8] ne "{?}") {
        $ud1[$udi] = $w[1];
        $ud2[$udi] = $w[8];
        $uds1[$udi] = $w[2];
        $uds2[$udi] = $w[9];
        $udi++;
        }
    }
close(IN);
open(IN,$u_sum) || die "open unaligned sum failed";
for ($i=0; $i<6; $i++) {
    <IN>;
    }
while (<IN>) {
    chop;
    @w = split;
    if ($w[1] ne "{?}" && $w[8] ne "{?}") {
        $us1[$usi] = $w[1];
        $us2[$usi] = $w[8];
        $uss1[$usi] = $w[2];
        $uss2[$usi] = $w[9];
        $usi++;
        }
    }
close(IN);
open(IN,$a_dif) || die "open aligned diff failed";
for ($i=0; $i<6; $i++) {
    <IN>;
    }
while (<IN>) {
    chop;
    @w = split;
    if ($w[1] ne "{?}" && $w[8] ne "{?}") {
        $ad1[$adi] = $w[1];
        $ad2[$adi] = $w[8];
        $ads1[$adi] = $w[2];
        $ads2[$adi] = $w[9];
        $adi++;
        }
    }
close(IN);
open(IN,$a_sum) || die "open aligned sum failed";
for ($i=0; $i<6; $i++) {
```

```
    <IN>;
    }
while (<IN>) {
    chop;
    @w = split;
    if ($w[1] ne "{?}" && $w[8] ne "{?}") {
        $as1[$asi] = $w[1];
        $as2[$asi] = $w[8];
        $ass1[$asi] = $w[2];
        $ass2[$asi] = $w[9];
        $asi++;
        }
    }
close(IN);
#for ($i=0; $i<$udi; $i++) {
#    print "$i   $ud1[$i]    $ud2[$i] \n";
#    }

$u=0;
for ($i=0; $i<$udi; $i++) {
    for ($j=0; $j<$usi; $j++) {
        if ($ud1[$i] eq $us1[$j] && $ud2[$i] eq $us2[$j]) {
            if($type == 3) {
                $unalign[$u] = abs($uds1[$i] - $uss1[$j]);
            } else {
                $unalign[$u] = abs($uds2[$i] - $uss2[$j]);
                }
            $unalignN[$u] = $us2[$j];
            $unalignH[$u] = $us1[$j];
#print "$u  $unalignN[$u]   $unalign[$u]    $uds2[$i] - $uss2[$j]
\n";
            $u++;
            }
        }
    }
$a=0;
for ($i=0; $i<$adi; $i++) {
    for ($j=0; $j<$asi; $j++) {
        if ($ad1[$i] eq $as1[$j] && $ad2[$i] eq $as2[$j]) {
            if($type == 3) {
                $align[$a] = abs($ads1[$i] - $ass1[$j]);
            } else {
                $align[$a] = abs($ads2[$i] - $ass2[$j]);
                }
            $alignN[$a] = $as2[$j];
            $alignH[$a] = $as1[$j];
#print "$a  $alignN[$a]   $align[$a]    $ads2[$i] - $ass2[$j]
\n";
            $a++;
            }
        }
    }
```

```
$err = 0.5;
$order = 1.0;
$count=0;
open(OUT,">$output");
for ($i=0; $i<$u; $i++) {
    for ($j=0; $j<$a; $j++) {
        if ($unalignN[$i] eq $alignN[$j]) {
            $rdc[$count] = ($align[$j] - $unalign[$i]) * $freq;
            $temp = index($alignN[$j],".") - 1;
            $temp2 = index($alignN[$j],"}") - 2;
            $rdcN1[$count] = substr($alignN[$j],1,$temp);
            $rdcA1[$count] = substr($alignN[$j],$temp+2,$temp2-
$temp);
            $temp = index($alignH[$j],".") - 1;
            $temp2 = index($alignH[$j],"}") - 2;
            $rdcN2[$count] = substr($alignH[$j],1,$temp);
            $rdcA2[$count] = substr($alignH[$j],$temp+2,$temp2-
$temp);
            if($type == 2) {
                $rdcA2[$count] = "C";
                $rdcN2[$count] = $rdcN1[$count] - 1;
                $rdc[$count] = $rdc[$count] / ($Jf+1);
                 if($norm == 1) {
                       $rdc[$count] = $rdc[$count] * 8.33; }
                 }
            if($type == 3) {
                $rdcA1[$count] = "C";
                $rdcN1[$count] = $rdcN2[$count] - 1;
                 if($norm == 1) {
                       $rdc[$count] = $rdc[$count] * (-3.2); }
                 }
           if($type == 4) {
                 $rdcA2[$count] = "CA";
                 if($rdcA1[$count]="N") {    # fix for 2D N-NH of
S3 exp
                     $rdcA1[$count] = "C";
                     $rdcN1[$count] = $rdcN1[$count]-1;
                     }
                 $rdcN2[$count] = $rdcN1[$count];
                 $rdc[$count] = $rdc[$count] / $lambda;
                 if($norm == 1) {
                       $rdc[$count] = $rdc[$count] * (-5.05); }
                 }
            if ($type == 1 || $type == 2 || $type ==5) {
                $rdc[$count] = -$rdc[$count]; }
            printf OUT "%-3i %3s %3i %3s  %8.3f    %3.1f
%4.2f\n",$rdcN1[$count], $rdcA1[$count], $rdcN2[$count],
$rdcA2[$count], $rdc[$count], $err, $order;
            $count++;
            }
        }
    }
```

```
print "$u unaligned and $a aligned crosspeaks \n";
print "$count matching peaks found \n";
```

C.2 Scripts used for calculating scattering profiles from pdb files

C.2.1 Script used for calculating Rc values from pdb coordinates

```
c program serc.f
c Stephen Edmondson, 1996
c
c Compute Rc for rod/helix from pdb file
c Axis must be aligned along X-, Y-, or Z- axis.
