

**THE ROLE OF PHOSPHATE NEUTRALIZATION IN ECORV-INDUCED DNA
BENDING**

by

Stephen P. Hancock

B.S. York College of Pennsylvania, 2001

Submitted to the Graduate Faculty of
Arts and Sciences in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH
FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Stephen Paul Hancock

It was defended on

August 4, 2008

and approved by

Karen Arndt, PhD Associate Professor, Department of Biological Sciences

Graham F. Hatfull, PhD Professor, Department of Biological Sciences

Craig L. Peebles, PhD Professor, Department of Biological Sciences

Gordon S. Rule, PhD Professor, Department of Biological Sciences Carnegie Mellon University

Dissertation Advisor: Linda Jen-Jacobson, PhD Professor, Department of Biological Sciences

Copyright © by Stephen Paul Hancock

2008

THE ROLE OF PHOSPHATE NEUTRALIZATION IN ECORV INDUCED DNA BENDING

Stephen Paul Hancock, PhD

University of Pittsburgh, 2008

DNA bending by DNA binding proteins is required to facilitate a myriad of essential cellular processes, including genome packaging, the formation of multi-protein complexes necessary for the expression of genetic information, the regulation of gene expression, and the correct positioning of recognition and catalytic elements required for site-specific hydrolysis of DNA. One mechanism proposed for protein induced DNA bending is that protein mediated neutralization of negatively charged phosphates on one face of the DNA produces a collapse of the DNA toward the neutralized surface. In order to further understand the role of phosphate neutralization in DNA bending, I have manipulated the energy required for achieving the 50° axial bend in the EcoRV endonuclease-DNA complex by either removing cationic protein side chains that contact DNA phosphates and/or replacing the charged phosphate with an uncharged methylphosphonate. I present evidence that neutralization of particular phosphates, positioned on the concave face of the bound DNA, can contribute favorably not only to the formation of the EcoRV-recognition complex but also to cleavage of the GATATC site. In addition, synergistic effects are observed when particular combinations of phosphates are neutralized. Fluorescence resonance energy transfer studies show that there is no significant difference in the degree of DNA bending in unmodified and modified complexes, implying that phosphate neutralization modulates the energetic cost of bending rather than the extent of bending. Further, van't Hoff analyses indicate that removal of interphosphate repulsion by phosphate neutralization contributes favorable enthalpy to EcoRV-DNA complex formation, and molecular dynamics

simulations show that this favorable enthalpy does not derive from the formation of new contacts between the introduced methyl group and the protein. Taken together, these results support the model that asymmetric phosphate neutralization by a site-specific protein promotes DNA bending. My work marks the first in depth thermodynamic analysis of the impact of phosphate neutralization on protein-induced bending by a DNA bending protein. Further, it shows that this is an energetic strategy employed by proteins to overcome the energetic cost of DNA bending.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	XVII
1.0 INTRODUCTION.....	1
1.1 LEVELS OF SEQUENCE SELECTIVITY FOR PROTEINS THAT INTERACT WITH DNA.....	2
1.2 RESTRICTION ENDONUCLEASES AS A MODEL FOR STRINGENT DISCRIMINATION OF DNA BINDING SITES.....	3
1.3 STRUCUTRAL AND ENERGETIC DETERMINANTS OF SPECIFICITY	4
1.3.1 Three types of binding sites for restriction endonucleases.....	4
1.3.2 Structural features of specific protein-DNA complexes	7
1.3.2.1 “Direct readout” and the structural determinants of specificity	7
1.3.2.2 “Indirect readout” and DNA deformation as a determinant of specificity	8
1.3.2.3 Protein conformational transitions in specific protein-DNA complexes.....	13
1.3.3 Structural features of non-cognate protein-DNA complexes.....	17
1.3.3.1 The non-specific complex is a ‘loose’ complex.	17
1.3.3.2 DNA distortion in non-cognate protein-DNA complexes.....	19

1.3.3.3	Conformational transitions and dynamics in non-cognate complex formation	20
1.3.4	Energetic features of specific protein-DNA interactions.....	21
1.3.5	Thermodynamic and kinetic basis for sequence discrimination by type II restriction endonucleases.....	25
1.3.5.1	Sequence discrimination in binding.....	25
1.3.5.2	Sequence discrimination in cleavage.....	29
1.3.6	Three different binding modes: pre-transition state, adaptive, and nonspecific complexes	32
1.4	THE ROLE OF DNA DISTORTION IN SPECIFICITY DETERMINATION.....	37
1.5	DNA BENDING INDUCED BY DNA BINDING PROTEINS	38
1.5.1	Sources of DNA rigidity.....	39
1.5.1.1	DNA helical rigidity and base stacking.....	39
1.5.1.2	Electrostatic origins of DNA rigidity	41
1.5.2	Two types of DNA bending proteins	43
1.5.2.1	“Type I” DNA bending proteins.....	45
1.5.2.2	Type II DNA bending proteins	46
1.6	ASYMMETRIC PHOSPHATE NEUTRALIZATION INDUCES DNA BENDING	47
1.6.1	Free DNA bending by APN measured experimentally.....	50
1.6.2	Free DNA bending by asymmetric phosphate neutralization measured computationally	55

1.7	INVESTIGATING THE ROLE OF PHOSPHATE NETRALIZATION IN PROTEIN INDUCED DNA BENDING: ECORV AS A MODEL	56
2.0	CHAPTER 2: THE EFFECTS OF PHOSPHATE NEUTRALZATION ON FORMATION OF THE ECORV-DNA COMPLEX	58
2.1	INTRODUCTION	58
2.1.1	Manipulating the electrostatic landscape at the EcoRV-DNA interface	58
2.1.2	Measuring the effects of electrostatic perturbations on formation of the EcoRV-DNA complex	62
2.1.3	Thermodynamic pseudocycles	63
2.2	RESULTS AND DISCUSSION	66
2.2.1	Basic side chain residues stabilize the specific EcoRV-DNA complex...	66
2.2.2	Single P _{Me} modifications impact EcoRV-DNA complex formation	74
2.2.3	Phosphate neutralization contributes favorable ΔG^0_{bind} to EcoRV-DNA complex formation	80
2.2.4	Electrostatic modifications affect the dependence of EcoRV-DNA binding on salt concentration.....	85
2.2.5	Synergistic, additive, and more than additive effects of multiple P _{Me} substitutions	89
2.2.6	Position dependence for the benefits of phosphate neutralization on specific EcoRV-DNA complex formation	92
2.2.7	Förster resonance energy transfer as a probe for DNA bending.	95
2.2.8	Substrate design and fluorophore labeling strategy	97

2.2.9	Modified and unmodified EcoRV-DNA complexes show similar extents of DNA bending.....	109
3.0	CHAPTER 3: INSIGHT INTO THE ENTHALPIC AND ENTROPIC CONSEQUENCES OF ASYMMETRIC PHOSPHATE NEUTRALIZATION	112
3.1	INTRODUCTION	112
3.1.1	Thermodynamic dissection of ΔG for the EcoRV-DNA interaction	112
3.2	RESULTS AND DISCUSSION	127
3.2.1	Effects of phosphate neutralization on ΔH^0 , ΔS^0 , and ΔC_p^0	129
3.2.1.1	Thermodynamics parameters: K119A-GATATC vs. wt-GATATC..	130
3.2.1.2	Thermodynamic parameters: K119A-Rp-P _{Me-2} vs. K119A-GATATC	131
3.2.1.3	Thermodynamic parameters wt-Rp-P _{Me-2} vs wt-GATATC.....	131
3.2.2	Molecular dynamics simulations as a probe for the effects of electrostatic modifications on the specific EcoRV-DNA complex	134
3.2.3	Global structural features of molecular dynamic simulated structural models	137
3.2.4	Local structural effects of modified EcoRV-DNA complexes.....	148
3.2.5	Relating structural observations to thermodynamic data	157
3.2.5.1	Phosphate neutralization contributes favorable ΔH^0_{bind} to EcoRV-DNA complex formation	162
3.2.5.2	Molecular factors that contribute to the thermodynamic parameters for K119A-TAG complex formation	163

3.2.5.3	Molecular factors that contribute to the thermodynamics of wt-Rp- P_{Me-2} complex formation	167
3.2.5.4	Molecular factors that affect the thermodynamics of K119A-Rp- P_{Me-2} complex formation relative to that for wt-Rp- P_{Me-2}	171
3.2.6	Phosphate neutralization by Rp- P_{Me-2} substitution impacts the electrostatic potential at the EcoRV-DNA interface	174
3.3	CONCLUSION AND DISCUSSION	179
3.3.1	On the magnitude of the energetic effects of phosphate neutralization relative to the total energetic requirement for DNA bending	180
4.0	THE EFFECTS OF PHOSPHATE NEUTRALIZATION ON ECORV-DNA TRANSITION STATE FORMATION.....	184
4.1	INTRODUCTION	184
4.2	RESULTS AND DISCUSSION	189
4.2.1	Measuring the effects of electrostatic perturbation on the formation of the EcoRV-DNA transition state complex	189
4.2.2	Basic side chain residues stabilize the transition state complex	191
4.2.3	Effects of P_{Me} substitutions on wt EcoRV-DNA transition state formation.....	193
4.2.4	Phosphate neutralization contributes favorable $\Delta G^{0\ddagger}$ to EcoRV-DNA transition state formation	194
4.2.5	Effects of multiple P_{Me} modifications on EcoRV-DNA transition state formation.....	197

4.2.6	The effects of phosphate neutralization on the efficiency of the EcoRV restriction endonuclease.	197
4.2.7	Conclusion	200
5.0	DISCUSSION	201
6.0	METHODS	213
6.1	SYNTHESIS AND PURIFICATION OF OLIGONUCLEOTIDE SUBSTRATES	213
6.2	DNA CONCENTRATION DETERMINATION AND DUPLEX FORMATION.....	215
6.3	5' END LABELLING	216
6.4	DETERMINATION OF EQUILIBRIUM BINDING CONSTANTS	216
6.4.1	Direct Equilibrium binding.....	217
6.4.2	Specific activity determination by converse binding	218
6.4.3	Equilibrium Competition	220
6.5	DNA CLEAVAGE KINETICS	222
6.6	FÖRSTER RESONANCE ENERGY TRANSFER.....	223
6.7	MOLECULAR DYNAMICS SIMULATIONS	224
6.8	VAN'T HOFF ANALYSIS	225
APPENDIX A	226
APPENDIX B	229
APPENDIX C	231
APPENDIX D	236
BIBLIOGRAPHY	238

LIST OF TABLES

Table 1.1 Discrimination at the protein-DNA interface: Three types of binding modes ^a	36
Table 2.1 Energetic effects of basic residue mutations on EcoRV-DNA complex formation.	69
Table 2.2 Origins of the energetic penalties observed for mutant EcoRV complex formation....	74
Table 2.3 Effects of single P _{Me} modifications on EcoRV-DNA pre-transition state complex formation.....	76
Table 2.4 Phosphate neutralization contributes favorable free energy to EcoRV-DNA complex formation.....	83
Table 2.5 Energetic effects of multiple P _{Me} substitutions.	90
Table 2.6 FRET efficiency calculations for various substrates.	105
Table 2.7 Relative extent of bending for modified and unmodified EcoRV-DNA complexes..	110
Table 3.1 Factors affecting thermodynamic contributions to specific protein-DNA complex formation.....	115
Table 3.2 Estimated ΔC_p^0 calculated from accessible surface area values	121
Table 3.3 Thermodynamic parameters for EcoRV-DNA binding obtained by a multi-parametric fit.....	133
Table 3.4 Thermodynamic parameters for EcoRV-DNA binding using the method of Clark and Glew.....	133
Table 3.5 Structural asymmetries in various EcoRV-DNA X-ray crystal models.	148

Table 3.6 Predicted effects on the thermodynamics of complex formation: K119A-Rp-P _{Me-2} vs. K119A-TAG.	163
Table 3.7 Predicted effects on the thermodynamics of complex formation: K119A-TAG vs. wt-GATATC.	166
Table 3.8 Predicted effects on the thermodynamics of complex formation: wt-Rp-PMe-2 vs. wt-GATATC.	168
Table 3.9 Solvent accessibility in modified and unmodified EcoRV-DNA complexes.	171
Table 3.10 Predicted effects on the thermodynamics of complex formation: wt-Rp-P _{Me-2} vs. K119A- Rp-P _{Me-2}	173
Table 4.1 Effects of basic side chain mutations on EcoRV-DNA transition state stabilization.	192
Table 4.2 Effects of PMe substitutions on wt EcoRV cleavage.	194
Table 4.3 Effects of phosphate neutralization on wt and mutant EcoRV cleavage rates.	196
Table 4.4 Effects of phosphate neutralization on $\Delta\Delta G_1^{\ddagger}$	199

LIST OF FIGURES

Figure 1.1 Three types of sites for the EcoRV, EcoRI, and BamHI restriction endonucleases.	6
Figure 1.2 Direct readout at the specific EcoRI-DNA interface.....	11
Figure 1.3 Phosphate contacts networked to base functional groups in EcoRV.	12
Figure 1.4 Conformational transitions upon specific EcoRV-DNA complex formation.	16
Figure 1.5 Energetic contributions to specific and non specific EcoRI-DNA complex.	24
Figure 1.6 Reaction pathway for DNA binding and hydrolysis by type II restriction endonucleases.	31
Figure 1.7. Ethylation interference footprints for specific, miscognate, and non-specific restriction endonuclease-DNA-DNA complexes.....	35
Figure 1.8 Two types of DNA bending proteins.....	44
Figure 1.9 DNA bending induced by asymmetric charge neutralization.....	49
Figure 1.10 Chemical structure and stereochemistry of neutral methylphosphonates (P _{Me}).....	52
Figure 1.11 Asymmetric phosphate neutralization induced DNA bending-A tract phasing.	53
Figure 2.1 Phosphate contacts at the specific EcoRV-DNA interface.....	60
Figure 2.2 Three-dimensional arrangement of phosphates at the specific EcoRV-DNA interface.	61
Figure 2.3 Thermodynamic pseudocycle for protein binding to modified and unmodified DNA substrates.....	64

Figure 2.4 Four basic side chain residues at the EcoRV-DNA interface.....	68
Figure 2.5 Intraprotein interactions to the R226 side chain	71
Figure 2.6 Free energy diagram for EcoRV binding to single P _{Me} substituted substrates.	77
Figure 2.7 Free energy diagram for EcoRV mutants binding to modified and P _{Me} substituted substrates.....	84
Figure 2.8 Salt dependence for the formation of modified and unmodified EcoRV-DNA complexes	88
Figure 2.9 Modified phosphate positions in multiple P _{Me} substituted substrates.	91
Figure 2.10 FRET as a measures of DNA bending.	96
Figure 2.11 FRET efficiency as a function of donor-acceptor distance relative to R ₀	99
Figure 2.12 Fluorescence spectra used to calculate ratio _A	101
Figure 2.13 Fluorescence spectra for unmodified FRET substrates.....	107
Figure 3.1 van't Hoff plots for the formation of modified and unmodified EcoRV-DNA complexes.	132
Figure 3.2 Electrostatic modifications and solvation of EcoRV-DNA complexes in preparation for molecular dynamics simulations.	136
Figure 3.3 RMSD for modified and unmodified EcoRV-DNA complexes relative to their starting coordinates.	139
Figure 3.4 Side chain dynamics in minimized and continuous MD simulations.....	142
Figure 3.5 Backbone RMSD for modified and unmodified EcoRV-DNA complexes.....	145
Figure 3.6 Per residue RMSD for K119A and P _{Me-2} modified complexes relative to the unmodified complex.	147
Figure 3.7 Structural consequences of the Rp-P _{Me-2} modification.....	150

Figure 3.8 Structural consequences of the K119A mutation.	152
Figure 3.9 Structural features of the K119A-Rp-P _{Me-2} complex.....	155
Figure 3.10 A119-CH ₃ distances to Rp-P _{Me-2} -CH ₃ in the K119A-Rp-P _{Me-2} complex.....	156
Figure 3.11 Thermodynamic pseudocycle for modified and unmodified complex formation... ..	158
Figure 3.12 Thermodynamic pseudocycle for thermodynamic consequences K119A mutation.	161
Figure 3.13 K119A:Nε to T111:O distance in the free wt EcoRV and in the wt-GATATC complex.....	165
Figure 3.14 Grid-mapping of the EcoRV-DNA complex by Delphi.....	176
Figure 3.15 Electrostatic potentials for modified and unmodified EcoRV-DNA complexes. ..	178
Figure 4.1 Reaction coordinate diagrams for the progression from free substrate to the transitions state.	186
Figure 4.2 D74 and D90 positions in the specific and non-specific EcoRV-DNA complexes. .	188
Figure 4.3 Parallel sequential model for EcoRV-DNA cleavage.	190

ACKNOWLEDGEMENTS

I would like to extend my most sincere gratitude to my advisor, Dr. Linda Jen-Jacobson for her guidance, wisdom, patience, and passion for science. My experience with Linda has been invaluable, not only for my future endeavors in science, but in all that I do.

I would like to thank several collaborators for their generous gifts of materials and time. I would like to thank Dr. Lauren Ernst at the Molecular and Biosensor Imaging Center (MBIC) and Carnegie Mellon University, from whom I learned fluorescence spectroscopy, and whose time and patience are greatly appreciated. I would also like to thank Dr. Richard Hogrefe and Dr. Wojciech Stec for the generous gift of chirally pure methylphosphonate substituted DNA substrates, and Dr. John Perona for the gift of EcoRV mutant protein and plasmids.

I have also received much guidance from members of the Jen-Jacobson laboratory over the years, and I would especially like to thank Mike Kurpiewski, Dr. Lisa Engler, and Dr. Paul Sapienza, who have provided me with helpful discussion, instruction, and mentorship. These individuals have also become good friends and have made the laboratory environment a productive and entertaining place to be.

I would like to thank all of the friends that I have made during my graduate studies. We have enjoyed many laughs, good times, and beers. We have also endured some sadness. I would like to acknowledge Megan Deitz, who passed away in September of 2005. Megan was a bright light and will always be remembered for her hard work and contagious smile.

I would like to thank my family for their support in this achievement especially my father, Richard Hancock, my mother Teresa Hancock and my siblings, Kevin, Daniel, and Sarah. I love you all very much.

Finally I would like to thank my companion, and best friend, Dr. Laura Jane Marinelli. She has been a constant source of motivation, and her strength and positive attitude were invaluable during the process of writing this dissertation. She has been caring and supportive in too many ways to enumerate and for that I am truly grateful.

1.0 CHAPTER 1: INTRODUCTION

Protein-DNA interactions are essential for many diverse biological processes. Importantly, DNA bending induced by DNA binding proteins is required for many of these processes, including regulation of gene expression, DNA packaging, and precise assembly of protein-DNA active sites. DNA bending is also thought to be a structural and energetic determinant for site specificity displayed by many DNA binding proteins. The bending of DNA is an energetically unfavorable process due to the requirements for base destacking, and the compression of phosphates on the concave face of a bent DNA molecule. Thus, DNA bending proteins must overcome the large cost of DNA bending in order to bind to their target DNA sites and facilitate their corresponding biological function. One strategy proposed to facilitate DNA bending is the asymmetric neutralization of phosphate charge on one lateral face of a substrate DNA molecule. This asymmetric charge neutralization will result in relaxation of interphosphate repulsion on one face of the DNA, while the interphosphate repulsion on the opposite face will produce a compressive force facilitating a collapse (bending) of the DNA toward the neutralized face. An understanding of the strategies employed by proteins to facilitate DNA bending is required if we are to understand the structural and energetic role of protein-induced DNA bending in the formation of protein-DNA complexes and in determining DNA site specificity.

In this introductory chapter I will review the current understanding of specificity in protein DNA interactions, discuss how DNA bending is thought to be a determinant in DNA

binding specificity, examine the origin of the energetic stability of the DNA double helix, and provide evidence that DNA bending is facilitated by asymmetric neutralization of the DNA double helix. I will then introduce the EcoRV restriction endonuclease as a model for understanding the structural and energetic role of asymmetric phosphate neutralization in protein induced DNA bending and specificity.

1.1 LEVELS OF SEQUENCE SELECTIVITY FOR PROTEINS THAT INTERACT WITH DNA

DNA binding proteins facilitate a range of diverse biological functions, many of which require that these proteins exhibit stringent discrimination against closely related sites. However, functional diversity requires that DNA binding proteins exhibit various levels of binding specificity in accordance with each specific function. For example, DNA polymerases and histone octamers must interact relatively non-specifically with their target DNA in order to package and replicate a wide range of DNA base sequences in a particular genome. Conversely, transcriptional activators and repressors must display a high degree of sequence specificity when interacting with their target sites in order to efficiently control the regulation of a few operator sites in a large molar excess of similar DNA sites. For example, the *lac* repressor binds to specific operator sites with 10 billion-fold greater affinity than non-operator sites (Jen-Jacobson, 1997). Interestingly bacteriophage λ repressors, cI and cro, display graduated affinity to a series of operator sites that vary in sequence by one or more base pairs (permissive discrimination), while retaining the ability to stringently discriminate between the operator sites and non-specific DNA sites (Harrison and Aggarwal, 1990). The graduated affinity of these transcriptional

regulator proteins for the different operators is carefully tuned to their function as “genetic switches” in the life cycle of the bacteriophage (Albright and Matthews, 1998). Finally, Type II restriction endonucleases, a family of bacterially encoded enzymes that catalyze the hydrolysis of double-stranded DNA, display the extremely stringent discrimination between their specific recognition sites and sites that differ by one or more base pairs. This discrimination is required in order to effectively hydrolyze foreign DNA, while avoiding potentially lethal cleavage of cellular DNA. Specifically, the second order rate constant (which takes into account both binding and cleavage) for the restriction endonuclease EcoRI is as much as 1.2×10^{11} -fold lower for a site with one incorrect base relative to the cognate site (Jen-Jacobson, 1997). This extremely high level of discrimination reflects the fact that restriction endonucleases, unlike non-catalytic DNA binding proteins, exercise discrimination in both binding and cleavage.

1.2 RESTRICTION ENDONUCLEASES AS A MODEL FOR STRINGENT DISCRIMINATION OF DNA BINDING SITES

One of the postulates of the Jen-Jacobson lab has been that by studying the DNA binding proteins that display the highest degree of sequence discrimination, we can understand the largest number of rules and constraints that govern specificity. In addition, it is suggested that the structural and energetic determinants of nonspecific or permissively discriminant protein-DNA interactions can be understood in terms of the relaxation of the factors that govern stringent specificity. For this reason our lab uses three restriction endonucleases, EcoRI, BamHI, and EcoRV, as models to understand the structural and energetic principles that govern sequence specificity in protein-DNA interactions.

1.3 STRUCTURAL AND ENERGETIC DETERMINANTS OF SPECIFICITY

The strategies used by site-specific DNA binding proteins to discriminate between their specific target sequence and a large molar excess of chemically similar sequences has been an area of interest for some time. To this end, many structural and thermodynamic studies have sought to elucidate the molecular strategies that govern the formation of both specific and non-specific protein-DNA complexes. In the following sections the structural and energetic features that distinguish the formation of specific, miscognate (one incorrect base-pair), and non-specific (generally >1 incorrect base-pair) protein-DNA complexes will be discussed (see below for definition of miscognate and non-specific sites). These features are generally relevant to the three type II restriction endonucleases studied in the Jen-Jacobson lab, thus data from all three systems will be used to highlight these features. When appropriate, the EcoRV restriction endonuclease will be used for illustrative purposes, as this is the model that I have used to investigate the role of phosphate neutralization in protein-induced DNA bending.

1.3.1 Three types of binding sites for restriction endonucleases

Restriction endonucleases interact with three types of DNA sites: cognate, miscognate, and non-specific. EcoRI, EcoRV, and BamHI bind as homodimers to their specific recognition sequences GAATTC, GATATC, and GGATCC, respectively. EcoRV hydrolyzes its specific substrate in a blunt end fashion at the center TA base step in each strand, while EcoRI and BamHI make staggered cuts between the first and second base pairs of the recognition sequence, resulting in four base pair “overhangs” at the 3' end of the cleaved substrate (Figure 1.1). Due to the symmetry in these sites, the rates of hydrolysis are generally equivalent in each half-site.

Further, because the rate of complex dissociation is extremely slow, cleavage often occurs in one binding event (Halford and Johnson, 1983). DNA sites that differ from the specific site by only one base pair are known as miscognate sites (historically referred to as EcoRV*, BamHI*, and EcoRI* sites). This base pair change disrupts the palindromic symmetry of the site (Figure 1.1) and leads to the formation of “adaptive” interfaces (see below) which result in asymmetric first order cleavage rate constants for the two DNA strands; a faster rate is observed for the cognate half-site). Finally, EcoRI and BamHI endonucleases bind to sites that differ from the specific recognition sequence at two or more base pairs (non-specific sites), albeit with much lower affinity than to specific sites. These non-specific sites are completely resistant to cleavage (Lesser et al., 1990). Interestingly, the EcoRV restriction endonuclease binds to and cleaves the GAATTC site, which differs from the EcoRV specific recognition site, GATATC, at *both* central base pairs, albeit with 10^4 -fold lower catalytic efficiency (Hiller et al., 2005). Even though the GAATTC site contains 2 incorrect bp, the fact that it can be cleaved means that it is a miscognate (not a nonspecific) site. It is not clear whether simultaneous changes of two base-pairs in the outer positions of the GATATC site would also produce cleavable substrates (i.e. miscognate rather than nonspecific sites). In the following, “non-cognate” site will refer to non-specific sites, miscognate sites, and sites containing base analog substitutions.

EcoRV	<p style="text-align: center;">Cognate Site</p> $\begin{array}{c} k_1 \\ \downarrow \\ 5' \text{--} \text{GAT} \text{ATC} \text{--} 3' \\ \uparrow \\ 3' \text{--} \text{CTA} \text{TAG} \text{--} 5' \\ k_2 \end{array}$	<p style="text-align: center;">Miscognate Site</p> $\begin{array}{c} k_1 \\ \downarrow \\ 5' \text{--} \text{GTT} \text{ATC} \text{--} 3' \\ \uparrow \\ 3' \text{--} \text{CAA} \text{TAG} \text{--} 5' \\ k_2 \end{array}$	<p style="text-align: center;">Non-specific</p> $\begin{array}{c} 5' \text{--} \text{CTATAG} \text{--} 3' \\ 3' \text{--} \text{GATATC} \text{--} 5' \end{array}$
EcoRI	<p style="text-align: center;">Cognate Site</p> $\begin{array}{c} k_1 \\ \downarrow \\ 5' \text{--} \text{GAA} \text{TTC} \text{--} 3' \\ \uparrow \\ 3' \text{--} \text{CTT} \text{AAG} \text{--} 5' \\ k_2 \end{array}$	<p style="text-align: center;">Miscognate Site</p> $\begin{array}{c} k_1 \\ \downarrow \\ 5' \text{--} \text{GAA} \text{TTT} \text{--} 3' \\ \uparrow \\ 3' \text{--} \text{CTT} \text{AAA} \text{--} 5' \\ k_2 \end{array}$	<p style="text-align: center;">Non-specific</p> $\begin{array}{c} 5' \text{--} \text{CTTAAG} \text{--} 3' \\ 3' \text{--} \text{GAATTC} \text{--} 5' \end{array}$
BamHI	<p style="text-align: center;">Cognate Site</p> $\begin{array}{c} k_1 \\ \downarrow \\ 5' \text{--} \text{GGATCC} \text{--} 3' \\ \uparrow \\ 3' \text{--} \text{CCTAGG} \text{--} 5' \\ k_2 \end{array}$	<p style="text-align: center;">Miscognate Site</p> $\begin{array}{c} k_1 \\ \downarrow \\ 5' \text{--} \text{GGTTCC} \text{--} 3' \\ \uparrow \\ 3' \text{--} \text{CCAAGG} \text{--} 5' \\ k_2 \end{array}$	<p style="text-align: center;">Nons-pecific</p> $\begin{array}{c} 5' \text{--} \text{CCTAGG} \text{--} 3' \\ 3' \text{--} \text{GGATCC} \text{--} 5' \end{array}$

Figure 1.1 Three types of sites for the EcoRV, EcoRI, and BamHI restriction endonucleases.

The cognate sites for EcoRV, EcoRI and BamHI are shown in the first column. Red triangles represent the axis of rotational symmetry. An example of one out of nine possible asymmetric miscognate sites for each enzyme is shown in column two. Note that each miscognate site contains a cognate half-site (i.e. for EcoRV $\frac{GAT}{CTA}$) and a miscognate half-site ($\frac{GTT}{CAA}$). The constants k_1 and k_2 refer to the rates of cleavage for the cognate and miscognate half-sites, respectively. A non-specific site is shown in panel C; non-specific sites are resistant to cleavage.

1.3.2 Structural features of specific protein-DNA complexes

In order to understand how site-specific DNA-binding proteins distinguish their specific recognition sequence from non-cognate sequences, it is helpful to consider the structural features that govern the formation of the specific complex. To this end, many X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy experiments have elucidated the structures of DNA binding proteins both free in solution and in complex with their specific DNA substrates. Specifically, for the ~3600 known restriction endonucleases that have been identified, the field has solved 114 crystal structures. 73 of these structures are of 38 distinct enzymes in complex with specific DNA substrates, while only 4 structures exist for non-cognate complexes. Interestingly 2 of these 4 non-cognate structures are for EcoRV (Berman et al., 2000). The following sections will summarize the structural features that characterize specific protein-DNA complexes and describe how they are distinct from those that characterize non-specific complexes.

1.3.2.1 “Direct readout” and the structural determinants of specificity

Structures of specific protein-DNA complexes contributed to our view of the structural features that underlie the formation of these complexes. Specific protein-DNA complexes are characterized by exquisite steric and electrostatic complementarity with the edges of DNA bases and with the sugar-phosphate backbone. Early views of sequence discrimination by DNA binding proteins focused on “direct readout” of the sequence specific arrangement of functional groups in the major groove of the DNA double helix. Indeed, a number of hydrogen bonds and van der Waals interactions are made to the bases (Aggarwal et al., 1988; Harrison and Aggarwal, 1990; Pabo et al., 1990; Steitz, 1990; Rosenberg, 1991; Kim and Burley, 1994) primarily in the

major groove but also in the minor groove (Kim and Burley, 1994) of the specific DNA substrate. Protein-base interactions can be made by constrained elements of the protein, such as the protein backbone or short polar side chains, as well as by longer more flexible side chains. These flexible side chains are often held in a precise position by hydrogen bonds and van der Waals interactions to nearby side chains (“buttressing”), (Harrison and Aggarwal, 1990) or due to contacts made to two functional groups of the same base (“bidentate” recognition) (Seeman et al., 1976; Pabo et al., 1990) or adjacent bases (“bridging”) (Harrison and Aggarwal, 1990; Rosenberg, 1991) (Figure 1.2). The coordination of these interactions suggests that protein-base interactions are not made in isolation, but rather are part of cooperative networks of contacts that require the correct juxtaposition of protein and DNA functional groups. Waters bound at specific protein-DNA interfaces can mediate contacts to DNA-bases and can be part of these cooperative networks (Otwinowski et al., 1988; Gehring et al., 1994; Newman et al., 1995; Wolberger, 1996; Schwabe, 1997) (Figure 1.2). While the “direct readout” of the sequence specific arrangement of base functional groups is thought to play a key role in distinguishing between cognate and non-cognate DNA sites, it is clear that it is not the only structural feature important for sequence discrimination.

1.3.2.2 “Indirect readout” and DNA deformation as a determinant of specificity

Unlike interactions with the functional groups of the DNA bases, contacts to the phosphate backbone were at one time thought to contribute only non-specifically to protein-DNA interactions, as DNA phosphates are a common element of all DNA sites. However, in many specific protein-DNA complexes, key phosphates are tightly constrained by protein moieties that interact with additional phosphate and base functional groups simultaneously (Figure 1.2) or are networked to protein side chains that make other phosphate or base contacts (Figures 1.3). It is

thought that these phosphate contacts play a key role in specific protein-DNA recognition. Indeed, particular phosphates in the specific EcoRI, EcoRV, and BamHI restriction endonuclease complexes are important for complex formation and are proposed to act as “clamps”, which help to position recognition elements and may stabilize distortions in the DNA (Lesser et al., 1990; Engler, 1997; Engler et al., 2001). Ethylation interference data show no localized footprint for non-specific protein-DNA complexes suggesting that these “clamp” contacts are absent in the non-specific complexes (Figure 1.7), and the precise positioning of key recognition elements is not achieved.

Many specific DNA binding proteins distort their target DNA from the energetically optimized B-form upon complex formation (Kim et al., 1990; Schultz et al., 1991; Winkler, 1993; Kim and Burley, 1994; Dickerson and Chiu, 1997). Crystallographic and solution studies suggest that both free DNA structure and the penchant for DNA distortion are a function of sequence (Dickerson and Drew, 1981; Dickerson and Chiu, 1997; Olson et al., 1998). It has become generally recognized that sequence dependent conformational properties of the DNA double helix play an important role in site specific recognition by DNA binding proteins (Harrison and Aggarwal, 1990; Lesser et al., 1990; Steitz, 1990). Detection of this sequence dependent conformational information by a specific DNA binding protein comprises the “indirect readout” of DNA sequence information (Otwinowski et al., 1988). “Indirect readout” has been implicated as a factor important for sequence discrimination based on the idea that the sequence specific conformational properties of a particular DNA site invariably influence the geometric juxtaposition of phosphate and base functional groups relative to other DNA sites. Recognition of these spatial relationships (or the ability to confer a certain spatial relationship) by sequence-specific binding proteins and the subsequent formation of extensive cooperative

networks of contacts to DNA bases and phosphate groups suggests an intimate link between “direct” and “indirect readout”.

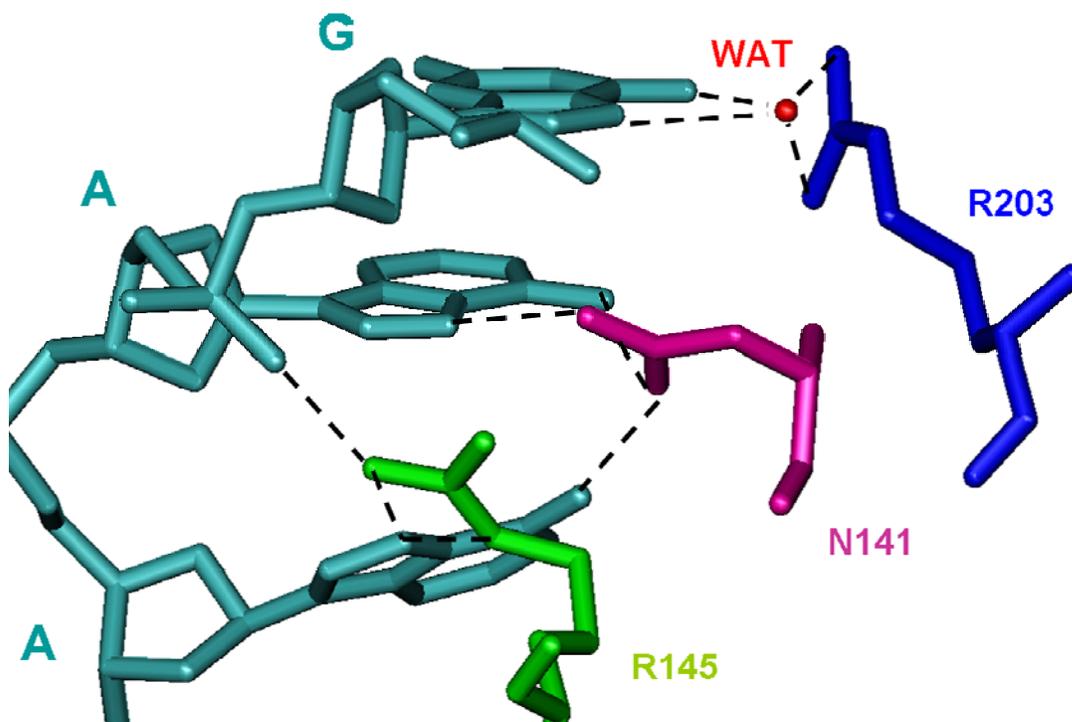


Figure 1.2 Direct readout at the specific EcoRI-DNA interface.

One strand of the EcoRI specific recognition half-site is shown in cyan. The “kink” in the EcoRI bound protein is observed as the A/A base step. Hydrogen bonds are represented as dotted lines. R200 (blue) makes a water-mediated (red) contact to the N7 and O6 of the first base of the specific recognition sequence GATATC. N141 (pink) forms a “bidentate” interaction with the N7 and N6 at the GATATC base, while “bridging” this base to the GAATTC base (hydrogen bonds to N6). R145 (green) bridges the specific base GAATTC to the phosphate at GpA.

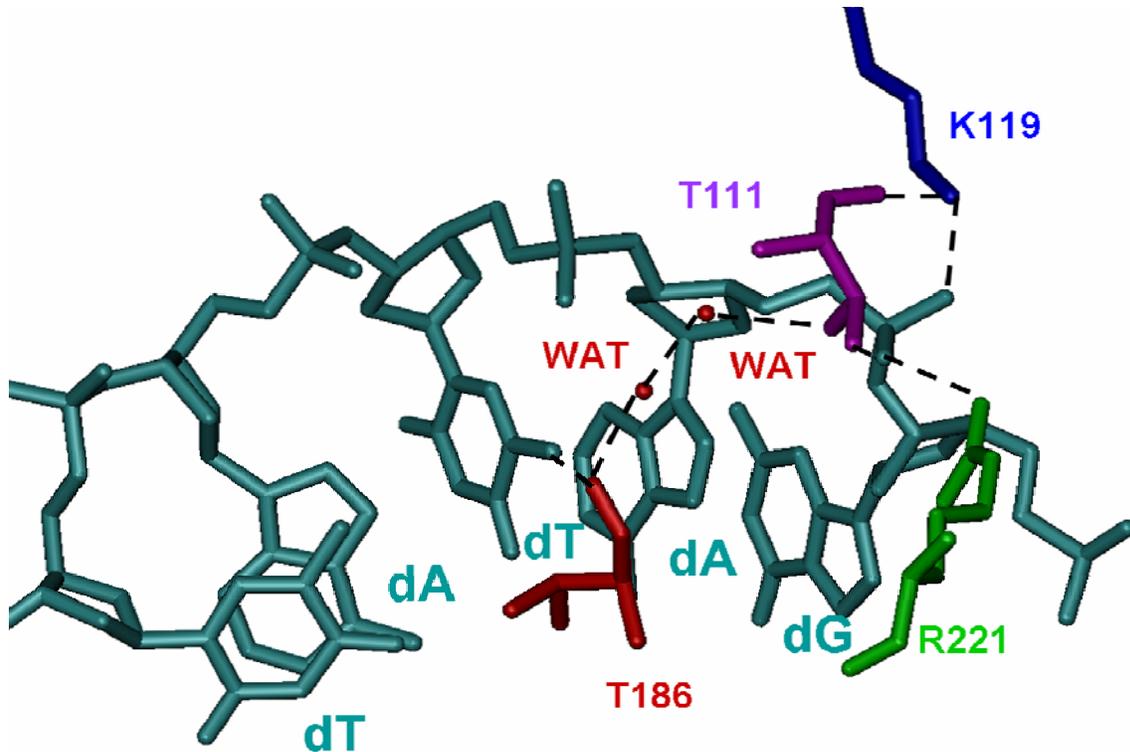


Figure 1.3 Phosphate contacts networked to base functional groups in EcoRV.

One strand of the EcoRV specific half-site (**GATATC**) is shown (cyan). The phosphate at G_pAT is contacted by K119 (blue), R221 (green), T111 (pink), as well as by one stably bound water (not shown). T111 is networked to T186 (red) by several stably bound waters; T189 makes a base specific interaction with the GAT base of the specific half-site. The correct geometric juxtaposition of all of these atoms is required for cooperative formation of these contacts.

1.3.2.3 Protein conformational transitions in specific protein-DNA complexes

Site-specific DNA binding proteins often undergo conformational rearrangements upon binding to their specific DNA targets. These rearrangements seem to play a key role in the precise positioning of recognition elements and can include DNA binding induced movement of domains, subunits, or secondary structural elements (Brennan et al., 1990; Perona and Martin, 1997; Horton and Perona, 2000), as well as local ordering of disordered regions of the free protein. Structures of free site site-specific binding proteins reveal that regions of the protein are disordered. Many of these disordered regions become ordered upon DNA binding. This has been observed for many site-specific binding proteins including EcoRV (Winkler, 1993; Perona and Martin, 1997), EcoRI (Grigorescu, 2003), GCN4 (Weiss et al., 1990), LEF-1 (Love et al., 2004), and the yeast ADR1 DNA binding domain (Hyre and Klevit, 1998).

EcoRV undergoes both adjustments in subunit arrangements as well as local disorder to order transitions upon binding to its specific recognition sequence GATATC (Winkler, 1993; Perona and Martin, 1997; Horton and Perona, 2000). Comparison of the free and specific DNA bound crystal structures of EcoRV reveals that there is an approximately 25° rotation of the two DNA binding domains (DBD) with respect to one another, accompanied by the correlated movement of two minor groove α -helices (B-Helices; Figure 1.4A). This comparison also indicates that a large C-terminal section of the protein (residues 218 to 245) and a surface loop (residues 183-187) both become ordered upon DNA binding (Figure 1.4A). The surface loop contains all of the residues that make base contacts to the specific DNA substrate and is referred to as to recognition loop or R-loop (Figure 1.4A). The flexibility observed in the free EcoRV protein is thought to be important for specific site localization by means of facilitated diffusion along non-specific DNA, allowing the protein to respond to the gross structural features of the

double helix at non-cognate sites until the specific recognition sequence is encountered, at which time a smooth transition to the specific complex can be envisaged (Winkler, 1993). Further, the flexibility in the quaternary structure and in the R-loop is necessary for sufficient opening of the DNA binding cleft during the initial association of protein and DNA, as the direct docking DNA into the binding cleft without this flexibility would be impossible (Winkler, 1993).

Local folding of the C-terminal region is thought to be initiated by the formation of intraprotein contacts within a mobile surface loop (residues 221-228) contained within this region, that directly interacts with the DNA backbone via residues R221, S223 and R226 (Perona and Martin, 1997). Once stabilized, this loop is suggested to trigger the folding of the remaining C-terminal region through stabilizing interactions with adjacent segments including, residues 229-232 and 218-220. Residues in two minor groove binding loops also become ordered upon DNA binding, they are the Q loop (residues 67 to 70) and the linker II loop (residues 35 to 39) (Figure 1.4A) (Perona and Martin, 1997). Q loop residues Q68 and Q69 contact the phosphates 3' of the specific recognition site (GATATC**pTp**), and N70 contacts the minor groove edge of recognition site cytosine (GATATC). Linker II residues T37 and K38 directly contact the phosphate at GATAT**pC** and the sugar moiety of the 3' most adenosine in the recognition sequence (GATATC), respectively. Finally, residues 183 to 187 become ordered upon formation of the specific complex. Residues in the R-loop that become ordered upon specific complex formation include the protein backbone of G184, which contacts the guanosine in the specific recognition site (GATATC), the N185 side chain, which contacts the guanosine and the 5' most adenine (GATATC), and finally the T186 side chain, which contacts the 5' most thymine of the specific recognition sequence (GATATC).

Coupling of structural transitions to site-specific binding can contribute to specificity if these transitions occur only when encountering cognate DNA sites. Free energy associated with the formation of a highly complementary protein-DNA interface is thought to drive the folding of protein regions that make up key components of this interface. (Spolar and Record Jr, 1994), as seen with R-loop and Q-loop residues which make key contacts to the phosphate backbone in the specific complex. While the favorable binding free energy released from the formation of the specific complex may drive these local folding transitions, non-cognate interactions likely do not provide the necessary free energy to pay the entropic cost associated with these transitions. Indeed, segments of the EcoRV restriction endonuclease that become ordered upon specific complex formation remain disordered upon formation of the non-specific complex (Winkler, 1993) (Figure 1.4A; see below).

One key to an enzyme achieving functional specificity is the coupling of specific site recognition to catalysis, such that efficient catalysis occurs only when the protein-DNA interface is correctly assembled. Correct assembly of the specific EcoRV-DNA interface is required for the precise positioning of catalytic residues and efficient hydrolysis of specific DNA sites. Taken together it appears that correct formation of the specific EcoRV-DNA recognition interface, which includes severe DNA distortion and a coordinated conformational transition, is intimately linked to efficient catalysis.

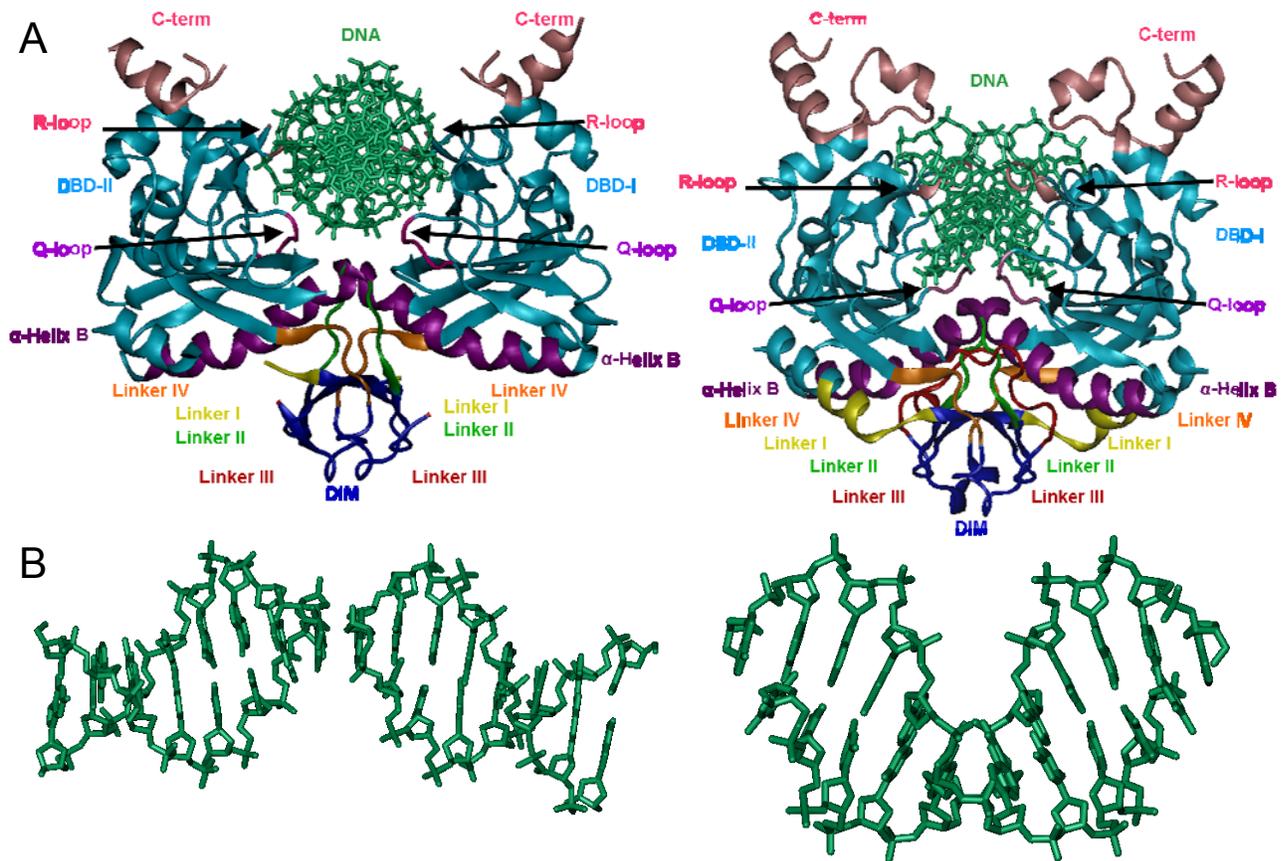


Figure 1.4 Conformational transitions upon specific EcoRV-DNA complex formation.

(A) Crystal models of the non-specific (left) and specific (right) EcoRV complexes are shown. The C-terminal domain (residues 218 to 245; pink) and the recognition loops (R-loop; residues 183 to 187; pink) are disordered in the unbound protein molecule and become ordered upon binding to the specific DNA site (light green). These regions are missing in the non-specific complex. The two DNA binding domains (cyan), dimerization domain (blue), Q-loops (magenta), α -helix B (purple), and linkers I (yellow), II (green), III (red), and IV (orange) are also shown. (B) DNA from the non-specific (left) and specific (right) complexes are shown; The DNA is rotated 90° about the y-axis relative to its orientation in the corresponding EcoRV-DNA complex.

1.3.3 Structural features of non-cognate protein-DNA complexes

The structures of specific protein-DNA complexes have elucidated some general features associated with these complexes; however, a complete structural understanding of site discrimination by DNA binding proteins also requires structural data for non-cognate complexes. Only four structures of type II restriction endonucleases bound to non-cognate DNA sites (Winkler, 1993; Viadiu and Aggarwal, 2000; Hiller et al., 2005; Townson et al., 2007) have been solved. In addition to the non-cognate structures for type II restriction endonucleases mentioned above, three other non-cognate complexes for site-specific DNA binding proteins have been solved. These are the glucocorticoid receptor DNA binding domain bound to a non-cognate DNA element (Luisi et al., 1991), the monomeric bacteriophage λ Cro repressor bound to a non-consensus half operator (Albright et al., 1998) and the dimeric DNA binding domain of *lac* repressor with non-specific DNA, solved by nuclear magnetic resonance (NMR) techniques (Kalodimos et al., 2004). An examination of these complexes, along with those for the non-cognate type II restriction endonuclease complexes, reveals that non-cognate complexes do not share the structural features that characterize specific complexes.

1.3.3.1 The non-specific complex is a 'loose' complex.

Extensive geometrically constrained networks of hydrogen bonds and van der Waals interactions to the DNA bases are observed in specific protein-DNA complexes (Aggarwal et al., 1988; Harrison and Aggarwal, 1990; Pabo et al., 1990; Steitz, 1990; Rosenberg, 1991; Kim and Burley, 1994); These networks are absent from the complexes that sequence-specific proteins make with non-cognate sites. (Luisi, Xu et al. 1991; Albright, Mossing et al. 1998; Kalodimos, Biris et al. 2004; Winkler 1993; Viadiu and Aggarwal 2000). Also absent from non-cognate complexes are

the geometrically constrained, cooperative “clamp” interactions to DNA phosphates, which have been shown to be important for the formation of the specific EcoRI, EcoRV, and BamHI complexes (Lesser et al., 1990; Engler, 1997; Engler et al., 2001). Instead, these complexes are characterized by more flexible non-specific (Coulombic) interactions with the DNA phosphates, often mediated by water. This looser mode of binding is thought to play a role in site localization by site-specific DNA binding proteins through a process by which they translocate along the linear sequence of non-specific DNA until they encounter their own specific recognition sequence (Von Hippel and Berg, 1989; Shimamoto, 1999). Indeed, it has been observed that, in the presence of non-specific DNA, site-specific DNA binding proteins locate their specific site much faster than the rate permitted by free diffusion alone, suggesting a facilitated diffusion (Jack et al., 1982; Terry et al., 1983; Ehbrecht et al., 1985; Halford and Marko, 2004a)

The loose association between site-specific DNA binding proteins and non-cognate sites is characterized by reduced steric complementarity and by reduced burial of solvent accessible surface area (ASA) relative to that observed for the specific complex (Luisi, Xu et al. 1991; Albright, Mossing et al. 1998; Kalodimos, Biris et al. 2004; Winkler 1993; Viadiu and Aggarwal 2000). For example, the total surface area buried upon specific EcoRV-DNA complex formation is $>1800 \text{ \AA}^2$ larger than that for the formation of the non-specific complex (Winkler, 1993). Similarly, 4900 \AA^2 of accessible surface area is buried for the specific *lac* repressor interaction with its specific operator site *O_I*, compared to the 3500 \AA^2 buried in the non-specific complex (Kalodimos et al., 2004).

1.3.3.2 DNA distortion in non-cognate protein-DNA complexes.

Protein-induced DNA distortion is a common feature of specific protein-DNA complexes. In contrast, very little DNA distortion is observed in non-specific complexes (Winkler, 1993; Albright et al., 1998; Kalodimos et al., 2004), while there can be a range of DNA distortion in miscognate complexes (Hiller et al., 2005). The DNA bound in the specific EcoRV-DNA complex is bent $\sim 55^\circ$ toward the major groove, while no such bend is observed for the DNA in the nonspecific complex, in which the DNA exists in its native B form (Figure 1.4B) (Winkler, 1993). As mentioned above, the nonspecific DNA in this complex is composed of two DNA octamers that abut one another such that the phosphate opposite the EcoRV active site (scissile phosphate) is missing (Winkler, 1993). The absence of the scissile phosphate in the non-specific complex does not seem to influence DNA bending (or lack thereof), as a specific complex in which the scissile phosphate is absent *does* show significant DNA bending resembling that of the specific EcoRV-DNA complex with a continuous DNA substrate (Horton and Perona, 1998b). There is also no DNA bending observed in the nonspecific *lac* repressor or λ Cro repressor-DNA complexes, while DNA bending is observed when these proteins are bound to their specific operator sites (Albright et al., 1998; Kalodimos et al., 2004). These observations implicate DNA distortion as a structural determinant for sequence discrimination by site-specific DNA binding proteins. The energetic requirements for DNA bending and their implications for the formation of specific protein-DNA interactions will be discussed below (Section 1.4).

Interestingly, significant DNA bending is observed for two non-cognate EcoRV-DNA complexes. Specifically, structural and Förster resonance energy transfer (FRET) studies of EcoRV bound to the non-specific GAATTC site show significant DNA bending relative to the specific complex which was $\sim 60\%$ of that observed on the specific complex (Hiller et al., 2005).

A steric clash between the exocyclic methyl groups of the two GAATTC bases restrict DNA bending in this complex. The EcoRV-GAAUTC complex, in which these two exocyclic methyl groups are absent, shows the full extent of DNA bending relative to the specific complex (Hiller et al., 2005). The observation that EcoRV can induce bending at these non-cognate sites is consistent with thermodynamic data that suggest that there is considerable complementarity at the miscognate interfaces which is distinct from the loose non-specific interface (Engler et al., 1997).

1.3.3.3 Conformational transitions and dynamics in non-cognate complex formation

The DNA binding induced conformational transitions that are observed with specific protein-DNA complex formation are generally not observed, or are less dramatic for the formation of non-cognate complexes. For example, the local folding transitions that accompany the formation of specific EcoRV-DNA interactions (Section 1.3.2.3) are not observed upon formation of the non-specific EcoRV-DNA complex. The R-loops and the C-terminal domain of the EcoRV restriction endonuclease, which become ordered upon formation of the specific EcoRV-DNA complex, remain disordered in the non-specific complex (Winkler, 1993) (Figure 1.4A), suggesting that the interaction with non cognate DNA sites does not provide sufficient binding free energy to pay the entropic cost associated with the folding of these regions.

In addition to variation in the conformational transitions observed for specific vs. nonspecific protein-DNA interactions, it has also been shown that there is variation in protein dynamics observed in these two types of complexes. Nuclear magnetic resonance (NMR) experiments suggest that *lac* repressor DNA binding domain becomes more rigid when in complex with its specific operator site relative to the free dimer; however, upon formation of the non-specific complex, *lac* repressor remains as dynamic as in the free state (Kalodimos et al.,

2004). These data agree with results from molecular dynamics (MD) simulations which show that the motions of the protein and DNA are more dynamic in the non-specific EcoRV and EcoRI-DNA complexes relative to the specific complexes (Duan et al., 1996)

Conformational transitions and restriction of dynamic motions distinguish specific from non-specific binding. During specific site location, the association between the protein and non-specific DNA sites remains loose and dynamic however, when sequence specific cues are encountered, such as the precise spatial arrangement of DNA functional groups or a high degree of DNA helical deformability, the protein switches to a specific binding mode, in which conformational transition are triggered, and water is squeezed out of the complex, and the specific, highly complementary protein-DNA interface is assembled resulting in the restriction of dynamic flexibility.

1.3.4 Energetic features of specific protein-DNA interactions

From an energetic perspective, site-specific DNA-binding proteins distinguish their specific recognition sites from closely related sites because the free energy change (ΔG^0) for the formation of the specific complex is more favorable than that for the formation of non-cognate complexes. Thermodynamic characterization of the factors that contribute to the free energy of specific and non-specific complex formation have been undertaken by many groups and, although a complete thermodynamic analysis is lacking, we can gain some insight by considering some extensively studied systems. These include the *lac* (Record Jr et al., 1976; Mossing and Record Jr, 1985; Record Jr et al., 1991; Frank et al., 1997), *trp* (Jin et al., 1993; Ladbury et al., 1994), λ cI (Sauer et al., 1990; Senear and Ackers, 1990; Strahs and Brenowitz, 1994; Merabet and Ackers, 1995), λ Cro (Takeda et al., 1989; Takeda et al., 1992), Arc (Brown and Sauer,

1993; Smith and Sauer, 1995), and Mnt (Waldburger and Sauer, 1995) repressors and the EcoRI (Terry et al., 1983; Lesser et al., 1990; Lesser et al., 1993; Jen-Jacobson, 1995), BamHI (Engler et al., 2001), and EcoRV (Engler et al., 1997) restriction endonucleases. Although there are several significant differences between these systems some general principles do emerge. One of these principles is that, for specific complex formation, binding free energy ($\Delta G^{\circ}_{\text{bind}}$) is the net of opposing, favorable and unfavorable factors, though the magnitudes of the contribution from each of these factors may vary from system to system (Figure 1.5) (Jen-Jacobson, 1997). Favorable contributions derive from the formation of an extraordinarily intimate interface involving, hydrogen bonds and van der Waals interactions with DNA bases (Lesser et al., 1990; Lesser et al., 1993), hydrogen bonds and Coulombic interactions with phosphate groups (Johnson and Benkovic, 1990), and the release of bound water from non polar surfaces of the DNA and protein molecules (hydrophobic effect) (Ha et al., 1989b; Spolar and Record Jr, 1994). These favorable contributions are balanced by unfavorable factors, which include desolvation of polar surfaces, loss of translational and rotational freedom for free protein and DNA molecules (Finkelstein and Janin, 1989; Spolar and Record Jr, 1994), induced folding/ordering of protein segments upon DNA binding, restriction of the conformational-vibrational freedom of protein and DNA elements in the complex (Sturtevant, 1977b; Duan et al., 1996; Jen-Jacobson, 2000a), and restriction of tightly bound water molecules that become trapped at the interface (Ladbury et al., 1994; Morton and Ladbury, 1996), DNA distortion (Kumar et al., 1994; Jen-Jacobson, 2000a), and electrostatic repulsion at active sites (Engler et al., 2001).

Non-specific protein-DNA complexes differ in that there are few or no contacts to DNA bases (Luisi et al., 1991; Winkler, 1993; Kalodimos et al., 2004). Fewer waters are released from both polar and nonpolar surfaces, and there are no significant DNA distortions or coupled

folding reactions that accompany DNA binding. Thus, the energetic factors that contribute to the formation of non-specific protein-DNA complexes derive primarily from favorable Coulombic interactions with the protein backbone, and are opposed by the cost of restricting the rotational and translational freedoms of both protein and DNA molecules. Estimates have been made as to the magnitude of the energetic contributions to specific and non-specific complex formation using the EcoRI system as an example (Figure 1.5) (Jen-Jacobson, 1997).

Formation of specific protein-DNA complexes is also accompanied by a large negative heat capacity change (ΔC_p^0) (Ha et al., 1989a; Ladbury et al., 1994; Hyre and Spicer, 1995; Merabet and Ackers, 1995; Petri et al., 1995; Berger et al., 1996; Frank et al., 1997; Jen-Jacobson, 2000a). Contributions to the large negative ΔC_p^0 include the release of water from nonpolar surfaces of both protein and DNA molecules (Sturtevant, 1977a; Murphy and Freire, 1992; Spolar et al., 1992; Spolar and Record Jr, 1994; Makhatadze and Privalov, 1995), immobilization of water molecules that become trapped at the protein-DNA interface (Ladbury et al., 1994; Conelly, 1997; Jen-Jacobson, 2000a), and the restriction of conformational-vibrational freedom of protein and DNA functional groups (Duan et al., 1996; Jen-Jacobson, 2000a), and linked equilibria such as protonation, ion binding and induced folding. All of these features are consistent with a highly immobilized and intimate interface. Conversely the ΔC_p^0 associated with the formation of non-specific protein-DNA complexes is ~ 0 (Takeda et al., 1992; Ladbury et al., 1994; Merabet and Ackers, 1995; Berger et al., 1996; Frank et al., 1997; Jen-Jacobson, 2000a), suggesting that formation of the non-specific complex is not accompanied by release of a substantial amount of water from nonpolar surfaces, immobilization of interfacial waters, or significant restriction of vibrational and configurational freedoms of protein and DNA side chains and functional groups.

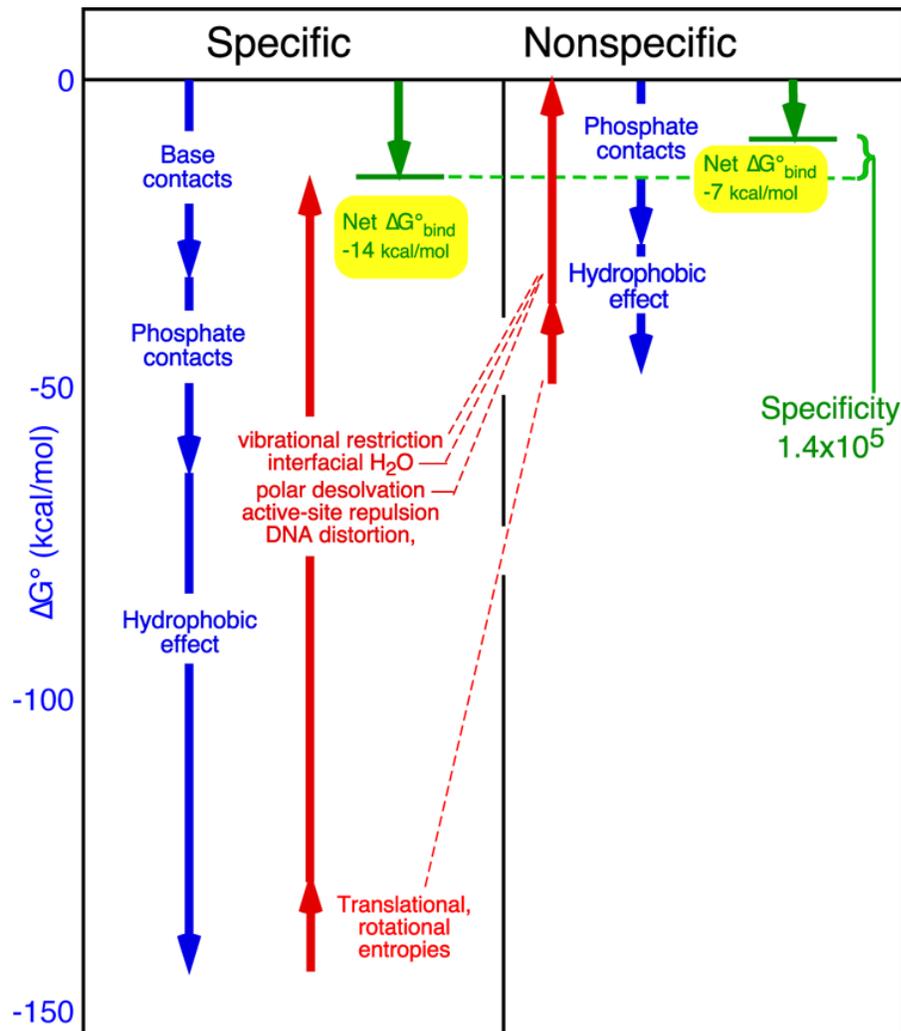


Figure 1.5 Energetic contributions to specific and non specific EcoRI-DNA complex.