c
      parameter (nmax=50000)
      dimension x(nmax),y(nmax),z(nmax),fact(nmax),xr(nmax)
      dimension x1(nmax),x2(nmax)
      character*4 atom(nmax),ctmp,dtmp,aname,res
      character*80 input,output

      write(6,10)
10    format('input pdb file: ',$)
      read(5,11)input
11    format(a80)
      write(6,15)
15    format('helix axis (1=x,2=y,3=z): ',$)
      read(5,*)iaxis
      if(iaxis.lt.1.OR.iaxis.gt.3) goto 999
      open(unit=10,file=input)
      i=1
      do 40 k=1,nmax
        call pdbrd(10,iatom,atom(i),res,nres,x(i),y(i),z(i))
c     write(31,31)atom(i),x(i),y(i),z(i)
c31   format(a4,3(1x,f10.4))
        if(iatom.eq.0)goto 50
        if(atom(i)(1:2).ne.'LP') then
           call wtrad(atom,fact(i),xr(i))
           i=i+1
        endif
40    continue
50    natoms=i-1
      close(10)
      write(6,52)natoms
52    format('number of atoms: ',i5)
      if(iaxis.eq.1) then
          do 60 i=1,natoms
            x1(i)=y(i)
```

```
60          x2(i)=z(i)
      else
          if(iaxis.eq.2) then
            do 62 i=1,natoms
                x1(i)=x(i)
62              x2(i)=z(i)
          else
            do 64 i=1,natoms
                x1(i)=x(i)
64              x2(i)=y(i)
          endif
      endif
c
      sx=0
      sy=0
      f=0
      do 70 i=1,natoms
          sx=sx+x1(i)*fact(i)
          sy=sy+x2(i)*fact(i)
          f=f+fact(i)
70    continue
      sx=sx/f
      sy=sy/f
c
      dd=0
      f=0
      do 100 i=1,natoms
      dis=(x1(i)-sx)**2 + (x2(i)-sy)**2
      dd = dd + dis*fact(i)
      f = f+fact(i)
100   continue
      rc = sqrt(dd/f)
      write(6,120)rc
120   format('Rc = ',f6.3)
c
999   continue
      end
ccccccccccccccccccccccccccccccccccccccccccccccccc
      subroutine wtrad(atom,f1,xr)
c
c Get weight and radius of atoms.
c
      character*4 atom

      f1=0
      xr=1.0
          if(atom(1:1).eq.'H') then
            f1=1
            xr=1.2
          endif
          if(atom(1:1).eq.'C') then
            f1=6
            xr=1.90
```

```
          endif
          if(atom(1:1).eq.'N') then
            f1=7
            xr=1.5
          endif
          if(atom(1:1).eq.'P') then
            f1=31
            xr=1.90
          endif
          if(atom(1:1).eq.'O') then
            f1=8
            xr=1.40
          endif
          if(atom(1:1).eq.'S') then
            f1=16
            xr=1.85
          endif
          if(atom(1:2).eq.'NA') then
            f1=11
            xr=0.95
          endif
          if(atom(1:2).eq.'CL') then
            f1=17
            xr=1.81
          endif
      if(f1.eq.0) then
          write(6,20)atom
20        format('ATOM NOT FOUND -> ',a4)
      endif
      return
      end
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
      subroutine pdbrd(in,iatnum,atnam,rsnam,irsnum,x,y,z)
C
c
      integer in,iatnum,irsnum
      character*4 atnam,rsnam
      real x,y,z
c
      character*80 buf
c        ... buffer for input line
      character*4 cat
c        ... temporary string for revised atom name
c
c     --- read lines until 'ATOM' or 'HETATM' record, or until
EOF ---
  100 continue
           read(in,'(a)',end=1000) buf
      if (buf(1:4) .ne. 'ATOM' .and. buf(1:4) .ne. 'HETA') go to
100
c     --- skip atoms that represent alternate conformers
      if (buf(17:17) .ne. ' ' .and. buf(17:17) .ne. 'A') go to
100
```

```
c     --- test for column number of start of atom name ---
      if (buf(13:13) .eq. ' ') then
c           --- heavy-atom name, starts in col. 14 ---
            read(buf,9000,err=2000)
iatnum,atnam,rsnam,irsnum,x,y,z
      else
c           --- H/D name, starts in col. 13 ---
            read(buf,9010,err=2000)
iatnum,atnam,rsnam,irsnum,x,y,z
      endif
c  --- now fix up atom name for hydrogens if Brookhaven wierdo
      if (buf(13:13).eq.'1' .or. buf(13:13).eq.'2'
     +    .or. buf(13:13).eq.'3') then
        cat(1:1) = buf(14:14)
        cat(2:2) = buf(15:15)
        cat(3:3) = buf(16:16)
        if (cat(3:3).eq.' ') then
          cat(3:3) = buf(13:13)
          cat(4:4) = ' '
        else
          cat(4:4) = buf(13:13)
        end if
        read (cat,'(a4)') atnam
      end if
      return
c
c     EOF
 1000 continue
      iatnum=0
      return
c     --- error reading PDB record ---
 2000 continue
      write(5,'(/2x,a,/,a)') 'Error reading PDB line:',buf
 9000 format(6x,i5,2x,a3,1x,a3,2x,i4,4x,3f8.3)
 9010 format(6x,i5,1x,a4,1x,a3,2x,i4,4x,3f8.3)
      end
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc
```

APPENDIX D

CHEMIACL CROSS-LINKERS, CHEMICAL MODIFICATIONS AND PROTEIN SEQUENCES

D.1 Chemical Cross-linkers used for cross-linking cysteine mutants of Sac7d

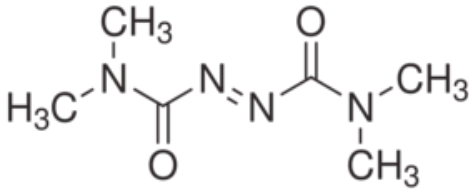
$BM(PEG)_2$ – 1,8 Bis maleimidodiethyleneglycol

BMB – 1,4-Bis maleimidobutane

BMDB – 1,4-Bis maleimidyl-2,3-dihydroxybutane

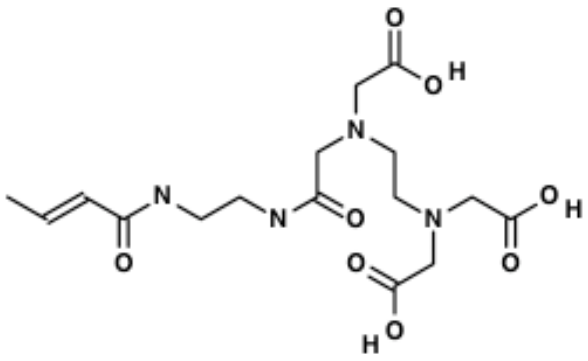
BMH – 1,6-Bis maleimidohexane

pPDM – p-phenylenedimaleimide



1,1'-Azobis(*N,N*-dimethylformamide)

D.2 Structure of EDTA modification to the dT in DNA for PRE assay.