Energetic components of the specific and non-specific EcoRI-DNA interactions are shown. The components of the interactions are applicable to protein-DNA interactions in general, although the magnitudes of these contributions certainly vary. Overall, binding free energy (ΔG°_{bind} ; green arrow) is the net of favorable (blue) and unfavorable (red) contributions to protein-DNA complex formation. Non-specific complex formation is largely Coulombic with virtually no contributions from base contacts or water release. Adapted from Jen-Jacobson *et al.* (1997)

1.3.5 Thermodynamic and kinetic basis for sequence discrimination by type II restriction endonucleases

The structural and energetic features that distinguish the formation of specific and non-specific protein-DNA complexes can be thought of as molecular strategies that facilitate the stringent discrimination displayed by type II restriction endonucleases. In order to successfully carry out their biological function of inactivating foreign DNA molecules, Type II restriction endonucleases must bind to and cleave both strands of their specific target DNA. Cleavage in only one DNA strand results in single strand nicks, which are rapidly repaired by DNA ligase *in vivo* (Heitman et al., 1989). Further, they must make these double strand cuts only at their specific recognition site while avoiding potentially lethal cleavage events at closely related sites (5/6 bp match) that are left unprotected by their accompanying methyltransferase. In order to meet these functional requirements, many type II restriction endonucleases bind as homodimers to two distinct half-sites in a palindromic recognition site. Hydrolysis occurs in two symmetrical active sites formed from elements of both protein subunits such that the rate of catalysis in each half-site depends on the correct assembly of the recognition interface in both half-sites. In the following sections I will describe the thermodynamic and kinetic features of specific, miscognate, and non-specific restriction endonuclease-DNA interactions which lead to the stringent discrimination that these enzymes display against non-cognate sites.

1.3.5.1 Sequence discrimination in binding

Type II restriction endonuclease complexes display stringent discrimination between their specific recognition sites and non-specific sites in the binding step of the overall reaction (Figure

1.6A). Solution studies have shown that in the absence of divalent metal, EcoRI binds to its specific recognition sequence (GAATTC) 4.3×10^3 to 2.6×10^6 -fold better than to an inverted site (CTTAAG), depending on flanking context (Jen-Jacobson, 1997). This corresponds to a difference in binding free energy change ($\Delta\Delta G_{\text{bind}}^{\circ}$) for the non-specific site of approximately +5 to +8.7 kcal/mol relative to the specific site. Similarly, non-specific site discrimination by BamHI ranges from 1.0×10^2 to 3.2×10^4 -fold ($\Delta\Delta G_{\text{bind}}^{\circ} \approx +2.7$ to +6.1 kcal/mol) and is about 2.8×10^2 -fold for EcoRV at pH 7.3 ($\Delta\Delta G_{\text{bind}}^{\circ} \approx +3$ kcal/mol, depending on flanking sequence) (Engler et al., 1997; Cao, 2002). Importantly, specific binding is dramatically increased in the presence of calcium (Ca^{2+}), while non-specific complex formation is unaffected (Table 1.1).

Interestingly, it had been reported that EcoRV was able to bind to both specific and non-specific sites with equal affinity in the absence of divalent metal (Taylor et al., 1991; Vipond and Halford, 1995), while binding discrimination is dramatically enhanced in the presence of the presence of calcium (Ca^{2+}) or magnesium (Mg^{2+}) (Thielking et al., 1992; Vermote and Halford, 1992; Vipond and Halford, 1995). This led to the proposal that EcoRV represented a new paradigm for site-specific binding proteins in which discrimination is achieved in the catalytic step and not in binding (Vipond and Halford, 1993). However, it was shown by Engler *et al.* (1997) that the inability to detect site specific binding by EcoRV in the absence of divalent metal resulted from suboptimal experimental conditions such as pH, salt concentration, and choice of DNA substrate. By optimizing these parameters Engler *et al.* (1997) have shown that EcoRV can indeed bind specific sites (GATATC) with significantly greater affinity than to non-specific sites (280 fold; see above) in the absence of Ca^{2+} .

Binding discrimination by EcoRV in the absence of divalent metal is weaker than that observed for EcoRI and BamHI. This difference may reflect the relative magnitudes of the

unfavorable energetic factors that must be overcome in order to achieve specific complex formation for each of these three proteins relative to those required for non-specific binding. For example, formation of the specific EcoRV-DNA complex requires dramatic DNA bending which is not required for the formation of the non-specific complex (Winkler, 1993) (see Figure 1.4). DNA bending is an energetically costly process which contributes unfavorable free energy to specific EcoRV-DNA complex formation. Because DNA bending is not required for the formation of the non-specific EcoRV-DNA complex there is no requirement to pay the energetic cost for DNA bending. Thus, the difference in the free energy for specific and non-specific EcoRV-DNA complex formation is smaller than that for EcoRI and BamHI. Formation of the specific EcoRI-DNA complex also requires DNA distortion in the form of a base pair kink (Frederick et al., 1984) while formation of the specific BamHI-DNA complex does not require any DNA distortion (Viadiu and Aggarwal, 2000). It may be that the energetic cost required to deform the DNA in the EcoRI-DNA complex is smaller than that required for EcoRV-DNA complex formation.

In addition, formation of each of these three specific complexes requires the apposition of several negatively charged functional groups at the active site, which includes the scissile phosphate and a number of functionally conserved negatively charged side chain residues (see Section 1.3.4). The juxtaposition of these functional groups creates a high affinity binding site for divalent metals. In the absence of divalent metal, unfavorable electrostatic repulsion at the active site must be overcome in order for the correct assembly of the specific recognition interface. A differential energetic requirement for active site repulsion in these three complexes, along with the differential requirement for DNA distortion are consistent with the differential discrimination observed for EcoRV in the absence of divalent metal. Thus the aggregate of these

two factors likely results in a reduced binding free energy for the formation of the specific EcoRV-DNA complex relative to that for the formation of the specific BamHI-DNA and EcoRI-DNA complexes.

Interestingly, specific binding affinity is stimulated in the presence of Ca^{2+} by 700-fold, 375-fold, and 380-fold, for EcoRV, BamHI, and EcoRI respectively, while non-specific binding is unaffected by Ca^{2+} . Additionally, specific binding is also stimulated at low pH, while non-specific binding is unaffected (Engler et al., 1997). The differential effects of Ca^{2+} and pH on specific and non-specific complex formation result in increased discrimination against non-specific sites in the presence of Ca^{2+} and at lower pH for all three restriction endonucleases. Specific binding stimulation by Ca^{2+} is, at least in part, due to the relief of electrostatic repulsion at the active site of all three restriction endonuclease-DNA complexes. At the specific EcoRV-DNA interface, multiple metals per active site (Horton and Perona, 2004) suggesting the formation of multiple metal binding sites, likely resulting in increased electrostatic repulsion. These observations are in support of the idea that increased active site repulsion for the specific EcoRV-DNA complex results in a smaller degree of discrimination displayed by this enzyme against non-specific sites relative to EcoRI and BamHI in the absence of Ca^{2+} .

Type II restriction endonucleases also show stringent discrimination against miscognate sites. Specifically, the reduction in binding affinity for miscognate sites relative to specific sites ranges from 9.0×10^2 to 2.5×10^4 -fold for EcoRI ($\Delta\Delta G_{\text{bind}}^0 \approx +4$ to $+6$ kcal/mol) (Lesser et al., 1990; Cao, 2002), from 3.0×10^2 to 2×10^3 -fold for BamHI ($\Delta\Delta G_{\text{bind}}^0 \approx +3.4$ to $+4.5$) (Engler et al., 2001; Cao, 2002), and from 2.0×10^2 to 2.4×10^2 fold for EcoRV (Cao, 2002). As will be shown in Section 1.3.6, factors such as Ca^{2+} and pH have differential effects on the formation of

specific, miscognate, and non-specific complexes, thus discrimination against non-cognate sites varies depending on solution conditions.

1.3.5.2 Sequence discrimination in cleavage

As enzymes that catalyze the hydrolysis of double stranded DNA, Type II restriction endonucleases can also discriminate between DNA sites in the catalytic step. First order cleavage rate constants have been measured for EcoRV, BamHI, and EcoRI with both cognate and miscognate sites (Lesser et al., 1990; Alves, 1995; Engler et al., 1997; Engler et al., 2001). Cleavage rates are reduced 70 to 1.7×10^6 -fold for EcoRI, 1.6×10^4 to 2.5×10^7 fold to for BamHI, and 1.0×10^4 to 1.3×10^6 -fold for EcoRV. Additionally, first order cleavage rate constants at miscognate sites are reduced in both the modified and unmodified half-sites (Figure 1.1) relative to those for specific sites suggesting that a single base pair change can have global effects on the endonuclease-DNA complex such that cleavage in the distal half-site is also affected.

Penalties observed for cleavage of miscognate sites by EcoRI are more severe than those associated with subtle perturbations by certain base analog substitutions, which can show very small (if any) reduction in cleavage rate at the modified half-site and no reduction at the cognate half-site (Lesser, Kurpiewski et al. 1990; Jen-Jacobson 1995). Specifically, for EcoRI, base analog substitutions at inner base pair positions of the GAATTC specific recognition site produce subtle effects in binding and cleavage, while substitutions to the outer bp positions (particularly the GAATTC position) show non-additivity in binding affinity leading to large reductions in cleavage rate for both the modified and unmodified half sites. Cleavage by EcoRV is also dramatically affected by base analog substitutions at each of the base pairs in the specific recognition sequence (GATATC). Specifically, substitutions made to only one of the guanine

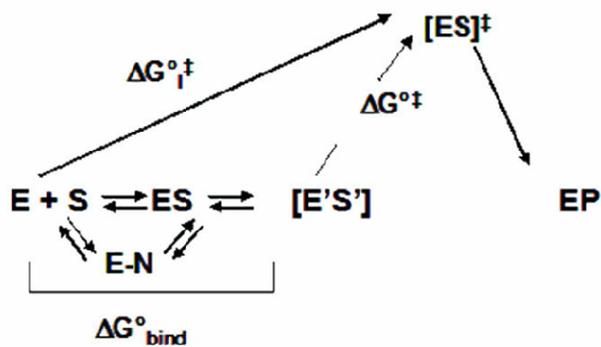
bases of the specific recognition sequence (GATATC) result in significant penalties for cleavage in both the modified, and unmodified half-sites, suggesting that there is “communication” between the two active sites (Parry et al., 2003). Non-specific sites are completely resistant to cleavage for all three endonucleases (Lesser et al., 1990; Engler, 1998; Engler et al., 2001).

Overall discrimination between specific sites and non-cognate sites occurs in both binding and cleavage. A useful metric to describe the overall discrimination displayed by type II restriction endonucleases is the change in transition state interaction free energy ($\Delta G_{I^\ddagger}^0$), described as the probability of going from the free macromolecules to the activated complex in which first strand cleavage occurs (Lesser et al., 1990). $\Delta G_{I^\ddagger}^0$ takes into account both the assembly of the protein-DNA complex (ΔG_{bind}^0) and the activation of this complex (ΔG^\ddagger) to the transition state ($[\text{ES}]^\ddagger$) (Figure 1.6). The difference in interaction free energy change ($\Delta\Delta G_{I^\ddagger}^0$) between cognate and non-cognate (modified) sites can be described as

$$\Delta\Delta G_{I^\ddagger}^0 = \Delta G_{I^\ddagger(\text{mod})}^0 - \Delta G_{I^\ddagger(\text{cognate})}^0, \quad (1.1)$$

where $\Delta G_{I^\ddagger(\text{mod})}^0$ is the transition state interaction free energy for any sequence that differs from the specific recognition sequence and $\Delta G_{I^\ddagger(\text{cognate})}^0$ is that for the specific site.

A



B

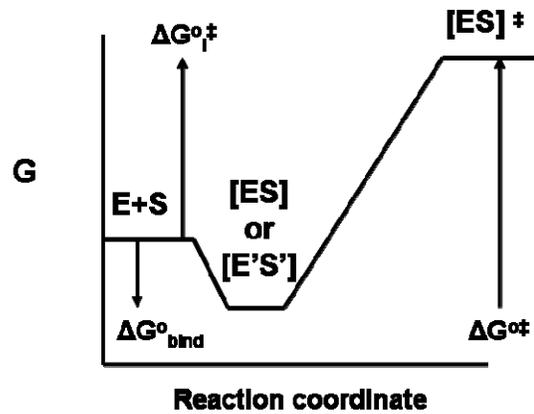


Figure 1.6 Reaction pathway for DNA binding and hydrolysis by type II restriction endonucleases.

(A) Transition state interaction free energy ($\Delta G_{\text{I}}^{\circ \ddagger}$) is represented as the probability of going from the free protein (E) and free DNA (S) to the transition state ($[\text{E}'\text{S}]^{\ddagger}$) and takes into account both the free energy of binding ($\Delta G_{\text{bind}}^{\circ}$) and free energy of transition state formation $[\text{ES}]^{\ddagger}$. Transition state formation ultimately leads to hydrolysis, and the formation of an enzyme-product complex is formed. (B) The components of the reaction pathway in (A) are shown in a free energy diagram.

Lesser *et al.* (1990) have shown that any one incorrect base-pair substitution in the GAATTC site produces a large cost in transition state interaction energy $\Delta\Delta G_1^{\ddagger}$. Penalties for the 9 possible EcoRI* sites range from +6.6 to +13 kcal/mol; these are much more severe than those observed for particular base analog substitutions, which remove only a single hydrogen bond or van der Waals interaction at the interface and result in a $\Delta\Delta G_1^{\ddagger}$ of 0 to +2.5 kcal/mol (Jen-Jacobson, 1995; Jen-Jacobson, 1997). (Note that there are only a few cases where base analog substitutions perturb only a single hydrogen bond, or van der Waals interaction). Thus, the large unfavorable $\Delta\Delta G_1^{\ddagger}$ observed for miscognate sites suggests that discrimination against these sites can not derive from the disruption of protein-base interactions alone, leading to the proposal that stringent discrimination against these sites is an aggregate of factors including, disruption of protein-base interactions ($\Delta\Delta G_{\text{base}}$), disruption of phosphate interactions ($\Delta\Delta G_{\text{phos}}$), and differences in conformational changes, entropic factors, and the cost of driving variant DNA sites to alternate conformations ($\Delta\Delta G_{\text{reorg}}$) (Lesser et al., 1990).

1.3.6 Three different binding modes: pre-transition state, adaptive, and nonspecific complexes

Thermodynamic profiles for the interactions of type II restriction endonucleases with specific, miscognate, and non-specific sites describe three classes of binding modes (Table 1.1). Subtle perturbations to the specific DNA-substrate that disrupt a single interaction to either the DNA bases or phosphates or that modulate the energy required to distort the DNA substrate, result in inhibition of binding ($\Delta\Delta G_{\text{bind}}^{\circ} > 0$) while they do not affect hydrolysis ($\Delta\Delta G^{\ddagger} = 0$) (Jen-Jacobson, 1995; Jen-Jacobson, 1997). This observation suggests that these factors (protein-base, protein-phosphate, and DNA distortability) contribute identically to the formation of the specific

recognition and the transition state complexes, such that the specific recognition complex closely resembles the transition state complex. As a result, binding free energy is efficiently utilized to drive the formation of the transition state complex. Thus, the specific recognition complex is referred to as a “pre-transition state” complex (Jen-Jacobson, 1997). Formation of the pre-transition state complex is characterized by several features, including strong symmetric ethylation interference footprints that highlight the importance of certain phosphates for binding (Figure 1.7) (Lesser et al., 1990; Engler et al., 1997; Jen-Jacobson, 1997; Engler et al., 2001), strong dependence on salt concentration (Engler et al., 1997; Engler et al., 2001) and pH for binding (Engler et al., 1997), binding enhancement by Ca^{2+} (Engler et al., 1997; Engler et al., 2001; Kurpiewski et al., 2004; Sapienza, 2005) which binds at the active site of specific EcoRV, EcoRI, and BamHI complexes but does not support cleavage, a large negative heat capacity change (ΔC_p°) (see below), and symmetric strand cleavage upon progression to the transition state (Table 1).

By contrast, formation of miscognate protein-DNA complexes is characterized by large penalties for binding relative to the specific complex, reduced and asymmetric rates of cleavage, asymmetric ethylation interference footprints (Figure 1.7) (Lesser et al., 1990), reduced dependence on salt concentration and pH (Engler, 1998) and a reduced amount of water release (Cao, 2002), and reduced magnitude in ΔC_p° . Taken together, these data suggest major structural differences in miscognate complexes relative to specific pre-transition state complexes. These complexes are said to be “adaptive” in that they have altered their conformation in various ways in order to optimize binding free energy and as a result contain poorly assembled active sites that are not on the correct path to the transition state. This concept is reflected in the

reduced first order cleavage rate constants observed for miscognate (adaptive) complexes and provides an explanation for discrimination against these sites in both binding and cleavage.

As described above, the structural and energetic features that govern non-specific complex formation are quite distinct from those that contribute to the formation of the pre-transition state complex. The thermodynamic properties associated with the formation of non-specific complexes are in accordance with these distinctions. Unlike specific complexes, non-specific complexes do not display strong ethylation interference footprints (Figure 1.7E) suggesting that the protein molecule is delocalized at these sites and can interact with all phosphates equally. Further, non-specific complexes show an equal or greater dependence on salt concentration relative to pre-transition state complexes highlighting the importance of electrostatic interactions with the phosphate backbone for these interactions. In addition, nonspecific complex formation is not enhanced by the addition of Ca^{2+} , varies much less with pH, and shows a reduced amount of water release. All these observations support the notion that the cleavage-incompetent nonspecific complex is distinct from both pre-transition state and adaptive miscognate complexes.

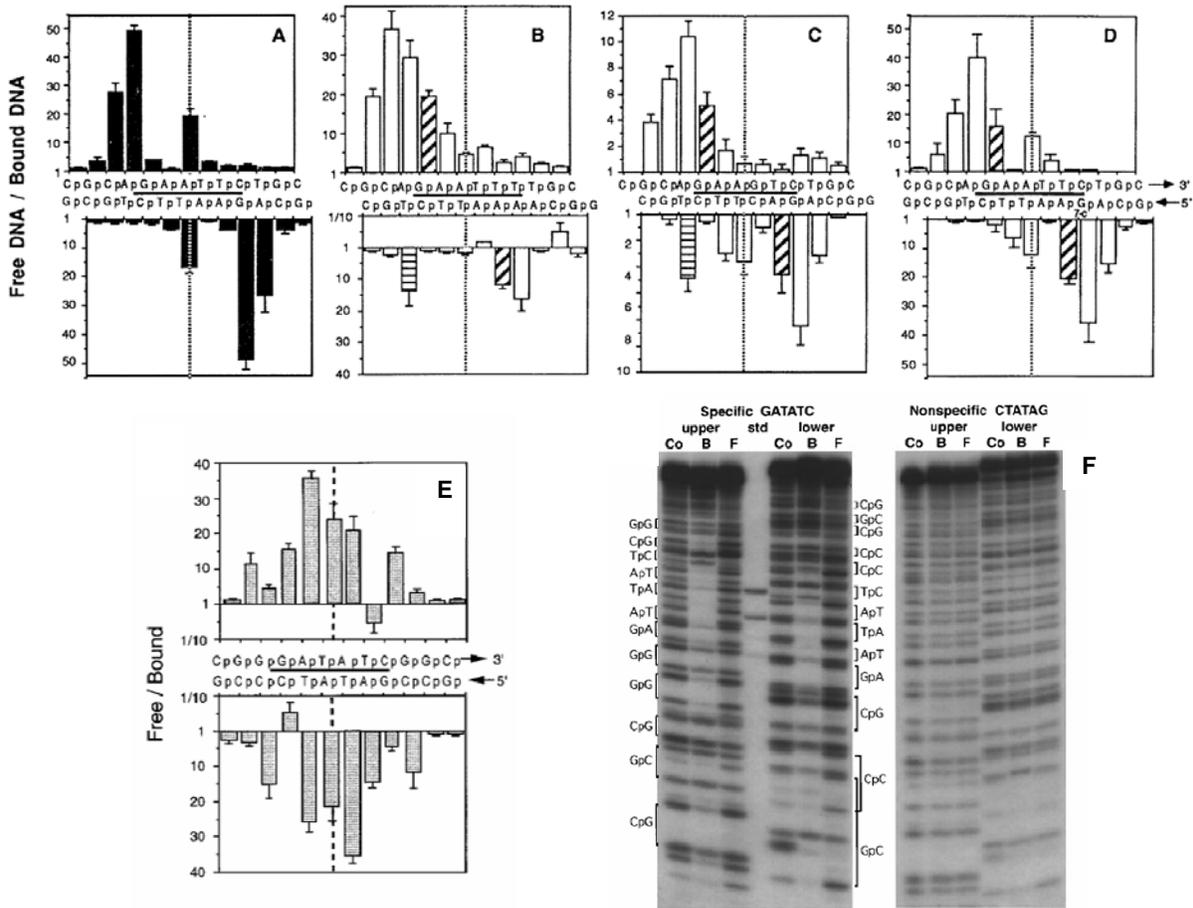


Figure 1.7. Ethylation interference footprints for specific, miscognate, and non-specific restriction endonuclease-DNA-DNA complexes.

Ethylation interference profiles for EcoRI endonuclease binding to specific (GAATTC; A), miscognate (AAATTC; B and GACTTC; C), and the specific sites with 7-deaza-guanosine substituted for GATATC (D). Ethylation interference footprints are also shown EcoRV binding to specific (GATATC; E), and non-specific sites Bar heights represent magnitudes of the perturbation caused by ethylation of particular phosphates. Magnitudes > 1 suggest that ethylation perturbs binding while those that are < 1 suggest that ethylation improved binding. There is a strong symmetric ethylation interference pattern for the specific EcoRI-DNA complex. In B-C, bottom panel shows the ethylation interference pattern in the modified half-site (i.e. AAA half-site in B) while the top panel represents that for the specific half-site. Base substitutions (B and C) result in ethylation interference patterns that are asymmetric and distinct from those observed for the specific complex suggesting adaptation in these complexes (see text). Base analog substitutions result in much more subtle effects on the ethylation interference pattern relative to that for the unmodified substrate. Ethylation interference profiles are also shown for the specific (E and F) and non-specific (F; right) EcoRV-DNA interactions. Autoradiograms show footprints on the top and bottom strands of the specific substrate and no footprint on the two strands of the non-specific substrate (F). Panels A to D taken directly from Lesser *et al.* (1990). Panels E and F taken directly from Engler *et al.* (1997).

Table 1.1 Discrimination at the protein-DNA interface: Three types of binding modes^a.

	Specific Complex	Miscognate Adaptive Complex	Non-specific Complex
DNA site	GATATC (EcoRV) GAATTC (EcoRI) GGATCC (BamHI)	One incorrect base pair	≥2 incorrect base pairs
Specificity ratio^b	1	1.3x10 ⁵	2.5x10 ⁶
Site Localization (Footprinting)^c	Strong	Moderate	Weak
pH effect^d	Strong	Weak	Very weak
Ca²⁺ effect^e	Strong	Weak	None
Salt dependence^f	Steep	Intermediate	Very steep
Heat capacity change (ΔC_P^o)^g	Large and negative (≥ -1.2 kcal/mol·K)	Moderate and negative (~ -0.2 – 0.5 kcal/mol·K)	~ 0
Number of waters released^h	430	130-150	80
Cleavage of DNA strandsⁱ	Symmetric rates	Reduced, asymmetric rates	None
DNA conformation^j	Symmetric distortion (mild to dramatic)	Asymmetric distortion?	None?

^a General features apply qualitatively to all three type II restriction endonuclease complexes (EcoRV, EcoRI, and BamHI); exact numbers are for the EcoRV endonuclease unless otherwise noted.

^b Specificity Ratio = $\frac{K_{specific}}{K_{non-cognate}}$ Calculated for EcoRI (Jen-Jacobson, 1997)

^c (Lesser et al. 1990; Engler et al. 1997; Engler et al. 2001)

^d (Engler et al. 1997)

^e (Engler et al. 1997; Engler et al. 2001; Kurpiewski et al. 2004; Sapienza 2005a)

^f (Engler et al 1997; Lesser, unpublished work)

^g (Jen-Jacobson 2000; Grigorescu 2003)

^h Water released upon complex formation (Cao, 2002)

ⁱ (Lesser, 1990)

Table taken directly from Cao 2002

1.4 THE ROLE OF DNA DISTORTION IN SPECIFICITY DETERMINATION

DNA distortion plays both a structural and energetic role in sequence discrimination by site specific binding proteins. Thus, this phenomenon often accompanies specific complex formation, whereas non-specific complexes contain largely undistorted DNA (Winkler, 1993; Albright et al., 1998; Kalodimos et al., 2004). From a structural perspective, DNA distortion is required in order to precisely orient different recognition elements so as to increase steric and electrostatic complementarity between protein and DNA molecules. Indeed, in the specific EcoRI-DNA complex, a central DNA kink is required to open the major groove of the DNA such that the major groove functional groups are accessible to protein side chains. Further, DNA distortion brings phosphate groups into novel positions relative to those in B-form DNA and is likely important for the formation of phosphate “clamp” interactions. From an energetic point of view, favorable contributions to specific complex formation deriving from protein-phosphate interactions, protein-base interactions, and from water release are thought to “pay” the energetic cost of DNA distortion, while non-specific complex formation, which does not feature favorable contributions from protein-DNA contacts or from water release, cannot pay the cost of DNA distortion (Figure 1.4).

DNA distortion is tightly correlated to biological function of many DNA binding proteins, including the assembly of nucleoprotein complexes required for regulation of transcription (Schultz et al., 1991; Kim, 1993a; Kim, 1993b; Goosen, 1995), recombination (Johnson et al. 2005), DNA compaction (Luger, 1997; Luijsterburg, 2006) and correct positioning of functional elements required formation of catalytic active sites (Winkler, 1993). The functional requirement for DNA distortion is exemplified by the fact that enhancement of recombination and transcriptional repression, which require DNA bending, can be achieved by

replacing either integration host factor (IHF) or phage $\Phi 29$ protein p4 binding sites with DNA sequences that are intrinsically bent or can be bent by alternative binding proteins (Goodman and Nash, 1989; Snyder et al., 1989; Rojo and Salas, 1991). Thus, functional discrimination between specific and non-specific sites can in part be due to the functional requirements for DNA distortion. A particular type of DNA distortion that is of interest is the linear bending of the DNA about its helical axis. This type of DNA bending is observed in many specific protein-DNA complexes, notably those involving catabolite activating protein (CAP) (Schultz et al., 1991), *lac* repressor (Spronk et al., 1999), and the EcoRV restriction endonuclease (Winkler, 1993). DNA bending is also observed for protein-DNA complexes involving binding proteins that exhibit reduced discrimination such as the histone octamer (Luger, 1997). Interestingly, sequences with a higher degree of “bendability” are more stably incorporated into nucleosomes (Spronk et al., 1999).

1.5 DNA BENDING INDUCED BY DNA BINDING PROTEINS

DNA bending over short distances requires energy. This unfavorable cost of DNA bending is overcome by DNA bending proteins by way of favorable van der Waals contacts, burial of hydrophobic surfaces, formation of hydrogen bonds, and Coulombic interactions that release condensed counterions. The following sections will discuss the sources of the energetic requirement for DNA bending as well as two different classes of proteins that fulfill this requirement in order to facilitate DNA bending.

1.5.1 Sources of DNA rigidity

Short segments of DNA are usefully described by the worm like-chain (WLC) model as stiff elastic polymers (Bustamante et al., 2003). In this continuum model, DNA rigidity (or flexibility) is described by its persistence length (A), which is defined as the distance over which two segments of the chain remain directionally correlated, or the length at which the time averaged angle made between the two ends of a DNA molecule is equal to 1 radian (57.3°) (Schellman, 1980; Travers and Thompson, 2004). For DNA in dilute cationic solutions ($M^+ \sim 1.0$ mM), A is found to be in the range of 800 to 1000 Å (Hagerman, 1988; Baumann et al., 1997). The theory of Manning *et al.* (1989) suggests that the rigidity of the DNA double helix is due to a balance between attractive and repulsive forces including, but not limited to, base stacking and stretching due to interphosphate repulsion (Manning et al., 1989). DNA bending thus requires destacking of DNA bases and the compression of phosphate charge on the concave surface of the bend DNA molecule. Indeed, both of these factors have been experimentally and theoretically implicated in determining the rigidity and structure of free DNA molecules and will be discussed in the following sections.

1.5.1.1 DNA helical rigidity and base stacking

DNA rigidity has been attributed to favorable vertical stacking of DNA bases, and experimental support for this model comes from several sources. Thermodynamic measurements suggest that DNA base stacking interactions are associated with a free energy change (ΔG) of about -1.0 to -2.0 kcal/mol (Searle and Williams, 1993). Additionally, free bases or nucleosides in solution have been shown to form stacked arrays (Chan et al., 1964; Solie and Schellman, 1968). Theoretical studies suggest that electrostatic properties of DNA bases create an energetically

favorable orientation for stacking interactions, resulting in strong tendencies to form helically stacked structures in the absence of a phosphate backbone suggesting that base stacking is a key determinant in the helical structure of the DNA double helix (Sarai et al., 1988). Finally, work by Hagerman's group has shown that a gapped duplex, consisting of a 24 base segment of single stranded dT (dT₂₄) flanked by two 38 bp regions of dT:dA duplex, shows hydrodynamic properties consistent with considerable flexibility of this species relative to a full duplex as measured by transient electric birefringence. The addition of free purine bases results in the formation of "meroduplex" in which the purine bases pair with the single-stranded dT region in the gapped duplex and stack on top of one another. The "meroduplex" displays hydrodynamic properties not statistically different from the full duplex, suggesting that base stacking of the purine:dT base pairs contributes considerably to the persistence length of the meroduplex independent of a full phosphate backbone (Hagerman and Hagerman, 1996; Mills and Hagerman, 2004).

An interesting feature of DNA base stacking interactions as they pertain to DNA helical structure and stability is the sequence dependence of these interactions. In contrast to the worm like chain treatment of the DNA double helix, which portrays DNA as a uniformly rigid elastic rod independent of base sequence, it has becoming clear that DNA structure and rigidity are not uniform, but rather, are heterogeneous (anisotropic) due to unique contributions from different DNA base steps. This concept is based on the analysis of the X-ray crystal model for a DNA dodecamer (Drew and Dickerson, 1981). This structure revealed that the values of local structural parameter including, backbone torsion angles, helical twist angles, propeller twist angles, and degree of vertical base overlap (stacking) depended on the sequence of dinucleotide steps (Dickerson and Drew, 1981). As more free and protein bound structures of duplex DNA

were solved, El Hassan and Calladine extended this analysis by considering the conformational characteristics of DNA base pair steps (i.e. sequential linear stacked bases) in 60 DNA complexes containing only Watson-Crick base pairs (El Hassan and Calladine, 1997). In general, it was found that particular base pair steps (AA/TT, AT and GA/TC) occupy a restricted region of conformational space, while other steps (AC/GT, TA, and TG/CA) occupy a broader range of conformational space. When only B form DNA is considered (as defined by sugar conformation), conformation of the TA and TG/CA are the most variable, suggesting that there is increased flexibility at base steps that sample a broad region of conformational space relative to those that are more conformationally restricted. Indeed, DNA molecules of the sequence 5'dAAA_(n)-3' show longer persistence lengths (more rigid) than 5'dATATAT_(n)-3' (Hogan et al., 1983; Travers, 2004). This sequence dependent rigidity or flexibility of DNA molecules has, therefore, been suggested to play a key role in site specific protein-DNA recognition (Schultz et al., 1991; Winkler, 1993; Horton and Perona, 1998b).

1.5.1.2 Electrostatic origins of DNA rigidity

In addition to base stacking interactions, DNA rigidity is conferred by the normal distribution of phosphate charge about the surface of the DNA double helix. This concept is illustrated by the fact that persistence length A , and thus DNA rigidity, is dramatically dependent on the ionic environment. That is, increasing monovalent cation concentration ($[M^+]$) from 0 to ~ 20 mM reduces the persistence length of double helical DNA from about 950 Å to about 500 Å, while there is little effect on A at $[M^+] > 20$ mM (Cairney and Harrington, 1982; Hagerman, 1988; Porschke, 1991; Sobel and Harpst, 1991; Baumann et al., 1997). Further, the presence of divalent and trivalent cations result in a reduction of A from ~ 500 Å to as little as 250 Å (Baumann et al., 1997). A physical view of the effect of ionic strength on A comes from

Manning's theory of counterion condensation, which suggests that the high density of negatively charged phosphates associated with the DNA backbone results in a concentrated cloud of mobile hydrated counter ions that reside within ~ 7 Å of the DNA surface (Manning, 1978). The ionic strength of this cloud approaches 1.0 M for monovalent cations and results in the neutralization of $\sim 76\%$ of the DNA charge, thus reducing the effective charge of each phosphate to $-0.24e$. The fact that there is no dependence of A on ionic strength for $[M^+] > 20$ mM suggests that the "cloud" of counterions around the DNA phosphates is saturated at these $[M]^+$. Divalent and trivalent cations reduce phosphate charge even further to $-0.12e$ and $-0.08e$, respectively (Manning, 1978), consistent with the observation that these cations can reduce A to a larger degree relative to monovalent cations. Finally, a variety of multivalent cations have been shown to facilitate condensation of DNA molecules into folded toroidal structures (Haynes et al., 1970; Laemmli, 1975; Gosule and Schellman, 1978; Bloomfield, 1997). Taken together, these observations highlight the contribution to DNA persistence length made by the electrostatic stretching forces between phosphate moieties even at high monovalent cation concentrations ($[M^+] > 20$ mM).

A can be affected by both dynamic or static bends in the DNA double helix (Schellman and Harvey, 1995). It is generally thought that the effect of monovalent cations results from the uniform neutralization of phosphate charge and uniform increase in flexibility resulting in dynamic fluctuations in the DNA double helix and a consequent reduction in A . More dramatic effects on A are observed for increasing multivalent cation concentrations. Multivalent cations have been shown to localize to the minor groove of specific DNA sequences, in which case it is thought that the charge neutralization resulting from multivalent cation binding leads to a static DNA bend toward the site of cation binding. Alternatively, for random DNA sequences, it has

been shown by nuclear magnetic resonance imaging (NMR) that multivalent cations are more mobile and are only loosely associated with the DNA backbone (Rose et al., 1982; Wemmer et al., 1985; Berggren et al., 1992). This observation lead Bauman *et al.* (1997) to suggest that the effect of increased multivalent cation concentration on A derives mainly from enhancement of dynamic bends, which are skewed toward the concentrated positive charge at the site of a transiently associated multivalent cation (Baumann et al., 1997). These studies suggest that relaxation of interphosphate stretching forces by mono- and multivalent cations can reduce A by means of either static or dynamic DNA bending.

1.5.2 Two types of DNA bending proteins

In order for a DNA binding protein to bend its DNA substrate, the protein must either provide enough free energy during the binding reaction to overcome the free energy required for DNA bending, or it must reduce the cost of DNA bending by destabilizing the B form helix. In this light, two types of DNA bending proteins have been identified that use different strategies to facilitated DNA bending. These proteins are referred to as type I and type II DNA bending proteins (Maher, 1998) and are classified on the basis of the direction of the induced bend relative to the surface of the protein. That is, type I bending proteins bend their substrate DNA away from the surface of the protein while type II DNA bending proteins typically bend their substrate DNA toward the face of the protein (Maher, 1998) (Figure 1.8). It should be noted that not all DNA bending proteins can be classified as purely type I or type II, as they contact both surfaces of the DNA and use a combination of structural and energetic strategies to facilitate DNA bending.

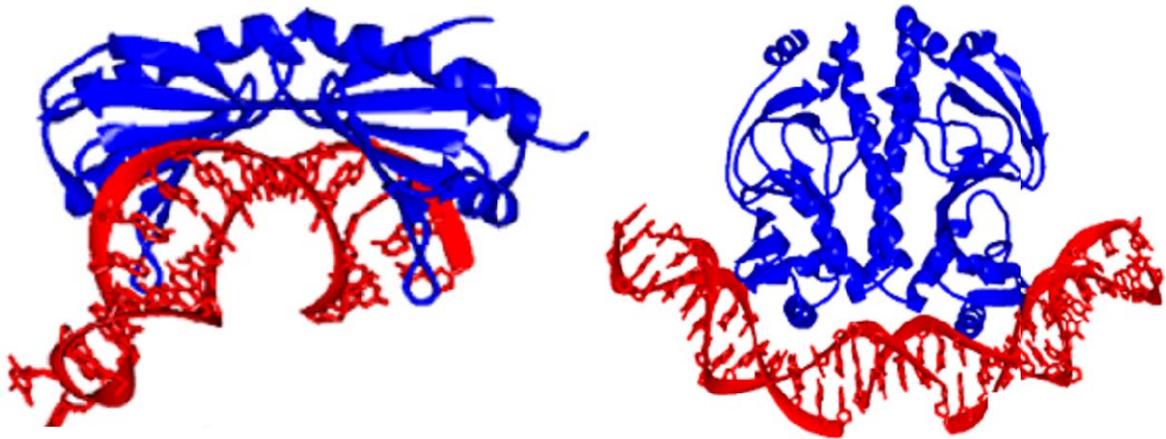


Figure 1.8 Two types of DNA bending proteins.

DNA bending proteins (blue) TBP (type I; left) and CAP (type II; right) bend their substrate DNA (red) away from the face of the protein and toward the face of the protein, respectively. This figure was adapted from Jen-Jacobson *et al.* (2000).

1.5.2.1 “Type I” DNA bending proteins

DNA bending proteins that facilitate bending away from the surface of the protein have been referred to as type I DNA bending proteins, a hallmark example of which is TATA binding protein (TBP; Figure 1.8). TBP is a key component of the eukaryotic transcription machinery that is required for the initiation of RNA synthesis by all three eukaryotic RNA polymerases (Tansey and Herr, 1997). At class II nuclear promoters, TBP, along with its associated factors, facilitates the formation of the multi-protein complex required for proper transcription initiation and recruitment of RNA polymerase II (Hisatake et al., 1993; Ruppert et al., 1993). Dramatic DNA bending of $\sim 90^\circ$ is observed in the TBP-DNA complex (Kim et al., 1993a; Kim et al., 1993b), and this is partially facilitated by the insertion of two phenylalanine (Phe) side chains (residue numbers 57 and 148) into the minor groove of the promoter DNA, thus stabilizing two kinks at the 5' most T-A:A-T and the A-T:G-C base steps of the TBP binding sequence TATAAAAG. The energy required to destack these bases is compensated for by the stacking of the Phe side chain against the A and T bases at the two aforementioned base steps. It has been proposed that intercalation of these bulky hydrophobic groups facilitates DNA bending by opening the minor groove and altering the helical trajectory (Maher, 1998). Several other proteins seem to use this strategy to bend DNA, including the human male sex determining factor (SRY) (Werner et al., 1995b), the human lymphoid enhancer binding protein 1 (hLef-1) (Love et al., 1995), human ETS1 (Werner et al., 1995a), and the *Escherichia coli* purine repressor protein PurR (Schumacher et al., 1994).

In addition to opening the minor groove by inserting a bulky hydrophobic residue, DNA bending proteins have been suggested to facilitate DNA bending away from the protein surface by an electrostatic mechanism (Travers, 1995). It has been shown theoretically that a protein-

shaped region having a low dielectric that is designed to fit snugly into the minor groove but that forms no favorable interactions with the DNA (e.g. no hydrogen bonds or van der Waals interactions) will facilitate minor groove opening and DNA bending when brought into close proximity (7.5Å) of the DNA molecule (Elcock and McCammon, 1995). This effect is said to derive from reduced shielding of electrostatic repulsion between cross-strand phosphates (Elcock and McCammon, 1995) and highlights the importance of these forces in stabilizing the DNA helix such that an imbalance can lead to significant DNA bending. This effect may also underscore the “desire” for phosphates to remain solvated upon approach of the low dielectric “protein” due to the fact that the solvation energy of the DNA remains constant as the protein comes closer and the DNA is deformed (Elcock and McCammon, 1995). This is distinct from a real protein-DNA interaction in which desolvation is typically accompanied by compensating interactions (electrostatic, hydrogen-bond, or van der Waals).

1.5.2.2 Type II DNA bending proteins

Type II DNA bending proteins contact the concave face of the bent DNA and facilitate bending toward the surface of the protein; an example of this is the *E. coli* catabolite activator protein (CAP). CAP is a transcriptional activator which, when bound to its allosteric effector cAMP (adenosine 3',5'-monophosphate), regulates the expression of more than 20 different promoters (de Crombrughe et al., 1984). CAP also induces dramatic DNA bending of ~90° that is largely due to two kinks at the second TG/CA base step of the specific recognition half-sequence TGTGA one on each side of the dyad axis (Schultz et al., 1991). Bending itself is thought to be important for transcription initiation as DNA elements with intrinsic curvature (A⁶ tracts) that replace CAP binding sites upstream of the *lac* promoter stimulate the rate of transcription initiation *in vitro* by 10-fold relative to promoters without these elements (Gartenberg and

Crothers, 1991). In addition, intrinsically curved DNA elements were able to restore expression at the *gal* promoter in the absence of a CAP binding site (Bracco et al., 1989). Together, these data suggest that DNA curvature can stimulate initiation of gene expression, which result from an increased binding affinity for the α subunit of RNA polymerase to these sequences (Aiyar et al., 1998). In addition to CAP, several other proteins facilitate DNA bending toward the protein surface, including histones (Luger, 1997), *E. coli* integration host factor (IHF) (Rice et al., 1996), the pu.1 ETS domain (Pio et al., 1996), the *Saccharomyces cerevisiae* MATa1/MATa2 homeodomain (Li et al., 1995), and the serum response factor (Pellegrini et al., 1995). While all of these proteins contact the concave face of their bent DNA substrates, the actual DNA bend can vary in magnitude and in topology (e.g. isolated kinks as in CAP or a more smoothly distributed bend as in the nucleosome). The surfaces of these proteins that are engaged with the DNA typically consist of multiple basic side chain residues, resulting in a positive surface that has electrostatic compatibility with the surface of the DNA. Thus, it has been proposed that this positive surface will effectively neutralize one face of the substrate DNA. This is thought to induce DNA bending toward the neutralized face by making use of interphosphate repulsive forces on the opposite surface of the DNA molecule. The phenomenon of asymmetric phosphate neutralization and its role in DNA bending will be discussed in the next section.

1.6 ASYMMETRIC PHOSPHATE NEUTRALIZATION INDUCES DNA BENDING

In accordance with the view that electrostatic repulsive forces contribute to the rigidity of the DNA double helix (see Section 1.5.1.2), Mirzabekov and Rich proposed a model for DNA bending suggesting that it could be induced by lateral asymmetric neutralization of the fractional

negative charge associated with DNA phosphates. This neutralization would lead to an imbalance of phosphate repulsion on one face of the DNA and result in the compression of the DNA toward the neutralized surface, in part, due to the repulsive forces on the convex face of the DNA molecule (Mirzabekov, 1979) (Figure 1.8). Manning *et al* (1989) estimated that DNA bending required for nucleosome formation could result from asymmetric neutralization of 10% of the total phosphate backbone (Manning et al., 1989). While this model for DNA bending was inspired by DNA bending in the nucleosome core particle, it has been implicated as strategy used for inducing or reducing the cost of DNA bending by several DNA binding proteins including CAP and the EcoRV restriction endonuclease.

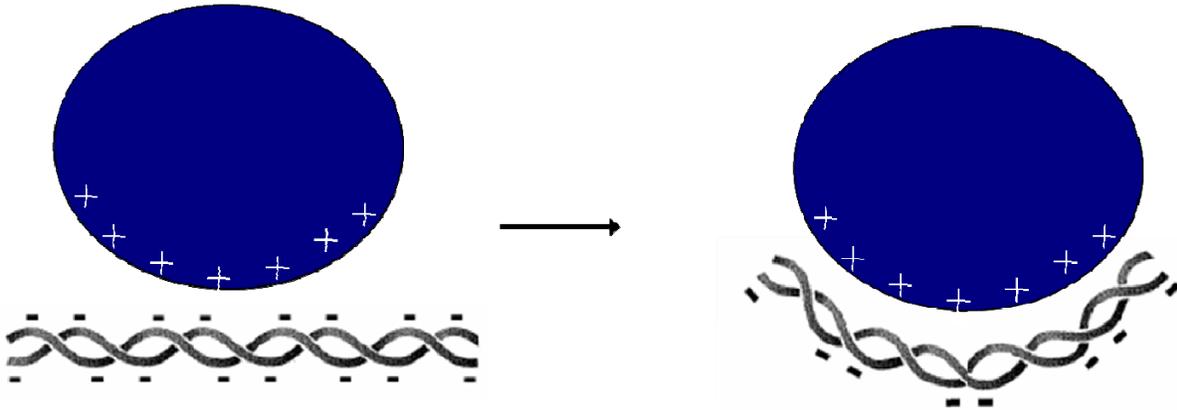


Figure 1.9 DNA bending induced by asymmetric charge neutralization.

Representation of protein-induced DNA bending by asymmetric phosphate neutralization. The protein (blue sphere) presents an electropositive surface (+) to one face of the polyanionic DNA (black tubes), causing a relaxation in interphosphate repulsion on the neutralized face of the DNA that results in a collapse of the DNA toward the surface of the protein. This collapse is stabilized by the compressive force generated from interphosphate repulsion on the DNA face that is not contacted by the protein.

1.6.1 Free DNA bending by APN measured experimentally

To test the applicability of the model proposed by Mirzabekov and Rich (see above), Maher's group undertook a series of experiments to probe the effects of asymmetric phosphate neutralization on free DNA conformation. Strauss and Maher (1994) employed an uncharged phosphate analog (methylphosphonate P_{Me}), in which one of the non-bridging phosphoryl oxygens is replaced with a methyl group (Figure 1.9). P_{Me} are electrochemically neutral and can be placed at any desired position during oligonucleotide synthesis. As shown by anomalous gel mobility, neutralization of a patch of six phosphates (three per strand) that are disposed across the minor groove facilitated bending of the DNA double helix. By placing this patch of P_{Me} s at different positions relative to a tract of six consecutive adenine bases (A^6 -tract), a DNA element with known intrinsic curvature toward the minor groove (Figure 1.10), it was shown that phosphate neutralization induced spontaneous bending *toward* the neutralized face of the DNA. Specifically, when the neutralized patch was placed ten bp (1 helical turn) away from the center of the A^6 tract, apparent DNA bending was *increased* whereas it was *decreased* when the neutralized patch was placed five bp (1/2 helical turn) from the center of intrinsic curvature. Quantification of these data suggested that neutralization of a patch of six phosphates induces DNA bending by approximately 20° . P_{Me} induced bending toward the neutralized face of the free DNA was inhibited by increasing amounts of multivalent cations, and was sequence independent. Importantly, P_{Me} substitutions themselves had no effect on the gross helical structure of the free DNA molecule, as inferred from the observation that *symmetric* phosphate neutralization about the helical axis did not produce a detectable effect on helical geometry. (Strauss and Maher, 1994). A more diffuse pattern of phosphate neutralization by P_{Me}

substitution made across the minor groove (Figure 1.10) induced a more subtle bend in free DNA of $\sim 13^\circ$ (Strauss-Soukup et al., 1997), while phosphate neutralization via covalent tethering of ammonium ions (electropositive at neutral pH) to the phosphate backbone (instead of by P_{Me} substitution) induced DNA bending by $\sim 8^\circ$ for a propyl tether (Strauss et al., 1996a) and by $\sim 4^\circ$ when a longer hexamethylene tether was employed (Strauss et al., 1996b). Time resolved Förster resonance energy transfer (trFRET) experiments verify that two adjacent tethered cations can induce free DNA bending toward the neutralized surface, while this is not the case for a single tethered cation (Williams et al., 2006). Taken together, these data suggest that phosphate neutralization induces DNA bending toward the neutralized surface, and the amount of DNA bending induced is directly correlated with the number of phosphates neutralized.

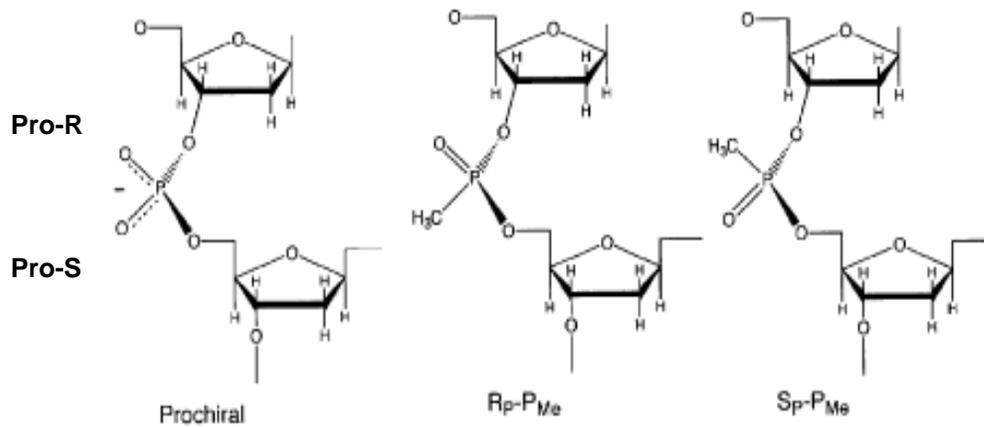


Figure 1.10 Chemical structure and stereochemistry of neutral methylphosphonates (P_{Me}).

Anionic phosphate diester (left) is shown along with two stereospecific P_{Me}. Replacement of the O1P non-bridging phosphoryl oxygen with a methyl group results in an R_p-P_{Me} stereoisomer (center), while that for the O2P phosphoryl oxygen results in the S_p-P_{Me} stereoisomer (right).

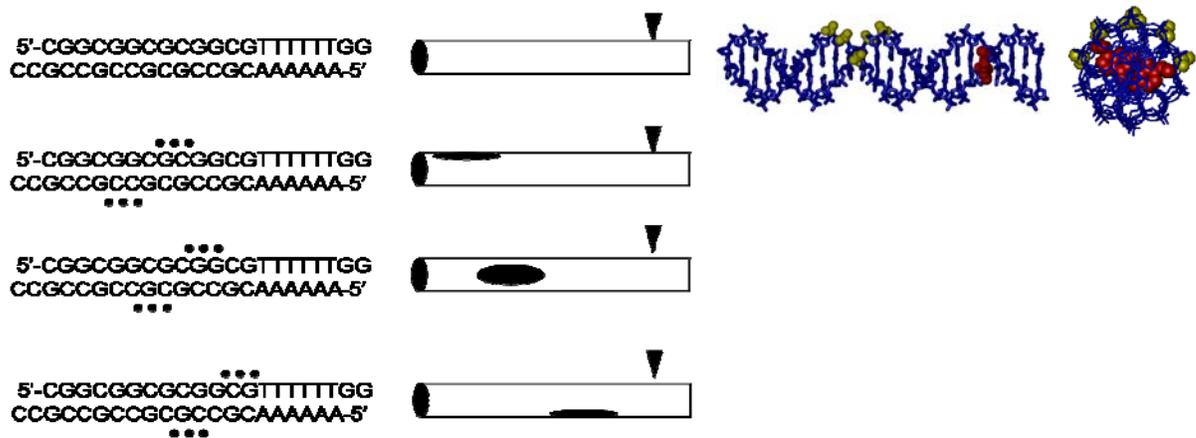


Figure 1.11 Asymmetric phosphate neutralization induced DNA bending-A tract phasing.

Substrates used to measure the effects of phosphate neutralization on free DNA. The DNA base sequence of the unmodified and P_{Me} modified substrates are shown. A patch of six P_{Me} analogs (filled circles; first column and black ovals; second column) were placed across the minor groove either *cis*, *ortho*, or *trans* to the center of an A^6 tract (black triangle) with known intrinsic curvature toward the major groove. Three-dimensional model depicting the position of the P_{Me} substitutions (yellow spheres; third and fourth columns) relative to the center of the A^6 tract (red); DNA bending by the A^6 tract is in the upward direction in these representations. Side views (third column) and head on views (fourth column) are shown. Figure was adapted from and describes work by Strauss *et al.* 1994.

In order to determine if the phosphate neutralization pattern employed by DNA binding proteins was sufficient to induce free DNA bending Maher's group used X-ray crystal models to identify putative electrostatic interactions between both CAP and pu.1 basic side-chain residues and phosphates at their specific DNA binding sites (Strauss-Soukup and Maher, 1997). Once identified, appropriate phosphates were neutralized by P_{Me} substitution, and the effects of phosphate neutralization on the shape of the free DNA molecules were evaluated by electrophoretic gel mobility. Interestingly, phosphate neutralization at the seven positions predicted to be contacted by basic side chain residues in the PU.1-DNA complex induced DNA bending by 28°, relative to the unmodified DNA substrate. P_{Me} induced DNA bending contrasts both the magnitude (8°) and direction of the DNA bend observed in the crystal model (Strauss-Soukup and Maher, 1997). P_{Me} substitution at phosphates predicted to be neutralized in the CAP binding site increased the intrinsic curvature of this site by 23% to (~26° to ~32°) in the same direction observed in the CAP-DNA crystal model (Hardwidge et al., 2002). These data suggest that phosphate neutralization provided by DNA bending proteins can contribute favorably to the DNA bending that is induced by these proteins, but cannot account for the degree or direction of the observed bend alone. In a similar study, Tomky *et al.* (1998) show that increasing phosphate neutralization at the AP-1 site (binding sites for transcription factors containing the basic leucine zipper domain, bZIP) results in increased bending by ~3.5° per neutralized phosphate in the predicted direction (i.e. toward the neutralized face) (Tomky et al., 1998). This observation is in quantitative agreement with the amount of bending observed for variant bZIP domains that have been engineered to neutralize various amounts of phosphate charge upon binding to the AP-1 site.

1.6.2 Free DNA bending by asymmetric phosphate neutralization measured computationally

Computational methods also provide evidence in support of the hypothesis that DNA bending is induced by asymmetric phosphate neutralization. Kosikov *et al.* (2002) simulated phosphate neutralization via P_{Me} substitution and by uniform neutralization (i.e. phosphate diester is defined to have a charge of 0) and found that neutralization of a patch of six phosphates (three per strand) results in bending of a free DNA (poly(CG)-poly(CG)) when disposed across the minor groove (7° bend) or the major groove (~11° bend). The authors suggest that the larger magnitude of DNA bending observed for neutralization across the major groove reflects the fact that the major groove is wider and provides more room for DNA bending without steric clash. This study also demonstrated that DNA bending induced by phosphate neutralization via P_{Me} substitutions shows stereospecific effects. All Rp-P_{Me} substitutions (Figure 1.10) induce a smaller amount of DNA bending (~9.3° across the minor groove and ~ 7.6° across the major groove) than do all Sp-P_{Me} modifications (~12.2° across the minor groove and ~ 22.5° across the major groove) (Kosikov *et al.*, 2002). These stereospecific effects observed computationally are consistent with experimental observations that suggest that Sp-P_{Me} substitutions may induce a slightly larger degree of bending.

In a separate set of experiments, the JUNCTION MINIMIZATION OF NUCLEIC ACIDS (JUMNA) program for nucleic acid modeling developed by the Lavery group (Lavery *et al.*, 1995) was used to predict DNA bending due to asymmetric phosphate neutralization. For the CAP binding site, phosphates involved in electrostatic interactions with positively charged protein side chains were neutralized, and energy optimized conformations of this DNA site were obtained. It was shown that the phosphate neutralization pattern corresponding to the CAP-DNA crystal model increased

the bend angle of the CAP binding site from 38° to 51° (Gurlie et al., 1999). When the pattern of neutralization was applied to a DNA site with an alternating GC sequence, bending was shown to increase by only 3°, from 4° for the non-neutralized DNA to 7° for the neutralized site, suggesting that the effects of this pattern of phosphate neutralization are highly sequence specific (Gurlie et al., 1999). This observation is distinct from that made from experiments employing patches of phosphate neutralization, which have been shown to induce free DNA bending independent of DNA sequence (Strauss and Maher, 1994; Sanghani et al., 1996). This analysis was extended to 24 protein-DNA complexes for which structural data show distinct DNA bending. It was observed that neutralization of phosphates involved in electrostatic interactions with positively charged protein side chains results in bending of the free DNA molecule for many of the DNA proteins examined, including many of the homeodomain proteins and HLH/bZIP family members. Calculation of the energetic cost required to bend a particular DNA site suggests that phosphate neutralization often can reduce this cost as much as -26 kcal/mol for one HLH/bZIP family member (Gurlie and Zakrzewska, 2001).

1.7 INVESTIGATING THE ROLE OF PHOSPHATE NETRALIZATION IN PROTEIN INDUCED DNA BENDING: ECORV AS A MODEL

The studies described above provide substantial support that phosphate neutralization at targeted positions can have a significant impact on DNA curvature. However, all of these studies used either a "phantom protein" approach (i.e. chemical neutralization of phosphate positions on one face of the DNA to mimic charge neutralization by a DNA binding protein) or cations tethered by hexamethylene linkers to the DNA backbone. Therefore, I wished to probe the role of

phosphate neutralization in DNA bending by a "real" DNA binding protein. In my dissertation work, I used the interaction of EcoRV with its cognate site as a model for understanding the structural and energetic contributions of neutralization at particular phosphate positions in protein-induced DNA bending.

The interaction of restriction endonuclease EcoRV with its cognate site is a particularly useful model because: 1) EcoRV dramatically bends its substrate DNA (Figure 1.4) by $\sim 50^\circ$ by imparting a positive roll at the center base step of its specific recognition sequence GATATC; 2) six basic side chain residues (three per monomer) contact the concave face the bent DNA substrate in the specific EcoRV-DNA crystal model (Winkler, 1993; Thomas et al., 1999; Horton and Perona, 2000) (Figure 1.12), suggesting that phosphate neutralization by the protein relaxes interphosphate repulsion and reduces the energetic cost of DNA bending; 3) There is a wealth of structural data available for the EcoRV-DNA interaction, including structures of the free protein, the specific EcoRV-DNA complex in the absence and presence of various active site metals, the specific EcoRV-DNA complex in various stages of DNA bending, and two non-cognate EcoRV-DNA complexes; 4) There are comprehensive thermodynamic and kinetic studies on binding and catalysis by EcoRV. These studies form a rich database that can aid in the interpretation of energetic consequences of perturbations at the protein-DNA interface.

2.0 CHAPTER 2: THE EFFECTS OF PHOSPHATE NEUTRALIZATION ON FORMATION OF THE ECORV-DNA COMPLEX

2.1 INTRODUCTION

DNA bending by site-specific binding proteins has been proposed as a factor that contributes to discrimination between specific and non-specific DNA sites (Section 1.3.2.2). A possible strategy used by DNA bending proteins to facilitate DNA bending is the asymmetric neutralization of phosphates along one lateral face of a DNA molecule, which induces bending toward this neutralized face. Asymmetric phosphate neutralization has been shown to induce DNA bending *in vitro* and *in silico* (Sections 1.6.1 and 1.6.2); however, an experimental investigation of the energetic consequences of phosphate neutralization and its role in protein-induced DNA bending has yet to be undertaken. In this chapter, I will discuss the experimental strategy used to understand the role of phosphate neutralization in EcoRV-induced DNA bending, as well as a proposed role for phosphate neutralization in reducing the cost of DNA bending by EcoRV.

2.1.1 Manipulating the electrostatic landscape at the EcoRV-DNA interface

In order to probe the effects of phosphate neutralization on the formation of the specific EcoRV-DNA complex, the charge landscape at the interface was manipulated by 1) mutating cationic

protein side chains that contact DNA phosphate groups to alanine residues (Figures 2.1 and 2.3) and/or 2) substituting negatively charged phosphate groups with uncharged methylphosphonates (P_{Me} ; Figure 1.10). A combinatorial approach has allowed the effects of phosphate neutralization to be separated from the confounding effects of EcoRV mutations and P_{Me} substitutions alone. For example, a stereospecific Rp- P_{Me-2} substitution (in which the P_{Me} methyl group replaces only the pro-S phosphoryl oxygen; see Figure 1.10 for P_{Me} naming conventions) removes favorable, attractive interactions to the K119 side chain (Figure 2.1), as well as unfavorable repulsive interactions with nearby phosphate groups on the concave face of the bent substrate DNA (Figure 2.2). Thus, the energetic effects of phosphate neutralization via Rp- P_{Me-2} substitution are obscured by the loss of the K119- P_{-2} interaction. However, in the context of a K119A mutant, the energetic effects of a stereospecific Rp- P_{Me-2} modification would only derive from phosphate neutralization, as the potential for a K119- P_{-2} interaction has been abolished by mutation. Thus, the energetic role of phosphate neutralization on the EcoRV-DNA binding reaction can be dissected using a combination of basic residue mutations and P_{Me} substitutions.