The structure was generated using NCBI PubChem sketcher. In the above figure, the left end $-CH_2$ is attached to thiamine ring. The chelating center is on the right.

D.3 Amino acid sequence of Sac7d

MVKVKFKYKG EEKEVDTSKI KKVWRVGKMV SFTYDDNGKT

GRGAVSEKDA PKELLDMLAR AEREKK

REFERENCES

Aeling, K. A., M. L. Opel, et al. (2006). "Indirect recognition in sequence-specific DNA binding by Escherichia coli integration host factor: the role of DNA deformation energy." J Biol Chem **281**(51): 39236-39248.

Agback, P., H. Baumann, et al. (1998). "Architecture of nonspecific protein-DNA interactions in the Sso7d-DNA complex." Nature Structural Biology **5**(7): 579-584.

Aggarwal, A. K., D. W. Rodgers, et al. (1988). "Recognition of a DNA operator by the repressor of phage 434: a view at high resolution." Science (New York, N.Y.) **242**(4880): 899-907.

Allemann, R. K. and M. Egli (1997). "DNA recognition and bending." Chemistry & Biology **4**(9): 643-650.

Baase, W. A. and W. C. Johnson, Jr. (1979). "Circular dichroism and DNA secondary structure." Nucleic Acids Res **6**(2): 797-814.

Bak, M., E. Hanna, et al. (1995). "[Multidrug resistance of testicular cancers. (Detection of P-glycoprotein and MDR1 gene expression and their clinical connection)]." Orv Hetil **136**(1): 19-25.

Battiste, J. L. and G. Wagner (2000). "Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data." Biochemistry **39**(18): 5355-5365.

Bell, C. E. and M. Lewis (2000). "A closer view of the conformation of the Lac repressor bound to operator." Nature Structural Biology **7**(3): 209-214.

Bevington, P. R. and D. K. Robinson (2003). Data reduction and error analysis for the physical sciences, McGraw-Hill.

Bewley, C. A., A. M. Gronenborn, et al. (1998). "Minor groove-binding architectural proteins: structure, function, and DNA recognition." Annual Review of Biophysics and Biomolecular Structure **27**: 105-131.

Biyani, K., M. A. Kahsai, et al. (2005). "Solution structure, stability, and nucleic acid binding of the hyperthermophile protein Sso10b2." Biochemistry **44**(43): 14217-14230.

Bloembergen, N. and L. O. Morgan (1961). "Proton Relaxation Times in Paramagnetic Solutions. Effects of Electron Spin Relaxation." The Journal of Chemical Physics **34**(3): 842-842.

Bloemendal, M. and W. C. Johnson, Jr. (1995). "Structural information on proteins from circular dichroism spectroscopy possibilities and limitations." Pharm Biotechnol **7**: 65-100.

Brukner, I., M. Dlakic, et al. (1993). "Evidence for opposite groove-directed curvature of GGGCCC and AAAAA sequence elements." Nucleic Acids Research **21**(4): 1025-1029.

Brukner, I., R. Sánchez, et al. (1995). "Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides." The EMBO Journal **14**(8): 1812-1818.

Bujalowski, W. and T. M. Lohman (1987). "Limited co-operativity in protein-nucleic acid interactions. A thermodynamic model for the interactions of Escherichia coli single strand binding protein with single-stranded nucleic acids in the "beaded", (SSB)65 mode." J Mol Biol **195**(4): 897-907.

Buttinelli, M., A. Minnock, et al. (1998). "The exocyclic groups of DNA modulate the affinity and positioning of the histone octamer." Proceedings of the National Academy of Sciences **95**(15): 8544-8549.

Card, P. B., P. J. A. Erbel, et al. (2005). "Structural basis of ARNT PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization." J Mol Biol **353**(3): 664-677.

Cavanagh, J., W. J. Fairbrother, et al. (1995). Protein NMR Spectroscopy: Principles and Practice, Academic Press.

Chasman, D. I., K. M. Flaherty, et al. (1993). "Crystal structure of yeast TATA-binding protein and model for interaction with DNA." Proceedings of the National Academy of Sciences of the United States of America **90**(17): 8174-8178.