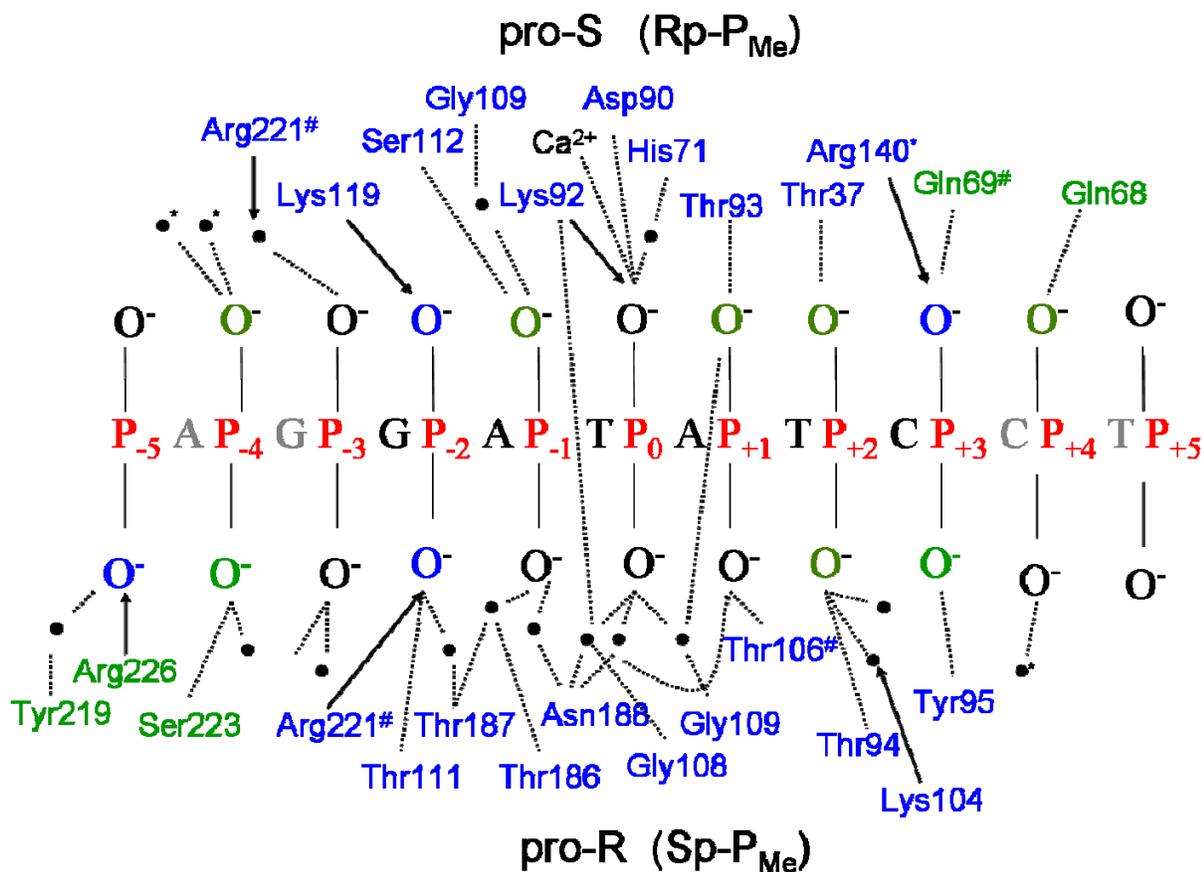


Figure 2.1 Phosphate contacts at the specific EcoRV-DNA interface.

Two-dimensional representation of the phosphate contacts to one strand of the specific EcoRV-DNA substrate. The DNA base sequence corresponds to that used in our thermodynamic experiments (while the sequence is AAAGATATCTT in the crystal model; 1B94.pdb). The specific recognition sequence (black) and the two flanking bases (grey) are shown as one letter text. Phosphates (red P) and phosphate phosphoryl oxygens (O⁻) are also shown. The subscripts associated with phosphate groups correspond to the phosphate numbering used in the body of this document. Contacts made to the pro-S and pro-R phosphoryl oxygens by polar side chain residues and water molecules (black circles) are represented as dotted lines, and those made by basic side chain residues are depicted as arrows. Contacts are based on an X-ray crystal model (1B94) with calcium (Ca²⁺) bound at the active site. All illustrated contacts are <3.5 Å in length. Phosphate contacts that make asymmetric contacts to only DNA chain C (#) or chain D (*) are distinguished from those that make symmetric contacts to phosphates in both DNA strands. Substitution of the pro-R phosphoryl oxygen with a methyl group results in an Sp-P_{Me} while substitution of the pro-S phosphoryl oxygen with a methyl group results in an Rp-P_{Me} substitution.

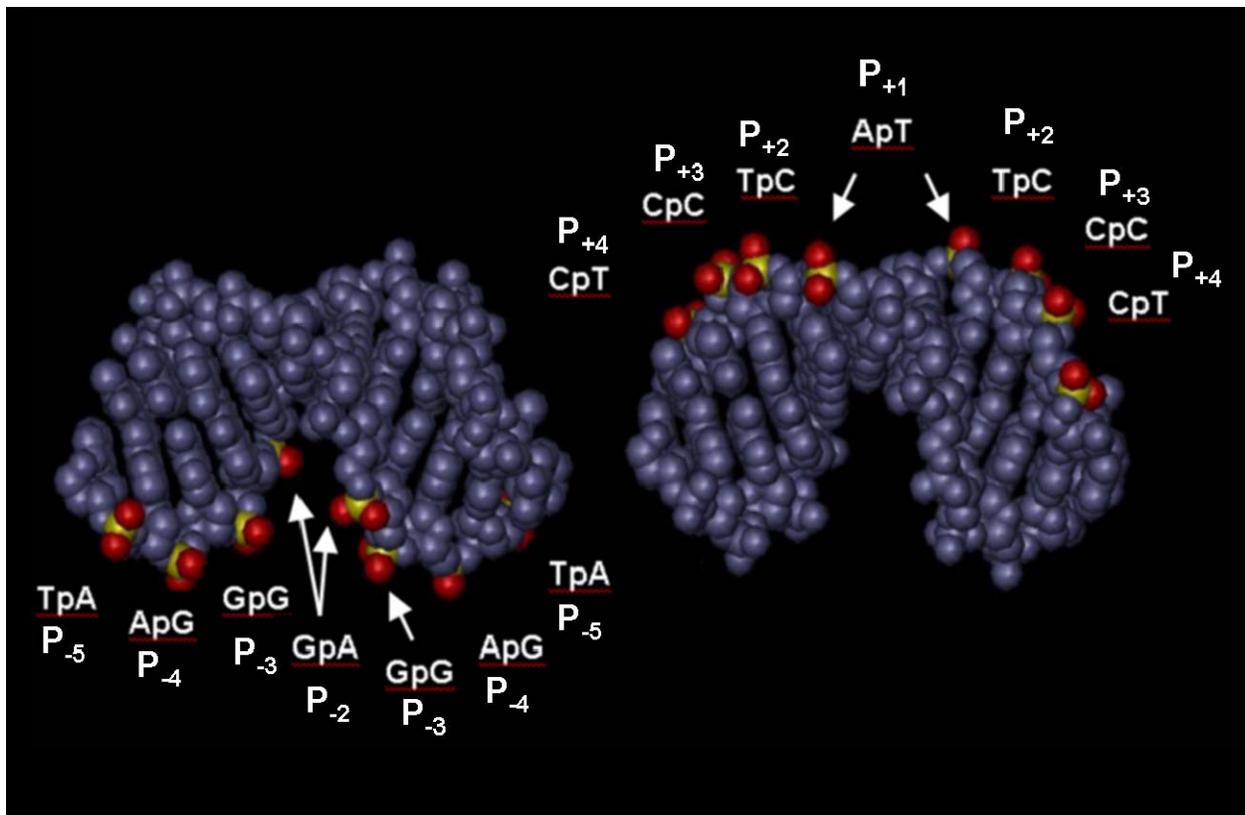
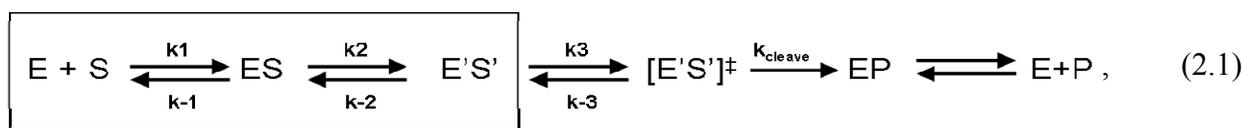


Figure 2.2 Three-dimensional arrangement of phosphates at the specific EcoRV-DNA interface.

Structural representation of the DNA in the specific EcoRV-DNA complex. The protein has been removed for clarity. Phosphate groups are highlighted (phosphorous atoms are yellow and oxygen atoms are red) and labeled with both base step position and number. The numbering scheme corresponds with that in Figure 2.2 and is used in the body of this document. Concave phosphate positions (relative to the direction of DNA bending) are highlighted on the left, and convex phosphate positions are on the right.

2.1.2 Measuring the effects of electrostatic perturbations on formation of the EcoRV-DNA complex

The effects of electrostatic perturbations have been measured for the formation of the specific EcoRV-DNA complex using the quantitative measures that will be discussed below. The progression of the EcoRV-DNA reaction can be represented schematically as



where E and S represent the free EcoRV and DNA molecules. Formation of the specific EcoRV-DNA complex, E'S', is thought to proceed through an intermediate, ES, which represents the non-specific encounter complex (Jen-Jacobson, 1997; Hiller et al., 2003).

In order to measure the formation of the specific EcoRV-DNA complex in solution, the essential catalytic cofactor magnesium (Mg^{2+}) has been replaced with calcium (Ca^{2+}), a divalent metal that binds to the EcoRV-DNA active site and inhibits DNA cleavage, restricting the reaction to the boxed region in Scheme 2.1 (Engler et al., 1997). The observed equilibrium constant K_{obs} for formation of the E'S' complex is defined by the equations:

$$K_{\text{sp,obs}} = \frac{[E'S']}{[E][S]}, \quad (2.2)$$

and

$$K_{\text{sp,obs}} = K_1 K_2, \quad (2.3)$$

where $K_1 = k_1/k_{-1}$ and $K_2 = k_2/k_{-2}$

The stability of the specific EcoRV-DNA complex can be described by the standard state Gibbs free energy change (ΔG°), given by

$$\Delta G^\circ = -RT \ln K_{\text{obs}}, \quad (2.4)$$

where R is the ideal gas constant (in kcal^oK/mol) and T is temperature (K). The difference between the free energy change for formation of the modified complex containing an electrostatic perturbation and the free energy change for formation of an unmodified complex is given by

$$\Delta \Delta G^\circ = \Delta G^\circ_{\text{bind,mod}} - \Delta G^\circ_{\text{bind,unmod}} = -RT \ln \frac{K_{\text{mod}}}{K_{\text{unmod}}}, \quad (2.5)$$

where $\Delta G^\circ_{\text{bind,mod}}$ and $\Delta G^\circ_{\text{bind,unmod}}$ are the standard state Gibbs free energies for the formation of modified and unmodified EcoRV-DNA complexes, respectively. Thus, the energetic effects of a particular electrostatic modification can be measured.

2.1.3 Thermodynamic pseudocycles

In evaluating the energetic consequences associated with a particular electrostatic perturbation, let us consider a thermodynamic pseudocycle, represented in Figure 2.3, for the binding of EcoRV to unmodified (DNA) or P_{Me} modified DNA (DNA-P_{Me}) to form the unmodified (EcoRV•DNA) and modified (EcoRV•DNA-P_{Me}) EcoRV-DNA complexes, respectively. This is

a “pseudocycle” because the vertical arrows do not correspond to *observable* physical processes. A similar pseudocycle can also be written for mutation of the protein.

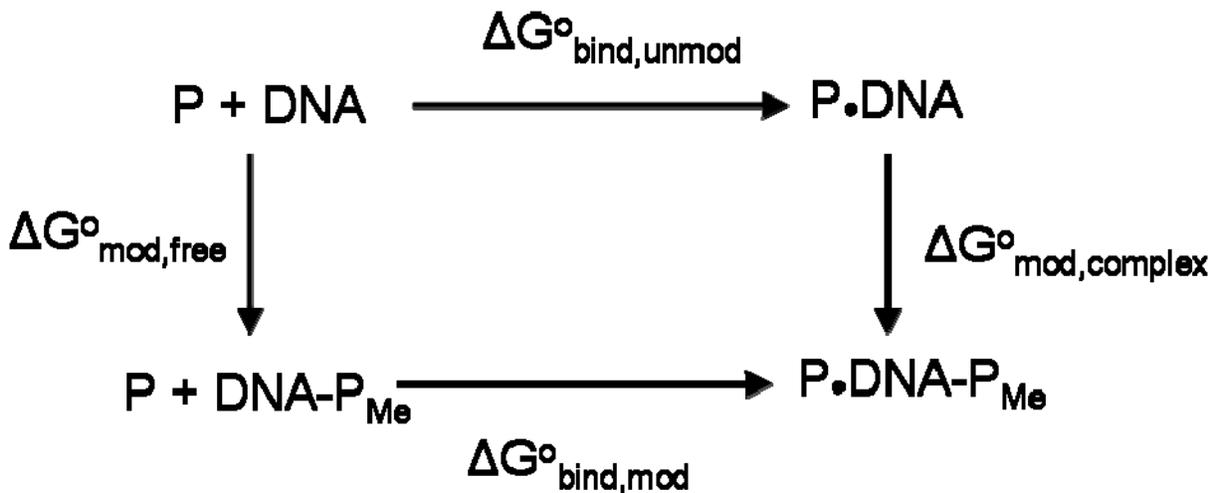


Figure 2.3 Thermodynamic pseudocycle for protein binding to modified and unmodified DNA substrates. Binding free energies ($\Delta G^{\circ}_{\text{bind,x}}$) for a protein interacting with both modified (DNA-P_{Me}) or unmodified (DNA) DNA substrates are represented for the progression from the free macromolecules (P + DNA or P + DNA-P_{Me}) to the complex (P•DNA or P•DNA-P_{Me}). These binding free energies can be measured experimentally. The free energies for making a P_{Me} substitution in the free DNA ($\Delta G^{\circ}_{\text{mod,free}}$) and in the complex ($\Delta G^{\circ}_{\text{mod,complex}}$) are also represented. These processes do not correspond to experimentally observable physical processes.

From this pseudocycle it can be written that

$$\Delta G^o_{bind,unmod} + \Delta G^o_{mod,complex} = \Delta G^o_{bind,mod} + \Delta G^o_{mod,free}, \quad (2.6)$$

where the subscript mod refers to the hypothetical free energy change of making a particular modification in either the free molecule ($\Delta G^o_{mod,free}$), or in the complex ($\Delta G^o_{mod,complex}$). It is not possible to directly measure the impact of these perturbations, either in the free molecules, or in the complex, as these modifications do not correspond to an experimentally observable physical process. From Equation 2.6, it follows that the experimentally observed difference in binding free energy

$$\Delta\Delta G^o_{bind} = \Delta G^o_{bind,mod} - \Delta G^o_{bind,unmod}, \quad (2.7)$$

is given by

$$\Delta\Delta G^o_{bind} = \Delta G^o_{mod,complex} - \Delta G^o_{mod,free}. \quad (2.8)$$

In principle, each modification can influence the entropy, enthalpy and free energy of the protein-DNA complex, of the free macromolecules, or both; thus, it is important to consider $\Delta G^o_{mod,free}$ when considering the thermodynamic consequences of making a particular modification. Structural data can provide information as to the features that influence $\Delta G_{mod,complex}$, while structural information for the free protein and DNA molecules (both modified and unmodified) would be required to understand the impact of these modifications on $\Delta G^o_{mod,free}$. Fortunately, structural models of the free EcoRV protein (Winkler, 1993; Perona and Martin, 1997), the EcoRV-DNA complex in the absence of metal (Winkler, 1993; Kostrewa and

Winkler, 1995; Perona and Martin, 1997; Horton and Perona, 1998a; Thomas et al., 1999) and the EcoRV-DNA complex in the presence of various divalent metals (Kostrewa and Winkler, 1995; Thomas et al., 1999; Horton and Perona, 2004) are available. These structural models, along with molecular dynamics (MD) simulations (see Chapter 3), can potentially provide insight into the possible structural consequences of a particular perturbation with the caveat that there is no structural model for the free DNA.

2.2 RESULTS AND DISCUSSION

2.2.1 Basic side chain residues stabilize the specific EcoRV-DNA complex

The energetic effects of mutating four basic side chain residues at the EcoRV-DNA interface have been investigated. The residues that were mutated to alanine are K119, R221, and R226, which interact with phosphates on the concave face of the bent DNA, and R140, which interacts with a phosphate on the convex face (Figures 2.2 and 2.4B). Equilibrium binding constants (K_A) for wild-type (wt) and mutant EcoRV were measured for unmodified DNA substrates in the presence of saturating amounts of Ca^{2+} using a nitrocellulose filter binding assay (see Methods). Large binding penalties were observed for all four of the mutant proteins relative to wt EcoRV, ranging from +3.1 to +5.5 kcal/mol (Table 2.1). This corresponds to a 180- to 13000-fold decrease in binding affinity. In support of these findings, Hiller *et al.*, (2008; unpublished) have shown by stopped-flow Förster resonance energy transfer (FRET) techniques using a 14 bp oligonucleotide containing the EcoRV specific recognition sequence GATATC that all four mutants result in a 4- to 20-fold reduction in association rate and a 120 to 1700-fold increase rate

of dissociation relative to wt EcoRV. Further, Wenz *et al.* (1996) have shown that K119A, R226A, and R140A mutants have decreased enzymatic activity as measured by steady state cleavage rates of 20 bp oligonucleotide substrates in the presence of 10 mM Mg^{2+} (Wenz *et al.*, 1996). Interestingly, the R221A mutation was reported to have little effect on K_M or k_{cat} in these assays (Wenz *et al.*, 1996). Taken together, these data suggest that these four basic side chains stabilize the specific EcoRV-DNA complex.

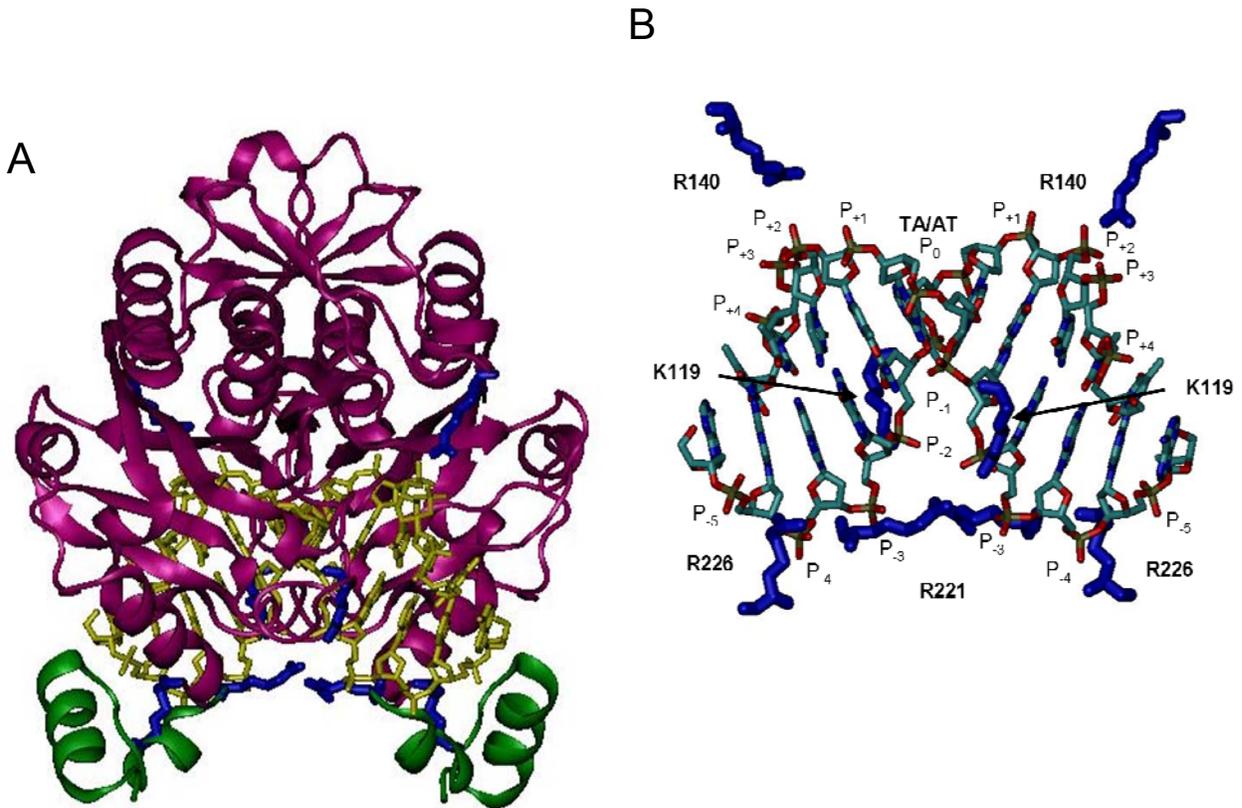


Figure 2.4 Four basic side chain residues at the EcoRV-DNA interface.

(A) Four basic side chain residues (blue) from the EcoRV protein (ribbons) contact each half-site of the bent DNA (yellow) in the EcoRV-DNA complex (1B94). R226 and R221 (blue; labeled in B) are part of the C-terminal domain (green), which becomes ordered upon formation of the specific EcoRV-DNA complex. (B) DNA from the EcoRV-DNA complex is shown (carbons are teal, phosphates are brown, oxygens are red and nitrogens are blue). The majority of the protein has been removed for clarity. Basic side chain residues (blue) that contact the concave (K119, R221, and R226) and convex (R140) face of the bent substrate DNA are shown. Phosphates are also labeled corresponding to the nomenclature in the text and in Figures 2.1 and 2.2.

Table 2.1 Energetic effects of basic residue mutations on EcoRV-DNA complex formation.

Complex^a	KA(M⁻¹)^b	$\Delta\Delta G^{\circ}_{\text{bind}}$ (kcal/mol)^b	Fold Difference^d
wt EcoRV + GATATC	1.0(\pm 0.1) x 10 ¹¹	0	1
CONCAVE			
K119A + GATATC	1.5(\pm 0.1) x 10 ⁷	+5.5(\pm 0.1)	6,500
R221A+ GATATC	5.7(\pm 0.5) x 10 ⁸	+3.1(\pm 0.1)	180
R226A + GATATC	7.9(\pm 0.3) x 10 ⁶	+5.5(\pm 0.1)	13,000
CONVEX			
R140A + GATATC	9.3(\pm 2.0) x 10 ⁷	+4.2(\pm 0.2)	1100

^aThe double stranded 24 bp oligodeoxyribonucleotide d(TGTGTTGTAG**GATATC**CCTACAGGT) was used for all equilibrium binding experiments.
^bEquilibrium binding constants were probed in binding buffer (BB) + 0.24 M KCl, 10 mM CaCl₂, 20 mM cacodylic acid (pH 7.25), 0.01% sodium azide, 100 mM dithiothreitol, and 100 mg/ml BSA at 21°C.
^cRelative to the unmodified substrate; Eqn. 2.5.
^dRelative to the unmodified substrate.

The energetic effects observed for the four basic residue mutants likely derive from multiple sources, including (but not limited to) the removal of favorable phosphate contacts, the removal of stabilizing intraprotein interactions, and increased electrostatic repulsion due to the removal of a positive charge (Table 2.2). Of the four mutations, R226A exacts the largest energetic penalty for binding ($\Delta\Delta G^{\circ}_{\text{bind}} = +5.5$ kcal/mol; Table 2.1). The R226 side chain makes Coulombic interactions with the phosphate backbone (position P₅; Figures 2.1 and 2.4B) and contacts the main chain carbonyl oxygens of E220 and T222 (Figure 2.5). The R226 side chain is also disordered in the free protein, suggesting that these interactions form upon DNA binding and contribute favorable $\Delta G^{\circ}_{\text{bind}}$ to the formation of the wt complex. Further, R226 contributes positive electrostatic potential to the phosphate at position P₅, which is located on the concave face of the bent DNA substrate (Figures 2.2 and 2.4B). Charge neutralization by cationic side chains on the concave face of the bent DNA is expected to relax interphosphate repulsion and

reduce the energy required to bend DNA, thus contributing favorable free energy to complex formation. Eliminating the potential for protein-phosphate and intraprotein interactions, as well as removing the positive charge that relieves interphosphate repulsion on the concave face of the bent substrate DNA, is expected to contribute unfavorably to $\Delta\Delta G^{\circ}_{\text{bind}}$.

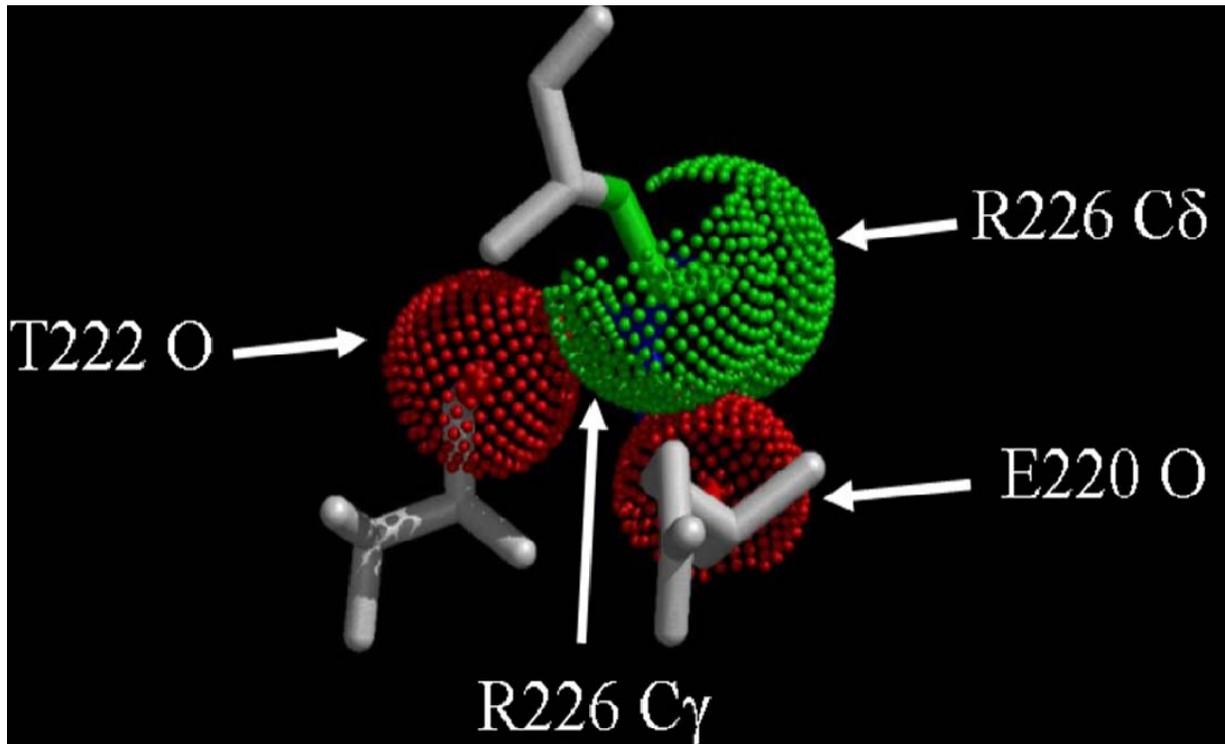


Figure 2.5 Intraprotein interactions to the R226 side chain .

The R226 side chain methylenes C γ and C δ (green) form intraprotein interactions to the main chain carbonyl oxygens (red) of E220 and T222.

There is also a large energetic penalty in binding observed for the K119A mutation ($\Delta\Delta G_{\text{bind}}^{\circ} = 5.2$ kcal/mol; Table 2.1). In the specific EcoRV-DNA complex, K119 packs against several neighboring protein residues, including T111, S112, and N116. Molecular dynamics simulations with the free EcoRV enzyme (see Section 3.2.2) show that these interactions are observed in both the free protein and in the K119A-DNA complex, suggesting that they do not contribute to $\Delta G_{\text{bind}}^{\circ}$. In the wt complex, K119 forms a stable interaction with the P₋₂ phosphate upon DNA binding (Figure 2.1 and 2.4B). The P₋₂ phosphate is located on the concave face of the bent DNA substrate, more proximal to the apex of the DNA bend relative to P₋₅ (Figure 2.2), where it is expected that interphosphate repulsion is more severe. Thus, the unfavorable effects of the K119A mutation likely derive from removing a phosphate contact to P₋₂ in each half-site and increasing interphosphate repulsion that accumulates upon DNA bending.

Of the four mutations, R221A exacts the smallest energetic penalty for binding relative to formation of the wt complex. This agrees qualitatively with the observation of Wentz *et al.* (1996) that R221A does not reduce the activity of EcoRV nearly as much as the other three mutations (K119A, R226A, and R140A). The R221 side chain makes a Coulombic interaction to the phosphate backbone (Figure 2.1) but is not involved in any intraprotein interactions, which is consistent with the observation that this side chain is rather flexible and can sample a large range of conformational space during a 6 nanosecond (ns) MD simulation (Section 3.2.3). MD simulations also suggest that R221 contacts a phosphate on the concave face of the bent DNA, at either P₋₂ (corresponding to the contact observed in the X-ray crystal model; Figures 2.1 and 2.4B) or P₋₃, which implies a role for R221 in relieving interphosphate repulsive forces on this face of the bent DNA. The transient nature of the R221 phosphate contact may reflect a lesser role for this interaction in stabilizing the EcoRV-DNA complex, which might explain the

reduced energetic cost observed for the formation of this mutant complex relative to that observed for the three other mutants examined.

The R140 side chain contacts the convex face of the bent DNA substrate in the EcoRV-DNA complex, where interphosphate repulsive forces are thought to contribute to the compressive force that drives DNA bending toward an asymmetrically neutralized face. Thus, increased interphosphate repulsion due to the R140A mutation is expected to contribute favorably to complex formation by increasing the driving force for DNA bending. This favorable contribution is in opposition to the unfavorable effect of removing a phosphate contact in each half-site. In addition, the R140 side chain forms an intersubunit interaction with the side chain of Q69, such that R140:NH₂ is 2.53 Å from Q69:NE₂. The R140A mutation may abolish the potential for this interaction and destabilize the specific EcoRV-DNA complex, contributing to the observed energetic penalty of +4.2 kcal/mol (Table 2.1).

Table 2.2 Origins of the energetic penalties observed for mutant EcoRV complex formation.

Complex	$\Delta\Delta G^{\circ}_{\text{bind}}$ (kcal/mol)	Possible Origins
Wt-GATATC	0.0	
R226A-GATATC	+5.5(±0.1)	<ul style="list-style-type: none"> • Removal of Coulombic interactions (P₅) (Unfavorable) • Removal of intraprotein interactions (Unfavorable) • Increased electrostatic repulsion (Unfavorable)
K119A-GATATC	+5.2(±0.1)	<ul style="list-style-type: none"> • Removal of Coulombic interactions (P₂) (Unfavorable) • Increased electrostatic repulsion (Unfavorable)
R221A-GATATC	+3.1(±0.1)	<ul style="list-style-type: none"> • Removal of Coulombic interactions (P₂/P₃ ?) (Unfavorable) • Increased electrostatic repulsion (Unfavorable)
R140A-GATATC	+4.2(±0.2)	<ul style="list-style-type: none"> • Removal of Coulombic interactions (P+3) (Unfavorable) • Removal of intersubunit stabilizing interaction to Q69 (Unfavorable) • Removal of electrostatic repulsion (convex face) (Favorable?)

2.2.2 Single P_{Me} modifications impact EcoRV-DNA complex formation

In order to investigate the role of phosphate neutralization via methylphosphonate (P_{Me}) substitutions in wt EcoRV complex formation, DNA substrates were synthesized that contain a single racemic P_{Me} substitution (replaces the pro-S and pro-R phosphoryl oxygens with a methyl group with equal probability) at particular phosphates within a 24 base pair (bp) parent harboring the specific recognition sequence GATATC. I also performed experiments using stereospecific (Rp and Sp) substitutions for each of the non-esterified phosphoryl oxygens at the P₂ position. P_{Me} substitutions were made in each strand of the duplex substrate and were symmetrically disposed about the apex of the DNA bend (Figures 2.2 and 2.6B). The energetic effects of these

substitutions were probed by equilibrium filter binding techniques. P_{Me} produced a range of energy differentials [from 0.0(\pm 0.1) to +4.1(\pm 0.1) kcal/mol; Table 2.3; Figure 2.6]. This reflects the unique nature of each of the phosphates in terms of 1) the number and type of phosphate contacts made at each position (Figure 2.1), as well as 2) the three-dimensional juxtaposition of each phosphate relative to the DNA bend (Figure 2.2). Large energetic penalties were observed for the P_{Me-2} (racemic and stereospecific), P_{Me-4} and P_{Me+3} substitutions ($\Delta\Delta G^{\circ}_{bind} = +3.0$ to +5.1 kcal/mol; Table 2.3; Figure 2.6) suggesting that the contacts made to these phosphates are important for the stability of the specific EcoRV-DNA complex. By contrast, much smaller energetic penalties were observed for the P_{Me-3} , P_{Me-5} , P_{Me+1} and P_{Me+2} substitutions ($\Delta\Delta G^{\circ}_{bind} = 0.0$ to +1.6 kcal/mol; Table 2.3; Figure 2.6).

The energetic effects observed for particular P_{Me} substitution agree qualitatively with observations made by others. Ethylation interference footprinting has shown that ethylation of phosphates at P_{-1} , P_{+1} and the scissile phosphate P_0 (not probed in this investigation) dramatically perturbs EcoRV-DNA complex formation, while a smaller but still very dramatic effect is observed for ethylation at P_{-4} , P_{-2} , and P_{+3} (Engler et al., 1997). Further, work done in collaboration with the lab of Dr. John Perona at the University of California, Santa Barbara, has shown that that racemic P_{Me} substitutions at P_{-4} (P_{Me-4}), P_{-2} , and P_{+3} result in a 2 to 6.5-fold reduction in association rate and a much more dramatic, 200 to 1000-fold *increase* in dissociation rate, as measured by stopped flow FRET (Hiller et al., 2008). These data are in agreement with our equilibrium observations that P_{Me} modifications at these positions destabilize the EcoRV-DNA complex. In accordance with the findings reported here, both studies report very modest effects associated with either ethylation or P_{Me} substitution at the P_{-5} phosphate (Engler, Welch et al, 1997; Hiller, 2008).

Table 2.3 Effects of single P_{Me} modifications on EcoRV-DNA pre-transition state complex formation.

Substrate^a	K_A(M⁻¹)^b	ΔΔG^o_{bind} (kcal/mol)^c	Fold difference^d
GATATC	1.0(±0.1)x10 ¹¹	0	1
CONCAVE			
P _{Me-5}	9.1(±1.0)x10 ¹⁰	0(±0.1)	1
P _{Me-4}	6.3(±1.3)x10 ⁸	+3.0(±0.1)	160
P _{Me-3}	8.3(±1.0)x10 ⁹	+1.5(±0.1)	12
P _{Me-2}	8.4 (±1.3)x10 ⁷	+4.1(±0.1)	1200
Rp-P _{Me-2}	2.4(±0.5)x10 ⁸	+3.5(±0.1)	420
Sp-P _{Me-2}	3.3(±0.2)x10 ⁷	+4.7(±0.1)	3000
CONVEX			
P _{Me+1}	9.01 (±1.1)x10 ⁹	+1.4(±0.1)	11
P _{Me+2}	4.3(±0.3)x10 ¹⁰	+0.5(±0.1)	2
P _{Me+3}	1.51(±0.35)x10 ⁷	+5.1(±0.1)	6600

^a The double stranded 24 bp oligodeoxyribonucleotide d(TGTGTTGTAGGATATCCTACAGGT) was used for all equilibrium binding experiments.

^b Equilibrium binding constants were probed in binding buffer (BB) + 0.24 M KCl, 10 mM CaCl₂, 20 mM cacodylic acid (pH 7.25), 0.01% sodium azide, 100 mM dithiothreitol, and 100 mg/ml BSA at 21°C.

^c Relative to the formation of the unmodified complex; Eqn. 2.5.

^d Relative to the formation of the unmodified complex.

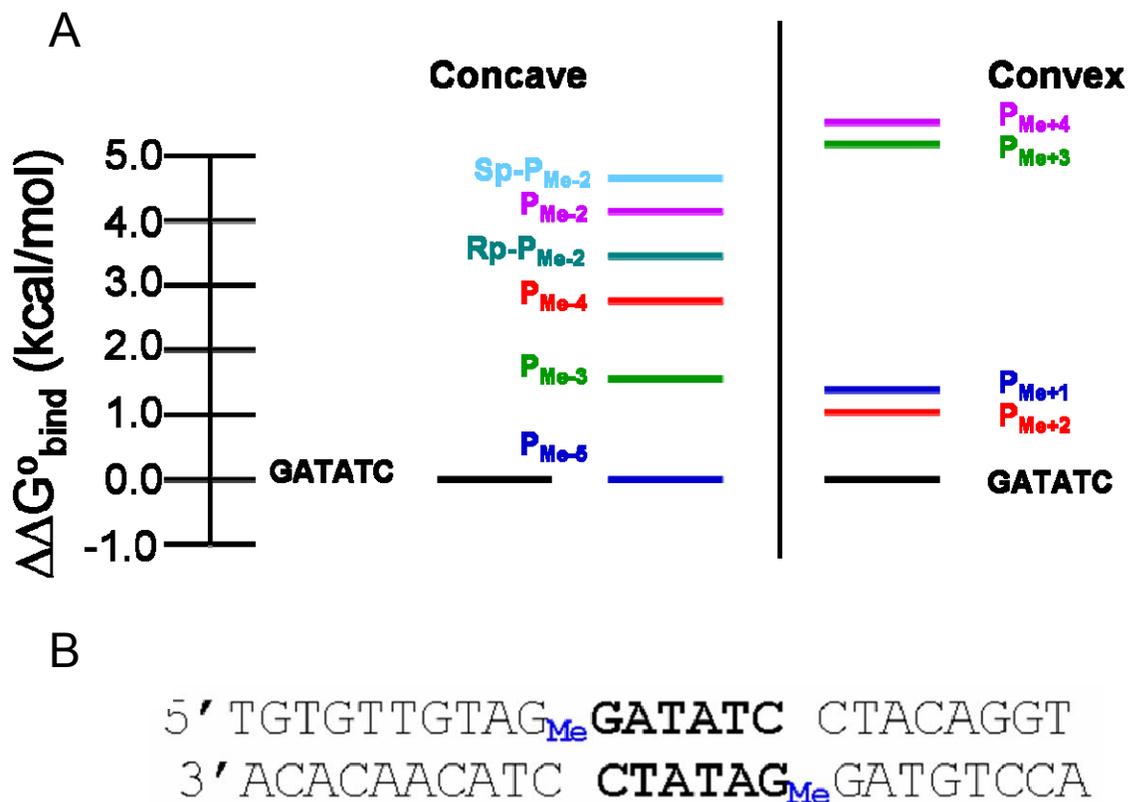


Figure 2.6 Free energy diagram for EcoRV binding to single P_{Me} substituted substrates.

Difference in free energy change for binding to 24 bp oligonucleotides containing the EcoRV specific recognition site GATATC, which are either unmodified or harbor a single P_{Me} substitution at the position specified. Note that P_{Me} substitutions are made on both strands of the double stranded substrate and are disposed symmetrically about the apex of the bend (Figures 2.3 and 2.5B). $\Delta\Delta G^{\circ}_{bind}$ was calculated using Equation 2.5 with unmodified substrate as the reference (K_{unmod}). (B) P_{Me} substitutions are made on both strands of the double stranded substrate and are disposed symmetrically about the apex of the bend (Figure 2.2).

$\Delta\Delta G_{\text{bind}}^{\circ}$ observed for single P_{Me} substitutions in forming the wt EcoRV-DNA complex is likely the net of multiple factors, such as removal of phosphate contacts, destabilization of bound waters, disruption of intraprotein interactions, modified polar and nonpolar solvation, as well as removal of phosphate charge. For example, the phosphate at position P_{-2} the pro-S phosphoryl oxygen is contacted by K119 while the pro-R phosphoryl oxygen is contacted by R221 (transiently) and T111, and is part of a water mediated network of contacts that bridges all of the phosphates of the specific recognition sequence (Figure 2.1). A racemic $P_{\text{Me-2}}$ substitution results in the disruption of the contacts to the pro-S phosphoryl oxygen in 50% of the complexes and disruption of the contacts to the pro-R phosphoryl oxygen in the other 50%, so that the penalty for the racemic modification reflects a rather complex combination of perturbations. To de-convolute the effects of the $P_{\text{Me-2}}$ substitution, stereospecific P_{Me} substitutions were used (See Figure 1.10 for P_{Me} stereochemistry and naming convention). The Rp- $P_{\text{Me-2}}$ (replaces only the pro-S phosphoryl oxygen with a methyl group) substitution results in a less unfavorable energetic penalty relative to the racemic substitution (Table 2.3; Figure 2.6), reflecting the fact that perturbing the interaction with K119 exacts a smaller energetic penalty than perturbing those with T111, R221 and water-mediated contact to T187 (Figure 2.1). The unfavorable consequence of disrupting the interactions to the pro-R phosphoryl oxygen is highlighted by the large penalty observed for the Sp- $P_{\text{Me-2}}$ modification, which is more severe than either the Rp or racemic $P_{\text{Me-2}}$ substitutions (Table 2.3; Figure 2.6). The unfavorable effects of removing protein-phosphate contacts likely mask the effects of phosphate neutralization at P_{-2} , which is proposed to reduce electrostatic repulsive forces and stabilize DNA bending

The racemic $P_{\text{Me-3}}$ substitution results in an unfavorable contribution to wt complex formation ($\Delta\Delta G_{\text{bind}}^{\circ} = +1.5$ kcal/mol), while there is no penalty observed for this substitution in

R221A complex formation ($\Delta\Delta G_{\text{bind}}^{\circ}$ 3.1 kcal/mol vs. 3.3 kcal/mol; Tables 2.3 and 2.4; Figures 2.6 and 2.7). The fact that the $P_{\text{Me-3}}$ substitution makes an unfavorable contribution to $\Delta G_{\text{bind}}^{\circ}$ was unexpected because upon bending, the largest decrease in interphosphate distance occurs between the P_{-3} phosphates (from 18.5 Å in B-form DNA to 13.8 Å in the bent DNA). One would have presumed that interphosphate repulsive forces would be significant at this phosphate position and its neutralization would yield significant energetic benefits. The R221 side chain is able to transiently interact with both P_{-2} and P_{-3} phosphates in the unmodified EcoRV-DNA complex. Thus, the $P_{\text{Me-3}}$ substitution would abolish the potential for the R221- P_{-3} interaction and may render the R221 side chain more constrained to the P_{-2} phosphate, resulting in an unfavorable reduction in conformational freedom. The observation that there is no penalty for complex formation in absence of the R221 side chain supports the existence of a favorable interaction between R221 and P_{-3} , but it is still puzzling as to why there is not an energetic benefit for the $P_{\text{Me-3}}$ substitution in R221A binding. It may be that there is an additional unfavorable energetic consequence associated with the $P_{\text{Me-3}}$ substitution that has not been identified.

The $P_{\text{Me-5}}$ substitution does not affect wt complex formation (Table 2.3, Figure 2.6). It may be that the benefits of phosphate neutralization balance the unfavorable effect of removing the Coulombic interaction between the P_{-5} phosphate and the R226 side chain (Figure 2.1), such that there is no net effect on $\Delta\Delta G_{\text{bind}}^{\circ}$. Alternatively, the effect of removing this Coulombic interaction may be minimal, in which case, the large energetic penalty observed for the R226A mutation (Table 2.1) would derive from the loss of the interprotein interactions with T222:O and E220:O, as well as the increase in electrostatic repulsion on the concave face of the bent DNA.

2.2.3 Phosphate neutralization contributes favorable $\Delta G^{\circ}_{\text{bind}}$ to EcoRV-DNA complex formation

Charge neutralization by cationic side chains on the concave face of the bent DNA is expected to relax interphosphate repulsion and reduce the energy required to bend the substrate DNA. Thus, phosphate neutralization on the concave face of the bent DNA substrate via P_{Me} substitution would be predicted to rescue some of the unfavorable effects of the K119A, R221A, and R226A mutations. Therefore, the effect of P_{Me} substitution on mutant EcoRV-DNA binding was probed, again using an equilibrium nitrocellulose filter binding assay (see Methods). In all cases, P_{Me} substitutions were made to both strands of the duplex DNA substrate, resulting in symmetrically disposed substitutions about the apex of the DNA bend (Figure 2.7).

Both racemic and stereospecific substitution at position P_{-2} increased K119A mutant binding affinity 14- and 50-fold respectively, relative to unmodified DNA. This corresponds to a favorable contribution to $\Delta\Delta G^{\circ}_{\text{bind}}$ of -1.5 and -2.3 kcal/mol, respectively (Table 2.4; Figure 2.7). The stereospecific $R_p\text{-}P_{\text{Me-}2}$ substitution is a more precise probe for the effects of phosphate neutralization at the P_{-2} position, as this modification replaces only the pro-S phosphoryl oxygen with a methyl group and is predicted to conserve the contacts made to the pro-R phosphoryl oxygen (Figure 2.1), whereas a racemic $P_{\text{Me-}2}$ substitution results in the loss of these interactions in 50% of the complexes. Taken together, these data suggest that phosphate neutralization at the P_{-2} position contributes as much as -2.3 kcal/mol to EcoRV-DNA complex formation.

Similarly, a racemic $P_{\text{Me-}5}$ substitution increases the binding affinity for the R226A mutant protein > 4 -fold relative to unmodified DNA substrates (Table 2.4; Figure 2.7). Note that there are no contacts made to the pro-S phosphoryl oxygen at P_{-5} , while in addition to the direct

hydrogen bond made by R226, Y219 makes a water-mediated contact to the pro-R phosphoryl oxygen (Figure 2.1). The energetic effect of this modification may derive from a combination of phosphate neutralization and the removal of this water-mediated contact to the pro-R phosphoryl oxygen. Regardless, these data suggest that phosphate neutralization at position P₅ contributes at least -0.9 kcal/mol to EcoRV-DNA complex formation.

In contrast, both racemic P_{Me-2} and stereospecific Sp-P_{Me-2} substitutions reduce R221A binding affinity by 90- and 250-fold, respectively (Table 2.4; Figure 2.7). This is likely due to the additional loss of favorable contacts to the P₂ phosphate by K119 for the racemic substitution, and by T111 and T187 (mediated through a stably bound water) for the Sp-P_{Me-2} substitution (Figure 2.1). Molecular dynamics simulations suggest that the Sp-P_{Me-2} substitution does not result in a large scale disruption of the water mediated network of contacts which includes the pro-R phosphoryl oxygen at P₂ and bridges all of the phosphates within the specific recognition sequence (not shown), suggesting that the severe energetic penalties for binding derive largely from the removal of phosphate contacts. Interestingly, the racemic P_{Me-5} substitution improves R221A binding by 2-fold, while it had no effect on wt binding (Table 2.4; Figure 2.7). In mutating the R221 side chain to alanine, an interaction to the phosphate backbone at P₂ or P₃ is removed, and interphosphate repulsion has likely been increased. The fact that the P_{Me-5} substitution contributes favorable free energy to complex formation suggests that phosphate neutralization at P₅ affects the global charge landscape on the concave face of the substrate DNA, resulting in an overall reduction in electrostatic repulsive forces. Further they suggest a role for the R221 in the neutralization of phosphates at the EcoRV-DNA interface in that there is a larger energetic benefit for phosphate neutralization in the absence of the R221 side chain.

In order to investigate the role of phosphate neutralization on the convex face of the bent DNA, I have examined the effects of phosphate neutralization at the P₊₃ position that is contacted by R140 in the unmodified complex (Figures 2.1 and 2.2). The R140A mutation is expected to increase interphosphate repulsive forces on the convex face of the DNA, which is expected to favor DNA bending, while a P_{Me} substitution at this position is expected to reduce interphosphate repulsion. The racemic P_{Me+3} substitution reduces R140A binding 11-fold (Table 2.4; Figure 2.7), an effect that likely results from either perturbing the interaction between the pro-S phosphoryl and the Y95 side chain (Figure 2.1) or relaxing interphosphate repulsion on the convex face of the substrate. These two possibilities are difficult to distinguish for the racemic P_{Me+3} substitution and would require a stereospecific Rp-P_{Me+3} substitution to resolve.

Table 2.4 Phosphate neutralization contributes favorable free energy to EcoRV-DNA complex formation.

Complex^a	K_A(M⁻¹)^b	ΔΔG^o_{bind} (kcal/mol)^b	Fold Difference^d
wt-EcoRV + GATATC	1.0(±0.1)x10 ¹¹	0	1
CONCAVE^e			
K119A + GATATC	1.5(±0.1)x10 ⁷	+5.2(±0.1)	6,500
K119A + P _{Me-2}	2.1(±0.1)x10 ⁸	+3.7(±0.1)	476
K119A + Rp-P _{Me-2}	7.7(±1.4)x10 ⁸	+2.9(±0.2)	133
R221A+ GATATC	5.7(±0.5)x10 ⁸	+3.1(±0.1)	180
R221A + P _{Me-2}	6.1(±3.0)x10 ⁶	+5.7(±0.4)	16,393
R221A + Sp-P _{Me-2}	2.2(±0.2)x10 ⁶	+6.3(±0.1)	45,454
R221A + P _{Me-3}	3.7(±0.2)x10 ⁸	+3.3(±0.1)	270
R221A + P _{Me-5}	1.2(±0.2)x10 ⁹	+2.6(±0.1)	80
R226A + GATATC	7.9(±0.3)x10 ⁶	+5.5(±0.1)	13,000
R226A + P _{Me-5}	3.6(±0.5)x10 ⁷	+4.6(±0.4)	2,800
CONVEX^e			
R140A + GATATC	9.3(±2.0)x10 ⁷	+4.2(±0.2)	1100
R140A + P _{Me+3}	8.2(±0.2)x10 ⁶	+5.5(±0.4)	12,200

^aThe double stranded 24 bp oligodeoxyribonucleotide d(TGTGTTGTAGGATATCCTACAGGT) was used for all equilibrium binding experiments.

^bEquilibrium binding constants were probed in binding buffer (BB) + 0.24 M KCl, 10 mM CaCl₂, 20 mM cacodylic acid (pH 7.25), 0.01% sodium azide, 100 mM dithiothreitol, and 100 mg/ml BSA at 21°C.

^cRelative to the formation of the unmodified complex; Eqn. 2.5.

^dRelative to the formation of the unmodified complex.

^eConcave and convex faces of the DNA in the complex.

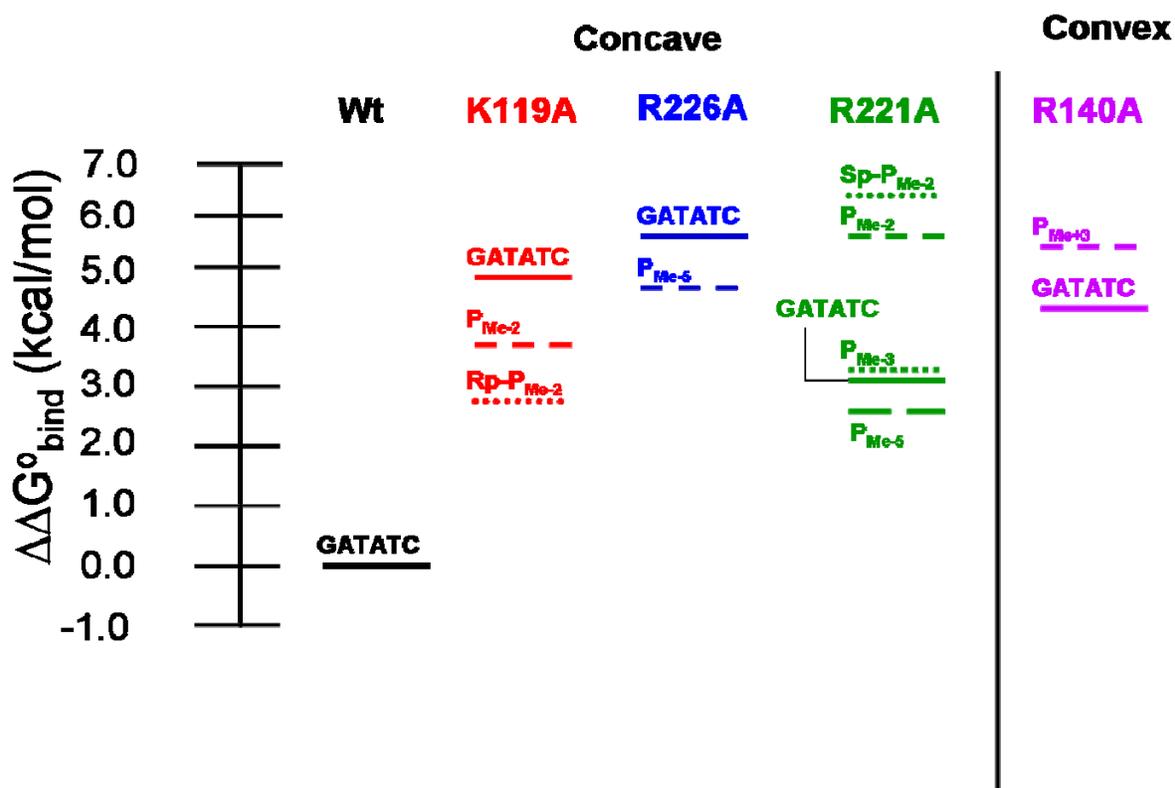


Figure 2.7 Free energy diagram for EcoRV mutants binding to modified and P_{Me} substituted substrates.

Difference in free energy change for wt (black) and mutant (red, blue, green and pink) EcoRV binding to 24 bp oligonucleotides containing the EcoRV specific recognition site GATATC, which was either unmodified (GATATC) or contained a single P_{Me} substitution at the position specified (see Figure 2.1 for phosphate positions). The position of each bar represents the change in binding free energy associated with the formation of each EcoRV-DNA complex relative to that for the unmodified complex. Each bar is labeled corresponding to the substrate which it represents. $\Delta\Delta G^{\circ}_{\text{bind}}$ was calculated using wt-GATATC complex formation as a reference (Equation 2.5).

2.2.4 Electrostatic modifications affect the dependence of EcoRV-DNA binding on salt concentration.

It is well established that protein-DNA interactions are strongly dependent on salt concentration such that binding affinity is reduced as salt concentration increases (Record Jr et al., 1991). Dependence of K_A on salt concentration is typically interpreted in terms of two competing models; the counterion condensation theory (Manning, 1977; Manning, 1978) and a model based on using the non-linear Poisson Boltzmann equation to solve for the electrostatic potential over the surface of a macromolecule (Sharp et al., 1995). For the purposes of this discussion I will focus on interpretation in terms of the counterion condensation theory.

Counterion condensation theory treats DNA as a rigid anionic rod-like species that is surrounded by two distinct layers of counterions (M^+). A concentrated layer of hydrated counterions “condense” around the DNA surface (Manning, 1978). The concentration of condensed counterions approaches 1.0 M $[M^+]$ and is *independent* of the bulk $[M^+]$, above 0.02M. This condensed layer of counterions is surrounded by an outer Debye-Hückel layer of counterions, which is treated as a classical ionic environment and is *dependent* on $[M^+]$. Thus, the dependence of K_A on $[M^+]$ derives from the differential entropy associated with releasing counterions from the condensed counterion layer into the bulk solvent. At low $[M^+]$ the entropy gained from this process is more favorable relative to that at high $[M^+]$. The number of thermodynamically bound cations released (or salt-links formed) upon protein binding can be estimated from a plot of $\log K_A$ vs. $\log [\text{salt}]$ according to equation.

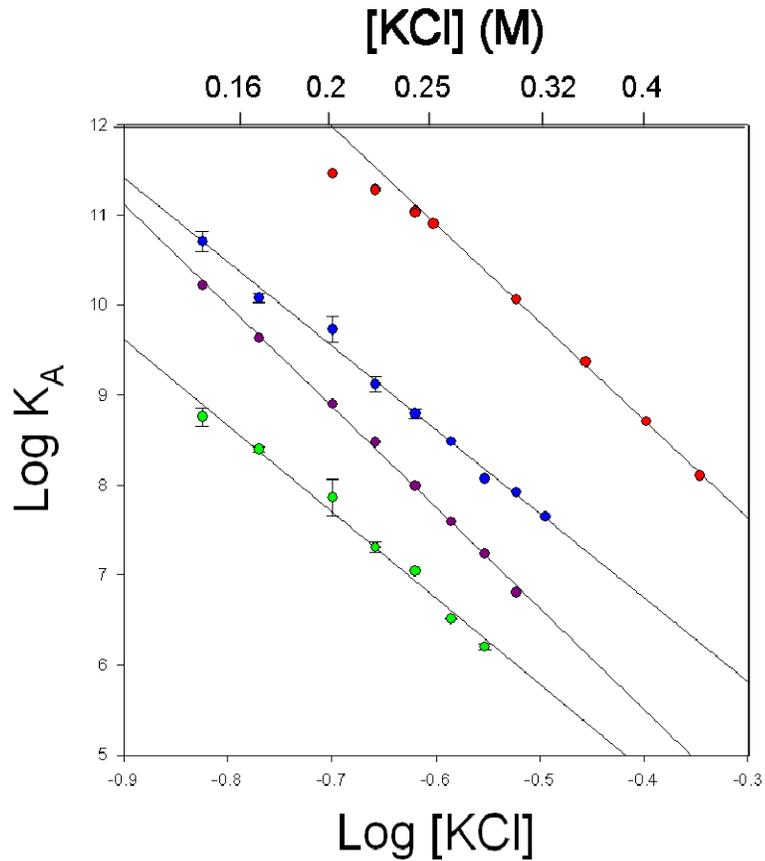
$$\text{Slope} = m^* \psi, \quad (2.9)$$

where m' is the number of cations released upon protein binding, and $\psi = 0.88$, referring to the number of cations thermodynamically bound to each phosphate.

In order to investigate how particular electrostatic perturbations affect specific EcoRV-DNA complex formation, I have measured the dependence of binding affinity on salt concentration for the formation of various EcoRV-DNA complexes including the unmodified complex, as well as complexes containing the K119A mutation, the Rp-P_{Me-2} substitution, or both of these modifications (Figure 2.8). The dependence of the formation of the wt-GATATC complex on increasing [KCl] shows a slope of $-10.9 (\pm 0.2)$. Applying Equation 2.9 to this slope suggests that approximately $12.4 (\pm 0.2)$ thermodynamically bound cations are released upon specific EcoRV-DNA complex formation. Additional data from the Jen-Jacobson lab show that the dependence of the formation of the wt-GATATC complex on increasing [NaCl] has a slope of $-11.3 (\pm 0.2)$, suggesting that $12.8 (\pm 0.2)$ thermodynamically bound cations are released upon complex formation (Pham, O and Jen-Jacobson, L., unpublished). This is similar to the number of thermodynamically bound cations released upon formation of the non-specific EcoRV-DNA complex (slope = $-11.2 (\pm 0.5)$; equivalent to the release of approximately $12.7 (\pm 0.5)$ thermodynamically bound cations; Engler et al., 1997; Kurpiewski and Jen-Jacobson, unpublished), highlighting the importance of Coulombic interactions between positively charged protein side chains and DNA phosphates for the stability of *both* specific and non-specific EcoRV-DNA complexes.

The K119A-GATATC complex shows a slope of $-9.6 (\pm 0.4)$ suggesting that the formation of this complex is accompanied by the release of $1.5 (\pm 0.4)$ fewer cations than for formation of the unmodified complex relative to that expected for a slope of $10.9 (\pm 0.4)$. This observation is consistent with the fact that this mutation results in the removal of approximately

two stable phosphate contacts between the K119 side chain and the P₂ phosphate. Interestingly, the Rp-P_{Me-2} modification has very little effect on the salt dependence for the formation of either the wt or K119A complexes. For the formation of the K119A-GATATC complex, the salt dependence is equal to $-9.6 (\pm 0.4)$ while that for the formation of the K119A-Rp-P_{Me-2} complex is $-9.3 (\pm 0.2)$. Similarly, the slope of salt depended for the formation of the wt-GATATC complex is $-10.9 (\pm 0.2)$ while that for the formation of the wt-Rp-P_{Me-2} complex is $-11.2 (\pm 0.2)$. This observation suggests that the Rp-P_{Me-2} modification does not affect the number of monovalent cations that are released upon EcoRV-DNA complex formation. Further, this observation shows that all the comparisons for column 2, Fig. 2.7 (K119A) have the same salt dependence such that that the magnitude of the energetic benefits of phosphate neutralization on EcoRV-DNA complex formation via Rp-P_{Me-2} substitution ($\Delta\Delta G^{\circ}_{\text{bind}} = -2.3$; Table 2.4, Figure 2.7) is the same at all salt concentrations tested (Figure 2.8). This is in contrast to the effects of phosphate neutralization on the magnitude of free DNA bending which has been shown to be quite dependent on the concentration and valence of cations in solution due to the differential screening of phosphates by these cations (Strauss and Maher, 1994). The fact that the energetic effects of phosphate neutralization are [salt] independent suggests that the electrostatic environment at the EcoRV-DNA interface is largely the result of the electrostatic character of the protein and DNA moieties, and not of the ionic conditions of the solvent (for the range of salt concentrations tested).



- wt-GATATC $m = -10.9 (\pm 0.2)$ ● K119A-GATATC $m = -9.6 (\pm 0.4)$
- wt-Rp-P_{Me-2} $m = -11.2 (\pm 0.2)$ ● K119A-Rp-P_{Me-2} $m = -9.3 (\pm 0.2)$

Figure 2.8 Salt dependence for the formation of modified and unmodified EcoRV-DNA complexes
 Equilibrium association constants (K_A) were measured in binding buffer (pH 7.25 at 21°C) at different KCl concentrations (see x-axis at the top of the figure for actual KCl concentrations). Slopes from linear regression analysis of these data are given in the legend. DNA substrates contained the GATATC site embed in the same parent 24mer that is used for all binding experiments.

2.2.5 Synergistic, additive, and more than additive effects of multiple P_{Me} substitutions

Multiple P_{Me} substitutions at particular sites can result in energetic penalties that are additive, less than additive or more than additive relative to the sum of the energetic penalties that are observed for the individual P_{Me} substitutions for wt EcoRV-DNA complex formation (Table 2.5). Multiple P_{Me} modifications at positions P₂/P₃ and P₃/P₄ show smaller energetic penalties relative to the sum of the penalties associated with each of the individual modifications (Table 2.5). These data suggest both a favorable role for phosphate neutralization at certain positions on the concave face of the bent DNA and that increased phosphate neutralization can result in a synergistic reduction in the cost of DNA bending. Figure 2.9 shows that these phosphate positions are consecutive, close to the apex of the DNA bend, and are located directly across the major groove from one another. These phosphates get “pushed” closer to one another during DNA bending and may represent “hot spots” for the accumulation of inter-phosphate repulsion, suggesting a significant benefit for phosphate neutralization at these positions.

Multiple P_{Me} substitutions at phosphate positions that are non-consecutive or are more distal to the apex of the DNA bend show energetic penalties that are equal to (P₃/P₅), or greater than (P₂/P₅, P₄/P₅, and P₃/P₄/P₅) those expected for the sum of the individual P_{Me} substitutions (Table 2.5). Interestingly, the P_{Me-3}/P_{Me-4}/P_{Me-5} substrate produces a greater than additive energetic penalty while that for the P_{Me-3}/P_{Me-4} substrate produces a less than additive penalty. The only difference between these two substrates is the P_{Me-5} substitution, which shows little effect on wt binding and a favorable contribution to R226A binding. Furthermore, the wt cleavage rate for P_{Me-3}/P_{Me-4}/P_{Me-5} substrate is only 3-fold slower than that for the unmodified substrate (Section 4.2.5; Table 4.3). First order cleavage rates are a sensitive probe for the correct assembly of the EcoRV-DNA interface due to the requirement for the precise geometric

coordination of critical active site residues for catalysis. Thus, it appears that there are no large scale structural adaptations in the wt- P_{Me-3}/P_{Me-4}/P_{Me-5} complex that would explain a greater than additive binding penalty. It may be that there is a threshold beyond which, the benefit for phosphate neutralization is masked by the removal of an increasing number of phosphate contacts.