Chastain, P. D. and R. R. Sinden (1998). "CTG repeats associated with human genetic disease are inherently flexible." Journal of Molecular Biology **275**(3): 405-411.

Chen, C. Y., T. P. Ko, et al. (2005). "Probing the DNA kink structure induced by the hyperthermophilic chromosomal protein Sac7d." Nucleic Acids Res **33**(1): 430-438.

Chen, S., J. Vojtechovsky, et al. (2001). "Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: DNA binding specificity based on energetics of DNA kinking." Journal of Molecular Biology **314**(1): 63-74.

Cherney, L. T., M. M. Cherney, et al. (2009). "The structure of the arginine repressor from Mycobacterium tuberculosis bound with its DNA operator and Co-repressor, L-arginine." Journal of Molecular Biology **388**(1): 85-97.

Choli, T., B. Wittmann-Liebold, et al. (1988). "Microsequence analysis of DNA-binding proteins 7a, 7b, and 7e from the archaebacterium Sulfolobus acidocaldarius." J. Biol. Chem. **263**(15): 7087-7093.

Clore, G. M. and A. M. Gronenborn (1991). "Two-, three-, and four-dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution." Annu Rev Biophys Biophys Chem **20**: 29-63.

Clore, G. M. and C. D. Schwieters (2006). "Concordance of residual dipolar couplings, backbone order parameters and crystallographic B-factors for a small alpha/beta protein: a unified picture of high probability, fast atomic motions in proteins." J Mol Biol **355**(5): 879-886.

Clore, G. M., C. Tang, et al. (2007). "Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement." Curr Opin Struct Biol **17**(5): 603-616.

Couch, G. S., D. K. Hendrix, et al. (2006). "Nucleic acid visualization with UCSF Chimera." Nucleic Acids Research **34**(4): e29-e29-e29-e29.

Cowman, M. K. and G. D. Fasman (1978). "Circular dichroism analysis of mononucleosome DNA conformation." Proc Natl Acad Sci U S A **75**(10): 4759-4763.

Crothers, D. M. (1994). "Upsetting the balance of forces in DNA." Science **266**(5192): 1819-1820.

Dame, R. T. (2005). "The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin." Mol Microbiol **56**(4): 858-870.

Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-293.

Dickerson, R. E., D. S. Goodsell, et al. (1994). ""...the tyranny of the lattice..."." Proceedings of the National Academy of Sciences of the United States of America **91**(9): 3579-3583.

Donaldson, L. W., N. R. Skrynnikov, et al. (2001). "Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy." J Am Chem Soc **123**(40): 9843-9847.

Dosset, P., J. C. Hus, et al. (2001). "A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings." J Biomol NMR **20**(3): 223-231.

Drew, H. R., R. M. Wing, et al. (1981). "Structure of a B-DNA dodecamer: conformation and dynamics." Proceedings of the National Academy of Sciences **78**(4): 2179-2183.

Dvoretsky, A., V. Gaponenko, et al. (2002). "Derivation of structural restraints using a thiol-reactive chelator." FEBS Lett **528**(1-3): 189-192.

Edmondson, S. P., L. Qiu, et al. (1995). "Solution structure of the DNA-binding protein Sac7d from the hyperthermophile Sulfolobus acidocaldarius." Biochemistry **34**(41): 13289-13304.

Edmondson, S. P. and J. W. Shriver (2001). "DNA binding proteins Sac7d and Sso7d from Sulfolobus." Methods in Enzymology **334**: 129-145.

Ellenberger, T. E., C. J. Brandl, et al. (1992). "The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex." Cell **71**(7): 1223-1237.

Escalante, C. R., E. Nistal-Villán, et al. (2007). "Structure of IRF-3 bound to the PRDIII-I regulatory element of the human interferon-beta enhancer." Molecular Cell **26**(5): 703-716.

Fairall, L., J. W. Schwabe, et al. (1993). "The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition." Nature **366**(6454): 483-487.

Falcon, C. M. and K. S. Matthews (2001). "Engineered disulfide linking the hinge regions within lactose repressor dimer increases operator affinity, decreases sequence selectivity, and alters allostery." Biochemistry **40**(51): 15650-15659.

Gao, Y. G., S. Y. Su, et al. (1998). "The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA." Nature Structural Biology **5**(9): 782-786.

Gaponenko, V., J. W. Howarth, et al. (2000). "Protein global fold determination using site-directed spin and isotope labeling." Protein Science: A Publication of the Protein Society **9**(2): 302-309.

Gartenberg, M. R. and D. M. Crothers (1988). "DNA sequence determinants of CAP-induced bending and protein binding affinity.", Published online: 30 June 1988; | doi:10.1038/333824a0 **333**(6176): 824-829.

Garvie, C. W. and C. Wolberger (2001). "Recognition of specific DNA sequences." Molecular Cell **8**(5): 937-946.

Geggier, S. and A. Vologodskii (2010). "Sequence dependence of DNA bending rigidity." Proceedings of the National Academy of Sciences of the United States of America **107**(35): 15421-15426.

Gennis, R. B. and C. R. Cantor (1972). "Optical studies of a conformational change in DNA before melting." J Mol Biol **65**(3): 381-399.

Goddard, T. D., C. C. Huang, et al. (2005). "Software extensions to UCSF chimera for interactive visualization of large molecular assemblies." Structure (London, England: 1993) **13**(3): 473-482.

Goddard, T. D., C. C. Huang, et al. (2007). "Visualizing density maps with UCSF Chimera." J Struct Biol **157**(1): 281-287.

Grant, R. A., M. A. Rould, et al. (2000). "Exploring the Role of Glutamine 50 in the Homeodomain−DNA Interface: Crystal Structure of Engrailed (Gln50 → Ala) Complex at 2.0 Å†,‡." Biochemistry **39**(28): 8187-8192.

Guarnaccia, C., B. Raman, et al. (2004). "DNA-mediated assembly of weakly interacting DNA-binding protein subunits: in vitro recruitment of phage 434 repressor and yeast GCN4 DNA-binding domains." Nucleic Acids Res **32**(17): 4992-5002.

Hagerman, P. J. (1988). "Flexibility of DNA." Annual Review of Biophysics and Biophysical Chemistry **17**: 265-286.

Halford, S. E. and J. F. Marko (2004). "How do site-specific DNA-binding proteins find their targets?" Nucleic Acids Research **32**(10): 3040-3052.

Hamès, C., D. Ptchelkine, et al. (2008). "Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins." The EMBO Journal **27**(19): 2628-2637.

Hanes, S. D. and R. Brent (1989). "DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9." Cell **57**(7): 1275-1283.

Hansen, D. F., M. A. S. Hass, et al. (2003). "Detection of short-lived transient protein-protein interactions by intermolecular nuclear paramagnetic relaxation: plastocyanin from Anabaena variabilis." J Am Chem Soc **125**(23): 6858-6859.

Haran, T. E. and U. Mohanty (2009). "The unique structure of A-tracts and intrinsic DNA bending." Quarterly Reviews of Biophysics **42**(1): 41-81.

Harrison, S. C. and A. K. Aggarwal (1990). "DNA recognition by proteins with the helix-turn-helix motif." Annual Review of Biochemistry **59**: 933-969.

Hegde, R. S. (2002). "The papillomavirus E2 proteins: structure, function, and biology." Annual Review of Biophysics and Biomolecular Structure **31**: 343-360.

Hegde, R. S., S. R. Grossman, et al. (1992). "Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target." Nature **359**(6395): 505-512.

Honig, B. and A. Nicholls (1995). "Classical electrostatics in biology and chemistry." Science **268**(5214): 1144-1149.

Horton, N. C., L. F. Dorner, et al. (2001). "Sequence selectivity and degeneracy of a restriction endonuclease mediated by DNA intercalation." Nat Struct Mol Biol **9**(1): 42-47.

Iwahara, J., D. E. Anderson, et al. (2003). "EDTA-derivatized deoxythymidine as a tool for rapid determination of protein binding polarity to DNA by intermolecular paramagnetic relaxation enhancement." J Am Chem Soc **125**(22): 6634-6635.

Iwahara, J. and G. M. Clore (2006). "Detecting transient intermediates in macromolecular binding by paramagnetic NMR." Nature **440**(7088): 1227-1230.

Iwahara, J., C. D. Schwieters, et al. (2004). "Characterization of nonspecific protein-DNA interactions by 1H paramagnetic relaxation enhancement." J Am Chem Soc **126**(40): 12800-12808.

Iwahara, J., C. D. Schwieters, et al. (2004). "Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule." J Am Chem Soc **126**(18): 5879-5896.

Iwahara, J., C. Tang, et al. (2007). "Practical aspects of (1)H transverse paramagnetic relaxation enhancement measurements on macromolecules." J Magn Reson **184**(2): 185-195.

Iwahara, J., M. Zweckstetter, et al. (2006). "NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA." Proc Natl Acad Sci U S A **103**(41): 15062-15067.

Iwahara, T., T. Akagi, et al. (2004). "CrkII regulates focal adhesion kinase activation by making a complex with Crk-associated substrate, p130Cas." Proc Natl Acad Sci U S A **101**(51): 17693-17698.

Jain, N. U., A. Venot, et al. (2001). "Distance mapping of protein-binding sites using spin-labeled oligosaccharide ligands." Protein Sci **10**(11): 2393-2400.

Jayaram, B., K. A. Sharp, et al. (1989). "The electrostatic potential of B-DNA." Biopolymers **28**(5): 975-993.

Johnson, B. A. (2004). "Using NMRView to visualize and analyze the NMR spectra of macromolecules." Methods Mol Biol **278**: 313-352.

Johnson, B. B., K. S. Dahl, et al. (1981). "Correlations between deoxyribonucleic acid structural parameters and calculated circular dichroism spectra." Biochemistry **20**(1): 73-78.

Johnson, P. E., E. Brun, et al. (1999). "The cellulose-binding domains from Cellulomonas fimi beta-1, 4-glucanase CenC bind nitroxide spin-labeled cellooligosaccharides in multiple orientations." J Mol Biol **287**(3): 609-625.

Johnson, W. C., Jr. (1988). "Secondary structure of proteins through circular dichroism spectroscopy." Annu Rev Biophys Biophys Chem **17**: 145-166.

Jones, J. M., M. Montcouquiol, et al. (2006). "Inhibitors of differentiation and DNA binding (Ids) regulate Math1 and hair cell formation during the development of the organ of Corti." The Journal of Neuroscience: The Official Journal of the Society for Neuroscience **26**(2): 550-558.

Joshi, R., J. M. Passner, et al. (2007). "Functional specificity of a Hox protein mediated by the recognition of minor groove structure." Cell **131**(3): 530-543.

Juo, Z. S., T. K. Chiu, et al. (1996). "How proteins recognize the TATA box." Journal of Molecular Biology **261**(2): 239-254.

Kalodimos, C. G., N. Biris, et al. (2004). "Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes." Science **305**(5682): 386-389.

Kalodimos, C. G., R. Boelens, et al. (2004). "Toward an integrated model of protein-DNA recognition as inferred from NMR studies on the Lac repressor system." Chem Rev **104**(8): 3567-3586.