Table 2.5 Energetic effects of multiple P_{Me} substitutions.

Sequence ^a	$\Delta\Delta G^{\circ}_{\text{bind}}$ (kcal/mol) ^b	$\Delta\Delta G^{\circ}_{\text{bind}}$ Predicted (additivity) ^c
P _{Me-5}	0 (±0.1)	-
P _{Me-4}	+3.0(±0.1)	-
P _{Me-3}	+1.5(±0.1)	-
P _{Me-2}	+4.1(±0.1)	-
P _{Me-2} /P _{Me-3}	+5.0(±0.3)	+5.7(±0.1)
P _{Me-3} /P _{Me-4}	+3.9(±0.2)	+4.5(±0.1)
P _{Me-3} /P _{Me-5}	+1.7(±0.1)	+1.5(±0.1)
P _{Me-4} /P _{Me-5}	+3.5(±0.2)	+3.0(±0.1)
P _{Me-2} /P _{Me-5}	+5.8(±0.2)	+4.4(±0.1)
P _{Me-3} /P _{Me-4} /P _{Me-5}	+5.2(±0.1)	+4.5(±0.1)

Example:

Multiple substitution	Single substitutions	$\Delta\Delta G^{\circ}_{\text{bind}}$ (single) kcal/mol	Predicted $\Delta\Delta G^{\circ}_{\text{bind}}$ (kcal/mol)	Observed $\Delta\Delta G^{\circ}_{\text{bind}}$ (kcal/mol)
P _{Me-3} /P _{Me-4}	P _{Me-3}	+3.0(±0.1)	+4.5(±0.1)	+3.9(±0.2)
	P _{Me-4}	+1.5(±0.1)		

^aThe double-stranded 24 bp oligodeoxyribonucleotide used was d(TGTGTTGTAGGATATCCTACAGGT for all experiments; P_{Me-x} denotes the position of substitution.

^bFor each modified complex, the observed difference in standard binding free energy between the modified and unmodified (as reference) complexes is calculated as $\Delta\Delta G^{\circ}_{\text{bind}} = -RT\ln(K_{A,\text{mod}}/K_{A,\text{unmod}})$ at 294K. Equilibrium binding constants were measured in buffer +0.24 M KCl, 10 mM CaCl₂ and pH 7.25 at 21°C.

^c $\Delta\Delta G^{\circ}_{\text{bind}}$ Predicted (additivity) is calculated as the addition of $\Delta\Delta G^{\circ}_{\text{bind}}$ for each of the individual P_{Me} modifications (see example; last two rows).

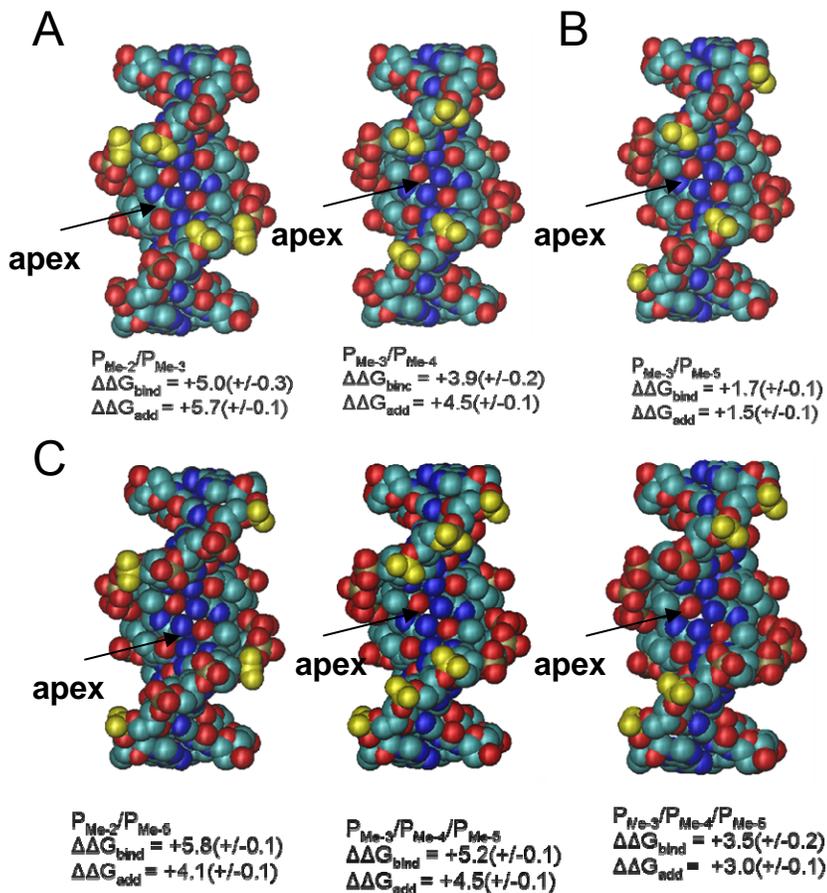


Figure 2.9 Modified phosphate positions in multiple P_{Me} substituted substrates.

The position of P_{Me} modifications (yellow) are shown for substrates with multiple P_{Me} substitutions. The direction of DNA bending in this view is toward the reader, and its apex is at the GATATC base step indicated by “apex” above. DNA substrates are shown as space filling models in which carbon (cyan), oxygen (red), nitrogen (blue), and phosphorous (brown) atoms are highlighted. The experimentally observed energetic effects of multiple P_{Me} modifications ($\Delta\Delta G_{bind}$) and those expected for the addition of the individual P_{Me} modifications ($\Delta\Delta G_{add}$) are shown. Multiple P_{Me} substitutions can result in less than additive (A), additive (B) and greater than additive (C) penalties for complex formation.

2.2.6 Position dependence for the benefits of phosphate neutralization on specific EcoRV-DNA complex formation

Phosphate neutralization via P_{Me} substitution has a greater effect on complex formation at phosphate positions nearer to the apex of the bend in the bound specific DNA substrate (Figure 2.2). Specifically, the magnitude of energetic benefit observed for phosphate neutralization at P_{-2} ($\Delta\Delta G_{bind}^{\circ} = -2.3$ kcal/mol) is greater than that for phosphate neutralization at P_{-5} ($\Delta\Delta G_{bind}^{\circ} = -0.9$ kcal/mol) for K119A and R226A complex formation, respectively (Table 2.2, Figure 2.6A). It is attractive to propose that this observation is due to increased electrostatic repulsion (strain) at P_{-2} in the K119A complex relative to that at P_{-5} in the R226A complex, resulting in a larger energetic benefit for phosphate neutralization at P_{-2} . Investigation of the EcoRV-DNA- Ca^{2+} ternary crystal model reveals that the major groove of the bent DNA substrate becomes more narrow, while phosphates across the major groove are pushed close to one another (Winkler, 1993), suggesting an increase in interphosphate repulsion. Specifically, the P_{-3} phosphate moves 4.7 Å closer to the P_{-3} phosphate on the opposite DNA strand located across the major groove (18.5 Å to 13.8 Å). Due to the proximity of the P_{-2} phosphate to P_{-3} , it is likely that there is a larger buildup of unfavorable electrostatic repulsion at P_{-2} relative to P_{-5} , which is located much more distally; this is consistent with a greater benefit for phosphate neutralization at P_{-2} .

Further support for the position dependent effect of phosphate neutralization on EcoRV-DNA complex formation comes from the analysis of the effects of multiple P_{Me} substitutions on wt EcoRV-DNA complex formation. It has been shown that certain combinations of P_{Me} substitutions show a reduced energetic penalty for wt complex formation relative to the sum of the energetic penalty for the individual substitutions (Table 2.5; Figure 2.9). This effect is specific to the P_{Me-2}/P_{Me-3} , and P_{Me-3}/P_{Me-4} substrates, both of which contain P_{Me} substitutions

made at consecutive phosphates and include the P₃ phosphate. When multiple substitutions are made at phosphates that are nonconsecutive or are more distal to the apex of the DNA bend, no such synergistic effect is observed (Table 2.5; Figure 2.9). Two possibilities may explain these observations. First, the consecutive P_{Me} modifications may promote free DNA bending, resulting in a DNA conformation that more closely resembles that of the bound DNA substrate, such that the amount of DNA bending required for complex formation is lessened. Second, the cost for fully bending the substrate DNA may be cooperatively lessened due to neutralization of a consecutive patch of phosphates rather than a single phosphate or diffuse phosphates.

Strauss and Maher (Strauss and Maher, 1994) have shown that neutralization of three consecutive phosphates per DNA strand, disposed across the minor groove of a free DNA molecule, will induce a spontaneous DNA bend of 20° toward the neutralized face; however, when these substitutions are made non-consecutively, the induced bend is only 13° (Strauss-Soukup, 1997). In these experiments, P_{Me} substitutions were made across the minor groove of a free DNA molecule. In the experiments considered here, phosphate neutralization is across the major groove. Kosikov *et al.* (2002) have shown by computational methods that neutralization of six phosphates, three per strand disposed across the major groove, results in a larger degree of spontaneous DNA bending than when disposed across the minor groove. This is likely due to the fact that there is more room in the major groove, and bending toward this groove is not as opposed by steric constraints as it is toward the minor groove. Taken together, the above observations suggest that multiple P_{Me} substitutions at consecutive positions near the site of the EcoRV-induced bend may affect the shape of the free DNA substrate, such that it is pre-bent in the direction observed in the EcoRV-DNA complex, thus requiring less DNA bending by the EcoRV protein to achieve the full DNA bend. This would result in a more favorable $\Delta G^{\circ}_{\text{bind}}$ for

these substrates relative to those with single P_{Me} substitutions or those with multiple P_{Me} substitutions that are made more distally to the site of DNA bending. However, it is unclear to what extent our multiple P_{Me} substitutions would impact the free DNA molecules. Analysis of free DNA bending for molecules with or without P_{Me} substitutions would be required to investigate the possibility of P_{Me-2}/P_{Me-3} and/or P_{Me-3}/P_{Me-4} induced bending of free DNA molecules. Regardless, the observation that multiple P_{Me} substitutions show synergistic effects for the formation of the EcoRV-DNA complex is consistent with the hypothesis that neutralization of a cluster of appropriately positioned phosphates can relax the energetic cost required for the compression of these phosphates due to DNA bending by EcoRV.

Taken together the data presented thus far suggest that phosphate neutralization on the concave face of EcoRV-induced bent DNA contributes favorable free energy to EcoRV-DNA complex formation. These favorable effects are position dependent and are synergistic for certain combinations of phosphates that are consecutive and close to the apex of the DNA bend. From these observations, it is proposed that phosphate neutralization on the concave face of the bent substrate DNA reduces unfavorable repulsive interactions between DNA phosphates, resulting in a reduced energetic cost of achieving the full extent of DNA bending. This hypothesis is distinct from the idea that the electrostatic modifications used to probe the effects of phosphate neutralization on DNA bending result in a reduced extent of bending, and thus bypass the energetic requirement for this process.

2.2.7 Förster resonance energy transfer as a probe for DNA bending.

To investigate whether the electrostatic modifications discussed above modulate cost of DNA bending or the degree of DNA bending, Förster Resonance Energy Transfer (FRET) was used to measure the extent of DNA bending in modified and unmodified EcoRV-DNA complexes. FRET is the transfer of excited state energy from a donor fluorophore to an acceptor fluorophore as a function of distance. The efficiency of FRET (E) can be described by the relationship

$$E = \frac{R_0^6}{(R_0^6 + R^6)}, \quad (2.10)$$

where R_0 is the Förster distance for a specific donor-acceptor pair (Figure 2.11), and R is the distance that separates the two fluorophores. It is expected that DNA bending by EcoRV will result in a decrease in the distance between the ends of the DNA substrate, such that when donor and acceptor fluorophores are covalently attached to the ends of this substrate, EcoRV-induced DNA bending will result in a decrease in the distance between these two fluorophores. Thus, steady state fluorescence spectra would show a decrease in donor emission and an increase in acceptor emission relative to that for the free DNA (Figure 2.10A). These fluorescence spectra can then be used to calculate E and R , from which a DNA bend angle can then be estimated. Thus, FRET can be used to measure the relative extent of DNA bending in modified and unmodified EcoRV-DNA complexes.

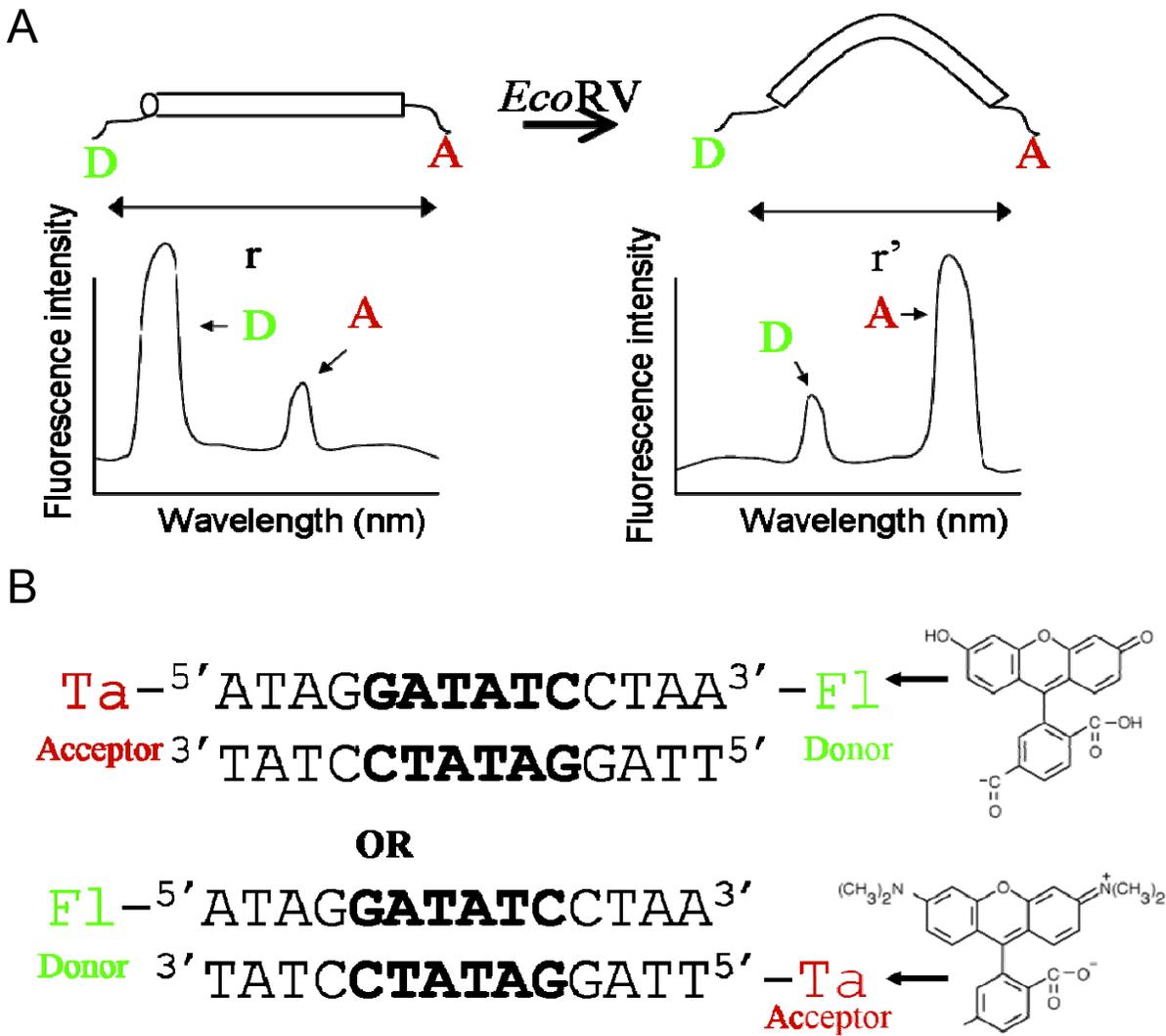


Figure 2.10 FRET as a measure of DNA bending.

(A) Schematic representation of the effects of DNA (rod) bending on interfluorophore distance and donor (D) and acceptor (A) fluorescence intensity. Upon addition of EcoRV, DNA is bent, bringing the two covalently linked fluorophores closer in proximity to one another. This results in a transfer of excited state energy from the donor fluorophore to the acceptor, which causes a decrease in fluorescence intensity of the donor and a concomitant increase in fluorescence intensity of the acceptor. (B) Two fluorophore labeling methods are shown, one in which both donor (Fl) and acceptor (Ta) fluorophores are appended to the “top” strand of the duplex, and one in which the donor is attached to the 5' end of the “top” strand, and the acceptor is attached to the 5' end of the “bottom” strand.

2.2.8 Substrate design and fluorophore labeling strategy

In order to achieve the highest sensitivity to the degree of DNA bending in modified and unmodified EcoRV-DNA complexes, it was desirable to use a DNA substrate that gave the greatest possible difference in FRET efficiency upon EcoRV-induced DNA bending. FRET efficiency is most sensitive to changes in interfluorophore distance when the fluorophores are separated by the Förster distance (R_0) of the specific donor-acceptor pair used, which is the distance at which the efficiency of FRET is 50% (Figure 2.11). Fluorescein (Fl) and carboxytetramethylrhodamine or TAMRA (Ta) were chosen as the donor-acceptor pair (Figure 2.10B) for which, R_0 is approximately 52.5 Å (Parkhurst et al., 1996; Hiller et al., 2003). To identify the optimum parent substrate, steady state fluorescence spectra were taken for a series of fluorophore labeled unmodified DNA duplexes containing the EcoRV recognition sequence GATATC. These substrates varied in their length, flanking sequence context (i.e. the three base pairs immediately 5' GATATC), and method of fluorophore labeling as follows. Both fluorophores were covalently attached to a particular DNA strand by a six carbon linker. Either both strands of the substrate were labeled at the 5' end with Fl linked to the “top” strand and Ta to the “bottom” (Fl-DNA-Ta), or only the “top” strand of the duplex was labeled with Ta linked to the 5' end and Fl to the 3' end of the single strand (Fl-DNA-Ta-top) (Figure 2.10B). By using these two methods, the helical phase, and thus, the distance between the two fluorophores, was subtly modulated in an attempt to maximize the difference in FRET efficiency between the free and bound/bent states.

DNA substrates contained one of three flanking contexts including, TAG, AAA, and GAA/CAA. These contexts were chosen for various reasons; TAG flanks the GATATC

recognition site in all the DNA constructs used in equilibrium binding and cleavage assays, AAA is the flanking triplet present in almost all crystal structures of EcoRV-cognate DNA complexes (Thomas et al., 1999) and is also the flanking context of the DNA in molecular dynamics (MD) simulations of unmodified and modified EcoRV-DNA complexes (see Section 3.2.3). Finally, the GAA/CAA flanking sequence corresponds to that used for a previously published FRET study of EcoRV binding and bending (Hiller et al., 2003). Notably, both AAA and GAA are the flanking triplets most preferred (amongst 64 triplets) by EcoRV in binding to its specific GATATC site, whereas a DNA sequence containing the TAG flanking triplet shows only 1.6-fold worse binding relative to the AAA triplet (Jen-Jacobson, unpublished). R_0 of the F1-Ta fluorophore pair is 52.5 Å, thus a 14 bp substrate represents the most appropriate length substrate for investigating FRET by EcoRV-induced DNA bending, as the F1-Ta distance in a 14 bp substrate is estimated to be close to R_0 . Nevertheless, FRET was measured for substrates of varying length to ensure that this would be the case. Substrates containing the TAG flanking sequence were 14, 15, 16, 20 and 21 base bp in length, while those containing the AAA flanking sequence were 14 and 16 bp in length. Only a 14 bp substrate was used for the GAA/CAA flanking sequence (Table 2.6).

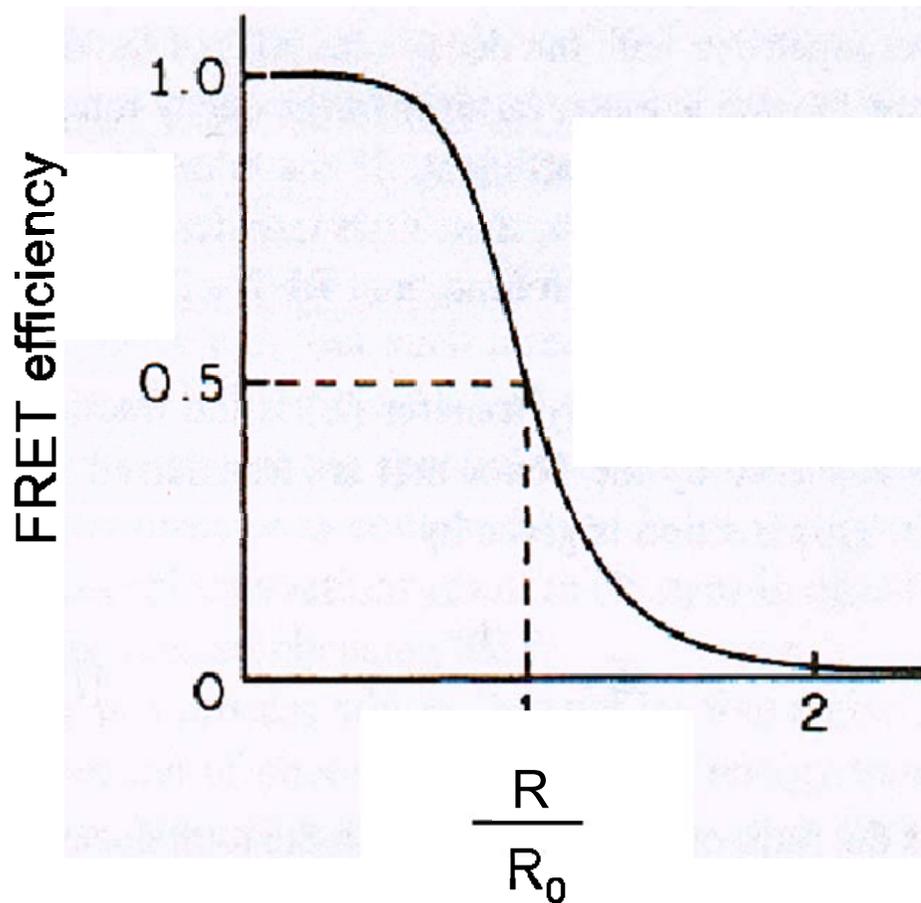


Figure 2.11 FRET efficiency as a function of donor-acceptor distance relative to R_0 . FRET efficiency is shown as a function of the ratio of interfluorophore distance (R) divided by the Förster distance (R_0). FRET efficiency is most sensitive to changes in R when R is equal to R_0 (Figure taken from Lakowicz, 1999).

FRET efficiency was measured as the increase in acceptor fluorescence intensity due to the presence of the donor fluorophore for both free and EcoRV-bound DNA substrates. To make these measurements, normalized spectra of the substrates labeled with Fl alone (Fl-DNA, black spectra; Figure 2.12) and excited at 485 nm (donor excited) were subtracted from the spectra of the doubly labeled substrate (Fl-DNA-Ta), also excited at 485 nm (pink spectra; Figure 2.12), in order to isolate the acceptor emission due to FRET (blue spectra; Figure 2.12). The area under this corrected spectra from 575 to 585 nm (shaded region) divided by the area under the acceptor excited spectra (555 nm) for the Fl-DNA-Ta (Red spectra; Figure 2.12) from 575 to 585 nm gave the value $ratio_A$. $ratio_A$ was then used to calculate FRET efficiency (E) using the expression

$$E = \left[\frac{\varepsilon^A(555)}{\varepsilon^D(485)} \right] \left[ratio_A - \frac{\varepsilon^A(485)}{\varepsilon^A(555)} \right], \quad (2.11)$$

where $\varepsilon^D(\lambda)$ and $\varepsilon^A(\lambda)$ are the extinction coefficients at wavelength λ for Fl and Ta, respectively. This method for determining FRET efficiency best compensates for uncertainties in the degree of labeling of the acceptor, quantum yield of the acceptor, and the concentration of DNA used (Clegg et al., 1992).

Steady state fluorescence spectra to measure ratio_A

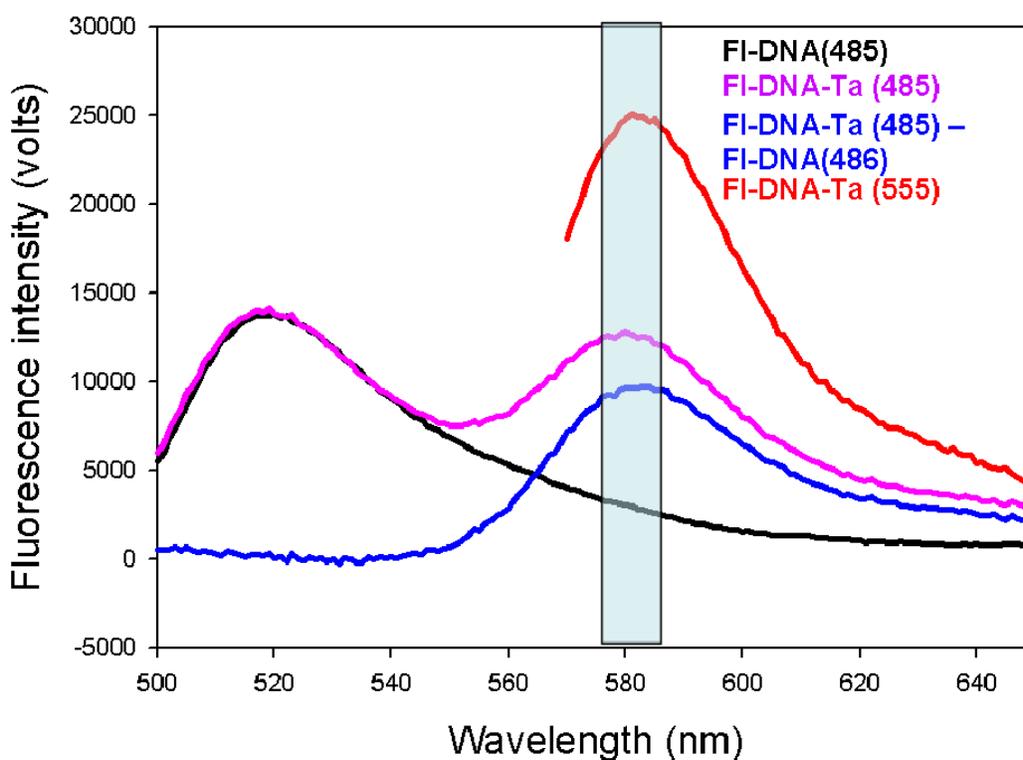


Figure 2.12 Fluorescence spectra used to calculate ratio_A .

Fluorescence spectra used to calculate ratio_A are shown. A normalized, donor excited ($\lambda^{\text{ex}} = 485$), spectra of the FI-DNA substrate (black) is subtracted from the spectra of the FI-DNA-Ta (pink) to isolate the acceptor emission (blue). The area under these spectra from 575 to 585 nm (shaded region) is divided by the area under the directly excited ($\lambda^{\text{ex}} = 555$) spectra (red) to yield ratio_A .

Table 2.6 shows FRET efficiencies that were measured for substrates labeled at the 5' end of both “top” and “bottom” strands. FRET efficiencies were not calculated for FI-DNA-Ta-top substrates due to the lack of a substrate labeled with Ta alone, as FI-DNA-Ta-top substrates were synthesized on FI-phosphoramidite columns, such that FI was incorporated into all “top” strands during synthesis. A Ta only labeled sample is required for measuring $\epsilon^A(485)$ and $\epsilon^A(555)$ and thus, is required for calculating FRET efficiency from ratio_A (Eqn. 2.7). The utility of these substrates, in terms of the degree of difference in FRET between the free and bound DNA substrate, was instead determined from differences in ratio_A (Table 2.6).

FRET experiments for FI-DNA-Ta substrates show that, for free substrates containing the TAG flanking sequence, FRET efficiency decreases from 0.52 to 0.14 as substrate length increases from 14 to 21 bp (Table 2.6). This was expected, as FRET efficiency is dependent on distance (Eqn. 2.6); a longer substrate would result in a greater interfluorophore distance. Further, longer substrates were much less sensitive to DNA bending by EcoRV due to the large interfluorophore distances and the dependence of E on R/R_o (Figure 2.11). It was also observed that substrates with tracts of four or five adenines separated by one helical turn (~10 bp; AAAs14 and 5As16, respectively) resulted in much greater FRET efficiency for both free and EcoRV-bound substrates ranging from 0.66 to 0.86 and from 0.94 to ~1.0, respectively (Table 2.6). This observation likely reflects the intrinsic curvature inherent in these sequences, which bend ~20° toward the minor groove (Koo et al., 1986). The intrinsic curvature is, thus, predicted to result in a reduced interfluorophore distance, both free in solution and when bound/bent by EcoRV, resulting in larger FRET efficiencies for these substrates. EcoRV binding produced a dramatic increase in FRET efficiency for the AAAs14 and 5As16 substrates; however, they were not chosen as parent substrates for P_{Me} modifications. This is because FRET efficiencies of the

magnitude observed in the presence of EcoRV (0.94 and 1.0) are relatively insensitive to differences in R (Figure 2.11) and, therefore, are insensitive to differences in the degree of EcoRV induced DNA bending.

Finally, these experiments have shown that the 14 bp substrates with AAA14 and GAA/CAA flanking sequences, labeled on both strands (Figure 2.10B; referred to as FI-AAA14-Ta and FI-GAA/CAA-Ta, respectively), result in the largest difference in FRET efficiency due to EcoRV-binding (difference in $E = 0.28$ and 0.13 , respectively; Table 2.6). Steady state fluorescence spectra show that FI intensity of the free FI-AAA14-Ta substrate is reduced only 7% relative to the substrate labeled with FI alone (FI-AAA), suggesting that, in the absence of EcoRV, FRET is not as efficient for this substrate as it is for the free GAA/CAA substrate, which shows a 38% decrease in FI intensity in the presence of Ta. This observation agrees qualitatively with the FRET efficiencies calculated for these substrates, as free FI-AAA14-Ta shows a FRET efficiency of 0.22, while that observed for the FI-GAA/CAA-Ta substrate is 0.62. This observation may be due to a difference in the intrinsic shape of these two substrates when they are free in solution. The AAA14 substrate includes an A_4 tract associated with the top strand and an A_3 tract on the bottom strand. While the intrinsic curvature of an A_3 tract is much milder (or non-existent) than that for an A_4 tract, the orientation of these sequence motifs may result in DNA curvature in opposite directions leading to an increased the interfluorophore distance. This is distinct from the 5AS16 and AAAs14 substrates for which the A tracts are in the same helical phase and are expected to exhibit DNA curvature in the same direction (Table 2.6). Regardless, upon EcoRV induced bending there is a significant increase in FRET for FI-AAA14-Ta (Table 2.6) characterized by a reduction in FI emission and increase in Ta emission, as is the case for the GAA/CAA substrate (compare red spectra to blue spectra; Figure 2.13). For all substrates the

addition of EcoRV resulted in an increase in FRET efficiency suggesting that EcoRV *induces* DNA bending and does not rely on the capture of a pre-bent substrate for binding. This is in contrast to a model for DNA binding in which the free bent substrate is the entity that is recognized.

Table 2.6 FRET efficiency calculations for various substrates.

Substrate name	Substrate sequence	Labeling method ^a	[EcoRV] (μM)	Ratio A ^b	FRET efficiency (E) ^c
TAG14	5'-ATAGGATATCCTAA TATCCTATAGGATT-5'	Fl-DNA-Ta-top	0	0.34	-
			0.5	0.37	-
		Fl-DNA-Ta	0	0.30	0.52
			0.5	0.37	0.61
TAG15	5'-AGTAGGATATCCTA TCATCCTATAGGAT-3'	Fl-DNA-Ta-top	0	0.29	-
			0.5	0.31	-
		Fl-DNA-Ta	0	0.33	0.44
			0.5	0.38	0.51
TAG16	5'-AGTAGGATATCCTAGA TCATCCTATAGGATCT-3'	Fl-DNA-Ta-top	0	0.25	-
			0.5	0.25	-
		Fl-DNA-Ta	0	0.26	-
			0.5	0.27	-
TAG20	5'-AAGTAGGATATCCTACAGAA TTCATCCTATAGGATGTGTT	Fl-DNA-Ta-top	0	0.18	-
			0.5	0.19	-
		Fl-DNA-Ta	0	0.21	0.24
			0.5	0.23	0.26
TAG21	5'-AATGTAGGATATCCTACAGA TTACATCCTATAGGATGTCT-5'	Fl-DNA-Ta-top	0	0.16	-
			0.5	0.17	-
		Fl-DNA-Ta	0	0.20	0.14
			0.5	0.22	0.15
AAA14	5'-AAAAGATATCTTTA TTTTCTATAGAAAT-5'	Fl-DNA-Ta-top	0	0.38	-
			0.5	0.40	-
		Fl-DNA-Ta	0	0.22	0.28
			0.5	0.33	0.60
AAAs14	5'-AAAAGATATCAAAA TTTTCTATAGTTTT-5'	Fl-DNA-Ta-top	0	0.34	-
			0.5	0.39	-
		Fl-DNA-Ta	0	0.33	0.86
			0.5	0.38	~1.0
5As16	5'-AAAAAGATATCAAAAA TTTTTCTATAGTTTTT-5'	Fl-DNA-Ta-top	0	0.21	-
			0.5	0.25	-
		Fl-DNA-Ta	0	0.29	0.66
			0.5	0.32	0.94

Substrate name	Substrate sequence	Labeling method ^a	Enzyme (μ M)	Ratio A ^b	FRET efficiency (E) ^c
GAA14	5'-TGAAGATATCTTCT ACTTCTATAGAAGA-5'	FI-DNA-Ta-top	0	-	-
			0.5	-	-
		FI-DNA-Ta	0	0.28	-
			0.5	0.33	-
GAA/CAA	5'-AGAAGATATCTTGA TCTTCTATAGAACT-5'	FI-DNA-Ta-top	0	-	-
			0.1	-	-
		FI-DNA-Ta	0	0.39 (\pm 0.01)	0.62 (\pm 0.01)
			0.1	0.47 (\pm 0.01)	0.75 (\pm 0.01)

^aLabeling method corresponds to those depicted in Figure 3.1B. FI-DNA-TA-top refers to the method in the top panel in which both fluorophores are appended to the top strand of the DNA duplex and FI-DNA-Ta refers to that in which the FI is linked to the top strand and the Ta is linked to the bottom strand.

^bMeasured as described in the text for free and bound DNA substrates in 0.1 M KCl, 10 mM CaCl₂, at pH 7.25.

^cCalculated as described in 3.2.

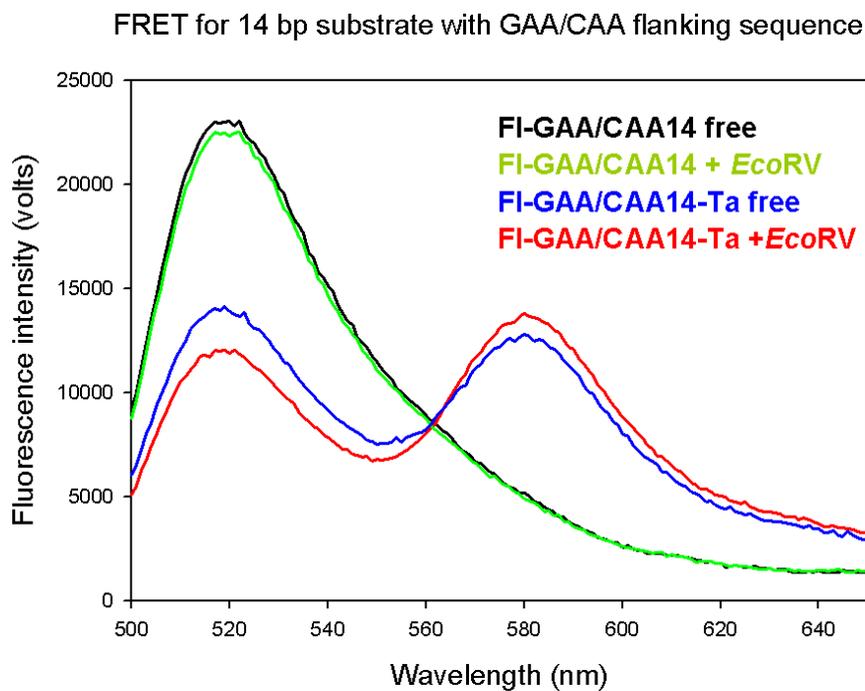
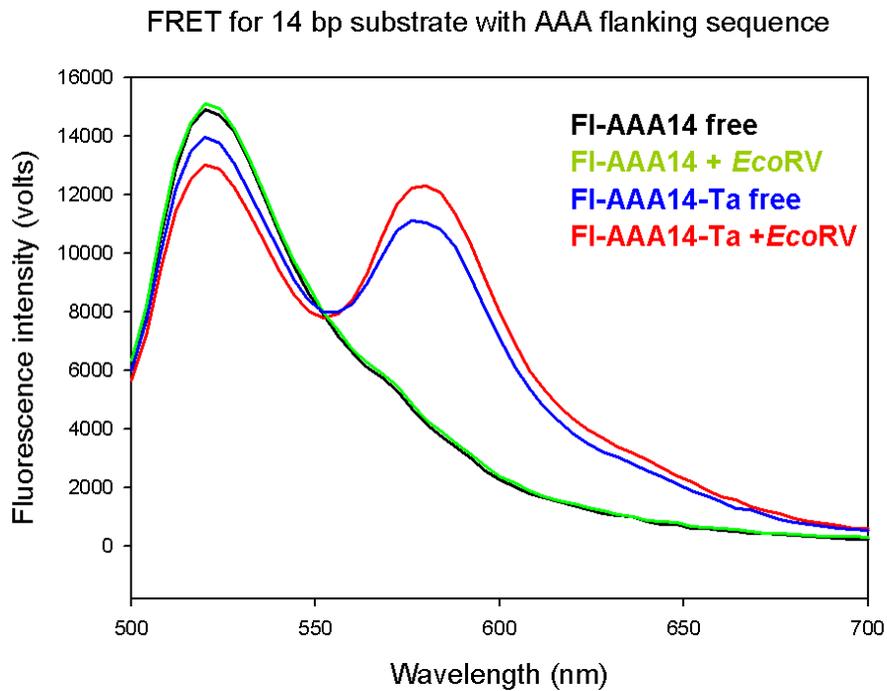


Figure 2.13 Fluorescence spectra for unmodified FRET substrates.

Steady state fluorescence emission spectra for AAA14 (top) and GAA/CAA (bottom) are shown. Included are spectra for FI-DNA substrates in the presence (green) and absence (black) of EcoRV as well as spectra for FI-DNA-Ta in the presence (red) and absence (blue) of EcoRV.

Importantly, the addition of EcoRV does not affect the fluorescence intensity of either the Fl-AAA14 or Fl-GAA/CAA-Ta substrate labeled with Fl alone (Figure 2.13; compare black and green spectra). This suggests that for the doubly labeled substrates, the reduced Fl intensity observed in the presence of EcoRV is due to FRET, and not quenching by EcoRV (Figure 2.13; compare blue and red spectra). In addition, EcoRV binding does not affect the fluorescence intensity of Ta (not shown) suggesting that the EcoRV-induced increase in Ta intensity observed for these substrates is due to FRET.

Based on these preliminary results, and in accordance with collaborative studies carried out by John Perona's lab group at the University of California at Santa Barbara, the GAA/CAA substrate was chosen as the parent substrate in which all P_{Me} substitutions were made. This sequence seems to be very suitable as a parent substrate in which all of the P_{Me} substitutions were made. GAA/CAA shows very little difference in equilibrium binding affinity (K_A) relative to the AAA14 and TAG14 substrates ($\Delta\Delta G^\circ \approx 0$; not shown), and, although it was not probed for the GAA/CAA sequence, K_A for wild-type EcoRV binding to the TAG14 substrate was not affected by the addition of the two hexamethyl tethered fluorophores. hexamethyl tethered fluorophores. Presumably this is also the case for the GAA/CAA substrate. Further, FRET efficiency for the free and EcoRV bound GAA/CAA substrate correlates to distances of 48.3(± 0.2) Å and 43.6(± 0.3) Å, respectively. When a simple model for DNA bending is applied to these distances such that the DNA is completely straight in the free state and is bent at the center of the molecule in the EcoRV-bound state, these distances correspond to a bend angle of approximately 50.8(± 2.0)°. This angle agrees well with those observed for EcoRV-induced bending in X-ray crystal models (Winkler, 1993; Thomas et al., 1999; Horton and Perona, 2000),

suggesting that bending of this substrate in solution resembles that observed in the crystalline state.

2.2.9 Modified and unmodified EcoRV-DNA complexes show similar extents of DNA bending.

In order to probe the extent of DNA bending for EcoRV-DNA complexes with electrostatic modifications, FRET was measured for various modified and unmodified DNA substrates in the presence and absence of wild-type or mutant EcoRV. For these experiments, 9nM of the DNA substrate was used to attain an appropriate fluorescence signal. EcoRV was added to a final concentration that was 10-fold greater than the DNA concentration (90 nM). At these concentrations, the EcoRV in each sample was at least 300-fold above the estimated equilibrium dissociation constant (K_D ; estimates based on the K_D values at 0.24M KCl and the dependence of a particular interaction on [KCl]) suggesting that the DNA was completely bound by EcoRV. Table 2.7 shows FRET efficiencies as well as interfluorophore distances, and bend angles calculated for the DNA substrates in these complexes. As mentioned above, the unmodified complex shows a bend angle of 50.8°. For the most part, modified complexes show bend angles that are similar to that observed for the unmodified complex and range from 46 to 56° (except for the R221A-P_{Me-2} complex; Table 2.7), in support of the model that the energetic effects of these modifications are due to differences in the cost of achieving the full DNA bend and not to a difference in extent of bending.

The lone exception is observed for the R221A-P_{Me-2} complex which shows a bend angle of 39° (Table 2.7). This is consistent with the observation that the first-order cleavage rate constant for this complex is dramatically reduced relative to that observed for the unmodified

complex (Section Table 4.2.3) suggesting that the protein-DNA interface, and thus the active site, is not properly assembled for in this complex. Proper formation of the EcoRV active site requires DNA bending (Winkler, 1993) thus a reduced extent of DNA bending in the R221A-P_{Me-2} complex would lead to a reduction in cleavage rate.

Table 2.7 Relative extent of bending for modified and unmodified EcoRV-DNA complexes

Complex^a	E (Free substrate)^{b,c}	E (+EcoRV)	Distance (Å) (free substrate)^d	Distance (Å) (+ EcoRV)	Bend Angle (o)
Wt-GATATC	0.62(±0.01)	0.75(±0.01)	48.3(±0.1)	43.6(±0.3)	50.8(±2.0)
Wt-P _{Me-5}	0.63(±0.02)	0.76(±0.02)	48.1(±0.6)	43.4(±0.6)	51.0(±3.6)
Wt-P _{Me-2}	0.55(0.01)	0.68(±0.03)	50.8(±0.5)	46.5(±1.0)	47.8(±3.3)
Wt-P _{Me+3}	0.51(±0.03)	0.64(±0.04)	52.0(±1.0)	47.8(±1.3)	46.6(±1.8)
K119A-GATATC	0.61(±0.01)	0.77(±0.01)	48.6(±0.1)	42.9(±0.4)	56.0(±2.0)
K119A-P _{Me-2}	0.56(±0.01)	0.76(±0.05)	50.6(±0.6)	43.3(±1.9)	55.8(±8.0)
R221A-GATATC	0.61(±0.01)	0.74(±0.02)	48.9(±0.4)	44.1(±0.9)	50.9(2.8)
R221A-P _{Me-2}	0.54(±0.01)	0.62(±0.03)	51.5(±0.5)	48.4(±0.5)	39.0(±3.2)
R226A-GATATC	0.61(±0.01)	0.73(±0.03)	48.8(±0.4)	44.4(±1.0)	48.7(±3.6)
R226A-P _{Me-5}	0.63(±0.02)	0.74(±0.04)	48.2(±0.6)	44.0(±1.3)	48.0(±5.0)
R140A-GATATC	0.61(±0.01)	0.73(±0.03)	48.6(±0.2)	44.3(±1.0)	48.1(±6.0)
R140A-P _{Me+3}	0.51(±0.02)	0.63(±0.03)	52.2(±0.6)	48.0(±1.0)	46.2(±2.7)

^aThe double stranded 14 bp oligonucleotide d(AGAAGATATCTTGA) was used as the parent sequence context for all FRET experiments.

^bFRET was probed in BB + 0.1 M KCl, 20 mM cacodylate, 10 mM CaCl₂, at pH 7.25.

^cMeans and standard deviations are for at least 2 determinations of FRET efficiency.

^dFRET efficiency was calculated from the acceptor emission as described in the text.

^eCalculated from Eqn. 3.1.

Overall, it appears that the electrostatic modifications used to probe the role of phosphate neutralization in EcoRV-DNA complex formation, do not result in reduced bending in the complex, and instead are associated with the formation of a specific complex in which the DNA

is fully bent. This suggests that the favorable effects of phosphate neutralization on complex formation are due to a reduced cost of DNA bending and not a reduced extend of bending.

3.0 CHAPTER 3: INSIGHT INTO THE ENTHALPIC AND ENTROPIC CONSEQUENCES OF ASYMMETRIC PHOSPHATE NEUTRALIZATION

3.1 INTRODUCTION

3.1.1 Thermodynamic dissection of ΔG for the EcoRV-DNA interaction

To further characterize how asymmetric phosphate neutralization reduces the energetic cost of DNA bending, I undertook a series of experiments to measure the enthalpic and entropic contributions to $\Delta G^{\circ}_{\text{bind}}$ and the associated heat capacity change $\Delta C^{\circ}_{\text{p}}$. The standard state Gibbs free energy can be expressed as the sum of its enthalpic and entropic components by the relationship:

$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ}, \quad (3.1)$$

where T is temperature (K), and ΔH° and $T\Delta S^{\circ}$ are the standard enthalpy and entropy changes that occur upon complex formation, respectively. As noted in Section 1.3.4, $\Delta G^{\circ}_{\text{bind}}$ is the net of opposing factors: favorable contributions from protein-base and protein-phosphate interactions and the release of water from nonpolar surfaces (hydrophobic effect) (Ha et al., 1989a; Spolar and Record Jr, 1994) are opposed by unfavorable contributions from restriction of rotational, translational, and configurational degrees of freedom upon complex formation, desolvation of

polar surfaces, and DNA distortion (Jen-Jacobson, 1997). These opposing factors are characteristic of any protein-DNA interaction, but for restriction endonucleases there is an additional component from electrostatic repulsion where active site side chains are closely opposed to DNA phosphates. This repulsion, taken alone, contributes a positive increment to both ΔG° and ΔC_p° (see below).

While measuring and comparing $\Delta G_{\text{bind}}^\circ$ can give us information concerning the stability of different protein-DNA complexes, dissecting ΔG° into ΔH° and $T\Delta S^\circ$ can provide insight into the molecular origins of the energetic differences between modified and unmodified EcoRV-DNA complex formation. The molecular factors that contribute to ΔH° and ΔS° are interrelated and can oppose one another. For example, formation of protein-base and protein-phosphate contacts contributes favorable ΔH° to complex formation; however, these interactions restrict the conformational and vibrational flexibility of the interacting partners and thus, contribute unfavorable ΔS° (Dunitz, 1995; Jen-Jacobson, 2000b). Unfavorable enthalpic contributions derive from desolvation of polar functional groups (Makhatadze and Privalov, 1993; Avbelj and Baldwin, 2004) (desolvation of nonpolar groups also contributes to ΔH° , but to a lesser extent (Makhatadze and Privalov, 1995) and molecular strain (Jen-Jacobson, 2000b) (Table 3.1). The term molecular strain refers to the fact that in order to create a properly formed protein-DNA interface, atoms, functional groups, entire protein side chain residues and DNA bases adopt positions in the complex that do not correspond to their positions of lowest potential energy. Strain may arise from bond bending, stretching, or rotation and/or from unfavorable non-bonded interactions, such as steric clashes and electrostatic repulsion (Jen-Jacobson, 2000b).

A major source of molecular strain derives from DNA distortion (Jen-Jacobson, 2000b). Investigation of the values of ΔH° and ΔS° for the formation of ten diverse protein-DNA

complexes reveals that the amount of DNA distortion observed in the X-ray crystal model is correlated to ΔH° for complex formation (Jen-Jacobson, 2000b). For example, GCN4 and λ cI complexes display little DNA bending (Jones et al., 1999), and ΔH° is most favorable for the formation of these complexes (Merabet and Ackers, 1995; Berger et al., 1996). For TBP and CAP, however, DNA distortion is dramatic (Schultz et al., 1991; Kim et al., 1993a; Kim et al., 1993b), and ΔH° is the most unfavorable (Ebright et al., 1989; Petri et al., 1995). It is proposed here that increased interphosphate repulsion due to DNA bending is a source of electrostatic strain and that phosphate neutralization may reduce strain, thus contributing favorable ΔH° to specific EcoRV-DNA complex formation.

Favorable ΔS° derives from the release of water (Spolar and Record Jr, 1994) and counterions (Sharp, 1995; Record et al., 1998) from protein and DNA surfaces, while restriction of rotation-translation of the macromolecules, loss of conformational and vibrational freedoms, and restriction of interfacial waters all contribute unfavorable ΔS° to specific protein-DNA interactions (Jen-Jacobson 2000b) (Table 3.1). Thus, dissecting $\Delta G^\circ_{\text{bind}}$ into ΔH° and ΔS° can help to elucidate the molecular origins of the energetic effects associated with a particular electrostatic perturbation.

Table 3.1 Factors affecting thermodynamic contributions to specific protein-DNA complex formation.

	ΔH°	ΔS°
Favorable	<ul style="list-style-type: none"> • Attractive interactions (H-bonds, charge-charge, nonpolar) 	<ul style="list-style-type: none"> • Water release • Counterion release
Unfavorable	<ul style="list-style-type: none"> • Molecular Strain • Repulsive interactions • Polar desolvation 	<ul style="list-style-type: none"> • Restricted rotation/translation • Reduced configurational freedom • Loss of vibrational freedom of trapped water
	ΔC_p°	
Negative	<ul style="list-style-type: none"> • Burial and desolvation of nonpolar surfaces • Loss of configurational/vibrational freedom (Interface restrains side chains, bases, backbone) • Restricted freedom of interfacial H₂O 	
Positive	<ul style="list-style-type: none"> • Burial and desolvation of polar surfaces • Molecular Strain (steric, electrostatic, etc.) 	

A third, informative thermodynamic parameter is the heat capacity change (ΔC_p^o) for the formation of protein-DNA complexes; ΔC_p^o is the thermodynamic parameter that describes the temperature dependence of the change in enthalpy and entropy at constant pressure given by:

$$\Delta C_p^o = \left(\frac{\partial \Delta H^o}{\partial T} \right)_P \quad (3.2a)$$

$$\Delta C_p^o = T \left(\frac{\partial \Delta S^o}{\partial T} \right)_P \quad (3.2b)$$

It has been shown for several systems including, EcoRI (Ha et al., 1989b; Jen-Jacobson, 2000a), BamHI (Engler, 1998; Jen-Jacobson, 2000a), EcoRV (this work; Engler, L.E. unpublished), cro repressor (Takeda et al., 1992), λ cI repressor (Merabet and Ackers, 1995), and *trp* repressor (Ladbury et al., 1994) that a large negative ΔC_p^o is a thermodynamic signature of specific protein-DNA complex formation, reflecting the assembly of a highly complementary intimate interface.

For protein folding, it has been shown that $\Delta C_{p,\text{fold}}^o$, determined in calorimetric studies, can be quantitatively predicted from thermodynamic data for the transfer of hydrocarbons and amides to the pure liquid phase (Murphy and Freire, 1992; Spolar et al., 1992; Makhatadze and Privalov, 1995). The successful parameterization of ΔC_p^o from computational analysis of the amount of buried polar and nonpolar surfaces during protein folding led Spolar & Record (1994) to propose that surface burial (with concomitant water release) was also the major contributor to ΔC_p^o for formation of specific protein-DNA complexes. This parameterization suggests that burial of nonpolar surface makes a strong negative contribution to ΔC_p^o , whereas burial of polar surface area makes a smaller but significant positive contribution – see Eq. 3.3 (Murphy and Freire, 1992; Spolar et al., 1992; Makhatadze and Privalov, 1995).

However, several groups noted that use of the empirical relationship between ΔC_p^o and changes in solvent accessible surface area are insufficient to account for the magnitude of the ΔC_p^o for these interactions (Ladbury et al., 1994; Petri et al., 1995; Berger et al., 1996; Engler, 1998). For example, using the X-ray co-crystal structure of the EcoRI-DNA complex to calculate the amount of polar and nonpolar surface area that is buried upon complex formation, it was estimated that the corresponding contribution to ΔC_p^o would be ≈ -0.31 kcal/molK. ΔC_p^o for specific EcoRI-DNA complex formation has been measured to be $-2.5(\pm 0.1)$ kcal/molK. Thus, burial molecular surface area can not be the dominant contributor to the large negative ΔC_p^o that is observed for the formation of the specific EcoRV-DNA complex, but additional factors must make significant contributions to this large negative ΔC_p^o . These additional factors include immobilization of interfacial waters and stiffening of soft vibrational frequencies (Table 3.1)(Sturtevant, 1977a; Ladbury et al., 1994; Jen-Jacobson, 2000a). The factors that contribute to a more negative ΔC_p^o for protein-DNA interactions are opposed by factors that make positive contributions to ΔC_p^o , including desolvation of polar surfaces (see above) and molecular strain (Jen-Jacobson, 2000a) (Table 3.1). Thus, the ΔC_p^o that is measured for protein-DNA interactions results from the aggregate of factors that make both negative and positive contributions to ΔC_p^o for complex formation (Table 3.1).

ΔC_p^o for the formation of the specific EcoRV-DNA complex is equal to -1.2 kcal/molK (Tables 3.3 and 3.4). What are the factors that contribute to this ΔC_p^o , and what are the magnitudes of these contributions? Factors which contribute to a more positive ΔC_p^o include the desolvation of polar groups, and molecular strain (Table 3.1). Two possible sources of molecular strain exist for the formation of the specific EcoRV-DNA complex including 1) dramatic DNA bending that is required for the formation of this complex (Figure 1.4) and from 2) the increase

in electrostatic repulsion at the EcoRV-DNA active site in which several negatively charged side chains become directly apposed to the anionic scissile phosphate (See Section 1.3.4). These positive contributions to ΔC_p^0 are offset by several factors that contribute to a more negative ΔC_p^0 for EcoRV-DNA complex formation such that the net ΔC_p^0 is large and negative. These factors include desolvation of nonpolar groups, immobilization of bound waters, and the stiffening of soft vibrational frequencies (Sturtevant, 1977a; Ladbury et al., 1994; Jen-Jacobson, 2000a)(Table 3.1)

The contribution of polar and nonpolar desolvation to ΔC_p^0 for the EcoRV-DNA interaction can be estimated based on the amount of surface area that is buried upon formation of this complex. Briefly, the amount of surface area buried upon complex formation (ΔASA) was calculated by subtracting the solvent accessible surface area of the EcoRV-DNA complex (ASA_{complex}) from the sum of solvent accessible surface areas of the free protein (ASA_{protein}) and the free DNA (ASA_{DNA}), such that

$$\Delta ASA = (ASA_{\text{protein}} + ASA_{\text{DNA}}) - ASA_{\text{complex}}. \quad (3.3)$$

Solvent accessible surface areas were calculated using the Naccess program (Hubbard et al., 1991) and were partitioned into those for polar and nonpolar surfaces (Table 3.2). ASA_{DNA} was calculated for the crystal structure of a 10bp B-form DNA molecule containing the specific GATATC recognition sequence (PDB entry 1ZFC), ASA_{complex} was calculated from the crystal structure of the EcoRV-DNA complex (Thomas et al., 1999; PDB entry 1B94), and various models were used to calculate ASA_{protein} . For Model A in Table 3.2, ASA_{protein} was calculated for the EcoRV assuming a rigid body interaction such that the EcoRV molecule was simply

“undocked” from the DNA in the specific EcoRV-DNA co-crystal model (PDB entry 1B94)(Thomas, 1999).

Crystallographic analysis shows that the free EcoRV molecule is structurally distinct from that in the EcoRV-DNA complex and includes several regions where there is a lack of appreciable electron density (Perona and Martin, 1997). The presence of these “disordered” regions suggests that a rigid body model may be inappropriate for estimating the amount of molecular surface area buried upon formation of the EcoRV-DNA complex. Thus, three additional models, which take into account the disorder-order transition, were used to estimate ASA_{protein} . These were: 1) The model of Lee and Richards (1971) (Model B, Table 3.2) which assumes that the disordered regions of the free protein are in a fully extended conformation. In this model, the ASA for disordered regions is approximated by estimating the ASA of each residue, “X” in the extended tripeptide Ala-X-Ala, and then summing the contributions of each residue over the composition of the disordered region of the protein. This represents the *maximum* ASA estimate for the disordered EcoRV regions. 2) The model of Livingstone et al. (1991)(Model C, Table 3.2), which treats the disordered region as an intact extended β -form chain and takes into account the primary sequence of the segment. 3) The model of Creamer and Rose (Creamer et al., 1995; Creamer et al., 1997)(Model D; Table 3.2) which uses peptides of varying length (3 to 45 residues) that have been excised from 43 different protein molecules in their ordered conformation. The ASA of the central residue in these peptides was calculated and averaged by residue type over all of the peptides tested. This model assumes that there is a considerable amount of residual structure in disordered protein regions and represents a *minimum* ASA estimate for these regions. Using these three models, I have calculated the upper (Model B) and lower (Model D) limits for the total amount of ASA that is buried upon formation of the

specific EcoRV-DNA complex, assuming a contribution from the ordering of the disordered regions in the free EcoRV protein (Table 3.2).

Table 3.2 Estimated ΔC_p^o calculated from accessible surface area values

Protein-DNA Interaction ^a	Total accessible surface area (Å ²) ^b	Polar accessible Surface area (Å ²)	Nonpolar accessible surface area (Å ²)	ΔC_p^o (CalcI) ^c	ΔC_p^o (CalcII) ^d
Model A: free protein	22359	12151	10208		
+free DNA	3957	2572	1385		
–complex	<u>21816</u>	<u>11702</u>	<u>10114</u>		
Net change (ΔASA)	4500	3021	1479	-0.05	-0.5
Model B: free protein	30490	17939	12551		
+free DNA	3957	2572	1385		
–complex	<u>21816</u>	<u>11702</u>	<u>10114</u>		
Net change (ΔASA)	12631	8809	3822	+0.01	-0.4
Model C: free protein	27553	16544	11009		
+free DNA	3957	2572	1385		
–complex	<u>21816</u>	<u>11702</u>	<u>10114</u>		
Net change (ΔASA)	9694	7414	2280	+0.3	-0.1
Model D: free protein	26193	15653	10540		
+free DNA	3957	2572	1385		
–complex	<u>21816</u>	<u>11702</u>	<u>10114</u>		
Net change (ΔASA)	8334	6523	1811	+0.3	-0.1
Model E: free protein	23183	12351	10832		
+free DNA	3957	2572	1385		
–complex	<u>21816</u>	<u>11702</u>	<u>10114</u>		
Net change (ΔASA)	5324	3221	2103	-0.02	-0.7

Model F:	free protein	24535	13634	10901		
	+free DNA	3957	2572	1385		
	-complex	<u>21816</u>	<u>11702</u>	<u>10114</u>		
	Net change (Δ ASA)	6676	4504	2172	-0.06	-0.5

^a Δ ASA was calculated using Equation 3.3 in the text. ASA for the free DNA was calculated for the X ray crystal structure of a 10bp B-DNA duplex containing the EcoRV specific recognition sequence GATATC. ASA for the free protein was calculated by various methods (see text). ASA for the complex was calculated for the EcoRV-DNA-Ca²⁺ co-crystal structure (PDB entry 1B94).

^bASA was calculated and partitioned into polar and nonpolar surface using the Naccess program of Hubbard et al. (1991) with an update to handle DNA.

^c ΔC_p^o was calculated based on ASA calculations using the parameterization of Spolar et al. (1992) described in Equation 3.4.

^d ΔC_p^o was calculated for burial of DNA surfaces based on ASA calculations and parameterization for DNA of Madan and Sharp (2001) described in Equation 3.5. ΔC_p^o calculated from ASA associated with protein surfaces was again calculated using the parameterization of Spolar et al. (1992)

I have used the Δ ASA calculations in Table 3.2 to estimate ΔC_p^o based on the burial of polar and nonpolar surface area upon formation of the specific EcoRV-DNA using semi-empirical relationship developed from model compound transfer data (Spolar et al. 1992) such that

$$\Delta C_p^o = -0.32\Delta A_{np} + 0.14\Delta A_p \text{ cal/molK}, \quad (3.4)$$

where ΔA_{np} is the amount of nonpolar surface area buried upon protein-DNA complex formation and ΔA_p is the amount of polar surface area buried upon complex formation. From this relationship, I have estimated that ΔC_p^o would equal to -0.05 kcal/molK using a rigid body assumption to calculate Δ ASA for the formation of the specific EcoRV-DNA complex (Table 1; Model A). Again, I have measured ΔC_p^o for the formation of the specific EcoRV-DNA complex to be $-1.2(\pm 0.1)$ kcal/mol^K (Tables 3.3. and 3.4), suggesting that the burial of molecular surface

area is insufficient to account for the large negative ΔC_p^0 observed for the formation of the specific EcoRV-DNA complex

Spolar and Record (1994) suggested that the deficit in the magnitude of ΔC_p^0 calculated for protein-DNA interactions predicted from burial of solvent accessible surface area results from underestimating the amount of surface area that is buried upon the complex formation. They suggested that taking into account the “coupled folding” (folding of disordered protein regions) that accompanies the formation of a protein-DNA complex would lead to the burial of additional surface area. Indeed, as mentioned above, several disordered regions of the free EcoRV protein become ordered upon formation of the specific EcoRV-DNA complex, including the C-terminal domains (residues 218 to 245), and recognition loops (R loop, residues 183 to 187; see Section 1.3.2.3 and Figure 1.4). These regions are expected to have a greater molecular surface area in the disordered state than in the ordered state such that coupled folding upon formation of the specific EcoRV-DNA complex results in an increase in ΔASA relative to that estimated from the rigid body interaction with a fully ordered protein.