Kalodimos, C. G., G. E. Folkers, et al. (2001). "Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices." Proc Natl Acad Sci U S A **98**(11): 6039-6044.

Kim, J. L., D. B. Nikolov, et al. (1993). "Co-crystal structure of TBP recognizing the minor groove of a TATA element." Nature **365**(6446): 520-527.

Kim, Y., J. H. Geiger, et al. (1993). "Crystal structure of a yeast TBP/TATA-box complex." Nature **365**(6446): 512-520.

King, C. Y. and M. A. Weiss (1993). "The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity." Proc Natl Acad Sci U S A **90**(24): 11990-11994.

Klass, J., F. V. t. Murphy, et al. (2003). "The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity." Nucleic Acids Res **31**(11): 2852-2864.

Koo, H. S., H. M. Wu, et al. (1986). "DNA bending at adenine . thymine tracts." Nature **320**(6062): 501-506.

Kosa, P. F., G. Ghosh, et al. (1997). "The 2.1-Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box." Proceedings of the National Academy of Sciences **94**(12): 6042-6047.

Kosen, P. A. (1989). "Spin labeling of proteins." Methods Enzymol **177**: 86-121.

Kosower, N. S., E. M. Kosower, et al. (1969). "Diamide, a new reagent for the intracellular oxidation of glutathione to the disulfide." Biochem Biophys Res Commun **37**(4): 593-596.

Krueger, J. K., B. S. McCrary, et al. (1999). "The solution structure of the Sac7d/DNA complex: a small-angle X-ray scattering study." Biochemistry **38**(32): 10247-10255.

Laity, J. H., H. J. Dyson, et al. (2000). "DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers." Journal of Molecular Biology **295**(4): 719-727.

Lefstin, J. A. and K. R. Yamamoto (1998). "Allosteric effects of DNA on transcriptional regulators." Nature **392**(6679): 885-888.

Lesser, D. R., M. R. Kurpiewski, et al. (1990). "The energetic basis of specificity in the Eco RI endonuclease--DNA interaction." Science **250**(4982): 776-786.

Lewis, M., G. Chang, et al. (1996). "Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer." Science **271**(5253): 1247-1254.

Liang, B., J. H. Bushweller, et al. (2006). "Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy." J Am Chem Soc **128**(13): 4389-4397.

Lohman, T. M. and D. P. Mascotti (1992). "Nonspecific ligand-DNA equilibrium binding parameters determined by fluorescence methods." Methods Enzymol **212**: 424-458.

Love, J. J., X. Li, et al. (1995). "Structural basis for DNA bending by the architectural transcription factor LEF-1.", Published online: 31 August 1995; | doi:10.1038/376791a0 **376**(6543): 791-795.

Lu, X. J., Z. Shakked, et al. (2000). "A-form conformational motifs in ligand-bound DNA structures." Journal of Molecular Biology **300**(4): 819-840.

Luijsterburg, M. S., M. C. Noom, et al. (2006). "The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective." J Struct Biol **156**(2): 262-272.

Lurz, R., M. Grote, et al. (1986). "Electron microscopic study of DNA complexes with proteins from the Archaebacterium Sulfolobus acidocaldarius." The EMBO Journal **5**(13): 3715-3721.

Luscombe, N. M., S. E. Austin, et al. (2000). "An overview of the structures of protein-DNA complexes." Genome Biology **1**(1): REVIEWS001-REVIEWS001.

Luscombe, N. M. and J. M. Thornton (2002). "Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity." Journal of Molecular Biology **320**(5): 991-1009.

Mack, D. R., T. K. Chiu, et al. (2001). "Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: a comparative analysis of T-A and A-T steps within A-tracts." Journal of Molecular Biology **312**(5): 1037-1049.

Macvanin, M. and S. Adhya (2012). "Architectural organization in E. coli nucleoid." Biochimica Et Biophysica Acta.

Maestre, M. F., D. M. Gray, et al. (1971). "Magnetic circular dichroism study on synthetic polynucleotides, bacteriophage structure, and DNA's." Biopolymers **10**(12): 2537-2553.

Marenduzzo, D., C. Micheletti, et al. (2010). "Biopolymer organization upon confinement." Journal of Physics. Condensed Matter: An Institute of Physics Journal **22**(28): 283102-283102.

McAfee, J. G., S. P. Edmondson, et al. (1995). "Gene cloning, expression, and characterization of the Sac7 proteins from the hyperthermophile Sulfolobus acidocaldarius." Biochemistry **34**(31): 10063-10077.

McAfee, J. G., S. P. Edmondson, et al. (1996). "Equilibrium DNA Binding of Sac7d Protein from the Hyperthermophile Sulfolobus acidocaldarius: Fluorescence and Circular Dichroism Studies†,‡." Biochemistry **35**(13): 4034-4045.

McGhee, J. D. and P. H. von Hippel (1974). "Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice." J Mol Biol **86**(2): 469-489.

Meng, E. C., E. F. Pettersen, et al. (2006). "Tools for integrated sequence-structure analysis with UCSF Chimera." BMC Bioinformatics **7**(1): 339-339.

Morris, J. H., C. C. Huang, et al. (2007). "structureViz: linking Cytoscape and UCSF Chimera." Bioinformatics (Oxford, England) **23**(17): 2345-2347.

Murphy, F. V. t., R. M. Sweet, et al. (1999). "The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition." EMBO J **18**(23): 6610-6618.

Murzin, A. G. (1993). "OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences." EMBO J **12**(3): 861-867.

Nadassy, K., S. J. Wodak, et al. (1999). "Structural features of protein-nucleic acid recognition sites." Biochemistry **38**(7): 1999-2017.