When coupled folding of disordered protein regions is considered, (Models B through D) the total ΔASA (ΔASA_{total}) for specific EcoRV-DNA complex formation increases by 3800 \AA^2 to 8100 \AA^2 relative to that calculated using an $ASA_{protein}$ for the fully ordered protein which was simply “undocked” from the EcoRV-DNA complex (Table 3.2; Model B-C). Further, the changes in both polar accessible surface area (ΔASA_{polar}) and nonpolar surface area (ΔASA_{np}) increase by 3500 to 5800 \AA^2 and from 300 to 2300 \AA^2 respectively. Applying these ΔASA values to Eq. 3.4 yields an estimated ΔC_p^0 of +0.01 to +0.3 kcal/molK, suggesting that burial of molecular surface area upon the coupled folding of disordered regions in the free EcoRV protein results in an *increase* in the estimated ΔC_p^0 (Table 3.2) and can not account for the large negative

ΔC_p^o that is measured for the formation of this complex. Thus the deficit between the calculated ΔC_p^o and that measured for the formation of the specific EcoRV-DNA complex is increased.

One luxury that is provided by the EcoRV system is that a crystal model of the free protein exists for which the disordered C-terminal regions and the R loops have been modeled into the structure (Winkler, 1993). This free EcoRV structure, and an MD simulated structure of this free protein were also used to calculate ASA_{protein} , and ΔASA for the formation of the specific EcoRV-DNA complex. ΔC_p^o calculations based on the burial of molecular surface area for these free protein models show values of -0.02 and -0.06 kcal/molK, respectively. Thus, estimated ΔC_p^o values range from -0.06 to $+0.3$ when Equation 3.4 is applied to the ΔASA that is calculated for the formation of the specific EcoRV-DNA complex using various models to treat ASA_{protein} . These values are also much less negative than that determined experimentally.

An alternative method for estimating the contribution DNA desolvation to ΔC_p^o has been put forth by Madan and Sharp (Madan and Sharp, 2001), which suggests that the heat capacity of hydration ($\Delta C_{p,\text{hydration}}$) for individual DNA bases, sugars, and the phosphate backbone is more positive than previously predicted. Thus, desolvation of both polar and nonpolar components of these groups will contribute to a more negative ΔC_p^o for protein-DNA interactions such that,

$$\Delta C_p^o = -0.173\Delta A_{\text{np}} - 0.171\Delta A_{\text{p}} = -0.17\Delta A_{\text{total}} \text{ cal/molK.} \quad (3.5)$$

If this model is applied to the surface area calculations in Table 3.2, a more negative ΔC_p^o is estimated for the formation of the specific EcoRV-DNA complex ranging from -0.1 to -0.7 kcal/molK depending on the model used for the ΔASA calculation. This observation suggests that burial of DNA accessible surface area may contribute more negatively to ΔC_p^o for specific

EcoRV-DNA binding than originally appreciated. It should be noted that this model for the contribution of ΔASA on ΔC_p^o has been parameterized based on observations with individual components of the DNA molecule, and not a duplex oligonucleotide and, while it is informative, must be used cautiously. Regardless, use of the Madan and Sharp model for burial of molecular surface area suggest that this factor still can not account for the large negative ΔC_p^o (-1.2 kcal/molK) that has been measured for the formation of the specific EcoRV-DNA complex (Table 3.3 and 3.4).

In addition to desolvation of molecular surfaces, the immobilization of water molecules at the protein-DNA interface is thought to contribute negative ΔC_p^o to protein-DNA interactions (Sturtevant, 1977a; Ladbury et al., 1994). Morton and Ladbury (1996) assign an average ΔC_p^o of -6 cal/K·(mol of water), while work by Connelly based on crystalline hydrate data (Connelly et al., 1994; Connelly, 1997) estimates a ΔC_p^o for the incorporation of a water molecule into the interface of a protein-ligand complex to be -8.2 cal/K (mol of water). There are 57 waters that appear to be “trapped” at the specific EcoRV-DNA interface. Using the value of -8.2 cal/K(mol of water), an upper bound of the contribution of interfacial water immobilization can be estimated and is approximately equal to -0.5 kcal/molK. If this value is combined with the range of values estimated for the contribution of polar and nonpolar desolvation, ΔC_p^o is approximately equal to -0.3 to -1.2 kcal/molK. The upper range for ΔC_p^o estimated in this way is similar to that measured by experiment ($\Delta C_p^o = 1.2 \pm 0.1$ kcal/molK).

It should be kept in mind that the estimates for ΔC_p^o due to immobilization of interfacial waters in crystalline hydrate are only applicable to the most tightly bound water molecules (Dunitz, 1994), thus the estimate of ΔC_p^o based on the number of “trapped” waters at the EcoRV-DNA complex represents an absolute maximum value. Further, ΔC_p^o for the formation

of the specific EcoRV-DNA complex is highly dependent on the concentration and type of anion in solution. Specifically ΔC_p^0 for complex formation probed at 0.17M NaCl was shown to equal -1.8 kcal/molK (Engler, L.E., and Jen-Jacobson, L., unpublished) while in the experiments done here, I observe a ΔC_p^0 of -1.2 kcal/molK in the presence of 0.24M KCl (Tables 3.3 and 3.4). From both of these considerations, and from the uncertainty associated with the Madan and Sharp (2001) model for the effects of ΔASA on ΔC_p^0 it can be inferred that burial of molecular surface area, and immobilization of interfacial waters likely do not account for the all of the large negative ΔC_p^0 that is observed experimentally. Thus, the stiffening of vibrational motions of interfacial elements of the protein and DNA, must make a significant, if not a major contribution to ΔC_p^0 for EcoRV-DNA complex formation. (Sturtevant, 1977a; Ladbury et al., 1994; Jen-Jacobson, 2000a).

3.2 RESULTS AND DISCUSSION

In order to measure the thermodynamic parameters ΔH° , ΔS° , and ΔC_p° for EcoRV-DNA complex formation, the temperature dependence of the equilibrium affinity constant (K_A) was measured for the binding of both wt and K119A to unmodified and Rp-P_{Me-2} modified DNA substrates. These data were fit to two models; the first is a multi-parametric model that assumes ΔC_p° is independent of temperature; the second is based on the methods of Clarke and Glew (Clarke, 1966; Dec, 1985) (henceforth referred to as the Clarke and Glew method). For the multi-parametric method, T_H , T_S and ΔC_p° were obtained from a fit of the temperature dependence of the K_A to the equation:

$$\ln K_{obs} = \left(\frac{\Delta C_p^\circ}{R} \right) \left[\left(\frac{T_H}{T} \right) - \ln \left(\frac{T_S}{T} \right) - 1 \right], \quad (3.6)$$

where T_H and T_S are the temperatures at which ΔH° and ΔS° are zero, respectively. The values for ΔH° and ΔS° were then calculated from the parameters T_H , T_S , and ΔC_p° , using the following equations:

$$\Delta H^\circ = \Delta C_p^\circ (T - T_H) \quad (3.7)$$

$$\Delta S^\circ = \Delta C_p^\circ \ln \left(\frac{T}{T_S} \right) \quad (3.8)$$

to yield ΔH° and ΔS° at 25°C.

The Clarke and Glew method was used in order to evaluate the temperature dependence of ΔC_p° , which the multi-parametric fit assumes is zero. Clarke and Glew showed that the parameters ΔH° , ΔS° , ΔG° and ΔC_p° can be expressed at temperature T, as the perturbation of their values at a reference temperature θ using the Taylor series expansion, such that:

$$R \ln K(T) = -\frac{\Delta G^\circ_\theta}{\theta} + \frac{\Delta H^\circ_\theta}{\theta} \left[\frac{x}{1+x} \right] + \Delta C_p^\circ \theta \left[x^2 \sum_{n=1}^{\infty} \frac{n}{n+1} (-x)^{n-1} \right] \\ + \frac{\theta}{2} \left(\frac{d\Delta C_p^\circ}{dT} \right)_\theta \left[x^3 \sum_{n=1}^{\infty} \frac{n}{n+2} (-x)^{n-1} \right], \quad (3.9)$$

which is truncated after taking the first temperature derivative of the heat capacity change. In Equation 3.9, $x = (T-\theta)/\theta$, and ΔG°_θ , ΔH°_θ , and $\Delta C_p^\circ_\theta$ represent these thermodynamic parameters at θ . Equation 3.9 can be written as:

$$R \ln K(T) = b_0 + b_1 t_1 + b_2 t_2 + b_3 t_3 \quad (3.10)$$

with

$$t_i = \sum_{n=1}^{\infty} \frac{n}{n+(i-1)} (-x)^{n-1}, \quad (3.11)$$

$$b_0 = -\frac{\Delta G^\circ_\theta}{\theta}, \quad (3.12)$$

$$b_1 = \frac{\Delta H^\circ_\theta}{\theta}, \quad (3.13)$$

$$b_2 = \Delta C_p^\circ \theta, \quad (3.14)$$

$$b_3 = \frac{\theta}{2} \left[\frac{d\Delta C^{\circ}_{P\theta}}{dT} \right]_{\theta} \quad (3.15)$$

Thus, the temperature dependent coefficients t_i ($i = 1-4$) around the reference temperature $\theta = 298.15$ K were calculated, and the equilibrium association constants were fit to Equation 3.10 in order to obtain the b_i partial regression coefficients. Propagation of errors in these coefficients were obtained from the diagonal elements of the error matrix as described (Clarke, 1966; Bevington, 1969; Dec, 1985).

3.2.1 Effects of phosphate neutralization on ΔH° , ΔS° , and ΔC°_P

Plots of $\ln K_A$ vs. temperature are shown in Figure 2.1 for the formation of wt and K119A complexes with unmodified and Rp-P_{Me-2} modified DNA substrates. Importantly, the specific activities, in regards to substrate binding, for both the wt and K119A enzymes were measured to be 100% at both ends of the experimental temperature range (4°C and 40°C; Section 6.2.4). The thermodynamic parameters calculated for the multiparametric fit agree very well with those calculated using the method of Clarke and Glew (Tables 3.3 and 3.4), and therefore, subsequent discussion will refer to values calculated from the Clarke and Glew method (Table 3.4), as the propagated errors for this method are smaller than those calculated for the multiparametric fit (Tables 3.3 and 3.4). Formation of all four complexes is characterized by a nonlinear dependence of $\ln K_A$ on temperature, resulting in a nonzero ΔC°_P (Table 3.4). As stated in the previous section, a large negative ΔC°_P is a thermodynamic signature of specific protein-DNA interactions and is attributable to the formation of a tight interface (Jen-Jacobson, 2000a). ΔC°_P

values reported here differ dramatically from those measured for the formation of the EcoRV-DNA complex under different anion conditions. It has been shown in the Jen-Jacobson lab that ΔC_oP for the formation of specific protein-DNA complexes is extremely dependent on chloride concentration (L.E. Engler, L. Jen-Jacobson, unpublished).

Further, all four complexes are characterized by an unfavorable $\Delta H^{\circ}_{\text{bind}}$ at 25°C, which is compensated by favorable $T\Delta S^{\circ}_{\text{bind}}$ (Table 3.4). This thermodynamic profile is characteristic of site-specific protein-DNA interactions in which there is a high degree of DNA distortion/bending (Jen-Jacobson, 2000b). Interphosphate repulsion is one likely contributor to the net unfavorable ΔH° observed for EcoRV-DNA complex. Thus, it is expected that phosphate neutralization will reduce the enthalpic cost of DNA bending due to interphosphate repulsion, and will, therefore, contribute favorable ΔH° to EcoRV-DNA complex formation.

3.2.1.1 Thermodynamics parameters: K119A-GATATC vs. wt-GATATC

Formation of the K119A-GATATC complex results in less favorable ΔH° and $T\Delta S^{\circ}$ relative to unmodified complex formation ($T\Delta\Delta S = -3.0 (\pm 0.8)$ kcal/mol; $\Delta\Delta H^{\circ} = +2.5 (\pm 0.8)$ kcal/mol; Table 3.4). Thus, the decrease in binding affinity ($\Delta\Delta G^{\circ}_{\text{bind}} = +5.5$ kcal/mol; Table 3.4) derives from a combination of unfavorable enthalpic and entropic factors. Both $\Delta\Delta H^{\circ}$ and $T\Delta\Delta S^{\circ}$ are constant over the experimental temperature range as the temperature dependence of these parameters (that is ΔC°_p) is the same for the formation of the K119A-GATATC complex relative to that for the unmodified complex. Further, ΔC°_p does not vary with temperature ($d\Delta C^{\circ}_p/dt \approx 0$; Table 3.4). The molecular factors that influence these thermodynamic observations will be considered below.

3.2.1.2 Thermodynamic parameters: K119A-Rp-P_{Me-2} vs. K119A-GATATC

Interestingly, formation of the K119A-Rp-P_{Me-2} complex results in a more favorable ΔH° relative to that for K119A-GATATC complex formation ($\Delta\Delta H^\circ = +6.3$ vs. 9.5 kcal/mol; Table 3.4), suggesting the phosphate neutralization at P₋₂ contributes -3.2 kcal/mol favorable ΔH° to K119A-DNA complex formation, while $T\Delta S^\circ$ is slightly less favorable ($T\Delta S^\circ = +18.4$ vs. $+19.3$ kcal/mol, respectively; Table 2.8). Notably, ΔC_p° for the formation of the K119A-Rp-P_{Me-2} is not significantly different than that observed for either wt or K119A-GATATC complex formation, suggesting that either the factors that contribute to ΔC_p° for these interactions are not significantly different from one another, or that there are compensating factors that contribute to ΔC_p° such that there is no *net* difference in ΔC_p° .

3.2.1.3 Thermodynamic parameters wt-Rp-P_{Me-2} vs wt-GATATC

The energetic penalty observed for wt-Rp-P_{Me-2} complex formation relative to that for wt-GATATC ($\Delta\Delta G_{\text{bind}}^\circ = 4.1$ kcal/mol; Table 3.4) results from a dramatic unfavorable enthalpic contribution to complex formation ($\Delta\Delta H_{\text{bind}}^\circ = +6.7(\pm 0.8)$ kcal/mol), which is partially compensated for by a favorable entropic contribution ($T\Delta\Delta S^\circ = +2.2$ kcal/mol). Further, wt-Rp-P_{Me-2} complex formation is associated with a significant increase in ΔC_p° ($\Delta\Delta C_p^\circ = +0.6$ kcal/mol·K), suggesting that the molecular factors that influence ΔC_p° are different for the formation of the wt-Rp-P_{Me-2} complex.

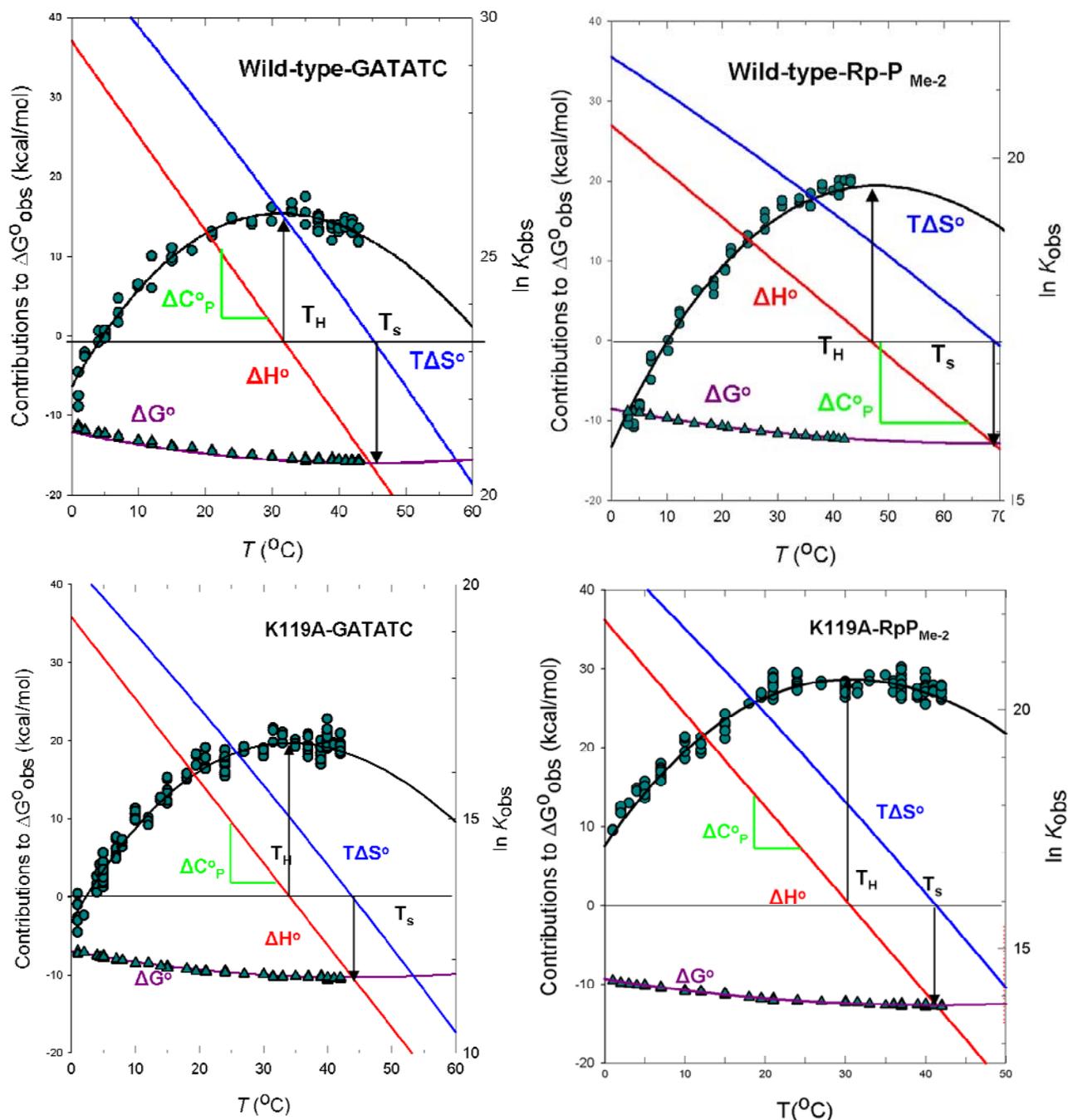


Figure 3.1 van't Hoff plots for the formation of modified and unmodified EcoRV-DNA complexes.

The equilibrium association constant for binding to the 24 bp oligonucleotide d(TGTGTTGTAGGATATCCTACAGGT) was measured as a function of temperature in 20 mM cacodylate, 0.24 M KCl, pH 7.25. The raw data for K_{obs} (green circles) are fit to the temperature dependence of K_{obs} (black line) using the van't Hoff equation (Eq. 2.8, see text). The enthalpy (red line) and entropy changes (blue line) were calculated with Equations 2.9 and 2.10, respectively. The experimental ΔG° values (green triangles) are shown and a fit over all temperatures (purple line) is calculated from $\Delta G^\circ_{bind} = -RT \ln K_{obs}$.

Table 3.3 Thermodynamic parameters for EcoRV-DNA binding obtained by a multi-parametric fit.

	Wt-GATATC	Wt-Rp-P _{Me-2}	K119A- GATATC	K119A- Rp-P _{Me-2}
ΔG° (kcal/mol) ^a	-15.3 (±1.4)	-11.2(±2.5)	-9.7(±1.3)	-12.1(±1.0)
ΔH° (kcal/mol)	+7.3(±0.6)	+13.4(±1.3)	+9.6(±0.7)	+6.5(±0.5)
$T\Delta S^\circ$ ((kcal/mol)	+22.6 (±1.3)	+24.6(±2.1)	+19.4(±1.1)	+18.6(±0.9)
ΔC°_P (kcal/mol ^o K)	-1.2(±0.1)	-0.7(±0.1)	-1.1(±0.1)	-1.1(±0.1)
$\Delta\Delta G^\circ$ (kcal/mol) ^b	0	+4.1(±2.9)	+5.6(±1.9)	+3.2(±1.7)
$\Delta\Delta H^\circ$ (kcal/mol)	0	+6.1(±1.4)	+2.3(±0.9)	-0.8(±0.8)
$T\Delta\Delta S^\circ$ (kcal/mol)	0	+2.0(±2.5)	-3.2(±1.7)	-5.0(±1.6)
$\Delta\Delta C^\circ_P$ (kcal/mol ^o K)	0	+0.5(±0.1)	+0.1(±0.1)	+0.1(±0.1)

^aThe temperature dependence of the equilibrium association constant K_A was measured for the double stranded 24 bp oligonucleotide d(TGTGTTGTAGGATATCCTACAGGT) in 20 mM cacodylate, 10 mM CaCl₂, 0.24 M KCl, at pH 7.25.

^bAll difference parameters (e.g. $\Delta\Delta G^\circ$) are referenced to the unmodified complex.

Table 3.4 Thermodynamic parameters for EcoRV-DNA binding using the method of Clark and Glew.

	Wt-GATATC	Wt-Rp-P _{Me-2}	K119A- GATATC	K119A- Rp-P _{Me-2}
ΔG° (kcal/mol) ^a	-15.3(±0.1)	-11.2(±0.1)	-9.8(±0.1)	-12.1(±0.1)
ΔH° (kcal/mol)	+7.0(±0.5)	+13.7(±0.6)	+9.5(±0.6)	+6.3(±0.6)
$T\Delta S^\circ$ ((kcal/mol)	+22.3(±0.5)	+24.9(±0.6)	+19.3(±0.6)	+18.4(±0.6)
ΔC°_P (kcal/mol ^o K)	-1.2(±0.1)	-0.6(±0.1)	-1.2(±0.1)	-1.1(±0.1)
$d\Delta C^\circ_P/dt$ (kcal/mol ^o K ²)	+0.01(±0.01)	-0.01(±0.01)	-0.01(±0.01)	+0.01(±0.01)
$\Delta\Delta G^\circ$ (kcal/mol) ^b	0	+4.1(±0.1)	+5.5(±0.1)	+3.2(±0.1)
$\Delta\Delta H^\circ$ (kcal/mol)	0	+6.7(±0.8)	+2.5(±0.8)	-0.7(±0.8)
$T\Delta\Delta S^\circ$ (kcal/mol)	0	+2.6(±0.8)	-3.0(±0.8)	-3.9(±0.8)
$\Delta\Delta C^\circ_P$ (kcal/mol ^o K)	0	+0.6(±0.1)	+0.0(±0.1)	+0.1(±0.1)

^aThe temperature dependence of the equilibrium association constant K_A was measured for the double stranded 24 bp oligonucleotide d(TGTGTTGTAGGATATCCTACAGGT) in 20 mM cacodylate, 10 mM CaCl₂, 0.24 M KCl, at pH 7.25.

^bAll difference parameters (e.g. $\Delta\Delta G^\circ$) are referenced to the unmodified complex.

3.2.2 Molecular dynamics simulations as a probe for the effects of electrostatic modifications on the specific EcoRV-DNA complex

Molecular dynamics (MD) simulations were also performed in order to gain insight into the structural differences between modified and unmodified complexes. I have performed these simulations for the free, wt and K119A proteins, as well as for the wt-GATATC, wt-Rp-P_{Me-2}, K119A-GATATC, and K119A-Rp-P_{Me-2} complexes with Ca²⁺ bound at the active site. The starting coordinates used for these simulations are from the crystal structure of an EcoRV-DNA-Ca²⁺ ternary complex (PDB entry 1B94) solved at 1.9 Å resolution (Figure 2.4) (Thomas et al., 1999). The desired structural perturbations were made *in silico* (Figure 3.2A), using the molecular graphics program, Midas, on a Silicon Graphics workstation. Hydrogen atoms were then added, and the complexes were immersed in explicit solvent containing water and enough Na⁺ and Cl⁻ ions to achieve a 0.24 M of each ion, which is the same as the concentration of these ions used in all of the thermodynamic experiments in this dissertation. Hydrogens and salt ions were added using the xLeaP module of Amber (Case, 2004) (Figure 3.2B). Each system was then subjected to an all-atom energy minimization to allow hydrogen and solvent molecules to find positions of lowest potential energy. Following this minimization, a short MD simulation was employed in which the protein and DNA were constrained, while the water molecules were allowed to equilibrate. This was followed by an all-atom MD simulation at constant pressure to allow the entire system to equilibrate. Upon completion of these equilibration steps, each system underwent a simulated annealing step in order to release the system from any local energy minimum. This was accomplished by raising the temperature of the system to 373° K (100°C) and allowing it to cool slowly back to room temperature (298 K; 25°C) (see Methods). The complexes were then subjected to production MD, during which all of the atoms were allowed to

move over time. Production dynamics were carried out in two ways; either the system was allowed to proceed for 6 ns continuously (written as wt-GATATC-continuous), or the system was minimized after every 0.5 ns over the course of a 6 ns simulation (written as wt-GATATC-min) (see Methods). In the following, the general and local structural features of these simulated models will be described.

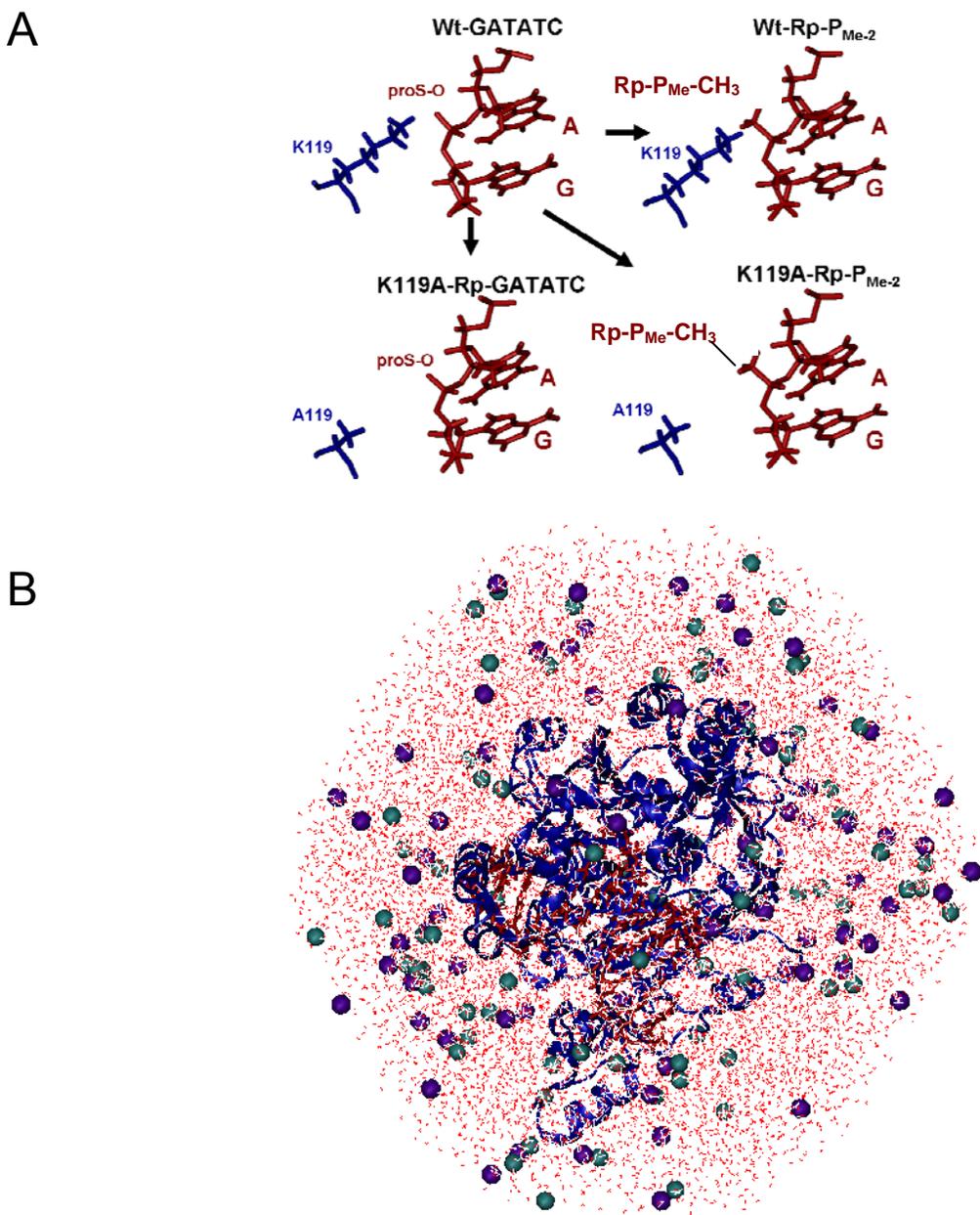


Figure 3.2 Electrostatic modifications and solvation of EcoRV-DNA complexes in preparation for molecular dynamics simulations.

(A) The *in silico* modifications to the specific EcoRV-DNA complex are shown. These graphical depictions represent the starting coordinates for the protein (blue) and DNA (red) in each of the four complexes, before MD simulation. Modified complexes contain either the K119 to A mutation (bottom left), the Rp-P_{Me-2} substitution (top right), or both (bottom right). (B) EcoRV-DNA (blue and red, respectively) complexes containing the modifications in (A) were immersed in solvent containing ~12,000 water molecules (red dots) and enough Na⁺ (cyan spheres) and Cl⁻ ions (purple spheres) to attain electroneutrality + 0.24 M NaCl in preparation for MD simulations.

3.2.3 Global structural features of molecular dynamic simulated structural models

Root mean squared deviations (RMSD) for protein backbone atoms were calculated for each of the simulated complexes over the course of the 6 ns production MD simulation relative to the backbone coordinates for each complex at the beginning of the simulation (i.e. prior to equilibration runs and simulated annealing; Figure 3.3). Both wt-GATATC simulations and the wt-Rp-P_{Me-2}-min simulation show RMSD values of about 1 Å relative to the starting model at the beginning of the 6 ns simulation and remain between approximately 1.0 and 1.25 Å over the course of the production MD simulation (Figure 3.3A). However, the wt-Rp-P_{Me-2}-continuous simulation shows a steadily increasing RMSD after 4 ns, to approximately 1.6 Å. For both K119A-GATATC simulations, RMSD values also remained between 1.0 and 1.25 Å over the course of the 6 ns production MD, whereas RMSD for the two K119A-RpP_{Me-2} simulations were slightly larger and ranged from 1.25 to 1.50 Å (Figure 3.3B). The fact that the change in RMSD stabilizes, even before production dynamics were begun, suggests that the positions of the backbone atoms have reached equilibrium in each of the complexes with the exception of the wt-Rp-P_{Me-2}-continuous simulation for which there is a steady increase in RMSD over the course of the simulation (Figure 3.3A; green plot). It should be noted that the lack of equilibration observed for the wt-Rp-P_{Me-2}-continuous simulation represents an anomaly relative to all other continuous simulations performed. Each of the other continuous simulations show that backbone RMSD fluctuations equilibrate quickly (even before the production MD step is begun), and remain equilibrated over the course of the 6 ns MD simulation. Further, continuous simulations performed on all four complexes under different solution conditions have shown that the backbone RMSD remains stable over 4.5 ns of production MD simulation time (not shown). These observations suggest that the increase in backbone RMSD for the wt-Rp-P_{Me-2}-continuous

simulation does not represent a systematic difference in the two different methods used for MD simulation (continuous vs. periodic minimizations), but rather, represents an isolated irregularity in the equilibration of this complex during the continuous MD simulation. However, because of the increase in backbone RMSD in the wt-Rp-P_{Me-2}-continuous simulation, I decided to use the simulations for which periodic all-atom energy minimizations were performed for the analysis of the global and local structural effects of each of the electrostatic perturbations (See below).

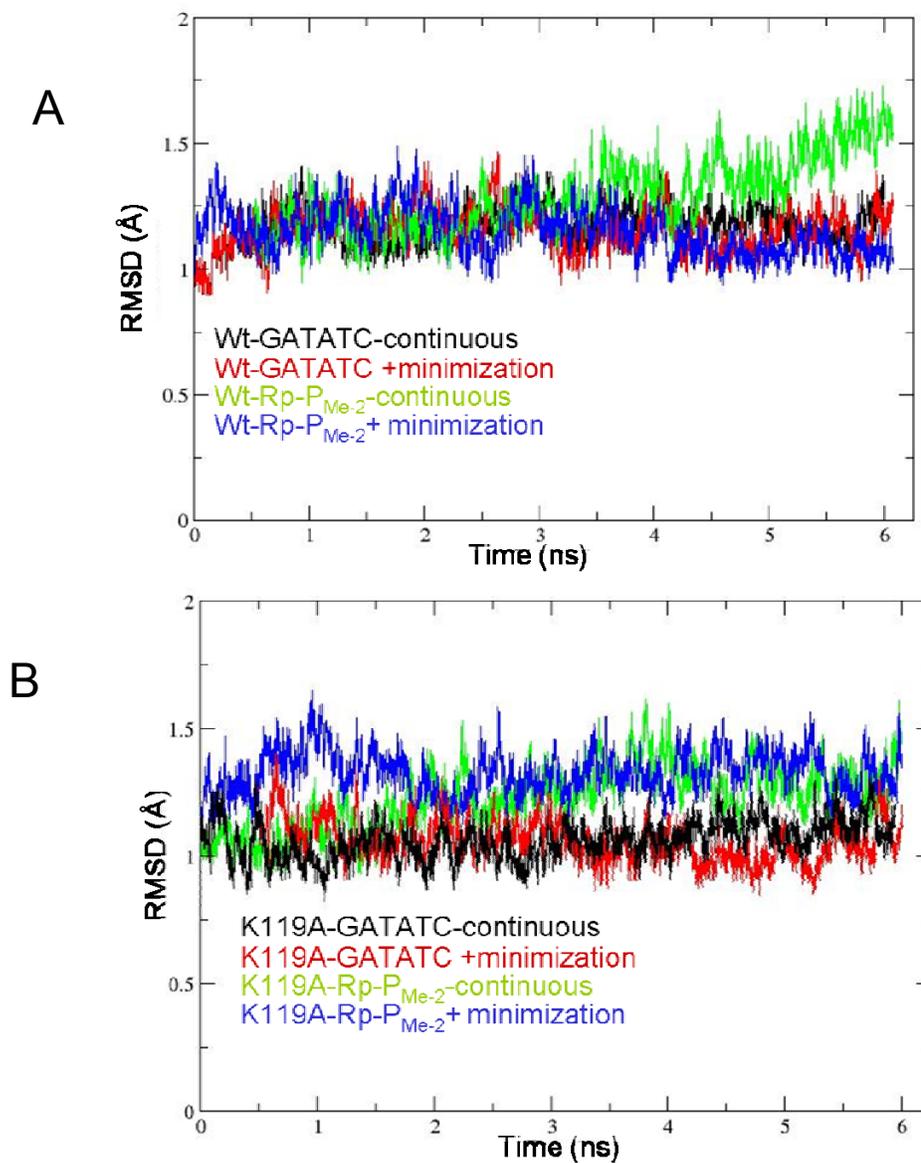


Figure 3.3 RMSD for modified and unmodified EcoRV-DNA complexes relative to their starting coordinates. RMSD plotted as a function of time for wt (A) and K119A (B) MD simulated complexes. In each panel, data for complexes with unmodified DNA, simulated continuously (black) or with minimization (red), accompany that for complexes with Rp-P_{Me-2} modified DNA, simulated continuously (green) or with minimization (blue).

Further support for the use of MD simulations in for which periodic all-atom energy minimization were conducted comes from an analysis of distance dynamics between protein side chains and DNA phosphates in both minimized and continuous MD simulations. I have observed side chain-DNA distance dynamics that describe both stable protein-phosphate interactions, as well as transient protein-phosphate interactions (Figure 3.4). For example, the interaction between the side chain hydroxyl of T111 (T111:OG1) and the pro-R phosphoryl oxygen at P₂ (see Figure 2.1 for phosphate nomenclature) is very stable in the wt-GATATC-minimized simulation (Figure 3.4A). In contrast the R221 side is able to transiently interact with both the P₂ and P₃ phosphates while also adopting positions that contact neither of these phosphates (Figure 3.4C). This is consistent with the observation that in various X-ray crystal models the R221 side chain interacts directly with P₂ in half-site A, but also adopts a position that makes a water mediated interaction with P₃, suggesting that there is a degree of conformational flexibility associated with this sidechain. Recall that the P_{Me-3} substitution shows an unexpected penalty for wild type EcoRV complex formation ($\Delta\Delta G_{\text{bind}}^{\circ} = +1.5$ kcal/mol). The unexpectedness of this modification comes from the fact that there are no contacts made to the P₃ phosphate in the crystal model. The fact that the R221 side chain is able to transiently contact the P₃ phosphate is consistent with the penalty observed for the P_{Me-3} substitution, which may result from removal of the R221-P-3 interaction, or from reduced conformational flexibility of the R221 side chain. The ability to detect both stable and transient protein DNA interactions suggests that periodic minimizations allow for sufficient dynamics of protein side chains and do not function to restrain atomic positions. In fact, the R221 side chain position relative to the P₂ and P₃ phosphates may be even more dynamic in the minimized simulation than in the continuous simulation (Figure 3.4 compare B vs. D). Further, distance plots for the minimized

MD simulations are similar to those observed for continuous simulations (Figures 3.4 B and D), suggesting that there is similar side chain behavior in both types of MD simulation. These observations validate the use of minimized simulations in understanding the effects of electrostatic perturbations (K119A and Rp-P_{Me-2}) on the structure and dynamics of specific EcoRV-DNA complexes.

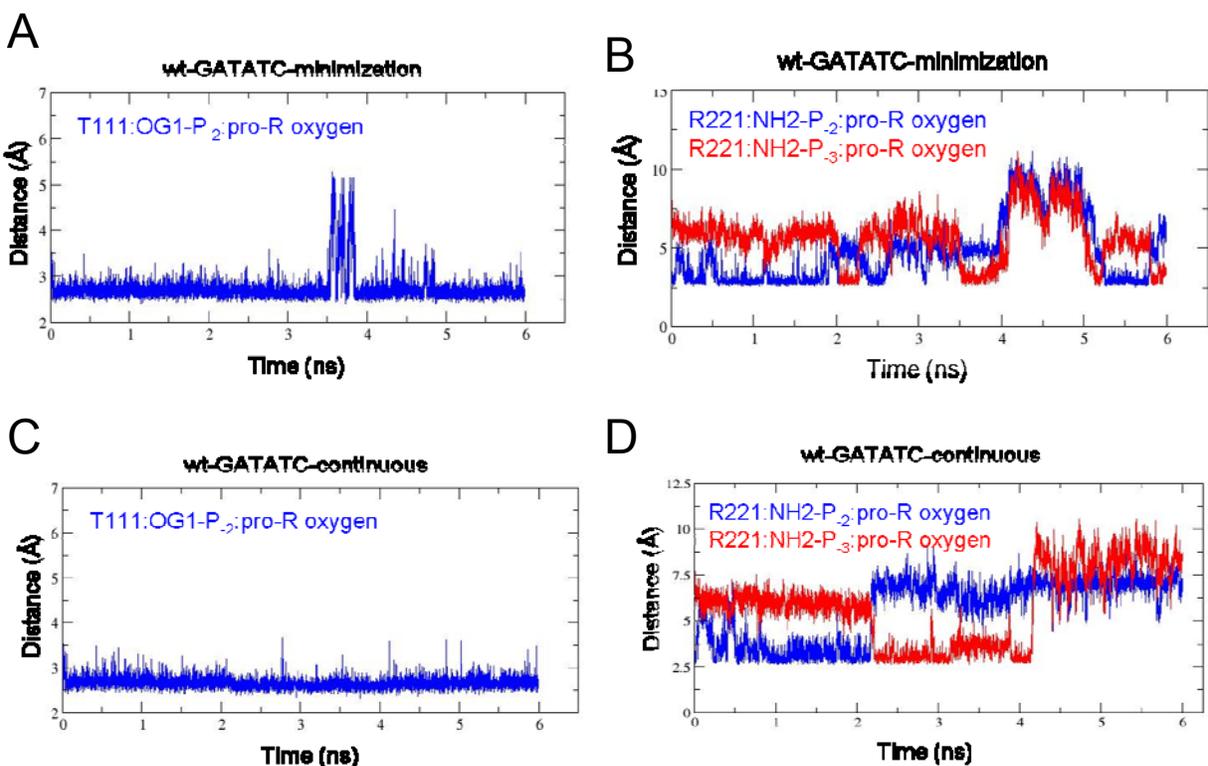


Figure 3.4 Side chain dynamics in minimized and continuous MD simulations.

Plots of interatomic distance as a function of simulation time are shown for either minimized (A and B), or continuous (C and D) MD simulations of the wt-GATATC complex. These plots show the distance between the T111:OG1 side chain atom and the P₂ pro-R phosphoryl oxygen (A and C) as well as the distance between the R221:NH2 side chain amine and either the P₂ pro-R phosphoryl oxygen (blue) or the P₃ pro-R phosphoryl oxygen (red) (B and D). In both the continuous and minimized MD simulations, the R221 side chain is able to adopt positions such that it is not in contact with either P₂ or P₃. In this case it is able to make water mediated interactions to both phosphates, but does not contact them directly.

In order to investigate the global structural effects of the Rp-P_{Me-2} and K119A modifications, snapshots of the wt-Rp-P_{Me-2}, K119A-GATATC, and K119A-Rp-P_{Me-} complexes were averaged over the final 1 ns of the 6 ns MD simulation. The backbone atoms from these averaged structures were then aligned, and the RMSD was calculated relative to the averaged structure of the unmodified complex (Figure 3.5). The RMSD for the protein backbone atoms of the modified complexes ranged from 1.02 to 1.08 Å relative to the unmodified complex, suggesting a relatively high degree of structural similarity between the modified complexes and the unmodified complex.

RMSD were also calculated on a per residue basis for all four MD simulated complexes relative to the starting model (Figure 3.6). These per residue backbone RMSD seem to be at least somewhat systematic, in that for all four complexes, there is a peak in the RMSD value at residue 90 in chain A, and for three of the four complexes (wt-GATATC, wt-Rp-P_{Me-2}, and K119A-Rp-P_{Me-2}) there is a peak in the RMSD value at residue 30. This may suggest that certain segments of the EcoRV-DNA complex were released from possible crystal lattice constraints in these MD simulations, allowing for significant conformational adaptation. Further, asymmetric RMSD were observed between half-site A and B for each complex, leading to slight structural differences that are observed between the two half-sites of a particular complex. This observation is not unexpected in light of structural asymmetries observed in a sampling of four X-ray crystal models of the EcoRV-DNA complex. These structural models have been solved by different research groups at 1.9 or 2.0 Å resolution, both in the presence and absence of Ca²⁺

(a divalent metal that binds at the EcoRV-DNA active site, but does not permit cleavage) (Table 3.5). In several cases, these asymmetries are revealed as a difference in the distance between basic side chain residues and the DNA phosphate backbone in half-site A relative to half-site B (Table 3.5). In the case of the 1BGB and 1AZO models, the K119 side chain has not even been assigned, suggesting that it has significant mobility in these crystals. Subsequent discussion will focus on the local structural impact of the K119A and Rp-P_{Me-2} modifications in half-site A, although in general; similar inferences of structural and energetic correlates can be made for both half-sites.

Backbone RMSD

Wt-GATATC	0.00
Wt-Rp-PMe ₂	1.02
K119A-GATATC	1.04
K119A-Rp-PMe ₂	1.08

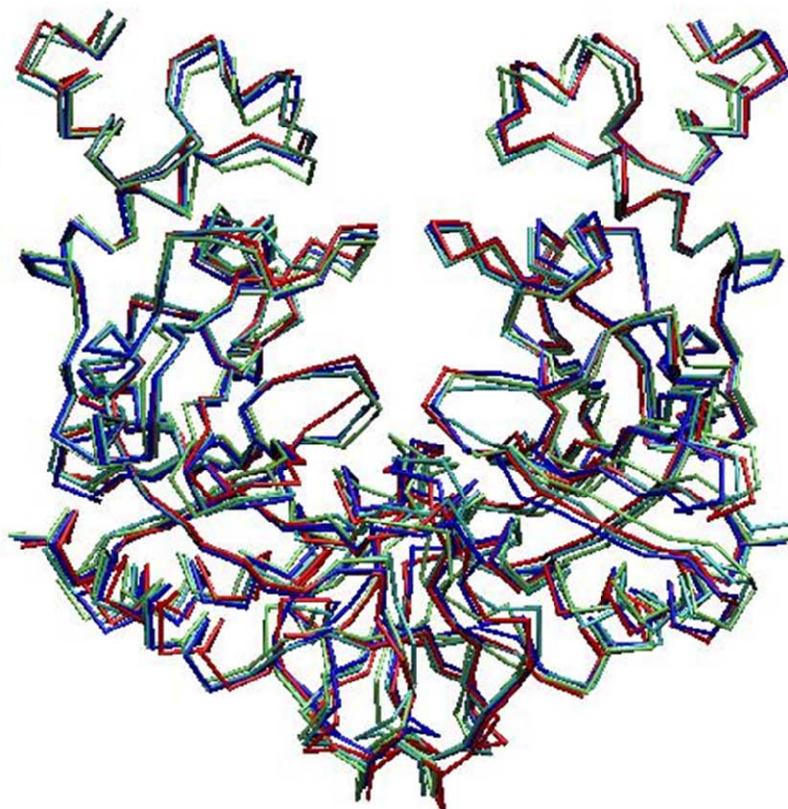
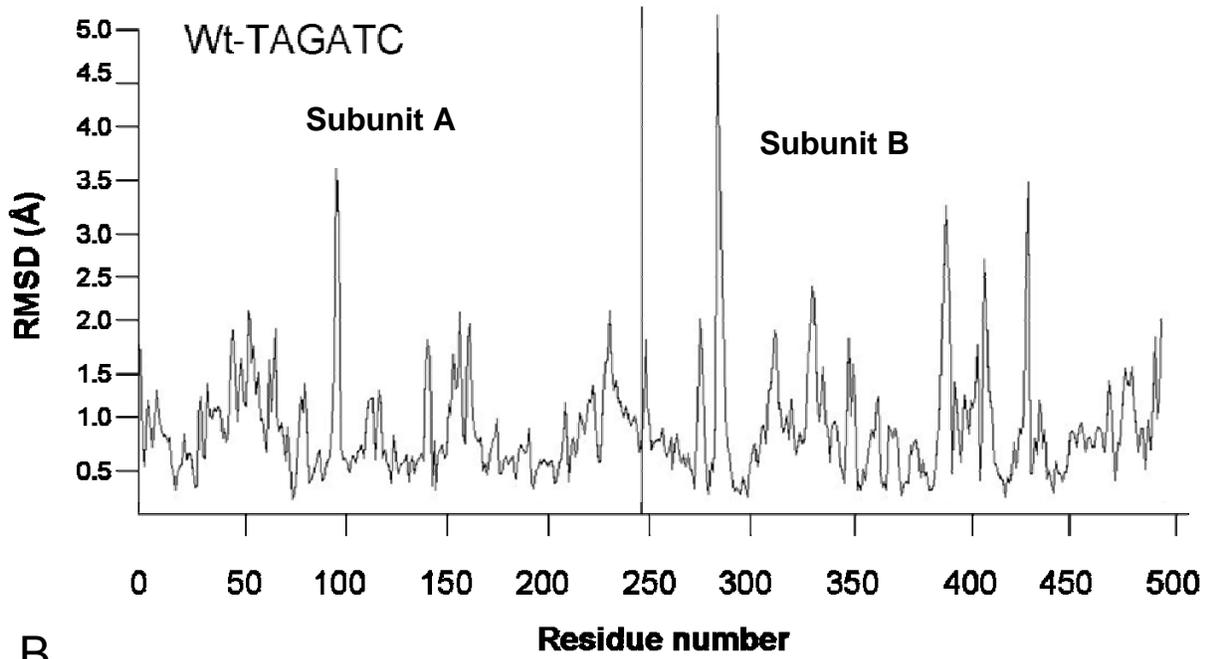


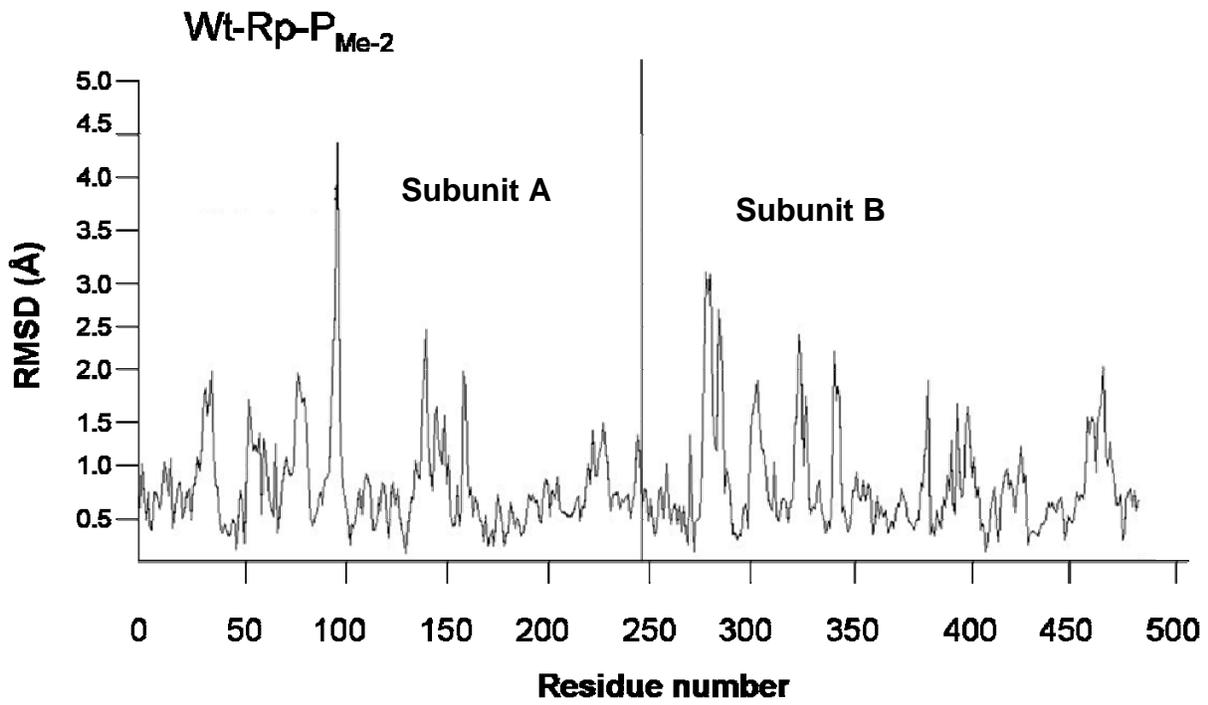
Figure 3.5 Backbone RMSD for modified and unmodified EcoRV-DNA complexes.

(A) An overlay of the backbone traces for the average structure from the final 1 ns of a 6 ns molecular dynamics simulation is shown for the wt-GATATC (cyan), wt-Rp-PMe₂ (blue), K119A-GATATC (red), and K119A-Rp-PMe₂ (light green) EcoRV-DNA complexes. Backbones were superimposed with an RMSD equal to that shown for each complex relative to the unmodified complex.

A



B



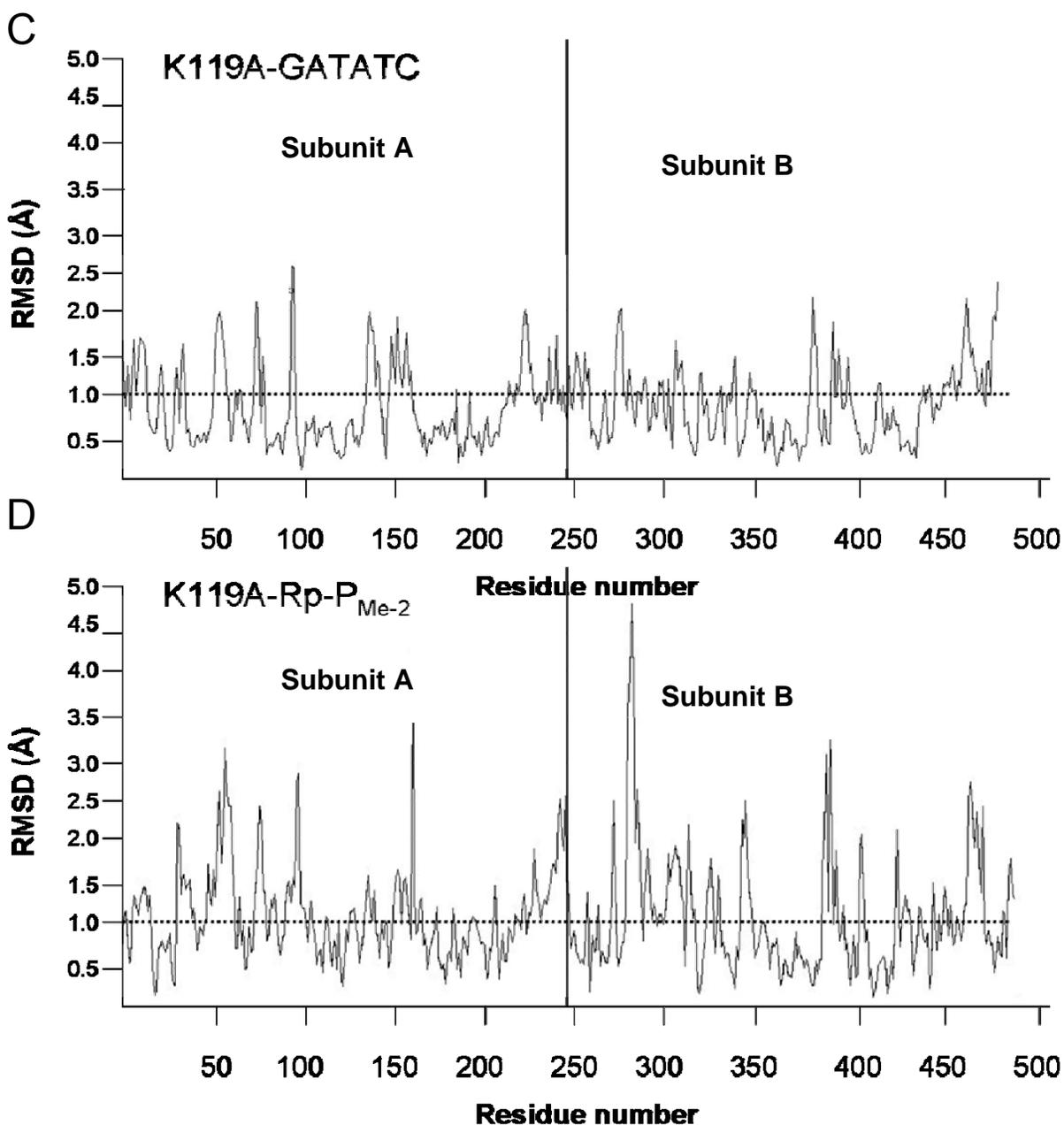


Figure 3.6 Per residue RMSD for K119A and P_{Me-2} modified complexes relative to the unmodified complex. Plots of RMSD vs. residue number are shown for the averaged structures of the wt-GATATC (A) wt-Rp- P_{Me-2} (B) K119A-GATATC (C), and K119A-Rp- P_{Me-2} (D) complexes relative to the unmodified complex. In these plots, residues 1 to 244 comprise subunit A of the protein molecule, while residues 245 to 488 comprise subunit B. The dotted line represents the average RMSD as calculated for the averaged structure (Figure 3.5), and the solid vertical line represents the boundary between protein subunits A and B.

Table 3.5 Structural asymmetries in various EcoRV-DNA X-ray crystal models.

PDB entry	Metal	Resolution	K119 to P ₋₂ (Å)		R226 to P ₋₅ (Å)		R221 to P ₋₂ (Å)		R140 to P ₊₃ (Å)	
			Chain A	Chain B						
1B94 ^a	Ca ²⁺	1.90Å	2.68	2.65	3.89	3.32	3.18	5.31	5.25	3.60
1RVA ^b	None	2.00Å	2.57	3.06	3.87	3.87	3.37	5.50	4.82	3.60
1AZ0 ^c	Ca ²⁺	2.00Å	2.67	NA	3.73	4.02	2.89	5.40	NA	NA
1BGB ^d	None	2.00Å	3.42	NA	3.65	4.27	5.49	5.65	NA	NA

^a(Thomas et al., 1999)
^b(Kostrewa and Winkler, 1995)
^c(Perona and Martin, 1997)
^d(Horton and Perona, 1998b)

3.2.4 Local structural effects of modified EcoRV-DNA complexes.

Representative snapshots for each of the simulated complexes were examined in order to elucidate the structural effects of each of the modified complexes relative to the unmodified complex. The wt-RpP_{Me-2} complex reveals several structural differences relative to the unmodified complex. In the wild type complex with unmodified DNA, the pro-S phosphoryl oxygen at the P₋₂ position (see Figure 1.10 for P_{Me} stereochemistry and naming convention) is contacted by the K119-Nε, which also makes a buttressing interaction with the main-chain carbonyl oxygen of T111 (Figure 3.7A). In the wt-Rp-P_{Me-2} complex, the pro-S phosphoryl oxygen has been replaced by a methyl group, and thus, the potential for a stabilizing interaction between the K119-Nε and a phosphoryl oxygen has been removed. As a consequence, the K119 side chain migrates toward the minor groove of the bound DNA substrate, leading to a loss of the interaction with the main-chain carbonyl oxygen of T111 (Figure 3.7B) and a significant increase in the degree of atomic fluctuation for the K119 side chain relative to that observed in the

unmodified complex. Further, in the wt-Rp-P_{Me-2} complex, the K119:Ne is now completely solvated by mobile water molecules, while the methylene groups are shielded from solvent by the backbone of T111 and S112 (not shown). In addition, the Rp-P_{Me-2}-CH₃ is shielded from solvent by the K119 side chain (Figure 3.7B), as well as by the backbone of T111 and S112 (not shown). Interestingly, the interactions between the pro-R phosphoryl oxygen and the T111-O γ are preserved in the wt-Rp-P_{Me-2} complex, as is the water-mediated interaction between T187-O γ and the pro-R phosphoryl oxygen (Figure 3.7A and B). In both the wt-GATATC and wt-Rp-P_{Me-2} complexes, the R221 side chain has moved away from Rp-P_{Me-2} to contact the phosphate at position P₃ (not shown). The R221 side chain has been observed to be quite dynamic and is able to sample a wide range of conformational space in all of the MD simulated complexes. Finally, in the unmodified complex, the P₂ phosphate is networked to the GATATC base by several water-mediated contacts that proceed from GATATC-O4 to the pro-R phosphoryl oxygen through T186-O γ 1, two water molecules, and T111 (Figure 3.7C). This network is completely conserved in the wt-Rp-P_{Me-2} complex (Figure 3.7D). Taken together, these observations suggest that the structural effects of the Rp-P_{Me-2} modification are quite subtle and manifest, primarily, in the migration of the K119 side chain toward the minor groove, resulting in increased dynamics for this side chain, the loss of a stabilizing interaction to the main chain of T111, and the loss of an interaction between K119A-Ne and the pro-S phosphoryl oxygen at P₂. The contributions of these structural effects and their thermodynamic consequences will be discussed in Section 3.2.5.

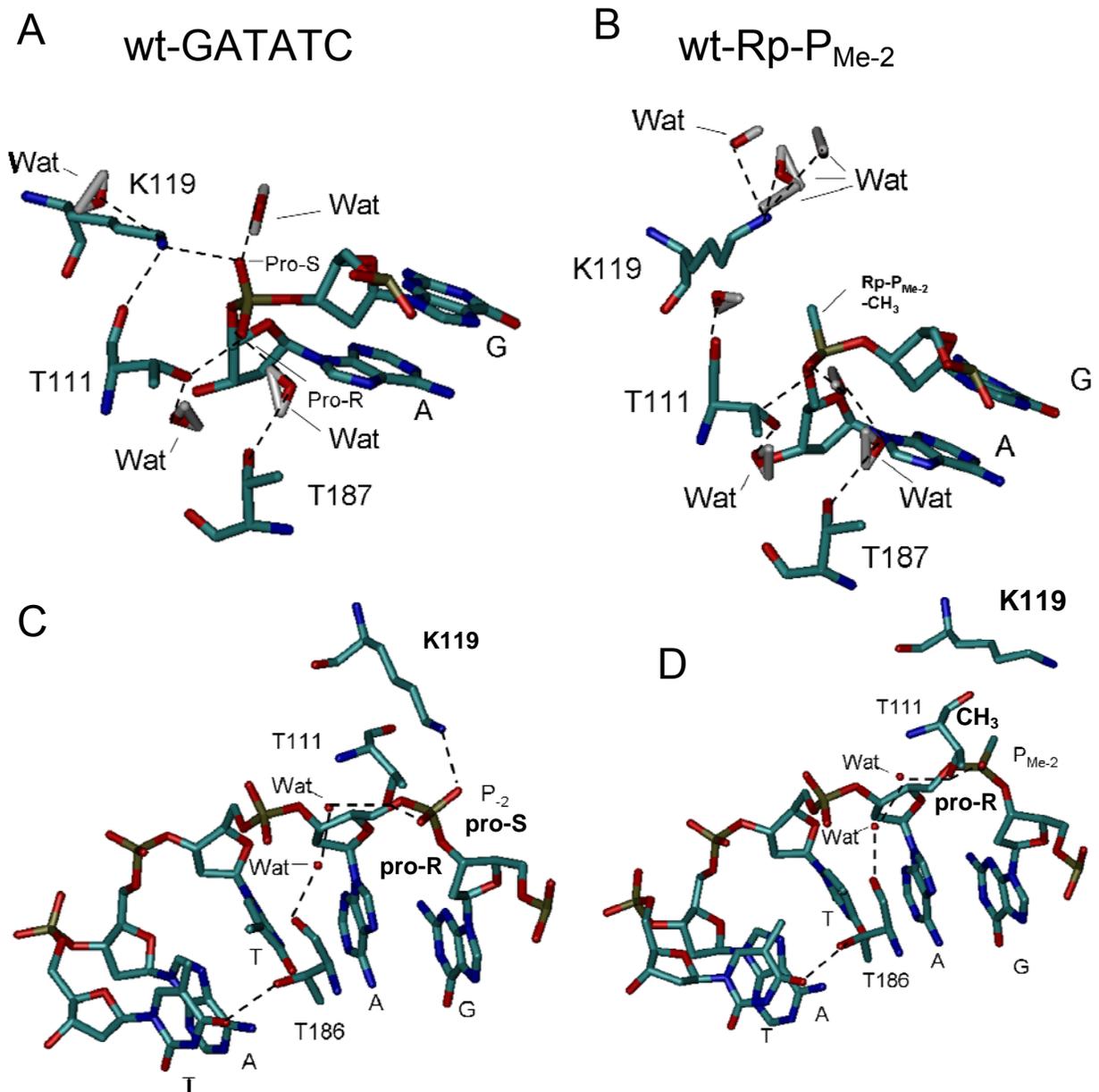


Figure 3.7 Structural consequences of the Rp-P_{Me-2} modification.

(A and C) Wire frame models of the wt-GATATC and (B and D) the wt-Rp-P_{Me-2} complexes. (A) The P₂ phosphate is contacted by K119, T111, and T187 via water in the wt-GATATC complex. (B) Interactions to the pro-R phosphoryl oxygen are conserved in the half-site A of the wt-Rp-P_{Me-2} modified complex, while the K119 side chain moves away from the Rp-P_{Me-2}-CH₃ and toward the minor groove. (C) Water-mediated network of contacts bridging the GATATC base to the P₂ phosphates via T186. (D) These interactions are completely conserved in half-site A of the wt-Rp-P_{Me-2} complex.