Nadeau, J. G. and D. M. Crothers (1989). "Structural basis for DNA bending." Proceedings of the National Academy of Sciences of the United States of America **86**(8): 2622-2626.

Nelson, H. C., J. T. Finch, et al. (1987). "The structure of an oligo(dA).oligo(dT) tract and its biological implications." Nature **330**(6145): 221-226.

Okumura, H., M. Yoshimura, et al. (2012). "Regulation of chromosomal replication initiation by oriC-proximal DnaA-box clusters in Bacillus subtilis." Nucleic Acids Research **40**(1): 220-234.

Olah, G. A., D. M. Gray, et al. (1995). "Structures of fd gene 5 protein.nucleic acid complexes: a combined solution scattering and electron microscopy study." J Mol Biol **249**(3): 576-594.

Olson, W. K., A. A. Gorin, et al. (1998). "DNA sequence-dependent deformability deduced from protein–DNA crystal complexes." Proceedings of the National Academy of Sciences of the United States of America **95**(19): 11163-11168.

Olson, W. K. and V. B. Zhurkin (2011). "Working the kinks out of nucleosomal DNA." Current Opinion in Structural Biology **21**(3): 348-357.

Otwinowski, Z., R. W. Schevitz, et al. (1988). "Crystal structure of trp repressor/operator complex at atomic resolution." Nature **335**(6188): 321-329.

Pabo, C. O. and R. T. Sauer (1992). "Transcription factors: structural families and principles of DNA recognition." Annual Review of Biochemistry **61**: 1053-1095.

Panne, D. (2008). "The enhanceosome." Current Opinion in Structural Biology **18**(2): 236-242.

Panne, D., T. Maniatis, et al. (2007). "An atomic model of the interferon-beta enhanceosome." Cell **129**(6): 1111-1123.

Parekh, B. S. and G. W. Hatfield (1996). "Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model." Proceedings of the National Academy of Sciences of the United States of America **93**(3): 1173-1177.

Pérez-Martín, J. and M. Espinosa (1993). "Protein-induced bending as a transcriptional switch." Science (New York, N.Y.) **260**(5109): 805-807.

Peters, W. B., S. P. Edmondson, et al. (2004). "Thermodynamics of DNA binding and distortion by the hyperthermophile chromatin protein Sac7d." Journal of Molecular Biology **343**(2): 339-360.

Peters, W. B., S. P. Edmondson, et al. (2005). "Effect of mutation of the Sac7d intercalating residues on the temperature dependence of DNA distortion and binding thermodynamics." Biochemistry **44**(12): 4794-4804.

Pettersen, E. F., T. D. Goddard, et al. (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." Journal of computational chemistry **25**(13): 1605-1612.

Pintilie, G. D., J. Zhang, et al. (2010). "Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions." J Struct Biol **170**(3): 427-438.

Privalov, P. L., A. I. Dragan, et al. (2009). "The cost of DNA bending." Trends Biochem Sci **34**(9): 464-470.

Protozanova, E., P. Yakovchuk, et al. (2004). "Stacked–Unstacked Equilibrium at the Nick Site of DNA." Journal of Molecular Biology **342**(3): 775-785.

Ramos, A. and G. Varani (1998). "A New Method To Detect Long-Range Protein−RNA Contacts: NMR Detection of Electron−Proton Relaxation Induced by Nitroxide Spin-Labeled RNA." J. Am. Chem. Soc. **120**(42): 10992-10993.

Rastinejad, F., T. Wagner, et al. (2000). "Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1." The EMBO Journal **19**(5): 1045-1054.

Ren, J., T. C. Jenkins, et al. (2000). "Energetics of DNA Intercalation Reactions†." Biochemistry **39**(29): 8439-8447.

Rice, P. A., S. Yang, et al. (1996). "Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn." Cell **87**(7): 1295-1306.

Robinson, H., Y. G. Gao, et al. (1998). "The hyperthermophile chromosomal protein Sac7d sharply kinks DNA." Nature **392**(6672): 202-205.

Rohs, R., X. Jin, et al. (2010). "Origins of specificity in protein-DNA recognition." Annual Review of Biochemistry **79**: 233-269.

Rohs, R., H. Sklenar, et al. (2005). "Structural and energetic origins of sequence-specific DNA bending: Monte Carlo simulations of papillomavirus E2-DNA binding sites." Structure (London, England: 1993) **13**(10): 1499-1509.

Rohs, R., S. M. West, et al. (2009). "The role of DNA shape in protein-DNA recognition." Nature **461**(7268): 1248-1253.

Roosild, T. P., J. Greenwald, et al. (2005). "NMR structure of Mistic, a membrane-integrating protein for membrane protein expression." Science (New York, N.Y.) **307**(5713): 1317-1321.

Rückert, M. and G. Otting (2000). "Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments." J. Am. Chem. Soc. **122**(32): 7793-7797.

Sanner, M. F., A. J. Olson, et al. (1996). "Reduced surface: an efficient way to compute molecular surfaces." Biopolymers **38**(3): 305-320.

Savitzky, A. and M. J. E. Golay (1964). "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." Analytical Chemistry **36**(8): 1627-1639.

Schumacher, M. A., K. Y. Choi, et al. (1994). "Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices." Science (New York, N.Y.) **266**(5186): 763-770.

Schwieters, C. D., J. J. Kuszewski, et al. (2003). "The Xplor-NIH NMR molecular structure determination package." J Magn Reson **160**(1): 65-73.

Serdyuk, I. N., T. N. Tsalkova, et al. (1987). "Determination of radii of gyration of particles by small-angle neutron scattering: calculation of the effect of aggregates." J Mol Biol **194**(1): 126-128.