In the K119A-GATATC complex, the pro-S phosphoryl oxygen, which is contacted by K119 in the unmodified complex (Figure 3.8), becomes highly solvated. The interaction between K119A-N ϵ and the pro-S phosphoryl oxygen has been replaced by a water molecule, as has the interaction between K119A-N ϵ and the main chain oxygen of T111 (Figure 3.8). The water molecules that surround the pro-S phosphoryl oxygen a P₂ are not stably bound, and have very short residence times (25-100 ps). Again both the T187 water-mediated interaction with the pro-R phosphoryl oxygen (Figure 3.8) and the water-mediated networking of pro-R phosphoryl oxygen to GATATC:O4 (not shown) are conserved in the K119A-GATATC complex.

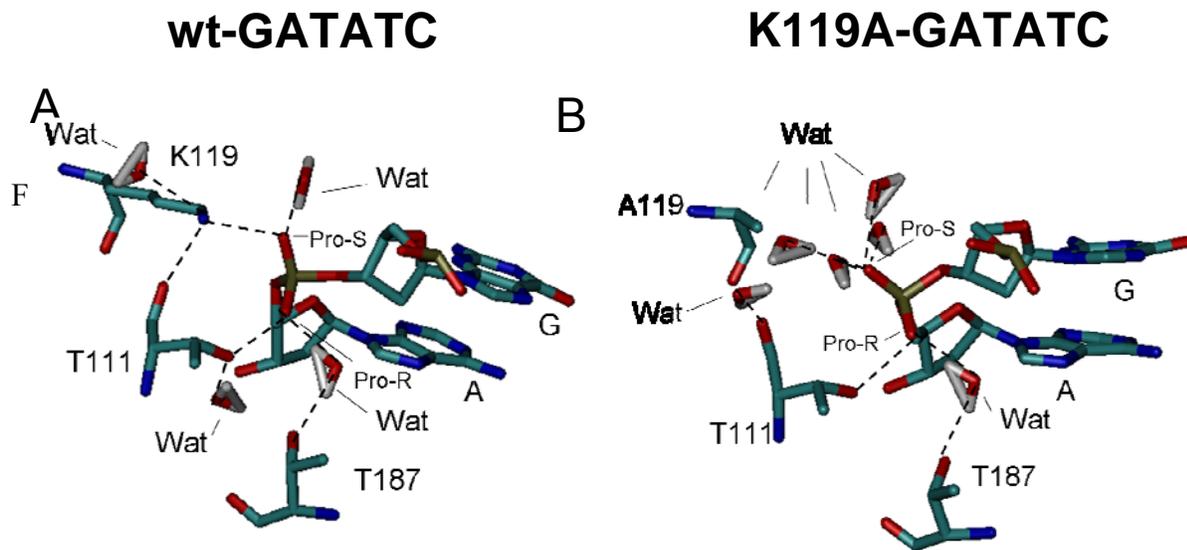


Figure 3.8 Structural consequences of the K119A mutation.

(A) Wire frame model for wt-GATATC half-site A. The P₂ phosphate is contacted by K119, T111 and water; the T187 water-mediated contact is not shown. K119 also interacts with the main chain of T111. (B) K119A-GATATC half-site A. In the absence of the K119 side chain, the pro-S phosphoryl oxygen is contacted by several waters, and the K119-Ne interaction with T111-O has been replaced by a water molecule. Contacts to the pro-R phosphoryl oxygen have been preserved.

The only chemical difference between the K119A-Rp-P_{Me-2} and K119A-GATATC complex is the replacement of the pro-S phosphoryl oxygen with a methyl group, yielding the Rp-P_{Me-2}. In the K119A-Rp-P_{Me-2} complex, the Rp-P_{Me-2}-CH₃ is surrounded by a shell-like structure of water molecules that interact with one another via hydrogen bonds (Figure 3.9A), whereas in the K119A-GATATC complex, the polar pro-S phosphoryl oxygen at the P₂ hydrogen bonds directly to water molecules that surround it (Figure 3.9B). In both complexes, waters surrounding the Rp-P_{Me-2}-CH₃ or pro-S phosphoryl oxygen are dynamic and exchange with the bulk solvent readily. Both the interaction between T111-O_{γ1} and the Rp-P_{Me-2} phosphoryl oxygen (Figure 3.9A) and the water-mediated interaction between T187-O_{γ1} and the Rp-P_{Me-2} phosphoryl oxygen are maintained in the K119A-Rp-P_{Me-2} complex (not shown), as is the water/T186-mediated network of interactions that bridge the GATATC:O4 to the phosphoryl oxygen of the Rp-P_{Me-2} (not shown). Importantly, the Rp-P_{Me-2} modification does not result in the formation of a new interaction between the Rp-P_{Me-2}-CH₃ and the A119CH₃. The distance between these atoms ranges from approximately 4.5 to 8.5 Å over the course of the 6 ns MD simulation (Figure 3.10; black plot), which is greater than the distance expected for a methyl-methyl van der Waals interaction (~4.0 Å). The case is similar for half-site B, in which this distance begins at approximately 6 Å and fluctuates between 4 and 6 Å over the course of the trajectory (Figure 3.10; red plot). The distance distribution for the A119-CH₃ and the Rp-P_{Me-2}-CH₃ in both half-sites differs dramatically from that observed between the T186-CH₃ and the GATATC:C5-CH₃ (Figure 3.10; green plot). The T186-CH₃ to GATATC:C5-CH₃ distance trajectory serves as an example of the distance distribution expected for a stable van der Waals interaction over the course of the MD simulation. These trajectories seem to flatten out, after 5 ns in half-site A and after only approximately 2 ns in half-site B, suggesting that the distance

between A119-CH₃ and the Rp-P_{Me-2}-CH₃ has stabilized in these simulations. Importantly, a 4.5 ns MD simulations done under more dilute salt conditions (only 12 Na⁺ and 12 Cl⁻ ions), and a 6ns MD simulation done in a periodic cube instead of the truncated octahedron that was used in the above simulations, also suggest that there is no stable van der Waals interaction formed between the K119A-CH₃ and the Rp-P_{Me-2} modification in the K119A-Rp-P_{Me-2} complex (not shown). Taken together, these observations suggest that in the context of the K119A mutant protein, the subtle chemical perturbation of replacing the pro-S phosphoryl oxygen with a methyl group results in differences in the interactions of this methyl group with solvent and differences in the electrostatics at this position, and not the addition or subtraction of interactions between protein and DNA functional groups.

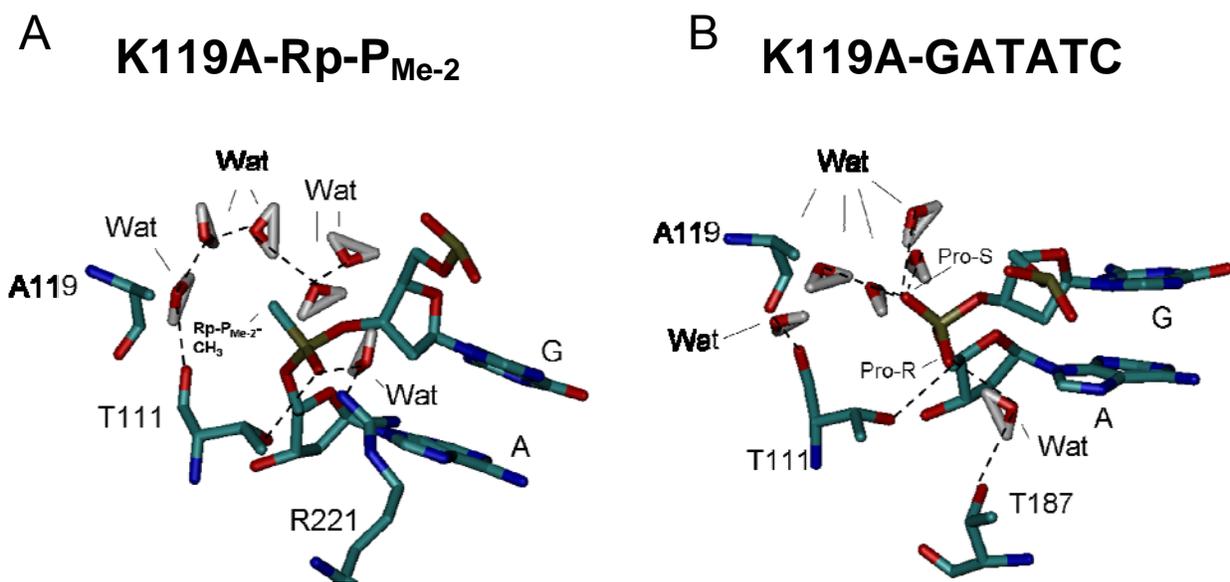


Figure 3.9 Structural features of the K119A-Rp-P_{Me-2} complex.

A) The Rp-P_{Me-2}-CH₃ is surrounded by water molecules, which form an ordered shell-like structure. The T111-O_γ contact to the phosphoryl oxygen is maintained, as is the T187, water-mediated interaction. Comparison between the K119A-Rp-P_{Me-2} and the K119A-GATATC (B) structures suggests that no new contacts are made to the Rp-P_{Me-2}-CH₃, and the structural differences derive, instead, from interactions with water.

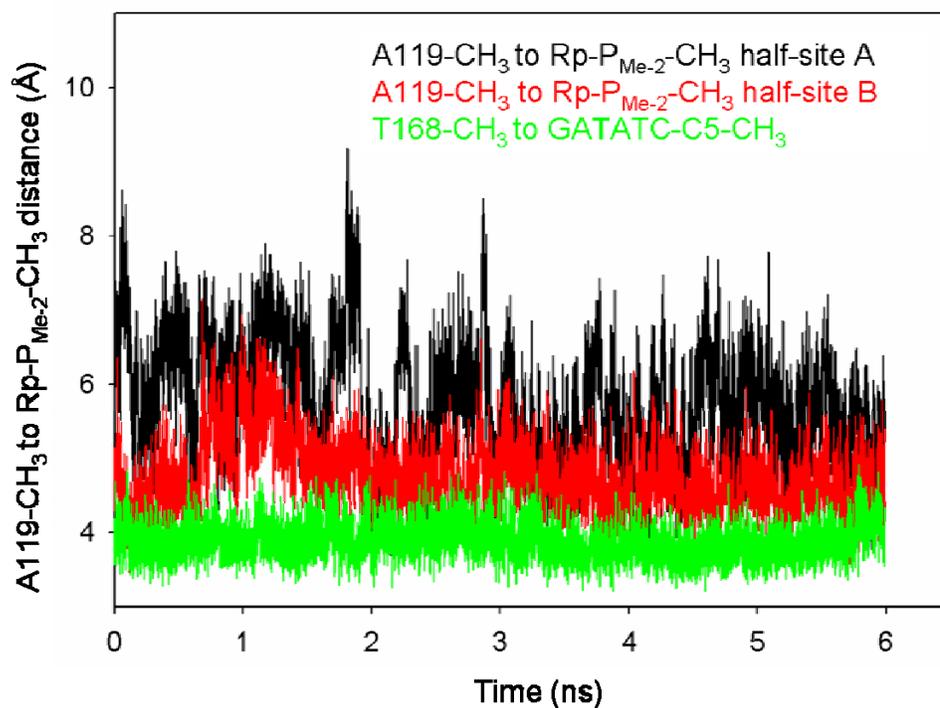


Figure 3.10 A119-CH₃ distances to Rp-P_{Me-2}-CH₃ in the K119A-Rp-P_{Me-2} complex.

Plot of the A119-CH₃ distance to Rp-P_{Me-2}-CH₃ in the K119A-Rp-P_{Me-2} complex as a function of time. Distances for half-site A (black line) and half-site B (red line), and the distance between the T186-CH₃ and the C5-CH₃ of GATATC base (green line) are indicated.

3.2.5 Relating structural observations to thermodynamic data

Data from MD simulations have shown various structural differences between modified and unmodified EcoRV-DNA complexes at the sites of modification. While it is difficult to attribute differences in the observed thermodynamic parameters to a specific structural difference between a modified and an unmodified complex, it is often possible to understand whether a particular structural feature will influence ΔH° , ΔS° , and ΔC_p° , and whether this contribution will be favorable or unfavorable. To this end, the following will address how these structural changes can give insight into the molecular origins of the thermodynamic differences observed for the formation of the modified EcoRV-DNA complexes.

One method of displaying the thermodynamic consequences of the K119A and Rp-P_{Me-2} modifications on EcoRV-DNA complex formation is by utilizing the type of thermodynamic pseudocycle shown in Figure 3.11. Each leg of this pseudocycle describes the energetic effects of making a particular structural modification on the thermodynamics of forming the modified complex relative to the complex at the origin of the arrow. The thermodynamic parameters shown in this pseudocycle correspond to those in Table 3.4, as calculated by the method of Clark and Glew. This pseudocycle will serve as a guide in the following discussion concerning the relationship between structural observations and the thermodynamic consequences of a particular modification.

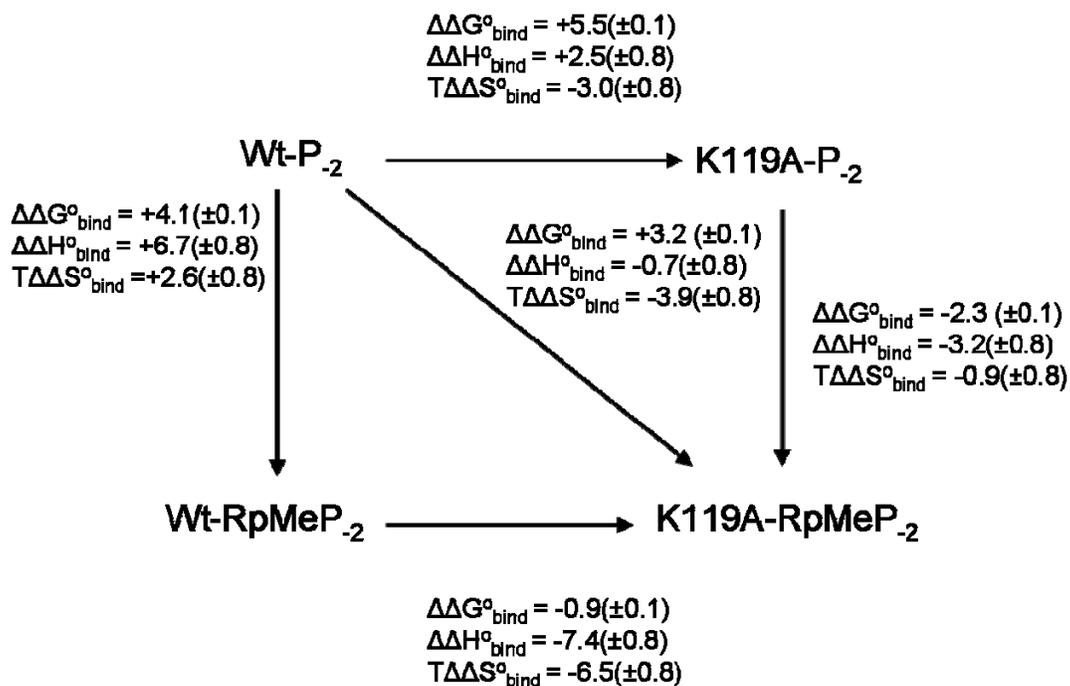


Figure 3.11 Thermodynamic pseudocycle for modified and unmodified complex formation.

Each leg of the cycle represents the thermodynamic consequences of forming a modified complex relative to the complex at the origin of the arrow. For example, the right hand vertical leg of the cycle represents the difference in the thermodynamic parameters for the formation of the K119A-Rp-PMe-2 relative to those for the formation of the K119A-GATATC complex.

Each corner of the pseudocycle in Figure 3.11 represents one of the four EcoRV-DNA complexes, which are either unmodified, or contain one or both of the K119A and Rp-P_{Me-2} modifications. It must be kept in mind that the thermodynamic consequences of these modifications result from the energetic differences in going from the free macromolecules to the complex. Thus, the measured $\Delta\Delta G^{\circ}_{bind}$ that appears on each leg represents the effect of modification on the protein-DNA complex *minus* the effect on the free molecules; that is (as noted in Chapter 2),

$$\Delta G^{\circ}_{bind,unmod} + \Delta G^{\circ}_{mod,complex} = \Delta G^{\circ}_{bind,mod} + \Delta G^{\circ}_{mod,free}, \quad (3.16)$$

where the subscript mod refers to the hypothetical free energy change of making a particular modification in either the free molecule ($\Delta G^{\circ}_{mod,free}$) or in the complex ($\Delta G^{\circ}_{mod,complex}$). Recall that it is impossible to directly measure the impact of these structural modifications directly, either in the free molecules, or in the complex, as this change does not correspond to any experimentally observable physical process. Instead, the change in binding free energy (ΔG°_{bind}) is measured for the modified and unmodified complexes. From Equation 3.4 it follows that the experimentally observed differences in binding free energy

$$\Delta\Delta G^{\circ}_{bind} = \Delta G^{\circ}_{bind,mod} - \Delta G^{\circ}_{bind,unmod}, \quad (3.17)$$

can be given by

$$\Delta\Delta G^{\circ}_{bind} = \Delta G^{\circ}_{mut,complex} - \Delta G^{\circ}_{mut,free}. \quad (3.18)$$

Corresponding equations can also be written for the thermodynamic parameters ΔH° and ΔS° that are associated with these processes. Each of the modifications can, in principle, influence the entropy, enthalpy and free energy of either, the protein-DNA complex, the free macromolecules, or both. Structural data from the MD simulations for modified and unmodified complexes can provide information as to the features that influence $\Delta G_{\text{mod,complex}}$, while structural information for the free protein and DNA molecules (both modified and unmodified) would be required to understand the impact of these modifications on $\Delta G_{\text{mod,free}}^\circ$. Again it should be noted that structural models of the free EcoRV protein (Winkler, 1993; Perona and Martin, 1997), the EcoRV-DNA complex in the absence of metal (Winkler, 1993; Kostrewa and Winkler, 1995; Perona and Martin, 1997; Horton and Perona, 1998a; Thomas et al., 1999) and the EcoRV-DNA complex in the presence of various divalent metals (Kostrewa and Winkler, 1995; Thomas et al., 1999; Horton and Perona, 2004) are available. These structural models, along with structural data from MD simulations can potentially provide insight into the possible structural consequences of a particular perturbation. It should be kept in mind that there is no data for the effects of particular P_{Me} substitutions on the structure of the free DNA molecule.

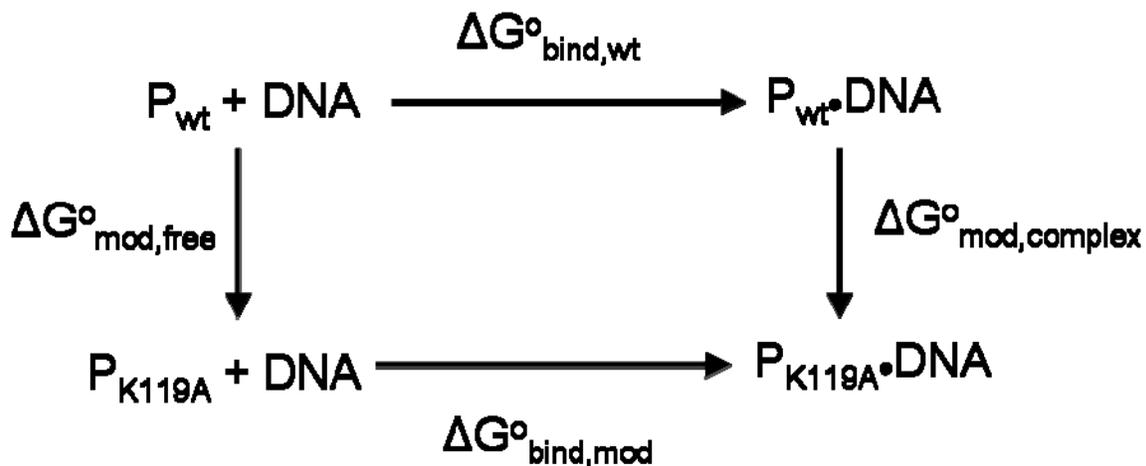


Figure 3.12 Thermodynamic pseudocycle for thermodynamic consequences K119A mutation.

Binding free energies ($\Delta G^{\circ}_{bind,x}$) for either wt (P_{wt}) or K119A (P_{K119A}) proteins interacting with the unmodified DNA substrate (DNA). This pseudocycle represents the progression from the free macromolecules ($P_{wt} + DNA$ or $P_{K119A} + DNA$) to the complex ($P_{wt} \bullet DNA$ or $P_{K119A} \bullet DNA$). Binding free energies ($\Delta G^{\circ}_{bind,wt}$ and $\Delta G^{\circ}_{bind,mod}$) can be measured experimentally. The free energies for making the K119A substitution in the free protein ($\Delta G^{\circ}_{mod,free}$) and in the complex ($\Delta G^{\circ}_{mod,complex}$) are also represented. These processes are thought to be nonphysical and cannot be probed by experimentation.

3.2.5.1 Phosphate neutralization contributes favorable $\Delta H_{\text{bind}}^{\circ}$ to EcoRV-DNA complex formation

The K119A mutant protein binds more tightly to DNA containing the Rp-P_{Me-2} substitution than to the unmodified substrate ($\Delta\Delta G_{\text{bind}}^{\circ} = -2.3$ kcal/mol; Figure 3.11, right leg). The energetic benefit of the Rp-P_{Me-2} substitution derive from a favorable enthalpic contribution to complex formation ($\Delta\Delta H_{\text{bind}}^{\circ} = -3.2$ kcal/mol), which is offset by a modest unfavorable entropic contribution ($T\Delta\Delta S^{\circ} = +0.9$ kcal/mol; Figure 3.11). In the context of the K119A mutation, the Rp-P_{Me-2} substitution has very subtle structural consequences and results in neither the loss, nor the formation of any of additional protein-DNA or protein-protein interactions. However, the Rp-P_{Me-2} modification does result in the ordering of water molecules around the nonpolar methyl group (Figure 3.9A). This is expected to be the case in both the free and bound DNA molecules, and therefore, would not contribute to the thermodynamics of complex formation. Taken together, these observations suggest that phosphate neutralization contributes favorably to $\Delta H_{\text{mod,complex}}^{\circ}$ presumably due to the relaxation of electrostatic repulsion on the concave face of the bent DNA substrate leading to a reduced enthalpic cost for DNA bending by -3.2 kcal/mol (Table 3.6).

Interestingly, $\Delta C_{\text{p}}^{\circ}$ for the formation of the K119A-Rp-P_{Me-2} complex is not significantly different than that observed for K119A-GATATC complex formation (Table 3.4). A reduction in electrostatic strain is expected to contribute to a more negative $\Delta C_{\text{p}}^{\circ}$; however, it may be that neutralization at only one phosphate position does not produce a detectable effect on $\Delta C_{\text{p}}^{\circ}$. Alternatively, a more negative $\Delta C_{\text{p}}^{\circ}$ from reduced electrostatic strain may be opposed by an additional factor that contributes to a more positive $\Delta C_{\text{p}}^{\circ}$, such that there is no *net* difference in $\Delta C_{\text{p}}^{\circ}$.

Table 3.6 Predicted effects on the thermodynamics of complex formation: K119A-Rp-P_{Me-2} vs. K119A-TAG.

Structural differences: K119A-Rp-P _{Me-2} vs. K119A-TAG	Predicted effect on $\Delta\Delta H^{\circ}_{\text{bind}}$	Predicted effect on $\Delta\Delta S^{\circ}_{\text{bind}}$
Reduced interphosphate repulsion	More favorable	Little effect or more unfavorable

3.2.5.2 Molecular factors that contribute to the thermodynamic parameters for K119A-TAG complex formation

Inspection of the top leg of the thermodynamic pseudocycle in Figure 3.1 shows that the unfavorable $\Delta G^{\circ}_{\text{bind}}$ (+5.5 kcal/mol) observed for the formation of the K119A-GATATC mutant complex – relative to the unmodified complex – is due to a combination of both unfavorable $T\Delta S^{\circ}_{\text{bind}}$ (-3.0 kcal/mol) and unfavorable $\Delta H^{\circ}_{\text{bind}}$ (+2.5 kcal/mol). The K119A mutation results in the loss of two interactions, one between the K119-N ϵ and the main-chain carbonyl oxygen of T111, and the other between the K119-N ϵ and the pro-S phosphoryl oxygen at the P₂ phosphate (Figure 3.8, Table 3.7). The loss of these interactions is expected to contribute unfavorable ΔH° to complex formation. MD simulations for the free EcoRV enzyme in the absence of DNA show that the distance between K119:N ϵ and T111:O is not stable over the course of the 6 ns MD simulation suggesting that this interaction is not very stable, and is completely absent after 2.5 ns of simulation time (Figure 3.13; blue plot). In contrast this distance is quite stable at about 2.5 Å in the wt-GATATC complex varying only over short intervals at which it increases to about 5 Å. This observation suggests that the K119A:N ϵ -T111:O interaction is stable in the EcoRV-DNA complex (Figure 3.13; red plot). Interpreting the energetic impact of removing the intraprotein interaction between K119-N ϵ and T111:O on the formation of the K119A mutant complex highlights the importance of considering both the free and bound protein and DNA

molecules. For example, if the K119:N ϵ -T111:O interaction was formed in the free wt EcoRV protein, then removing this interaction via the K119A mutation would contribute equally to $\Delta H^{\circ}_{\text{mod,free}}$ and $\Delta H^{\circ}_{\text{mod,complex}}$ and would not manifest as an observable difference in the overall $\Delta\Delta H^{\circ}_{\text{bind}}$. As it is, the K119:N ϵ -T111:O interaction is much more stable upon DNA binding (Figure 3.13; red plot) relative to that in the free enzyme (Figure 3.13; blue plot), and therefore likely contributes favorably to $\Delta H^{\circ}_{\text{bind,wt}}$. Thus, removal of this interaction upon formation of the K119A mutant complex results in an unfavorable contribution to $\Delta H^{\circ}_{\text{mod,complex}}$.

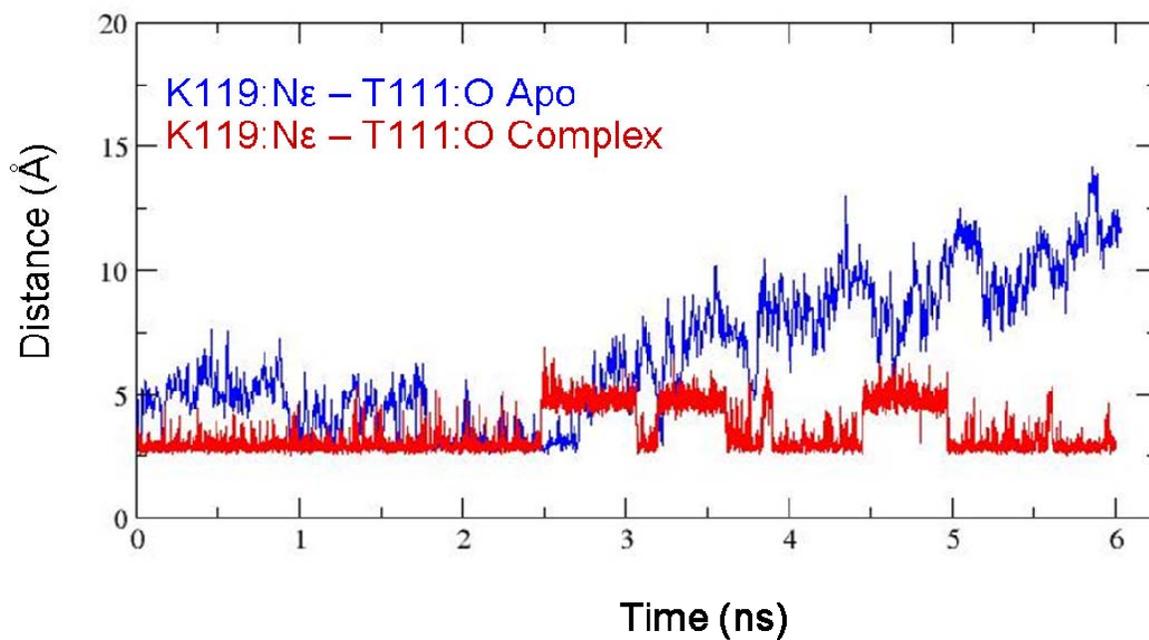


Figure 3.13 K119A:Ne to T111:O distance in the free wt EcoRV and in the wt-GATATC complex
 Plots of the distance (Å) between the K119:Ne and the T111:O is plotted as a function of MD simulation time (ns) for the free enzyme (Apo; blue trace) and for the wt-GATATC complex (red trace).

The unfavorable $\Delta H^{\circ}_{\text{mod,complex}}$ that results from the removal of the interactions involving the K119:N ϵ is somewhat offset by increased solvation of the pro-S phosphoryl oxygen upon formation of this complex relative to formation of the unmodified complex, which is expected to make a favorable contribution to $\Delta H^{\circ}_{\text{mod,complex}}$ (Figure 3.8, Table 3.7). Reduced polar desolvation at the P₂ pro-S phosphoryl oxygen is also expected to contribute unfavorably to $\Delta S^{\circ}_{\text{mod,complex}}$. This unfavorable $\Delta S^{\circ}_{\text{mod,complex}}$ may be partially off-set by a favorable $\Delta S^{\circ}_{\text{mod,complex}}$ from the increased conformational/vibrational flexibility of the T111, which is stabilized by the interaction with K119-N ϵ in the wt-GATATC complex. However, analysis of the atomic fluctuations for T111 in both the wt-GATATC and K119A-GATATC suggests that there is not a significant difference in the dynamics of T111 in these two complexes. Table 3.7 summarizes the structural effects of the K119A-GATATC modification as well as how these effects might influence the thermodynamic parameters observed for the formation of this complex.

Table 3.7 Predicted effects on the thermodynamics of complex formation: K119A-TAG vs. wt-GATATC.

Structural differences: K119A-TAG vs. wt-GATATC	Predicted effect on $\Delta\Delta H^{\circ}_{\text{bind}}$	Predicted effect on $\Delta\Delta S^{\circ}_{\text{bind}}$
Loss of K119:N ϵ -P ₂ pro-S phosphoryl oxygen interaction	More unfavorable	Little effect or more favorable
Loss of K119:N ϵ -T111:O interaction	More unfavorable	Little effect or more favorable
Increased solvation of P ₂ -pro-S phosphoryl oxygen	More favorable	Less favorable

3.2.5.3 Molecular factors that contribute to the thermodynamics of wt-Rp-P_{Me-2} complex formation

The left leg of the thermodynamic pseudocycle in Figure 3.11 represents the thermodynamic consequences of wt-Rp-P_{Me-2} complex formation relative to formation of the unmodified complex. Formation of the wt-Rp-P_{Me-2} complex results in the loss of the interaction between the K119:Nε and the pro-S phosphoryl oxygen at P₋₂, the loss of the interaction between K119:Nε-T111:O buttressing interaction, and the movement of the K119 side chain toward the minor groove of the bound DNA (Figure 3.7B). The Rp-P_{Me-2}-CH₃ group is shielded from solvent by the K119 Cγ and Cδ methylenes; however, these groups are not close enough to the Rp-P_{Me-2}-CH₃ to be considered favorable van der Waals interactions (Figure 3.7B). Unfavorable $\Delta\Delta G_{\text{bind}}^{\circ}$ for the formation of the wt-Rp-P_{Me-2} complex ($\Delta\Delta G_{\text{bind}}^{\circ} = +4.1$ kcal/mol) results from a large, unfavorable $\Delta\Delta H_{\text{bind}}^{\circ}$ ($\Delta\Delta H_{\text{bind}}^{\circ} = +6.7$ kcal/mol), which is significantly off-set by a favorable $T\Delta\Delta S_{\text{bind}}^{\circ}$ ($T\Delta\Delta S_{\text{bind}}^{\circ} = +2.6$). Perturbations of favorable K119 side chain interactions with the P₋₂ phosphate and T111 main chain carbonyl oxygen are likely to contribute to a more unfavorable $\Delta H_{\text{mod,complex}}^{\circ}$ relative to that for the unmodified complex. This unfavorable contribution may be partially compensated for by reduced desolvation of the K119:Nε, which is fully solvated in the wt-Rp-P_{Me-2} complex, but is partially desolvated upon formation of the wt-GATATC complex (Figure 3.7 A and B). Removal of the buttressing interactions to the K119 side chain also results in increased K119 side chain dynamics in the wt-Rp-P_{Me-2} complex relative to the wt-GATATC complex, as measured by its three-dimensional atomic fluctuation (not shown). This increase in dynamics is expected to result in a more favorable $T\Delta\Delta S_{\text{mod,complex}}^{\circ}$ for wt-Rp-P_{Me-2} complex formation relative to wt-GATATC complex formation. Finally, introduction of molecular strain may contribute to the unfavorable $\Delta\Delta H_{\text{mod,complex}}^{\circ}$ observed for formation of the wt-Rp-P_{Me-2} complex. Molecular strain may derive from the requirement of the K119 side chain to adopt a

novel position due to the presence of the Rp-P_{Me-2}-CH₃. While molecular strain can be difficult to identify, a tentative case is made for its contribution to the thermodynamics of wt-Rp-P_{Me-2} complex formation below. Table 3.8 summarizes the predicted effects on the ΔH° and ΔS° for formation of the wt-Rp-P_{Me-2} complex relative to that of the wt-GATATC complex arising from the Rp-P_{Me-2} substitution.

Table 3.8 Predicted effects on the thermodynamics of complex formation: wt-Rp-PMe-2 vs. wt-GATATC.

Structural differences: wt-Rp-P_{Me-2} vs. wt-GATATC	Predicted effect on $\Delta\Delta H^\circ_{\text{bind}}$	Predicted effect on $\Delta\Delta S^\circ_{\text{bind}}$
Loss of K119:N ϵ to P ₂ pro-S phosphoryl oxygen interaction	More unfavorable	Little effect or more favorable
Loss of K119:N ϵ -T111:O interaction	More unfavorable	Little effect of more favorable
Increased solvation of the K119 side chain	More favorable	Little effect or more unfavorable
Increased conformational/vibrational freedom of the K119 side chain	Little effect or more unfavorable	More favorable
Increased molecular strain	More unfavorable	More favorable

Interestingly, ΔC°_p for wt-Rp-P_{Me-2} complex formation is significantly less negative than that for the formation of the unmodified complex (Table 3.4; $\Delta\Delta C^\circ_p = +0.6$ kcal/mol·K). Table 3.1 outlines the factors that influence ΔC°_p . An increase in ΔC°_p may be due to: 1) reduced burial of nonpolar surface area, 2) increased burial of polar surface area, 3) increased molecular strain, 4) increased vibrational and configurational freedom, and/or 5) reduced restriction of interfacial waters in the wt-Rp-P_{Me-2} complex. The following will describe how the structural features that accompany wt-Rp-P_{Me-2} complex formation may influence ΔC°_p .

Solvent accessible surface area (ASA) was calculated for modified and unmodified complexes in order to ascertain whether a modulated burial of molecular surface area could contribute to the less negative ΔC_p^0 observed for wt-Rp-P_{Me-2} complex formation. ASA was calculated as described in Section 3.11 (Equation 3.3). The total surface area buried upon formation of the wt-Rp-P_{Me-2} complex is 152 Å² greater than that buried upon formation of the unmodified complex; however, approximately 330 Å² *more* nonpolar surface area and 150 Å² *less* polar surface area is buried upon formation of the modified complex (Table 3.9). This suggests that differential burial of molecular surface area upon formation of the wt-Rp-P_{Me-2} complex would result in a *more negative* ΔC_p^0 relative to that for the formation of the unmodified complex. Thus, this factor is ruled out as being responsible for the increase in ΔC_p^0 .

It is possible that the perturbation of the interaction between K119:Nε and the P₂ pro-R phosphoryl oxygen, as well as that between K119:Nε and the T111 main chain carbonyl oxygen, may result in an increase in the conformational-vibrational freedom in the wt-Rp-P_{Me-2} complex. This would be expected to contribute to a more positive ΔC_p^0 . As mentioned above, removal of these interactions to the K119 side chain results in increased K119 side chain dynamics in the wt-Rp-P_{Me-2} complex relative to the wt-GATATC complex, as measured by its three-dimensional atomic fluctuation (not shown). However, removal of the K119:Nε-T111:O interaction due to the migration of the K119 side chain away from the P_{Me-2}-CH₃, has little effect on the atomic fluctuation of the T111 side chain, nor does it affect the atomic fluctuations of side chains which interact with either the P₂, or neighboring phosphates. It is not clear to what extent the increase in K119 side chain dynamics would increase ΔC_p^0 observed for wt-Rp-P_{Me-2} complex formation.

An increase in the amount of molecular strain would also result in an increase in ΔC_p^0 . However, there is only an indirect observation that might suggest the existence of such strain in the wt-Rp-P_{Me-2} complex, which is that the K119 side chain must move toward the minor groove of the substrate DNA in order to avoid a steric clash with the Rp-P_{Me-2}-CH₃. Notably, the hypothesis that increased molecular strain at the K119 side chain results in an increase in ΔC_p^0 is supported by the fact that K119A-Rp-P_{Me-2} complex formation, for which the K119 side chain is absent, results in more negative ΔC_p^0 , equivalent to that observed for the formation of the unmodified complex (Table 3.4). This would not be the case if the increase in ΔC_p^0 were due only to increased conformational/vibrational flexibility upon removal of the K119 interactions, as these are also removed in the K119A-Rp-P_{Me-2} complex. An increase in molecular strain in the wt-Rp-P_{Me-2} complex would also contribute unfavorable $\Delta\Delta H_{\text{bind}}^0$ to complex formation relative to unmodified complex formation; which I also observe (Figure 3.11; Table 3.4). Still, the idea that the structural adaptation of the K119 side chain position creates strain in the wt-Rp-P_{Me-2} complex is speculative, and therefore, its invocation as a factor that increases the ΔC_p^0 for the formation of the wt-Rp-P_{Me-2} complex must also be viewed as speculative.

Finally, it has been observed that the T186 and T187 water-mediated interactions to the pro-R phosphoryl oxygen (Figure 3.7A and B) are preserved in the wt-Rp-P_{Me-2} complex. Further, the waters that mediate these interactions are as, if not more, stable in the wt-Rp-P_{Me-2} complex (residence times can range from approximately 1 to 3 ns) than in the unmodified complex (residence times can range from approximately 0.5 to 1.5 ns) suggesting that a reduced degree of immobilization of waters can not account for the increase in ΔC_p^0 observed for the formation of the wt-Rp-P_{Me-2} complex.

Table 3.9 Solvent accessibility in modified and unmodified EcoRV-DNA complexes.

EcoRV Complex	ASA _{Protein} (Å ²)		ASA _{DNA} (Å ²)		ASA _{Complex} (Å ²)		ΔASA (Å ²)		
	NP	P	NP	P	NP	P	Total	Np	P
Wt-GATATC	10901	13634	1385	2572	10232	12315	5955	2054(35%)	3901(65%)
wt-Rp-P _{Me-2}	10901	13634	1497	2472	10015	12382	6107	2383(39%)	3724(61%)
K119A-GATATC	10901	13634	1385	2572	10090	12690	5712	2196(38%)	3516(62%)
K119A -Rp-P _{Me-2}	10901	13634	1497	2472	10771	12582	5151	1627(32%)	3524(68%)

3.2.5.4 Molecular factors that affect the thermodynamics of K119A-Rp-P_{Me-2} complex formation relative to that for wt-Rp-P_{Me-2}

Finally, the bottom leg of the pseudocycle in Figure 3.11 shows the thermodynamic consequences of forming the K119A-Rp-P_{Me-2} complex relative to forming the wt-Rp-P_{Me-2} complex. The relatively small $\Delta\Delta G^{\circ}_{\text{bind}}$ (-0.9 kcal/mol) is the net of a large favorable $\Delta\Delta H^{\circ}_{\text{bind}}$ (-7.4 kcal/mol) and a large unfavorable $T\Delta\Delta S^{\circ}_{\text{bind}}$ (-6.5 kcal/molK). The major structural difference between the wt-Rp-P_{Me-2} and K119A-Rp-P_{Me-2} complexes is the presence of the K119 side chain, which, as stated previously, moves toward the minor groove while shielding the Rp-P_{Me-2}-CH₃ from solvent. These dramatic thermodynamic differences may derive from an aggregate of factors, each of which contributes to a varying degree.

One factor that might result in a more favorable $\Delta\Delta H^{\circ}_{\text{bind}}$ for K119A-Rp-P_{Me-2} complex formation – relative to that for the wt-Rp-P_{Me-2} complex – is the retention of the shell-like structure of water surrounding the Rp-P_{Me-2}-CH₃ in the K119A-Rp-P_{Me-2} complex that is released into the bulk solvent upon formation of the wt-Rp-P_{Me-2} complex (Figure 3.7B and 3.9A). Water release from the Rp-P_{Me-2}-CH₃ upon wt-Rp-P_{Me-2} complex formation results in the disruption of dipole-induced dipole interactions between water and the Rp-P_{Me-2}-CH₃, as well as the disruption

of hydrogen bonds between the waters themselves. It should be noted that disruption of these interactions might result in an unfavorable $\Delta\Delta H_{\text{bind}}^{\circ}$ only if they are not completely compensated for by favorable van der Waals interactions with the methylene groups of the K119 side chain, which it appears, remain only slightly farther away from the Rp-PMe-2-CH₃ (4.2 and 4.6 Å) than the distance required for a favorable van der Waals interaction (~3.8 Å). Thus, it is expected that if that retention of this shell-like structure of water does influence $\Delta\Delta H_{\text{bind}}^{\circ}$ for K119A-Rp-P_{Me-2} complex formation relative to that for wt-Rp-P_{Me-2} complex formation, then this effect is probably small in magnitude and, can not account for the *dramatic* favorable $\Delta\Delta H_{\text{bind}}^{\circ}$ that is observed. On the other hand, the retention of structured water at the Rp-P_{Me-2}-CH₃ may contribute to the unfavorable $T\Delta\Delta S_{\text{bind}}^{\circ}$ associated with K119A-Rp-P_{Me-2} complex formation relative to that for the formation of the wt-Rp-P_{Me-2} complex.

A second factor that could contribute to the more favorable $\Delta H_{\text{bind}}^{\circ}$ observed for K119A-Rp-P_{Me-2} complex formation relative to that for the formation of the wt-Rp-P_{Me-2} complex is the removal of molecular strain that is potentially present in the wt-Rp-P_{Me-2}. As mentioned above, upon wt-Rp-P_{Me-2} complex formation, the K119 side chain moves away from the Rp-P_{Me-2}-CH₃ in order to avoid a steric clash. I have proposed that this movement may be associated with increased in molecular strain in this complex resulting in an increase in $\Delta C_{\text{p}}^{\circ}$ for the formation of the wt-Rp-P_{Me-2} complex relative to that for the formation of the unmodified complex ($\Delta C_{\text{p}}^{\circ} = -0.6$ and -1.2 kcal/molK, respectively). Removal of the K119 side chain “rescues” this increase in $\Delta C_{\text{p}}^{\circ}$ as the formation of the K119A-Rp-P_{Me-2} complex shows a $\Delta C_{\text{p}}^{\circ}$ of -1.2 kcal/molK, possibly due to the removal molecular strain in this complex. Similarly, increased molecular strain is expected to result in a more unfavorable $\Delta H_{\text{bind}}^{\circ}$ for wt-Rp-P_{Me-2} complex formation. Thus, removal of the K119 side chain would relieve this strain, resulting in a more favorable

$\Delta\Delta H^{\circ}_{\text{mod,complex}}$ The observations that there is an unfavorable $\Delta\Delta H^{\circ}_{\text{bind}}$ for the left leg of the cycle in Figure 3.11 and a favorable $\Delta\Delta H^{\circ}_{\text{bind}}$ for the bottom leg, and that the magnitudes of the $\Delta\Delta H^{\circ}_{\text{bind}}$ for these two legs are comparable, support a role for reduced molecular strain in the thermodynamics of K119A-Rp-P_{Me-2} complex formation relative to that for wt-Rp-P_{Me-2} complex formation.

Table 3.10 Predicted effects on the thermodynamics of complex formation: wt-Rp-P_{Me-2} vs. K119A- Rp-P_{Me-2}.

Structural differences: wt-Rp-P_{Me-2} vs. K119A- Rp-P_{Me-2}	Predicted effect on $\Delta\Delta H^{\circ}_{\text{bind}}$	Predicted effect on $\Delta\Delta S^{\circ}_{\text{bind}}$
Reduced molecular strain	More favorable	More unfavorable
Increased solvent interactions to Rp-P _{Me-2} -CH ₃ ^a	More favorable or little effect	More unfavorable

^aThis would only contribute unfavorably to $\Delta\Delta H^{\circ}_{\text{bind}}$ if these interactions were not compensated for by K119 side chain interactions to Rp-P_{Me-2}-CH₃.

Thus, insight into the molecular origins of the thermodynamic consequences of each of the electrostatic modifications can be gained by simulation and analysis of structural differences between unmodified and modified complexes. For the sake of simplicity, only half-site A was considered here; however, a similar analysis can be done for half-site B of these complexes, likely yielding similar insights. The combinatorial approach used to dissect out the role of phosphate neutralization in EcoRV-DNA complex formation has been shown to be particularly effective by this analysis, as phosphate neutralization is the only perceived molecular factor that contributes to the thermodynamic differences observed for K119A-Rp-P_{Me-2}-CH₃ binding relative to K119A-GATATC binding (Table 3.6). These data suggest that phosphate neutralization contributes favorable $\Delta H^{\circ}_{\text{bind}}$ to EcoRV-DNA complex formation by reducing electrostatic

repulsive forces in the bent DNA conformation, thus reducing the energetic cost of DNA bending.

3.2.6 Phosphate neutralization by Rp-P_{Me-2} substitution impacts the electrostatic potential at the EcoRV-DNA interface

I have presented evidence that P_{Me}-mediated phosphate neutralization at the GpATATC position produces favorable energetic effects, presumably because relaxation of interphosphate repulsion reduces the cost of DNA bending. Thus, it is expected that the Rp-P_{Me-2} substitution will affect the electrostatic potential at the EcoRV-DNA interface. In order to understand the impact of both the K119A mutation and Rp-P_{Me-2} substitutions on the electrostatic potential at the EcoRV-DNA interface, the Delphi v.4 program (Rocchia et al., 2002) was used to calculate these electrostatic potentials for the four MD simulated EcoRV-DNA complexes mentioned above. Delphi solves the Poisson-Boltzmann (PB) equation by finite difference methods. The PB equation can be written as,

$$f(\phi) = -\epsilon\kappa^2 \sinh(\phi(r)) + \frac{4\pi e\rho^f(r)}{kT}, \quad (3.19)$$

to give the electrostatic potential (ϕ) over all three-dimensional space, where ϵ is the dielectric constant, κ is the Debye-Huckel parameter, ρ is the interior charge density, k is Boltzmann's constant and T is temperature. Essentially, the molecular components of the EcoRV-DNA complex were mapped onto a three-dimensional grid, and the parameters ϵ and ρ were defined at each grid point (Figure 3.14). The number of grid points per Å is directly correlated with the accuracy of the calculation, and an acceptable value of two grid points per Å was used in these

calculations. Representative structural snapshots near the end of the 6 ns MD simulations were used for these calculations, for which the atomic parameters (van der Waals radius and charge) correspond to those used in the MD simulations as defined by the Amber parameter set. Waters were removed from the complexes, while the explicit salt ions were retained. The following will describe the results of these calculations and will address the electrostatic effects of the K119A and Rp-P_{Me-2} modifications.

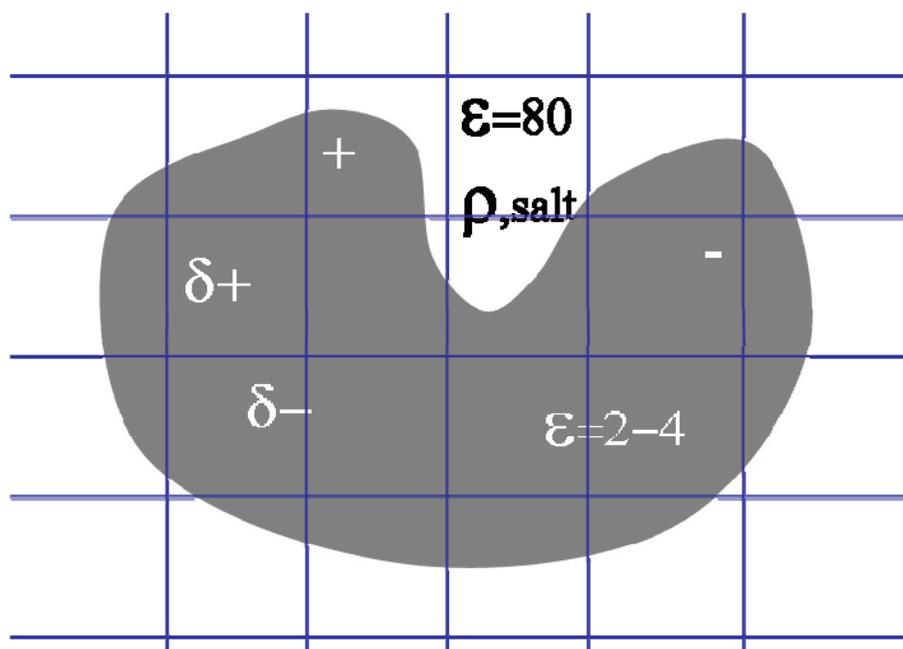


Figure 3.14 Grid-mapping of the EcoRV-DNA complex by Delphi.

Graphical representation of the grid definitions used by Delphi to calculate the finite-difference Poisson-Boltzmann equations is shown. The macromolecule (grey) is mapped onto a grid (blue lines). The interior and exterior dielectric constants (ϵ), the salt concentration of the solvent, and the interior charge density (ρ) are defined as are the charge and size of each atom in the macromolecule.

Electrostatic potentials (EP) have been calculated by Delphi and have been mapped onto the molecular surface of the DNA molecule for the modified and unmodified EcoRV-DNA complexes (Figure 3.15). These calculations suggest that, in the unmodified complex, the EP of the P₋₂ phosphate is partially positive (blue) (Figure 3.15A and B). This is likely due to the presence of the K119 side chain, which forms a stable interaction with this phosphate. Mutation of K119 to alanine results in a significantly more negative (red) EP at the P₋₂ phosphate, which is consistent with the effects expected for the removal of a positively charged side chain (Figure 3.15A and B), suggesting that there is increased interphosphate repulsion in this mutant complex. Interestingly, the electrostatic profile of phosphates neighboring the P₋₂ phosphate is not significantly affected by the K119A mutation. In the K119A complex, the electrostatic potential at the P₋₂ position is less negative upon P_{Me-2} substitution, suggesting that the Rp-P_{Me-2} reduces interphosphate repulsion at this position (Figure 3.15D). Finally, the Rp-P_{Me-2} substitution in the wt complex results in an electrostatic potential at P₋₂ that likely reflects neutralization of the phosphate by the P_{Me} substitution and the redistribution of the positively charged lysine side chain (Figure 3.15C). Overall these data suggest that the P_{Me-2} substitution has a definite impact on the electrostatic potential at the EcoRV-DNA interface, supporting the hypothesis that this modification can reduce the energetic cost of DNA bending in the K119A complex.

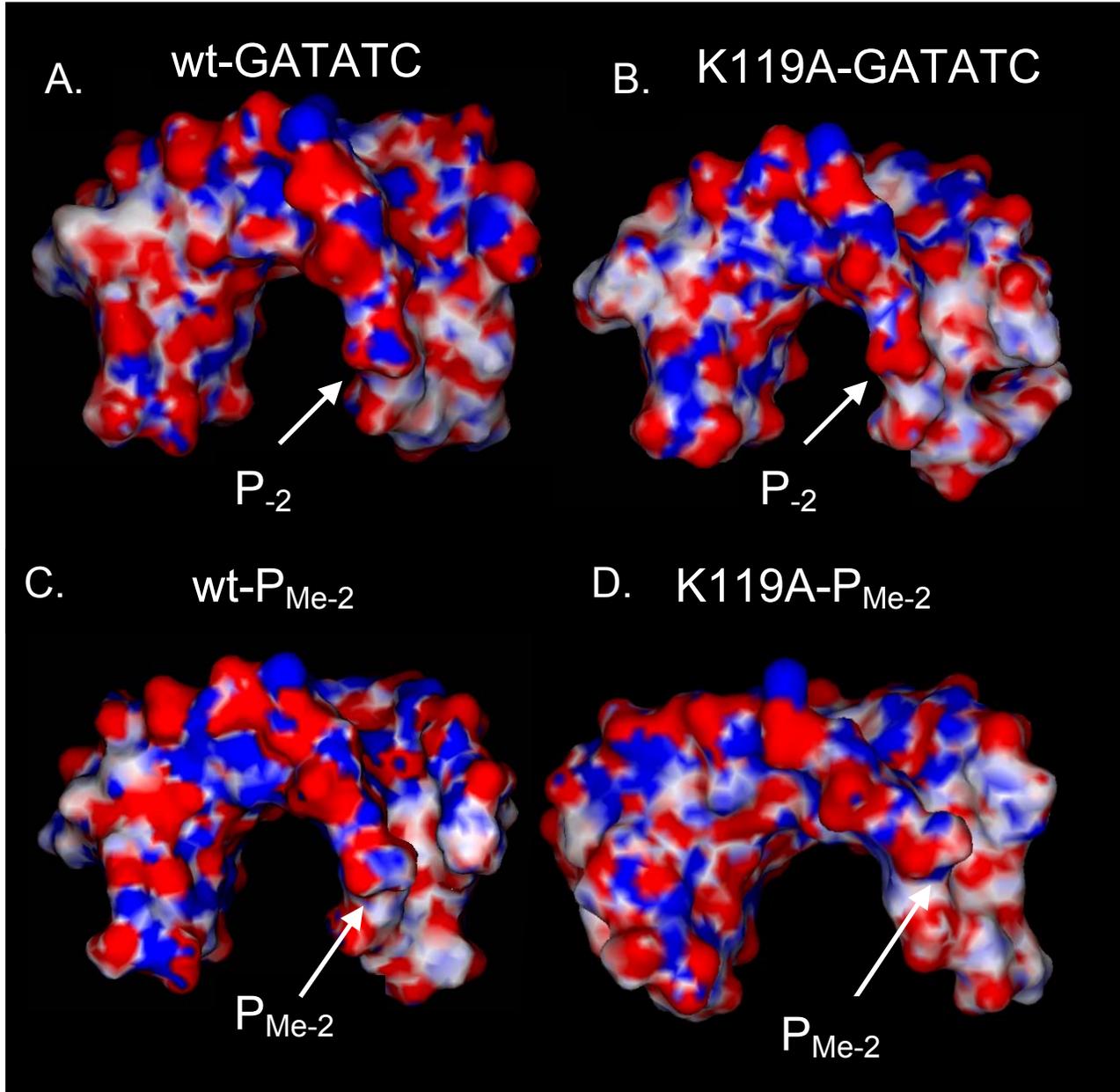


Figure 3.15 Electrostatic potentials for modified and unmodified EcoRV-DNA complexes. Electrostatic potentials (blue is positive; red is negative) have been mapped onto surface representations of the DNA molecules from the modified and unmodified EcoRV-DNA complexes. The protein has been removed for clarity.

3.3 CONCLUSION AND DISCUSSION

I have measured the effects of phosphate neutralization on the thermodynamic parameters ΔH° , ΔS° , and ΔC_p° . These data show that the formation of the EcoRV-DNA complex is characterized by a large negative ΔC_p° , a thermodynamic signature of specific protein-DNA interactions. Further, an unfavorable ΔH° for EcoRV-DNA complex formation is compensated for by favorable ΔS° at 25°C. An unfavorable ΔH° is characteristic of specific protein-DNA interactions in which DNA is dramatically distorted/bent which, is thought to derive from the large enthalpic cost for DNA distortion/bending (Jen-Jacobson et al., 200b). Interestingly, I find that phosphate neutralization on the concave face of the substrate DNA contributes favorable ΔH° to EcoRV-DNA complex formation, thus I propose that phosphate neutralization reduces the enthalpic cost of DNA bending due to relaxation of interphosphate repulsive forces on the concave face of the DNA bend.

In addition, using molecular dynamics simulations as a probe, I have investigated the structural effects of particular modifications to the EcoRV-DNA DNA interface. These effects have provided insight into the molecular factors that influence the thermodynamic parameters that were measured for these modified complexes. In general thermodynamic differences in modified complexes derive from altered interactions involving protein, DNA and solvent molecules, and from an altered electrostatic environment. Importantly these data show that the favorable contribution to ΔH° observed for the Rp-P_{Me-2} substitution does not derive from the introduction of a new protein-DNA interaction, but is due to phosphate neutralization, in support of the proposed role for this modification in reducing the cost of DNA bending.

3.3.1 On the magnitude of the energetic effects of phosphate neutralization relative to the total energetic requirement for DNA bending

I showed in Chapter 2 that phosphate neutralization via the P_{Me-2} substitution results in a favorable contribution to $\Delta G^{\circ}_{\text{bind}}$ for the formation of K119A-Rp-P_{Me-2} relative to K119A-GATATC complex formation, and that the magnitude of this favorable contribution is equal to – 2.3 kcal/mol. I attributed this favorable effect to the relief of electrostatic repulsive forces on the concave face of the DNA substrate upon EcoRV-induced DNA bending. How does the magnitude of the favorable energetic effects of phosphate neutralization relate to magnitude of the energetic requirement to achieve the EcoRV-induced DNA bend? To put this magnitude into perspective at the free energy (ΔG) level, I will consider a simplified relationship that estimates the free energy required for EcoRV-induced DNA bending as follows,

$$\Delta G_{\text{bend}} = 0.0135 \frac{(\Delta\Theta_{\text{deg}})^2}{L_{\text{bp}}} (\text{kcal} / \text{mol}), \quad (3.20)$$

where the DNA bending free energy ΔG_{bend} is expressed in kcal/mol assuming a DNA persistence length of 150bp (the persistence length at $[M^+] > 0.02 \text{ M}$), and $\Delta\Theta_{\text{deg}}$ is the degree of DNA bending over a contour length equal to L_{bp} . Applying this equation to the DNA bend induced by EcoRV suggests that the cost of DNA bending is approximately +20.4 kcal/mol assuming that $\Delta\Theta_{\text{deg}} = 55^\circ$ and that $L_{\text{bp}} = 2$ base pairs. If one considers that a range of DNA bend angles is observed in various X-ray crystal models (42° to 60°) (Horton and Perona, 2000), then the cost of EcoRV-induced DNA bending would range from +11.9 to 24.3 kcal/mol. These estimates assume that all of the EcoRV-induced DNA bend occurs at the central GATATC base

step, which displays a dramatic positive role of approximately 50° into the major groove (Winkler, 1993; Horton and Perona, 2000)(See Section 1.7; Figure 1.4). Use of this model to estimate ΔG_{bend} suggests that phosphate neutralization at P₋₂ to can reduce the cost of DNA bending by approximately 10 to 20 percent of the total free energy required for EcoRV-induced bending of the specific DNA substrate. It should be noted that nature has designed EcoRV to neutralize phosphates at multiple positions on the concave face of the bound DNA (e.g. P₋₅ and perhaps P₋₃; see Figure 3.4 B and D) thus it is an aggregate of this neutralization that must be considered when assessing the overall role of phosphate neutralization in reducing the cost of DNA bending. If the effects of phosphate neutralization at *both* P₋₂ ($\Delta\Delta G_{\text{bind}}^0 = -2.3$ kcal/mol; Table 2.4) and P₋₅ ($\Delta\Delta G_{\text{bind}}^0 = -0.9$ kcal/mol) are considered, the cumulative benefit for phosphate neutralization is approximately -3.2 kcal/mol, suggesting that phosphate neutralization at these two phosphates can reduce the cost of DNA bending by approximately 15 to 25 percent. Due to confounding energetic factors, it is not possible to deconvolute the effects of phosphate neutralization by the R221 side chain, or at additional phosphate positions however, neutralization by R221 and at these additional positions almost certainly contributes to the overall reduction in the cost of DNA bending.

While the x-ray crystal structure suggests a model for EcoRV-induced DNA bending in which bending is facilitated at the central TA base pair step, two additional models for DNA bending can be envisioned. First DNA bending might be distributed over the entire length of the protein-DNA interface which contains approximately 12 base pairs. In this case $\Delta\Theta_{\text{deg}}$ would equal 42° to 60° , while L_{bp} would equal 12bp, and ΔG_{bend} would equal $+2.0$ to $+4.1$ kcal/mol suggesting that phosphate neutralization can contribute as much as 100 percent of the total cost of EcoRV induced DNA bending assuming this model for DNA bending. This model likely

represents a lower limit for the energetic cost of EcoRV-induced DNA bending ($\Delta G_{\text{bend}}^{\circ}$) therefore phosphate neutralization likely does not provide all of the free energy required for DNA bending. If one were to assume a model for DNA bending in which EcoRV-induced DNA bend was facilitated by a positive roll at two distinct positions in the DNA substrate, or if the DNA bend were distributed evenly over a shorter length of the substrate then I would estimate an intermediate ΔG_{bend} with values that are in between +24.3 and +2.0 kcal/mol. Thus, it is apparent that independent of the model chosen for EcoRV-induced DNA bending, phosphate neutralization can contribute significantly to the reduction of the energy required to achieve this bend.

In addition, I have shown in this Chapter that the favorable effects of phosphate neutralization at the P₋₂ position derive from a favorable *enthalpic* contribution to EcoRV-DNA complex formation. Here we distinguish the effects of phosphate neutralization on the $\Delta G_{\text{bind}}^{\circ}$, which is an aggregate of the entropic and enthalpic effects of phosphate neutralization (recall that $\Delta G_{\text{bind}}^{\circ} = \Delta H_{\text{bind}}^{\circ} - T\Delta S_{\text{bind}}^{\circ}$) from its effects on ΔH° which is a more direct measure of the energy required for DNA distortion. As mentioned above, data in this Chapter suggest that the phosphate neutralization at position P₋₂ reduces the enthalpic cost of DNA bending -3.2 kcal/mol. Again, EcoRV-induced DNA bending is facilitated through a 50° roll at the center TA base step of the GATATC resulting in the destacking of these base pairs (Winkler, 1993; Horton and Perona, 2000). It has been estimated that the enthalpic cost of destacking of a single base pair step is approximately +7 to +11 kcal/mol (Searle and Williams, 1993; Jen-Jacobson, 2000b) with the TA base pair step being the easiest to destack (El Hassan and Calladine, 1997; Olson et al., 1998). This suggests that the favorable enthalpy contributed by phosphate neutralization at P₋₂ can account for almost half of the enthalpy required for TA base pair destacking.

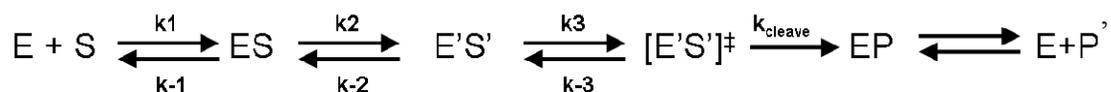
How does the magnitude of the favorable enthalpy contributed by phosphate neutralization at P₋₂ relate to the total energy required achieve the EcoRV-induced DNA bend conformation? Work has been done using internal coordinate energy optimization to estimate the energy required to achieve a variety of protein induced DNA conformations (Zakrzewska, 2003). In this work the energy required for DNA deformation (E_{def}) was examined for 71 different protein-DNA complexes. Essentially, the energy required for E_{def} was equal to the difference between the total optimized energies (E) of the DNA molecule isolated from the protein-DNA complex and the energy optimized free DNA molecule. Factors that were considered to contribute to E included the bonded factors (E_{bond}) such as torsional energies, as well as non bonded factors ($E_{\text{non-bonded}}$) such as Leonard-Jones and electrostatic energies (Zakrzewska, 2003). E_{def} for EcoRV-induced DNA bending was calculated to equal 44.6 kcal/mol (Zakrzewska, 2003), suggesting that phosphate neutralization at position P₋₂ may contribute approximately 7 percent of the total energy required to achieve the EcoRV-induced DNA bend. Phosphate neutralization at additional phosphate positions along the concave face of the bent DNA likely further reduces the cost for DNA bending by EcoRV while additional sources of favorable free energy are likely required to stabilize DNA bending by EcoRV, including attractive protein-base, and protein-phosphate interactions.

4.0 THE EFFECTS OF PHOSPHATE NEUTRALIZATION ON ECORV-DNA TRANSITION STATE FORMATION

I established in Chapter 2 that phosphate neutralization on the concave face of the bent DNA substrate improves EcoRV-DNA binding affinity by reducing the energetic cost of DNA bending. This chapter addresses how modulation of the electrostatic landscape at the EcoRV-DNA interface affects the progression of the EcoRV-DNA ground state complex to the transition state, in which DNA bending is a prerequisite for the correct positioning of the catalytic elements

4.1 INTRODUCTION

EcoRV catalyzes the hydrolysis of its specific DNA site, GATATC. This reaction is proposed to proceed as described in the scheme



where E and S represent the free protein and DNA molecules, ES represents the non-specific EcoRV-DNA encounter complex (See Chapter 2), E'S' represents the specific recognition complex, and $[E'S']^\ddagger$ represents the transition state complex. By coupling correct formation of

the recognition interface to correct assembly of the active site, EcoRV achieves extremely high site selectivity (Winkler, 1993; Perona and Martin, 1997).

Previous work in the Jen-Jacobson lab has shown that a variety of structural perturbations can alter the binding free energy for specific complex formation (i.e. $\Delta\Delta G^{\circ}_{\text{bind}} > 0$) and yet they do not affect the energy required to achieve the transition state (i.e. $\Delta\Delta G^{\circ\ddagger} = 0$) (Figure 4.1 panel C). These observations suggest that certain energetic factors contribute equally to the formation of both the specific recognition complex (E'S') and the transition state complex (ES[‡]). This leads to the proposal that the specific recognition complex is a “pre-transition state” complex that closely resembles the transition state complex, such that binding free energy is efficiently utilized in achieving the transition state (Jen-Jacobson, 1997). This concept is supported by the observation that catalysis can be achieved by soaking divalent metals into crystals containing metal free EcoRV-DNA complexes (Perona and Martin, 1997). Further, crystallographic analysis of EcoRV-DNA complexes in both the presence and absence of divalent metal shows that the overall structure of these complexes is superimposable to within 0.4 Å, and that both active sites are similarly assembled with only a minor displacement of the scissile phosphate in a + Mg²⁺ structure (Kostrewa and Winkler, 1995).

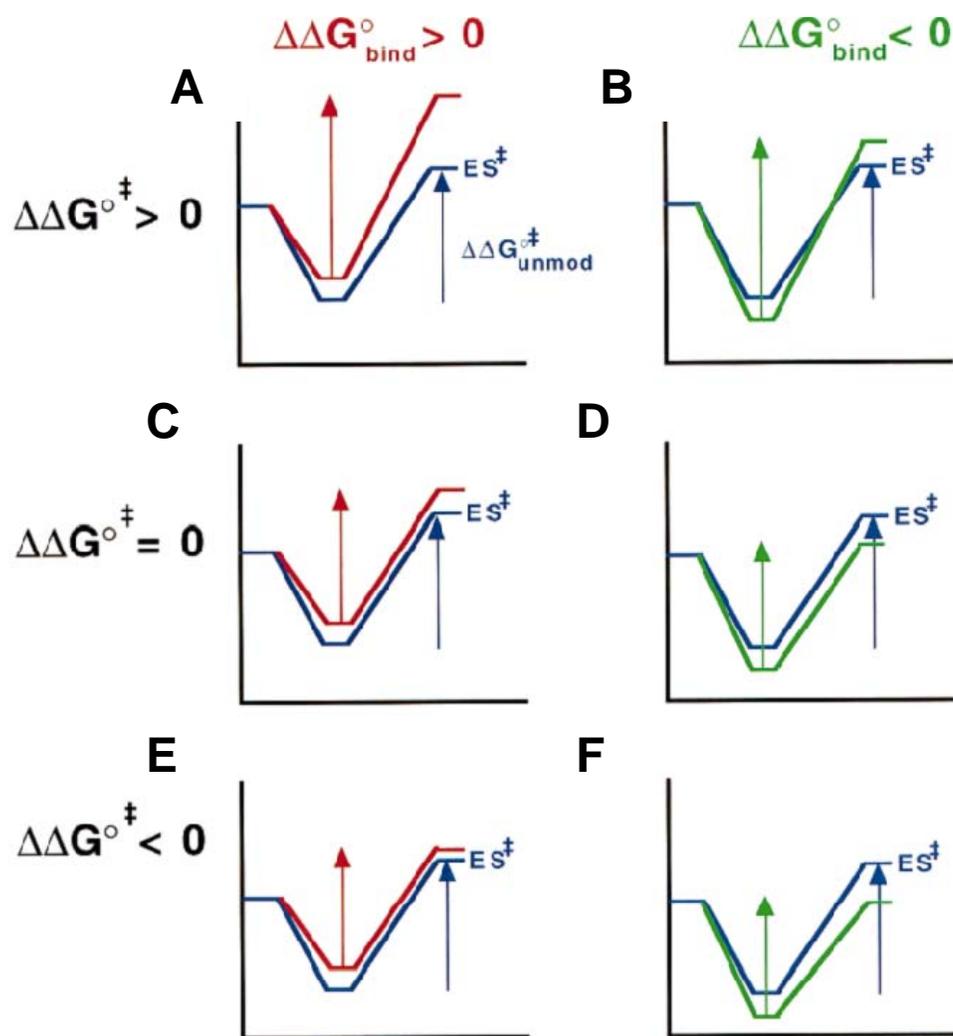


Figure 4.1 Reaction coordinate diagrams for the progression from free substrate to the transition state.

Each progress curve shows only one enzyme–substrate complex (e.g. E'S') at its free energy minimum; any binding intermediates (e.g., ES preceding E'S') are omitted for clarity. In each panel, the reaction profile of the unmodified substrate is in blue, and standard free energy of activation, (ΔG^{\ddagger}) for the unmodified substrate is indicated by the blue arrow. Reaction profiles are in red for modifications that inhibit binding ($\Delta\Delta G^{\circ}_{\text{bind}} > 0$) and in green for modifications that improve binding ($\Delta\Delta G^{\circ}_{\text{bind}} < 0$). Activation free energies for modified substrates are indicated by corresponding red and green arrows. Figure was taken directly from (Jen-Jacobson, 1997).

Interestingly, the precise spatial arrangement of correctly assembled catalytic elements depends on DNA bending (Winkler, 1993; Perona and Martin, 1997). The phosphate backbone in the bent DNA of the specific EcoRV complex is displaced by approximately 3 Å relative to the unbent DNA in the non-specific complex, and thus precisely apposes the scissile phosphate to two key catalytic side chains, D74 and D90 (Figure 4.2) (Winkler, 1993). These residues, along with the scissile phosphate, coordinate a Mg^{2+} ion that is required for catalysis (Figure 4.2A and C). However, while the scissile phosphate is missing in the non-specific complex, it is readily apparent that D74 and D90 are much farther from the DNA than in the specific complex. For example, in the non-specific complex, D74:C γ and D90:C γ are 14.4 Å and 13.1 Å from the oxygen atom 5' to the missing scissile phosphate are, respectively, while in the specific complex, D74:C γ and D90:C γ are only 7.2 and 6.2 Å from this O5'. Thus without DNA bending, neither coordination of the Mg^{2+} cofactor nor catalysis would occur. A link between DNA bending and DNA catalysis is also provided by the observation that the magnitude of DNA bending in four different crystal structures of the specific EcoRV-DNA ground state complex varies by approximately 50%, from 42° to 60°. Interestingly, experiments in which Mg^{2+} is soaked into these crystals suggests that only the crystal form that displays the largest magnitude of DNA bending retains catalytic activity (Winkler, 1993; Kostrewa and Winkler, 1995; Perona and Martin, 1997).

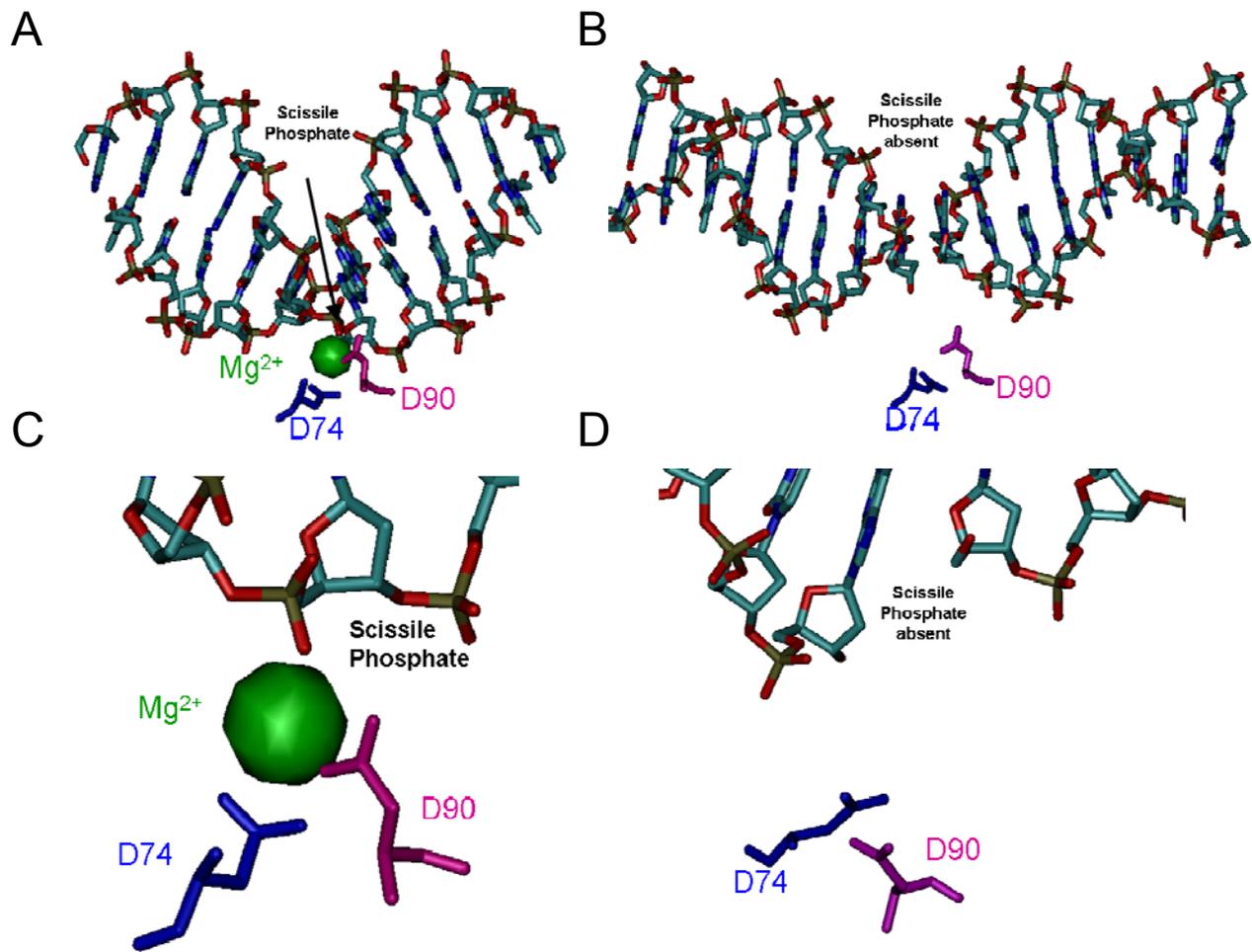


Figure 4.2 D74 and D90 positions in the specific and non-specific EcoRV-DNA complexes.

Active site residues, D74 (blue) and D90 (pink), are shown relative to the scissile phosphate in the specific EcoRV-DNA complex (A and C; coordinates from 1RVB), and relative to the DNA phosphate backbone in the non-specific EcoRV-DNA complex (B and D; coordinates from 2RVE). An active site Mg²⁺ (green) is coordinated by D74, D90, and the scissile phosphate in the specific complex. The non-specific complex was solved in the absence of Mg²⁺ and is missing the scissile phosphate.

4.2 RESULTS AND DISCUSSION

4.2.1 Measuring the effects of electrostatic perturbation on the formation of the EcoRV-DNA transition state complex

First order cleavage rate constants were measured for modified EcoRV-DNA complexes in which the electrostatic landscape at the interface had been modulated. (See Methods, Section 6.6 for details of single turnover cleavage assays). EcoRV catalyzes the hydrolysis of each strand of the DNA backbone in two distinct half-sites. The scheme in Figure 4.3 depicts parallel-sequential cleavage of the EcoRV GATATC site (embedded in 24 bp oligonucleotide); k_1 and k_2 represent rate constants for cleavage of the top and bottom strands of the DNA substrate, respectively, while k_3 and k_4 represent rate constants for cleavage of the nicked intermediates. Symmetrical modifications in the EcoRV-DNA complex produce equal rates in both active sites; that is, $k_1 = k_2$. The first order cleavage rate constant k_1 , is used to define the activation free energy $\Delta G^{0\ddagger}$ for formation of the EcoRV-DNA transition state complex by the expression

$$\Delta G^{0\ddagger} = -RT \ln k_1. \quad (4.1)$$

The difference in activation free energy $\Delta\Delta G^{0\ddagger}$ between modified and unmodified EcoRV-DNA complexes is given by

$$\Delta\Delta G^{0\ddagger} = -RT \ln(k_{\text{cleave}}^{\text{mutatnt}}/k_{\text{cleave}}^{\text{wild-type}}). \quad (4.2)$$

As was the case for measuring the effects of phosphate neutralization on EcoRV-DNA ground state formation, EcoRV mutants were used to isolate the effect of phosphate neutralization on transition state stabilization from other consequences of P_{Me} substitutions.

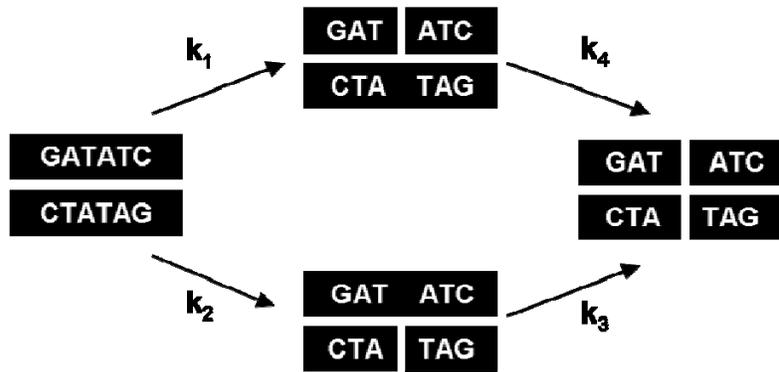


Figure 4.3 Parallel sequential model for EcoRV-DNA cleavage.

Parallel-sequential pathway for double-strand cleavage of GATATC by EcoRV. Rate constants k_1 , k_2 and k_3 , k_4 can be distinguished experimentally using eccentric site location in oligonucleotides (Lesser et al., 1990) (Section 6.6). Figure adapted from Jen-Jacobson (1997).

4.2.2 Basic side chain residues stabilize the transition state complex

First order cleavage rate constants are dramatically reduced for K119A, R226A, and R140A mutants with unmodified DNA substrates ($k_1 = 0.06 \text{ s}^{-1}$ to 0.37 s^{-1} ; Table 4.1) relative to those for the wt enzyme ($k_1 = 1.8 \text{ s}^{-1}$). In addition, there is a large penalty in binding for these mutants ($\Delta\Delta G_{\text{bind}}^{\circ} = +4.2$ to $+5.5 \text{ kcal/mol}$; Table 2.1). Thus, progression from the free protein and DNA molecules to the transition state corresponds to the reaction coordinate diagram in Figure 1.4A. Equilibrium FRET data suggest that these mutant complexes are able to achieve the same degree of DNA bending that is observed in the unmodified EcoRV-DNA ground state complex (Table 2.7), suggesting that the penalties in cleavage rate result from an additional energetic cost of going from the specific ground state complex to the transition state. As noted above, it has been previously inferred (Horton and Perona, 2000) that only fully bent DNA permits progression to the transition state complex. It is possible that the K119A and R226A mutants produce increased electrostatic repulsion on the concave face of the DNA molecule, thus resulting in an increased cost of achieving the transition state complex in which the DNA is fully bent. In this case, it is expected that phosphate neutralization by P_{Me} substitution will reduce interphosphate repulsion on the concave face of the DNA, which will favor the formation transition state complex. Section 4.2.4 will address the effects of P_{Me} substitutions on cleavage rate in the context of the EcoRV basic residue mutants. The R140 side chain contacts the P_{-3} DNA phosphate, and makes an intersubunit interaction with the side chain of Q69. Horton and Perona (2000) have shown the EcoRV intersubunit orientation varies with increasing DNA bending, suggesting that the R140-Q69 interaction may stabilize the transition state by stabilizing the intersubunit orientation in the fully bent complex which is required for

progression to the transition state. Thus, the R140A mutant may destabilize this interaction and result in a reduced cleavage rate.

There is only a small penalty for the R221A mutation in cleavage relative to that for the wt enzyme ($k_1 = 1.4 \text{ s}^{-1}$; Table 4.1), while there is a significant penalty in binding ($\Delta\Delta G_{\text{bind}}^{\circ} = +3.1 \text{ kcal/mol}$; Table 2.1). This circumstance corresponds to the reaction coordinate profile shown in Figure 4.1C, and suggests that once the penalty for binding is paid in the ground state, it does not to be paid again in the transition state. Thus, the energetic contribution of the R221 side chain in ground state stabilization is conserved in the EcoRV-DNA transition state as there is little additional energetic effect for the R221A mutation in cleavage.

Table 4.1 Effects of basic side chain mutations on EcoRV-DNA transition state stabilization.

Enzyme/Sequence^a	$k_{1,\text{cleave}} \text{ (s}^{-1}\text{)}^{\text{b}}$	$\Delta\Delta G_1^{\circ \ddagger}$ (kcal/mol)^c	$k_{2,\text{cleave}} \text{ (s}^{-1}\text{)}$	$\Delta\Delta G_2^{\circ \ddagger}$ (kcal/mol)
wt-GATATC	1.8(±0.1)	0	1.8(±0.1)	0
CONCAVE				
K119A + GATATC	0.13(±0.01)	+1.6(±0.1)	0.11(±0.02)	+1.7(±0.1)
R221A + GATATC	1.4(±0.3)	+0.1(±0.2)	1.3(±0.3)	+0.1(±0.2)
R226A + GATATC	0.06(±0.01)	+2.0(±0.1)	0.06(±0.01)	+2.0(±0.1)
CONVEX				
R140A + GATATC	0.37(±0.1)	+0.9(±0.2)	0.32(0.1)	+1.0(±0.2)

^aFirst order cleavage rate constants were measured for the oligonucleotide d(TGTGTTGTAGGATATCCTACAGGT) in 0.1 M KCl, pH 7.25 and 10 mM MgCl₂.
^b $k_{1,\text{cleave}}$ and $k_{2,\text{cleave}}$ denote first order cleavage rate constants for the top and bottom strand, respectively.
^c $\Delta\Delta G_1^{\circ \ddagger} = -RT\ln(k_{1,\text{modified}}/k_{1,\text{unmodified}})$

4.2.3 Effects of P_{Me} substitutions on wt EcoRV-DNA transition state formation

Individual P_{Me} substitutions exact a variety of effects on wild-type cleavage rates (Table 4.3). The P_{Me-5} substitution shows only a small effect on cleavage rate and little effect on binding ($\Delta\Delta G^{\ddagger} = +0.3$ kcal/mol; $\Delta\Delta G^{\circ}_{\text{bind}} = 0.0 (\pm 0.1)$ kcal/mol Tables 4.2 and 2.3). The P_{Me-3} substitution shows only a small penalty for cleavage, but a significant penalty in binding ($\Delta\Delta G^{\ddagger} = +0.2 \pm (0.1)$ kcal/mol, $\Delta\Delta G^{\circ}_{\text{bind}} = +1.5 (\pm 0.1)$ kcal/mol; Tables 4.2 and 2.3). This scenario corresponds to the reaction coordinate diagram in Figure 4.1 C. Finally, there is a penalty in both binding and cleavage for the Rp-P_{Me-2}, P_{Me-2} and P_{Me-4} substitutions (Tables 4.2 and 2.3) corresponding to the reaction coordinate diagram in Figure 4.1 panel A. These effects suggest that interactions to these phosphates may help to stabilize the transition state in addition to stabilizing the ground state complex. The data for multiple P_{Me} substitutions in Table 4.2 will be discussed in Section 4.2.5.

Table 4.2 Effects of P_{Me} substitutions on wt EcoRV cleavage.

Complex	k ₁ (s ⁻¹) ^a	ΔΔG ^{0‡} _{observed} (kcal/mol) ^b	ΔΔG ^{0‡} _{predicted} (kcal/mol) ^c
wt +GATATC	1.8(±0.1)	0	-
wt + P _{Me-5}	1.1(±0.08)	+0.3(±0.1)	-
wt + P _{Me-4}	0.26(±0.04)	+1.1(±0.1)	-
wt + P _{Me-3}	1.3 (±0.2)	+0.2(±0.1)	-
wt + P _{Me-2}	0.05(±0.004)	+2.1(±0.1)	-
wt + Rp-P _{Me-2}	0.08(±0.02)	+1.8 (±0.1)	-
wt + P _{Me-2} /P _{Me-3}	0.02(±0.007)	+2.7(±0.1)	+2.3(±0.1)
wt + P _{Me-2} /P _{Me-4}	0.006(±0.002)	+3.4(±0.1)	+3.2(±0.1)
wt + P _{Me-3} /P _{Me-4}	0.12(±0.01)	+1.6(±0.1)	+1.3(±0.1)
wt + P _{Me-3} /P _{Me-5}	1.3(±0.2)	+0.2(±.2)	+0.5(±0.1)
wt + P _{Me-4} /P _{Me-5}	0.53(±0.05)	+0.7(±0.1)	+1.4 (±0.1)
wt + P _{Me-3} /P _{Me-4} /P _{Me-5}	0.56(±0.01)	+0.7(±0.1)	+1.6 (±0.1)

^aFirst order cleavage rate constants were measured for the oligonucleotide. d(TGTGTTGTAGGATATCCTACAGGT) in 0.1 M KCl, pH 7.25 and 10 mM MgCl₂.
^bΔG^{0‡}₁ = -RTln(k_{1,modified}/k_{1,unmodified})
^cΔΔG^{0‡}_{predicted} is calculated from the addition of the ΔΔG^{0‡}_{observed} for each individual P_{Me} modification.

4.2.4 Phosphate neutralization contributes favorable ΔG^{0‡} to EcoRV-DNA transition state formation

K119A cleavage of the modified Rp-P_{Me-2} site is approximately 4-fold faster than that for the unmodified GATATC site (compare 0.56 s⁻¹ to 0.13 s⁻¹), thus the activation energy is

approximately -0.9 kcal/mol more favorable for the K119A-Rp-P_{Me-2} complex than that for the K119A-GATATC complex (Table 4.2). Similarly, R226A cleavage of the P_{Me-5} site is almost 5-fold faster than that of the unmodified site (compare 0.28 s⁻¹ to 0.06 s⁻¹), thus phosphate neutralization at P₋₅ also improves ΔG^{\ddagger} by -0.9 kcal/mol. These results suggest that reducing interphosphate repulsion on the concave face of the bent DNA substrate increases the likelihood of achieving the transition state in which the DNA must be fully bent.

It is interesting to note that the favorable effect of the Rp-P_{Me-2} substitution, in the context of the K119A mutant, is less dramatic for transition state stabilization ($\Delta\Delta G^{\ddagger} = -0.9$ kcal/mol; Table 4.2) than for ground state formation ($\Delta\Delta G^{\circ}_{\text{bind}} -2.3$ kcal/mol; Table 2.3). Because most of the cost of DNA bending has already been paid in forming the exquisitely complementary ground state complex, the additional (smaller) benefit from phosphate neutralization in the transition state may derive from either further adjustments to optimize the complementarity between the bent DNA and protein in the transition state complex and/or shifting the conformational ensemble toward the bent state optimal for catalysis.

The chiral Sp-P_{Me-2} substitution reduces the R221A cleavage rate 230-fold relative to that for an unmodified DNA substrate, resulting in an unfavorable contribution to $\Delta\Delta G^{\ddagger}$ of $+3.4$ kcal/mol (Table 4.2). In addition, the Sp-P_{Me-2} results in a huge penalty for binding ($\Delta\Delta G^{\circ}_{\text{bind}} = +6.3 (\pm 0.1)$; Table 2.3), and shows reduced DNA bending relative to the unmodified complex as measured by FRET (39° vs. 50.8°). This observation supports the model that disruption of the interactions to the P₋₂ pro-R phosphoryl oxygen by T111:O γ 1 and T187 (mediated by water) shifts the ground state complex off of the path to the transition state and into an adaptive binding mode.

The racemic P_{Me+3} substitution reduces the cleavage rate for the R140A mutant ($\Delta\Delta G^{\ddagger} = +2.7(\pm 0.1)$; Table 4.3). It is possible that this modification relieves interphosphate repulsion on the convex face of the bent DNA, which is predicted to stabilize the transition state complex. Alternatively, this modification may also perturb additional protein-phosphate interactions that stabilize the transition state.

Table 4.3 Effects of phosphate neutralization on wt and mutant EcoRV cleavage rates.

Enzyme/Sequence ^a	$k_{1,\text{cleave}} \text{ (s}^{-1}\text{)}^b$	$\Delta\Delta G_1^{\ddagger}$ (kcal/mol) ^c	$k_{2,\text{cleave}} \text{ (s}^{-1}\text{)}$	$\Delta\Delta G_2^{\ddagger}$ (kcal/mol)
Wt-GATATC	1.8(± 0.1)	0	1.8(± 0.1)	0
CONCAVE				
R226A + GATATC	0.06(± 0.01)	+2.0(± 0.1)	0.06(± 0.01)	+2.0(± 0.1)
R226A + P _{Me-5}	0.28(± 0.5)	+1.1(± 0.2)	0.27(± 0.4)	+1.1(± 0.2)
K119A + GATATC	0.13(± 0.01)	+1.6(± 0.1)	0.11(± 0.02)	+1.7(± 0.1)
K119A + Rp-P _{Me-2}	0.56(± 0.03)	+0.7(± 0.1)	0.61(± 0.05)	+0.6(± 0.2)
R221A + GATATC	1.4(± 0.8)	+0.2(± 0.2)	1.3(± 0.7)	+0.2(± 0.2)
R221A + Sp-P _{Me-2}	0.006(± 0.001)	+3.4(± 0.1)	0.006(± 0.001)	+3.4(± 0.1)
CONVEX				
R140A + GATATC	0.37(± 0.1)	+0.9(± 0.2)	0.32(0.1)	+1.0(± 0.2)
R140A + P _{Me+3}	0.02(± 0.001)	+2.7(± 0.1)	0.02(± 0.01)	+2.7(± 0.1)

^aFirst order cleavage rate constants were measured for the oligonucleotide d(TGTGTTGTAGGATATCCTACAGGT) in 0.1 M KCl, pH 7.25 and 10 mM MgCl₂.
^b $k_{1,\text{cleave}}$ and $k_{2,\text{cleave}}$ denote first order cleavage rate constants for the first strand cleavage of the top and bottom strand of the substrate, respectively.
^c $\Delta\Delta G_1^{\ddagger} = -RT\ln(k_{1,\text{modified}}/k_{1,\text{unmodified}})$

4.2.5 Effects of multiple P_{Me} modifications on EcoRV-DNA transition state formation.

Multiple P_{Me} modifications show both additive and synergistic effects on wt cleavage rates (Table 5). Specifically, the P_{Me-4}/P_{Me-5} and P_{Me-3}/P_{Me-4}/P_{Me-5} substrates show $\Delta\Delta G^{\ddagger}$ values that are less unfavorable than those expected for the sum of the $\Delta\Delta G^{\ddagger}$ values for the individual modifications ($\Delta\Delta G^{\ddagger}_{\text{predicted}}$; Table 4.2). Another way to view these data is in light of the effect of the P_{Me-5} substitution which “rescues” the wt cleavage rate for the P_{Me-4} and P_{Me-3}/P_{Me-4} substrates by 2 and 5-fold respectively (Table 4.2). These data further support a favorable role for phosphate neutralization at P₅ in stabilizing the EcoRV-DNA transition state, which is likely due to the relaxation of interphosphate repulsive forces. In contrast, all other substrates containing multiple P_{Me} substitutions show penalties for $\Delta\Delta G^{\ddagger}$ that are equal to that expected for the sum of the individual substitutions (Table 4.2). Interestingly, the multiple P_{Me} substituted substrates that show synergistic effects for binding (Table 2.5) are not the same as those that show synergistic effects for cleavage, suggesting that phosphate neutralization may contribute differently to binding and cleavage. The physical basis for this effect is not immediately apparent.

4.2.6 The effects of phosphate neutralization on the efficiency of the EcoRV restriction endonuclease.

An informative parameter in assessing the effects of phosphate neutralization on the function of the EcoRV restriction endonuclease is the difference in transition state interaction free energy ($\Delta\Delta G^{\ddagger}_1$). This parameter is related to the probability of going from the free protein and DNA

molecules to the activated EcoRV-DNA complex in which first strand cleavage occurs (Figure 1.6) such that,

$$\Delta\Delta G_{I}^{\ddagger} = \Delta\Delta G_{\text{bind}}^{\circ} + \Delta\Delta G^{\circ\ddagger},$$

where, as defined previously, $\Delta\Delta G_{\text{bind}}^{\circ}$ is the difference in binding free energy and $\Delta\Delta G^{\circ\ddagger}$ is the difference in transition state free energy associated with a particular EcoRV-DNA complex. Because $\Delta\Delta G_{I}^{\ddagger}$ takes into account both binding and cleavage it represents a measure of the overall efficiency of a particular EcoRV protein (wt or mutant) in the hydrolysis of a particular DNA substrate (unmodified, or P_{Me} substituted), analogous to the $k_{\text{cat}}/K_{\text{M}}$ parameter as measured by steady-state kinetic analysis (Lesser et al., 1990; Wolfenden and Snider, 2001). Table 4.4 shows that all four mutant complexes result in penalties in $\Delta\Delta G_{I}^{\ddagger}$. For K119A, R226A and R140A this penalty results from significant reduction in both binding affinity and cleavage (Figure 4.1 panel A) and ranges from +5.1 to +7.5 kcal/mol (Table 4.4). The penalty in $\Delta\Delta G_{I}^{\ddagger}$ for the R221A mutation is less severe and largely reflects the reduction in binding for this modification ($\Delta\Delta G_{I}^{\ddagger} = +3.3$ kcal/mol; Table 4.4; Figure 4.2 panel C). Phosphate neutralization via the Rp-P_{Me-2} and P_{Me-5} substitutions “rescues” the penalty in $\Delta\Delta G_{I}^{\ddagger}$ for the K119A and R226A mutants by contributing -3.2 and -1.8 kcal/mol to $\Delta\Delta G_{I}^{\ddagger}$ respectively (Table 4.4). These observations suggest that phosphate neutralization contributes significantly to the overall efficiency of the EcoRV restriction endonuclease in facilitating hydrolysis of a specific substrate, and the magnitude of this effect can be more fully appreciated when the effects of phosphate neutralization on both binding and cleavage are considered.

Table 4.4 Effects of phosphate neutralization on $\Delta\Delta G_1^{\ddagger}$

Enzyme/Sequence^a	$\Delta\Delta G_{\text{bind}}^{\circ}$ (kcal/mol)^b	$\Delta\Delta G^{\circ\ddagger}$ (kcal/mol)^c	$\Delta\Delta G_1^{\circ\ddagger}$ (kcal/mol)^d
wt-GATATC	0	0	1.8(±0.1)
CONCAVE			
R226A + GATATC	+5.5(±0.1)	+2.0(±0.1)	+7.5(±0.1)
R226A + P _{Me-5}	+4.6(±0.4)	+1.1(±0.2)	+5.7(±0.4)
K119A + GATATC	+5.2(±0.1)	+1.6(±0.1)	+6.8(±0.1)
K119A + Rp-P _{Me-2}	+2.9(±0.2)	+0.7(±0.1)	+3.6(±0.2)
R221A + GATATC	+3.1(±0.1)	+0.2(±0.2)	+3.3(±0.2)
R221A + Sp-P _{Me-2}	+6.3(±0.1)	+3.4(±0.1)	+9.7(±0.1)
CONVEX			
R140A + GATATC	+4.2(±0.2)	+0.9(±0.2)	+5.1(±0.3)
R140A + P _{Me+3}	+5.5(±0.4)	+2.7(±0.1)	+8.2(±0.4)

^aBoth binding and cleavage were measured for the oligonucleotide d(TGTGTTGTAGGATATCCTACAGGT). Binding was measured in 0.24M KCl, pH 7.25, and 10mM CaCl₂ while first order cleavage rates were measured in 0.1 M KCl, pH 7.25 and 10 mM MgCl₂.

^bRelative to the unmodified complex; Eqn. 2.5.

^cRelative to the unmodified complex; Eqn. 4.2.

^dRelative to the unmodified complex; Eqn. 4.3

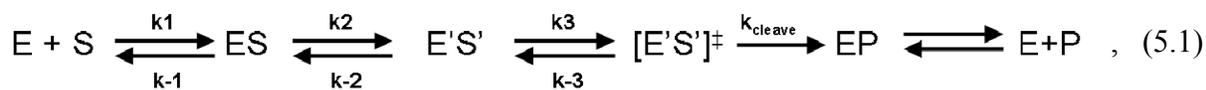
4.2.7 Conclusion

The data presented in this chapter show that phosphate neutralization at particular positions contributes favorable free energy to the EcoRV-DNA transition state formation. In particular, phosphate neutralization at the P₋₂ and P₋₅ positions increases the rates of cleavage for the K119A and R226A mutant enzymes, respectively. An additional role for phosphate neutralization at the P₋₅ phosphate is inferred from the effects of the P_{Me-5} substitution in partially “rescuing” (i.e. increasing) the reduced rate of wt cleavage for the P_{Me-4} and P_{Me-3}/P_{Me-4} substrates. I propose that the favorable effect of phosphate neutralization on $\Delta\Delta G^{\ddagger}$ derives from the relaxation of interphosphate repulsion in the EcoRV-DNA transition state which results in either further adjustments to optimize the complementarity between the bent DNA and protein in the transition state complex and/or shifting the conformational ensemble toward the bent state optimal for catalysis. These data provide insight into the role of phosphate neutralization in the overall efficiency of the EcoRV restriction endonuclease which contributes -3.2 and -1.8 kcal/mol favorable $\Delta\Delta G^{\ddagger}$ to K119A and R226A activity respectively.

5.0 DISCUSSION

The work presented in this dissertation has shown that the EcoRV restriction endonuclease can reduce the large energetic cost required for dramatically bending its specific DNA site by selectively neutralizing negatively charged phosphates on the concave face of the bent DNA molecule thereby relaxing electrostatic repulsive forces on this surface. Specifically, I have shown that phosphate neutralization contributes favorable free energy to both binding and cleavage by EcoRV. Further, phosphate neutralization at multiples sites on the concave face of the bent DNA can have synergistic effects on binding cleavage. Additionally, the energetic benefits for phosphate in binding derive from a favorable enthalpic contribution to complex formation. This investigation marks the first energetic analysis that demonstrates a role for phosphate neutralization in protein-induced bending by a protein in solution.

Aside from stabilizing the EcoRV-DNA ground and transition state complexes, it is interesting to propose that phosphate neutralization may also be important for the transition from the loose, non-specific EcoRV-DNA binding mode, to the specific binding mode. Formation of the specific EcoRV-DNA interaction is thought to be described by the scheme,



where E and S represent the free protein and DNA molecules respectively. Formation of the specific protein-DNA complex, E'S', is thought to proceed through an intermediate, ES, which represents a “loose” non-specific encounter complex (Jen-Jacobson, 1997; Hiller et al., 2003). Interactions with non-specific DNA are thought to play a key role in specific-site localization by sequence-specific DNA binding proteins through a process by which the DNA binding protein diffuses by “sliding” along the linear sequence of non-specific DNA, or by three-dimensional hopping or jumping, until the specific recognition sequence is encountered (Riggs et al., 1970a; Von Hippel and Berg, 1989; Shimamoto, 1999; Gowers and Halford, 2003; Halford and Marko, 2004b). Evidence for this non-specific intermediate comes from the EcoRI system. Specifically it has been shown that the apparent rate of association for specific EcoRI-DNA complex formation ($k_a^{app} = \frac{k_1 k_2}{k_{-1}}$; assuming that $k_2 \gg k_{-1}$ and the process described by k_2 is negligible; see scheme 5.1) is strongly dependent on salt concentration (decreases with decreasing [salt]) such that the slope of this dependence ($d \log k_a^{app} / d \log [\text{salt}]$) is greater than 1 (Jen-Jacobson, 1997). This observation suggests that there is an intermediate for specific complex formation. If k_a^{app} were a true elementary second-order rate constant which describes a single step association, it would show only a weak dependence on salt concentration ($d \log k_a^{app} / d \log [\text{salt}] \approx 1.5$) due to screening of DNA phosphates by associated counterions (Record Jr et al., 1991; Jen-Jacobson, 1997). A strong dependence of k_a^{app} on salt concentration is indicative of an intermediate for specific complex formation as the rate, k_{-1} , for the dissociation step $ES \rightarrow E + S$ in scheme 5.1 is strongly salt dependent (Lohman et al., 1978; Record Jr et al., 1991; Jen-Jacobson, 1997).

As discussed in Section 1.3.3.1 the non-specific complex is thought to be stabilized primarily through Coulombic interactions between positively charged protein side-chain

residues, and the anionic phosphate backbone of the DNA. The interaction between site-specific binding proteins and non-specific DNA sites show a very strong dependence on salt concentration (Lesser et al., 1990; Ha et al., 1992; Engler et al., 1997; Engler et al., 2001). Specifically the slope of $\log K_A$ vs $\log [KCl]$ for the formation of the non-specific EcoRV-DNA complex is approximately equal to -11.3 (Engler et al., 1997) while that for the formation of the specific ground state complex is approximately -10.9 (see Section 2.2..). According to counterion condensation theory (Manning, 1977) (see Section 1.5.1.2), the number of thermodynamically bound cations released (or salt-links formed) upon protein binding can be determined from the slope of $\log K_A$ vs $\log [KCl]$ according to the equation,

$$\text{Slope} = m' \cdot \psi, \quad (5.2)$$

where m' is the number of cations released upon protein binding, and $\psi = 0.88$, referring to the number of cations thermodynamically bound to each phosphate. Using the slope for the non-specific EcoRV-DNA interactions, it is estimated that approximately 13 monovalent cations are released upon formation of the non-specific complex. Interestingly, inspection of the non-specific EcoRV-DNA crystal structure reveals that there are potentially 12 to 13 positively charged side chain residues within 7\AA of the phosphate backbone of DNA in the non-specific EcoRV-DNA complex (i.e. within the first layer of condensed counterions; see Section 1.5.1.2) (Manning, 1977; Williams and Maher, 2000). These residues are shown in Figure 5.1. The EcoRV C-terminal domain is disordered in the non-specific EcoRV-DNA complex and is missing from the X-ray crystal structure (Winkler, 1993; Perona and Martin, 1997)(Figures 5.1 and 5.2). This region contains the positively charged side chains R221 and R226 (Figure 5.2),

which are thought to make long range Coulombic interactions to the DNA phosphates in the non-specific complex, but may not facilitate the release of a thermodynamically bound cation. Both the structure of the non-specific EcoRV-DNA complex (particularly the electrostatic complementarity between the positively charged surface of the protein molecule and the negatively charged surface of the DNA) and strong salt dependence observed for the formation of this complex, supports the model for EcoRV sliding along non-specific DNA sites. This sliding thus seems to be facilitated by favorable Coulombic interactions until the specific recognition sequence is encountered and a transition to the specific binding mode is triggered.

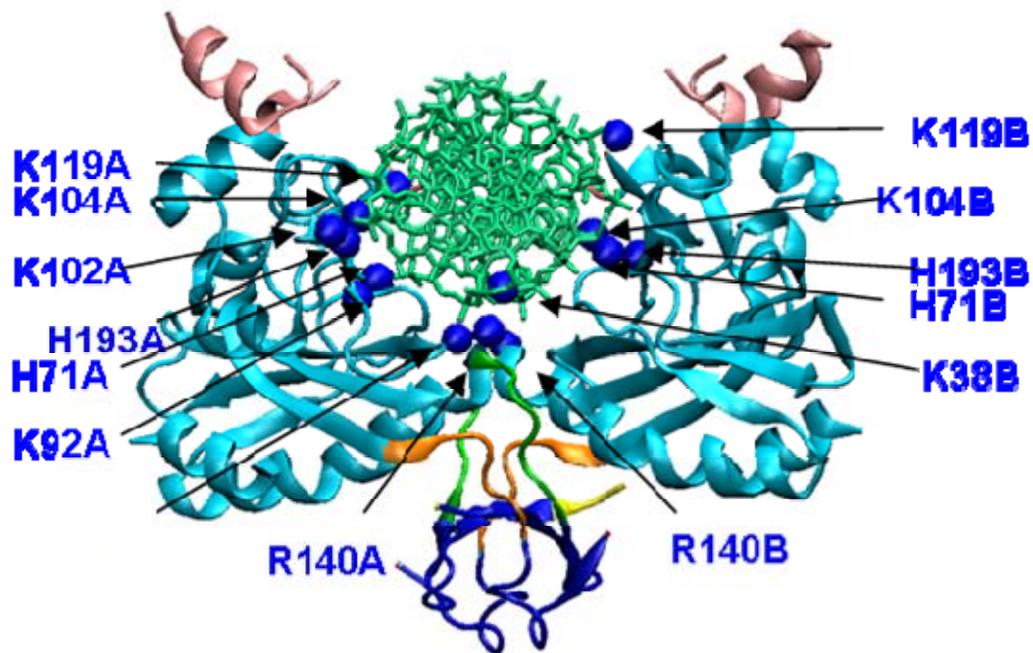
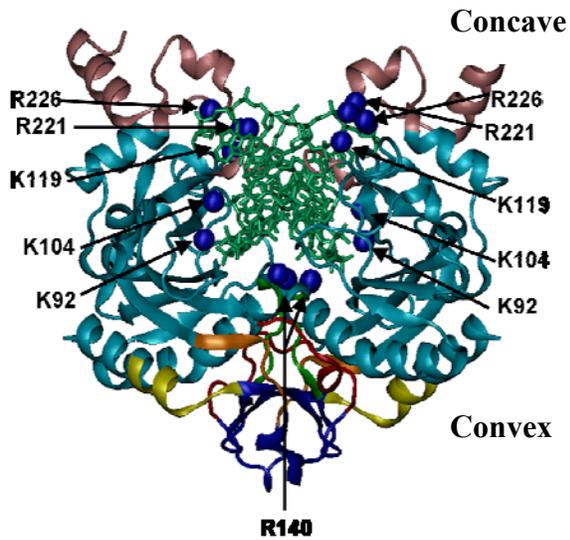


Figure 5.1 Phosphates within 7Å of the DNA phosphate backbone in the non-specific EcoRV-DNA complex
 The non-specific EcoRV-DNA complex is shown with the protein represented as multi-colored ribbons and the DNA as green licorice. Positively charged moieties of basic side chain residues (blue spheres) within 6 Å of the DNA backbone are also shown and labeled. A and B distinctions following the residue number indicate with which protein subunit each side chain is associated.

How is this transition from the non-specific to the specific binding mode triggered? This question can be viewed in a temporal sense by considering that site specific binding proteins bind with similar affinities to various non-specific DNA sites, suggesting that these non-specific sites represent an isopotential surface for protein sliding. The specific recognition site represents a potential well at which greater binding affinity is achieved. The increased binding affinity achieved at specific sites is largely the result of a slower rate of dissociation from these sites. Thus, a site-specific binding protein can rapidly scan non-specific sites until it happens upon its specific recognition site where it will pause, at which time it will transition into the specific binding mode. Interestingly it has been observed experimentally that EcoRI will even pause at sites that differ from its specific recognition site (GAATTC) by one incorrect base pair (Jeltsch et al., 1994).

From a structural perspective, the transition from the non-specific to the specific binding mode requires the unique positioning of recognition elements that are required for the formation of a tight, properly assembled, specific protein-DNA interface. For EcoRV, this transition is accompanied by a dramatic conformational transition which includes subunit re-orientation, folding of the C-terminal regions of the protein, and dramatic DNA bending as deduced from inspection of the specific and non-specific EcoRV-DNA complexes (Figure 5.2; Figure 1.4). This structural rearrangement functions to align all of the appropriate recognition elements and results in the formation of specific protein-base and protein-phosphate interactions as well as the squeezing out of water molecules from the protein-DNA interface (see Section 1.3.4)

A



B

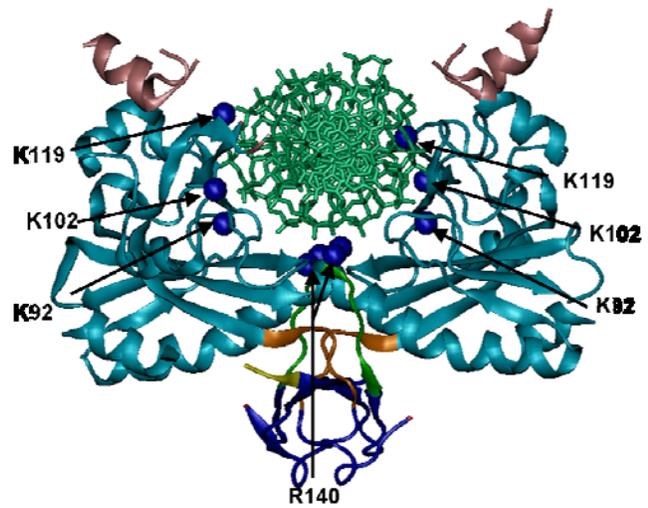


Figure 5.2 Distribution on basic side chain moieties in the specific and non-specific EcoRV-DNA complexes. Specific (A) and non-specific (B) EcoRV-DNA complexes. EcoRV (multi-colored ribbons) and DNA (green) are shown. Positively charged moieties for basic side chain residues (blue spheres) within 4 Å of the DNA backbone are also shown and labeled. DNA is bent toward the top of the page in the specific complex and the concave and convex faces are marked. There is no DNA bending in the non-specific complex

Figure 5.2 also shows that EcoRV uses the same binding cleft in order to interact with both specific (left), and non-specific (right) sites. It is currently thought that EcoRV recognition elements are “poised” to help facilitate the structural transition from the specific to non-specific complex, which includes DNA bending. Indeed it has been shown by stopped-flow FRET and stopped-flow fluorescence anisotropy measurements, that the conformational transition that accompanies the transition to the specific binding mode occurs either simultaneously, or much faster than the rate of bimolecular association (Hiller et al., 2003). This was deduced from the fact that the observed rate of EcoRV-DNA association, as measured by stopped-flow anisotropy experiments (a measure of the rotational correlation time associated with a DNA conjugated fluorophore; anisotropy increases upon EcoRV binding) and by stopped-flow FRET experiments was equal to the a rate expected from a diffusion controlled process suggesting that the overall k_{on} includes both the initial bimolecular association and bending (Hiller et al., 2003). Further, the rates of association measured by both techniques were found to be identical. From these observations it was inferred that bending must be faster than the initial bimolecular association event (Hiller et al., 2003). In addition, the increase in fluorescence anisotropy and FRET both fit well to a single exponential fit, yielding identical first-order rate constants (Hiller et al., 2003) suggesting that bending (and the associated conformational transition to the specific binding mode) occurs at a much faster rate than the initial bimolecular association event, otherwise the observed kinetic behavior would be expected to be more complex (Hiller et al., 2003). One might be able to understand which elements of the EcoRV-DNA interaction are important for DNA bending, and consequently, the transition between the non-specific and specific binding modes using the kinetic studies mentioned above. For example, a role in the transition from the non-specific to the specific binding mode would be implied if a certain perturbation to the

EcoRV-DNA interaction resulted in a reduced rate for this transition, which would manifest as a reduced rate of DNA bending relative to the unmodified interaction.

Investigation of the crystal structures of the specific and non-specific EcoRV-DNA complexes may shed light on the elements of the EcoRV-DNA interaction that may be important for the transition from the non-specific to the specific binding modes. Interestingly, the conformational transition associated with the switch from the non-specific to specific binding modes results in a differential distribution of positively charged side chains in the specific and non-specific complexes. Figure 5.2 shows the positively charged moieties that reside within 3.5Å of the phosphate backbone in both the specific and non-specific EcoRV-DNA complexes. This figure illustrates the fact that R221 and R226 are disordered in the non-specific (right) complex and become ordered upon transition to the specific complex (left) where they interact with phosphates on the concave face of the bent DNA. K119 is positioned close to the DNA backbone in both the specific and non-specific complexes, but is shifted to position more directly on the concave surface of the DNA in the specific complex. Thus, it is interesting to speculate that these side chain residues may play a role in the transition between the specific and non-specific binding modes, and that part of that role is due phosphate neutralization on the concave face of the bent DNA which reduces the energetic cost of DNA bending required for this transition.

Experiments designed to measure the rate of specific EcoRV-DNA complex formation may give insight into the role of phosphate neutralization in facilitating the transition from the specific to non-specific binding modes. Using stopped-flow FRET, the effects of phosphate neutralization on the association rate of EcoRV-DNA complexes have been analyzed (Hiller et al., 2008 unpublished). As mentioned in Section 2.2.2 this work was done in a collaboration

between our lab and that of Dr. John Perona at the University of California, Santa Barbara. The observed association rate (k_{obs}) was measured as a function of increasing concentrations of EcoRV ($[P]_{\text{tot}}$). Data obtained from these experiments were fit to the expression

$$k_{\text{obs}} = k_{\text{off}} - k_{\text{on}}[\text{DNA}]_{\text{tot}} + k_{\text{on}}[P]_{\text{tot}}, \quad (5.2)$$

where k_{on} and k_{off} represent a composite of elementary constants that includes forward and reverse rates defined in scheme 5.1 such that $k_{\text{on}} = k_1(k_2+k_{-2})/k_{-1}+k_2+k_{-2}$ and $k_{\text{off}} = k_{-1}k_2/(k_{-1}+k_2+k_{-2})$. $[\text{DNA}]_{\text{tot}}$ is the total DNA concentration and $[P]_{\text{tot}}$ is the concentration of EcoRV. Using Equation 5.2, k_{on} is equal to the slope of the dependence of k_{obs} on $[P]_{\text{tot}}$.

The observed association rate (k_{obs}) for modified and unmodified EcoRV-DNA complexes was shown to be linearly dependent on increasing EcoRV concentrations suggesting that these stopped-flow FRET experiments measure the second-order bimolecular association process which limits the rate of complex formation. If the rate of DNA bending were limiting, there would be no such dependence of k_{obs} on EcoRV concentration or linearity would break down at higher concentrations of EcoRV (Hiller et al., 2003; Sugimura and Crothers, 2006). This suggests that if there is a separable bending step, then it must be much faster than the bimolecular association event, even for modified complexes containing electrostatic perturbations to either the protein or the DNA. k_{obs} measured for modified EcoRV-DNA complexes was slower (3-70 s^{-1} ; Hiller et al. 2008) than the estimated rate of DNA bending in the unmodified EcoRV-DNA complex (100-1000 s^{-1})(Hiller et al., 2003), suggesting that the rate of DNA bending may be affected in these modified complexes, but is not observed by stopped flow FRET experiments as it did not become rate limiting.

It should be noted that the linearity of the plots of k_{obs} vs. $[\text{P}]_{\text{tot}}$ are associated with a degree of uncertainty due to a more shallow dependence of k_{obs} on $[\text{P}]_{\text{tot}}$, and large standard deviation associated with k_{obs} measured at higher $[\text{P}]_{\text{tot}}$ for several modified complexes (Hiller et al., 2008). Thus, these experiments have not clearly demonstrated that the rate limiting step has not changed (becoming bending and not bimolecular association) for each of the modified EcoRV-DNA pairs. In addition, for the formation of some complexes, the rate of FRET increase is described well by a multiple exponential fit suggesting that the rate of DNA bending may not be significantly faster than the bimolecular association rate for some EcoRV-DNA complexes. With this in mind it is interesting to note that k_{on} for the formation of the R226A-GATATC complex is 18-fold slower than that observed for the formation of the unmodified complex which may suggest a role for this side chain in facilitating DNA bending and the transition from the non-specific to the specific EcoRV-DNA binding mode. Interestingly, phosphate neutralization at the P₋₅ position *increases* k_{on} by 5-fold for the R226A mutant suggesting that if the R226A mutation does result in an impaired ability to facilitate DNA bending, then this impairment is somewhat rescued by phosphate neutralization. Thus, phosphate neutralization may play a role in increasing the rate of DNA bending by EcoRV, and may be involved in the transition between the non-specific and specific EcoRV-DNA binding modes. However, as mentioned above the experiments done in this study are unable to clearly detect a difference in the rate of DNA bending for EcoRV-DNA complexes with various electrostatic perturbations, thus the role of phosphate neutralization in the transition from the specific to non-specific binding modes for EcoRV-DNA complex formation is still not well understood. A more advanced technique with greater time resolution is required to distinguish the rates of DNA bending from the rates of bimolecular association for modified and unmodified EcoRV-DNA complexes.

One experimental technique that may be able to inform us about the role of phosphate neutralization in the transition from the non-specific to specific EcoRV-DNA binding modes is the laser induced temperature jump technique. This technique can be used to rapidly increase solution temperature which would perturb the bound EcoRV-DNA complex equilibrium and initiate EcoRV-DNA unbending and dissociation. Kinetics for the relaxation of this laser induced perturbation, can be measured by time resolved FRET techniques which would report on the rate EcoRV-induced “re-DNA bending”. Together, these technologies have been used to directly observe the rates of integration host factor (IHF) induced DNA bending, which can range from 200-4000s⁻¹ (Kuznetsov et al., 2006). These rates are in the range of those expected for EcoRV induced DNA bending upon both unmodified (100-1000s⁻¹)(Hiller et al., 2003) and modified (>3-70s⁻¹) (Hiller, et al., 2008) EcoRV-DNA complex formation. Thus these techniques may be sufficient to directly observe EcoRV-induced DNA bending rates for both unmodified EcoRV-DNA complexes and complexes which have been modified to contain electrostatic perturbations including mutations of key positively charged protein side chains (including K119, R226, and R221) and P_{Me} substituted phosphates. This information may shed light on the role of phosphate neutralization in the kinetics of DNA bending which reflect the transition between the non-specific and specific EcoRV-DNA complexes.

6.0 METHODS

6.1 SYNTHESIS AND PURIFICATION OF OLIGONUCLEOTIDE SUBSTRATES

Oligonucleotides were synthesized on the Expedite 8909 synthesizer at the University of Pittsburgh DNA synthesis facility. Methylphosphonate substituted (P_{Me}) Oligonucleotides were synthesized on Q supports (Glen Research). Methylphosphonamidites (Glen Research) were incorporated at desired positions using manufacturer's recommendations except that a coupling time of 6 minutes was used. It should be noted that dG phosphonamidite were diluted with fresh tetrahydrofuran instead of acetonitrile. Tetrahydrofuran was purchased fresh each time dG phosphonamidite was diluted. Because P_{Me} linkages are much more labile than the standard phosphodiester linkage, P_{Me} substituted Oligonucleotides needed to be cleaved from their solid supports and deprotected with more mild conditions than those used for cleavage and deprotection of standard, non substituted substrates. Thus, P_{Me} substituted oligonucleotides were cleaved from Q-supports as in (Hogrefe et al., 1993), by adding 0.8mL acetonitrile : Ethanol : ammonium hydroxide (45:45:10) to the support, and incubated at room temperature for 30 min. Following cleavage, P_{Me} substituted oligonucleotides were deprotected with an equal volume of ethylenediamine at room temperature for 6 hours. It should be noted that Due to concerns of "cross amination" dC phosphonamidites were protected with x protecting group instead of the standard x protecting group. The mixture was then brought up to 15 ml with water and the pH was adjusted to 7.0 with 6 M hydrochloric acid in acetonitrile and water (1:9, v/v). P_{Me}

substituted Oligonucleotides were then loaded onto activated C18 columns (Alltec) and washed with 3-5 volumes of Buffer A (Appendix C) followed by 3-5 column volumes of water. Oligonucleotides were then eluted from the column in 30% acetonitrile. Un substituted oligonucleotides were cleaved and deprotected by the University of Pittsburgh DNA synthesis facility.

Both unmodified and P_{Me} substituted Oligonucleotides used for equilibrium Förster Resonance Energy Transfer experiments were of the sequence d(AGAAGATATCTTGA). Fluorescein (Fl;donor) was incorporated to the 5' end of these oligonucleotides using a 5'-fluorescein phosphoramidite (Glen Research) with a coupling time of 3 min. Tetramethylrhodamine (Ta; acceptor) was incorporated to oligonucleotides post-synthetically by coupling an amine reactive TAMRA (molecular probes) to an amino modified 6 carbon linker (glen research) incorporated at the 5' end of the oligonucleotide. Substrates labeled with both donor and acceptor fluorophores will be denoted Fl-14mer-Ta while those singly labeled with donor will be denoted Fl-14mer. P_{Me} incorporation, cleavage, and deprotection of these substrates was carried out as above.

All oligos were purified by electrophoresis on 16% polyacrylamide gels (1x TBE, pH 8.3, 8.3 M Urea). Gels were run at constant current (80 milliamps) until the bromophenol blue dye-front traveled 25-30 cm. DNA was visualized by UV shadowing (Hassur and Whitlock, 1974). Gel slices containing the full-length DNA were excised from the gel and the DNA was eluted from the gel slice into a solution of buffer A (Appendix B) by shaking at 37°C for at least 15 hours. The DNA solution was again loaded onto an activated Alltech C-18 column. Columns were washed with 1 volumes of Buffer A and 2 volumes of water (relative to sample volume).

P_{Me} substituted DNA was eluted from the column with 30% acetonitrile while un-substituted DNA was eluted with 30% ethanol.

6.2 DNA CONCENTRATION DETERMINATION AND DUPLEX FORMATION

Concentrations of single-stranded oligonucleotides were determined by UV spectroscopy (260nm) on a Cary UV 100 spectrophotometer from extinction coefficients calculated by the nearest neighbor method (Warshaw and Cantor, 1970) as elaborated by Senior et al. (Senior et al., 1988). Stoichiometric amounts of complementary single strands were mixed in filter buffer (Appendix B) plus 0.24 M KCl. The strands were annealed by heating to >95°C for 15 minutes and cooling slowly overnight to 21°C. DNA was confirmed to be >98% in duplex form by non-denaturing polyacrylamide gel electrophoresis (12% polyacrylamide, 1X TBE, pH 8.3). Concentrations of the DNA duplex was determined using extinction coefficients for DNA base pairs, and compared to concentrations expected based on the previously determined concentration of the single stranded DNA. In some cases, a third method was used to check the concentration of DNA duplexes, we quantified UV visualized bands on non-denaturing gels and compared the band intensity to the intensity of bands from a concentration standard of the same length. DNA concentrations determined by the three methods were equivalent within 10%. Fluorophore labeled oligonucleotides were quantified as above, accounting for the contribution of fluorescein and TAMRA to A₂₆₀.

6.3 5' END LABELLING

Single stranded or duplex DNA was 5'-phosphorylated with γ - ^{32}P -ATP and T4 polynucleotide kinase (PNK) as previously described (Jen-Jacobson et al., 1983), with the following modification. Typically 10 picomoles of DNA were incubated with 200% excess γ - ^{32}P -ATP to maximize specific activity of the labeled DNA. In addition, polyethylene glycol (PEG 8000) was included in the labeling reaction (15% v/v) to increase labeling efficiency (Harrison et al., 1986). Reactions were terminated by adding 35 volumes of buffer A (Appendix x) and loaded onto an activated Alltech C-18 column to separate labeled DNA from unincorporated nucleotide. Purified P_{Me} labeled DNA was eluted from the column with 30% acetonitrile, while unmodified labeled DNA was eluted from the column with 30% ethanol. Both types of DNA was dried under vacuum or using a TurboVap (Zymark Inc.). Dried DNA was then dissolved in filter buffer plus 0.24 mM KCl and annealed as described above. Specific activity of labeled DNA was typically 5×10^{18} to 2×10^{19} cpm/mole.

6.4 DETERMINATION OF EQUILIBRIUM BINDING CONSTANTS

Equilibrium binding constants (K_A) were determined by one of two filter binding techniques, direct binding or equilibrium competition (Riggs et al., 1970b; Lin and Riggs, 1972) which are described in the next two sections.