Shakked, Z., G. Guzikevich-Guerstein, et al. (1994). "Determinants of repressor/operator recognition from the structure of the trp operator binding site." Nature **368**(6470): 469-473.

Shakked, Z. and D. Rabinovich (1986). "The effect of the base sequence on the fine structure of the DNA double helix." Progress in Biophysics and Molecular Biology **47**(3): 159-195.

Sharp, K. A., B. Honig, et al. (1990). "Electrical potential of transfer RNAs: codon-anticodon recognition." Biochemistry **29**(2): 340-346.

Shriver, J. W., W. B. Peters, et al. (2001). "Calorimetric analyses of hyperthermophile proteins." Methods in Enzymology **334**: 389-422.

Sierk, M. L., Q. Zhao, et al. (2001). "DNA deformability as a recognition feature in the reverb response element." Biochemistry **40**(43): 12833-12843.

Siggers, T. W., A. Silkov, et al. (2005). "Structural alignment of protein--DNA interfaces: insights into the determinants of binding specificity." Journal of Molecular Biology **345**(5): 1027-1045.

Sreerama, N., S. Y. Venyaminov, et al. (1999). "Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy." Protein Sci **8**(2): 370-380.

Studier, F. W., A. H. Rosenberg, et al. (1990). "Use of T7 RNA polymerase to direct expression of cloned genes." Methods Enzymol **185**: 60-89.

Swinger, K. K. and P. A. Rice (2004). "IHF and HU: flexible architects of bent DNA." Curr Opin Struct Biol **14**(1): 28-35.

Tang, C., J. Iwahara, et al. (2006). "Visualization of transient encounter complexes in protein-protein association." Nature **444**(7117): 383-386.

Tang, C. L., E. Alexov, et al. (2007). "Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides." Journal of Molecular Biology **366**(5): 1475-1496.

Taylor, W. H. and P. J. Hagerman (1990). "Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. II. NaCl-dependence of DNA flexibility and helical repeat." Journal of Molecular Biology **212**(2): 363-376.

Travers, A. (1998). "Transcription: activation by cooperating conformations." Current Biology: CB **8**(17): R616-618-R616-618.

Travers, A. and G. Muskhelishvili (2005). "Bacterial chromatin." Current Opinion in Genetics & Development **15**(5): 507-514.

Travers, A. A. (1989). "DNA conformation and protein binding." Annual Review of Biochemistry **58**: 427-452.

Travers, A. A. (2004). "The structural basis of DNA flexibility." Philos Transact A Math Phys Eng Sci **362**(1820): 1423-1438.

Treisman, J., P. Gönczy, et al. (1989). "A single amino acid can determine the DNA binding specificity of homeodomain proteins." Cell **59**(3): 553-562.

Tucker-Kellogg, L., M. A. Rould, et al. (1997). "Engrailed (Gln50-->Lys) homeodomain-DNA complex at 1.9 A resolution: structural basis for enhanced affinity and altered specificity." Structure (London, England: 1993) **5**(8): 1047-1054.

Tullius, T. (2009). "Structural biology: DNA binding shapes up." Nature **461**(7268): 1225-1226.

Ukiyama, E., A. Jancso-Radek, et al. (2001). "SRYand architectural gene regulation: the kinetic stability of a bent protein-DNA complex can regulate its transcriptional potency." Molecular Endocrinology (Baltimore, Md.) **15**(3): 363-377.

Varani, L., S. I. Gunderson, et al. (2000). "The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein." Nature Structural Biology **7**(4): 329-335.

von Hippel, P. H. and O. G. Berg (1986). "On the specificity of DNA-protein interactions." Proceedings of the National Academy of Sciences of the United States of America **83**(6): 1608-1612.

Watkins, D., S. Mohan, et al. (2010). "Sequence recognition of DNA by protein-induced conformational transitions." J Mol Biol **396**(4): 1145-1164.

Werner, M. H., A. M. Gronenborn, et al. (1996). "Intercalation, DNA Kinking, and the Control of Transcription." Science **271**(5250): 778-784.

Werner, M. H., J. R. Huth, et al. (1995). "Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex." Cell **81**(5): 705-714.

West, S. M., R. Rohs, et al. (2010). "Electrostatic interactions between arginines and the minor groove in the nucleosome." Journal of Biomolecular Structure & Dynamics **27**(6): 861-866.

White, M. F. and S. D. Bell (2002). "Holding it together: chromatin in the Archaea." Trends in Genetics: TIG **18**(12): 621-626.

Wing, R., H. Drew, et al. (1980). "Crystal structure analysis of a complete turn of B-DNA.", Published online: 23 October 1980; | doi:10.1038/287755a0 **287**(5784): 755-758.

Woody, R. W., H. Sugeta, et al. (1996). "[Circular dichroism of proteins: recent developments in analysis and prediction]." Tanpakushitsu Kakusan Koso **41**(1): 56-69.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids, Wiley-Interscience.

Yang, J. M. and A. H. Wang (2004). "Engineering a thermostable protein with two DNA-binding domains using the hyperthermophile protein Sac7d." J Biomol Struct Dyn **21**(4): 513-526.

Yang, W. and T. A. Steitz (1995). "Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site." Cell **82**(2): 193-207.

Zahran, M., I. Daidone, et al. (2010). "Mechanism of DNA recognition by the restriction enzyme EcoRV." Journal of Molecular Biology **401**(3): 415-432.

Zimmer, C. and G. Luck (1974). "Conformation and reactivity of DNA. VI. Circular dichroism studies of salt-induced conformational changes of DNAs of different base composition." Biochim Biophys Acta **361**(1): 11-32.

Zimmerman, S. B. (2006). "Shape and compaction of Escherichia coli nucleoids." J Struct Biol **156**(2): 255-261.