6.4.1 Direct Equilibrium binding

Direct equilibrium binding assays were performed by titrating fixed concentrations of ^{32}P labeled DNA with a series of endonuclease concentrations. Reactions were carried out in binding buffer (Appendix B). Volumes of the binding reactions varied depending on the expected value of K_A ; if tight binding (large value of K_A) was expected or observed, reaction size tended to be larger to allow for a sufficient number of counts in the reaction. To ensure that binding reactions had reached equilibrium, reactions were incubated for different durations. Reactions were deemed at equilibrium if measurements at two successive times resulted in the same binding constant. Reactions were filtered (flow rate 0.5 ml/10 sec with vacuum applied from below), through a 25 mm nitrocellulose filter that had been soaked in filter buffer (FB), followed by a wash with 400 μl of FB to removed free DNA from the filter. Each filter was then placed in liquid scintillation fluid (Scintsafe 30%, Fisher Biotech, Pittsburgh, PA), and radioactivity (cpm) was measured with a liquid scintillation counter (Packard 1600 TR). The ratio of counts bound to the filter at each protein concentration to the maximum cpm (all DNA bound to protein) was then used to determine the concentration of protein-DNA complex for each binding reaction. Data were fit to a binding isotherm using the non-linear regression analysis program, Sigmaplot (Sigma Plot, Jandel Inc.; San Rafael, CA), with the following equation:

$$\frac{[ED]}{[D]_f} = \frac{K_A[E]_f}{1 + K_A[E]_f} \quad (6.1)$$

where $[D]_f$ and $[E]_f$ are the free concentrations of DNA and enzyme respectively. $[ED]$ is the concentration of the Enzyme-DNA complex, and K_A is the observed binding constant. If the

fixed concentration of DNA is less than 20% of the K_D ($K_D = 1/K_A$), and the total enzyme concentration $[E]$ is much higher than the DNA concentration $[D]$, all of the enzyme can be treated as unbound ($[E]_t$), and equation 6.1 can be simplified into:

$$\frac{[ED]}{[D]_t} = \frac{K_A[E]_t}{1 + K_A[E]_t} \quad (6.2)$$

The ratio of $[ED]/[D]_t$ is obtained as the ratio of radioactivity for any given binding reaction versus the maximum cpm at the plateau (where all DNA is bound by enzyme). Fitting this equation with Sigmaplot yields the observed equilibrium association constant, K_A .

6.4.2 Specific activity determination by converse binding

Converse binding experiments were performed in tandem with direct binding experiments to assay the specific activity of either mutant or wt EcoRV enzymes at various temperatures. For converse binding experiments, a constant amount of protein is titrated with increasing amounts of ^{32}P labeled DNA in binding buffer (Appendix B). The reaction was allowed to equilibrate for 90' and was then filtered through a nitrocellulose membrane to which the EcoRV-DNA complex bound. The amount of radiation on each filter was measured as described in Section 6.4.1 and was a metric for the amount of EcoRV-DNA complex had been retained on the filter. The amount of EcoRV-DNA complex formed for each binding reaction was fit as a function of DNA concentration by the expression.

$$[ED] = \frac{[ED]_{\max} K_A [D_t]}{1 + K_A [D_t]}, \quad (6.3)$$

where $[ED]$ is the amount of EcoRV-DNA complex formed, and $[D]_t$ is the amount of DNA in the reaction (D_t was $< 20\%$ of the K_D for the binding reaction thus $[D]_t = [D]_f$; see Section 6.4.1), $[ED]_{\max}$ is the amount of EcoRV-DNA complex achieved when all of the protein is bound by DNA, and K_A is the equilibrium binding constant.

The specific activity of a particular protein stock, at a certain environmental condition, was calculated using both direct and converse binding data. For direct binding measurements, $[ED]_{\max}$ was taken to represent the $[ED]$ at which all of the DNA is bound thus, $[ED]_{\max} = D_t$. The concentration of DNA bound at $[ED]_{\max}$ in the converse experiments can be calculated from the expression

$$\frac{CPM_{\max,direct}}{CPM_{\max,converse}} = \frac{[ED]_{\max,direct}}{[ED]_{\max,converse}}, \quad (6.4)$$

where $CPM_{\max,x}$ is the maximum radioactivity corresponding to that at plateau of the hyperbolic fit of the direct or converse binding data, and $[ED]_{\max,x}$ is the concentration of protein-DNA complex when maximum binding is achieved. For direct binding $[ED]_{\max,direct}$ is equal to $[D]_t$. The specific activity (SA) of the protein of interest at a particular environmental condition can then be calculated from the expression

$$SA = \frac{[E]_t}{[ED]_{\max,converse}}, \quad (6.5)$$

where $[E]_t$ is the total protein concentration in used in the converse experiment.

6.4.3 Equilibrium Competition

Because the retention of protein-DNA complexes on nitrocellulose filters is dependent on the dissociation rate (k_d) of the complex, equilibrium association constants for some modified sequences or solution conditions (poor flanking context, high salt, or high pH) must be determined by competition methods (Lin and Riggs, 1972) as modified by (Jen-Jacobson et al., 1986). The half-lives of complexes with these sites/under these solution conditions can be in the millisecond time range (Engler et al., 1997) such that the complexes are poorly retained on the filter. This problem is circumvented by the equilibrium competition method where the half-life of the complex formed with a radio labeled specific DNA 'reference probe' is much greater than the filtering time ($t_{1/2} \gg \gg 6$ sec.). Reactions contained the radio labeled reference probe (usually at a concentration of $6 \cdot K_D$), and the enzyme at a concentration of 5.4-fold over the K_D (K_D is the equilibrium dissociation constant of the radio labeled probe), and varying amounts of an unlabeled competitor DNA. All of the reaction components except for the enzyme were mixed; enzyme was then added, and the reaction was incubated for another 90 minutes. Ten individual reactions were used to obtain the binding constant, including one reaction without competitor, and one where the probe was 100% bound by the enzyme. Equilibrium competition was analyzed as described previously (Lin and Riggs, 1972). The concentration of the DNA probe-enzyme complex (in each reaction with different competitor concentrations) was determined by the following quadratic equation:

$$[ED_1] = \frac{-b - \sqrt{b^2 - 4ac}}{2} \quad (6.6)$$

where $[ED_1]$ is the complex between the enzyme and the radio labeled probe DNA,

$$b = -\frac{K_1 + [D_2]_t * K_1}{K_2 + [E]_t + [D_1]_t} \quad (6.7)$$

$$a = 1, \quad (6.8)$$

$$c = [D_1]_t * [E]_t, \quad (6.9)$$

and $[D_1]_t$, $[D_2]_t$, K_1 , and K_2 represent the total concentration of the probe DNA, the concentration of the competitor DNA, the binding constant of the reference probe, and the binding constant of the competitor, respectively. The non-linear regression analysis program Sigmaplot (Sigma Plot, Jandel Inc.; San Rafael, CA) was used to fit the curve generated by plotting $[ED_1]$ versus $[D_2]_t$ to obtain the best fit value for K_2 . Note that for fitting, K_1 must be known. Usually a set of direct equilibrium filter binding (Section 5.5.1) reactions were included to determine K_1 on the same day as the equilibrium competition experiments were performed.

6.5 DNA CLEAVAGE KINETICS

All cleavage rate constants were obtained under single-turnover conditions. EcoRV (1.0 μ M) was pre-equilibrated with radiolabeled DNA (0.5 μ M) in binding buffer (BB) pH7.25 plus 0.1MKCl. 10 μ l aliquots of the Pre-equilibrated reaction mixture were transferred to microfuge tubes. Cleavage reactions were initiated in each tube by the addition of the Mg₂₊ cofactor such that the final MgCl₂ concentration was 10mM unless otherwise noted. The reaction was quenched at various times by the addition of ten volumes of stop buffer (9M urea, 50 mM EDTA). Cleavage products were separated by electrophoresis on 16% polyacrylamide gels (8.3M urea, 1.5X TBE, pH 8.3); gels were then exposed to single sided emulsion Biomax at –80°C or to fujifilm imaging plates for times tailored to the number of cpm loaded per lane. Band intensities were determined with a 3XC flat-bed scanner or with a fuji scanner (Fla-5100). Best fitting values for the four cleavage rate constants were computed simultaneously using non-linear least squares fits to the observed data (% Product1, % Product2, and % uncleaved substrate at each time point), using the computer program, Scientist (Micromath Software, Salt Lake City, Utah), using the following equation (Lesser et al., 1990):

$$P1 = \frac{S_0}{2} \left[1 - e^{-(k_1+k_2)t} - \frac{k_2}{k_4 - k_1 - k_2} \left(e^{-(k_1+k_2)t} - e^{-k_4t} \right) \right] \quad (6.10)$$

$$P2 = \frac{S_0}{2} \left[1 - e^{-(k_1+k_2)t} - \frac{k_2}{k_3 - k_1 - k_2} \left(e^{-(k_1+k_2)t} - e^{-k_3t} \right) \right] \quad (6.11)$$

where S_0 is the total concentration of the DNA substrate, P1 and P2 are the two cleaved products. k_1 and k_2 are the first order rate constants for cleavage of the top and bottom strands. k_3 and k_4 denote

the subsequent rate constants for cleavage of the nicked intermediates on the top and bottom strands respectively (see Figure 4.1).

6.6 FÖRSTER RESONANCE ENERGY TRANSFER

Steady state FRET was carried out as described (Clegg et al., 1992; Hiller, 2003) with slight modifications. Steady-State fluorescence measurements were performed on a Photon Technologies International fluorimeter at 21oC. Excitation slit widths were set to 1.5nm and emission slit widths were set to 3.0nm for all emission spectra. The donor fluorophore (fluorescein (FL)) was excited at 488nm and emission spectra were collected from 500-625nm. The acceptor fluorophore (TAMRA (TA)) was directly excited at 555 nm and the emission spectra were collected from 570-650nm. The efficiency of energy transfer was extracted from the induced emission of the acceptor as in (Clegg et al., 1992; Hiller, 2003). A normalized FL-14mer scan was subtracted from the donor excited FL-14mer-Ta spectra. The area under this peak from 575-585nm was divided by area under the directly excited emission spectra, also integrated from 575-585nm. This gave the value Ratio A. Ratio A was then used to calculate the efficiency of energy transfer (E) using

$$E = [\varepsilon^A(555) / \varepsilon^D(485)][ratio_A - \varepsilon^A(485) / \varepsilon^A(555)] \quad (6.12)$$

where $\varepsilon^A(\lambda)$ and $\varepsilon^D(\lambda)$ are the extinction coefficients at wavelength λ for the donor and acceptor fluorophores respectively.

6.7 MOLECULAR DYNAMICS SIMULATIONS

Coordinates from PDB ID code 1B94 (Thomas, 1999) were used as the starting coordinates for simulations of the wild-type complex. K119 was mutated to A and the pro-s phosphoryl oxygen of P₂ was replaced by a methyl group using the Midas silicon graphics package (Reference). This was done in combination such that the coordinates for four complexes were available. These were the unmodified EcoRV-DNA complex, K119A with unmodified DNA, Wt EcoRV with a P_{Me-2} substitution, and the K119A complex with a P_{Me-2} substitution. Hydrogen atoms were built onto the protein, DNA, and crystallographic water molecules using the AMBER 8 package (Case, 2004) with the all atom force field described by Cornell et al (Cornell, 1995) and modified by (Wang, 2000). TIP3 water molecules were added to extend 10Å beyond any atoms of the EcoRV-DNA complex, creating a 514,000 Å³ octahedral water box (about 12,000 water molecules in total). 74 Na⁺ and Cl⁻ ions were added to give a salt concentration of 0.24 M, supplemented by additional Na⁺ ions to achieve electroneutrality (8-4Na⁺ depending on the complex). SHAKE (Rykeart, 1977) was used to restrain bonds involving hydrogens (tolerance 0.0001Å). Equations of motion were integrated using a 0.0015 ps time step in an isothermal-isobaric ensemble. Electrostatic interactions were treated with an Ewald sum (de Leeuw et al., 1980), and the temperature was regulated using the bath coupling method (Berendsen et al., 1984). Positions of the added salt and water atoms were refined through energy minimization followed by successive rounds of molecular dynamics to equilibrate T and V. In order to escape any local minima, we heated the system to 373K and annealed back to 300K and allowed P, V, and r to equilibrate. Production dynamics were then initiated on the Cray XT3 MPP system at the Pittsburgh Supercomputing Center. Following every 0.5ns of production dynamics the system was re-minimized prior to the next 0.5ns of production dynamics. Atomic parameters used for

methylphosphonate moieties P_{Me} are those defined by Hamelberg *et al.* (2002)(Hamelberg *et al.*, 2002).

6.8 VAN'T HOFF ANALYSIS

The effect of temperature on the equilibrium association constant was measured by both direct and competition equilibrium binding techniques (Section 5.5.1 and 5.5.2) in a temperature controlled room over a temperature range of 1 to 45°C \pm 0.2°C. Binding reactions were conducted in Binding Buffer (20 mM Cacodylate, 0.24 M KCl, 1 mM EDTA, 0.01% NaN₃, 100 μ M dithiothreitol, 100 μ g/ml bovine serum albumin, pH 7.25). The pH of the buffer was checked at each individual temperature. very few adjustments in pH were necessary as expected for a buffer with a negligible heat of ionization (Christensen *et al.*, 1976). A minimum of three measurements of the binding constant were made at each temperature. In order to confirm that changes in the binding constants with temperature were due to changes in the binding equilibrium, and not hot or cold denaturation, converse binding experiments were performed at temperatures near the hot and cold extremes (4 and 40 ° respectively). The specific activity was found to be 100% in both cases. In K_A were fit to two different models (Multiparametric fit and the Clarke and Glew fit) in order to calculate the parameters ΔH° , ΔS° , and ΔC_p° . These models are outlined in the body of the text (Sections 3.2), and were fit with Sigmaplot 10 (Sigma Plot, Jandel Inc.; San Rafael, CA).

APPENDIX A

DNA SUBSTRATES USED IN THIS WORK

Sequence ^a	Length	Studies used in
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage

5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-TGTGTTGTAGGATATCCTACAGGT ACACAACATCCTATAGGATGTCCA-5'	24 bp	Equilibrium binding, Cleavage
5'-FI-ATAGGATATCCTAA-Ta TATCCTATAGGATT-5'	14 bp	FRET
5'-FI-ATAGGATATCCTAA TATCCTATAGGATT-Ta-5'	14 bp	FRET
5'-FI-AGTAGGATATCCTAA-Ta TCATCCTATAGGATT-5'	15 bp	FRET
5'-FI-AGTAGGATATCCTAA TCATCCTATAGGATT-Ta-5'	15 bp	FRET
5'-FI-AGTAGGATATCCTAGA-Ta TCATCCTATAGGATCT-5'	16 bp	FRET
5'-FI-AGTAGGATATCCTAGA TCATCCTATAGGATCT-Ta -5'	16 bp	FRET
5'-FI-AAGTAGGATATCCTACAGAA-Ta TTCATCCTATAGGATGTGTT-5'	20 bp	FRET
5'-FI-AAGTAGGATATCCTACAGAA TTCATCCTATAGGATGTGTT-Ta-5'	20 bp	FRET
5'-FI-AAGTAGGATATCCTACAGAGA TTCATCCTATAGGATGTGTCT-Ta-5'	21 bp	FRET

5'-Fl-AAGTAGGATATCCTACAGAGA-Ta TTCATCCTATAGGATGTGTCT-5'	21 bp	FRET
5'-Fl-AAAAGATATCTTTA-Ta TTTTCTATAGAAAT-5'	14 bp	FRET
5'-Fl-AAAAGATATCTTTA TTTTCTATAGAAAT-Ta-5'	14 bp	FRET
5'-Fl-AAAAGATATCAAAA-Ta TTTTCTATAGTTTT-5'	14 bp	FRET
5'-Fl-AAAAGATATCAAAA TTTTCTATAGTTTT-Ta -5'	14 bp	FRET
5'-Fl-AAAAAGATATCAAAA-Ta TTTTTCTATAGTTTT-5'	16 bp	FRET
5'-Fl-AAAAAGATATCAAAA TTTTTCTATAGTTTT-Ta-5'	16 bp	FRET
5'-Fl-TGAAGATATCTTCT-Ta ACTTCTATAGAAGA-5'	14 bp	FRET
5'-Fl-TGAAGATATCTTCT ACTTCTATAGAAGA-Ta-5'	14 bp	FRET
5'-Fl-AGAAGATATCTTGA-Ta TCTTCTATAGAACT-5'	14 bp	FRET
5'-Fl-AGAAGATATCTTGA TCTTCTATAGAACT-Ta -5'	14 bp	FRET

^a P_{Me} substitution were made at the 5' to the bolded base; The specific EcoRV recognition site is underlined; Fl and Ta represent Fluorescein and Tamara.

APPENDIX B

BUFFER RECIPES

Buffer A: Buffer containing counter-ion used in DNA purification over Alltech C18 columns
(see methods)

- 100 mM Bis-Tris-Propane (BTP) , pH 7.5 at 21°
- 10 mM Triethylammonium acetate (TEA)
- 1 mM ethylenediaminetetraacetate (EDTA)

Binding Buffer

- 20 mM cacodylic acid, pH 7.25 at 21°C
- 10mM calcium chloride (CaCl₂)
- 0.01% sodium azide (NaN₃)
- 100 mg/ml bovine serum albumin (BSA)
- 100 μM dithiothreitol (DTT)
- Salt and pH as stated in the text

Filter Buffer

- 20 mM cacodylic acid, pH 7.25 at 21°C
- 10mM calcium chloride (CaCl₂)
- 0.01% sodium azide (NaN₃)
- Salt and pH as stated in the text

1XTBE (Tris Borate EDTA)

- 100 mM Tris
- 89 mM Boric Acid
- 2.5 mM EDTA
- pH 8.3 without titration

Enzyme Storage Buffer

- 20 mM Na₂HPO₄ pH 7.0, 21°C
- 0.3 to 0.4 M NaCl
- 1 mM EDTA
- 0.01% w/v NaN₃
- 10% v/v Glycerol

Cleavage Stop Buffer

- 50mM EDTA
- 10mM Urea

APPENDIX C

ERROR ANALYSIS

The following details the equations used to calculate the mean and standard deviation for the parameters in this dissertation.

- 1) Mean $\bar{\chi}$ and standard deviation $\bar{\sigma}\chi$ are given by the following equations:

$$\bar{\chi} = \sum_{i=1}^n x_i \quad (\text{C.1})$$

$$\bar{\sigma} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (\text{C.2})$$

- 2) $\Delta G^{\circ}_{\text{bind}}$ and its standard deviation $\sigma_{\Delta G^{\circ}_{\text{bind}}}$ are calculated by the following equations.

$$\Delta G^{\circ}_{\text{bind}} = -RT \ln \bar{K}_A \quad (\text{C.3})$$

$$\sigma_{\Delta G^o_{bind}} = -RT \left(\frac{\sigma_{K_A}}{K_A} \right) \quad (C.4)$$

Calculation of $\Delta G^{o\ddagger}$ and its standard deviation simply involves replacing \bar{K}_A with \bar{k}_{cleave} and σ_{K_A} with $\sigma_{k_{cleave}}$ in Equations C.3 and C.4 respectively.

3) $\Delta\Delta G^o_{bind}$ and its standard deviation $\sigma_{\Delta\Delta G^o_{bind}}$ are calculated by the following equations:

$$\Delta\Delta G^o_{bind} = \Delta G^o_{bind2} - \Delta G^o_{bind1} = -RT \ln \frac{\bar{K}_{A2}}{\bar{K}_{A1}} \quad (C.5)$$

$$\sigma_{\Delta\Delta G^o_{bind}} = RT \sqrt{\left(\frac{\sigma_{\bar{K}_{A1}}}{\bar{K}_{A1}} \right)^2 + \left(\frac{\sigma_{\bar{K}_{A2}}}{\bar{K}_{A2}} \right)^2} \quad (C.6)$$

4) The transition state interaction free energy, $\Delta\Delta G^o_{I\ddagger}$ and its standard deviation are calculated by the following equations:

$$\Delta\Delta G^o_{I\ddagger} = \Delta\Delta G^o_{bind} + \Delta\Delta G^o_{\ddagger} \quad (C.7)$$

$$\sigma_{\Delta\Delta G^o_{I\ddagger}} = \sqrt{\left(\sigma_{\Delta\Delta G^o_{bind}} \right)^2 + \left(\sigma_{\Delta\Delta G^o_{\ddagger}} \right)^2} \quad (C.8)$$

5) The slope (S), intercept (I), σ_{slope} , and $\sigma_{intercept}$ from linear regression analysis of the dependence of the equilibrium association constant on monovalent salt concentration are calculated with the following equations:

$$slope = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \quad (C.9)$$

$$intercept = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \quad (C.10)$$

$$\sigma_{slope} = \sqrt{\frac{nS^2}{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}}, \quad (C.11)$$

$$\sigma_{intercept} = \sqrt{\frac{S^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}}, \quad (C.12)$$

where S^2 in equation C.11 and C.12 represents the covariance given by:

$$S^2 = \frac{1}{n-2} \left[\sum_{i=1}^n \left(y_i - intercept - slope \times x_i \right)^2 \right] \quad (C.13)$$

6) Propagation of errors for ΔH°_{obs} and ΔS°_{obs} parameters obtained from van't Hoff analysis is as follows:

$$\Delta H^{\circ}_{obs} = \Delta C^{\circ}_{P,obs} (T - T_H), \quad (C.14)$$

$$\Delta S^{\circ}_{obs} = \Delta C^{\circ}_{P,obs} \ln\left(\frac{T}{T_S}\right), \quad (C.15)$$

$$\sigma_{\Delta H^{\circ}_{obs}} = \sqrt{\left[\Delta C^{\circ}_{P} \times (T - T_H)\right]^2 \times \left[\left(\frac{\sigma_{\Delta C^{\circ}_{P}}}{\Delta C^{\circ}_{P}}\right)^2 + \frac{\sigma_{(T-T_H)}^2}{(T - T_H)^2} + 2 \frac{\left[\sigma_{\Delta C^{\circ}_{P}} \times \sigma_{(T-T_H)}\right]^2}{\Delta C^{\circ}_{P} (T - T_H)} \right]} \quad (C.16)$$

$$\sigma_{\Delta S^{\circ}_{obs}} = \sqrt{\left[\Delta C^{\circ}_{P} \times \ln\left(\frac{T}{T_S}\right)\right]^2 \times \left[\left(\frac{\sigma_{\Delta C^{\circ}_{P}}}{\Delta C^{\circ}_{P}}\right)^2 + \frac{\sigma_{\ln\left(\frac{T}{T_S}\right)}^2}{\ln\left(\frac{T}{T_S}\right)^2} + 2 \frac{\left[\sigma_{\Delta C^{\circ}_{P}} \times \sigma_{\ln\left(\frac{T}{T_S}\right)}\right]^2}{\Delta C^{\circ}_{P} \times \ln\left(\frac{T}{T_S}\right)} \right]} \quad (C.17)$$

7) Propagation of errors for the parameters $\Delta G^{\circ}_{\text{obs}}$, $\Delta H^{\circ}_{\text{obs}}$, and $\Delta C^{\circ}_{P,\text{obs}}$ and $\frac{\Delta C^{\circ}_{P,\text{obs}}}{dT}$ for Clarke and Glew fit were obtained from the diagonal elements of the error matrix as described (Clarke, 1966; Bevington, 1969; Dec, 1985)

APPENDIX D

VERIFYING THE WT-GATATC FIRST ORDER CLEAVAGE RATE CONSTANT

In order to verify the relatively fast cleavage rate observed for the unmodified EcoRV-DNA complex ($1.8(\pm 0.1)\text{s}^{-1}$), I have employed two different methods for equilibrating the reaction components prior to the initiation of cleavage, and two different methods for mixing the reaction components in order to initiate and stop the cleavage reaction. First, the unmodified protein-DNA complex was pre-equilibrated in binding buffer in the absence of Mg^{2+} to form the specific recognition complex [E'S']. Mg^{2+} was then added to initiate the reaction, which was stopped at the desired time intervals, at which times cleavage products were quantified (see Section 6.6). Alternatively, EcoRV and DNA were equilibrated separately in the presence of Mg^{2+} . In this case the cleavage reaction was initiated upon addition of the DNA- Mg^{2+} mixture to an EcoRV- Mg^{2+} mixture. Identical cleavage rates were achieved independent of equilibration method. Further, in order to mix the reaction components, I used both rapid quench instrumentation, which facilitates rapid mixing and allows reaction points to be taken at 0.005 second intervals, and hand mixing techniques, for which I was able to achieve 0.3 second reaction point intervals. Both of these techniques yielded identical cleavage rate for the wt-GATATC complex ($1.8(\pm 0.1)\text{s}^{-1}$; Table 4.1). It should also be noted that rapid quench was not particularly fruitful in

measuring modified EcoRV-DNA cleavage rates which were found to be 2.5 to 3 fold slower as measured by rapid quench relative to those measured by hand mixing techniques. This is possibly due to perturbation of the stability of the complexes upon dilution and/or defective delivery/mixing of $MgCl_2$ in the rapid quench experiments, thus all rates reported for modified EcoRV-DNA complexes have been measured by hand mixing techniques for which the enzyme-DNA complex was pre-equilibrated in the absence of Mg^{2+} . This technique is particularly amenable to reactions with first order cleavage rates as fast as $1.8s^{-1}$.

BIBLIOGRAPHY

- 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 **242**(4880): 899-907.
- Aiyar, S. E., R. L. Gourse, et al. (1998). "Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase σ subunit." Proceedings of the National Academy of Sciences of the United States of America **95**(25): 14652-14657.
- Albright, R. A. and B. W. Matthews (1998). "How Cro and λ -repressor distinguish between operators: The structural basis underlying a genetic switch." Proceedings of the National Academy of Sciences of the United States of America **95**(7): 3431-3436.
- Albright, R. A., M. C. Mossing, et al. (1998). "Crystal structure of an engineered Cro monomer bound nonspecifically to DNA: Possible implications for nonspecific binding by the wild-type protein." Protein Science **7**(7): 1485-1494.
- Alves, J. (1995). "Accuracy of the EcoRV restriction endonuclease: binding and cleavage studies with oligodeoxynucleotide substrates containing degenerate recognition sequences." Biochemistry **34**(35): 11191-11197.
- Avbelj, F. and R. L. Baldwin (2004). "Origin of the neighboring residue effect on peptide backbone conformation." Proceedings of the National Academy of Sciences of the United States of America **101**(30): 10967-10972.
- Baumann, C. G., S. B. Smith, et al. (1997). "Ionic effects on the elasticity of single DNA molecules." Proceedings of the National Academy of Sciences of the United States of America **94**(12): 6185-6190.
- Berendsen, H. J. C., J. P. M. Postma, et al. (1984). "Molecular dynamics with coupling to an external bath." The Journal of Chemical Physics **81**(8): 3684-3690.
- Berger, C., I. Jelesarov, et al. (1996). "Coupled folding and site-specific binding of the GCN4-bZIP transcription factor to the AP-1 and ATF/CREB DNA sites studied by microcalorimetry." Biochemistry **35**(47): 14984-91.
- Berggren, E., L. Nordenskiöld, et al. (1992). "Interpretation of 25Mg spin relaxation in Mg-DNA solutions: Temperature variation and chemical exchange effects." Biopolymers **32**(10): 1339-1350.
- Berman, H. M., J. Westbrook, et al. (2000). "The Protein Data Bank." Nucleic Acids Research **28**(1): 235-242.
- Bevington, P. R. (1969). Data reduction and error analysis for the physical sciences. New York, McGraw-Hill.
- Bloomfield, V. A. (1997). "DNA condensation by multivalent cations." Biopolymers **44**(3): 269-282.
- Bracco, L., D. Kotlarz, et al. (1989). "Synthetic curved DNA sequences can act as transcriptional activators in Escherichia coli." EMBO Journal **8**(13): 4289-4296.

- Brennan, R. G., S. L. Roderick, et al. (1990). "Protein-DNA conformational changes in the crystal structure of a λ Cro-operator complex." Proceedings of the National Academy of Sciences of the United States of America **87**(20): 8165-8169.
- Brown, B. M. and R. T. Sauer (1993). "Assembly of the Arc repressor-operator complex: cooperative interactions between DNA-bound dimers." Biochemistry **32**(5): 1354-63.
- Bustamante, C., Z. Bryant, et al. (2003). "Ten years of tension: Single-molecule DNA mechanics." Nature **421**(6921): 423-427.
- Cairney, K. L. and R. E. Harrington (1982). "Flow birefringence of T7 phage DNA: dependence on salt concentration." Biopolymers - Peptide Science Section **21**(5): 923-934.
- Cao, D. (2002). Cosolute effects as probes of the role of water in DNA binding and catalysis by three restriction endonucleases (PhD Thesis). Biological Sciences, University of Pittsburgh: Pittsburgh, PA.
- Case, D. A., Darden, T.A., Cheatham, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Wang, B., Pearlman, D.A., et al. (2004). Amber 8. University of California, San Francisco.
- Chan, S. I., M. P. Schweizer, et al. (1964). "Interaction and association of bases and nucleosides in aqueous solutions. III. A nuclear magnetic resonance study of the self-association of purine and 6-methylpurine." Journal of the American Chemical Society **86**(19): 4182-4188.
- Christensen, J. J., L. D. Hansen, et al. (1976). Handbook of proton ionization heats and related thermodynamic quantities. New York, Wiley.
- Clarke, E. C. W., Glew, D.N. (1966). "Evaluation of thermodynamic functions from equilibrium constants." Trans. Faraday Soc. **62**: 539-547.
- Clegg, R. M., A. I. Murchie, et al. (1992). "Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction." Biochemistry **31**(20): 4846-56.
- Connelly, P. R., Ed. (1997). In: Thermodynamics and structure based drug design. R.G. Landes: Austin, TX.
- Connelly, P. R. (1997). The cost of releasing site-specific, bound water molecules from proteins: Toward a quantitative guide for structure-based drug design. Structure-based drug design. J. E. Ladbury and P. R. Connelly. Georgetown, Springer.
- Connelly, P. R., R. A. Aldape, et al. (1994). "Enthalpy of hydrogen bond formation in a protein-ligand binding reaction." Proceedings of the National Academy of Sciences of the United States of America **91**(5): 1964-1968.
- Cornell, W. D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson Jr., D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A. (1995). "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules." J. Am. Chem. Soc. **117**: 5179-5197.
- Creamer, T. P., R. Srinivasan, et al. (1995). "Modeling unfolded states of peptides and proteins." Biochemistry **34**(50): 16245-16250.
- Creamer, T. P., R. Srinivasan, et al. (1997). "Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility." Biochemistry **36**(10): 2832-2835.
- de Crombrughe, B., S. Busby, et al. (1984). "Cyclic AMP receptor protein: role in transcription activation." Science **224**(4651): 831-8.
- de Leeuw, S. W., J. W. Perram, et al. (1980). "Ewald summations and dielectric constants." Proceedings of the Royal Society of London. Series A **373**: 27-56.

- Dec, S. F., Gil, S.J. (1985). "Temperature dependence of errors in parameters derived from van't Hoff studies." J. Chem. Educ. **62**: 879-881.
- Dickerson, R. E. and T. K. Chiu (1997). "Helix Bending as a Factor in Protein/DNA Recognition." Biopolymers - Biospectroscopy Section **44**(4): 361-403.
- Dickerson, R. E. and H. R. Drew (1981). "Structure of a B-DNA dodecamer II. Influence of base sequence on helix structure." Journal of Molecular Biology **149**(4): 761-786.
- Drew, H. R. and R. E. Dickerson (1981). "Structure of a B-DNA dodecamer. III. Geometry of hydration." Journal of Molecular Biology **151**(3): 535-556.
- Duan, Y., P. Wilkosz, et al. (1996). "Dynamic contributions to the DNA binding entropy of the EcoRI and EcoRV restriction endonucleases." J Mol Biol **264**(3): 546-55.
- Dunitz, J. D. (1994). "The Entropic Cost of Bound Water in Crystals and Biomolecules." Science **264**(5159): 670.
- Dunitz, J. D. (1995). "Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions." Chem Biol **2**(11): 709-12.
- Ebright, R. H., Y. W. Ebright, et al. (1989). "Consensus DNA site for the Escherichia coli catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the E. coli lac DNA site." Nucleic Acids Res **17**(24): 10295-305.
- Ehbrecht, H. J., A. Pingoud, et al. (1985). "Linear diffusion of restriction endonucleases on DNA." Journal of Biological Chemistry **260**(10): 6160-6166.
- El Hassan, M. A. and C. R. Calladine (1997). "Conformational characteristics of DNA: Empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **355**(1722): 43-100.
- Elcock, A. H. and J. A. McCammon (1995). "Sequence dependent hydration of DNA: Theoretical results." Journal of the American Chemical Society **117**(40): 10161-10162.
- Engler, L. E. (1998). Specificity determinants in the BamHI-DNA interaction (PhD Thesis). Biological Sciences, University of Pittsburgh: Pittsburgh, PA.
- Engler, L. E., P. Sapienza, et al. (2001). "The energetics of the interaction of BamHI endonuclease with its recognition site GGATCC." Journal of Molecular Biology **307**(2): 619-636.
- Engler, L. E., K. K. Welch, et al. (1997). "Specific binding by EcoRV endonuclease to its DNA recognition site GATATC." Journal of Molecular Biology **269**(1): 82-101.
- Engler, L. E., Welch, K.K., Jen-Jacobson, L. (1997). "Specific binding by EcoRV endonuclease to its DNA recognition site GATATC." Journal of Molecular Biology **269**: 82-101.
- Finkelstein, A. V. and J. Janin (1989). "The price of lost freedom: entropy of bimolecular complex formation." Protein Engineering **3**(1): 1-3.
- Frank, D. E., R. M. Saecker, et al. (1997). "Thermodynamics of the interactions of Lac repressor with variants of the symmetric Lac operator: Effects of converting a consensus site to a non-specific site." Journal of Molecular Biology **267**(5): 1186-1206.
- Frederick, C. A., J. Grable, et al. (1984). "Kinked DNA in crystalline complex with EcoRI endonuclease." Nature **309**(5966): 327-31.
- Gartenberg, M. R. and D. M. Crothers (1991). "Synthetic DNA bending sequences increase the rate of in vitro transcription initiation at the Escherichia coli lac promoter." Journal of Molecular Biology **219**(2): 217-230.

- Gehring, W. J., Q. Van Qian, et al. (1994). "Homeodomain-DNA recognition." Cell **78**(2): 211-223.
- Goodman, S. D. and H. A. Nash (1989). "Functional replacement of a protein-induced bend in a DNA recombination site." Nature **341**(6239): 251-4.
- Goosen, N. V. d. P., Pieter (1995). "The regulation of transcription initiation by integration host factor." Molecular Microbiology **16**(1): 1-7.
- Gosule, L. C. and J. A. Schellman (1978). "DNA condensation with polyamines I. Spectroscopic studies." J Mol Biol **121**(3): 311-26.
- Gowers, D. M. and S. E. Halford (2003). "Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling." Embo J **22**(6): 1410-8.
- Grigorescu, A. (2003). Structural and Energetic Determinants of the Binding Specificity of EcoRI Endonuclease. Molecular Biophysics. Pittsburgh, PA, University of Pittsburgh.
- Gurlie, R., T. H. Duong, et al. (1999). "The role of DNA-protein salt bridges in molecular recognition: a model study." Biopolymers **49**(4): 313-27.
- Gurlie, R. and K. Zakrzewska (2001). "Protein-induced DNA bending: The role of phosphate neutralisation." Theoretical Chemistry Accounts **106**(1-2): 83-90.
- Ha, J. H., M. W. Capp, et al. (1992). "Thermodynamic stoichiometries of participation of water, cations and anions in specific and non-specific binding of lac repressor to DNA. Possible thermodynamic origins of the "glutamate effect" on protein-DNA interactions." J Mol Biol **228**(1): 252-64.
- Ha, J. H., R. S. Spolar, et al. (1989a). "Role of the hydrophobic effect in stability of site-specific protein-DNA complexes." Journal of Molecular Biology **209**(4): 801-816.
- Ha, J. H., R. S. Spolar, et al. (1989b). "Role of the hydrophobic effect in stability of site-specific protein-DNA complexes." J Mol Biol **209**(4): 801-16.
- Hagerman, K. R. and P. J. Hagerman (1996). "Helix rigidity of DNA: The meroduplex as an experimental paradigm." Journal of Molecular Biology **260**(2): 207-223.
- Hagerman, P. J. (1988). "Flexibility of DNA." Annual review of biophysics and biophysical chemistry **17**: 265-286.
- Halford, S. E. and N. P. Johnson (1983). "Single turnovers of the EcoRI restriction endonuclease." Biochemical Journal **211**(2): 405-415.
- Halford, S. E. and J. F. Marko (2004a). "How do site-specific DNA-binding proteins find their targets?" Nucleic Acids Research **32**(10): 3040-3052.
- Halford, S. E. and J. F. Marko (2004b). "How do site-specific DNA-binding proteins find their targets?" Nucleic Acids Res **32**(10): 3040-52.
- Hamelberg, D., L. D. Williams, et al. (2002). "Effect of a neutralized phosphate backbone on the minor groove of B-DNA: Molecular dynamics simulation studies." Nucleic Acids Research **30**(16): 3615-3623.
- Hardwidge, P. R., J. M. Zimmerman, et al. (2002). "Charge neutralization and DNA bending by the Escherichia coli catabolite activator protein." Nucleic Acids Res **30**(9): 1879-85.
- 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.
- Hassur, S. M. and H. W. Whitlock, Jr. (1974). "UV shadowing--a new and convenient method for the location of ultraviolet-absorbing species in polyacrylamide gels." Anal Biochem **59**(1): 162-4.
- Haynes, M., R. A. Garrett, et al. (1970). "Structure of nucleic acid-poly base complexes." Biochemistry **9**(22): 4410-6.

- Heitman, J., N. D. Zinder, et al. (1989). "Repair of the Escherichia coli chromosome after in vivo scission by the EcoRI endonuclease." Proceedings of the National Academy of Sciences of the United States of America **86**(7): 2281-2285.
- Hiller, D. A., J. M. Fogg, et al. (2003). "Simultaneous DNA Binding and Bending by EcoRV Endonuclease Observed by Real-Time Fluorescence." Biochemistry **42**(49): 14375-14385.
- Hiller, D. A., Fogg, J.M., Martin, A.M., Beecham, J.M., Reich, N.O., Perona, J.J. (2003). "Simultaneous DNA binding and bending by EcoRV endonuclease observed by real-time fluorescence." Biochemistry **42**: 14375-14385.
- Hiller, D. A., S. P. Hancock, et al. (2008). Asymmetric Phosphate Neutralization Assists DNA Binding and Bending by EcoRV Restriction Endonuclease. unpublished.
- Hiller, D. A., A. M. Rodriguez, et al. (2005). "Non-cognate enzyme-DNA complex: Structural and kinetic analysis of EcoRV endonuclease bound to the EcoRI recognition site GAATTC." Journal of Molecular Biology **354**(1): 121-136.
- Hisatake, K., S. Hasegawa, et al. (1993). "The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle regulatory protein CCG1." Nature **362**(6416): 179-181.
- Hogan, M., J. LeGrange, et al. (1983). "Dependence of DNA helix flexibility on base composition." Nature **304**(5928): 752-754.
- Hogrefe, R. I., M. M. Vaghefi, et al. (1993). "Deprotection of methylphosphonate oligonucleotides using a novel one-pot procedure." Nucleic Acids Res **21**(9): 2031-8.
- Horton, N. C. and J. J. Perona (1998a). "Recognition of flanking DNA sequences by EcoRV endonuclease involves alternative patterns of water-mediated contacts." Journal of Biological Chemistry **273**(34): 21721-21729.
- Horton, N. C. and J. J. Perona (1998b). "Role of protein-induced bending in the specificity of DNA recognition: Crystal structure of EcoRV endonuclease complexed with d(AAAGAT) + d(ATCTT)." Journal of Molecular Biology **277**(4): 779-787.
- Horton, N. C. and J. J. Perona (2000). "Crystallographic snapshots along a protein-induced DNA-bending pathway." Proceedings of the National Academy of Sciences of the United States of America **97**(11): 5729-5734.
- Horton, N. C. and J. J. Perona (2004). "DNA cleavage by EcoRV endonuclease: Two metal ions in three metal ion binding sites." Biochemistry **43**(22): 6841-6857.
- Hubbard, S. J., S. F. Campbell, et al. (1991). "Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors." J Mol Biol **220**(2): 507-30.
- Hyre, D. E. and R. E. Klevit (1998). "A disorder-to-order transition coupled to DNA binding in the essential zinc-finger DNA-binding domain of yeast ADR1." Journal of Molecular Biology **279**(4): 929-943.
- Hyre, D. E. and L. D. Spicer (1995). "Thermodynamic evaluation of binding interactions in the methionine repressor system of Escherichia coli using isothermal titration calorimetry." Biochemistry **34**(10): 3212-3221.
- Jack, W. E., B. J. Terry, et al. (1982). "Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence." Proceedings of the National Academy of Sciences of the United States of America **79**(13): 4010-4014.
- Jeltsch, A., J. Alves, et al. (1994). "Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA." Biochemistry **33**(34): 10215-10219.

- Jen-Jacobson, L. (1995). "Structural-Perturbation approaches to thermodynamics of site-specific protein-DNA interactions." Methods in Enzymology **259**: 305-344.
- Jen-Jacobson, L. (1997). "Protein-DNA recognition complexes: Conservation of structure and binding energy in the transition state." Biopolymers **44**(2): 153-180.
- Jen-Jacobson, L., Engler, L.E., Ames, J.T., Kurpiewski, M.R., Grigorescu, A. (2000a). "Thermodynamic parameters of specific and nonspecific protein-DNA binding." Supramolecular Chemistry **12**: 143-160.
- Jen-Jacobson, L., Engler, L.E., Jacobson, L.A. (2000b). "Structural and thermodynamic strategies for site specific DNA binding proteins." Structure **8**(10): 1015-1023.
- Jin, L., J. Yang, et al. (1993). "Thermodynamics of ligand binding to trp repressor." Biochemistry **32**(28): 7302-9.
- Johnson, K. A. and S. J. Benkovic, Eds. (1990). *The Enzymes*. New York, Academic Press.
- Jones, S., P. van Heyningen, et al. (1999). "Protein-DNA interactions: A structural analysis." J Mol Biol **287**(5): 877-96.
- Kalodimos, C. G., N. Biris, et al. (2004). "Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes." Science **305**(5682): 386-389.
- Kim, J. L. and S. K. Burley (1994). "1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG." Nature Structural Biology **1**(9): 638-653.
- Kim, J. L., D. B. Nikolov, et al. (1993a). "Co-crystal structure of TBP recognizing the minor groove of a TATA element." Nature **365**(6446): 520-527.
- Kim, J. L. N., Dimitar B. Burley, Stephen K. (1993a). "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. (1993b). "Crystal structure of a yeast TBP/TATA-box complex." Nature **365**(6446): 512-520.
- Kim, Y., J. C. Grable, et al. (1990). "Refinement of Eco RI endonuclease crystal structure: A revised protein chain tracing." Science **249**(4974): 1307-1309.
- Kim, Y. G., James H.; Hahn, Steven; Sigler, Paul B. (1993b). "Crystal structure of a yeast TBP/TATA-box complex." Nature **365**(6446): 512-520.
- Koo, H. S., H. M. Wu, et al. (1986). "DNA bending at adenine · thymine tracts." Nature **320**(6062): 501-506.
- Kosikov, K. M., A. A. Gorin, et al. (2002). "Bending of DNA by asymmetric charge neutralization: All-atom energy simulations." Journal of the American Chemical Society **124**(17): 4838-4847.
- Kostrewa, D. and F. K. Winkler (1995). "Mg²⁺ binding to the active site of EcoRV endonuclease: A crystallographic study of complexes with substrate and product DNA at 2 Å resolution." Biochemistry **34**(2): 683-696.
- Kumar, S., Y. Duan, et al. (1994). "Molecular dynamics simulations suggest that the Eco RI kink is an example of molecular strain." Journal of Biomolecular Structure and Dynamics **12**(3): 487-525.
- Kurpiewski, M. R., L. E. Engler, et al. (2004). "Mechanisms of coupling between DNA recognition specificity and catalysis in EcoRI endonuclease." Structure **12**(10): 1775-88.
- Kuznetsov, S. V., S. Sugimura, et al. (2006). "Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF." Proceedings of the National Academy of Sciences of the United States of America **103**(49): 18515-18520.
- Ladbury, J. E., J. G. Wright, et al. (1994). "A thermodynamic study of the trp repressor-operator interaction." Journal of Molecular Biology **238**(5): 669-681.

- Laemmli, U. K. (1975). "Characterization of DNA condensates induced by poly(ethylene oxide) and polylysine." Proc Natl Acad Sci U S A **72**(11): 4288-92.
- Lavery, R., K. Zakrzewska, et al. (1995). "JUMNA (junction minimisation of nucleic acids)." Computer Physics Communications **91**(1-3): 135-158.
- 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.
- Lesser, D. R., M. R. Kurpiewski, et al. (1993). "Facilitated distortion of the DNA site enhances EcoRI endonuclease-DNA recognition." Proc Natl Acad Sci U S A **90**(16): 7548-52.
- Li, T., M. R. Stark, et al. (1995). "Crystal structure of the MATA1/MAT?2 homeodomain heterodimer bound to DNA." Science **270**(5234): 262-269.
- Lin, S. Y. and A. D. Riggs (1972). "Lac repressor binding to non-operator DNA: detailed studies and a comparison of equilibrium and rate competition methods." Journal of Molecular Biology **72**(3): 671-690.
- Lohman, T. M., P. L. DeHaseth, et al. (1978). "Analysis of ion concentration effects of the kinetics of protein-nucleic acid interactions. Application to lac repressor-operator interactions." Biophys Chem **8**(4): 281-94.
- Love, J. J., X. Li, et al. (1995). "Structural basis for DNA bending by the architectural transcription factor LEF-1." Nature **376**(6543): 791-795.
- Love, J. J., X. Li, et al. (2004). "The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA." Biochemistry **43**(27): 8725-8734.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J. (1997). "Crystal structure of the nucleosome core particle at 2.8Å resolution." Nature **389**(6648): 251-260.
- Luijsterburg, M. S., Noom, Maarten C., Wuite, Gijs J.L., Dame, Remus Th. (2006). "The architectural role of nucleoid associated proteins in the organization of bacterial chromatin: A molecular perspective." Journal of Structural Biology **156**: 11.
- Luisi, B. F., W. X. Xu, et al. (1991). "Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA." Nature **352**(6335): 497-505.
- Madan, B. and K. A. Sharp (2001). "Hydration heat capacity of nucleic acid constituents determined from the random network model." Biophys J **81**(4): 1881-7.
- Maher, L. J., 3rd (1998). "Mechanisms of DNA bending." Curr Opin Chem Biol **2**(6): 688-94.
- Makhatadze, G. I. and P. L. Privalov (1993). "Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration." Journal of Molecular Biology **232**(2): 639-659.
- Makhatadze, G. I. and P. L. Privalov (1995). "Energetics of protein structure." Adv Protein Chem **47**: 307-425.
- Manning, G. S. (1977). "Limiting laws and counterion condensation in polyelectrolyte solutions. IV. The approach to the limit and the extraordinary stability of the charge fraction." Biophysical Chemistry **7**(2): 95-102.
- Manning, G. S. (1978). "The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides." Q Rev Biophys **11**(2): 179-246.
- Manning, G. S., K. K. Ebraldise, et al. (1989). "An estimate of the extent of folding of nucleosomal DNA by laterally asymmetric neutralization of phosphate groups." J Biomol Struct Dyn **6**(5): 877-89.
- Merabet, E. and G. K. Ackers (1995). "Calorimetric analysis of lambda cI repressor binding to DNA operator sites." Biochemistry **34**(27): 8554-63.

- Mills, J. B. and P. J. Hagerman (2004). "Origin of the intrinsic rigidity of DNA." Nucleic Acids Res **32**(13): 4055-9.
- Mirzabekov, A. D., Rich, A. (1979). "Asymmetric lateral distribution of unshielded phosphate groups in nucleosomal DNA and its role in DNA bending." Proceedings of the National Academy of Science USA **76**(3): 1118-1121.
- Morton, C. J. and J. E. Ladbury (1996). "Water mediated protein-DNA interactions: The relationship of thermodynamics to structural detail." Protein Science **5**(10): 2115-2118.
- Mossing, M. C. and M. T. Record Jr (1985). "Thermodynamic origins of specificity in the lac repressor-operator interaction. Adaptability in the recognition of mutant operator sites." Journal of Molecular Biology **186**(2): 295-305.
- Murphy, K. P. and E. Freire (1992). "Thermodynamics of structural stability and cooperative folding behavior in proteins." Adv Protein Chem **43**: 313-61.
- Newman, M., T. Strzelecka, et al. (1995). "Structure of Bam HI endonuclease bound to DNA: Partial folding and unfolding on DNA binding." Science **269**(5224): 656-663.
- 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.
- 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., A. K. Aggarwal, et al. (1990). "Conserved residues make similar contacts in two repressor-operator complexes." Science **247**(4947): 1210-1213.
- Parkhurst, K. M., M. Brenowitz, et al. (1996). "Simultaneous binding and bending of promoter DNA by the TATA binding protein: Real time kinetic measurements." Biochemistry **35**(23): 7459-7465.
- Parry, D., S. A. Moon, et al. (2003). "DNA recognition by the EcoRV restriction endonuclease probed using base analogues." Journal of Molecular Biology **331**(5): 1005-1016.
- Pellegrini, L., S. Tan, et al. (1995). "Structure of serum response factor core bound to DNA." Nature **376**(6540): 490-498.
- Perona, J. J. and A. M. Martin (1997). "Conformational transitions and structural deformability of EcoRV endonuclease revealed by crystallographic analysis." Journal of Molecular Biology **273**(1): 207-225.
- Petri, V., M. Hsieh, et al. (1995). "Thermodynamic and kinetic characterization of the binding of the TATA binding protein to the adenovirus E4 promoter." Biochemistry **34**(31): 9977-84.
- Pio, F., R. Kodandapani, et al. (1996). "Crystal structure of pu.1 ets domain-dna complex: a new pattern for helix-turn-helix recognition." FASEB Journal **10**(6).
- Porschke, D. (1991). "Persistence length and bending dynamics of DNA from electrooptical measurements at high salt concentrations." Biophysical Chemistry **40**(2): 169-179.
- Record Jr, M. T., J. H. Ha, et al. (1991). "Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA." Methods in Enzymology **208**: 291-343.
- Record Jr, M. T., T. M. Lohman, et al. (1976). "Ion effects on ligand nucleic acid interactions." Journal of Molecular Biology **107**(2): 145-158.
- Record, M. T., Jr., W. Zhang, et al. (1998). "Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and

- interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts." Adv Protein Chem **51**: 281-353.
- Rice, P. A., S. W. Yang, et al. (1996). "Crystal structure of an IHF-DNA complex: A protein-induced DNA U-turn." Cell **87**(7): 1295-1306.
- Riggs, A. D., S. Bourgeois, et al. (1970a). "The lac repressor-operator interaction. 3. Kinetic studies." J Mol Biol **53**(3): 401-17.
- Riggs, A. D., H. Suzuki, et al. (1970b). "Lac repressor-operator interaction. I. Equilibrium studies." Journal of Molecular Biology **48**(1): 67-83.
- Rocchia, W., S. Sridharan, et al. (2002). "Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects." Journal of Computational Chemistry **23**(1): 128-137.
- Rojo, F. and M. Salas (1991). "A DNA curvature can substitute phage phi 29 regulatory protein p4 when acting as a transcriptional repressor." Embo J **10**(11): 3429-38.
- Rose, D. M., C. F. Polnaszek, et al. (1982). "25Mg-NMR investigations of the magnesium ion-DNA interaction." Biopolymers **21**(3): 653-664.
- Rosenberg, J. M. (1991). "Structure and function of restriction endonucleases." Current Opinion in Structural Biology **1**(1): 104-113.
- Ruppert, S., E. H. Wang, et al. (1993). "Cloning and expression of human TAF(II)250: A TBP-associated factor implicated in cell-cycle regulation." Nature **362**(6416): 175-179.
- Rykeart, J. P., Ciccotti, G., and Brenderson, H.J.C. (1977). "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes." Journal of Computational Physics **23**: 327-341.
- Sanghani, S. R., K. Zakrzewska, et al., Eds. (1996). Modeling DNA Bending Induced by Phosphate Neutralisation. In Biological Structure and Dynamics. New York, Adenine press.
- Sapienza, P. J. (2005). Thermodynamic, kinetic, and structural basis for the relaxed DNA sequence specificity of "promiscuous" mutant *EcoRI* endonucleases (PhD Thesis). Biological Sciences, University of Pittsburgh: Pittsburgh, PA.
- Sarai, A., J. Mazur, et al. (1988). "Origin of DNA helical structure and its sequence dependence." Biochemistry **27**(22): 8498-8502.
- Sauer, R. T., S. R. Jordan, et al. (1990). "Lambda repressor: a model system for understanding protein-DNA interactions and protein stability." Adv Protein Chem **40**: 1-61.
- Schellman, J. A. (1980). "The flexibility of DNA. I. Thermal fluctuations." Biophysical Chemistry **11**(3-4): 321-328.
- Schellman, J. A. and S. C. Harvey (1995). "Static contributions to the persistence length of DNA and dynamic contributions to DNA curvature." Biophysical Chemistry **55**(1-2): 95-114.
- Schultz, S. C., G. C. Shields, et al. (1991). "Crystal structure of a CAP-DNA complex: The DNA is bent by 90°." Science **253**(5023): 1001-1007.
- Schumacher, M. A., K. Y. Choi, et al. (1994). "Crystal structure of LacI member, PurR, bound to DNA: Minor groove binding by ? helices." Science **266**(5186): 763-770.
- Schwabe, J. W. R. (1997). "The role of water in protein-DNA interactions." Current Opinion in Structural Biology **7**(1): 126-134.
- Searle, M. S. and D. H. Williams (1993). "On the stability of nucleic acid structures in solution: Enthalpy - entropy compensations, internal rotations and reversibility." Nucleic Acids Research **21**(9): 2051-2056.

- Seeman, N. C., J. M. Rosenberg, et al. (1976). "Sequence specific recognition of double helical nucleic acids by proteins." Proceedings of the National Academy of Sciences of the United States of America **73**(3): 804-808.
- Senear, D. F. and G. K. Ackers (1990). "Proton-linked contributions to site-specific interactions of lambda cI repressor and OR." Biochemistry **29**(28): 6568-77.
- Senior, M., R. A. Jones, et al. (1988). "Influence of dangling thymidine residues on the stability and structure of two DNA duplexes." Biochemistry **27**(10): 3879-85.
- Sharp, K. A. (1995). "Polyelectrolyte electrostatics: Salt dependence, entropic, and enthalpic contributions to free energy in the nonlinear Poisson-Boltzmann model." Biopolymers **36**(2): 227-243.
- Sharp, K. A., R. A. Friedman, et al. (1995). "Salt effects on polyelectrolyte-ligand binding: Comparison of Poisson- Boltzmann, and limiting law/counterion binding models." Biopolymers **36**(2): 245-262.
- Shimamoto, N. (1999). "One-dimensional diffusion of proteins along DNA. Its biological and chemical significance revealed by single-molecule measurements." Journal of Biological Chemistry **274**(22): 15293-15296.
- Smith, T. L. and R. T. Sauer (1995). "P22 Arc repressor: role of cooperativity in repression and binding to operators with altered half-site spacing." J Mol Biol **249**(4): 729-42.
- Snyder, U. K., J. F. Thompson, et al. (1989). "Phasing of protein-induced DNA bends in a recombination complex." Nature **341**(6239): 255-7.
- Sobel, E. S. and J. A. Harpst (1991). "Effects of Na⁺ on the persistence length and excluded volume of T7 bacteriophage DNA." Biopolymers **31**(13): 1559-1564.
- Solie, T. N. and J. A. Schellman (1968). "The interaction of nucleosides in aqueous solution." Journal of Molecular Biology **33**(1): 61-77.
- Spolar, R. S., J. R. Livingstone, et al. (1992). "Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water." Biochemistry **31**(16): 3947-55.
- Spolar, R. S. and M. T. Record Jr (1994). "Coupling of local folding to site-specific binding of proteins to DNA." Science **263**(5148): 777-784.
- Spronk, C. A. E. M., A. M. J. J. Bonvin, et al. (1999). "The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator." Structure **7**(12): 1483-1492.
- Steitz, T. A. (1990). "Structural studies of protein-nucleic acid interaction: The sources of sequence-specific binding." Quarterly Reviews of Biophysics **23**(3): 205-280.
- Strahs, D. and M. Brenowitz (1994). "DNA conformational changes associated with the cooperative binding of cI-repressor of bacteriophage lambda to OR." J Mol Biol **244**(5): 494-510.
- Strauss-Soukup, J. K. and L. J. Maher, 3rd (1997). "Role of asymmetric phosphate neutralization in DNA bending by PU.1." J Biol Chem **272**(50): 31570-5.
- Strauss-Soukup, J. K., M. M. Vaghefi, et al. (1997). "Effects of neutralization pattern and stereochemistry on DNA bending by methylphosphonate substitutions." Biochemistry **36**(29): 8692-8698.
- Strauss-Soukup, J. K., Vaghefi, M.M., Hogrefe, R.I., Maher, L.J.III (1997). "Effects of neutralisation pattern and stereochemistry on DNA bending by methylphosphonate substitutions." Biochemistry **36**: 7.
- Strauss, J. K. and L. J. Maher, 3rd (1994). "DNA bending by asymmetric phosphate neutralization." Science **266**(5192): 1829-34.

- Strauss, J. K., T. P. Prakash, et al. (1996a). "DNA bending by a phantom protein." Chemistry and Biology **3**(8): 571-678.
- Strauss, J. K., C. Roberts, et al. (1996b). "DNA bending by hexamethylene-tethered ammonium ions." Proceedings of the National Academy of Sciences of the United States of America **93**(18): 9515-9520.
- Sturtevant, J. M. (1977a). "Heat capacity and entropy changes in processes involving proteins." Proc Natl Acad Sci U S A **74**(6): 2236-40.
- Sturtevant, J. M. (1977b). "Heat capacity and entropy changes in processes involving proteins." Proceedings of the National Academy of Sciences of the United States of America **74**(6): 2236-2240.
- Sugimura, S. and D. M. Crothers (2006). "Stepwise binding and bending of DNA by Escherichia coli integration host factor." Proceedings of the National Academy of Sciences of the United States of America **103**(49): 18510-18514.
- Takeda, Y., P. D. Ross, et al. (1992). "Thermodynamics of Cro protein-DNA interactions." Proc Natl Acad Sci U S A **89**(17): 8180-4.
- Takeda, Y., A. Sarai, et al. (1989). "Analysis of the sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments." Proceedings of the National Academy of Sciences of the United States of America **86**(2): 439-443.
- Tansey, W. P. and W. Herr (1997). "TAFs: Guilt by association?" Cell **88**(6): 729-732.
- Taylor, J. D., I. G. Badcoe, et al. (1991). "EcoRV restriction endonuclease binds all DNA sequences with equal affinity." Biochemistry **30**(36): 8743-8753.
- Terry, B. J., W. E. Jack, et al. (1983). "Thermodynamic parameters governing interaction of EcoRI endonuclease with specific and nonspecific DNA sequences." Journal of Biological Chemistry **258**(16): 9820-9825.
- Thielking, V., U. Selent, et al. (1992). "Mg²⁺ confers DNA binding specificity to the EcoRV restriction endonuclease." Biochemistry **31**(15): 3727-3732.
- Thomas, M. P., R. L. Brady, et al. (1999). "Structural analysis of a mutational hot-spot in the EcoRV restriction endonuclease: A catalytic role for a main chain carbonyl group." Nucleic Acids Research **27**(17): 3438-3445.
- Thomas, M. P., Brady, R.L., Halford, S.E., Sessions, R.B., Baldwin, G.S. (1999). "Structural analysis of a mutational hot-spot in the EcoRV restriction endonuclease: a catalytic role for a main chain carbonyl group." Nucleic acids research **27**(17): 3438-3445.
- Tomky, L. A., J. K. Strauss-Soukup, et al. (1998). "Effects of phosphate neutralization on the shape of the AP-1 transcription factor binding site in duplex DNA." Nucleic Acids Res **26**(10): 2298-305.
- Townson, S. A., J. C. Samuelson, et al. (2007). "BstYI Bound to Noncognate DNA Reveals a "Hemispecific" Complex: Implications for DNA Scanning." Structure **15**(4): 449-459.
- Travers, A. A. (1995). "Reading the minor groove." Nature Structural Biology **2**(8): 615-618.
- Travers, A. A. (2004). "The structural basis of DNA flexibility." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **362**(1820): 1423-1438.
- Travers, A. A. and J. M. T. Thompson (2004). "An introduction to the mechanics of DNA." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **362**(1820): 1265-1279.
- Vermote, C. L. M. and S. E. Halford (1992). "EcoRV restriction endonuclease: Communication between catalytic metal ions and DNA recognition." Biochemistry **31**(26): 6082-6089.

- Viadiu, H. and A. K. Aggarwal (2000). "Structure of BamHI bound to nonspecific DNA: A model for DNA sliding." Molecular Cell **5**(5): 889-895.
- Vipond, I. B. and S. E. Halford (1993). "Structure-function correlation for the EcoRV restriction enzyme: from non-specific binding to specific DNA cleavage." Mol Microbiol **9**(2): 225-31.
- Vipond, I. B. and S. E. Halford (1995). "Specific DNA recognition by EcoRV restriction endonuclease induced by calcium ions." Biochemistry **34**(4): 1113-9.
- Von Hippel, P. H. and O. G. Berg (1989). "Facilitated target location in biological systems." Journal of Biological Chemistry **264**(2): 675-678.
- Waldburger, C. D. and R. T. Sauer (1995). "Domains of Mnt repressor: roles in tetramer formation, protein stability, and operator DNA binding." Biochemistry **34**(40): 13109-16.
- Wang, J., Cieplak, P., Kollman, P.A. (2000). "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?" Journal of Computational Chemistry **21**: 1049-1074.
- Warshaw, M. M. and C. R. Cantor (1970). "Oligonucleotide interactions. IV. Conformational differences between deoxy- and ribodinucleoside phosphates." Biopolymers **9**(9): 1079-103.
- Weiss, M. A., T. Ellenberger, et al. (1990). "Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA." Nature **347**(6293): 575-578.
- Wemmer, D. E., K. S. Srivenugopal, et al. (1985). "Nuclear magnetic resonance studies of polyamine binding to a defined DNA sequence." Journal of Molecular Biology **185**(2): 457-459.
- Wenz, C., A. Jeltsch, et al. (1996). "Probing the indirect readout of the restriction enzyme EcoRV: Mutational analysis of contacts to the DNA backbone." Journal of Biological Chemistry **271**(10): 5565-5573.
- Werner, M. H., G. M. Clore, et al. (1995a). "The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain intercalation." Cell **83**(5): 761-771.
- Werner, M. H., J. R. Huth, et al. (1995b). "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.
- Williams, L. D. and L. J. Maher, 3rd (2000). "Electrostatic mechanisms of DNA deformation." Annu Rev Biophys Biomol Struct **29**: 497-521.
- Williams, S. L., L. K. Parkhurst, et al. (2006). "Changes in DNA bending and flexing due to tethered cations detected by fluorescence resonance energy transfer." Nucleic Acids Research **34**(3): 1028-1035.
- Winkler, F. K., Banner, D.W., et al (1993). "The crystal Structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments." The EMBO Journal **12**(5): 1781-1795.
- Wolberger, C. (1996). "Homeodomain interactions." Current Opinion in Structural Biology **6**(1): 62-68.
- Wolfenden, R. and M. J. Snider (2001). "The depth of chemical time and the power of enzymes as catalysts." Accounts of Chemical Research **34**(12): 938-945.
- Zakrzewska, K. (2003). "DNA Deformation Energetics and Protein Binding." Biopolymers **70**(3): 414-423.